## Getting Started & Introduction to Data Visualization

Introduction to Data Visualization

- Video: Lecture Introduction to Data Visualization
- Reading: Reading Introduction to Data Visualization
- Video: Lecture What is Data Visualization and Why Do We Do It?
- Discussion Prompt: Data Visualization Challenges
- Video: Lecture Tools for Data Visualization
- Discussion Prompt: Visualization Tools
- Discussion Prompt: Python and R
- Video: Lecture Installing Tableau
- Discussion Prompt: Installing Tableau
- Video: Lecture Creating Your First Visualization
- Discussion Prompt: Creating Your First Visualization (Activity)

- Video: Lecture Exploring and Navigating Tableau Overview
- Video: Lecture Importing Visualizations
- Discussion Prompt: Downloading a Visualization
- Video: Lecture Navigating the Tableau Public's Software - Part 1
- Video: Lecture Navigating the Tableau Public's Software - Part 2
- Video: Lecture Navigating the Tableau Public's Software - Part 3
- Discussion Prompt: Navigating Tableau
- Reading: Reading Additional Readings
- Video: Lecture Making Data Connections Overview
- Video: Lecture Preparing Your Data For Import
- Video: Lecture Primary Types of Connections
- Video: Lecture Preparing Your Data
- Video: Lecture Connecting and Merging Multiple Data Sources
- Discussion Prompt: Tableau Public Sample Data Sets
- Discussion Prompt: Tips to Connecting to Data

Readings and Review

- Reading: Reading TDE or Live? When to Use Tableau Data Extracts (or not)
- Reading: Reading Tableau - Extract Your Data
- Peer-graded Assignment: Connecting to Multiple Data Sources

- Review Your Peers: Connecting to Multiple Data Sources
- Quiz: Making Data Connections

Context of Data Visualization

- Video: Lecture Context of Data Visualization Overview
- Video: Lecture The "Who, What, and How" of Data Visualization
- Reading: Reading Resonate - Present Visual Stories that Transform Audiences
- Video: Lecture Illustration of the "Who, What, and How"
- Discussion Prompt: Visualization of the Day Exercise
- Video: Lecture Context of Data Visualization
- Discussion Prompt: Context of Data Visualization (Activity)

Welcome to this first module, where you will begin to discover the power of data visualization. You will define the meaning and purpose of data Review

Peer-graded Assignment: Storyboarding Your Visualization
2h
Review Your Peers: Storyboarding Your Visualization
Quiz: Context of Data Visualization
10 questionssualization and explore the various types of data visualization tools, beyond Tableau. You will install Tableau on your own device and create your first visualization.

## Lernziele

- Discuss why we visualize data
- Define terminology related to visualization
- Identify what types of software options are available to do visualizations
- Operate installation procedures to install the Tableau Public software on your computer
- Create a visualization

This is the first course with the data visualization and storytelling specialization brought to you by UC Davis Extension.

I use data visualization to identify potential fundraising opportunities, that were not evident from simply looking at spreadsheets. A famous American baseball player and coach, Yogi Berra, once said that, you can see a lot by looking. And, he's so right.

This is what data visualization is all about. **It's taking a bunch of numbers and texts and transforming them into something our brain can process**. This specialization will be a hands on examination of data visualization and storytelling.

We're going to assume that you're new to the field of data visualization. But once you've completed the various courses within the specialization, you'll have a solid base that you can use with in your career. This particular courses been broken into four modules.

- In the first module, we're going to identify what types of software options are available to do visualizations and define common terminology. You will **install Tableau Public** software on your own computer and do a visualization.
- In the second module, you will navigate and practice the nuts and bolts of Tableau Public. It's a pretty intuitive piece of software, but it's important for us to be on the same page with the basics.
- The third module is being taught by my colleague Suk Brar. Suk has over ten years of experience in statistical and predictive analysis, data forecasting, and data visualization. He brings with him a unique perspective, as he has worked in various industries throughout his career. In that module you will **prepare and import data into Tableau Pub**lic.

- In the final module of this course I want to take a step back and help you think about how to contextualize your questions so that the data visualizations can answer those questions. You will be **able to determine how your visualization will address the who, what, and how of a data context**. Again, I'm really excited you're taking this course and I think that you'll learn a lot. So let's dive right in and get started.

Data visualization is an essential toolkit to everyone's career. It gives you that edge when presenting your results. But sometimes it's not intuitive. So, to get us started in building this skill, we will discuss why we visualize data? And define common terminology. Also, within this module, we will identify what

types of software options are available to do visualizations. You will install the Tableau Public software on your computer and do your first visualization. I'm so excited to have you join me for this, so let's get started.

- https://www.tableau.com/sites/default/files/media/designing-great-visualizations.pdf (pages 3-8). This article presents some iconic and important visualizations from the past. Some of the techniques employed in these visualizations are still in wide use today. Why do we visualize quantitative data? This blog post by noted data visualization author Stephen Few provides a good primer on why we visualize quantitative data. Who Does What (Infographic) Check out this infographic about the various roles associated with data analytics in larger organizations. Note which of these professions are most associated with visualization of data.

In this lesson, we're going to answer the questions, what is data visualization and why do we do it? I want you to be able to articulate three specific ways visualization can aid in business analytics. But first, we need to ensure we all recognize and can define the **basic terminologies of data visualization and business analytics**. I'll touch on the key terms in this lesson but I've also included a couple of readings that are essential to help you define other terms.

What is data visualization and why do we do it? Simply put, **data visualization is using pictures to represent data**. It's been around for hundreds of years, but recent technological advances have made it easy to summarize large volumes of data very quickly. For example, one can take 50 years of crime data in the United States and illustrate patterns and relationships across time, geography and types of crime. In theory, it could have been done before we had computers, but now we can churn through millions of rows of data in seconds.

Plus, we can have interactivity that wasn't possible in those iconic and historic visualizations. Business analytics is about **asking questions and using statistical and quantitative tools for explanatory and predictive analysis that answer questions**. Data visualization is at its heart a business analytics tool because it lends itself well to what's called data-driven or fact-based decision making.

Decision makers can click through visualizations and drill down to the details that they need really quickly. If a decision maker can't drill down and the information isn't readily available, they'll have to ask the analyst for more

information, which could take weeks or months before it is compiled. Why should we do data visualization when the same information might be available in say, Excel tables or in spreadsheets? The first reason is that it just looks nicer, but more importantly, **we can't really process data or see insights using just spreadsheets**.

This is a phenomenon known as **the picture superiority effec**t. For example, suppose you were given a spreadsheet like this. Let's see if you can find the five highest revenue product regions from the data here. You can even pause the video and see if you can do it. Even if you were able to find those five, I'm sure it wasn't that easy. But, if I as an analyst decided to add just a smallest amount of color using Excel's very simple and straightforward conditional formatting, all of a sudden, it's completely evident. All with the click of a button and you've provided the decision maker the information very quickly and easily. But this is just for small amounts of information that gets compiled into spreadsheets.

But what about the large amount of information that we get from a wide variety of sources in every field imaginable? **This large amount of data has spun off a specialization around what is known as big data**.

