

Pairs Plots

September 20, 2019

One of my favorite data sets contains a bunch of information about movies, including how the movie scores on the Bechdel test. A movie passes the Bechdel test if it satisfies 3 rules: 1. it has at least two women; 2. the women talk to each other; and 3. they talk to each other about something or someone other than a man.

Next week, we will build some models for a movie's international gross earnings in inflation-adjusted 2013 dollars (`intgross_2013`) based on the following 5 explanatory variables:

1. `budget_2013`
2. `run_time_min`
3. `imdb_rating`
4. `mpaa_rating`
5. `bechdel_test_binary`

```
head(movies)
```

```
## # A tibble: 6 x 14
##   year title bechdel_test bechdel_test_bi~ budget domgross intgross budget_2013 domgross_2013 intgross_2013 imdb_rating
##   <dbl> <chr> <ord>         <ord>         <dbl>   <dbl>   <dbl>   <dbl>         <dbl>   <dbl>         <dbl>
## 1  2013  21 &~ notalk      FAIL      1.30e7 25682380 4.22e7 13000000 25682380 42195766 5.9
## 2  2012 Dred~ ok        PASS      4.50e7 13414714 4.09e7 45658735 13611086 41467257 7.1
## 3  2013 12 Y~ notalk      FAIL      2.00e7 53107035 1.59e8 20000000 53107035 158607035 8.1
## 4  2013  2 Gu~ notalk      FAIL      6.10e7 75612460 1.32e8 61000000 75612460 132493015 6.7
## 5  2013 42   men        FAIL      4.00e7 95020213 9.50e7 40000000 95020213 95020213 7.5
## 6  2013 47 R~ men        FAIL      2.25e8 38362475 1.46e8 225000000 38362475 145803842 6.3
## # ... with 3 more variables: num_imdb_ratings <dbl>, mpaa_rating <ord>, run_time_min <dbl>
```

The first thing to do is always to make some plots. When we're thinking about multiple regression type problems with a reasonably small number of variables, the go-to plot is a pairs plot.

```
library(ggplot2) # general plotting functionality
```

```
library(GGally) # includes the ggpairs function, pairs plots via ggplot2
```

```
# I find it helpful to put the response variable last in this list, and to put quantitative variables next to each other.
vars_to_use <- c("mpaa_rating", "bechdel_test_binary", "budget_2013", "run_time_min", "imdb_rating", "intgross_2013")
ggpairs(movies %>% select(vars_to_use))
```

