10/31/2021

World University Rankings

The World Universities Rankings is utilizing information about Universities Rankings from various sources: Center for World University Rankings (CWUR), Shanghai University Rankings (AWUR), Times Higher Education University Rankings (THE). Every one of these sources is utilizing an alternate hierarchisation framework. Other than the World University Ranking frameworks, there are additionally data about the schools and nations, about the instruction use (per nation and year) and training fulfillment (per nation and year, various dimensions).

Datasets used from World University Rankings

```
timesData<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/timesData.csv')
shanghaidata<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/shanghaiData.csv')
school_country<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/school_and_country_table.csv')
cwurData<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/cwurData.csv')
e_a_s_data<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/educational_attainment_supplementary_data.
csv')
educationExpenditure<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/education_expenditure_supplement
ary_data.csv')
```

```
library(tidyverse)
library(dplyr)
library(kableExtra)
library(ggplot2)
library(formattable)
library(plotly)
library(gtable)
library(grid)
library(gridExtra)
library(tidyr)
library(magrittr)
library(stringr)
library(lubridate)
library(lemon)
library(knitr)
library(corrplot)
library(tidyverse)
library(reshape2)
library(tidyverse)
library(readr)
```

Summary of the data

We have 6 data files in University Rankings dataset, as following: * timesData.csv (2603 rows, 14 columns) * cwurData.csv (2200 rows, 14 columns) * education_expenditure_supplementary_data.csv (333 rows, 9 columns) * educational_attainment_supplementary_data.csv (79055 rows, 29 columns) * school_and_country_table.csv (818 rows, 2 columns) * shanghaiData.csv (4897 rows, 11 columns)

```
count(timesData)
```

```
##      n
## 1 2603
```

```
count(cwurData)
```

```
##      n
## 1 2200
```

```
count(educationExpenditure)
```

```
##     n
## 1 333
```

```
count(e_a_s_data)
```

```
##       n
## 1 79055
```

```
count(shanghaidata)
```

```
##      n
## 1 4897
```

Center for World University Rankings (CWUR)

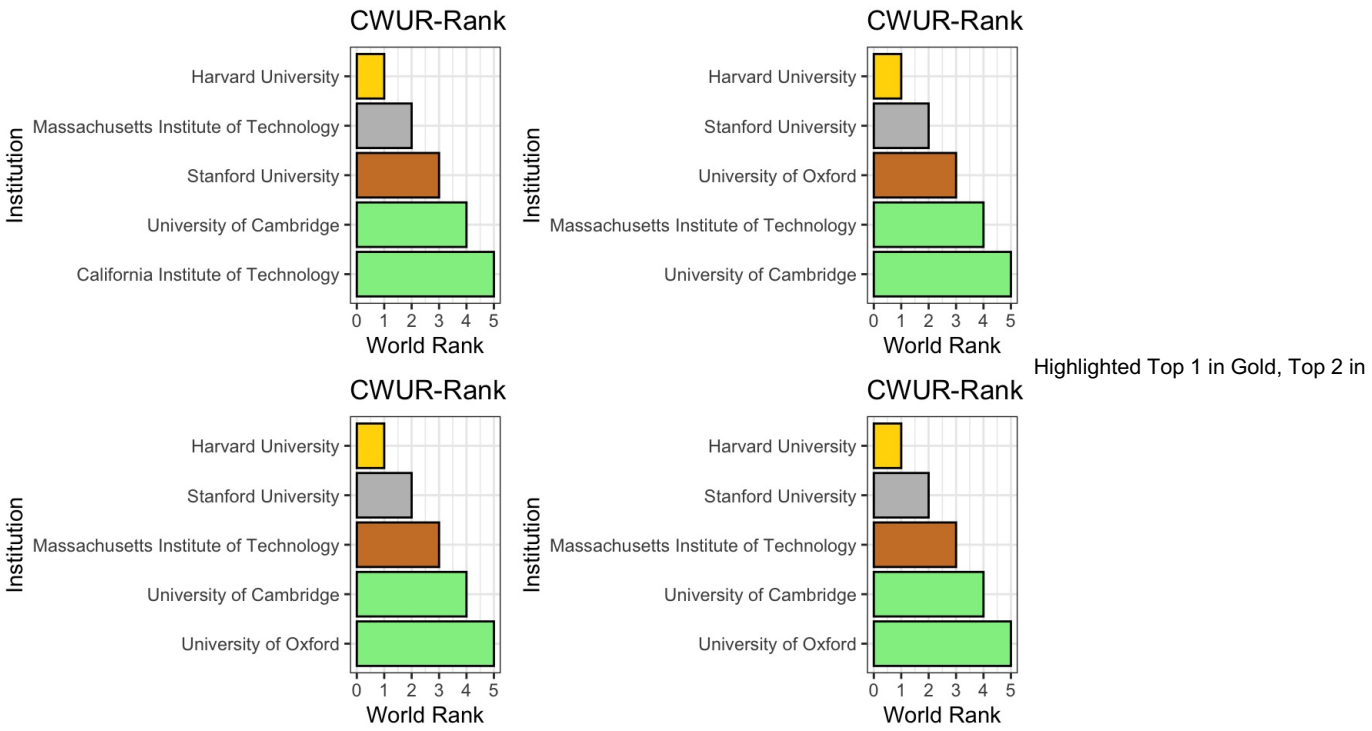
The Center for World University Rankings provides information for top universities in the World.

