

**Final independent project**  
**Data Science Fundamentals in R**  
**Dr. Nyssa Silbiger**

You will create and present one final independent project that will be due on **December 10<sup>th</sup> at 1pm** and your presentations will be on the 10<sup>th</sup> and 17<sup>th</sup>. The goal of the independent project is to *tell a story* with your data. You will have multiple possible options on how you plan to tell that story, listed below. Regardless of which option you choose the code for your final project with its associated data and output must be pushed onto GitHub in a **new stand-alone public repository on the OCN 682 organization**. You must also have **at least 3 outputs** associated with your project. The term output is used loosely: it can be a plot, table, animation, unique useful functions, etc. I want you to be creative and communicate your results in an informative way.

You will give an 8 – 10 minute presentation (there is no time to go over, so make sure you practice) to the class explaining your dataset, the questions that you intended to answer with your data, and walk your classmates through the steps you took to get to your final set of visualizations/outputs. Your visuals should be publication quality and easy to understand. You can use any platform for your presentation

Choose one of the options listed below or feel free to make a suggestion. I have added some associated resources to get you started.

You will need to propose an idea and dataset to me by **October 22<sup>nd</sup>** via slack, email, in person, or office hours. Let me know what data you would like to use and any preliminary ideas that you have for your project. Use this as an opportunity to get advice on how to move forward. If you do not have your own dataset to work with you can use any public dataset available (there are hundreds of data sets in TidyTuesday).

----- Everything below we learned in class -----

**Rmarkdown or Quarto document** (will need to add a bit of “flair” if using Rmarkdown to advance it slightly past what we did in class).

- You could use multiple data sources
- Make animations with your plots
- Make a map
- Include a function
- You can use package *flair* <https://github.com/r-for-educators/flair>
- Or other more advanced coding techniques of your choosing

**Shiny app**

- <https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/>
- <https://deanattali.com/blog/building-shiny-apps-tutorial/>

----- Everything below is one step up from what we did in class -----

**Create a Package** with your own set of canned functions specific to your research needs (at least 5 functions). Your package must include a test dataset with examples of how to use the different functions.

- <https://r-pkgs.org/index.html>
- [https://kbroman.org/pkg\\_primer](https://kbroman.org/pkg_primer)

**Xaringan presentation** – This package is what I use to make the slides for lectures in class.

- <https://slides.yihui.org/xaringan/>
- [http://www.favstats.eu/post/xaringan\\_tut/](http://www.favstats.eu/post/xaringan_tut/)

**Bookdown document** – This is a compilation of multiple markdown files into a book.

- <https://bookdown.org/yihui/bookdown/>

**Github pages website** – Free version-controlled website through GitHub.

- <https://guides.github.com/features/pages/>
- [https://www.youtube.com/watch?v=BA\\_c3bGQXIQ](https://www.youtube.com/watch?v=BA_c3bGQXIQ)

**Other ideas are welcome, just let me know!**

**Grading rubric on next page.**

**You grade will be determined based on the criteria below:**

1 – poor job

10 – excellent job

**Do your datasheets follow best practices?**

1      2      3      4      5      6      7      8      9      10

**Do you have appropriate metadata and data dictionaries in your repository?**

1      2      3      4      5      6      7      8      9      10

**Is the file structured appropriately for your project?**

1      2      3      4      5      6      7      8      9      10

**Does it have a readme file with relevant information?**

1      2      3      4      5      6      7      8      9      10

**Is your code commented appropriately?**

1      2      3      4      5      6      7      8      9      10

**Is your code clean and easy to understand?**

1      2      3      4      5      6      7      8      9      10

**Are your graphics clear and easy to follow?**

1      2      3      4      5      6      7      8      9      10

**Is your presentation clear and easy to follow (and did it stay within the required timeframe)?**

1      2      3      4      5      6      7      8      9      10

**Did you answer questions from the audience appropriately?**

1      2      3      4      5      6      7      8      9      10

**How creative is your set of visuals/platform used to communicate your data?**

1      2      3      4      5      6      7      8      9      10