FIT3179 DATA VISUALISATION

Tutorial Week 2: Introduction to Tableau

Semester 2, 2019

Contents

Activity 1: Overview of *Tableau*

Tableau is an advanced visualisation software that enables visualisations to be created with little to no programming. We have chosen this as the focus for this class as there are a wide range of programming skills amongst the students we are teaching and we'd rather focus on the *theory* rather than *programming*.

A limitation in using Tableau is that we are **restricted** in terms of the visualisation choices we can make. We are limited to the visualisation types offered by Tableau. *However,* Tableau is used across multiple industries and the practical content that we're covering is just scratching the surface. It's worth looking at building your skills in this software, even beyond what you're doing for the assignment. There are Tableau specific jobs out there!

One benefit of using a widely respected and well-established product is that there are also many online resources you can draw on. An example can be found here.

Activity 2: Practice with Tableau

You can download the free public version of Tableau (so you can build your assignment at home) here.

Creating a visualisation in Tableau has several steps:

1. Connecting to a data source

2. Building visualisation sheets

3. Creating a dashboard and putting the visualisations into it

In this tutorial, you will learn some of Tableau's basic workflow. Our first step in making meaningful data visualisation would be *designing the visualisation*, because you don't want to just jump in and build something without a plan. Before creating any visualisation, you need to ask **three basic questions**: who, why, and how.

As an example, we want to find a way to visualise world population by country.

WHAT

We located our data source from [a population dataset](#) from the World Bank.

The data covers a lot! It has:
- Country Name
- Country Code
- Total Country Population from 1960 to 2017

WHY

So what is our visualisation macro narrative? We want to allow *comparison* between countries and years, and maybe *identification* of differences.
For this, let's make a *stacked bar graph*, with *filters* for *country* and *year*. This will allow an overview, then filtering to get details.
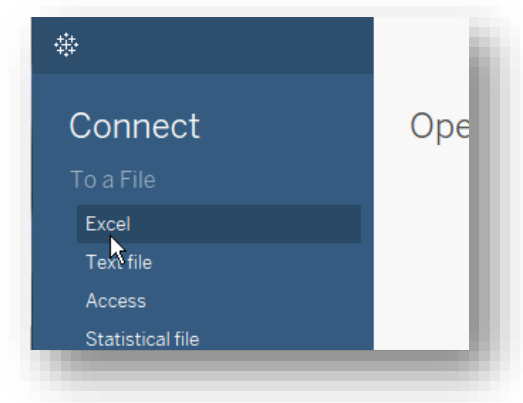
HOW
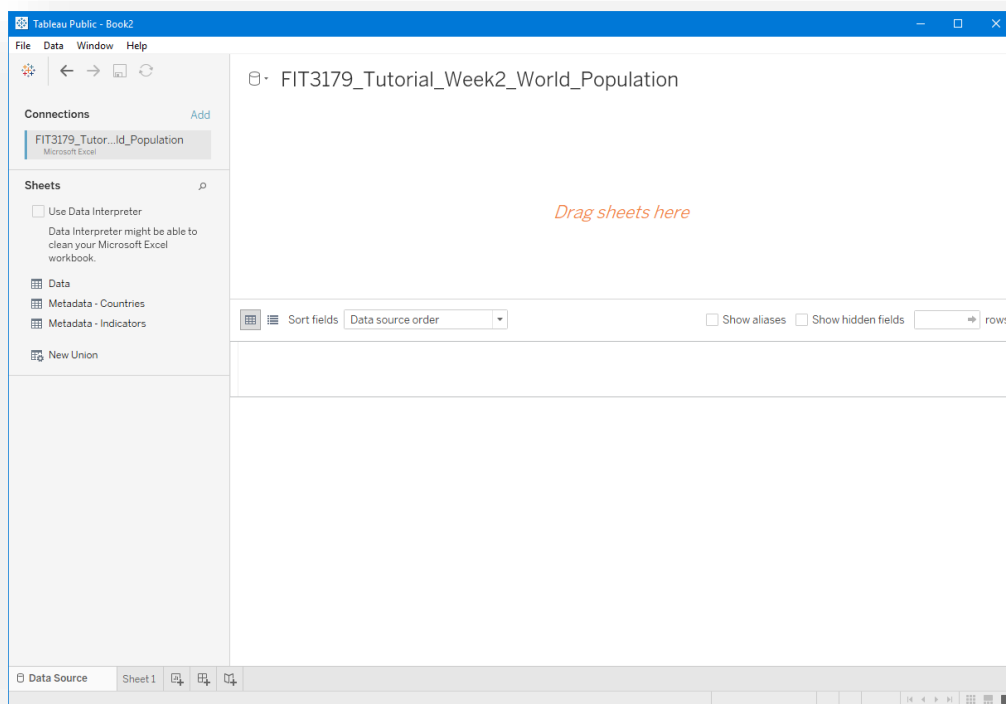
Applying your Tableau skills!

2.1 Importing the dataset
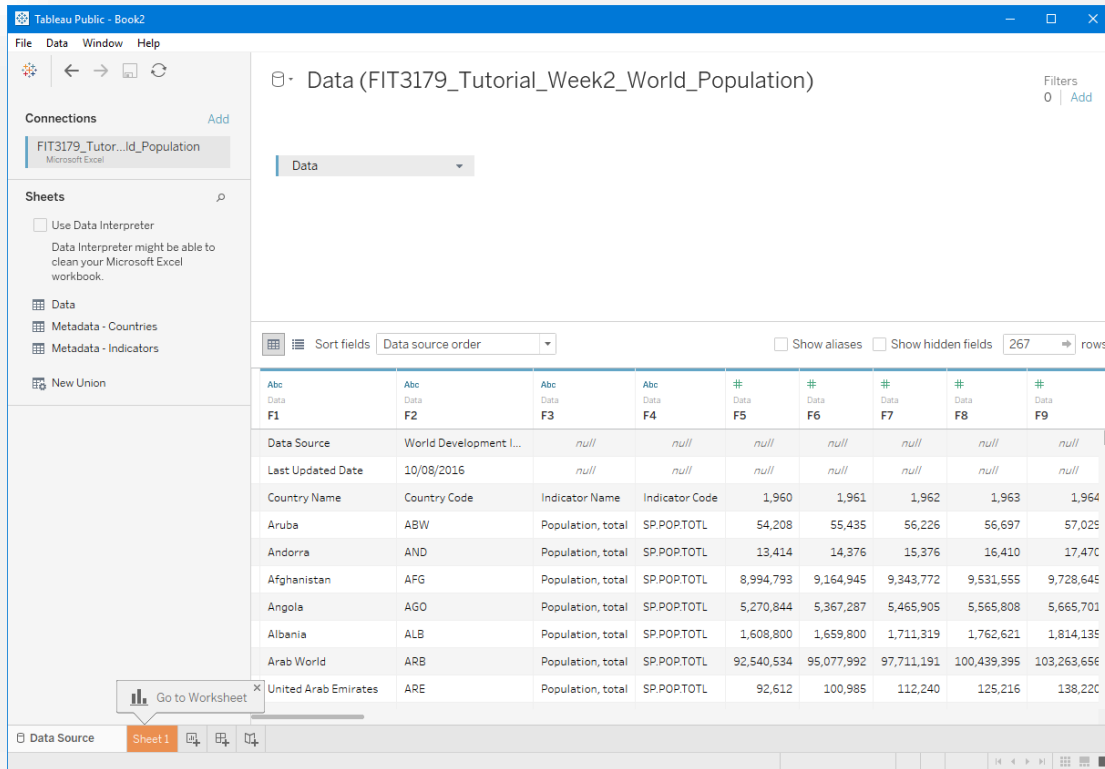
2.1.1 Cleaning the Dataset

1. Open Tableau. We need to provide the project with a **data source**.

2. Let's start a new project and import the excel data from the World Bank: **FIT3179_Tutorial_Week2_World_Population** (available on Moodle).

3. Here you can see the sheets inside the Excel file you just opened. Drag and drop the Data sheet.

4. You should be able to see that there is a problem with the way Tableau interpreted the dataset. The column of your data is not recognised (F1, F2, F3, etc). Because you cannot edit the file on Tableau, you need to make sure your data is clean before importing it.

5. Let's open Excel and clean this up. We just need to delete the first few rows of the first sheet.

	A	B	C	D	E	F
				B18	▾	fx	BEN
1	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961
2	Aruba	ABW	Population, tota	SP.POP.TOTL	54208	554
3	Andorra	AND	Population, tota	SP.POP.TOTL	13414	143
4	Afghanistan	AFG	Population, tota	SP.POP.TOTL	8994793	91649
5	Angola	AGO	Population, tota	SP.POP.TOTL	5270844	53672
6	Albania	ALB	Population, tota	SP.POP.TOTL	1608800	16599

6. Tableau won't refresh our data source even after we have saved it. We'll need to close the connection, then reattach the data source.
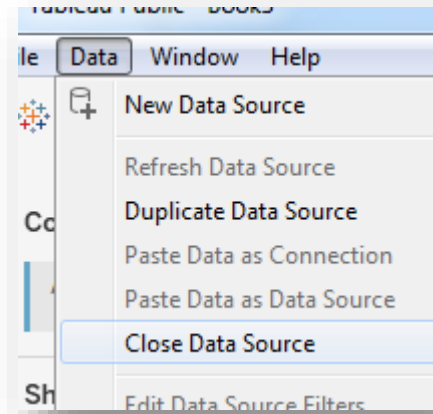

