Categorical Data

Topics in this session

Categorical data
- Visualizing nominal or ordinal categories
- Bar Charts

How to access key features in the Power BI User Interface
- Aggregation
- Sorting
- Filtering
- Visual formatting

Finding the message in the data

Common questions

What does our data mean?

What question can the data help us answer?

What decision or action should we take next?

Exploratory data visualization a process to answer these questions by looking at visual representations of measures of the data.

Other terms you will come across similar to message:

Signal – separating the signal from the noise in the data.

Story – what is the story in the data.

Categories

Very often we have data organised by category:

For example these can be names or codes:

Type of coffee   Espresso, Latte, Cappuccino

Travel to work area   Berwick, Newcastle, Gateshead

Standard Industrial Classification Code   6201, 6202, 9001

These are alphanumeric *labels* rather than numeric values.

It can be useful to think of them as text values you can count but it usually makes no sense to add them together.

Categorical Data Sources

When you read data into Power BI it assigns a *type* to each column

Load in the excel file   *NE_CompanyData_v00.xlsx*
Click on **Edit Queries** to enter the **Query Editor**

The type is displayed at the top of each column.
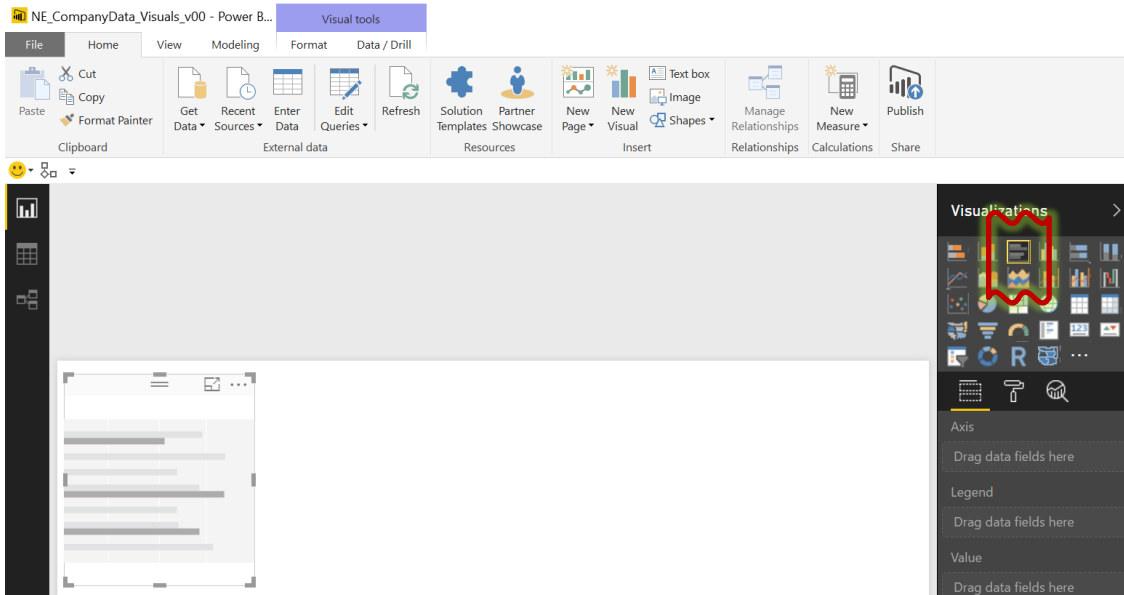Right clicking a column header gives the option to change type.

Bar Charts
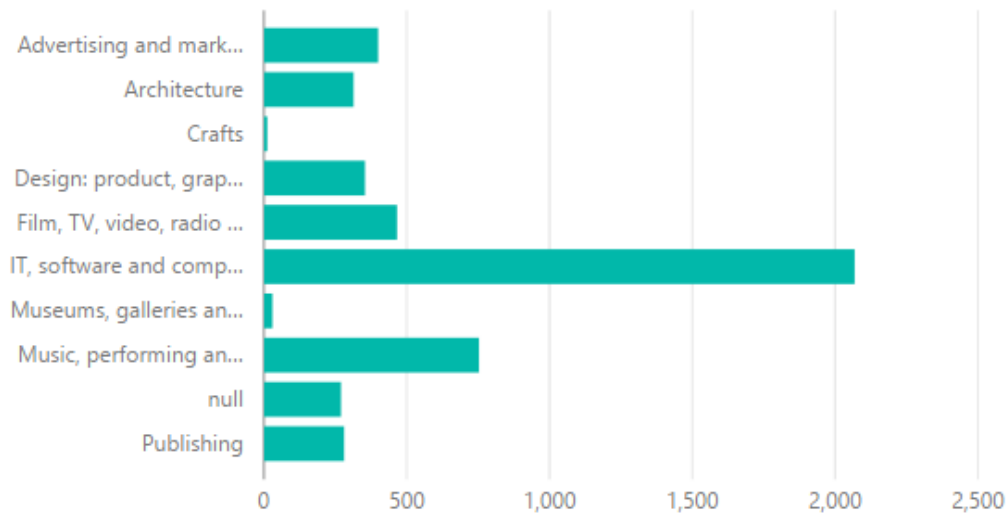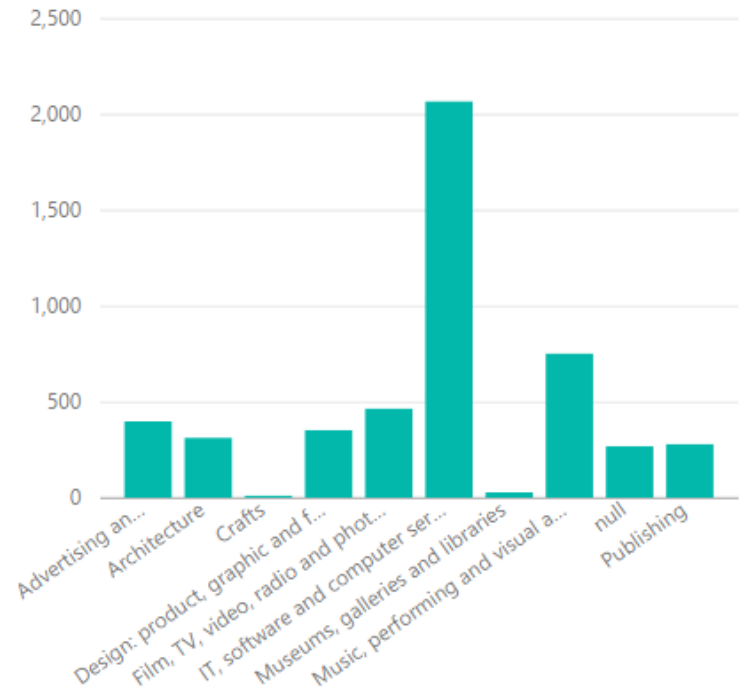
Good for direct comparisons of values across categories.

