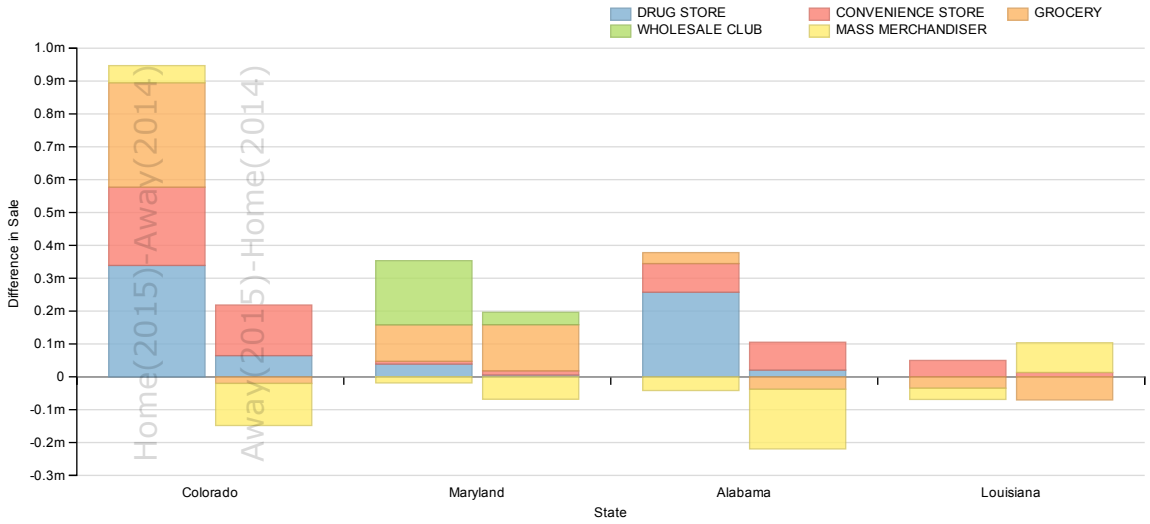
Football Games and Candy Sales in Four States

Data -- This data has been provided by the Hershey Chocolate company. The data contains the sales of candy stores during 15 consequetive saturdays (with football games), the sales difference between 2014 and the corresponding date in 2015 with different type of game (i.e. Home and Away), as well as the average of weekly sales in both years.

The ultimate goal of this visualization is to discover this dataset to see how different market channels and states affect the sales. Also, we want to see if there is any difference between those weeks with home game (2015) and away game (2014) versus the opposite.

To evaluate the effectiveness of the visualization, I first prepared two visualizations and I designed a survey on google forms to ask a number of respondents to provide suggestions that could improve their understanding of the patterns in the dataset. Three respondets provided suggestions and I changed the charts according to their opinions. Both charts have been provided in this write up.



The bar plot above is intended to show to what extent different states and channels are different in terms of average sale. For this case, bar plot is one of the best options, as we only have two categorical variables (i.e. state and channel) and one numerical (average sale). One categorical variable can be presented in one axis and the other can be presented by coloring different levels of the variable. As we can see here, we have managed to show that:

1- The whole sale club sells the highest on average and it can only be found in Maryland.