7. If the first row is not recognised as column names, you can manually specify it by clicking on the small triangle icon on data source. And then select "Field names are in the first row".
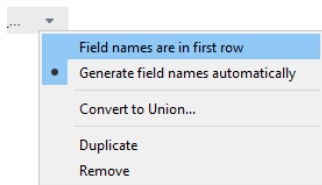
|	Abc	Abc	#	#	#
P.POP.TOT...	API\SP.POP.TOTL.IDS...	API\SP.POP.TOTL...	API\SP.POP.T...	API\SP.POP.T...	API\SP.POP.T...
ntry Code	Indicator Name	Indicator Code	F5	F6	F7
	Indicator Name	Indicator Code	1,960	1,961	1,962
/	Population, total	SP.POP.TOTL	54,211	55,438	56,225
	Population, total	SP.POP.TOTL	8,996,351	9,166,764	9,345,868
'	Population, total	SP.POP.TOTL	5,643,182	5,753,024	5,866,061
	Population, total	SP.POP.TOTL	1,608,800	1,659,800	1,711,319
'	Population, total	SP.POP.TOTL	13,411	14,375	15,370
	Population, total	SP.POP.TOTL	92,490,932	95,044,497	97,682,294
	Population, total	SP.POP.TOTL	92,634	101,078	112,472
	Population, total	SP.POP.TOTL	20,619,075	20,953,077	21,287,682
l	Population, total	SP.POP.TOTL	1,874,120	1,941,491	2,009,526

8. Much better! All of the columns in your data are recognised by Tableau!
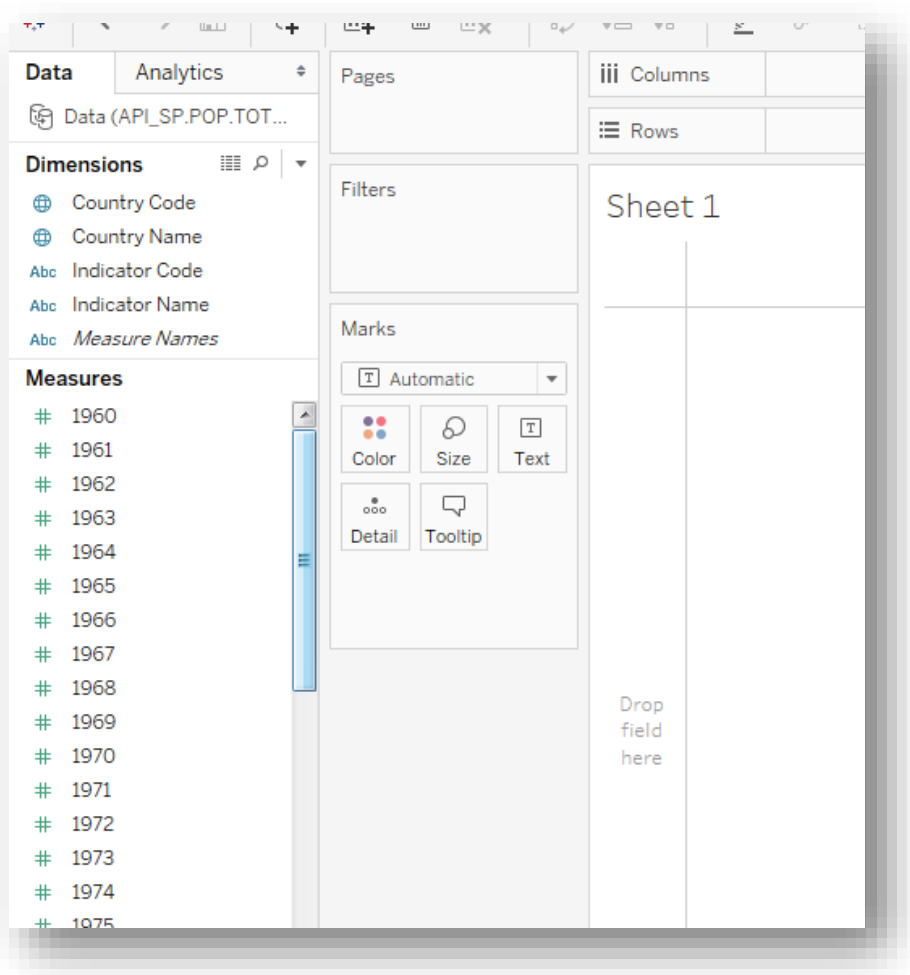

Question

Is there any faster way to clean up this dataset in Tableau?

2.1.2 Importance of the Dataset Shape

1. Once we have done imported the dataset, we can start making visualisations! We can make multiple visualisations with a single dataset. Tableau has already made a sheet for us; let's use that.
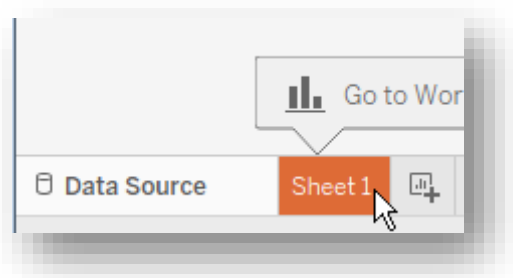
Look at the Data panel. You can see that the fields of the table are categorised into Dimension and Measure.

What are the differences between these two categories?

2. Some important interfaces (panels) that you need to pay attention to are: Dimensions, Measures, Filters, Marks, Columns and Rows. Try to think / discuss/ discover the functionality of each panels.

3. To create our visualisation, what we want is to have **country names** as the **columns**, and the **year data** stacked above each country name. Let's try putting that together.
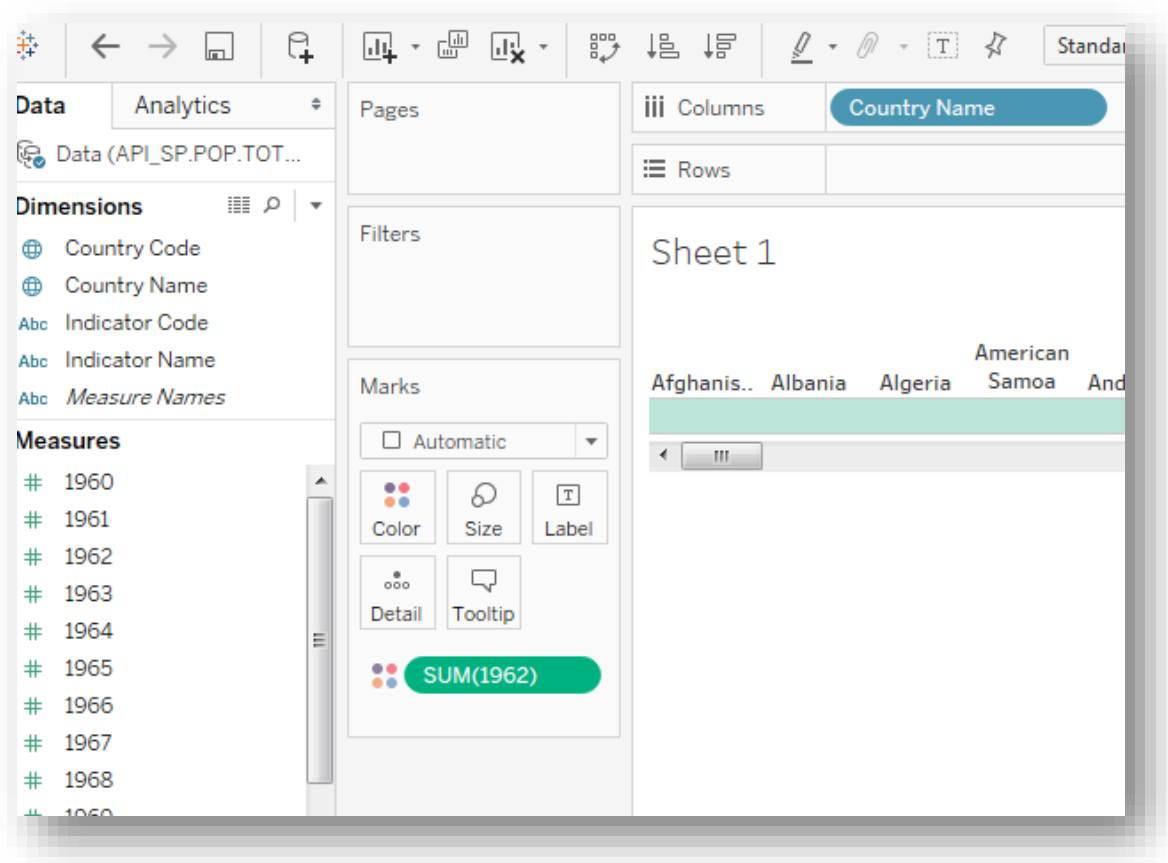
4. We can easily get the country name into the columns. But when we try to add multiple years to the Marks, we can't!

Question

Before moving on to the next step, can you explain why the shape of this dataset is not suitable for analysis?

5. We MUST **RESHAPE** our data! In fact, a good rule-of-thumb is that each piece of data should be in a separate column.

   So what we want is:

BEFORE

Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965
Aruba	ABW	Population, total	SP.POP.TOTL	54,208	55,435	56,226	56,697	57,029	
Andorra	AND	Population, total	SP.POP.TOTL	13,414	14,376	15,376	16,410	17,470	
Afghanistan	AFG	Population, total	SP.POP.TOTL	8,994,793	9,164,945	9,343,772	9,531,555	9,728,645	9,
Angola	AGO	Population, total	SP.POP.TOTL	5,270,844	5,367,287	5,465,905	5,565,808	5,665,701	5,
Albania	ALB	Population, total	SP.POP.TOTL	1,608,800	1,659,800	1,711,319	1,762,621	1,814,135	1,
Arab World	ARB	Population, total	SP.POP.TOTL	92,540,534	95,077,992	97,711,191	100,439,395	103,263,656	106,
United Arab Emirates	ARE	Population, total	SP.POP.TOTL	92,612	100,985	112,240	125,216	138,220	
Argentina	ARG	Population, total	SP.POP.TOTL	20,619,075	20,953,079	21,287,682	21,621,845	21,953,926	22,
Armenia	ARM	Population, total	SP.POP.TOTL	1,867,396	1,934,239	2,002,170	2,070,427	2,138,133	2,
American Samoa	ASM	Population, total	SP.POP.TOTL	20,012	20,478	21,118	21,883	22,701	
Antigua and Barbuda	ATG	Population, total	SP.POP.TOTL	54,681	55,403	56,311	57,368	58,500	

AFTER

Country Name	Country Code	Year	Population
Aruba	ABW	1960	54208
Aruba	ABW	1961	55435
Aruba	ABW	1962	56226
Aruba	ABW	1963	56697
Aruba	ABW	1964	57029
Aruba	ABW	1965	57360
Aruba	ABW	1966	57712
Aruba	ABW	1967	58049
Aruba	ABW	1968	58385
Aruba	ABW	1969	58724
Aruba	ABW	1970	59065
Aruba	ABW	1971	59438
Aruba	ABW	1972	59849
Aruba	ABW	1973	60239
Aruba	ABW	1974	60525
Aruba	ABW	1975	60655
Aruba	ABW	1976	60589
Aruba	ABW	1977	60366

What you can do

You can read a lot more about cleaning and reshaping data for Tableau here.

