

Foundations 1

Introduction to Data Visualization

DataVis 2020
<http://datavis2020.github.io>

Dr. Benjamin Bach



THE UNIVERSITY
of EDINBURGH

Why is visualization **working?**

Find every '5'

9176867991960386255930486551443
9353652502752141394912668766013
5095444663473545993604636448078
6425620171155906710121069218701
0584448205561385902503573068845
1172102391879794238892669764599
6987227951015948926017759166604
8436204036527029291707742717051
3280506428553158863136868380421
4055135906770329783671748392874

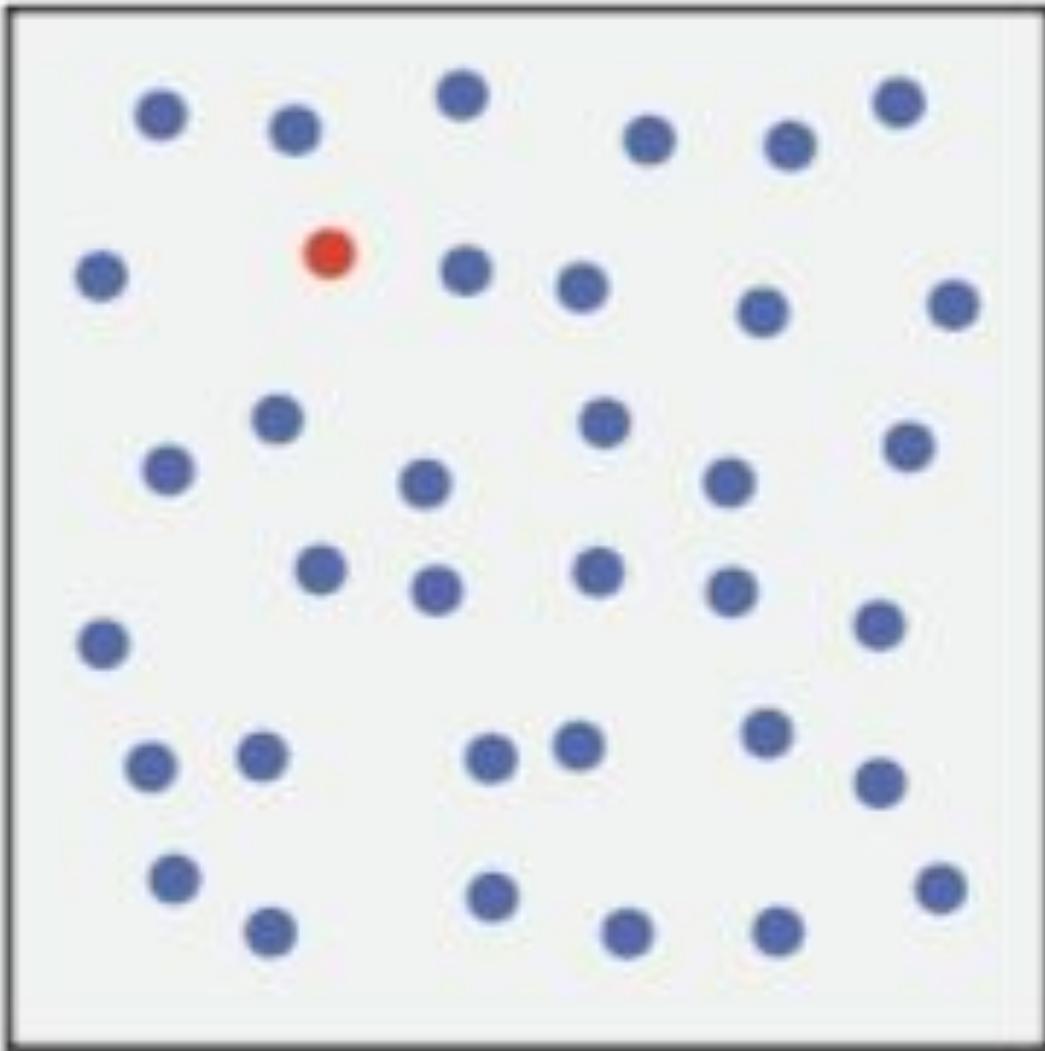
Pre-attentiveness:

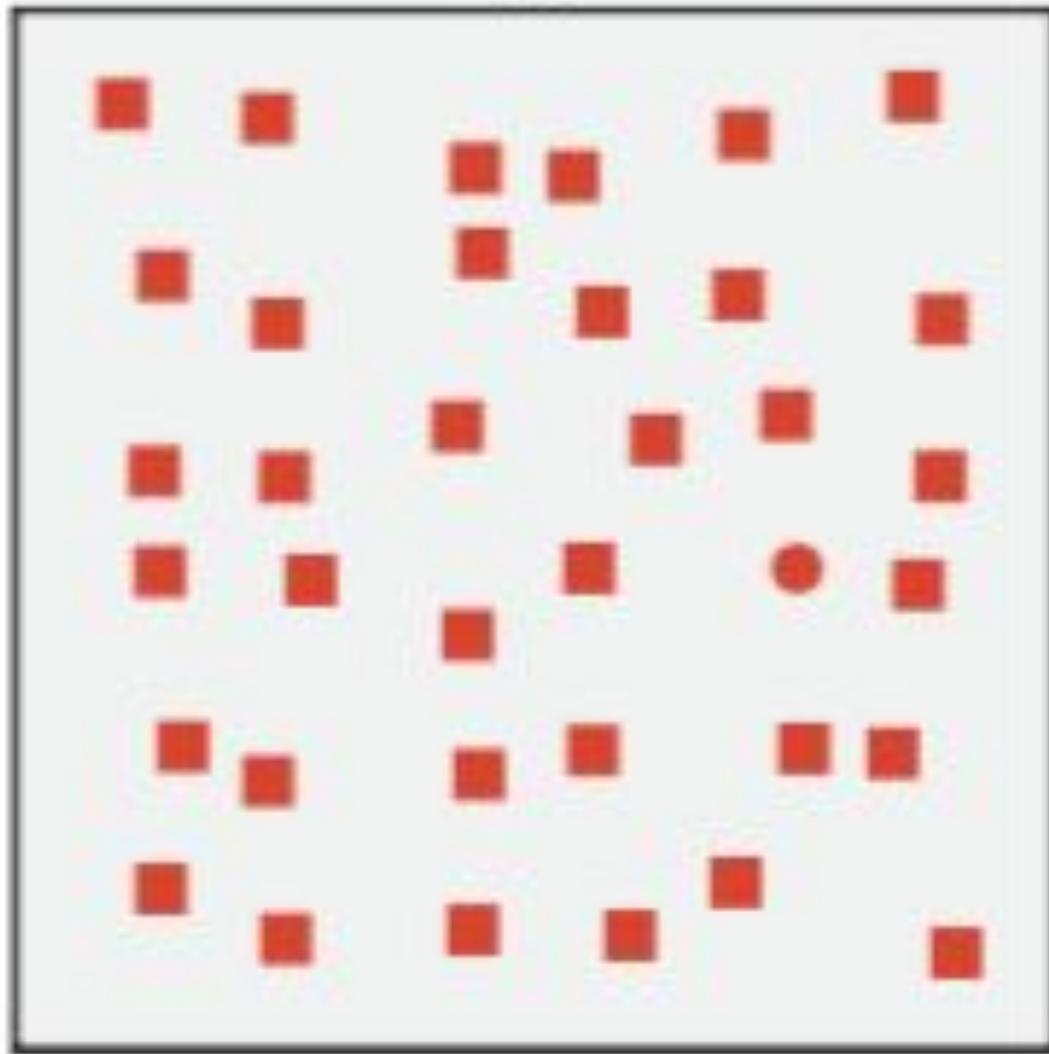
- "is the subconscious accumulation of information from the environment"
- realizing something before you think

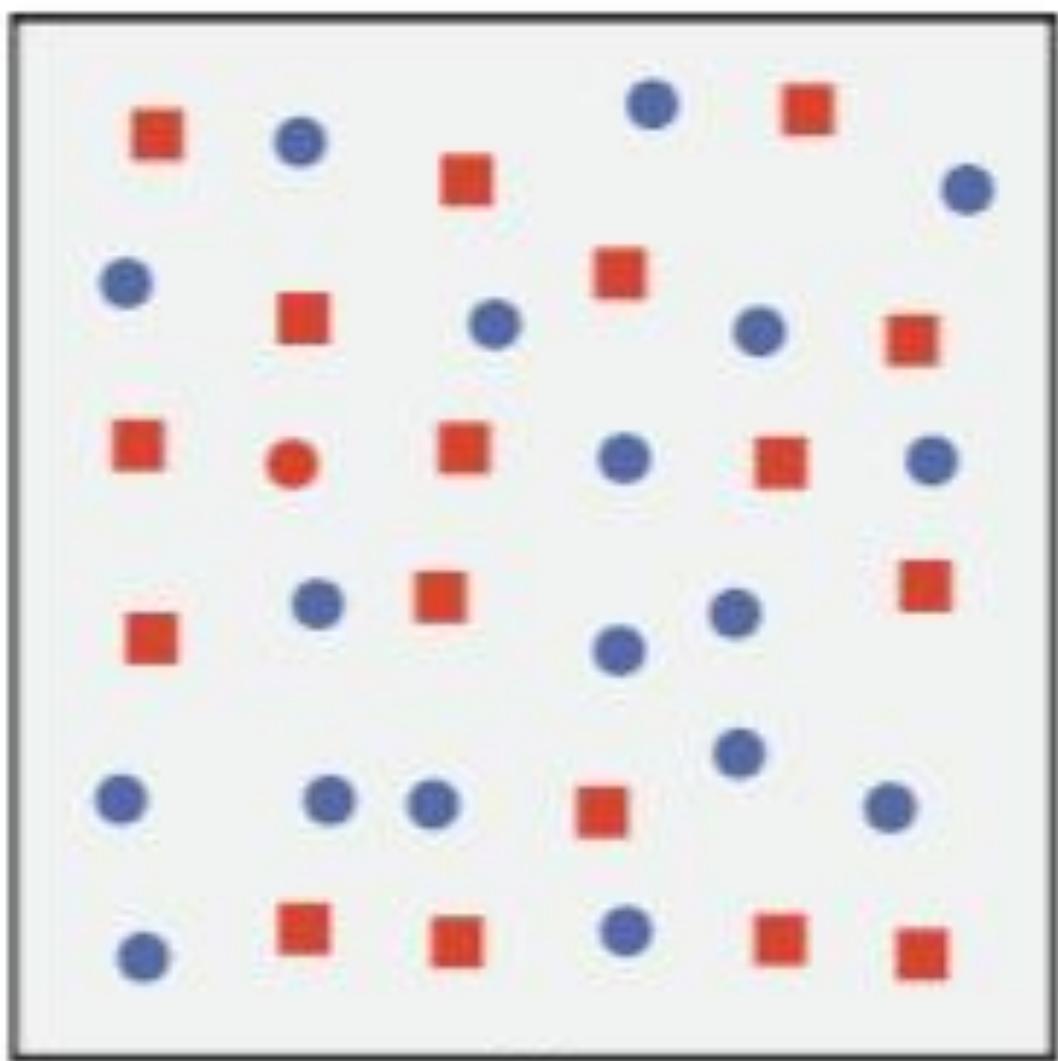
Parallelity:

- Perceiving stimuli in parallel



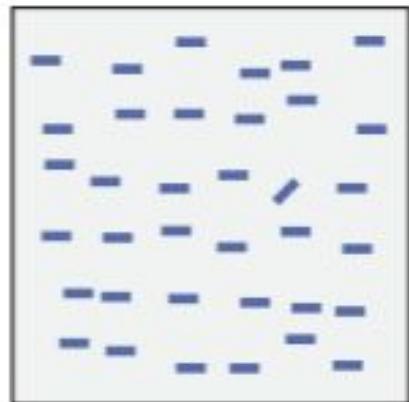




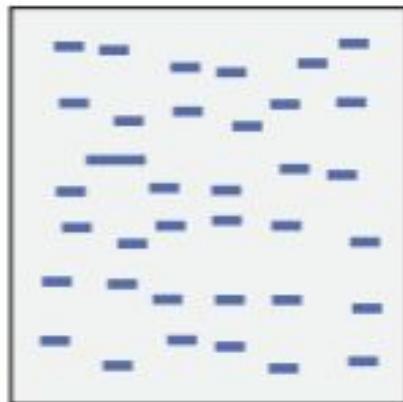


c c co co
c c c c c c
c c c c c c
c o c c c c
c c c c c c
c c c c c c

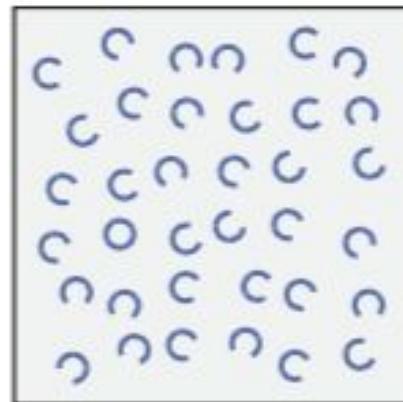
Preattentiveness



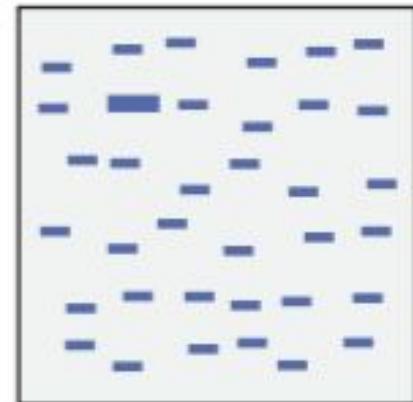
orientation
[16], [17], [18], [19]



