Iran Dental Schools Ranking

Ahmad Sofi-Mahmudi

Aim

In this project, I aimed to rank dental schools in Iran using principal component analysis of five variables:

- Average mark of students;
- Number of all professors;
- Number of full and associate professors;
- Median mark in residency exam; and,
- The proportion of participants in the residency exam who were allowed to choose their specialty.

Analysis

First, we load the data:

```
dental_data = read.csv("iran_dental_schools_data.csv")
```

This data set has data for these 34 schools:

```
dental_data$School
```

```
 [1] "Ahwaz"          "Alborz"         "Ardabil"        "AzadKhorasgan"
 [5] "AzadShiraz"     "AzadTehran"     "Babol"          "Bandarabbas"
 [9] "Birjand"        "Bushehr"        "Gilan"          "Golestan"
[13] "Hamadan"        "Isfahan"        "Kerman"         "Kermanshah"
[17] "Kurdistan"      "Lorestan"       "Mashhad"        "Mazandaran"
[21] "Urmia"          "Qazvin"         "Qom"            "Rafsanjan"
```

```
[25] "Semnan"          "Shahed"          "ShahidBeheshti" "Shiraz"
[29] "Tabriz"          "Tehran"          "Yasuj"          "Yazd"
[33] "Zahedan"         "Zanjan"
```

The variables are:

```
names(dental_data)
```

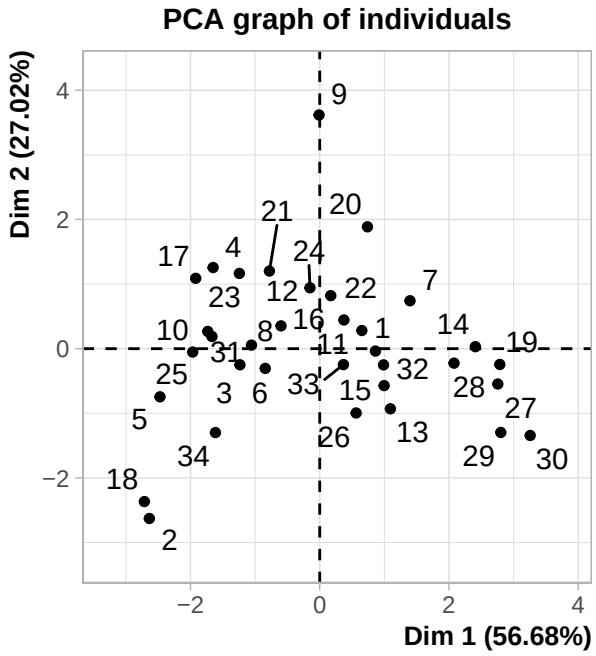
```
[1] "School"
[2] "students_average"
[3] "n_professors"
[4] "full_associate_professors_proportion"
[5] "residency_mark_median"
[6] "allowed_to_choose_residency_proportion"
```

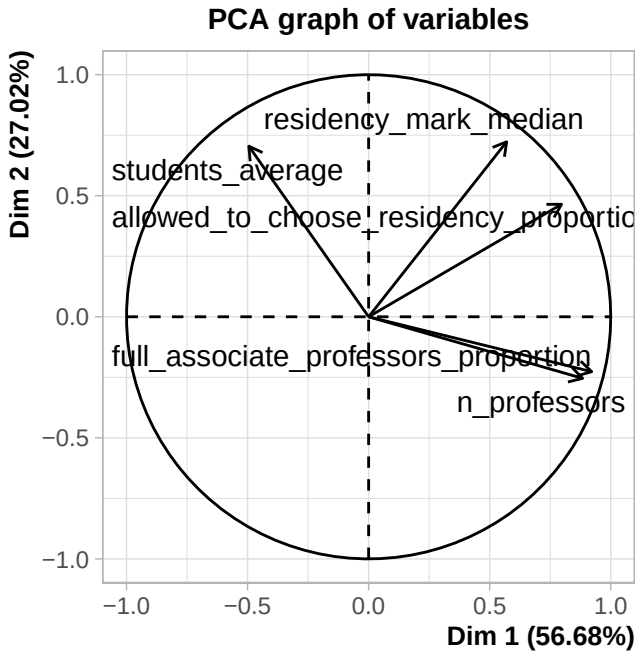
Now, we run the PCA analysis:

```
library(FactoMineR)
library(dplyr)
```

```
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
```

```
pca_out = PCA(dental_data[,-1])
```






```
score0 = pca_out$ind$coord[,1]
pca_score = 100*(score0 - min(score0))/(max(score0) - min(score0))
score = as.data.frame(pca_score)
pca_data = bind_cols(dental_data, score)
```

