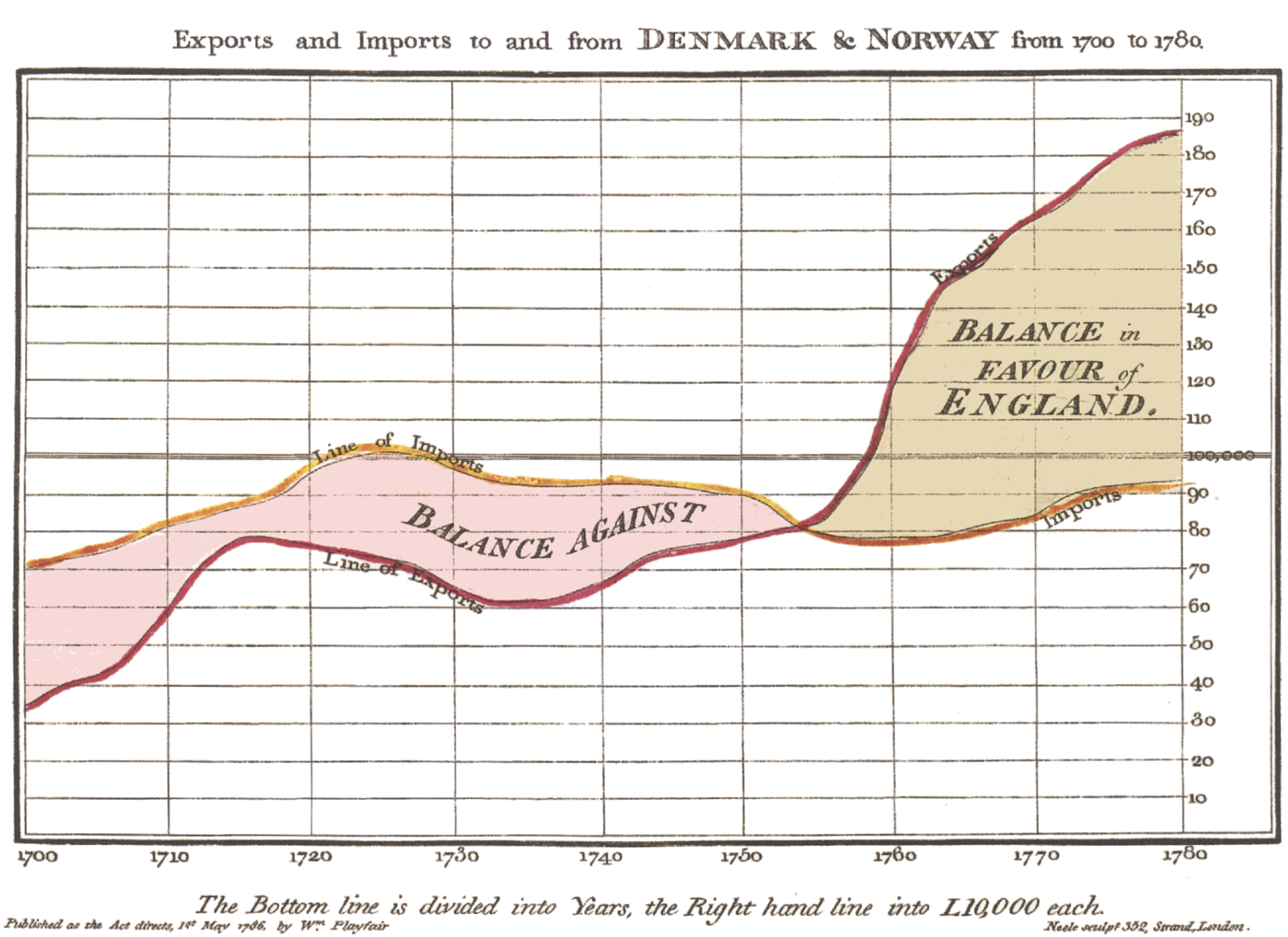
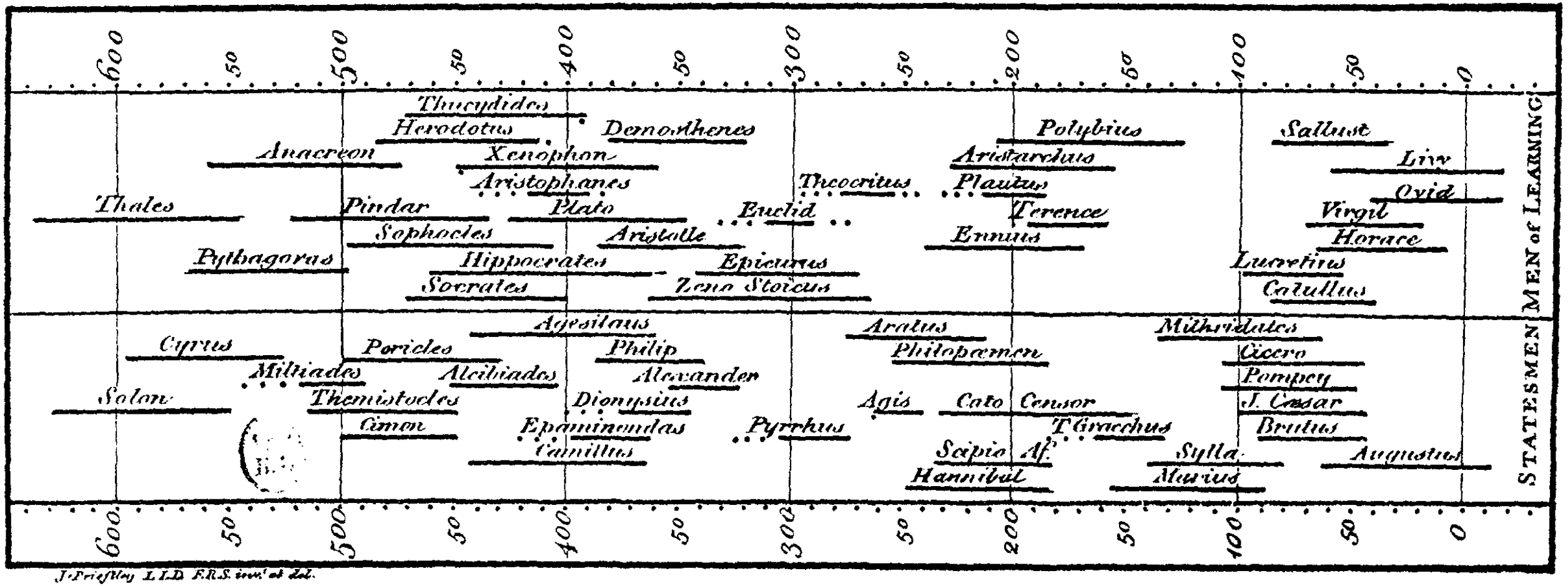
VISUALISATION

Interaction with Machine Learning

Cambridge MPhil ACS 2017-2018

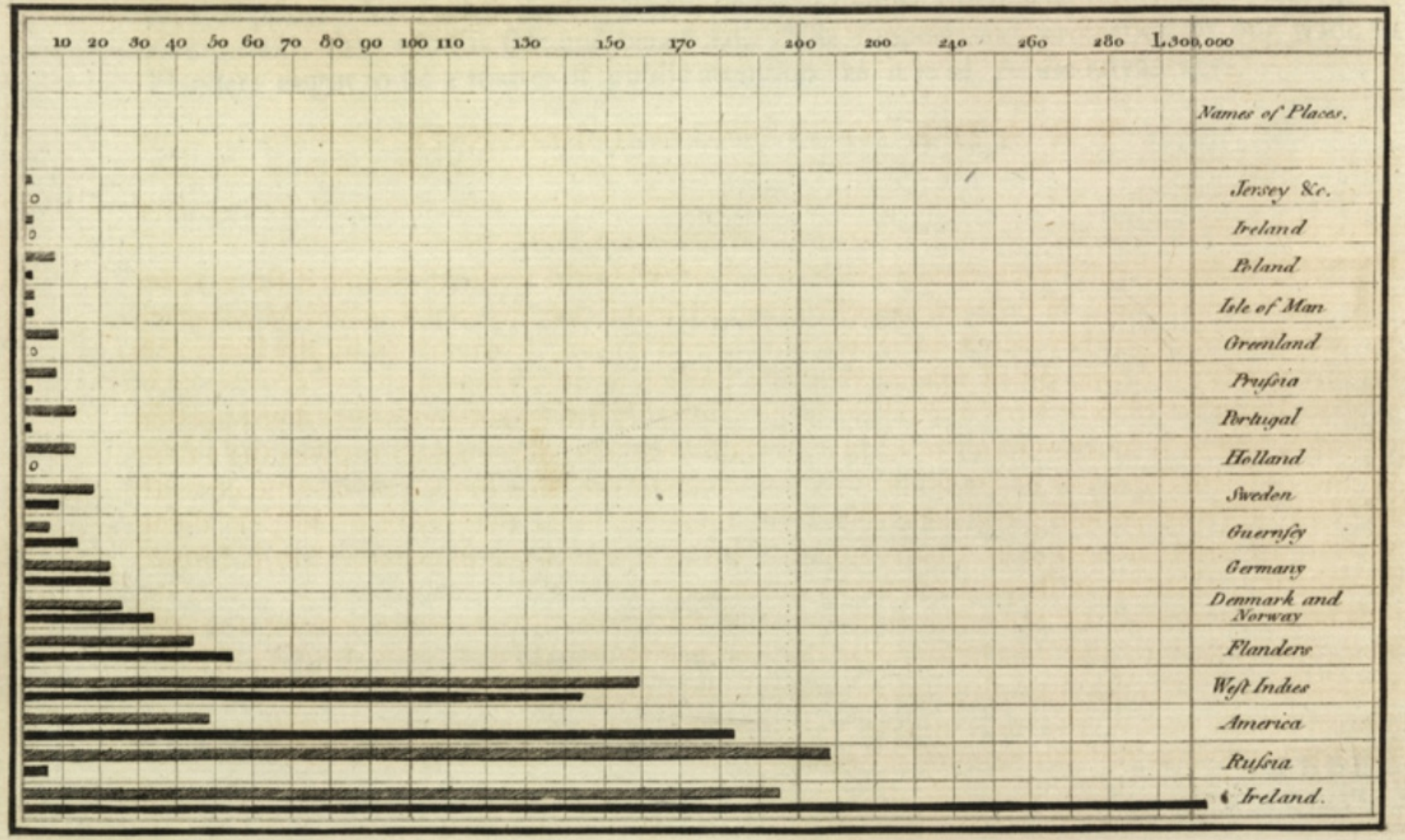


Time Series area chart
William Playfair, c1786



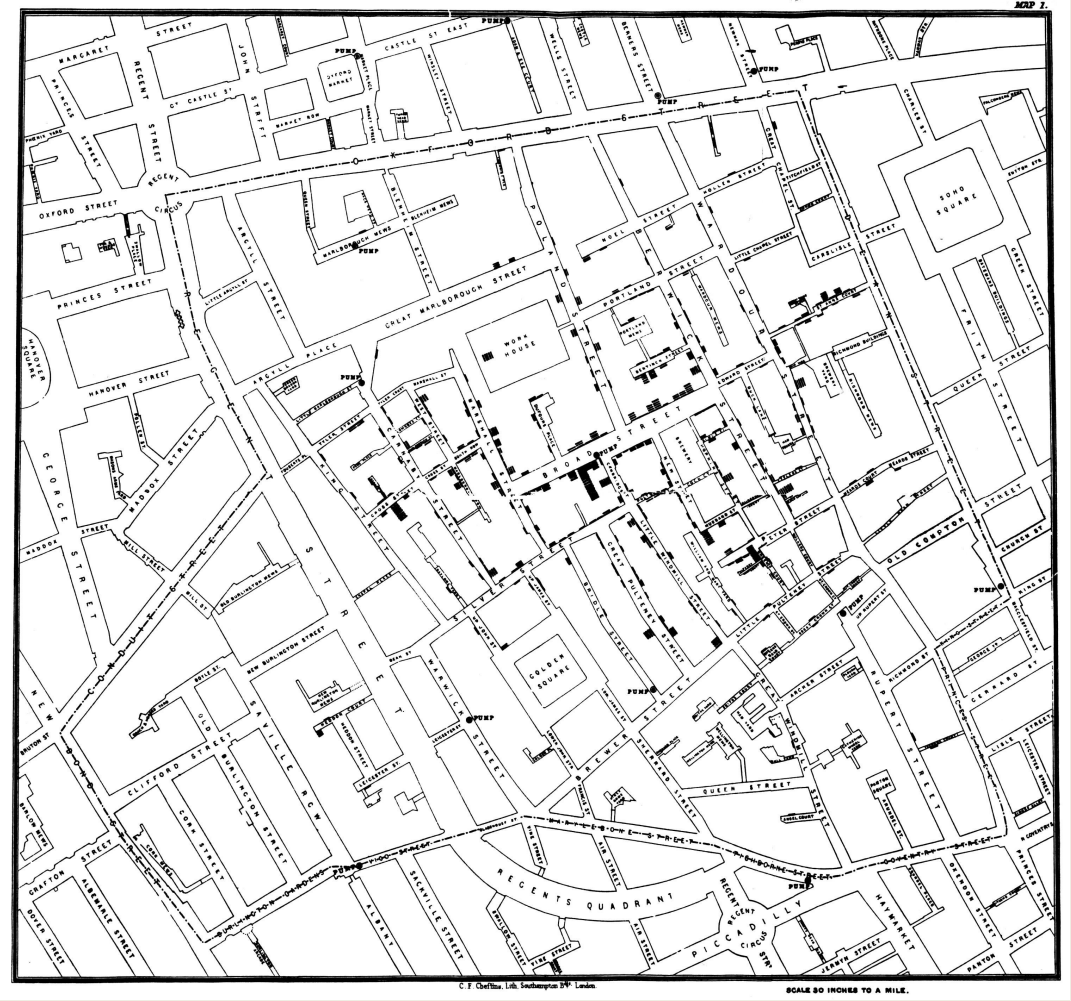
Lifespan chart
Joseph Priestly, c1765

Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.



The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed lines Imports.

