

# London

*Ana Real*

*June 4, 2019*

This work separates the universities in tops from the overall Rank variable:

- Top10 1-10.
- Top20 11-20.
- Top30 21-30.
- Top40 31-40.
- Top50 41-50.
- Top60 51-60.
- Top70 61-70.
- Top80 71-80.
- Top90 81-90.
- Top100 91-100.

Then some correlation plots were exported using this variables as groups.

## Read data

```
library(xlsx)
library(ggpubr)

## Loading required package: ggplot2
## Loading required package: magrittr
library(DescTools)
library(xtable)
library(caret)

## Loading required package: lattice
##
## Attaching package: 'caret'
## The following objects are masked from 'package:DescTools':
##
##      MAE, RMSE
library(DMwR)

## Loading required package: grid
library(ggcorrplot)
library(corrplot)

## corrplot 0.84 loaded
data <- read.xlsx2("../final.xlsx", sheetIndex = 1)

for(i in c(1,2,4:10)){
  data[,i] <- as.numeric(as.character(data[,i]))
}
```

```

}

colnames(data) <- c("Year", "Rank", "City", "Overall", "Rankings", "Student.Mix", "Desirability", "Emp.

# Adding variable for top universities

data$Top <- data$Rank

for (i in 1:376){
  if (data$Rank[i] <= 10){
    data$Top[i] <- "Top10"
  } else if (data$Rank[i] > 10 & data$Rank[i] <= 20){
    data$Top[i] <- "Top20"
  } else if (data$Rank[i] > 20 & data$Rank[i] <= 30){
    data$Top[i] <- "Top30"
  } else if (data$Rank[i] > 30 & data$Rank[i] <= 40){
    data$Top[i] <- "Top40"
  } else if (data$Rank[i] > 40 & data$Rank[i] <= 50){
    data$Top[i] <- "Top50"
  } else if (data$Rank[i] > 50 & data$Rank[i] <= 60){
    data$Top[i] <- "Top60"
  } else if (data$Rank[i] > 60 & data$Rank[i] <= 70){
    data$Top[i] <- "Top70"
  } else if (data$Rank[i] > 70 & data$Rank[i] <= 80){
    data$Top[i] <- "Top80"
  } else if (data$Rank[i] > 80 & data$Rank[i] <= 90){
    data$Top[i] <- "Top90"
  } else {
    data$Top[i] <- "Top100"
  }
}

data$Top <- as.factor(data$Top)
str(data)

```

```

## 'data.frame':   376 obs. of  11 variables:
## $ Year          : num  2018 2018 2018 2018 2018 ...
## $ Rank          : num   1  2  3  4  5  6  7  8  9 10 ...
## $ City          : Factor w/ 110 levels "Aberdeen","Adelaide",...: 50 99 56 62 76 65 12 110 96 89 ...
## $ Overall       : num  482 479 476 467 463 461 457 454 453 449 ...
## $ Rankings      : num  100 84 68 57 93 54 49 63 64 93 ...
## $ Student.Mix   : num   92 55 100 94 80 74 75 83 97 67 ...
## $ Desirability  : num   80 97 91 89 80 89 88 94 95 67 ...
## $ Emp.Activity  : num   93 100 86 80 88 78 80 90 84 92 ...
## $ Affordability : num   25 54 33 47 38 67 71 42 23 44 ...
## $ Student.View  : num   92 89 98 100 84 99 94 82 90 86 ...
## $ Top           : Factor w/ 10 levels "Top10","Top100",...: 1 1 1 1 1 1 1 1 1 1 ...

```

Group Top10, top 10 cities.

