Homework 3: Data Visualization

100 points total [4% of your final grade]
Due: November 10, 2019 by 11:59pm
Delivery: Submit via ecampus

- *Goals of this homework:* In this homework, you will create a data visualization based on an exploration of a dataset we provide. Your visualization should capture Tufte's principles of graphical excellence: "complex ideas communicated with clarity, precision, and efficiency."
- *Submission Instructions:* You should create and submit two files: your image and a brief discussion. Your discussion file should be in plain text, named `your-uin_hw#.txt`. For example, my homework submission would be: `555001234_hw3.txt`. Your image should be in PNG, PDF, JPG, or other reasonable format. Submit these two files via ecampus.
- *Late submission policy:* For this homework, you may use up to two of your late days, meaning that no submissions will be accepted after November 12 at 11:59pm.

(100 points) Visualize This

We're going to take a look at a sample of data from American Community Survey 2010-2012 Public Use Microdata Series about college majors and basic earning information.

Data: Get the csv file on Piazza.

There are 21 columns and 174 rows in the file. 174 rows represent 174 different college majors, and the descriptions of the 21 columns are:

Header	Description
Rank	Rank by median earnings
Major_code	Major code, FO1DP in ACS PUMS
Major	Major description
Major_category	Category of major
Total	Total number of people with major
Sample_size	Sample size (unweighted) of full-time, year-round ONLY (used for earnings)
Men	Male graduates

Women	Female graduates
ShareWomen	Women as share of total
Employed	Number employed
Full_time	Employed 35 hours or more
Part_time	Employed less than 35 hours
Full_time_year_round	Employed at least 50 weeks and at least 35 hours
Unemployed	Number unemployed
Unemployment_rate	Unemployed / (Unemployed + Employed)
Median	Median earnings of full-time, year-round workers
P25th	25th percentile of earnings
P75th	75th percentile of earnings
College_jobs	Number with job requiring a college degree
Non_college_jobs	Number with job not requiring a college degree
Low_wage_jobs	Number in low-wage service jobs

Your goal is to (i) identify an interesting story or insight from this data; and (ii) carefully design a data visualization that communicates this story. You may assume that your audience is the general public.

You may clean, manipulate, and transform the data as you like. You are free to use data from a single file we provide or multiple files. The choice is yours. You may augment what we have provided with other sources if you like.

Ultimately, you should produce:

- A single image (e.g., PNG, PDF, JPG) that effectively communicates the data. You **must use a Python visualization toolkit** as your key tool, but you are welcome to augment with other tools if you like (e.g., Photoshop, MS Paint, GIMP). For Python viz tools, you may use matplotlib, plotly, seaborn, geoplotlib, bokeh, networkx, or whatever else you can find.
- A brief explanation describing your design. (a few paragraphs, 1/2 page or so).

Your explanation should justify the design choices you have made. Why did you choose the particular visualization type? What motivated your choice of size, color, and scale? Why are they appropriate for what you are trying to communicate? Additionally, your explanation should document what you are trying to communicate, the benefits of your choice of visualization, as well as any downsides to your viz (e.g., are certain connections downplayed due to your choices?).

Grading Breakdown

- 10 points: Interestingness and clarity of question you ask

- 20 points: Appropriateness of data, scale, and other viz design choices
- 10 points: Legends, context, and titles of your viz
- 20 points: Overall quality of your viz: in the extreme, every pixel serves a purpose
- 20 points: Design rationale (writeup)
- 20 points: Pros and cons of your design (writeup)