# K304 project 3/4: Basic summary measures and graphs; R for data wrangling, summarizing, and graphing

# 50 points

## File submission requirements

Submit your R script file; nothing else.

Your work must be submitted via Canvas by 10:59 p.m. on November 20, 2020. Submissions uploaded to Canvas before 11:59 p.m. will be graded without penalty.

After 11:59 p.m., submissions can be made to the late submission assignment on Canvas. See the syllabus for more details regarding penalties.

Work not submitted via Canvas will not be graded.

## The Doing Business dataset and methodology

In this project you will analyze data collected in the Doing Business survey over the years 2004-2018. Doing Business is a global study, run by the World Bank, that aims to assess the ease of doing business in different economies through measures such as the number of procedures required to start a business, the thoroughness of credit monitoring, the extent to which investors are protected, and so on.

For each survey topic (e.g., Starting a Business), business and legal practitioners in the economies being surveyed are asked to answer questions about a single, detailed set of circumstances. Independent experts are then used to validate their responses. The use of a single, standard example for each topic area allows results to be compared across economies. Data regarding costs are likewise collected in a manner allowing for comparison across economies; for example, the cost of starting a business is reported as a percentage of the economy’s per-capita income, the idea being to represent the relative ease or difficulty of starting a business for a person living and earning wages in that economy.

More information about the study, methodology, the data variables gathered, and the cases created to gather those variables is available at [www.doingbusiness.org](http://www.doingbusiness.org).

## Problem set

See the R script file, and complete the problems for this project using the packages mentioned (dplyr, ggplot2, ggrepel, and ggthemes).

Most of what you will need to complete this project can be found in the class exercises. There will likely be a few things that you need to research. When you do, for best results, include either “dplyr” or “ggplot” in your search terms, depending on where the question lies.

Sketches and notes for the reporting objects of interest are included with the project materials. Here’s how to think of these—

* The sketches are rough, not exact, representations of the desired reporting objects—these would have been developed either by or in collaboration with the client/management, which here might be the World Bank. In an ideal world, you (the team of analysts working on this project) would have gotten the client/management to sign off on these sketches before you started in-program development work.
* The post-it notes are from Aaron and Sarah. (Your client is not likely to know about the ggthemes package, for example, or that there is an economist theme, or a tufte theme, or a fivethirtyeight theme.)