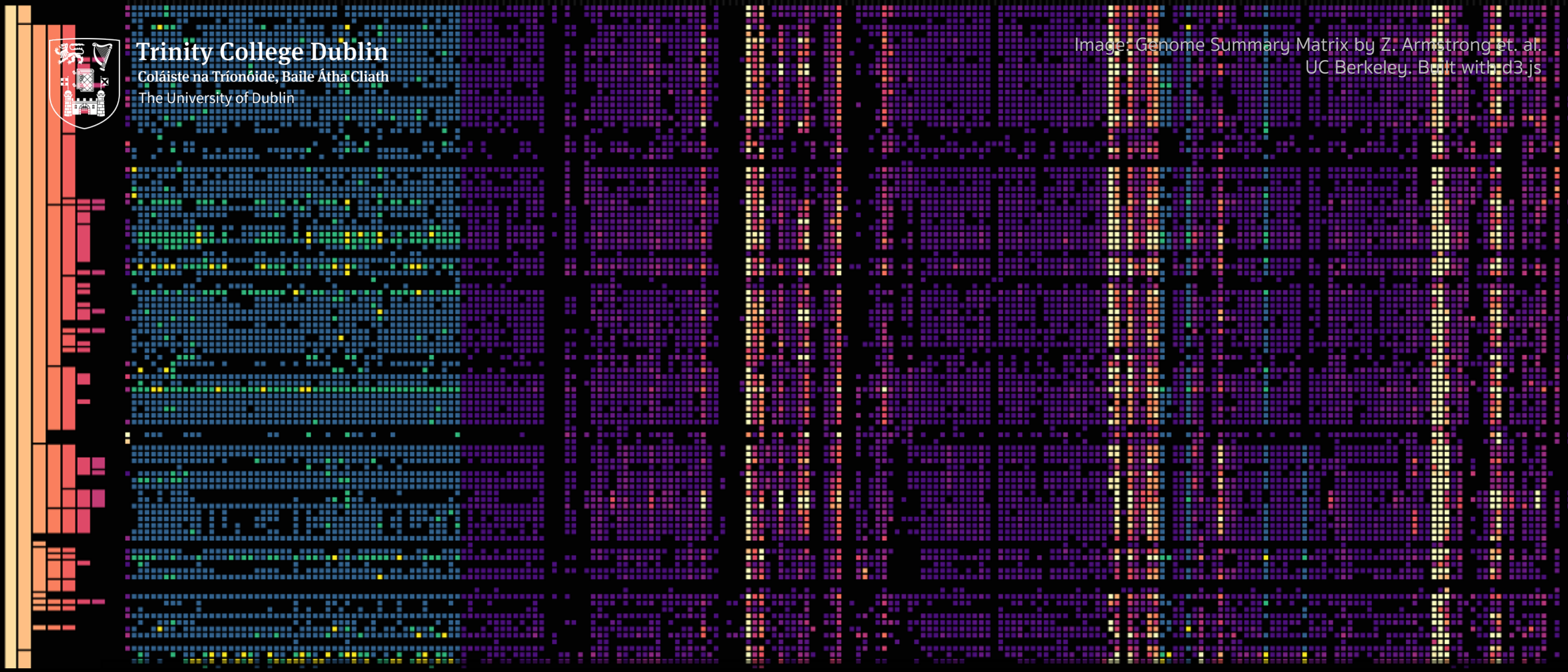


Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Image: Genome Summary Matrix by Z. Armstrong et. al. UC Berkeley. Built with d3.js

Visualization Idioms

05/10/2023

Part 1: Idioms that use position and size

Preview: **A2 - Visualization Analysis and Design**

N.B. This is a draft preview. Details will be released next week

ANALYSIS: Discuss the following visualizations with regard to the concepts of data, tasks and visual encoding channels as discussed in lectures and related readings.

- [How] Outline encoding channels and idioms employed in the visualization
- [What] Outline what appears to be the input main data/dataset being presented
- [Why] Outline one or more tasks that are supported by this visualization; what outcomes/conclusions/information can be derived using this visualization

DESIGN: Choose an interesting dataset (some suggested sources/datasets provided).

- [What] Outline what are the main data and dataset types comprised in the dataset
- [Why] Outline one or more visualization tasks that would be relevant (and you wish to support) for this dataset.
- [How] Outline a proposal for visualizing this data set in order to support the stated tasks, justifying the encoding channels and idioms you choose.

You may [but are not required to] use this dataset also for Assignment 3

Preview: **A3 - Novel Visualization**

N.B. This is a draft preview. Minor details will be added when officially released in week 8.

Objectives

- Design and implement a visualization for a data set, with the objective of supporting one or more specific tasks. You may use any tool.
- Choose an interesting dataset (some suggested sources/datasets provided). You may use the dataset from A2, but the data or task should be complex enough that it necessitates visualization using something more than a single standard chart
- Justify the design, and analyze if the final implementation effectively supports the stated task

The visualization is considered sufficiently "novel" if either one of the following requirements is met

- Visualization employs multiple known idioms that are combined e.g., a "dashboard"
- Visualization is unique in some way e.g., combining several encoding channels or a completely new chart type

Possible considerations in advance:

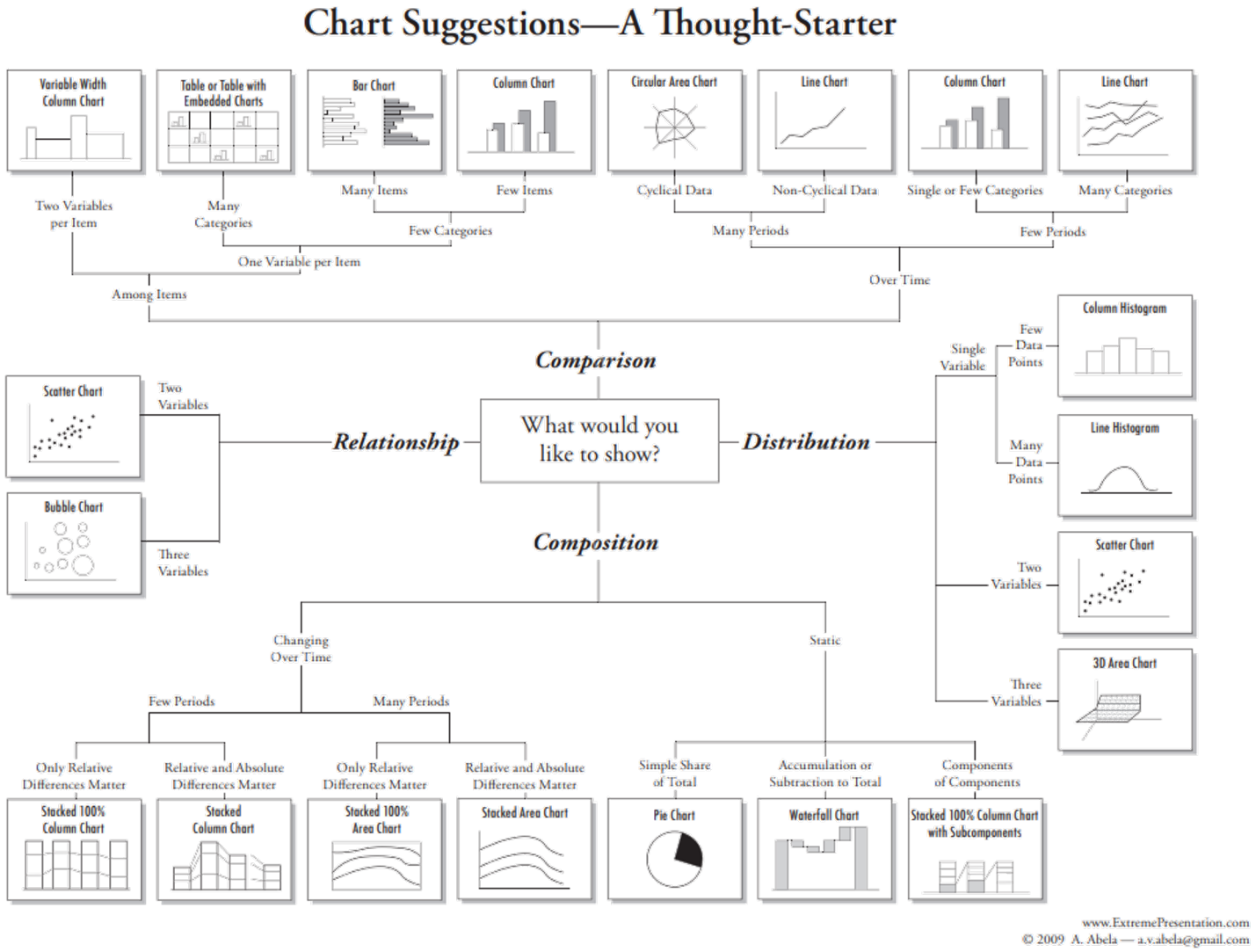
- Look for a data set you would be interested in visualizing [suggested sources will be posted shortly on blackboard],
- Think about the type/complexity of the data,
- Think about tasks that would motivate visualization
- Possibly, make use of assignment 1.2/1.3 to help decide what tool(s) you would use,
- Possibly use assignment 2 to plan the visualization

Visualization Idioms

A flowchart guide by A. Abela that helps choose an effective chart based on task

Also see:

- Juice Analytics' chart chooser based on this, which interactively chooses appropriate charts for excel and powerpoint [https://www.juiceanalytics.com/chartchooser]
- A colorized version by Qlik's Patrick Lundbald including a textual discussion: https://www.qlik.com/blog/third-pillar-of-mapping-data-to-visualizations-usage



[https://extremepresentation.com/design/7-charts/]

Visualization Idioms

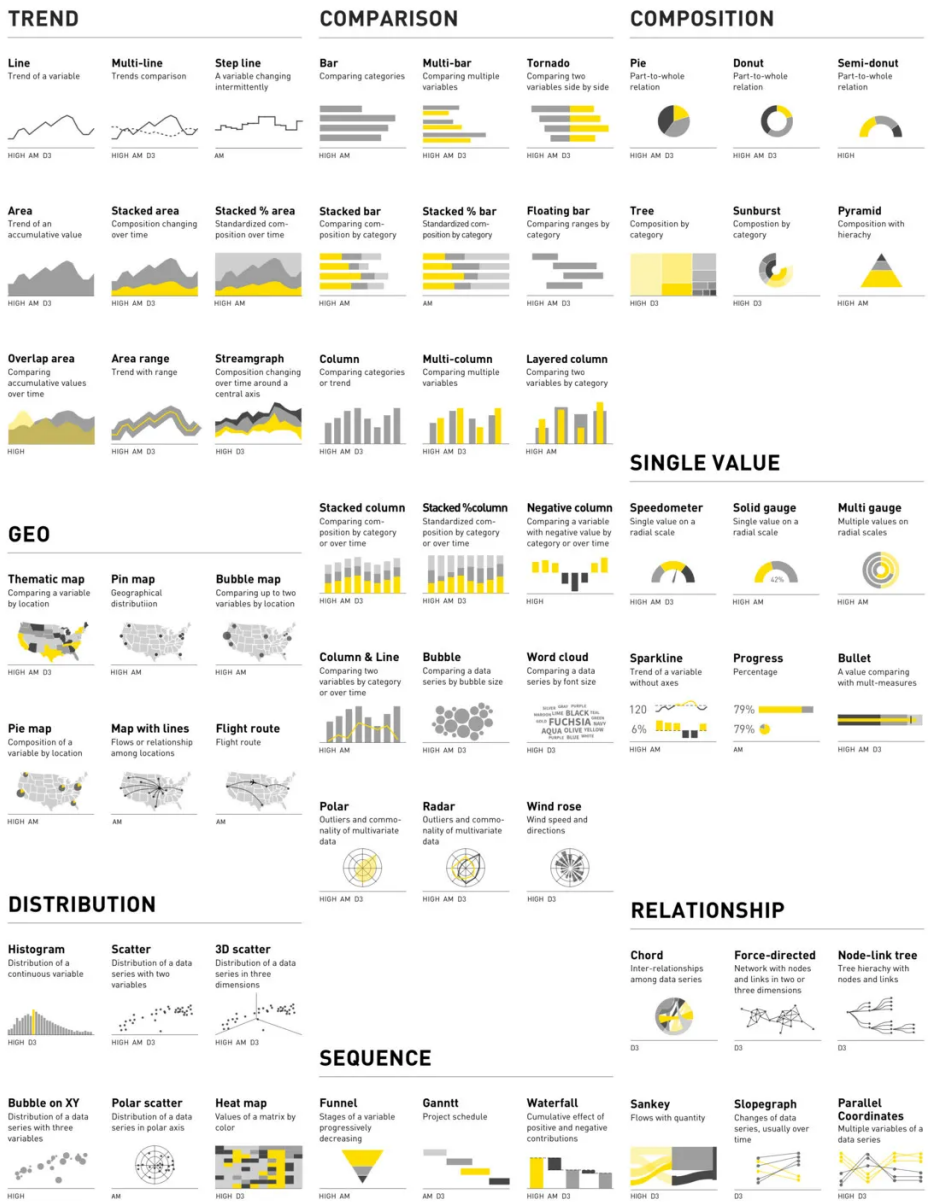
60 Charts to Visualize Your Data - A cheatsheet for UX designers. Wenjin Ju.
https://medium.com/@wenjunwu/60-charts-to-visualize-your-data-51344d7178d1

Please note: this comprises subjective individual opinions of the author, and mostly pertain to **explanatory** visualization but may provide some useful starting points.

1. Understand your data
2. Know your users
3. Be clear about your story
4. Pick most effective chart. Iterate.
5. Validate with real data

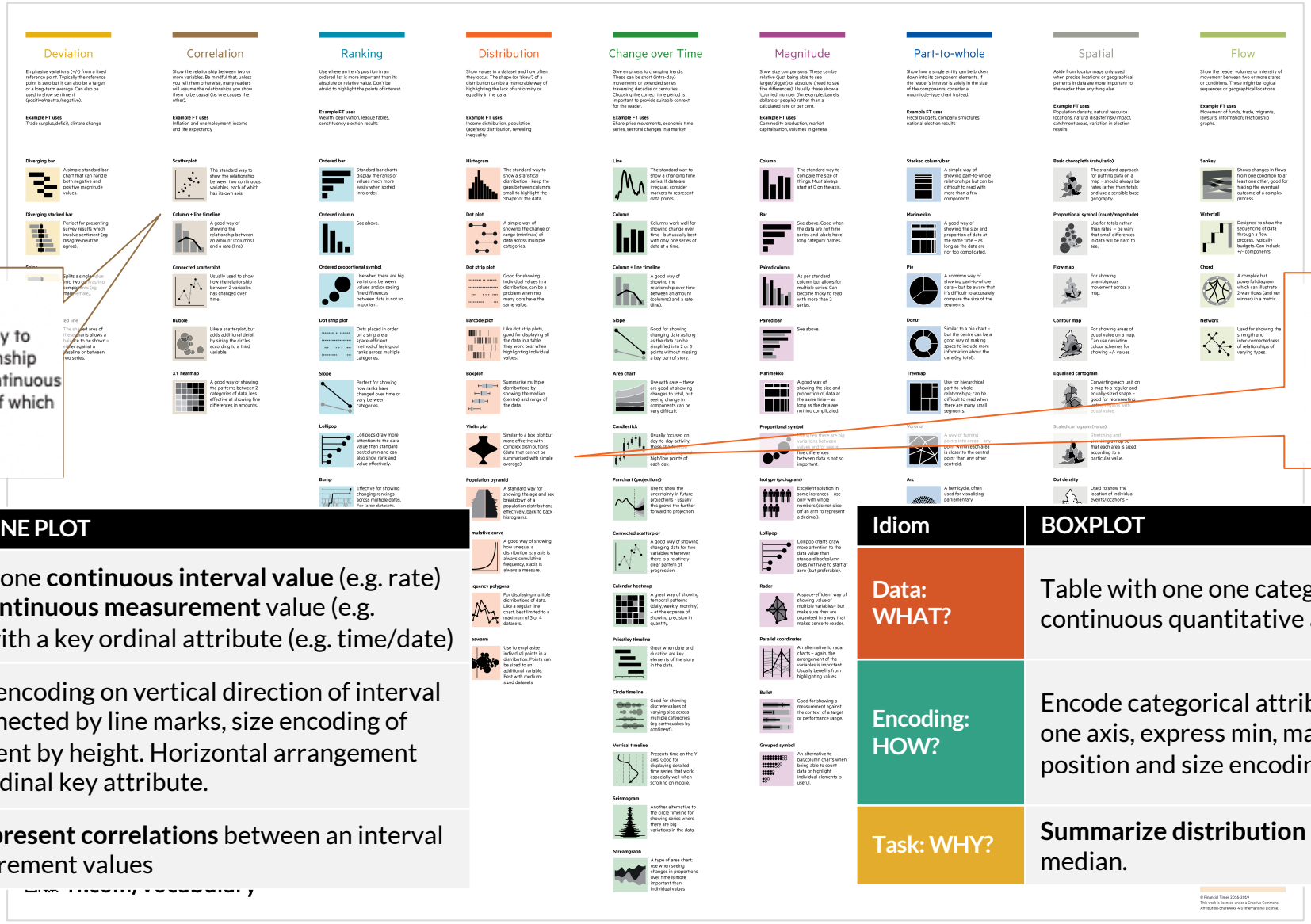
Chart-type suggestions:

- Is the data related to **time**? — Trend
- Is the data **comparable**? — Comparison
- Does the data have the **part-to-whole relationship**? — Composition
- Is it a data series with an interesting **distribution**? — Distribution
- Can the data be shown in **geographical** context? — Geo
- Is there any **relationship** among the datasets? — Relationship
- Does the data have a **sequence**? — Sequence
- Is there only 1 to 3 values to show? — Single value



