20.1 Lesson Plan - Time to Tableau Some Minds

4. Instructor Do: Loading and Exploring Data (0:05)

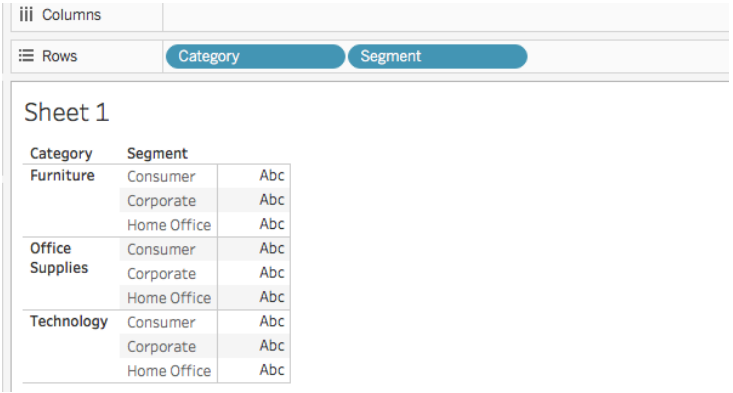
Tableau does not allow its users to change the values contained within the cells of a dataset

Filtering data is very simple, however, as all Tableau users need to do is click on the "Add" button beneath the `Filters` text in the top-right corner of the application and select what column they would like to filter by.

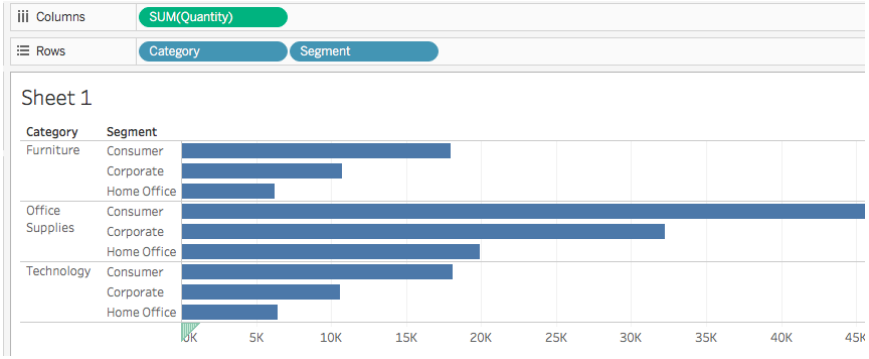
After selecting which column to filter by, the values to filter are then chosen manually or based upon some kind of condition.

5. Instructor Do: Building Basic Visuals (0:10)

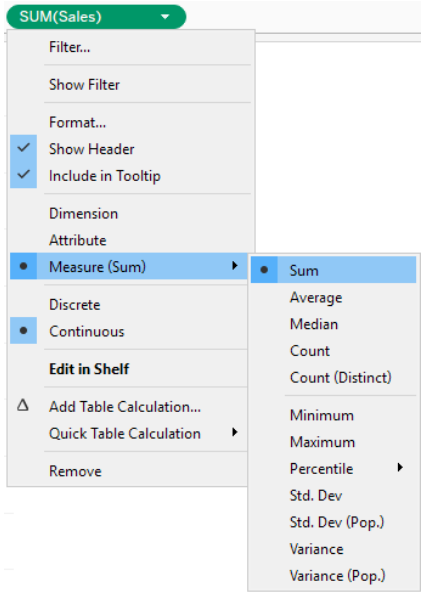
- Create a new Tableau project and connect it with the GlobalSuperstoreOrders2016.xlsx file provided.
- Drag the `Orders` sheet into the main area.
- Place Category in Rows
- Drag `Segment` into `Rows` and placing it after the `Category` pill. Now each category within the visualization has been split into three distinct parts.



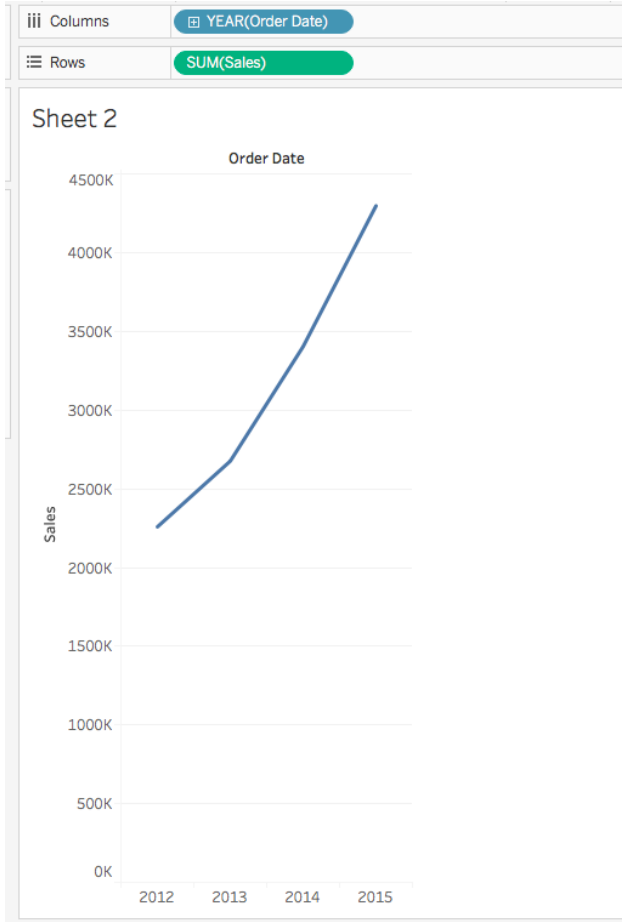
- From measures place SUM(QUANTITY)



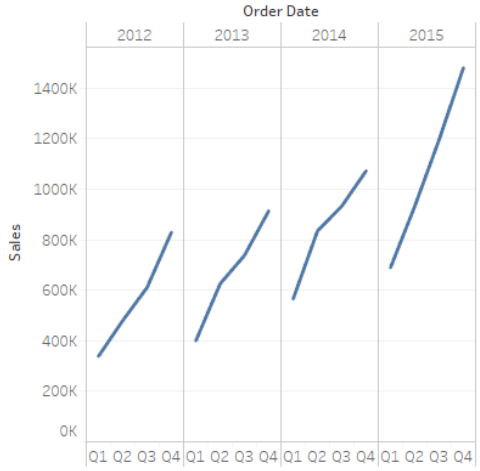
- The chart can then be made more detailed by adding more elements. By adding `Market` into `Columns` , for example, multiple charts are created to show the quantity of orders per segment per category within each geographic market.
- Create a new worksheet within Tableau.
- Drag `Sales` into the `Rows` section.
  - Point out that a bar chart was created that visualizes the total sales made by the company in question. This is because the `Sales` pill is being measured by its sum by default.
- The type of calculation performed on a `Measures` pill can be changed by clicking on the pill, selecting "Measure" from the drop-down menu, and then picking one of the calculation types present.



- Now drag `Order Date` into the `Columns` field to create a very basic line chart.

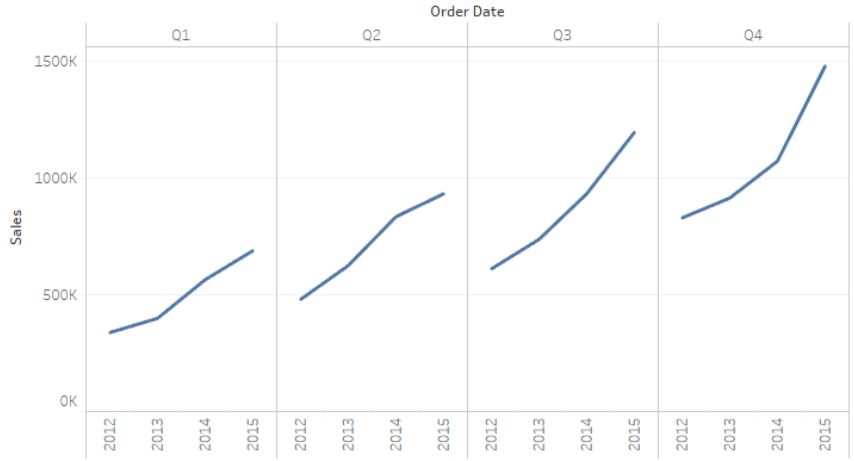


- Point out to the class how Tableau has aggregated the dates at the year level. In order to expand this to include quarters, simply click on the plus symbol within the `YEAR` pill.



- Explain that Tableau, by default, visualizes at the **least** granular level. In this case, it displays the yearly results by default.

- In order to compare how Q1 has performed over the years, simply move the `QUARTER` pill before `YEAR`.



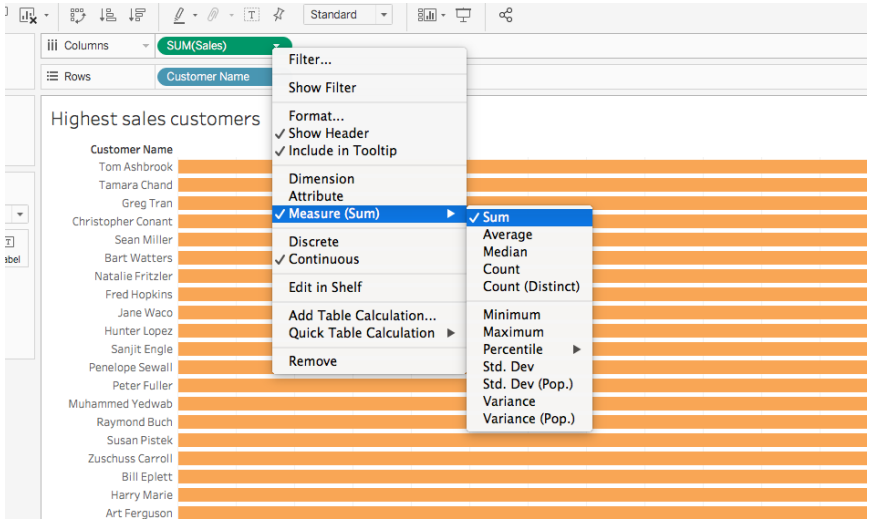
- Explain to the class how the best way to learn Tableau is to dive into the application and test it out manually.

6. Students Do: Explore Data (0:15)

- In this activity, students will be given visualizations, which they will attempt to recreate using Tableau.
- **Instructions**:
  - Readme.pdf

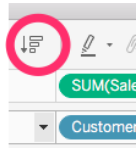
7. Instructor Do: Review Activity (0:10)

- **File**: Activities/03-Stu_Exploration/Solved/sales.twbx
- The first visualization, of the customers with the highest sales, requires dragging the `Customer Name` pill to Rows, and the `Sales` pill to Columns.

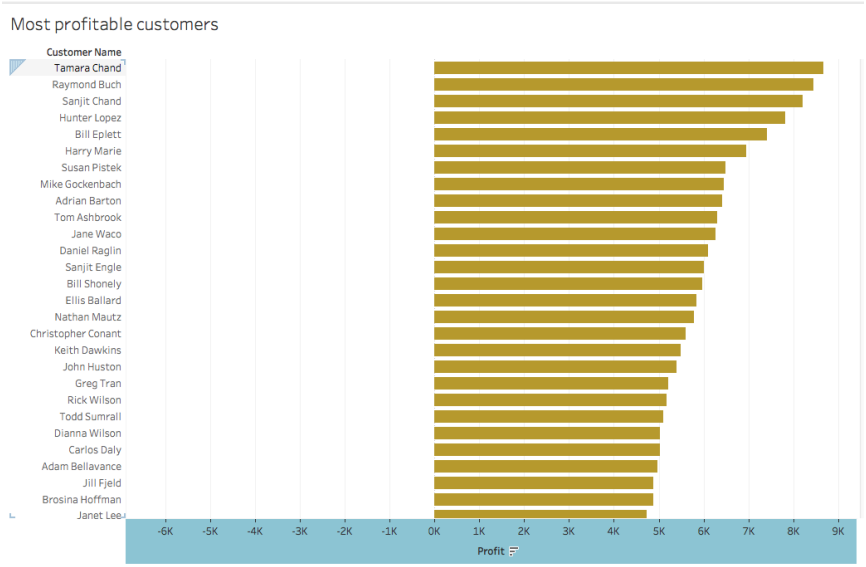


  - Clicking on the arrow on the `Sales` pill, selecting `Measure`, then `Sum` aggregates the data into a sum.

