

FIT3179 DATA VISUALISATION

Tutorial Week 4

Tableau: Brushing and Linking, Dashboard and Multiple Data Sources

Data Visualisation 1: 5DS Methodology

1. Overview

This week's Tableau tutorial consists of four activities.

1. Creating an Interactive Visualisation in Tableau
2. Creating a Dashboard and Combining Visualisations in Tableau
3. Working with Multiple Data Sources
4. Blending Data Sources

After the Tableau exercise, you should start designing your data visualisation 1. Please discuss the domain and ideas with your tutor.

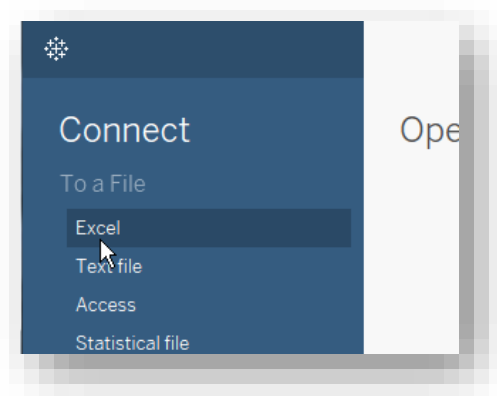
2. Creating an Interactive Visualisation in Tableau

In the previous Tableau tutorial, we learned some basic workflow in creating a visualisation with Tableau. We also learned to create a simple filtering mechanism as an interaction. In this tutorial, we will explore more of the interactivity that Tableau provides. We will create multiple visualisations and add interactive filtering using mouse click on a map.

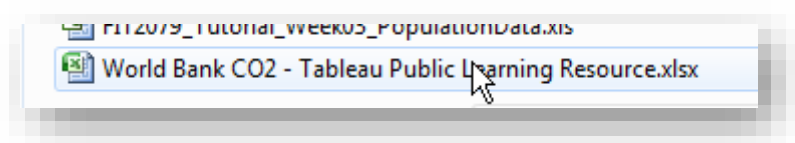
First, let's import the data we need and recall some of the knowledge that we learned over the previous weeks.

2.1. Connecting to a Data Source

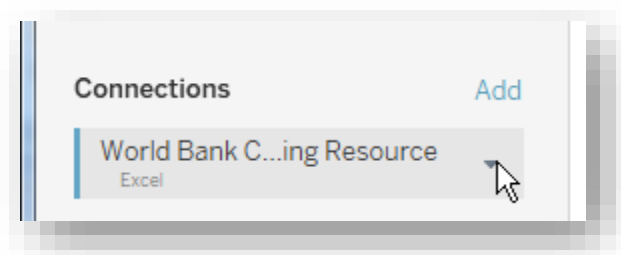
1. Open Tableau. We need to provide the project with a data source.



2. Check Moodle for the Country emission dataset for this week. The file name is World Bank CO2 – Tableau Public Learning Resource.xlsx.



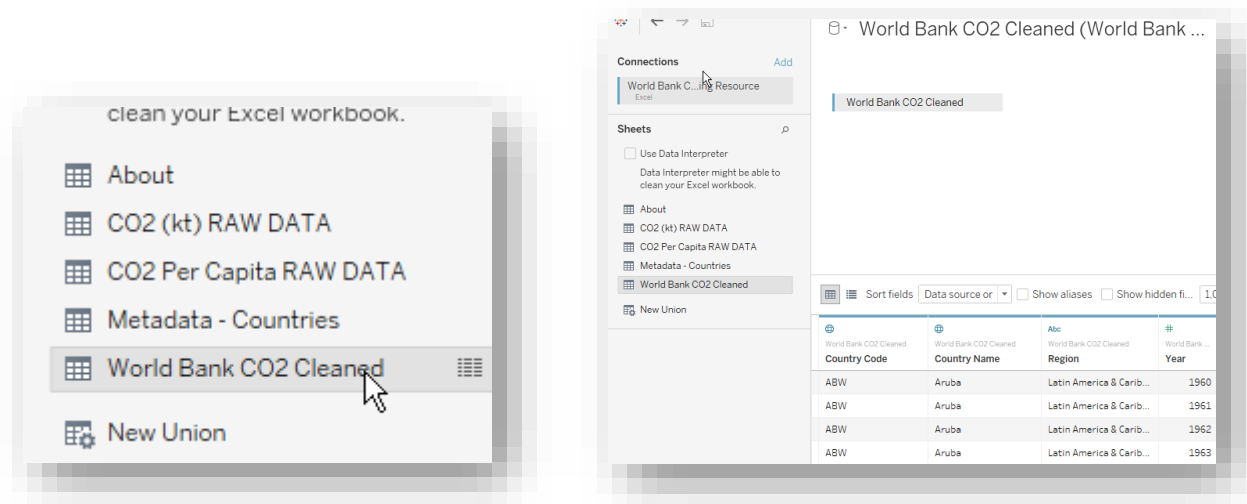
3. Once we have added this, we can add extra datasets if we want more data (this is useful!). We can also remove datasets, but we must have at least one dataset. So, we cannot delete the one we just added until we have added another.



Question

Can you find out how to switch between data source connections if you have multiple data sources?

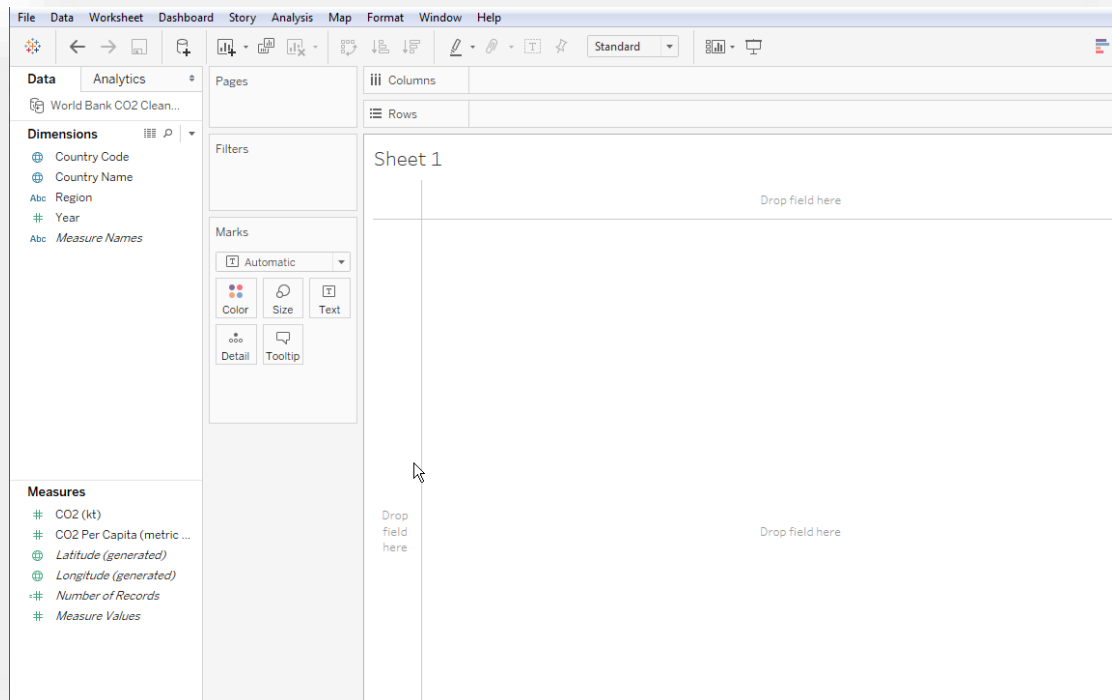
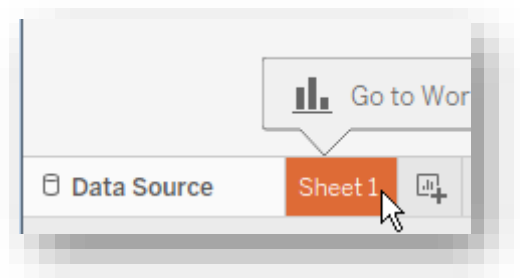
- There are multiple sheets in the excel file, but we only want to use the World Bank CO2 Cleaned sheet. As per the name, this data has already been cleaned for us. Note that for your assignment, you may need to clean your dataset before you can use it.



Question

- Can you find out the feature on Tableau that indicates the data type of each field on World Bank CO2 Cleaned sheet?
- What are the data types of Country Name, Region, and Year?
- What will happen if you change the data type of Year to date? Change it back to the original data type before you continue.

- Once we have done this, we can now start making visualisations! We can make multiple visualisations from a single dataset. Tableau has already made a sheet for us; let's use that.

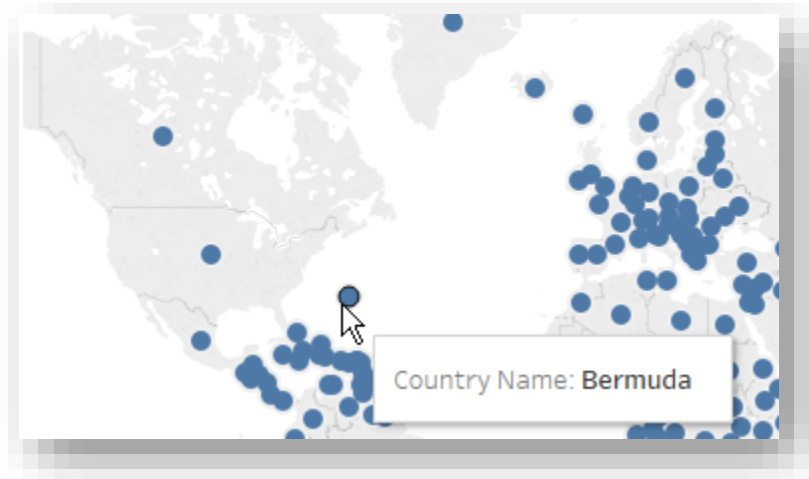


Question

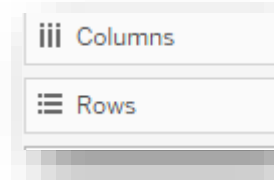
1. Let's recall what you learned in the week 2 tutorial. Look at the **Data** panel. You can see that the fields of the table are categorised into **Dimension** and **Measure**.
What is the difference between the two categories?
2. Suppose we have a new field named Average Temperature, what category do you think this field will fall into?

