Project: University Entrance Exam Analysis

Dol_R

8/24/2020

Contents

1 Introduction and Key Takeaways

In Turkey, every year millions of students take the university entrance exam. After the announcement of results, participants list their university and department choices and they are placed according to their ranks.

This analysis focuses on universities and departments popularity over the years. Hopefully, it would help future participants in their decision making process.

Key Takeaways:

- Medicine maintained its popularity over the years and added mean of 577 more quotas every year.
- Koç University has a significantly higher popularity with its scholarship options.
- Computer Engineering has gained popularity over the years.
- Civil Engineering has lost its popularity over the years.

2 Summary of the Data and Explanations

Using **University Exam** data from *Hacettepe University's Website*, we obtained university results of years 2016-2020. Each year is on a separate Excel spreadsheet. Since every year some departments are opened and some are shut down, datasets will have different number of rows. There are 9 variables and more than 10000 rows for each dataset.

- **university**: Name of the university
- **city**: University's location
- **department**: Name of the department
- **type**: Type of the exam points.
- **quota**: Maximum number of participants to be accepted, upper bound for accepted_number
- **accepted_number**: Accepted number of participants to the selected university's selected department
- **lowest_score**: Lowest score of the accepted participant's scores
- **highest_score**: Highest score of the accepted participant's scores
- **lowest_ranking**: Lowest ranking of the accepted participant's scores, last accepted person's ranking

Objectives:

- Analyze University Exam Entrance data
- Cleaning and manipulation of datasets
- Exploration of the popularity trends of the universities and departments
- Comparison by visualization
- Finding top universities and departments

2.1 Data Preprocessing

In order to ease the reading process, functions mani97 and mani86 are created. These functions remove the columns related to the percentage of change regarding the previous year. They also rename the columns and remove the last empty rows. The raw data for some years have an additional column at the end, therefore two functions were needed.

```r
knitr::opts_chunk$set(echo = TRUE, message=FALSE, warning=FALSE) # applies to all chunks
mani97 <- function(data){
  newdata <- data %>%
  rename(
    number = c(1),
    difference = c(2),
    university = c(3),
    city = c(4),
    department = c(5),
    type = c(6),
    quota = c(7),
    difference_quota = c(8),
    accepted_number = c(9),
    lowest_score = c(10),
    highest_score = c(11),
    lowest_ranking = c(12),
    difference_ranking = c(13)
  )  %>% select(-number, -difference, -difference_quota, -difference_ranking) %>%
    slice_head(n=nrow(data)-9)
  return(newdata)
}
mani86 <- function(data){
  newdata <- data %>%
  rename(
    number = c(1),
    difference = c(2),
    university = c(3),
    city = c(4),
    department = c(5),
    type = c(6),
    quota = c(7),
    difference_quota = c(8),
    accepted_number = c(9),
    lowest_score = c(10),
    highest_score = c(11),
    lowest_ranking = c(12)
  )  %>% select(-number, -difference, -difference_quota) %>%
    slice_head(n=nrow(data)-9)
  return(newdata)
}
```

2.2 Loading the Libraries and Datasets

After loading the required libraries, we download the .xlsx file from our project repository. After reading it, we remove the file.

```r
library(tidyverse)
library(lubridate)
library(tinytex)
library(readxl) # read_excel
library(tidyr)
library(httr) # GET
library(reshape2)
library(ggforce) # circle data points
library(xaringan)
library(kableExtra)
library(webshot)
library(knitr)
```

```r
url<-’https://github.com/pjournal/boun01g-dol-r/blob/gh-pages/uni_exam_project/uni_exam.xlsx?raw=true’
GET(url, write_disk(tf <- tempfile(fileext = ".xlsx")))
```

```r
raw_df20 <- read_excel(tf, sheet="20", skip=21)
raw_df19 <- read_excel(tf, sheet="19", skip=21)
raw_df18 <- read_excel(tf, sheet="18", skip=21)
raw_df17 <- read_excel(tf, sheet="17", skip=21)
raw_df16 <- read_excel(tf, sheet="16", skip=21)
file.remove(tf)
```

We have 5 datasets for the last 5 years. Each dataset has 9 columns and they all have different number of rows. For example there are 10617 rows in dataset for the year 2020 and 11402 rows in dataset for the year 2019. Some departments or even universities may have been opened or closed.

```r
data2020 <- mani97(raw_df20)
data2020 <- mani97(raw_df20)
data2019 <- mani97(raw_df19)
data2018 <- mani86(raw_df18)
data2017 <- mani97(raw_df17)
data2016 <- mani97(raw_df16)
data2020 %>% summarise(exam20=n()) %>% mutate(data2019 %>% summarise(exam19=n())) %>%
  mutate(data2018 %>% summarise(exam18=n())) %>%
  mutate(data2017 %>% summarise(exam17=n())) %>%
  mutate(data2016 %>% summarise(exam16=n()))
```

exam20	exam19	exam18	exam17	exam16
10617	11402	11958	11484	10657

As an example, let's observe dataset for the year 2018. First four variables are categorical and the rest is numerical. We could also see that there are 11958 rows for this year.

```
data2018 %>% arrange(desc(lowest_score)) %>% glimpse()
```

```
## Rows: 11,958
## Columns: 9
## $ university      <chr> "İSTANBUL MEDİPOL ÜNİVERSİTESİ", "KOÇ ÜNİVERSİTESİ"...
## $ city            <chr> "İSTANBUL", "İSTANBUL", "İSTANBUL", "İSTANBUL", "İS...
## $ department      <chr> "Tıp Fakültesi (İngilizce) (Burslu)", "Tıp Fakültes...
## $ type            <chr> "SAY", "SAY", "SAY", "SAY", "SAY", "SAY", "SAY", "S...
## $ quota           <dbl> 10, 15, 8, 11, 52, 82, 11, 30, 82, 50, 9, 175, 50, ...
## $ accepted_number <dbl> 10, 15, 8, 11, 52, 82, 11, 30, 82, 50, 9, 175, 50, ...
## $ lowest_score    <dbl> 549.1749, 548.1738, 543.2350, 542.8008, 538.5325, 5...
## $ highest_score   <dbl> 562.9543, 556.2859, 546.1979, 548.5940, 559.5140, 5...
## $ lowest_ranking  <dbl> 56, 84, 207, 221, 400, 496, 510, 511, 622, 637, 739...
```

3 Exploratory Data Analysis

When the exam procedure changes, the types of grading systems also change. Therefore, there is no point of comparing `type` variable for different years. But it can be an identifier in each year.

3.1 Top Universities and Departments

Some universities' departments may be selected by the participants with lowest scores in different types of grading systems. After grouping by `university` and `department`, lowest score column is put in decreasing order. Lowest score is the last entering person's score to a specific university and department. By putting it into decreasing order, we get the highest scores of last entering people. Let's see how the top 10 departments changed over the years.

To do this, we count unique values and use `full_join` on all of the datasets. Finally, we replace `NULL` values with zero, which means this university or department was not in top 10 for that specific year.