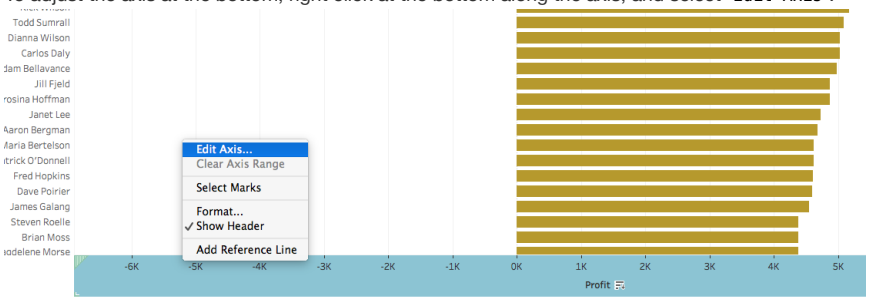
- To sort the data, click on the sort button:



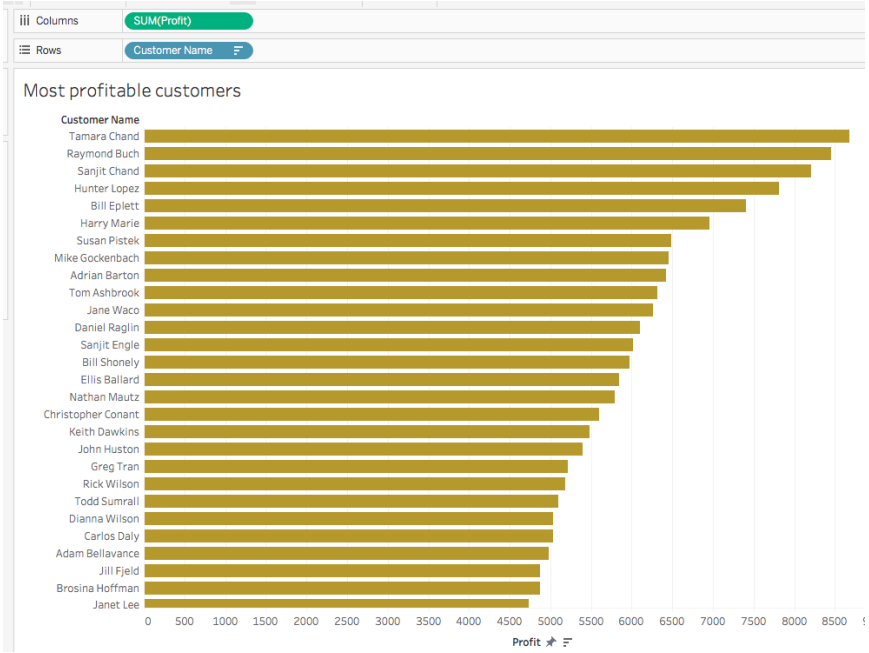
- In the next tab, in order to chart the most profitable customers, simply do the same as above, this time with the `Profit` pill:



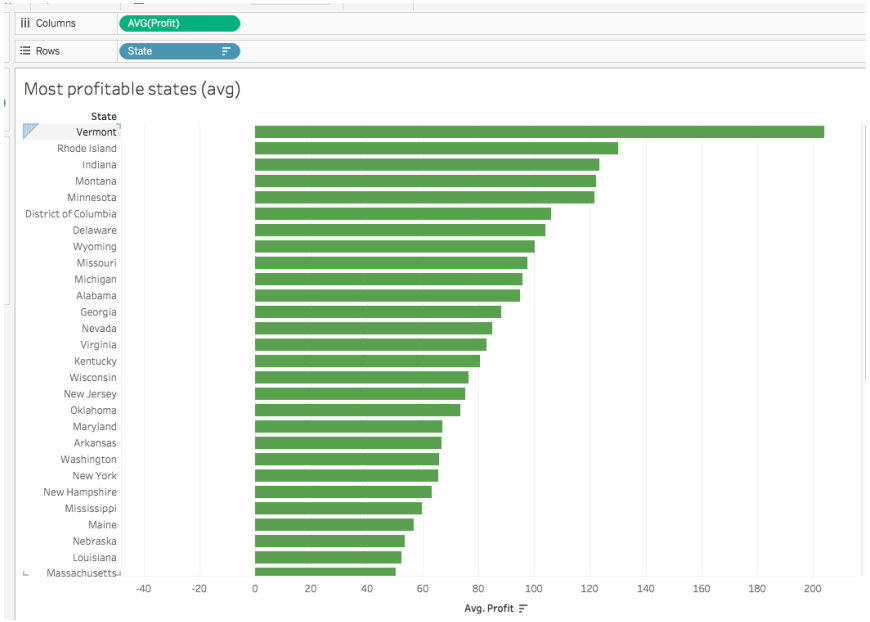
- To adjust the axis at the bottom, right click at the bottom along the axis, and select `Edit Axis` :

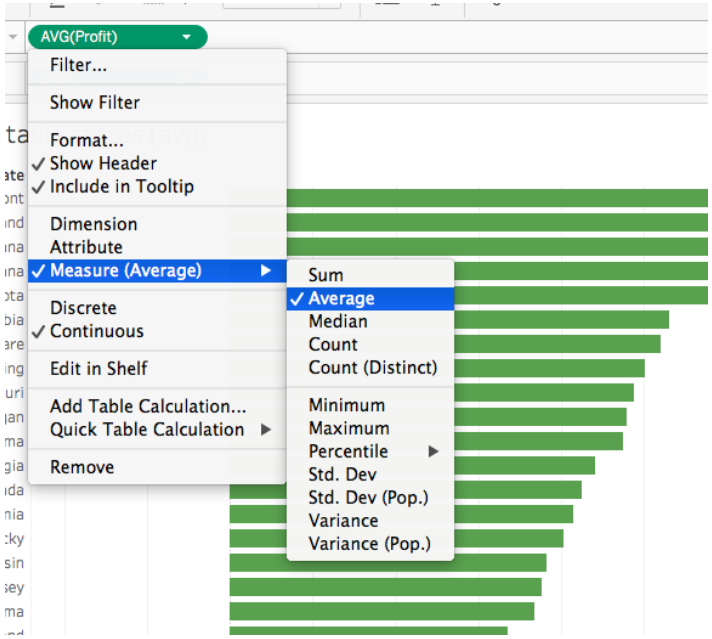


- After filtering out negative profit figures, the chart should now look like this:

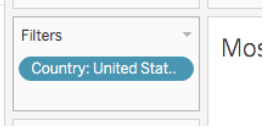


- And to chart the states with the highest average profit, choose the `Profit` pill again, then `Average` under `Measure` :

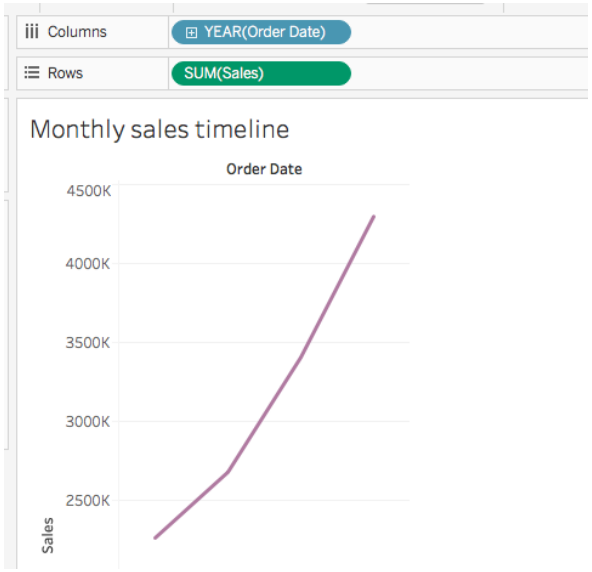




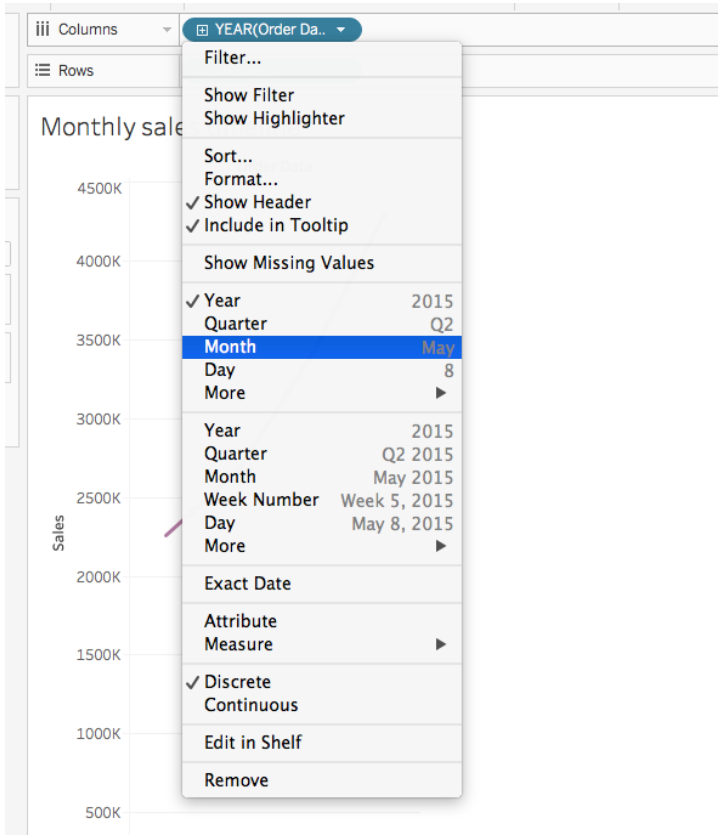
- The filter should be set to the United States:

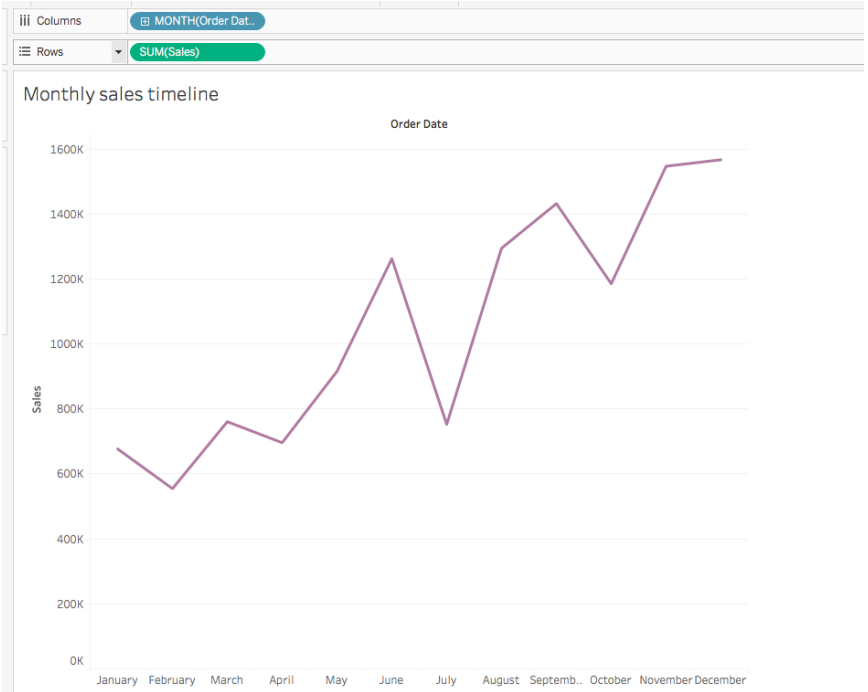


- Next, to display a monthly timeline, drag the `Sales` pill to Rows, then obtain its sum. Then drag `Order Date` to columns.

