

DATA SOCIETY®

The premiere data science training for professionals

“One should look for what is and not what he thinks should be.”

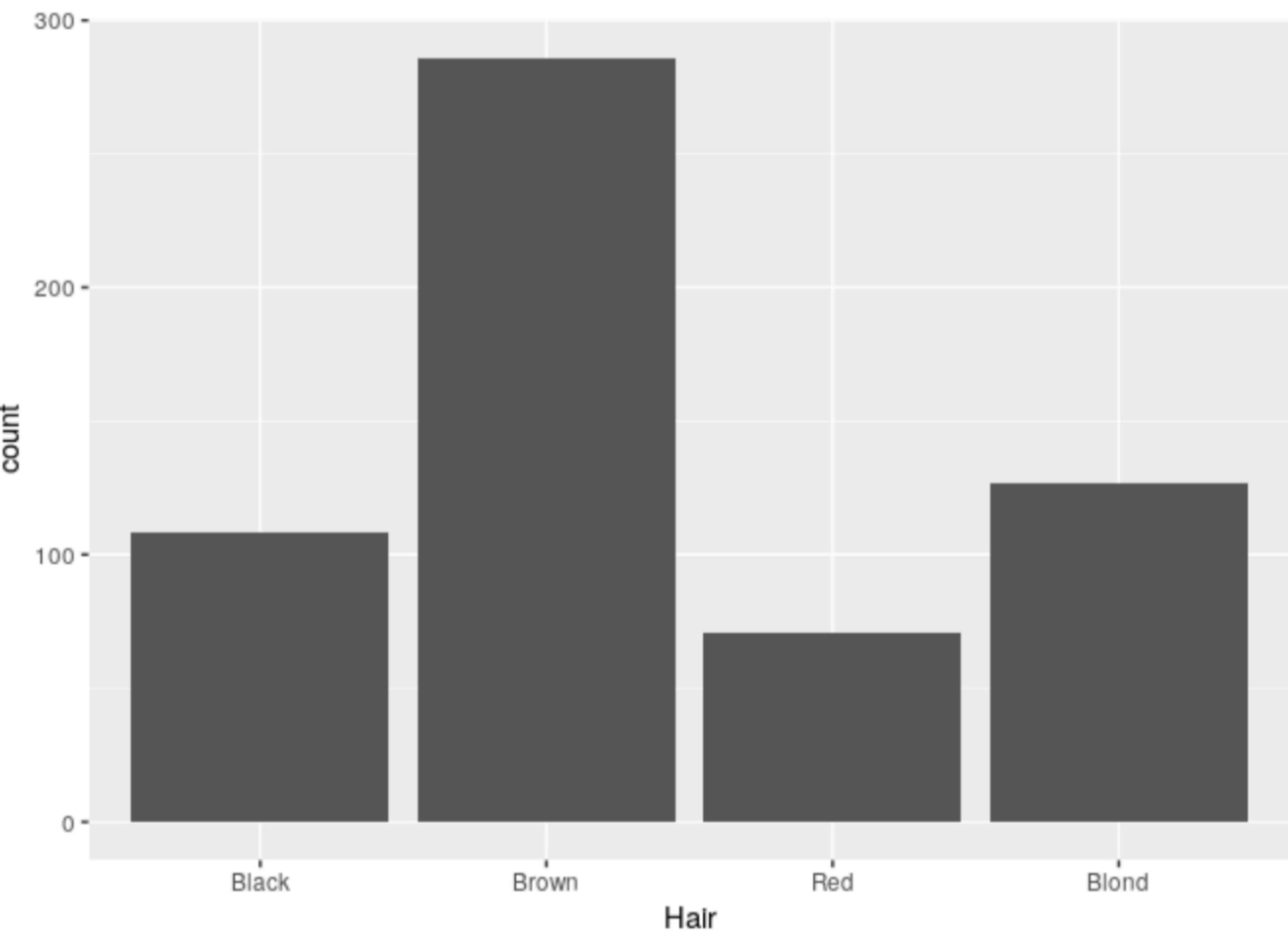
- Albert Einstein

Recap: 10 Design Principles

- As you're building out your visualization, keep these principles in mind:
 1. Ensure an appropriate resolution / file type
 2. Use high contrast colors meaningfully
 3. Use labels and descriptions
 4. Make sure your chart is readable
 5. Use accurate and proportionate scales
 6. Make sure comparative sizes are accurate
 7. Remove unnecessary information
 8. Clear out dense charts
 9. Double check data reliability and accuracy
 10. Include citations for your data sources

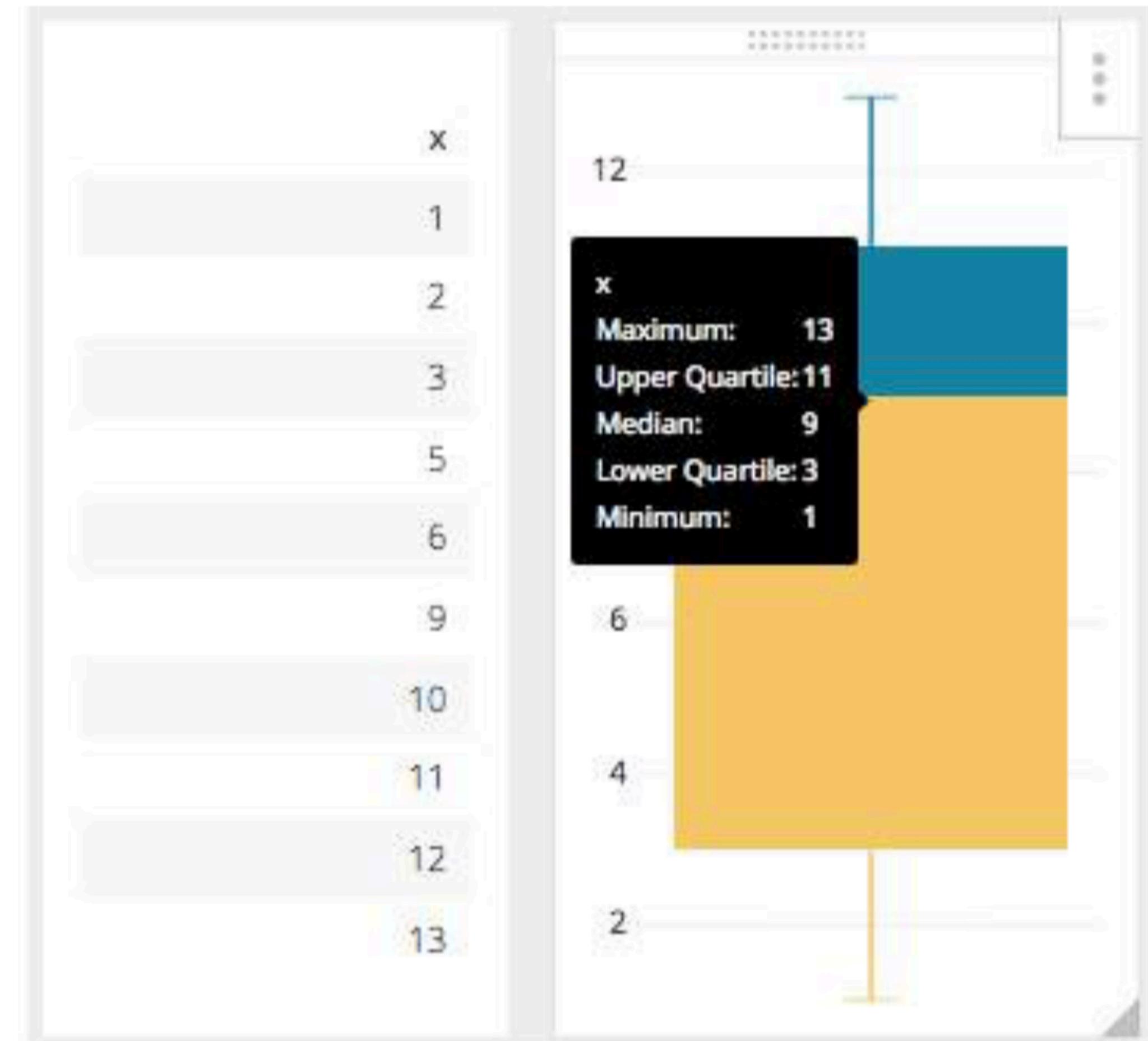
Recap: Visualization

- Insight - comparisons and proportions
 - Chart types include vertical bar, column bar, horizontal bar, pie, bullet charts, stacked bar, and tree maps
 - Categorical data (i.e. non-numeric or qualitative data)



Recap: Visualization

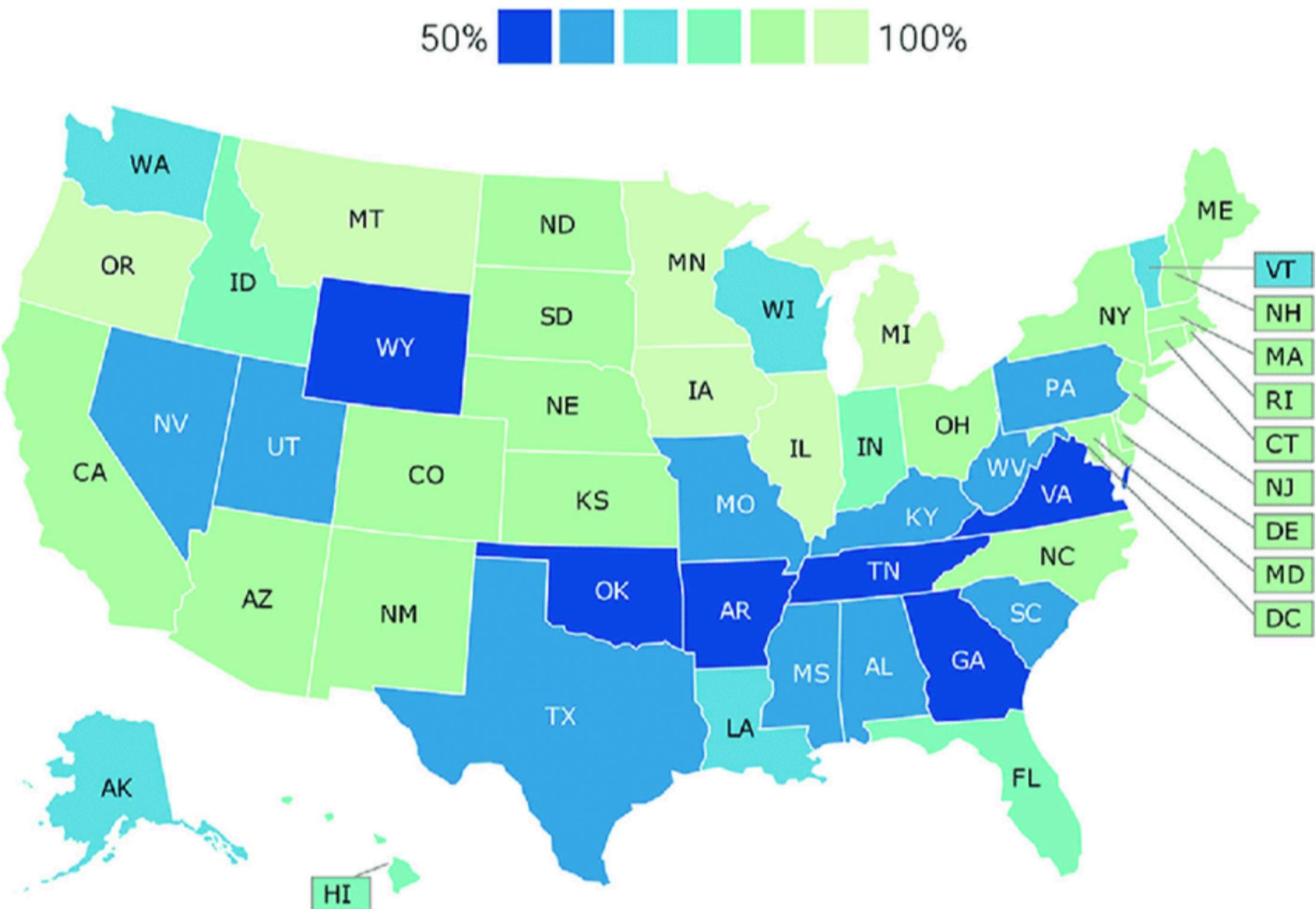
- Insight - distributions, proportions, frequencies
 - Chart types include histogram, density, and box plots
 - Univariate data (one numeric variable)



Recap: Visualization

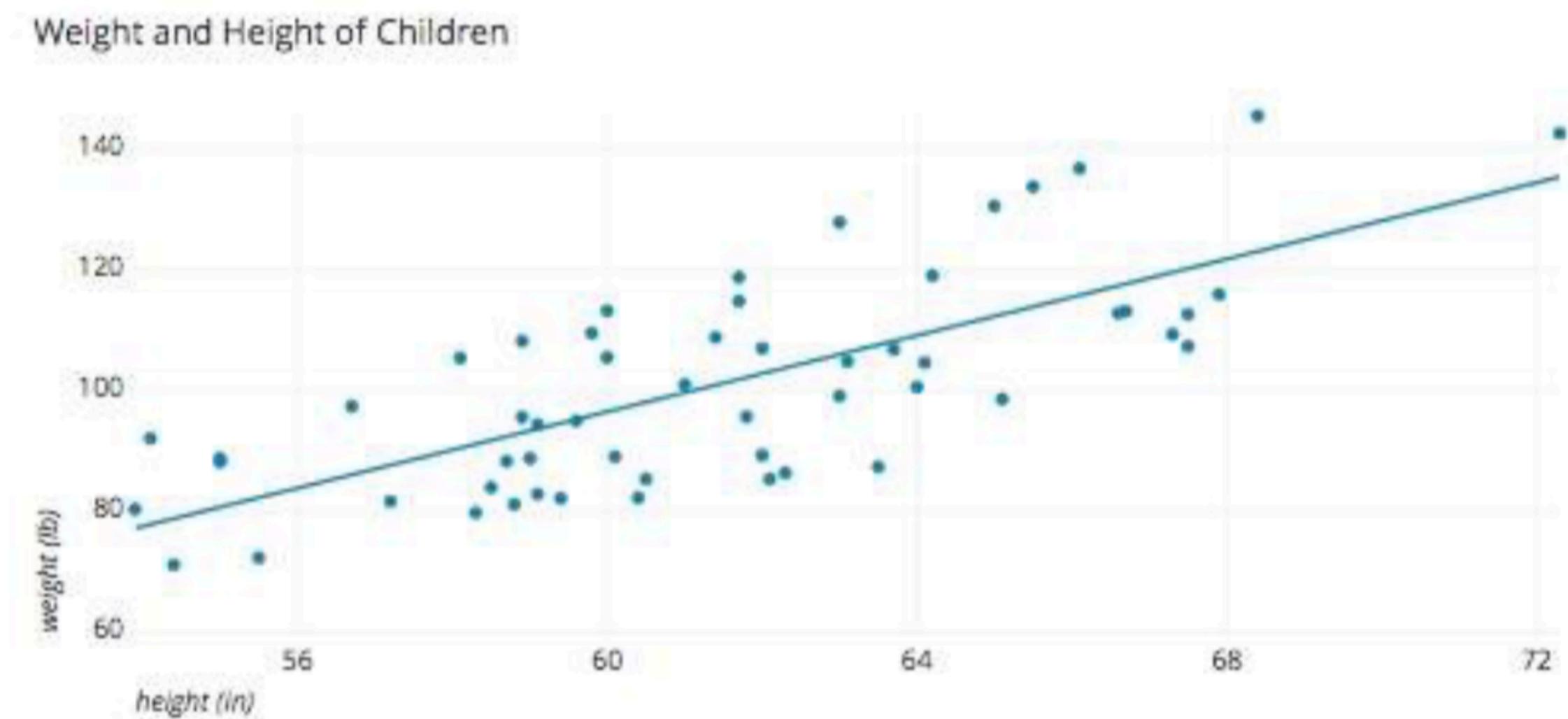
- Insight - locations, comparisons, trends
 - Chart types include choropleth filled map, point map, connection map, and isopleth map
 - Geospatial data (specific locations)

Smoke-free air law coverage by state (2017)



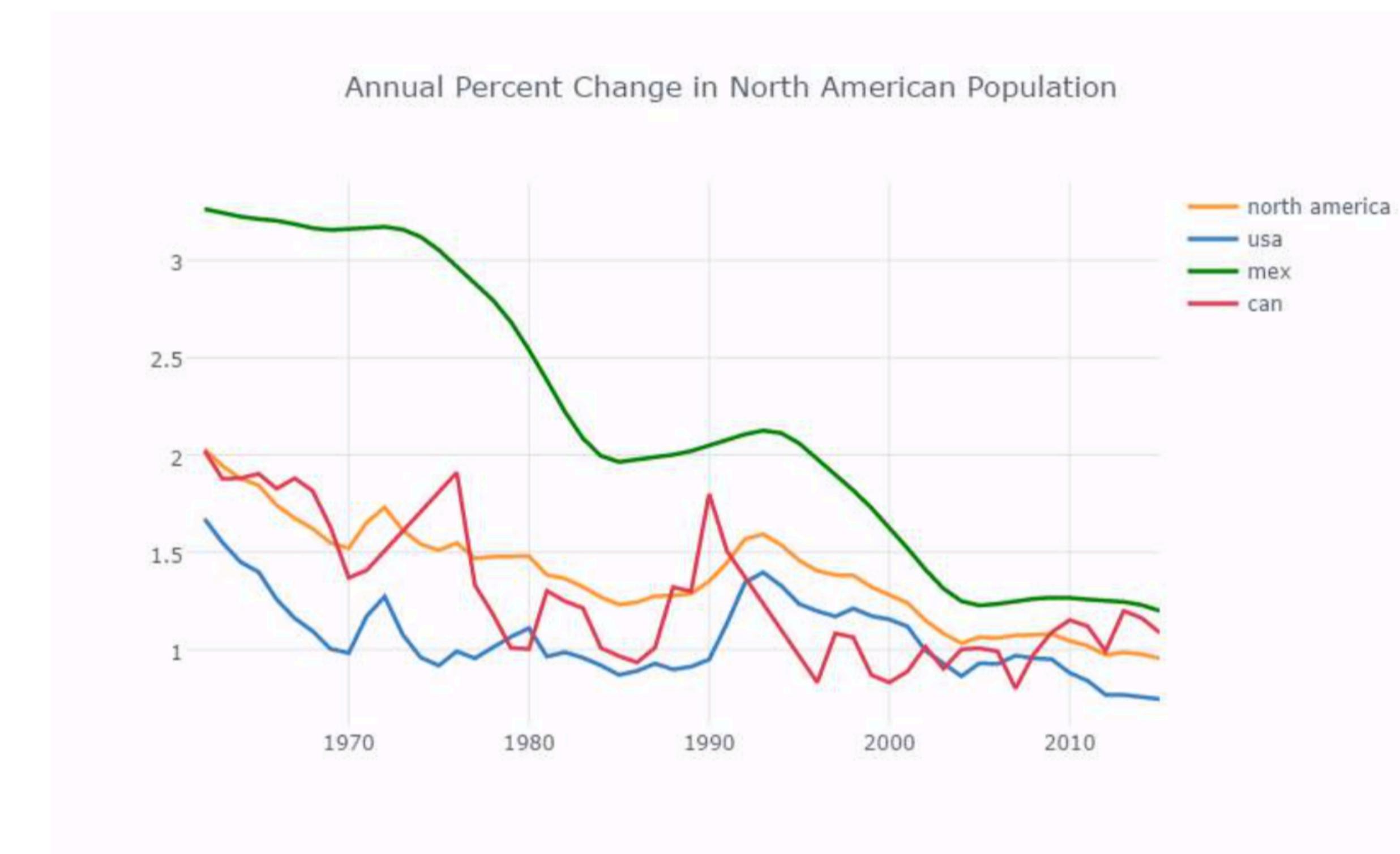
Recap: Visualization

- Insight – relationships, correlation, proportions, and frequencies
 - Chart types include scatterplot, bubble, parallel, radar, bullet, and heat
 - Two or more numeric variables (i.e. weight, height, and diabetes)