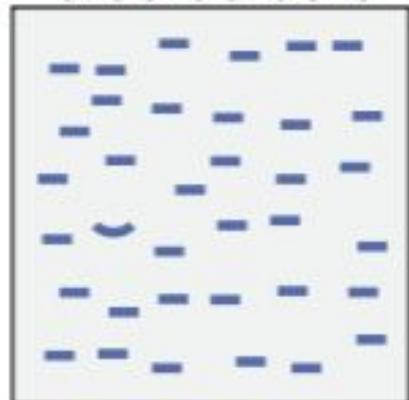
length
[20], [21]



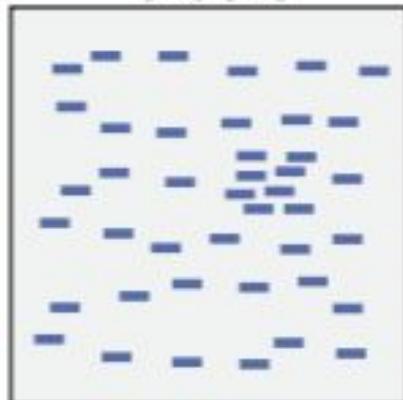
closure
[16]



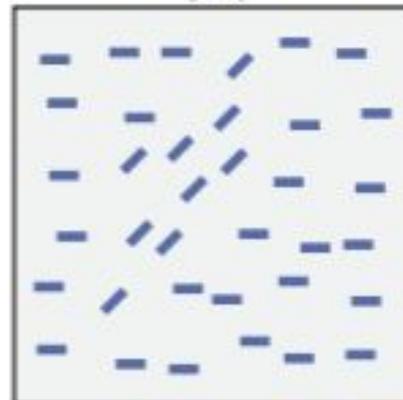
size
[22], [23]



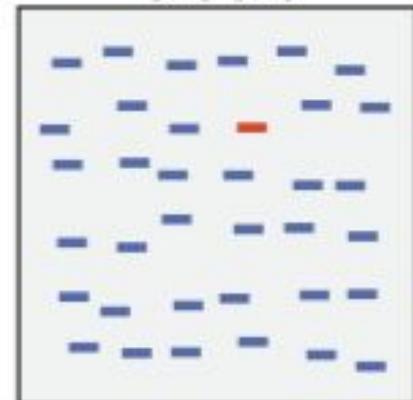
curvature
[21]



density
[23]

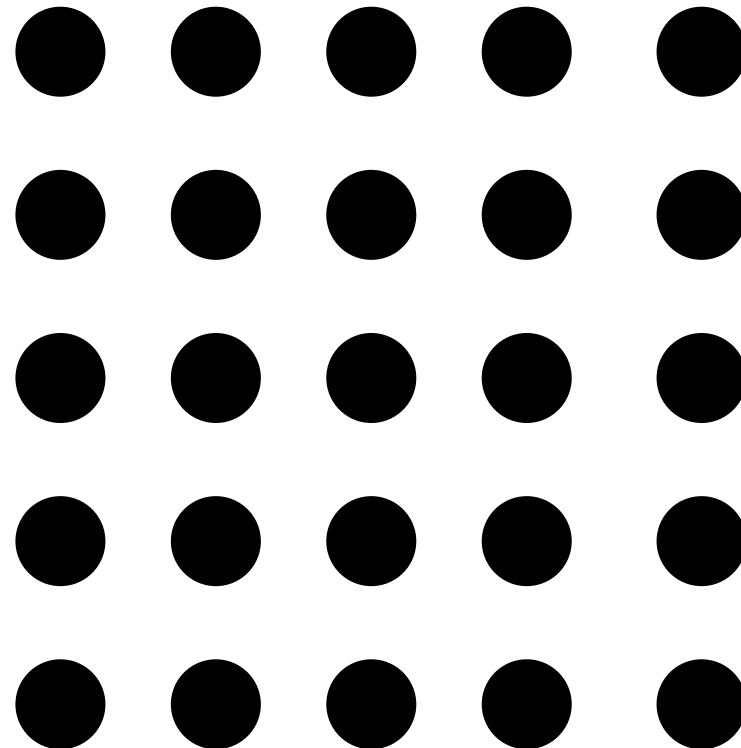


number
[20], [24], [25]

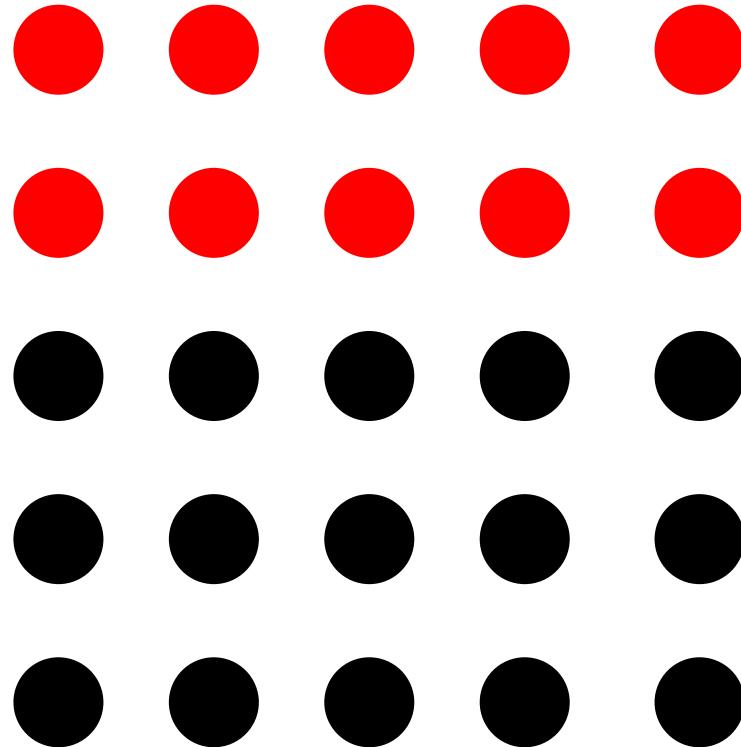


hue
[23], [26], [27], [28], [29]

