

Actividad 3. Dando respuestas a través de la visualización – 20-27/11/2019

Material: En ALUD están colgados 3 datasets con información productos de supermercado/grandes superficies.

Paso 0. Revisar la información existente en cada uno de ellos y ver qué acciones de Data Cleaning son necesarias.

Dar respuesta a las siguientes cuestiones mediante visualizaciones.

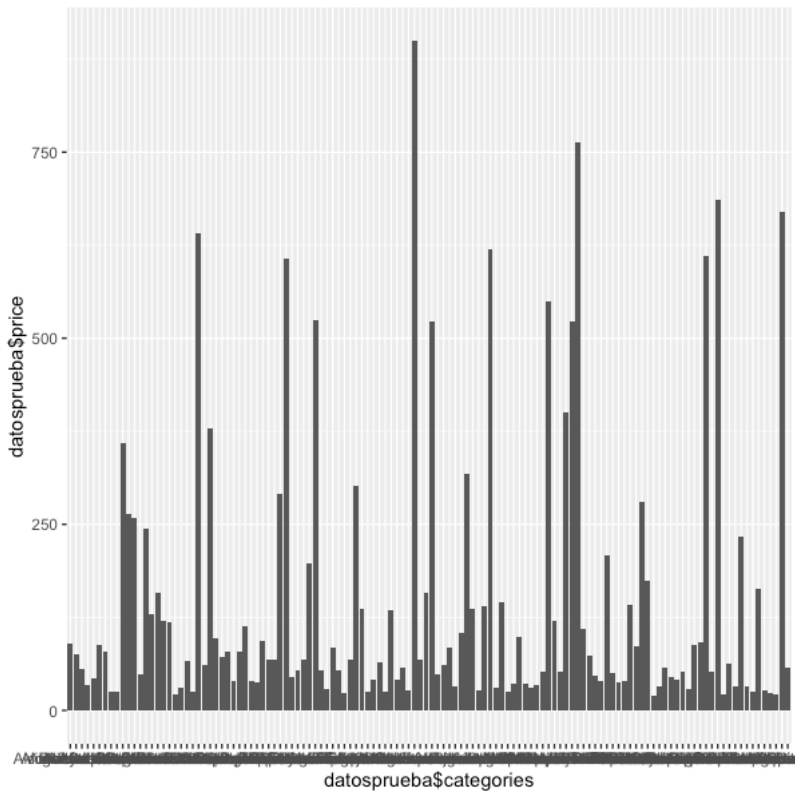


- Gráfico de Barras
- Gráfico de Líneas
- Box plot
- Pie Chart
- ...

Dataset Coto Digital 3/6/2016.

Reto 1. ¿Cuál es la categoría de productos cuyos precios son más altos? Muestra una comparativa mediante un gráfico. [0,25]

Heladeras Y Freezer → 899.00000



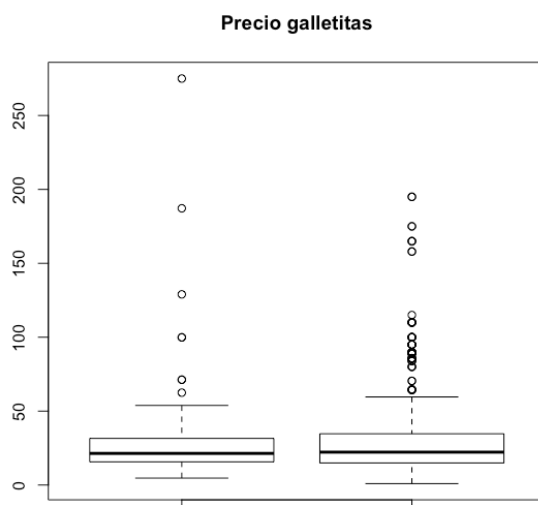
Reto 2. Vamos a estudiar el producto mermelada, ¿cuánto cuestan los 100g en cada referencia? [0,25]

Dataset Jumbo Online 3/6/2016

DataFrame : mermeladaArt\$precio_100gramos

Reto 3. ¿Podemos comparar cuánto valen las referencias de “galletitas” en ambos dataset? ¿Cuál sería la visualización adecuada? [0,5]

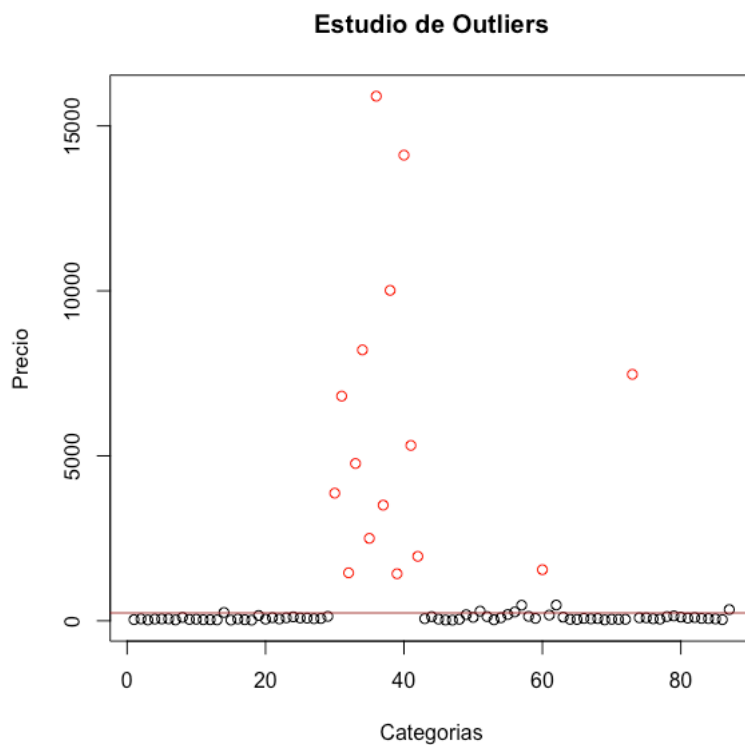
Si. Con un boxplot podemos diferenciar los precios de coto con los de Jumbo.



Reto 4. Estudio de outliers. ¿Qué visualización nos dará información sobre los outliers? Haz una visualización que muestre qué categorías de productos tienen precios muy diferentes al resto? [0,5]

Una forma de detectar outliers puede ser:

```
for (i in 1:ncol(datos)-1){
  columna = datos[,i]
  if (is.numeric(columna)){
    media = mean(columna)
    desviacion = sd(columna)
    datos$outlier = (datos$outlier | columna > (media + 3 * desviacion) | columna < (media - 3 * desviacion))
  }
}
```



Reto 5. Siguiendo con el punto anterior, ¿podríamos visualizar qué marcas son outliers con respecto al resto?. Hacerlo utilizando los cuantiles [0,5]

Dataset Veja Digital 3/6/2016

Reto 6. ¿Podemos comparar cuánto valen las referencias de “leche” en todos dataset? ¿Cuál sería la visualización adecuada? [0,5]

Reto 7. ¿Son suficientes los pasos anteriores para caracterizar al supermercado? ¿Qué más podríamos hacer? [0,5]