**Situación Problemática:**

Conforme a los años que han pasado podemos notar que el precio de la canasta básica ha cambiado de precio. Gracias a las fuentes de información: WFP, RRD, Program Unit SO Tindouf Algeria,VAM Unit SO Tindouf, ONASA, SIMA, el gobierno no Benin, INE, entre otras pudimos encontrar la información respecto al precio de diferentes años de los alimentos básicos en países alrededor del mundo. Ya que como seres humanos es vital ingerir alimento, el cual es comprado; surge la inquietud de saber si este precio subirá, se mantendrá o bajará en los próximos años.

**Problema científico:**

Se desea saber si el precio de la canasta básica va a cambiar en los próximos años. La moneda alrededor del mundo, el valor que tiene se va deteriorando haciendo que todo lo que tiene que ver con el movimiento consumista aumente en su valor. Con esto se encontraran personas sin la capacidad de cubrir como mínimo los alimentos básicos para una buena alimentación con esto se aumentará la tasa de desnutrición alrededor del mundo.

**Objetivos:**

Crear un análisis predictivo en el precio de la canasta básica por si esta subirá o bajará porcentualmente de forma global en los próximos años. Para ser específicos se buscará hacer una predicción de los precios de los alimentos básicos del año actual a 5 años adelante.

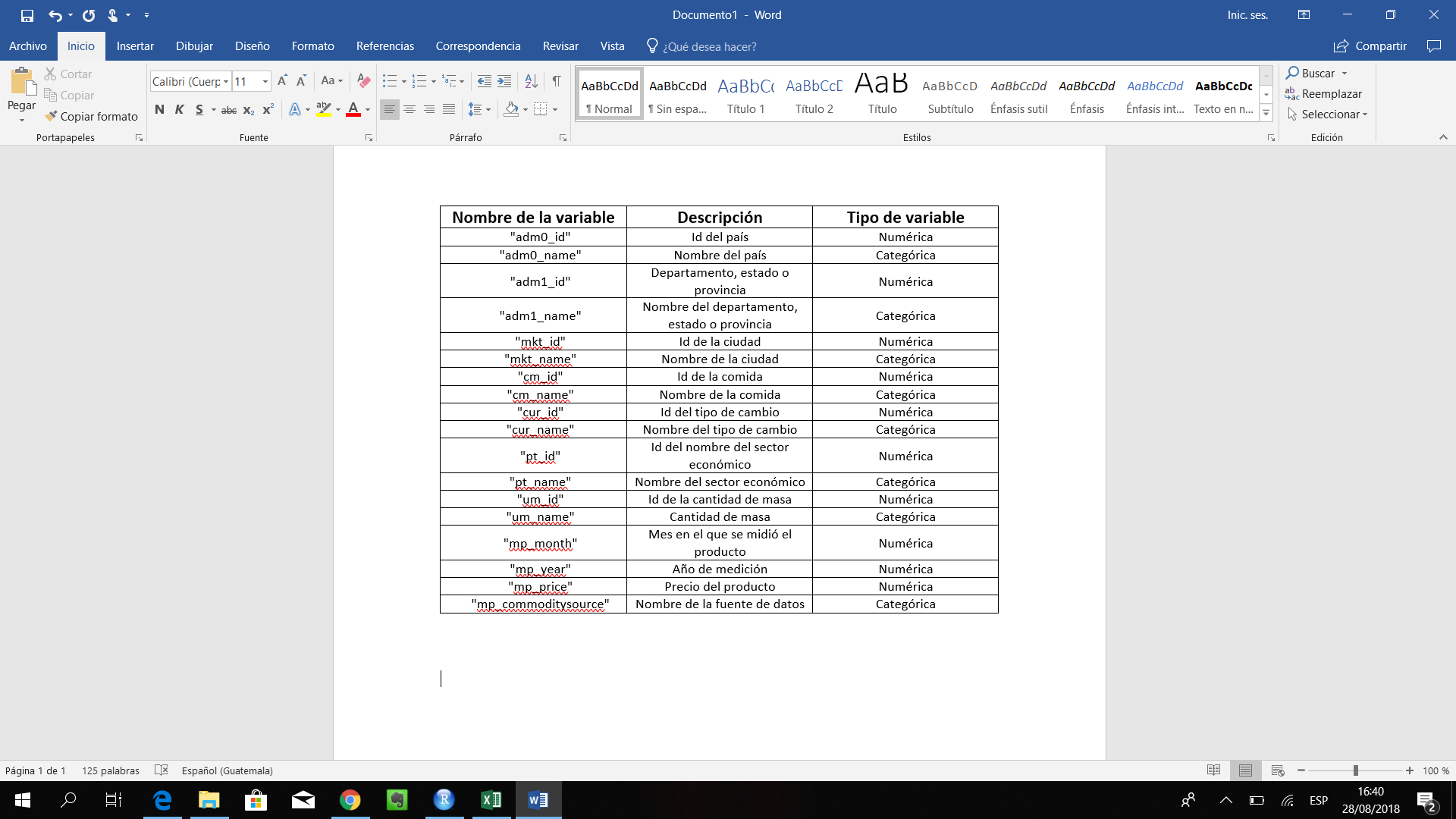
**Descripción de datos**

Estos datos cuentan con variables bastante independientes entre sÍ. solamente se encontró que el coeficiente de correlación más elevado es entre la ciudad y el tipo de cambio el cual es de 0.558. Además de este todo es independiente entre sÍ. Este dataset tiene un acantidad de 18 variables, con un total de 783788 observaciones. Se cuentan tanto con variables cuantitativas como cualitativas.

Los datos encontrados son datos recaudados alrededor del mundo acerca de los precios de algunos alimentos en diferentes países con su respectivo peso, moneda, ciudad, etc. Para comenzar se realizó un análisis de datos con una tabla de correlación la cual se muestra a lo largo del documento.

