



# Exploratory Analysis of Bookshop Data

Sri Lakshmi Tammineni

CSCE 5320 - 001

Summer 2021

# Summary of Findings

- Etaoin Shrdlu Press is largest publisher with 30 publications.
- Sci-Fi / Fantasy is the most popular genre and 2 x Mystery.
- More publications = More sales

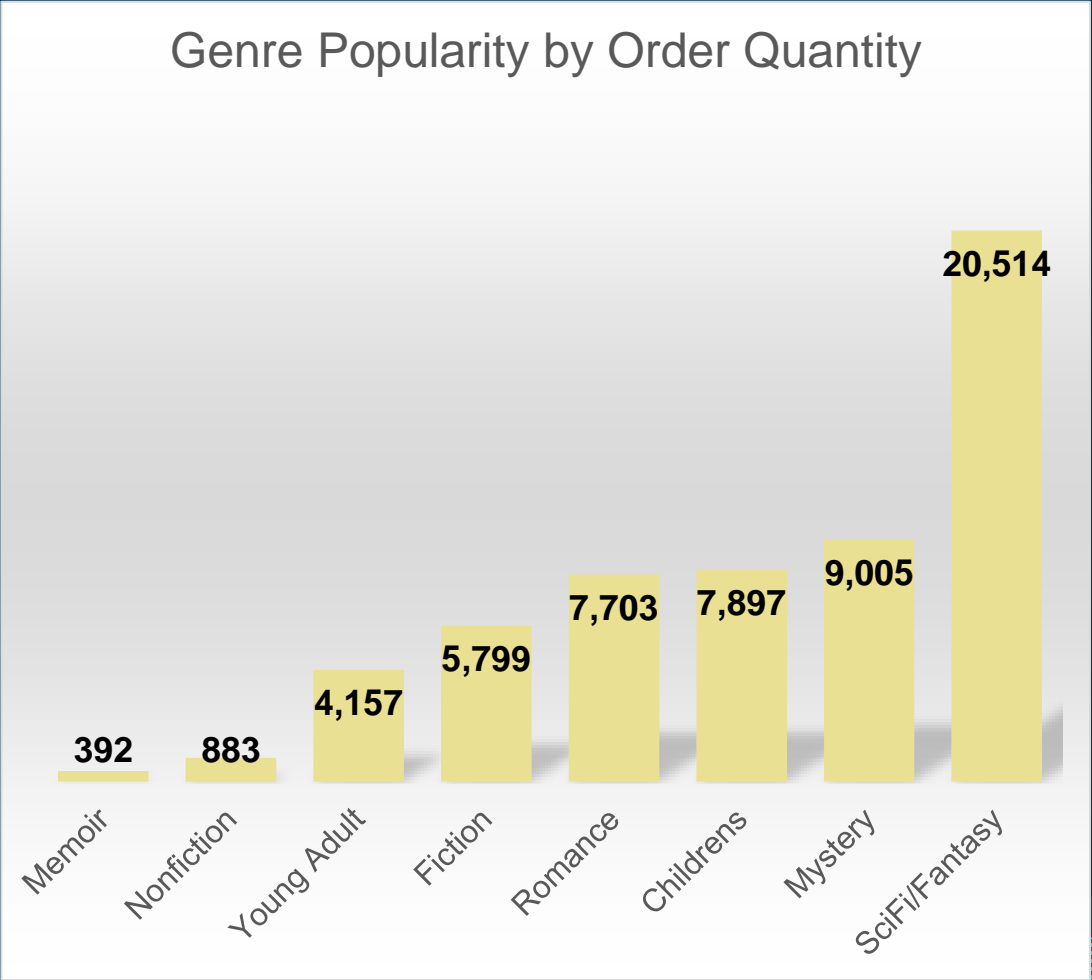
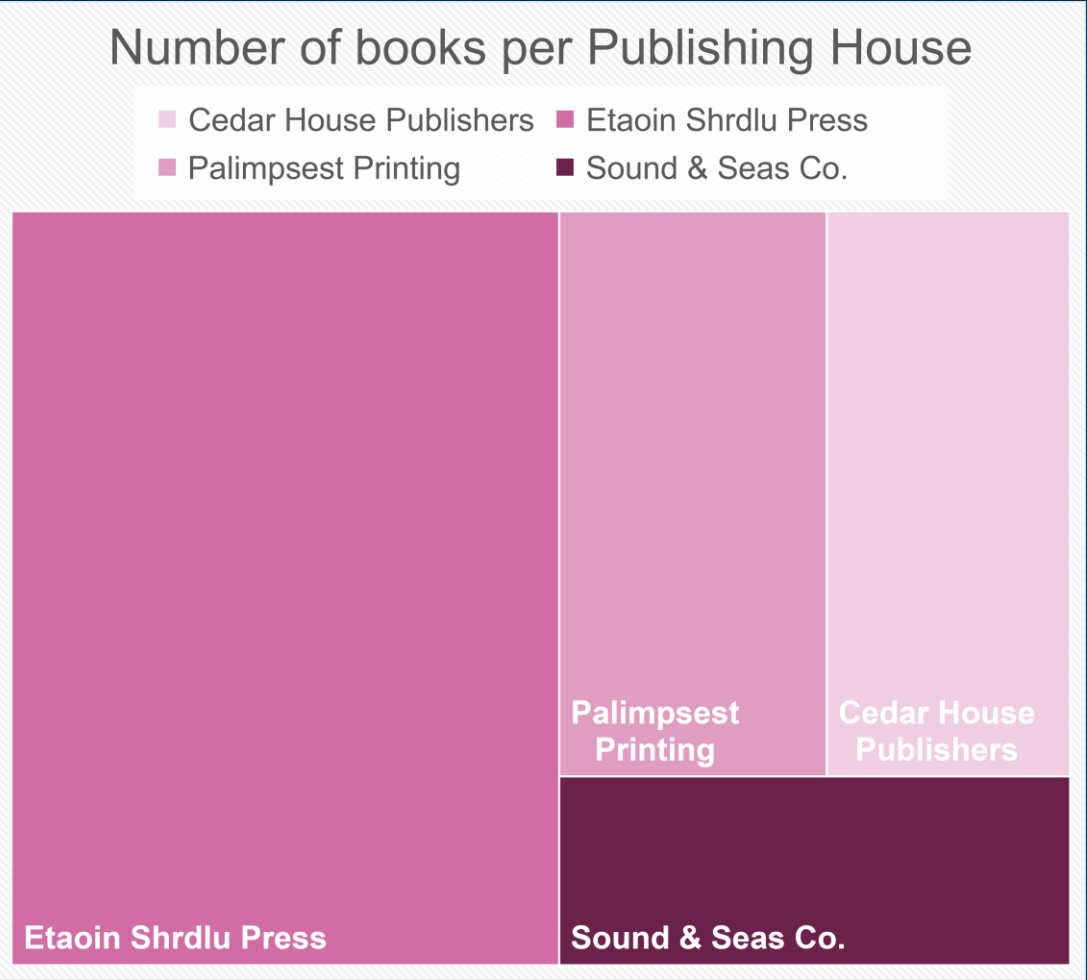
2185 and 2188 see highest number of sales and highest number of publications.

