Introduction to Business Analytics

Lecture 5: Visualization in Tableau

Iegor Vyshnevskyi

Woosong University

March 28 / April 3, 2023

Agenda

1. Introduction to Data Visualization
2. Importance of Data Visualization
3. Chart Types
4. Visual Cues
5. Introduction to Tableau
6. Loading the Table
7. First Chart
8. Filtering
9. Dashboards
10. More filter examples
11. In-class assignment

1. Introduction to Data Visualization

What is Data Visualization

- *Data visualization* is the process of representing data and information in a graphical format.

- The <u>goal of data visualization</u> is to communicate insights and patterns in a more effective and meaningful way.

- Data visualization allows analysts, researchers, and decision-makers to easily understand complex data sets.

- Effective data visualization leverages design principles such as color, shape, and layout to make information more accessible and understandable.

- Data visualization enables users to quickly and efficiently gain insights from data.

2. Importance of Data Visualization

Effective data visualization:

- helps users to better understand patterns, trends, and relationships in data

- helps to identify outliers and anomalies in data that might be missed otherwise

- reveal hidden insights and relationships that are not immediately apparent in raw data

- helps to communicate findings and insights to stakeholders and decision-makers in a clear and compelling way

- helps to support data-driven decision-making by providing an intuitive and accessible view of data.

Poor data visualization can lead to:

- misleading interpretations and conclusions;
- oversimplification or obscuring of important details;
- confusion or misinterpretation of the data;
- biases or misrepresentations based on design choices;
- difficulty in visualizing certain types of data effectively;
- inaccessibility for users with visual impairments;
- incomplete or insufficient analysis due to a lack of context or nuance.

Four Elements of Good Data Visualization
under the David McCandless method

Source: www.informationisbeautiful.net
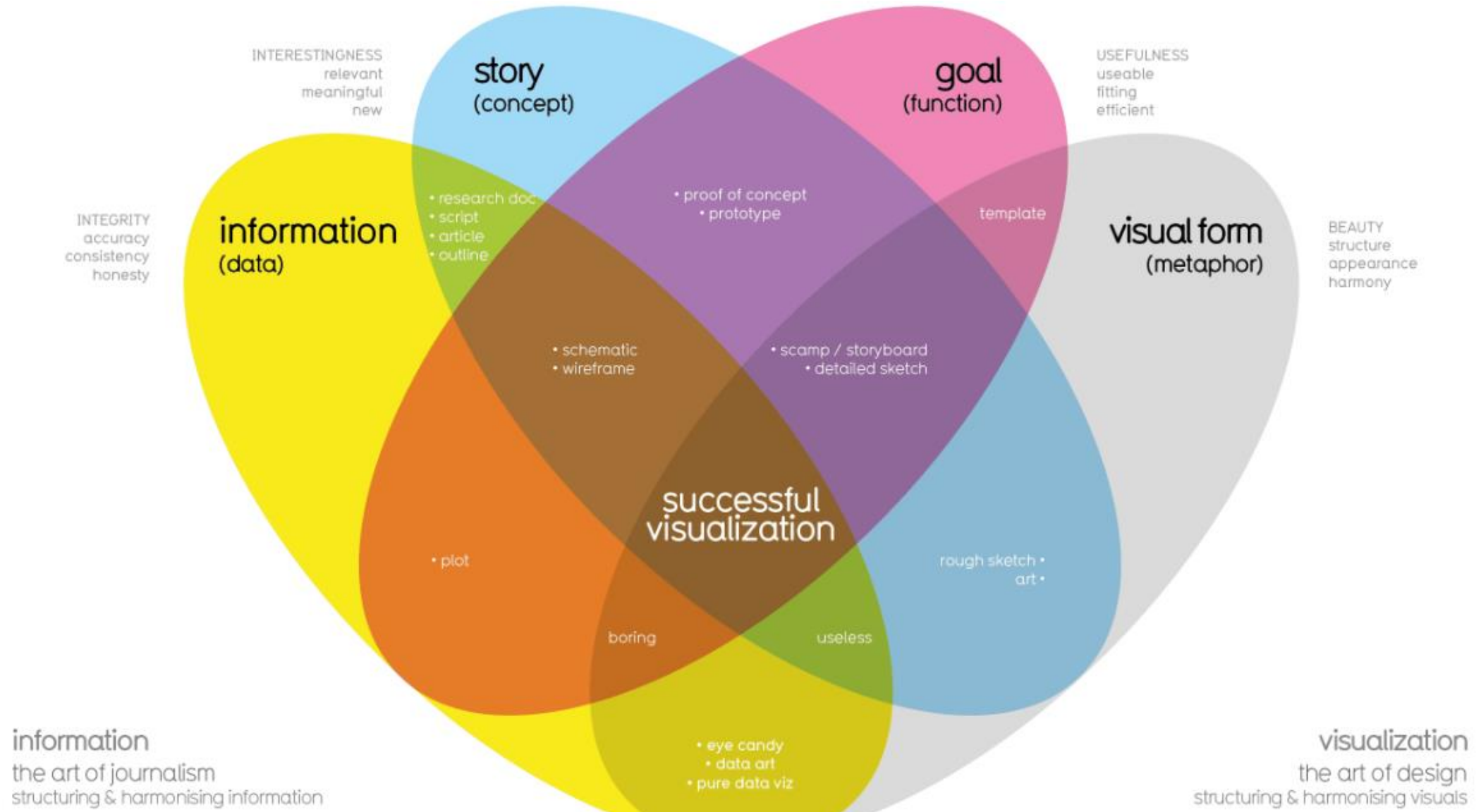
- **Information (data):** The information or data that you are trying to convey is a key building block for your data visualization. Without information or data, you cannot communicate your findings successfully.

- **Story (concept):** Story allows you to share your data in meaningful and interesting ways. Without a story, your visualization is informative, but not really inspiring.

- **Goal (function):** The goal of your data visualization makes the data useful and usable. This is what you are trying to achieve with your visualization. Without a goal, your visualization might still be informative, but can't generate actionable insights.

- **Visual form (metaphor):** The visual form element is what gives your data visualization structure and makes it beautiful. Without visual form, your data is not visualized yet.

Kaiser Fung's Junk Charts Trifecta Checkup

to estimate the effectiveness of data visualization
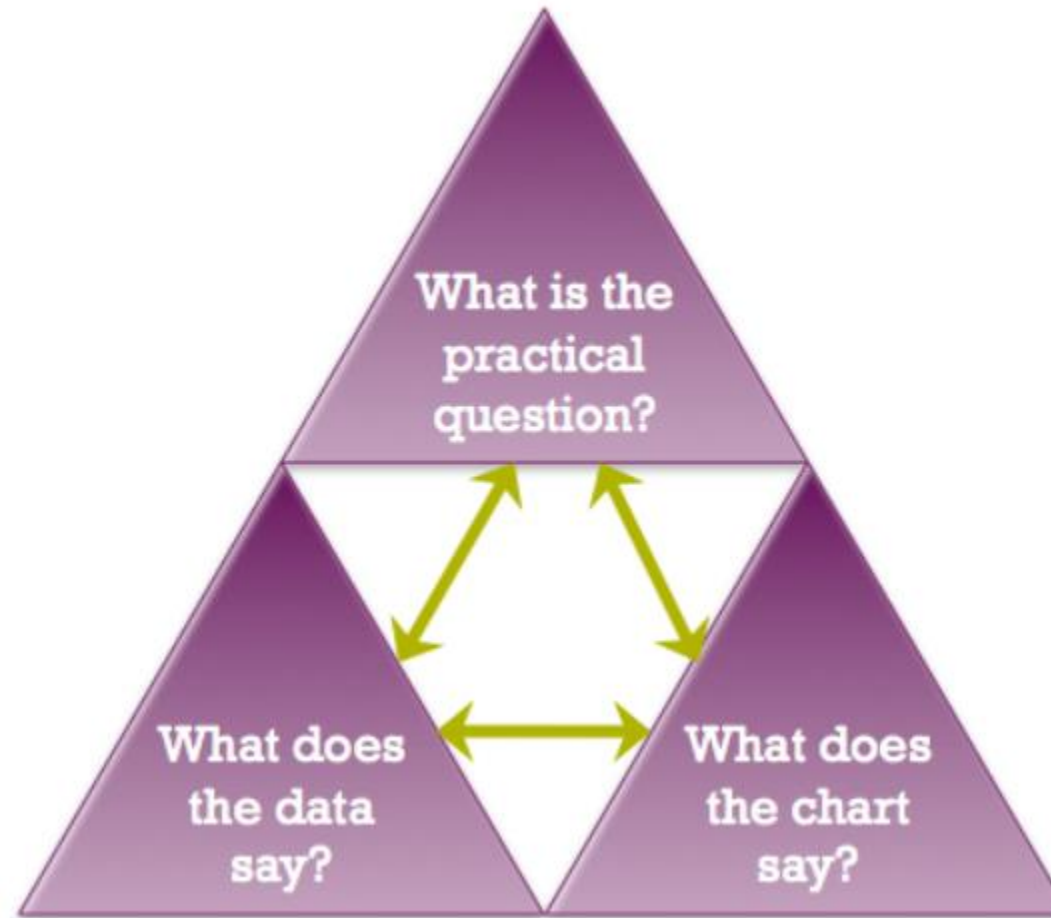
What to Avoid

Cutting off the y-axis	Changing the scale on the y-axis can make the differences between different groups in your data seem more dramatic, even if the difference is actually quite small.
Misleading use of a dual y-axis	Using a dual y-axis without clearly labeling it in your data visualization can create extremely misleading charts.
Artificially limiting the scope of the data	If you only consider the part of the data that confirms your analysis, your visualizations will be misleading because they don't take all of the data into account.
Problematic choices in how data is binned or grouped	It is important to make sure that the way you are grouping data isn't misleading or misrepresenting your data and disguising important trends and insights.
Using part-to-whole visuals when the totals do not sum up appropriately	If you are using a part-to-whole visual like a pie chart to explain your data, the individual parts should add up to equal 100%. If they don't, your data visualization will be misleading.
Hiding trends in cumulative charts	Creating a cumulative chart can disguise more insightful trends by making the scale of the visualization too large to track any changes over time.
Artificially smoothing trends	Adding smooth trend lines between points in a scatterplot can make it easier to read that plot, but replacing the points with just the line can actually make it appear that the point is more connected over time than it actually was.

Source: Coursera course "Share Data Through Art of Visualization"

3. Chart Types

- Different types of visualization are better suited to different types of data and communication goals

- Choosing the right visualization can help you communicate your insights more effectively and support decision-making.