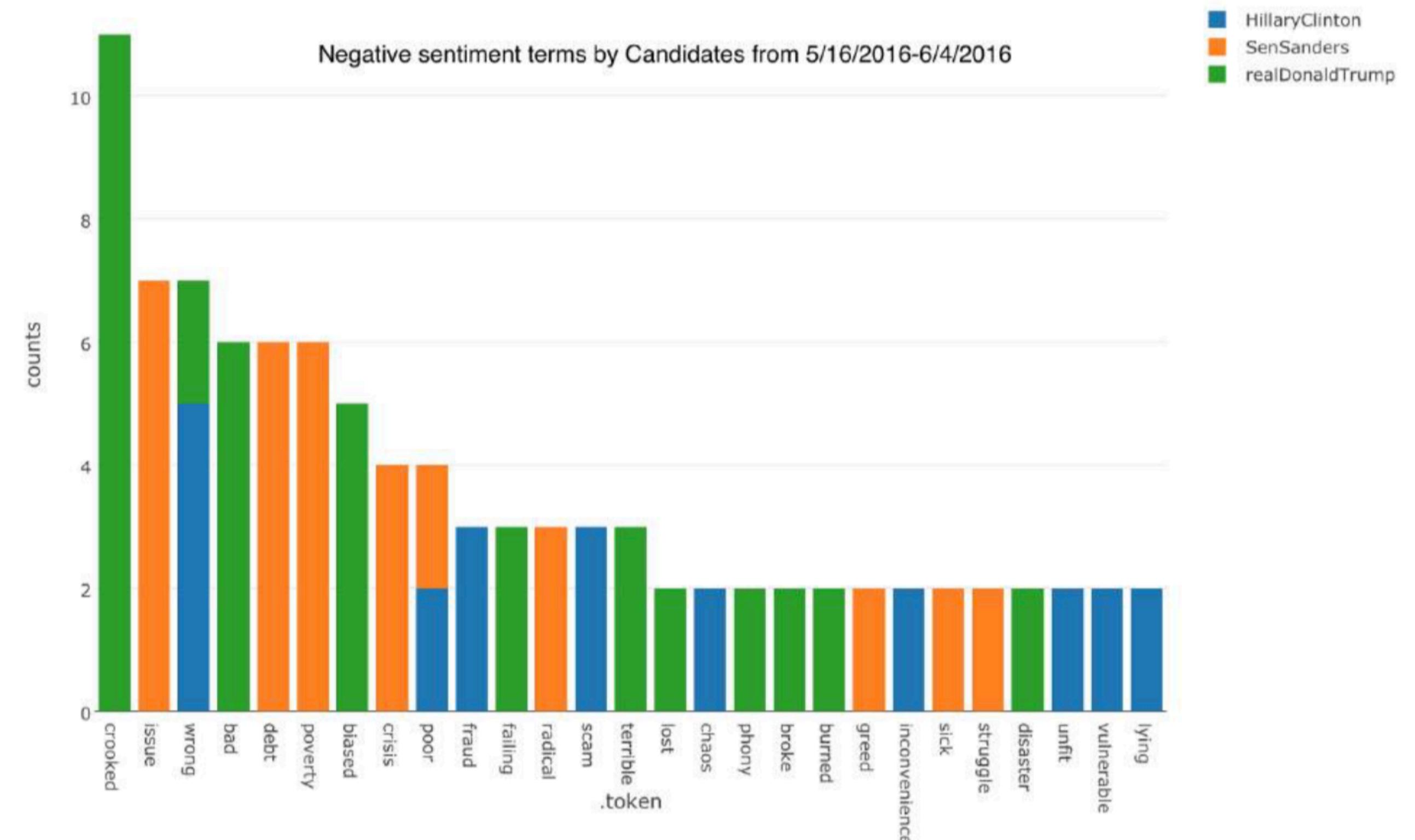
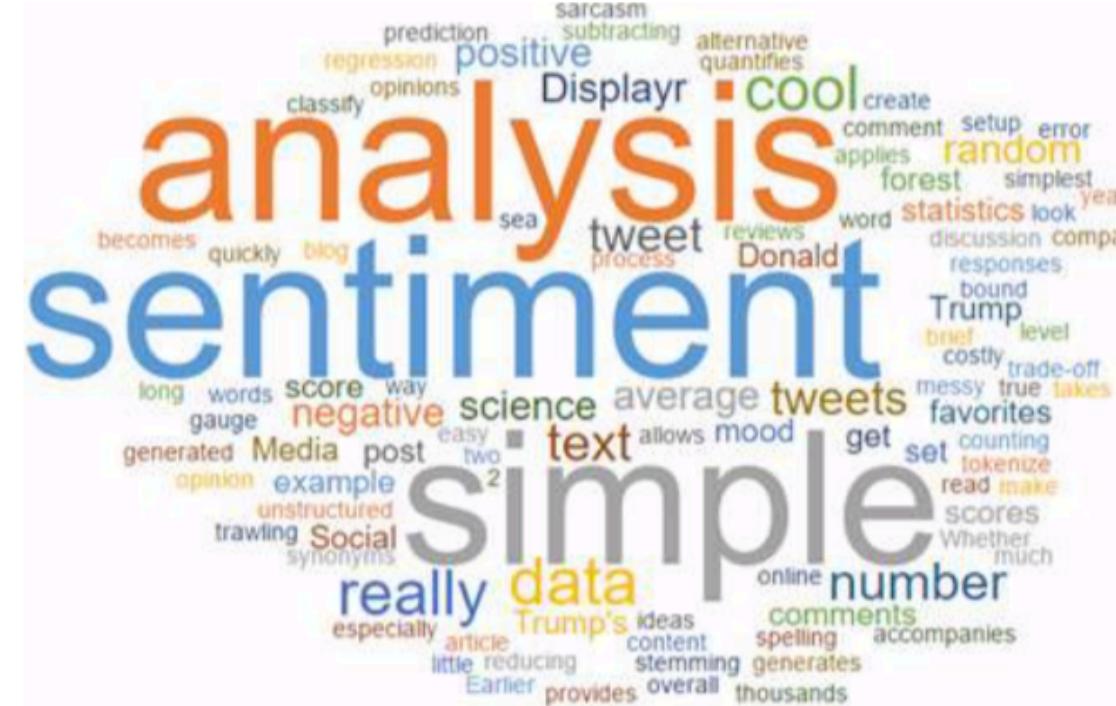
Recap: Visualization

- Insight - trends, comparisons and cycles
 - Chart types include line, area, bubble, and vertical bar charts
 - Years, months, days, hours, etc.



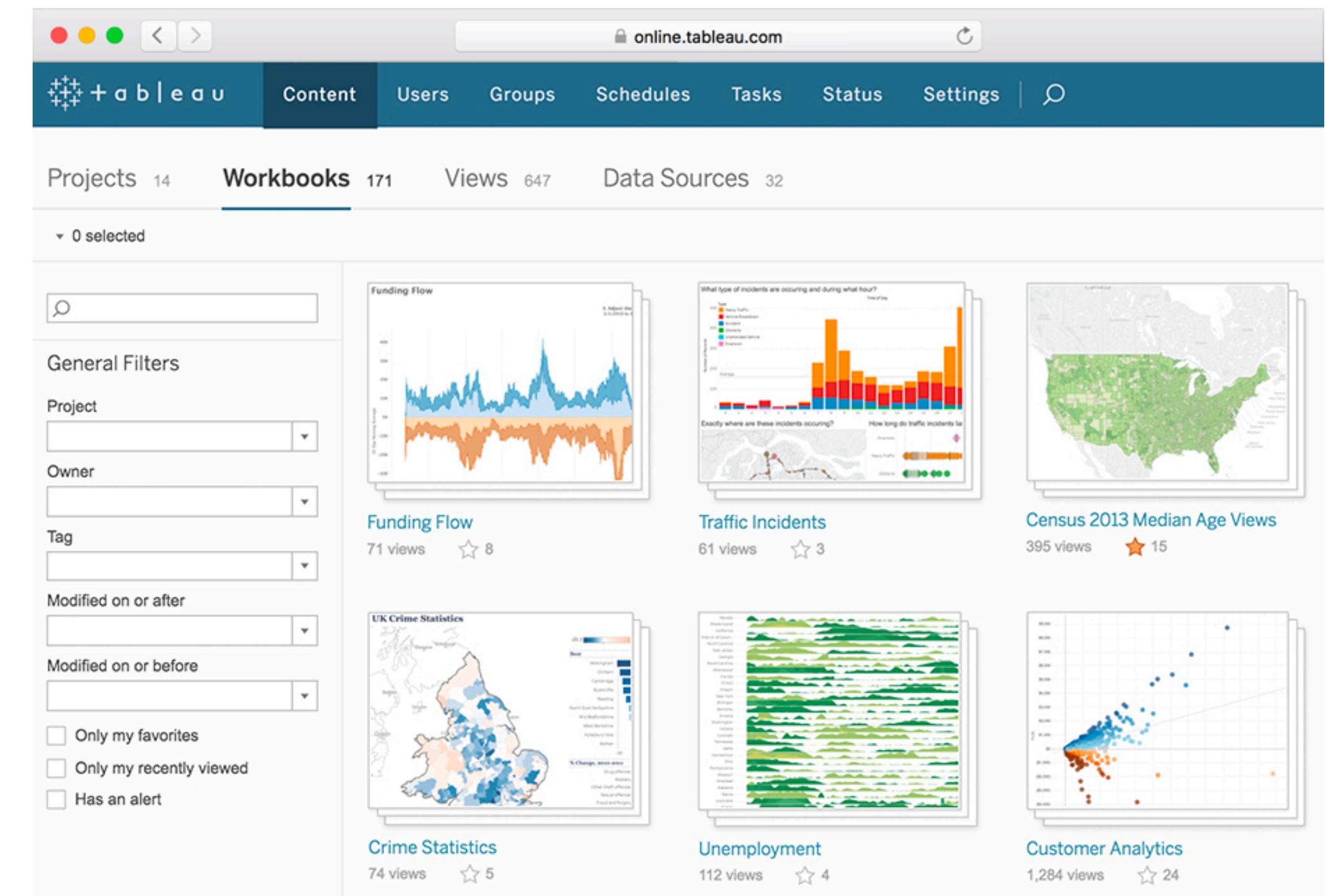
Recap: Visualization

- Insight - sentiment, comparisons, frequency
 - Chart types word cloud, histogram, bar chart
 - Single words or phrases (keywords)



Why Tableau?

- It offers a quick and easy way to create interactive visualizations and explore data.
- It is easy to integrate with multiple data sources.
- Compatible with OS X, Windows and Linux.
- Integrates with R and Python for more advanced analysis.



Excel vs Tableau

Parameters	Excel	Tableau
Purpose	Spreadsheet application used for manipulating the data.	Perfect visualization tool used for analysis.
Usage	Most suitable for statistical analysis of structured data.	Most suitable for quick and easy representation of big data which helps in resolving the big data issues.
Performance	Moderate speed with no option to quicken.	Moderate speed with options to optimize and enhance the progress of an operation.
Security	The inbuilt security feature is weak when compared to Tableau. The security update needs to be installed on a regular basis.	Extensive options to secure data without scripting. Security features like row level security and permission are inbuilt.

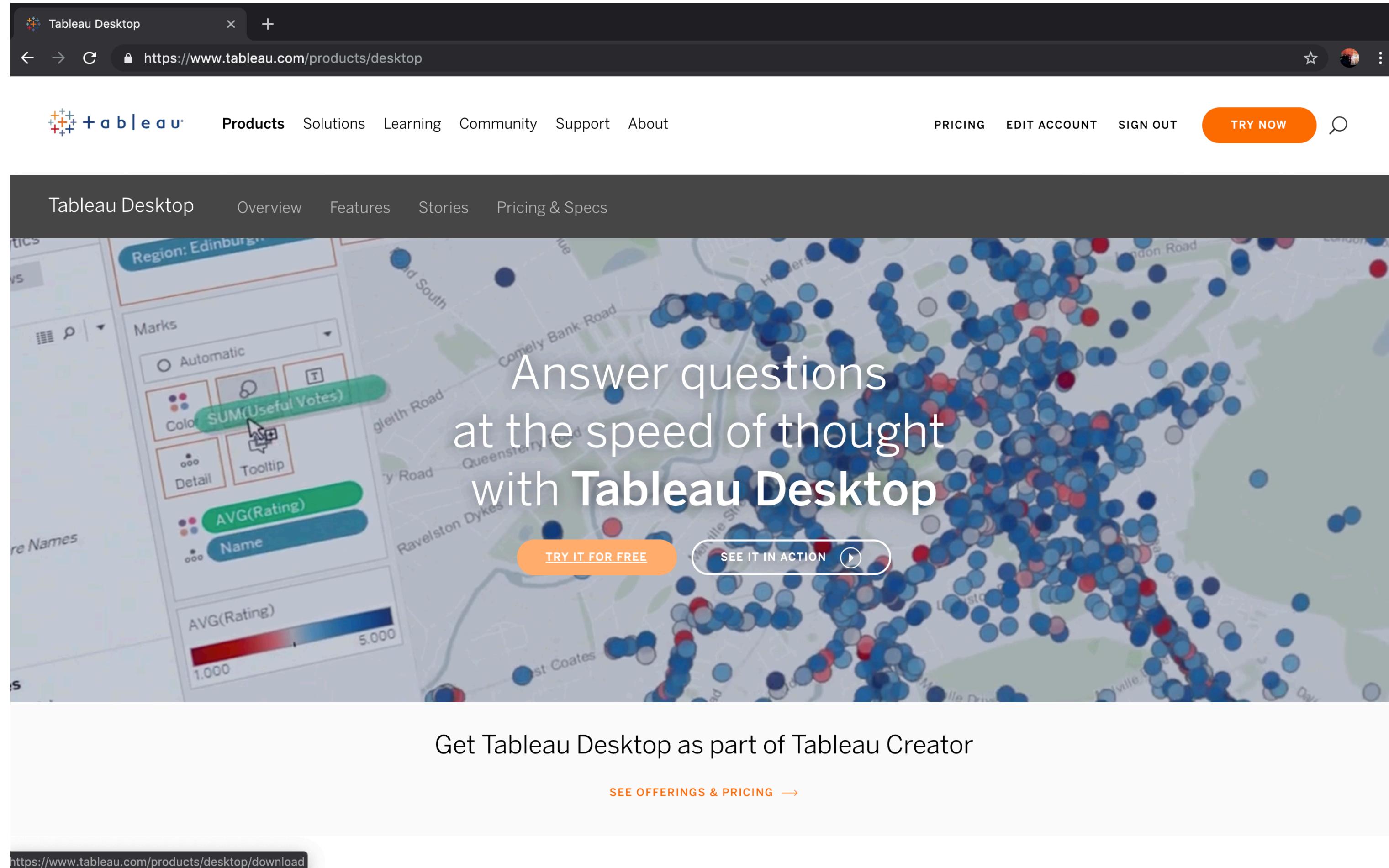
Excel vs Tableau

User Interface	To utilize excel to full potential, macro and visual basic scripting knowledge is required	The tool can be used without any coding knowledge.
Business need	Best for preparing on-off reports with small data	Best while working with big data.
Products	Bundled with MS Office tools	Comes with different versions such as the Tableau server, cloud, and desktop.
Integration	Excel integrates with around 60 applications	Tableaus integrated with over 250 applications
Real time data exploration	When you are working in excel, you need have an idea of where your data takes you to get to know the insights	In Tableaus, you are free to explore data without even knowing the answer that you want. With the in-built features like data blending and drill-down, you will be able to determine the variations and data patterns.

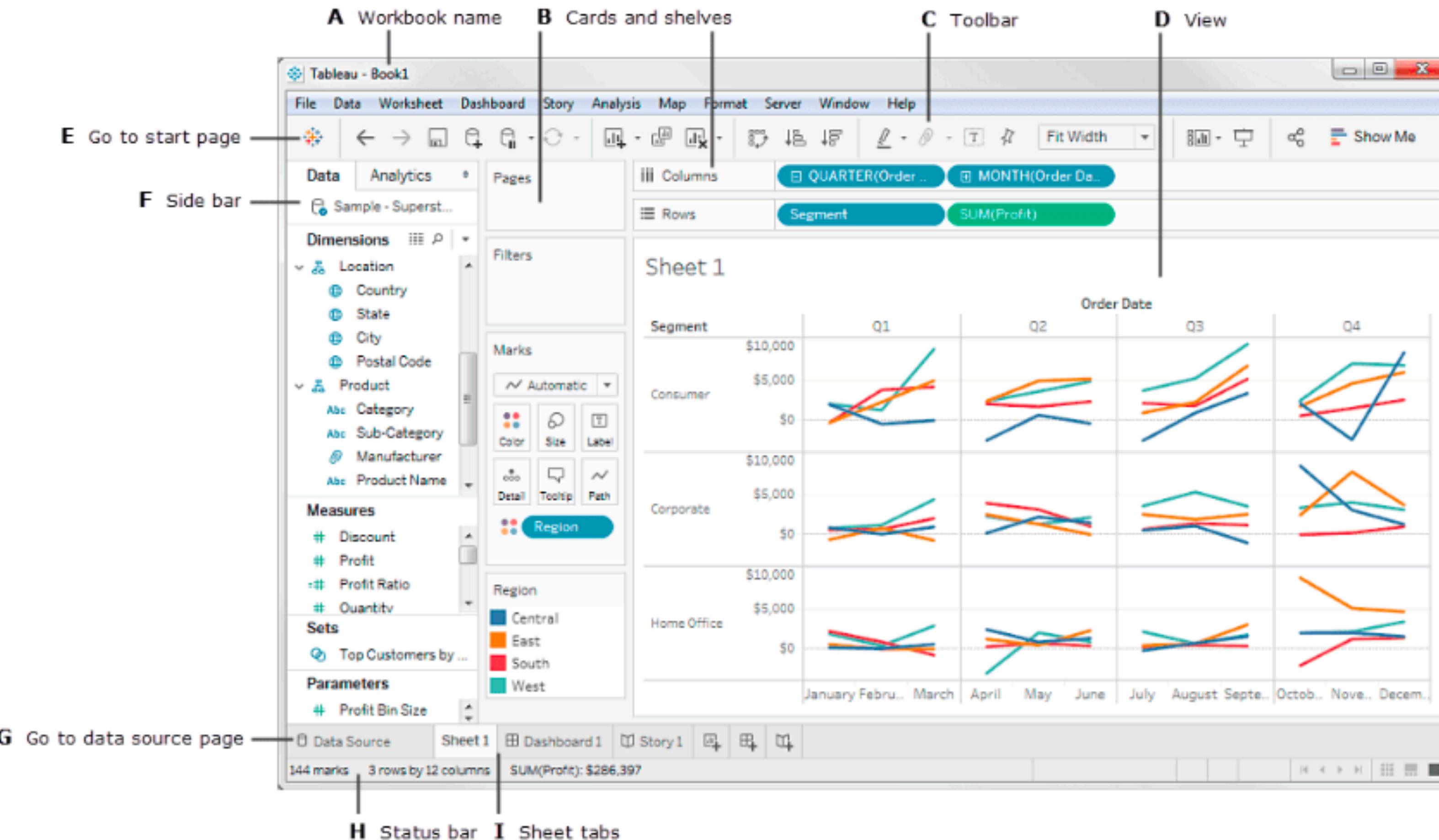
Tableau Desktop vs Tableau Public

	Tableau Desktop	Tableau Public
Pricing	Personal Edition - \$35 per month Professional Edition - \$70 per month	Free and Open Source
Data Source Connection	Connect to any data source files including databases, web applications and more	Connect to Excel and text files
Publish/Save	Can be saved as a local file and can be published in Tableau Server and Tableau Online	Can be published only in Tableau Public profile
OS	Windows and Mac	Windows and Mac
Security	Data and Visualizations are secured	Data and visualizations are not secured as it is available in public
Usage Details	Professional and Enterprise level	Personal level
Data Extract	Data can be extracted from various data sources and stored as Tableau extract file	Data cannot be obtained from different data sources as it is limited to connect only excel and text files.

