

07 Data Visualisation: Encoding Data

DMV

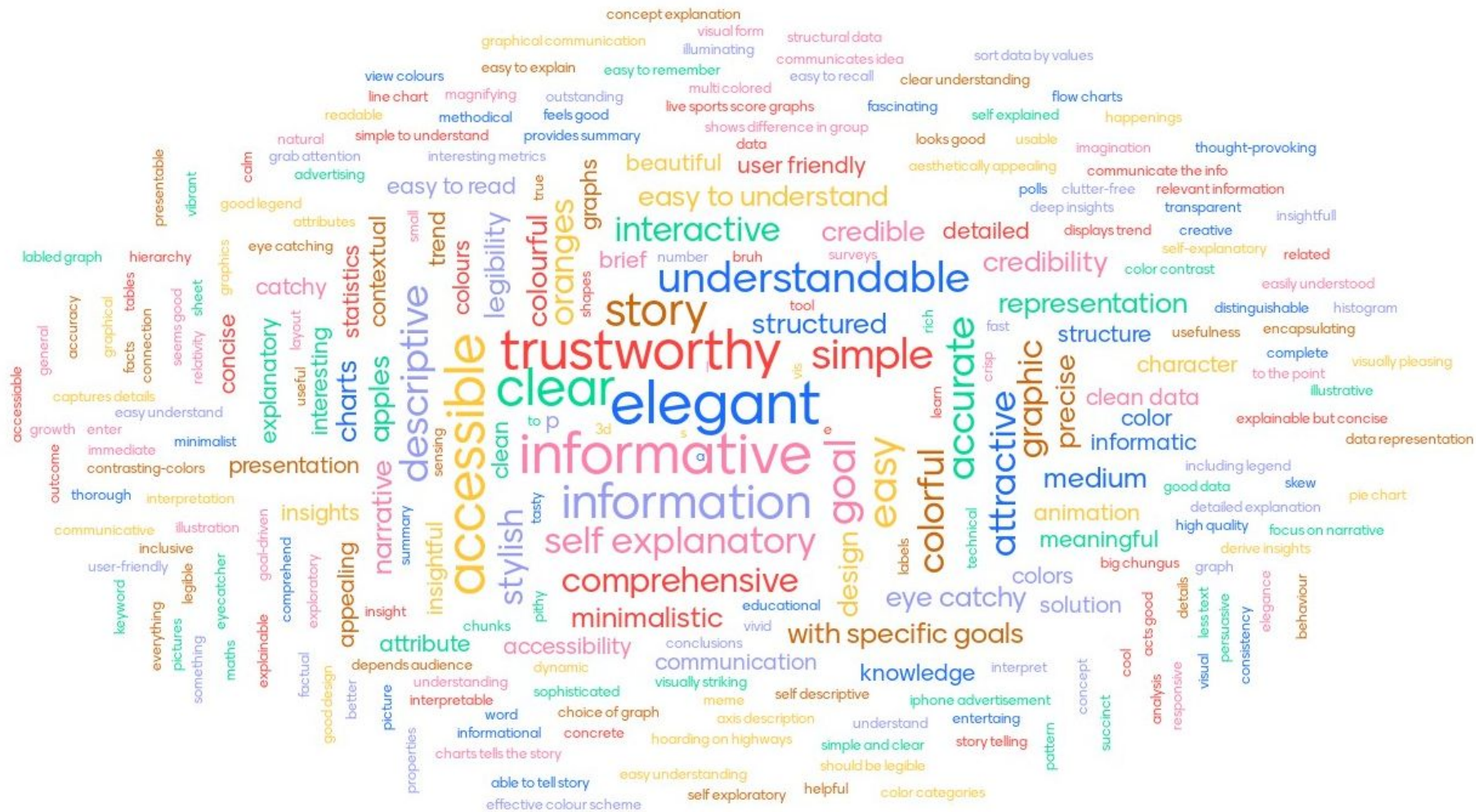
suzanne.little@dcu.ie

Today: Data Visualisation (part II)

