Scientific and Large Data Visualization
22 November 2017
High Dimensional Data

Massimiliano Corsini

Visual Computing Lab, ISTI - CNR - Italy

Overview

- Graphs Extensions
- Glyphs
  - Chernoff Faces
  - Multi-dimensional Icons
- Parallel Coordinates
- Star Plots
- Dimensionality Reduction
  - Principal Component Analysis (PCA)
  - Locally Linear Embedding (LLE)
  - IsoMap
  - t-SNE

Multi-set Bar Charts

- Also called *grouped bar charts*.
- Like bar chart for more datasets.
- Allow accurate numerical comparisons.
- It can be used to show mini-histograms.

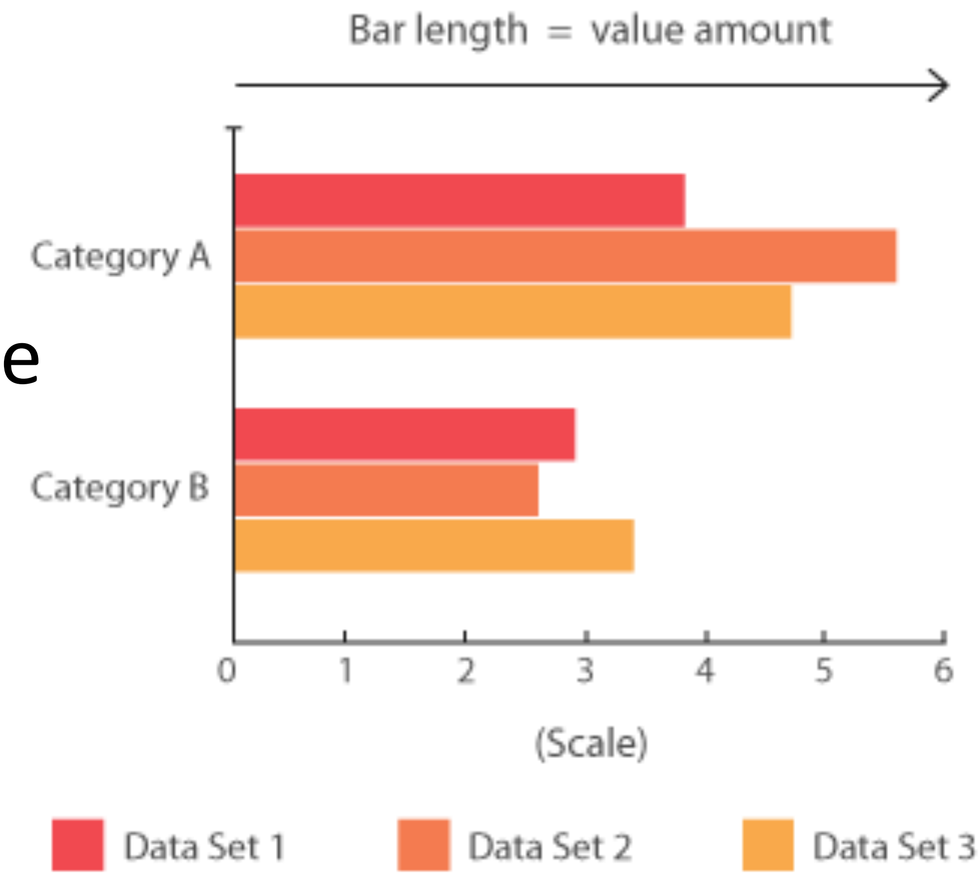
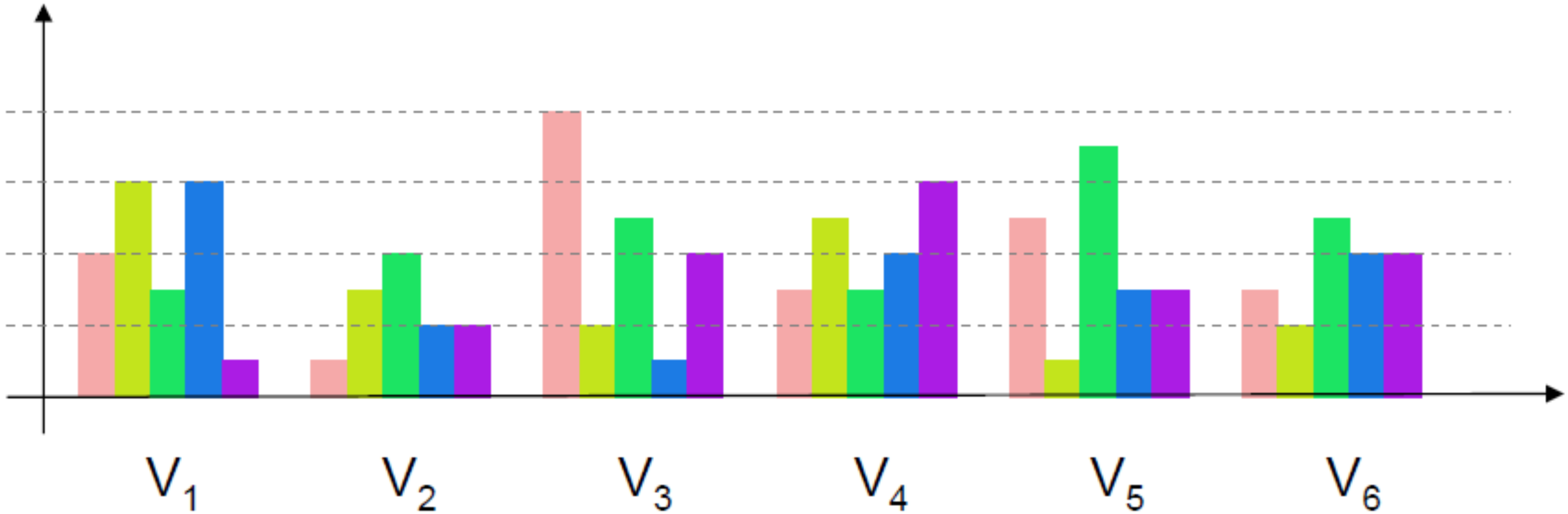


