Tableau Visualization - Geographical Exploration

Example geographical questions

- Planning for the “Hydrogen Economy”
  - How many hydrogen fuel stations need to be built so that 70% of the U.S. population lives within 2 miles of one?
  - How many hydrogen fuel stations need to be built so that there is at least one every 50 miles on all of the interstate highways in the U.S.?

More examples

- “Point decisions”
  - Where should we locate our distribution centers? How many are needed to cover the population?
  - Where should we build cell phone towers?
  - Where should we build fire stations?

- “Line decisions”
  - What route should we follow to get from A to B?
  - Where should we build our next road? What roads should we widen?

And some more…

- "Area Decisions"
  - How should we divide up the country into sales regions?
  - How should we define the boundaries of voting districts?

- Many decisions have a geographical aspect to them
  - If we see the decisions displayed on a map we can judge their effectiveness more quickly
  - We need to see the data driving the decisions displayed geographically, too

Our Goals in This Section

- Understand Tableau basics
- Convey information about the data attached to points
- Aggregate sales data by zipcode and display on a map
- Explore data for geographical insights
- Use other visualizations to validate findings
- Explore data with visualization

To Get Credit For This Recitation...

- Screenshot your completed visualizations when instructed to in these slides
- Your final submission should include 3 visualizations.
- Submit a pdf with the screenshots onto Gradescope

Instead of Tableau.. you could potentially use GIS packages

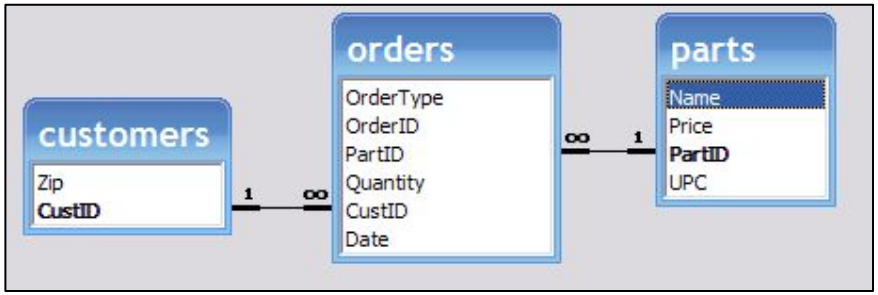
For GIS instructions, see here:

https://docs.google.com/presentation/d/1SAJ_Bi0I0Sal6fuhVpSJkdsU-JNyIde0SAx1D01mxLw/edit?usp=sharing

Benefit: GIS is very robust for map-making

Cons: Tableau has a more modern UX, and is more likely to be used in workplaces

Database Schema



- Zip = 5 digit zipcode of dealer
- OrderType = “RR” (replenishment order)
  or “EO” (emergency order)
- Quantity = quantity ordered by dealer
- Date = date ordered by dealer
- Price = unit price of part
- Name = part description
- UPC = uniform product code
  (a useful classification of parts)

Downloads

Download the appropriate csv files from the course webpage (we have 5), and put them in an easy to access directory.

If you haven't already, download the student version of Tableau - it is a free 1 year license.

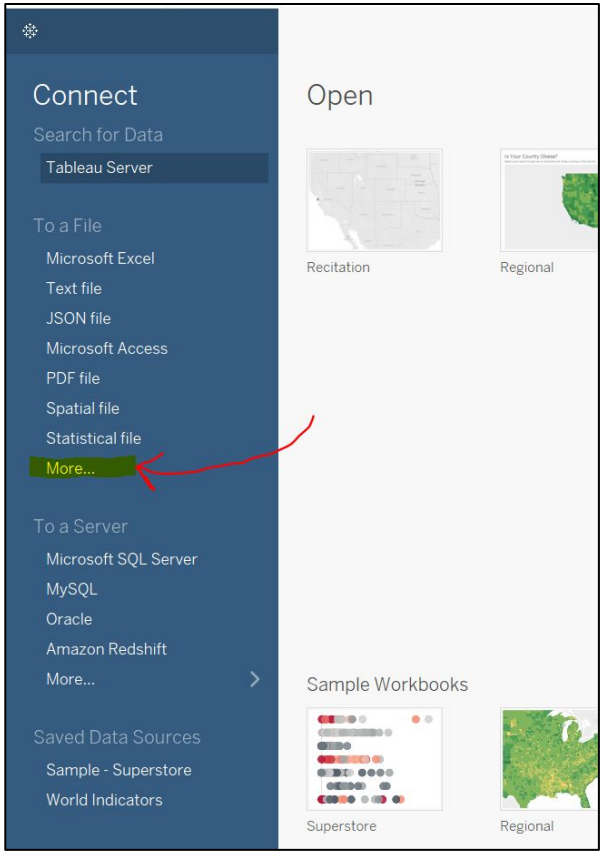
After downloading, open the program.

Preliminary look at csv files

Open your csv files…

See what columns of each file may be able to join with columns of other tables. We will need to bring all csvs together!

Importing data

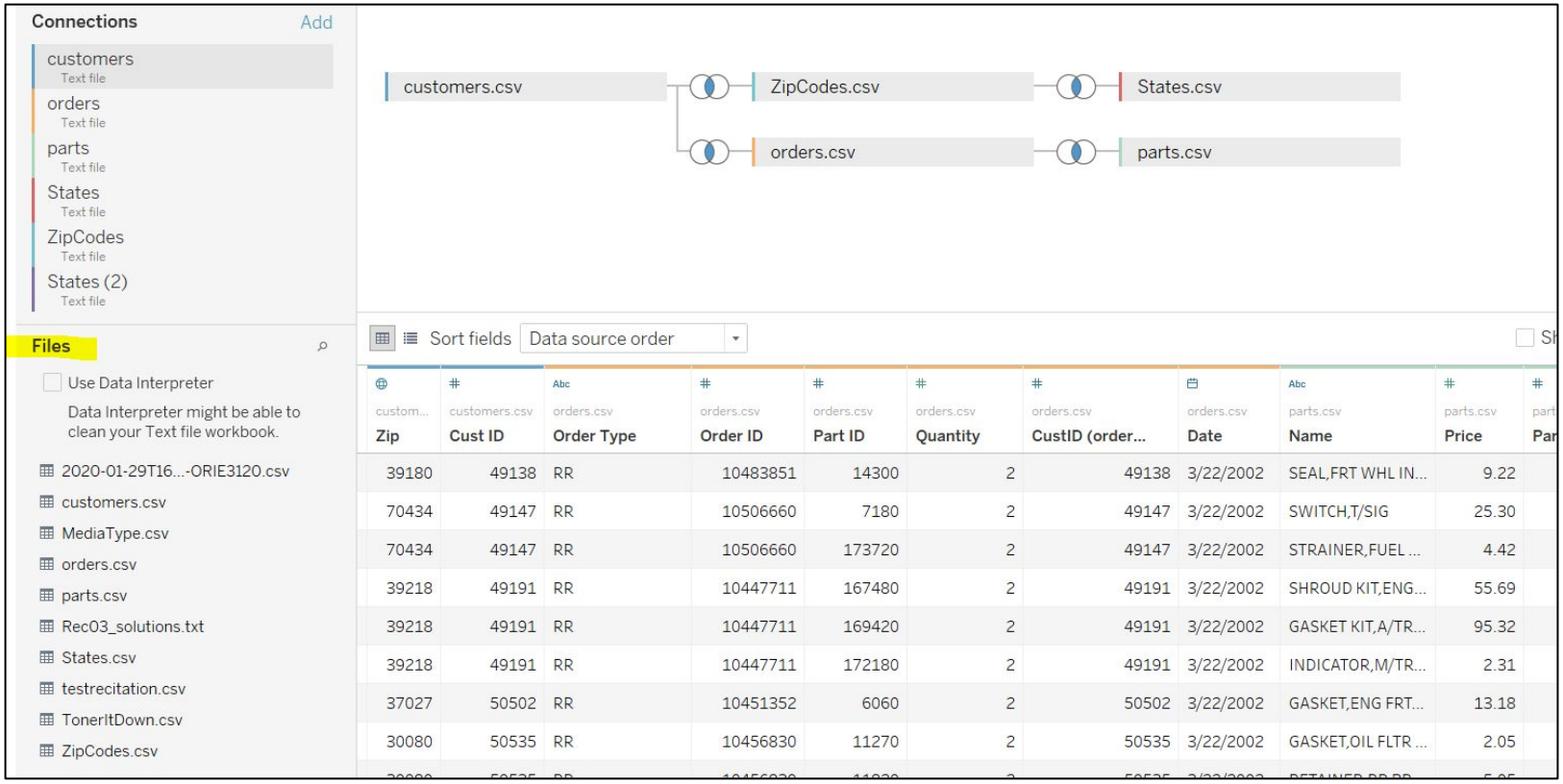


On the left side of Tableau's menu, choose the type of data you are trying to import.