The premise of big data isn't that it's just a lot of rows and columns in a data table but it's actually something very different. To be real big data in the way that it is commonly understood by professionals, data must be of three components.

- One, increasing in volume,
- two, increasing at a very high velocity and
- three, with a lot of variety.

Big data also drops the assumption of statistics to uncover relationships that may seem arbitrary, but after it's been analyzed through a computational process known as machine learning, the **relationships begin to gather structure**.

The wind map that you see on the screen is an example of one way that big data can be visualized. Another example of big data is real time monitoring of credit card transactions. When you use a credit card in a certain way that may be unusual, you may received a text message or a phone call asking you to confirm if the transaction is valid. That's because big credit card companies have analyzed patterns of use and can predict what a fraudulent transaction is and what it isn't.

There are times when big data and data visualization get conflated with other essential terms, for example, business intelligence. **Business intelligence is a set of tools and techniques that data from various databases gets stored in data warehouses**. And gets converted into meaningful information through the use of queries and reports that are often developed by IT professionals. Ideally, analysts will take information stored in various formats by the IT professionals and coax it into data visualizations that could be **used for decision making**.

That's enough of the terminology for now. In the next lesson, we'll be looking at some software options for doing data visualizations. Make sure you review the readings to get more familiar with. The terminology that I touched upon and with other terminology you'll hear around data visualization and analytics.

Welcome back. Because there are so many data visualization tools out there, it can get rather confusing. In this lesson, we're going to examine the types of technology tools used to do data visualization. We will be using the public version of Tableau for our course. But keep in mind the things you will learn in this course are very much software neutral. So it doesn't actually matter what piece of software you use. First, let's look at the apps that everyone has some passing familiarity with, which is **Microsoft Excel**.

For the versions beginning in 2010, Microsoft included a tool called PowerPivot that allows advanced data connections and improved pivoting capabilities far beyond of what was on offer before in Excel, which was the standard pivot tables. If you don't see it immediately available in your version of Excel, check the add-ins and see if it's activated. There's a real learning curve for the tool called PowerPivot. But it is in fact a very powerful option for those who are tied into the Microsoft architecture, especially through SharePoint and PowerBI.

If you don't have those products, then sharing it across teams will still have the same limitations as before with Excel, when you were using standard pivot tables. However, you have much improved data connections in ways that simply were impossible before.

Secondly, when you're searching for data visualizations on the web, you often run across either **Python or R as data visualization tools**. For machine learning and big data, which we sort of defined in the last lesson, those are possibly good options. But I would strongly caution you to learn and become an expert in either Excel or Tableau before diving into Python or R. Even though both of this products are free, they are both **incredibly difficult and overwhelming to**

**learn** if you're a beginner in this area. It will be very frustrating for you unless you're a strong programmer. That being said, both Python and R are in fact tools that are extremely powerful. And once programmed, can process unstructured data and big data much more quickly and efficiently than any other tool out there today.

Finally, we come **to Tableau's products**. They're different flavors of the software, which can be confusing to a lot of people. So I'm going to go over them one by one so that you have a sense of them.

First is Tableau Public. That's the software we're going to be using in the course. It's free but it has a notable limitations. First of all, all of the information, including your raw data are stored in the Tableau cloud. And can be downloaded by anyone, with no controls possible. That means you have to be extra careful with your data before it's saved. The other caveat is that the files cannot be saved locally. You have to save it to the cloud. If you have data that's not anonymized in any way, you'll have to prep it ahead of time in Excel or some other program before importing it into Tableau Public. Because once the data are imported into Tableau, it becomes a public document that can be downloaded by anyone. **Tableau Desktop**, on the other hand, which is not free, can connect to an amazing array of data sources and has very strong security for your data. It's the full version of Tableau. And you can get data from your own data sources, from your own databases, your own data warehouses, SQL Server, Google Analytics, Oracle, Amazon Web Services. Virtually any data source can be connected through Tableau Desktop.

And it really is a very powerful tool if you can get a copy of that license. **Tableau Server** allows the most flexibility of all of the Tableau products. But it is only available to those who can support a server that is dedicated to Tableau. As such, there's a very large financial investment for an organization having this version. But if your organization does, count yourself lucky because you could harness the full capability of the software in ways that no other version can have. For example, you can control how others see your data, even when it's embedded on a public website. Then you can have more secure access to data and prevent people from downloading those data sets.

And the final type of Tableau product is called **Tableau Online**. Tableau Online is intended to be a way to allow control of your data when embedding data visualizations. But it doesn't require a large investment of money into a server. The disadvantage of this approach is that in order to view the visualization online, the viewer needs a Tableau Online account as well. So visualizations

produced with Tableau Online cannot be shared with the public like it can with Tableau Server or Tableau Public, unless you save it as a Tableau Public file. Whatever piece of software or app you choose, you can be assured that this course and this specialization can address your needs. So say tuned for the other parts of the class.

Welcome back. After all of the preliminaries, it's time for the fun part, the first visualization we're going to do in this course. I don't want you to worry about some of the details for now. We'll delve into those later. What I want you to do is follow along. You could always pause it or even go back a little bit if you miss out on something, but don't worry too much about it. Just watch and replicate. From your web browser, go to the URL shown on the screen. Make sure you type it in exactly as it's written there, otherwise it won't work. If you need to, feel free to pause the video so that you can make sure that you have it onscreen before you continue. Click on the Download Workbook in the upper right-hand side of the website, as shown in the screen in front of you. Again, make sure that you have it downloaded fully before continuing so that we're able to be on the same page. The file will download. Depending on your browser setting, you will either be able to open the file immediately, or you'll have to open it from your saved location. In either case, make sure you open it in Tableau Public and come back to the video. Once opening Tableau Public, I want you to drag the country field away like I'm doing. I'll show you here a few times so that your screen will look just like mine. The reason the country field is there in the document is because Tableau Public won't allow me to save a file without at least a little visualization. So this is just giving a small visualization so that you all could download it. And I want to drag it away so that it's a completely blank slate, and we could start from there. Now we'll go very slowly so you can replicate exactly what I'm doing here. But if I'm going too fast, which sometimes happens, just hit the pause button on the screen and then you can catch up. You can even go backwards and see something that you may have missed.

But the goal of this is really just to see how quickly we can actually do a meaningful visualization with data that we already have. Okay, now let's do this.

The first thing we're going to do is to set up our map. I'm grabbing the longitude field with my mouse and dragging it to the column shelf. Next, I'll grab the latitude field and drag it to the row shelf. If you make a mistake, just click on the back arrow as seen on the screen. That undoes the last action that you did. Again, you're dragging the longitude field to the column shelf and the

latitude to the rows shelf. Drag the country field and drop it over the map. To do this, click on the country field and hold the left mouse button down, drag it over, and just let go of the mouse button. Look for the folder that says development. Click on the folder and expand the options as I'm doing. Select the CO2 emissions field. Drag it and drop it on the color button, as I'm doing here. On my screen, it seems to show that the higher the CO2 emissions, the greener the color. That doesn't make sense, since CO2 is a type of pollution. So we need to change the color of the map, for sure. To edit the color, I'm going to just click on color and edit the color to an orange color range instead of the green that I have on my screen. Now I'm going to add a filter. Let's do a filter on years so that we can highlight specific years instead of showing all the years In one map. What you do is you right-click on years and say show filter. Then you go all the way to the right-hand side, you click on the little arrow, and then select the single value list so that we can see the years and how things change from year to year. We're going to be using this dataset a fair bit in our exploration of data visualization.