2- The average sale in Maryland and Colorado are higher than th eother two

3- Grocery stores are the most successful channels except for Maryland's case

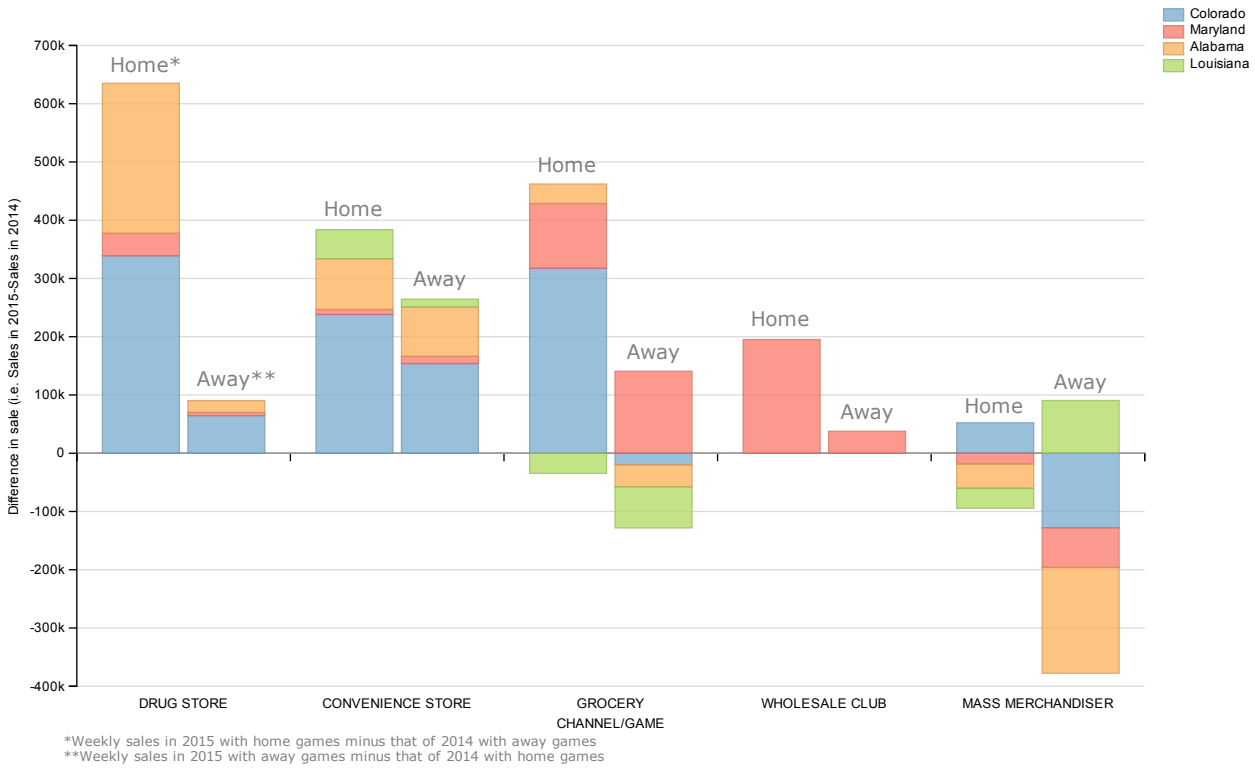
The following s respondents' comments for this chart:

Respondent 1: "Overall, the bar chart is informative but can use some improvements. what stroke me first is the two vertical faded texts on the first two columns in the bar chart. While I understand the intuition behind this, it looks like a mistake. It would be better if you choose another way to show that the two columns are different. "

Respondent 2: "First, it took me some time to understand what home-away means. Second, your bar chart uses five colors and the border lines are pretty close to one another which makes it hard to get any useful information."

Respondent 3: "I think the bar chart is fine overall. Because you are trying to show aggregation on three categorical variables (Home vs.Away, Channels,States) you should think of a way to use the best option for each variable. I would use color for States, because people are often more familiar with States and you have only four states, and I would use text for Channels. This way you will have less colors and hence a less confusing bar chart. I would also explain what home-away means in a footnote. "

After considering their comments, I ended up with the following chart:

