Right Plot for Visualization

SATHYANARAYANAN RV

CONTENTS

Introduction

- Data viz is the communication of data in a visual manner, or turning raw data into insights that can be easily interpreted by your readers.
- The numeric statistics might give you the essence of your data, a graph or visualization can uncover a whole new dimension of underlying information within your dataset.
- When it comes to presenting your data, especially to clients, it's always good to use visualization tools that can help bring out the scope and purpose of your work.
- You wouldn't want just to show data files or code, rather a neat set of graphs to make your story seem more interesting and easily most grabbable.

Types of Visualizations

- Temporal
  - They are linear; they are single-dimensional in nature
- Hierarchical
  - They order groups within larger groups
- Network
  - Visualizations show how they relate to one another within a network
- Multidimensional
  - There are always 2 or more variables in this type
- Geospatial
  - Relates to real life physical locations, overlaying familiar maps with different data points

Temporal

Use a box plots for the following reasons:

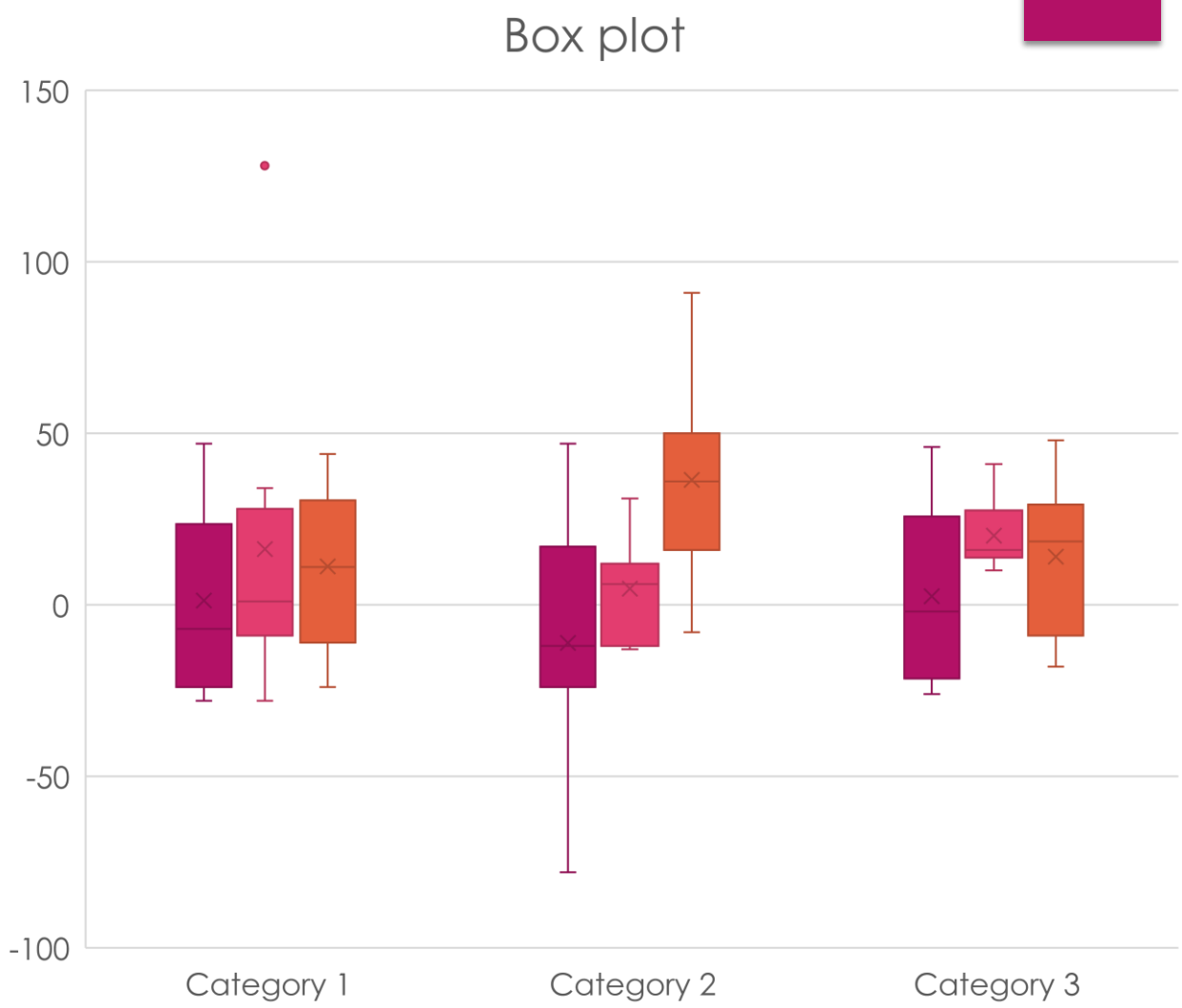
- To display or compare a distribution of data
- To identify the minimum, maximum and median of data

