1.a

Logo, company name

Description automatically generated

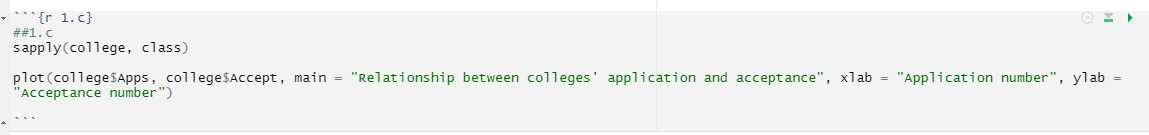
1.b

Graphical user interface, text, application, email

Description automatically generated

1.c

I firstly checked which features are numeric then I plot a scatterplot to show the relation ship between colleges’ application and acceptance.

Table

Description automatically generated with low confidence

Chart, scatter chart

Description automatically generated

1.d

Graphical user interface, text

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

1.e

Text

Description automatically generated

Chart, box and whisker chart

Description automatically generatedDiagram, box and whisker chart

Description automatically generated

Totally there are 26 top universities.

1.f

A picture containing scatter chart

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

I would like to take a deep look in graduate rates of top schools and other schools. My hypotheses was top schools had lower graduate rate since they are intuitively 'better' than other schools. However, the interesting fact is that top schools generally have higher graduate rate (around 90%) than others (around 60%). The reason probably is the students in top schools are better at studying so even it is more difficult to graduate, top schools sill have higher graduate rates.

2.a

A picture containing text

Description automatically generated

'FFMC', 'DMC', 'DC', 'ISI', 'temp', 'RH', 'wind', 'rain', 'area' can be considered as quantitative predictors. 'month' and 'day' can be considered as qualitative predictors, but they are easily represented as quantitative predictors."

2.b

A screenshot of a computer

Description automatically generated with medium confidence

Table

Description automatically generated with medium confidence

2.c

Text

Description automatically generated with low confidence

Graphical user interface, table

Description automatically generated with medium confidence

2.d

A picture containing text

Description automatically generated

Chart

Description automatically generated

As the bar plot shows, august has the most count of forest fires.

2.e

I firstly transferred all “month” and “day” data to numeric data.

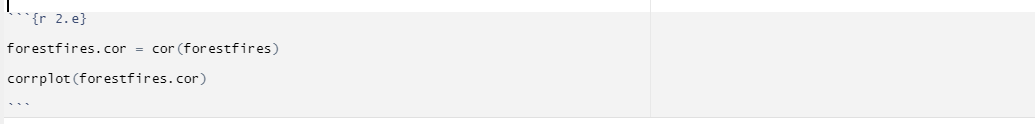
Table

Description automatically generated with medium confidence

Graphical user interface, text, application, Word

Description automatically generated

Then create correlation matrix for all relevant variables.



Chart, bubble chart

Description automatically generated

2.f

As above figure shown, temperature in degrees Celsius(temp) and relative humidity(RH ) might be useful in predicting area since the area burned by the forest fire has higher positive correlation coefficient with temp, and higher negative correlation coefficient with RH, which means the burned area goes up with the higher temperature and lower relative humidity in some cases.