Visualization Idioms

A more comprehensive “cheat sheet” for choosing charts, created by the Financial Times:
https://raw.githubusercontent.com/ft-interactive/chart-doctor/master/visual-vocabulary/poster.png



Correlation

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Distribution

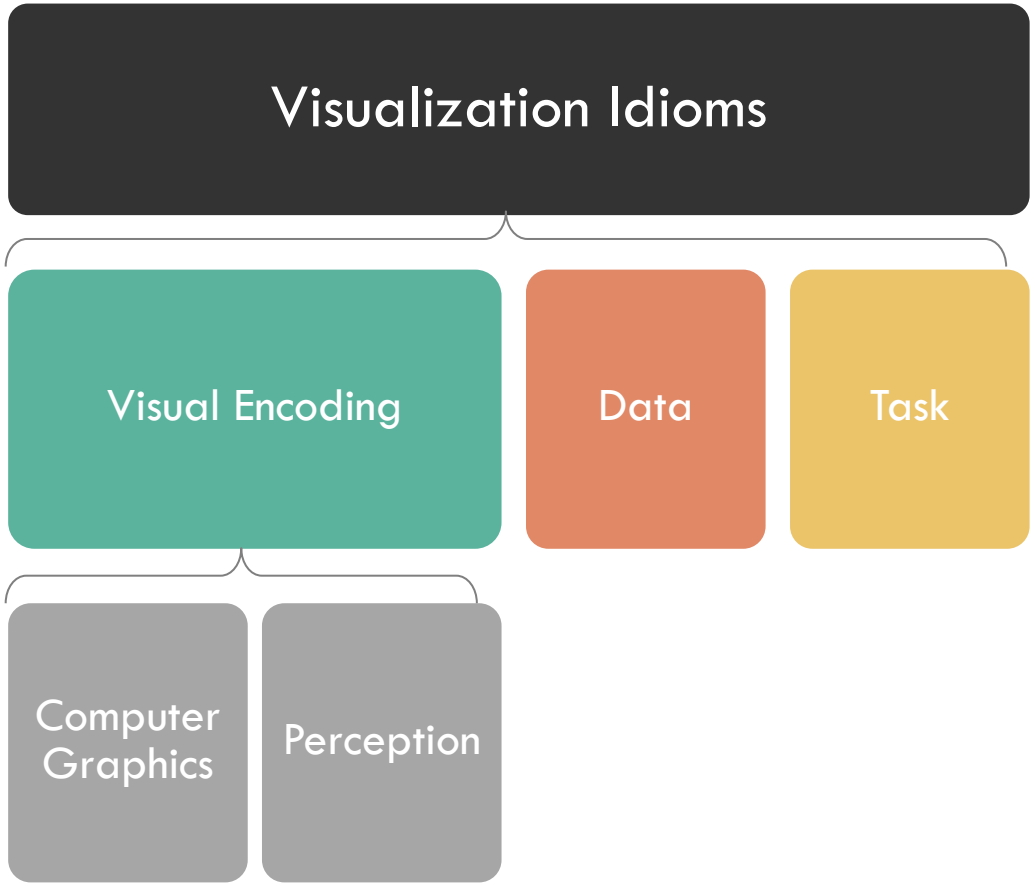
Boxplot

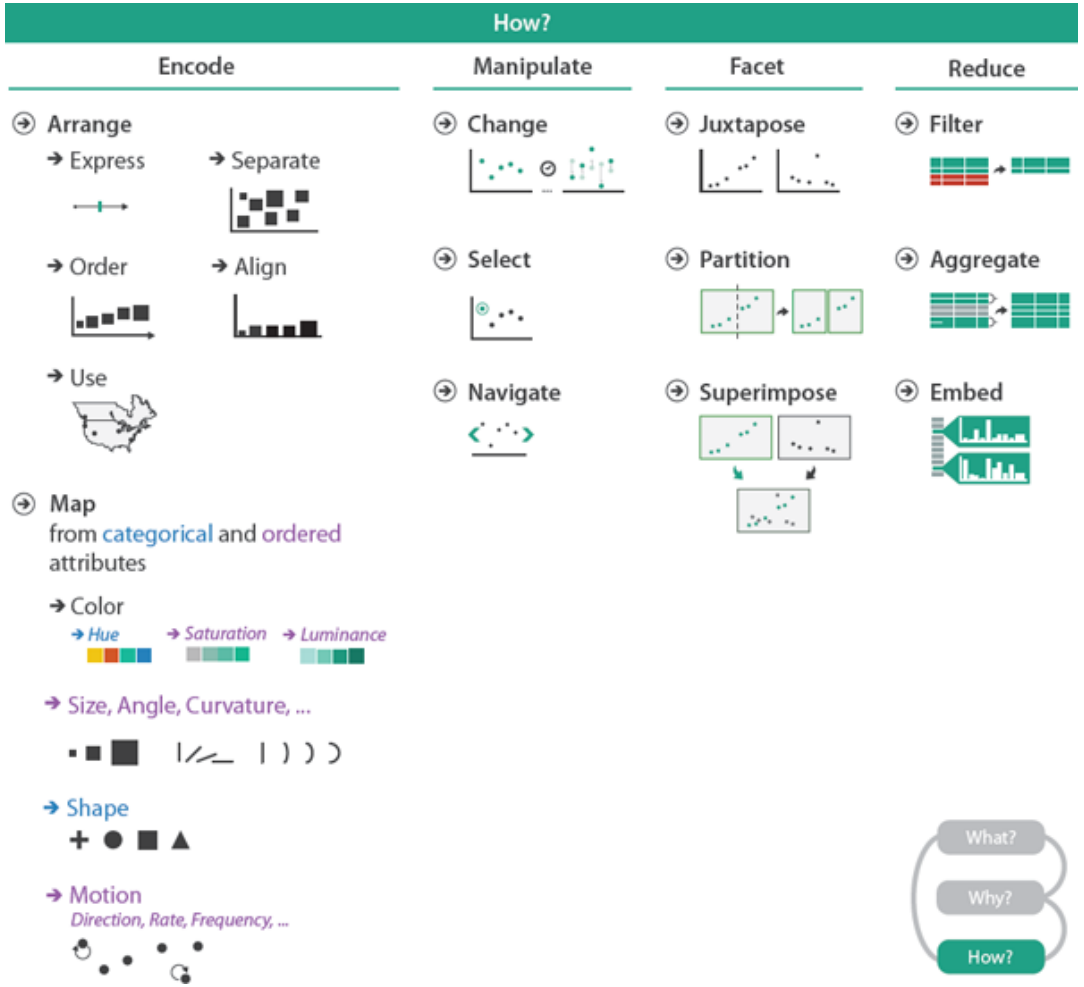
Summarise multiple distributions by showing the median (centre) and range of the data

Idiom	BAR AND LINE PLOT
Data: WHAT?	Table with one **continuous interval value** (e.g. rate) and **one continuous measurement** value (e.g. amount), with a key ordinal attribute (e.g. time/date)
Encoding: HOW?	Positional encoding on vertical direction of interval values connected by line marks, size encoding of measurement by height. Horizontal arrangement encodes ordinal key attribute.
Task: WHY?	**Discover/present correlations** between an interval and measurement values

Idiom	BOXPLOT
Data: WHAT?	Table with one one categorical attribute and one continuous quantitative attribute.
Encoding: HOW?	Encode categorical attribute by arrangement along one axis, express min, max, median and quartiles by position and size encoding along the other axis.
Task: WHY?	**Summarize distribution** of values by range and median.

Effective Visualization Idioms





Position based encoding: “Arranging” Data

Spatial encoding dominates mental model

- Position is the most effective encoding channel
- Non-spatial channels are good for some specific cases but not always effective across all attribute types

Visual Encoding using space. Two main uses:

Directly encode
quantitative/ordinal attributes

Express Values

Arrange visual elements and
encoding categorical attributes

Separate, Order, Align Regions

Separate

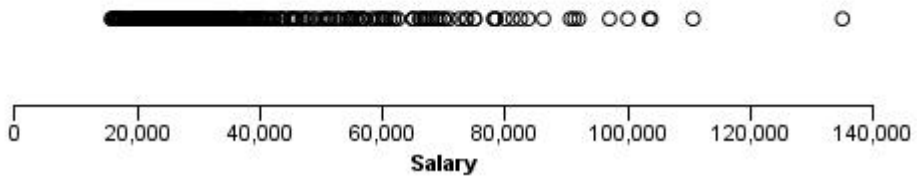
Order

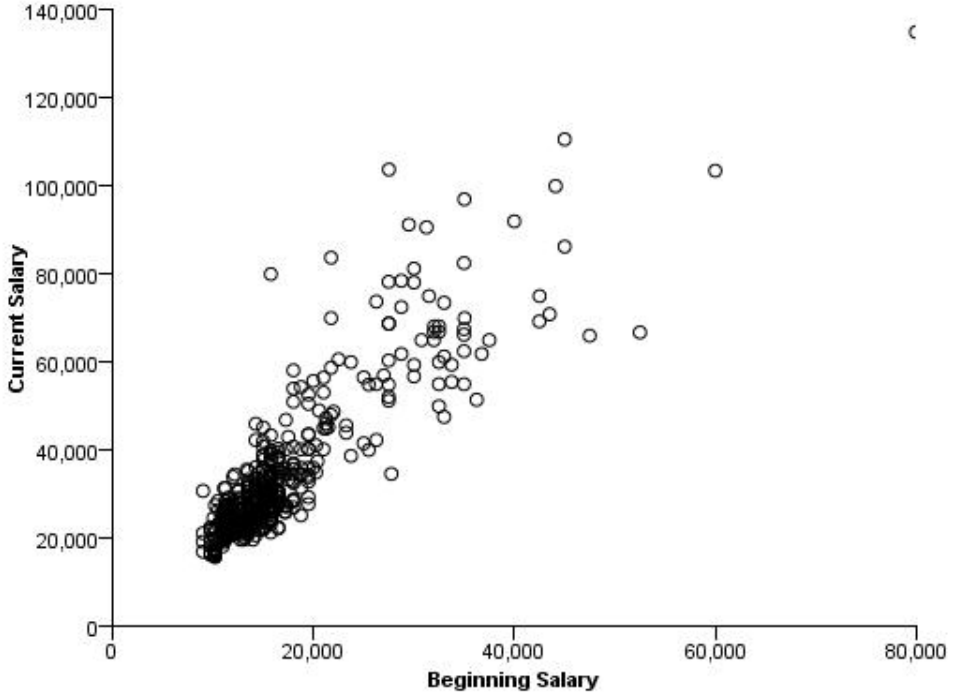
Align

Expressing Quantitative Data with Position

Spatial encoding is used to directly express the <u>value</u> of a quantitative attribute

- Value of item mapped to position along axis

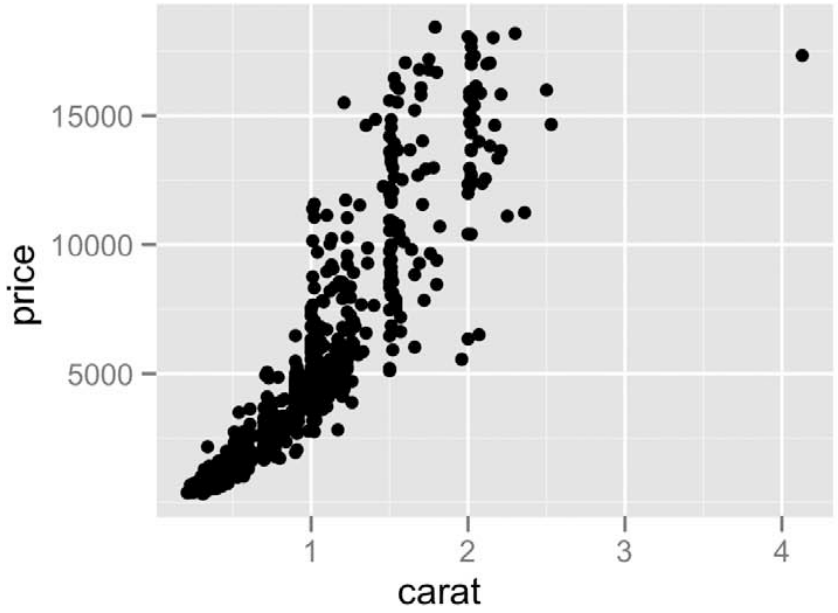


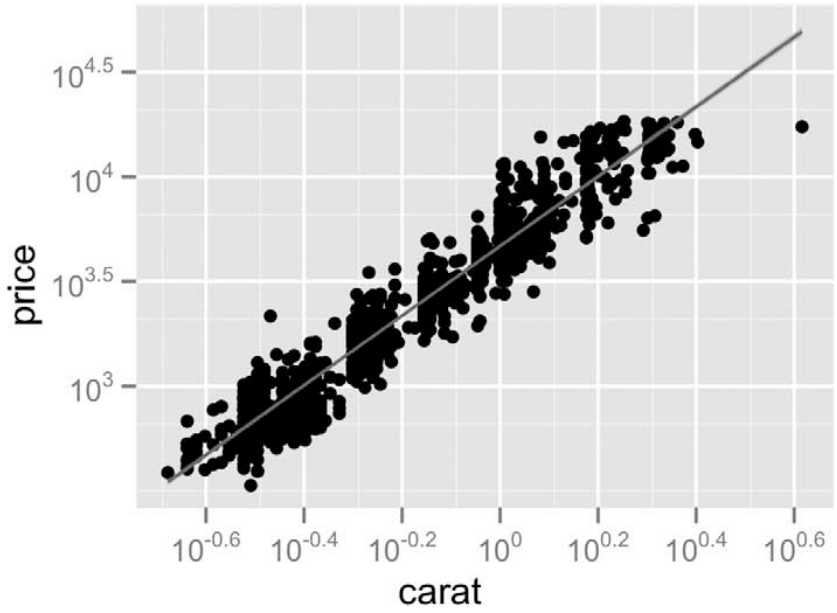


- Most commonly: a 2D scatterplot based on flat-table with 2 quantitative attributes mapped to 2 spatial dimensions

Images © IBM Knowledge Center

Scatterplots

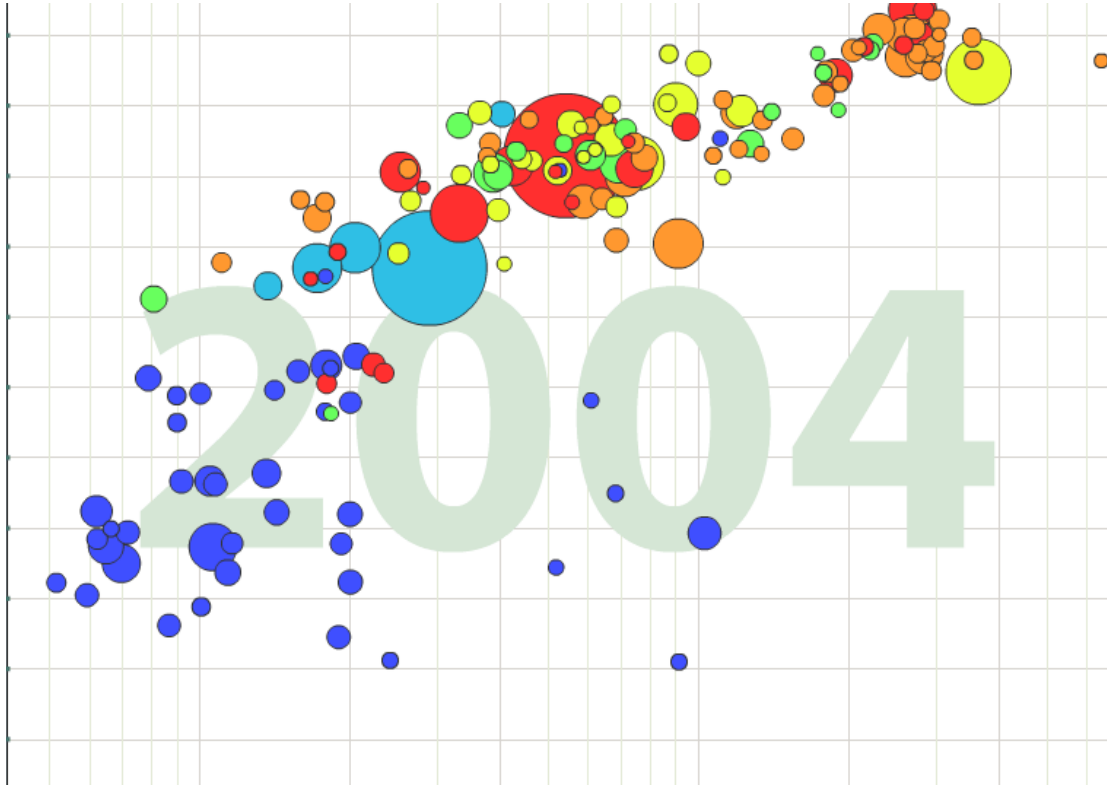




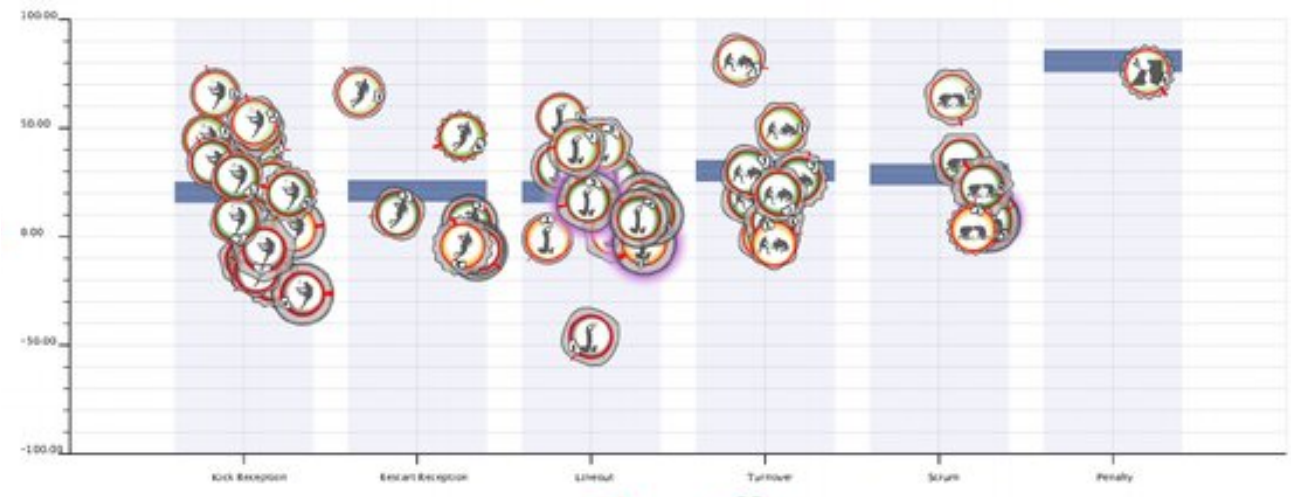
Idiom	SCATTERPLOT
Data: WHAT?	Table Dataset, 2 quantitative numerical attributes
Encoding: HOW?	Point marks. **Express** with horizontal and vertical position
Task: WHY?	Discover trends/outliers. Locate clusters. Explore distribution. Identify correlation.

