Data Visualization in Data Science
Iowa Liquor Sales in 2020

May 2021

1 Project description

The data set is provided by the Iowa Department of Commerce, Alcoholic Beverages Division. The Iowa Department of Commerce requires that every store that sells alcohol in bottled form for off-the-premises consumption must hold a class "E" liquor license (an arrangement typical of most of the state alcohol regulatory bodies). All alcoholic sales made by stores registered thusly with the Iowa Department of Commerce are logged in the Commerce department system, which is in turn published as open data by the State of Iowa. The data points in 2020 were chosen to study some interesting questions about Iowa's liquor sales in this year. More information can be found through clicking the link.

2 Data Description

The data set contains information on 24 variables (columns) such as the name, type, price, quantity, and location of sale of a single alcoholic beverage container or package in Iowa. The dataset covers 20.7 million rows and 24 columns, and each row is an individual product purchase. There are various types of columns in this dataset, such as Invoice/Item number are plain text, date presents date & time, Category Name is categorical, Bottle Volume is numeric and so on.

The table below shows several variables that are of value to the research questions:

Tab 1 Explanation of part of data sets

Column Name	Description	Type
Invoice/Item	Concatenated invoice and line number associated with the liquor order. This provides a unique identifier for the individual liquor products included in the store order	Plain Text
Date	Date of order	Date & Time
Store Location	Location of store who ordered the liquor. The Address, City, State and Zip Code are geocoded to provide geographic coordinates.	Point （Latitude ,longitude）

County	County where the store who ordered the liquor is located	Plain Text
Category Name	Category of the liquor ordered.	Plain Text
Item Description	Description of the individual liquor product ordered.	Plain Text
State Bottle Cost	The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered	Number
State Bottle Retail	The amount the store paid for each bottle of liquor ordered	Number
Bottles Sold	The number of bottles of liquor ordered by the store	Number
Volume Sold (Liters)	Total volume of liquor ordered in liters.	Number
Volume Sold (Gallons)	Total volume of liquor ordered in gallons.	Number

Data are summarized by R studio and the results are as follows:

```
Item_Description        Pack         Bottle_Volume_ml State_Bottle_Cost State_Bottle_Retail
Length:2614365    Min.   : 1.00   Min.   :  20.0   Min.   :   0.33   Min.   :   0.50
Class :character  1st Qu.: 6.00   1st Qu.: 375.0   1st Qu.:   6.00   1st Qu.:   9.00
Mode  :character  Median :12.00   Median : 750.0   Median :   8.74   Median :  13.11
                  Mean   :12.15   Mean   : 841.9   Mean   :  10.85   Mean   :  16.28
                  3rd Qu.:12.00   3rd Qu.:1000.0   3rd Qu.:  13.50   3rd Qu.:  20.25
                  Max.   :60.00   Max.   :6000.0   Max.   :1850.00   Max.   :2775.00

 Bottles_Sold      Sale_Dollars       Volume_Sold_Liters  Volume_Sold_Gallons
Min.   :    1.00   Min.   :     1.30   Min.   :    0.020   Min.   :   0.000
1st Qu.:    3.00   1st Qu.:    38.97   1st Qu.:    1.500   1st Qu.:   0.390
Median :    6.00   Median :    81.60   Median :    4.800   Median :   1.260
Mean   :   11.41   Mean   :   151.72   Mean   :    9.261   Mean   :   2.441
3rd Qu.:   12.00   3rd Qu.:   157.56   3rd Qu.:   10.500   3rd Qu.:   2.770
Max.   :13200.00   Max.   :250932.00   Max.   :13200.000   Max.   :3487.070
```

Fig 1 summary of part of data sets

3 Research Questions

To explore our dataset, the following five questions are used as the starting points to create the visualizations.

1. How much liquor is sold and consumed in each county? Which county has the most orders for liquor?

2. How many items are in each category? What are the most popular and least popular categories?
3. Which 10 stores have the most liquor sales and how does the monthly sales change?
4. Which order has the most profit? How many bottles were sold on the order? What is the price difference between the Beverage Division and the Store?
5. Is there any relationship between liquor sales and time, such as month and season?