This first example is just a taste of what we will see. It'll be exciting to see all of the different ways we can look at the data to allows us to answer questions and discover interesting insights. And we'll be doing this for the next several classes, so stay tuned. And we're going to do even more advanced visualizations, more advanced dashboards, more advanced story boards. So stay tuned.

## Exploring and Navigating Tableau

With the last module, you were able to create your first visualization through guided practice. The secret to doing visualizations is really knowing the tool you will be using. For this module, you will explore and navigate the Tableau interface and be able to use specific tools as you begin your visualization journey.

### Lernziele

- Examine and navigate the Tableau Public workspace
- Develop a visualization
- Use the "show me" feature

Hello and welcome to the second module which is exploring and navigating Tableau. In the last model I introduced you to Tableau and you were able to

do your first visualization. In this module I want to really examine the different tools that Tableau has within their interface. It will make things a lot easier if you can really identify and memorize what each tool is for as you can begin to work quicker with your projects. For this module you will navigate the Tableau Public workspace while doing some visualization. We'll tackle rows, columns, and mark shelves. In addition, you will use the show me button to make some quick and dramatic visual changes. So let's get started.

Welcome back. This lesson is going to walk you through importing visualizations from Tableau Public's website into our version Tableau Public that we previously installed on our computer. You can pull any visualization from the Tableau Public website onto your computer version.

Today we're going to download the visualization that replicates John Snow's 1854 cholera study, which I showed you in the first module. Let's dive in! This is a very well designed visualization that combines a historical context of epidemiology and public health, with the design principles needed to tell a story. It isn't an extremely difficult project, from a statistical point of view, but much thought clearly went into the original design in 1854. As well as the redesign done very recently.

The visualization is assembled as a story. **For data people, a story is a way to explain your data in a systematic way**. This visualization starts with a hypothesis that cholera isn't transmitted through the air as was widely assumed in The 1800s. I want you to click on the collecting the data tab in this visualization and you'll see the sort of schematic of the door to door sampling that was done in 1854. Now, click on mapping the results. That's a tab there. It describes the process, that you as a data visualization expert, after completing this specialization, of course, and others go through in thinking through how best to show the data to tell the right story.

He showed all of the deaths from cholera, and added the locations of the water pumps in the neighborhood. Because of data visualization, specifically, in this case, mapping, you can see in the tab labeled Snow's Analysis Focus on Broad Street, that there is a pocket of deaths around one water pump.

This is an example of both the data driven and as an actionable visualization. It's data driven because there are specific elements of the location of wells and homes along with the number of baths in each of the homes. It's

actionable because Snow could then take your irrefutable evidence and get the government to immediately shut down the well. Let's now download this into our own computer. Scroll down on the website and you'll see download on the bottom right of the visualization. When you click download you will get options that you see on the screen as well. Click on Tableau Workbook, then choose Tableau Public as the app use to download the Workbook. Your computer will then open Tableau Public and the visualization will be displayed as shown on the screen. What you'll see, Is definitely not how it looks on the web site, but it's more or less that way. That's because Tableau hasn't opened the file for editing. On the left-hand side, you will see a list of the Dashboards and Worksheets. If you double-click on, for example, Map of Outbreak, you will see that particular visualization.

Below that, you will see a section labeled Navigator, and one labeled Story. Don't worry too much about these yet, we'll learn all about them in a future lesson. Now I want you to click on the Cholera Cases tab at the very bottom. This is called a worksheet and is where all your visualization are first going to be created. Here you see what the author did. He use the geography field as the column and row to represent latitude and longitude. Then he dragged the number of deaths at address to both color and size to highlight larger numbers of deaths. Finally I want you to click on the little icon to the right of the dimensions text at the upper left area. It opens the Underlying data that's used to illustrate the visualization. As you can see, there's latitude and longitude represented as decimals and thus shows exactly the location, the spatial location, where those dots are supposed to go. And there are other data fields, including the number of deaths. That sort of illustrate the issue.

Now, I want you to spend some time navigating with the tabs and other aspects of the work space. Don't worry about messing up the visualization. It won't be saved, so you can take away the stuff and put it back in, add some other things if you want. You don't have to worry about it messing up on the website. This is in your own computer. And this is just how, this is just to get you to see how it's done. But in the next lesson we're going to go through the specifics of the workspace. Again just please make sure you go through, navigate some of these things understand where it's coming from and then we're going to go through the specifics of the workspace. I'll see you there, thanks.

Welcome back! For the next three lessons, you'll be doing some hands-on exploring and navigating through Tableau Public. You'll create a visualization of mobile phone usage, and **assess how different categories of data behave as you input and change settings**. I'd like to encourage you to explore on your own and share what you discover as you work your way through these lessons. We'll begin by explaining some very important Tableau Public limitations that will directly affect how you manage your files and protect data privacy. You'll then see how to use the helpful describe feature to assess information. As you begin the following lesson, you'll start building a visualization on mobile phone usage. You'll learn some specifics about shelves. I'll tell you about a quick and easy way to use the undo feature.

Then you'll get to modify data and illustrate different visualizations on separate worksheets. Finally, in the last lesson, you'll work with other important places on the Tableau Public workspace, including **control properties like type, color, size, and shape, and the very important Show Me button.**

These lessons will be intense, hands on explorations of the Tableau workspace. It may be frustrating at first, but you'll be a stronger user of the software after you're done. So let's get started. First of all, as I mentioned in the very first lesson, there are some notable limitations with Tableau Public.

The first significant limitation is

- **you cannot save your file to your local computer**. The reason, I suppose, from Tableau's perspective, is that they want your visualization to be public. But from your perspective, it means you cannot save your files offline and you must be connected in some way to the Internet. If you're not connected to the Internet and your computer crashes, then you lose all unsaved changes. The other significant Tableau Public limitation is that
- **all of the data that you import into Tableau Public is fully available for anyone to use**. If you have any information that can identify someone, it should be removed before importing into Tableau Public so that the data are fully anonymous. When I talk more about data prep, I'll show you how to prepare your data in Excel so that it is in fact anonymous. There are a couple of other limitations that are less important to worry about, but still worth thinking about. One is that there are **fewer ways to connect the data in Tableau Public**. Other

Tableau versions like Tableau Desktop can directly connect to virtually any data source that you can think of.

- Tableau Public, however**, is limited** just a few. It is free, after all. It doesn't really matter too much, because it can connect to Microsoft Excel, and Excel can connect to almost anything. Sometimes it just takes an extra step, but it is an extra step.

