Stat 541 Final Portfolio

Table of contents

Independent Learning (IL):

These objectives show your ability to seek out new information and adapt to new tools to solve data analysis problems.

[IL-1] Adding new skills:

- I can find and adopt new packages to accomplish tasks.
- I can adapt to different syntax styles (tidy, base, formula style, data.table).
- I can use tutorials, etc. to enhance my understanding of new concepts

Level: 3

Justification:

Lab2: I read the documentation for the ggpubr package to be able to enhance my data visualization.

Lab4: I read in the data when it was a txt file (base R).

Project1: I followed a Youtube tutorial to enhance the UI/UX of our Shiny app.

3

[IL-2] Online resources:

- I can use online resources (Google, ChatGPT, StackOverflow) to solve problems, debug, or find new tools.
- I can find source code for similar projects to use as starting points for my own
- I can read the documentation of an API to figure out how to access data.

Level: 3

Justification:

Github: I used online resources to learn how to use git commands. Additionally, I learned about best practices for Github during collaboration.

Project1: We used ChatGPT to set up the RShiny app

Project1: I watched a Youtube video titled "How to bring modern UI to your Shiny apps" in order to make our app look prettier.

Lab4: We had to look at the information/documentation ont he satellite passes website to figure out how to use the API to access the JSON data.

Reproducible Workflow (RW):

These objectives show your ability to produce artifacts and deliverables that are organized, documented, version tracked, and responsibly designed. _____

[RW-1] File, code, and data management:

- I can use Git and GitHub to track my progress and collaborate (creating repos, cloning, forking, pull requesting).
- I always use R Projects and the {here} package to organize my scripts, notebooks, data, and applications.

Level: 4

Justification:

Github: Before this class, I never really used Git and Github. I'm glad that one of the goals since the beginning of the class was to learn about it, and I have been taking that very seriously. I learned git commands, so that I can use Github features outside of R. My partner, Eva, was not very well versed in data management, so I helped her answer questions about Github related stuff. I ran into merge conflicts, and I was able to help solve that.

Project1: I set up this Github repository, and organized the files in its correct directory such as "Data" and "Scripts". I used the "here" package when reading in my data.

Lab4: Another example of organizing files in appropriate directories and using the "here" package.

[RW-2] Notebooks:

- I can use Quarto and/or R Markdown to produce a reproducible notebook and polished rendered document.
- I can use appropriate chunk options (echo, error, cache, etc.) to render my qmd/Rmd quickly and cleanly.

Level: 3

Justification:

Lab2: I used Quarto formatting styles to organize all of the bullet points presented in this analysis

Lab3: I added to the YAML to intentionally set the global chunk options(echo, include, message, warning).

Lab4: I added labels to my code chunks to make it easier to navigate in my notebook.

[RW-3] Code style

- My code is clear, readable, well-organized, and well-commented.
- I can use a package-based workflow to organize my analyses

Level: 2

Justification

Lab2: I fixed the formatting from the work that my teammates did. The bullet points were not consistent, so I made sure they were.

Project2: I made comments to every part of my code to indicate what is happening. For example, I labeled where the stopping conditions began and what kind of statistics are the valid inputs.

Technical Communication (TC):

These objectives show your ability to communicate the processes you have implemented in your code, as well as the data conclusions and results.

[TC-1] Project summaries:

- I can clearly and succinctly summarize the contributions of my project.
- I accurately interpret statistical or modeling results.
- I consider the appropriate scope and impact of my project results.

Level: 4

Justification:

Lab1: I interpreted the p-values for the predictors.

(Project1)[https://github.com/lana1218/STAT-541-Project1/commit/1c666a4502ce04d60321eb9d84908587a559 We wrote this in class together, but essentially, this README file contained information about an overview of our dataset, data cleaning steps, scrips and files in our folder, and the capabilities of the app.

Project2: I created a README file for our package using R markdown. This allowed the user to see a high-level overview of the capabilities of the package. This README file included a description of the package overall as well as a few use cases and sample code.

[TC-2] Documentation:

- I provide ample documentation and tutorials for my custom functions.
- I provide user-friendly guides for my tools and software

Level: 3

Justification:

Project2: I wrote a custom function for the package and provided documentation in the form of a docstring, so that the user is able to prompt the "help" in R and understand how to use the function.

Project2: I created a README file for our package using R markdown. This allowed the user to see a high-level overview of the capabilities of the package. Additionally, there were examples of use cases and sample code, so that the user is able to follow along and replicate.

Data Manipulation (DM):

These objectives relate to the collection, cleaning, processing, and preparing of datasets for analysis.

[DM-1] Data preparation:

- I can read in datasets to R, including untidy ones.
- I can clean datasets to deal with missing data, typos, poor formatting, etc.

Level: 3

Justification:

Lab1: I had to clean up the "date of birth" variable since all of the observations had many different types of formats.

Lab2: We read in the dataset (xlsx format) using read_excel. Since we need to change the coding of the variable from numeric to the actual region, I suggested using a "mutate case when" function in order to reassign the different regions.