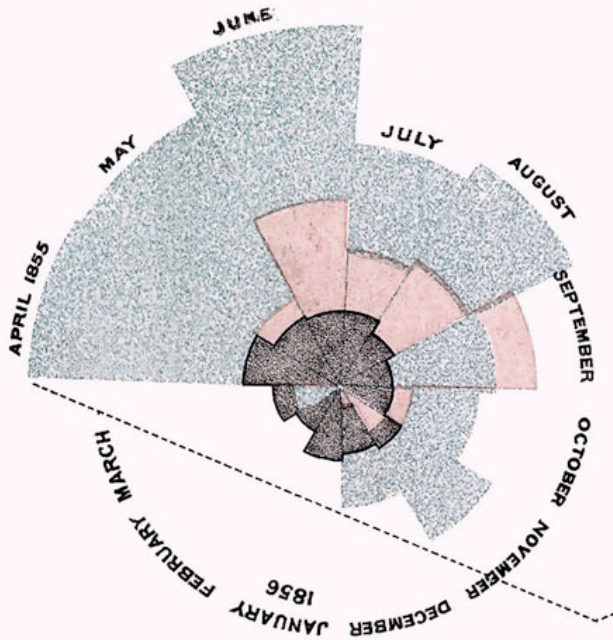
Bar chart
William Playfair, c1786



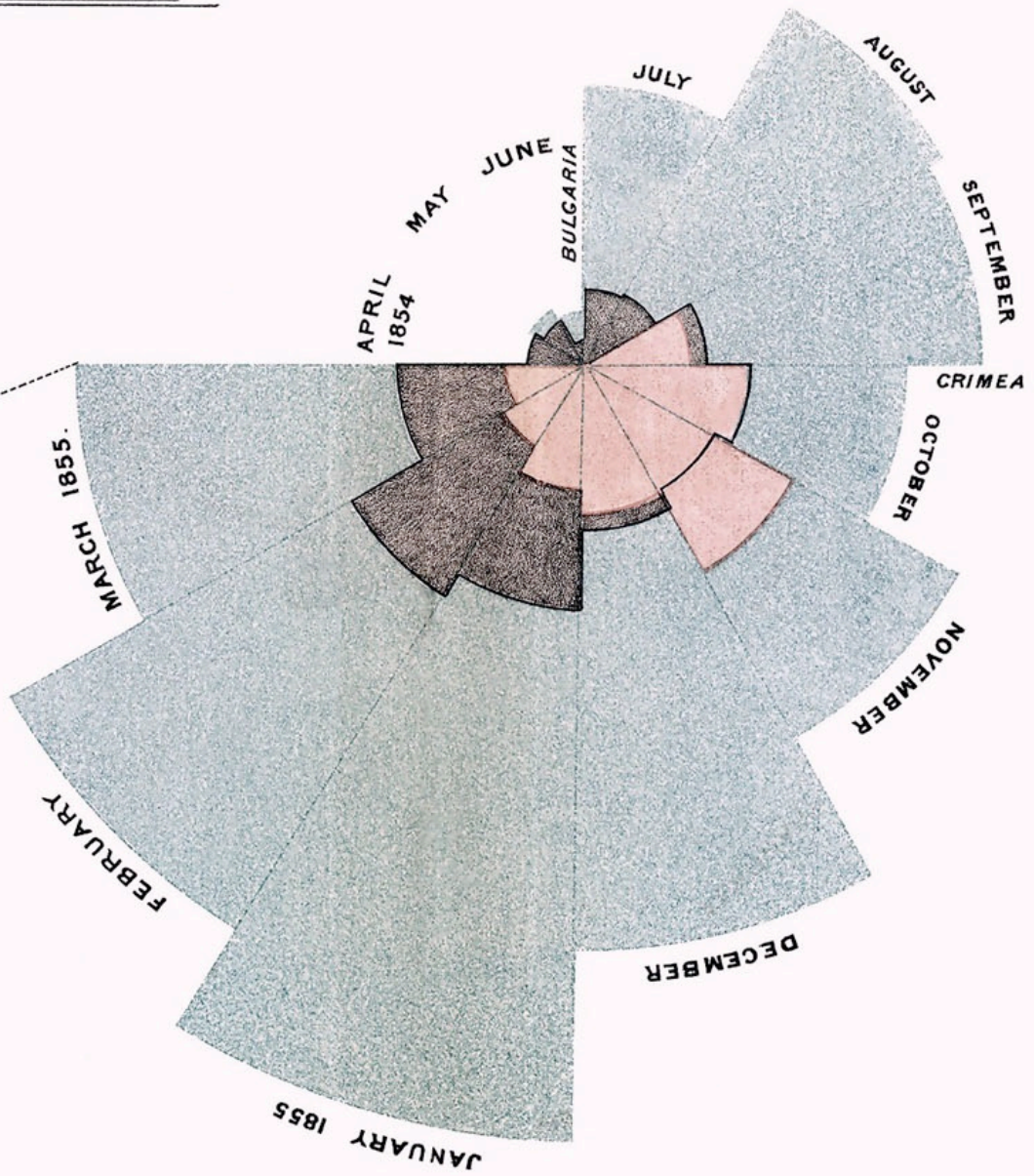
Cholera map
John Snow, 1854

DIAGRAM of the CAUSES of MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.

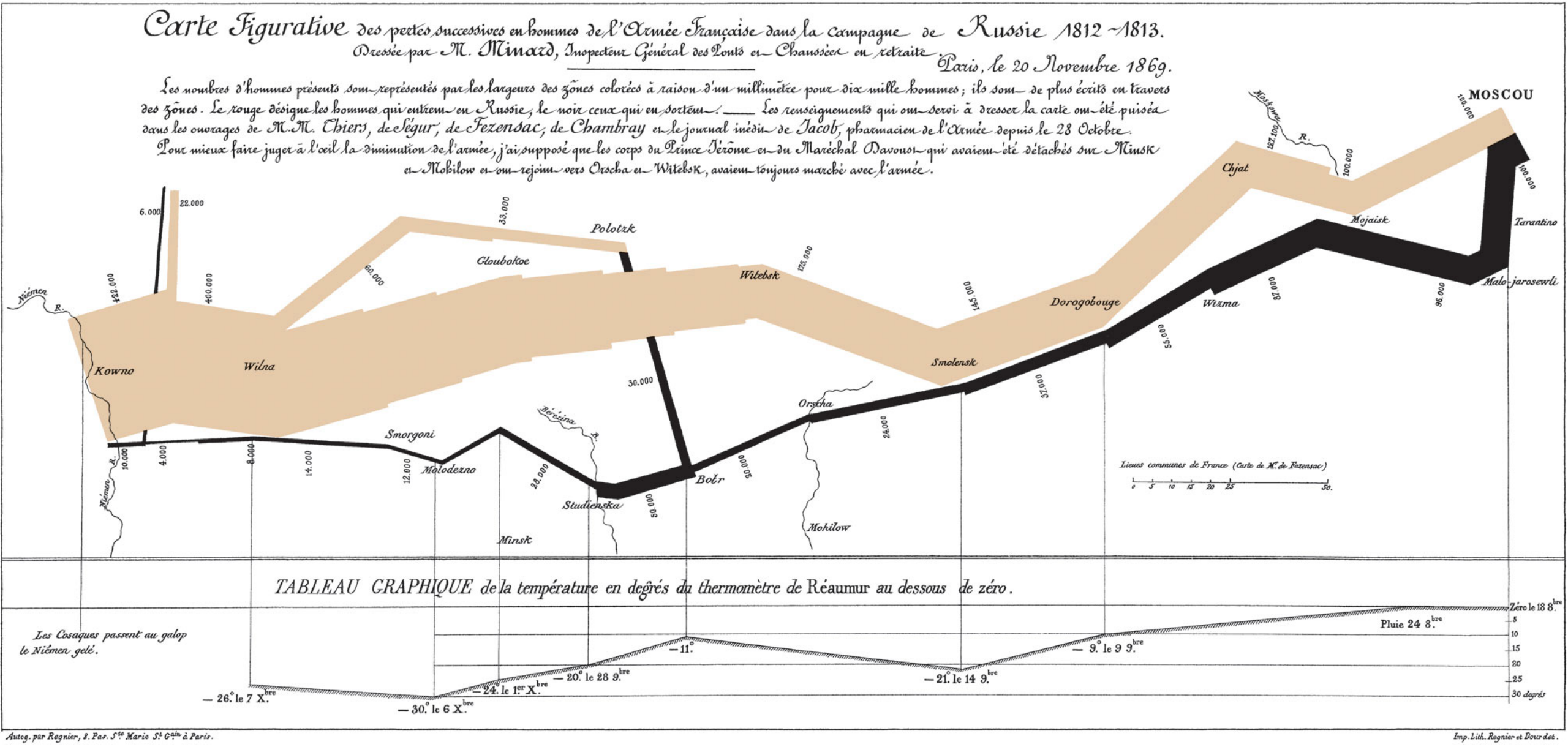


1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Rose diagrams
Florence Nightingale, c1858

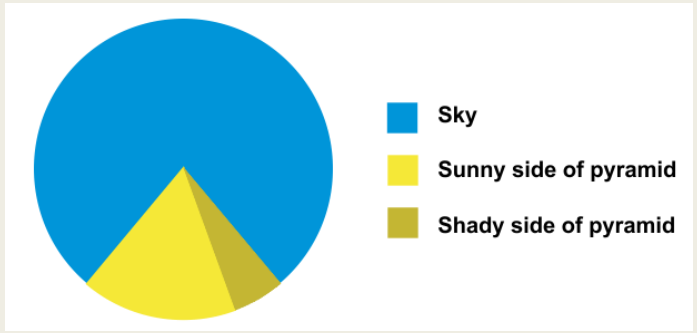


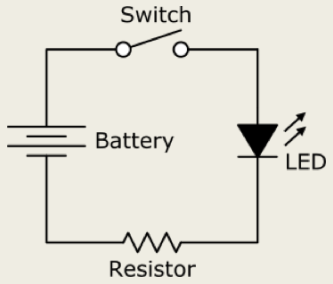
Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

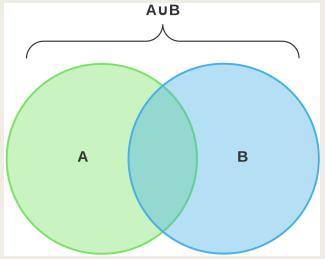
Drawn by Charles Minard, Inspector General of Bridges and Roads (retired). Paris, November 20, 1869.

What is visualisation?

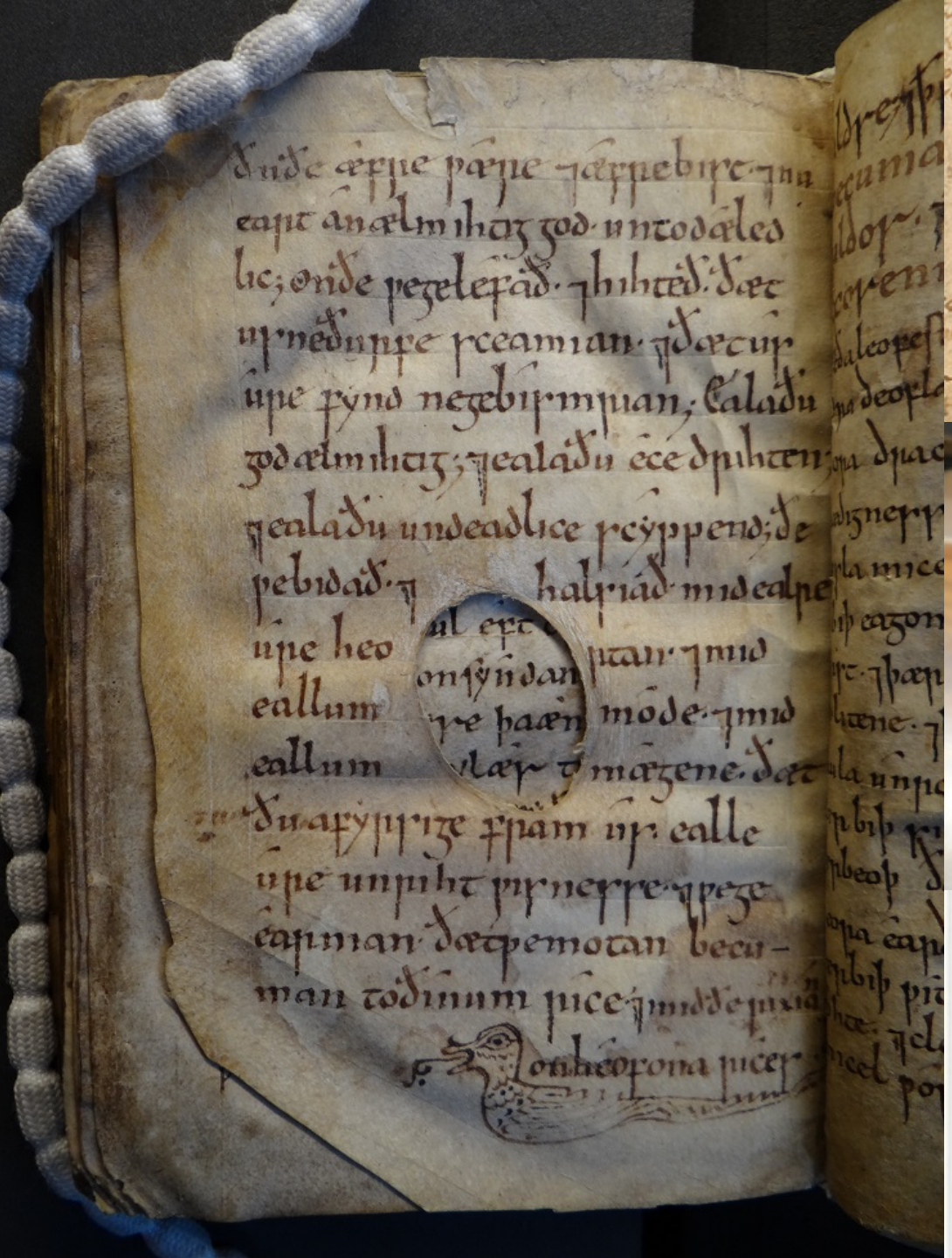
- Charts & statistical visualisations
- Typography & typesetting
- Diagrams
- Illustrations and drawings
- Infographics
- Symbols
- Marks



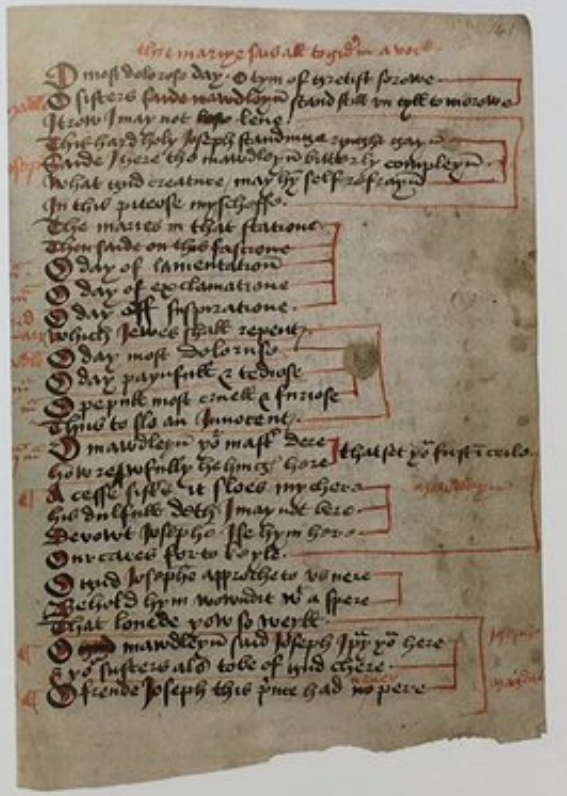





Bodleian Libraries
UNIVERSITY OF OXFORD
WESTON LIBRARY
1 December 2017 – 22 April 2018 ADMISSION FREE
DESIGNING
English
Graphics on the Medieval Page