Información de los datos:

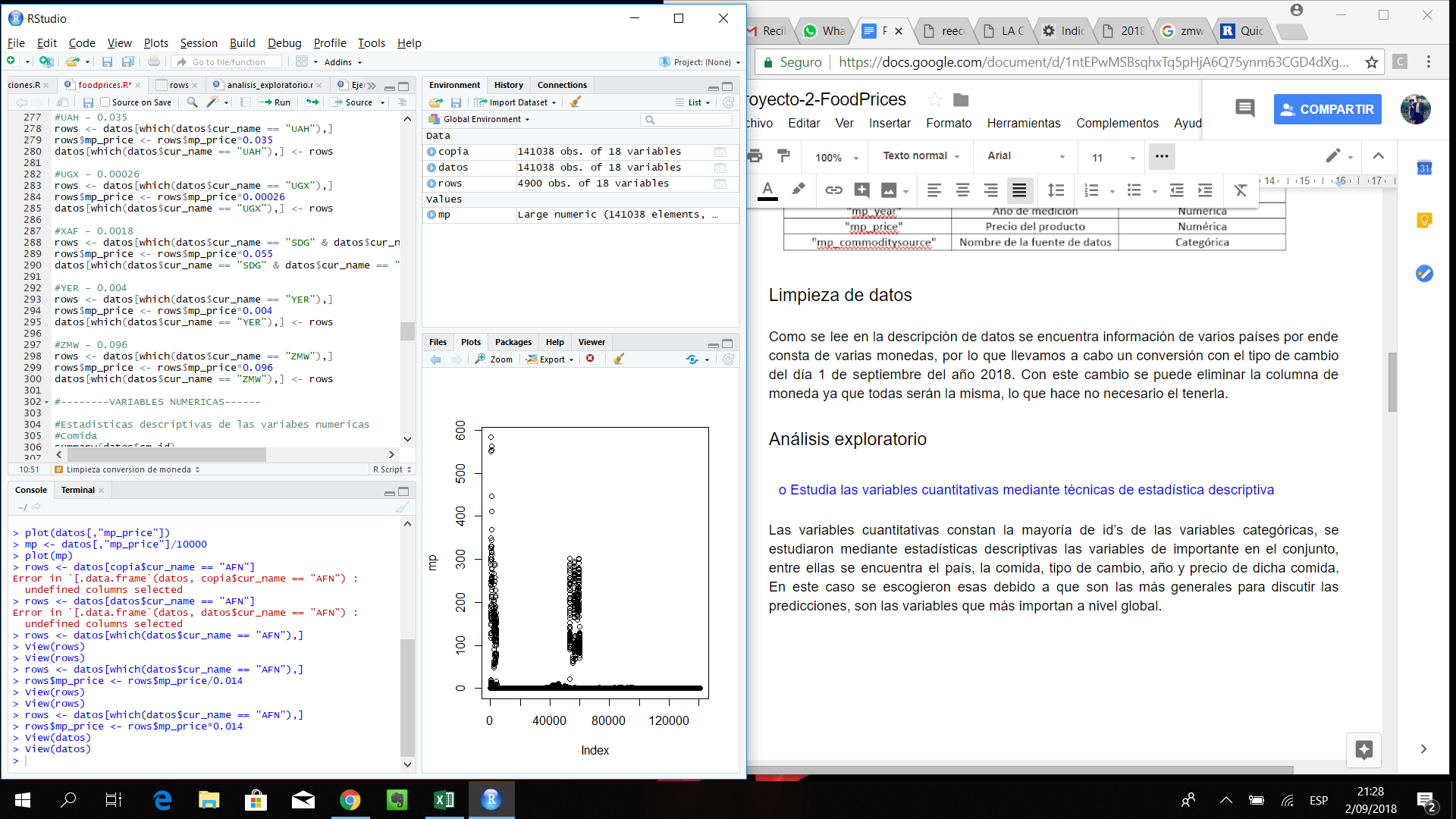
Las variables de las cuales consta el documento son las siguientes junto con 783,788 filas de datos, a continuación se encuentra el significado de cada una de ellas.



