

Data Visualization

Guide to Tableau Public

Pittsburgh Data Jam
Pittsburgh DataWorks, Pitt Science Outreach, and Pittsburgh
Penguins Foundation

Guide Version 1
Authors: Emily Mitchell, Jackie Choffo, Jennifer Lundahl,
Margaret Farrell, Zach Gazze

I. Introduction	3
A. General Overview	3
B. Websites for Data	3
C. Introduction to the Scientific Method	3
II. Opening a Tableau Public Account	6
A. Creating a New Account	6
B. Saving your Work	6
C. Accessing Saved Projects	6
III. Formatting Data in Excel for use in Tableau Public	7
A. General	7
B. Column Headings	7
C. Using One Data Point per Row	8
D. Formatting Dates	8
IV. Importing Data to Tableau Public	10
A. General	10
V. Creating Visualizations in Tableau Public	11
A. Types of Visualizations	11
B. Using Tableau Public's Features	18
C. Using Tableau Public's Dashboard Feature	19
VI. Sharing Your Work	20
A. Introduction to Presentations	20
VII. Online Resources	21
A. Tutorial Videos	21
B. Helpful Links	21
VIII. Step-by-Step Example	22
A. Asking a Question	22
B. Research and Background Information	22
C. Constructing a Hypothesis	22
D. Methods	23
E. More Methods: Obtaining and Organizing Data	23
F. Data and Analysis	28
G. Interpreting Data, Drawing Conclusions, and Communicating Results	33

I. Introduction

A. General Overview

Tableau Public is a free data visualization tool that allows users to import data and arrange it in a variety of ways. Data can be used to create displays, which can then be shared publicly.

This guide will help with using Tableau Public and the scientific method to answer data questions. There are many tutorial videos available (see section VII. Online Resources for links), but this guide will provide tips and information that we found while learning to use the programs.

We will use a step-by-step example (using Pittsburgh employment data downloaded from the U.S. Department of Labor Bureau of Labor Statistics) to show how the program can be used with the scientific method for Data Jam projects. This is found in Section VII.

B. Websites for Data

Many governmental departments and national foundations have links to statistics on their websites, which offer data that can be downloaded for free. This data can be used for practice. The specific files we use in our example can also be downloaded - links are provided in Section VII.

National Science Foundation - <http://www.nsf.gov/>

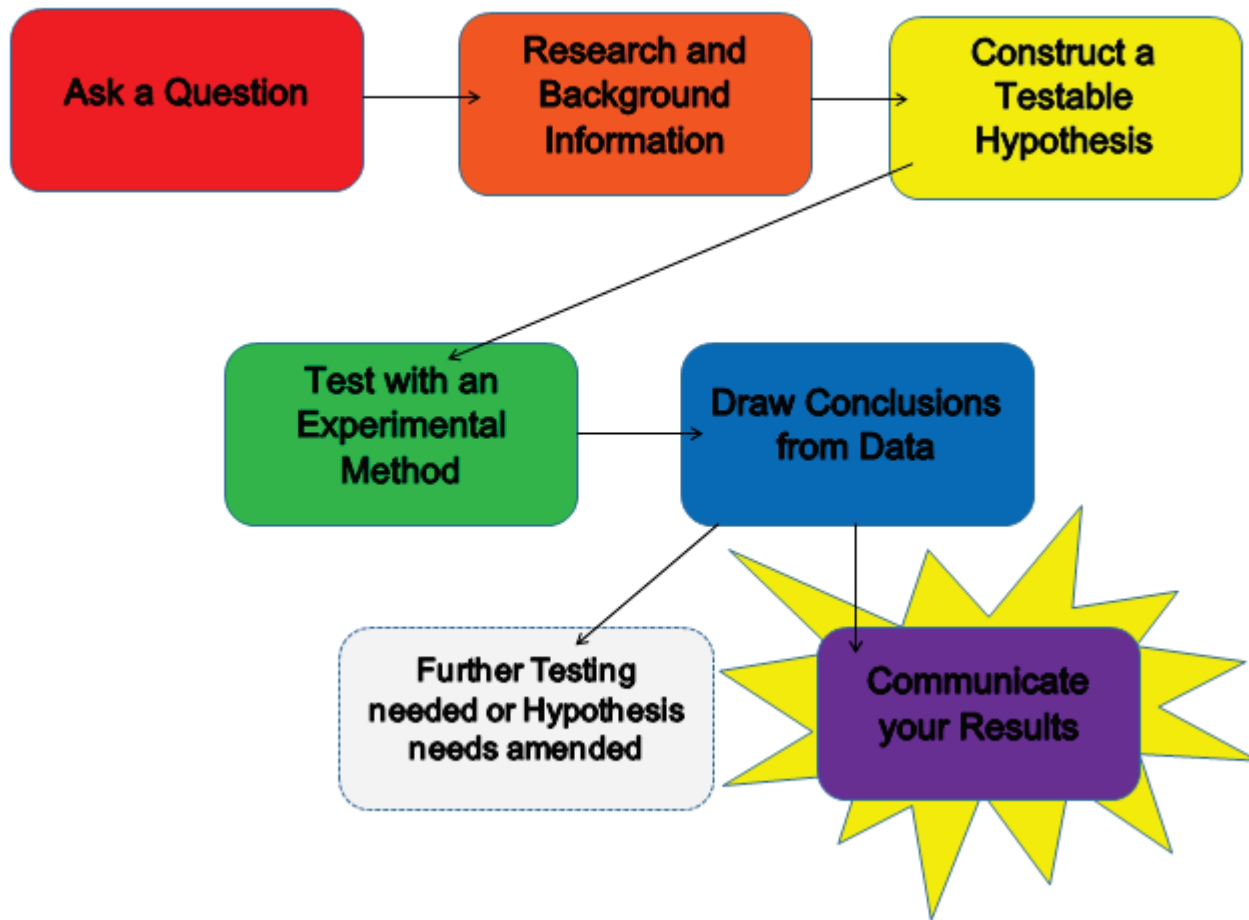
United States Census Bureau - <http://www.census.gov/en.html>

Bureau of Labor Statistics - <http://www.bls.gov/>

C. Introduction to the Scientific Method

Many students encounter the scientific method in early science classes. How can this process be applied to a data study? In the Step-by-Step Example section, we will follow the scientific method to answer a simple data question. The files used are available for downloading as practice, but we encourage you to try working with your own data using our example as a guide.

Fig. I.1 Scientific Method Flowchart



Scientific Method Steps Explained

Step 1: Ask a Question - We take in observations around us everyday, even if we do not realize we are doing this consciously. All of the observations we take in lead us to ask questions as to why and how the world we live in operates. The first step in the scientific method is to postulate why you observe the things you do. Your question should be clear and specific to a single topic of interest. You do not have to have an idea of how to answer your question just yet, only ask a question that you find puzzling and intriguing!

Step 2: Research and Background Information - After you form a question that you are interested in investigating, it is now time to gather as much information you can on the subject at hand. Read up on any and all information you can get your hands on, especially past experiments to see how researchers have gone about investigating your question before. Even if you come across information that you do not think pertains exactly to your question, take it into account because you never know how it could help you along the way of your experimentation. Also during this step, look into the different variables that you will be accounting for in your experiment, in other words, the causal factors you will be investigating.

Step 3: Construct a Testable Hypothesis - This is one of the most fundamental steps of the scientific method. The hypothesis is what is going to lead your experimentation and guide your research. Think of it as analogous to the thesis of your English paper, the compass that provides you with direction. There are many ways to craft the actual statement, using an “if...then” statement is a popular choice. No matter what way you word your hypothesis, you have to word it in a way that it can be proven right or wrong. It has to be testable, so include both the variable you are manipulating (independent variable) and the response you expect to see as an outcome of this manipulation (dependent variable). To see if you have a testable hypothesis, you should be able to write a null hypothesis, or what would be the opposite of what you expect to happen.

Step 4: Testing and Methods - Here is where the bulk of the written work takes place and the majority of the time is spent on experimentation. You will have to come up with a plan on how to investigate your hypothesis. The experiment does not have to take place in the traditional setting of a laboratory; many social science researchers collect data from controlled experiments they create to be conducted in public. Sports Medicine researchers go onto the athletic fields and training facilities to test athletes. Don't limit yourselves; the possibilities are virtually endless. This step also includes gathering data, so make sure you get yourself organized and take detailed notes so you can understand what data you collected when you go back to analyze your results.

Step 5: Draw Conclusions - Now that you have collected your raw data, you need to make sense of what the numbers and observations mean. There are many ways to analyze your findings and seeing if there are positive correlations between the two variables in question. For instance, if you need to compare your results to those of another researcher or the true values, you would want to calculate the percent error in your calculation and the standard deviation. One visually appealing way to interpret your data is to create charts and graphs. This way you can see the correlation between data points and also use your created charts and figures when you go to present your findings.

Step 6: Communicate Results - In the world of science, there is a broad scientific community. It is unique in the fact that scientific discoveries build off of each other and can only occur when there is a sharing of thoughts, findings, and knowledge. Another document will go into much further detail on how to create and give a scientific presentation, but remember to be willing to contribute your findings to the scientific community. Also, in creating your presentation, keep in mind who your audience is and characteristics about them. Will you be writing a research paper only to be seen by your teacher or are you presenting your research to an audience of 10,000 at a national conference? Knowing your audience will help you tailor the type of presentation you will create and also how you will present your information.

***Further Testing needed or Hypothesis needs amended** - Many times, the first trial in an experiment does not go exactly as planned. The data may be inconclusive, not enough data was collected, or you figured out along the way that there was a better method to conduct your experiment. It is always beneficial to critically analyze your experiment as a whole and its parts after it is complete. You may find that you have changed your hypothesis from the start of the experiment and it needs amended. This is fine! Science is not about being right all of the time on the first try. Where would be the fun in that if we already knew the answers to all of our question without investigation? Be open to tweaking your experiment to make sure that your conclusions are logical and are backed up with repeatable, consistent results.

II. Opening a Tableau Public Account

A. Creating a New Account

The free desktop product you will be using saves everything on the Tableau Public web servers. In other words, nothing can be saved locally to your personal computer or USB device (think of it as a private cloud just for Tableau users). When you go to save your first Viz, you will be prompted to sign into your Tableau account. Simply enter the required information with your primary email address and create a password and click the “create account” button located at the bottom of the pop-up window. Now, whenever you begin a new Viz or want to access an old one, you can sign into your account either through the Tableau Public application or on the web at <http://www.tableau.com/public/> and click “log in” in the upper right hand corner.

B. Saving your Work

When you are finished editing your data for the time being, make sure to save your Viz before exiting the program. At the top of the program there is a toolbar with the usual File menu. Click on the first heading “File” and select “Save to Web as” from the drop down menu. Again, you will be prompted to sign into your Tableau account if you are not logged in already. After signing in, type the name of your project in the designated area and click “Save”. This file will be saved as a thin AJAX based JavaScript application. Your data on Tableau Public is now accessible on the world-wide web and is downloadable by anyone. We strongly suggest you only publish data you are willing to share with anyone.