There are several choices in Power BI
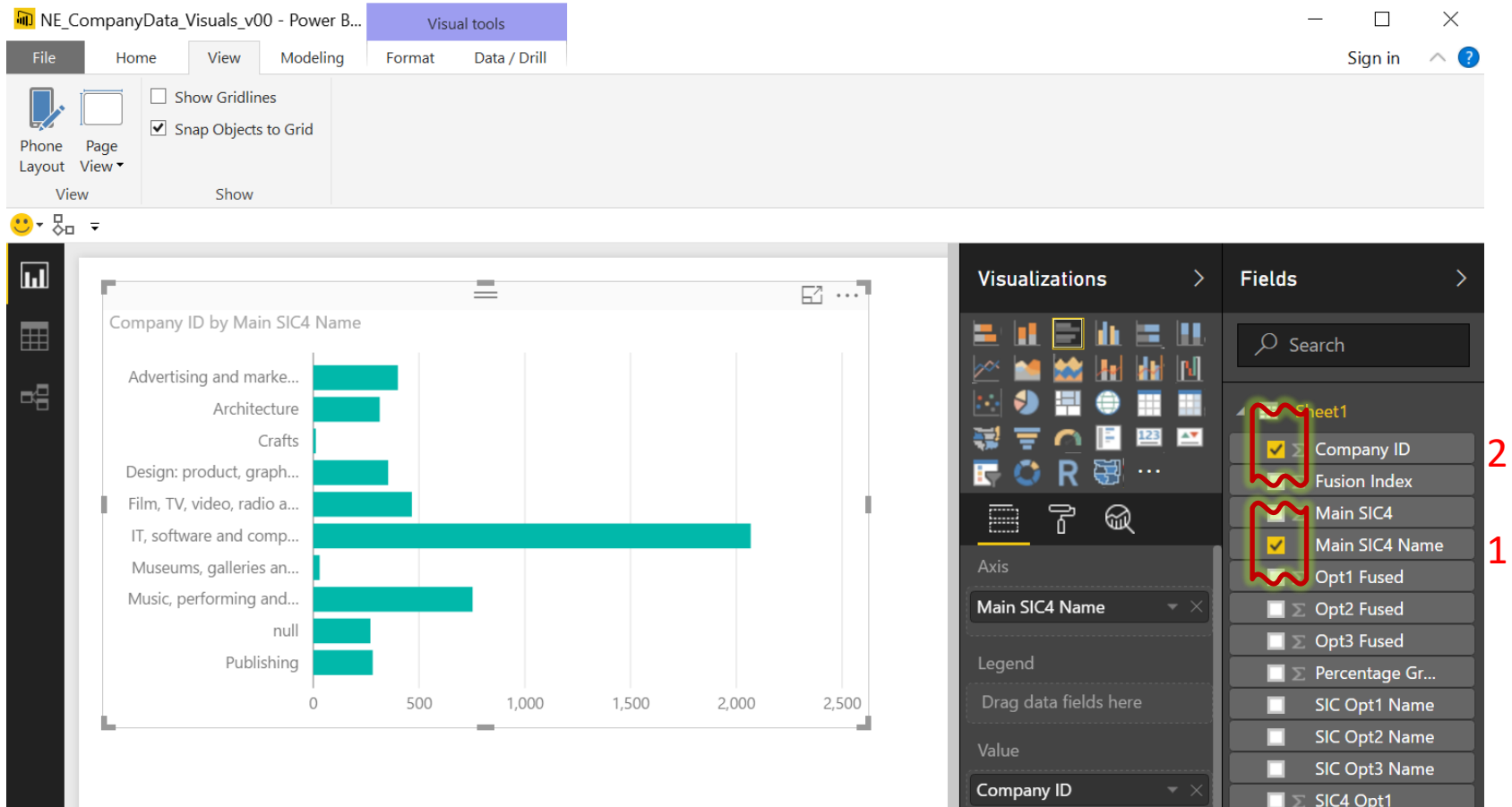   horizontal layout for bar charts is often easier to read.

Clustered Bar Chart

Start creating a simple bar chart visualization by clicking *Clustered Bar Chart*