Limpieza de datos

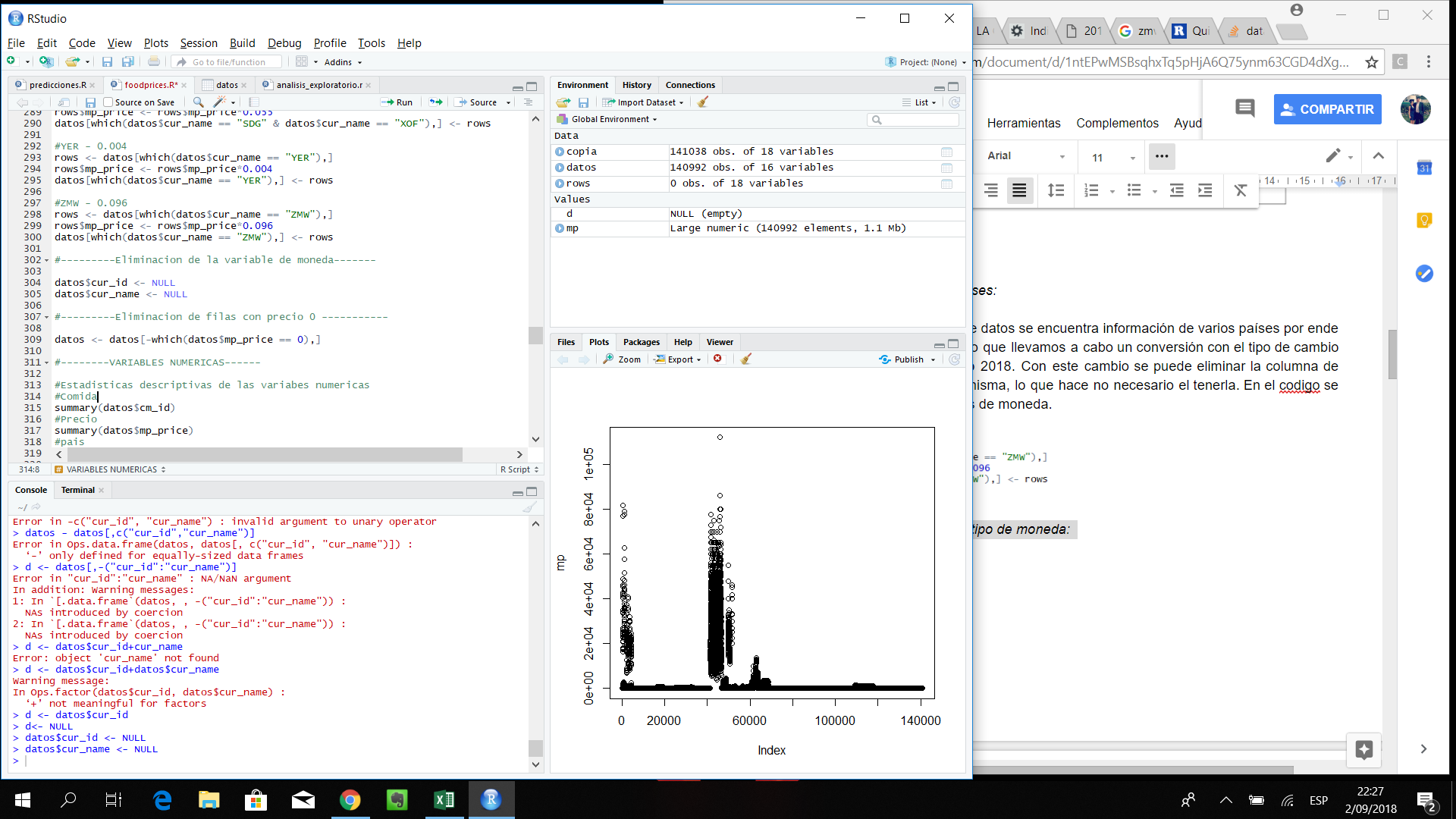
*Precios en dólares estadounidenses:*

Como se lee en la descripción de datos se encuentra información de varios países por ende consta de varias monedas, por lo que llevamos a cabo un conversión con el tipo de cambio del día 1 de septiembre del año 2018. Con este cambio se puede eliminar la columna de moneda ya que todas serán la misma, lo que hace no necesario el tenerla. En el código se usó el mismo para todos los tipos de moneda.



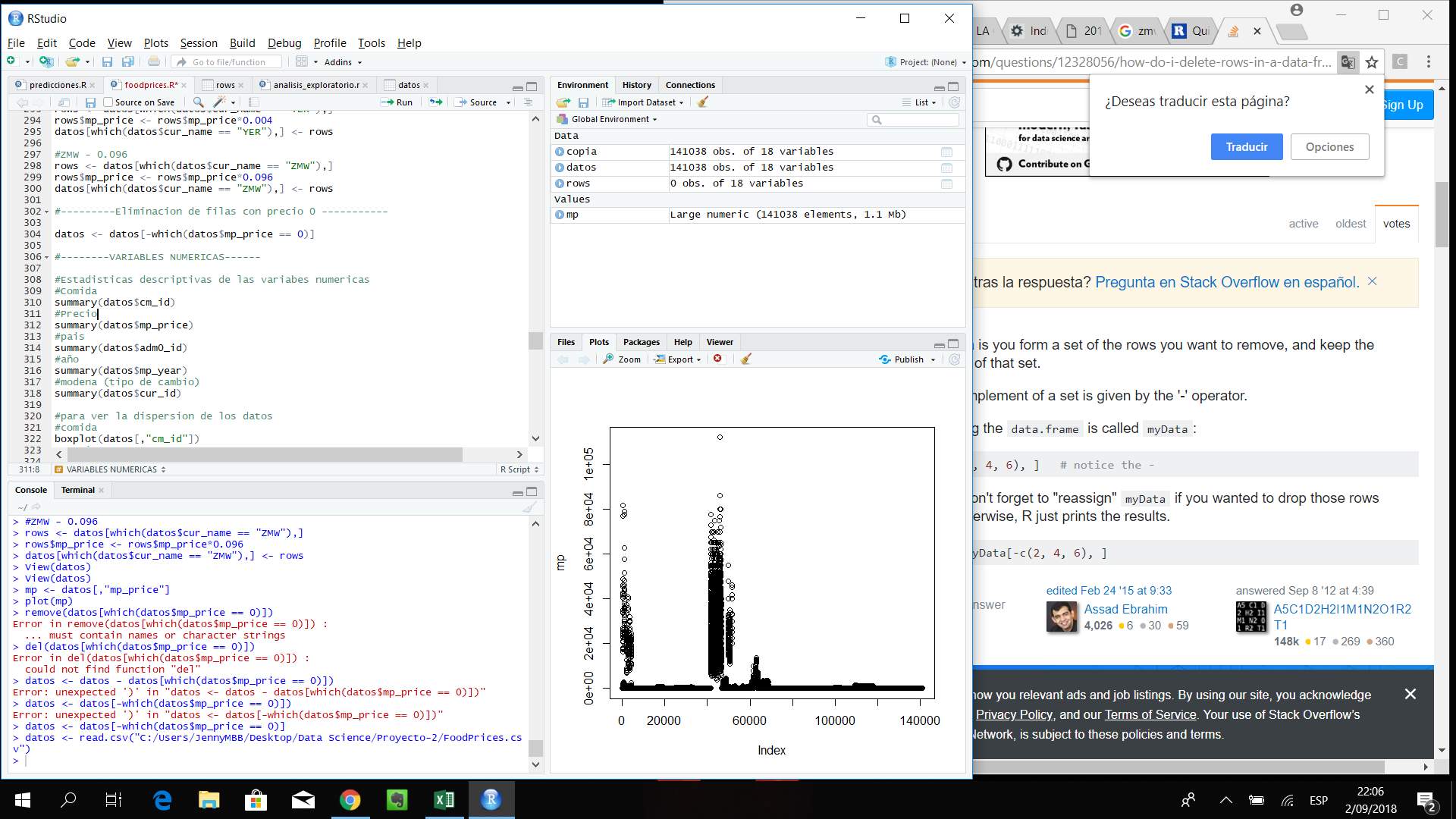
*Eliminación de las columnas del tipo de moneda:*

Las mayoría de las variables son conforme el nombre de las variables y el id de las mismas, por lo que al eliminar la variable de la moneda también sus id´s.