2.2. Building Visualisation Sheets: World CO2 Emission in a Map

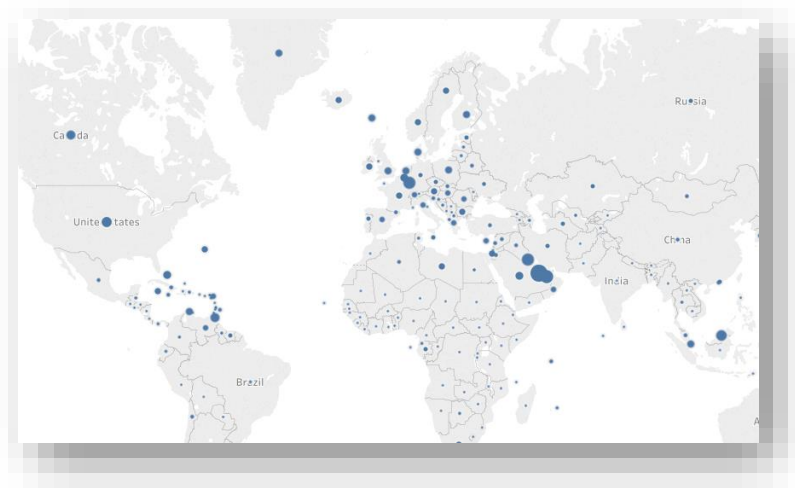
1. Now we want to have a map showing the data. Easy! Drag the **Country Name** dimension into the main window. Tableau will recognise each country for us and will automatically generate a map! It will furthermore put a dot on each country in our map (how easy is that?!).



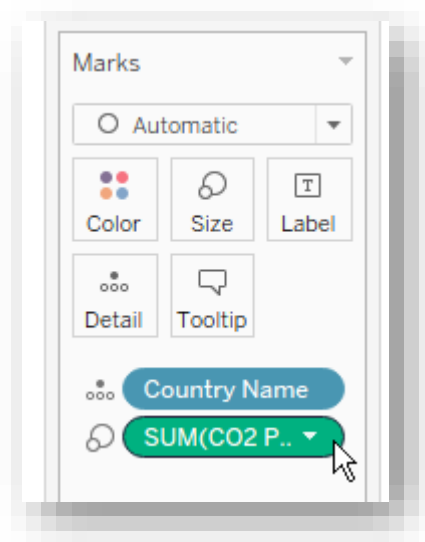
Look at the **Columns** and **Rows** shelves; do you see any changes after dropping the **Country Name** to the sheet?



2. The map is just showing the country names and *nothing else*. Can you find a way to encode **CO2 Per Capita** into the visualization so that the size of the circles represent the **CO2 Per Capita** value (see the image below)?



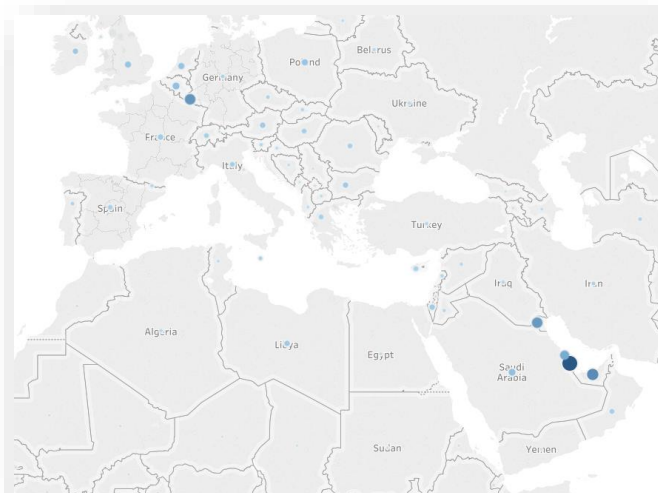
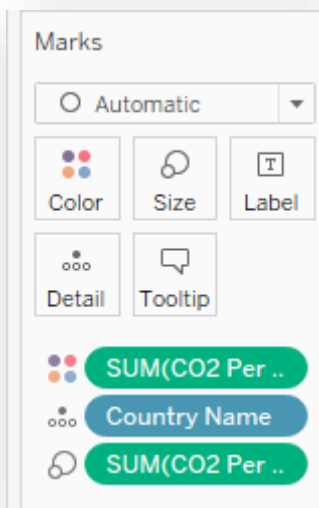
- Look at the **Marks** panel. There are two marks used so far: **Country Name** for details, and **SUM(CO2 Per Capita)** for the size.



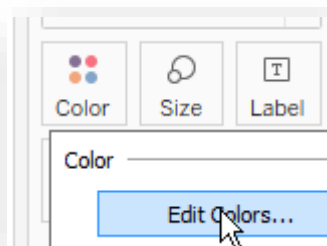
Question

- What does the *SUM* in the **Marks** panel means?
- Can you change the mark such that the size represents different measure of the CO2 per capita such as average, maximum, minimum, etc.? How?
- Once you have done, revert the measure to *SUM*.

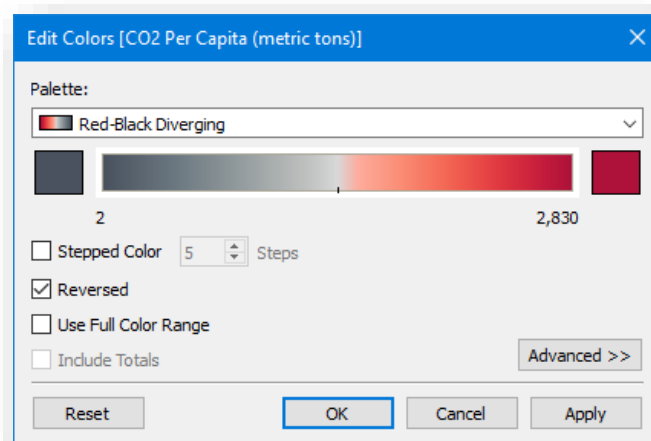
- Let's do something more complicated. We'll have the emissions denoted as colour. We can do this by dragging the **CO2 Per Capita** onto the **Marks** <colour> icon.



5. Now we have 3 sections in **Marks**! The colour is from white to dark blue based on emissions. This isn't easy to read. Let's change this colour palette. We can <left click> on the **Marks** <colour> icon and edit the colour.



6. We will change this to be red-blue diverging, flipping the colours so they start at blue and end in red. Why do you think we reversed this?



Question

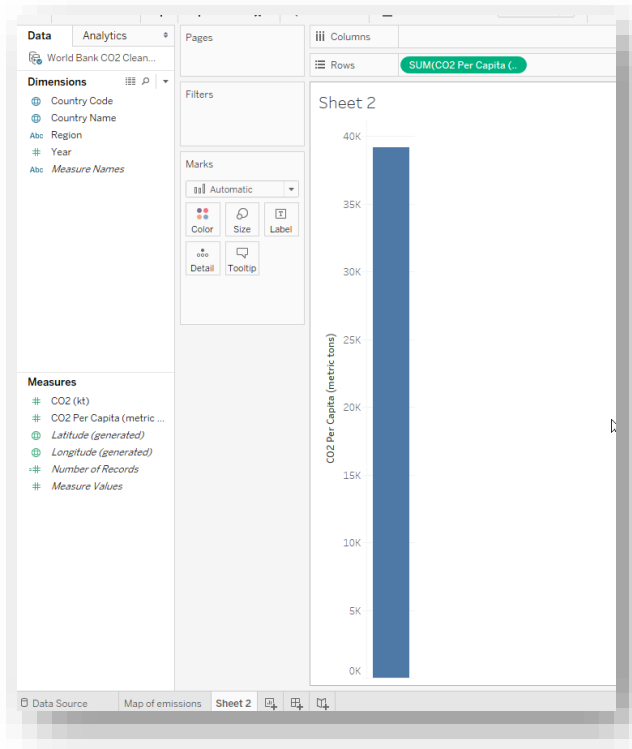
1. Recall again the theory in the lecture about encoding value into colours.
What are the colour elements (Hue, Saturation, Value) that are involved in the blue-black diverging colour scheme?
2. *Do you think it is a good idea to use blue-black diverging? If it is not, what would be a better colour scheme to use in this case? Why?*
3. *Remember, this exercise is designed to introduce Tableau functionalities that you can use for your project. This proportional symbol map may not be the best idiom to visualise this data. Can you think of a better visualisation idiom in this case? Why is that better?*

Much better! Suppose we were happy with this. Let's name this sheet by <right clicking> on the sheet name. Rename it as "World CO2 Emission – Map".

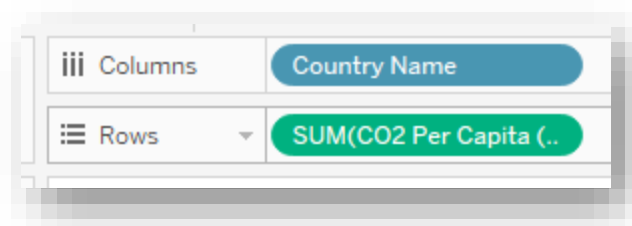
2.3. World CO2 Emission – Line Chart

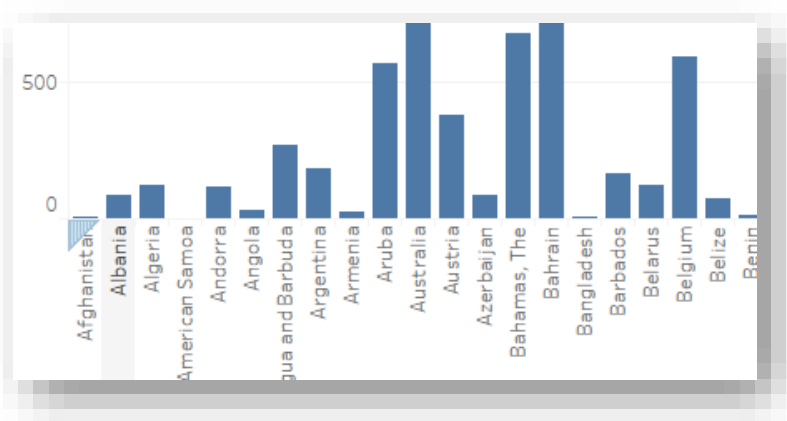
Now let's make a 2nd sheet

1. Instead of dragging anything onto the new sheet, let's <double click> on the **CO2 Per Capita** data. This will automatically create a bar chart for us. Note that at the top of the sheet we can see the **rows** (i.e. left axis) are being driven by the **total sum of emissions**. Not very interesting!