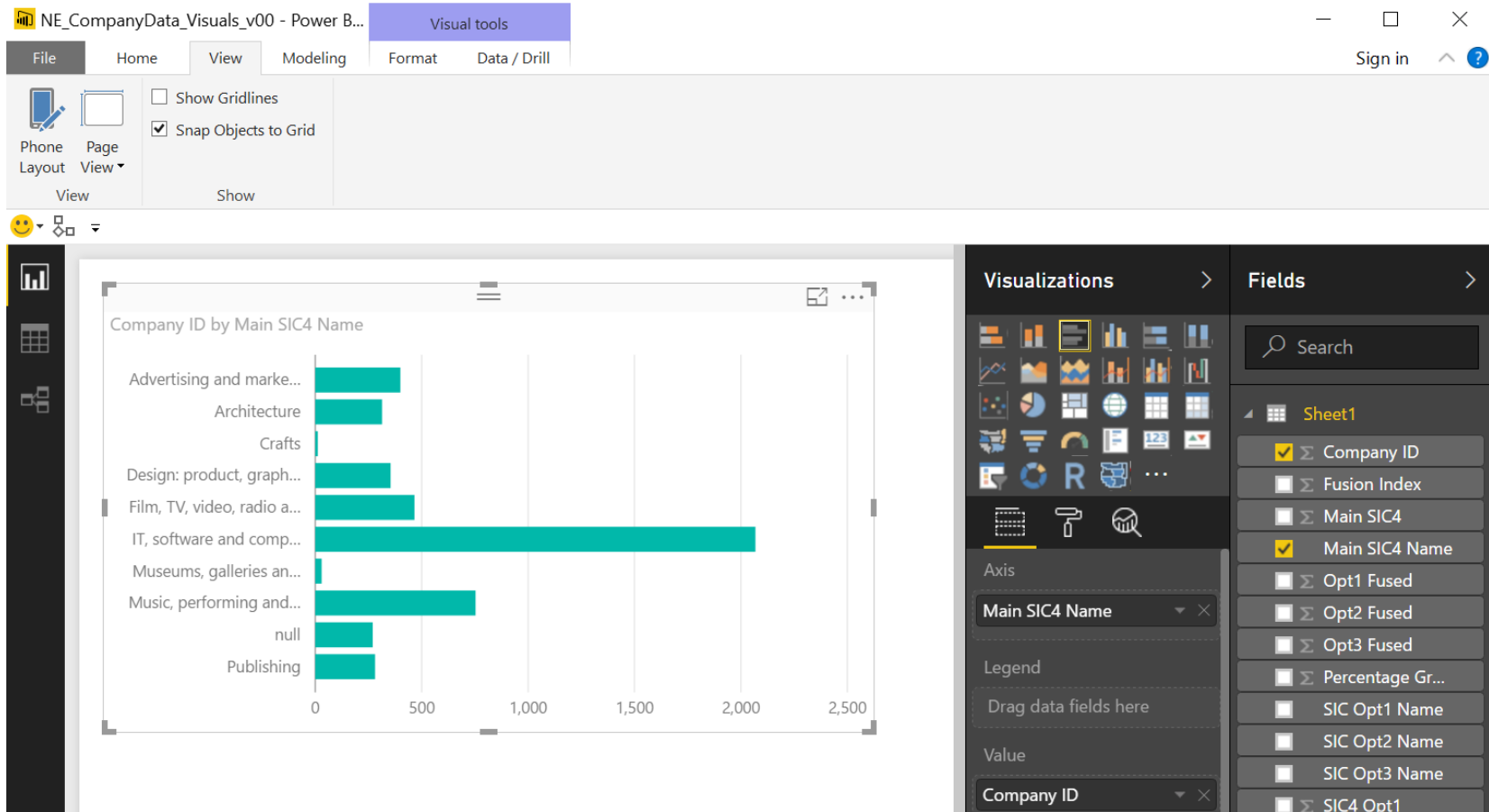This creates an empty chart.

Clustered Bar Chart

To create the visualization we add data
    select a categorical data set for the axis (1)
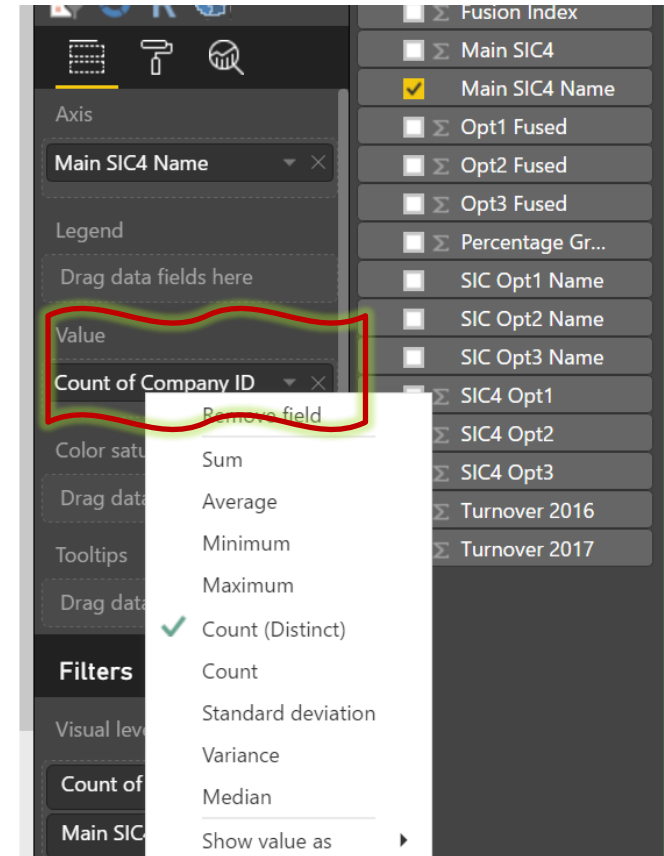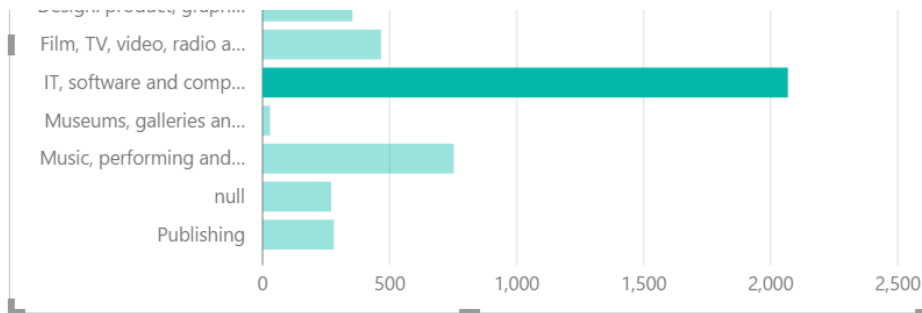   select the categorical labels to count and plot per category (2)

