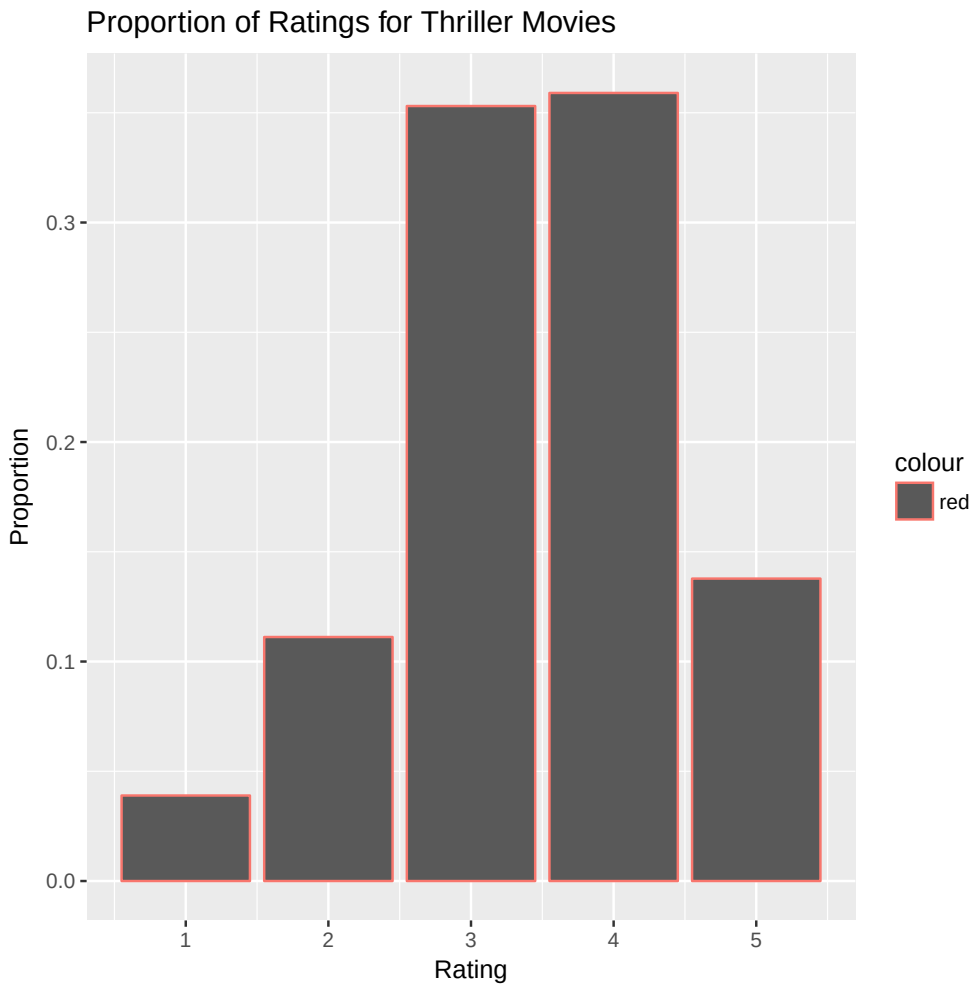
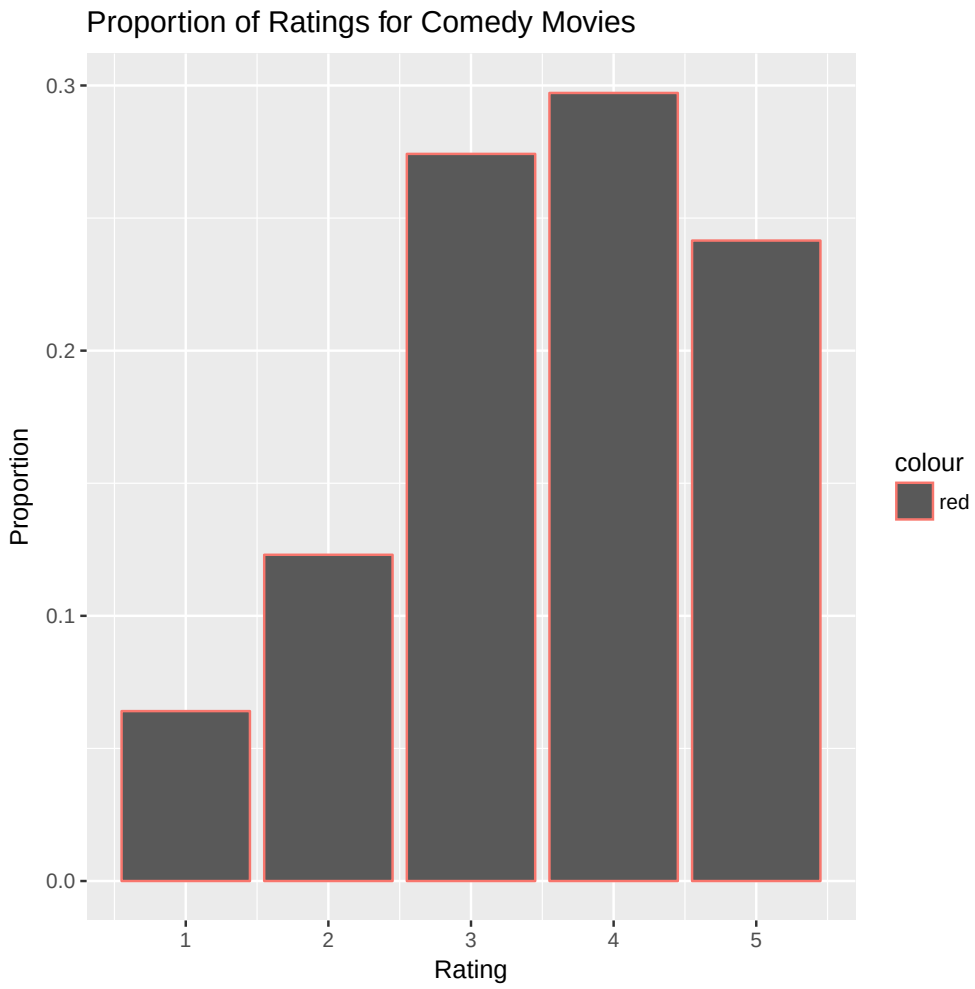
Statistics 1559
Final Project
Alex Wassel and Alana McBane

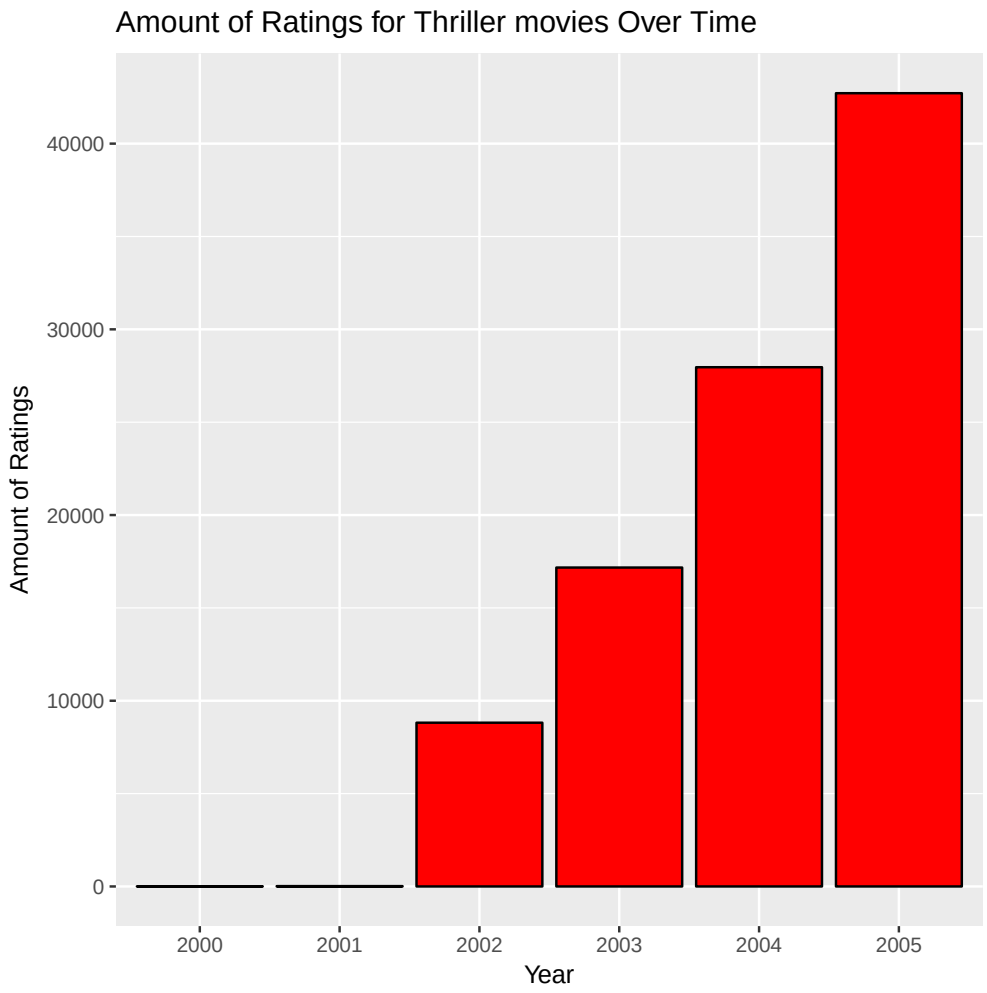
```
> setwd("/Users/alanamcbane/Documents/Data_1559")
> data <- read.csv("Reviews_requested.csv", header = TRUE)
> library(ggplot2)
> #Subsetting the entire data based off each individual movies
> scarsofdracula<-data[1:60,]
> montypython<-data[61:19677,]
> astrozombies<-data[19678:19788,]
> markofthedevil<-data[19789:19991,]
> americanpie2<-data[19992:110236,]
> theoffice<-data[110237:136080,]
> behindenemylines<-data[136081:215637,]
> zoolander<-data[215638:266507,]
> jeeperscreepers<-data[266508:283269,]
> shrek<-data[283270:286310,]