- Mass market > Hardcover > Trade Paperback in net Sales.
- Upward sales trend from spring to early fall. Jumpstart to sales during holiday season.
- The Mallemaroking, most sold book.
- Americans and Kiwis are most published authors.

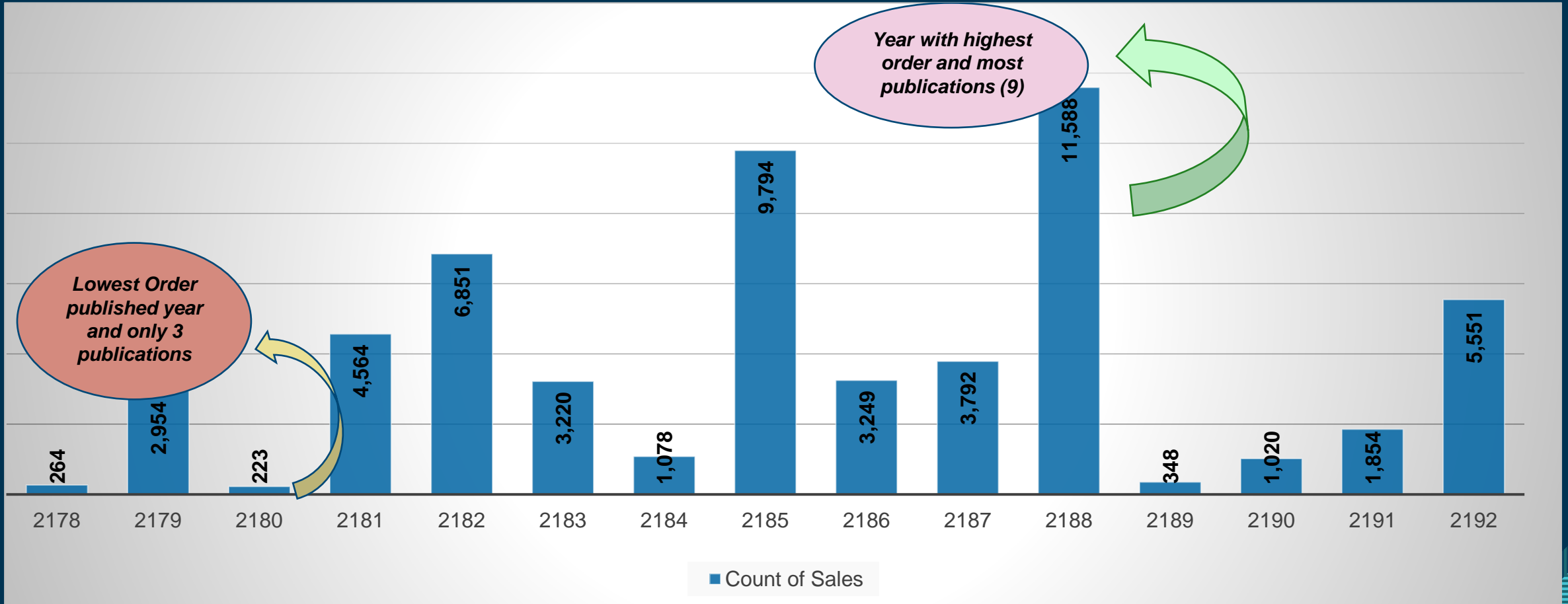
## 2e. What is causing disparity between Actual raw data vs what Tableau is Calculating?

- ❑ The disparity between SUM (Number of Checkouts) vs Raw excel data is due to how Tableau processed joins into one single table. **Joins are defined upfront, before analysis and into a single table resulting in missing unmatched values or duplicate aggregate values.** In our case, this resulted in duplication of rows from Checkout table. Example, there are several rows for the combination: BookID (Checkouts) = AD222 and Checkout Month Num = 10 when there is only one entry for this combination in raw data resulting in duplication of Number of checkouts = 25 several times. **This is because there are several sales for each Book and Edition, join creating a single table resulted in duplication of Checkouts table contents across multiple rows.**
- ❑ LOD or Level of detail expression computes the values requested at both data source level and visualization level and it also offers control on the level of granularity. In the case of calculating Monthly checkouts, we specify the granularity to be at Book ID and Checkout Month Num. An English translation of the LOD we used is:
- ❑ **For each [Book ID] and [Checkout Month Num] calculate the minimum [Number of Checkouts]. This is result in elimination of these duplicate rows into the calculation.**