Well, having gotten all of that out the way, let's start looking at some of the workspace. Let's use the map visualization that we used earlier. The URL where it's located is on the screen in front of you. Please pause the recording until you have it so that it's easy to access. You should be on Sheet 1 with a bunch of pictures of countries shown. First, let's drag the Country field off the Rows shelf for now. On the left-hand side, you'll see data in two categories, Dimensions and Measures. Dimensions help to categorize the data. Measures help to determine the scale of a category. Next to each field, you'll see either a hash sign, a paper clip, an ABC symbol, a small globe, or a hash sign with an equal sign in front of it. There are also folders. When you click on them like I'm doing here, other fields show up. In this case, there are mostly fields with hash signs. Let's look at an example. Let's look at the field called Population Total in the Population folder. This field has a hash sign next to it. That means, plain and simple, that it is a number. We also know that it's a number because of where it's placed in the Tableau workspace. That is, it's a measure. But what else can we learn about this field? Using your mouse, right-click on the population total, scroll all the way down to where it says Describe, and click. You'll see a lot of information here.

For example, the role is what's called a continuous measure. That means that the field can take on any number. We also see in this box information about where the data were originally sourced. Click on the Load button, then scroll down. You'll see a summary of the data ranges from about 18 or 19,000 to 1.3 billion. Now, click on Close. It's a good idea to do that for other fields as well, and we're going to go through two more. Now let's go to one of the Dimensions, and right click on Region. Click on Describe, then click on Load. We see that there are six possible values that categorizes a country into one of six possible regions. And that's what I was going to show you. So now you can click on Close. Let's look at one more. Right-click on Year, then click on Describe and load the domain. We see that the data are from 2000 to 2012. So this is a pretty good amount of data that we're working with here. That's really great, because I'll be using this data set to illustrate some of the cool things we can learn from data through

visualization, and having a large number of years really helps in that regard. We'll continue our exploration of the Tableau workspace in the next video.

We're going to continue our survey of the Tableau workspace, so let's just get into it. The best way to learn an app or software in general is to use it. So we'll go through the process of doing some visualizations, that way we can have some of the details in the Tableau work space. So first, let's drag the longitude field to columns. Now let's drag the latitude field to the rows. What you should see is an unfilled map like you see on the screen here. There's a reason why I'm showing a math visualization here. Longitude and latitude are used in geography in the same way as rows and columns in data. Columns go up and down. Rows go left to right. How you place the data fields will determine how the visualizations show up.

It's not always intuitive at first. So I encourage you to mess around as you're doing your own visualizations to ensure that you're getting the visualizations in the way that you want and need. The nice thing about Tableau is that there is the all important undo feature. In Windows you can use the standard Ctrl+Z shortcut to go back to what you had before. But you can also click the left arrow as I'm demonstrating on the screen. And often, because you're doing a lot of manipulation with your mouse, using the left arrow is probably what you're used to most. But just know that you do have in Windows the Ctrl+Z option. When you place fields on a row or column shelf, it matters if the field is a dimension or a measure. A dimension will be placed on a shelf as is with Tableau splitting them into categories. With measures, Tableau will do a calculation automatically. And add that based on any dimensions that were added. Let's see an example. Click on the new worksheet button below. A new worksheet pops up.

You can add **unlimited numbers of worksheets** here to test different ways to do visualizations without messing up the ones that you've already done. New worksheets are Intended to illustrate different visualizations, different ways of looking at things, different types of experiments on the data, just testing out things. That's what the different worksheets are for. So feel free to have tons of worksheets as you're going through and learning.

Now, let's add a year to the row shelf by dragging the years field over to rows. Then click on the development folder if it's not all ready opened and drag the GDP field to the columns Notice that for Years, Tableau keeps the field as Years. That's because it's a dimension. But for the GDP field which is considered a measure, Tableau calculates an average. In this case, an

average is appropriate, but it's not always the case. Sometimes Tableau might calculate a sum or account. Tableau is trying to interpret the data for you.

Sometimes it gets it wrong, sometimes it gets it right. You just have to make sure that what it's doing is right. Otherwise, you can change it, which I'll show you in a future video. Tableau shows the data as a line chart but it looks really weird. How can we fix this? It's easy, we simply have to click on the swap fills button. Watch carefully as I press the swap button. The swap button allows me to easily fix how the visualization is displayed. Now it's displayed in a way that's similar to other line charts. There's so much to learn in the Tableau workspace, so we're going to continue our exploration of the workspace in the next video. My recommendation is to keep everything the same and join me there immediately so that we can just continue our exploration. See you then.

So we're still working in the same space as in the last video. I want you to double-click on the worksheet tab we just created, and type in Line Chart of World GDP. Navigate to the other worksheet. Let's rename this World Map. We should have Latitude in the Rows shelf and Longitude in the Columns shelf. We see the Country field here, but what should we do with it? This is where the Marks shelf becomes important. With your mouse, click and drag the Country field to the Marks shelf, and drop it exactly how I'm doing it on the screen. I'm repeating it a few times just so that you know what I'm doing in the process. The map should change colors. On my screen it's blue, but it could depend on the computer and the display resolution. Having taken care of the Country field and putting it where it belongs, let's click and drag the Mobile Phone Usage field over to the Marks shelf and drop it over the Color button. You should see a change to graduated colors. I want to take a moment to explain what is happening when you add information to the Marks shelf. The Mark shelf is the spot where you drag fields to control properties like type, color, size, and shape. You can add the same field to another button to get more information. If we take the field Mobile Phone Usage, the same field I was doing with the color, click and drag it to the Label button and drop it there, we'll get values. I mean, it's frankly pretty ugly and not necessary so we could drag it away like this. However, the value of this is to actually be able to add more texture to your information and to your visualization by using the same fields in different ways to get different visualizations. So let's take a peek at the attributes of one of the fields, specifically the Year field. Right-click and follow along. It looks like it ranges from 2000 to 2012. That's a really nice range. But at the moment what you see on the screen is all of the years combined in an average. That doesn't

make any sense intuitively or otherwise. It's just plain wrong. So this is a good opportunity to introduce what's called the filter. I want you to right-click on the Year field. You see a bunch of choices we'll explore over the time that we're together, but for now, I want you to click on the Show Filter. Now on the right-hand side, the Year shows up as a bunch of check marks. Just hover over the Filter as you see on the screen. Click on the down arrow and select Single value list. Now only one Year is shown at a time unless you select All. You could then click on each of the years to notice that the color shading changes as you click the years. My aim is to show you the workspace. However, as someone who does visualizations for a living and has been doing it for a long time, I could tell you that there's some issues with how this shading is done. There's something going on there, and I would have to delve into it more. It will need to be fixed, but we're not going to get into that right now. But maybe when the video is done, you could always take a moment and see if you can fix the issues yourself, because I think you could probably do it. I'd like to show you one more thing in our survey of the workspace.