NOTE: We *strongly* recommend that you read this! It will make your assignment a lot easier to put together.

It would be quite painful to fix this manually.

There is a good online resource for it Pivot Data from Columns to Rows. We suggest you to take a look at that online tutorial on your own time. For the purpose of this tutorial, we provide you with a reshaped dataset.
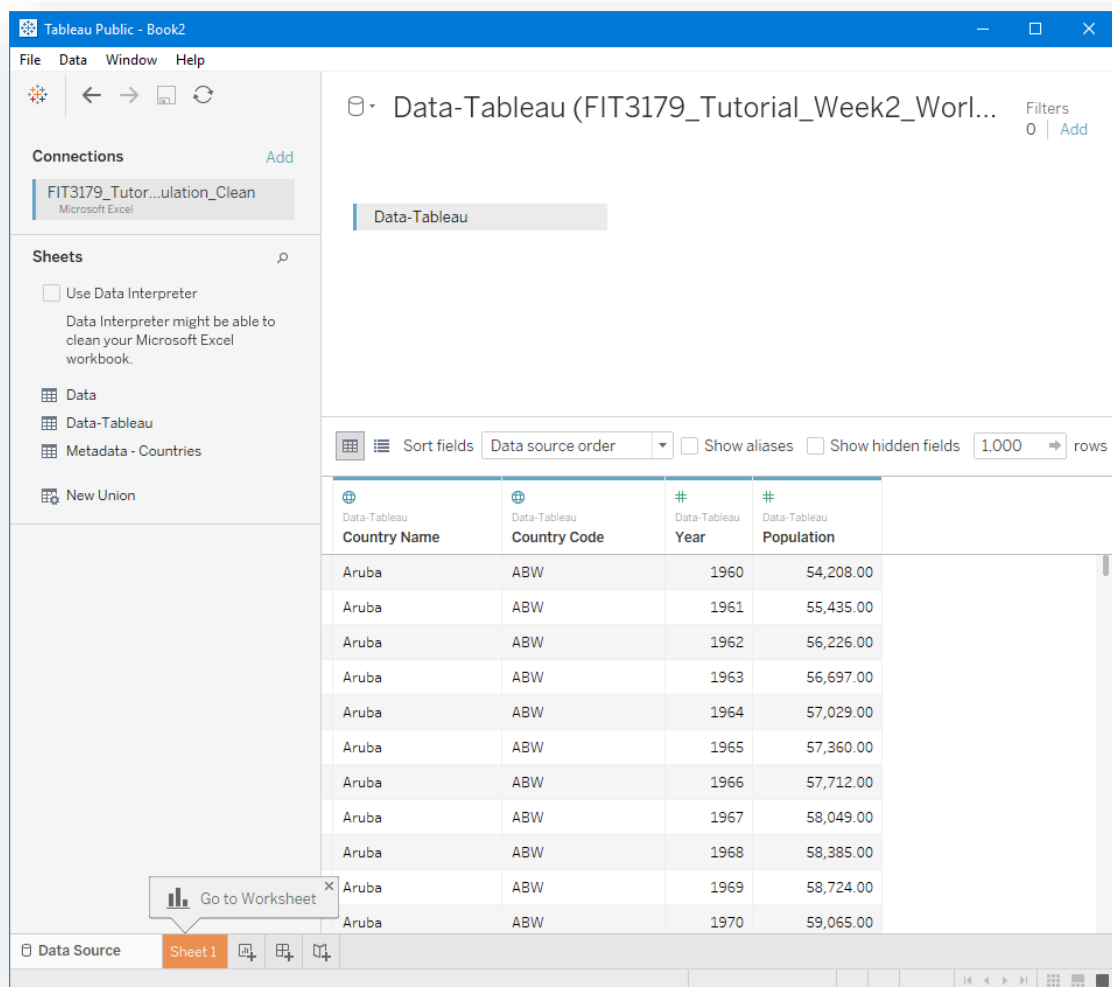
You could also use other data analysis tools in R or Python to reshape your data. However, that is beyond the scope of our tutorial. So let's start again.

2.2   Creating Stacked Bar Charts

Our data is now cleaned and well-structured. It's time to make the visualisation that can deliver the message we want in the WHY step of our design process.

In this activity, we will try to create a stacked bar chart and to add basic interactive filtering mechanism.
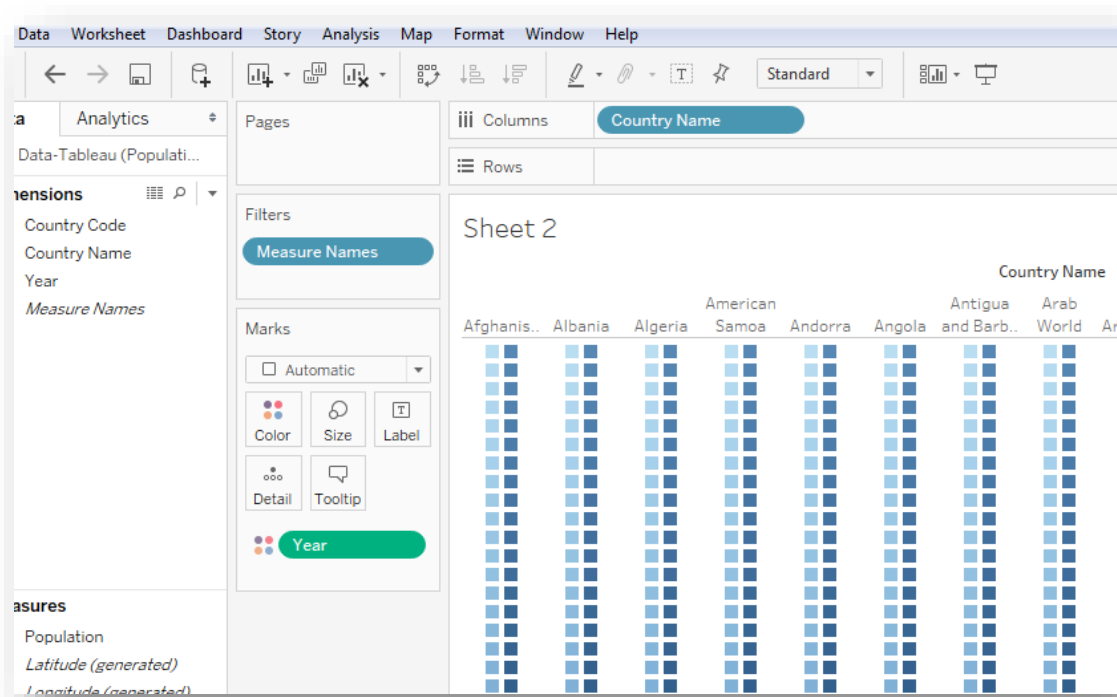
1. Import **FIT3179_Tutorial_Week2_World_Population_Clean csv** file into Tableau. Since csv files are text files, you need to open **Text file** instead of **Microsoft Excel** file. Once you do that, you can see a new data structure similar to the figure bellow.
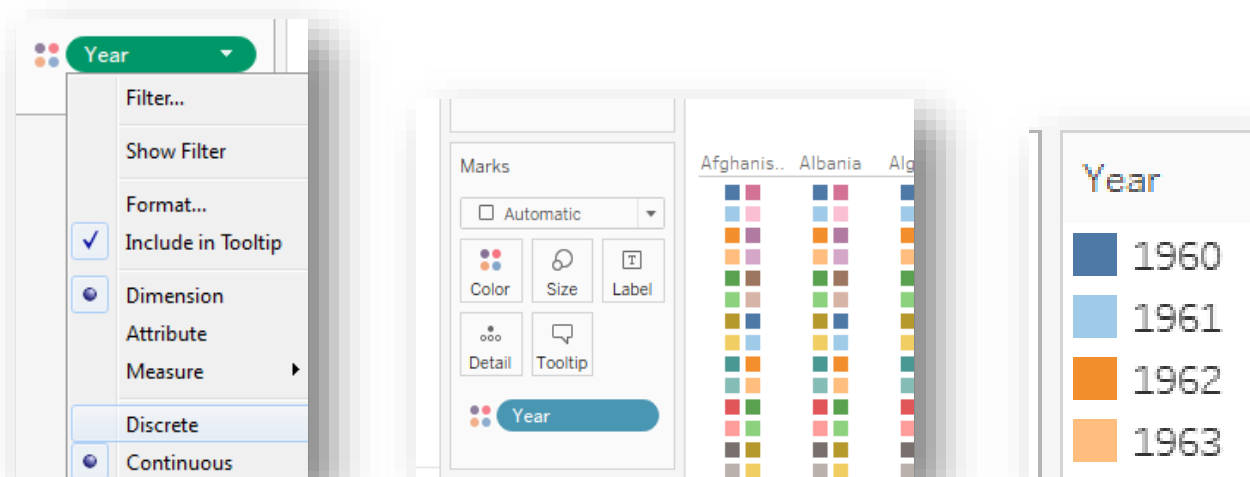

2. Now we can start the worksheet and try to add the country as columns and year as the marks.

3. Much nicer! We will want to easily compare different years. At the moment, everything is in a shade of blue that starts very pale and gets darker.