*Eliminación de filas con precio 0:*

Si nuestro objetivo es predictivo, se debe eliminar la mayor cantidad de desviación que se pueda de la base de datos. En algunos datos se encuentra en 0 el precio de los productos, sin embargo, ningún producto es gratis según una investigación realizada, por lo que se procede a eliminar dichos datos y así reducimos el ruido en el análisis. Con esto también se eliminan las filas vacías las cuales constaban solo del id de algunos países correspondientes.



*Eliminación de algunos artículos alimenticios irrelevantes:*

* *Livestock:*

Podemos observar al analizar los precios, habían algunos datos atípicos por lo que al evaluar existían artículos irrelevantes como animales para consumo humano como pollos, cerdos, reses, etc. los precios de dicho artículos se colocaban tomando en cuenta la edad del animal y estos aun con vida. Este tipo de artículos no pueden ser medidos para un estudio de alimentos básicos para una persona, lo cual hace factible la eliminación de los mismos.

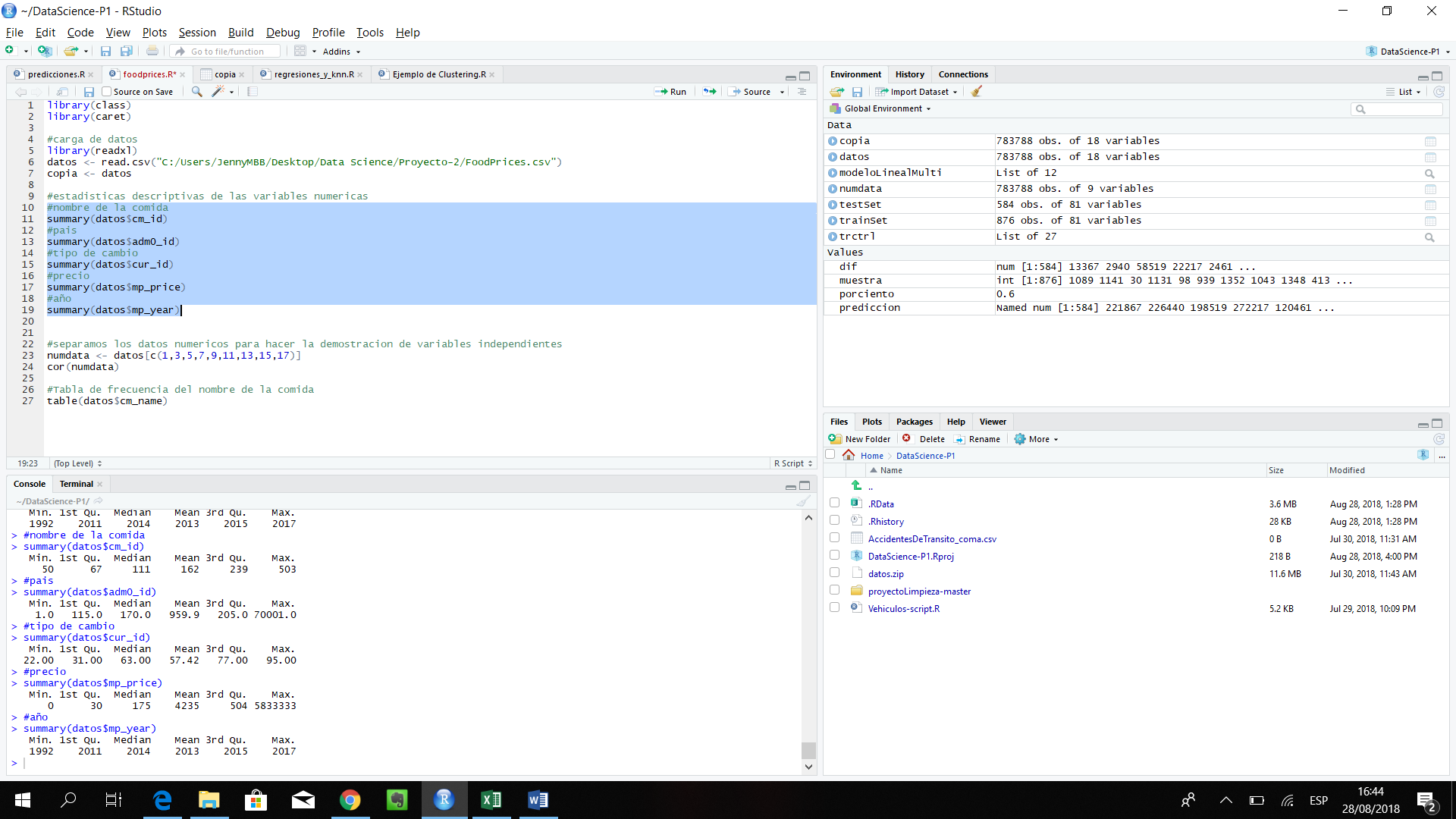
* *Unión de elementos*

En artículos como frutas, verduras, legumbres y lacteos se encuentran repetidos pero con nombres o cantidades diferentes, por lo que se procede a juntarlos.

Análisis exploratorio

o Estudia las variables cuantitativas mediante técnicas de estadística descriptiva

Las variables cuantitativas constan la mayoría de id’s de las variables categóricas, se estudiaron mediante estadísticas descriptivas las variables de importante en el conjunto, entre ellas se encuentra el país, la comida, tipo de cambio, año y precio de dicha comida. En este caso se escogieron esas debido a que son las más generales para discutir las predicciones, son las variables que más importan a nivel global.