> #Subsetting each individual movies into the thriller genre
> thriller<-rbind(scarsofdracula, astrozombies, markofthedevil, behindenemylines,
+ jeeperscreepers)
> #Subsetting each individual movies into the comedy genre
> comedy<-rbind(montypython, americanpie2, theoffice, zoolander, shrek)
>

> #Graph showing the proportion of ratings for thriller movies
> ggplot(thriller, aes(x=rating, color = "red")) + geom_bar(aes(y=..prop..)) +
+ labs(title = "Proportion of Ratings for Thriller Movies", x = "Rating",
+ y = "Proportion")
>
```



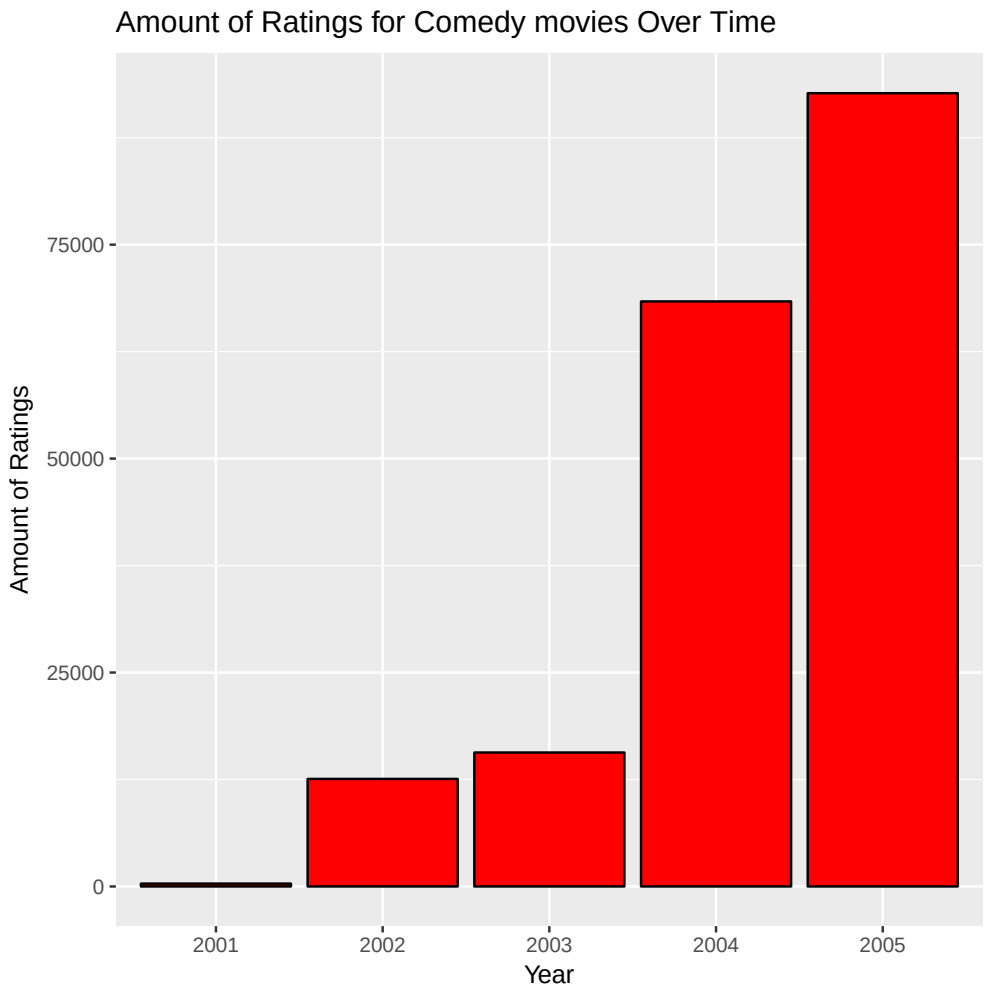

```
> #Graph showing the proportion of ratings for comedy movies
> ggplot(comedy, aes(x=rating, color = "red")) + geom_bar(aes(y=..prop..)) +
+ labs(title = "Proportion of Ratings for Comedy Movies", x = "Rating",
