

Tarea - Transformaciones Estadísticas con ggplot

Jesus Mudarra Luján

2022-09-07

Ejercicio 1

¿Qué hace el parámetro `geom_col`? ¿En qué se diferencia de `geom_bar`?

`geom_bar` usa `stat_count` por defecto, es decir, cuenta el número de ocurrencias para cada categoría de la variable. Por otro lado, `geom_col` es como si se usara un `geom_bar(stat = "identity")`.

Ejercicio 2

La gran mayoría de geometrías y de stats vienen por parejas que siempre se utilizan en conjunto. Por ejemplo `geom_bar` con `stats_count`. Haz una pasada por la documentación y la chuleta de ggplot y establece una relación entre esas parejas de funciones. ¿Qué tienen todas en común?

Tienen en común que todas pertenecen a la familia de los diagramas de barras.

Ejercicio 3

¿Qué variables calcula la función `stat_smooth`? ¿Qué parámetros controlan su comportamiento?

Aesthetics:

- x
- y
- alpha
- colour
- fill
- group
- linetype
- size
- weight
- ymax
- ymin

Computed variables:

- y or x: predicted value
- ymin or xmin: lower pointwise confidence interval around the mean
- ymax or xmax: upper pointwise confidence interval around the mean
- se: standard error

Ejercicio 4

Cuando hemos pintado nuestro diagrama de barras con sus proporciones, necesitamos configurar el parámetro `group = 1`. ¿Por qué?

Para evitar que cada columna sume 1 y sea la suma de todas las variables del eje X las que sumen 1.

Ejercicio 5

¿Qué problema tienen los dos siguientes gráficos?

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, y = ..prop..))
```

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = color, y = ..prop..))
```

Como se ha dicho en el ejercicio anterior, en un `geom_bar`, siempre que se utilice `y = ..prop..` debe ir siempre acompañado de un `group = 1` para que la suma de las variables del eje X de igual a 1.