Don't use a box plots for the following reason:

- To visualize individual, unconnected data sets

Best practices for a box plots visualization

- Ensure font sizes for labels and legends are big enough and line widths are thick enough to understand the findings easily
- If plotting multiple datasets, use different symbols, line styles or color to differentiate each
- Always remove unnecessary clutter from the plots



Other Temporal Charts

Line Charts

Use a line chart for the following reasons:

- You want to understand trends, patterns, and fluctuations in your data
- You want to compare different yet related data sets with multiple series
- You want to make projections beyond your data

Don't use a line chart for the following reason:

- You want to demonstrate an in-depth view of your data

Spark lines

Use a sparkline for the following reasons:

- You can pair it with a metric that has a current status value tracked over a specific time period
- You want to show a specific trend behind a metric

Don't use a sparkline for the following reasons:

- You want to plot multiple series
- You want to illustrate precise data points (i.e. individual values)

Bar Charts

Use a bar chart for the following reasons:

- You want to compare two or more values in the same category
- You want to compare parts of a whole
- You don't have too many groups (less than 10 works best)
- You want to understand how multiple similar data sets relate to each other

Don't use a bar chart for the following reasons:

- The category you're visualizing only has one value associated with it
- You want to visualize continuous data

Hierarchical

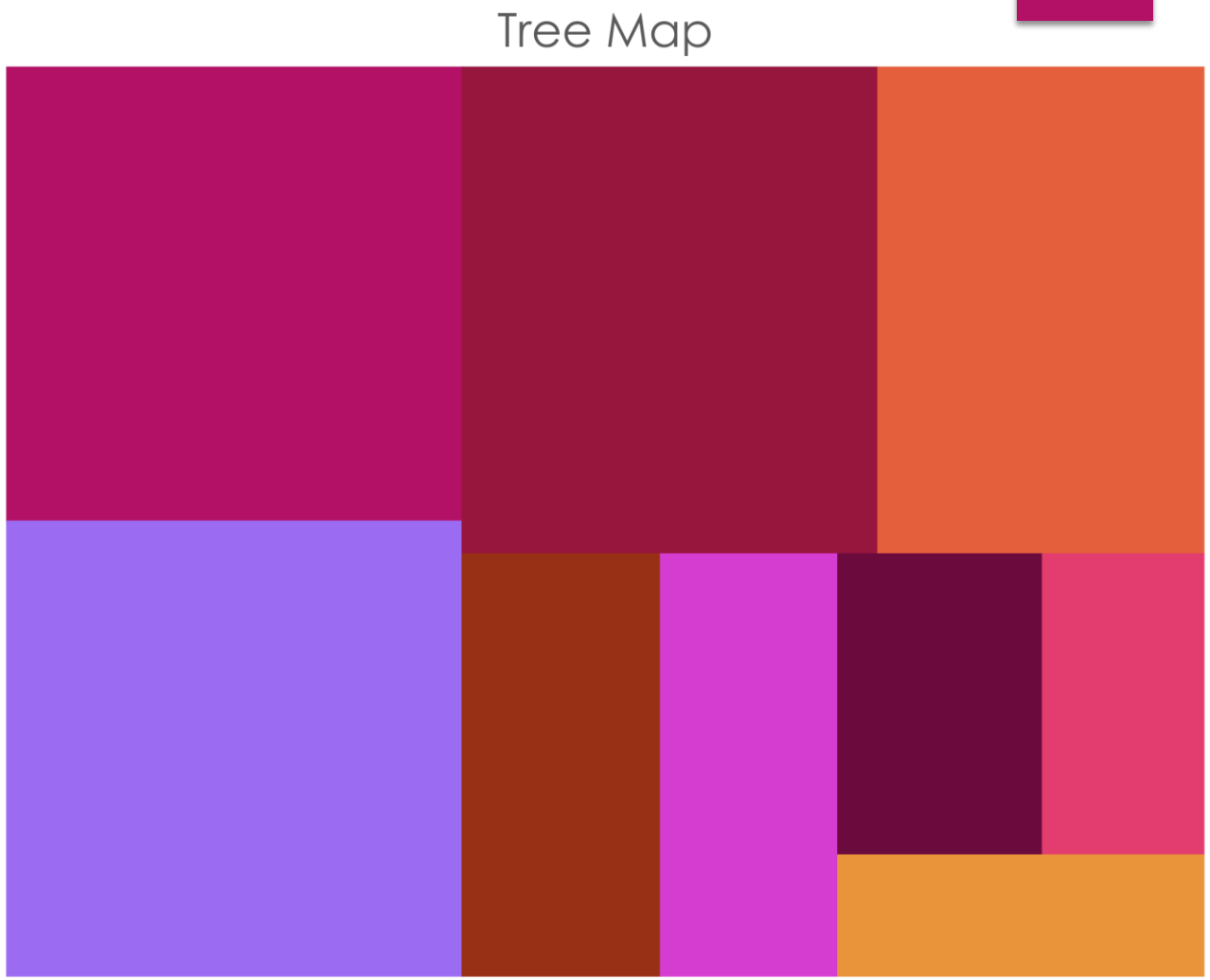
Use tree maps to display data in nested rectangles. You use dimensions to define the structure of the tree map, and measures to define the size or color of the individual rectangles. Tree maps are a relatively simple data visualization that can provide insight in a visually attractive format.

