Homework 4: Stats with R

Mike Lopez

September 2017

General instructions for homeworks:

- Make a new R Markdown file (.Rmd) referring to the assignment on the course Github page
- Change the heading to include your author name
- Save the R Markdown file (named as: [MikeID]-[Homework01].Rmd – e.g. "mlopez-Lab01.Rmd") to somewhere where you'll be able to access it later (zip drive, My Documents, Dropbox, etc)
- Your file should contain the code/commands to answer each question in its own code block, which will also produce plots that will be automatically embedded in the output file
- **Each answer must be supported by written statements (unless otherwise specified) as well as any code used**: In other words, if the answer is `24`, you should write "The answer is 24" (as opposed to just showing the code and output).
- Include the names of anyone you collaborated with at the top of the assignment
- I recommend copying the raw .Rmd code from the Github page as a start
- Homeworks are due at the start of class – please print the HTML and hand in.

The data set that we'll be working with for this assignment is found at:

```
library(fivethirtyeight)
library(tidyverse)
glimpse(bechdel)
?bechdel
```

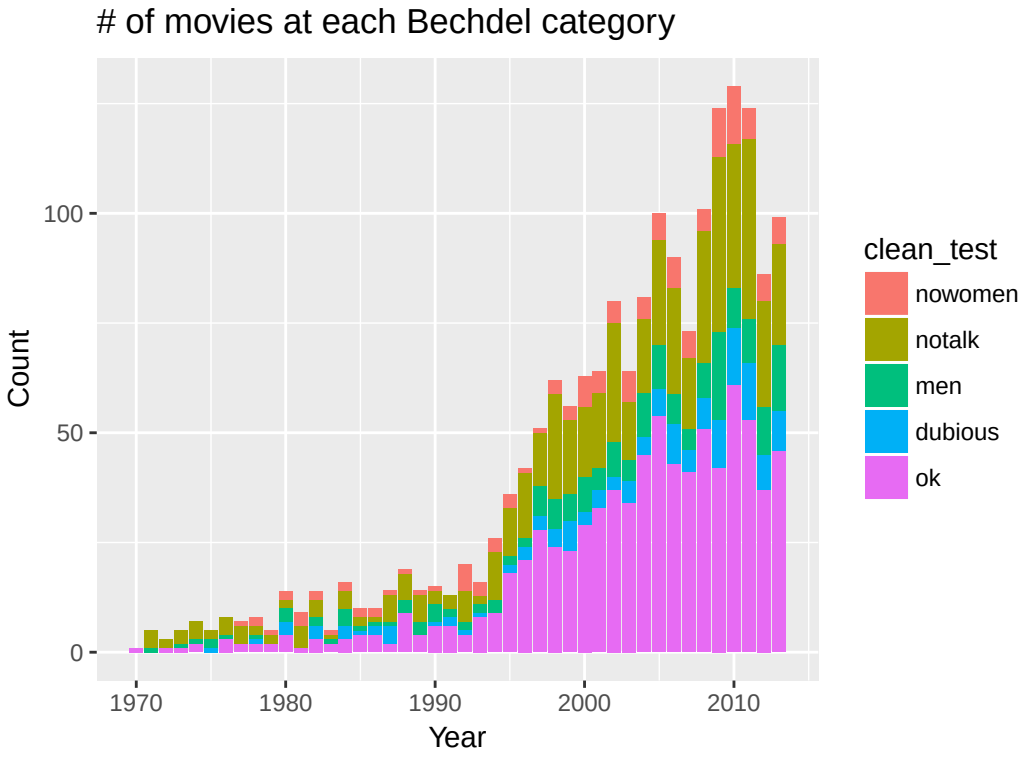
Refer to the help screen for information regarding each variable.