- Recap
- Data representation
- Chart Types
- Choosing your chart

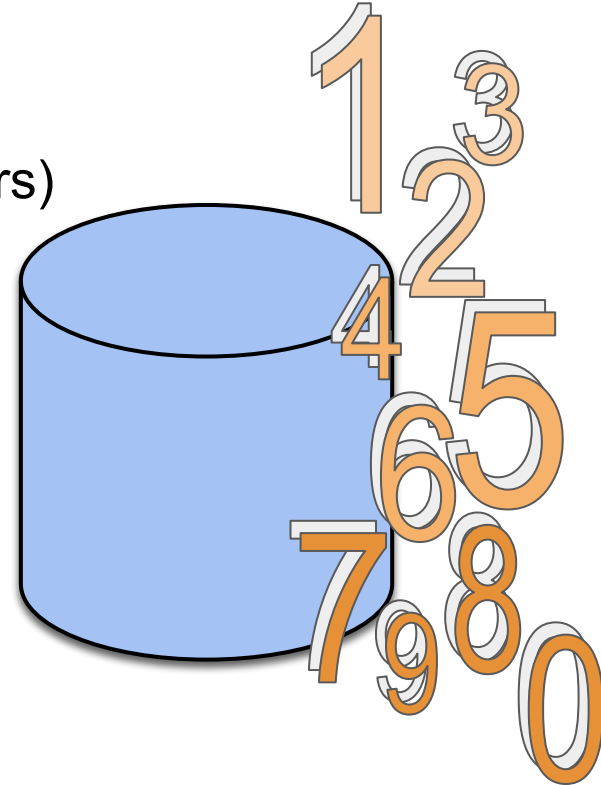
Recap

- What is a “good” visualisation?
- 4 components of communication
 - Sender, Message, Medium, Receiver
- 4 types of Graphic Communication
 - Information, Persuasion, Education, Entertainment
- Opinions on good visualisation

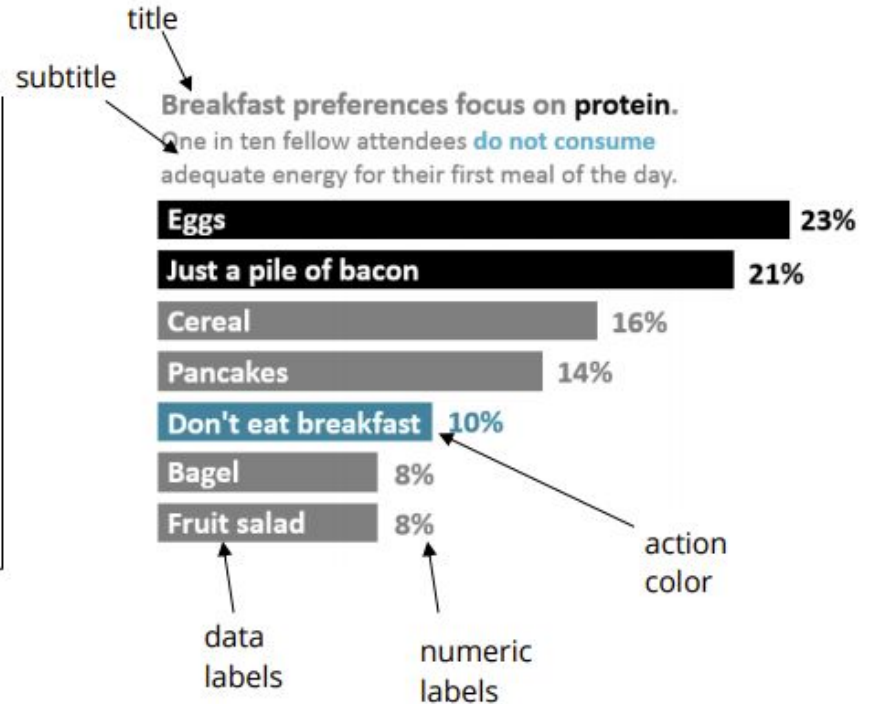
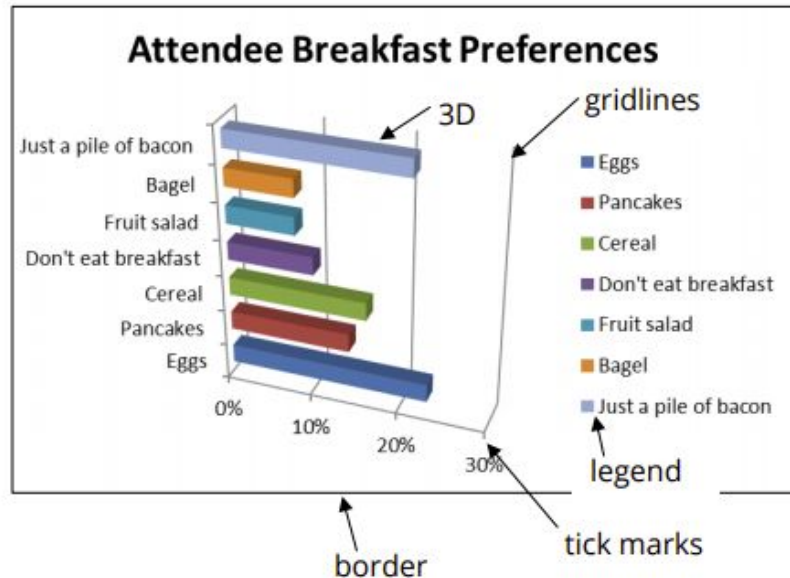


