

Data Visualization II: Multivariate Plots - Group Work

Radhika Ananth and Gaea Morales

Summer 2022

Multivariate Plots

This groupwork assignment deals with multivariate plots. The dataset we will use is `wdi_development_data.csv`, which can be found in the [Training Data folder](#). The dataset is a compilation of developmental indicators from the World Bank WDI data for 202 countries between the years 1960-2005.

For each of your graphs, please include a title and a subtitle and make sure to label your axes. Seek to make your graph as informative as possible: the goal is to have a plot that can (more or less) stand on its own!

Exercise 1: For a country of your choosing, plot the infant mortality (given per 1000 life births) against time for the years 1960-2005 in a line graph. Briefly explain what trend you observe in the infant mortality over time.

Exercise 2: For five countries of your choosing, plot the infant mortality against time for the years 1960-2005 in a line graph. Distinguish each country by color and make sure to include a legend. Comment on the differences across the 5 countries with time.

BONUS: Instead of a legend, add the names of the five countries as labels on the plot. This will require locating labels at specific coordinates that make it clear which line the label refers to. Note that this will only be possible if the countries selected have Y-axis values that are sufficiently distinct.

Exercise 3: In a scatterplot, plot infant mortality, for all countries, against time for the years 1985-2000. Include a jitter, choose an appropriate opacity, and include a trend line in your graph. What are the general trends that you observe in infant mortality? How does adjusting the opacity improve how informative your graph is?

Exercise 4: Lastly, create a dummy variable called `wealthy`. Classify countries with a `gdppc > mean gdppc` as `wealthy` (`dummy = 1`) and those with a `gdppc < mean gdppc` as `not wealthy` (`dummy = 0`). Distinguish data points in your graph from exercise 3 using this dummy variable using both shape and color, so that there is a clear distinction between the data points from the two groups of countries respectively. Comment on trends across time.