Data Visualisation

CMP020L013A

Week 6: Chart Types Guide

Mohammad Javaheri
(Dr Mohammad Ali Javaheri Javid)

Chart Types

- Selecting the right visualisation to present data is complicated:
  - the number of chart choices can distract you from the goal of communicating the key insight
  - Each type of chart is designed to show a type of data in a particular way

Comparisons of Categories and Time

- *Questions:*
  - 1. What's the best? What's the worst? Compared to what?
  - 2. Who's ranked the highest? The lowest?
  - 3. How does performance compare to the target or goal? For example, did total sales exceed the forecast?

Comparisons of Categories and Time

- **Insight:**
  - use comparisons to illustrate the similarities and differences among categories. This includes the minimum value, maximum value, rank, performance, sum, totals, counts, and quantities.
- **Data:**
  - aggregated categorical data, such as the number of books sold by author. Time series data can be shown as a categorical variable. For example, each year can be a category.
- **Chart options:**
  - vertical bar, column bar, horizontal bar, and bullet charts.

Chart types to present categorical data

- **Vertical Bar**
- Bars are arranged vertically on the x-axis. Each bar represents a category or sub-category. The bar height measures the quantity (count) or sum.
  - Keep bars the same color and shade when they measure the same variable
  - Use a zero baseline for the y-axis.
  - Show negative values below the baseline.
  - Keep the width of the bar about twice the width of the space between the bars.

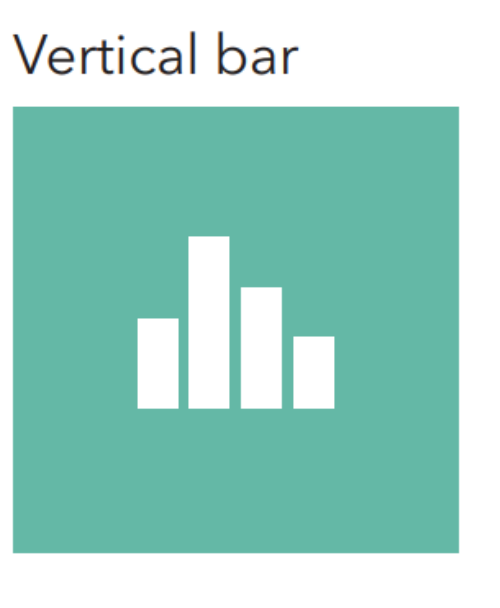


Chart types to present categorical data

- **Column Bar**
- Column bar charts present two series for each category.
- Use different color shading for each series.
- Shade bars from lightest to darkest.

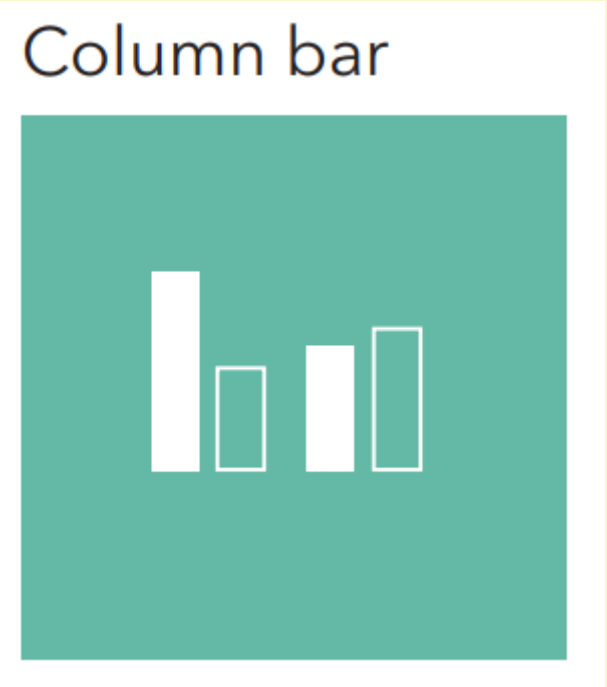


Chart types to present categorical data

- **Horizontal Bar**
- Bars are arranged horizontally, rather than vertically.
- Best used for ranking, such as first place, second place, third place.
- Arrange bars in descending order, from largest to smallest.

