

Module 1B:

From Question to Answer

Brandon NG
brandon.ng@nus.edu.sg

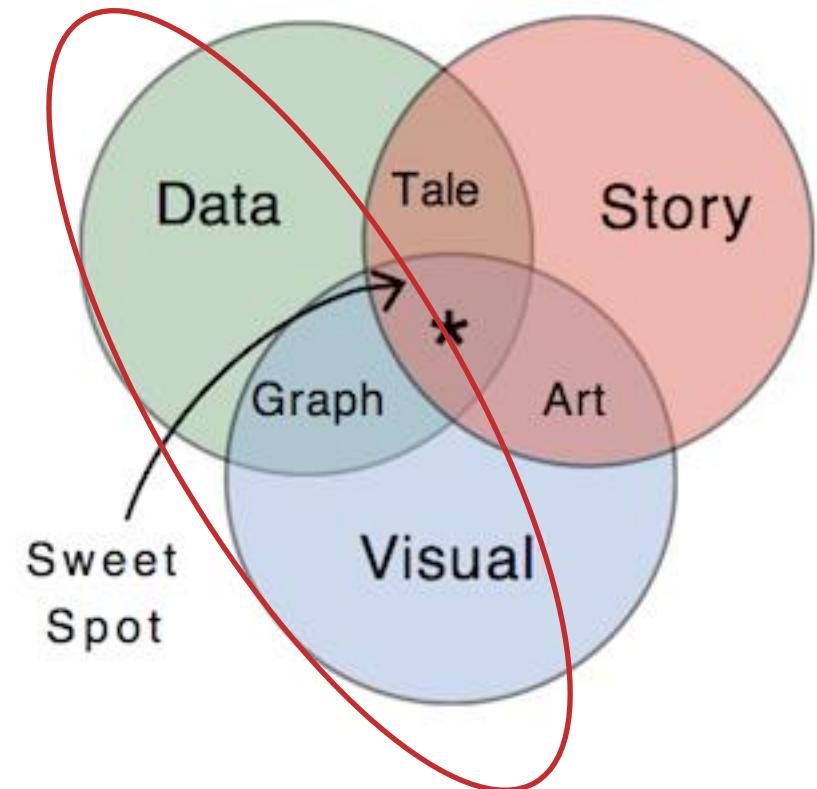
Institute of Systems Science
National University of Singapore



© 2018 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of NUS ISS, other than for the purpose for which it has been supplied.

Introduction

- Type of Analytics Patterns/Graphs for Different type of Question
 - Analytical Patterns
 - Method Classification
- Characters = Charts + Metrics
- Implementation using visualization tools

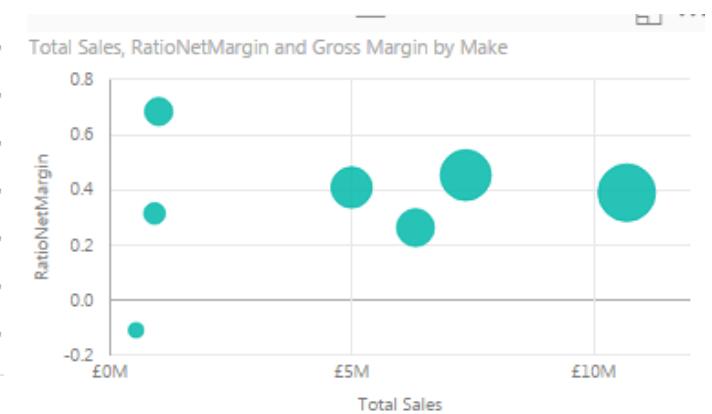
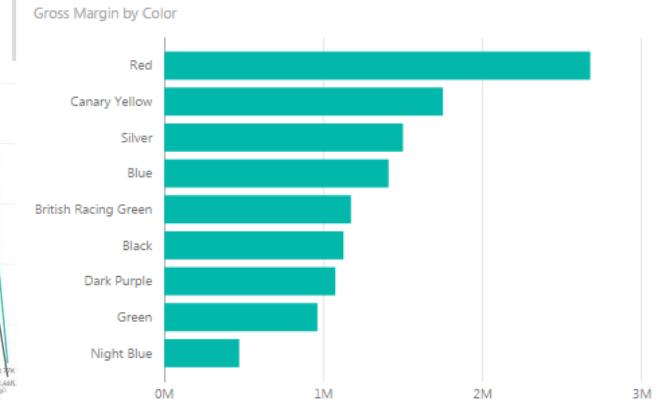
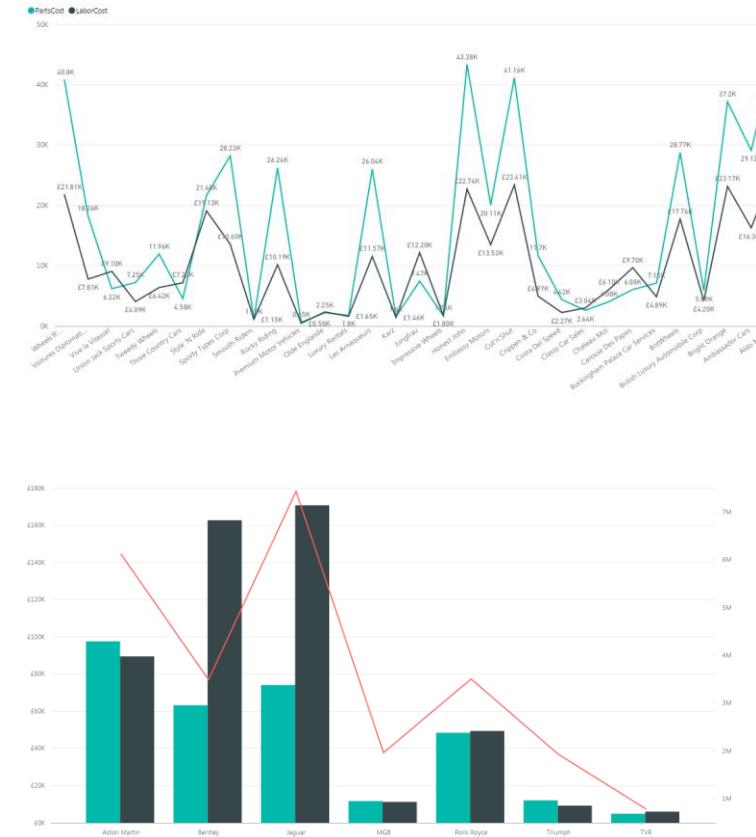
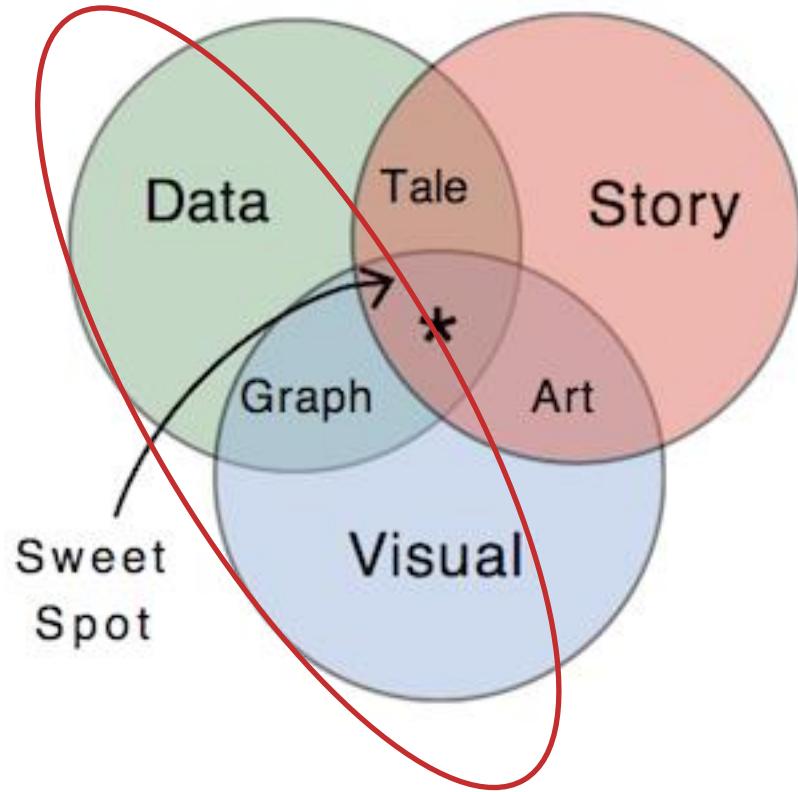




Types of Analytics Patterns/Graphs for Different Types of Question

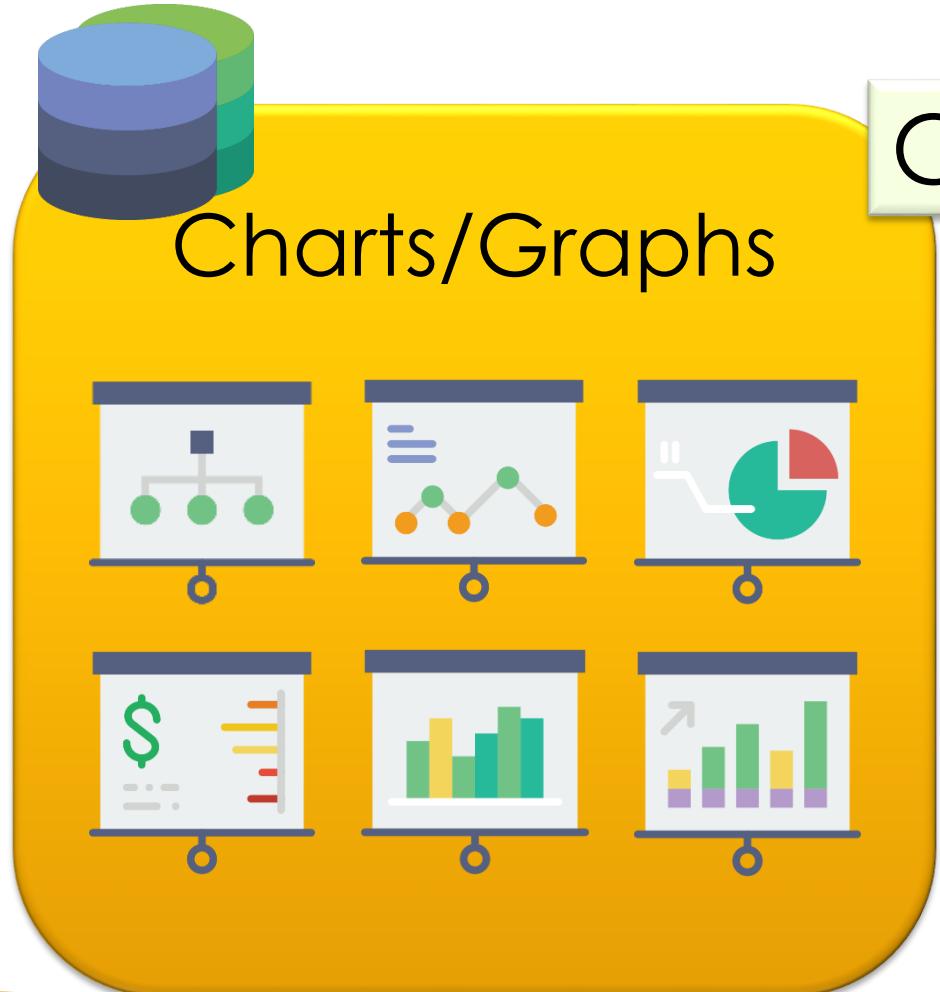
Visualization gives you answers to questions you didn't know you had
- Ben Schneiderman

Data + Visual = Graph



Characters : Charts and Metrics

data to be represented & presented with **Charts/Graphs** &
to be measured with **Metrics**



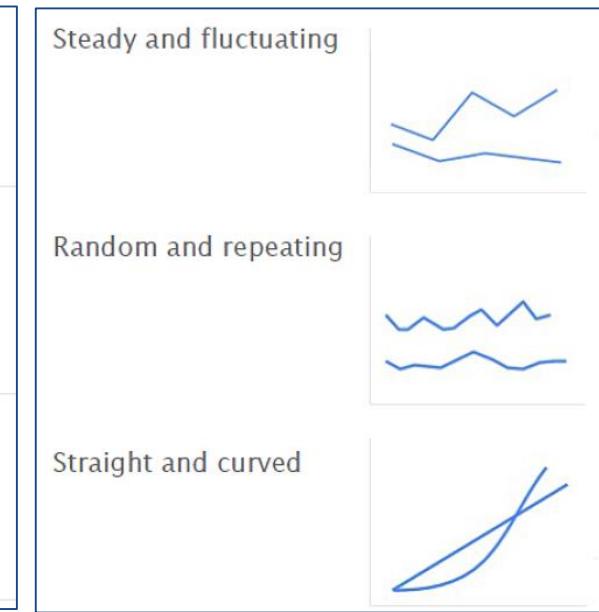
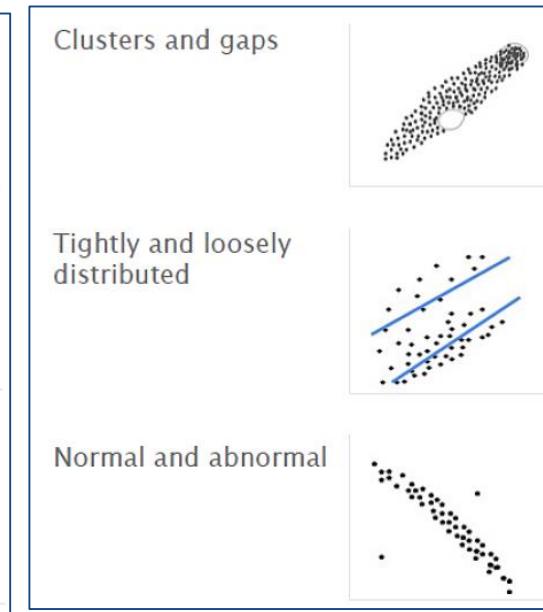
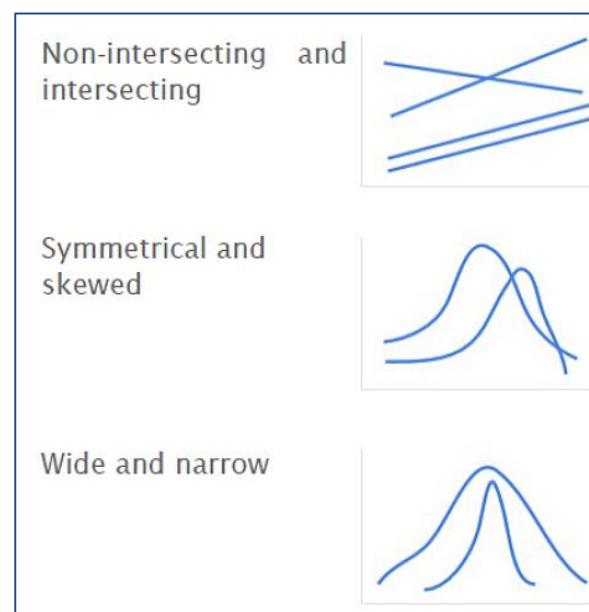
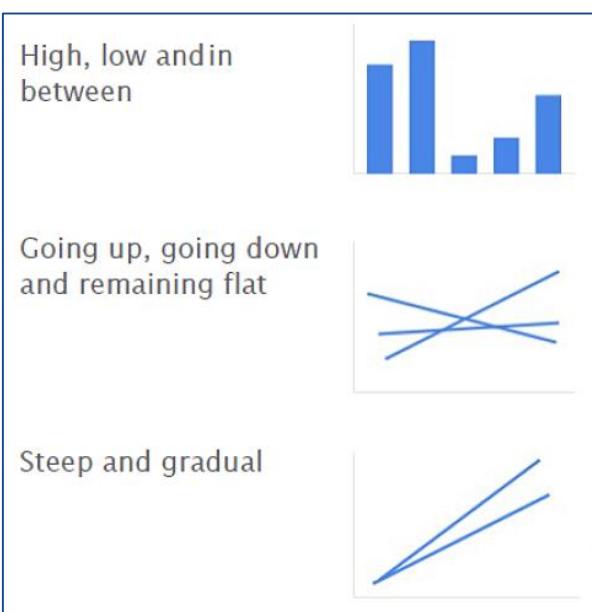
Characters



Analytical Patterns

We can use 'Pre-Attentive Attributes' to form 'Analytical Patterns'...

The 'Analytical Patterns' provide us with a visual language that allows us to communicate information visually.



Method Classification



Data summarisation



Comparing categories



Assessing hierarchies and part-to-whole relationships



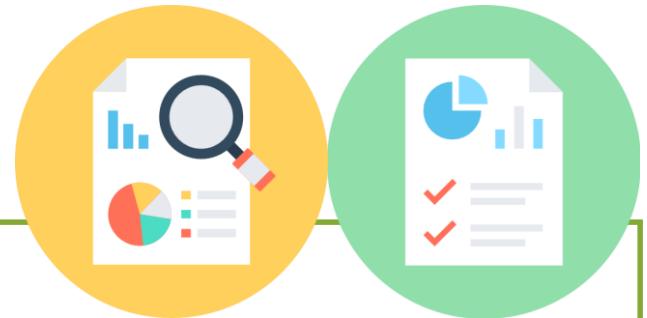
Showing changes over time



Plotting connections and relationships



Mapping geo-spatial data



Data Summarisation

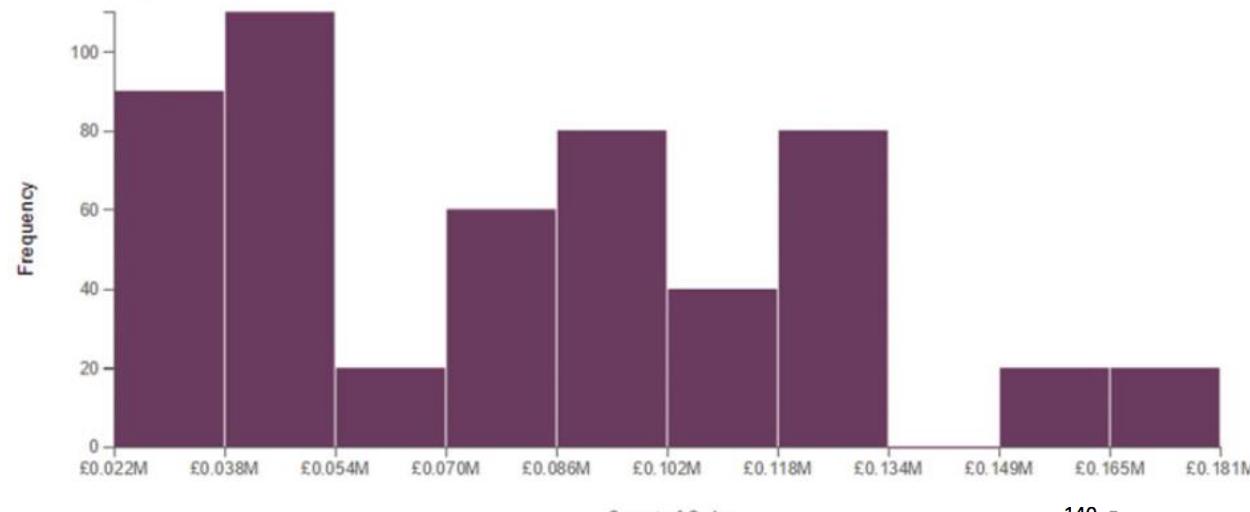
The following examples present chart types that facilitate the overview of data.