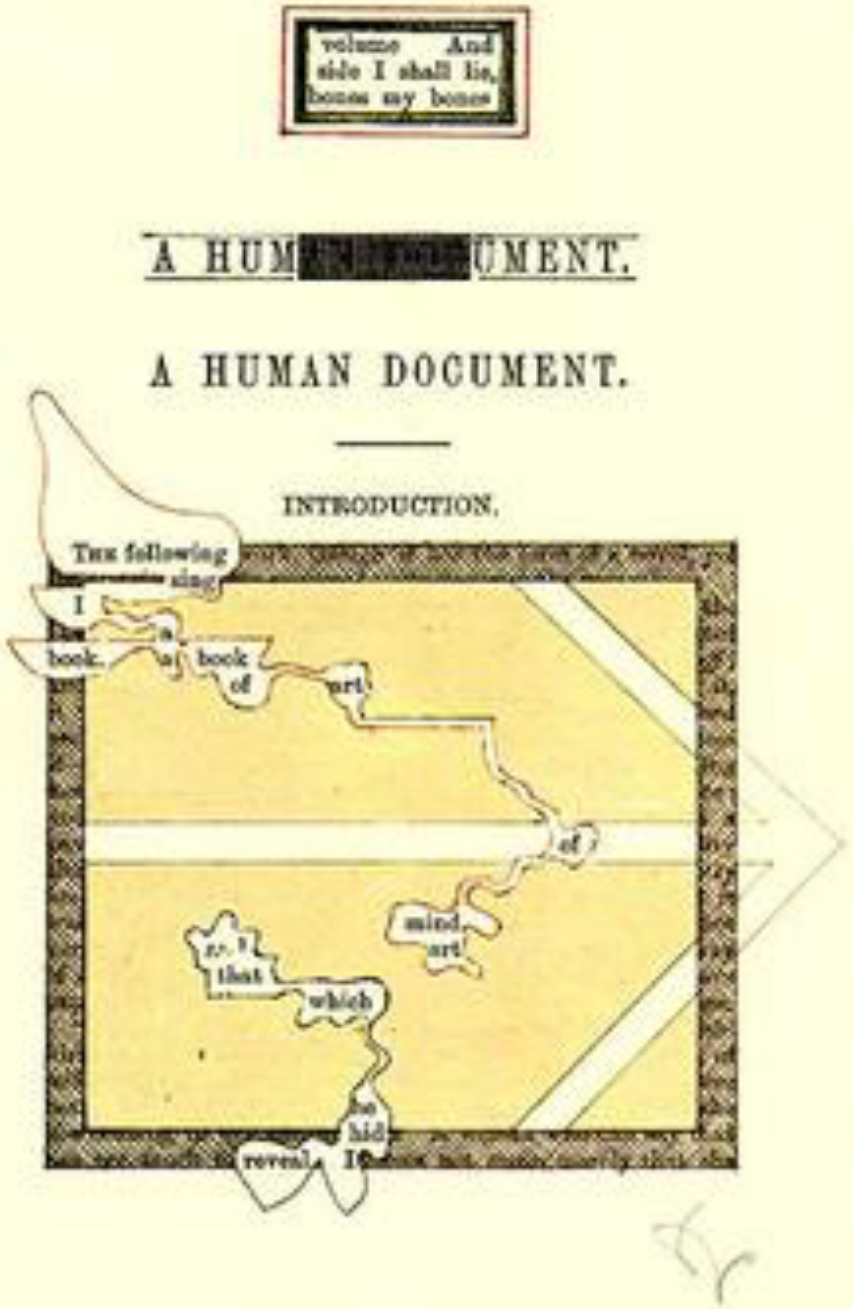


he adds further empty phrases to fill out the verse.[74] And he redesigns the layout, adding speech prefixes in red. After thirteen pages, he seems to have realized what he wanted, and from there on the text is set out as a play only.[75] The first few pages of revision, though, show him consciously redesigning a book with the conventions of drama. He was able to revise his layout to guide people in voicing and performing words.



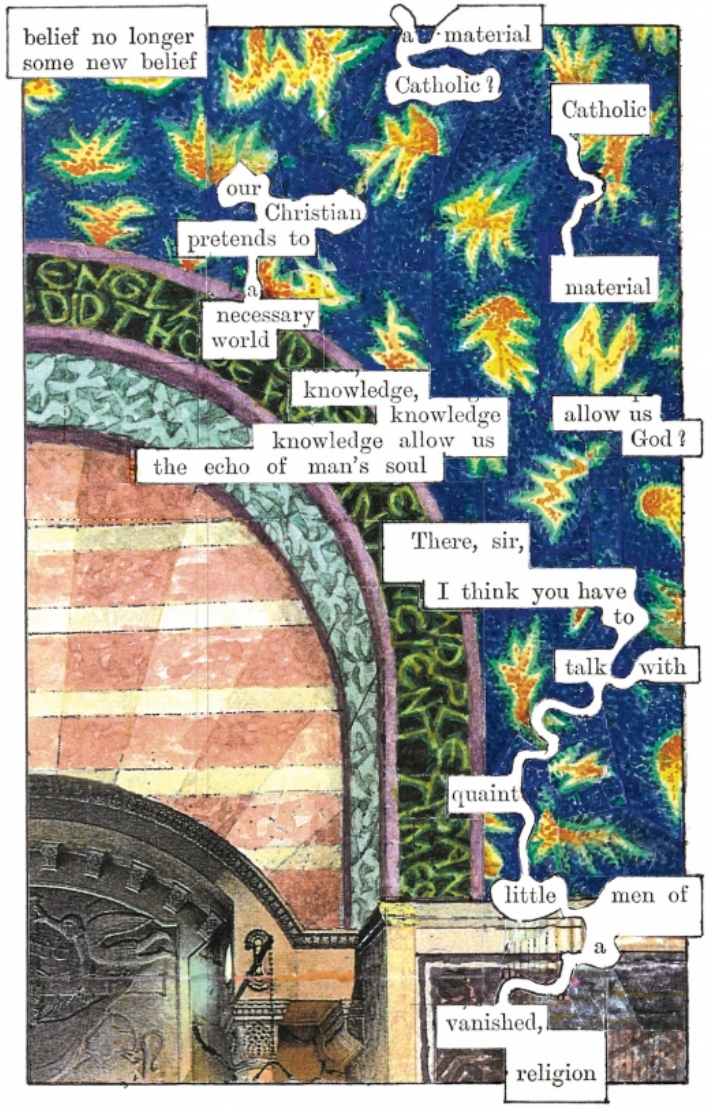
73. 'First entereth Wisdom' –Elaborate red stage directions but visual braces on the verse dialogue in a late fifteenth-century copy of a morality play *Wisdom*. MS. Digby 133, fol. 158r.

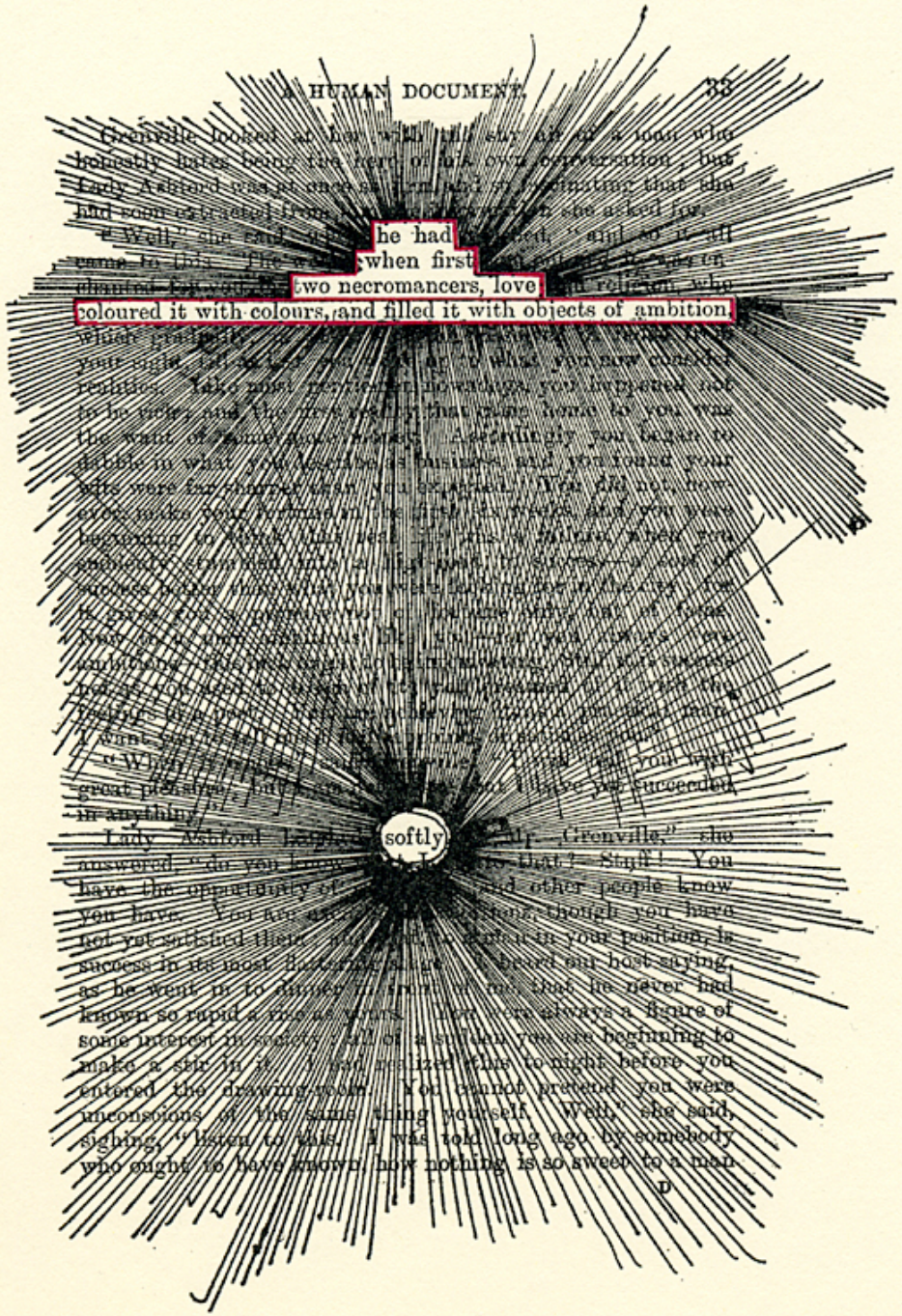
74. 'Certain lines which should not be said if it be played' – A scribe working c.1518–20 uses red ink to add speech prefixes and rewrite the dialogue, thereby turning a poem into a play. *The Burial of Christ*. MS. e Musaeo 160, fol. 141r.

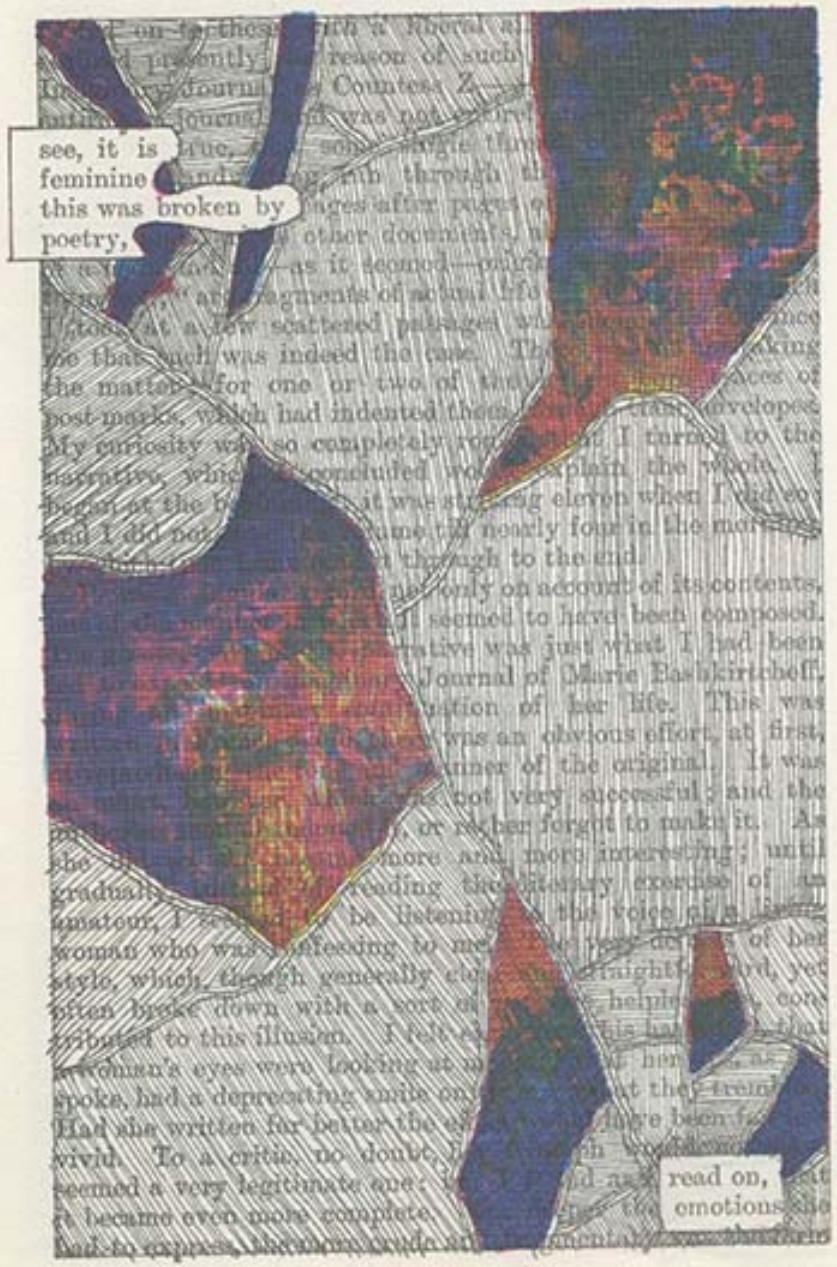


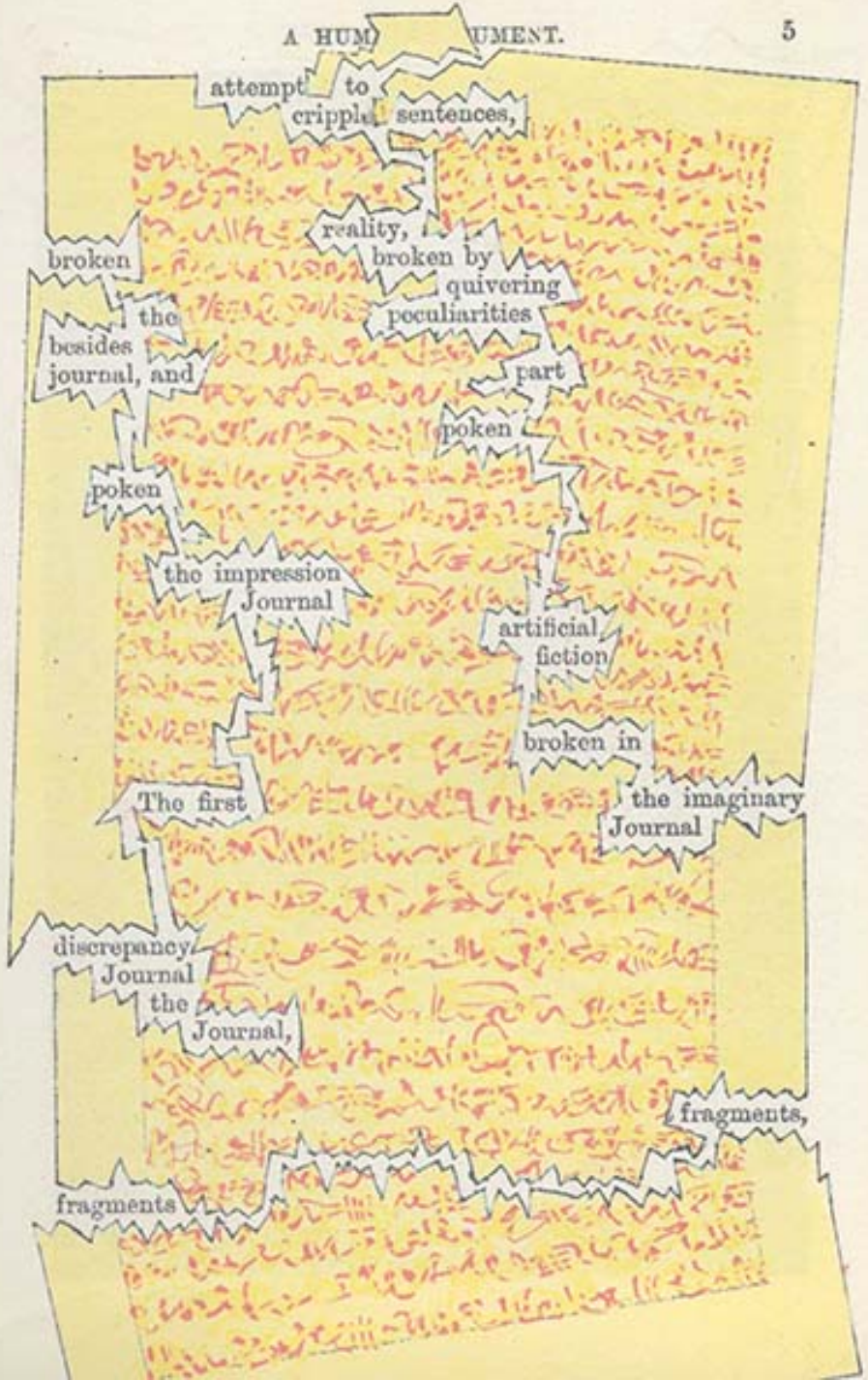
A Humument
Tom Phillips, 1960s

“It is a forgotten Victorian novel found by chance … plundered, mined, and undermined its text to make it yield the ghosts of other possible stories, scenes, poems and replaced the text [he'd] stripped away with visual images of all kinds.”

