SC4024/CZ4124
Data Visualization

Assistant Professor WANG Yong
yong-wang@ntu.edu.sg
CCDS, Nanyang Technological University

1

Chapter 12
High-dimensional Data Visualization

2

Outline

- What is high-dimensional data?
- High-dimensional data visualization approaches
  - Scatter-plot Matrix
  - Parallel Coordinates
  - Glyph-based Methods
  - Small Multiples
  - Parallel Sets

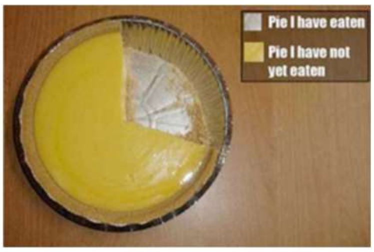
Data Dimension

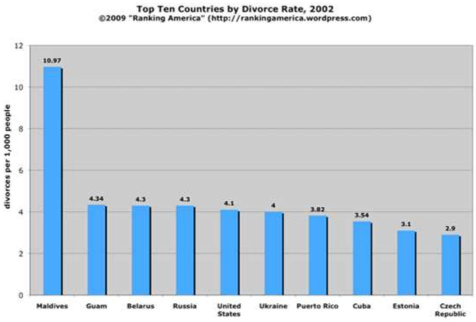
- Dimension (Number of attributes)
  - One Dimension
  - Two Dimension
  - Three Dimension
  - High Dimension

1-D Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Eaten or not





Top ten divorce rates

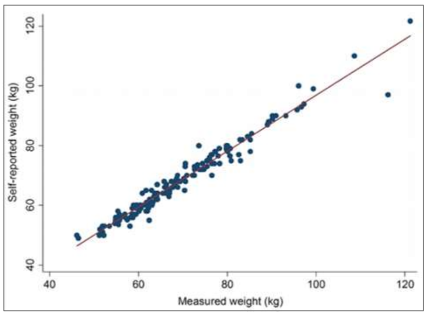
5

5

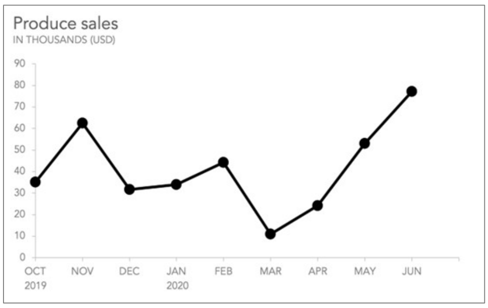
2-D Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

2-D plots



Self-reported weights *vs.* measured weight

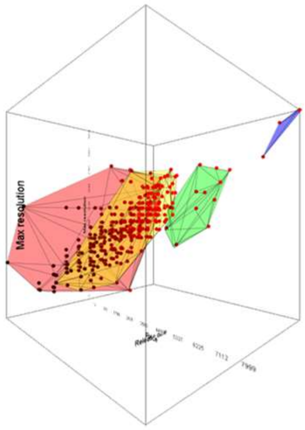


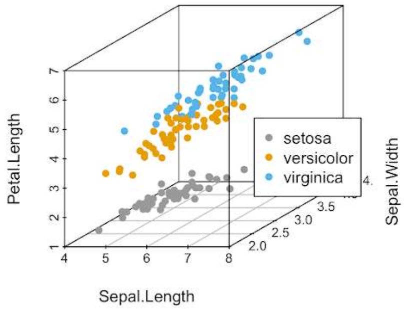
Produces sales *vs.* time

6

6

3-D Data





3D data can be represented by the 3D coordinates of each dot in a 3D scatterplot.

[1] Elmqvist et al. "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation." IEEE transactions on Visualization and Computer Graphics 14.6 (2008): 1539-1148.
[2] http://www.sthda.com/english/wiki/scatterplot3d-3d-graphics-r-software-and-data-visualization

7

High-dimensional Data

- How to visualize high-dimensional data in visual space (2-D or 3-D) ?