Histogram

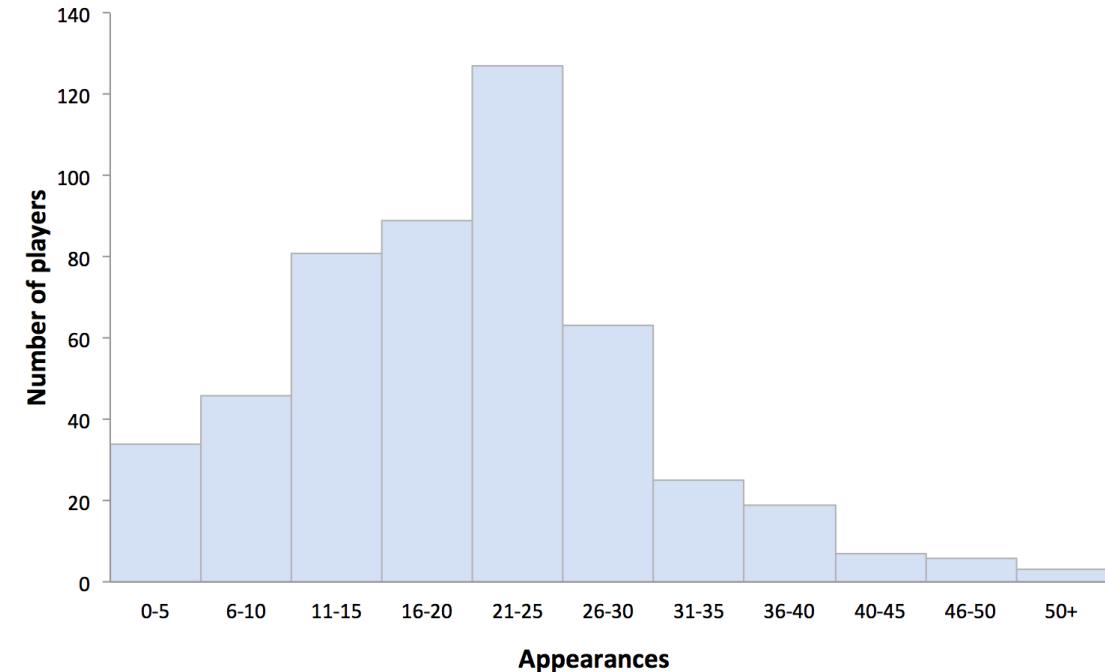
- Used for showing **distribution** of your data and makes categorizing easier.
 - Bin size is important:
 - Too small – you get spurious patterns
 - Too large – you might miss important patterns
- Histogram vs. Bar Chart
 - Bar Chart is used to display **categorical** data
 - Histogram is used to display a range of **numeric** values

Histogram

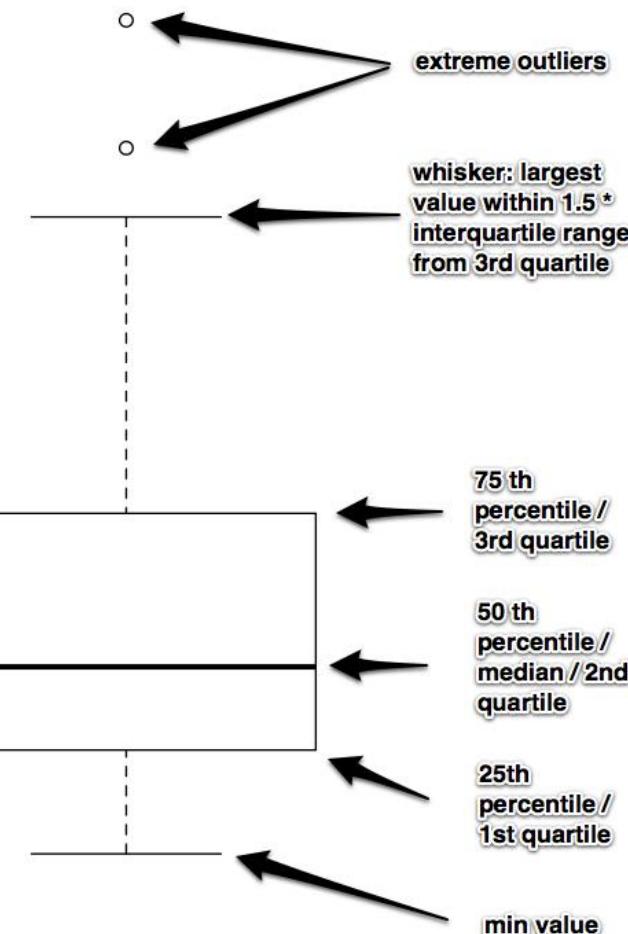
Count of Color by SalePrice



Count of Color



Word Cloud vs Box Plot



The five-number summary can be represented graphically using a boxplot

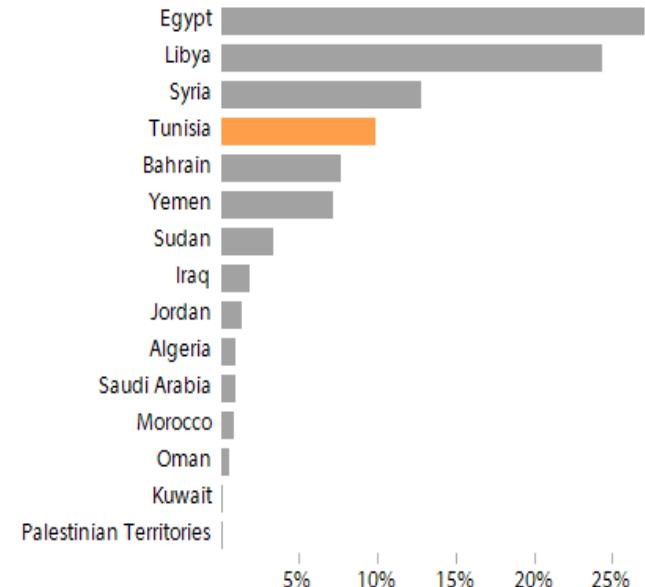


Comparing Categories

The following examples present chart types that facilitate the comparison of categorical values.

Bar Chart

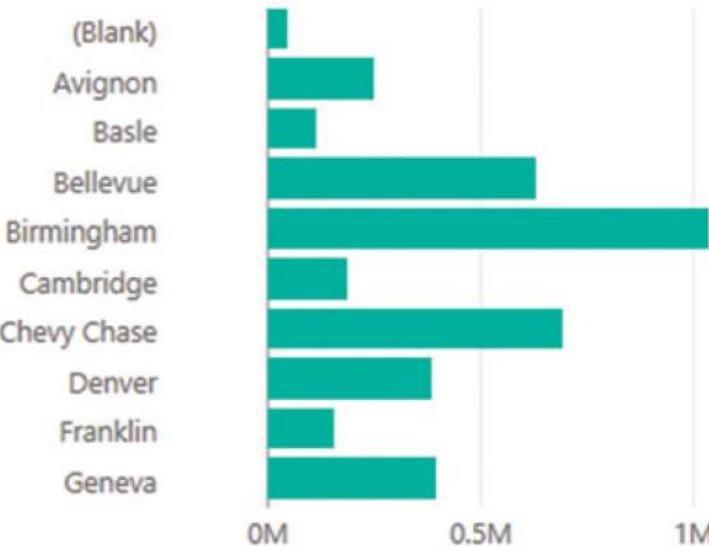
- Used for **comparing categories** or examining sales volume across time periods
 - How sales are distributed across products?
 - Which product produces the highest sales
 - How sales have changed over time?
 - Which months were good for sales
- Probably the most useful and most commonly used visualization of data
- Able to quickly **draw initial conclusions** about your data and leads to asking more questions than you can find answers for



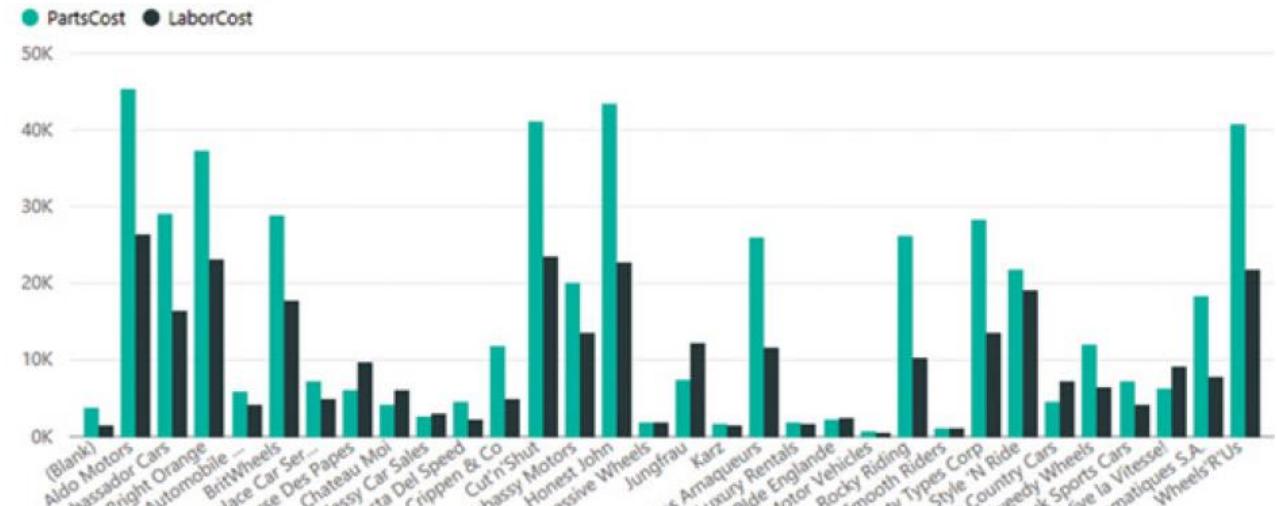
Bar Chart

Column / Clustered Column & Simple Stacked Column Chart

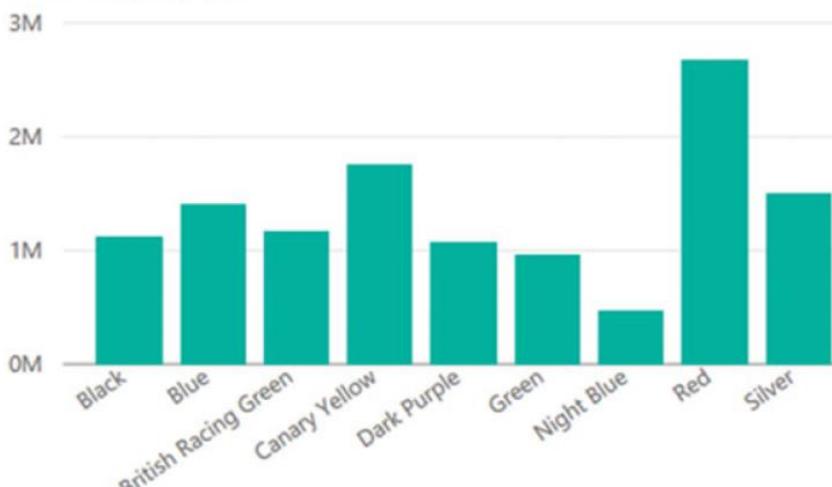
Gross Margin by Town



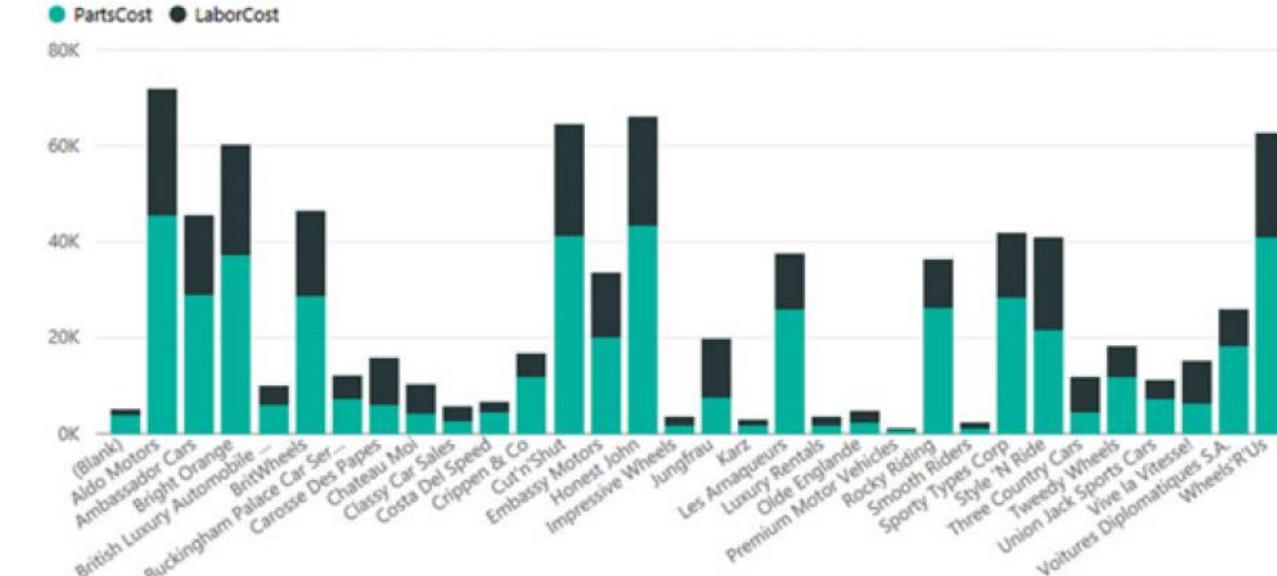
PartsCost and LaborCost by ClientName



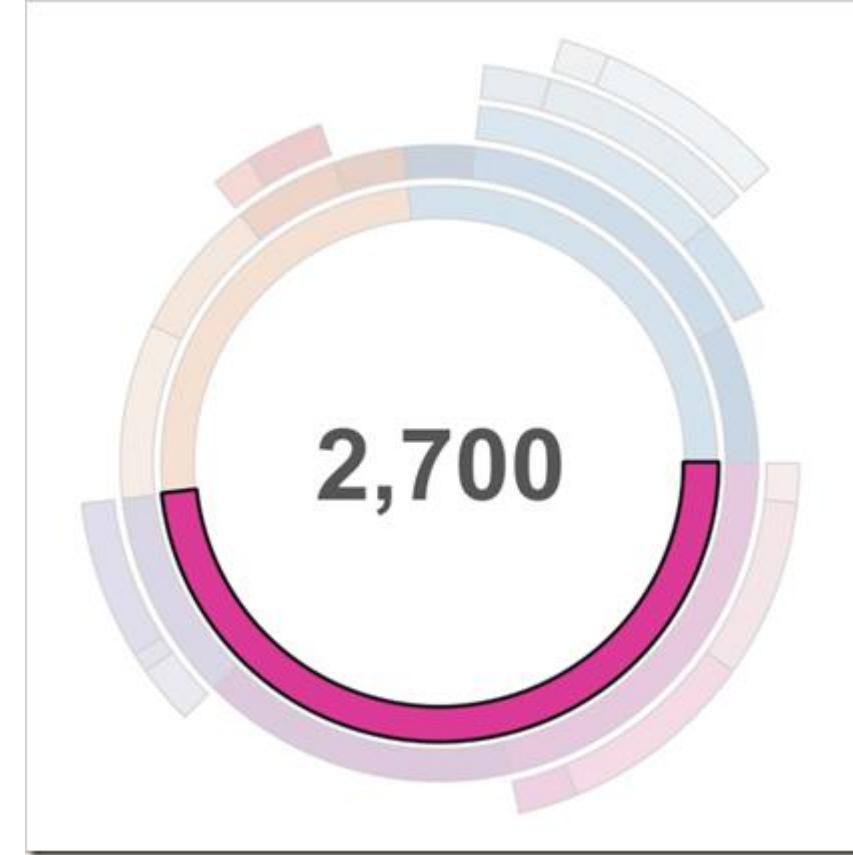
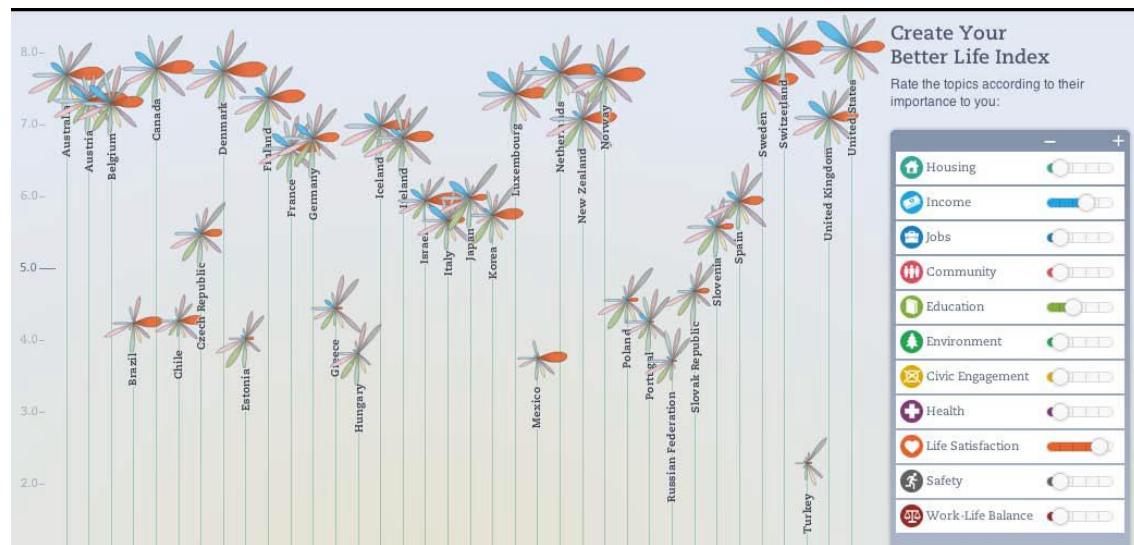
Gross Margin by Color