Installation



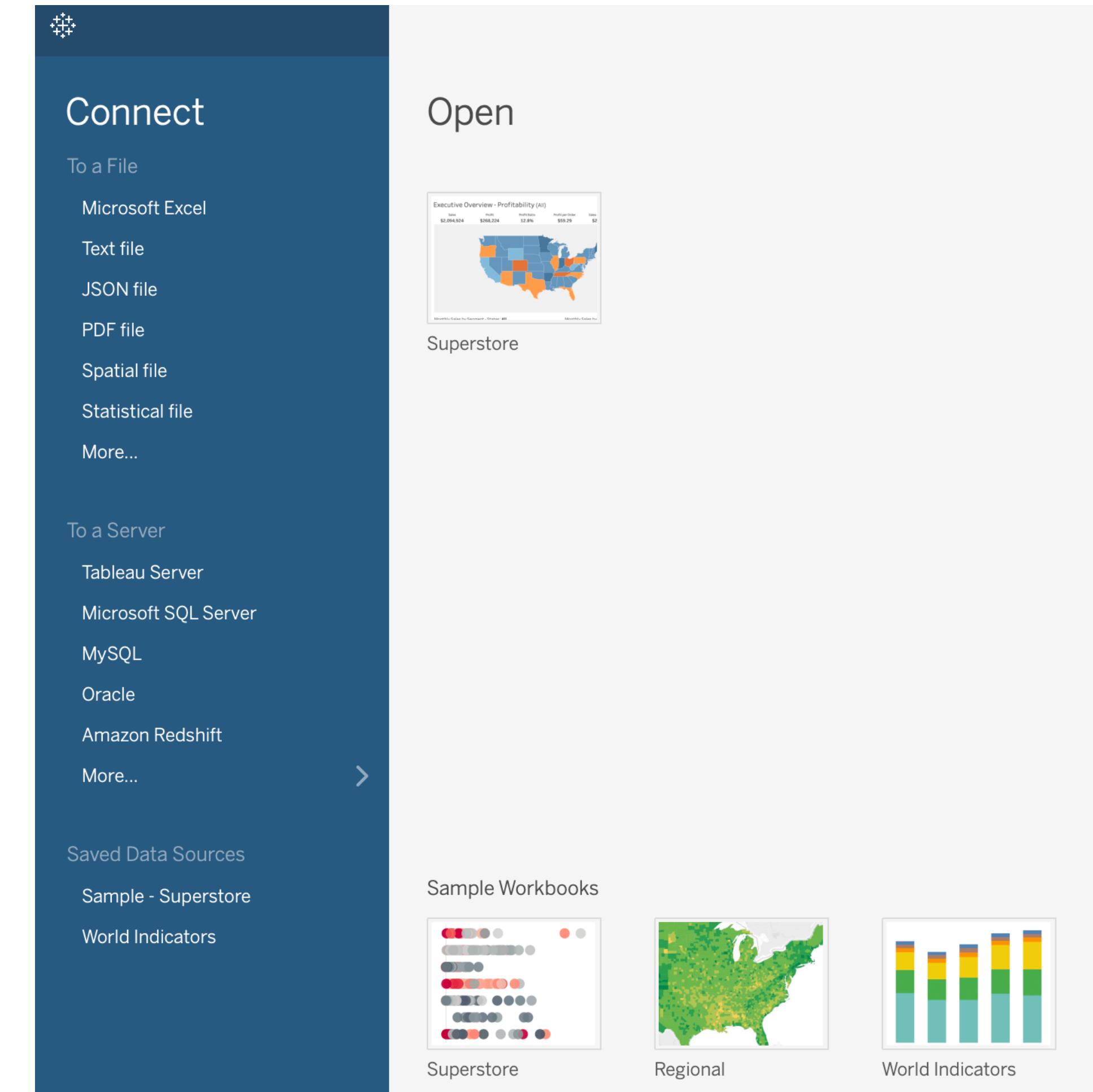
Overview



source: [tableau.com](https://www.tableau.com)

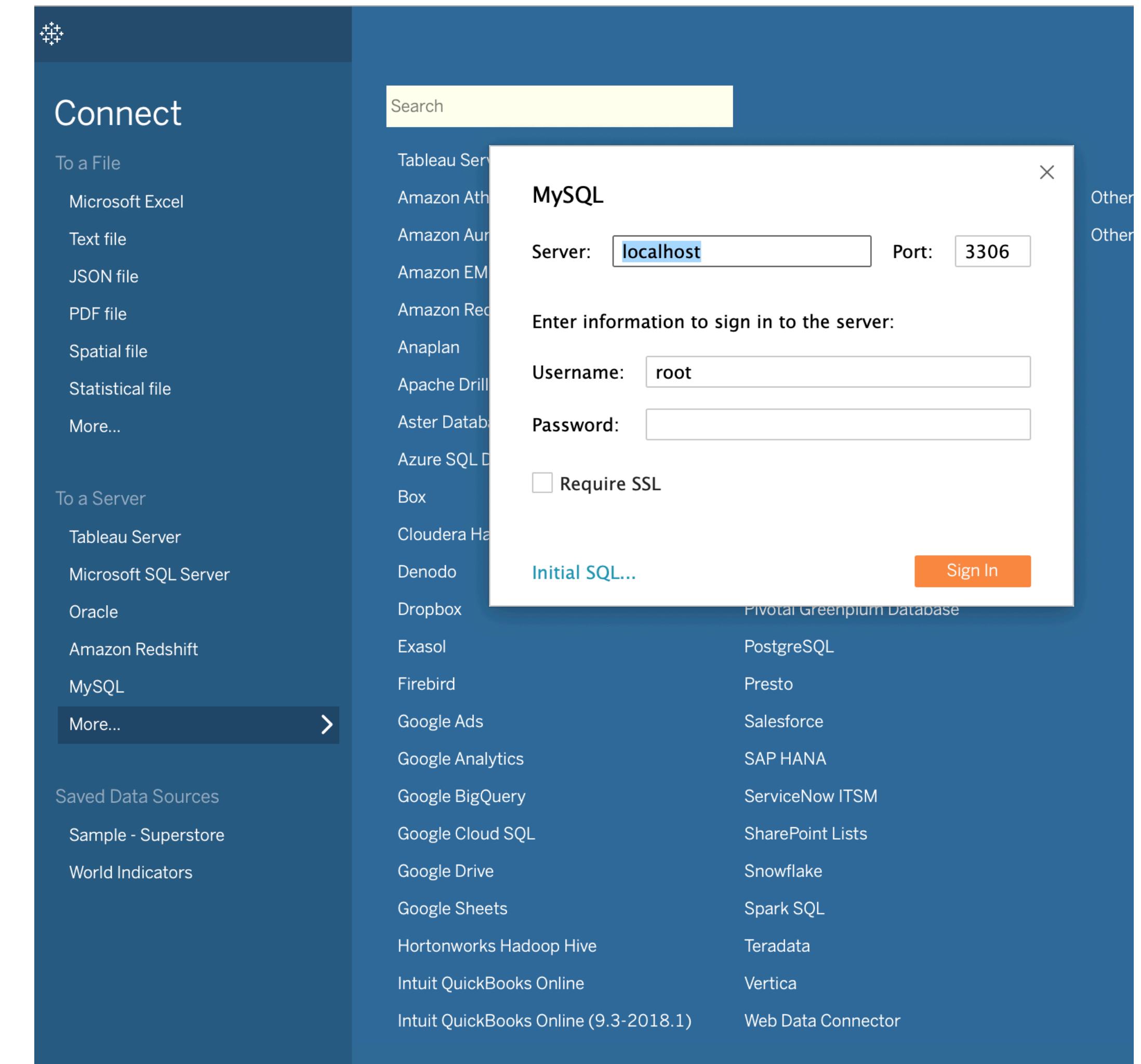
Importing Data

- Supports multiple formats such as:
Microsoft Excel (.xlsx), Text (.txt,
.csv), JSON (.json), PDF (.pdf), R data
format (.RData), etc.
- Supports Database Connections
(MySQL, Oracle, Redshift, etc.)



Connecting to a DB Server

- Tableau supports database connections to popular databases such as MySQL, Postgres, MongoDB, etc.
- It's necessary to install database drivers first. Tableau will point you to the drivers that you need to install for whatever database you are using.



Joining Data in Tableau

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' section lists a MySQL connection named 'localhost'. The 'Database' dropdown is set to 'employees'. The 'Table' section lists several tables: current_dept_emp, departments, dept_emp, dept_emp_latest_date, dept_manager, employees, salaries, and titles. A 'New Custom SQL' and 'New Union' option are also present. In the center, a diagram illustrates a join between three tables: 'employees', 'salaries', and 'titles'. A modal window titled 'Join' shows four options: Inner, Left, Right, and Full Outer. The 'Left' option is selected, with the condition 'Emp No = Emp No (Titles)'. Below the join diagram, a preview of the data is shown in a grid format. The columns include: Employee Number (Emp No), Birth Date (Birth Date), First Name (First Name), Last Name (Last Name), Gender (Gender), Hire Date (Hire Date), Emp No (Salaries), Salary, From Date (Salaries), and To Date (Salaries). The data shows multiple rows for employee 10,001, each with different salary records. The 'titles' table is not yet joined. The top right corner shows 'Connection Live' and 'Filters 0 | Add'.

Show Me

- Show Me palette makes it easy to choose the visualization that you want.
- Tableau automatically adjusts dimensions and measures to better fit your data to the map.



Dimensions and Measures

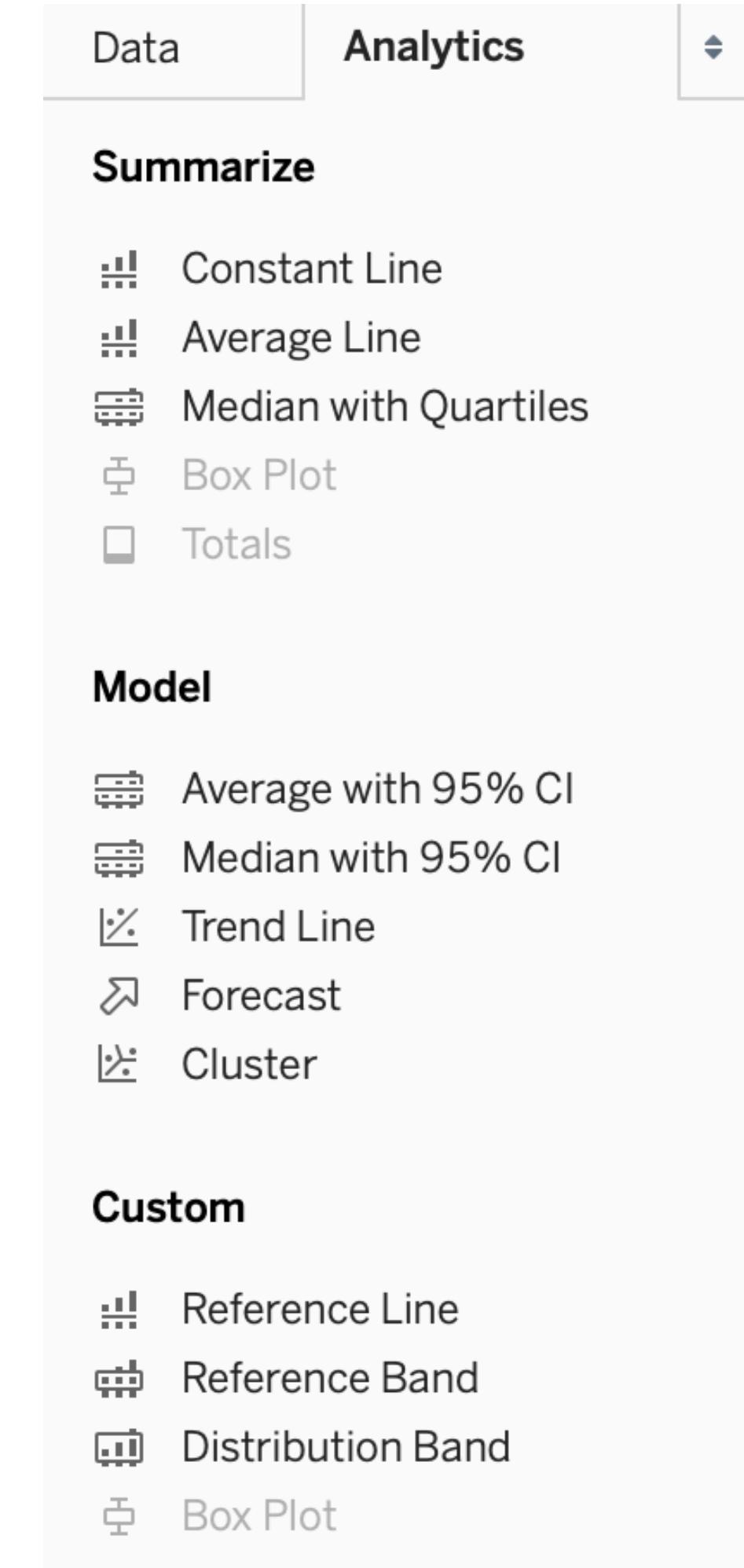
- **Dimensions** are used for categorical data.
- **Measures** are used for quantitative data.

Dimensions	
📅	Ship Date
Abc	Ship Mode
▼	Location
🌐	Country
🌐	State
🌐	City
🌐	Postal Code
▼	Product
Abc	Category
Abc	Sub-Category
👤	Manufacturer
Abc	Product Name
.bars.	Profit (bin)
Abc	Region
Abc	Measure Names

Measures	
#	Quantity
#	Sales
🌐	Latitude (generated)
🌐	Longitude (generated)
-#	Number of Records
#	Measure Values

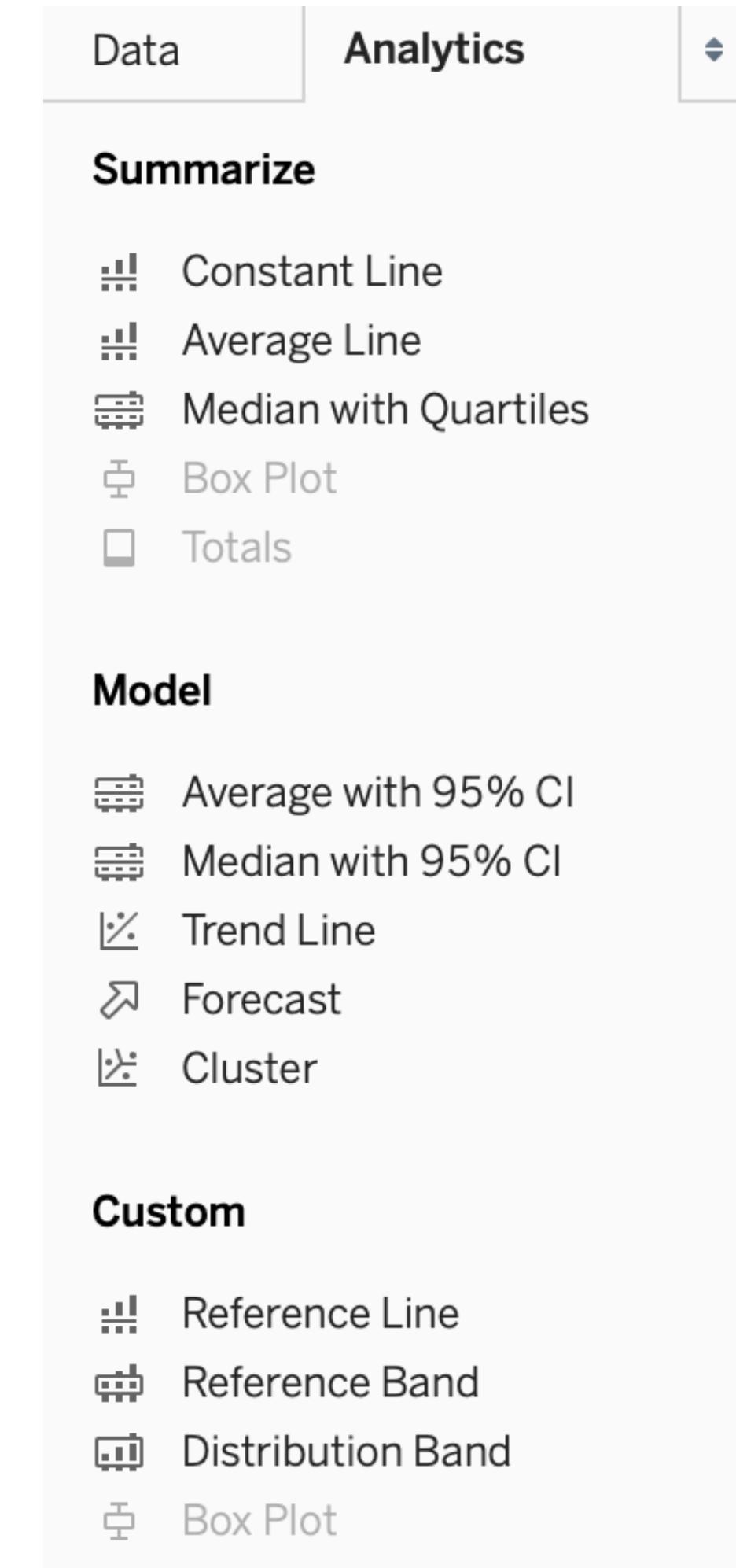
Analytics

- Drag reference lines, box plots, trend lines forecasts, and other items into your view from the **Analytics** pane, which appears on the left side of the workspace.



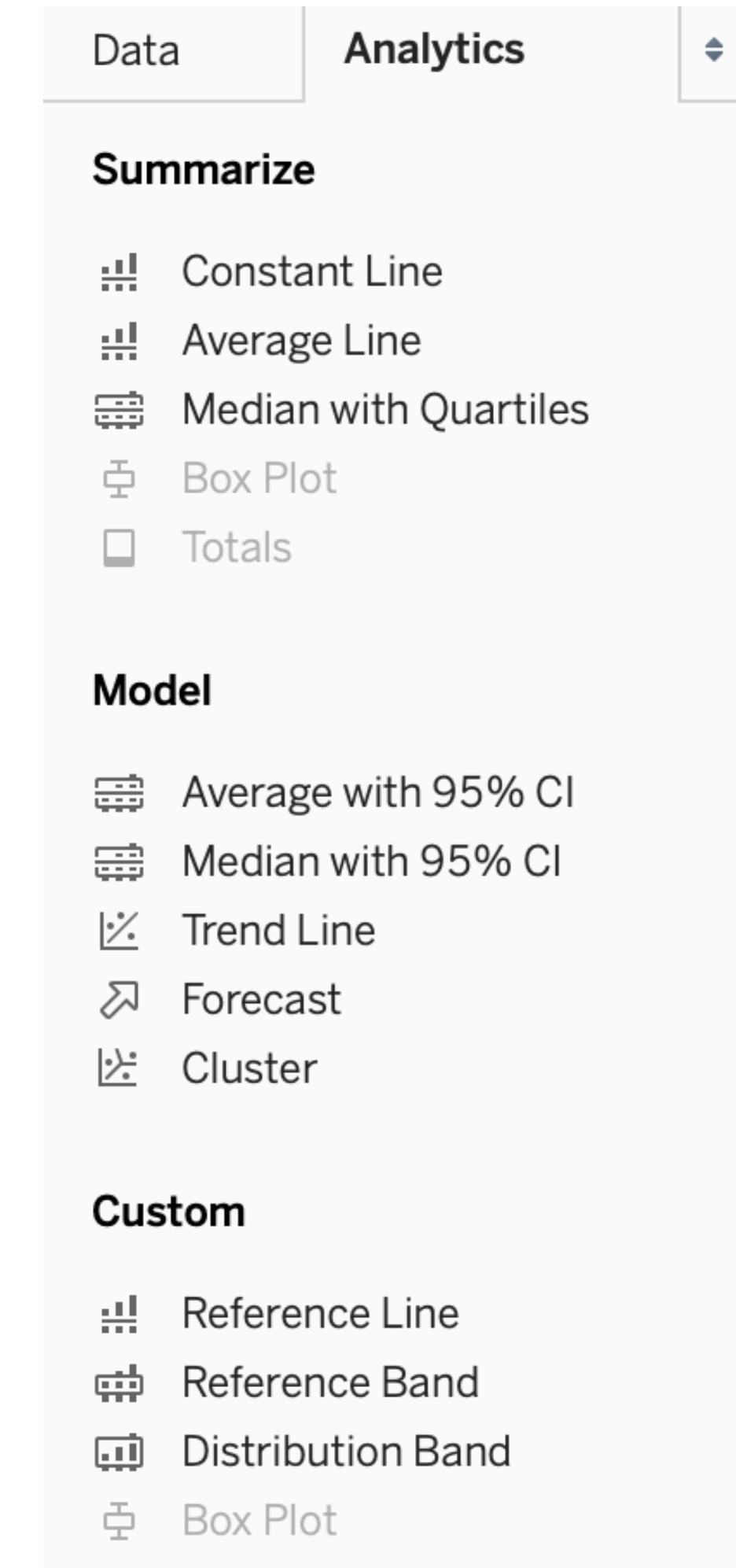
Analytics: Constant Line

- Adds one or more constant lines to the view. You can add a constant line for a specific measure, for all measures, or for date dimensions. When you add a constant line, Tableau displays a Value prompt where you specify the value for the constant line.



