Peer Group Project 1: Shiny App

Presentations + Main Deliverables due Thursday April 7 in class

In collaboration with your Peer Group, you will create an interactive Shiny application that explores a topic of interest to your group.

This project will focus on data wrangling, interactive data visualization, and effective communication. The question(s) you address need not be complex, and the answers might involve only basic summary statistics. However, the dataset(s) you work with should provide a data ingestion and/or wrangling challenge. That is, tidying up the data and getting it into the format needed to produce even simple statistics of interest may be complex. The interactivity of the Shiny application should enhance the message you're trying to convey.

This project is deliberately open-ended to allow your group to explore your creativity and interests. There are only three main rules that must be followed:

1. *Your project must be centered around data.* Preferably, you will work with large, complex and/or messy data. Alternatively, you may work with a dataset that isn't very messy, but is challenging to obtain (e.g., a challenging web scraping task).
2. *You must create an interactive Shiny application.* The interactivity of the visualizations should enhance the message you're trying to convey.
3. *Each group member must create at least one component of the Shiny application.*

Learning objectives

The learning objectives of the Shiny App project are to demonstrate your ability to:

1. identify a set of questions that can be addressed with data available to you;
2. wrangle a large, messy dataset (including gathering, reshaping, and cleaning the data) into a format necessary to answer the question(s) at hand;
3. deploy interactive data visualizations using Shiny;
4. effectively communicate results via visualizations and oral presentation;
5. effectively collaborate with your peers and identify the value in teamwork; and
6. demonstrate awareness of ethical considerations related to data acquisition, management, and communication.

Components

Initial proposal (Friday, March 11)

The initial proposal should be submitted to me via email following the provided template, compiled to .pdf. All group members should be cc-ed on the email.

The proposal should contain the following content:

- Group name: This will be used as the name of your group repo for both group projects.

- Title: The tentative title of your project
- Purpose: Describe the general topic/phenomena you want to explore, as well as some questions that you hope to address.
- Data: Identify one or more data sources that could be used in the project. What form is the data in (downloadable csv file? needs to be scraped from web?)? What do you imagine will be challenging about ingesting and/or wrangling the data?
- Shiny app: Describe some visualizations, tables, and/or other components you envision including in your Shiny app. What will the interactive components be?

Feel free to include more than one direction if you would like feedback on different ideas you have. Part of the process here is learning how to refine questions, and how to evaluate whether you'll actually be able to answer the question you set out to answer. It is expected that these proposals will be revised after the break.

Note: I will construct group project repos after receiving initial proposals, from which you will work for the rest of the semester on the group projects. It will become very important to check what repo you are working in as you complete various tasks - the class repo, your personal class repo, or the group repo.

Revised proposal (Thursday, March 24)

At this point, you should have landed on a finalized plan for what data you'll be using and details about the vision for your Shiny app. The revised proposal should be uploaded as a .pdf in your group Github repo, should be in the same format as the initial proposal (given above), and should incorporate any feedback received on the initial proposal.

Shiny tutorial - Individual (Wednesday, March 30)

To help build your shiny skills, everyone is required to work through the online shiny tutorial video referenced earlier in your Prep 4 assignment. No work is required to be submitted in regards to this. However, working through the tutorial will aid you in the creation of your shiny app greatly.

Relevant files, scripts, and instructions for the tutorial designed to assist you are provided in a separate file / folder for your convenience.

You may start working through the tutorial at any time, following the suggested breakdown in the separate instructions for it, or doing it all in one sitting.

Wednesday, March 30th is the deadline for completing the tutorial, as the data set for the project is submitted soon after, and your focus should be entirely on the app by the end of that week. Completing the tutorial earlier (perhaps you want to consider it your Prep for the week when there isn't one!) could be useful.

Wrangled dataset (Thursday, March 31)

The (reproducible!) code to create the dataset(s) should be saved in an R or Rmd file within your group repository and named "data-wrangling." I will be running the code, so make sure your repository is organized (so I can easily find the "data-wrangling" file) and be sure your code is reproducible, readable, organized logically, and documented with informative comments.

At this stage, your data should be in the format needed to create the visualizations and summaries you'll present in the Shiny app. You want to have little or no wrangling code within your Shiny app program, so it's best to save your wrangled datasets as permanent files that you can load in at the top of the Shiny app program.

You may have more than one wrangled dataset! For instance, if there are some outputs in your Shiny app that require the data to be in long format and other outputs in your Shiny app that require the data to be in wide format, then you'll want to create both a long format dataset and a wide format dataset in your "data-wrangling" file.

Include a note at the start of the data wrangling file to indicate each person's contribution to the wrangling process.

A note on reproducibility

A common error you may encounter (and may have already encountered this semester) is having an object saved in your local environment that is not defined in the code. To help prevent this, I strongly recommend that you should never save the ".RData" file when prompted (the workspace), and you should occasionally clear your environment and re-run your code.