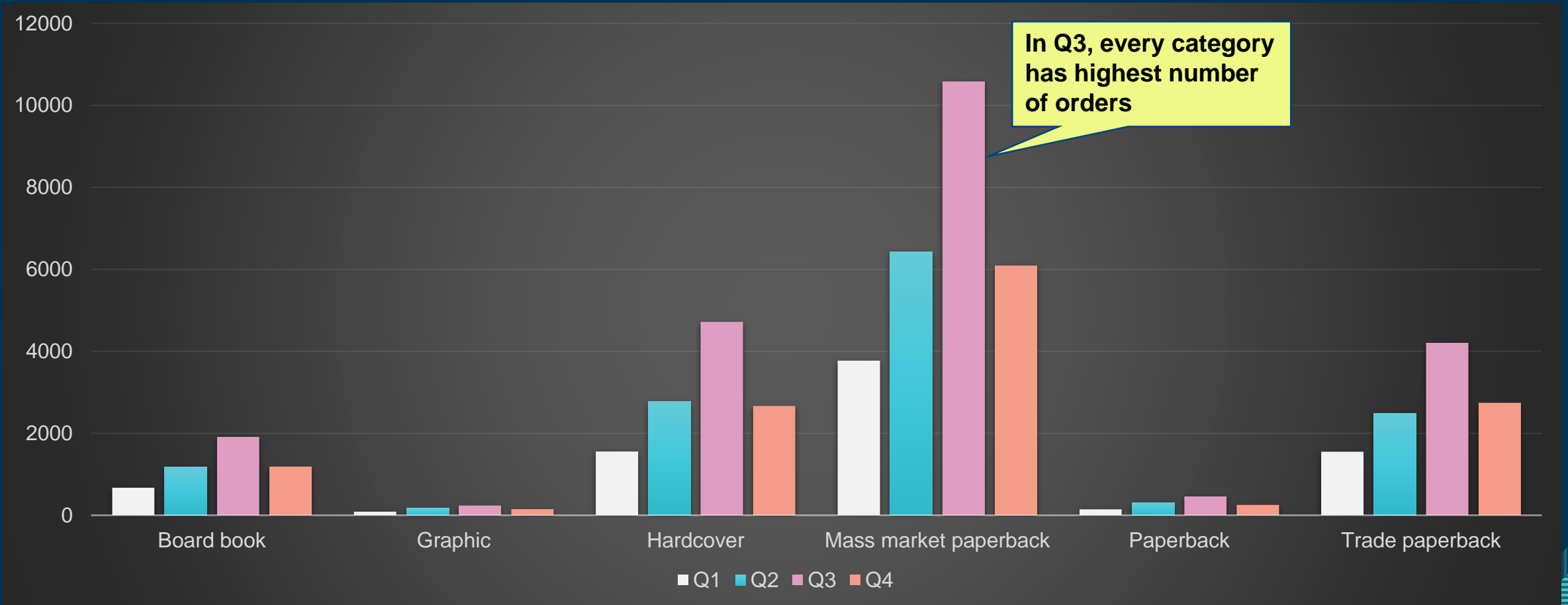
# Share of Books per Publisher & Popularity of each Genre