C. Accessing Saved Projects

From the Tableau Public application, you can access your saved projects by simply launching the program and clicking on “File” from the menu bar at the top of the screen on the start page. Select the title of the file you want to continue working on and it should open up. You can also see all of your published projects on Tableau’s website after you login. On your account screen, you will be able to see all of the projects you have created and saved to your Tableau account. From here you can open up a project or find all of the information you need to publish the Viz to another online source.

III. Formatting Data in Excel for Use in Tableau Public

A. General

Use of this section depends on the format of the data you obtain or are given. If the Excel file is already formatted for a data program, you will not need these steps.

Most Excel files downloaded from the Internet will be formatted for viewing as Excel files - not for use in data visualization software. They often have titles, notes, and other information that must be deleted before importing to Tableau Public.

They also may use formatting that makes information clear to a reader, but will not make sense when connecting to Tableau Public (see **Fig. III.1**). The data must be reorganized into individual points. This can be tedious, but is essential for using the data in the program.

If you are comfortable with Excel, you may only need to look over this section. It is written to make details clear to someone formatting a data set for the first time (like the Science Mentor writing this section).

B. Column Headings

This may sound obvious, but will become very important in the next section. The heading of each column must describe exactly what is in the column, and only what is in that column. The column headings should occupy Row 1.

	A	B	C	D	E	F	G	H	I	
1	TABLE 6. Science and engineering research space in academic institutions, by state, control, institution, and field: FY 2011									
2	(Net assignable square feet in thousands)									
3			Agricultural	Biological	Computer		Health and			
4			and natural	and	and		clinical	Mathematics	Physical	
5			resources	biomedical	information		sciences	and statistics	sciences	
6	State, control, and institution	All fields	sciences	sciences	sciences	Engineering				
8	Alabama									
10	Public									
11	AL A&M U.	104	70	7	1	20	0	1	5	
12	AL State U.	147	0	70	2	0	29	35	11	
13	Auburn U. main campus	443	157	28	0	150	60	*	44	
14	U. AL Birmingham, The	953	0	317	1	28	571	2	18	
15	U. AL Huntsville, The	260	1	15	22	120	13	5	83	
16	U. AL Tuscaloosa, The	192	0	33	7	72	0	0	62	
17	U. South AL	185	1	99	2	16	37	*	6	
19	Private									
20	Tuskegee U.	309	96	28	13	153	19	0	0	
22	Alaska									
24	Public									
25	U. AK Fairbanks	336	130	89	2	31	0	0	80	
26	U. AK Southeast	11	*	6	0	0	0	0	4	
28	Arizona									
30	Public									
31	AZ State U.	848	3	276	2	272	39	*	153	
32	Northern AZ U.	171	30	75	*	8	3	1	41	

Fig. III.1 The data above reports square footage of research space in numerous fields at many institutions. It is arranged in headings; under “State, control, and institution” the data lists a State, the institution type (public/private) in subheading, and the institution in a second subcolumn. This will not translate into Tableau Public. This column must be separated into three columns - State, Institution, and Institution Type, as shown in **Fig. III.2**.

	A	B	C
1	State	Institution	Institution Type

Fig. III.2 Each column heading shows exactly and only what the column will show. The title and other information have been removed and the column headings now occupy Row 1.

C. Using One Data Point per Row

Many of the tutorial videos state this, but do not go into very much detail. This is related to the column headings - it basically means that every row must be explicitly clear.

Another problem with the data shown in **Fig. III.1** is that the horizontal headings describe the scientific field to which the values below apply, i.e. the headings do not describe exactly and only what is listed below them. Tableau Public needs to know, through an explicit heading, what the numbers mean. Also, the scientific field is another value, and Tableau Public needs it grouped under a heading. See **Fig. III.3** below.

	A	B	C	D	E
1	State	Institution	Institution Type	Field	Number of Research Facilities
2	Alabama	AL A&M U.	Public	All	104
3	Alabama	AL A&M U.	Public	Agricultural and Natu	70
4	Alabama	AL A&M U.	Public	Biological and Biomer	7
5	Alabama	AL A&M U.	Public	Computer and Inform	1
6	Alabama	AL A&M U.	Public	Engineering	20
7	Alabama	AL A&M U.	Public	Health and Clinical Sc	0
8	Alabama	AL A&M U.	Public	Mathematics and Sta	1
9	Alabama	AL A&M U.	Public	Physical Sciences	5
10	Alabama	AL A&M U.	Public	Psychology	0
11	Alabama	AL A&M U.	Public	Social Sciences	0
12	Alabama	AL A&M U.	Public	Other	0

Fig. III.3 Each heading now describes only the information in the column below it. This expands the data significantly - there is now a row for every data point (the square footage of research facilities). Values in the other columns will repeat because they describe more than one number, but because the numbers are unique, each data point is unique. The data begins in Row 2.

D. Formatting Dates

If the data contains a column for a day, month, year, etc, Tableau Public can use this information to visualize data over time. However, for Tableau Public to recognize the number as a date, it must be formatted as a date in Excel.

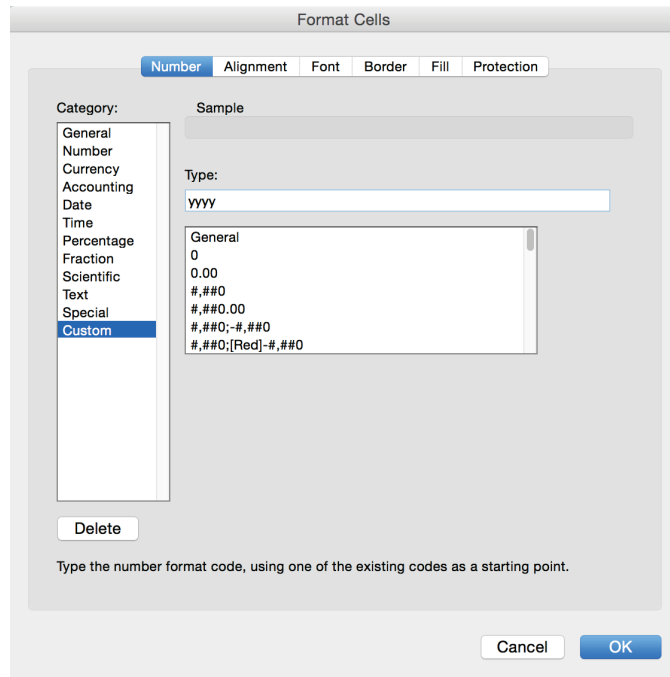


Fig. III.4 Format cells option window

To change the cell type, select the column, right click, and find **Format Cells**.

In the Category box, select **Date**, choose the format, and click OK.

Note: If the date is a year only, you must create a custom format. In the **Date** category, select the format closest to the one you want (d/m/yyyy will work). Then, select **Custom** in the category box, and in the **Type** box, type yyyy. Click OK.

Note: If the downloaded data has dates that are too difficult and tedious to reformat, it may be useful to create a list of dates from scratch. Use this link: <https://support.office.microsoft.com/en-us/article/Create-a-list-of-sequential-dates-aa1c0fa7-c76a-4762-8bc9-46f1691defec?CorrelationId=0c418051-9c3d-4b73-96c5-22426ebbb967&ui=en-US&rs=en-US&ad=US> Be careful not to lose any data or information if you choose to do this.

IV. Importing Data to Tableau Public

A. General

Most frequently the data that will be used for the Pittsburgh Data Jam will be organized in an Excel file, but Tableau Public can also import data from Microsoft Access and text files. When uploading a new data spreadsheet, simply pick the corresponding category the file fits into, whether it be Microsoft Access, Microsoft Excel, or a text file.

On the left side bar, Tableau Public will display the title **Workbook** and the name of the uploaded file. (See Fig. IV.2) It will also display the title **Sheets** and the names of the different sheets found in the uploaded file. Some data files may have several sheets containing different variations of the same data set. To work with one specific sheet, click and drag the name of the sheet to the label **Drag sheets here**.

If the data file has more than one sheet, the data can be joined. This means that the data from multiple sheets can be accessed simultaneously and used in one visualization. This is useful when there is additional related data stored in different sheets of the same workbook.

Note: Tableau Public's join function will only join by columns (the sheets will join if the second adds new columns of data to the first). They will not join if you are stacking rows of data (if the two sheets contain the same headings and different data values). Tableau's desktop product has a custom join feature that allows this.

After uploading and picking the data that is able to be accessed by the program, clicking on **Go to Worksheet** will allow visualizations to be created from this data.

V. Creating Visualizations

A. Types of Visualizations

Tableau Public has several different visualizations to choose from to display data. They are text tables, heat maps, highlight tables, symbol maps, filled maps, pie charts, bar graphs, treemaps, circle views, side-by-side circles, lines, dual lines, area charts, dual combinations, scatter plots, histograms, box and whisker plots, Gantt, bullet graphs and packed bubbles. Each of these relates different numbers of dimensions and measures to display data relationships.

1. Text Tables

Description: Text Tables are basically Tableau Public's version of the Excel spreadsheet.

Data Type: To use a text table, choose at least one dimension and at least one measure.

Pros: The advantage of using the text table is being able to pick what measures and dimensions to display and being able to change their position within the table. This allows emphasis to be put on certain relationships within the data set.

Cons: The text table doesn't allow for much visualization for the audience. It would be a nice supplement to a graph or other visualization, but isn't great for audience visualization on its own.

Year of Date	# Employed	Unemployed	#
2004	13,588,617	818,271	
2005	13,650,839	754,664	
2006	13,762,313	675,789	
2007	13,838,965	627,540	
2008	13,985,209	755,616	
2009	13,632,898	1,062,264	
2010	13,524,040	1,149,120	
2011	13,693,418	1,082,620	
2012	13,936,018	1,083,517	
2013	13,963,449	1,011,455	
2014	12,928,609	737,374	












2. Heat Maps

Description: Heat Maps utilize the size and color marks to show the differences between data points.

Data Type: To use a heat map, choose 1 or more dimensions and 1 or 2 measures.

Pros: The heat map allows data points to be associated with different colors and sizes based on their value, which can facilitate a better visual understanding of the data for an audience.

Cons: The number of measures that a heat map can display is very limited, so more complex relationships in the data should not be displayed using a heat map.

Year of Date	
2004	
2005	
2006	
2007	
2008	
2009	
2010	
2011	
2012	
2013	
2014	

3. Highlight Tables

Description: Highlight Tables are similar to text tables, but include the use of the color mark to provide information about the data displayed in the table.

Data Type: To use a highlight table, choose 1 or more dimensions and 1 measure.

Pros: The highlight table adds a visual component to the text table, allowing for the structure of a text table, but incorporating a visual element relating to the data.

Cons: Because of the added feature of incorporating color into the table, only one measure can be used in this table limiting options to display multiple data connections and relations.

