R Notebook

Hide

```r
library(dplyr)
library(tidyr)
library(tidyverse)
library(corrplot)
```

Hide

```r
##Q3.2
adm <- read_csv("admission.csv")
```

```
Parsed with column specification:
cols(
  `Serial No.` = [32mcol_double()[39m,
  `GRE Score` = [32mcol_double()[39m,
  `TOEFL Score` = [32mcol_double()[39m,
  `University Rating` = [32mcol_double()[39m,
  SOP = [32mcol_double()[39m,
  LOR = [32mcol_double()[39m,
  CGPA = [32mcol_double()[39m,
  Research = [32mcol_double()[39m,
  `Chance of Admit` = [32mcol_double()[39m
)
```

Hide

```r
#Q3.3
##Research from numeric to factor
adm$Research <- factor(adm$Research, levels = c(0,1))
##Drop all observations for which research is 0 (because why not)
adm_research <- adm %>% filter(Research == 1)
```

Hide

```r
#Q3.4
write_csv(adm_research, "output.csv")
```

Hide

```r
#Q3.5
summary(adm)
```

```
   Serial No.       GRE Score      TOEFL Score    University Rating      SOP
 Min.   :  1.0   Min.   :290.0   Min.   : 92.0   Min.   :1.000    Min.   :1.0
 1st Qu.:100.8   1st Qu.:308.0   1st Qu.:103.0   1st Qu.:2.000    1st Qu.:2.5
 Median :200.5   Median :317.0   Median :107.0   Median :3.000    Median :3.5
 Mean   :200.5   Mean   :316.8   Mean   :107.4   Mean   :3.087    Mean   :3.4
 3rd Qu.:300.2   3rd Qu.:325.0   3rd Qu.:112.0   3rd Qu.:4.000    3rd Qu.:4.0
 Max.   :400.0   Max.   :340.0   Max.   :120.0   Max.   :5.000    Max.   :5.0
      LOR            CGPA        Research   Chance of Admit
 Min.   :1.000   Min.   :6.800   0:181    Min.   :0.3400
 1st Qu.:3.000   1st Qu.:8.170   1:219    1st Qu.:0.6400
 Median :3.500   Median :8.610            Median :0.7300
 Mean   :3.453   Mean   :8.599            Mean   :0.7244
 3rd Qu.:4.000   3rd Qu.:9.062            3rd Qu.:0.8300
 Max.   :5.000   Max.   :9.920            Max.   :0.9700
```

Hide

```r
cat("Mean of SOP is: ", mean(adm$SOP))
```

```
Mean of SOP is:  3.4
```

Hide

```r
cat("\nDataset has ", sum(is.na(adm))," missing variables.")
```

```
Dataset has  0  missing variables.
```

Hide

```r
#Q3.6
adm_num <- which(sapply(adm,is.numeric))
adm_num <- adm_num[-1]
cor_num <- cor(adm[,adm_num])
corrplot.mixed(cor_num)
```