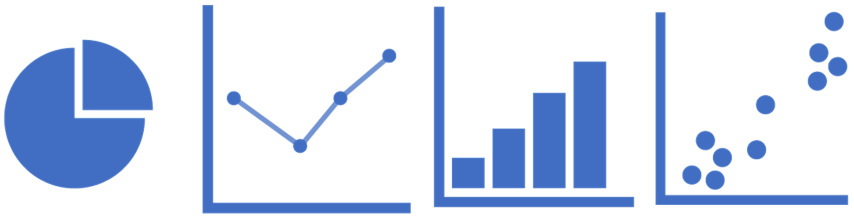
A curated selection of charts for DMV

(or graphs, tables, figures, maps, plots, diagrams, ...)

Suzanne Little, suzanne.little@dcu.ie



You may already be familiar with the main three or four charts -- pie, line, bar and perhaps a scatter plot. We've seen examples of most of these in the courses so far. You may also know about histograms and boxplots from your maths and statistics courses. There are many, many more types of graphs and variations of these.

This document contains a curated reference list of charts. There are many other types of charts that you could include and many of the charts given here have slight variations. You can also create grids or dashboards of small charts ("small multiples") that have a different impact. The charts I have included are an illustration of the variety and represent the set of graphs that I will reasonably choose from for any assessment. Therefore, I expect you to be able to recognise the chart types provided in the next five steps, what category they belong to and how they work. For a full discussion and more chart options for each category please refer to chapter 6 of Kirk: Data Visualisation.

Why is it useful to group and classify charts? In my opinion the actual label or category of the chart is less important than you taking the time to consider the type of data you are graphing and the message you are trying to communicate. Ask yourself two key questions:

1. Is my data categorical or quantitative? Or multiple types?
2. Do you want the audience to remember a single number, a comparison, a trend, a distribution or a breakdown?

Once you have some answers to these two questions then the chart categories will help you narrow down the most suitable choice.

Finally, while it's fun to look at the variety of graphs and charts in the next steps and useful to understand how to read them, you'll probably find that you create only a small subset of these in practical regular use (unless you become a data journalist or graphic designer).

Categories of charts -- CHRTS

In this course I am using the grouping proposed by Andy Kirk (2016) that divides charts into the following categories or families:

- **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

Appropriately this forms the mnemonic: CHRTS :-) Kirk acknowledges the convenience of this acronym.

Quote from Kirk

> "I know what you're thinking: 'well that's a suspiciously convenient acronym'! Honestly, if it was as intentional as that I would have tried harder to somehow crowbar in an 'A' family. OK, I did spend a lot of time, but I couldn't find it and it's now my life's ambition to do so. Only then will my time on this planet have been truly worthwhile. In the meantime, CHRTS is close enough. Besides, vowels are hugely overrated."

I've chosen to use this grouping as it provides a fairly complete selection, is easy to remember and the associated section in the book (Chapter 6) has a very thorough set of examples and explanations. It's also relatively application agnostic -- that is, it's useful for business use, data journalism, academic research, etc.

There are other ways to group or analyse chart types but this grouping might help you to "chunk" the information which is a way of improving memory and understanding.

Categorical examples

Comparing *categories* and *distributions* of *quantitative* values

Comparison

The classic: Bar graph.
A bar graph provides comparison of *quantitative* values from different *categories*.
AKA: column chart (generally used for vertical bar). NOT a histogram (see the next step).
Common variations: horizontal or vertical, stacked (see part-to-whole) or clustered.

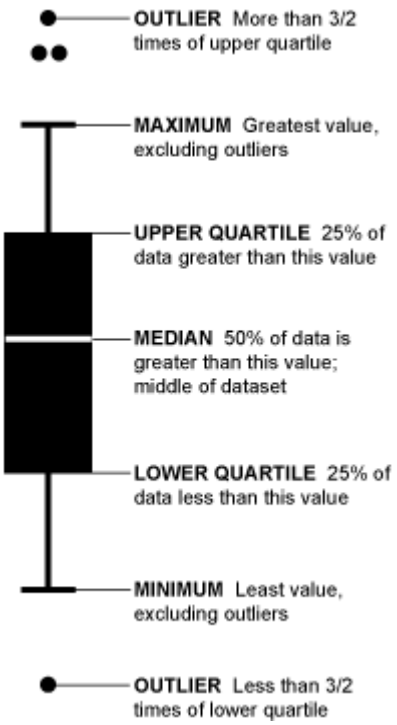
Bar charts are a workhorse for visualisation. They suit many situations and are generally well understood. They do lack a little bit of the “cool” factor so good design choices will improve the impact of your graph.

