

BIOL351: Data Visualization & Communication for Biologists

THE FINAL PROJECT

OVERVIEW:

In your capstone project for this class, you and your partners will select a biological topic of interest to the general public and develop visual representations from publicly accessible data. This is an individual project, but you will work with similar datasets to those in your group so that you have classmates to discuss the data, material, and visualizations. Think of your group as your lab mates or work colleagues; they understand the system you are working in and can offer advice and troubleshooting help. However, all the material that you turn in should be your own and will be graded individually. This assignment is an opportunity for you to apply your ability to source, organize, and communicate using publicly available sources of data. Each person will be responsible for producing two different visualizations.

You will tell a story from the data with your visualizations through a written report on your analysis (made in R markdown) and you will prepare **one slide** to share with the class during our final exam period. This project is worth 25% of your final grade (10% on the code, graphs, and slide; 10% on the written analysis; 5% on your participation in the lightening poster session). A rubric for this project will be provided.

DATA TOPICS:

1. COVID data ([New York Times](#) live data on cases and other measures)
2. Natural disasters
 - a. Wildfires (National [Burn Severity Index](#) or [Fire Program Analysis](#) data)
 - b. Hurricanes and Typhons ([NOAA location and wind speed](#) data)
 - c. Earthquakes ([USGS significant earthquakes](#) data)
3. Coral Bleaching
 - a. Sea Surface Temperature (NOAA [Coral Reef Watch](#))
 - b. Bleaching events ([Reefbase](#) data)
4. Continue exploring National Earth Observatory Network data ([NEON](#))
5. Clinical data on cancer:
 - a. [Gene mutations in breast cancer patients](#) data set
 - b. [American Cancer Society](#) has data on case number and death rate
6. Opioid use:
 - a. DEA's database tracking where opioid pills go ([Washington Post](#) gives good overview of ways to access this data). This is a big dataset, so here is a [subset of the data](#) that you can look at before choosing to dive in.
 - b. This [Kaggle dataset](#) on 25K opiate prescriptions is a subset of a much larger (1 million entries!) dataset from the [Centers for Medicare and Medicaid Services](#). We can't guarantee that the smaller dataset is representative of the whole dataset, but you may use the subset to save processing power, if you'd like.
7. Malaria data from many sources has been compiled by the [Malaria Atlas Project](#)

Source: Grayson, K., Hilliker, A. (2021). [Teaching Data Viz and Communication as an Undergraduate Biology Course: Assignments and Projects. Calling Bull - a resource sharing and teaching community](#), QUBES Educational Resources. [doi:10.25334/5C87-YE71](#)

8. Data on the International Trade of Endangered species from [CITES](#). This data covers documented, legal trading. There is an R package, [citesdb](#), to help explore the data, if you are interested in exploring it.

If you are looking for some other data to complement what you use above, try [Google's Dataset Search](#) engine.

Within each topic, each person will present different aspects of the data so you need to coordinate to avoid overlap and produce figures that address different questions. While we have provided recommended data sources, you can run other ideas for topics past your instructors. Group members can use different datasets or parts of datasets within your topic.

The requirements for your two figures:

- At least one visualization must be made in Program R with annotated code
- The second visualization can be made in Program R or in Tableau
- You must make two different graph types. The type of graph will depend on the data, but we are looking for you to move beyond a simple bar chart here, so make sure your data set is rich and interesting enough to create some interesting visualizations.
- The graphs should not be two different ways to show the exact same variables of data. They should explore different aspects of the data set.
- The data must come from a reputable source and you must have access to the details of the data collection methods.
- You will turn in your proposed data source and the question(s) you are addressing on Monday April 19th for approval from your instructors.

THE STORY:

You will be presenting your data story in two ways— via a written analysis submitted to your instructors as a R Markdown file and a “lightening” slide to share with your classmates.

THE VISUALIZATIONS:

The audience for your figures will be a highly interested, scientifically literate lay person. Think of someone reading the New York Times or [FiveThirtyEight](#) . They are interested in science and are willing to spend some time looking at complex visualizations. But being eye catching and clear is important. Your visualizations should be embedded in your written analysis as well as your presentation slide.

