

SPEC REU R Resources: Multivariate Plots – Group Work

Radhika Ananth, Gaea Morales, Claudia Salas Gimenez and Ben Graham

Summer 2024

Welcome to the groupwork assignment on multivariate plots! In this document, you'll apply what you've learned in the previous four walkthroughs on multivariate plots by visualizing data trends across countries using key developmental indicators.

For each graph, be sure to include a title, subtitle, and appropriately labeled axes to clearly convey the information on its own. Additionally, challenge yourself to enhance the plot with extra aesthetic elements for high-quality, polished visuals.

Initial Setup

Before we dive into the exercises, set up your environment and load the necessary libraries and dataset. For this assignment, we'll use a compilation of developmental indicators from the World Bank WDI data for 202 countries from 1960 to 2005, saved as `wdi_development_data.csv`.

For more details on the indicators used, check the [World Development Indicators data catalog](#).

Exercise 1: Plotting Infant Mortality for a Selected Country

Select a country and create a line plot showing infant mortality rates (per 1,000 live births) over time, from 1960 to 2005. Then, briefly interpret the trends for the country you selected.

Exercise 2: Comparing Infant Mortality Across Five Countries

Choose five countries and compare their infant mortality rates over time (1960-2005). Differentiate each country by color, and include a legend. Comment on the observed trends.

Bonus Question (Exercise 2)

Add a label for each country directly on the plot instead of using a legend, if the trends allow clear differentiation.

Note: This will only be feasible if the selected countries have sufficiently distinct y-axis values.

Exercise 3: Global Trends in Infant Mortality

Create a scatterplot to display global infant mortality rates over time for the years 1985 to 2000. Adjust the opacity and add jitter to the points for clarity, and be sure to include a trend line to highlight overall trends. Feel free to incorporate additional aesthetic elements to make the visualization more engaging and informative.

What are the general trends that you observe in infant mortality? How does adjusting the opacity improve how informative your graph is?

Exercise 4: Impact of Wealth on Infant Mortality

For this last exercise, create a dummy variable called `wealthy` based on GDP per capita (`gdppc`), classifying countries as high-income if their `gdppc` is above the mean, assigning them a value of 1, while countries below the mean receive a 0. Use this `wealthy` variable to distinguish data points in the scatterplot from Exercise 3 by shape and color in a scatter plot, then interpret the observed trends.

Bonus Question

For an extra challenge, try creating a new visualization by selecting variables from the `wdi_development_data.csv` file or data from one of R's preloaded datasets (use `data()` to list them, and `?dataset_name` to see details about each). Choose a plot type that best represents your variables and aligns with the purpose of the figure. Additionally, include a brief description explaining why you chose this type of plot—what insights are you aiming to reveal, and how does the plot effectively highlight that information?