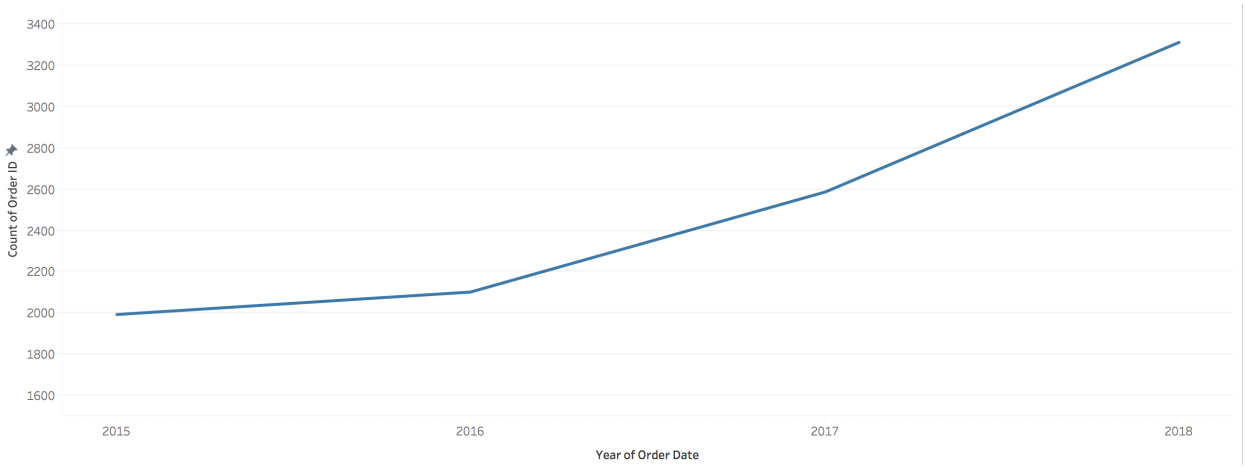
Data analysis of superstore sales data

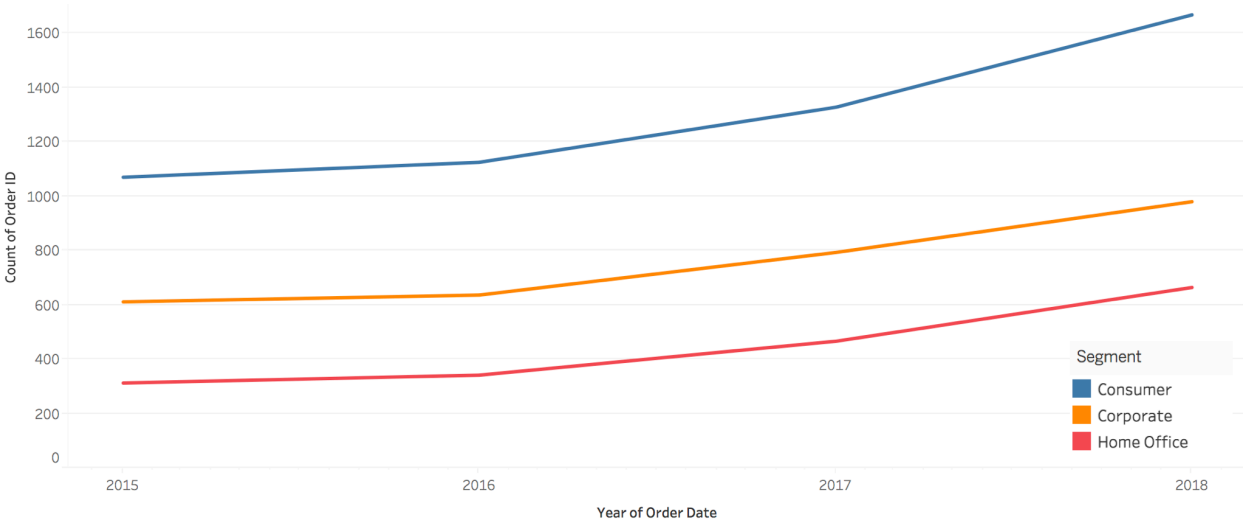
Introduction:

This is the superstore sales data which is an example dataset offered by Tableau. I choose this dataset because the dataset has various kind of data based on which many kinds of plot could be made. And I'm looking for data analyst job, so I'd like to use this example to make some basic training of Bussiness Analysis. The question I want to explore is very general: what is the trend of sales over time? Which category contributes more to the sales/profit? Which state contributes more to the sales/profit?

Firstly, let's have a high-level big picture of the sales data: we can see from the below plot that the number of orders continues rising up from 2015, and the increasing rate also increases with time.



if we split the total number of count into three segments, we can see more details: the consumer segment share the largest proportion of orders and home office segment is the smallest part.