Use a tree maps for the following reasons:

- Comparing variables in categorical data.

Don't use a tree maps for the following reason:

- Non-categorical data.



Other Hierarchical Charts

Sunburst

- The sunburst charts are ideal for displaying hierarchical data. Each level of the hierarchy is represented by one ring or circle with the innermost circle as the top of the hierarchy.

When to use

- The sunburst charts are most effective at showing how one ring is broken into its contributing pieces.

When Not to use

- However, a sunburst charts with multiple levels of categories shows how the outer rings relate to the inner rings
- A sunburst charts without any hierarchical data (one level of categories), looks similar to a doughnut chart

Ring(Donut) Charts

- A single ring chart is much like a pie chart, except arrayed as a ring (or "donut").
- A multi-ring chart - with concentric rings of data - can display data distributions for multiple sub-groups simultaneously.

When to use:

- The primary use of a Ring Chart is to display the proportional relationship of various components to a whole, but for more than one data series simultaneously.

When not to use:

- Negative data cannot be identified unless marked. It is not advised to use a doughnut chart when negative values are included

Network

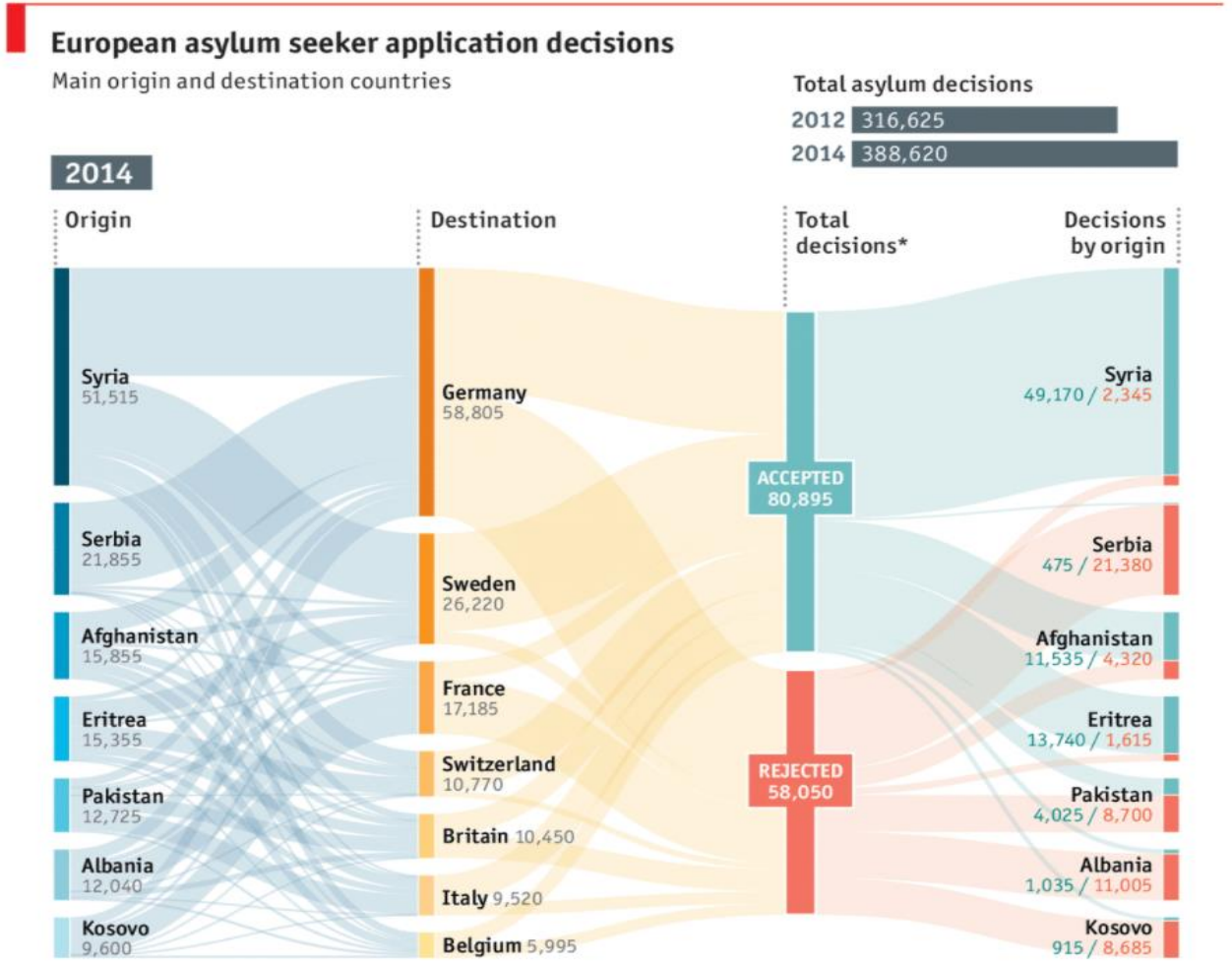
Network Visualization is about connecting multiple datasets with each other and showing how they relate with one another in a network where each variable is connected.

Alluvial diagrams are a type of flow diagram originally developed to represent changes in network structure over time. It is a variant of the Parallel Sets but for categorical variables and often to display trends over time and phases.

When to use

They can be used to visualize any type of change in group composition between states or over time and include statistical information to reveal significant change.

Alluvial Diagram



Multi-dimensional

Use a scatter plots for the following reasons:

- You want to show the relationship between two variables
- You want a compact data visualization