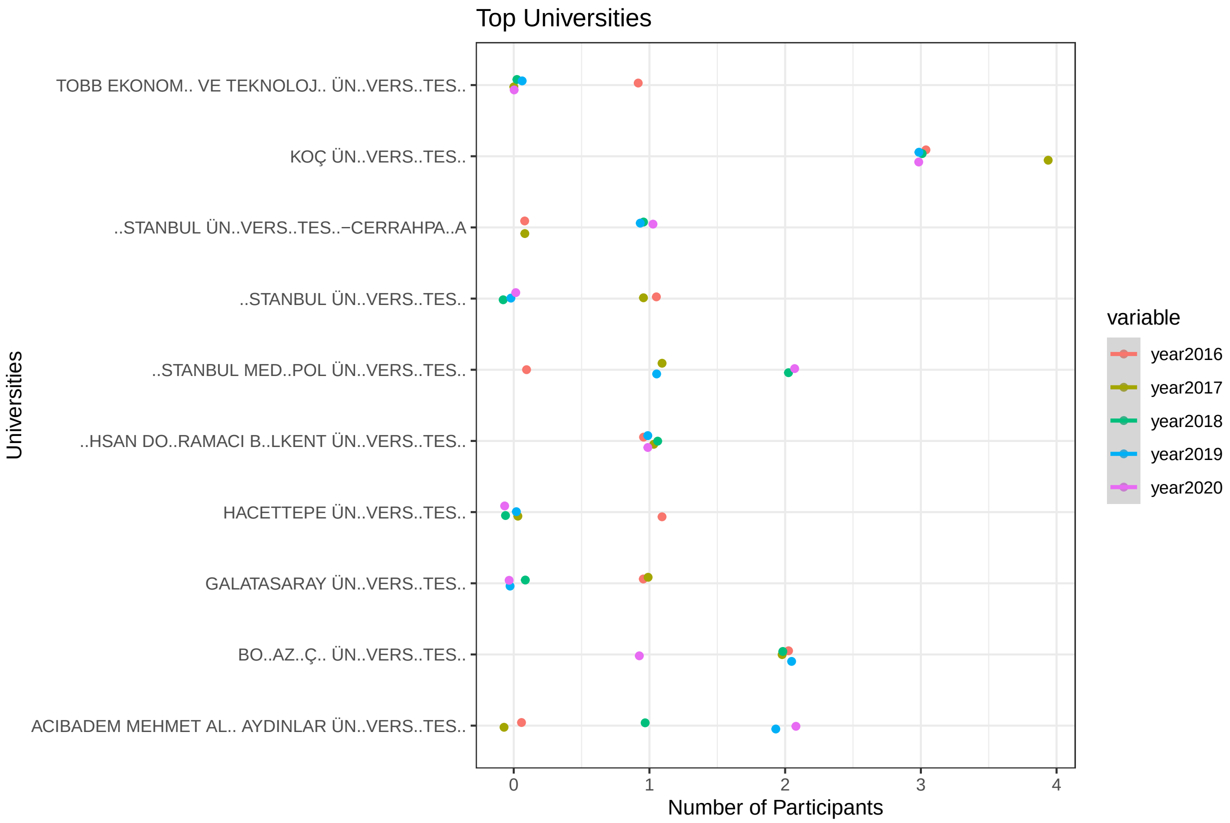
```
score <- function(data){
  data %>% group_by(university, department) %>% select(-type, -city) %>%
    arrange(desc(lowest_score)) %>% head(10) %>%
    mutate(department=replace(department, str_detect(department,"Tıp"), "Tıp")) %>%
    mutate(department=str_replace(department, "Burslu", "")) %>%
    mutate(department=str_replace(department, "İngilizce", "")) %>%
    mutate(department=str_replace(department, " \\([^()]{0,}\\)", "")) %>%
    mutate(department=str_replace(department, " \\([^()]{0,}\\)", "")) %>% ungroup()
}
dept <- function(data){
  data %>% count(department) %>% arrange(desc(n))
}
uni <- function(data){
  data %>% count(university) %>% arrange(desc(n))
}
```

This plot shows how many times a university occurred in our top 10 calculation in each year. For example, Bilkent University occurred one time each year and Koç University has the highest number of occurrences.

```r
top_uni <- full_join(uni(score(data2016)), uni(score(data2017)), by="university",
                     suffix=c(".2016", ".2017")) %>%
  full_join(., uni(score(data2018)), by="university", suffix=c(".2016", ".2018"))
top_uni <- full_join(top_uni, uni(score(data2019)), by="university",
                     suffix=c(".2018", ".2019"))
top_uni <- full_join(top_uni, uni(score(data2020)), by="university",
                     suffix=c(".2019", ".2020"))
names(top_uni)[names(top_uni)==’n.2016’]<-’year2016’
names(top_uni)[names(top_uni)==’n.2017’]<-’year2017’
names(top_uni)[names(top_uni)==’n.2018’]<-’year2018’
names(top_uni)[names(top_uni)==’n.2019’]<-’year2019’
names(top_uni)[names(top_uni)==’n’]<-’year2020’

top_uni <- top_uni %>%
  replace_na(list(year2016 = 0, year2017 = 0, year2018 = 0, year2019 = 0, year2020 = 0))

top_uni.long <- melt(top_uni)
plot_uni <- ggplot(top_uni.long, aes(x=value, y=university, col=variable)) +
  geom_jitter(width=0.1, height=0.1)  + stat_smooth() + ggtitle("Top Universities") +
  xlab("Number of Participants") + ylab("Universities") +
  theme(plot.caption=element_text(hjust = 0.5)) + theme_bw()
plot_uni
```