They use an overall rank per university and individual values for the National Rank, Quality of Education, Alumni Employment, Quality of Faculty, Publications, Influence, Citations, Broad Impact, Patents, and an Total Score. All these values are for a certain year. For example, Harvard University has the first position for 2012, with 100.00 overall score and ranks 1 for most of the dimensions.

Best Universities

Let's see the top 5 Universities for the years 2012-2015, according to CWUR

```
cwur <- function(nYear) {
  cwurData %>% filter(year==nYear) %>% top_n(5,-world_rank) %>%
  ggplot(aes(x=reorder(institution,-world_rank), y=world_rank)) + geom_bar(stat="identity", aes(fill=reorder(inst
itution,-world_rank)), colour="black") +
    theme_bw() + coord_flip() +  scale_fill_manual(values=c(rep("lightgreen",2), "#CD7F32", "grey", "gold")) + gu
ides(fill=FALSE) +
    labs(x="Institution", y="World Rank",
        title=paste("CWUR-Rank in ",nYear))
}

cwur(2012) -> d1
cwur(2013) -> d2
cwur(2014) -> d3
cwur(2015) -> d4
grid.arrange(d1,d2,d3,d4, ncol=2)
```

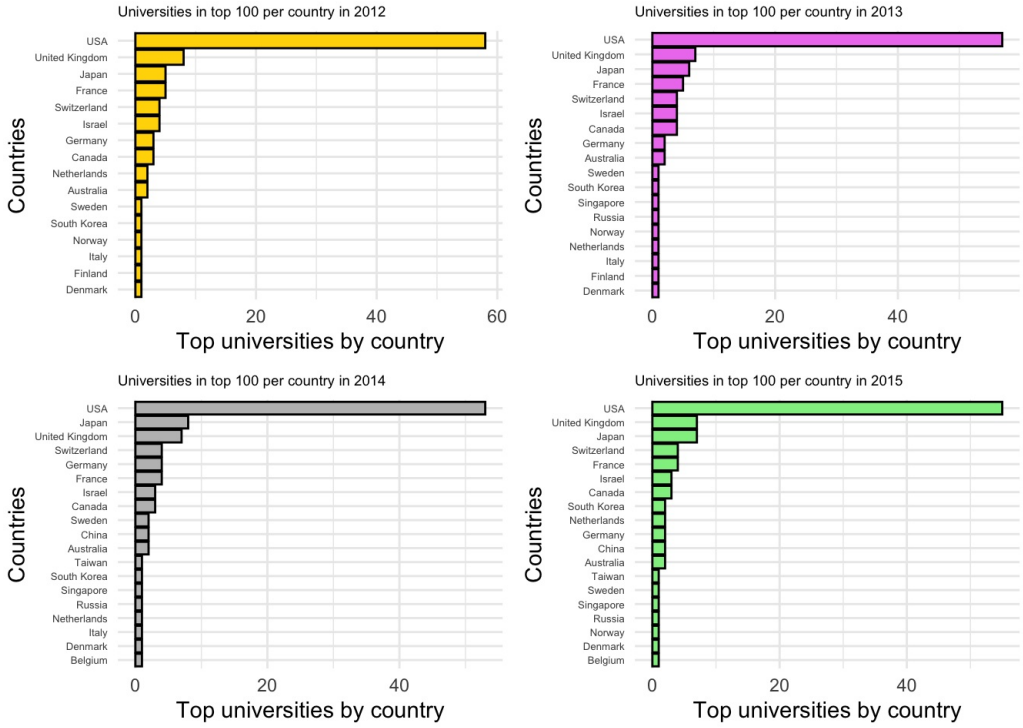


Highlighted Top 1 in Gold, Top 2 in silver and Top 3 in Bronze.

Harvard is on the first place on all the years, followed by MIT in 2012 and Stanford in 2013-2015; the third place is changing between Stanford (2012), Oxford (2013), with MIT holding it for 2014 and 2015.

