**Tableau**

* Overview

**Introduction to Tableau**

**Tableau Products**

* 1. [Tableau Public](https://public.tableau.com/en-us/s/): Tableau Public is completely free to use. Data can be downloaded by anyone, so do NOT use confidential data. On the same token, it's a good place to find data to play with. In fact, all of the details behind the public visualizations you see can be accessed by anyone. If you want help getting started on a project: find something you like, open it in your Tableau instance, and modify it to make it your own. It's a great way to learn Tableau. Limitations of Tableau Public are that your workbooks cannot be saved locally, there are fewer ways to connect data, and you are limited to 10 million rows per workbook.
  2. [Tableau Desktop](https://www.tableau.com/products/desktop): Tableau Desktop is used to create dashboards or stories, like Tableau Public. It connects to data sources directly and workbooks can be saved locally. It also is not free... or cheap.
  3. [Tableau Server](https://www.tableau.com/products/server): Tableau Server is used when a company wants to publish data and dashboards internally. The dashboards are created using Tableau Desktop and pushed to Tableau Server, where they are shared across the internal network. The server must be dedicated to Tableau.
  4. [Tableau Cloud](https://www.tableau.com/products/cloud-bi):Tableau Cloud is a mix of Tableau Server and Tableau Public. Like Tableau Public, the server is hosted by Tableau. Unlike Tableau Public, your data and visualizations are private unless you specify otherwise. Like Tableau Server, you would build visualizations on Tableau Desktop and deploy them to a server.
  5. [Tableau Prep](https://www.tableau.com/products/prep): Tableau Prep is a newer product. It is a GUI-based data wrangling application. It includes Tableau Prep Builder and Tableau Prep Conductor. The builder is for creating data flows. The conductor is for sharing flows and managing them across the organization. The ideal user for this software is a data analyst with limited programming skills.

**Tableau File Types**

* 1. .twb - Tableau Workbook: Stores visualization without source data.
  2. .tds - Tableau Data Source: Stores the server address, password, and other information required to access a datasource.
  3. .tbm - Tableau Bookmark: Stores a connection to a worksheet in another Tableau workbook.
  4. .tde - Tableau Data Extract: Stores Tableau data as a filtered and aggregated extract.
  5. .twbx - Tableau Packaged Workbook: Stores extracted data and visualizations for viewing in Tableau or Tableau Reader.

**Download Public Dashboards**

Tableau Public allows you to download it on your computer as Tableau Workbook (TWBX), PDF, Crosstab, Data and Image. To do so, hover over the download button at the bottom right corner and select file type. Not all workbooks allow for downloading as a Tableau workbook.

**Tableau Resources**

* 1. [Install Tableau Public](https://public.tableau.com/en-us/s/)
  2. [How-to Videos, Sample Data, & Live Training](https://public.tableau.com/en-us/s/resources)
  3. [Viz Gallery](https://public.tableau.com/en-us/s/gallery)
  4. [Leading Authors & Artists](https://public.tableau.com/en-us/s/authors#!/)

**Examples of Charts**

[Horizon Chart: Unemployment](https://public.tableau.com/profile/technical.product.marketing#!/vizhome/UnemploymentHorizionChart_1/HorizonChart)

[Positive + Negative Bar Chart: Gender Pay Gap](https://public.tableau.com/profile/iting#!/vizhome/170414_tax2014-15_genderpaygap/malejobs)

[Arrow Chart & Comet Chart: Sales](https://public.tableau.com/profile/tableaubims#!/vizhome/ArrowChartCometChart/Arrowchart)

**Examples of Dashboards**

[Dashboard: Twitter Analytics](https://public.tableau.com/en-us/s/gallery/adams-twitter-analytics?gallery=featured)

[Dashboard: Top Places to Retire](https://public.tableau.com/profile/zillow.real.estate.research#!/vizhome/TopPlacestoRetire/Dashboard1)

[How-to: Custom Filters](https://public.tableau.com/s/blog/2015/06/using-custom-shapes-dashboard-filters)

[How-to: Add Visual Elements](https://public.tableau.com/s/blog/2017/12/how-add-illustrations-your-dashboard-and-why-you-should-care)

**Examples of Stories**

[2018 World Series](https://public.tableau.com/en-us/s/gallery/journey-2018-world-series?gallery=votd)

[St. Mungo Hospital Annual Report](https://public.tableau.com/profile/sean.oslin#!/vizhome/StMungoHospitalAnnualReportPart1/Dashboard1)

[Mobile example: Which countries has the most opiod prescriptions in 2015?](https://public.tableau.com/profile/raycom.news.network#!/vizhome/Whichcountieshadthemostopioidprescriptionsin2015MobileVersion/Dashboard2)

[Deloitte Technology Fast 500](https://public.tableau.com/en-us/s/gallery/2018-deloitte-technology-fast-500?gallery=featured)

[Holiday Shopping](https://public.tableau.com/en-us/s/gallery/online-holiday-shopping?gallery=featured)

[IT Trends 2018](https://public.tableau.com/en-us/s/gallery/it-trends-2018?gallery=featured)