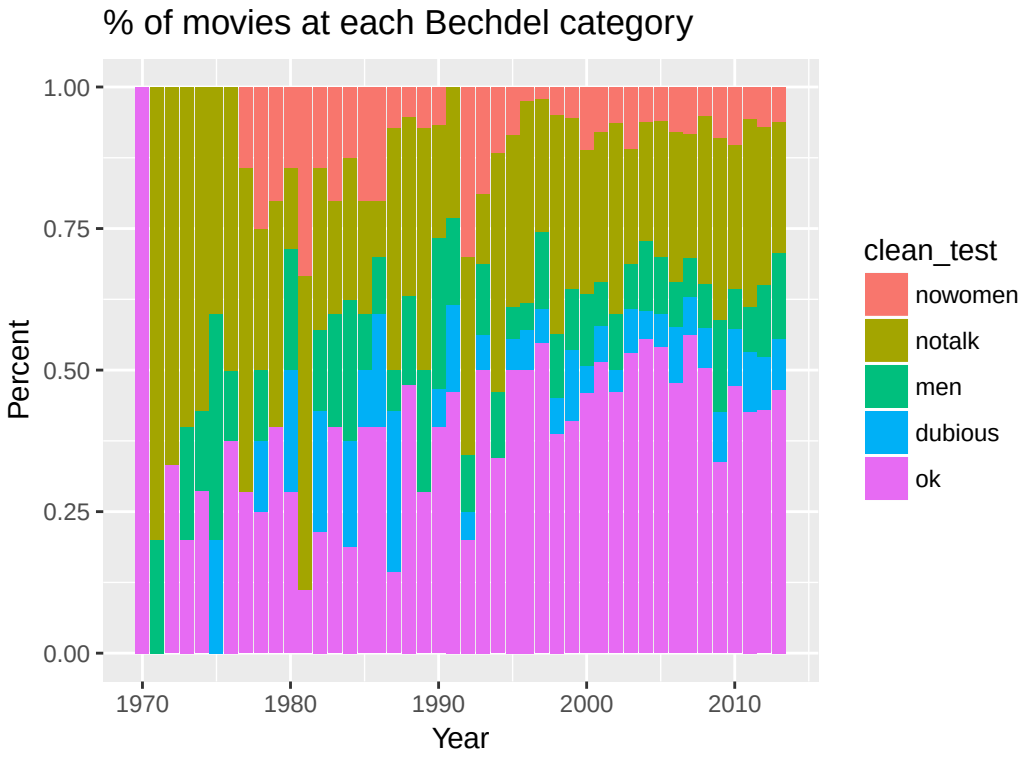
Part I: commands

1. Use R code to answer the following questions

- Identify the movies that passed the Bechdel test in 1980 (using the `binary` variable).
- Identify the movies that grossed more than 2 billion dollars (using 2013 adjusted dollars)
- Identify the percent of movies with `test` result `men-disagree`
- Identify the median domestic gross (using 2013 adjusted dollars) of movies made since 2000 that have passed the Bechdel test (using the `binary` variable).

2. Replicate the following graphs. Note that to obtain percentages on a bar chart, you can use `geom_bar(position = "fill")`.





3. Review the two charts above. Identify advantages and disadvantages to each visualiation.

4. Add a line of code to either plot above to change the legend title on the right hand side of the graph (which currently says `clean_test`). You may change to whatever label you think is more appropriate – and you will need to consult code outside of class to figure this out.

5. Make histograms of `intgross_2013` and `log(intgross_2013)`. What are the differences? What are

the benefits and disadvantages to working on a log scale?

6. Explore **at least one** of the following tools/packages using `ggplot()`. First, describe what the package is doing (or can do). Second make an interesting graph using the `bechdel` data. Points for creativity!

- ggthemes, https://cran.r-project.org/web/packages/ggthemes/vignettes/ggthemes.html
- ggrepel package, https://github.com/slowkow/ggrepel/blob/master/vignettes/ggrepel.md
- GGally package, http://ggobi.github.io/ggally/
- ggridges, https://cran.r-project.org/web/packages/ggridges/vignettes/gallery.html
- scale colour brewer, http://ggplot2.tidyverse.org/reference/scale_brewer.html