4. The green colour of the mark interprets year as a **continuous variable**. Since year is a **discrete variable**, we need to change it.


5. Now each year has a separate colour. Much easier to see! However, we still do not have exactly what we want; a stacked bar graph. It should look like:
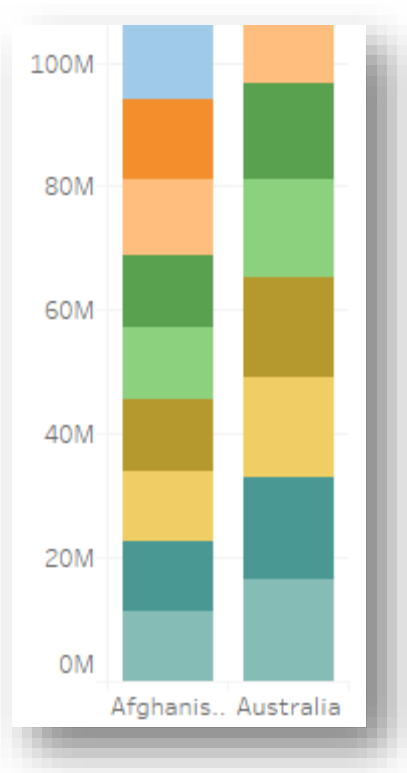
6. Let's try to make this happen. We want to add something to the rows. We need to find data that represents the amount of population in each year. Luckily, Tableau automatically generates this for us:
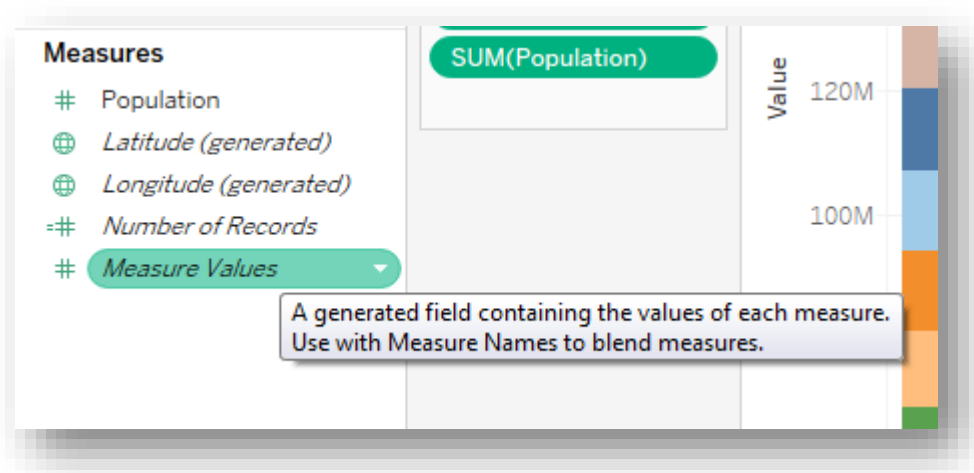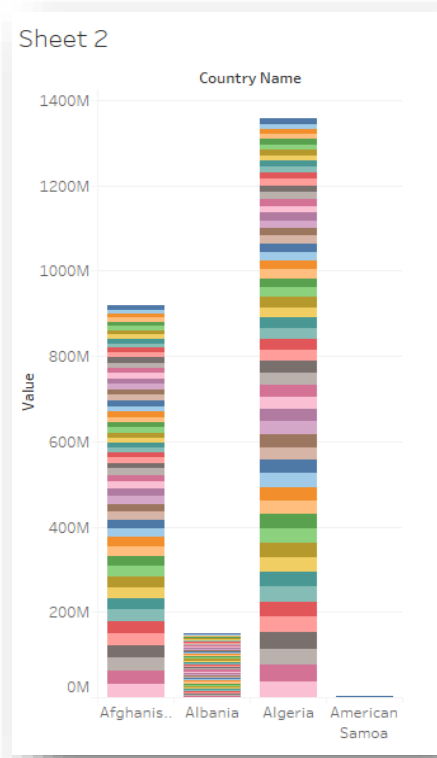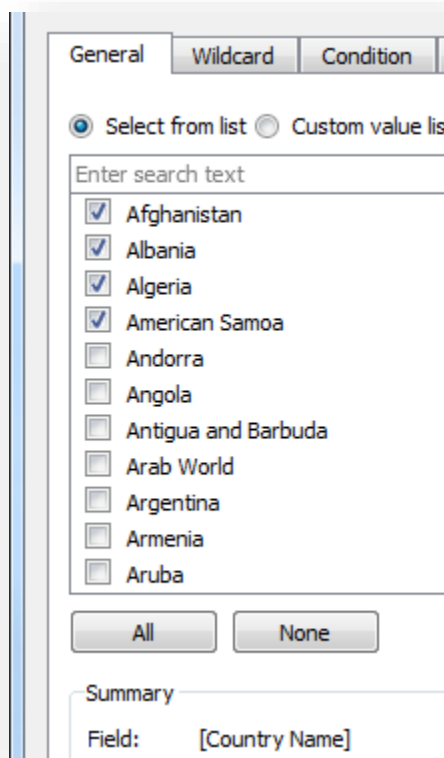

7. This is the value of each measure (in our case, each year). Perfect! Let's add that to the rows.
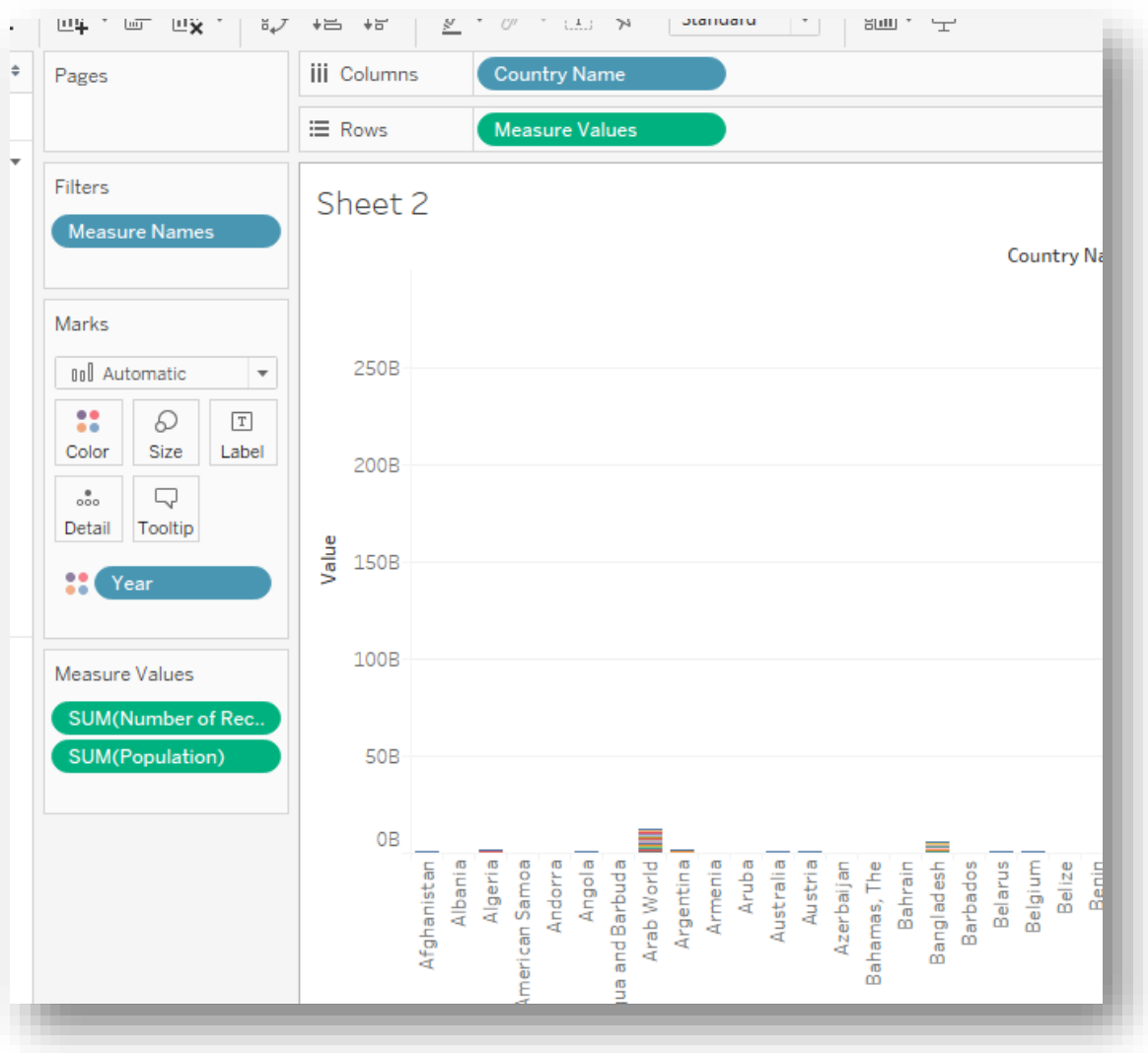
8. It's a bit hard to see what is going on here. The total world population is moving the scale so much that everything else looks small.

