

## Tarea 4

### Descripción

Se le ha compartido el archivo *diamonds.csv* que es una colección de datos de diamantes con detalles de 1000 diamantes. Entre las variables incluidas están el quilate (*carat*) y el precio (*price*).

Construya la variable *diamonds* que contenga la lectura del CSV, luego explore el data frame con la función *str()*.

**Agregando geometrías:** Para cada geometría utilice el data frame *diamonds*.

#### 1. Dispersión

- 1.1. Construya un gráfico de dispersión colocando en la estética *carat* (eje x) versus *price* (eje y), luego devuelva el gráfico.
- 1.2. Edite la estética y mapee el color con la variable *clarity*.
- 1.3. Devuelva el gráfico con esta modificación.

#### 2. Histograma

- 2.1. Construya el tipo de geometría histograma para la variable *price*. Configure el parámetro *bins* con el valor 40 y devuelva el gráfico.
- 2.2. Agregue una escala logarítmica de base 10 y devuelva el gráfico para ver la diferencia.

#### 3. Boxplots

- 3.1. Construya un gráfico de cajas colocando en la estética la variable *cut* (eje x) versus *price* (eje y). Agregue una escala logarítmica por defecto y por último devuelva el gráfico.
- 3.2. Edite el color de cada caja con el parámetro *fill* dentro de la estética y asignando la variable *cut*. Por último devuelva el gráfico.

#### 4. Faceting

- 4.1. Filtre *diamonds* por la variable *price* que sea mayor a 1000 y por *cut* que contenga las categorías *Good*, *Ideal* y *Premium*. Luego asigne el resultado a la variable *filtered\_price*. Dado que *cut* debe tomar en cuenta tres categorías específicas puede hacer uso del pipe *%in%*, a continuación un ejemplo de su uso:

```
filter (var %in% c ("value1", "value2", "value3"))
```

- 4.2. Utilice la variable *filtered\_price* para construir un gráfico de barras, en la estética configure  $x = cut$ ,  $y = depth$ .
- 4.3. Por último aplique *faceting* para la variable *clarity* y devuelva la parcela de gráficos.
- 4.4. Edite la estética y configure el parámetro *fill* con la variable *clarity*. Luego devuelva la parcela de gráficos con esta modificación.

### Grouping and summarizing

1. Utilice *diamonds* para construir un agrupamiento por la variable *cut*, luego aplique *summarize* para encontrar la media de las variables  $x$ ,  $y$ ,  $z$  (los nombres de los valores sumariados serán *mediaX*, *mediaY*, *mediaZ*, respectivamente). Asigne el resultado a la variable *median\_dimension\_diamond*, por último imprima la variable.
2. Utilice *diamonds* para filtrar por la variable *cut* igual a *Premium*. Luego agrupe por la variable *clarity* y utilice *summarize* para encontrar mínimo y máximo de *price* (los nombres de los valores sumariados serán *minPrice* y *maxPrice*, respectivamente). Asigne el resultado a la variable *cut\_premium* e imprima el valor de la variable.
3. Construya un gráfico de barras y utilice *cut\_premium* como set de datos, configure  $x=clarity$ ,  $y=maxPrice$  y el color de relleno (*fill*) con la variable *clarity*. Por último devuelva el gráfico.

### Entrega

- Crear un notebook con el nombre **Homework-4.Rmd** y desarrollar cada uno de los incisos expuestos anteriormente.
- Cargar el notebook al aula virtual.
- Se envía a más tardar el **21-junio hasta las 23:59**