HW 2

Marion Geary

```
library(fivethirtyeight)
library(tidyverse)
library(knitr)
library(ggthemes)
data("bechdel")
```

Exercise 1

```
bechdel90.13 <- filter(bechdel, between(year, 1990, 2013))
```

Exercise 2

```
budget_summary_binary <- bechdel90.13 %>% group_by(binary) %>% select(budget_2013, domgross_2013, intgr
budget_summary_binary
```

```
## # A tibble: 2 x 4
##   binary med_budget med_domgross med_intgross
##   <chr>       <dbl>        <dbl>        <dbl>
## 1 FAIL    48385984.    57318606.    104475669
## 2 PASS    31070724     45330446.     80124349
```

Though these analysis, we see that movies that fail the Bechdel test have higher median budgets and higher median international and domestic gross.

Exercise 3

```
budget_summary_clean <- bechdel90.13 %>% group_by(clean_test) %>% select(budget_2013, domgross_2013, in
budget_summary_clean
```

```
## # A tibble: 5 x 4
##   clean_test med_budget med_domgross med_intgross
##   <ord>           <dbl>        <dbl>        <dbl>
## 1 nowomen     43373066     44891296.    89509349
## 2 notalk      56570084.    63890455    123102194
## 3 men         39737690.    56392786     99578022.
## 4 dubious     35790994     49173429     89883201
## 5 ok          31070724     45330446.    80124349
```

Dubious means that some contributors to the data were skeptical about whether the film passed the test.

Through grouping by the specific results of the Bechdel test, we find that movies that have women that never talk to each other have the highest median budgets, domestic gross, and international gross. These movie's numbers are significantly higher than those in the categories. Movies that pass the test still have the lowest budgets and lowest international gross. They have the second lowest domestic gross by a small margin.

Exercise 4

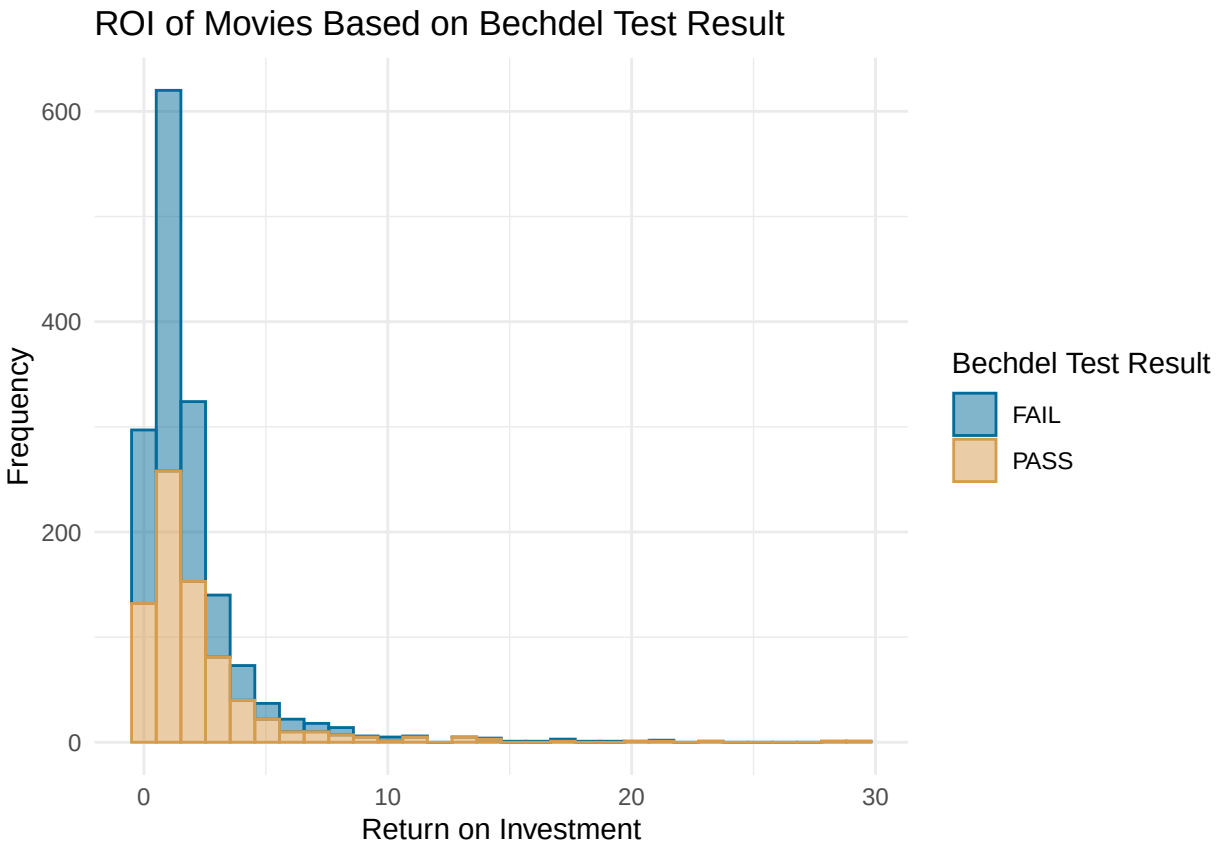
```
bechdel90.13 <- bechdel90.13 %>% mutate(roi = round(domgross_2013 / budget_2013, digits = 2), .after = y
```

Exercise 5

```
roi_over_30 <- bechdel90.13 %>% filter(roi > 30) %>% select(year, roi, title, clean_test, binary, budget

kable(roi_over_30)
```

year	roi	title	clean_test	binary	budget_2013	domgross_2013
2010	36.01	Insidious	ok	PASS	1602348	57694294
2004	45.99	Saw	men	FAIL	1479831	68053804
1998	47.37	Pi	men	FAIL	97181	4603452
1997	48.03	Chasing Amy	ok	PASS	362810	17424311
2002	48.29	My Big Fat Greek Wedding	ok	PASS	6475896	312705749
2012	53.26	The Devil Inside	ok	PASS	1014639	54041622
1991	53.37	Slacker	ok	PASS	39349	2100070
2004	60.68	Primer	notalk	FAIL	8632	523811
2006	62.97	Once	men	FAIL	173369	10917487
2008	66.91	Fireproof	men	FAIL	541128	36208310
2004	111.35	Napoleon Dynamite	notalk	FAIL	493277	54927590
1994	113.83	Clerks.	notalk	FAIL	42435	4830398
1997	115.35	In the Company of Men	notalk	FAIL	36281	4184879
1999	234.23	The Blair Witch Project	ok	PASS	839077	196538593
2007	239.82	Paranormal Activity	dubious	FAIL	505595	121251476
1992	291.57	El Mariachi	nowomen	FAIL	11622	3388636

```
roi_under_30 <- bechdel90.13 %>% filter(roi < 30) %>% select(year, roi, title, clean_test, binary, budg

library(wesanderson)
my_colors <- wesanderson::wes_palette("Darjeeling2")[2:4]

ggplot(roi_under_30, aes(x = roi, fill = binary, color = binary)) + geom_histogram(alpha = 0.5)  + theme
  scale_fill_manual(values = my_colors) + labs(x = "Return on Investment", y = "Frequency", title = "RO
```



Through this analysis, I noticed that a higher proportion of movies with lower ROIs fail the Bechdel test. Fewer of the movies with high ROIs pass the Bechdel test. Of the high ROI (`roi > 30`) movies that do not pass the Bechdel test, only one has no women. Overall, most movies have low ROIs, but the data is very right skewed.