Introduction to Data Visualization Workshop

Download the zip file from http://uoft.me/dataviz and save it in a folder on the desktop. Make sure to extract the files from the zip file.

If you are unsure how to do this, you will find detailed instructions in the following *Set Up* section.

If you are familiar with file management and working with zip files, skip the *Set Up* section and move on to the *Tableau Demo* section on the next page.

Set Up

I recommend you use *Google Chrome* to download all the files for this workshop. To have *Chrome* ask you where to save a file whenever you download something (to avoid having to find it in your Downloads directory and move it to your workshop folder):

1. Start up *Chrome* (if not already running).

2. Click on the customize icon (3 dots arranged vertically) on the top right.

3. Select **Settings**. Scroll down to the bottom and expand the **Advanced** settings. Under **Downloads**, turn on **Ask where to save each file before downloading** and then close the tab.

Next, you should create a folder to store all your files. You are welcome to organize the files as you wish, but if you need some file management assistance, I recommend the following:

4. Right click on the desktop and select **New** and then **Folder**. Name this folder *Workshop*.

Finally, download all the files to this newly created directory.

5. Using *Chrome*, go to http://uoft.me/dataviz (case sensitive)

6. Click on the *WorkshopFiles.zip* file. From the pop-up window, browse to the *Workshop* folder you just created on the Desktop. Then click on **Save**.

7. Go to the *Workshop* folder. Right click on the zip file called *WorkshopFiles* and select **7-Zip->Extract Here**

8. Now you should have two folders within the Workshop folder – *ExerciseDataFiles* and *TableauDataFiles*. Each folder should contain a number of MS Excel files.

9. The *Slides.pdf* file is a copy of the slides for your reference.

10. The *DataVizExamples.pdf* file contains examples of visualizations we will use in the *Improving Visualizations* activity.

11. The *Handout.pdf* file is the handout for the workshop with not only this first page of instructions, but all of the Tableau Desktop demo instructions.

Tableau Demo

Creating a Simple Bar Graph

1. Start by opening up *Tableau Desktop* by clicking on its icon on the Desktop (looks like a white square with plus symbols in it).
2. Let's start by loading some data. Our data is an Excel file, so click on **Excel** under **Connect To a File**, and select the *2015RainfallByMonthByCountry.xls* and click **Open**.
3. Click on **Sheet 1** (at the bottom, the leftmost tab) to open up a worksheet and start creating your visualization.
4. On the left you can see our variables listed, categorized by **Dimensions** and **Measures**. Relating it back to what we've been talking about, dimensions are roughly qualitative data and measures are roughly quantitative data. The centre area is where you will be dragging and dropping your variables onto different areas, such as rows and columns, or to vary marks like colour or size by your variable, or to filter by a variable. In terms of Tableau terminology, those areas that say filters or pages are called shelves, the marks area is called a card, and when the variables show up in those areas, they are called pills, as they are shaped like a pill.
5. Ok let's start by making a simple bar graph. We loaded in some data that shows average monthly rainfall by country (looking at data from 1901 to 2015). So let's create a bar graph with month along the x-axis and average rainfall along the y-axis. So drag the **Month** variable (under the **Dimensions** section on the left) to the **Columns** section and **Rainfall (mm)** (under the **Measures** section on the left) variable to the **Rows** section.
6. You can see that when we dragged rainfall, it automatically summarized it by adding up all the rainfall averages for all the countries. We can change this to average. Right click on the **SUM(Rainfall (mm))** pill in the **Rows** section, go to the **Measure (Sum)** menu and pick **Average**.
7. Right now it is showing data combined for all of our countries, but let's say we just want it to show one of them. Drag the **Country** variable over to the **Filters** shelf and select one country from the list – let's pick **India** and click **OK**.
8. Now we have a bar graph showing the average rainfall in millimetres for India by month.
9. Next let's look at the **Marks** card. You should see 5 boxes labelled Color, Size, Label, Detail, and Tooltip. You can use these to customize your visualization.
10. Click on **Color**, and change the colour of the bars to a different shade of blue.
11. Click on **Size**, and use the slider to make the bars wider or narrower.
12. Click on **Label**, and select **Show Mark Labels** to see the values for each bar.
13. Click on **Tooltip** to adjust the text that shows up in the pop-up you get when you hover over the data in your graph. Add to the bottom of the text "Data from WorldBank". Then click on **OK**. Now hover over the data to see your changes.

