



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Caracterización del mercado de la distribución minorista de gasolina**

**Libardo Acero García**

Universidad Nacional de Colombia  
Facultad de Ciencias Económicas  
Maestría en Ciencias Económicas – Profundización  
PL3AN 2938  
Bogotá D.C, Colombia  
2019



# **Caracterización del mercado de la distribución minorista de gasolina**

**Libardo Acero García**

Trabajo final de maestría presentado como requisito parcial para optar al título de:

**Magister en Ciencias Económicas**

Director:

Ph. D Gustavo Adolfo Junca

Línea de Investigación:

Análisis Económico Regional y Formación de precios

Universidad Nacional de Colombia

Facultad de Ciencias Económicas

Grupo de Investigación en Economía Regional y Urbana- GIERU

Maestría en Ciencias Económicas – Profundización

PL3AN 2938

Bogotá D.C, Colombia

2019



*"El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad"*

*Víctor Hugo (1802-1885)*

*Dedicado a:*

*Mis padres quienes con su dedicación y amor me motivaron a siempre continuar transitando el camino del conocimiento.*

*A Gloria quien con su amor y compañía fue un soporte en este trabajo.*

*A Santiago, el ejemplo muestra el camino...*



## Resumen

En este trabajo se analizan las relaciones y las variables significativas en el entorno de las estaciones de servicio que modelan el precio final, aplicando regresiones tipo panel, con una base de datos de 478,114 observaciones y 15 variables explicativas, asimismo se desarrollaron modelos de conglomerados para clasificar las estaciones de servicio.

En el estudio, se parte de un marco teórico que incluye, la teoría de mercados oligopólicos, en los cuales se modelan restricciones de oferta, considerando las capacidades instaladas en el corto plazo, así como la diferenciación de producto por efecto de las distancias de los consumidores a las empresas, caso similar al mercado de gasolina específicamente en poblaciones donde hay baja densidad de estaciones de servicio.

Se utilizaron varios elementos en el modelamiento entre los cuales se destacan el desarrollo de la matriz de distancias entre las estaciones de servicio, dicha matriz permitió incluir la variable espacial en las regresiones de datos de panel y el análisis de conglomerados para la clasificación de estaciones de servicio.

La matriz de ponderación geográfica permitió incluir el efecto de las distancias entre estaciones de servicio en los precios y cantidades vendidas, para ser incluidos en los modelos panel, se consideraron dieciocho escenarios de simulación con diferentes combinaciones de variables.

Considerando los resultados de las regresiones iniciales y observando que las banderas no son el principal factor explicativo significativo en el precio de la gasolina vendida en las estaciones de servicio, se hizo un ejercicio de análisis de conglomerados para definir grupos de precios homogéneos considerando variables adicionales a la bandera de la estación, tales como, la cantidad vendida, el PIB municipal y el número de estaciones por municipio.

La agrupación por conglomerados se desarrolló utilizando los métodos jerárquicos, dentro de los cuales se incluyeron, centroide, vecino más cercano, vecino más lejano, Ward e indicadores para la elección del número de conglomerados, tales como kl, cindex y db, entre otros y utilizando la semilla de los resultados de los métodos jerárquicos en el método

no jerárquico para formar dos grupos específicos y diferenciados en cuanto al comportamiento del precio.

Respecto al comportamiento de la elasticidad a nivel de estaciones de servicio para los departamentos, se encontró un predominante comportamiento elástico, con algunos departamentos que posiblemente por una baja oferta de estaciones y dispersión en las mismas mantienen un comportamiento inelástico.

Dentro de los resultados de los conglomerados propuestos, se destaca que los grupos determinados, son significativos frente al precio de la gasolina en las estaciones de servicio al realizar las regresiones de panel; asimismo en el 35% de los departamentos analizados, existe una relación significativa entre la demanda de gasolina motor frente al precio en las estaciones. El precio en las estaciones de servicio en el 45% de los departamentos tiene una relación significativa con el producto interno bruto per cápita municipal. El porcentaje de ventas por bandera a nivel municipal tiene una relación significativa con el precio cobrado en las estaciones en el 10% de los departamentos.

La estructuración del método propuesto podría utilizarse por entes de política pública y de regulación para diseño de nuevas políticas y señales al mercado, además en trabajos posteriores definir algunas reglas de mercado considerando los grupos determinados de estaciones de servicio.

**Palabras clave:** Gasolina motor, conglomerados, centroide, vecino más cercano, vecino más lejano, Ward, kl, cindex, db, métodos jerárquicos, métodos no jerárquicos.

## Abstract

In this paper we analyze the relationships and significant variables in the environment of service stations that model the final price, applying panel type regressions, with a database of 478,114 observations and 15 explanatory variables, as well as conglomerate models for classify service stations.

In the study, i starts from a theoretical framework that includes the theory of oligopolistic markets, in which supply constraints are modeled, considering the installed capacities in the short term, as well as the differentiation of product by effect of the distances of the consumers to companies, a case similar to the gasoline market specifically in populations where there is low density of service stations.

Several elements were used in the modeling, among which the development of the distance matrix between the service stations, the use of panel models for the econometric regressions and the analysis of conglomerates for the classification of service stations stand out.

The geographical weighting matrix allowed to include the effect of the distances between service stations in the prices and quantities sold, to be included in the panel models, ten simulation scenarios with different combinations of variables were considered.

Considering the results of the initial regressions and observing that the “flags” are not the only significant explanatory factor in the price of gasoline sold at service stations, a cluster analysis exercise was carried out to define homogeneous price groups considering additional variables to the flag of the station, such as, the quantity sold, the municipal GDP and the number of stations per municipality.

Clustering was developed using the hierarchical methods, within which were included, centroid, nearest neighbor, furthest neighbor, Ward and indicators for the choice of the number of clusters, such as kl, cindex and db, among others and using the seed of the results of the hierarchical methods in the non-hierarchical method to form two specific and differentiated groups in terms of price behavior.

Regarding the behavior of the elasticity at the level of service stations for the departments, a predominantly elastic behavior was found, with some departments possibly maintaining a non-elastic behavior due to a low supply of stations and dispersion in them.

Within the results it is highlighted that, for the conglomerates proposed when calculating the regressions of the panels, these are significant compared to the price. In 35% of the departments analyzed, there is a significant relationship between the demand for motor gasoline and the price at the stations. The price at the service stations in 45% of the

departments has a significant relationship with the gross domestic product per capita. The percentage of sales per flag at the municipal level has a significant relationship with the price charged at the stations in 10% of the departments.

The structuring of the proposed method could be used by public policy and regulatory bodies to design new policies and signals to the market, and in subsequent works to define some market rules considering the groups determined in the market.

**Keywords:** gasoline, centroid conglomerates, single linkage, complete linkage, Ward, kl, cindex, db, hierarchical methods, non-hierarchical methods

# Contenido

	<u>Pág.</u>
1 Introducción .....	17
2 Marco teórico.....	21
3 Metodología .....	49
4 Resultados.....	102
5 Conclusiones y recomendaciones.....	107
6 Bibliografía .....	113
7 Anexo 1 Solicitud de información SICOM .....	115
8 Anexo 2 Mapas de las estaciones.....	118
9 Anexo 3 Resultados clúster por departamento.....	123
10 Anexo 4 Códigos R+ y matlab .....	207
11 Anexo 5 matriz W .....	242

## Lista de figuras

	<u>Pág.</u>
Figura 1-1 Cadena de valor de combustibles líquidos .....	19
Figura 2-1 Comportamiento de la demanda en presencia de restricciones.....	26
Figura 2-2 Modelo Ciudad lineal.....	28
Figura 2-3 Decisiones de compra. (a) algunos no compran, (b) todos compran .....	29
Figura 2-4 Equilibrio de Nash en ciudad lineal y ciudad circular con 5 empresas.....	32
Figura 2-5 Equilibrio de Cournot y Stackelberg.....	35
Figura 2-6 Proceso de realización de un análisis de conglomerados.....	39
Figura 2-7 Dendograma método centroide .....	40
Figura 2-8 Dendograma método vecino más cercano .....	41
Figura 2-9 Dendograma método vecino más lejano .....	42
Figura 2-10 Dendograma método vinculación promedio.....	43
Figura 2-11 Dendograma método Ward .....	44
Figura 2-12 Proceso análisis no jerárquico.....	45
Figura 2-13 Elasticidad curva demanda .....	46
Figura 2-14 Demanda perfectamente elástica .....	47
Figura 2-15 Demanda perfectamente inelástica .....	48
Figura 3-1 variables caracterización mercado .....	49
Figura 3-2 Porcentaje de estaciones según bandera.....	56
Figura 3-3 Decil <10%.....	57
Figura 3-4 Decil > 75%.....	59
Figura 3-5 Porcentaje de estaciones por Departamento.....	62
Figura 3-6 Estaciones de servicio gasolina Bogotá, Rionegro .....	63
Figura 3-7 PIB por municipio .....	64
Figura 3-8 Estructura del precio de la gasolina motor.....	65
Figura 3-9 Precios promedio de gasolina motor .....	66
Figura 3-10 Distribución precios p por departamento .....	67
Figura 3-11 Caracterización precios por bandera .....	68
Figura 3-12 Precios por bandera .....	68
Figura 3-13 Caracterización de la variación de precios mensuales por bandera .....	70
Figura 3-14 participación consumo de gasolina motor por departamento.....	71
Figura 3-15 Cantidades por departamento .....	72
Figura 3-16 Cantidades por bandera .....	73
Figura 3-17 Caracterización demanda por bandera.....	74
Figura 3-18 Caracterización de la variación de demanda .....	76

---

Figura 3-19 Elasticidad precio -Demanda.....	77
Figura 3-20 Comportamiento de elasticidad precio de la demanda a nivel de estación de servicio para diferentes años.....	77
Figura 3-21 histograma precio demanda .....	80
Figura 3-22 Modelos ponderados con matriz W .....	84
Figura 3-23 Agrupación con método centroide .....	95
Figura 3-24 Agrupación con método Ward .....	96
Figura 3-25 Agrupación con método vecino más cercano .....	97
Figura 3-26 Agrupación con método vecino más lejano .....	97
Figura 3-27 Agrupación con método vinculación promedio .....	98
Figura 3-28 Índices de elección del número de conglomerados .....	98
Figura 3-29 Grupos de estaciones y el comportamiento del precio.....	101
Figura 5-1 Efecto de la cantidad de estaciones de servicio vs precio .....	109

## Lista de tablas

	<u>Pág.</u>
Tabla 3-1 Resumen de variables de matriz panel.....	52
Tabla 3-2 Base de datos total (N=478,114) .....	54
Tabla 3-3 Percentiles número de estaciones.....	57
Tabla 3-4 Deciles número estaciones.....	57
Tabla 3-5 Caracterización variación precios .....	69
Tabla 3-6:Caracterización variación Demandas .....	74
<b>Tabla 3-7 Comportamiento de elasticidad precio de la demanda a nivel de estación de servicio a nivel departamental .....</b>	<b>78</b>
Tabla 3-8 Modelos sin ponderación con matriz W .....	81
Tabla 3-9 Test de Hausman sin considerar ponderaciones .....	87
Tabla 3-10 Test de Hausman con variables ponderadas por matriz de distancias W .....	87
Tabla 3-11 Resultados regresiones modelos panel variables sin ponderar por distancia	89
Tabla 3-12 Regresión efectos fijos con variables ponderadas por matriz de distancias W .....	91
Tabla 3-13 Información base de datos para determinar conglomerados en el departamento de Antioquia. (n=3,770) .....	93
Tabla 3-14 Test de Hausman con grupos 1 y 2 a partir del análisis no jerárquico .....	100
Tabla 3-15 Antioquia significancia conglomerado.....	100
Tabla 4-1 Significancia conglomerados por departamento .....	105

## Lista de Símbolos y abreviaturas

<b>Símbolo</b>	<b>Término</b>
<i>COP</i>	Pesos colombianos
kBD	Kilo barriles día



# 1 Introducción

El presente estudio busca caracterizar el mercado de la gasolina motor, relevante en el ámbito económico nacional, con unas ventas de 126 kBD promedio, durante el año 2018, acorde al Sistema de Información de Combustibles líquidos - SICOM (Minminas, 2017). Esta cifra representa cerca del 37% de los combustibles líquidos comercializados en el país, además es uno de los combustibles catalogados de abastecimiento estratégico en Colombia por la Unidad de Planeación Minero Energética – UPME, en el Plan de Abastecimiento de Combustibles Líquidos (UPME, 2016).

Desde la perspectiva de producto, la gasolina se puede considerar como un producto no diferenciado, definido a partir de normas técnicas promulgadas por la Superintendencia de Industria y Comercio – SIC decreto 1074 de 2015 (SIC, 2015), el Ministerio de Minas y Energía – Minminas, en el Decreto 948 de 1995, la Resolución 898 de 1995 y la Resolución 1180 de 2006 (Minminas, 2006). Desde la perspectiva de mercado y salvo algunos casos donde las empresas buscan presentar un producto diferenciado agregando algunos aditivos e impulsar su marca o “banderas” como comúnmente se les conocen (por ejemplo: Terpel, Texaco, Puma, entre otras), en general, y como hipótesis inicial podría aproximarse el mercado de la gasolina a un oligopolio.

Otro elemento que se destaca para el análisis de la distribución de la gasolina, es la dispersión de las estaciones a lo largo de la geografía nacional. Existen actualmente departamentos como Cundinamarca donde se encuentran ubicadas 6.221 estaciones, mientras que otros departamentos como Arauca, se tienen 149 estaciones, situación que deriva a que a la fecha se presente casos extremos como poblaciones con una sola estación hasta poblaciones con 404 (Minminas, 2017). Dada la dispersión de las estaciones, el considerar la ubicación es un elemento destacado en el análisis, el cual se modelará a partir de la información que se solicite al Ministerio de Minas y Energía. Dicha información, será complementada por la información de PIB municipal, calculable a partir de la información del Departamento Nacional de Estadística -DANE.

Respecto a la competencia, se puede considerar este aspecto como un factor relevante, dado que pueden existir poblaciones con un mercado en competencia, hasta poblaciones que se asimilen a un monopolio; de tal manera que pueden presentarse modelos de diferenciación de precios basados en la localización (Varian, 2012).

Desde la perspectiva de la composición del mercado existen cinco jugadores destacados: Terpel con el 37% de las estaciones, Biomax con el 16%, Mobil con el 15%, Texaco con el

9% y Petromil con el 7%. Estos participantes cubren el 84% de las estaciones a nivel nacional, lo cual podría reforzar la hipótesis de modelar como un mercado bajo la estructura de oligopolio.

Ahora bien, la cadena de valor que entrega la gasolina a los usuarios opera con tres grandes procesos: *upstream*, *midstream* y *downstream*, desarrollados secuencialmente así:

*Upstream*: es la exploración, perforación, y extracción del petróleo, actividades desarrolladas por empresas como Ecopetrol, Exxon, BP, entre otras.

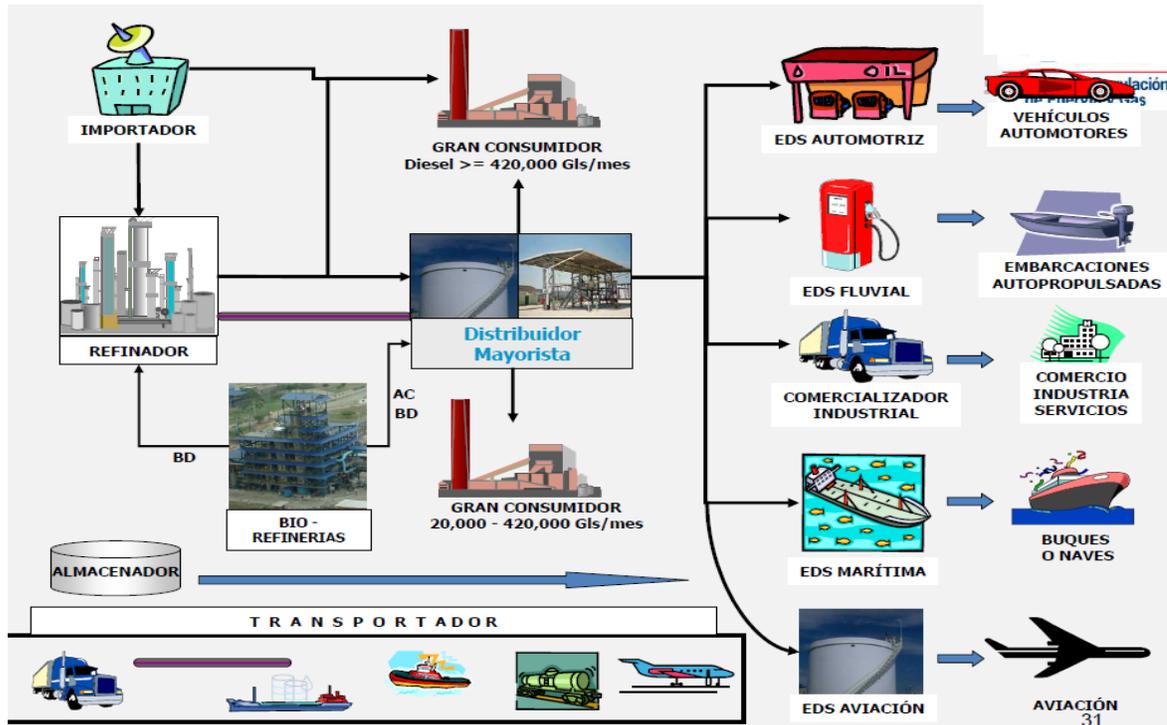
El *midstream*, cuyas actividades son el transporte y almacenamiento del petróleo. En Colombia se destaca la empresa Cenit, filial de Ecopetrol, que cuenta con la mayor red de poliductos del país.

Finalmente, el *downstream* donde se incluyen las actividades de refinación, distribución y comercialización. En este segmento se incluyen los distribuidores mayoristas, a partir de los cuales, surgen las marcas “banderas” que observan los consumidores de gasolina, cuyos comercializadores minoristas son lo que generalmente se conocen como estaciones de servicio. En la Figura 1-1 se presenta un diagrama general de la cadena de valor.

Cabe anotar, que dicha cadena de valor distribuye otros productos, clasificados como derivados y no derivados del petróleo. Entre los primeros se encuentran el ACPM, jet fuel, GLP y queroseno dentro de los segundos se encuentra el etanol.

De manera general, la regulación existente para la cadena de valor está definida así: a. El precio del suministro, es decir, de la gasolina producida o importada es regulado por el Ministerio de Minas y Energía, Decreto 4130 de 2011 (Minminas, 2011). b. El transporte de la gasolina a las plantas de los distribuidores mayoristas es regulado y definido por la Comisión de Regulación de Energía y Gas -CREG, Decreto 1260 de 2013 (Minminas, 2013). y c. El margen en las estaciones de servicio (definido como margen de comercialización minorista) es libre en varias ciudades del territorio nacional, como se presentará en el numeral 1.1.

Figura 1-1 Cadena de valor de combustibles líquidos



Fuente CREG

## 1.1 Antecedentes y justificación

En los años recientes, el petróleo ha tenido una fuerte caída de precio en el mercado internacional, entre otras causas, por la producción de petróleo de esquisto – shale gas en Estados Unidos. Muestra de esto es que el precio tuvo una variación de 130US\$/bl en 2008 a 50 US\$/bl para finales del 2018 (Investing, 2018).

Consecuencia de este comportamiento, se afecta directamente el precio de los productos refinados tales como la gasolina, debido a que Colombia tiene como referencia el precio internacional (Muñoz Morales, 2014).

Las implicaciones del precio de la gasolina para el hogar promedio, es un elemento relevante, debido a que los hogares colombianos destinan casi 3% de sus gastos mensuales para dicho rubro, de acuerdo con el ponderador de gastos básicos (DANE, 2018). Asimismo, en la canasta de productos que utiliza el DANE, que contiene 181 rubros básicos, las expensas en combustibles para el transporte son el sexto grupo, con ponderaciones cercanas a las expensas en servicios de telefonía y en energía eléctrica.

El mercado de gasolina motor incluye ciudades cuyo régimen tarifario establece libertad vigilada, es decir, el margen de comercialización minorista es libre, acorde a la Resolución 1264 del Ministerio de Minas y Energía (Minminas, 2012). Esta condición aplica para las

ciudades de Bogotá D.C., Santiago de Cali (Valle del Cauca), Barranquilla (Atlántico), Medellín (Antioquia), Montería (Córdoba), Cartagena de Indias (Bolívar), Pereira (Risaralda), Ibagué (Tolima), Bucaramanga (Santander), Sincelejo (Sucre), Manizales (Caldas), Villavicencio, (Meta), Popayán (Cauca), Santiago de Tunja (Boyacá), Barrancabermeja (Santander), Chía (Cundinamarca), Palmira (Valle del Cauca), Soacha (Cundinamarca), Tuluá (Valle del Cauca), Cartago (Valle del Cauca), Armenia (Quindío), Rionegro (Antioquia), Neiva (Huila), Santa Marta (Magdalena), Sogamoso (Boyacá).

El Ministerio de Minas y Energía basado en el Indicador Herfindahl-Hirschman (HHI), señala en la Resolución 1264 (Minminas, 2012), que se está generando una efectiva competencia en las estaciones de servicio de algunas ciudades capitales del país, sus áreas metropolitanas y en varios municipios del país, sin embargo, en la documentación soporte no se evidencia el análisis ni los valores asociados al indicador HHI. Cabe anotar, que no necesariamente este indicador recoge las variables de entorno para las estaciones a lo largo de la geografía nacional, de tal manera, que existe la oportunidad de revisar alternativas en la caracterización del mercado cuyos resultados pueden ser de utilidad en el desarrollo regulatorio futuro.

## 1.2 Pregunta de Investigación

¿Será factible caracterizar el mercado de la gasolina motor para determinar el grado de competitividad entre las estaciones de servicio, de tal manera que se puedan identificar agrupaciones de estaciones a las cuales se les puede aplicar reglas de regulación diferenciales acorde a su nivel de precios teniendo en cuenta información geográfica?

## 1.3 Objetivo general y específicos

El objetivo general es caracterizar el mercado de la gasolina a nivel departamental para determinar el grado de competitividad entre las estaciones de servicio, mediante agrupaciones por precio de estaciones de servicio, que permitan definir aspectos regulatorios y de política. Los objetivos del trabajo son:

- a) Análisis general del comportamiento del mercado de gasolina de al menos un departamento, Antioquia, a partir de la información del Sistema de Información de Combustibles – SICOM (Minminas, 2017).
- b) Caracterización de al menos un departamento, Antioquia, considerando aspectos de localización geoespacial de oferta y demanda (THISSE, 1987).
- c) Revisión de elasticidades
- d) Modelamiento incluyendo variables de ubicación geográfica (LABYS, 1986).
- e) Presentar con agrupaciones diferenciadas por precios para al menos un departamento, Antioquia.

## 2 Marco teórico

En el presente numeral se incluyen los principales elementos de la teoría económica que van a servir de herramientas para el análisis en el presente estudio, empezando por referencias a mercados en competencia, oligopolios, así como los supuestos que se harán en la depuración de información y la formación de conglomerados.

Se incluye dentro de la teoría de oligopolio, las referencias al modelo de Bertrand, incluyendo algunos ajustes en el enfoque considerando la cantidad como una variable estratégica, el modelo de Cournot, a partir del cual se revisan restricciones de capacidad y finalmente se desarrollan un par de casos donde presenta la diferenciación de producto por la distancia, considerando modelos de ciudad lineal y circular, los cuales se complementan con el desarrollo de la matriz de distancias y ponderación geográfica.

Las matrices de ubicación geográfica servirán para ponderar variables como precios y cantidades por la distancia entre las estaciones de servicio, las cuales se utilizarán de entrada para crear variables proxy en los modelos panel cuyos aspectos teóricos se incluyen en el numeral 2.4, así como para observar la significancia estadística de las variables.

Finalmente, la caracterización de las estaciones en grupos, mediante conglomerados cuya descripción teórica está incluida en el numeral 2.5, busca incluir factores tales como, la ubicación de las estaciones, el entorno económico que las rodea mediante el PIB municipal, el número de estaciones y el porcentaje de ventas, con el fin de dar una clasificación que pudiera ser utilizada dentro de las señales política que podrían afectar los precios.

### 2.1 Mercados

En el modelo competitivo, se supone que todos los consumidores y productores actúan como tomadores de precio, en efecto comportándose como si las funciones de demanda o suministro que enfrentan son infinitamente elásticas a los precios de mercado. Sin embargo, este supuesto puede no ser adecuado cuando solo hay unos pocos agentes en la oferta, ya que estos agentes a menudo poseen poder de mercado, es decir la capacidad de alterar los precios de manera rentable lejos de niveles competitivos.

El ejemplo más simple de poder de mercado surge cuando hay un solo vendedor, es decir un monopolista, si la demanda del mercado de este bien disminuye de manera continua

en función del precio, el resultado natural es que el monopolista, haciendo un pequeño aumento en su precio por encima del nivel competitivo reduce de manera poco significativa sus ventas y concluirá que vale la pena elevar su precio por encima del nivel competitivo.

A pesar de las características opuestas entre los mercados en competencia y los monopolios, comparten un aspecto en común, el cual es que no tienen que preocuparse de las repercusiones que su comportamiento puede tener sobre los demás agentes del mercado (Fernandez & Tugores, 1999).

En competencia perfecta al ser el precio un parámetro dado y al carecer las unidades productivas de cualquier poder de mercado para influir en el mismo, solo tienen que determinar la cantidad óptima a producir, con la seguridad que cuando el precio es de equilibrio podrían vender cualquier cantidad que desearan y que fuera rentable producir.

Si por otro lado la empresa es un monopolio al no tener competidores reales no tendría sentido preocuparse por los mismos.

Cuando hay más de un agente productor, lo más común es que las empresas interactúen en mercados oligopolísticos estando interrelacionados los posibles cursos de acción que puedan emprender para maximizar su utilidad.

El mercado de la gasolina en general no es un producto que puede considerarse como diferenciado salvo algunos casos donde las empresas buscan presentar un producto diferenciado agregando algunos aditivos e impulsar su marca, caso en el cual podría aproximarse al mercado de la gasolina a un oligopolio, adicional a ello dada la dispersión de las estaciones de servicio es posible que se presenten modelos de diferenciación de producto basados en la localización como se presenta en el numeral 2.2.4.1.

## **2.2 Mercados oligopólicos**

Como se presenta en el numeral 3.2.2, es posible formular como hipótesis inicial un mercado oligopólico con cinco jugadores que cuentan con más del 84% de las estaciones de combustible en el mercado nacional, cuyas zonas de influencia podrían obedecer a la estrategia de marca, abastecimiento y demanda a atender.

La interdependencia de un pequeño número de empresas en el mercado se denomina oligopolio (Varian, 2012), la cual al analizar el mercado de la gasolina se podría considerar aproximación factible. En los modelos de oligopolios existen casos especiales de equilibrio por una parte el de Cournot donde se compite con cantidades y el de Bertrand donde se compite con precios.

### **2.2.1 Modelo Bertrand: Competencia de precios**

En este modelo la variable estratégica es el precio, donde una empresa escoge los precios a los cuales va a cobrar sus productos (Mas-Colell & Whinston, 1995). Para ilustrar es

factible considerar la curva de demanda que enfrentan las empresas, en el caso restringido de dos empresas cuya función de demanda viene dada por  $x(p)$ , se asume que  $x(\cdot)$  es continua y estrictamente decreciente en todos  $p$ , de modo que  $x(p) > 0$  y que existe  $p < \infty$  tal que  $x(p) = 0$  para todo  $p \geq \bar{p}$ . Las dos empresas tienen rendimientos constantes de tecnologías de escala con el mismo costo,  $c > 0$ , por unidad producida. Asumimos que  $x(c) \in (0, \infty)$ , lo que implica que el nivel de salida socialmente óptimo (competitivo) en este mercado es estrictamente positivo y finito.

La competencia tiene lugar así: las dos empresas nombran simultáneamente sus precios  $p_1$  y  $p_2$ . Las ventas para la empresa  $j$  son entonces dadas por:

$$\text{Ecuación 1 } x_j(p_j, p_k) = \begin{cases} x_j(p_j) & \text{Si } p_j < p_k \\ \frac{1}{2} x_j(p_j, p_k) & \text{Si } p_j = p_k \\ 0 & \text{Si } p_j > p_k \end{cases}$$

Las empresas producen a pedido y, por lo tanto, incurren en costos de producción solo para un producto, nivel igual a sus ventas reales, dados los precios  $P_i$  y  $P_k$ , las ganancias de la empresa  $j$  son por lo tanto igual a  $(P_j - c) x_j(P_j; P_k)$ .

El modelo de Bertrand constituye un juego definido de movimientos simultáneos al cual en este caso, ambas firmas establecen que los precios iguales al costo:  $P_j^* = P_k^* = c$ . (Mas-Colell & Whinston, 1995).

En principio parece poco intuitivo el hecho de que el precio sea igual al coste marginal en una industria formada por dos empresas. El problema se halla, en parte, en que el juego de Bertrand es un juego que sólo se juega una vez: los jugadores eligen los precios y concluye el juego, algo que no suele ser una práctica habitual en los mercados del mundo real. (Varian, 2012).

Adicionalmente, el modelo de Bertrand predice que las distorsiones que surgen del ejercicio del poder del mercado se limitan al caso especial del monopolio.

Es factible examinar tres cambios en el modelo de Bertrand que debilitan considerablemente esta fuerte conclusión:

- 1) Hacer que la cantidad producida por las empresas sea una variable estratégica
- 2) Introducir restricciones de capacidad (o, más generalmente, rendimientos decrecientes a escala).
- 3) Permitir la diferenciación del producto.

## 2.2.2 Modelo de Cournot: Competencia de cantidades

Se podría asumir que la competencia entre las dos empresas toma otro enfoque, las dos empresas deciden simultáneamente cuánto producir,  $q_1$  y  $q_2$ . Dado estas opciones de cantidad, el precio se ajusta al nivel que atiende el mercado,  $p(q_1 + q_2)$ , donde  $p(\cdot) =$

$x - 1(\cdot)$  es la función de demanda inversa. Este modelo es conocido como el Modelo de Cournot.

Se asume que  $p(\cdot)$  es diferenciable con  $p'(q) < 0$  para todo  $q \geq 0$ , ambas empresas producen una producción a un costo de  $c > 0$  por unidad. También se asume que  $p(0) > c$  y que existe un nivel de salida único  $q^0 \in (0, \infty)$  tal que  $p(q^0) = c$ .

La Cantidad  $q^0$  es, por lo tanto, el nivel de producción socialmente óptimo (competitivo) en este mercado. Para encontrar un equilibrio de Nash (estrategia pura) de este modelo, para la firma  $j$ , la maximización del problema dado un nivel de salida  $\bar{q}_k$  de la otra empresa,  $k \neq j$ :

$$(Ecuación 2): \underset{q_j \geq 0}{Max} p(q_j + \bar{q}_k) \cdot q_j - cq_j$$

En la solución del problema (Ecuación 2), la firma  $j$  actúa exactamente como un monopolista que se enfrenta a la función inversa de demanda  $p'(q_j) = p(q_j + \bar{q}_k)$ . Una elección de cantidad óptima para la empresa  $j$  dada la producción de la empresa rival  $\bar{q}_k$ , por lo tanto, debe satisfacer la condición de primer orden.

$$(Ecuación 3) p' \cdot (q_j + \bar{q}_k) q_j + p(q_j + \bar{q}_k) \cdot \leq c \text{ con igualdad si } q_j > 0$$

Para cada  $\bar{q}_k$ , permitimos que  $b_j(\bar{q}_k)$  denota el conjunto de opciones de cantidad óptima de la empresa  $j$ ;  $b_j(\cdot)$  es la empresa  $j$  con mejor respuesta.

Un par de opciones de cantidad  $(q_1^*, q_2^*)$ , es un equilibrio de Nash si y solo si  $q_j^* \in b_j(q_k^*)$  para  $k \neq j$  y  $j = 1, 2$ . Por lo tanto, si  $(q_1^*, q_2^*)$ , es un equilibrio de Nash, estas cantidades deben satisfacer:

$$(Ecuación 4) p' \cdot (q_1^* + q_2^*) q_1^* + p(q_1^* + q_2^*) \cdot \leq c \text{ con igualdad si } q_1^* > 0$$

$$(Ecuación 5) p' \cdot (q_1^* + q_2^*) q_2^* + p(q_1^* + q_2^*) \cdot \leq c \text{ con igualdad si } q_2^* > 0$$

Se puede demostrar que bajo las suposiciones iniciales, que debemos tener  $(q_1^*, q_2^*) \gg 0$ , y así las condiciones (Ecuación 4) y la (Ecuación 5) deben mantenerse con igualdad en cualquier equilibrio de Nash. Al agregar estas dos igualdades, se debe tener:

$$(Ecuación 6) p' \cdot (q_1^* + q_2^*) \left( \frac{q_1^* + q_2^*}{2} \right) + p \cdot (q_1^* + q_2^*) = c$$

La condición (*Ecuación 6*) permite llegar a la conclusión: En cualquier equilibrio de Nash de un modelo de duopolio de Cournot con costo  $c > 0$  por unidad para las dos firmas y una función de demanda inversa  $p(\cdot)$  que satisface  $p'(q) < 0$  para todos  $q \geq 0$  y  $p(0) > c$ , el precio de mercado es mayor que  $c$  (el precio competitivo) y menor que el precio de monopolio.

En contraste con el modelo de Bertrand, el modelo de Cournot muestra una reducción del poder de mercado a medida que aumenta el número de empresas. Sin embargo, el ajuste en las cantidades producidas no necesariamente es la estrategia que escogen las empresas y prefieren ajustar los precios.

La opción al ajuste de cantidades es el modelo de Bertrand que se presenta a continuación y ofrece una interpretación alternativa del modelo de Cournot. La idea básica es que se puede considerar en las opciones de cantidad en el modelo de Cournot como opciones de capacidad a largo plazo, siendo la determinación del precio de la función de demanda inversa un proxy para el resultado de la competencia de precios a corto plazo dadas estas opciones de capacidad.

### 2.2.3 Restricciones de capacidad y rendimientos decrecientes a escala

En muchos entornos, es natural suponer que las empresas operan en condiciones que eventualmente disminuyen los rendimientos a escala, al menos en el corto plazo cuando el capital es fijo.

Un caso especial de rendimientos decrecientes ocurre cuando una empresa tiene una restricción de capacidad, eso evita que produzca más de una cantidad máxima, digamos  $\bar{q}$ . Se puede considerar, cómo afecta la introducción de restricciones de capacidad a la predicción del modelo de Bertrand presentado inicialmente.

Con restricciones de capacidad, ya no es sensato asumir que una empresa con un precio definido representa un compromiso de proporcionar cualquier cantidad demandada, ya que los costos de un pedido mayor que la capacidad instalada de la empresa son infinitos.

Por eso se hace un mínimo ajuste a las reglas del modelo de Bertrand tomando los precios para abastecer la demanda solo hasta la capacidad instalada. También se asume que las capacidades son comúnmente conocidos entre las empresas. (Mas-Colell & Whinston, 1995).

Para ver cómo las restricciones de capacidad pueden afectar el resultado de la fijación de precios del duopolio se supone que cada una de las dos empresas tiene un costo marginal constante de  $c > 0$  y una restricción de capacidad de  $\bar{q} = \frac{3}{4}x(c)$ . Como antes, la función de demanda del mercado  $x(\cdot)$  es continua, está disminuyendo estrictamente en todos  $p$ , de modo que  $x(p) > 0$ , y tiene  $x(c) > 0$ .



hacen. Así, para determinar las funciones de demanda para las empresas, se necesita establecer una regla que especifique qué consumidores le pueden comprar a la firma de bajo precio cuando la demanda exceda su capacidad.

De hecho, la elección de una regla de racionamiento puede tener efectos importantes en el comportamiento de equilibrio y existen múltiples opciones para fijar las reglas tales como:

- a) Atender a la demanda más “valiosa” como en la Figura 2-1, la cual en general produce los mejores resultados.
- b) Un sistema de asignación de unidades por agendamiento.
- c) Un sistema de asignación aleatoria de unidades, tal como puede ocurrir en el mercado de la gasolina donde los consumidores pueden comprar dependiendo su ruta de manera aleatoria.

Hasta este punto se ha asumido el nivel de capacidad de una empresa como exógena, sin embargo, normalmente las empresas eligen sus niveles de capacidad, de tal manera que, si primero eligen sus niveles de capacidad y luego compiten en precios, el único subjuego en perfecto equilibrio de Nash en este modelo de dos etapas es el resultado de Cournot. (Mas-Colell & Whinston, 1995).

Este resultado, el cálculo del precio a partir de la curva de demanda inversa en el modelo de Cournot puede ser considerado como un proxy para esta competencia de precios de segunda etapa. De hecho, para un amplio rango de opciones de capacidad  $(\bar{q}_1, \bar{q}_2)$ , el equilibrio único del subjuego de precios implica que ambas empresas establezcan sus precios iguales a  $p(\bar{q}_1 + \bar{q}_2)$ .

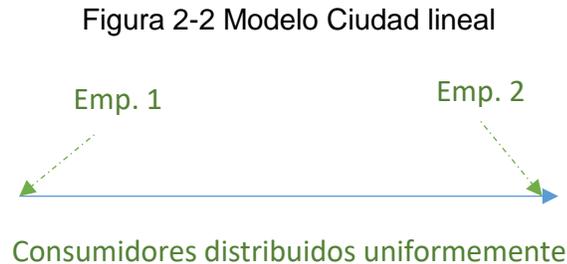
Por lo tanto, este modelo de dos etapas de la elección de capacidad y competencia de precios nos permite una reinterpretación del modelo de Cournot: es factible considerar que la cantidad de Cournot modela la competencia a largo plazo a través de la elección de capacidad y competencia con precio en el corto plazo, dados los niveles de capacidad elegidos.

## 2.2.4 Diferenciación de producto

En el modelo de Bertrand, las empresas enfrentaron una curva de demanda infinitamente elástica en equilibrio, con un diferencial de precio arbitrariamente pequeño, todos los consumidores preferirían comprar a la empresa cuyos productos se ofrecían al menor precio.

Sin embargo, en el caso de que los consumidores perciban diferencias entre los productos, cuando exista diferenciación de producto, cada empresa poseerá algún poder de mercado como resultado de la singularidad de su producto. Por ejemplo, si hay  $j > 1$  empresas,

cada una produce a un costo marginal constante de  $c > 0$ . La demanda del producto de la empresa  $j$  viene dada por la función continua  $x_j(p_j, -p_{-j})$  donde  $p_{-j}$  es un vector de precios de los rivales a la empresa  $j$ .



Fuente: (Mas-Colell & Whinston, 1995)

En una solución de precio simultáneo, cada empresa toma las opciones de precio conocido de sus rivales  $p_{-j}$  y elige  $p_j$  para resolver:

$$(Ecuación 7) \quad \underset{p_j}{Max} \quad (p_j - c) \cdot x_j(p_j - \bar{p}_{-j})$$

Tenga en cuenta que siempre que  $x_j(c, \bar{p}_{-j}) > 0$ , la mejor respuesta de la empresa  $j$  implica necesariamente un precio en exceso de sus costos ( $p_j > c$ ) porque puede asegurarse ganancias estrictamente positivas fijando su precio ligeramente por encima de  $c$ . Así, en presencia de diferenciación de producto, los precios de equilibrio estarán por encima del nivel competitivo. Como con la competencia de cantidad y limitaciones de capacidad, la presencia de diferenciación de productos suaviza fuertemente resultado competitivo del modelo de Bertrand.

Entre los modelos de diferenciación de producto se destacan el de ciudad lineal y una variación al mismo, en el modelo circular, los cuales se presentan a continuación.

#### 2.2.4.1 El modelo de ciudad lineal de diferenciación de productos.

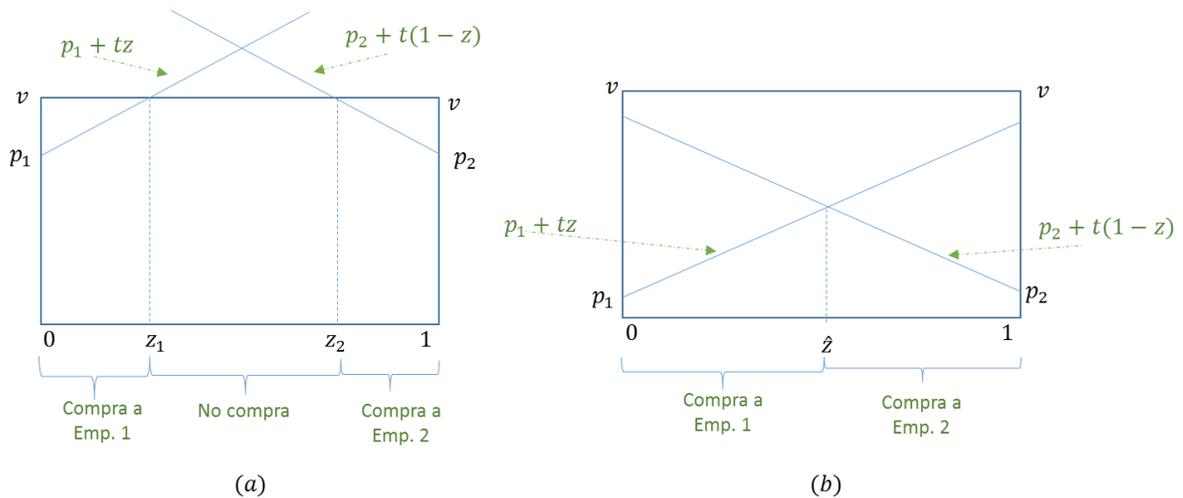
Considera una ciudad que se puede representar como tendido en un segmento de línea de longitud 1, como se muestra en Figura 2-2. Hay un continuo de consumidores cuyo número total es  $M$  y se supone que están ubicadas uniformemente a lo largo de este segmento de línea.

La ubicación de un consumidor está indexada por  $z \in [0, 1]$ , la distancia desde el extremo izquierdo de la ciudad a cada extremo de la ciudad se encuentra un proveedor de productos: La empresa 1 está en el extremo izquierdo; mientras que la empresa 2, a la derecha. Los bienes se producen a un costo unitario constante de  $c > 0$ . Cada consumidor desea como máximo 1 producto y obtiene un beneficio bruto de  $v$  de su consumo.

El costo total de compra de la empresa  $j$  para un consumidor ubicado a una distancia  $d$  desde la empresa  $j$  es  $p_j + td$ , donde  $\frac{t}{2} > 0$  puede considerarse como el costo o desutilidad por unidad de distancia recorrida por el consumidor al ir y venir desde la ubicación de la empresa  $j$ . La presencia de costos de viaje introduce la diferenciación entre los productos de las dos empresas porque los consumidores pueden preferir estrictamente comprar de una de las dos empresas incluso cuando los bienes se vendan al mismo precio.

La Figura 2-3 (a) ilustra las decisiones de compra de los consumidores ubicados en varios puntos en la ciudad, para un par dado de precios  $p_1$  y  $p_2$ .

Figura 2-3 Decisiones de compra. (a) algunos no compran, (b) todos compran



Fuente: (Mas-Colell & Whinston, 1995)

Al observar las ubicaciones es factible concluir que:

- Considerando que  $p_1 + tz < p_2 + t(1 - z)$ , los consumidores prefieren comprar a la firma 1.
- Los consumidores que están ubicados en  $[0, z_1]$  compran en la firma 1.
- Para  $v - p_1 - tz > 0$  comprar en la firma 1 es mejor que no comprar.
- En la ubicación  $z_1$ , un consumidor es indiferente entre comprando de la firma 1 y no comprar; es decir,  $z_1$  satisface  $v - p_1 - tz_1 = 0$ .
- En la Figura 2-3 (a), los consumidores en el intervalo  $(z_1, z_2)$  no compran de ninguna de las empresas, mientras que los del intervalo  $(z_2, 1)$  compran de la firma 2.
- La Figura 2-3 (b), por el contrario, representa un caso en el que, dados los precios  $p_1$  y  $p_2$ , todos los consumidores pueden obtener un excedente estrictamente

positivo al comprar el bien de una de las empresas. La ubicación del consumidor que es indiferente entre las dos empresas, es el punto  $z$  tal que:

$$(Ecuación 8) p_1 + t\hat{z} = p_2 + t(1 - z)$$

$$(Ecuación 9) \hat{z} = \frac{t+p_2-p_1}{2t}$$

En general, el análisis de este modelo es complejo debido a que dependiendo de los parámetros  $(v, c, t)$ , los equilibrios pueden involucrar áreas de mercado para las empresas que no sean alcanzables (como en la Figura 2-3 (a)), o puede hacer que las empresas luchen por los consumidores en el medio del mercado (como en la Figura 2-3 (b)).

Por simplicidad se supone que el valor del producto es mucho mayor que los costos de viaje, o que  $v > c + 3t$ . En este caso se puede demostrar que la empresa prefiere establecer su precio a un nivel que evite que algunos consumidores no compren, por lo tanto, se ignora la posibilidad de no comprar.

A partir de  $p_1$  y  $p_2$ , sea  $z$  definido como en (Ecuación 9), de tal manera que la demanda de la empresa, esta dada por un par de precios  $(p_1, p_2)$ , iguales a  $M\hat{z}$  cuando  $z \in [0,1]$ ,  $M$  cuando  $\hat{z} > 1$ , y 0 cuando  $\hat{z} < 1$ .

Sustituyendo por  $\hat{z}$  de (Ecuación 9) tenemos:

$$(Ecuación 10) x_1(p_1, p_2) = \begin{cases} 0 & \text{Si } p_1 > p_2 + t \\ \frac{M}{2t}(t + p_2 - p_1) & \text{Si } p_1 \in [p_2 - t, p_2 + t] \\ M & \text{Si } p_1 < p_2 - t \end{cases}$$

La función de demanda para la empresa 2 de manera simétrica es:

$$(Ecuación 11) x_2(p_1, p_2) = \begin{cases} 0 & \text{Si } p_2 > p_1 + t \\ \frac{M}{2t}(t + p_1 - p_2) & \text{Si } p_2 \in [p_1 - t, p_1 + t] \\ M & \text{Si } p_2 < p_1 - t \end{cases}$$

Considerando la Ecuación 10 y la Ecuación 11, que cada empresa  $j$ , en busca de su mejor respuesta a cualquier elección de precio  $\bar{p}_{-j}$  por su rival, puede restringirse a los precios en el intervalo  $[\bar{p}_{-j} - t, \bar{p}_{-j} + t]$ .

Cualquier precio  $p_j > \bar{p}_{-j} + t$  produce los mismos beneficios que el ajuste  $p_j = \bar{p}_{-j} + t$  (es decir, cero), y cualquier precio  $p_j < \bar{p}_{-j} - t$  produce ganancias más bajas que ajuste  $p_j = \bar{p}_{-j} - t$ . Así, la firma  $j$  responde a  $p_j$  solucionando la siguiente maximización

$$(Ecuación 12) \underset{p_j}{Max} (p_j - c) \cdot (t + \bar{p}_{-j} - p_j) \cdot \frac{M}{2t}$$

Sujeto a

$$(Ecuación 13) p_j \in [\bar{p}_{-j} - t, \bar{p}_{-j} + t]$$

La condición de primer orden para este problema es

$$(Ecuación 14) t + \bar{p}_{-j} + c - 2p_j = \begin{cases} \leq 0 & \text{Si } p_j > \bar{p}_{-j} - t \\ = 0 & \text{Si } p_j \in [\bar{p}_{-j} - t, \bar{p}_{-j} + t] \\ \geq 0 & \text{Si } p_j = \bar{p}_{-j} + t \end{cases}$$

Resolviendo se obtiene para cada firma la mejor función de respuesta:

$$(Ecuación 15) \bar{b}(p_{-j}) = \begin{cases} \bar{p}_{-j} + t & \text{Si } \bar{p}_{-j} \leq c - t \\ \frac{t + \bar{p}_{-j} + c}{2} & \text{Si } \bar{p}_{-j} \in [c - t, c + 3t] \\ \bar{p}_{-j} - t & \text{Si } \bar{p}_{-j} \geq c + 3t \end{cases}$$

Cuando  $\bar{p}_{-j} < c - t$ , la empresa  $j$  se valora de una manera que hace que sus ventas sean iguales a cero (no puede obtener ganancias porque no puede realizar ventas a cualquier precio por encima de  $c$ ). Cuando  $\bar{p}_{-j} > c + 3t$ , los precios de la empresa  $j$  hacen que capture todo el mercado. En el caso intermedio, la mejor respuesta de la empresa  $j$  a  $\bar{p}_{-j}$  deja a ambas empresas con niveles positivos de venta.

Dada la simetría del modelo, se busca un equilibrio simétrico, es decir, un equilibrio en el que  $p_1^* = p_2^* = p^*$ . En cualquier equilibrio simétrico,  $p^* = b(p^*)$ . Al examinar (Ecuación 15), se observa que esta condición solo puede satisfacerse en el punto medio (teniendo en cuenta que este es el único caso en el que ambas empresas pueden tener resultados estrictamente positivos de ventas, como deben ser en cualquier equilibrio simétrico). Por lo tanto,  $p^*$  debe satisfacer

$$(Ecuación 16) p^* = \frac{1}{2}(t + p^* + c)$$

Entonces:

$$(Ecuación 17) p^* = \frac{1}{2}(t + c)$$

Cada empresa tiene ventas de  $\frac{M}{2}$  y una ganancia de  $\frac{tM}{2}$ . teniendo en cuenta que a medida que  $t$  se acerca a cero, los productos de las empresas se vuelven completamente indiferenciados y los precios de equilibrio se acercan a  $c$ , como en el modelo de Bertrand.

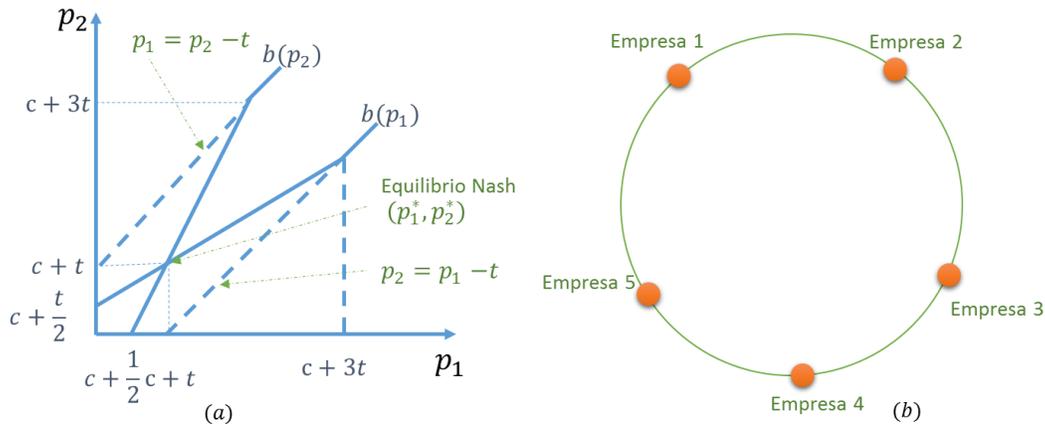
En la otra dirección, como el costo de viaje  $t$  se hace mayor, lo que aumenta la diferenciación entre los productos de las empresas, los precios de equilibrio y las ganancias aumentan. Tal como pueden aumentar en los precios en las zonas alejadas de la geografía nacional donde llevar la gasolina puede implicar métodos multimodales de transporte como el terrestre y el fluvial.

La Figura 2-4 (a) describe las funciones de mejor respuesta para las dos empresas (para precios mayor que o igual a  $c$ ) y el equilibrio de Nash. Como de costumbre, el equilibrio de Nash se encuentra en la intersección de los gráficos de estas funciones de mejor respuesta se observa además que aquí no hay equilibrios asimétricos.

Los indicadores se vuelven más complicados cuando  $v < c + 3t$  porque las empresas pueden querer establecer precios en el que algunos consumidores no quieren comprar de ninguna de las dos empresas. Se puede mostrar, sin embargo, que el equilibrio que se acaba de obtener sigue siendo válido mientras  $v \geq c + \frac{3}{2}t$ . En contraste, cuando  $v < c + t$ , en equilibrio, las áreas de mercado de las empresas no se tocan (las empresas son como "locales monopolistas ") como puede ser el caso de estaciones de servicio en poblaciones aisladas.

En el caso intermedio en el que  $v \in [c + t, c + \frac{3}{2}t]$ , las empresas se encuentran en la curva de sus funciones de demanda y el consumidor en la ubicación indiferente  $\hat{z}$  no recibe ningún excedente de su compra.

Figura 2-4 Equilibrio de Nash en ciudad lineal y ciudad circular con 5 empresas



Fuente: (Mas-Colell & Whinston, 1995)

Las características esenciales del modelo de ciudad lineal pueden extenderse al caso en el que el número de empresas sea mayor a dos ( $J > 2$ ). Al hacerlo, es conveniente para los propósitos analíticos considerar un modelo de una ciudad circular, para que las empresas puedan mantenerse en posiciones simétricas. En este modelo, los consumidores están distribuidos uniformemente a lo largo de un círculo de circunferencia

igual a 1, y las empresas se posicionan a intervalos iguales entre sí. La Figura 2-4 (b) representa un caso donde  $j = 5$ .

Los modelos de ciudad lineales y circulares se conocen como modelos espaciales de diferenciación de producto porque cada empresa se identifica con una "dirección" en el espacio de distribución del producto.

Con mayor generalidad es factible imaginar productos de empresas ubicados en algún espacio de características N-dimensionales, con las "direcciones" de los consumidores (sus puntos ideales de consumo) distribuidos en este espacio.

Los modelos espaciales comparten la característica de que cada empresa compite por clientes solo a nivel local, es decir, únicamente con las empresas que ofrecen productos similares. (Mas-Colell & Whinston, 1995)

El mercado de la gasolina, donde las estaciones de servicio están dispersas a lo largo de la geografía nacional, son un ejemplo práctico de la diferenciación de productos, entre otros aspectos por la distancia y la capacidad instalada.

Las estaciones de servicio a partir del desarrollo teórico presentado, pueden modelarse a partir de las consideraciones de modelos oligopólicos presentados hasta ahora y complementarse con modelos espaciales (Arbia, 2014), cuyos aspectos teóricos se presentan en el numeral 2.3.

### **2.2.5 Interacción repetida y costos de entrada**

En un juego infinitamente repetido, cualquier pago descontado es factible que otorgue a cada jugador, por período, más que el pago más bajo que podría garantizarse en una sola partida del juego de componentes de movimiento simultáneo, si los jugadores descuentan las inversiones a futuro en un grado suficientemente pequeño, tal como puede ocurrir en el mercado de gasolina atendido por las estaciones donde se podrían considerar múltiples repeticiones.

Desde el punto de vista de entrada en el mercado las estaciones de servicio podrían asemejarse a un modelo de dos etapas como un modelo de una empresa que incurre en un costo de entrada hundido al menos en parte, que le permite contar con períodos posteriores de interacción competitiva.

Cuando una empresa debe incurrir en un costo irrecuperable al ingresar, debe considerar la reacción de otras empresas a su entrada, parte de ello es por lo que los análisis de mercado son relevantes al escoger la ubicación de las estaciones de servicio.

### 2.2.6 Modelo Stackelberg: líder - seguidor

Otros modelos que se podrían considerar en la aproximación inicial serían del tipo líder – seguidor tales como el desarrollado por Stackelberg, en los cuales se modelan dos fases de eventos, donde hay interdependencia, a partir de la decisión del líder, el seguidor define las cantidades a producir en este caso podría modelarse el líder Terpel y los seguidores como el resto de comercializadores mayoristas, otra aplicación podría considerarse en poblaciones con baja densidad de estaciones de venta de gasolina, donde exista una estación líder y estaciones seguidoras que fijen su venta objetivo aplicando el modelo líder -seguidor.

Se trata esencialmente de un modelo que consta de dos fases y en el que una de las empresas mueve primero. La otra puede observar el nivel de producción que ha elegido la primera y elegir su propio nivel óptimo de producción.

El modelo del liderazgo en la elección de la cantidad es un juego consecutivo, se resuelve, así, se supone que la empresa 1 es la líder y la 2 la seguidora. En ese caso, el problema de la empresa 2 es sencillo: dado el nivel de producción de la 1, la 2 desea maximizar sus beneficios  $p(q_1 + q_2)q_2 - c_2(q_2)$ . La condición de primer orden de este problema es la siguiente:

$$(Ecuación 18) \quad p(Q) + p'(Q)q_2 - c'_2(q_2).$$

Esta ecuación es exactamente igual que la condición de Cournot y puede utilizarse para obtener la función de reacción de la empresa 2,  $f_2(q_1)$ .

Pasando a la primera fase del juego, ahora la empresa 1 desea elegir su nivel de producción, mirando hacia adelante y teniendo en cuenta cómo responderá la 2.

Por lo tanto, la empresa 1 desea maximizar

$$(Ecuación 19) \quad p(q_1 + f_2(q_1))q_1 - c_1(q_1).$$

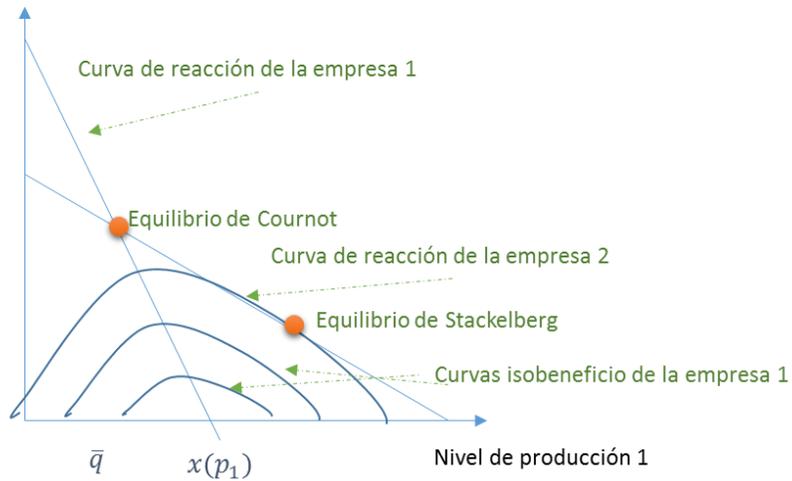
La condición de primer orden de este problema es

$$(Ecuación 20) \quad p(Q) + p'(Q) [1 + f'_2(q_1)]q_1 = c'_1(q_1).$$

Las (Ecuación 18) y la (Ecuación 20) son suficientes para hallar los niveles de producción de las dos empresas. En la Figura 2-5 se muestra gráficamente el equilibrio de Stackelberg. En este caso, las curvas isobeneficio de la empresa 1 indican las combinaciones de niveles de producción que generan unos beneficios constantes. Cuanto más bajas sean las curvas isobeneficio, mayores son los niveles de beneficio. Como muestra la Figura 2-5, la empresa 1 quiere producir en el punto de la curva de reacción de la empresa 2 que le genere los mayores beneficios, el líder en el equilibrio de Stackelberg normalmente tiene mayores beneficios que los que obtendría en el equilibrio de Cournot del mismo juego.

Figura 2-5 Equilibrio de Cournot y Stackelberg

Nivel de producción 2



Fuente: (Hal R. Varian, 1992)

### 2.3 Matriz de distancias y ponderación geográfica

El proceso de modelamiento contemplará una evaluación de pruebas de hipótesis nula  $H_0$  donde se consideran irrelevantes las variables de ubicación en el margen minorista y mayorista, así como una hipótesis alternativa  $H_1$  donde la variable de ubicación sea un elemento relevante en la formación del precio, o en su defecto altera las variables significativas dentro del panel.

Para considerar las distancias entre las estaciones de servicio, es necesario construir una matriz de ponderaciones o matriz de conectividad. (Arbia, 2014).

Para el ejercicio que se plantea, se le asignan pesos relativos asociados a las distancias entre las diferentes estaciones de servicio para lograrlo se utilizó el algoritmo incluido en Anexo 4, el cual toma las direcciones de las estaciones de servicio y las convierte en coordenadas geoespaciales de latitud y longitud (Bonivento, 2007), complementario a ello con las coordenadas se estructuró la matriz  $W$  a partir de la distancia euclídea utilizando Matlab (Otto & Denier, 2005).

La distancia se puede expresar de manera matricial mediante la matriz  $W$  así:

$$(Ecuación 21) W = \begin{bmatrix} w_{11} & \dots & w_{n1} \\ \vdots & w_{ik} & \vdots \\ w_{1n} & \dots & w_{nn} \end{bmatrix}$$

Donde los elementos se definen como:

$$(Ecuación 22) w_{i,k} = \begin{cases} 1 & \text{si } k \in N(i) \\ 0 & \text{otro} \end{cases}$$

Donde:

$N(i)$ : estaciones de servicio vecinas en la locación  $k$ . Considerando que  $w_{ii}$  es cero.

La matriz  $W$  se construye basada en la distancia (que es  $k \in N(i)$  si  $d_{i,k} < d_{max}$ ,  $d_{i,k}$  es la distancia entre la ubicación  $i$  y la ubicación  $k$ ).

A partir de la matriz de distancias  $W$ , es factible normalizar los valores considerando:

$$(Ecuación 23) w_{i,k}^* = \frac{w_{i,k}}{\sum_{j=1}^n w_{i,k}}; w_{i,k}^* \in W^*$$

Ahora bien, se podrían considerar los precios de las estaciones cercanas por medio de la función  $L(P_m)$  el cual puede convertirse en un insumo destacado en el análisis, mediante la siguiente ecuación es factible plantear el cálculo:

$$(Ecuación 24) L(p_i) = \frac{w_{i,j} * P_i}{\sum_{j=1}^n w_{i,k}}$$

De tal manera que es factible modelar la condición de mercado considerando la función:

$$(Ecuación 25) Mc = F(p_{i,t}, q_{i,t}, W_i)$$

Donde:

$Mc$ : condición de mercado monopolístico, mercado competitivo, u oligopólico.

$P_i$ = precio para la estación  $i$  en el periodo  $t$

$q_{i,t}$ = cantidad para la estación  $i$  en el periodo  $t$

$W_i$ = ponderador de distancia para estación  $i$

A partir de la matriz  $W$  se ponderan variables como precios y cantidades por la distancia entre las estaciones de servicio las cuales se utilizarán de entrada para que se puedan crear variables proxy en los modelos panel cuyos aspectos teóricos se incluyen en el numeral 2.4, y observar la significancia estadística de las variables

## 2.4 Modelos de panel de datos

Se escogió el modelo de panel de datos por las siguientes ventajas, (Damodar & Gujarati, 2010):

- a) Los datos de panel se refieren a individuos, empresas, estados, países, etc., a lo largo del tiempo, donde lo más probable es que exista heterogeneidad en estas variables, es decir que tengan unidades de medida distintas. Las técnicas de estimación de datos de panel toman en cuenta de manera explícita tal heterogeneidad, como es el caso de las variables en el análisis donde hay información de precios, cantidades, PIB municipal, número de estaciones.
- b) Al combinar las series de tiempo de las observaciones de corte transversal, los datos de panel proporcionan una mayor cantidad de datos informativos, más variabilidad, menos colinealidad entre variables, más grados de libertad y una mayor eficiencia.
- c) Al estudiar las observaciones en unidades de corte transversal repetidas, los datos de panel resultan más adecuados para estudiar la dinámica del cambio.
- d) Los datos de panel detectan y miden mejor los efectos que no se observan en datos puramente de corte transversal o de series de tiempo.
- e) Los datos de panel permiten estudiar modelos de comportamiento más complejos, como fenómenos con economías de escala y el cambio tecnológico.
- f) Al hacer disponibles datos para varios miles de unidades, los datos de panel reducen el sesgo posible si se agregan individuos o empresas en conjuntos numerosos.

El modelo típico de datos panel con efectos fijos es

$$C_{i,t} = \beta_{1,t} + \beta_2 Q_{i,t} + \beta_3 PF_{i,t} + \beta_4 LF_{i,t} + u_{i,t} \text{ Ecuación 26}$$

Donde:

$i=1 \dots n$  variable

$t=1 \dots n$  periodo

$\beta_{1,t}$ : intercepto efectos fijos

Q, PF, LF: variables cuya hipótesis es que están relacionadas con el comportamiento de  $C_{i,t}$ , entre las cuales pueden existir variables dicotómicas.

En nuestro caso se destaca que a partir del test de Hausman como se verá más adelante la regresión es de efectos fijos, es decir el intercepto varía entre las variables explicativas, pero no así en el tiempo.

La hipótesis nula en que se basa la prueba de Hausman es que los estimadores del modelo de regresión de efectos fijos y el modelo de componentes aleatorios no difieren considerablemente. El estadístico de prueba desarrollado por Hausman tiene distribución asintótica  $\chi^2$ . Si se rechaza la hipótesis nula, la conclusión es que el modelo de regresión de componentes aleatorios no es apropiado porque es probable que los efectos aleatorios

estén correlacionados con una o más regresoras y se prefiere el modelo de efectos fijos, tal como ocurre en los casos que se analizan más adelante.

## 2.5 Análisis de conglomerados

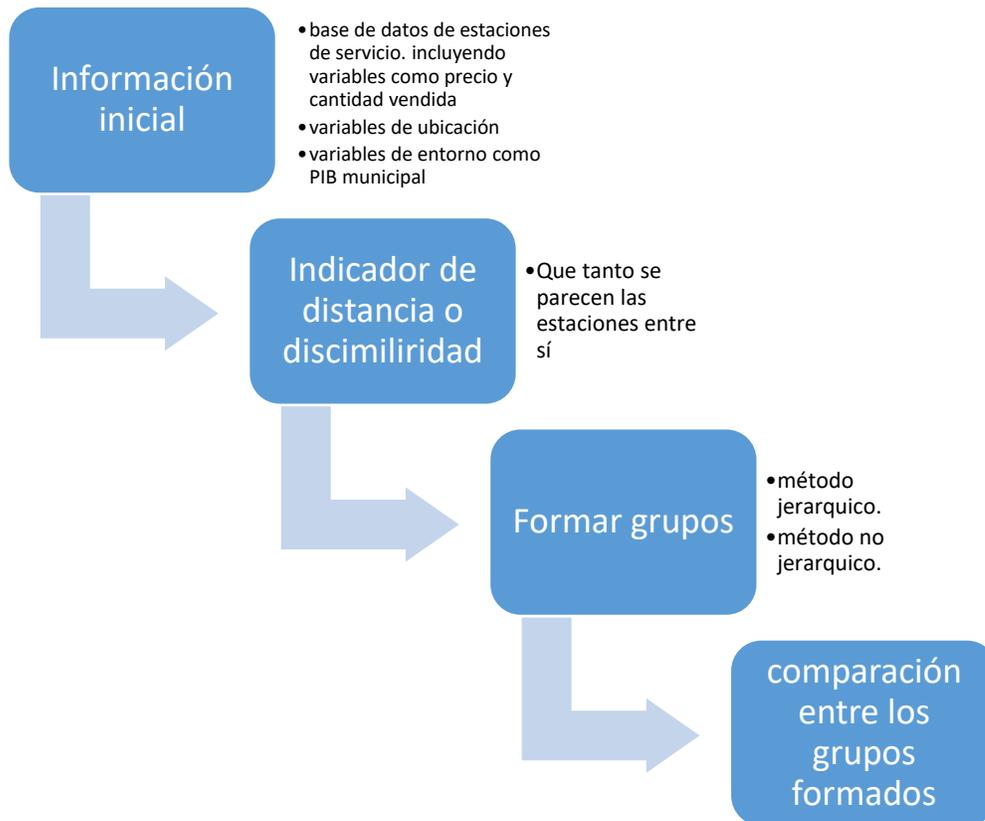
El análisis de conglomerados es un método de agrupamiento que permite determinar la cercanía o lejanía de un conjunto de unidades de observación mediante la comparación de medidas de distancia. El método consiste en formar grupos en donde se incluya la distancia entre los elementos del grupo y se analice el efecto en el precio en nuestro caso.

Típicamente un análisis de conglomerados o clúster es una técnica diseñada para clasificar distintas observaciones en grupos. (Aldás & Uriel, 2017). De tal forma que:

- a) Cada grupo (conglomerado o clúster) sea homogéneo respecto a las variables utilizadas para caracterizarlo, es decir, que cada observación contenida en él sea parecida a todas las que estén incluidas en ese grupo.
- b) Que los grupos sean lo más distintos posible unos de otros respecto a las variables consideradas.

El análisis de conglomerados se puede ilustrar bajo una secuencia lógica, tal como se presenta a continuación

Figura 2-6 Proceso de realización de un análisis de conglomerados



Dentro de la información que incluye la base de datos de estaciones de servicio, existen variables como precio y cantidad vendida, variables de ubicación, variables de entorno, tales como PIB municipal, se incluye la medida de similitud que en nuestro caso es la distancia euclídea.

$$\text{Ecuación 27 } D_{i,j} = \sqrt{\sum_{p=1}^k (x_{i,p} - x_{j,p})^2}$$

Donde:

$D_{i,j}$  = Distancia euclídea entre las variables

$x_{i,p}$  = Valor que toma la variable  $p$  en la observación  $i$

$x_{j,p}$  = Valor que toma la variable  $p$  en la observación  $j$

$k$  = observaciones totales

Además de ello es necesario estandarizar los datos para evitar que por tener ordenes de magnitud asociados a la unidad de medida generen distorsiones en la regresión (Aldás & Uriel, 2017). Una vez se cuenta con la información estandarizada y se dispone de las distancias euclídeas, es factible estructurar los grupos.

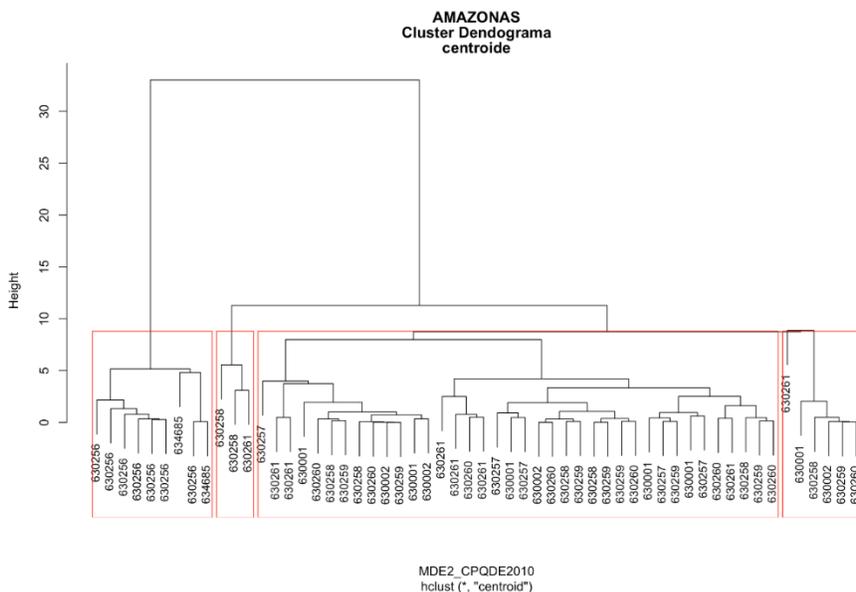
Para la estructuración de los grupos se consideran dos métodos, el método jerárquico donde se hace un proceso de aglomeración por similitud y el no jerárquico donde se definen grupos iniciales y los individuos se van clasificando en cada uno de esos grupos, dichos métodos se presentan a continuación.

### 2.5.1 Métodos jerárquicos

En este método se agrupan desde un individuo formando grupos considerando su similitud es decir de manera aglomerativa o se agrupan todos los individuos y se van separando en un método desagregativo. Entre los algoritmos para formar los grupos jerárquicos se destacan:

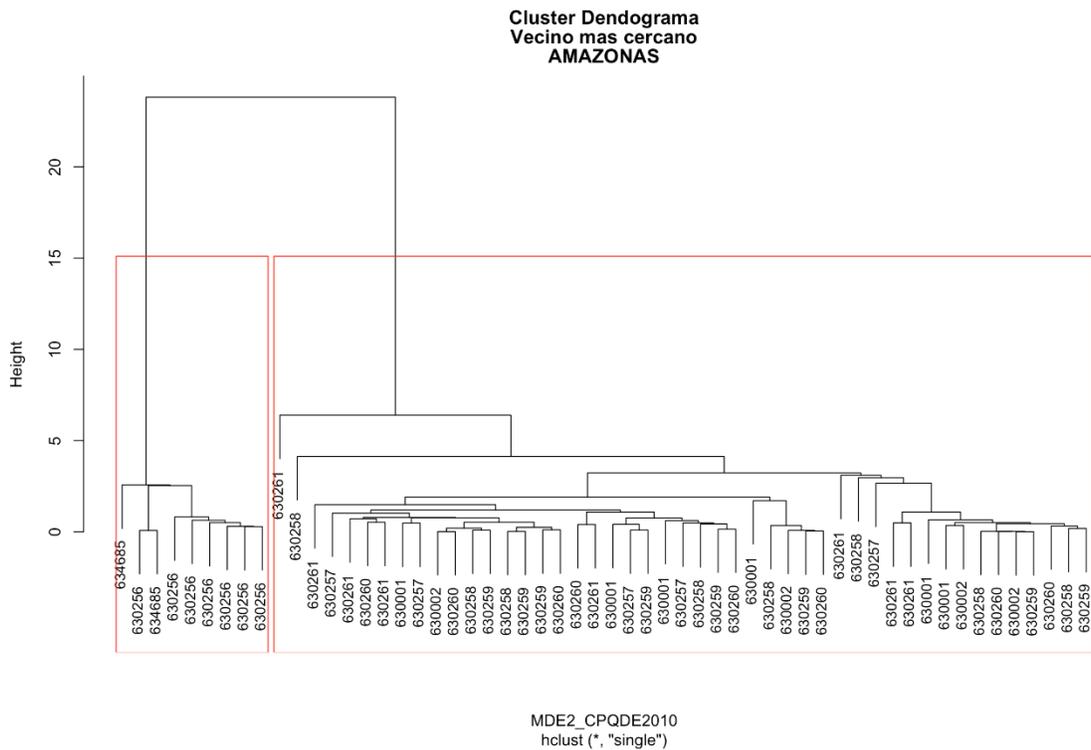
- a) Método del centroide: El cual parte de observaciones cercanas y recalcula el centro (centroide) en la medida que va agregando observaciones y las distancias se calculan a partir de dicho centroide. A continuación, se presenta un ejemplo considerando el departamento del Amazonas.

Figura 2-7 Dendograma método centroide



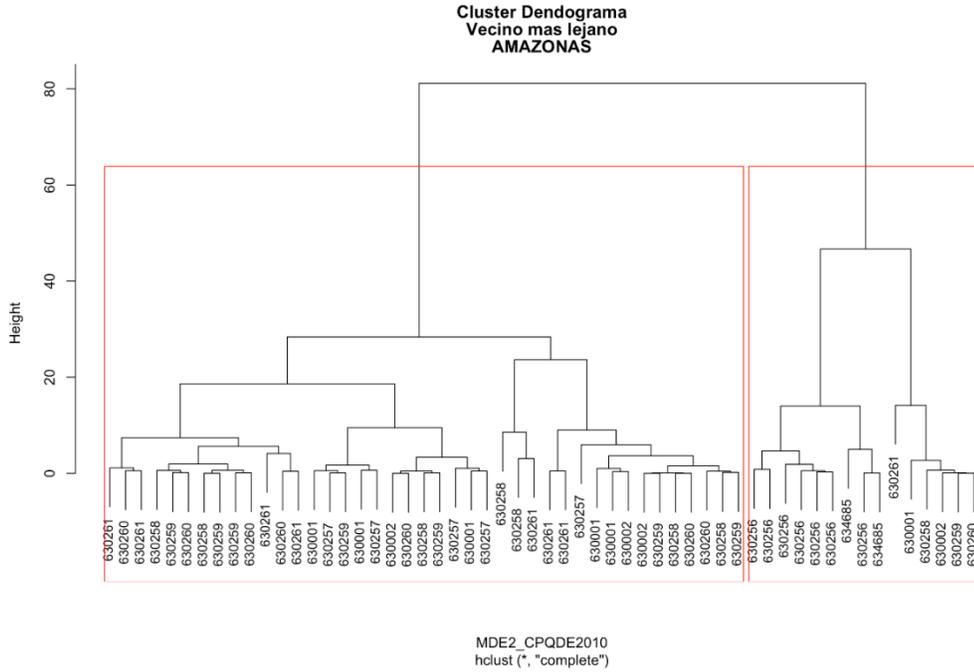
- b) Método del vecino más cercano: En el método anterior, la distancia entre los grupos se obtenía calculando las distancias entre sus centroides. En el método del vecino más cercano la distancia entre dos grupos es aquella que se da entre los dos miembros más cercanos de esos grupos. A continuación, se presenta un ejemplo para el departamento del Amazonas

Figura 2-8 Dendograma método vecino más cercano



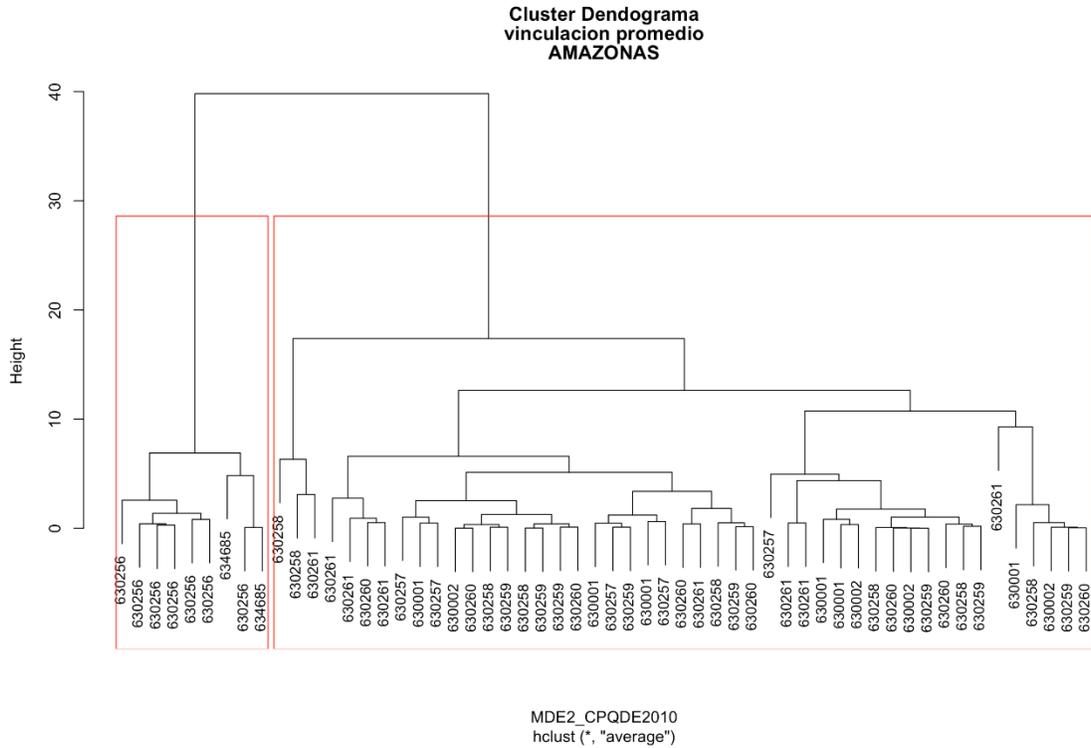
- c) Método del vecino más lejano: El método es análogo al anterior, pero con la diferencia de que la distancia entre dos grupos se mide por la distancia entre sus miembros más alejados. A continuación, se presenta un ejemplo considerando el departamento del Amazonas

Figura 2-9 Dendograma método vecino más lejano



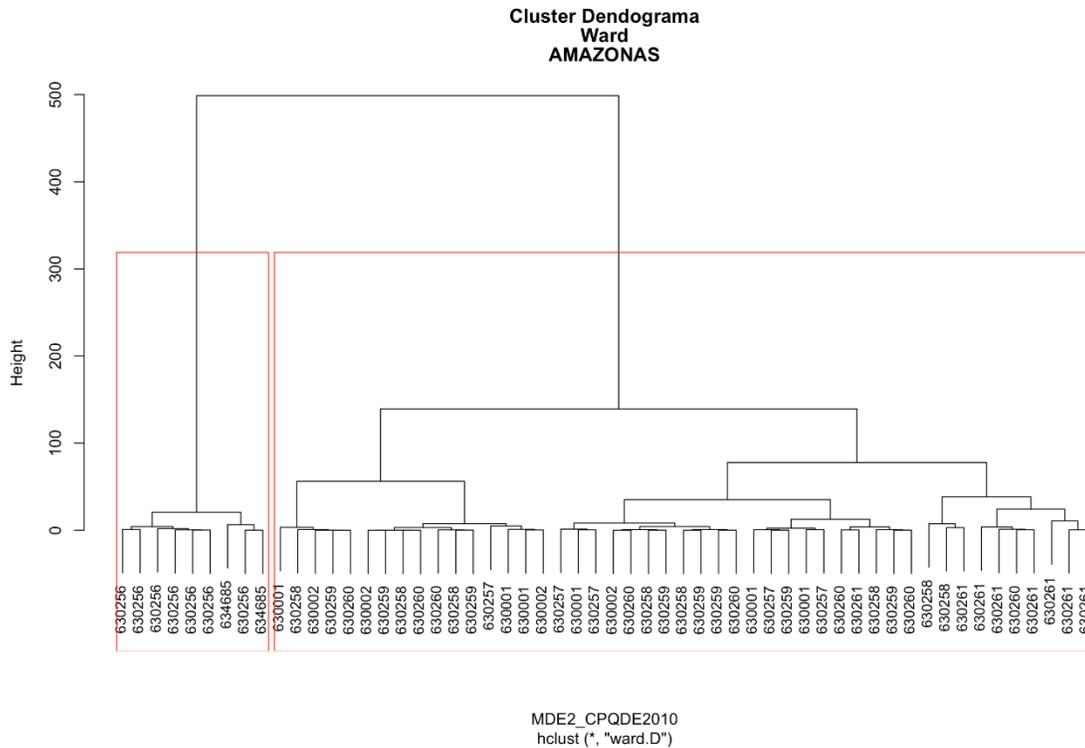
- d) Método de la vinculación promedio: En este procedimiento, la distancia entre dos grupos se obtiene calculando la distancia promedio entre todos los pares de observaciones que pueden formarse tomando un miembro de un grupo y otro miembro del otro grupo. A continuación, se presenta un ejemplo considerando el departamento del Amazonas

Figura 2-10 Dendograma método vinculación promedio



- e) Método de Ward: El método de Ward (Ward, 1963) no calcula la distancia entre los distintos conglomerados para decidir cuáles se deben fusionar, ya que su objetivo es maximizar la homogeneidad dentro de cada conglomerado. Para ello plantea todas las posibles combinaciones de observaciones para el número de grupos que se esté considerando en cada etapa concreta. A continuación, se presenta un ejemplo considerando el departamento del Amazonas

Figura 2-11 Dendograma método Ward



Respecto a estos métodos se puede señalar acorde a Aldás & Uriel (2017) que:

- a) El método del vecino más cercano es más sensible a la presencia de observaciones anómalas que el método del vecino más lejano.
- b) El método del vecino más lejano identifica habitualmente grupos muy homogéneos, en los que las observaciones son muy parecidas unas a otras.
- c) El método de Ward tiende a encontrar conglomerados no solo muy compactos, sino también de tamaño similar.
- d) El método del vecino más cercano tiene tendencia a crear menos grupos que el del vecino más lejano.