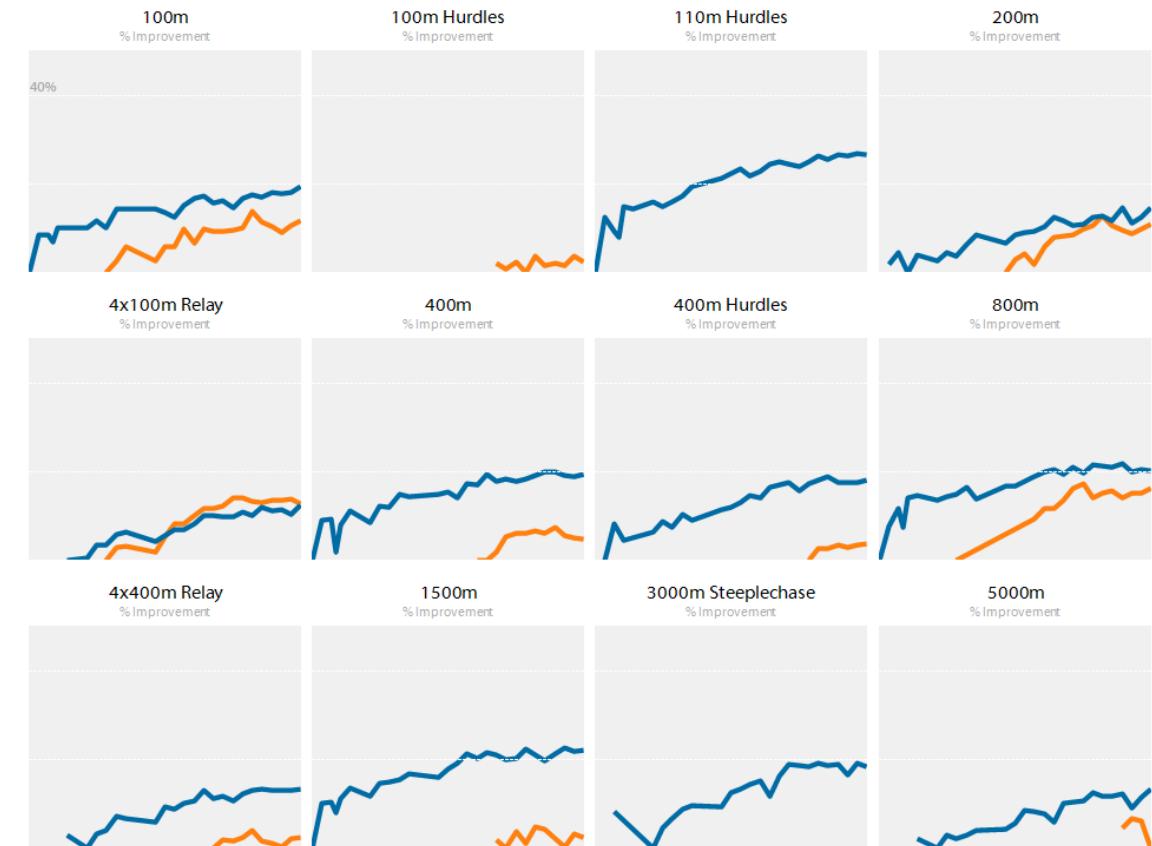
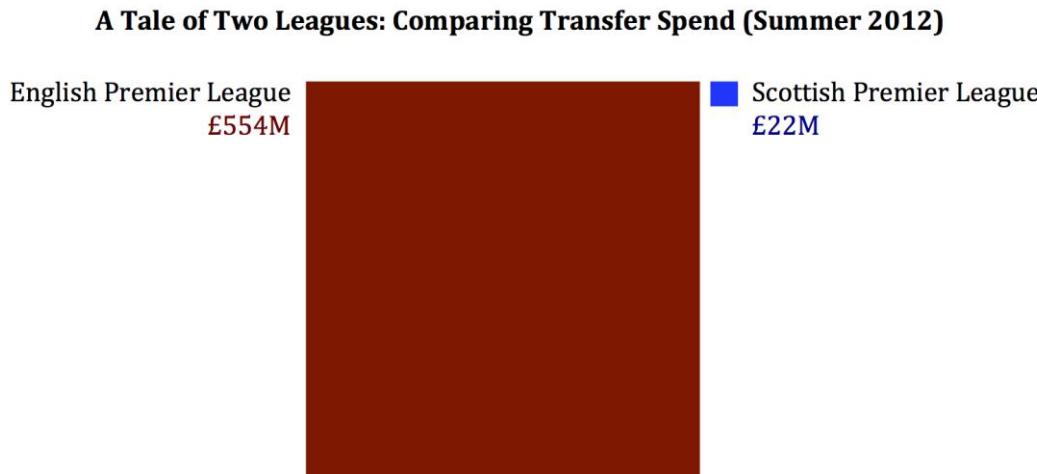
PartsCost and LaborCost by ClientName

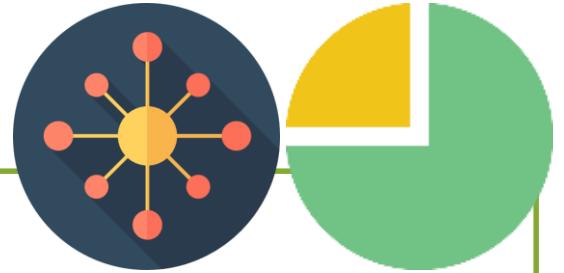


Glyph Chart vs Radial Chart



Area Size Chart vs Small Multiples / Trellis Chart





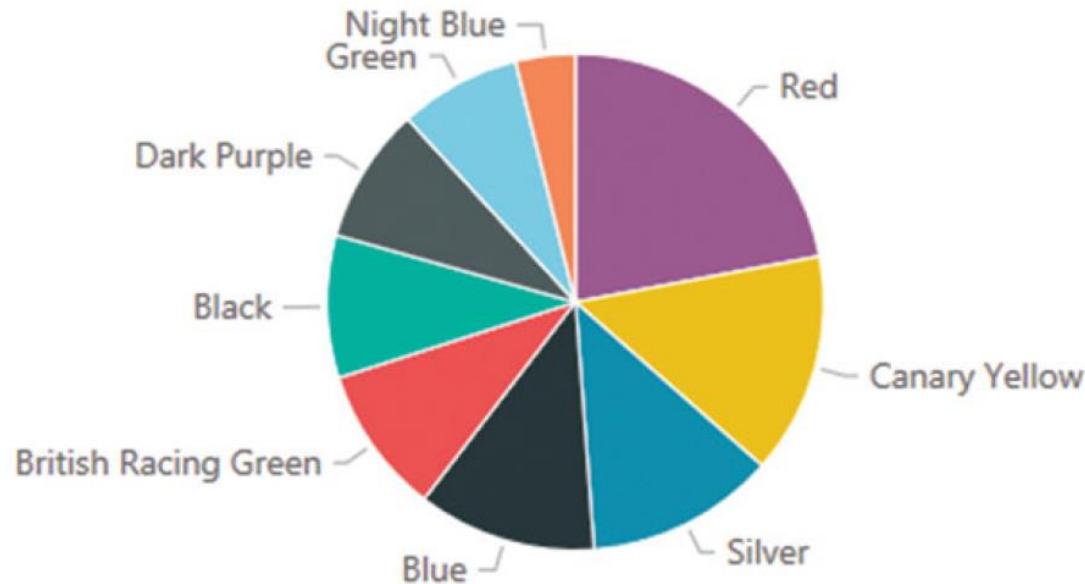
Assessing Hierarchies and Part-to-Whole Relationships

The following examples present chart types that help us to assess hierarchical and part-to-whole relationships.

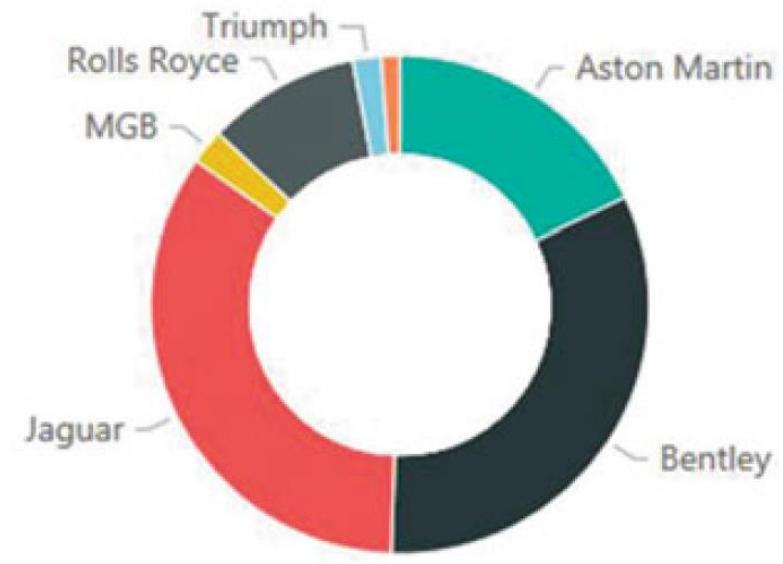
Pie Chart / Donut Chart

To display a limited set of data for a single series

Gross Margin by Color

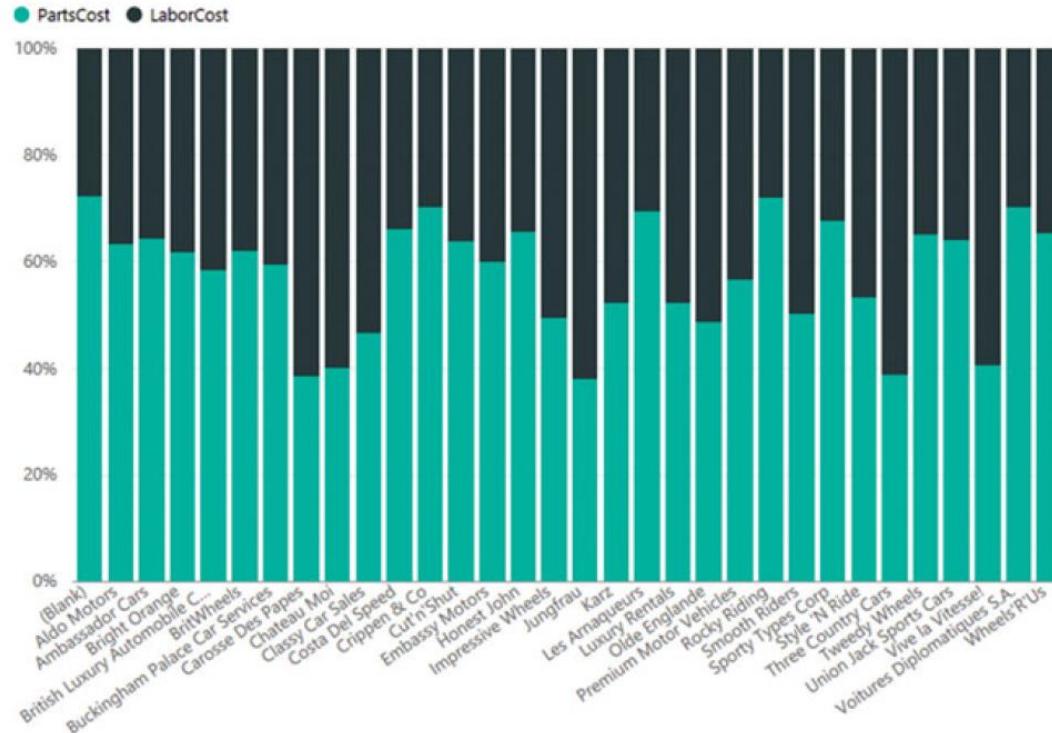


PartsCost by Make



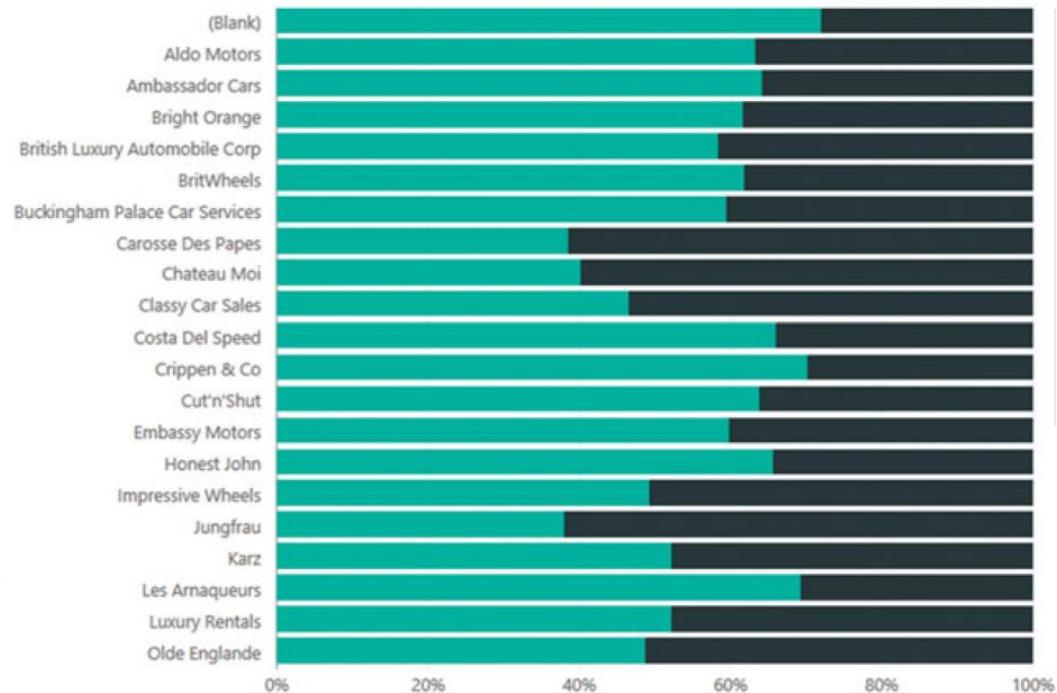
100% Stacked Column & 100% Stacked Column Chart

PartsCost and LaborCost by ClientName



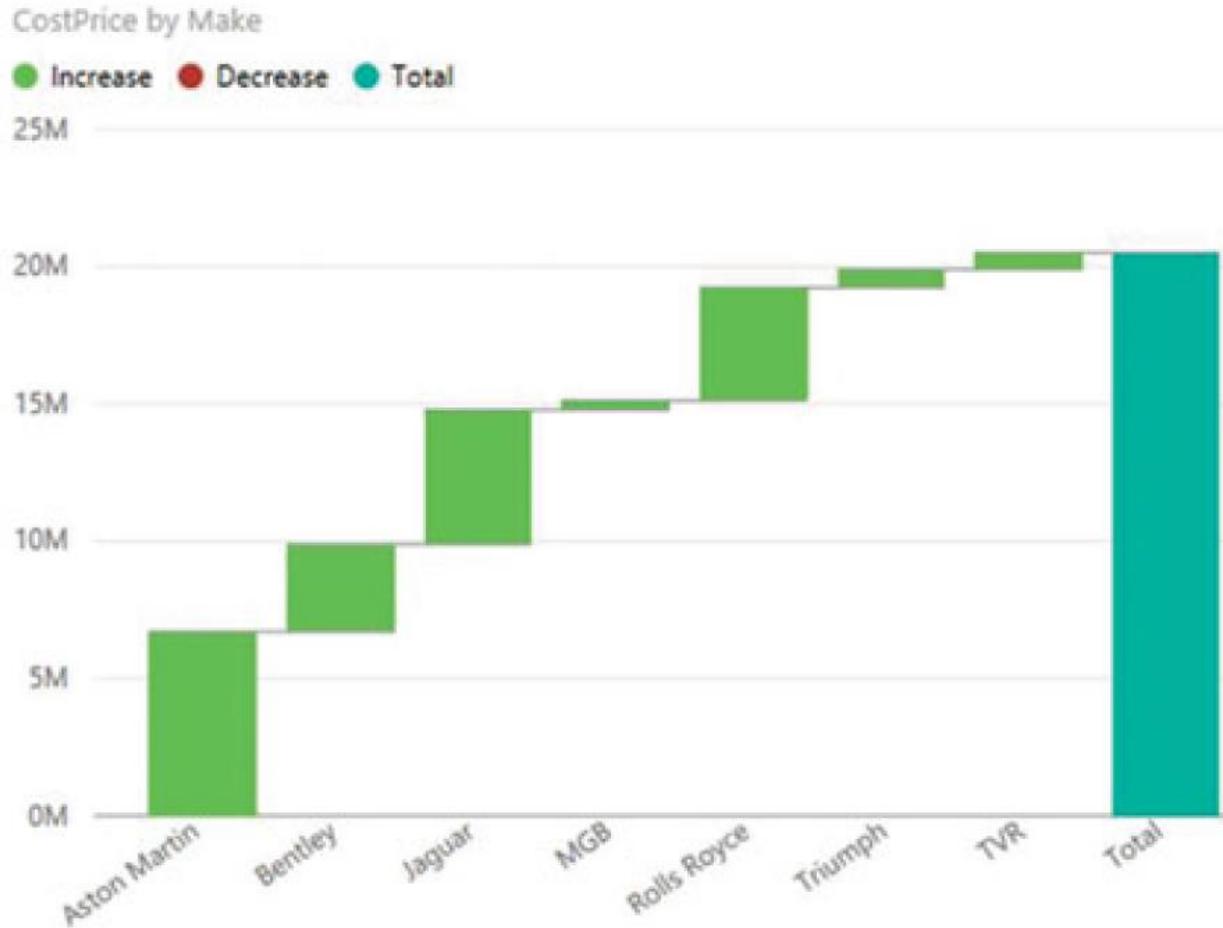
PartsCost and LaborCost by ClientName

● PartsCost ● LaborCost



To compare data from multiple datasets is to present each individual data series as a percentage of the total