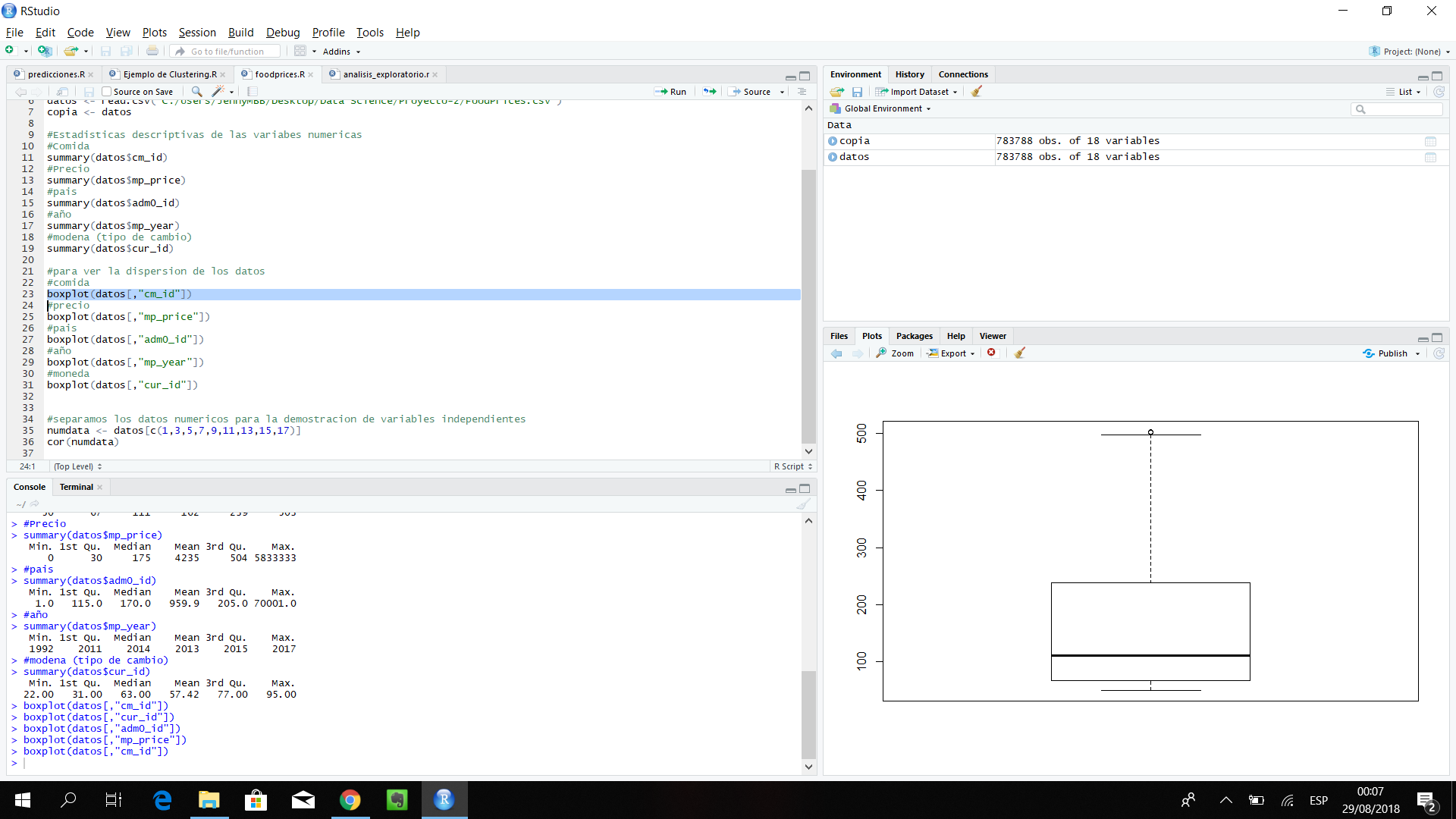


o Hace gráficos exploratorios como histogramas, diagramas de cajas y bigotes, gráficos de dispersión que ayudan a explicar los datos

Para hacer una revisión de los datos atípicos encontrados en los datos se van a realizar gráficos de caja y bigotes. Donde la línea en medio de la caja es la mediana la cual divide en 50 50 los datos.