THE WRITTEN ANALYSIS:

The audience for your written analysis is the same audience as above, plus your instructors. You will turn in a knitted PDF from an R markdown file with the following components:

- An interesting and truthful **title**.
- **Background:** Brief, essential background information (~2 paragraphs). You should include any information critical to understanding the specifics of your question and the

subfield. Also contextualize why the audience should care about the topic and include references to external sources.

- Use in-text citations throughout your text (Armado et al., 2013). <- as an example.
 - A strong background will include information from primary research papers and other reputable sources of data.
- **Data Overview:** A description of the data (1 paragraph). Your description should include who gathered the data, where you accessed the data, number of variables (columns) in the data, sample size (number of rows), and which parts of the data you used in your analyses.
- **Data Visualization:**
 - Clear statement of the question(s) or hypotheses being tested.
 - Include all the code used to clean, process, and visualize your data. Please make sure that only your final code is included (not code that you tried but isn't used in your final product). Your code should include comments to explain what how you are processing your data.
 - Make sure that both of your figures are presented clearly in your markdown file. Follow each figure with a figure legend that helps the reader understand what is shown in the data. Be sure to define any shorthand, symbols, etc.
 - If you use Tableau, a description of how you designed your second figure (and please upload your data and workbook to Tableau online). Your Tableau figure should be saved as an image file and included in your markdown file. Upload the image file in the file pane of RStudio and use the following code to include it in your markdown file:
 - ``
 - Where _____ is the name of the file – make sure to include the file extension
 - You can include commands to format the display of the image such as:
`{#id .class width=50% height=50%}`
- **Data Analysis:** An analysis section that includes (2 - 3 paragraphs):
 - A description of the key trends in your data (results)
 - What you've learned from the data (conclusions)
 - Analysis of the strengths/weaknesses of the data collection.
 - Proposal for future analyses or experiments to address questions raised by this data.
 - A description of your aesthetic choices and how they improve the interpretation of the data
- **Discussion:** Your discussion section should summarize the current view of your topic in the literature and how your results compare to the work of others (~2 paragraphs and referencing at least three papers)
- **References:** Include all external sources listed in [APA format](#). Please make sure all links are active and format your links as DOIs for journal articles (not the web address).

- Your PDF should be formatted clearly and attractively, so consult this [R markdown cheat sheet](#) for help with formatting.
 - If your comments are long, they will be cut off in the knitted PDF. In that case, knit to a word doc and then export as a PDF. Make sure any links still work in the PDF.

THE LIGHTENING SLIDE PRESENTATION:

Your audience is your classmates and you have one slide to show your figures and any other graphics that you want to display. You will display your slide during our final exam period.

Your slide should include:

- A title
- Your two figures
- Other visuals if you'd like
- A slide format that you find aesthetically pleasing

We will broadcast 6 posters at a time for roughly 20 minutes. Your classmates will mingle around the room and ask you questions. You should be prepared to give them a short informal walk through to share your data viz story (2-3 minutes) or be prepared to ask questions. The presentation is not formal, but a chance for you to celebrate what you've learned and show off your data viz skills.

If you'd like to see some examples of what one can do to show multiple facets of data in one attractive slide, check out the "past examples" [here](#).

SUBMISSION:

You will submit:

- A knitted PDF of your markdown file, submitted to your data viz folder (please make a subfolder called Final Project), named LastNameFirstInitial_FinalProject.pdf
- Your slide (.ppt or google slide), submitted to your data viz folder (in the Final Project sub folder), named LastNameFirstInitial_Slide
- Your R markdown file should be on R StudioCloud, uploaded into our class repository as a new project, named LastNameFirstInitial_FinalProject.rmd. We should be able to click "run all" and run your entire code without errors.
- If you used Tableau, upload your worksheet and data to Tableau Online.

Your files are due by the beginning of our exam period, Friday April 30th, at 7 pm. We will have the lightening poster session during our final exam period and celebrate all the skills you've learned. There will not be an exam or any other assignment during the final exam period. However, your attendance and participation in the lightening poster session (as presenter and audience member) are required and are part of your project grade.