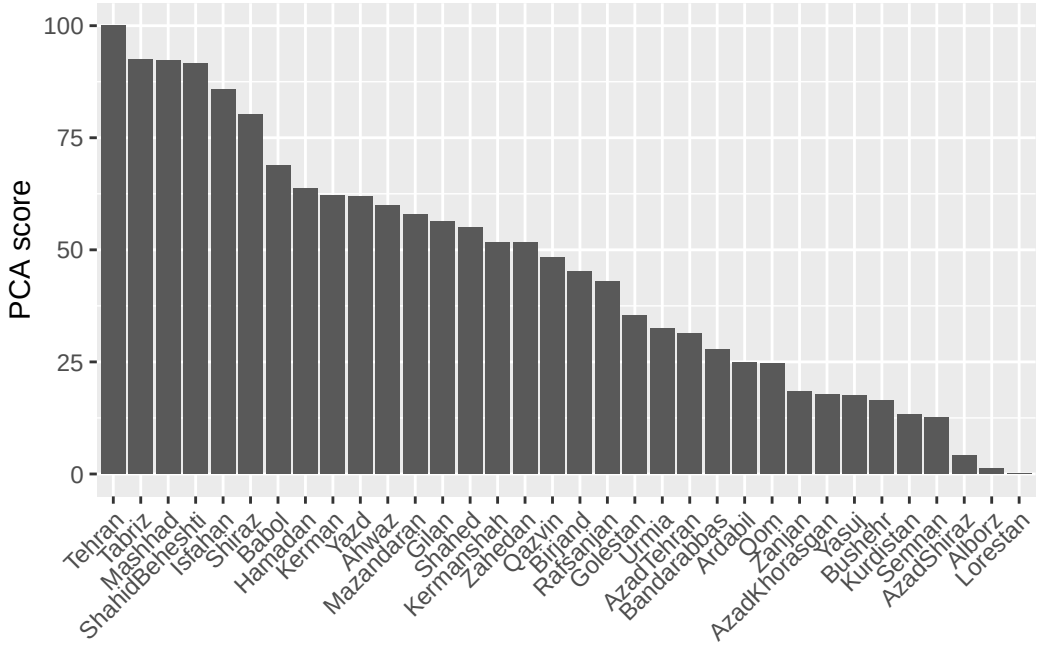
Here are the scores:

```
sorted_pca = pca_data[, c(1, 7)][order(pca_data$pca_score, decreasing = T),]
sorted_pca
```

```
           School  pca_score
30         Tehran 100.000000
29         Tabriz  92.388246
19        Mashhad  92.126739
27 ShahidBeheshti  91.607487
14        Isfahan  85.793467
28         Shiraz  80.267929
7           Babol  68.844185
13        Hamadan  63.765460
15         Kerman  62.132963
32           Yazd  61.980979
1           Ahwaz  59.893235
20     Mazandaran  57.821395
11          Gilan  56.365350
26         Shahed  54.882374
16     Kermanshah  51.717437
33        Zahedan  51.589248
22         Qazvin  48.309183
9         Birjand  45.270591
24      Rafsanjan  42.923197
12       Golestan  35.417481
21          Urmia  32.407703
6      AzadTehran  31.330858
8     Bandarabbas  27.755195
3         Ardabil  24.764206
23            Qom  24.627724
34         Zanjan  18.418912
4   AzadKhorasgan  17.813388
31          Yasuj  17.514947
10        Bushehr  16.436427
17      Kurdistan  13.303971
25         Semnan  12.526615
5      AzadShiraz   4.044774
2          Alborz   1.284014
18       Lorestan   0.000000
```

We can also plot it:

```r
library(ggplot2)

sorted_pca %>%
        ggplot() +
        aes(x = reorder(School, pca_score), y = pca_score) +
        geom_bar(stat = "identity") +
        scale_fill_gradient2(high = "red", low = "white") +
        theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
        scale_x_discrete(limits = sorted_pca$School) +
        xlab(NULL) +
        ylab("PCA score")
```




```r
# ggsave("PCA_scores.png", width = 16, height = 8, units = "in", dpi = 300)
```

Now, we can save the data:

```r
write.csv(sorted_pca, "dental_schools_ranking_data.csv", row.names = F)
```