The example of detail interactive decision tree to make decisions based on key questions that you can ask yourself I highly recommend: https://www.data-to-viz.com/

The Most Common Chart Types
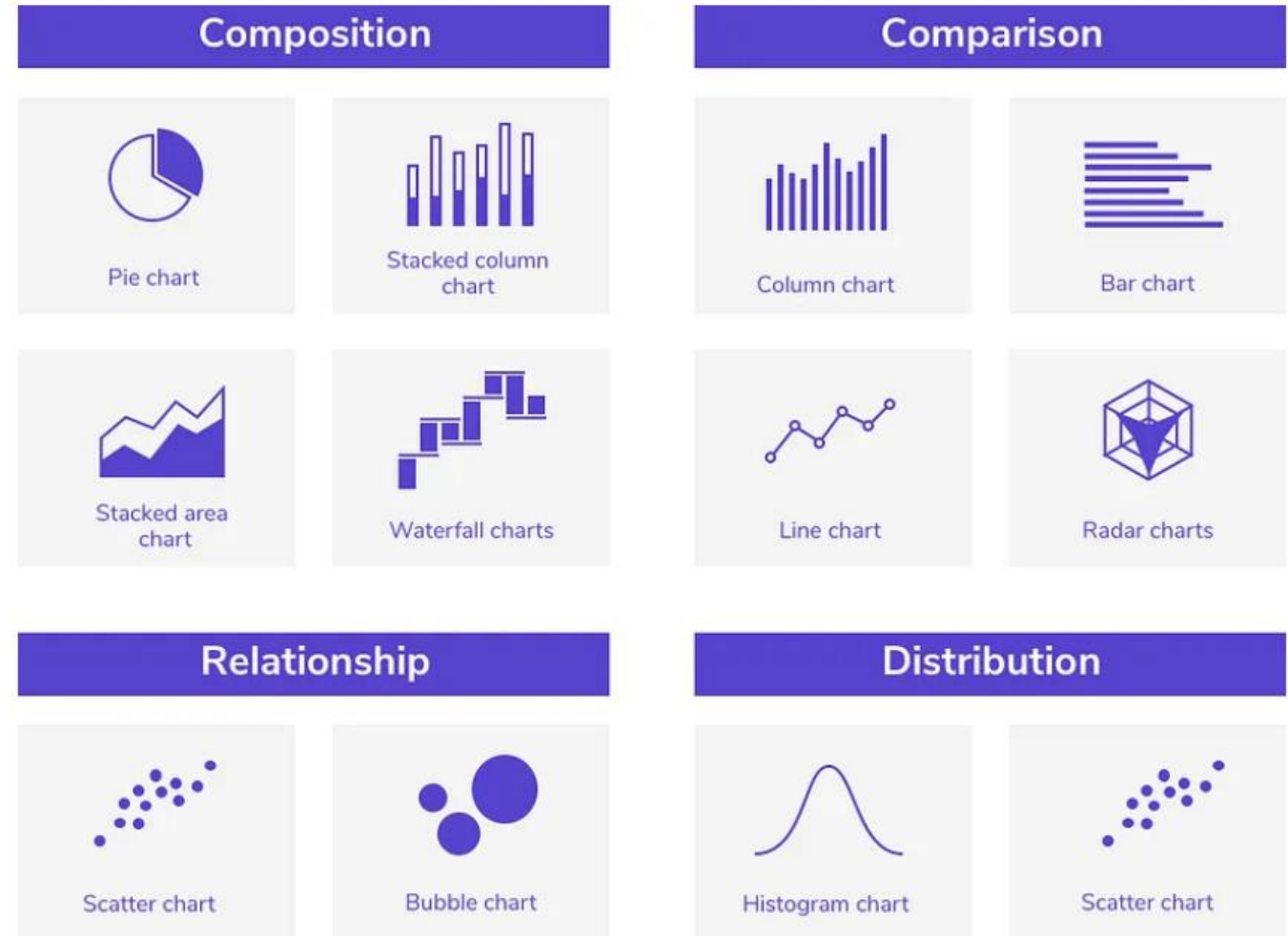
Source: https://uxplanet.org/data-heavy-applications-how-to-design-perfect-charts-c0c893fef6de

- Different types of visualization are better suited to different types of data and communication goals

- Choosing the right visualization can help you communicate your insights more effectively and support decision-making.

The example of detail interactive decision tree to make decisions based on key questions that you can ask yourself I highly recommend:
https://www.data-to-viz.com/

4. Visual Cues

If you want to invite your audience into your presentation and keep them engaged, you have only **5 seconds** to catch their interest.

They should be able to process and understand the information you are trying to share with this extremely short time frame.

Effective visual cues are highly valuable for this purpose.

Effective Visual Cues

- **Headlines that pop**

Average Rents in the Tri-City Area
Oceanside, Vista and Carlsbad

Source: Coursera course "Share Data Through Art of Visualization"
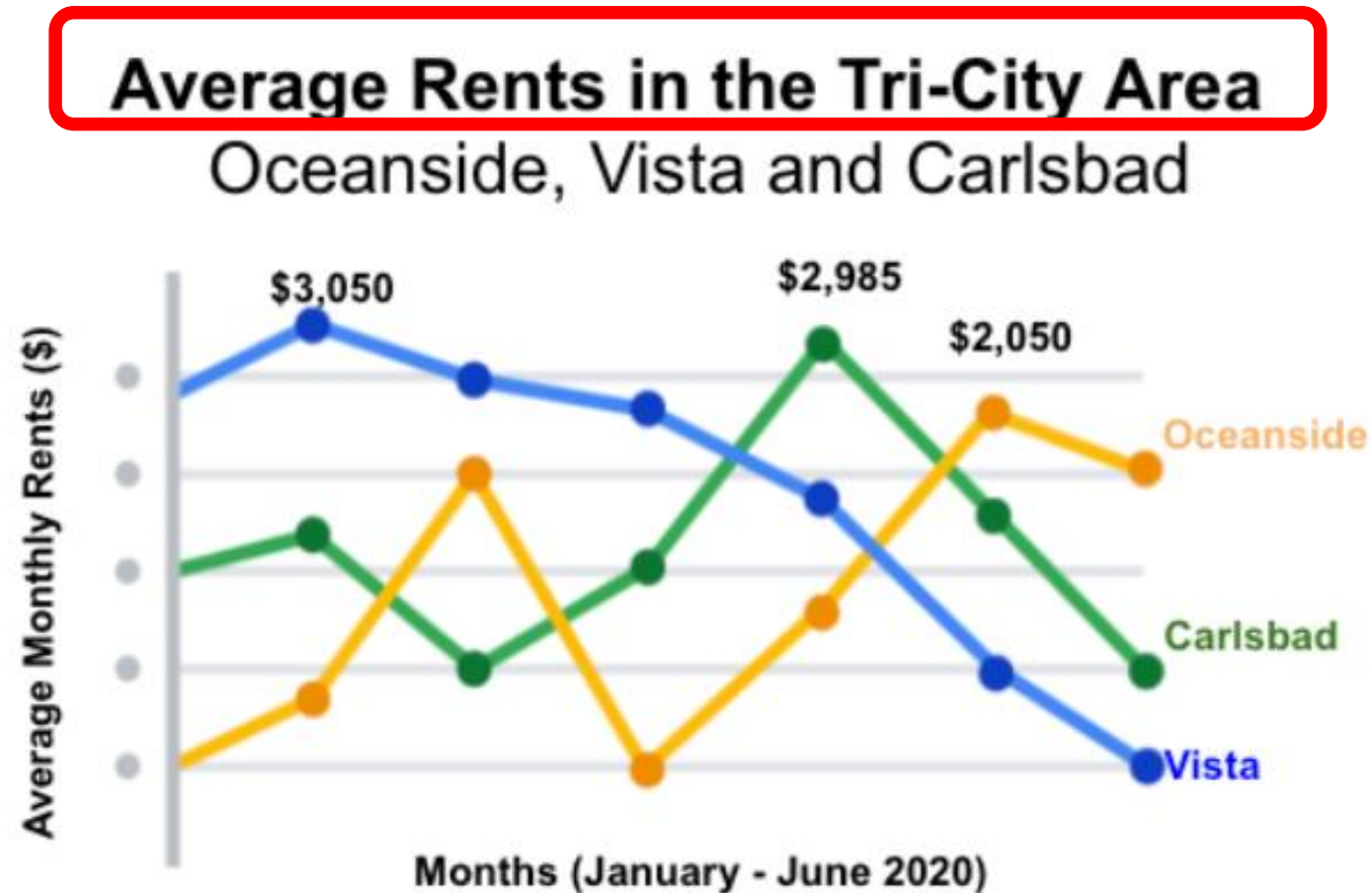
Effective Visual Cues

- **Subtitles that clarify**

Source: Coursera course "Share Data Through Art of Visualization"
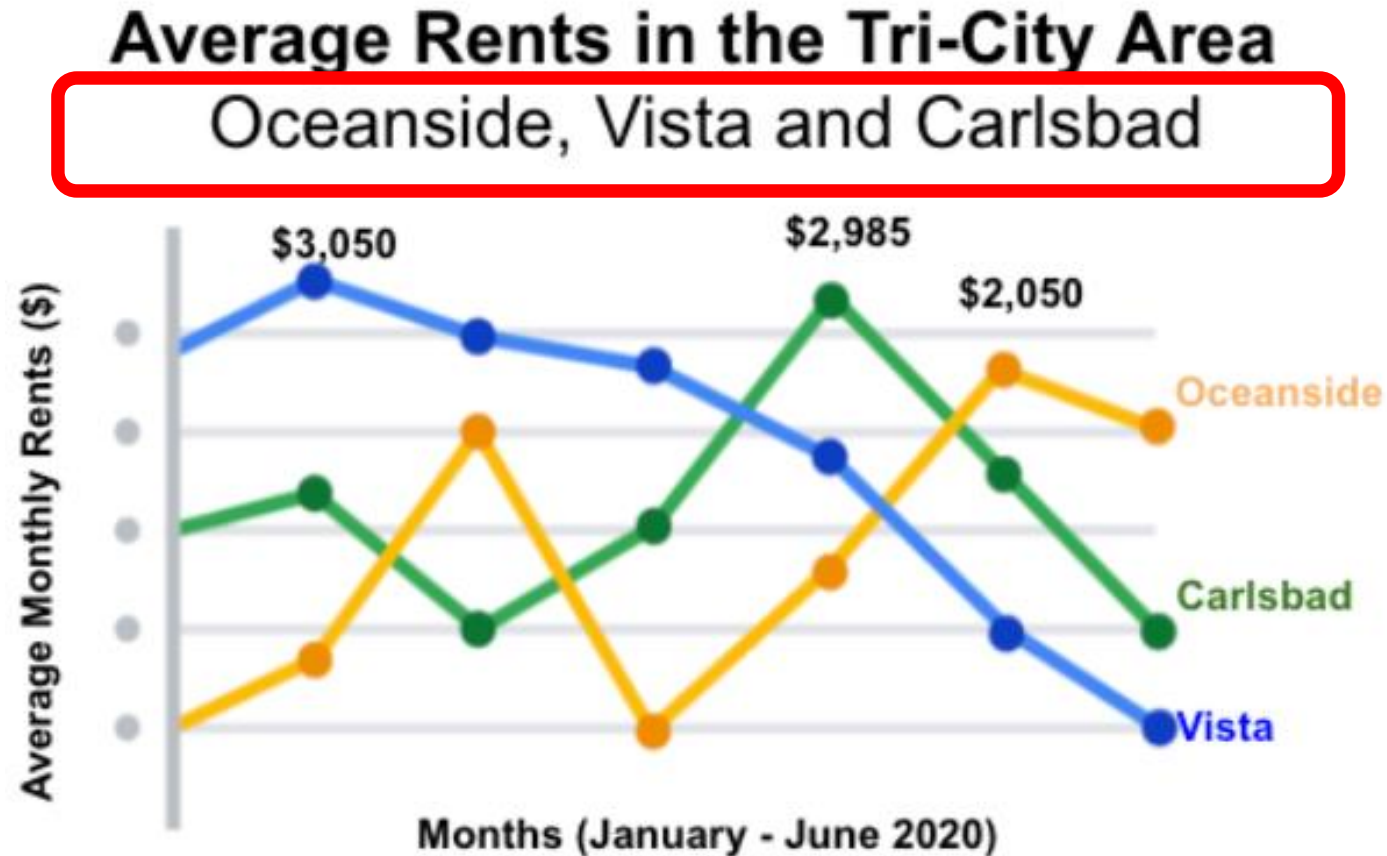
Effective Visual Cues