Aggregation

By default Power BI aggregates (in this case counted) the number of companies per SIC code
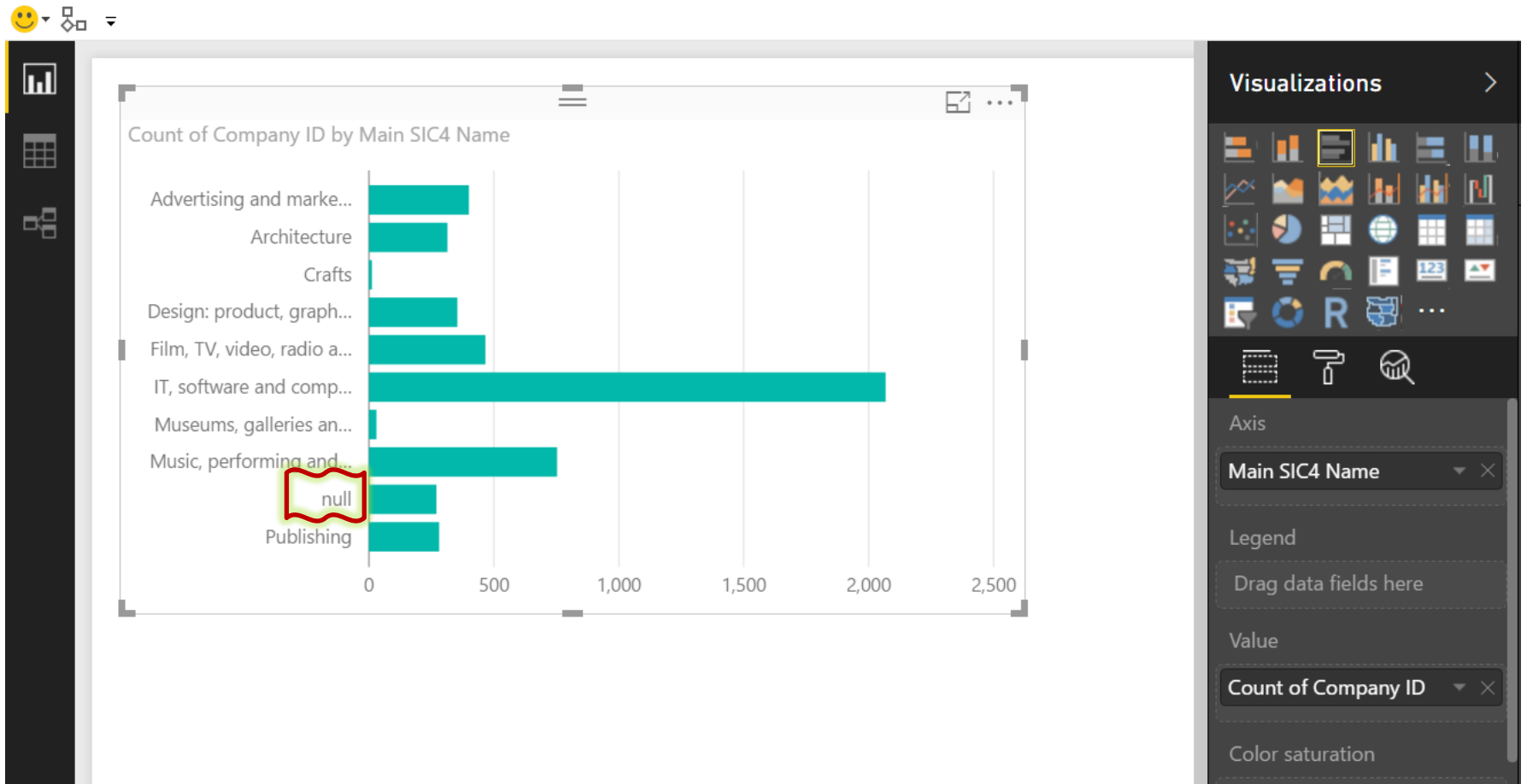
Aggregation Options

Right clicking on the field under **Value** pops up aggregation options change the option to: **Count (Distinct)**
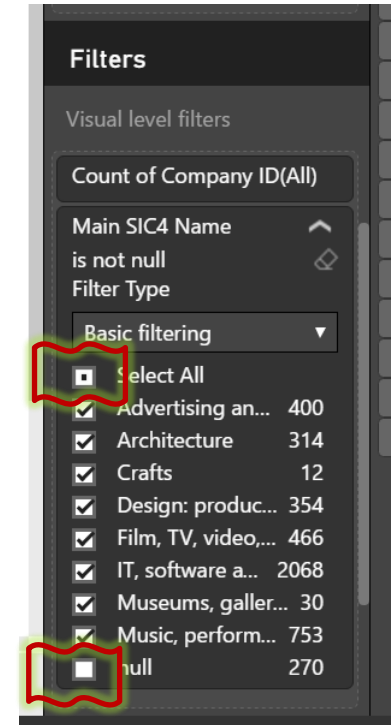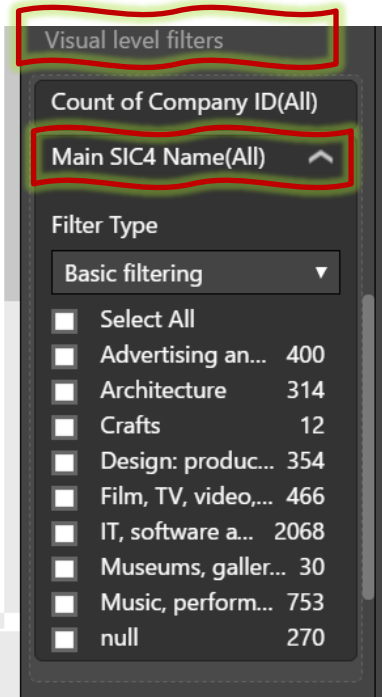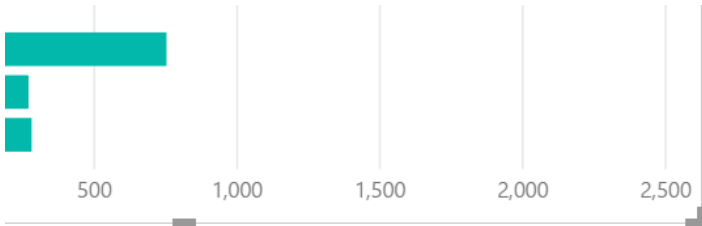This will not change this visualization but it is our logical intent.

