Lab 16 Introduction to Tableau

Learning objectives:

- connecting to a data source
- dimensions and measures
- creating a view
- aggregation SUM(), AVG(), etc.
- aliases
- bar and scatter-plot charts
- filtering
- creating a calculated measure
- creating a table calculation
- dashboards

1. Connecting to a data source

Tableau can connect to many types of data including .csv or excel. We will connect to data stored on a local Microsoft SQL server. After you open Tableau select Microsoft SQL Server. Complete the following connection:

a. server name: **econsql**
b. database: **bibitor**
c. click use specific username and password
d. username: **student**
e. password: **UnionEconomics!**

The bibitor database contains information on transactions in a chain of wine and liquor stores. It includes several tables. For this lab, let's drag tables *sales* and *products* onto the canvas. Let's check how Tableau *joined* the two tables. Let's also check the *data types* that Tableau assigns to each column.

2. Dimensions and measures

Click on *sheet1*. Under *Data tab* we see two panes: Dimensions and Measures. The Dimensions pane includes qualitative variables (e.g. *Description*). The Measures pane includes quantitative variables (e.g *Sales Dollars*).

3. Aggregation

Let's create our first view by dragging *Sales Dollars* into the view. Tableau created one mark, i.e. one data point in the view. It is a number that is equal to the SUM of *Sales Dollars* variable across the entire database. There are many transactions in the database, each transaction has a Sales Dollars associated with it. If we just ask for *Sales Dollars*, Tableau needs to figure out

how to aggregate across all transactions and display a number. SUM is the default aggregation. We can change it to average, AVG, and many other aggregation methods.

IN-CLASS EXERCISE 1: What happens if we change the aggregation from SUM() to AVG()? What about COUNT()?

4. Adding dimensions to our view

Let's switch the aggregation method back to SUM, and add *Description* to the rows shelf. We now generated 11 thousand marks. The level of detail or granularity in the view now is *Description.* The numbers are the sums of *Sales Dollars* for each product description. We can click on Description in the rows shelf and select "sort", select sort by the *Sales Dollars* field in order to see products from the highest sales to the lowest. Which product is the top seller?

Let's add *Sales Price* to the view. Does it make sense to aggregate price using SUM? Let's switch the aggregation to AVG. How much does Jack Daniels normally sell for?

IN-CLASS EXERCISE: Let's add *Sales Quantity* to the view. Does the product of average price and sum of quantity equal total sales?

5. Creating a bar chart

Let's open a new sheet and create a bar chart with sales for each year and a quarter. We can do this by dragging *Sales Dollars* into the rows shelf and then dragging *Sales Date* onto the columns shelf. The default date granularity is year, but we can change this by clicking on the + icon in the year. We can switch from marks displayed as lines to marks displayed as bars by switching selecting bar in the Marks card.

Would you say that liquor and wine sales display seasonal fluctuations?

IN-CLASS EXERCISE 2: Split the quarter into months. Are there months that have particularly high sales?

6. Creating Aliases

The dimension *Classification* takes value 1 if the transaction was liquor and 2 if the transaction was wine. Let's create an alias so that in our visualizations 1 appears as "liquor" and 2 appears as "wine". We can now drag *Classification* onto the rows shelf. Tableau creates a new set of rows each row corresponding to each distinct values of *Classification*. This allows us to see the time-series variation in sales for wine and liquor separately.

Are sales generally higher for wine than for liquor?

IN -CLASS EXERCISE 3: Let's drag Classification from the rows shelf to color in the marks card. How does the viz change? Do you like this viz better? Did the number of marks change?

IN -CLASS EXERCISE 4: In the previous viz the liquor bar was stacked on top of the wine bar. Can you figure out a way to have the bars next to each other instead?

IN-CLASS EXERCISE 5: Go back to the stacked bars. Add product *Description* instead of *Classification* to color. What happens? Can you sort the products within each bar by their sales?

7. Creating a scatterplot

Start a new sheet. Let's put SUM of quantity in the columns shelf and AVG price on the rows shelf. How marks did we generate? Without any dimensions Tableau will generate just one mark - aggregating quantity and price across all of the transactions. We can increase the level of detail by adding various dimensions to the detail option in the marks card. Let's add Product id, Store and Date. Let's also change the axes' scales to logarithmic by clicking on the axes and selecting logarithmic scale. We can add a best fit line by clicking on the analytics tab and adding a logarithmic line.

How many marks are we plotting?
What is the nature of the relationship between price and quantity?

8. Filtering

Sometimes we want to view only certain observations. For example, if we wanted to see sales of Jack Daniels No. 7 Black only, we can drag Product Description to filter and select filter. Clicking on the variable in the filter carn we can "show" the filter in view and control it in the view.

IN-CLASS EXERCISE 6: Can you figure out why the data points are clustered around different price points?

9. Creating a calculated field

Tableau allows us to create new fields as functions of existing fields. Right clicking in the data pane opens a menu from which we can select 'Create a calculated field'. Let's create our own measure of unit price which we define as *Sales Dollars/Sales Quantity*. Compare this measure to *Sales Price*. Are they different?

IN-CLASS EXERCISE 7: What would be the best way to make sure that our calculated price is identical to *Sales Price*?

10. Creating a table calculation

Let's create a table with sales for wine and liquor in different rows and date in columns. Thus we should have *Classification* in the rows shelf, *Year* in the columns shelf, and SUM of *Sale Dollars* in the marks card.

Suppose we wanted to calculate growth in sales. We can click on the measure of SUM of sales and click on 'add table calculation'. In the window we can select to calculate percent difference from 'Specific Dimension' and click on 'Year of Sale Date'. Does that give us what we want? Let's add total sales to the table to verify the growth calculations.

We can drill down *Year* into *Quarters* by clicking the '+' symbol next to *Year*. Is this a more meaningful table? How is the wine business compared to the liquor business?

IN-CLASS EXERCISE 8: Can you format the color of the numbers such that negative growth is red and positive growth is green?

11. Creating a dashboard

Visualizations in different sheets can be combined into a dashboard. Tableau is especially intuitive in this part - just click on dashboard tab, drag sheets into view. You can edit, resize, add textboxes titles, etc.

Exercises:

For each question below create a sheet in your Tableau workbook. Put the sheets in a visually pleasing dashboard. Include a textbox with answers to the questions below. Print the dashboard and bring to class.

1. Connect to the bibitor database. Create a sheet that shows sales of liquor and wine in each store. Which store number had the highest total sales? Which store had the highest sales of liquor?
2. Which store had the highest growth in sales in 2017 Q2 compared to the same period last year?
3. Connect the store table to your data. What is the relationship between store square footage and store dollar sales? (Hint: Be careful about how square footage is aggregated in your visualization.) Which stores would you say underperform in terms of its sales given the store's square footage?
4. Create a table with each store's sales per square foot. Does your table agree with the store you identified in question 3?