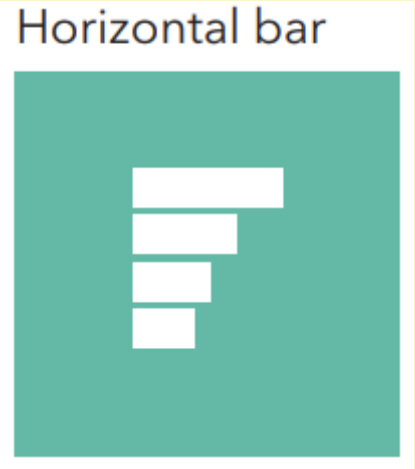
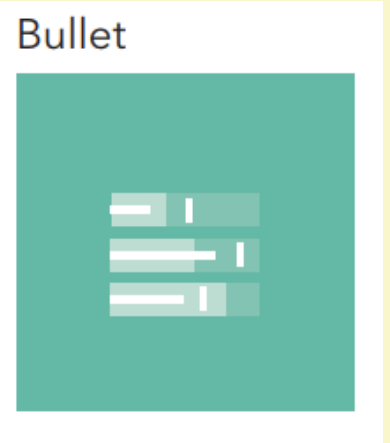


Chart types to present categorical data



- **Bullet**
- Bullet charts display performance of a variable as a horizontal bar compared to a target or goal, represented by a vertical line.
- For example, a bullet chart could show whether the actual sales for a given period(s) are above/ below target sales.
- The performance measure (horizontal bar) overlays several shaded rectangles that represent qualitative ranges (e.g., 40% to the target goal, to indicate the performance progress).

Distributions

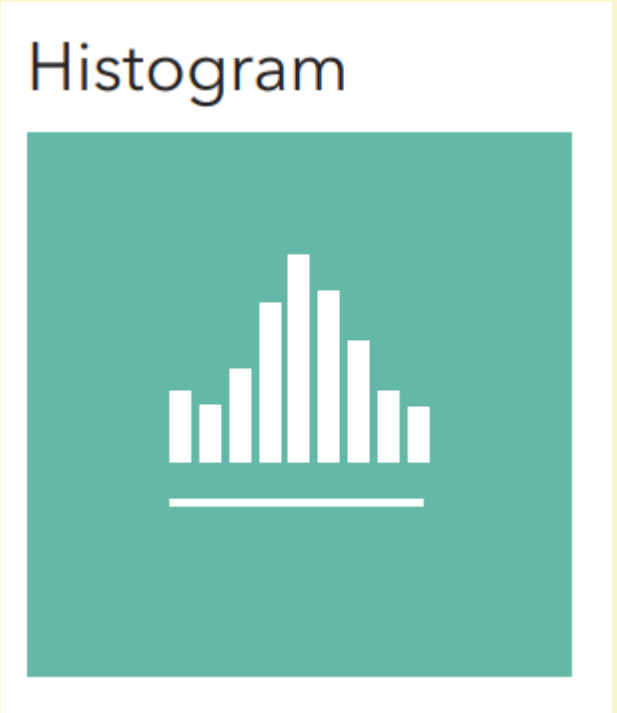
- For showing possible values (or intervals) of the data and how often they occur. These types of charts can reveal the minimum and maximum values, median, outliers, median, frequency, and probability densities.
- *Questions:*
- 1. What are the highest, middle, and lowest values?
- 2. Does one thing stand out from the rest?
- 3. What does the shape of the data look like?

Distributions

- **Insight:**
- use to distributions charts reveal outliers, the shape of the distribution, frequencies, range of values, minimum value, maximum value, and the median.
- **Data:** univariate or a single numeric variable.
- **Chart options:**
  - histogram, density plot, and a boxplot.

Distributions

- **Histogram**
- Histograms show frequencies of a single variable grouped into bins or frequency ranges on the x-axis.
- The y-axis of the histogram shows the frequency count or percentage.
- A large bin size can obscure the data.
- Adjust the size of the bins to best reveal the shape of the frequency distribution.

