In [1]:

```
df = read.csv('C:\\Users\\hp\\Downloads\\Chrome\\Admission_Predict.csv')
df
```

A data.frame: 400 × 9

Serial.No.	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA	Research	Chance.of.A
<int>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1	337	118	4	4.5	4.5	9.65	1	0.92
2	324	107	4	4.0	4.5	8.87	1	0.76
3	316	104	3	3.0	3.5	8.00	1	0.72
4	322	110	3	3.5	2.5	8.67	1	0.80
5	314	103	2	2.0	3.0	8.21	0	0.65
6	330	115	5	4.5	3.0	9.34	1	0.90
7	321	109	3	3.0	4.0	8.20	1	0.75
8	308	101	2	3.0	4.0	7.90	0	0.68
9	302	102	1	2.0	1.5	8.00	0	0.50
10	323	108	3	3.5	3.0	8.60	0	0.45
11	325	106	3	3.5	4.0	8.40	1	0.52
12	327	111	4	4.0	4.5	9.00	1	0.84
13	328	112	4	4.0	4.5	9.10	1	0.78

In [2]:

```
str(df)
```

```
'data.frame':   400 obs. of  9 variables:
 $ Serial.No.       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ GRE.Score        : int  337 324 316 322 314 330 321 308 302 323 ...
 $ TOEFL.Score      : int  118 107 104 110 103 115 109 101 102 108 ...
 $ University.Rating: int  4 4 3 3 2 5 3 2 1 3 ...
 $ SOP              : num  4.5 4 3 3.5 2 4.5 3 3 2 3.5 ...
 $ LOR              : num  4.5 4.5 3.5 2.5 3 3 4 4 1.5 3 ...
 $ CGPA             : num  9.65 8.87 8 8.67 8.21 9.34 8.2 7.9 8 8.6 ...
 $ Research         : int  1 1 1 1 0 1 1 0 0 0 ...
 $ Chance.of.Admit  : num  0.92 0.76 0.72 0.8 0.65 0.9 0.75 0.68 0.5 0.45 ...
```

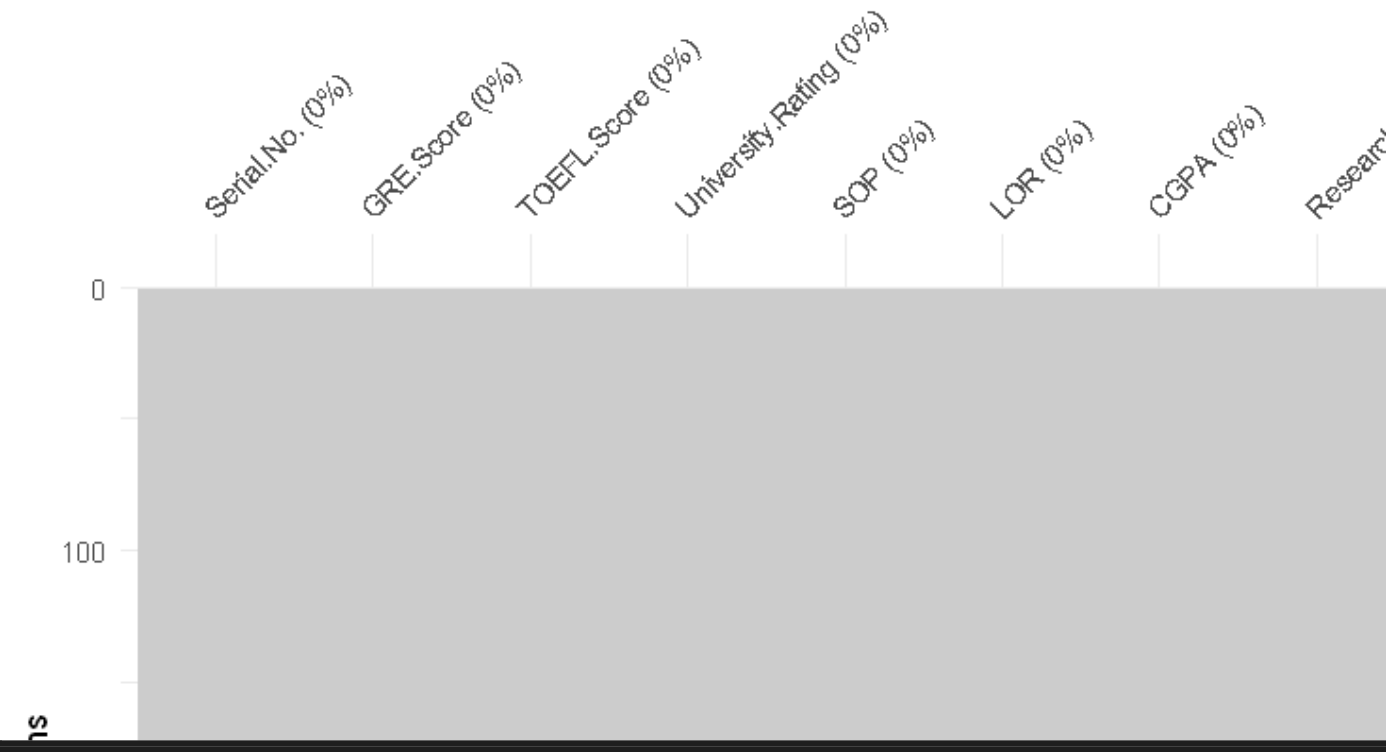
In [3]:

```
library(naniar)
```

```
Warning message:
"package 'naniar' was built under R version 3.6.3"
```