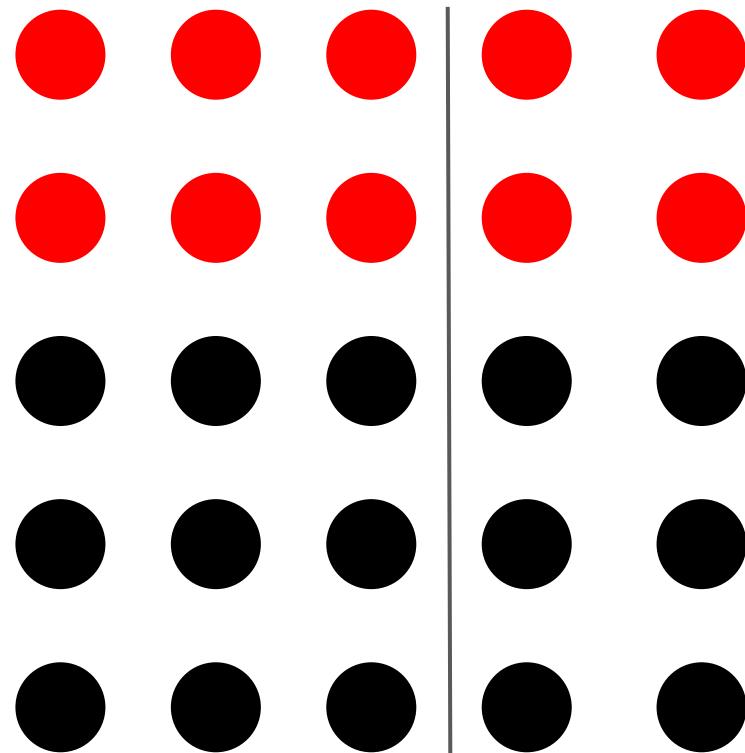
Gestalt Laws



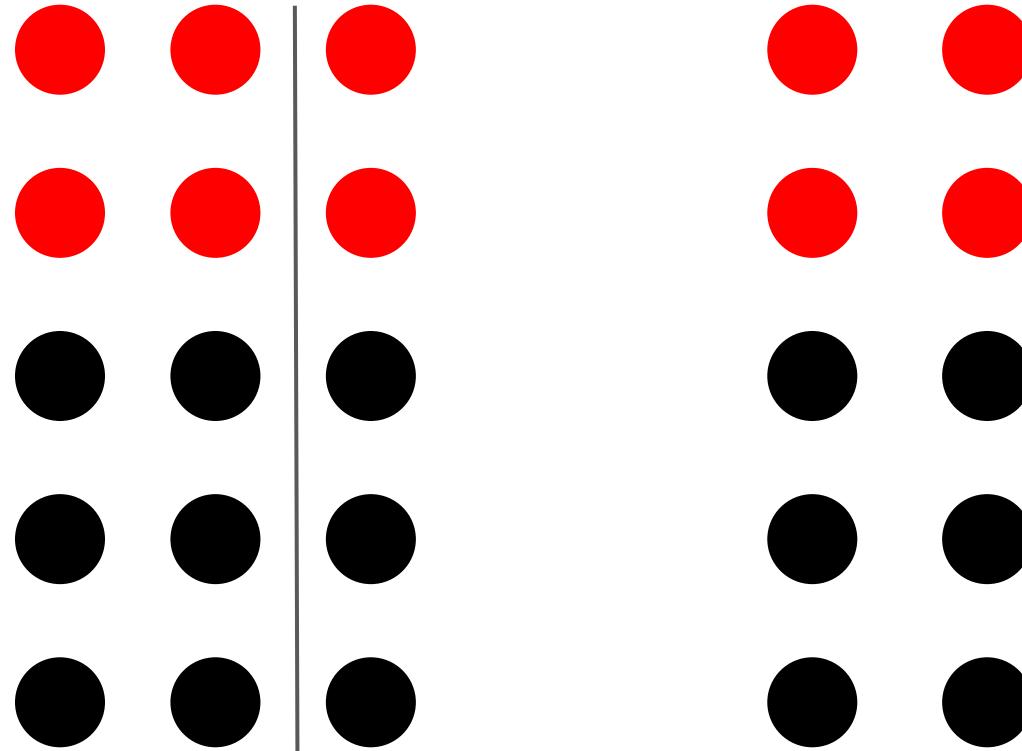
Law of Similarity



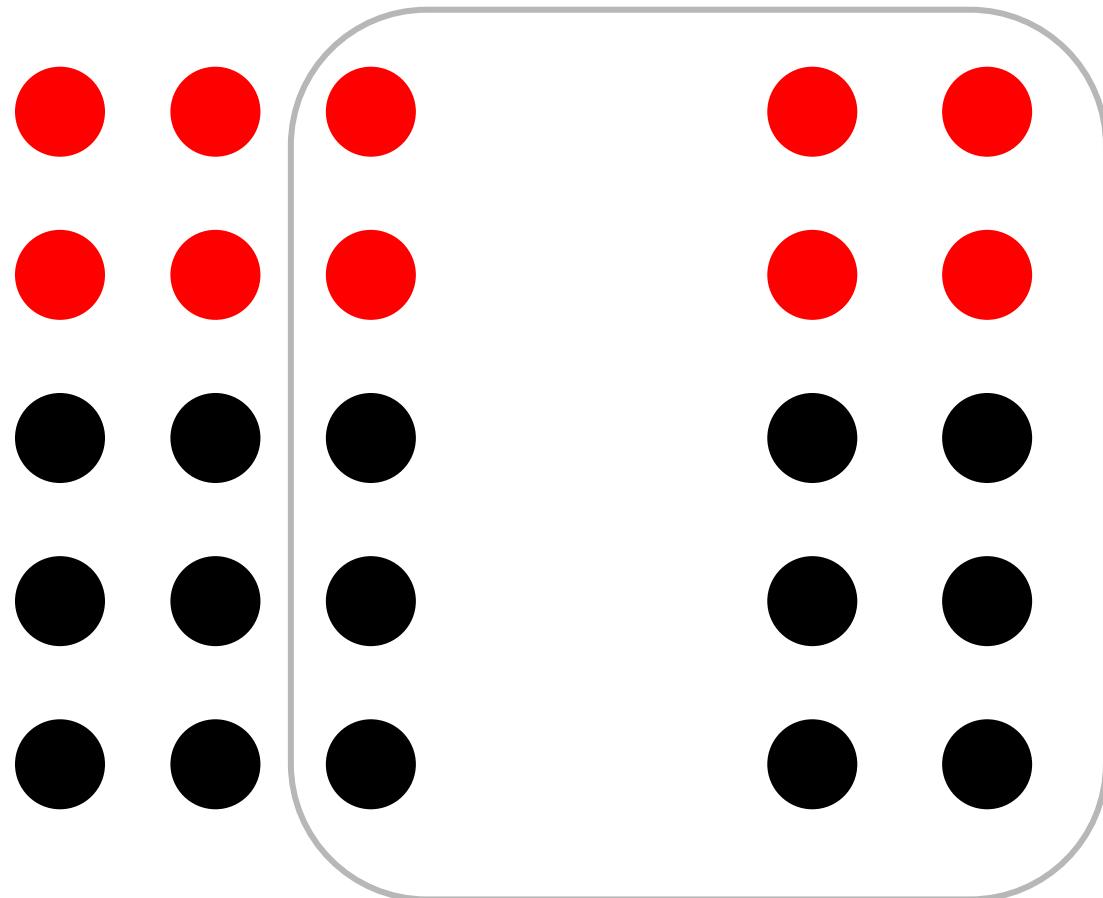
Law of Closure



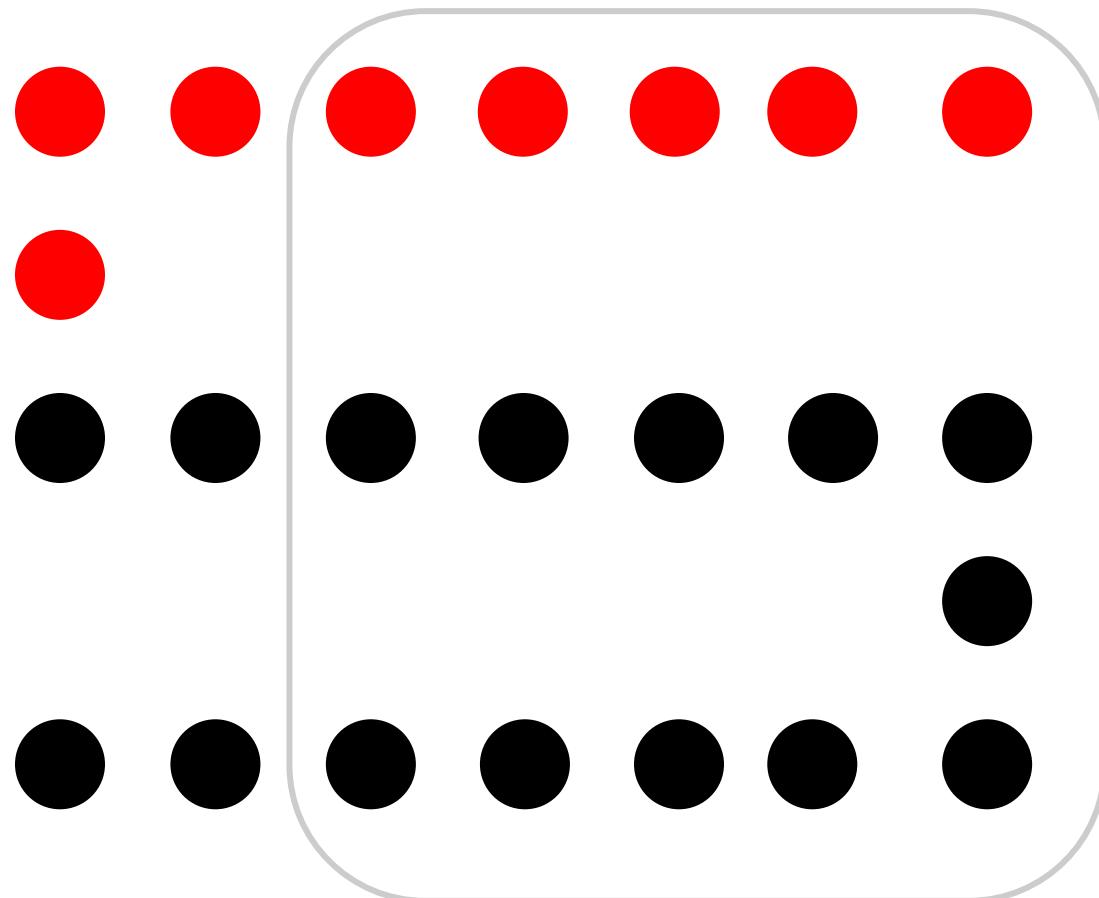
Law of Proximity



Law of Enclosure

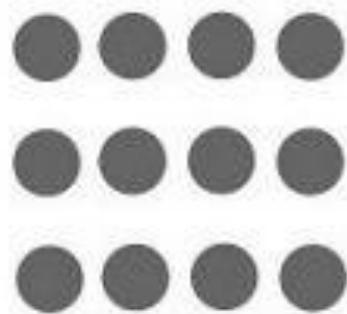


Law of Continuity

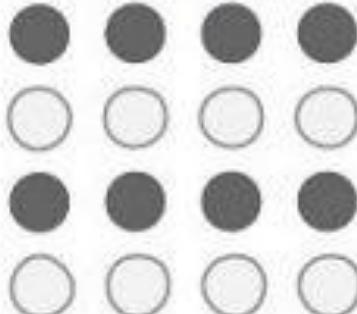


Gestalt Laws

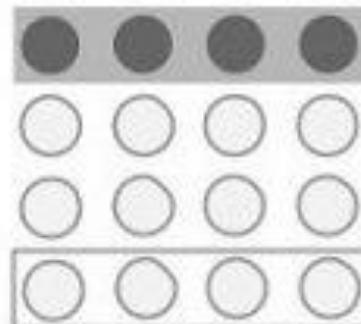
Proximity



Similarity



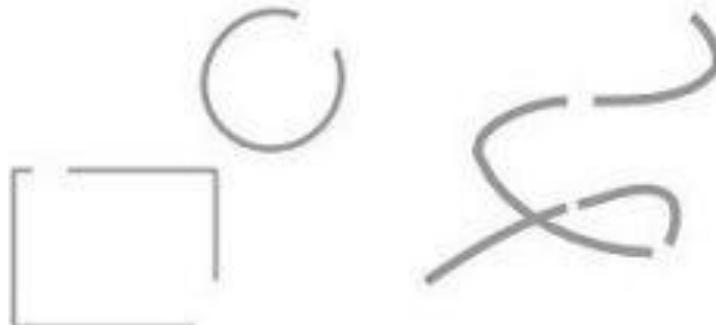
Enclosure



Symmetry



Closure



Continuity

Connection

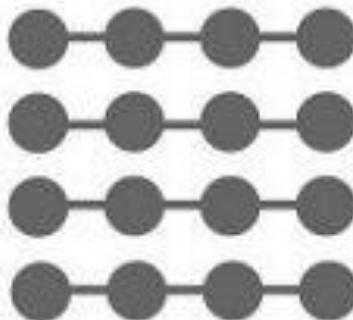


Figure & ground



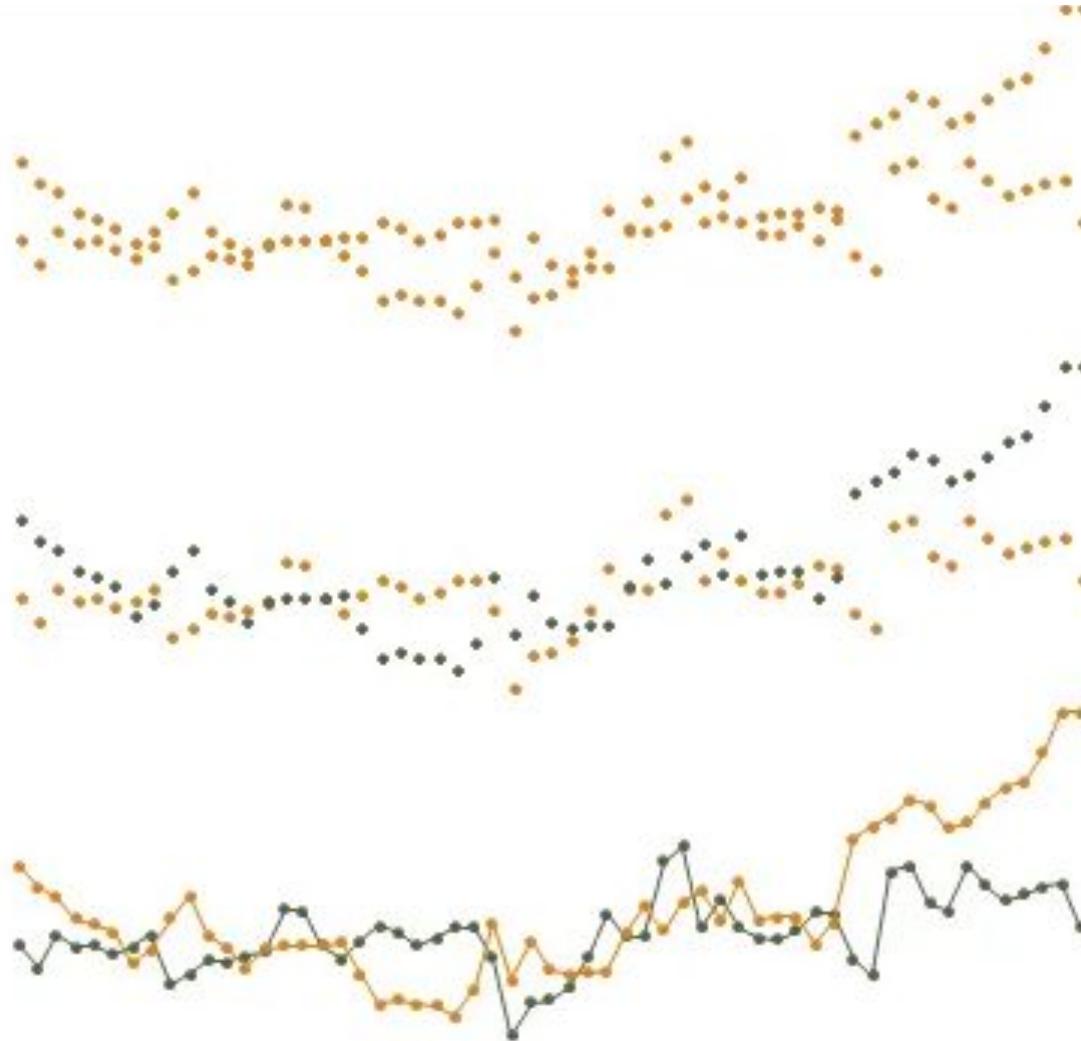
Gestalt Laws



Gestalt Laws



Gestalt Laws

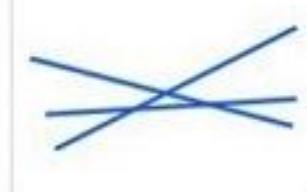


Visual Patterns

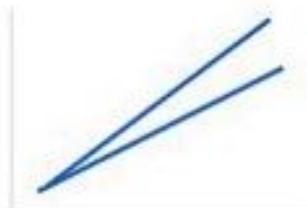
High, low and in between



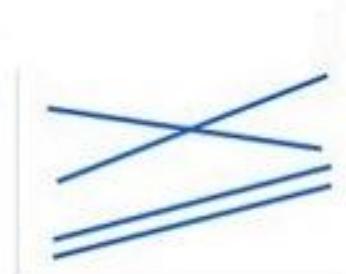
Going up, going down and remaining flat



