

# Manipulating Data and ggplot2

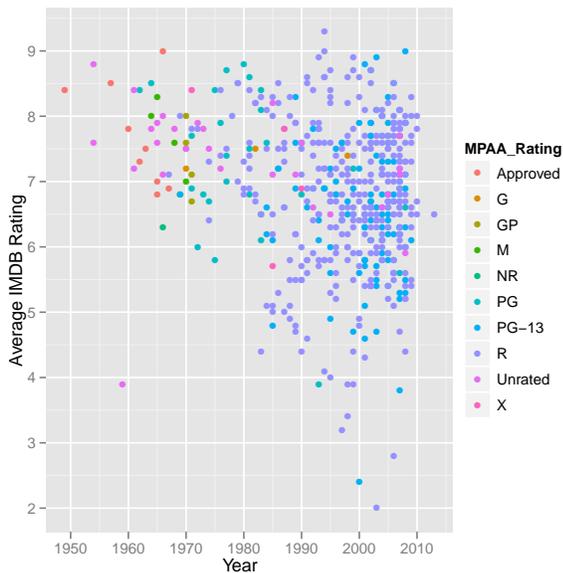
Jordan Boyd-Graber  
Digging into Data

February 10, 2014

After downloading these data, <http://goo.gl/0NQB3D>, create these plots using ggplot2. For each of the plots, I give hints on commands that might be useful in creating the plots. Use Google and R's built-in help liberally! Also try to build up these plots slowly, changing one thing at a time.

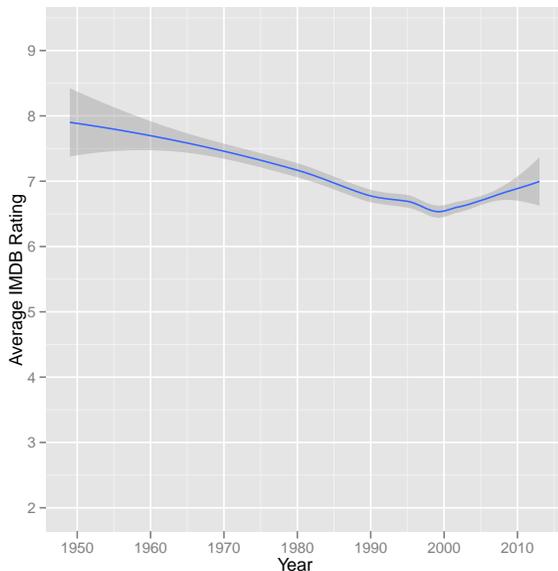
We won't be using Rattle much today, as its graphics aren't as pretty as ggplot2. So you'll be mostly interacting with R through the command line. This will also help you see how R works so you can see what Rattle does behind the scenes.

After you've read in the data (`read.csv`), you should be able to create this plot using:



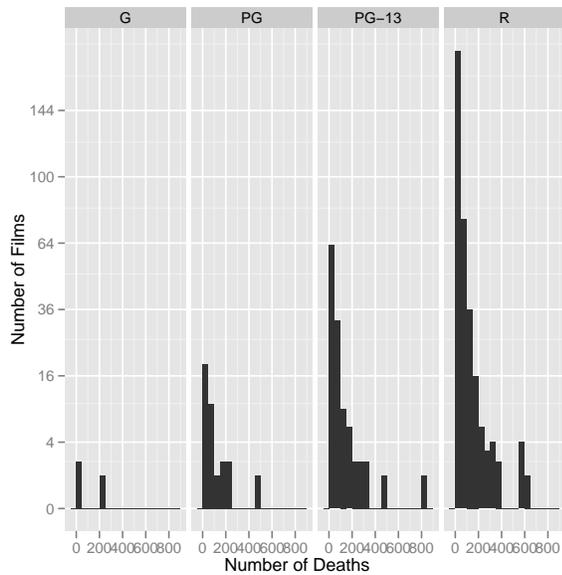
- `geom_point`
- `ylab`
- `ggplot`
- `aes`

This plot is very similar to the previous one, but you'll need:



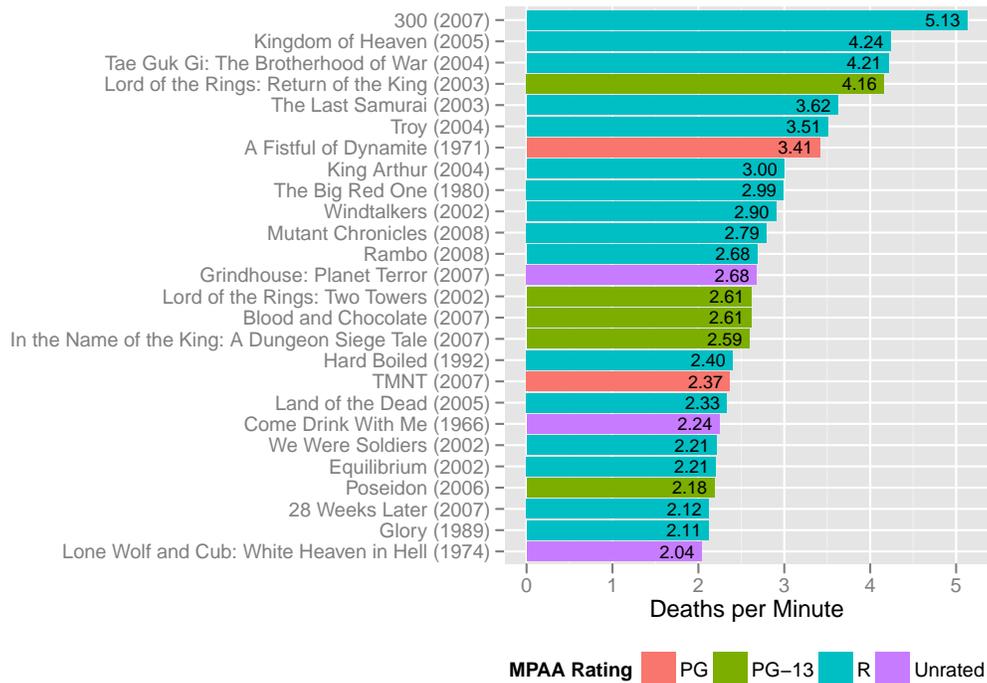
- `geom_smooth`

You'll want to use these commands:



- subset
- geom\_histogram
- facet\_grid
- scale\_y\_sqrt
- xlab

This plot is much more complicated. First, you'll need to create new columns to contain the number of deaths per minute and the extended film title. You can do that in R or in your favorite program. Then, try to create a bar chart with the **top** movies in terms of the deaths per minute. After that, flip the coordinates, order the movies, and add the text.



- geom\_text (with "label" and "hjust" arguments)
- geom\_bar
- fill
- coord\_flip
- labs (with the "fill" argument)
- opts (with the "legend.position" argument)
- sprintf
- reorder