Categorising data

- Structured or Unstructured
 - Tabulated or Raw
- Qualitative (description) or Quantitative (numbers)
- Text (categorical), Numbers (numeric)
 - Documents, Images, Video, Audio, 3D
- Discrete or Continuous
- Nominal, Ordinal, Interval, Ratio
- Temporal (or Time Series)
- Geographic (or Spatial)



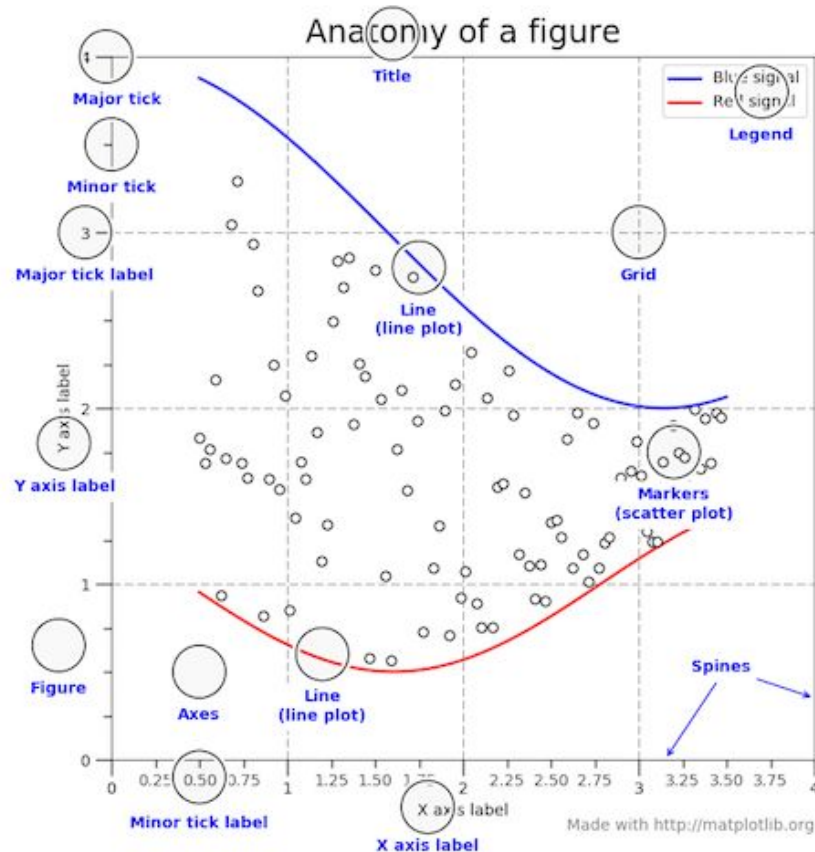
What is a graph? (or chart)



<http://stephanieevergreen.com/updated-data-visualization-checklist/>

What is a graph? (or chart)

<https://matplotlib.org/examples/showcase/anatomy.html>

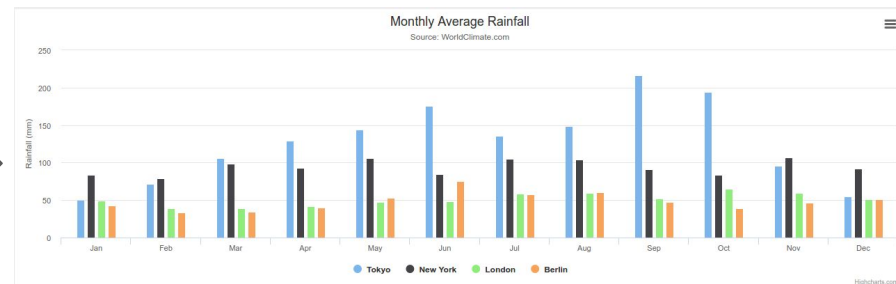
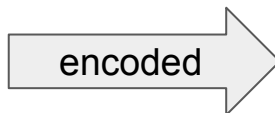