It's really important, it's called the Show Me button. Just click on that. It opens up a list of possible other visualizations besides a map. I want you to click on the heat map, which is in the top middle row. To show you some of the flexibility with the Tableau workspace, I'm going to drag the Year field away from the Filter, like so. Hmm, that's very interesting. I'm going to make the Columns and Rows a bit bigger, so that it's a bit more visible. Next, I'm increasing the Size of the boxes a bit, for a bit of an aesthetic improvement. Let's find the field called Regions. Let's right-click on that and click on Show filter. Again, change the Filter to a list as we did with Year previously. What you're seeing, as I'm going through a very basic visualization, is each of the pieces of the Tableau workspace get coordinated into a visualization.

I started this visualization with a map, but now it's not a map at all. I'm going to leave you with this. The purpose of this lesson is to give you a short, but intense, look at the Tableau workspace. There's going to be some readings and links that will help you to learn more about the Tableau workspace. But in doing so, we went through some visualizations, and saw there are different ways to show the same information.

We'll be spending a lot of time determining which way is better, but this should whet your appetite. In addition, as you noticed, it's often easier to learn about the Tableau workspace, and frankly any other type of program, by actually doing the thing that you need to be doing. Whether in Tableau it's

visualizations, or in Python whether it's programming some cool new thing, doing it often helps you to learn the program.

It's frustrating at first, but you'll actually be a stronger person, stronger programmer, stronger user of the software after you're done. My colleague, Sikh Brar, will actually be going through how to do the data preps. I'm not going to be doing that, however I'm going to come back and actually discuss the context of visualization. Funnily enough, you're actually not going to be needing Tableau or any other software in that particular section of the course, but it's going to be talking about the context, which is actually extremely important. So stay tuned, and thank you so much for joining me.

- Watch the Tableau Public Overview training video (Video 1) on the Tableau Public Resources page.

Read this blog post by Dr. Mu Lin on Why Data Visualization Matters

## Making Data Connections

Creating visualizations require data and in this module, you will discuss the various data sources for visualization and specifically what can be used in Tableau. You will prepare your data and identify the types of data connections possible with Tableau. You will be able to connect and merge to multiple data sources which can help make your visualizations more powerful.

### Lernziele

- Prepare your data for import
- Differentiate between live data connection and a data extract
- Identify types of data
- Practice and connect to different data sources

So what are data connections in Tableau and why are they important? Data connections allow Tableau access to your data so you can create beautiful and interactive dashboards. Making the correct data connections is very important

in Tableau because data can be stored in various types, like Microsoft Excel, text files, or statistical files or in various locations such as your computer or on a server.

Understanding what type of file stores your data and where the data is located will dictate what type of data connection you will need to make. In this module we'll focus on preparing you data for import and how to connect to different types of data using Tableau Public.

Then, we will distinguish between a live data connection and a data extract. After that, we'll identify the types of data that Tableau Public can and cannot connect to and look at the ways to get around the limitations of Tableau Public.

Finally, we'll use Tableau's data source section to connect to different data sources and merge multiple sources of data together using common variables.

In this lesson, we'll focus on preparing your data for import, and I'll illustrate how to connect to different types of data using Tableau Public. Before connecting to your data, you need to make sure that the data is correctly prepared. If the data is not prepared correctly, it will be very difficult for Tableau to recognize and utilize for analyses. Let me show you what I mean. For Tableau Public, there are three steps to format your data correctly. First, make sure that your data begins in column A1 and that the first row of data only includes column headers. Second, make sure each column of data contains the same type of data throughout the column and each row of data is one data item. Finally, remove any column or row totals. Once your data is correctly formatted, you are ready to import the data into Tableau Public. When you first open Tableau Public, you will be directed to the Tableau Public home page. On the left side of the screen, you will notice that Tableau Public gives you the option to connect to an Excel file, Text file, Access file, or Statistical file located on your PC or Mac, or the option to connect to an OData or Web Data Connector Server. The Excel, Text, Access, and Statistical files will be local databases, such as files located on your own computer, while the OData and Web Data Connector will allow you to connect to an online database. We will discuss first how to connect to one of your local databases and then I will show you how to connect to the OData and Web Data Connector.

So let's first connect to your local database or **an Excel, Text, Access, or Statistical data file** by clicking on the file type. For this example, we will use data that is stored in a Microsoft Excel file. From the open file window, navigate to where the file is stored and click "Open". Tableau Public will import

every sheet that resides within the Excel workbook and each sheet will appear under sheets here at the bottom of the worksheet, similar to tab worksheet file management people have come to expect from Excel. Click on the sheet that contains the data you would like to analyze and drag it to the drag sheets here area. You now have a visual representation of your data, and it should look very similar to the original Excel data. As you can see, each column has a header name and the data type is automatically identified by Tableau through the use of icons. An ABC icon represents string data, for example, data that includes names. A globe icon represents geographical data, for example, data that includes cities, states, zip codes, et cetera. A calendar icon represents dates, and the pound icon represents numeric data.

If Tableau incorrectly assigned a data type, you can click on the associated icon and choose the correct data type. If everything looks correct, you can click on sheet one and begin analyzing your data. Shifting gears, let's take a look at Tableau's Open Data Protocol Connection. An **OData connection defines a protocol for acquiring and updating of data utilizing existing web protocols**, such as Microsoft SharePoint server lists. OData is intended to be used to access information from a variety of sources including, but not limited to, relational databases, file systems, content management systems, and traditional websites. If you want more information on OData connections, I have included the link to visit the Microsoft OData within the resources. To make an OData connection, first we click on OData from the Tableau Public home page under servers. From the OData connection window, enter the URL to where the data is located. If no authentication information is necessary, click on "None". If authentication information is needed, enter your account key for Windows Azure Marketplace Data Market or your username and password for the URL you entered. We can click "Okay", and the connection will pull in the available data for you to analyze. For more information on Tableau's OData connection, I have included a link within the resources.

**Finally, another connection type that Tableau Public provides is a Web Data Connector**. A Tableau Web Data Connector gives you a way to connect to data that doesn't already have a connector. Using a Web Data Connector, you can create and use a connection to almost any data type that is accessible over HTTP. This can **include Internal Web Services, websites, JSON data, XML data, REST APIs**, and many other sources. Because you control how the data is retrieved, you can even combine data from multiple sources. You will need to use the Web Data Connector when you need to connect to any data source that is accessible via JavaScript, but not currently listed in the native tab flow connections. In order to make use of a Web Data Connector, you don't need

programming experience whatsoever. If you are aware of a Web Data Connector you want to use, all you need to know is a URL to access that. Simply select the Web Data Connector from the connection list, enter or paste the URL of the Web Data Connector into the appropriate window and follow the instructions presented. The URL of the Web Data Connector may be different from the site that holds your data. If the connector displays a web page, enter any information that you are prompted for and then submit the page. Wait while the connector retrieves your data and imports it into Tableau as an extract. When Tableau opens a new worksheet, you can begin your analysis. Each Web Data Connector is unique and its interface depends on how the author has designed it. For additional information on Tableau's Web Data Connector, please visit the links in the resources. However, if you do not have a Web Data Connector, you will need to create one. This involves programming and knowledge of JavaScript. Additional information on creating a Web Data Connector can be found in the resources. Now that we have our data imported, in the next lesson, we will discuss the two primary types of data connections for Tableau, a live connection and a data extract.