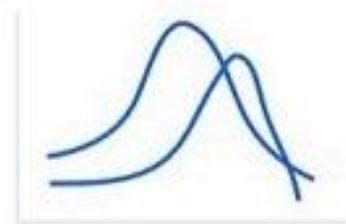
Steep and gradual



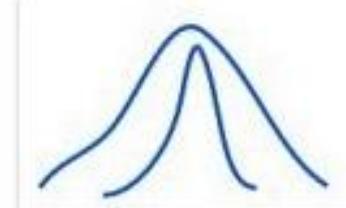
Non-intersecting and intersecting



Symmetrical and skewed



Wide and narrow

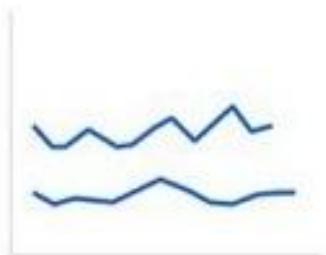


Visual Patterns

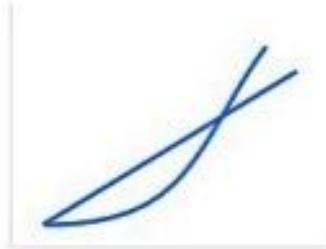
Steady and fluctuating



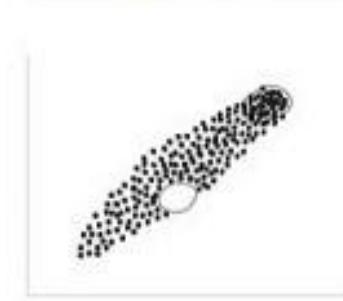
Random and repeating



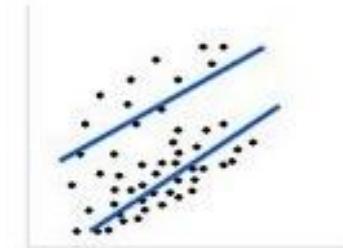
Straight and curved



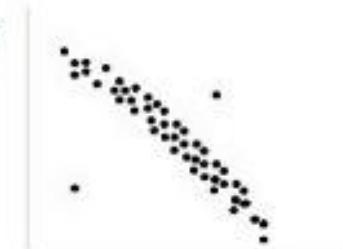
Clusters and gaps



Tightly and loosely distributed

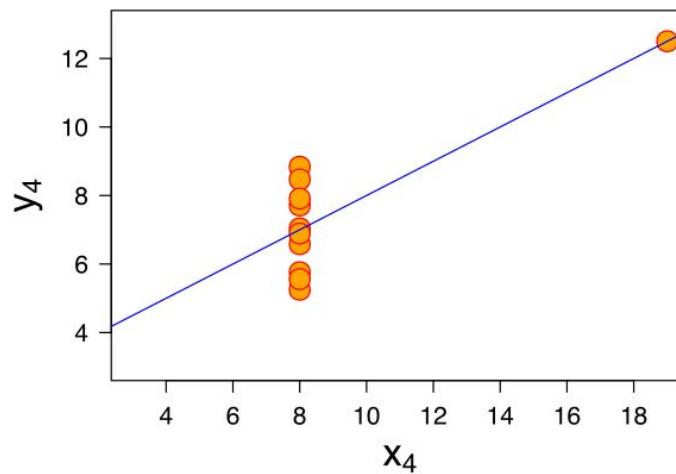
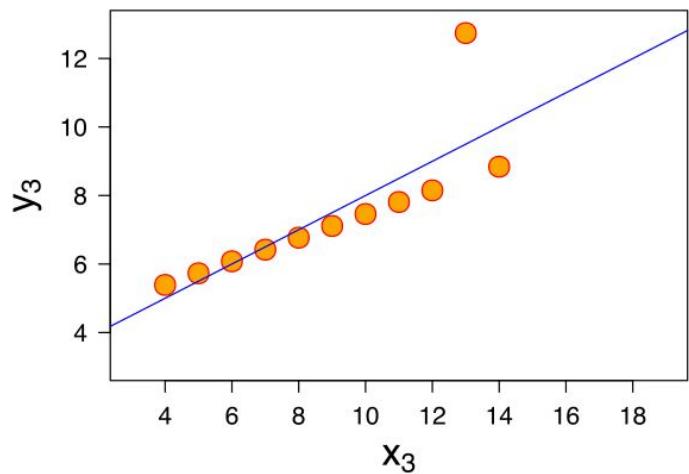
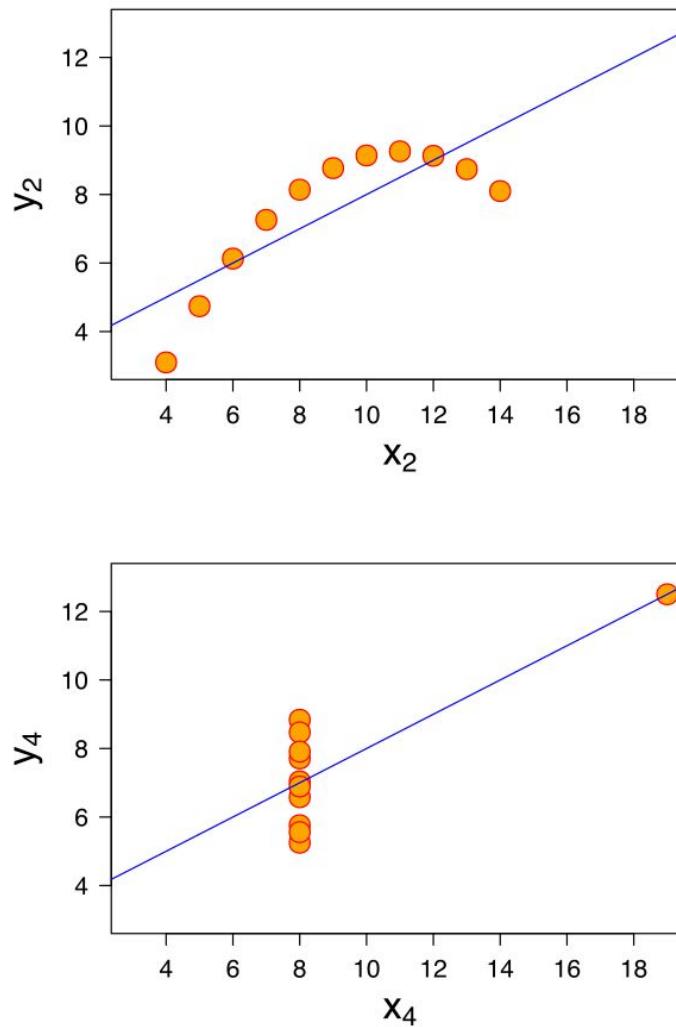
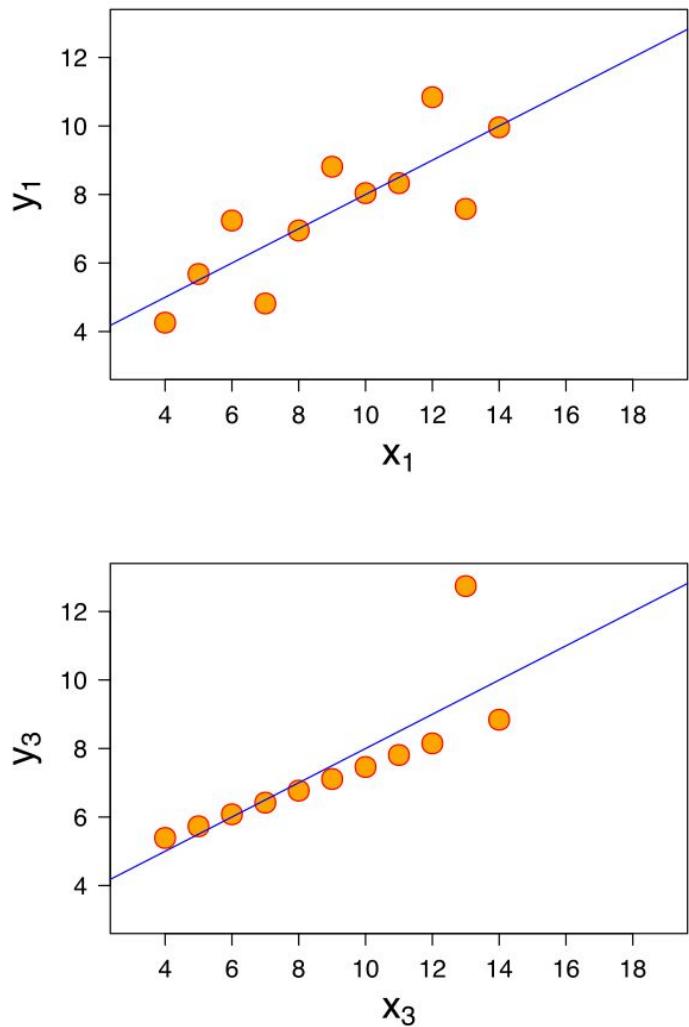


Normal and abnormal

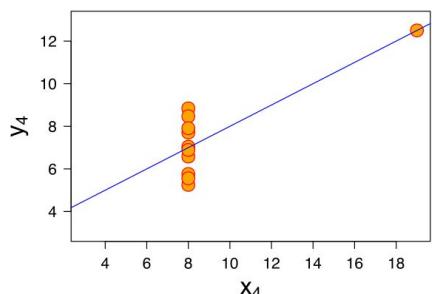
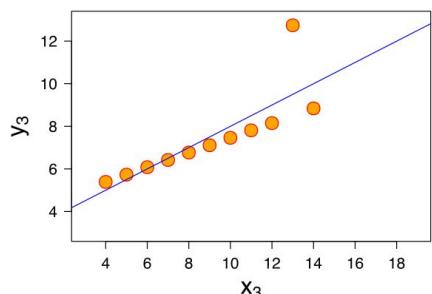
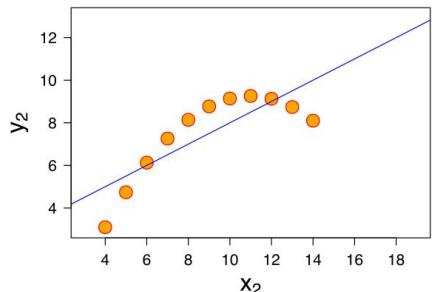
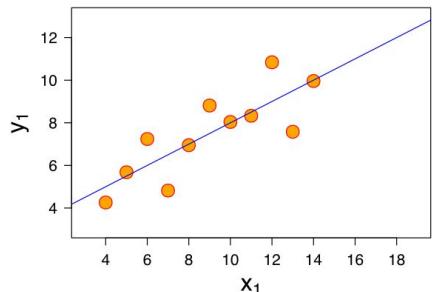


Why are we visualizing?