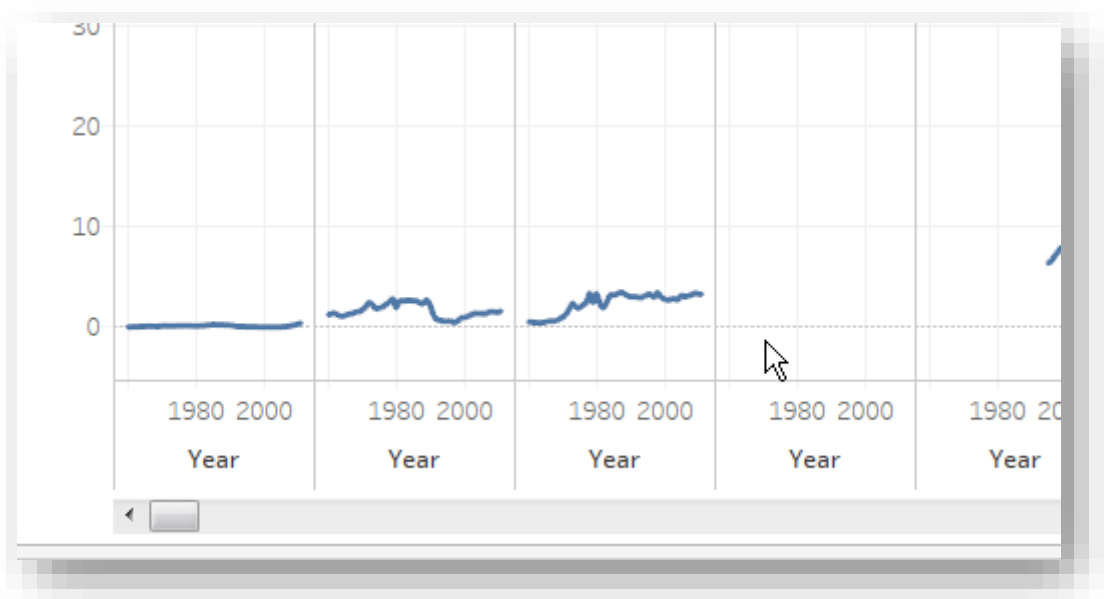
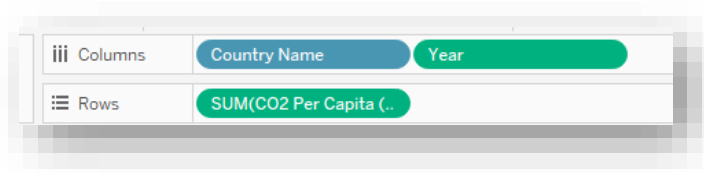


2. What would happen if we add **Country Name** to the **columns** section by <double clicking> it?

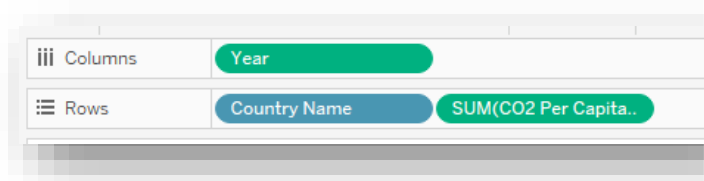


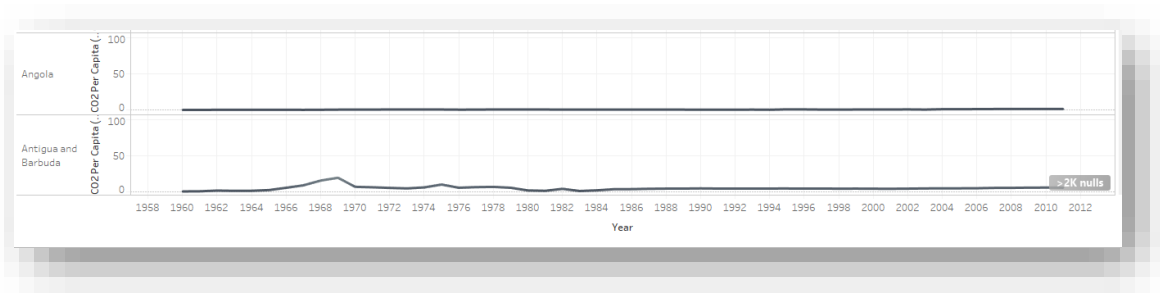


3. It is interesting, isn't it? Now, let's add <double click> **Year** as well.

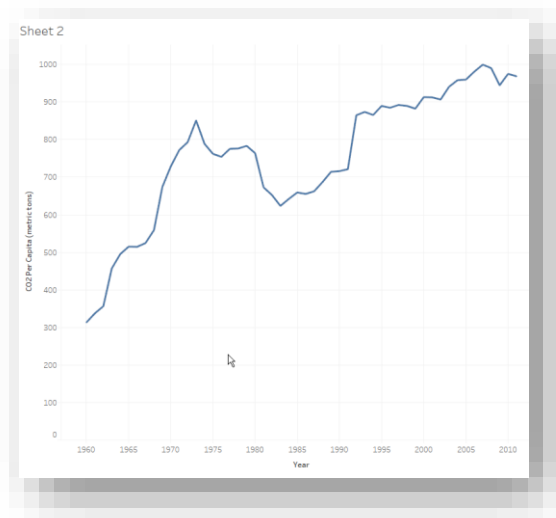


4. Cool! It adds more data to the columns. Try to move the **Country Name** to **Rows** shelf and see how the visualisation changes. Which do you think is better?





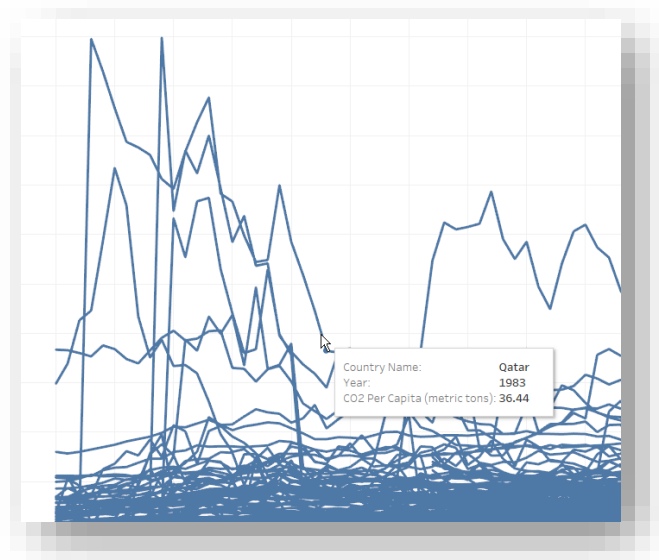
- Adding **Country Name** to Rows and Columns makes the visualisation entirely hard to see because of the enormous number of countries. That isn't what we wanted! Let's remove that by selecting **Country Name** and pressing delete.



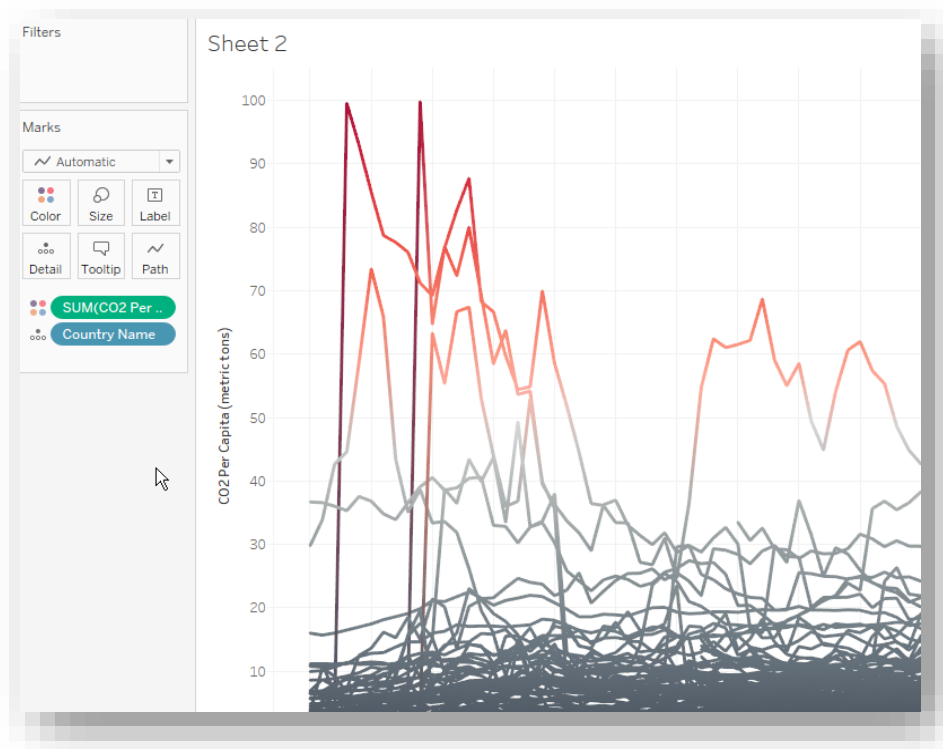
Question

The line chart shows the total CO2 per capita of the entire world over time. Can you change it such a way that it shows the average?

- The line chart now shows the **global total emission per year**, which might be useful on its own. Suppose we want to show the country data overlapping, how do we do that? We can just drag **Country Name** into the main window!



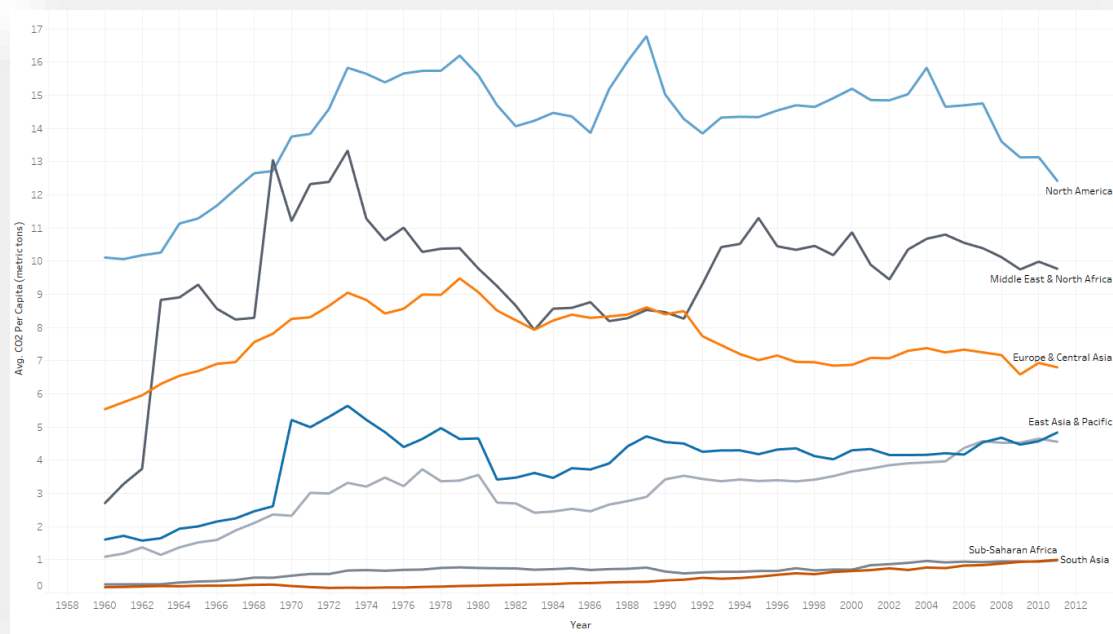
7. Now we have a visualisation showing overlapping lines representing each country. In our final visualisation, we will show this together with the map we did earlier. It would therefore be a good idea to have the colours matching (otherwise, it will be confusing!). Use the same process that we did before in adding a colour *Mark* for the **Country Name**.