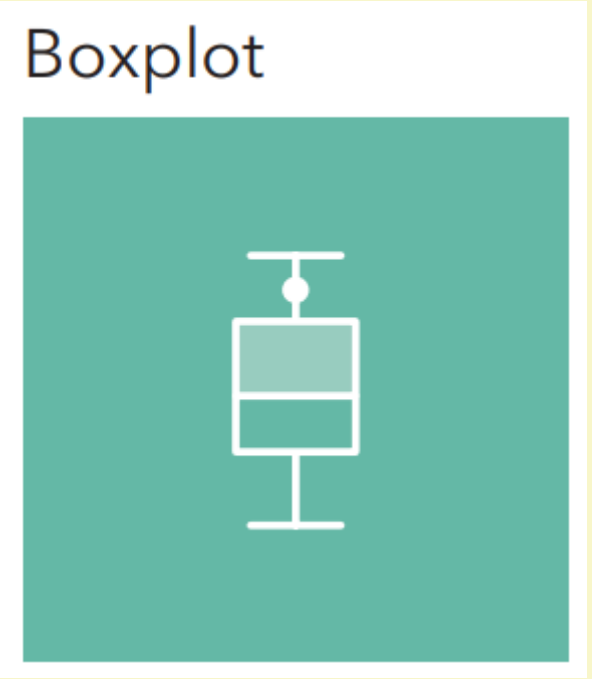


Distributions

- **Density Plot**
- Density plots show probability densities and the distribution of a single variable. The area under the curve emphasizes the shape of the distribution of data.
- Annotate the mean to draw attention to the center of the distribution.

Distributions

- **Boxplot**
- Boxplots show the range of a single variable including the minimum, 25th percentile, 50th percentile, median (not the mean), 75th percentile, and the maximum value.
- Boxplots are helpful to spot outliers.



Proportions

- Presents options for displaying individual parts of a whole.
- This enables comparisons among subcategories by evaluating relative proportions, for example, demographics by neighborhood.
- *Questions:*
- 1. What are the parts that make up the whole?
- 2. What part is the largest or smallest?
- 3. What parts are similar or dissimilar?

Proportions

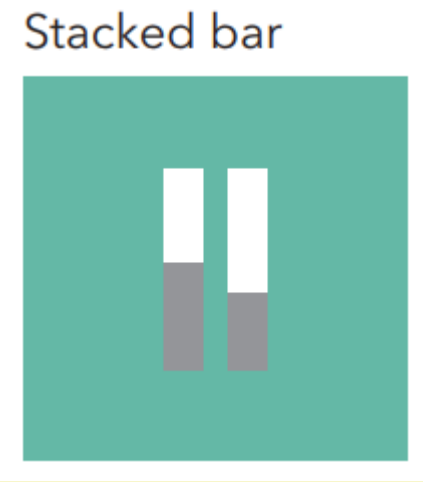
- **Insight:** use to show summaries, similarities, anomalies, percentage related to the whole (by category, subcategory, and over time).
- **Data:** single categorical variable with subcategories, two or more variables. A time dimension can also be included.
- **Chart options:**
  - pie, stacked bar, stacked area, tree map, and doughnut chart.

Proportions

- **Pie Chart**
- Pie charts show proportions within a whole. The slices are subcategories of a single category. Slices add up to 100% or 1.
- Avoid using pie charts if all the slices are similar in size.
- Limit pie charts to eight slices or less.
- Label directly on the pie slices, rather than using a legend.
- Keep pie slices the same color. Use the whitespace between slices to differentiate the slices.