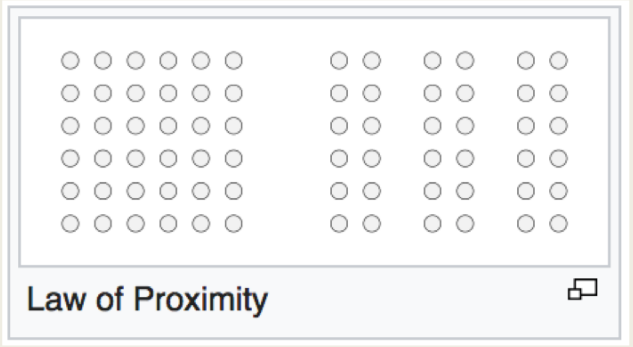

belief no longer
some new belief
material
Catholic?
Catholic
our
Christian
pretends to
material
a
necessary
world
knowledge,
knowledge
allow us
God?
knowledge allow us
the echo of man's soul
There, sir,
I think you have
to
talk
with
quaint
little
men of
a
vanished,
religion


he had
when first
two necromancers, love
coloured it with colours, and filled it with objects of ambition,
softly


see, it is feminine
this was broken by poetry,
read on, emotions

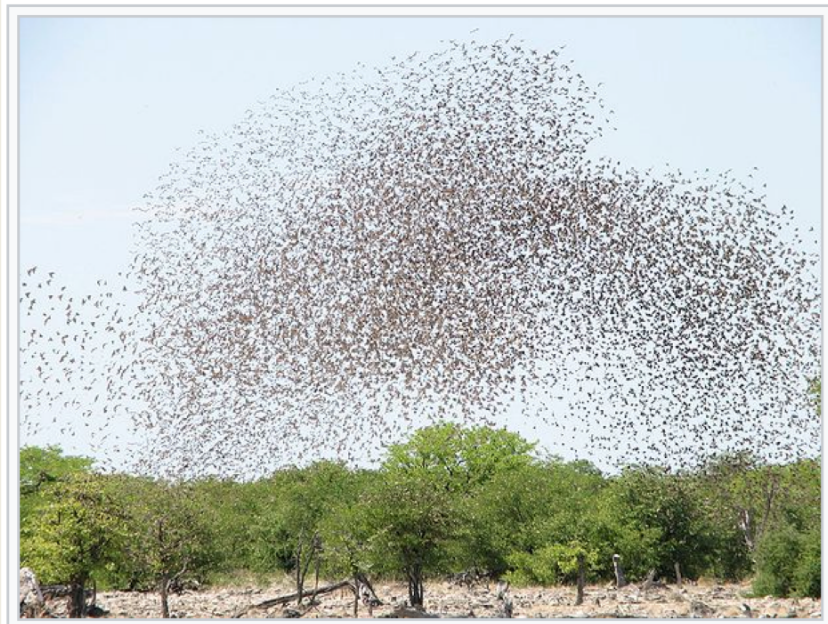

attempt to cripple sentences,
reality, broken by quivering peculiarities
broken
the
besides journal, and
part
poken
poken
the impression Journal
artificial fiction
broken in
The first
the imaginary Journal
discrepancy Journal
the Journal,
fragments,
fragments

Gestalt principles of visual perception

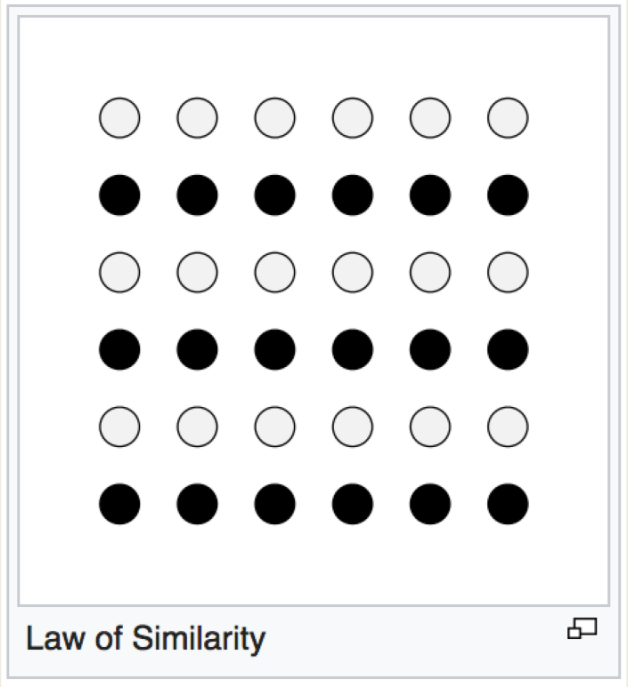


Law of Proximity

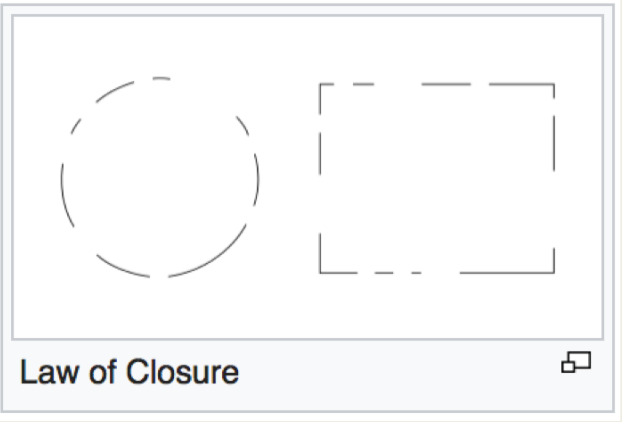
Law of good continuation



Law of common fate—a flock of birds.



Law of Similarity



Law of Closure

	Graphic Resources	Correspondence	Design Uses
Marks	Shape Orientation Size Texture Saturation Colour Line	Literal (visual imitation of physical features) Mapping (quantity, relative scale) Conventional (arbitrary)	Mark position, identify category (shape, texture colour) Indicate direction (orientation, line) Express magnitude (saturation, size, length) Simple symbols and colour codes
Symbols	Geometric elements Letter forms Logos and icons Picture elements Connective elements	Topological (linking) Depictive (pictorial conventions) Figurative (metonym, visual puns) Connotative (professional and cultural association) Acquired (specialist literacies)	Texts and symbolic calculi Diagram elements Branding Visual rhetoric Definition of regions
Regions	Alignment grids Borders and frames Area fills White space Gestalt integration	Containment Separation Framing (composition, photography) Layering	Identifying shared membership Segregating or nesting multiple surface conventions in panels Accommodating labels, captions or legends
Surfaces	The plane Material object on which marks are imposed (paper, stone) Mounting, orientation and display context Display medium	Literal (map) Euclidean (scale and angle) Metrical (quantitative axes) Juxtaposed or ordered (regions, catalogues) Image-schematic Embodied/situated	Typographic layouts Graphs and charts Relational diagrams Visual interfaces Secondary notations Signs and displays

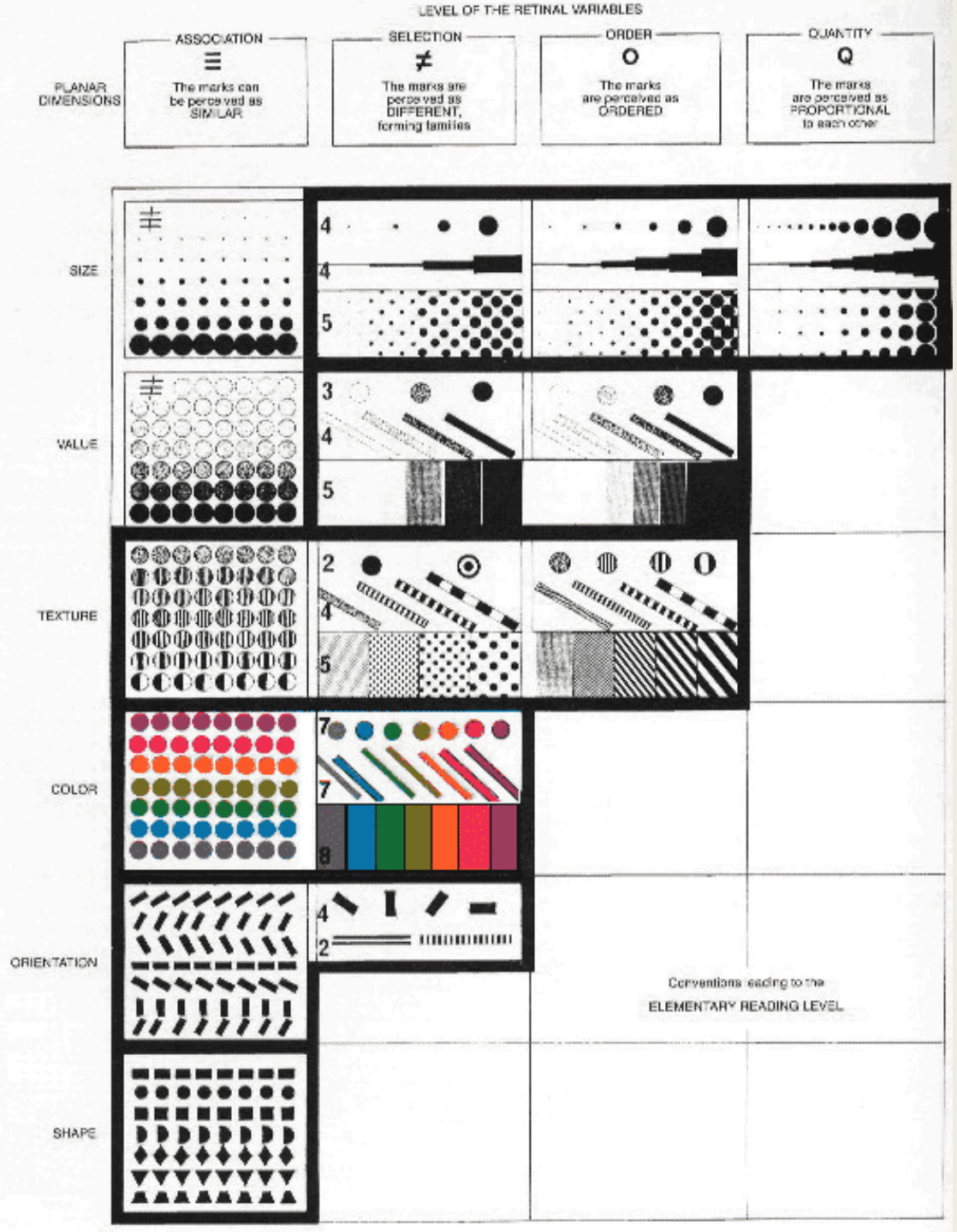
Bertin, J. (1967). Semiologie graphique. Paris: Editions Gauthier-Villars. English translation by WJ. Berg (1983)as Semiology of graphics, Madison, WI: University of Wisconsin Press

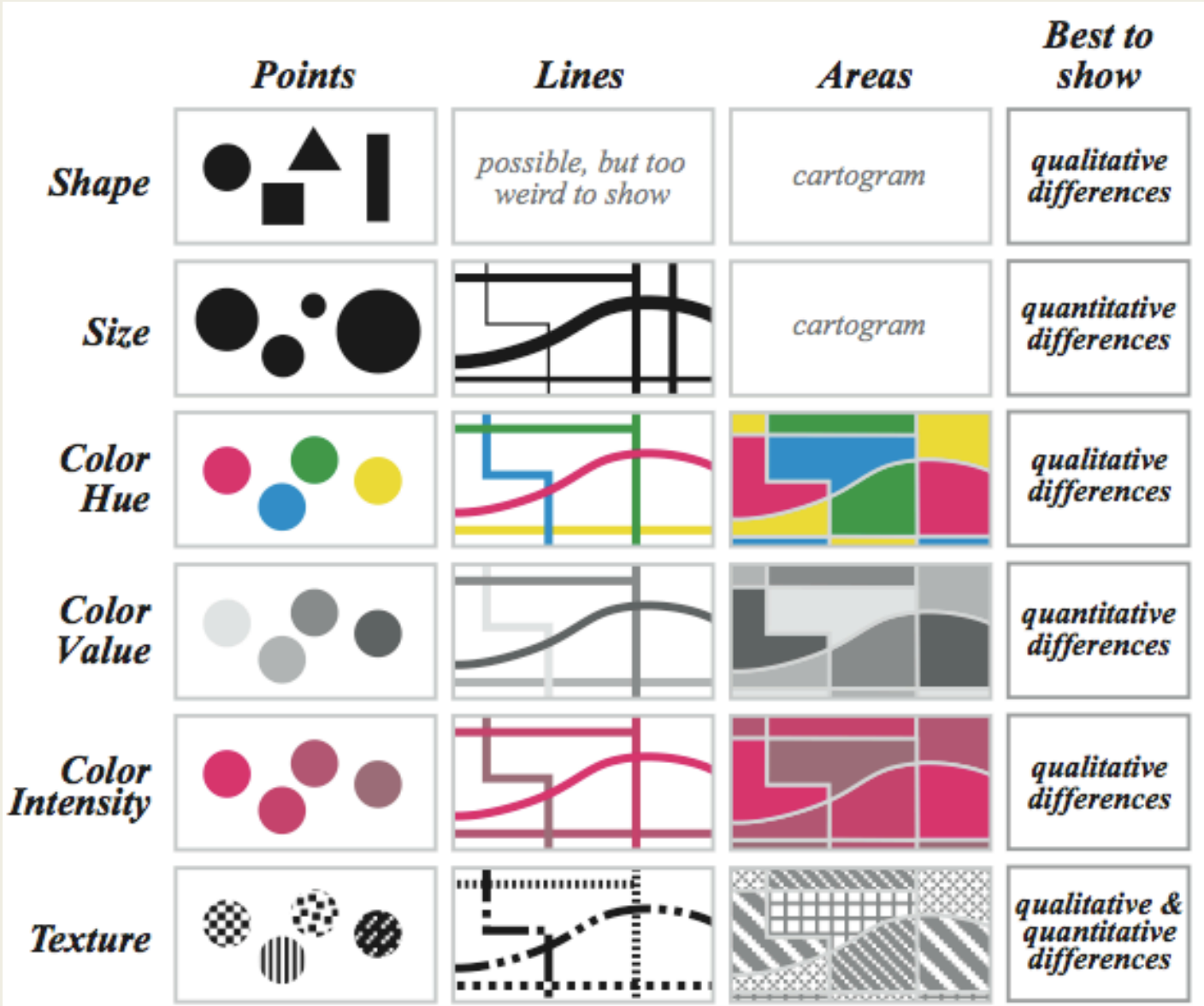
Blackwell, A.F. and Engelhardt, Y. (2002). A meta-taxonomy for diagram research. In M. Anderson&B. Meyer&P. Olivier (Eds.), Diagrammatic Representation and Reasoning, London: Springer-Verlag, pp. 47-64.