8. Our last step is to name this visualisation “World CO2 Emission – Line Chart”

2.4. World Regional CO2 Emission – Line Chart

1. It's time to apply what we have learned from the previous activity to create a similar visualisation. This time, you need to create a line chart of the average CO2 emission per capita of different **regions** of the world, as shown in the sample below. The colour palette used in this line chart is the **Colour Blind** palette.

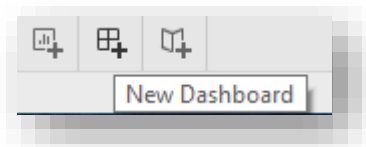


2. Once you are satisfied, rename the sheet to “World Regional CO2 Emission – Line Chart”.

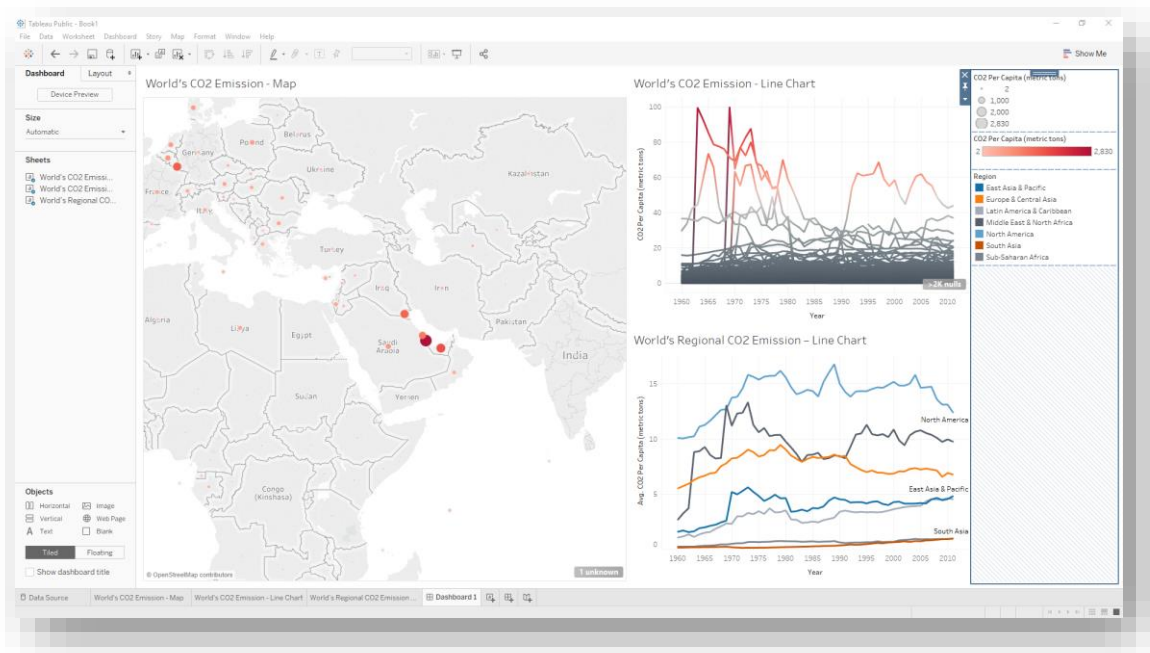
3. Creating a Dashboard and Combining Visualisations in Tableau

3.1. Combining Your Visualisations Together

1. As the last step, we will see how to combine all of the visualisation / sheets that we have created into a single interactive visualisation *dashboard*.
2. First, create new dashboard.



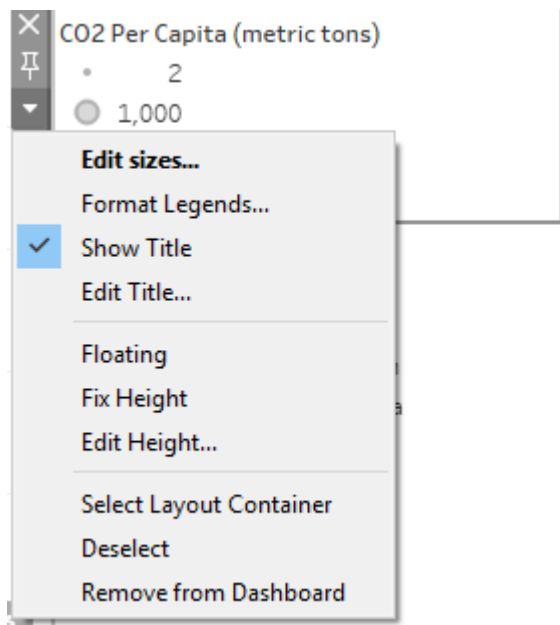
3. In the Dashboard panel, you will see **Size**, **Sheets**, and **Objects**. You can adjust the size of the dashboard in the **Size** menu. Select **automatic** in the **Size** menu so that the dashboard will fit in any device size. **Sheets** show all the sheets you have made. **Objects** show different object we can put into dashboard. We will use it later.
4. We can now drag the visualisations from **Sheets** into our dashboard. Choose any layout that you think the best for all three visualisations. One of the possible layouts is shown below.



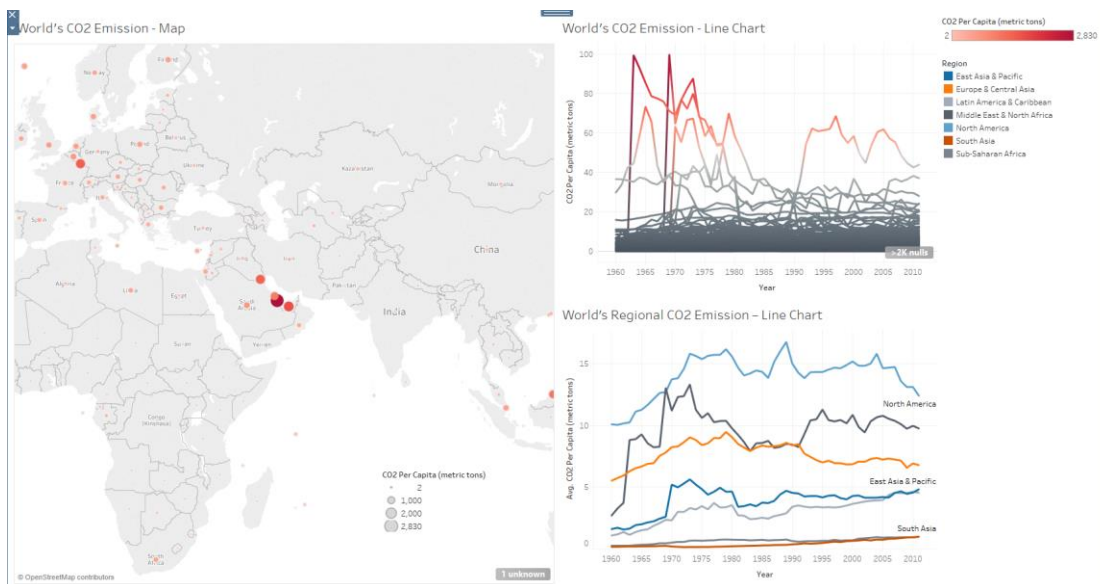
5. This looks nice. However, the legend of the map is too far from it! The viewers might need to switch focus when they want to estimate the actual CO2 per capita emission from a circle.

To minimise focus switch, it is a good idea to move the legend nearer or on the map.

6. To do that, we need to make the legend “float”. Go to **More Options → Floating**.



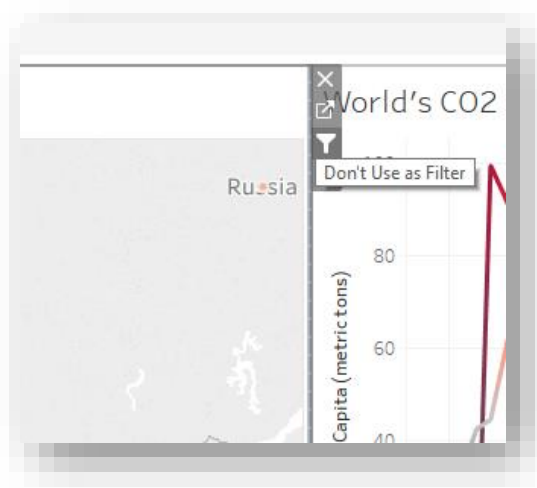
7. Now that the legend is detached from the layout, position it on the **bottom right corner** of the map.



8. This is nicer, but the visualisations still do not affect each other. The **World CO2 Emission – Line Chart** is very cluttered. This is where **interactive filtering** will come into play.

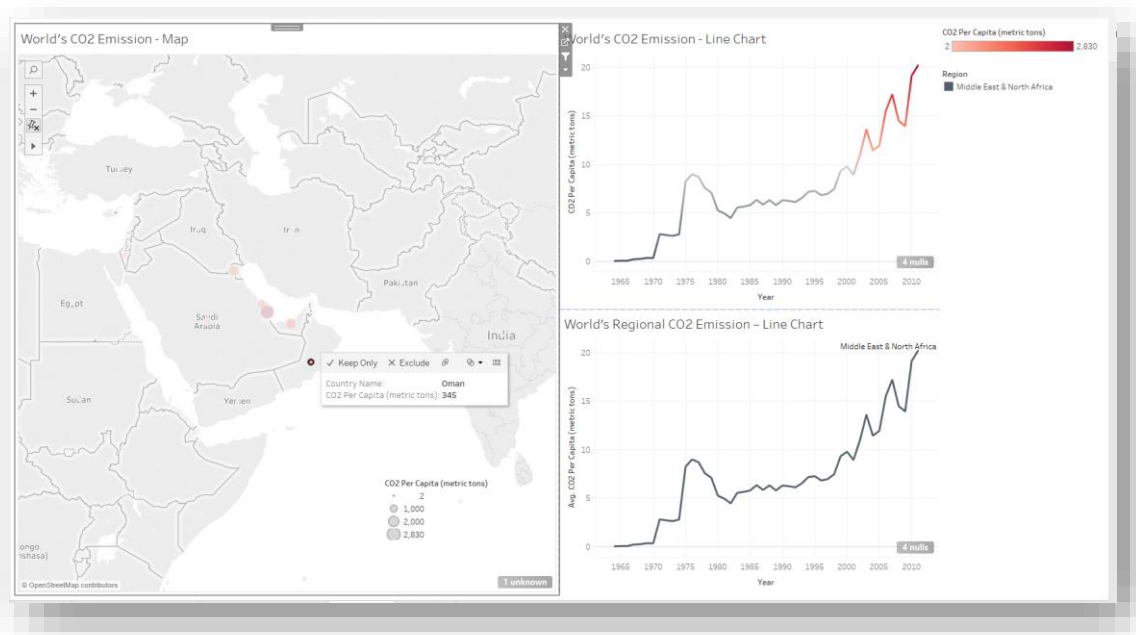
Let's make the dashboard more interactive. We want to use the map as a filter for the cluttered line chart.