### 2.5.1.1 Selección del número de conglomerados de la solución

Considerando que es necesario definir el número de conglomerados que forman una solución razonable, para ello existen algunos índices:

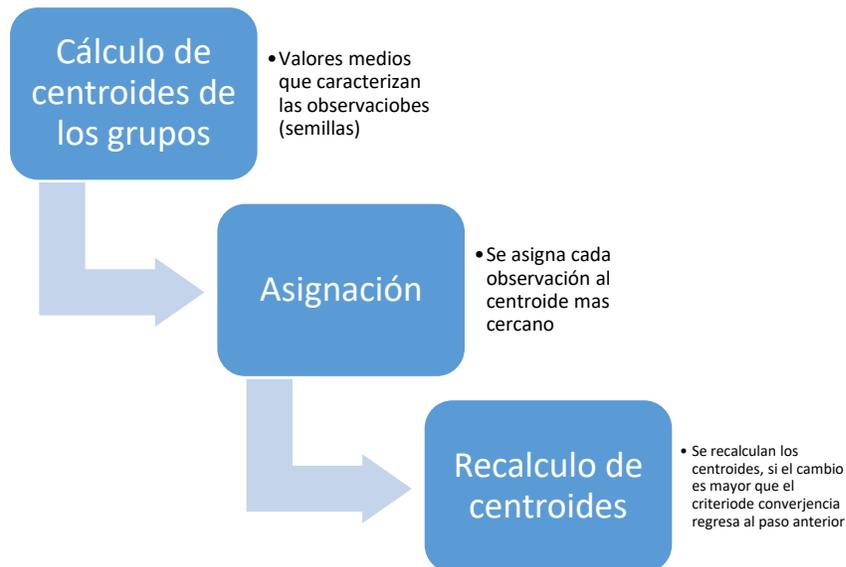
- a) Índice CH: El índice de Calinski y Harabasz
- b) Índice CCC: El índice CCC (Cubic Clustering Criterion);

- c) Índice Pseudo t2, El índice Pseudo t2 propuesto por Duda y Hart (1973) solo es aplicable a métodos jerárquicos.
- d) Índice DB El índice DB Davies y Bouldin, es una función de la ratio entre la dispersión interna de cada clúster y la separación entre ellos.
- e) Índice de Dunn El índice de Dunn (Dunn, 1974) define la ratio entre la mínima distancia intraconglomerado y la máxima distancia interconglomerado.
- f) Estadístico de Hubert: El estadístico es el coeficiente de correlación serial entre dos matrices, por un lado, la matriz de proximidad del conjunto de datos y una matriz con la distancia entre dos puntos representativos de los conglomerados a los que pertenecen.
- g) Índice Dindex: El índice se basa en la ganancia de inercia dentro del conglomerado, que mide el grado de homogeneidad .de los datos asociados con ese grupo. Calcula las distancias y las compara con un punto de referencia del perfil, normalmente el centroide.

## 2.5.2 Métodos no jerárquicos

En estos métodos se establece inicialmente un número de grupos a priori y los individuos se van clasificando en cada uno de esos grupos. En la siguiente figura se ilustran los pasos para construirlos:

Figura 2-12 Proceso análisis no jerárquico



Fuente: propia

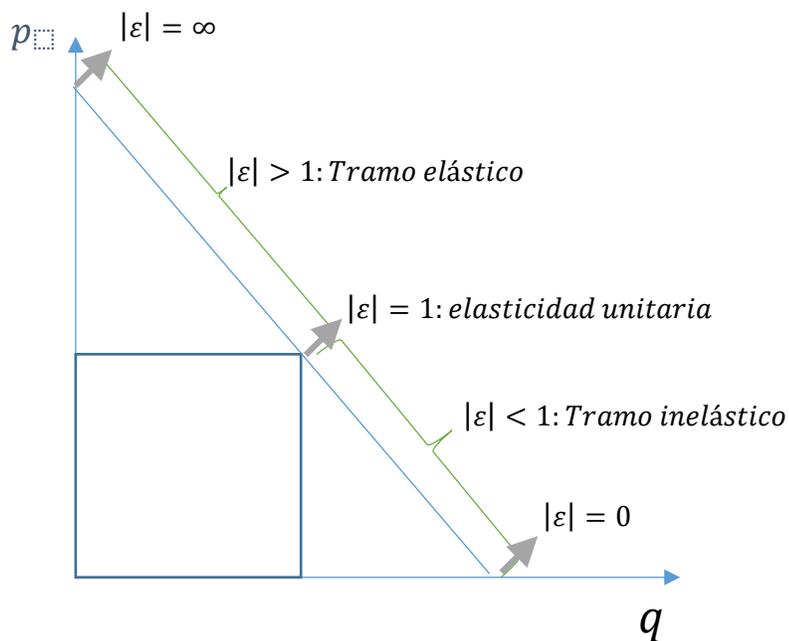
## 2.6 Elasticidad precio de la demanda

La elasticidad precio de la demanda, se puede definir como la variación porcentual de la cantidad dividida por la variación porcentual del precio, la elasticidad se puede escribir así

$$\text{Ecuación 28 } \varepsilon = \frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}}$$

Para interpretar el significado de la elasticidad se puede señalar que cuando esta es mayor a 1 en valor absoluto es una demanda elástica, mientras que si esta es menor que 1 en valor absoluto esta es inelástica, y si tiene una elasticidad de -1 es unitaria. (Varian, 2010). De manera gráfica es factible presentar la elasticidad de una curva de demanda lineal así:

Figura 2-13 Elasticidad curva demanda

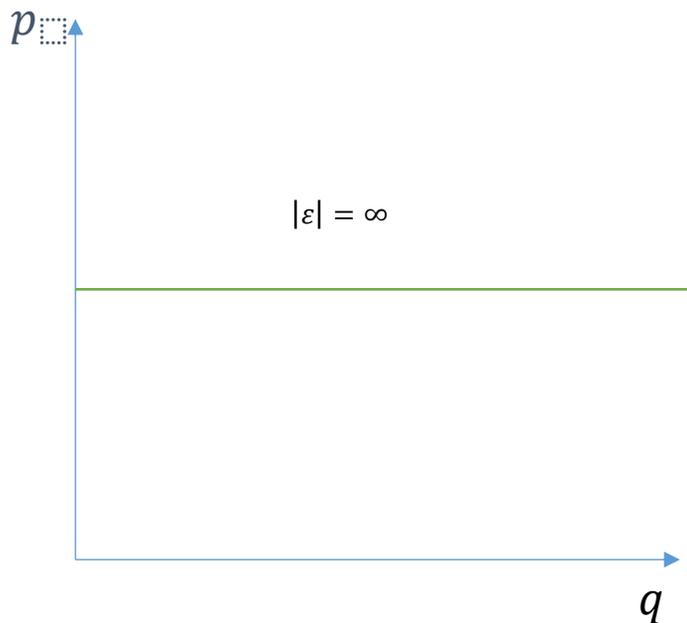


Fuente:(Varian, 2010)

### 2.6.1 Demanda perfectamente elástica:

Para el caso extremo en el que un pequeño cambio del precio ocasiona un cambio importante en la demanda, la elasticidad torna valores cercanos a  $|\varepsilon| \rightarrow \infty$ , asimismo la curva de precios vs demanda se torna horizontal, cuya consecuencia es que cualquier incremento en el precio producirá tal disminución en la cantidad, la demanda se volverá nula, es decir, a cualquier precio diferente a  $P$  no hay demanda por el bien.

Figura 2-14 Demanda perfectamente elástica

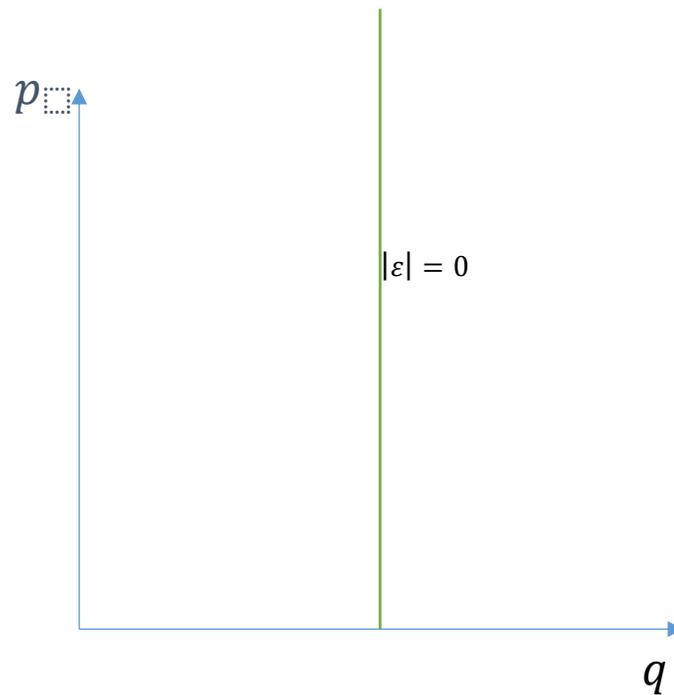


Fuente: (Nicholson, 2005), elaboración propia

### 2.6.2 Demanda perfectamente inelástica:

Cuando una variación en el precio del bien no ocasiona ningún cambio en la demanda, se podría afirmar que es una demanda perfectamente inelástica.  $|\varepsilon| = 0$

Figura 2-15 Demanda perfectamente inelástica



Fuente: (Nicholson, 2005), elaboración propia

En el presente trabajo se estudia el comportamiento de la variación de las cantidades de la gasolina motor, vendida en las estaciones de servicio, frente a la variación del precio; los principales resultados se abordan en el numeral 3.2.8

### 3 Metodología

Dentro de los elementos que se abordan en la metodología de este estudio se destacan:

- a) Investigación general del mercado de gasolina a nivel nacional, caracterizando la información estadística del número de estaciones de servicio, los precios, las cantidades vendidas, el PIB municipal y la distribución geográfica de las mismas.
- b) Regresiones econométricas del precio incluyendo la variable de ubicación geográfica, variables de entorno como el PIB municipal, incluyendo dieciocho modelos de regresión cuya estimación tiene como variable dependiente el precio de la gasolina en cada estación (LABYS, 1986).
- c) Estimaciones de conglomerados por departamento.(Aldás & Uriel, 2017).

En principio es factible estructurar la metodología a partir de la conjunción de las variables de caracterización del mercado presentadas en el siguiente diagrama:

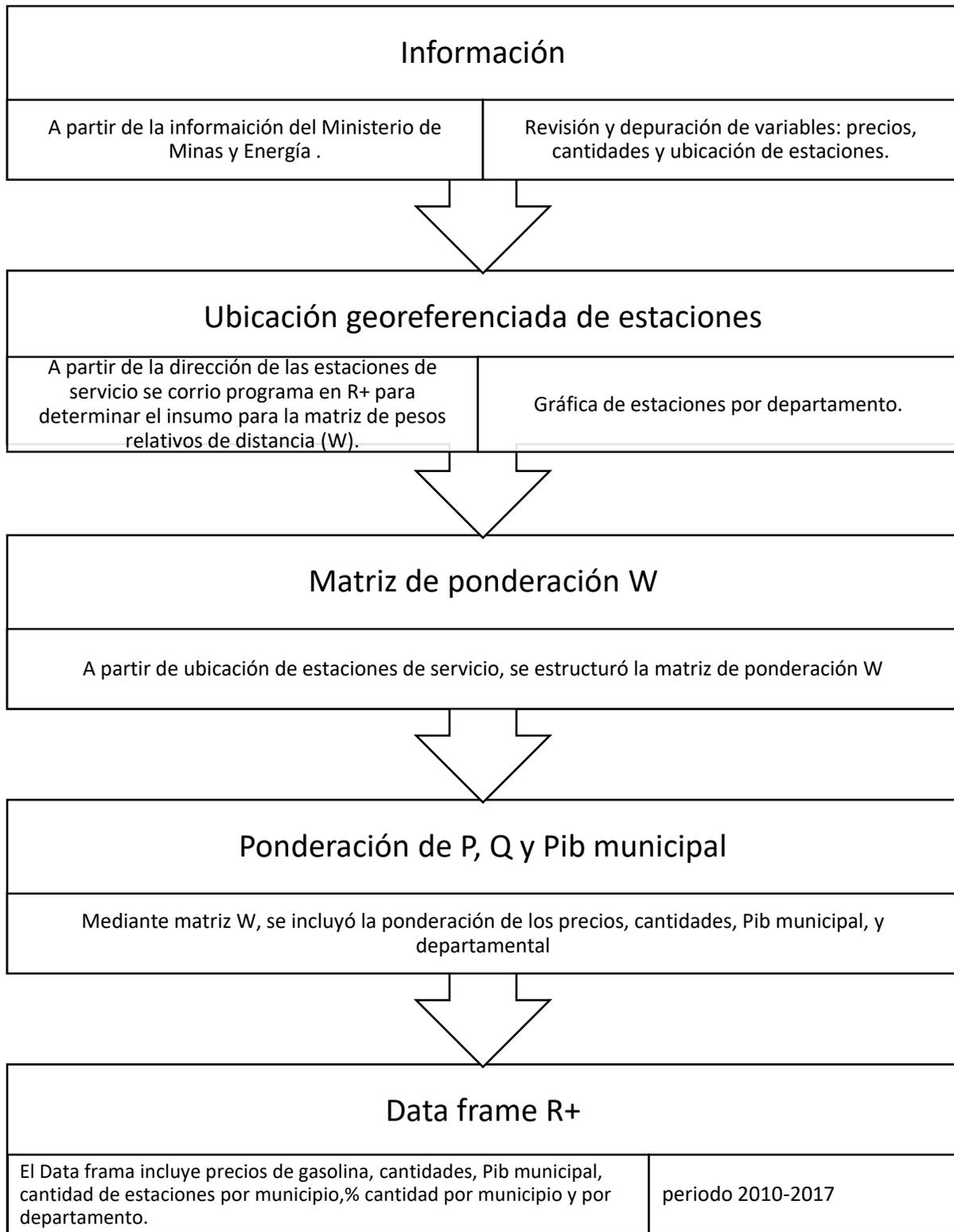
Figura 3-1 variables caracterización mercado

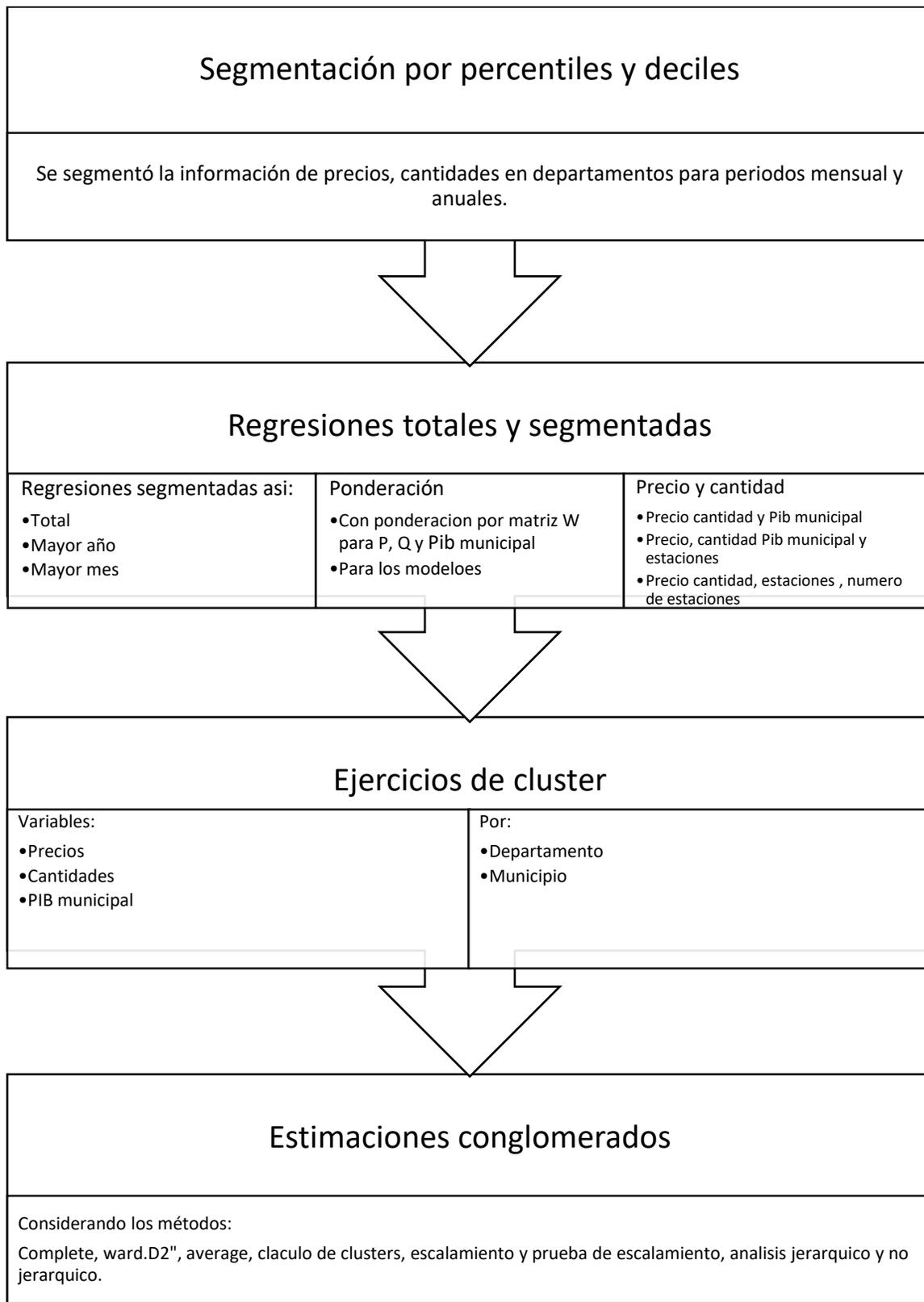


Fuente: propia

Para estructurar el modelo econométrico y los conglomerados se consideró el proceso descrito en el siguiente diagrama:

Figura 3-1 Principales hitos en el desarrollo del modelo econométrico





### 3.1 Depuración de información

En la depuración de información se eliminaron los datos con información faltante considerando que la eliminación de casos por parejas, bajo el supuesto en que contengan un valor perdido en las variables que se están utilizando en un análisis determinado. La limitación de este enfoque es el tamaño muestral, provocando serios problemas en algunas técnicas, debido a que la muestra podría reducirse de manera notable y podría dejar de ser aplicable el método de regresión. (Aldás & Uriel, 2017).

Sin embargo, en presente caso con una base de datos de 478.114 datos en el panel no se considera que dicho aspecto sea un problema relevante dado el tamaño de la base de datos con el que se cuenta.

Respecto a los outliers en la base de datos se restringió el rango de precios en la muestra entre 1.000 COP a 60.190 COP considerando las experiencias del autor en las visitas que se hacen a campo en zonas alejadas en el país. (UPME, 2015).

### 3.2 Análisis de la información

La información incluida en el análisis fue consultada del SICOM, Sistema de Información de Combustibles Líquidos del Ministerio de Minas y Energía, en el cual se organiza, controla y sistematiza la comercialización, distribución, transporte y almacenamiento de combustibles líquidos derivados del petróleo, alcohol carburante y biodiesel. (Minminas, 2017).

Este sistema requiere unos niveles de acceso para operar y entrar al mismo por ello dentro de este trabajo se prefirió hacer una solicitud formal al Ministerio de Minas y Energía de la información para el modelamiento, acorde al Anexo 1.

#### 3.2.1 Matriz de información consolidada y depurada

Se estructuró una matriz de información de 478.114 filas incluyendo 42 variables para adelantar las corridas principalmente al panel de datos, las cuales se describen en la siguiente tabla:

Tabla 3-1 Resumen de variables de matriz panel

Variable	Descripción
SICOM	Código Sistema de Información de Comercialización de Combustibles
fecha	Mes y año de la información
alo	Año de la información
mes	Mes de la información
P	Precio en COP/galón
Q	Cantidad, galones/mes

Tabla 3-1 Resumen de variables de matriz panel

Variable	Descripción
PIB_municipal	Producto interno bruto municipal
PP_R	Precio ponderado por matriz de distancias W
QQ_R	Cantidad ponderado por matriz de distancias W
PIBP_municipal_R	Producto interno bruto municipal ponderado por matriz de distancias W
BANDERA	Marca estación de servicio
N_estaciones	Número de estaciones por municipio
PorcQ_bandera_T_alo	porcentaje del total de cantidad vendida en cada municipio por la estación de servicio
PorcQ_bandera_municipio_alo	Porcentaje del municipio de cantidad vendida en cada municipio por la estación de servicio
AYATAWACOOP	Variable dicotómicas (1-0) existencia de la marca AYATAWACOOP en el municipio
BIOMAX	Variable dicotómicas (1-0) existencia de la marca BIOMAX en el municipio
BRIO	Variable dicotómicas (1-0) existencia de la marca BRIO en el municipio
CASAMOTOR	Variable dicotómicas (1-0) existencia de la marca CASAMOTOR en el municipio
COOMULPINORT	Variable dicotómicas (1-0) existencia de la marca COOMULPINORT en el municipio
CORPOIL	Variable dicotómicas (1-0) existencia de la marca CORPOIL en el municipio
DISCOWACOOP	Variable dicotómicas (1-0) existencia de la marca DISCOWACOOP en el municipio
ECOS	Variable dicotómicas (1-0) existencia de la marca ECOS en el municipio
ESSO	Variable dicotómicas (1-0) existencia de la marca ESSO en el municipio
GULF	Variable dicotómicas (1-0) existencia de la marca GULF en el municipio
MOBIL	Variable dicotómicas (1-0) existencia de la marca MOBIL en el municipio
OCTANO	Variable dicotómicas (1-0) existencia de la marca OCTANO en el municipio
PETROBRAS	Variable dicotómicas (1-0) existencia de la marca PETROBRAS en el municipio
PETROMIL	Variable dicotómicas (1-0) existencia de la marca PETROMIL en el municipio
PROXXON	Variable dicotómicas (1-0) existencia de la marca PROXXON en el municipio
PUMA	Variable dicotómicas (1-0) existencia de la marca PUMA en el municipio
SAVE	Variable dicotómicas (1-0) existencia de la marca SAVE en el municipio

Tabla 3-1 Resumen de variables de matriz panel

Variable	Descripción
TERPEL	Variable dicotómicas (1-0) existencia de la marca TERPEL en el municipio
TEXACO	Variable dicotómicas (1-0) existencia de la marca TEXACO en el municipio
ZAPATA_Y_VELASQUEZ	Variable dicotómicas (1-0) existencia de la marca ZAPATA_Y_VELASQUEZ en el municipio
ZEUSS	Variable dicotómicas (1-0) existencia de la marca ZEUSS en el municipio

El resumen de la información del panel se consigna en la siguiente tabla:

Tabla 3-2 Base de datos total (N=478,114)

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
SICOM			630,001			636,323
Año			2,010	2,011	2,015	2,017
mes			1	3	9	12
P	8,504.312	2,307.410	1,000.000	7,840.000	8,760.000	60,190.000
Q	26,495.880	30,526.590	1.000	7,148.000	34,660.000	569,000.000
PIB_ municipal	2,139,956.000	3,866,945.000	3,216.828	112,423.000	1,874,596.000	22,650,614.000
PP_R	46,796,227.000	217,768,188.000	10,103.000	72,758.000	92,572.000	1,118,500,000.000
QQ_R	227,724,487.000	540,004,781.000	5,156.000	52,295.000	88,545.000	3,377,300,000.000
PIBP_ municipal_R	357,755,352.018.000	168,115,814.628.000	15,100.000.000	258,460,000.000.000	458,370,000.000.000	893,980,000.000.000
AYATAWACOOP			0			1
BIOMAX			0			1
BRIO			0			1
CASAMOTOR			0			1
COOMULPINORT			0			1
CORPOIL			0			1
DISCOWACOOP			0			1

ECOS			0			1
ESSO			0			1
GULF			0			1
MOBIL			0			1
OCTANO			0			1
PETROBRAS			0			1
PETROMIL			0			1
PROXXON			0			1
PUMA			0			1
SAVE			0			1
TERPEL			0			1
TEXACO			0			1
ZAPATA_Y_VELASQ UEZ			0			1
ZEUSS			0	0	0	1
N_estaciones	51.992	106.380	1	5	34	404
PorcQ_bandera_T_al o	0.214	0.146	0.00001	0.100	0.372	0.391
PorcQ_bandera_mun icipio_alo	0.467	0.311	0.001	0.218	0.683	1.000

Fuente SICOM

Se destaca del panel que cuenta con 478.114 observaciones además de incluir como variables dicotómicas las “banderas” para las diferentes estaciones.

Como abordaje inicial entre los aspectos de caracterización del mercado de gasolina se pueden considerar

- a) Aspectos de composición del mercado asociado a la bandera o distribuidor mayorista al que está afiliado.
- b) La cantidad y distribución geográfica de estaciones de servicio.
- c) El comportamiento de precios y cantidades

### 3.2.2 Estaciones de servicio y distribuidor mayorista

Es factible observar la estructura de este a partir de la distribución del número de estaciones y la “bandera” a la cual pertenecen, lo cual se presenta en la siguiente gráfica

Figura 3-2 Porcentaje de estaciones según bandera



Fuente: SICOM, información con corte a febrero de 2017

Lo que en principio indica una fuerte participación a nivel nacional de estaciones afiliadas a banderas tales como Terpel, Biomax, Mobil, Texaco y Petromil las cuales suman un 84% de participación.

### 3.2.3 Estaciones por municipio

Una de las variables destacadas es la asociada a la cantidad de estaciones a nivel municipal, debido a que las condiciones geográficas pueden hacer que el mercado en una población apartada pueda estar sujeto a un número restringido de estaciones las cuales independiente de la bandera podrían considerar mayores precios en ausencia de competencia.

Para analizar dicho fenómeno se separaron las estaciones en percentiles y en deciles considerando el número de estaciones por municipio, para observar si existen niveles de precios superiores en zonas con bajo número de estaciones.

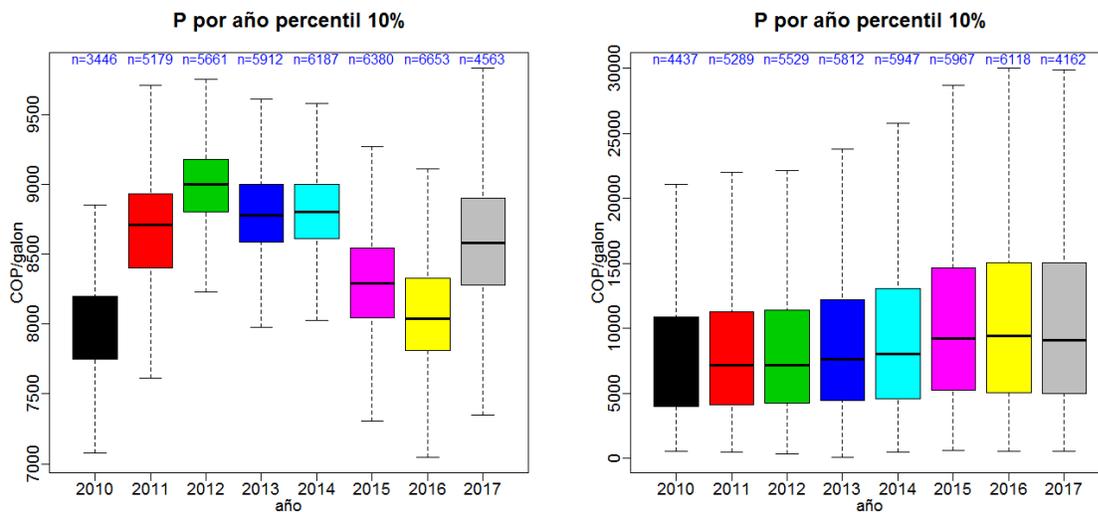
En primera instancia se consideran la caracterización por percentiles y el decil 10% para el análisis base.

Tabla 3-3 Percentiles número de estaciones						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
1	5	11	91	34	404	

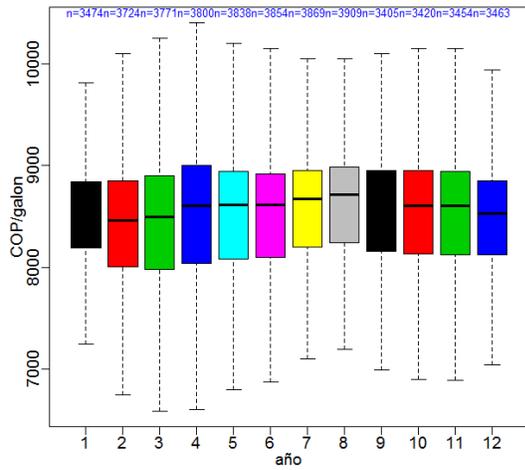
Tabla 3-4 Deciles número estaciones										
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1	2	4	5	8	11	17	24	45	153	404

### 3.2.3.1 Decil <10%

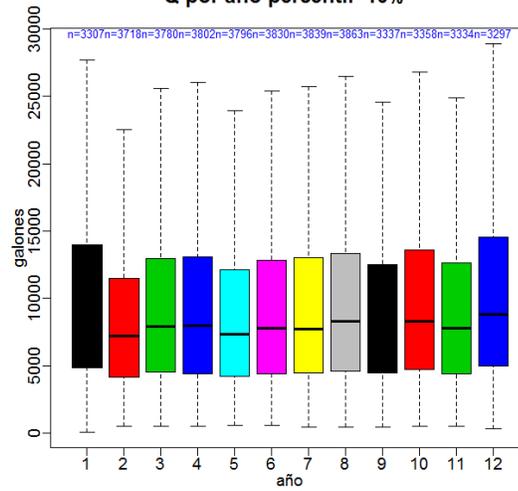
Figura 3-3 Decil <10%



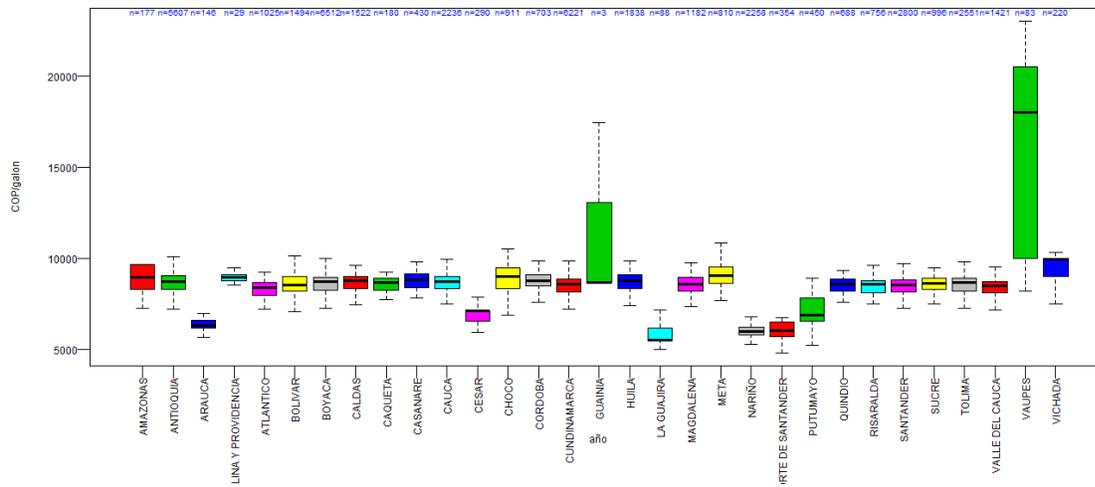
P por año percentil 10%

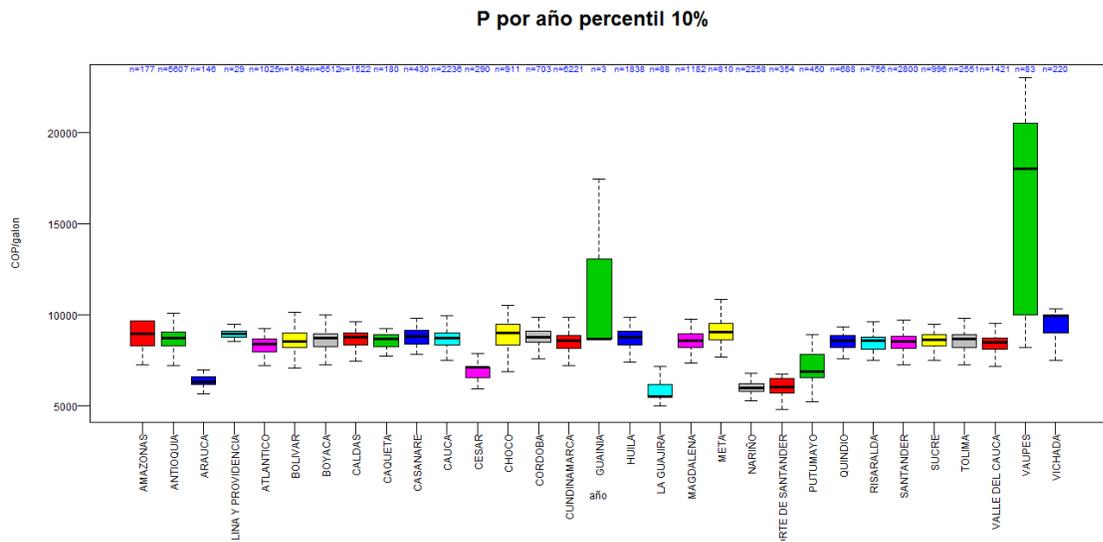


Q por año percentil 10%



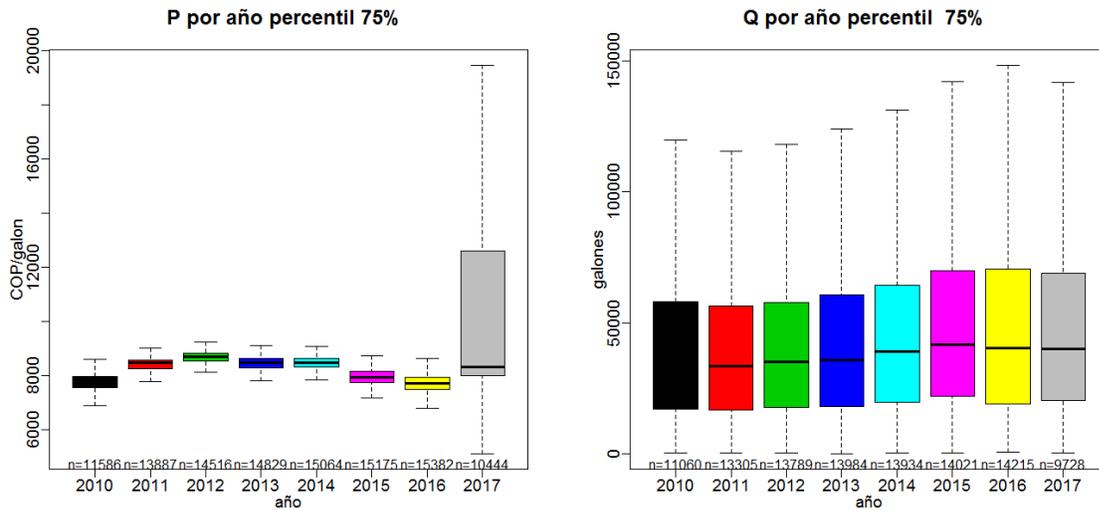
P por año percentil 10%

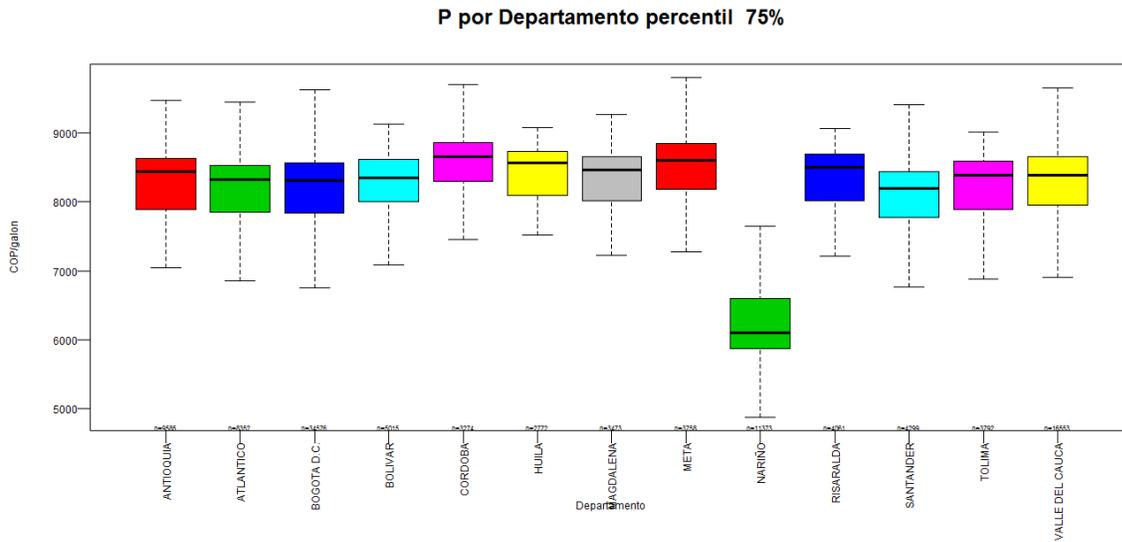
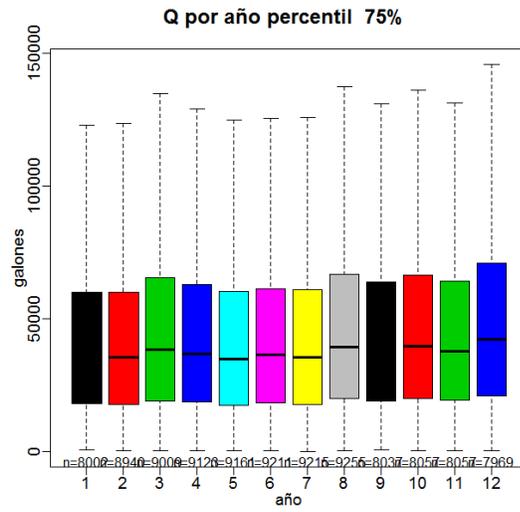
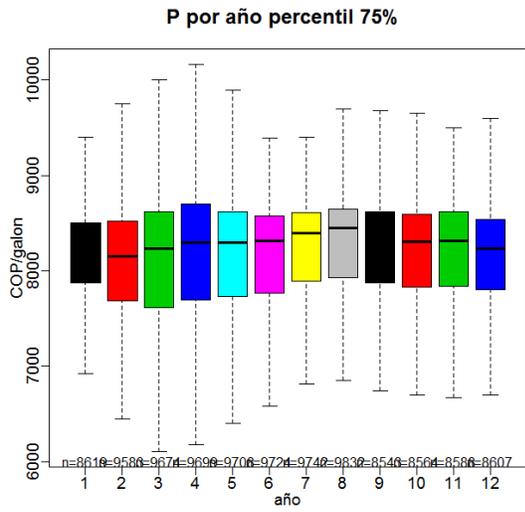


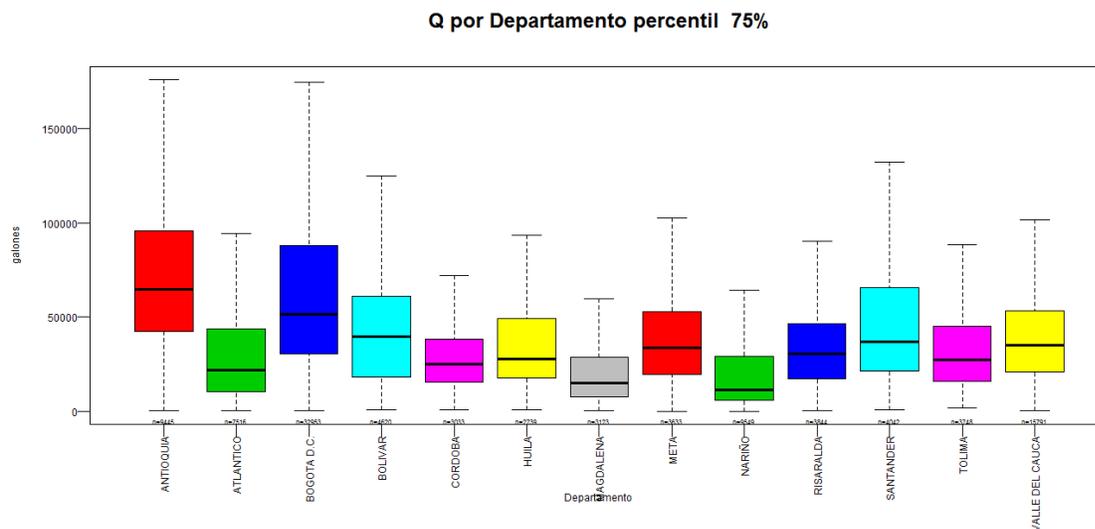


**3.2.3.2 Percentil >75% cantidad de estaciones**

Figura 3-4 Decil > 75%







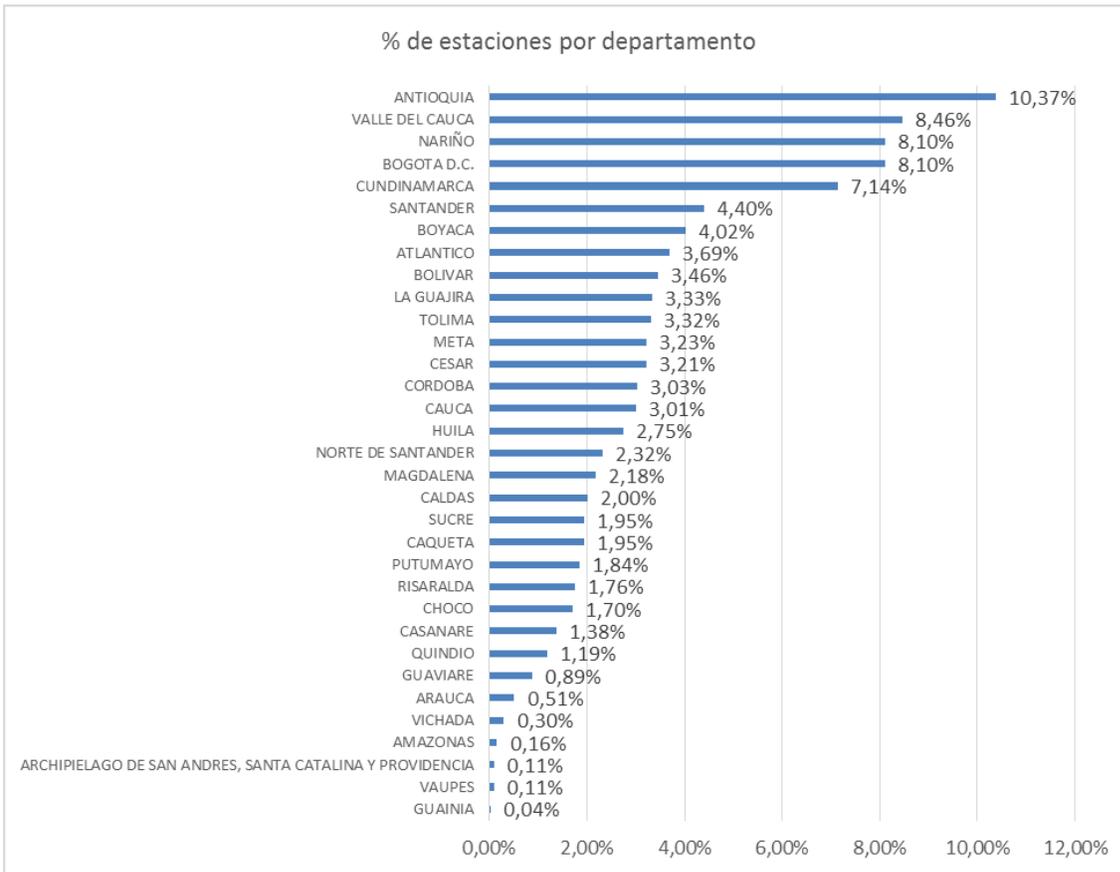
A partir de los resultados es factible señalar que:

- a) Al comparar el comportamiento agregado por año de los precios entre el decil 10% y 75% no se puede observar que el supuesto de que con la menor cantidad de estaciones por municipio existan mayores precios de gasolina.
- b) Al comparar el comportamiento agregado por mes de los precios entre el decil 10% y 75% se puede observar el supuesto de que con la menor cantidad de estaciones por municipio existan mayores precios de gasolina, aunque con unas variaciones cercanas al 3%.
- c) Al comparar el comportamiento agregado por departamento de los precios entre el decil 10% y 75% se puede observar que el supuesto de que con la menor cantidad de estaciones por municipio existan mayores precios de gasolina, aunque con unas variaciones cercanas al 5.8%.

### 3.2.4 Distribución geográfica estaciones de servicio

Respecto al número de estaciones por departamento en la siguiente figura se pueden observar la distribución porcentual:

Figura 3-5 Porcentaje de estaciones por Departamento

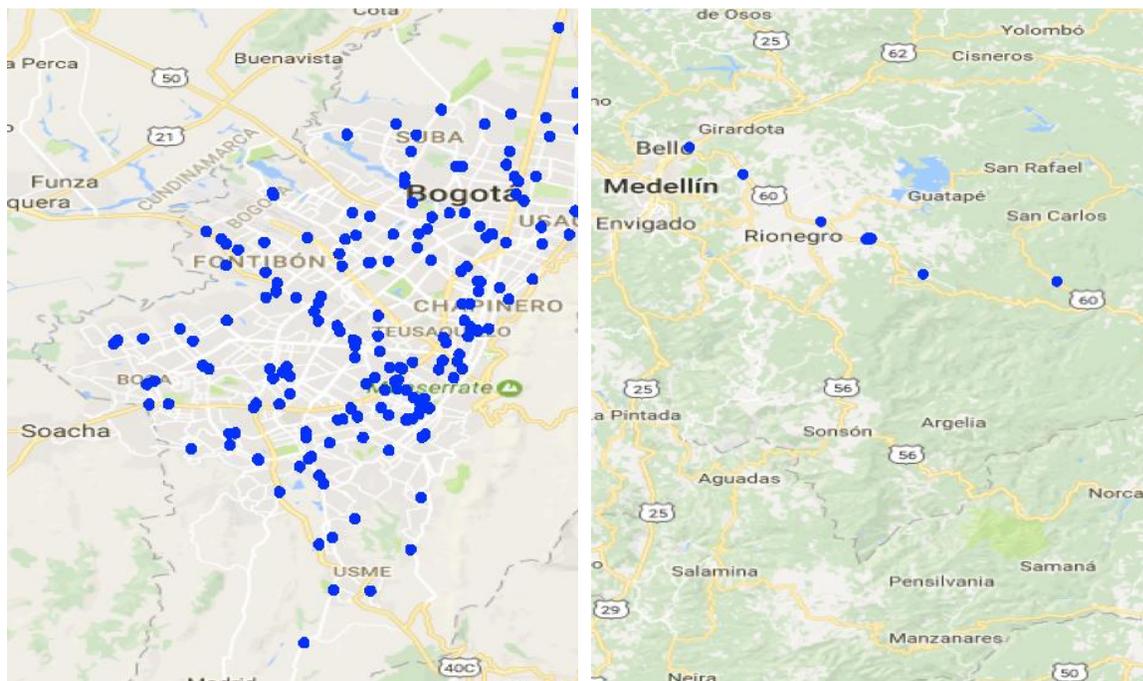


Fuente: SICOM, información con corte a febrero de 2017

Existen departamentos como Antioquia, Valle del Cauca, Nariño, Bogotá, Cundinamarca, Santander y Boyacá que cuentan con 50.6% de las estaciones de servicio en el país.

Desde la perspectiva de distribución geográfica es factible observar la distribución de las estaciones de servicio que distribuyen gasolina motor, en la siguiente gráfica se presenta la distribución geoespacial para el mercado de Bogotá D.C

Figura 3-6 Estaciones de servicio gasolina Bogotá, Rionegro



Fuente: SICOM, información mes de febrero de 2017, realizado en R+

Al observar la Figura 3-6 se presenta un mercado con amplia densidad de estaciones de servicio tal como Bogotá y un mercado con baja densidad como el existente en la vía Medellín, la Dorada, estas diferencias de distribución confirman el supuesto de las diferencias en la dispersión de las estaciones de servicio.

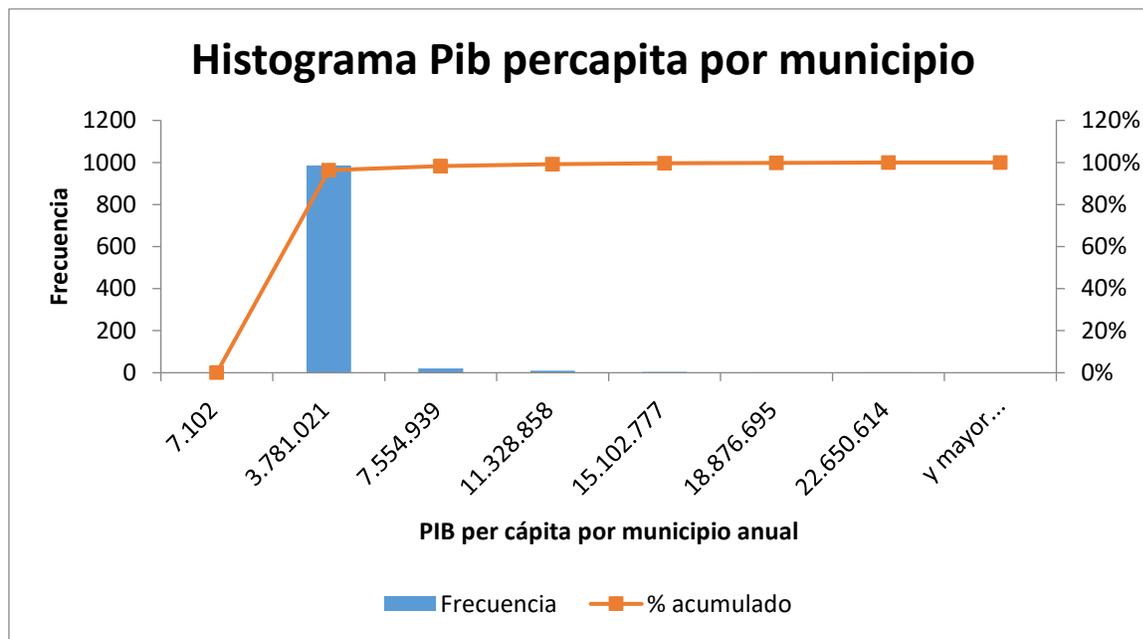
En el Anexo 2 Mapas de las estaciones, se incluyen algunos mapas de las ubicaciones de las estaciones de servicio, por departamento.

### 3.2.5 PIB municipal

Dado que se consideró deseable contar con una variable que representara el ingreso en los diferentes municipios para analizarlo a la luz de su efecto en los precios en las diferentes estaciones se construyó un PIB a nivel de municipio.

A partir de información del indicador de importancia económica municipal publicado por el (DANE, 2017), y de la información del Producto interno bruto per cápita departamental publicado también por el DANE (DANE, 2017), se realizó una ponderación para determinar de manera mensual para cada uno de los municipios. En el siguiente histograma se presenta la distribución del PIB per cápita por municipio.

Figura 3-7 PIB por municipio



Fuente: DANE

Como se puede observar el 97% de los municipios tiene un PIB menor a 3,800,000 COP al año.

### 3.2.6 Precios

En presente aparte se revisan los principales aspectos frente a la estructura de precios, así como la estadística del comportamiento.

#### 3.2.6.1 Estructura de precios

Desde la perspectiva de estructura de precios de la gasolina, como se puede observar en la Figura 3-8, las principales componentes del precio de la gasolina son el ingreso al productor que representa el valor del producto, y la carga impositiva, mientras que el distribuidor minorista incluido el valor de la pérdida por evaporación alcanza el 9%.(Minminas, 2017).

Figura 3-8 Estructura del precio de la gasolina motor



Fuente: SICOM, información mes de enero de 2017,

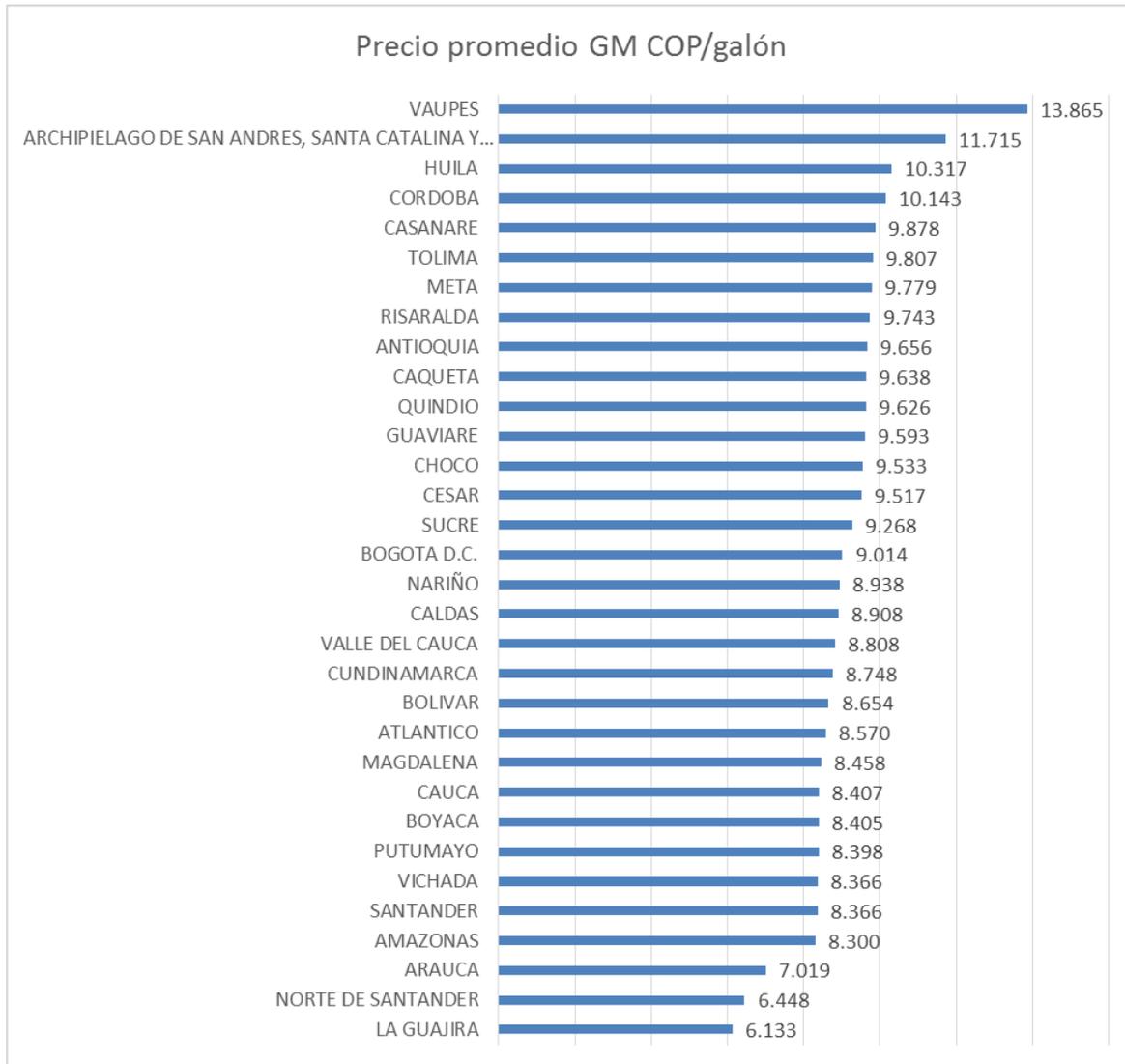
### 3.2.6.2 Promedio precios Estaciones de servicio por departamento

Frente al comportamiento de los precios promedio por departamento se puede observar su distribución en la Figura 3-9, para el mes de febrero de 2017, donde se destacan los mayores precios en departamentos aislados tales como Vaupés y San Andrés Isla y sorpresivamente en zonas del interior con tradición petrolera como el Meta.

Asimismo, existen departamentos como la Guajira y Norte de Santander con menores precios, probablemente producto de su vecindad con Venezuela y los posibles efectos del contrabando de combustibles, probablemente en el análisis que se desarrolle en el trabajo exista la necesidad de excluir algunos departamentos.

Existen mercados como Bogotá que como se verá posteriormente en el análisis de cantidad presentan un consumo importante y que sin embargo están en la mitad de la tabla de precios probablemente restringidos por el resto de las componentes de costos en la cadena de valor tales como el transporte y el ingreso al productor

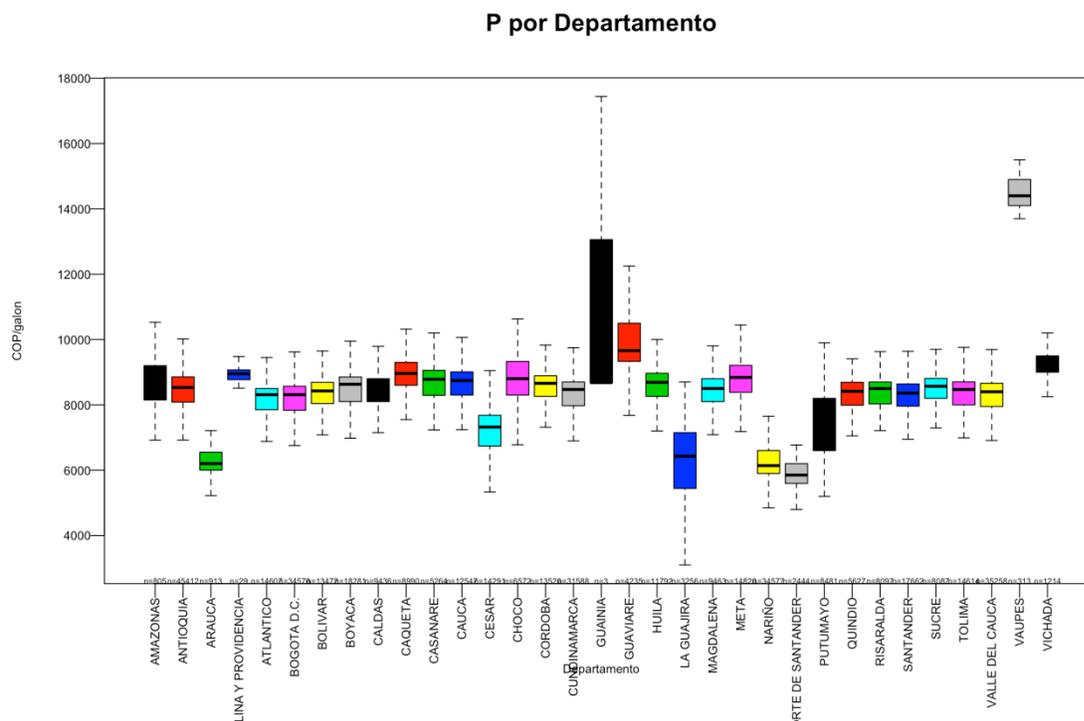
Figura 3-9 Precios promedio de gasolina motor



Fuente: SICOM, información mes de febrero de 2017,

Al observar la Figura 3-10 donde se incluye la mediana y los percentiles del precio de la gasolina por departamento se destaca el valor en la zona de frontera tales como Norte de Santander, Arauca, Nariño y la Guajira y existen departamentos como Guainía que tienen una mediana cercana a los 13,000COP y Vaupés cercana a los 15,000COP

Figura 3-10 Distribución precios p por departamento

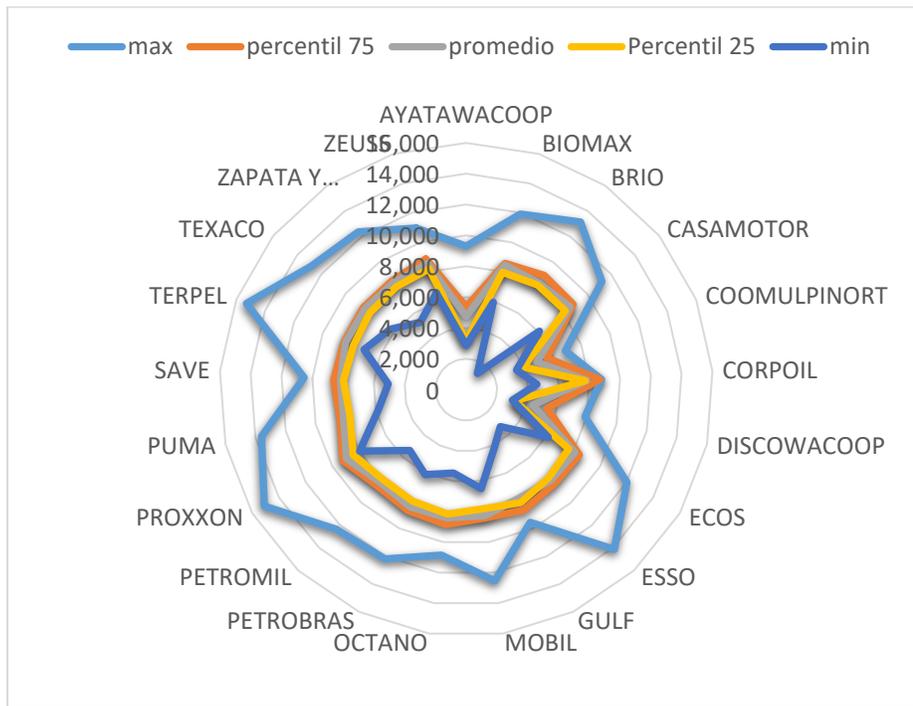


### 3.2.6.3 Caracterización de precios por bandera

Dado que uno de los factores destacados es evaluar el comportamiento de precios a nivel de bandera, en el siguiente aparte se analiza el comportamiento de los precios en el tiempo para cada una de las firmas de distribución mayorista “banderas” en el argot del sector, para observar comportamientos sistemáticamente prevalentes acorde a la estrategia de cada una de las empresas.

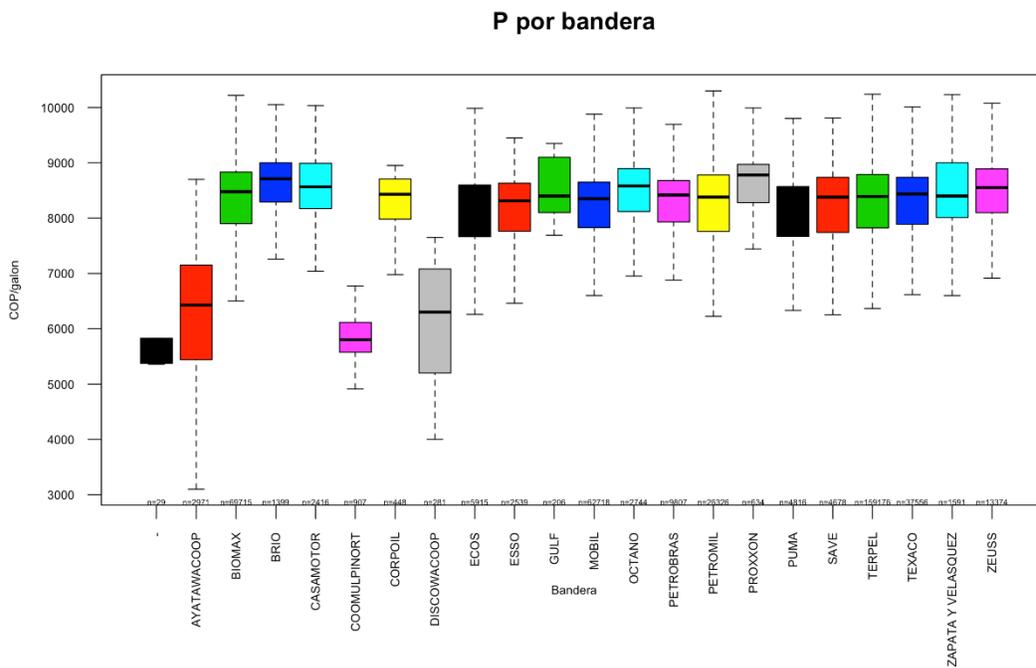
Graficando el valor máximo, los percentiles 75 y 25 así como el valor mínimo de la serie de precios como presenta en la Figura 3-11 en la Tabla 3-5, es factible observar el comportamiento diferenciado de jugadores con mayor participación del mercado frente a jugadores de menor participación.

Figura 3-11 Caracterización precios por bandera



Fuente: SICOM, información mes de enero de 2017,

Figura 3-12 Precios por bandera



Desde la perspectiva de los precios por bandera se observan que las empresas que atienden en zonas de frontera tienen unos precios con medianas menores en parte por los cupos de combustibles y trato diferencial a dichos mercados.

Por otra parte, respecto a los precios de jugadores importantes como Terpel, Biomax y Mobil, no es factible afirmar que los precios tienen una distribución con media y percentiles más altos frente a otros jugadores en el mercado con menor participación.

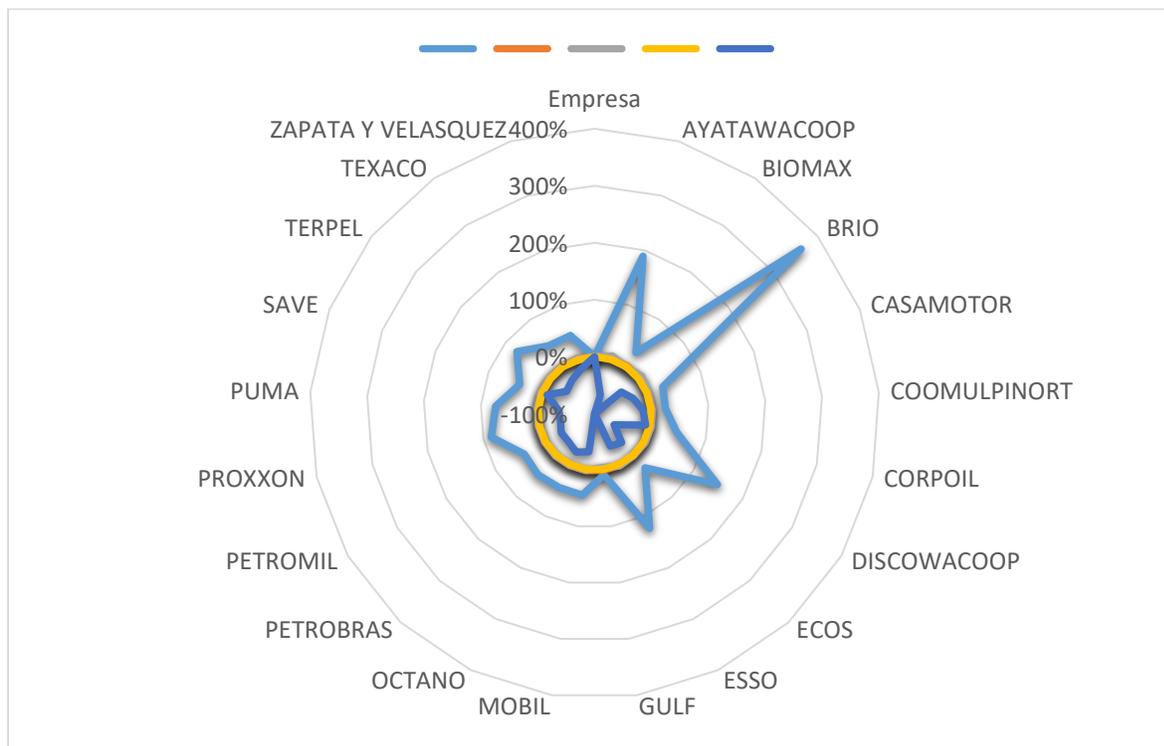
En el comportamiento de precios en el tiempo existen jugadores como GULF que dada su entrada en el mercado sistemáticamente mediante una campaña agresiva en precios su promedio es negativo en la serie, mientras que jugadores con participación importante en el mercado como TERPEL mantiene tasas de variación de precios cercanas al 0,61% como se puede apreciar en la Tabla 3-5.

Empresa	máx.	percentil 75	promedio	Percentil 25	min
AYATAWACOOP	189.24%	1.12%	5.98%	-0.77%	-66.0%
BIOMAX	29.37%	1.04%	-1.72%	-1.05%	-100.0%
BRIO	363.48%	1.17%	5.90%	-1.27%	-39.6%
CASAMOTOR	28.19%	1.08%	1.15%	-1.28%	-27.1%
COOMULPINORT	24.19%	1.90%	0.77%	-1.08%	-14.9%
CORPOIL	47.76%	2.01%	2.19%	-0.85%	-8.1%
DISCOWACOOP	148.90%	2.15%	1.98%	-0.85%	-60.4%
ECOS	29.67%	0.86%	0.61%	-0.85%	-30.3%
ESSO	123.32%	1.35%	2.02%	-1.26%	-36.6%
GULF	8.32%	0.47%	-4.41%	-0.77%	-100.0%
MOBIL	44.61%	1.26%	0.79%	-0.98%	-32.0%
OCTANO	44.16%	1.04%	0.69%	-1.05%	-25.0%
PETROBRAS	46.27%	1.16%	0.80%	-1.17%	-32.6%
PETROMIL	42.43%	0.92%	0.81%	-0.89%	-32.4%
PROXXON	85.76%	0.81%	0.65%	-1.03%	-39.0%
PUMA	73.88%	1.15%	1.11%	-1.09%	-37.2%
SAVE	40.84%	1.12%	0.96%	-0.83%	-10.5%
TERPEL	74.51%	1.29%	0.65%	-1.24%	-37.3%
TEXACO	43.60%	1.26%	0.93%	-0.98%	-29.5%
ZAPATA Y VELASQUEZ	43.23%	1.52%	1.05%	-1.69%	-18.3%
ZEUSS	34.02%	0.97%	0.61%	-0.82%	-24.5%

Fuente: SICOM, información mes de enero de 2017.

La caracterización de la variación de precios se puede observar en la Figura 3-13.

Figura 3-13 Caracterización de la variación de precios mensuales por bandera



Fuente: SICOM, información mes de enero de 2017,

Como se puede observar en la Figura 3-13 la variación promedio y lo percentiles 25 y 75 están muy cercanas es decir dicha distribución es altamente concentrada en la media, con algunos outliers.

### 3.2.7 Cantidades

#### 3.2.7.1 Participación gasolina motor por departamento

Respecto a las cantidades vendidas, es factible caracterizarlas acorde al departamento, como se observa en la Figura 3-14, donde se destacan los principales mercados son Bogotá, Antioquia, Valle del Cauca y Cundinamarca, los cuales suman el 50.27% de la demanda total a nivel país.

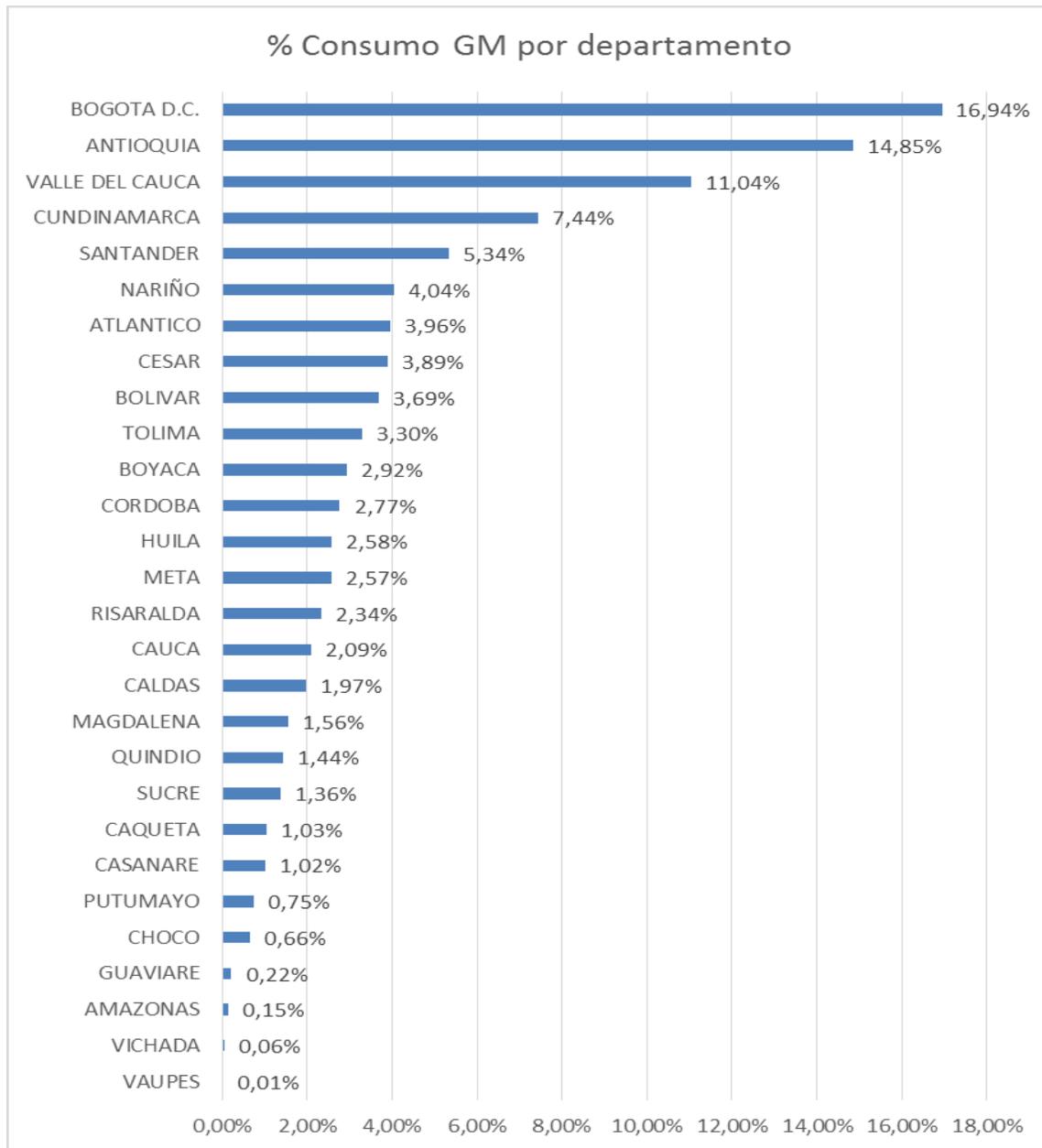
Las zonas alejadas del centro del país tales como Vaupés, Vichada y Amazonas no alcanzan a sumar el 0.21%.

Vale la pena resaltar que en Santander departamento que alberga la refinería de Barrancabermeja se demanda apenas un 5.34% del total, mientras que en el departamento del Atlántico donde se encuentra instalada la refinería de Cartagena se consume apenas un 3.96%, estos elementos ponen de relevancia la red de poliductos que conecta al país

entre los centros de consumo y los centros de oferta, infraestructura proveída por la filial de Ecopetrol, CENIT(Cenit, 2017).

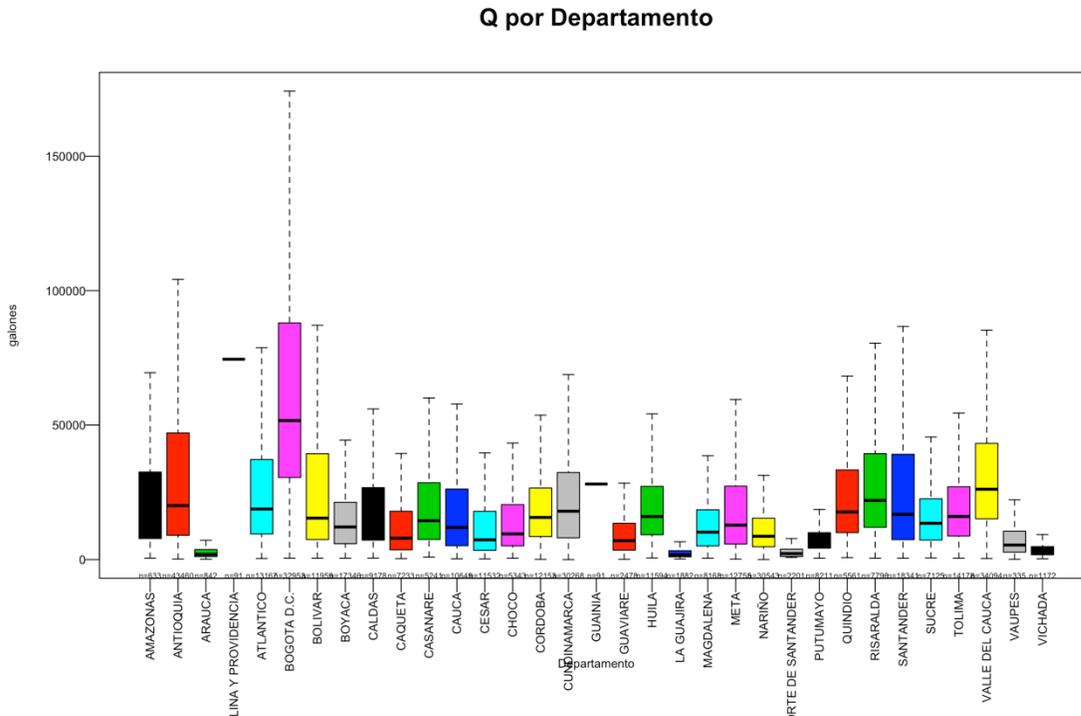
Por otro lado, como se verá más adelante el comportamiento de dichos consumos ha tenido una dinámica para los diferentes distribuidores mayoristas en los cuales algunos han ganado participación dentro del mercado como es el caso de TERPEL.

Figura 3-14 participación consumo de gasolina motor por departamento



Fuente: SICOM.

Figura 3-15 Cantidades por departamento



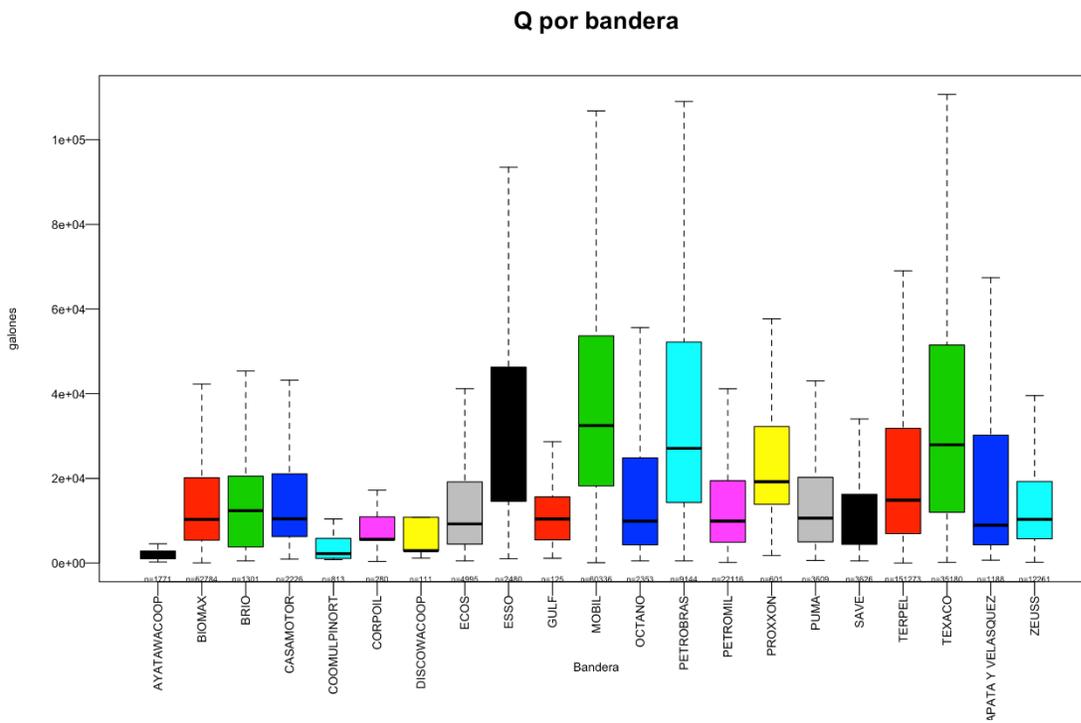
Fuente: SICOM.