Date										
2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
0.6820	0.6290	0.5620	0.5220	0.6160	0.8690	0.9400	0.8780	0.8660	0.8100	0.5940

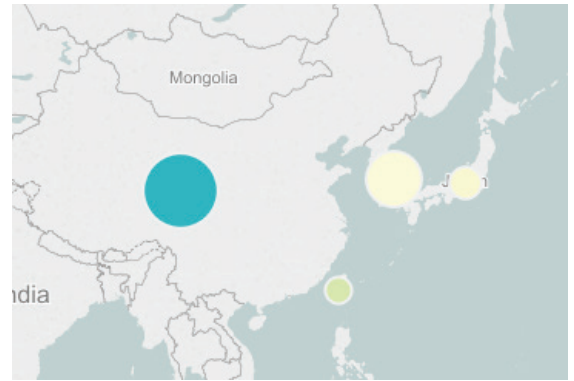
4. Symbol Maps

Description: Symbol Maps are geographic maps with symbols on each specified location that represents the data.

Data Type: To use a symbol map, choose 1 geo dimension and 0-2 measures.

Pros: The symbol maps allows the visualization of the data with respect to size and color, but also incorporate the relationship between the data and its origin.

Cons: Because of the nature of the size and color details, and the limitation of not being able to directly quantifying the data with axis, if a lot of the data is similar with some outliers, the data can be hard to interpret using this method.



5. Filled Maps

Description: Filled Maps are geographic maps with colored shading on each specified location that the data comes from.

Data Type: To use a symbol map, choose 1 geo dimension and 1 measure.

Pros: The filled map allows direct comparison of the data points and their relation to different regions.

Cons: The filled map only allows the use of one measure, limiting the ability to show relationships between multiple measures and location on the same map.



6. Pie Charts

Description: Pie Charts are circles that are sectioned off based on how much a dimension contributes to a measure.

Data Type: To use a symbol map, choose 1 dimension and 1-2 measures.

Pros: The pie chart allows the visualization of the contribution of the dimension to the data point as a whole, which allows the audience to understand where the whole data point comes from.

Cons: The pie chart only allows the data to be viewed as a portion of the whole data point, it gives no insight into the relationship between separate measures.



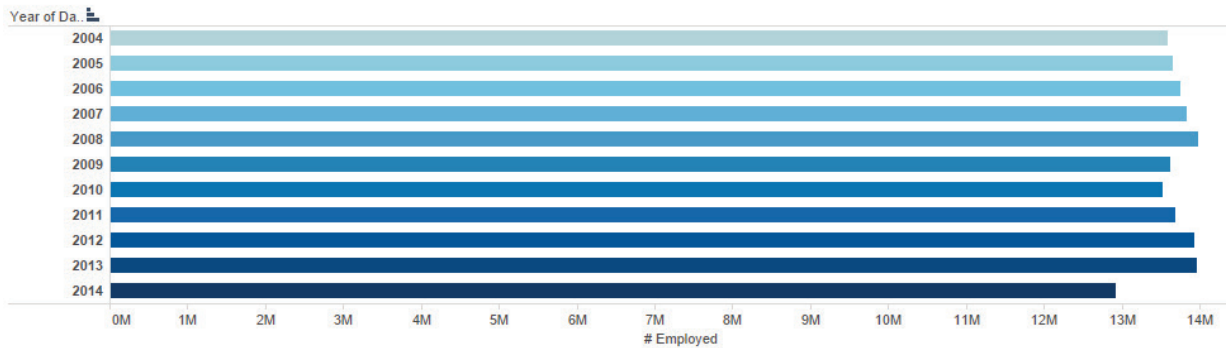
7. Horizontal Bars

Description: Horizontal Bars are like your typical bar graph with the bars going from left to right rather than up and down.

Data Type: To use a symbol map, choose 1 or more dimensions and 1 or more measures.

Pros: The horizontal bars allow the comparison of any number of measure and any number of dimensions, which allows complex relationships across multiple dimension and measures to be visualized.

Cons: After adding a lot of dimensions and measures to the horizontal bars, the graph can become very hard to interpret.



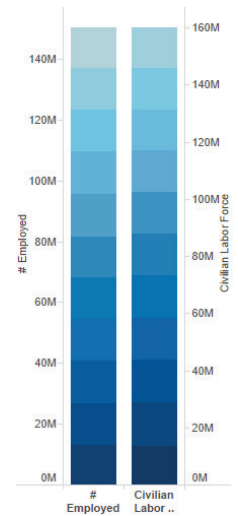
8. Stacked Bars

Description: Stacked Bars show the contribution of different dimensions to the sum of a measure through colored bars stacked directly on top of each other.

Data Type: To use a symbol map, choose 1 or more dimensions and 1 or more measures.

Pros: The stacked bars allow the visualization of the contributions of each separate dimension to the measure as a whole, while still being able to related different measures to each other and also comparing the contributions of different dimensions to each measure.

Cons: Using too many dimensions and/or measures at a time in this visualization can make the relationships and data very hard to understand.



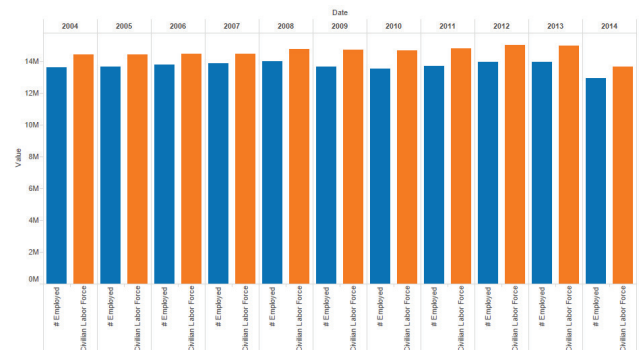
9. Side-By-Side Bars

Description: Side-By-Side Bars are multiple bar graphs grouped by a dimension.

Data Type: To use a symbol map, choose 1 or more dimensions and 1 or more measures.

Pros: The side-by-side bars allow conclusions to be drawn from comparing data that has been split up within the dimension.

Cons: With too many measures or dimensions, conclusions may be hard to be see from the visualization due to the related bars being so far from each other.



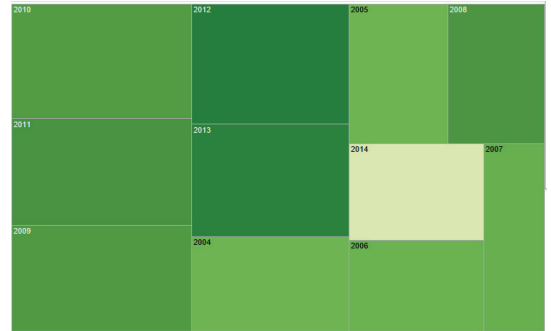
10. Treemaps

Description: Treemaps use color and size of boxes to show the same kind of relationships as the symbols maps, but without using any geographical data.

Data Type: To use a text table, choose at least one dimension and 1 or 2 measures.

Pros: Treemaps allow two measures to be compared through the use of different color and size boxes to emphasize different relationships between the two measures.

Cons: The treemap is limited to only relating two measures.



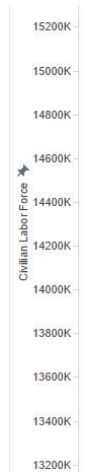
11. Circle Views

Description: Circle Views use circles in a set up similar to a stacked bar graph, but the circle lies at the top of where the bar would be.

Data Type: To use a text table, choose 1 or more dimensions and 1 or more measures.

Pros: Circle views allow the data to be compared, while telling where some point come close to being the same, shown through the overlapping of the circles.

Cons: The data points on a circle view aren't as explicitly defined as they are in other visualizations.



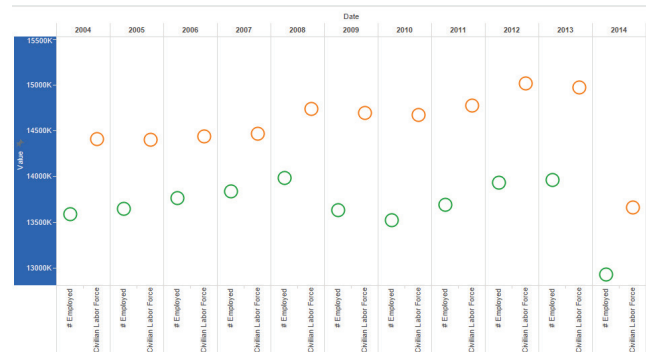
12. Side-by-side Circles

Description: Side-By-Side Circles are set up similar to the side-by-side bars, but have circles at the top of where the bars would be.

Data Type: To use a text table, choose 1 or more dimensions and 1 or more measures.

Pros: Side-By-Side Circles have the same advantages as the side-by-side bars, but also have the added feature of being more like points on a graph to put emphasis on the trends of the data.

Cons: The data points on the side-by-side circles aren't as explicitly defined as they are in other visualizations.



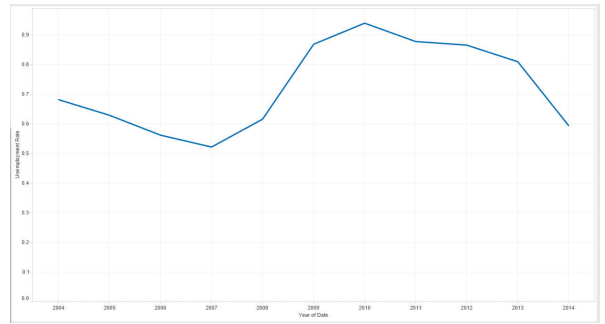
13. Lines

Description: Line graphs reflect the fluctuation of a value across progressing intervals.

Data Type: To use a line graph, choose 1 date, and 1 or more measures.

Pros: Line graphs are very beneficial in regards to the ease of understanding them, and their familiarity to many people.

Cons: In order to compare line graphs, it is necessary to have multiple plots on the same graph. After just a few of these, the chart becomes very crowded and difficult to decipher.



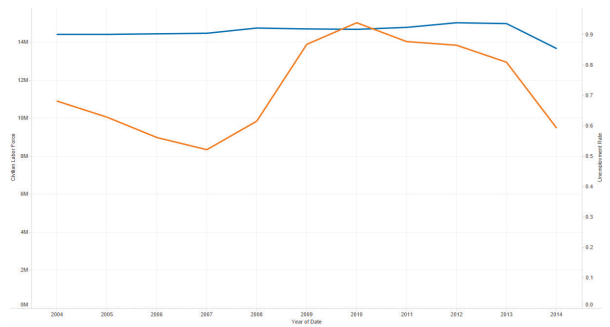
14. Dual Lines

Description: Dual lines is a version of line graph that emphasizes the comparison between two related sets of data.

Data Type: A Dual line graph requires 1 date, and 2 measures.

Pros: Effective way of comparing two sets of data.

Cons: There is a limit of two sets of data.



