

Movie profitability and the Bechdel test

```
head(iris)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

Some folks in Hollywood claim that movies with female characters make less in the box office, FiveThirtyEight showed us that's not the case. The `fivethirtyeight` package includes the dataset behind this analysis, and many more.

Let's take a look at the top 20 highest grossing films published after 1990, 80% of the films pass the Bechdel test¹.

```
library("tidyverse")
library("fivethirtyeight")
bechdel %>%
  filter(year > 1990) %>%
  select(title, year, clean_test, binary, intgross_2013) %>%
  arrange(desc(intgross_2013)) %>%
  slice(1:20)
```

title	year	clean_test	binary	intgross_2013
Titanic	1997	ok	PASS	3171930973
Avatar	2009	men	FAIL	3022588801
Jurassic Park	1993	ok	PASS	1669905770
The Lion King	1994	notalk	FAIL	1497608078
The Lord of the Rings: The Return of the King	2003	notalk	FAIL	1445122625
Star Wars: Episode I - The Phantom Menace	1999	ok	PASS	1408313741
Harry Potter and the Deathly Hallows: Part 2	2011	notalk	FAIL	1375512553
Harry Potter and the Sorcerer's Stone	2001	ok	PASS	1282627914
Pirates of the Caribbean: Dead Man's Chest	2006	men	FAIL	1225855881
Independence Day	1996	ok	PASS	1213975115
Iron Man 3	2013	dubious	FAIL	1212692272
The Lord of the Rings: The Two Towers	2002	notalk	FAIL	1200694576
The Lord of the Rings: The Fellowship of the Ring	2001	notalk	FAIL	1167441808
Transformers: Dark of the Moon	2011	ok	PASS	1163903170
Shrek 2	2004	ok	PASS	1155511754
Finding Nemo	2003	notalk	FAIL	1147664217
The Lost World: Jurassic Park	1997	dubious	FAIL	1141669701
Harry Potter and the Chamber of Secrets	2002	ok	PASS	1138436153

¹Read more about the Bechdel Test on bechdeltest.com.

title	year	clean_test	binary	intgross_2013
Toy Story 3	2010	dubious	FAIL	1136341723
Skyfall	2012	dubious	FAIL	1124923772

When comparing the profitability of films, films that pass the Bechdel have a higher median profitability than any other category!

```

bechdel %>%
  # filter(year > 1990) %>%
  mutate(profitability = intgross_2013 / budget_2013) %>%
  group_by(clean_test) %>%
  summarise(median_profitability = median(profitability, na.rm = TRUE)) %>%
  arrange(desc(median_profitability))

```

clean_test	median_profitability
ok	2.701904
men	2.663718
notalk	2.646896
dubious	2.595882
nowomen	2.383619

Boxplots provide more information than the summary statistics in the table above are able to communicate.

```

bechdel %>%
  filter(year > 1990) %>%
  mutate(profitability = intgross_2013 / budget_2013) %>%
  filter(profitability < 50) %>%
  group_by(clean_test) %>%
  ggplot(aes(x = clean_test,
             y = profitability)) +
  geom_boxplot() +
  coord_flip() +
  theme_bw()

```

