Data Analysis

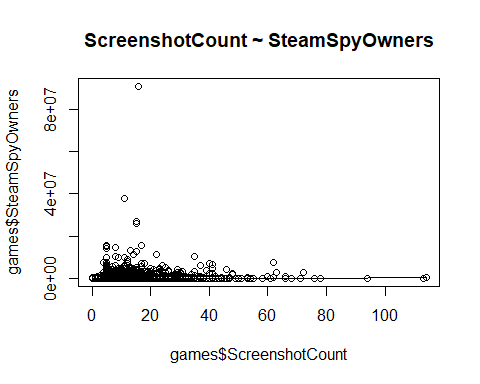
Tim Abramov

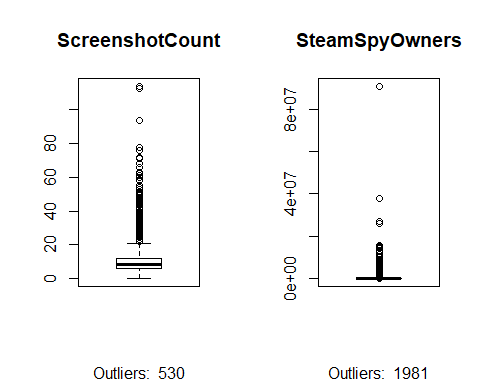
## Question 1

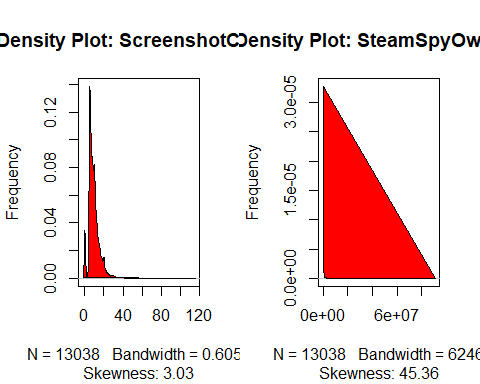
**“Are there any qualitative differences that can be measured between games that received a Metacritic score and those that didn’t?”**

## Question 2

**“Does the number of game screenshots have a measurable effect on the sales of the game?”**

To visualize the relationship, let’s graph a scatterplot first: 

Then, to see if we have any outliers, let’s graph boxplots for both the “ScreenshotCount” and “SteamSpyOwners”: 

Next, let’s look at distributions of our variables. Ideally our predictor variable(number of screenshots) should be close to being normally distributed: 

Finally, let’s …

## [1] 0.07601871

##   
## Call:  
## lm(formula = SteamSpyOwners ~ ScreenshotCount, data = games)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1558814 -165885 -119448 -62582 90431160   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 41296 17839 2.315 0.0206 \*   
## ScreenshotCount 13445 1545 8.705 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1137000 on 13036 degrees of freedom  
## Multiple R-squared: 0.005779, Adjusted R-squared: 0.005703   
## F-statistic: 75.77 on 1 and 13036 DF, p-value: < 2.2e-16

## Question 3

**“Is there an association between how many copies of the game were sold and the price of the game?”**