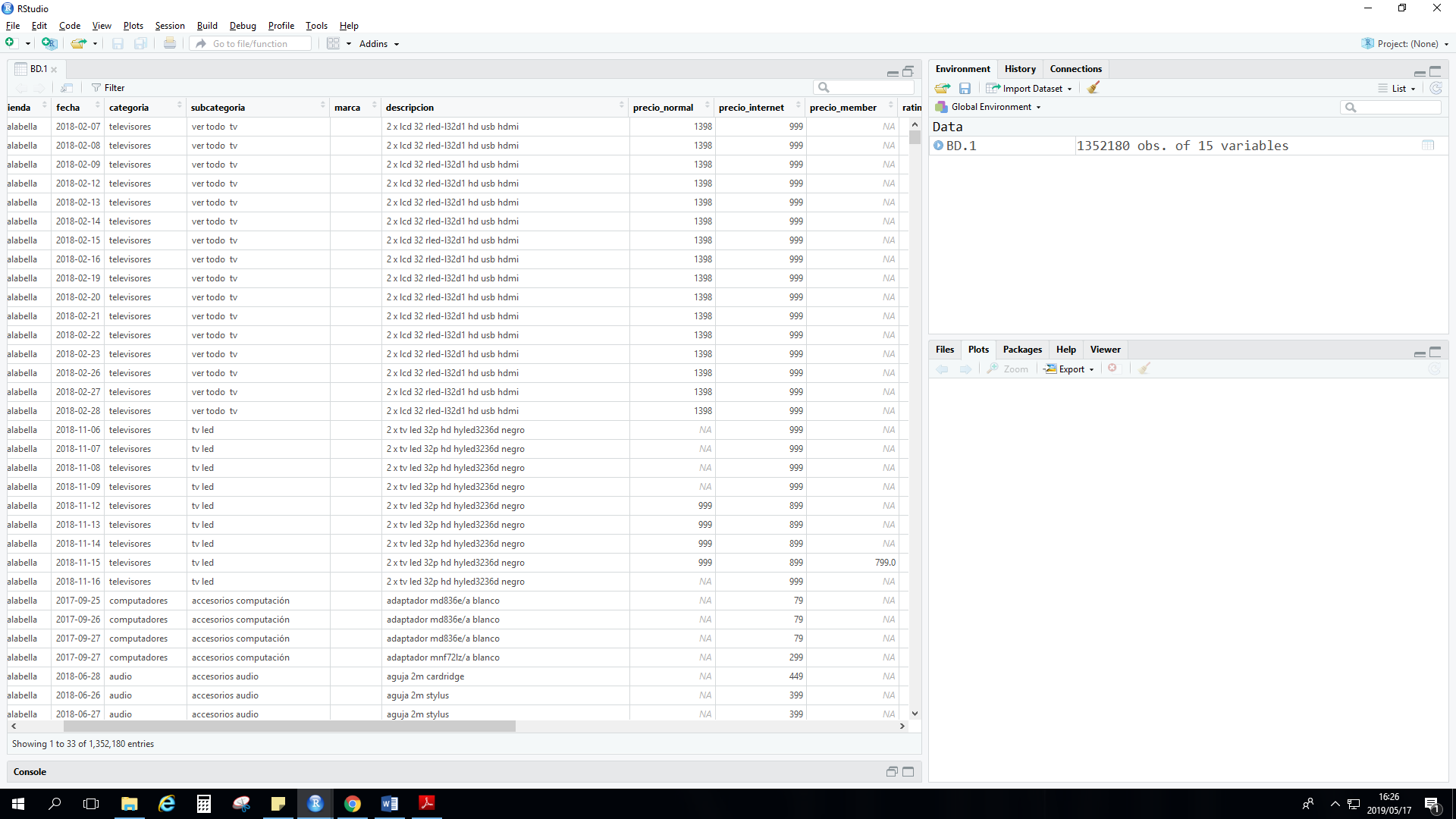
PROCESAMIENTO Y ACTUALIZACION DE LA BASE DE DATOS: SAGA FALABELLA[[1]](#footnote-1)

El siguiente documento detalla el proceso realizado a la base de datos *Base\_Final\_Peru.RData*. Esta base clasificada como base de datos panel dado que realiza el seguimiento de los precios de los bienes a lo largo del periodo de estudio para la tienda minorista Saga Falabella.

La base tiene una fecha de inicio registrada el *19 de setiembre del 2016 (2016-09-19)*  y una fecha final el 30 de abril del 2019 (2019-04-30), esta fecha final está determinada por la última actualización realizada a la base de datos. Esta puede seguir alterando y aumentando con el programa **scraping**.

Una vez actualizado la base de datos hasta la fecha final especificada, se realiza el proceso de limpieza que consiste en la creación de un identificador único a cada producto de la base, la corrección de categorías y subcategorías, así como el ordenamiento de la base por ítem en orden ascendente.

Se menciona que esta base no hace seguimiento únicamente a un solo precio por ítem sino que dado las características de la página web, podemos scrapear 3 diferentes tipos de bienes. Se tiene el precio de tienda física al que se le denominada precio normal, el precio de internet y el precio por ser socio de alguna tarjeta de la tienda, denominada precio member.

Dado que el código de scrapeo no es perfecto, este puede no obtener la información de los bienes por algunos días. Para estos casos donde no hay registro de los precios por determinado momento se realiza una regla para completarlos (Revisa script Creación ID).

*Ejemplo de la base final*

**Proceso de filtrado**

Esta base de datos realiza un seguimiento de aprox 10 mil bienes identificados únicamente a lo largo del periodo de estudio. Sin embargo, debido a diversas circunstancias como caída del scrapeo o el agotamiento de un bien, estos ítems pueden que no tengan la información de precios completa. Por ejemplo, para un determinado bien con el ID ***j***, este se solo aparece entre 4 o 5 veces en los dos años y medio.

Al considerar el problema anterior, y con la finalidad de poder tener una base donde los ítems que permanezcan tengan precios continuos en el periodo de estudio, entonces se procede a conocer el estado actual de nuestra base con la obtención de algunos estadísticos, distribuciones gráficos, entre otros. Este proceso se realiza en la carpeta **H:\SGInvEco\ALMACEN\_TRABAJO\Alexandra\Pass-through\Analisis\_base**

1. Todos los resultados se encuentran en la carpeta de H:\SGInvEco\ALMACEN\_TRABAJO\Alexandra\Pass-through [↑](#footnote-ref-1)