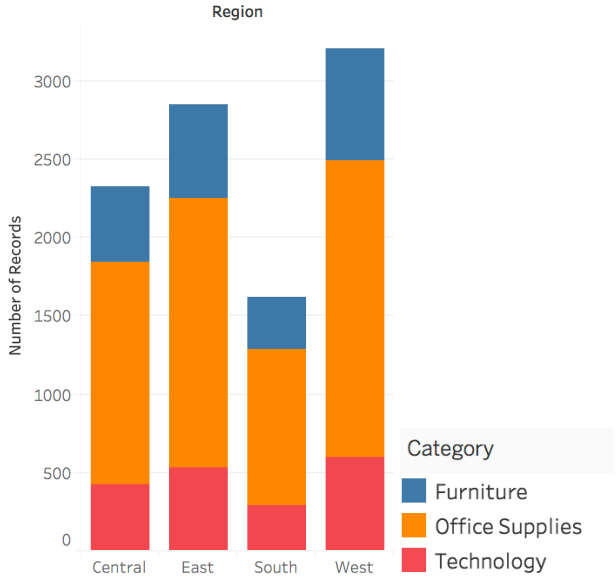


Besides the number of orders, if we go deeper with the profit, the box plot can tell us that the technology category has the highest profit, in terms of absolute value and variance. The furniture category, on the other end, has the lowest profit with nearly 40% of orders having negative profit.

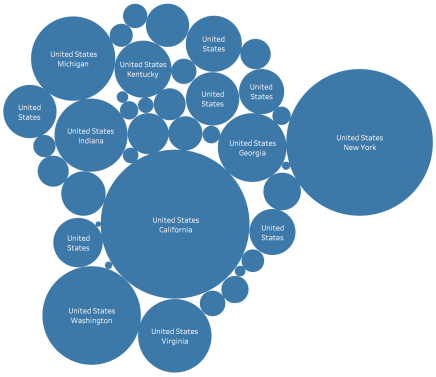
Then, if we compare the change of sales of these three categories over time (year), only furniture category kept increase over the four years, there's a clear drop in sales for other two categories in the year 2016.

	Order Date			
Category	2015	2016	2017	2018
Furniture	$157,193	$170,518	$198,901	$215,387
Office Supplies	$151,776	$137,233	$183,940	$246,097
Technology	$175,278	$162,781	$226,364	$271,731

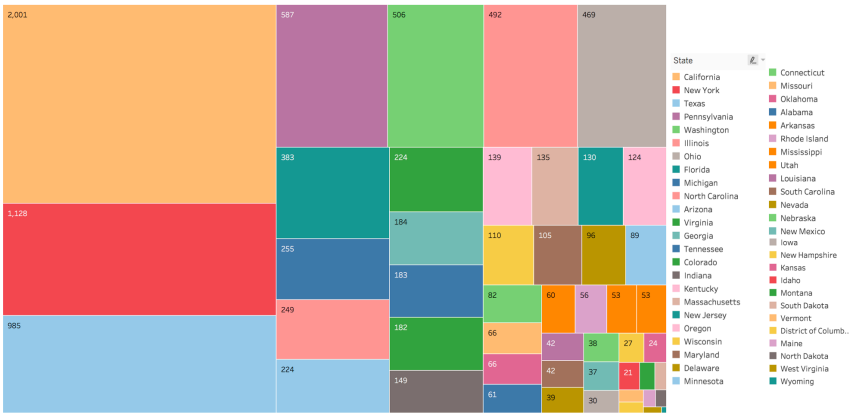
Next, the below bar plot shows the contribution of three categories on the total number of orders of each region. The west region has the largest number of orders while the south region has the smallest. In all four regions, the office supplies category always making the largest contribution to total orders.



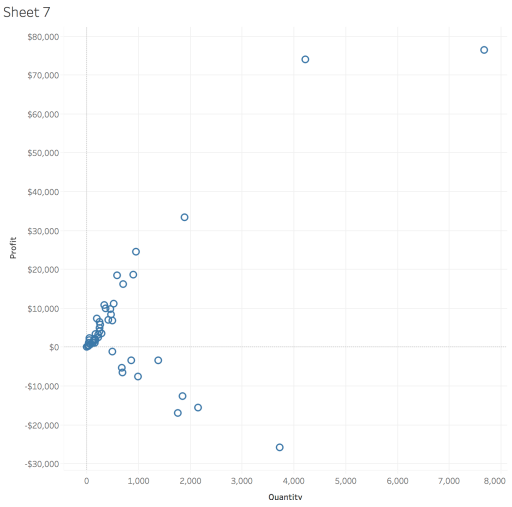
More specifically, if we want to see the individual contribution of each state, we can see from the below bubble plot that California, New York and Washington three states make the most contribution to the total number of orders.

