Lecture 2 – HOMEWORK

- GO TO: http://tutorials.iq.harvard.edu/R/Rintro/Rintro.html and find similar examples there. This page is very helpful for your homework.
  There are some examples and information there. You can try them in R just to have fun
- Our homework is about the baby names! I would like to know what the most popular baby names are. In the course of answering this question we will learn to call R functions, install and load packages, assign values to names, read and write data, and more.
  (Note: The examples in this workshop use the baby names data provided by the governments of the United States and the United Kingdom. A cleaned and merged version of these data is in dataSets/babyNames.csv.)
- Our first goal is to read these data into R. In order to do that we need to learn how to *call functions*, *install packages*, set out *working directory*, *read* as .csv file, and *assign* the result to a name. Lets get to it.

1. First, set your working directory
2. Install and load the tidyverse and stringr packages
3. Go to http://tutorials.iq.harvard.edu/R/Rintro/dataSets/ and see the data set already saved as a csv file. You should call it from R and save it as a new data frame called "baby_names" like:

baby_names <- read_csv ("http://tutorials.iq.harvard.edu/R/Rintro/dataSet/ babyNames.csv")

Note: it takes a very long time since it is a really big data. Be patient

1. You alreadry have a "Percent" column in your dataframe. Create a new column called "Proportion" which is equal to Percent divided by 100.
2. There are 20 years of data saved in the csv file. Filter baby_names that are given in 2015 and save it as a data frame called "baby_names_2015" like:
   baby_names_2015 <- …..
3. Filter the data to extract only Massachusetts (Location "MA"), and calculate the total number of children born in Massachusetts in 2015.
4. Filter baby names to show only names given to at least 5 percent of new-borns in 2015 and then save it as a new data frame called "popular_names_2015".
5. Save that data frame as a csv file called "popular_names_2015.csv" .
6. Calculate the total number of new-borns in 2015.
7. Calculate the total number of babies named as "zeynep" in all years.

Ok. This is enough. We will do the rest next week in the class:

8. Create a new column called "name_length" and write the length of names to that column
9. Find the average length of names and plot them by year. Has the average length of names changed over time?