

A literacy-based approach to learning visualization with R's ggplot2 package¹

Angela M. Zoss, Ph.D. (angela.zoss@duke.edu)
Duke University Libraries

Author Name and Brief Biography

Angela Zoss is an Assessment & Data Visualization Analyst with the Duke University Libraries. She has been teaching visualization practices and tools for over 10 years, including over 100 workshops, guest lectures, and conference presentations. In her role as Duke University's first Data Visualization Coordinator, she created new library workshops on visualization; hosted an annual student data visualization contest; co-organized a weekly talk series on visualization topics; consulted with students, researchers, and faculty members on research projects; and helped to introduce visualization concepts and tools into a variety of undergraduate and graduate courses. She holds an M.S. in Communication from Cornell University and a Ph.D. in Information Science from Indiana University. She is a certified instructor for both RStudio (tidyverse) and The Carpentries.

Nutrition Information

This recipe describes the process for delivering a 2-hour workshop on visualization in R using ggplot2. The workshop is designed to promote literacy around ggplot2 and data visualization creation. Over the course of the workshop, participants learn about the theory behind the ggplot2 package and its approach to visualization, the typical syntax used to create a visualization with ggplot2, and a few features for visualization design offered by ggplot2. The workshop combines conceptual material with hands-on exercises to build participants' vocabulary, expose participants to the range of features available in

ggplot2, and reinforce learning with repeated practice.

All teaching materials and several videos of past workshops are [available publicly](#).

Target Audience & Number Served

This workshop is designed for individuals who have some prior experience with R, no experience with the ggplot2 package is expected. While the workshop has largely been taught to higher education audiences (especially students), the materials may also be appropriate for advanced high school students.

The workshop works best when participants are able to follow along with the instructor's pace and ask clarifying questions, which tends to be easier with a maximum of about 30 attendees. Having helpers available for technical problems is especially useful for larger groups.

Learning Objectives

Students will:

- recall and arrange the basic components of the code required to produce a ggplot2 visualization
- use ggplot2 documentation and cheat sheets to look up syntax information and additional helpful chart components
- critique the output of ggplot2 based on an understanding of the dataset and a general knowledge of best practices for visualization design

¹ This is an author-produced version of a book chapter. The final version will be published at in the *ACRL Data Literacy Cookbook*, edited by Kelly Getz and Meryl Brodsky. The chapter thus adopts a "cookbook" style for section headings.

- compare different visualization tools based on their properties and their suitability for a particular project
- identify patterns and useful keywords in warnings and errors generated while developing code for ggplot2 visualizations

Cooking Time

This workshop runs for two hours, but it may not be possible to cover all exercises in that time period. It's important to gauge the needs of the audience in real time and prioritize activities based on the topics you think will be most important or relevant to them.

Dietary Guidelines

This recipe outlines a workshop that is designed to promote critical reflection on the process of creating visualizations using the ggplot2 package in R. The recipe incorporates the following frames from the ACRL Framework:¹ “Information Creation as a Process” (personal choices are part of visualization creation), “Authority is Constructed and Contextual” (software is constructed and not neutral), “Research as Inquiry” (visualization assist in data-based inquiry), “Scholarship as Conversation” (visualization is a learned communication skill), and “Search as Strategic Exploration” (there are specific search strategies for software troubleshooting).

Ingredients

- Some way to share files with students (e.g., a GitHub repository, Google Drive or Box)
- RStudio with tidyverse installed on both the instructor machine and the participant machines
- Workshop files: datasets, R Markdown exercise files and answer keys, slides, handout
- Feedback forms
- Technology to sharing the instructor screen with participants (e.g., a classroom with a projector, Zoom)

Preparation

- Finalize exercises, slides, and handout
 - Tailor exercises to audience as needed
 - Test all exercises to make sure they work as expected
 - Update the answer key files with any corrections/additions
 - Practice slides
 - Update slides with any corrections/additions
 - Review and update handout based on slides and exercises
- Host final versions of files in a place that can be easily accessed by participants (e.g., GitHub, Box, Google Drive)
- Prepare feedback form for assessment
- Set up physical or virtual meeting space (including recording, if desired)
- Send reminders to registrants, including a link to the files and instructions for installing or accessing RStudio software prior to the workshop

Instructions

Begin the workshop by sharing the link to the shared workshop materials. Distribute feedback form early in the session in case participants have to leave early.

Next, share some background on ggplot2. This gives people context for the material and also provides buffer time in case participants arrive late. After the context slides, make sure everyone can open RStudio and set up the workshop files. I prefer to walk through this slowly to make sure everyone is set up before doing any hands-on exercises, even if some participants have to wait for others to catch up. The slides also include suggestions about how to debug ggplot2 code in general and how to look for help on ggplot2 visualizations, and the subsequent activities provide many opportunities to reinforce those techniques.

When teaching ggplot2, the syntax (or the composition of the code components into a valid sequence) tends to be the most difficult material.

