Workshop 2 – 2/11/2020

ggplot2 and data visualization

Exercise 1

From iris data:

1a) Using the variable 'Sepal.Length" create an histogram.

1b) Draw the histogram for the variable "Sepal.Length", with 50 blue bins, where the y-axis represents the densities. Add a density red line to the plot.

1c) Draw a scatterplot of Sepal.Length and Sepal.Width where color and shape depend on the Species

1d)Add a separate regression line for each group.

1e) Then overall a smooth line (method = “loess”)

1f) Draw a separate scatter plot with a regression line, one for each level of the variable Species

Exercise 2

- 2a) From “mpg” dataset draw a scatter plot for `displ` (x-axis) and `hwy` (y-axis)
- 2b) Modify the previous scatter plot in such a way that the colour depends on the class and the shape on the year
- 2c)display the same data conditionally on one categorical variable (here the class variable)- Hint: ?mapping
- 2d) Load the diamonds dataset. Draw a scatter plot of 'carat' ` (x-axis) and 'price’ (y-axis) where the colour depends on the variable 'cut', add also a smooth line (‘lm’) and display conditionally on the variable 'color’.

Exercise 3

Load the starwars dataset (dplyr::starwars).

- 3a) Which variable (column) has the highest number of missing values?
- 3b) How many humans are contained in starwars dataset? show them by gender.
- 3c) From which homeworld do the most individuals (rows) come from?

- 3d) Create a barplot of the gender distribution of the starwars Universe, set the title : "Gender distribution of the sw Universe". Make the colours of the columns depend on the gender, modify the colour using the command : scale_fill_manual.

- 3f) Draw the densities for the height variable of feminines and masculines only.

- 3g) Draw a segmented barplot for the variable 'sex'. The colors depend on the hair colours. Show the proportions [0,1] on the y-axis.