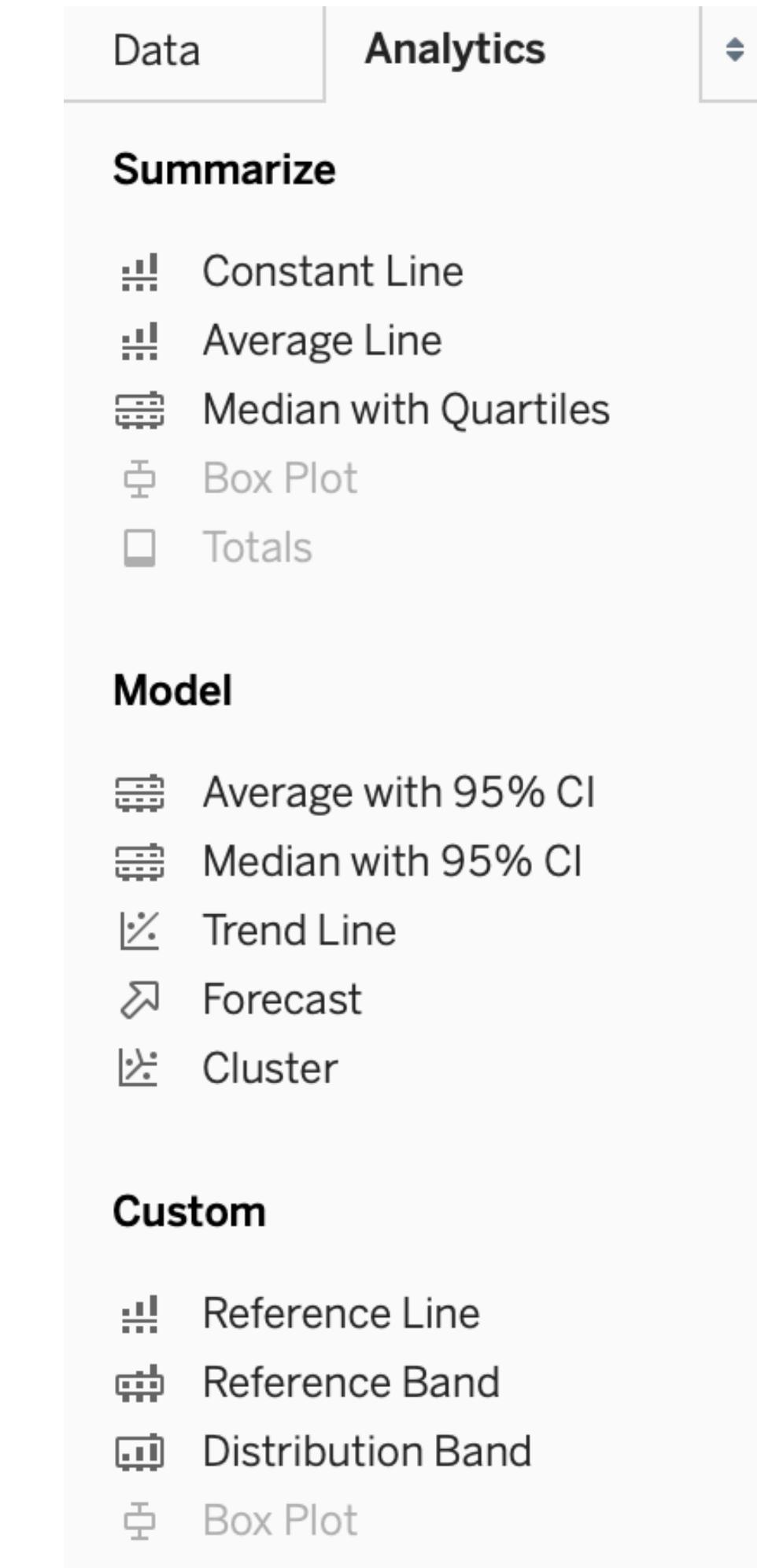
Analytics: Average Line

- Adds one or more average lines to the view. You can add an average line for a specific measure or for all measures



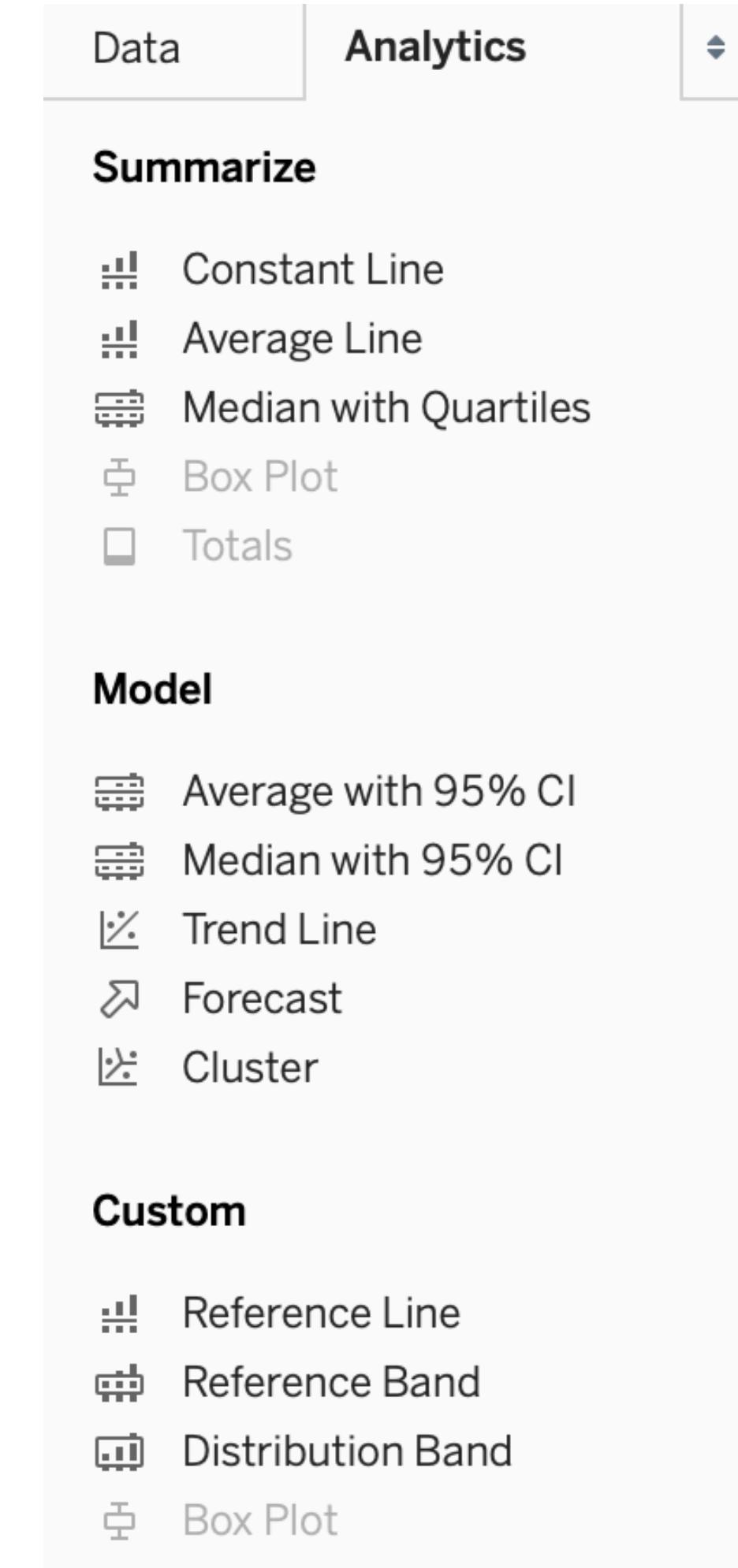
Analytics: Median w/ Quartiles

- Adds one or more sets of median lines and distribution bands to the view. You can add a median with quartiles for a specific measure or for all measures.
- The distribution bands are computed as quartiles; the middle two quartiles are shaded.



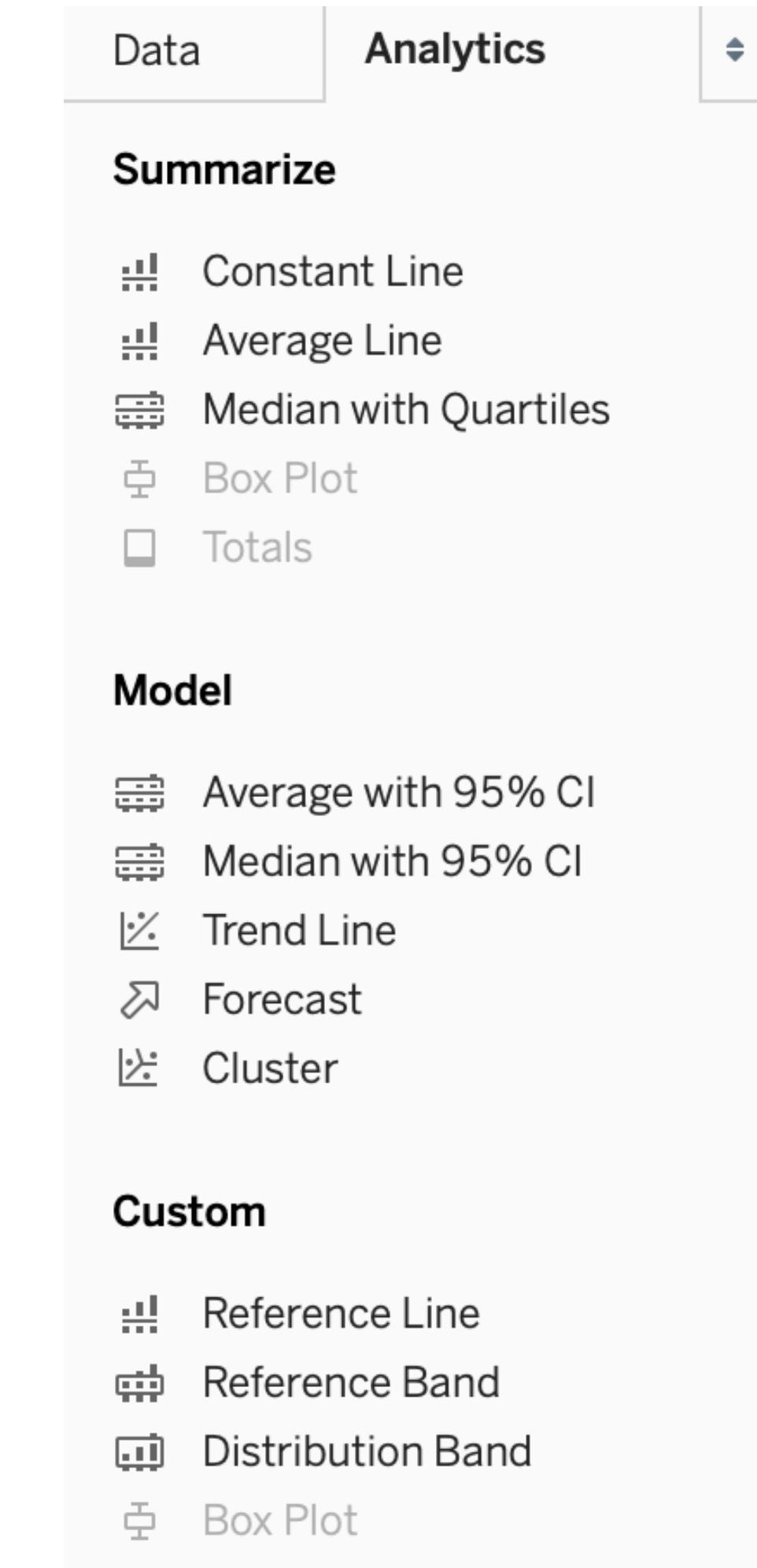
Analytics: Box Plots

- Adds one or more box plots to the view. You can add box plots for a specific measure or for all measures. The scope for a box plot is always **Cell** (and never **Table** or **Pane**).



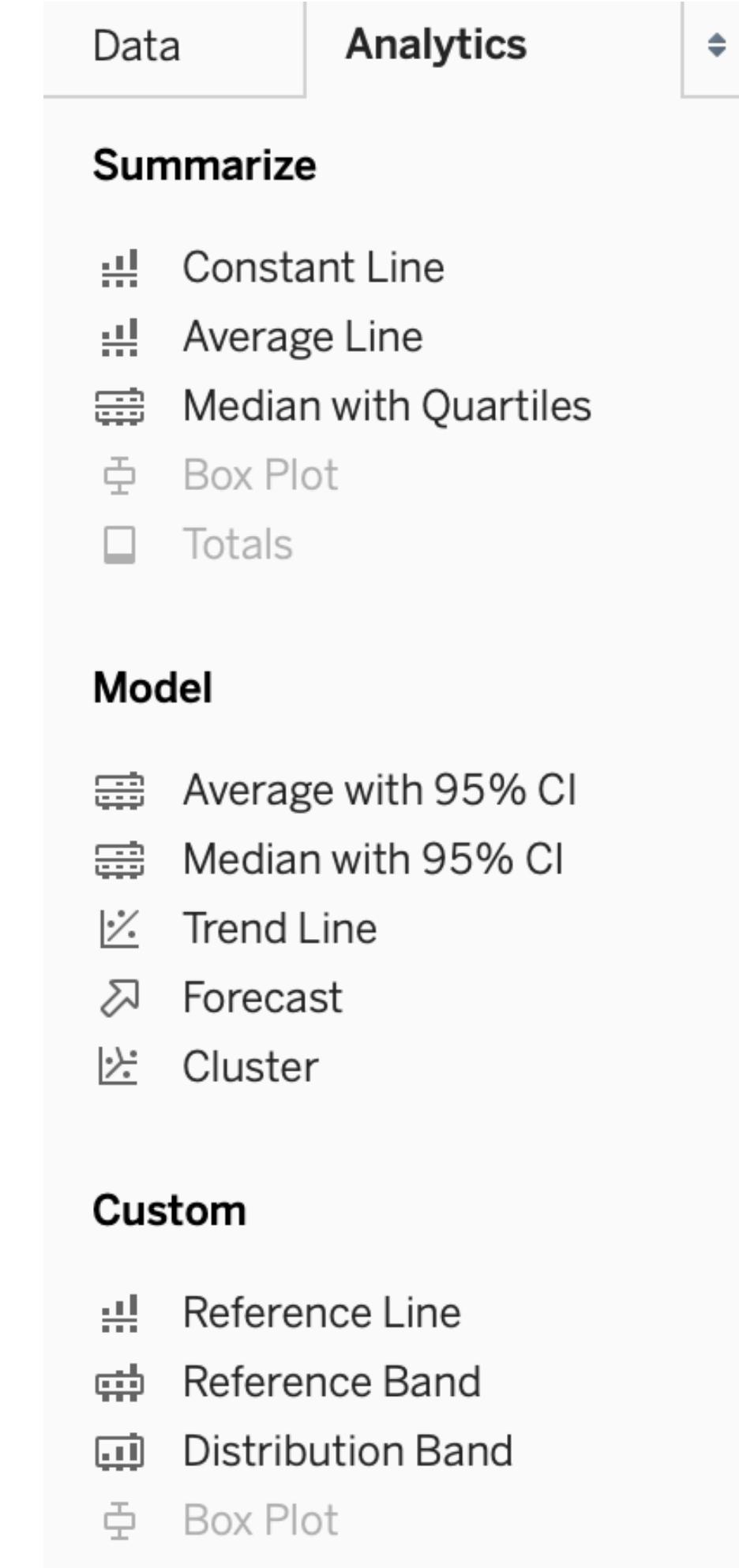
Analytics: Average w/ 95% CI

- Adds one or more sets of average lines with distribution bands; the distribution bands are configured at a 95% confidence interval. You can add these items for a specific measure or for all measures.



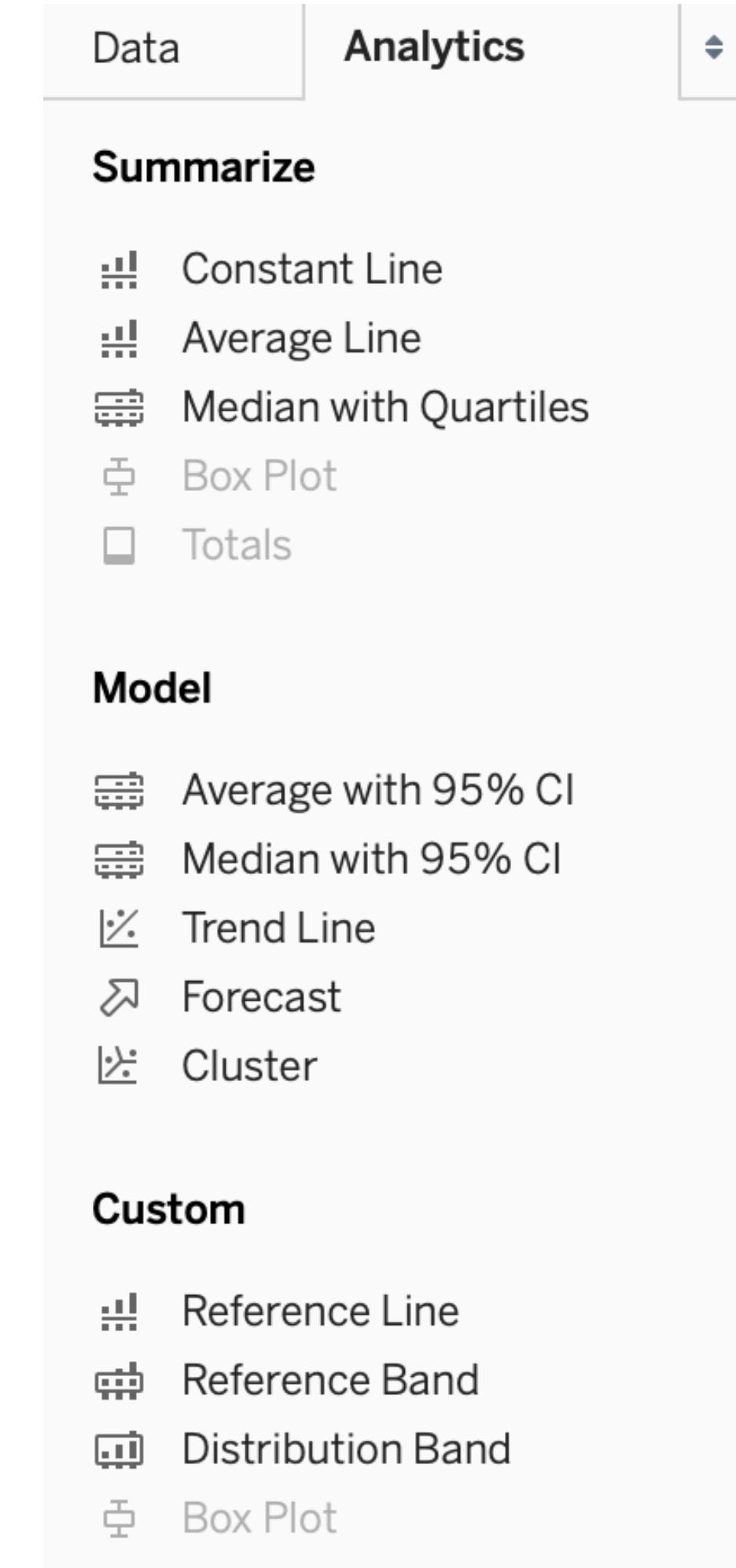
Analytics: Trend Line

- Adds one or more trend lines to the view. When you add trend lines, the drop options identify the trend line model types available in Tableau: **Linear, Logarithmic, Exponential, and Polynomial**. For some views, only a subset of these options is available.