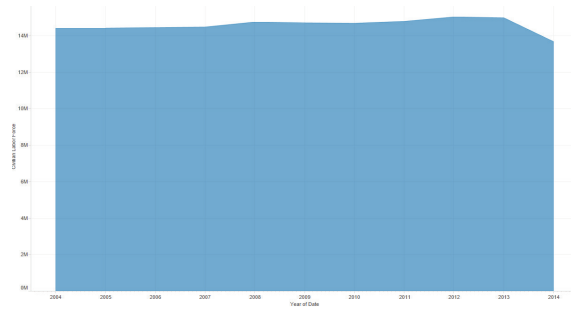
15. Area Charts

Description: Area charts combine a line and bar graph.

Data Type: An Area chart requires 1 date, and 1 or more measures.

Pros: This is an effective graph to use when an emphasis is wanted to be put upon the overlap, and gaps between, 2 line graphs.

Cons: After 2 measures are implemented, the chart becomes very complicated and difficult to read.



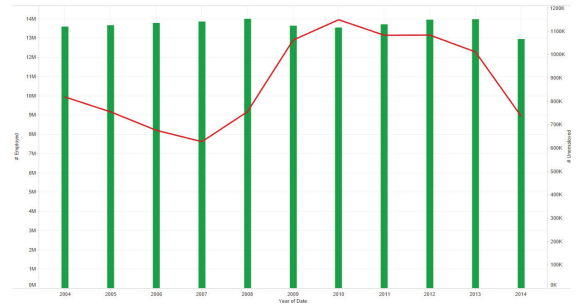
16. Dual Combinations

Description: Dual Combinations is another way of showing a correlation between 2 separate groups of data over time.

Data Type: A Dual Combinations chart requires 1 date, and 2 measures.

Pros: There is a distinction between the two sets of data, therefore it is easier to keep track of which is which.

Cons: The chart is limited to only 2 measures, and time must be used as the interval along the x-axis.



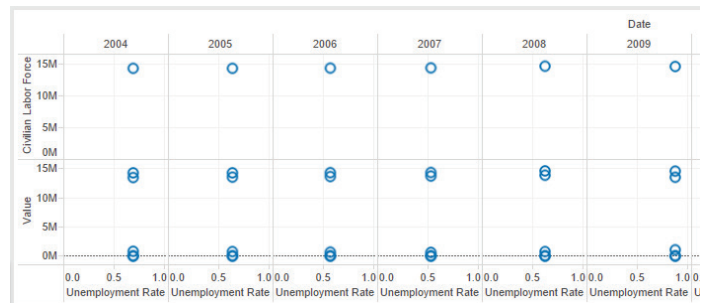
17. Scatter Plots

Description: Scatter plots are similar to line graphs in that the same data points are plotted, but there is no line connecting them.

Data Type: 2 to 4 measures.

Pros: Displays only raw data, which is a strong component to scientific presentations.

Cons: No analysis or interpretation of the data is provided.



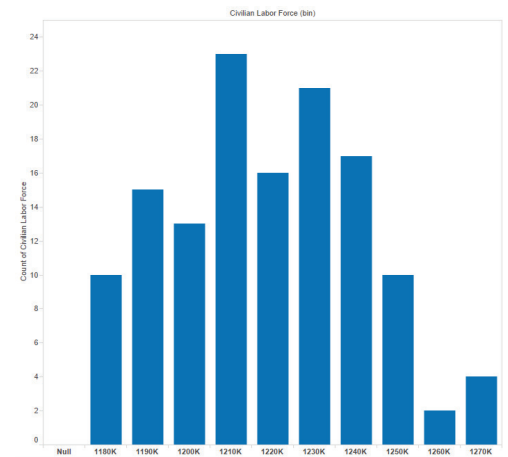
18. Histograms

Description:

Data Type: Histograms require 1 (appropriate) measure

Pros: Histograms create a unique bell-curve that, in itself, communicates a specific correlation between data.

Cons: Histograms can only be used for a select type of data.



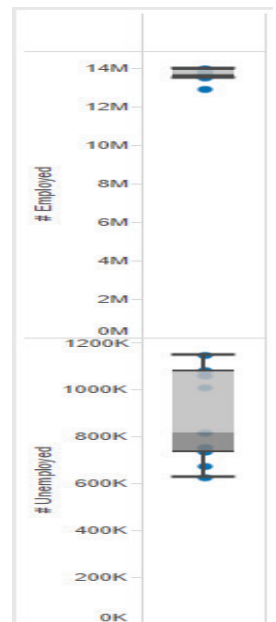
19. Box and Whisker Plots

Description: Box and Whisker Plots show a variety of statistical values such as the maximum, minimum, average, median, etc..

Data Type: 1 or more measures

Pros: Provides a variety of statistical measures.

Cons: If the data is not spread out, the plots can become very condensed and illegible.



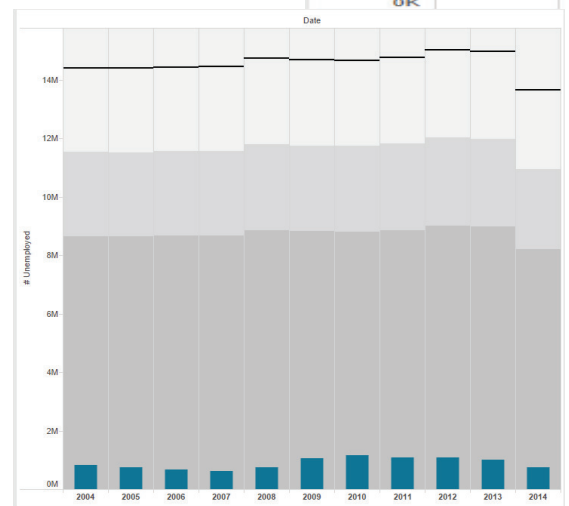
20. Bullet Graphs

Description: The bullet graph features a single, primary measure (for example, current year-to-date revenue), compares that measure to one or more other measures to enrich its meaning (for example, compared to a target), and displays it in the context of qualitative ranges of performance, such as poor, satisfactory, and good.

Data Type: Bullet graphs require 2 measures

Pros: Allows a comparison of 2 measures

Cons: Difficult to understand without prior experience or explanation



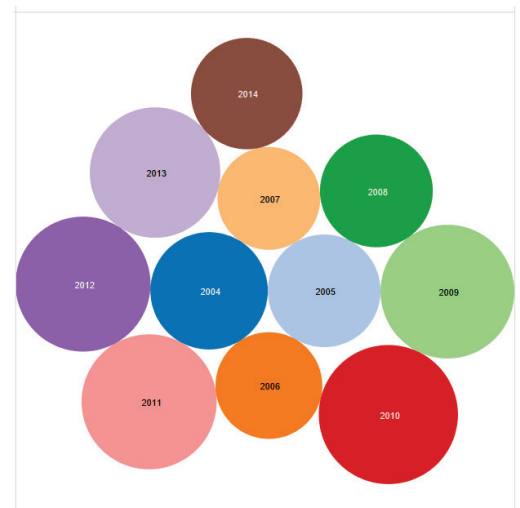
21. Packed Bubbles

Description: Scanning the mouse tracker over a particular bubble shows a text box containing all of the entered correlating data for that particular year.

Data Type: 1 or more dimensions, 1 or 2 measures

Pros: Very clean, appealing to the eye, and easy to understand.

Cons: Not the best chart to demonstrate comparisons between data - more like raw data.



B. Using Tableau Public's Features

Once you have imported your data, Tableau Public uses a drag and drop format to arrange it. The four main features you will use are described below.

Dimensions and Measures

Tableau Public automatically categorizes your column headings from the data you imported into dimensions or measures. Along the left hand side of your sheet you will see the titles in two separate boxes: **dimensions** and **measures**. **Dimensions** are categorical, they include non-numerical data, for instance: vehicle type, months, and region. If your data set included geographic locations, such as zip codes or boroughs, Tableau will recognize this as a geolocation and sort it into the dimension category as well. The other type of data would be **measures**. This includes all of your numerical data headings, anything that includes a number in essence. Unemployment rate, percentages, total number of people in the work force, and the like would be included under this section. If for some reason you believe that Tableau sorted your data incorrectly, you can right click on the title and select "**Convert Dimension**" and place it into the appropriate category as long as Tableau Public agrees with you.

Show Me

Tableau Public has a unique feature in which the program will suggest appropriate visualization for the type of data you want to display. First, you will need to decide what dimensions and/or measures you would like to display. You need to select a minimum of two column headings to begin. Simply click and drag the data title from the left hand side of the screen to the space after "**Columns**" that you wish to be displayed on the x-axis of your visualization. Then click and drag the data title from the left hand side of the screen to the space after "**Rows**" that you wish to be displayed on the y-axis of your visualization. For both columns and headings, you may click and drag more than one data title into the space to include more data in your visualization. After you have selected all of the data you want to include, click on the button "**Show Me**" at the top right hand corner of the page if it does not automatically pop up. The types of visualization that can be generated from the number and type of data you have selected to include will be highlighted. Select the appropriate visualization after referring the above section (**A. Types of Visualizations**). From here your visualization will be displayed in the middle of your screen and you are able to make the necessary changes and modification to it to show exactly what you want it to.

Marks: Color, Size, etc.

The **Marks** box at the left of the canvas contains several features.

- **Color** When you drag a **Dimension** to the **Color** box, you can separate your data by color. This is helpful when you have two categories for the same type of data (e.g. in the Step-by-Step example, we want to compare national and Pittsburgh employment data, so we use **Color** to separate one color for Pittsburgh and one for national.)
- **Size** does something similar. You can drag a **Measure** to the **Size** box and the marks on your graph will reflect that measure (for example, if we wanted to use a map to show public school


enrollment, you could make the marks reflect the amount of enrollment - they would appear larger on the areas where the schools with higher enrollment are located).

Filters

A **Filter** can be created in two ways: drag a dimension to the **Filter** box above the **Marks** box, or (this is the simpler way) right click on a dimension and click **Show Quick Filter**. It will appear to the right of the canvas. You can customize the filter by clicking the inverted triangle in the upper right hand corner of the filter. You can adjust how you use the filter - **lists**, **sliders**, and **dropdowns** can be chosen as appropriate to your data. If you'd like to see this in more detail, we use a **Quick Filter** in the Step-by-Step Example.

C. The Dashboard Feature

After you have created several visualizations on multiple Sheets in Tableau Public, you can build these into a Dashboard - a canvas for arranging and displaying your charts and graphs together.

Click this icon  at the bottom of the screen. This will open a new Dashboard.

At the bottom of the leftmost column, you can adjust the overall dimensions of the Dashboard. The **Range** drop down menu has multiple automatic dimension settings.

To get started, look at the leftmost column and drag a Sheet name to the Dashboard to place its contents on the Dashboard.

The default setting for placing Sheets on a Dashboard is **Tiled** - Tableau Public will automatically adjust sizes to create a spread of your visualizations. To manually adjust sizes, select the Sheet in the leftmost column and check the box next to **Floating** at the bottom of the leftmost column. The visualization can now be moved and adjusted manually.