9. We can add a filter in to hide some of the data and improve the scale. For now, let's hide everything but the first few countries.
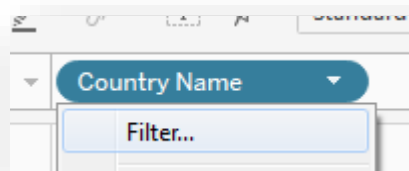
10. Much better. You will notice that when we move the mouse over these values, we are seeing a popup. That is enough data for us; we can remove the legend. Let us also add a manual country filter, so the user can select whichever countries they want to see.
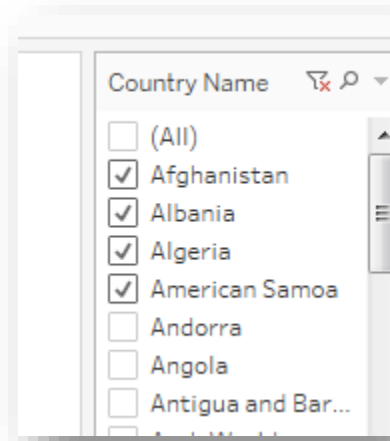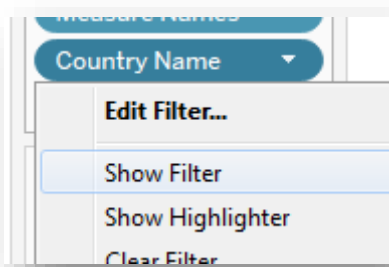

11. Almost done. The last thing we want to do is add a way to limit the years. This allows the user to just select the years they are interested in.
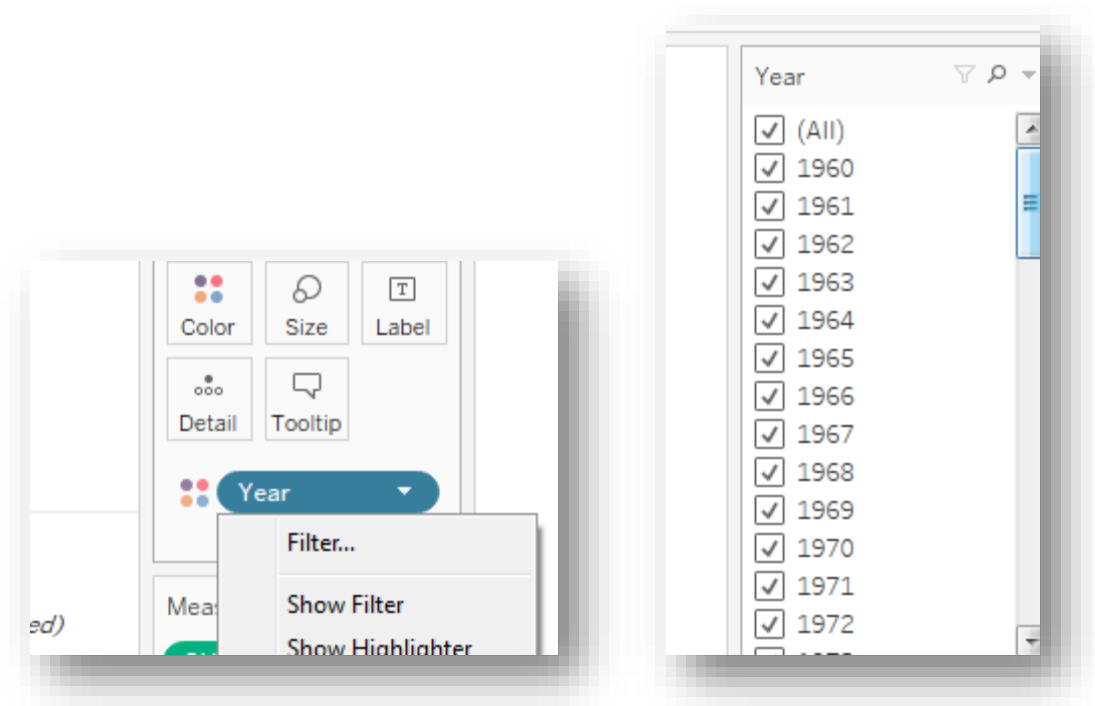
12. The choice of filter is driven by the kind of narrative we are telling. If we want the user to compare two different years (eg. 1960 and 1980) then a list that can be ticked is best.
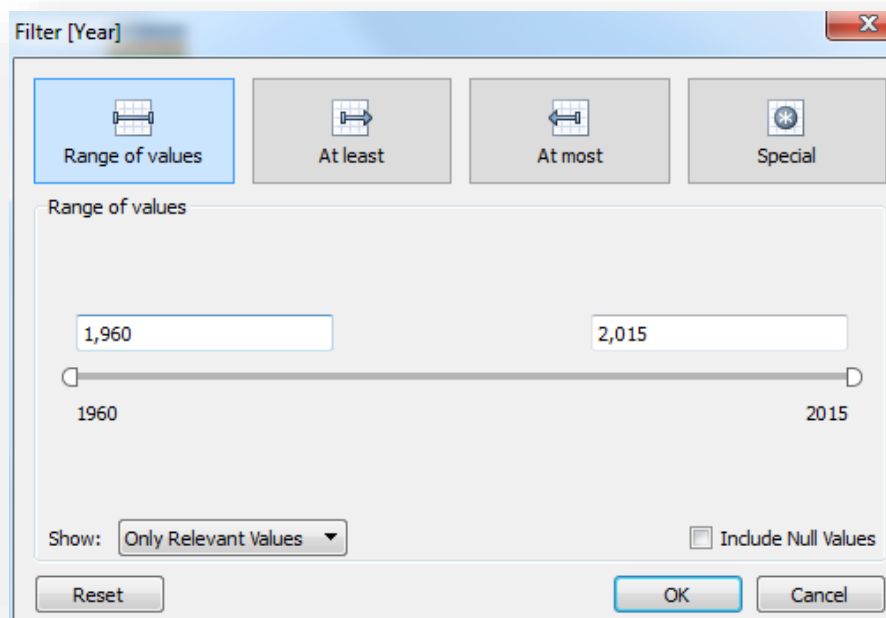
Question
What will be a better interface to have if you want to filter the year by a certain range?
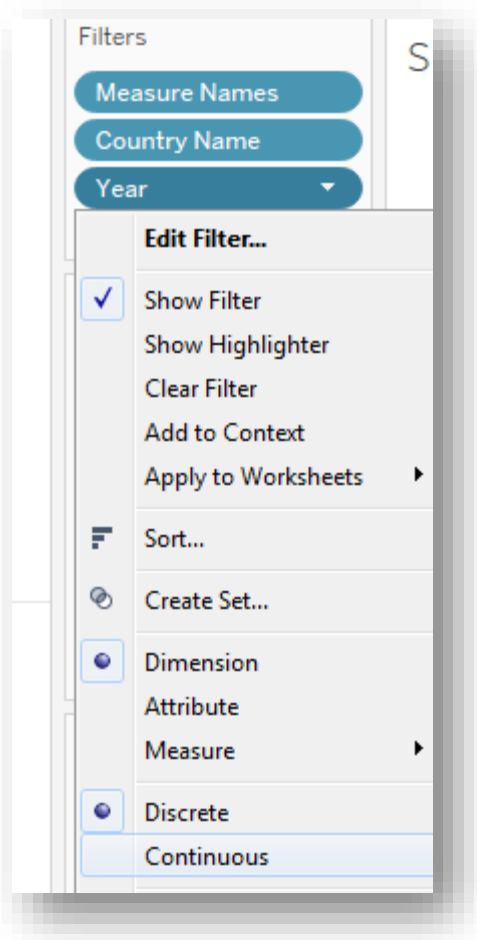
If your answer for previous question is **slider**, you are right!

In order to have a slider, we need to set year as a *continuous* dataset again. We can just change this on the year filter!

13. This will pop a window that asks how you want this continuous dataset to function. Here we are happy with the default settings. We just need to tell Tableau to show the filter, and voila! We are done.

Activity Summary

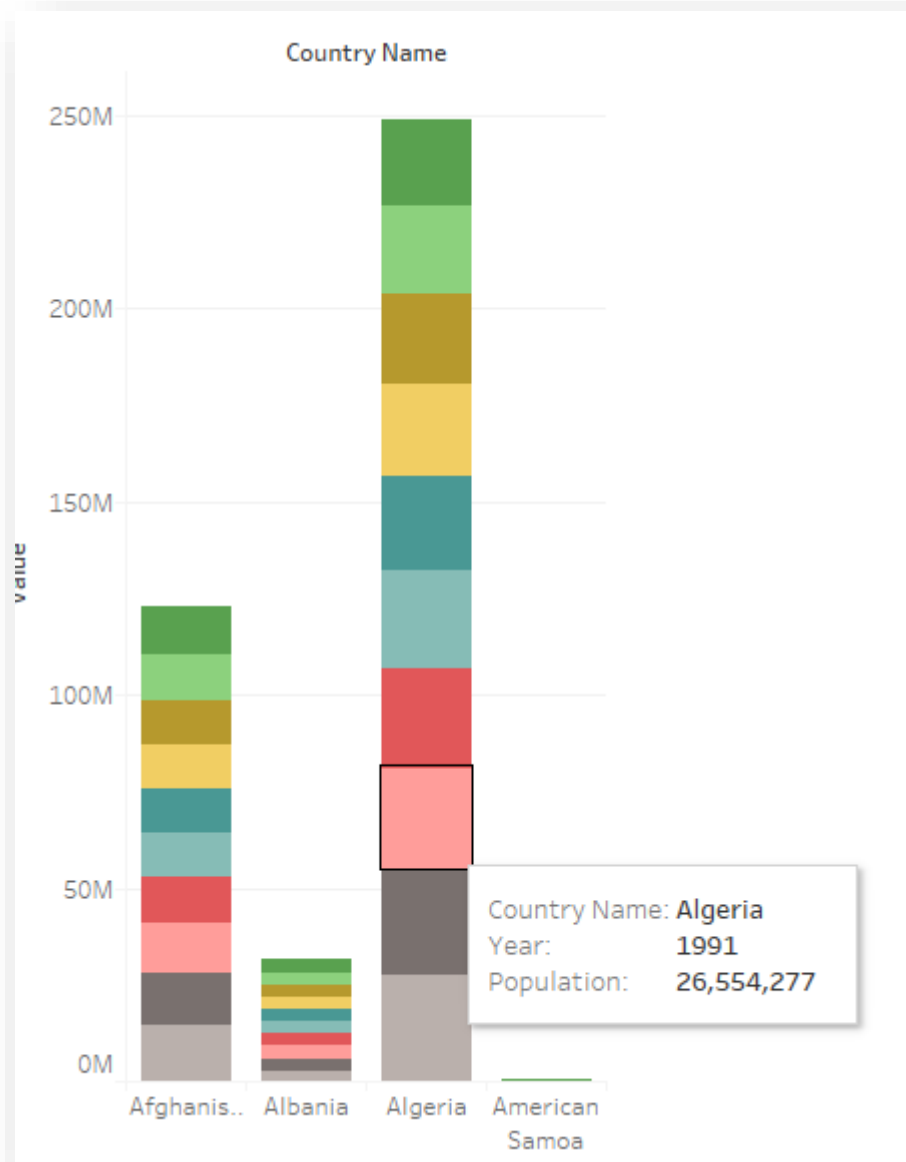*We now have a visualisation that allows direct comparison of any two countries by filtering, allows a range of years to be shown and provides details with a popup. However, please note that this part of the tutorial*