Engelhardt, Y. (2002). The Language of Graphics. A framework for the analysis of syntax and meaning in maps,charts and diagrams. PhD Thesis, University of Amsterdam.

MacEachren, A.M. (1995). How maps work: Representation, visualization, and design. Guilford.

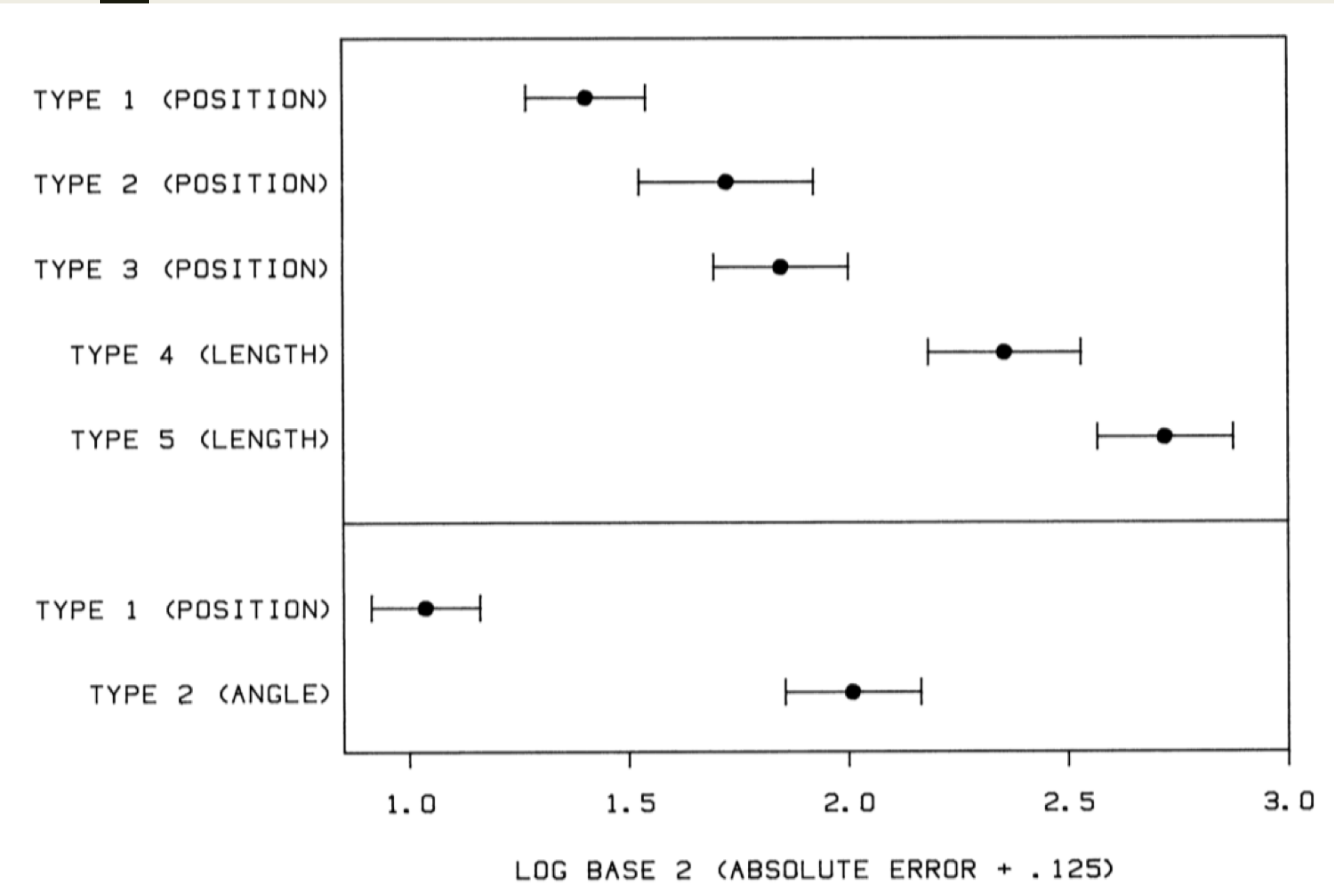
Jacques Bertin, Semiology of Graphics, 1967





Cleveland, W. S.,&McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.*Journal of the American Statistical Association*,*79*(387), 531–554. https://doi.org/10.2307/2288400

Heer, J.,&Bostock, M. (2010). Crowdsourcing graphical perception: using {Mechanical Turk} to assess visualisation design.*ACM Human Factors in Computing Systems (CHI)*, 203–212.



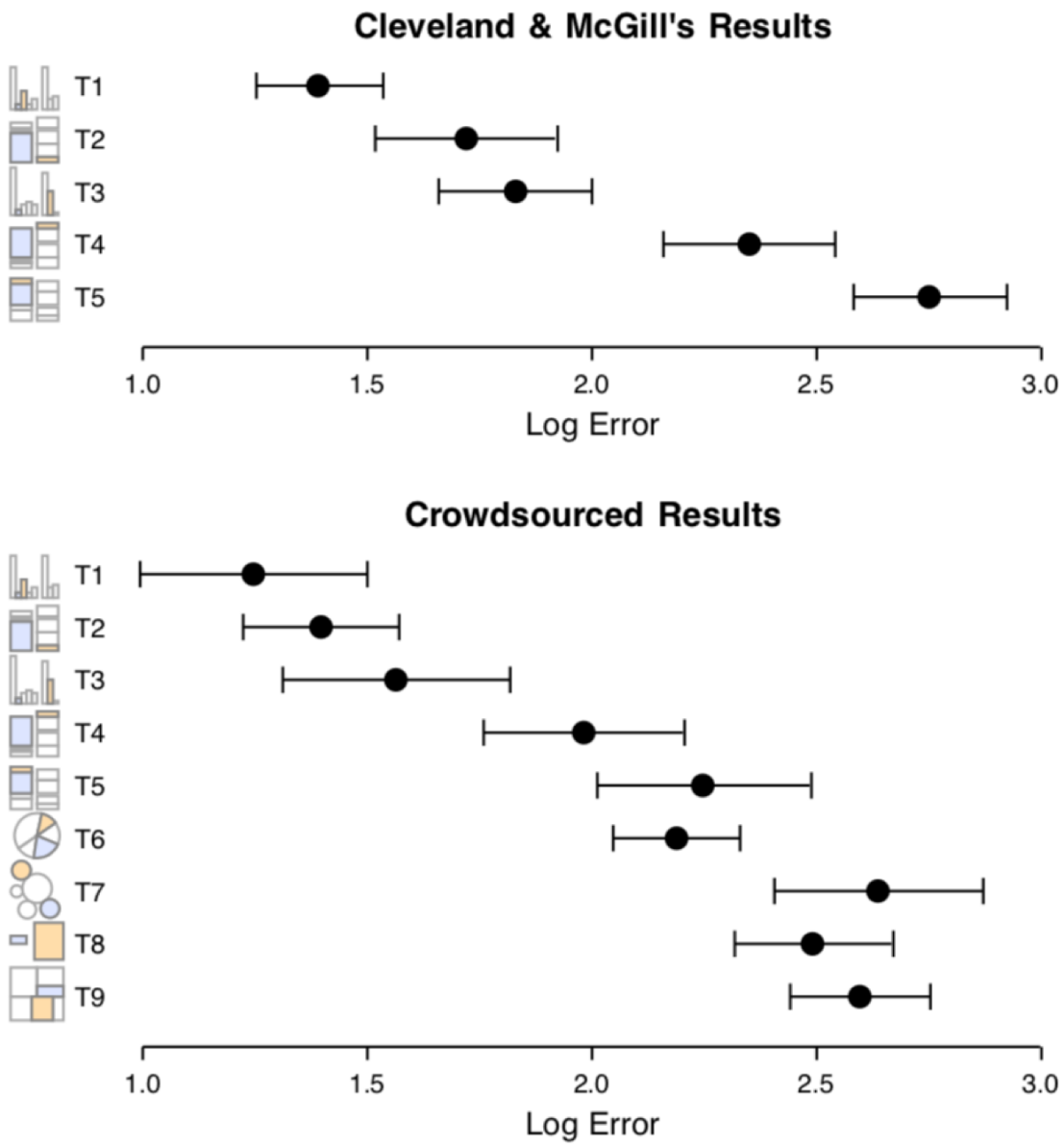


Figure 4: Proportional judgment results (Exp. 1A & B). Top: Cleveland & McGill's [7] lab study. Bottom: MTurk studies. Error bars indicate 95% confidence intervals.

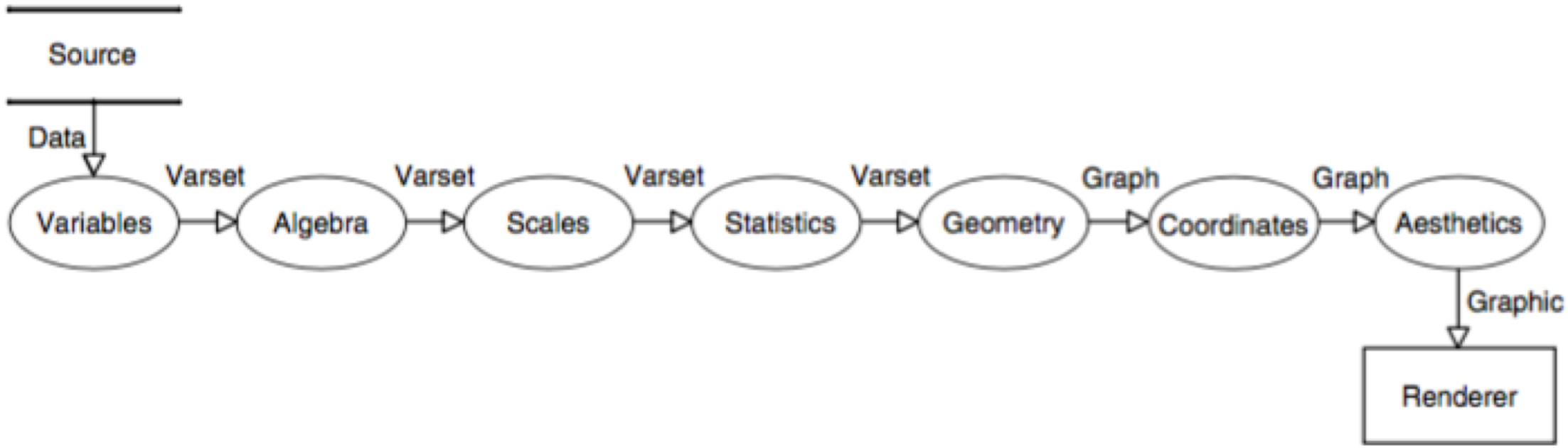
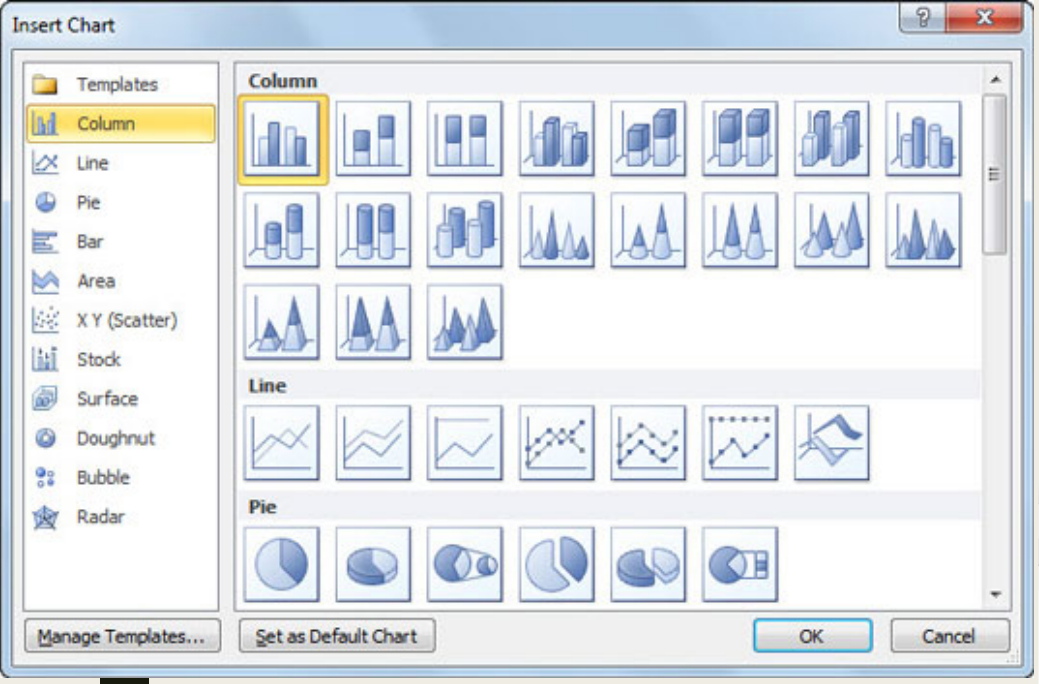


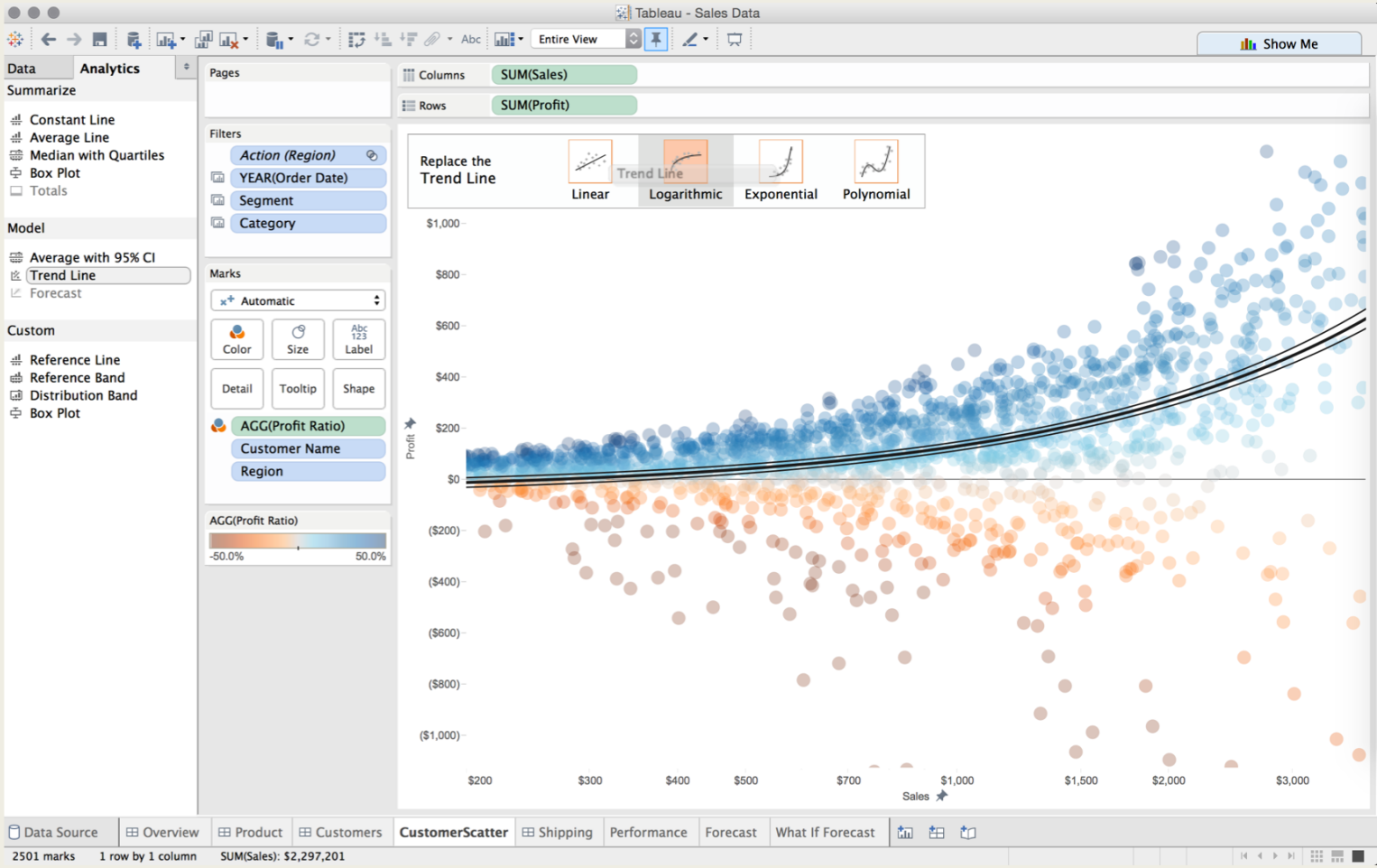
FIGURE 1 | The grammar of graphics data flow.