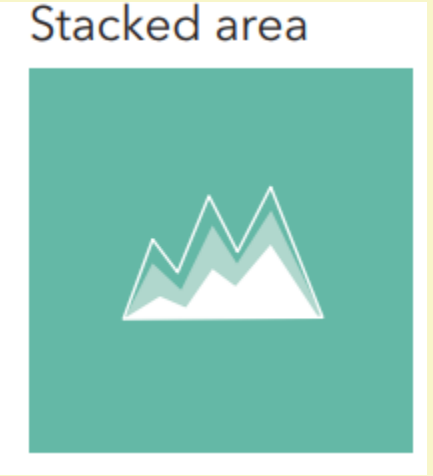
Proportions

- **Stacked Bar**
- Stacked bar charts show proportions and quantities within a whole category. They show absolute and relative differences.
- Limit the number of subcategories to four or less.
- Use stacked bars that add up to 100% to show the relative differences between quantities within each group



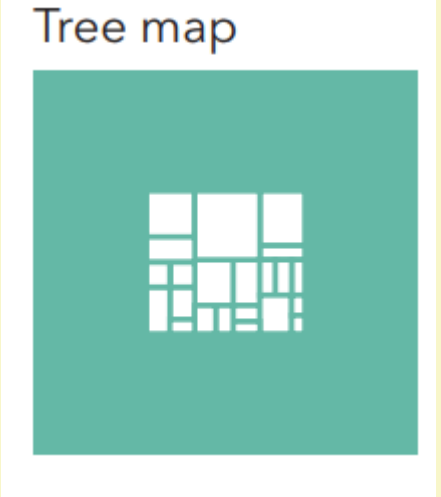
Proportions

- **Stacked Area**
- Stacked area charts highlight the absolute and relative differences between two or more series. They are line charts with the area below the line filled in with color.
- To show relative differences use a 100% stacked area chart. Label each series directly, if possible over using a legend



Proportions

- **Tree Map**
  - Tree maps show parts of the whole by using nested rectangles. Each rectangle is designated a size and a shade of a color.
  - This enables to emphasize both the importance (usually shown by size) and urgency (usually represented by color) of a data point.
  - Used often for portfolio analysis to highlight similarities and anomalies.
  - Usually require interactivity such as mouse-over, to read the subcategory labels for the smallest rectangles.
  - This chart type is best used for analysis and exploration rather than presentation.



Proportions

- **Doughnut Charts**
- Doughnut charts present proportions of a whole through slices of a doughnut shaped graphic It is just a pie chart with the center missing.
- This type of chart can contain multiple series, represented as doughnuts arranged inside one another.

Relationships

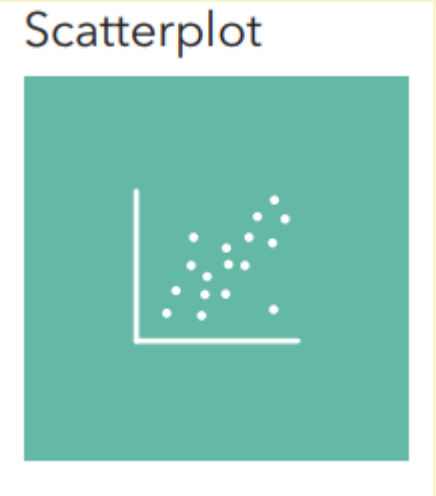
- Presents options for displaying multivariate data.
- These charts show how one or more variables relates to other variables.
- For example, how do sales affect profitability by region?

Relationships

- ***Questions:***
  - 1. Is the relationship positive, negative, or neither?
  - 2. How are x and y related to each other?
  - 3. What makes one group or cluster different from another?
- **Insight:** use to show outliers, correlations, positive, and negative relationships among two or more variables.
- **Data:** two or more numeric variables.
- **Chart options:**
  - scatterplot, scatterplot matrix, bubble, parallel coordinates, radar, bullet, and a heat map

Relationships

- **Scatterplot**
- Scatterplots show relationships between two variables. For example, they show the change in x given y.
- Use to show positive or negative correlations, or linear and nonlinear relationships between variables.
- Labeling of every data point reduces readability but increases interpretation.

