

Week 2 Day 3

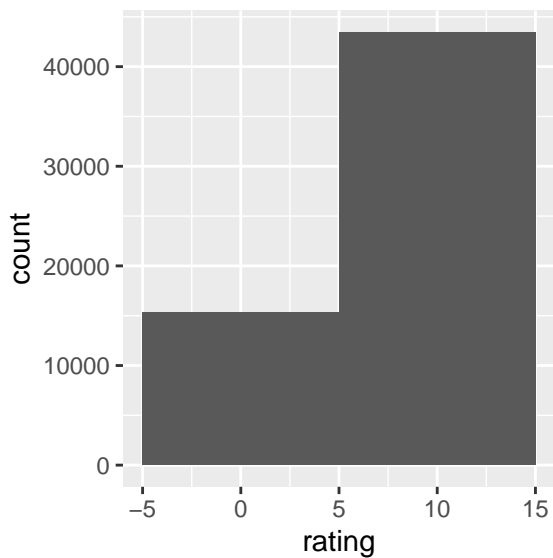
Stat140-04

Shan Shan

Part I: Shape of a histogram

The `movies` dataset consists of 58,788 movies that came out before 2013. The following code block plots the histogram of the `rating`. The current binwidth of the histogram is set to be 10.

```
library(ggplot2movies)
library(tidyverse)
ggplot(data = movies, mapping = aes(x = rating)) +
  geom_histogram(binwidth = 10)
```



1. Copy the following code to <https://rdr.io/snippets/>. Modify the code block above, and change the binwidth to an appropriate value. Note that the rating is ranged from 0 to 10. What do you expect the shape of the histogram of the `rating` would look like?

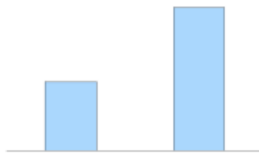
2. Describe the distribution using the histogram you created. Discuss whether it is unimodal or multi-modal, symmetric or skewed, and whether there are any outliers.

Part II: Reverse engineering

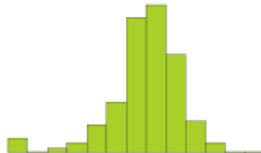
Match the following variables with the histograms and bar plots given below. These data represent students from students of Stat140 at Mount Holyoke. [Hint: Think about how each variable should behave.]

- (a) the number of hours students spent on homework lastnight
- (b) left/right handed breakdown of students
- (c) the time it took students to get on the MHC VPN
- (d) the number of hours sleep students got last night
- (e) whether or not students live off campus
- (f) the number of piercings students have

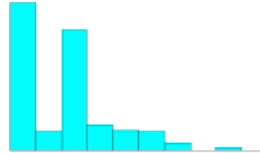
(1)



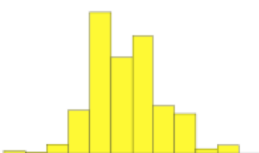
(3)



(5)



(2)



(4)



(6)