- **Labels that identify**

Average Rents in the Tri-City Area
Oceanside, Vista and Carlsbad

$3,050 $2,985

$2,050

Average Monthly Rents ($)

Oceanside

Carlsbad

Vista

Months (January - June 2020)
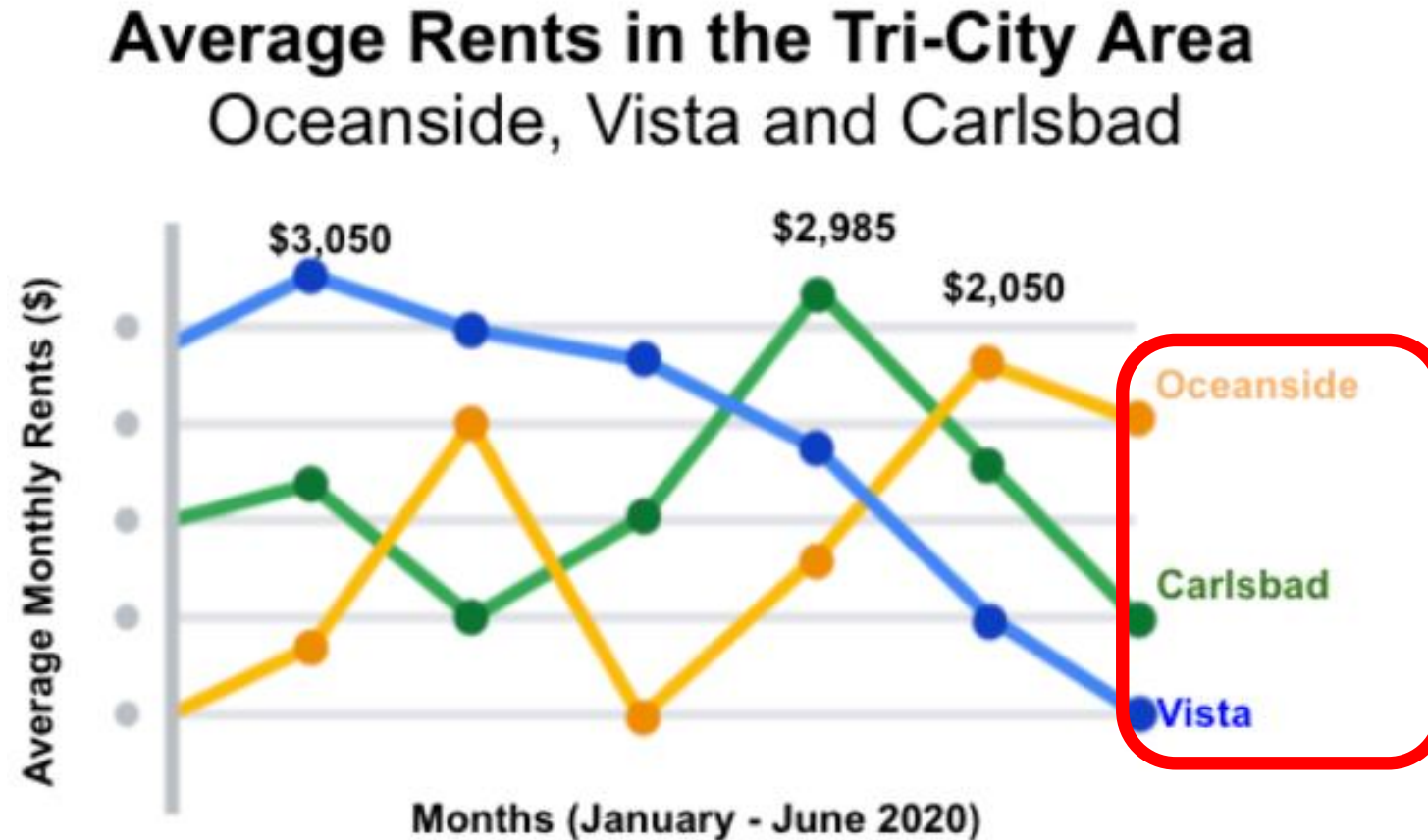
Effective Visual Cues

- **Annotations that focus**

Source: Coursera course "Share Data Through Art of Visualization"
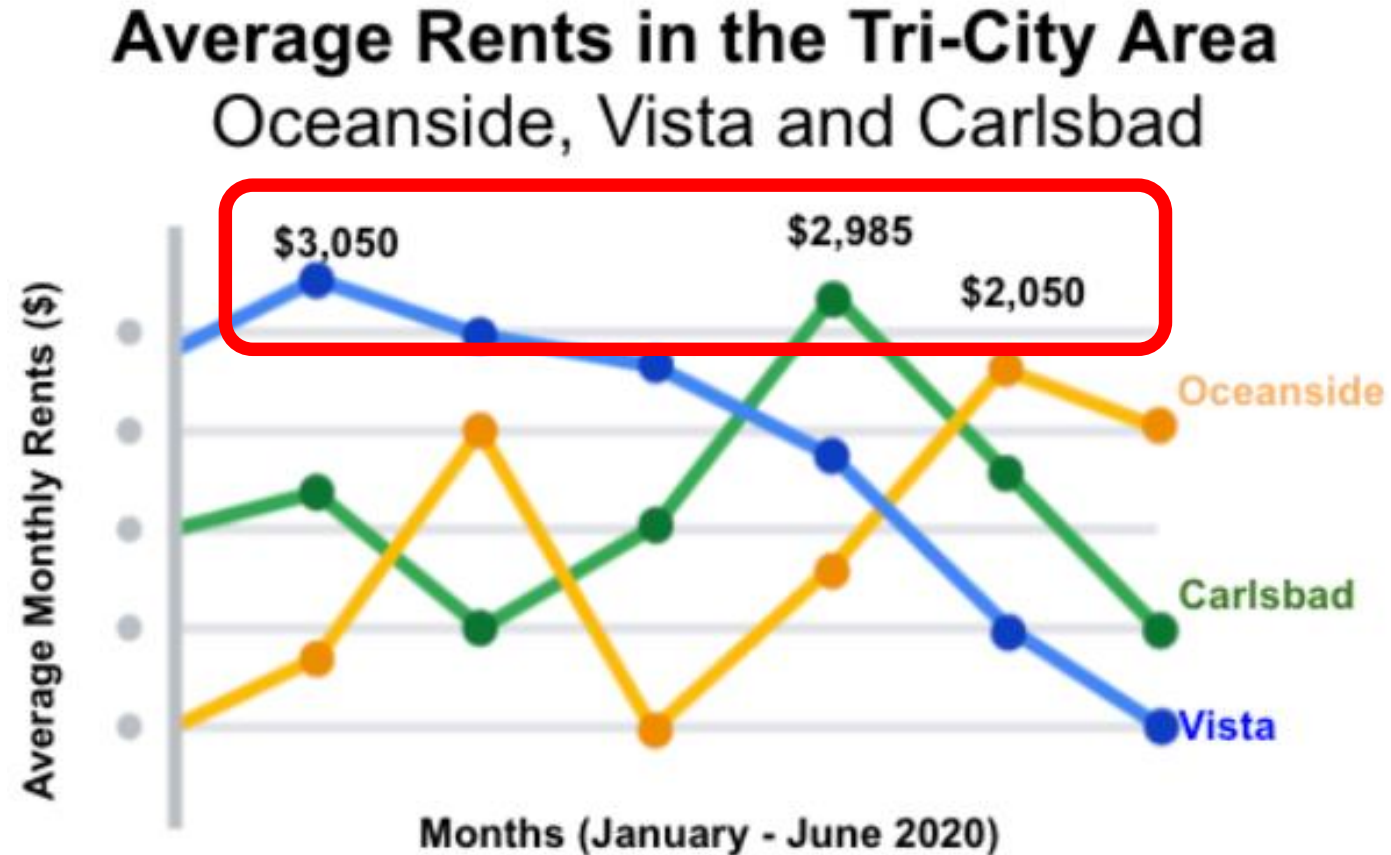
5. Introduction to Tableau

Tableau is a free platform to publicly share and explore data visualizations online.

Key features of Tableau include:

- ✓ drag-and-drop interface,

- ✓ ability to connect to a wide range of data sources,

- ✓ interactive dashboards and visualizations, and

- ✓ support for collaboration and sharing.

Tableau is popular because it allows users to easily create visualizations and explore data in an intuitive and interactive way, without requiring extensive technical knowledge.

6. Loading the Table

1. Log in to Tableau Public.

2. Go to your profile. Hover over the circle in the upper-right corner and click **My Profile**.

3. Under the Getting Started header, click **Create a Viz**.

4. Load the file clicking **Upload from computer**.

5. Double-click on the file **amsterdam.csv** to load that file's data into the main part of the screen. You can also drag and drop the sheet into the area where it says **Drag tables here**.

6. Click **Update Now** to view the data of the file.

7. **Sheet 1** in the left bottom corner.

7. First Chart

Tableau interprets the type of data in each column. The following icons, which are above in the column name, refer to how Tableau interprets the data type in the column.

Numeric data   String data   Geographic data

# amsterdam.csv F1	# amsterdam.csv Id	Abc amsterdam.csv Neighbourhood	⊕ amsterdam.csv Latitude	⊕ amsterdam.csv Longitude	Abc amsterdam.csv Room Type	# amsterdam.csv Price
0	2818	Oostelijk Havengebied - Indis...	52.3657545	4.941419	Private room	59
1	3209	Westerpark	52.3902251	4.873924	Entire home/apt	160
2	20168	Centrum-Oost	52.3650870	4.893541	Entire home/apt	80
3	25428	Centrum-West	52.3731144	4.883668	Entire home/apt	125
4	27886 25428 trum-West		52.3867273	4.892078	Private room	150
5	28658	Bos en Lommer	52.3753422	4.857289	Private room	65