4 Visualisation development and results

4.1 Exploration of design space

4.1.1 Online reference

The data set we selected is a huge data set, and some interesting works have been done which can give us some inspiration.

In Kaggle, some interesting work gives us some ideas as shown in Fig 2. You can find more information through this link.
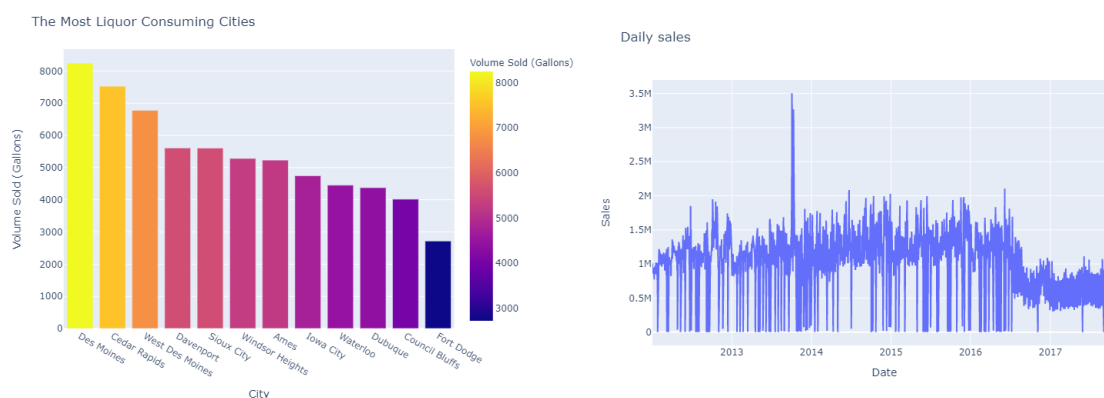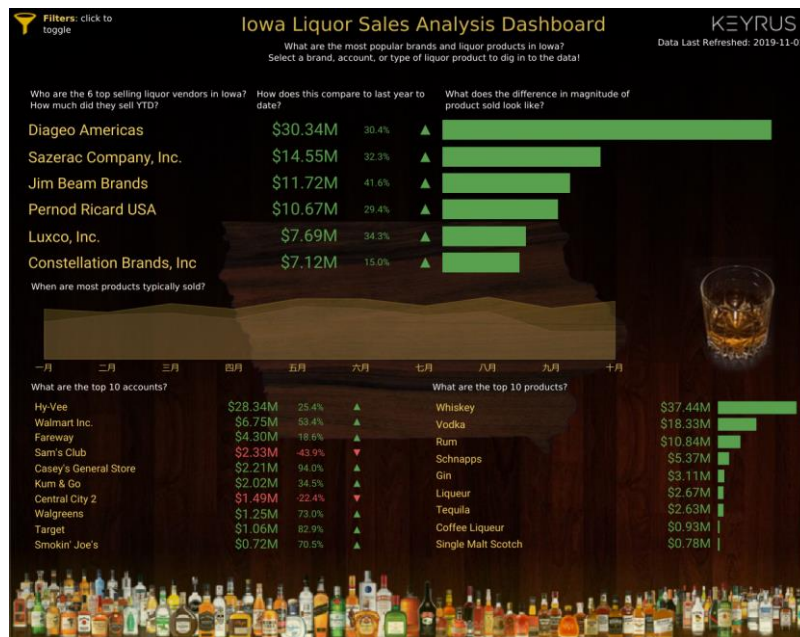

