MICA

MVIZ5303
Statistics and Analytics

Final Projects

What is being assessed?

- Your ability to prepare and present a summary of your data that involves
  - Preparing a data set
  - Statistical analysis
  - Creation & Presentation of Visual Representations
- Effectiveness of communicating the essence of the data
- Plan and manage time

Presentation of the Final Project

- For your final project you will give a short (~15mins) presentation to the class about a data source.

- Broken into several parts:
  - Data Choice
    - What? Where? Why?
  - Data Cleaning
    - What you did to get & prepare the data
  - Exploratory Summary
    - Basic summary statistics
  - Presentation
    - "Tell a Story"
  - Critiques

Requirements: Data Choice

- You have your choice of data file
- Suggestions
  - Do not wait until the last minute to start looking
  - This is not a major life decision, find something that works and go with it
  - If you have a question about something you find, ask.
- Expectations
  - Size:
    - ~tens to hundred rows
    - ~several columns
    - You don't have to use all the sheets (.xlsx) or rows
    - Ideally it will involve more than one file, so that you have to merge or join
  - Presentations should provide a discussion of the entire original data file(s),how you gathered and processed the data, and a statistical summary of the selected records (rows) and measures (columns) you analyzed

- http://libguides.mica.edu/infovis -> Find Data

Submission Details: Data Choice

- PDF
  - How and where you found your data set
  - How you got it, and the format
  - Why you chose it
  - Brief written description of the data set
    - Subject matter, #rows, #columns, etc…

I need access to your data

- In both cases, include the resulting data file with your PDF in a single .ZIP file
  - NOTE: If the file is too large, please contact me so that we can make some arrangement

- Case-1: Link to Simple Data Source
  - Include a link in the PDF to the data source. This link should allow for a direct download of the data file you intend to use.
- Case-2; Selections and Options for Data Source
  - Some data sources are available after a number of selections or options have been made. In this case, the PDF should contain a list of these options

Submission Details: Exploratory Data Analysis

- Same format as other HW assignments
  - A .ZIP file with multiple files contained within a directory:
    - .RMD file
    - .HTML file which results from 'knitting'
    - any data files or R data objects you saved
      - I need to be able to run the .RMD
        - Be careful, this may involve accessing / downloading files
  - Proper naming convention, and good organization of files

- Expectations
  - Basic summaries of the files
    - Sizes, number of rows & columns
    - Types of data in columns
      - Numeric, categorical, etc..
    - Connections between files
      - How can information be combined

Submission Details: Presentation

- .RMD/.HTML and/or PDF
  - Your presentation
    - HTML & Markdown and/or PowerPoint is suggested
      - Minimal animations
    - Designing a large page, e.g. Adobe Illustrator is difficult to present
    - Ask if you have other ideas
  - “Tell a Story”
    - Why this data? Where did find it? How did you get it?
    - What did you do with it?
    - What are some interesting things about it?

Presentation Details

- Class will be split into 2-3 groups
- Each person will present to their group
- All will be recoded and posted
- Plan on ~15 mins max
  12/group*15mins = 3 hrs or 8/group*14mins = 2hrs
  - We will try passing control to you
    - This is often fraught with obstacles such as audio and sharing of your desktop
    - Back-up will be for host to keep control, and you direct
      - E.g. you ask us to advance

Submission Details: Critique

- What is being assessed
  - Your ability to
    - Evaluate and critique
    - Give and receive constructive comments
      - Why you enjoyed specific parts
      - How specific parts can be improved

- Submission
  - Your critique of two other presentations (randomly assigned)
    - Submit assignment
    - Share via course forum

Ideally: I can take your submitted .R or .RMD and re-create all the artifacts you use in your presentation

See Some Examples:
https://www.kaggle.com/headsortails/tidy-titarnic
https://www.kaggle.com/mrisdal/exploring-survival-on-the-titanic

Any questions…. ?