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	type
3.8	0.31	0.02	11.1	0.036	20	114	0.99248	3.75	0.44	12.4	6	white
3.9	0.225	0.4	4.2	0.03	29	118	0.989	3.57	0.36	12.8	8	white
4.2	0.17	0.36	1.8	0.029	93	161	0.98999	3.65	0.89	12	7	white
4.2	0.215	0.23	5.1	0.041	64	157	0.99688	3.42	0.44	8	3	white
4.4	0.46	0.1	2.8	0.024	31	111	0.98816	3.48	0.34	13.1	6	white
4.4	0.32	0.39	4.3	0.03	31	127	0.98904	3.46	0.36	12.8	8	white
4.4	0.54	0.09	5.1	0.038	52	97	0.99022	3.41	0.4	12.2	7	white
4.5	0.19	0.21	0.95	0.033	89	159	0.99332	3.34	0.42	8	5	white
4.6	0.52	0.15	2.1	0.054	8	65	0.9934	3.9	0.56	13.1	4	red
4.6	0.445	0	1.4	0.053	11	178	0.99426	3.79	0.55	10.2	5	white
4.7	0.6	0.17	2.3	0.058	17	106	0.9932	3.85	0.6	12.9	6	red
4.7	0.67	0.09	1	0.02	5	9	0.98722	3.3	0.34	13.6	5	white
4.7	0.455	0.18	1.9	0.036	33	106	0.98746	3.21	0.83	14	7	white
4.7	0.145	0.29	1	0.042	35	90	0.9908	3.76	0.49	11.3	6	white
4.7	0.335	0.14	1.3	0.036	69	168	0.99212	3.47	0.46	10.5	5	white

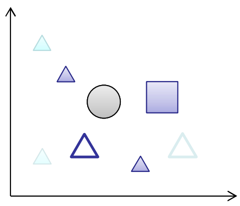
Wine Dataset

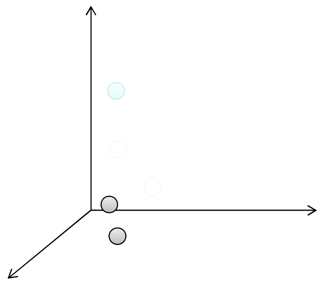
Source: UCI Machine Learning Repository https://archive.ics.uci.edu/ml/datasets/Wine

8

Simple Solutions

- Add more channel on 2-D or 3-D plots
  - Shape/fill style/color/size of points





9

9

Simple Solutions

- Multiple coordinated views: present some attributes of objects in a view

[1] Andrew et al. "Leveraging Interaction History for Intelligent Configuration of Multiple Coordinated Views in Visualization Tools
LIVVIL: Logging Interactive Visualizations & Visualizing Interaction Logs (2016).
[2] https://square.github.io/crossfilter/

10

10

High-Dimensional Data Visualization

- More solutions?

One Concrete Example – Show Me Your Visualization Designs?

- Suppose you are given a dataset describing 100 students' profile (anonymized) and academic information, which includes the following attributes:
  - Age
  - Height
  - Weight
  - Travel time to campus
  - Math score
  - Physics score
  - Visual analytics score
  - English language score
  - Algorithm design score
  - Python programming score
  - Art design score
  - Chemistry score

One Concrete Example – Show Me Your Visualization Designs?

Task 1: Show the correlation between any two attributes?

Task 2: Show the overall distribution of all these attributes of all students and help viewers visually identify the similarities between students

- Age
- Height
- Weight
- Travel time to campus
- Math score
- Physics score
- Visual analytics score
- English language score
- Algorithm design score
- Python programming score
- Art design score
- Chemistry score

13

13

Outline

- What is high-dimensional data?
- High-dimensional data visualization approaches
  - Scatter-plot Matrix
  - Parallel Coordinates
  - Glyph-based Methods
  - Small Multiples
  - Parallel Sets

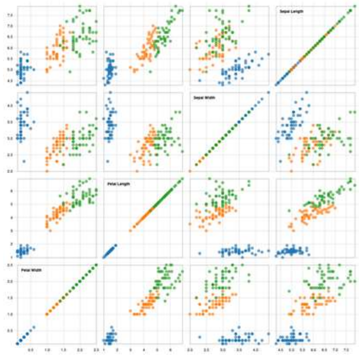
14

14

Scatter-plot Matrix

- 2-D plot for each dimension pair
- Display correlations between dimensions
- The number of 2-D plots is proportional to square of dimensions

Implemented by D3.js



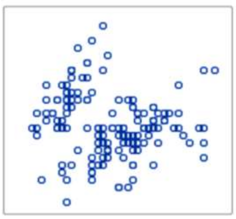
Iris Dataset: https://observablehq.com/@d3/brushable-scatterplot-matrix

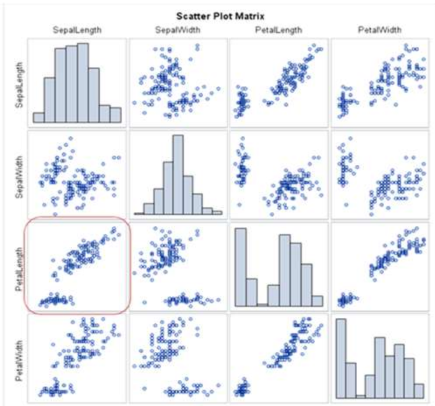
15

15

Scatter-plot Matrix

- Scatter-plot matrix can be further enhanced:
  - It can be combined with dimension reduction based projection
  - The diagonal cell can be used to show the distribution of each attribute/variable