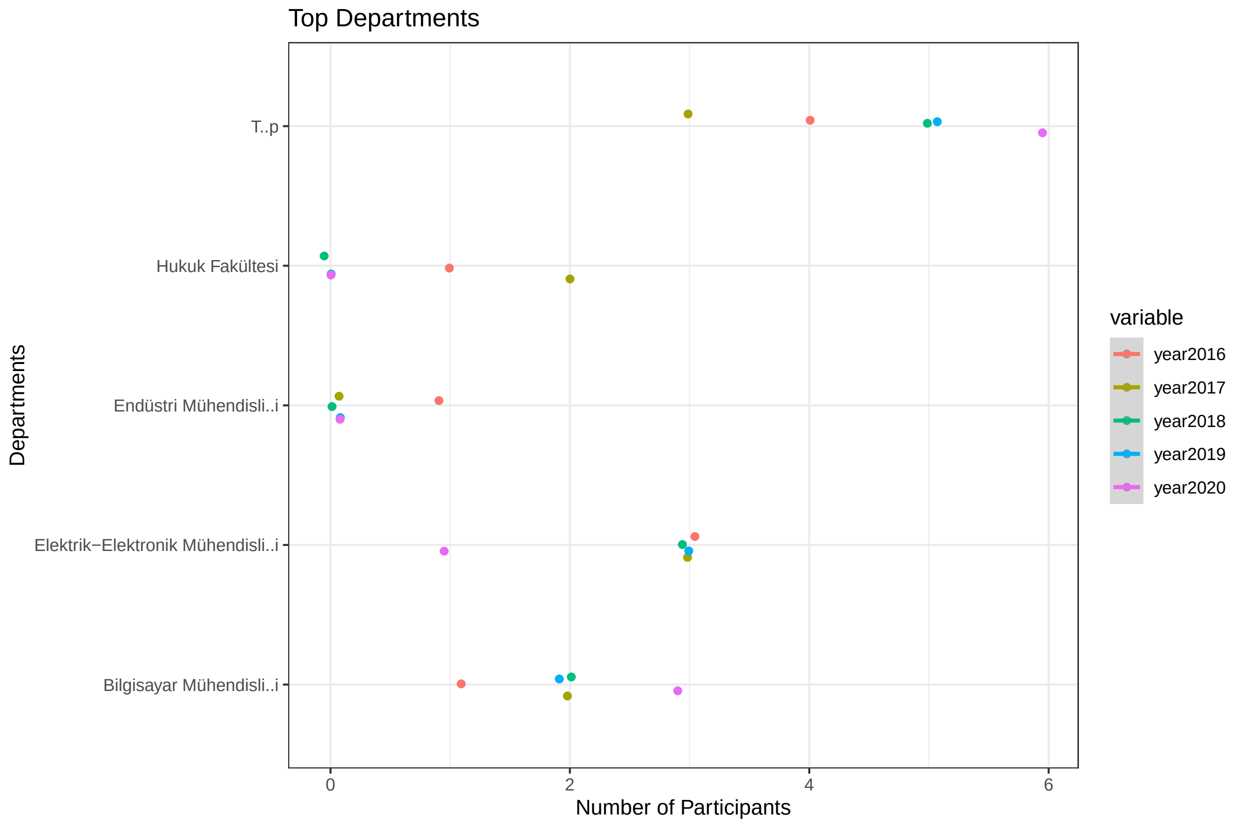


This plot shows how many times a department occurred in our top 10 calculation in each year.For departments we remove "(Burslu)" and "(İngilizce)" information in order to count them correctly.

```
top_dept <- full_join(dept(score(data2016)), dept(score(data2017)), by="department",
                      suffix=c(".2016", ".2017")) %>%
  full_join(., dept(score(data2018)),by="department", suffix=c(".2016", ".2018"))
top_dept <- full_join(top_dept, dept(score(data2019)), by="department",
                      suffix=c(".2018", ".2019"))
top_dept <- full_join(top_dept, dept(score(data2020)), by="department",
                      suffix=c(".2019", ".2020"))
names(top_dept)[names(top_dept)==’n.2016’]<-’year2016’
names(top_dept)[names(top_dept)==’n.2017’]<-’year2017’
names(top_dept)[names(top_dept)==’n.2018’]<-’year2018’
names(top_dept)[names(top_dept)==’n.2019’]<-’year2019’
names(top_dept)[names(top_dept)==’n’]<-’year2020’

top_dept <- top_dept %>%
  replace_na(list(year2016 = 0, year2017 = 0, year2018 = 0, year2019 = 0, year2020 = 0))

top_dept.long <- melt(top_dept)
plot_dept <- ggplot(top_dept.long, aes(x=value, y=department, col=variable)) +
  geom_jitter(width=0.1, height=0.1) + stat_smooth() + ggtitle("Top Departments") +
  xlab("Number of Participants") + ylab("Departments") +
  theme(plot.caption=element_text(hjust = 0.5)) + theme_bw()
plot_dept
```



3.2 Top Cities

Cities play a key role while selecting university.

```
total_data <- full_join(data2016,data2017,by=c(’university’,’department’,’city’,’type’),
                        suffix = c(’.2016’,’.2017’))%>%
  full_join(. ,data2018,by=c(’university’,’department’,’city’,’type’))%>%
  full_join(. ,data2019,by=c(’university’,’department’,’city’,’type’),
            suffix = c(’.2018’,’.2019’))%>%
  full_join(. ,data2020,by=c(’university’,’department’,’city’,’type’))
names(total_data)[names(total_data)==’type’]<-’type.2020’
names(total_data)[names(total_data)==’quota’]<-’quota.2020’
names(total_data)[names(total_data)==’accepted_number’]<-’accepted_number.2020’
names(total_data)[names(total_data)==’lowest_score’]<-’lowest_score.2020’
names(total_data)[names(total_data)==’highest_score’]<-’highest_score.2020’
names(total_data)[names(total_data)==’lowest_ranking’]<-’lowest_ranking.2020’
```

3.2.1 University and City Distribution of the First Thousand Students in the University Exam

In the first table below, we calculated how many listings of cities appeared in the first thousand ranking.

```
bin_tr<- total_data%>%
  filter(lowest_ranking.2020<1000)
tab2<-table(bin_tr$city,bin_tr$type.2020)
kable(tab2,caption = "Table.1. City  Distribution of the First Thousand Students")%>%
  kable_styling(latex_options = "hold_position")
```

Table 1: Table.1. City Distribution of the First Thousand Students

	DİL	EA	SAY	SÖZ
ANKARA	0	4	2	1
İSTANBUL	1	15	8	5

Similarly, the second table shows the distribution of universities listed in the first thousand.

```r
tab<-table(bin_tr$university,bin_tr$type.2020)
kable(tab,caption = "Table.2. University Distribution of the First Thousand Students ")%>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 2: Table.2. University Distribution of the First Thousand Students

	DİL	EA	SAY	SÖZ
ACIBADEM MEHMET ALİ AYDINLAR ÜNİVERSİTESİ	0	0	2	0
BOĞAZİÇİ ÜNİVERSİTESİ	0	3	2	1
GALATASARAY ÜNİVERSİTESİ	0	1	0	0
İHSAN DOĞRAMACI BİLKENT ÜNİVERSİTESİ	0	3	2	0
İSTANBUL 29 MAYIS ÜNİVERSİTESİ	0	0	0	1
İSTANBUL MEDİPOL ÜNİVERSİTESİ	0	0	1	0
İSTANBUL ÜNİVERSİTESİ	0	1	0	0
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA	0	0	1	0
KOÇ ÜNİVERSİTESİ	1	6	2	2
ÖZYEĞİN ÜNİVERSİTESİ	0	1	0	1
SABANCI ÜNİVERSİTESİ	0	2	0	0
TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ	0	1	0	1
YEDİTEPE ÜNİVERSİTESİ	0	1	0	0

3.2.2 University and City Distribution of the First Five Thousand Students in the University Exam

Table 3 shows how many listings of cities appeared in the first five thousand ranking.