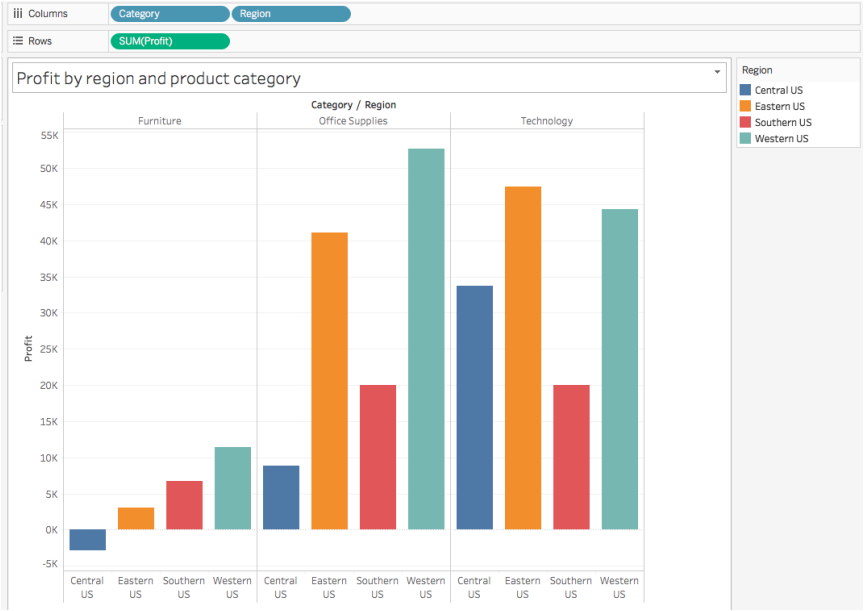


- This, however, is not what we want. There are several ways to drill down to the monthly level. One is to click on the `+` symbol on the `YEAR (Order Date)` pill, then getting rid of undesired levels. The other is to click on the arrow in the pill and select `Month` :





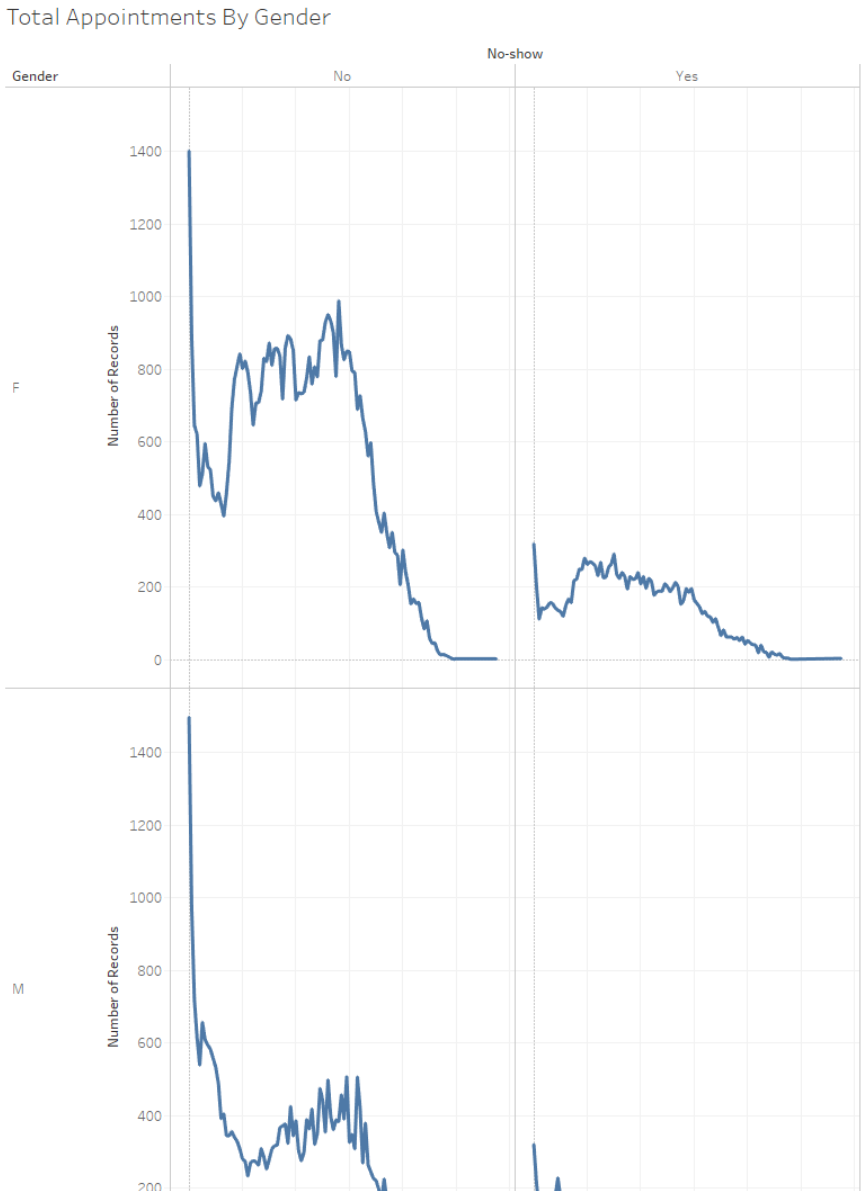
- Finally, to visualize profit by region and product category, drag `Category` and `Region` pills to Columns, and create a sum of the `Profit` pill in Rows:

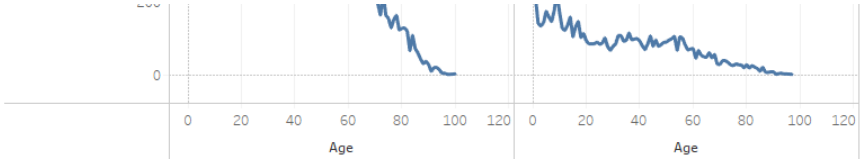


- The side-by-side bar chart can be created using the `Show Me` option:

8. Students Do: No Shows (0:20)

- Students will now spend some time creating a series of visualizations that will answer some questions as to what kinds of people are more/less likely to show up to doctor's appointments.

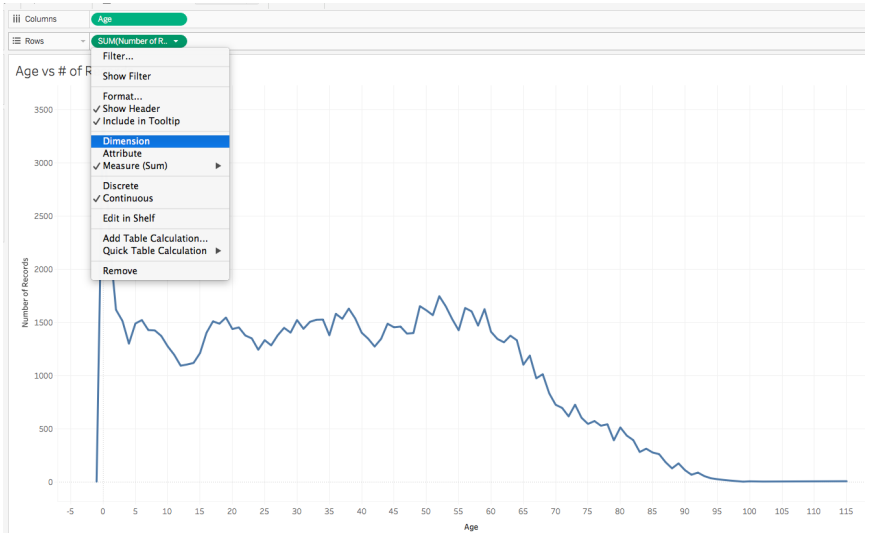




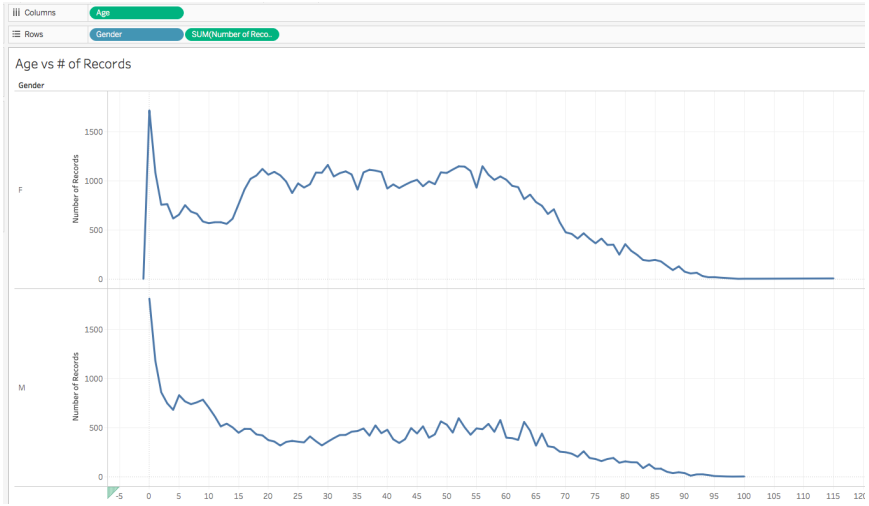
- **File:**
  - no_shows.csv
- **Instructions:**
  - Create a line chart that compares the ages of patients against the total number of appointments. Then split this graph based upon gender and whether the patient showed up to their appointment. For this first step, you'll need to convert `Age` from a measure to a dimension.
  - Create a pair of bar charts that compare how many patients showed up to appointments versus how many were no-shows in different neighborhoods.
  - Create a stacked bar chart that compares no-shows to those who made it to appointment based upon the day of the week.
  - Create a pair of line graphs that compare age versus diabetes in both men and women.
  - Create a pair of line graphs that compare age versus alcoholism in both men and women.
- **Bonus:**
  - Figure out how to create filters and manually exclude non-significant values from your charts using the `Filters` panel.
  - Now, using filters, modify your charts so that they are more visually understandable.

9. Everyone Do: No Shows Review (0:10)

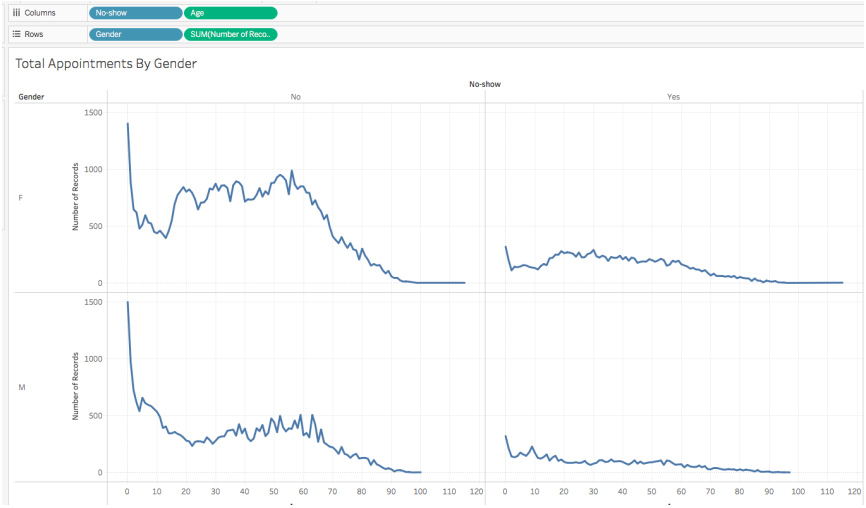
- Open up 04-Stu_NoShow within Tableau and walk through the application with the class, answering whatever questions students may have.
- The first step for this activity is to drag `Age` to Columns, and `Number of Records` to Rows. `Age` must also be converted from measure into dimension by clicking on the arrow on the pill.



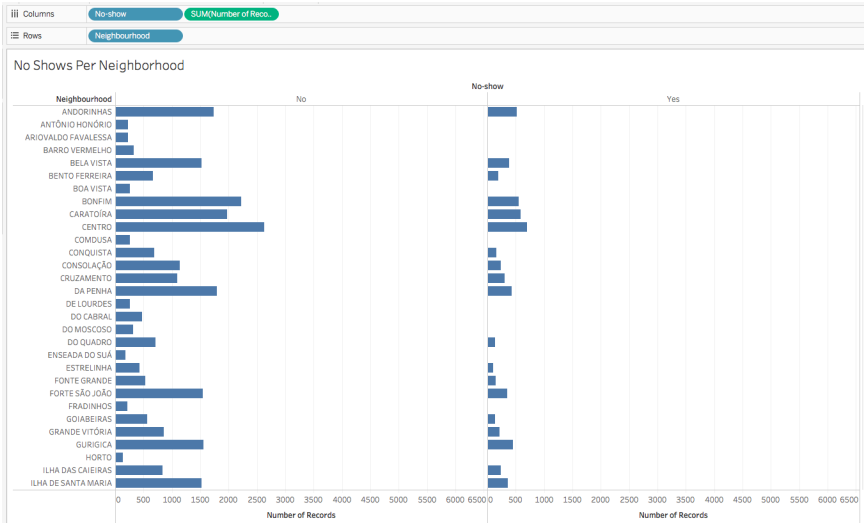
- To split up the results by gender, drag `Gender` into Rows:



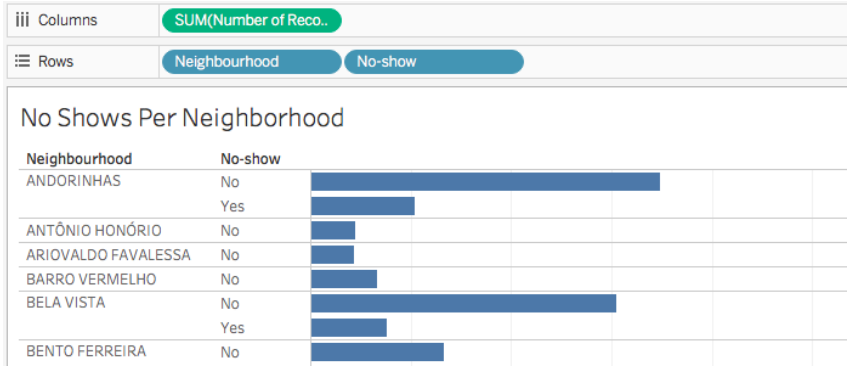
  - Male patients, overall, seem to have fewer appointments across age groups!
- Finally, to stratify the results by no-show appointments, drag `No-show` to columns:



