Names: Sami Bajwa, Omid Nikjoo, Cristian Rodriguez
Stage 1:
Due: Friday 6/25, 8pm

1. Please submit a draft report that describes the data wrangling steps already completed [PDF]
2. A listing of your R code in one file [.R file]

1. The data wrangling steps completed consists of us doing a few things.
   First of all, in part B we did 2 steps. For the covid19 dataset we used the “pivot_longer” function on all columns excluding the first four columns to condense columns 5-50 into two columns consisting of a date and the amount of cases that day. We also created a new variable in part B that properly shows the confirmed deaths across the specific days.

   For part C we tidied up the covid19 dataset again by using a “group_by” and “summarise” to have a table that displays the Country/Region and the max amount of cases it has had. We did the same thing with the confirmed deaths dataset, and we renamed the country name variables in hospitalbeds and demographics datasets so they match the covid19 dataset, making it easier to use inner_join in the future.

   For part D, to create a new column in the covid19 dataset showing the death rate, we did an inner_join between the covid19 dataset and the dataset with confirmed covid19 deaths, and then we used the mutate function to create a new variable containing the death rate, calculated by using deaths/cases.

   For part E we tidied up the hospitalbeds dataset by using “group_by” and “filter” functions so that it clearly shows the country and the recent max amount of beds available for the country and we renamed the column to beds .

   For part F we changed the NA values to zero so we can be able to add all of the male and female population numbers together in the demographics dataset using the “mutate” function more easier and removed any unnecessary columns and we changed the values that were zero to NA values so we could drop those NA rows.

   And finally we did an “inner_join” with the covid19, hospitalbeds, and demographics datasets so that they are all in one table so it can be ready for linear regression.