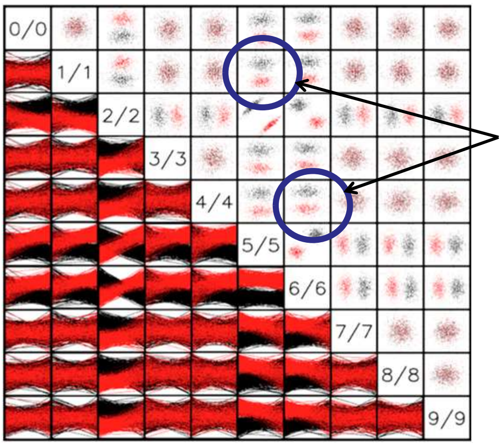
16

16

Scatter-plot Matrix – Insight Discovery

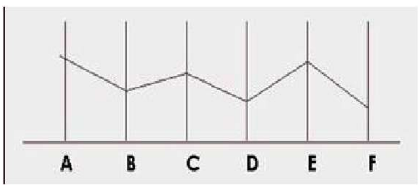
- Scatter-plot matrix can be further enhanced:

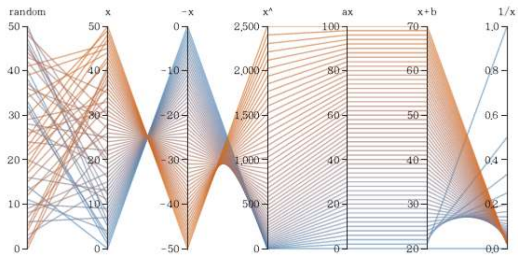
  - The cells of the upper triangle and bottom triangle can be used to serve for different analysis tasks, e.g., clustering and correlation



Parallel Coordinates

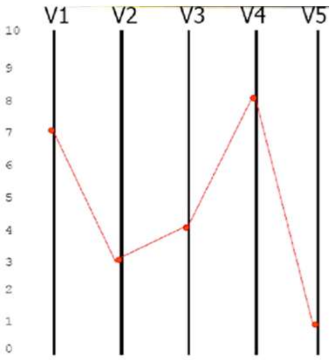
- It was proposed by Prof. Alfred Inselberg in 1985 for high-dimensional geometry
- Parallel axes
- Data points represented by lines





Inselberg, Alfred. "The plane with parallel coordinates." *The visual computer* 1, no. 2 (1985): 69-91.

Parallel Coordinates



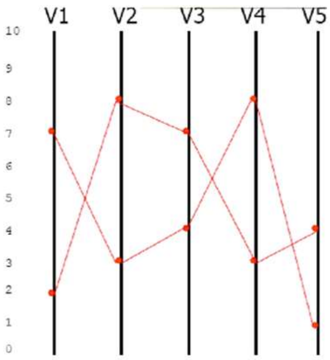
D1

7 3 4 8 1

19

19

Parallel Coordinates



D2

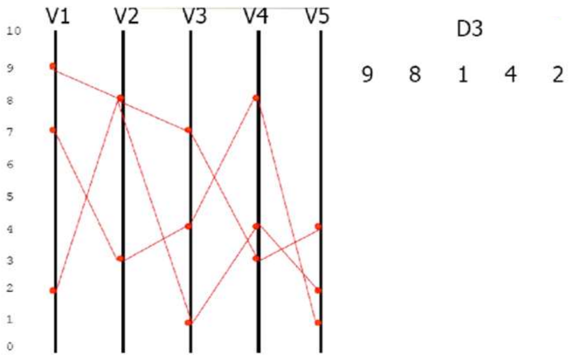
2 7 6 3 4

20

20

Parallel Coordinates

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

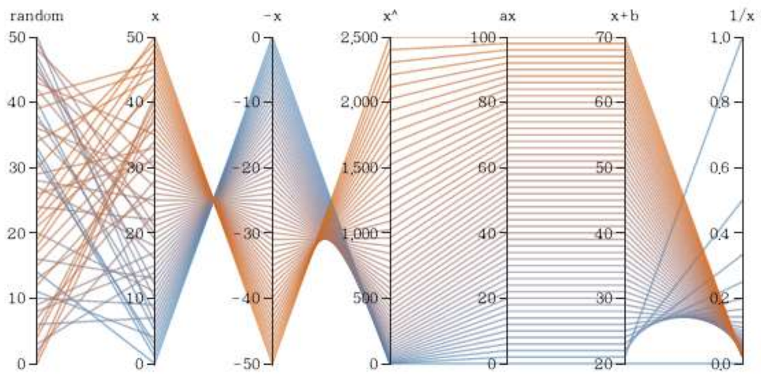


21

21

Correlations

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE



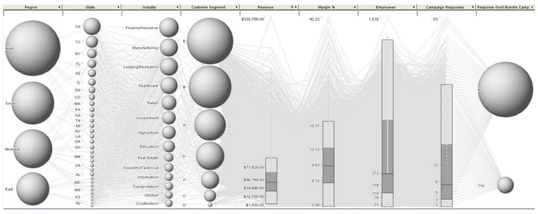
- The parallel coordinate illustrates the correlation between adjacent variables
- The order of the axis or variables really matters, as appropriate orders can reveal the correlation between adjacent axes more clearly