Lab3: We read in the states_capital.txt dataset.

Project1: The "911 calls" dataset from the Seattle website consisted of almost 2 million rows. I took a sample of that dataset and saved it to a new csv file. I cleaned our dataset by making sure our variables are in Datetime format. Even though our sampled file was significantly smaller than before, I still needed to filter it by the year 2023 so that the app does not lag as much.

Lab9: I cleaned the "surf" dataset to only include numeric values, so that I can generate art from it.

[DM-2] Data wrangling

- I can cleverly use pivoting, grouping, and joining to wrangle data.
- I can use mapping ({purrr}), applying (tapply, lapply, …), and/or iteration (for loops) to perform repeated tasks.

Level: 2

Justification:

Lab1: I used a lot of of tidyverse's data wrangling functions to find the answer to Part 2's questions. This includes functions such as group_by, summarize, slice_max, mutate, etc. Additionally, we received an expert level grade for this portion of the lab.

Lab3: We used the map_dfr function to "map" the JSON data to a data frame.

[DM-3] Data formats

- I can use API urls to access JSON data and convert it to a data frame

Level: 2

Justification

Lab3: We had to look at the information/documentation ont he satellite passes website to figure out how to use the API to access the JSON data. We then used the map_dfr to return it as a dataframe.

[DM-4] Data collection

- I can webscrape simple tables and information

Level: 2

Justification

Lab4: We used web scraping to get the information about each of the ten cheeses, with the help of Chat GPT.

Professional Visualization (PV):

[PV-1] ggplot: grammar of graphics

- I can use less common geometries, including those from ggplot extension packages.
- I can use the correct aesthetics to map variables
- I understand how geometries inherit aesthetics I can add annotations to my plot

Level: 3

Justification:

Lab2: When recreating the "ugly" plot from this lab, I had to make a new data frame in order to manipulate the variables for plotting. I ordered the categorical variables by using the factor function.

Project2: Here is an example where I set up a function to plot a histogram, given a bootstrap distribution. Another parameter that I added to the function was "statistic," so I can use the user's input to give the plot a title.

Lab9: I used ggplot to plot my points from an existing dataset in polar coordinates to make art!

Lab9: I used ggplot from data points that were generated from uniform distributions in order to make art!

[PV-2] ggplot: theme

- I can edit the titles, subtitles, captions, axis labels, etc. to create a clearly labelled plot
- I can choose colors ("scales") and themes to make a visually pleasing and accessible plot

Level: 4

Justification:

Lab2: To clearly label my plot, I gave it a title and subtitle. I also removed the default axis labels since we did not need them for the plot to be clear.

Lab3: To recreate the "ugly" plot from the original report, I utilized a variety of theme functions to make the plot look as nice as possible. For this plot in particular, I used "theme_pubr" so that it is ready for publication. Additionally, I made sure the texts were the correct size so that it is readable. I gave the plot a title to the plot to indicated what the variables are, and I made a subtitle to indicate what trend the reader should be seeing.

Lab9: Here is an example where I chose a color palette called "Surf and Turf" for my art piece. I also gave it a title.

[PV-3] Dynamic visualizations

- I can use a package like {gganimate} to create self-contained gifs
- I can use a package like {plotly}, {ggplotly}, {leaflet}, {ggirafe}, etc. to make interactive html widgets

Level: 3

Justification:

Project1: I made a leaflet plot. https://github.com/lana1218/STAT-541-Project1/pull/2/files

Lab4: We made a dynamic visualization to animate the route of the ISS. https://github.com/Cal-Poly-Advanced-R/lab-3-using-apis-joshblank73/pull/7/files

---

[PV-4] Shiny

- I can create a functional Shiny app.
- I understand the principle of reactivity, and how to use it.

Level: 4

Justification:

Project1: I created a Shiny App that displays a map of 911 accidents in Seattle, a time-series plot, and the average number of accidents per day. The filters that the user can toggle with include the type of accident and the month of accident. Not only can I implement these features and explain them thoroughly, but I can make a beautiful user-interface. I made a toggle for the filters, so that the user can close them if they wanted full-screen the output. Additionally, I intentionally chose appealing themes, fonts, and colors. https://github.com/lana1218/STAT-541-Project1/pull/4/files https://github.com/lana1218/STAT-541-Project1/pull/2/files

Lab4: We created a Shiny app to display the route of the ISS. I made the app look prettier using the UI skills that I learned from the previous project. https://github.com/Cal-Poly-Advanced-R/lab-3-using-apis-joshblank73/pull/7/files

---

Software Development (SD):

These objectives relate to your ability to develop correct, usable, well-designed, and sophisticated software in the R language.

[SD-2] Package creation:

- I can create a folder that is installable as an R package, possibly using {usethis} helper functions
- I can document my functions using {roxygen2} style commenting
- I can write and run unit tests using {testthat}
- I can design a package that is user-friendly and has well-designed functions.