For this lab, we are using csv files. Choose "more" to find that option.

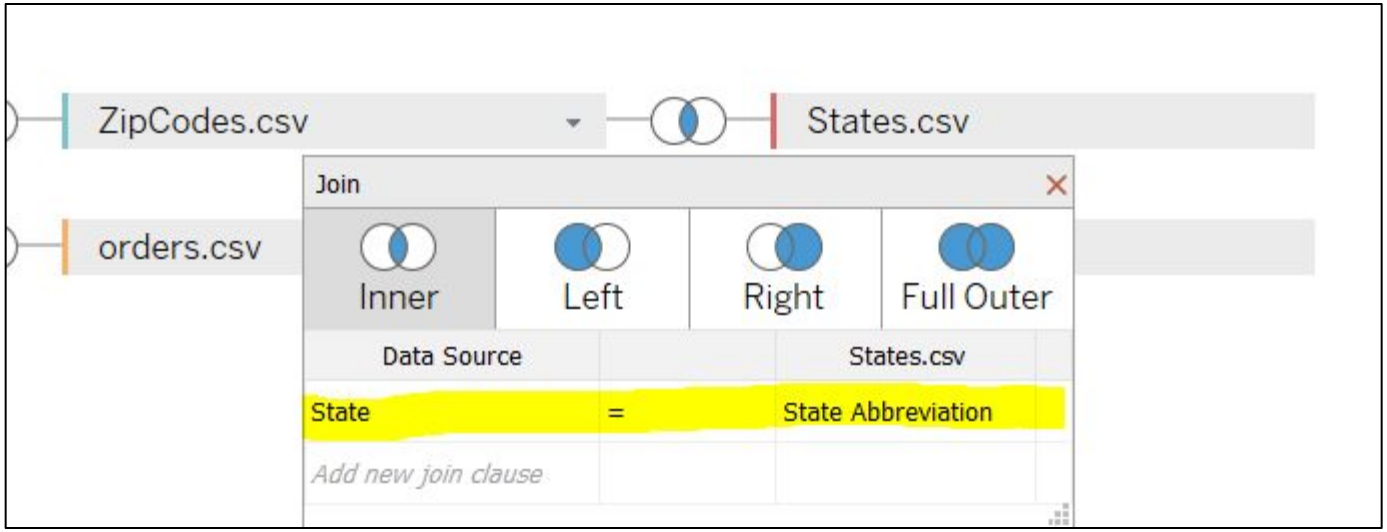
Import all csv files for the lab.

Joining data



Drag a csv from files (highlighted) to the diagram on the top right one at a time, and inner join them all on valid connections (see next slide).

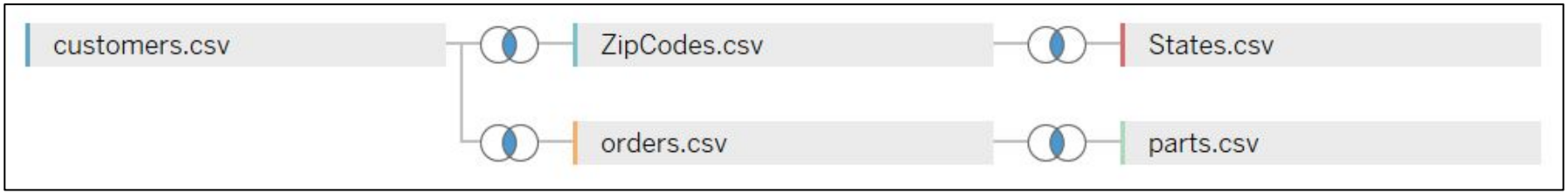
Joining data sets



Example above: inner join ZipCodes and States

We look at the specifications, and know that "State" in ZipCodes corresponds to "State Abbreviation" in States - and select these two fields from the respective dropdowns. Tableau might auto-join some tables for you.

Suggested join ordering

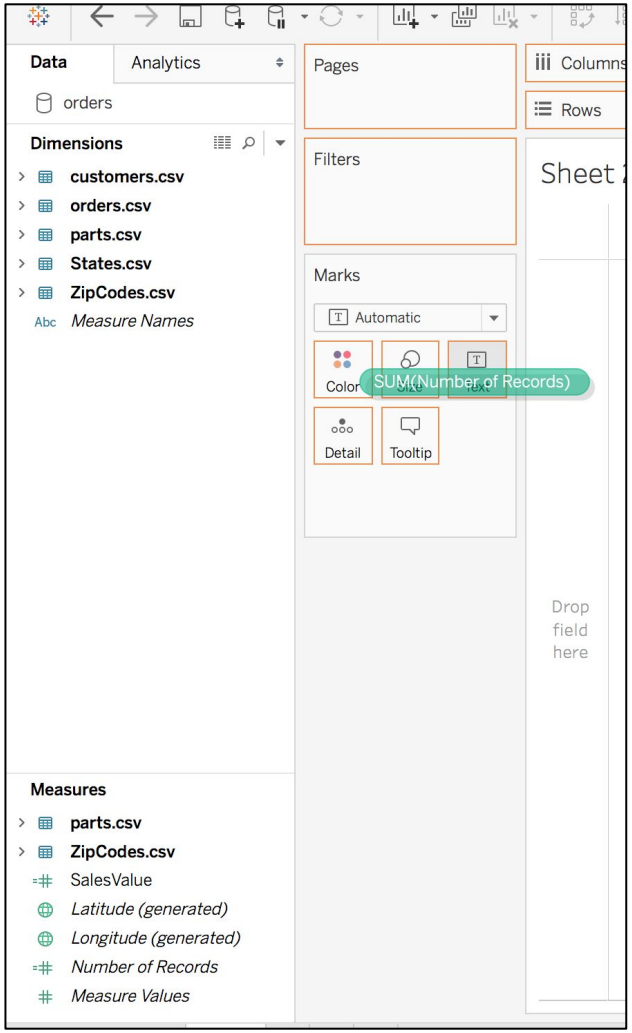


Our workbook has the following set of inner joins:

- customers.Zip = ZipCodes.Zip
- ZipCodes.State = States.State Abbreviation
- customers.Cust ID = orders.CustID
- orders.Part ID = parts.PartID

We want inner joins here (as opposed to left/right/outer/unions) because we want each entry to be unique and complete with data

Check the number of rows



Click to the “Sheet 1” tab (or any blank sheet) on bottom.

Then, under “Measures”, drag “Number of Records” to the Marks box called “Text” as demonstrated to the left.

Verify number of rows

If you do not have 545,370 rows, please check your work with peers or a TA.

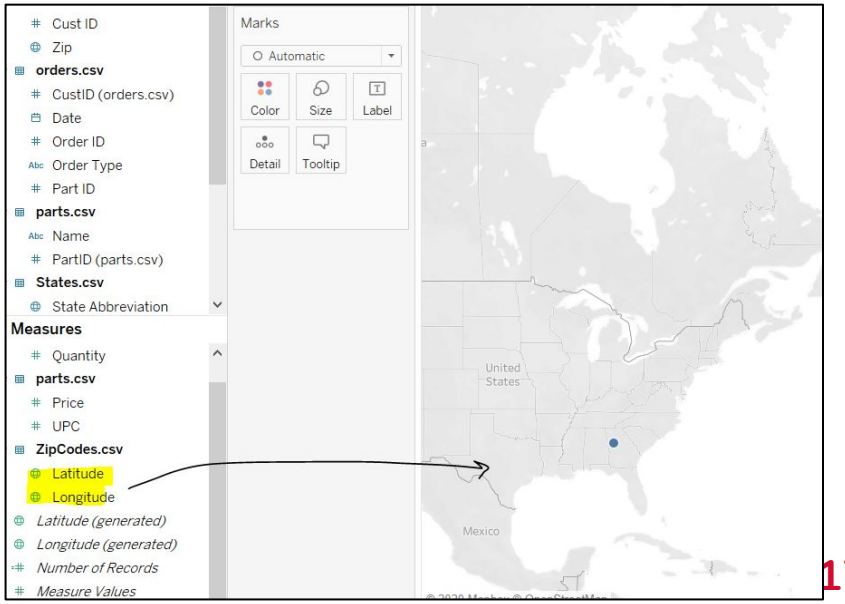
Begin Mapping

Drag “Number of Records” off the page to remove it and return to a blank sheet.