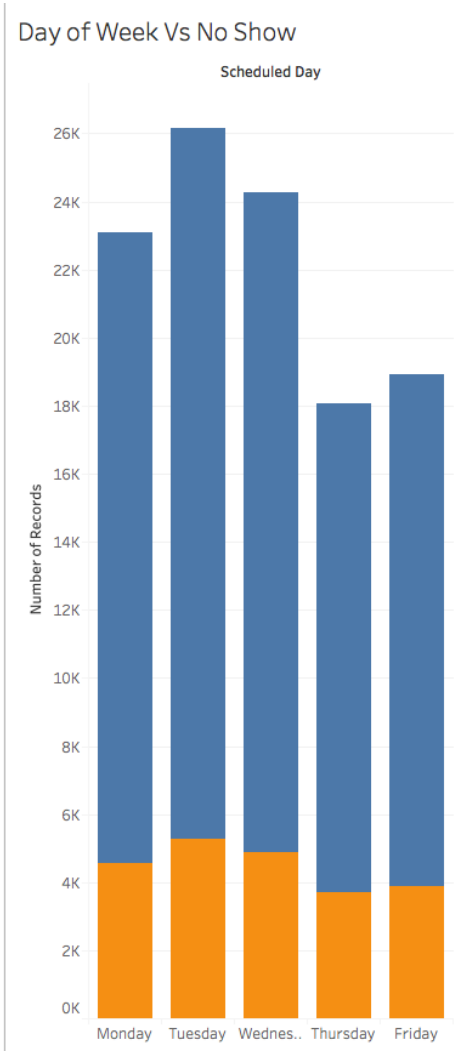
- In the next visualization, students were asked to compare no-shows by neighborhood. This can be done in the following way:



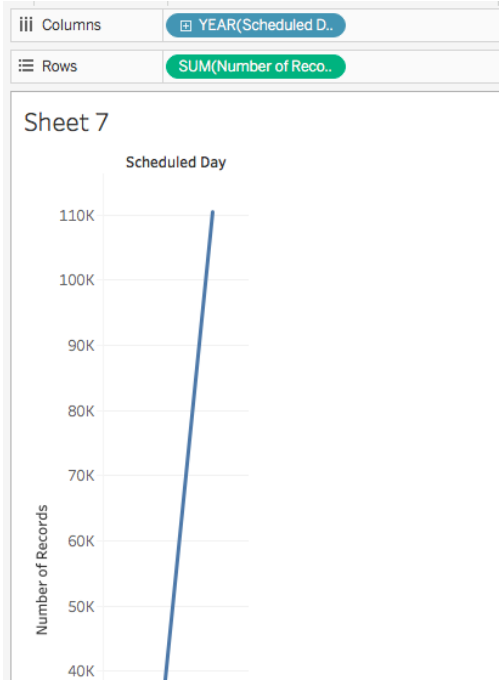
  - `No-show` and `Number of Records` are dragged to Columns, and `Neighbourhood` to Rows.
- It can also be visualized thus:



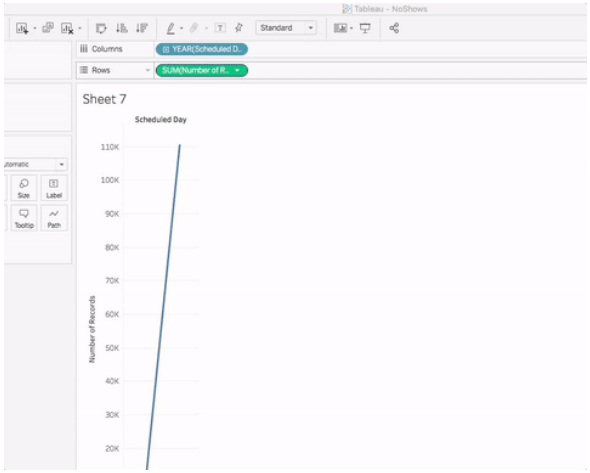
  - `No-show` is moved to Rows instead of Columns.
- In the next visualization, students were asked to visualize the number of no-show patients by the day of the week:



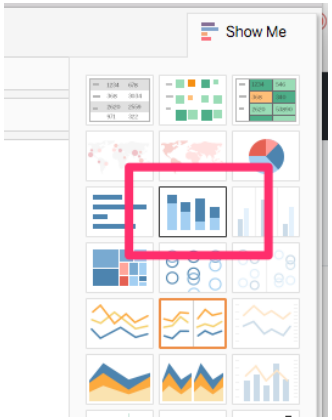
- Since we're counting the number of no-show appointments, it makes sense to to drag `No-show` to Rows, and visualize this measure vertically. And since we're tallying the number of no-shows by the day of the week, to drag `Scheduled Day` into Columns:



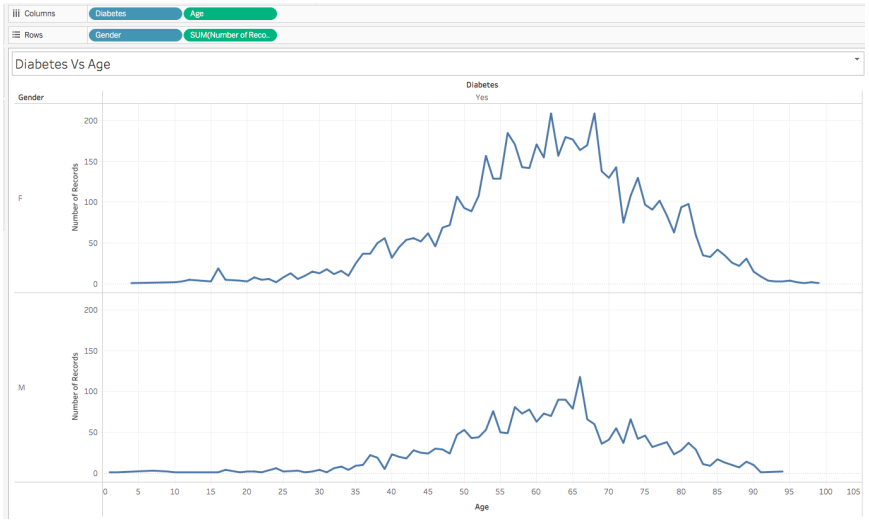
- However, this isn't quite what we want. We're shown results by year, instead of the day of the week. This can be selected by clicking on the arrow on the `Scheduled Day` pill, More, then Weekday.



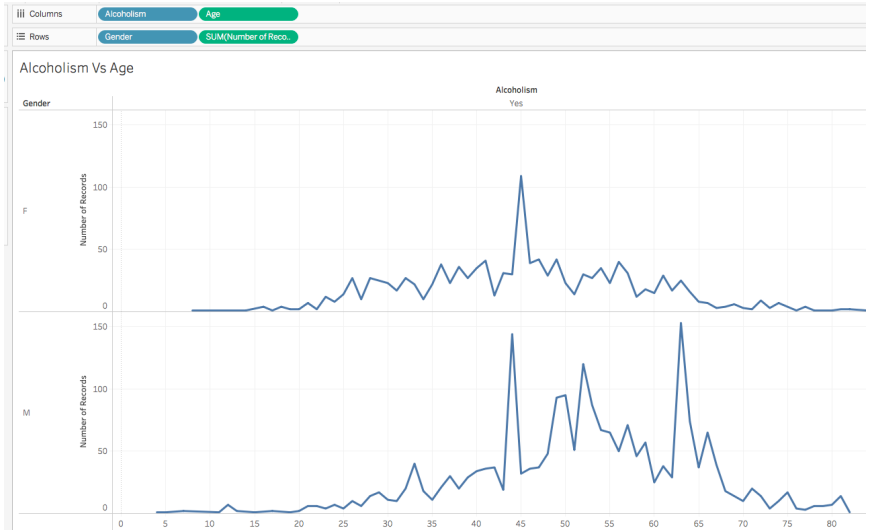
- To display a bar chart instead of a line chart, select `Show Me`, then the stacked bar chart option:



- In the next visualization, students are asked to display the number of diabetics by gender and across age groups. One way to visualize this is by stacking `Gender` in Rows.

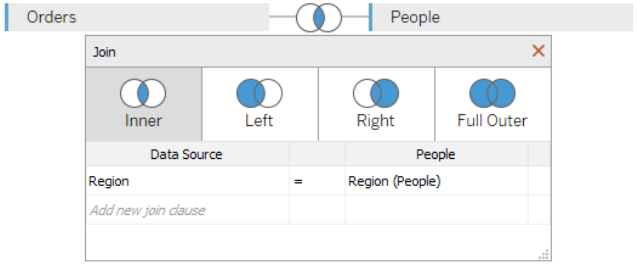


- The final visualization is very similar to the previous one, visualizing alcoholism instead of diabetes.

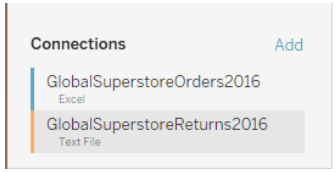


10. Instructor Do: Joins and Splitting Made Easy (0:05)

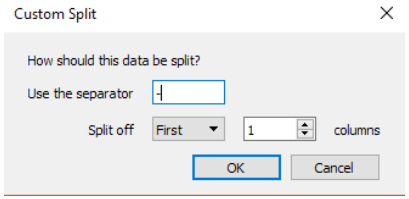
- Joins are an inescapable aspect of data science and are often thought to be both tedious and complex. Tableau, however, trivializes joins to such a degree that even complex joins can be performed in just a few clicks.
- Open up GlobalSuperstoreOrders2016.xlsx one more time within Tableau and drag the "Orders" sheet into the main area.
    - In order to merge these two datasets together, click and drag the "People" sheet into to main area of Tableau alongside the "Orders" sheet.
    - Tableau will automatically create an inner join on the columns that contain matching values. In this case, the join is on the "Region" columns.
    - To change what type of join is used, simply click on the interlacing circles at the top of the application and select what form of join to use from the menu that appears. This same menu can be used to modify what columns to merge on.



- It is also possible to create joins across data sources.
    - To do this, click on the "Add" button in the `Connections` panel and add the secondary data source desired. For the purposes of this demonstration, that is GlobalSuperstoreReturns2016.csv.



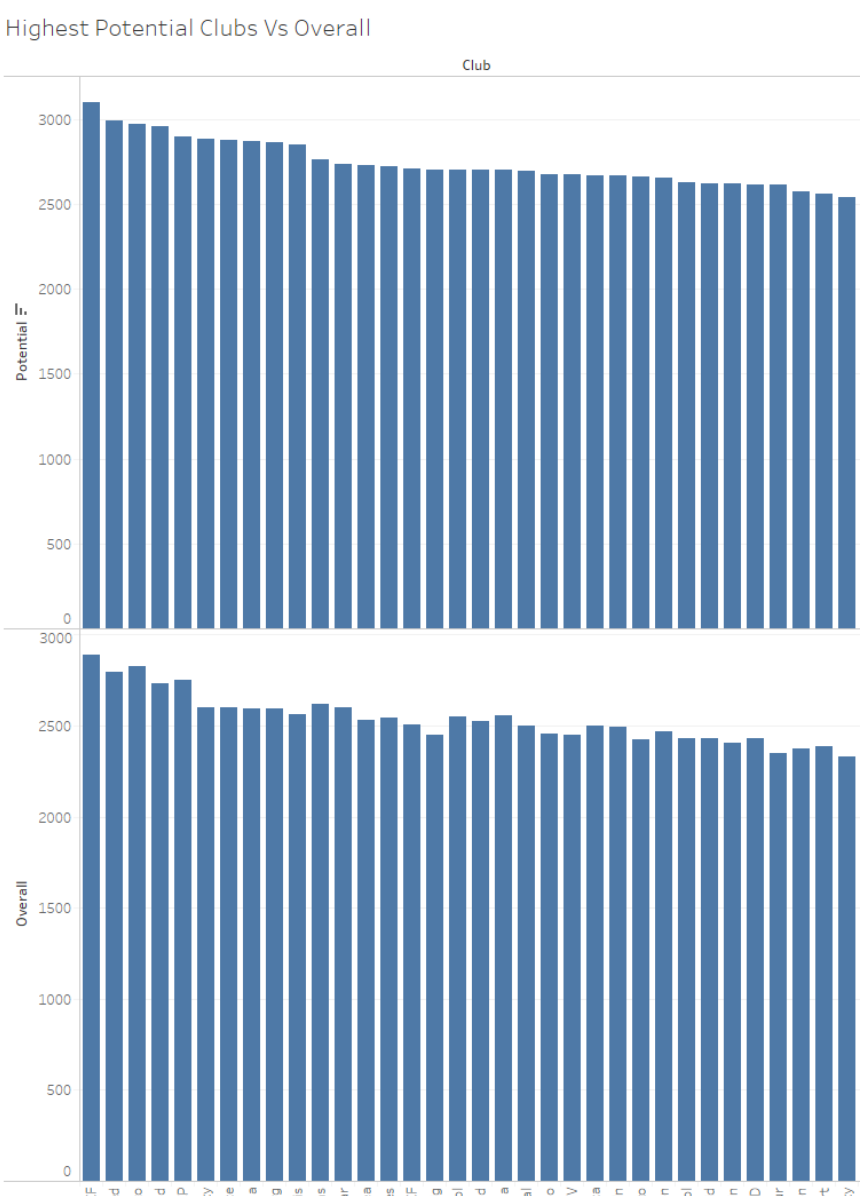
    - After the data source has been added, it can then be joined with the other data files desired using the method mentioned before.
- Another interesting feature of Tableau is that columns containing text can be split so as to extract data.
    - To do this, select the column header whose values should be split, right click, and select "Custom Split" from the drop down menu.
    - Select what character to split the text on, whether to split from the beginning or end of the string, and then how many times the text should be split.
    - Show this off by splitting the "Order ID" column on the first hyphen one time. This will extract the state in which a sale was made from the initial string.