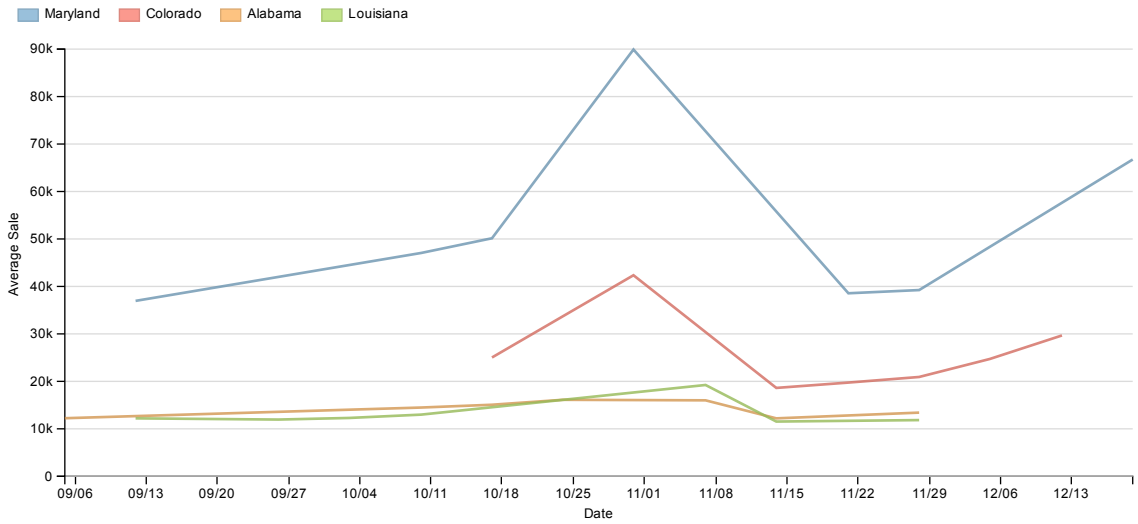


We also wanted to investigate the trend of sales in the four states in this 15 weeks. Line chart is a good option when it comes to time series. Our surveys indicated that lines alone while indicative of overall trend, cannot fully show the extent of diffeence between break points. To address this, I added a bubble chart to accentuate the average sale in each data for each state. the chart above successfuly shows the following:

1- The hierarchy of sales for different states is more evident here.

2- We can see that the week of October's end there is a jump in average sale (probably because of Holloween)

3- The jumps for Maryland and Colorado are more noticable compared the other two states.



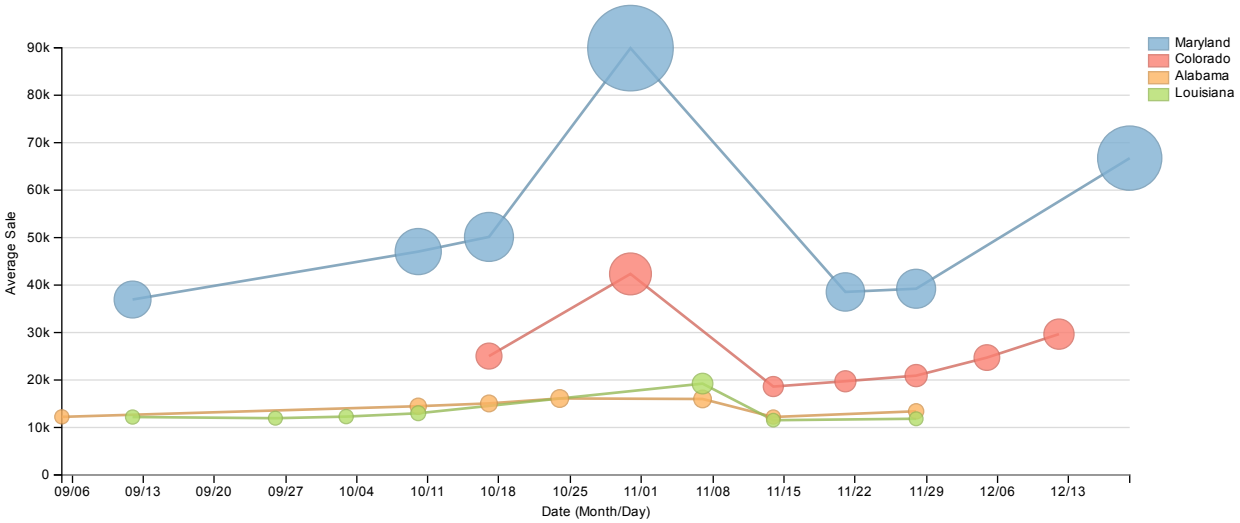
The following s respondents' comments for this chart:

Respondent 1: "I understand the sales trend in each state easily by looking at this graph. However, visually, this is too simple compared to your previous chart. I think it can improve by providing more information on this graph."

Respondent 2: "As an international student, I get confused when I see date formats as such because this is specific to the US. I should hover through all the dates and see which one is more than 12 and assume that it is day and hence the other one is month. I suggest you provide your date format on your X axis. "

Respondent 3: "This is a good diagram and it shows some clear patterns for sales in the four states. One thing that could possibly make this look better to emphasize on the break points in some manner so that we can see the differences between the states where breaks occur."

After considering their comments, I ended up with the following chart:



It would be interesting to also check out the perfomance of different channels. Since the variables are rather the same, we use the same type of visualization. We have already learned how to make a goo visualization based on the respondents' comments for this types of charts. The chart below shows the following:

1- Generally we can see how sales' trends for different channels compare to one another.

2- Convenience store are not as successful during the Holloween week compared to other channels.

3- Unlike other channels, Wholesale clubs are not as successful during the last week of November. Although, recall that our datapoints were less for Wholesale clubs as they can only be found in Maryland.