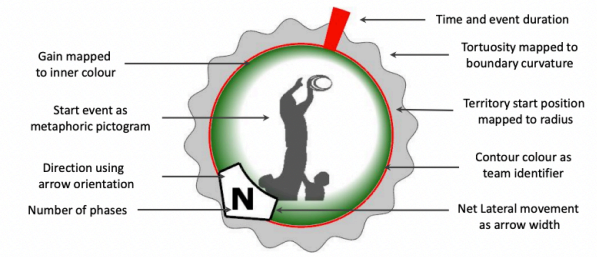
Interaction with other Channels

Encode value with spatial position. Then additional attributes may be encoded on the same mark with other channels or other spatial dimensions



Gapminder trendalyzer © Rosling 2003 URL:
https://www.gapminder.org/about-gapminder/

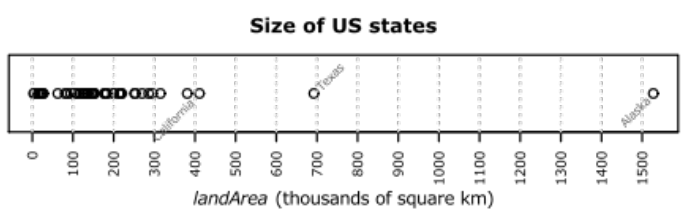


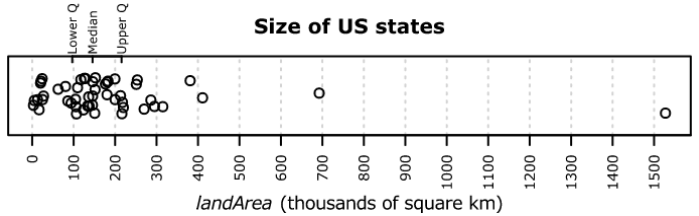


Visualization of a rugby match [Chung et al 2013]

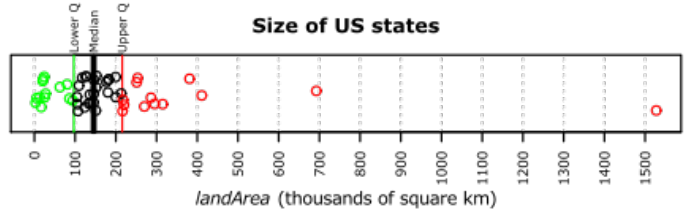
Scale / Over-plotting

Images © Carifio and Perla 2007

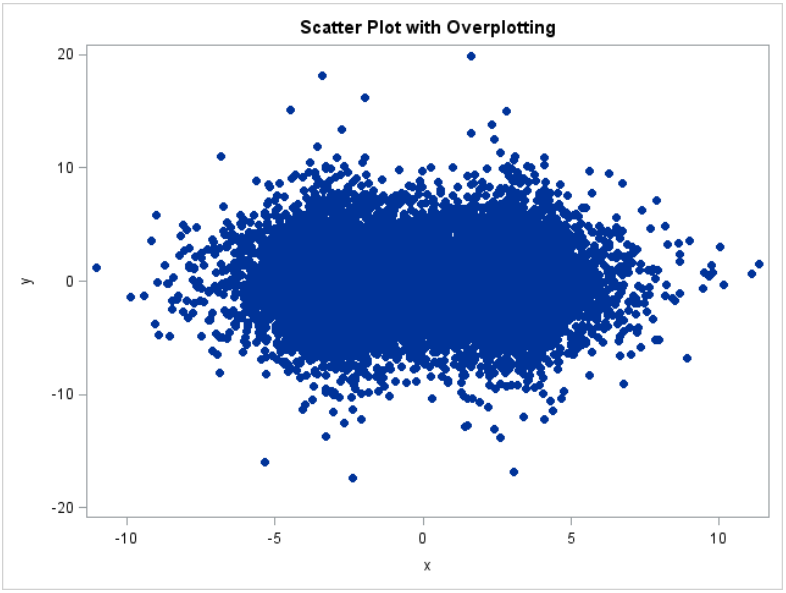


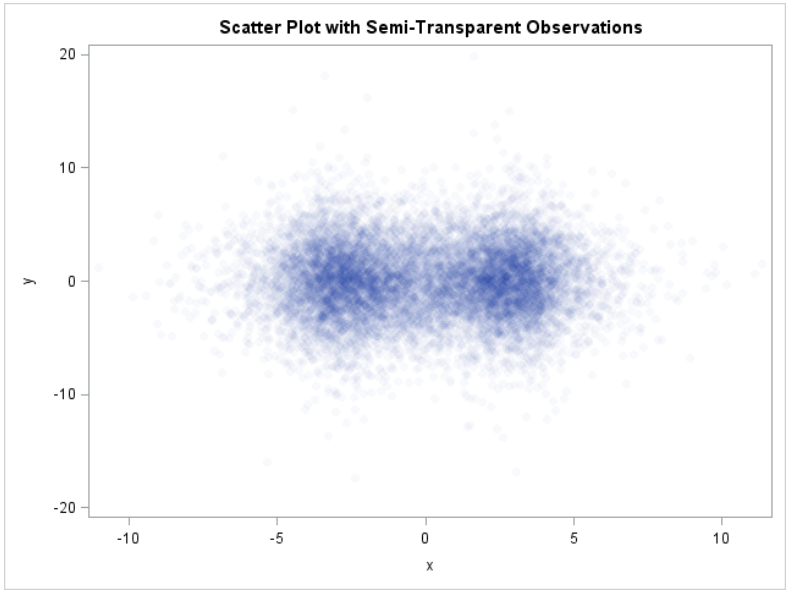


With jittering

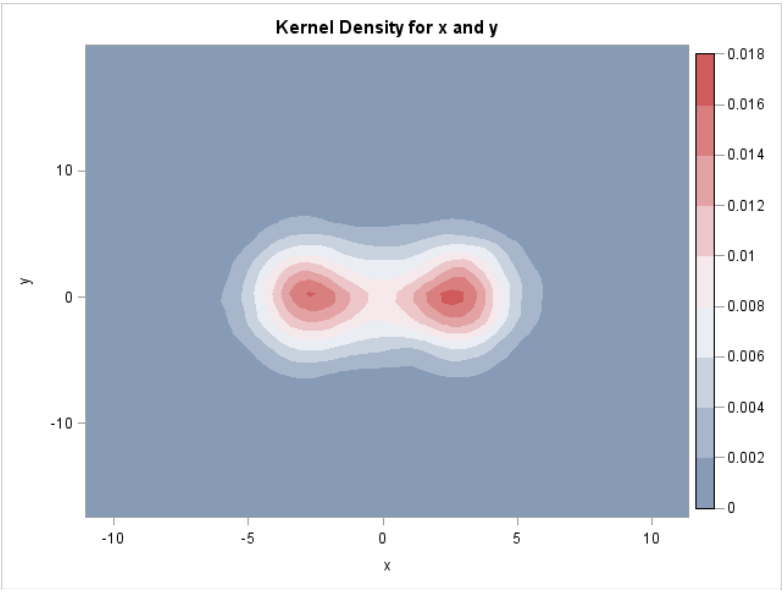


With quartiles mapped to colour





With translucency



With derived attribute: density, mapped to colour

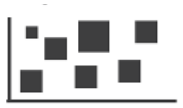
Dealing with overplotting (Wicklin 2011) URL: https://goo.gl/kWY21E

Arranging Categorical Data by Region

Not possible to directly encode Categorical data meaningfully with position

- Instead encode with spatial region
  - region = distinct bounded area/spatial partition
- Spatial proximity used to encode similarity

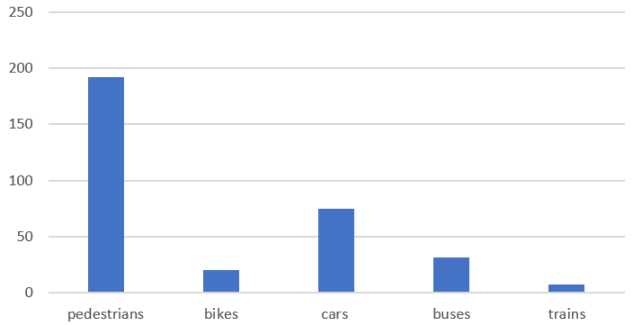
Three key spatial operations:

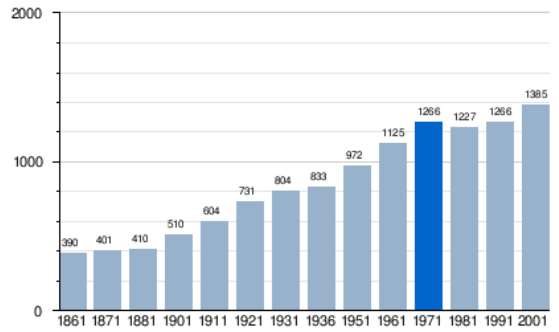


- **Separate** into regions: based on a categorical attribute

- **Order** the regions: based on an ordered attribute

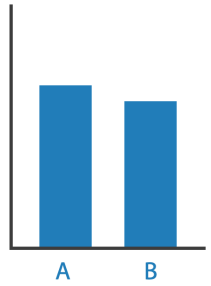
- **Align** regions: ordered relationships between values of regions





Data from ISTAT

Unaligned



Aligned

Arranging Data by Spatial Regions

Basic case Single key: One region per item - one-dimensional list arrangement

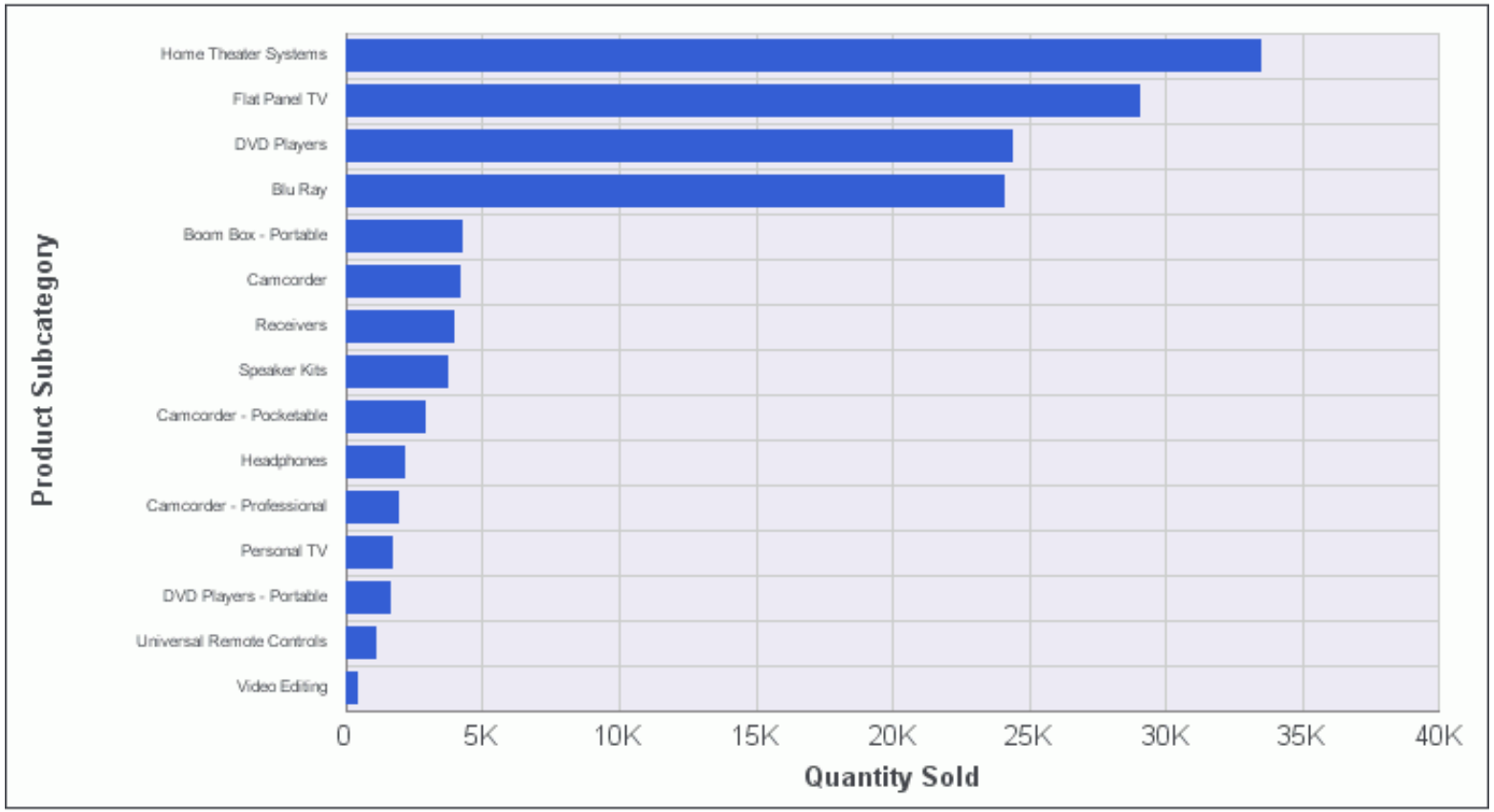
Apples	40
Bannanas	60
Grapes	30
Oranges	50
Pears	20

Depicted in a two-dimensional area

- Categorical regions **separated**, **aligned** and/or **ordered** in one dimension,
- Values **expressed** in the other dimension

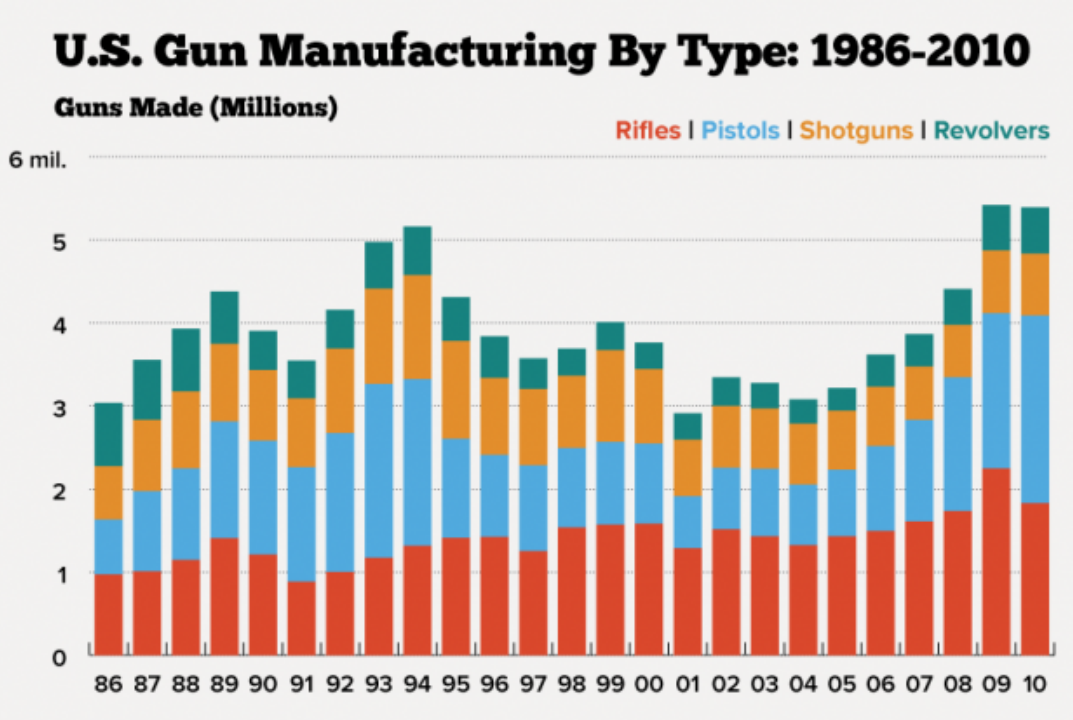
Bar charts

A	40
B	60
C	30
D	50
E	20



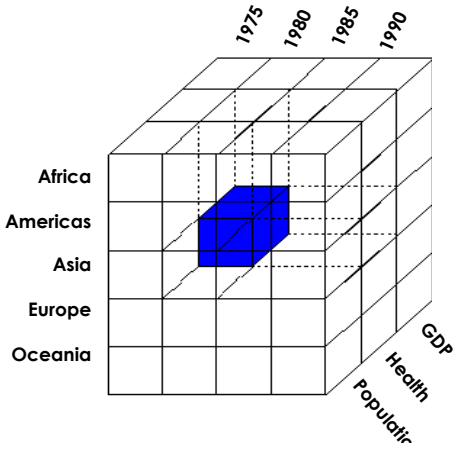
Idiom	BAR CHART
Data: WHAT?	Table: one numerical + one categorical attribute
Encoding: HOW?	**Express** value with aligned position/length in primary direction, **separate** key attribute with position in second direction
Task: WHY?	Lookup and compare values