> *“There’s a strand of the data viz world that argues everything could be a bar chart. That’s possibly true but also possibly a world without joy.”* Amanda Cox, Editor, The Upshot

Other examples:

- Dot plot: Similar to a bar but use a point or symbol to indicate the value so can include colour, area, shape to capture extra dimensions. Variation is the connected dot plot. AKA: 1 dimensional scatterplot
- Circle packing: comparisons of values using area, shape, colour, layout
- Polar chart: (also radar or spider) radially plotted bar chart showing 3+ quantitative measures. Radar charts are able to display multivariate observations with an arbitrary number of variables.

Distribution



- Box-and-Whisker plot (above): common in statistical analysis
- Histogram (**not** a bar chart): frequency and distribution
- Word cloud: frequency of concepts

You've already seen word clouds in step 3.2.6. Box plots or box-and-whiskers plots and histograms are very common in statistical analysis. They have a few unique points to be aware of if you choose to use them as a visual communication tool. Remember the following:

Do not confuse a histogram with a bar chart. The y-axis for a histogram is the count of frequency and the x-axis is the bins (groups the instances are divided into). A typical vertical bar chart looks similar but has columns on the categories along the x-axis and the values on the y-axis. The gap between the bars is often more pronounced in the bar chart and non-existent in the histogram.

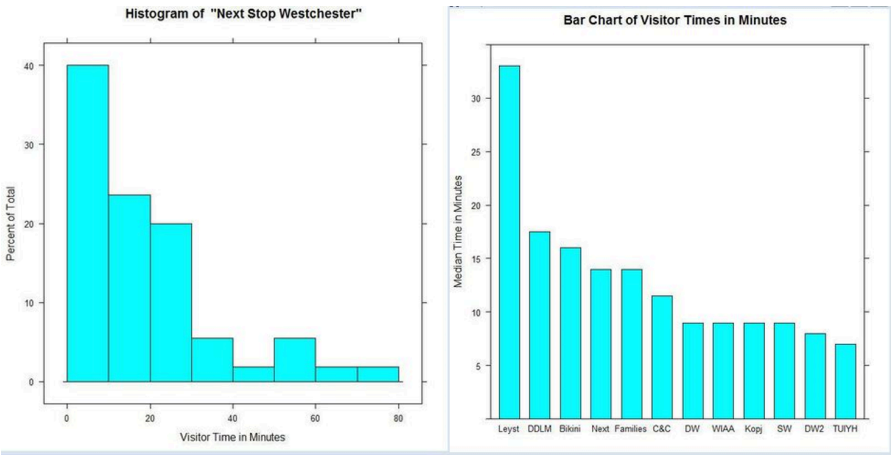


Fig: Histogram (Left) Bar chart (Right) (from: https://www.forbes.com/sites/naomirobbins/2012/01/04/a-histogram-is-not-a-bar-chart/)

Learn how to read a box/whisker plot (diagram above) but remember non-experts often struggle with them. And a “non-expert” is anyone who doesn’t use them regularly. They are very useful but you might need to explain and educate your audience.

Use these plots when you need:

- Histogram → rich visualisation of distributions
- Boxplot or Boxplots → comparing distributions between several groups

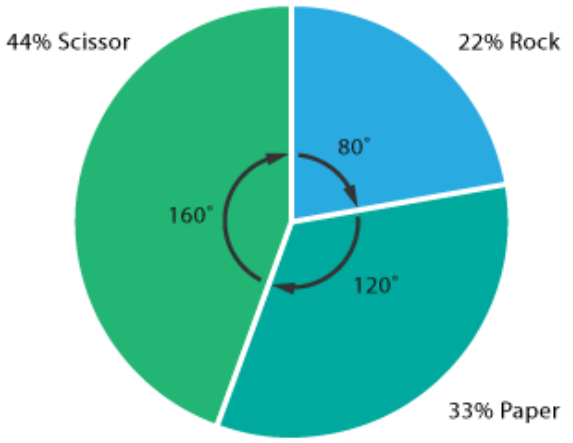
Hierarchical examples

Charting *part-to-whole* relationships and *hierarchies*

Used where you need to show the composition or break down of something. Especially where you have a category and percentages of a total (ie, they add up to 100%).