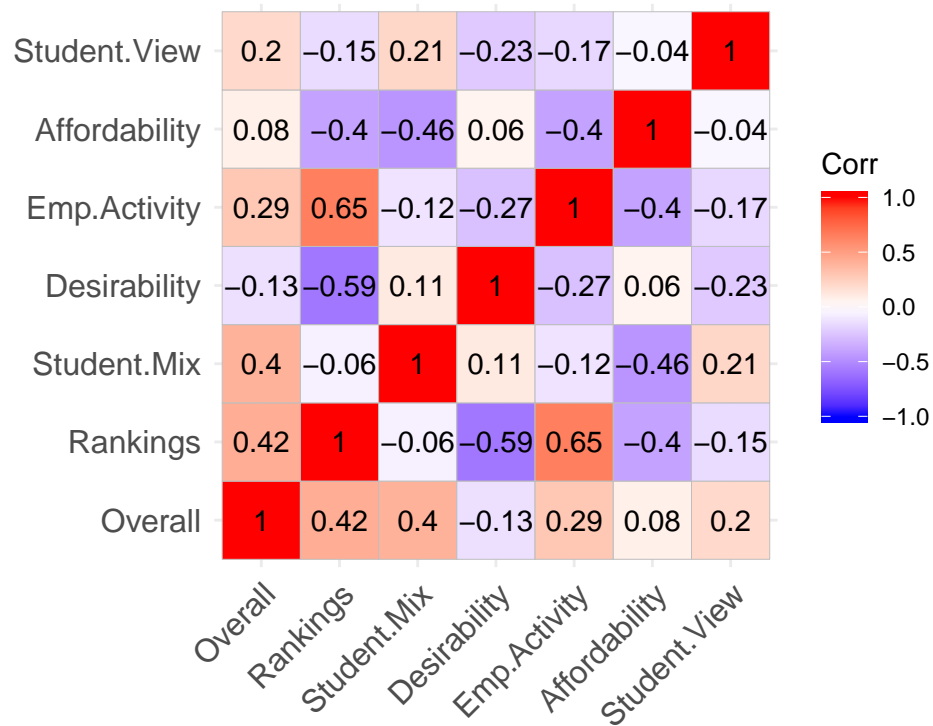
```

# Using spearman

corr1 <- cor(data[data$Top=="Top10",c(4:10)], method="spearman")

```

```
ggcorrplot(corr1, lab = TRUE)
```



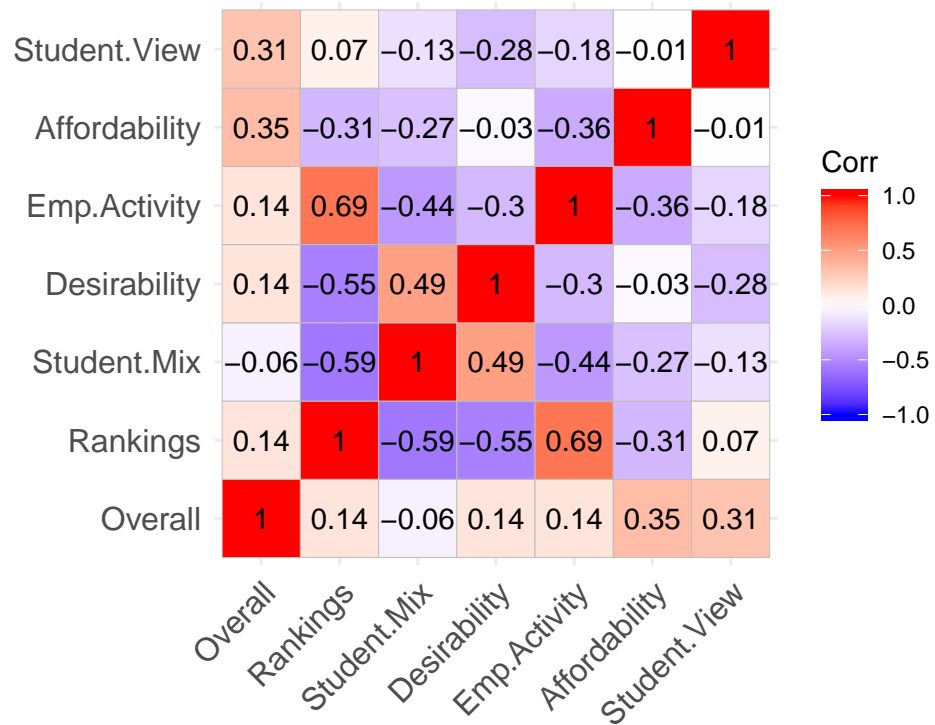
Some cities in this group: \* London \* Paris \* Tokyo