Stacked Bar Chart



Idiom	STACKED BAR CHART
Data	Multi-dimensional table: one quantitative attribute, two categorical key attributes
Encoding	**Bar** glyph with length-encoded components of value attribute for each category of secondary key attribute. Bars are horizontally separated by category of primary attribute.
Task	Lookup and compare values, find trends, part-to-whole relationship

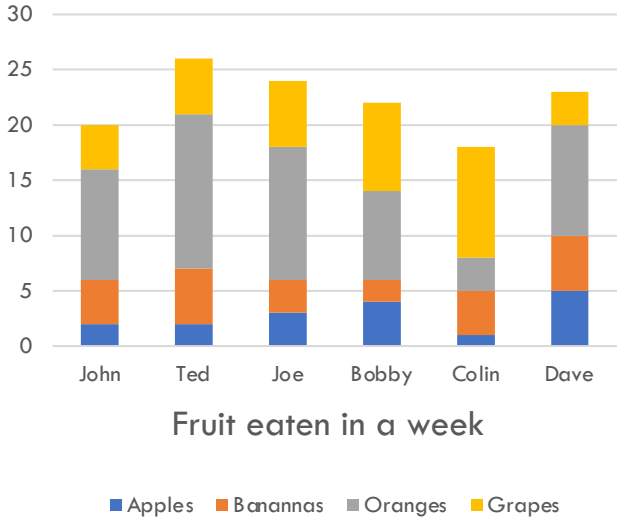
Aside: Multidimensional Table

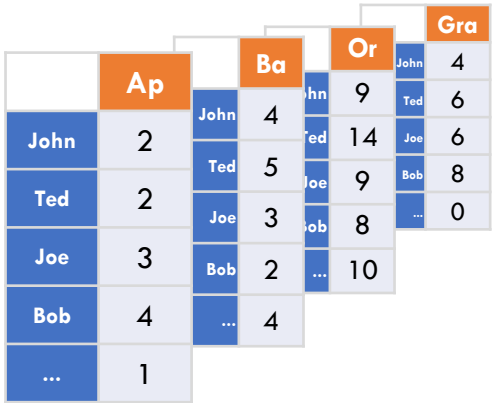


3D Multidimensional table: unique key for each dimension, used to look up a cell

	Apples	Bannanas	Oranges	Grapes
John	2	4	9	4
Ted	2	5	14	6
Joe	3	3	9	6
Bob	4	2	8	8
...	1	4	10	0

2 Dimensional Table: Can be treated as any other flat table.
This is essentially because we have only one attribute for one of the dimensions



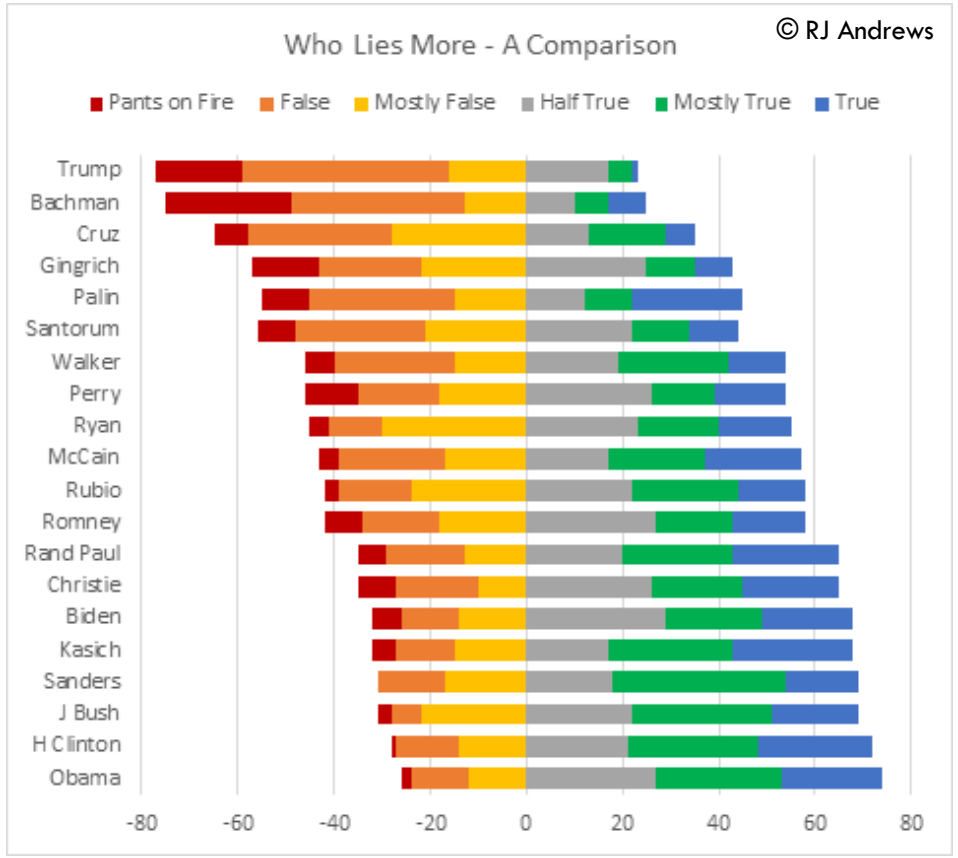


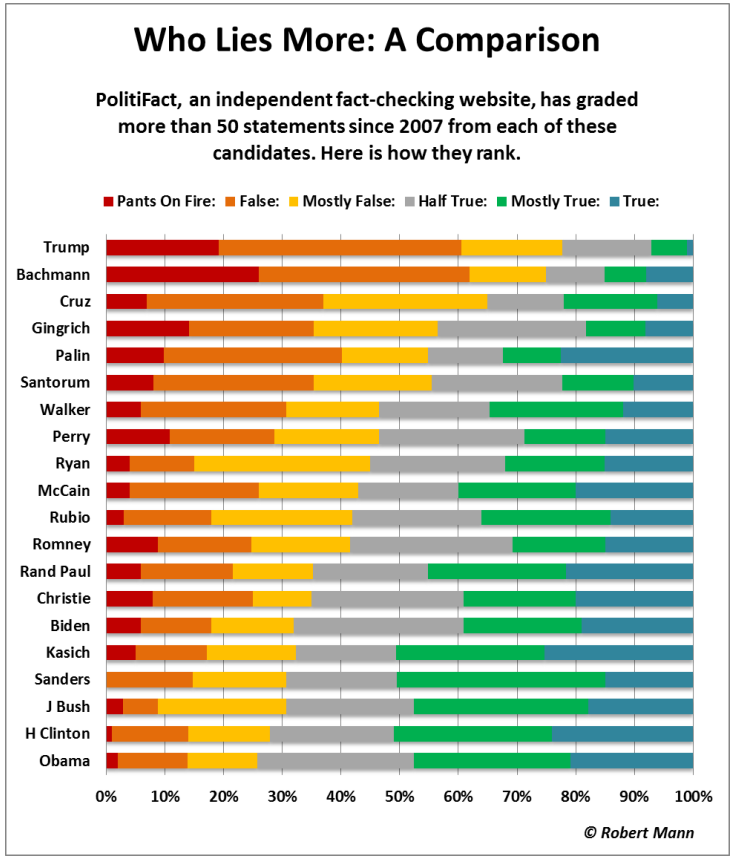
Alternate view

Applicant no	Degree Grade	Gender	Age
11001	64.5	M	23
11003	77.2	F	22
11050	91	M	33
33555	55.5	M	21
90210	80	F	19

Not all 2D tables can be treated in this way. In contrast, we don't consider the above multidimensional (as the type, units, scales of each attribute are different)

Stacked Bar Charts: Alternative Alignment

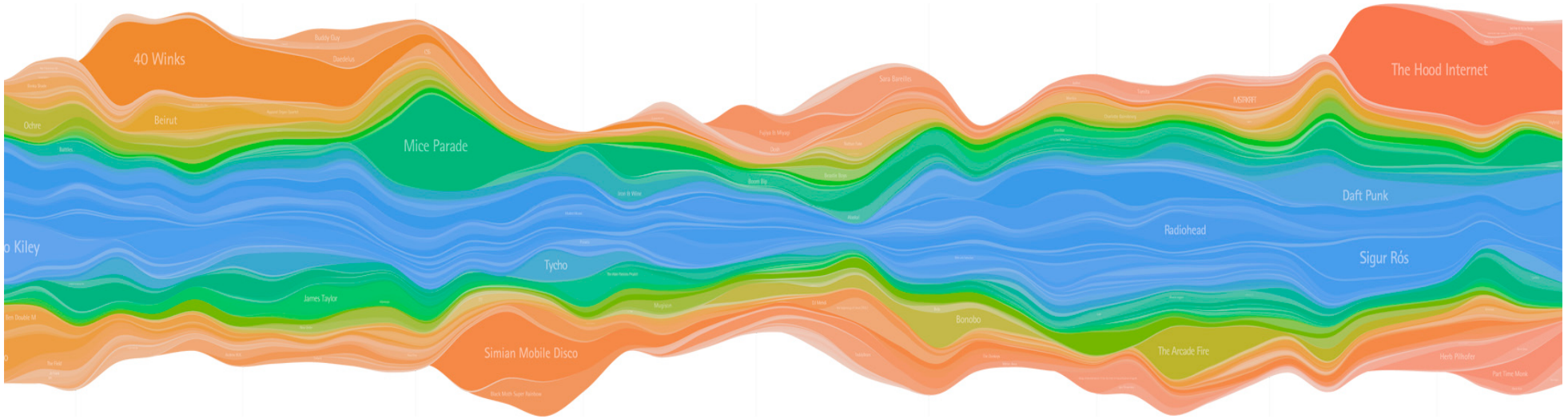




Idiom	NORMALIZED STACKED BAR CHART
Data	Multidimensional table: one quantitative value attribute, two categorical key attributes.
Derived data	One quantitative value attribute (normalized version of original attribute).
Encode	Bar marks with length channel encoding attribute value of each category of secondary key; rectilinear layout.
Task	Lookup an and compare part–whole relationship.

Steam Graphs

	1	2	3	4	5
A	18	96	50	42	44
B	8	6	12	14	4
C	56	58	74	98	38
E	20	32	36	60	22
F	44	38	64	90	24

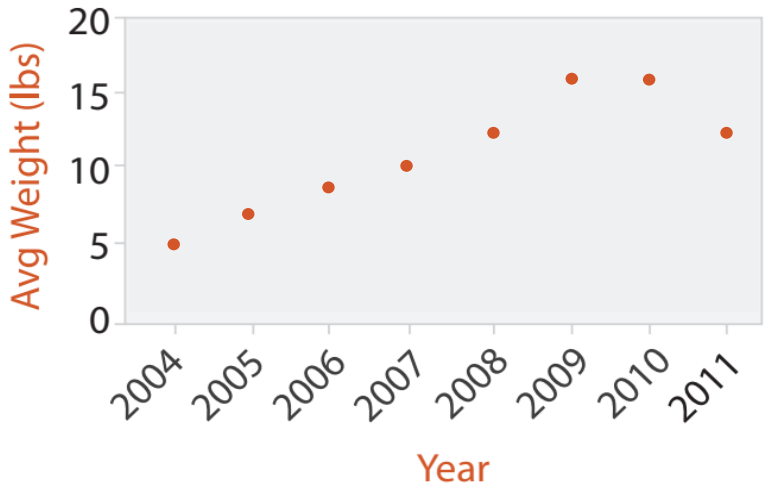


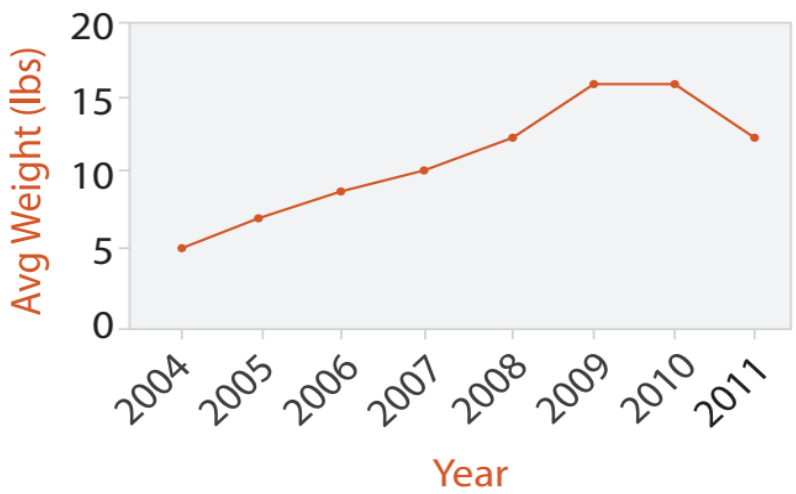
Visualization of Trends in Music Listening - by Byron and Wattenberg (2008)

Idiom	STEAM GRAPH
Data	Multi-dimensional Table: one quantitative value attribute (e.g. count), one ordered key attribute (e.g. time), one categorical key attribute (e.g. artist)
Encoding	Layer height encodes count, colour encodes categorical attribute and arranged along vertical axis, time is encoded along horizontal axis,
Task	Compare continuity of overall value/trend of component items

Dot and Line Charts

1	40.0
2	60.0
3	30.0
4	50.0
5	20.0

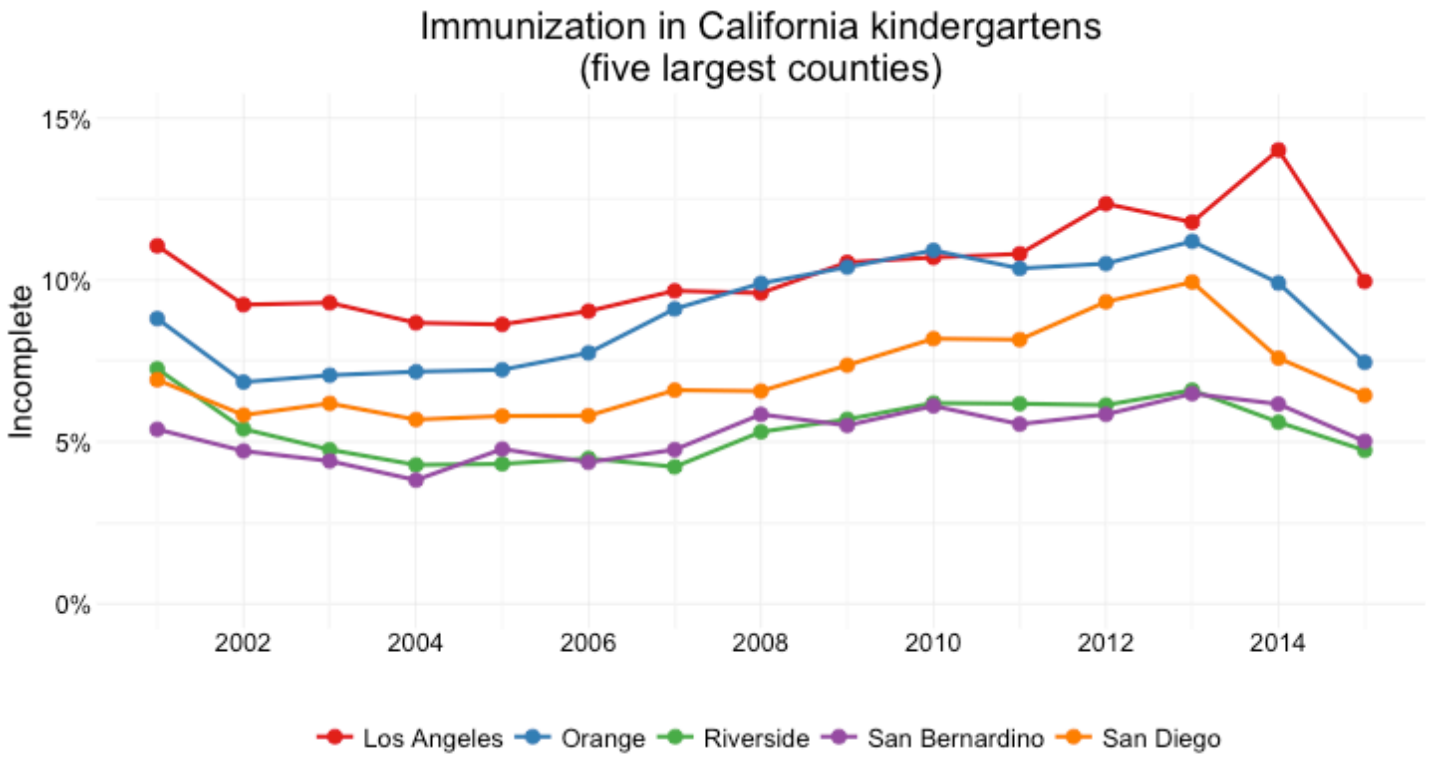




Idiom	DOT CHART
Data	Table: one quantitative attribute, one ordered key attribute
Encoding	Express value attribute with aligned vertical position and point marks. Separate/order into horizontal regions by key attribute
Task	Look up and compare values