    - New columns created this way can then be used when creating visualizations later on.

11. Students Do: FIFA Analysis (0:20)

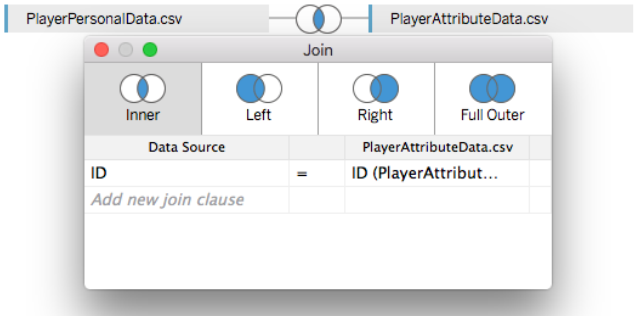
- Students will now create some tables based upon FIFA video game's player datasets. This will require them to merge multiple data sources together and then create visualizations off of the newly made dataset.



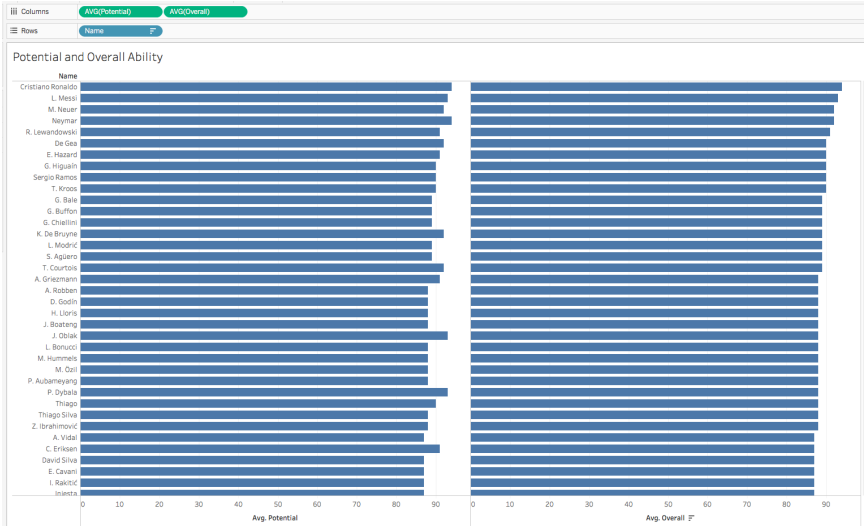
- **Files:**
  - PlayerPersonalData.csv
  - PlayerAttributeData.csv
  - PlayerPlayingPositionData.csv
- **Instructions:**
  - Create a join between each of the charts so that each player's data is matched up correctly.
  - Create a pair of charts that compare the potential of a club's players to their overall ability (`Overall` column). Then sort them from best to worst.
  - Create a chart that determines which soccer club is the most aggressive overall.
  - Create a chart that determines which nationality has the greatest accelaration on average, making sure to note how many players are from each nation in a second chart.
  - Create a chart that determines which nationality has the greatest long passing on average.
  - Create a chart that marks the potential of a player over time as they age.

12. Everyone Do: FIFA Analysis Review (0:10)

- Open up 06-Stu_FIFAPlayers within Tableau and walk through the application with the class.
- In order to join the two CSVs, drag them to the main pane in the `Data Source` tab, then, select an inner join:

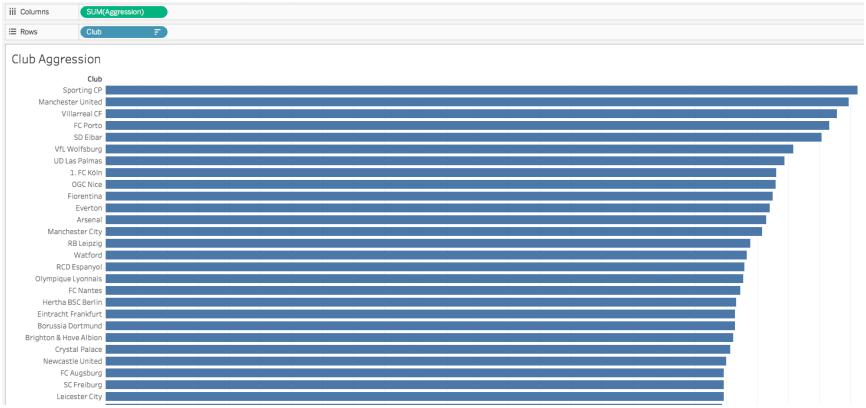


- The first visualization is of each player's potential, as well as overall ability, sorted in descending order:



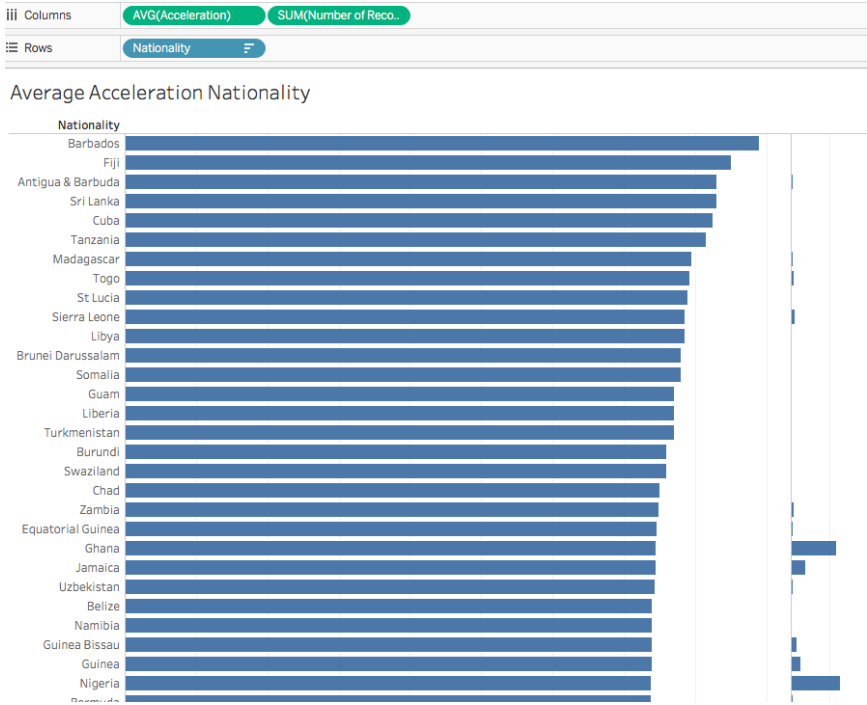
- The `Potential` and `Overall` pills are dragged to Columns, and `Names` to Rows.
- By default, players' potential and overall ability values be aggregated as sums, and will therefore exceed 100 for players with multiple rows. To correct for this, click on the pills, and from `Measure` choose `Average`.
- The results are sorted in descending order. This must be done for each chart.

- The second visualization tallies the `Aggression` of each club.



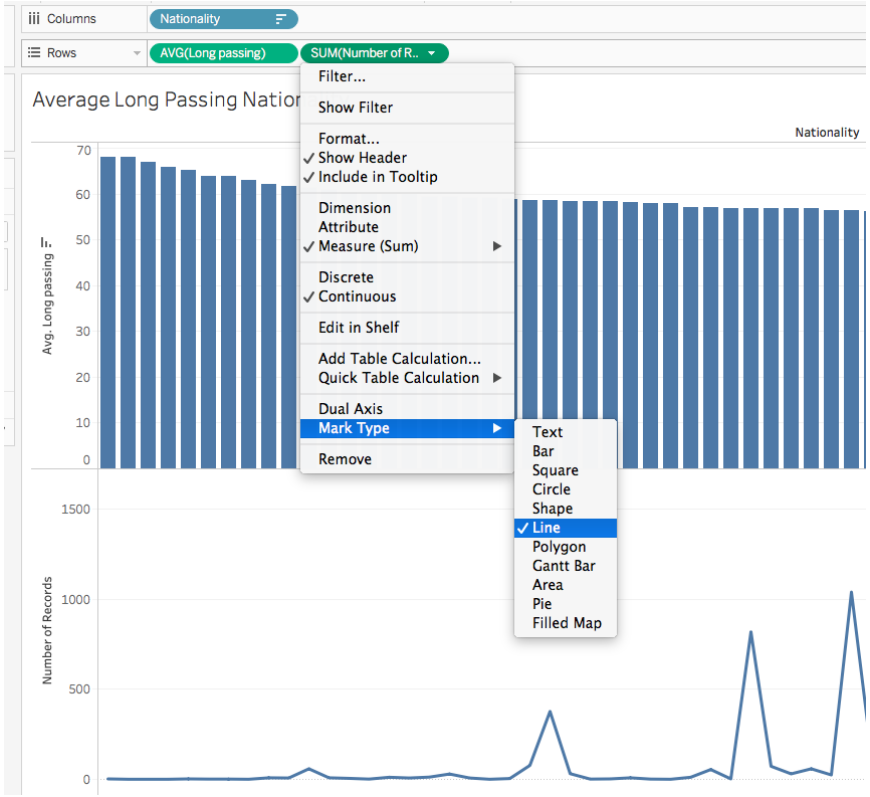
- This chart is simply an aggregation of the `Aggression` column, displayed by clubs in Rows.
- In this case, it makes sense to aggregate the sum of aggression, comprising the total aggression ratings of all the players in a club.

- The next visualization is of average acceleration by country, as well as the number of records from each country.



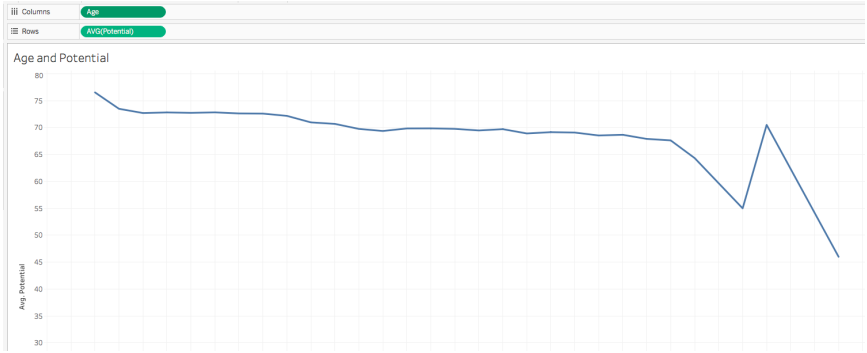
- `Acceleration` is plotted against `Nationality` in the left chart.
- In the right chart, the total `Number of Records` is plotted against `Nationality`.

- The next visualization is of average long passing by country, as well as the number of players from that country.

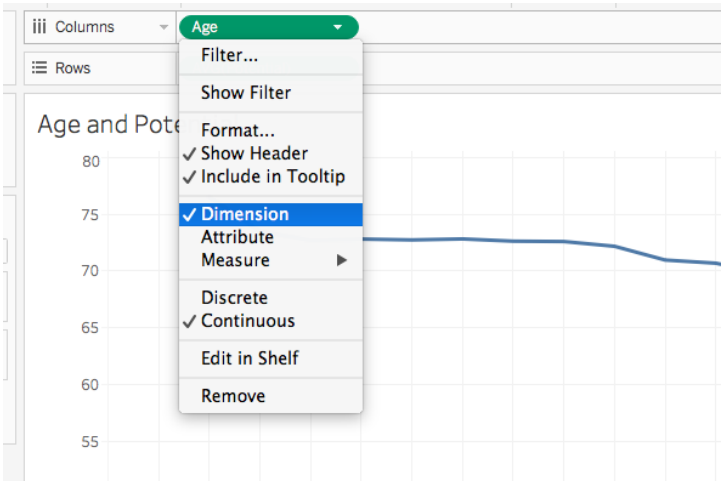


- This time, the two charts were stacked vertically (though they could have been placed horizontally, side by side, as well).
- The bottom chart, Number of Records by Nationality, is a bar chart by default. To change it to a line chart, click on the `Number of Records` pill and select "Line" under `Mark Type`.

