

SEMINARI 2. Comparacions i Distribucions

1. OBJECTIUS

La part 1 d'aquest seminari introdueix les geometries de *ggplot* que ens permeten visualitzar comparacions de dades categòriques (nominals i ordinals). La part 2, introdueix les geometries de *ggplot* que ens permeten visualitzar distribucions. A més, el seminari té com objectiu també, introduir algunes eines que ens permetran fer una mínima edició dels nostres gràfics.

2. PART 1. Comparacions mitjançant els diagrames de barres

Continuem amb el conjunt de dades *mtcars* del seminari 1 que conté informació de 32 cotxes. Com el conjunt de dades és petit, poseu `view(mtcars)` per veure el *dataframe* en forma de taula. La funció `view()` ens permet obtenir informació sobre el conjunt de dades de manera complementària a les dues maneres que vam veure l'altre dia: l'ajuda `?mtcars` o la funció mostrant l'estructura, `str(mtcars)`.

Si obriu R de nou, primer de tot recordeu que heu de tornar a carregar la llibreria *tidyverse*: `library(tidyverse)`.

EXERCICIS:

1.- Veient una sola gràfica, volem comparar el nombre de cotxes que hi ha de tres grups/nivells diferents segons el nombre de cilindres (4 cilindres, 6 cilindres i 8 cilindres). Feu un gràfic de barres on l'alçada de les barres sigui proporcional al nombre de cotxes de cada grup/nivell. Quina informació us dona la gràfica?

- a) Poseu la etiqueta als eixos utilitzant `xlab()` i `ylab()`
- b) Pinteu les barres segons el nombre de cilindres utilitzant `fill` dins d'`aes()`. Ens dona alguna informació extra pintar-les? Per què?
- c) Poseu títol i etiquetes a la llegenda que us ha sortit anteriorment per defecte
- d) Feu les barres més estretes especificant la `width` de les barres. Nota: `width` actua com argument de `geom_bar()`

2.- Feu un gràfic on pugueu veure la proporció de vehicles amb transmissió manual versus automàtica segons el nombre de cilindres. Què observeu?

Utilitzeu l'argument `dodge` de manera que cada tipus de transmissió aparegui en l'eix x, una al costat de l'altra d'acord amb el nombre de cilindres. Per fer això últim especificarem l'argument posició en `geom_bar` (`position="dodge"`). Observeu més coses?

Ara que hem vist com comparar algunes dades mitjançant diagrames de barres, veiem com mostrar les distribucions a la part 2 del seminari.

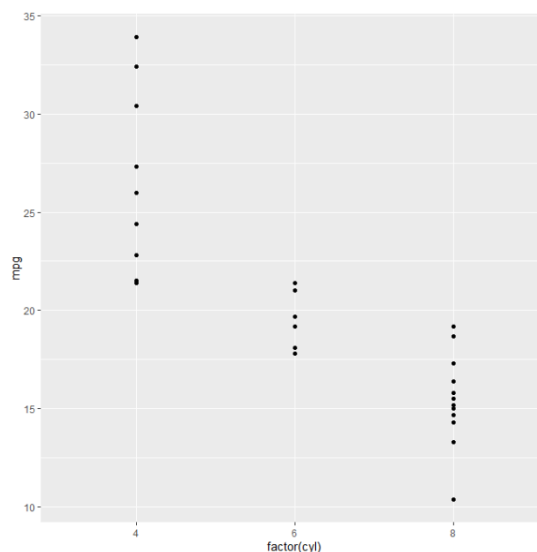
3. PART 2. Distribucions

EXERCICIS:

1.- Mostreu la distribució de la variable 'hp'. Quina informació en podeu extreure?

2.- Feu ús de `geom_boxplot()` per fer un gràfic en que l'eix de les x correspongui a la variable 'cyl' (cilindres) i l'eix y a la variable 'mpg' (milles de galó). Un cop tingueu el gràfic, compareu aquesta gràfica amb la resultant de l'exercici 2 del seminari 1 (Quina informació tenim aquí que no teníem abans?). A més:

- Poseu la llegenda dels eixos utilitzant `xlab()` i `ylab()`
- Gireu els eixos de coordenades, fent que els boxplots quedin en horitzontal utilitzant la comanda `coord_flip()`
- Compareu la gràfica resultant amb la gràfica de l'exercici 2 seminari 1.



Gràfica de l'exercici 2 seminari 1

d) Pinteu els boxplots de tres maneres diferents:

d1) Escollint un únic color amb `color`, `fill` i `alpha` per als tres grups:

```
geom_boxplot(color="red", fill="blue", alpha=0.2)
```

Fixeu-vos en el resultat de canviar els tres paràmetres `color`, `fill`, `alpha`. Què fa cadascun?

d2) De manera que el color es trobi en el contorn del boxplot per a cada grup (cotxes amb els mateixos cilindres)

d3) De manera que el color es trobi en l'interior del boxplot per a cada grup (cotxes amb els mateixos cilindres)

Els canvis fets en d2) i d3) ens aporten alguna informació nova?

NOTA: Per d2) i d3) al pintar diferenciant per grups recordeu que afegim la informació dins d'`aes()` com ja havíem fet en el seminari 1.

3.- Mostreu en un mateix gràfic un polígon de freqüències i un histograma del temps en quarts de milla (*qsec*). Afegiu títol i etiquetes als eixos. Poseu un color per cada geometria. Què es mostra? **NOTA:** Ajusteu els bins.

Un cop tingueu el gràfic afegiu la capa `theme_minimal` (), què us fa?

4.- Com depèn la distribució del temps en quarts de milla (*qsec*) segons la cilindrada? Feu una imatge que permeti visualitzar-ho. Imagineu que heu de posar aquesta gràfica en una web corporativa de tons verds, poseu-hi un color verd per tal que estèticament quedi més en el context de la pàgina web.