Q) How many universities(top 100) are there in the world by country in 2012,2013,2014 and 2015?

```r
#Top 100 Universities per Country using cwur Ranking
library(gridExtra)
cwur<-read.csv('/Users/diyasree/Downloads/Diya/IE 5374/Project 1/cwurData.csv')

A <- cwur %>% filter(year == 2012,world_rank <= 100) %>%
     select(country) %>% table() %>% data.frame()
B <- cwur %>% filter(year == 2013,world_rank <= 100) %>%
  select(country) %>% table() %>% data.frame()
C <- cwur %>% filter(year == 2014,world_rank <= 100) %>%
  select(country) %>% table() %>% data.frame()
D <- cwur %>% filter(year == 2015,world_rank <= 100) %>%
  select(country) %>% table() %>% data.frame()

names(A)[names(A) == '.'] <-"country"
names(B)[names(B) == '.'] <-"country"
names(C)[names(C) == '.'] <-"country"
names(D)[names(D) == '.'] <-"country"

p1<-A %>% ggplot(aes(x =reorder(country,Freq), y=Freq), fill=country) +
 geom_bar(stat="identity", fill="gold", colour="black", position=position_dodge(0.1))+ theme_bw() +
  ggtitle("Universities in top 100 per country in 2012")+
  labs(x="Countries", y="Top universities by country")+
  theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 5))

p2<-B %>% ggplot(aes(x =reorder(country,Freq), y=Freq), fill=country) +
 geom_bar(stat="identity", fill="violet", colour="black", position=position_dodge(0.1))+ theme_bw() +
  ggtitle("Universities in top 100 per country in 2013")+
  labs(x="Countries", y="Top universities by country")+
  theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 5))

p3<-C %>% ggplot(aes(x =reorder(country,Freq), y=Freq), fill=country) +
 geom_bar(stat="identity", fill="grey", colour="black", position=position_dodge(0.1))+ theme_bw() +
  ggtitle("Universities in top 100 per country in 2014")+
  labs(x="Countries", y="Top universities by country")+
  theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 5))

p4<-D %>% ggplot(aes(x =reorder(country,Freq), y=Freq), fill=country) +
 geom_bar(stat="identity", fill="lightgreen", colour="black", position=position_dodge(0.1))+ theme_bw() +
  ggtitle("Universities in top 100 per country in 2015")+
  labs(x="Countries", y="Top universities by country")+
  theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 5))

grid.arrange(p1,p2,p3,p4,ncol=2)
```



As shown above we have presented the bar graph of the universities in top 100 per country for the years 2012, 2013, 2014 and 2015. This graphs gives the detailed information as mentioned in the dataset for comparing the university ranking in different countries from 2012 to 2015.

According to the observation we can say that the most of the top universities are from USA(United States of America). There is a huge rise of the ranking universites in Japan comparred from 2012 to 2015. A similar pattern has been observed in the change of ranking in universities of Netherland.

Shanghai Academic Rankings for World Universities (ARWU)

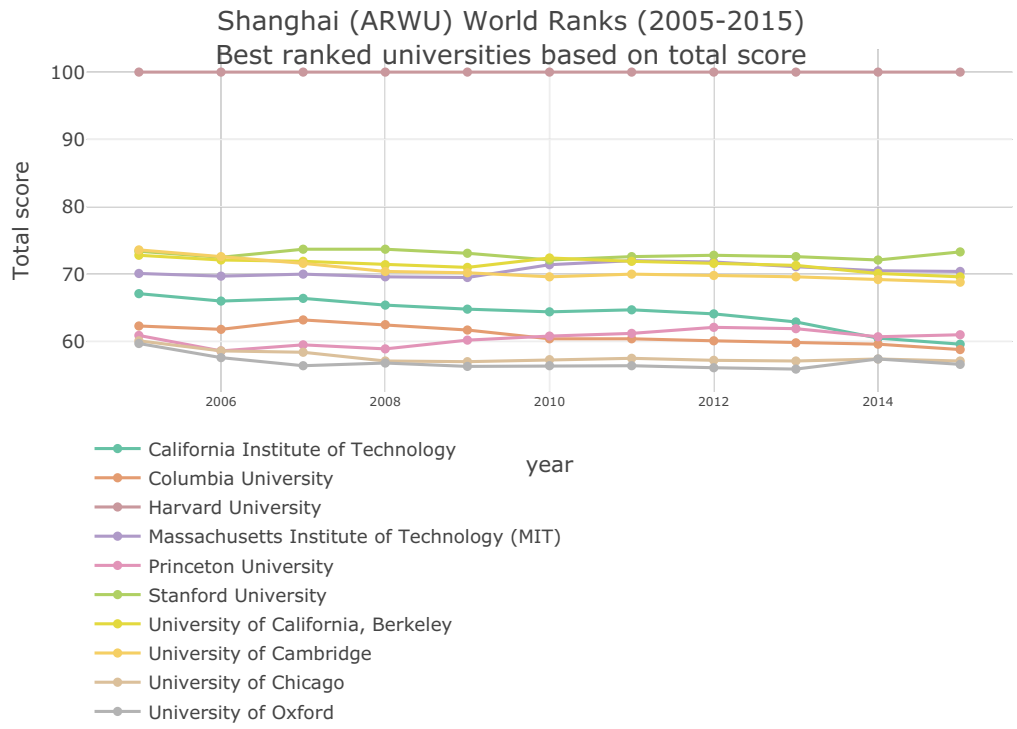
The Shanghai World Rankings for colleges have a World Rank (more modest worth is awesome), a National Rank (per country, more modest worth is ideal) and scores from 0 to 100

