Lab 02

due January 19th at 11:59 PM

Dav King

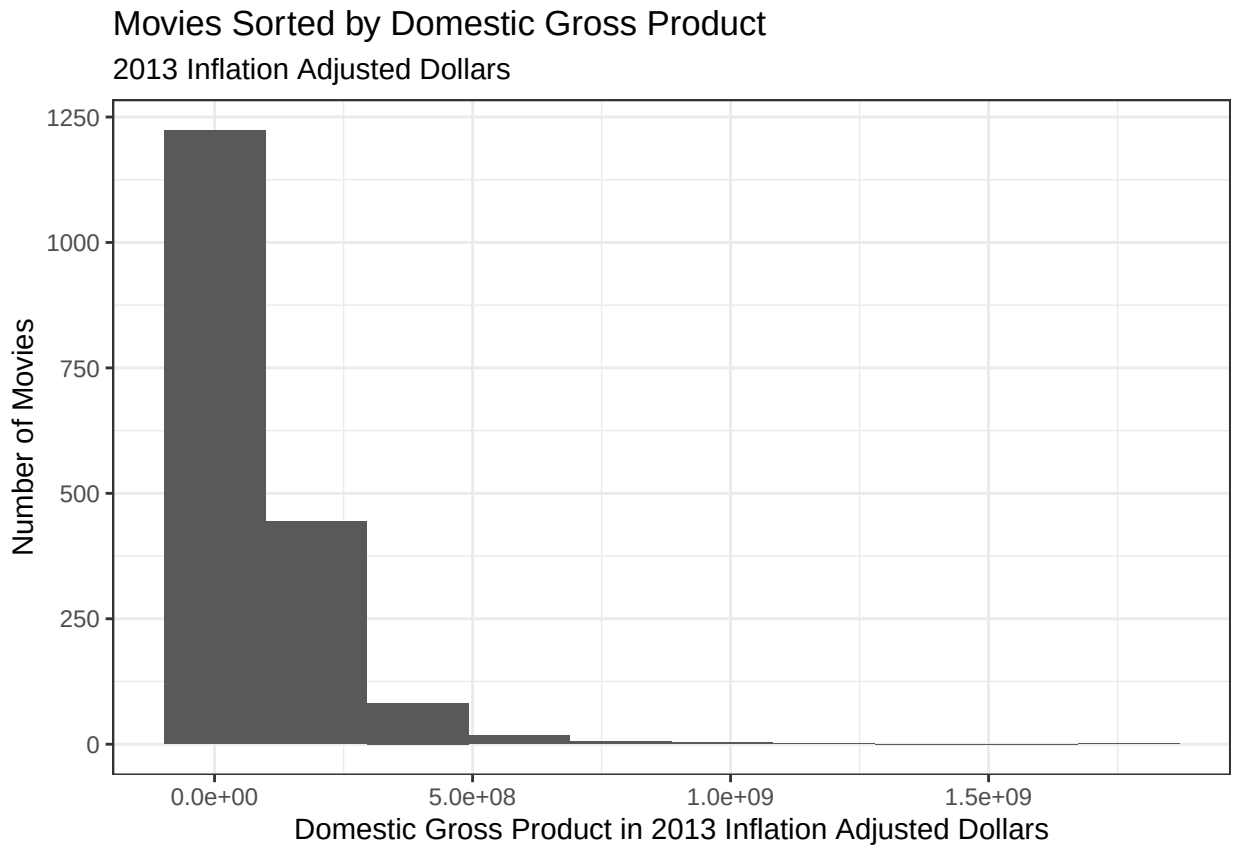
1/10/22

Load Packages

```
library(tidyverse)
library(viridis)
library(fivethirtyeight)
library(scales)
```