Idiom	DOT AND LINE CHART
Data	Table: one quantitative attribute, one ordered key attribute
Encoding	As with dot chart with connection marks in between
Task	Show trend

Multi-Line Charts



	I	II	III	IV	V
A	6.9	4.8	6	10.2	7.8
B	2.7	9.3	7.8	9	7.8
C	8.1	9.3	6.9	5.4	1.8
E	5.1	0.9	3.9	9.6	0.6
F	3	4.2	1.2	9.6	8.1

	MULTI-LINE CHART
Data	Multi-dimensional Table: one quantitative value attribute, one ordered key attribute, one categorical key attribute
Encoding	Encode ordered key attribute by horizontal arrangement. Encode categorical attribute by colour. Express quantitative attribute with aligned vertical position and point marks connected by lines.
Task	Look up and compare trends and values

Aside: **Bar vs Line (rule of thumb)**

Based on type of key attribute

- bar charts if categorical key attribute
- line charts if ordered key attribute

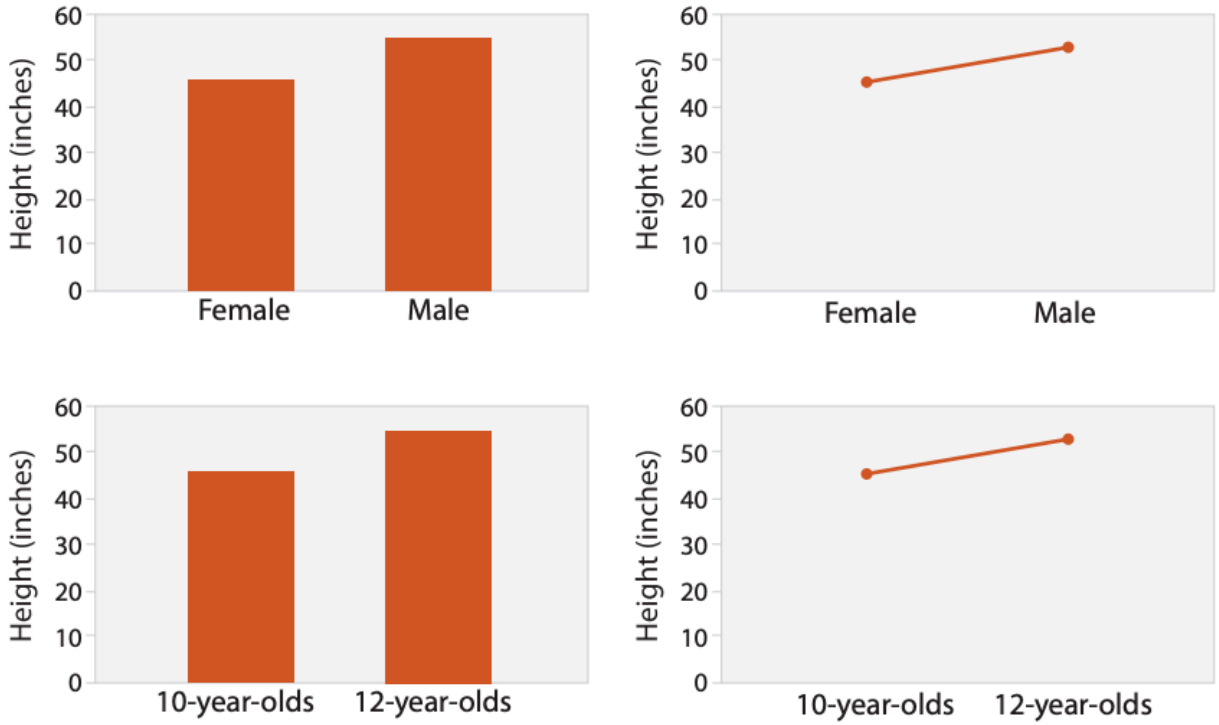
Do not use line charts for categorical key attributes. Implication of trend so strong that it overrides semantics. e.g. “The more male a person is, the taller he/she is”

Based on type of quantitative value attribute

- bar chart not ideal for intervals e.g. longitude, latitude

Based on task

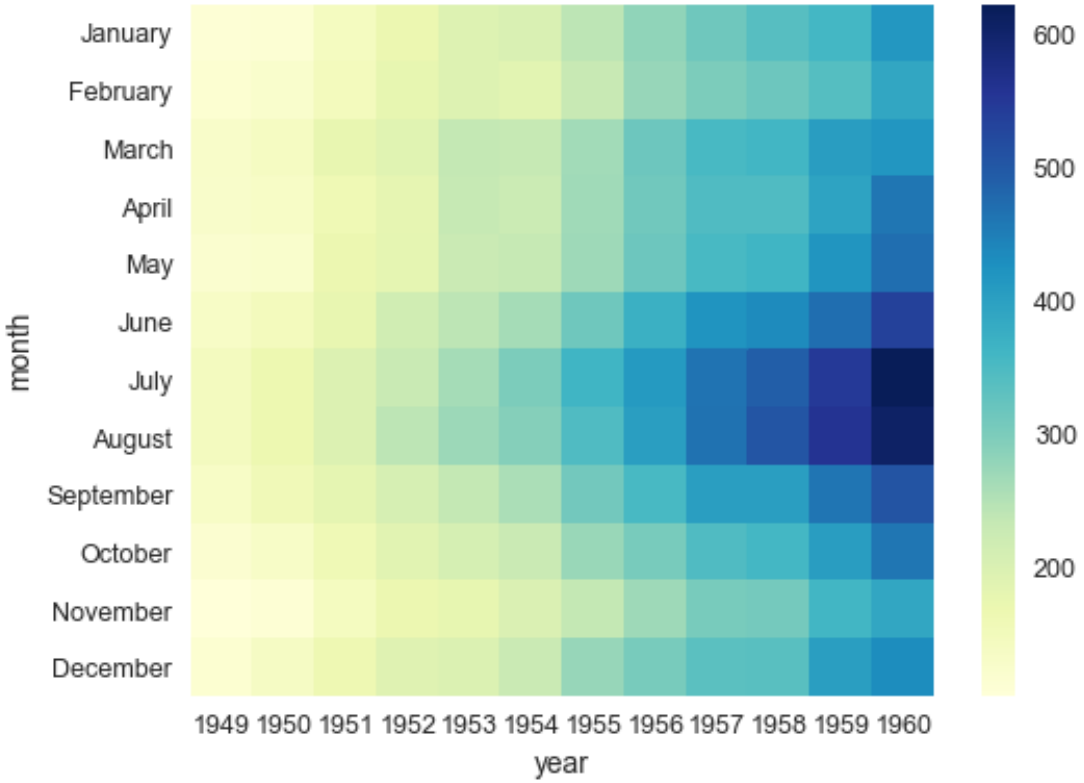
- Continuous line encoding better, for analysing trends or features



from Munzner(2014)

Matrix Arrangement

	I	II	III	IV	V
i	6.9	4.8	6	10.2	7.8
ii	2.7	9.3	7.8	9	7.8
iii	8.1	9.3	6.9	5.4	1.8
iv	5.1	0.9	3.9	9.6	0.6
v	3	4.2	1.2	9.6	8.1

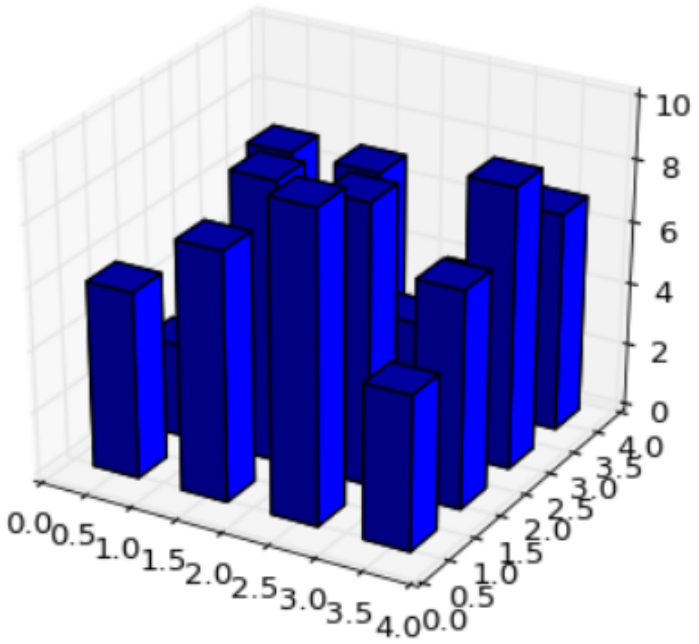


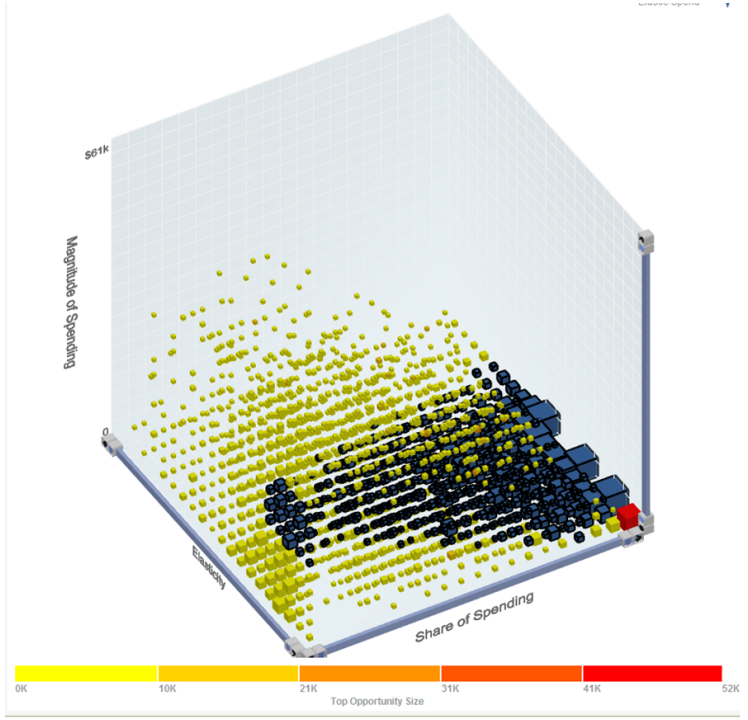
Idiom	HEATMAP
Data	Table: two categorical key attributes (e.g. Month, Year), one quantitative attribute
Encoding	2D matrix alignment of rectangular marks arranged on horizontal and vertical axes based on categorical key attributes. Diverging color-map encodes the quantitative attribute.
Task	Locate/Browse/Explore or summarize clusters, outliers, trends

Higher Dimensions of Arrangement

In case of multiple ordered/numerical keys, 2 spatial dimensions is sometimes not sufficient.
Why not use 3D space/position?

	I	II	III	IV	V
i	6.9	4.8	6	10.2	7.8
ii	2.7	9.3	7.8	9	7.8
iii	8.1	9.3	6.9	5.4	1.8
iv	5.1	0.9	3.9	9.6	0.6
v	3	4.2	1.2	9.6	8.1

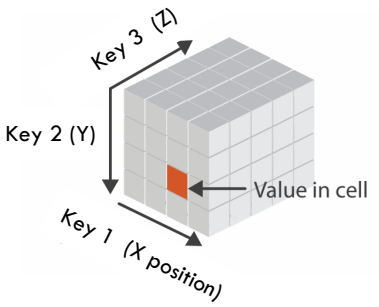


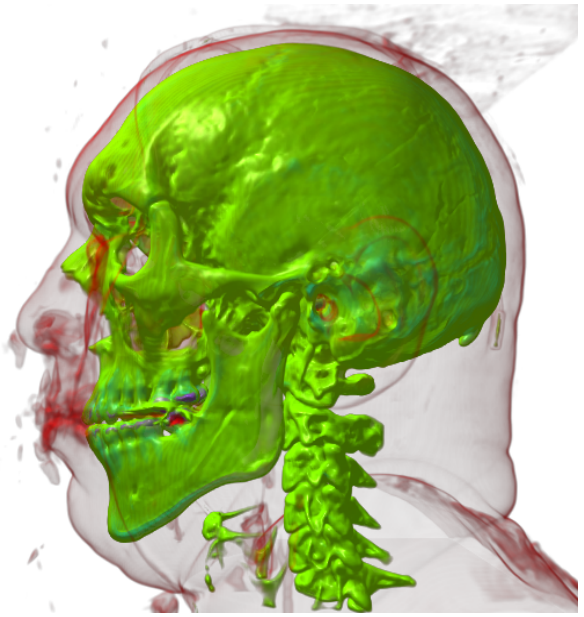


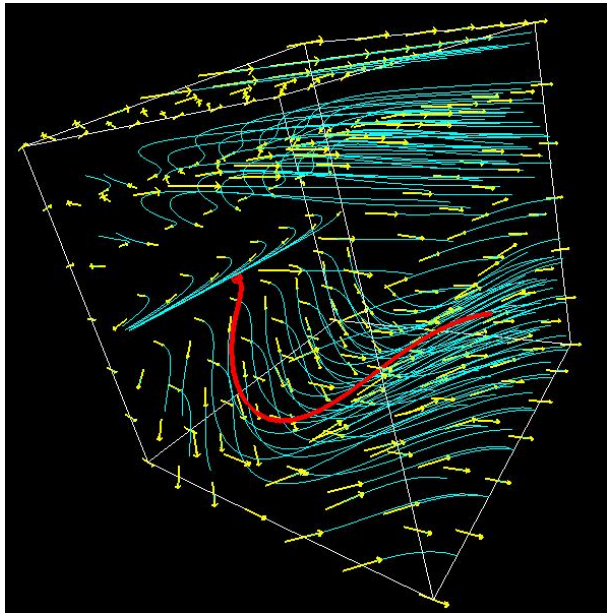
3D arrangement is possible but typically not recommended for non-spatial data due to depth ambiguity.

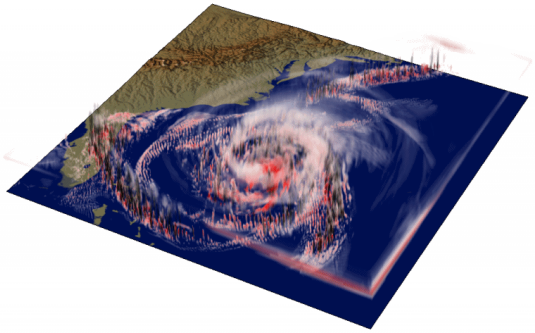
Higher Dimensions of Arrangement

Some exceptions when the data is spatial and 3D or higher dimensional. Discussed in more detail later









Idiom	Volumetric Rendering
Data	3D Grid with one quantitative atrribute per grid position (aka Scalar Field)
Encoding	Essentially, data points arranged by key attributes of x, y, z, position. Colour (+ opacity) denotes quantitative attribute. This is flattened to 2D by projection
Task	Search and query for features, outliers, shape

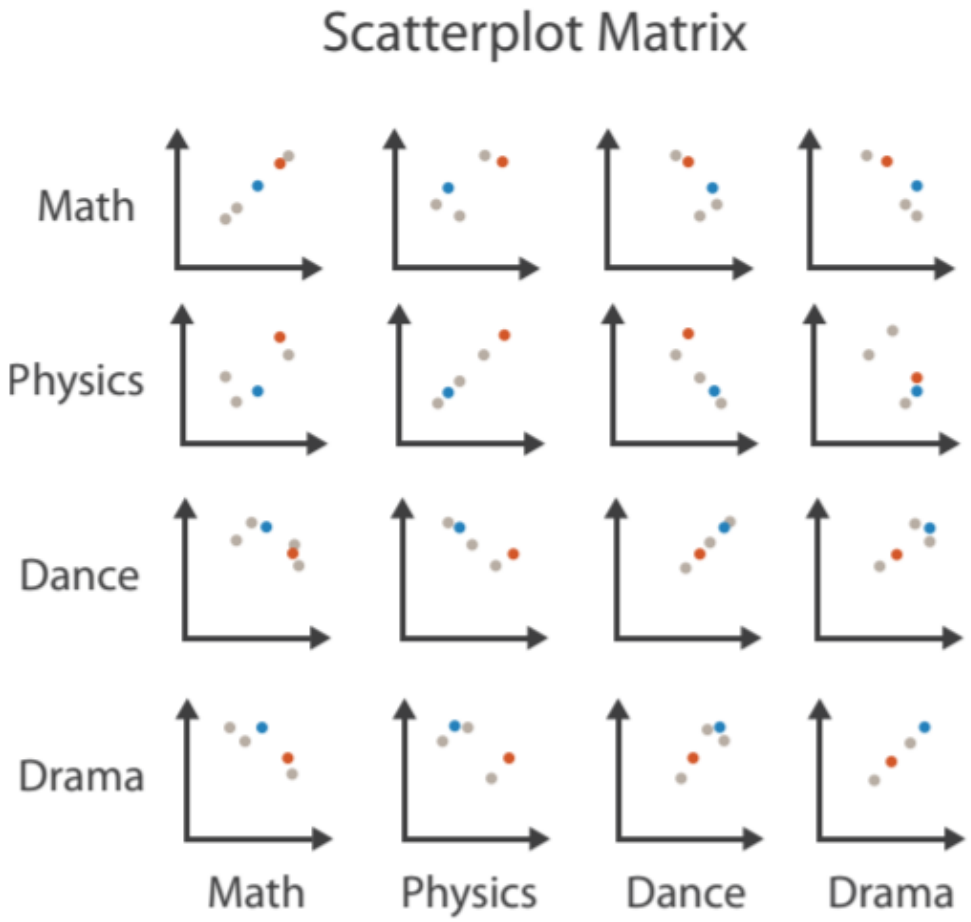
Idiom	3D Discretised Vector Field Visualization
Data	3D Grid with 3D vector at each position
Encoding	Angle encoding denotes vector direction, length and colour denote magnitudes of vector.
Task	Search and query for feature, outliers, shape, trends