22

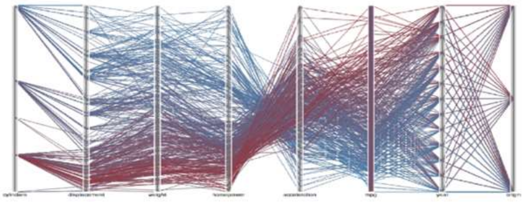
22

Parallel Coordinates in Visualization Software or Programming Packages

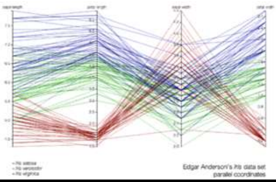
Advizor



Protovis



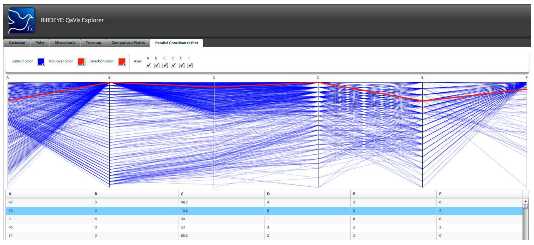
D3



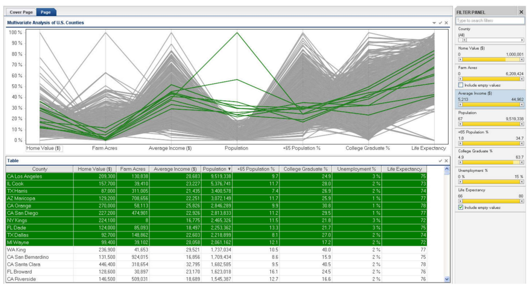
https://www.advizorsolutions.com/#home-demo
https://mbostock.github.io/protovis/

Parallel Coordinates + Interactions

- Selection



- Brush and Filter

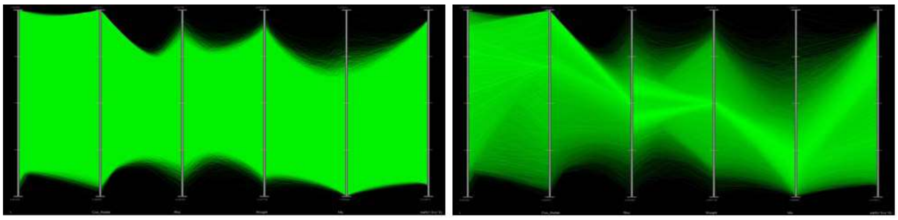


Multivariate Analysis Using Parallel Coordinates. Stephen Few September 12, 2006
https://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf

Any Other Variants of Parallel Coordinates?

Transparent Parallel Coordinates

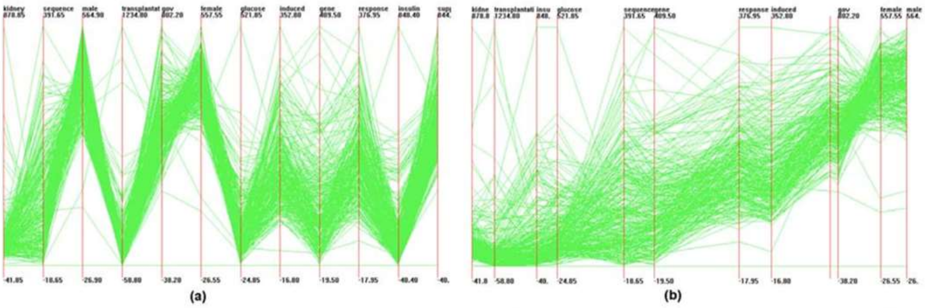
- When there are a huge number of lines in the parallel coordinates, visual clutters occur and it looks messy. A good way to mitigate it is to adjust the opacity of the lines, for example, make those lines half-transparent.



Chad Jones et al. "An Integrated Exploration Approach to Visualizing Multivariate Particle Data. " Computing in Science & Engineering (2008).

Parallel Coordinates: Re-ordering Axes to Reduce Visual Clutter

- By varying the dimension order in a display, it is possible to **reduce clutter** without reducing information content or modifying the data in any way.