[A Data Viz Storyboard about Data Visualizations](https://public.tableau.com/profile/andy.kriebel#!/vizhome/VisualVocabulary/VisualVocabulary)

**Exercises**

* 1. [Create](https://public.tableau.com/app/discover) a Tableau Public account.
  2. [Download](https://www.tableau.com/products/public/download) Tableau Public.
  3. Log in to Tableau Public.
  4. [Explore](https://public.tableau.com/app/discover) public dashboards and download one.
* Connecting to Data

**Connecting to Data**

Tableau can connect to a broad range of data storage systems, softwares and file types. Tableau Public limits the connections, but the most common are available and should satisfy most of what you want to do. The first step in building a visualization in tableau is connecting to the data. Tableau offers flexibility in how you connect (maintain a live connection vs. create a data extract), in the data types that can be read, as well as in the merging of multiple data sources and datasets. Upon opening the Tableau app (and not a workbook), you will start on the connection page. We will spend this lesson here and on the *Data Source* tab, which you will see after completing a connection.

**Lesson Goals**

* 1. Know where to get help
  2. Connect to a data source
  3. Filter your data
  4. Prepare your data

**Primary Types of Data Sources**

**Local Files**

* 1. Excel: .xls, .xlsx
  2. Text: .csv, .txt, .tsv, .tab
  3. JSON Files: .json
  4. PDF Files: .pdf
  5. Spatial Files: Esri File Geodatabases (gdb\*.zip), Esri Shapefiles (.shp), MapInfo Tables (.tab), MapInfo Interchange Format (.mif), GeoJSON (.geojson), TopoJSON (.json, .topojson), KML (.kml)
  6. Statistical Files: SAS (.sas7bdat), SPSS (.sav), R (.rda, .rdata)

**Servers**

* 1. Google Sheets
  2. OData: mysql database server, e.g.
  3. Web data connector

**Other data connectors** (not available in Tableau Public)

* 1. [Other Data Sources](https://www.tableau.com/products/desktop?_ga=2.168784937.943005315.1582736711-305547270.1582736711&_fsi=g1y4KoAD#data-sources)

**Connection Method: Live vs. Data Extract**

"Tableau Data Extracts are snapshots of data optimized for aggregation and loaded into system memory to be quickly recalled for visualization. Extracts tend to be much faster than live connections, especially in more complex visualizations with large data sets, filters, calculations, etc." (Medrano, 2016). If data is being updated regularly, refreshing the extract is necessary in order to stay up to date with the most recent data.On the other hand, live connections will provide real-time updates and those changes will be reflected in your visualizations, with any changes in the data source reflected in Tableau. While extracts are always optimized for performance, the databases you may access may not be. That is to say that With live connections, your queries will only perform as well as the database can. Generally, you will be using extracts. When you publish a dashboard that needs real-time updates (which is the small minority of the time), you may decide to use a live connection. However, even in those cases, when building the visualizations on your desktop, you will likely use an extract. All of that said, with Tableau Public, you do not have the option to create a data extract.

**Connect to a File**

* 1. When connecting to a file, select the file type, navigate to the file you wish to connect to in your directory, and *open*.
  2. On the left side, you will see Connections, Files & New Union. Also, listed under *Files*, you will also see any other files in your current working directory that can be read by tableau.
  3. In the main section, you will see a sample (1,000 rows default) of the file you have connected to. Think about this sample size like running a SQL query that limits your results to 1,000. You can increase that number, and it will also increase the time required to load the data into view. You really just want a sample large enough to understand the data that is contained in each field. For this lesson, we will connect to cc\_institution\_details.csv.

**Join Files**

**Join Columns** You can connect and merge multiple data sources, such as multiple files or spreadsheets. Tableau joins are similar to joining tables in a mySQL database. You can do an inner, left, right, or full outer join. A box will display asking for the field(s) to join on. For this lesson, we will add a new connection to the file, merged\_2013\_PP.csv.

**Union Rows** You can append rows from 2 sources if your columns map to each other.

**Filter Data**

In this section, we will be focused on the filter option box. You can get there by looking in the top right of the *Data Source* page. There you can add a filter.

Let's work through an example. Say we want to add a new filter to keep only 4 year universities.

* 1. To do this, we will add a filter to the *Level* field.
  2. Filter (in top right) -> Add -> Add... -> Select field *Level* -> ...
  3. Set the criteria.

**General & Wildcard Filters**

There are many ways to identify the criteria for your filters. Under the **General** tab, you have ways to set criteria for exact matches to values in the *Level* field. Under the **Wildcard** tab, you can set criteria for wildcard string matching. We will look at the Condition tab in our next example. For now, let's filter to include only 4-year institutions.

**Conditional Filters**

For the next example, we will filter states. Add a new filter, and select *State* as the field. Under the **Condition** tab, let's add a condition 'By field'. Let's say we want to only include states which have a median *Student Count* of less than 1,000. We will select 'By field' and fill in the parameters to say *Student Count* Median << 1000. You should end with only Florida and Utah remaining. What you have done is filter out *States*. You have not filtered out schools that have a *Student Count* > 1000. To test this, complete your filter by selecting OK. You should have 206 rows remaining. Now, go back to your filters, edit the existing filter to not filter by a condition, but instead filter by name of state. Include only Florida and Utah. You will see you end up with the same number of observations.