*is designed to demonstrate Tableau functionalities. The result of the visualisation is, indeed, **not a perfect** visualisation.*

Question

Discuss these following questions with the student next to you!
1. *What are the advantages of using a stacked bar chart in this case? Support your argument with examples.*
2. *What are the disadvantages of using a stacked bar chart in this case? Support your argument with examples.*

2.3 Improving Our Time-Series Data Visualisation

2.3.1 Creating Line Charts

1. In this step, we will investigate a different way of communicating our narrative visualisation. Let's start fresh by creating a new sheet.

2. We want to compare the population over the years. Let's plot Year and Population. Put **year** in the <u>columns</u>, and **population** in the <u>rows</u>.
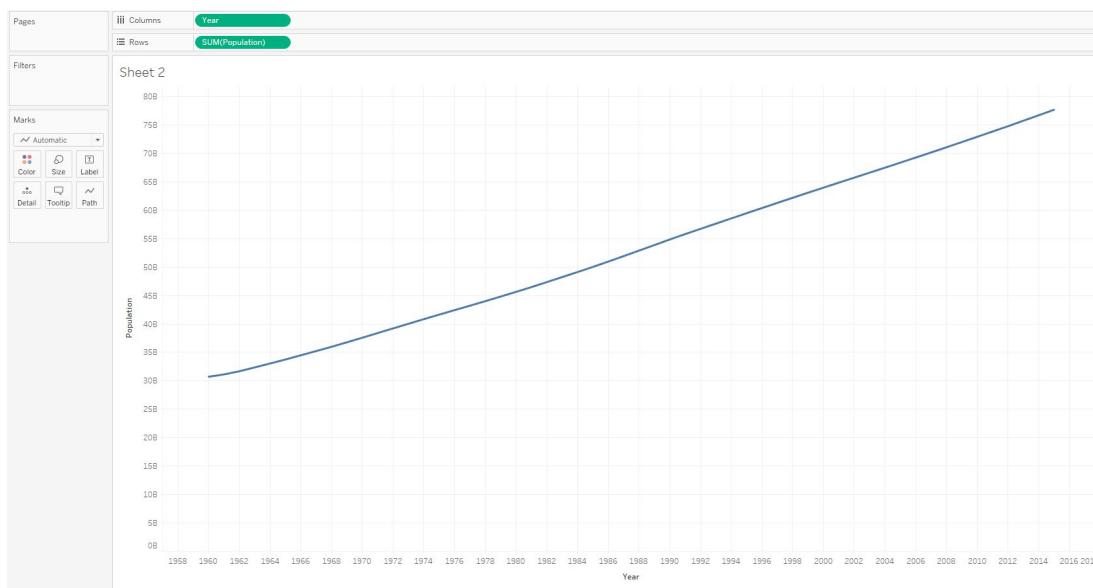
3. It produced a line chart showing the SUM of <u>the population of all countries</u> over the years. We can see the global trend here, but this is not what we want to achieve. We want to see the population of each individual country. To do it, let's map the **country name** to **colour channel**. Drag and drop **country name** to **Colour Marks**.
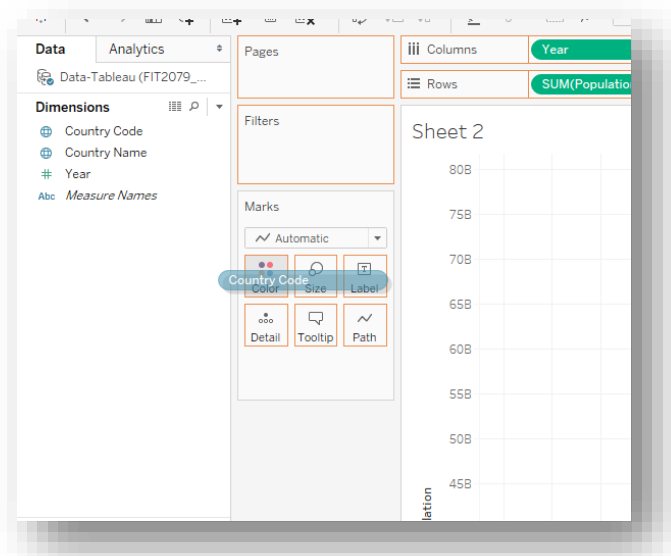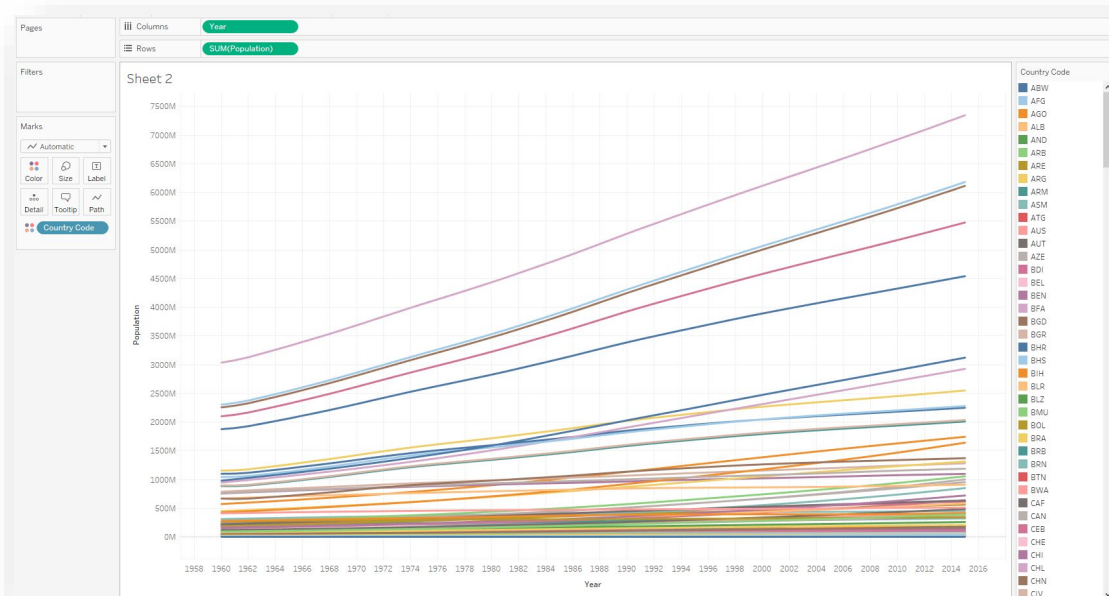
4. It will trigger a warning pop up menu. For now, let's ignore it and click on **Add all members**.
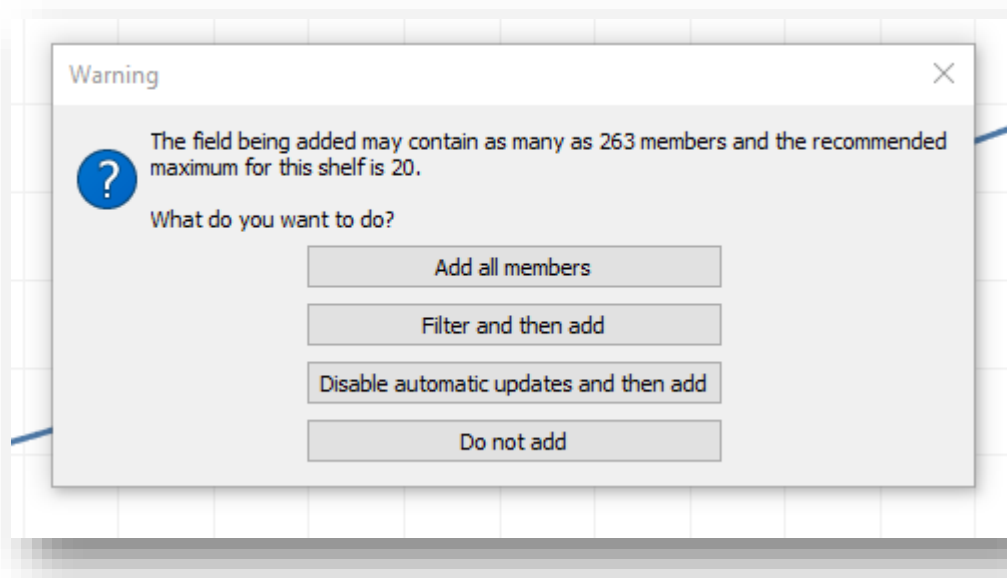
Question

What is this warning trying to tell us?

5. After you add all members, you will see multiple lines that represent **all** countries in your dataset. This way, you can compare the trend of the population for each country. However, this is too cluttered. Let's apply a filter.

