

The goal of this assignment is explore and manipulate a data set with the dplyr package, make and discuss three findings, and support those findings with nice bar and line plots from ggplot2. You should work independently. Use the data “4_Flash_Proj_1_Data.csv” which includes seven variables: country, continent, year, lifeExp (life expectancy in years), gdpPercap (GDP per capita in \$US), gdpBillions (GDP in billions of \$US), and popThous (population in thousands). Submit your findings in an HTML file created by RMarkdown and your full code by 11:00 AM Wednesday, Oct 12, according to the directions below. I recommend starting simple then building from there—don’t start by trying to do a complex analysis with the perfect write up.

Your Task:

1. Explore and manipulate the data in dplyr. I want you to use a number of different dplyr functions to subset, summarise, group and make new variables—be adventurous! Use exploratory visualizations in either base or ggplot2, whatever works best.
2. Determine your three most interesting findings and produce a clear, well titled and labeled graph in ggplot2 to support each finding. Make sure to follow the guidelines we have discussed in class. I expect you to have 3 nicely formatted graphs—at least one bar graph and one line graph.
3. Write up a five paragraph summary of the project: an introduction, a paragraph and nicely formatted graph for each finding and a conclusion/summary in RMarkdown with just text, tables and graphs (no code) and save as an HTML.
4. Submit your full code and your HTML. Make sure your code is clear and documented, you may want to label sections by the question you are exploring.

Data: 4_Flash_Proj_1_Data.csv

Software: Use R (and [Rstudio](#)).

Flash Project Grade Rubric		
Category - Percent of grade	Score	Comments
Introductory Paragraph – 2 points Brief findings, clear, no jargon		
Finding 1 – 4 points Description, Stat, Visual, no jargon		
Finding 2 – 4 points Description, Stat, Visual, no jargon		
Finding – 4 points Description, Stat, Visual, no jargon		
Conclusion – 1 point Clear, Stat, no jargon		
Code – 10 points Clear, labeled, complete, documented, variety of commands, appropriate use of techniques		
Total	/25	