14. You can also customize your axes. Right click on your y-axis title, and select **Edit Axis...** Change the title (under **Axis Titles** in the **General** tab) to write out the word "Average" instead of "Avg." Then close the window. The change will be applied automatically.

15. You can also annotate your visualization. Perhaps you want to point out that the summer months are Monsoon season for India, which may be why there is such a spike in average rainfall. Right click on the graph above those months and select **Annotate** and pick **Area...** Then type in "Monsoon Season", change its font size to 12, bold it and click **OK**. Now you can resize and move the box and place it where you want in the graph.

16. Finally, we can give our visualization a title by double clicking on **Sheet 1** at the top and replacing the text with our title "Average Monthly Rainfall for India" and click **OK**. Done!

Creating a Simple Line Graph

1. Okay, let's create a new visualization. A line graph of average monthly temperature data by country (looking again at the same range of years, 1901-2015). First we need to load some more data. Go to the top **Data** Menu and select **New Data Source**. Select **Excel** and choose the *2015TemperaturesByMonthByCountry.xls* file.

2. From this screen, you can see what types of variables Tableau has detected based on small icons above the variables name, and you can make changes if you'd like. For example, you can see that **Year** and **Month** columns have been identified as a string (the small Abc above it). If I want to plot data over time using a line graph, it would be best to change **Month** to date format. To do that, I just click on the **Abc icon** above the **Month** column and select **Date** instead. You'll see the data has changed format and the icon now looks like a calendar.

3. Once I'm happy with my data, I can create a new worksheet to start building a new visualization by clicking on the tab next to where it says Sheet 1. This is a **new worksheet icon**.

4. Let's drag the **Month** variable to the **Columns** section again. This time it is a date variable, so we have more options in our drop-down menu. We want is to make sure we're displaying months, not years, so right click on the **Month** pill and select the first option for **Month**.

5. Next drag the **Temperature** variable to the **Rows** section. You'll notice that our graph is summing the temperatures for all the countries. We can again right click on the **Temperature** pill, and select **Measure (Sum)**, then pick **Average**.

6. We would like to see each country's data separately, so let's drag the **Country** variable onto the **Color** box in the **Marks** card. You can see that Tableau has assigned a qualitative colour palette scheme to represent our countries, but we do have a lot of them, so it is a bit overwhelming.

7. One way to simplify this would be to show 2 countries to compare their temperature distributions (remember this goes back to figuring out what story you want to tell). Drag the **Country** variable over to the **Filters** shelf. Click on the **None** button to first clear the selections. Then select two countries only – let's pick **Canada** and **Brazil**. Now we are just filtering the data to show only Canada and Brazil.

8. Another way we could do this, would be to allow the user to filter it themselves based on what countries they are interested in. To do that, go back to the **Filters** shelf, right click on the **Country** pill and pick **Edit Filter…** Select the **All** button to re-select all the countries and then click **OK**. Then right click on the **Country** pill again, but this time select **Show Filter**. Now you can see the filters show up on the right. We can select or deselect as we like and the graph changes. (If you don't see the filters on the right, click on **Show Me** on the top right to close its options panel that might be blocking the view.)

9. To further help the user read your graph, you could also add a highlighter. Go back to the **Filters** shelf, right click on the **Country** pill, but this time pick **Show Highlighter**. Now the **Highlight Country** box shows up on the right. The user can pick a country and the graph emphasizes that country. To try it out, make sure you aren't filtering any of the countries first, then click on the **Highlight County** search box to the see the list of countries, and then hover over one – try **Canada**. You should see it emphasized in the graph.

10. Let's adjust a bit more of the formatting on this graph. For one thing, I don't like how the months are displayed at the bottom. We can fix that. Right click on one of the months and select **Format**… The **Format** pane should show up on the left. From the **Header** tab, under **Default**, where it says **Dates**, select from the drop-down menu, **Abbreviation**. Then click anywhere on the graph to save it.

11. Finally, let's say you didn't like these colours. Click on the **Color** Box and select **Edit Colors**… Here you can pick from a drop-down menu of various qualitative palettes, including a colour blind safe palette. Select one you like and click on **Assign Palette**. Then click on **OK**.

12. Again, we can give our visualization a more meaningful title by double clicking on **Sheet 2** at the top and replacing the text with our title "Average Monthly Temperature for Various Countries" and click **OK**.

