

Data 227 - Autumn 2023 HW 1
Data Visualization and Communication - Trimble
Due: Friday October 6, 2023 11:59pm

Homework 1, Population pyramid + longitudinal cigarette sales

Upload your responses to Canvas as PDF or HTML files; please put your answers to both questions in a single file. Attach the most relevant pieces of your code in any convenient format, but if your answers are not in PDF or HTML we will send your homework back ungraded.

1. 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. (This dataset was published in the textbook Stock, James H. and Mark W. Watson (2003) Introduction to Econometrics, Addison-Wesley Educational Publishers, and has been distributed as part of data-containing R packages for years.

B. Create a scatterplot comparing the number of packs sold per capita, Packpc, to the year, Year. Color each state the same color and put labels sufficient to identify a few of the states with the most extreme per-capita cigarette sales. Overlapping, illegible labels are ok for this assignment.
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 <https://www.ipums.org/> 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).

data/census1900-2000.csv on canvas

<https://canvas.uchicago.edu/courses/38293/files/folder/data>

<https://canvas.uchicago.edu/files/6084647/download>

It looks like this:

```
Sex,Year,Age,People
1,1900,0,4619544
1,2000,0,9735380
1,1900,5,4465783
1,2000,5,10552146
...
```

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, Statistics of the Population

<http://www2.census.gov/library/publications/decennial/1900/volume-2/volume-2-p2.pdf>

Examine the data, ask a question of the data, and design a visualization that answers the question.

Include a figure caption and at most one paragraph describing your findings, and any code you used to produce the visualization.

You do not have to use any specific tools to produce the visualization but you should to find an interesting question and answer it effectively.