

Analyzing Video Game Sales

Sebastián Soriano Pérez

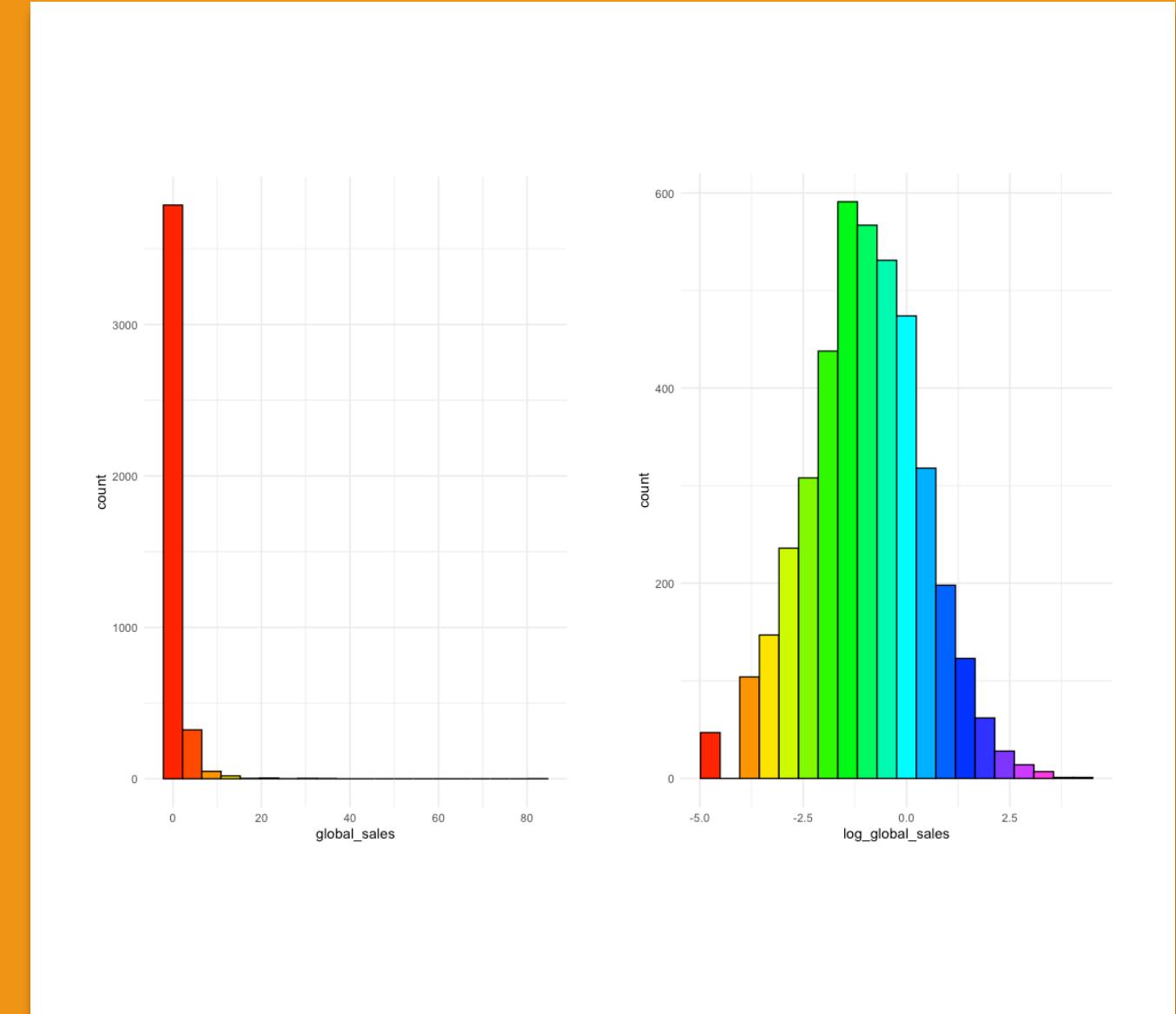


Introduction

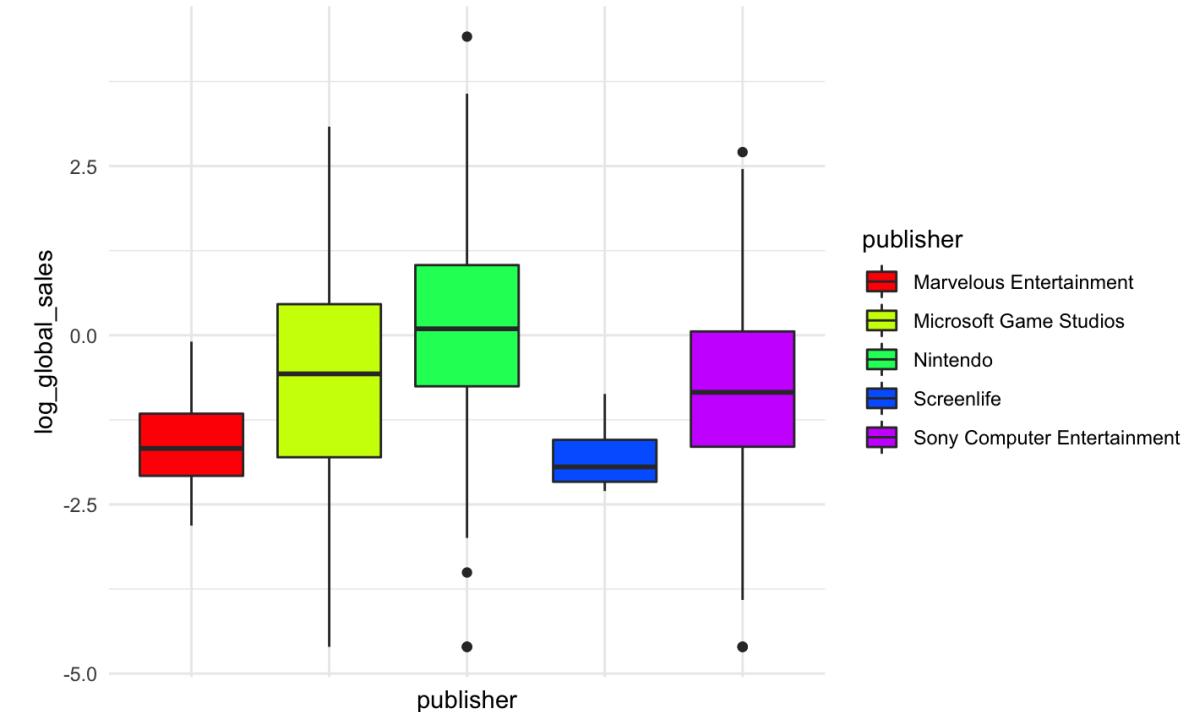
- Video game software sales from 2000 to 2016
- Original dataset had 14700 rows
 - Used 6951 rows with complete data
 - Missing data seemed to be at random
- Columns:
 - name
 - platform
 - year_of_release
 - genre
 - publisher
 - na_sales
 - eu_sales
 - jp_sales
 - other_sales
 - **global_sales**
 - critic_score
 - critic_count
 - user_score
 - user_count
 - developer
 - rating