Fig 2 Reference on Kaggle

In the public site of Tableau, we also find some interesting references of the similar dataset. In Fig 3, we can know how to visualize the rising sales. More details can be found by this link.
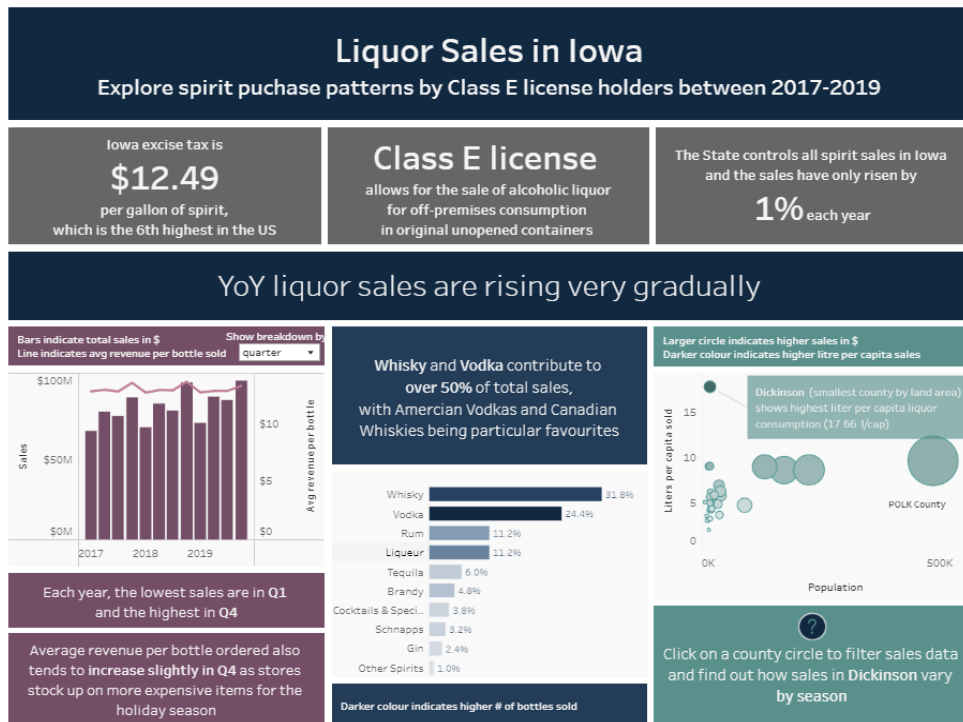
Fig 3 Reference on Tableau（1）

There are some other great examples such as the work shown in Fig 4 and Fig5. These works focus on different problems and they give us some ideas to carry out our project. For example, how to show more information in a graph? How to design the interaction to help users find more results in the visualization?

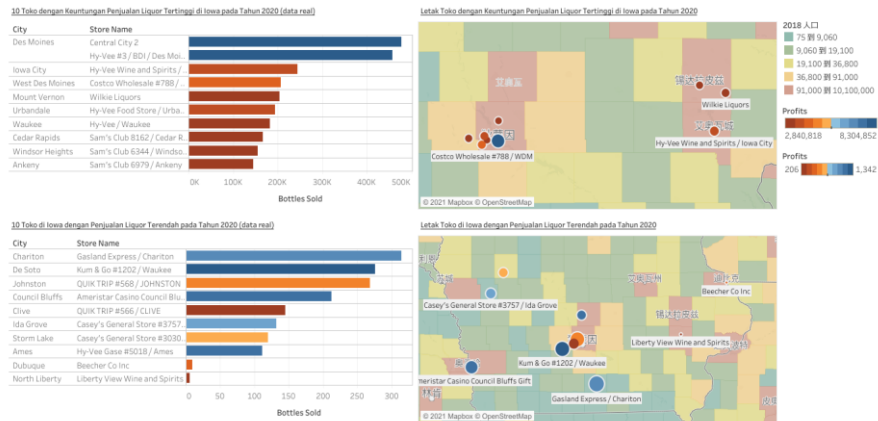**Fig 4 Reference on Tableau(2) (link)**

Fig 5 Reference on Tableau(3) ([link](#))