Waterfall Chart

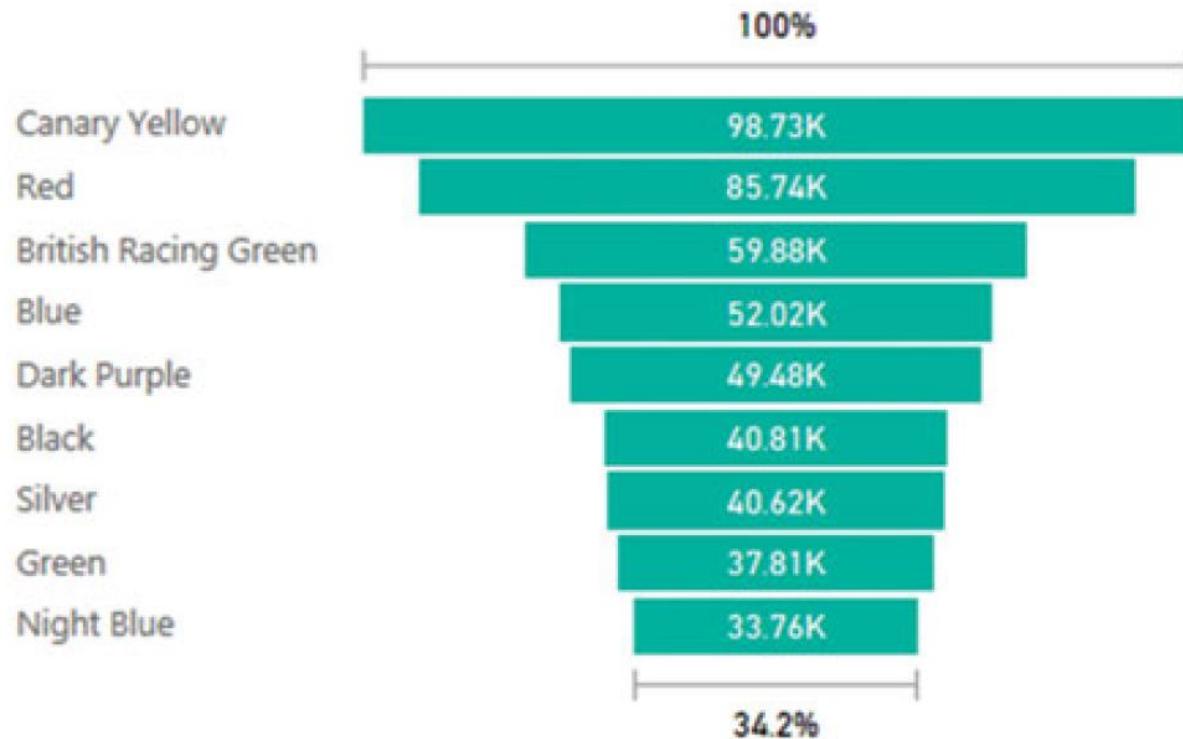


To display the component parts of a final figure

- useful when analyzing the constituent elements of a whole

Funnel Chart

PartsCost by Color



To compare the relative values of a single series of figures

Tree Map

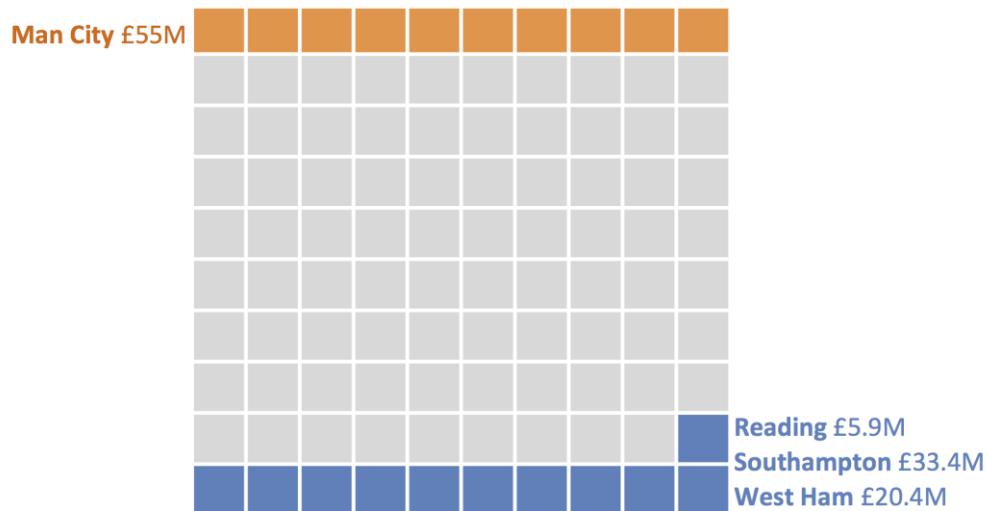


To see the way that individual values relate to a total is the tree map.

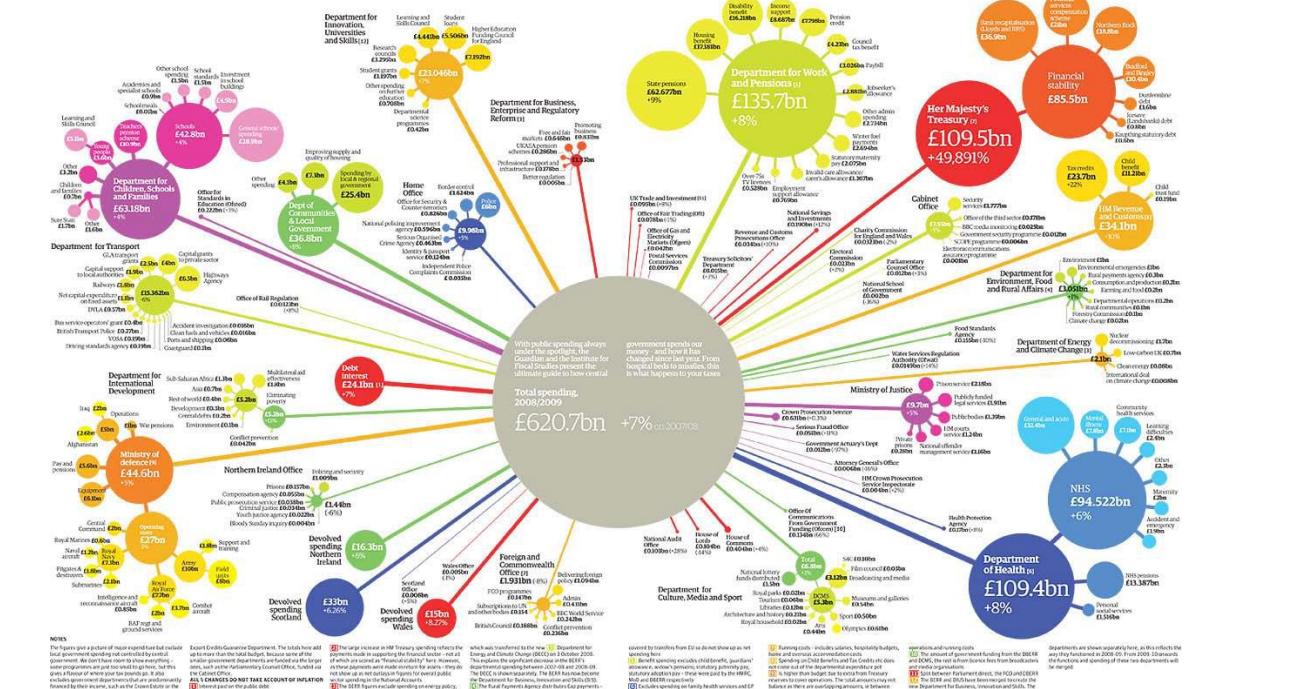
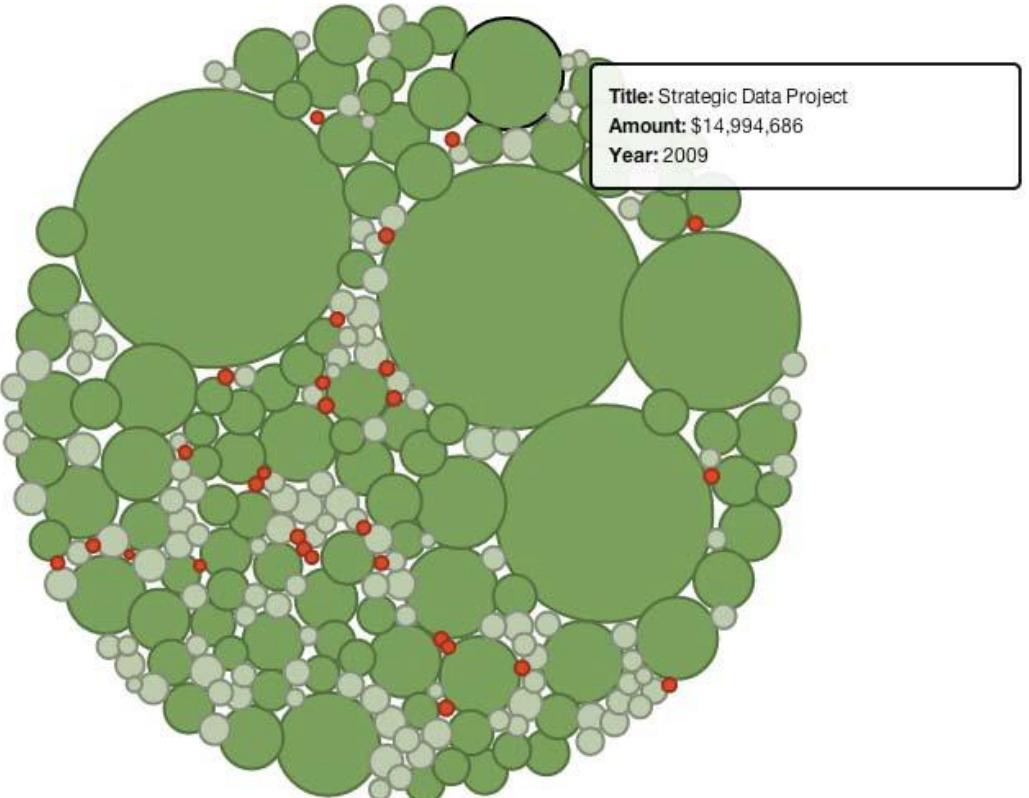
While this is not a map in any geographical sense, it can certainly assist users in finding their way around a set of figures.

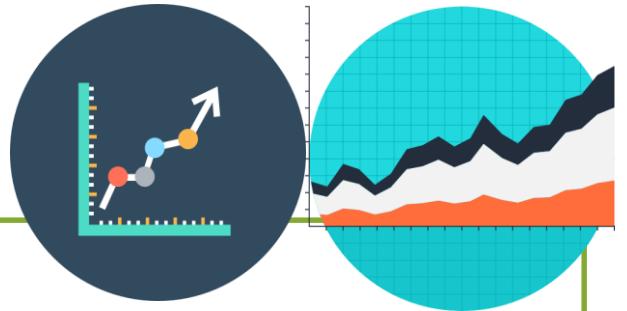
Square Pie / Unit Chart / Waffle Chart vs Tree Hierarchy

Champions vs. Promoted Teams: Total Transfer Spend, Premier League 2012



Circle Packing Diagram vs Bubble Hierarchy





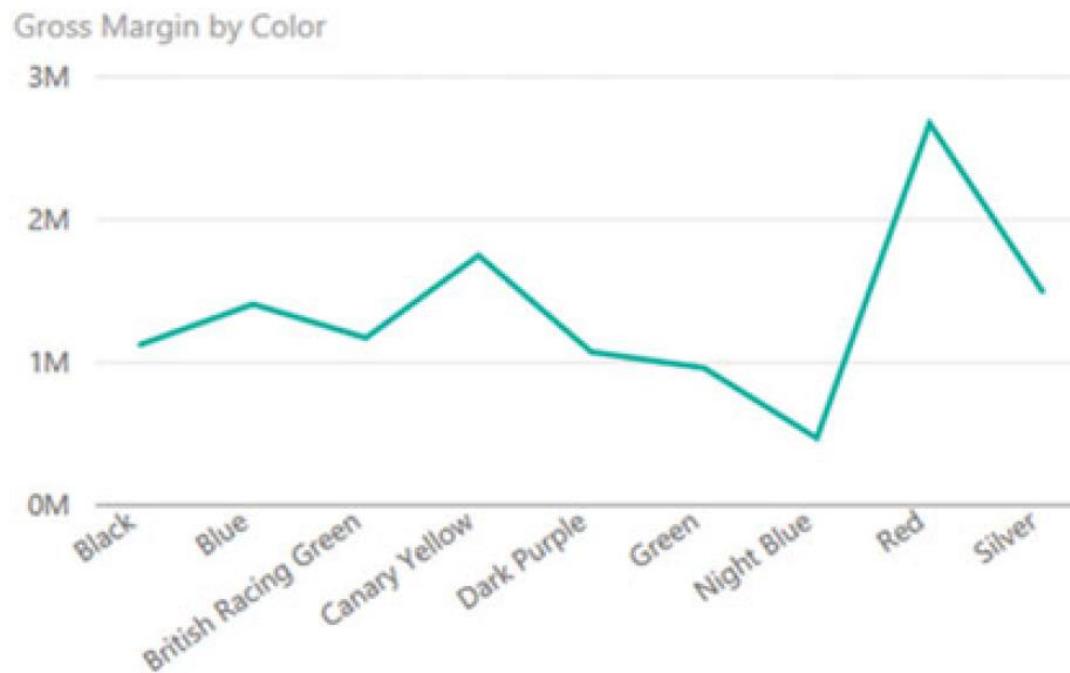
Showing Changes Over Time

The following examples show alternative ways of graphically showing changes over time

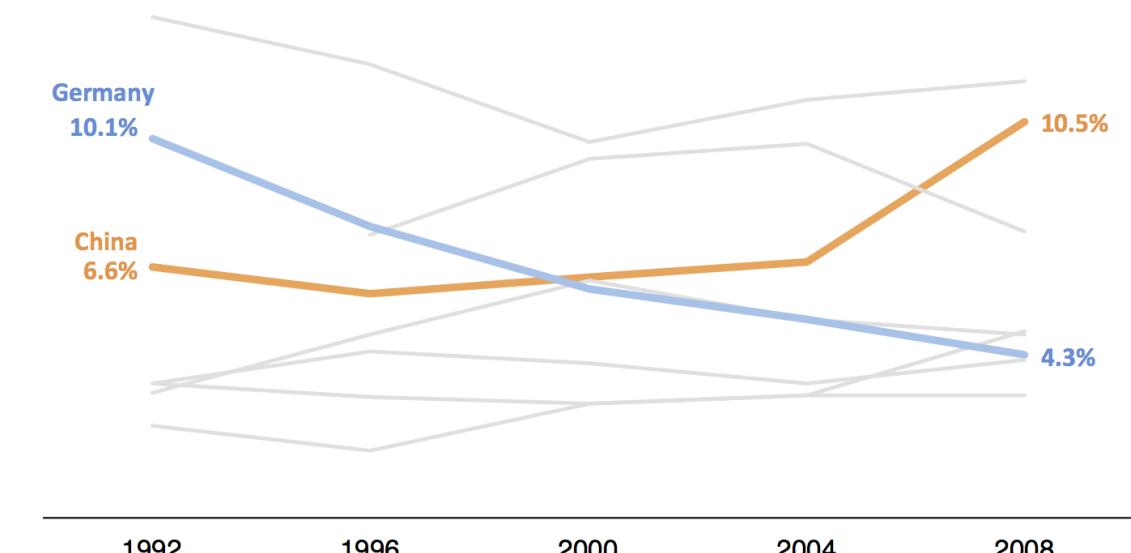
Line Chart

- displays the data as a set of points joined by a line
- Used to show the direct **relationship between 2 variables**, and how one data variable **changes over time**.
 - show data over a series of numeric values
 - Show how sales have changed over the last x years/quarters/month
- Line Plots are useful for **examining trends** over time periods. You can find peaks and dips – are sales moving up or down or stagnant; “What is happening?”; “Why is this happening?”; “Is this unusual?”; “Is this seasonal?”; etc.