Idiom	Animated Scalar Field
Data	4D data with one value in each 3D position in timestamp
Encoding	Data points arranged by key attributes of x, y, z position and across animation frames by timestamp. Colour (+ opacity) denotes quantitative attribute. This is flattened to 2D by projection
Task	Search and query for feature, outliers, shape, trends

Scatterplot Matrix

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

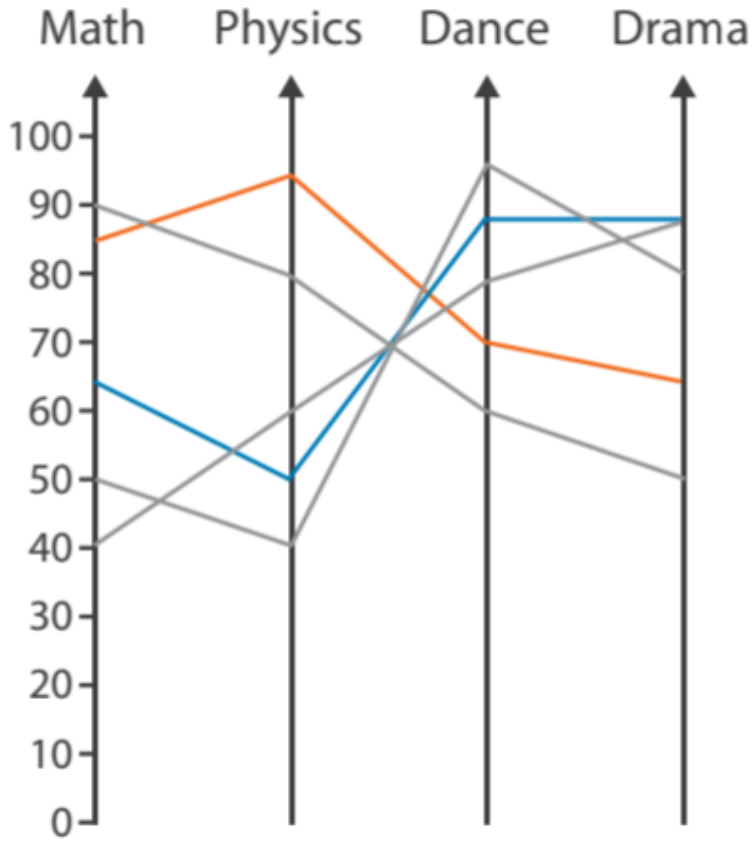
Idiom	Scatterplot Matrix (SPLOM)
Data	Multidimensional Table
Encoding	Scatterplots in 2D matrix alignment.
Task	Discover correlations between attributes. Identify and look up trends/outliers



Parallel Coordinates

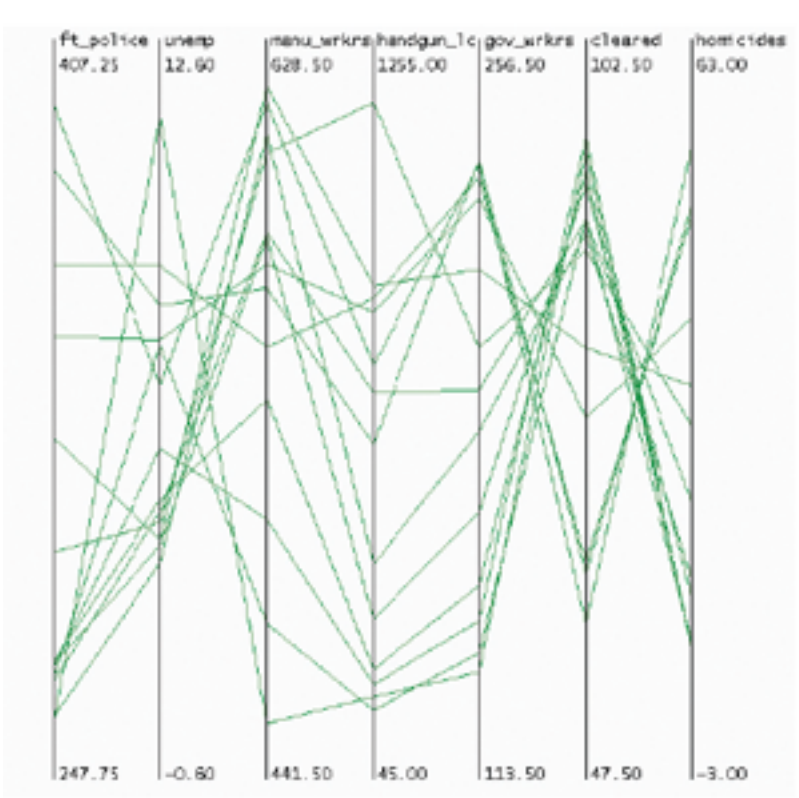
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Idiom	PARALLEL COORDINATES
Data	Table: many value attributes.
Encoding	Parallel layout: horizontal spatial position used to separate axes, vertical spatial position used to express value along each aligned axis with connection line marks as segments between them.
Task	Find trends, outliers, extremes, correlation.

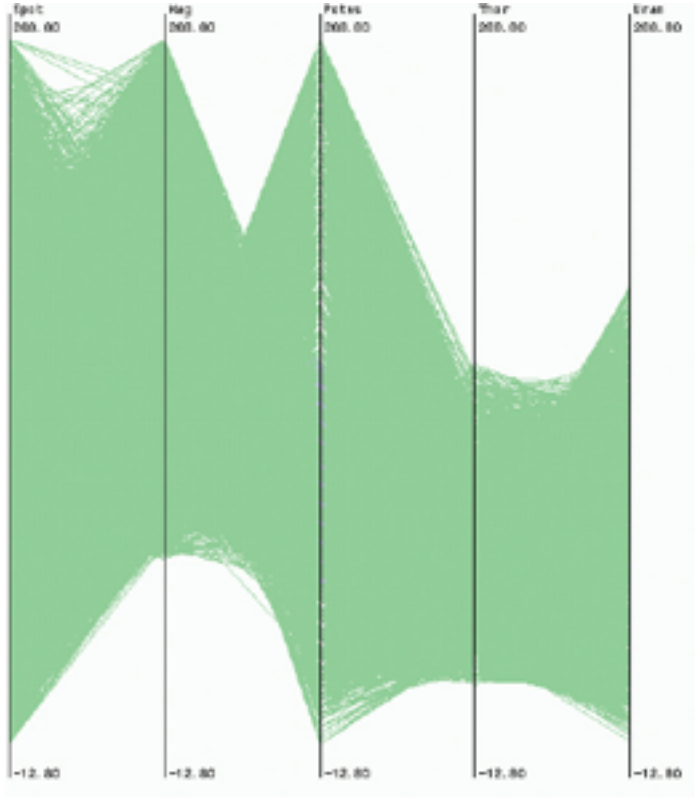


N.B. this goes slightly against what we said about superfluous use of dot and line charts. But here the line depicts not a continuous interpolatable value but a link between data points

Scalability



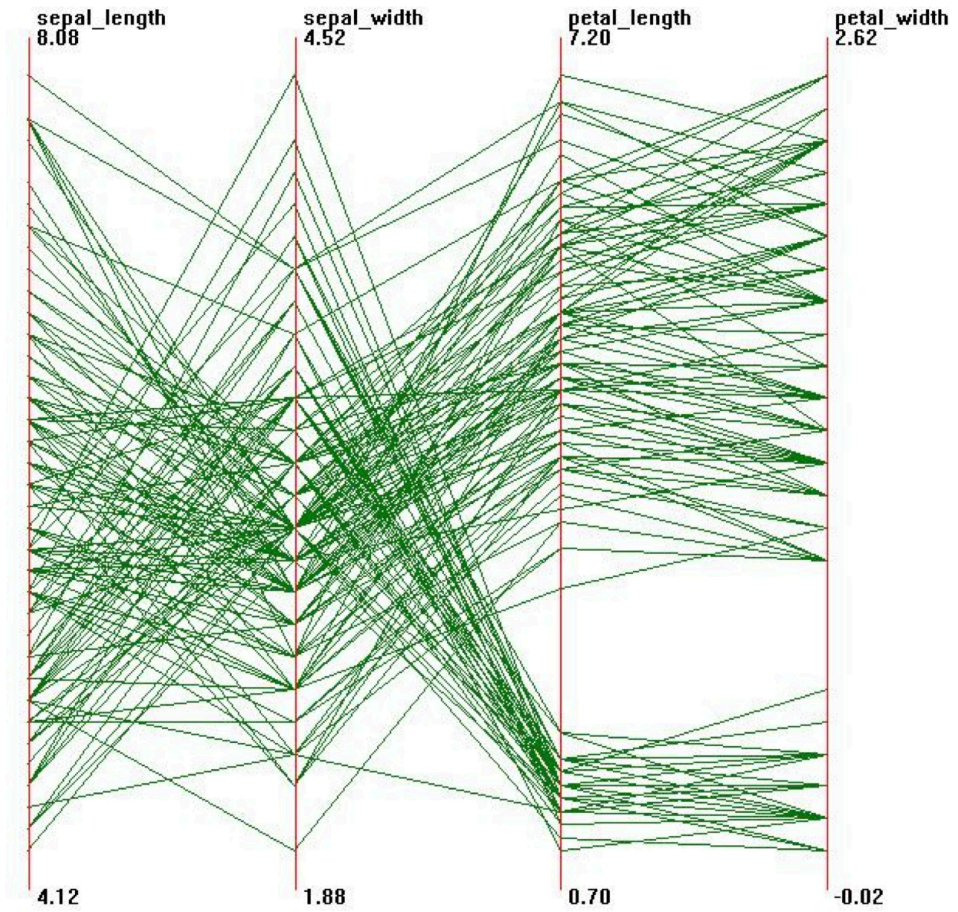
(a)

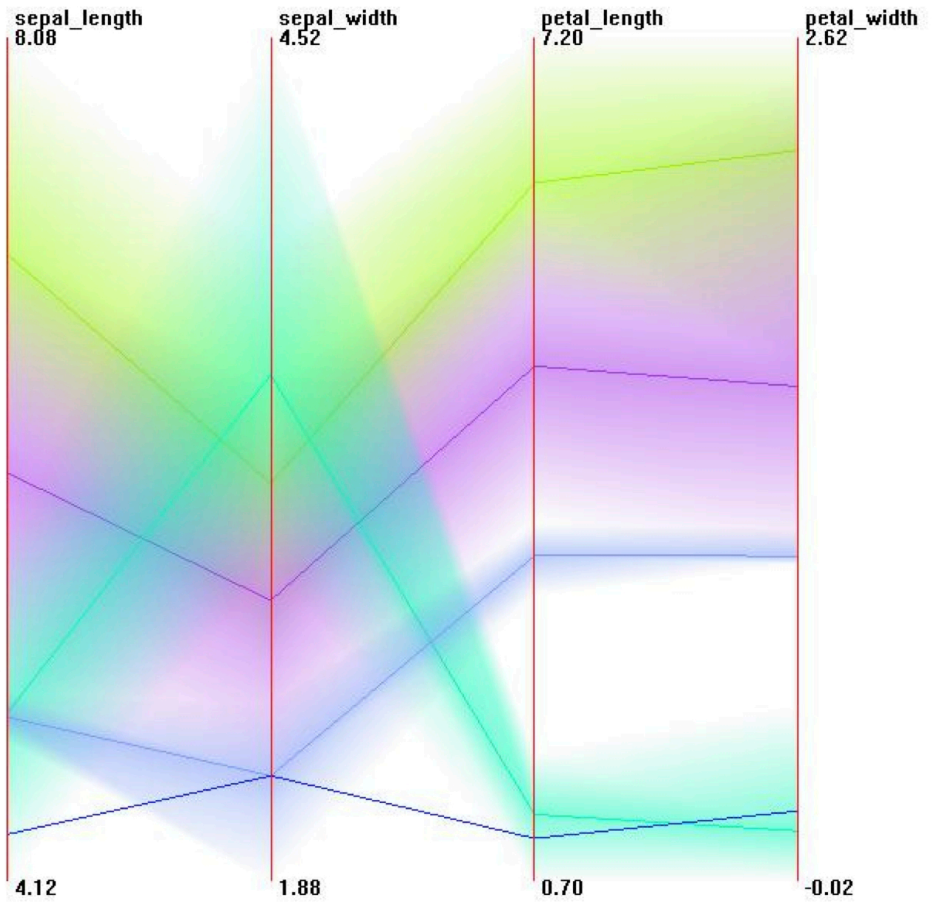


(b)

Suffers from overplotting when larger than ~12 dimensions

Reduction

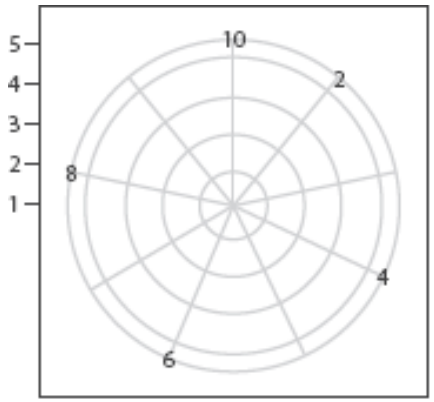


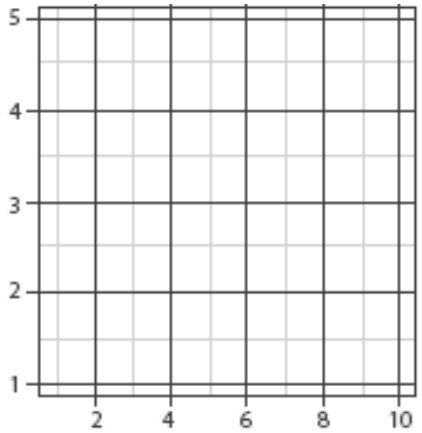


Ideally suited for tasks involving analysis of trends, outliers, extremes, correlation. Thus some visual abstraction techniques can address overplotting problems and enhance visualization for such tasks (reduction is discussed in more detail later in the module).

Radial Layout

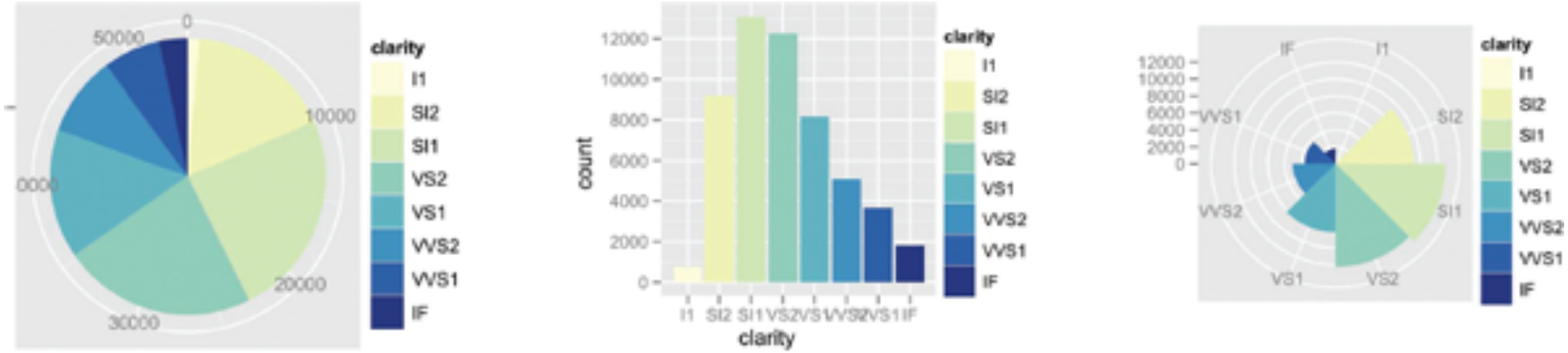
Items distributed around a circle using angle channel (angular position for arrangement. And separation) and one or more linear spatial channels (e.g. radial distance)





- Not as effective as linear position. Humans don't perceive angles as accurately as position
- Inherently cyclic
- Orientation can also make comparison of the linear dimensions more difficult

Pie Charts (Angle and area encoding)

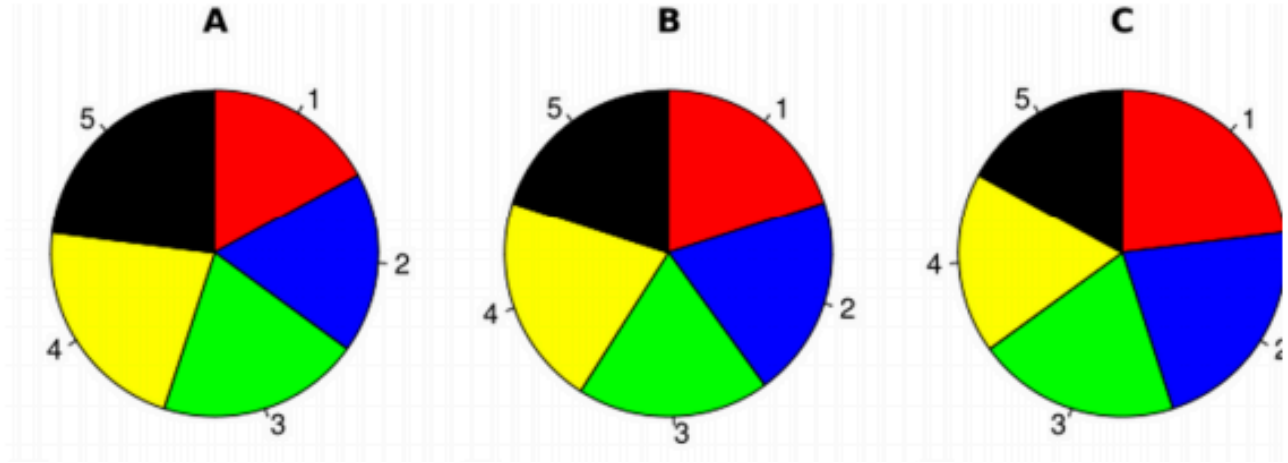


Three visualizations of the clarity distribution of diamonds, where I1is worst and IF is best.(Wickham 2010)

	PIE CHARTS
Data	Table: one quantitative attribute, one categorical attribute.
Encoding	Area marks (wedges) with angle channel; radial layout.
Task	Part–whole relationship.