Creating a Series of Simple Scatterplots (and Merging Datasets)

1. Sometimes you have to pull data from multiple sources instead of having it all in one excel file. So, let's add a couple more datasets, but this time we're going to match them up or join them together to create one large dataset to work from.

2. Go to the top **Data** Menu and select **New Data Source**. Select **Excel** and choose the *AuthorDataCitationsGrants.xls* file. This is the dataset we've been looking at in the examples.

3. Now click on **Add** next to **Connections** on the top left to add a second dataset. Select **Excel** and then pick the *AuthorDataExperience.xls* file. This dataset has just author names, along with how many years of experience they have as a researcher.

4. At the top you can see the two datasets listed, and below you'll see that Tableau has already joined the two together, matching on the author name. If you want to double check this, click on the **circles** between the two dataset names at the top. You can see from the icons that Tableau has done an inner join, matching the column Author in one dataset with the Author column in the second dataset. It has copied pasted that information about the author's years of experience next to that author wherever the author shows up in the first table. The inner join

means that it only keeps data on an author if that author shows up in both tables. (Note: If database joins are new to you, see the link listed at the end of the handout for more info.)

5. Again, once I'm happy with my data, I can create a new worksheet to start building a new visualization by clicking on the **new worksheet icon**.
6. This time, let's take a look at creating a scatterplot using the **Show Me** feature. Hold down the **CTRL** button (on your keyboard) and click on **Grants** and **Years of Experience**, and then click on the **Show Me** tab to expand it. I see that a Scatterplot is one of the recommendations (i.e., not greyed out) – select it. If you're not sure which one it is, hover over the images and then read the description below to find out which one is the Scatterplot.
7. Click on **Show Me** again to hide its options panel.
8. First, instead of summing these variables, let's take the averages, as we've done before. So right click on each pill in the **Rows** and **Columns** sections, and select **Average** from the **Measure (Sum)** menu.
9. Our next problem is that it is plotting just one x/y pair – the averages of the whole dataset. We need to plot the points either for each author, country, or institution. Let's do it by author. Drag the **Author** variable over the **Details** box; it explodes out the aggregation to plot it by author. When you hover over each point, you can see the details.
10. You may notice there is now also another box on the Marks card called Shapes. If you want to add another categorical variable to your scatterplot, you could do so by using different shape to represent different categories. Drag **Institution** on to the **Shapes** box. Now you should see that there is a legend on the right, using different shapes for different institutions.
11. If you want to add a trend line, right click in the centre of the graph and select **Trend Lines**, and then **Show Trend Lines**. Hovering over the lines gives you statistical information, such as p-values.
12. Finally, one interesting feature of Tableau is to create not just one visualization, but a series of them, using the **Pages** shelf. Drag **Year** on to the **Pages** shelf. Now you should see some controls on the right. The user can scroll through three different scatterplots, one for each year, or they can click on the play button to have it animate through the years.

Creating a Simple Treemap

1. Go to the top **Data** Menu and select **New Data Source**. Select **Excel** and choose the *2016PopulationbyRegion.xls* file. This dataset lists population totals for various regions on earth.
2. Again, once I'm happy with my data, I can create a new worksheet to start building a new visualization by clicking on the **new worksheet icon**.
3. Treemaps help show hierarchical divisions of parts within a whole. To create a treemap in Tableau, first drag the **Region** variable onto the **Text** box in the **Marks** card, as we're going to separate and label each box with the Region name. Next drag the **Population** variable onto the

Size box as we're going to size these regions blocks by their population. Finally, drag the **Continent** onto the **Color** box to colour code the blocks by continent.

4. You can hover over the blocks to get more information on the populations, or you could label it as well. Drag the **Population** variable again over the **Text** box to include that information under the region name.

Creating a Stacked Bar Chart (and Using Parameters)

1. Go to the top **Data** Menu and select **New Data Source**. Select **Excel** and choose the *RomeoAndJulietWordFrequenciesByAct.xls* file. This dataset lists word frequencies in Romeo and Juliet by Act.

2. Again, once I'm happy with my data, I can create a new worksheet to start building a new visualization by clicking on the **new worksheet icon**.

3. Hold down the **CTRL** button (on your keyboard) and click on the **Term** variable and the **Count** variable, then click on the **Show Me** tab to expand it. I see that a horizontal bar graph is one of the recommendations (i.e., not greyed out) – select it. Click on **Show Me** again to hide its options panel.