6. Do the same process as you did in the previous visualisation (stacked bar chart) to create a filter for each country. Drag and drop **country name** to **Filters** and choose the first four countries.
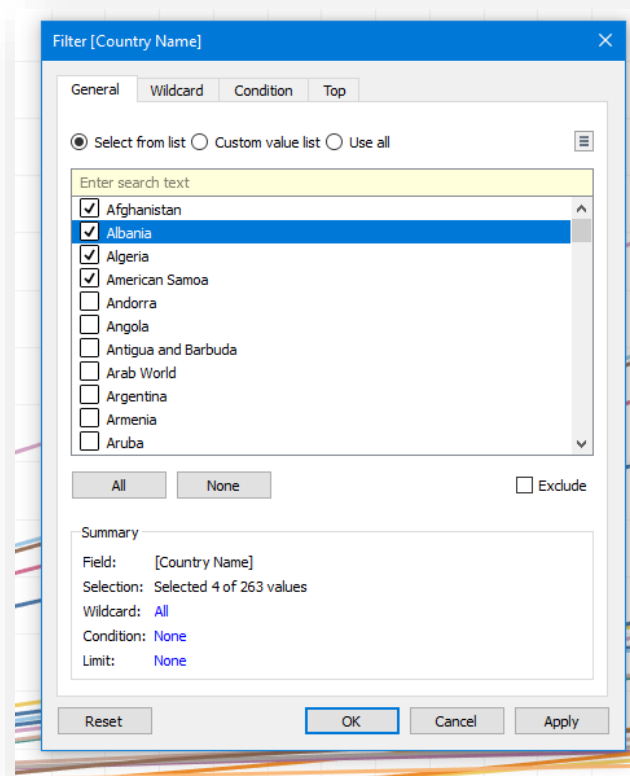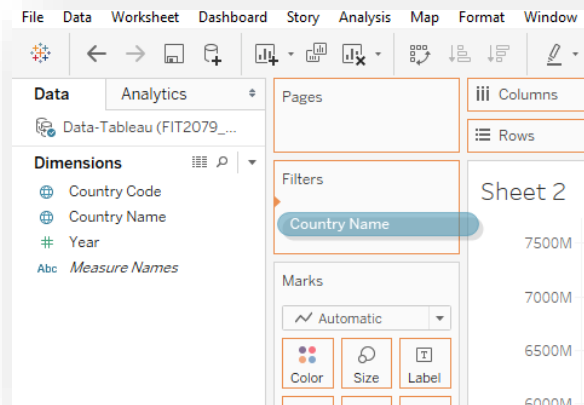
7. Now you can easily compare the trend in population among those countries.
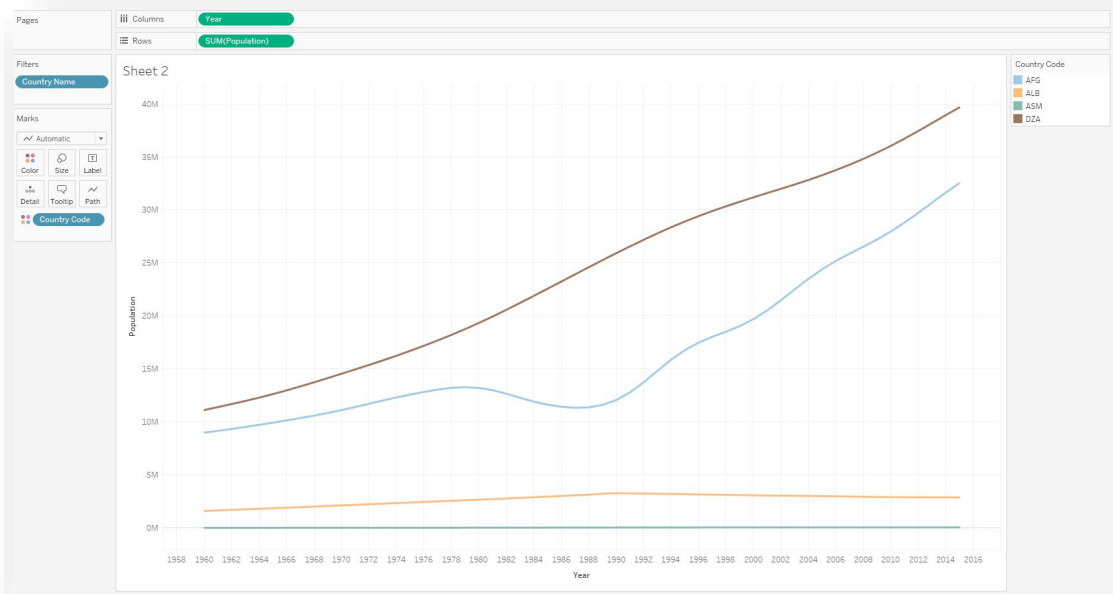
8. It is still missing the year slider and country checkboxes though. Add those filters!
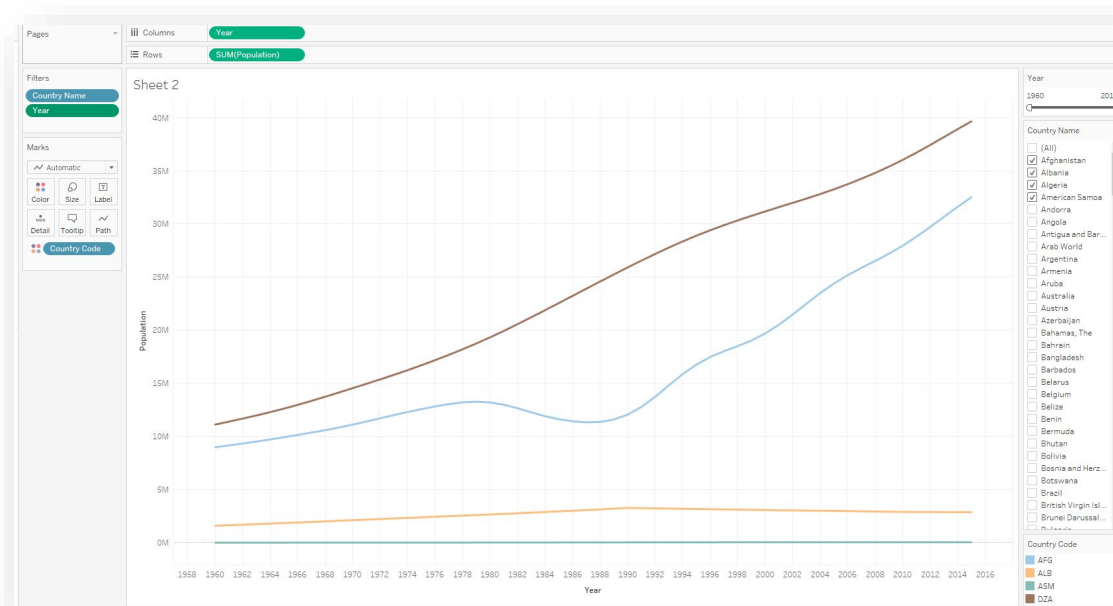

9. Well done! You have finished your interactive line chart.

2.3.2 Exploring the Marks

Lines look like a nice fit for this visualisation. Nevertheless, don't you think it is a bit difficult to compare the population of multiple countries in a single year? Can we easily switch to stacked bar chart to show the population ratio among countries in a single year? Yes, we can!

Look at Marks panel. Try to change the Line to Bar. You should see this visualisation!
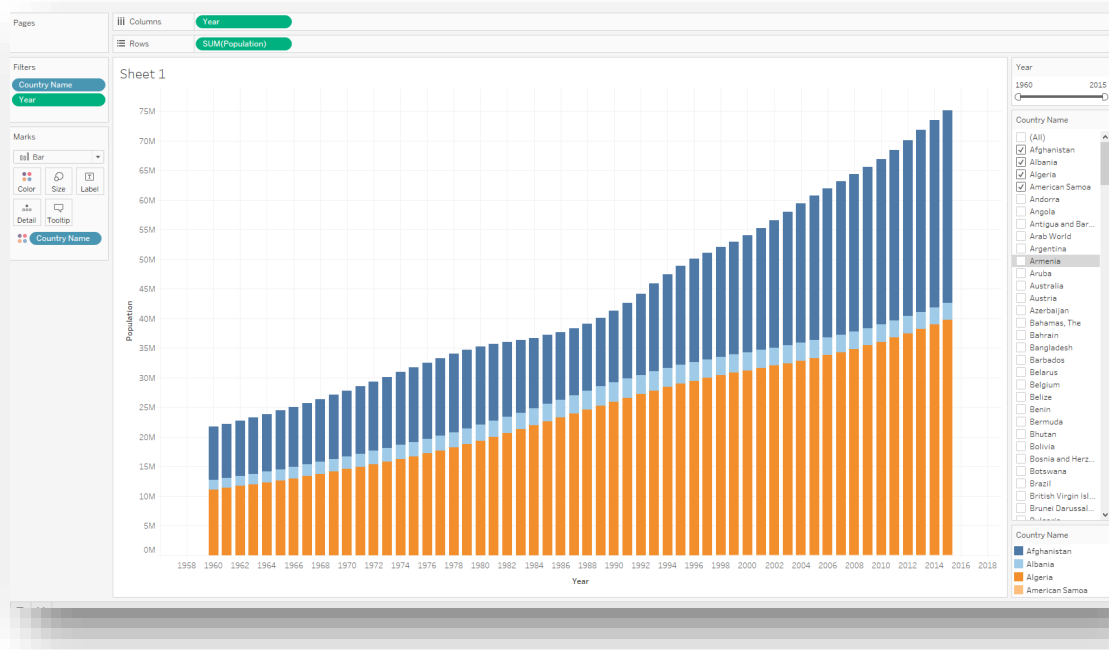
Explore other mark types. You might discover something interesting!

<div style="background:black; color:white">Activity Summary</div>

In this activity, you have learned how to create interactive stacked bar chart and line chart visualisation in Tableau. You also learned the characteristics of each visualisation. Line charts are good to show the trend of an interval or ratio (population) data from an ordinal variable (year). We could also use colour channels to map other variables such as country (a categorical variable). Bar charts are good to compare the trend of a discrete data across an ordinal (year) or categorical variable (country).

