Exercises and Guided Projects

DATA VISUALIZATION AND DASHBOARDS

Between Sessions

Session 1 to Session 2

- complete the exercises of session 1
- download the datasets from the website

Session 2 to Session 3

- complete the exercises of session 2

After Session 4

- complete the exercises of session 4
- attempt the guided projects

Session 3 to Session 4

- complete the exercises of session 3
- install R / RStudio (Posit)
- read Programming Primer (sections 1 – 4)
- read Basic Visualizations in R
- install the following R packages: tidyverse, ggplot2, dplyr

Guided Project I

Identify a scenario for which a dashboard could prove useful.

Determine specific questions that the dashboard could help answer or insights that it could provide. Identify data sources and data elements that could be fed into your dashboard.

Design a display (with pen and paper) with mock charts. What are the strengths and limitations of your dashboard? Is it functional? Elegant?

Use some of the storytelling concepts described in the course, as required.

Guided Project II

This project uses the Gapminder Tools (there is also an offline version).

1. At what point in the data science workflow do you think that visualizations of this nature could be useful?
2. What are the ways in which observations could be anomalous? Have you found any such anomalies? Do you have explanations for them? In particular, consider the case of South Africa in 2012, which appears to be a clear outlier. Follow the path of the South African bubble from 1975 to 2020, in relation to the general pattern. Does the apartheid/income inequity explanation suggested in the text (DUDADS) still make sense?
3. Pick 2+ "definitive" visualizations (methods, variables, etc.) other than the default configuration. What are some important insights?
4. How would you describe the insights of step 3 without resorting to visual vocabulary?
5. Can you think of ways in which the data of interest to you in your day-to-day activities could benefit from the same treatment? What situations could you explore in such a scenario? How would that help your team better understand the system under consideration?

Guided Project III

Select a dataset from the list below (or any other set of interest to you):

- GlobalCitiesPBI.csv
- 2016collisionsfinal.csv
- polls_us_election_2016.csv
- HR_2016_Census_simple.xlsx
- custdata.tsv

For your dataset(s):

1. Create a data dictionary for this dataset. Establish a list of variables that you think are crucial to a good understanding of the dataset. Justify your choices.
3. Create (at least) 5 bivariate/univariate visualizations that can help you understand the dataset.
4. Produce (at least) 3 "definitive" visualizations for the dataset. Use the principles discussed in class (including documentation, legends, annotations, Multiple I's, etc.). Emphasis should be placed on content AND on presentation (suggestions: consider creating a reasonably high number of charts using a random selection of a random number of variables in order to minimize the odds of missing out on useful information). Use storytelling concepts where appropriate.