Filtering Options

Notice we have a row called "**null**" as this is not a valid SIC code classification for companies we choose to remove it.
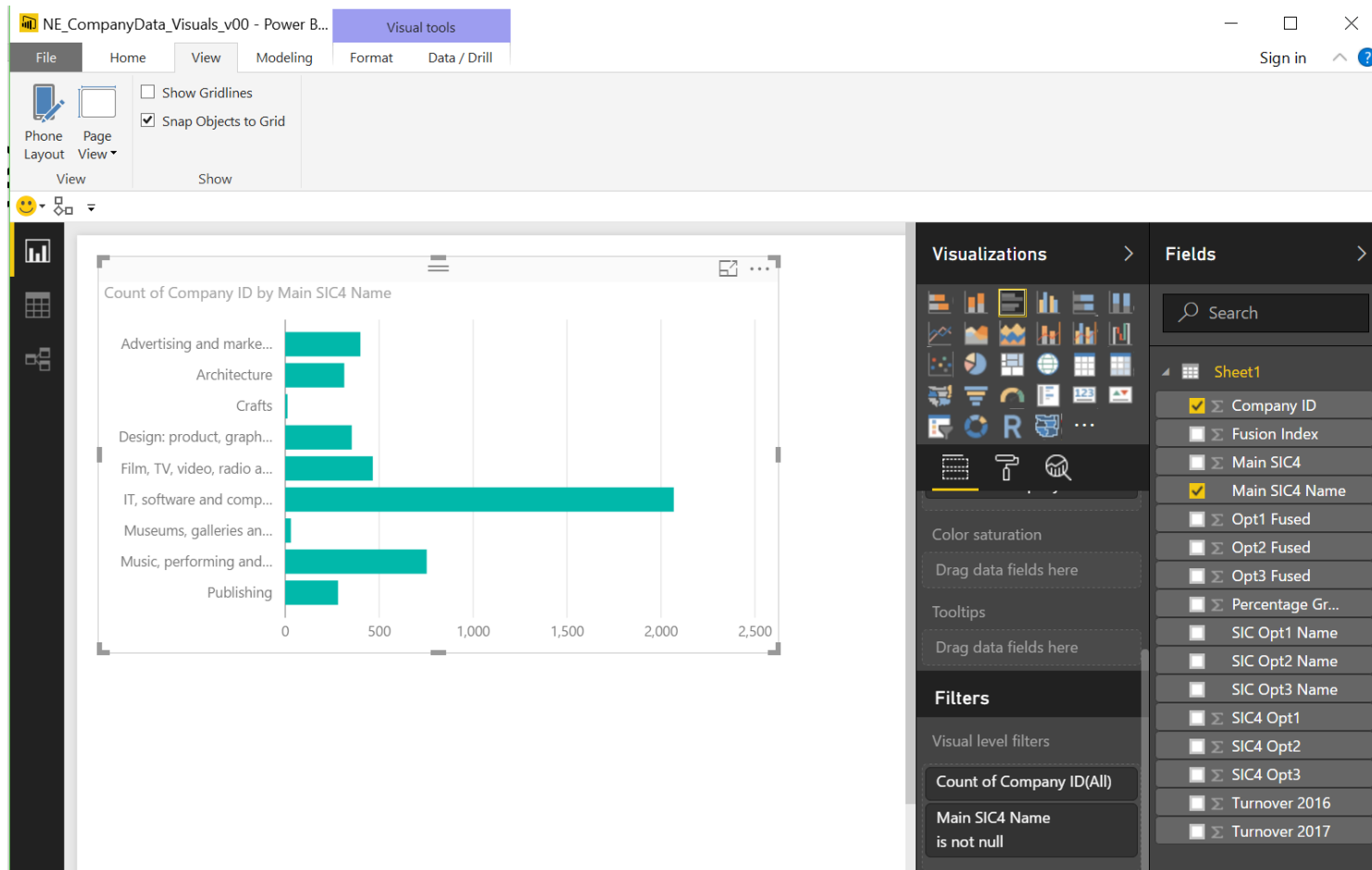
Filtering Options

Click on the field **Main SIC4 Name(All)** in **Visual level Filters** section
This should open a list of all values for this field
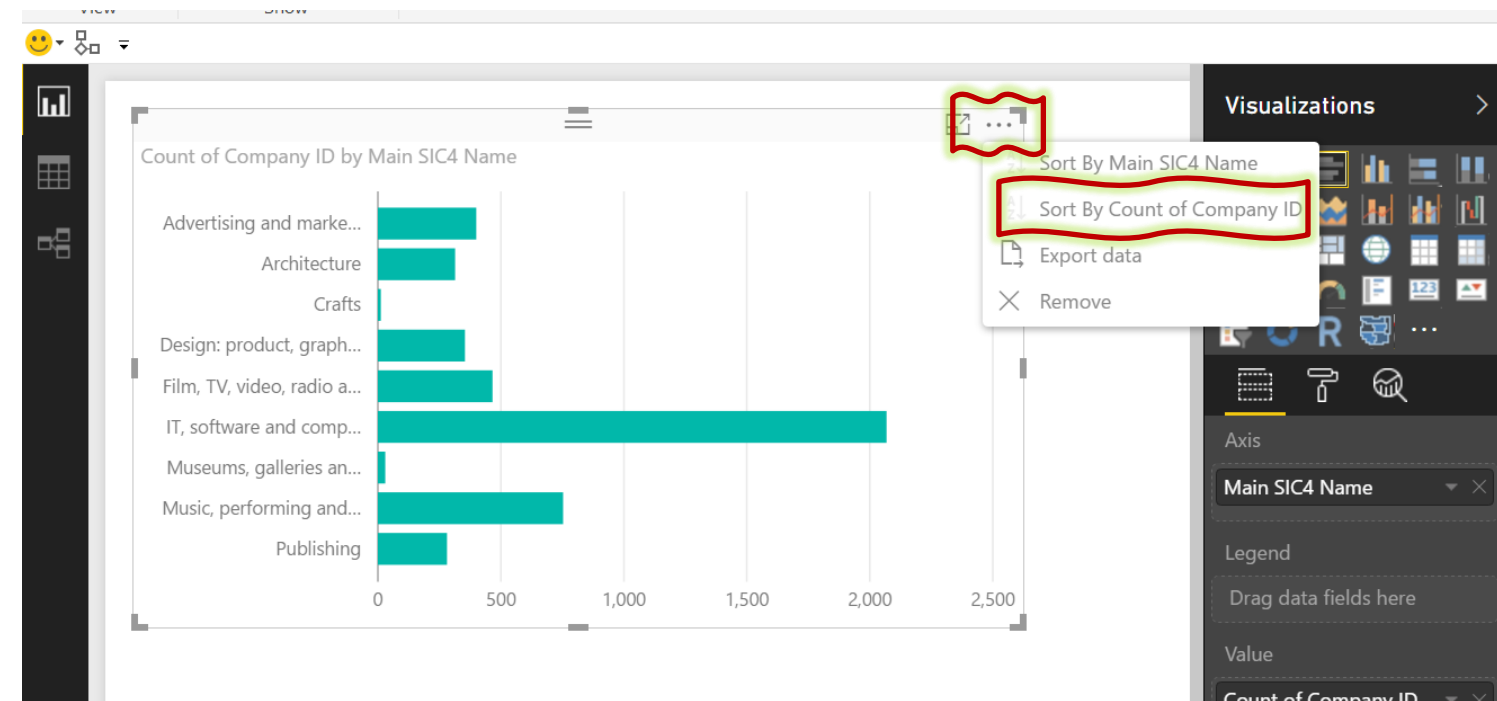
Click **Select All** then uncheck **null**

