

Assignment 05

This homework will help you to practice and self assess your knowledge and ability around the course concepts. Use the answer key to assess your work.

1. Install the **nycflights13** package. Load the **tidyverse** & **nycflights13** packages.
2. View the following data sets:
 - A. **starwars** from **dplyr** (part of **tidyverse** package)
 - B. **diamonds** from **ggplot2** (part of **tidyverse** package)
 - C. **flights** data set from **nycflights13**

Look at the help files for these data sets in order to understand their variables.

Starwars Data Set

3. Nominal Comparison Relationship: Drop non-humans from the **starwars** data set. Also drop missing height observations. Then make a plot to compare character heights. Sort the characters from tallest to shortest (note you'll likely need to make a factor in order to generate this). Make sure the character names are readable.
4. Use this plot and map gender in a way that allows for comparison of heights within genders. Does this reveal anything unexpected?

Diamonds Data Set

5. One would expect that as diamonds weigh more they cost more. Make a plot to examine this. What patterns/trends do you notice?
6. What is the distribution of prices?

Flights Data Set

7. We would expect that flights that depart later will arrive late as well. Investigate this relationship (note that overplotting may be an issue; choose geoms and/or aesthetics to counter this issue). What pattern(s) do you notice?
8. For flights leaving the JFK airport, how does the median arrival delay change across months? What relationship(s)/trend(s)/pattern(s) do you notice?