R Data Anlaysis Course: Final Project Instructions

Ronnie Bailey-Steinitz

2025-11-03

Week 8: Independent Data Exploration and Analysis

Self-Guided Exercise

This week, you will apply all the skills you've learned so far to analyze new data.

Your goal is to create a *clear*, *reproducible* R Markdown report that walks through your full analysis: **from data import to final visualization.**

You'll share one plot from your analysis in next week's *Presentation Day*.

1. Create a New R Markdown File

- Open RStudio and create a new .Rmd file called Week8_YourName_DataAnalysis.Rmd.
- Save it inside your course project folder.
- Add a title, author, and date in the header.
- Use headings (`#`, `##`) to organize your markdown file into sections.

2. Load Required Libraries

- Determine which libraries you need and load them.

3. Import Your Dataset

- Choose one dataset: **your own research dataset** or a **demo dataset** (e.g., `penguins`, `gapminder`, `diabetes`, or other from the shared folder).
- Import your data `read_csv()` or `read_xlsx()`.
- Clean the variable names for easier coding. Remember you can always use `names()` to view a list of your column names.
- View or glimpse your dataset to make sure it loaded properly.

4. Explore and Clean the Data

- What are the variable types (e.g., numeric, character, factor)?
- Do any variables need to be converted (e.g., from text to numeric)?
- Are there missing values? Do existing NAs make sense?
- Are variable names consistent and clear?

5. Summarize and Visualize

- Compute basic summary statistics for key variables.
- Report your summary statistics as **text between your code chunks!**.
- Create initial visualizations to explore patterns or distributions.

6. Conduct 1–2 Statistical Analyses

Choose analyses appropriate for your data:

- Explore relationships -> linear or logistic regression
- Compare two groups -> t-test
- Compare more than two groups -> ANOVA

7. Visualize Your Results

Create at least one final visualization that summarizes your main result. Examples:

- Boxplot with group means and p-values
- Bar plot with error bars
- Stacked columns with different fill colors
- Scatterplot with regression line

In your plot, you should:

1. Give your plot a title
2. Label axes clearly.
3. Adjust geom parameters to best reflect your data (e.g., remove standard error ribbon if it is distracting; use alpha to set transparency for clustered points or stacked geoms).
4. Control colors and themes where appropriate.
5. Export your graph.

8. Write a Brief Interpretation.

For each analysis, include a few sentences below the analysis code chunk:

1. What question were you testing?
2. What type of data did you use (e.g., numeric and also categorical)
3. What did you find (in plain language)?
4. How does your visualization help explain it?
5. **What issues did you encounter** when analyzing or visualizing your data, and **how did you resolve them?**

9. Knit and Submit

- Knit your file to HTML or PDF.
- Make sure your code runs start-to-finish without errors.
- Bring your favorite plot and a short explanation to class next week. You may upload the plot and a few sentences to Google Slides.