- The next visualization plots age against potential:



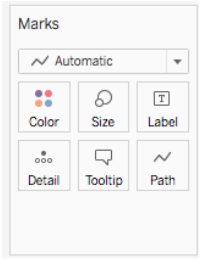
- To be able to chart each age year as a discrete quantity, click on the `Age` pill and select `Dimension`.



13. BREAK (0:15)

14. Instructor Do: Sizing, Coloring, and Labels (0:02)

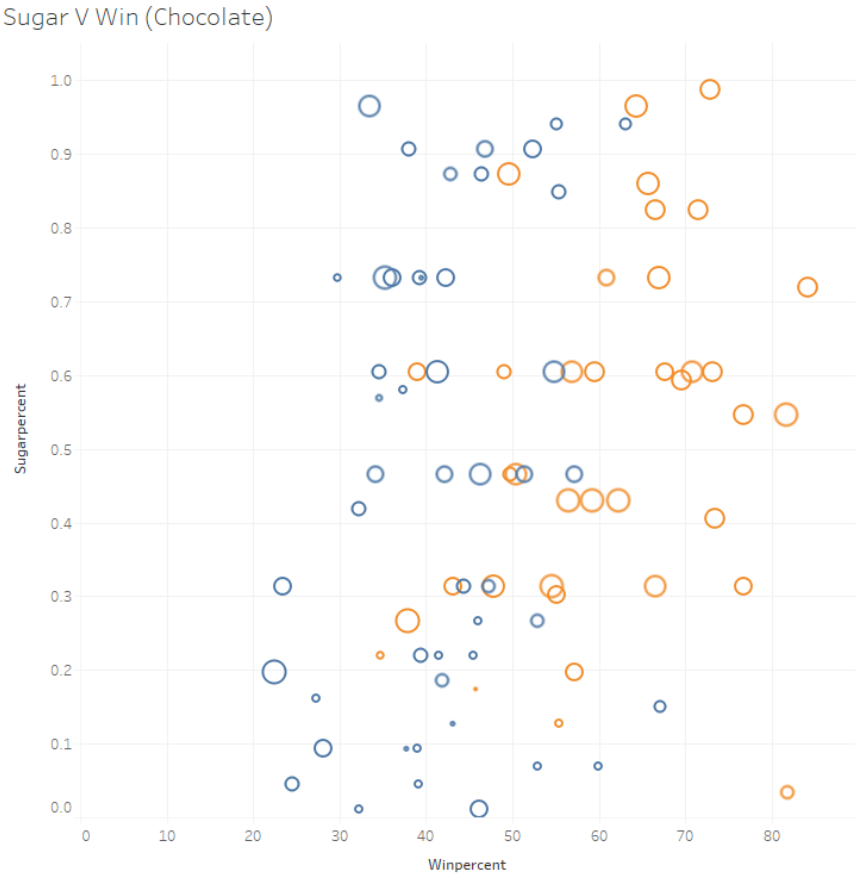
- Students have likely noticed by now that there are panels on the left side of the application that they have yet to touch. These marks can be used in order to differentiate or add details to a chart's visuals.



  - `Color` : Modifies the colors of a chart so that elements are colored according to the values passed.
  - `Size` : Modifies the sizing of elements to that they are bigger or smaller depending upon the values passed.
  - `Label` : Places text next to points on a chart that correspond with the values passed.
  - `Detail` and `Tooltip` : Acts much like labels, but only appear when the cursor hovers over the associated point/element on a chart.
  - `Shape` : Changes the shape of an element/point depending upon the values passed.
- Explain to students that they can drag pills to these marks to create visual effects. They will have an opportunity to do just that in the next activity.

15. Students Do: The Ultimate Candy (0:10)

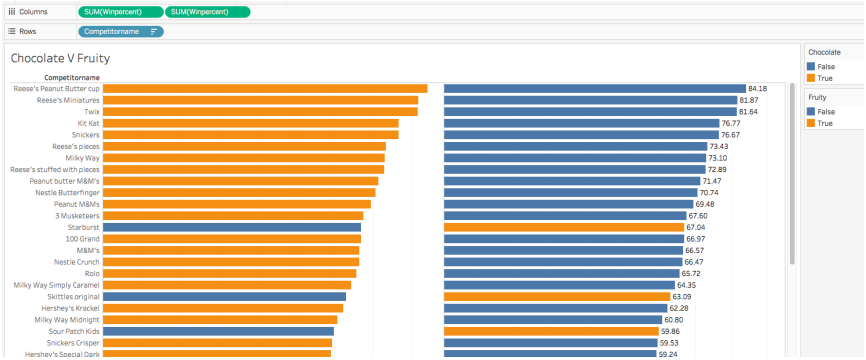
- Students will now take some time to create charts that compare candies against one-another. The charts themselves are quite basic but will be made more complex using marks.



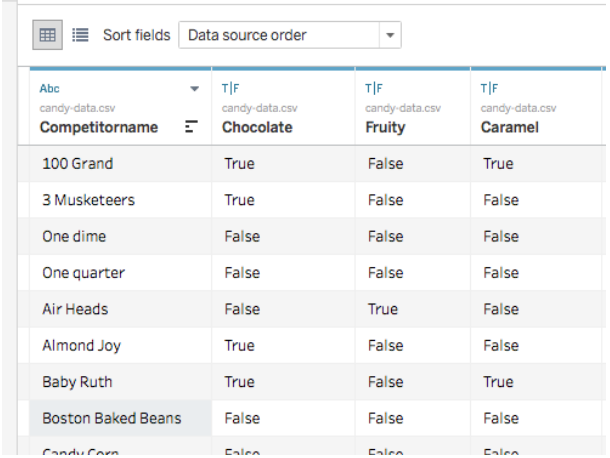
- **File:**
  - candy-data.csv
- **Instructions:**
  - Create a pair of bar graphs that chart the win percent of each candy, then color the bars according to whether they are fruity and/or chocolatey.
  - Create a scatter plot comparing the sugar percentage against the win percentage. Color the points based upon whether they are chocolatey and size them according to price.
  - Create one more scatter plot comparing the sugar percentage against the win percentage. Color the points based upon whether they are fruity and size them according to price.

16. Everyone Do: The Ultimate Candy Review (0:05)

- Open up 07-Stu_UltimateCandy within Tableau and walk through the application with the class, answering whatever questions students may have.
- The first visualization is a pair of bar graphs that chart the `Winpercent` of each candy:

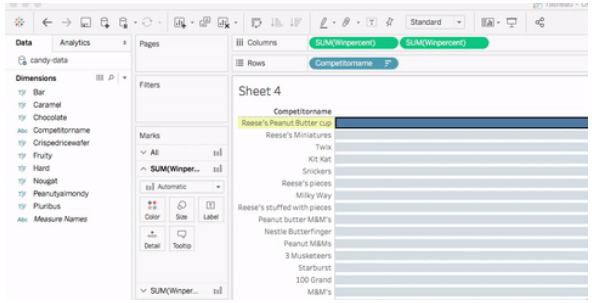


  - The bars in the left chart are colored by whether the candy is chocolate-flavored.
  - The bars in the right chart are colored by whether the candy is fruity.
- In the `Data Source` tab, `Chocolate` and `Fruity` columns hold true or false values.

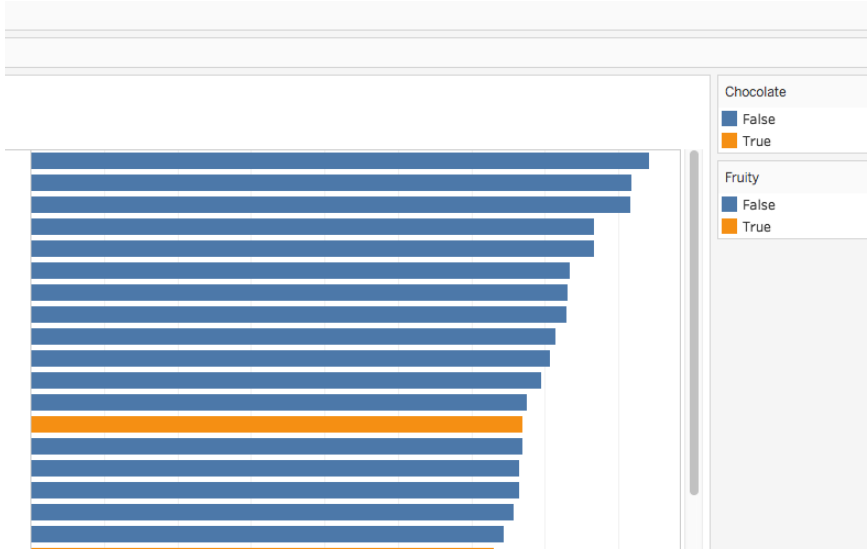


Abc candy-data.csv Competitorname	T|F candy-data.csv Chocolate	T|F candy-data.csv Fruity	T|F candy-data.csv Caramel
100 Grand	True	False	True
3 Musketeers	True	False	False
One dime	False	False	False
One quarter	False	False	False
Air Heads	False	True	False
Almond Joy	True	False	False
Baby Ruth	True	False	True
Boston Baked Beans	False	False	False

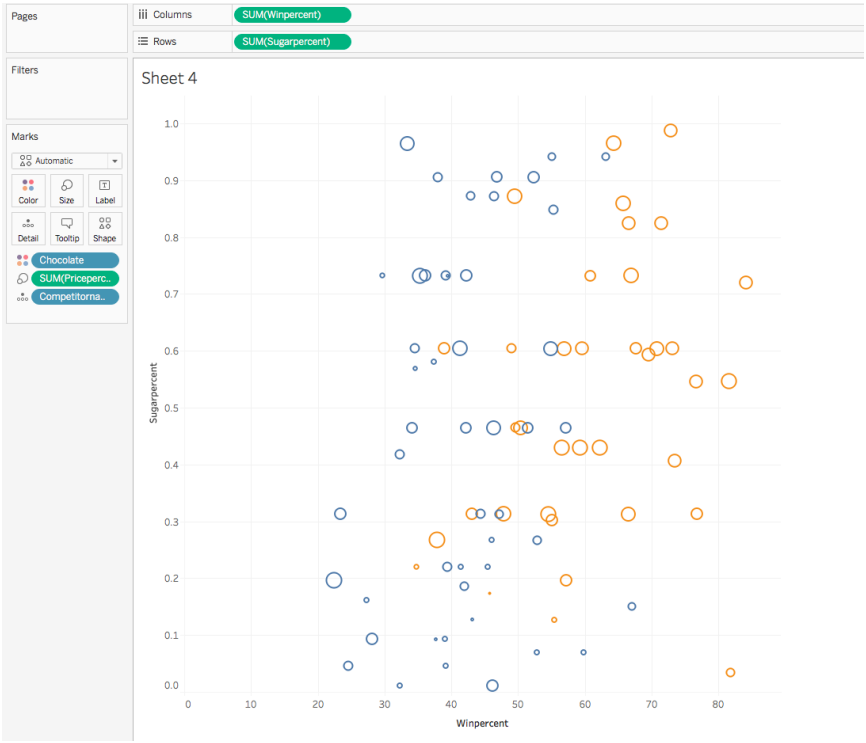
- Back to the charts. The bars to the left can be colored simply by dragging the `Chocolate` pill to the Color mark.



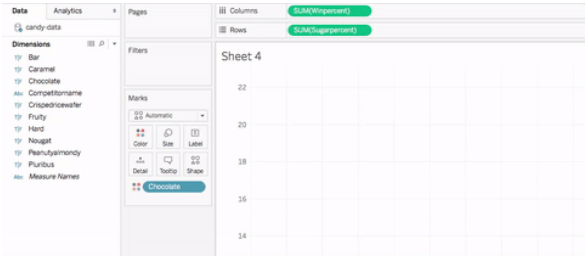
- Do the same for the bar chart on the right side. Tableau automatically colors the bars, and creates a legend.



- In the second visualization, a scatterplot is created that plots `Winpercent` against `Sugarpercent`.

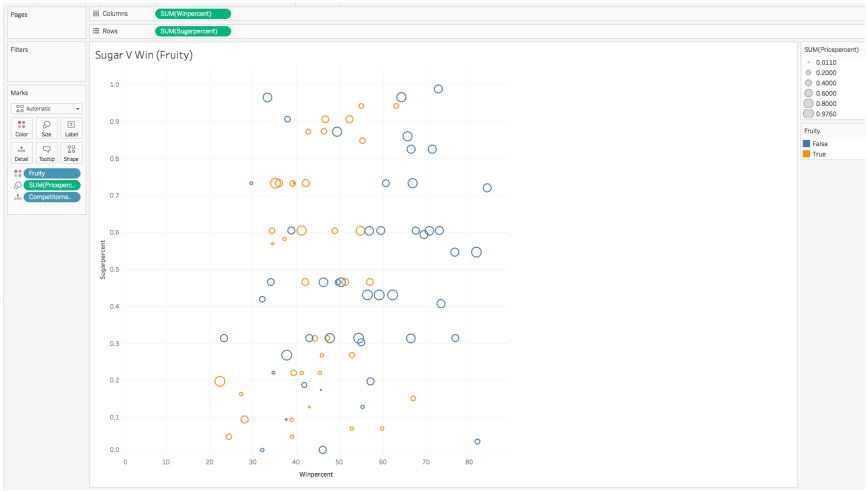


- `Chocolate` is dragged to the Colors mark, `Competitorname` is dragged to the Detail mark, and `Pricepercent` is dragged to the Size mark.



