Initial Analysis

Akshi Chaudhary

2017-12-03

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.2
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
```

```
## Warning: package 'ggplot2' was built under R version 3.4.2
```

```
## Warning: package 'tibble' was built under R version 3.4.2
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
## Warning: package 'readr' was built under R version 3.4.2
```

```
## Warning: package 'purrr' was built under R version 3.4.2
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(ggplot2)
```

```
shang <- read_csv('../data/raw/shanghaiData.csv', trim_ws = TRUE)
times <- read_csv('../data/raw/timesData.csv', trim_ws = TRUE)
cwu <- read_csv('../data/raw/cwurData.csv', trim_ws = TRUE)
expenditure <- read_csv('../data/raw/education_expenditure_supplementary_data.csv', trim_ws = TRUE)
```

```
times %>% filter(university_name == 'University of British Columbia')
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.2
```

```
## # A tibble: 6 x 14
##   world_rank                university_name country teaching international
##        <chr>                          <chr>   <chr>    <dbl>         <chr>
## 1         30 University of British Columbia  Canada     65.1          93.3
```

```
## 2         22 University of British Columbia  Canada     68.6          88.7
## 3         30 University of British Columbia  Canada     69.1          82.1
## 4         31 University of British Columbia  Canada     59.9          84.2
## 5         32 University of British Columbia  Canada     60.5          84.8
## 6         34 University of British Columbia  Canada     60.2          90.5
## # ... with 9 more variables: research <dbl>, citations <dbl>,
## #   income <chr>, total_score <chr>, num_students <dbl>,
## #   student_staff_ratio <dbl>, international_students <chr>,
## #   female_male_ratio <chr>, year <int>
```

```
#shang %>% filter(university_name == 'University of British Columbia')
```

```
#cwu %>% filter(institution == 'University of British Columbia')
```

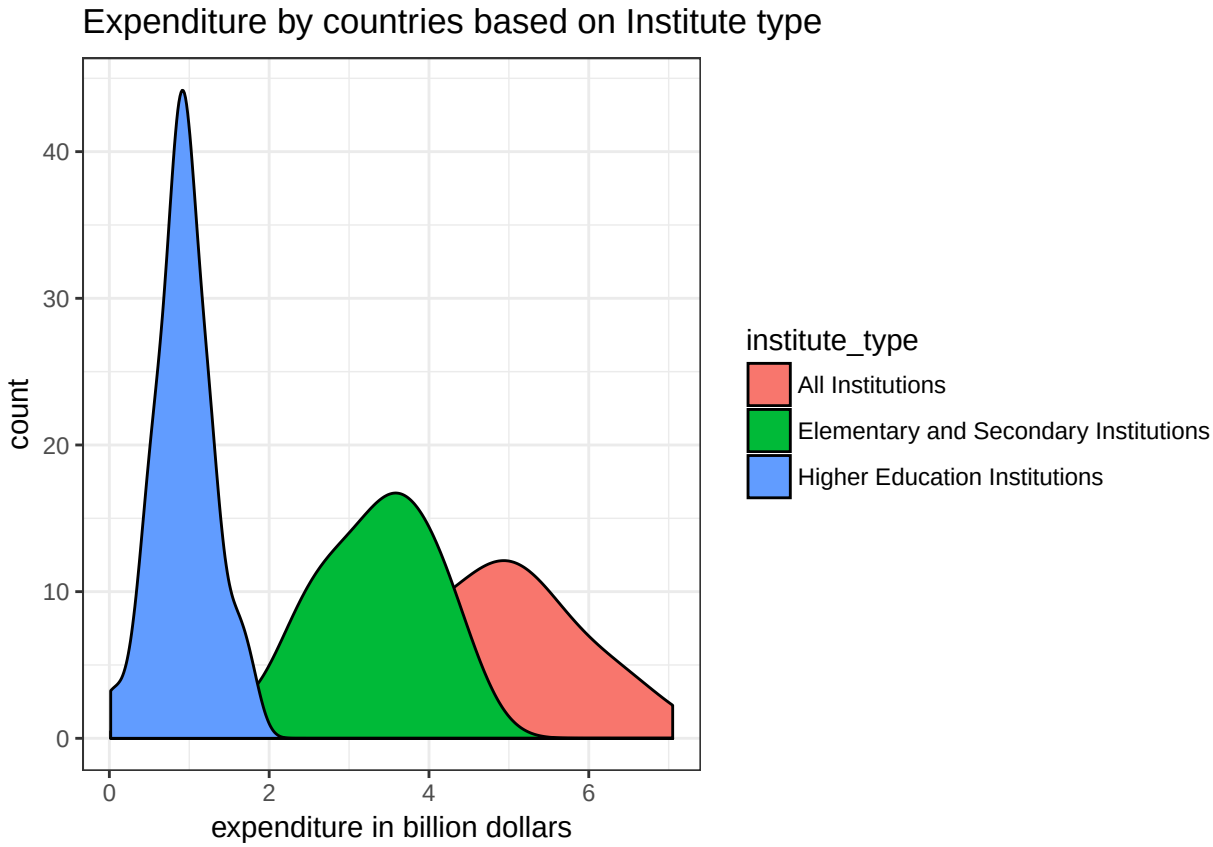
```
country_score <- times %>% filter(total_score != '') %>% group_by(country) %>% summarise(best_score = ma
```