7. To create chart click **Sheet 1** in the left bottom corner.

Dimensions

Measures that you can track for these dimensions

8. Drop *Neighborhood* dimension to *Color* icon for Tableau to assign the color for each *Neighborhood*.

10. Double-click of Latitude and Longitude to see the *Neighborhood* spots on the map.

11. Double-click *Number of reviews* measure to assign the value for each *Neighborhood.*

Change the size of the drops

Add/modify the labels

Add/modify the comments

If we double-click *Room Type*, we will see the distribution of *Number of Reviews* for each type of room.
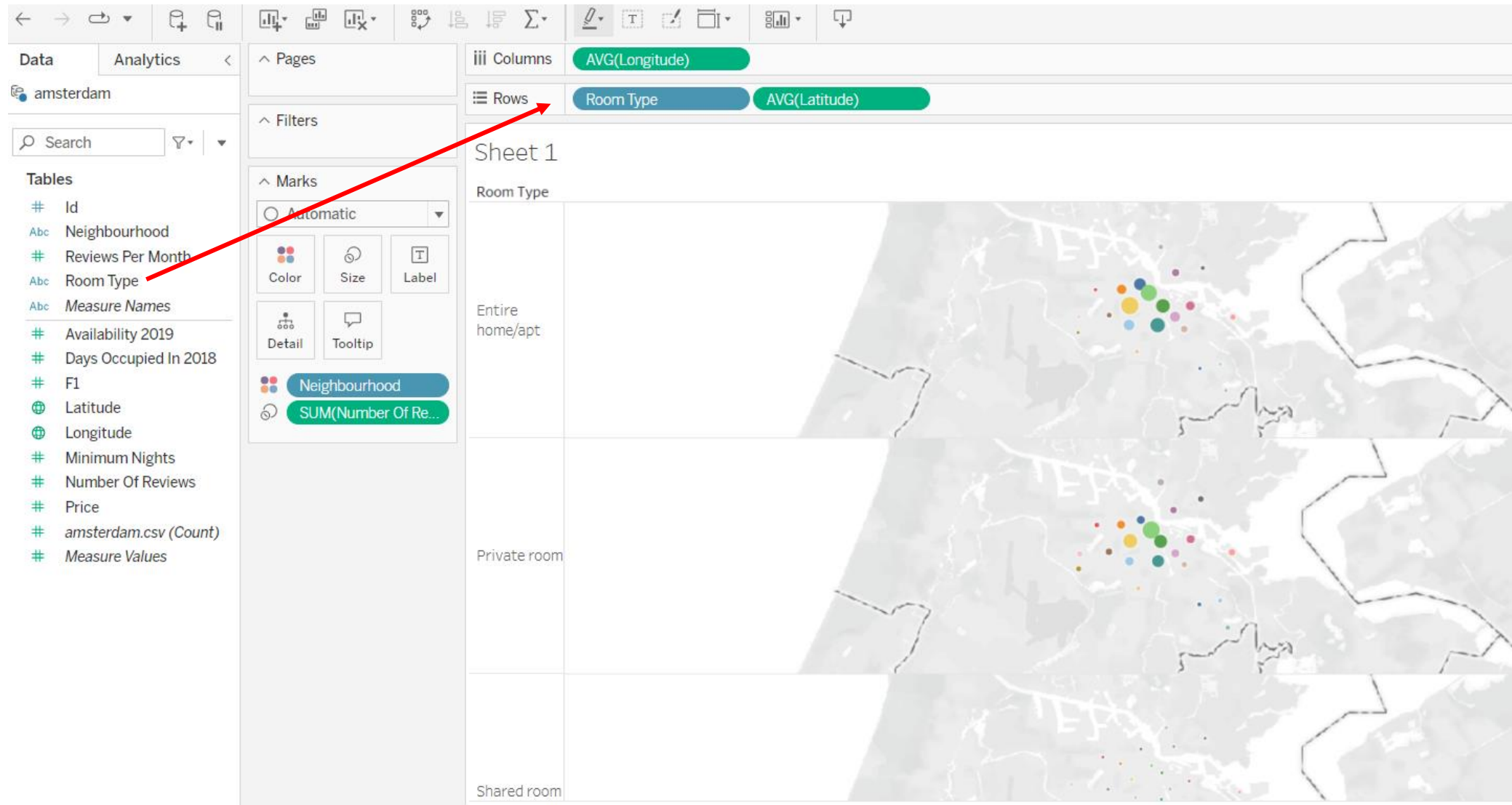
8. Filtering

Let's filter data by Room Type.
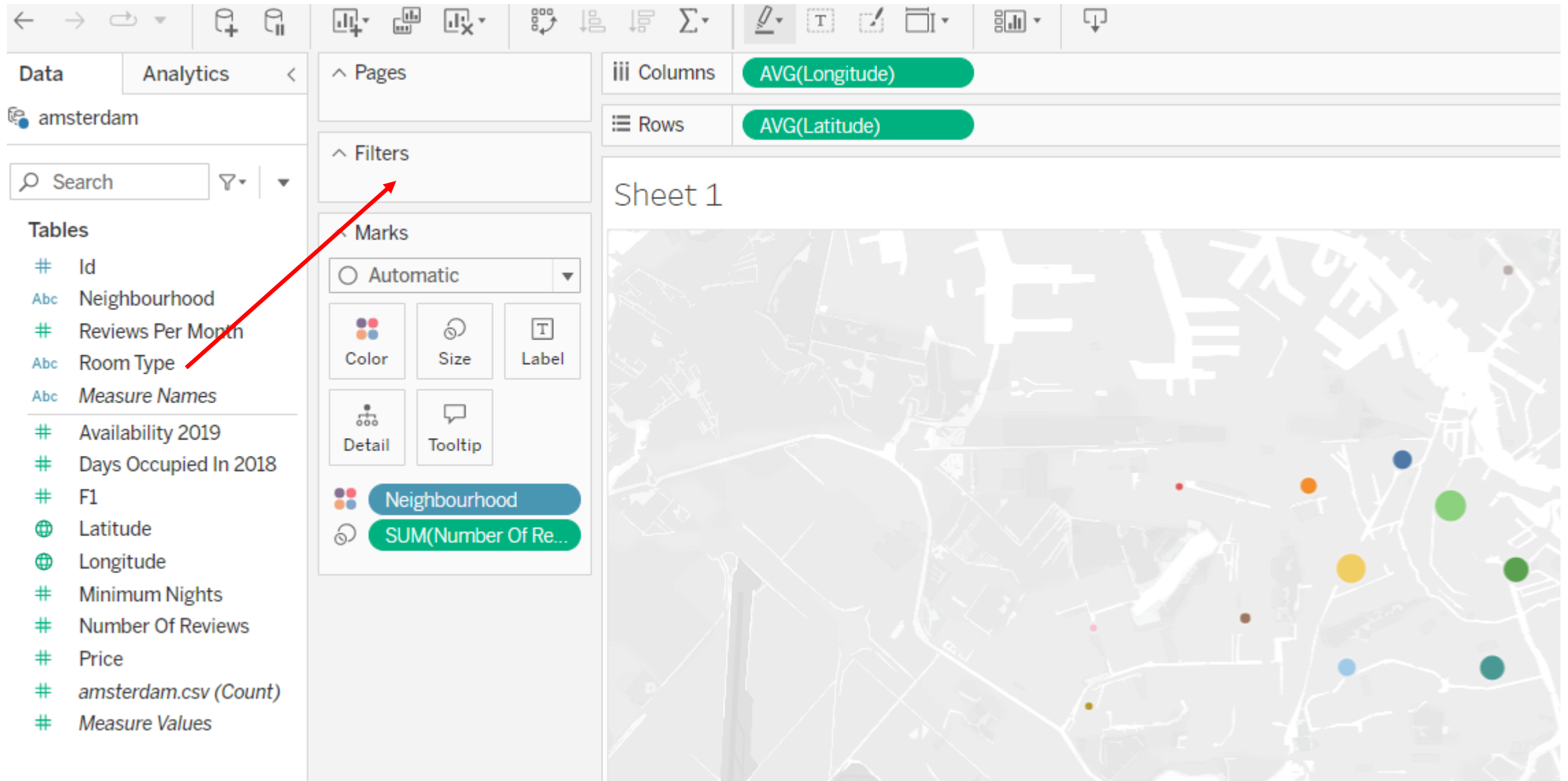
Check the required Type of the Room.

Make this filter option to the available to the user

You can leave only values you need by manually choosing the region on the map.
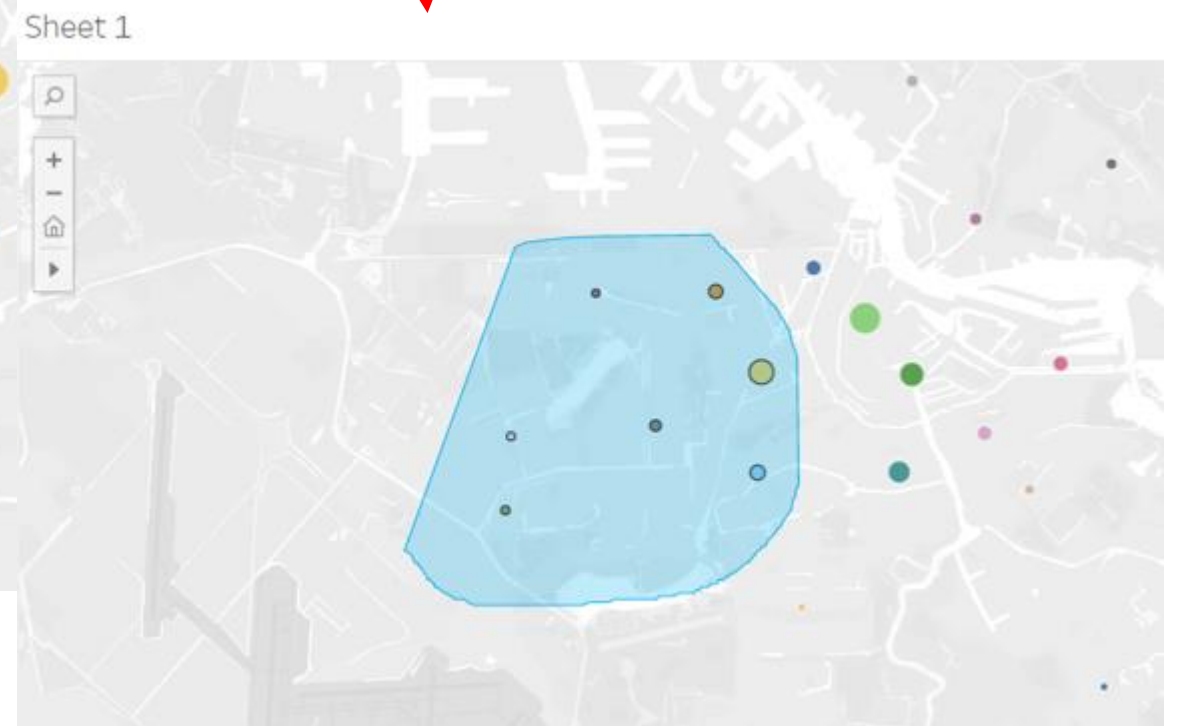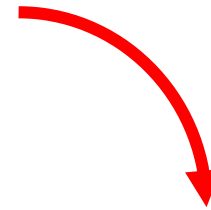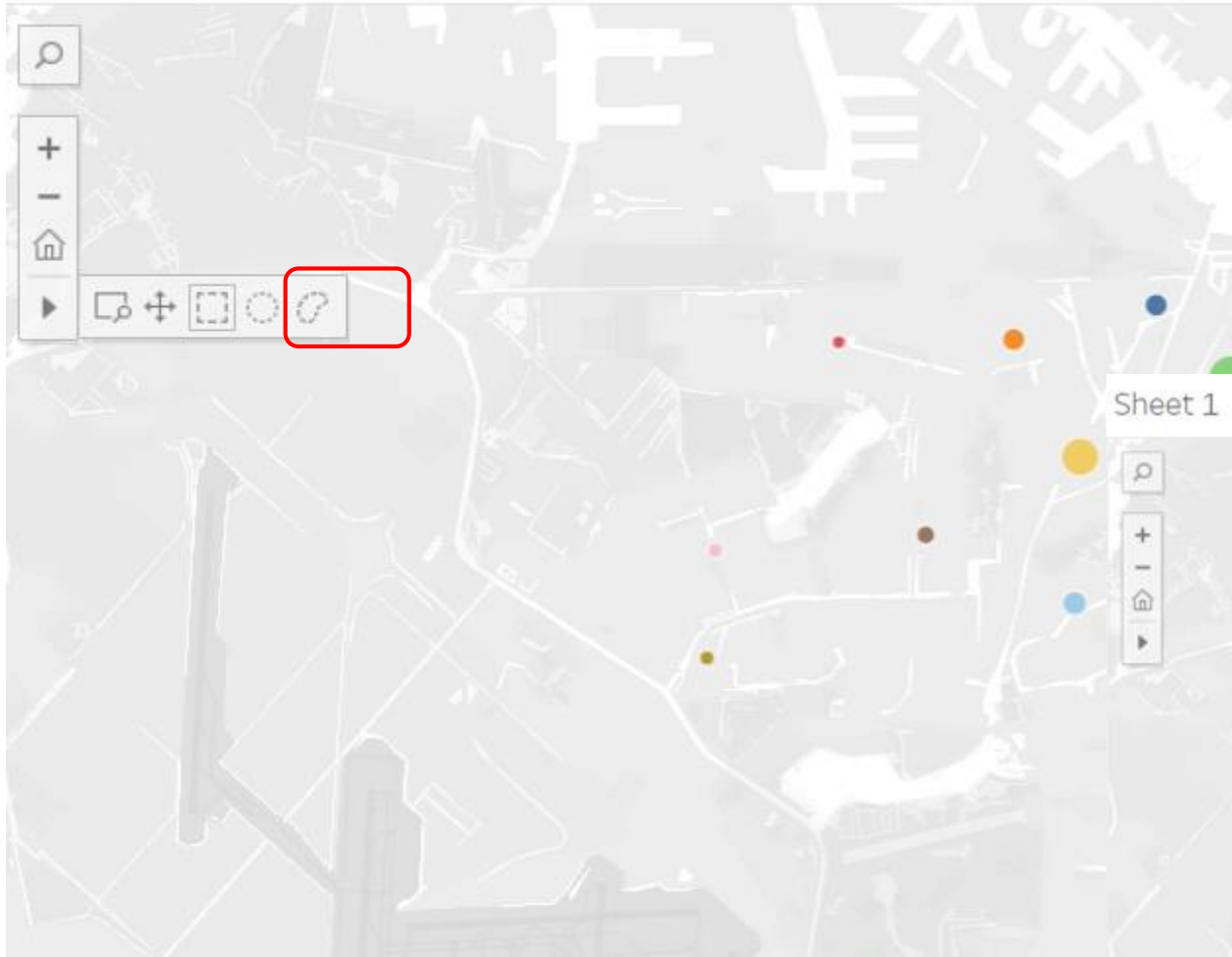
9. Dashboards