Anscombe's Quartet



Anscombe's Quartet



Property	Value
Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression	0.67

Exploration

Insights

Explanation

Data centered

Human centered

Domain experts

Non-experts

Generating
Insights

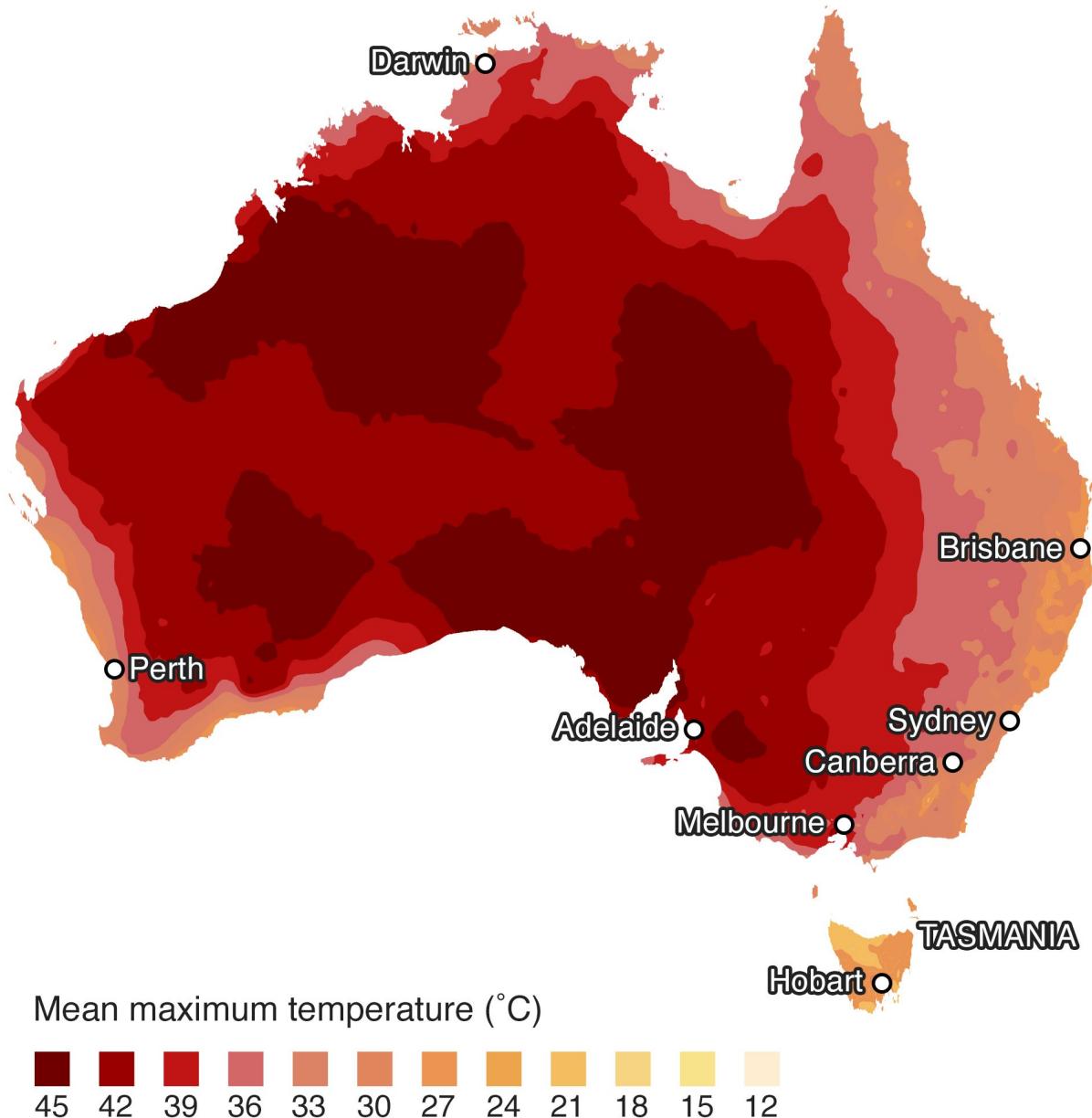
Conveying
messages

Lab Setting

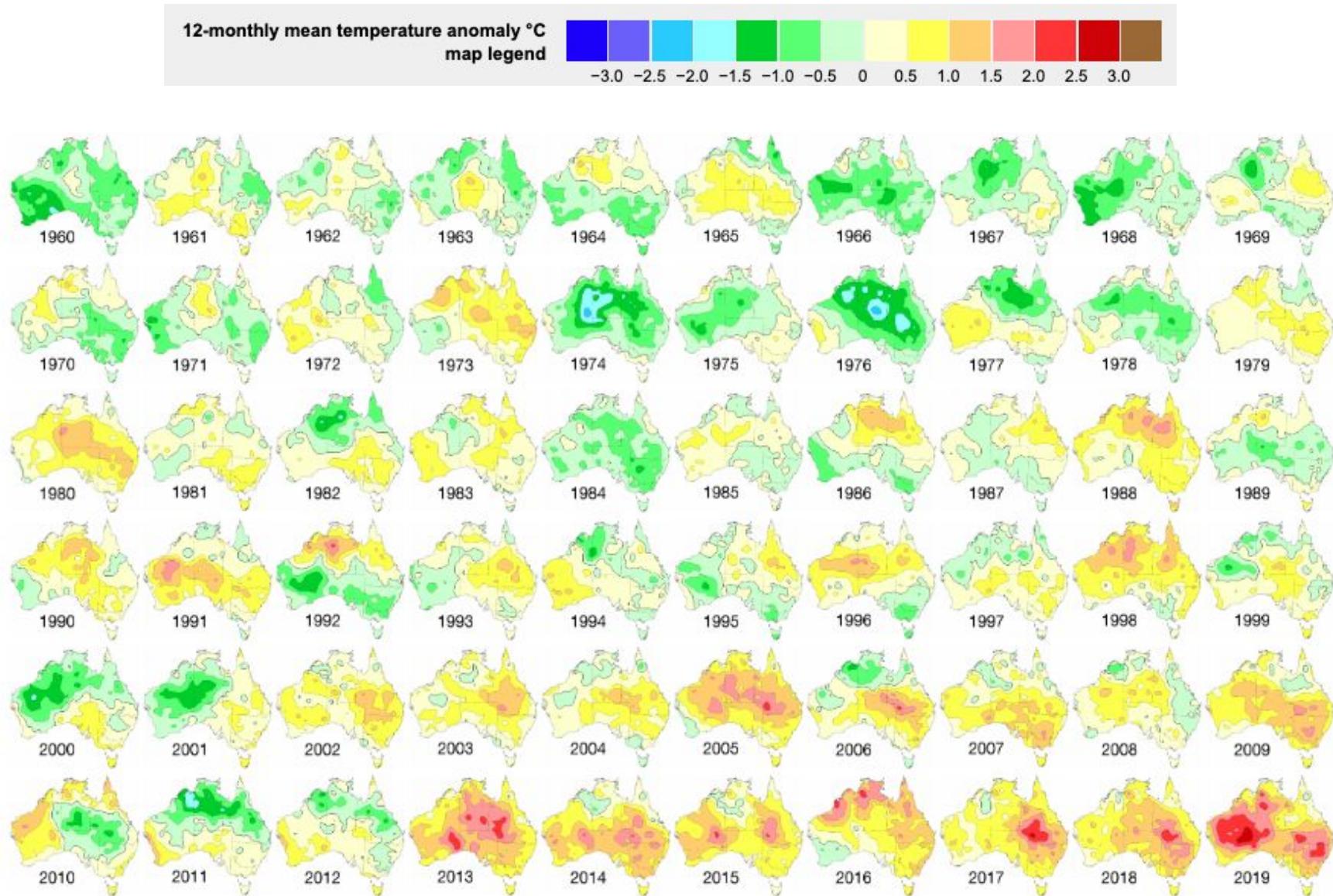
In-the-wild /
diverse

Exploratory

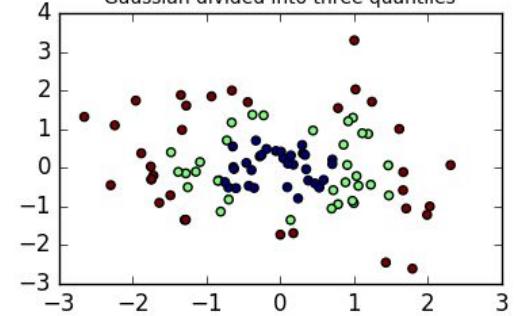
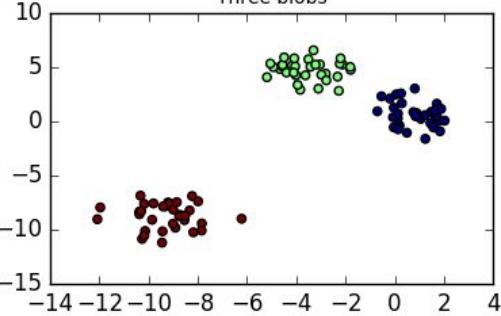
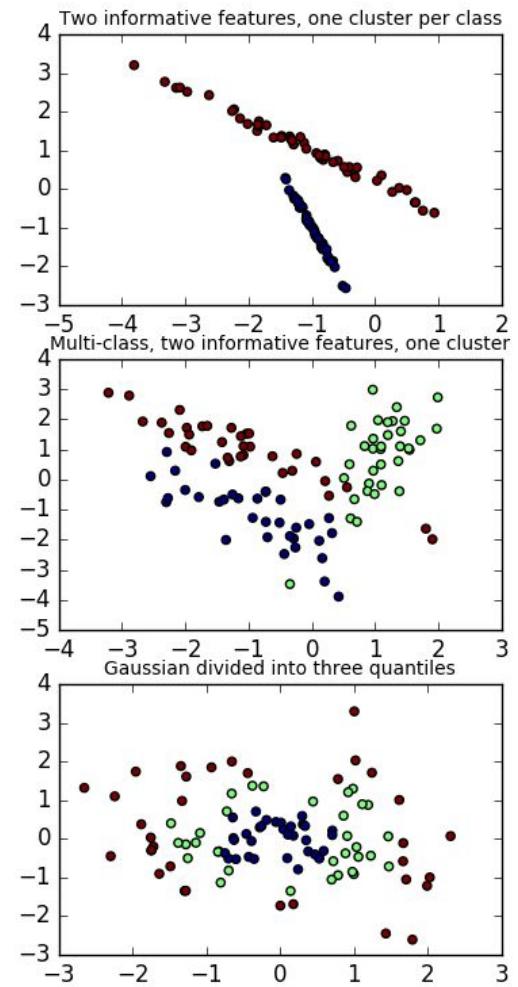
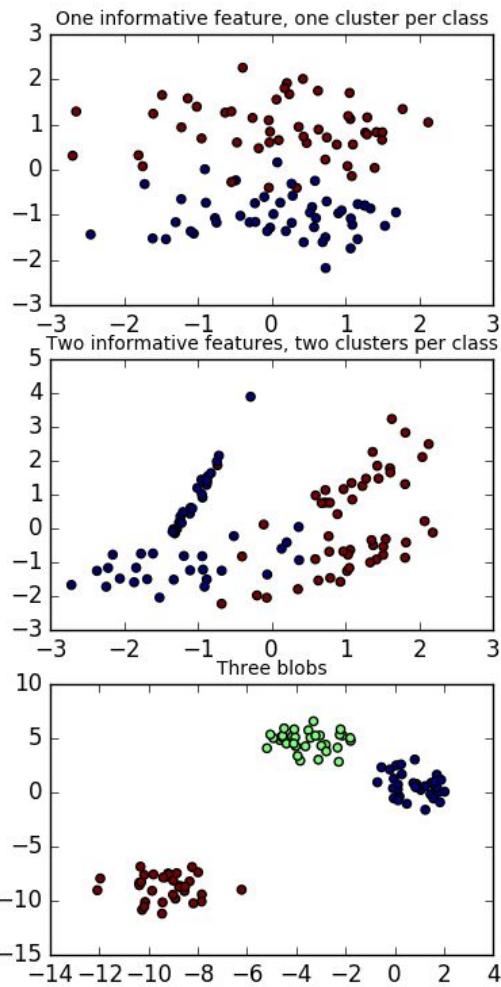
Mean maximum temperature, 19 Dec



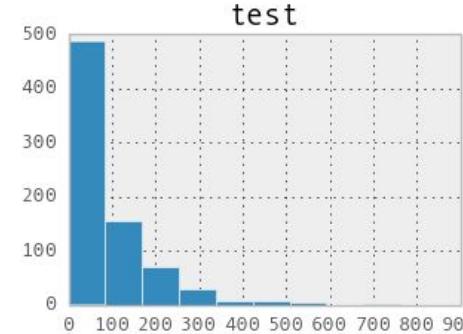
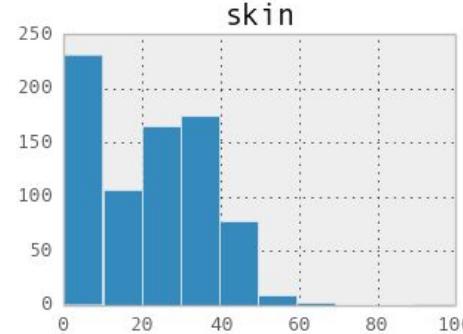
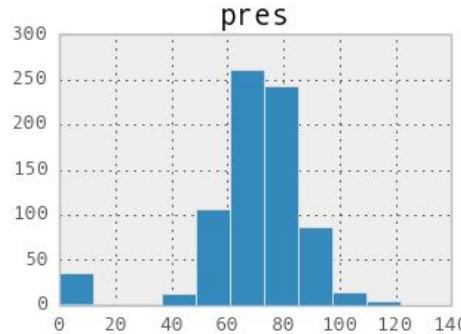
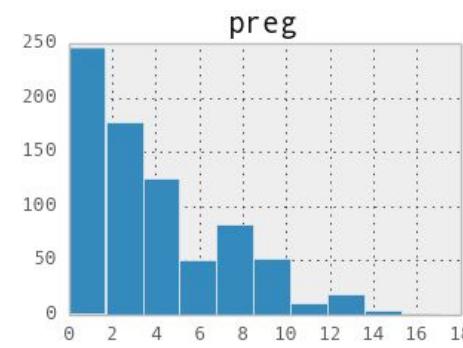
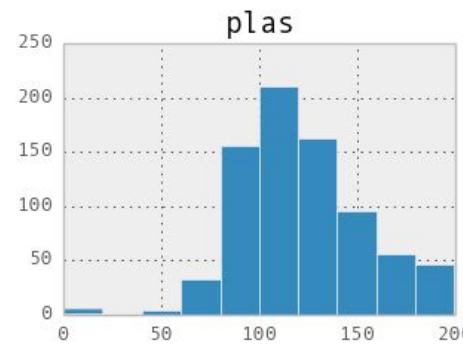
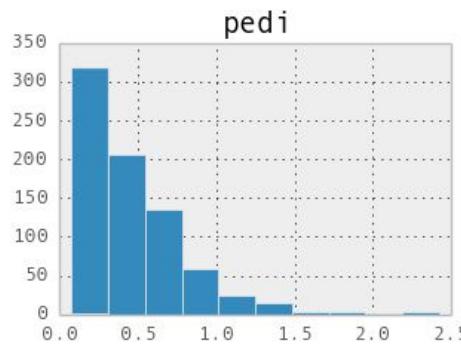
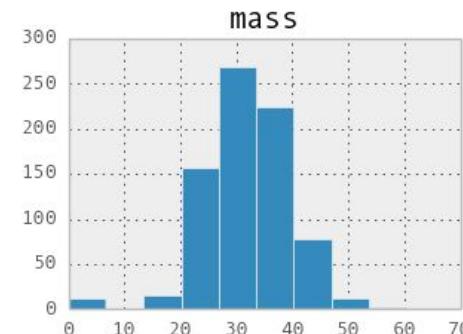
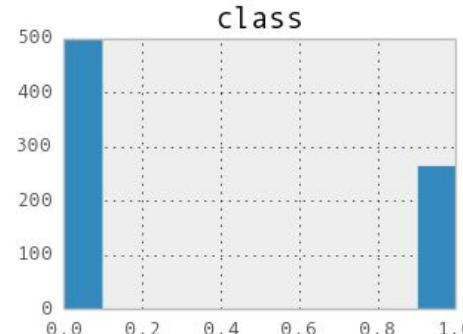
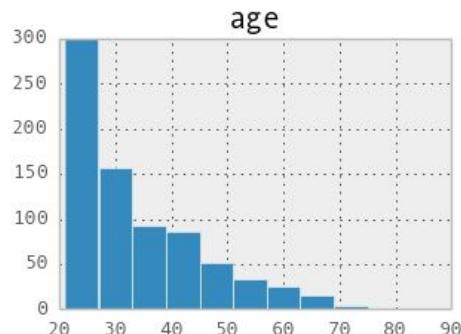
Exploratory: trends + outliers