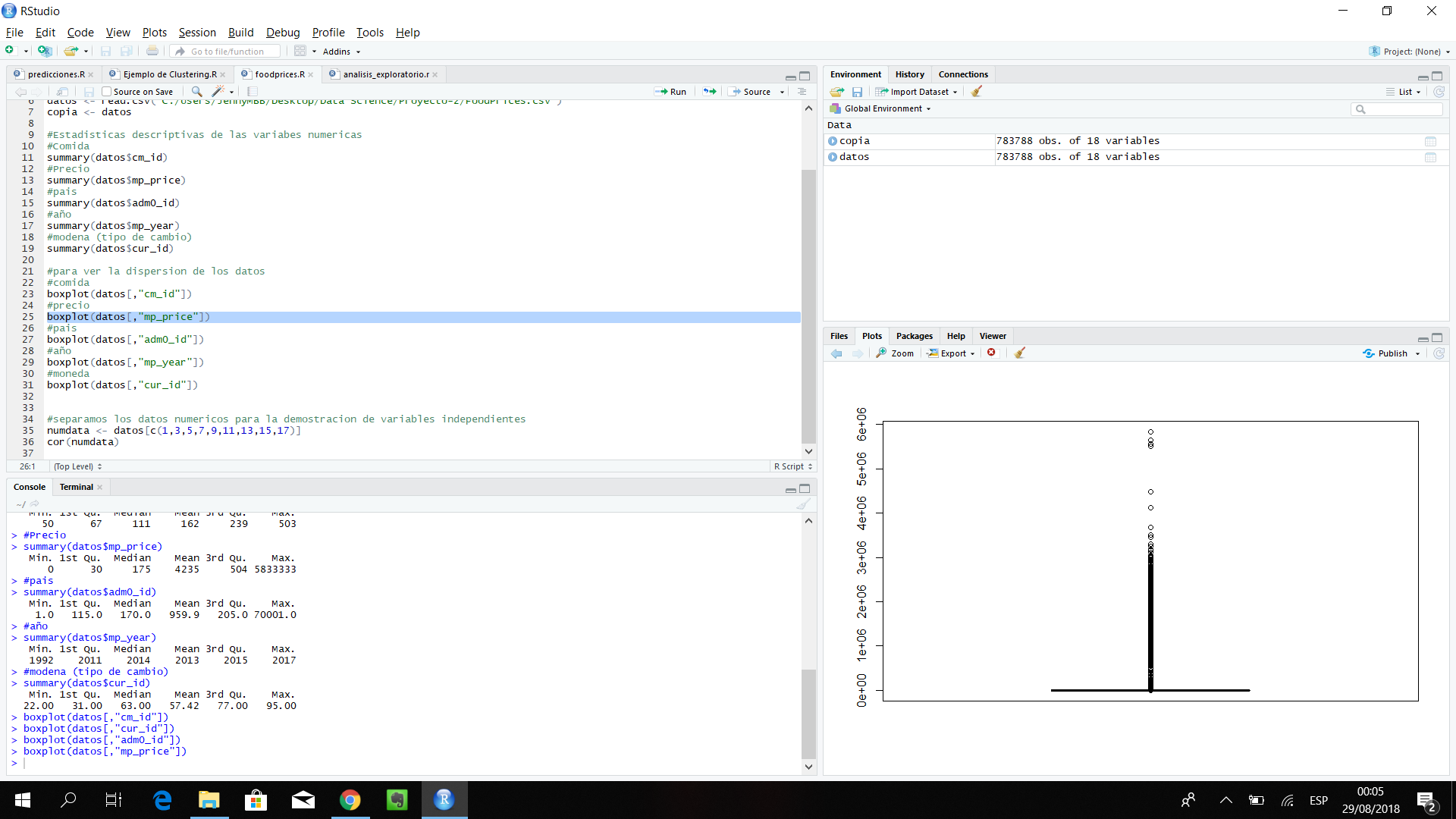
Comida

En esta variable se encuentra que en una mitad están menos dispersos que la otra y en esa misma se encuentra un dato atípico en la parte entre 400 - 500.



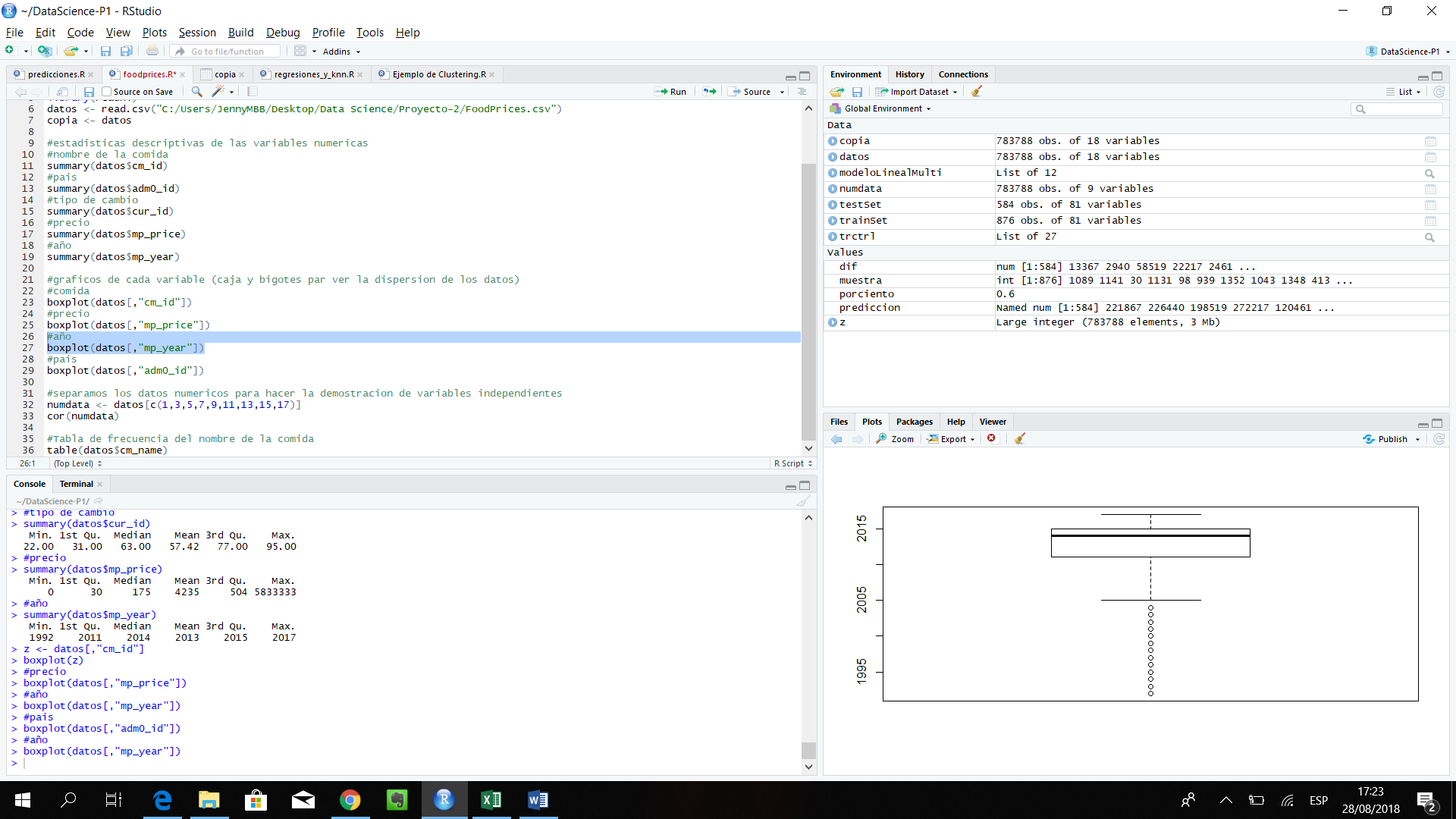
Precio

Los precios están muy dispersos sin embargo la mayoría se encuentran entre el precio de 2-25, pero hay que tomar en cuenta que también los datos de la comida van dispersos en cantidad de masa lo cual afecta directamente a la dispersión.