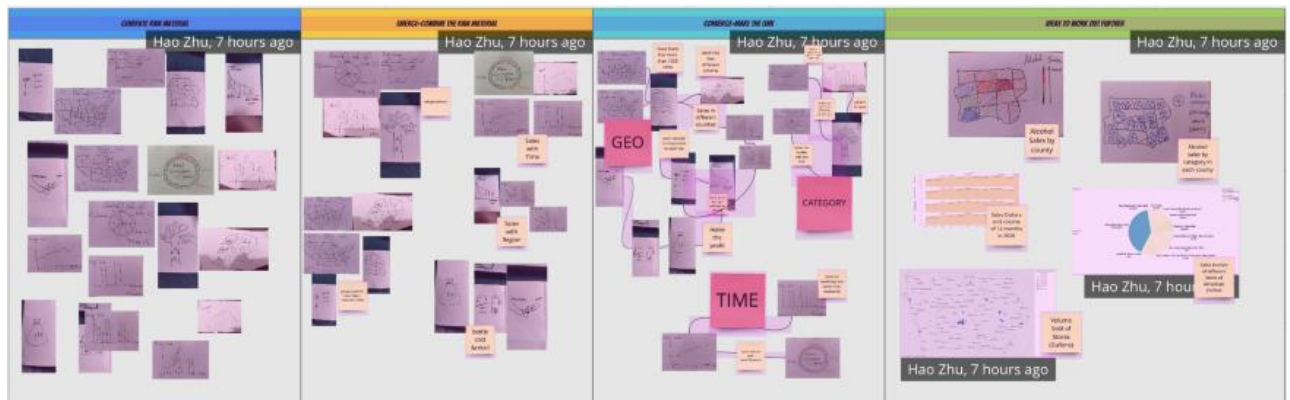## 4.1.2 Results of design exercises

Fig 6 A Miro Board Overview ([link](#))

As shown in Fig 6, our design work was conducted by the Miro Board. There are 4 phases totally, which are Generate Raw Material, Emerge-Combine the Raw Material, Converge-Make the Link, Ideas to Work out Further.

Phase1 Generate Raw Material

In this phase, we generate a lot of sketches from our raw data, which covers categories, cost, profits etc. There are some interesting sketches in diverge as shown in Fig 7. Left graph is about categories and items. The big branches represent the categories of the liquor, and the witches are items within the category. Right graph represents the map, the density of the dots is how many gallons were sold.
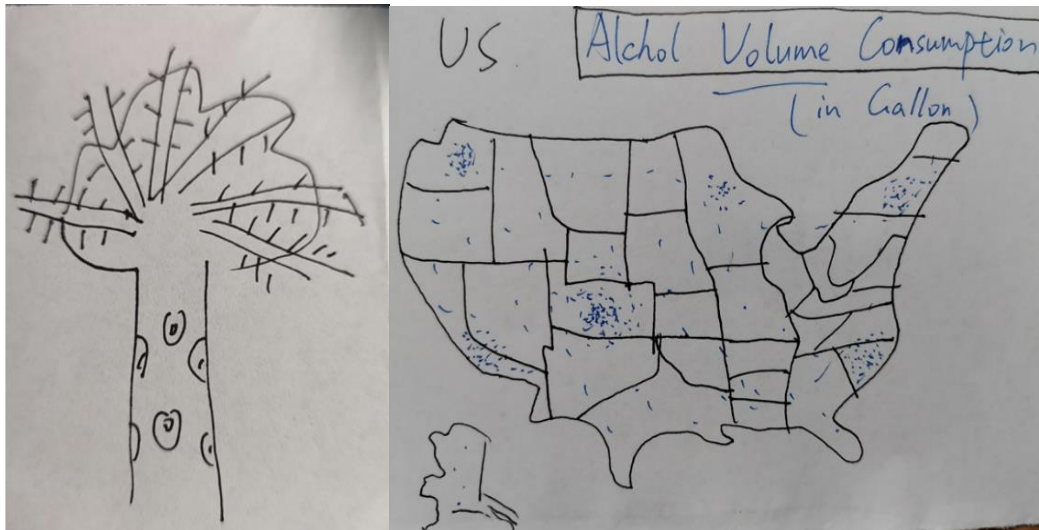
Fig 7 Categories(left) and Volume Consumption (right)

Phase2 Emerge-Combine the Raw Material

In this phase, we emerge from the combination of our raw material according to category&item，geography(US), sales with time, sales with regions and bottle cost&retail. In this combination, we combine all the sketches about the time. As the last combination, the scale of time is also getting changed. There are four sketches respectively representing the consumption in seasons, months, weekends, or workdays. You can figure out the trend with time series clearly.