Respecto al comportamiento de las cantidades vendidas por departamento es importante agregar que existe una importante varianza en los principales mercados como Bogotá y el departamento de Antioquia.

### 3.2.7.2 Evolución demanda por distribuidor mayorista

Respecto a la demanda del volumen promedio por bandera en el periodo analizado (febrero de 2009 a enero 2016) se observa dispersiones en meses puntuales que afectaron en general a los distribuidores mayoristas en diferentes proporciones, no observando un comportamiento estratégico en la variación para capturar mayor mercado de un agente en particular como se observa en la siguiente figura

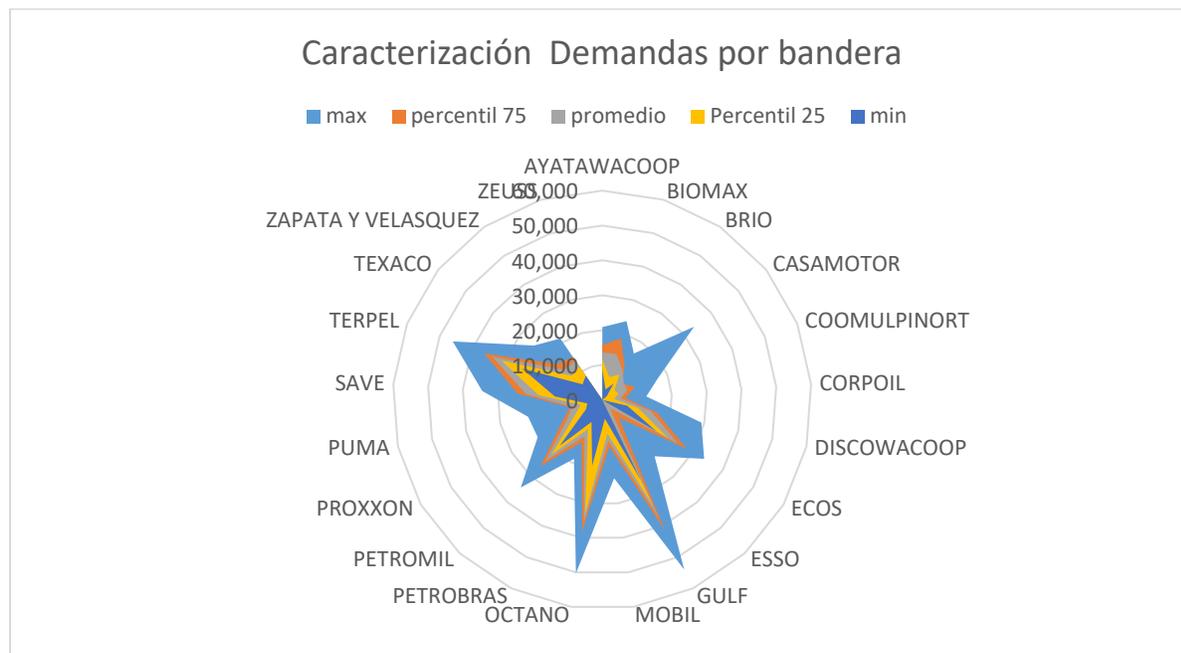
Figura 3-16 Cantidades por bandera



Fuente: SICOM.

Respecto a la caracterización de la demanda promedio para cada uno de los agentes existen agentes con un mejor desempeño tales como TERPEL, OCTANO, GULF y PETROMIL, como se puede observar en la Figura 3-17.

Figura 3-17 Caracterización demanda por bandera



Fuente: SICOM.

### 3.2.7.3 Variación de la demanda mensual por distribuidor mayorista

La caracterización de las variaciones de la demanda representadas en la Tabla 3-6, destaca que en promedio ninguna empresa ha tenido una variación negativa en su demanda, es decir es un mercado cuyo crecimiento permite que puedan crecer en general todos los agentes, aunque a diferentes ritmos

Empresa	Max.	percentil 75	promedio	Percentil 25	min
AYATAWACOOP	227.68%	18.59%	7.10%	-18.65%	-71.2%
BIOMAX	28.30%	6.91%	1.13%	-4.36%	-26.5%
BRIO	63.53%	11.07%	2.93%	-8.01%	-42.5%
CASAMOTOR	45.11%	8.62%	0.87%	-8.24%	-46.6%
COOMULPINORT	165.20%	12.35%	6.74%	-12.59%	-52.7%
CORPOIL	264.26%	42.16%	14.13%	-27.18%	-100.0%
DISCOWACOOP	211.38%	37.07%	9.80%	-26.83%	-100.0%
ECOS	37.58%	10.80%	1.33%	-7.59%	-44.4%
ESSO	33.93%	7.62%	0.74%	-7.81%	-31.2%
GULF	120.72%	30.07%	10.70%	-22.74%	-55.6%
MOBIL	34.19%	7.45%	0.78%	-5.56%	-31.0%
OCTANO	55.87%	10.35%	2.11%	-9.29%	-42.8%

Empresa	Max.	percentil 75	promedio	Percentil 25	min
PETROBRAS	49.90%	6.53%	1.06%	-4.17%	-42.6%
PETROMIL	28.06%	6.74%	0.95%	-4.62%	-32.4%
PROXXON	30.31%	7.24%	0.86%	-5.09%	-27.4%
PUMA	54.71%	11.64%	2.22%	-12.42%	-34.4%
SAVE	51.81%	14.11%	2.46%	-8.23%	-27.1%
TERPEL	31.07%	7.23%	0.91%	-5.46%	-27.1%
TEXACO	22.38%	6.86%	0.42%	-6.74%	-37.7%
ZAPATA Y VELASQUEZ	135.06%	13.91%	2.30%	-13.24%	-46.3%
ZEUSS	54.91%	5.70%	0.97%	-4.14%	-45.0%

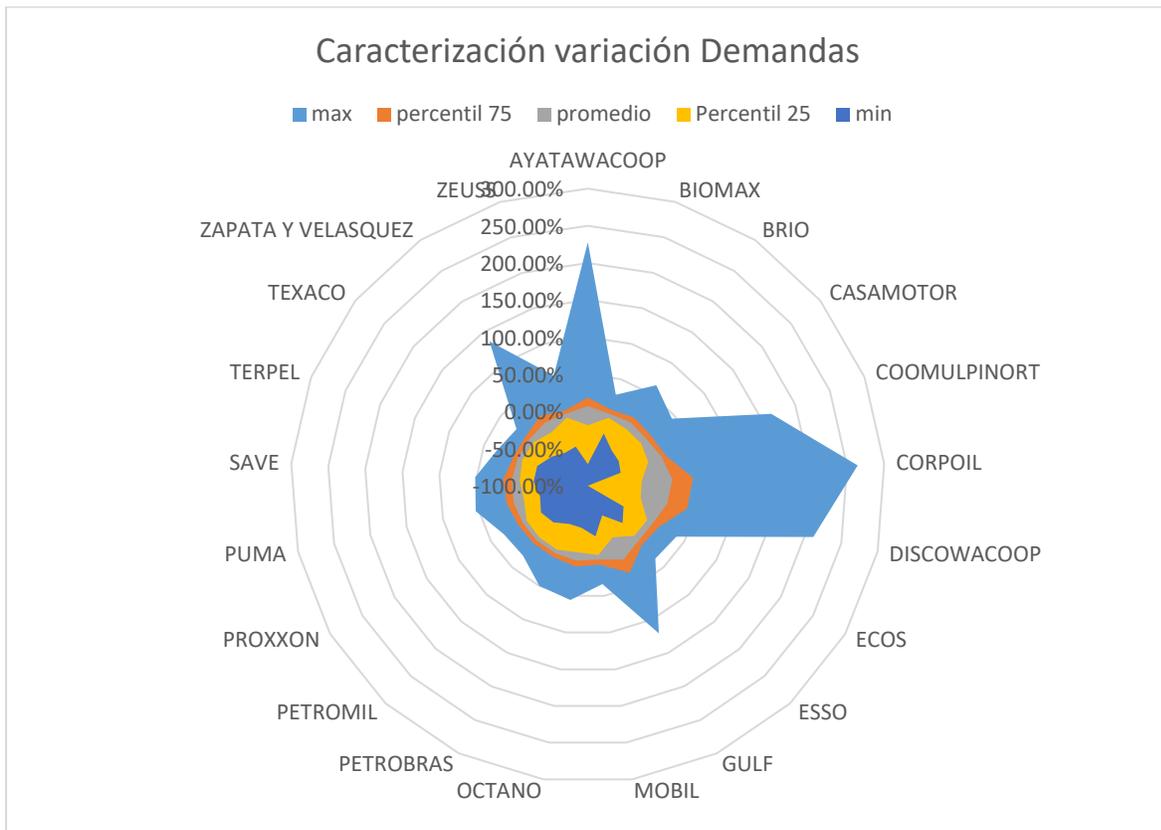
Fuente: SICOM.

La caracterización de la variación mensual de la demanda se puede observar en la Figura 3-18, donde se destaca que salvo casos puntuales de distribuidores mayoristas con mercados de baja participación o cautivos tiene fuertes variaciones, aunque igual que en el caso de los precios pueden ser debidos a falencias en el reporte de información.

Respecto a las demandas de los agentes con mayor participación del mercado sus variaciones son concentradas en la media dadas las cercanías de los percentiles 25 y 75.

En promedio las variaciones de las demandas en las estaciones de servicio de agentes con amplia participación en el mercado son cercanas al 1%.

Figura 3-18 Caracterización de la variación de demanda



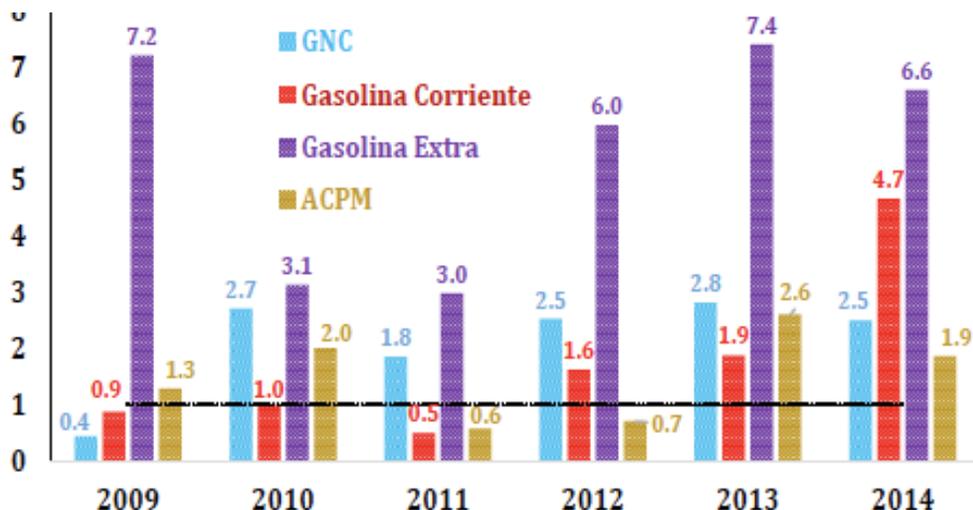
Fuente: SICOM.

El comportamiento de la evolución en el periodo analizado de la demanda promedio por bandera, agentes como Terpel, Texaco y Mobil, han mantenido en el periodo analizado su participación, contrario a lo que ha ocurrido con la firma Zapata y Velásquez.

### 3.2.8 Elasticidad precio demanda

En el presente trabajo se revisó el comportamiento de la elasticidad precio de la demanda a nivel de estación de servicio, considerando dos análisis, para cada uno de los años 2010-2017 y para cada uno de los departamentos. Existen trabajos sobre el comportamiento de la demanda (UPME, 2015) cuyos resultados se presentan en la Figura 3-19.

Figura 3-19 Elasticidad precio -Demanda

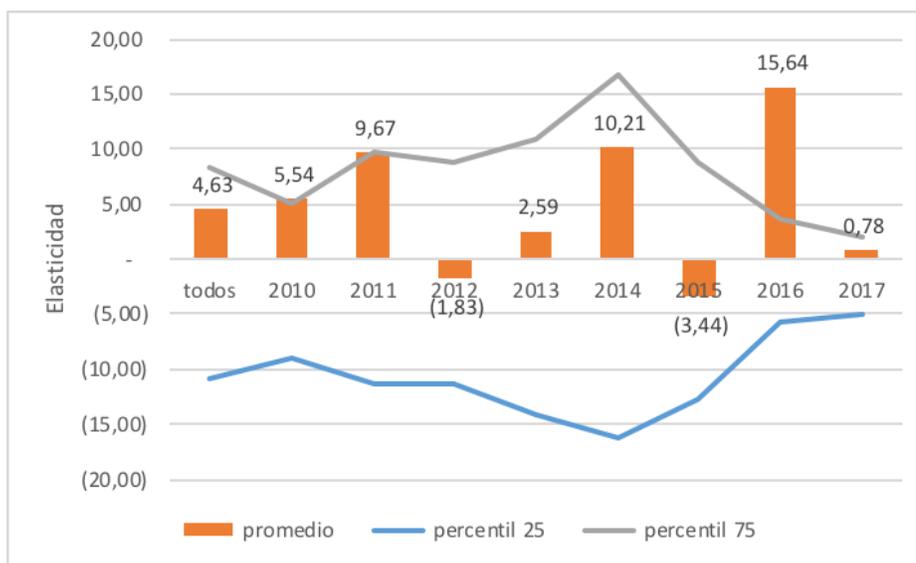


Fuente: UPME (2015)

Ahora bien, en los trabajos de la UPME cuyo resultado se expone en la Figura 3-19, el cálculo se hace a partir de una muestra de precios en las estaciones de servicio recopilados en campo, en un mes determinado del año, existen otros estudios cuyo comportamiento agregado presentan un comportamiento inelástico para la gasolina motor. (García, Pérez, Orrego, Castaño, & Mauro, 2016).

En el presente este trabajo se considera el reporte mensual para cada uno de los años y todas estaciones de servicio habilitadas en el SICOM, por ello es probable que los valores de elasticidades difieran entre los estudios.

Figura 3-20 Comportamiento de elasticidad precio de la demanda a nivel de estación de servicio para diferentes años



Fuente: SICOM, elaboración propia

A partir de la Figura 3-20, se puede señalar que:

- a) La elasticidad precio de la demanda de la gasolina motor a partir de la información de precios y cantidades a nivel de estación de servicio para el agregado nacional, se comporta de manera elástica excepto para el año 2017, teniendo como un promedio de 4,63.
- b) El comportamiento de la elasticidad promedio presenta sus mayores valores los años 2014 y 2016.
- c) Al comparar los resultados frente al reporte que presenta la UPME, se observan coincidencias en que años como el 2013 y 2014, respecto al comportamiento elástico de la gasolina motor, sin embargo, es importante tener en cuenta que en el caso de la UPME se estima a partir de una muestra de estaciones en un mes definido del año mientras que en el presente estudio se calcula a partir de la base de datos del SICOM.

Ahora bien, adicionalmente se calculó la elasticidad por departamento considerando la información de cada una de las estaciones de servicio tal como se presenta en la **Tabla 3-7**.

<b>Tabla 3-7 Comportamiento de elasticidad precio de la demanda a nivel de estación de servicio a nivel departamental</b>			
<b>Departamento</b>	<b>percentil25</b>	<b>promedio</b>	<b>percentil75</b>
Total	(10.82)	4.63	8.25
NORTE DE SANTANDER	-	0.08	-
ARCHIPIELAGO DE SANANDRES, SANTA CATALINA Y PROVIDENCIA	-	0.22	-
GUAINIA	-	0.45	0.67
VALLEDELCAUCA	(10.15)	(0.68)	7.51
ARAUCA	-	(0.68)	-
CAUCA	(13.93)	1.03	10.39
TOLIMA	(13.63)	(1.05)	8.16
RISARALDA	(8.91)	(1.21)	7.13
SUCRE	(12.46)	1.70	10.91
CALDAS	(11.62)	1.86	8.94
BOLIVAR	(9.78)	2.13	8.60
QUINDIO	(10.78)	2.26	8.19
SANTANDER	(11.02)	(2.51)	7.43
BOGOTAD.C.	(10.13)	2.56	6.56
HUILA	(11.91)	3.24	8.03
PUTUMAYO	(5.18)	4.27	5.01
ANTIOQUIA	(10.12)	4.31	7.01
MAGDALENA	(13.64)	5.15	10.52

<b>Tabla 3-7 Comportamiento de elasticidad precio de la demanda a nivel de estación de servicio a nivel departamental</b>			
<b>Departamento</b>	<b>percentil25</b>	<b>promedio</b>	<b>percentil75</b>
CORDOBA	(10.63)	7.29	9.18
CESAR	(4.80)	7.86	4.57
NARIÑO	(6.67)	8.39	6.67
ATLANTICO	(8.00)	8.53	7.71
CUNDINAMARCA	(13.63)	8.68	10.67
CHOCO	(14.15)	8.69	13.61
META	(15.33)	8.94	11.80
VAUPES	(18.03)	(9.26)	31.02
CAQUETA	(14.72)	9.32	10.68
CASANARE	(15.09)	9.46	10.57
VICHADA	(10.82)	10.10	8.78
GUAVIARE	(21.27)	11.48	18.88
BOYACA	(15.48)	15.38	13.77
LAGUAJIRA	-	(89.64)	-

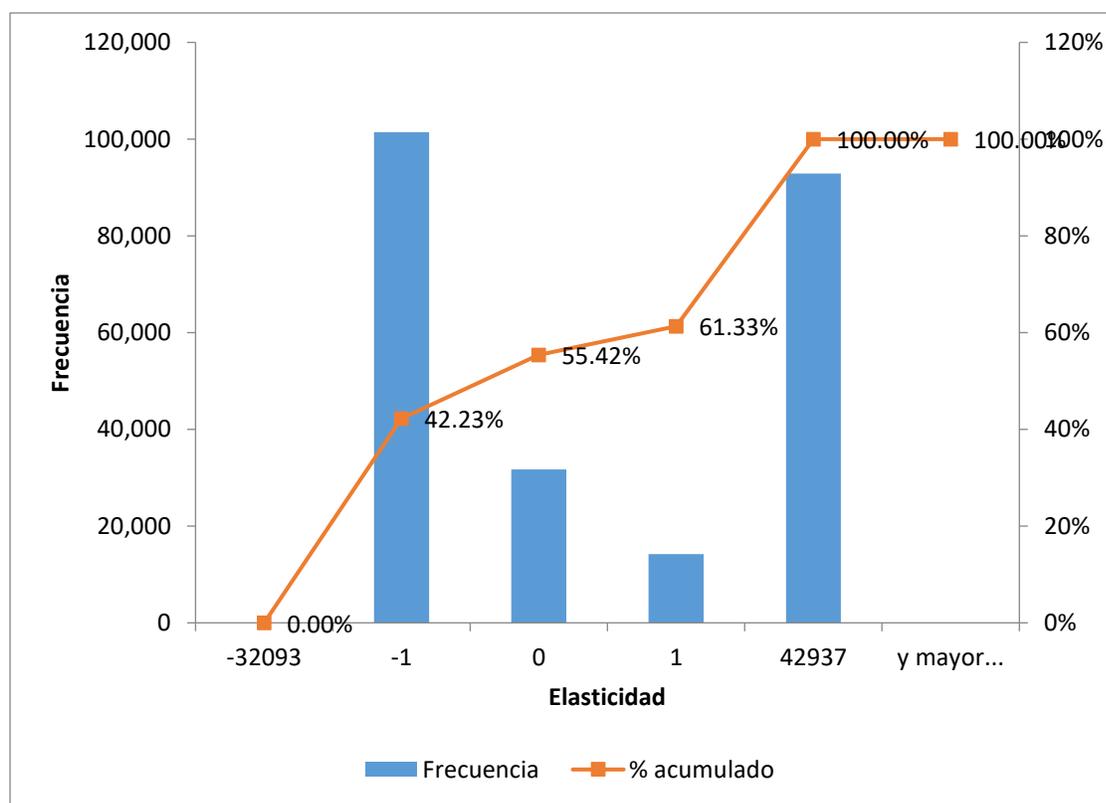
Considerando los resultados de las elasticidades a nivel departamental consignados en la **Tabla 3-7**, donde las celdas en verde muestran departamentos cuya elasticidad precio de la demanda está dentro del rango elástico y departamentos cuyo comportamiento es inelástico señalados en rojo es factible señalar que:

- a) Respecto a los departamentos que presentan un comportamiento elástico, tales como Antioquia y Cundinamarca con un amplio número de estaciones y además con puntos urbanos de alta concentración de estaciones de servicio que comercializan gasolina motor, es previsible que el consumidor pueda seguir un comportamiento elástico, debido a la gran oferta de estaciones de servicio y a las cortas distancias entre las mismas es factible desplazarse buscando una estación con mejor precio además de aplicaciones móviles (WAZE, 2018) que permiten ubicar fácil mente la estación de servicio permiten al consumidor profundizar el comportamiento elástico.
- b) Existen departamentos tales como, Norte de Santander, Archipiélago de San Andrés, Santa Catalina y Providencia, Guainia, Arauca, cuyo comportamiento de la elasticidad en el periodo analizado (2010-2017) manifiesta un comportamiento inelástico a nivel de estación de servicio lo cual podría explicarse por el bajo número de estaciones de servicio en departamentos como Guainia con apenas un 0.04% del total nacional o norte de Santander con apenas el 2.32% de las estaciones; además de ello la dispersión es un elemento relevante en la oferta de gasolina, lo cual hace que el consumidor ante la escasas de estaciones termine comprando en la que le queda cerca.

- c) En el Valle del Cauca por otro lado a pesar de contar con casi 9% de las estaciones, también se presenta un comportamiento inelástico el cual puede ser producto de la dispersión en las estaciones, además de las condiciones de la zona como puerto y complejidades sociales que presenta, los cuales pueden influir en la elasticidad precio demanda de la gasolina motor.

Finalmente al analizar el histograma del comportamiento de la elasticidad para el periodo analizado (2010-2017) considerando las estaciones registradas en el SICOM, muestra que solo el 19% de las observaciones a nivel de estaciones de servicio, presenta un comportamiento inelástico, mientras que el 81% presenta un comportamiento elástico, tal como se puede observar en la Figura 3-21.

Figura 3-21 histograma precio demanda



Fuente: SICOM.

### 3.3 Modelamiento

Durante el desarrollo se utilizaron varios modelos, entre los cuales se destacan:

- Paneles de datos in R+ con la librería plm
- Paneles con variables ponderados por distancias de matriz
- Ejercicios de conglomerados en R+

En los siguientes apartes se describen los modelos.

#### 3.3.1 Formación de matriz W y ponderación de P, Q, y PIB municipal

La estructura de la matriz W y la ponderación de las variables considerando lo señalado en el numeral 2.3, se aplicó en Matlab cuyo código está en el anexo 4 dado que la matriz cuadrada de 5.254 filas, la cual se incluye en anexo digital en el Anexo 5

#### 3.3.2 Modelos de análisis con paneles de datos

Los modelos de para adelantar las regresiones en el panel de datos se simularon en una primera instancia sin considerar las variables ponderadas por la distancia geográfica, se incluyen en la siguiente tabla:

Tabla 3-8 Modelos sin ponderación con matriz W	
Base de información	Modelo
Total	Ecuación 29 $Caso1.th = (\log(P) \sim \log(Q) - 1$

Tabla 3-8 Modelos sin ponderación con matriz W	
Base de información	Modelo
	<p>Ecuación 30</p> $\begin{aligned} \text{caso2.PA.W} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) \\ & + \text{AYATAWACOOP} + \text{BIOMAX} + \text{BRIO} + \text{CASAMOTOR} \\ & + \text{COOMULPINORT} + \text{CORPOIL} + \text{DISCOWACOOP} \\ & + \text{ECOS} + \text{ESSO} + \text{GULF} + \text{MOBIL} + \text{OCTANO} \\ & + \text{PETROBRAS} + \text{PETROMIL} + \text{PROXXON} + \text{PUMA} \\ & + \text{SAVE} + \text{TERPEL} + \text{TEXACO} \\ & + \text{ZAPATA\_Y\_VELASQUEZ} + \text{ZEUSS} + \log(\text{N\_estaciones}) \\ & + (\text{PorcQ\_bandera\_T\_alo}) \\ & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1 \end{aligned}$
	<p>Ecuación 31</p> $\begin{aligned} \text{caso2.PA.W} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) + \text{AYATAWACOOP} \\ & + \text{BIOMAX} + \text{BRIO} + \text{CASAMOTOR} + \text{COOMULPINORT} \\ & + \text{CORPOIL} + \text{DISCOWACOOP} + \text{ECOS} + \text{ESSO} + \text{GULF} \\ & + \text{MOBIL} + \text{OCTANO} + \text{PETROBRAS} + \text{PETROMIL} \\ & + \text{PROXXON} + \text{PUMA} + \text{SAVE} + \text{TERPEL} + \text{TEXACO} \\ & + \text{ZAPATA\_Y\_VELASQUEZ} + \text{ZEUSS} + \log(\text{N\_estaciones}) \\ & + (\text{PorcQ\_bandera\_T\_alo}) + (\text{PorcQ\_bandera\_municipio\_alo}) \\ & - 1) \end{aligned}$
	<p>Ecuación 32</p> $\begin{aligned} \text{caso3.th.WR} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) + \text{TERPEL} \\ & + \log(\text{N\_estaciones}) - 1 \end{aligned}$
Mayor año	<p>Ecuación 33</p> $\text{caso4.th} = \text{phtest}(\log(P) \sim \log(Q) - 1,$

Tabla 3-8 Modelos sin ponderación con matriz W	
Base de información	Modelo
	<p>Ecuación 34</p> $\begin{aligned} \text{caso5.th.WR} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) \\ & + \text{AYATAWACOOP} + \text{BIOMAX} + \text{BRIO} + \text{CASAMOTOR} \\ & + \text{COOMULPINORT} + \text{CORPOIL} + \text{DISCOWACOOP} \\ & + \text{ECOS} + \text{ESSO} + \text{GULF} + \text{MOBIL} + \text{OCTANO} \\ & + \text{PETROBRAS} + \text{PETROMIL} + \text{PROXXON} + \text{PUMA} \\ & + \text{SAVE} + \text{TERPEL} + \text{TEXACO} \\ & + \text{ZAPATA\_Y\_VELASQUEZ} + \text{ZEUSS} + \log(\text{N\_estaciones}) \\ & + (\text{PorcQ\_bandera\_T\_alo}) \\ & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1 \end{aligned}$
	<p>Ecuación 35</p> $\begin{aligned} \text{caso6.th.WR} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) \\ & + \log(\text{N\_estaciones}) + (\text{PorcQ\_bandera\_T\_alo}) \\ & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1 \end{aligned}$
Mayor mes	<p>Ecuación 36</p> $\text{caso7.th} = \text{phtest}(\log(P) \sim \log(Q)) - 1$
	<p>Ecuación 37</p>

Tabla 3-8 Modelos sin ponderación con matriz W	
Base de información	Modelo
	$  \begin{aligned}  \text{caso8.th.WR} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) \\  & + \text{AYATAWACOOP} + \text{BIOMAX} + \text{BRIO} + \text{CASAMOTOR} \\  & + \text{COOMULPINORT} + \text{CORPOIL} + \text{DISCOWACOOP} \\  & + \text{ECOS} + \text{ESSO} + \text{GULF} + \text{MOBIL} + \text{OCTANO} \\  & + \text{PETROBRAS} + \text{PETROMIL} + \text{PROXXON} + \text{PUMA} \\  & + \text{SAVE} + \text{TERPEL} + \text{TEXACO} \\  & + \text{ZAPATA\_Y\_VELASQUEZ} + \text{ZEUSS} + \log(N\_estaciones) \\  & + (\text{PorcQ\_bandera\_T\_alo}) \\  & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1  \end{aligned}  $
	<p>Ecuación 38</p> $  \begin{aligned}  \text{caso9.th.WR} = & (\log(P) \sim \log(Q) + \log(\text{PIB\_municipal}) \\  & + \log(N\_estaciones) + (\text{PorcQ\_bandera\_T\_alo}) \\  & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1  \end{aligned}  $

Ahora bien, se incluyeron los modelos para las variables ponderadas por distancia con la matriz W, las cuales fueron el precio, la cantidad, el PIB municipal, los cuales se presentan en la siguiente tabla

Figura 3-22 Modelos ponderados con matriz W	
Base de información	Modelo
Total	<p>Ecuación 39</p> $  \text{caso1P.th} = (\log(\text{PP\_R}) \sim \log(\text{QQ\_R}) - 1  $
	<p>Ecuación 40</p> $  \text{caso2P.th.WR} = (\log(\text{PP\_R}) \sim \log(\text{QQ\_R}) + \log(\text{PIBP\_}(municipal\_R)))  $

Figura 3-22 Modelos ponderados con matriz W	
Base de información	Modelo
	$  \begin{aligned}  &+AYATAWACOOP + BIOMAX + BRIO + CASAMOTOR \\  &+ COOMULPINORT + CORPOIL + DISCOWACOOP \\  &+ ECOS + ESSO + GULF + MOBIL + OCTANO \\  &+ PETROBRAS + PETROMIL + PROXXON + PUMA \\  &+ SAVE + TERPEL + TEXACO \\  &+ ZAPATA_Y_VELASQUEZ + ZEUSS + \log(N_{estaciones}) \\  &+ (PorcQ_{bandera\_T\_alo}) \\  &+ (PorcQ_{bandera\_municipio\_alo}) - 1  \end{aligned}  $
	<p>Ecuación 41</p> $  \begin{aligned}  caso3P.th.WR = &(\log(PP\_R) \sim \log(QQ\_R) + \log(PIBP\_municipal\_R) \\  &+ TERPEL + \log(N_{estaciones}) + (PorcQ_{bandera\_T\_alo}) \\  &+ (PorcQ_{bandera\_municipio\_alo}) - 1)  \end{aligned}  $
Mayor año	<p>Ecuación 42</p> $  caso4P.th = (\log(PP\_R) \sim \log(QQ\_R) - 1)  $
	<p>Ecuación 43</p> $  \begin{aligned}  caso5P.th.WR = &(\log(PP\_R) \sim \log(QQ\_R) + \log(PIBP\_municipal\_R) \\  &+ AYATAWACOOP + BIOMAX + BRIO + CASAMOTOR \\  &+ COOMULPINORT + CORPOIL + DISCOWACOOP \\  &+ ECOS + ESSO + GULF + MOBIL + OCTANO \\  &+ PETROBRAS + PETROMIL + PROXXON + PUMA \\  &+ SAVE + TERPEL + TEXACO \\  &+ ZAPATA_Y_VELASQUEZ + ZEUSS + \log(N_{estaciones}) \\  &+ (PorcQ_{bandera\_T\_alo}) \\  &+ (PorcQ_{bandera\_municipio\_alo}) - 1)  \end{aligned}  $
	<p>Ecuación 44</p>

Figura 3-22 Modelos ponderados con matriz W	
Base de información	Modelo
	$\text{caso6P.th.WR} = (\log(\text{PP}_R) \sim \log(\text{QQ}_R) + \log(\text{PIBP\_municipal}_R) + \log(\text{N\_estaciones}) + (\text{PorcQ\_bandera\_T\_alo}) + (\text{PorcQ\_bandera\_municipio\_alo}) - 1$
Mayor mes	Ecuación 45  $\text{caso7P.th} = (\log(\text{PP}_R) \sim \log(\text{QQ}_R) - 1$
	Ecuación 46  $\begin{aligned} \text{caso8P.th.WR} = & (\log(\text{PP}_R) \sim \log(\text{QQ}_R) + \log(\text{PIBP\_municipal}_R) \\ & + \text{AYATAWACOOP} + \text{BIOMAX} + \text{BRIO} + \text{CASAMOTOR} \\ & + \text{COOMULPINORT} + \text{CORPOIL} + \text{DISCOWACOOP} \\ & + \text{ECOS} + \text{ESSO} + \text{GULF} + \text{MOBIL} + \text{OCTANO} \\ & + \text{PETROBRAS} + \text{PETROMIL} + \text{PROXXON} + \text{PUMA} \\ & + \text{SAVE} + \text{TERPEL} + \text{TEXACO} \\ & + \text{ZAPATA\_Y\_VELASQUEZ} + \text{ZEUSS} + \log(\text{N\_estaciones}) \\ & + (\text{PorcQ\_bandera\_T\_alo}) \\ & + (\text{PorcQ\_bandera\_municipio\_alo}) - 1 \end{aligned}$
	Ecuación 47  $\text{caso9P.th.WR} = (\log(\text{PP}_R) \sim \log(\text{QQ}_R) + \log(\text{PIBP\_municipal}_R) + \log(\text{N\_estaciones}) + (\text{PorcQ\_bandera\_T\_alo}) + (\text{PorcQ\_bandera\_municipio\_alo}) - 1$

### 3.3.2.1 Test efectos fijos o variables en el panel

En primera instancia, se adelantó el test de Hausman para determinar si existen efectos fijos o aleatorios en el panel, cuyos resultados se presentan en la siguiente tabla

Tabla 3-9 Test de Hausman sin considerar ponderaciones

	<b>Estadístico</b>	<b>p.value</b>	<b>efectos</b>
Total base: caso1.th	87808.583	0	fijos
Total base: caso2.th.WR	3569.373	0	fijos
Total base: caso3.th.WR	196623.656	0	fijos
Max. year: caso4.th	87808.583	0	fijos
Max. year: caso5.th.WR	3569.373	0	fijos
Max. year: caso6.th.WR	82594.914	0	fijos
Max. mes: caso7.th	87808.583	0	fijos
Max. mes: caso8.th.WR	3569.373	0	fijos
Max. mes: caso9.th.WR	82594.914	0	fijos

Así mismo para los casos con ponderación por la matriz de distancias W se adelantó el test de Hausman, donde al igual que en el caso sin ponderar los modelos cuentan con efectos fijos

Tabla 3-10 Test de Hausman con variables ponderadas por matriz de distancias W

	<b>Estadístico</b>	<b>p.value</b>	<b>efectos</b>
Total <u>base</u> : caso1P.th	87808.583	0	fijos

Tabla 3-10 Test de Hausman con variables ponderadas por matriz de distancias W

---

	<b>Estadístico</b>	<b>p.value</b>	<b>efectos</b>
Total base: caso2P.th.WR	3569.373	0	fijos
Total base: caso3P.th.WR	196623.656	0	fijos
Max. year: caso4P.th	87808.583	0	fijos
Max. year: caso5P.th.WR	3569.373	0	fijos
Max. year: caso6P.th.WR	82594.914	0	fijos
Max. mes: caso7P.th	87808.583	0	fijos
Max. mes: caso8P.th.WR	3569.373	0	fijos
Max. mes: caso9P.th.WR	82594.914	0	fijos

---

### 3.3.2.2 Resultados Panel

En las siguientes tablas se resumen los resultados de las regresiones lineales del panel tanto para las variables sin ponderación como las variables ponderadas.

Tabla 3-11 Resultados regresiones modelos panel variables sin ponderar por distancia

	model Total base:caso1 .P.A.W (1)	model Total base:caso2 .P.A.W (2)	model Total base:caso3 .P.A.W (3)	model Max. year:caso4 .P.A.W (4)	model Max. year:caso5 .P.A.W (5)	model Max. year:caso6 .P.A.W (6)	model Max. mes:caso7. P.A.W (7)	model Max. mes:caso8. P.A.W (8)	model Max. mes:caso9. P.A.W (9)
log(Q)	0.021*** (0.001)	0.014*** (0.001)	0.014*** (0.001)	0.021*** (0.001)	0.014*** (0.001)	0.014*** (0.001)	0.021*** (0.001)	0.014*** (0.001)	0.014*** (0.001)
log(PIB_municipal)		0.101*** (0.001)	0.101*** (0.001)		0.101*** (0.001)	0.101*** (0.001)		0.101*** (0.001)	0.101*** (0.001)
PorcQ_bandera_T_a lo		-0.961*** (0.036)	-0.961*** (0.036)		-0.961*** (0.036)	-0.961*** (0.036)		-0.961*** (0.036)	-0.961*** (0.036)
PorcQ_bandera_mu nicipio_alo		-0.014*** (0.004)	-0.014*** (0.004)		-0.014*** (0.004)	-0.014*** (0.004)		-0.014*** (0.004)	-0.014*** (0.004)
Observations	370,181	367,252	367,252	370,181	367,252	367,252	370,181	367,252	367,252
R <sup>2</sup>	0.004	0.023	0.023	0.004	0.023	0.023	0.004	0.023	0.023
Adjusted R <sup>2</sup>	-0.010	0.010	0.010	-0.010	0.010	0.010	-0.010	0.010	0.010

Tabla 3-11 Resultados regresiones modelos panel variables sin ponderar por distancia

	model Total base:caso1 .PA.W (1)	model Total base:caso2 .PA.W (2)	model Total base:caso3 .PA.W (3)	model Max. year:caso4 .PA.W (4)	model Max. year:caso5 .PA.W (5)	model Max. year:caso6 .PA.W (6)	model Max. mes:caso7. PA.W (7)	model Max. mes:caso8. PA.W (8)	model Max. mes:caso9. PA.W (9)
F Statistic	1,284.541* ** (df = 1; 365060)	2,168.421* ** (df = 4; 362152)	2,168.421* ** (df = 4; 362152)	1,284.541* ** (df = 1; 365060)	2,168.421* ** (df = 4; 362152)	2,168.421* ** (df = 4; 362152)	1,284.541* ** (df = 1; 365060)	2,168.421* ** (df = 4; 362152)	2,168.421* ** (df = 4; 362152)

Note:

Tabla 3-12 Regresión efectos fijos con variables ponderadas por matriz de distancias W

	model Total base:caso1 P.PA.W (1)	model Total base:caso2 P.PA.W (2)	model Total base:caso3 P.PA.W (3)	model Max. year:caso4 P.PA.W (4)	model Max. year:caso5 P.PA.W (5)	model Max. year:caso6 P.PA.W (6)	model Max. mes:caso7 P.PA.W (7)	model Max. mes:caso8 P.PA.W (8)	model Max. mes:caso9 P.PA.W (9)
log(QQ_R)	0.110*** (0.001)	0.094*** (0.001)	0.094*** (0.001)	0.110*** (0.001)	0.094*** (0.001)	0.094*** (0.001)	0.110*** (0.001)	0.094*** (0.001)	0.094*** (0.001)
log(PIBP_municipal_R)		0.273*** (0.005)	0.273*** (0.005)		0.273*** (0.005)	0.273*** (0.005)		0.273*** (0.005)	0.273*** (0.005)
PorcQ_bandera_T_alo		20.008*** (0.397)	20.008*** (0.397)		20.008*** (0.397)	20.008*** (0.397)		20.008*** (0.397)	20.008*** (0.397)
PorcQ_bandera_municipio_alo		-0.747*** (0.042)	-0.747*** (0.042)		-0.747*** (0.042)	-0.747*** (0.042)		-0.747*** (0.042)	-0.747*** (0.042)

Tabla 3-12 Regresión efectos fijos con variables ponderadas por matriz de distancias W

	model Total base:caso1 P.PA.W	model Total base:caso2 P.PA.W	model Total base:caso3 P.PA.W	model Max. year:caso4 P.PA.W	model Max. year:caso5 P.PA.W	model Max. year:caso6 P.PA.W	model Max. mes:caso7 P.PA.W	model Max. mes:caso8 P.PA.W	model Max. mes:caso9 P.PA.W
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observations	369,170	362,871	362,871	369,170	362,871	362,871	369,170	362,871	362,871
R <sup>2</sup>	0.028	0.046	0.046	0.028	0.046	0.046	0.028	0.046	0.046
Adjusted R <sup>2</sup>	0.014	0.032	0.032	0.014	0.032	0.032	0.014	0.032	0.032
F Statistic	10,346.500* ** (df = 1; 364051)	4,275.641*** (df = 4; 357775)	4,275.641*** (df = 4; 357775)	10,346.500 *** (df = 1; 364051)	4,275.641*** (df = 4; 357775)	4,275.641*** (df = 4; 357775)	10,346.500 *** (df = 1; 364051)	4,275.641*** (df = 4; 357775)	4,275.641*** (df = 4; 357775)

Note:

\* \*\* \*\*\* p<0.01

A partir de las regresiones se puede observar que:

- a) Considerando la base de datos al aplicar el test de Hausman se determinó que solo se presentaron efectos fijos.
- b) El efecto de ponderar los valores de precios, cantidades PIB municipal, por la matriz de distancias  $W$ , no cambia las variables significativas al correr el panel, pero si su impacto en el precio.

### 3.3.3 Ejercicios de conglomerados

Considerando que la mera ponderación geográfica con la matriz  $W$  no permitía incluir algún tipo de caracterización de mercados que clasificara las estaciones a las que se les pueda aplicar algunas reglas para fijar el margen mayorista acorde al entorno al que están expuestas.

Se incluyó el presente capítulo, en el cual se analiza el caso de un departamento en el cual se ejecuta el método jerárquico, se optimiza el número de conglomerados y a partir de los mismos se toma la semilla de los centroides para crear dos grupos diferenciados de estaciones de servicio donde los precios sean significativamente diferenciales.

Desde luego en la determinación de los conglomerados se incluyen pruebas de un nuevo panel para determinar la significancia de la asignación de grupos a las estaciones de servicio, para así confirmar que la separación en los dos grupos cuenta con significancia estadística.

Adicionalmente y como valor agregado en el Anexo 3 Resultados clúster por departamento, se incluyen otros departamentos con la información de las corridas para adelantar el análisis de conglomerados tanto con el método jerárquico y no jerárquico como su optimización.

### 3.3.4 Análisis para el departamento de Antioquia

A continuación, se presenta el análisis de conglomerados para el departamento de Antioquia.

En primera instancia de la base de datos principal, se filtra la información para el departamento de Antioquia en la siguiente tabla

Tabla 3-13 Información base de datos para determinar conglomerados en el departamento de Antioquia. (n=3,770)						
Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
SICOM	631,574.40 0	1,880.606	630,262	630,409	633,779	636,301

Tabla 3-13 Información base de datos para determinar conglomerados en el departamento de Antioquia. (n=3,770)						
Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
alo	2,013.674	2.258	2,010	2,012	2,016	2,017
mes	6.000	0.000	6	6	6	6
P	8,719.212	1,742.337	6,050	8,030	8,800	34,140
Q	34,648.620	37,163.970	200	9,585	47,197.5	280,810
PIB_municipal	2,431,356.000	4,231,723.000	4,458.684	56,424.090	1,356,601.000	14,518,250.000
PP_R	21,021,412.000	148,379,313.000	10,103.000	47,850.000	84,254.000	1,108,800,000.000
QQ_R	469,538,525.000	679,166,603.000	10,314.000	52,600.750	1,194,100,000.000	3,050,200,000.000
PIBP_municipal_R	289,739,532,960.000	141,846,715,356.000	12,860,000,000.000	221,505,000,000.000	344,440,000,000.000	876,320,000,000.000
AYATAWACOOP			0			0
BIOMAX			0			1
BRIO			0			0
CASAMOTOR			0			0
COOMULPINOR T			0			0
CORPOIL			0			1
DISCOWACOOP			0			0
ECOS			0			0
ESSO			0			1
GULF			0			1
MOBIL			0			1
OCTANO			0			0
PETROBRAS			0			0
PETROMIL			0			1
PROXXON			0			1

Tabla 3-13 Información base de datos para determinar conglomerados en el departamento de Antioquia. (n=3,770)						
Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
PUMA			0			0
SAVE			0			0
TERPEL			0			1
TEXACO			0			1
ZAPATA_Y_VEL ASQUEZ			0			1
ZEUSS			0			1
PIB_municipal.1	1,581,888.0 00	2,648,901.0 00	3,216.828	41,748.380	849,086.20 0	6,575,947.0 00
N_estaciones	30.190	41.530	1	4	21	108
PorcQ_bandera_ T_alo	0.208	0.146	0.0002	0.031	0.372	0.391
PorcQ_bandera_ municipio_alo	0.442	0.262	0.001	0.278	0.550	1.000

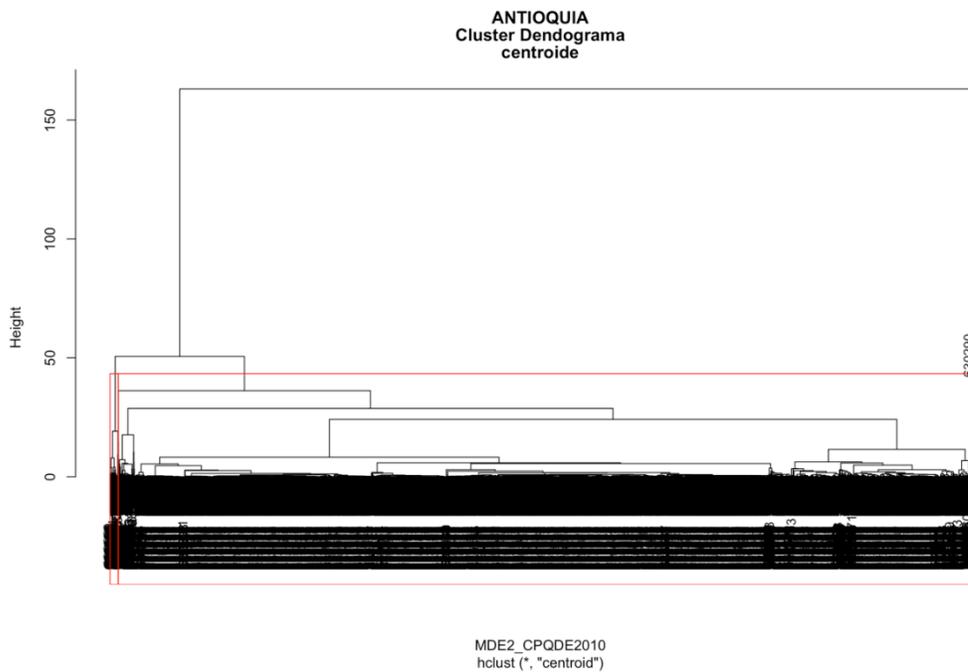
A partir de esta información es factible aplicar los métodos de formación de conglomerados descritos en el numeral 2.5.1.

### 3.3.4.1 Formación de conglomerados

A continuación, se corren los métodos:

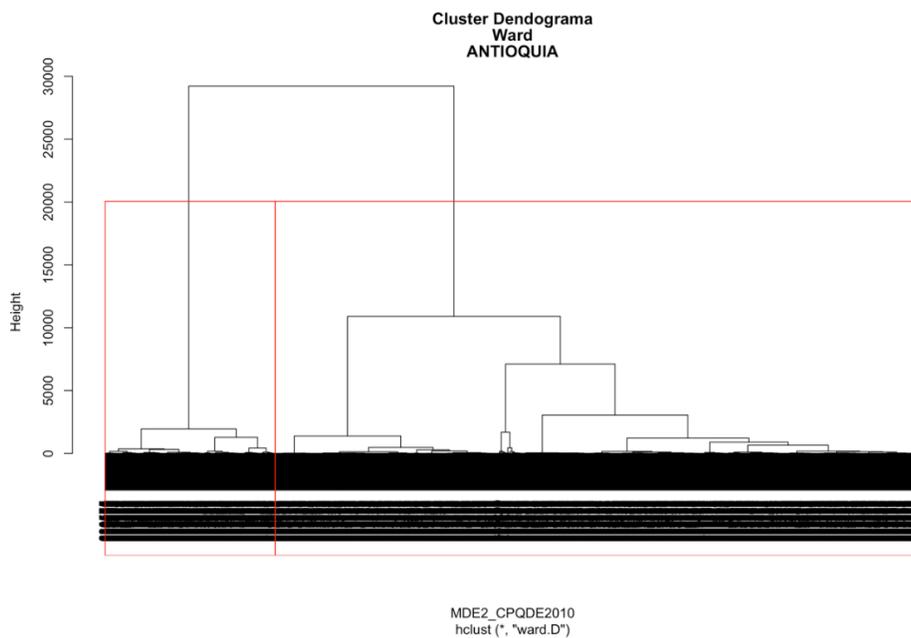
- a) Método del centroide
- b) Método del vecino más cercano
- c) Método del vecino más lejano.
- d) Método de Ward.

Figura 3-23 Agrupación con método centroide



Como se puede observar se puede caracterizar en esta agrupación básicamente dos grupos donde apenas deja un par de estaciones separadas en un grupo particular.

Figura 3-24 Agrupación con método Ward



En el método Ward se consideran también dos grupos, aunque con una distribución de tamaño similar.

Figura 3-25 Agrupación con método vecino más cercano



En esta agrupación básicamente el método muestra un solo grupo, al igual que el vecino más lejano y la vinculación promedio

Figura 3-26 Agrupación con método vecino más lejano

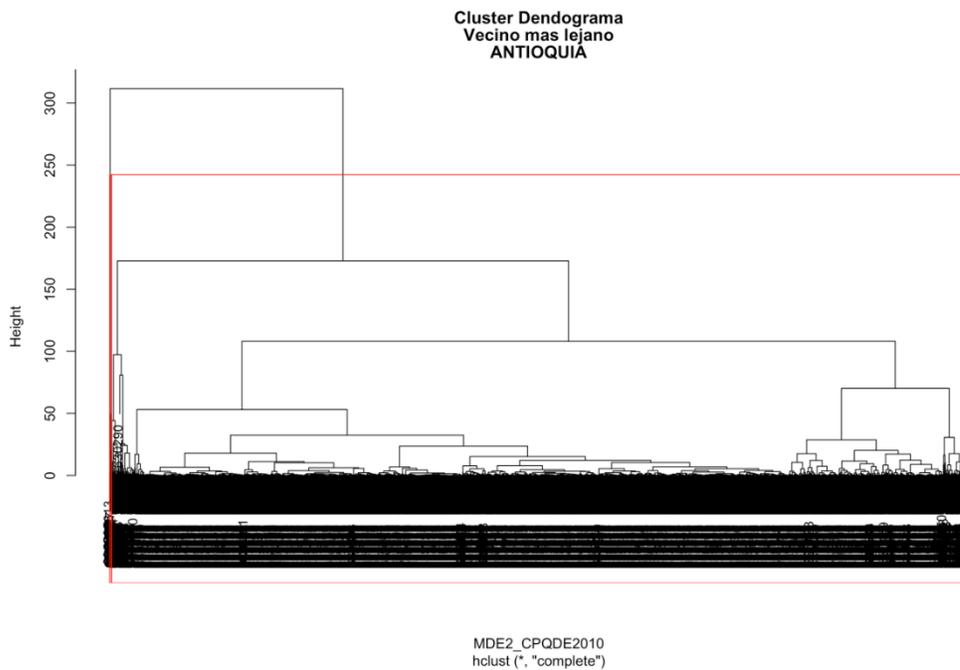
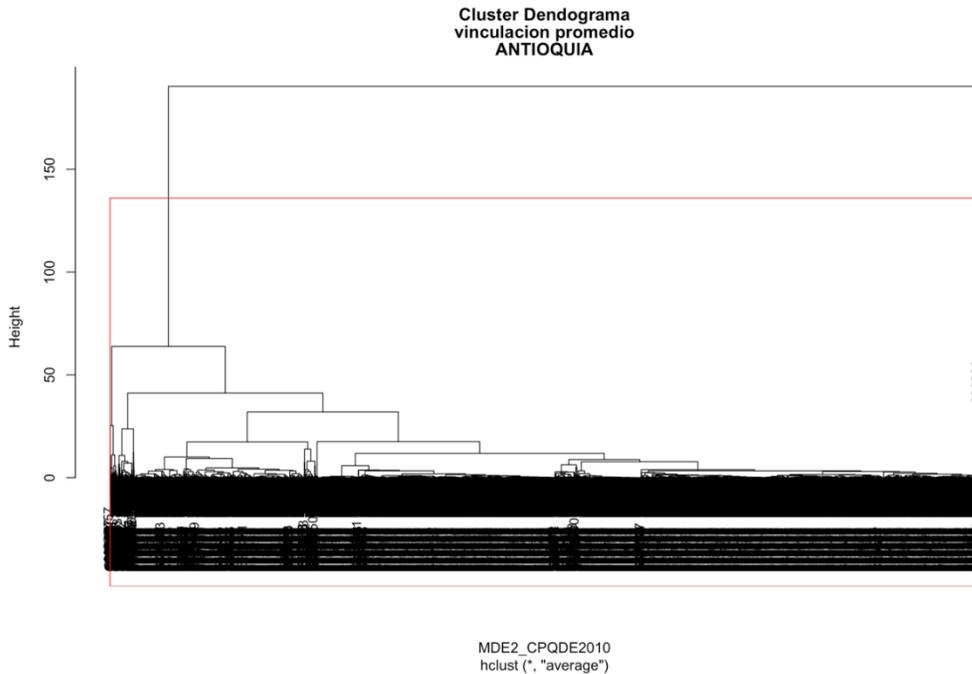


Figura 3-27 Agrupación con método vinculación promedio

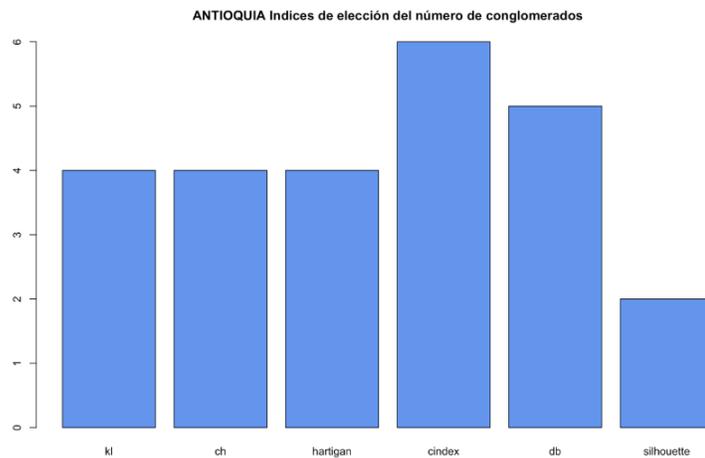


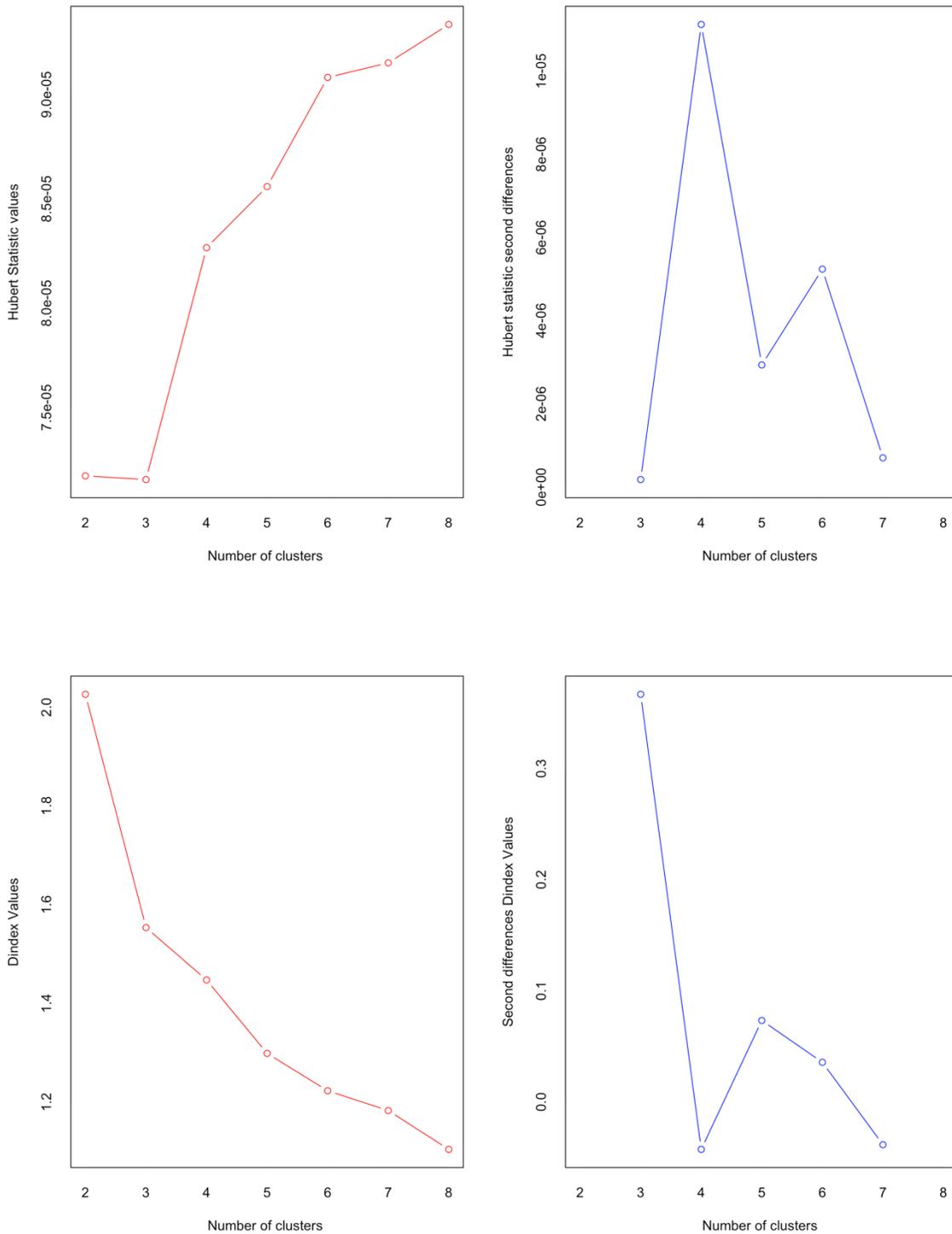
En conclusión, gráficamente el método del centroide y Ward dan indicios de formar dos grupos, en el siguiente aparte se hacen las simulaciones para seleccionar el mejor número de conglomerados.

**3.3.4.2 Selección mejor número de conglomerados método jerárquico**

Considerando los índices de CH, kl, hartigan, index, db, silhouette y los métodos gráficos Dindex y Humber, es factible tener aun mejor información para definir los grupos.

Figura 3-28 Índices de elección del número de conglomerados





A partir de los métodos de optimización se consideran que cuatro grupos arrojaría una distribución con los grupos mejor diferenciados con base en as variable incluidas y presentadas previamente.

### 3.3.4.3 Definición de dos grupos con método no jerárquico

Ahora bien tomando los centroides de la corrida de los métodos jerárquicos se aplican los métodos no jerárquicos (Aldás & Uriel, 2017), para definir dos grupos diferenciados las estaciones de servicio, se adiciona una variable a las descritas en la Tabla 3-13 y se forma el panel de datos, inicialmente se vuelve revisar si se presentan efectos fijos, aplicando el test de Hausman tal como se presenta en la siguiente tabla

Tabla 3-14 Test de Hausman con grupos 1 y 2 a partir del análisis no jerárquico

	<b>statistic</b>	<b>p.value</b>	<b>efectos</b>
Cluster_ ANTIOQUIA	12145.399	0	fijos

Se vuelve a correr el panel cuya significancia se observa en la siguiente tabla con la variable CPQDE2010.kmeans.cluster cuyos resultados muestran que es significativa, tal como se incluye a continuación.

Tabla 3-15 Antioquia significancia conglomerado

	<i>Dependent variable:</i>
	log(P)
	<i>Panel</i>
	<i>Linear</i>
log(Q)	-0.017*** (0.003)
log(PIB_municipal)	-0.017*** (0.004)
PorcQ_bandera_T_alo	0.261** (0.124)
PorcQ_bandera_municipio_alo	0.002 (0.016)
CPQDE2010.kmeans.cluster	-0.769*** (0.007)
Observations	3,770
R <sup>2</sup>	0.819
Adjusted R <sup>2</sup>	0.788
F Statistic	2,908.326*** (df = 5; 3219)

Tabla 3-15 Antioquia significancia conglomerado

*Dependent variable:*

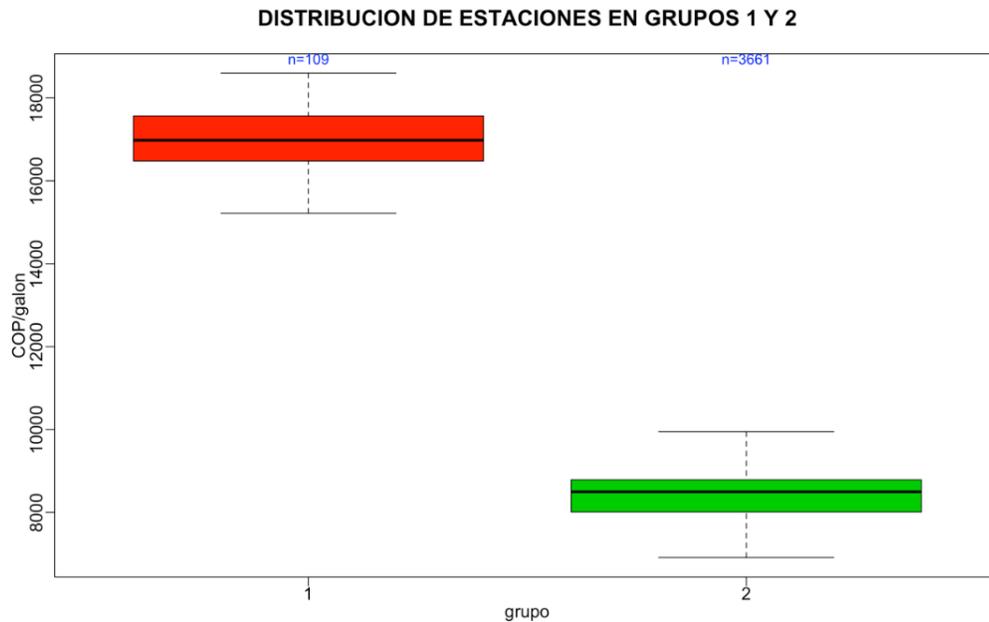
---

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

A partir del proceso de estructurar conglomerados se puede incluir una forma de clasificación con las variables de entorno para definir dos grupos de estaciones de servicio.

En la siguiente gráfica se presentan los dos grupos donde se observa claramente diferenciado el comportamiento del precio, en el 97% de las estaciones, del grupo dos, cuentan con una mediana cercana a los 8.500 COP mientras que el grupo uno con apenas el tres por ciento de las estaciones tiene una mediana de 17.000 COP.

Figura 3-29 Grupos de estaciones y el comportamiento del precio



Fuente propia

## 4 Resultados

En los resultados se destaca que:

- Dentro de la investigación desarrollada en el presente estudio, se identificaron las interfaces de programación de aplicaciones – API, así como el algoritmo para determinar las coordenadas de latitud y longitud para lograr estructurar una matriz de distancias  $W$ , es decir se ubicaron espacialmente 5.254 estaciones de servicio dentro de las guías de desarrollador de Google.(Google, 2018).
- A partir de la información entregada por el Ministerio de Minas y Energía, la cual incluye la serie de información de precios, cantidades, la “bandera” y dirección de las estaciones de servicio, se realizó un proceso de la depuración de información el cual arrojó una base de datos panel de 478,114 observaciones (ver numeral 3.1), el cual permitió construir la matriz de distancia  $W$ .
- Se construyó la matriz de distancia entre las estaciones de servicio, la cual permite integrar la perspectiva espacial a los modelos econométricos. Dicha matriz cuadrada cuenta con 5.254 filas y columnas, la cual además de la aplicación en este trabajo, puede considerarse en trabajos posteriores para temas tales como, análisis de la cadena logística en el abastecimiento de las estaciones de servicio.
- Se construyó la serie de producto interno bruto per cápita municipal, considerando el indicador de importancia económica municipal y la serie de producto interno bruto per cápita departamental, con el fin de incluirlo dentro de las variables explicativas en el panel de datos. Esta construcción muestra que el 96% de los municipios tiene un PIB per cápita inferiores a 3.781.000 COP por año, resultado concordante con el indicador de desigualdad GINI del 0,49. (Banco Mundial, 2018).

Este indicador al igual la matriz de distancias puede convertirse en un insumo para múltiples análisis posteriores en el planteamiento de señales de política, específicamente en temas de transferencias.

- Se ajustaron y aplicaron los algoritmos en R+ (Aldás & Uriel, 2017) para modelar los conglomerados para agrupar las estaciones en el mercado de la gasolina motor. Se utilizaron los métodos jerárquicos y no jerárquicos, utilizando la variable precios para realizar dicha caracterización, lo que permite mostrar que las estaciones se pueden agrupar en dos grandes grupos en cada departamento, uno donde se esperaría un comportamiento más competitivo por la cercanía entre las estaciones de servicio los cuales permitieron definir grupos de estaciones de servicio que

pueden ser un insumo desde la perspectiva regulatoria, en la fijación de reglas de comercialización minorista.

- Dentro de la caracterización estadística de las estaciones de servicio, se puede señalar que en el mercado colombiano, la ciudad con mayor cantidad de estaciones es Bogotá con 404 estaciones; el promedio de estaciones por municipio es de 91 y existen poblaciones solo con una estación de servicio. Combinado este efecto con la distribución del mercado donde cinco jugadores concentran el 84% de las estaciones, a la luz de la teoría de oligopolios descrita en el numeral 2.2.4, es factible considerar los casos de diferenciación por distancia, los cuales están modelados en los conglomerados determinados acorde al numeral 3.3.3 y cuyo marco teórico se presentó en el numeral 2.5, de tal manera que los resultados muestran la clara diferenciación de precio en las estaciones.
- Como resultados del modelo panel, se destaca que las cantidades vendidas de gasolina a nivel de estaciones de servicio es significativo como variable explicativa en los modelos estudiados y contribuye entre un 1,4% y 2,21% en el precio final de la gasolina, mientras que el PIB per cápita municipal aporta un 10% para las series que no están ponderadas por distancia mediante la matriz  $W$ . Cuando se ponderan por la distancia utilizando la matriz  $W$ , las cantidades pueden tener un impacto en el precio entre un 9,4% y un 11% y el PIB per cápita municipal de un 27,3% en el precio acorde a lo que se presentó en el numeral 3.3.2.
- Al comparar el comportamiento agregado por departamento de los precios, entre el decil 10% y 75%, en el número de estaciones por municipio se puede observar el supuesto de que con la menor cantidad de estaciones existan mayores precios de gasolina, con unas variaciones cercanas al 5.8% frente al precio promedio.
- Al contrastar el comportamiento agregado por año de los precios entre el decil 10% y 75% en el número de estaciones por municipio, no se puede observar que el supuesto de que con la menor cantidad de estaciones existan mayores precios de gasolina, es decir en la escala de tiempo anual no se detecta el fenómeno analizado y es necesario hacer el análisis en periodos con resoluciones temporales menores.
- Al confrontar el comportamiento agregado por mes de los precios entre el decil 10% y 75%, del número de estaciones que atienen en una población, se puede observar que el supuesto de que con la menor cantidad de estaciones existan mayores precios de gasolina, con unas variaciones cercanas al 3%.

Como valor agregado y adicional al alcance del estudio se calcularon los conglomerados para los demás departamentos, generando dos grupos de estaciones claramente diferenciados en su nivel precio final de la gasolina motor, los cuales se resumen en la Tabla 4-1 y se integran en el Anexo 3 Resultados clúster por departamento, en dicho anexo se incluyeron las gráficas de agrupación con los diferentes métodos de formación de conglomerados.

- En todos los casos fue significativa en las regresiones, la variable clúster (CPQDE2010.kmeans.cluster) que representa el grupo al cual pertenecen las estaciones, mostrando así que la agrupación estimada es robusta en la formación de los grupos.
- En la medida que los mercados son más grandes tales como Antioquia y Tolima se encontraron mejores ajustes del  $r^2$  con valores superiores a 0.82.
- Respecto al PIB per cápita municipal como variable explicativa del precio de la gasolina, en los ejercicios de conglomerados a nivel departamental, se esperaba que tuviera significancia al hacer las regresiones de los modelos panel, lo cual se confirmó en el presente estudio. Además de ello, se esperaba que tuviera una relación positiva con el precio de la gasolina es decir su coeficiente fuera positivo en las regresiones, sin embargo, se observa que existen algunos departamentos cuya relación es inversa. La hipótesis para explicar este fenómeno es que el 91% de los componentes de formación de precios son regulados, salvo el margen minorista, el cual sólo tiene el 9% del total, lo cual restringe el margen de ajuste en el precio final.

Además, dada la inelasticidad precio de la demanda en algunos departamentos, es probable que el consumidor use el combustible aun si su ingreso real sea menor, también, existen otras externalidades como el consumo de la gasolina en cultivos ilícitos. (Serrano, 2014), las cuales impactan en la relación entre el PIB per cápita municipal y el precio.

- Respecto a las cantidades vendidas, se destaca que los principales mercados son Bogotá, Antioquia, Valle del Cauca y Cundinamarca, los cuales suman el 50.27% de la demanda total del país.
- Las zonas alejadas del centro del país tales como Vaupés, Vichada y Amazonas suman menos del 0.21% de las ventas mensuales de gasolina.
- Respecto a las variables incluidas en los conglomerados es factible afirmar que:
  - En el 35% de los departamentos analizados, existe una relación significativa entre la demanda de gasolina motor frente al precio en las estaciones. Además de ello esta variable tiene un impacto entre un 1,7% y un 4,4% en el precio de la gasolina.
  - En el 45% de los departamentos, el precio en las estaciones de servicio tiene una relación significativa con el producto interno bruto per cápita municipal, e impacta el precio entre un 1,7% y un 11% a nivel departamental.
  - El porcentaje de ventas por bandera a nivel municipal tiene una relación significativa con el precio cobrado en las estaciones en un 10% de los departamentos y tiene un impacto en el precio entre un 10% y un 18%.

Tabla 4-1 Significancia conglomerados por departamento

CASO	log(Q)		log(PIB_municipal)		PorcQ_bandera_T_alo		PorcQ_bandera_municipio_alo		Variable Clúster (CPQDE2010. kmeans.cluster)		Observations	R2	Adjusted R2	F Statistic	intervalo de confianza (+)	intervalo de confianza (-)
AMAZONAS	0.005	(0.007)	0.101**	(0.047)	0.173	-1.190			-0.158***	(0.008)	56	0.901	0.873	97.692*** (df = 4; 43)	0,014	-0,014
ANTIOQUIA	-0.017***	(0.003)	-0.017***	(0.004)	0.261**	(0.124)	0.002	(0.016)	-0.769***	(0.007)	3770	0,819	0.788	2,908.326*** (df = 5; 3219)	0,006	-0,006
ARCHIPIELAGO DE SAN ANDRES	0.010	(0.009)	0.015	(0.019)	3.235**	-1.373			0.078***	(0.010)	63	0,763	0.650	33.797*** (df = 4; 42)	0,018	-0,018
ATLANTICO	0.003	(0.005)	-0.045***	(0.009)	.466	(0.318)	0.030	(0.038)	-0.960***	(0.015)	1,169	0,808	0.772	829.186*** (df = 5; 988)	0,010	-0,010
BOGOTA D.C.	-0.026***	(0.003)			-0.332**	(0.166)	0.028	(0.111)	-0.767***	(0.009)	2,861	0,756	0.716	1,906.549*** (df = 4; 2461)	0,006	-0,006
BOLIVAR	0.008	(0.005)	0.012	(0.010)	-0.691**	(0.296)	0.004	(0.025)	0.835***	(0.013)	1,051	0,842	0.809	924.869*** (df = 5; 868)	0,010	-0,010
BOYACA	0.004	(0.007)	0.074***	(0.009)	-1.644***	(0.365)	-0.052*	(0.031)	-0.113***	(0.005)	1,492	0,316	0.195	117.005*** (df = 5; 1268)	0,014	-0,014
CALDAS	-0.020**	(0.010)	0.044***	(0.015)	0.070	(0.392)	0.074**	(0.037)	-0.109***	(0.007)	795	0,301	0.182	58.363*** (df = 5; 679)	0,019	-0,020
CAQUETA	0.002	(0.007)	0.043**	(0.020)	0.810	(0.571)	0.019	(0.045)	-0.110***	(0.008)	609	0,274	0.119	37.854*** (df = 5; 501)	0,014	-0,014
CASANARE	-0.016**	(0.007)	0.117***	(0.008)	0.049	(0.321)	0.039	(0.029)	-1.075***	(0.029)	445	0,813	0.773	318.788*** (df = 5; 366)	0,014	-0,014
CAUCA	-0.018**	(0.008)	0.023**	(0.010)	0.549	(0.393)	0.046	(0.029)	-0.108***	(0.007)	928	0,265	0.108	55.031*** (df = 5; 764)	0,016	-0,016
CESAR	-0.043***	(0.005)	-0.040**	(0.019)	1.254**	(0.585)	0.183***	(0.038)	0.867***	(0.018)	1,006	0,759	0.707	521.397*** (df = 5; 827)	0,010	-0,010
CHOCO	0.007	(0.007)	-0.015*	(0.009)	0.553	(0.601)	0.032	(0.029)	-0.869***	(0.030)	458	0,691	0.615	164.248*** (df = 5; 367)	0,014	-0,014
CORDOBA	0.010	(0.007)	-0.071***	(0.012)	0.338	(0.329)	0.044	(0.031)	0.902***	(0.015)	1,072	0,818	0.782	804.840*** (df = 5; 896)	0,014	-0,014
CUNDINAMARCA	-0.013***	(0.005)	0.006	(0.006)	0.079	(0.217)	0.016	(0.021)	-0.823***	(0.013)	2,623	0,638	0.574	785.854*** (df = 5; 2225)	0,010	-0,010
GUAVIARE	0.005	(0.011)	0.002	(0.024)	1.667*	-1.002	-0.148**	(0.065)	0.529***	(0.041)	230	0,502	0.367	36.322*** (df = 5; 180)	0,022	-0,022
HUILA	0.025***	(0.005)	0.002	(0.010)	-0.451*	(0.233)	0.014	(0.021)	0.803***	(0.017)	999	0,731	0.681	457.272*** (df = 5; 842)	0,010	-0,010
LA GUAJIRA	0.044**	(0.021)	0.095**	(0.043)	-101.398	-78.718	0.071	(0.115)	0.281***	(0.020)	149	0,675	0.542	43.580*** (df = 5; 105)	0,042	-0,040
MAGDALENA	0.006	(0.005)	-0.085***	(0.013)	0.063	(0.481)	-0.103***	(0.028)	0.860***	(0.021)	716	0,732	0.677	324.221*** (df = 5; 594)	0,010	-0,010
META	0.006	(0.004)	0.085***	(0.004)	0.165	(0.224)	0.011	(0.020)	-0.805***	(0.013)	1,106	0,825	0.791	874.355*** (df = 5; 925)	0,008	-0,008

Tabla 4-1 Significancia conglomerados por departamento

CASO	log(Q)		log(PIB_municipal)		PorcQ_bandera_T_alo		PorcQ_bandera_municipio_alo		Variable Clúster (CPQDE2010. kmeans.cluster)		Observations	R2	Adjusted R2	F Statistic	intervalo de confianza (+)	intervalo de confianza (-)
NARIÑO	-0.029***	(0.005)	0.020***	(0.008)	-0.497*	(0.264)	0.139***	(0.027)	-0.851***	(0.013)	2,654	0,7	0.636	1,021.171*** (df = 5; 2191)	0,010	-0,010
NORTE DE SANTANDER	-0.455*	(0.237)			-2.318	-3.544	0.615	(0.970)	0.895***	(0.061)	86	0,804	0.697	56.451*** (df = 4; 55)	-0,357	0,572
PUTUMAYO	-0.044***	(0.009)	0.048***	(0.016)	-1.707***	(0.491)	0.215	(0.274)	0.903***	(0.032)	714	0,593	0.520	176.235*** (df = 5; 605)	-0,017	0,018
QUINDIO	-0.042***	(0.012)			0.361	(0.412)	0.041	(0.048)	0.782***	(0.021)	484	0,776	0.739	358.985*** (df 4; 414)	-0,023	0,024
RISARALDA	-0.032***	(0.008)	-0.039***	(0.010)	0.014	(0.292)	0.064	(0.050)	0.748***	(0.018)	683	0,762	0.722	373.619*** (df = 5; 583)	-0,015	0,016
SANTANDER	-0.011**	(0.005)	-0.021***	(0.008)	-1.183***	(0.236)	0.059**	(0.025)	0.717***	(0.015)	1,514	0,639	0.571	450.226*** (df = 5; 1272)	-0,010	0,010
SUCRE	0.006	(0.007)	-0.044***	(0.012)	-0.850**	(0.390)	0.016	(0.027)	-0.980***	(0.019)	630	0,848	0.815	577.399*** (df 5; 517)	-0,014	0,014
TOLIMA	0.033***	(0.004)	-0.050***	(0.008)	0.144	(0.213)	-0.039*	(0.024)	-0.806***	(0.012)	1,228	0,822	0.792	968.537*** (df 5; 1046)	-0,008	0,008
VALLE DEL CAUCA	-0.039***	(0.004)			0.057	(0.156)	0.009	(0.026)	0.753***	(0.010)	2,983	0,686	0.634	1,398.289*** (df 4; 2556)	-0,008	0,008
VAUPES	0.017	(0.016)	0.125	(0.077)	-4.827**	-2.119			0.815***	(0.041)	25	0,974	0.961	148.795*** (df 4; 16)	-0,032	0,031
VICHADA	0.009	(0.017)	0.383*	(0.201)	-6.397*	-3.764	0.008	(0.089)	0.157***	(0.023)	85	0,437	0.261	9.923*** (df 5; 64)	-0,033	0,033

NOTA: \*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

## 5 Conclusiones y recomendaciones

Finalmente, a la pregunta de investigación ¿Será factible caracterizar el mercado de la gasolina motor para determinar el grado de competitividad entre las estaciones de servicio, de tal manera que se puedan identificar agrupaciones de estaciones a las cuales se les puede aplicar reglas de regulación diferenciales acorde a su nivel de precios teniendo en cuenta información geográfica?, se determina que es factible y su solución se desarrolló a lo largo del presente estudio, destacando la caracterización del mercado de gasolina, incluyendo modelos de datos panel en las regresiones y modelos de conglomerados, los cuales podrán utilizarse como insumo en el análisis de señales de política y regulatorias.

### 5.1 Conclusiones

- a) En este estudio se desarrolló una matriz de distancias, que permite incluir en el análisis las variables desde la perspectiva espacial, tal como se presentó en el marco teórico en los modelos de diferenciación de productos en virtud de la distancia y estructura de la matriz de distancias (numerales 2.2.4 y 2.3). Es un desarrollo innovador y pionero en el análisis del mercado de la gasolina, el cual, podrá ser un insumo para temas complementarios, tales como logística de operaciones en el sector de hidrocarburos y específicamente en el abastecimiento de gasolina, pudiendo incorporarse en trabajos como los que desarrolla la UPME.
- b) El ejercicio de conglomerados desarrollado en el estudio, podrá servir como insumo en los análisis que ejecuta tanto el Ministerio de Minas y Energía como la CREG, en los planteamientos de señales políticas y regulatorias, pudiendo aplicar reglas diferenciales, considerando las características de los grupos definidos de estaciones de servicio para cada departamento.
- c) A la luz del marco teórico sobre oligopolios, se observa el fenómeno de diferenciación de precios por el efecto de las distancias, acorde a lo expuesto en el numeral 2.5 al estructurar los conglomerados.
- d) Respecto al PIB per cápita municipal como variable explicativa del precio de la gasolina, al realizar el análisis a nivel global, se esperaba que tuviera significancia estadística al hacer las regresiones de los modelos panel, lo cual se confirmó en el presente estudio. Además de ello, se esperaba que tuviera una relación positiva con el precio de la gasolina, es decir, su coeficiente fuera positivo en las

regresiones, de tal manera que, si a nivel municipal los consumidores recibían mayor ingreso estuvieran dispuestos a pagar un mayor precio. Esta hipótesis fue confirmada en presente estudio con un impacto entre un 10% y un 27%, tal como se presenta en el numeral 3.3.2; sin embargo, al realizar los modelos de conglomerados por departamento, esta variable no necesariamente se comporta así tal como se presentó en los análisis adicionales integrados en los resultados del numeral 4 y cuya hipótesis de explicación esta asociada al comportamiento inelástico - precio de la demanda y a externalidades.