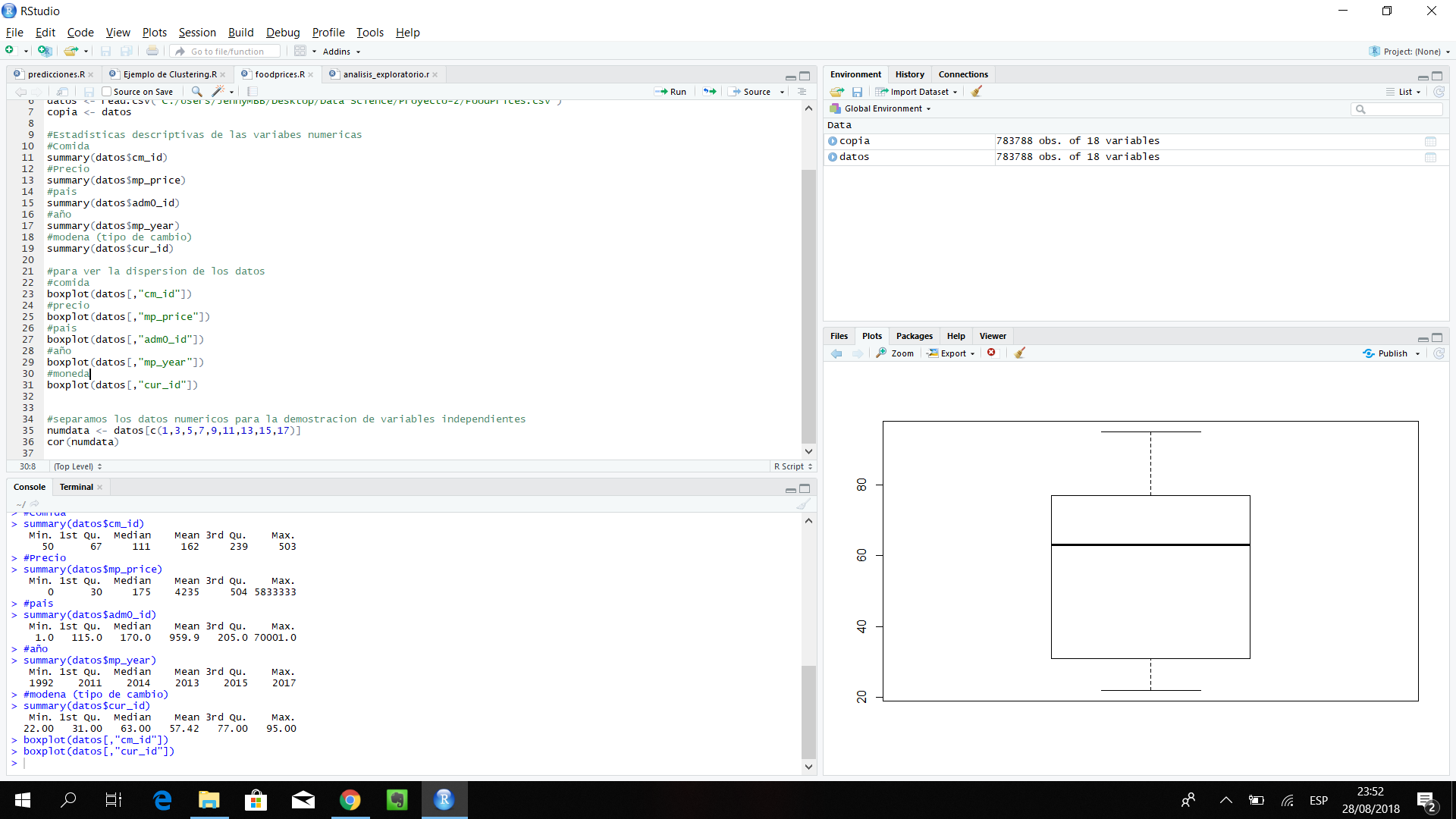
Año

En la variable de años según el diagrama de caja y bigotes podemos observar que los años que nos ayudarían más es del 2005 hacia el 2017 debido a que por debajo de estos se tiene muy poca información