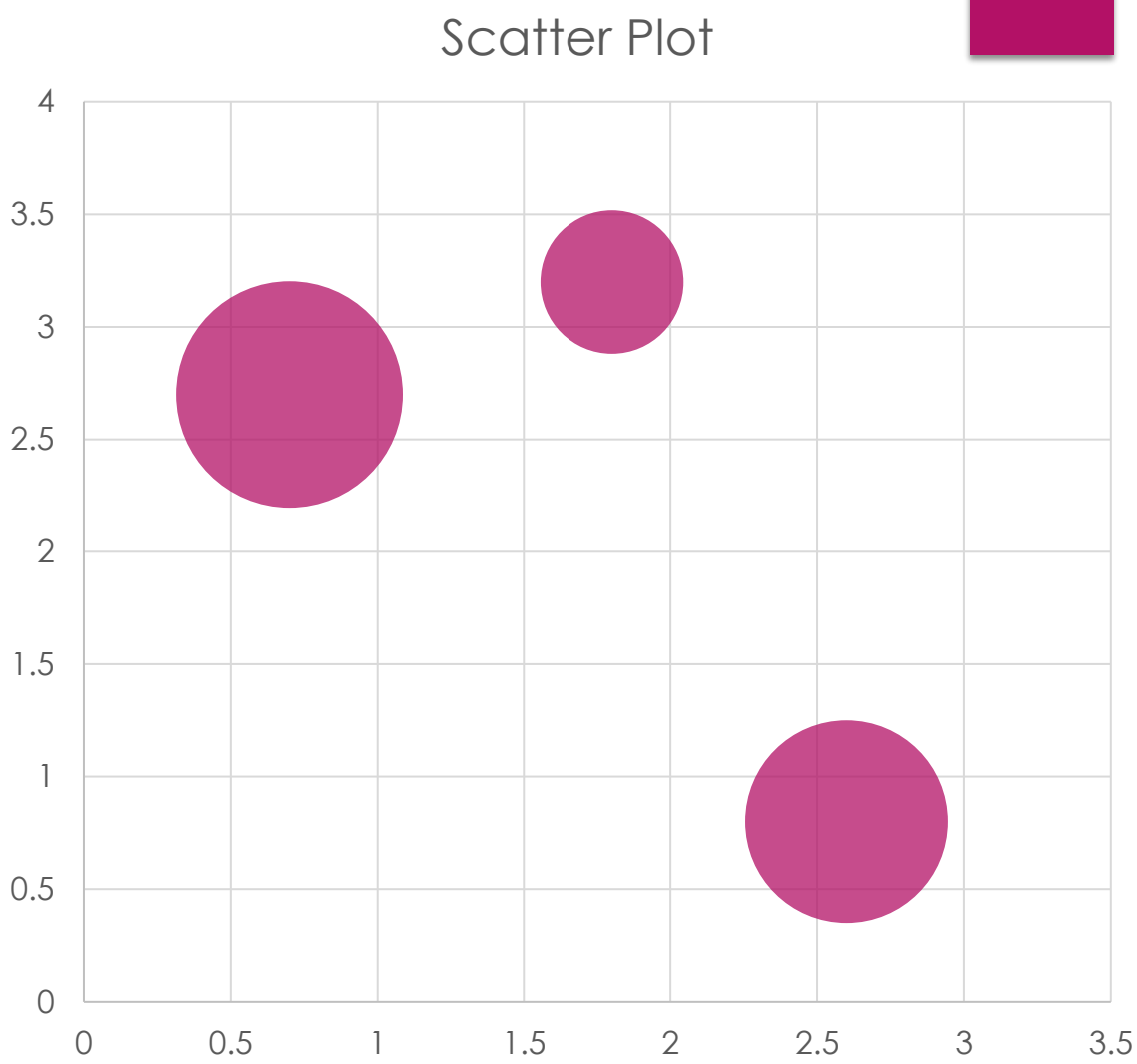
Don't use a scatter plots for the following reasons:

- You want to rapidly scan information
- You want clear and precise data points

Best practices for a scatter plots visualization

If you use a scatterplot, here are the key design best practices:

- Although trend lines are a great way to analyze the data on a scatterplot, ensure you stick to 1 or 2 trend lines to avoid confusion
- Don't forget to start at 0 for the y-axis



Other Multi-dimensional Charts

Pie Charts

Use a pie chart for the following reasons:

- You want to compare relative values
- You want to compare parts of a whole
- You want to rapidly scan metrics

Don't use a pie chart for the following reason:

- You want to precisely compare data

Make sure that the pie slices **add up to 100%**. To make this easier, add the numerical values and percentages to your pie chart

Stacked bar charts

Use a stacked bar chart for the following reasons:

- If the goal is to show sizes between individual categories, use a grouped column or bar chart.
- If the goal is to show the total sizes of groups, use a regular stacked bar chart.
- If the goal is to show relative differences within each group, use a stacked percentage column chart.

Don't use a stacked bar chart for the following reasons:

- If you have negative values then you should steer clear of this chart, as there is no way to show negative spaces in a positive bar

Histograms

Use a histogram for the following reasons:

- Histograms are great when we would like to show the distribution of the data we are working with. This allows us to group continuous data into bins and hence, provide a useful representation of where observations are concentrated.

Don't use a histogram for the following reasons:

- Be careful when the data you are working with contains multiple categories or variables. Multi-column histograms are among the chart types to be avoided when they look like this.