A graph *encodes* data

encode? represent in an alternative way

2 100 39 91 93 98 94 89 30 82

name, age, id, colour, language



→ **Marks & Attributes**

Data representation: **Marks**

Point



No variation

Eg. Quantity through position
(scatter plot)

Data representation: **Marks**

Point



No variation

Eg. Quantity through position
(scatter plot)

Line



1 dimension

Eg. Quantity through variation in
size (bar chart)

Data representation: **Marks**

Point



No variation

Eg. Quantity through position
(scatter plot)

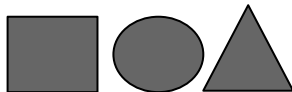
Line



1 dimension

Eg. Quantity through variation in
size (bar chart)



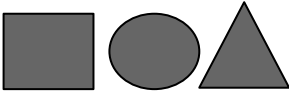
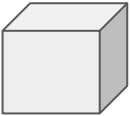
Area



2 dimensions

Eg. Quantity through size and
position (bubble chart)

Data representation: **Marks**

Point		No spatial variation	Eg. Quantity through position (scatter plot)
Line		1 spatial dimension	Eg. Quantity through variation in size (bar chart)
Area		2 spatial dimensions	Eg. Quantity through size and position (bubble chart)
Form		3 spatial dimensions	Eg. Quantity through variation in size/volume (proportional shape)

Data representation: **Attributes**

Quantitative

Position

Size (length, area, volume)

Angle/Slope

Quantity

Colour: Saturation

Colour: Lightness

Pattern

Motion

Categorical

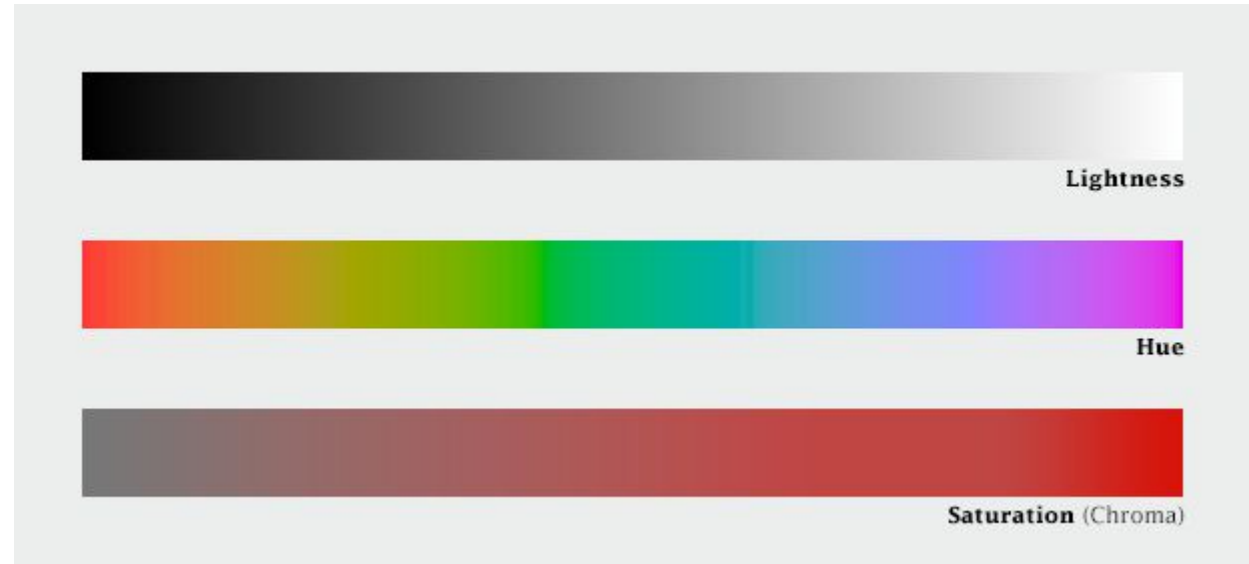
Colour: Hue

Symbol/Shape (😊 😞 ♻️ Ω)