Leland Wilkinson, The Grammar of Graphics, 1999
Later extended by Hadley Wickham



Excel chart picker

Tableau chart designer



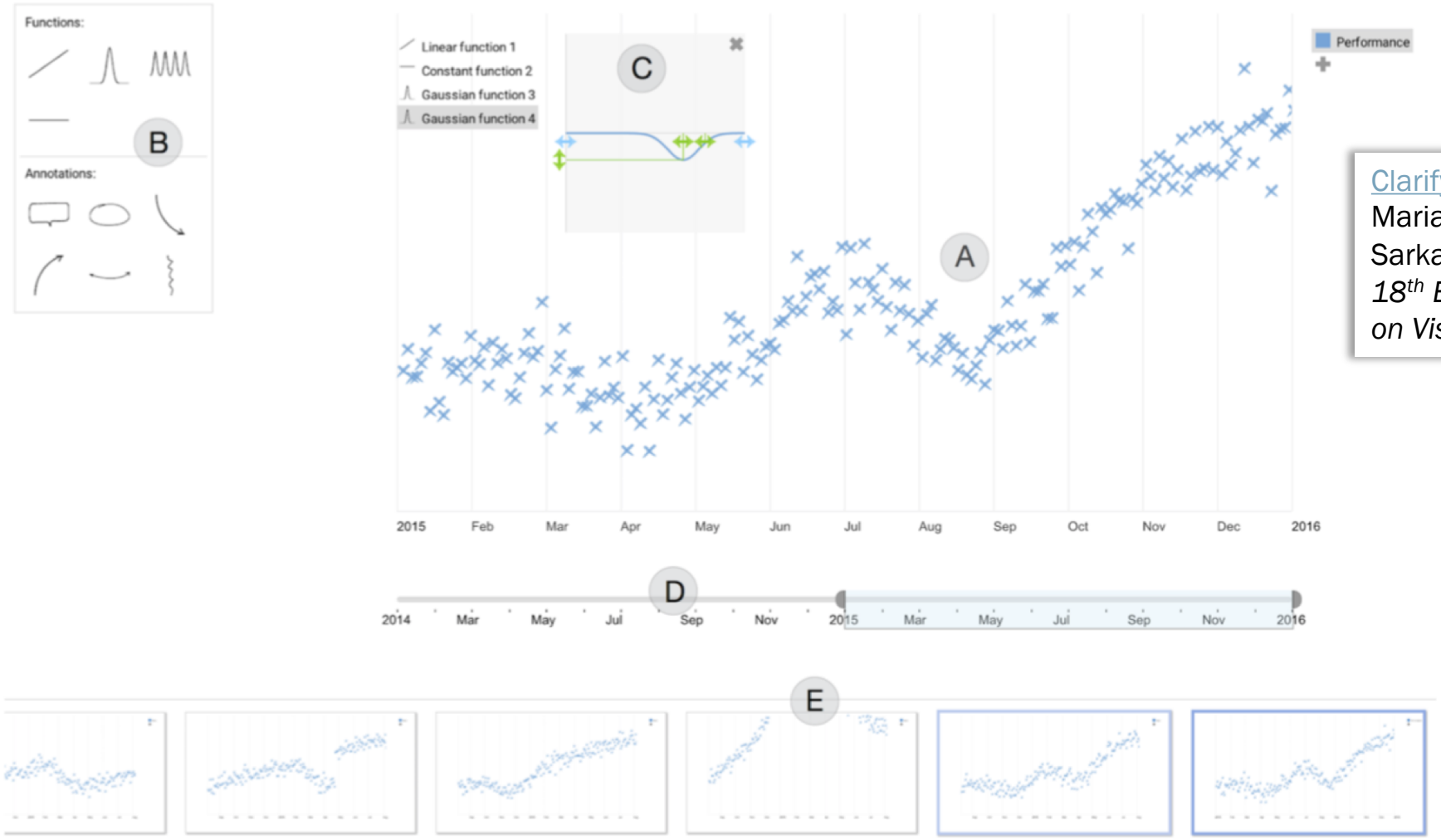


Figure 1: The SelfRaisingData prototype with the main components highlighted. (A) The time series chart with the fictional data points generated around the shape described through function composition, as presented in Section 4.1. (B) The tool panel containing functions and annotations (Section 4.2). (C) The function editor allows interactive modification of the mathematical parameters of the function and the time range for which it applies, as discussed in Section 4.3. (D) The time axis range selector (see Section 4.4). (E) Graphical history using a comic strip metaphor allows branching and visualising previous states (see Section 4.5).

Clarifying hypotheses by sketching data
Mariana Mărășoiu, Alan F. Blackwell, Advait Sarkar, Martin Spott
18th Eurographics/IEEE VGTC Conference on Visualization, 2016 (EuroVis 2016)

Principles of visualisation

- Structural: e.g., Bertin, Wilkinson/Wickham
- Perceptual/cognitive: e.g., Bertin, Cleveland & McGill
- Aesthetic/designerly: e.g., Tufte

Interaction and visualisation

- Shneiderman’s mantra: Overview, zoom, filter, detail-on-demand
- Yi et al (2007)

LATENT SEMANTIC ANALYSIS

A

$$
\mathrm{M}=\begin{array}{c} \\ T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ \vdots \\ T_{\mathrm{m}} \end{array}
\begin{array}{c}
\begin{array}{cccccccc} D_1 & D_2 & D_3 & D_4 & D_5 & D_6 & \cdots & D_{\mathrm{n}} \end{array} \\
\begin{pmatrix}
0.00060 & 0.00012 & 0.00003 & 0.00003 & 0.00333 & 0.00048 & \cdots & a_{1n} \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & a_{2n} \\
0 & 2.98862 & 0 & 0 & 0 & 1.49431 & \cdots & a_{3n} \\
0 & 0 & 0 & 13.32555 & 0 & 0 & \cdots & a_{4n} \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & a_{5n} \\
1.03442 & 1.03442 & 0 & 0 & 0 & 3.10326 & \cdots & a_{6n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & a_{m4} & a_{m5} & a_{m6} & \cdots & a_{mn}
\end{pmatrix}
\end{array}
$$

B

$$
\mathrm{U}=\begin{array}{c} \\ T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ \vdots \\ T_m \end{array}
\begin{array}{c}
\mathrm{U}_k \\
\begin{array}{cccccc} C_1 & C_2 & C_3 & \cdots & C_m \end{array} \\
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\
a_{31} & a_{32} & a_{33} & \cdots & a_{3m} \\
a_{41} & a_{42} & a_{43} & \cdots & a_{4m} \\
a_{51} & a_{52} & a_{53} & \cdots & a_{5m} \\
a_{61} & a_{62} & a_{63} & \cdots & a_{6m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm}
\end{pmatrix}
\end{array}
$$

$$
\Sigma=\begin{array}{c} \\ T_1 \\ T_2 \\ T_3 \\ T_4 \\ \vdots \\ T_m \end{array}
\begin{array}{c}
\sum_k \\
\begin{array}{ccccc} D_1 & D_2 & D_3 & \cdots & D_n \end{array} \\
\begin{pmatrix}
a_{11} & 0 & 0 & \cdots & 0 \\
0 & a_{22} & 0 & \cdots & 0 \\
0 & 0 & a_{33} & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & a_{mm}
\end{pmatrix}
\end{array}
$$

$$
\mathrm{V}^T=\begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ \vdots \\ C_n \end{array}
\begin{array}{c}
\mathrm{V}_k^T \\
\begin{array}{ccccc} D_1 & D_2 & D_3 & \cdots & D_n \end{array} \\
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\
a_{41} & a_{42} & a_{43} & \cdots & a_{4n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}
\end{pmatrix}
\end{array}
$$

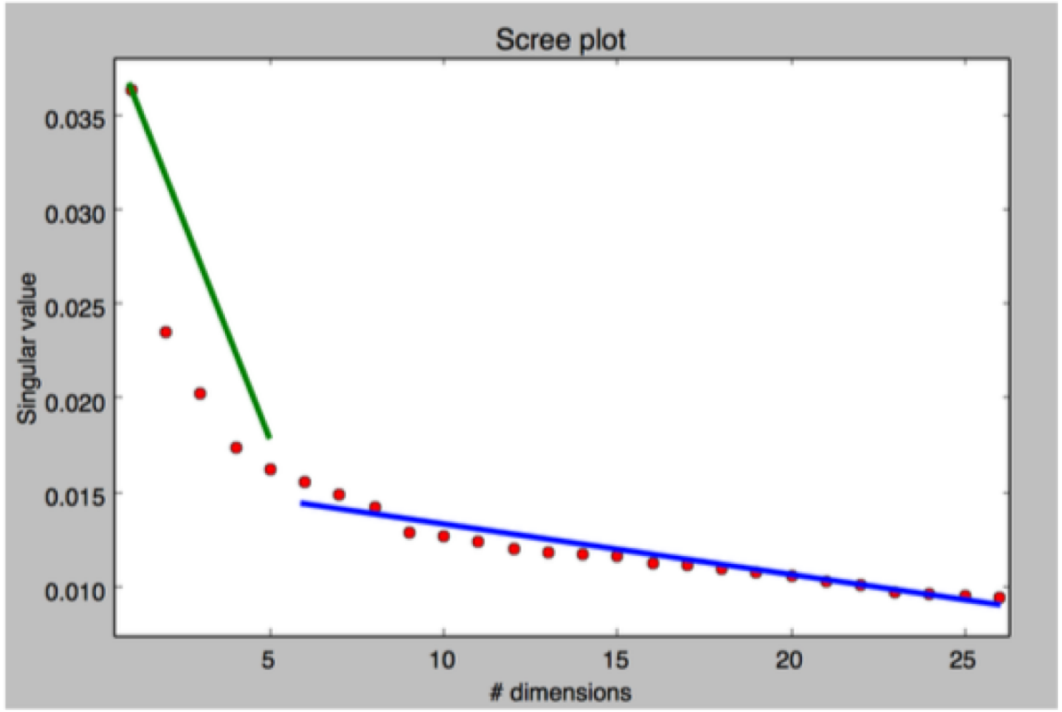
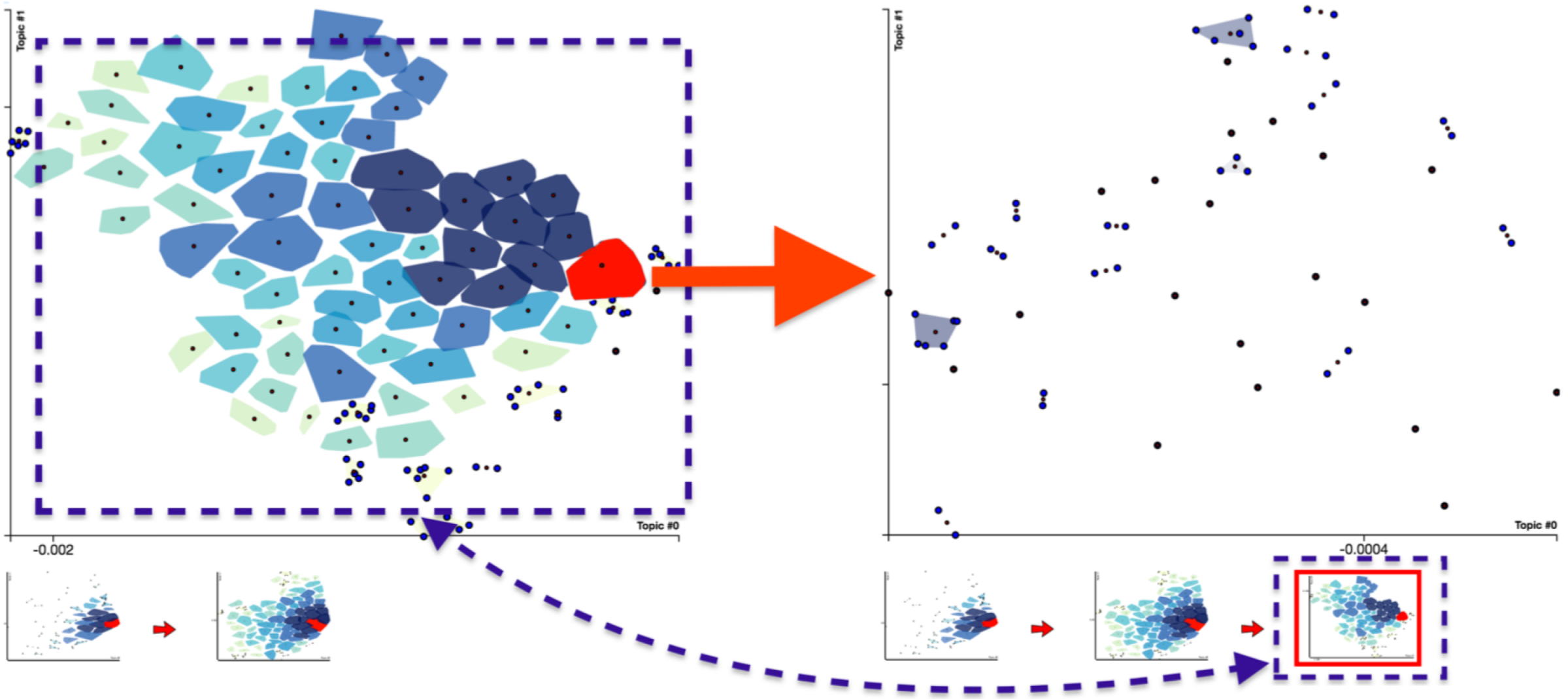
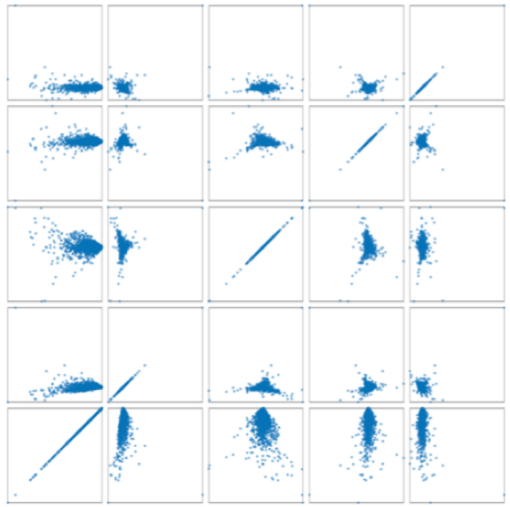


Figure 3.1: Singular value scree plot with a knee found by L-method at the 5^{th} singular value



Overview + detail, Semantic zooming, Graphical interaction histories



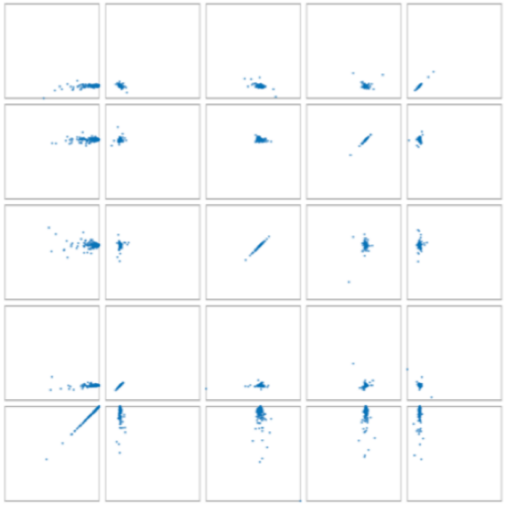


Figure 3.9: Random sampling performed in each scatte
very dense areas and a number of potentially interestin
and the shape is distorted. Sampling more values in w
performance.