**'Top' Filters**

Similar to Conditional filtering, you can filter the states by ranking in any of the columns. For example, say you want to explore colleges that are in the top 10 states in terms of *Median SAT Value*. You would filter *State* by going to the **Top** tab, and then filter Top 10 by *Median SAT Value* followed by your aggregate method, such as average. Note: Median SAT Value is the median SAT value for that school. When we are aggregating by average, we are taking the average of all of the colleges' *Median SAT Value* in that state. Then Tableau ranks those averages and returns the top 10 states.

**Prepare Data**

In this section, we will be focused on the table containing a sample view of our data. Each column header has 3 groups of activities contained in that tiny little box. The first is the icon in the top left of the column header representing the type of data. Second is the drop down arrow in the top right of the column header. Finally is the sort option in the bottom right of the column header box. The drop down menu arrow and the sort icon on the right side will not appear until you hover over the column header.

**Data Types**

For each column, there is an icon that represents the type of data it contains, or, more accurately, the type of data Tableau has concluded it contains. The options are: Number (decimal), Number (whole), Date & Time, Date, String, and Boolean. When you see a globe icon, this means the data type is a string or a number with a *Geographic Role* assigned to it. Possible geographic roles include airport, area code, city, country/region, county, state/province, zip code, latitude, and longitude. If the data type is a string, you will not see latitude or longitude as options under geographic role. To update the data type of a column, click the icon and select the new data type. The most common types that need correcting are ID columns that default to a number or a year that also defaults to a number. A good way to think about what data type a column should be is to consider how the column will be used in your visualization. If it might make sense to aggregate the field by summing or averaging, for example, then that field should be a number type. If the only way you would aggregate that column would be through a count, then it is likely that column should be in a string. If the column represents a date, such as year, then, obviously, change it to date format. What happens when you change a column *year* to a date data type?

**Other Column Options**

The quick menu in the top right of each column provides the following options...

* 1. **Rename**: Rename the column name
  2. **Reset Name**: Reset the name of the column to the name in the source data. Tableau renames the fields, upon import, to a more audience friendly name. Often, when building the visualizations, and especially calculated fields, it is easier to have names not separated by spaces. In these cases, it might be easier to revert back to the original names.
  3. **Copy Values**: By selecting the column and then 'copy values', you are copying the values in the column to the clipboard. It is the same as typing command-C.
  4. **Hide**: Hide will hide the column from view. You can still reference the column, such as in calculated fields.
  5. **Aliases** (strings only): Aliases are useful when you want a value in a field to be a more user-friendly value. For example, if *State* was in abbreviated form, but you want to make sure the entire state is spelled out on your visualizations, then aliasing would be useful. The alias does not change anything about the data in the background, only what is displayed.
  6. **Create Calculated Field...**: Create a new field based on existing fields. As an example, we will create a new field, award\_per\_natl\_delta, that is the difference between awards\_per\_natl\_value and awards\_per\_value. In the column *Awards Per Natl Value*, select Create Calculated Field from the drop down menu in the top right corner of the column header. Add the new column name, Awards Per Natl Delta. In the formula box, enter the formula: [Awards Per Value] - [Awards Per Natl Value] and click OK. You now have a new column where the data type icon, instead of '#' for number, it shows '=#' to represent a calculated number. You will also notice the lack of a blue bar bordering the top of the column header, indicating the column was not in the original data.
  7. **Create Group...**: Created Group is used for grouping categories into larger, higher hierarchical groups. For example, if I wanted to have a way to identify states by region, I could use the *State* column to create groups and then add each state to a group, such as south, northeast, southwest, northwest, and central. To do so, go to "Create Group" in the drop down menu for the *State* column. Begin selecting the states you wish to go in the first group. To select multiple states, hold the Command key down. Once your first group values are selected, click 'group'. You can then name the group. Move on to the next group. When you are done, title the new field and click ok. You can know a field is a result of grouping another field by the paperclip that has been added to the data type icon.
  8. **Split** (strings only): Split strings at a common delimiter into multiple columns. Let's take the field, *Counted Pct*, as an example. *Counted Pct* is defined as the "share entering undergraduate class who were first-time, full-time, degree-seeking students, meaning that they generally would be part of a tracked cohort of potential graduates. The entering class of 2007 is displayed for 4-year institutions; 2010 for 2-year institutions" (<https://data.world/databeats/college-completion/workspace/data-dictionary>). Currently, the data is in the format pct|yy, such at 54.8|10. 54.8% of those entering that 2-yr college's undergraduate class of 2010 were first-time, full-time, degree-seeking students. I would like to be able to use that percentage, so if I split the column into 2, then I can turn one into numeric and the other into date. I can then hide the original column.
  9. **Custom Split** (strings only): With custom split, you can specify the delimiter to split on. You can also specify how many columns to split off.
  10. **Create Bins...** (numbers only): Bins are useful when you have a continuous variable and you want to reduce the noise and bin the values close to each other together. For example, if you wanted to create a field that bins the number of students into equally sized groups, you could use "Create Bins" to do that.
  11. **Pivot** (when selecting multiple fields): Use as you would when creating a pivot table in excel or Python.
  12. **Merge Mismatched Fields** (when selecting multiple fields): Does what you would think! Use this when you want to merge fields into a single field, such as *City* and *State* into *City State*.
  13. **Describe**: Describe is a useful resource for a quick view into the column. The domain is not loaded by default. To see the domain of the field, click 'Load' in the Describe window. If you look at the description for *Awards Per Natl Delta*, a calculated field, you will see the formula that is used to calculate that field.