4. Let's filter it so we're only see the top 10 words mentioned. Drag the **Term** variable to the **Filters** shelf. Go to the **Top** tab, and select **By Field**. By default it is going to use the Count variable and sum up the instances to get the top 10. Look's good. So click **OK**.

5. Now we have our top 10 terms, but in alphabetical order by term. It might be nicer to sort them by count. We can do this by right clicking on the **Term** pill in the **Rows** section, and selecting **Sort**. In the **Sort By** section, select **Field** and leave the defaults, as it is using the sum of count. Select **Descending** at the top, and then click **OK**, to get the most cited term first.

6. Now I'd like to add the Act information to make a colour-coded stacked bar chart. If we look at what the variable types Tableau has identified, you'll notice that it thinks the **Act** variable is a numeric variable in the measures section, when really it is categorical in this case. So let's change it to move it to our dimensions section. Drag the **Act** variable into the **Dimensions** section.

7. Now it is a category, so it makes sense to visualize it through colour. Drag the **Act** variable onto the **Color** box in the **Marks** card. Looking better, but it seems to be backwards. We can fix that by right clicking on the **Act** pill, and selecting **Sort**. Select **Descending** and then click **OK**.

8. So now we have our top 10 terms, subdivided by Act; however, what if our audience would rather just see top 5 terms, or would like to expand it out to top 20 or 30 terms. We can get an audience's input into our visualizations using parameters. Right click on **Term** in the **Filters** shelf, and select **Edit Filter...** Go to the **Top** tab and click on the drop-down arrow next to where 10 is specified. Select **Create a New Parameter...** Give it a name, such as **Top Number**. Go down to the **Range of Values** section. Set the minimum to **5,** the maximum to **30**, and the **step size** to **5**.

Then click on **OK**, and click **OK** again on the **Filter** window. Now you should see a control on the right that your audience can use to adjust how many terms to see in their top terms list.

Creating a Dashboard

1. Now so far we've been creating visualizations within worksheets. But you can also create dashboards that combine a number of worksheets together. Let's click on the **create new dashboard** icon at the bottom, next to the **new worksheet** icon.
2. You then can drag and drop various sheets to layout a dashboard. You can also drag other objects (bottom left), such as text or images to create your dashboard. Let's drag **Sheet 1** to the top, and **Sheet 2** to the bottom.
3. We can also take the filters we had for one worksheet, in this case the temperature graph, and then apply it to the all the sheets on the dashboard. We can do that in this case, with just a few adjustments to our Sheet 1. Go back to Sheet 1 by clicking on its tab at the bottom. Select the annotation and press the **Delete** key on your keyboard to remove it. **Adjust the title** removing the mention of a particular country. Drag the **Country** filter, set to India, back towards the **Dimensions** and **Measures** section to remove it (which you can do with any variables you want to remove from your visualization).
4. Go back to the dashboard by selecting its tab at the bottom. Right click on the title of the filter, **Country**, select **Apply to Worksheets**, and then pick **All Using Related Data Sources**. Right click again on the **Country** filter title, but this time select **Single Value (list**). This will make sure that your audience can only view one country at the time. Now you'll see that when you select one country, both graphs refer to that country.

Publishing Visualizations

1. If you're working on some visualizations, like we have here, you can save work-in-progress as a Tableau Workbook file. Go to the **File** menu and click on **Save**. Give it a name and select where you want to save the file. (Note: If you do this, it doesn't save the underlying data, so you have keep the data file(s) and workbook together). Or you can **Export Packaged Workbook…** (also from the **File** menu). Then you could share that file with others who have Tableau (and this time it includes the data). In both cases, you can then come back to revisit your work later by reopening it in Tableau Desktop.
2. If you want to export one of your worksheets to an image you can use in a report or article, you can do so from the **Worksheet** menu. First select a worksheet tab you want to save, then there are options in the **Worksheet** menu to either export it as an image or copy it and paste it into a document or image editing software to work with it further. You can do the same thing for your whole Dashboard from the Dashboard menu. Try it out!
3. Finally, you can publish your worksheets and dashboards so that readers can interact with them. Your free option is to create a Tableau Public account and publish it there using that option in the **Server** menu (but note it is public; however, you can adjust the settings, if you don't want readers to download the underlying data and/or workbook). There are also pay options (Tableau

Online and Tableau Server), if you want to be able to limit permissions on who can view your visualizations or to publish to your own server; use the **Publish Workbook** option from the **Server** menu to access these options.