

Homework 1

Due date: Friday, September 16

Logistical details

Complete the following exercises and submit your assignment via gradescope (linked on the course webpage). To do this, take the following steps:

- Complete your assignment.
- Scan your assignment as a PDF document.
- Upload your PDF to gradescope.
- Specify which problems are on each page.

Problems to start after class Sept 12

Q1.

The `dplyr` R package provides a `starwars` data set containing data on 87 Star Wars characters. Load the data and necessary packages using the following code chunk:

```
library(ggformula)
library(mosaic)
data("starwars", package = "dplyr")
```

Explore the data by doing the following:

- a. Use the `dplyr::glimpse()` to get an overview of the data set.
- b. Pick a single numerical variable and make a histogram of it. Select a reasonable binwidth for it. Write a sentence summarizing what you see.

(A little bit of starter code is provided below, you can use it as a starting point and fill in the blanks.)

```
gf_histogram(~___, data = starwars, binwidth = ___)
```

- c. Calculate summary statistics for the variable you selected in part (c) using `favstats()`.
- d. Pick a numerical variable and a categorical variable and make a visualization (you pick the type!) to visualize the relationship between the two variables. Along with your code and output, provide an interpretation of the visualization.
- e. Pick two numerical variables and make a scatterplot to visualize the relationship between the two variables. Along with your code and output, provide an interpretation of the visualization.
- f. Pick a single categorical variable from the data set and make a bar plot of its distribution.
- g. Pick two categorical variables and make a visualization to visualize the relationship between the two variables. Along with your code and output, provide an interpretation of the visualization.