You can also select **Floating** under **New Objects**: and all new objects dragged into the Dashboard will be floating by default.

Quick Tips

Double click the title of a visualization to edit it.

Right click a point on a mark - from this menu you can show or hide **trend lines** and **mark labels**, or add **annotations** to marks or points.

Select a sheet and click the inverted triangle in the upper right-hand corner. From this menu you can add **captions**, **color legends**, and **filters**.

VI. Sharing Your Work

A. Introduction to Presentations

A Presentation Guide will be available and will go into more detail about creating presentations. Here, we will just note some important things to consider when using Tableau Public.

Questions to Consider

- What kind of visualization is appropriate for the kind of data that you have?
- Which of those is the clearest visualization to use to present this data to an audience?
- Can you explain each of your visualizations in detail?
- How can you build the rest of your presentation in order for your visualizations to be a useful focal point?

Note: Because Tableau Public does not allow users to save work locally, screenshots are the best method for saving visualizations for use in posters.

VII. Online Resources

A. Tutorial Videos

Check out the link <http://www.tableau.com/public/training> and scroll down a bit for loads of helpful Tableau tutorials, including:

1. [Introduction to Tableau Public](#)
 - a. explains how to open data, create visuals, and publish your work to the web
2. [Introduction to the Interface](#)
 - a. explains columns and rows (and what to put in them), filters, formatting, and pages
3. [The Viz Types - "Show Me"](#)
 - a. talks about the different visualization options/chart types to choose from
4. [Maps - Dimensions and Measures](#)
 - a. step-by-step instructions to building maps in Tableau
5. [How to Structure Your Data](#)
 - a. explains how to format your data in Excel so that it is Tableau friendly!
6. [Dashboard Design and Formatting](#)
 - a. tips for designing your dashboard and formatting sheets
7. [Filters and Actions](#)
 - a. explains how to filter and highlight information in your dashboard
8. [Dashboard Elements](#)
 - a. tips on adding images, websites, and text boxes to your dashboard
9. [Publishing and Embedding](#)
 - a. how to publish your work to the Tableau Public website and other sites

For extra help with formatting data taken from the web in Excel, check out this video!

Microsoft Excel 2013 Tutorial - 11 - Formatting Your Worksheet

<https://www.youtube.com/watch?v=7U1af-AAPhU>

Just keep in mind that...

- even if you are not using Excel 2013, the instructions in the tutorial should still work!
- each row should only have **one** column heading! Note that the dataset that the narrator copies and pastes is already unambiguous - each column heading describes exactly what is in that column. Total Carbs contains a number of carbs, Total Fat contains a measurement of fat, etc. Your final data set should look like that!

B. Helpful Links

To get some inspiration for your project, check out the Gallery on Tableau Public's website! (Also, if you want to see how others made their dashboards, you can download them and open them in Tableau!)

<http://www.tableau.com/public/gallery>

If you're in need of data sets to practice with, click these links to download some! There are tons to choose from!

<http://www.tableau.com/public/community/sample-data-sets>

<http://www.tableau.com/public/community/data-catalogue>

Finally, if Excel is giving you some trouble, Microsoft's website has lots of how-to's and tutorials!

<http://www.microsoft.com/mac/how-to/excel>

VIII. Step-by-Step Example

We will now use the Scientific Method described in the introduction and Tableau Public to answer a data question. This is the basic idea of what a team might do to answer a Data Jam question, but do not limit your questions and ideas to this specific template; creativity and critical thinking are essential in science!

A. Asking a Question

For our example, we will use the question, *“How does Pittsburgh employment data compare to national employment data from 2004 to 2014?”*

This is a straightforward and somewhat surface-level question, but addressing it would allow us to explore deeper questions like, “What factors are the most influential on the relationship between Pittsburgh and national employment trends?” or “How do these national and regional employment trends impact trends in other areas of life, such as housing, education, and income level?” These are things we might explore in a full Data Jam project.

B. Research and Background Information

Background information for our question deals with known information about Pittsburgh’s employment status. Here are some important details we will consider when we construct our hypothesis about Pittsburgh employment trends:

- Between 1995 and 2005 (overlapping with our period of study), Pittsburgh’s job growth was positive, but a lower rate than that of the national average.
 - Contrast with national trends was likely due to differences in population change, as the areas in which national job growth was concentrated were construction, retail, and government - population-driven sectors.
(<http://www.pittsburghfuture.com/economy.html>)
- Pittsburgh’s economy is highly dependent on healthcare, education, technology and financial services. (<http://www.forbes.com/places/pa/pittsburgh/>)
- In 2012, the national economy depended more on retail than on education and financial services combined. State and local government sectors provide the most jobs.
(http://www.bls.gov/emp/ep_table_201.htm)
- Pittsburgh’s retail trade increased by over 3% between December 2013 and December 2014.
(http://pittsburghtoday.org/view_economy_job_growth_view1.html)

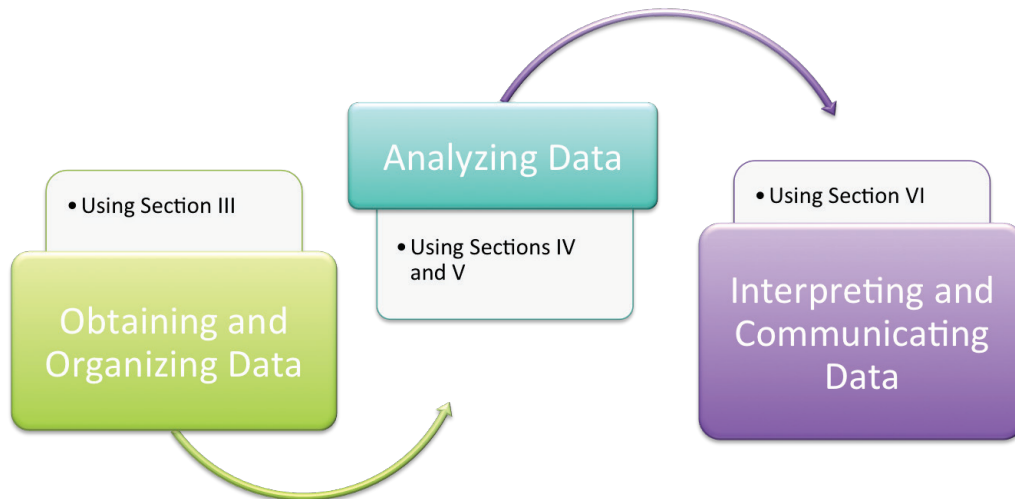
C. Constructing a Hypothesis

We will use the following hypothesis:

Because employment rates are heavily dependent on the retail sector, and Pittsburgh’s retail growth is lower than the national average at the beginning of our interval, the employment rate will begin at a lower growth rate than the national trends. Because the retail sector in Pittsburgh strengthens near the end of our interval, the employment rate will approach or surpass the national average.

D. Methods

In some scientific studies, the next step would be to construct experimental steps to test the hypothesis. This will be slightly different in our case, as the actual collection of the data has already been completed. Our methods will cover these main points:



E. More Methods: Obtaining and Organizing Data

Having completed the first steps of the scientific method, we will begin the first step of our methods: *Obtaining and Organizing Data*. In this section we use the steps from A-D to reformat data in Excel.

The datasets used in this walkthrough are available [here](#).

Note: This step-by-step example will show formatting for the four Pittsburgh data sets only. The same process is employed for organizing the national data. The result screenshot will display the final dataset, including the national data.

The Original Data Files

The Bureau of Labor Statistics chose to present different aspects of the data in separate Excel files - a screenshot of each is shown below. We have four files each for Pittsburgh and the national data (8 files total). To put all of this data into Tableau Public, we will need to consolidate the data into one dataset.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
2004	1193308	1186391	1192606	1188600	1194434	1215312	1225340	1209632	1194722	1198716	1204700	1203127	1200574
2005	1194584	1187881	1189406	1188446	1199000	1213579	1224274	1211837	1198869	1196885	1201199	1199543	1200459
2006	1188935	1189676	1193721	1190755	1199607	1219777	1224310	1214594	1202513	1204076	1205849	1204289	1203175
2007	1196640	1188260	1189569	1183276	1196870	1221283	1227254	1220130	1210163	1208704	1214097	1210259	1205542
2008	1211581	1200677	1207399	1211252	1226907	1243014	1252702	1246395	1234564	1238647	1235949	1231738	1228402
2009	1226369	1218583	1215936	1217930	1228254	1242660	1249534	1236159	1218287	1214550	1216542	1210358	1224597
2010	1213390	1206644	1205388	1211936	1220812	1236732	1246327	1241601	1226349	1219047	1223703	1221231	1222763
2011	1224569	1212806	1212872	1213930	1229360	1245373	1249240	1245512	1234378	1236764	1236475	1234759	1231336
2012	1232537	1231062	1239199	1239048	1253047	1270894	1272662	1261757	1253910	1259992	1253993	1251434	1251628
2013	1251938	1240788	1236418	1241116	1252862	1270647	1272320	1256114	1247157	1235258	1239958	1230328	1247909
2014	1226567	1234572	1233855	1236635	1242979	1253570	1261889	1249209	1241553	1243649	1241505		

Fig. III.5 Original data table for Civilian Labor Force

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
2004	1114706	1110492	1116785	1125324	1128695	1143192	1154040	1142165	1134318	1137374	1141179	1140347	1132385
2005	1121280	1113231	1118176	1129441	1138173	1148367	1159653	1150688	1142441	1142176	1143043	1144170	1137570
2006	1124360	1123967	1132251	1135951	1142569	1158797	1163294	1156638	1153819	1157709	1156864	1156094	1146859
2007	1136476	1130561	1137349	1137546	1148166	1167027	1171172	1165545	1162810	1161428	1164567	1156318	1153247
2008	1146558	1136626	1145517	1159421	1168623	1178435	1185552	1178824	1175345	1178336	1171173	1160799	1165434
2009	1139842	1129868	1128862	1139113	1142694	1150282	1154061	1141361	1131485	1126987	1128278	1120065	1136075
2010	1106494	1095167	1101358	1118868	1126576	1140122	1148326	1145805	1140679	1133848	1134472	1132325	1127003
2011	1124440	1115690	1120341	1130147	1141711	1150359	1153143	1150239	1149622	1153934	1153198	1150594	1141118
2012	1137779	1136145	1148496	1157660	1165088	1174275	1173545	1164434	1169753	1175659	1169983	1163201	1161335
2013	1148980	1146802	1148332	1161898	1169850	1179135	1182265	1168976	1170926	1160350	1166023	1159912	1163621
2014	1146147	1156152	1158669	1178468	1177041	1184031	1188234	1175708	1184849	1191240	1188070		

Fig. III.6 Original data table for number of the Civilian Labor Force that is employed