In [4]:

```
1 vis_miss(df)
```



In [5]:

```
1 attach(df)
```

In [6]:

```
1 cor(df)
```

A matrix: 9 × 9 of type dbl

	Serial.No.	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA
Serial.No.	1.00000000	-0.09752579	-0.1479317	-0.1699479	-0.1669324	-0.08822139	-0.0456(
GRE.Score	-0.09752579	1.00000000	0.8359768	0.6689759	0.6128307	0.55755452	0.83306
TOEFL.Score	-0.14793170	0.83597680	1.0000000	0.6955898	0.6579805	0.56772092	0.82841
University.Rating	-0.16994786	0.66897585	0.6955898	1.0000000	0.7345228	0.66012345	0.74647
SOP	-0.16693236	0.61283074	0.6579805	0.7345228	1.0000000	0.72959254	0.71814
LOR	-0.08822139	0.55755452	0.5677209	0.6601235	0.7295925	1.00000000	0.67021
CGPA	-0.04560845	0.83306045	0.8284174	0.7464787	0.7181440	0.67021130	1.00000
Research	-0.06313754	0.58039064	0.4898579	0.4477825	0.4440288	0.39685926	0.52165
Chance.of.Admit	0.04233586	0.80261046	0.7915940	0.7112503	0.6757319	0.66988879	0.87328