Exercise 1

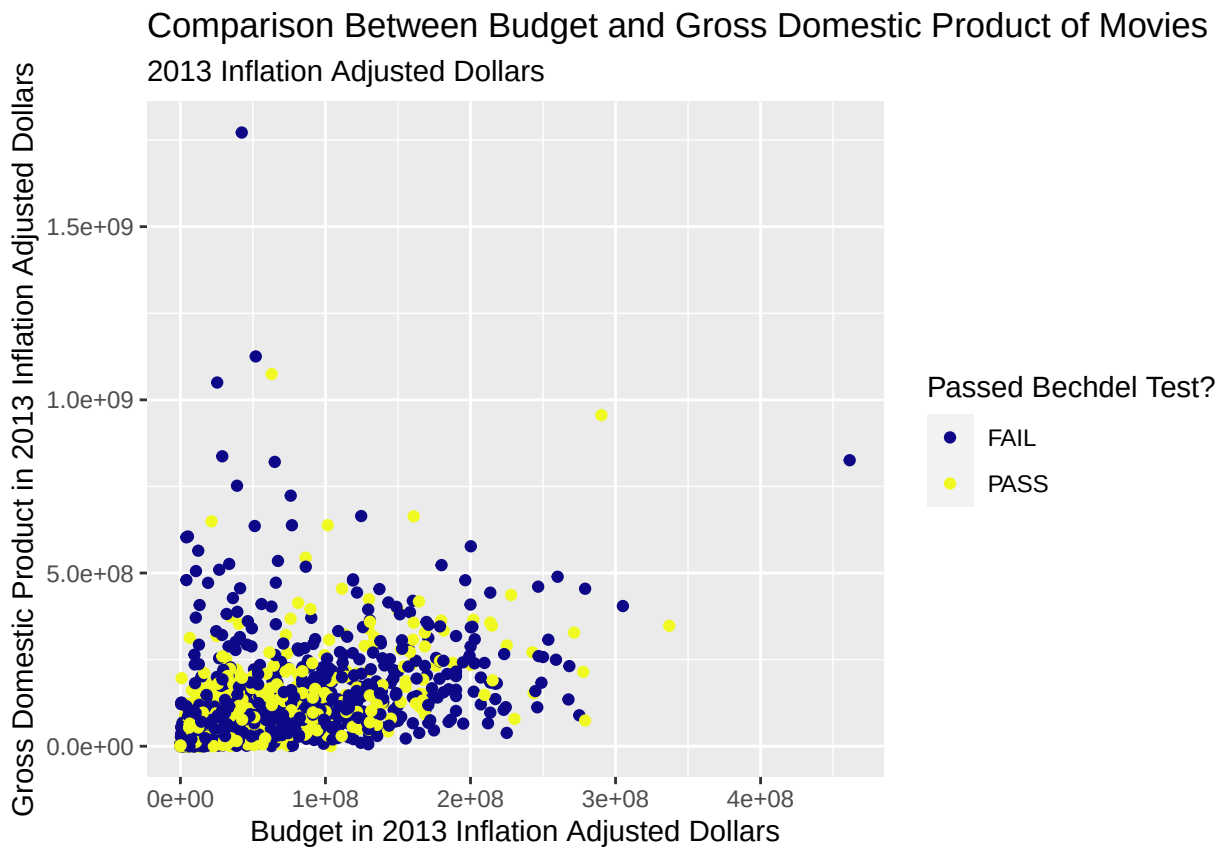
```
ggplot(bechdel, aes(domgross_2013))+
  geom_histogram(bins = 10)+
  labs(title="Movies Sorted by Domestic Gross Product",
  subtitle="2013 Inflation Adjusted Dollars",
  x="Domestic Gross Product in 2013 Inflation Adjusted Dollars",
  y="Number of Movies")+
  theme_bw()
```



The data do appear to have at least one major outlier, as at least one bin very far to the right in the histogram has a movie with a very large gross domestic product. Additionally, the data are clearly skewed right, which suggests that more of these data might be considered outliers (depending on the results of a statistical test).

Exercise 2

```
ggplot(bechdel, aes(budget_2013, domgross_2013, color=binary))+geom_point()+
  scale_color_viridis(discrete = TRUE, option = "C",
                      name = "Passed Bechdel Test?")+
  labs(title="Comparison Between Budget and Gross Domestic Product of Movies",
       subtitle="2013 Inflation Adjusted Dollars",
       x="Budget in 2013 Inflation Adjusted Dollars",
       y="Gross Domestic Product in 2013 Inflation Adjusted Dollars")
```