Figure from Data Visualization Catalogue.

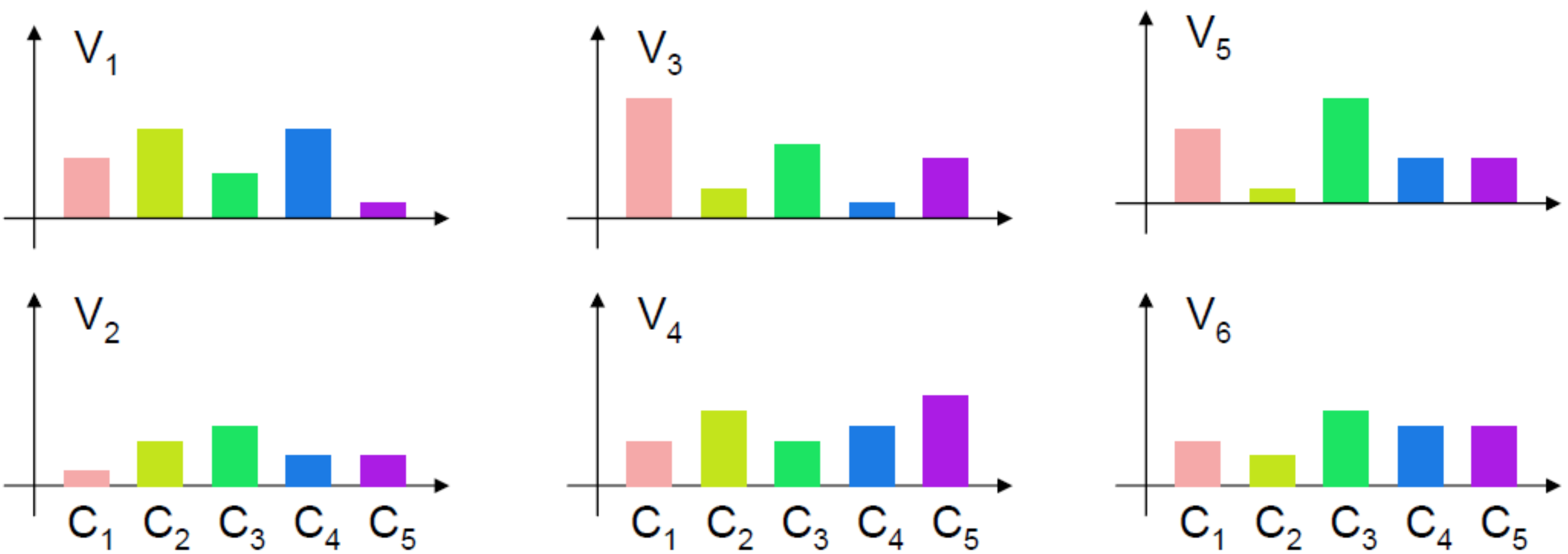
Multiple Bars

- Distributions of each variable among the different categories/data points.



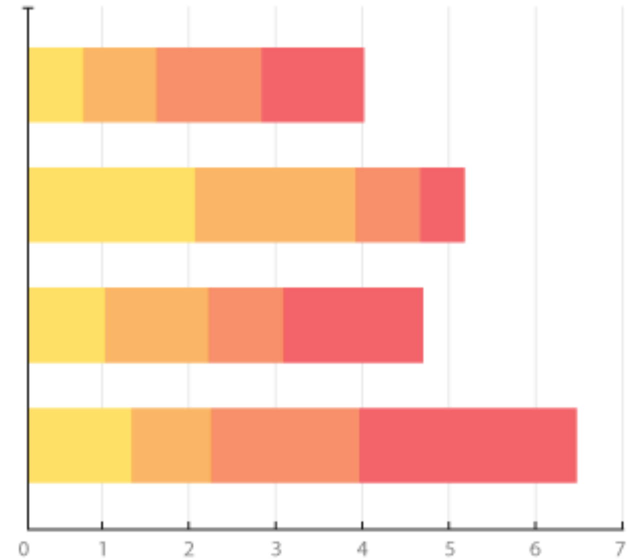
Multiple Views

- Distributions of each variable among the different categories/data points (*each variable has its own display*).



Stacked Bar Charts

- Each dataset is drawn on top of each other.
- Two types:
  - *Simple Stacked Bar Charts*
  - *Percentage Stacked Bar Charts*.
- More segments each bar more difficult to read.



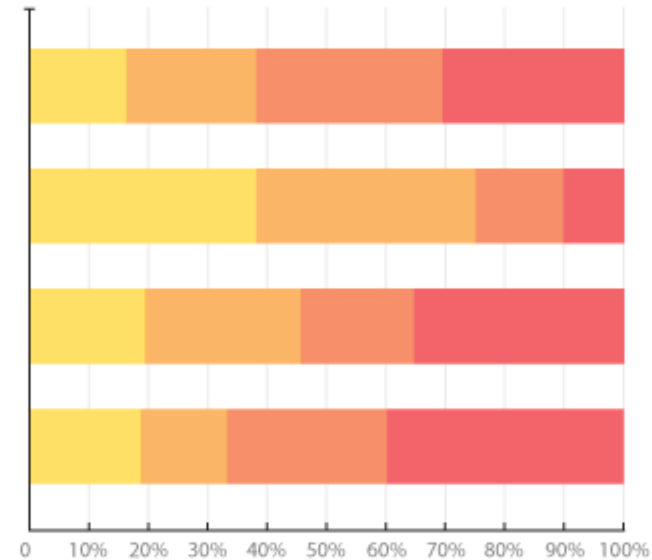


Figure from Data Visualization Catalogue.

Simple vs Percentage Bar Charts

- Simple Stacked Bar Charts
  - Useful if the visualization of the absolute values (and their sum) is meaningful.
- Percentage Bar Charts
  - Better to show the relative differences between quantities in the different groups.

