CSCI 444/544 – Data Visualization

ASSIGNMENT #: 1
DUE DATE: 1/28 Visualizations
1/31 Documents and Presentation
POINTS: 50
TOPIC: Bureau of Labor Statistics: Consumer Expenditures

Overview

This assignment introduces students to the process of creating compelling visuals based upon datasets consisting of ordered pairs. This project will also introduce students to an inquiry based method for doing projects that will be followed this semester. We will begin with scatter plots, linear fits, and time series. These are among the most common formats for data visuals, and you will need to master them. The data source for this assignment has volumes of data that lend themselves to these sorts of bivariate display. You are free to use more sophisticated means of data display, but make certain to offer examples of the basics before doing anything fancier. So, of the four visuals you prepare, make at least 2 bivariate plots with a clear x and y axis.

Data

Use the Consumer Expenditure data found on the website, https://www.bls.gov/cex/

Analysis

A simple way of demonstrating a trend, and testing the strength of a trend, is a linear fit. In this assignment you must fit some of the data to a linear model.

Assignment

Do the following:

- Form a hypothesis from the data. State the hypothesis as clearly as possible.
- Identify two visuals on the Internet, in books, magazines, or other media. One should be a good example of how a hypothesis can be supported with similar data (e.g. a time series plot that is conclusive, and especially well done). The other visual should be a uniquely bad visual that uses similar data, but is difficult or impossible to draw conclusions from. In a sentence: one visual for how to do it, and one for how not to do it. Obtain electronic copies of the visuals and submit them with the assignment. (I found several sites with data visualizations, here is one: https://www.visualcapitalist.com/money-spent-different-income-groups/)
- Write *one half to one* page about the visuals you found, where you found them, and why you believe they are consistent with lecture material and appropriate to your objectives.

- Produce three visuals of your own, from the data, that support your hypothesis. At least one visual must be a scattered plot with a best fit line.
- Write *one half to one* page detailing how your visuals are consistent with the examples, support the hypothesis, and are consistent with the material covered in lecture.
- Produce *one* visual that refutes your hypothesis and provide a paragraph explaining why.
- Summarize your findings in a few overhead slides and be prepared to defend them to the class.

Submission

The four figures you produce, and commentary on them, should be one document. Take care to see that high-quality reproductions of the graphics are turned in; watch the dpi and the embedding of font in whatever software you use.

A second document should be your examples of good and bad visuals, as well as commentary on them.

A third document should show the source code required for the analysis; in this case a least squares fit.

All documents including your data, should be placed into a directory titled <LASTNAME>_1 and then zip (or tar/gzip) that directory and upload it to the Moodle.

Presentation

Your presentation slides should be very brief (in most cases just two, your hypothesis and the updated visualization). Focus on what you changed from the critique, or why you did not make changes.

Critique – performed on 1/28/20, report on 1/31/20

Everyone will be responsible for critiquing another student's visualization. The critiquing student should answer the questions: 1) What were the strengths of the visualization in each category? 2) How could it have been improved? (You should not assign the visualization a number score.)