Activity 1: Exploratory Data Analysis
Sridhar Sampath (ssampath30)

As the exercise suggested, initial analysis of the data was performed. The data was loaded using

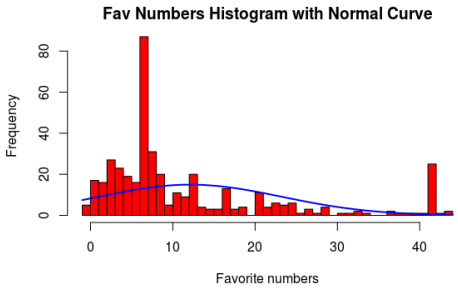
survey = read.csv('ac1_survey.csv', stringsAsFactors=FALSE)

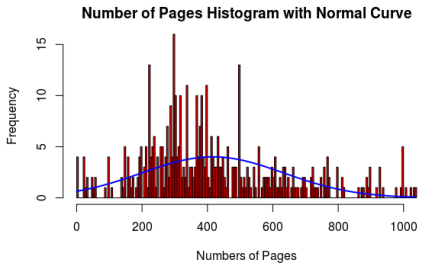
The data had dimension of [530, 8]. Here “8” represents the features columns. The following were the feature columns present in the dataset.

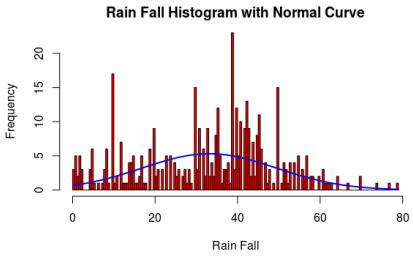
```
"Feature (Column) Information"
"Timestamp"
"What.is.your.favorite.number."
"What.was.the.maximum.temperature.in.your.area.yesterday."
"What.is.the.name.of.your.favorite.comic.book.character."
"What.genre.of.music.do.you.like.to.listen.to."
"How.much.total.rainfall.did.your.city.region.receive.in.2016...in.inches."
"What.is.the.title.of.the.book.that.you.are.reading..or.last.read.."
"How.many.pages.are.there.in.that.book."
```

In order to understand more about the features, head and str commands were used to display sample data information of each features. To proceed furthere with analysis, data cleaning is required. Here for this activity, only features with integers are considered. To filter data to contain only integers, ***na.omit and is.finite*** was used.

Sample hist plot of the data is shown below.

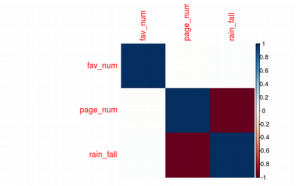






Further, these information was used to find the correlation between them. The correlation information between favorite number, Number of pages in the book and rain fall in the region is shown below.

	fav_num	page_num	rain_fall
fav_num	1	-0.003247092	0.003147507
page_num	-0.003247092	1	-0.999665371
rain_fall	0.003147507	-0.999665371	1



As seen from the table, there is no correlation between favorite number and page number. But there is a negative correlation between page number and rain_fall. This implies that if there is a increase rain fall, then number of pages of the book used for reading decreased and vice versa.