

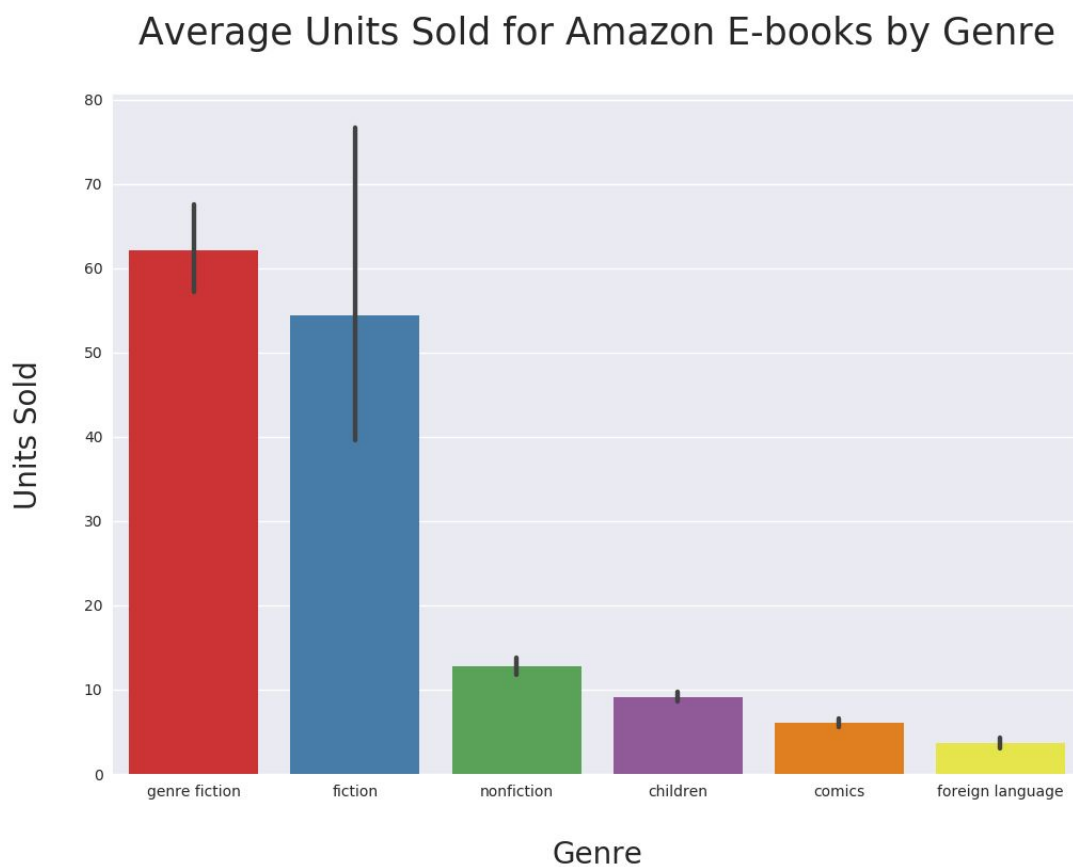
Assignment 2: Introduction to Data Visualization with Seaborn