The visualisation clutter can be reduced by introducing interactivity to your visualisation. In this case, we provided an interactive filter for both year and country.

In the next tutorial, you will learn more exciting features that Tableau provide to support you in making your visualisation meaningful as well as beautiful.

Activity 3: Using Visualisation Templates

You might have or might not have noticed that Tableau provides you with some basic templates under the **Show Me** button.
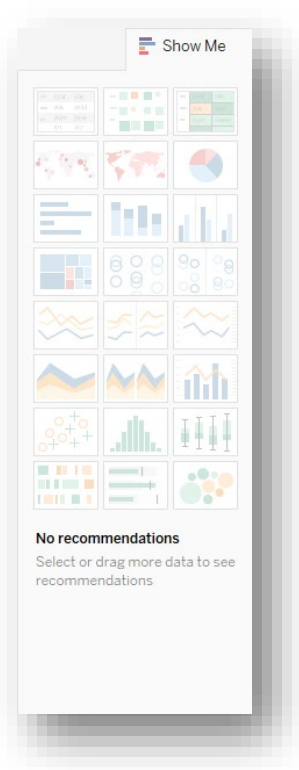

If you hover your cursor on one of the templates, you will see the data type requirement to create a visualisation. For example, a bar chart requires **0 or more Dimensions** and **1 or more Measures**.
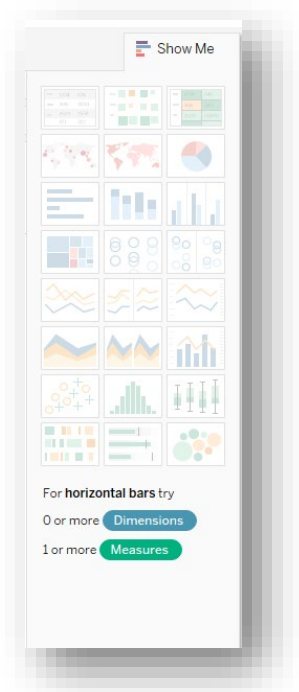
Next, you will explore other ways of visualising your data and think about the characteristic of each chart types. Create a new sheet and explore various charts provided in the templates!
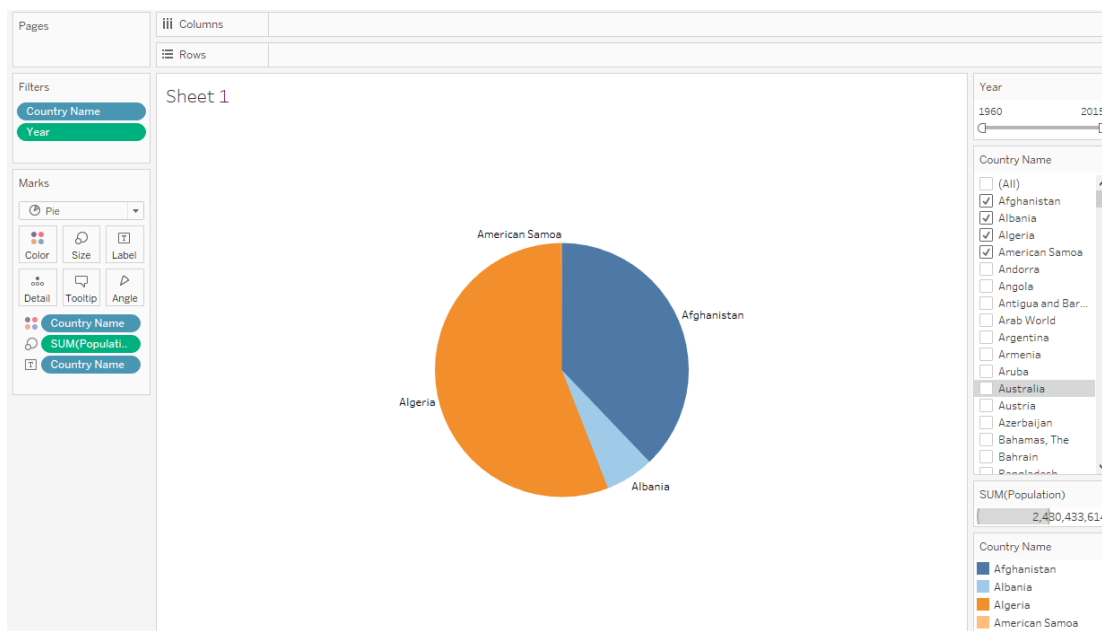
3.1 Creating Tree Maps

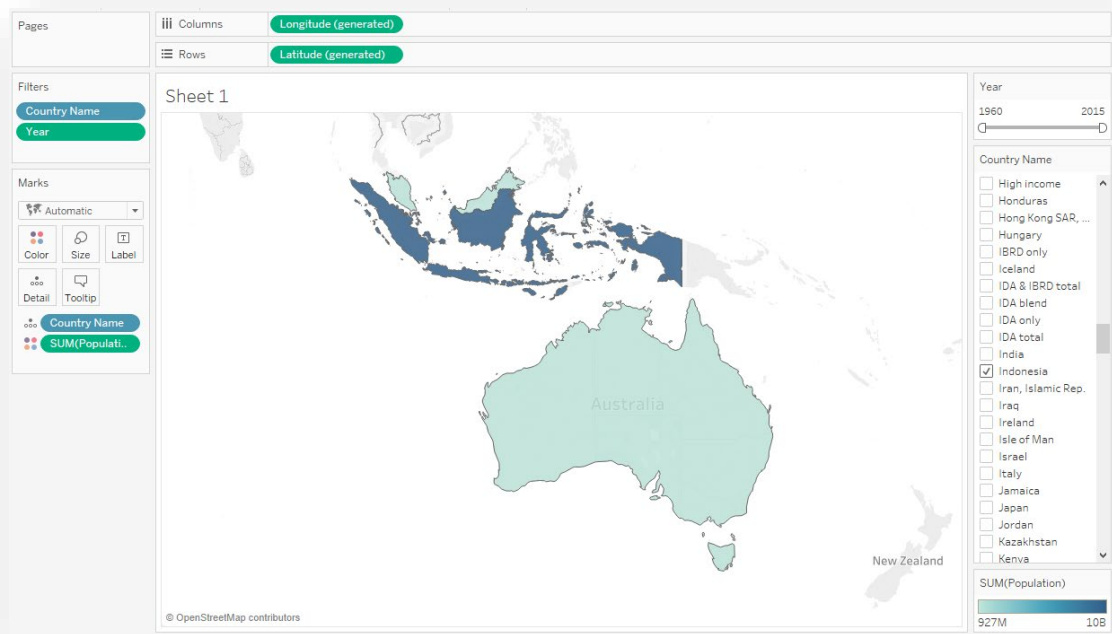Try to create a tree map and discuss what it is good for!
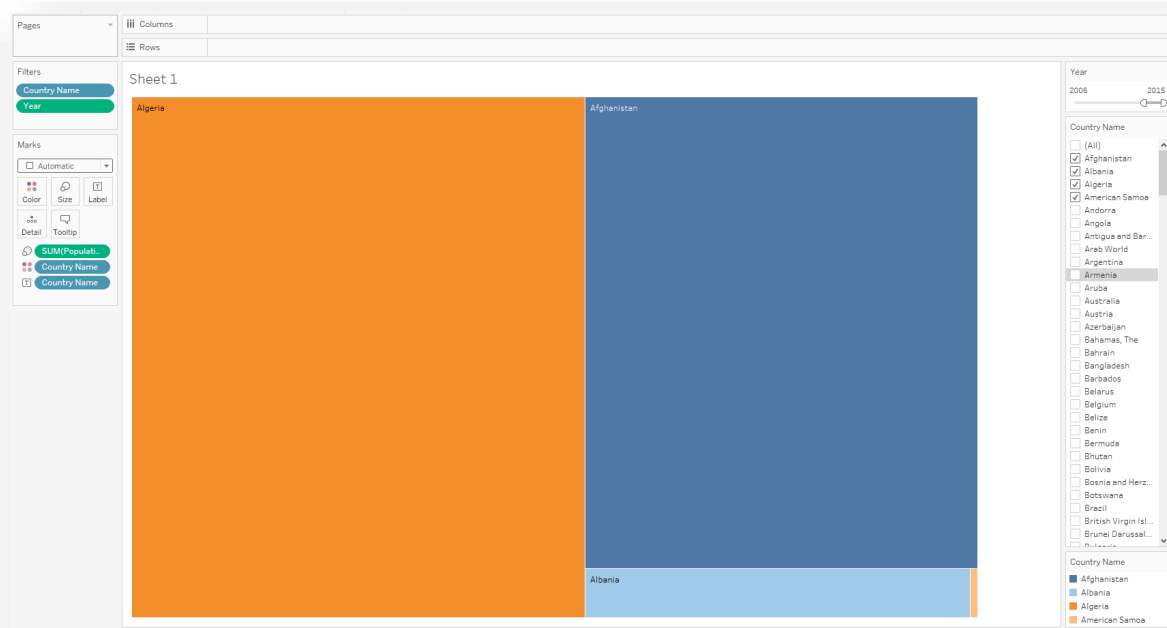
3.2 Creating Pie Charts

Try to create a pie chart and discuss what it is good for!

3.3   Creating Maps

You can also create maps! We will learn more about maps in another tutorial.

Once you finished exploring different marks, save your work so that you can open it up later.

3.4   Saving Your Work

Unfortunately, we can't save our work offline in Tableau Public. We can only save our project to our online Tableau profile.

You will be able to open this again (using the online version) if you need to edit it further.

1. To save your work on Tableau Public, click File → Save to Tableau Public
2. Name your workbook
3. Done. Your web browser will open up and show your visualisation online.

If you are using Tableau Desktop, you can save the workbook file offline by clicking File -> Save.