```r
besbin_tr<- total_data%>%
  filter(lowest_ranking.2020<5000)
tab3<-table(besbin_tr$city,besbin_tr$type.2020)
kable(tab3,caption = "Table.3. City Distribution of the First Five Thousand Students") %>%
  kable_styling(latex_options = "hold_position")
```

Table 3: Table.3. City Distribution of the First Five Thousand Students

	DİL	EA	SAY	SÖZ
ANKARA	9	19	21	10
ESKİŞEHİR	0	0	1	0
İSTANBUL	14	42	35	47
İZMİR	2	0	1	2

Table 4 below shows the distribution of universities listed in the first thousand.

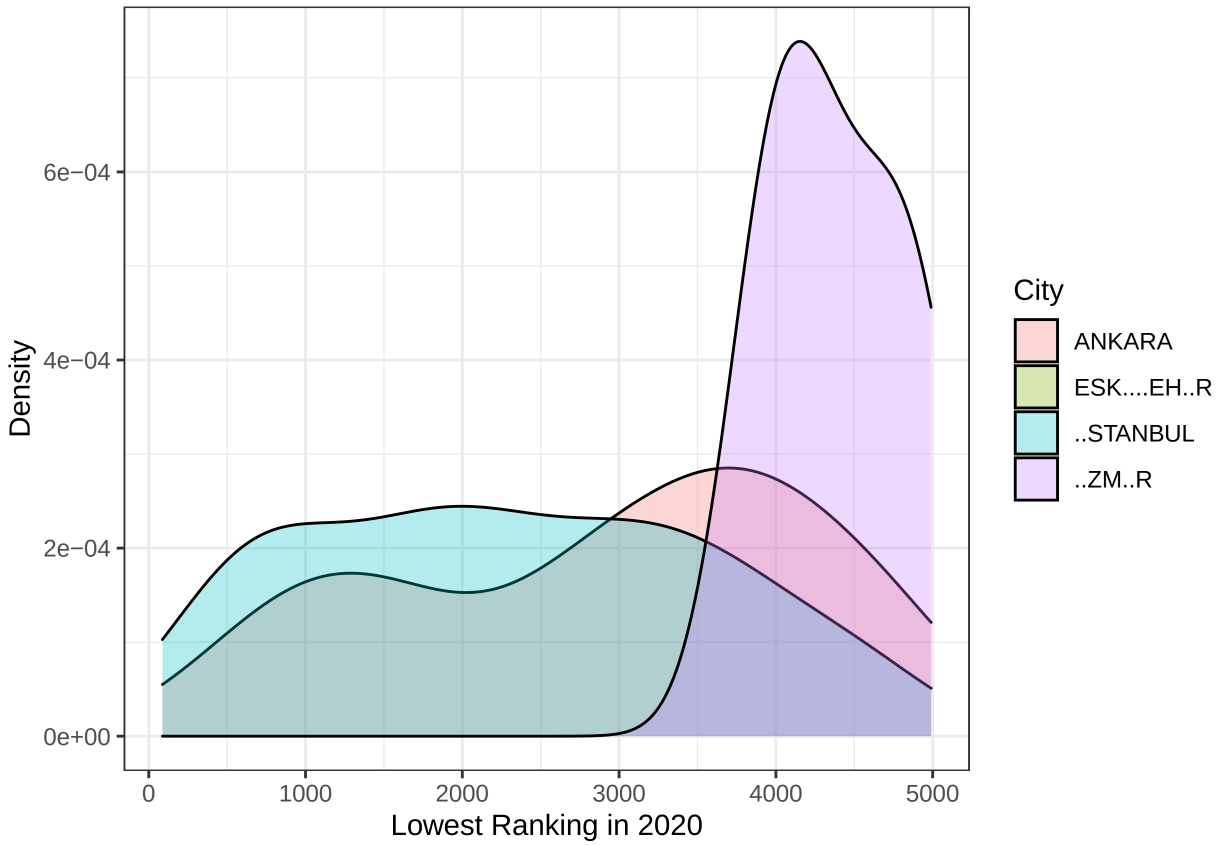
```r
tab4<-table(besbin_tr$university,besbin_tr$type.2020)
kable(tab4,
      caption = "Table.4. University Distribution of the First Five Thousand Students")%>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 4: Table.4. University Distribution of the First Five Thousand Students

	DİL	EA	SAY	SÖZ
ACIBADEM MEHMET ALİ AYDINLAR ÜNİVERSİTESİ	0	0	2	0
ANKARA ÜNİVERSİTESİ	0	1	2	0
BAHÇEŞEHİR ÜNİVERSİTESİ	0	2	1	7
BAŞKENT ÜNİVERSİTESİ	1	0	2	1
BEYKENT ÜNİVERSİTESİ	0	0	0	1
BEZM-İ ÂLEM VAKIF ÜNİVERSİTESİ	0	0	1	0
BOĞAZİÇİ ÜNİVERSİTESİ	4	7	4	3
EGE ÜNİVERSİTESİ	0	0	1	0
ESKİŞEHİR TEKNİK ÜNİVERSİTESİ	0	0	1	0
GALATASARAY ÜNİVERSİTESİ	1	5	0	1
GAZİ ÜNİVERSİTESİ	0	0	2	0
HACETTEPE ÜNİVERSİTESİ	2	1	2	2
İBN HALDUN ÜNİVERSİTESİ	0	0	0	3
İHSAN DOĞRAMACI BİLKENT ÜNİVERSİTESİ	3	8	7	1
İSTANBUL 29 MAYIS ÜNİVERSİTESİ	1	0	0	4
İSTANBUL BİLGİ ÜNİVERSİTESİ	2	2	0	6
İSTANBUL MEDİPOL ÜNİVERSİTESİ	0	0	1	0
İSTANBUL OKAN ÜNİVERSİTESİ	1	0	0	0
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ	0	0	0	2
İSTANBUL TEKNİK ÜNİVERSİTESİ	0	1	3	0
İSTANBUL ÜNİVERSİTESİ	1	2	2	0
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA	0	0	2	0
İSTİNYE ÜNİVERSİTESİ	0	0	2	0
İZMİR EKONOMİ ÜNİVERSİTESİ	1	0	0	2
KADİR HAS ÜNİVERSİTESİ	0	0	0	3
KOÇ ÜNİVERSİTESİ	2	10	11	3
MALTEPE ÜNİVERSİTESİ	0	0	0	1
MARMARA ÜNİVERSİTESİ	0	0	1	1
ORTA DOĞU TEKNİK ÜNİVERSİTESİ	2	3	2	2
ÖZYEĞİN ÜNİVERSİTESİ	0	9	3	2
SABANCI ÜNİVERSİTESİ	0	2	1	0
TED ÜNİVERSİTESİ	0	0	0	1
TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ	1	6	4	3
YAŞAR ÜNİVERSİTESİ	1	0	0	0
YEDİTEPE ÜNİVERSİTESİ	2	2	1	10

We can monitor the density of cities for first five thousand ranking. Obviously, Istanbul has dominated the first five thousand while Izmir starts to attract students from three thousand. Ankara is somewhere in between and mostly left-skewed.

```
besbin_tr%>%
  ggplot(aes(x = lowest_ranking.2020, fill = city)) +
  geom_density(alpha = 0.3)+theme_bw() + labs(x='Lowest Ranking in 2020', y= 'Density',
                                              fill='City' )
```



3.3 Quota Informations

In 2020, the occupancy rate in quotas is 92.63 %. Which is greater than 2019 value (90.64 %). You can find the departments with the highest quotas in recent years in Table 5.

```
all_years_data<-bind_rows(mutate(data2016,year=2016),
mutate(data2017,year=2017),
mutate(data2018,year=2018),
mutate(data2019,year=2019),
mutate(data2020,year=2020))
  dep<-all_years_data%>%
  group_by(year,department)%>%
  summarise(department_quota=sum(quota))%>%
   arrange(desc(department_quota))%>%head(20)
  kable(dep,caption = "Table.5. Department Quotas in Recent Years")%>%
    kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 5: Table.5. Department Quotas in Recent Years

year	department	department_quota
2020	Hemşirelik	12469
2019	Hemşirelik	11994
2017	İşletme (Açıköğretim)	11788
2018	Hemşirelik	11629
2017	Hemşirelik	11306
2016	Hemşirelik	10921
2020	Tıp Fakültesi	10702
2019	Tıp Fakültesi	10203
2018	Tıp Fakültesi	9648
2017	İşletme	9628
2016	İşletme	9481
2016	İşletme (Açıköğretim)	9226
2017	Tıp Fakültesi	9029
2017	İktisat	8840
2019	İlahiyat	8698
2018	İşletme	8593
2016	İktisat	8591
2020	İlahiyat	8573
2018	İlahiyat	8558
2017	İlahiyat	8501

You can find the cities with the highest quotas in recent years in Table 6.