Welcome back. In the last lesson, we discussed how to connect two primary sources of data within Tableau. In this lesson, I will define the two primary types of connections for Tableau. A live connection and a data extract. Let's get started.

A live data connection means that Tableau sends queries to your underlying database and retrieves up to date data whenever the Tableau dashboard is updated**. A live data connection is best and should be used when the live connection between the Tableau dashboard and your data is possible**. Sometimes your data maybe stored on a server and a live connection is not permitted. And when the connection is not slow. Sometimes a live data connection becomes very slow when the data is not stored locally on your own computer or when the size of the data files is very large.

The other type is a Tableau data extract connection. Which means that Tableau sends queries to your extracted or static database and not the underlying data. Any updates to the underlying database will not update the Tableau dashboard. You will have to refresh the Tableau extract to get the up-to-date data.

**TDE should be used when a live connection to your data is not possible, when the live data connection is too slow, or in such cases where the size of the data is extremely large. In general, TDE should be used when your data is**

**located on your own computer, a real data feed to your dashboard is not needed and when the data size is sufficiently large.** For more information on when to use a live connection versus a Tableau data extract, please visit the resources section. So when should you use a live connection versus a Tableau data extract. Well, it depends. A Tableau data extract is recommended when your database is too slow for interactive analytics or when you need to be offline and will not always have an internet or network connection to your data. A live connection is recommended when you have a fast database or when you need up to the minute data. If you would like additional resources on the material that we've covered in this lesson, please visit the links below. The Tableau website provide very helpful instruction videos including introduction to working with the Tableau Public interface, connecting to data, preparing your data for import, different types of data connections etc. Now that you have prepared your data for import and have connected to different types of data, in the next lesson we will prepare your data for Tableau Public.

In previous lessons we focus on preparing your data for import. Connecting to different types of data using Tableau Public and distinguishing between a live data connection and a data extract. In this lesson we will cover preparing your data for Tableau Public. We will identify and demonstrate the different types of data that Tableau Public can and cannot connect to. And we will also demonstrate ways to get around some of the limitations of Tableau Public. When you first open Tableau Public you will be directed to the Tableau Public homepage. On the left side of the screen, you will notice that Tableau Public gives you the option to connect to an Excel file, a text file, an Access file, or a statistical file located on your computer. Or the option to connect to OData or web data connector server. The Excel, text, Access, and statistical files will be local databases, while the OData connection will allow you to connect to an online database To connect to your Excel, text, Access or statistical data, click on the file type. So for example, when you click on Excel, you will notice that you have the option to connect to a .xls, .xlxsx or xlsm file. When you click on text, you will have the option to connect to a character delimate file.csv tab delimate files .tab or .tsv, or text files.txt. When you click on Access you will have the option to connect to a Microsoft Access database .mdb or .accdb. And when you click on Cisco file, you will have the option to connect to a SAS file .sas7b.spsfile.save or our files .rdata or .rda Let's take some time and use Tableau to connect to a data file. For this example, we will use data that is stored in a Microsoft Excel file. From the Open file window, navigate to where the file is stored and click Open. Tableau Public will import every sheet that resides within the Excel Workbook and each sheet will appear under Sheets here at the bottom of the worksheet. Similar to tab worksheet file

management, people come to expect from Excel. Click on the sheet that contains the data you would like to analyze and drag it to the Drag Sheets Here area. You now have a visual representation of your data and it should look very similar to the original Excel data. As you can see, each column has a header name and the data type is automatically identified by Tableau through the use of icons. An ABC icon represents string data. For example data that includes names. A globe icon represents geographical data. For example data that includes states,cities, ZIP codes et cetera. A calendar icon represents dates and the pound icon represents numeric data. If Tableau incorrectly assigned a data type, you can click on the associated icon and choose a correct data type. If everything looks correct you can click on Sheet One and begin analyzing your data. [BLANK AUDIO].

If your data resides online you can make a connection through an OData or web data connector. A Tableau web data connector gives you a way to connect to data that doesn't already have a connector. Using a web data connector, you can create and use a connection to almost any data that is accessible over HTTP. This can include internal web services, JSON data, XML data, Rest APIs and many other sources. Because you can control how the data is retrieved, you can even combine data from multiple sources. You create a web data connector by writing a webpage that contains JavaScript and optional HTML. Please visit your resources section for more information on web data connectors and JavaScript. After you've written a web data connector you can share it with other Tableau users by publishing it to the Tableau server. To help you create web data connectors, Tableau has created a software development kit that concludes templates, example code and a simulator that lets you test web data connector. This documentation also includes a tutorial that walks you through how to create a web data connector from scratch. At this time Tableau Public can only connect to Microsoft Excel, Microsoft Access, multiple text file formats, statistical files and Web Data Connectors. All other data sources such as Tableau servers, SAS servers, MySQL servers, Amazon Redshift servers, etc at this time are not supported by Tableau Public. If you currently have data in a data source that is not supported by Tableau Public, I would recommend transferring your data into one of the data sources that is supported by Tableau Public for analysis. Another limitation of Tableau Public is that currently it is limited to working with 10 million rows of data per workbook. If your data exceeds this limit, I would suggest trying to break it into smaller data sets so Tableau Public is able to analyze your data. Before we conclude this session, I would like you to practice connecting to different data sources and importing your data into Tableau Public for analysis. In part four of this module,

we will use Tableau to connect to different data sources and connecting different data sources together. Good luck.

Welcome back, in the last lesson, we discussed the types of data that Tableau Public can and cannot connect to. In this lesson, we will use Tableau Public to connect to multiple data sources and **merge different data tables together**, using different types of joins. When your data is located in two or more data sources, such as multiple sheets from an Excel workbook or multiple files from different sources, you will need to merge or join the data sets together through specific fields to access data from all of the data sets. With Tableau, you can use multiple data sources for a single worksheet by joining multiple tables. One data source is primary, while all others are secondary. While there are no set requirements for which data set should be primary versus secondary, I recommend making the data source with the majority of the data you will be working with the primary data source. For the primary data source, Tableau selects a data source of the first field you add to the worksheet. The primary data source creates relationships with fields of the same name and data type that it finds in the other data sources. If you have the same variable in two or more databases, make sure the variable name and data type are exactly the same across all data sources. Otherwise, Tableau will not be able to recognize that the variables are the same across all data sources. You will first want to connect to your data and create your data source, or data sources, as we discussed in the other lessons of this module. After you select the database, double-click or drag a table to the join area of the data source page. Then double-click or drag another table to the same join area. Tableau will join the data sources together by finding similar data fields in all sources. Before we continue, let's take a moment to discuss the different types of joins. There are a few different types of joins that Tableau will provide. An inner join, left join, right join, or a full outer join. Let's talk about each one of these and how or why you would use them. An inner join will show the records or rows present in both tables, if there is at least one match between columns. You will always use an inner join when you want to display all of the data that is stored in both of the data sources, using a merging variable. A left join returns all of the records or rows present in the left table and matching rows from the right table. You would use a left join when you want to keep all of the data in the left table and only merge in limited data from the table on the right. A right join returns all of the records or rows present in the right table and matching rows from the left table. You would use a right join when you want to keep all of the data in the right table and only merge in limited data from the table on the left. A full outer join were shown all of the records or rows present in both left and the right table. You will use a full outer join when you wanted display all of the data

