

# Data Cleaning - Assignment 7

*Bryana Benson*

*Conner Bryan*

*2/5/2019*

## Libraries and Data

```
library(readr)
library(stringr)
library(dplyr)
library(janitor)
library(stringr)
library(knitr)
```

## Read the data file

```
data_clean<-read.csv("cleaning.csv")
```

Above is the data preview.

## Clean the file - convert to numeric and normalize city names

```
data_clean$Zip <- str_replace(data_clean$Zip, "631o8", "63108")

data_clean$Income<-as.numeric(data_clean$Income)

data_clean$City<-case_when(
  data_clean$Zip == "63108" ~ "Kirkwood",
  data_clean$Zip == "63122" ~ "Kirkwood",
  data_clean$Zip == "63123" ~ "Kirkwood",
  data_clean$Zip == "63125" ~ "Kirkwood",
  data_clean$Zip == "63126" ~ "Kirkwood",
  data_clean$Zip == "63127" ~ "Kirkwood",
  data_clean$Zip == "63128" ~ "Kirkwood",
  data_clean$Zip == "63131" ~ "Kirkwood",
  data_clean$Zip == "63108" ~ "Saint Louis",
  data_clean$Zip == "63110" ~ "Saint Louis",
  data_clean$Zip == "63124" ~ "U City",
  data_clean$Zip == "63130" ~ "U City",
  data_clean$Zip == TRUE ~ 'NA'
)
```

## Filter out NAs and calculate average income by city

```
filteredData <- filter(data_clean, City!= 'NA')

meanIncome<-filteredData%>%
  group_by(City)%>%
  summarize(mean=round(mean(Income, na.rm=TRUE),0))
```

## Creating a well formatted table

```
kable(meanIncome, col.names=c("City", "Mean Salary"))
```

City	Mean Salary
Kirkwood	533
Saint Louis	559
U City	612