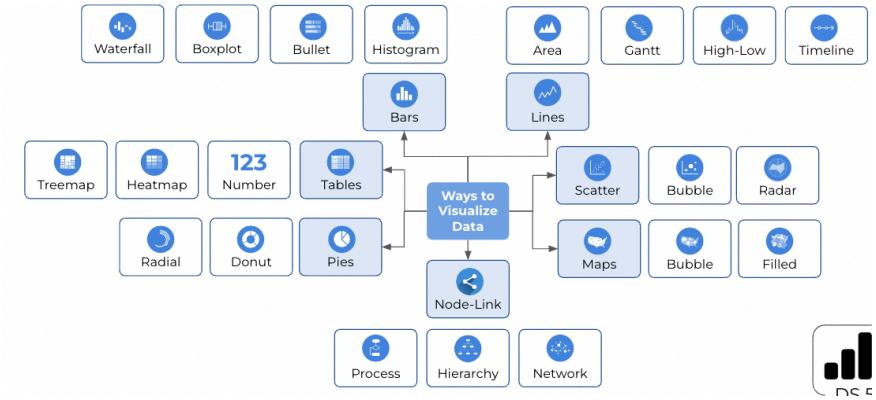
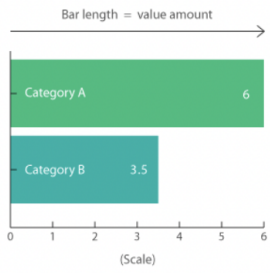
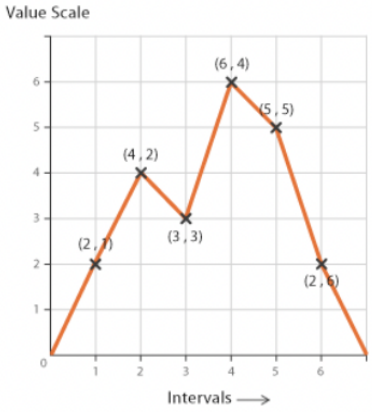
CDS DS 594
Lec 06

Chart Types

1. Terminology
    a. Dimension: A data element that is represented by categorical values (i.e. north, east, west, south)
    b. Metric: A Data element that is represented by numerical values (i.e. 314, 5001)
    c. Temporal: A data element that is represented by a date, time or both (i.e. 2023-09-26 15:30:00, 12-DEC-2012)
    d. Geographic: A data element that is represented by a physical location (i.e. latitude and longitude, city, coordinates)
2. Chart Selection Guide
    a. 

3. Bars
    a. Good for comparing categories across a dimension
    b. A bar chart shows discrete, metric comparisons across categories. One axis of the chart shows the specific categories being compared and the other axis represents a discrete metric
    c. Bar charts discrete data is categorical and therefore answers the question of "how much" in each category
    d. Bar charts can be either horizontal or vertical bars (column chart). Horizontal bar charts should be used when there are a large number of categories, and labels are long
    e. 

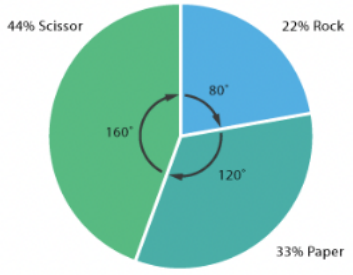
4. Bar Chart Variations
    a. Histogram
        i. The distribution of data across a metric to show how data are concentrated and dispersed.
        ii. Useful for estimating probability distribution
    b. Bullet
        i. Comparison of multiple metrics, normalized against metric targets and thresholds
    c. Boxplot
        i. Distribution of a metric, within groups, using quantile ranges.
        ii. Outliers are flagged separately
    d. Waterfall
        i. The cumulative effect of sequentially introduced positive or negative values
        ii. These intermediate values can either be temporal or dimension based
5. Lines
    a. Good for showing trends over time
    b. A Line Graph is most frequently used to show temporal trends and analyse how the data has changed over time.
    c. Typically, the y-axis has a quantitative value, while the x-axis is a timescale or a sequence of intervals. Negative values can be displayed below the x-axis.
    d. The line's journey across the graph can create patterns that reveal trends in a dataset.
    e.



6. Line Chart Variations
    a. Area
        i. Line charts with area filled below the line.
        ii. Useful for showing relative distance between lines.
    b. High-Low
        i. Used to show the high, low and range of a metric within a time period, and overall trend across time periods.

c. Timeline
    i. Discrete events shown chronologically, either to scale or in sequence.
    ii. Allows for detailed description of each event.
d. Gantt
    i. Progress towards a goal measured with the start date and duration of various activities.

7. Pies
    a. Good for showing percentages between categories
    b. By dividing a circle into proportional segments, each arc length represents a proportion of each category, while the full circle represents the total sum of all the data, equal to 100%.
    c. Pie Charts are unsuitable for large datasets with many categories, and take up more space than alternatives such as 100% Stacked Bar Charts.
    d. Because it is harder to distinguish the size of a category's arc it is difficult to make accurate comparisons. Comparing a given category (one slice) to the total can often be more effective.



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°

    e.

8. Pie Variations
    a. Donut
        i. Similar to Pie Chart, but with space in the middle for supporting statistics or cleaner display.
    b. Radial
        i. Displays relative progress of multiple metrics towards a goal or whole.
        ii. Technically a bar chart displayed radially.