Load development_indicators file.

Let's say, we want to see the relation of GDP per capital with the level of CO2 Emission and GDP per capital with Life Expectancy in the countries of Asia.

Filter the continent
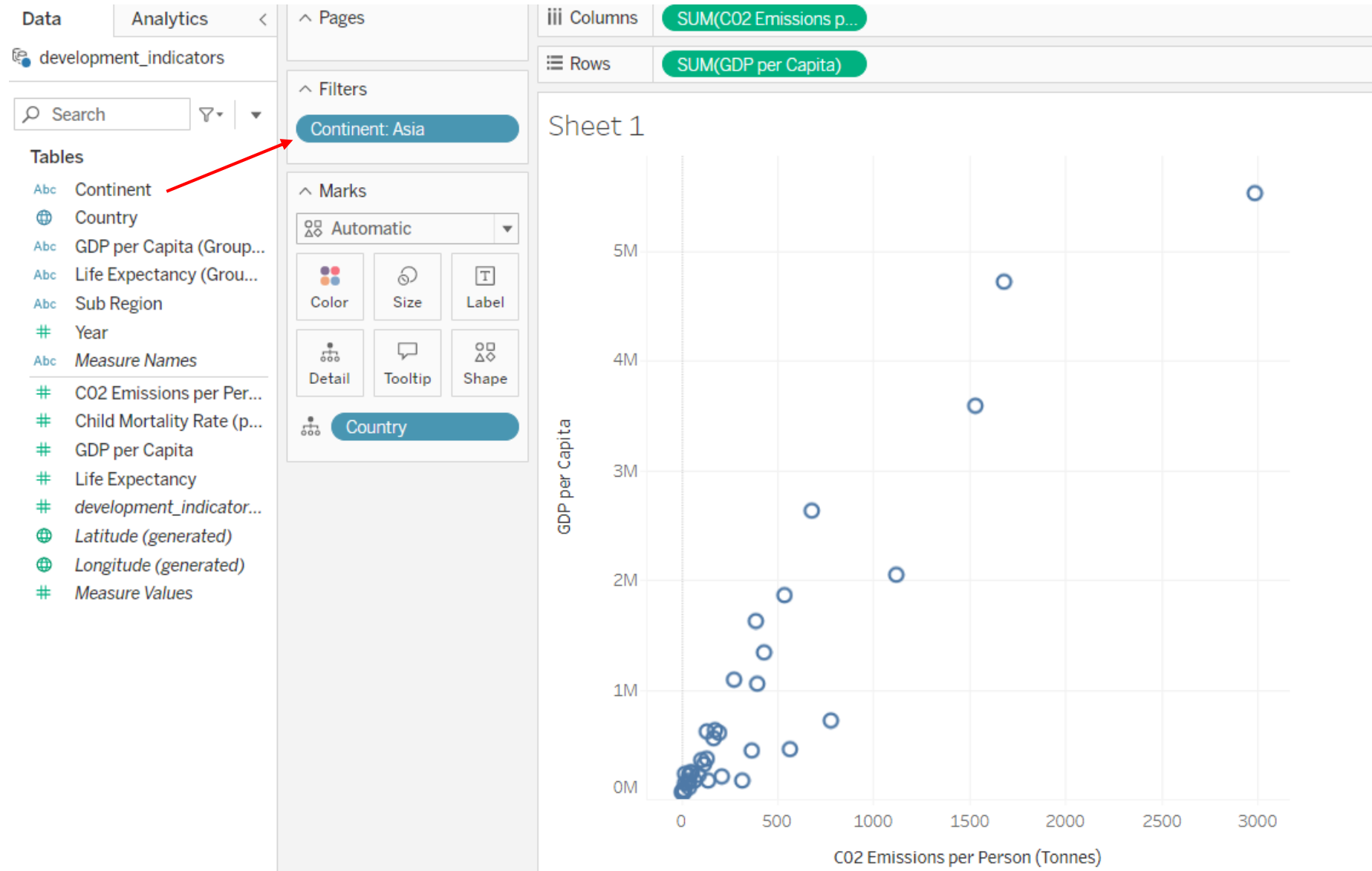
Filter the continent.