Total Score:

Alumni of an institution winning Nobel Prizes and Fields Medals (alumni); Staff of an institution winning Nobel Prizes and Fields Medals (awards); Highly Citated researchers (hici); Number of papers in Nature & Science papers (ns); Papers indexed in Science Citation Index-expanded and Social Science Citation Index (pub); Per capita academic performance of an institution (pcp); All these values are for a corresponding year between 2005 and 2015.

Q) What are the best ranked universities based on total score ?

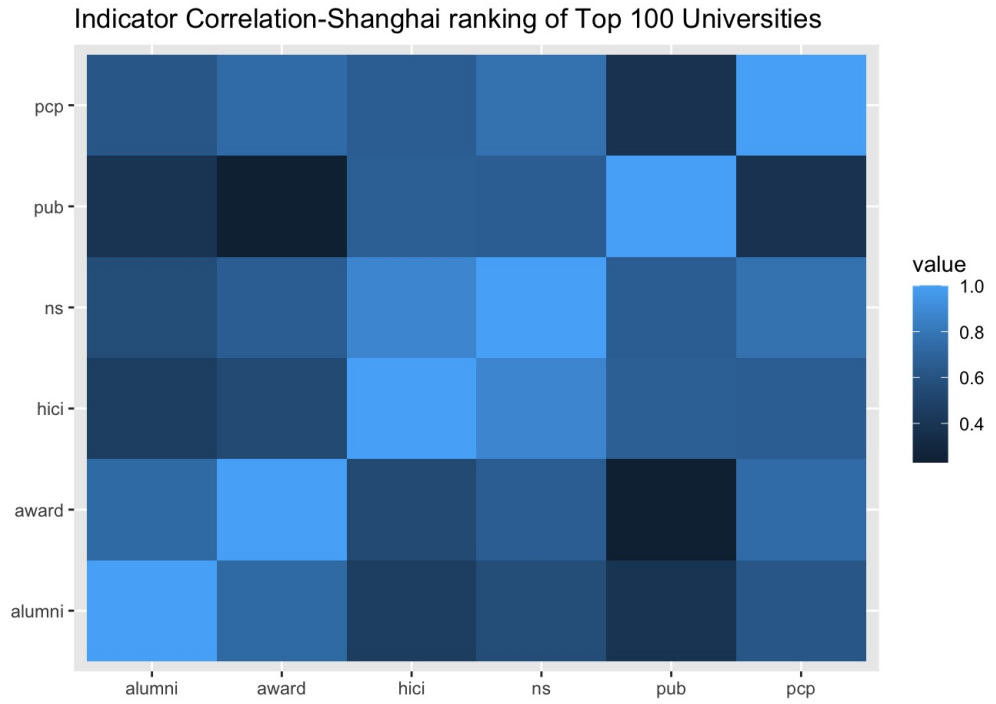
```
shanghaiDataCld <- shanghaidata

#Calculate the total score
shanghaiDataCld$t_score =
  0.1 * shanghaiDataCld$alumni + 0.2 * shanghaiDataCld$award + 0.2 * shanghaiDataCld$hici +
  0.2 * shanghaiDataCld$ns + 0.2 * shanghaiDataCld$pub + 0.1 * shanghaiDataCld$pcp

shanghaiDataCld$total_score[is.na(shanghaiDataCld$total_score)] = shanghaiDataCld$t_score[is.na(shanghaiDataCld$t
otal_score)]
#eliminate the additional data

shanghaiDataCld <- shanghaiDataCld[complete.cases(shanghaiDataCld),]
#Fix the duplicate name for University of California-Berkeley

shanghaiDataCld$university_name[shanghaiDataCld$university_name=="University of California-Berkeley"] <- "Univers
ity of California, Berkeley"

shanghaiDataCld %>% group_by(year) %>%
  top_n(10, wt = total_score) %>% select(year,university_name,total_score,alumni, award, hici, ns, pub, pcp) %>%
ungroup() -> top10univ


plot_ly(top10univ, x = ~year) %>%
  add_trace(y = top10univ$total_score, name = top10univ$university_name, showlegend=TRUE, type = 'scatter', mode
= 'lines+markers', color= top10univ$university_name) %>%
  layout(title="Shanghai (ARWU) World Ranks (2005-2015)<br>Best ranked universities based on total score", legend
= list(orientation = 'h'),
         xaxis = list(showticklabels = TRUE, tickangle = 0, tickfont = list(size = 8)),
         yaxis = list(title = "Total score"),
         hovermode = 'compare')
```



The same 10 universities (University of California at Berkley appears with 2 different entries) appears in top 10. Harvard University is leading with a flat 100 points as total score.