**Group Top20, cities from 11-20.**

```
# Using spearman
```

```
corr2 <- cor(data[data$Top=="Top20",c(4:10)], method="spearman")
```

```
ggcorrplot(corr2, lab = TRUE)
```



Some cities in this group:

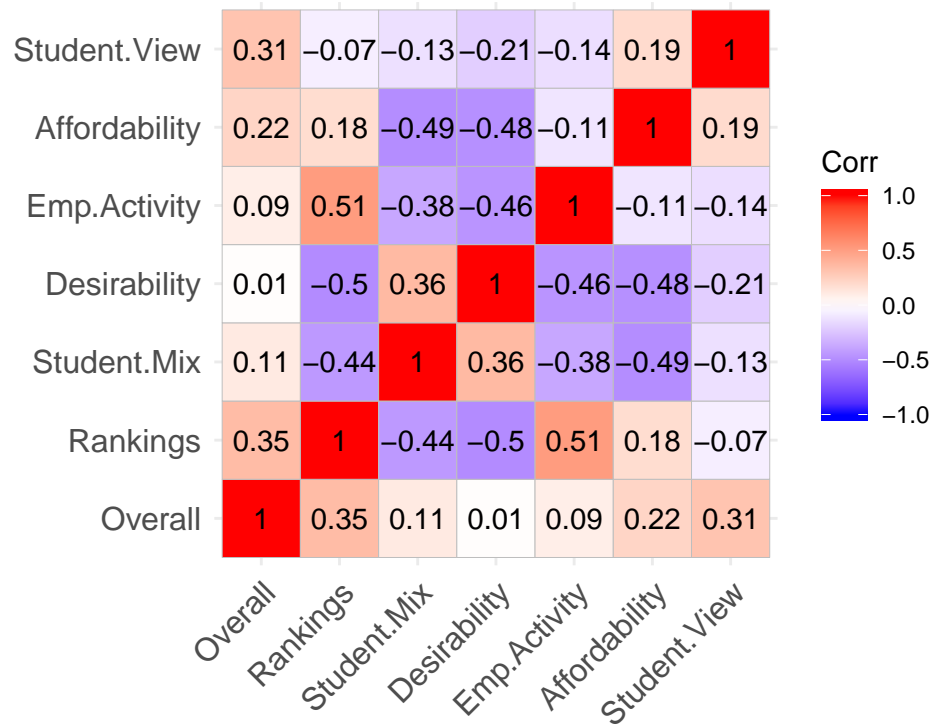
- Seoul
- Toronto
- New York

**Gruop Top30, cities from 21-30.**

```
# Using spearman

corr3 <- cor(data[data$Top=="Top30",c(4:10)], method="spearman")

ggcorrplot(corr3, lab = TRUE)
```



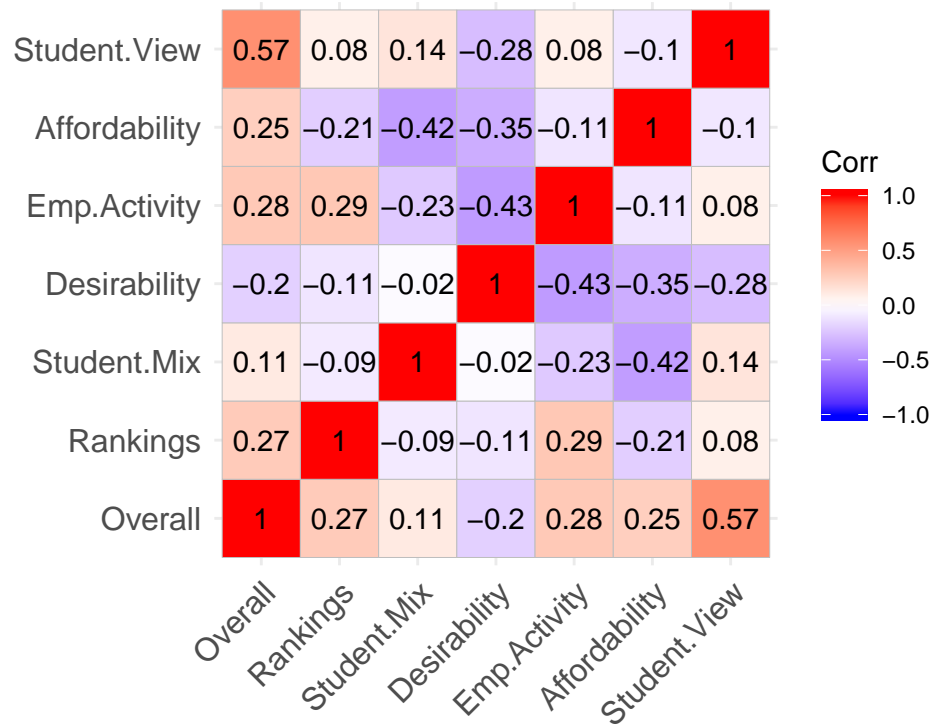
Some cities in this group: \* Canberra \* Auckland \* Prague