Line Chart

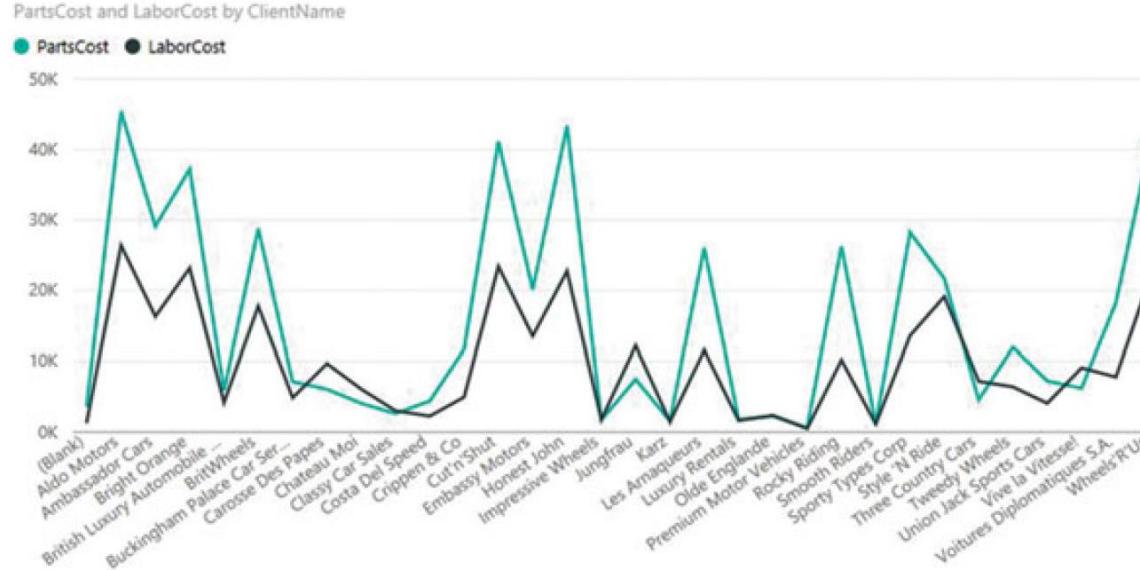


The Contrasting Fortunes of German and Chinese Olympic Success
Percentage of total medals won across past five Olympics (eight countries selected based on ranking at 2008)

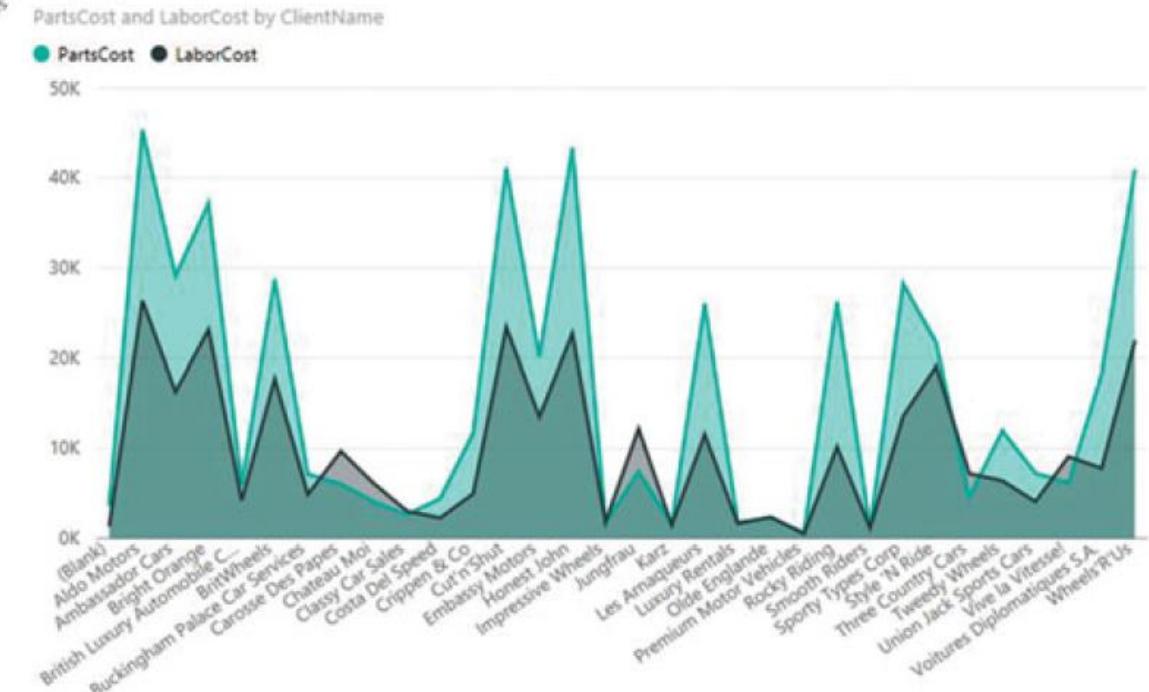


displays the data as a set of points joined by a line

Line Chart & Area Chart for Multiple Values

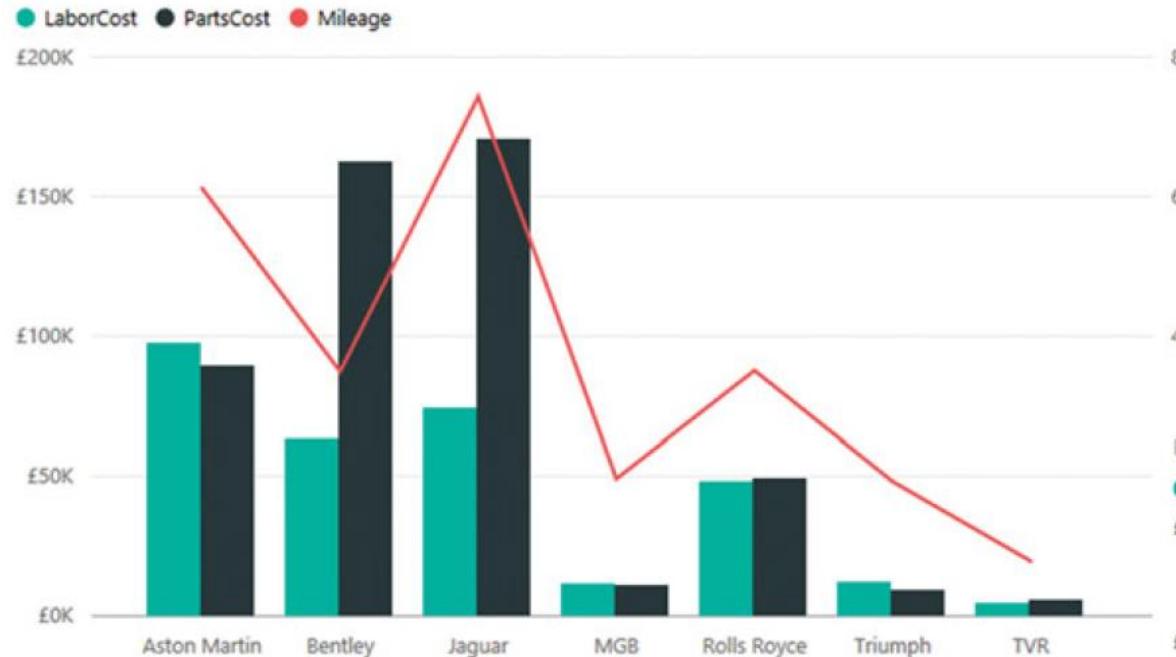


displays the data as a set
of points joined by a line



Line and Clustered Column Chart & Line and Stacked Column Chart

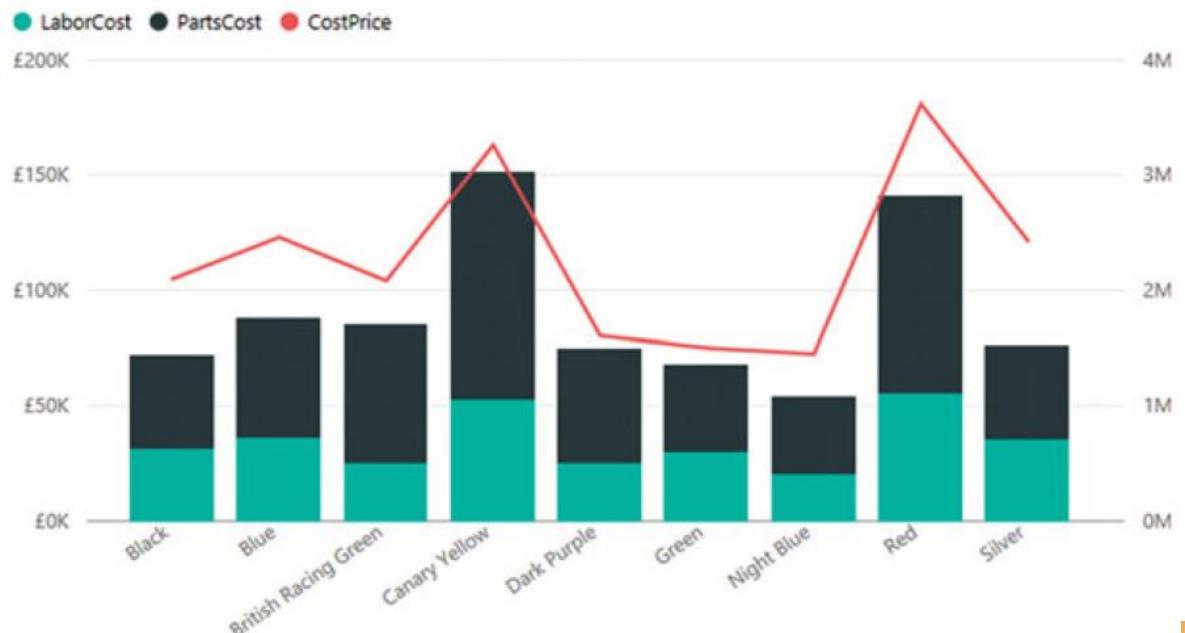
LaborCost, PartsCost and Mileage by Make



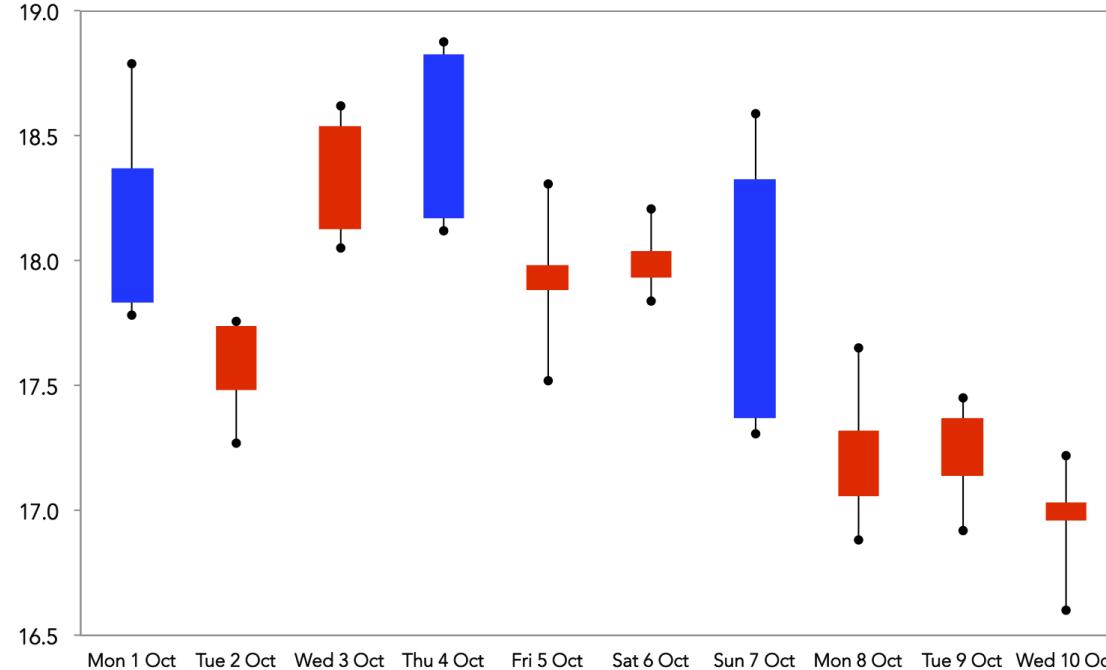
to isolate any interesting correlations

- key advantages of dual-axis charts is that the two X axes can have vastly different scales

PartsCost, LaborCost and CostPrice by Color

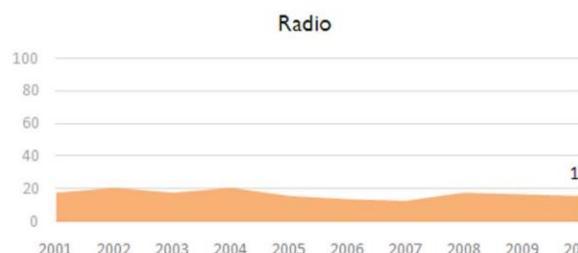
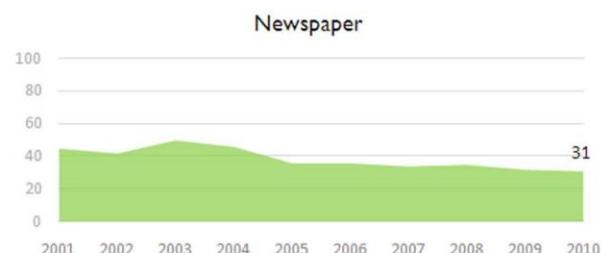
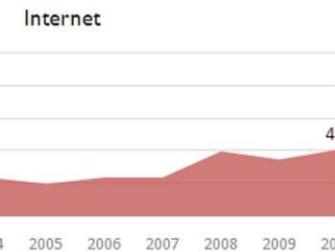
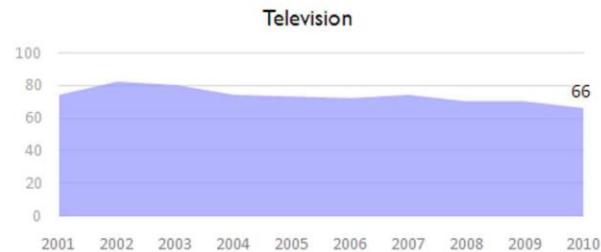


Candlestick Chart / Box and Whiskers Plot vs Sparklines

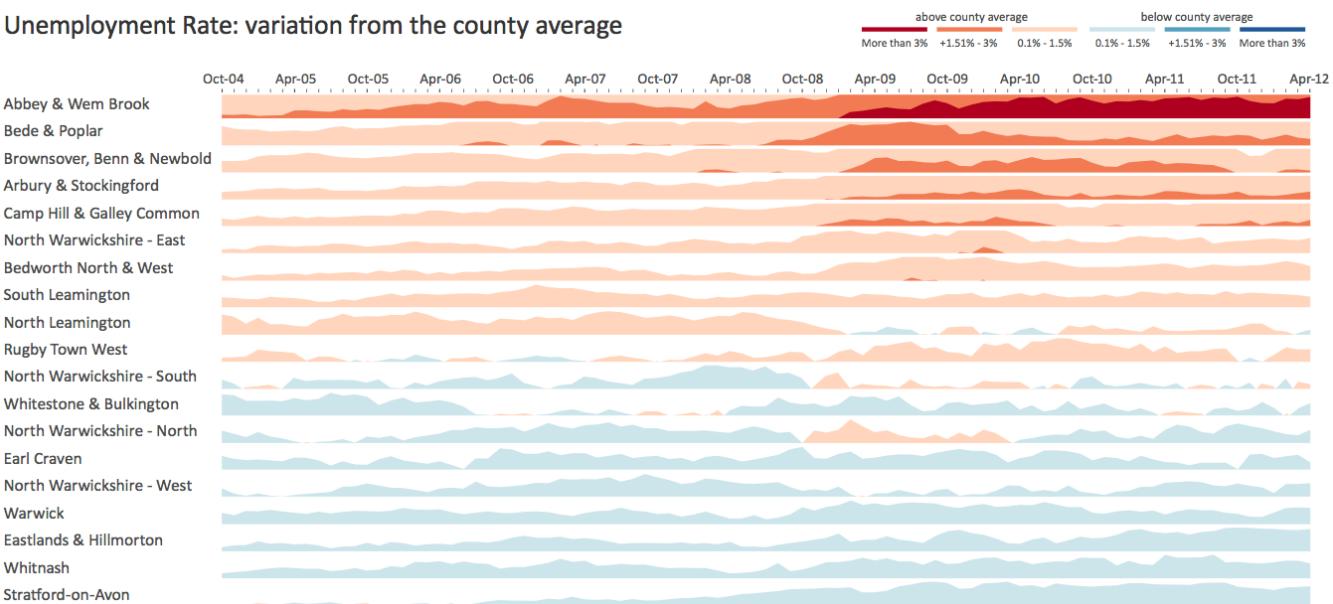


	Monthly £	Average	52 Week Sales	Best
Team A	■■■■■	£55.5	Sparkline	73
Team B	■■■■■	£50.0	Sparkline	70
Team C	■■■■■	£51.3	Sparkline	66
Team D	■■■■■	£53.2	Sparkline	73
Team E	■■■■■	£38.8	Sparkline	75
Team F	■■■■■	£56.0	Sparkline	72
Team G	■■■■■	£48.0	Sparkline	70