Relational

Connection/Edge

Containment

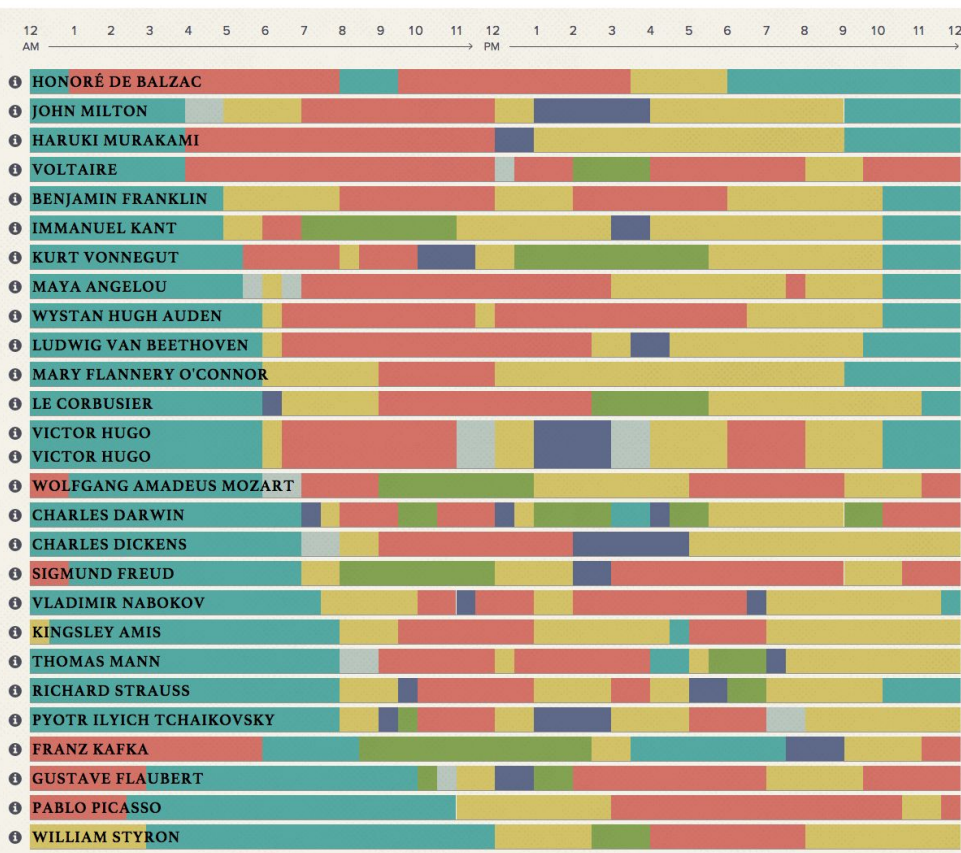


THE DAILY ROUTINES OF

FAMOUS CREATIVE PEOPLE

Turns out great minds don't think alike. Discover how some of the world's most original artists, writers and musicians structured their day, based on 'Daily Rituals' by Mason Currey. Filter the different categories by toggling on or off, and hover over the colored bars to learn more about the daily routines.

SLEEP CREATIVE WORK DAY JOB/ADMIN FOOD/LEISURE EXERCISE OTHER



<https://podio.com/site/creative-routines>

(charts, graphs, tables, figures, maps, plot, diagram, ...)

Chart types

Categorical: comparing categories and distributions of quantitative values

Hierarchical: Charting part-to-whole relationships and hierarchies

Relational: Graphing relationships to explore correlations and connections

Temporal: Showing trends and activities over time

Spatial: Mapping spatial patterns through overlays and distortions

Charts (a curated selection) - Categorical

- Bar graph: comparisons of quantitative values from different categories
- [Dot plot](#): Like bar but use a point or symbol to indicate the value so can include colour, area, shape to capture extra dimensions.
- [Circle packing](#): comparisons of values using area, shape, colour, layout
- [Polar chart](#): (also [radar or spider](#)) radially plotted bar chart showing 3+ quantitative measures

comparison

Charts (a curated selection) - Categorical

- Bar graph: comparisons of quantitative values from different categories
 - [Dot plot](#): Like bar but use a point or symbol to indicate the value so can include colour, area, shape to capture extra dimensions.
 - [Circle packing](#): comparisons of values using area, shape, colour, layout
 - [Polar chart](#): (also [radar or spider](#)) radially plotted bar chart showing 3+ quantitative measures
-
- Box-and-Whisker plot: common in statistical analysis
 - Histogram (not a bar chart): frequency and distribution
 - Word cloud: frequency of concepts

comparison

distribution

Quick word on histograms and box/whiskers

Statistical graphs - very useful and powerful!

