

# Data Science 311

## Lab 2 (10 points)

*Due at 10pm on April 19, 2022*

**Read all of the instructions. Late work will not be accepted.**

### Overview

Now that we have the language to talk about, and the tools to create, effective visualizations, in this lab you will spend some effort producing some nice visualizations.

### Collaboration

For this lab, you are encouraged to spend the lab period working together with a partner. Together means synchronously and collaboratively: no divide and conquer. After the lab period ends, you will work independently and submit your own solution, though you may continue to collaborate the same partner if you wish. Your submission must acknowledge which person you worked with, if any, and for what parts of the lab (this should be included as a statement at the top of your notebook).

### Details

#### Data

The data you choose to visualize is mostly up to you – you are free to revisit any of the datasets we have worked with in class, or in Lab 1, or the datasets built into Seaborn (accessible via `sns.load_dataset`, discoverable via `sns.get_dataset_names()`). Note that, as described below, you will need to use at least three different datasets in all.

#### Tasks

In a single notebook titled `lab2.ipynb`, create the following plots. Use Seaborn unless otherwise noted, and spend some time to make each look as nice as you can. Not all plots can use the same dataset: use at least three different datasets in this lab.

1. Make a dot plot or a line plot.
2. Make a box and whiskers plot.
3. Make a scatter plot (or set of scatter plots).
4. Make a bar, column or pie chart.
5. Make a histogram.

For each of your plots, include:

- Any data processing code needed
- The code to produce the plot and plot itself
- A caption describing what the plot shows and what the reader should focus on

- A discussion of the design decisions you made when creating your plot

Your plots should be carefully designed to tell a specific story. This does not preclude you from making rich, data-dense plots (indeed, this is encouraged!), but the effect that motivated you to make the plot should be easy to see. Your design discussion should justify the choices you made in terms of the principles we talked about in class. You need not address every principle, but any that apply to your plot should be discussed.

Finally, there is one last plot to make:

6. Make the worst graph you can using the same data as one of the good graphs above. Break all the rules. Aim for something that still technically represents all of the right numbers while actually being totally misleading or unreadable. Prefer being misleading to merely unreadable.

## Submitting Your Work

Please zip your lone ipynb file (using the naming conventions stated above, where spelling, spacing and capitalization matter) and upload the zip via Canvas.

## Grading

Each plot will be graded based on the following:

- 50% based on the quality of the plot, including both correctness of the code to generate it and how well the result agrees with the six principles discussed in class. (For the 6th plot, your grade will be based on how well your plot *disagrees* with the six principles.)
- 50% based on the completeness of the information provided, including the clarity of the caption and discussion of design decisions.

## Acknowledgments

*This work is largely based upon a lab assignment created by Scott Wehrwein in Fall 2021.*