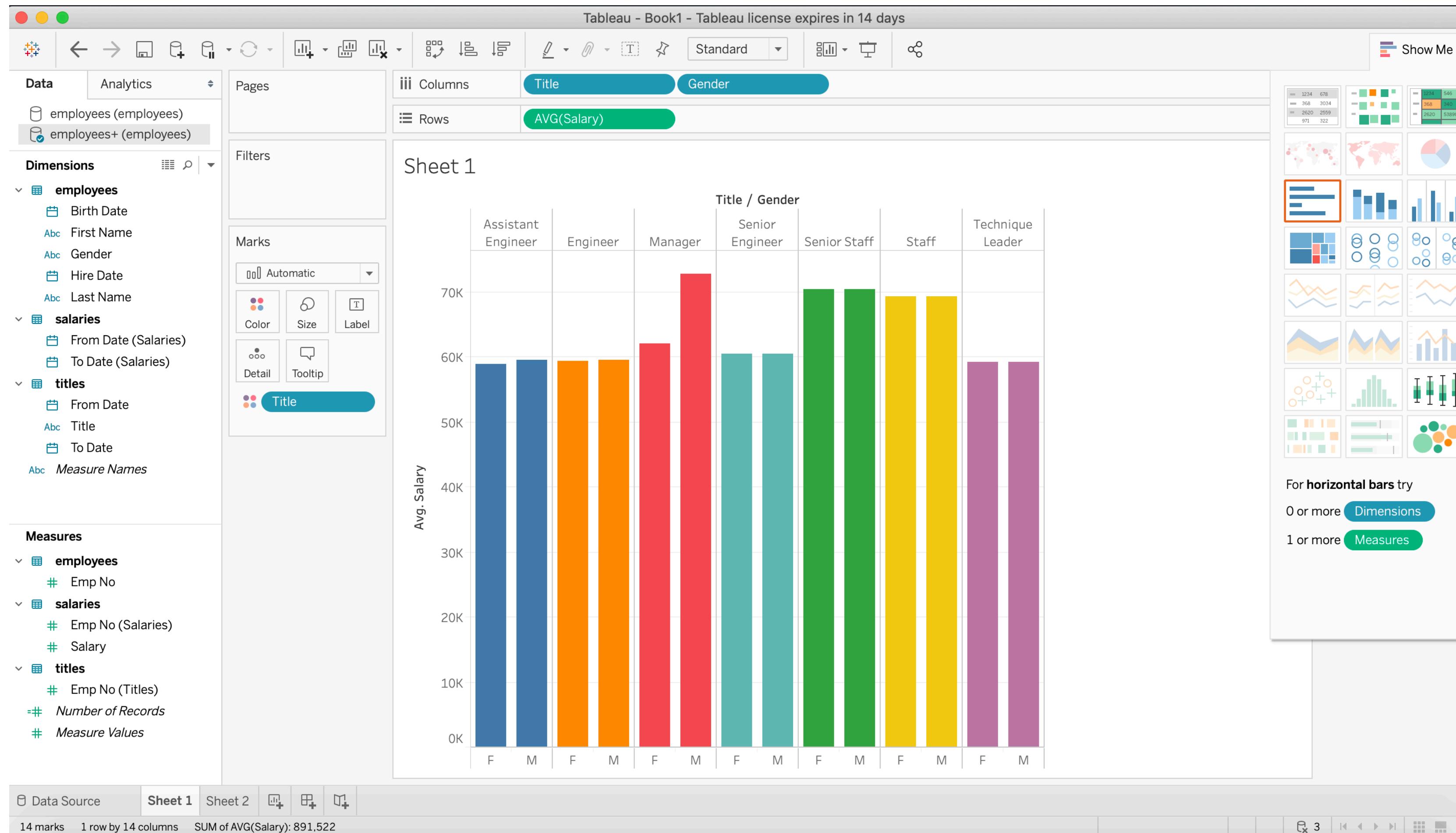


Analytics: Forecast

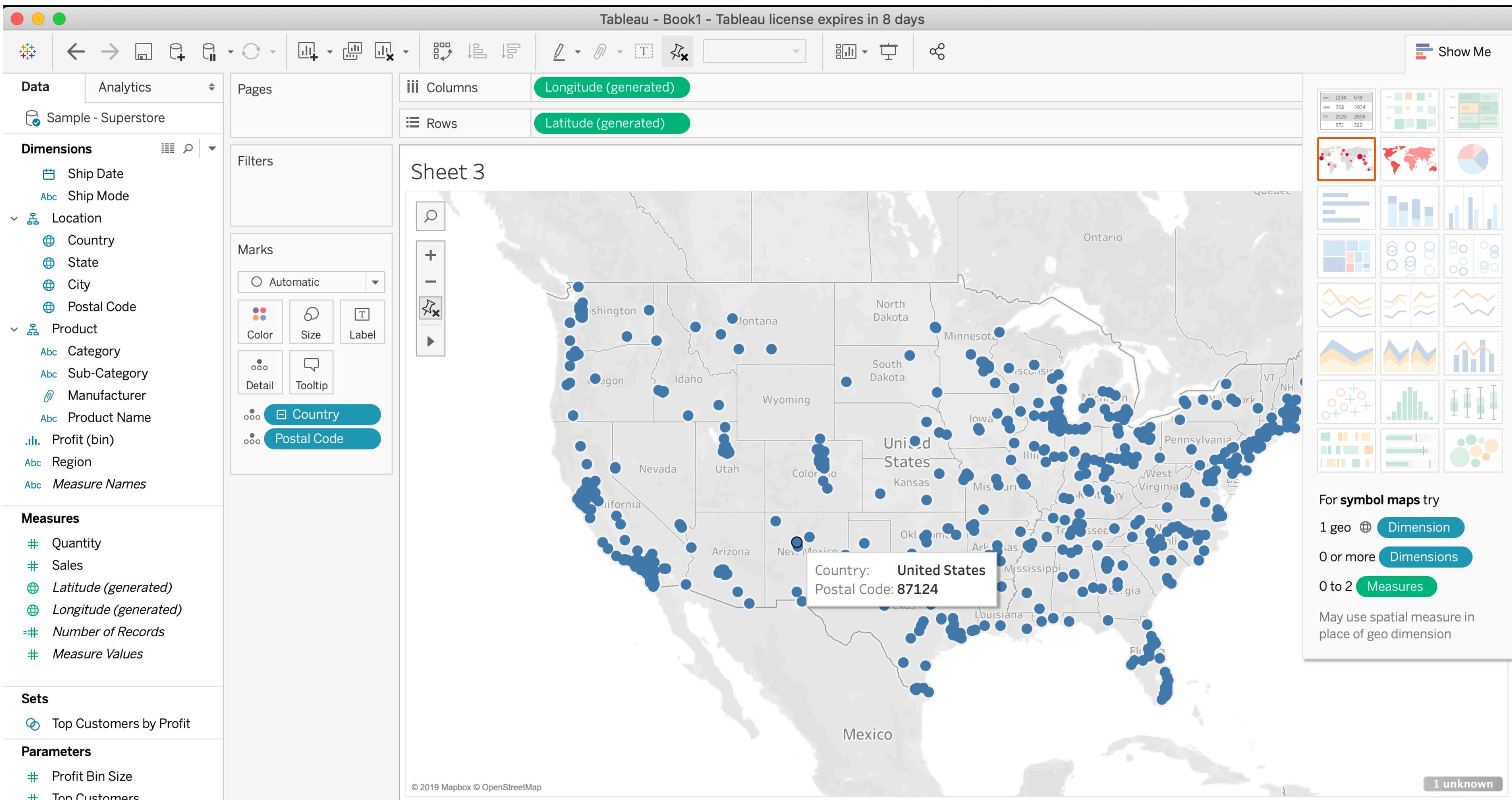
- Adds a forecast to the view. This option is only available in Tableau Desktop—not when you edit a view on the web. Forecasting is only possible when there is at least one measure in the view.