Geospatial

Use a map for the following reason:

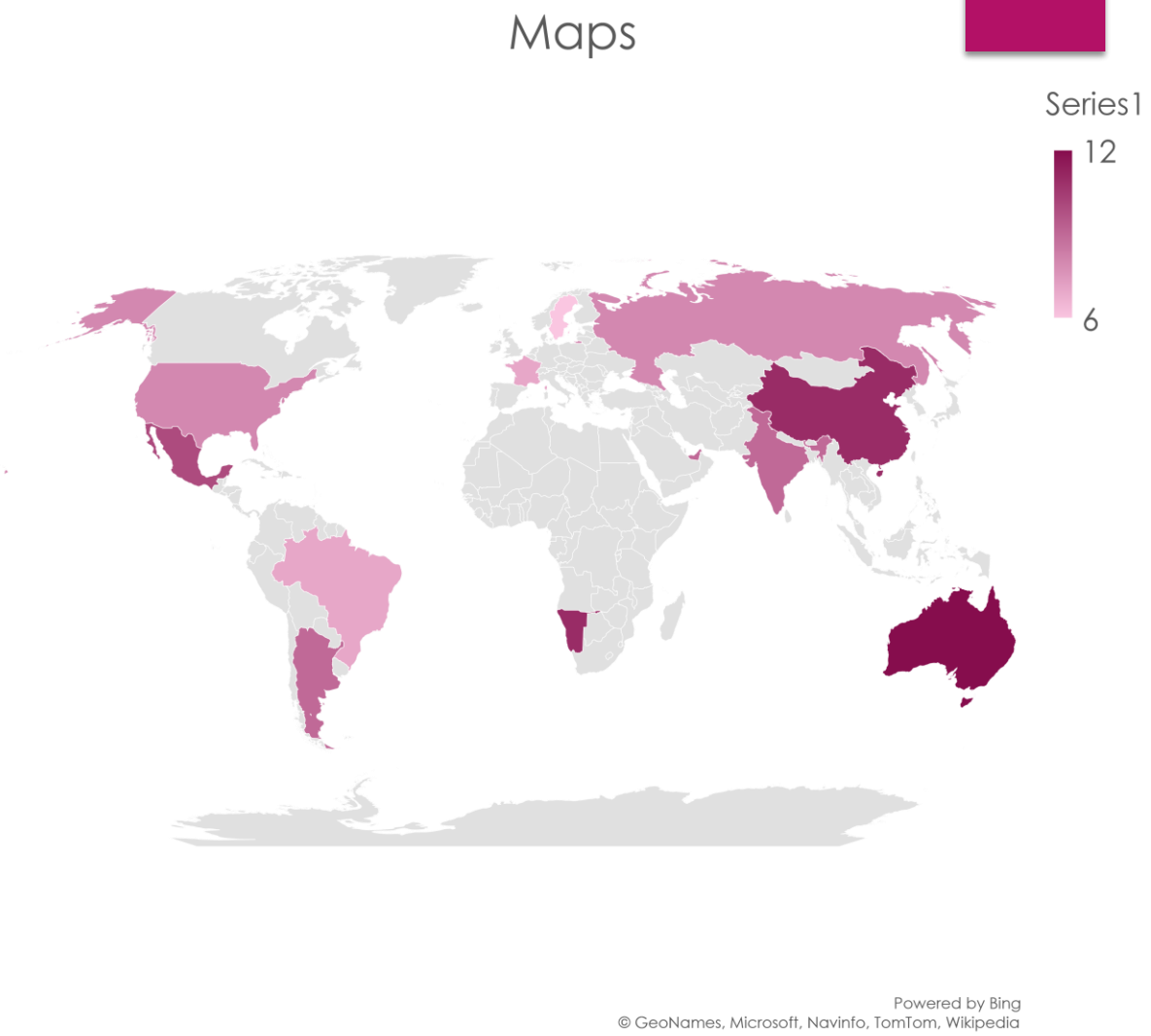
- Geography is an important part of your data story

Don't use a map for the following reasons:

- You want to show precise data points
- Geography is not an important element of the dashboard's overarching story

Best practices for a map visualization

- Avoid using multiple colors and patterns on your map. Use varying shades of the same color instead
- Make sure to include a legend with your map, so that everyone understands what the data means



Other Geo Spatial Charts

Geo Heat Maps

- **Use a heat map for the following reasons:**
- To show a relationship between two measures
- To illustrate an important detail
- To use a rating system
- **Don't use a heat map for the following reason:**
- To visualize individual, unconnected metrics

Choropleth Maps

- **Use a choropleth map when your data**
  - Attached to enumeration units (e.g., counties, provinces, countries)
  - Standardized to show rates or ratios (never use choropleth with raw data/counts), and
  - Have a continuous statistical surface, in other words, you could conceptually measure the phenomena anywhere in space (n.b. 'zero' is still a valid measurement)

ThAnK yOu