Phase3 Converge-Make the Link

As for the convergence of our combinations, we draw some arrows to link up the sketches to clarify our design.

First, we can see the geography combinations, which starts with the US map, and then link to the Iowa State and its cities. Next, we add the volumes in different volumes and sales in different counties, which leads to the sales profits finally. Similarly, in the category and item combinations, we start with a tree representing all the categories and items. It leads to different types and categories of liquid.

Phase4 Ideas to Work out Further

At the last phase, we show more ideas to work out further, including alcohol sales by county and alcohol sales by category in each county.
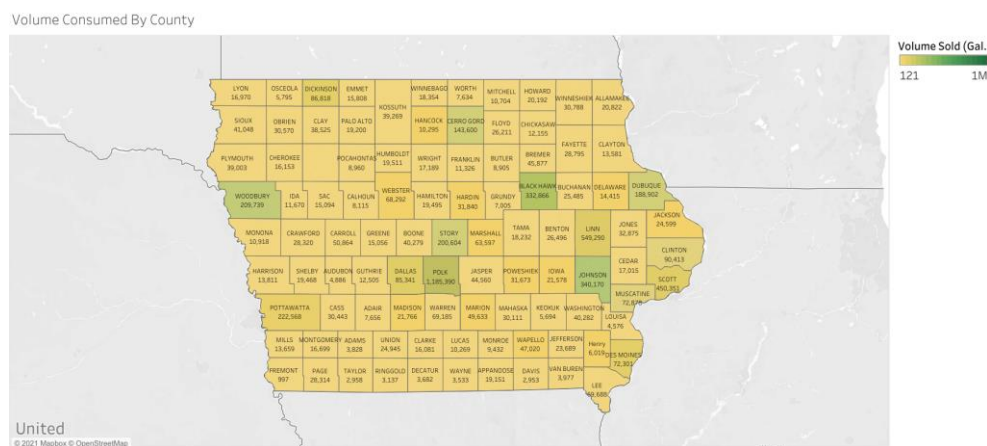
Fig 8. Liquor Sales(left) and Sales by Category in Each County(right)

4.2 Final design

Q1: How much liquor is sold in each county? Which county has the most orders of liquor?

Fig 9 Volume(Gallon) sold in counties (link)

In Fig 9, we can see counties which consumed the most liquor and the color on the map represents the volume sold. So, it is obviously figured out that Polk county sells liquid most. To investigate the sales more deeply in each county, we build another similar graph in Fig 10 in which the color on the map represents sales and the pie in the county represents the volume of different liquors. To find more details, you can click the links below the figures.
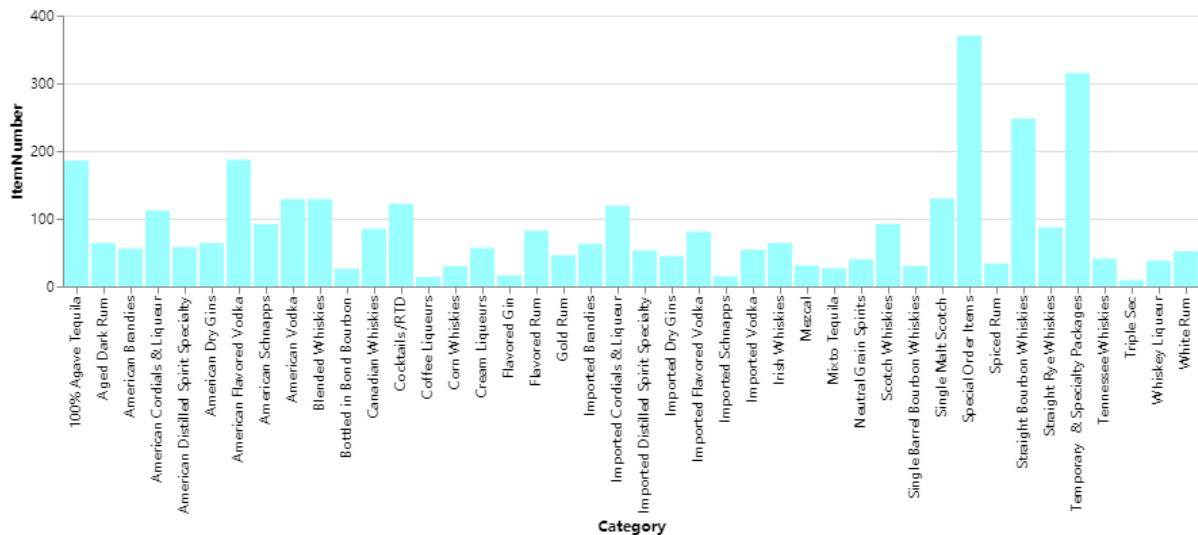
Fig 10 Sales in counties (link)