Add GDP per Capita to Rows shelf and CO2 Emissions to Columns shelf respectively.
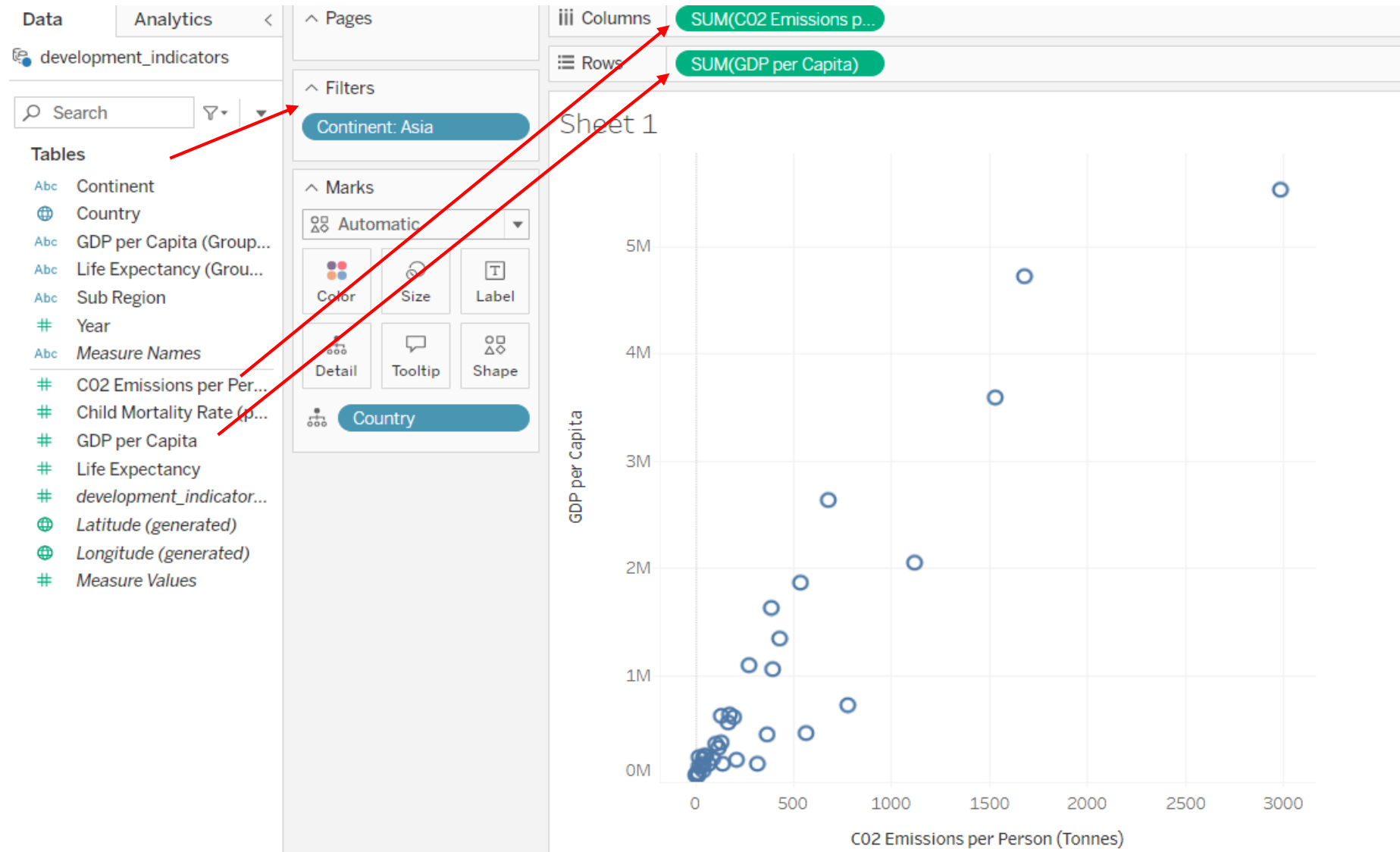
Add trend line to the chart
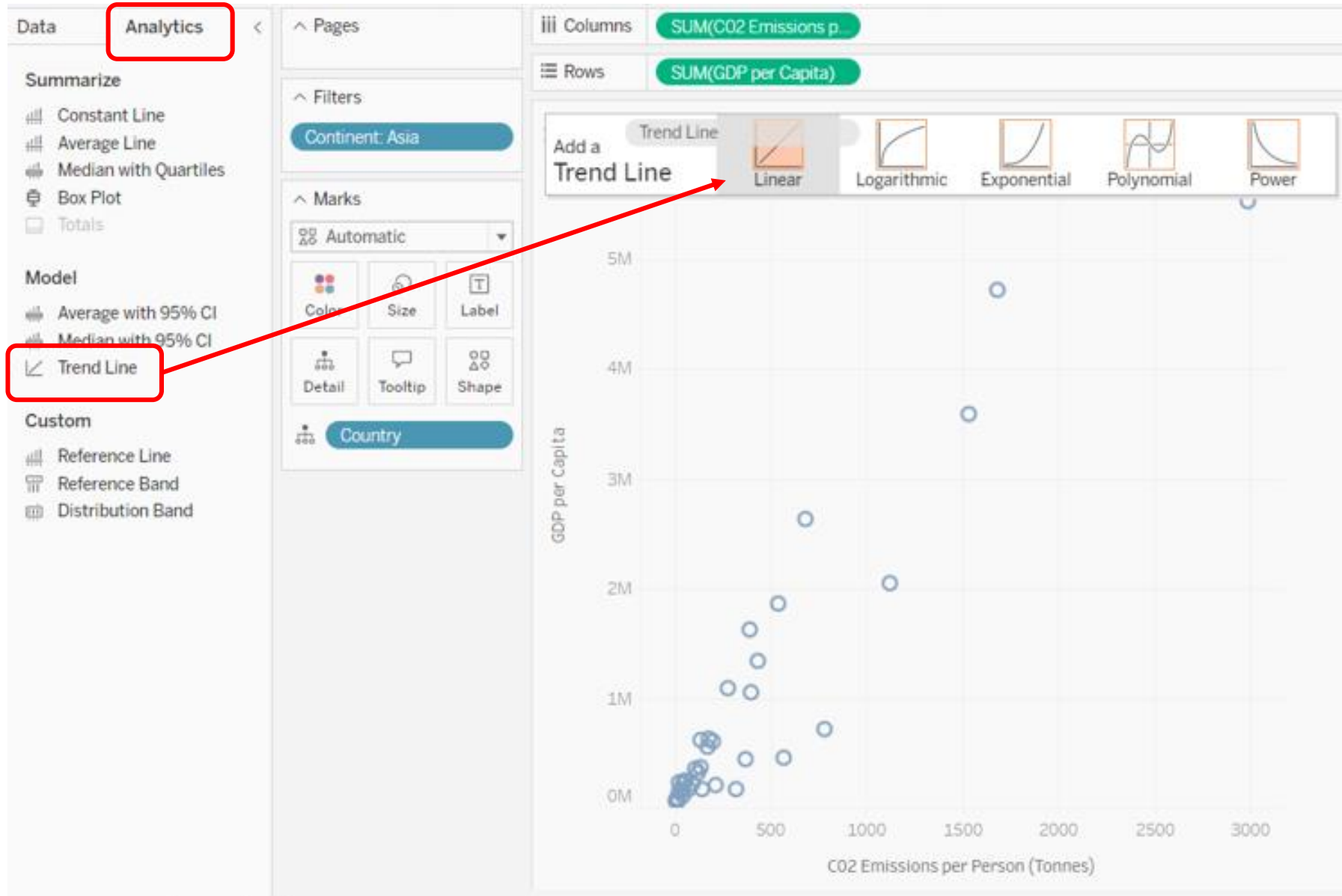
Make a copy of the chart.
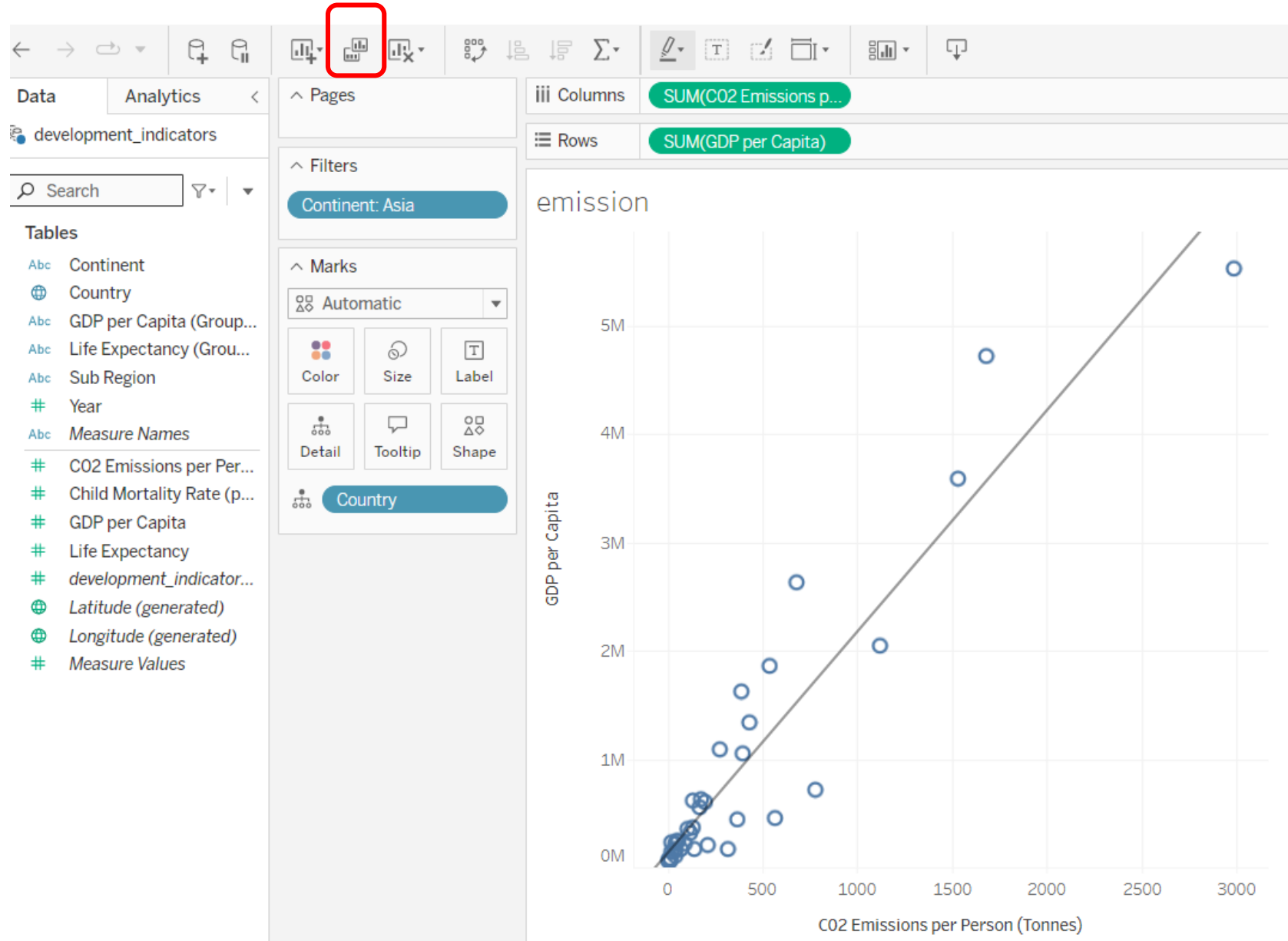
In the copied sheet change drag Life Expectancy instead of CO2 Emission in the Columns shelf.

