

SPEC REU R Resources: Basic Data Visualization with ggplot2 – Group Work

Alix Ziff, Valeria Flores-Cadenas, Therese Anders, Aaron Gong, Claudia Salas Gimenez, Ben Graham

Summer 2024

For this groupwork assignment, you will build on the skills covered in previous walkthroughs by focusing on creating detailed and professional-quality figures using the `tidyverse` and `ggplot2` packages.

First, your task will be to recreate two specific visualizations, focusing on aspects such as scaling, labeling, line types, plot background, and legends. Then, you will also create your own graph using either unused variables or preload datasets in R to practice making visualizations on your own, a key skill for research. Please note that some of the detailed skills needed to accurately recreate these figures may not have been covered in the walkthroughs. However, effective troubleshooting and the use of online resources are important components of working with R.

This exercise will not only reinforce your ability to create clean, univariate plots but also give you the opportunity to develop more polished and professional-quality visuals of your own. Enjoy the challenge!

Initial Setup

Ensure your working directory is set correctly and load the necessary libraries. For this assignment, you'll use the `wdi_cleaned_part2.csv` file located in the Training Data folder. This dataset provides all the information needed for the visualizations: for Exercise 1 you'll use the percentage of energy from renewable sources, and in Exercise 2 you'll plot energy use per capita.

Remember to refer to the [World Development Indicators data catalog](#) for detailed descriptions of the indicators. Keep in mind that since these datasets have been modified for this walkthrough, the variable names may differ slightly from those in the original codebooks.

Exercise 1: Visualizing Trends in Renewable Energy Consumption by Year

Utilize the percentage of energy from renewable sources variable to generate a line density plot showcasing the distribution of renewable energy consumption across 1992, 2002 and 2012. Upgrade the plot by differentiating each year with unique colors and line types. Ensure the plot has a title, subtitle, and axis labels. Also, implement a black and white theme for the plot background.

Helpful Hint: Remember the rough structure to produce a graph in `ggplot2`: `ggplot(data = dat, aes(x = , y = , color = , linetype =)) + geom_point() + labs(title = "Your Title Here")`

Exercise 2: Visualizing Per Capita Energy Use by Country

For the second figure, focus on the energy use per capita variable to produce a series of density plots, separated by country. Include data specifically for the United States, Germany, Brazil, Russian Federation, India, and China, implement a black and white theme for the plot background, and ensure the density plot has a blue

fill with controlled transparency. The figure should also include a title, subtitle, and axis labels, with a black and white theme for the background.

Bonus Question

For an extra challenge, create a new visualization using either a variable from the `wdi_cleaned_part2.csv` file that hasn't been used before, or any of the preloaded datasets in R. To see the available datasets in R, use the `data()` command to list them. You can then use `?dataset_name` (replace "dataset_name" with the actual dataset name) to get details about the variables. Remember that we've worked with some of these datasets before, like `hflights` in the Data Management I module. Make sure to choose a visualization that fits the data and its purpose (see the supplementary document on 'What Makes a Good Visualization?' for references on which visualization to use).

While working on this, pay close attention to the aesthetics to create professional-quality figures. This includes adding a clear title, adjusting colors or line types, and experimenting with settings like transparency, among other things. Feel free to also explore different features that we might not have covered in class, such as adding different backgrounds or themes. This task will help you explore new data, be creative, and improve your attention to detail. Good luck!