Chris Dong

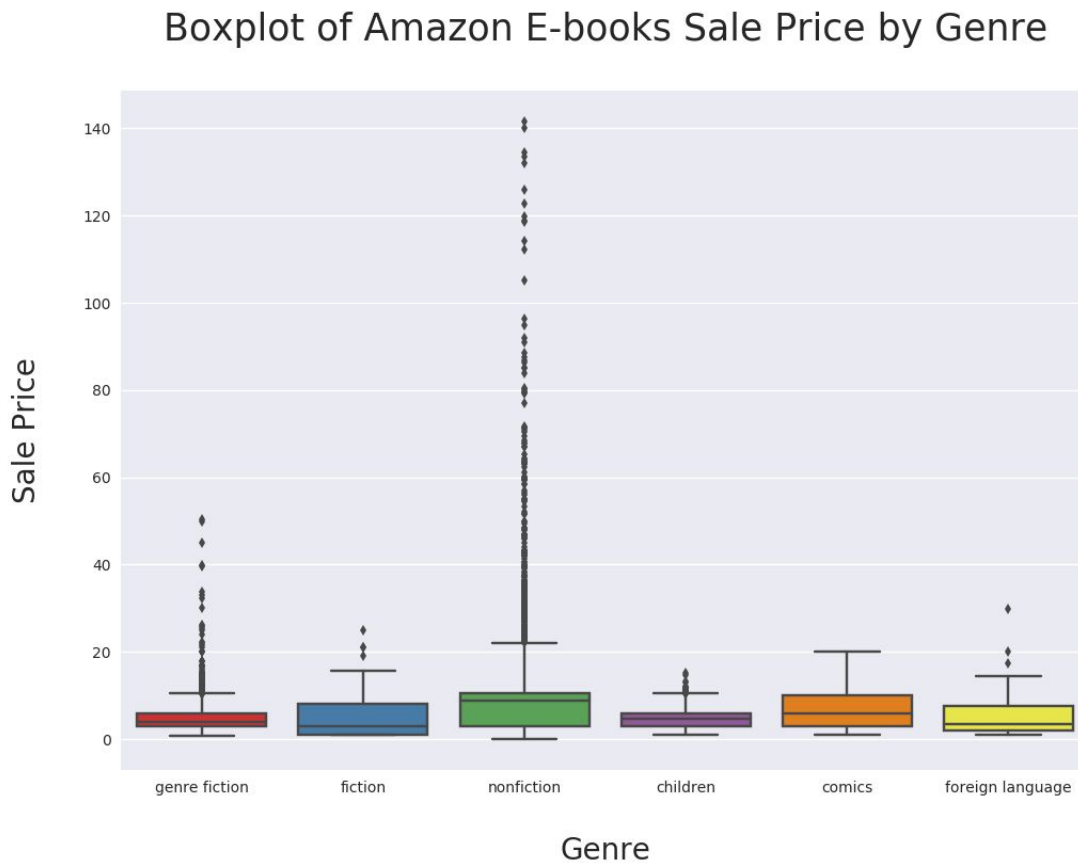
MSDS 622: DATA VISUALIZATION

Dataset: Publishers

BARPLOT



The majority of E-books sold on Amazon consist of the genres: *fiction* and *genre fiction*.

BOXPLOT

Nonfiction Amazon E-books is much more expensive than the other genres. Both *Nonfiction* and *genre fiction* seem to have high variance and many outliers (outside the whiskers of the boxplot).

BIVARIATE SCATTERPLOT

Authors seem to get the most revenue from the genre *genre fiction*. There is a lot of variance for *genre fiction* as well. *Comics* and *foreign language* generate the least revenue for authors.

TRIVARIATE SCATTERPLOT



Indie and *Big Five* publishers seem to dominate the market for E-books for most genres except *Comics*. Most publishers for *Comics* are small/medium. For *foreign language*, some single authors can be seen.