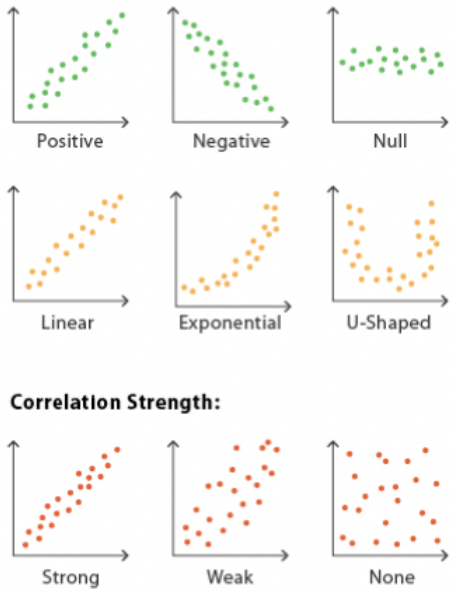
9. Tables
    a. Good for showing many rows and columns of data
    b. When exact values and underlying detail of an analysis needs to be displayed, tables are an excellent way to quickly display a large array of data.

c. Data can be scanned in a way that is often masked with other chart types.
d. Tables may also contain totals and subtotals, offering additional information.

	Total defects	A	B	C	D	E
A4636	131	37	21	28		45
A2524	86	20	24	21	1	20
A3713	75	17	13	18		27
A4452	73	5	33	17		18
A4088	72	14	16	12	2	28
A2103	68	14	13	14	1	26
A2156	68	16	13	19	2	18
A3681	66	12	16	9	1	28
A1366	50	11	15	12		12
A2610	39	5	7	12		15
Total	**728**	**151**	**171**	**162**	**7**	**237**

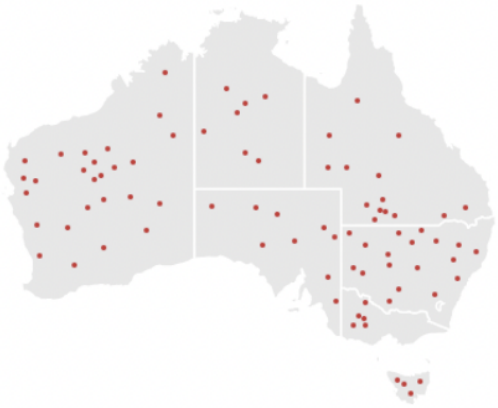
e.

10. Table Variations
    a. Number
        i. A single number representing a metric value of importance.
        ii. Numbers come with labels describing the meaning and units.
    b. Heatmap
        i. A cross tabulation with two dimensions and a metric.
        ii. Colors show relative value of metrics within each table cell.
    c. Treemap
        i. Displays hierarchical data using nested rectangles with an area proportional to the metric associated with a dimension.
        ii. Colors can distinguish dimensions.
11. Scatter Plots
    a. Good for showing relationships between metrics
    b. A Scatterplot places points on a Cartesian Coordinates system to display all the values between two variables. By having an axis for each variable, you can detect if a relationship or correlation between the two exists.
    c. Scatterplots are ideal when you have paired numerical data and you want to see if one variable impacts the other. However, do remember that correlation is not causation and another unnoticed or indirect variable may be influencing the results.
    d. The kind of correlation can be positive , negative or null. The shape of the correlation can be linear, exponential and U-shaped. The strength of the correlation can be determined by how closely packed the points are to each other on the graph. Points that end up far outside the general cluster of points are known as outliers.

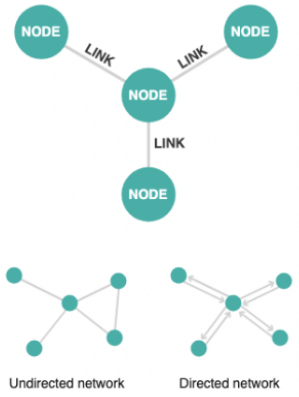


e.

12. Scatter Plot Variations
    a. Bubble
        i. A scatter plot that displays a third metric, showing relative value of by the size of a dot.
    b. Radar
        i. Comparison of three or more metrics across one or more groups.
13. Maps
    a. Good for showing location data
    b. Maps are a way of detecting spatial patterns or the distribution of data across a location, and can reveal patterns when the points cluster on the map.
    c. Maps are easy to grasp and are better at giving an overview of the data, but are not great for retrieving exact values.
    d. Maps can represent geographic (countries, cities, etc.) or physical (floor plans, circuit boards) locations



e.

14. Map Variations
    a. Bubble
        i. Similar to bubble charts, showing the relative value of a metric by the size of a dot.
    b. Filled
        i. Uses color or shade to represent the relative value or a metric, or the value of a dimension.
15. Node-Link Diagrams
    a. Good for showing relationships between entities
    b. Node-Link Diagrams consist of nodes and links (also called arcs and edges), with each edge directed from one node to another.
    c. Nodes and links can both display metrics related to frequency and magnitude.
    d. Node-Link Diagrams have numerous scientific and computational applications, ranging from biology (evolution, family trees, epidemiology) to information science (citation networks) to computation (scheduling).
    e. Node-Link Diagrams are used in process mining.
    f. 

16. Node-Link Variations
    a. Process
        i. Shows frequency and duration of steps in a process as defined by recorded activity.
    b. Hierarchy
        i. The hierarchical relationship between entities in a system.
    c. Network
        i. Non-hierarchical relationship between between entities in a system.
17. Chart of Charts
    a. Grouped
        i. Bar chart with an additional dimension displayed as multiple sets of bars.
    b. Stacked
        i. Bar, line or area charts stacking values of an additional dimension on top of each other.
        ii. Can represent proportional or absolute values.

c. Combo
    i. Often a combination of a bar and line chart used in a time series.
    ii. The bar represents a count and line a metric.
d. Sparklines
    i. A table to line or bar charts, with each row representing a dimension value.
    ii. Often accompanied by a number related to the chart.
e. Trellis
    i. A grid of charts with cell representing values of one or two dimensions.
    ii. Used to show results across a large number of dimension values.