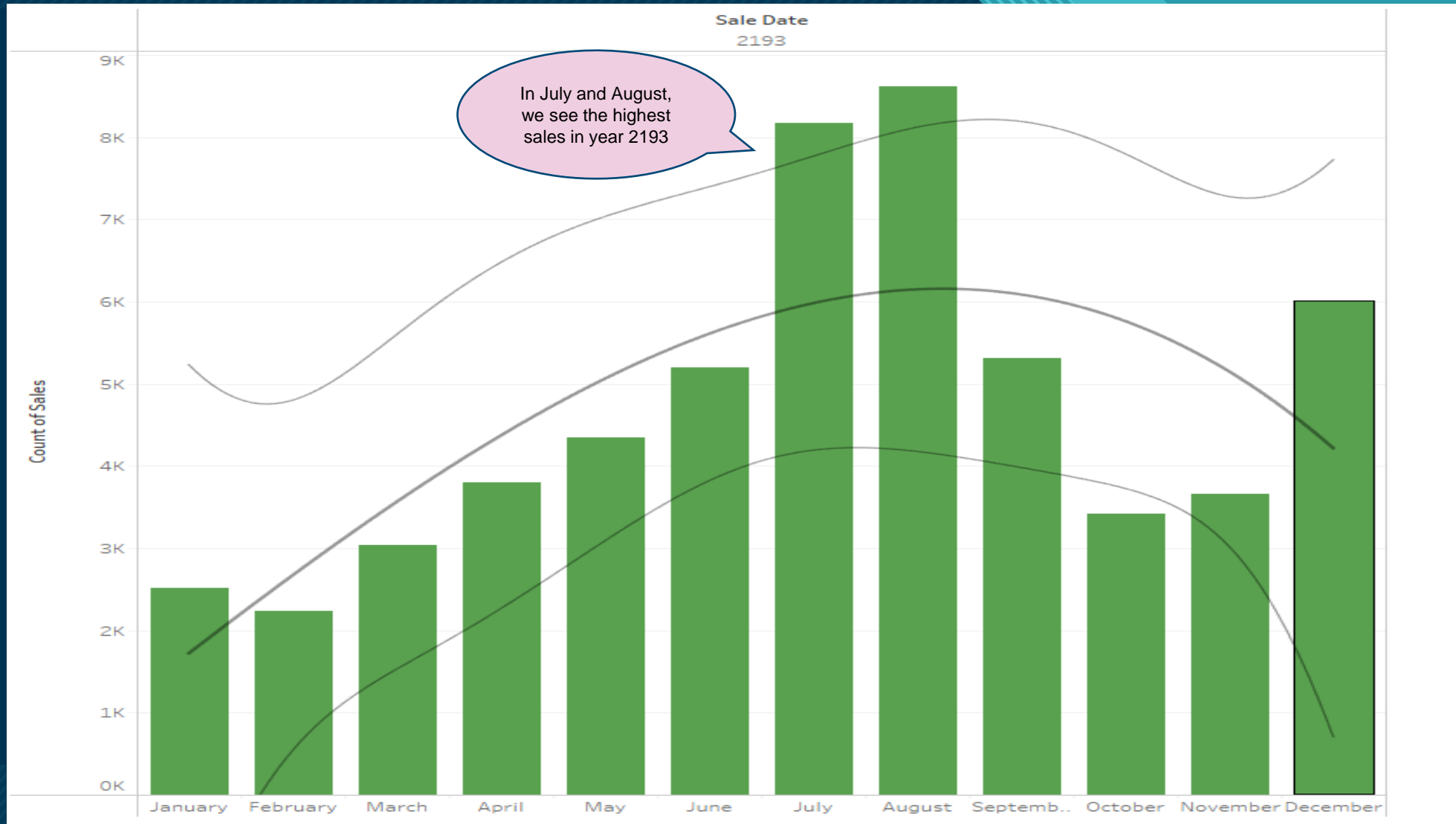
# Number of Orders per Publication Year



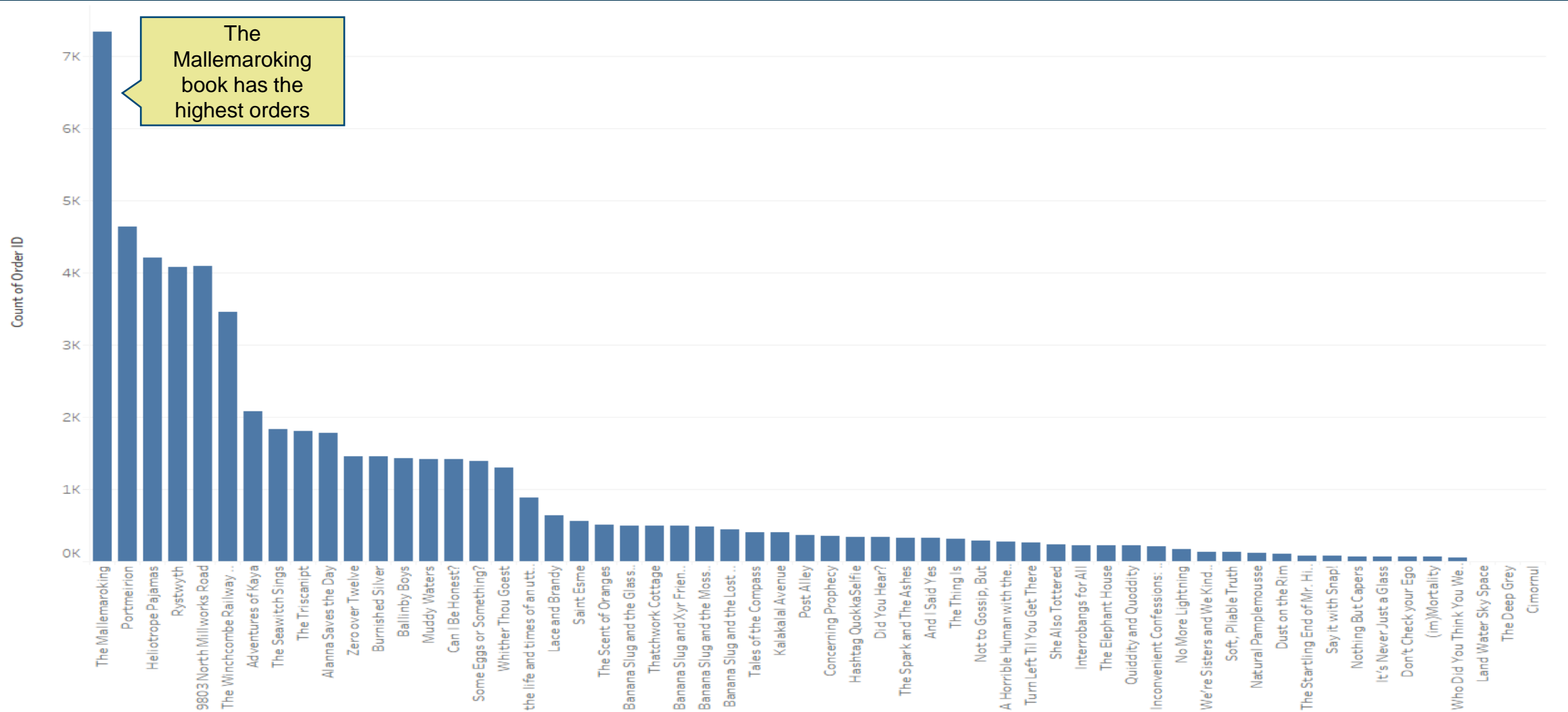
# Number of Order per Book Format