Exploratory: Trends + Outliers

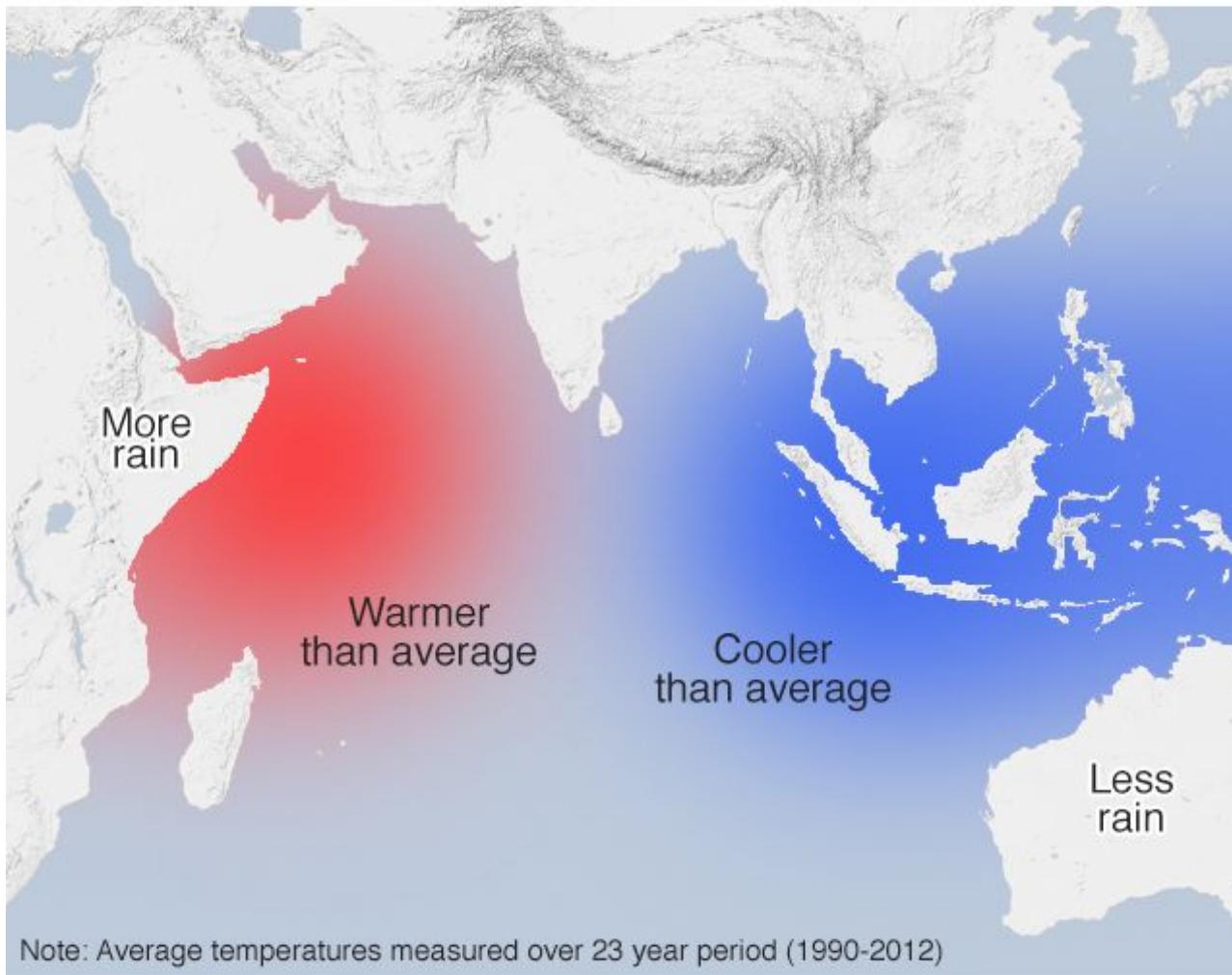


Exploratory: Data Quality



Explanatory

A positive Indian Ocean Dipole means a wetter west and drier east

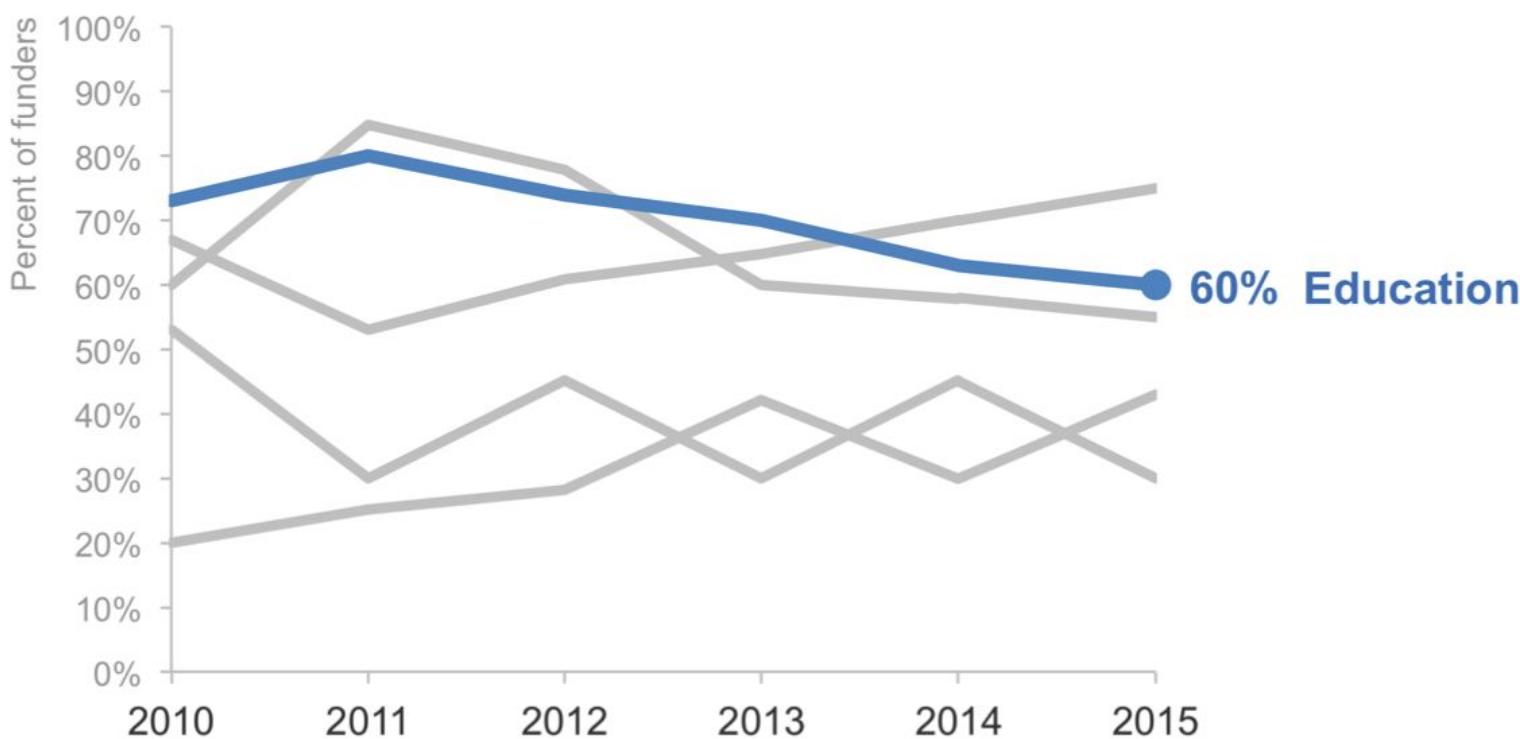


Explanation: Presentation

Clarity, highlight elements, direct user attention, ..

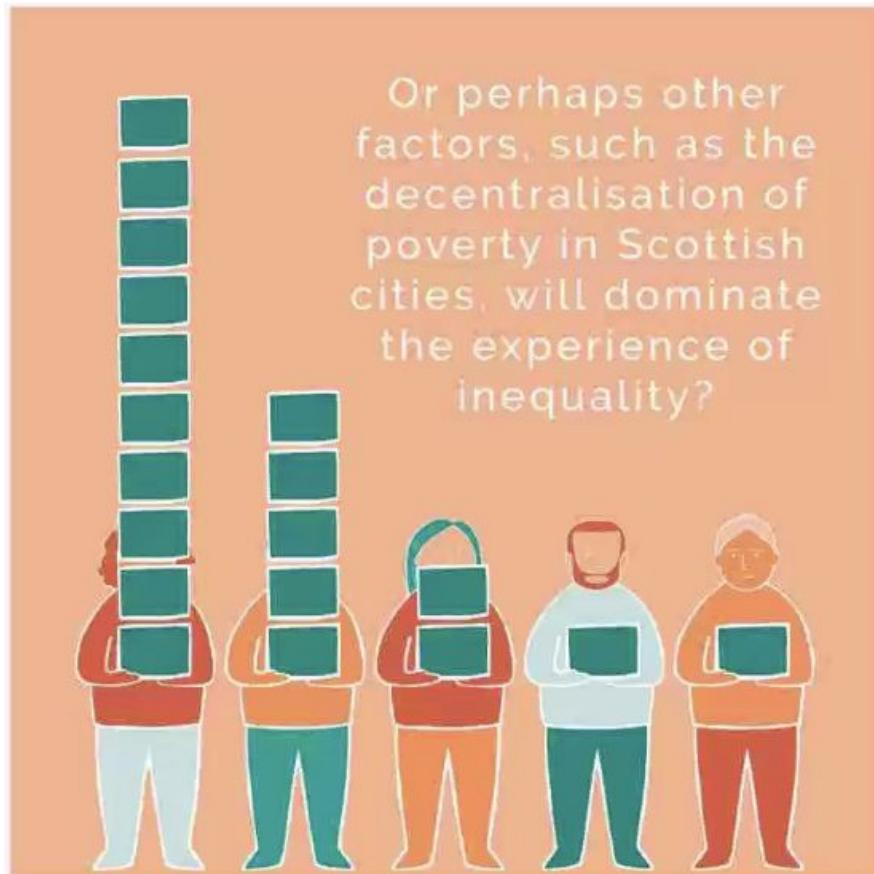
Support for education non-profits has decreased markedly over the past 5 years.

Types of non-profits supported by area funders

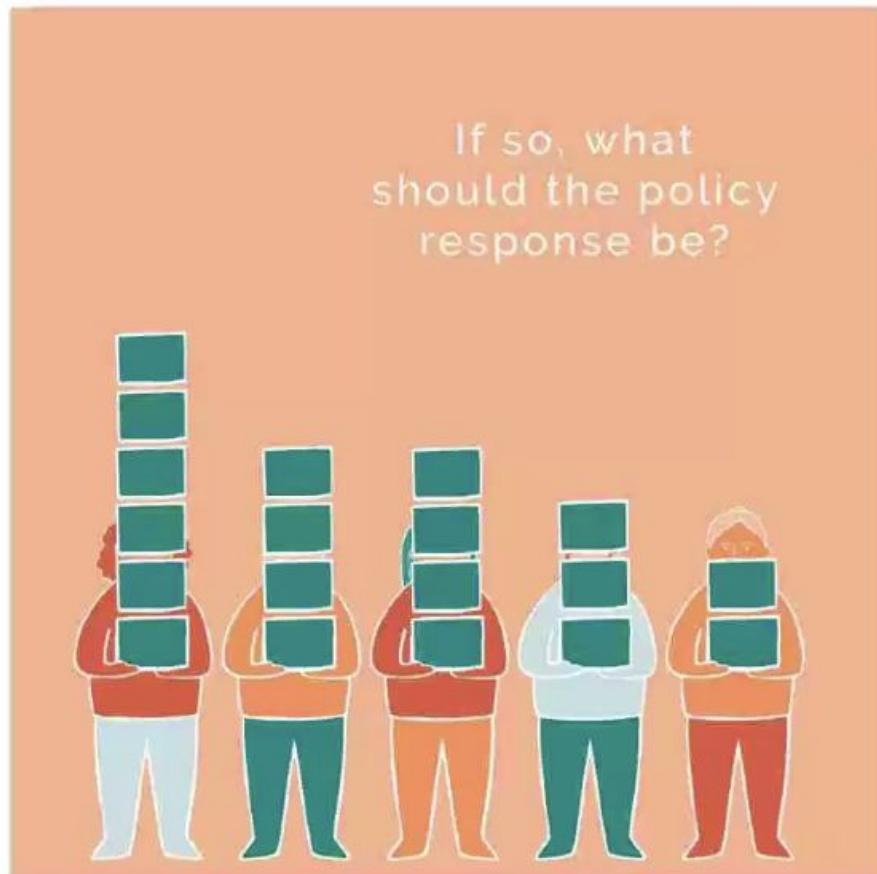


Explanation: Storytelling

Attract attention, drag readers in, communicate message, engage, entertain ...



Or perhaps other factors, such as the decentralisation of poverty in Scottish cities, will dominate the experience of inequality?

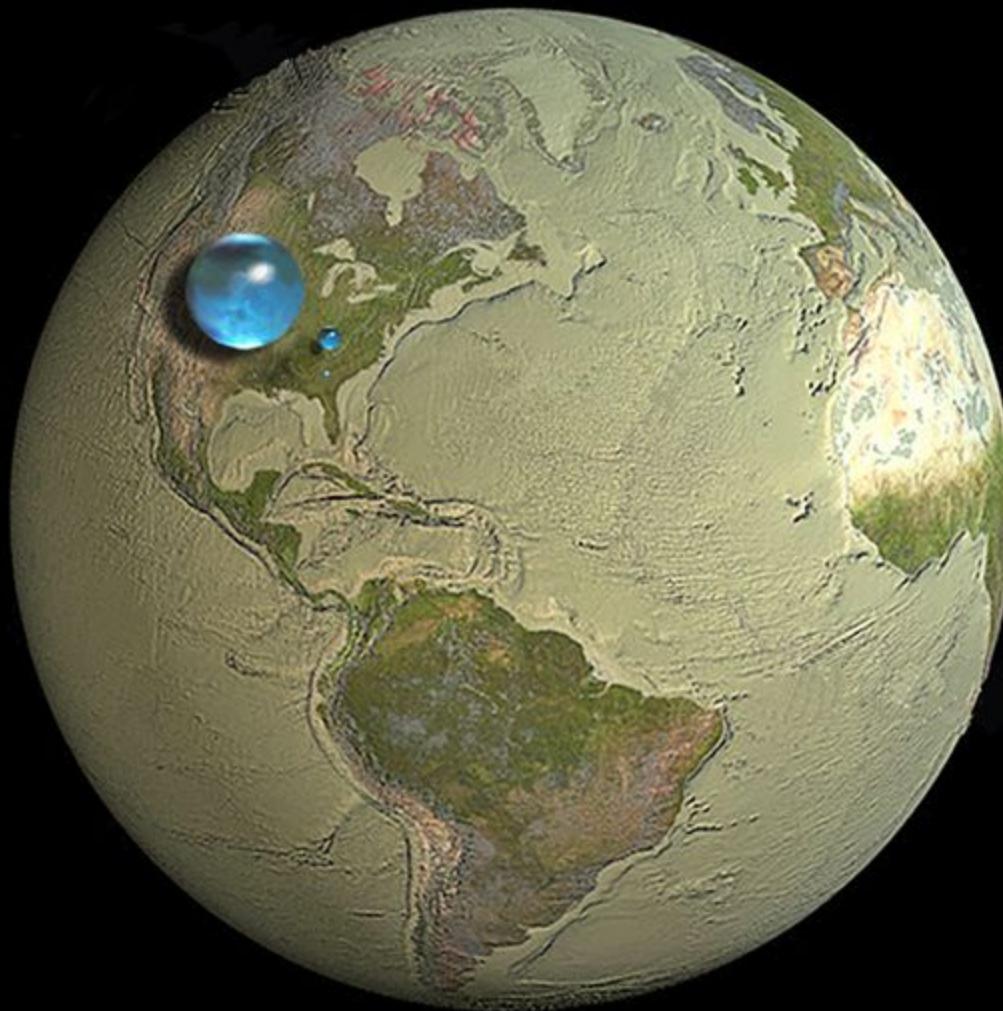


If so, what should the policy response be?