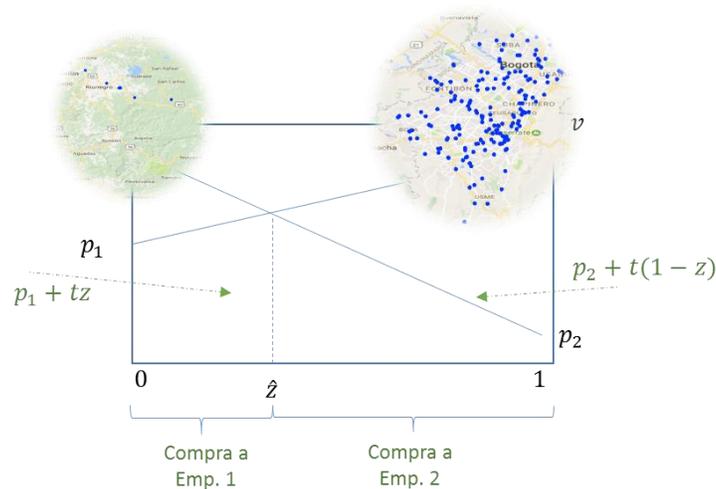
- e) Desde la perspectiva de la elasticidad - precio de la demanda, de la gasolina motor a partir de la información de precios y cantidades a nivel de estación de servicio para el agregado nacional, muestra que solo el 19% de las observaciones a nivel de estaciones de servicio, presenta un comportamiento inelástico, mientras que el 81% presenta un comportamiento elástico, tal como se puede observar en la Figura 3 21.

A nivel anual el mercado de la gasolina motor se comporta de manera elástica excepto para el año 2017, teniendo como un promedio de 4,63, el comportamiento de la elasticidad promedio presenta sus mayores valores los años 2014 y 2016. Al comparar los resultados frente al reporte que presenta la UPME, se observan coincidencias en que años como el 2013 y 2014, respecto al comportamiento elástico de la gasolina motor, sin embargo, es importante tener en cuenta que en el caso de la UPME se estima a partir de una muestra de estaciones en un mes definido del año mientras que en el presente estudio se calcula a partir de la base de datos del SICOM.

- f) A nivel departamental se presenta en general un comportamiento elástico, como es el caso de Antioquia y Cundinamarca los cuales cuentan con un amplio número de estaciones y además con puntos urbanos de alta concentración de estaciones de servicio que comercializan gasolina motor, es previsible que el consumidor pueda seguir un comportamiento elástico, debido a la gran oferta de estaciones de servicio y a las cortas distancias entre las mismas es factible desplazarse buscando una estación con mejor precio, además con aplicaciones móviles (WAZE, 2018) que permiten ubicar fácilmente la estación de servicio permiten al consumidor profundizar el comportamiento elástico.
- g) Existen departamentos tales como, Norte de Santander, Archipiélago de San Andrés, Santa Catalina y Providencia, Guainia, Arauca, cuyo comportamiento de la elasticidad en el periodo analizado (2010-2017) manifiesta un comportamiento inelástico a nivel de estación de servicio lo cual podría explicarse por el bajo número de estaciones de servicio en departamentos como Guainia con apenas un 0.04% del total nacional o norte de Santander con apenas el 2.32% de las estacione; además de ello la dispersión es un elemento relevante en la oferta de gasolina, lo cual hace que el consumidor ante la escases de estaciones termine comprando en la que le queda cerca.

- h) .El comportamiento de las cantidades vendidas muestra estacionalidad, cuyas mayores ventas se presentan en el mes de diciembre, por otra parte restringiendo la muestra a poblaciones con menor número de estaciones (decil 10%), esta estacionalidad se observa también en el mes de enero. Esta condición, en principio es atribuible a la temporada de vacaciones, dados los viajes de turismo utilizando el vehículo particular. Vale la pena mencionar que la gasolina motor es principalmente usada por vehículos livianos y es probable que los consumidores que usualmente se abastecen en las grandes ciudades, donde existe una mayor oferta de estaciones, al desplazarse deban abastecerse en poblaciones donde exista menor oferta de estaciones como los muestra el numeral 3.2.3.
- i) Considerando que el margen de comercialización minorista de las estaciones representa un 9% del valor total, es relevante el efecto de la cantidad de estaciones en una población, el cual puede representar un aumento del 64% en el precio final al consumidor, pudiendo explicarse por los costos adicionales en transporte y un margen adicional para la estación, de manera gráfica y recurriendo a lo incluido en el marco teórico en el numeral 2.2.4.1, se puede considerar mercados con una alta densidad de estaciones de servicio versus mercados de baja densidad de estaciones de servicio y su efecto en el precio, tal como se presenta en la Figura 5-1.

Figura 5-1 Efecto de la cantidad de estaciones de servicio vs precio



Elaboración propia

- j) En la medida que los mercados son más grandes tales como Bogotá y Antioquia, las cantidades vendidas son más relevantes y son más elásticas-precio de la demanda que en otros departamentos.
- k) Las señales desde la regulación y política energética podrían enfocarse a ajustar y complementar los supuestos, tales como el uso del Indicador Herfindahl-Hirschman (HHI), a la luz de la estructura de los conglomerados propuestos. De tal manera que se pueda hacer un modelo más robusto en la definición de libertad regulada o

de libertad vigilada, en las estaciones de venta de gasolina de las diferentes poblaciones.

- l) Existen cinco jugadores destacados en el mercado de gasolina, Terpel con el 37% de las estaciones, Biomax con el 16%, Mobil con el 15%, Texaco con el 9% y Petromil con el 7%, los cuales cubren el 84% de las ventas nacionales. Sin embargo, no es factible, afirmar que el comportamiento de los precios en las estaciones de servicio este principalmente relacionado con las “banderas”, considerando que los resultados de la regresión de los datos de panel donde el porcentaje de venta por bandera representaba apenas un impacto en el precio de apenas un 1,4%. Este resultado está acorde a lo presentado en el numeral 3.3.2.2, de tal manera que no es factible considerar que el precio en las estaciones a nivel nacional responde principalmente a una estrategia de marca.
- m) El efecto de ponderar los precios, cantidades y PIB per cápita municipal por la matriz de distancias  $W$ , no modifica cuáles variables son significativas al hacer la regresión de los datos de panel, sin embargo, sí cambia su impacto frente al precio pasando del 10% a un 27% tal como se presenta en el numeral 3.3.2.
- n) Desde el punto de vista de entrada en el mercado, las estaciones de servicio podrían caracterizarse como un modelo de dos etapas, como se consideró en el numeral 2.2.5, asemejando el modelo de una empresa que incurre en un costo de entrada hundido, que le permite contar con períodos posteriores de interacción competitiva. Desde la perspectiva de competencia, las empresas propietarias de las estaciones deben considerar la reacción de otras empresas a su entrada, por ello los análisis de mercado son relevantes al escoger la ubicación de las estaciones de servicio.
- o) Podría considerarse, en poblaciones con baja densidad de estaciones de venta de gasolina, que exista una estación líder y estaciones seguidoras y que su venta objetivo se determine aplicando el modelo líder -seguidor presentado en el numeral 2.2.6.

## 5.2 Recomendaciones

- a) Considerando que la información que entregó el Ministerio de Minas y Energía tiene corte de 2017, se recomienda que en caso de aplicarse el método desarrollado para agrupar estaciones y dar señales diferenciales regulatorias, se incluya información actualizada.
- b) Una variable destacada en la caracterización de las estaciones es la ubicación, por tanto, se recomienda que cada una de las estaciones sean ubicadas mediante GPS con mediciones en sitio. de tal manera que no se requiera correr un algoritmo

aproximado para determinar las coordenadas de latitud y de longitud en el caso de que se requiera desarrollar para análisis complementarios una matriz W.

- c) Considerando el trabajo de depuración de la información realizado e incluido en el numeral 3.1, es importante señalar que el SICOM, podría incluir módulos de crítica de datos para que las empresas que reportan detecten fácilmente errores de digitación y en el cargue masivo de información.



## 6 Bibliografía

- Aldás, J., & Uriel, E. (2017). *ANÁLISIS MULTIVARIANTE APLICADO CON R*.
- Arbia, G. (2014). *A Primer for Spatial Econometrics With Applications in R*.
- Banco Mundial. (2018). Índice de Gini | Data. Retrieved June 9, 2019, from <https://datos.bancomundial.org/indicador/SI.POV.GINI?end=2017&locations=CO&start=1992&view=chart>
- Bonivento, C. (2007). *Advances in control theory and applications*. Springer.
- Cenit. (2017). Cenit. Retrieved May 17, 2017, from <https://www.cenit-transporte.com/quienes-somos/>
- Damodar, & Gujarati. (2010). *Econometría*.
- DANE. (2017). Indicador de Importancia Económica Municipal. Retrieved June 22, 2018, from <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales/indicador-de-importancia-economica-municipal#información-histórica-indicador-de-importancia-económica-municipal>
- DANE. (2018). IPC - Ponderaciones gastos básico y números índice total nacional.
- Fernandez, J., & Tugores, J. (1999). *Fundamentos de Microeconomía*.
- García, J. J. ;, Pérez, D. ;, Orrego, P., Castaño, M. ;, & Mauro, J. (2016). *Un modelo Casi Ideal de Demanda de Combustibles para la Industria de Transporte*. Retrieved from [https://repository.eafit.edu.co/bitstream/handle/10784/8021/2016\\_03\\_Jhon\\_Jairo\\_Garcia.pdf?sequence=2&isAllowed=y](https://repository.eafit.edu.co/bitstream/handle/10784/8021/2016_03_Jhon_Jairo_Garcia.pdf?sequence=2&isAllowed=y)
- Google. (2018). Guía del desarrollador | Google Maps Geocoding API | Google Developers. Retrieved May 6, 2018, from <https://developers.google.com/maps/documentation/geocoding/intro?hl=es-419>
- Hal R. Varian. (1992). *Análisis microeconómico*.
- Investing. (2018). Brent Oil Futures Overview. Retrieved from <https://www.investing.com/commodities/brent-oil>
- LABYS, T. T. and W. C. (1986). Spatial Equilibrium Analysis. In *REGIONAL ECONOMICS*.
- Mas-Colell, A., & Whinston, M. (1995). *Microeconomic Theory*.

- Minminas. Resolución 1180 de 2006 (2006).
- Minminas. (2011). Decreto 4130-2011, 1–4.
- Minminas. Resolución 181254 2012 (2012).
- Minminas. (2013). Decreto 1260 de 2013 “Por el cual se modifica la estructura de la Comisión de Regulación de Energía y Gas - CREG,” 9. Retrieved from [http://wsp.presidencia.gov.co/Normativa/Decretos/2013/Documents/JUNIO/17/DECRETO\\_1260\\_DEL\\_17\\_DE\\_JUNIO\\_DE\\_2013.pdf](http://wsp.presidencia.gov.co/Normativa/Decretos/2013/Documents/JUNIO/17/DECRETO_1260_DEL_17_DE_JUNIO_DE_2013.pdf)
- Minminas. (2017). Sistema de información de combustibles líquidos. Retrieved March 13, 2017, from <http://www.sicom.gov.co/index.shtml>
- Muñoz Morales, Y. (2014). *Efecto de la liberación de los precios en el mercado de combustibles de Ecopetrol S.A.* Retrieved from <http://ezproxy.unal.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ir00238a&AN=unal.46660&lang=es&site=eds-live>
- Nicholson, W. (2005). *Teoría Microeconómica*.
- Nicholson, W. (2008). *Teoría Microeconómica. Principios Básicos y Ejercicios*. Retrieved from <http://latinoamerica.cengage.com>
- Otto, S. R., & Denier, J. P. (2005). *An introduction to programming and numerical methods in MATLAB. An Introduction to Programming and Numerical Methods in MATLAB*. <https://doi.org/10.1007/1-84628-133-4>
- Serrano, M. (2014). Cultivos ilícitos de coca y bienestar en las regiones productoras: Un análisis desde el enfoque de capacidades. *Universidad Nacional de Colombia*. Retrieved from [http://fce.unal.edu.co/media/files/documentos/Doctorado/Tesis/Tesis\\_Miguel\\_Serrano.pdf](http://fce.unal.edu.co/media/files/documentos/Doctorado/Tesis/Tesis_Miguel_Serrano.pdf)
- SIC. Decreto 1074 de 2015 (2015).
- THISSE, M. J. B. and J. F. (1987). The Location of Production Activities. In *REGIONAL ECONOMICS*.
- UPME. (2015). *Proyección de Demanda de Combustibles en el Sector Transporte en Colombia*.
- UPME. (2016). *Avance Plan de Abastecimiento de Combustibles Líquidos*. Retrieved from [http://www.upme.gov.co/SeccionHidrocarburos\\_sp/Publicaciones/2016/Avance\\_Combustibles\\_Liquidos\\_Abril\\_2016.pdf](http://www.upme.gov.co/SeccionHidrocarburos_sp/Publicaciones/2016/Avance_Combustibles_Liquidos_Abril_2016.pdf)
- Varian, H. R. (2010). *Microeconomía intermedia Un* (9th ed.).
- Varian, H. R. (2012). *análisis microeconómico*.
- WAZE. (2018). No Title. Retrieved from <https://www.waze.com>

## 7 Anexo 1 Solicitud de información SICOM

Señor  
Carlos David Beltrán  
Director de Hidrocarburos  
Ministerio de Minas y Energía  
(...)

Asunto: Derecho de petición: Solicitud de información estudio de caracterización mercados de gasolina y diésel a nivel municipal.

Respetado señor Beltrán

El grupo de investigación de geografía económica el cual dirijo, dentro de los múltiples temas de estudio ha enfocado su análisis en el mercado de combustibles líquidos, específicamente el de gasolina y diésel.

Dicho enfoque lo ha hecho a través trabajo de maestría que se encuentra en desarrollo de caracterización a nivel regional y municipal de los mercados de gasolina motor y diésel y análisis de modelos de regresión lineal espacial, así como algunos modelos no lineales, dentro de los temas a analizar y la aplicación de dichos modelos está el comportamiento de precios y cantidades con una desagregación a nivel geográfico.

Considerando que la fuente de información más relevante en estos temas es el Sistema de Información de Comercialización de Combustibles-SICOM le, que en los términos establecidos en la Constitución Política de Colombia y en especial de acuerdo con lo dispuesto en su artículo 23, solicitamos amablemente su apoyo para contar con la información de precios y cantidades a nivel municipal de cada una de las estaciones de la geografía nacional, incluyendo la siguiente caracterización:

- 1) Tipo de entrega: digital
- 2) Periodo 2010 a la fecha
- 3) Resolución temporal: mensual
- 4) Alcance geográfico: a nivel nacional
- 5) Formato: Excel
- 6) Variables:
  - a) Información de Precios:
    - i) Código SICOM
    - ii) Departamento
    - iii) Municipio
    - iv) Código DANE municipio
    - v) Productos:
      - (1) ACEM - DIESEL ECOLOGICO

- (2) ACPM - DIESEL
- (3) ACPM DIESEL - IMPORTADO
- (4) BIOACEM AL 9%
- (5) BODIESEL CORRIENTE
- (6) BODIESEL EXTRA
- (7) GASOLINA CORRIENTE
- (8) GASOLINA CORRIENTE - IMPORTADO
- (9) GASOLINA CORRIENTE OXIGENADA
- (10) GASOLINA EXTRA
- (11) GASOLINA EXTRA OXIGENADA
- vi) Precios: en COP/galón (corrientes)
- b) Información de cantidades:
  - i) Código SICOM
  - ii) Departamento
  - iii) Municipio
  - iv) Código DANE municipio
  - v) Productos:
    - (1) ACEM - DIESEL ECOLOGICO
    - (2) ACPM - DIESEL
    - (3) ACPM DIESEL - IMPORTADO
    - (4) BIOACEM AL 9%
    - (5) BODIESEL CORRIENTE
    - (6) BODIESEL EXTRA
    - (7) GASOLINA CORRIENTE
    - (8) GASOLINA CORRIENTE - IMPORTADO
    - (9) GASOLINA CORRIENTE OXIGENADA
    - (10) GASOLINA EXTRA
    - (11) GASOLINA EXTRA OXIGENADA
  - vi) Volúmenes: galones (corrientes)

Les agradezco la atención prestada y puede responder el derecho de petición a los correos electrónicos: [gajuncar@unal.edu.co](mailto:gajuncar@unal.edu.co), [gjuncar@gmail.com](mailto:gjuncar@gmail.com), [lacerog@unal.edu.co](mailto:lacerog@unal.edu.co).

(...)

La información fue remitida por parte del ministerio mediante la siguiente comunicación

Ministerio de Minas y Energía  
 Origen: DIRECCIÓN DE HIDROCARBUROS  
 Rad: 2017066327 05-10-2017 12:19 PM  
 Anexos: ARCHIVO VIRTUAL  
 Destino: JUNCA GUSTAVO  
 Serie:



383

Bogotá, D.C.

Señor  
**GUSTAVO ADOLFO JUNCA RODRIGUEZ**  
 Profesor Asociado  
 Director del Grupo de Investigación Geográfica Económica  
[gajuncar@unal.edu.co](mailto:gajuncar@unal.edu.co)

Asunto: Derecho de petición radicado No. 2017061115 del 15 de septiembre de 2017. Petición de Información.

Respetado Señor Junca:

En atención al oficio con radicado Minminas 2017061115 del 15 de septiembre de 2017 radicado por usted, se indica lo siguiente:

1. Se solicita la información correspondiente a los precios de venta por producto a nivel municipal de las estaciones de servicio para los periodos 2010 a la fecha.

La información de los precios de venta por producto de las estaciones de servicio para los años de 2010 a 2017, será enviada junto con esta comunicación al correo electrónico [gajuncar@unal.edu.co](mailto:gajuncar@unal.edu.co), dicha información fue extraída de los registros realizados por las estaciones de servicio en el SICOM.

2. Se solicita la información correspondiente a los despachos de combustibles realizados a los departamentos y municipios para los periodos 2010 a la fecha.

La información de los despachos realizados a los diferentes departamentos y municipios del territorio nacional para los años de 2010 a septiembre de 2017, será enviada junto con esta comunicación al correo electrónico [gajuncar@unal.edu.co](mailto:gajuncar@unal.edu.co), dicha información está expresada en galones/año y fue extraída del módulo de órdenes de pedido del SICOM.

Cordialmente,

  
**CARLOS DAVID BELTRÁN QUINTERO**  
 Director de Hidrocarburos

C.C. Participación Ciudadana

Presente: Edgar Andrés López  
 Recibe: Juan Pablo Pineda  
 Apellido: Carlos David Beltrán  
 ID: 111611

Calle 43 No. 57-31 CAN Bogotá, Colombia  
 Conmutador (57 1) 2200 300  
 Código postal 111321  
[www.minminas.gov.co](http://www.minminas.gov.co)

Página 1 de 1



## 8 Anexo 2 Mapas de las estaciones

En este anexo se incluyeron algunos mapas con la distribución geográfica por departamento, sin embargo, por restricciones en el uso de la licencia del paquete RgoogleMaps para graficar las funciones no fue factible incluir más mapas.

Figura 6-1 Estaciones País

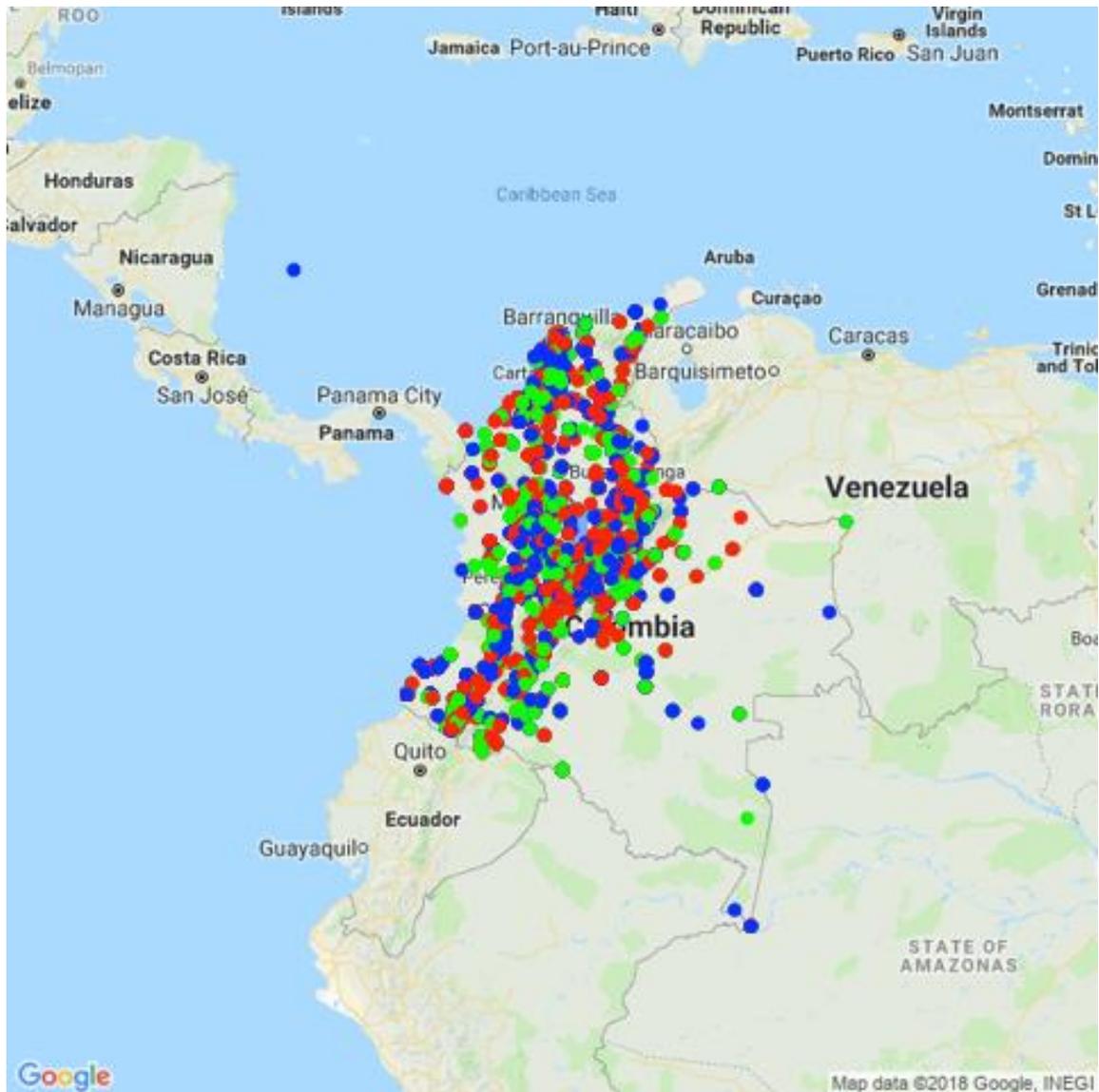


Figura 6-2 Estaciones Bolívar

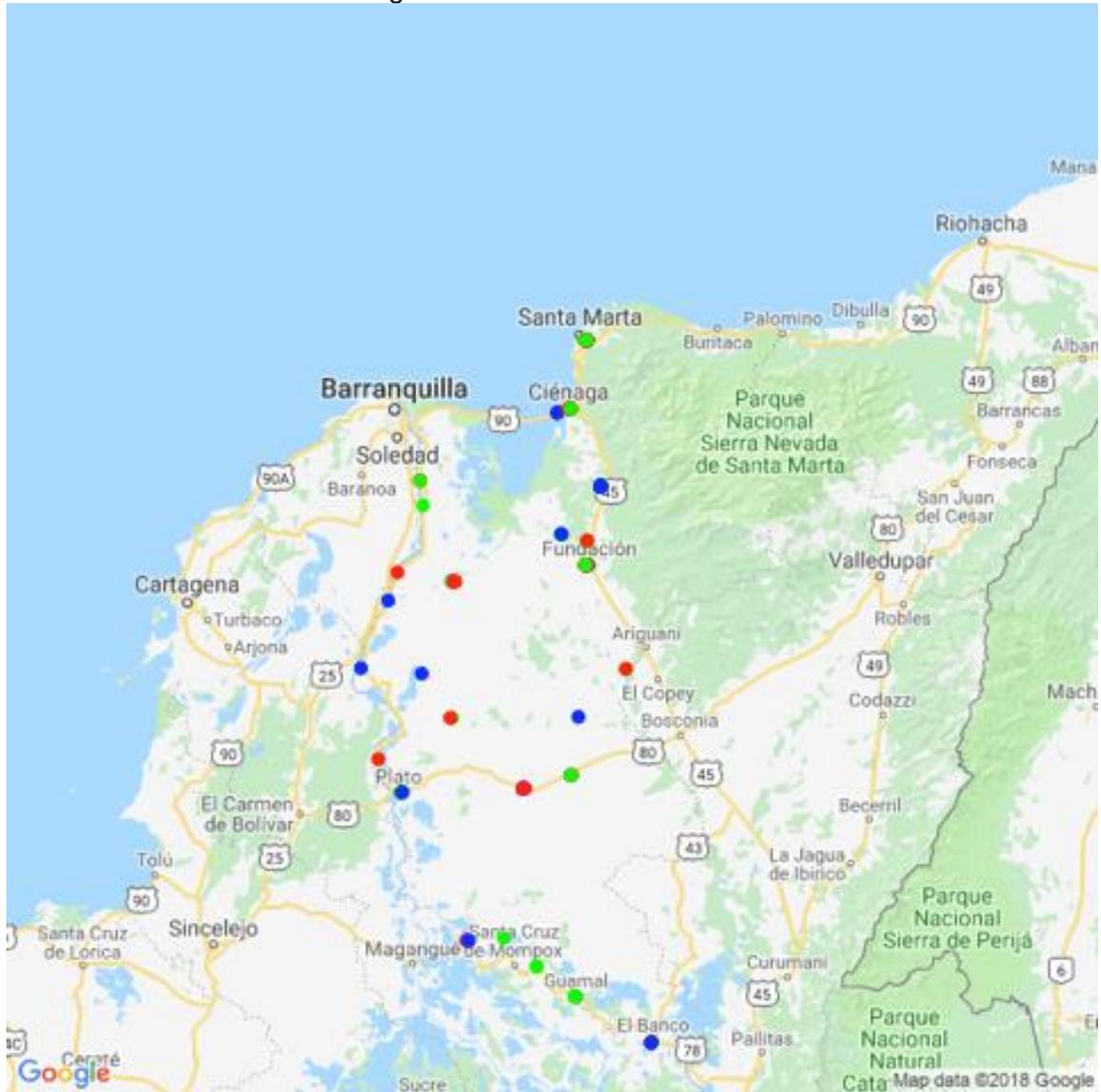


Figura 6-3 Estaciones Antioquia

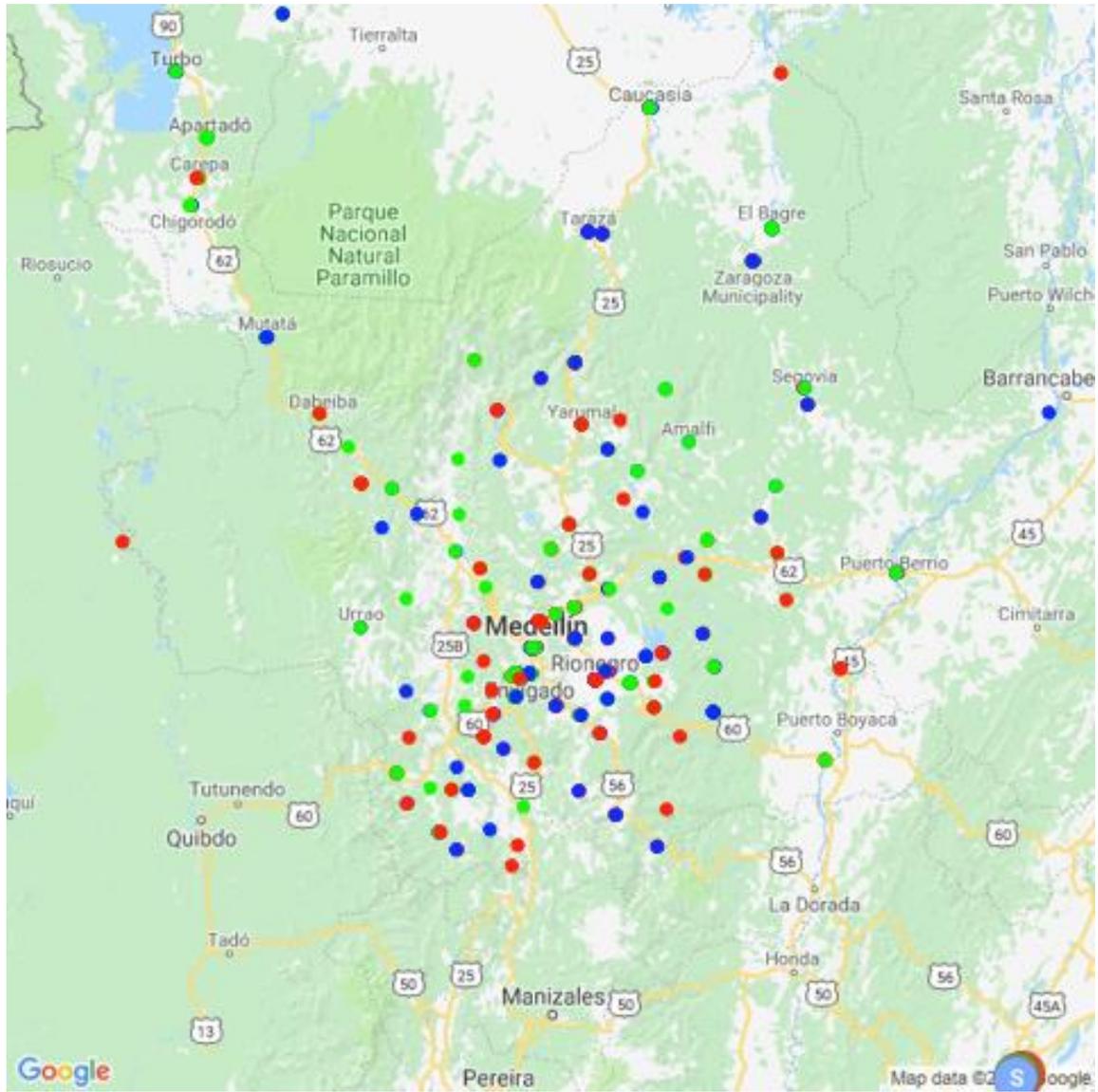


Figura 6-4 Estaciones Cundinamarca

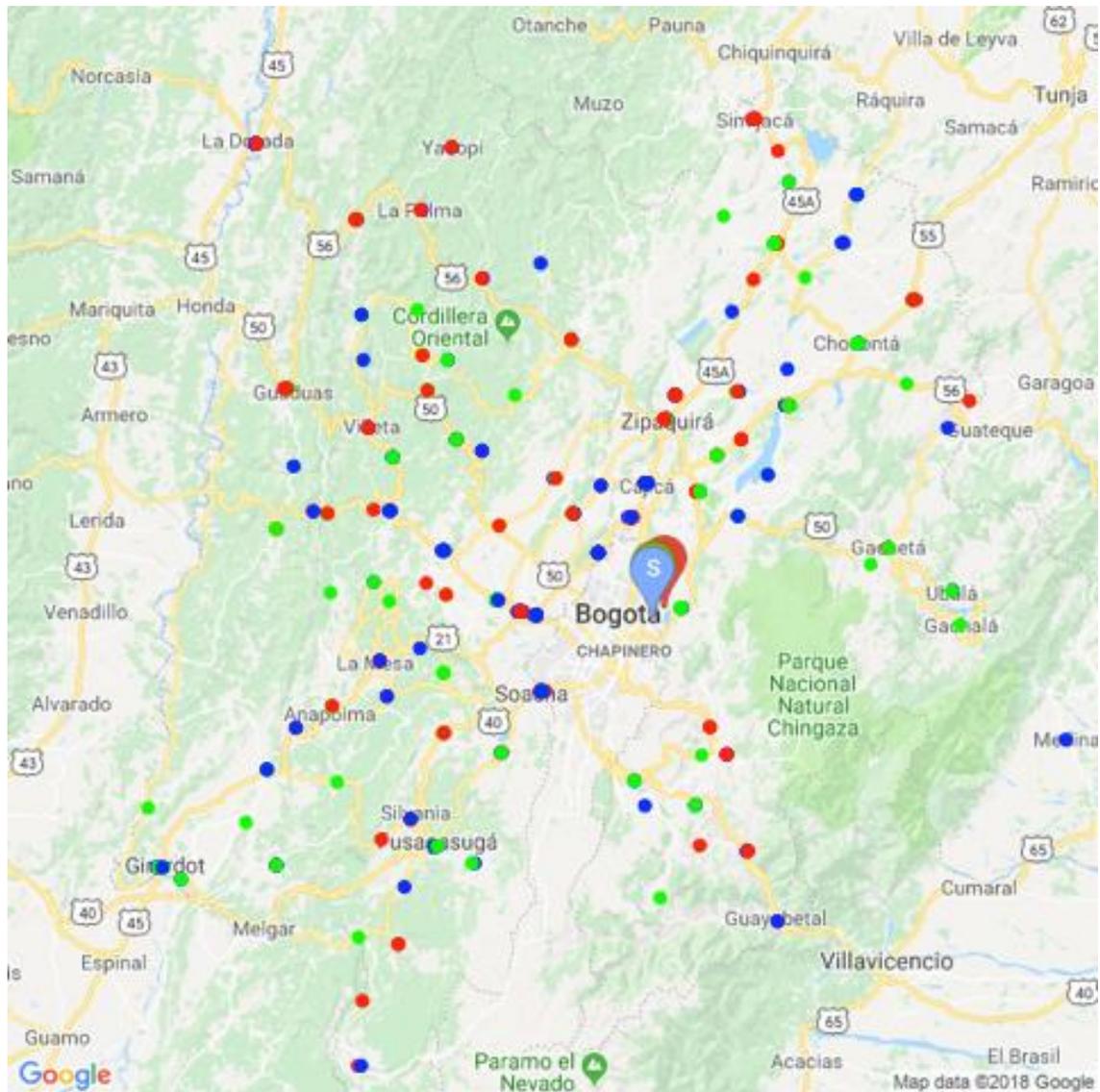
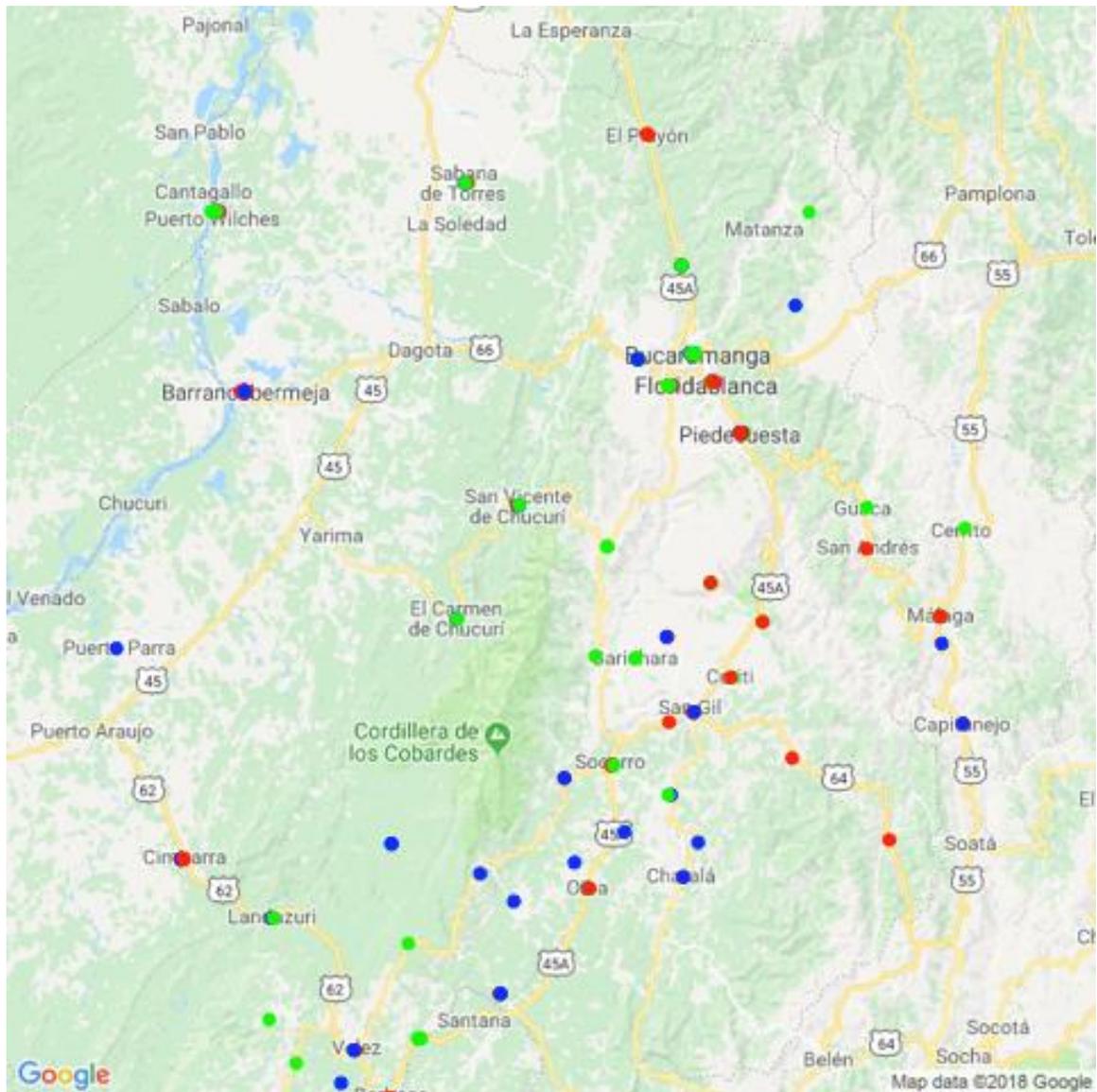


Figura 6-5 Santander

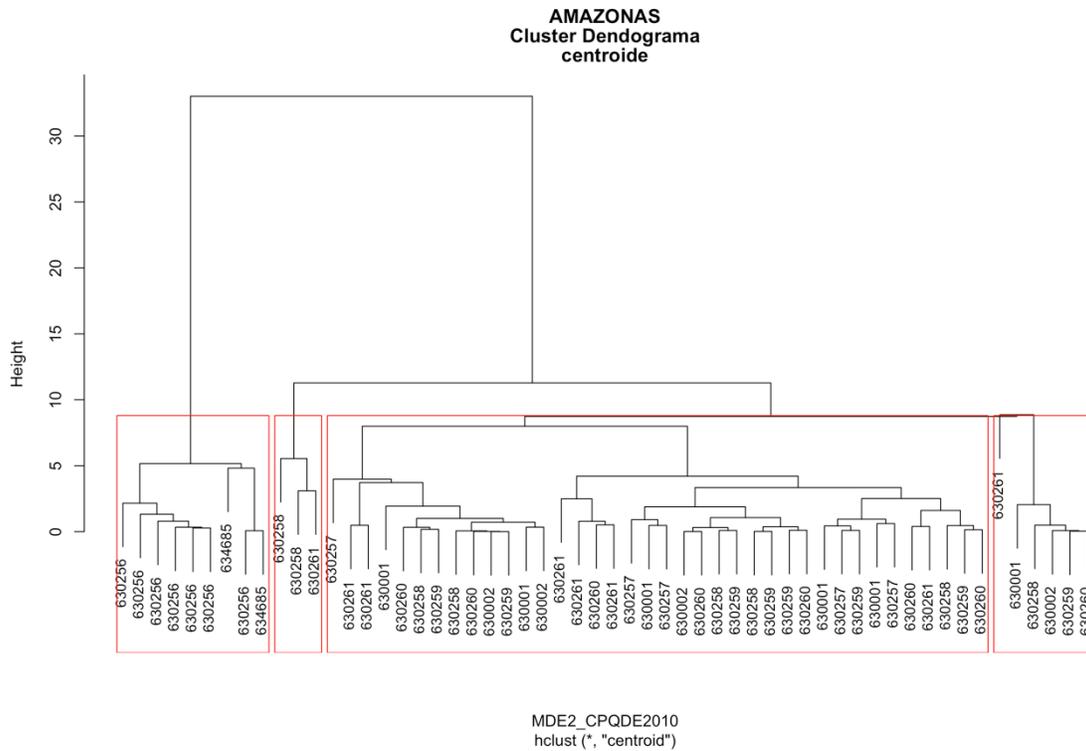


## 9 Anexo 3 Resultados clúster por departamento

En el presente anexo se incluyen los datos complementarios tanto gráficas como distribuciones finales de las estaciones de servicio para los diferentes departamentos, agrupados mediante los métodos jerárquicos y no jerárquicos en la formación de los conglomerados. Este anexo se incluye únicamente en formato digital

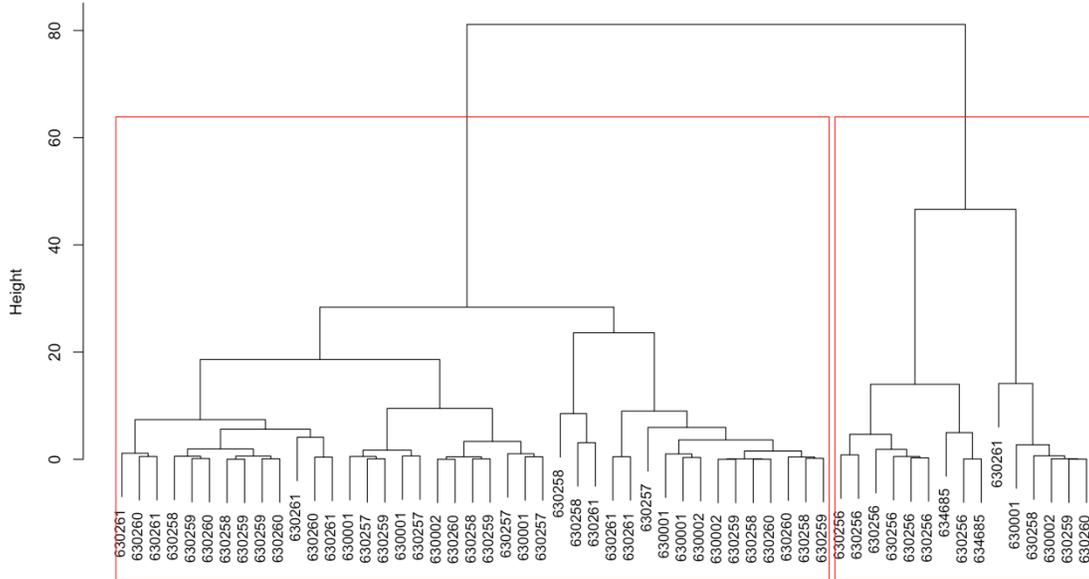
## 9.1 AMAZONAS

### 9.1.1 Resultados métodos conglomerados



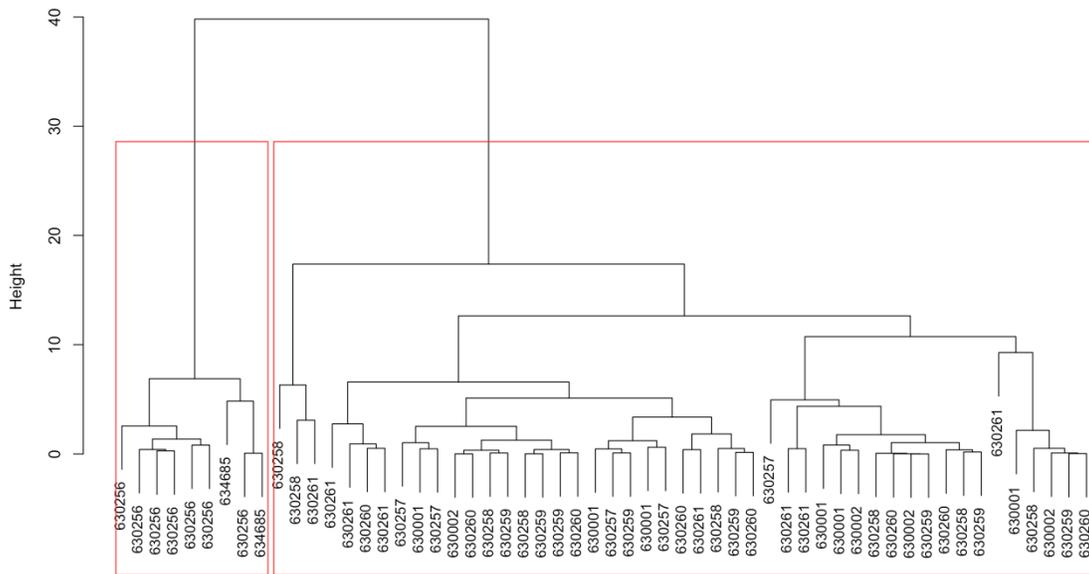


Cluster Dendrograma  
Vecino mas lejano  
AMAZONAS



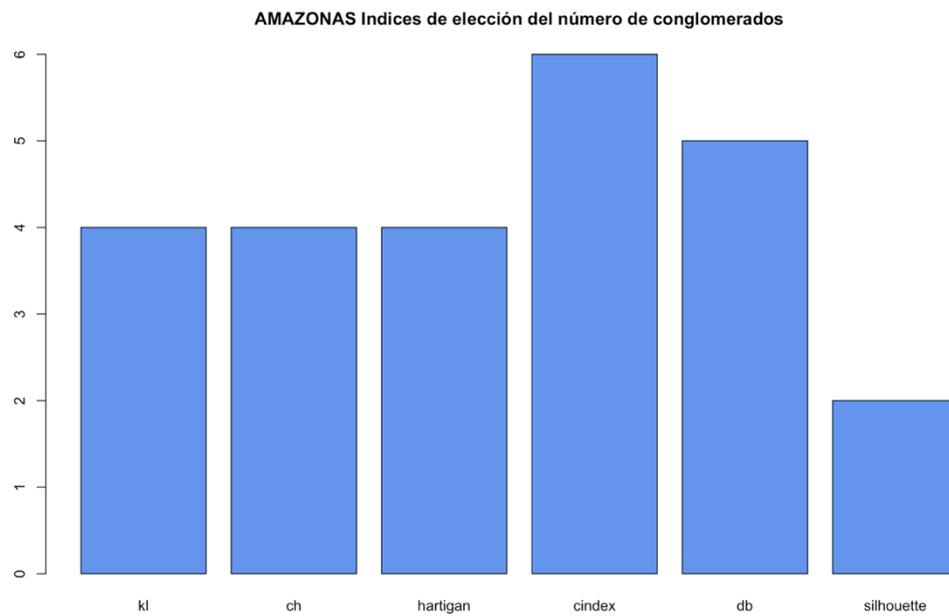
MDE2\_CPQDE2010  
hclust (\*, "complete")

Cluster Dendrograma  
vinculacion promedio  
AMAZONAS

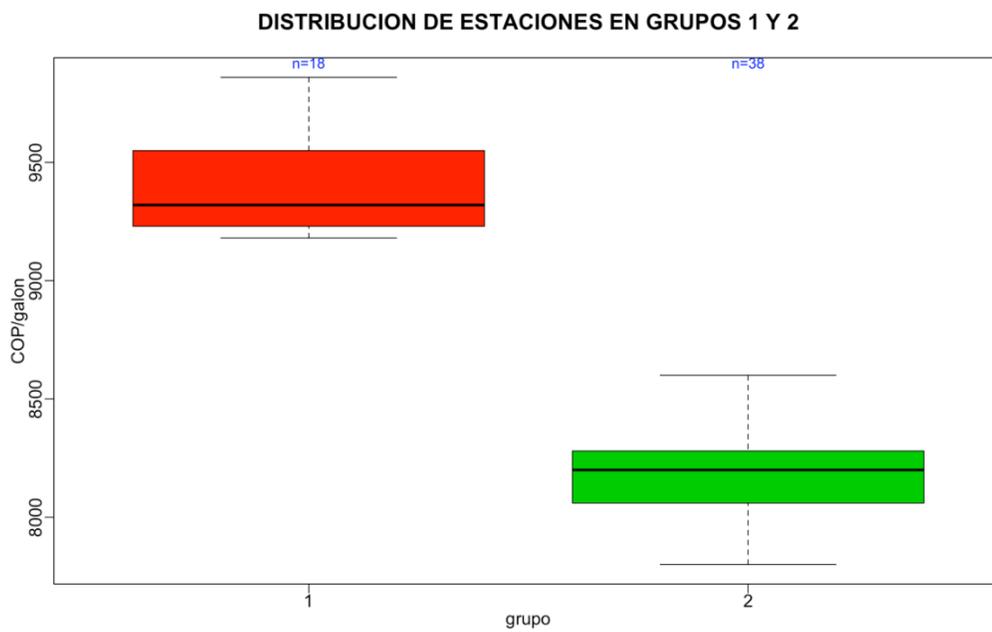


MDE2\_CPQDE2010  
hclust (\*, "average")

### 9.1.2 Selección mejor número de conglomerados método jerárquico y no jerárquico

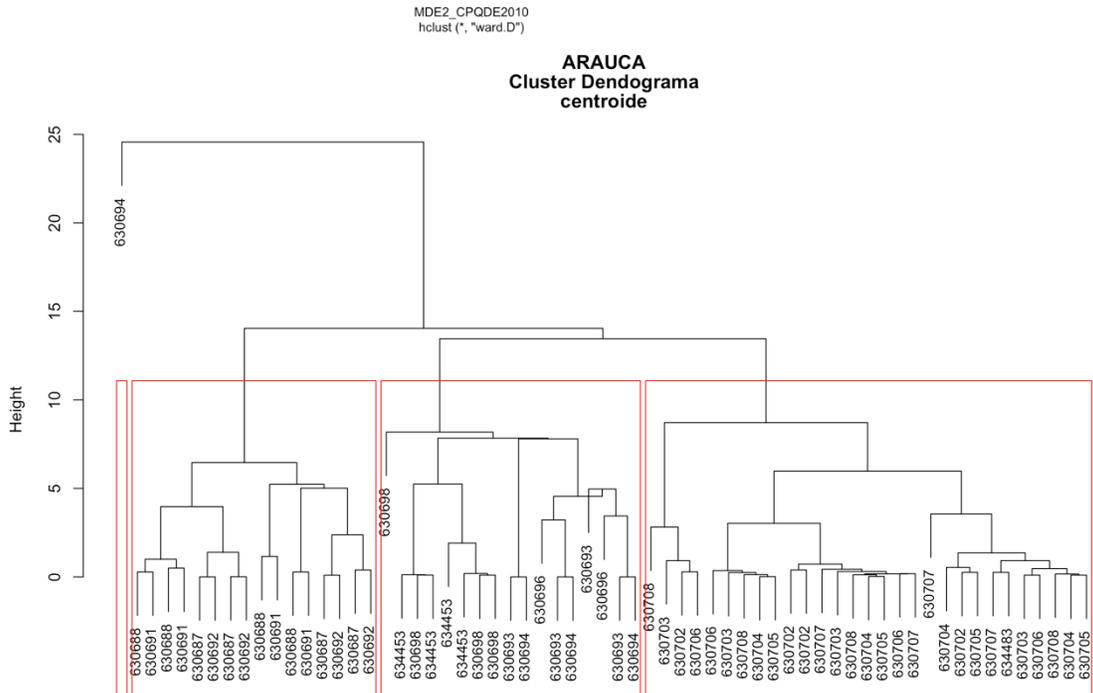
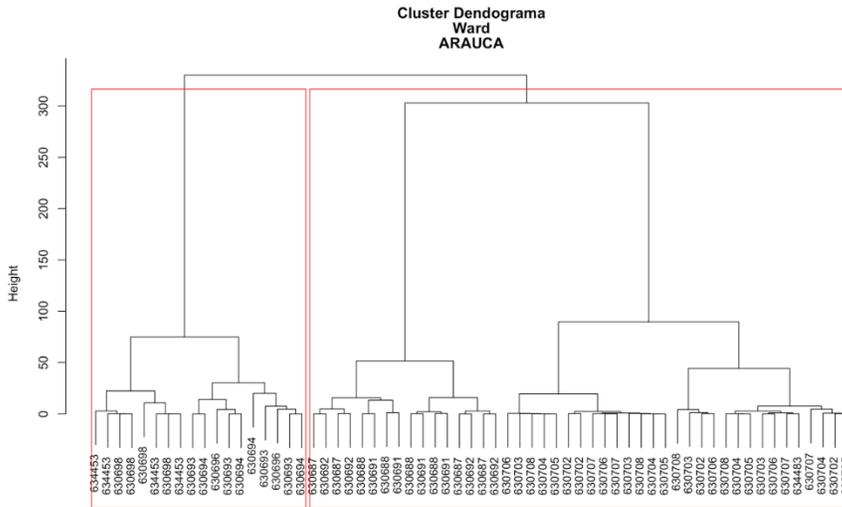


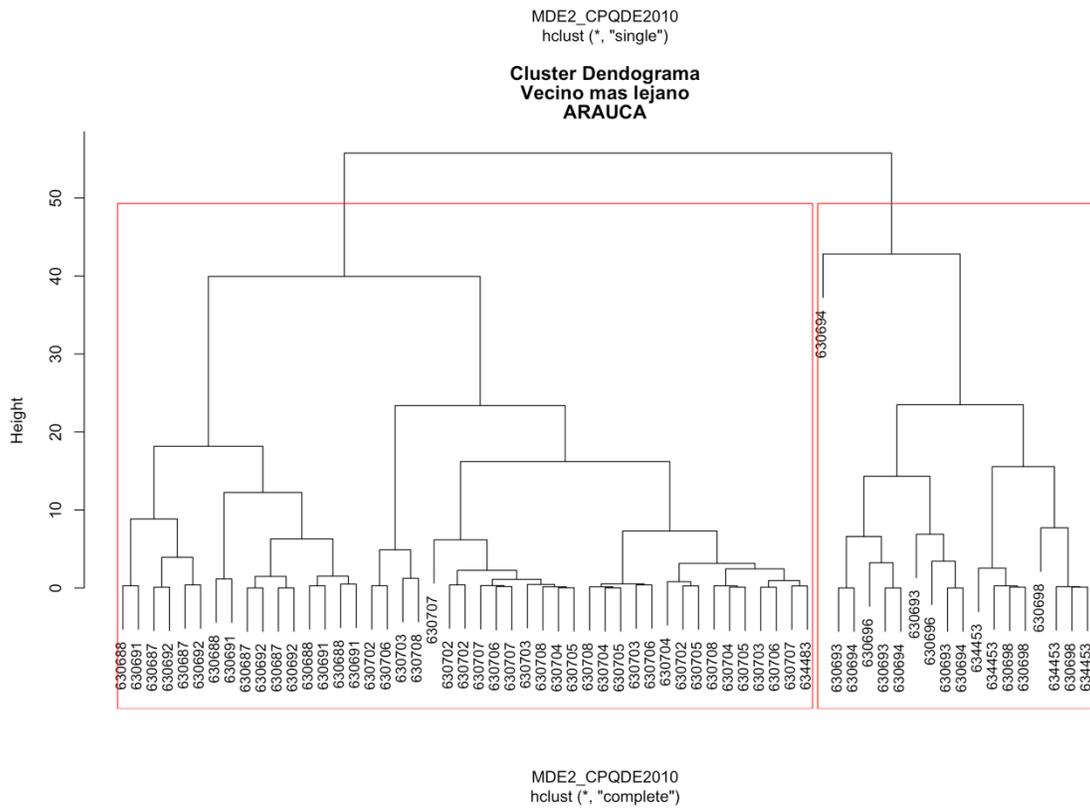
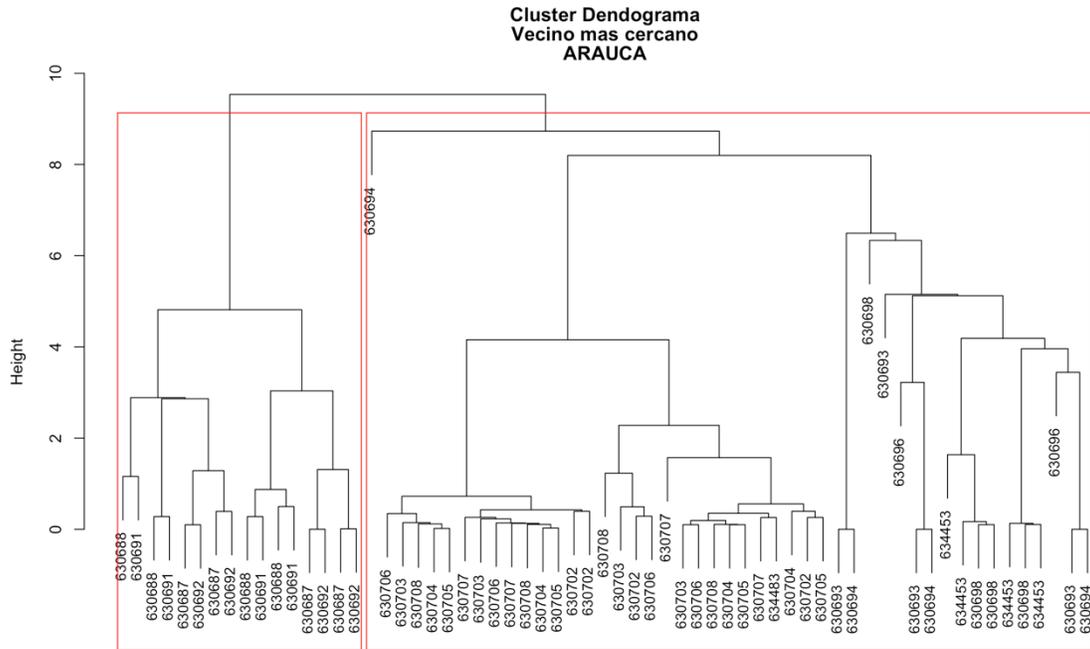
### 9.1.3 Grupos de estaciones y el comportamiento del precio

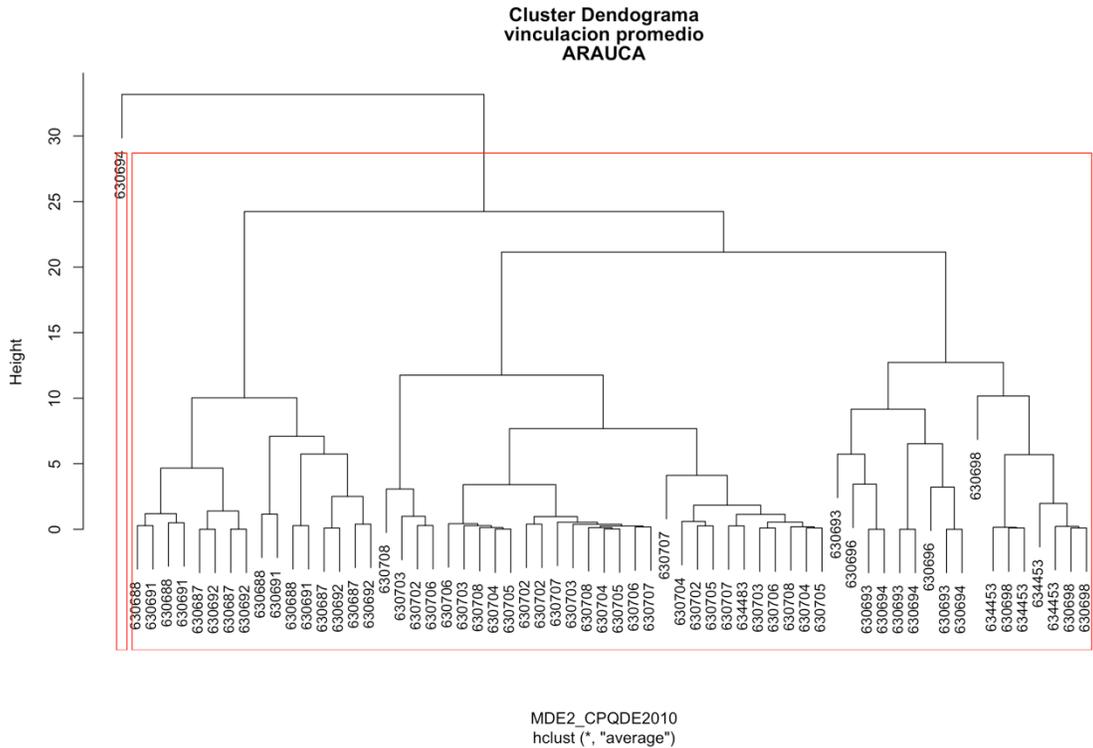


## 9.2 ARAUCA

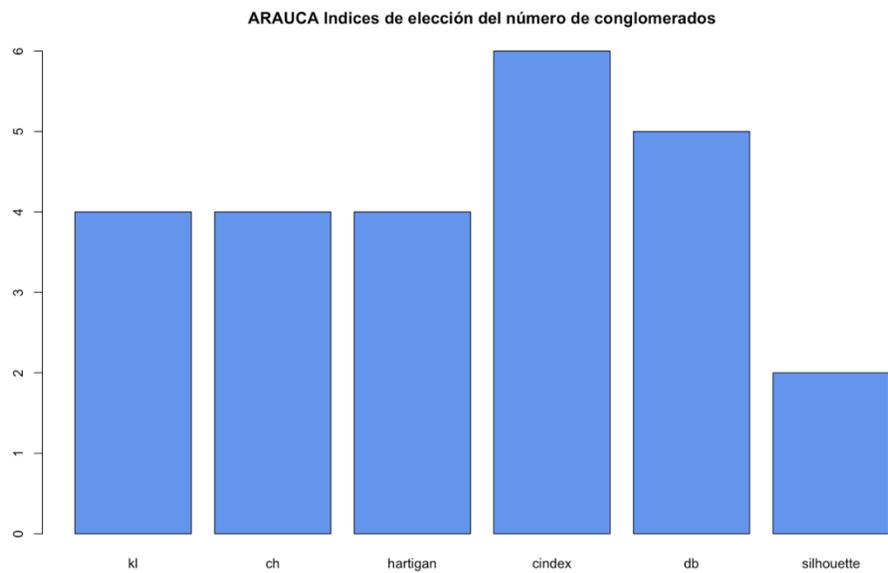
### 9.2.1 Resultados métodos conglomerados



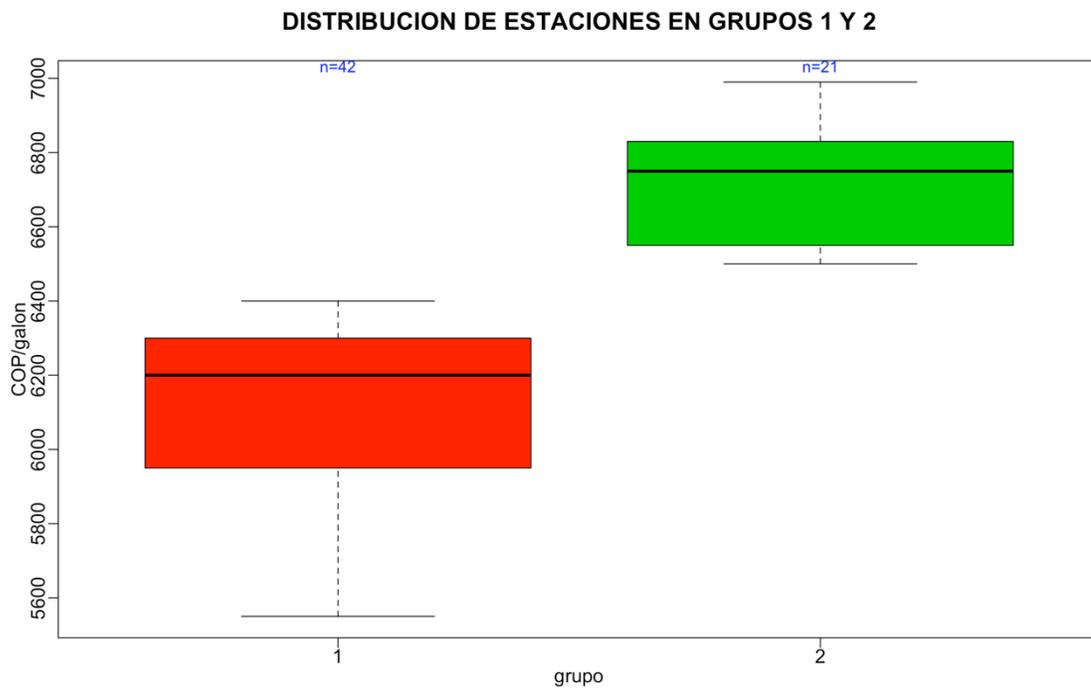




### 9.2.2 Selección mejor número de conglomerados método jerárquico y no jerárquico

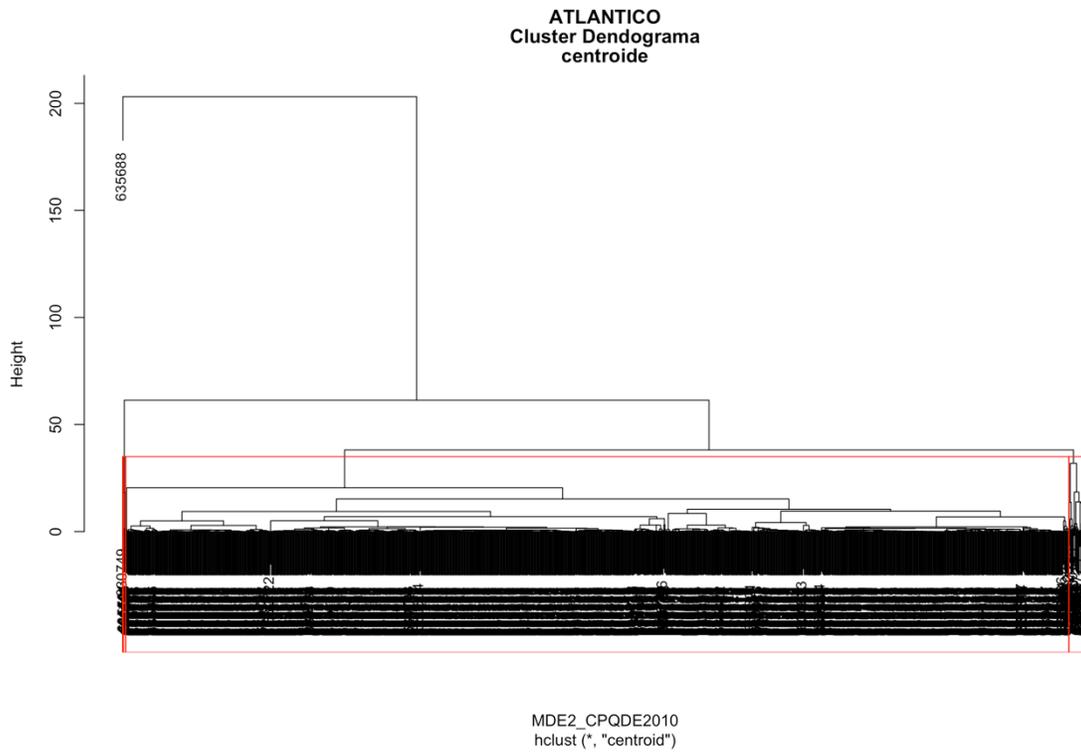


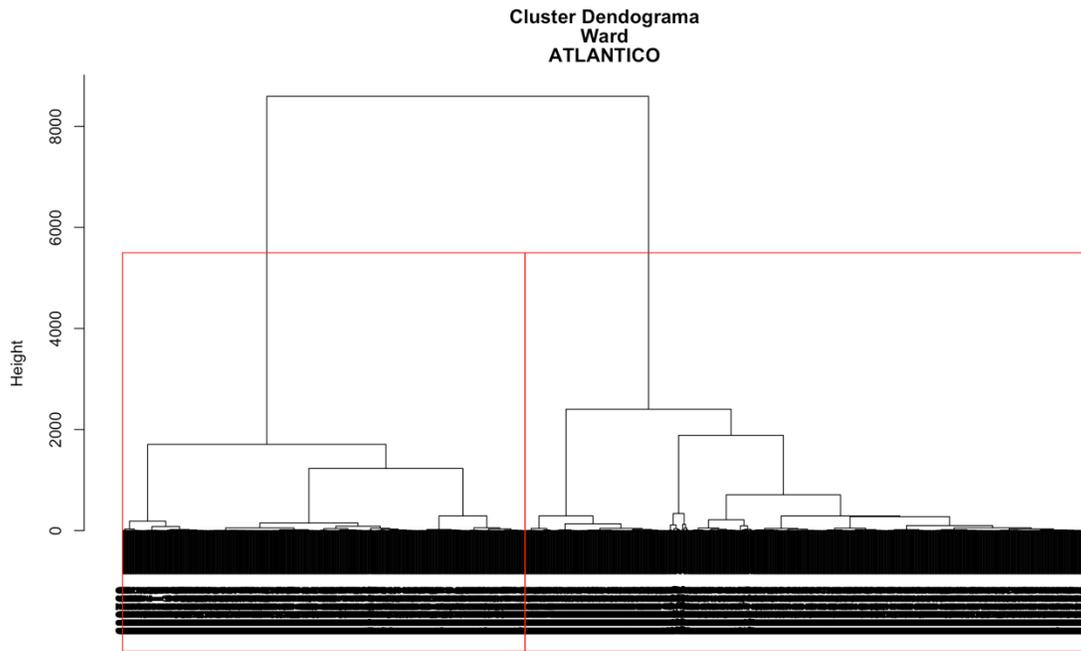
### 9.2.3 Grupos de estaciones y el comportamiento del precio



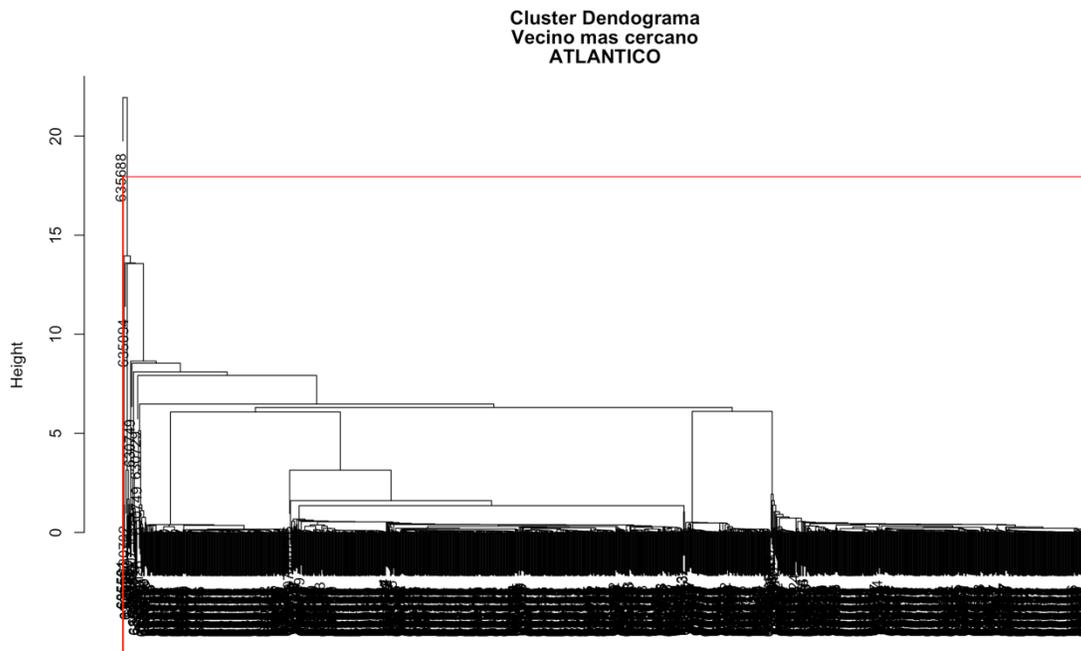
### 9.3 ATLANTICO

#### 9.3.1 Resultados métodos conglomerados



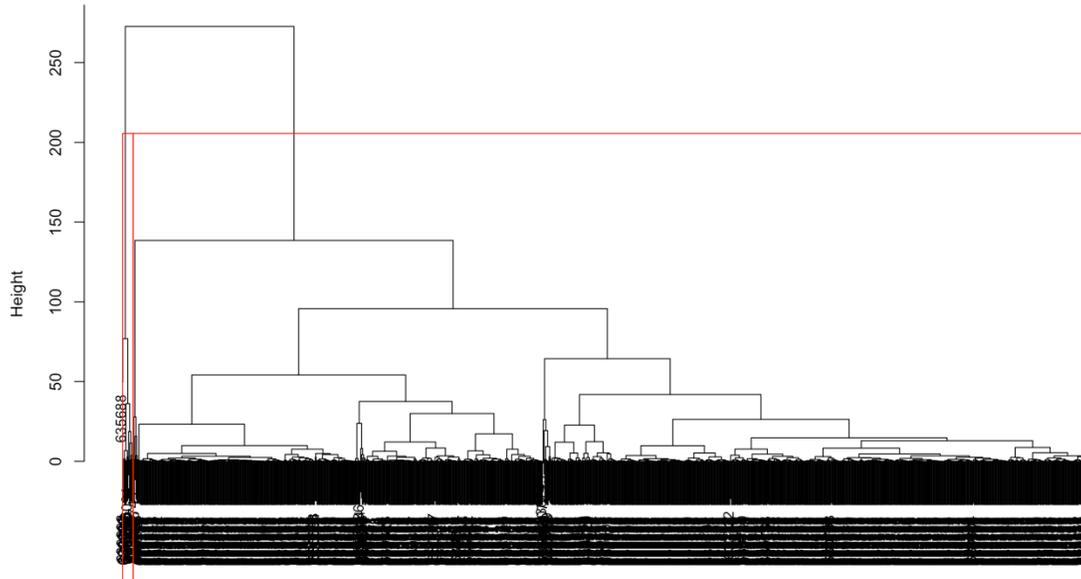


MDE2\_CPQDE2010  
hclust (\*, "ward.D")



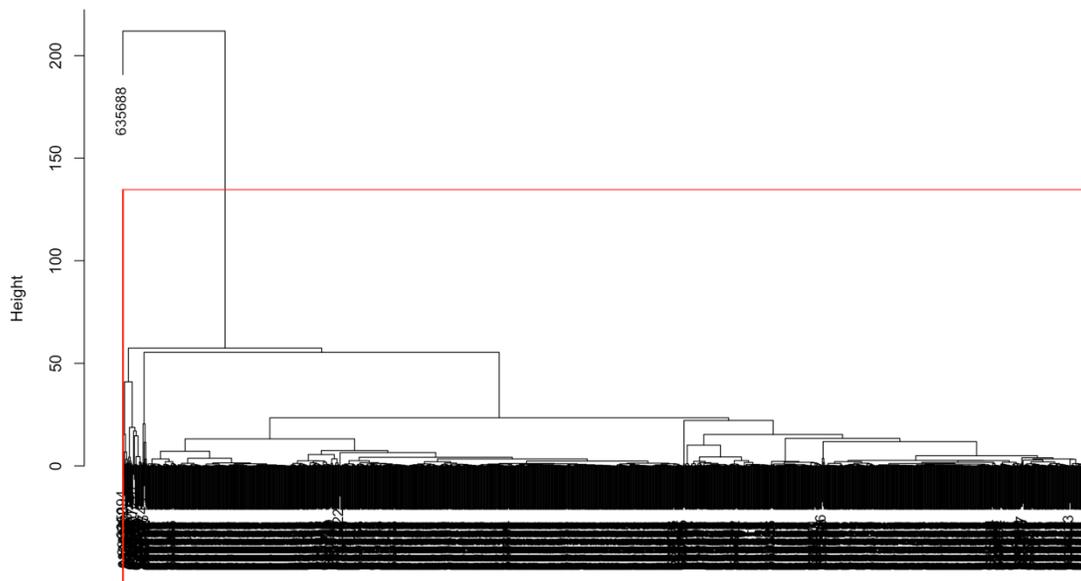
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
ATLANTICO**



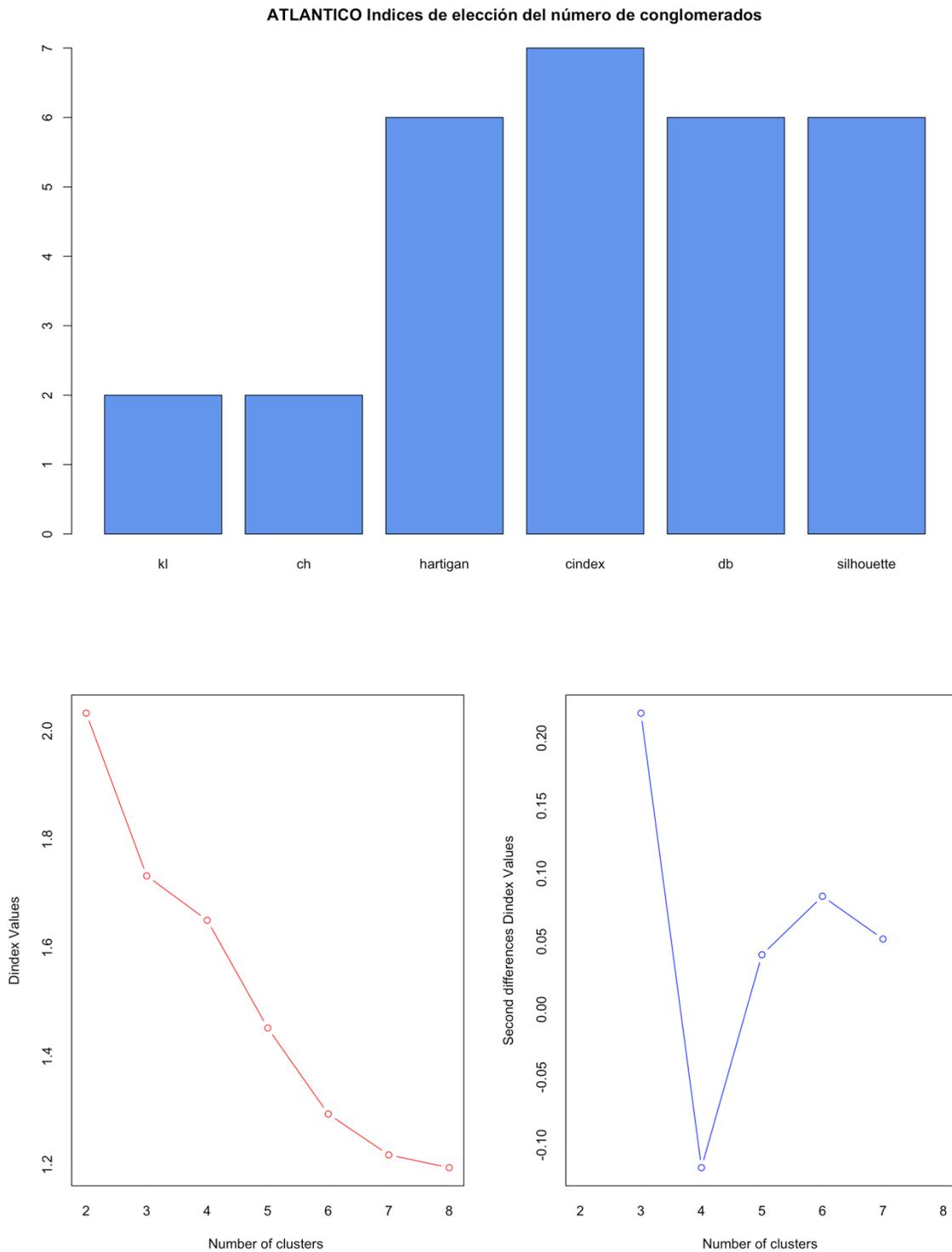
MDE2\_CPQDE2010  
hclust (\*, "complete")

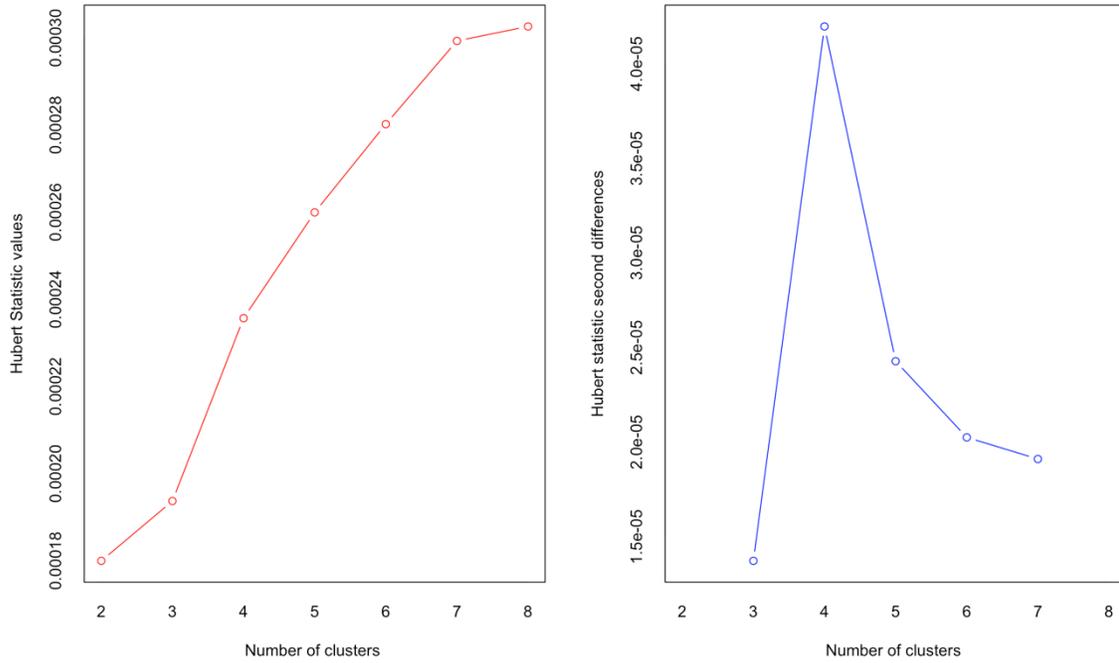
**Cluster Dendograma  
vinculacion promedio  
ATLANTICO**