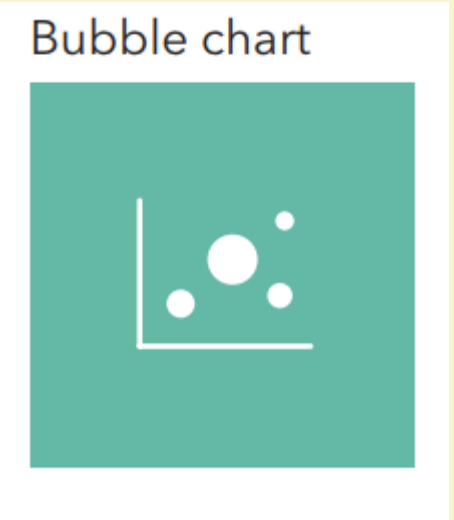


Relationships

- **Scatterplot Matrix**
- Scatterplot matrices help identify a correlation between multiple variables. It makes it easy to observe the relationship between pairs of variables in one set of plots.
- This chart type is best reserved for exploration versus presentation.

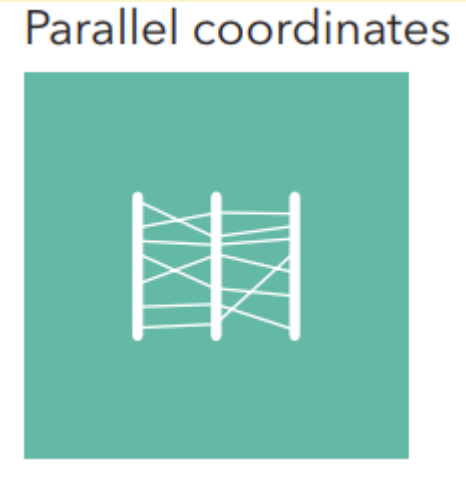
Relationships

- **Bubble Chart**
- A bubble chart is a scatterplot that shows relationships between three or four variables.
- The position of the bubble shows the relationship between the x and y variables.
- The bubble size is based upon a numerical variable, such as population, or sales.
- The bubble color is best reserved for categorical data, such as region.
- Bubble charts are best when the bubble sizes vary significantly.



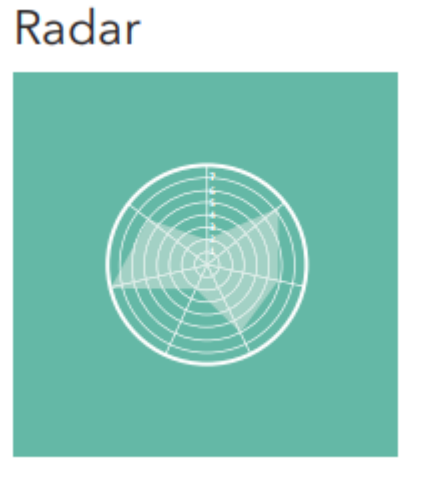
Relationships

- **Parallel Coordinates**
- Parallel coordinates map each column in a data table as a vertical parallel line with its own axis. Each observation (row) is represented by a point on the parallel line. That point is then connected to the next point on the next parallel line by a horizontal line.
- Use the technique of highlighting the lines that touch any number of values in either of the categories, called brushing, to provide data context while focusing on select series.
- This chart type is best reserved for exploration over presentation.



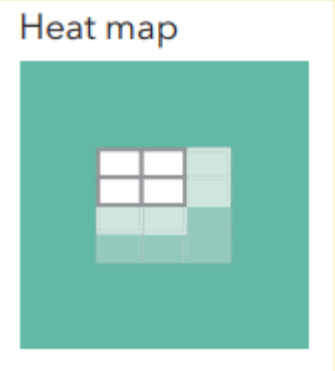
Relationships

- **Radar**
- Radar charts compare multiple numerical variables.
- They show which variables have similar values, and to spot outliers, high values, and low values. Each variable is provided its own individual axis, but the axes are arranged radially. Every observation connects to form a shaded polygon.
- Limit the number of variables to reduce the number of axes to increase readability.
- Scaling is affected when variables have dissimilar minimum and maximum ranges.



Relationships

- **Heat Map**
- A heat map is a graphical representation of a table of data.
- The individual values are arranged in a table/matrix and represented by colors.
- Use grayscale or gradient for coloring.
- Sorting of the variables changes the color pattern.