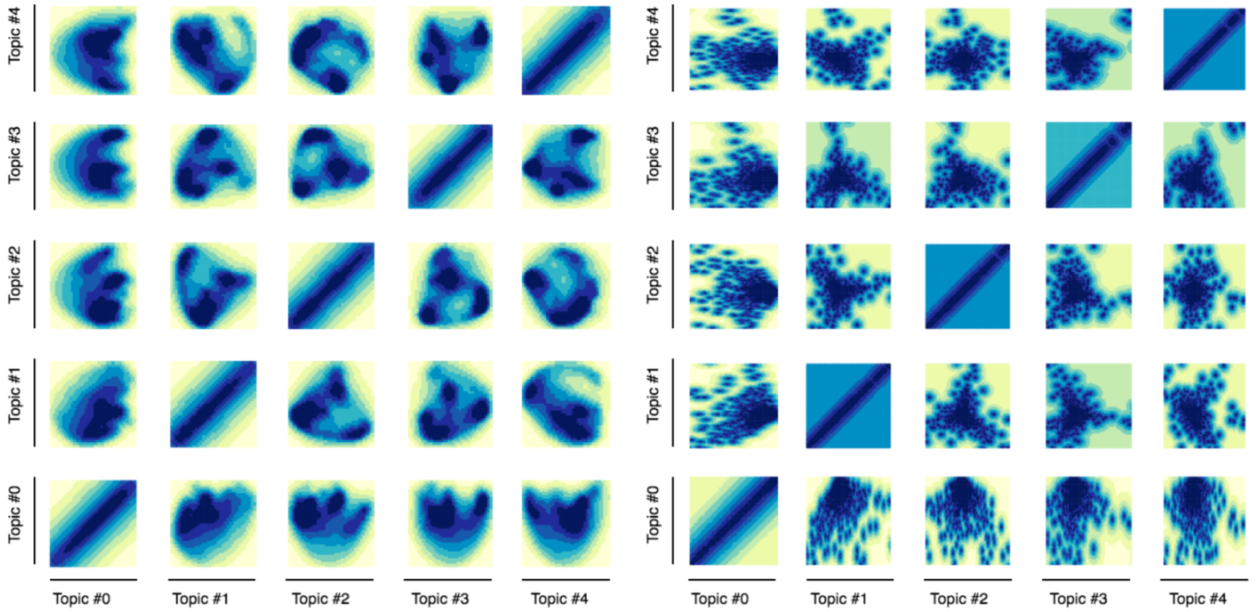


Figure 3.10: Two examples of heat map matrices. The colour scale ranging from light yellow to dark blue indicates the estimated probability density of the data distribution. Blue areas indicate higher probabilities of data points at that position.

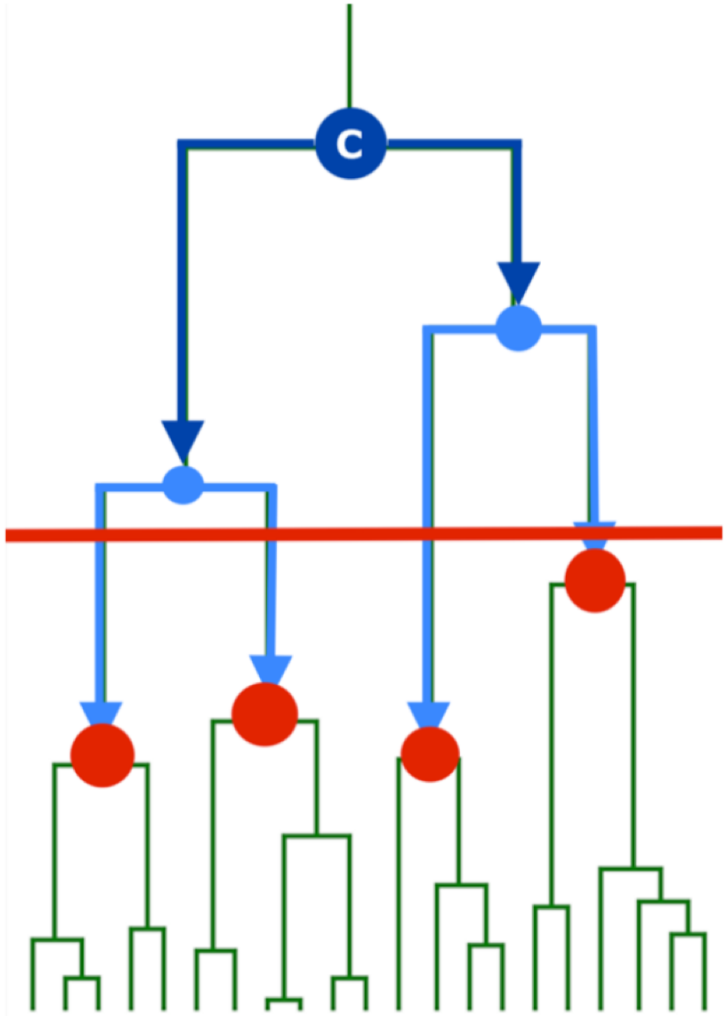


Figure 3.15: Obtaining the expansion of a cluster. To determine which clusters would become C's children in the expansion tree, a cut (in red) is made at the height corresponding to the minimum displayable distance between clusters. C's children are then expanded until the clusters immediately below the cut are reached; these are then chosen as C's expansion.

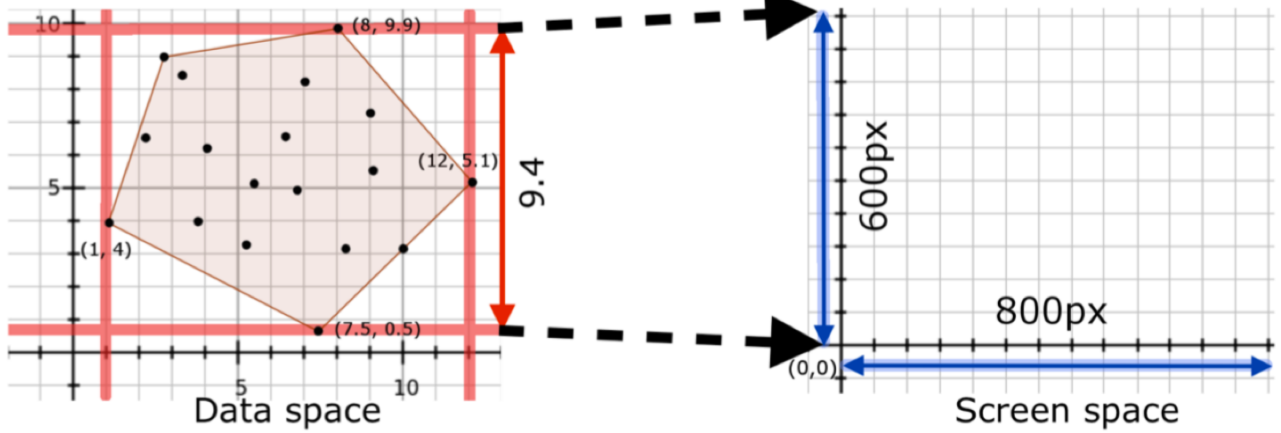
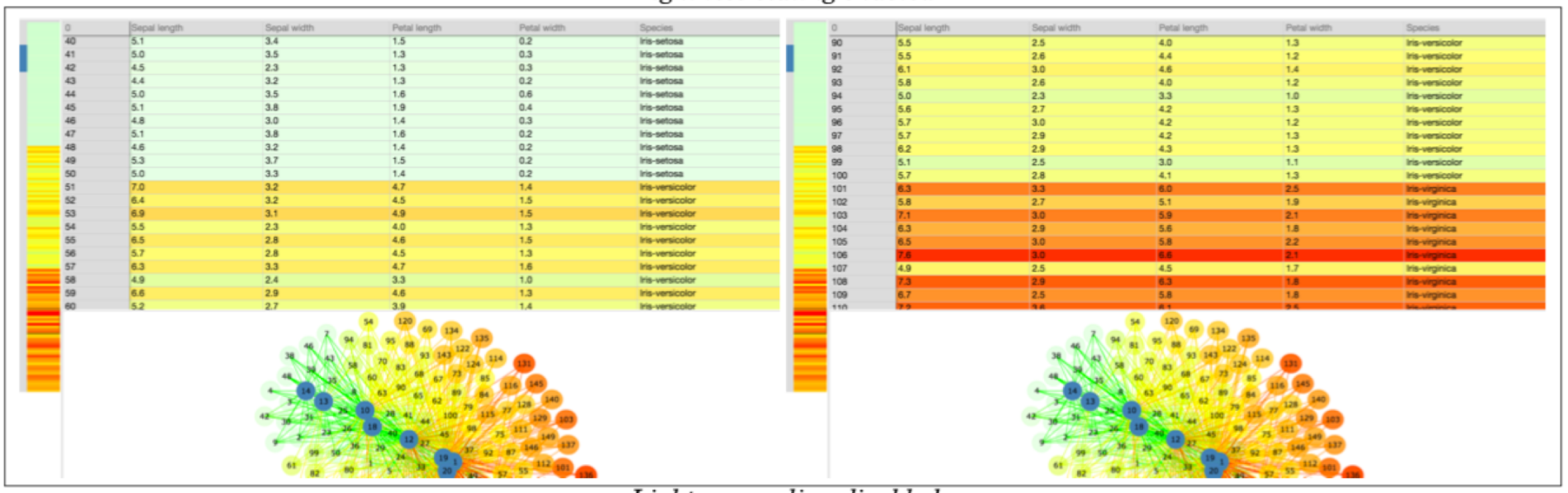


Figure 3.16: Mapping from data to screen space. The cluster shown is a cluster we want to expand and will be fragmented into its descendant clusters. By knowing the extent on one dimension in data space and the size of the y-axis in screen space, we can obtain a linear mapping between the two spaces. We can do the same for the other data dimension and x-axis.

BRAINCEL

Lightness scaling enabled



	Sepal length	Sepal width	Petal length	Petal width	Species
40	5.1	3.4	1.5	0.2	Iris-setosa
41	5.0	3.5	1.3	0.3	Iris-setosa
42	4.5	2.3	1.3	0.3	Iris-setosa
43	4.4	3.2	1.3	0.2	Iris-setosa
44	5.0	3.5	1.6	0.6	Iris-setosa
45	5.1	3.8	1.9	0.4	Iris-setosa
46	4.8	3.0	1.4	0.3	Iris-setosa
47	5.1	3.8	1.6	0.2	Iris-setosa
48	4.6	3.2	1.4	0.2	Iris-setosa
49	5.3	3.7	1.5	0.2	Iris-setosa
50	5.0	3.3	1.4	0.2	Iris-setosa
51	7.0	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
54	5.5	2.3	4.0	1.3	Iris-versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor
56	5.7	2.8	4.5	1.3	Iris-versicolor
57	6.3	3.3	4.7	1.6	Iris-versicolor
58	4.9	2.4	3.3	1.0	Iris-versicolor
59	6.6	2.9	4.6	1.3	Iris-versicolor
60	5.2	2.7	3.9	1.4	Iris-versicolor

	Sepal length	Sepal width	Petal length	Petal width	Species
90	5.5	2.5	4.0	1.3	Iris-versicolor
91	5.5	2.6	4.4	1.2	Iris-versicolor
92	6.1	3.0	4.6	1.4	Iris-versicolor
93	5.8	2.6	4.0	1.2	Iris-versicolor
94	5.0	2.3	3.3	1.0	Iris-versicolor
95	5.6	2.7	4.2	1.3	Iris-versicolor
96	5.7	3.0	4.2	1.2	Iris-versicolor
97	5.7	2.9	4.2	1.3	Iris-versicolor
98	6.2	2.9	4.3	1.3	Iris-versicolor
99	5.1	2.5	3.0	1.1	Iris-versicolor
100	5.7	2.8	4.1	1.3	Iris-versicolor
101	6.3	3.3	6.0	2.5	Iris-virginica
102	5.8	2.7	5.1	1.9	Iris-virginica
103	7.1	3.0	5.9	2.1	Iris-virginica
104	6.3	2.9	5.6	1.8	Iris-virginica
105	6.5	3.0	5.8	2.2	Iris-virginica
106	7.6	3.0	6.6	2.1	Iris-virginica
107	4.9	2.5	4.5	1.7	Iris-virginica
108	7.3	2.9	6.3	1.8	Iris-virginica
109	6.7	2.5	5.8	1.8	Iris-virginica

