Please complete each problem and submit a PDF with your solutions. If you use code to solve the problems, you should include it in your submission.

Problem 1: Tidy Data

A. The Cigarette.csv file contains an "untidy" dataset with different variables related to cigarette sales from 1985-1995 in the 48 contiguous United States. There are three mistakes–fix them so that the dataset would be considered tidy. Your final dataset should have 528 rows and 9 columns.

B. Create a scatterplot comparing the number of packs sold per capita, `Packpc`, to the year, `Year`. Color each state a different color (do not worry about the clunky legend for this portion of your assignment). Connect each point with a line. Briefly describe the findings from your dataset.

R Solutions

Python Solutions

C. In the spirit of learning more about your dataset, identify the two states with the highest cigarette sales and the state with the lowest. Is there any reason that might explain why these are the highest and the lowest?

D. In your reading for this week, Edward Tufte talks about the time series plot as an innovation. Do you think he would say this is a good use of a time series plot?

Problem 2:

For this assignment, you will create a static visualization of the population histograms from the 1900 and 2000 US. Census. These came from the U.S. Census Bureau via IPUMS and give numbers of people enumerated in 5 year bins, age 0-4, 5-9, 10-14... up to 90+, and reported sex (1 for male, 2 for female). The data can be found in the census.csv on Canvas.

It looks like this (in R, similar in Python):

```
##   Sex Year Age   People
## 1   1 1900   0  4619544
## 2   1 2000   0  9735380
## 3   1 1900   5  4465783
## 4   1 2000   5 10552146
## 5   1 1900  10  4057669
## 6   1 2000  10 10563233
```

You can read about this data, and some light commentary about age histograms in the Gilded Age in 1900 Census: Volume II. Population, Part 2.

Examine the data, ask a question of the data, and design a visualization that answers the question. Include one paragraph to go with the visualization, like a figure caption, and include one to three paragraphs explaining your question and your graphical design, and any code you used to produce the visualization.

You do not have to use any specific tools to produce the visualization (you could even draw it by hand) but you need to find an interesting question and answer it effectively.