**Gruop Top40, cities from 31-40.**

```
# Using spearman

corr4 <- cor(data[data$Top=="Top40",c(4:10)], method="spearman")

ggcorrplot(corr4, lab = TRUE)
```



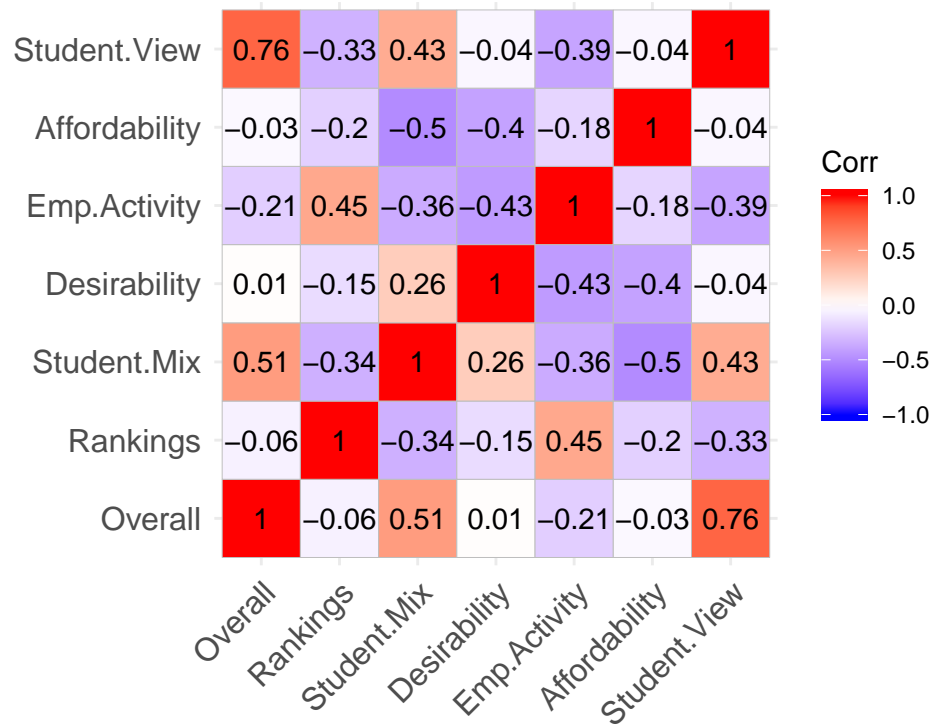
Some cities in this group: \* Barcelona \* Los Angeles \* Milan

**Gruop Top50, cities from 41-50.**

```
# Using spearman

corr5 <- cor(data[data$Top=="Top50",c(4:10)], method="spearman")

ggcorrplot(corr5, lab = TRUE)
```