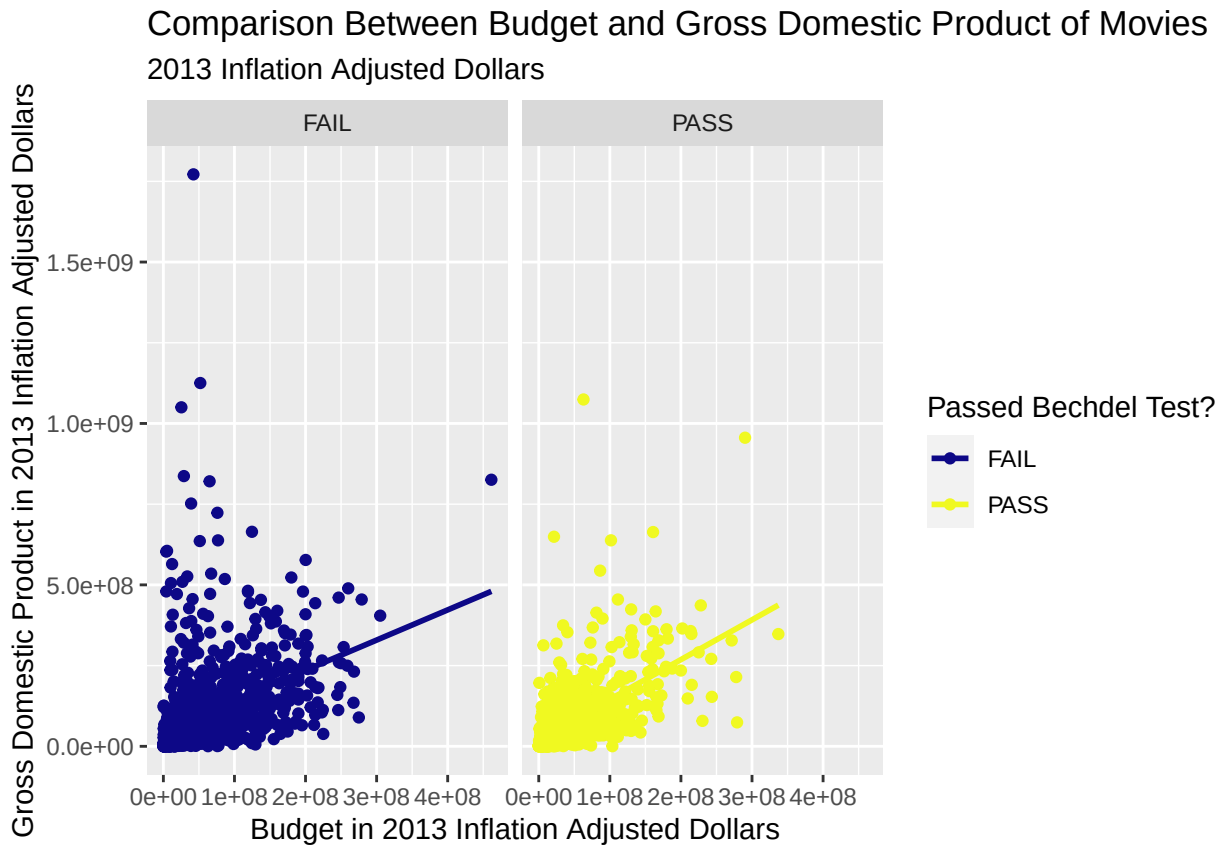
Exercise 3

By and large, these two overlain plots display the same relationship between budget and domestic product for movies that pass and fail the Bechdel test. However, there are a fair number of low-budget movies which made a large gross domestic product, almost all of which failed the Bechdel test. There also may be slightly more movies that failed the Bechdel test among movies which had a very large budget but failed to turn much gross domestic product. Without conducting statistical tests, however, both of these claims are somewhat spurious - there isn't too much difference in this relationship moderated by the Bechdel test.

Exercise 4

```
ggplot(bechdel, aes(budget_2013, domgross_2013, color=binary))+geom_point()+
  facet_grid(cols=vars(binary))+
  geom_smooth(method = lm, se=FALSE)+
  scale_color_viridis(discrete = TRUE, option = "C",
                      name = "Passed Bechdel Test?")+
  labs(title="Comparison Between Budget and Gross Domestic Product of Movies",
       subtitle="2013 Inflation Adjusted Dollars",
       x="Budget in 2013 Inflation Adjusted Dollars",
       y="Gross Domestic Product in 2013 Inflation Adjusted Dollars")
```