**Column Sorting**

Click the horizontal bar graph icon in the column header to sort that column. It will cycle through ascending, descending, and original sort as you click the icon. When you hover over the column header, you can tell if it is sorted by the way the horizontal bar graph image is sorted.

**Getting Help**

* 1. F1 or the Help menu header: you have options to get support, watch videos, see sample workbooks, sample gallery, as well as customizing settings & preferences.
  2. [onlinehelp.tableau.com](http://onlinehelp.tableau.com)

**Exercises**

* 1. In your mySQL client, query the database telco\_normalized, joining all tables together into a single table which you will then export to a csv and save on your local drive. Do NOT use the telco\_churn database.
  2. Connect to your csv through the tableau client that is installed on your laptop. Reminder: a csv is considered a text file.
  3. Hide any redundant columns (payment\_type\_id, contract\_type\_id, internet\_service\_type\_id are all represented through the columns with their descriptive names).
  4. Dimensions are something you could group by to see aggregated measures, like average payment by gender, or total charges by customer\_id. Measures are the fields you would perform calculations on. Ensure all possible measures are stored number datatype, and all dimensions are NOT stored as number data types. E.g. Senior Citizen will need to be changed to a string.
  5. Create aliases for values in the following fields. Follow the examples to make similar aliases. The goal is to make the values easily interpretable to the user.
     + Senior Citizen: "Is Senior Citizen", "Not Senior Citizen"
     + Partner
     + Dependents
     + Paperless Billing
     + Phone Service
     + Multiple Lines: "No": "Single line", "Yes": "Multiple lines", "No phone service": "No Phone service"
     + Online Security: "No": "Internet without Online Security", "Yes": "Online Security", "No internet service": "No internet service"
     + Online Backup
     + Device Protection
     + Tech Support
     + Streaming TV
     + Streaming Movies
  6. Create a new grouped field from payment type. Group the automatic payments by selecting both of those payment types, and name the group "Automatic Payment". Then group the non-auto payments and name the group "Manual Payment". You will then have a new field at the end of your table titled "Payment Type (group)".
  7. Once your table is ready, export it to a csv for a backup in case your tableau file gets deleted! (data -> export data to csv)

**Data Resources**

* 1. <https://ds.codeup.com/appendix/open_data/>
  2. <https://public.tableau.com/en-us/s/resources>

**References**

Medrano, Diego (April 14, 2016), Tableau Online tips: Extracts, live connections, & cloud data, <https://www.tableau.com/about/blog/2016/4/tableau-online-tips-extracts-live-connections-cloud-data-53351>

* Creating Custom Fields

**Creating Custom Calculations & Fields**

Custom-calculated fields are used to segment data, convert data types, aggregate data, filter results, and calculate ratios.

Example scenarios include:

* 1. Data is missing, such as net profit, when you have revenue and expenses.
  2. Transform values, such as year-over-year growth. (Quick Table Calculation)
  3. Quickly categorize data, such as: IF [net profit] > 0 THEN 'growth' ELSE 'decline' END.

**Lesson Goal**

* 1. Create basic calculations

**Basic Calculations**

Basic calculations allow you to transform values or members at the data source level of detail (a row-level calculation) or at the visualization level of detail (an aggregate calculation).

* 1. Example row-level calculations, or those at the data source level of detail, include: splitting an email address into two columns, username & domain, computing a date difference between start date and end date, or computing the profit of each sale by taking the difference of cost and sales. The calculation in Tableau would look like: [sales\_price] - [cost]. In pandas, the same type of calculation would look something like: df['profit'] = df.sales - df.cost.
  2. Example aggregate calculations, or those at the visualization level of detail include: summing the profit or computing the median sales price. How these are grouped when aggregated depends on the dimensions in your visualization. So, the results of the calculation will change with your visualization. The calculation in Tableau would look like: MEDIAN([sales\_price]). What this concept could look like in pandas, assuming you want to compute the median sales price for each product category: df.groupby(['product\_category'])['sales\_price'].median(). In the basic calculations, though, the group by is updated every time you alter your visualization by adding or removing a dimension (i.e. a group by field).
     + *NOTE: This may be addressed in the calculation using a*[*FIXED Level of Detail*](https://help.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_lod.htm#fixed)*which computes an aggregate using only the specified dimension(s).*

**Simple Calculations**

Create a [simple calculated field](https://help.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_formulas.htm) with the following steps:

* 1. Analysis > Create Calculated Field OR in the field names on the left, click on the down arrow of a field you wish to use in the calculation and select 'Create Calculated Field'.
  2. Name the new field
  3. Enter a formula.