Explanation: Relating



Drop on the planet: visualizing the Earth's water



Water in, on, and above the Earth



- Liquid fresh water
- Freshwater lakes and rivers

Howard Perlman, USGS
Jack Cook, Adam Nieman
Data: Igor Shiklomanov, 1993

Roles of visualization

1. **Discovery**: overview, trends, outliers ...
2. **Data quality**: noisiness, completeness, certainty
3. **Hypotheses**: questions ...
4. **Inform**: analysis, transformations, cleaning, more data ...
5. **Present / explain**,
6. **Persuade / shock**
7. ..

What **is** visualization?

(Possible) Definitions

- "The use of computer-supported, interactive, **visual representations of abstract data to amplify cognition**"—Stu Card
- "... an accessible way to see and understand **trends, outliers, and patterns** in data."—Tableau
- ".. to help people carry out **tasks more effectively**"
— Tamara Munzner

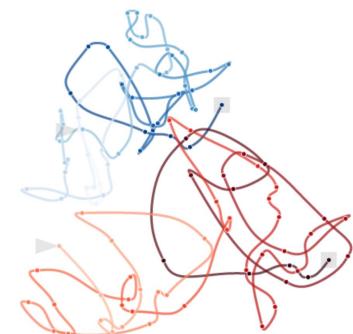
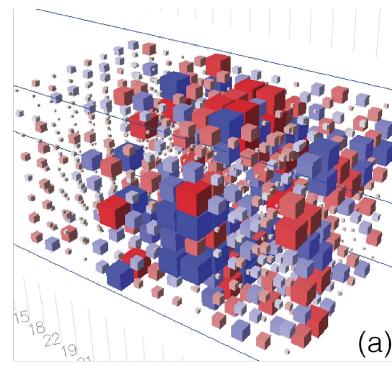
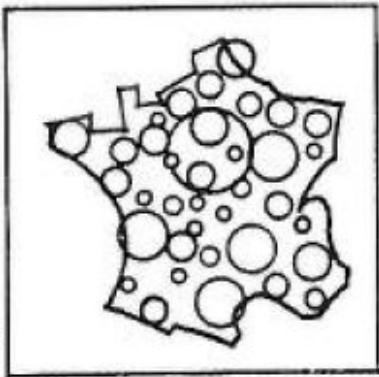
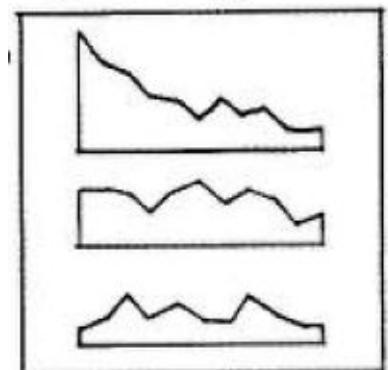
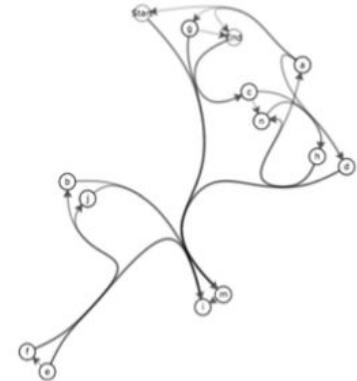
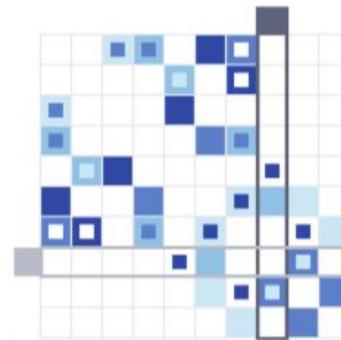
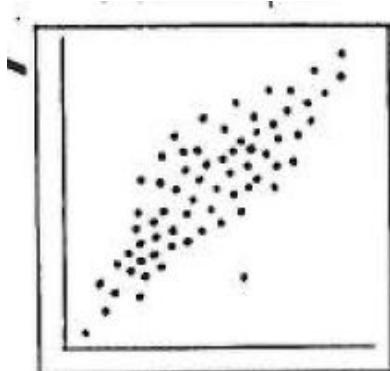
Theory Practise

- Visual perception
 - Design principles
 - Visual variables
 - Data structures
 - Task taxonomies
 - Visualization tax.
 - Presentation
 - Tools
 - ...
- Understanding problems
 - Communicating problems
 - Sketching ideas
 - Creating visuals
 - Seeing
 - Storytelling
 - ...

Common

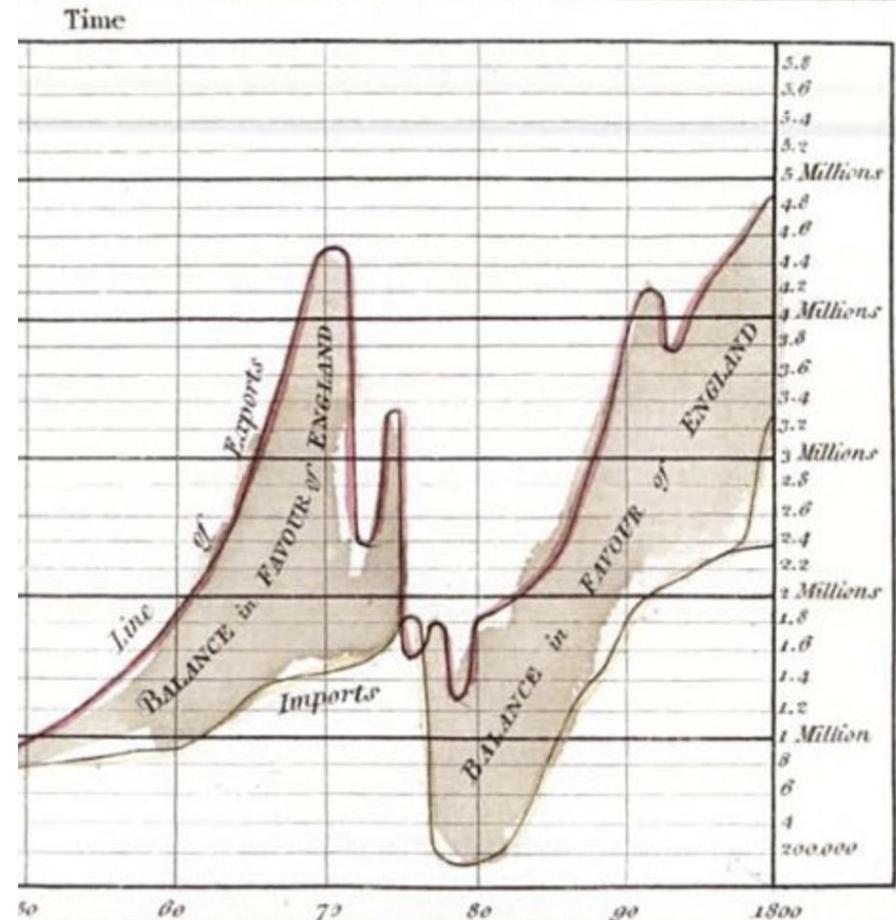


Uncommon



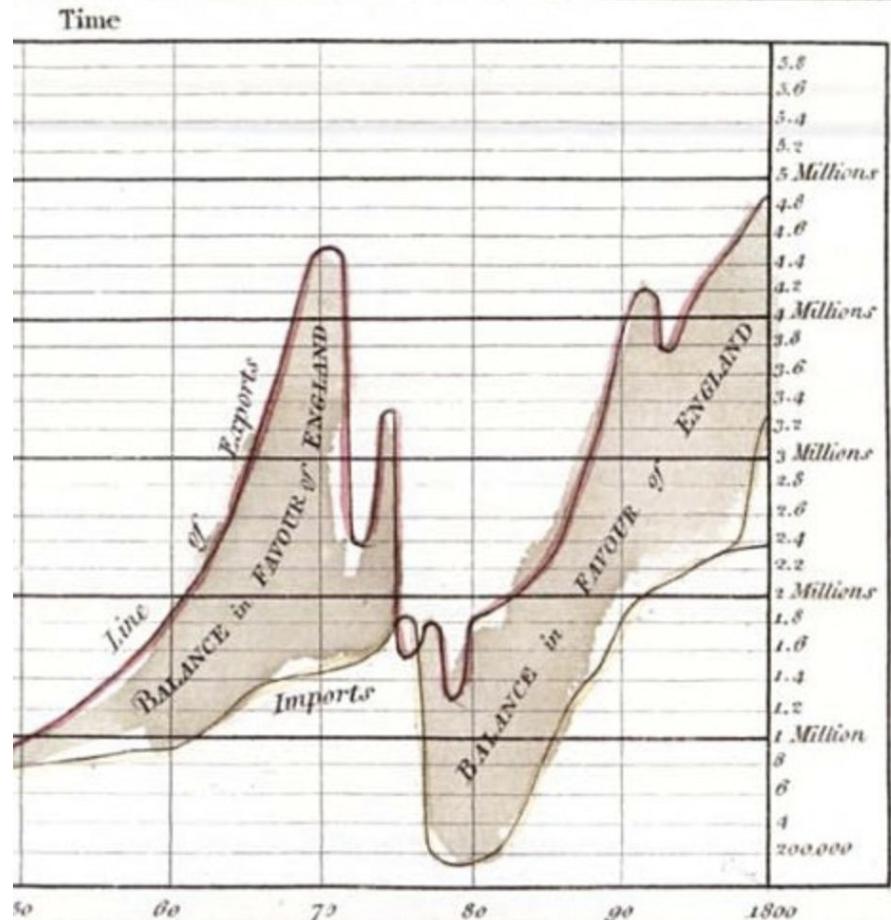
Low level

- Largest / smallest ...
- How many ...
- Which elements ...
- Where ...
- When ...
- What ...
- ...

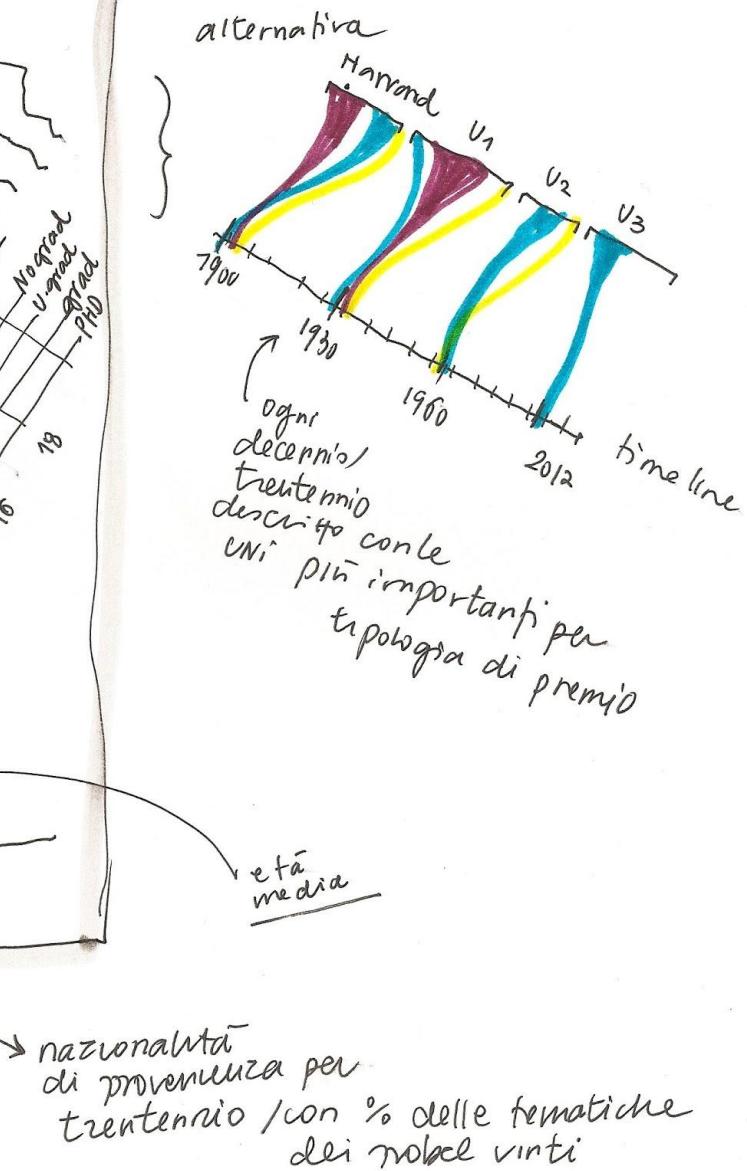


High-level

- Which trends
- Which outliers
- Correlations
- Anomalies
- Describe
- Compare
- Group
- ...

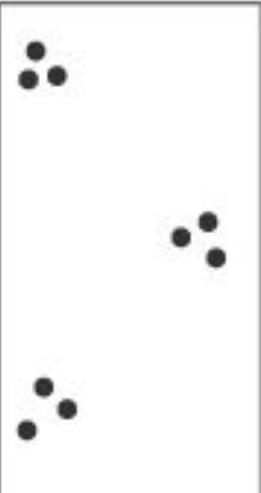


How are we visualizing?