Q) How alumni score of an institution winning nobel prizes and field medals, awards score of an institution winning nobel prizes in Physics-Chemistry-Medicine, HiCi score of highly cited researchers, N&S score of papers published in nature & Science, PUB score of papers indexed in the science citation Index-Expanded and Social Science Citation and PCP score of number of Full time academic staff are correalted?

```
Top100_shan<-head(shanghaidata,100)

shanghai_corr<- Top100_shan %>%
select(alumni,award,hici,ns,pub,pcp) %>%
drop_na() %>%
na.omit() %>%
na_if("-")

cormat_shan <- round(cor(shanghai_corr),2)
melted_cormat_shan <- melt(cormat_shan)
ggplot(data = melted_cormat_shan, aes(x=Var1, y=Var2, fill=value)) +
        geom_tile()+
    ggtitle("Indicator Correlation-Shanghai ranking of Top 100 Universities")+
  xlab("") +
  ylab("")
```



From the above correlation plot of Shanghai Ranking: Exploration capacity and instruction quality are correlated. Scholastic per capita perfomance is correlated with research capacity. Nature of workforce are additionally correlated with research capacity. Since Shanghai Ranking spotlight vigorously on distributed papers, investigates and global honors, these solid correlations can be anticipated.

Times Higher Education (THE) University Rankings

The Times Higher Education data has information from 2011 to 2016. There is a World Rank (smaller values are better) and coeficients (between 0 and 100) for:

Teaching (teaching/the learning environment); International Outlook (international/staff, students and research); Research (research/volume, income and reputation); Citations (citations/research influence); Industry Income (income/knowledge transfer, value not available for all); Overall (total_score).

Q) Which universities in Top 10 has the highest student to teacher ratio & Quality of Research, Teaching, Citations according to world ranking-times higher education?

```
top10university<-head(timesData,10)
top10university$student_staff_ratio<-as.numeric(as.character(top10university$student_staff_ratio))
top10university$teaching<-as.numeric(as.character(top10university$teaching))
top10university$research<-as.numeric(as.character(top10university$research))
top10university$citations<-as.numeric(as.character(top10university$citations))


Student_Staff<- top10university%>%
ggplot(aes(x=reorder(university_name,student_staff_ratio), y=student_staff_ratio),fill=year) +
  geom_bar(stat="identity", fill="lightblue", colour="black", position=position_dodge(0.2))+ theme_bw() + coord_f
lip() +
  labs(x="University", y="Student staff ratio",
       title="Student staff ratio (number of students per one staff)", subtitle="Times-Top 10 universities")+
theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 7))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one.
```

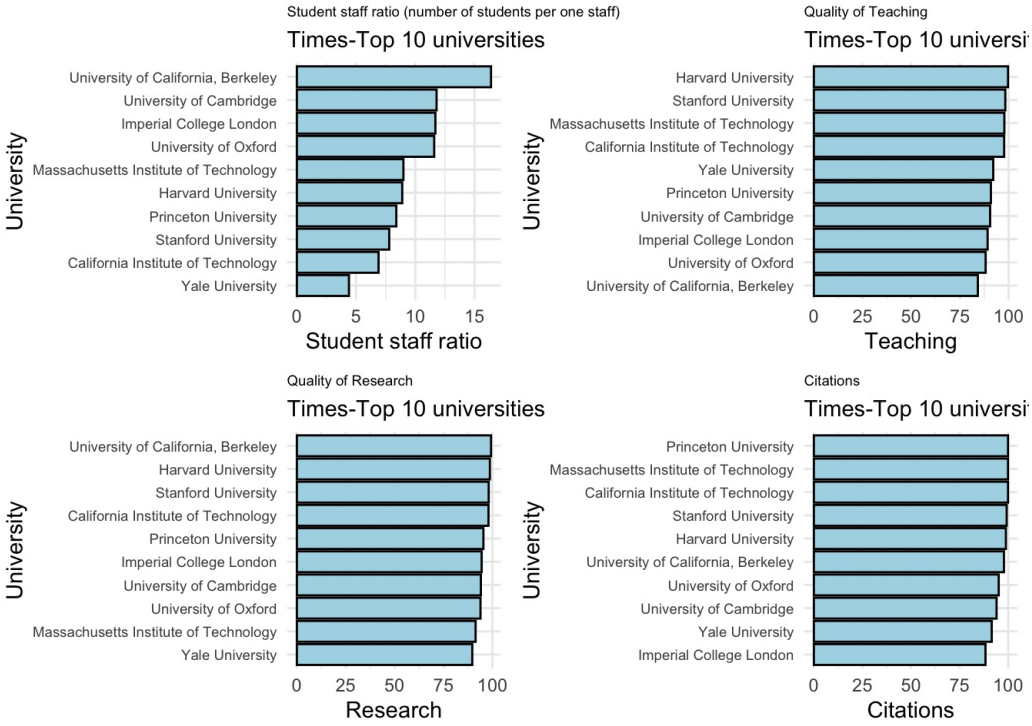
```
Teaching<- top10university%>%
ggplot(aes(x=reorder(university_name,teaching), y=teaching),fill=year) +
  geom_bar(stat="identity", fill="lightblue", colour="black", position=position_dodge(0.2))+ theme_bw() + coord_f
lip() +
  labs(x="University", y="Teaching",
       title="Quality of Teaching", subtitle="Times-Top 10 universities")+
theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 7))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one.
```

```
Research<- top10university%>%
ggplot(aes(x=reorder(university_name,research), y=research),fill=year) +
  geom_bar(stat="identity", fill="lightblue", colour="black", position=position_dodge(0.2))+ theme_bw() + coord_f
lip() +
  labs(x="University", y="Research",
       title="Quality of Research", subtitle="Times-Top 10 universities")+
theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 7))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one.
```