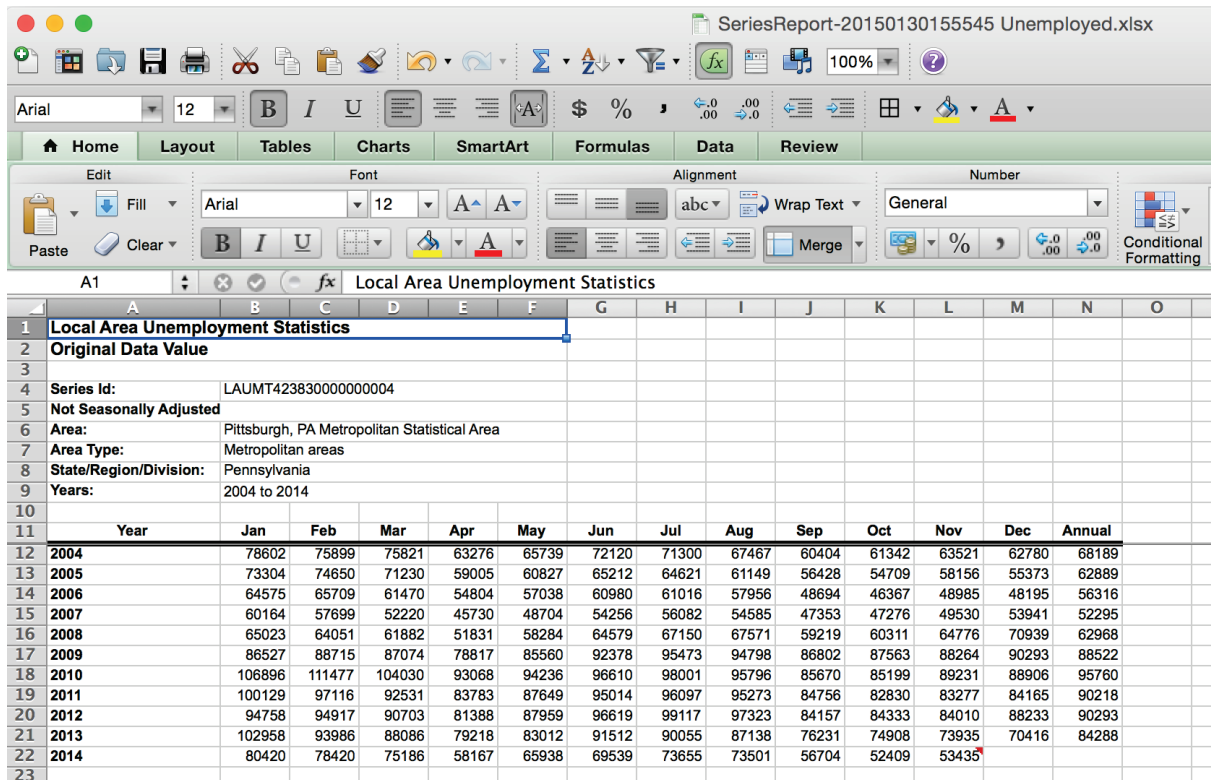


Fig. III.7 Original data table for the number of the Civilian Labor Force that is unemployed

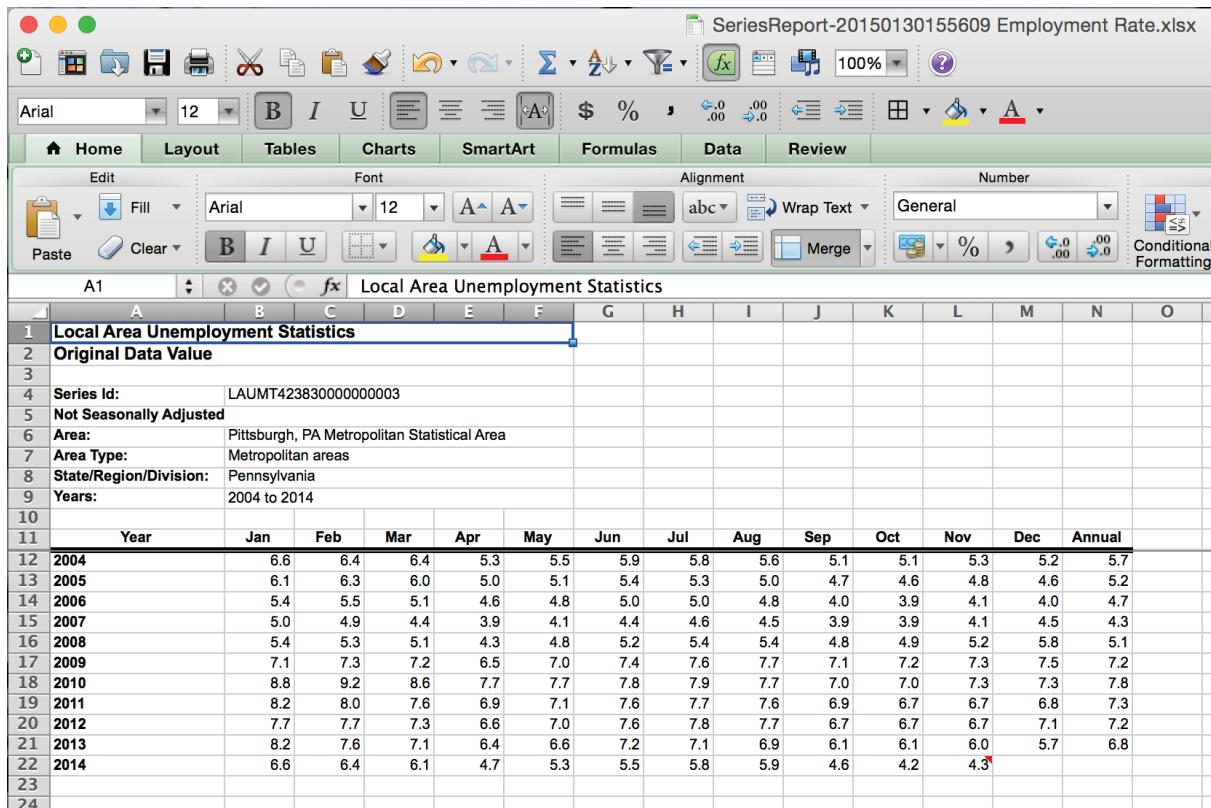


Fig. III.8 Original data table for the unemployment rate

Reformatting

1. The consolidated dataset will need column headings that cover all the information. In a new, blank Excel file, enter these column headings.

	A	B	C	D	E	F
1	Date	Civilian Labor Force	# Employed	# Unemployed	Unemployment Rate	Region

Fig. III.9 Column headings for the new dataset. “Date” will include both year and month. Region will specify whether that row’s data is from the Pittsburgh or national dataset.

2. Fill in dates to incorporate both the year and the month. Enter the date 01/01/04 in cell A2. Hover the mouse in the lower right-hand corner of the cell until the cursor changes to a black cross. Click and drag the mouse through row 132 – this will fill a date series. However, it has filled by daily increments. To change this to monthly, as our data requires, select the entire series of dates. Click **Edit** → **Fill** → **Series** and select **Month** under the category **Date Unit**. This fills the monthly dates for our entire Pittsburgh set of data.

3. For the purposes of this walkthrough, we will not use the “Annual” column. Delete this column in each data sheet.

4. **Copy** each **row** from the original data tables (this will select the monthly data in Civilian Labor Force, Employment, Unemployment, or Unemployment Rate for **one year at a time**). First select the Civilian Labor Force data for 2004. Now, go back to your consolidated data table, and right-click in the first cell of the Civilian Labor Force column. Select **Paste Special**, check the box for **Transpose**, and click **OK**. This pastes as a **column** what was previously a **row**.

5. Repeat this **copy/paste special** process for each year, so that the Civilian Labor Force column fills with the appropriate data. Then repeat this process for each year in Employment, Unemployment, and Employment rate.

6. Start a new date series under the end of the Pittsburgh data, following the same process outlined in Step 2. Then repeat the **copy/paste special** process for the national data, adding to your consolidated dataset.

7. We need to make a few more adjustments to the data. The original data tables report the unemployment rate as a number. We need to reformat this as a percentage. When Excel converts to a percentage, it multiplies the number in the cell by 100. To make the list compatible with this function, enter **100** in a blank cell, **copy** that cell, and then **select all the cells containing an unemployment rate**. Right click and select **Paste Special**. In the Paste Special window, under **Operations**, select **Divide**. Now, with the list still selected, right click and select **Format Cells**. In the **Number** category, select **Percentage**, and **reduce the number of places after the decimal** to 1.

8. Lastly, we need to correct the national numbers, which are expressed in thousands. Follow a similar procedure to Step 7 and enter **1,000** in a blank cell. **Copy** that cell, and select all the national numbers (excluding Unemployment Rate) and select **Paste Special** and **Multiply**. This will multiply all selected cells by 1,000.

The result should look like **Fig. III.10**

	A	B	C	D	E	F
114	05/01/13	155562000	143890000	11671000	7.5%	National
115	06/01/13	155761000	144025000	11736000	7.5%	National
116	07/01/13	155632000	144275000	11357000	7.3%	National
117	08/01/13	155529000	144288000	11241000	7.2%	National
118	09/01/13	155548000	144297000	11251000	7.2%	National
119	10/01/13	154615000	143453000	11161000	7.2%	National
120	11/01/13	155304000	144490000	10814000	7.0%	National
121	12/01/13	155047000	144671000	10376000	6.7%	National
122	01/01/14	155486000	145206000	10280000	6.6%	National
123	02/01/14	155688000	145301000	10387000	6.7%	National
124	03/01/14	156180000	145796000	10384000	6.6%	National
125	04/01/14	155420000	145724000	9696000	6.2%	National
126	05/01/14	155629000	145868000	9761000	6.3%	National
127	06/01/14	155700000	146247000	9453000	6.1%	National
128	07/01/14	156048000	146401000	9648000	6.2%	National
129	08/01/14	156018000	146451000	9568000	6.1%	National
130	09/01/14	155845000	146607000	9237000	5.9%	National
131	10/01/14	156243000	147260000	8983000	5.7%	National
132	11/01/14	156402000	147331000	9071000	5.8%	National
133	12/01/14	156129000	147442000	8688000	5.6%	National
134	1/1/04	1193308	1114706	78602	6.6%	Pittsburgh
135	2/1/04	1186391	1110492	75899	6.4%	Pittsburgh
136	3/1/04	1192606	1116785	75821	6.4%	Pittsburgh
137	4/1/04	1188600	1125324	63276	5.3%	Pittsburgh
138	5/1/04	1194434	1128695	65739	5.5%	Pittsburgh
139	6/1/04	1215312	1143192	72120	5.9%	Pittsburgh
140	7/1/04	1225340	1154040	71300	5.8%	Pittsburgh
141	8/1/04	1209632	1142165	67467	5.6%	Pittsburgh
142	9/1/04	1194722	1134318	60404	5.1%	Pittsburgh
143	10/1/04	1198716	1137374	61342	5.1%	Pittsburgh
144	11/1/04	1204700	1141179	63521	5.3%	Pittsburgh
145	12/1/04	1203127	1140347	62780	5.2%	Pittsburgh
146	1/1/05	1194584	1121280	73304	6.1%	Pittsburgh
147	2/1/05	1187881	1113231	74650	6.3%	Pittsburgh
148	3/1/05	1189406	1118176	71230	6.0%	Pittsburgh

Fig. III.10 A sample of the properly formatted dataset (our national data is listed first). Note the restart in date series at Row 134. This will not affect Tableau Public's reading of the date.