**Ad Hoc Calculations**

[Ad hoc calculations](https://help.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_adhoc.htm) are calculations that you can create and update as you work with a field on a shelf in the view. Ad-hoc calculations are also known as type-in or in-line calculations. Double-click on a field in your chart (in the rows or columns list, e.g.) and begin.

**Quick Table Calculations**

To create [quick table calculations](https://help.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations_quick.htm), such as year-over-year difference, click on the menu arrow of a field that exists in your chart/table and select 'Quick Table Calculation', then follow the options.

**The Format of Calculations**

To understand the format of calculations, we will use this example from the [Tableau help site](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm):

IF [Profit per Day] > 2000 THEN "Highly Profitable"

ELSEIF[Profit per Day] <= 0 THEN "Unprofitable"

ELSE "Profitable"

END

// this function labels all profits over 2000 as highly profitable, those <= 0 as unprofitable and all others as profitable.

* 1. **Functions**: [Functions](https://help.tableau.com/current/pro/desktop/en-us/functions.htm) transform the values or members in a field. The functions in the example include IF, THEN, ELSEIF, ELSE, and END.
  2. **Fields**: [Fields](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#Fields) are dimensions or measures from your data source, i.e. columns. The fields in the example include Profit per Day.
  3. **Operators**: [Operators](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#operator-syntax) are symbols to denote an operation. Operators in our example include > and <=.
  4. **Literal Expressions**: [Literal expressions](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#literal-expression-syntax) are constant values that are represented “as is”, such as a string you want to match or return in an if statement. The literal expressions in our example include "Profitable", "Unprofitable", "Highly Profitable", 2000, and 0.
  5. **Parameters**: [Parameters](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#add-parameters-to-a-calculation) are placeholder variables that can be inserted into calculations to replace constant values. See more on [creating parameters](https://help.tableau.com/current/pro/desktop/en-us/parameters_create.htm#create-a-parameter).
  6. **Comments**: [Comments](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#add-comments-to-a-calculation) You should add comments to any calculations beyond the very basic and simple. Comments are preceded by //. The comment in our example is: "// this function labels all profits over 2000 as highly profitable, those <= 0 as unprofitable and all others as profitable."
  7. **Data Types**: [Data Types in Calculations](https://help.tableau.com/current/pro/desktop/en-us/functions_operators.htm#understanding-data-types-in-calculations) include string, date/datetime, number, and boolean.

**Exercises**

Creating Custom Fields:

* 1. Create a new calculated field from total charges. Call it "Estimated Tenure (months)". It computes the number of months the customers were around from monthly charges and total charges. Round your tenure value to the nearest whole number.
  2. Create a new calculated field from "Churn Month". Call it "Customer Status". It computes whether or not a customer has churned based on the "Churn Month" field using an IF statement. Your function should look like: IF isnull([field1]) THEN 'churned' ELSE 'active' END.
  3. Create a new calculated field from "Tech Support". Call it "with Tech Support". Using an IF ELSE statement, write the function to return a 1 if the customer has tech support ([Tech Support = 'Yes']) else a 0.
  4. Repeat the process in number 3 above for the following fields: 'Streaming TV', 'Streaming Movies', 'Online Backup', 'Online Security'.
  5. Create a new calculated field from "with Tech Support". Call it "Add-On Count". In this field, you will add all the fields you created in numbers 3 and 4 above. It will look like: [with Tech Support] + [with Streaming TV] + [with ...]
  6. Convert all of your new fields to dimensions. (Click the down arrow to the right of the field name and select convert to dimension)

**Appendix**

**Other Types of Calculations**

* 1. Create LOD expression
  2. Create table calculation
  3. Become familiar with [Tableau functions](https://help.tableau.com/current/pro/desktop/en-us/functions.htm)
  4. Become familiar with Tableau documentation on [Calculated Fields](https://help.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_create.htm)

**Level of Detail Expressions**

[Level of Detail (LOD) expressions](https://help.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_lod.htm), like basic calculations, allow you to compute values at the data source level and the visualization level. However, LOD calculations give you even more control at the level of granularity you want to compute. They can be performed at a more granular level (INCLUDE), a less granular level (EXCLUDE), or an entirely independent level (FIXED) with respect to the granularity of the visualization. The function in Tableau would look like: { FIXED [product\_category]:(MEDIAN([sales\_price]))}. In pandas, this would look like: df.groupby(['product\_category'])['sales\_price'].median(). This is just like the example in basic calculations, but this time, in Tableau, the group by field stays static in the visualizations. So if you create a visualization that removes the product\_category dimension and adds product\_subcategory as a dimension, the value of the median sales\_price will remain at the product\_category level. So you will see multiple subcategories with the same value.

**Table Calculations**

[Table calculations](https://help.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations.htm) allow you to transform values at the level of detail of the visualization only. For example, if you wanted to add year-over-year growth to your visualization, you could do this in 2 ways. The first way is to add in the Create Calculated Field formula box. It would look like this: ATTR([sales]) - LOOKUP(ATTR([sales]), -1). You could also create this calculation by clicking on the menu arrow of the field sales and selecting 'Quick Table Calculation'. Common use cases for table calculations include ranking, cumulative total, rolling averages, and inter-row calculations such as year-over-year.