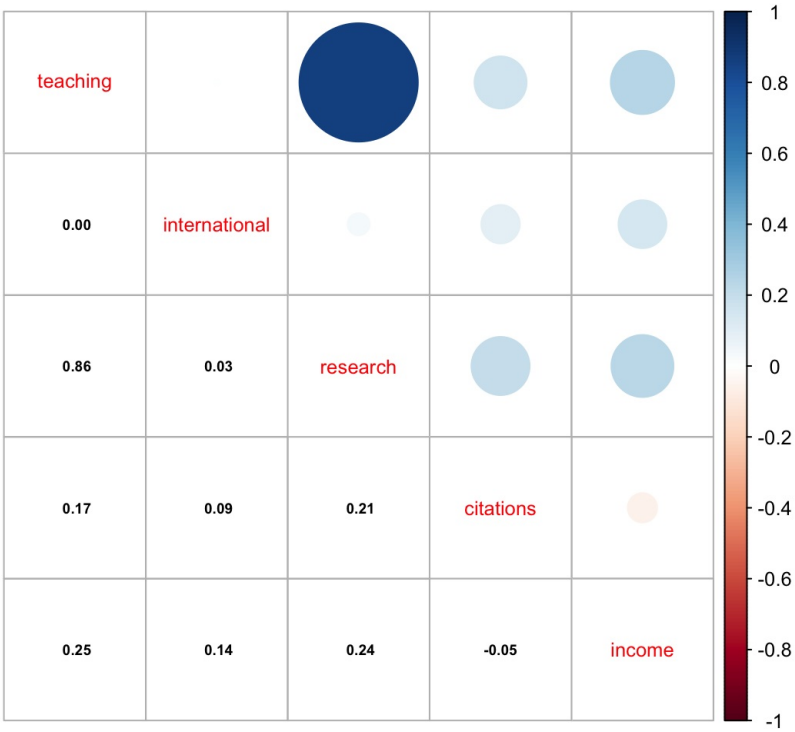
```
Citations<- top10university%>%
ggplot(aes(x=reorder(university_name,citations), y=citations),fill=year) +
  geom_bar(stat="identity", fill="lightblue", colour="black", position=position_dodge(0.1))+ theme_bw()+
  labs(x="University", y="Citations",
       title="Citations", subtitle="Times-Top 10 universities")+
theme(label.text=element_text(size= 0.2))+
  theme_minimal()+
  theme(plot.title = element_text(size = 7))+
  coord_flip()+theme(axis.text.y =element_text(size = 7))

grid.arrange(Student_Staff,Teaching,Research,Citations,ncol=2)
```



As shown above the graph represents the Student-Staff ratio,quaility of research, teaching and the citations of top 10 universities according to the world ranking. The highest student staff ratio and quality of research is performed in University of California,Berkeley and Yale University is least respectively. The quality of teaching is more in Harvard University compared to all other universities. Princetown university has the highest citations.

Q) How the learning environment score,international outlook score,volume-income-reputation score, research influence, industry income(knowledge transfer) score are correalted?