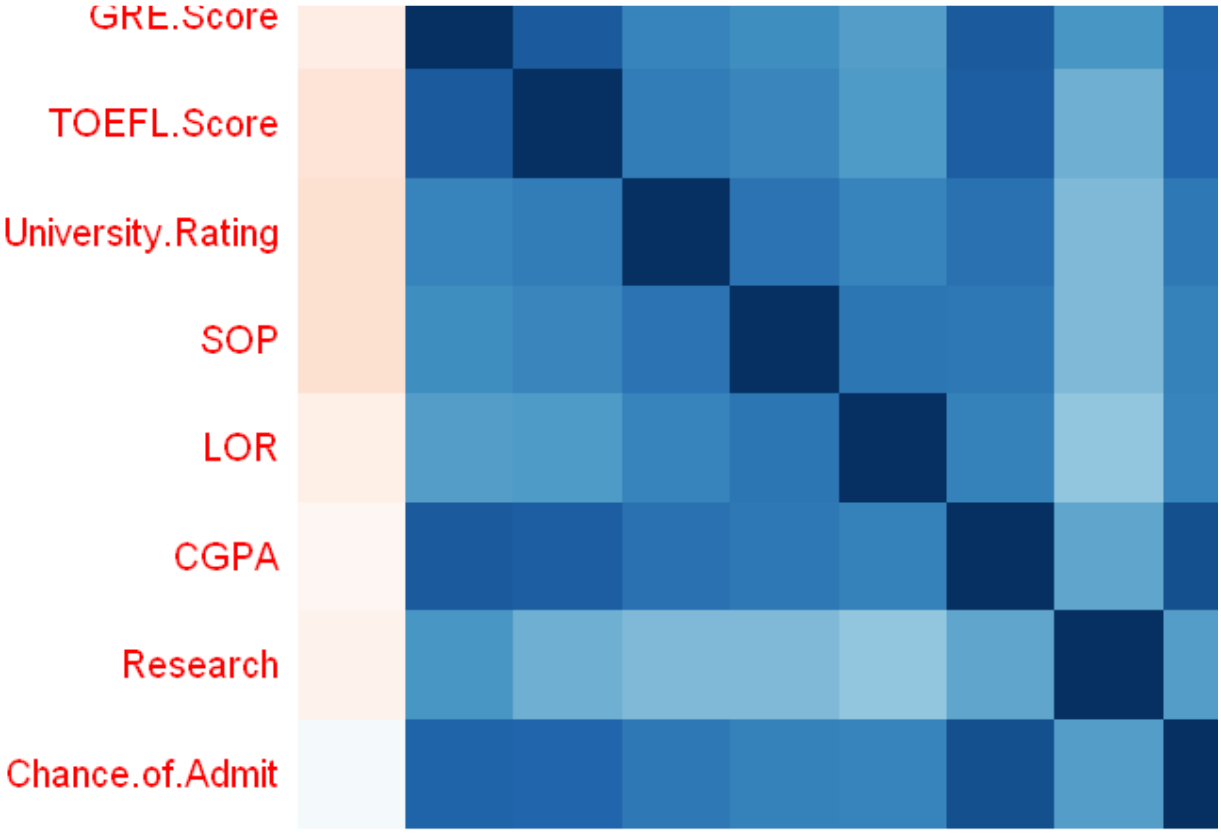
In [7]:

```
library(corrplot)
```

```
Warning message:
"package 'corrplot' was built under R version 3.6.3"corrplot 0.84 loaded
```

In [10]:

```
corrplot(cor(df),method = 'color')
```



In [13]:

```
model1 =lm(Chance.of.Admit~.,data=df)
model1
```

```
Call:
lm(formula = Chance.of.Admit ~ ., data = df)

Coefficients:
      (Intercept)         Serial.No.          GRE.Score         TOEFL.Score
        -1.294e+00          1.593e-04          1.799e-03          3.682e-03
University.Rating                SOP               LOR               CGPA
         8.785e-03          9.937e-05          2.154e-02          1.053e-01
         Research
         2.438e-02
```

In [14]:

```
summary(model1)
```

```
Call:
lm(formula = Chance.of.Admit ~ ., data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.233576 -0.026637  0.006226  0.038273  0.140252

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.294e+00  1.201e-01 -10.775  < 2e-16 ***
Serial.No.         1.593e-04  2.769e-05   5.753 1.77e-08 ***
GRE.Score          1.799e-03  5.749e-04   3.129 0.001885 **
TOEFL.Score        3.682e-03  1.056e-03   3.487 0.000543 ***
University.Rating  8.785e-03  4.617e-03   1.903 0.057821 .
SOP                9.937e-05  5.380e-03   0.018 0.985272
LOR                2.154e-02  5.330e-03   4.041 6.41e-05 ***
CGPA               1.053e-01  1.198e-02   8.786  < 2e-16 ***
Research           2.438e-02  7.653e-03   3.185 0.001561 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06132 on 391 degrees of freedom
Multiple R-squared:  0.8188,    Adjusted R-squared:  0.8151
F-statistic: 220.9 on 8 and 391 DF,  p-value: < 2.2e-16
```