that is stored within all of the tables If the data joined, chosen by Tableau, is not the join you wanted, you can simply click on the join to open the join dialog box. Add one or more join conditions by selecting a field from one of the available tables used in the data source, a join operator and a field from the added table. Inspect the join condition to make sure it reflects how you want to connect the tables. When finished, click the X icon to close the join dialogue box. You can delete an unwanted join condition by clicking the red X that displays when you hover over the right side of the condition. When you have joined tables in your data source, the data pane is automatically organized to use a group by table command. You can turn this feature off, or change how the data pane is sorted, by selecting one of the sort by options on the data pane menu. This option is only available if you have connected to the database live. If you have imported the data, the fields are no longer grouped by table. For additional information on merging multiple data sources together, please visit the resources section.

**Author helps a new Tableau user answer, when to use a Tableau Data Extract (TDE) vs. a live connection: http://drawingwithnumbers.artisart.org/tde-or-live-when-to-use-tableau-data-extracts/ A resource to help you to extract your data: https://onlinehelp.tableau.com/current/pro/online/mac/en-us/extracting_data.html Context of Data Visualization & Course Wrap-Up**

**Data visualization is about telling a story using data**. However, before you can be successful at data visualization, you must understand the

- "who",
- "what", and
- "how" of data context.

In this final module, you will be able to determine who your audience will be and what your relationship to them is. You will analyze a real world application of data context and be able to write out a visualization story based on data context.

### Lernziele

- Define the audience and your relationship
- Examine ways to define your project
- Discuss how you will respond

Welcome to the fourth and final module of the course. In this module, we'll take on the who, what, and how of data visualization. By the end of this module, you will be able to define the audience and your relationship to the audience, examine ways to define what you're going to do, for example a visualization, and discuss how you are going to respond to the who and the what. We will also work through an example of the who, what, and how. Let's get started.

Welcome back. In module three, Sue walked you through data preparation and importing. After those steps, how much can you tell me about the data? Do you know what the data are? How and when the data were collected? Do you know who will look at it, and what they will do with it? These are important questions to answer. In this lesson, we're going to focus on these and other questions you'll need to answer before you begin to visualize data.

Obviously, **it's good practice to be well versed with the data** you are working on. But for visualizing your data, it's essential, since you are visualizing for an audience. You need to know that the data can show certain things and not show other things. Prepping your data isn't a direct way to understand your data, but it will help a lot, even if you can make the most beautiful visualizations, they mean absolutely nothing unless you can explain why you visualized the data in the way that you did. Data visualizations are a way to explain your data, but you can't just start doing it until you understand the context of what you need to show.

Furthermore, poorly designed data visualizations can absolutely ruin what you're trying to convey.

The first thing you need to ask yourself is, **what's the audience**? It is your boss or a colleague? Is it a committee? And if it is a committee, what kind of a committee? Is it a big group? Is it for hundreds of people? Having defined the decision maker or your audience, you need to **think about your relationship with the audience**.

If the audience or decision maker is your boss, then you need to **know what her expectations** are for the visualizations. Are they a

**series of iterative drafts**?
Or is it **a final document**?

If the audience is a committee, the questions you have to ask yourself is, are you a **member of this committee, or are you a guest**?

These basic questions are by no means comprehensive, but I want you to start thinking about this in your work.

There are two components to the what of contextualizing your data visualization.

- The first is, what do you need your audience, boss, or committee to do? This is tough for a lot of people. We might think that the audience knows better, but they only know as much as you tell them. Most on a committee may make a lot of decisions, but they're trying to make these decisions based on the data you and others provide. If you provide unbiased information and state your recommendation, you'll be surprised at how often your recommendation is heeded. Your recommendation must be based on data, however.
- The second is, how you will communicate this need to your audience. This is often decided for you depending on the who. If it's a large group, then you'll have a lot of control over the presentation, but you can't offer a lot of detail. But if it's all in a written document or email, then you should provide a lot of detail, but the audience, the people who receive the email, control how they use the information. Finally, after we know the audience and what we need them to do, we can ask ourselves, how this will happen? In other words, what data are available to support the story we are going to tell? Once you have that, then it's time to visualize it.

In the last lesson, I discussed the importance of knowing the context of your data. It is so important that I didn't even show you any data visualizations. I'm not going to show them in this lesson either. Instead, I'll share with you an example of how context affected a situation in my professional experience. As I take you through the example, I want you to think of an example that's relevant to your expertise or interest. I encourage you to share your example with others. Let's get started. And my example is from higher education.

Suppose that I'm in a university assessment office, and the university just piloted a special web-based summer session that helped give incoming freshmen the tools that they need to succeed in introductory chemistry. What's the who here? There are **several potential audiences for the information**. You can brainstorm on this as an exercise, but I'll list out a few. The first might be

the chemistry department. They may want to know how well the students performed after taking the pilot course. The second one would be the budget office. They will want to know the value of the course, and if it received adequate funding. Also, they would want to know if the program needs to be funded again next year. And three is, prospective students might want to know if it's worth taking time to enroll in the summer before they start college. As you can see from the example, the story you will need to tell to each of these groups would be different.

I'm only going to take one of these on, that of the budget office, and see the what and the how. For the budget office, the **what would be to show that the program was a success and to provide the rationale to continue to offer it at the same, decreased, or increased level of funding**. How would you define success? There are a couple of ways to do so. You could survey participants in the summer program. But perhaps a better way would be to see if there's a difference between two groups, one that took the summer course and the second group that did not. You can present this and show that there is or isn't a difference. If there is a positive difference, then you could justify continued funding for the program. Don't forget the non-supporting data though. If you exclude that, then you will be painting a one-sided story. Contextualize the non-supporting data with the supporting data, so that decision makers have the full story. It may muddy the water, but it's better, because you build trust with your audience and build a constituency in your company organization for data driven decision making. What if there isn't a positive difference in any measure? In other words, despite your personal opinion of the session, if the data do not show that there are some value in a program, then you still need to present the data and base your recommendation on that data. There might be other more qualitative reasons to continue funding the sessions, and you can marshal that separately. **But the data must not be excluded just because it doesn't support your argument**. The final piece of the puzzle is what Nancy Duarte, in her book Resonate, calls **the big idea**.

The big idea has three pieces to it.

- One, it must articulate a unique point of view.
- Two, it must convey the stakes.
- And three, it must be a complete sentence.

That sounds super easy? Let's see if we can use this for our puzzle. I'll give you a moment to think about it in the context of our chemistry summer session. You might want to get out a sheet of paper and pen, and just write it down.

Now remember, there isn't one right answer here. But what I would boil it down to is this.

**The pilot web-based summer chemistry session was a success at improving outcomes for students who took introductory chemistry in the fall. Therefore, we recommend continuing to fund the program next year**.

