SocialComplexityLab added assignment 2 History

1 contributor

Raw Blame

106 lines (106 sloc) 5.12 KB

Assignment 2.

Formalia:

Please read the assignment overview page (https://github.com/suneman/socialdata2021/wiki/Assignment-1-and-2) carefully before proceeding. This page contains information about formatting (including formats etc), group sizes, and many other aspects of handing in the assignment.

If you fail to follow these simple instructions, it will negatively impact your grade!

Due date and time: The assignment is due on Monday April 5th, 2021 at 23:55. Hand in your files via `http://peergrade.io` (http://peergrade.io/).

Peergrading date and time: *Remember that after handing in you have a week to evaluate a few assignments written by other members of the class*. Thus, the peer evaluations are due on Monday April 12th, 2021 at 23:55.

Part 1: Questions to text and lectures.

A) Please answer my questions to the Segal and Heer paper we read during lecture 7 and 8.

- What is the *Oxford English Dictionary's* defintion of a narrative?
- What is your favorite visualization among the examples in section 3? Explain why in a few words.
- What's the point of Figure 7?
- Use Figure 7 to find the most common design choice within each category for the Visual narrative and Narrative structure (the categories within visual narrative are 'visual structuring', 'highlighting', etc).
- Check out Figure 8 and section 4.3. What is your favorite genre of narrative visualization? Why? What is your least favorite genre? Why?

B) Also please answer the questions to my talk on explanatory data visualization (https://www.youtube.com/watch?v=yHKYMGwefso)

- What are the three key elements to keep in mind when you design an explanatory visualization?
- In the video I talk about (1) *overview first*, (2) *zoom and filter*, (3) *details on demand*.
  - Go online and find a visualization that follows these principles (don't use one from the video).
  - Explain how it does achieves (1)-(3). It might be useful to use screenshots to illustrate your explanation.
- Explain in your own words: How is explanatory data analysis different from exploratory data analysis?

Part 2: Random forest and weather

The aim here is to recreate the work you did in Part 1-3 of the Week 7 lecture. I've phrased things differently relative to the exercise to make the purpose more clear.

Part 2A: Random forest binary classification.

- Using the and instructions and material from Week 7, build a *random forest* classifier to distinguish between two types (you choose) of crime using on spatio-temporal (where/when) features of data describing the two crimes. When you're done, you should be able to give the classifier a place and a time, and it should tell you which of the two types of crime happened there.
    - Explain about your choices for training/test data, features, and encoding. (You decide how to present your results, but here are some example topics to consider: Did you balance the training data? What are the pros/cons of balancing? Do you think your model is overfitting? Did you choose to do cross-validation? Which specific features