```
city_tab<- all_years_data%>%
  group_by(year,city)%>%
  summarise(city_quota=sum(quota))%>%
  arrange(desc(city_quota))%>%head(20)
  kable(city_tab,caption = "Table.6. City Quotas in Recent Years") %>%
    kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 6: Table.6. City Quotas in Recent Years

year	city	city_quota
2018	İSTANBUL	112514
2020	İSTANBUL	106565
2019	İSTANBUL	102296
2017	İSTANBUL	101551
2016	İSTANBUL	94271
2017	ESKİŞEHİR	65815
2016	ESKİŞEHİR	52348
2020	ANKARA	45092
2019	ANKARA	40273
2018	ANKARA	40069
2018	ESKİŞEHİR	38118
2017	ANKARA	37907
2016	ANKARA	35065
2019	ESKİŞEHİR	29768
2020	ESKİŞEHİR	26229
2020	İZMİR	23462
2019	İZMİR	22564
2018	İZMİR	22163
2020	ERZURUM	21999
2017	İZMİR	19236

You can find the universities with the highest quotas in recent years in Table 7.

```
uni<- all_years_data%>%
 group_by(year,university)%>%
 summarise(university_quota=sum(quota))%>%
 arrange(desc(university_quota))%>%head(20)
 kable(uni,caption = "Table.7. University Quotas in Recent Years")%>%
   kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 7: Table.7. University Quotas in Recent Years

year	university	university_quota
2017	ANADOLU ÜNİVERSİTESİ	61485
2016	ANADOLU ÜNİVERSİTESİ	48110
2018	ANADOLU ÜNİVERSİTESİ	32351
2019	ANADOLU ÜNİVERSİTESİ	24057
2018	İSTANBUL ÜNİVERSİTESİ	22687
2020	ATATÜRK ÜNİVERSİTESİ	21203
2020	ANADOLU ÜNİVERSİTESİ	20939
2017	İSTANBUL ÜNİVERSİTESİ	18309
2016	İSTANBUL ÜNİVERSİTESİ	18169
2018	ATATÜRK ÜNİVERSİTESİ	15735
2020	İSTANBUL ÜNİVERSİTESİ	15283
2019	İSTANBUL ÜNİVERSİTESİ	15115
2016	ATATÜRK ÜNİVERSİTESİ	14239
2019	ATATÜRK ÜNİVERSİTESİ	13663
2017	ATATÜRK ÜNİVERSİTESİ	13524
2017	SELÇUK ÜNİVERSİTESİ	9738
2016	SELÇUK ÜNİVERSİTESİ	8364
2020	ANKARA ÜNİVERSİTESİ	8196
2018	SELÇUK ÜNİVERSİTESİ	8162
2017	GAZİ ÜNİVERSİTESİ	8055

3.4 Quota Differences

In some cases we would wonder the departments whose quota more increased or decreased. However, there are some highly volatile departments ending with "(Açıköğretim) , (İÖ) or (..indirimli). they may be misleading, so we should eliminate the departments ending with these words in order to get more established departments. Also another criteria can be necessity of being founded at least 4 years ago. So that we can see trends of department quotas. The top 20 departments which have the highest quota increase can be found in Table 8. Note: Ranked by taking mean quota differences of last 4 years.

```r
quote_by_dept <- all_years_data%>%
  group_by(year,department)%>%
  summarise(department_quota=sum(quota))
quote_by_dept <- data.frame(quote_by_dept) #after group_by we should convert it to df.
quote_by_dept <- quote_by_dept %>%
    group_by(department) %>%
    mutate(Diff = department_quota - lag(department_quota)) %>% arrange(desc(Diff))
quote_by_dept <- data.frame(quote_by_dept) # Again convert to df
quote_by_dept <- quote_by_dept %>%
  subset(!substr(department,nchar(department)-1,nchar(department)) %in%
           c("i)","m)","Ö)")) # eliminating some departments
# Also eliminating new founded departments
quote_by_dept <- quote_by_dept %>% group_by(department) %>% filter(n()>= 4)
quote_by_dept <- data.frame(quote_by_dept) # Again convert to df
increase<-quote_by_dept %>% group_by(department) %>%
  summarise(mean_difference=mean(Diff,na.rm = T)) %>%
  arrange(desc(mean_difference)) %>% head(20)#top 20 dept increased quota
kable(increase,
      caption = "Table.8. Deparments which Their Quotas Increased Most in Recent Years")%>%
  kable_styling(font_size = 8, latex_options = "hold_position")
```

Table 8: Table.8. Deparments which Their Quotas Increased Most in Recent Years

department	mean_difference
Tıp Fakültesi	577.50
Psikoloji	464.25
Diş Hekimliği Fakültesi	461.75
Sağlık Yönetimi	460.75
Finans ve Bankacılık	454.50
Gastronomi ve Mutfak Sanatları	424.50
İslami İlimler	391.25
Bilgisayar Mühendisliği	388.00
Hemşirelik	387.00
Mimarlık	370.25
Beslenme ve Diyetetik	343.50
Fizyoterapi ve Rehabilitasyon	342.00
Radyo, Televizyon ve Sinema	316.75
İngiliz Dili ve Edebiyatı (İngilizce)	294.00
Rehberlik ve Psikolojik Danışmanlık	293.00
Sosyal Hizmet	282.50
Siyaset Bilimi ve Kamu Yönetimi	265.75
Matematik	257.25
Türk Dili ve Edebiyatı	235.50
Hukuk Fakültesi	223.50

The top 20 departments which have the highest quota decrease can be found in Table 9.

```
decrease<-quote_by_dept %>% group_by(department) %>%
  summarise(mean_difference=mean(Diff,na.rm = T)) %>%
  arrange(mean_difference) %>% head(20)#20 dept with most decrease in quota
kable(decrease,digits = 2,
      caption = "Table.9. Deparments which Their Quotas Decreased Most in Recent Years")%>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

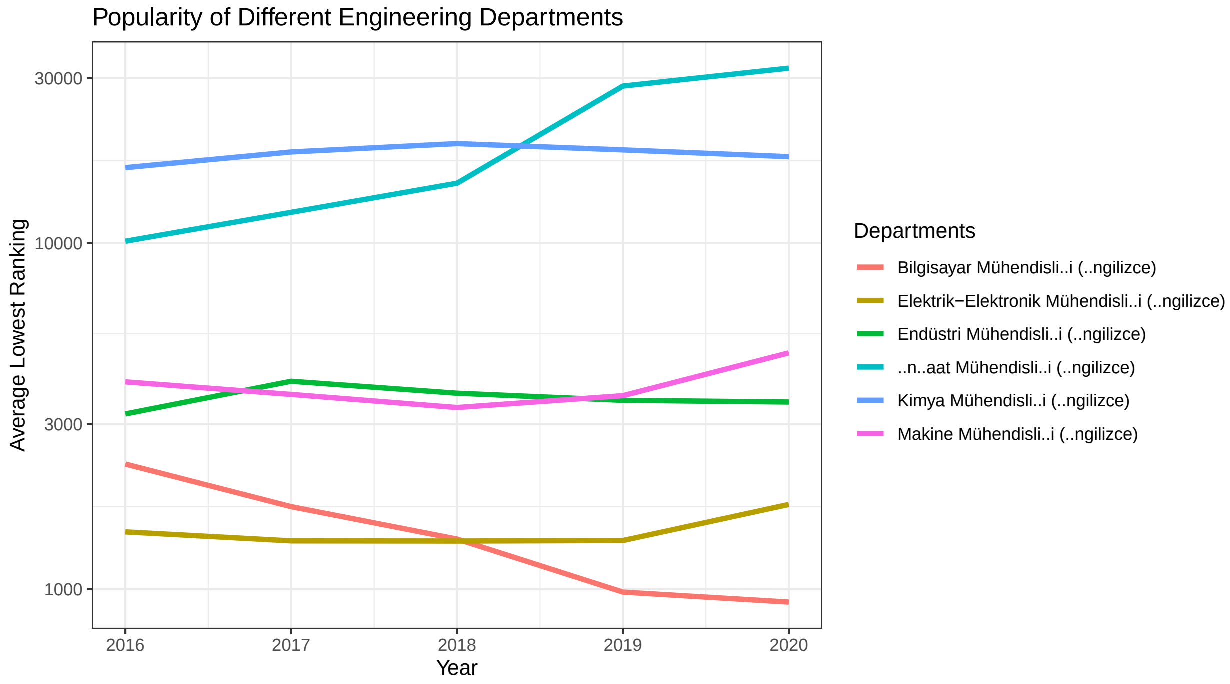
Table 9: Table.9. Deparments which Their Quotas Decreased Most in Recent Years