It sounds really simple, but is not something that's always intuitive. What can you come up with from your own work experience? I'm going to leave you with a quote now. Blaise Pascal, who's one of the most famous mathematicians in the world, once said, or supposedly once said, I would have written a shorter letter but I didn't have the time. In other words, **explaining something clearly and concisely is not intuitive and not easy to do, but it's absolutely necessary**. Thanks, and see you next time.

Welcome back, this is the last lesson in this module. In the previous lessons I'd talked about the importance of knowing the who, what, and how of your data.

Now I will link data context to visualization and story telling. What I'm going to do now is provide an example for my own work experience and expertise. And I'd like to really get you to think about one from your own expertise, work experience, or interest, as well, as we're going along to be able to get the context of data visualization. And what I'm not going to do is actually do a visualization on a computer. I'm going to do some pretend visualizations on a sheet of paper. And to be able to illustrate how we can set up the context to visualization so that we can get to the answer that we need based on the question.

And we're going to start with simply talking about, probably the simplest formula and data visualization which is, what is this context that we're talking about? And **the context is simply the question that we need to answer, with the data, which gives us the story**. Without the question, all of the data that we have can't answer any actual questions. And without data, we can't really get to how, or what, the information is that we're seeking to develop a story. So you need both, both are equally important. And usually in data visualization and data analysis in general, we often will have data, or we have a question.

And we struggle sometimes to be able to combine those two to really illustrate the story that we are looking at, how do we get to this answer of this question, with the data that we have. So it comes to now, which is the issue, and this is

really the issue that we are struggling with at a university, which is often types of questions that come up and higher ed, or even in high school, or wherever.

And education is, how can we ensure that the students have success based on the courses that they're taking. And so the issue here is, for us in this example, is we are concerned about the level of success that students who are taking introductory chemistry are having in their other classes. So when we're framing that as the issue, so I'm just going to write that on my sheet of paper here to be able to illustrate this. Notice this is one sentence that's stating just one issue in the sentence. If you feel like there's a second issue, or you need a second sentence, don't include it in the single issue that we're talking about here. Having defined the issue we want to actually be able to demonstrate that there is in fact this issue. So we have a before, in this case, and then we perhaps have a pilot web based course that allows us to look at the data afterwards. So for example, let's say in 2014 there was an introductory chemistry course, and we had a certain amount of data with exams scores and a certain average exam score and a distribution of these exam scores. We then initiated a pilot web based course that allowed students who are coming in in the summer to be able to take this course to be able to prepare them for introductory chemistry.

And then in 2015 we look and see if the exam scores are better about the same or worse that before. But to demonstrate that we want to be able to show you how it looks in 2014, before the introductory pilot web based program is initiated. And here I want you to do something sort of interesting, I would like you to actually think about what type of visualization might illustrate this particular piece of the puzzle.

**How are you going to show the people who are making the decision**? Some of the evidence about how well students were doing in the summer, or in the course before the pilot summer course. And the way we're going to do that is, we're going to actually just draw a visualization, and we're not going to worry about if it's the right visualization, or if it meets best practices, which we are going to talk about in a later course, etc.

What I want to do is just think about, okay, how would I want to describe and show exam scores for this course? And the way I want to do it, and again, this is perhaps not the way you are going to do it, is I am just going to draw a very small graph here with exam scores, And I'm just going to draw a simple histogram which shows the range of exam scores in 2014. In other words, how well students were doing before the pilot was initiated. Yours may look quite

different, this may not even be the best way to show it, but I just wanted to illustrate it to get a sense in my head.

How I might be able to describe the before the pilot. So given that we understand this as an issue, which is the incoming students do not seem to be as prepared to succeed in intro chemistry as we would like. What are some of the options available to us? One of the options could be this pilot web based summer program, which I have already spoken about, and for purposes of the report, or presentation, or the story, I would want to make sure that the reader or the decision maker has the context of it.

So for example, I would want to talk about how much this web based pilot summer course would cost.
I would want to describe what it does. For example, what's so good about the web based content? What's the faculty interaction? All these things that, questions that need to be answered about why we do these web based programs and why it might be a good thing. And so we gathered that information up and present some other options about what we could be doing.

And so, having in fact done this pilot summer web based course, we can then see, in fact if it made a difference next year Holding everything else constant. And now, what we're going to do is we're going to take this little graph that I drew here and see the, how it looks before the web based course and after the web based course. Here I've just simply overlayed the 2015 exam scores. Which by the way I don't know I'm just illustrating it just to show you how you might be able to compare the 2014, the before, and the 2015 after results.

And here I'm just showing you that there was a An increase in the exam score so that might in fact mean that we are in good shape here. That perhaps we need to have the pilot web based course for another year. And we need to present that. We need to show that based on All of these other issues with the cost etcetera is still worth it to be able to do that. If it wasn't, we would want to recommend that perhaps it's not worth it, it's what the data show and there are other contexts that we may not be getting here. But at least, we can see that with the data with the exam scores, at least we have a sense of one piece of information that, it is working. So that we're going to have a recommendation which is the pilot Study seems to show a success and so we're going to recommend that we continue funding for another year.

And this recommendation is based solely on the data we provide, that we visualize, that we set up through our study. There maybe other pieces of

information that the decision maker has that may not be based on data. It might be based on some other factor that's outside of our control. And so they may not choose to actually fund for another year but you could be assured by providing this data they'll have another piece of information that they'll be able to use, in order to ensure student success in this case or whatever it is that you are interested in.

And so that is, **the key to doing data visualizations is to understand the context first**. And if that means stepping away from your computer and writing it out, then do it. Because **if you don't understand the context first, your data visualizations are not going to be as crisp, concise and clear as they need to be**. And we're going to go over this particular thing in much more detail in further courses. But at least you will have a sense of, okay, let me just step back, let me write it out, let me sketch it out, maybe outline it, and then get into the data. And so that is the key to this particular lesson and the key to this course in fact. So, we'll see you next time.
With **the simple formula of context in mind (Question + Data = the Story)**, outline a potential issue or problem you want to answer. What is the issue? What data do you already have that can help you visualize the current situation? What recommendations have you already been thinking about and what other data do you need to help you reach a decision or to help you tell the story?

I wanted to take this time to thank you for joining me in this discussion of the fundamentals of visualization. We've discussed a great deal of content and I hope you've come away with a better understanding of the fundamentals and framework of the data visualization industry. We spoke early on about what is data visualization and why we do it? You installed the Tableau software and did your first data visualization. In the second module, you explored the software in more detail and I hope you now feel more comfortable in using the tools and are confident in your abilities to keep going. In the third module, Suk helped you make data connections and you now know how to prepare and import your into the software. In the final module, we talked about a framework for thinking through your data visualization projects, how to ask the right type of questions and why. While we covered a lot of material, we also only scratch of the surface of data visualization and I hope you'll continue your exploration of the topic. Whether you decide to continue on in this specialization and take the additional courses offered by me and my colleagues, or do your own independent research and development, I wish you the best of luck with your future in data visualization. Until next time.