Level: 3

Justification

Project2: There is no pull request, but I set up the R package using the {use this} helper functions

Project2: I made sure all of the functions that I wrote was well-documented using {roxygen2} style commenting.

https://github.com/lana1218/rbooty/pull/5/files

https://github.com/lana1218/rbooty/pull/8

Project2: I helped brainstorm what kind of unit tests were needed for our R package.

https://github.com/lana1218/rbooty/pull/13/files

Project2: I helped make our package more user friendly by adding use cases to our README file.

https://github.com/lana1218/rbooty/pull/11/files

Algorithms and Iteration (AI):

These objectives ask you to design code-based approaches to statistical computing problems, usually involving iteration to a stopping condition.

[AI-1] Creating an algorithm:

- I can invent and implement my own iterative algorithm.
- I have built in checks for possible problems or extreme cases in the algorithm.

Level: 2

Justification:

Lab3: I helped brainstorm how to get the url for the 3 closes future international space stations passes.

Project2: I created stopping conditions for the case that the user inputs something invalid.

[AI-2] Generative art:

- I can apply a variety of generative art functions to make a visually pleasing piece.
- I can explain why particular changes to the code result in particular differences in the visualization.

Level: 2

Justification:

Lab9: I created generative art by making functions to sample data, color palettes, and plot points. I explained how I utilized different angles and radii lengths to make my art.

Lab9: I also created art using an existing dataset.

Code Design (CD):

These objectives relate to making wise or clever choices in how you implement a procedure in code; including creating functions and objects, or thinking about the clarity and efficiency of processes.

[CD-1] Speed and Efficiency:

- I can recognize moments of possible slowdown in my code, and use built-in functions or parallelizing to speed them up.
- I always use and design vectorized functions whenever possible.

Level: 2

Justification:

Lab3: I suggested using sapply to vectorize our process.

Project2: I used an "if" and "else if" logic to minimizing the number of checks I needed to do (boolean zen).

[CD-2] Object handling:

- I can make reasonable choices in my code design about when to save intermediate objects.
- I can convert objects between types and structures as needed.

Level: 2

Justification:

[CD-3] Supporting functions:

- I write helper/shortcut functions to streamline repeated tasks and make my code easier to read.
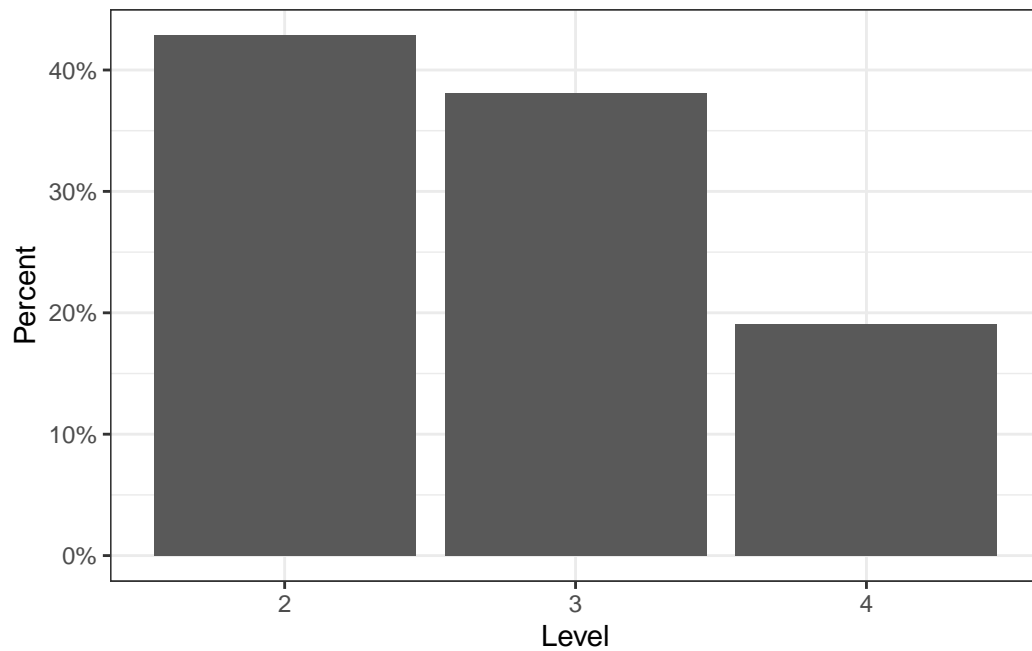- I use intermediate functions to streamline repeated or looping processes.

Level: 2

Justification:

Lab9: I used helper functions to sample color palettes and data.

Summary

```
Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(count)` instead.
```


Grade

Based on the summary plot above, I believe I have earned an A in STAT 541.

Justification

Based on the content of this course, I think I have put a lot of effort into learning and mastering the learning objectives for each week. Although I still have a lot of 1s and 2s at this time, I will continue to build upon my portfolio in the coming weeks to show that I fully mastered the material.