Number of liqour orders in different counties
Click to Zoom

Fig 11 Orders of each county (link)

The number of orders is another good way to judge which counties enjoy liquor. In Fig 11, we find POLK and LINN have the largest numbers of orders. If you click one of the counties, you can see a bigger and clearer graph.

Q2 How many items are in each category? What are the most popular and least popular categories?

Fig 12 Item number of Category (link)

Fig 13 Sales of different liquor categories (link)

In Fig 12, we can learn about the item number of categories. Clearly, Special order items has the most items of liquor. In Fig 13, each color represents each brand of the certain category. The bigger sector, the larger sales of the corresponding category. From this graph, we can figure out the most popular categories and least popular categories.

Q3: Which 10 stores have the most liquor sales and how does the monthly sales change?
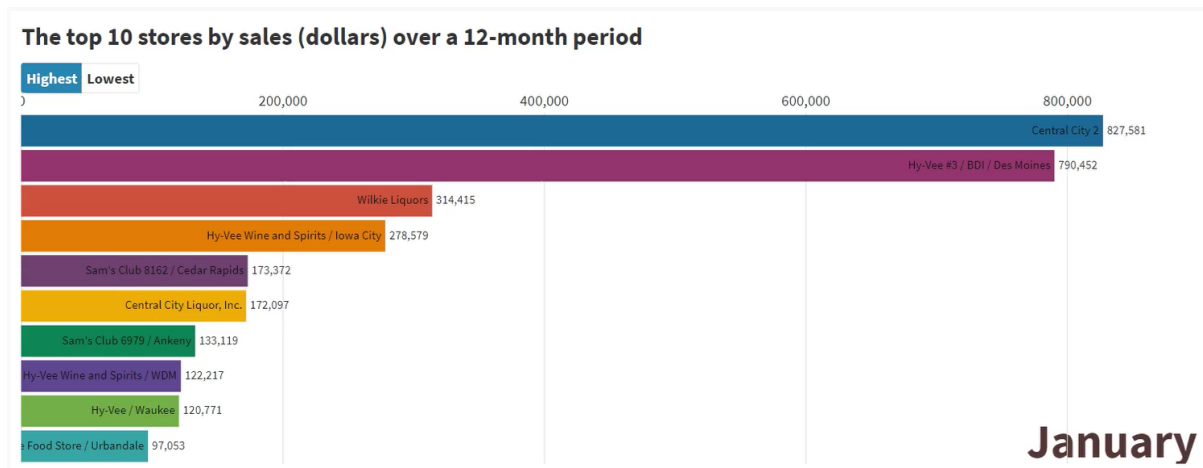
Fig 14 Top 10 stores by sales over a 12-month period (link)

We chose to use video dynamics to show how the top 10 stores' sales change over the next 12 months of the year. In the video, you can see the sales of ten stores running like a race. The final No.1 store was Hy-Vee #3 / BDI/Des Moines.

Q4: Which order has the most profit? How many bottles were sold on the order? What is the price difference between the Beverage Division and the Store?


Fig 15. Most profitable orders (link)

These are the top 15 orders profit by (retail-cost) times the number of sold bottles. The vertical axis represents (retail-cost), that is, the price difference; the horizontal axis represents income; the size of the bubble represents the number of sold bottles sold in this order.