  - This chart provides a handy view of trends.
  - For example, candies with a higher win percent tend to be chocolates (orange), and they tend to be pricier (larger circle size).

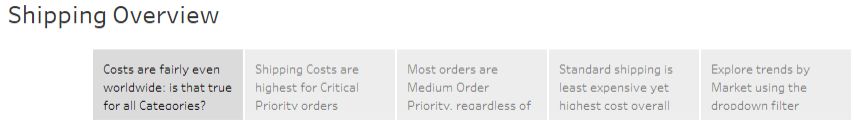
- The last visualization is virtually identical to the previous one, except that it compares fruity candies against non-fruity.



  - Overall, fruity candies appear to have a lower `Winpercent`, and tend to be less expensive.

17. Instructor Do: Storytelling (0:05)

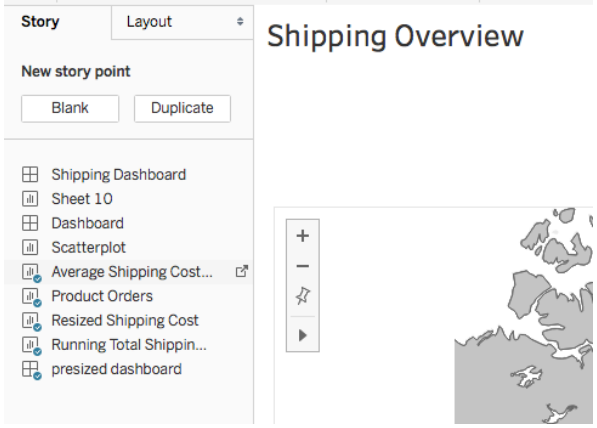
- Sometimes a single chart does not provide viewers with all of the information they might desire. In fact, visualizations are sometimes only truly helpful when placed alongside other charts/data.
  - Tableau makes the process of bringing multiple charts together in one place using stories.
- Open up 08-Ins_Storytelling within Tableau and navigate into the `Shipping Overview` tab, pointing out how there are a bunch of text-boxes at the top of this view.



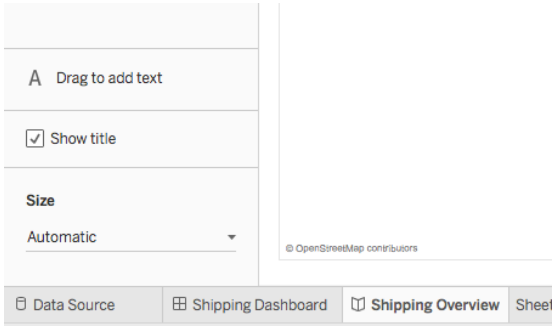
  - Click through a couple of the text boxes at the top of the view, discussing with the class how each text box is associated with a visualization from within the workbook.
- Create a new story within the workbook by selecting the `New Story` button from the bottom tabs of the application.

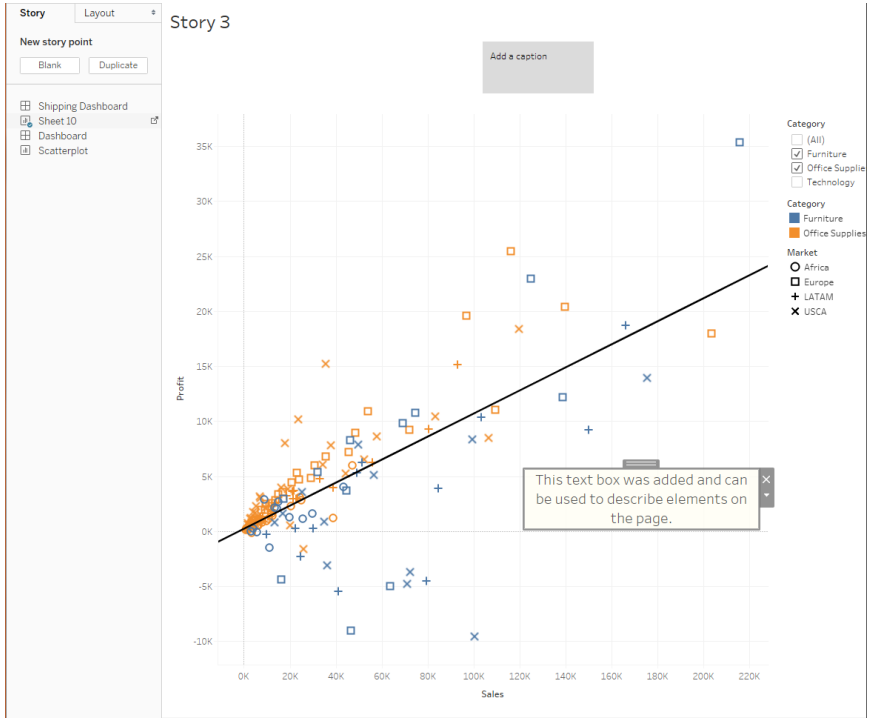
- The view on the left side of the page will now contain all of the sheets within the current workbook and can be added into the story by dragging them into the main area.
- Captions for the story point can be added/edited by clicking on the gray box at the top of the main view.

- To add a new page to a story, navigate to the `New Story Point` and select either `Blank` to create a blank page or `Duplicate` to create a page based upon an already existing page.



- Text boxes may also be added to the page by dragging the `Drag to Add Text` element onto the page so as to allow for more detailed explanations.

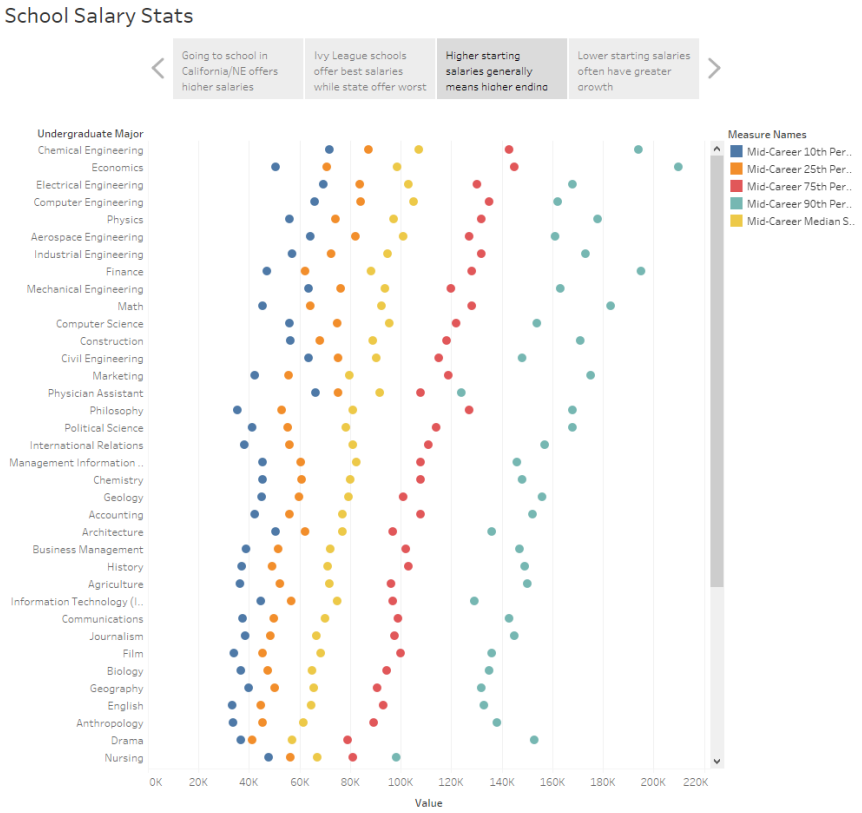




- Answer whatever questions the class may have regarding stories and then continue onto the next activity.

18. Students Do: Which Degrees Pay? (0:15)

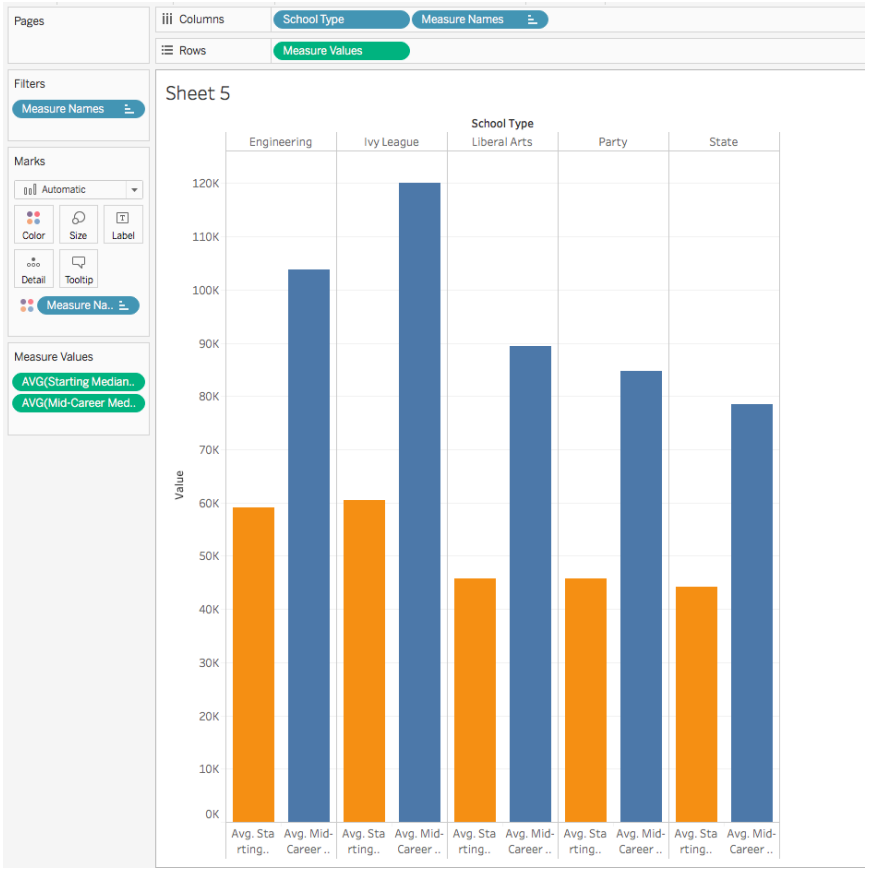
- The class will now build upon everything they have learned today in order to create a story in Tableau that visualizes what degrees/universities/regions pay out the best over time.



- **Files**:
  - degrees-that-pay-back.csv
  - salaries-by-college-type.csv
  - salaries-by-region.csv
- **Instructions**:
  - Create a story using the datasets provided and formulate graphs that might be used to explore the following hypotheses:
  - "Ivy League schools offer best salaries while state offer worst"
  - "Going to school in California/NE offers higher salaries"
  - "Higher starting salaries generally mean higher salaries mid-career"
  - Bonus: Create a chart that visualizes starting median salaries, by major, against mid-career median, 75th percentile, and 90th percentile salaries.

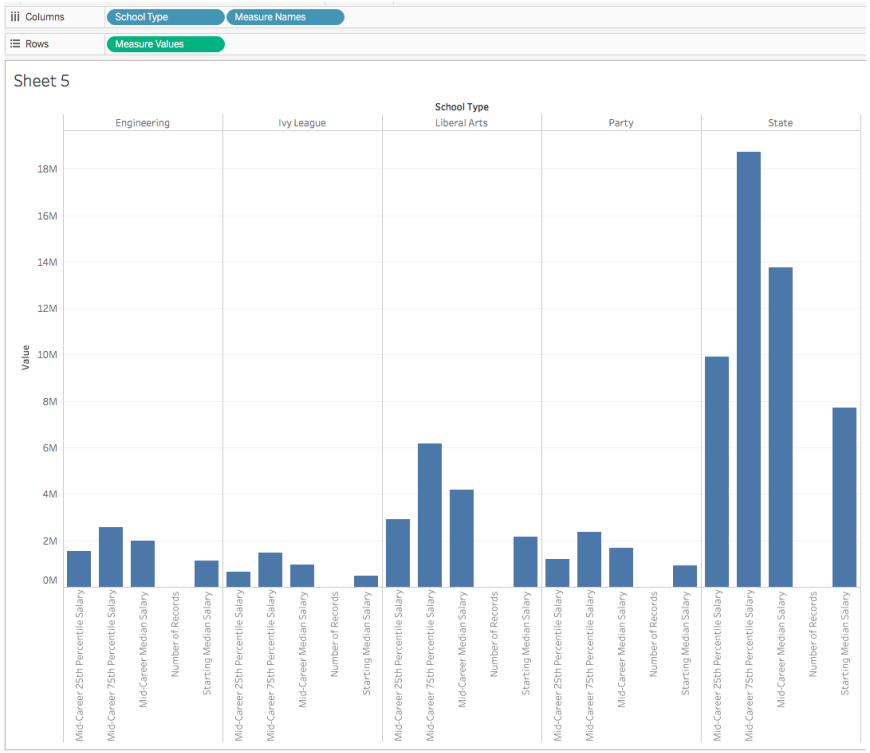
19. Everyone Do: What Degrees Pay? Review (0:10)

- Open up 09-Stu_DegreesPay within Tableau and walk through the application with the class, answering whatever questions students may have.
- Again, emphasize to students that there is no single correct solution, that there are many ways to create these visualizations.
- The first visualization address the first prompt: do Ivy League school graduates earn higher salaries than their counterparts from state schools?