```
colnames(expenditure)[4] <- "y1995"
colnames(expenditure)[5] <- "y2000"
colnames(expenditure)[6] <- "y2005"
colnames(expenditure)[7] <- "y2009"
colnames(expenditure)[8] <- "y2010"
colnames(expenditure)[9] <- "y2011"

expenditure
```

```
## # A tibble: 333 x 9
##           country   institute_type direct_expenditure_type y1995 y2000
##             <chr>            <chr>                   <chr> <dbl> <dbl>
##  1   OECD Average All Institutions                  Public   4.9   4.9
##  2      Australia All Institutions                  Public   4.5   4.6
##  3        Austria All Institutions                  Public   5.3   5.4
##  4        Belgium All Institutions                  Public   5.0   5.1
##  5         Canada All Institutions                  Public   5.8   5.2
##  6          Chile All Institutions                  Public    NA   4.2
##  7 Czech Republic All Institutions                  Public   4.8   4.2
##  8        Denmark All Institutions                  Public   6.5   6.4
##  9        Estonia All Institutions                  Public    NA    NA
## 10        Finland All Institutions                  Public   6.6   5.5
## # ... with 323 more rows, and 4 more variables: y2005 <dbl>, y2009 <dbl>,
## #   y2010 <dbl>, y2011 <dbl>
```

```
expenditure <- expenditure %>% mutate(y1995 = ifelse(is.na(y1995),0,y1995)
                                      ,y2000 = ifelse(is.na(y2000),0,y2000)
                                      ,y2005 = ifelse(is.na(y2005),0,y2005)
                                      ,y2009 = ifelse(is.na(y2009),0,y2009)
                                      ,y2010 = ifelse(is.na(y2010),0,y2010)
                                      ,y2011 = ifelse(is.na(y2011),0,y2011))

school_exp <- expenditure %>% filter(direct_expenditure_type !='Total') %>%mutate(avg_exp = (y1995+ y200
```