To do that, click the **World CO2 Emission – Map**. Click on **use as filter** icon.

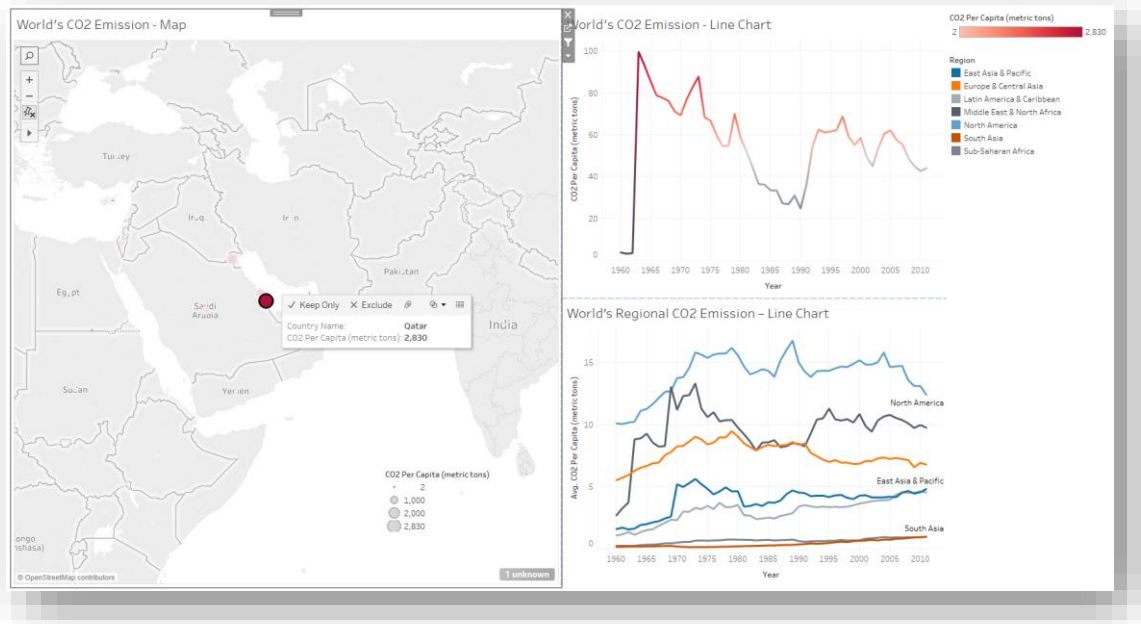


9. Now you have a **brushing and linking interaction** (this concept will be explained later in the interactive visualisation topic)! The line chart is now filtered out when you click on a country! However, it is also affecting the regional line chart (**World Regional CO2 Emission**). This could be something you want or something you didn't expect to happen.

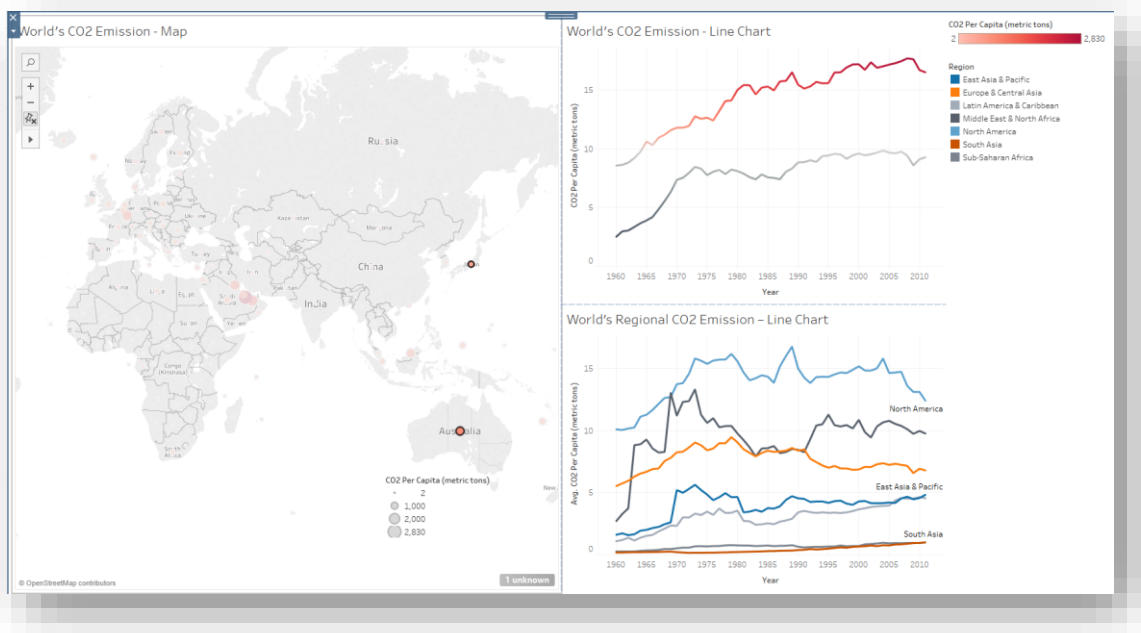
Ideally, you don't need to filter an uncluttered visualisation because it can be easily understood. Therefore, we need to modify the regional line chart so it does not react to the map click.



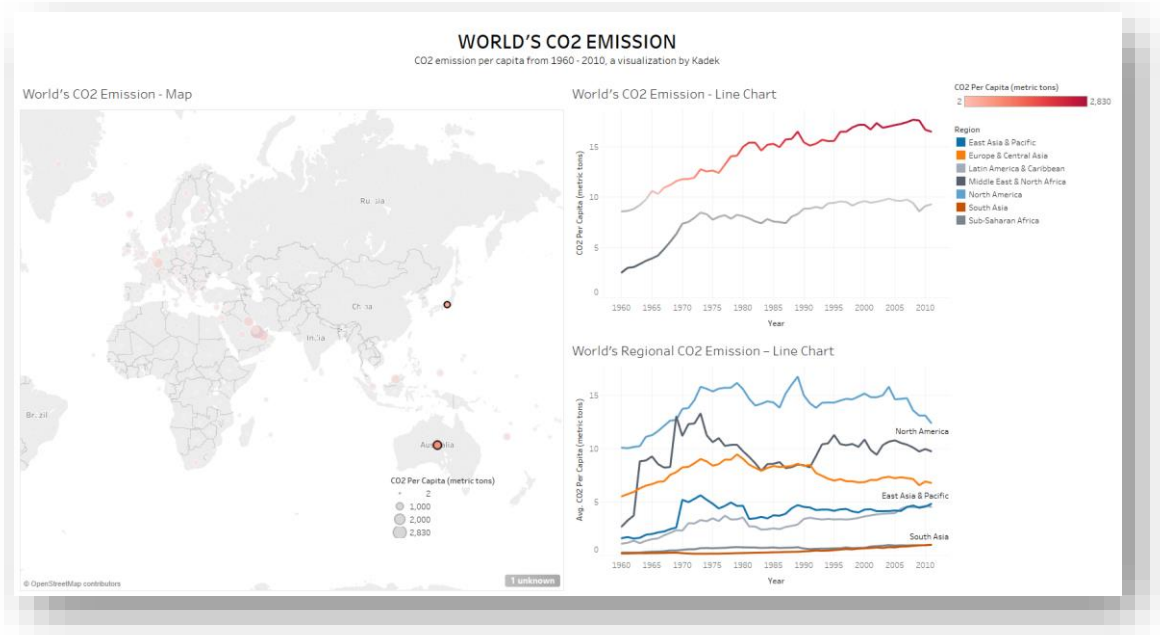
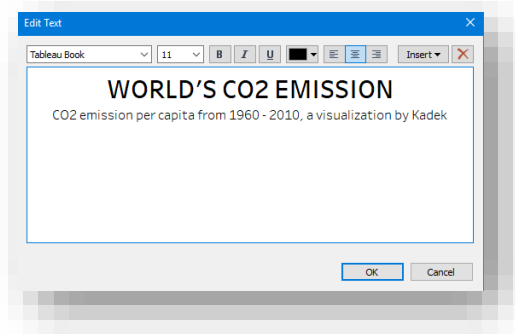
To do that, bring the regional line chart back to its previous form by **clicking an empty space** on the map. Then, click the **World's Regional CO2 Emission - Line Chart, More Options → Ignore Actions**. Now the filter will not affect it anymore!



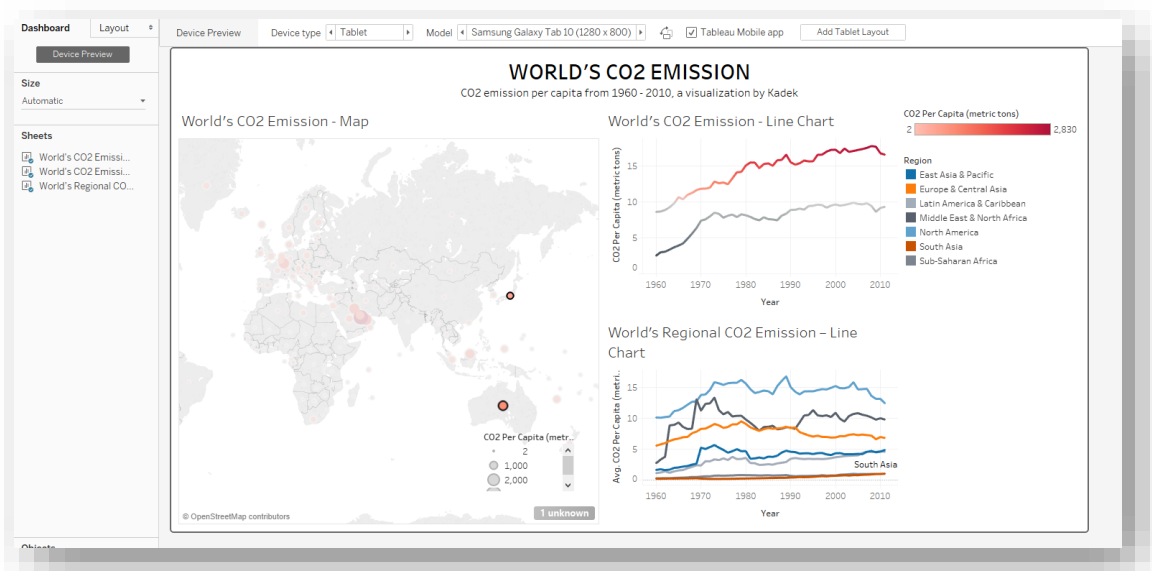
10.If you want to compare two countries, we just need to **<Ctrl> + click** them both. For example, **Japan** and **Australia** are selected in the image bellow.