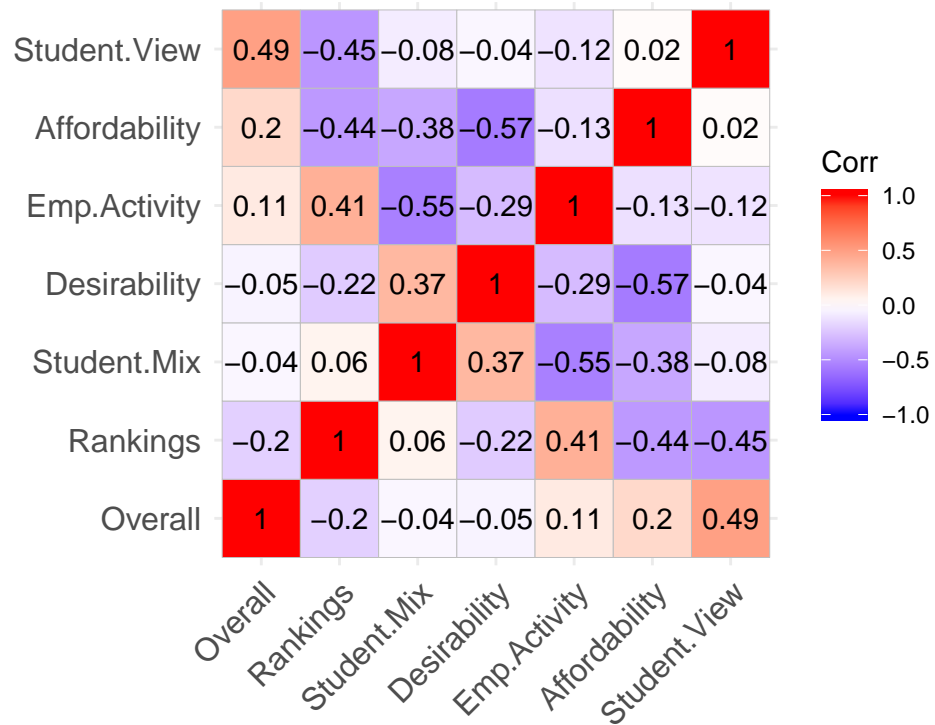
Some cities in this group: \* Lyon \* Ottawa \* Budapest

**Gruop Top60, cities from 51-60.**

```
# Using spearman

corr6 <- cor(data[data$Top=="Top60",c(4:10)], method="spearman")

ggcorrplot(corr6, lab = TRUE)
```



Some cities in this group: \* Mexico City \* Bangkok \* Philadelphia

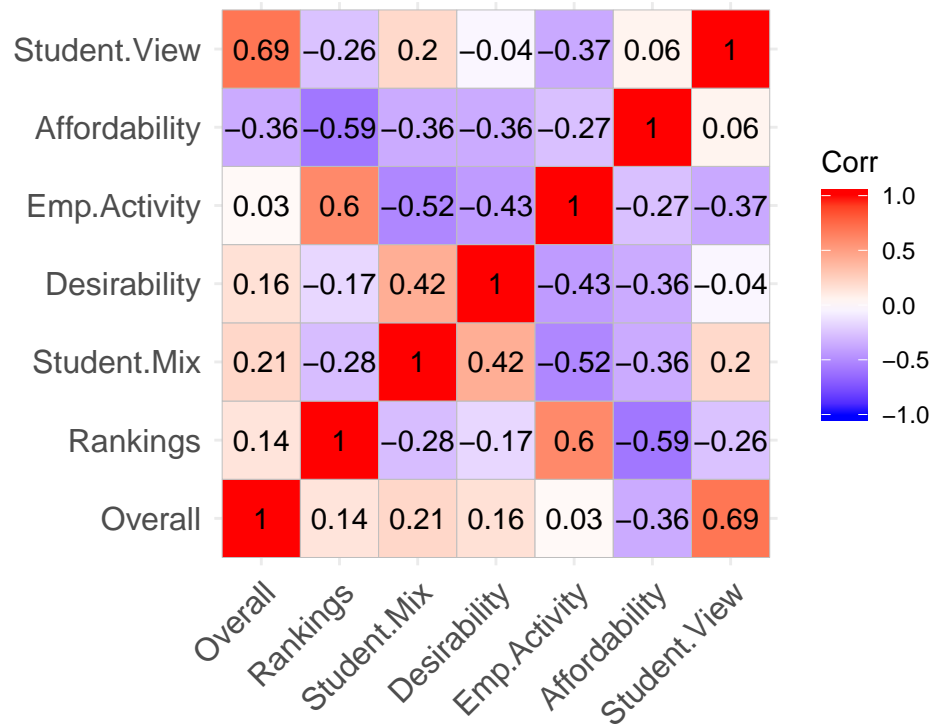
**Gruop Top70, cities from 61-70.**

```
# Using spearman

corr7 <- cor(data[data$Top=="Top70",c(4:10)], method="spearman")

ggcorrplot(corr7, lab = TRUE)
```