*Q5：Is there any relationship between liquor sales and time, such as month and season?*
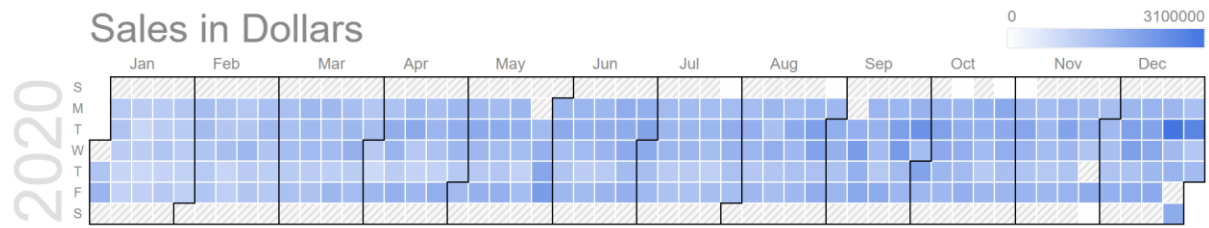
Fig 16 Calendar sales (link)
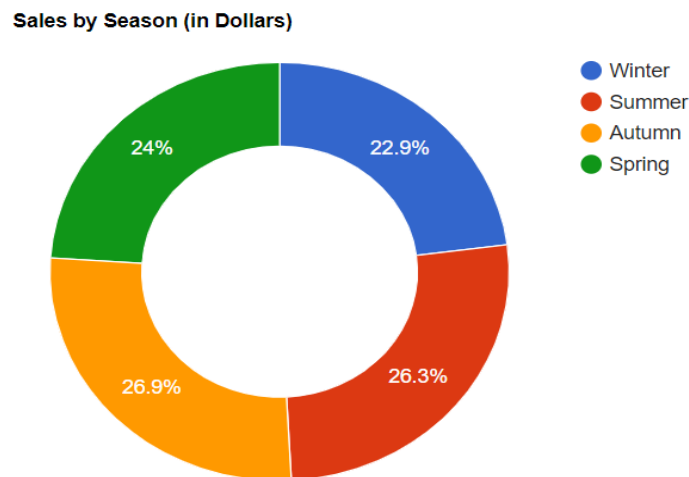
Fig 17 Sales by season (link)

To investigate the relationship between liquor sales and month, the calendar sales are shown in Fig 16 and you can interact with the calendar by moving your mouse through this link. Clearly, the color in December is darker than other months. In Fig 17, we can find the relationship between sales and seasons. To have more detailed information of sales, the curve of sales and the bar plots of volume are shown in Fig 18. At the start of the year, the sales of liquor in Iowa increased until June and the curve are flat from March to May. After June, a decrease was witnessed until August. There was a sharp decrease in November which may be caused by the COVID-19. In December, the sales of liquor rose sharply and reached the highest of the year which was consistent with the calendar.
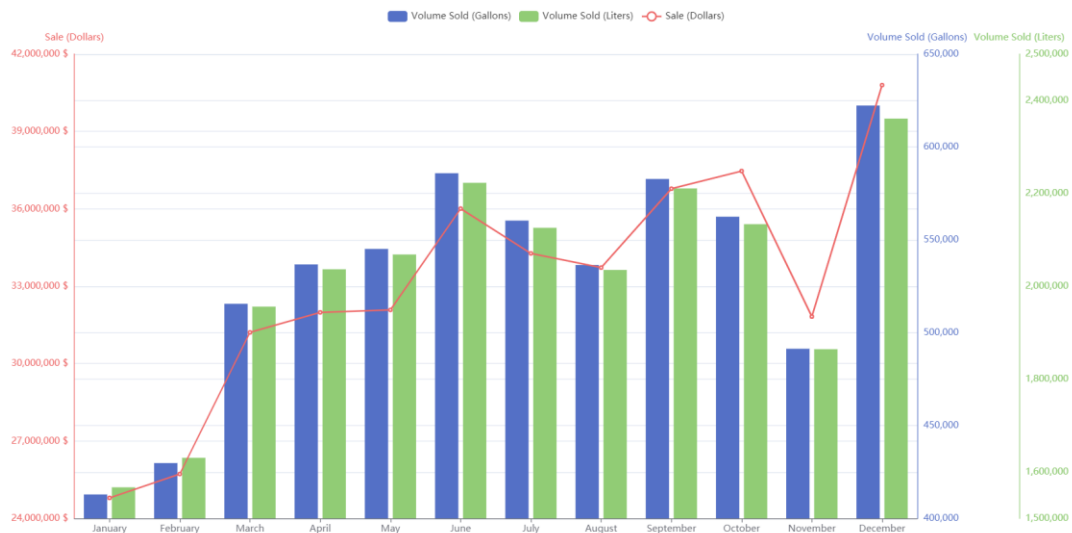
Fig 18 Relation between sales (volume sold) and month (link)