	POLAR AREA CHARTS
Data	Table: one quantitative attribute, one categorical attribute.
Encoding	Area marks (wedges) with length channel; radial layout.
Task	Part–whole relationship.

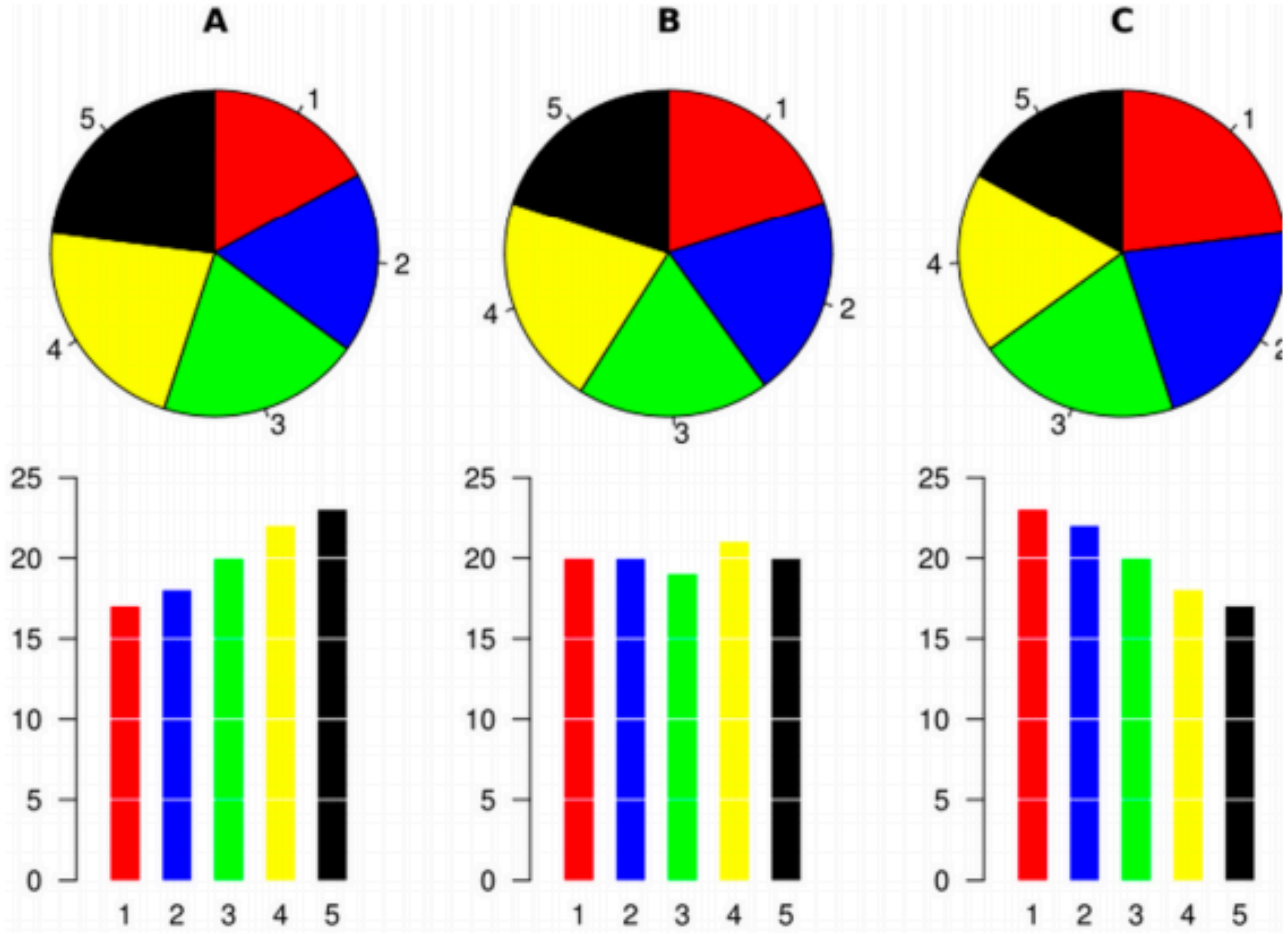
Aside: **Pie Charts**



Can you spot differences? (Alexander Lex 2014)

Aside: **Pie Charts**

This example by Lex (2014) shows disadvantages of pie charts (comparing areas subtended by angles is not as effective as one-dimensional length judgements)



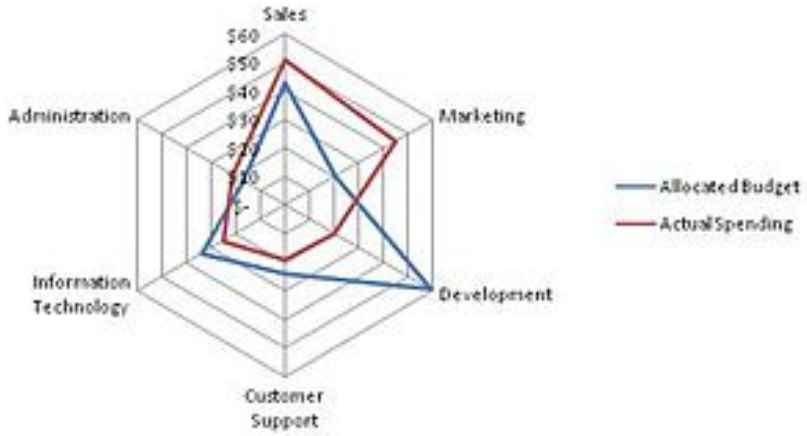
StarPlot

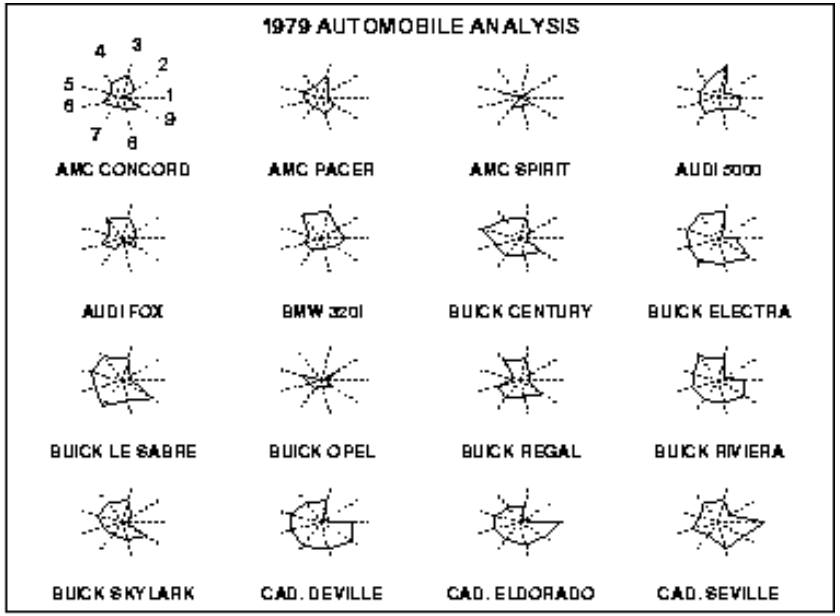
a.k.a. Radar Chart, Spiderweb Diagram

Data/task: multi-variate data, compare values, derive (overall value, similarties)

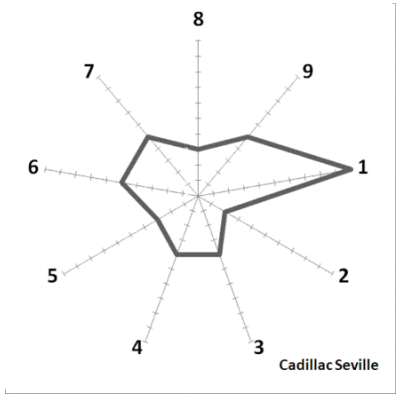
Encoding:

- Radially distributed axes for each quantitative attribute
- Radial line plot connecting each data point





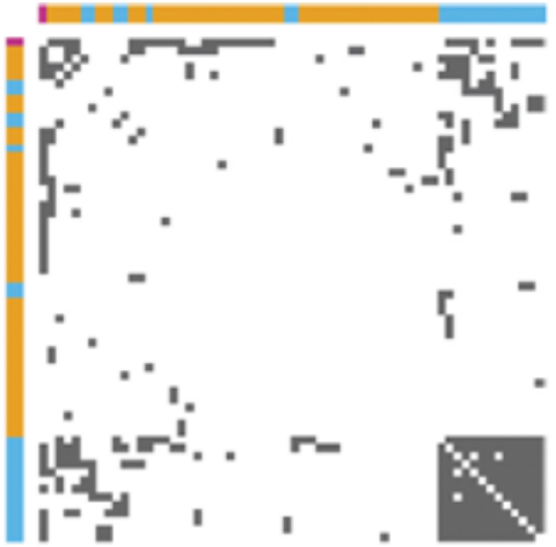
1. Price
2. Mileage (MPG)
3. 1978 Repair Record (1 = Worst, 5 = Best)
4. 1977 Repair Record (1 = Worst, 5 = Best)
5. Headroom
6. Rear Seat Room
7. Trunk Space
8. Weight
9. Length



© NIST – Information Technical Library
https://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm

In Brief: **Network and Tree arrangement –**

Spatial arrangement is an important factor in effective network visualizations (these will be discussed in detail later)



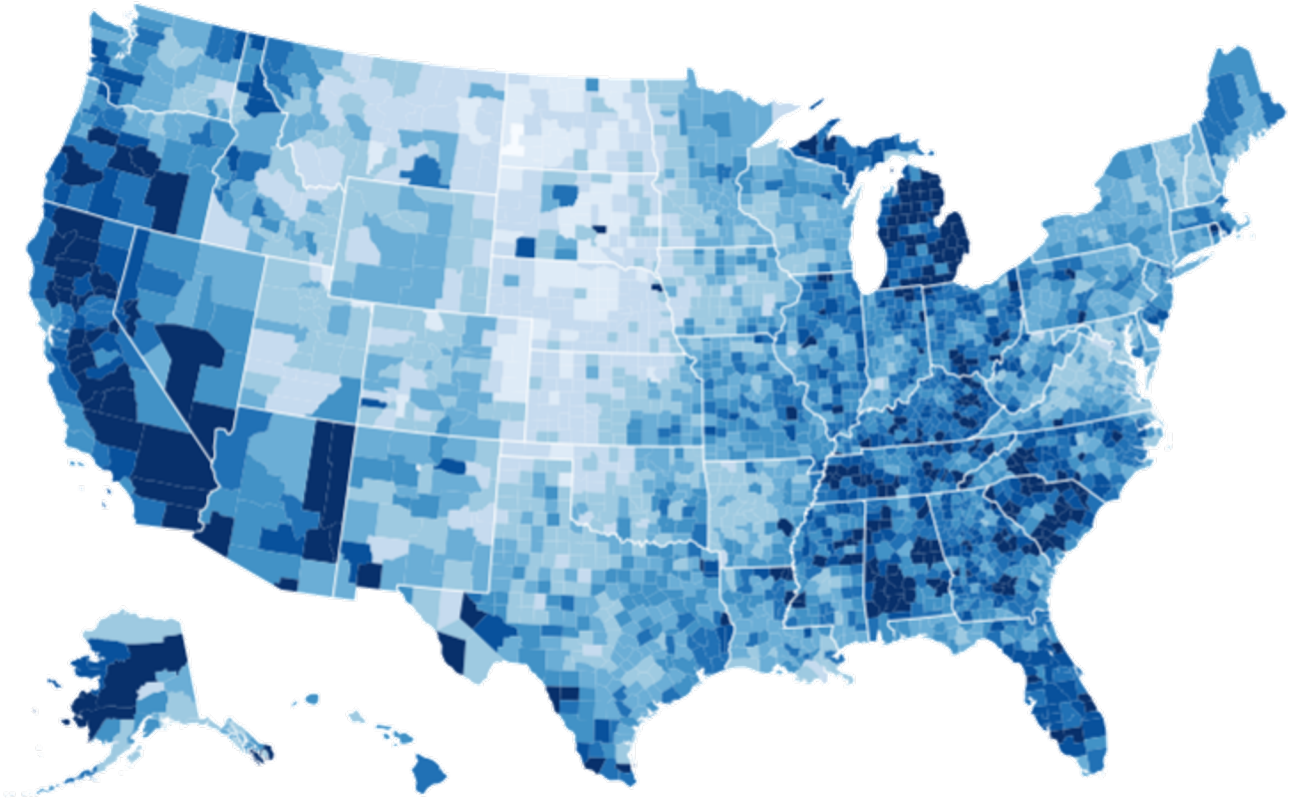
	NODE-LINK DIAGRAM
Data	Network.
Encoding	Point marks for nodes, connecting line marks for links.
Tasks	Explore topology, locate paths.

	Adjacency Matrix
Data	Network.
Derived	Table: network nodes as keys, link status between two nodes as values (binary).
Encode	Area marks in 2D matrix alignment.

In Brief: **Spatialized data arrangement**

Spatial arrangement is a key element of spatial data visualizations (these will be discussed in detail later)

	CHOROPLETH MAP
Data	Geographic geometry data with region ID. Table with one quantitative attribute per region ID.
Encoding	Space: use **given** geometry for area mark boundaries. Color: sequential segmented colormap.
Tasks	Analyze spatially distributed values, trends; comparison, search, query



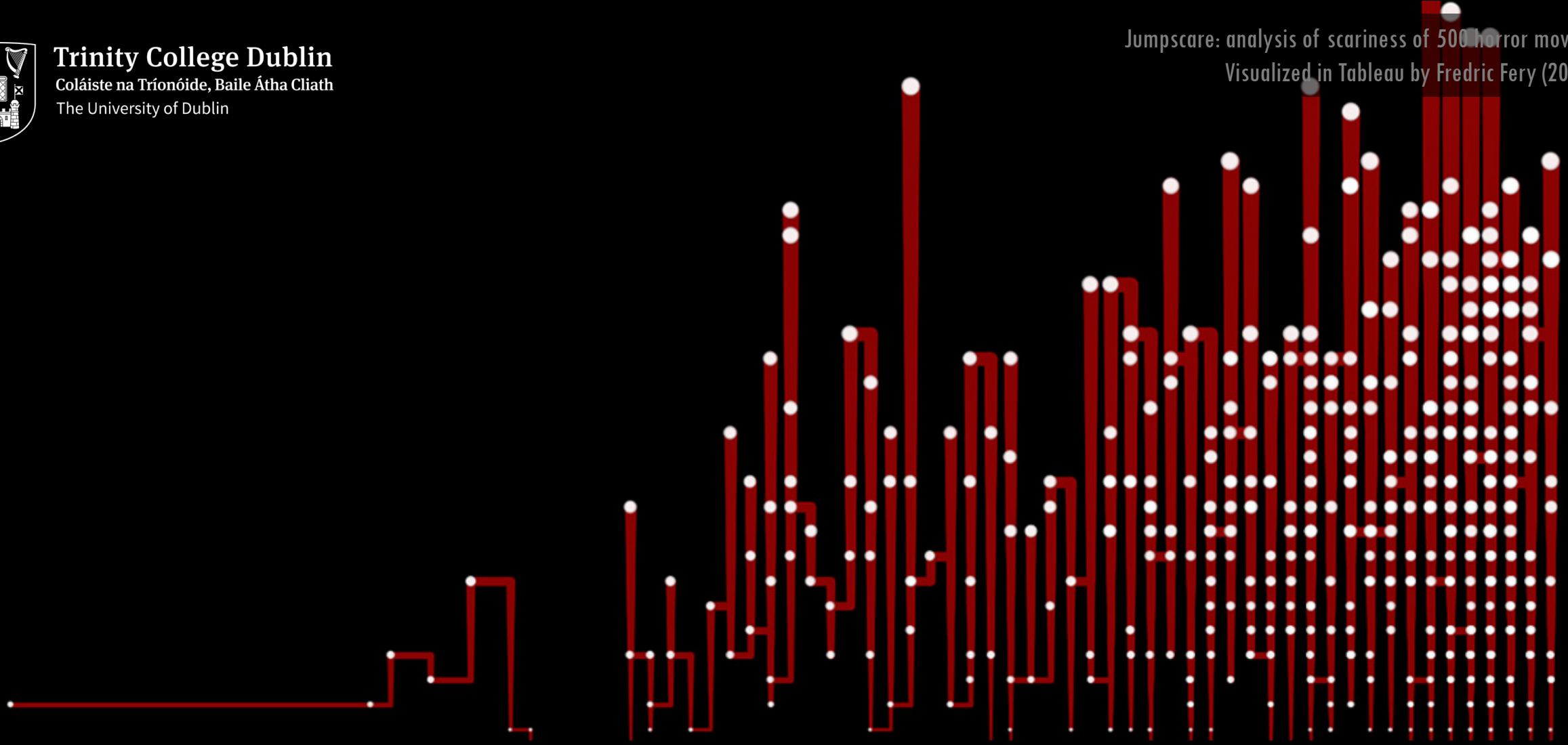
Recommended Reading

Idioms

- Chapter 7 – “Arrange Tables” in Visualization Analysis and Design, Tamara Munzner 2014 [Available as e-book in Library Reading Rooms]