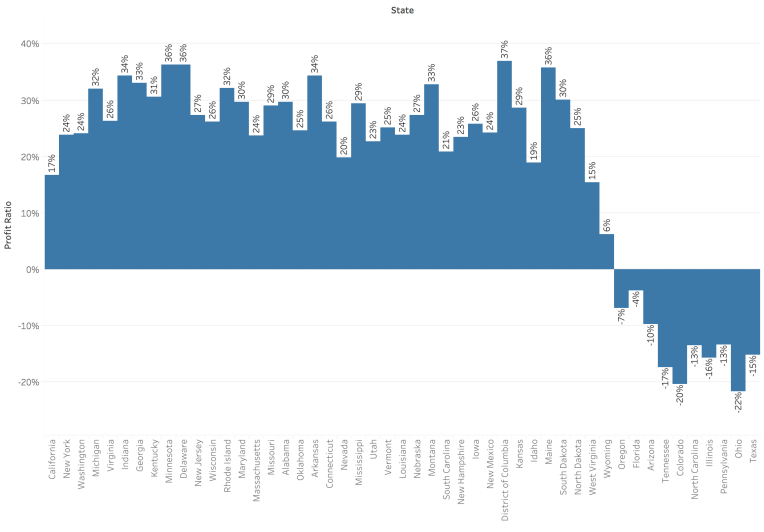


The below tree-mapping makes it easier to make direct comparison between states:

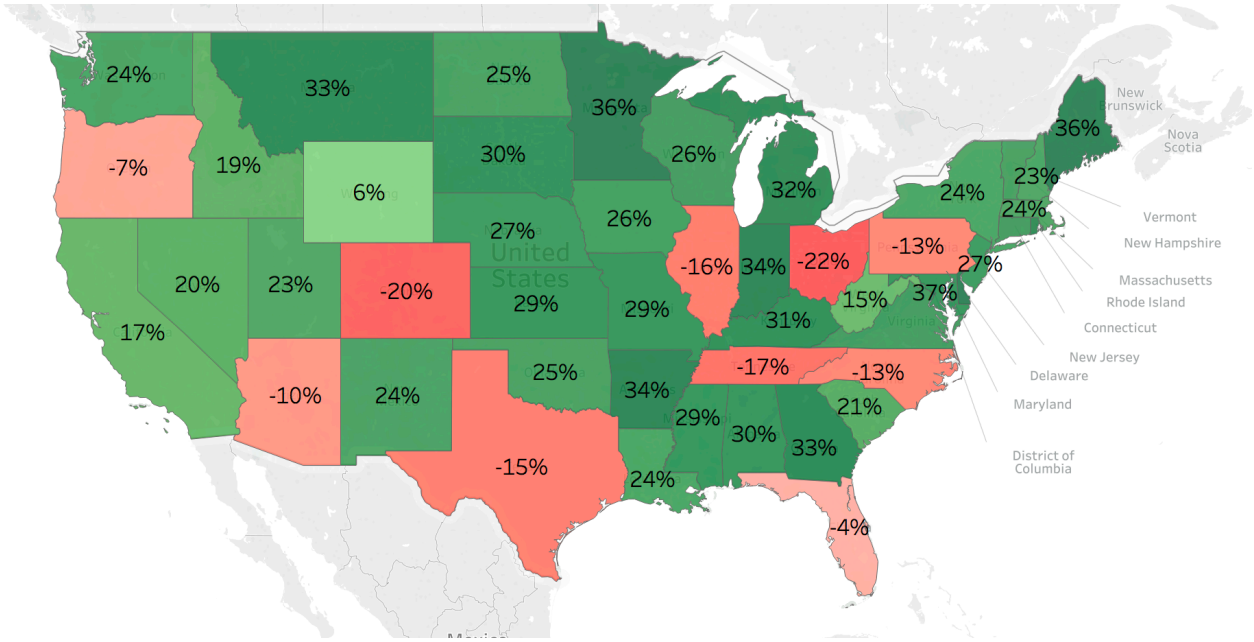


Besides the order counts, we are also interested in the relationship between order quantity and total profit: if it's positive or negative. The below left scatter plot present the quantity and total profit of each state. There are around 10 states where there's negative relationship, meaning that the average profit per order are negative, while for majority states the relationship is positive. More details of the profit rate for each state can be found from the below right histgram plot.





Finally, l use the below Chloropleth Map to present the geographical information: profit rate of each state, which is more informative than the above histogram plot. We can easily match the profit rate to each state and there is no clear geographic relationship between the states with negative profit rate. So, more economic analysis is needed to interpret their relationship.



Summary:

1) What is the trend of sales over time?
Generally speaking, it keeps increasing from 2015 to 2018.
2) Which category contributes more to the sales/profit?
The Office Supplies category contribute to the total sales/profit.
3) Which state contributes more to the sales/profit?
California contributes most to the sales/profit, while the DC has the highest profit rate.