Spineplots

- Generalization of stacked bar charts.
- Special case of *mosaic plot*.
- Permit to show both percentages and proportions between variables.

Spineplots – Example

- Car data:

Nationality	Repair records	# models
domestic	1	2
domestic	2	8
domestic	3	27
domestic	4	9
domestic	5	2
foreign	1	0
foreign	2	0
foreign	3	3
foreign	4	9
foreign	5	9

Spineplots – Example

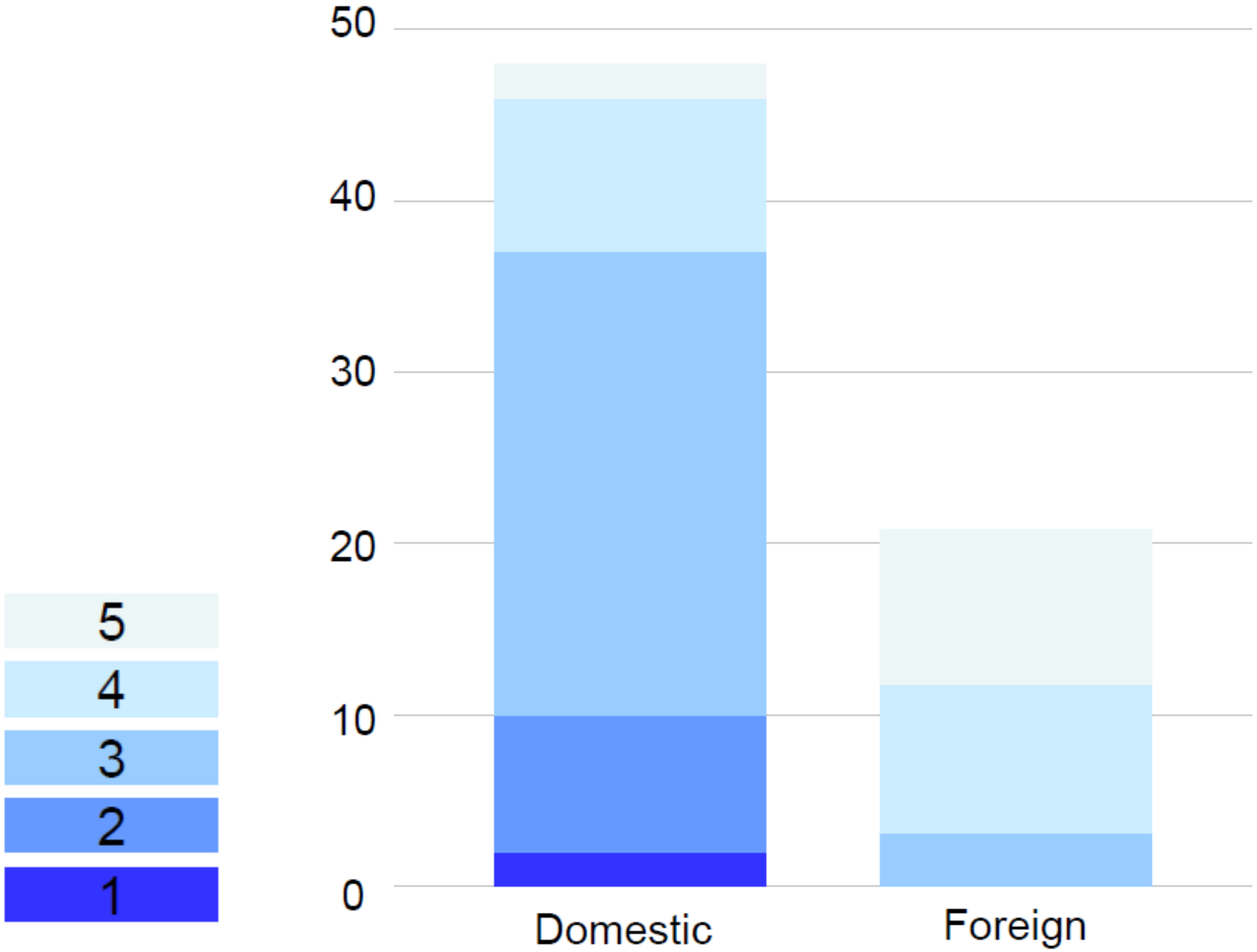


Figure from *Information Visualization Course*, Università Roma 3.