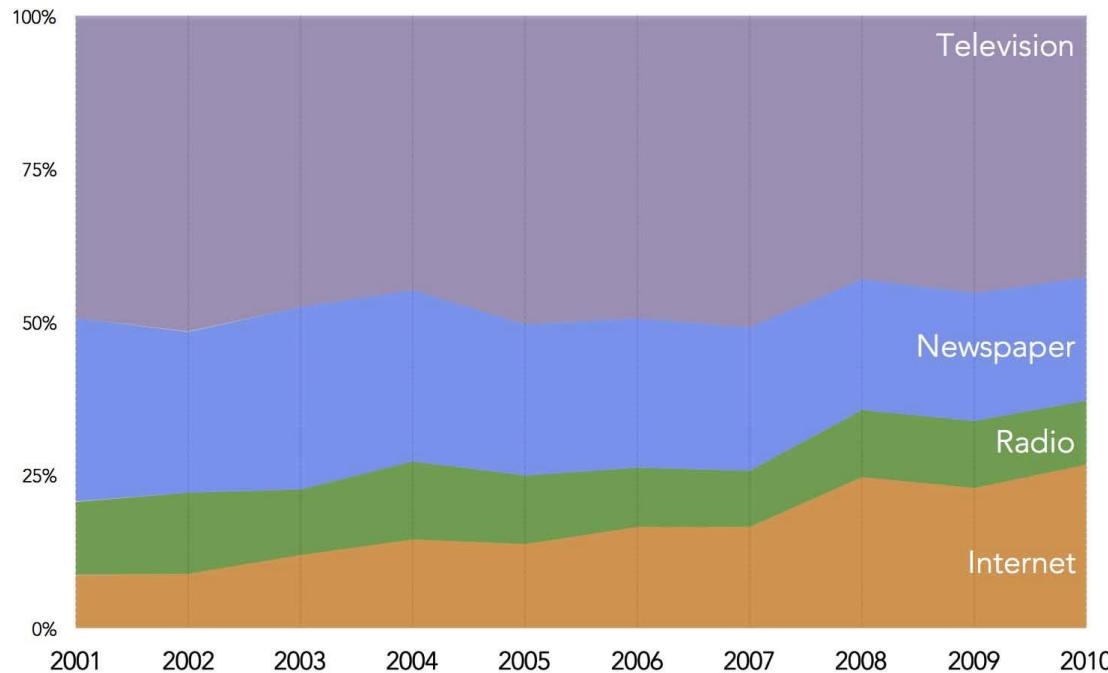
Area Chart vs Horizon Chart



Unemployment Rate: variation from the county average



Stacked Area Chart vs Stream Graph

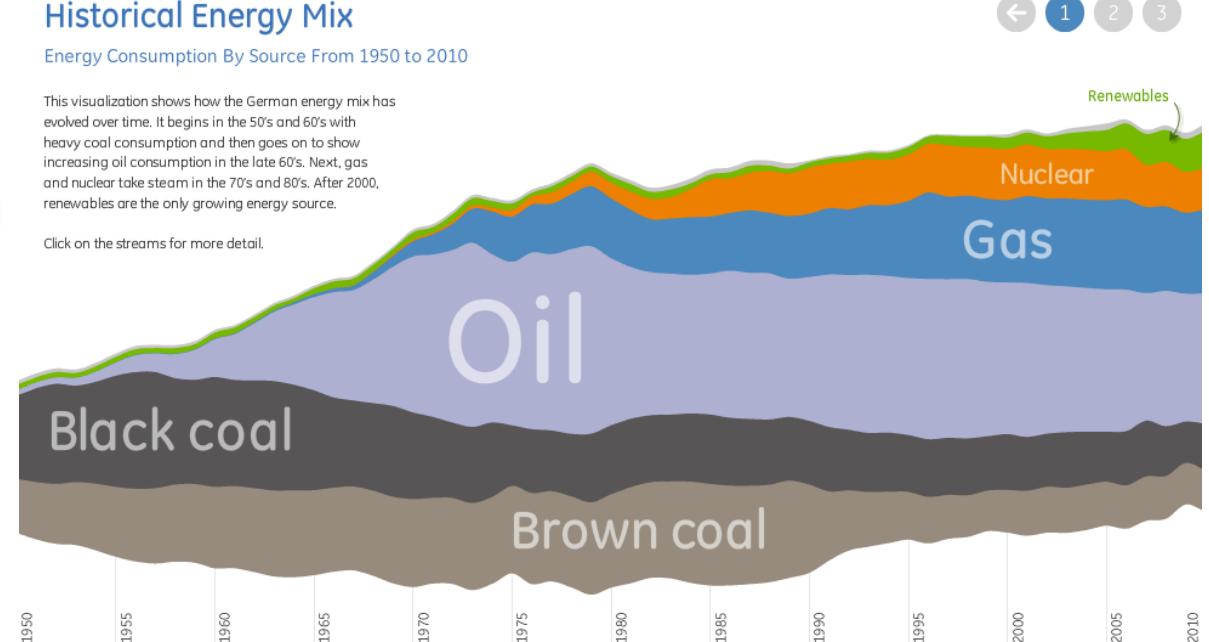


Historical Energy Mix

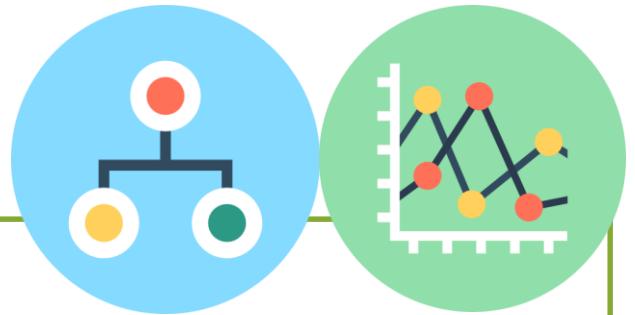
Energy Consumption By Source From 1950 to 2010

This visualization shows how the German energy mix has evolved over time. It begins in the 50's and 60's with heavy coal consumption and then goes on to show increasing oil consumption in the late 60's. Next, gas and nuclear take steam in the 70's and 80's. After 2000, renewables are the only growing energy source.

Click on the streams for more detail.



Sources: AG Energiebilanzen, Kalert, BPB

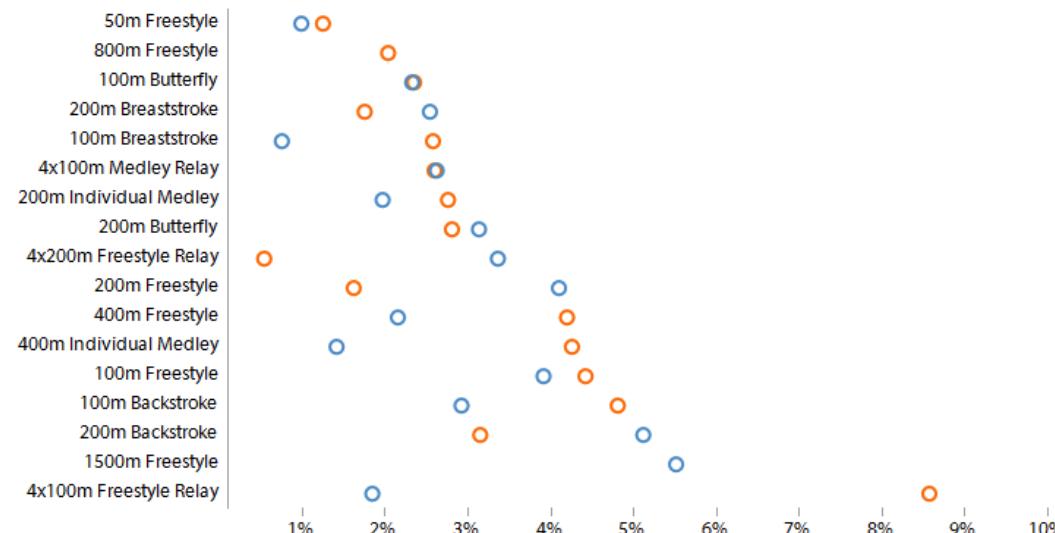


Plotting Connections and Relationships

We now look at the different visualization techniques used to plot connections and relationships

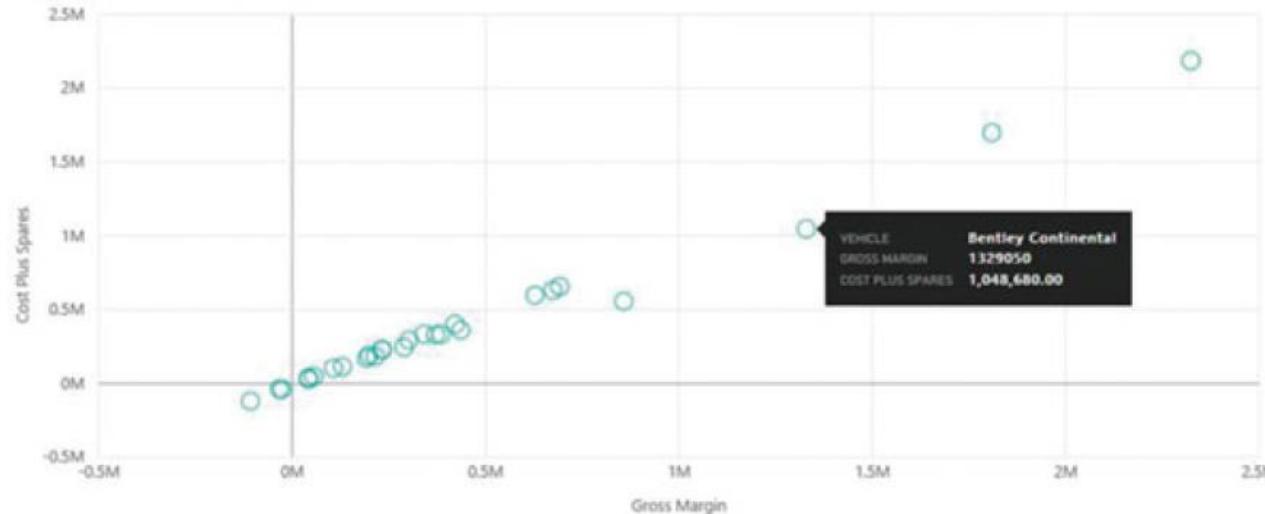
Scatter Plot / Bubble Plot Charts

- Used for **analyzing relationships** between two numeric data variables.
 - It is often used to prove or disprove a relationship between the two variables.
 - Example: see how sales and profit relate.
 - Allows a view of skewness of the data distribution.
 - Example – can easily **detect Outliers**.
- Scatter plots can be more useful if you draw **trend lines**.
- Bubble Plot - is a scatter chart with a third piece of data included