While `ggplot2` is an R package and provides functions like other R packages, the way to build a visualization in `ggplot2` is different from the typical syntax in both base R and `tidyverse`. I split the syntax into components and give names to the different components to try to help participants remember the components more easily, a practice inspired by techniques for explicit direct instruction.² I define the components as: the main function (`ggplot()`), a geometry or shape layer (e.g., `geom_point()`), the dataset, and the aesthetics mapping. The aesthetics mapping is the trickiest part of the syntax. I find it helpful to cover the mapping in a simple but reliable way the first time it is used and then get into more detail in later exercises.

The hands-on activities included cover different aspects of chart creation and can be delivered as “work on your own” activities, “work together with the instructor” activities, or some combination. The available activities are:

- *Iris*: This activity is captured in the slides, rather than an R Markdown file. It introduces the basic `ggplot2` syntax with a scatterplot using the built-in `iris` dataset.
- *Game of Thrones*: This activity uses a dataset of character traits from the television show *Game of Thrones*. It reinforces the syntax for a basic scatterplot, but it goes into more detail about aesthetics mapping, adding multiple geometry layers, and searching help documentation to learn about function options.
- *Star Wars characters*: This activity explores new geometry layers by looking at numerical distribution and linear trends (optional). It also introduces the concepts of aesthetics inheritance, overriding data in a shape layer, and incorporating data manipulation inside a plot.
- *Star Wars opinions*: Using opinion survey data, this activity introduces techniques for working with categorical data. It covers `geom_bar()`, setting axis limits, flipping the axis of bar charts, mapping a variable to custom colors, and creating facets.

- *Gapminder*: This optional activity can be included for especially advanced groups. In this activity, a finished chart is projected and participants are challenged to build the code to reproduce it. Allot at least 10 minutes for open work time on this activity, as well as a few minutes to “solve the problem” with the group.

Between activities, I switch back to the slides to show a few additional concepts to prepare for the next activity. For example, before the *Star Wars characters* activity, I introduce facets and a few useful `tidyverse` functions, like the `magrittr %>%` (pipe) symbol and the `dplyr filter()` function. I cover factors before the *Star Wars opinions* activity.

When teaching with hands-on exercises, I use best practices for live coding³ to scaffold the experience for participants. For example, it is crucial to type slowly and verbally narrate what you are typing so that participants can type along with you without watching your screen. I also recommend always typing out the names of arguments inside the `ggplot()` and geometry functions – for example, `ggplot(data = data_frame_name)` – to help beginners understand how the elements inside the functions are interpreted by the package.

Reviews/Assessment Strategy

Use the following questions to assess the effectiveness of the workshop.

- Free text response questions:
 - List two things that you learned in this library session
 - List two things that you still don't understand about the information covered in this session
 - List one thing that would improve this library session
 - Which topics would you like to see covered in future sessions?

- Likert-type questions using a five-point scale from Strongly Agree to Strongly Disagree:
 - My understanding has increased as a result of this program/training.
 - My interest in this subject has increased as a result of this program/training.
 - I am confident I can apply what I learned in this program/training.

Adapting the Recipe

This workshop can be adapted with new activities as needed. For example, it could be tailored to an audience by selecting datasets that are targeted to a particular scholarly domain, or different features of ggplot2 could be added to existing activities. For workshops open to anyone who might wish to attend, I often select datasets related to media and entertainment, but there are many other topics of general interest (e.g., weather, sports, politics). It can be difficult to find datasets that are a good fit for visualization activities – that is, datasets that are relatively small, easy to understand quickly, and that have properties that match up with the activities you want to develop. Be sure to budget plenty of extra time for finding and preparing datasets if you adapt that part of the recipe.

Additional Resources

Posner, Miriam. 2015. "[A Better Way to Teach Technical Skills to a Group.](#)"

Miriam Posner's advice about teaching technical skills has always been an inspiration. She has a post from 2015 about some of her techniques, including another use for Post-It notes, and she continues to generously share her insights.

Vaughn, Michael. 2016. [Effective Presentation Design.](#)

Michael Vaughn's *Effective Presentation Design* is a great resource for presentation design. It includes an activity for designing presentations using Post-It notes that improved the flow of this workshop immensely.

Zoss, Angela M. 2021. "[Ggplot2 Workshop.](#)"

This GitHub repository contains all of the workshop materials described in this chapter.

Funding Acknowledgement

This project was made possible in part by the [Institute of Museum and Library Services](#), RE-73-18-0059-18.

¹ "[Framework for Information Literacy for Higher Education.](#)" 2015. Chicago..

² Hermans, Felienne, and Marileen Smit. 2018. "[Explicit Direct Instruction in Programming Education.](#)" In *Psychology of Programming Interest Group 2018 - 29th Annual Workshop*, 86–93. London, UK.

³ Wilson, Greg. 2019. "Teaching as a Performance Art." In [Teaching Tech Together: How to Make Your Lessons Work and Build a Teaching Community around Them](#), 1st Ed., 73–85. Boca Raton: Chapman and Hall/CRC.