In [15]:

```
str(df)
```

```
'data.frame':   400 obs. of  9 variables:
 $ Serial.No.       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ GRE.Score        : int  337 324 316 322 314 330 321 308 302 323 ...
 $ TOEFL.Score      : int  118 107 104 110 103 115 109 101 102 108 ...
 $ University.Rating: int  4 4 3 3 2 5 3 2 1 3 ...
 $ SOP              : num  4.5 4 3 3.5 2 4.5 3 3 2 3.5 ...
 $ LOR              : num  4.5 4.5 3.5 2.5 3 3 4 4 1.5 3 ...
 $ CGPA             : num  9.65 8.87 8 8.67 8.21 9.34 8.2 7.9 8 8.6 ...
 $ Research         : int  1 1 1 1 0 1 1 0 0 0 ...
 $ Chance.of.Admit  : num  0.92 0.76 0.72 0.8 0.65 0.9 0.75 0.68 0.5 0.45 ...
```

In [16]:

```
library(dplyr)
```

```
Warning message:
"package 'dplyr' was built under R version 3.6.3"
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
```

In [24]:

```
df.data <- select(df,CGPA,GRE.Score,TOEFL.Score,Chance.of.Admit )
head(df.data)
```

A data.frame: 6 × 4

	CGPA	GRE.Score	TOEFL.Score	Chance.of.Admit
	<dbl>	<int>	<int>	<dbl>
1	9.65	337	118	0.92
2	8.87	324	107	0.76
3	8.00	316	104	0.72
4	8.67	322	110	0.80
5	8.21	314	103	0.65
6	9.34	330	115	0.90

In [26]:

```
cgpa<- factor(df.data$CGPA)
gre <- factor(df.data$GRE.Score)
Toefl <- factor(df.data$TOEFL.Score)
admission<- factor(df.data$Chance.of.Admit )
str(df.data)
```

```
'data.frame':   400 obs. of  4 variables:
 $ CGPA           : num  9.65 8.87 8 8.67 8.21 9.34 8.2 7.9 8 8.6 ...
 $ GRE.Score      : int  337 324 316 322 314 330 321 308 302 323 ...
 $ TOEFL.Score    : int  118 107 104 110 103 115 109 101 102 108 ...
 $ Chance.of.Admit: num  0.92 0.76 0.72 0.8 0.65 0.9 0.75 0.68 0.5 0.45 ...
```

In [29]:

```
library(ggplot2)
```

```
Warning message:
"package 'ggplot2' was built under R version 3.6.3"
```

In [30]:

```
library(caTools)
#set.seed(101)
split = sample.split(df.data$Chance.of.Admit ,SplitRatio = 0.70)
final.train<- subset(df.data,split ==TRUE)
final.test <- subset(df.data,split == FALSE)
final.log.model<- glm(formula = Chance.of.Admit~.,data =final.train )
summary(final.log.model)
```

```
Call:
glm(formula = Chance.of.Admit ~ ., data = final.train)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-0.282508  -0.022490   0.008799   0.040702   0.128255

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.5316117  0.1253767 -12.216  < 2e-16 ***
CGPA         0.1558598  0.0137298  11.352  < 2e-16 ***
GRE.Score    0.0016270  0.0007167   2.270  0.02396 *
TOEFL.Score  0.0036871  0.0013651   2.701  0.00734 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.004503143)

    Null deviance: 5.6774  on 281  degrees of freedom
Residual deviance: 1.2519  on 278  degrees of freedom
AIC: -717.39

Number of Fisher Scoring iterations: 2
```