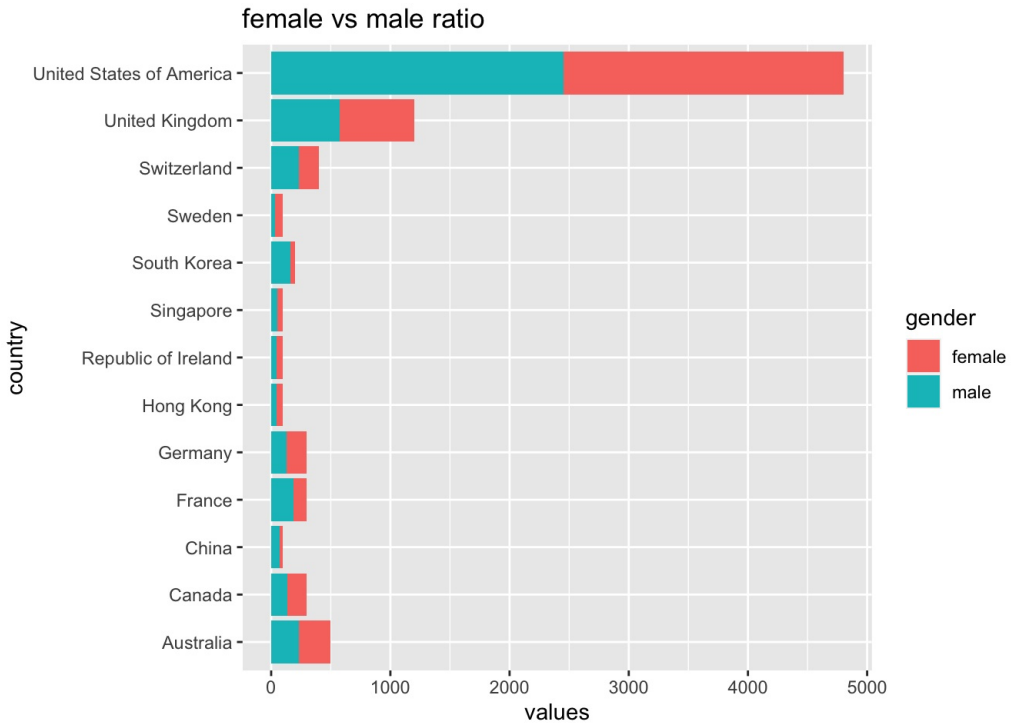
```
library(corrplot)

Top100_times<-head(timesData,100)

timesdata_corr<- Top100_times %>%
select(teaching,international,research,citations,income) %>%
drop_na() %>%
na.omit() %>%
na_if("-")
#changing the data type
timesdata_corr$international<-as.numeric(timesdata_corr$international)
timesdata_corr$income<-as.numeric(timesdata_corr$income)
timesdata_corr1<-data.frame(timesdata_corr)
timesdata_corr1[is.na(timesdata_corr1)]=0


cormat<- cor(timesdata_corr1)
corrplot_airlines<-corrplot.mixed(cormat, lower.col = "black",
                                  number.cex = .6, tl.cex=0.8)
```



As shown from the above correlation plot-From Times Ranking: Exploration capacity is firmly associated with educating quality. Worldwide viewpoint doesn't have solid connection with some other factors.

Q) Female and male ratio in Top 100 Countries

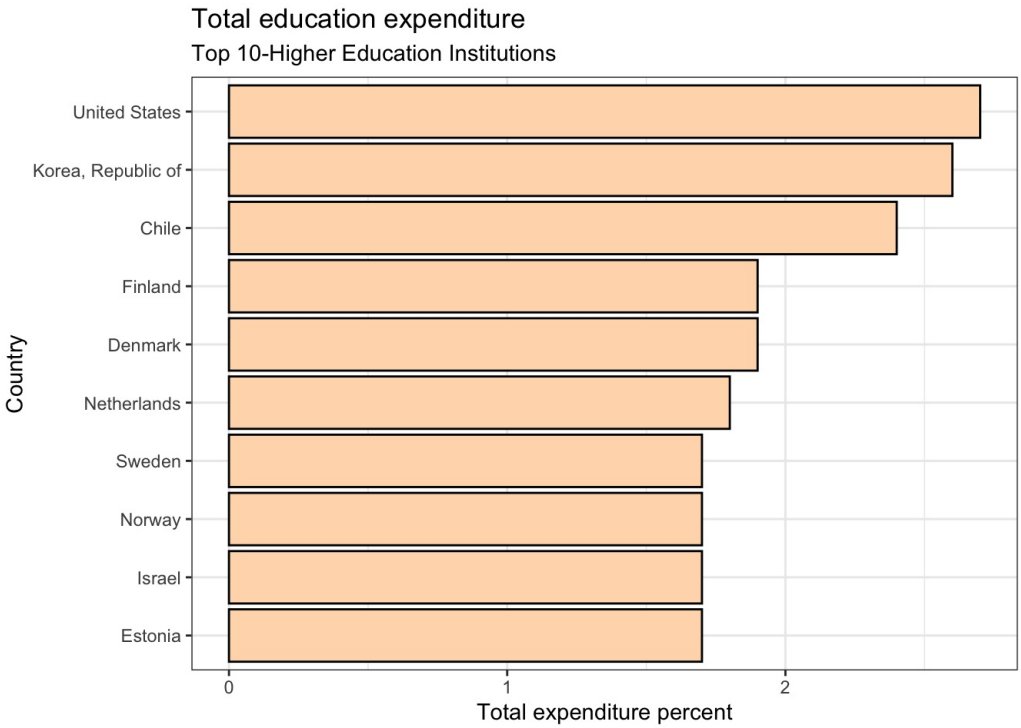
```
df1<-head(timesData,100)
df<-df1 %>%
  select(country,female_male_ratio,world_rank) %>%
  na_if("") %>%
  drop_na()
df_t<-separate(df,female_male_ratio, into=c("female","male"), sep=":",remove = FALSE,convert = TRUE,fill = "left"
)
times_col<-df_t %>%
  select(world_rank,country,female,male)
new_f<-times_col %>%
  select(world_rank,female)
new_f$gender<-"female"
new_m<-times_col %>%
  select(world_rank,male)
new_m$gender<-"male"
times1<-times_col %>%
  pivot_longer(c(female : male),
               names_to = 'gender',
               values_to = 'values',
               values_drop_na = TRUE)

female_male_ratio <- ggplot(data=times1, aes(x=values, y=country, fill=gender)) + geom_bar(stat="identity") +ggti
tle("female vs male ratio")
female_male_ratio
```