Filtering Options

We now have a correctly aggregated and null filtered bar chart ....now what/where is the message in this data?
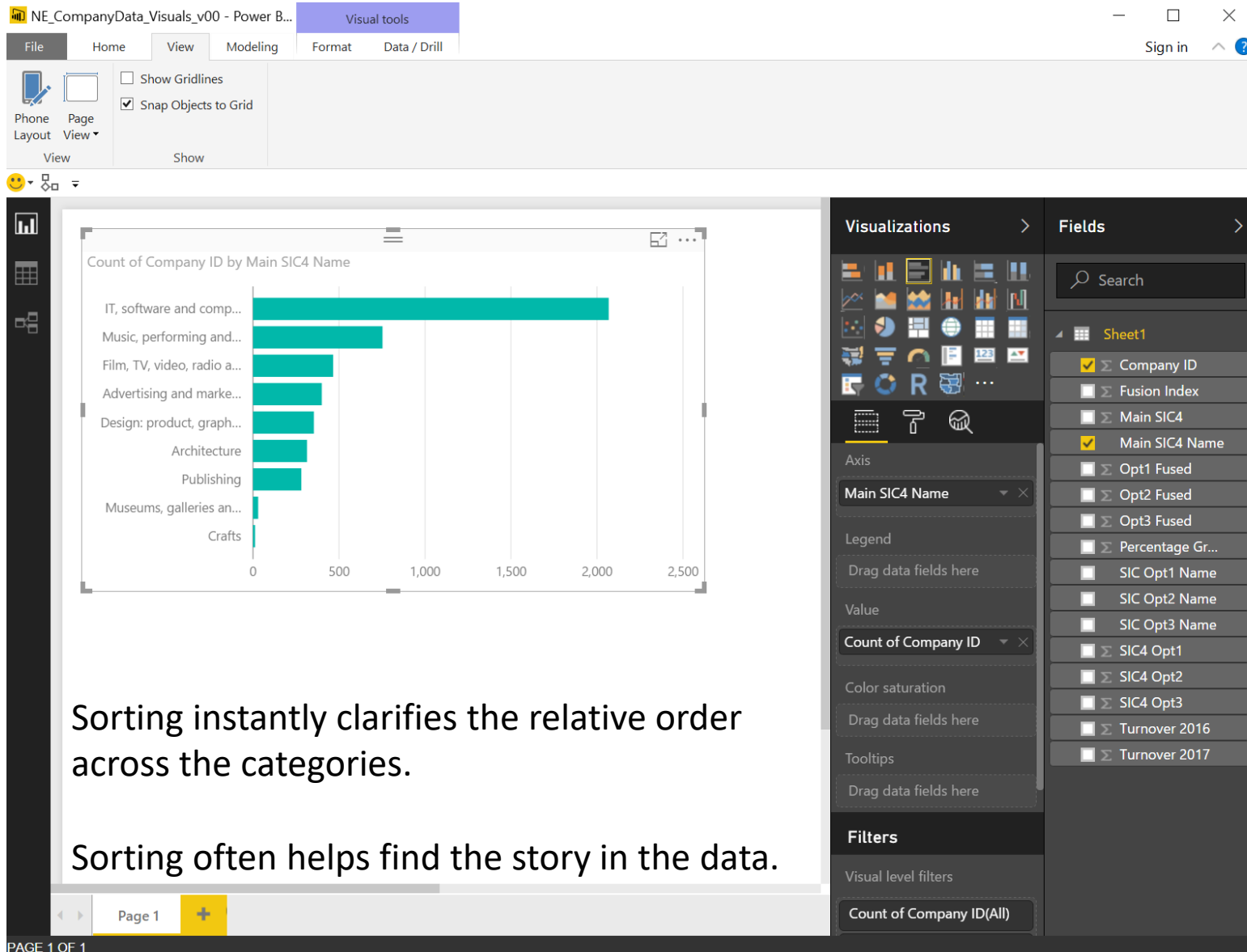
Sorting Data

Sorting data will clarify the context and message in the data
Which SIC code sector is biggest? Which is smallest?

Click on the **…** to bring up the sorting menu for this visual.
Then click on Sort by Count of Company ID
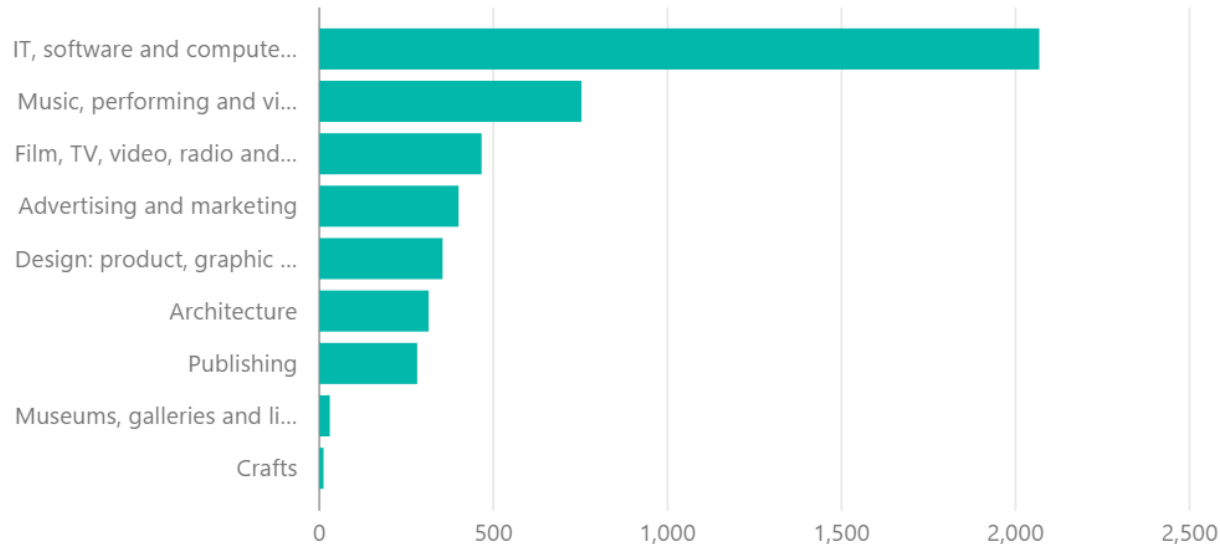
Sorting Data Result

Sorting instantly clarifies the relative order across the categories.

