SRT411-Assignment2-ldyer1

Lorraine Dyer

February 25, 2017

Assignment 2: Data Analysis at Home and on the Web

Student:Lorraine Dyer

Student Number: 014057111

Date Due: Feb 23 2017

The defined problem is as follows:

Which models of hard drives are more likely to fail?

Available Data:

Statistics about Hard Drives from the years 2013 and 2016.

Parsing Information - (See data below)

library(readr) X2016_12_31 <- read_csv("~/Lorraine/Winter2017-Semester7/SRT411/Data_Sets/Test/data_Q4_2016/2016-12-31.csv", trim_ws = FALSE) head(X2016_12_31)

```
library(readr)
X2016_12_31 <- read_csv("~/Lorraine/Winter2017-Semester7/SRT411/Data_Sets/Test/data_Q4_2016/2016-12-31.
                        trim_ws = FALSE)
```

Data Aggregation (see below)

2016 Results - Model versus failure

```
barplot(tcsvfile,main = " 2016 Results - Drive Failure according to model",xlab = "model",ylab = "Drive
names.arg("A","B","C","D","E","F","G")
```

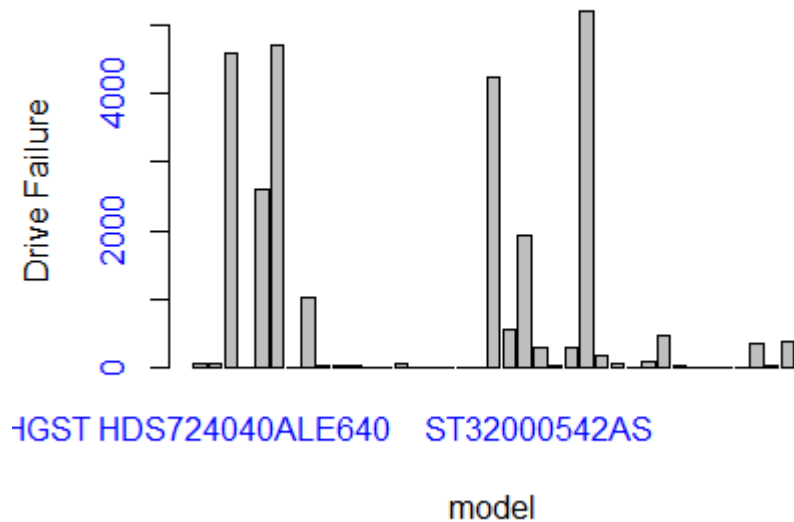## 2016 Results - Drive Failure according to mod

image:

```
barplot(X2016_12_31, main = "2016 Results - Model versus failure", xlab = "failure", ylab = "Model Seri
names.arg=c("A","B", "C","D","E","F","G")

X2013_12_31 <- read_csv("~/Lorraine/Winter2017-Semester7/SRT411/Data_Sets/Test/data_2013/2013/2013-12-3
                        trim_ws = FALSE)
```

Cleaning Data

csvfile<-X2013_12_31

colSums(is.na(csvfile[]))

csvfile.clean<-na.omit(csvfile) nrow(csvfile.clean)

head(csvfile)