```
school_exp %>% filter(direct_expenditure_type == 'Public') %>% ggplot(aes(total_exp,fill = institute_typ
```

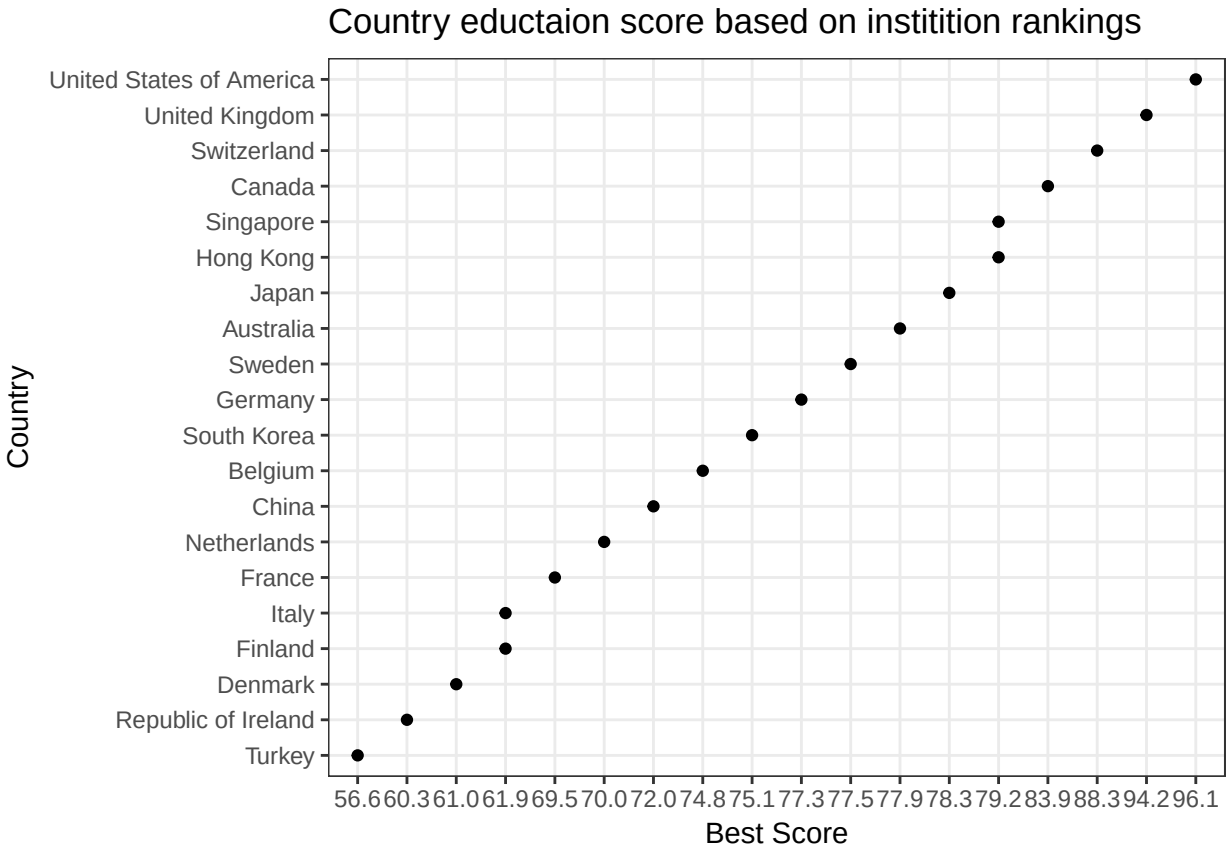



```
library(forcats)
```

```
## Warning: package 'forcats' was built under R version 3.4.2
```

```
country_score$country <- factor(country_score$country , levels = country_score$country [order(country_s

country_score %>% top_n(20) %>% arrange(desc(best_score)) %>% ggplot(aes(x = best_score, y = country)) +
        axis.title=element_text(size=14,face="bold")) +theme_bw()
```

```
## Selecting by best_score
```


Country eductaion score based on institition rankings
United States of America
United Kingdom
Switzerland
Canada
Singapore
Hong Kong
Japan
Australia
Sweden
Germany
South Korea
Belgium
China
Netherlands
France
Italy
Finland
Denmark
Republic of Ireland
Turkey
Country
56.6
60.3
61.0
61.9
69.5
70.0
72.0
74.8
75.1
77.3
77.5
77.9
78.3
79.2
83.9
88.3
94.2
96.1
Best Score