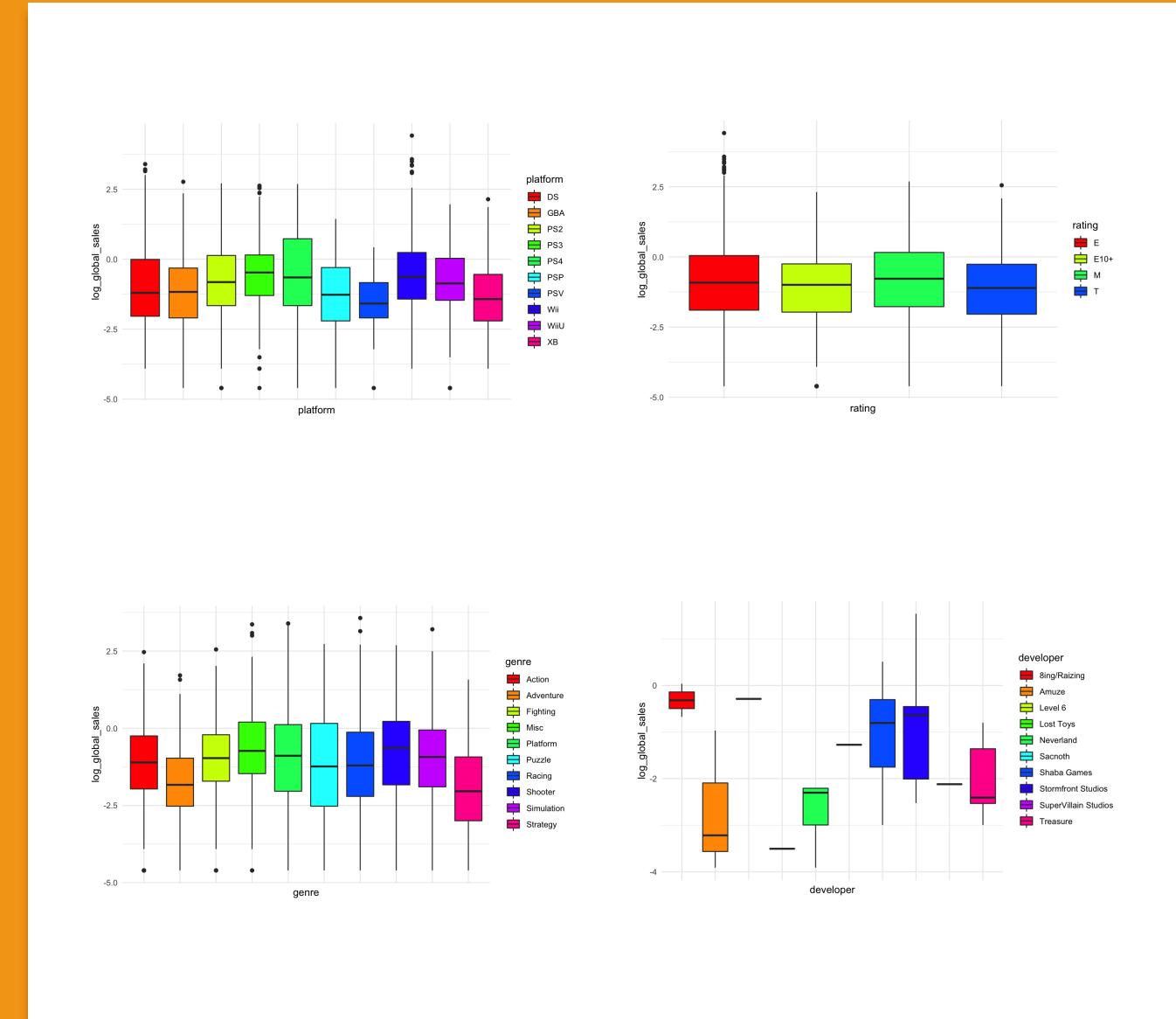
Exploratory Data Analysis



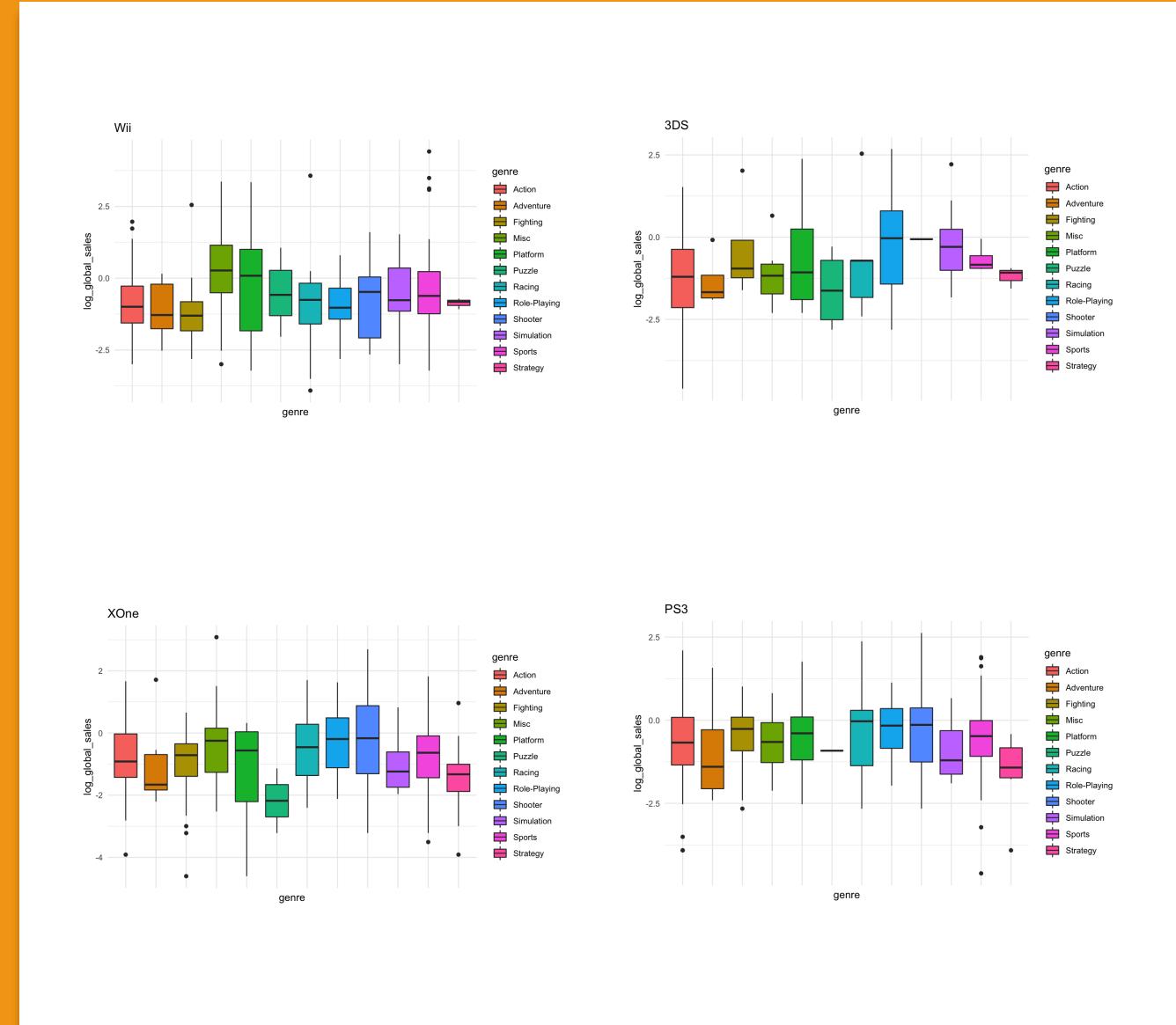
Exploratory Data Analysis



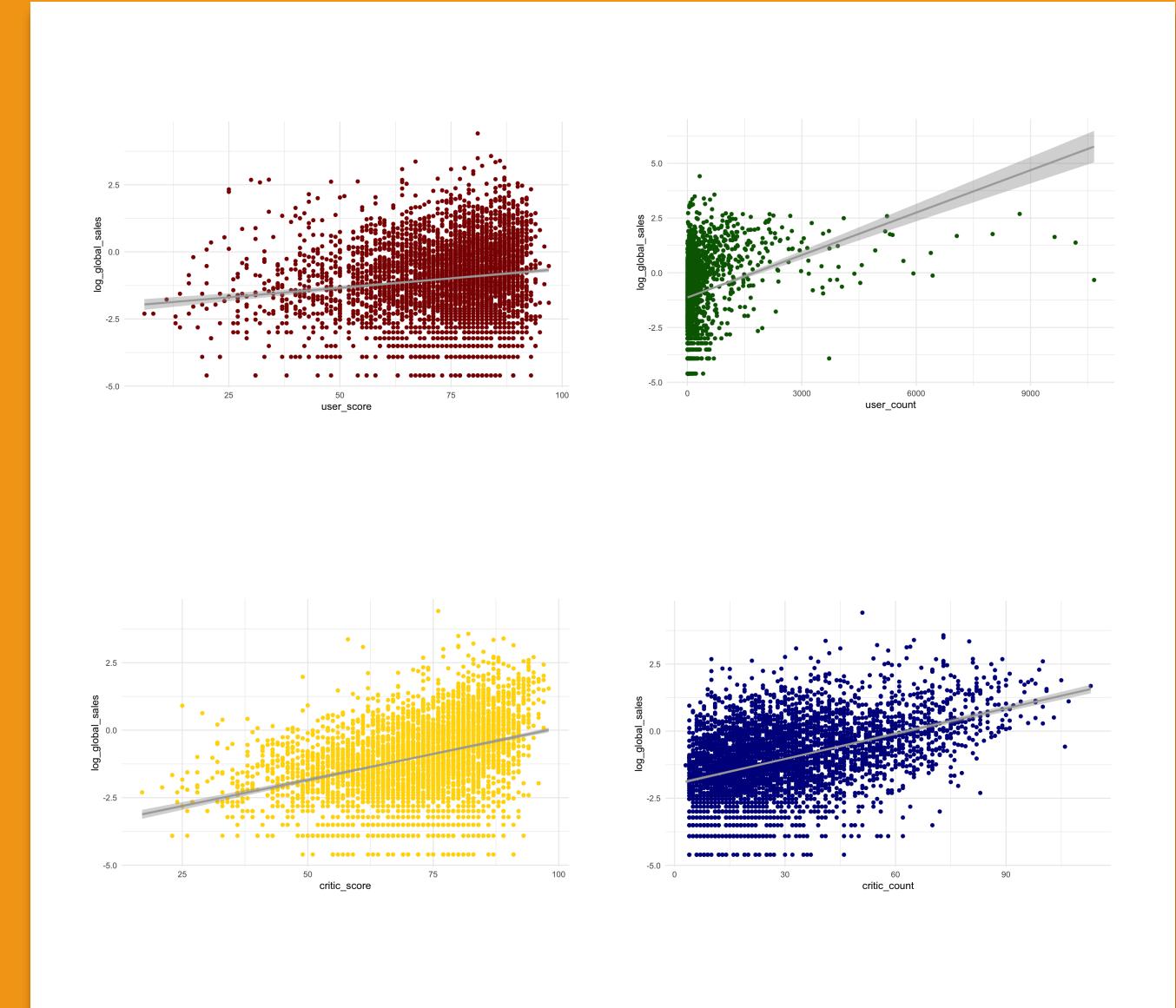
Exploratory Data Analysis



Exploratory Data Analysis



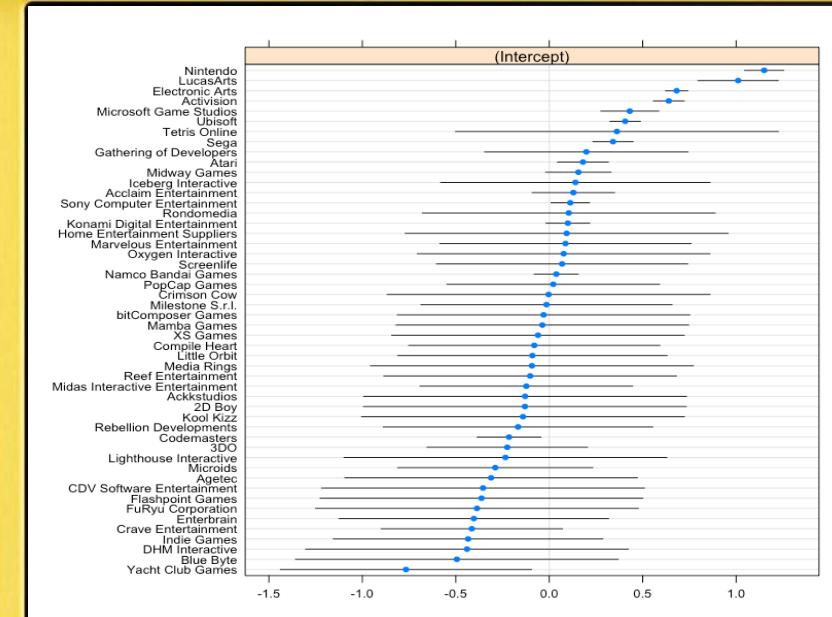
Exploratory Data Analysis



Model Selection

- Stepwise selection using AIC, BIC and conditional R-squared
- VIFs for fixed effects
 - All numerical v. < 5
 - Categorical v. (?)
- Proposed model:
 - Random intercept for publisher
 - Conditional R-squared: 0.57

```
log_global_sales ~  
platform + genre + critic_score  
+ critic_count + user_score  
+ user_count + rating  
+ platform:genre  
+ platform:rating + genre:rating  
+ (1 | publisher)
```



Results and Findings

- Nintendo, Lucas Arts, Activision, Microsoft, Ubisoft, Sega, Atari, Sony have an intercept greater than the mean
- Codemasters and Yatch Club Games have an intercept smaller than the mean
- A PlayStation, PlayStation 2, or Xbox 360 game would sell 590%, 260%, and 79% more units than a Nintendo 3DS game on average, respectively
- A puzzle game would sell 81% less than an action game
- A unit increase in `critic_score` would increase sales by 3.39%



Conclusions

- Try random intercepts by developer
- Try models with random slopes by publisher and developer
- In general, the model describes what variables are significant





Thank you!