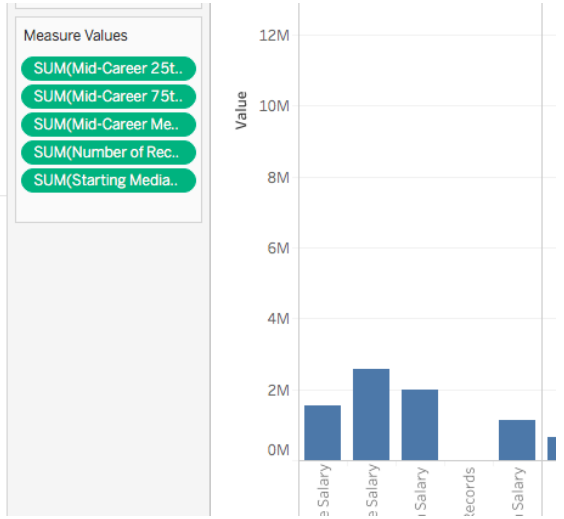


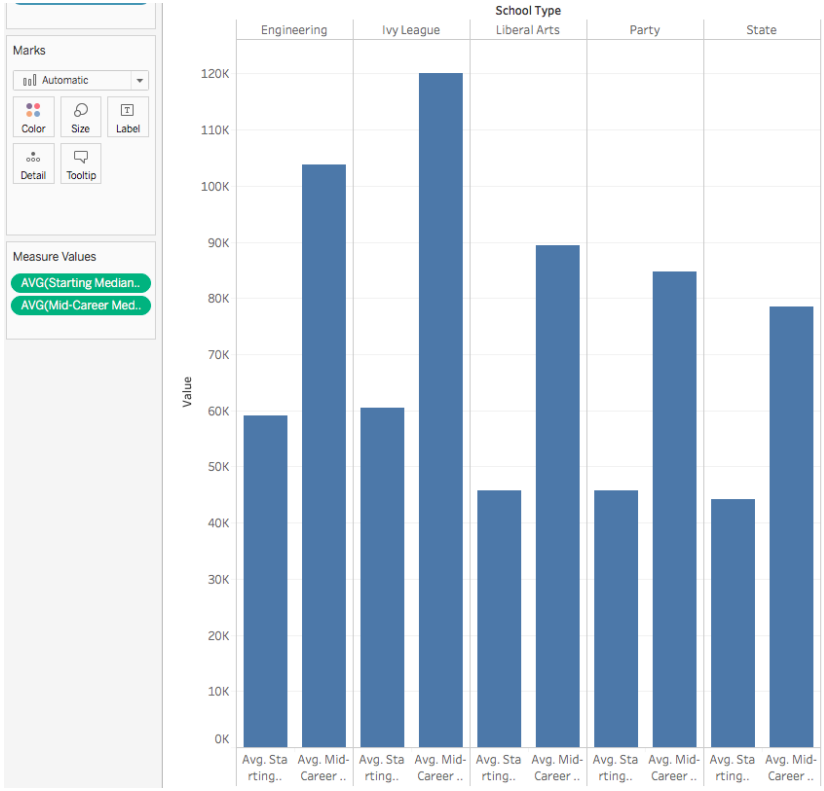
  - In this case, the average starting salaries, as well as mid-career salaries, were used.

- The chart can begin with `School Type` and `Measure Names` in Columns, and `Measure Values` in Rows.

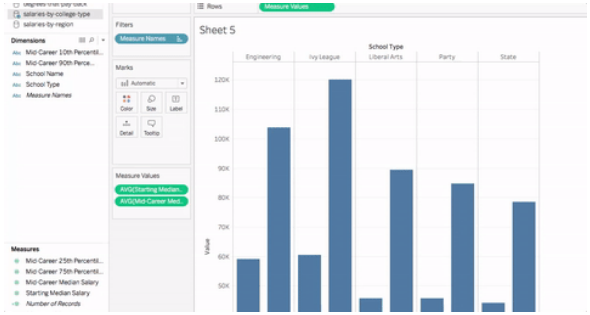


  - `Measure Names` and `Measure Values` are automatically generated by Tableau to enable building charts like this: Measure Names and Measure Values
- In the `Measure Values` pane, undesired pills can be removed, and aggregated as we wish: in this case, to averages.



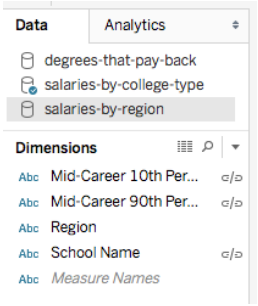


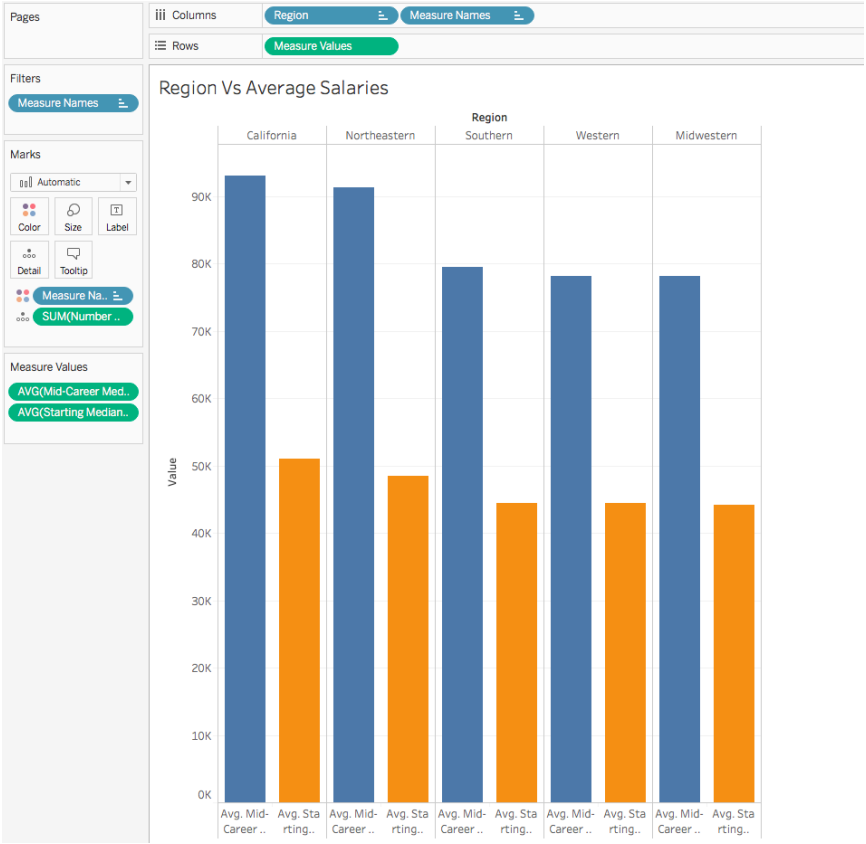
- The bars colors can be differentiated by dragging `Measure Names` to the Color mark:



  - The `Number of Records` pill can be moved to the Detail mark to include this detail in the tooltip.

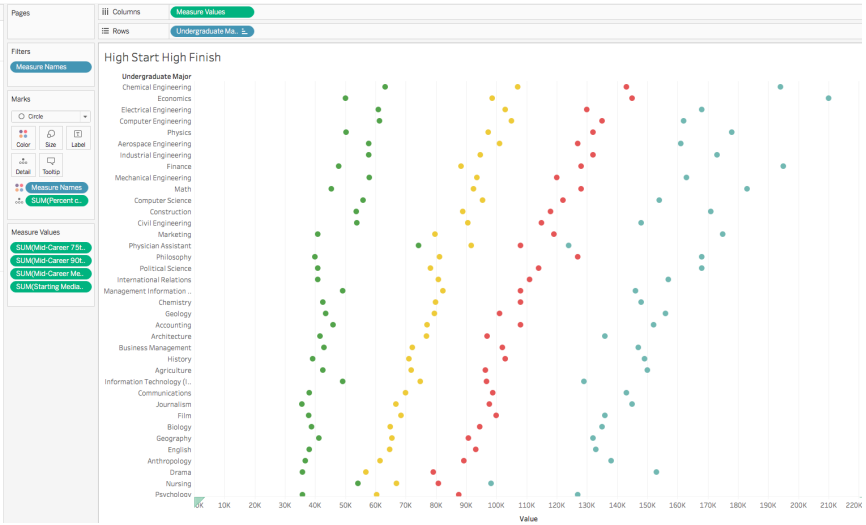
- The next visualization address whether grads of schools in the Northeast or California earn higher salaries than their counterparts in other regions. It is altogether similar to the previous one, substituting `Region` for `School Type`:



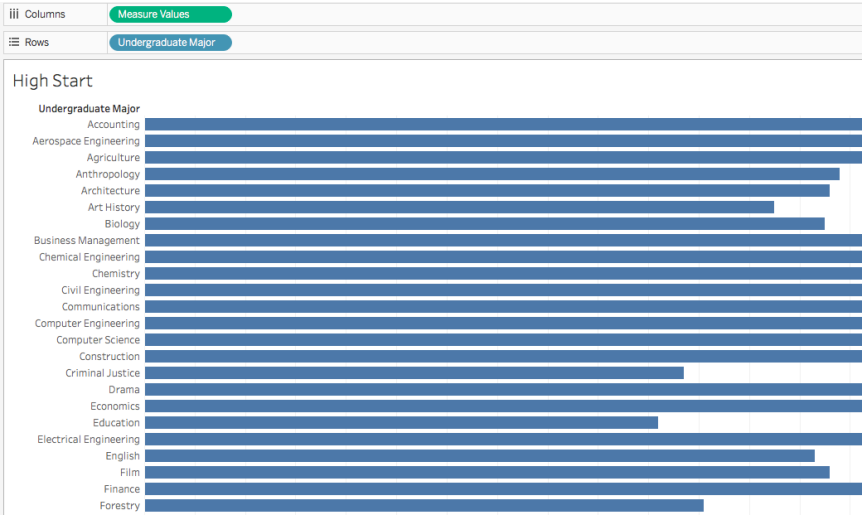


- In the next visualization, the question explored is whether higher starting salaries lead to higher salaries mid-career:

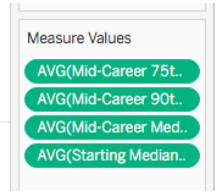
  - It plots the average starting median salary in Columns against average mid-career salary in Rows.
  - The `Undergraduate Major` pill is dragged to the Detail mark.
  - The `Percent change from Starting to Mid-Career Salary` pill is dragged to the Size mark.
- In the final visualization, the median starting salaries, on the left in green, are compared with mid-career salaries at 50th (median), 75th, and 90th percentiles.



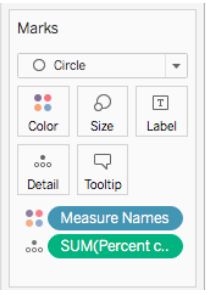
- This chart can be created by starting with the `Undergraduate Major` in Rows, and `Measure Values` in Columns.



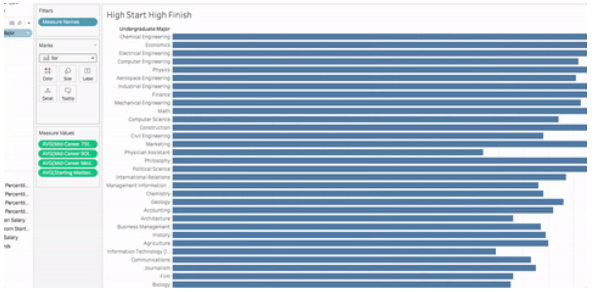
  - Again, `Measure Values` is automatically generated and allows multiple columns to exist in a single chart.
- `Measure Values` are modified to include the desired columns.



- Then `Measure Names` can be dragged to the Color mark, and `Percent change...` to the Detail mark.



- To change from a bar chart to circles, click on the drop-down menu in Marks, and choose `Circle` :



Copyright