According to the female-male proportion, here we see something trademark to previous socialist nations yet in addition an indication of advancement: female-male ratio is higher in USA, United kingdom stands in the second place,Sweden has the smallest percent of male ratio. From the above graph we can see that female-male ratio is almost the same in Singapore, Republic of Ireland and Hongkong.

Q) Which countries spend most on Higher education in Top 10 Universities in the world?

```
edu_longer<-educationExpenditure %>%
  pivot_longer(c(X1995 : X2011),
               names_to = 'Year_X',
               values_to = 'values',
               values_drop_na = TRUE)

edu_longer$Year<-gsub("X","",as.character(edu_longer$Year_X))

edu_year<-edu_longer %>% select(country,institute_type,direct_expenditure_type,Year,values) %>%
  filter(Year==2011,institute_type=="Higher Education Institutions ",direct_expenditure_type=="Total") %>% top_n(
10,values) %>%
ggplot(aes(x=reorder(country,values), y=values),fill=year) + guides(fill=FALSE) +
  geom_bar(stat="identity", fill="peachpuff", colour="black", position=position_dodge(0.1))+ theme_bw() + coord_f
lip() +
  labs(x="Country", y="Total expenditure percent",
    title="Total education expenditure", subtitle="Top 10-Higher Education Institutions")

edu_year
```



As per the above graph, USA has spent the highest expenditure on education, korea is the second highest expenditure spent and Sweden,Norway, Israel & Estonia all of these four contries almost spent equal expenditure for education.

Conclusion:

Through exploratory data analysis on the three rankings and monetary and schooling pointers of every university/country, We had the option to reach the accompanying inference: Harvard University is the top university in Center for World University Rankings, trailed by MIT in 2012 and Stanford in 2013-2015; the third spot is changing between Stanford (2012), Oxford (2013), with MIT holding it for 2014 and 2015.

USA had the most incredible number of colleges on the CWUR Top 10 Ranking, which is 52. United Kingdom had 8 colleges on the CWUR Top 100 Ranking. Denmark had the most un-number of colleges.

Harvard University is as yet driving with a level of 100 focuses as complete score. USA is a nation where establishments are different in understudy body, have high student staff proportion, Students can be taken acceptable consideration of. The research capacity, training quality is very acceptable. It is likewise an excellent country. College of California,Berkeley has extraordinary examination capacity and Harvard college has the incredible instruction quality.

Female-Male proportion is higher in USA, United kingdom remains in the second place,Sweden has the littlest percent of male proportion.We can see that female-male proportion is practically something similar in Singapore, Republic of Ireland and Hongkong.

Suggestions and Insights:

Times Ranking have a greater number of nations and more variety in establishments than Shanghai Ranking. Look more into schools in American, European and Asian there are a ton of schools that have extraordinary advancement capacity, and solid scholarly execution.

America is an extraordinary decision if you might want to concentrate abroad. It got most appraisals on student/staff proportion, and extraordinary variety in student body. Netherlands, Finland, Sweden, Norway, Israel and Estonia central area likewise have great use on training.

To get well-rounded schooling, allude to Shanghai Ranking with just one of the pointers on the positioning, since they are associated with each other. On the off chance that you might want to have some variety in picking schools and nations, Times Ranking is useful.

In case there are more factors that identified with other financial and political pointers of a country, then, at that point, we can construct a model on why there are that number of establishment of a specific nation is on the best 100 positioning.