11.A good visualisation always has a title. Let's add some text to the visualisation by dragging **Text** from **Object** panel.



12. We are finally done! Before saving it, let's see how the dashboard looks on a different device. Click **Device Preview** button on **Dashboard** panel. This is how it looks on Samsung Galaxy Tab 10.



Question

Once you have created a visualisation following the instructions above, can you create extra components?

*For instance, how can you improve World's CO2 Emission – Line Chart to be more understandable?
Or, how to improve regional emission chart and legend to reduce focus switching?*

Also consider what other visualisation sheets you can create, how you can combine them together to describe a narrative and what kind of questions you are seeking to answer.

Lastly, try to use data visualisation theories to critic the visualisation you just created and think about how to improve it.

4. Working with Multiple Data Sources

So far, you have learned to create visualisation based on an existing dataset. In some cases, you might need to start designing your visualisation without thinking about whether or not the dataset exists. To bring your design into realisation, you may need to hunt for datasets on the Internet. Stumbling upon dirty datasets which require cleaning is common. You might need to combine pieces of dataset into a single data that can satisfy your visualisation requirements.

In this tutorial, we will learn the basics of working with multiple data sources. Tableau provides a feature to join multiple data sources together.

For this exercise, we need a fresh new workbook. Save your previous work and create a new workbook by clicking **File -> New**.

4.1. Joining Datasets via a Single Field

For this activity, we will use two hypothetical datasets to help you recall different join types: **Inner Join**, **Left Join**, **Right Join**, **Full Outer Join**. Remember your Databases unit?

1. Download **camera_store_sales.xls** from the Moodle page.
2. Import the Excel file you just downloaded. There are three sheets in the dataset: **customer_data**, **transaction_detail**, and **transaction**.
3. Drag and drop the **transaction** sheet. Then, drag and drop the **customer_data** sheet.
4. Because both sheets have a common column, **Customer Id**, the **inner join** is automatically performed.

transactions+ (camera_store_sales)

Filters
0 | Add

transactions — customer_data

Sort fields Data source order ☐ Show aliases ☐ Show hidden fields 3 rows

#	customer_data	customer_data	transactions	transactions	transactions
	Customer Id (Cust...	Address	Store	Trans Id	Customer Id
	1001	EAST MELBOURNE, VIC	STORE1	1	1001
	1005	MELBOURNE, VIC	STORE1	2	1005
	2005	PORT MELBOURNE, VIC	STORE2	1	2005

5. If you click on the circle, you can see **Inner Join** is selected. You can also see that the Data Source (**transactions**) and **customer_data** are linked by the **Customer Id** field. The result of the join is immediately shown on the table below. **Inner Join** will return a set of rows that have common values on both datasets, as shown on the **Venn diagram**. In this case, only customer id **1001**, **1005**, and **2005** are present in **customer_data** sheet.

transactions — customer_data

Join

Inner Left Right Full Outer

Data Source customer_data

Customer Id = Customer Id (Customer!...

Add new join clause

Sort fields Data source order ☐ Show aliases ☐ Show hidden fields 3 rows

#	customer_data	customer_data	transactions	transactions	transactions
	Customer Id (Cust...	Address	Store	Trans Id	Customer Id
	1001	EAST MELBOURNE, VIC	STORE1	1	1001
	1005	MELBOURNE, VIC	STORE1	2	1005
	2005	PORT MELBOURNE, VIC	STORE2	1	2005

Question

Now, switch to another type of join by clicking the join icon. Inspect the resulting table. Make sure you understand the characteristics of each join!

What are the characteristics of the **Left Join**? **Right Join**? **Full Outer Join**?

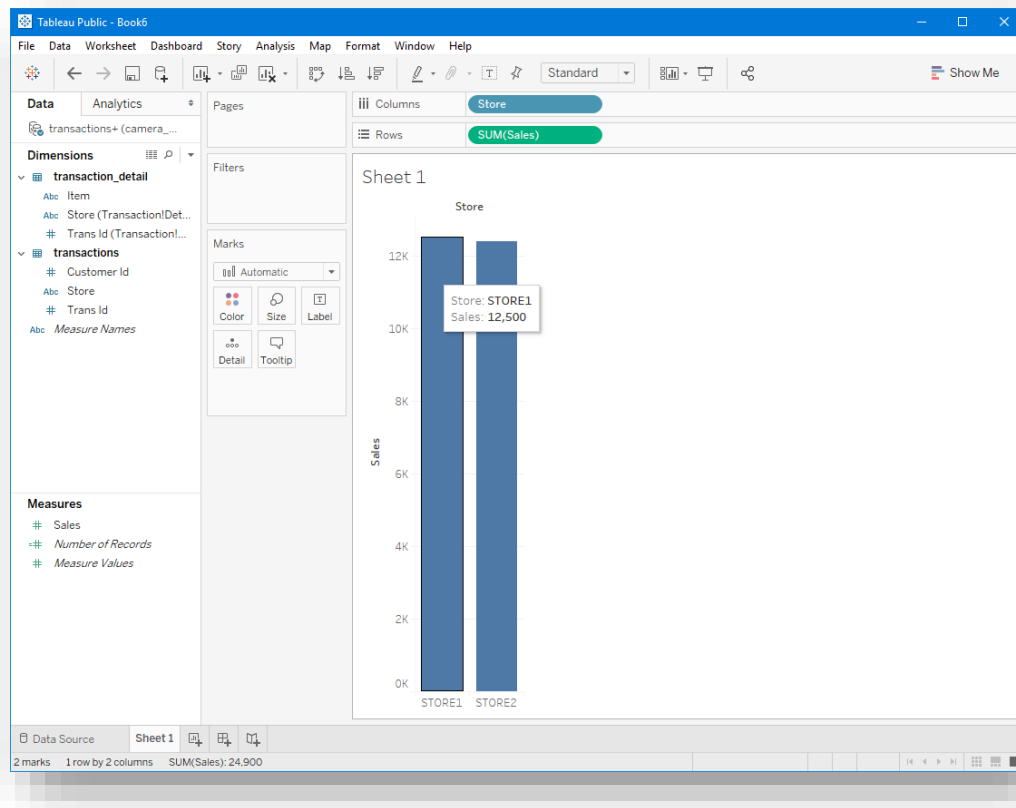
4.2. Joining Datasets via Multiple Fields

1. Once you understand how **joins** work, drop **customer_data** and put in **transaction_detail**.

The screenshot shows the Tableau Public interface with a data source named 'camera_store_sales' containing three sheets: 'customer_data', 'transaction_detail', and 'transactions'. A join dialog box is open, showing an Inner Join between 'transactions' and 'transaction_detail' on the 'Trans Id' field. The resulting data table is displayed below the join configuration.

transaction_detail Store (TransactionID)	transaction_detail Trans Id (TransactionID)	transaction_detail Item	transaction_detail Sales	transactions Store	transactions Trans Id	transactions Customer Id
STORE1	1	CAMERA	1,000	STORE2	1	2005
STORE1	1	CAMERA	1,000	STORE1	1	1001
STORE1	1	LENS KIT	1,200	STORE2	1	2005
STORE1	1	LENS KIT	1,200	STORE1	1	1001
STORE1	2	CAMERA	600	STORE2	2	2120
STORE1	2	CAMERA	600	STORE1	2	1005
STORE1	3	TRIPOD	100	STORE1	3	1023
STORE2	1	CAMERA	600	STORE2	1	2005
STORE2	1	CAMERA	600	STORE1	1	1001
STORE2	2	CAMERA	9,000	STORE2	2	2120
STORE2	2	CAMERA	9,000	STORE1	2	1005

2. Now, let's create a new sheet and visualise the sales of each store! Do you still remember how to do this?



- The bar chart below shows that the total sales of STORE1 is 12,500. However, if we check the data, STORE1 did not produce that much sales.

store	trans_id	item	sales
STORE1		1 CAMERA	1000
STORE1		1 LENS KIT	1200
STORE1		2 CAMERA	600
STORE1		3 TRIPOD	100
STORE2		1 CAMERA	600
STORE2		2 CAMERA	9000

- Let's go back to the data source and figure out what's wrong with it. Turns out there are many unnecessary repeating rows! This happened because the join condition only considers **trans_id**. In fact, both STORE1 and STORE2 have the same id for different sales. To fix it, we need to add another join condition.
- Click the **Inner Join** icon and add a new **condition**. In this case, we want to have **store** as another key.

transactions+ (camera_store_sales)

Filters 0 | Add

transactions — transaction_detail

Join

Inner Left Right Full Outer

Data Source		transaction_detail
Trans Id	=	Trans Id (Transaction!...
Store	=	Store (Transaction!De...