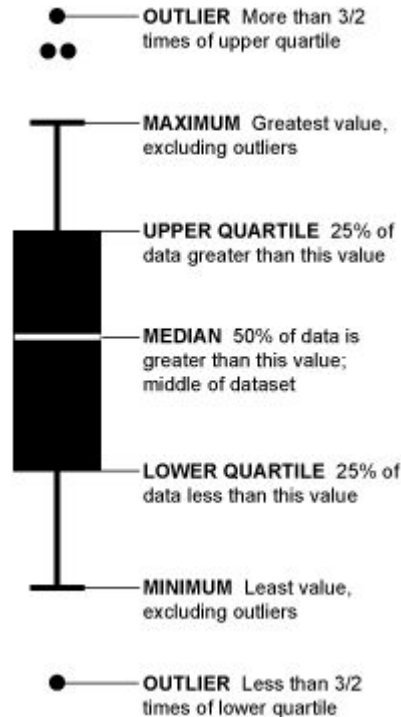
Do not confuse histogram with bar chart

Learn how to read a box/whisker plot

But remember non-experts often struggle with them

Histogram → rich visualisation of distributions

Boxplot → comparing distributions between several groups



Charts (a curated selection) - Hierarchical

- Pie charts: how quantities make up a whole
- [Waffle charts](#): aka square pie, coloured grid squares to show quantities
- [Stacked bar chart](#): breakdown values within bar
- [Treemap](#): enclosed hierarchical display
- Venn diagram: relationships between sets and collections

part-to-whole

Charts (a curated selection) - Hierarchical

- Pie charts: how quantities make up a whole
 - [Waffle charts](#): aka square pie, coloured grid squares to show quantities
 - [Stacked bar chart](#): breakdown values within bar
 - [Treemap](#): enclosed hierarchical display
 - Venn diagram: relationships between sets and collections
-
- [Dendogram](#): aka tree hierarchy, layout tree, clusters. Node-link diagram showing hierarchical relationships across multiple layers

part-to-whole

hierarchies

Charts (a curated selection) - Relational

- [Scatter plot](#): relationship between quantitative values for two categories
- [Bubble plot](#): relationship between 3 qualitative values (area, x position, y position)
- Heat map: quantitative values between 2 categorical dimensions (colour coded)
- Matrix chart: quantitative values between 2 categorical dimensions
- [Sankey diagram](#): categorical composition and qualitative flows

connections

Charts (a curated selection) - Temporal

- Line chart: change in quantitative values over time
- Area chart: coloured in line chart :-)
- [Stream graph](#): continuous changes in qualitative values in different categories over time

trends

Charts (a curated selection) - Temporal

- Line chart: change in quantitative values over time
 - Area chart: coloured in line chart :-)
 - [Stream graph](#): continuous changes in qualitative values in different categories over time
-
- [Gantt chart](#): start, finish & duration of difference categorical activities

trends

activities

Charts (a curated selection) - Spatial

Map projections - think about flattening an orange peel

<http://geoawesomeness.com/best-map-projection/>

https://youtu.be/KUF_Ckv8HbE

Charts (a curated selection) - Spatial

- Choropleth map: (aka heat map) quantitative values for distinct spatial regions
- Isarithmic map: (aka contour map) quantitative values linking spatial regions
- Proportional symbol map: represent values by proportionally sized areas overlayed on map