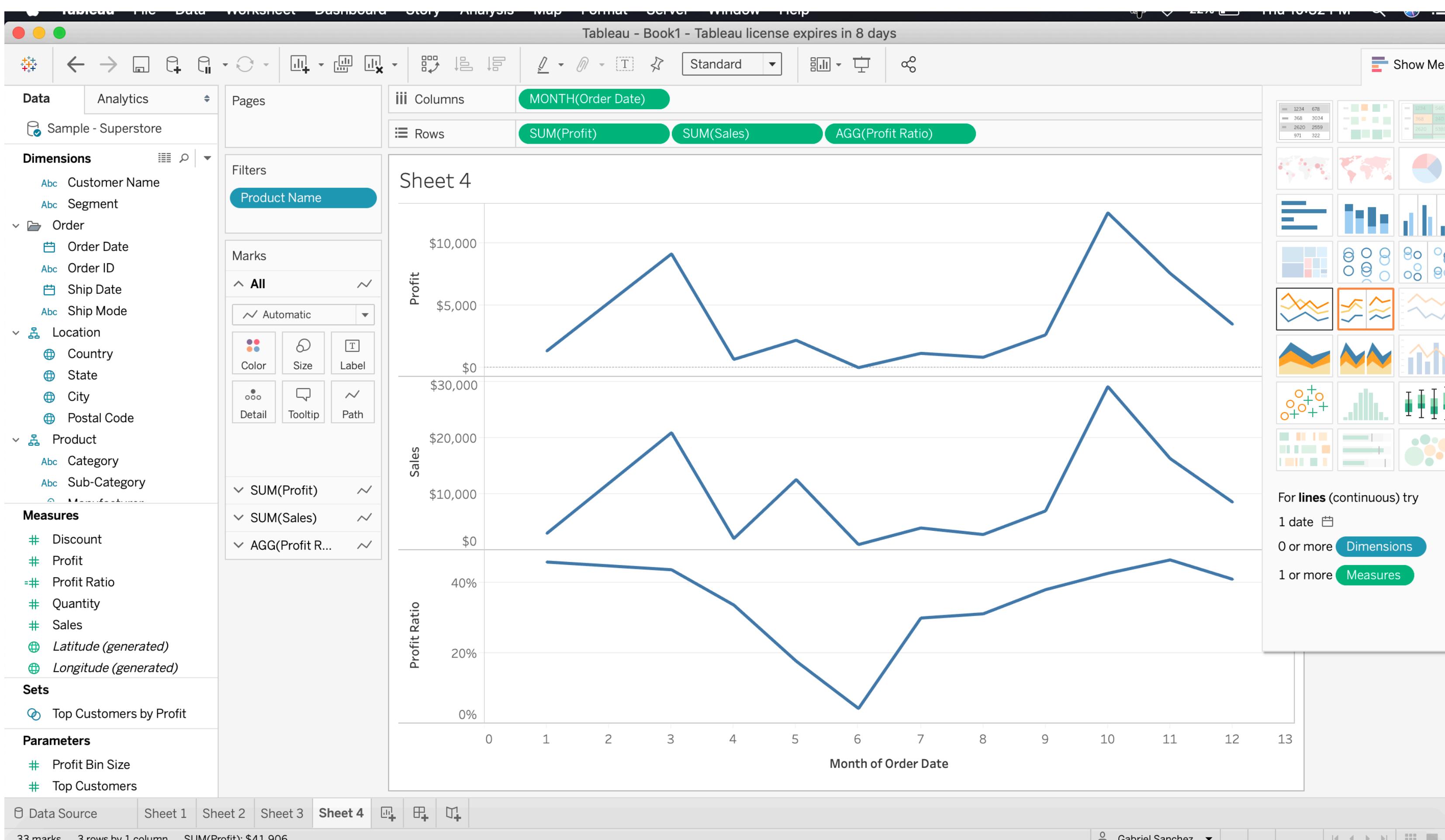
Data Visualization: Bar Charts



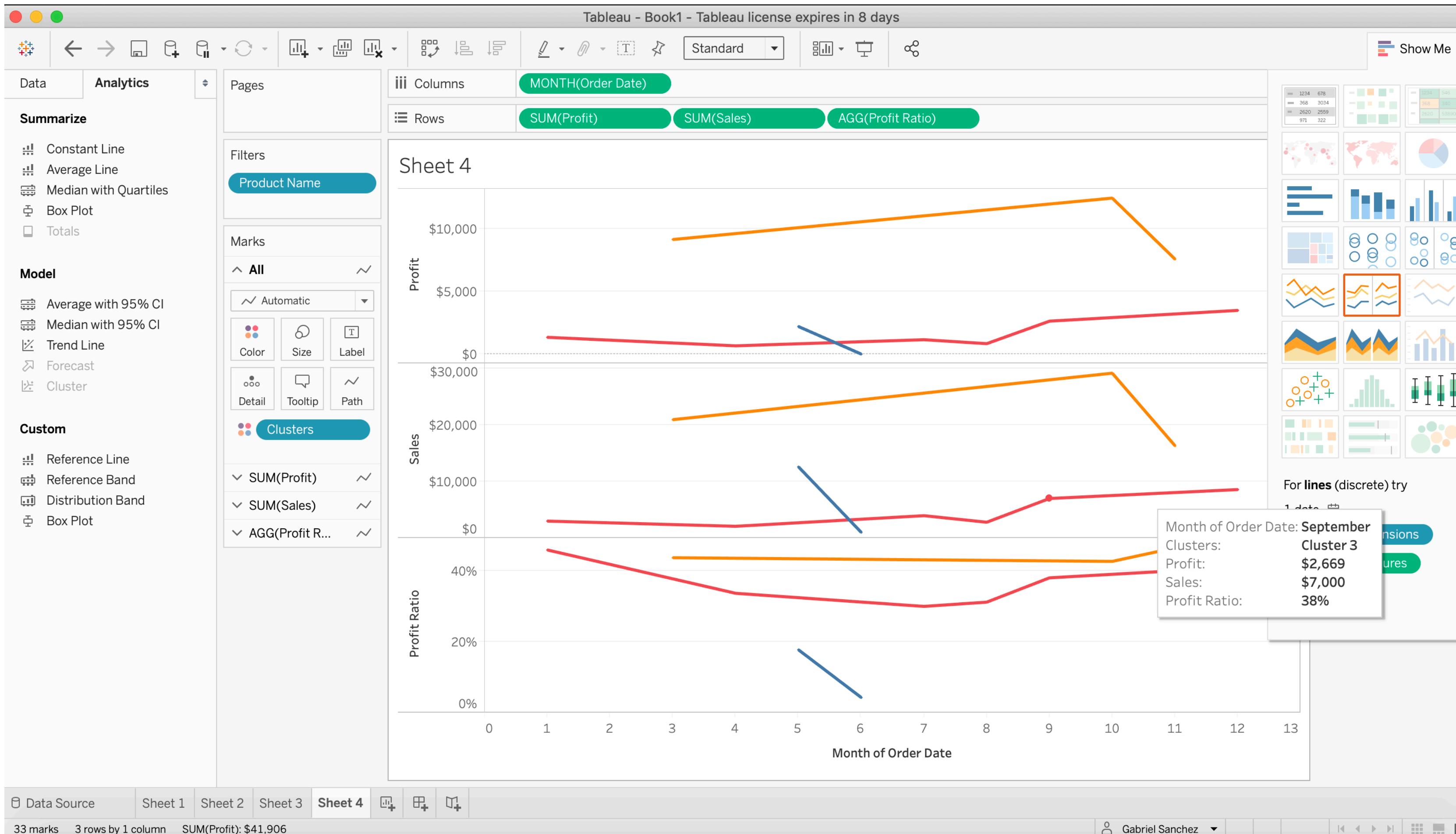
Data Visualization: Symbol Map



Data Visualization: Line Chart

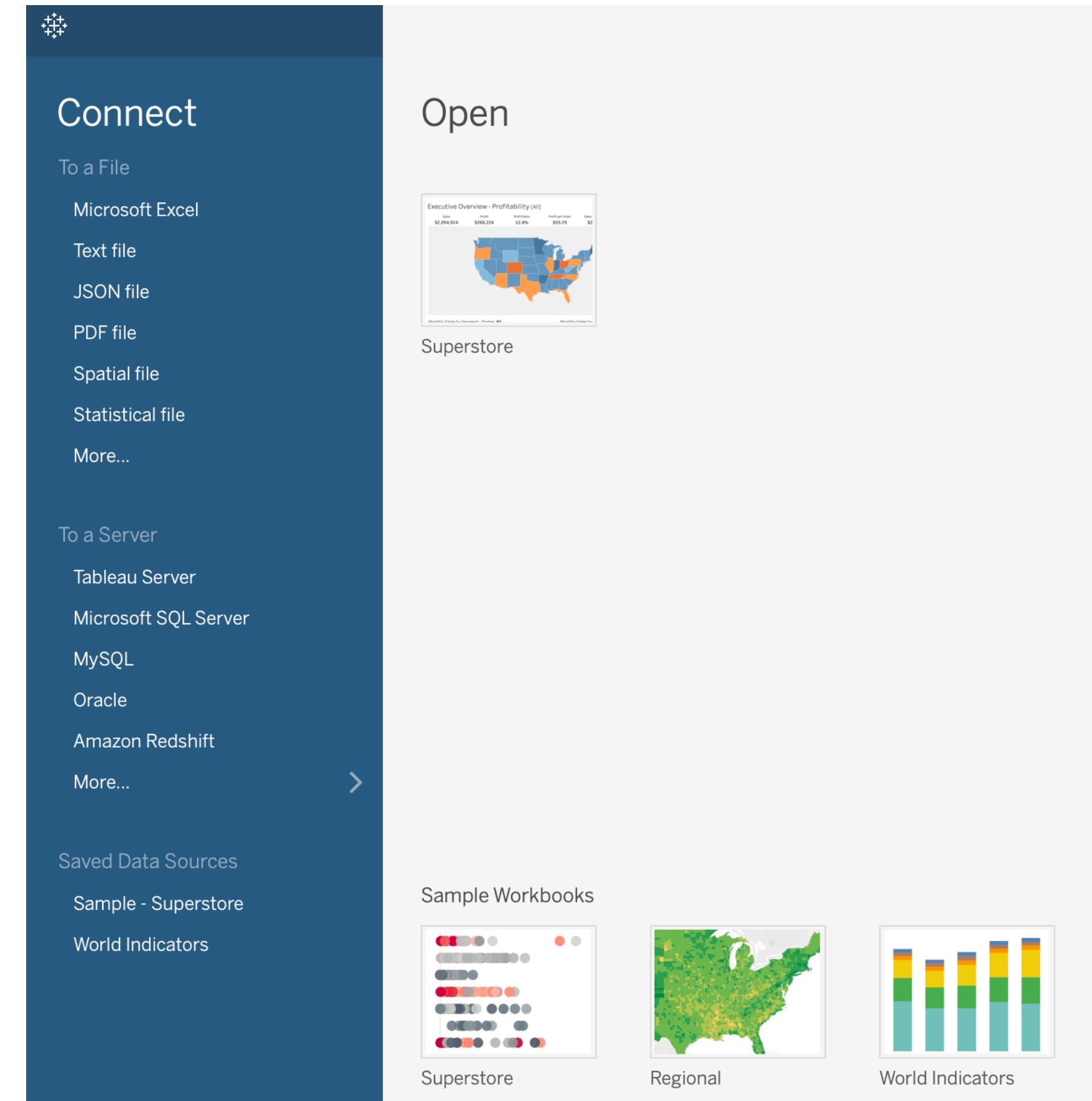


Data Visualization: Cluster

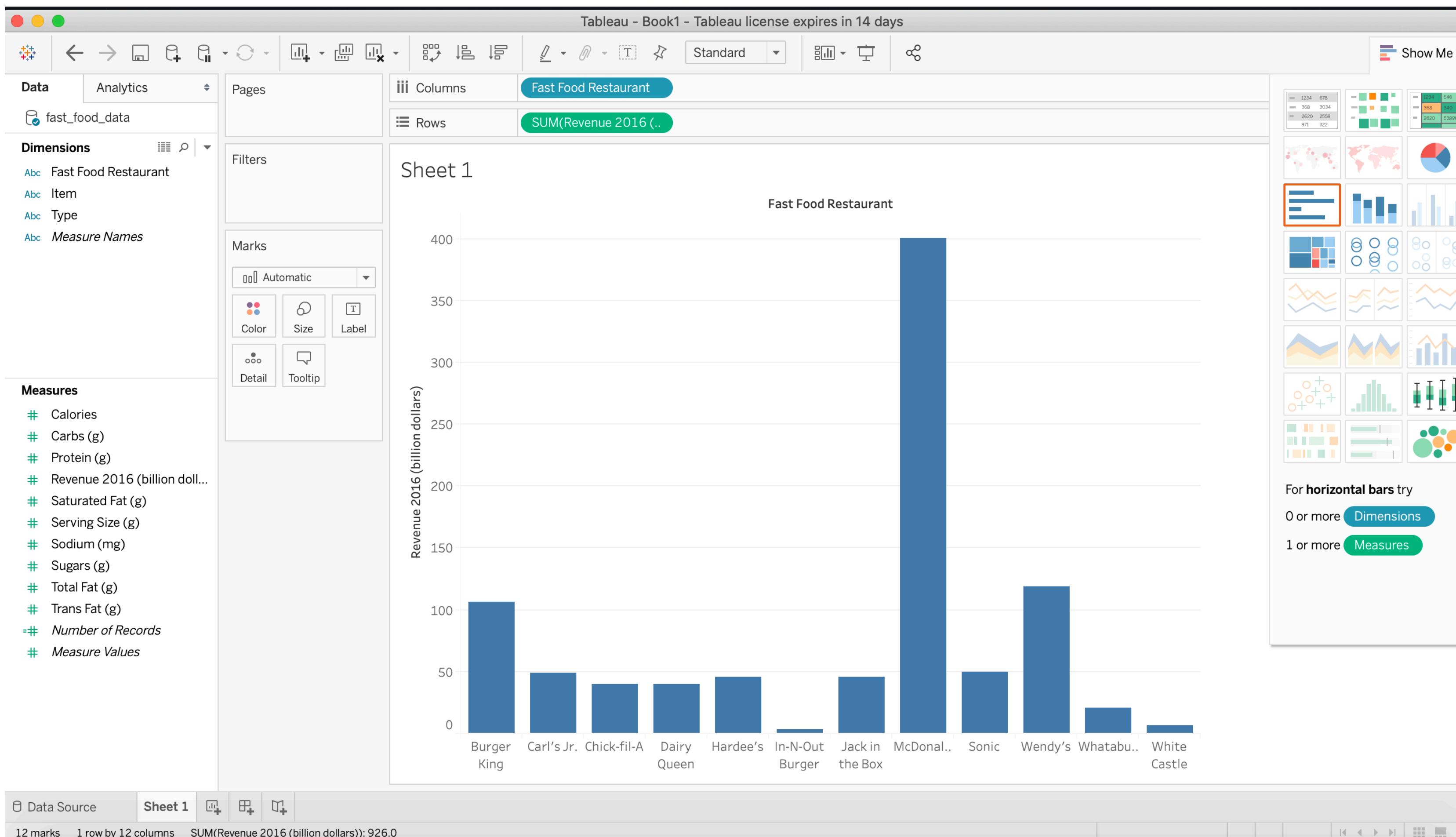


Fast Food Data: Import CSV

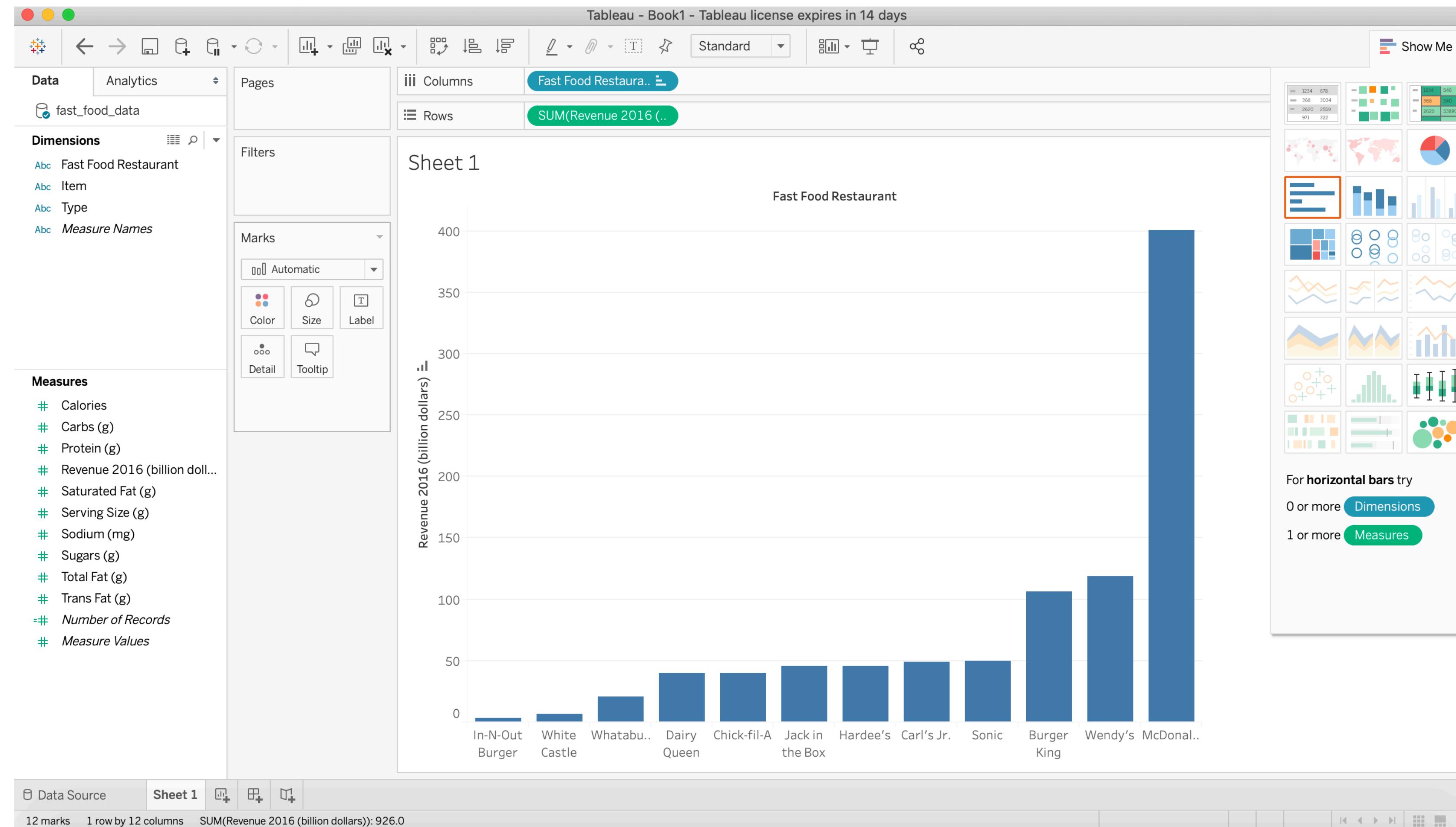
- Let's import the fast food dataset that we used in a previous session and let's see if we can get any insights from it.
- Suppose we want to visualize what restaurants make the most revenue.