Spineplots – Example

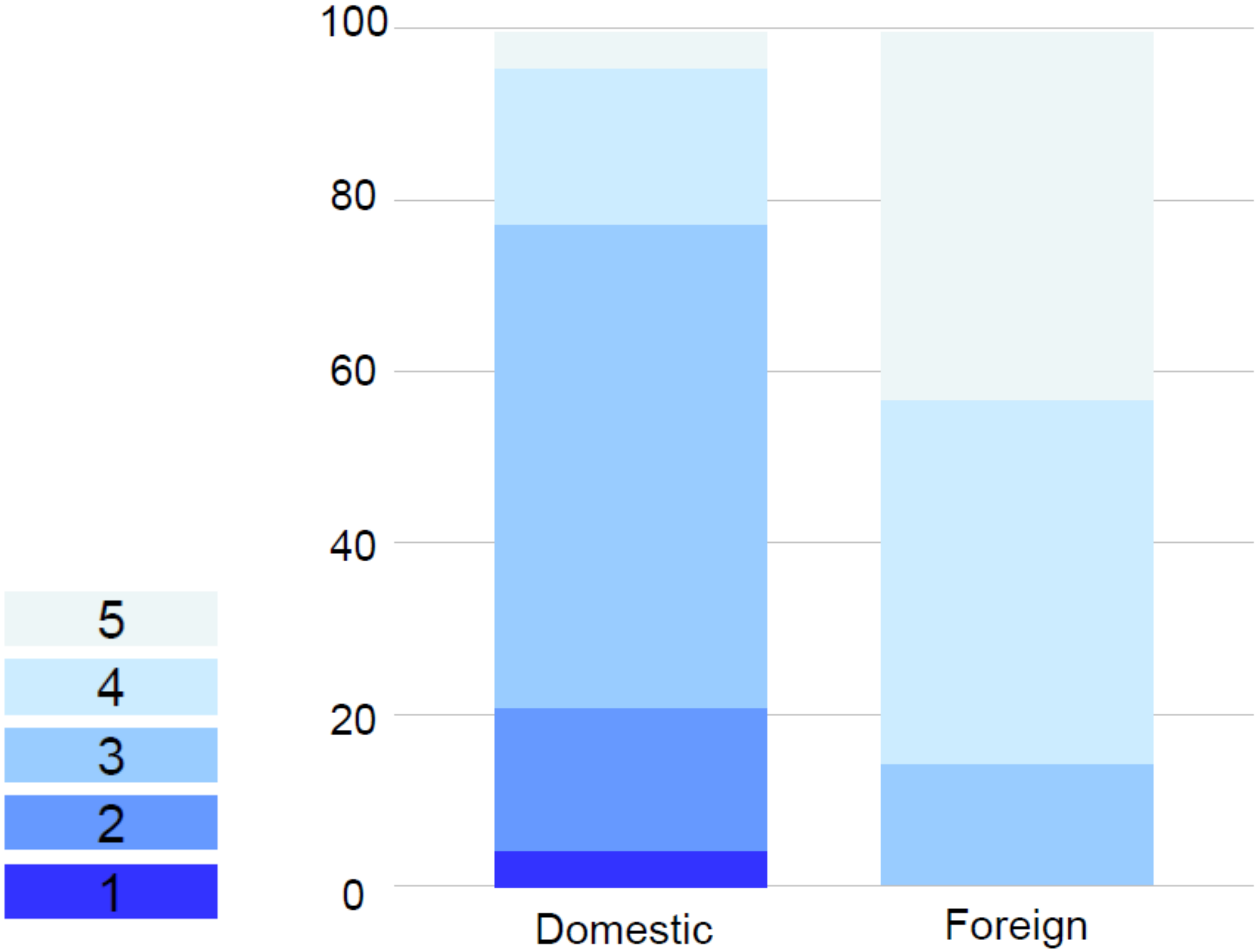


Figure from *Information Visualization Course*, Università Roma 3.