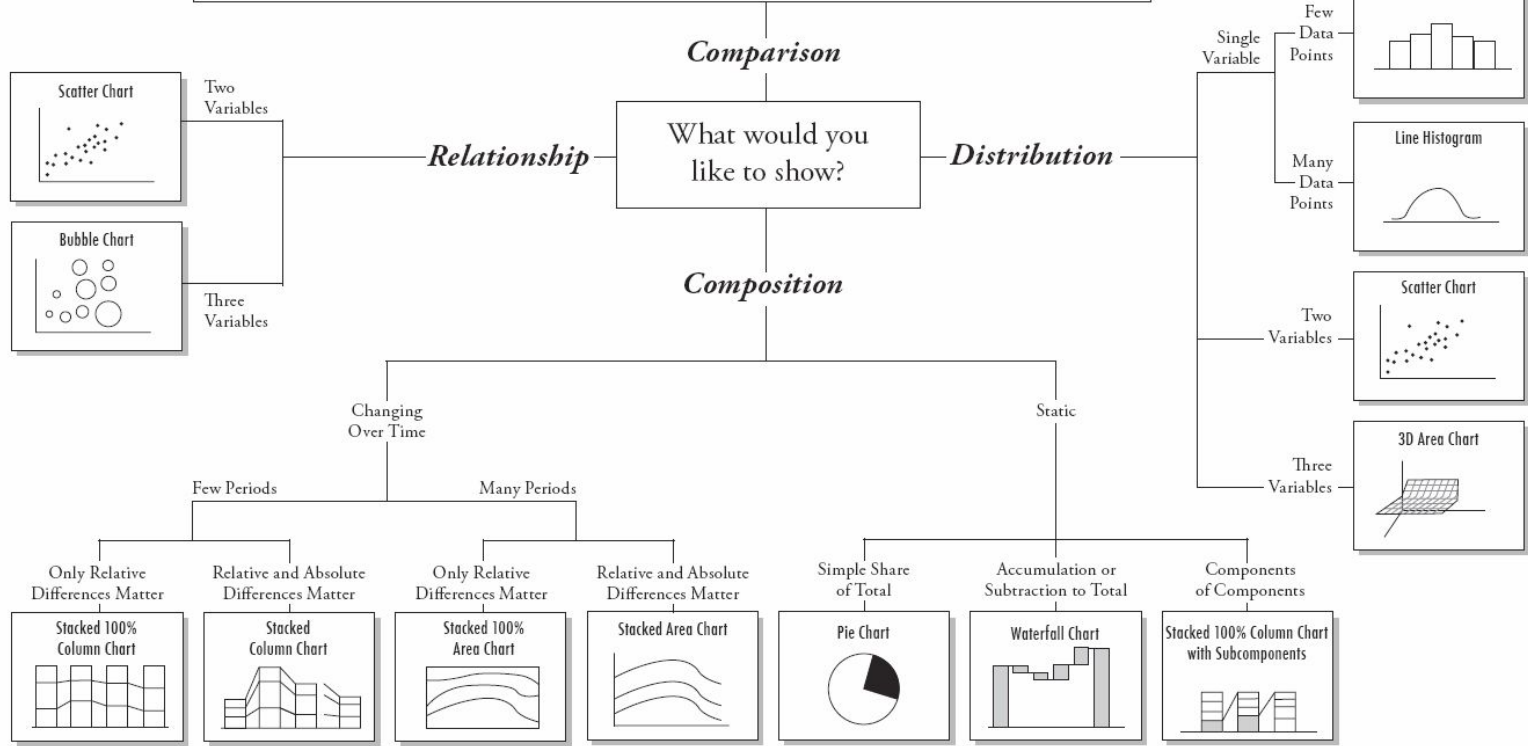
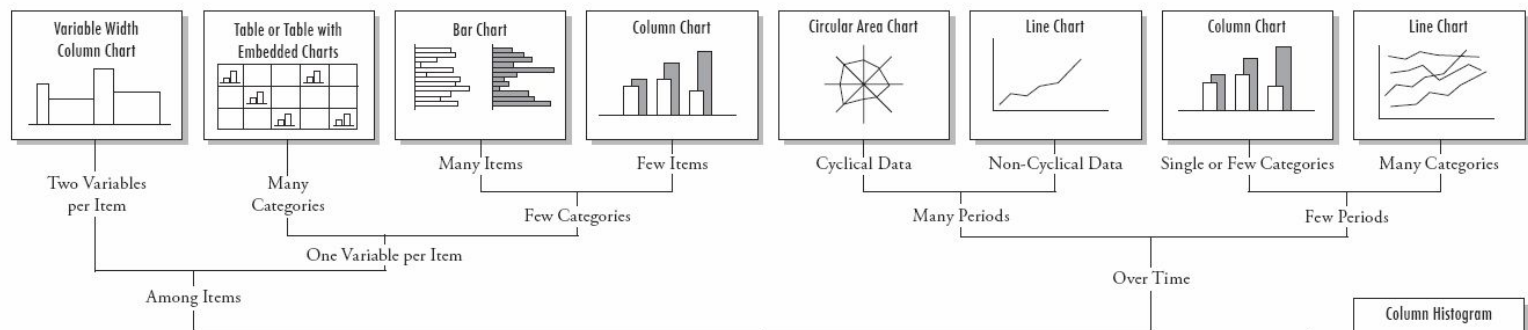
overlays

Charts (a curated selection) - Spatial

- Choropleth map: (aka heat map) quantitative values for distinct spatial regions
 - Isarithmic map: (aka contour map) quantitative values linking spatial regions
 - Proportional symbol map: represent values by proportionally sized areas overlayed on map
-
- Area cartogram: distort map spatial regions to show value
 - Dorling Cartogram/Grid map: arrange regular shapes into map using colour to indicate category

overlays

distortions



Exercise

In pairs/threes, discuss the best graph type for your question.

Categorical: Comparing categories and distributions of quantitative values

Hierarchical: Charting part-to-whole relationships and hierarchies

Relational: Graphing relationships to explore correlations and connections

Temporal: Showing trends and activities over time

Spatial: Mapping spatial patterns through overlays and distortions

Tools for Visualisation?