Relationships

- **Small Multiples**
- A series of similar graphs that use the same scale. This allows for easy comparisons between variables.
- A single chart represents each categorical variable, such as sales personnel; the individual charts are grouped together on a single display.
- Allows easy comparisons by using the same scale for each chart
- Avoid showing too much detail in any individual chart.

Locations

- Geospatial visualizations require data that corresponds to geography (e.g., latitude and longitude).
- Presents the options for showing these location types.
- These geospatial displays can identify places, population differences, concentrations, and distances.

Locations

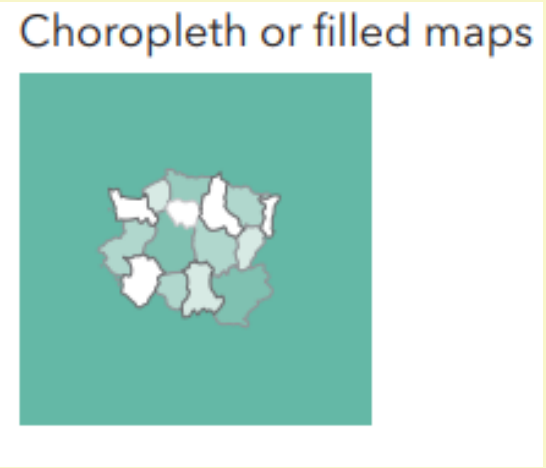
- *Questions:*
  - 1. Where can the most or least be found?
  - 2. How does one area compare to another?
  - 3. What is the distance from one place to another?
  - 4. How does a variable change by location?
- **Insight:** use to demonstrate similarities and differences by location, density, distance, and counts (such as population).
- **Data:** latitude and longitude, zip codes, census tracks, cities, states, countries, and regions.

Locations

- **Chart options:**
- use a choropleth-filled map for comparing regions; bubble or point maps to mark a location while showing quantities or categories; connection maps to show distances;
- and isopleth maps to show variables, such as temperature, that are unrestrained by boundaries (e.g., city, state, zip code)

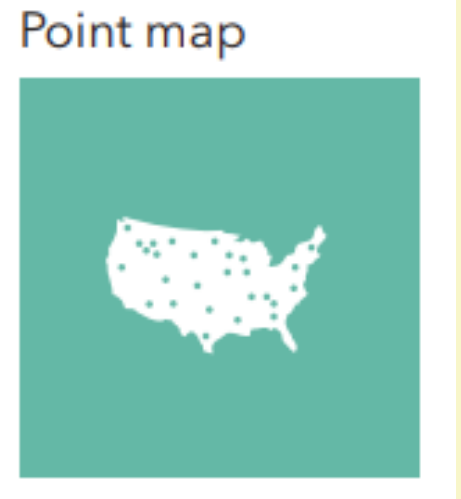
Locations

- **Choropleth (filled maps)**
- Choropleth maps fill regions with color. A color gradient and density distinguishes regions from one another.
- Use to compare different regions such as continents, countries, states, territories, zip codes, or census tracks.
- Provide a legend. Keep the gradient of colors within a limited range. This will allow the reader to easily compare the regions.



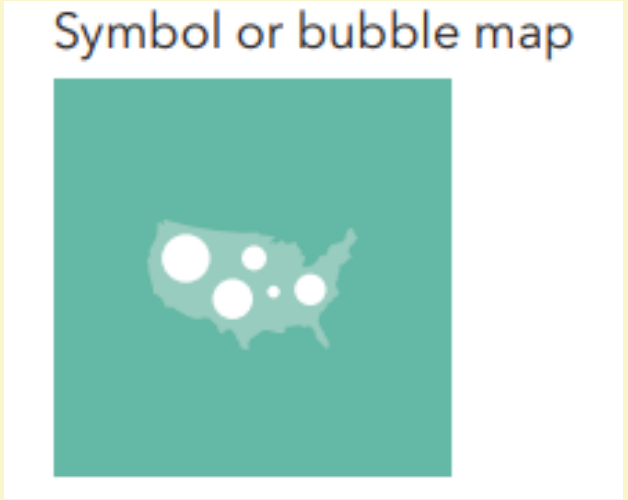
Locations

- **Point Map**
- Point maps show a specific location. These dots can vary in size, form, or color.
- Point maps illustrate density when the individual locations are easily distinguishable.
- Too many points can obscure the location.
- Consider the size of the points and the labeling of the points.



