Telling Stories with Data Final Project

PUBHLTH 460: Spring, 2022
taught by Prof. Nicholas Reich

Overview of the project assignment

For the project you will create, in small groups, a thorough analysis of one of several large datasets recommended by the instructor. The goal is to tell a story using data about the topic of your dataset. The project will have components completed by the group and by each individual student.

Each group will present their results during one of the last two days of class. In addition the group will hand in the following items on Moodle by 8pm Eastern time on Wednesday, May 4, 2022:

- A ten-slide presentation summarizing the key findings.
- A 500 word technical write-up describing the details of your group's analysis.

Additionally, each individual student will hand in a separate detailed write-up that describes the analyses that they contributed to their group analysis (see details below).

The group portion of the project will be worth 60% and the individual part 40% of the final grade.

Choices of a primary dataset and motivating question

- Data on historical causes of death from the CDC Wonder database
    - Use your analytical and data visualization skills to recommend measures that could take steps toward resolving or reducing the national opioid crisis.
    - Use your analytical and data visualization skills to analyze trends in deaths due to motor vehicles and recommend measures that could reduce the impact of these accidents on mortality in the US.
- COVID-19 data from COVIDcast
    - Use your analytical and data visualization skills to recommend measures that could reduce the impact of the COVID-19 pandemic in the US.
    - Use your analytical and data visualization skills to identify community-level risk-factors for high case, hospitalization, and death rates in the US.
- If you have a particular dataset that your group is interested in, you may petition the instructor to allow you to use it. In general, datasets should be associated with human health in some way and should arise from a real-world setting.

Teams are encouraged to use other external data sources to supplement their primary analyses.

Guidelines for the group write-up

Overall outline: Your team will create a story that contains a compelling central narrative. Your job is to convince the reader that your story is important and interesting. You must convey an understanding of the broader societal context of the data you are analyzing. As we have seen throughout the semester, using appropriate and simple visualization techniques is often the key to telling an effective story. While a regression or other modeling analysis may be a valuable supplement to your story, it is not a required element of the project.

Working as a group: Your group should assemble an outline of the key elements of the story that you want to tell and which team member will be responsible for each element. The general idea is that each individual element should focus on one key observation or insight about the dataset. Together, the individual elements should complement each other and tell a coherent story about your dataset. Elements could be data visualizations, regression analyses, integrations with other datasets, model evaluation, or some other quantitative piece of the story.

Evaluation: As a team, you are encouraged to pay close attention to the evaluation rubric provided for the project (see below).

Guidelines for the individual write-up

Each member of the group will serve as the lead on one or more elements that tell a story about the assigned dataset. In addition to incorporating these elements into the final group deliverables, each individual student will produce a separate write-up (including up to 4 total tables and figures) about their individual analyses. These individual analyses should provide more technical detail about the analyses performed as well as giving a brief introduction and conclusion to the analysis. Each write-up should stand on its own, providing tables and figures as necessary.

The individual data analysis write-ups will be due Friday May 6th at 8pm ET, to be handed in on Moodle. Individual analyses should be handed in as HTML files knitted using RMarkdown. Code-folding should be activated, so that no code is displayed by default, but code chunks could be opened and inspected by interested readers. Any figures and tables resulting from your analyses should be dynamically created by your code, not manually inserted from an external source.

Grading

Your project grade makes up 30% of your final grade for the class. The grading rubric for the group portion of the project will be evaluated based on the Rubric (see below).

Participation: To evaluate individuals' contributions to the group, I will be using the following approach to evaluate each of your contributions to the project. After you submit your group project, each student will be given 100 points to allocate among your teammates (excluding yourself). The more points you give to a teammate, the more you are indicating they contributed to the project. You cannot allocate the same number of points for any two team members. I reserve the right to intervene to correct gross imbalances in allocations if necessary. The number of points that you receive from your teammates will be summed, divided by 100, and then used as a multiplier on the final grade for the 60-point group component of the project.

Example: Your group receives 50/60 points for the "final product produced by the group". You have three teammates who give you scores of 35, 40 and 30, respectively. Therefore, you receive a total of 105 points from your teammates. So your final "group" grade is scaled up by 5%, as follows: (50/60) * (105/100) = 0.875 = 52.5/60. You then receive a 35/40 on the individual portion of the assignment, so your final project grade would be an 87.5/100.