Do not forget to change the measure of Life Expectancy to average.
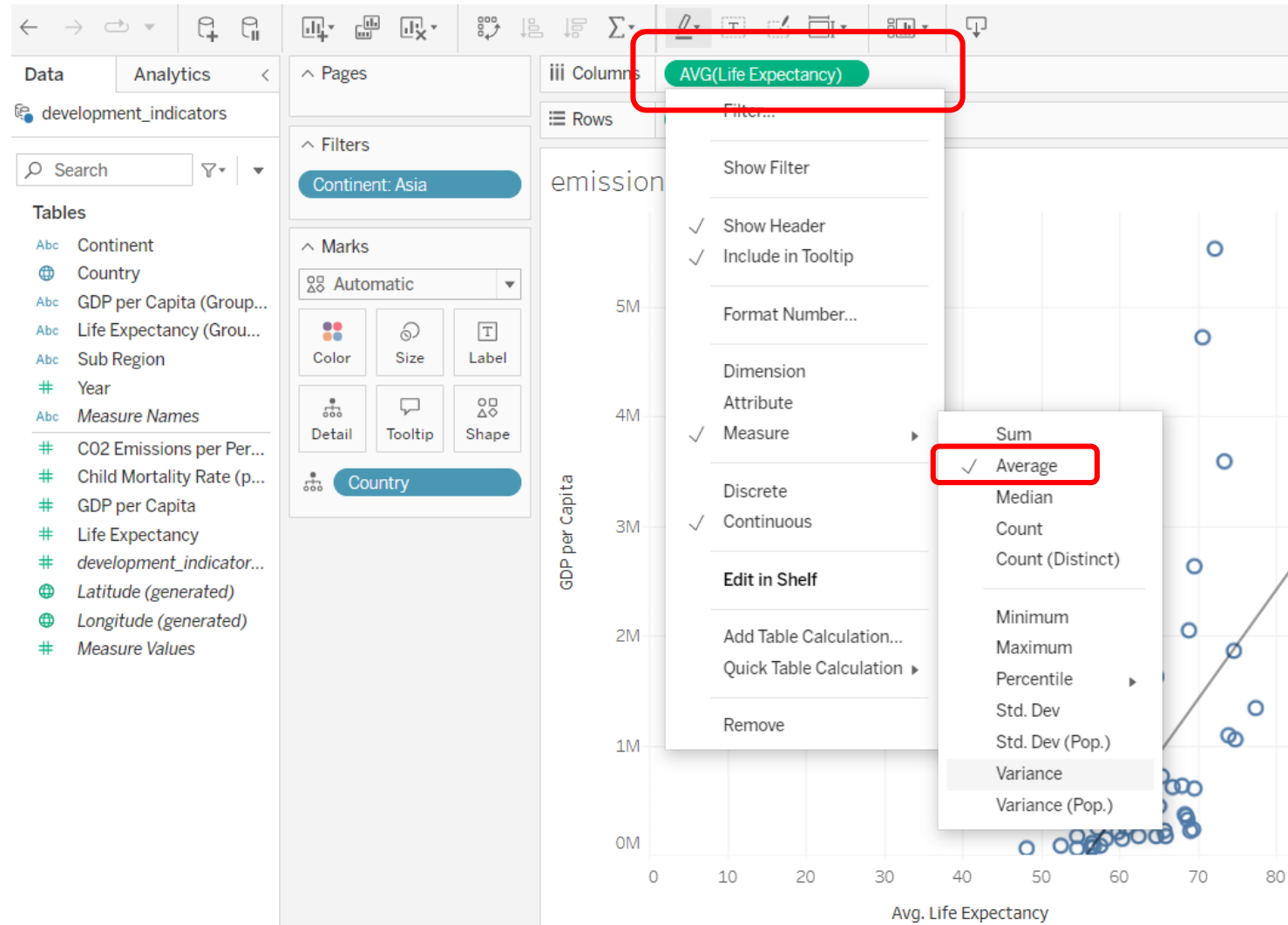
Change the trend line to Exponencial.
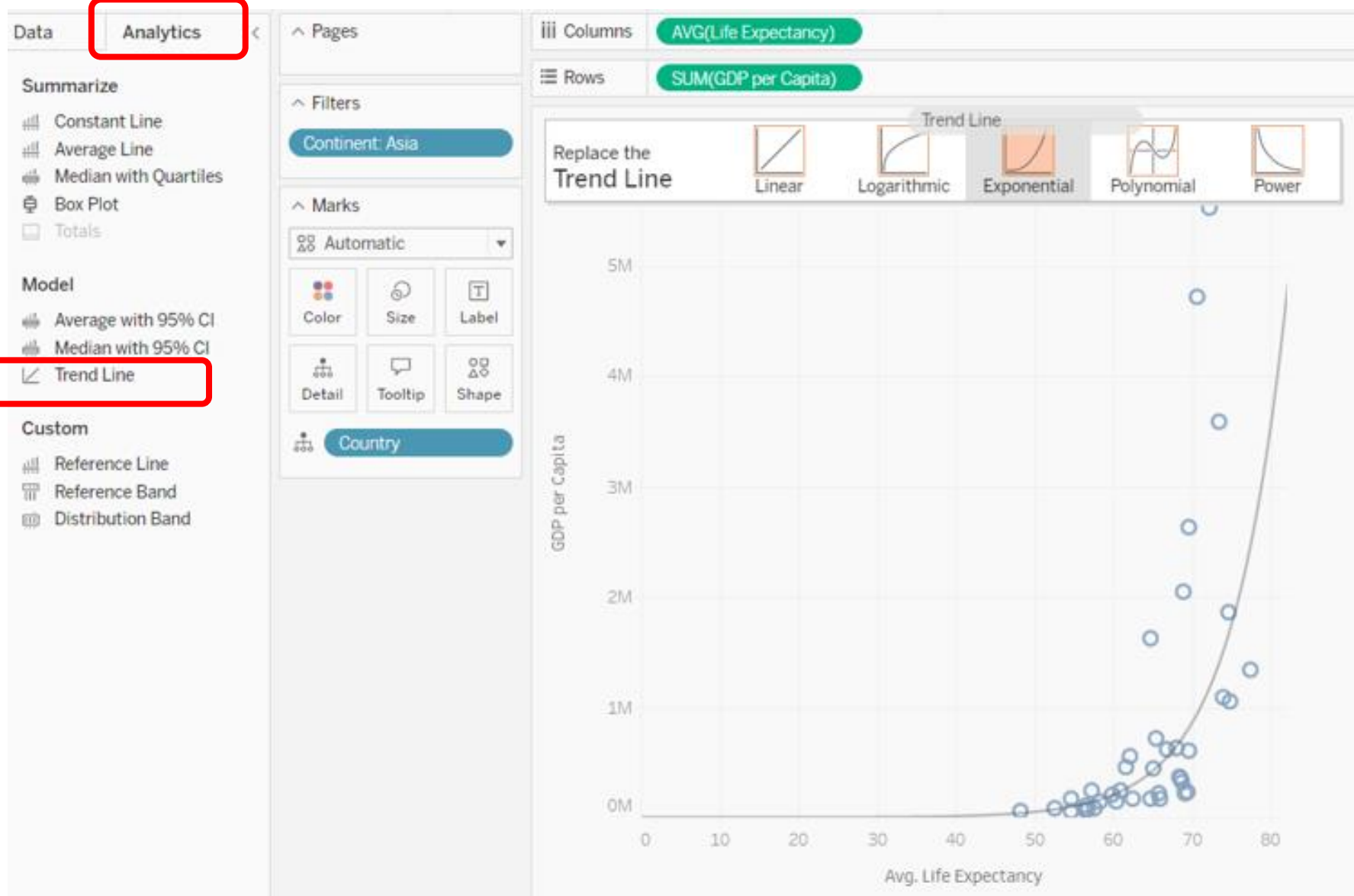
Rename the sheet 1 by double-clicking on it.
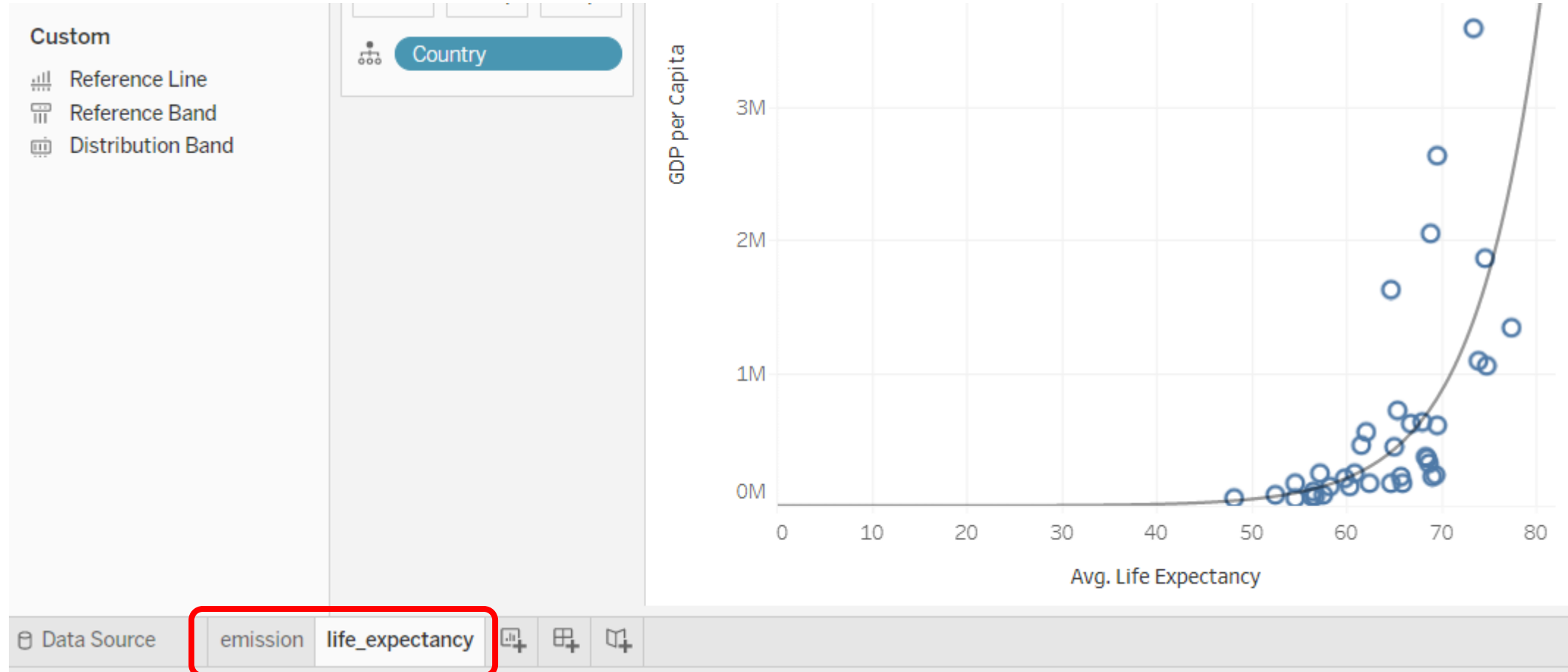
Start to create Dashboard

Drag Sheets on a Dashboard to create a multiple chart.

Let's create table with data on GDP, Emission and Life Expectancy in Asia, and make a Dashboard which include this table.

Basic Table

Country	CO2 Emissi..	GDP per Ca..	Life Expect..
Afghanistan	8	66,086	2,649
Armenia	87	218,840	3,797
Azerbaijan	363	443,040	3,571
Bahrain	1,116	2,050,000	3,788
Bangladesh	9	80,943	3,164
Bhutan	18	144,267	3,200
Cambodia	7	81,071	2,884
China	139	168,519	3,560
Cyprus	275	1,091,120	4,063
Georgia	117	323,820	3,776
India	40	113,090	3,109
Indonesia	51	250,600	3,351
Iraq	169	565,400	3,411
Israel	395	1,059,360	4,113
Japan	431	1,337,430	4,252
Jordan	128	373,390	3,758
Kazakhstan	779	722,190	3,593
Kuwait	1,527	3,592,000	4,033
Lebanon	175	628,750	3,738
Malaysia	192	608,270	3,824
Maldives	49	247,370	3,144
Mongolia	210	216,580	3,289
Myanmar	10	78,147	3,099
Nepal	4	71,130	3,120
Oman	387	1,626,220	3,560
Pakistan	32	153,270	3,314
Philippines	40	222,110	3,617
Qatar	2,985	5,527,400	3,968
Saudi Arabia	680	2,635,200	3,828
Singapore	535	1,867,470	4,101
Sri Lanka	21	232,490	3,816
Tajikistan	67	174,930	3,429
Thailand	100	359,530	3,762
Timor-Leste	3	63,434	3,006
Turkey	133	620,100	3,674
Turkmenistan	567	463,620	3,381

emission
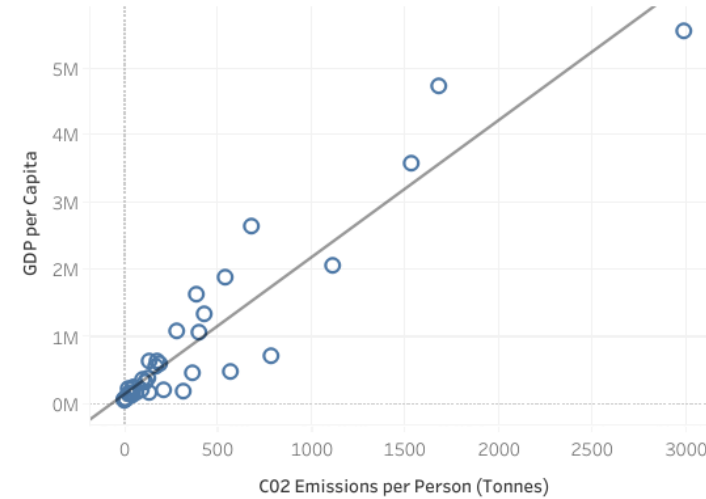
life_expectancy
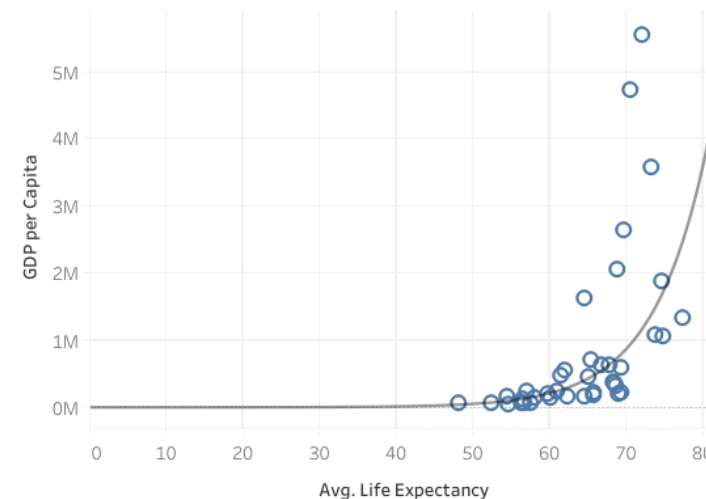
Create new worksheet by one of two ways.
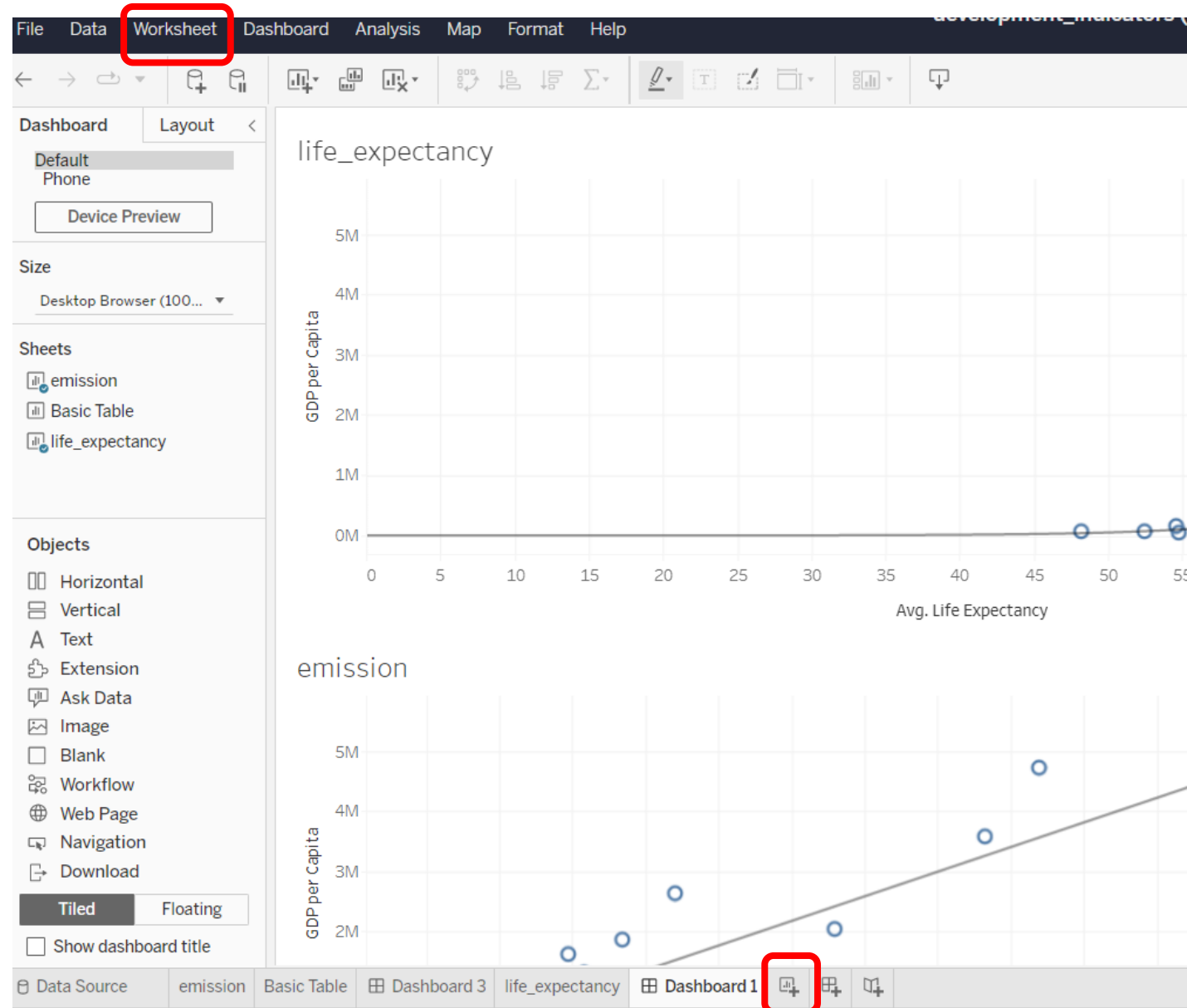
Drag Country Dimension to Rows shelf.

