Assignment 1: EDA

GG Force (6870945,)

23-11-2020

```r
library(ISLR)
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

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

```r
# popular_apps <-
#   read_csv("data/googleplaystore.csv") %>%
#   mutate(Downloads = parse_number(Installs)) %>%
#   filter(Downloads > 500000000) %>%
#   arrange(-Rating) %>%
#   distinct(App, .keep_all = TRUE) %>%
#   select(App,Rating,Category, Downloads)
#
# popular_apps
```

```r
cwur <- read_csv("data/cwurData.csv")
```

```
##
## -- Column specification ---------------------------------------------------------
## cols(
##   world_rank = col_double(),
##   institution = col_character(),
##   country = col_character(),
##   national_rank = col_double(),
##   quality_of_education = col_double(),
##   alumni_employment = col_double(),
##   quality_of_faculty = col_double(),
```

```
##   publications = col_double(),
##   influence = col_double(),
##   citations = col_double(),
##   broad_impact = col_double(),
##   patents = col_double(),
##   score = col_double(),
##   year = col_double()
## )

## Warning: 4 parsing failures.
##  row         col              expected actual                 file
## 1010 institution delimiter or quote      A 'data/cwurData.csv'
## 1010 institution delimiter or quote      B 'data/cwurData.csv'
## 1829 institution delimiter or quote      A 'data/cwurData.csv'
## 1829 institution delimiter or quote      O 'data/cwurData.csv'
```
```
View(cwur)
```

Explanation of the Data Set

For this assignment we have used the World University Ranking data set found on kaggle.
(https://www.kaggle.com/mylesoneill/world-university-rankings?select=cwurData.csv)

The data set contains various information about three World University Rankings: The
Times Higher Education Ranking, the Shanghai Ranking, and the lesser known Center for
World University Rankings. Comparing these different data sets we can see if there are
disambiguities in their rankings or if they lean towards specific measures or insitutions.

For simiplicity we generally focused on the CWUR rankings. The Center for World University
Rankings (CWUR) organisation enlists universities out of more than 20,000 universities
globally, and provides information about: the ranking globally, ranking nationally, country,
quality of education, employment of alumni, publications, influence, citations and of course
the score for years 2012-2015. The data was generally well structured and contains only a
few NA's in the broad_impact column.

Tables of Relevant Summary Statistics

Descriptive Plots

Including Plots

Explanation