# Sales Pattern for the Year

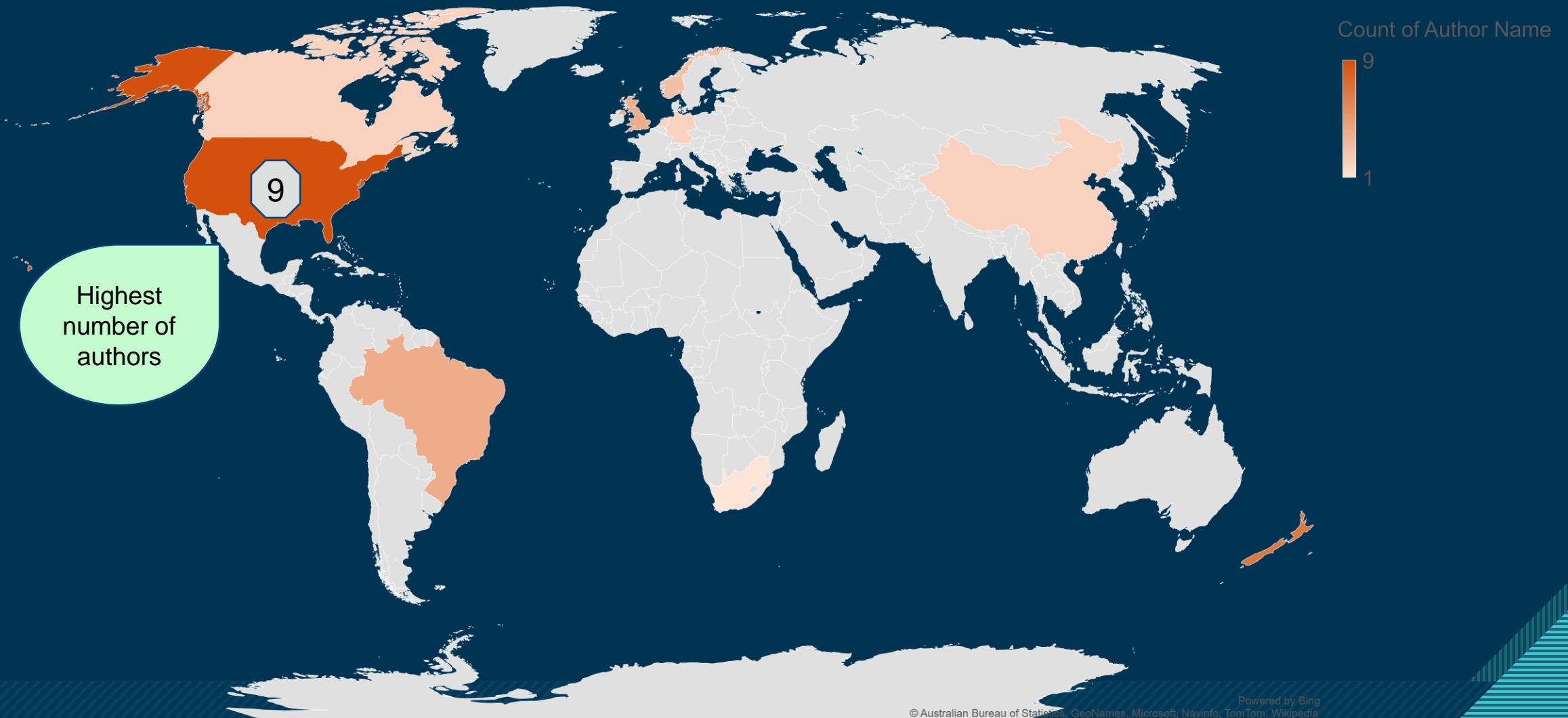


# Total Orders per Book





# Authors Geographic Distribution





**Thank You!**