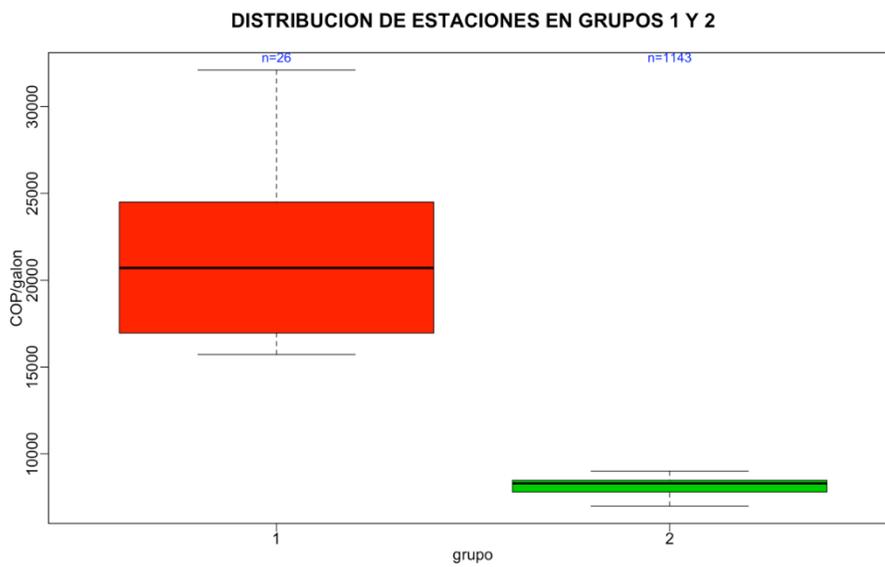
MDE2\_CPQDE2010  
hclust (\*, "average")

### 9.3.2 Selección mejor número de conglomerados método jerárquico



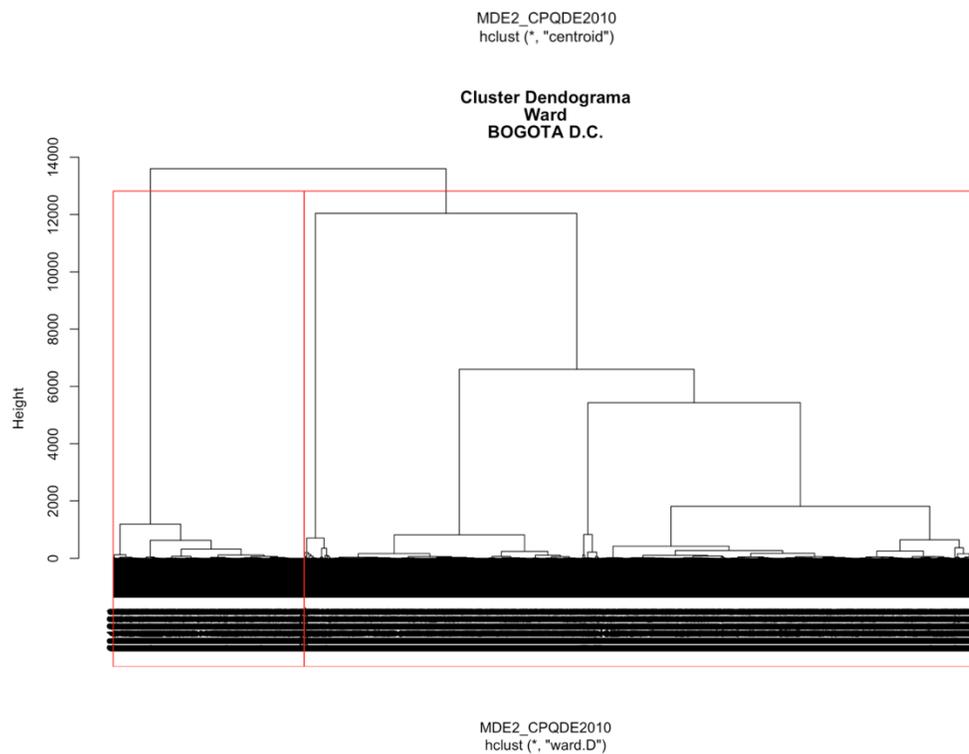
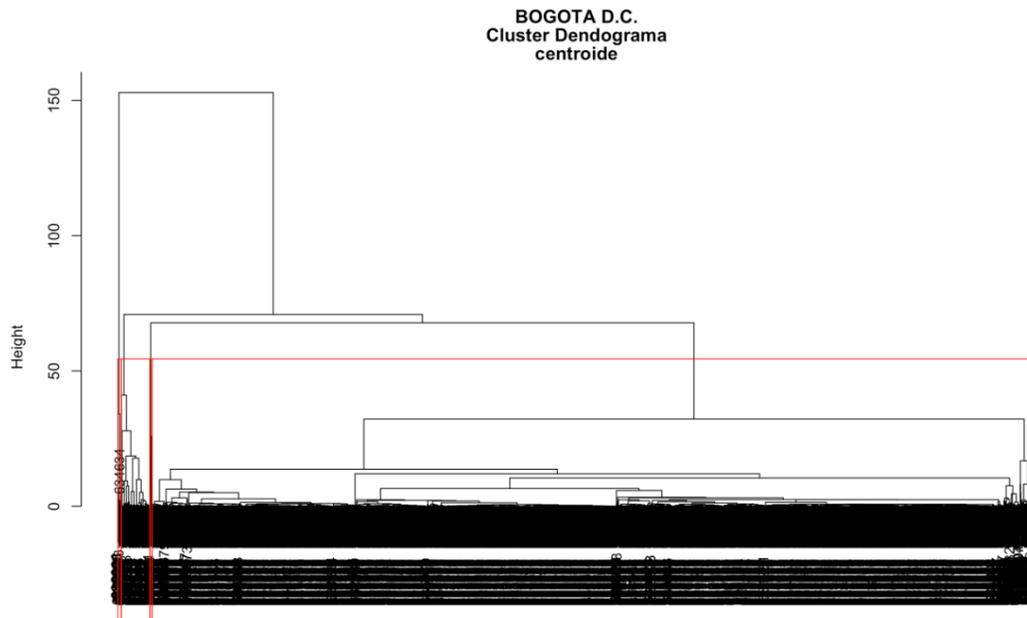


### 9.3.3 Grupos de estaciones y el comportamiento del precio

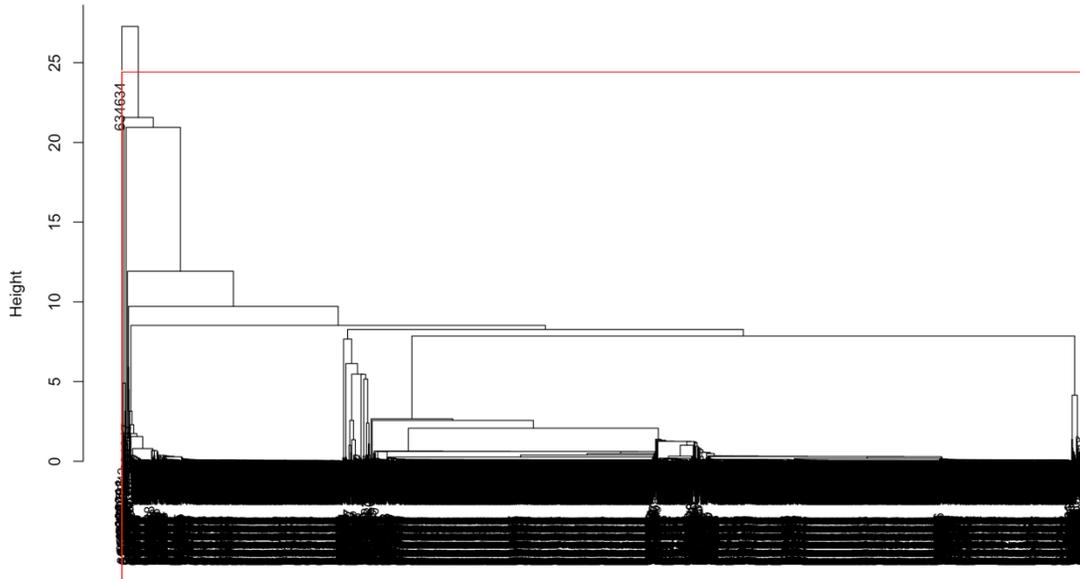


## 9.4 BOGOTA

### 9.4.1 Resultados métodos conglomerados

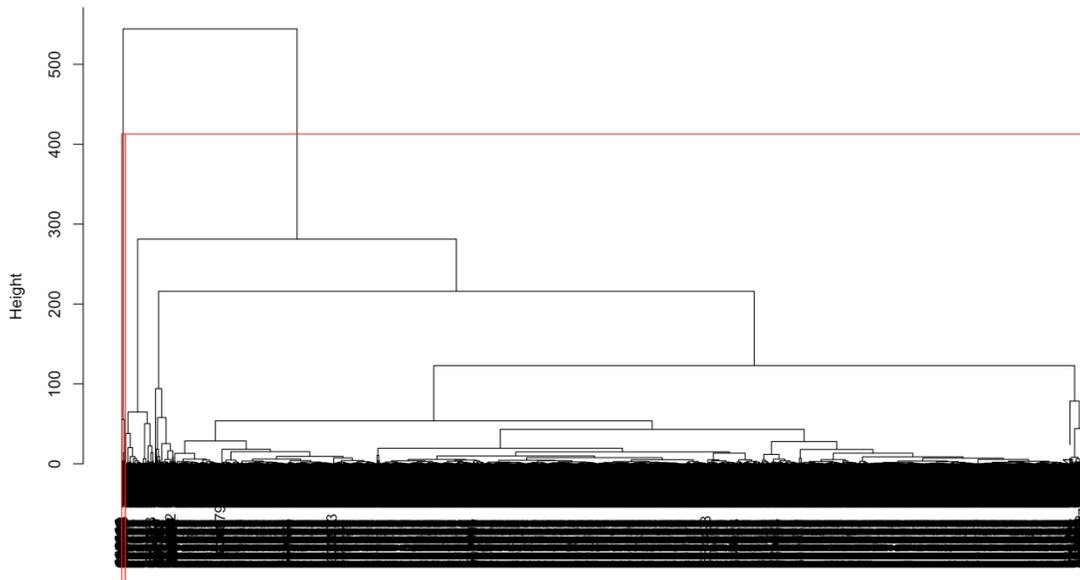


**Cluster Dendograma  
Vecino mas cercano  
BOGOTA D.C.**

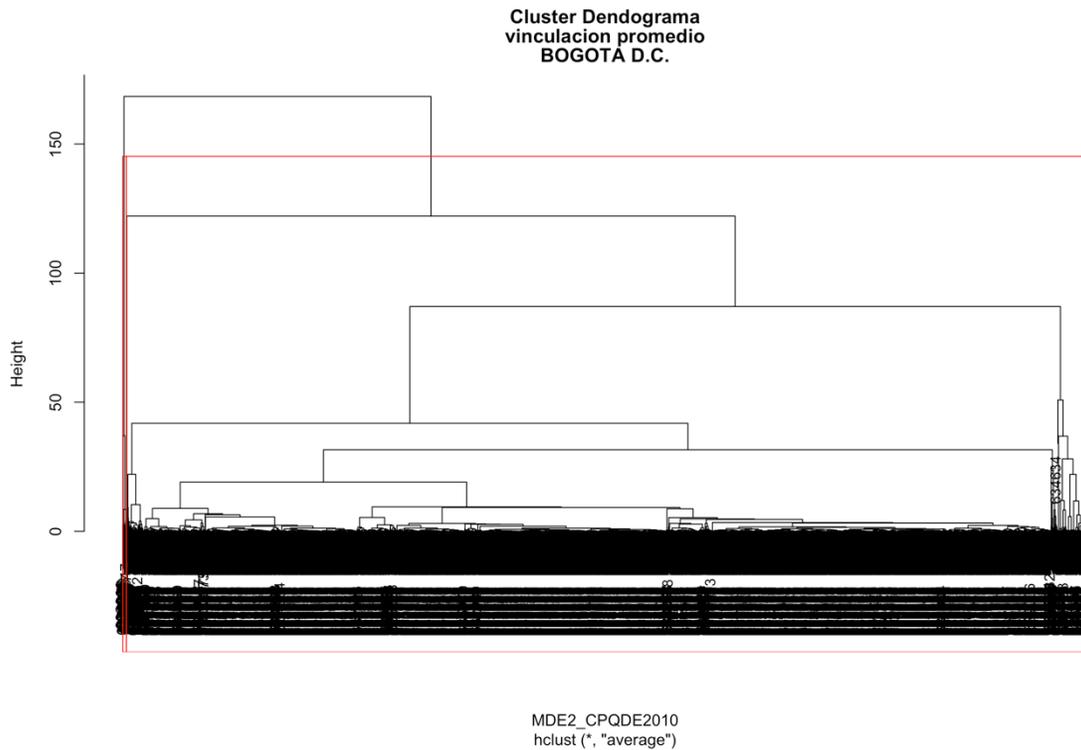


MDE2\_CPQDE2010  
hclust (\*, "single")

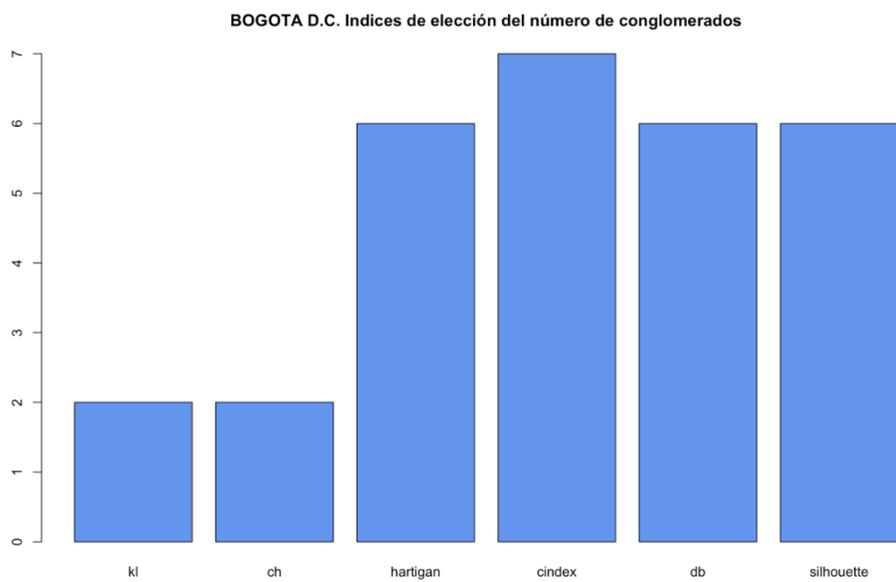
**Cluster Dendograma  
Vecino mas lejano  
BOGOTA D.C.**



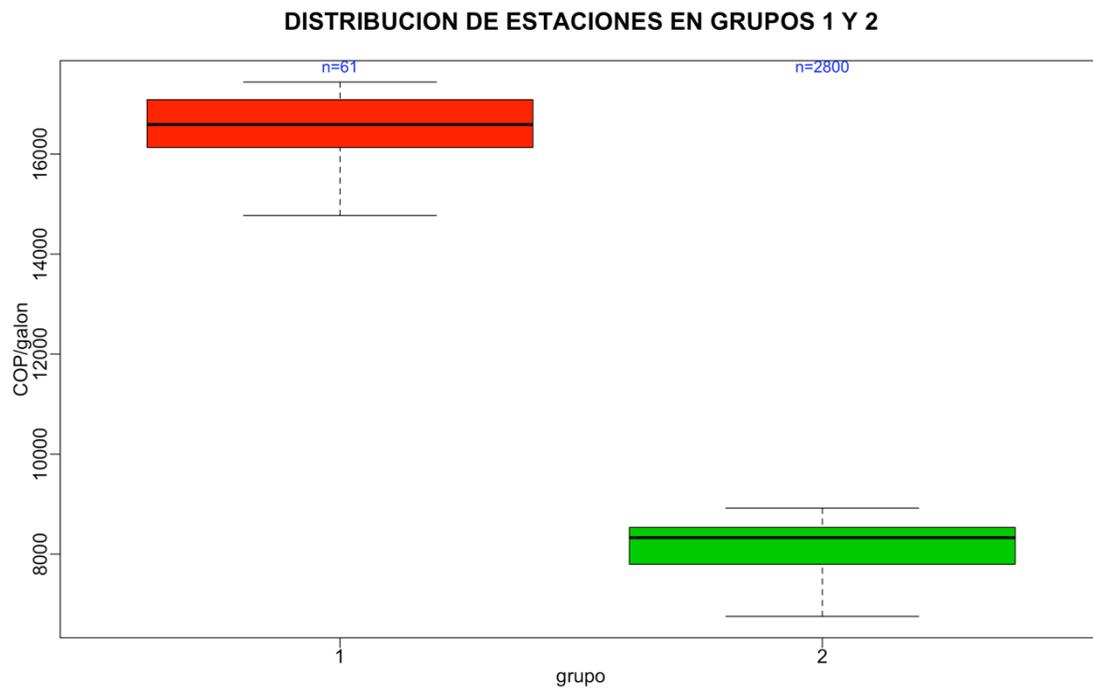
MDE2\_CPQDE2010  
hclust (\*, "complete")



### 9.4.2 Selección mejor número de conglomerados método jerárquico

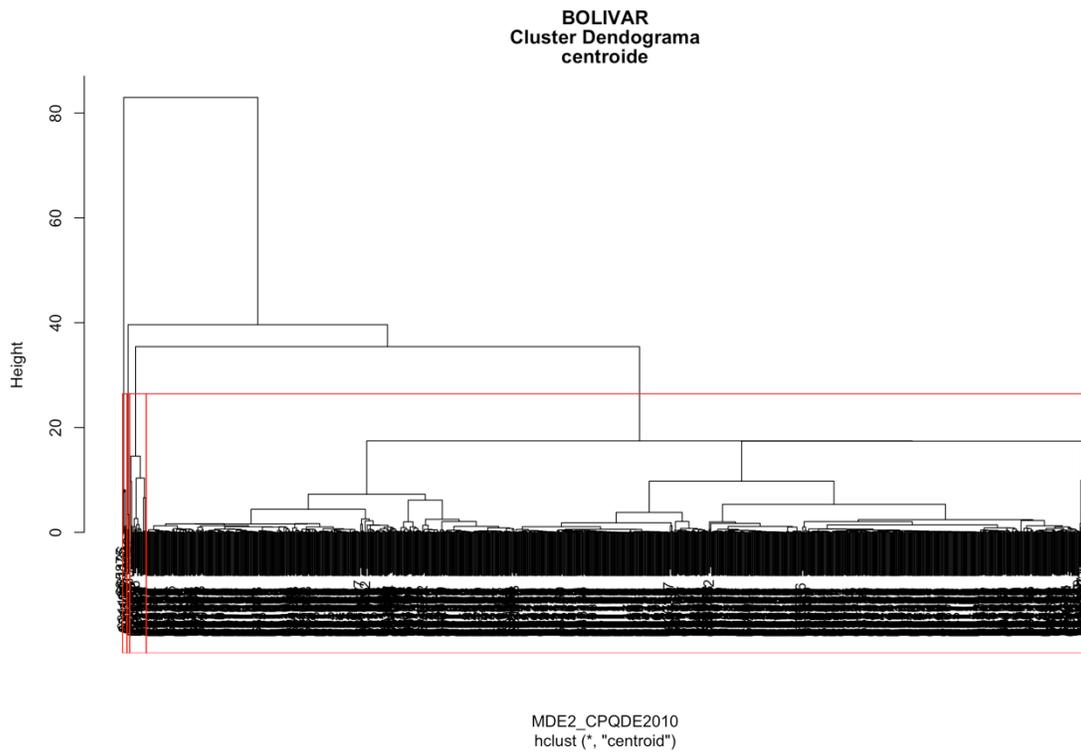


### 9.4.3 Grupos de estaciones y el comportamiento del precio



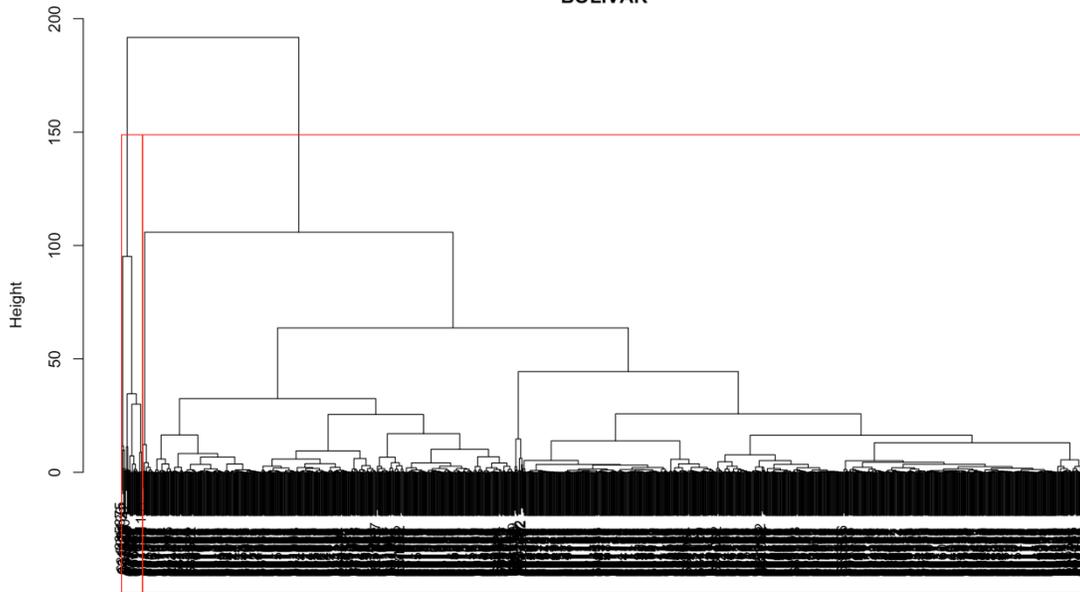
## 9.5 BOLIVAR

### 9.5.1 Resultados métodos conglomerados



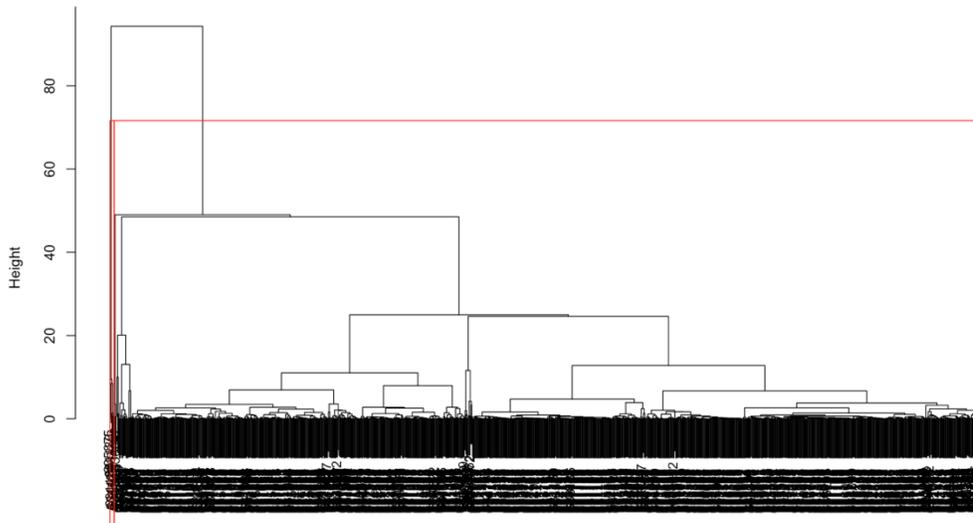


**Cluster Dendograma  
Vecino mas lejano  
BOLIVAR**



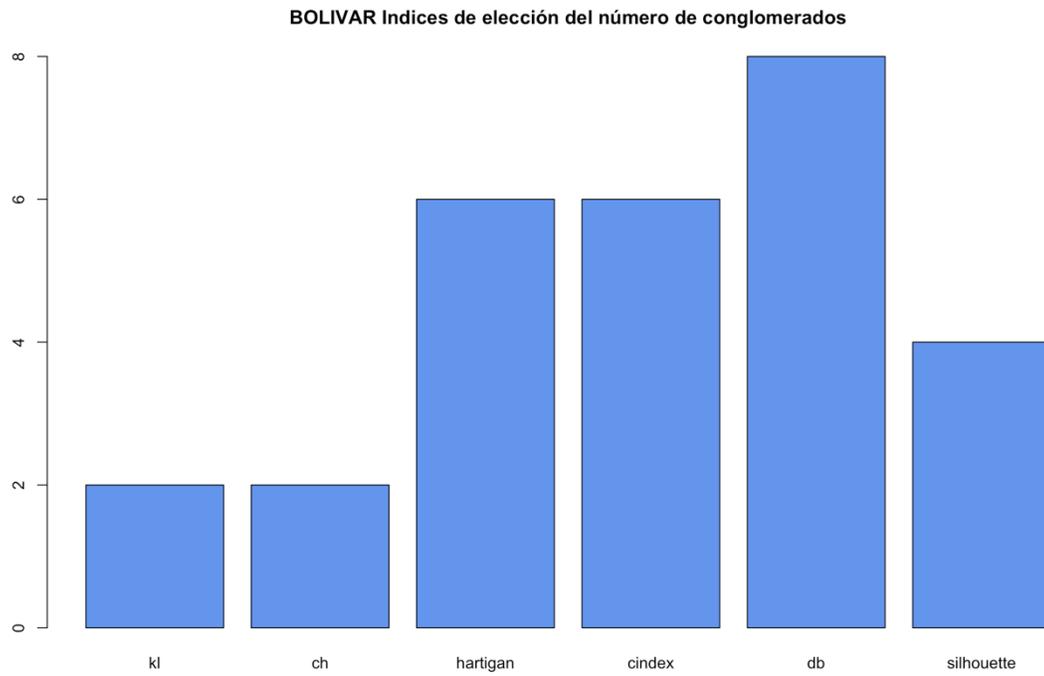
MDE2\_CPQDE2010  
hclust (\*, "complete")

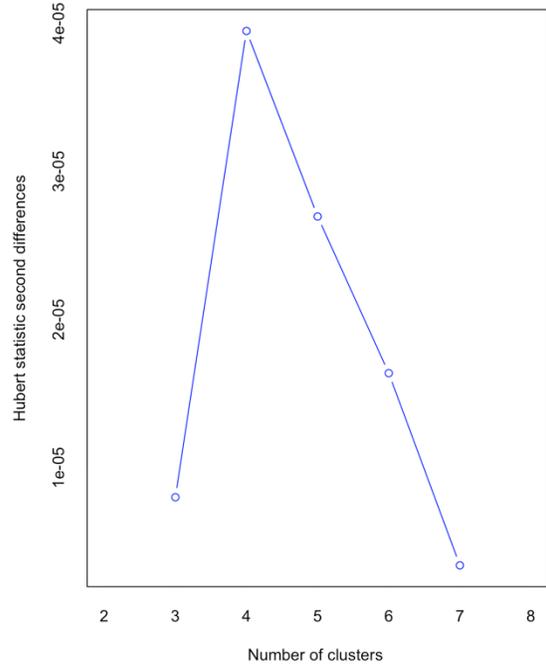
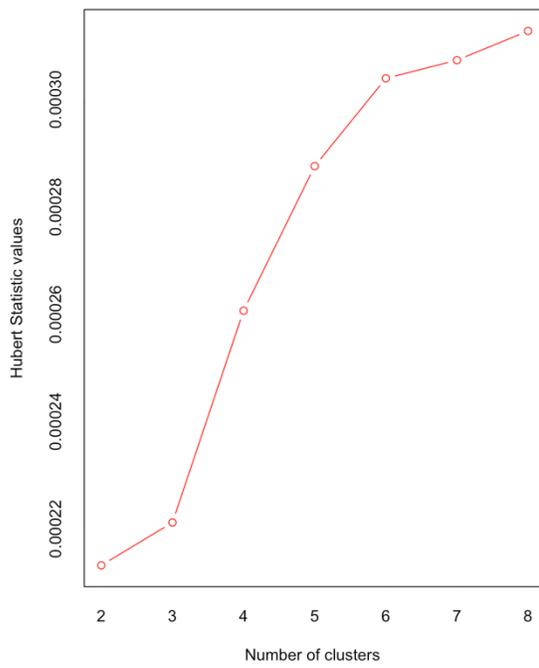
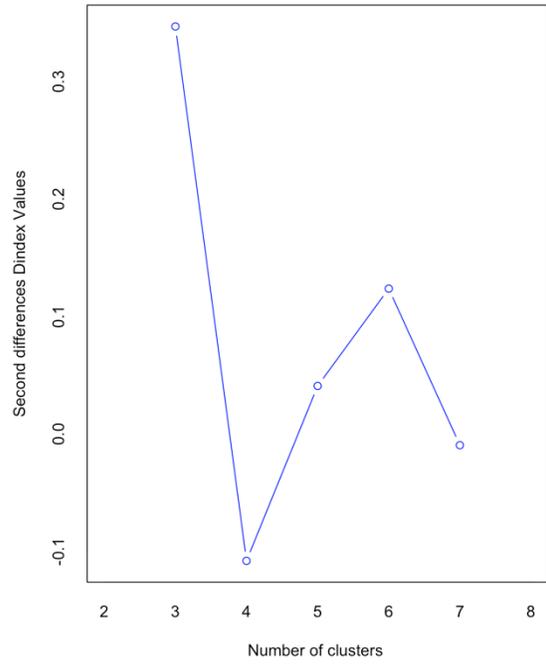
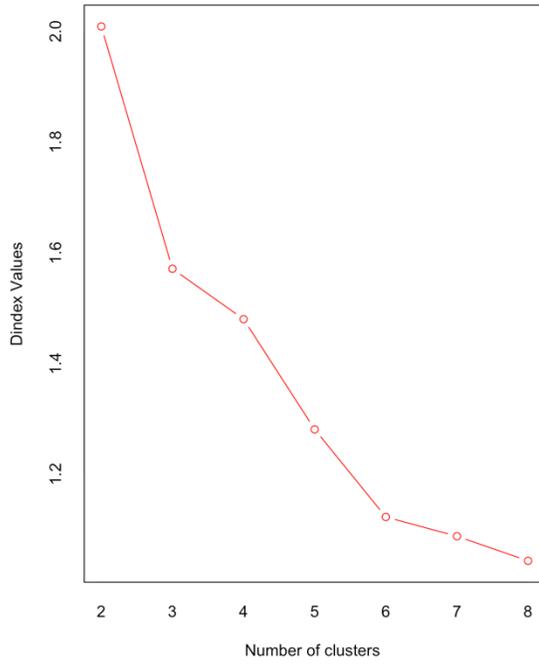
**Cluster Dendograma  
vinculacion promedio  
BOLIVAR**



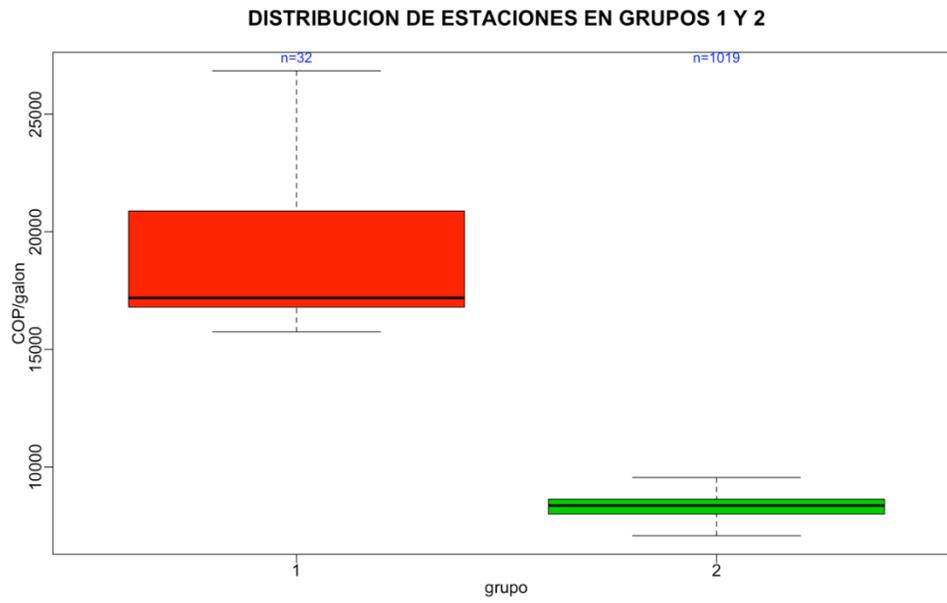
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.5.2 Selección mejor número de conglomerados método jerárquico



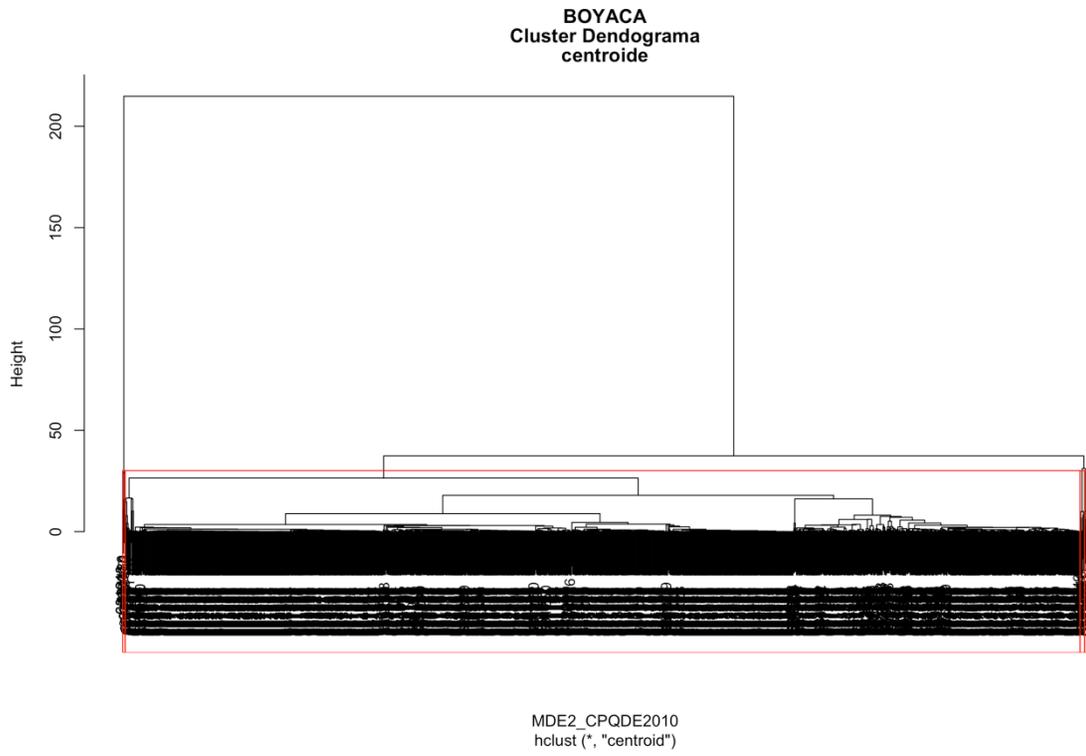


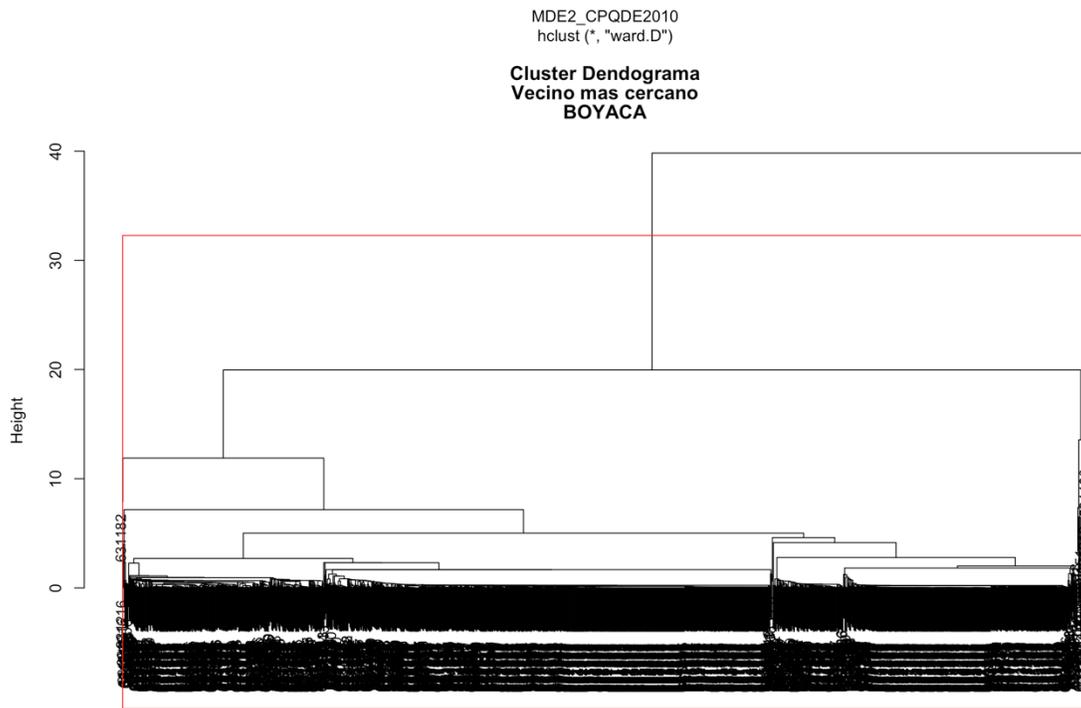
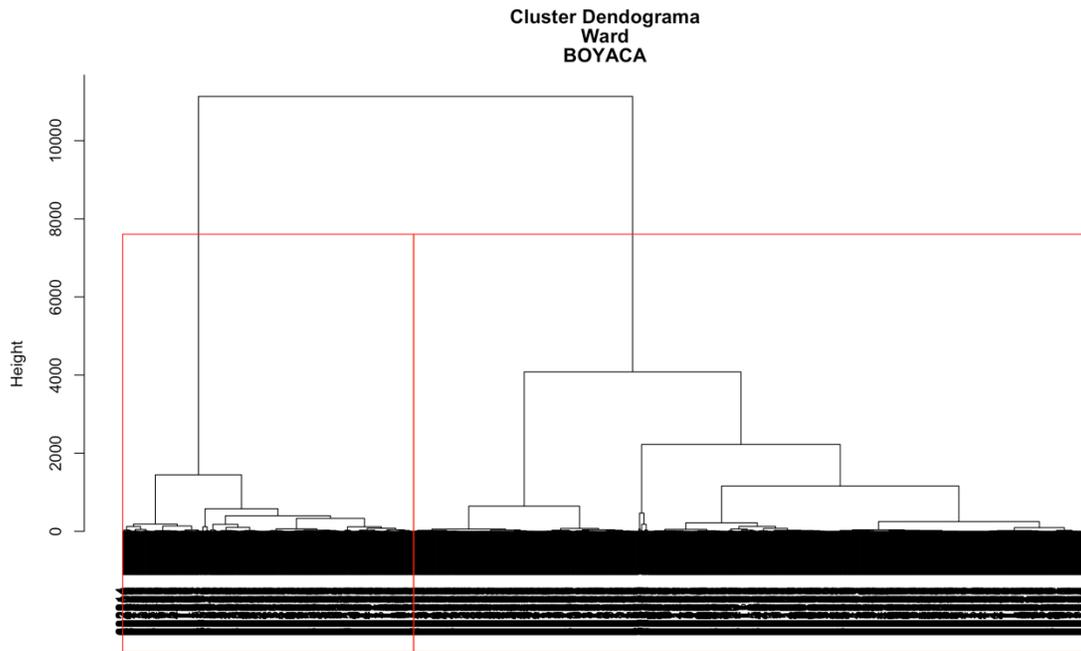
### 9.5.3 Grupos de estaciones y el comportamiento del precio



## 9.6 BOYACA

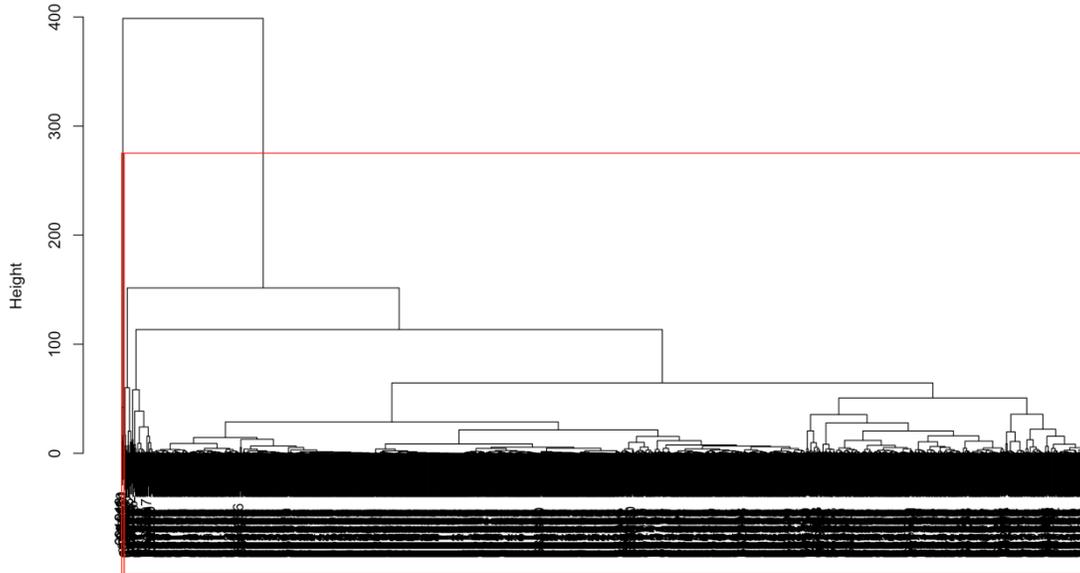
### 9.6.1 Resultados métodos conglomerados





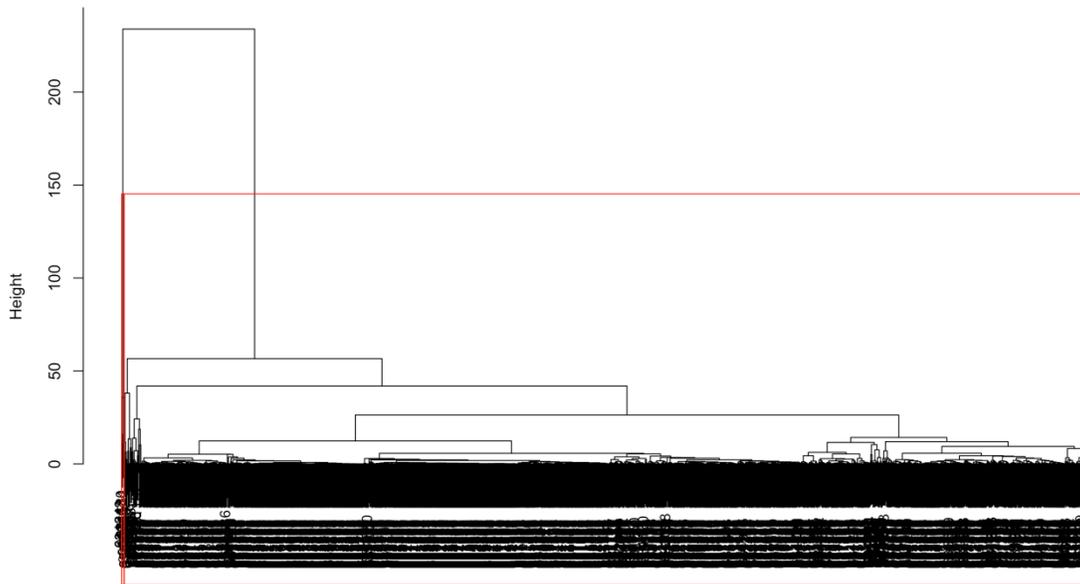
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
BOYACA**



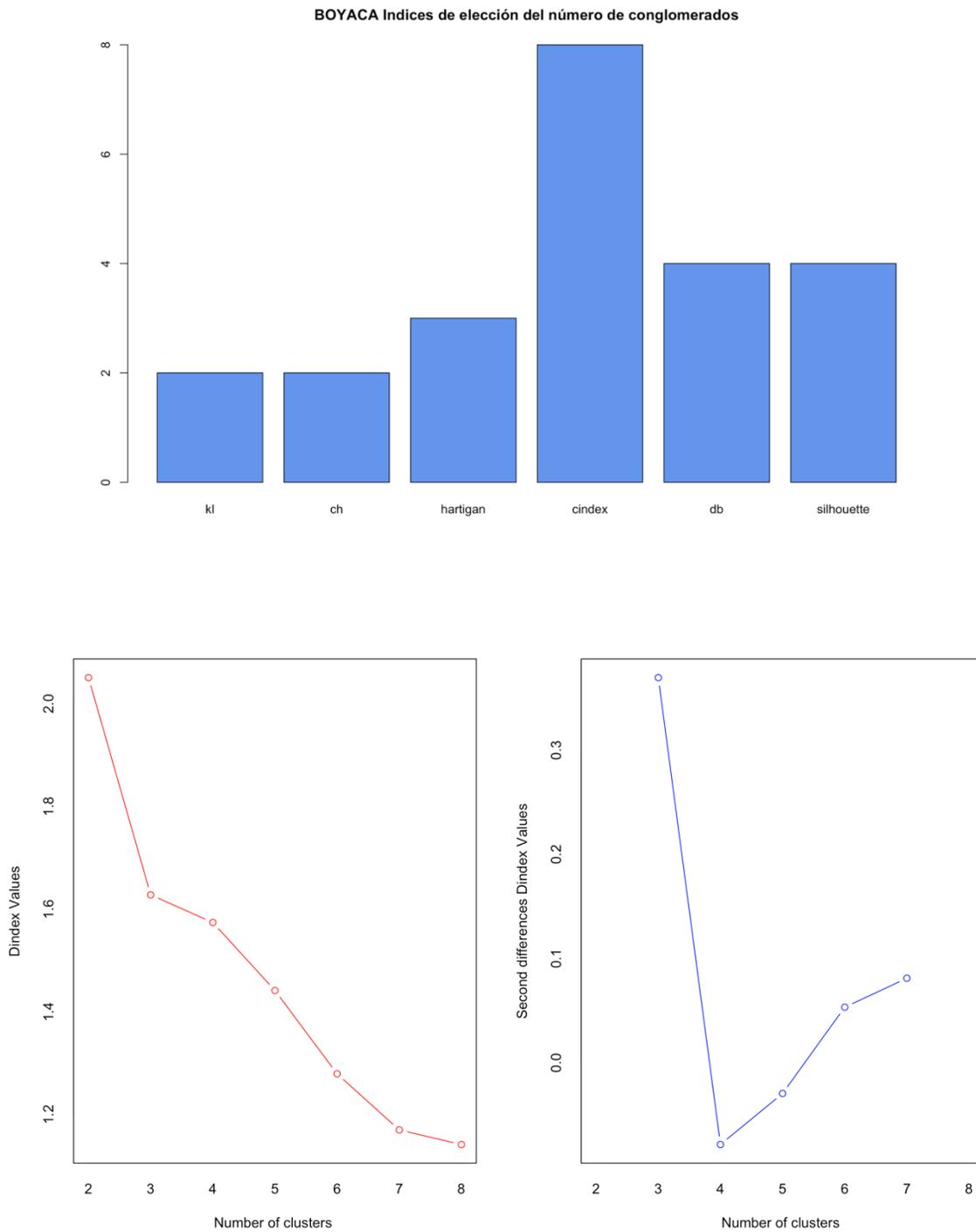
MDE2\_CPQDE2010  
hclust (\*, "complete")

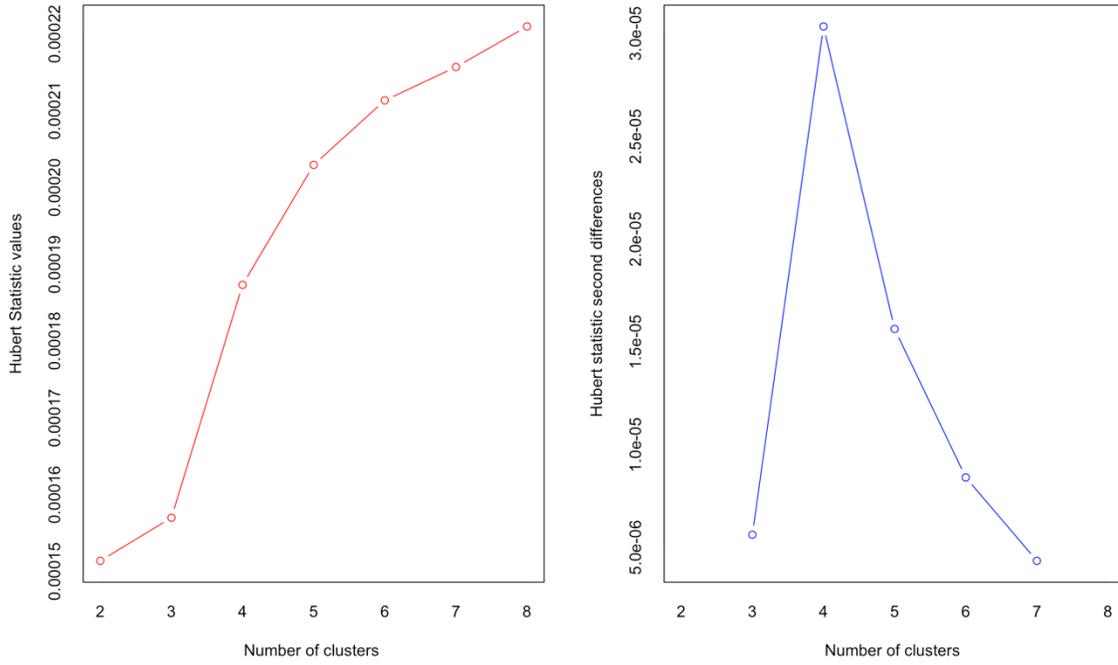
**Cluster Dendograma  
vinculacion promedio  
BOYACA**



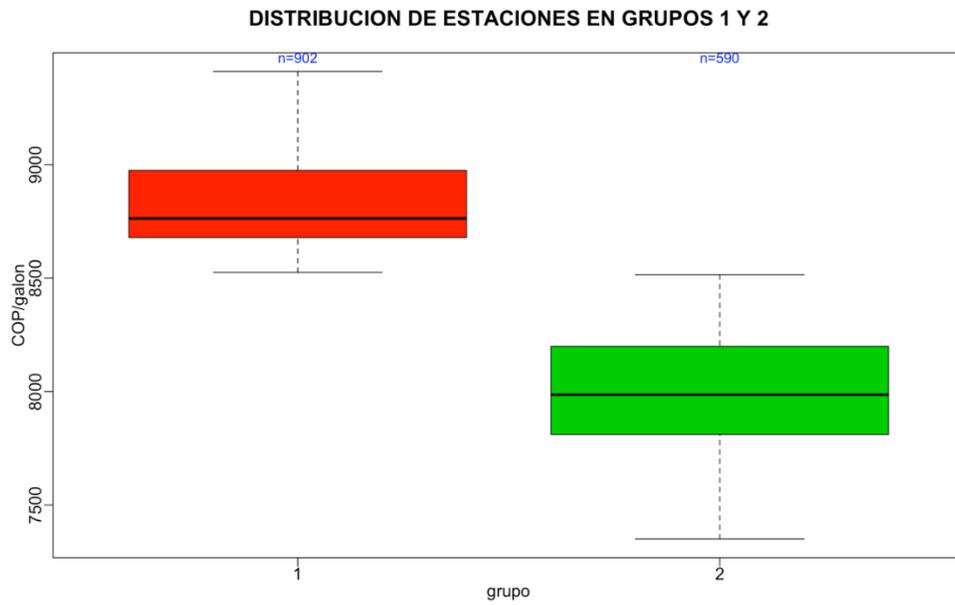
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.6.2 Selección mejor número de conglomerados método jerárquico



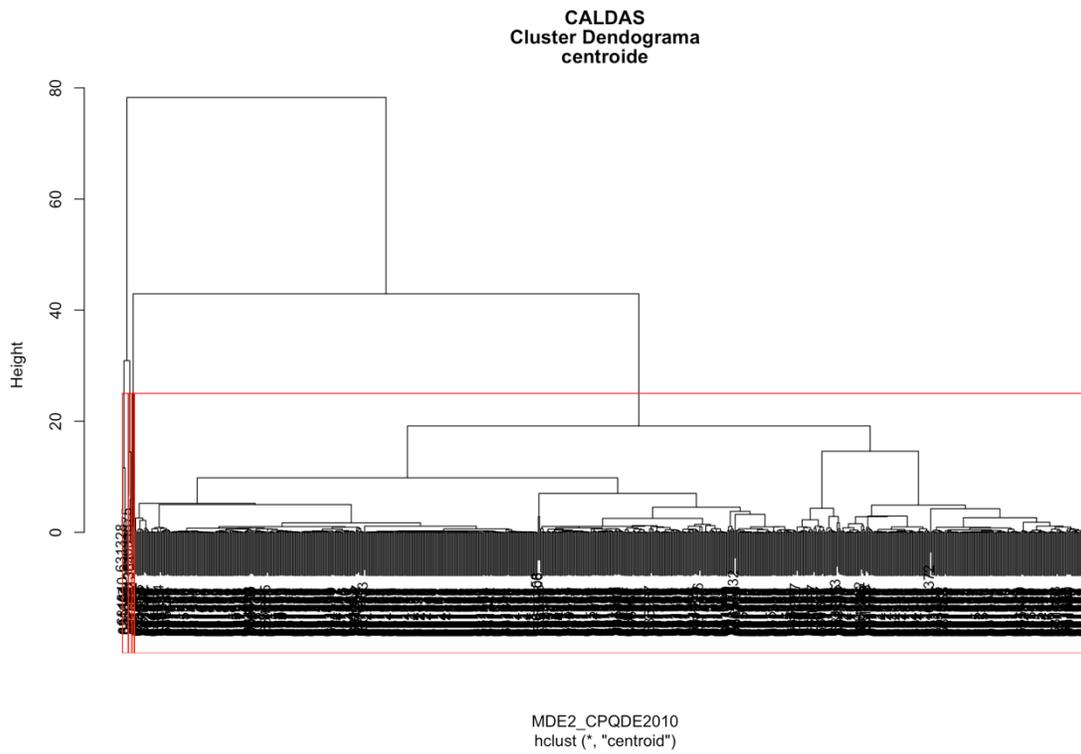


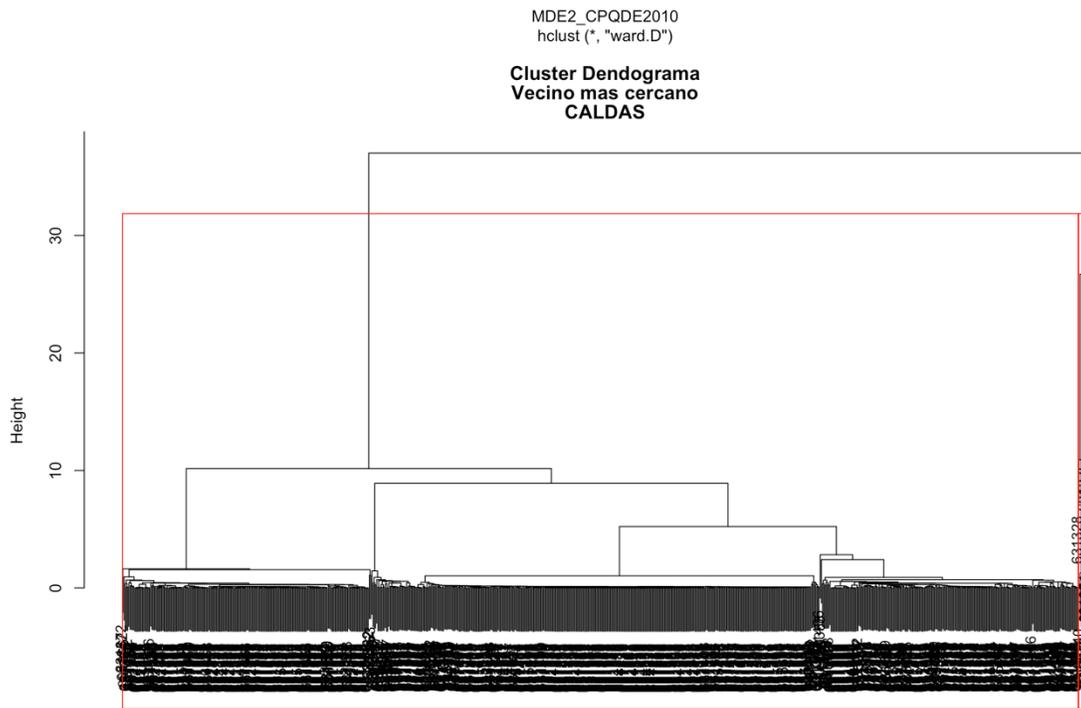
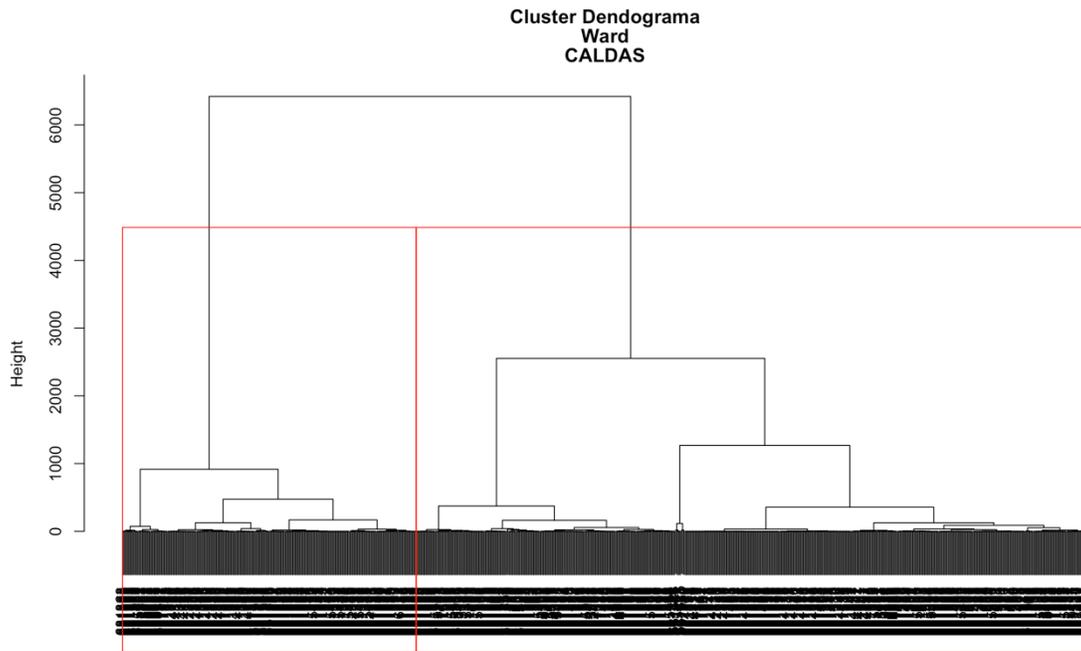
### 9.6.3 Grupos de estaciones y el comportamiento del precio



## 9.7 CALDAS

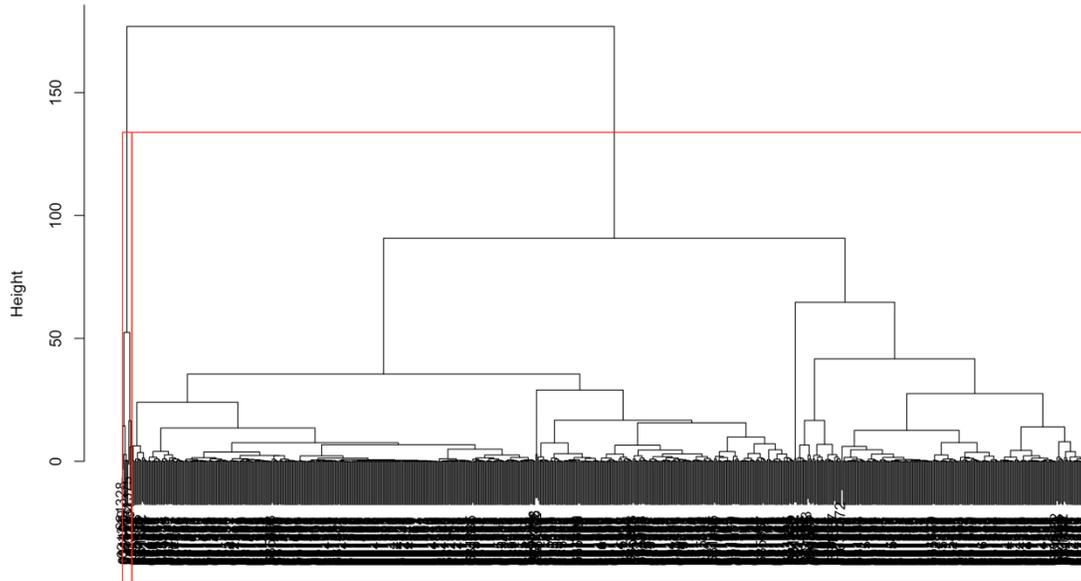
### 9.7.1 Resultados métodos conglomerados





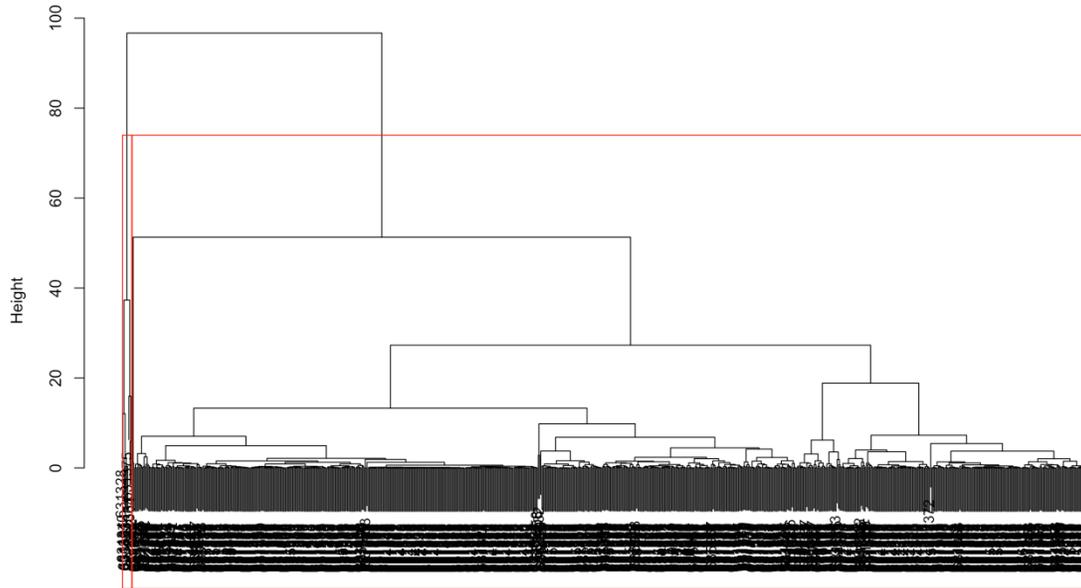
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
CALDAS**



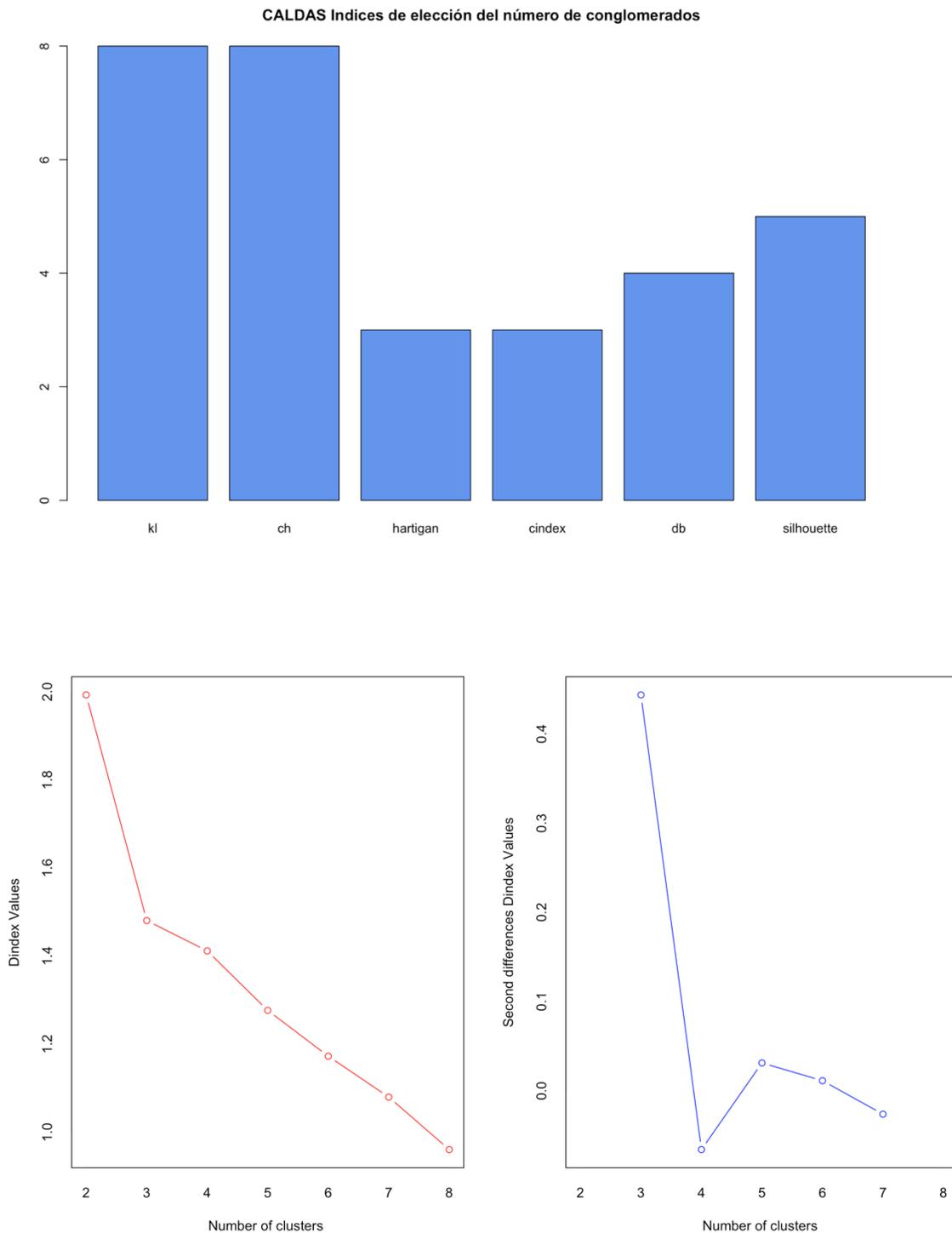
MDE2\_CPQDE2010  
hclust (\*, "complete")

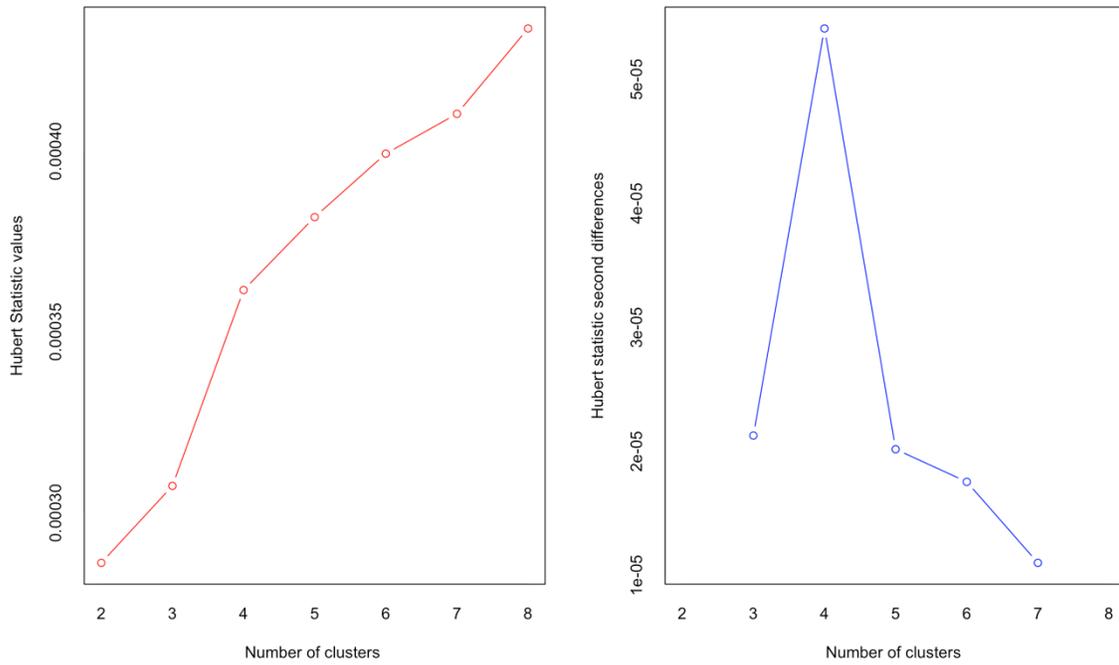
**Cluster Dendograma  
vinculacion promedio  
CALDAS**



MDE2\_CPQDE2010  
hclust (\*, "average")

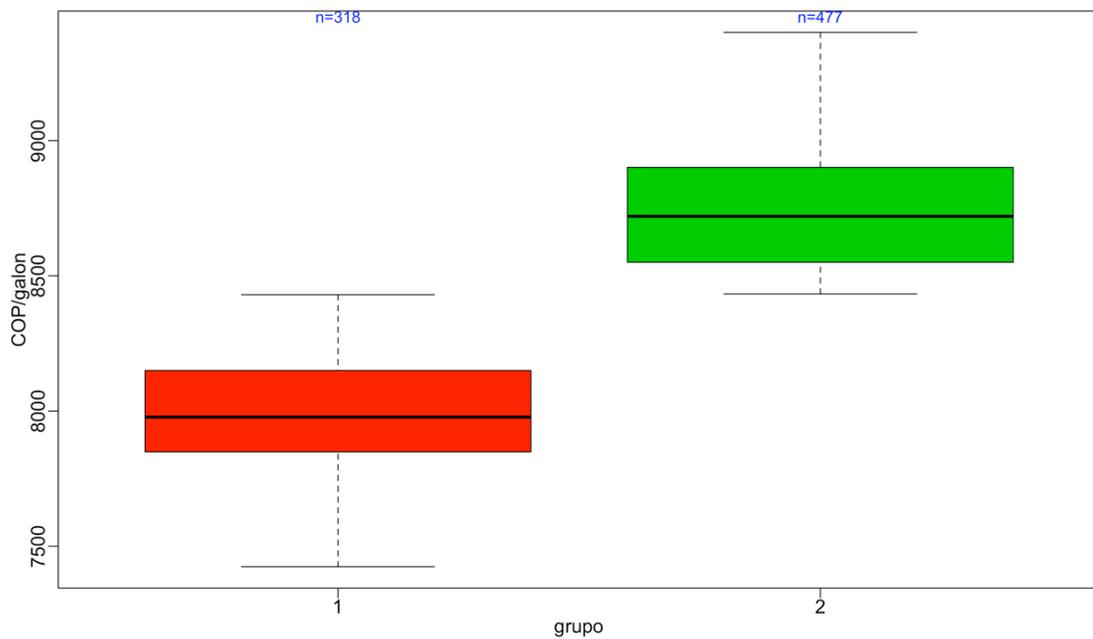
## 9.7.2 Selección mejor número de conglomerados método jerárquico





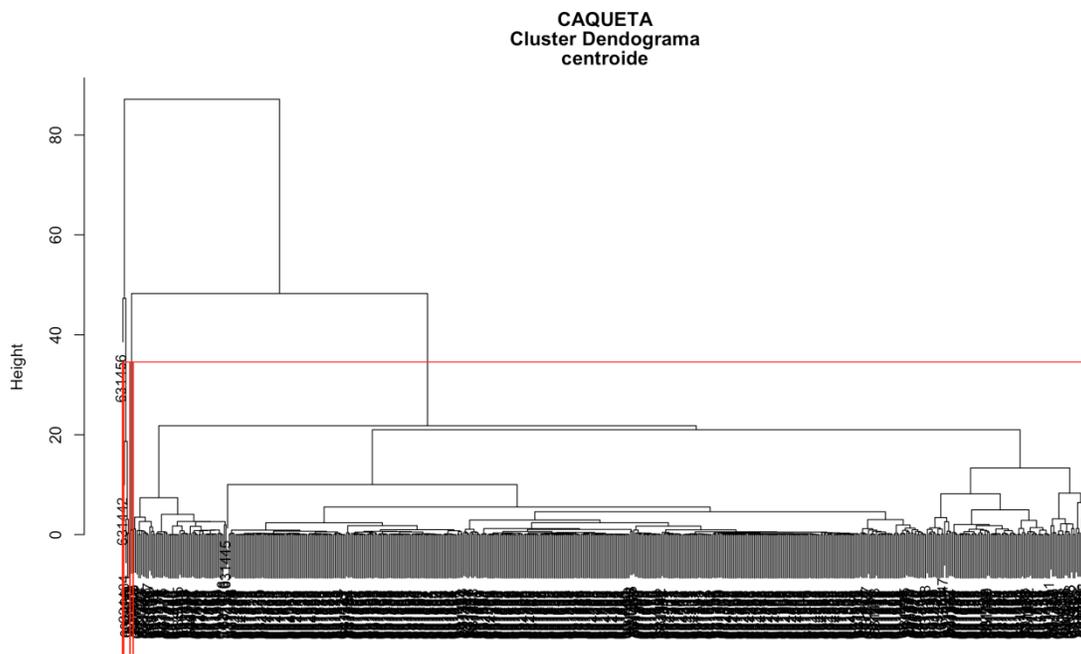
### 9.7.3 Grupos de estaciones y el comportamiento del precio

DISTRIBUCION DE ESTACIONES EN GRUPOS 1 Y 2

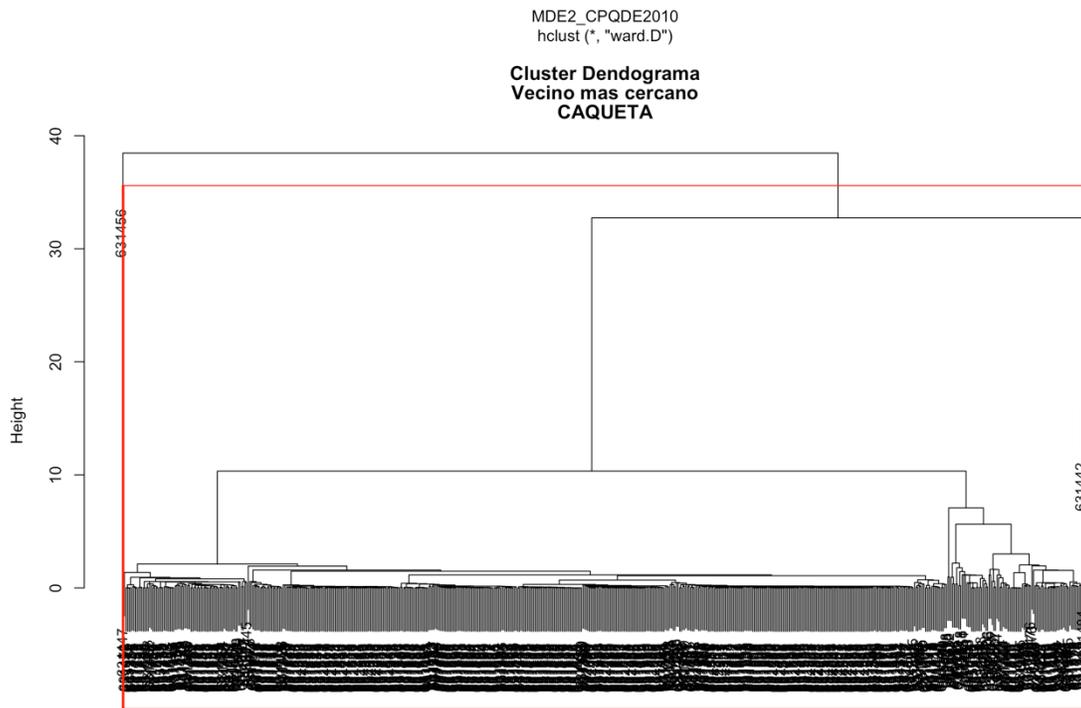
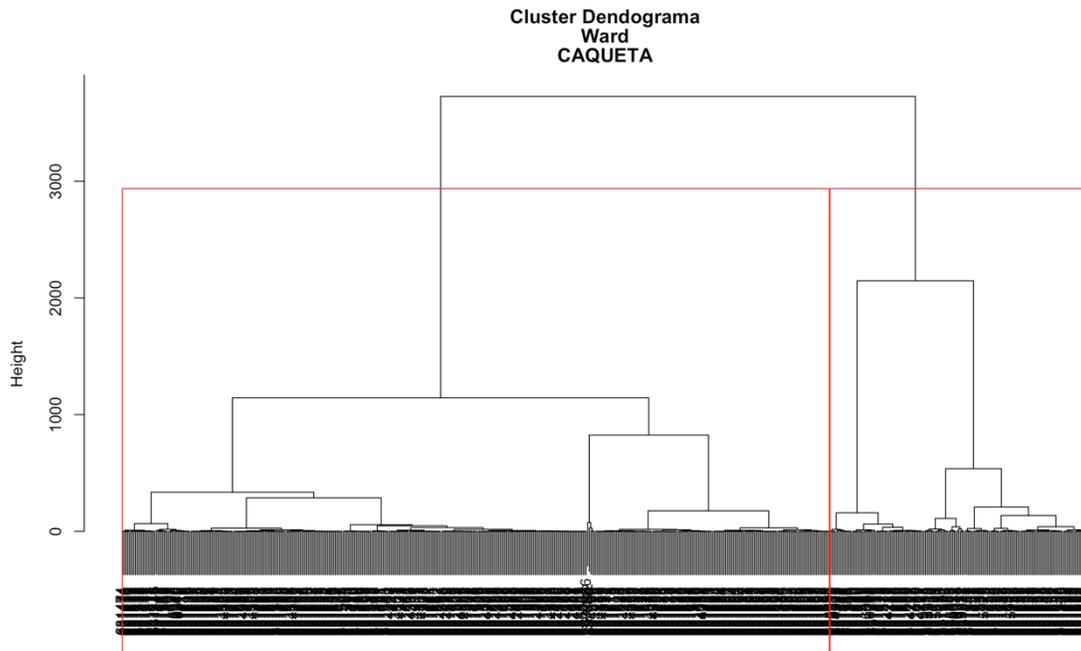


## 9.8 CAQUETA

### 9.8.1 Resultados métodos conglomerados

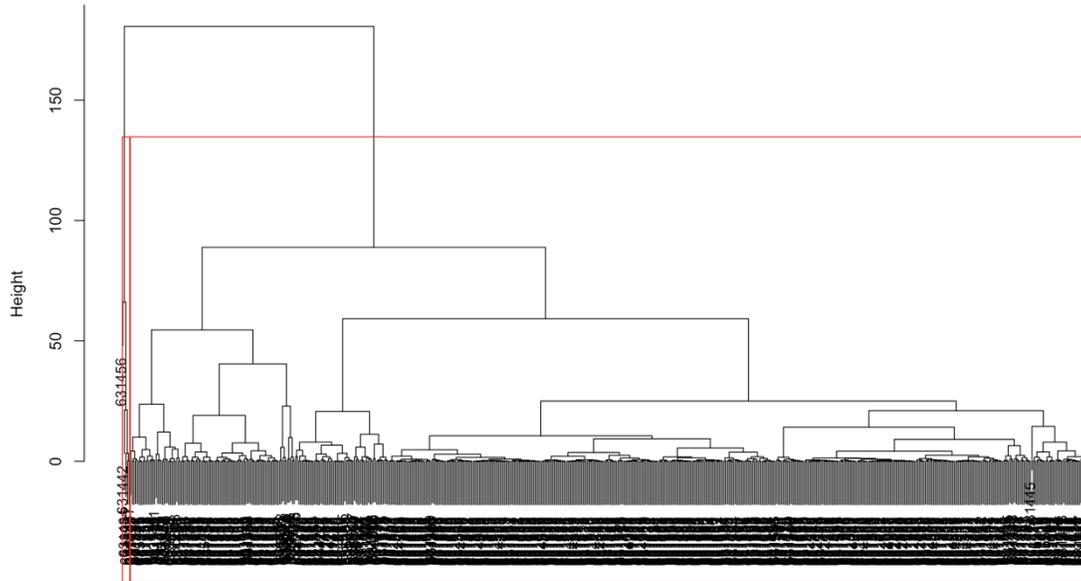


MDE2\_CPQDE2010  
hclust (\*, "centroid")