In [32]:

```
fitted.probabilities <- predict(final.log.model,newdata = final.test,type= 'respon
fitted.results <- ifelse(fitted.probabilities > 0.5,1,0)
```

In [37]:

```
misclassicerror <- mean(fitted.results!=final.test$Chance.of.Admit,na.rm =T )
```

In [38]:

```
print(paste('accuracy', 1- misclassicerror))
```

```
[1] "accuracy 0"
```

In [39]:

```
misclassicerror
```

```
1
```

In [40]:

```
table(final.test$Chance.of.Admit,fitted.probabilities>0.5)
```

	FALSE	TRUE
0.34	0	1
0.36	1	0
0.38	0	1
0.42	0	1
0.44	1	0
0.45	1	0
0.46	1	0
0.47	1	0
0.48	0	1
0.49	0	1
0.5	0	1
0.52	0	1
0.53	0	1
0.54	0	1
0.56	0	2
0.57	1	1
0.58	0	1
0.59	0	1
0.61	0	2
0.62	0	3
0.63	0	2
0.64	0	5
0.65	0	3
0.66	0	2
0.67	0	2
0.68	0	3
0.69	0	2
0.7	0	4
0.71	0	5
0.72	0	5
0.73	0	4
0.74	0	3
0.75	0	2
0.76	0	4
0.77	0	2
0.78	0	4
0.79	0	4
0.8	0	3
0.81	0	2
0.82	0	2
0.83	0	1
0.84	0	3
0.85	0	2
0.86	0	2
0.87	0	1
0.88	0	1
0.89	0	3
0.9	0	2
0.91	0	2
0.92	0	2
0.93	0	3
0.94	0	4
0.95	0	1

```
0.96     0    2
0.97     0    1
```

In [45]:

```
install.packages('Amelia')
```

```
Warning message:
"dependency 'foreign' is not available"also installing the dependency 'RcppArmadillo'


package 'RcppArmadillo' successfully unpacked and MD5 sums checked
package 'Amelia' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\hp\AppData\Local\Temp\RtmpWmiYJB\downloaded_packages
```

In [47]:

```
install.packages('RcppArmadillo')
```

```
package 'RcppArmadillo' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\hp\AppData\Local\Temp\RtmpWmiYJB\downloaded_packages
```

In [50]:

```
install.packages('foreign')
```

```
Warning message:
"package 'foreign' is not available (for R version 3.6.1)"
```

In [49]:

```
library(Amelia)
```

```
Warning message:
"package 'Amelia' was built under R version 3.6.3"


Error: package or namespace load failed for 'Amelia' in loadNamespace(i, c(lib.loc, .libPaths()), versionCheck
 there is no package called 'foreign'
Traceback:

1. library(Amelia)
2. tryCatch({
 .     attr(package, "LibPath") <- which.lib.loc
 .     ns <- loadNamespace(package, lib.loc)
 .     env <- attachNamespace(ns, pos = pos, deps, exclude, include.only)
 . }, error = function(e) {
 .     P <- if (!is.null(cc <- conditionCall(e)))
 .         paste(" in", deparse(cc)[1L])
 .     else ""
 .     msg <- gettextf("package or namespace load failed for %s%s:\n %s",
 .         sQuote(package), P, conditionMessage(e))
 .     if (logical.return)
 .         message(paste("Error:", msg), domain = NA)
 .     else stop(msg, call. = FALSE, domain = NA)
 . })
3. tryCatchList(expr, classes, parentenv, handlers)
4. tryCatchOne(expr, names, parentenv, handlers[[1L]])
5. value[[3L]](cond)
6. stop(msg, call. = FALSE, domain = NA)
```

In [52]:

```
test_roc = roc(final.test$Chance.of.Admit~ test_prob,plot= TRUE, print.auc = TRUE)
test_roc
```

```
Error in roc(final.test$Chance.of.Admit ~ test_prob, plot = TRUE, print.auc = TRUE): could not find function "r
Traceback:
```