+ y = "Proportion")
> #Subsetting dates into day, month, years
> date1<-do.call(rbind, strsplit(as.character(thriller$date), '-'))
> thriller$year<-date1[,1]
> thriller$month<-date1[,2]
> thriller$day<-date1[,3]
> date2<-do.call(rbind, strsplit(as.character(comedy$date), '-'))
> comedy$year<-date2[,1]
> comedy$month<-date2[,2]
> comedy$day<-date2[,3]
>
```




```
> #Graph showing the amount of ratings for thriller movies over time
> ggplot(thriller, aes(x=year)) + geom_bar(position="dodge", colour="black",
+ fill="red") + labs(title = "Amount of Ratings for Thriller movies Over Time",
+                    x="Year", y = "Amount of Ratings")
```




```
> #Graph showing the amount of ratings for comedy movies over time
> ggplot(comedy, aes(x=year)) + geom_bar(position="dodge", colour="black",
+ fill="red") + labs(title = "Amount of Ratings for Comedy movies Over Time",
+ x = "Year", y = "Amount of Ratings")
```




```
> #Two-way table for comedy and thriller movies (count) with movie ids as
> #rows (movie id key on report)
> options(digits = 2)
> twoway <- xtabs(~movie.id + rating, data = data)
> twoway
```

movie.id	rating 1	2	3	4	5
6743	10	7	27	10	6
7237	972	1208	3765	6613	7059
7546	41	36	23	8	3
8706	45	56	69	22	11
10170	4859	10944	27717	28372	18353
12494	2061	7750	27999	30434	11313
12792	1856	2273	3575	6691	11449
15233	4422	8829	16574	13507	7538
16516	1610	2900	6021	4242	1989
16615	41	75	361	1168	1396

>

References:

- Class Notes from the semester