Week 4 Lab: Geoms and Aesthetic Mappings

CSC3107 Information Visualization

Michael T. Gastner

2025-05-26

Introduction

In this lab, you will explore ggplot2's geoms and aesthetic mappings. In Section 1, you will apply various geoms to visualize passenger and crew data from the Titanic's ill-fated maiden voyage. In Section 2, you will source a real-world data set and implement meaningful aesthetic mappings—color, size (or line width), and shape (or line type)—according to each variable's semantic role. By the end of this lab, you will have a better understanding of how to use ggplot2 to create effective visualizations that convey information clearly and accurately.

Quarto Instructions

Submit your solution as a Quarto Markdown file and use R to perform all tasks. Follow these guidelines:

- Use one section header for each task.
- Include the code chunks and their output in the rendered HTML file. Unless stated otherwise, do not suppress warnings or errors, but please suppress messages and avoid excessive output.
- Break down your solutions into clear, reader-friendly steps, using Markdown to explain each one. For each step, display relevant output to facilitate reviewing your solution.
- Use only R—preferably the tidyverse packages—to complete the tasks. Do not modify the raw files outside of R. Ensure your code runs correctly for anyone who downloads the data sets from the links provided in this document.
- Avoid repetitive code and `for` loops.

1 Visualizing the Distribution of Travelers on the Titanic

The spreadsheet at `titanic.csv`, available from Prof. Gastner's website, contains information about the passengers and crew aboard the Titanic during its maiden voyage in 1912, which tragically ended when the ship hit an iceberg and sank. This spreadsheet is based on data available from the Encyclopedia Titanica (Hind, 2023).

1.1 Bar Chart of Travelers by Class

Create a bar chart displaying the number of travelers by class (1st, 2nd, 3rd, and crew). Include a plot title and a caption attributing the source as "Encyclopedia Titanica (2023)." Aim for a solution similar to Figure 1.

Hint: You may find it easier to aggregate the number of travelers using `count()` before calling `ggplot()`.

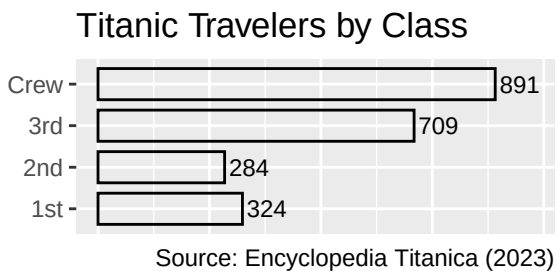

Figure 1: Suggested bar chart to be shown as a solution for Section 1.1.

1.2 Dodged Bar Chart of Travelers by Class and Survival

Use a dodged bar chart to visualize the dependence of the survival rate on the class in which a person traveled. Aim for a solution similar to Figure 2.

1.3 Mosaic Plot

When working with two categorical variables, such as class and survival, mosaic plots provide an alternative to segmented bar charts. Use the `geom_mosaic()` function from the ggmosaic package to create a mosaic plot of the data in `titanic`. Aim for a solution similar to Figure 3.

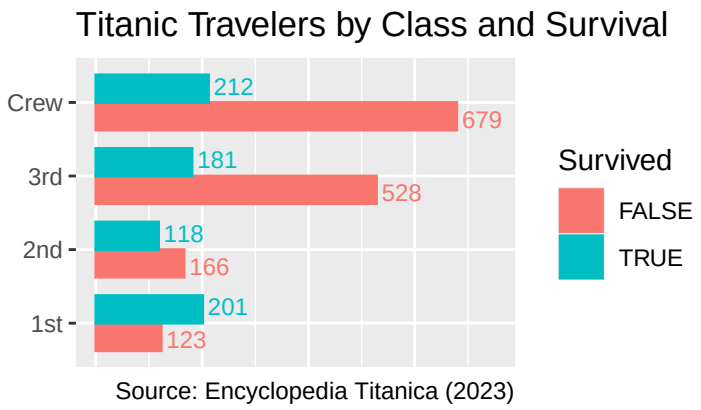

Figure 2: Suggested dodged bar chart to be shown as a solution for Section 1.2.

