

Final Project

Draft (subject to minor negotiations)

Idea

This summary may not exactly describe your process in a linear fashion. Welcome detours, distractions, or even a few dead ends. If you start soon, you have time to abandon ideas, get back on track, and still finish at a leisurely pace even *if* things go “wrong”.

You should find 1 to 3 publicly-available data sets (not used directly during class) related to a topic or issue, potentially of personal interest or curiosity. Explore the data using relevant R commands to produce summaries, tables, or other helpful outputs and generate 5 (or more) exploratory visualizations. Create 2 (or more) carefully-annotated, related, honest, and accessible data visualizations. Using software of your choice or physical media (e.g., printed paper) make a collage or “infographic” that represents your findings.

Other elements will probably include a

- descriptive title or phrase omit personal information unless it is relevant (e.g., name, date, semester),
- *brief* text statement,
- small summary table, or
- content stemming from your own ideas.

With respect to resources the only requirements are that you - use R/Rstudio, - do not use Artificial Intelligence in coding and design, and - cite and describe any special functions or packages that *you* used to *create* the data visualizations.

Considerations

The infographics will be shared during the final exam period during a presentation and “defense” of the data visualizations contained on the infographic in a *short* presentation. Because of this, it is important that you choose a topic that you are comfortable discussing with others. This is not meant to constrain available topics or sources of data, just a reminder that you will publicly share your results with others from class. You might even pick a data sets that you understand but *do not* care about, so that there are no personal feelings attached.

Presentations made during the final period may feature a mix of live and prerecorded presentations. This could occur on a voluntary basis or for all (e.g., a 2-minute live summary and a 4-5 minute recording). The decision will depend on the number of presentations and the likelihood of fairly hearing all live presentations given the number.

Specific requirements

Prior to the start of the final you should submit a zipped directory¹ that contains appropriately-named files of the following formats:

¹The name of the directory (and .zip file) should be of the form last-first-final. *Before* uploading your .zip file, *it is your responsibility* to ensure that it unzips to a directory with the appropriate name and all contents are appropriately named. Do not zip the directory last-first-final and *then* change the name of the file. Change the name of the folder before zipping.

- .csv files for all raw data (if your data set is excessively large, use your judgement and include a representative subset)
- .RProj and .qmd files for your project session and analysis
- an output file showing your work and documentation (either a .pdf or a .html *printed to* .pdf)
- your one-page infographic – it may be produced entirely as a digital file (e.g., data visualizations pasted into a document or slide) or made as a manual collage and scanned (you have access to very nice equipment in the library, or in a pinch I can do a favor)
- additional files as required

Additional files as required

Graduate students should submit a typed, approximately one-page reflection (as a .pdf) describing your process. You can address any of the following:

- challenges locating or manipulating raw data
- changes to your original plan in response to challenges
- what from your work you are most proud of
- what from your work you wish you could improve
- anything else you would like to share about this process or your results

This is *optional, but encouraged* for undergraduate students.

Submission

If appropriate, list *both* names in alphabetical order by last name in the `author` field of the .qmd header. Each partner should submit a full copy to D2L with their own name in the `last-first` positions of all files. Otherwise all contents *should* be identical. If relevant, self- and peer-evaluations will be submitted separately.

Tips

- If your data comes from a public dashboard, go beyond reproducing graphs that the dashboard provides, instead *create something new* instead.
- Once you have your data, consider archiving it at <https://www.archive.org>.
- Carefully consider the type of graph with respect to the data you are using.
- Do not use the same type of graph for the same type of data more than once, unless you have a good reason and your exploration contains more than 5 meaningful graphs.
- If there is something you would like to study but are unable to find data, email me as quickly as possible, but not before you have given yourself a chance to succeed. Be patient and open-minded in your original searches, you never know what you may discover.
- Ask for help early and often. Drop by during office hours or by chance. Schedule a visit or a Zoom call.
- Depending on what format you set as the output of your .qmd file, you may want to over-ride the default of automatically deleting image files when the .pdf is created. Alternatively, you could generate .html output and print that to .pdf as a backup.