

Hive - October Requests Histogram

Christopher Rusnak (UNI: cjr2176)

March 19, 2017

```
library(ggplot2)

# Read output files from Hive
file1 <- read.table("000000_0",header = FALSE, sep = "\t", col.names = c("freq", "day"))
file2 <- read.table("000001_0",header = FALSE, sep = "\t", col.names = c("freq", "day"))

# Merge output files into a single file
dat <- rbind(file1,file2)

# Histogram plot of number of requests made in a day for every day in the month of October
ggplot(data=dat, aes(x = dat$freq)) + geom_histogram(bins = 30) +
  xlab("Request Frequency") + ylab("Count of Days") +
  ggtitle("Number of Requests per Day: October 1995")
```