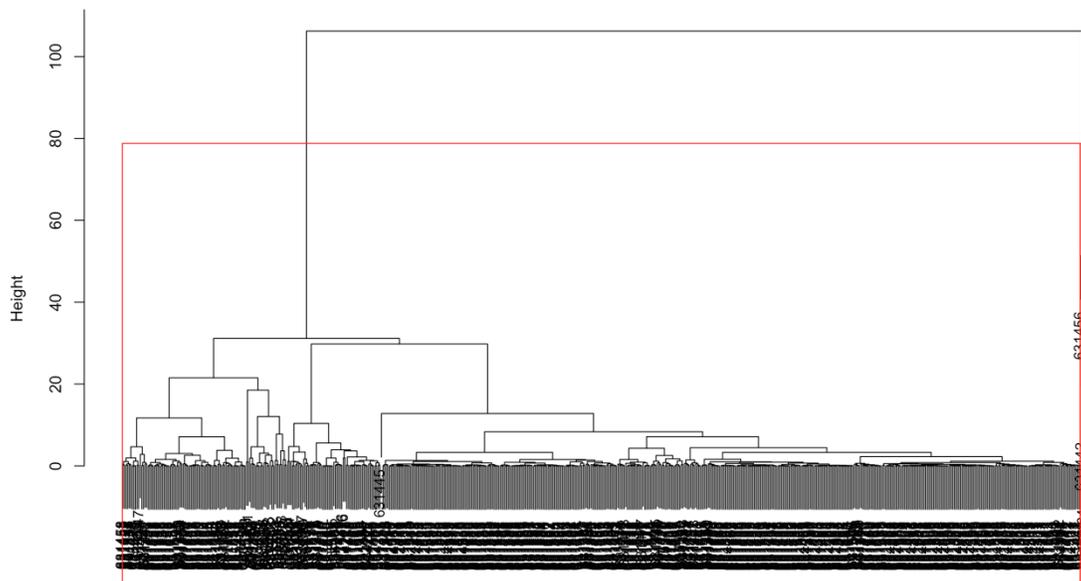
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
CAQUETA**



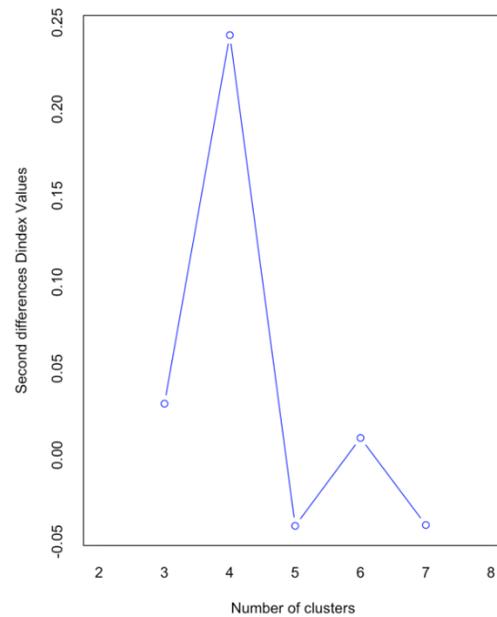
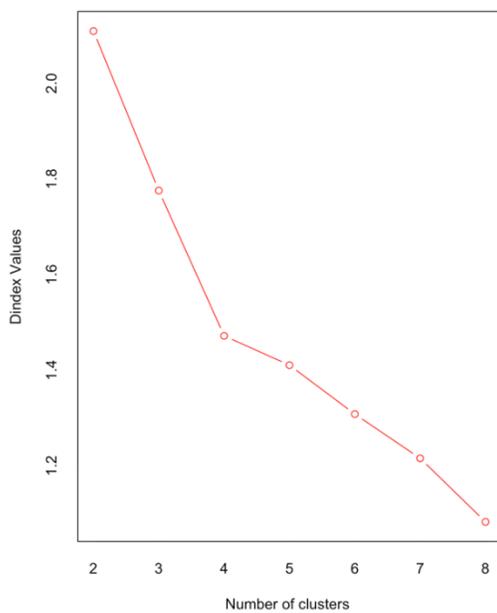
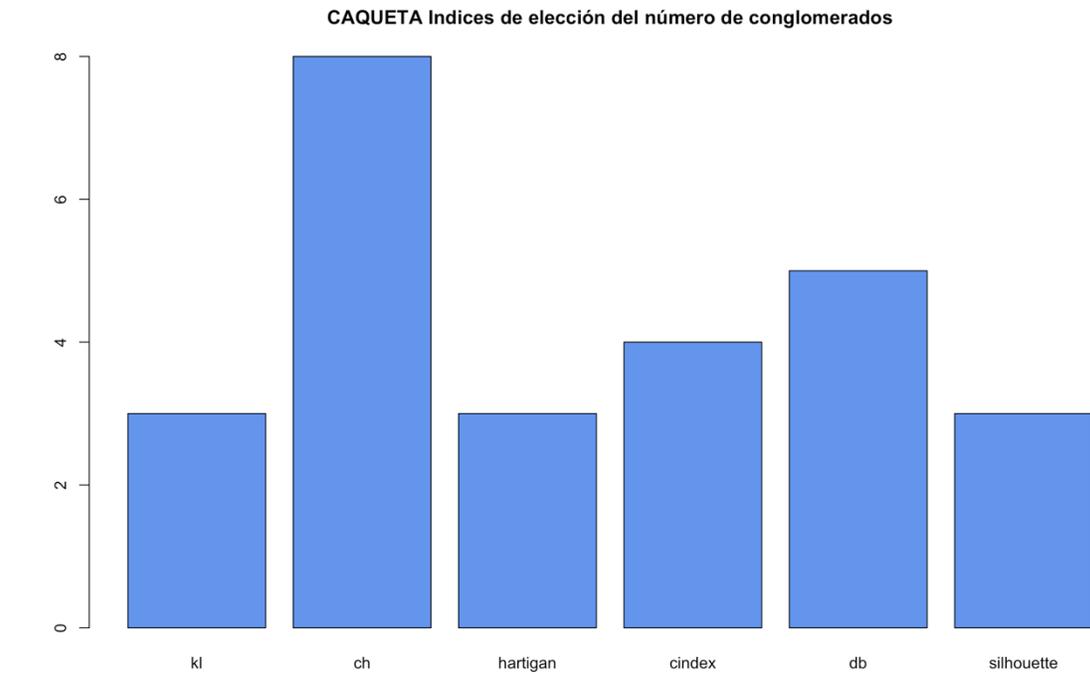
MDE2\_CPQDE2010  
hclust (\*, "complete")

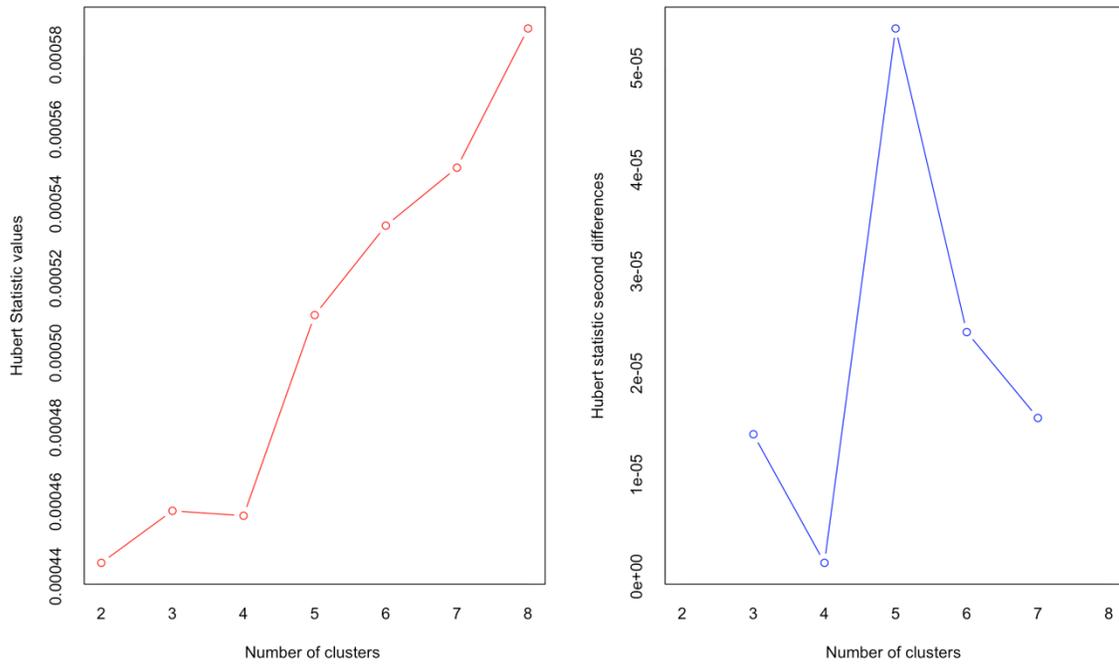
**Cluster Dendograma  
vinculacion promedio  
CAQUETA**



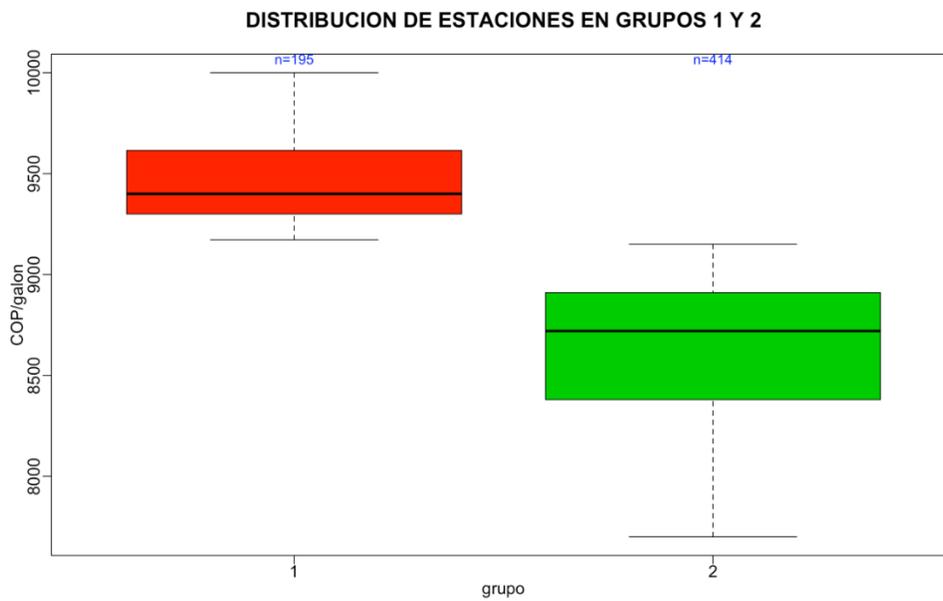
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.8.2 Selección mejor número de conglomerados método jerárquico



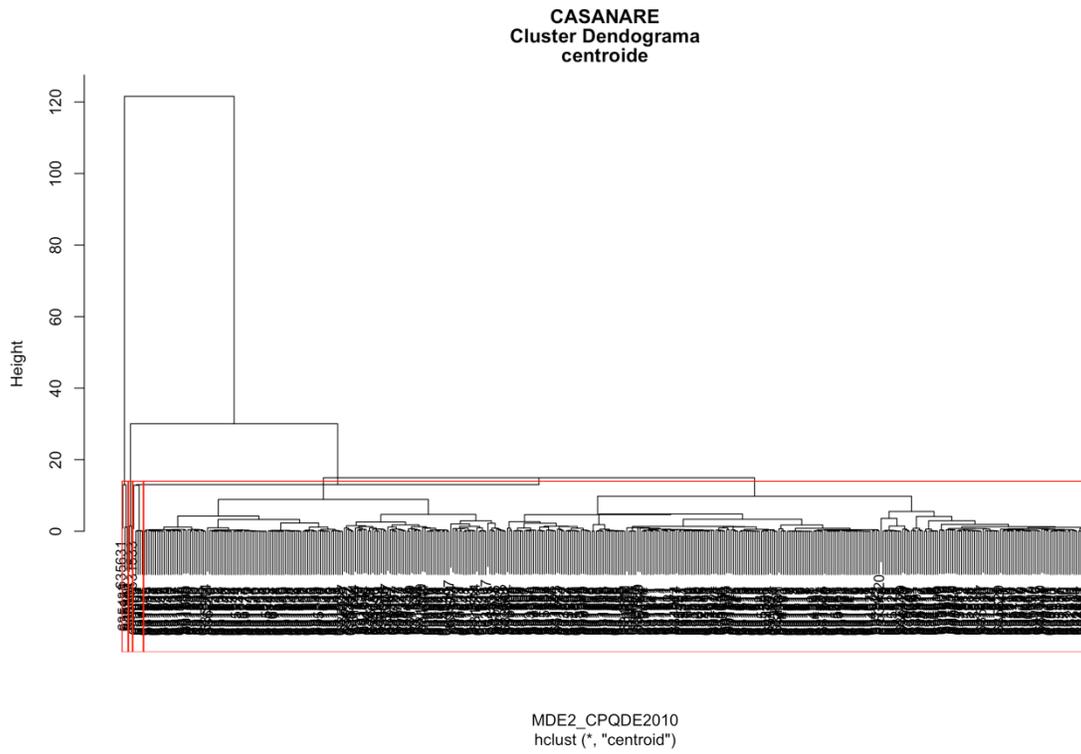


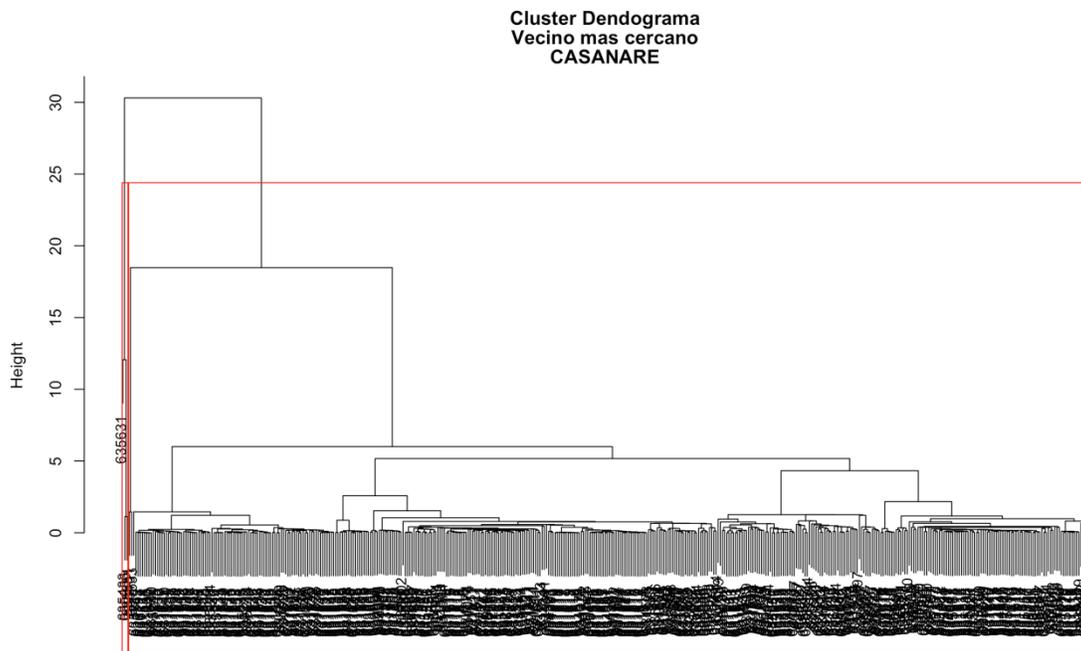
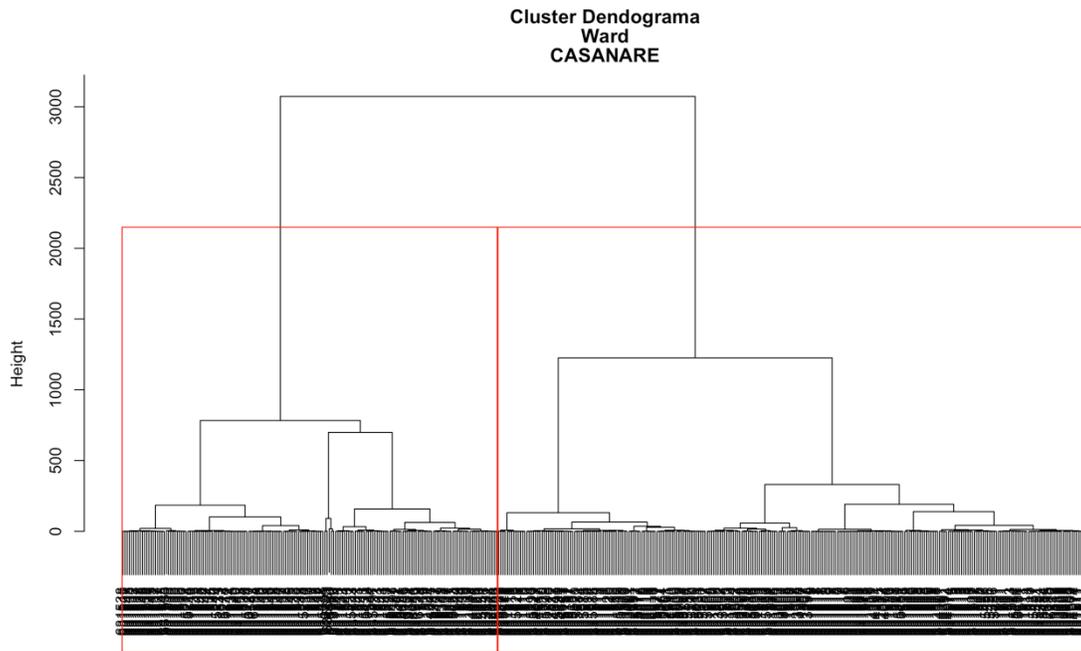
### 9.8.3 Grupos de estaciones y el comportamiento del precio



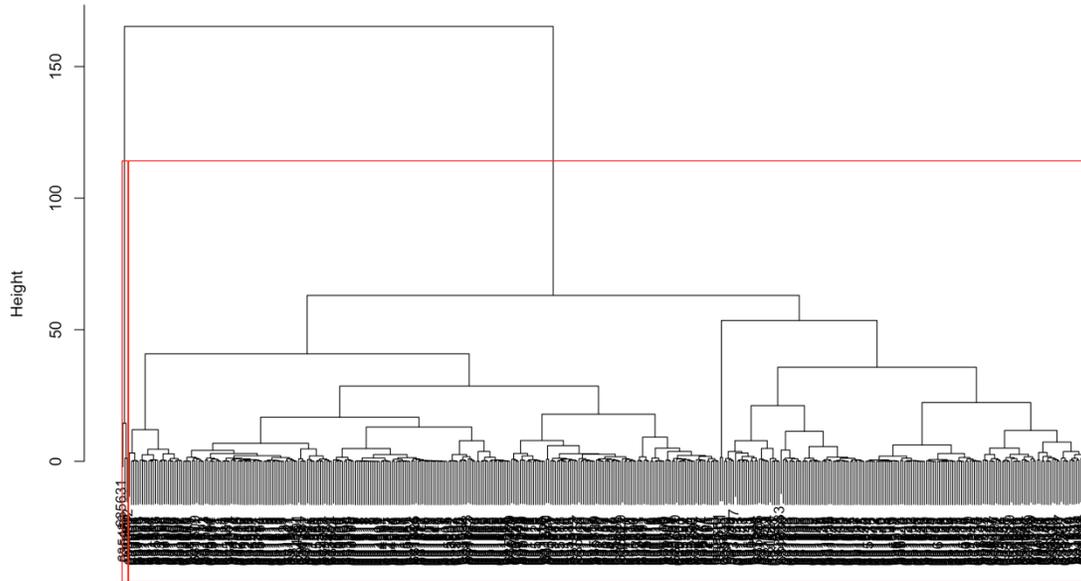
## 9.9 CASANARE

### 9.9.1 Resultados métodos conglomerados



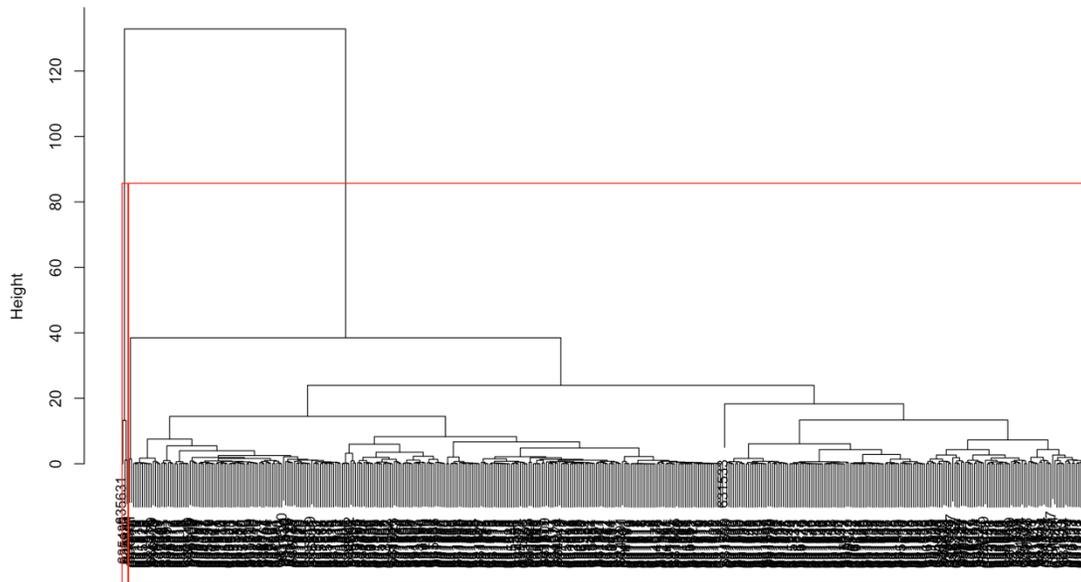


**Cluster Dendograma  
Vecino mas lejano  
CASANARE**



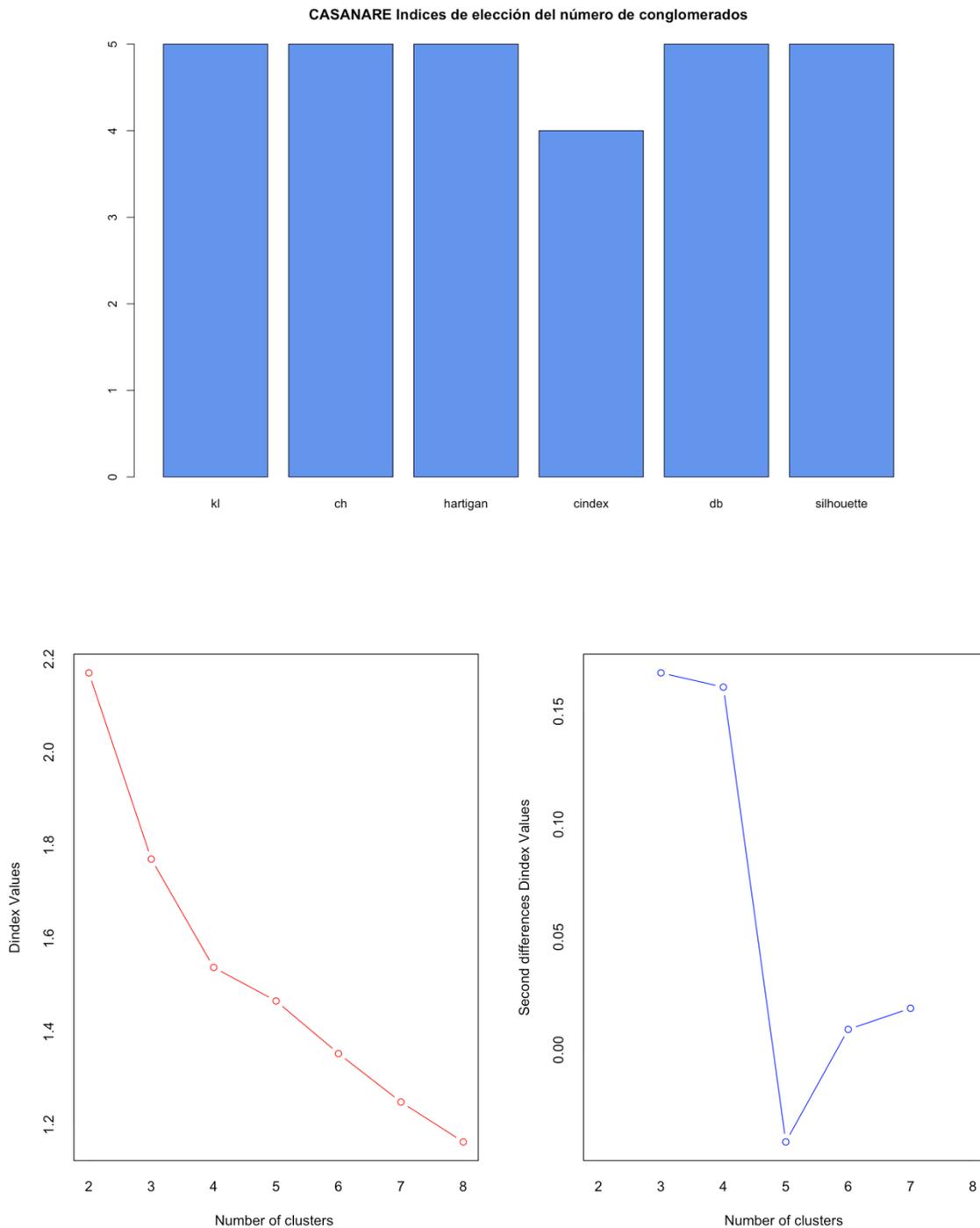
MDE2\_CPQDE2010  
hclust (\*, "complete")

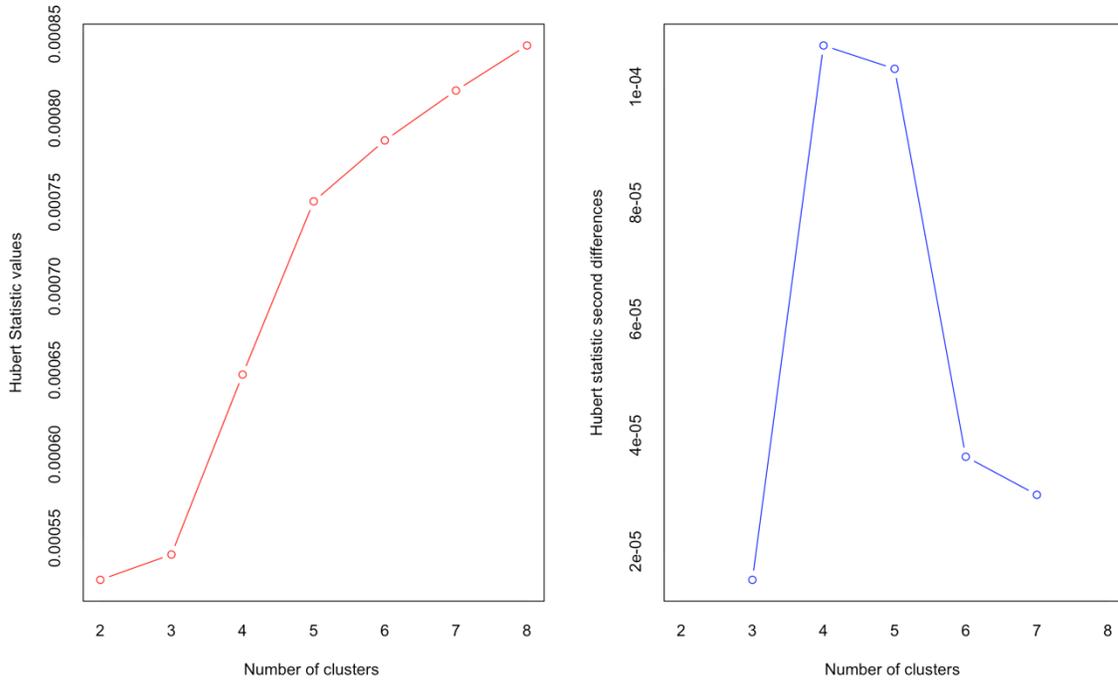
**Cluster Dendograma  
vinculacion promedio  
CASANARE**



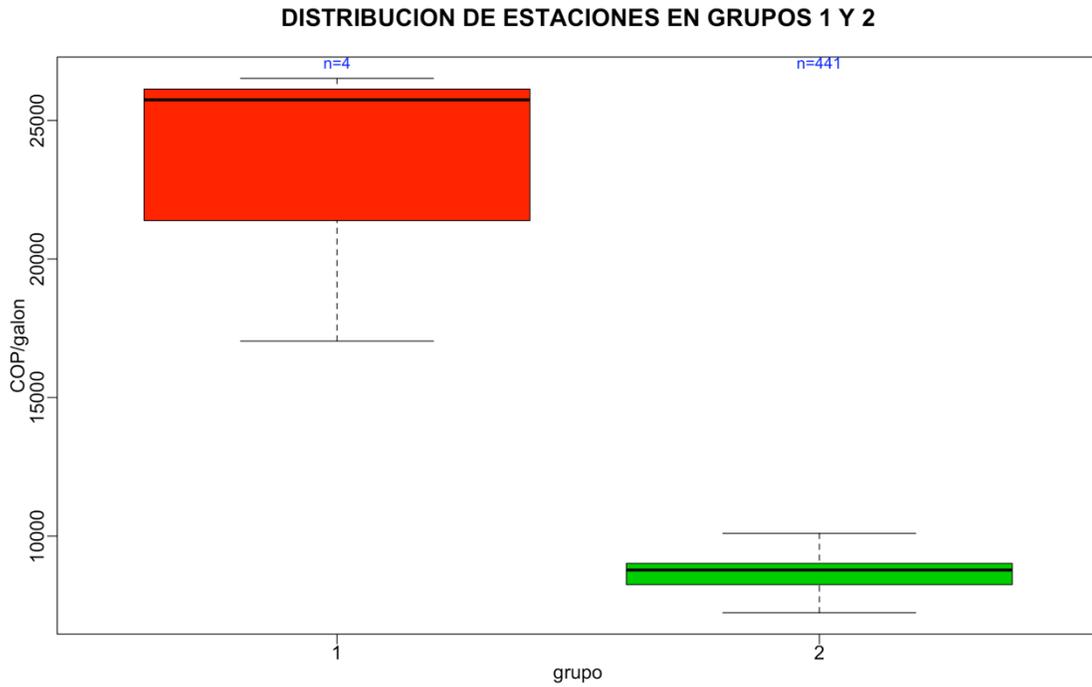
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.9.2 Selección mejor número de conglomerados método jerárquico



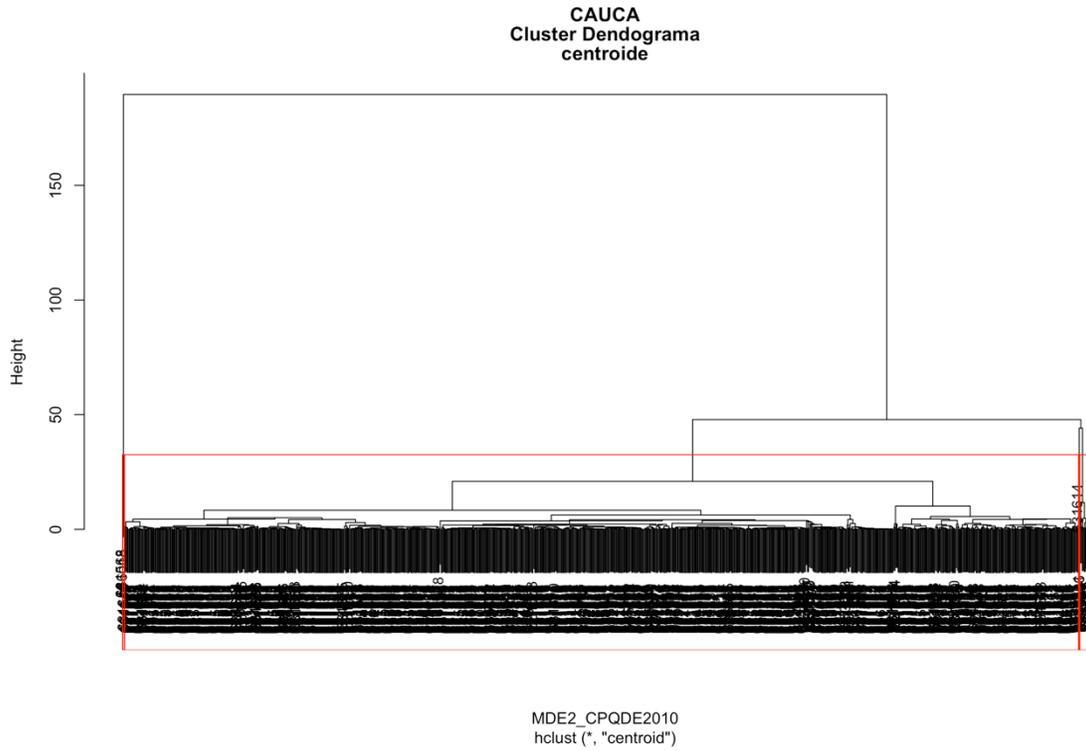


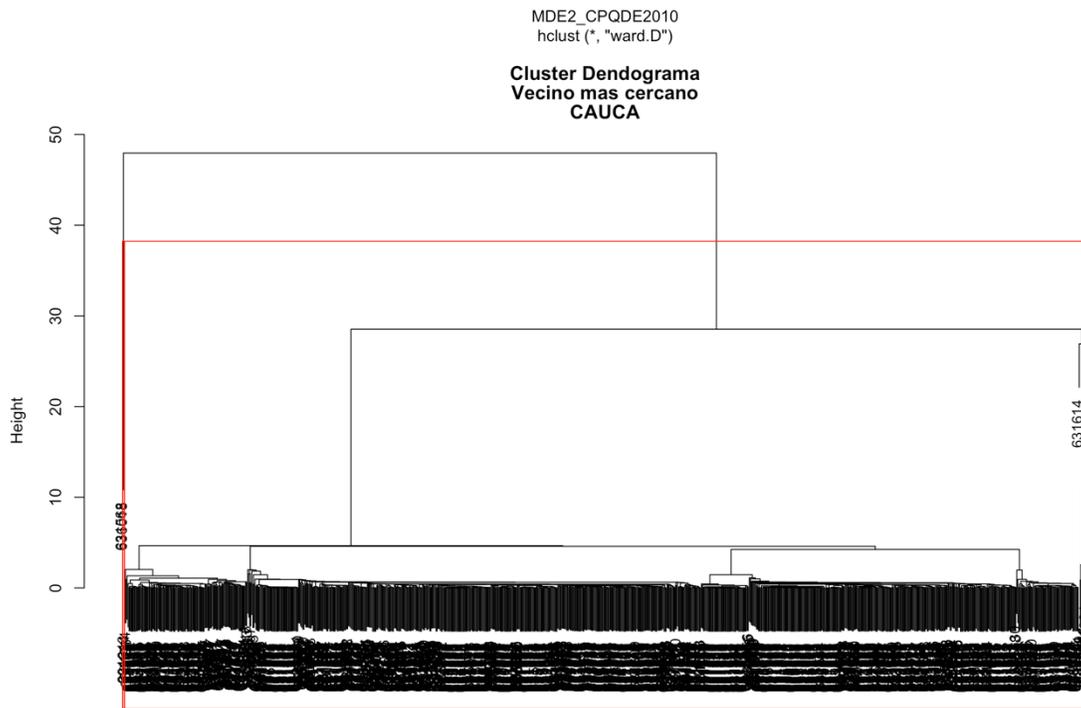
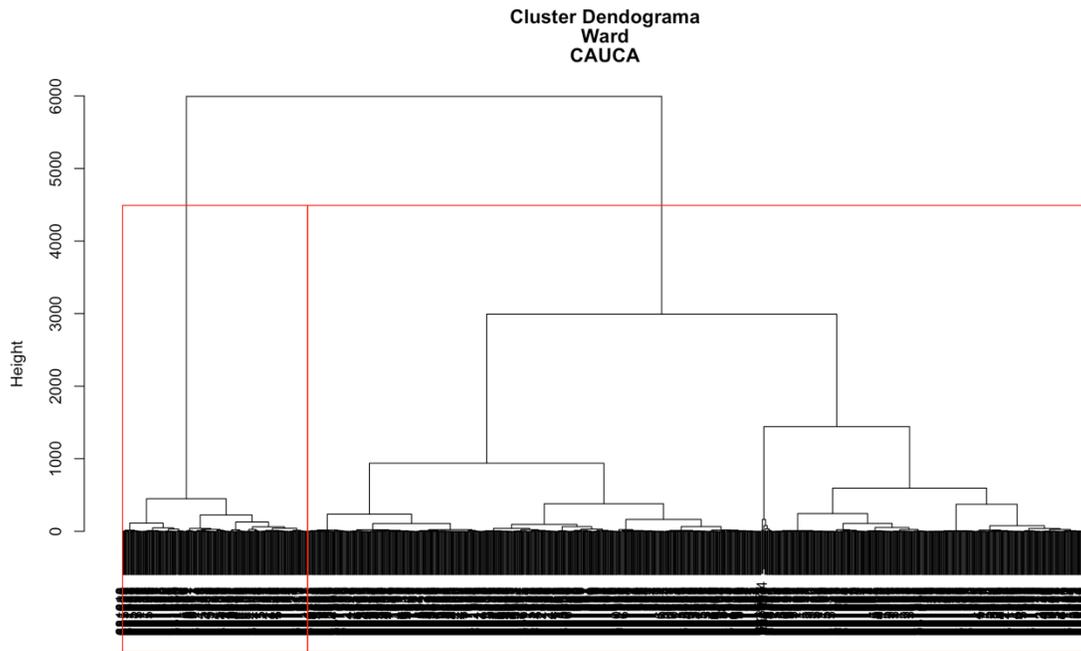
### 9.9.3 Grupos de estaciones y el comportamiento del precio



## 9.10 CAUCA

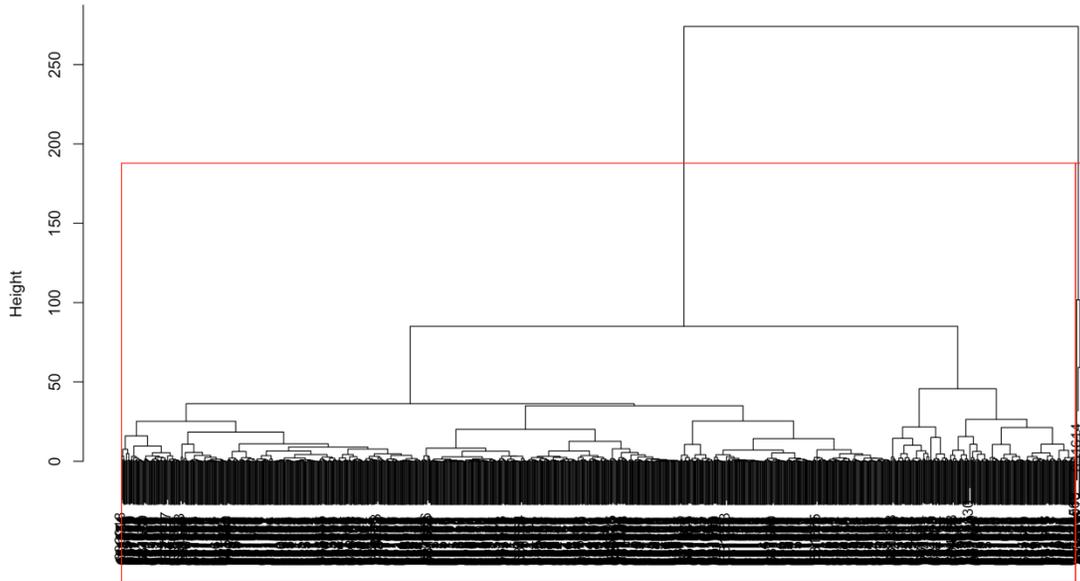
### 9.10.1 Resultados métodos conglomerados





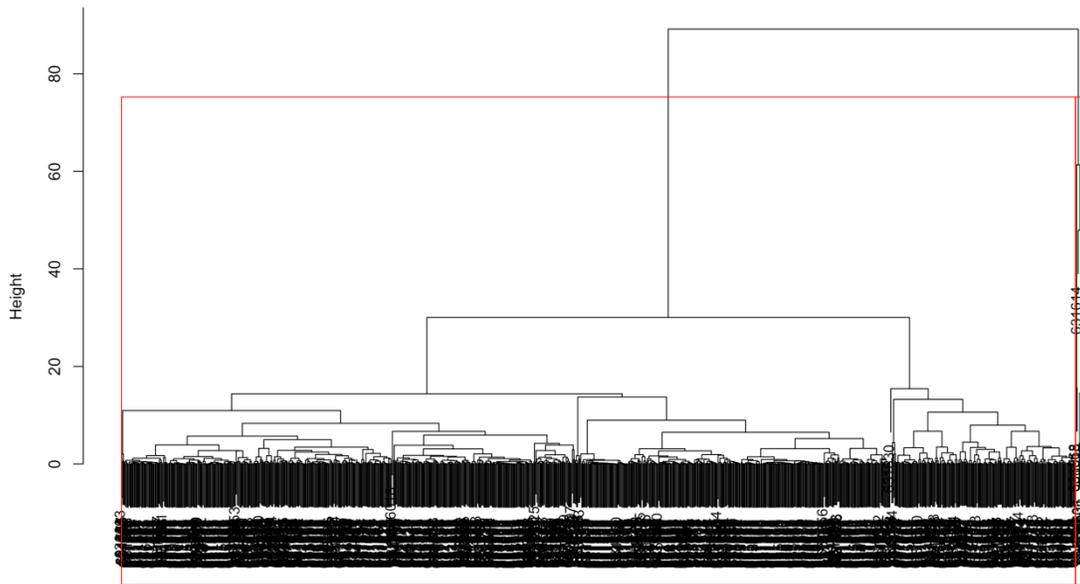
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
CAUCA**



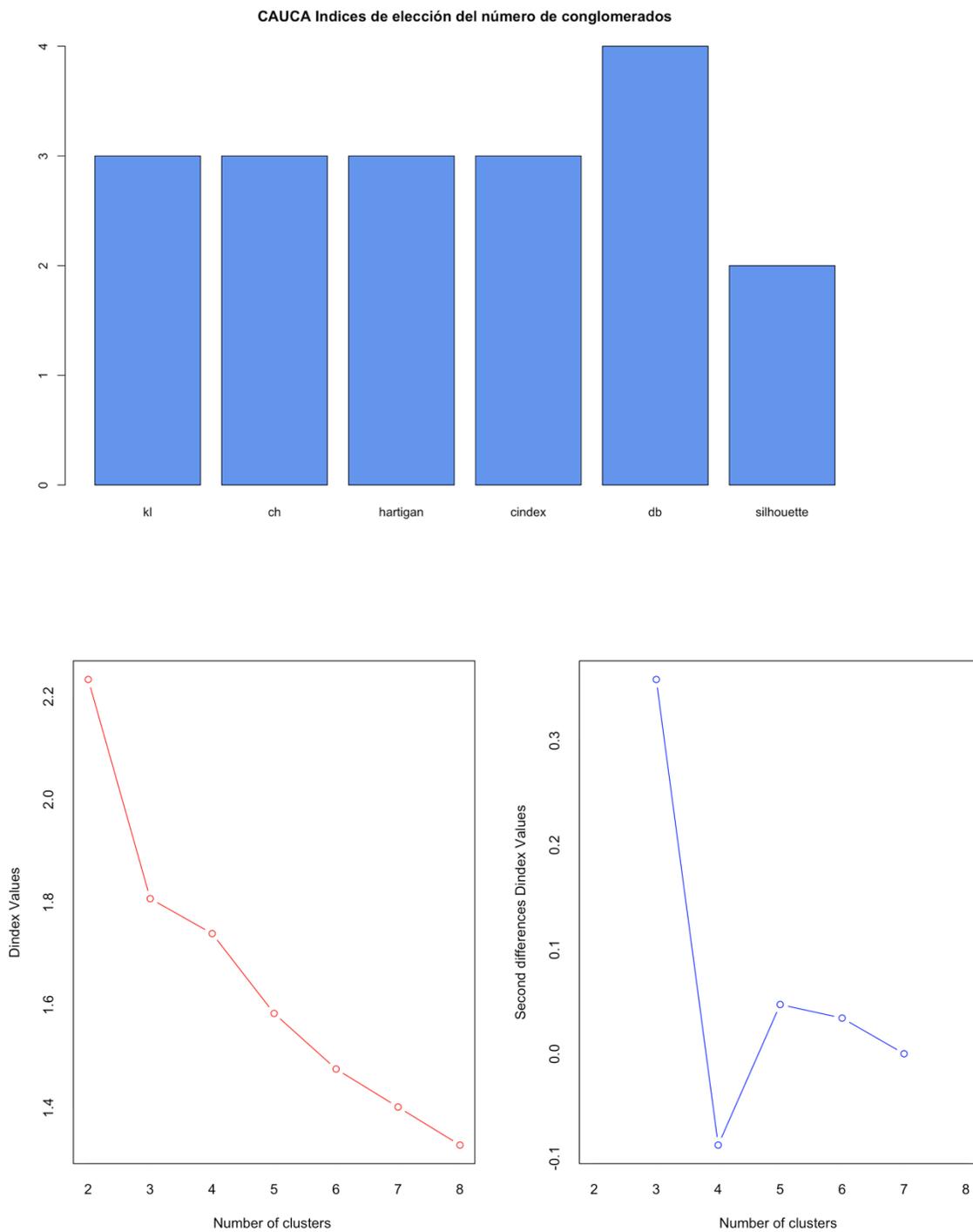
MDE2\_CPQDE2010  
hclust (\*, "complete")

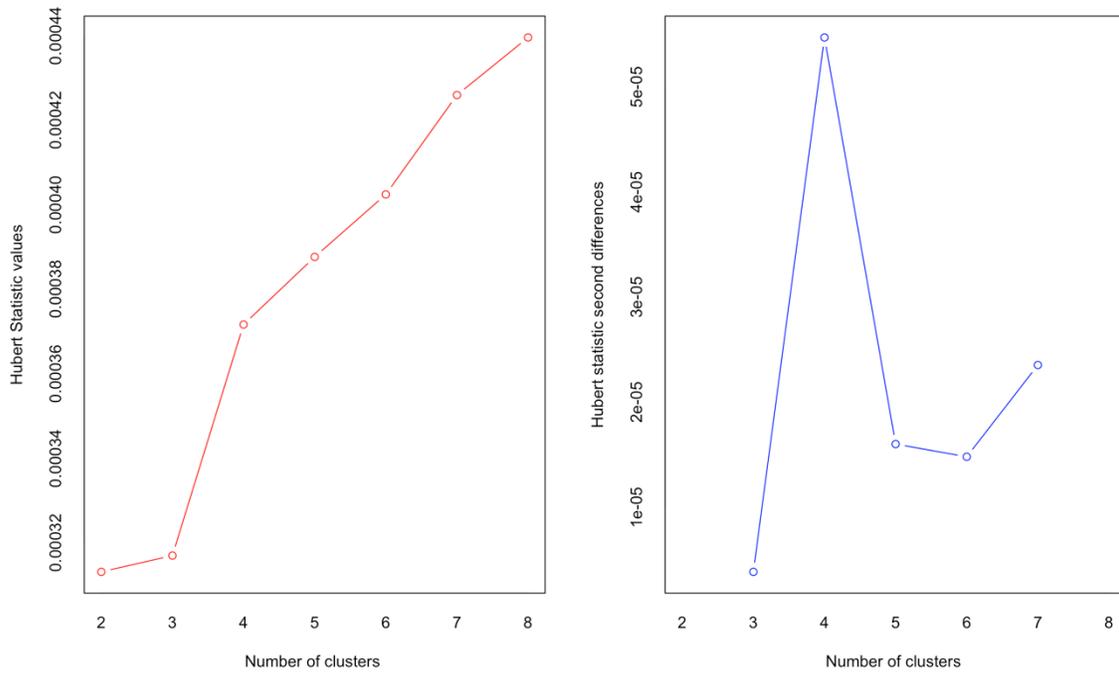
**Cluster Dendograma  
vinculacion promedio  
CAUCA**



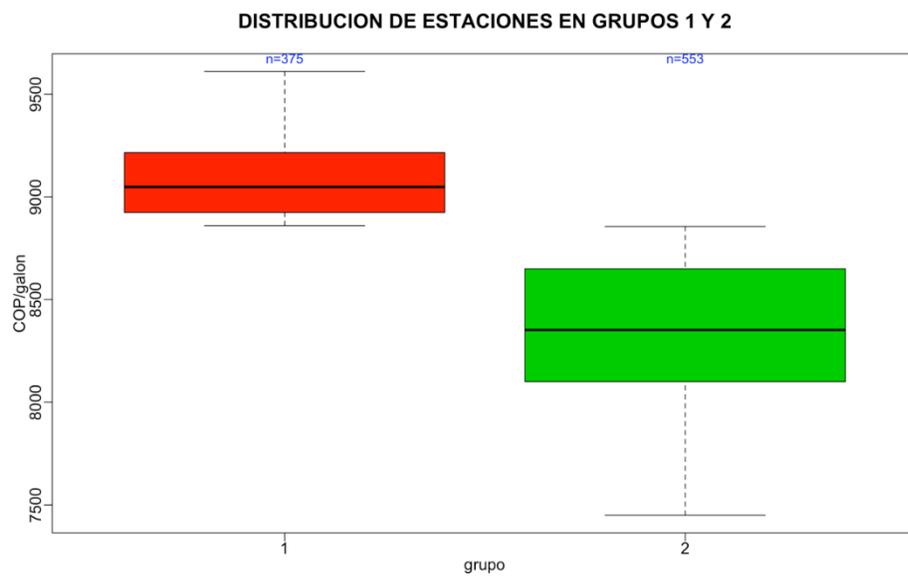
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.10.2 Selección mejor número de conglomerados método jerárquico



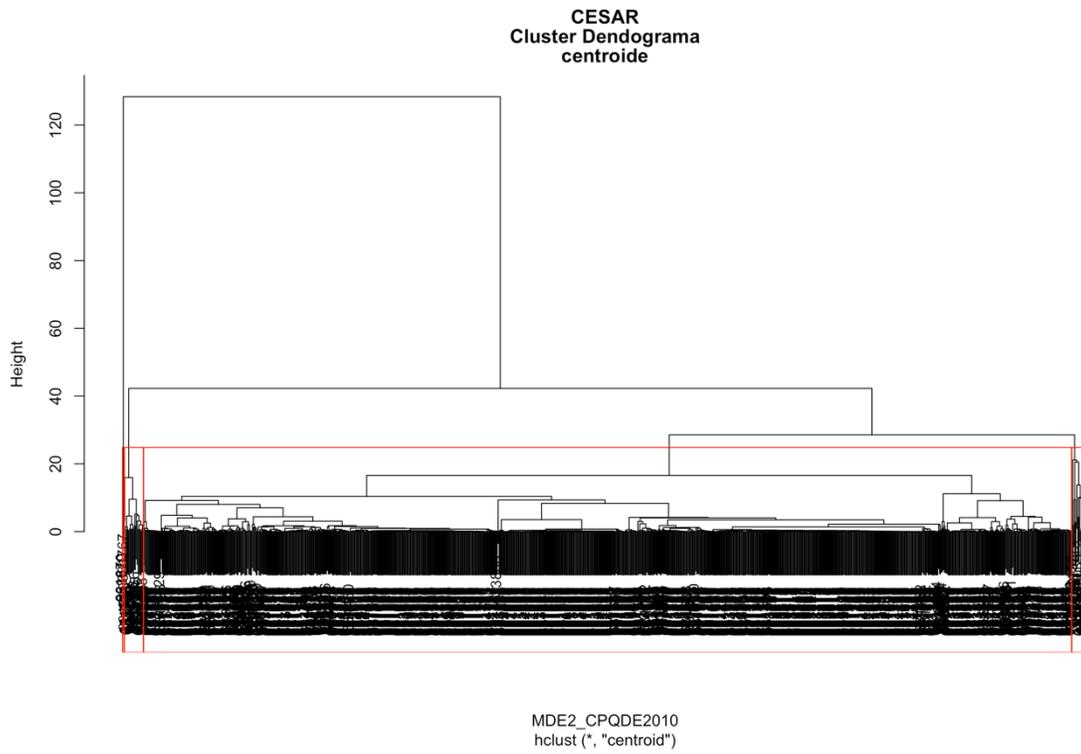


### 9.10.3 Grupos de estaciones y el comportamiento del precio

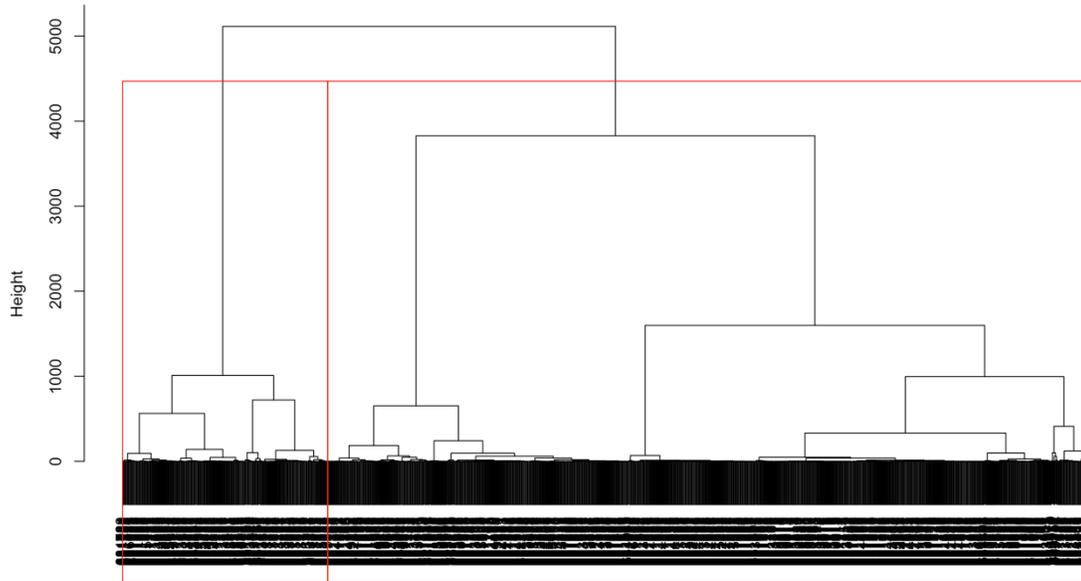


## 9.11 CESAR

### 9.11.1 Resultados métodos conglomerados

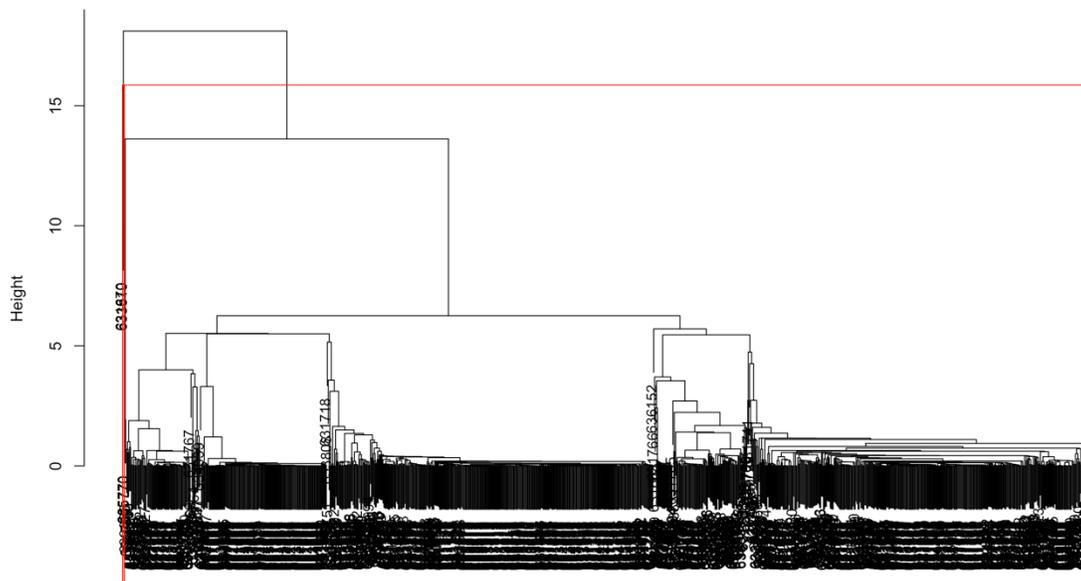


Cluster Dendrograma  
Ward  
CESAR



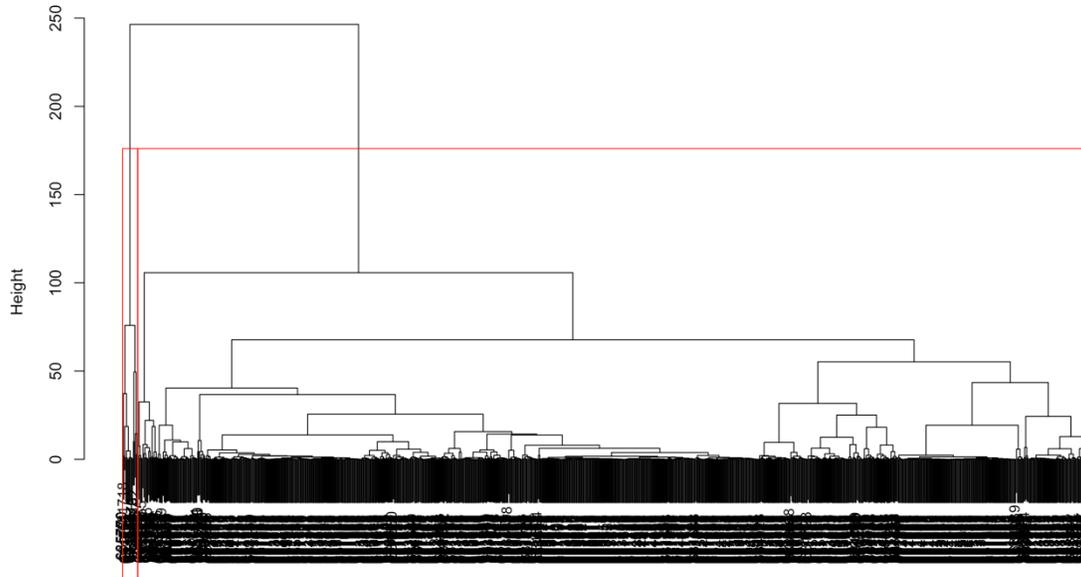
MDE2\_CPQDE2010  
hclust (\*, "ward.D")

Cluster Dendrograma  
Vecino mas cercano  
CESAR



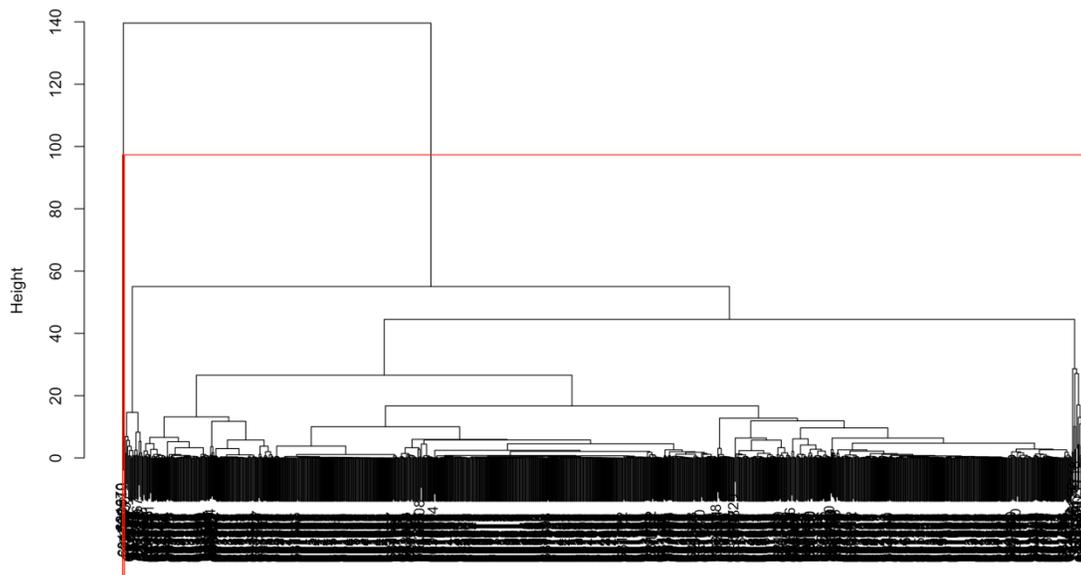
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
CESAR**



MDE2\_CPQDE2010  
hclust (\*, "complete")

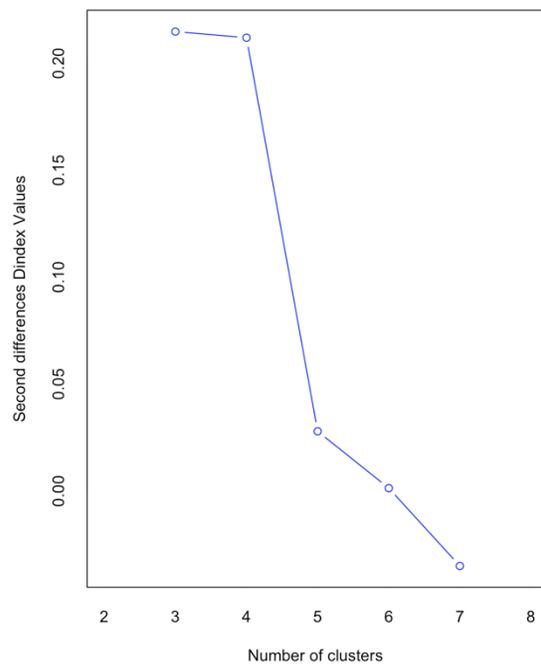
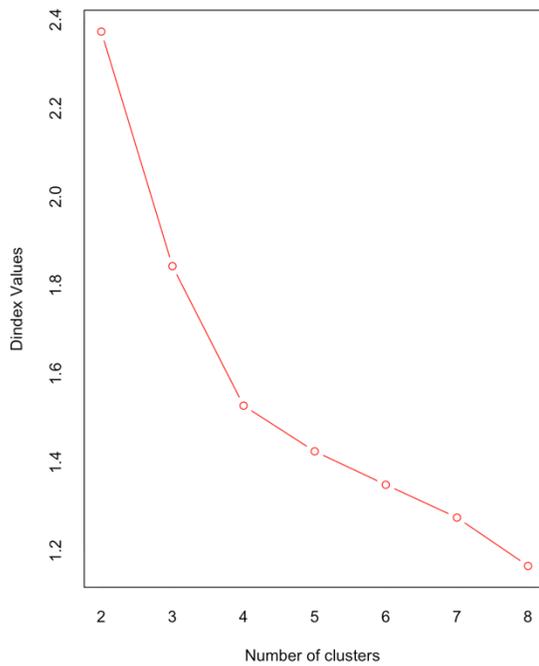
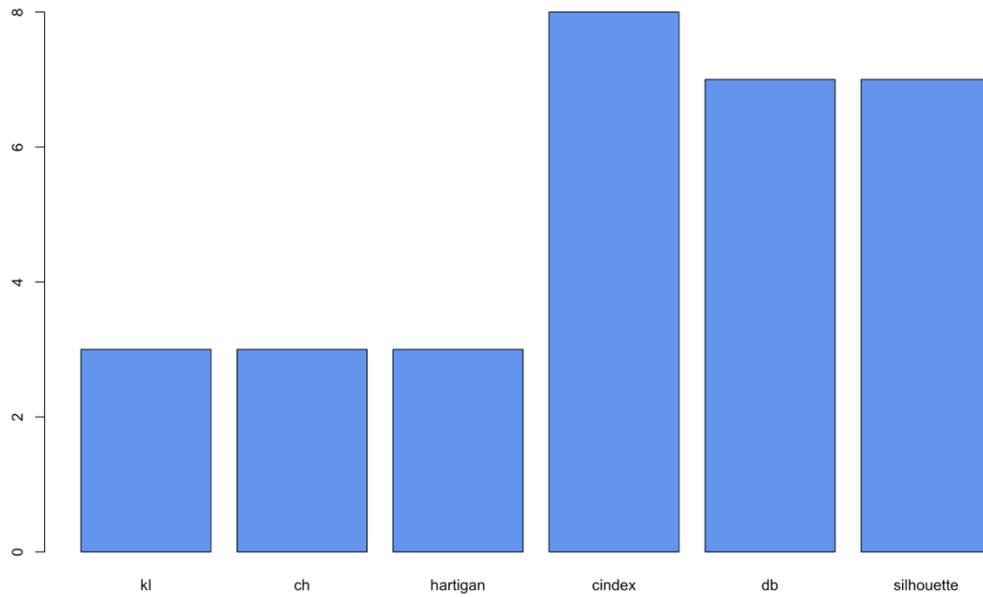
**Cluster Dendograma  
vinculacion promedio  
CESAR**

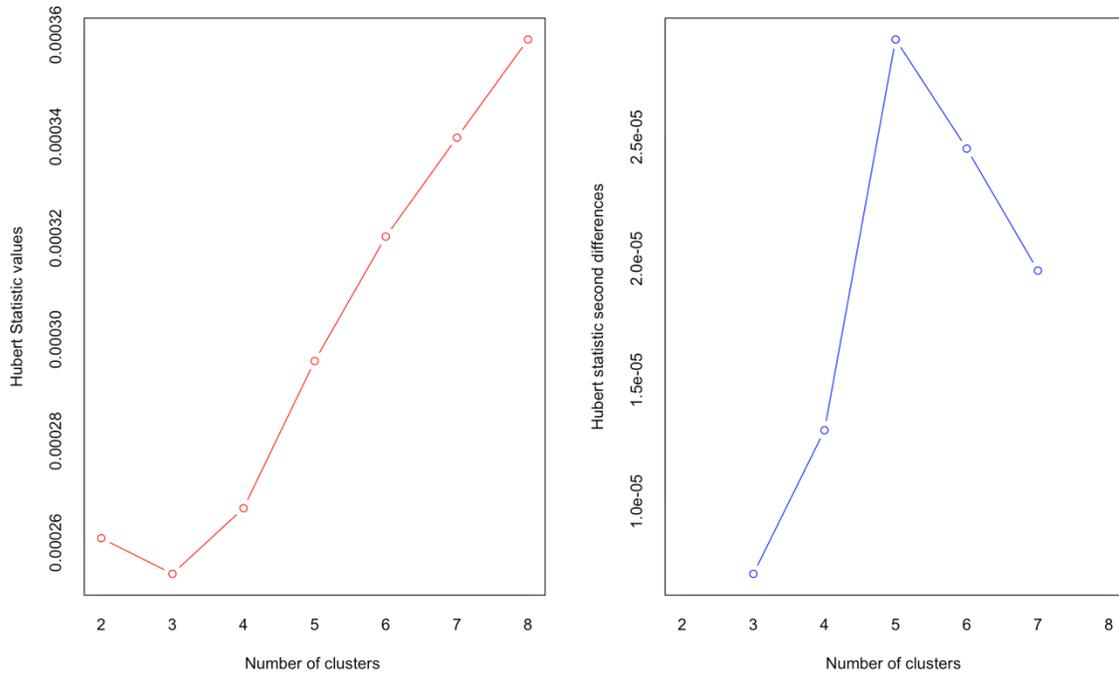


MDE2\_CPQDE2010  
hclust (\*, "average")

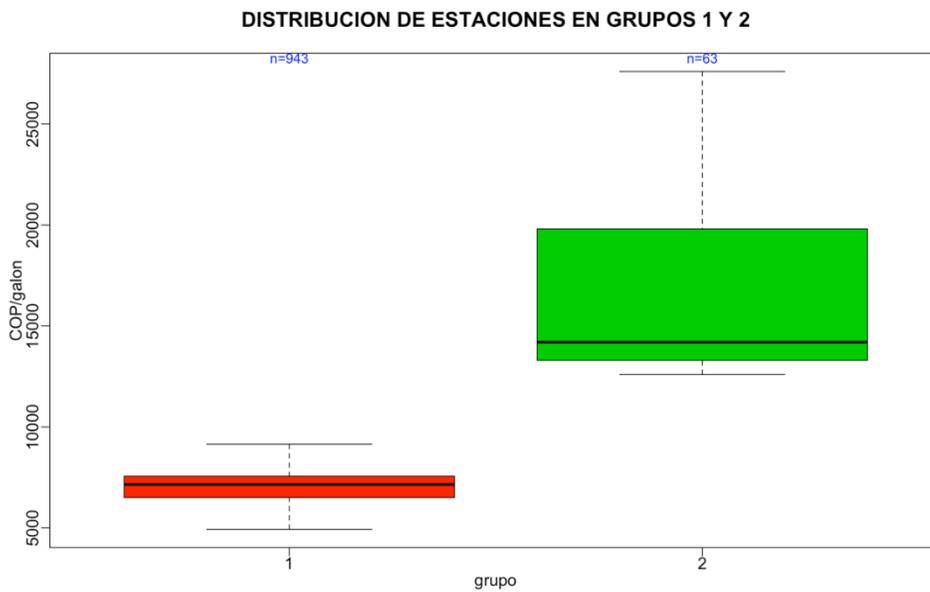
### 9.11.2 Selección mejor número de conglomerados método jerárquico

CESAR Indices de elección del número de conglomerados



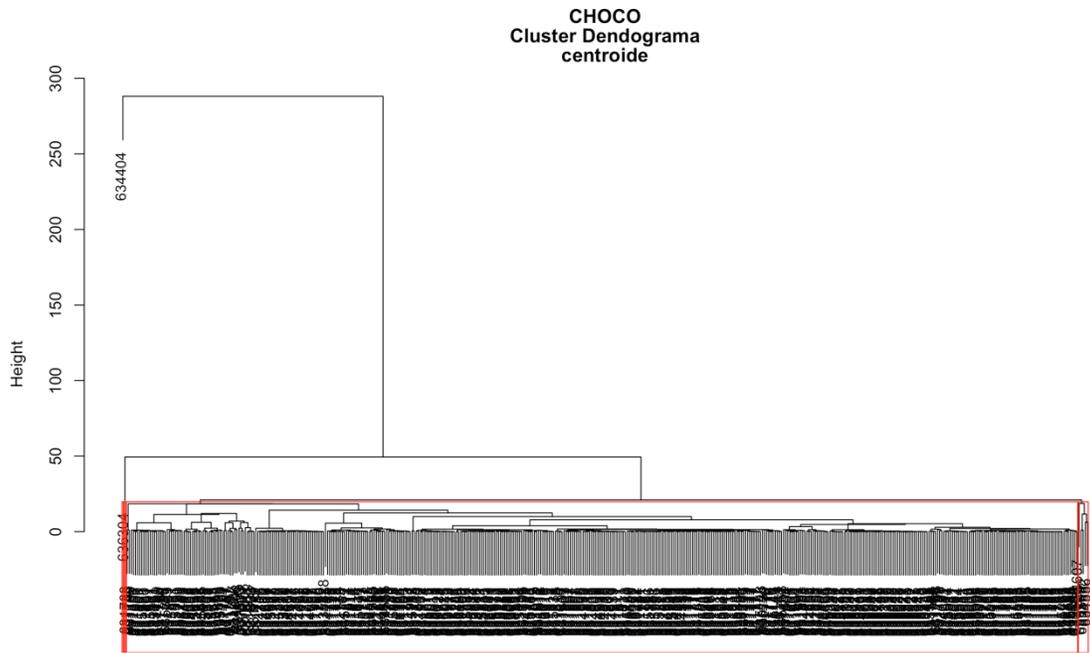


### 9.11.3 Grupos de estaciones y el comportamiento del precio

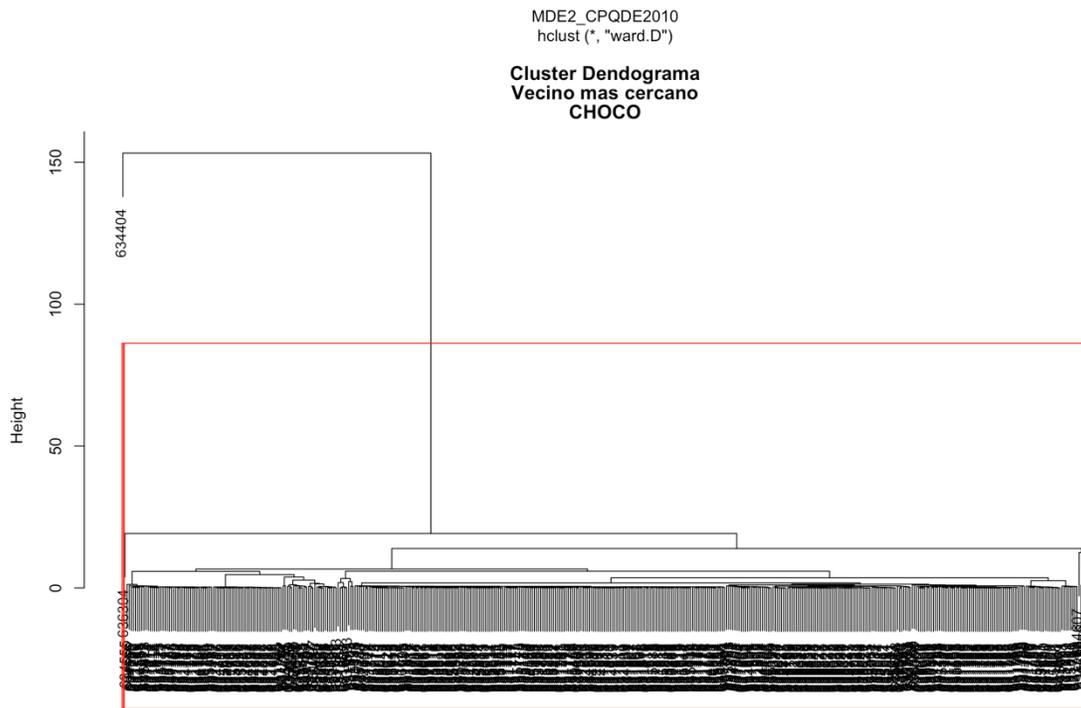
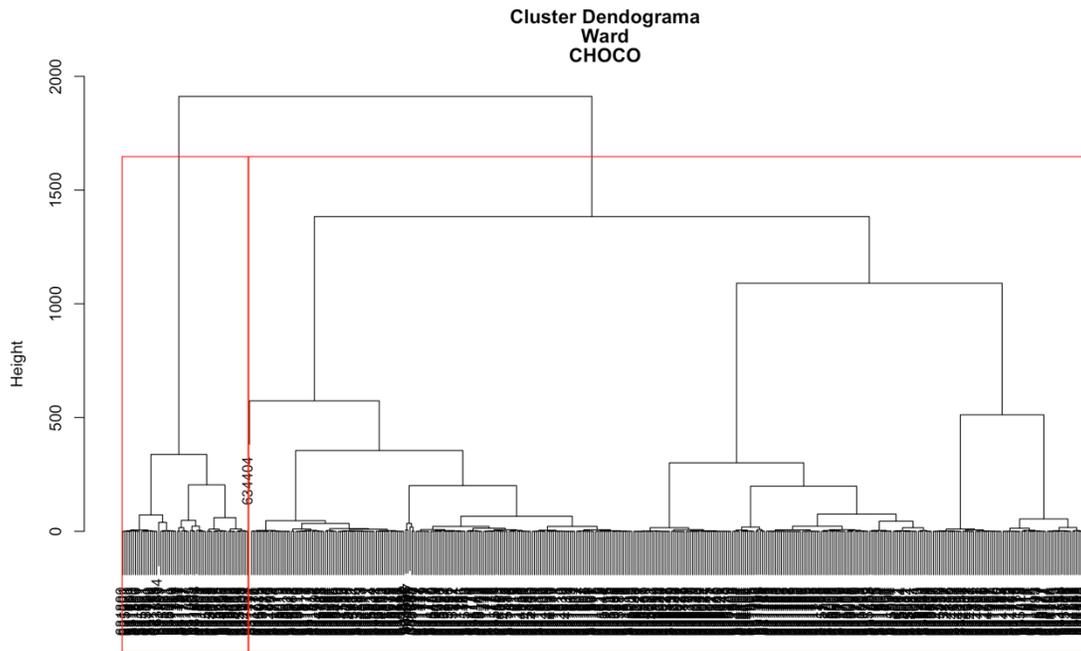


## 9.12 CHOCO

### 9.12.1 Resultados métodos conglomerados

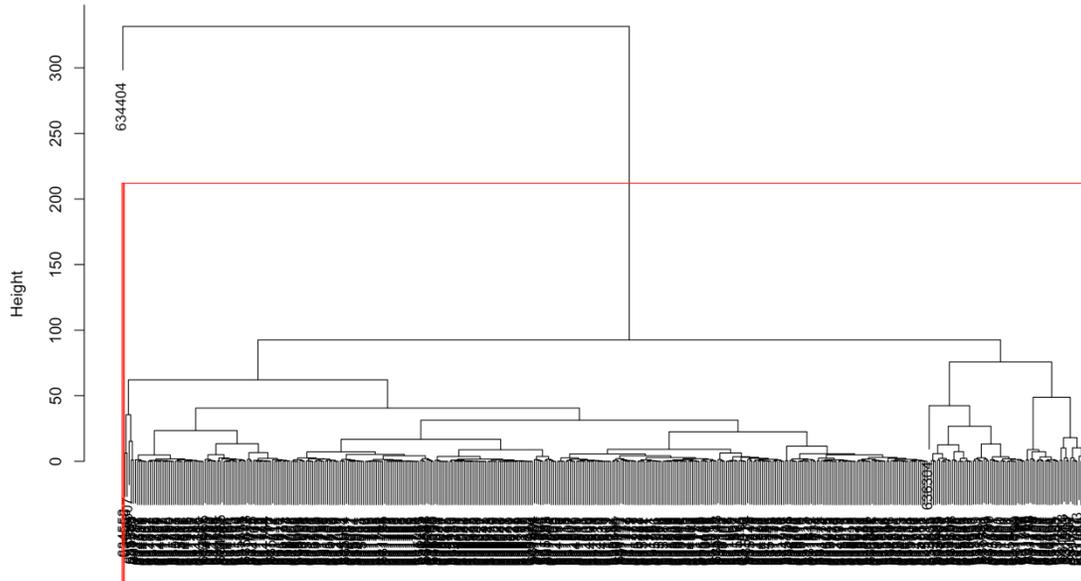


MDE2\_CPQDE2010  
hclust (\*, "centroid")



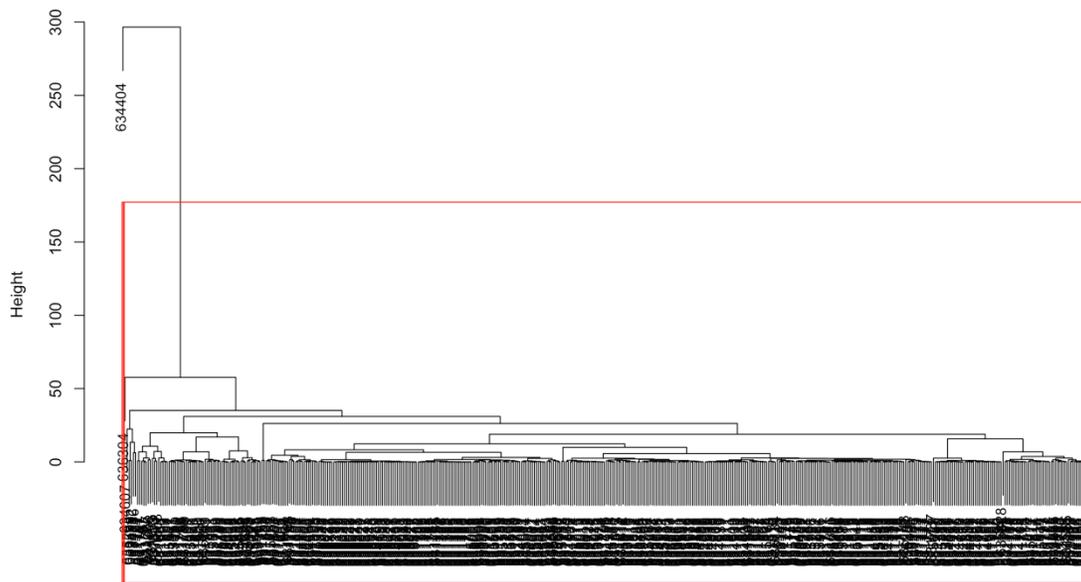
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendrograma  
Vecino mas lejano  
CHOCO**