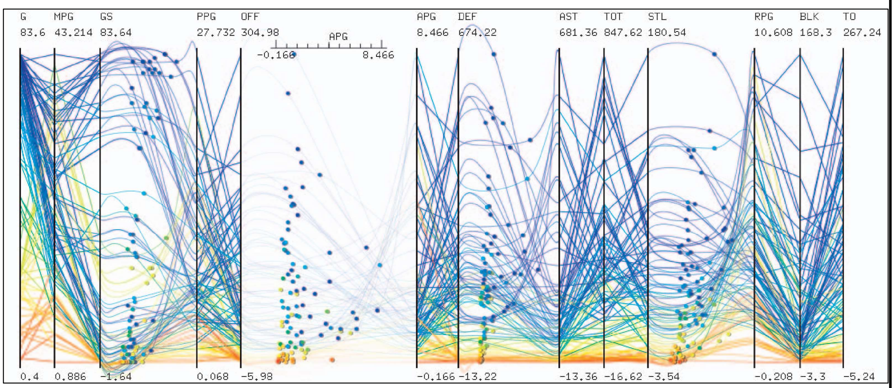
Peng, Wei , M. O. Ward et al. "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Re-ordering." IEEE Symposium on Information Visualization (2005).

27

27

Parallel Coordinates with Scatter-plots

- Augment parallel coordinates with scatterplots to facilitate **data selection** and **data clustering**

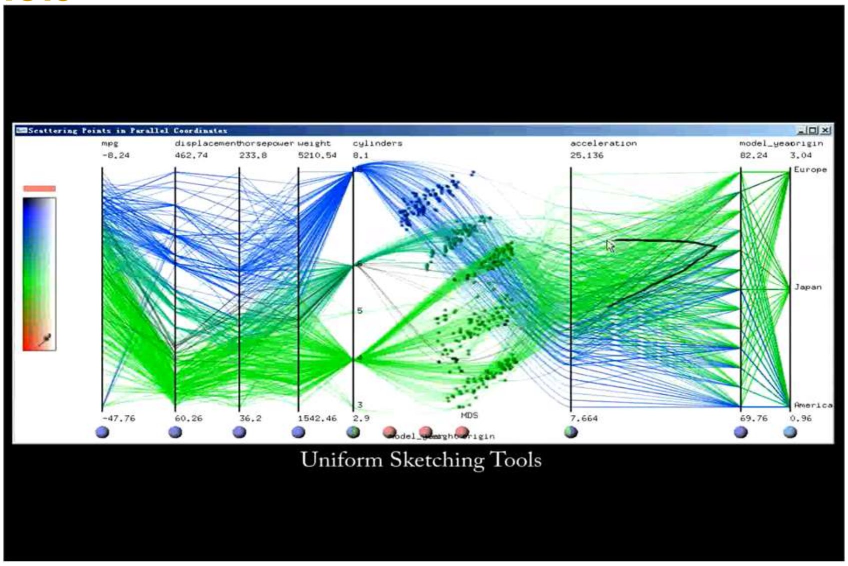


Yuan et al. "Scattering Points in Parallel Coordinates. " IEEE Transactions on Visualization and Computer Graphics (2009).

28

28

Parallel Coordinates with Scatter-plots



Yuan et al. "Scattering Points in Parallel Coordinates. " IEEE Transactions on Visualization and Computer Graphics (2009).

29

29

Outline

- What is high-dimensional data?
- High-dimensional data visualization approaches
  - Scatter-plot Matrix
  - Parallel Coordinates
  - Glyph-based Methods
  - Small Multiples
  - Parallel Sets

30

30

Glyph-based Methods

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Radar Chart (a.k.a.: star plot)
- Chernoff Faces

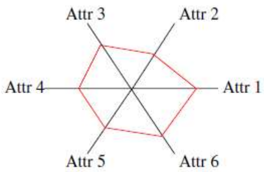
31

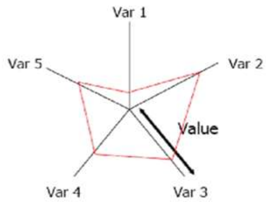
31

Star Plot

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Space variables around a circle
- Encode values on "spokes"
- Each data point is now a shape