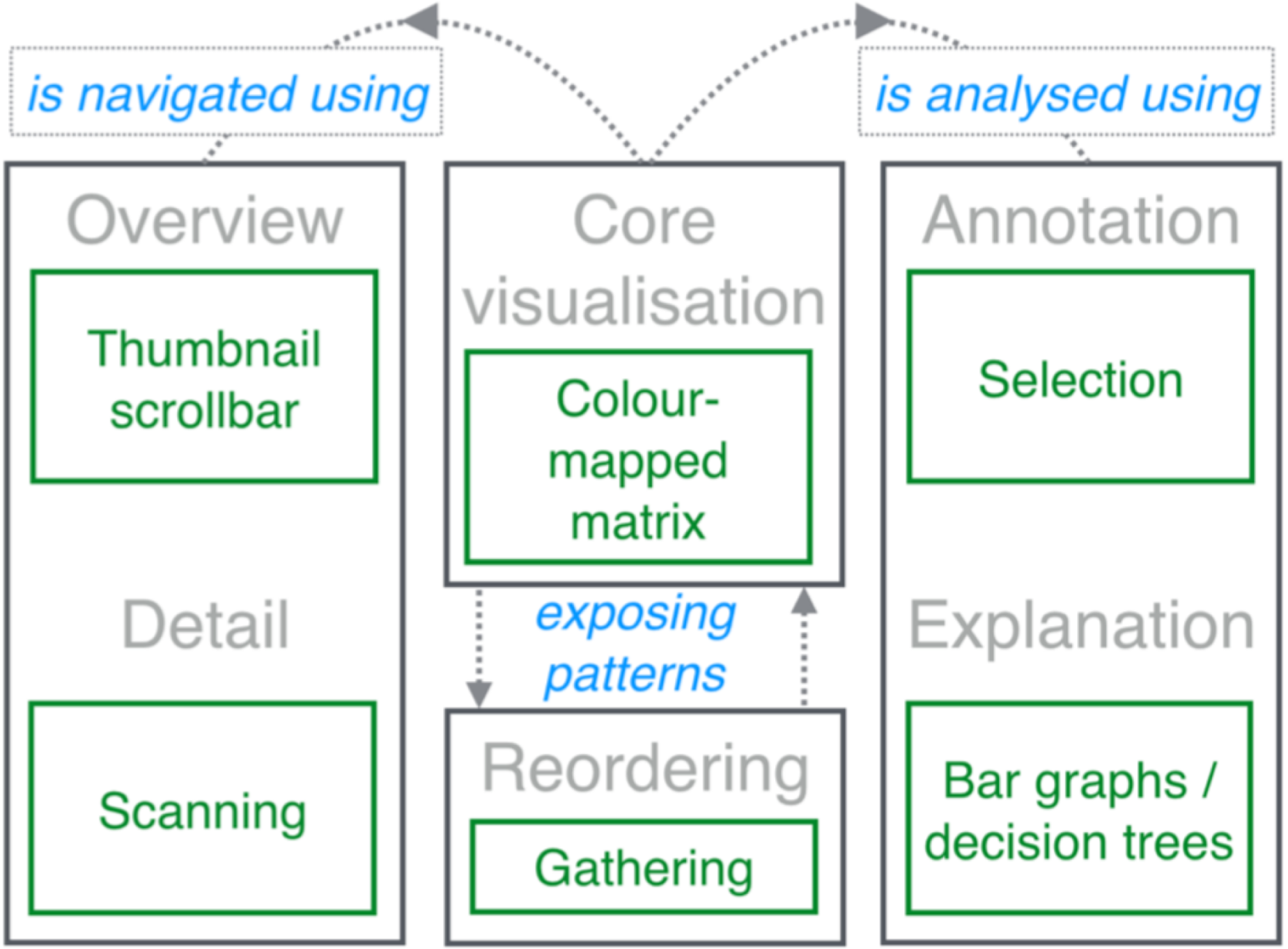
Lightness scaling disabled

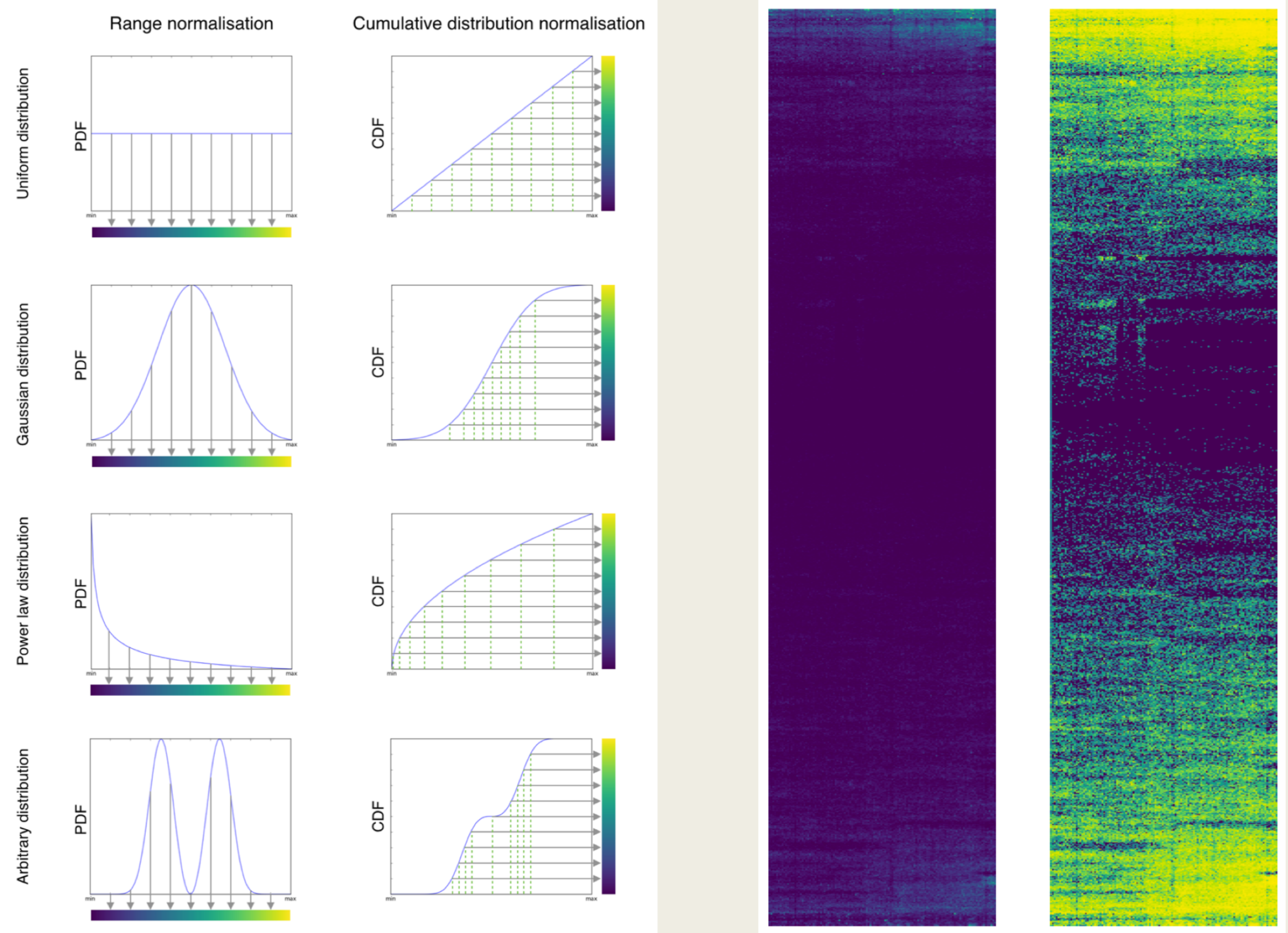
	Sepal length	Sepal width	Petal length	Petal width	Species
40	5.1	3.4	1.5	0.2	Iris-setosa
41	5.0	3.5	1.3	0.3	Iris-setosa
42	4.5	2.3	1.3	0.3	Iris-setosa
43	4.4	3.2	1.3	0.2	Iris-setosa
44	5.0	3.5	1.6	0.6	Iris-setosa
45	5.1	3.8	1.9	0.4	Iris-setosa
46	4.8	3.0	1.4	0.3	Iris-setosa
47	5.1	3.8	1.6	0.2	Iris-setosa
48	4.6	3.2	1.4	0.2	Iris-setosa
49	5.3	3.7	1.5	0.2	Iris-setosa
50	5.0	3.3	1.4	0.2	Iris-setosa
51	7.0	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
54	5.5	2.3	4.0	1.3	Iris-versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor
56	5.7	2.8	4.5	1.3	Iris-versicolor
57	6.3	3.3	4.7	1.6	Iris-versicolor
58	4.9	2.4	3.3	1.0	Iris-versicolor
59	6.6	2.9	4.6	1.3	Iris-versicolor
60	5.2	2.7	3.9	1.4	Iris-versicolor

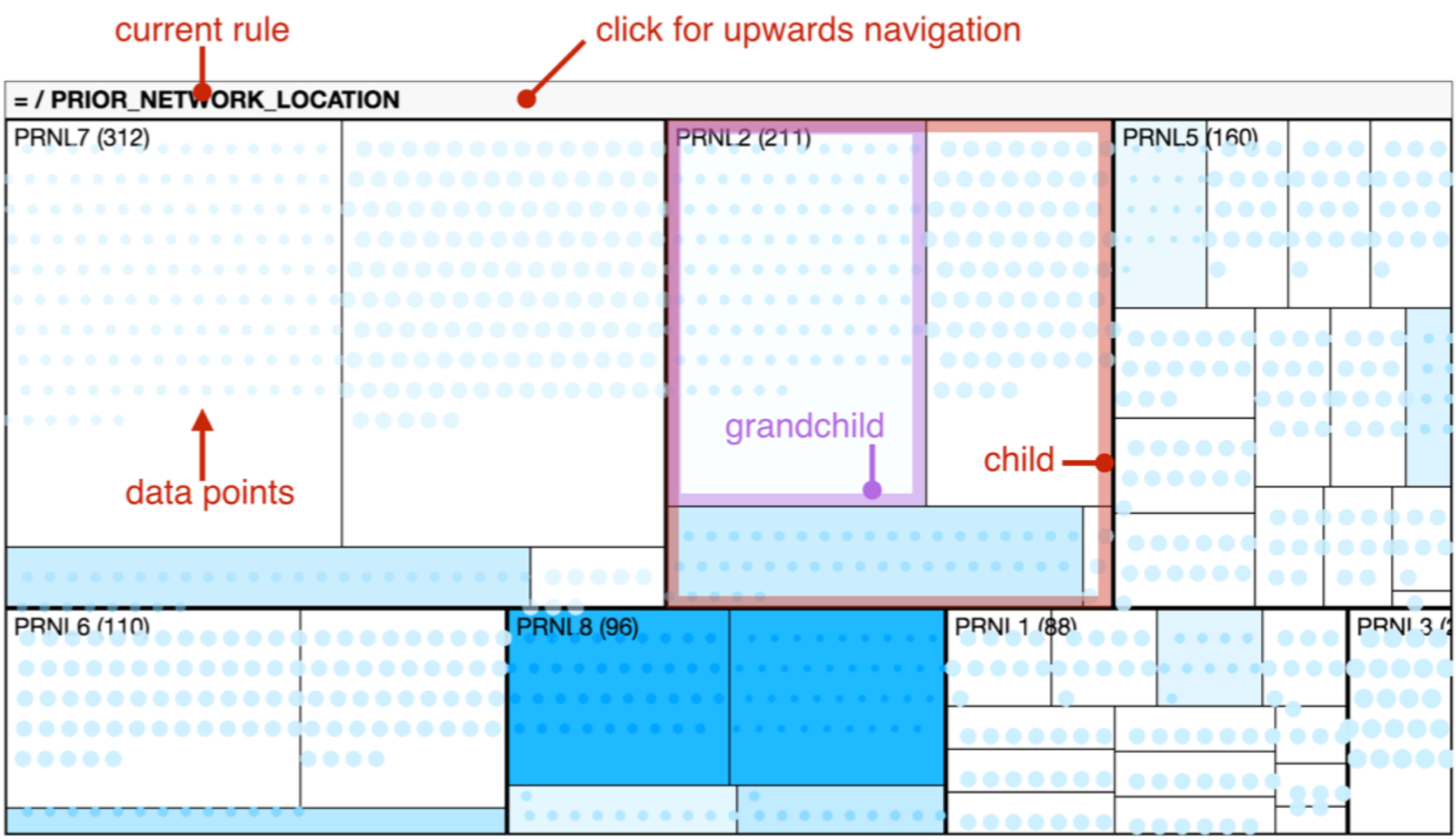
	Sepal length	Sepal width	Petal length	Petal width	Species
90	5.5	2.5	4.0	1.3	Iris-versicolor
91	5.5	2.6	4.4	1.2	Iris-versicolor
92	6.1	3.0	4.6	1.4	Iris-versicolor
93	5.8	2.6	4.0	1.2	Iris-versicolor
94	5.0	2.3	3.3	1.0	Iris-versicolor
95	5.6	2.7	4.2	1.3	Iris-versicolor
96	5.7	3.0	4.2	1.2	Iris-versicolor
97	5.7	2.9	4.2	1.3	Iris-versicolor
98	6.2	2.9	4.3	1.3	Iris-versicolor
99	5.1	2.5	3.0	1.1	Iris-versicolor
100	5.7	2.8	4.1	1.3	Iris-versicolor
101	6.3	3.3	6.0	2.5	Iris-virginica
102	5.8	2.7	5.1	1.9	Iris-virginica
103	7.1	3.0	5.9	2.1	Iris-virginica
104	6.3	2.9	5.6	1.8	Iris-virginica
105	6.5	3.0	5.8	2.2	Iris-virginica
106	7.6	3.0	6.6	2.1	Iris-virginica
107	4.9	2.5	4.5	1.7	Iris-virginica
108	7.3	2.9	6.3	1.8	Iris-virginica
109	6.7	2.5	5.8	1.8	Iris-virginica

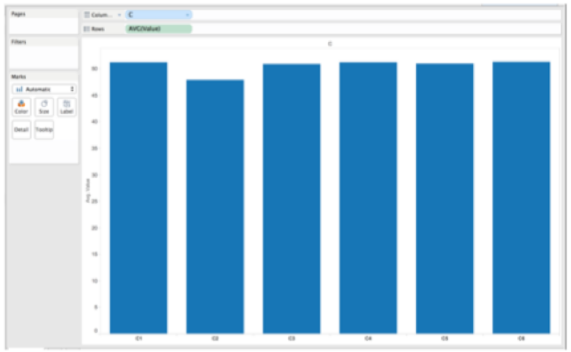
Figure 5.5: The effect of lightness scaling. Without lightness scaling, high-confidence (green) rows command disproportionately greater visual attention (the effect is most apparent onscreen).

GATHERMINER


is navigated using
is analysed using
Overview
Thumbnail scrollbar
Detail
Scanning
Core visualisation
Colour-mapped matrix
exposing patterns
Reordering
Gathering
Annotation
Selection
Explanation
Bar graphs / decision trees

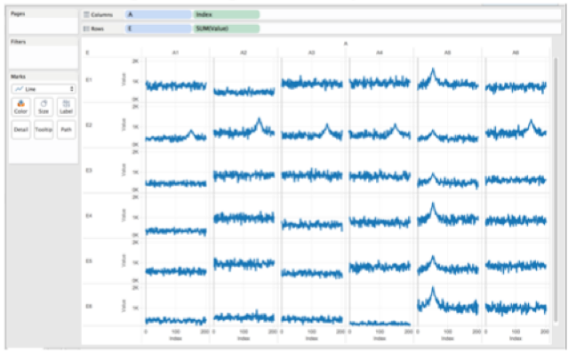

Range normalisation
Cumulative distribution normalisation
Uniform distribution
Gaussian distribution
Power law distribution
Arbitrary distribution
PDF
CDF
min
max


current rule
click for upwards navigation
= / PRIOR_NETWORK_LOCATION
PRNL7 (312)
PRNL2 (211)
PRNL5 (160)
grandchild
child
data points
PRNI 6 (110)
PRNI 8 (96)
PRNI 1 (88)
PRNI 3



(a) This strategy involved comparing bar charts of each attribute-value pairing, aggregated over the entire span of time. Since the interesting features in our time series consisted of unusual spikes/troughs, this usually reflected in a higher/lower overall sum or average for those series – easily spotted in an unusually tall or short bar.

(b) This strategy involved comparing aggregate line charts of each attribute-value pairing. Here, any attribute-value that caused spikes or dips was clearly reflected.

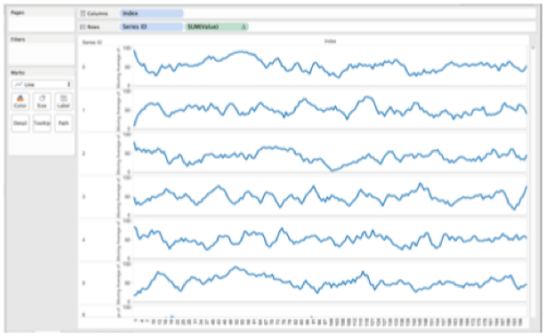


(c) This interesting strategy also compared aggregate line charts of each attribute-value pairing. Here, by creating a 2D matrix of small multiples, the analyst was able to investigate the interaction of any two attributes.

Figure 4.15: Three successful strategies in Tableau.

(a) This strategy involved inspecting a completely aggregated line graph. In this dataset, we prepared a number of time series that had spikes at about $1/3$ and $2/3$ the duration of the series, which are clearly visible in the aggregate chart. However, there are also a number of series which have an upward spike in the halfway mark, and an equal number which have an equal and opposite downward spike at the same position. The two cancel each other out and become invisible in the aggregate line graph, and so the analyst never discovers them.

(b) This strategy, similar to the first successful strategy, uses summary bar graphs to represent the time series. However, since the series are completely disaggregated (one bar is generated per series), it is impossible to seek out global patterns.



(c) This strategy involved scanning through the entire list of time series, represented as line graphs, and manually noting down the attributes of any which appeared interesting. Needless to say, this is extremely ineffective and led to several false correlations being "discovered".

Figure 4.16: Three unsuccessful strategies using Tableau.