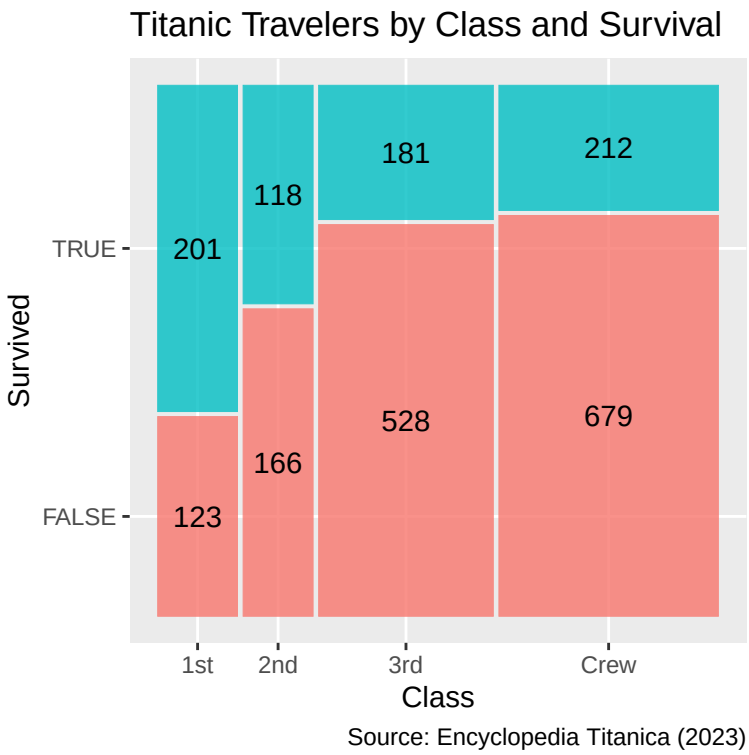

Figure 3: Suggested mosaic plot to be shown as a solution for Section 1.3.

1.4 Pros and Cons of Mosaic Plots

State one advantage and one disadvantage of using the mosaic plot in Figure 3 instead of the dodged bar chart in Figure 2 for the Titanic data.

1.5 Alternative Visualization

Using `ggplot2` (and any necessary add-on packages), create an alternative visualization to bar and mosaic plot that illustrates the same bivariate relationship between class and survival. Then, discuss the **advantages** and **limitations** of your chosen visualization compared to bar@#@! and mosaic plots.

2 Find a Data Set to Demonstrate Semantically Correct Aesthetic Mappings

2.1 Objective

Find a real-world data set and construct a single plot that exhibits **at least 2 out of 3** of the following aesthetic mappings. The visualization must include **both x and y coordinates**, and the aesthetic mappings should be semantically appropriate for the variable's type:

1. Color (via `color` or `fill`)
2. Size or line width
3. Shape or line type

2.2 Data Requirements

- **Novelty**: You may **not** recycle any data source previously used in this course.
- **Accessibility**: The data must be publicly available online in CSV or Excel format (e.g. a government portal scientific journal).
- **Provenance**: Provide the URL where you obtained the file, and include the raw data file with your submission.

2.3 Tasks

In your QMD file:

a. Provide the data URL.
b. Briefly summarize the scope, origin, and key variables of your data set in your own words.

c. Use ggplot2 (and any necessary add-on package) to produce a single plot incorporating at least two out of three aesthetic mapping categories listed in Section 2.1. Choose any plot type that best illustrates your variables.
d. For each aesthetic mapping, explain why it is appropriate, referencing the variable nature and perceptual best practices. For example: "We mapped the `population` variable to the size aesthetic because the variable is extensive: when the sample size changes, the expected value of `population` also changes by the same factor."
e. State the main insight or conclusion the reader should draw from your visualization.

3 Reflections

Write a brief team reflection (2–4 sentences each) addressing the following:

1. **Key takeaway:**
   What is one thing you learned in today's lab? Explain why this point stood out to you.
2. **Challenges:**
   What did you find confusing or difficult?
3. **AI assistance:**
   If you used any generative AI tools (e.g., ChatGPT or GitHub Copilot), explain how they helped.
4. **Teamwork:**
   How did your team collaborate on this assignment? Identify one strength and one area for improvement.
5. **Additional feedback** *(optional)*:
   Any other comments on this module or the exercise.

Your reflections should be specific to this lab assignment.

4 Submission Checklist

- Use RStudio's built-in spell checker (Edit → Check Spelling) to ensure that your Quarto Markdown file is free of spelling errors.
- Label all code chunks using the `#| label:` option.
- Use the styler and lintr packages to ensure that all R code chunks are compliant with the tidyverse style guide.
- Submit a ZIP file containing all files (including images) and directories in your RStudio project. Upload this file to xSITe (Dropbox → Week 4 Team Lab Assignment). Ensure that your project is self-contained and can be rendered without external files and dependencies.
- **Deadline**: Friday, 30 May 2025, 12:00 noon

Hind, P. (ed.) (2023) 'Encyclopedia Titanica'. Available at: https://www.encyclopedia-titanica.org/ (Accessed: 30 November 2023).