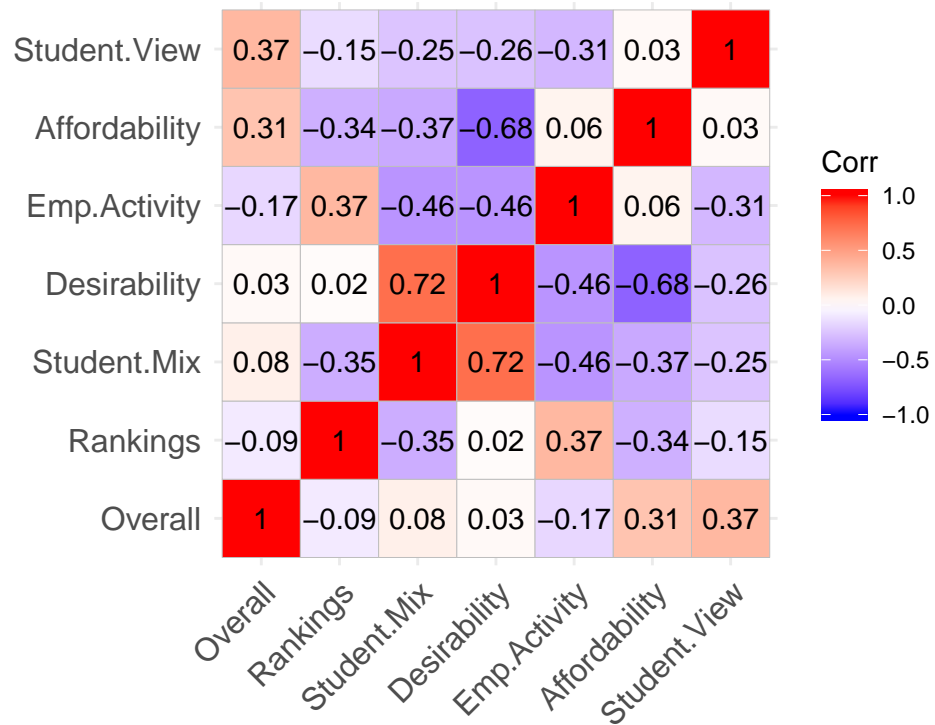
Some cities in this group: \* Dubai \* Oslo \* Rome

**Gruop Top80, cities from 71-80.**

```
# Using spearman

corr8 <- cor(data[data$Top=="Top80",c(4:10)], method="spearman")

ggcorrplot(corr8, lab = TRUE)
```