Grading Rubric: Individual component (40 points overall)

And two main criteria on which your individual components will be evaluated (40 total points):

- Overall quality of analysis (25 points)
    - Rubric:
        - correct implementation and interpretation of method(s) used,
        - appropriate use of equations to show what methods/models have been used (if applicable),
        - appropriate use of graphics/tables to support central results,
        - succinct summary of key results.
- Clarity and presentation (15 points)
    - Rubric:
        - clear statement/summary of goals and central results,
        - clear and accurate description of methods/models used,
        - use of figures rather than text to illustrate central ideas,
        - figures dynamically generated within the RMarkdown file,
        - figure limit adhered to.

Grading Rubric: Group component (60 points overall)

There are two main criteria on which your group project will be evaluated (60 total points):

- Narrative and organization (25 points)
- Evidence, methods, and graphics (35 points)

For judging each criteria there are point ranges for Poor, Fair, and Exceptional.

These criteria have been adapted from the 2018 Public Health Data Challenge Judging Criteria document.

Narrative and organization (25 points)

Your presentation should provide a narrative that engages the audience and holds their interest. This is done through constructing a strong theme that is driven with compelling facts and a connection to the overarching context of your datasest. Your presentation of the materials should be organized in a logical progression for the audience to follow. Additionally it will be important to use terminology and phrasing that are clearly understandable. Your visual materials used in the presentation should be well organized so that an audience could easily follow along. Your points for this section will be based the work in your presentation.

Point Range	1-8 Poor	9-17 Fair	18-25 Exceptional
Attributes	Poor Strength of Story - Did not convince audience of importance/impact -Presentation disconnected with intended audience - Does not display enthusiasm for their project Poor Progression - Difficult to follow train of thought - Results presented out of order Poor Lexicon - Use lots of unclear jargon, phrases or terms that would be unfamiliar to audience Clutter and Complication - Huge volume of words - Poorly organized items - Visual does not match presentation - Many typos	Fair Strength of Story - convinced audience of importance/impact, but could be stronger -Presentation showed moderate connection with intended audience - Displays moderate enthusiasm for their project Fair Progression - Fairly clear train of thought - Results presented in reasonable order Fair Lexicon - Use little unclear jargon, phrases or terms that would be unfamiliar to audience Little Clutter and Complication - reasonable volume of words - organized items - Visual matches presentation - Few typos	Exceptional Strength of Story - convinced audience of importance/impact - “wow factor” -Presentation well connected with intended audience - Clear communication Interesting - Displays strong enthusiasm for their project Great Progression - Intuitive order to presentation materials - Covered all important topics Great Lexicon - Use no jargon that would be unfamiliar to audience - clear/concise language used. No Clutter and Complication - Concise/clear wording - Well organized items - Visuals complement presentation - Few to no typos

Evidence, Methods, and Graphics (35 points)

The group presentation and overall narrative must be supported by the data through properly conducted, in-depth analysis and carry objectively correct interpretations of evidence. The score for these criteria will be based on the technical write-up and your presentation.

Point Range	1-15 Poor	16-27 Fair	28-35 Exceptional
Attributes	Poor Evidence - Statements not backed by data - Statements refuted by data Poor Depth - Only very simple data attributes plotted Poor Methods -inadequate/superficial or improper/overly-simplistic methods used - Incorrect interpretations of modeling/graphical items Poor Impact - Data visualization did not contribute in a meaningful way to the presentation Poor Aesthetics - Use continuous graphical aesthetic to display categorical data element, or vice-versa - Distractingly poor graphical Choices Poor Scaling - numeric scales improper for the context of the data	Fair Evidence - Questionable statements about the data - Some statements not supported by data Fair Depth - Relatively simplistic data relationships visualized Fair Methods - questionable methods used - Lacks sophistication - Some incorrect interpretations of modeling/graphical items Fair Impact - Data visualizations contribute moderately to the presentation Fair Aesthetics - Generally avoids using continuous graphical aesthetic to display categorical data element, or vice-versa Fair Scaling - numeric scales proper for the context of the data	Great Evidence - Statements backed by data - Legitimate findings Great Depth - Complex analysis and data relationships visualized Great Methods - Appropriate/sophisticated methods used properly - Correct interpretations of modeling/graphical items Great Impact - Data visualizations contribute in a major way to the presentation Great Aesthetics - Appropriately uses graphical Aesthetics Great Scaling - numeric scales proper for the context of the data