Visual Variables

POSITION



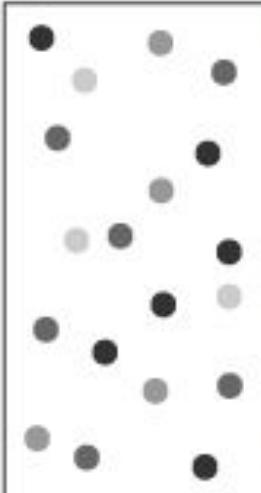
SIZE



SHAPE



VALUE



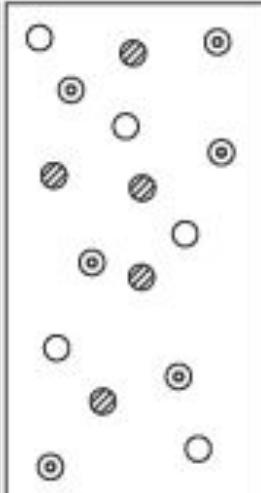
HUE



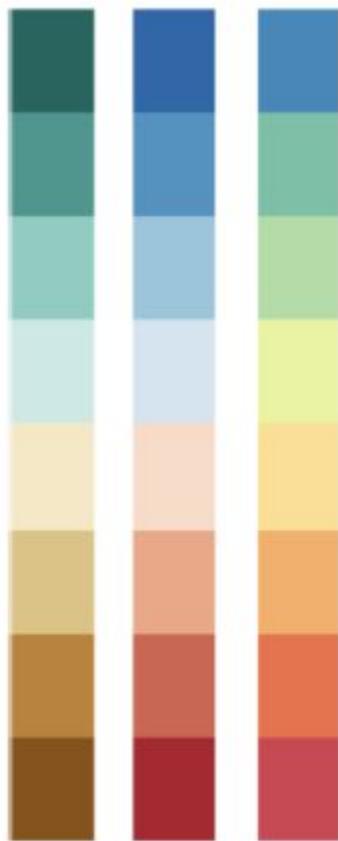
ORIENTATION



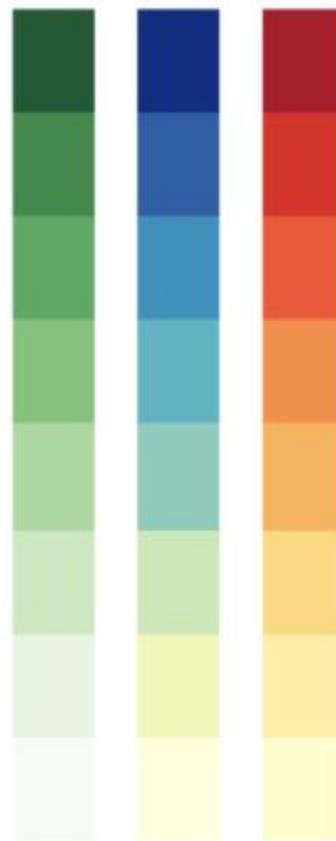
TEXTURE



Color Scales



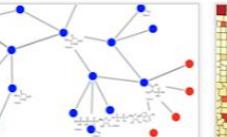
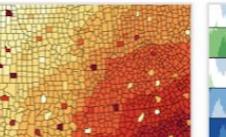
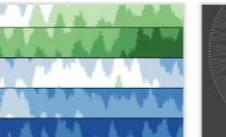
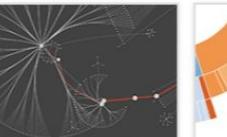
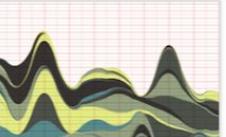
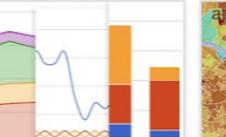
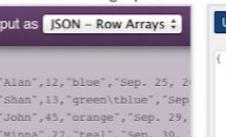
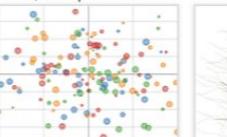
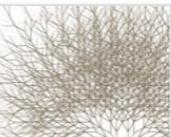
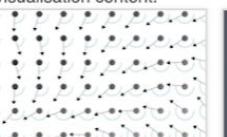
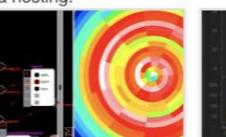
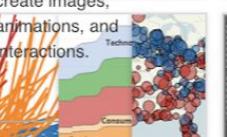
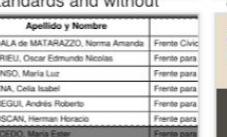
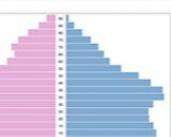
diverging



sequential

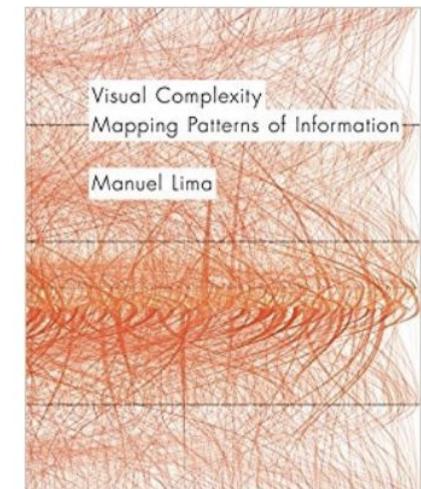
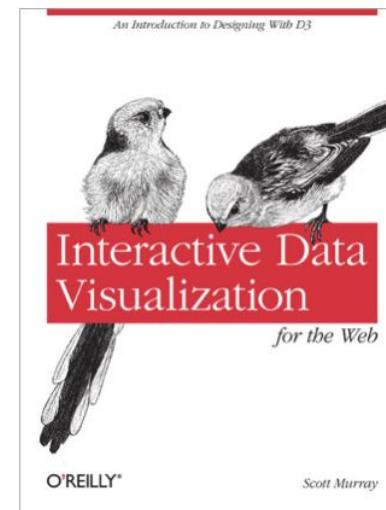
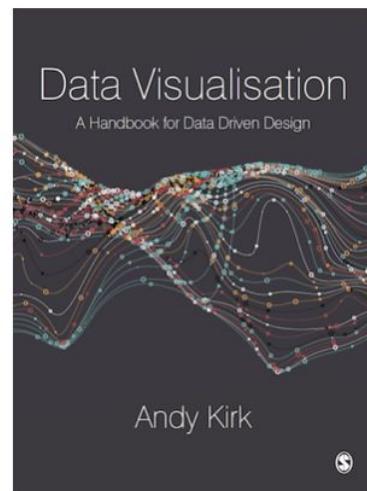
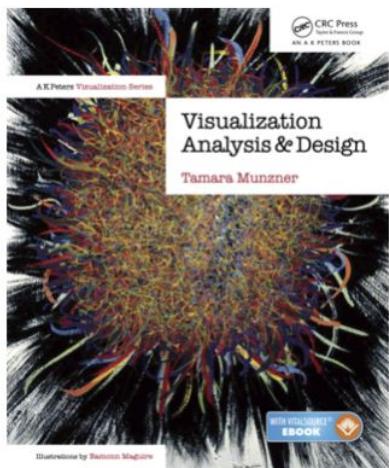
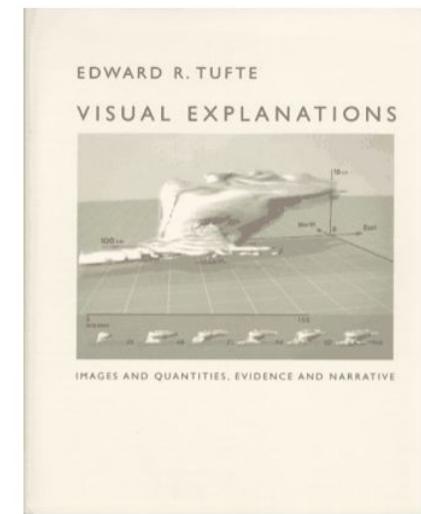
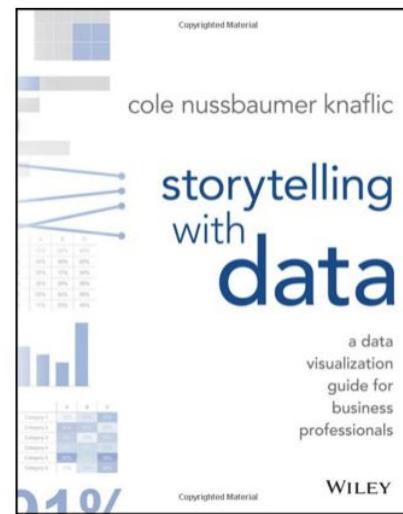
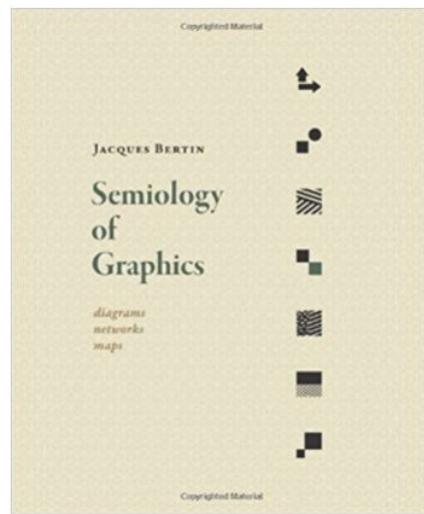
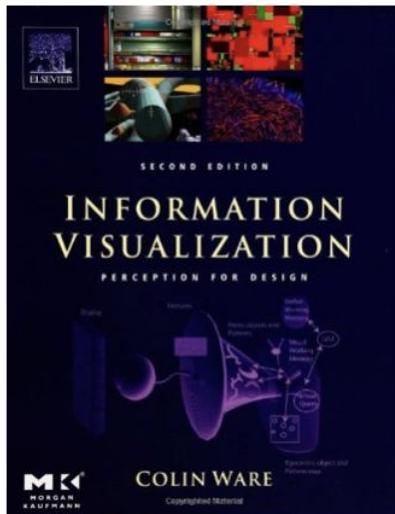


qualitative

 CartoDB A web service for mapping, analyzing and building applications with data.	 Chroma.js Interactive color space explorer that allows to preview a set of linear interpolated equidistant colors.	 Circos A software package for visualizing data in a circular layout.	 Cola.js A library for arranging networks using constraint-based optimization techniques.	 ColorBrewer A web tool for selecting colors for maps.	 Cubism.js A library for creating interactive time series and horizon graphs based on D3.js	 Cytoscape An application for visualizing complex networks and integrating these with any type of interactive documents.	 D3.js An small, flexible and efficient library to create and manipulate interactive documents.
 Degrafa A powerful declarative graphics framework for rich user interfaces, data visualizations and mapping.	 Envision.js A library for creating fast, dynamic and interactive time series visualizations.	 Flare A set of software tools for creating rich interactive data visualizations in ActionScript.	 GeoCommons A public community and set of tools to access, visualize and analyze data with compelling map.	 Gephi A visualization and exploration platform for networks with dynamic and hierarchical graphs.	 Google Chart Tools A collection of simple to use, customizable and free to use interactive charts and data tools.	 Google Fusion Tables A web application that makes it easy to host, manage, collaborate on, visualize, and publish.	 I Want Hue A web application to generate and refine palettes of optimally distinct colors.
 Many Eyes A web application to build, share and discuss graphic representation of user uploaded data.	 MapBox A web platform for hosting custom designed map tiles and a set of open source tools to	 Miso A toolkit to expedite the creation of interactive storytelling and data visualisation content.	 Modest Maps A display and interaction library for tile-based maps in Flash, JavaScript and Python.	 Mr. Data Converter A simple console that converts Excel data into web-friendly formats, including HTML, JSON and XML.	 Mr. Nester A simple console for learning and experimenting with d3.js data nesting.	 NVD3.js A collection of re-usable charts and chart components for d3.js.	 NodeBox A desktop application that lets you create generative, static, animated or interactive
 Polymaps A library for making dynamic, interactive maps with image- and vector-based tiles.	 Prefuse A set of software tools for creating rich interactive data visualizations in Java.	 Processing An open source programming language and environment to create images, animations, and interactions.	 Processing.js The sister project of Processing that makes projects work using web standards and without	 Protovis A library that composes custom views of data with simple marks such as bars and dots.	 Quadrigram A visual programming language aimed to gather, process and visualize information.	 R A software environment for statistical computing and graphical techniques.	 Raphaël A small library that simplifies working with vector graphics on the web.
 SVG Crowbar A bookmarklet that extracts SVG nodes from	 Sigma.js An open-source lightweight library to	 Tableau Public A desktop application to build and post interactive	 Tabula A tool to extract CSV formatted data from text	 Tangle A library that allow to interactively explore,	 Timeline.js A tool to create timelines with data and media from	 Unfolding A library to create interactive maps and	 Vega A visualization grammar, a declarative format for

```
106 //on initial load we are going to get workbook parameters and read the two field names from them
107 getCurrentWorkbook().getParametersAsync().then(function(parms) {
108     // to check on whether parms is loaded correctly
109     //console.log(parms);
110
111     //loop through array and set the xAxis and yAxis field names
112     for(var i =0; i < parms.length; i++)
113     {
114         if(parms[i].getName() === 'xAxisField')
115         {
116             xAxisField = parms[i].getCurrentValue().value;
117         }
118         else if(parms[i].getName() === 'yAxisField')
119         {
120             yAxisField = parms[i].getCurrentValue().value;
121         }
122     }
123
124     //on initial load get data and store it // we will need to recall this when the underlying data changes
125     getCurrentWorksheet().getSummaryDataAsync({
126         maxRows: 0,
127         ignoreSelection: true,
128         includeAllColumns: true,
129         ignoreAliases: true
130     }).then(onDataLoadOk, onDataLoadError);
```

General Readings



<https://visualinteractivedata.github.io/resources.html>