Part-to-whole

The classic (and highly controversial!): Pie charts
Pie charts break down quantities using the angles (wedges) in a circle.



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
2/9=22%	3/9=33%	4/9=44%	100%
Degrees for each "pie slice"			
(2/9) x 360 = 80°	(3/9) x 360 = 120°	(4/9) x 360 = 160°	360°

Fig: showing the how a pie chart is constructed from data (from: https://datavizcatalogue.com/methods/pie_chart.html)

Pie charts are generally disapproved of by most practitioners but are highly popular, especially in business situations. We'll talk more about why this is in the "Five simple tips" (spoiler: Tip 1 is don't use pie charts). For reference: Edward Tufte says that pie charts "should never be used," (1973, p178), Stephen Few (2012, p94) "I don't use pie charts and I strongly recommend you abandon them as well … pie charts communicate data poorly" while Knaflic (2016, p61) says "Pie charts are evil" though there's a moderating view from the Storytelling with data community. There are dissenting opinions (for example).

To avoid a pie chart check out these five alternatives.

Common variations: doughnut

Other examples:

- Waffle charts: aka square pie, coloured grid squares to show quantities make it easier to compare or evaluate via counting rather than angles.
- Stacked bar chart: breakdown values within bar. A variation of the stacked bar chart is the waterfall.
- Treemap: enclosed hierarchical display
- Venn diagram: relationships between sets and collections

Hierarchies

- Dendogram: aka tree hierarchy, layout tree, clusters. Node-link diagram showing hierarchical relationships across multiple layers

Relational examples

Relational charts are used to show the *connections* between data points.

The classic: scatterplot
Shows the relationship between *quantitative* values for two *categories*.
Often used to show correlations, it may include a line of best fit or a trend line.

Other examples:

- 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

This family could also include charts like table lens (example) and pyramid (often a back-to-back histogram).

Temporal examples

Trends

The classic: line chart
A line chart shows the change in *quantitative* values over *time* (or possibly a sequence of equal intervals). Be careful not to use this where your x-axis has categories!
AKA: stock chart

- 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
- Slope graph: before and after changes in quantity over a number of variables

Activities

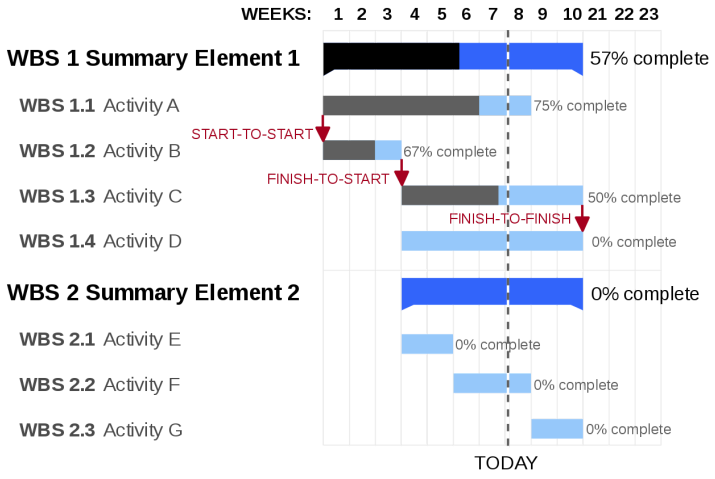


Fig: Example Gantt chart (from: https://en.wikipedia.org/wiki/Gantt_chart)

This is a special type of temporal chart best represented by the Gantt chart that you may have seen or used for project planning. This

encodes the start, finish and duration of different categorical activities often by different individuals or groups. Gantt charts are normally created using specialist project management software but can also be done in D3.js or using a spreadsheet.

Spatial examples

A word about geographic maps:
All world map projections are an approximation of reality. Think about trying to peel an orange and then completely flattening the peel.