- *Alternative reference*: the following early draft of a book by Munzner is freely available online on the authors page. Chapter 8 “ discusses spatial and non-spatial visualization idioms:
  - Chapter 8 – “Making Views” in Visualization Design and Analysis: Abstractions, Principles, and Methods (DRAFT). Tamara Munzner, 2012. [https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=142]

Jumpscare: analysis of scariness of 500 horror movies
Visualized in Tableau by Fredric Fery (2019)



Vis TooIS #3 : Tableau

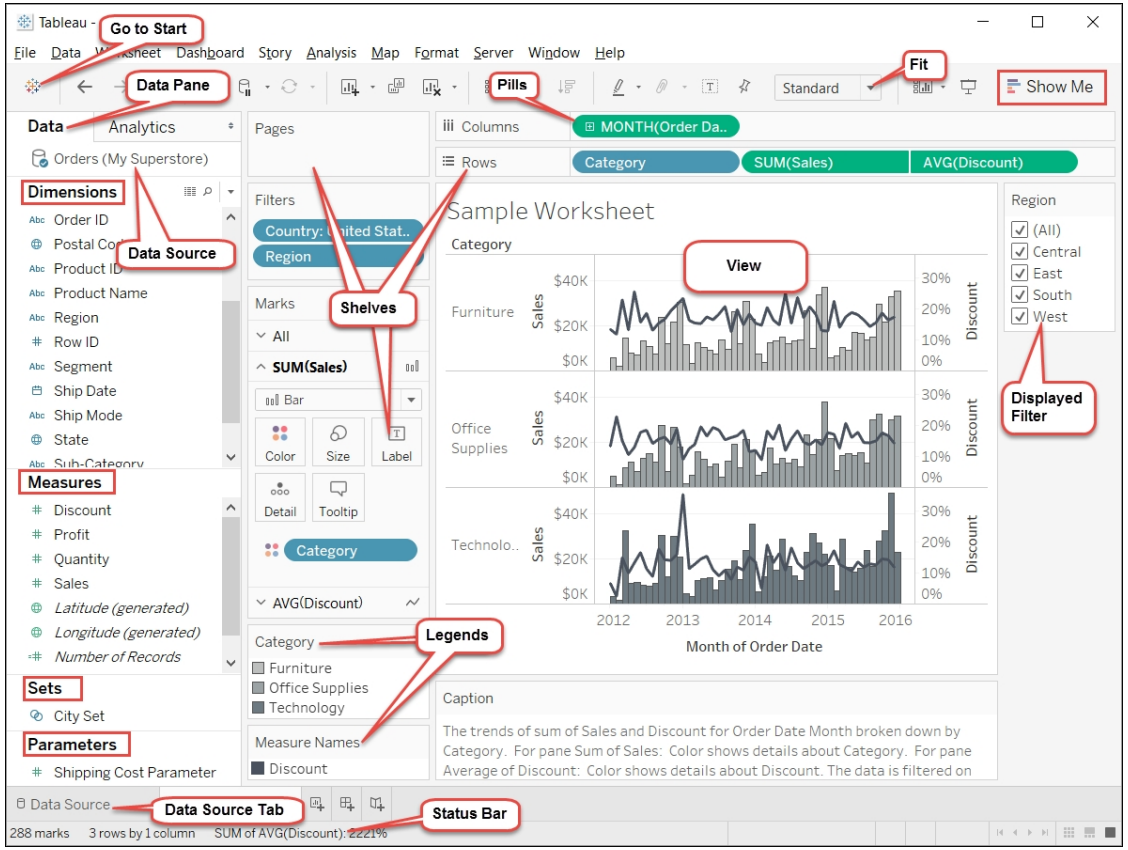
What is Tableau

A program for visual analytics that allows users to create complex visualizations

Easy to use: GUI-based; simple drag and drop. No need for any coding

Technical features include:

- Pre-set encoding idioms: data visualization, map visualization,
- Color maps, filters
- Calculating/deriving intermediate data
- Dashboards: combined visualization idioms
- Interactivity & GUI elements
- Loading & organizing data



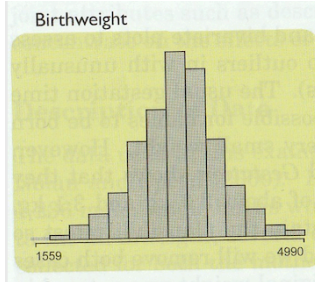
From Mastering Tableau by David Baldwin (2016)

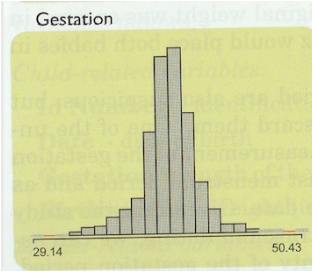
Try This at home: Birthweight vs smoking

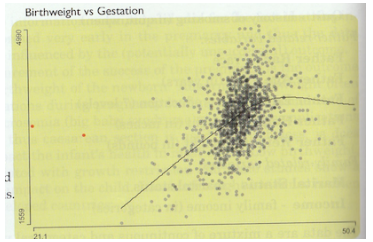
Get Data* from: http://www.interactivegraphics.org/Datasets.html

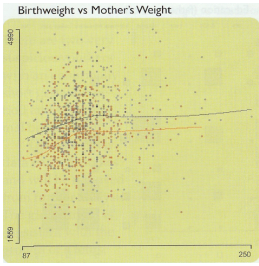
Try to re-create the following graphs (from the book):

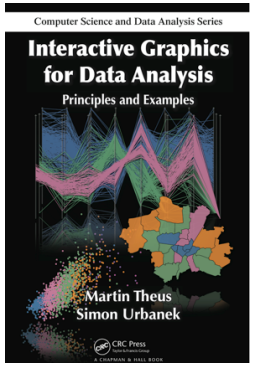
- Histogram of Birthweights in the data set
- Histogram of Gestation Periods
- Scatterplot of Birthweight vs Gestation
- Scatterplot and trendline of Birthweight vs Mothers Weight
- Box-and Whisker Plot of Birthweight by Smoking Patterns











N.B. the data in the printed book seems to be slightly different, so don't be too worried if it doesn't look identical.

* Supplementary materials fom "Interactive Graphics for Data Analysis". Theus and Urbanek. CRC-Press. 2009.

How to get Started

Get Tableau

- Tableau Desktop for students (1 year free)
  https://www.tableau.com/en-gb/academic/students
- Tableau Public (free, no local saving of workbooks):
  https://public.tableau.com/en-us/s/

Look at the Gallery to see what it can do:

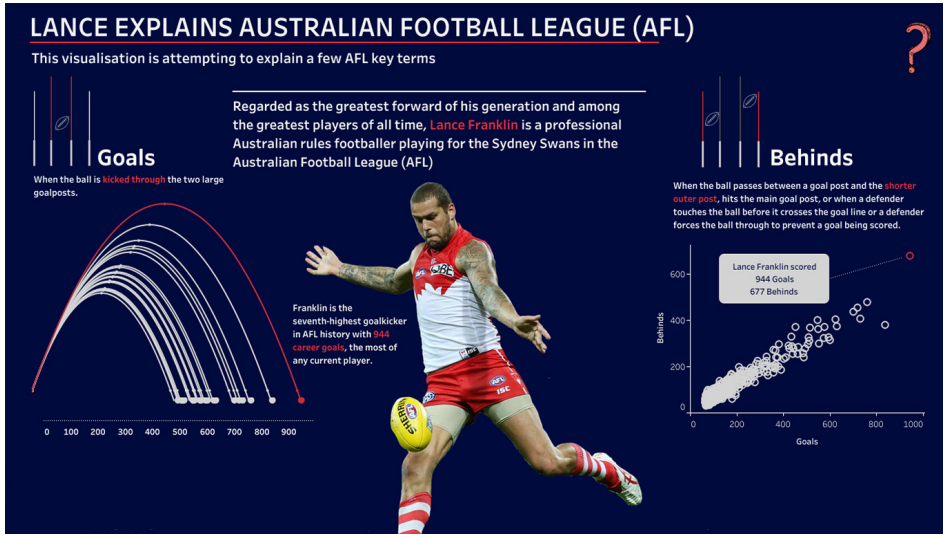
- https://public.tableau.com/en-gb/gallery

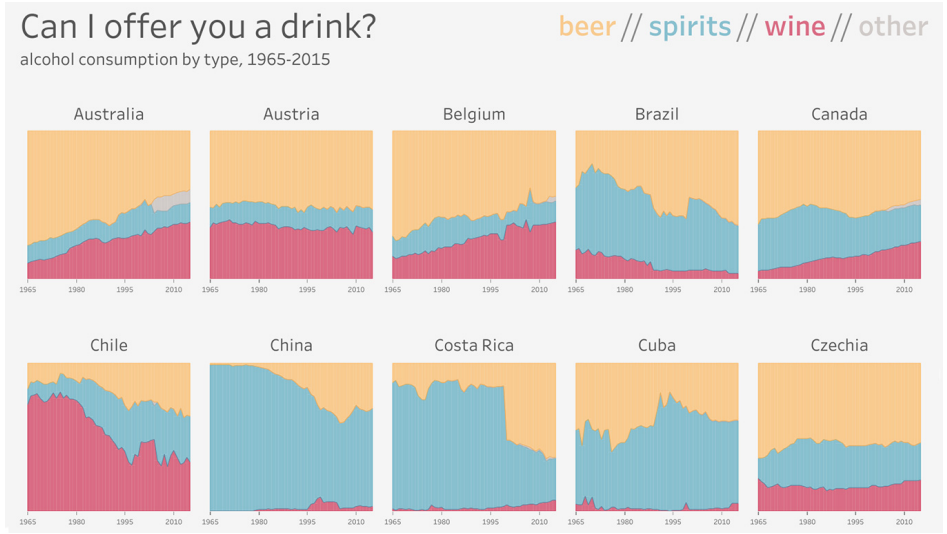
Try some Tutorials:

- https://public.tableau.com/en-us/s/resources

Lots of Sample Data:

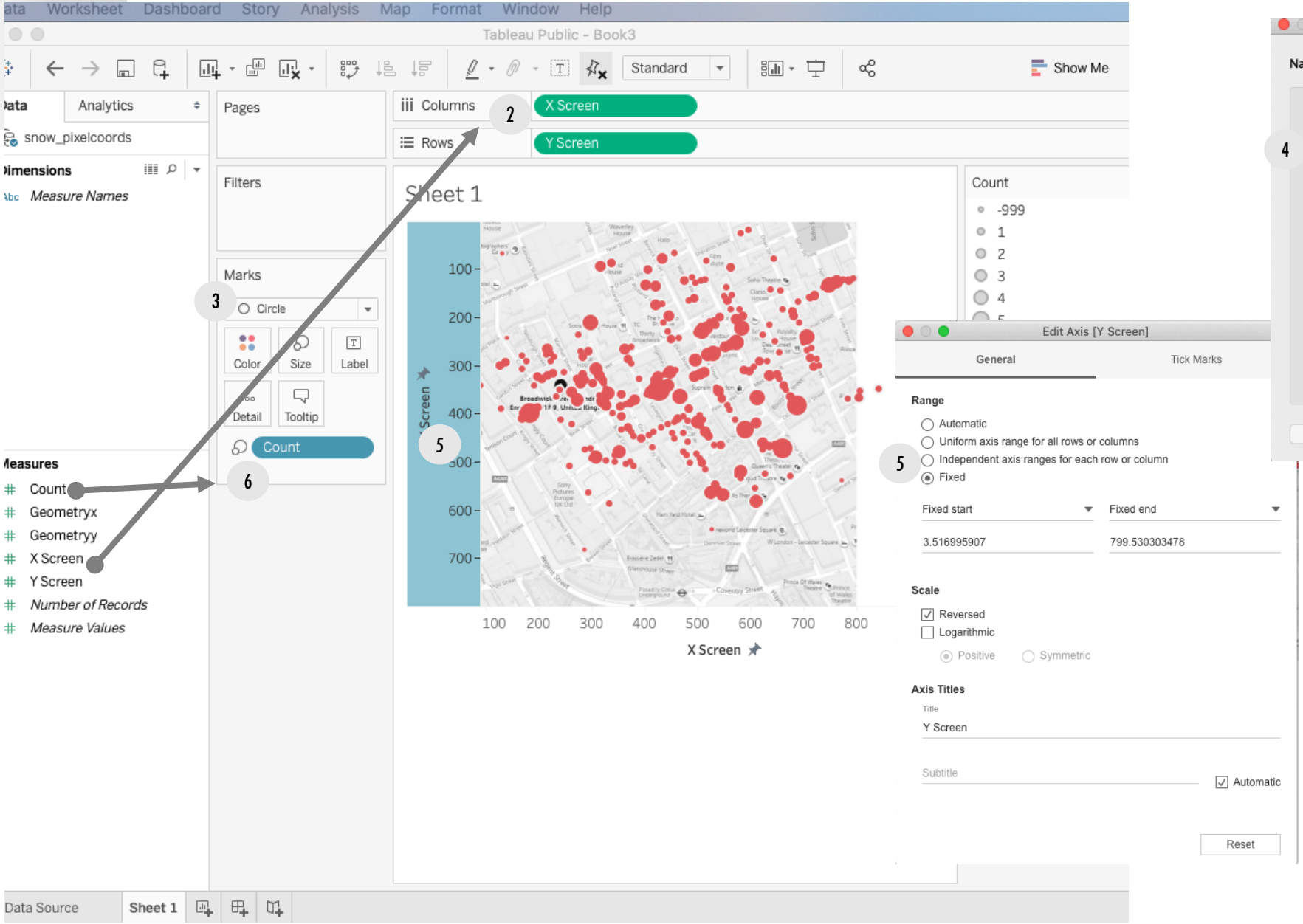
- https://public.tableau.com/en-us/s/resources



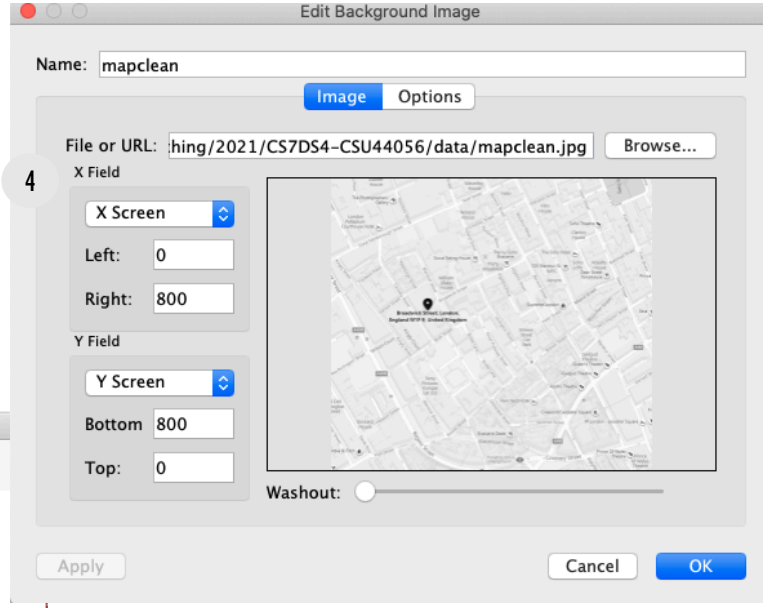


Snow Map Example: Walkthrough

1. Load the snow data into Tableau (available on blackboard as snow_pixelcoords.csv)
2. Drag the “X Screen” and “Y Screen” measures from the left Data Pane into the “Columns” and “Rows” shelves above the chart
   - Click on the green ‘pills’ and select “Dimension” in the pop-up menu
3. In the “Marks” pane, change drop down list to “Circle”
4. In the menu bar, go to Map → Background Image … →snow_pixelscoords.
   - Click on “Add Image …”. Find the map image “mapclean.jpg” (you’ll need to download this from blackboard)
   - Change the X Field to Left: 0; Right: 800
   - Change the Y Field to Bottom: 800; Top: 0
5. Right click on the Y axis label and click “Edit Axis..”
   - Set range to Fixed.
   - Set Fixed Start to 0 and Fixed End to 800
   - Select Reveresed under “Scale”
6. Drag the #Count measure from the left panel onto the “Size” box in the Marks Pane.
   - Click on the green pill for Count, tick the Dimension and Discrete options in the pop=up menu
   - Click on the “Color” box and choose red


Worksheet Dashboard Story Analysis Map Format Window Help
Tableau Public - Book3
Standard
Show Me
Analytics
snow_pixelcoords
Dimensions
Measure Names
Pages
Filters
Marks
Circle
Color
Size
Label
Detail
Tooltip
Count
Measures
Count
Geometryx
Geometryy
X Screen
Y Screen
Number of Records
Measure Values
Columns
X Screen
Rows
Y Screen
Sheet 1
Count
-999
1
2
3
4
100
200
300
400
500
600
700
800
Y Screen
X Screen
Data Source
Sheet 1
2
3
4
5
6

Edit Axis [Y Screen]
General
Tick Marks
Range
Automatic
Uniform axis range for all rows or columns
Independent axis ranges for each row or column
Fixed
Fixed start
Fixed end
3.516995907
799.530303478
Scale
Reversed
Logarithmic
Positive
Symmetric
Axis Titles
Title
Y Screen
Subtitle
Automatic
Reset
5


Edit Background Image
Name: mapclean
Image
Options
File or URL: hing/2021/CS7DS4-CSU44056/data/mapclean.jpg
Browse...
X Field
X Screen
Left: 0
Right: 800
Y Field
Y Screen
Bottom 800
Top: 0
Washout:
Apply
Cancel
OK
4