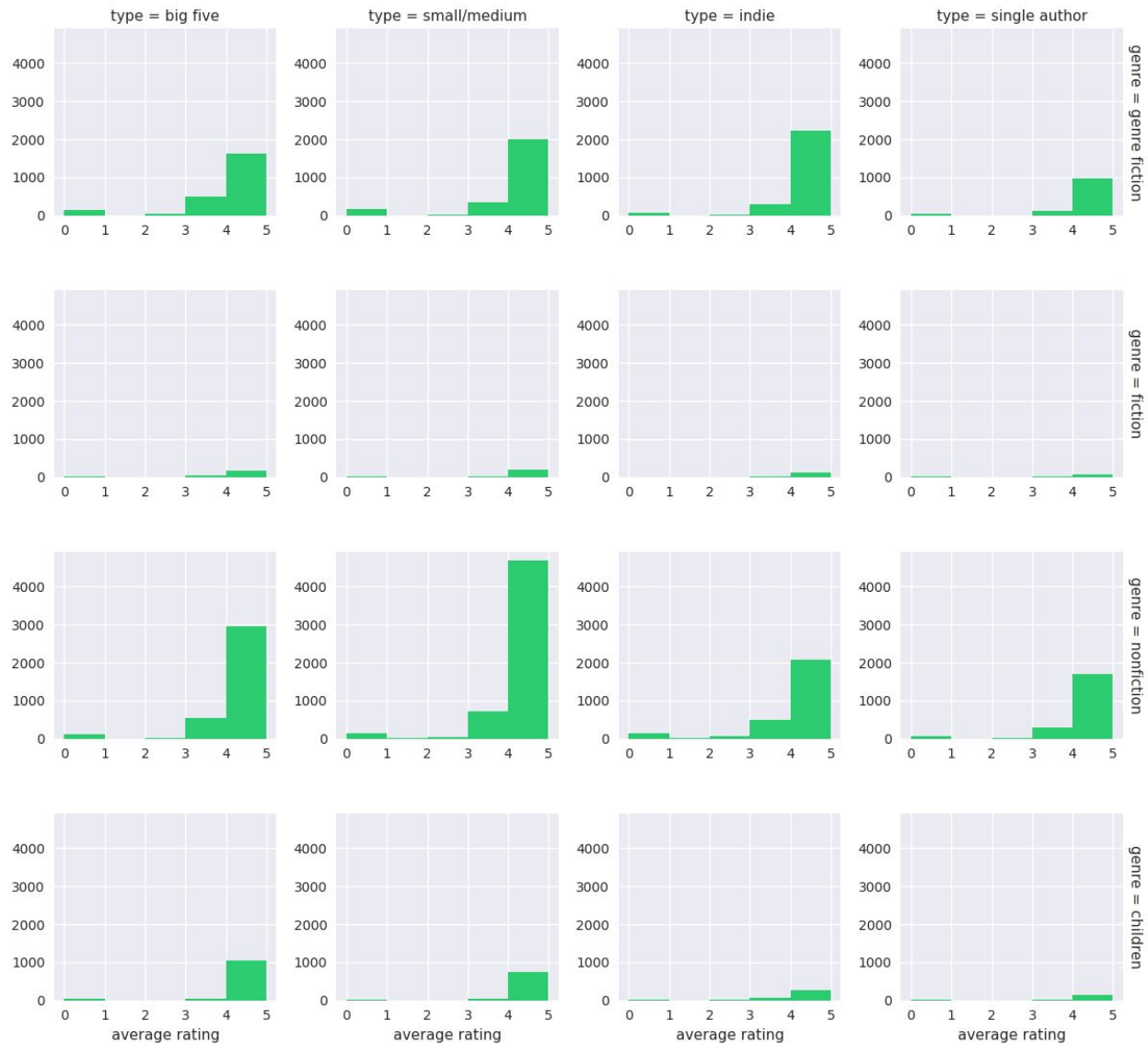
SWARM PLOT



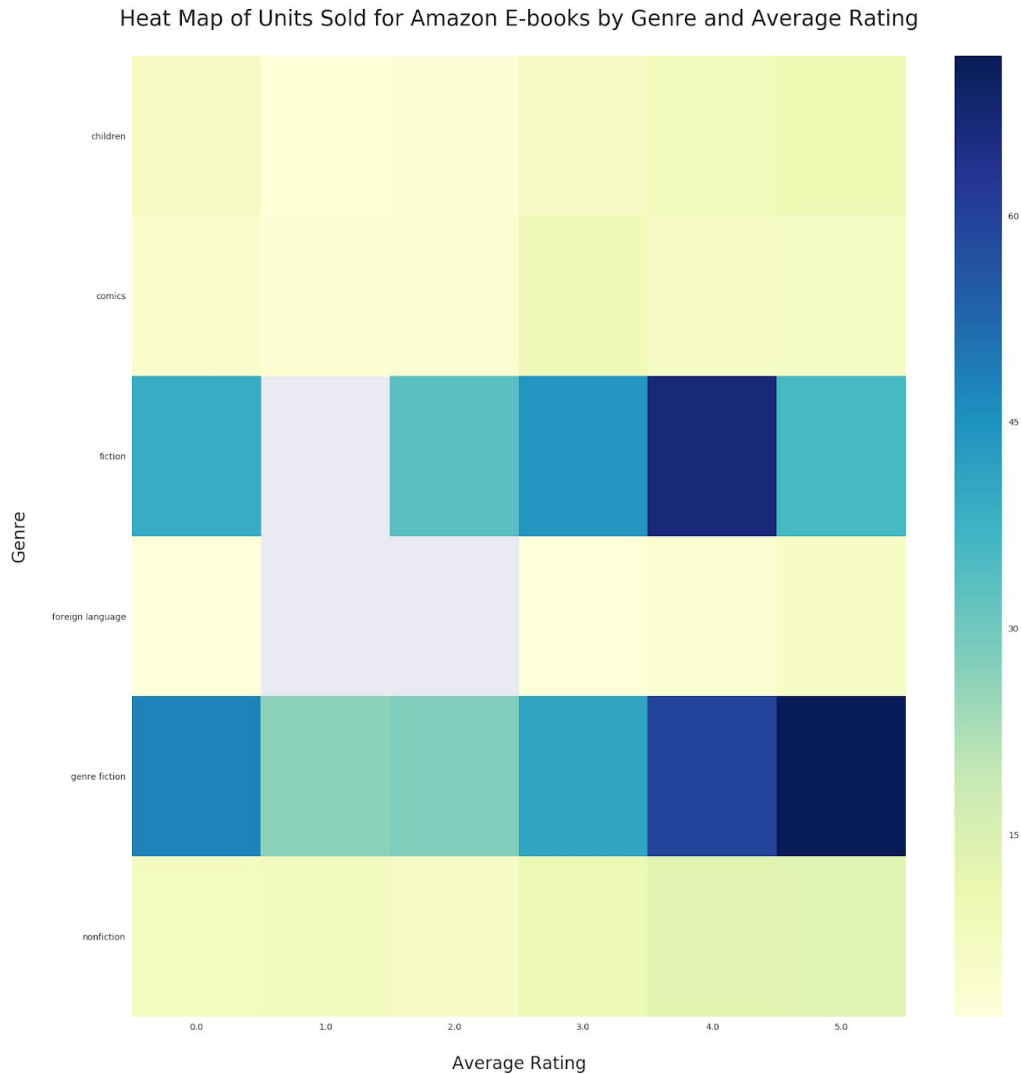
The publisher *Big Five* seems to dominate the market in terms of gross sales, much more than *single authors* and *amazon*.

FACETED HISTOGRAM

Histogram of Average Rating by Genre and Type



The most common publisher/genre combo consists of *Nonfiction* E-books published by small/medium publishers; they also provide the most satisfaction as seen by the high bars because the x-axis consists of Average Ratings. Some combinations are basically non-existent such as *single author/fiction*. Note that I filtered out the type *amazon* and genres *comics* and *foreign language* because their counts were even smaller in order to reduce clutter.

EXTRA: HEATMAP

The box for genre fiction and 5 star rating (note that the ratings are rounded to the nearest integer) has the most units sold, with second place being fiction /4stars. Combinations such as fiction/1star, foreign language/1star, foreign language/2stars are missing (NA).

EXTRA: VIOLIN PLOT

The sale price for E-books published by Amazon does not have much variance (seen by the thickness) and is usually around \$5. Small/medium has some outliers in terms of sale price.

EXTRA: REGRESSION PLOT

We can see the relationship between author and amazon revenue. In the middle are situations where Amazon and the Author makes the same amount of money per book. The points above the regression line display the books where amazon makes more than the author for the same book. Finally, points below the regression line, indicate when the author makes more than Amazon for a particular book.