Then look under ZipCodes.csv in the Measures section. Select Latitude and Longitude simultaneously and drag them onto the Sheet to see a map populated.

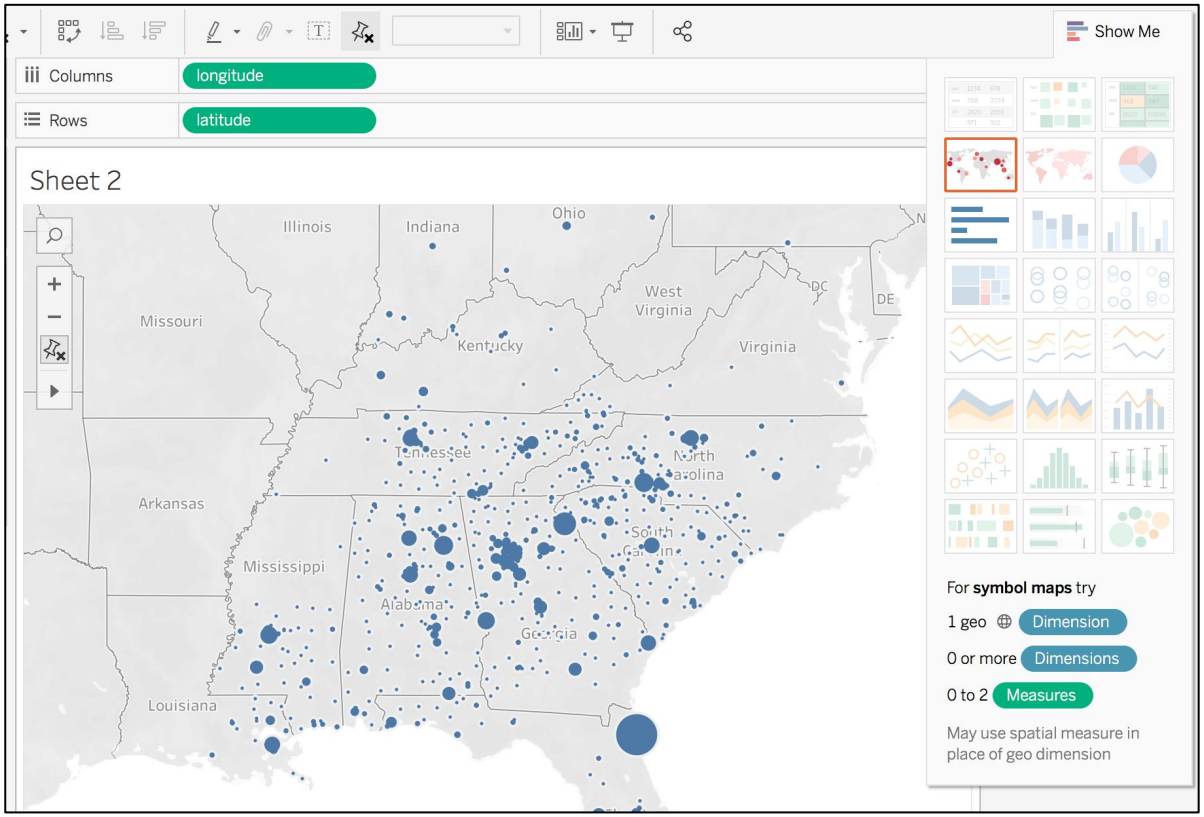
Troubleshooting:

- **DO NOT grab “Latitude (generated)” or “Longitude (generated)”.**
- **See next slide for alternative instructions if the map doesn’t appear.**



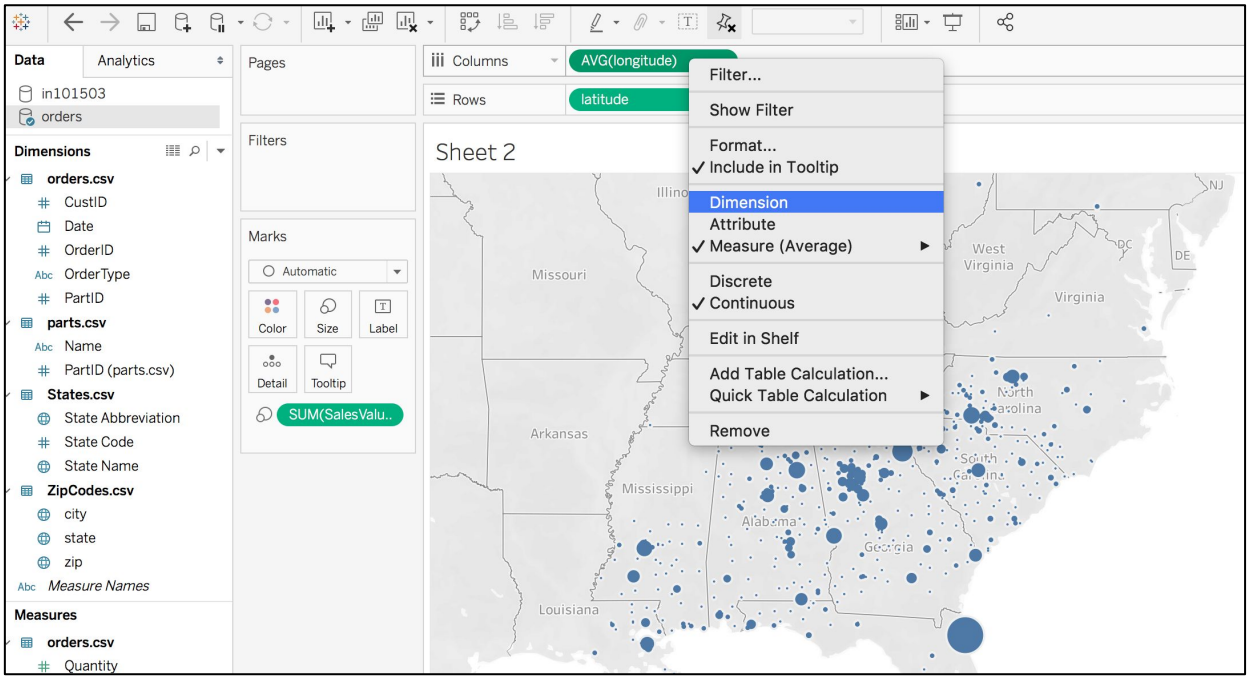
If Map Doesn't Appear

Click on "symbol maps" under "Show Me".



Dimensions vs Calculated Values

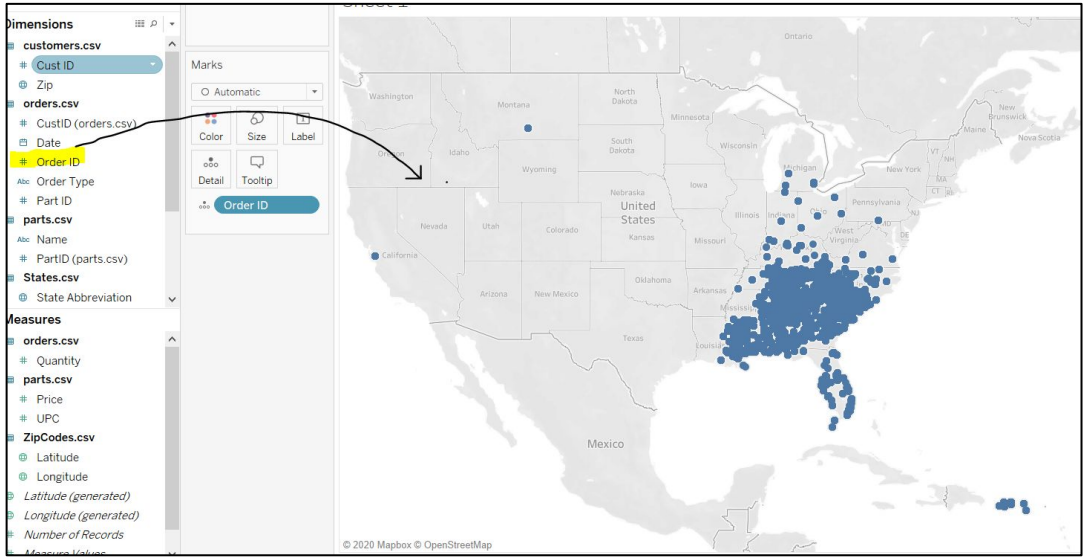
Tableau initializes measures to be their averages (rather than their individual values). Set latitude and longitude to be dimensions to get their individual values.