Drive Failure According to Model and Capacity (or Drive Failure)
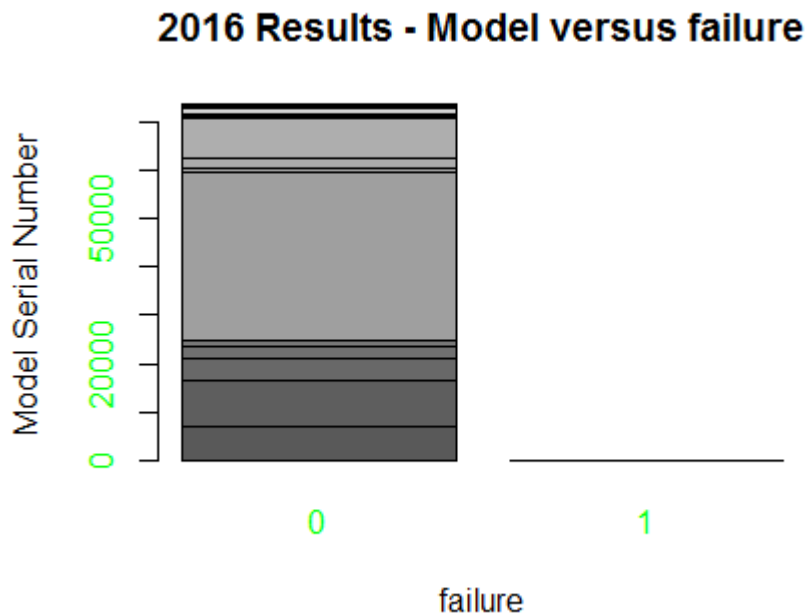
2016 Results - Model versus failure

image:

```
barplot(tcsvfile,main = " 2016 Results - Drive Failure accoring to model",xlab = "model",ylab = "Drive I
names.arg("A","B","C","D","E","F","G")
```
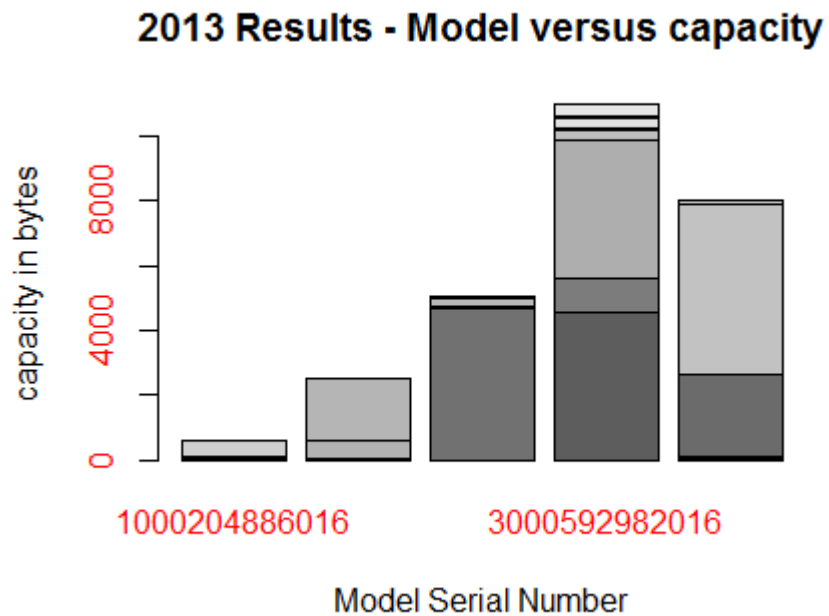
2013 Results - Model versus capacity

image:

```
barplot(scsvfile, main = "2013 Results - Model versus capacity", xlab = "Model Serial Number", ylab = "
names.arg=c("A","B", "C","D","E","F","G")
```

```
#2016 Model versus Capacity
image: ![](C:\Users\lorra\Documents\Lorraine\Winter2017-Semester7\SRT411-assign2-imgs\2016-Results-Modv

barplot(scsvfile, main = "2013 Results - Model versus capacity", xlab = "Model Serial Number", ylab = "
names.arg=c("A","B", "C","D","E","F","G")
```

Conclusions

2016 - Drive Failure According to Hard Drive Model

+ According to this bar graph the lower the alphabetical Hard Drive Model,

+ has almost the same amount of failure as the higher alphabetical Hard Drive Model.

2016 Drive failure according to Model Serial Number

+ Drives with a serial number less than 20000 were less likely to fail.

2013 Results - Model versus capacity of drive

+It was determined that the lower the model serial number the lesser the capacity of the Hard Drive.

Source of Data: https://www.backblaze.com/blog/hard-drive-failure-rates-q3

" '