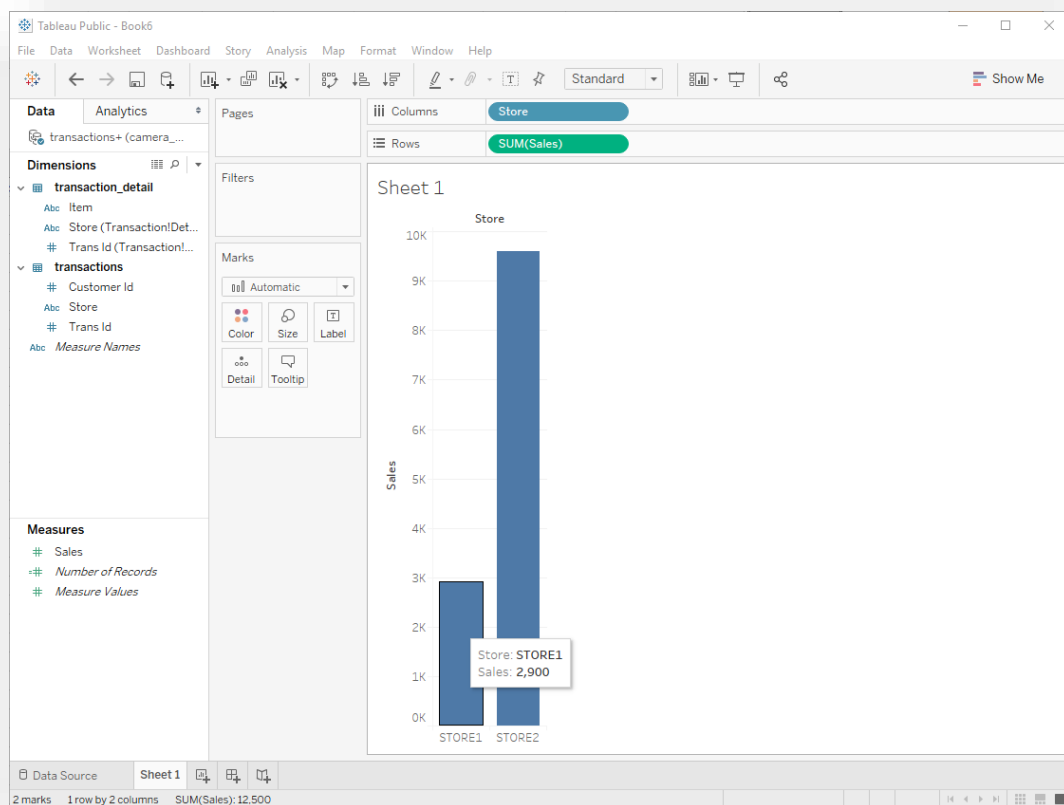
Add new join clause

Sort fields Data source order

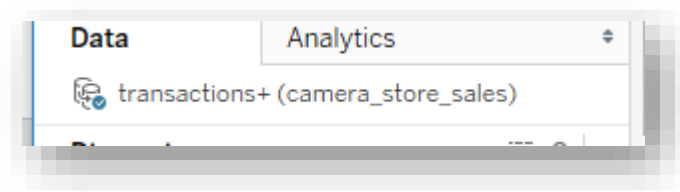
Show aliases Show hidden fields 6 rows

transaction/detail Store (Transactio...	# transaction/detail Trans Id (Transact...	transaction/detail Item	# transaction/d... Sales	transactions Store	# transactions Trans Id	# transactions Customer Id
STORE1	1	CAMERA	1,000	STORE1	1	1001
STORE1	1	LENS KIT	1,200	STORE1	1	1001
STORE1	2	CAMERA	600	STORE1	2	1005
STORE1	3	TRIPOD	100	STORE1	3	1023
STORE2	1	CAMERA	600	STORE2	1	2005
STORE2	2	CAMERA	9,000	STORE2	2	2120

6. This looks much better now. If you come back to Sheet1, you can see that the bar chart has been updated.



7. Joining datasets will produce a single data source. If you look at the **Data** tab, you can see there is only one data source named **transaction + (camera_store_sales)**. In the next activity, you will use multiple data sources without explicitly joining them together.

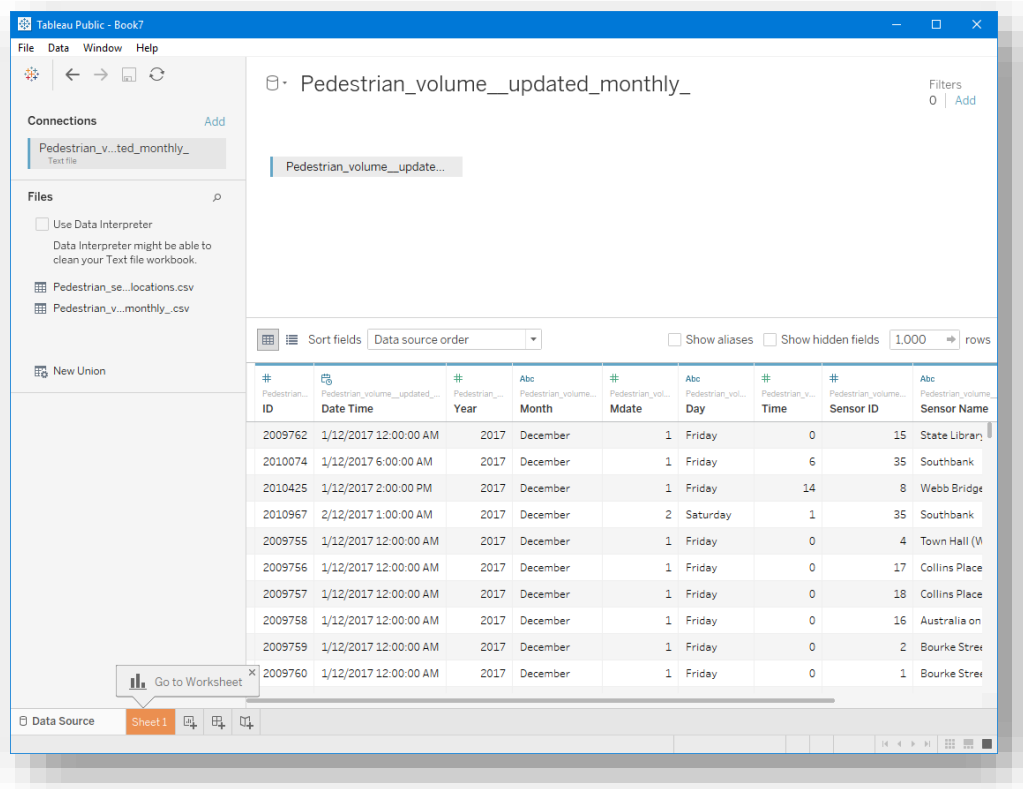


5. Blending Data Sources

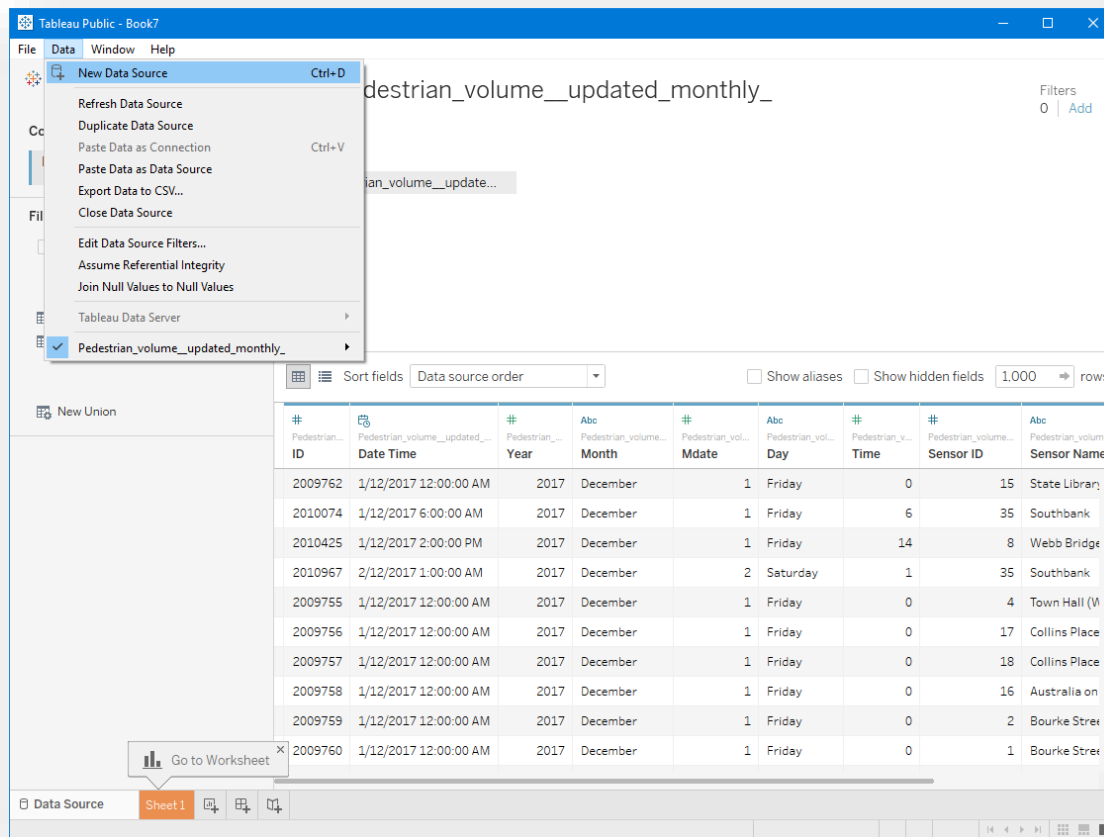
Blending means joining datasets on the fly. Instead of creating a single table that contains all the data we need, we can create links between data sources on demand. For this activity, we will use a different dataset.

Save the previous workbook and create a new one.

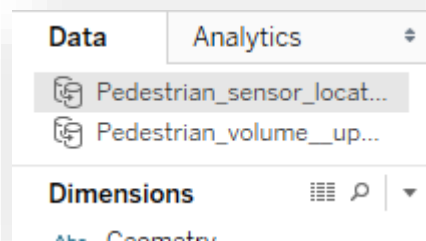
1. Download **Pedestrian_sensor_locations.csv** and **Pedestrian_volume__updated_monthly_.csv**
2. Open Tableau and import **Pedestrian_volume_update_monthly.csv**



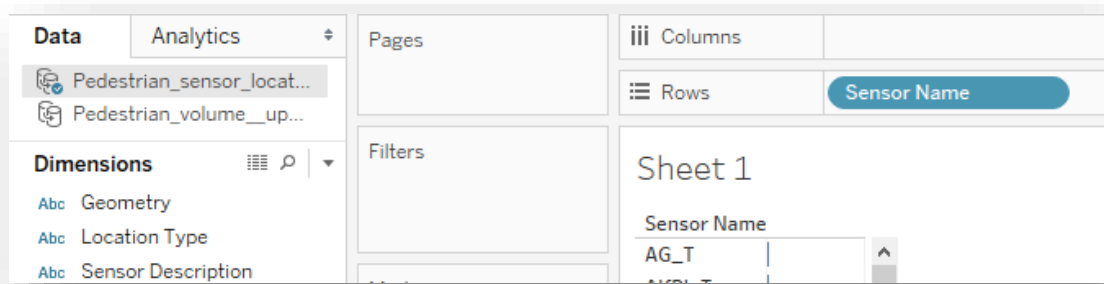
3. Add a new data source and select **Pedestrian_sensor_locations.csv**



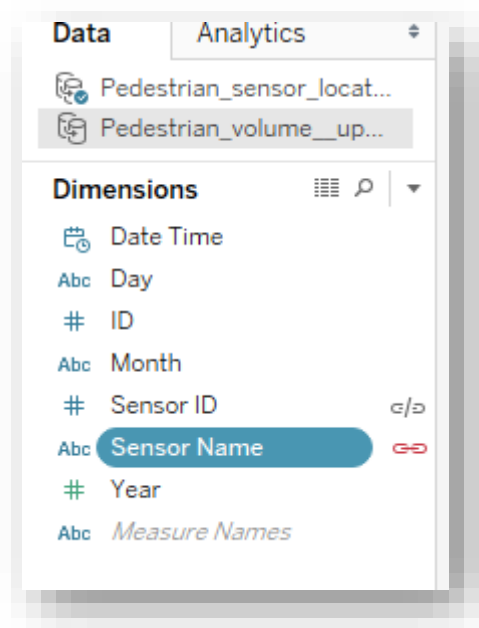
- Go to the **Sheet 1**. You will see two data sources on the **Data** tab. Both data sources have the same icon because we have not yet created any visualisation.