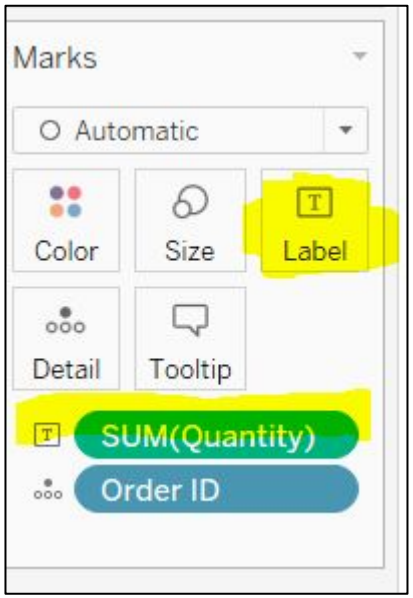
Populate the map

Drag “Order ID” to the map to see a dot representing each order.

You could also drag a number of different dimensions over to see the dots populated, as our tables are joined, and each has the unique location of the order tied to it through joins.



Adding more info



Drag a field you are interested in (such as Quantity) to the Label box.

Then mouse over any dot on the map, and see the sum of order quantity listed in the label flag.

Right click the green SUM(Quantity) bar, and change the Measure to “Count”, and see how the label flag then displays the count of the orders.

Creating calculated fields

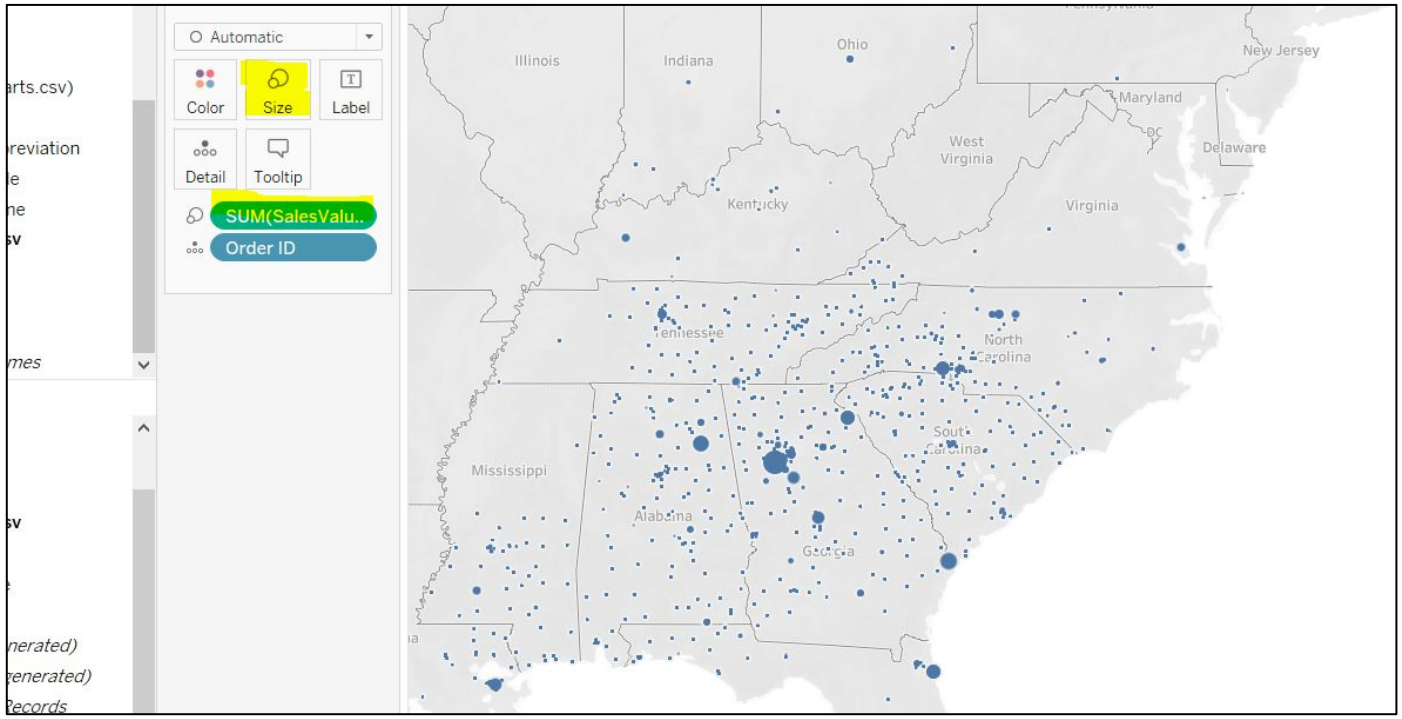
We want to know the total value of sales for each order, SalesValue.

To compute it, right click on any measure or dimension in the left “Data” sidebar, mouse over create, and select “Calculated Field”. Enter the following equation, click Apply, and see “SalesValue” pop up under Measures.

Displaying SalesValue

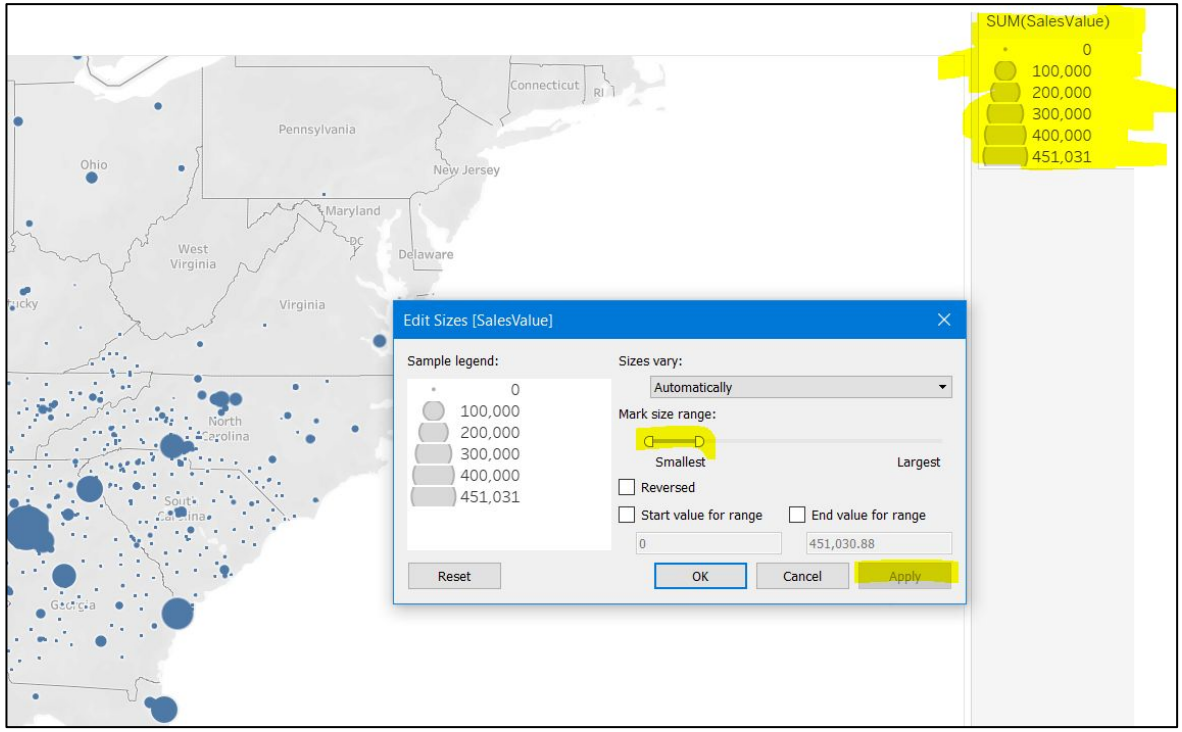
We want to quickly be able to see magnitudes of SalesValue by site.