Excel/Google sheets

Photoshop/GIMP

Powerpoint

Tableau

PowerBI

Qlikview

R - ggplot

Pandas .plot()

Python - matplotlib, seaborn,
bokeh

Plot.ly

D3.js & other javascript libraries

[Overview of Python Visualisation Libraries \(with example notebooks\)](#)

Tools to create visualisations -

<https://loop.dcu.ie/mod/page/view.php?id=1651443>

- Programming languages
 - Document discussing Python Libraries on Loop
<https://loop.dcu.ie/mod/url/view.php?id=1651494>
- Dedicated tools (many!)
 - Tableau - <https://www.tableau.com/academic/students>
- Web-based, interactive options like D3.js and other Javascript libraries (exercises will be on loop for next week)

D3.js

- <https://d3js.org/>
- Data Driven Documents
- JavaScript library
- Transform data to standard web formats (HTML, SVG, CSS)
- Good for interactive and dynamic browser visualisations
- “D3 does not replace the browser’s toolbox, but exposes it in a way that is easier to use”

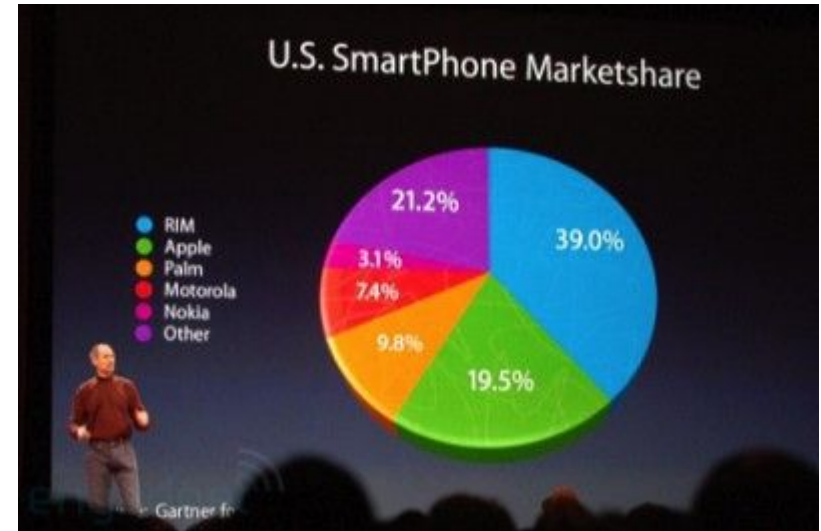
D3.js doesn't ...

- support older browsers
- generate prepared visualisations for you (unlike Excel, Tableau etc.)
 - so you generally don't do **Processing** or **Analytics** in D3.js
- ~~handle bitmaps (non-vector graphics) like the tiles on Google Maps (although there are ways around this)~~ Use [leaflet.js](https://d3js.org/leaflet.js)
- hide your original data - it's all sent to the browser (client) to do the graph generation. So be sure you want it exposed!

Critiquing designs - reading for week 8

<https://www.washingtonpost.com/graphics/politics/2016-election/trump-charts/>

<https://simonrogers.net/2013/03/15/a-conversation-with-stephen-few-about-data-visualisation-kind-of/>



Tonight's lab

Document: Python Libraries and other tools for visualisation → has links to three example notebooks for using Matplotlib; Seaborn and Bokeh

Review the linked notebooks

Your challenge: “Notebook: Create a graph using Python” You have the outputs of the cells. Can you work out what code is used to create them?

Resources

Chapter 6 of Data Visualisation (Andy Kirk) or Chapter 5 of ebook (<http://site.ebrary.com/lib/dublincu/Doc?id=10642563>) covers most of the types.

“A tour through the Visualization Zoo” (ACM publication)
<http://queue.acm.org/detail.cfm?id=1805128> - See also interactive presentation on loop (<https://loop.dcu.ie/mod/resource/view.php?id=79908>)

“The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations”, Ben Schneiderman,
http://www.interactiondesign.us/courses/2011_AD690/PDFs/Shneiderman_1996.pdf

More resources

<https://dsaber.com/2016/10/02/a-dramatic-tour-through-pythons-data-visualization-landscape-including-ggplot-and-altair/>

Some examples using python,

<https://towardsdatascience.com/5-quick-and-easy-data-visualizations-in-python-with-code-a2284bae952f>

<https://datavizcatalogue.com/> Interactive website with a catalogue of different chart types

<https://chaione.com/blog/building-blocks-graphs/> Good summary of main graph components