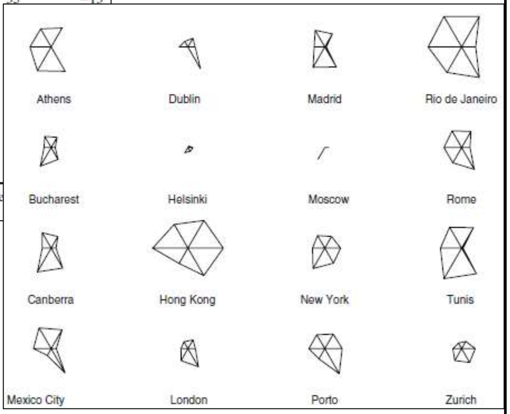
32

32

Multivariate Data: Climatic Variables

City	Precip. average	Temp. average	Temp. max average	Temp. min average	Record max	Record min
Athens	37	17	21	13	42	−3
Bucharest	58	11	16	5	49	−23
Canberra	62	12	19	6	42	−10
Dublin	74	10	12	6	28	−7
Helsinki	63	5	8	1	31	−36
Hong Kong	218	23	25	21	37	2
London	75	10	13	5	35	−13
Madrid	45	13	20	7		
Mexico City	63	17	23	11		
Moscow	59	4	8	1		
New York	118	12	17	8		
Porto	126	14	18	10		
Rio de Janeiro	109	25	30	20		
Rome	80	15	20	11		
Tunis	44	18	23	13		
Zurich	107	9	12	6		

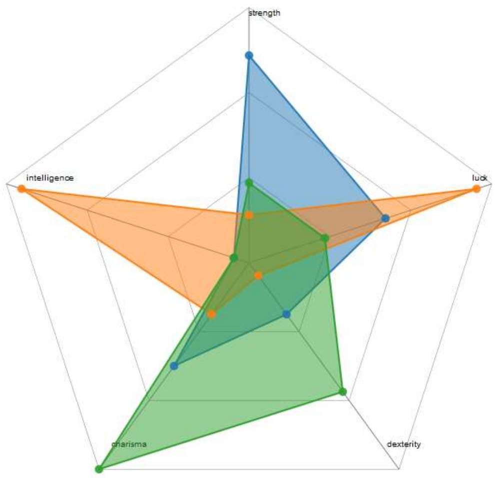
Table 4.1 Annual climatic values in Celsius of some world citie
http://www.weatherbase.com.



33

33

Star Plot in D3.js



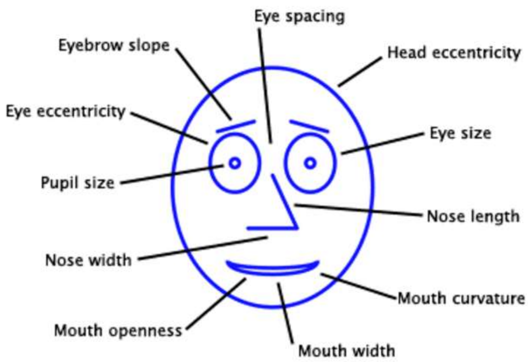
Source: http://graves.cl/radar-chart-d3/ and http://bl.ocks.org/nbremer/6506614

34

34

Chernoff Faces

- Use face encodings to represent different attributes



Facial Variables

Spinelli, Joseph G., and Yu Zhou. "Mapping quality of life with chernoff faces." In *Proceedings of 24th ESRI International User Conference.*. 2004.

35

35

Multivariate Data: Climatic Variables

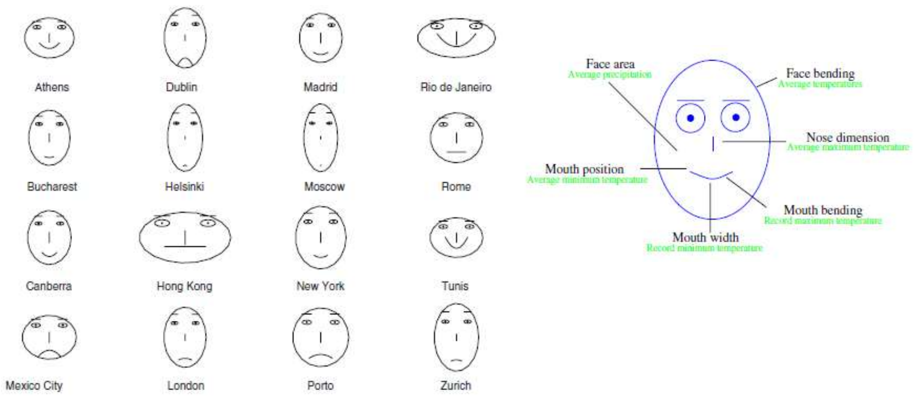
City	Precip. average	Temp. average	Temp. max average	Temp. min average	Record max	Record min
Athens	37	17	21	13	42	−3
Bucharest	58	11	16	5	49	−23
Canberra	62	12	19	6	42	−10
Dublin	74	10	12	6	28	−7
Helsinki	63	5	8	1	31	−36
Hong Kong	218	23	25	21	37	2
London	75	10	13	5	35	−13
Madrid	45	13	20	7	40	−10
Mexico City	63	17	23	11	32	−3
Moscow	59	4	8	1	35	−42
New York	118	12	17	8	40	−18
Porto	126	14	18	10	34	−2
Rio de Janeiro	109	25	30	20	43	7
Rome	80	15	20	11	37	−7
Tunis	44	18	23	13	46	−1
Zurich	107	9	12	6	35	−20

Table 4.1 Annual climatic values in Celsius of some world cities. Values from http://www.weatherbase.com.