Sorting often helps find the story in the data.

Formatting the visualization
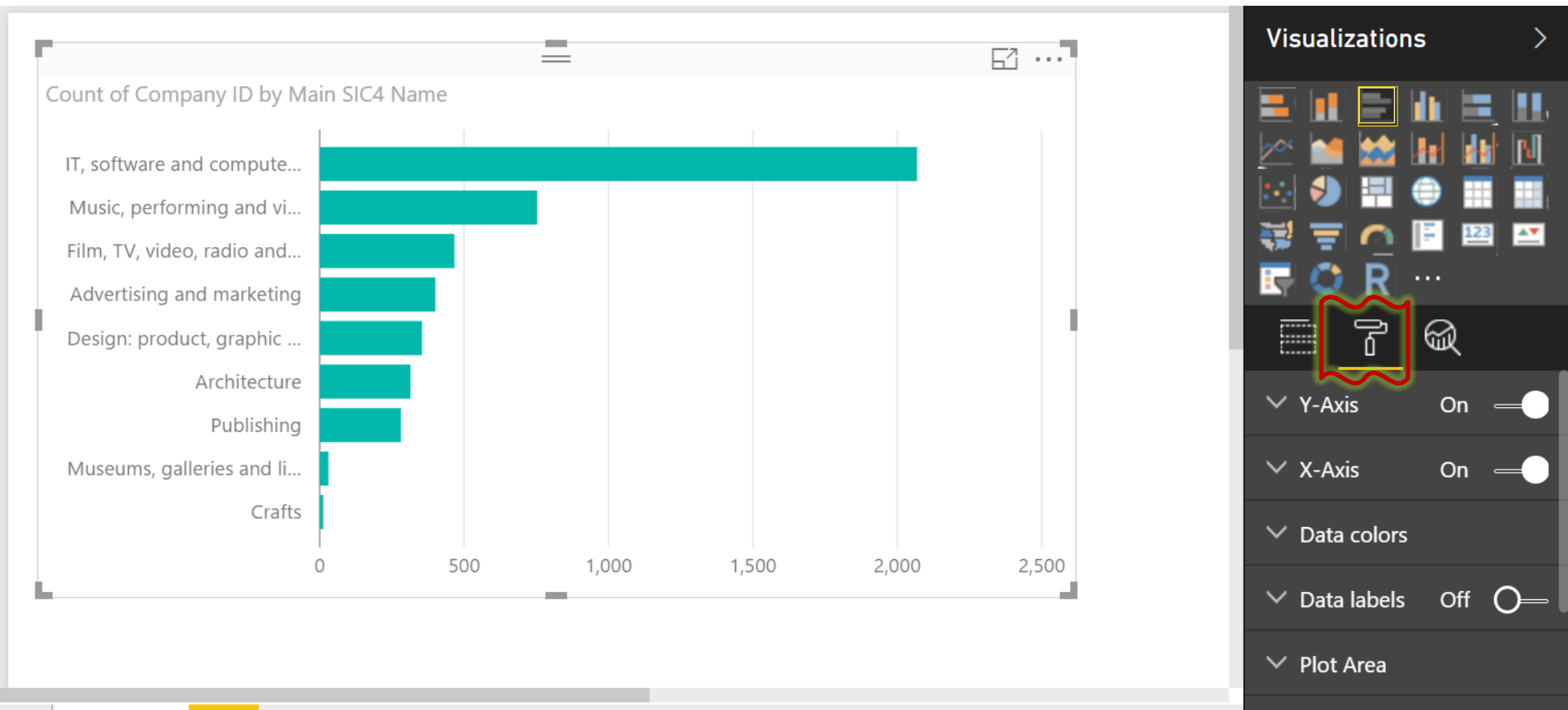

Now we can adjust how the visualization looks!
Click the background to deselect any visuals then click **Format**
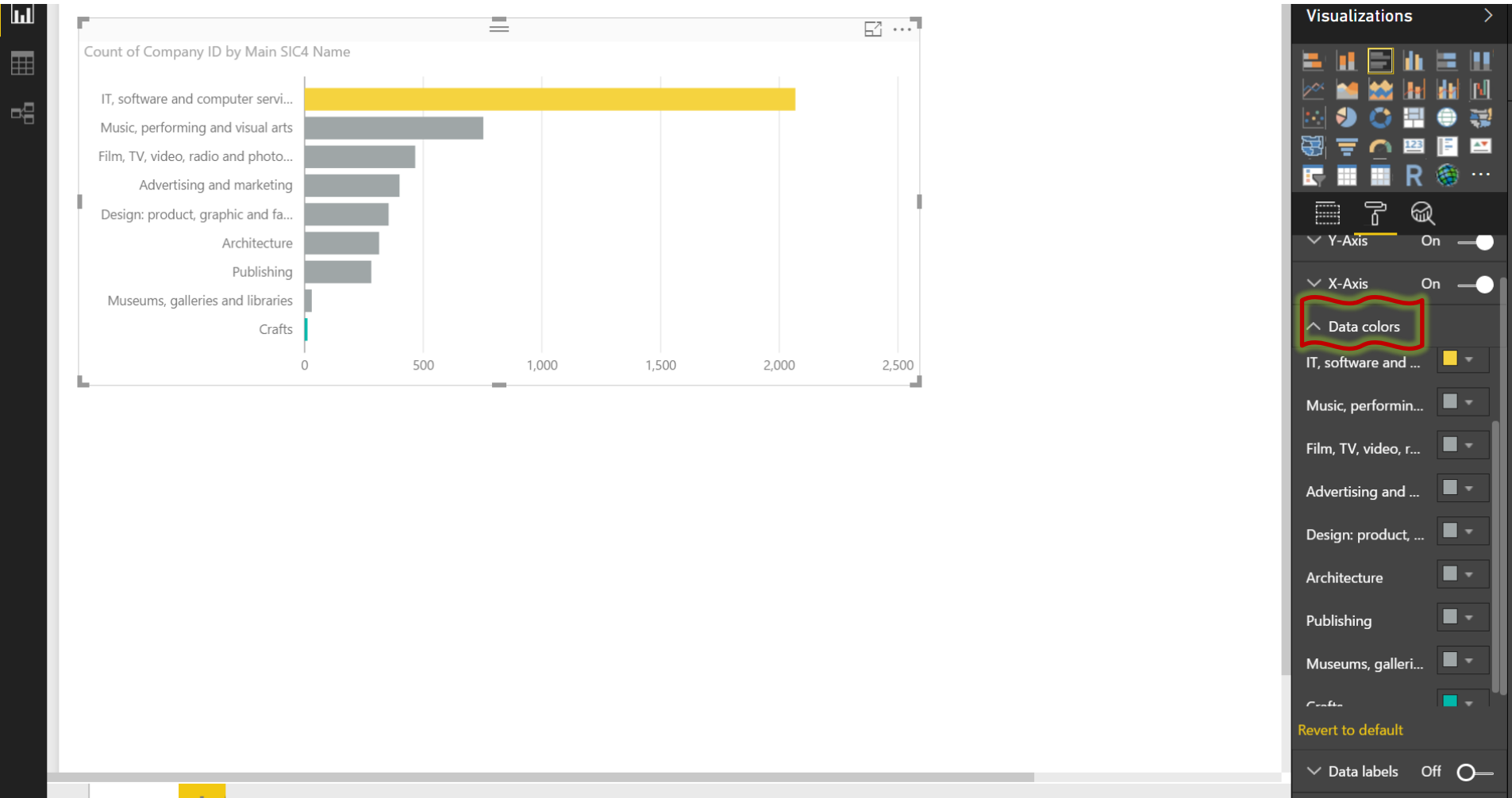
Formatting the visualization
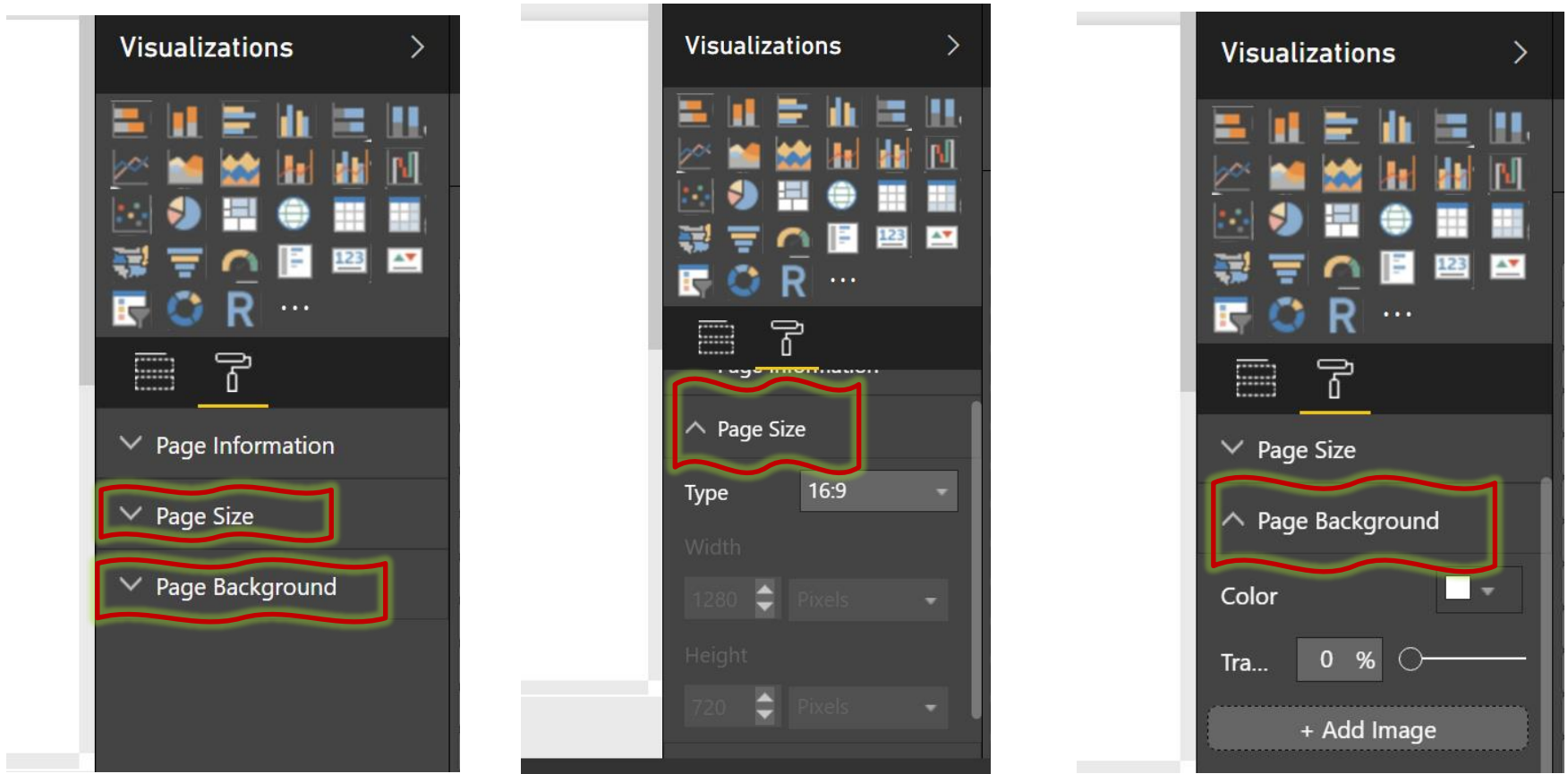
This shows standard settings for formatting the page, allowing you to adjust page size and fill the background.

Formatting the visualization

Now click to select the visual and again click **Format**
This brings up the format options for the selected visual.

Formatting the visualization

The **Data colors** menu lets you choose one colour per category.

Categorical Data Key Points

Categories are often alphanumeric labels for types/groups of data.

Bar charts are a good solution when there is a limited number of categories.

Key steps in creating a visual:
 Shape (wrangle) the data.
 Identify a field for the axis categories. (independent variable)
 Identify a field for the value. (dependent variable)
 Choose a visual type and then:
  Aggregate
  Filter
  Sort
  Format