Drag “SalesValue” into the Size box, and see the dots scale to size.



Editing the scale

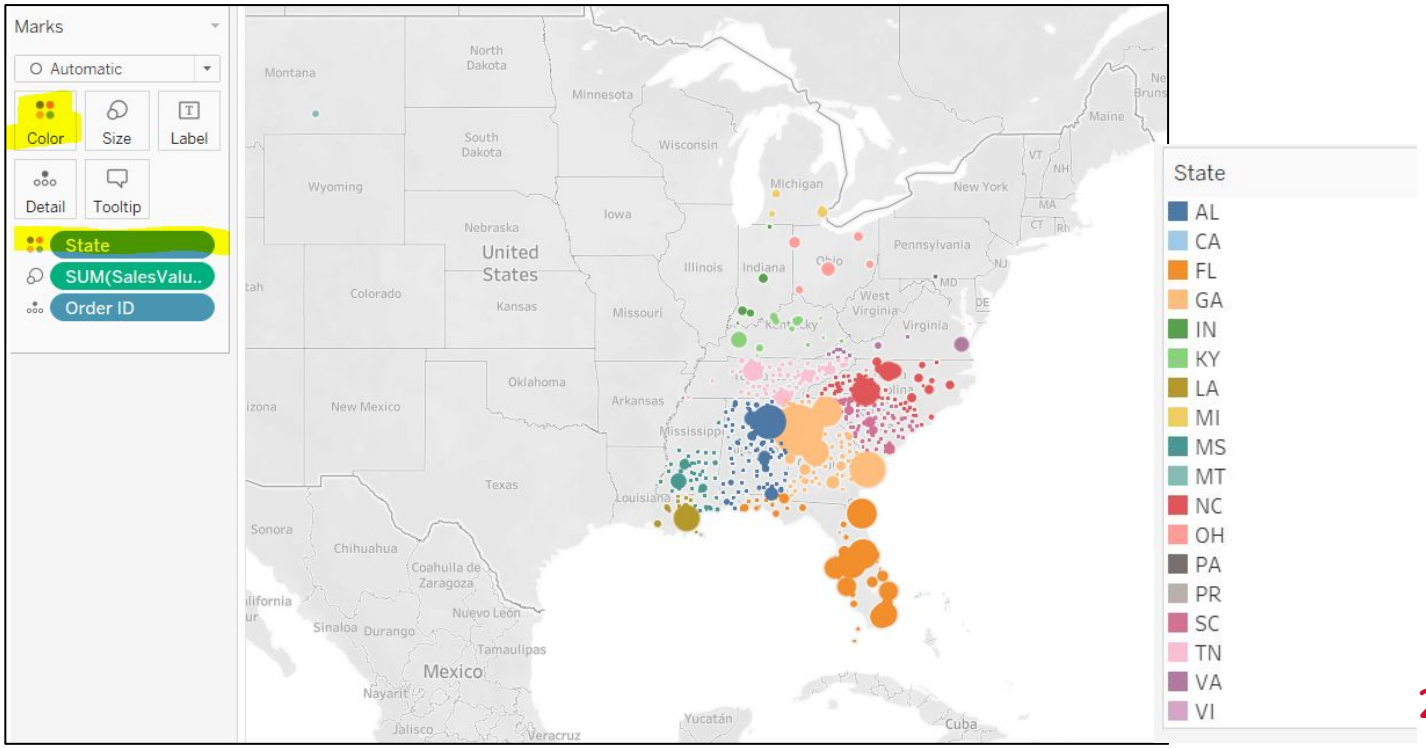
Right click the SUM(SalesValue) legend in the top right corner and click “Edit Sizes”. Drag the “Mark size range” slider and click apply to see how it impacts your map. Leave the slider at a spot you feel is appropriate.



Adding more visual cues - color

Drag “State” to the Color box to color each dot by state. Can you quickly tell which states have the highest sales volumes?

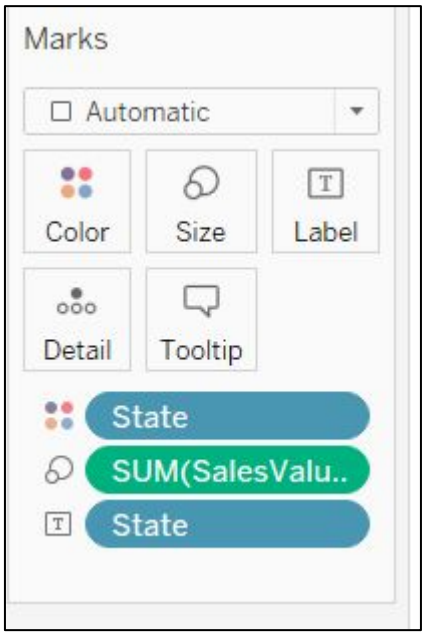
Screenshot your completed map.



Another visualization!

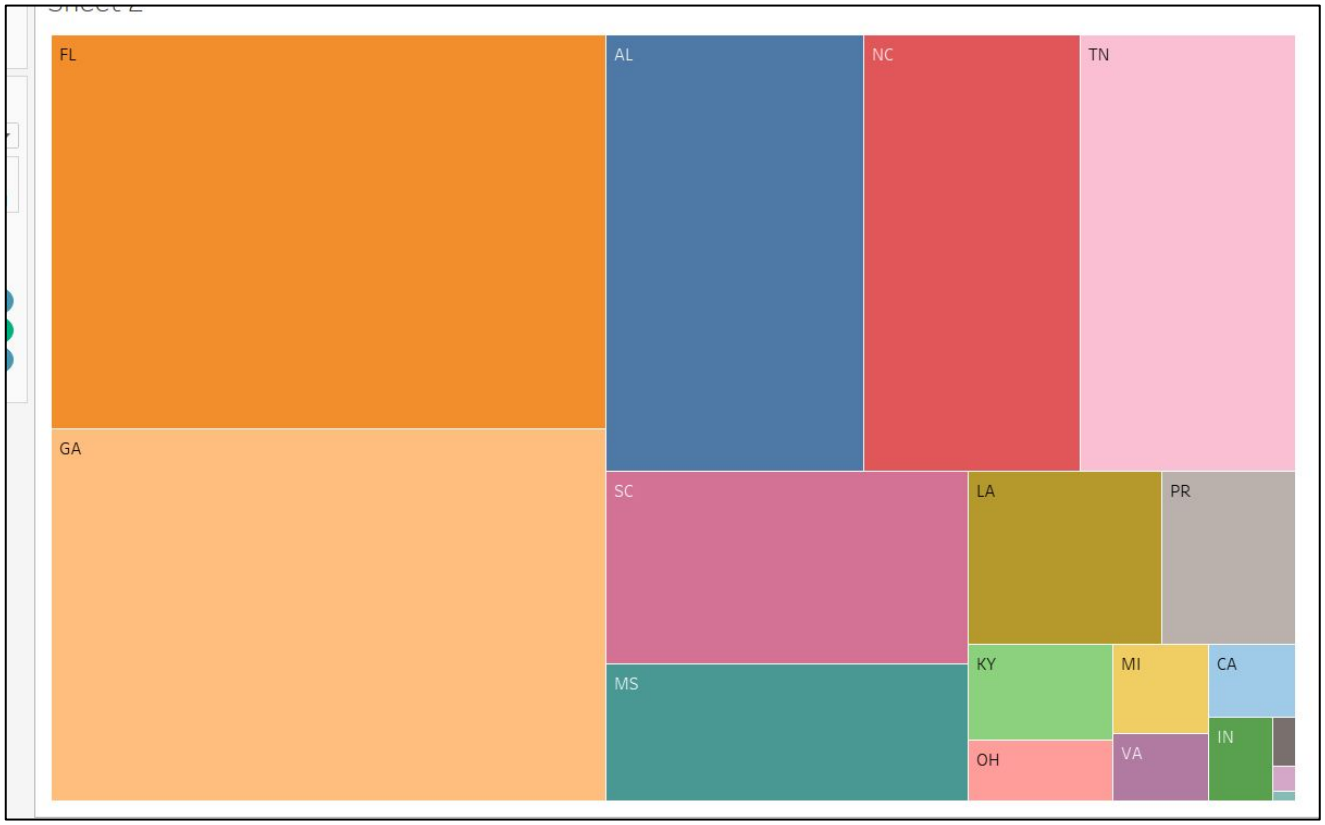
Now, we want to see the states by highest sales volume, but without a map. Click on the lower menu bar to open a new sheet.