Filter the data.

Drag required Measures to Text icon.

Choose the required type for data to be shown.

Don't forget to rename the sheet with the table.

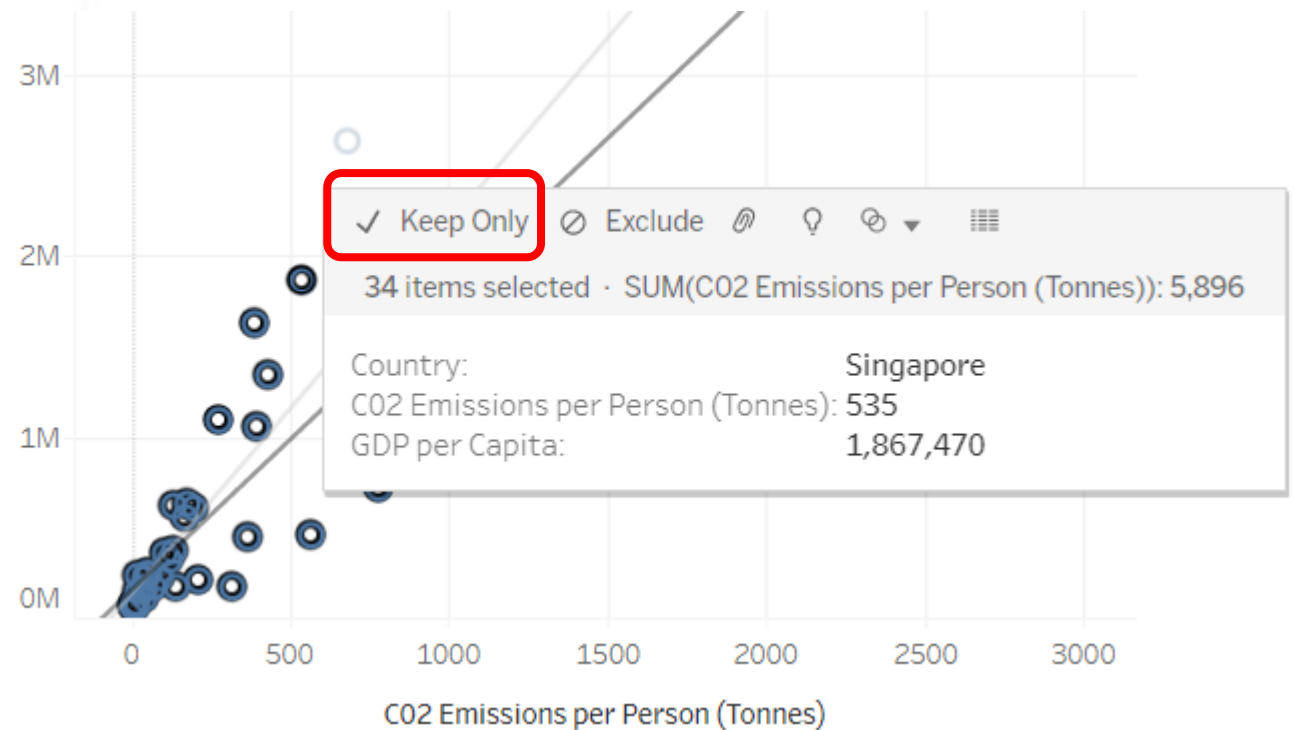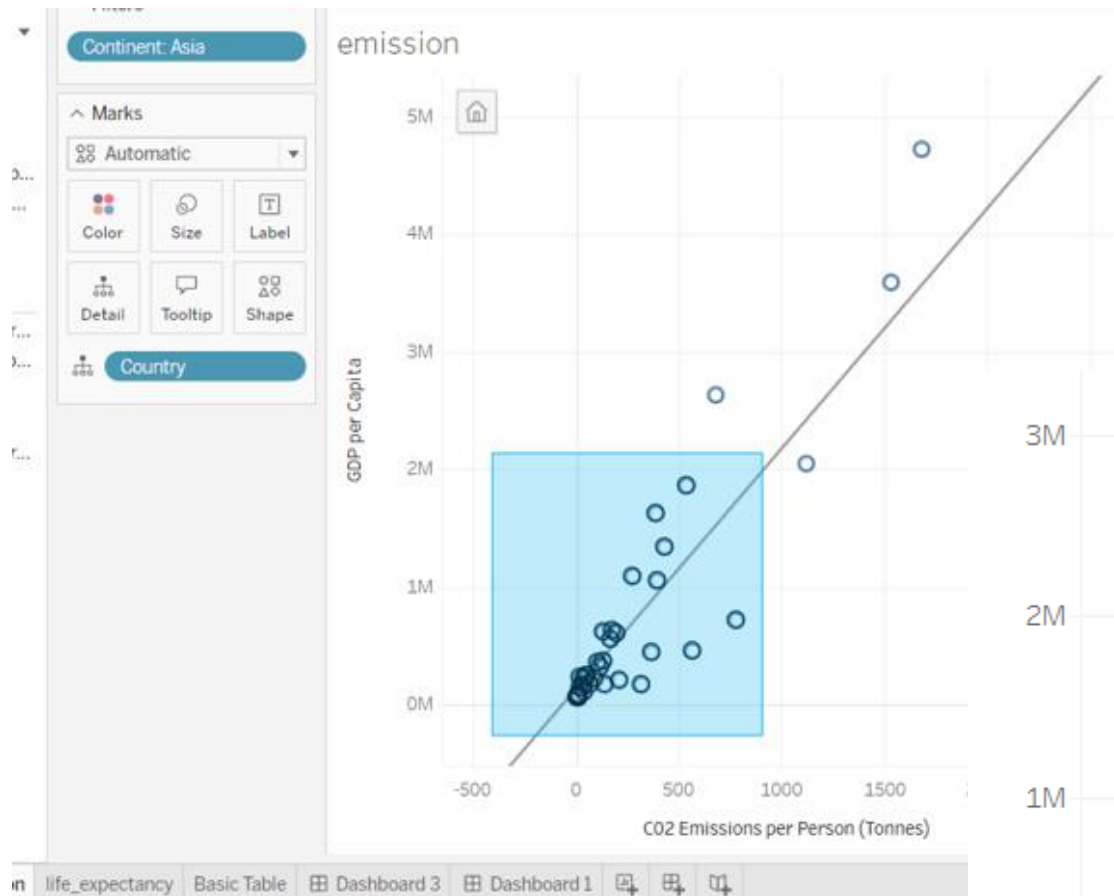Click Dashbord -> New Dashboard.

Drag the Sheets to Dashboard as required by the goals.

10. More filter examples

To delate the outlier double-click on the respective data point and choose *Exclude*.
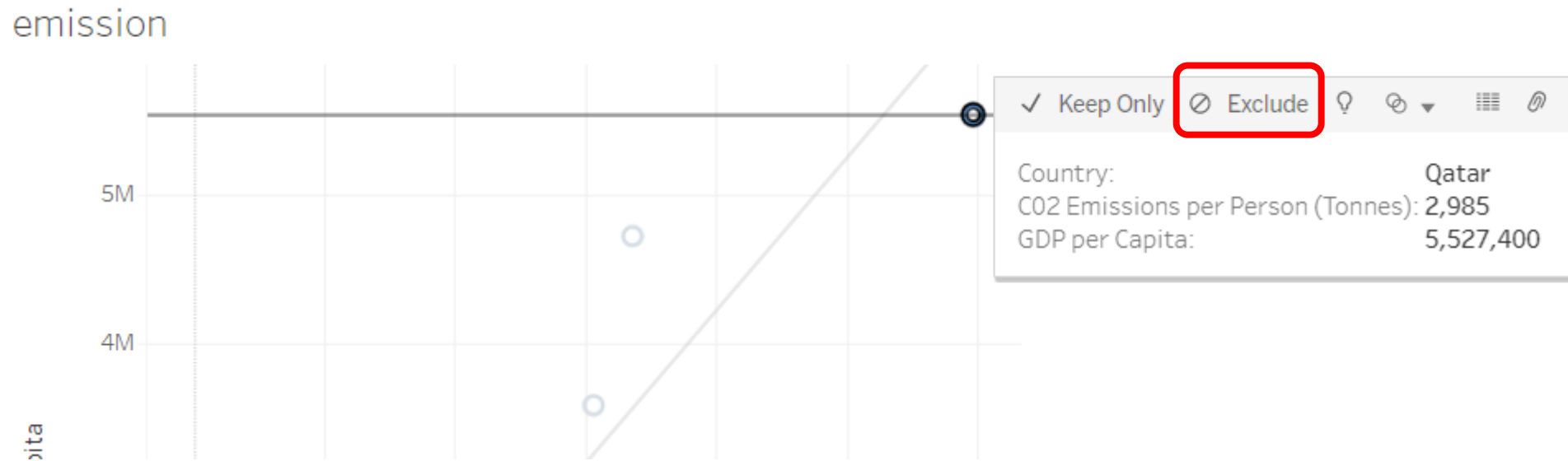
Select only specific data.

Select only specific countries.

Select only specific countries.

11. In-class assignment

- You will be working with the table **UNHealth**.
- Open the file and quickly look at the table to get a sense of what they contain.