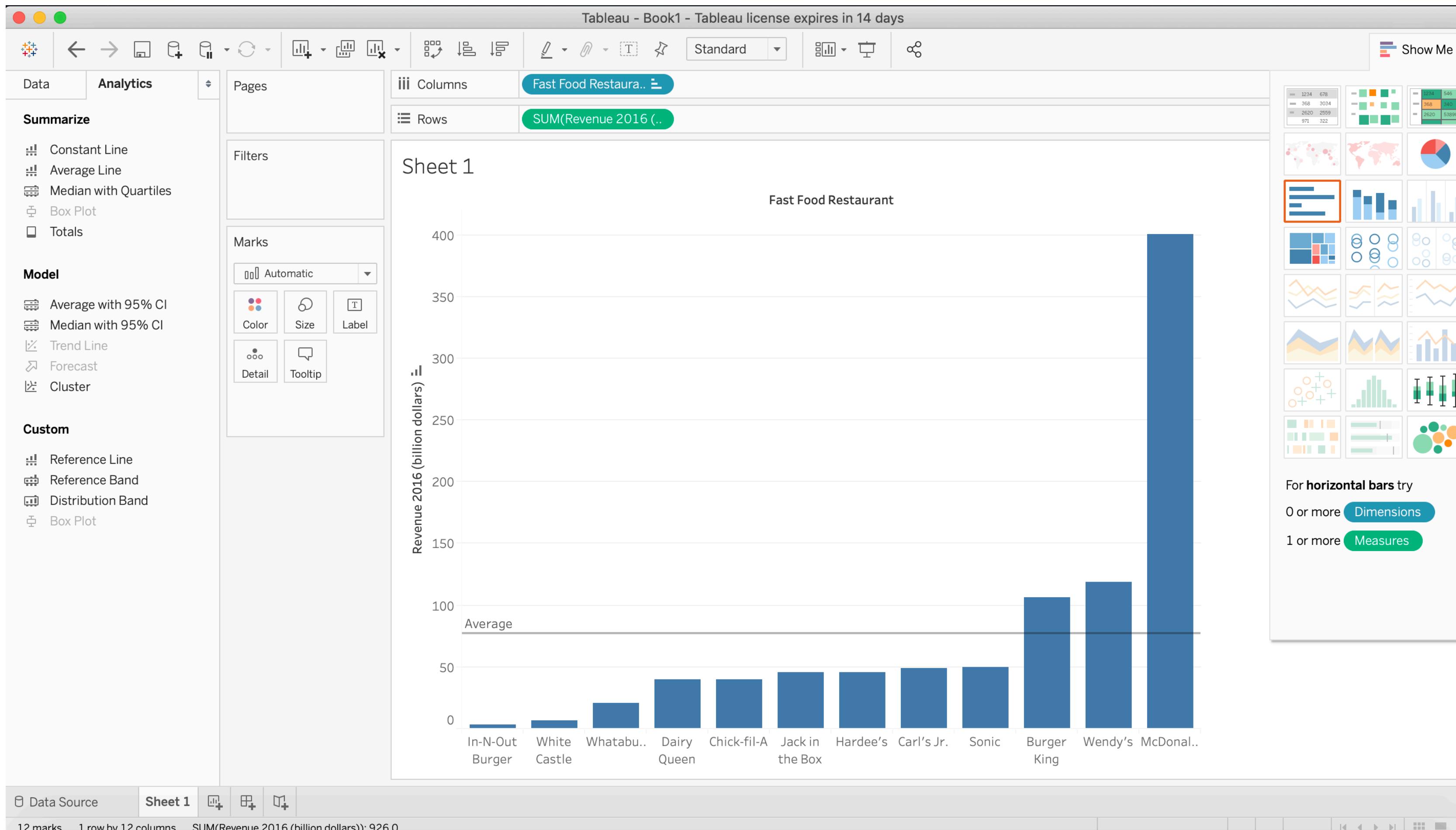
Fast Food Data: Visualization



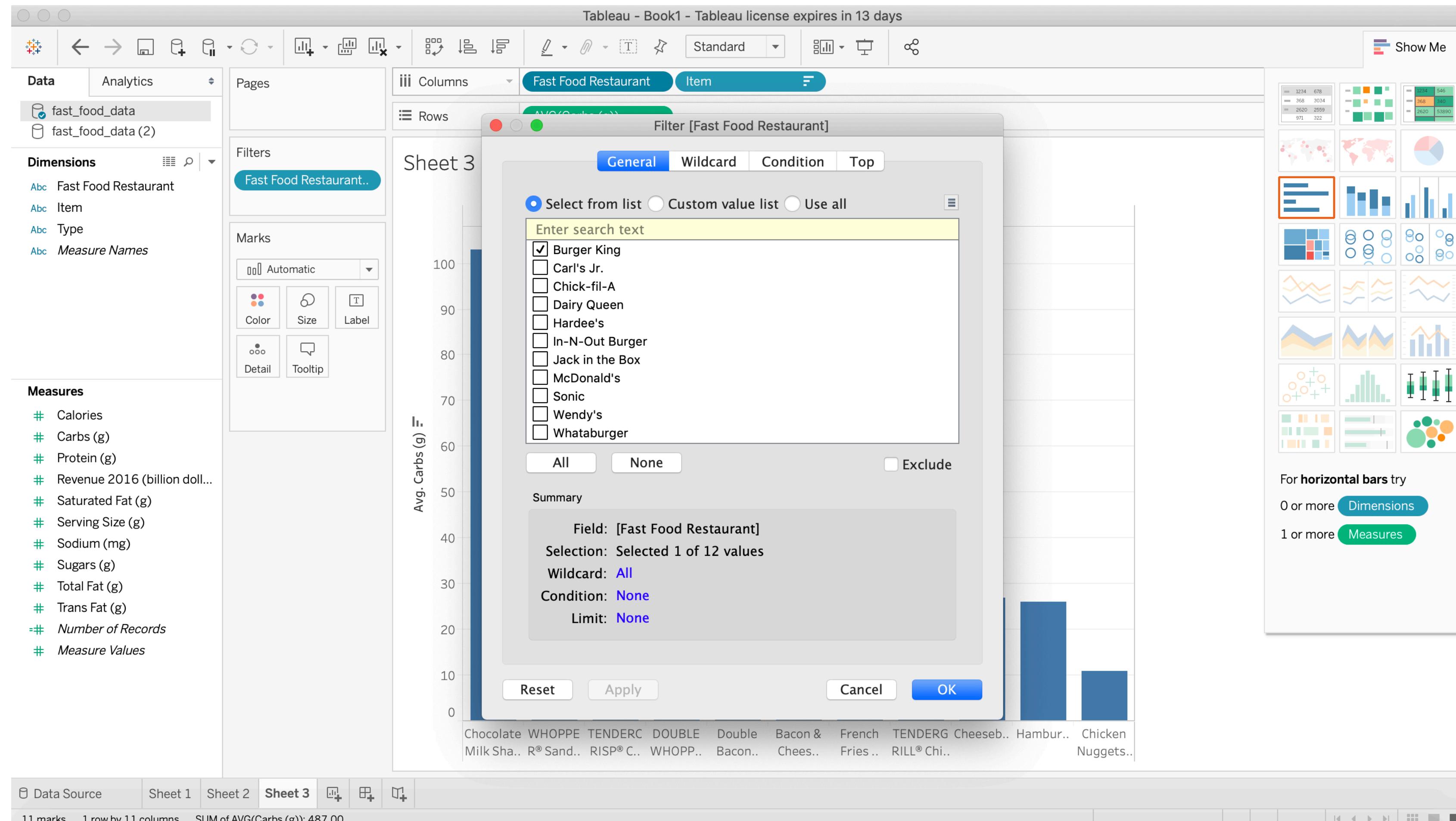
Fast Food Data: Sorting



Fast Food Data: Analytics



Fast Food Data: Filters



Knowledge check 1

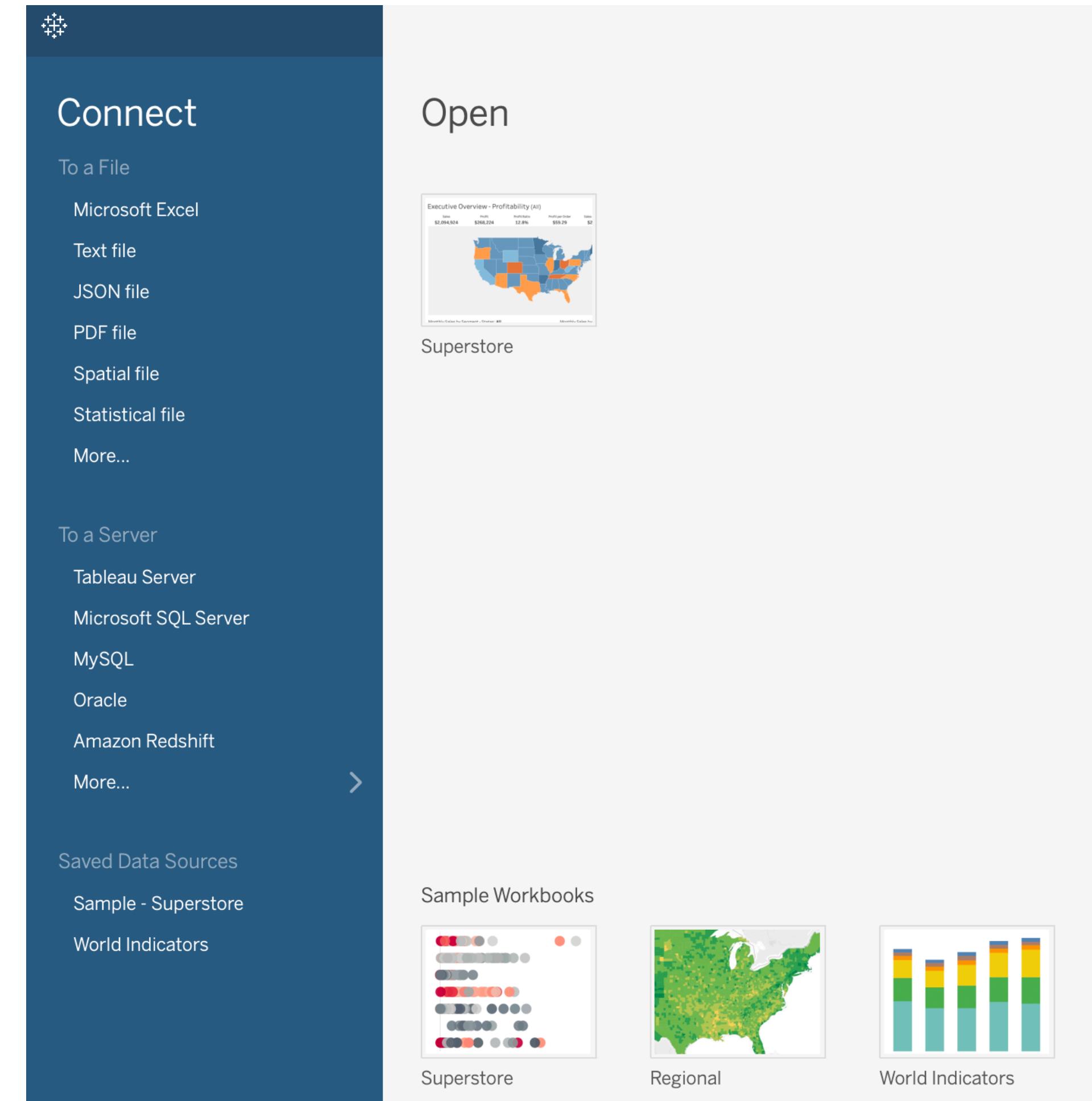


Exercise 1



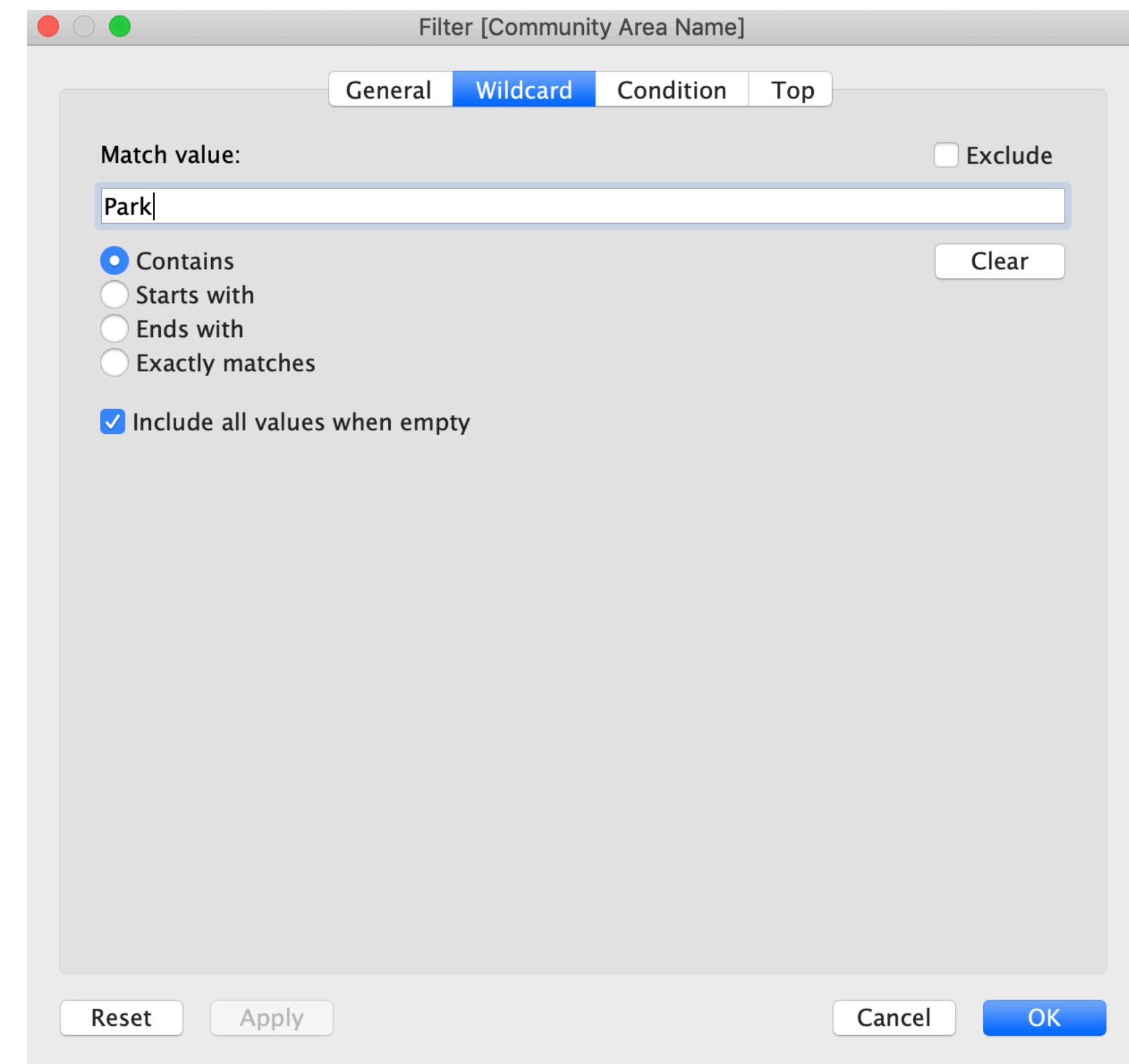
Chicago Census Data: Import CSV

- Let's import the Chicago Census dataset that we used in a previous session and let's see if we can get any insights from it.
- Suppose we want to visualize Per Capita Income for every Community Area



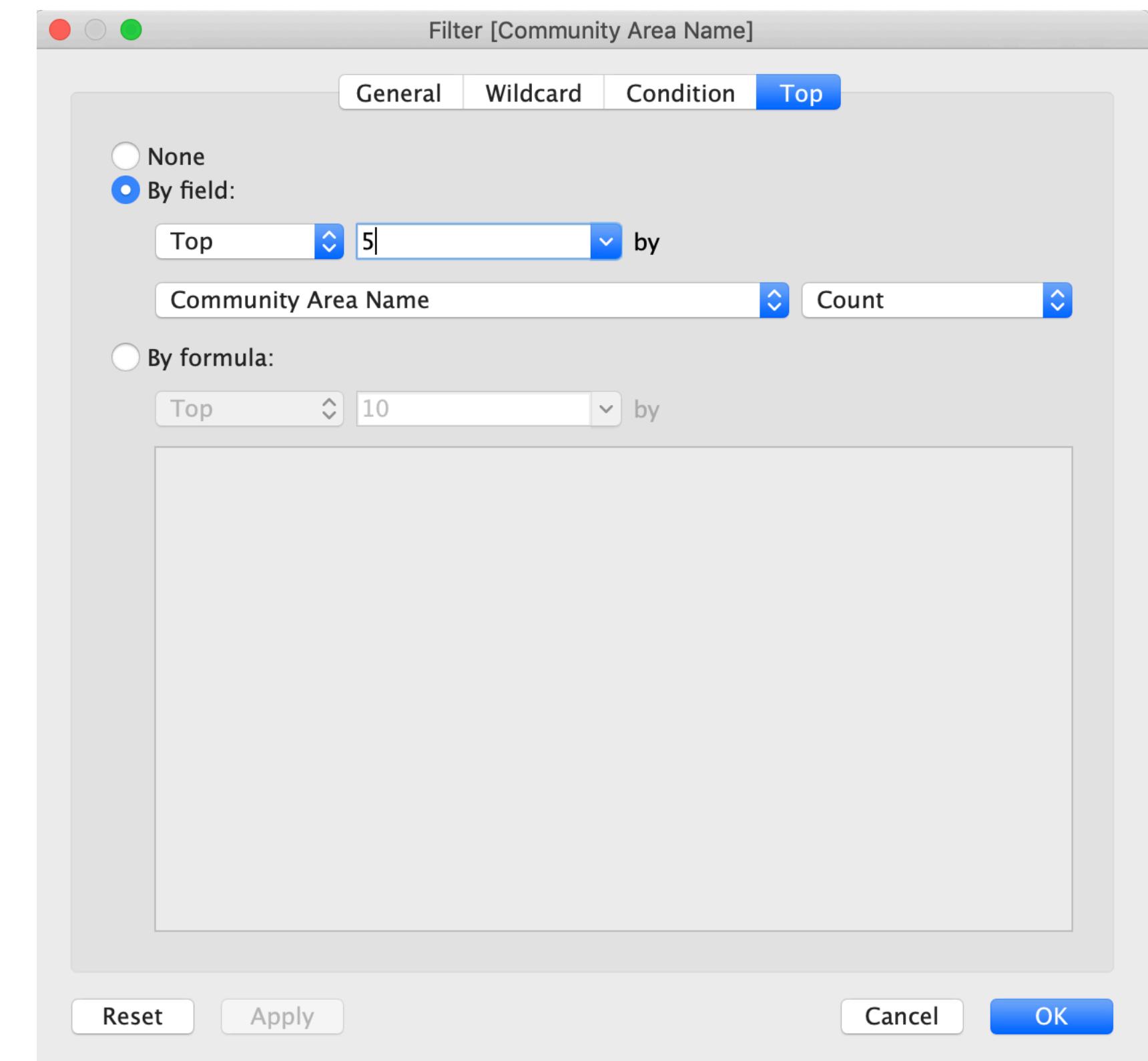
Chicago Census Data: More Filtering

- We can use different type of filters to get more specific data.
- Adding a match value will only include or exclude those columns that match that value.



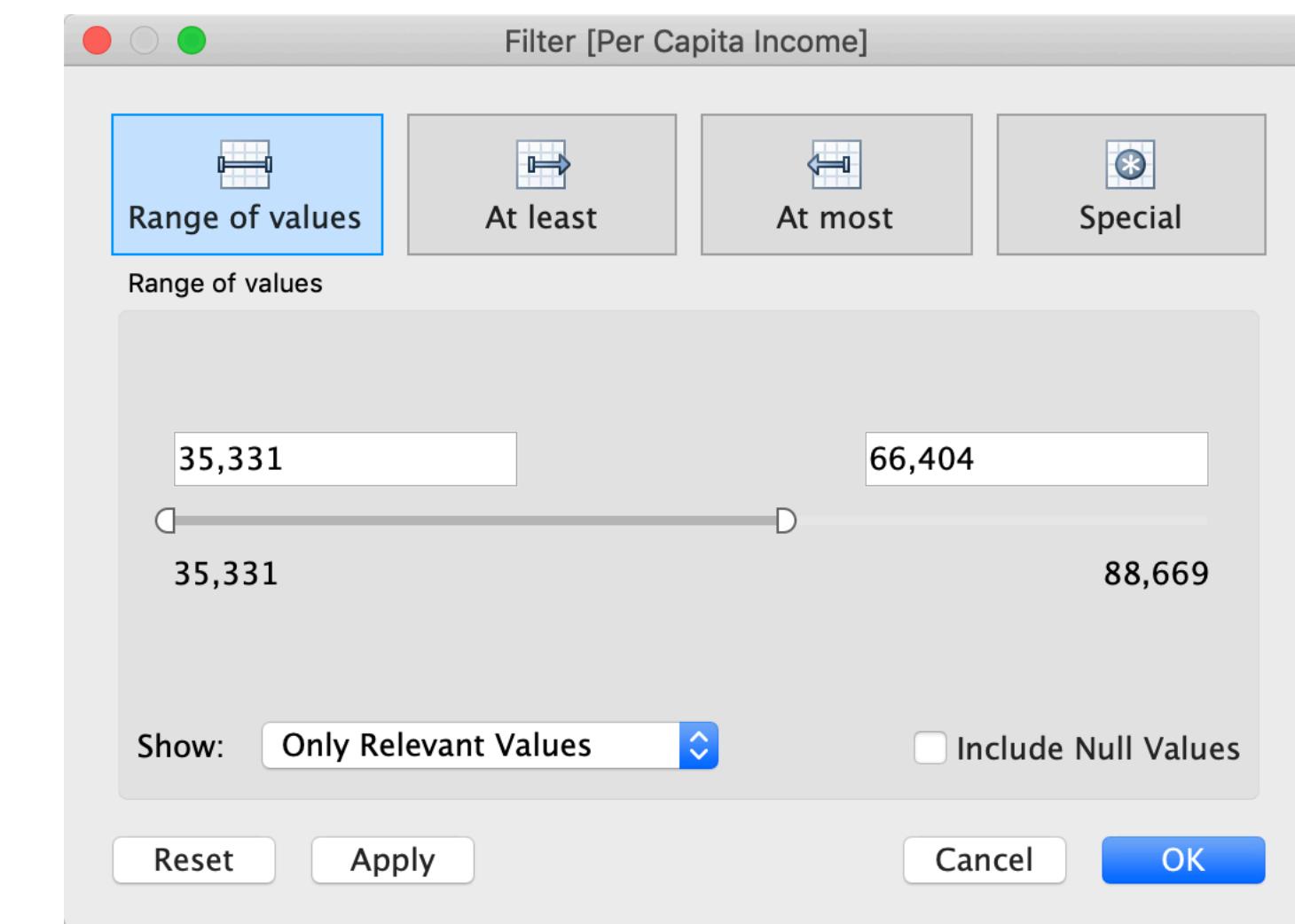
Chicago Census Data: More Filtering

- If we want to display only top results we can use Top option and enter the Top number of values that we want.

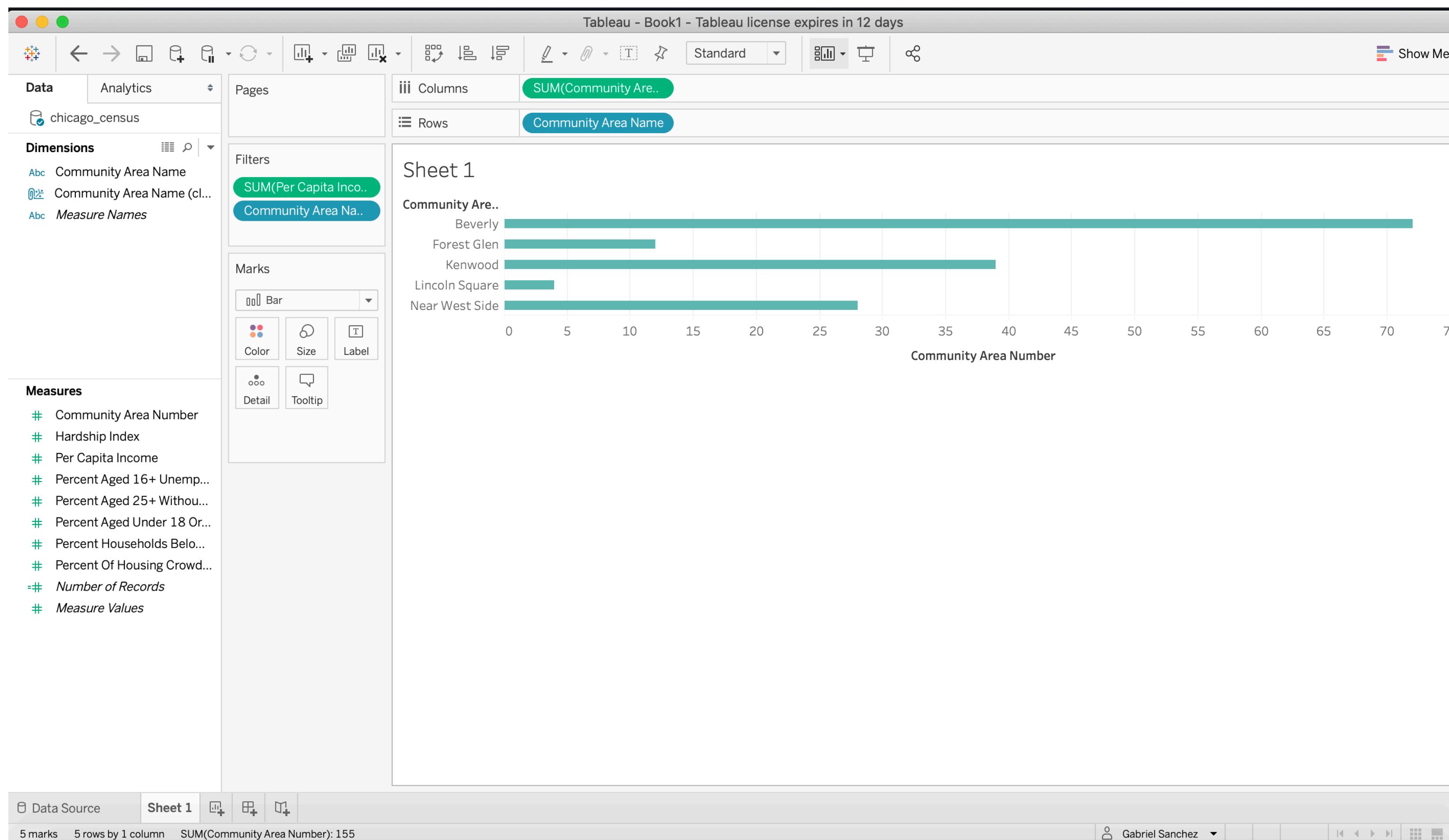


Chicago Census Data: More Filtering

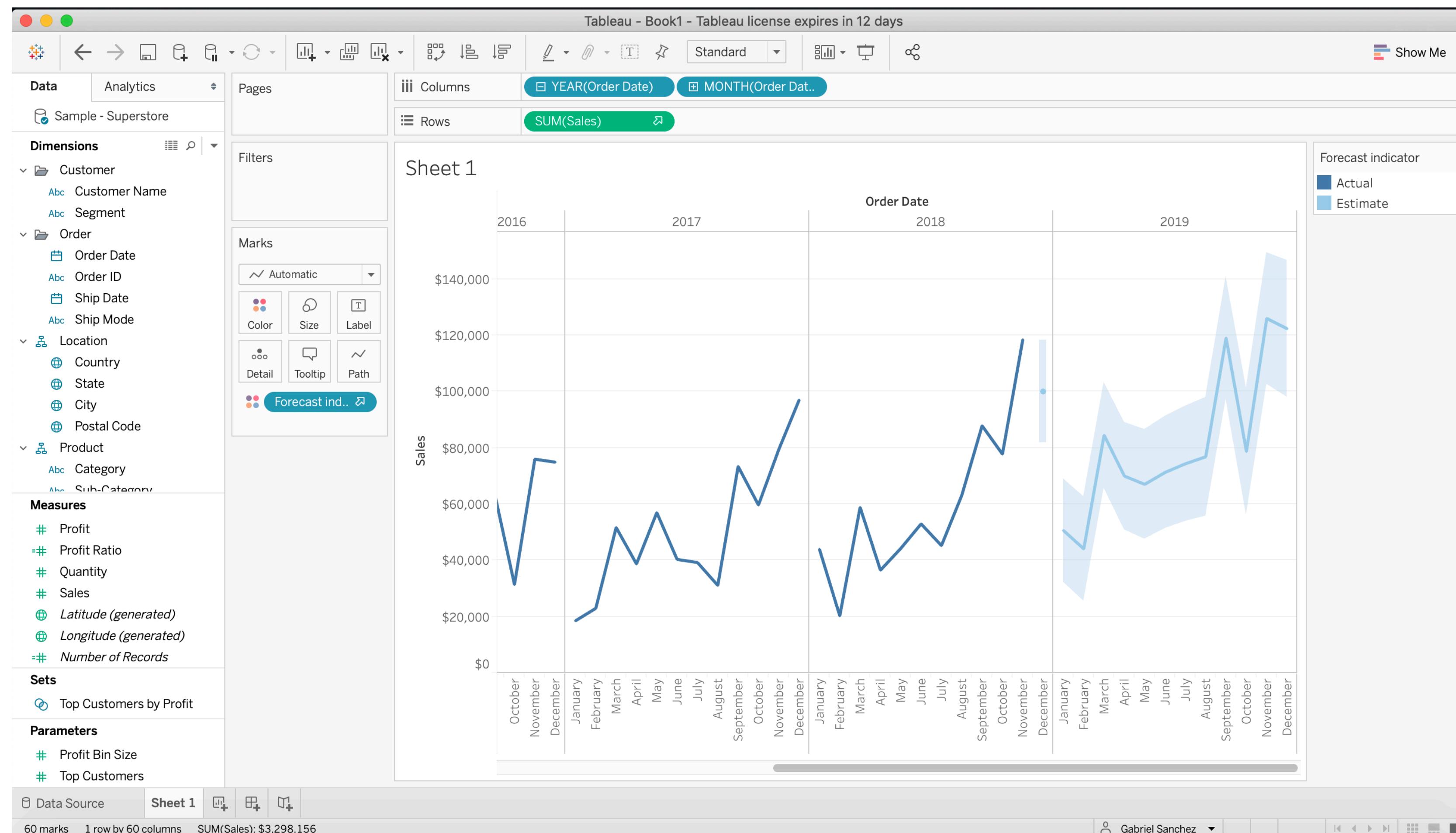
- We can also use specific ranges to filter our data. This only works when our data is quantitative.
- There are different ways of filtering by range. We can also choose to filter out null values.