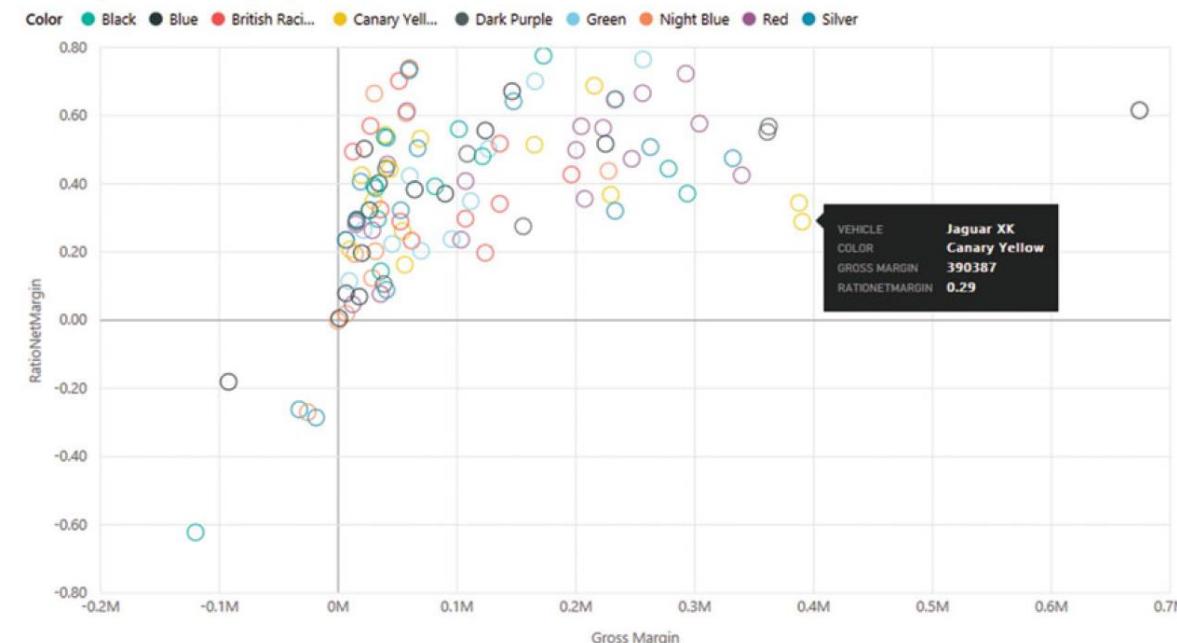


Scatter Plot

Gross Margin and Cost Plus Spares by Vehicle

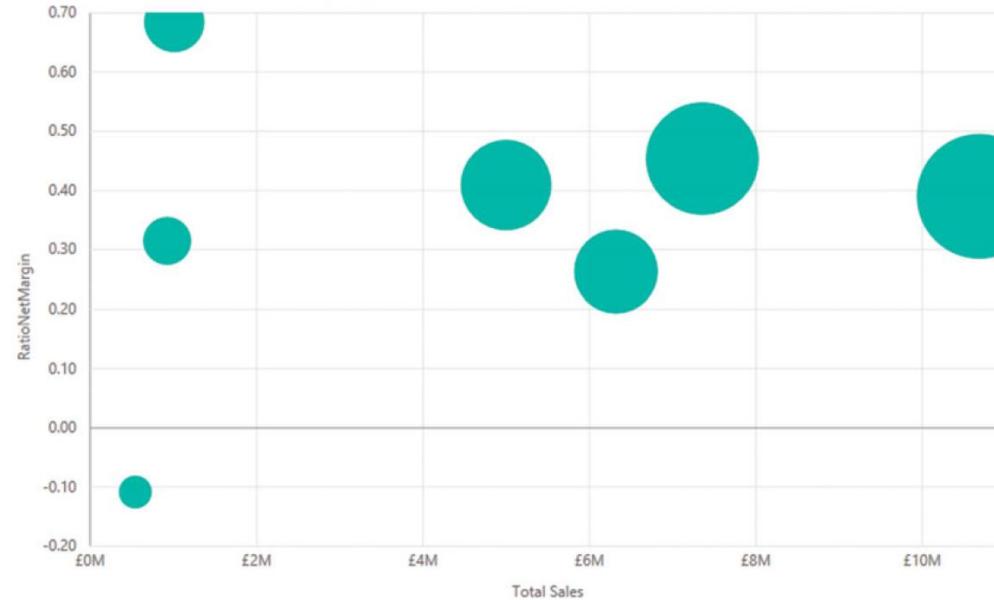


Gross Margin and RatioNetMargin by Vehicle and Color



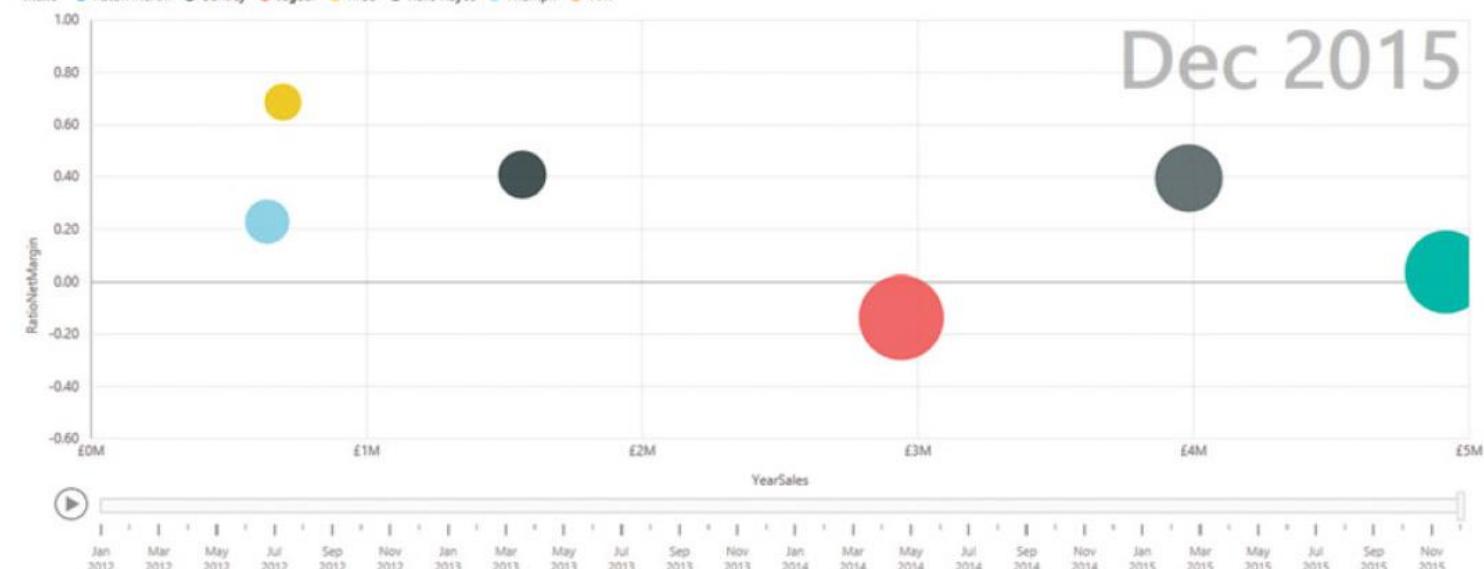
Bubble Plot Chart

Total Sales, RatioNetMargin and Gross Margin by Make



YearSales, RatioNetMargin and Direct Costs by Make and MonthAndYearAbbr

Make: Aston Martin, Bentley, Jaguar, MGB, Rolls Royce, Triumph, TVR



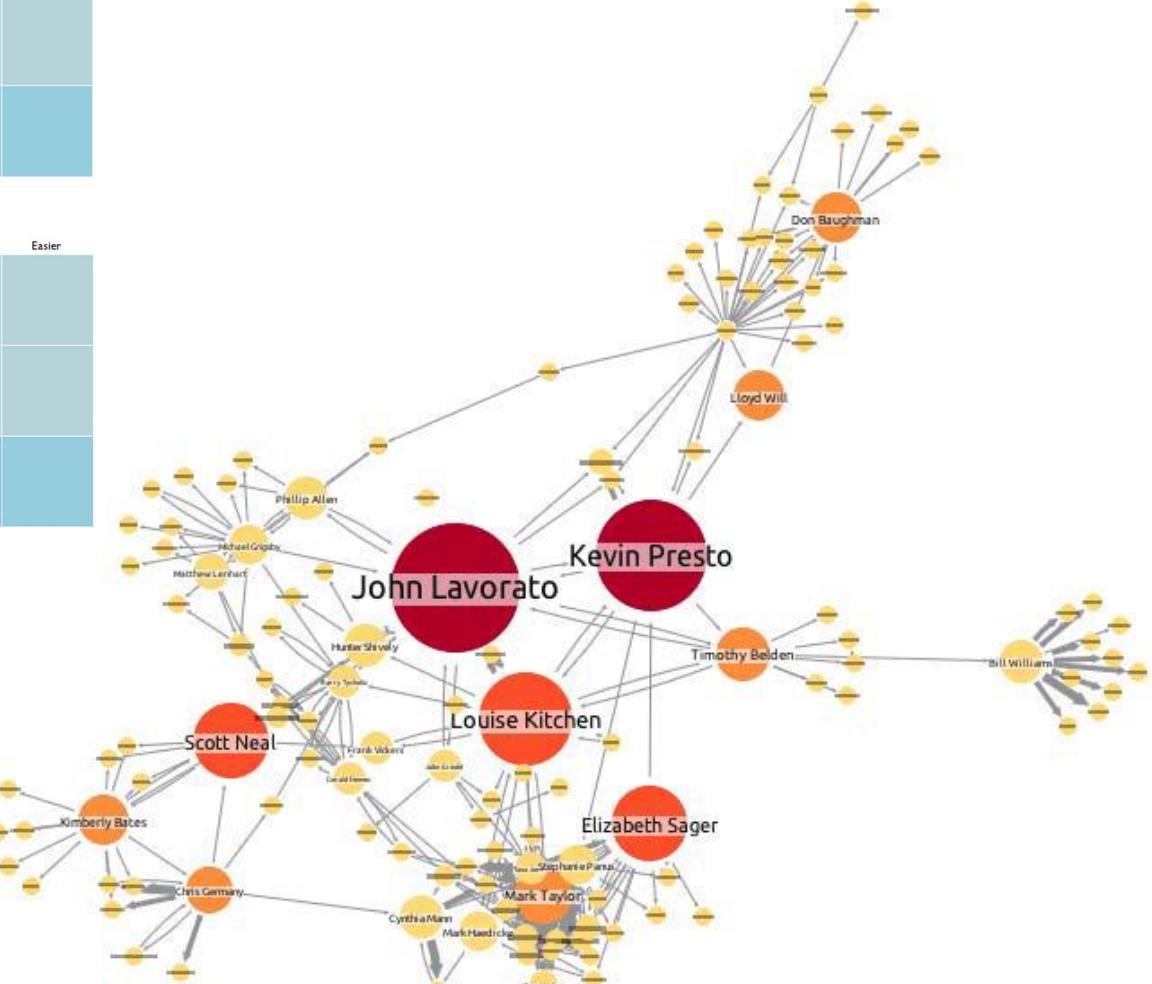
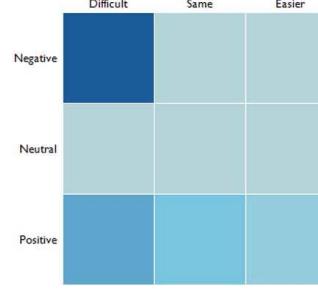
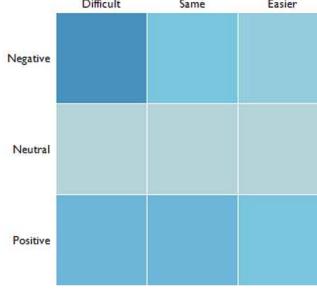
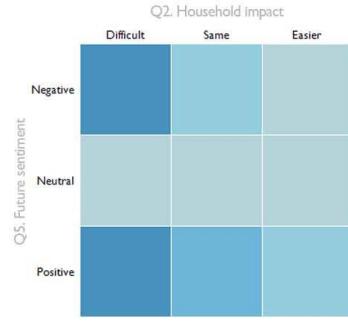
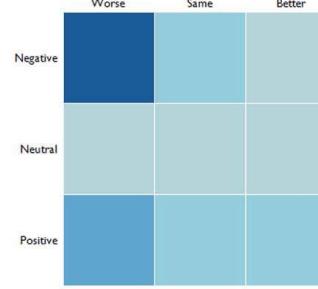
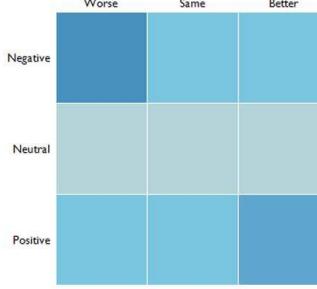
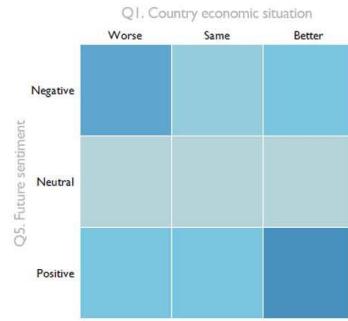
Gauges

Cost Plus Spares, SalePrice and TotalCostPlusSpares



to compare actuals to targets or see how a metric compares to a key performance indicator (KPI)

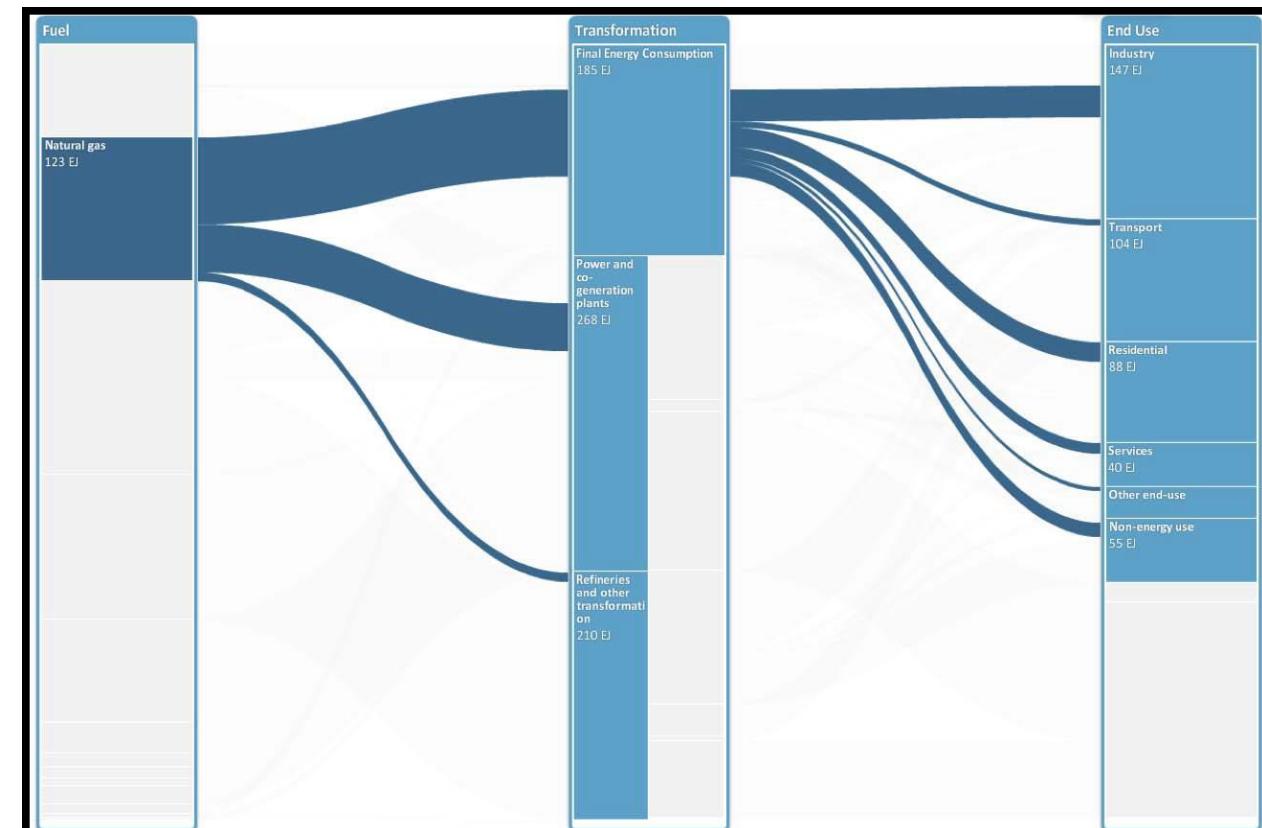
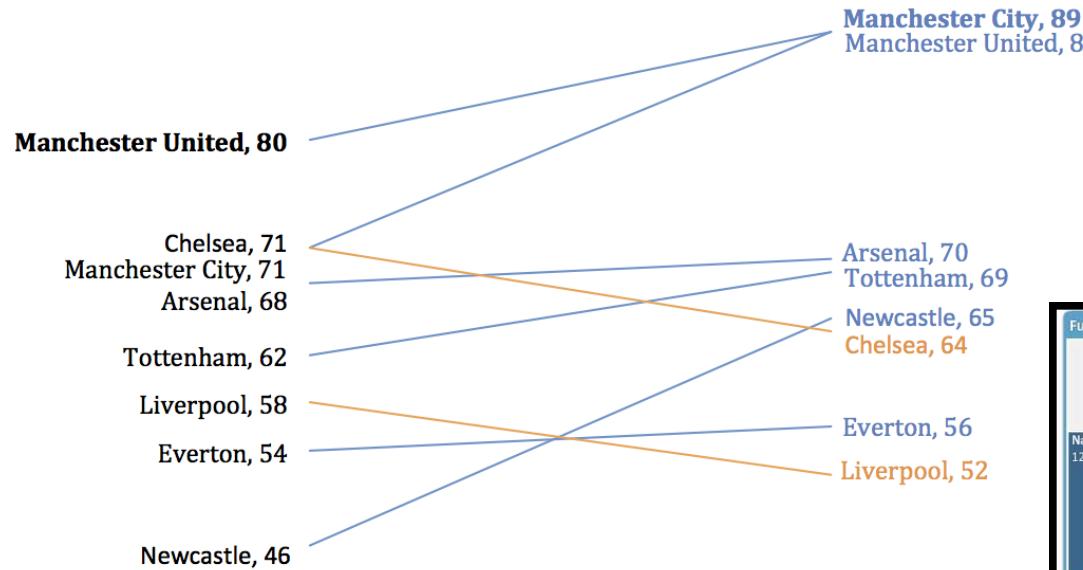
Heatmap / Matrix Chart vs Network Diagram / Force-Directed / Node-Link Network



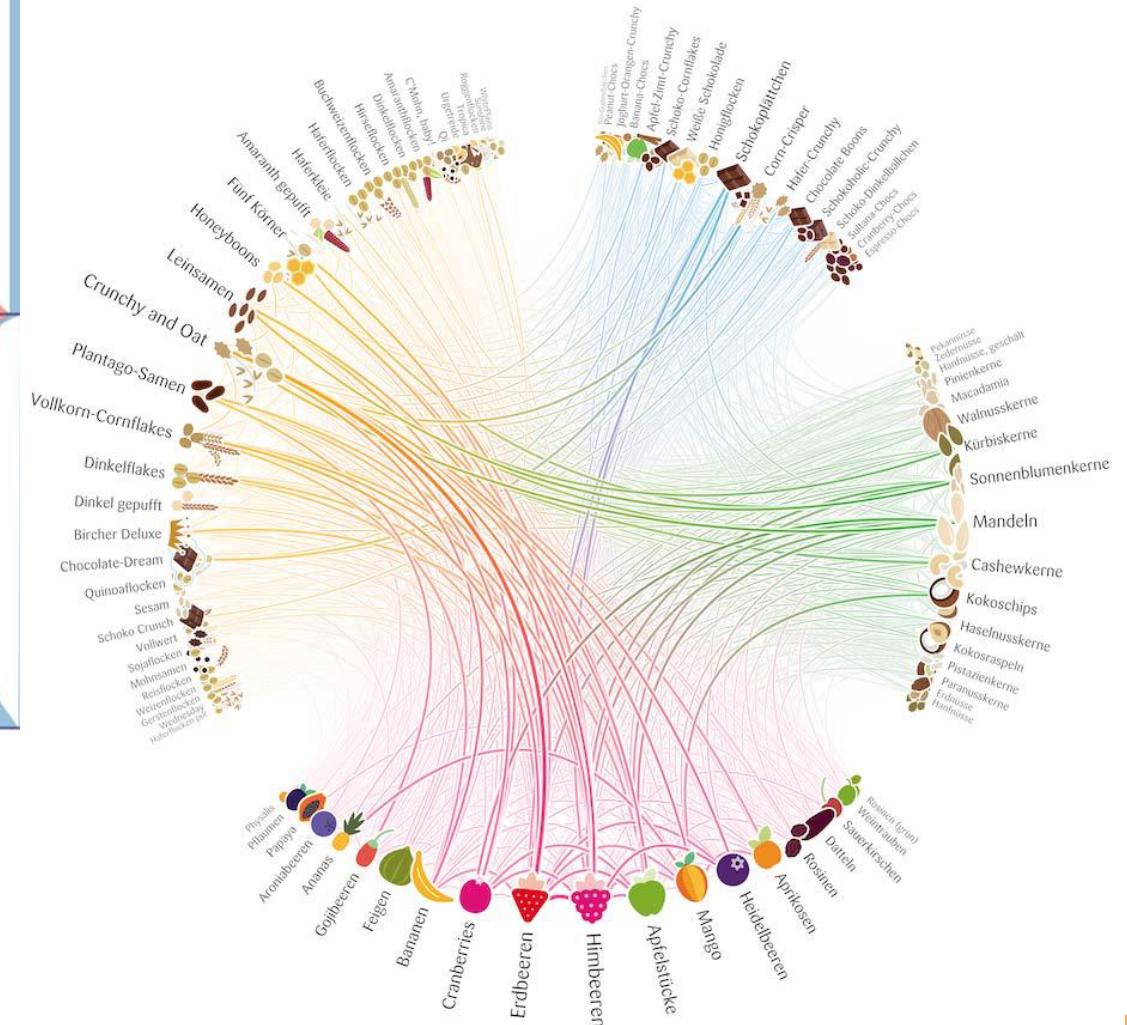
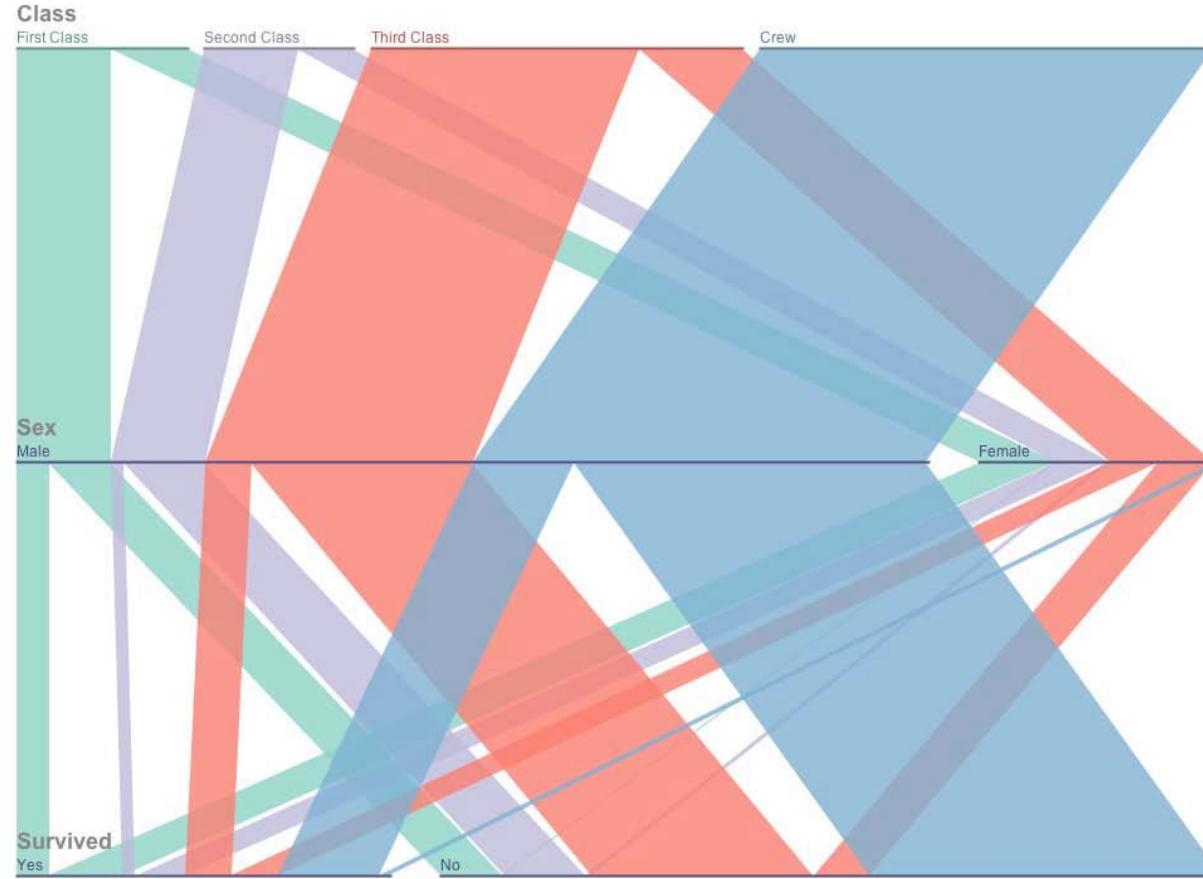
Slopegraph vs Sankey Diagram