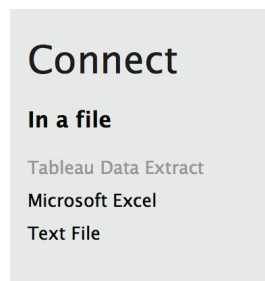
Save this new file, and the data is now ready to be imported into Tableau Public.

F. Data and Analysis

Import to Tableau Public



After opening Tableau Public, click the **Open Data** button shown in **Fig. IV.1**
Fig. IV.1 Open Data button.



Select **Microsoft Excel** from the left side **Connect** menu
Fig. IV.2 Connect to data from this menu.

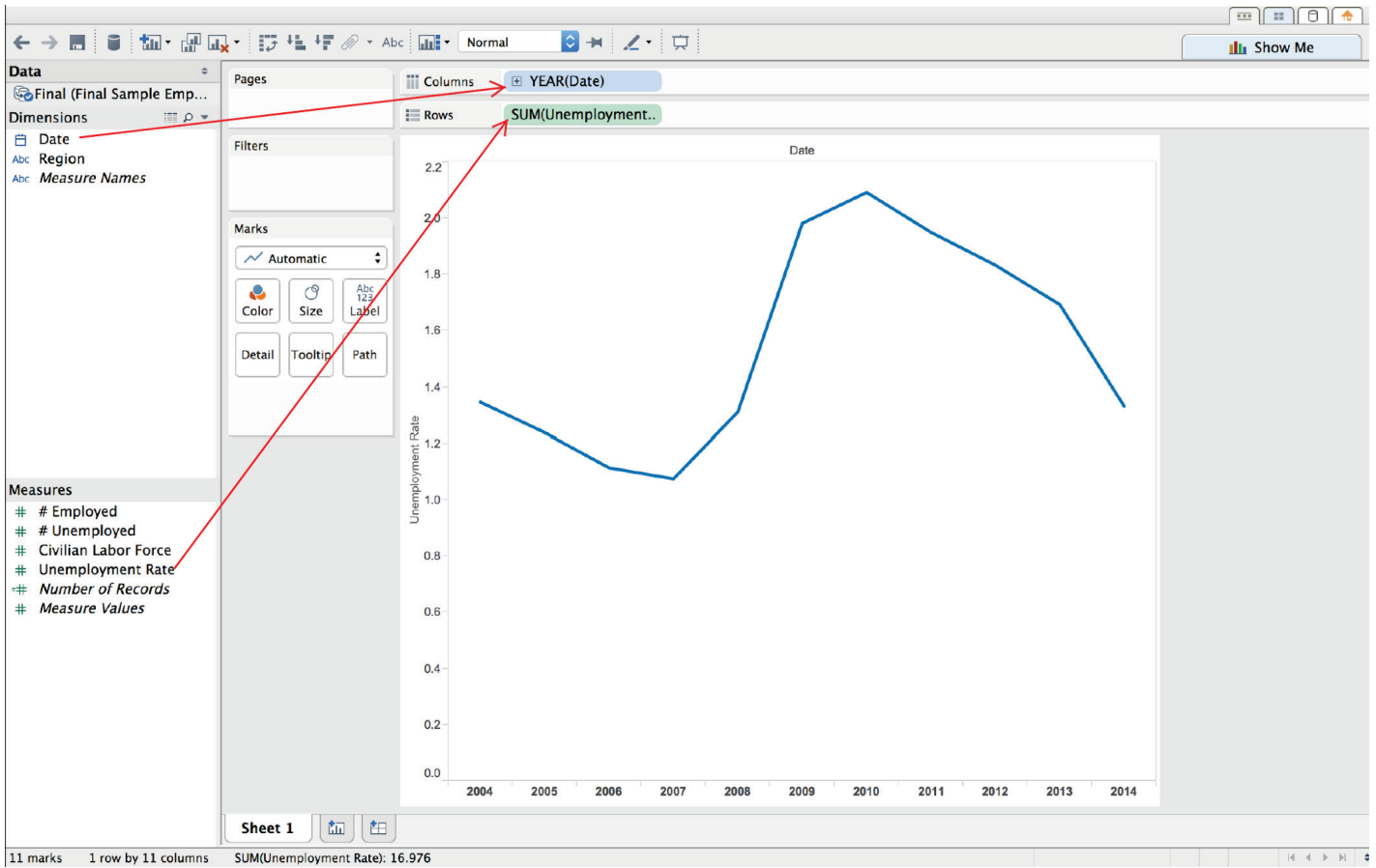
To start our visualizations, we'll use some of the questions from Section VI. to think about our data in Tableau Public.

What kind of visualization is appropriate for the kind of data that you have?

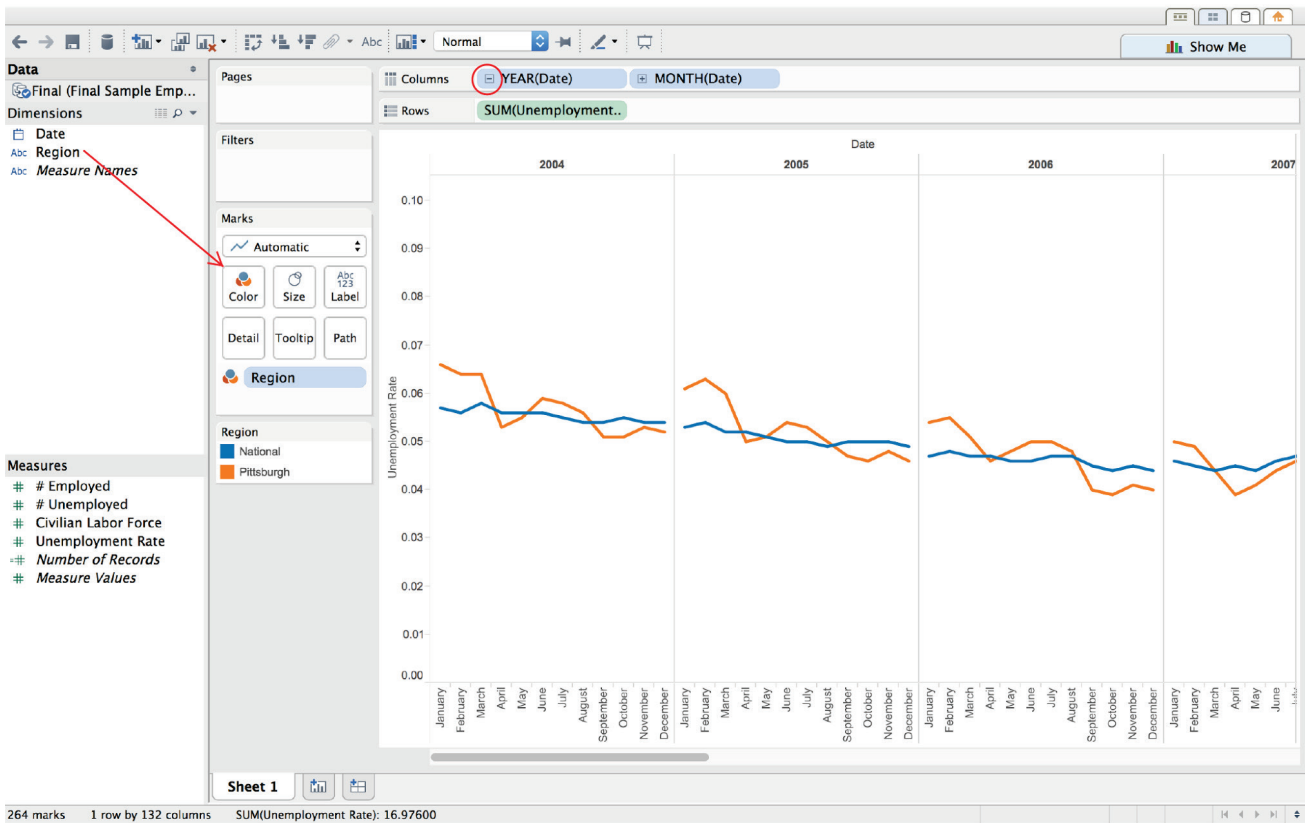
Because this data shows simple data over time, it is most appropriate to use a line graph. We can also use a bar graph to compare data points from specific years.

Which of those is the most clear visualization to use to present this data to an audience?

A line graph will be the most useful to an audience.

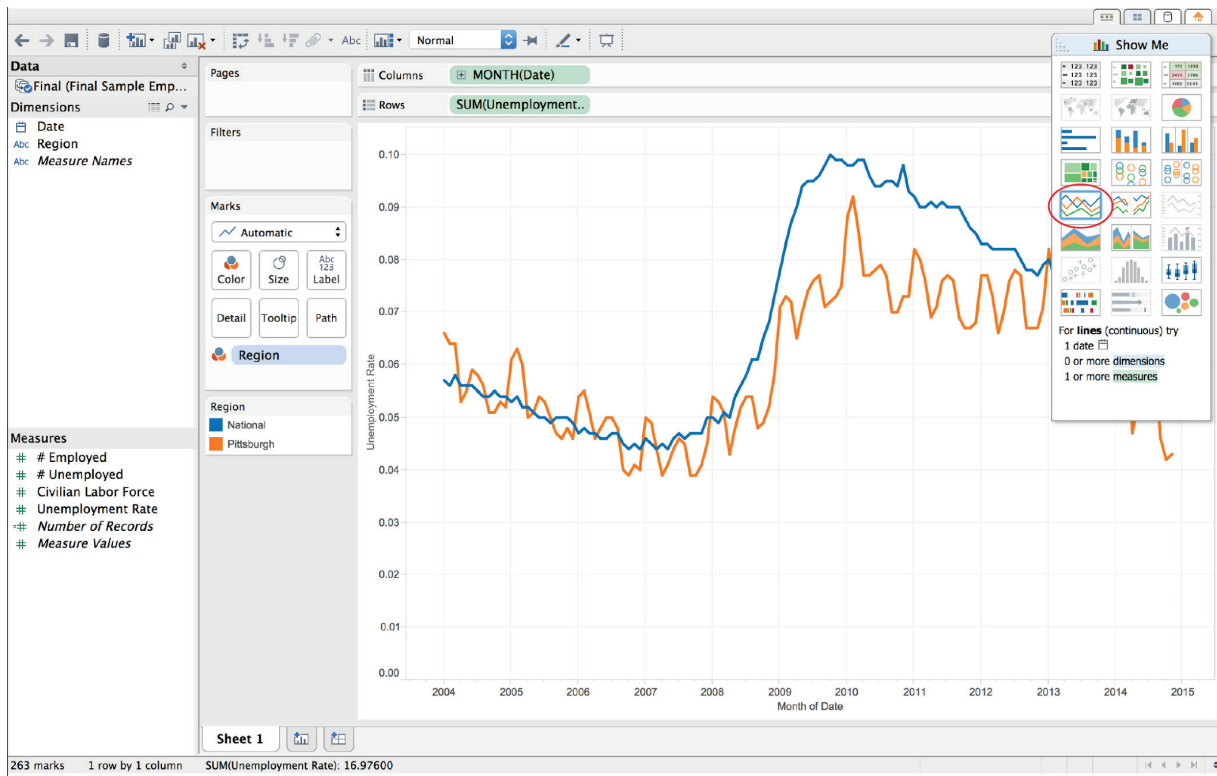


First, we'll compare the employment rates over the interval. We'll drag the **Dimension** "Date" to the **Columns** area and the **Measure** "Unemployment Rate" to the **Rows** area, as shown by the arrows above.