* Creating Charts

**Creating & Customizing Charts**

Before we can create dashboards and stories, we have to create charts. As of now, we should have one tab labeled *data source*, and a blank sheet labeled *sheet 1*. We will now move to *sheet 1*. Every chart we wish to create, we will do so on a new sheet such as this one. Tableau refers to the tabs used for indiviual charts as *worksheets*. In this lesson, we will create a variety of charts on multiple worksheets. One way to create a new tab is to click on the tab that displays a bar chart icon with a '+' in the bottom right corner. The two tabs that follow it are for creating a new dashboard and creating a new story, respectively. Another way you can create a new tab is by duplicating existing tab (right click on tab name) and then editing variables, chart types, titles, etc. So, let's get building.

**Lesson Goals**

* 1. Creating charts
  2. Sorting and filtering data
  3. Displaying the data underlying a workbook

**Data Table**

Tableau can create a basic table, like a pivot table you would create in Excel but with more flexibility, that groups and aggregates. You can specify how it aggregates, which fields to aggregate, and the format of the numeric data. If you explore the 'Analysis' tab, you can change the table values to return a 'percentage of' and then specify whether to compute percentage of row, column, or entire table, for example.

**Box Plot**

"A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum." (Khan Academy, 2020)

A purpose of box plot is to see how the points are distributed across the quartiles, as well as any outliers that exist in the data. For a box plot, you need 0 or more dimensions and 1 or more measures.

**Bar Plot**

**Heatmap**

To demonstrate the heatmap, we will duplicate our basic table tab, and then we can use the "Show Me" menu on the top right to change the visualization from a table to a heatmap or a highlight table (which I think of as a heatmap, also). In this case the highlight table gives us what we need.

**Scatter Plot**

A scatter plot needs 2 measures, or numeric variables. By adding dimensions the scatter plot goes from 1 data point to multiple data points, depending on the number of distinct values or combinations in the dimensions.

Dimensions:

For example, let's take retention rate (x) and cohort size (y). We have to select a way to aggregate each of those. Let's aggregate both retention rate and cohort size by median. That value really doesn't mean anything when everything is aggregated together into a single point. So, we need to bring in some dimensions. Think, *group by*. Dimensions can be visualized separately through color, shape, size, or other means, *or* they can just be pulled in so that individual points are not lost and the aggregate measure only goes as far as each value in that dimension. Back to our example, we can add the dimension of *Control* which indicates whether the school is public institution or a private one. That is a good one to indicate by color, as there are a limited number of distinct values. But what if we want to see a data point for each college? We can drag that dimension into our chart. It will add the college name in the info box that displays when hovering over the point and our chart will now have a dot for each college. The measure we are aggregating is now the actual value you would find in the original dataset.

Analysis:

On the top left of the window, there is a tab that exists behind the *Data* tab (where all the columns are listed) called *Analytics*. From here (or from Analysis Menu item at top of page) we can add trend lines, regression formula, among other bits of information.

**Line Chart**

A line chart needs 2 continuous variables. A line chart can be used to analyze the correlation of 2 variables. It can be used in time series analysis and many other use cases as well.

**Map**

Tableau chart types: **symbol maps**, **maps**

These both require a geographic element. They can take 0 or more dimensions. **Symbol maps** can take 0 to 2 measures, while **maps** take 0 to 1 measures.

Drag longitude to columns and latitude to rows. A default symbol map of a map will populate. Drag *Student Count* to the center of the map and aggregate by median. What's wrong? Why doesn't anything happen? (hint: group by) Add *City* to group by. Where is Portland, OR? I can't find the dot representing portland. Why could this be?

First we should confirm that the data contains Portland. There are multiple ways to do this. One way is to add *city* as a filter and search to include only Portland. Another way is to filter by State and drill closer into the state. Upon doing that, we see a very little dot of "Portland". That seems counter to what I would expect. Big city should mean bigger dot. This is one way maps can be confusing and/or deceiving. It really shouldn't when we think about how we are aggregating. We are taking the median, and larger cities are going to have more, smaller colleges, thus driving down the median. We could use sum if we want to see total students in the city. We could also count the number of colleges to get different information.

Let's change the **symbol map** to **map**. We will then colored the state by the number of colleges that exist in that state. What is wrong with this visualization (misleading, deceptive, confusing, ...)?

**Display Underlying Data**

To display or export the underlying data:

To create a **crosstab** from a chart, right click on tab and select "duplicate as crosstab"

**Sorting and filtering data**

**Filter**

**Sort**

**Titles, Labels, Captions, Summary**

**Titles**

You might notice that your title matches the name of your tab. Rename your tab, and you will notice that these two strings are connected, by default. Now double click on the title of the page, above the chart. Now you can see how these strings are linked. To customize the chart name, enter the new name in the box provided. If you would like to maintain the synchronization between the tab and the chart title, rename the tab, but keep the chart title as < Sheet Name >

**Captions**

To add a caption, go to *Worksheet* menu header, and select *Show Caption*. Captions are useful for key points, notes about the data, what you want the user to take from your chart, how to read the chart, etc. When you are answering a question through a viz, also answer the question in your caption.