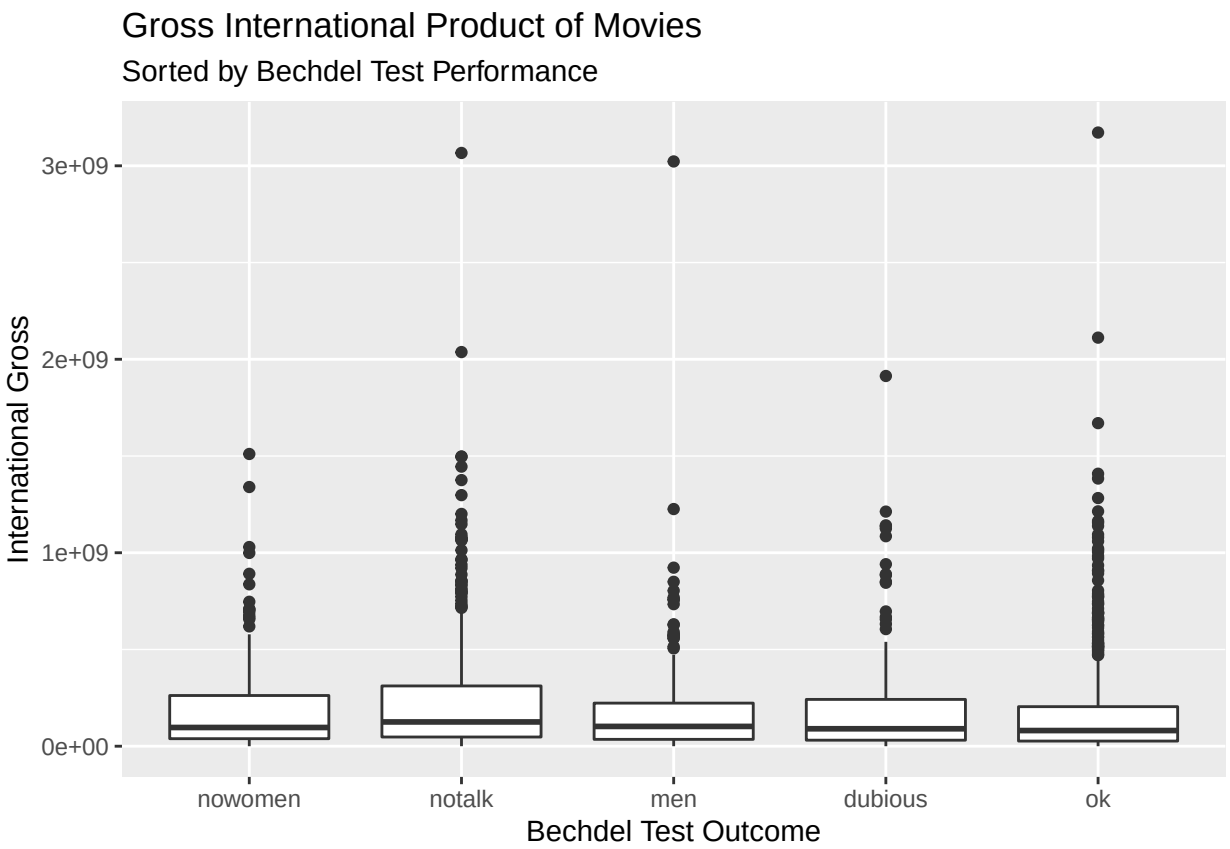
```
## `geom_smooth()` using formula 'y ~ x'
```



I definitely prefer this graph to the prior one - it's really hard to read data when that data is overlain on top of itself. This plot, with its separate regression lines, enables us to see more clearly that movies passing the Bechdel test in this sample have a slightly stronger relationship between budget and gross domestic product than movies that fail it do. The ease of reading and interpreting the data make it the better plot.

Exercise 5

```
ggplot(bechdel, aes(clean_test, intgross_2013))+geom_boxplot()+
  labs(title="Gross International Product of Movies",
       subtitle="Sorted by Bechdel Test Performance",
       x="Bechdel Test Outcome", y="International Gross")
```



Ultimately, this suggests that there is not very much difference in how much gross international product a movie makes depending on its Bechdel test performance. On average, movies do not make very much money - and this can be seen in this multiple box plot, where all charts have very low means and quartiles. In fact, the group of movies which seem to have performed the best are the ones in which there are female characters who don't talk - low on the Bechdel test scores here. Since the box plots list all outliers as discrete points, we can see that it is the categories in which women do not talk, talk but only about men, or pass the Bechdel test that have had a movie make at least 3 billion dollars in international gross - with one movie apiece for each category.