department	mean_difference
İşletme	-789.50
İktisat	-736.50
Bilgisayar ve Öğretim Teknolojileri Öğretmenliği	-577.00
Fen Bilgisi Öğretmenliği	-457.00
Gıda Mühendisliği	-450.00
Çevre Mühendisliği	-416.75
Sınıf Öğretmenliği	-237.25
Arkeoloji	-219.00
Kamu Yönetimi	-216.00
Enerji Sistemleri Mühendisliği	-191.50
Uluslararası Ticaret	-185.25
Metalurji ve Malzeme Mühendisliği	-171.50
Konaklama İşletmeciliği	-158.67
Makine Mühendisliği	-105.00
Ekonometri	-100.00
Turizm İşletmeciliği ve Otelcilik	-97.67
Jeoloji Mühendisliği	-94.75
Sosyal Bilgiler Öğretmenliği	-92.50
İmalat Mühendisliği	-90.50
Kimya Mühendisliği	-88.00

3.5 Popularity of Different Engineering Departments

In this part of the project, the two most popular state universities' data are taken into account. 6 different engineering departments that exist in both Boğaziçi University and Middle East Technical University are chosen to make a comparison their popularity over the years. Data shows that computer engineering gains popularity and civil engineering lose its popularity in the recent years. Other engineering branches such as industrial engineering and electrical&electronics engineering have relatively stable popularity.

```
meanRanking<-all_years_data%>%
  filter(department%in%c("İnşaat Mühendisliği (İngilizce)",
                         "Bilgisayar Mühendisliği (İngilizce)",
                         "Endüstri Mühendisliği (İngilizce)",
                         "Elektrik-Elektronik Mühendisliği (İngilizce)",
                         "Makine Mühendisliği (İngilizce)",
                         "Kimya Mühendisliği (İngilizce)"))%>%
  filter(university%in%c("BOĞAZİÇİ ÜNİVERSİTESİ","ORTA DOĞU TEKNİK ÜNİVERSİTESİ"))%>%
  group_by(year,department)%>%
  summarise(mean_ranking=mean(lowest_ranking))
ggplot(meanRanking,aes(x=year,y=mean_ranking,color=department)) + geom_line(size=1.3) +
  labs(x="Year",y="Average Lowest Ranking",color="Departments") +
  scale_y_log10() + theme_bw() +
  ggtitle("Popularity of Different Engineering Departments")
```

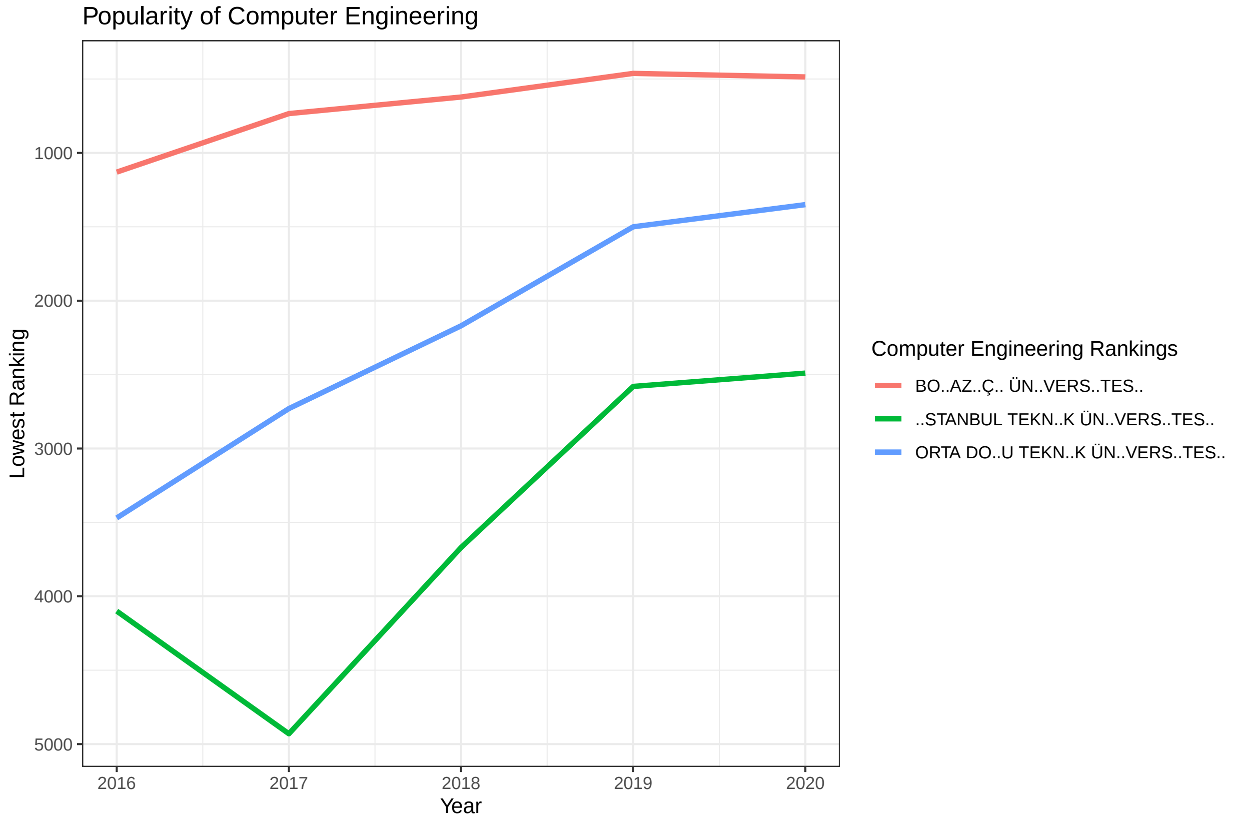


3.6 Popularity of Computer Engineering and Civil Engineering

Popularity of computer engineering and civil engineering in the top 3 state universities can be shown in plots below. There is an increasing trend of choosing computer engineering but negative trend of choosing civil engineering in the recent years.