36

36

Chernoff Faces



Source: The Use of Faces to Represent Points in k-Dimensional Space Graphically (http://www.jstor.org/stable/2284077) and A Critique of Chernoff Faces (http://eagereyes.org/criticism/chernoff-faces)

37

37

Outline

- What is high-dimensional data?

- High-dimensional data visualization approaches
  - Scatter-plot Matrix
  - Parallel Coordinates
  - Glyph-based Methods
  - Small Multiples
  - Parallel Sets

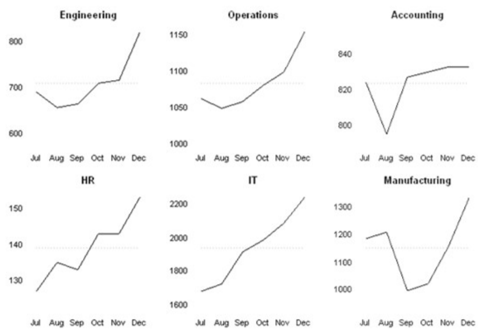
38

38

Small Multiples

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- A series or grid of small similar graphics or charts for an easy exploration and comparison of multiple dimensions



https://en.wikipedia.org/wiki/Small_multiple

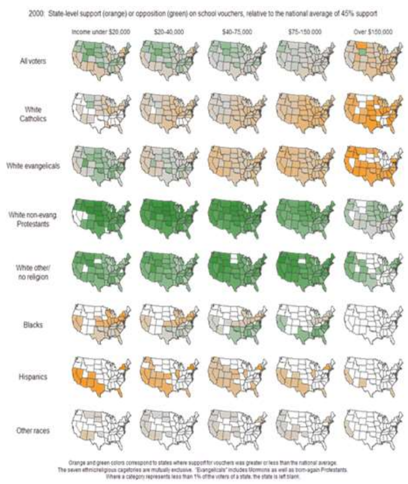
39

39

Small Multiples

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Example: Use small multiples to visualize state-level support or opposition on the school vouchers across different races and in-come levels in the united states.



http://andrewgelman.com/2009/07/hard_sell_for_b/

40

40

Outline

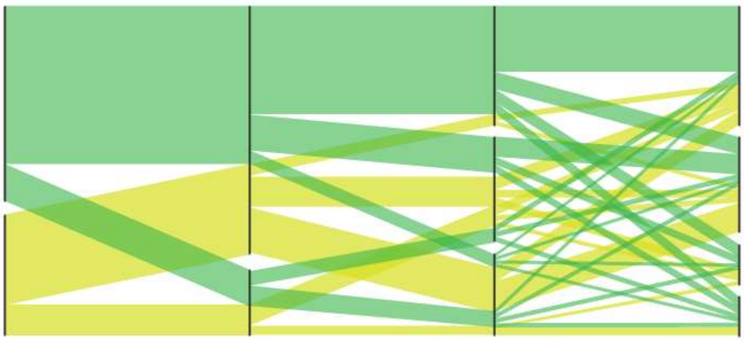
- What is high-dimensional data
- High-dimensional data visualization approaches
  - Scatter-plot Matrix
  - Parallel Coordinates
  - Glyph-based Methods
  - Small Multiples
  - Parallel Sets

41

41

Parallel Sets

- It is mainly designed for visualizing high-dimensional **categorical** data
- It shows data frequencies instead of the individual data points



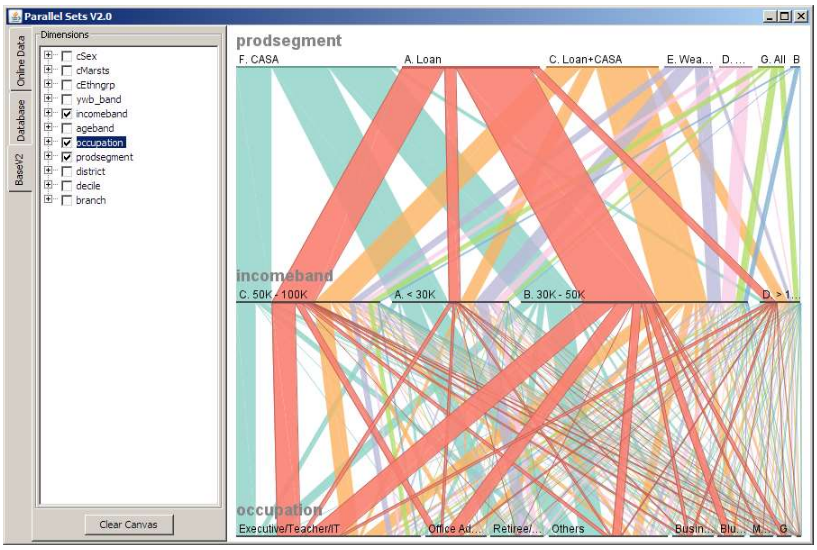
https://datavizcatalogue.com/methods/parallel_sets.html