**Axes**

You can customize the x & y axes by selecting one and right clicking.

**Summary**

Right click and select *Summary* to view the summary card. You can also view it by *Worksheet* -> *Show Summary*.

**Exercises**

**Plot 1: Scatter plot**

* 1. Create the plot:
     + Create a Scatter Plot with the X axis representing 'Estimated Tenure (months)' and y axis representing 'Monthly Charges'.
     + In order to create a scatter plot of all the customers' tenure x monthly charges, you need to add the customer ids to the chart. Drag the 'Customer ID' field to the "Detail" box in "Marks".
     + Next, add some color to represent whether or not a customer has churned. Do this by dragging 'Customer Status' to the "Color" box in "Marks".
  2. Add a caption:
     + What do you notice from your plot? Add the caption box to your window by right clicking on a gray area in your window until you see the option to select "Caption". Select it so that it has a check mark next to it.
     + Once the caption box appears, double click inside it to bring up the edit box. Keep the default text, but after it, add any takeaways you found in this visualization about how churn seems to relate to tenure and monthly charges.
  3. Format the axes:
     + Format the axis labels: right click in the area of the y-axis and select format.
     + To your left you will see "scale". Format the number labels to be currency (custom). Remove the decimals.
  4. Add a Title:
     + Double click in the title space, currently titles "Sheet 1", and edit your title.
     + Make your title indicative of what you want the user to take away from this chart. A question is often a useful way to title a chart. For example, “Are Monthly Charges Related to When/Whether a Customer Will Churn?”

**Plot 2: Bar plot**

* 1. Create the plot:
     + Make a bar plot showing how many customers have churned and how many are active.
     + In a bar plot, you will have 1-2 dimensions and 1 measure, such as count of total records. In this case, make the x-axis 'Customer Status' and the y-axis count of the total records.
     + Add a dimension to color, internet service type, so you can view number of customers and their status with respect to type of internet
  2. Add a caption with your takaways.
  3. Format the axes:
     + Change the title of 'count of .csv' to something meaningful, such as "Number of Customers". Do this by right clicking on the y-axis and selecting "Edit Axis". There you have the option to change the title.
  4. Add a title:
     + What do you want the user to takeaway from your chart? Make your point clear and hard to miss! An example is: "Customers with No Internet Service Do Not Stick Around"

**Plot 3: Duplicate Plot 2**

* 1. Right Click on sheet 2 and select "duplicate"
  2. Convert your y-axis into a percentage:
     + Select Analyze (top menu item) -> percentage of -> columns

**Plot 4: Explore Tech Support and Add-Ons**

Explore the number of add-ons with and without tech support and how these relate to churn. Control for internet service type.

* 1. Add a new chart.
  2. Add to the columns, Internet Service Type and Add-On Count.
  3. Add to rows 'with Tech Support' and your record count values (e.g. mine is titled 'Telco\_from\_sql.csv (Count)')
  4. Select the best type of plot for your data and insights.
  5. Give your chart a title that indicates what you take away from the chart.

Does having tech support have a relationship with churn? If so, what is it? Is it what you would expect?

**Save your workbook to Tableau Public**

**References**

Khan Academy, (2020), Retrieved from <https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/box-whisker-plots/a/box-plot-review>

* Creating Dashboards

**Creating a Dashboard**

A [dashboard](https://help.tableau.com/current/pro/desktop/en-us/dashboards.htm) is a collection of several views or charts. In Tableau, it is another tab, just like the worksheets for charts and the data tab, but specified for dashboards. When data is modified or a chart in a worksheet is modified, the dashboard is modified, and vice versa.

Remember all we have discussed on knowing your audience and best practices in storytelling as you create your dashboards. Who is your audience, what is the purpose of the dashboard, what is the setting or what kind of device will it be viewed on, etc.

**Lesson Goals**

* 1. Create a dashboard
  2. Set default display to be tablet or phone
  3. Add charts to dashboard
  4. Group layout containers
  5. Add text objects
  6. Add interactivity through highlighting and filters

**Dashboard Display**

* 1. [Dashboard Size](https://help.tableau.com/current/pro/desktop/en-us/dashboards_organize_floatingandtiled.htm#dashboard-size-options): options include fixed size (default: dashboard remains the same size), range (size scales between a min and max value), and automatic (resizes to fit the window used).
  2. [Dashboard Device Layouts](https://help.tableau.com/current/pro/desktop/en-us/dashboards_dsd_create.htm): You can make use of templates and create dashboard layouts based on the type of device the dashboard will be viewed on (phone, tablet, desktop).
  3. [Accessible Dashboards](https://help.tableau.com/current/pro/desktop/en-us/accessibility_dashboards.htm): Tableau also has features for creating accessible dashboards that you can look into further using the hyperlink.
  4. [Managing Sheets in Dashboard Display](https://help.tableau.com/current/pro/desktop/en-us/environ_workbooksandsheets_sheets_hideshow.htm): When you are deploying a dashboard or a story, you will want to manage your sheets so that you product is clean and simple.

