# Project 2: Data Preparation & Transformation Pt. I

*Albert Gilharry*

*March 8, 2018*

## Contents

## Data Set III - Child Mortality

This data set was posted by **Raj Kumar**. Childmortality.org is an organization that publishes Child Mortality estimates for all the countries around the world. They provide all available data and the latest child mortality estimates for each country based on the research of the UN Inter-agency Group for Child Mortality Estimation.

The data is in a very wide format and contains six variables with values of interest. Each of these variables are concatinated to each year from 1950 to 2016, resulting in 405 columns. These key variables are

Under-5 (0-4 years) mortality Infant (0-1 years) mortality Neonatal (0-1 month) mortality Number of under-5 deaths Number of infant deaths Number of neonatal deaths

As suggested by Raj, We can read this data into a data frame and subset the data set to the median estimate for each country. We need to also handle null values of the data. Then we can convert this data into long format with 4 variables country, year, category and their respective value. This will make it easier to analyze the data."

My goals with this data set are as follows:

(-) **Load, Tidy, and transform the data.** (-) **Map child mortality for a given year** (-) **Create a time series plot of child mortality.**

## Load Libraries

```r
library("tidyverse")
```

```
## -- Attaching packages -------------------------------------------------------
## v ggplot2 2.2.1      v purrr   0.2.4
## v tibble  1.4.2      v dplyr   0.7.4
## v tidyr   0.8.0      v stringr 1.2.0
## v readr   1.1.1      v forcats 0.2.0

## -- Conflicts ----------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library("stringr")
library("DT")
library("rworldmap")
```

```
## Loading required package: sp
```

```
## ### Welcome to rworldmap ###

## For a short introduction type :   vignette('rworldmap')
```

## Load Data

```r
data <- read.csv("data/ChildMortality.csv", sep = ",", header = TRUE, stringsAsFactors = FALSE)

data <- gather(data, "YearCat", "Rate", 4:405)

data$Category <- str_sub(data$YearCat,1,-6)
data$Year <- strtoi(str_sub(data$YearCat, -4))

d <- filter(data, `Category` == "U5MR", `Uncertainty.bounds.` == "Lower", Year == 2000)

data(countryExData)
sPDF <- joinCountryData2Map( d,joinCode = "ISO3", nameJoinColumn = "ISO.Code")
```

```
## 195 codes from your data successfully matched countries in the map
## 0 codes from your data failed to match with a country code in the map
## 48 codes from the map weren't represented in your data
```

```r
mapDevice() #create world map shaped window
mapCountryData(sPDF,nameColumnToPlot='Rate')
```

```r
d <- filter(data, `Category` == "U5MR", `Uncertainty.bounds.` == "Lower", `ISO.Code` == "NLD" | `ISO.Co
ggplot(d, aes(x = Year, y = Rate)) +
  geom_line(aes(color = `ISO.Code`), size = 1) +
  #scale_color_manual(values = c("#00AFBB", "#E7B800")) +
  theme_minimal()
```

```
## Warning: Removed 19 rows containing missing values (geom_path).
```