You can change your workspace settings in *Tools* → *Global Options..* → *General*. Make sure the box next to "Restore .RData into workspace at startup" is *unchecked* and "Save workspace to .RData on exit:" is set to "*Never*". You may want to adjust the options for the R history file as well. If all your code is in the .Rmd, you won't need the history file.

Shiny application (Thursday, April 7)

Each team member needs to take the lead on at least one component of the Shiny app. One way to make this clear is to utilize the tab or navigation list format and have each member take the lead on coding one tab or panel of content.

You will be assessed on both a group and individual basis for this portion of the project. In a comment at the top of your Shiny app program (not displayed in the app itself unless you want to), please indicate specifically which parts each team member took the lead on and/or contributed to, if it is not otherwise clear in the program, e.g.,

```
# Janet: Tab 1 Histogram
# Karl: Tab 2 Scatterplot
# Simar: Tab 3 Bar graph
```

To create the shiny app, you should include the Shiny app file (*app.R*) in its own folder along with any necessary datasets or other files read in by the *app.R* file. The name of the folder should be the name of the app (no spaces).

The Shiny app must be published and there must be a working link in the repo to it. (Only one group member needs to host it, but there needs to be one published version.)

Presentation (Thursday, April 7)

Each group will present their Shiny application to the class in a 6-7-minute oral presentation. You will be assessed on both a group and individual basis for this portion of the project.

An effective oral presentation is an integral part of this project. Communication is key as a data scientist. In their book *Build a Career in Data Science*, Emily Robinson and Jacqueline Nolis emphasize the importance of communication. Here are just three quotes:

- "employers are first and foremost looking for evidence that you can code and communicate about data"* (page 59)
- "Much of a data scientist's job is conveying information to nontechnical peers"* (page 141)
- "A data scientist needs to be able to communicate. Over and over, people we interviewed for the book mentioned that their success came from communicating their work effectively."* (page 280)

You want to show you can communicate your results clearly (with the audience in mind) and concisely. If your audience cannot understand your results or interpretations, then the technical merit of your project is irrelevant.

The intended audience for this presentation is our actual audience: a class of data science students. Your goal should be to convey to the audience a clear understanding of your topic, along with a basic understanding

of your project, and how well the Shiny app addresses the question(s) you posed. You should not tell us everything that you did, nor should you show a bunch of things that you tried that didn't work well.

There are always exceptions, but your talk should probably follow this general structure: brief background on topic, define question(s) of interest, explain data source(s), display Shiny app (including demonstrating the app's functionality and the main messages it's conveying), concluding remarks.

You don't have to display the app live, though that may work well for some groups. You can use slides that show screen shots of the app with settings you have selected for demonstration. This may be useful if the data set takes a long time to load or the visuals chosen in the app take a long time to render.

Executive Summary (Thursday, April 7)

Apart from the presentation, you will also prepare an executive summary (like a handout) of your app that covers the same basic ideas covered in the presentation. That is, it should convey to the audience a clear understanding of your topic, along with a basic understanding of your project, and how well the Shiny app addresses the question(s) you posed. You should not tell us everything that you did, nor should you show a bunch of things that you tried that didn't work well. It must also include at least one image of the app in order to demonstrate it's functionality. (This will help you learn to include an image in a .Rmd file).

This executive summary must be at most 2 pages long, and a template is provided. It should be easy to construct the summary from your presentation.

Why is this a useful exercise if we are already making the presentation? Well, an executive summary could be given to someone to review if they missed your presentation, or wanted to review key ideas about the project, etc. Presenting information concisely is also a great skill for data scientists to have.

The compiled .pdf of the executive summary should be submitted in your group repo for the project.

Reflection (Monday, April 11, part of prep 7)

The reflection will be completed individually, and consists of a series of questions designed to help you reflect on what worked well within your group and what could be improved upon for next time. One of the questions will involve assessing yourself and your team members using the American Association of Colleges and Universities' Teamwork Value Rubric (available in the projects folder and on Moodle). Check out the rubric early on in the project to see what qualities contribute to being a good team member!

Timeline and grade

The initial proposal will be sent to me via email with all group members cc-ed so that I may set up the repos with the group names you choose.

The Shiny app must be published and there must be a working link in the repo to it.

The peer group reflection is a Google form that will have a link on Moodle posted for it.

All other deliverables must be submitted to your group repo. This includes the revised proposal, wrangled dataset, app files, executive summary, and a copy of the presentation slides.

Activity	**Points**	**Timeline**
Initial proposal	10 points	Friday, March 11 by midnight ET (Email)
Revised proposal	10 points	Thursday, March 24 by midnight ET
Project work week		The week of March 28 in class will be devoted to project work
Shiny tutorial	Individual	Wednesday, March 30 by midnight
Wrangled dataset	30 points	Thursday, March 31 by midnight ET
Shiny app	50 points	Thursday, April 7 by start of class
Presentation	30 points	Thursday, April 7 by start of class
Executive summary	20 points	Thursday, April 7 by start of class
Peer group reflection	10 points	Monday, April 11 by midnight (reminder as part of Prep7)