Pre-class assignment #10

CMSE 402, Data Visualization Principles and Techniques
Spring 2022

This assignment is due the evening of Wednesday, February 9, 2022. Turn in the assignment via the GitHub classroom by committing your assignment to the repository and pushing it. **All writeups must be in PDF or plain-text format (which includes Markdown or RTF)** – Word, ODF, etc. will not be accepted.

Purpose: The purpose of this assignment is for you to learn about time series datasets, and experiment with visualizing them.

Instructions: This pre-class assigment comes in two parts: (1) reading about time series datasets and the types of visualizations used to show them, and (2) experimenting with a relatively simple example dataset.

Part 1: Readings. Do the following readings:

1. **Required:** The Truthful Art: Data, Charts, and Maps for Communication, by Alberto Cairo – Chapter 8 ("Revealing Change")
2. **Required:** Look at the Duke Library taxonomy of data visualization types and examine the "Temporal" section to see what types of visualizations they use.
3. **Required:** Look at the The Data Visualization Catalogue and search for the various visualizations that show how data is visualized over time (look in "search by function").

In the writeup, synthesize Cairo's chapter on revealing changes in datasets. What does he claim are the most important considerations when showing this type of data? How can they be commonly used to mislead an audience? (And thus how do you avoid that?) From the Duke taxonomy and Data Viz Catalogue, pick one example of a visualization from each that seems particularly specialized, and describe the circumstances (i.e., type of data and message that you want to communicate) where you might use it.

Part 2: Analysis. The repository for this pre-class assignment contains a file showing hourly temperature readings for a variety of cities around the world over several years, from 2012 through 2017 (extracted from this Kaggle dataset). Plot the temperature reading for Detroit for this time using a standard matplotlib line plot. In addition, calculate the average temperature for each month, as well as the minimum and maximum temperature for that month, and plot it **on top** of the raw data. What does this reveal, if anything, that the raw data does not?

Handing in the assignment: Turn in all this assignment via the GitHub classroom repository where you found these instructions. You can do this by committing your assignment and pushing it – i.e., in the directory that contains your assignment repository type "`git add (name of file containing writeup)`", then "`git commit (filename)`", and then "`git push`". The last step is critical, since that pushes your changes to GitHub where I can see and grade them!