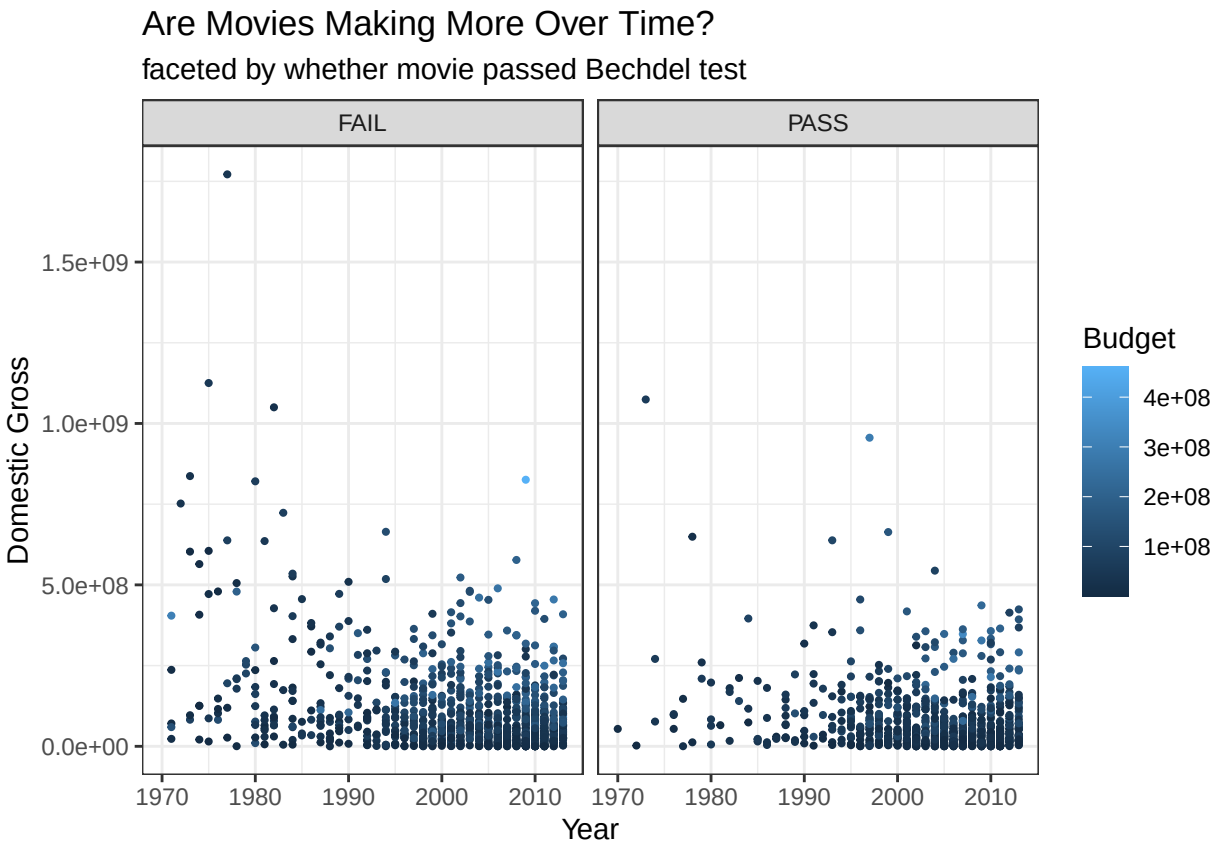
Exercise 6

```
ggplot(bechdel, aes(year, fill=binary))+geom_bar(position="fill")+
  labs(title="The Bechdel Test Over Time", x="Year", y="Prevalence")+
  scale_fill_discrete(name="Passed Bechdel Test?", labels=c("Fail","Pass"))
```

Over time, the percentage of movies in this sample passing the Bechdel test slowly increased. It declined in the last few years of the sample, such that a curve fitted atop the data would peak sometime between the mid-1990s and late 2000s. In most years, around 30-40% of movies in the sample passed the Bechdel test, although the peak years (with the exception of 1970) were just over 50%.

Exercise 7

```
ggplot(bechdel, aes(year, domgross_2013, color="budget_2013"))+
  geom_point(aes(color = budget_2013), size = .75)+
  facet_grid(cols=vars(binary))+
  scale_fill_distiller(palette="Blues")+
  labs(title="Are Movies Making More Over Time?",
       subtitle = "faceted by whether movie passed Bechdel test",
       x="Year", y="Domestic Gross", color="Budget")+
  theme_bw()
```


Are Movies Making More Over Time?
faceted by whether movie passed Bechdel test
FAIL
PASS
Domestic Gross
1.5e+09
1.0e+09
5.0e+08
0.0e+00
1970
1980
1990
2000
2010
Year
Budget
4e+08
3e+08
2e+08
1e+08