# Homework 6 – Intro to Probability and Statistics

Your name here

## Instructions:

**Due:** 05/21 at 11:59PM.

**What am I expecting?** An R Markdown with the answers.

**Have fun!**

The Brazilian Ministry of Health keep records of the daily number of Coronavirus cases, by dates and municipalities. The dataset can be found here: (https://covid.saude.gov.br/).

### Question 1

Take some time to familiarize yourself with the data. After that, please provide here a brief explanation of the variables and how the dataset is organized.

### Question 2

I saved the May 15 version of the dataset on the class GitHub, under the name `brazilcorona.csv`. Load the dataset in your computer and put a `head` of the dataset here.

### Question 3

Note that the dataset has the aggregated results by country, region, and state. Exclude all these aggregated results, keeping only the cases by the municipality. After that, exclude all cases before May 15. Put a `head` of the filtered dataset here. What is the dimension of the filtered dataset?

### Question 4

Use the municipal population to compute the average infections rate (the number of infections divided by population times 100000) and mortality rate (the number of casualties divided by municipal population times 100000). For each state, which municipalities have the lowest and the highest infections and death rates? Create a new dataset with the average infections and mortality rates by municipality.

### Question 5

In the dataset `brmayors.csv`, there are the party, gender, and schooling of the Brazilian mayors elected in 2016. I downloaded this dataset from the `CEPESP Data` repository (http://cepespdata.io/). Explore the variables making tables and/or barplots. Provide a brief interpretation of the results.

### Question 6

Merge this dataset with the mortality and infections rates dataset created in the previous question. Put a `head` of the new dataset.

## Question 7

Some experts showed that places ruled by women have lower mortality rates than areas ruled by men. Test this hypothesis using regression analysis. Provide a brief interpretation of your results.

## Question 8

Note that most municipalities have no cases of the disease. What happens if you put zero in the places that had no cases, and run the regression in *question 6*? Provide a brief interpretation of your results.

## Question 9

Create a variable that codes the ideological positions, based on the politician's parties. Use the three-way category: `left`, `center`, and `right`. Justify your choices. Provide a table with the frequency by each category. Run a regression for the new variable, to investigate whether ideological position influences the COVID casualties and infections. Provide a brief interpretation.

## Question 10

Using `ifelse` creates a variable `collegedegree`, which is `Yes` when the mayor has a college degree, and `No` otherwise. Run a regression for the new variable, to investigate whether education influences the COVID casualties and infections. Provide a brief interpretation.