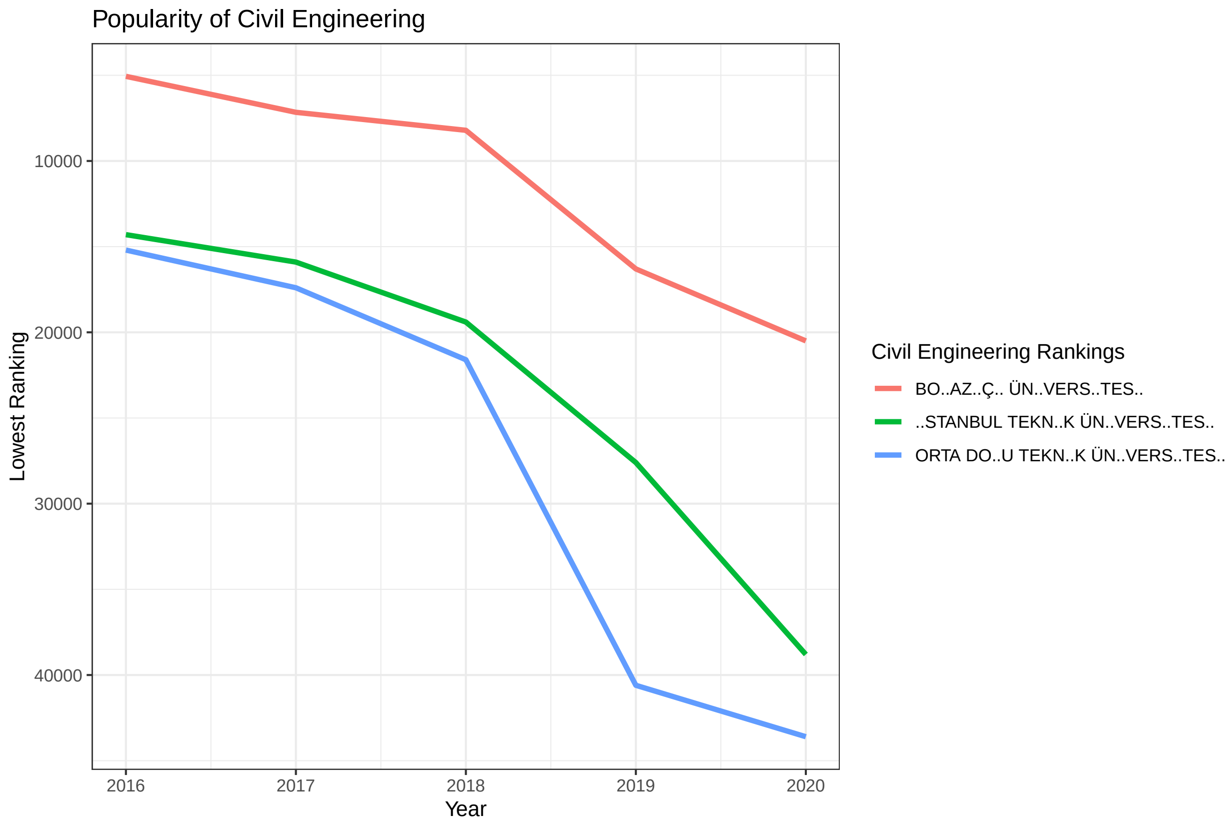
Below you can see the increasing trend of Computer Engineering.

```
cmpe<-all_years_data%>%
  select(year,university,department,lowest_ranking)%>%
  filter(university%in%c("ORTA DOĞU TEKNİK ÜNİVERSİTESİ",
                         "BOĞAZİÇİ ÜNİVERSİTESİ","İSTANBUL TEKNİK ÜNİVERSİTESİ"),
         department=="Bilgisayar Mühendisliği (İngilizce)")
ggplot(cmpe,aes(x=year,y=lowest_ranking,color=university)) + geom_line(size=1.3) +
  scale_y_reverse() +
  labs(x="Year",y="Lowest Ranking",color="Computer Engineering Rankings") +
  theme_bw() + ggtitle("Popularity of Computer Engineering")
```



Below you can see the increasing trend of Civil Engineering.

```
ce<-all_years_data%>%
  select(year,university,department,lowest_ranking)%>%
  filter(university%in%c("ORTA DOĞU TEKNİK ÜNİVERSİTESİ","BOĞAZİÇİ ÜNİVERSİTESİ",
                         "İSTANBUL TEKNİK ÜNİVERSİTESİ"),
         department=="İnşaat Mühendisliği (İngilizce)")
ggplot(ce,aes(x=year,y=lowest_ranking,color=university)) + geom_line(size=1.3) +
  scale_y_reverse() +
  labs(x="Year",y="Lowest Ranking",color="Civil Engineering Rankings") +
  theme_bw() + ggtitle("Popularity of Civil Engineering")
```