Drag the following attributes into the correct boxes in the Mark tab:



Results

The resulting visualization should look like:



Does this match what you discovered in the first part of this lab? Screenshot your new visualization.

Application

Say you're running this business, and on June 16th of 2002, you had record sales in Georgia. To determine how to continue great sales, you might ask why a large volume of sales happened in the first place.

Data visualizations are here to help.

Start a new sheet here.

Approaching the question

We only want data from 2002 and Georgia. We should use the Filter. Drag “State Name” to Filter, right click, Edit Filter, and only check the box next to Georgia. Apply a filter to “Date” as well, and select a range that shows you the summer of 2002 (exact dates don’t matter).

We also know we want to look at SalesValue over time.

So drag SalesValue to rows and Date to columns. On Date’s drop down menu, change Year to Week Number (make sure you choose Week Number and not Week).

Looking for more information?

Currently, we don't see much (your visualization should look somewhat like the image below). So let's see what information we can add - the Marks options are great for this.

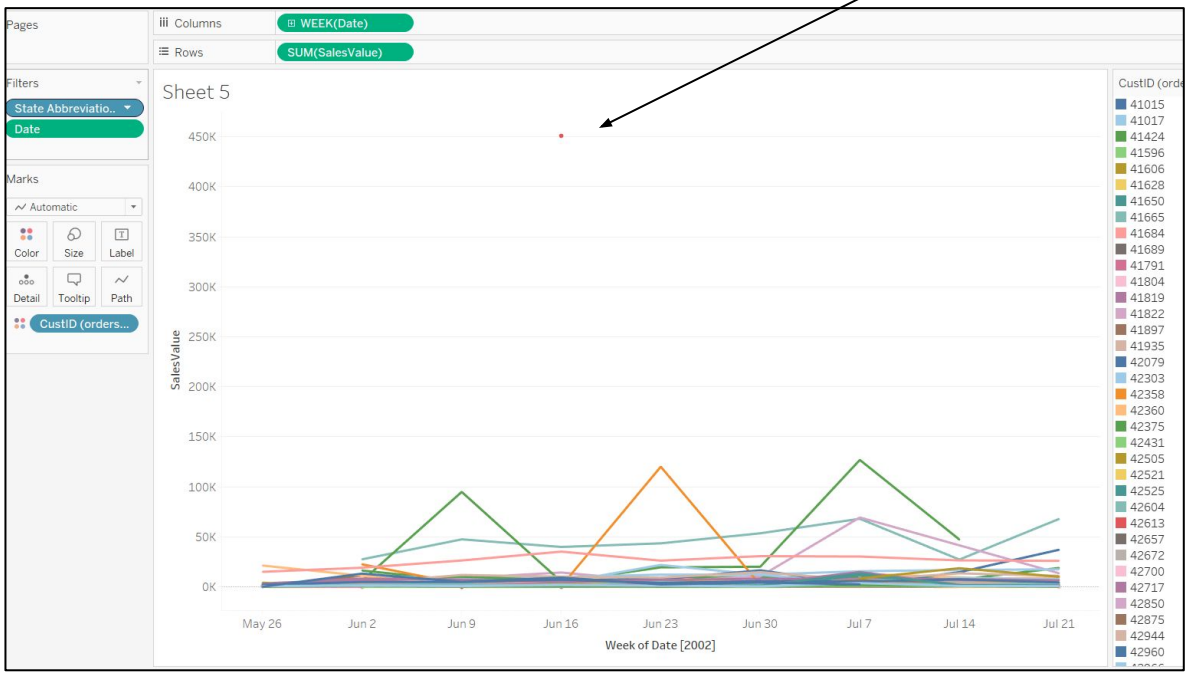
Start by dragging "CustID" from Dimensions (not Measures) to color.

If prompted, select "add all members".

Looking further…

We can see the chart now separated and colored by CustID. On June 16 - we see a single point much higher than the rest of our chart. While this may be hard to decipher, we can't write it off as nothing.

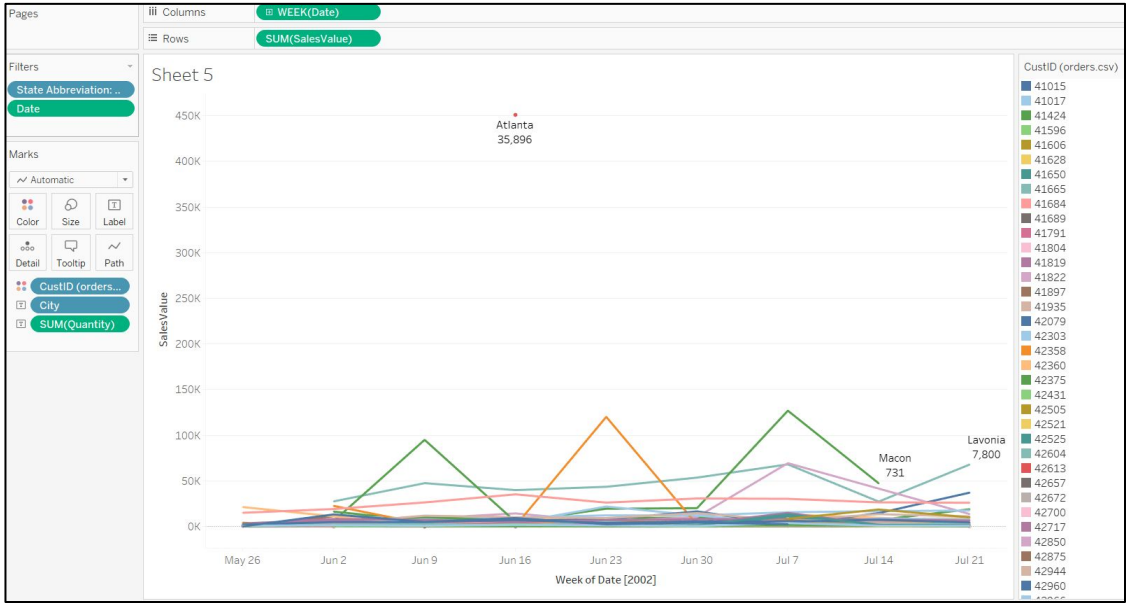
To find out more about this point - drag "city" and "quantity" the Label box.



Conclusions?

Now we can quickly read off the chart that on June 16th, Atlanta placed a one-time, high-quantity order. Click on “Atlanta” to ensure no other points highlight - then we know it is Atlanta’s only order for this date range.

This is as far as we will take this example, but think about what you would look at next to discover what is happening with Atlanta’s sales.

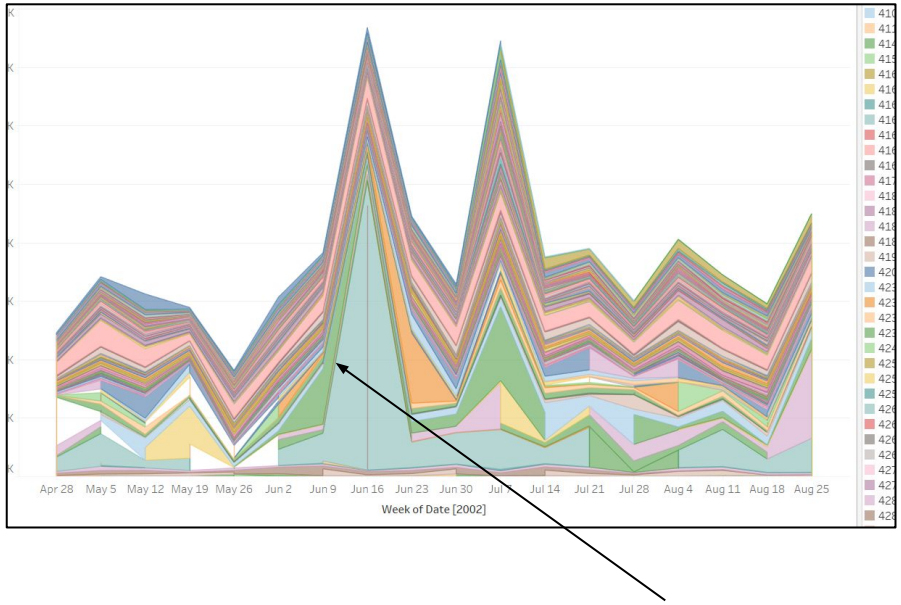


Screenshot this completed visualization.