Next, click the small box with a + sign in it next to “Year.” Another **Dimension** will appear - a box for “Quarter.” Click the + in the box next to “Quarter” and a box for “Months” will appear. We’ll right click on “Quarter” and **Remove** it, so that the graph shows years broken up into months (without showing quarters).

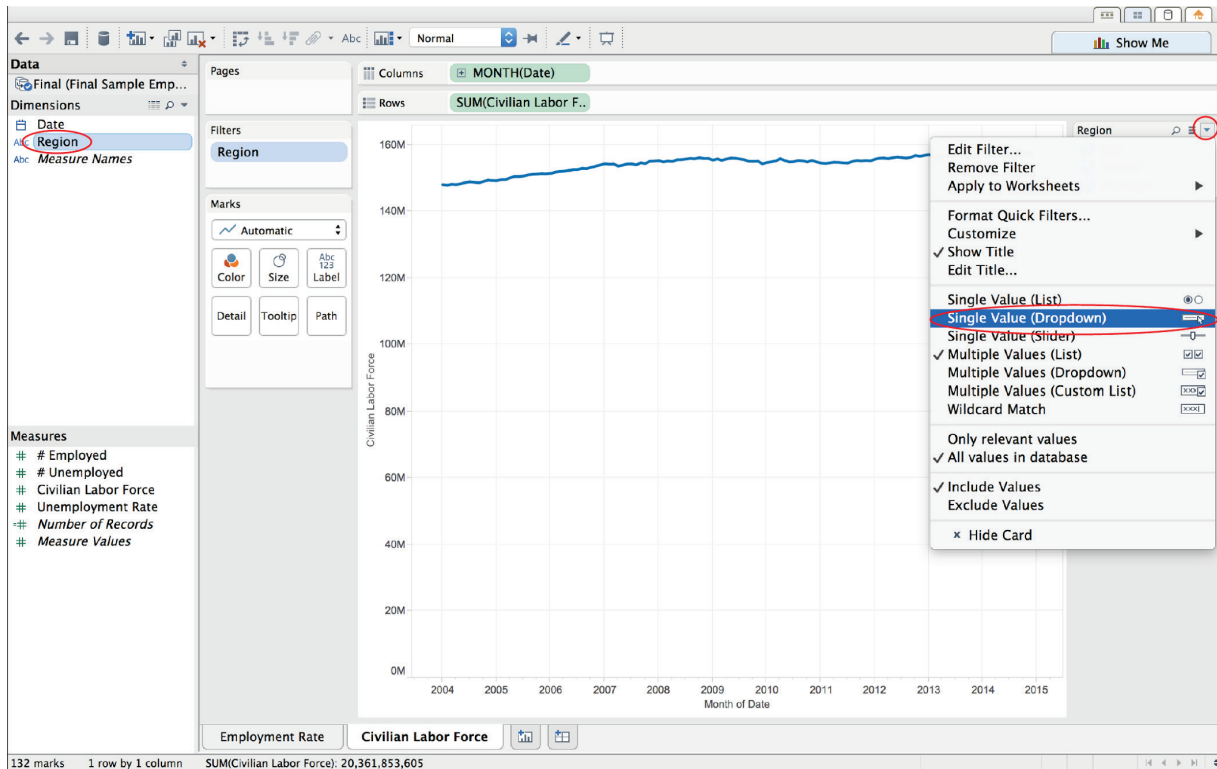
Also, the initial visualization shows Pittsburgh and national data together in one line. To separate them, we’ll use a **Filter**. Drag “Region” to the **Color** box as shown above.



Because Tableau Public used a discrete line graph as a default visualization, we'll click on **Show Me** and select the continuous line graph.

Next, we'll look at how the Civilian Labor Force has changed in both datasets. Add a new Sheet by clicking the bar graph icon at the bottom of the screen.

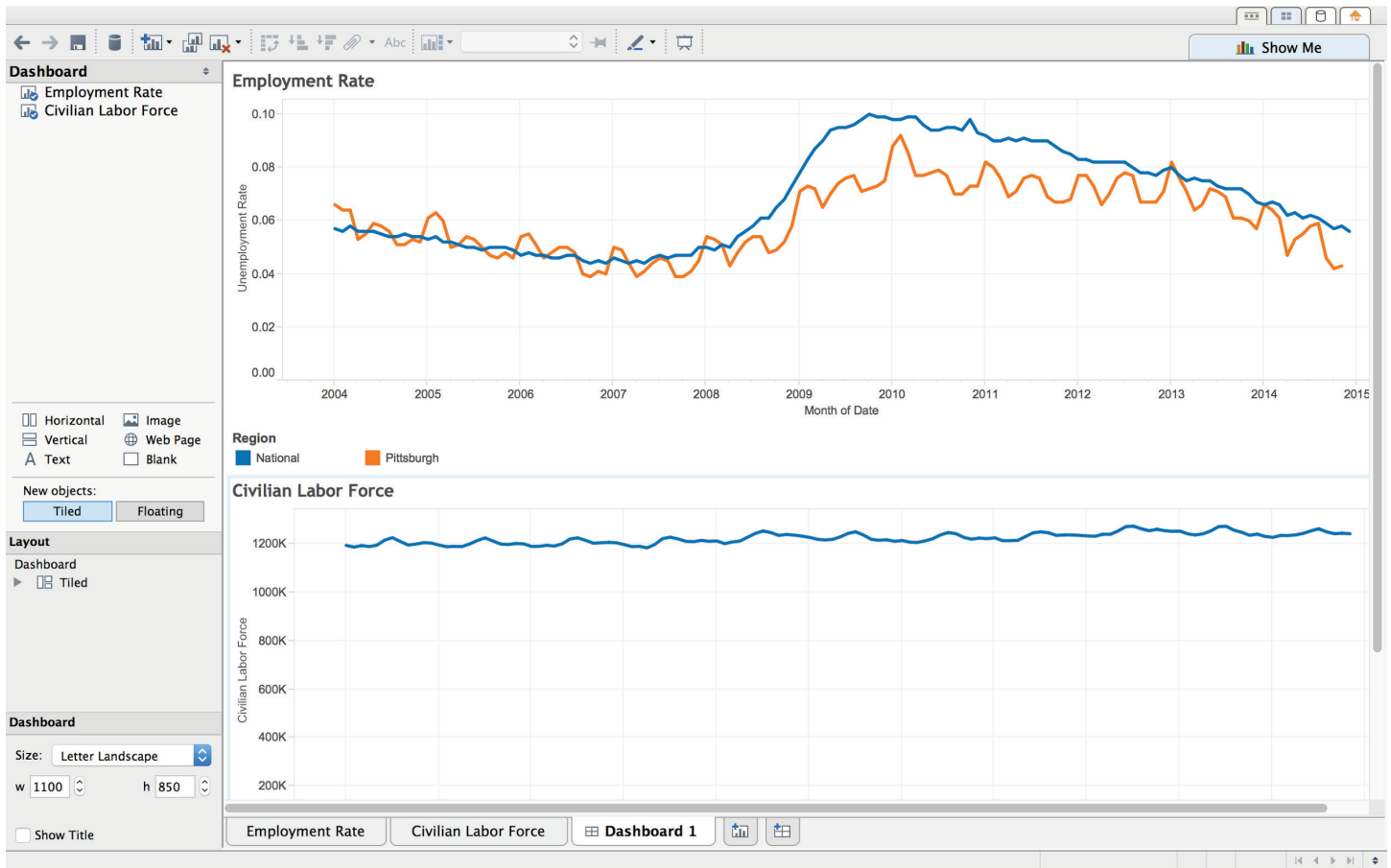
Drag "Date" to the **Columns** area as before, and "Civilian Labor Force" to the **Rows** area. Repeat the processes to change the date to "Year" and "Month" and make the line continuous.



This time, instead of using **Color** to show the region, we will use “Region” as a **Filter** (we are doing this so that the change in each can be compared - using **Color** to show the region places them on the same line graph and the difference in scale is so great that Pittsburgh change is difficult to see).

Right click on “Region” under **Dimensions** and click **Show Quick Filter**. Adjust it to show one region at a time by clicking the inverted triangle at the top of the **Region Quick Filter** box (as shown above) and selecting **Single Value (Dropdown)**. In the same menu, under **Customize**, deselect **Show “All” Value**.

Now we can show both graphs together by creating a **Dashboard**. Click the **New Dashboard** tab at the bottom of the screen.



All objects in this Dashboard are **Tiled**. This arrangement was obtained by dragging the objects and adjusting the size by hovering over the edges while it is selected.

G. Interpreting Data, Drawing Conclusions, and Communicating Results

We'll return to our questions:

Interpretation and Conclusions: Can you explain each of your visualizations in detail?

The section of the Dashboard titled "Employment Rate" shows the change in employment rate in Pittsburgh at the national level from 2004-2014. They follow a similar general trend, dipping between 2004 and 2008 and rising between 2009-2011, and then dropping again in 2012. The Pittsburgh line shows much greater variation, which is likely due to the difference in sample size.

We can understand this difference from the section titled "Civilian Labor Force." Switching back and forth between the Pittsburgh and national lines in the dropdown menu, we see the scale change on the graph's y-axis. This shows the difference in sample size and allows us to observe the similarities in trend between the two lines. We can conclude that the trends in Pittsburgh employment were analogous to the national trends. The next question will address where we go from here.

Communication: How can you build the rest of your presentation in order for your visualizations to be a useful focal point?

Since our hypothesis was based on limited research, it isn't really surprising that the actual data didn't follow the trend we predicted. However, this does not mean we haven't found anything useful.

Our presentation should focus on change over time. It should also provide background on different causes of employment fluctuation. It should provide context for the similar pattern between the two datasets. We'll ask things like, "Why did they both drop between 2004 and 2008?" and "What occurred during that time that might be responsible for the parallel events?" We'll want to communicate as much of the context as possible while staying relevant, focused, and accurate.