Fig: Different world map projections

Any use of a map in a visualisation relies heavily on the audience having a basic understanding of geography. Most people are familiar with the Mercator projection (the common map of the world). But if you show a map of Ireland to someone from the US will they understand that Dublin, the capital, is on the east coast? This is especially important for some of the map variations used to show spatial information.

If you are interested then here's some more about map projections:

- Which is the best map projection?
- Buzzfeed: Maps That Prove You Don't Really Know Earth (2:23)
- XKCD: What your favorite map projection says about you

There are two main ways to encode spatial data using a map: Overlays and Distortions.

Note: spatial visualisations don't only have to be applied at the scale of geographical maps. You could also use a floor plan or other local map. Imagine trying to show where people commonly sit in a lecture theatre. As long as your audience shares and understands your spatial model these maps are effective.

Overlays

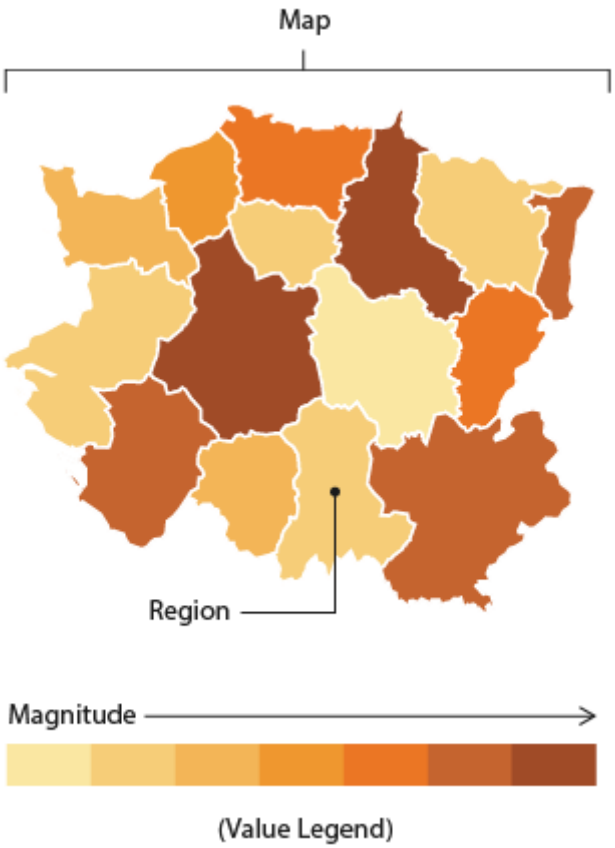


Fig: representation of a choropleth map (from: https://datavizcatalogue.com/methods/choropleth.html)

The classic: Choropleth map
A choropleth map encodes quantitative values for distinct spatial regions often using a colour progression.
AKA: heat map

- Isarithmic map: (aka contour map) quantitative values linking spatial regions. Commonly seen in weather maps showing the connections between areas of pressure.
- Proportional symbol map: represent values by proportionally sized areas (like circles) overlayed on map

Distortions

Distortions play with the sizes of regions in the map or build a visual representation of the map layout using symbols. This relies on your audience knowing what the map should look like to understand the data being conveyed.

- Area cartogram: distort map spatial regions to show value
- Dorling Cartogram/Grid map: arrange regular shapes into map using colour to indicate category

You can see some examples of different spatial visualisations applied to both US and UK elections - https://communityhub.esriuk.com/geoxchange/2016/11/1/us-election-2016-battle-of-the-maps.

Exercise

Find examples online of graphs that you aren't familiar with:

- https://datavizcatalogue.com/index.html
- http://visualizationuniverse.com/charts/
- https://blockbuilder.org/search
- https://www.highcharts.com/demo
- http://chartmaker.visualisingdata.com/

References

Evergreen, S. (2016) “Effective Data Visualization”, Sage.
Few, S. (2012) “Show Me the Numbers”, Analytics Press.
Kirk, A. (2016) “Data Visualisation”, Sage.
Nussbaumer Knaflic, C. (2015) “Storytelling with Data”, Wiley.
Yau, N. (2011) “Visualize This: The FlowingData Guide to Design, Visualization, and Statistics”, Wiley.