42

42

Exploring High-dimensional Categorical Data with Parallel Sets

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

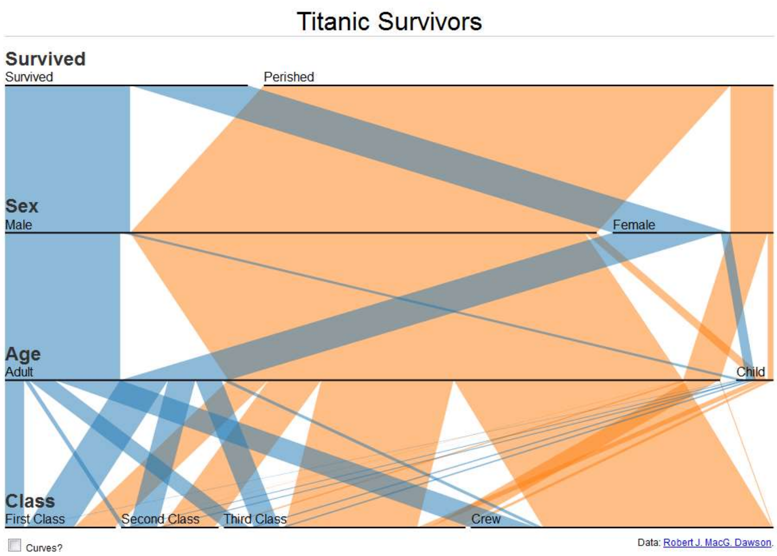


43

43

Parallel Sets – d3.js

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE



44

44

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

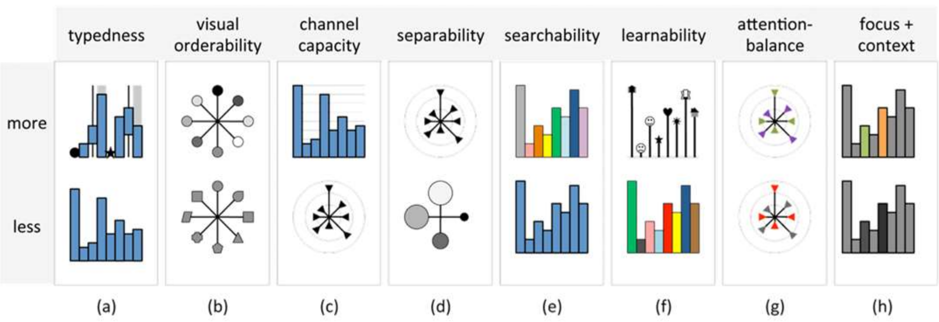
Questions?

Thank You!

45

A Survey on Glyph-based Visualization

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE



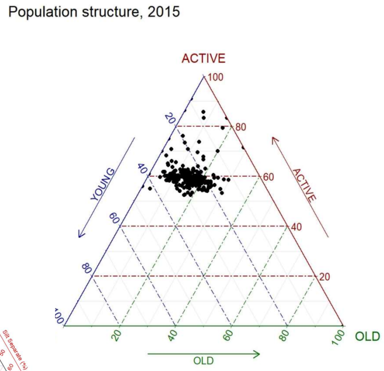
Borgo R, Kehrer J et al. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. Eurographics (STARs) (2013)

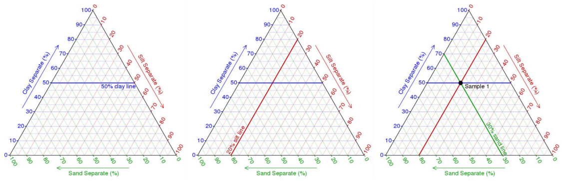
46

46

TERNARY PLOT

- Ternary plots are a way of displaying the distribution and variability of three-part compositional data.
- Its display is a triangle with three components. Each side represents one of the three components.





Reference: https://en.wikipedia.org/wiki/Ternary_plot

Dimensionality Reduction

- Project the high-dimensional data onto a lower-dimensional subspace using linear or non-linear transformations.
- Projection preserves important relations (e.g., no information loss, data discrimination).

$$x = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_N \end{pmatrix} \rightarrow reduce\ dimensionality \rightarrow \hat{x} = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_K \end{pmatrix} (K \ll N)$$

Dimensionality Reduction Methods

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Linear methods:
  - Principal Component Analysis (PCA)
  - Multidimensional Scaling (MDS)

- Nonlinear methods:
  - * ISOMAP
  - * Local Linear Embedding (LLE)

49