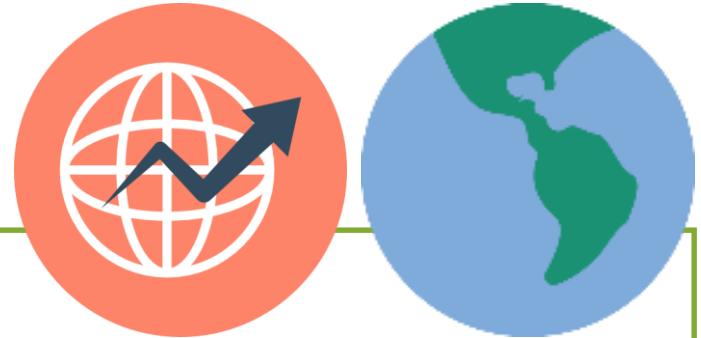
2010/2011

2011/2012



Parallel Sets / Parallel Coordinates vs Radial Network / Chord Diagram

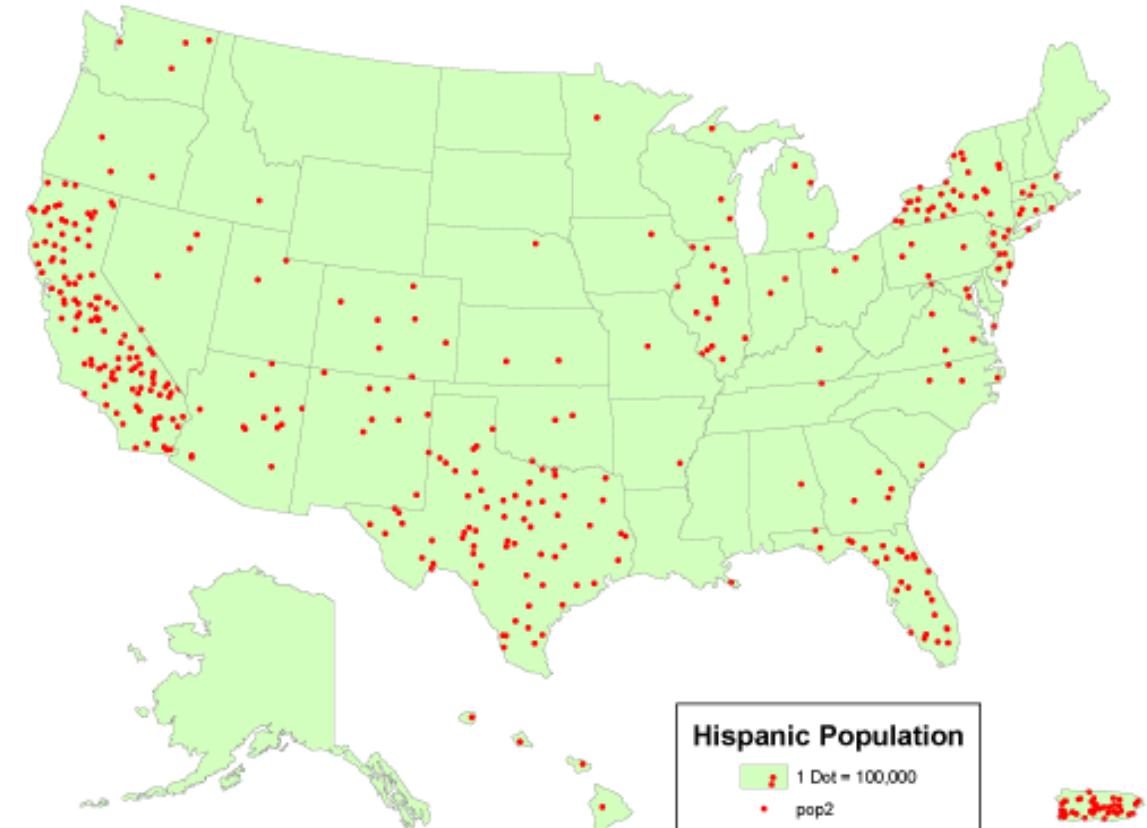
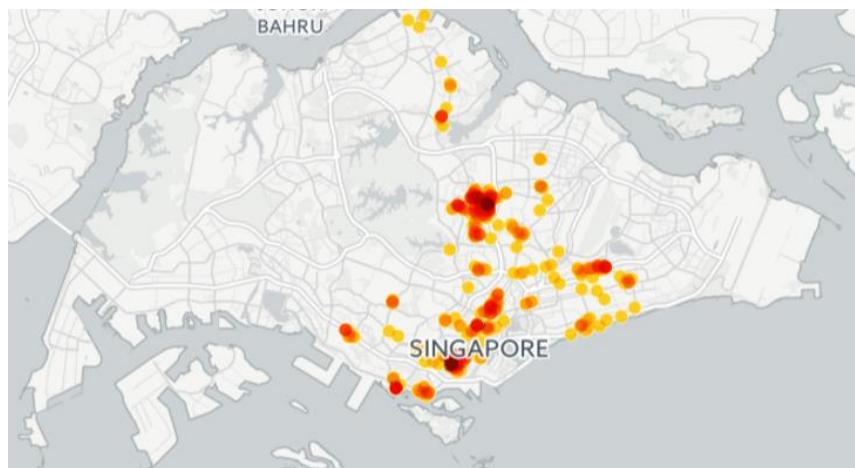
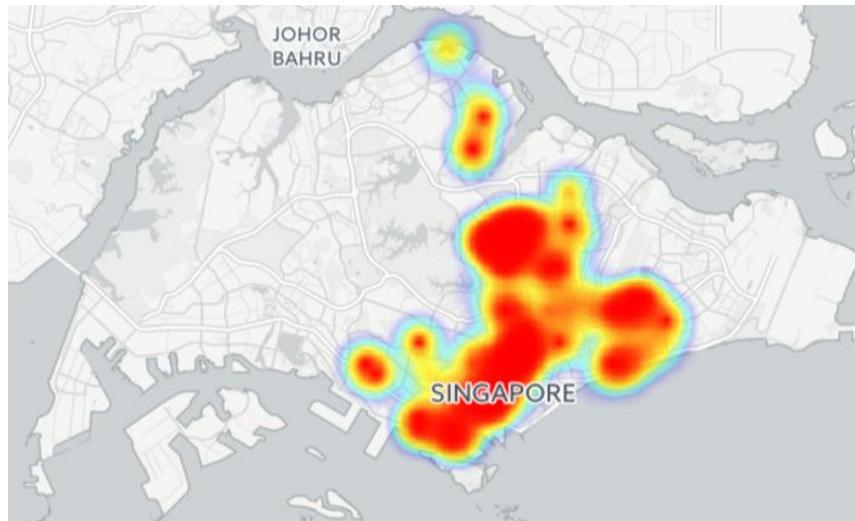




Mapping Geospatial Data

This final set of examples explores the different ways of mapping geo-spatial data

Choropleth Map (Heatmap) vs Dot Plot Map



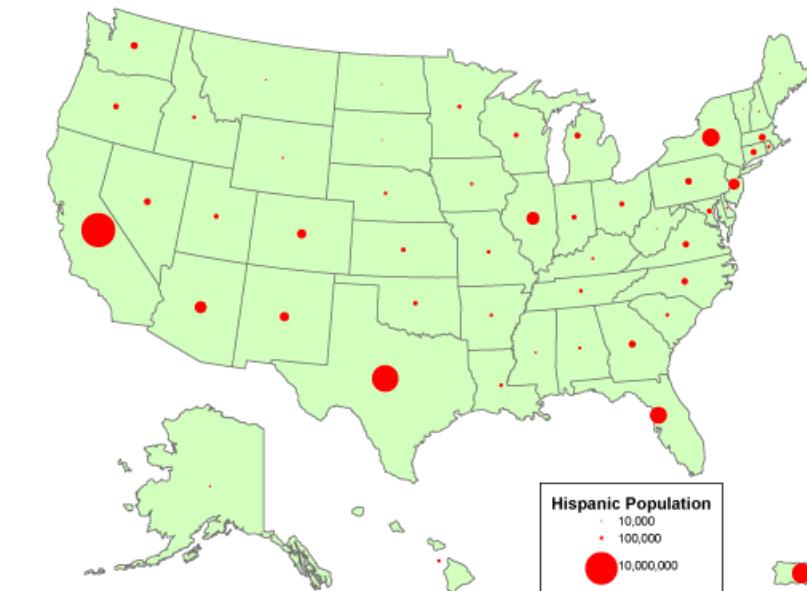


SalePrice by CountryName and Color

● Black ● Blue ● British Rac... ● Canary Yel... ● Dark Purple ● Green ● Night Blue ● Red ● Silver

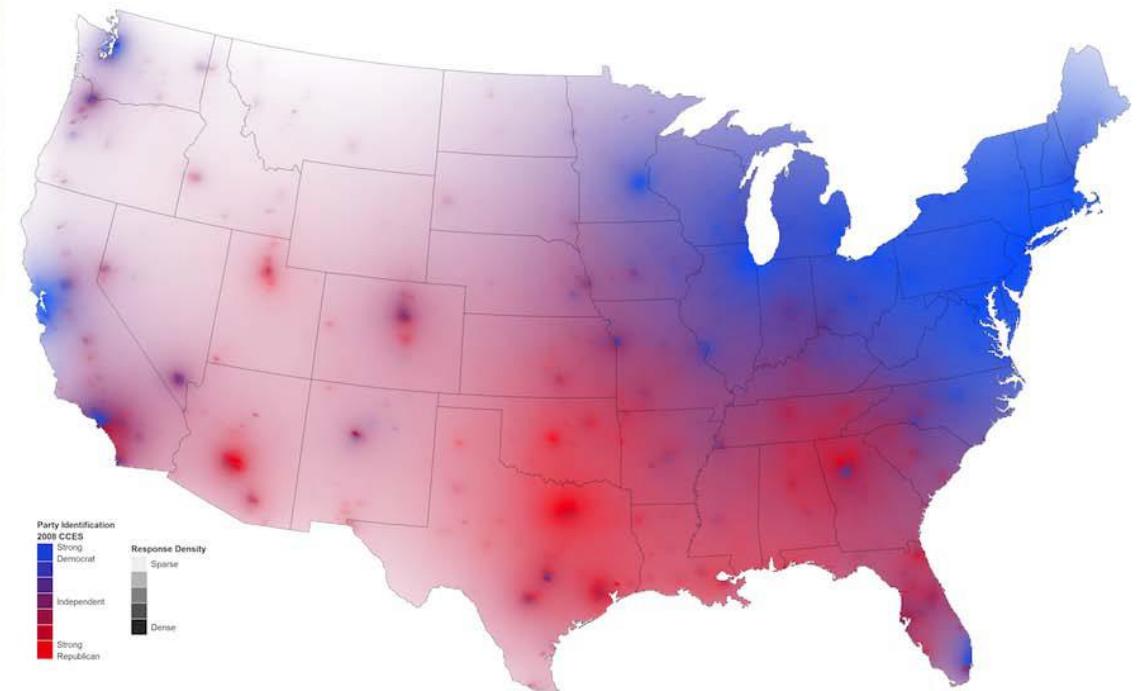


Bubble Plot Map



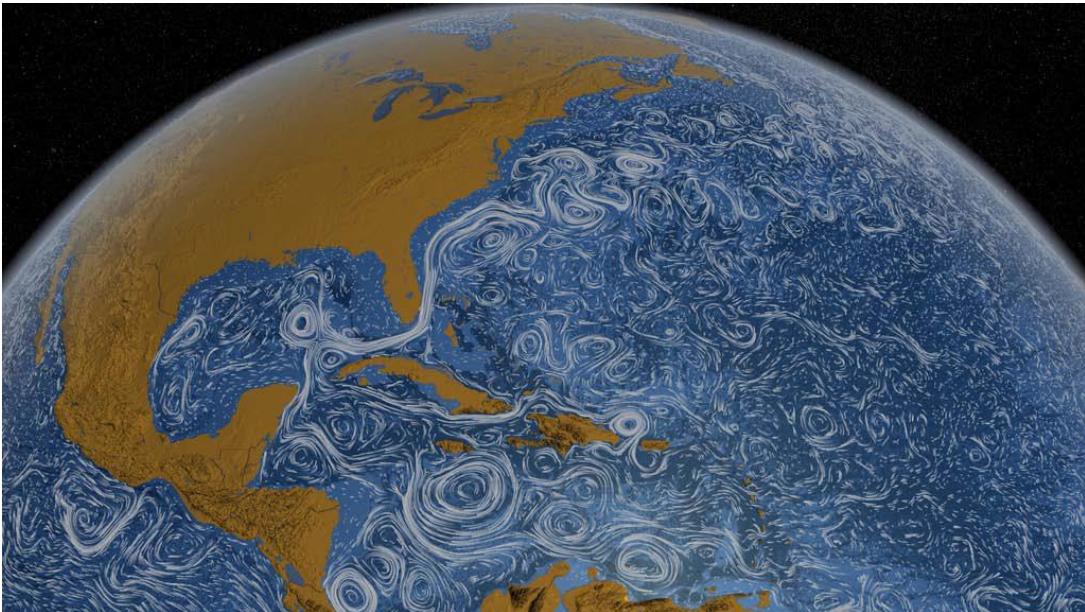
Filled Map vs Isarithmic Map / Contour Map / Topological Map

CostPrice by CountryName

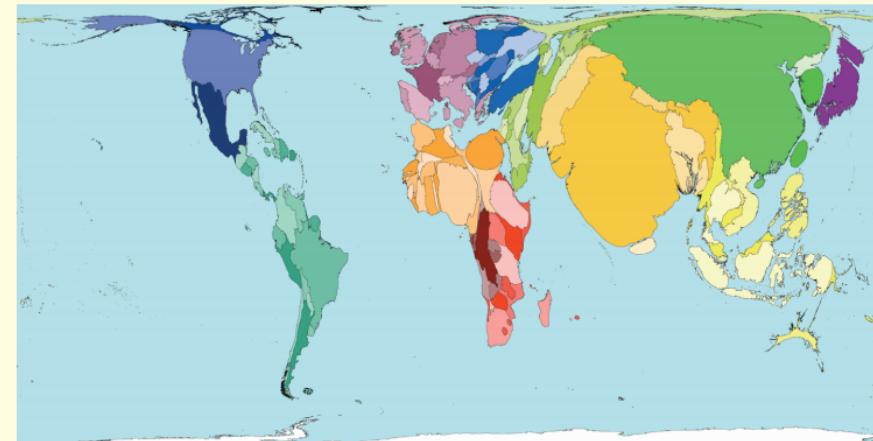


<https://dsparks.wordpress.com/2011/10/24/isarithmic-maps-of-public-opinion-data/>

Particle Flow Map vs Cartogram



Total Population



Technical notes
• Data source: United Nations Development Programme, 2004, Human Development Report.
• Population data is from 2004.
• The figure for Iraq's population is estimated as 2 to 3 million (see Appendix map 2).
• See website for further information.

MOST AND FEWEST PEOPLE	
Rank	Territory
1	China
2	India
3	United States
4	Indonesia
5	Brazil
6	Pakistan
7	Russian Federation
8	Bangladesh
9	Japan
10	Nigeria
Value	
1295	
1050	
291	
217	
176	
150	
144	
144	
128	
121	
millions	
42	
34	
33	
27	
20	
18	
13	
10	
2	
1	
thousands	

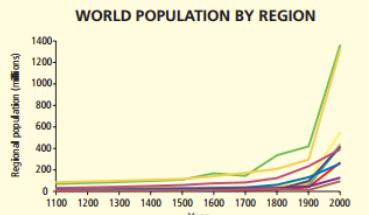
The University of Sheffield M The Leverhulme Trust G Geography Association

In Spring 2000 world population estimates reached 6 billion; that is 6 thousand million. The distribution of the earth's population is shown in this map.

India, China and Japan appear large on the map because they have large populations. Panama, Namibia and Guinea-Bissau have small populations so are barely visible on the map.

Population is very weakly related to land area. However, Sudan, which is geographically the largest country in Africa, has a smaller population than Nigeria, Egypt, Ethiopia, Democratic Republic of Congo, South Africa or Tanzania.

The size of each territory shows the relative proportion of the world's population living there.



"Out of every 100 persons added to the population in the coming decade, 97 will live in developing countries."

Hania Zlotnik, 2005

Map 002

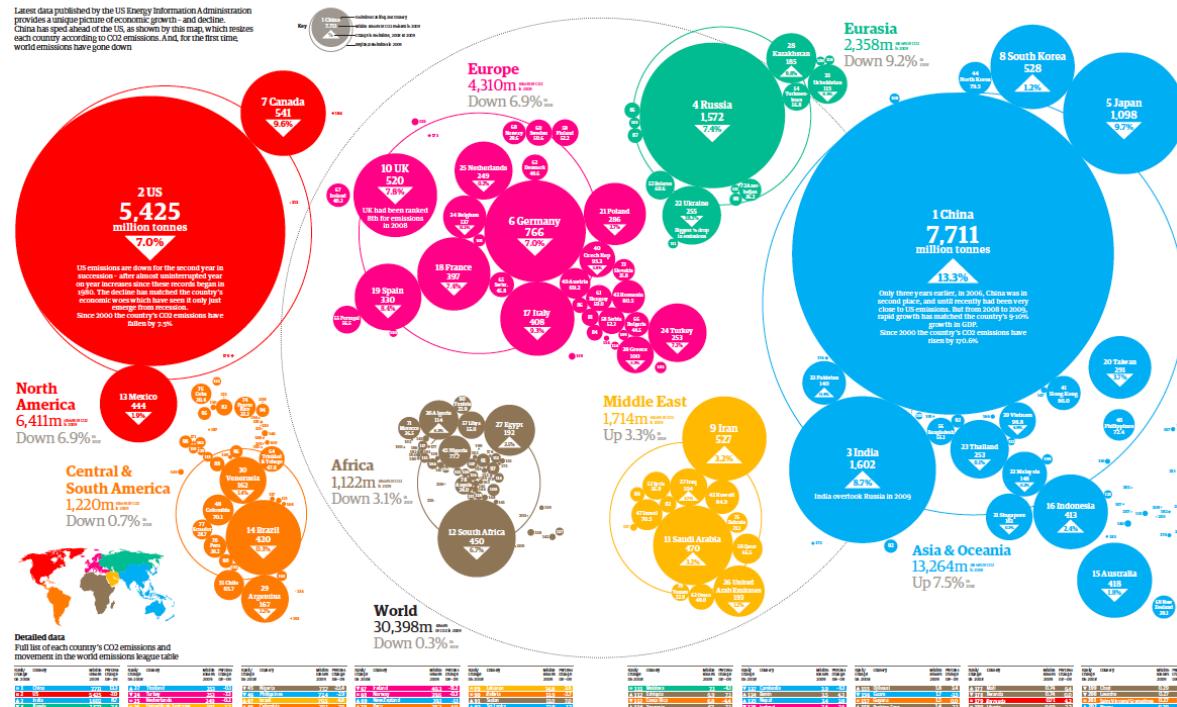
<http://www.worldmapper.org/>

http://www.worldmapper.org/posters/worldmapper_map2_ver5.pdf

Dorling Cartogram vs Network Connection Map

An atlas of pollution: the world in carbon dioxide emissions

Latest data published by the US Energy Information Administration provides a unique picture of economic growth - and decline. China has sped ahead of the US, as shown by this map, which resizes each country according to CO₂ emissions. And, for the first time, world emissions have gone down



Summary

- Implementation of data representation and graphical presentation with visualization tool
- Method Classification
- Types of Questions
- Types of Charts/Graphs
- Graph = Data + Visual