Area versus Line Chart

Here we've recreated the line chart we made, except as an area chart, and pasted it below.

A line charts quickly highlighted the anomaly; however, the point can be easy to miss, as it appears small, and near the top.

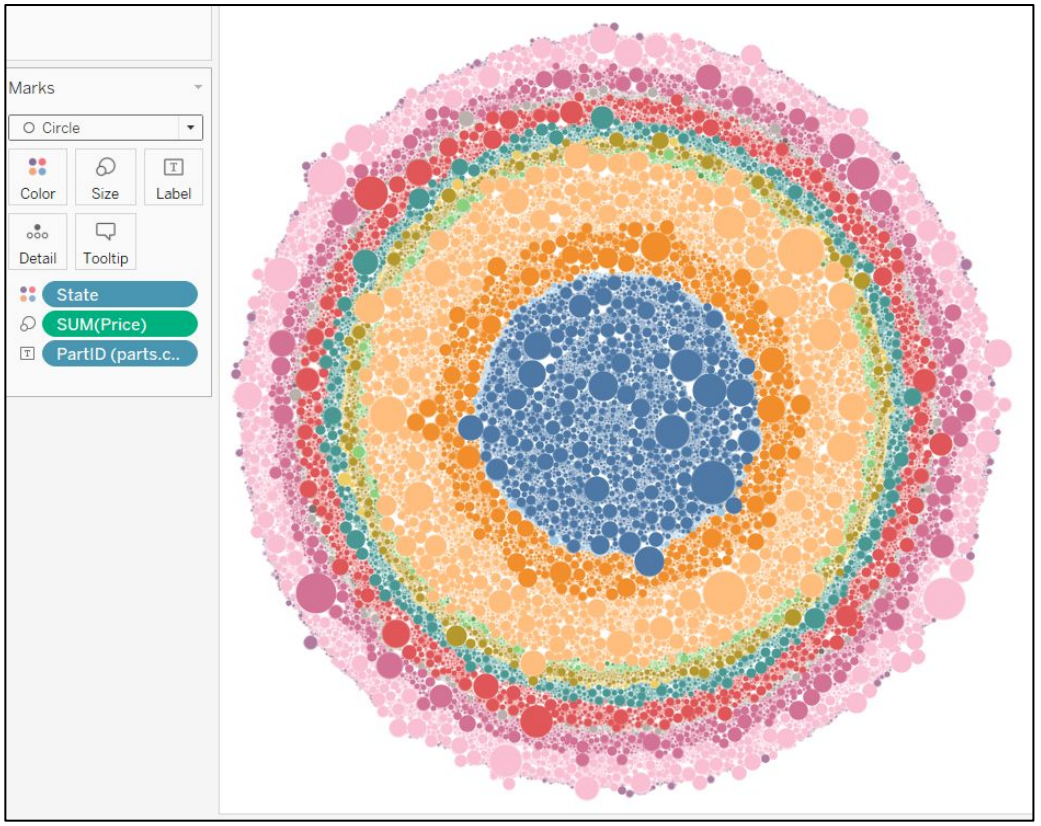


An area chart can point out the same observations. See the vertical line in the large teal area? Again, that should stand out to us. Further investigation would show you it is the same point as previous.

When visualizations turn to art

Look at the following visualization made from this lab's data set. While it is pretty, what does it tell us?

This course does not cover the fine details of visualization creation. But remember that proper labeling and logic - why is each attribute where it is, what are calculated values telling you - are the foundations of a useful visualization.

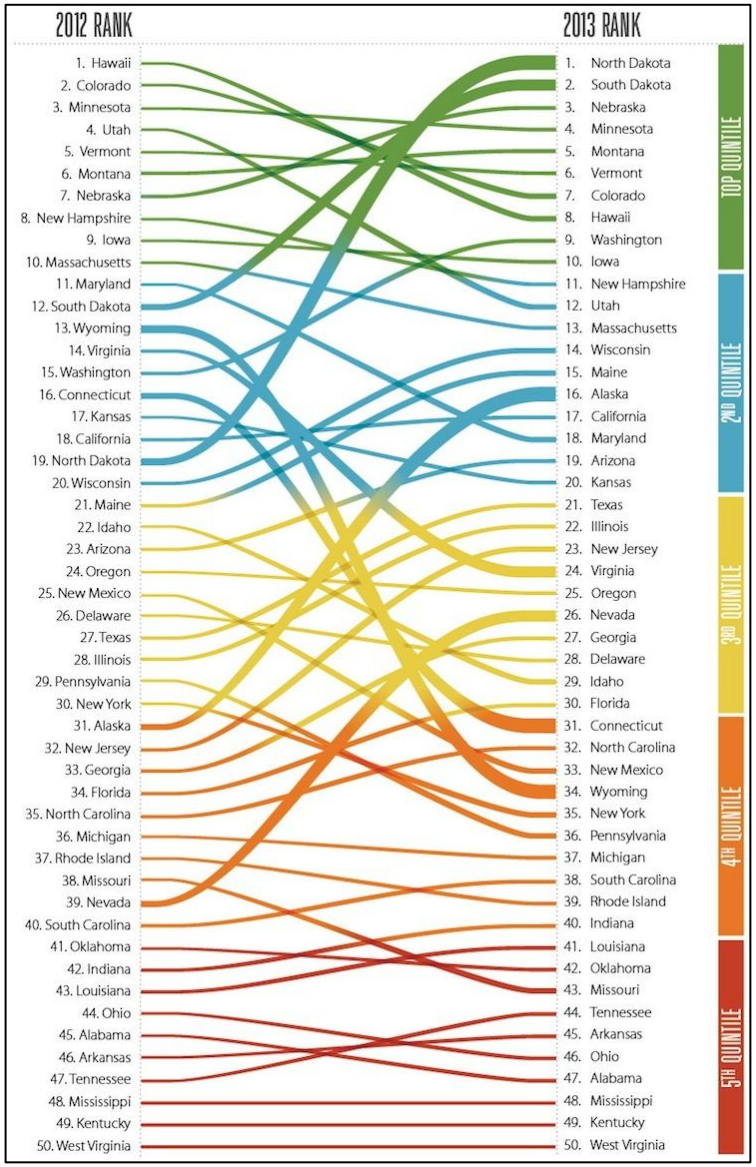


feel free to make this interesting visual with the specifications in the photo above if you'd like!

And more…

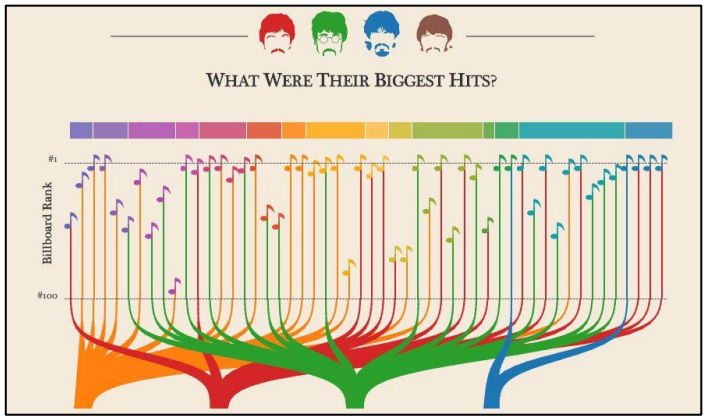
This was a <u>brief</u> introduction to Tableau.