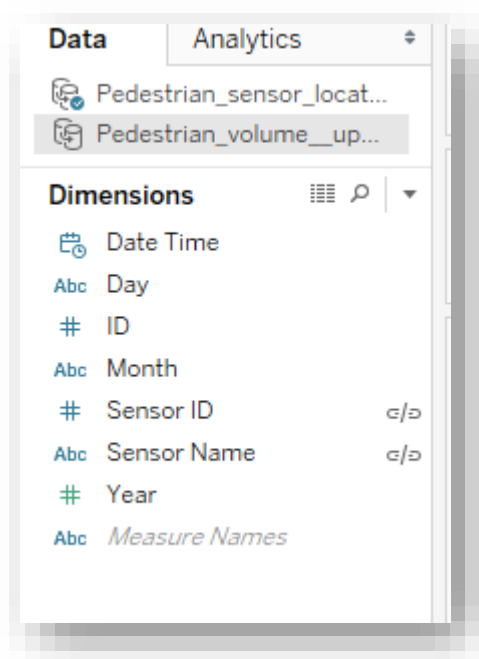
- The main data source is determined by the first field that you put into the visualisation. To illustrate this concept, let's drop **Sensor Name** from the first data source.



6. As you can see, the icon of the first data source has changed. Now, click on the second dataset. You will notice the link icons on both **Sensor ID** and **Sensor Name** dimensions. In this case, the **Sensor Name** link is active because Tableau automatically detected the same field name from the main data source (the first one).

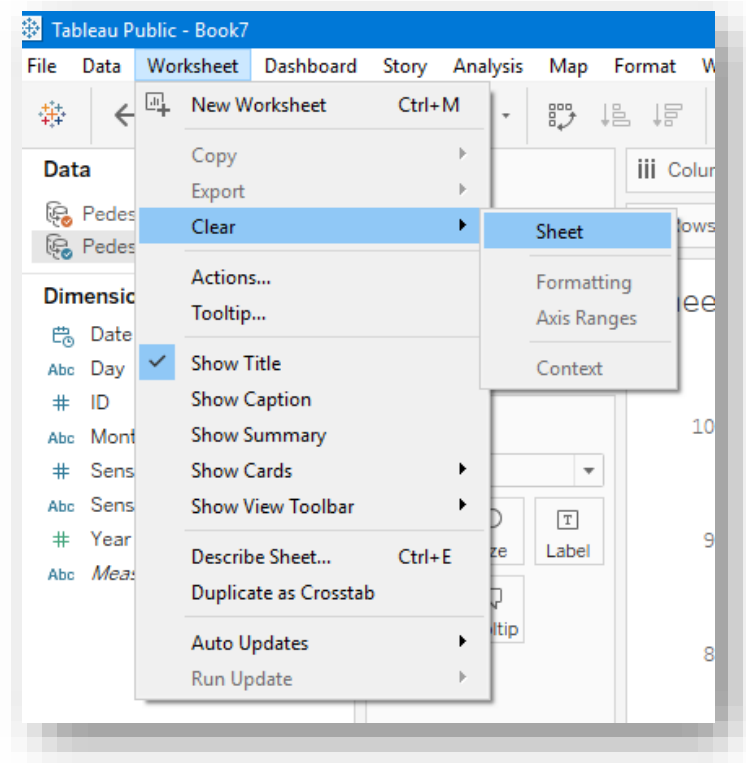


7. Now, remove **Sensor Name** field from the **Rows** and add **Location Type** from the first data source.
8. After that, click the second data source. The **Sensor Name** link is no longer active because Tableau cannot determine to which field **Location Type** is connected.

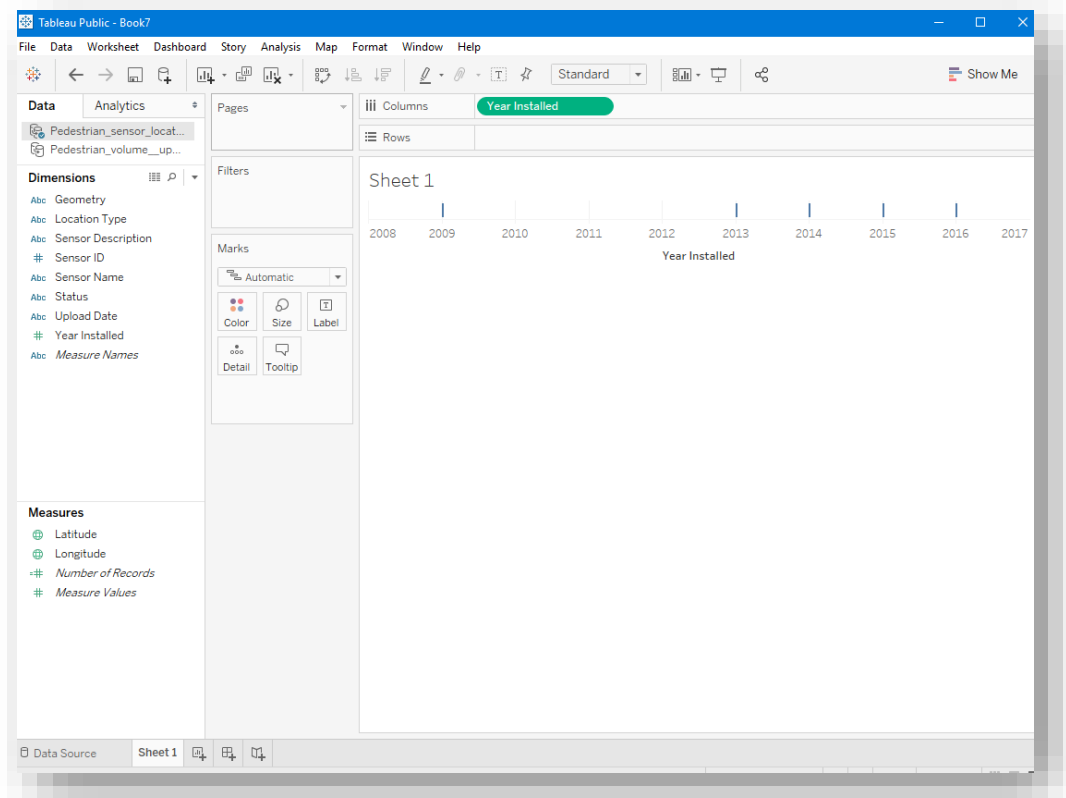


9. If you want to make the second data source as the main data source, **clear** your **Sheet** and drop a field from the second data source first.

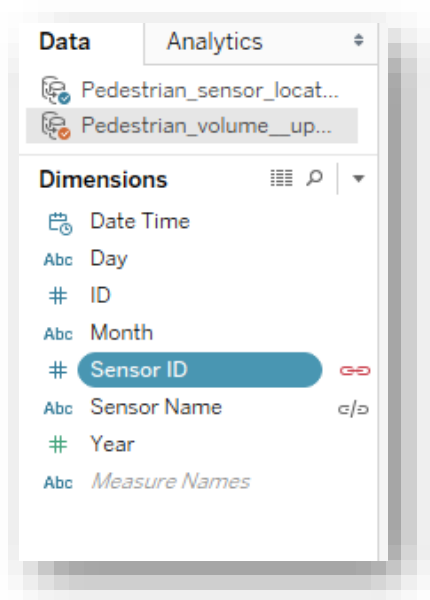
10. Let's create a simple bar chart using these two data sources. First, clear your Sheet by clicking **Worksheet → Clear → Sheet**



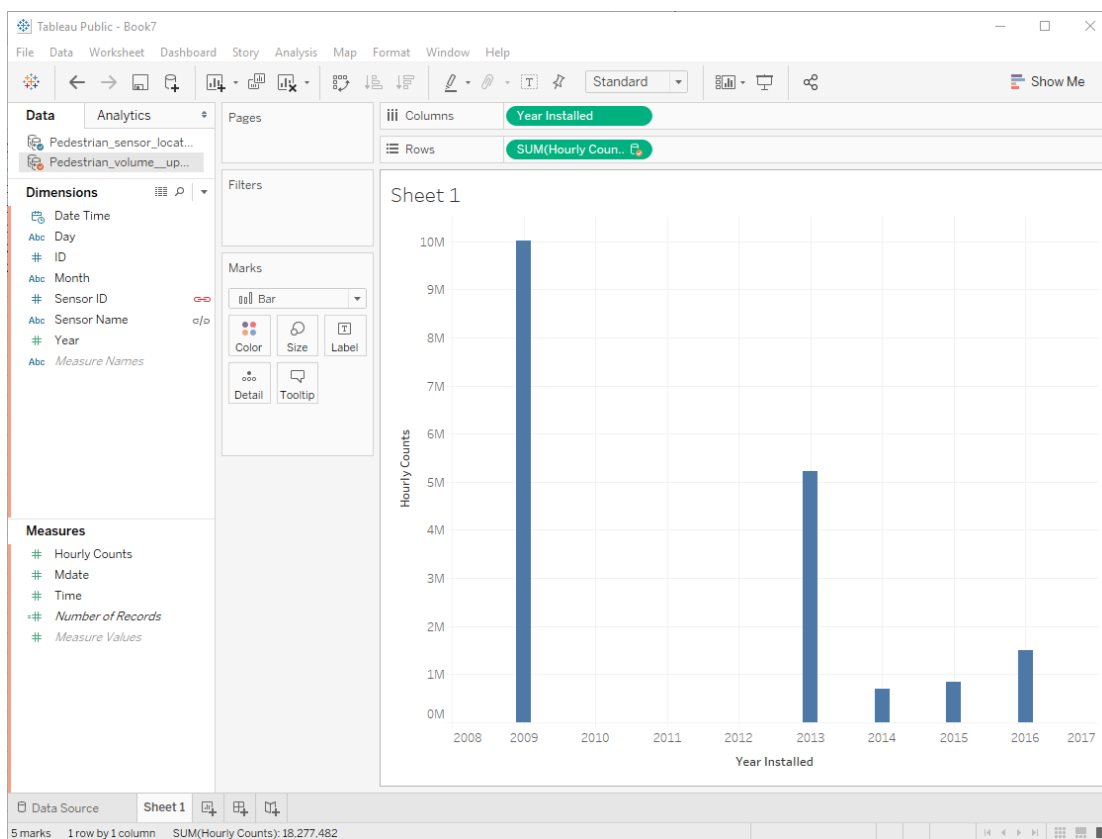
11. Suppose we want to know the **number of pedestrians** counted by the sensor on the first year of instalment. Drop the **Year Installed** dimension from the first data source to Columns.



12. Click the second data source and activate the **Sensor ID** link.



13. Drop the **Hourly Counts** measure to Rows and set the **Marks** to **Bar**.



14. Here you go. Now you have a simple bar chart showing the **total pedestrian count** on the first year the sensor was installed! This visualisation does not tell you much; however, the focus of this activity is to introduce Data Source Joining and Blending so that you can use it later in your fascinating visualisation project.

Question

Look at the Data Sources tab; can you recognise which one is the main source based on the icon?