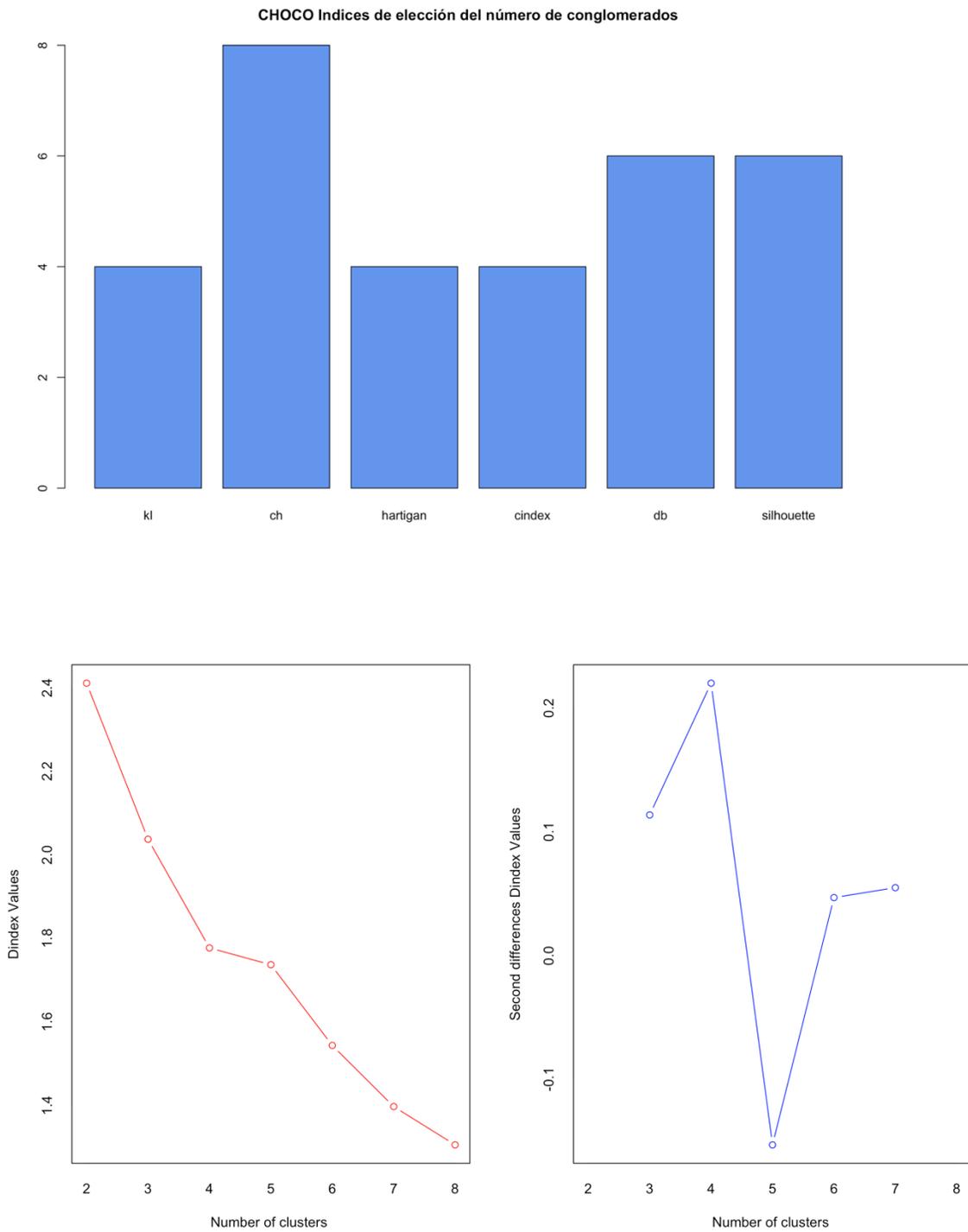
MDE2\_CPQDE2010  
hclust (\*, "complete")

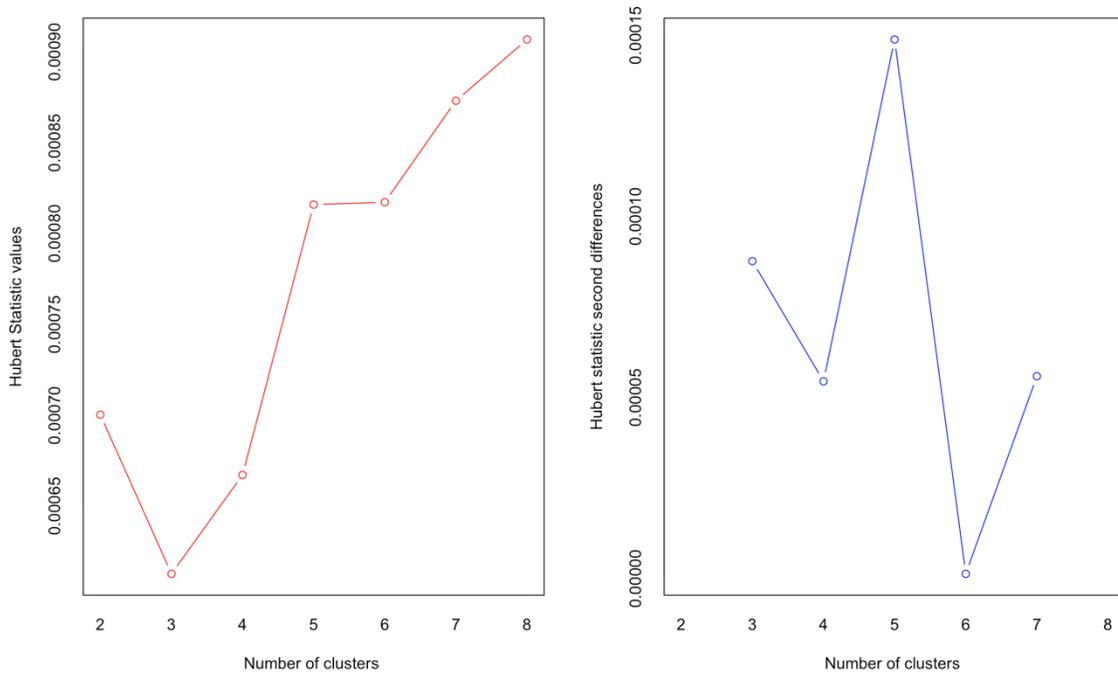
**Cluster Dendrograma  
vinculacion promedio  
CHOCO**



MDE2\_CPQDE2010  
hclust (\*, "average")

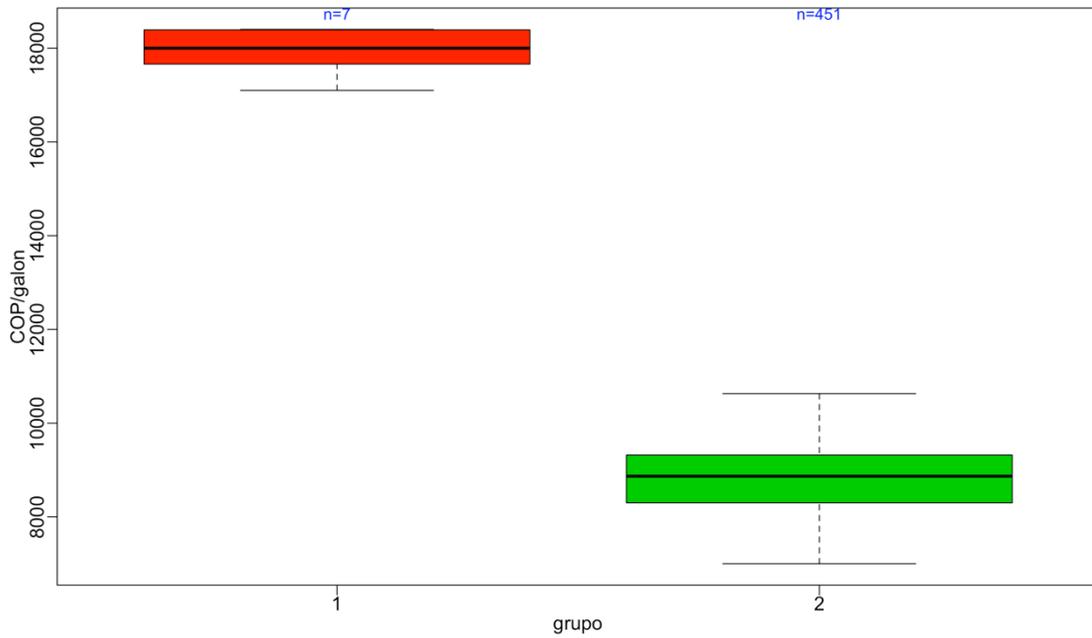
### 9.12.2 Selección mejor número de conglomerados método jerárquico





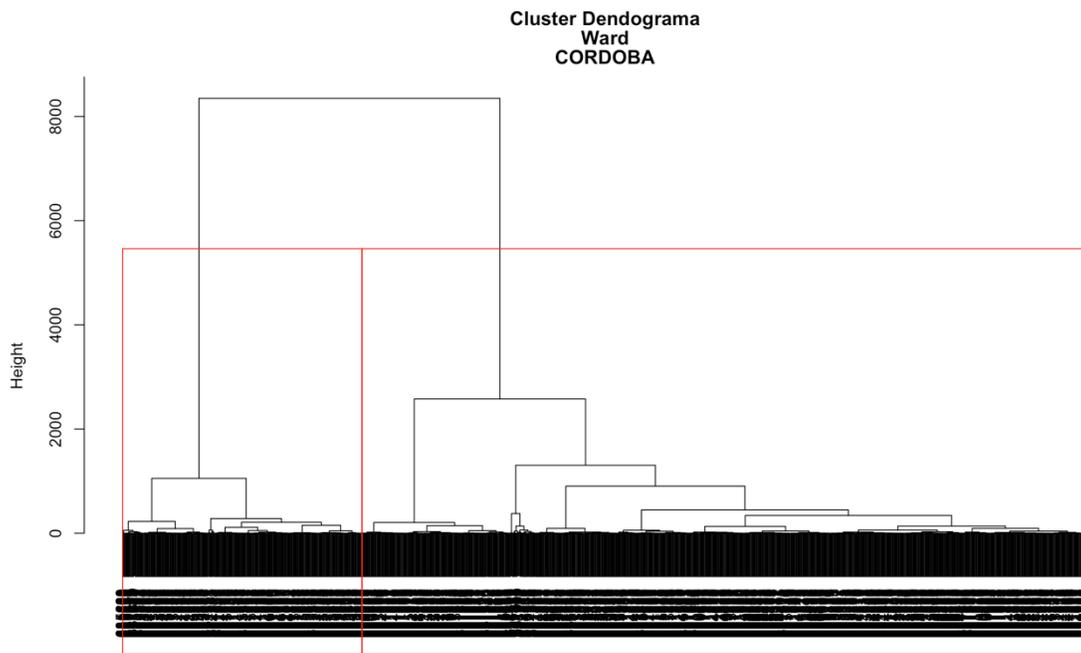
### 9.12.3 Grupos de estaciones y el comportamiento del precio

DISTRIBUCION DE ESTACIONES EN GRUPOS 1 Y 2



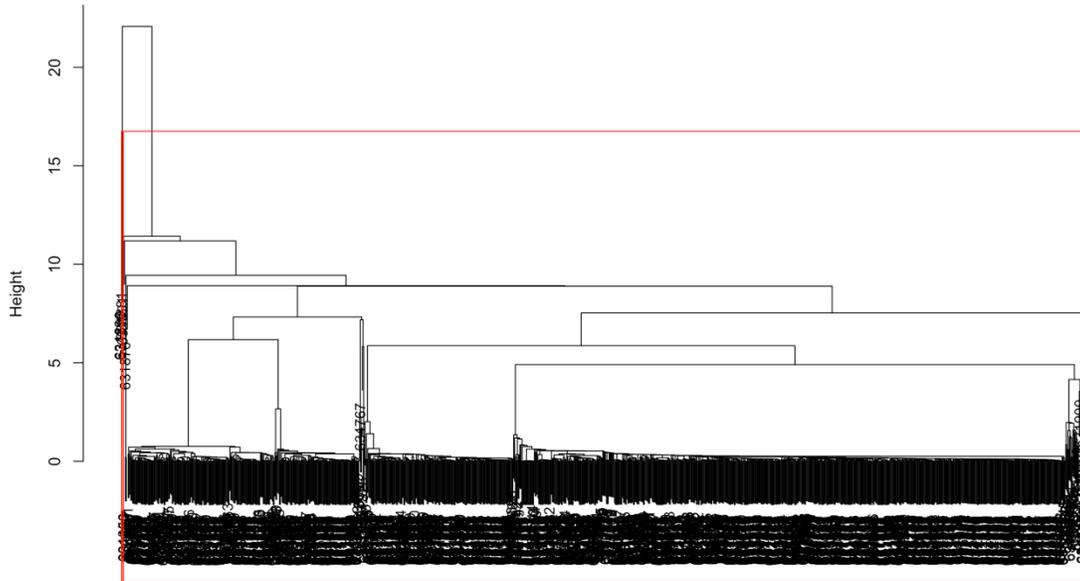
## 9.13 CORDOBA

### 9.13.1 Resultados métodos conglomerados



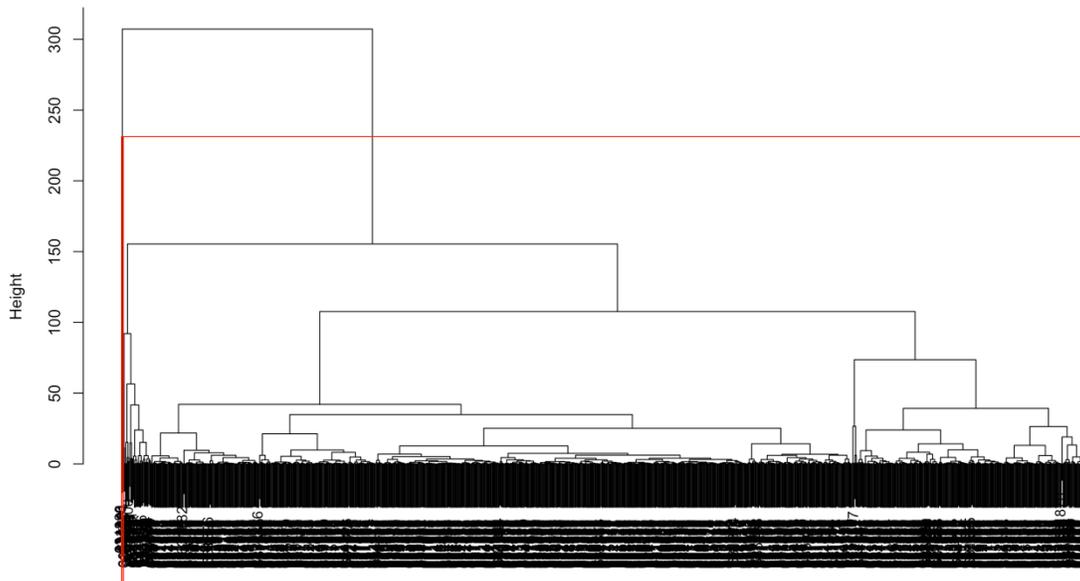
MDE2\_CPQDE2010  
hclust ("ward.D")

**Cluster Dendograma  
Vecino mas cercano  
CORDOBA**

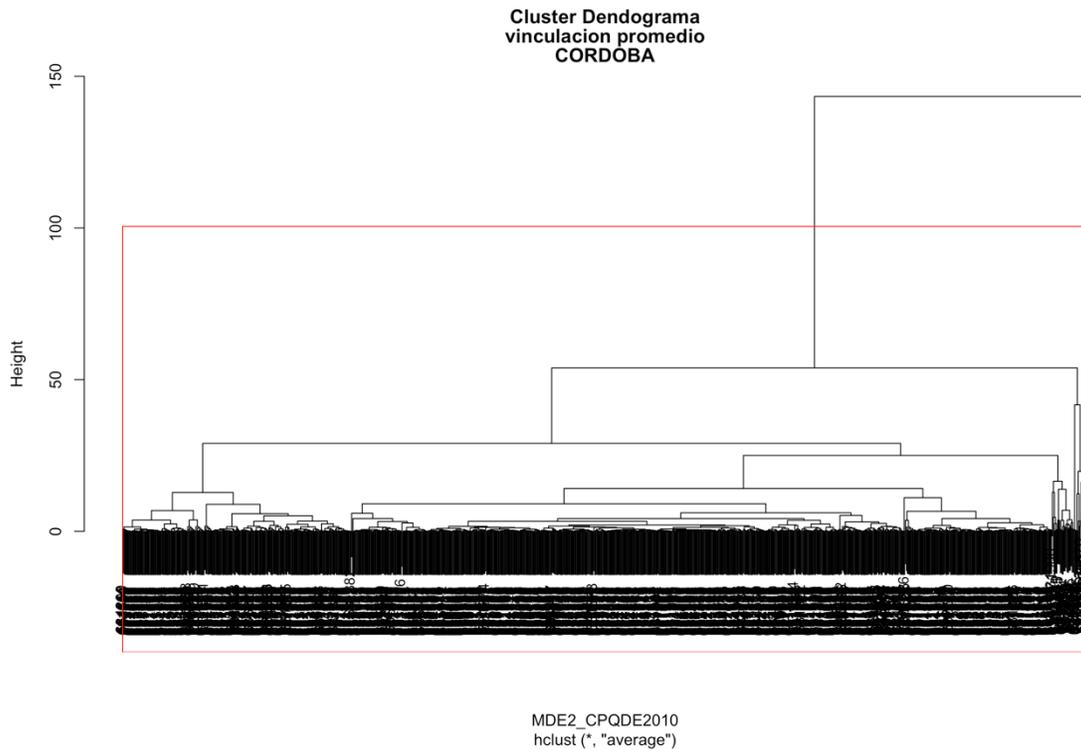


MDE2\_CPQDE2010  
hclust (\*, "single")

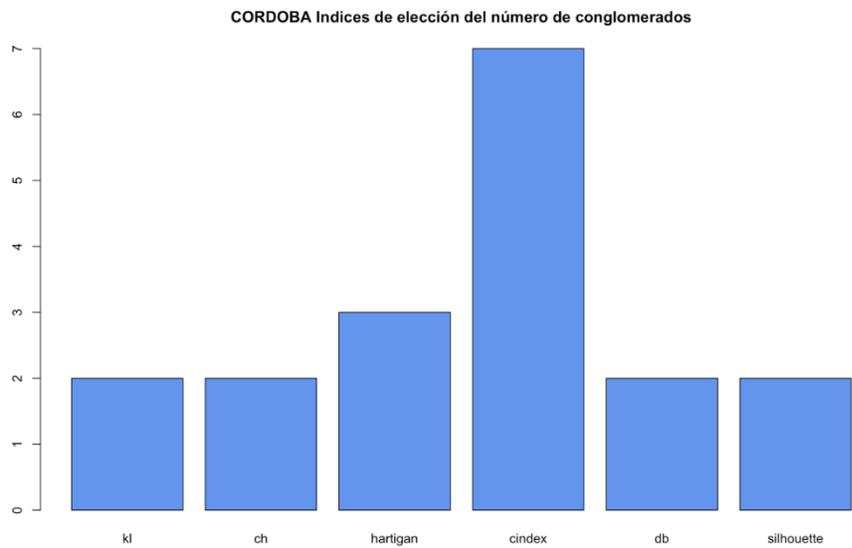
**Cluster Dendograma  
Vecino mas lejano  
CORDOBA**

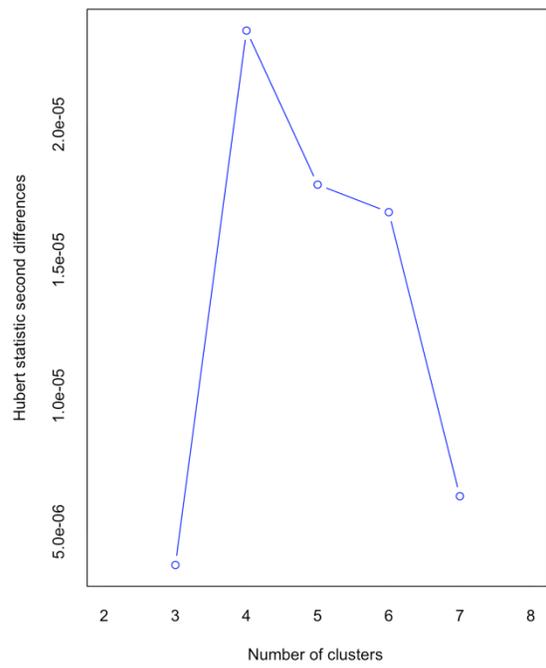
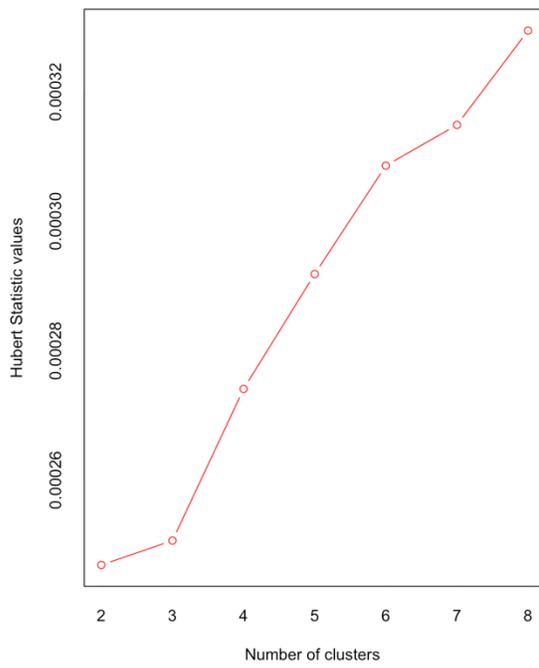
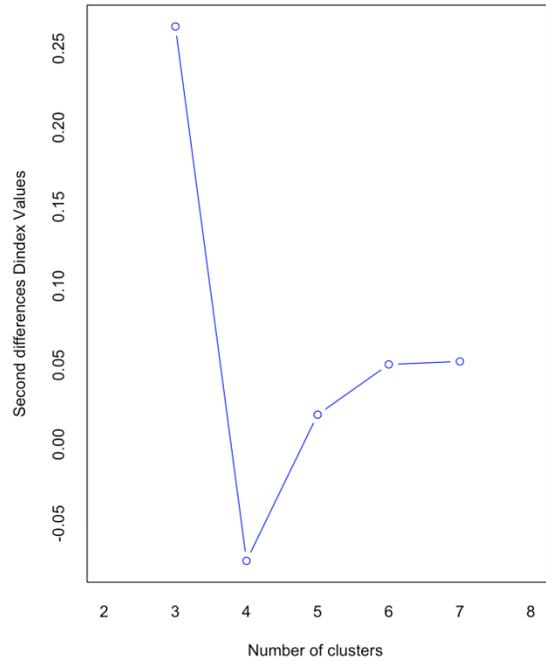
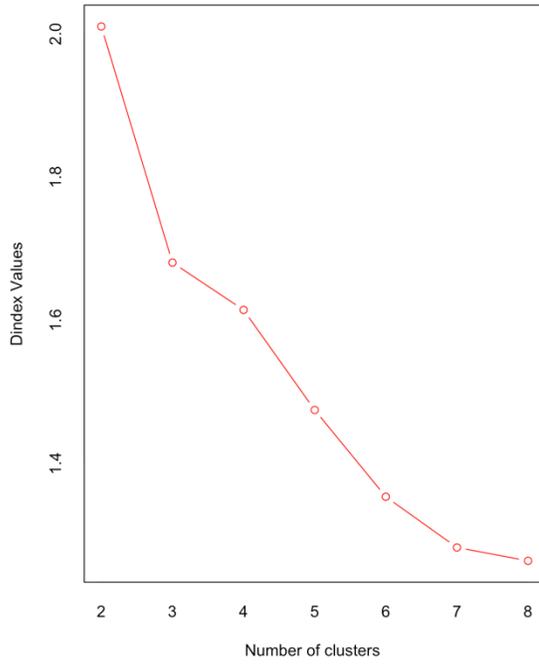


MDE2\_CPQDE2010  
hclust (\*, "complete")

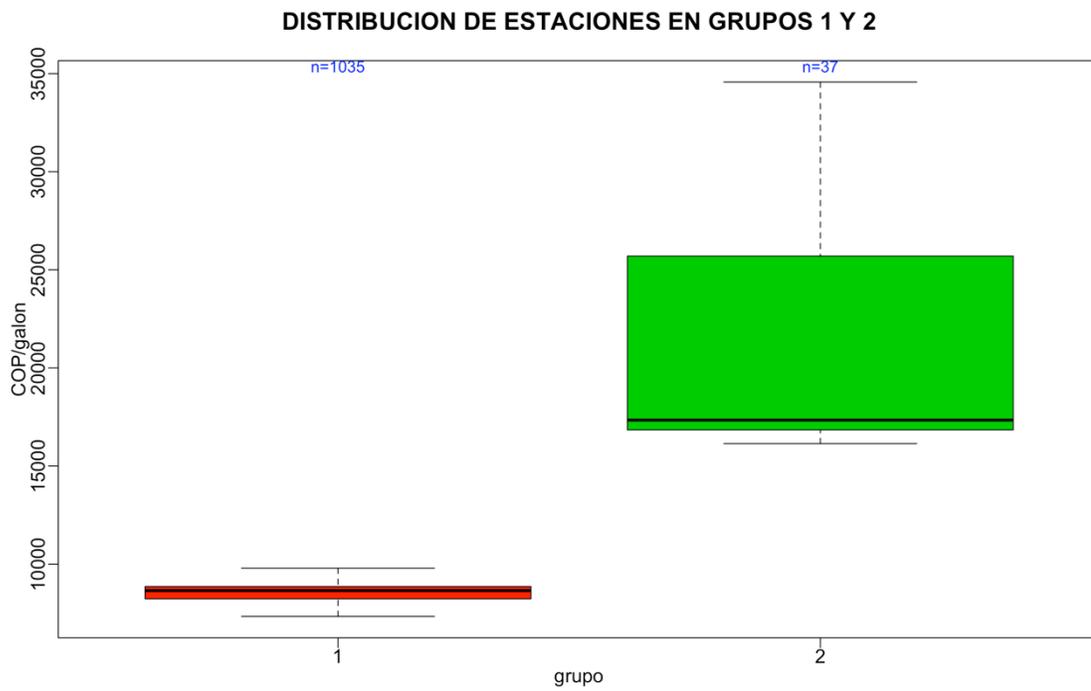


### 9.13.2 Selección mejor número de conglomerados método jerárquico



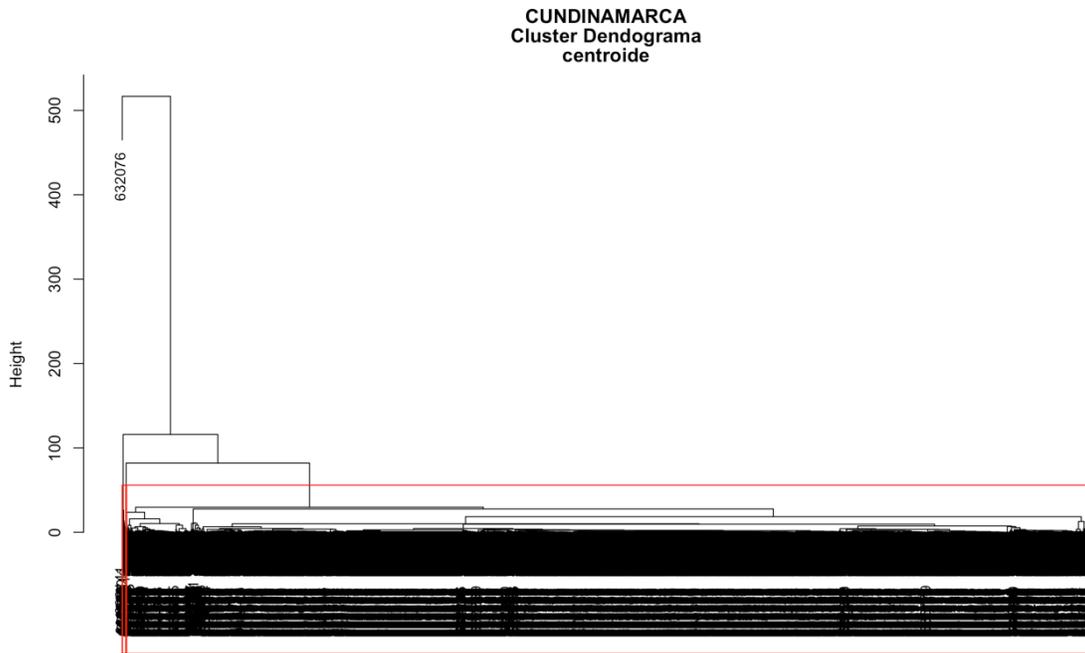


### 9.13.3 Grupos de estaciones y el comportamiento del precio

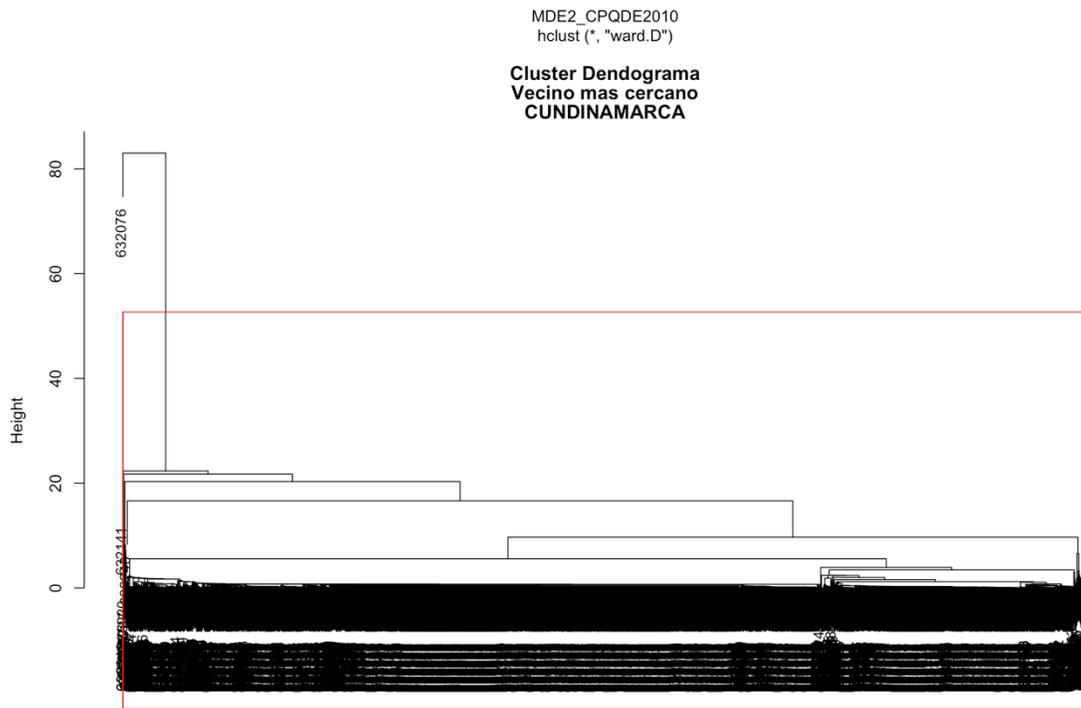
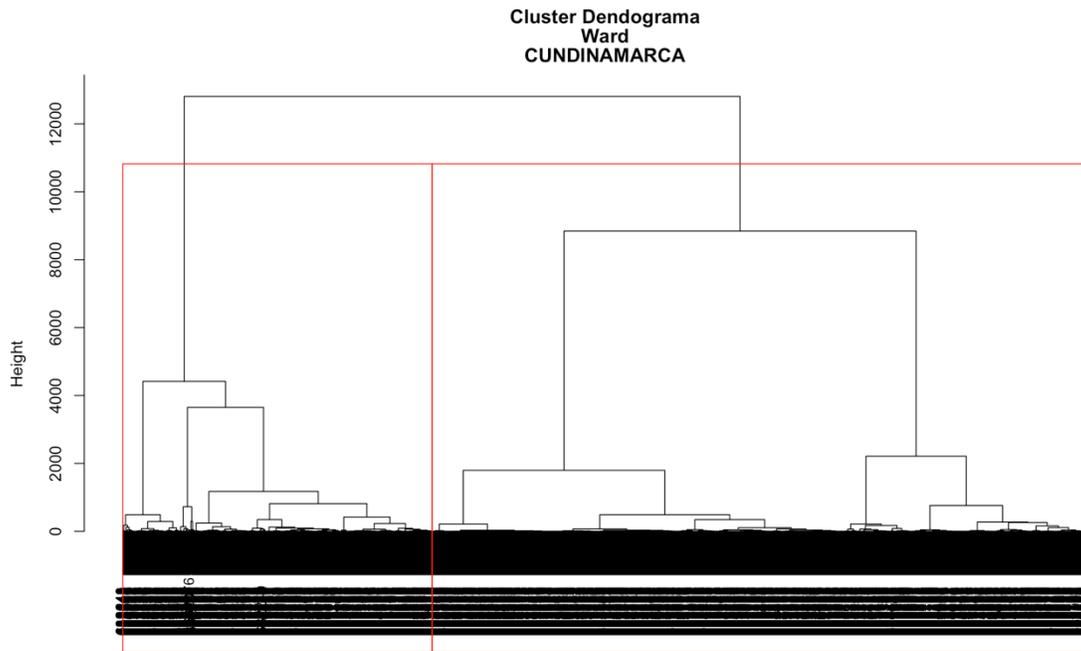


### 9.14 CUNDINAMARCA

### 9.14.1 Resultados métodos conglomerados

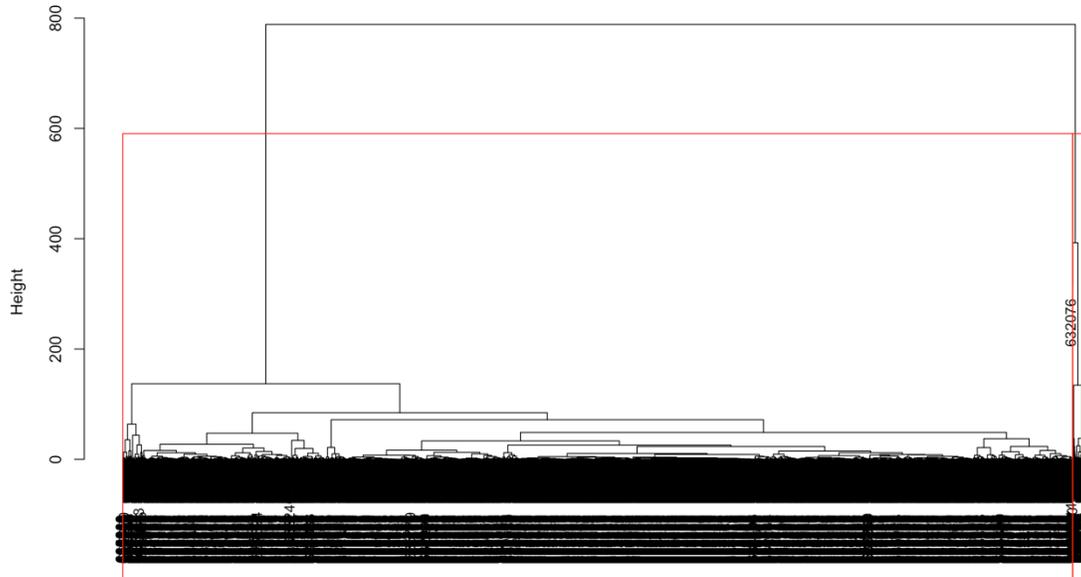


MDE2\_CPQDE2010  
hclust (\*, "centroid")



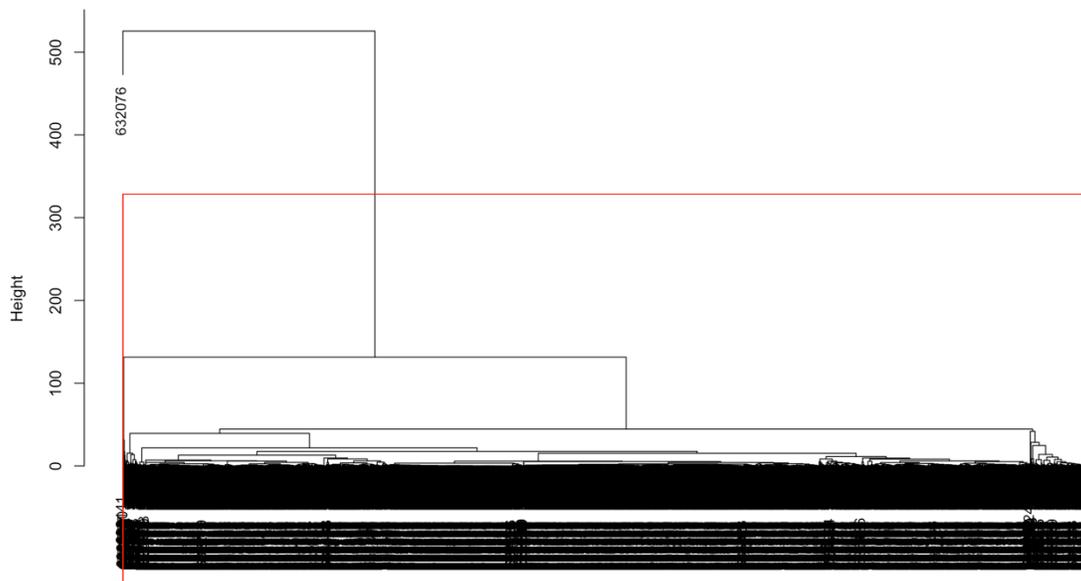
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
CUNDINAMARCA**



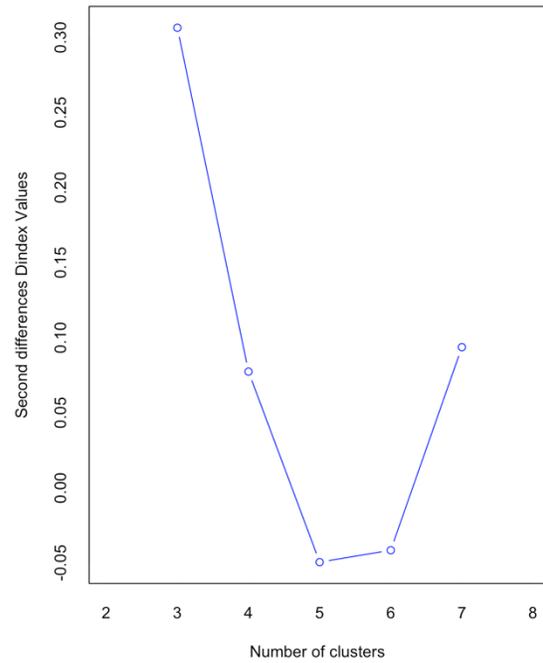
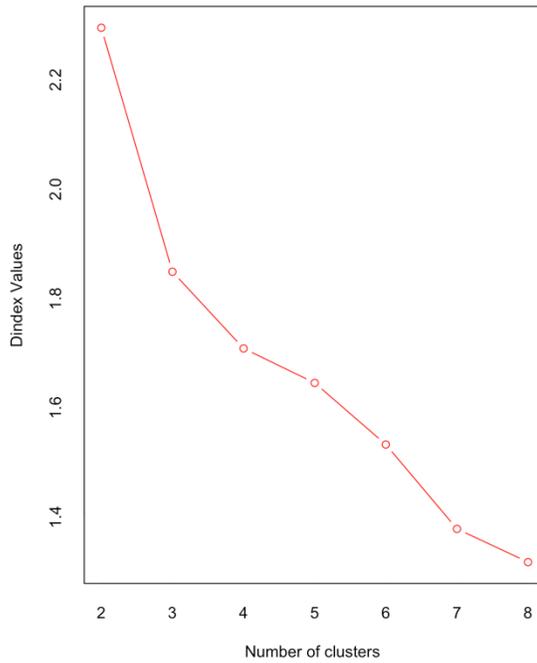
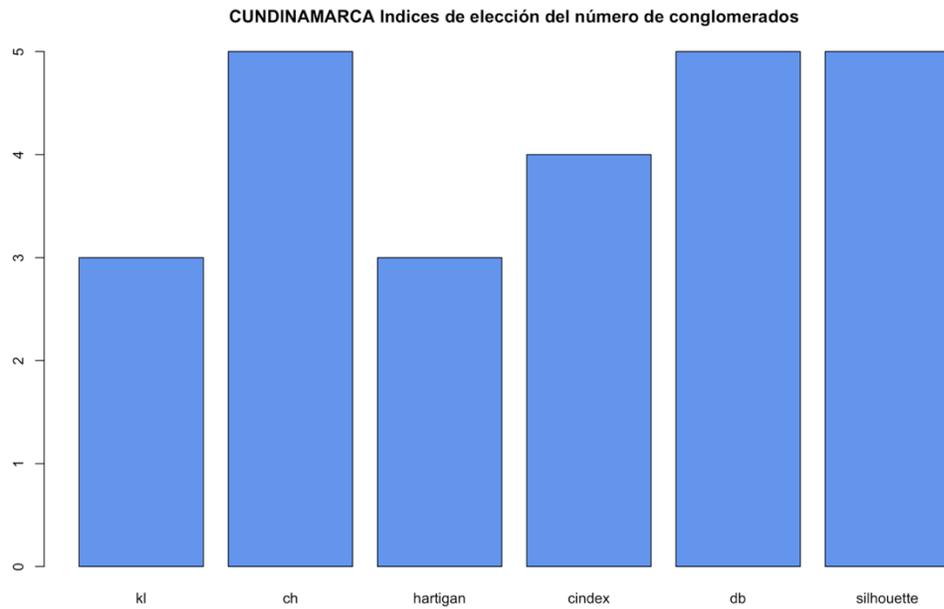
MDE2\_CPQDE2010  
hclust (\*, "complete")

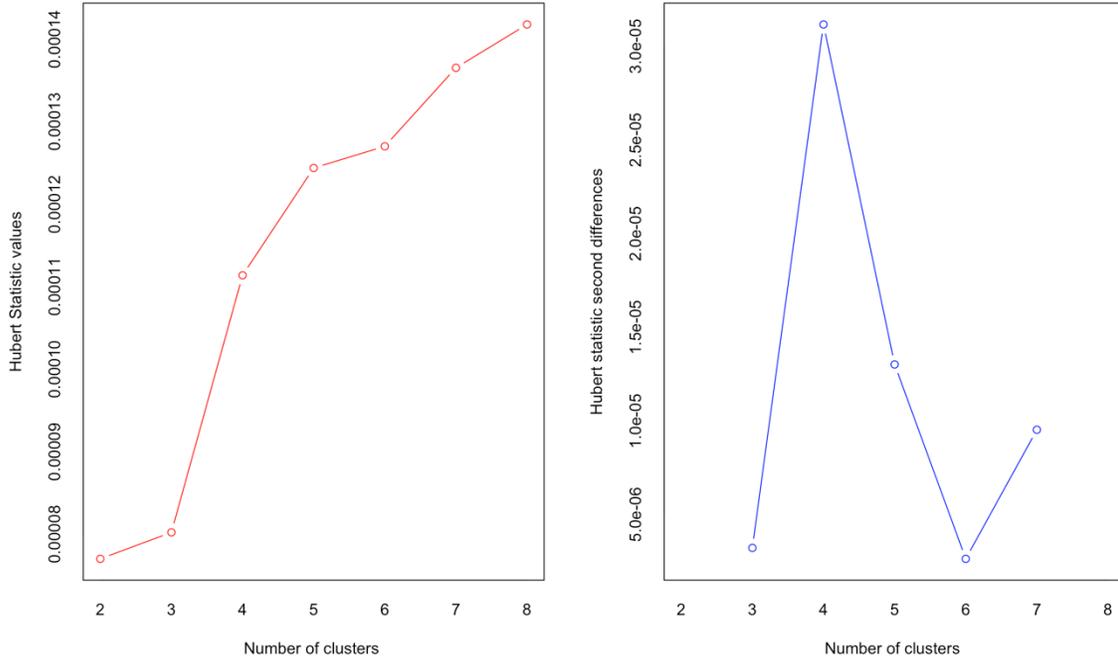
**Cluster Dendograma  
vinculacion promedio  
CUNDINAMARCA**



MDE2\_CPQDE2010  
hclust (\*, "average")

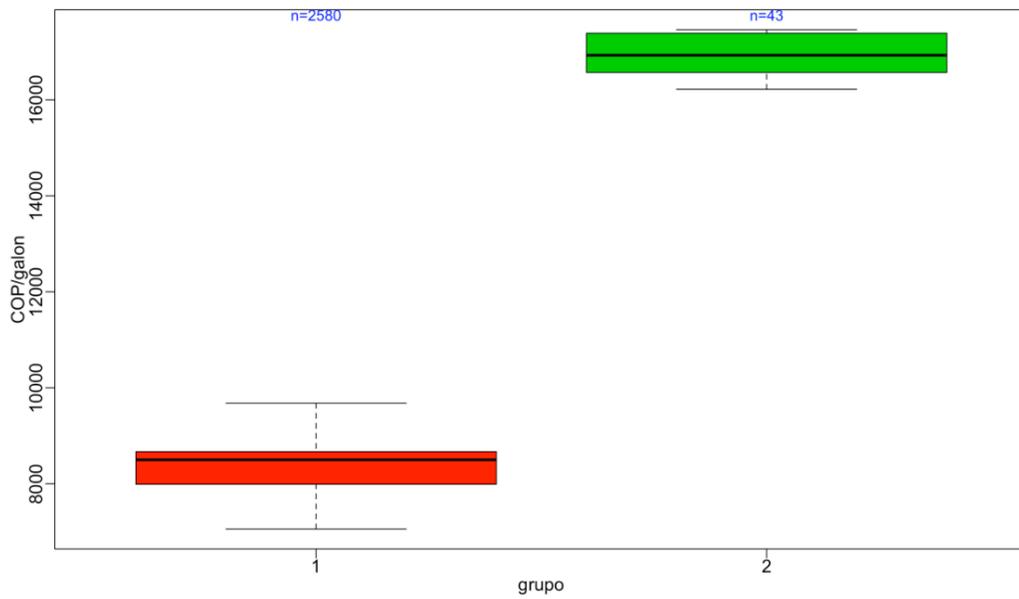
### 9.14.2 Selección mejor número de conglomerados método jerárquico





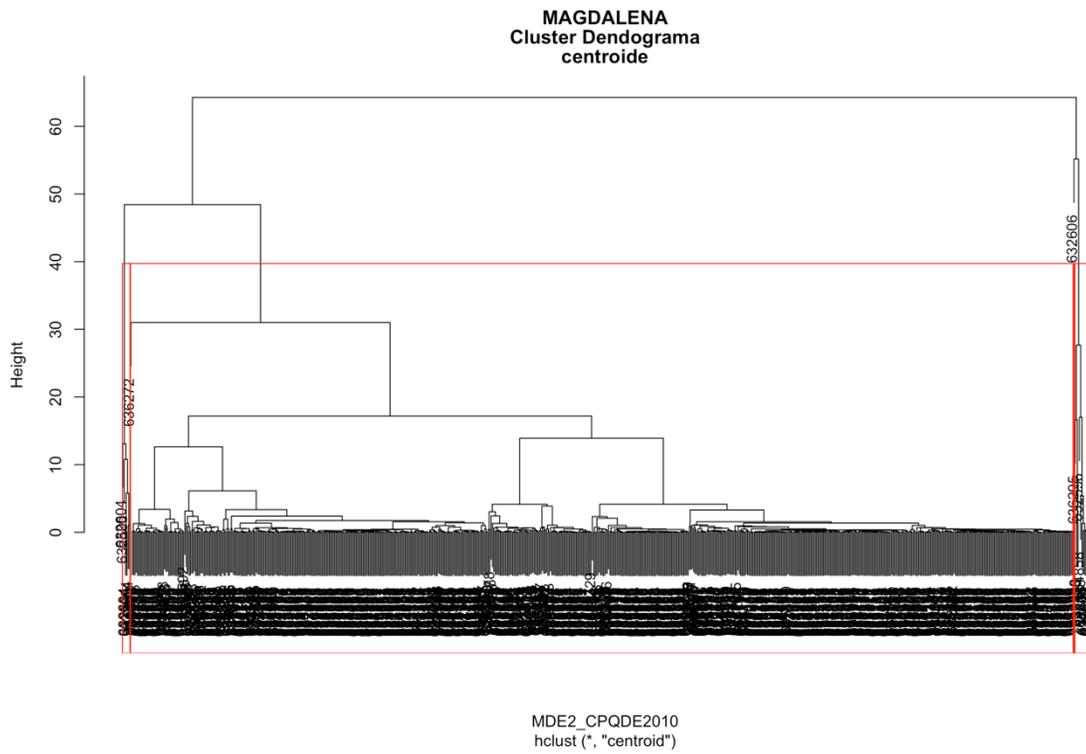
### 9.14.3 Grupos de estaciones y el comportamiento del precio

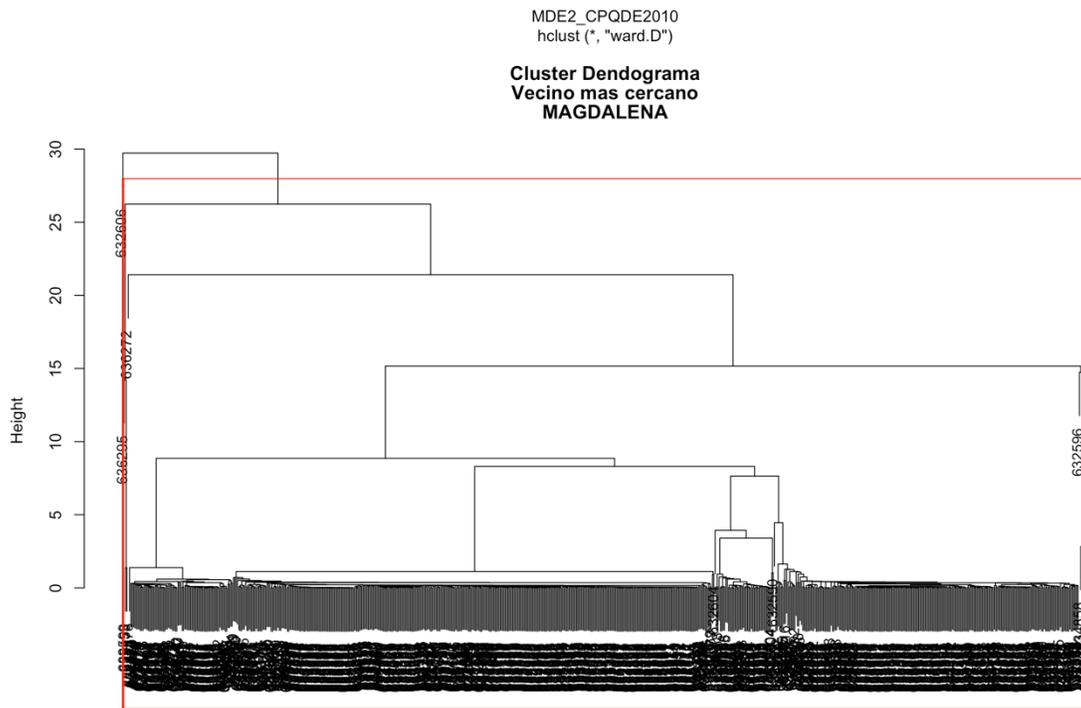
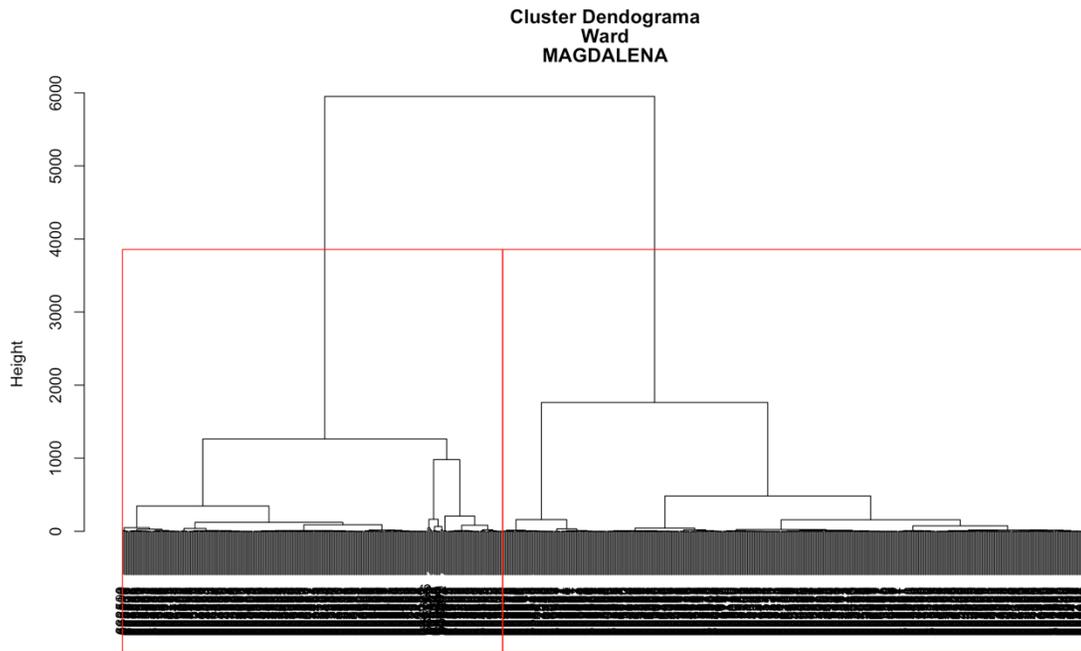
DISTRIBUCION DE ESTACIONES EN GRUPOS 1 Y 2



## 9.15 MAGDALENA

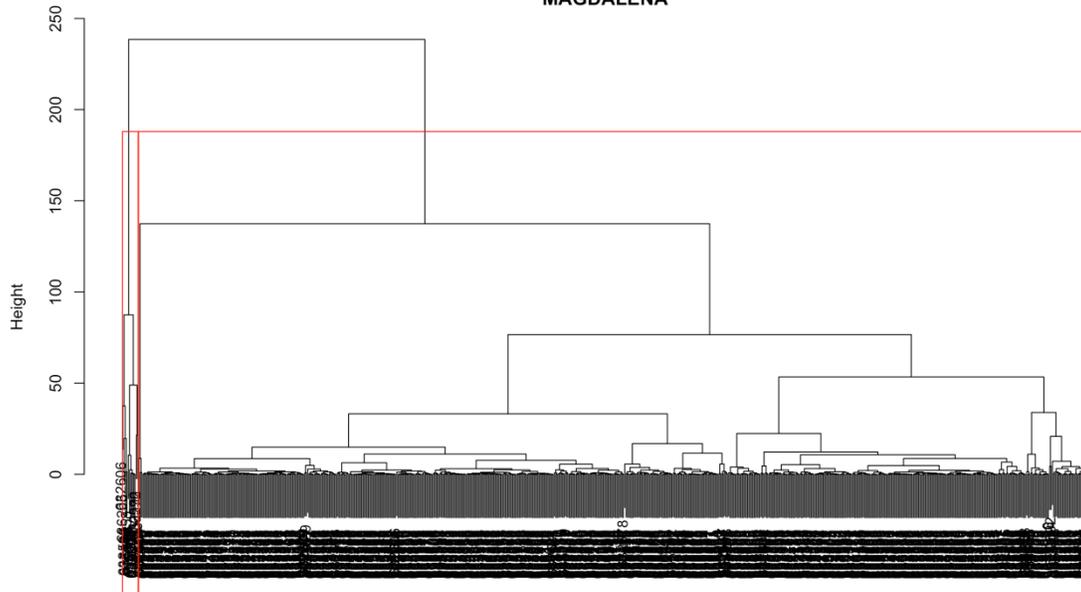
### 9.15.1 Resultados métodos conglomerados





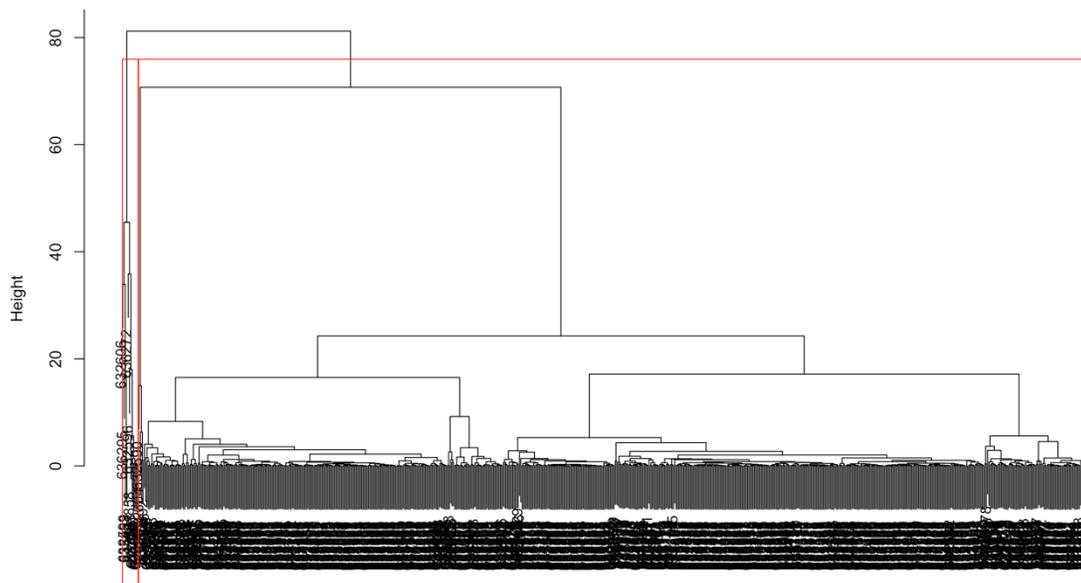
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendrograma  
Vecino mas lejano  
MAGDALENA**



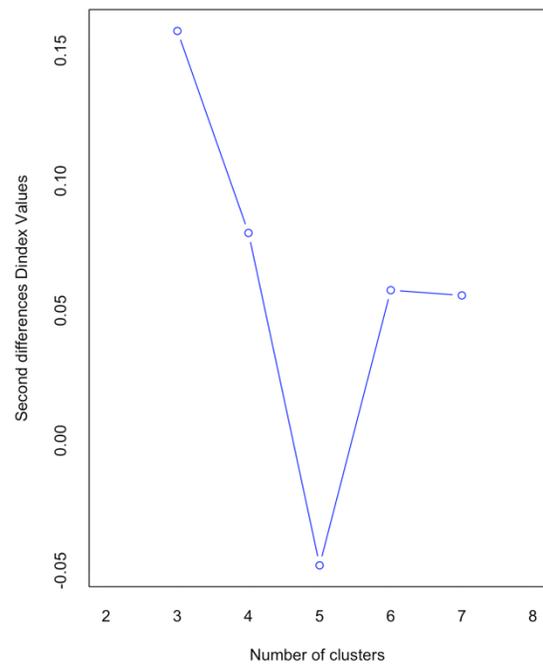
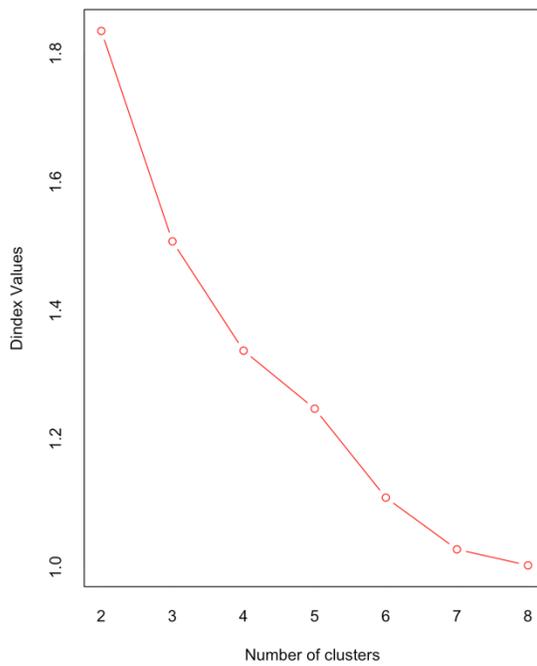
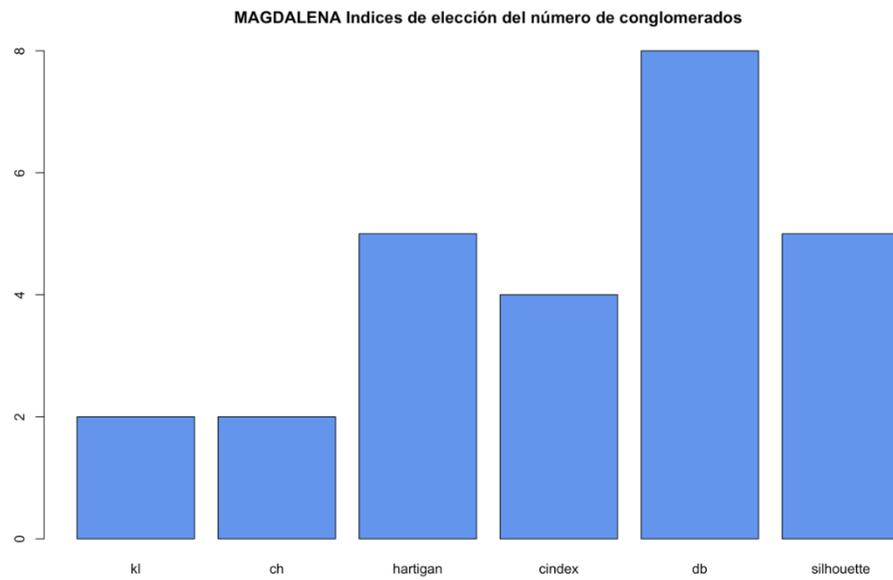
MDE2\_CPQDE2010  
hclust (\*, "complete")

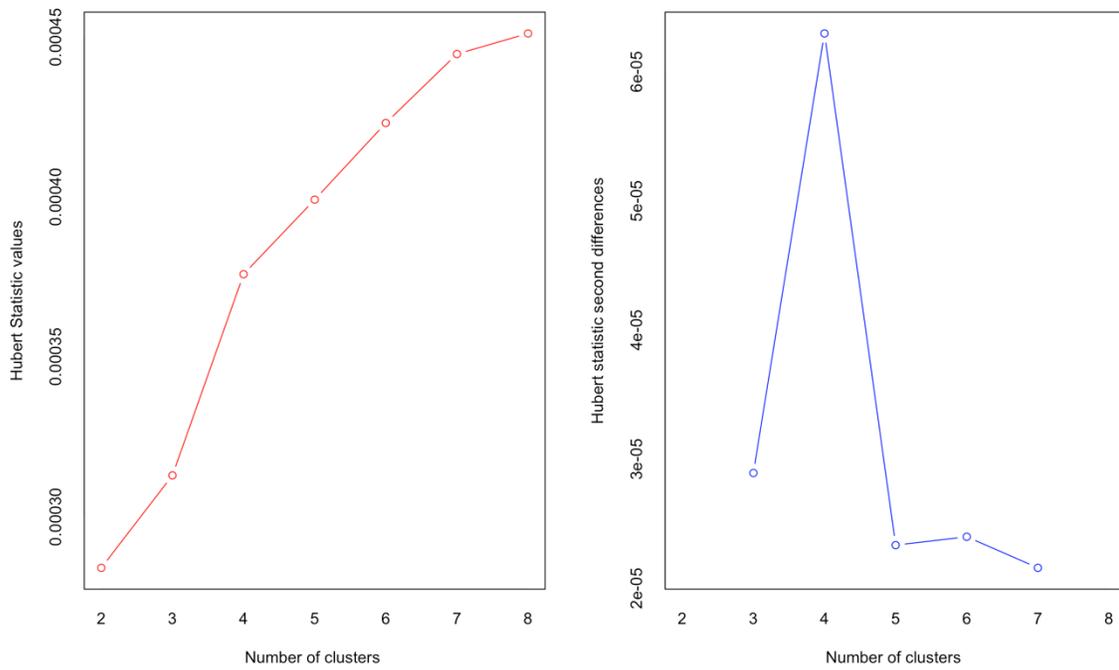
**Cluster Dendrograma  
vinculacion promedio  
MAGDALENA**



MDE2\_CPQDE2010  
hclust (\*, "average")

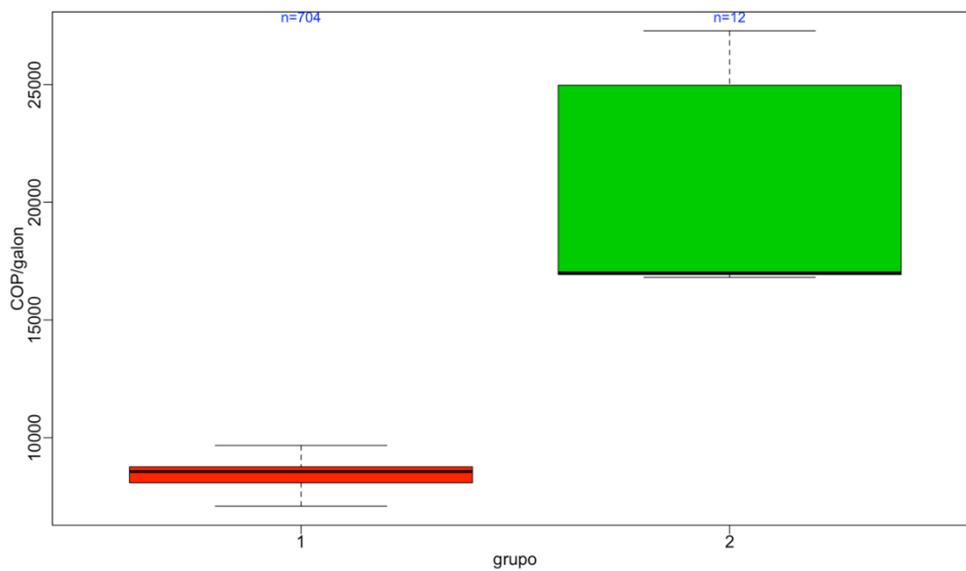
### 9.15.2 Selección mejor número de conglomerados método jerárquico





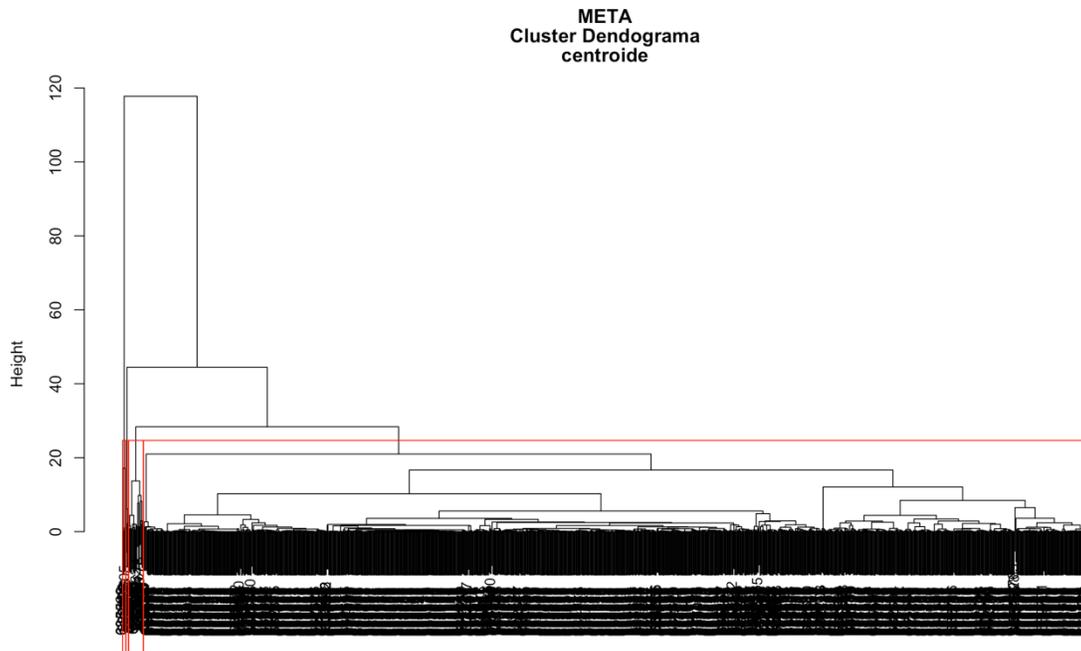
### 9.15.3 Grupos de estaciones y el comportamiento del precio

DISTRIBUCION DE ESTACIONES EN GRUPOS 1 Y 2

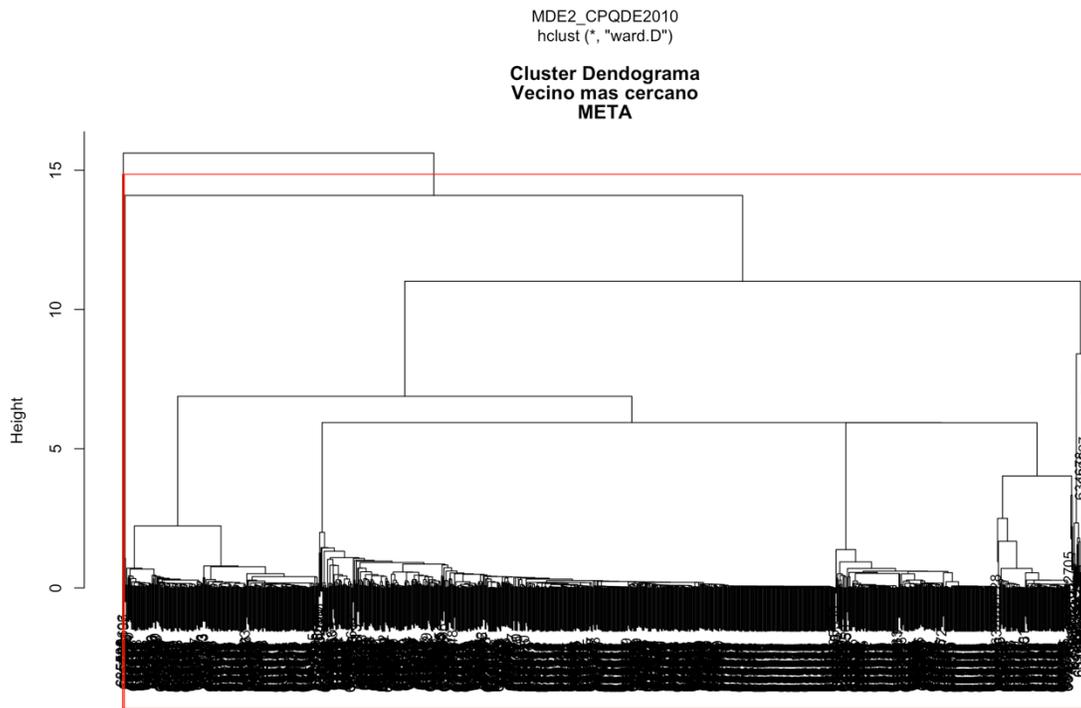
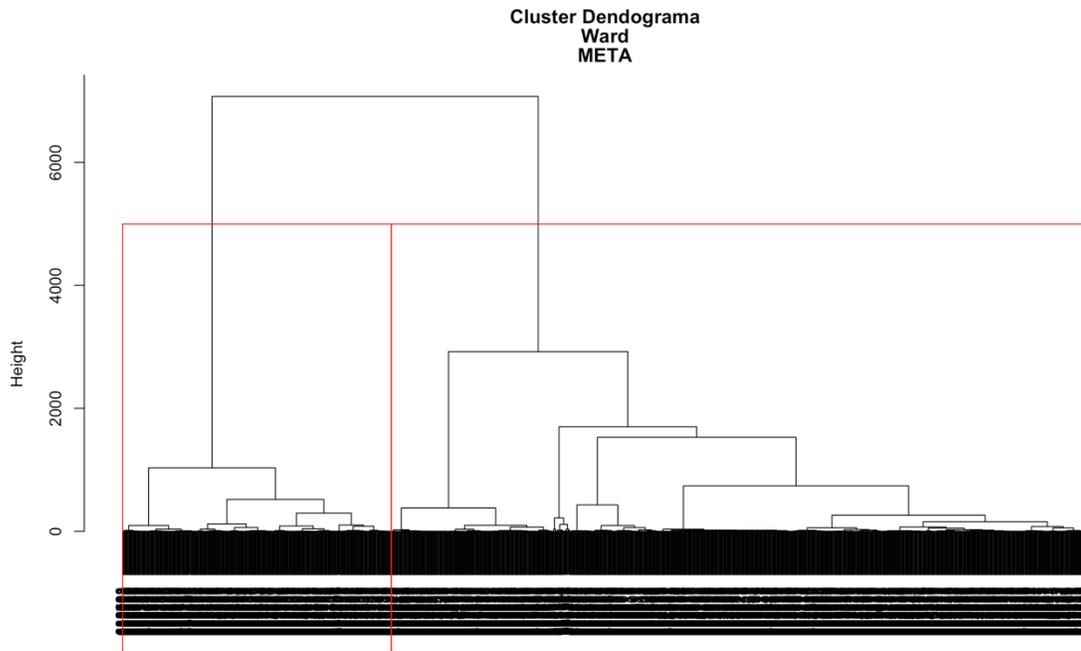


## 9.16 META

### 9.16.1 Resultados métodos conglomerados

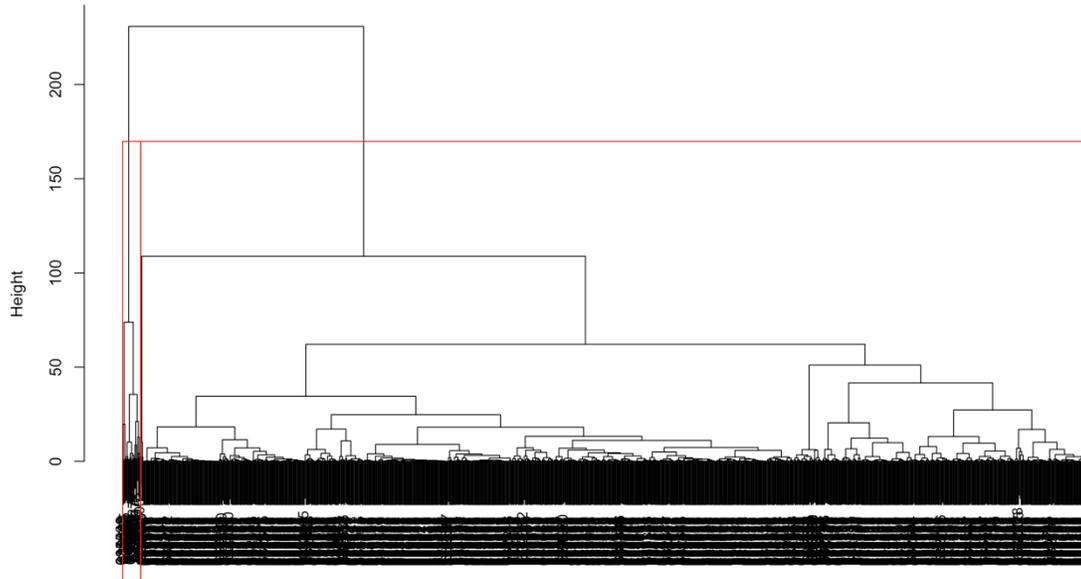


MDE2\_CPQDE2010  
hclust (\*, "centroid")



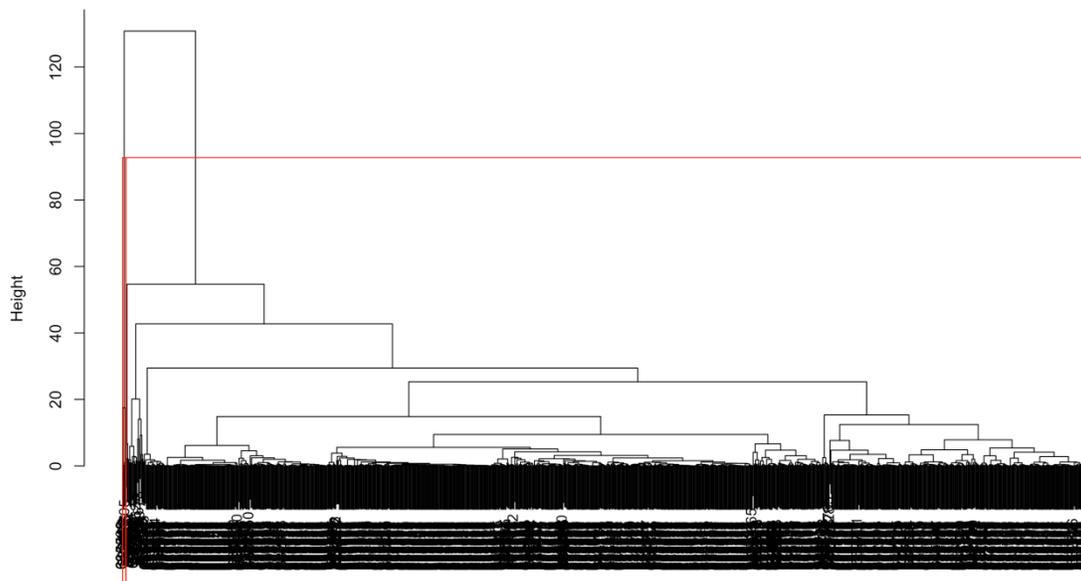
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
META**



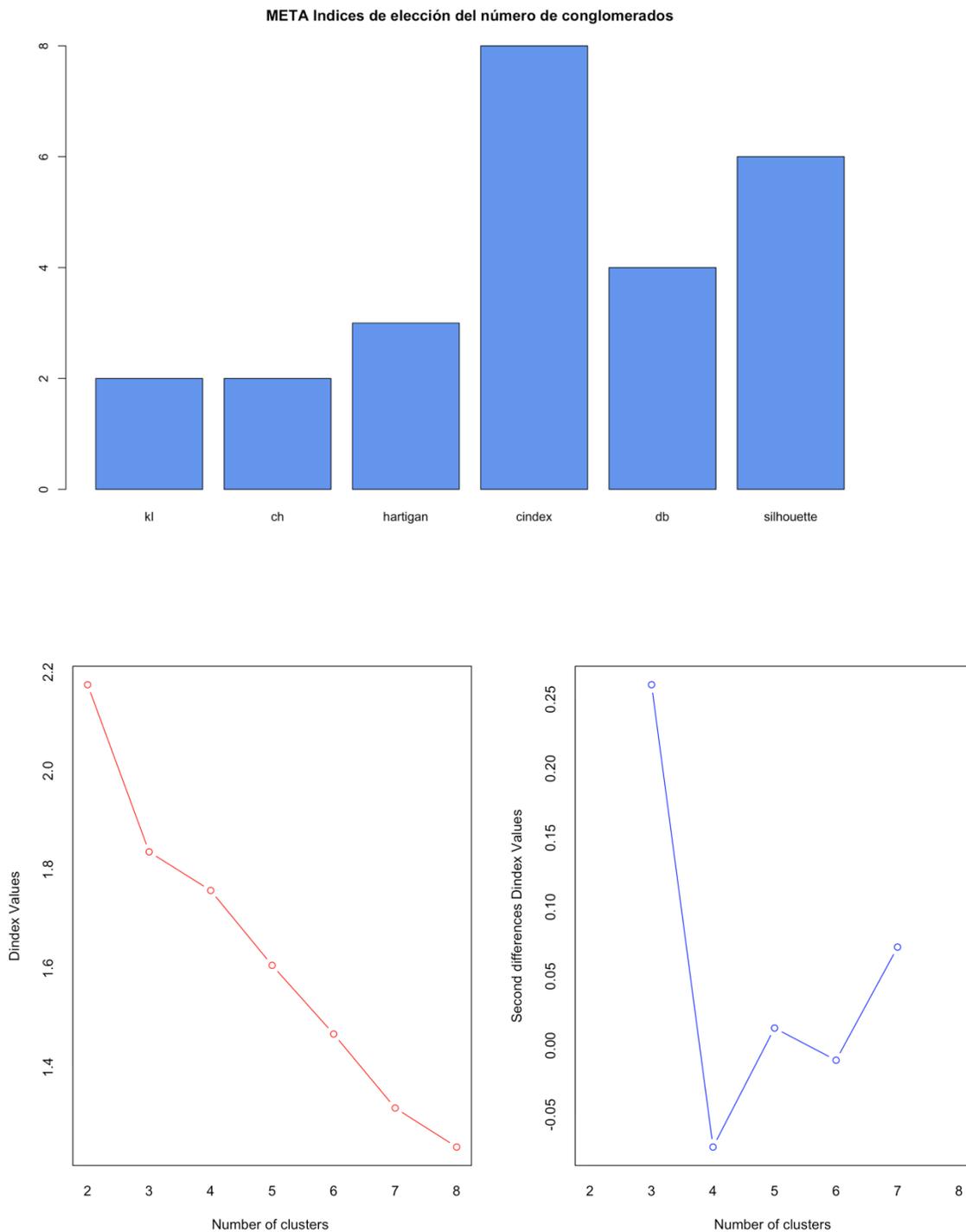
MDE2\_CPQDE2010  
hclust (\*, "complete")