4.3 Implementation

All the work is finished by 4 members in the group. There is no implementation of students from other groups.

We started with group discussion and brainstorming. Under the guidance of the professor and the teaching assistants, we made sketch drawing and preliminary design, followed the three processes of diverge, emerge and contrive, and generated our own ideas and insights. Then, we used a variety of visualization software and programming language for data visualization analysis, such as Vega, Echarts, Tableau, R Studio and so on. Each progress in the meantime will be updated on the blog as requested. Finally, we gathered all the achievements and had group discussions again to write the report and make the video.

4.4 Insights

Through the research and summary of this data set visualization, we have a comprehensive and in-depth understanding of the structure and content of this data set. For example, the correlation between the sold alcohol of volume, sales dollars and the geographical and time, the price difference between the beverage division and the store as well as the sales situation, and so on. At the same time, we also master a variety of data visualization tools and programming languages. For example, different visualization tools can give us various visual sense of research questions. Through

interaction (clicking, selection & scrolling) with the graph, we can reach the detailed information of data.

At the same time, through a semester of group cooperation (although it was online), we also fully realized the importance of team cooperation, especially in the design stage of data visualization, brainstorming is very useful. Through clear division of labor and cooperation, we successfully completed the tasks of each stage, which is of profound significance to the learning of other courses in the future.

4.5 Description of each student's contribution

Tab 2 Description of contribution

Student Information	Contribution
Hao Zhu(r0816422)	1. sketch design 2. design and completion of question3 3. Participation in video production 4. report writing
Shiqi Wang(r0815604)	1. sketch design 2. design and completion of question 5 3. Participation in video production 4. report writing
Xiao Wang(r0819511)	1. sketch design 2. design and completion of question 1 and 2 3. Participation in video production 4. report writing
Shichen Wu(r0862533)	1. sketch design 2. design and completion of question 4 3. Participation in video production 4. blog writing

4.6 Links of video and codes

You can view our video of the project through clicking the link attached with graphs. Also, you can click the links in the Tab 3 to check graphs, code, and video.

Tab 3 Links of the project

Content	Links
Data set	https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy
Qustion1	Fig 9: https://public.tableau.com/profile/xiaowang6870#!/vizhome/Volu

	meConsumedByCounty/VolumeConsumedByCounty?publish=yes Fig 10: https://public.tableau.com/profile/xiaowang6870#!/vizhome/xw-22/sheet1?publish=yes Fig 11: https://moodlunatic.github.io/datavis/numberOforders.html
Question 2	Fig 12: https://moodlunatic.github.io/datavis/CategoryAndItems.html Fig 13: https://moodlunatic.github.io/datavis/SalesCa.html
Question 3	Fig 14: https://moodlunatic.github.io/datavis/TopStores.html
Question 4	Fig 15: https://moodlunatic.github.io/datavis/bubble-gradient.html
Question 5	Fig 16: https://shiqiii.github.io/ Fig 17: https://shiqiii.github.io/sales_by_season.html Fig 18: https://moodlunatic.github.io/datavis/salesMonth.html
Code	https://github.com/moodlunatic/datavis https://github.com/shiqiii/shiqiii.github.io
Video	https://www.youtube.com/watch?v=Bb18X5vsQYs&t=68s