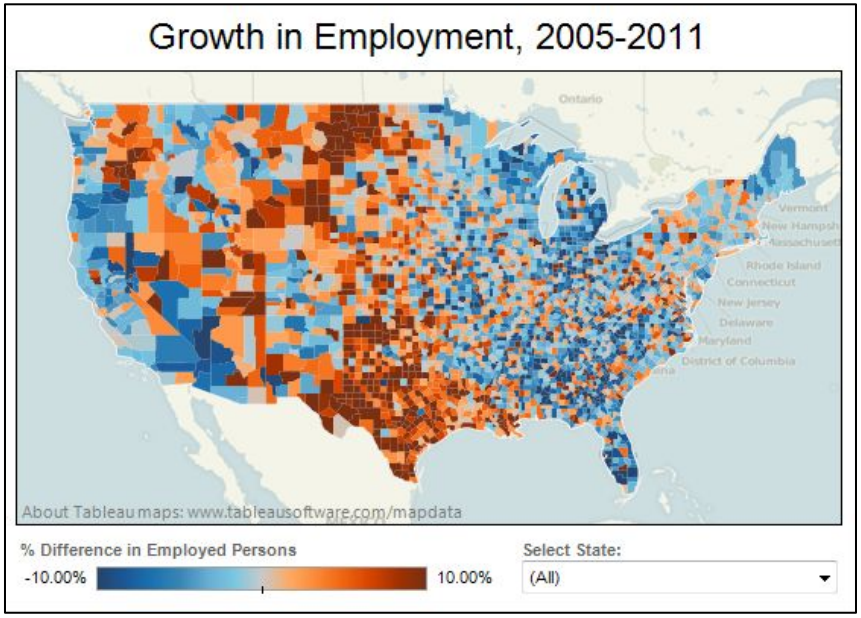
You can do incredible things with tableau like create a map of a states "happiness" ranking across years…

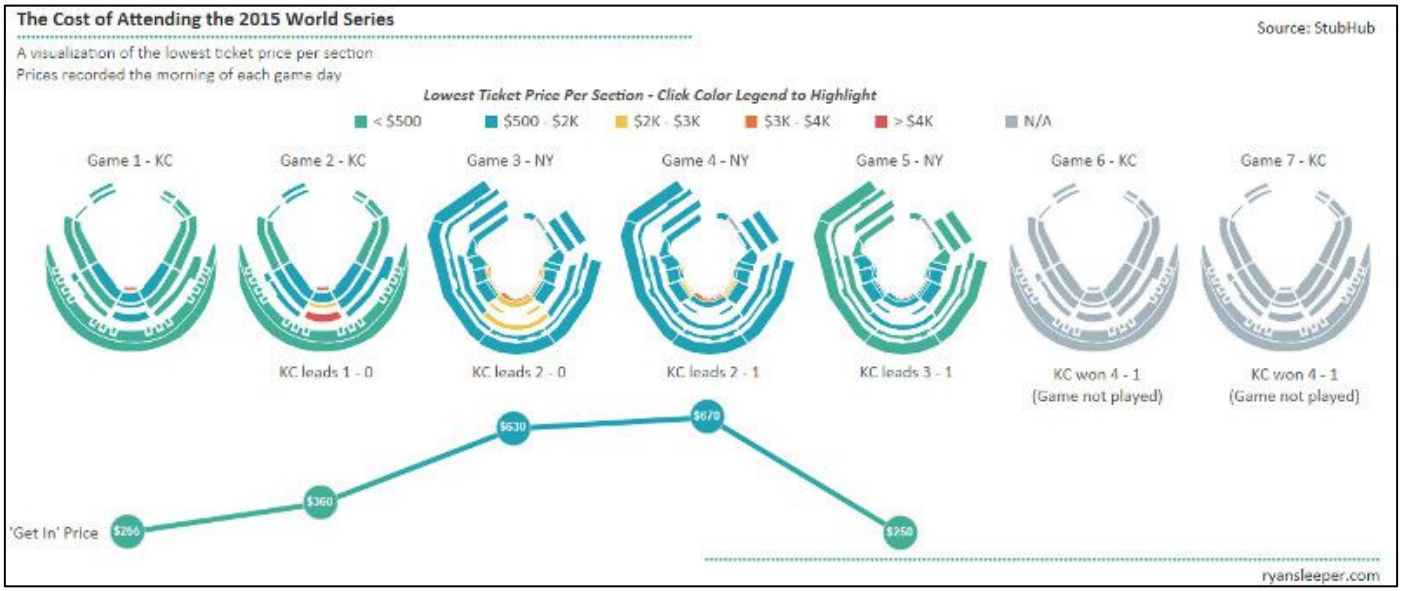


And…

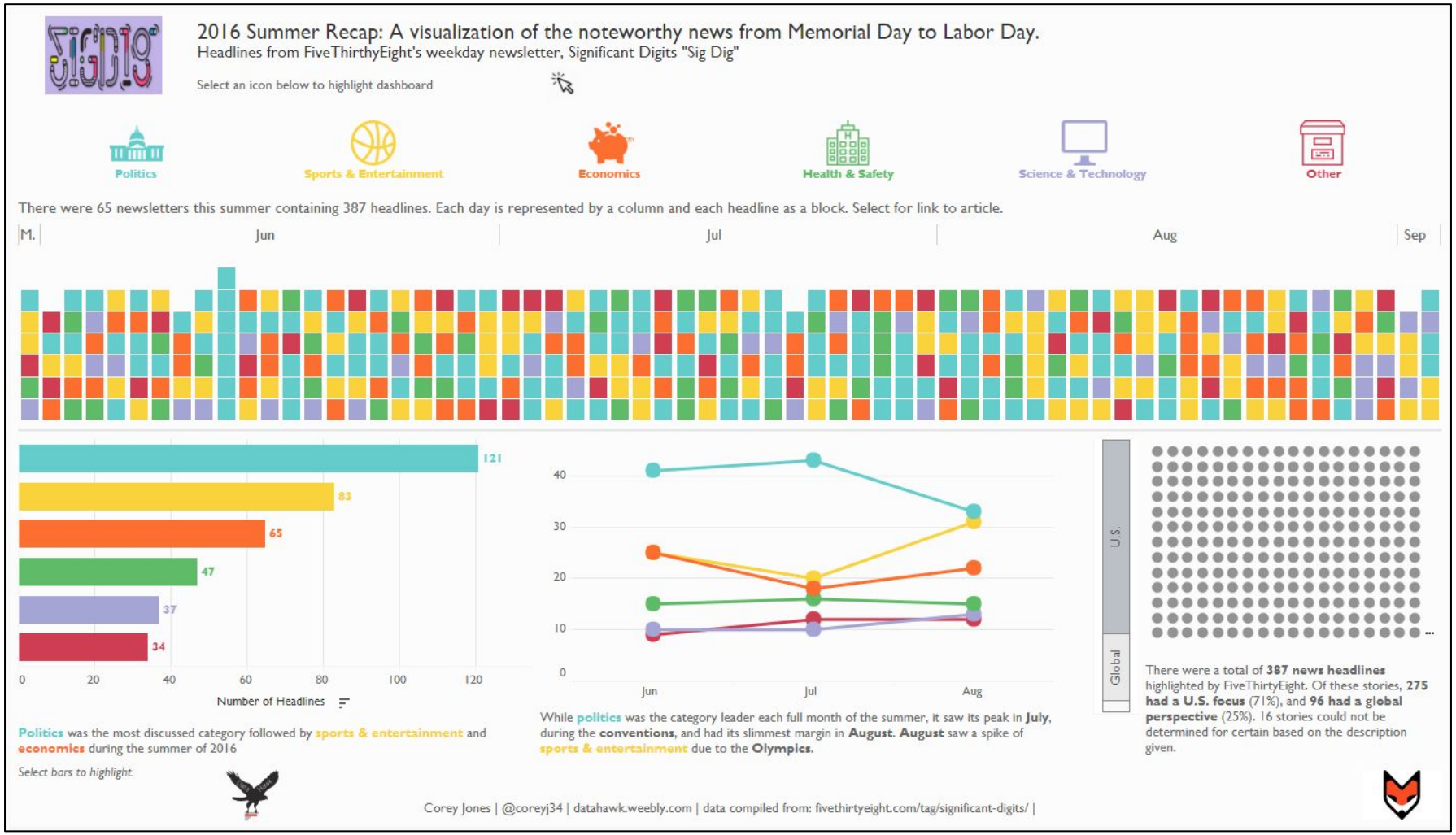
…visualizations such as:







Or entire dashboards such as:



Resources

There are many online resources and forums for Tableau communities - a great starting place is:

https://www.tableau.com/learn