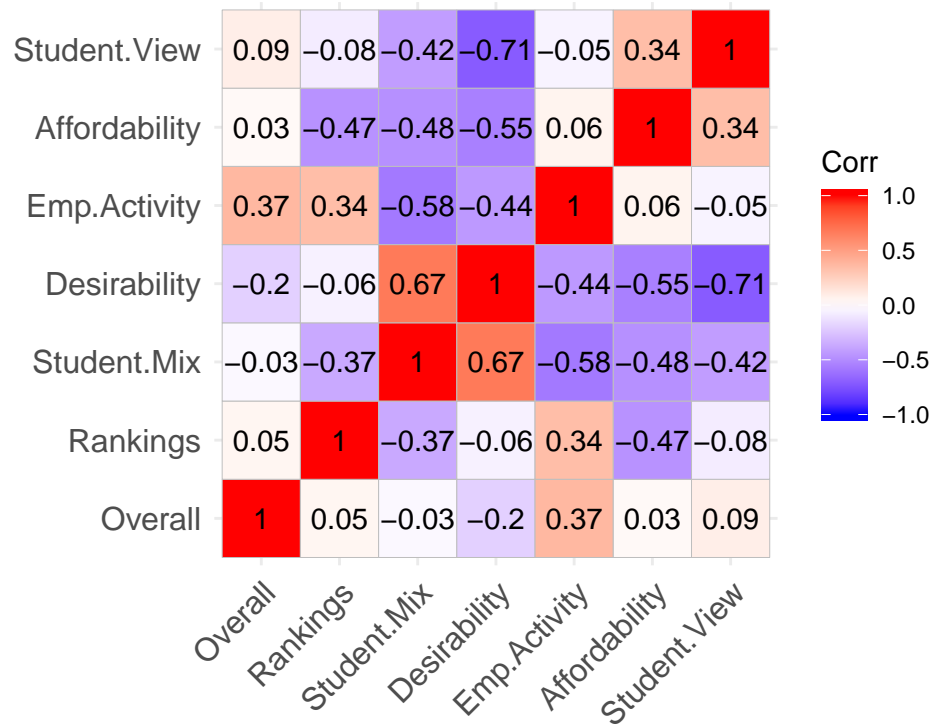
Some cities in this group: \* Cape Town \* Athens \* Toulouse

**Gruop Top90, cities from 81-90.**

```
# Using spearman

corr9 <- cor(data[data$Top=="Top90",c(4:10)], method="spearman")

ggcorrplot(corr9, lab = TRUE)
```



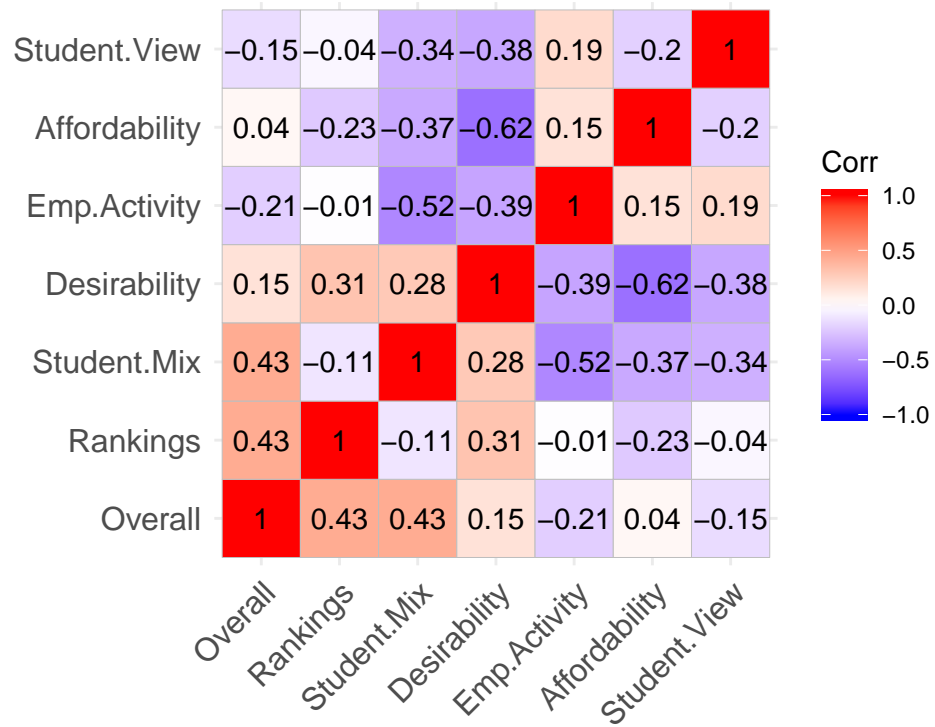
Some cities in this group: \* Bogota \* Monterrey \* Nanjing

**Gruop Top100, cities from 91-100.**

```
# Using spearman

corr10 <- cor(data[data$Top=="Top100",c(4:10)], method="spearman")

ggcorrplot(corr10, lab = TRUE)
```



Some cities in this group: \* Mainla \* Mumbai \* Rio de Janeiro