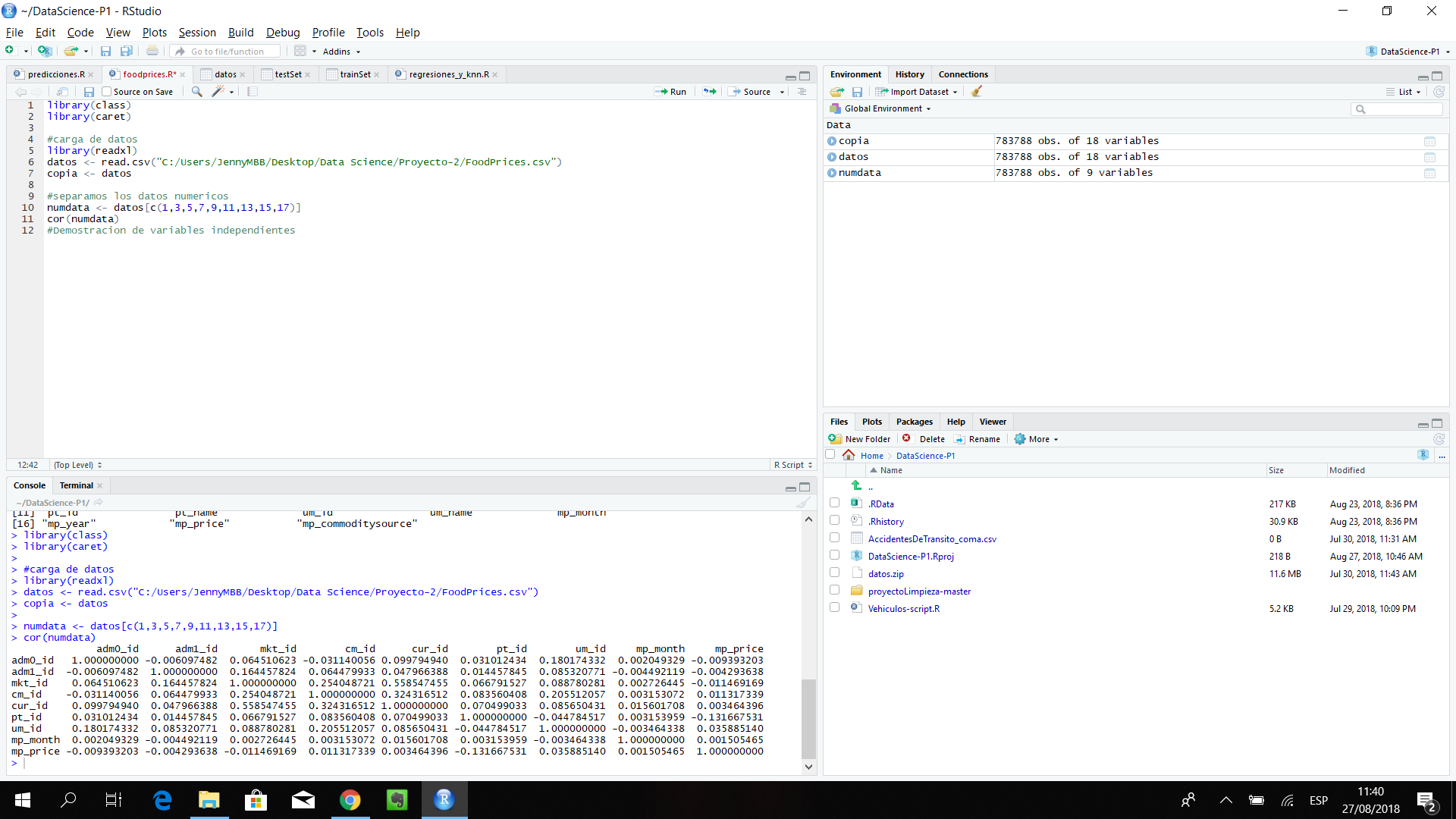
Moneda - tipo de cambio

Es la variable que tiene mejor dispersión a pesar que la mediana está un poco desviada del centro, lo que es un poco extraño si tomamos en cuenta la distribución de países.



o Analiza las correlaciones entre las variables, trata de explicar los outliers (puntos atípicos) y toma decisiones acertadas ante la presencia de valores faltantes.

En la siguiente imagen se demuestra que entre la variables numéricas se tiene independencia, siendo la mayoría de ellas id. El coeficiente de correlación más elevado es entre la ciudad y el tipo de cambio el cual es de 0.558.



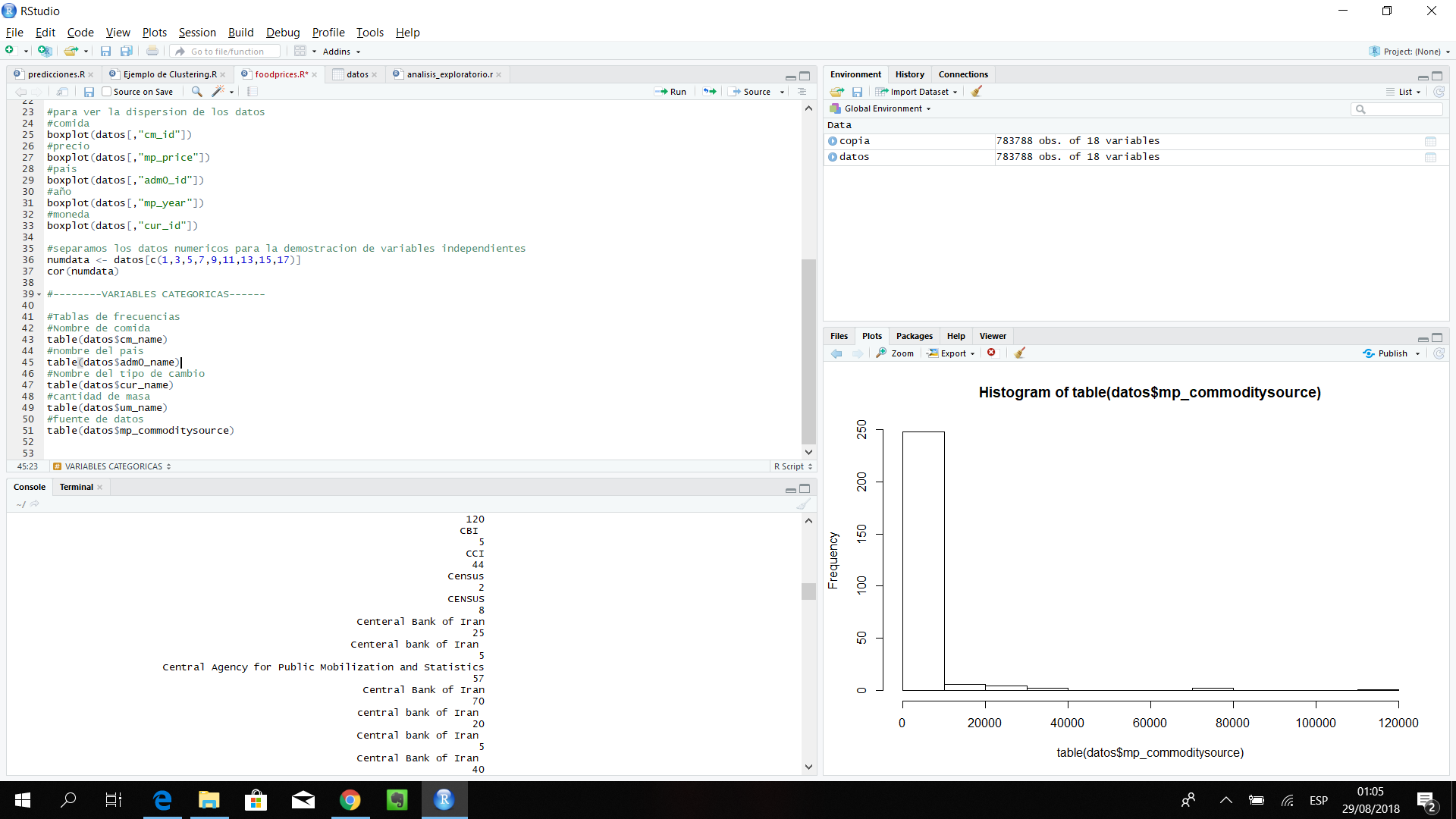
no pearson

o Estudia las variables categóricas

o Elabora gráficos de barra, tablas de frecuencia y de proporciones o Explica muy bien todos los procedimientos y los hallazgos que va haciendo. o Valora la posibilidad de hacer un análisis de componentes principales y si es factible aplica la técnica

Variables categóricas

Para evaluar las variables categóricas se hicieron tablas de frecuencia de las variables. A Continuación se mostraran las importantes:



o Determina el mejor número de clusters a utilizar usando el método de Ward

o Hace el agrupamiento con cualquiera de los algoritmos estudiados

o Verifica la calidad del agrupamiento usando el método de la silueta.

o Interpreta los grupos, usando para eso las variables numéricas y categóricas dentro de cada grupo. -