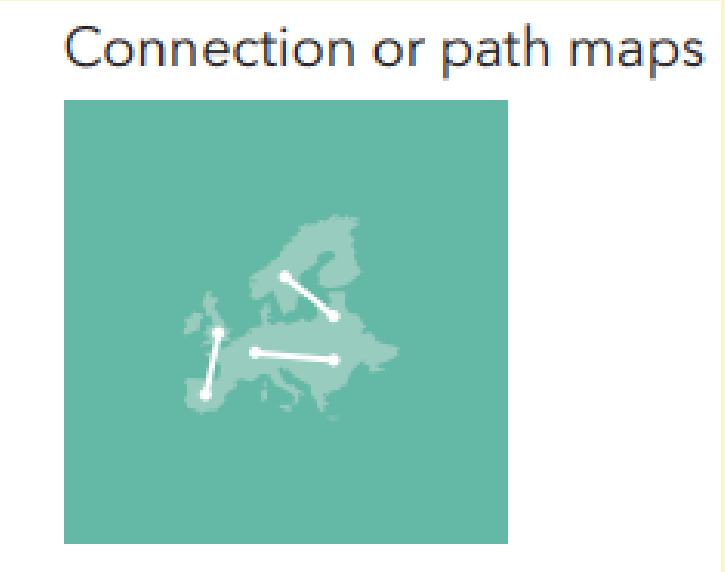
Locations

- **Symbol or Bubble Map**
- Symbol maps are point maps that use different sized bubbles or shapes to mark a location.
- These symbols are sized by a certain variable.
- Too many or too large bubbles can obscure the locations referenced.



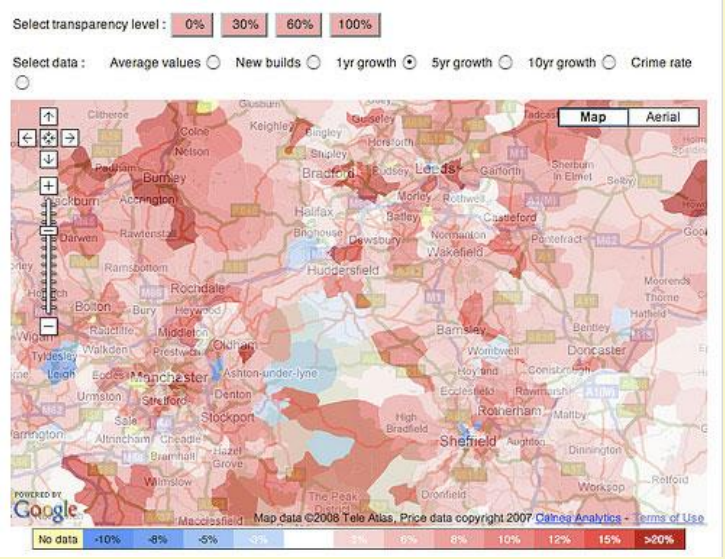
Locations

- **Connection or Path Maps**
- Connection maps graph a line from one or more points to another. Use to show distances or pathways between one or more locations.
- Use high contrasting colors for the map projection and the lines that connect the points.
- Avoid too many overlapping lines.



Locations

- Geographic Heat Map (Isopleth)
- Isopleth maps show gradual change over geography. This technique uses a color value (lightness/darkness) and hue to show density.
- The color value is not constrained by boundary lines (e.g., such as zip code).
- Use for events that are continuous and unbounded (e.g., such as temperature).



- There are many available resources to guide in determining the right chart for your data.
  - Choosing A Good Chart
  - The Data Visualisation Catalogue