**Dashboard Objects**

To [add an object](https://help.tableau.com/current/pro/desktop/en-us/dashboards_create.htm#add-an-object), select an item from the bottom left corner and drag to the main area on the page. To set the options, click the object container to select it and then click the arrow in the upper corner to open the shortcut menu.

* 1. **Horizontal and Vertical objects** provide [layout containers](https://help.tableau.com/current/pro/desktop/en-us/dashboards_refine.htm#Use_a_layout_container) that let you group related objects together and fine-tune how your dashboard resizes when users interact with them. Your layout containers can be customized in ways such as [grouping containers together](https://help.tableau.com/current/pro/desktop/en-us/dashboards_organize_floatingandtiled.htm#group-items-using-layout-containers), [evenly distributing layout containers' items](https://help.tableau.com/current/pro/desktop/en-us/dashboards_organize_floatingandtiled.htm#evenly-distribute-a-layout-containers-items), [tiling or floating items](https://help.tableau.com/current/pro/desktop/en-us/dashboards_organize_floatingandtiled.htm#tile-or-float-dashboard-items), and [adding padding, borders, or background colors to items](https://help.tableau.com/current/pro/desktop/en-us/dashboards_organize_floatingandtiled.htm#add-padding-borders-and-background-colors-around-items).
  2. **Text objects** can provide headers, explanations, and other information.
  3. **Image objects** add to the visual flavor of a dashboard, and you can link them to specific target URLs.
  4. **Web Page objects** display target pages in the context of your dashboard. Be sure to review these web security options, and be aware that some web pages don't allow themselves to be embedded—Google is one example.
  5. **Blank objects** help you adjust spacing between dashboard items.
  6. **Navigation objects** let your audience navigate from one dashboard to another, or to other sheets or stories. You can display text or an image to indicate the button's destination to your users, specify custom border and background colors, and provide informational tooltips.
  7. **Export objects** let your audience quickly create a PDF file, PowerPoint slide, or PNG image of a dashboard. Formatting options are similar to Navigation objects.
  8. **Extension objects** let you add unique features to dashboards or integrate them with applications outside Tableau ([more](https://help.tableau.com/current/pro/desktop/en-us/dashboard_extensions.htm))

**Dashboard Interactivity**

Dashboards can become interactive and more flexible through the use of filters and parameters. Also, enabling the use of highlighting allows for exploration in the dashboard. One was you can add interactivity is in upper corner of sheet, enable the Use as Filter option to use selected marks in the sheet as filters for other sheets in the dashboard. You can find other actions possible [here](https://help.tableau.com/current/pro/desktop/en-us/actions_dashboards.htm).

**Exercises**

* 1. Create a Dashboard with at least three visualizations that you’ve created so far.
  2. Create a text object as a header with the name “TelcoCo KPI Report” Add [this image](https://drive.google.com/file/d/1OmmavSYhDbDxZNLvqnDMQVoN9zm09qKj/view?usp=sharing) as an object on your dashboard.
  3. Determine which of your visualizations are most valuable, and remove at least one, then change the format into a phone layout.
     + How did you decide to drop one of your charts? Think about the ordination of value that each provides and describe your thought process.
* Creating Stories

**Creating a Story**

A Tableau Story is a sequence of visualizations that work together to convey information. You can create stories to tell a data narrative, provide context, demonstrate how decisions relate to outcomes, or to simply make a compelling case. ~ Stories on [help.tableau.com](http://help.tableau.com)

A story is a separate sheet, just like data, charts, and dashboards. Stories contain multiple pages within them, known as story points. Users can navigate between the points by clicking on the menu at the top of the story.

**Story Elements**

* 1. Adding a new story point: Select Blank to add a new blank point, or Duplicate to use a copy of the current point as the beginning of a new one.
  2. The Story pane: Use this pane to drag dashboards, sheets, and text descriptions to your story sheet. You can also show/hide the title and set the size of the story.
  3. The Layout pane: Change the style of the navigator object and choose whether to show/hide the navigation arrows.
  4. The Story menu: Use this menu to format the story font styles, shading, and borders.
  5. The Story toolbar: Hover your mouse over the navigation bar to make it appear. Use it to delete points, revert changes, update points, or save the point to a new story.
  6. The navigator: Use the navigator to select different story points. You can also change the order of story points by dragging and dropping them. Your audience will use the navigator to engage with your story.

[Source](https://help.tableau.com/current/pro/desktop/en-us/story_workspace.htm)

**7 Types of Data Stories**

Tableau talks about 7 types of data stories in a table you can see [here](https://help.tableau.com/current/pro/desktop/en-us/story_best_practices.htm#the-seven-types-of-data-stories).

**Build a Story**

For details on how to create and format stories, check out the [Tableau documentation](https://help.tableau.com/current/pro/desktop/en-us/story_create.htm).

**Exercises**

* 1. Create a story titled “TelcoCo Key Performance Presentation”.
  2. Create an initial point with two text objects, as a preview for your presentation.
  3. Make a second point with your most valuable (by personal assessment) visualization.
  4. Make a third point containing the desktop variant of your dashboard.