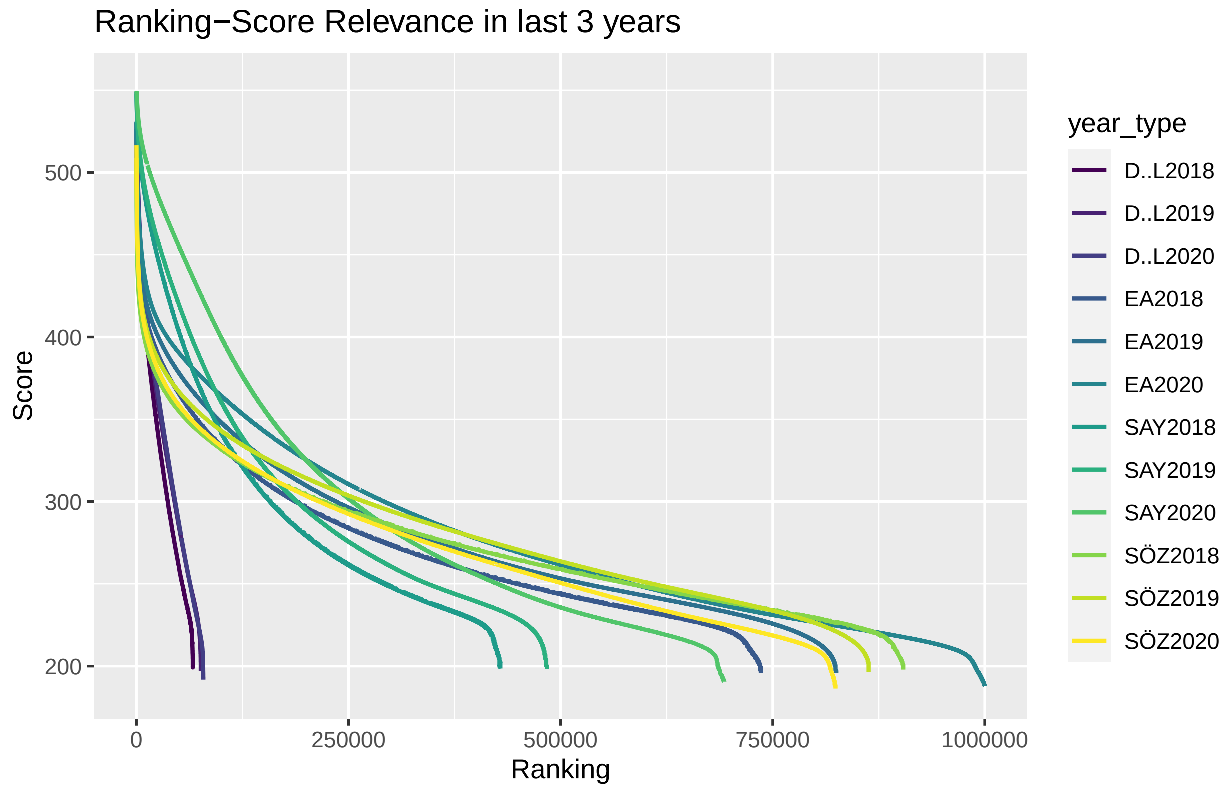


3.7 Equivalence of rankings to score in years 2018, 2019 and 2020

Among university exam applicants one of the biggest curiosity is the equivalence of exam score to rankings. Calculating possible exam scores is possible with the equations published by ÖSYM or via websites but ranking equivalence change every year by the toughness of the exam. Also another problem of the equivalence is ÖSYM does not publish the year by year equivalence of scores to rankings. The only indicator we have is the lowest ranking and lowest score equivalence table. So we aim to find the correlation between the scores and the ranking with the data we have and forecast possible minimum score to enter different ranking intervals. We only use 2018, 2019, 2020 data since before the ranking system was different.

```
YKS_years_data<-bind_rows(mutate(data2018,year=2018),
mutate(data2019,year=2019),
mutate(data2020,year=2020)) %>% mutate (year_type = paste0(type,as.character(year)))
SAY_quota_2020 <- YKS_years_data %>% filter(year_type== "SAY2020") %>%select(quota) %>%sum()
SAY_quota_2018 <- YKS_years_data %>% filter(year_type== "SAY2018") %>%select(quota) %>%sum()
SOZ_quota_2020 <- YKS_years_data %>% filter(year_type== "SÖZ2020") %>%select(quota) %>%sum()
SOZ_quota_2018 <- YKS_years_data %>% filter(year_type== "SÖZ2018") %>%select(quota) %>%sum()
```

```
ggplot(YKS_years_data,aes(x = (lowest_ranking) )) +
  geom_line(aes(color = year_type, y = lowest_score), cex = 0.8) +
  scale_color_viridis_d() + ggtitle("Ranking-Score Relevance in last 3 years") +
  xlab("Ranking") + ylab("Score")
```



From this graph we can observe that in top 100000 students, effect of the score on the ranking decreases. Also we can guess the effect of people's choises on rankings. In 2018 students with most ranking of 400,000 enter SAY departments but in 2020 it rises up to 700,000. The SAY quotas were 192792 in 2018 and 191155 in 2020. There is a decrease in quotas but it can be seen that people with higher rankings prefer SAY departments in 2020.

Also in from 2018 to 2020 people prefer SÖZ departments more and the lowest rankings decrease. The SÖZ quotas were 105445 in 2018 and 113362 in 2020. The quota increase by 8% but the rankings got lower.

DIL departments show really small fluctuations. There is slight change in those three years.

```r
dil_data <- YKS_years_data %>% filter(type == "DİL")
ea_data <- YKS_years_data %>% filter(type == "EA")
say_data <- YKS_years_data %>% filter(type == "SAY")
soz_data <- YKS_years_data %>% filter(type == "SÖZ")
temp <- dil_data %>% filter(lowest_ranking < 11000 & lowest_ranking > 9000 ) %>%
  select(lowest_score)
dil_10k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- dil_data %>% filter(lowest_ranking <52500  & lowest_ranking > 47500 ) %>%
  select(lowest_score)
dil_50k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- dil_data %>% filter(lowest_ranking < 105000 & lowest_ranking > 95000 ) %>%
  select(lowest_score)
dil_100k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- ea_data %>% filter(lowest_ranking < 11000 & lowest_ranking > 9000 ) %>%
  select(lowest_score)
ea_10k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- ea_data %>% filter(lowest_ranking < 52500 & lowest_ranking > 47500 ) %>%
  select(lowest_score)
ea_50k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- ea_data %>% filter(lowest_ranking < 105000 & lowest_ranking > 95000 ) %>%
  select(lowest_score)
ea_100k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- say_data %>% filter(lowest_ranking < 11000 & lowest_ranking > 9000 ) %>%
  select(lowest_score)
say_10k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- say_data %>% filter(lowest_ranking < 52500 & lowest_ranking > 47500 ) %>%
  select(lowest_score)
say_50k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- say_data %>% filter(lowest_ranking < 105000 & lowest_ranking > 95000 ) %>%
  select(lowest_score)
say_100k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- soz_data %>% filter(lowest_ranking < 11000 & lowest_ranking > 9000 ) %>%
  select(lowest_score)
soz_10k <- temp$lowest_score %>% mean()%>% signif(6)
temp <- soz_data %>% filter(lowest_ranking < 52500 & lowest_ranking > 47500 ) %>%
  select(lowest_score)
soz_50k <- temp$lowest_score %>% mean() %>% signif(6)
temp <- soz_data %>% filter(lowest_ranking < 105000 & lowest_ranking > 95000 ) %>%
  select(lowest_score)
soz_100k <- temp$lowest_score %>% mean() %>% signif(6)
dil_100k <- "-"
temp_vec <- c(dil_10k,dil_50k,dil_100k,ea_10k,ea_50k,ea_100k,say_10k,say_50k,say_100k,
              soz_10k,soz_50k,soz_100k)
temp_vec <-  matrix(temp_vec,nrow = 3,ncol = 4)
temp_df <-  as.tibble(temp_vec)
names(temp_df)=c("DİL","EA","SAY","SÖZ")
Table <- temp_df %>% mutate("Ranking" = c("10,000","50,000","100,000")) %>%
  select("Ranking","DİL","EA","SAY","SÖZ")
```

Even though there are significant fluctuations in the higher ranking area, the top 100,000 ranking shows stable behavior. In the university exam students mostly aim three ranking depends on their work and potential. These rankings are 100,000, 50,000 and 10,000.

We wanted to create a small guideline for future students to see their possible interval with their potential scores. We take the average of scores in three years with selecting an interval around the selected rankings. This interval is 9k to 11k in 10,000 ranking, 47.5k and 52.5k in 50,000 ranking and 95k and 105k in 100,000 ranking.

```
kable(Table)%>%
  kable_styling(font_size = 15, latex_options = "hold_position")
```

Ranking	DİL	EA	SAY	SÖZ
10,000	418.772	427.827	494.556	404.704
50,000	278.34	375.578	424.469	359.692
100,000	-	347.141	365.983	335.743

4 Conclusion

In Turkey, every year millions of students take the university entrance exam and enter university. Every year rankings of departments and universities change due to popularity and this affect the results significantly. We made analyses on the university entrance exam data and observe the outcomes. The data ww used was not in a clean form therefore we started with cleaning and manipulating the data.

We found the top universities and departments along with the most preferred cities. We knew the importance of the quota information so we observe the changes in the department quotas. Using both the ranking and quota information leads us to understand the cities and departments with the highest demand among entrants. We observe correlation between the quota and the ranking changes in several departments.

We observe the change in popularity of engineering departments and wanted go into detail in the popularity of engineering departments. We found out significant changes in the rankings of engineering departments such as Civil and Computer Engineering.

Furthermore we defined the correlation of scores with rankings and create a ranking-score table to be used as a guide by prospective students.

5 Shiny App

We also have a Shiny App called University Exams in Turkey.

6 References

- *Hacettepe University's Website.*
- *Xaringan Tutorial*
- *Stack Overflow*