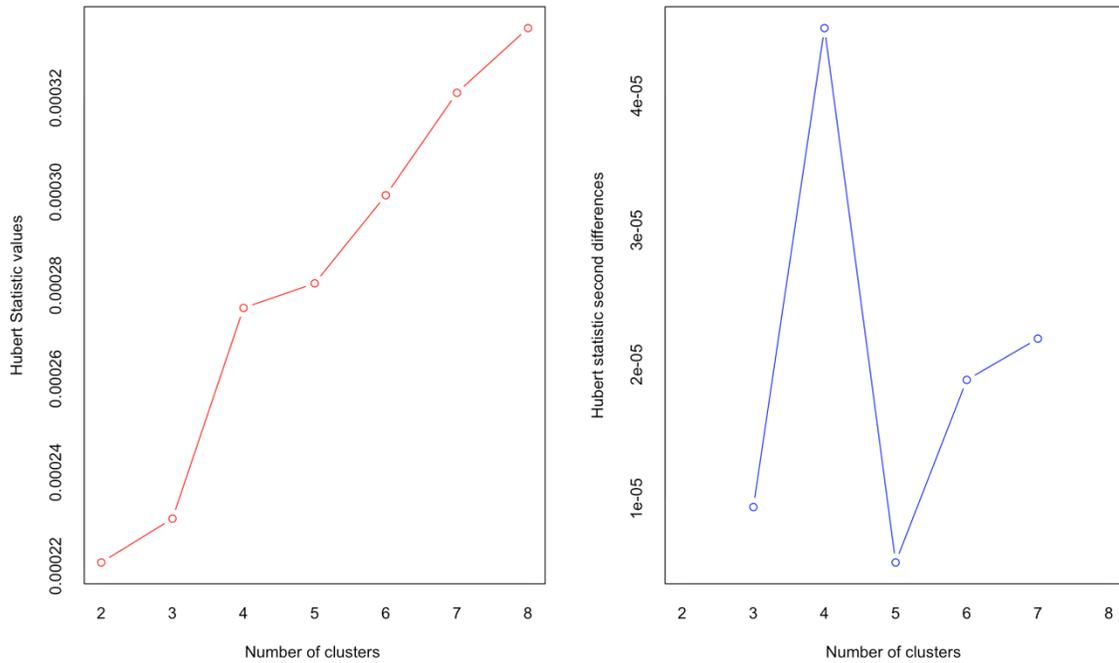
**Cluster Dendograma  
vinculacion promedio  
META**



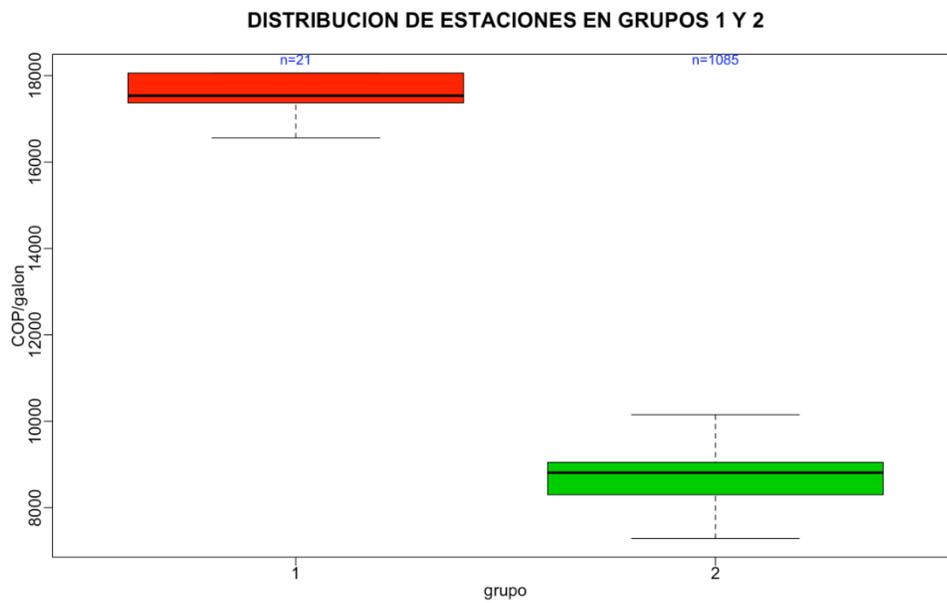
MDE2\_CPQDE2010  
hclust (\*, "average")

## 9.16.2 Selección mejor número de conglomerados método jerárquico



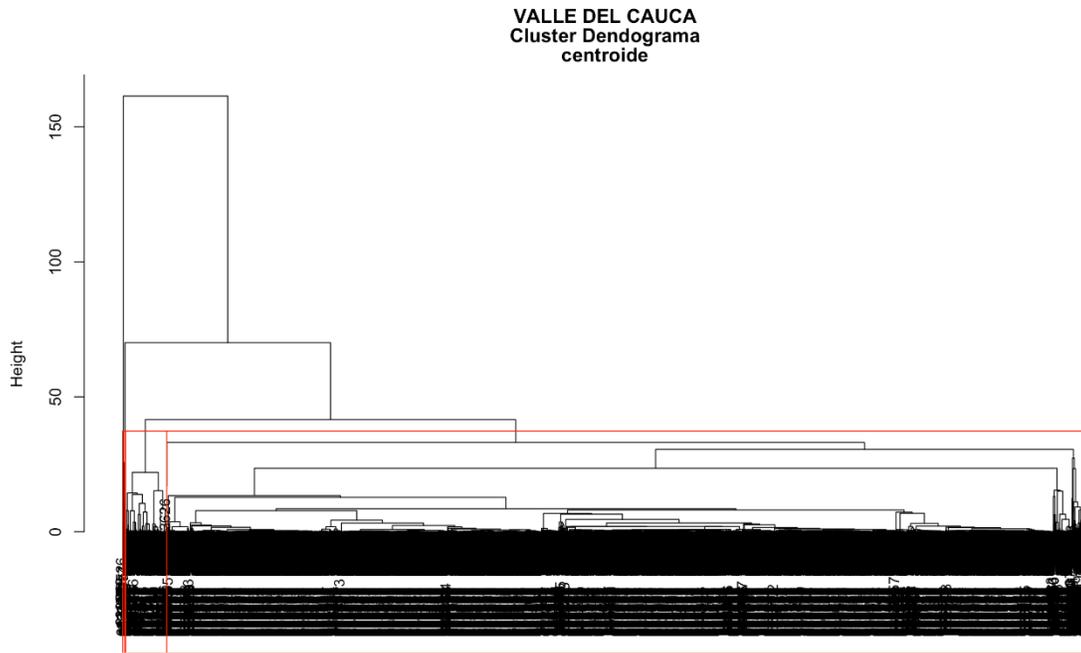


### 9.16.3 Grupos de estaciones y el comportamiento del precio



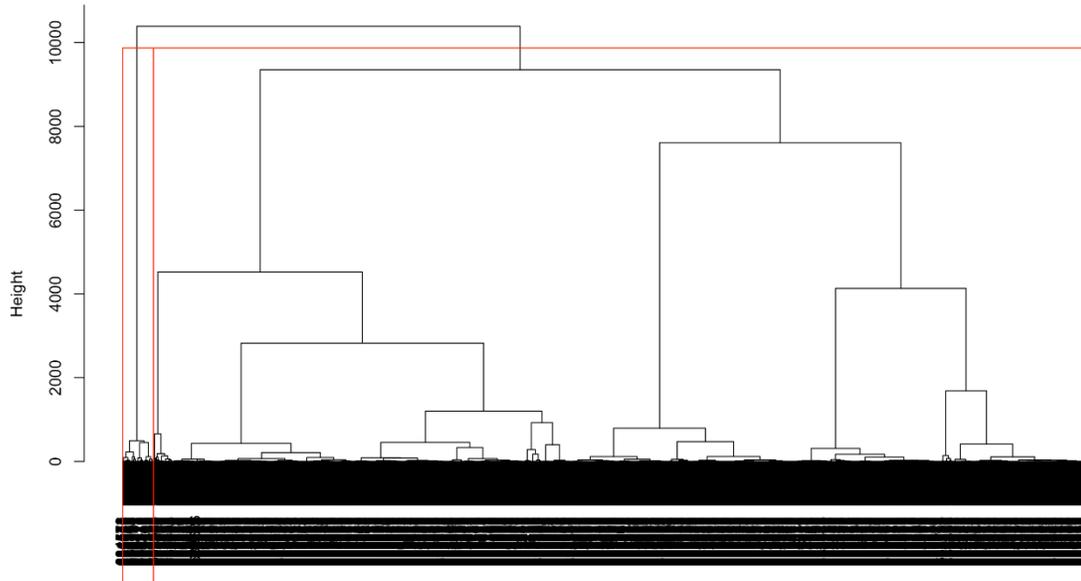
## 9.17 VALLE DEL CAUCA

### 9.17.1 Resultados métodos conglomerados



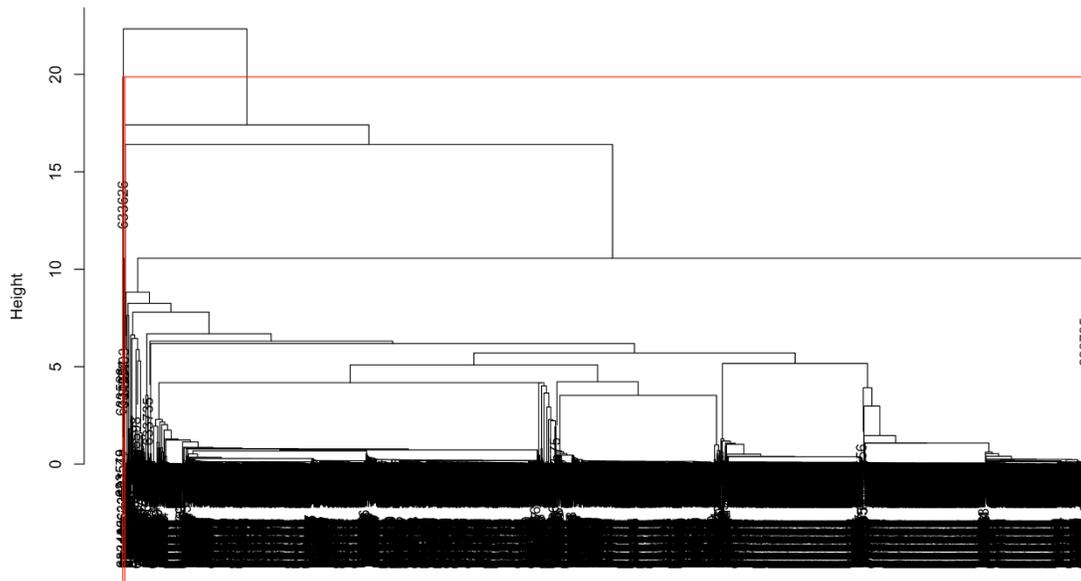
MDE2\_CPQDE2010  
hclust (\*, "centroid")

Cluster Dendrograma  
Ward  
VALLE DEL CAUCA



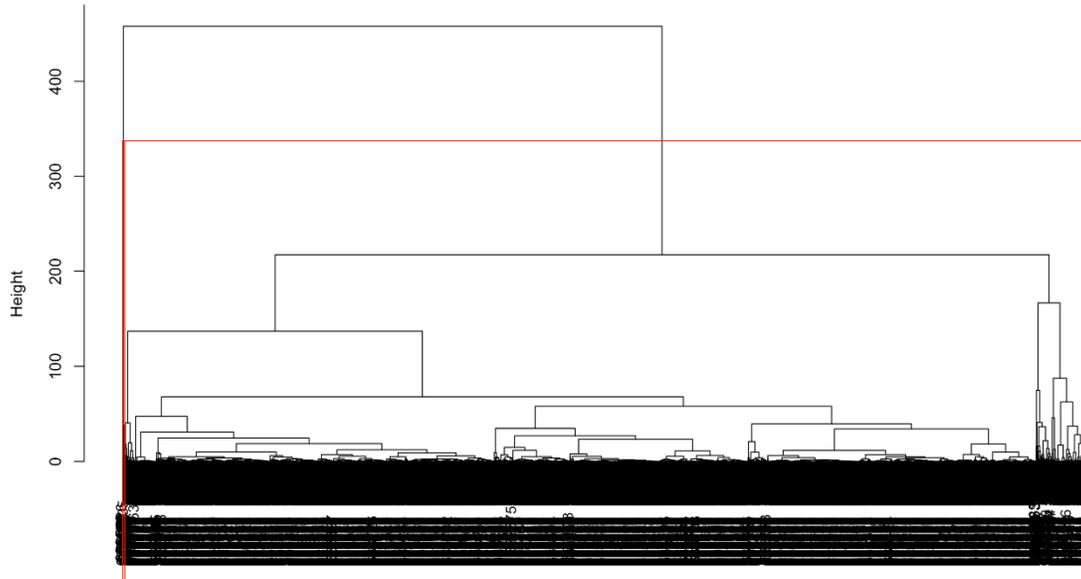
MDE2\_CPQDE2010  
hclust (\*, "ward.D")

Cluster Dendrograma  
Vecino mas cercano  
VALLE DEL CAUCA



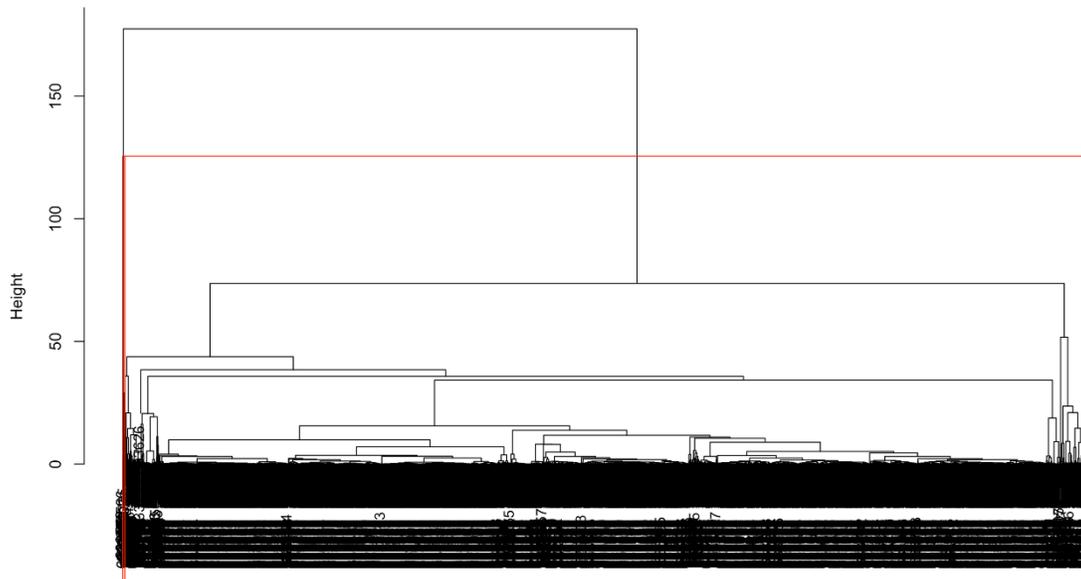
MDE2\_CPQDE2010  
hclust (\*, "single")

**Cluster Dendograma  
Vecino mas lejano  
VALLE DEL CAUCA**



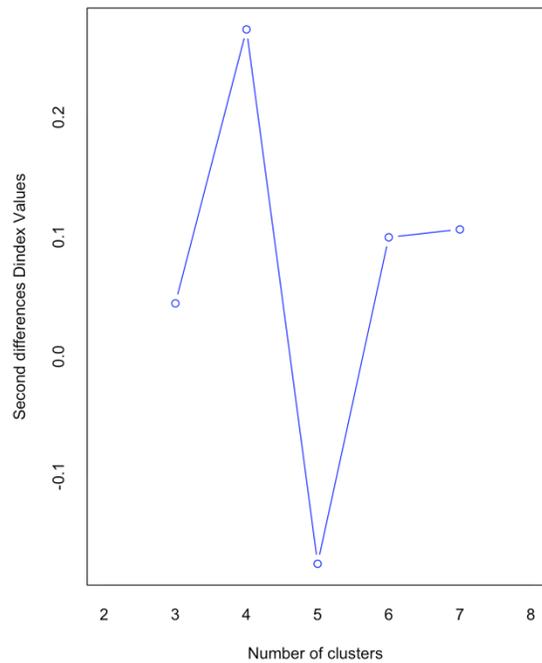
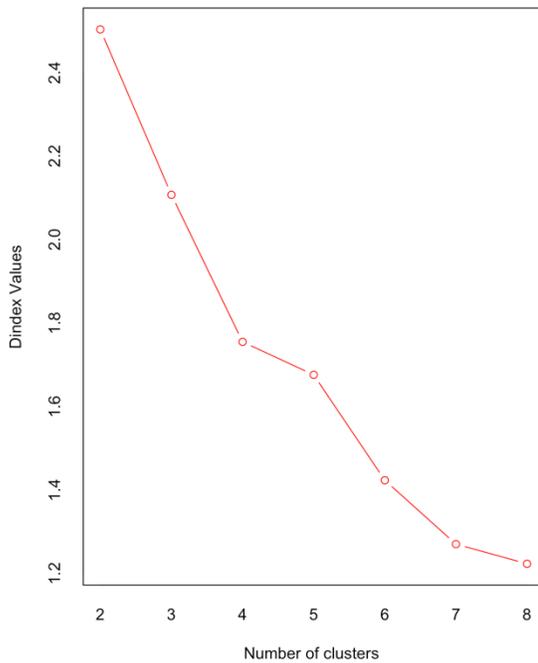
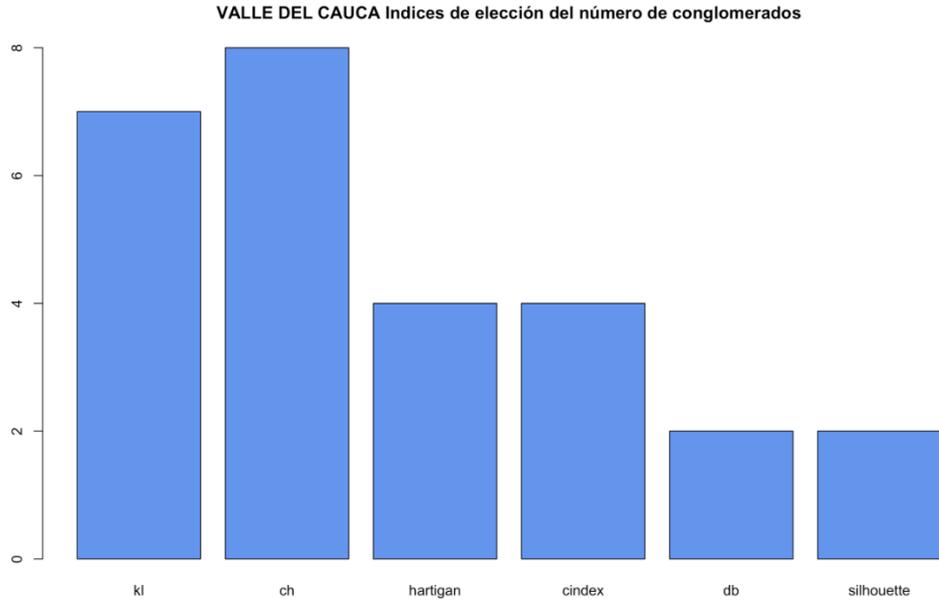
MDE2\_CPQDE2010  
hclust (\*, "complete")

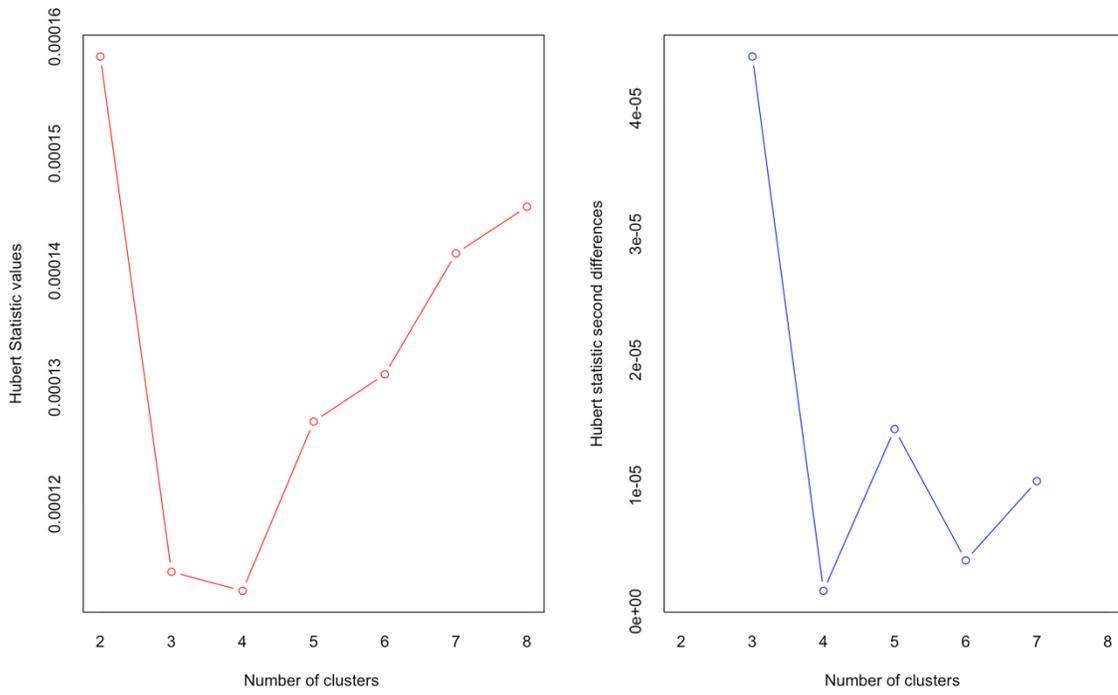
**Cluster Dendograma  
vinculacion promedio  
VALLE DEL CAUCA**



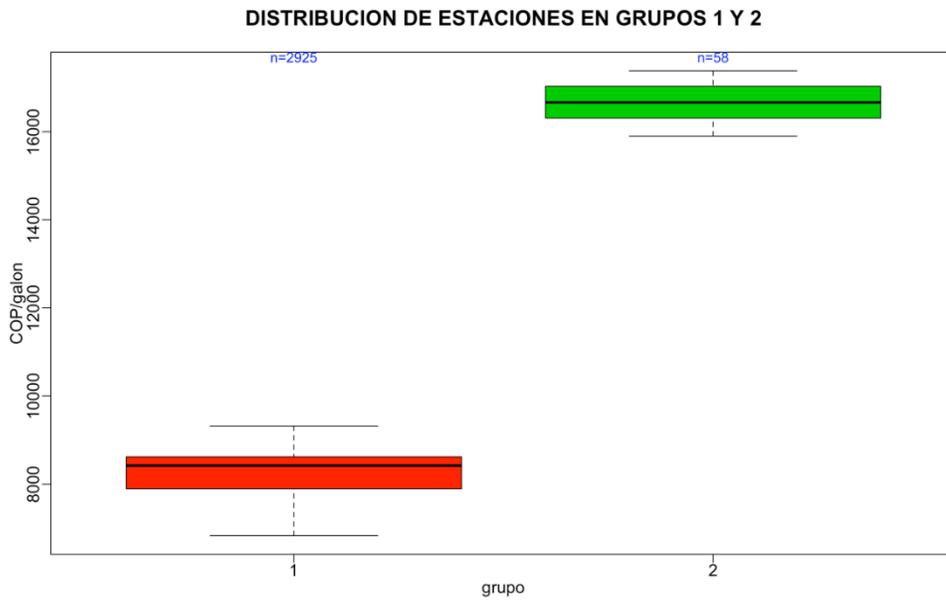
MDE2\_CPQDE2010  
hclust (\*, "average")

### 9.17.2 Selección mejor número de conglomerados método Grupos de estaciones y el comportamiento del precio





### 9.17.3 Grupos de estaciones y el comportamiento del precio



## 10 Anexo 4 Códigos R+ y matlab

Este anexo se incluye únicamente en formato digital.

## 10.1 Generación coordenadas estaciones

```

# R Tomar direcciones desde R

geocodeAddress <- function(address) {
  require(RJSONIO)
  url <- "https://maps.google.com/maps/api/geocode/json?address="
  # url <- URLEncode(paste(url, address, "&sensor=false", sep = ""))
  # url1 <- URLEncode(paste(url1, address, "&key=AlzaSyA6x7jCrXF-
8mYr4PKfHdVglaM27wDz4Eo"))
  url <- URLEncode(paste(url,
address, "&key=AlzaSyDF6kiXknTK236BnOoIKBCEo98cFzlux7U"))
  x <- fromJSON(url)#, simplify = FALSE)
  x <- fromJSON(url, simplify = FALSE)
  if (x$status == "OK") {
    out <- c(x$results[[1]]$geometry$location$lng,
            x$results[[1]]$geometry$location$lat)
  } else {
    out <- NA
  }
  Sys.sleep(0.2) # API only allows 5 requests per second
  out
}
#browser()

##### BOGOTA
library(readxl)
BOG_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx", sheet =
"AGENTES5_bogota") #, range = "a850:l1320")
View(BOG_GM)

#IMPORTANDO LONGITUD Y LATITUD
attach(BOG_GM)
direcciones=DIRECCION_AJUSTADA
rm(geocoded)
rm(BOG_GM1)
#geocoded <-0
geocoded <- data.frame()
iiii=length(DIRECCION_AJUSTADA)
for (ii in seq(1, length(DIRECCION))){ #seq(1, 100)}
  #print(paste("Working on index", ii, "of", 100))
  #query the google geocoder - this will pause here if we are over the limit.
  direccio=direcciones[ii]
  result = geocodeAddress(direcciones[ii])
  #print(result$status)

```

```

#result$index <- ii
#append the answer to the results file.
geocoded <- rbind(geocoded, result)
#save temporary results as we are going along
#saveRDS(geocoded, tempfilename)
}
#View(geocoded)
rm(lon_geocoded)
rm(lat_geocoded)
lon_geocoded=geocoded[1:length(DIRECCION),1]
lat_geocoded=geocoded[1:length(DIRECCION),2]

#PINTANDO ESTACIONES bogota de geocode
center = c(mean(lat_geocoded,na.rm=TRUE), mean(lon_geocoded,na.rm=TRUE));
zoom      <-      min(MaxZoom(range(lat_geocoded,na.rm=TRUE),
range(lon_geocoded,na.rm=TRUE)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap      <-      GetMap(center=center,      zoom=zoom,markers      =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "MyTile1.png");
tmp_geocoded<- PlotOnStaticMap(MyMap, lat = lat_geocoded, lon = lon_geocoded,
destfile = "geocode.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
BOG_GM1=data.frame(BO=BOG_GM,lat_g = lat_geocoded,lon_g = lon_geocoded)
#Guardando mapa
dev.copy(png,'GEO_BOGOTA.png')
dev.off()
#View(BOG_GM1)

##-----

##### medellin
library(readxl)
ME_GM      <-      read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet      =
"AGENTES5_medellin") #, range = "a850:l1320")
#View(ME_GM)

#IMPORTANDO LONGITUD Y LATITUD
attach(ME_GM)
direcciones=DIRECCION_AJUSTADA
rm(geocoded)
rm(ME_GM1)
#geocoded <-0
geocoded <- data.frame()
iiii=length(DIRECCION_AJUSTADA)
for (ii in seq(1, length(DIRECCION))){ #seq(1, 100)}
  #print(paste("Working on index", ii, "of", 100))

```

```

#query the google geocoder - this will pause here if we are over the limit.
direccio=direcciones[ii]
result = geocodeAddress(direcciones[ii])
#print(result$status)
#result$index <- ii
#append the answer to the results file.
geocoded <- rbind(geocoded, result)
#save temporary results as we are going along
#saveRDS(geocoded, tempfilename)
}
#View(geocoded)
rm(lon_geocoded)
rm(lat_geocoded)
lon_geocoded=geocoded[1:length(DIRECCION),1]
lat_geocoded=geocoded[1:length(DIRECCION),2]

#PINTANDO ESTACIONES medellin de geocode
center = c(mean(lat_geocoded,na.rm=TRUE), mean(lon_geocoded,na.rm=TRUE));
zoom      <-      min(MaxZoom(range(lat_geocoded,na.rm=TRUE),
range(lon_geocoded,na.rm=TRUE)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap      <-      GetMap(center=center,      zoom=zoom,markers      =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "MyTile1.png");
tmp_geocoded<- PlotOnStaticMap(MyMap, lat = lat_geocoded, lon = lon_geocoded,
destfile = "geocode.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
ME_GM1=data.frame(ME=ME_GM,lat_g = lat_geocoded,lon_g = lon_geocoded)
#Guardando mapa
dev.copy(png,'GEO_medellin.png')
dev.off()

#View(ME_GM1)
##-----
##### cali
library(readxl)
CA_GM      <-      read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet      =
"AGENTES5_cali") #, range = "a850:l1320")
#View(CA_GM)

#IMPORTANDO LONGITUD Y LATITUD
attach(CA_GM)
direcciones=DIRECCION_AJUSTADA
rm(geocoded)
rm(CA_GM1)
#geocoded <-0
geocoded <- data.frame()

```

```

iiii=length(DIRECCION_AJUSTADA)
for (ii in seq(1, length(DIRECCION))){ #seq(1, 100){
  #print(paste("Working on index", ii, "of", 100))
  #query the google geocoder - this will pause here if we are over the limit.
  direccio=direcciones[ii]
  result = geocodeAddress(direcciones[ii])
  #print(result$status)
  #result$index <- ii
  #append the answer to the results file.
  geocoded <- rbind(geocoded, result)
  #save temporary results as we are going along
  #saveRDS(geocoded, tempfilename)
}
#View(geocoded)
rm(lon_geocoded)
rm(lat_geocoded)
lon_geocoded=geocoded[1:length(DIRECCION),1]
lat_geocoded=geocoded[1:length(DIRECCION),2]

#PINTANDO ESTACIONES cali de geocode
center = c(mean(lat_geocoded,na.rm=TRUE), mean(lon_geocoded,na.rm=TRUE));
zoom <- min(MaxZoom(range(lat_geocoded,na.rm=TRUE),
range(lon_geocoded,na.rm=TRUE)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "MyTile1.png");
tmp_geocoded<- PlotOnStaticMap(MyMap, lat = lat_geocoded, lon = lon_geocoded,
destfile = "geocode.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
CA_GM1=data.frame(CA=CA_GM,lat_g = lat_geocoded,lon_g = lon_geocoded)
#Guardando mapa
dev.copy(png,'GEO_cali.png')
dev.off()
#View(CA_GM1)

##-----
##### barranquilla
library(readxl)
BA_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet =
"AGENTES5_barranquilla") #, range = "a850:l1320")
#View(BA_GM)

#IMPORTANDO LONGITUD Y LATITUD
attach(BA_GM)
direcciones=DIRECCION_AJUSTADA
rm(geocoded)

```

```

rm(BA_GM1)
#geocoded <-0
geocoded <- data.frame()
iiii=length(DIRECCION_AJUSTADA)
for (ii in seq(1, length(DIRECCION))){ #seq(1, 100)}
  #print(paste("Working on index", ii, "of", 100))
  #query the google geocoder - this will pause here if we are over the limit.
  direccio=direcciones[ii]
  result = geocodeAddress(direcciones[ii])
  #print(result$status)
  #result$index <- ii
  #append the answer to the results file.
  geocoded <- rbind(geocoded, result)
  #save temporary results as we are going along
  #saveRDS(geocoded, tempfilename)
}
#View(geocoded)
rm(lon_geocoded)
rm(lat_geocoded)
lon_geocoded=geocoded[1:length(DIRECCION),1]
lat_geocoded=geocoded[1:length(DIRECCION),2]

#PINTANDO ESTACIONES barranquilla de geocode
center = c(mean(lat_geocoded,na.rm=TRUE), mean(lon_geocoded,na.rm=TRUE));
zoom      <-      min(MaxZoom(range(lat_geocoded,na.rm=TRUE),
range(lon_geocoded,na.rm=TRUE)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap      <-      GetMap(center=center,      zoom=zoom,markers      =
paste0("&markers=color:blue|label:S|", "4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "MyTile1.png");
tmp_geocoded<- PlotOnStaticMap(MyMap, lat = lat_geocoded, lon = lon_geocoded,
destfile = "geocode.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
BA_GM1=data.frame(BA=BA_GM,lat_g = lat_geocoded,lon_g = lon_geocoded)
#Guardando mapa
dev.copy(png,'GEO_barranquilla.png')
dev.off()
#View(ME_GM1)

##### PAIS
library(readxl)
PAIS_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet = "AGENTES5")
#, range = "a850:l1320")
View(PAIS_GM)

#IMPORTANDO LONGITUD Y LATITUD
attach(PAIS_GM)

```

```

direcciones=DIRECCION_AJUSTADA
rm(geocoded)
rm(PAIS_GM1)
#geocoded <-0
geocoded <- data.frame()
iiii=length(DIRECCION_AJUSTADA)
for (ii in seq(1, length(DIRECCION))){ #seq(1, 100)}
  #print(paste("Working on index", ii, "of", 100))
  #query the google geocoder - this will pause here if we are over the limit.
  direccio=direcciones[ii]
  result = geocodeAddress(direcciones[ii])
  #print(result$status)
  #result$index <- ii
  #append the answer to the results file.
  geocoded <- rbind(geocoded, result)
  #save temporary results as we are going along
  #saveRDS(geocoded, tempfilename)
}
#View(geocoded)
rm(lon_geocoded)
rm(lat_geocoded)
lon_geocoded=geocoded[1:length(DIRECCION),1]
lat_geocoded=geocoded[1:length(DIRECCION),2]

#PINTANDO ESTACIONES bogota de geocode
center = c(mean(lat_geocoded,na.rm=TRUE), mean(lon_geocoded,na.rm=TRUE));
zoom      <-      min(MaxZoom(range(lat_geocoded,na.rm=TRUE),
range(lon_geocoded,na.rm=TRUE)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap      <-      GetMap(center=center,      zoom=zoom,markers      =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "MyTile1.png");
tmp_geocoded<- PlotOnStaticMap(MyMap, lat = lat_geocoded, lon = lon_geocoded,
destfile = "geocode.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
PAIS_GM1=data.frame(PAIS=PAIS_GM,lat_g = lat_geocoded,lon_g = lon_geocoded)
#Guardando mapa
dev.copy(png,'GEO_PAIS.png')
dev.off()
#View(BOG_GM1)

#guardando tablas
WriteXLS(list(BOG_GM1=BOG_GM1,
ME_GM1=ME_GM1,CA_GM1=CA_GM1,BA_GM1=BA_GM1,PAIS_GM1), "file.xlsx")

#####- inicio pintar estaciones

```

```
#ESTACIONES A NIVEL PAIS
library(readxl)
PaisGM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet = "AGENTES5") #,
range = "a850:l1320")
View(PaisGM)
#IMPORTANDO LONGITUD Y LATITUD
attach(PaisGM)
lat_PaisGM=latitud
lon_PaisGM=longitud

#PINTANDO ESTACIONES pais GM
center = c(mean(lat_PaisGM), mean(lon_PaisGM));
zoom <- min(MaxZoom(range(lat_PaisGM), range(lon_PaisGM)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "PaisGM.png");
tmp_PaisGM<- PlotOnStaticMap(MyMap, lat = lat_PaisGM, lon = lon_PaisGM, destfile =
"PaisGM1.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
#Guardando mapa
dev.copy(png,'E_Pais.png')
dev.off()
#Guardando tabla
WriteXLS(PaisGM,"resultado.xlsx",SheetNames = "PAIS", perl = "perl")

#ESTACIONES A NIVEL BOGOTA
library(readxl)
BOG_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet =
"AGENTES5_bogota") #, range = "a850:l1320")
View(BOG_GM)
#IMPORTANDO LONGITUD Y LATITUD
attach(BOG_GM)
lat_BOG_GM=latitud
lon_BOG_GM=longitud

#PINTANDO ESTACIONES BOGOTA GM
center = c(mean(lat_BOG_GM), mean(lon_BOG_GM));
zoom <- min(MaxZoom(range(lat_BOG_GM), range(lon_BOG_GM)));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "PaisGM.png");
tmp_PaisGM<- PlotOnStaticMap(MyMap, lat = lat_BOG_GM, lon = lon_BOG_GM, destfile
= "PaisGM1.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
```

```

#Guardando mapa
dev.copy(png,'E_BOGOTA.png')
dev.off()
#Guardando tabla
WriteXLS(BOG_GM,"resultado.xlsx",SheetNames = "E_BOGOTA", perl = "perl")

#ESTACIONES A NIVEL MEDELLIN GM
library(readxl)
ME_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet =
"AGENTES5_medellin") #, range = "a850:l1320")
View(ME_GM )
#IMPORTANDO LONGITUD Y LATITUD
attach(ME_GM )
lat_ME_GM=latitud
lon_ME_GM=longitud

#PINTANDO ESTACIONES MEDELLIN GM
center = c(mean(lat_ME_GM ), mean(lon_ME_GM ));
zoom <- min(MaxZoom(range(lat_ME_GM), range(lon_ME_GM )));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|",4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "PaisGM.png");
tmp_PaisGM<- PlotOnStaticMap(MyMap, lat = lat_ME_GM, lon = lon_ME_GM, destfile =
"PaisGM1.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
#Guardando mapa
dev.copy(png,'E_MEDELLIN.png')
dev.off()

#ESTACIONES A NIVEL CALI GM
library(readxl)
CA_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet =
"AGENTES5_cali") #, range = "a850:l1320")
View(CA_GM )
#IMPORTANDO LONGITUD Y LATITUD
attach(CA_GM )
lat_CA_GM=latitud
lon_CA_GM=longitud

#PINTANDO ESTACIONES CALI GM
center = c(mean(lat_CA_GM ), mean(lon_CA_GM ));
zoom <- min(MaxZoom(range(lat_CA_GM), range(lon_CA_GM )));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|",4.702147,-

```

```

74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "PaisGM.png");
  tmp_PaisGM<- PlotOnStaticMap(MyMap, lat = lat_CA_GM, lon = lon_CA_GM, destfile =
"PaisGM1.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
  #Guardando mapa
  dev.copy(png,'E_CALI.png')
  dev.off()

#ESTACIONES A NIVEL BARRANQUILLA GM
library(readxl)
BA_GM <- read_excel("2018.05.17_CONSULTA_LAG_V03.xlsx",sheet =
"AGENTES5_barranquilla") #, range = "a850:l1320")
View(BA_GM )
#IMPORTANDO LONGITUD Y LATITUD
attach(BA_GM )
lat_BA_GM=latitud
lon_BA_GM=longitud

#PINTANDO ESTACIONES BARRANQUILLA GM
center = c(mean(lat_BA_GM ), mean(lon_BA_GM ));
zoom <- min(MaxZoom(range(lat_BA_GM), range(lon_BA_GM )));
#this overhead is taken care of implicitly by GetMap.bbox();
MyMap <- GetMap(center=center, zoom=zoom,markers =
paste0("&markers=color:blue|label:S|", "4.702147,-
74.015794&markers=color:green|label:G|4.711614,-74.012318&markers=",
"color:red|color:red|label:C|4.718217,-73.998284"), destfile = "PaisGM.png");
  tmp_PaisGM<- PlotOnStaticMap(MyMap, lat = lat_BA_GM, lon = lon_BA_GM, destfile =
"PaisGM1.png", cex=1.5,pch=20, col=c('red', 'blue', 'green'), add=FALSE);
  #Guardando mapa
  dev.copy(png,'E_BARRANQUILLA.png')
  dev.off()

#Guardando tabla
WriteXLS(c("PaisGM","BOG_GM","ME_GM","CA_GM",
"BA_GM"),"resultado.xlsx",SheetNames =
c("E_PaisGM","E_Bogota","E_Medellin","E_CALI","E_BARRANQUILLA"), perl = "perl")

```

## 10.2 Multiplicación de matriz W en Matlab

```

%% Import the data
%base
%[~, ~, raw] =
xlsread('/Users/lacero/Documents/OneDrive/UNAL/Maestria_economia/0-
Tesis/BD/2018.11.02_BD_agentes_V02.xlsm', 'BD');
% base depurada

```

```

[~, ~, raw] =
xlsread('/Users/lacero/Documents/OneDrive/UNAL/Maestria_economia/0-
Tesis/BD/2018.11.02_BD_agentes_V03_1.xlsm', 'BD_D');
raw = raw(2:end, :);
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x), raw)) = {'');
cellVectors = raw(:, 7);
raw = raw(:, [1, 2, 3, 4, 5, 6, 8, 9]);

%% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x), raw); % Find non-numeric
cells
raw(R) = {NaN}; % Replace non-numeric cells
raw(R) = {0}; % Replace non-numeric cells

%% create indes
data = reshape([raw{:}], size(raw));
idata2010 = find( data(:, 3) == 2010 );
idata2011 = find( data(:, 3) == 2011 );
idata2012 = find( data(:, 3) == 2012 );
idata2013 = find( data(:, 3) == 2013 );
idata2014 = find( data(:, 3) == 2014 );
idata2015 = find( data(:, 3) == 2015 );
idata2016 = find( data(:, 3) == 2016 );
idata2017 = find( data(:, 3) == 2017 );
%% Create matrix filtrada
data2010 = data(idata2010, :);
data2011 = data(idata2011, :);
data2012 = data(idata2012, :);
data2013 = data(idata2013, :);
data2014 = data(idata2014, :);
data2015 = data(idata2015, :);
data2016 = data(idata2016, :);
data2017 = data(idata2017, :);
%% import matrix W
[~, ~, raw] =
xlsread('/Users/lacero/Documents/OneDrive/UNAL/Maestria_economia/0-
Tesis/BD/2018.11.02_BD_agentes_V02.xlsm', 'W');
raw(cellfun(@(x) ~isempty(x) && isnumeric(x) && isnan(x), raw)) = {'');

%% Replace non-numeric cells with NaN
R = cellfun(@(x) ~isnumeric(x) && ~islogical(x), raw); % Find non-numeric
cells
raw(R) = {NaN}; % Replace non-numeric cells

%% Create output variable
W = reshape([raw{:}], size(raw));
clearvars raw R ;

% borrar variables innecesarias
%load 2018.11.02_sw.mat
%%guardar variables
save dataDD_D.mat data data2010 data2011 data2012 data2013 data2014
data2015 data2016 data2017 W -v7.3;

```



```
dataDD_D.W2012=W2012
clearvars W2012

W2013=vertcat(W,W,W,W,W,W,W,W,W,W,W,W);
W2013=horzcat(W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013,W2013);
dataDD_D.W2013=W2013
clearvars W2013

W2014=vertcat(W,W,W,W,W,W,W,W,W,W,W,W);
W2014=horzcat(W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014,W2014);
dataDD_D.W2014=W2014
clearvars W2014

W2015=vertcat(W,W,W,W,W,W,W,W,W,W,W,W);
W2015=horzcat(W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015,W2015);
dataDD_D.W2015=W2015
clearvars W2015

W2016=vertcat(W,W,W,W,W,W,W,W,W,W,W,W);
W2016=horzcat(W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016,W2016);
dataDD_D.W2016=W2016
clearvars W2016

W2017=vertcat(W,W,W,W,W,W,W,W);
W2017=horzcat(W2017,W2017,W2017,W2017,W2017,W2017,W2017,W2017);
dataDD_D.W2017=W2017
clearvars W2017

%% precios cantidades y Pib municipal ponderado por W
%%precios
dataDD_D.PP2010=dataDD_D.W2010*dataDD_D.P2010;
dataDD_D.PP2011=dataDD_D.W2011*dataDD_D.P2011;
dataDD_D.PP2012=dataDD_D.W2012*dataDD_D.P2012;
datestr(now)
dataDD_D.PP2013=dataDD_D.W2013*dataDD_D.P2013;
dataDD_D.PP2014=dataDD_D.W2014*dataDD_D.P2014;
dataDD_D.PP2015=dataDD_D.W2015*dataDD_D.P2015;
dataDD_D.PP2016=dataDD_D.W2016*dataDD_D.P2016;
dataDD_D.PP2017=dataDD_D.W2017*dataDD_D.P2017;
datestr(now)
%%cantidades
datestr(now)
dataDD_D.QP2010=dataDD_D.W2010*dataDD_D.Q2010;
dataDD_D.QP2011=dataDD_D.W2011*dataDD_D.Q2011;
dataDD_D.QP2012=dataDD_D.W2012*dataDD_D.Q2012;
dataDD_D.QP2013=dataDD_D.W2013*dataDD_D.Q2013;
dataDD_D.QP2014=dataDD_D.W2014*dataDD_D.Q2014;
```

```

dataDD_D.QP2015=dataDD_D.W2015*dataDD_D.Q2015;
dataDD_D.QP2016=dataDD_D.W2016*dataDD_D.Q2016;
dataDD_D.QP2017=dataDD_D.W2017*dataDD_D.Q2017;
datestr(now)
%piib municipal
datestr(now)
dataDD_D.PibP2010=dataDD_D.W2010*dataDD_D.Pib2010;
dataDD_D.PibP2011=dataDD_D.W2011*dataDD_D.Pib2011;
dataDD_D.PibP2012=dataDD_D.W2012*dataDD_D.Pib2012;
dataDD_D.PibP2013=dataDD_D.W2013*dataDD_D.Pib2013;
dataDD_D.PibP2014=dataDD_D.W2014*dataDD_D.Pib2014;
dataDD_D.PibP2015=dataDD_D.W2015*dataDD_D.Pib2015;
dataDD_D.PibP2016=dataDD_D.W2016*dataDD_D.Pib2016;
dataDD_D.PibP2017=dataDD_D.W2017*dataDD_D.Pib2017;
datestr(now)

%% armado base de datos para correr en R+ regresiones

dataDD_D.PP=vertcat(dataDD_D.PP2010,dataDD_D.PP2011,dataDD_D.PP2012,data
DD_D.PP2013,dataDD_D.PP2014,dataDD_D.PP2015,dataDD_D.PP2016,dataDD_D.PP2
017);
dataDD_D.QP=vertcat(dataDD_D.QP2010,dataDD_D.QP2011,dataDD_D.QP2012,data
DD_D.QP2013,dataDD_D.QP2014,dataDD_D.QP2015,dataDD_D.QP2016,dataDD_D.QP2
017);
dataDD_D.PibP=vertcat(dataDD_D.PibP2010,dataDD_D.PibP2011,dataDD_D.PibP2
012,dataDD_D.PibP2013,dataDD_D.PibP2014,dataDD_D.PibP2015,dataDD_D.PibP2
016,dataDD_D.PibP2017);
%% colocando en data los vectores ponderados
%precios ponderados
dataDD_D.data(:,9)=dataDD_D.PP;
%cantidades ponderadas
dataDD_D.data(:,10)=dataDD_D.QP;
%Pib ponderado
dataDD_D.data(:,11)=dataDD_D.PibP;

```

## 10.3 Modelos econométricos

### 10.3.1 Corridas y análisis con paneles de datos

```

#regresiones sin efecto espacial
#####
# con datos totales
search()
loadedNamespaces
rm(list=ls())

```

```
BD <- read_excel("2018.09.04_BD_modelo_V04.xlsx",sheet = "BD_modelo") #, range = "a1:l1320")
#mac
setwd("/Users/lacero/Documents/OneDrive/UNAL/Maestria_economia/0-Tesis/BD")
#windows
getwd()
setwd("D:/G/UNAL/Maestria_economia/0-Tesis/BD")
```

```
BD <- read_excel("2018.11.24_BD_D_modelo_V06.xlsx",sheet = "BD_D_modelo") #,)
BD <- read_excel("2018.11.30_BD_D_modelo_V08.xlsx",sheet = "BD_D_modelo") #,)
```

```
summary(BD)
save(BD, file = "BD2.RData");
attach(BD)
#View(BD)
load("BD2.RData");
## graficas generales sin ponderacion
# box plot alo
par(mfrow=c(1,2))
boxplot2(subset(BD)$P~subset(BD)$alo,col="gold",xlab="alo",ylab="COP/galon",
  main="P por alo",boxfill = 1:8, notch = FALSE,show.names= TRUE, outline=FALSE, par( cex
= 1,cex.axis=1.2,cex.lab=1.2,cex.main=1.5,cex.sub=1,mgp= c(1.5, 0.5, 0), mex=1))

boxplot2(subset(BD)$Q~subset(BD)$alo,col="gold",xlab="alo",ylab="galones",
  main="Q por alo",boxfill = 1:8, notch = FALSE,show.names= TRUE, outline=FALSE, par( cex
= 1,cex.axis=1.2,cex.lab=1.2,cex.main=1.5,cex.sub=1,mgp= c(1.5, 0.5, 0), mex=1))
par(mfrow=c(1,1))
####

#aaa=list(subset(BD)$Q~subset(BD)$alo,col="gold",xlab="alo",ylab="galones",
#  main="Q por alo",boxfill = 1:8, notch = FALSE,show.names= TRUE, outline=FALSE, par(
cex = 1,cex.axis=1.2,cex.lab=1.2,cex.main=1.5,cex.sub=1,mgp= c(1.5, 0.5, 0), mex=1))

#do.call("boxplot2",aaa)
####

# box plot departamento
boxplot2(subset(BD)$P~subset(BD)$DEPARTAMENTO,boxfill = 1:8,show.names= TRUE,
outline=FALSE,horizontal=FALSE,las = 2,par( cex =
0.5,cex.axis=1.5,cex.lab=1.5,cex.main=3,cex.sub=3,mgp= c(3, 0.5, 0), mex=2.8),
xlab="Departamento",ylab="COP/galon",
  main="P por Departamento")
boxplot2(subset(BD)$Q~subset(BD)$DEPARTAMENTO,boxfill = 1:8,show.names= TRUE,
outline=FALSE,horizontal=FALSE,las = 2,par( cex =
0.5,cex.axis=1.5,cex.lab=1.5,cex.main=3,cex.sub=3,mgp= c(3, 0.5, 0), mex=2.8),
xlab="Departamento",ylab="galones",
```

```

main="Q por Departamento")

BD=data.frame(BD,N_estaciones=(0))
#####

#caracterizacion estadistica departamentos
#frecuencias de estaciones por municipio
#E.D=as.data.frame(table(MUNICIPIO))
# inclui?n numero de estaciones por alo y por municipio en BD
E.D=as.data.frame(table(MUNICIPIO, alo))
#agregadon la frecuencia de estaciones al MUNICIPIO
#BD2.index <- seq(1:10)
# argumentos de la funcion args(plm)
BD.index <- seq(1:length(BD[[1]]))
alos <- seq(2010,2017,1)
Sys.time()

for ( i in BD.index) {
  BD$N_estaciones[[i]]=subset(E.D,MUNICIPIO==BD$MUNICIPIO[[i]]&alo==BD$alo[[i]])$Freq
}

Sys.time()
save(BD, file = "BD2.RData");
#####

#####
#Casos estudio sin ponderacion geogr?fica
#caso 1 log(P) ~ log(Q)-1
##test haussman
caso1.th=phtest(log(P) ~ log(Q)-1, data = BD,model = c("within","random"), method =c("chisq",
"aux"), vcov = vcovHC)
caso1.th
## within transformation: efectos fijos
caso1.PA.W<- plm(log(P) ~ log(Q)-1, data = BD, model = "within")
summary(caso1.PA.W)
## random transformation: efectos aleatorios
caso1.PA.R<- plm(log(P) ~ log(Q)-1, data = BD, model = "random")
summary(caso1.PA.R)
#####
##rrr=list(log(P) ~ log(Q)-1, data = BD, model = "random")
##jj=new.env()
##rr=do.call("plm",rrr,envir=jj,quote=TRUE)
##summary(rr)
#####

#caso 2 (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+ BRIO+CASAMOTOR+
COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+ MOBIL+ OCTANO

```

```

+PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso2.th.WR=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso2.th.WR
caso2.th.RW=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso2.th.RW
## within transformation: efectos fijos
caso2.PA.W= plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso2.PA.W)
## random transformation: efectos aleatorios
caso2.PA.R<- plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso2.PA.R)
### ejecutando indirecto
##L=list(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+ BRIO+CASAMOTOR+
COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+ MOBIL+ OCTANO
+PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
##LL=do.call("plm",L)
#summary(LL)

###

```

```

#caso 3 (log(P) ~ log(Q)+log(PIB_municipal)+log(N_estaciones)-1
##test haussman
caso3.th.WR=phtest(log(P) ~ log(Q)+log(PIB_municipal)+ TERPEL+log(N_estaciones)-1, data =
BD,model = c("within","random"), method =c("chisq", "aux"), vcov = vcovHC)
caso3.th.WR
caso3.th.RW=phtest(log(P) ~ log(Q)+log(PIB_municipal)+ TERPEL+log(N_estaciones)-1, data =
BD,model = c("random","within"), method =c("chisq", "aux"), vcov = vcovHC)
caso3.th.RW
## within transformation: efectos fijos
caso3.PA.W=          plm(log(P)          ~          log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso3.PA.W)
## random transformation: efectos aleatorios
caso3.PA.R<-          plm(log(P)          ~          log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_munic
ipio_alo)-1, data = BD, model = "random")
summary(caso3.PA.R)

#####
# con datos mayor alo
#buscando el alo mas reciente y con mayor numero de estaciones
detach(BD)
summary.data.frame(BD)
hist(subset(BD)$alo, axes = TRUE, plot = TRUE, labels = TRUE)
BD.MA=subset(BD,alo==2017)
attach(BD.MA)
#caso 4 mayor alo log(P) ~ log(Q)-1
##test haussman
caso4.th=phtest(log(P) ~ log(Q)-1, data = BD,model = c("within","random"), method =c("chisq",
"aux"), vcov = vcovHC)
caso4.th
## within transformation: efectos fijos
caso4.PA.W<- plm(log(P) ~ log(Q)-1, data = BD, model = "within")
summary(caso1.PA.W)
## random transformation: efectos aleatorios
caso4.PA.R<- plm(log(P) ~ log(Q)-1, data = BD, model = "random")
summary(caso4.PA.R)
#caso 5 mayor alo (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso5.th.WR=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+

```

```

MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso5.th.WR
caso5.th.RW=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso5.th.RW
## within transformation: efectos fijos
caso5.PA.W= plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso5.PA.W)
## random transformation: efectos aleatorios
caso5.PA.R<- plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso5.PA.R)
#caso 6 mayor alo (log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1
##test haussman
caso6.th.WR=phtest(log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov = vcovHC)
caso6.th.WR
caso6.th.RW=phtest(log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov = vcovHC)
caso6.th.RW
## within transformation: efectos fijos
caso6.PA.W= plm(log(P) ~ log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso6.PA.W)

```

```

## random transformation: efectos aleatorios
caso6.PA.R<- plm(log(P) ~ log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "random")
summary(caso6.PA.R)
#####

# con datos mayor mes
detach(BD.MA)
summary.data.frame(BD)
hist(subset(BD)$fecha, axes = TRUE, plot = TRUE, labels = TRUE)
BD.MM=subset(BD,mes==8&alo==2017)
attach(BD.MM)

#caso 7 mayor mes log(P) ~ log(Q)-1
##test haussman
caso7.th=phtest(log(P) ~ log(Q)-1, data = BD,model = c("within","random"), method =c("chisq",
"aux"), vcov = vcovHC)
caso7.th
## within transformation: efectos fijos
caso7.PA.W<- plm(log(P) ~ log(Q)-1, data = BD, model = "within")
summary(caso7.PA.W)
## random transformation: efectos aleatorios
caso7.PA.R<- plm(log(P) ~ log(Q)-1, data = BD, model = "random")
summary(caso7.PA.R)
#caso 8 mayor alo (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso8.th.WR=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso8.th.WR
caso8.th.RW=phtest(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)

```

```

caso8.th.RW
## within transformation: efectos fijos
caso8.PA.W= plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso8.PA.W)
## random transformation: efectos aleatorios
caso8.PA.R<- plm(log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso8.PA.R)
#caso          9          mayor          alo          (log(P)          ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1
##test haussman
caso9.th.WR=phtest(log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov = vcovHC)
caso9.th.WR
caso9.th.RW=phtest(log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov = vcovHC)
caso9.th.RW
## within transformation: efectos fijos
caso9.PA.W=          plm(log(P)          ~          log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso9.PA.W)
## random transformation: efectos aleatorios
caso9.PA.R<-          plm(log(P)          ~          log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "random")
summary(caso9.PA.R)
#COMPILADO DE CASOS SIN PONDERACION
casos=list(

caso1.th=caso1.th,caso1.PA.W=caso1.PA.W,caso1.PA.R=caso1.PA.R,caso2.th.WR=caso2.th.WR
,caso2.th.RW=caso2.th.RW,caso2.PA.W=caso2.PA.W,caso2.PA.R=caso2.PA.R,caso3.th.WR=cas
o3.th.WR,caso3.th.RW=caso3.th.RW,caso3.PA.W=caso3.PA.W,caso3.PA.R=caso3.PA.R,

caso4.th=caso4.th,caso4.PA.W=caso4.PA.W,caso4.PA.R=caso4.PA.R,caso5.th.WR=caso5.th.WR

```

```
,caso5.th.RW=caso5.th.RW,caso5.PA.W=caso5.PA.W,caso5.PA.R=caso5.PA.R,caso6.th.WR=cas
o6.th.WR,caso6.th.RW=caso6.th.RW,caso6.PA.W=caso6.PA.W,caso6.PA.R=caso6.PA.R,

caso7.th=caso7.th,caso7.PA.W=caso7.PA.W,caso7.PA.R=caso7.PA.R,caso8.th.WR=caso8.th.WR
,caso8.th.RW=caso8.th.RW,caso8.PA.W=caso8.PA.W,caso8.PA.R=caso8.PA.R,caso9.th.WR=cas
o9.th.WR,caso9.th.RW=caso9.th.RW,caso9.PA.W=caso9.PA.W,caso9.PA.R=caso9.PA.R)
save(casos, file = "casos.RData");
load("casos.RData")
#####
#Casos estudio con ponderacion geogr?fica
## graficas generales con ponderacion
# box plot alo
par(mfrow=c(1,2))
boxplot2(subset(BD)$PP_R~subset(BD)$alo,col="gold",xlab="alo",ylab="COP/galon",
  main="P por alo",boxfill = 1:8, notch = FALSE,show.names= TRUE, outline=FALSE, par( cex
= 1,cex.axis=1.2,cex.lab=1.2,cex.main=1.5,cex.sub=1,mgp= c(1.5, 0.5, 0), mex=1))

boxplot2(subset(BD)$QQ_R~subset(BD)$alo,col="gold",xlab="alo",ylab="galones",
  main="Q por alo",boxfill = 1:8, notch = FALSE,show.names= TRUE, outline=FALSE, par( cex
= 1,cex.axis=1.2,cex.lab=1.2,cex.main=1.5,cex.sub=1,mgp= c(1.5, 0.5, 0), mex=1))
par(mfrow=c(1,1))

# box plot departamento
boxplot2(subset(BD)$PP_R~subset(BD)$DEPARTAMENTO,boxfill = 1:8,show.names= TRUE,
outline=FALSE,horizontal=FALSE,las = 2,par( cex =
0.5,cex.axis=1.5,cex.lab=1.5,cex.main=3,cex.sub=3,mgp= c(3, 0.5, 0), mex=2.8),
xlab="Departamento",ylab="COP/galon",
  main="P por Departamento")
boxplot2(subset(BD)$QQ_R~subset(BD)$DEPARTAMENTO,boxfill = 1:8,show.names= TRUE,
outline=FALSE,horizontal=FALSE,las = 2,par( cex =
0.5,cex.axis=1.5,cex.lab=1.5,cex.main=3,cex.sub=3,mgp= c(3, 0.5, 0), mex=2.8),
xlab="Departamento",ylab="galones",
  main="Q por Departamento")

#caso 1 log(P) ~ log(Q)-1
##test haussman
caso1P.th=phtest(log(PP_R) ~ log(QQ_R)-1, data = BD,model = c("within","random"), method
=c("chisq", "aux"), vcov = vcovHC)
caso1P.th
## within transformation: efectos fijos
caso1P.PA.W<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "within")
summary(caso1P.PA.W)
## random transformation: efectos aleatorios
caso1P.PA.R<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "random")
summary(caso1P.PA.R)
```

```

#caso 2 (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+ BRIO+CASAMOTOR+
COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+ MOBIL+ OCTANO
+PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso2P.th.WR=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso2P.th.WR
caso2P.th.RW=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso2P.th.RW
## within transformation: efectos fijos
caso2P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso2P.PA.W)
## random transformation: efectos aleatorios
caso2P.PA.R<- plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso2P.PA.R)
#caso 3 (log(P) ~ log(Q)+log(PIB_municipal)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1
##test haussman
caso3P.th.WR=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD,model = c("within","random"), method =c("chisq", "aux"), vcov = vcovHC)
caso3P.th.WR

```

```

caso3P.th.RW=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD,model = c("random","within"), method =c("chisq", "aux"), vcov = vcovHC)
caso3P.th.RW
## within transformation: efectos fijos
caso3P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso3P.PA.W)
## random transformation: efectos aleatorios
caso3P.PA.R<- plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "random")
summary(caso3P.PA.R)

#####
# con datos mayor alo
#buscando el alo mas reciente y con mayor numero de estaciones
detach(BD)
summary.data.frame(BD)
hist(subset(BD)$alo, axes = TRUE, plot = TRUE, labels = TRUE)
BD.MA=subset(BD,alo==2017)
attach(BD.MA)
#caso 4 mayor alo log(P) ~ log(Q)-1
##test haussman
caso4P.th=phtest(log(PP_R) ~ log(QQ_R)-1, data = BD,model = c("within","random"), method
=c("chisq", "aux"), vcov = vcovHC)
caso4P.th
## within transformation: efectos fijos
caso4P.PA.W<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "within")
summary(caso4P.PA.W)
## random transformation: efectos aleatorios
caso4P.PA.R<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "random")
summary(caso4P.PA.R)
#caso 5 mayor alo (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso5P.th.WR=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera

```

```

_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso5P.th.WR
caso5P.th.RW=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso5P.th.RW
## within transformation: efectos fijos
caso5P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso5P.PA.W)
## random transformation: efectos aleatorios
caso5P.PA.R<- plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso5P.PA.R)
#caso 6 mayor alo (log(P) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1
##test haussman
caso6P.th.WR=phtest(log(PP_R) ~
log(QQ_R)+log(PIBP_municipal_R)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso6P.th.WR
caso6P.th.RW=phtest(log(PP_R) ~
log(QQ_R)+log(PIBP_municipal_R)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso6P.th.RW
## within transformation: efectos fijos
caso6P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso6P.PA.W)
## random transformation: efectos aleatorios

```

```

caso6P.PA.R<-      plm(log(PP_R)      ~      log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "random")
summary(caso6P.PA.R)

#####
# con datos mayor mes
detach(BD.MA)
summary.data.frame(BD)
hist(subset(BD)$fecha, axes = TRUE, plot = TRUE, labels = TRUE)
BD.MM=subset(BD,mes==8&alo==2017)
attach(BD.MM)

#caso 7 mayor mes log(P) ~ log(Q)-1
##test haussman
caso7P.th=phtest(log(PP_R) ~ log(QQ_R)-1, data = BD,model = c("within","random"), method
=c("chisq", "aux"), vcov = vcovHC)
caso7P.th
## within transformation: efectos fijos
caso7P.PA.W<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "within")
summary(caso7P.PA.W)
## random transformation: efectos aleatorios
caso7P.PA.R<- plm(log(PP_R) ~ log(QQ_R)-1, data = BD, model = "random")
summary(caso7P.PA.R)
#caso 8 mayor alo (log(P) ~ log(Q)+log(PIB_municipal)+AYATAWACOOP + BIOMAX+
BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+ ECOS+ESSO+ GULF+
MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA + SAVE + TERPEL +
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1
##test haussman
caso8P.th.WR=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE
+
TERPEL
+
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso8P.th.WR
caso8P.th.RW=phtest(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE
+
TERPEL
+
TEXACO
+
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso8P.th.RW

```

```

## within transformation: efectos fijos
caso8P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "within")
summary(caso8P.PA.W)
## random transformation: efectos aleatorios
caso8P.PA.R<- plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+AYATAWACOOP +
BIOMAX+ BRIO+CASAMOTOR+ COOMULPINORT+ CORPOIL+ DISCOWACOOP+
ECOS+ESSO+ GULF+ MOBIL+ OCTANO +PETROBRAS+PETROMIL+ PROXXON + PUMA +
SAVE + TERPEL + TEXACO +
ZAPATA_Y_VELASQUEZ+ZEUSS+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD, model = "random")
summary(caso8P.PA.R)
#caso 9 mayor alo (log(PP_R) ~
log(Q)+log(PIB_municipal)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municip
io_alo)-1
##test haussman
caso9P.th.WR=phtest(log(PP_R) ~
log(QQ_R)+log(PIBP_municipal_R)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("within","random"), method =c("chisq", "aux"), vcov =
vcovHC)
caso9P.th.WR
caso9P.th.RW=phtest(log(PP_R) ~
log(QQ_R)+log(PIBP_municipal_R)+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera
_municipio_alo)-1, data = BD,model = c("random","within"), method =c("chisq", "aux"), vcov =
vcovHC)
caso9P.th.RW
## within transformation: efectos fijos
caso9P.PA.W= plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "within")
summary(caso9P.PA.W)
## random transformation: efectos aleatorios
caso9P.PA.R<- plm(log(PP_R) ~ log(QQ_R)+log(PIBP_municipal_R)+
TERPEL+log(N_estaciones)+(PorcQ_bandera_T_alo)+(PorcQ_bandera_municipio_alo)-1, data =
BD, model = "random")
summary(caso9P.PA.R)

#COMPILADO DE CASOS CON PONDERACION
casosP=list(

caso1P.th=caso1P.th,caso1P.PA.W=caso1P.PA.W,caso1P.PA.R=caso1P.PA.R,caso2P.th.WR=ca
so2P.th.WR,caso2P.th.RW=caso2P.th.RW,caso2P.PA.W=caso2P.PA.W,caso2P.PA.R=caso2P.P

```

```
A.R, caso3P.th.WR=caso3P.th.WR, caso3P.th.RW=caso3P.th.RW, caso3P.PA.W=caso3P.PA.W, ca
so3P.PA.R=caso3P.PA.R,
```

```
caso4P.th=caso4P.th, caso4P.PA.W=caso4P.PA.W, caso4P.PA.R=caso4P.PA.R, caso5P.th.WR=ca
so5P.th.WR, caso5P.th.RW=caso5P.th.RW, caso5P.PA.W=caso5P.PA.W, caso5P.PA.R=caso5P.P
A.R, caso6P.th.WR=caso6P.th.WR, caso6P.th.RW=caso6P.th.RW, caso6P.PA.W=caso6P.PA.W, ca
so6P.PA.R=caso6P.PA.R,
```

```
caso7P.th=caso7P.th, caso7P.PA.W=caso7P.PA.W, caso7P.PA.R=caso7P.PA.R, caso8P.th.WR=ca
so8P.th.WR, caso8P.th.RW=caso8P.th.RW, caso8P.PA.W=caso8P.PA.W, caso8P.PA.R=caso8P.P
A.R, caso9P.th.WR=caso9P.th.WR, caso9P.th.RW=caso9P.th.RW, caso9P.PA.W=caso9P.PA.W, ca
so9P.PA.R=caso9P.PA.R)
```

## 10.4 Ejercicios de conglomerados en R+

```
##### ejercicios cluster
#Escenario
ES=c("P", "N_estaciones")
DEPARTAMENTO1="SANTANDER"
##### Centroides
CPQDE2010=subset(BD,
alo==2010&DEPARTAMENTO==DEPARTAMENTO1&mes==6&P>=0&Q>=0, na.exclude=TRUE);
summary(CPQDE2010);
#matriz distancia euclidea
MDE_CPQDE2010<-dist(CPQDE2010[,ES], method="euclidean", diag=TRUE)
#Normalizamos los datos
N_CPQDE2010.norm<-scale(CPQDE2010[,ES])
#MDE_CPQDE2010.norm<-
dist(N_CPQDE2010.norm[,c("P", "Q")], method="euclidean", diag=TRUE)
#MDE_CPQDE2010.norm.R=round(print(MDE_CPQDE2010.norm), 2)
#calcula de la distancia euclídea
MDE_CPQDE2010.norm.D<-
dist(CPQDE2010[,ES], method="euclidean", diag=TRUE)#, label=DatosCaso3.1$nombre_empresa)
MDE2_CPQDE2010=(MDE_CPQDE2010.norm.D)^2
round(print(MDE2_CPQDE2010), 2)
#efectuamos el cluster con método todo centroides
CL_CPQDE2010<-hclust(MDE2_CPQDE2010, method="centroid")
#Sacamos el historial de aglomeración del objeto hclust.centroides
data.frame(CL_CPQDE2010[2:1])
#dendrograma centroides
par(mfrow=c(1, 1))
main1=c("Cluster
Dendrograma", "centroides", CPQDE2010$DEPARTAMENTO[1], "Escenario", ES)
```

```

plot.hclust<-plot(CL_CPQDE2010,labels=CPQDE2010$MUNICIPIO, cex = 0.9, main =
main1)
rect.hclust(CL_CPQDE2010, k = 2, border = "red")
### #efectuamos el cluster con método vecino más cercano
CPQDE2010.cercano<-hclust(MDE2_CPQDE2010,method="single")
data.frame(CPQDE2010.cercano[2:1])
View(data.frame(CPQDE2010.cercano[2:1]))
##### dendograma vecino más cercano
main1=c("Cluster                               Dendograma", "Vecino
cercano",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)
plot(CPQDE2010.cercano,labels=CPQDE2010$MUNICIPIO, cex = 0.9, main = main1)
rect.hclust(CPQDE2010.cercano, k = 2, border = "red")
#efectuamos el cluster con método todo vecino mas lejano
CPQDE2010.lejano<-hclust(MDE2_CPQDE2010,method="complete")
data.frame(CPQDE2010.lejano[2:1])
#dendograma vecino mas lejao
main1=c("Cluster                               Dendograma", "Vecino
lejano",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)
plot(CPQDE2010.lejano,labels=CPQDE2010$MUNICIPIO, cex = 0.9, main = main1)
rect.hclust(hclust.lejano, k = 2, border = "red")

#efectuamos el cluster con método vinculacion promedio
CPQDE2010.promedio<-hclust(MDE2_CPQDE2010,method="average")
data.frame(CPQDE2010.promedio[2:1])
#dendograma vecino vinculacion promedio
main1=c("Cluster                               Dendograma", "vinculacion
promedio",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)
plot(CPQDE2010.promedio,labels=CPQDE2010$SICOM, cex = 0.9, main = main1)
rect.hclust(CPQDE2010.promedio, k = 2, border = "red")

#efectuamos el cluster con método todo de Ward
CPQDE2010.ward<-hclust(MDE2_CPQDE2010,method="ward.D")
data.frame(CPQDE2010.ward[2:1])
#dendograma método de Ward
main1=c("Cluster                               Dendograma", "vinculacion
promedio",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)
plot(CPQDE2010.ward,labels=CPQDE2010$SICOM, cex = 0.9, main = main1)
rect.hclust(CPQDE2010.ward, k = 2, border = "red")

#### indicadores nivel gerarquico
#Pedimos los indicadores para un jerarquico con distancia euclídea y método todo de
ward
CPQDE2010.NJ=scale(data.frame(CPQDE2010$P,CPQDE2010$N_estaciones))# se
incluyó escalamiento
#=#

```

```

#estimacion con promedio

MDE_CPQDE2010.norm.D<-
dist(CPQDE2010[,ES],method="euclidean",diag=TRUE)#,label=DatosCaso3.1$nombre_empresa)

#matriz.dis.euclid.caso3<-
dist(Datos_3_3_Caso[,c("automovi","tvcolor","video","microond","lavavaji","telefono")],method="euclidean",diag=TRUE)

hclust.average.CPQDE2010<-hclust(MDE_CPQDE2010.norm.D,method="average")

data.frame(hclust.average.CPQDE2010[2:1])

par(mfrow=c(1,1))

main1=c("Cluster                                Dendograma","vinculacion
promedio",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)

plot(hclust.average.CPQDE2010,labels=CPQDE2010$SICOM,xlab = NULL, ylab =
"Height", sub=NULL , cex = 1,cex = 0.5,cex.axis=0.5,cex.lab=0.5,cex.main=1,cex.sub=1, main =
main1)

rect.hclust(hclust.average.CPQDE2010, k = 2, border = "red")

#estimacion con complete

hclust.complete.CPQDE2010<-hclust(MDE_CPQDE2010.norm.D,method="complete")

data.frame(hclust.complete.CPQDE2010[2:1])

main1=c("Cluster                                Dendograma","vinculacion
completa",CPQDE2010$DEPARTAMENTO[1],"Escenario",ES)

plot(hclust.complete.CPQDE2010,labels=CPQDE2010$SICOM,cex = 1,cex =
0.5,cex.axis=0.5,cex.lab=0.5,cex.main=1,cex.sub=1, main = main1)

rect.hclust(hclust.complete.CPQDE2010, k = 2, border = "red")

## CON method = "ward.D2"

CPQDE2010.w2.res<-NbClust(CPQDE2010.NJ, distance = "euclidean", min.nc=2,
max.nc=6, method = "ward.D2", index = "allong")

CPQDE2010.w2.all=CPQDE2010.w2.res$All.index

CPQDE2010.w2.best=CPQDE2010.w2.res$Best.nc

CPQDE2010.w2.best_part=CPQDE2010.w2.res$Best.partition

## CON method = "average"

```

```
CPQDE2010.av.res<-NbClust(CPQDE2010.NJ, distance = "euclidean", min.nc=2,
max.nc=15, method = "average", index = "alllong")

CPQDE2010.av.all=CPQDE2010.av.res$All.index

CPQDE2010.av.best=CPQDE2010.av.res$Best.nc

CPQDE2010.av.best_part=CPQDE2010.av.res$Best.partition

## CON method = "complete"

CPQDE2010.co.res<-NbClust(CPQDE2010.NJ, distance = "euclidean", min.nc=2,
max.nc=6, method = "complete", index = "alllong")

CPQDE2010.co.all=CPQDE2010.co.res$All.index

CPQDE2010.co.best=CPQDE2010.co.res$Best.nc

CPQDE2010.co.best_part=CPQDE2010.co.res$Best.partition

numero.cluster<-seq(2,length(ward.ccc)+1,1)

ward.ccc <-CPQDE2010.w2.res$All.index[,4]

average.ccc <-CPQDE2010.w2.res$All.index[,4]

complete.ccc<-CPQDE2010.co.res$All.index[,4]

CPQDE2010.fig<-data.frame(numero.cluster,ward.ccc,average.ccc,complete.ccc)

metodo<-
c("ward","ward","ward","ward","ward","average","average","average","average","average","complet
e","complete","complete","complete","complete")

ccc<-c(ward.ccc ,average.ccc,complete.ccc)

cluster<-seq(2,(length(ward.ccc))*3+1,1)

CPQDE2010.fig_1<-data.frame(metodo,ccc,cluster)

library(ggplot2)

ggplot(data=CPQDE2010.fig_1, aes(x=cluster, y=ccc, group=metodo)) +
```

```
geom_line(aes(linetype=metodo))+
geom_point()+
scale_x_continuous(breaks = round(seq(min(data=CPQDE2010.fig_1$cluster),
max(data=CPQDE2010.fig_1$cluster), by = 1),1))

plot(hclust.complete.CPQDE2010,labels=CPQDE2010$SICOM,xlab = NULL, ylab =
"Height", sub=NULL, hang=-1,cex = 0.6)

rect.hclust(hclust.complete.CPQDE2010, k = 2, border = "red")

plot(hclust.average.caso3,labels=Datos_3_3_Caso$id,xlab = NULL, ylab = "Height",
sub=NULL, hang=-1,cex = 0.6)

rect.hclust(hclust.average.caso3, k = 2, border = "red")

plot(hclust.complete.caso3,labels=Datos_3_3_Caso$id,xlab = NULL, ylab = "Height",
sub=NULL, hang=-1,cex = 0.6)

rect.hclust(hclust.complete.caso3, k = 2, border = "red")

#efectuamos el cluster con método todo centroide
BP=3# apartir de resultado anterior
kmeans.CPQDE2010<-kmeans(CPQDE2010.NJ, BP) # solución de dos conglomerados
#obtenemos las medias
aggregate(CPQDE2010.NJ,by=list(kmeans.CPQDE2010$cluster),FUN=mean)
# añadimos la pertenencia al cluster
CPQDE2010.NJ<- data.frame(CPQDE2010.NJ, kmeans.CPQDE2010$cluster)
print(kmeans.CPQDE2010$cluster)
print(kmeans.CPQDE2010$centers)
print(kmeans.CPQDE2010$totss)
print(kmeans.CPQDE2010$withinss)
print(kmeans.CPQDE2010$tot.withinss)
```

```
print(kmeans.CPQDE2010$betweenss)
print(kmeans.CPQDE2010$size)
print(kmeans.CPQDE2010$iter)

# Paquete SMACOF
#=====

fit_CPQDE2010<-mds(delta=MDE_CPQDE2010,ndim=2,type="interval")
#Generaci3n del gr3fico
x2 <- fit_CPQDE2010$conf[,1]
y2 <- fit_CPQDE2010$conf[,2]
qplot(x2, y2, colour = l("black"), xlab=ES[1], ylab=ES[2])
# Coordinadas
print(fit_CPQDE2010$conf)
# Disparidades
print(fit_CPQDE2010$dhat)
# Distancias entre configuraciones
print(fit_CPQDE2010$confdist)
# Stress
print(fit_CPQDE2010$stress)
# Stress por punto
print(fit_CPQDE2010$spp)
#RSQ
print(1-fit_CPQDE2010$rss)
dist<-cbind(c(fit_CPQDE2010$dhat))
dism<-cbind(c(fit_CPQDE2010$confdist))
lm_0_CPQDE2010=data.frame("lm(dist~dism)","lm(dist~dism)","lm(dist~dism)")
library( data.table)
```

```
library(broom)
library(tidyr)
library(dplyr)

lm_1_CPQDE2010=tidy(lm(dist~dism))
lm_2_CPQDE2010=tidy(lm(dist~dism))
lm_3_CPQDE2010=tidy(lm(dist~dism))
summary(lm_CPQDE2010)
all_models <- rbind(lm_0_CPQDE2010,lm_1_CPQDE2010,
lm_2_CPQDE2010,
lm_3_CPQDE2010,use.names=TRUE)

tidy(mod,quick=F,exponentiate=F,conf.level=0.95,conf.int = TRUE)

all_models <- rbindlist(
mutate(lm_1_CPQDE2010),
mutate(lm_2_CPQDE2010),
mutate(lm_3_CPQDE2010))

par(mfrow=c(2,3))
# gráfico: "Shepard"
plot(fit_CPQDE2010, plot.type="Shepard", plot.dim = c(1,2), sphere = TRUE,
```

```
      subscale = 0.1, col = 1, label.conf = list(label = TRUE, pos = 3, col = 1, cex = 0.8),
shepard.x = NULL, identify = FALSE, type = "p", pch = 20, asp = 1, col.hist = NULL)

# gráfico: "confplot"

plot(fit_CPQDE2010, plot.type="confplot", plot.dim = c(1,2), sphere = TRUE,

      subscale = 0.1, col = 1, label.conf = list(label = TRUE, pos = 3, col = 1, cex = 0.8),
shepard.x = NULL, identify = FALSE, type = "p", pch = 20, asp = 1, col.hist = NULL)

# gráfico: , "resplot"

plot(fit_CPQDE2010, plot.type="resplot", plot.dim = c(1,2), sphere = TRUE,

      subscale = 0.1, col = 1, label.conf = list(label = TRUE, pos = 3, col = 1, cex = 0.8),
shepard.x = NULL, identify = FALSE, type = "p", pch = 20, asp = 1, col.hist = NULL)

# gráfico: "stressplot"

plot(fit_CPQDE2010, plot.type="stressplot", plot.dim = c(1,2), sphere = TRUE,

      subscale = 0.1, col = 1, label.conf = list(label = TRUE, pos = 3, col = 1, cex = 0.8),
shepard.x = NULL, identify = FALSE, type = "p", pch = 20, asp = 1, col.hist = NULL)

# gráfico: "bubbleplot"

plot(fit_CPQDE2010, plot.type="bubbleplot", plot.dim = c(1,2), sphere = TRUE,

      subscale = 0.1, col = 1, label.conf = list(label = TRUE, pos = 3, col = 1, cex = 0.8),
shepard.x = NULL, identify = FALSE, type = "p", pch = 20, asp = 1, col.hist = NULL)
```

## 11 Anexo 5 matriz W

Se incluye link al archivo digital con la matriz W, dado que es una matriz de 5254 filas que pesa 374MB y si se imprime serian 400 hojas en PDF que probablemente no generen mayor valor en el documento debido a ello se incluyó el link del archivo al google drive de la cuenta de la universidad el cual es:

[Matriz W.xlsx](#)<sup>1</sup>

---

<sup>1</sup> [https://drive.google.com/file/d/145M0nxzUHe2OMZZ\\_3XyDIra3E6A3tiKj/view?usp=sharing](https://drive.google.com/file/d/145M0nxzUHe2OMZZ_3XyDIra3E6A3tiKj/view?usp=sharing)