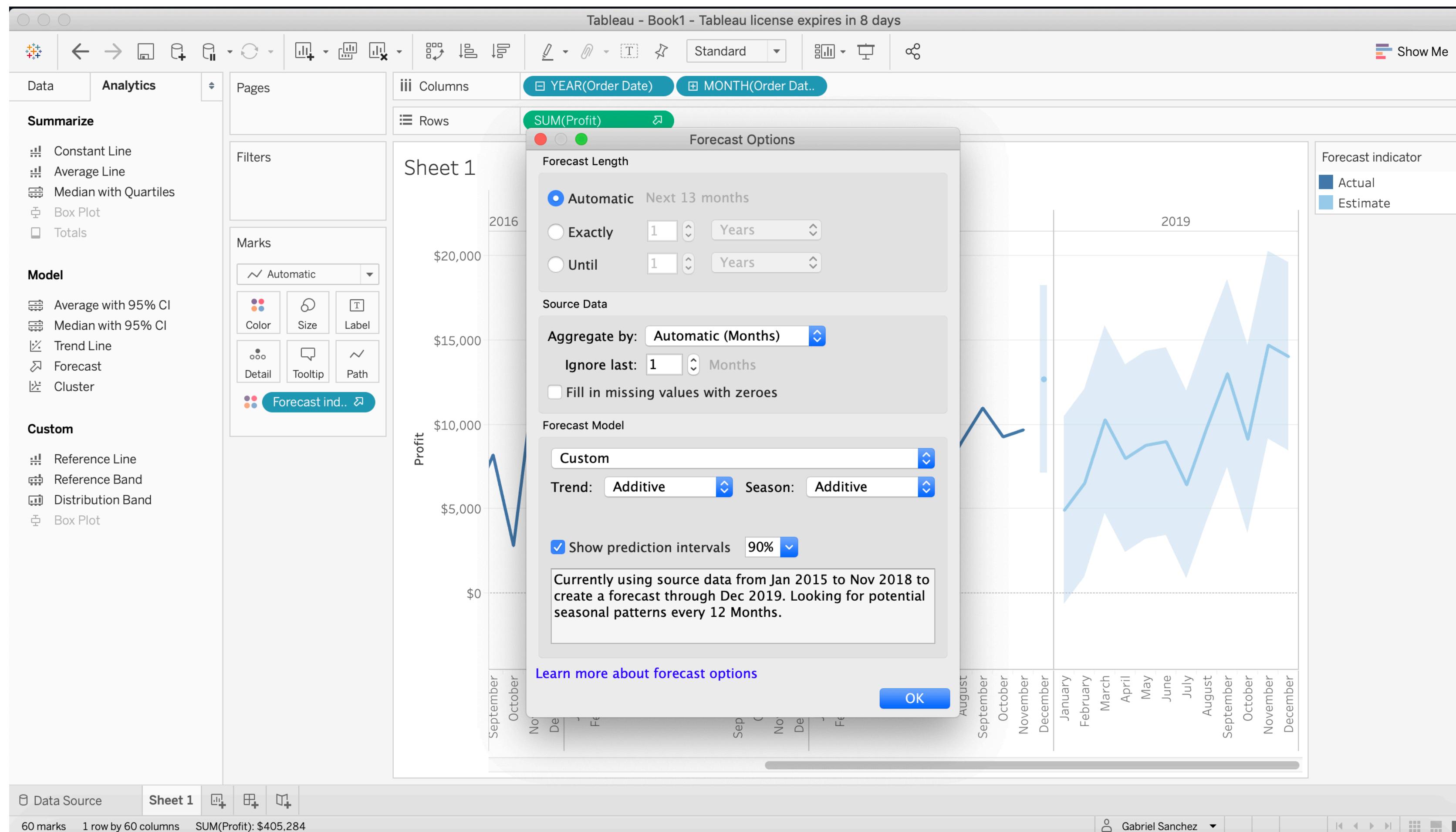
Chicago Census Data: Marks



Superstore Data: Forecast

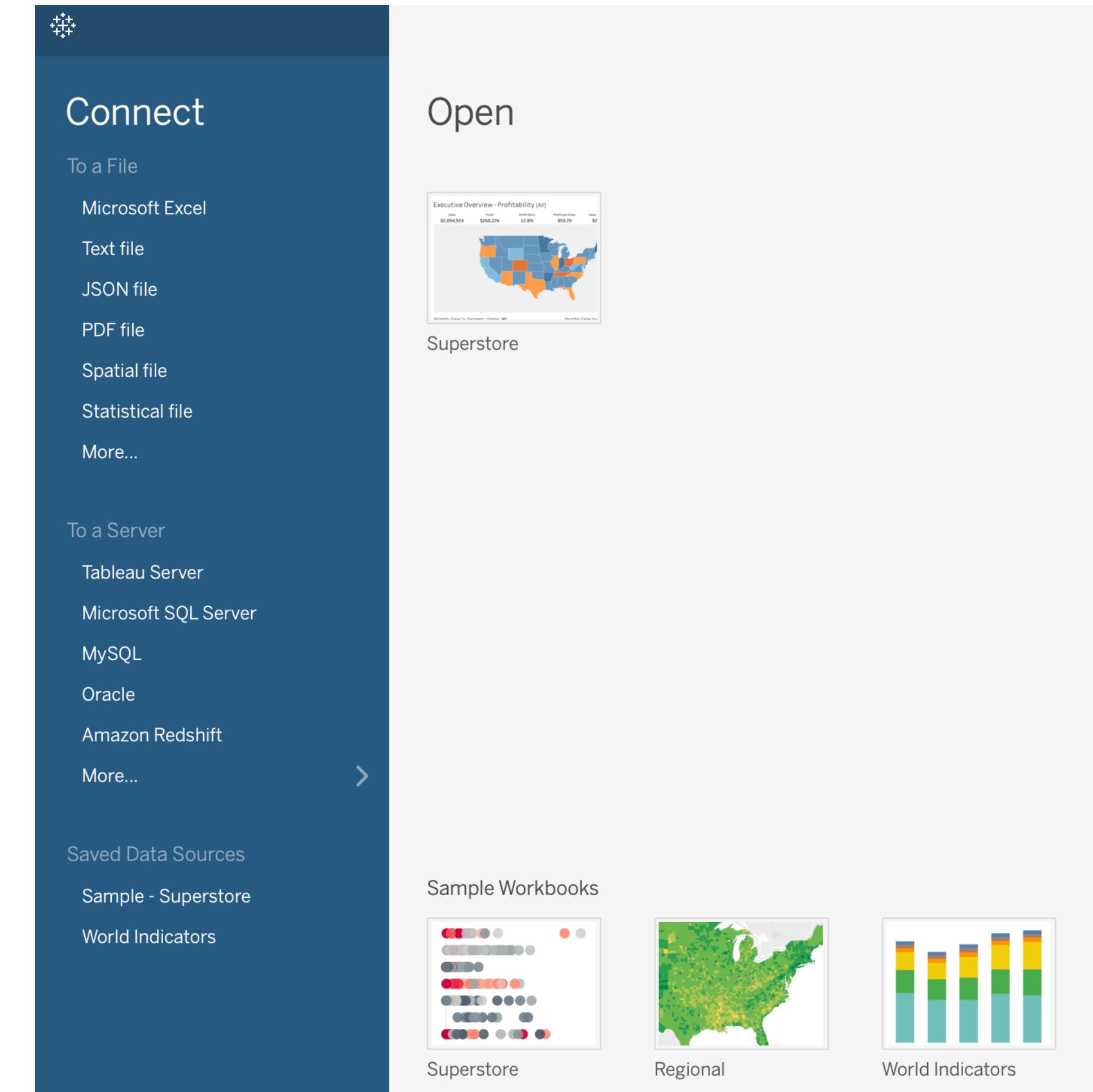


Superstore Data: Forecast Options



Superstore Data: Import CSV

- Let's import the Chicago Census dataset that we used in a previous session and let's see if we can get any insights from it.
- Suppose we want to visualize Per Capita Income for every Community Area



Knowledge check 2

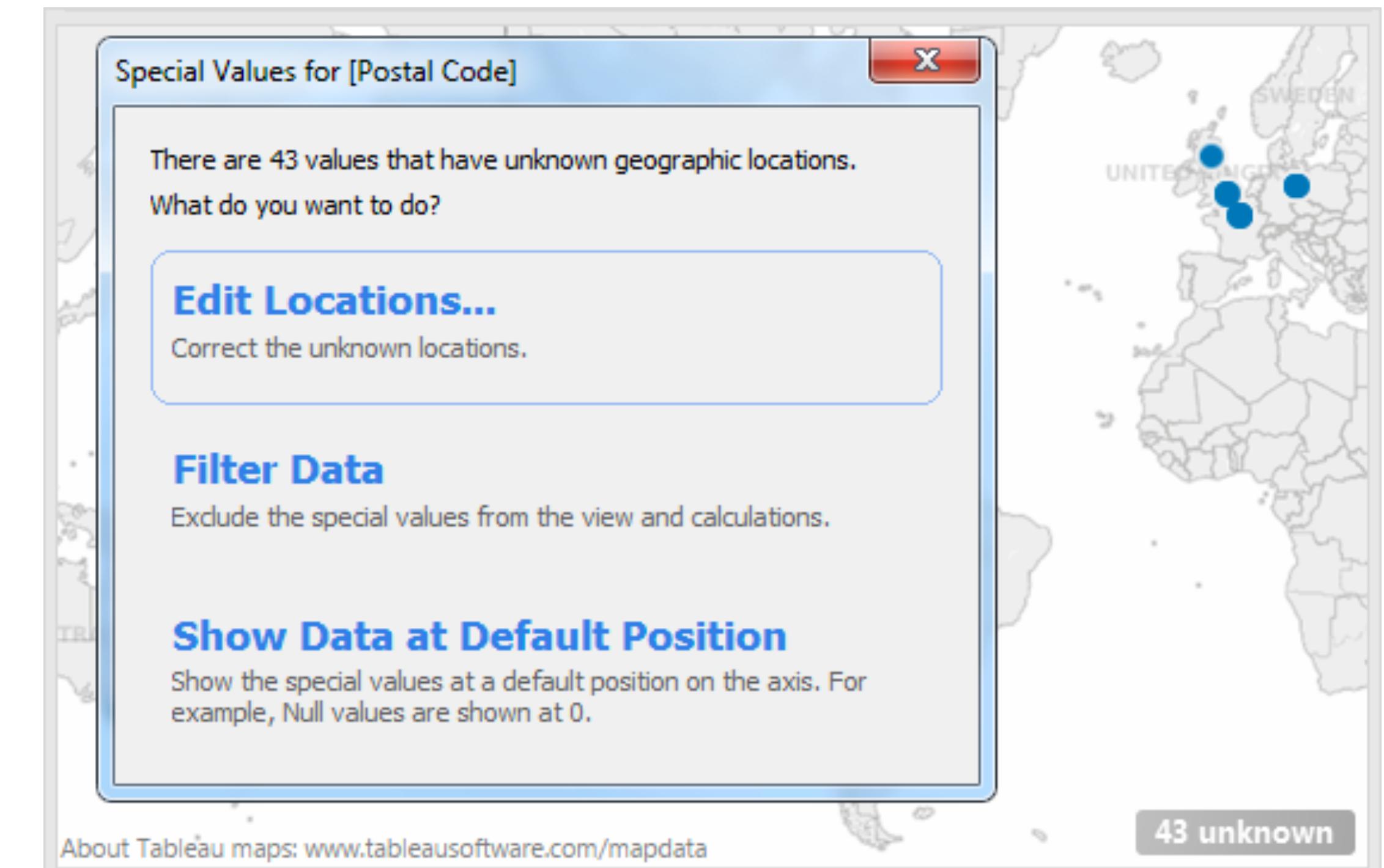


Exercise 2



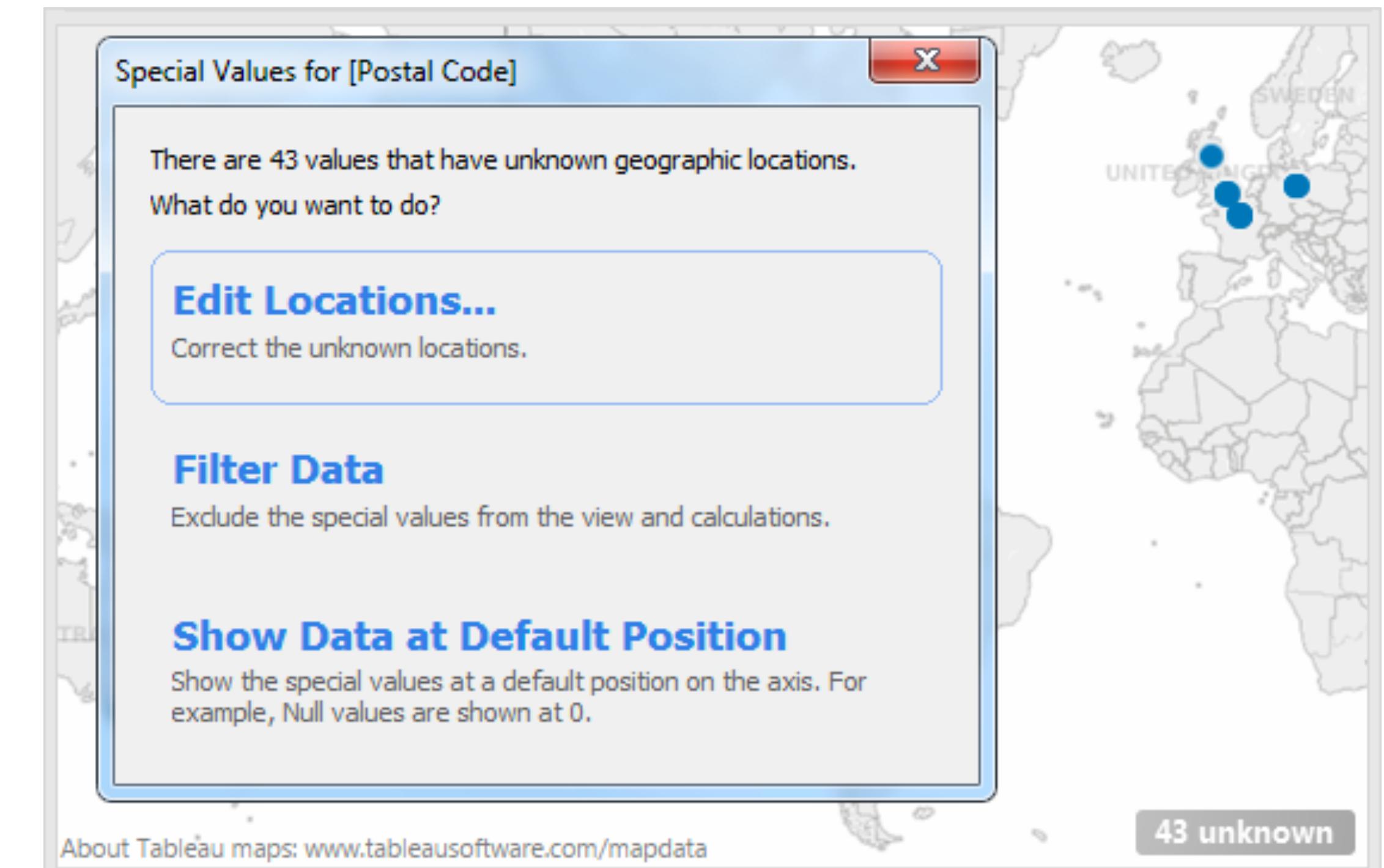
Handling Null Values

- When your data contains any of these special values, Tableau cannot plot them in the view. Instead, it displays an indicator in the lower right corner of the view. Click the indicator to see options for how to handle these values.
- Filter Data** - exclude the null values from the view using a filter. When you filter data, the null values are also excluded from any calculations used in the view



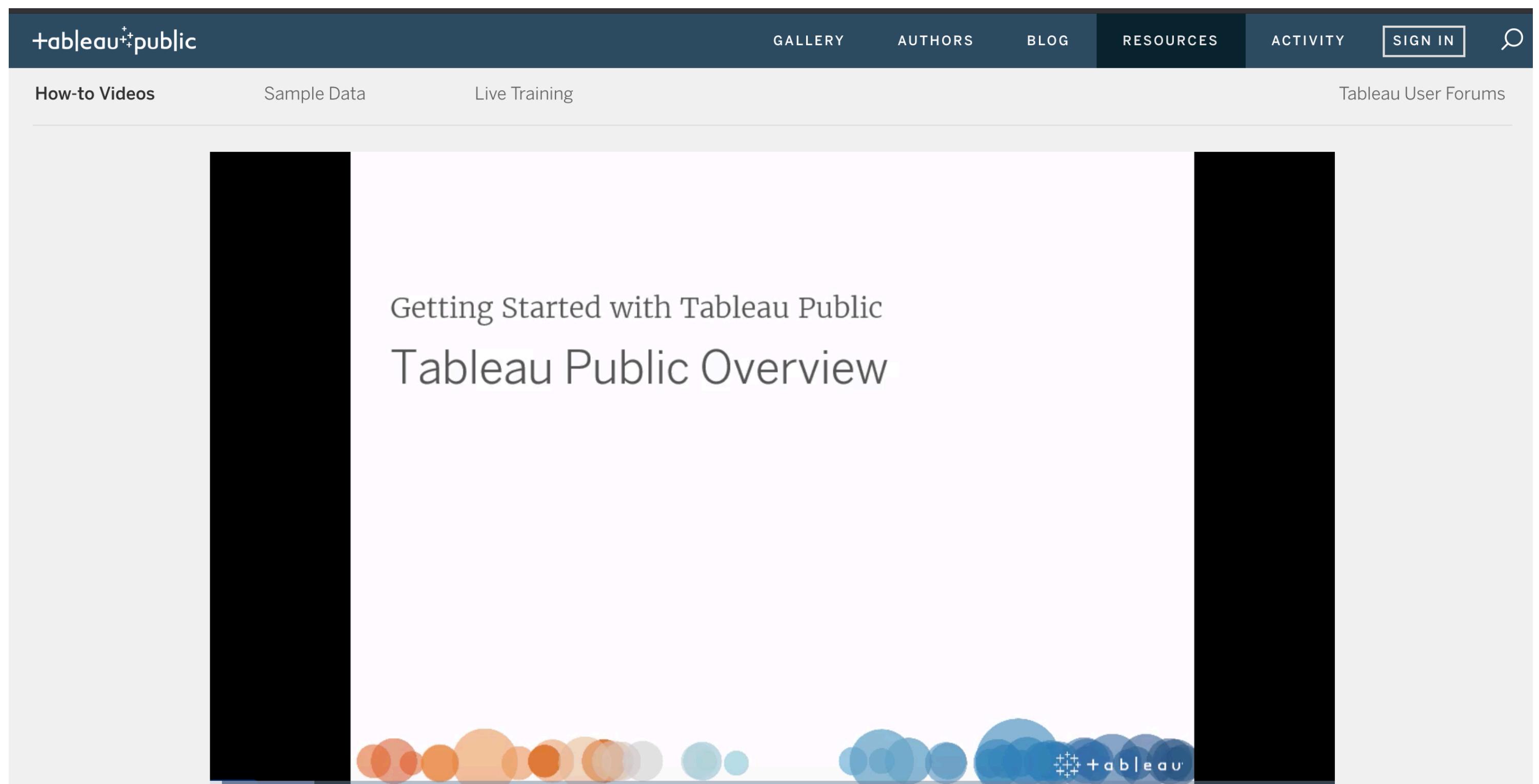
Unknown Geographic Locations

- When working with maps and geographic fields, unknown or ambiguous locations are identified by the indicator in the lower right corner of the view. Click the indicator and choose from the following options:
- Edit Locations** - correct the locations by mapping your data to known locations.
- Filter Data** - exclude the unknown locations from the view using a filter. The locations will not be included in calculations.
- Show Data at Default Position** - show the values at the default position of (0, 0) on the map.



Resources

- <https://public.tableau.com/en-us/s/resources>



Knowledge check 3



Exercise 3