Spineplots – Example

- Convey both percentages and proportions.

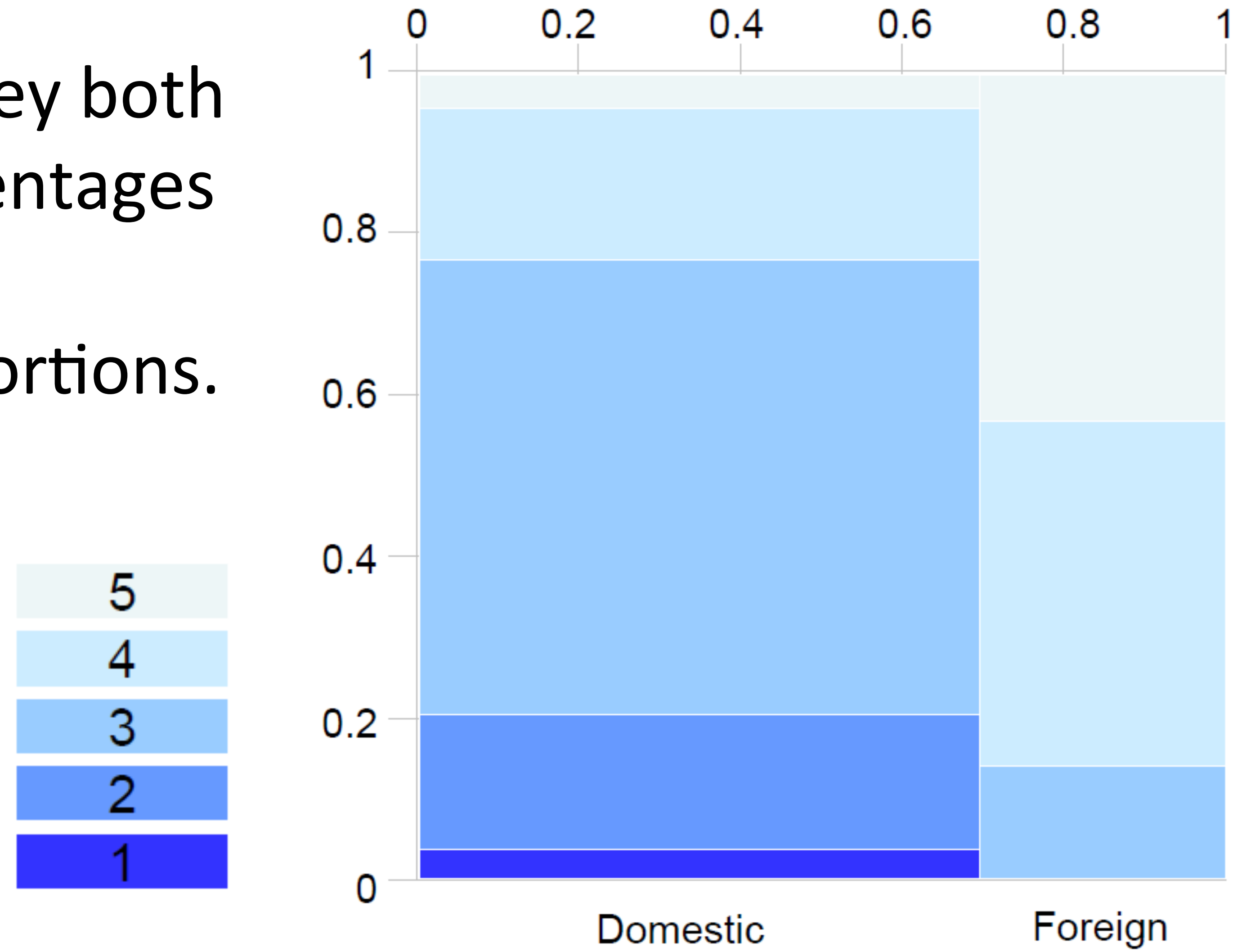
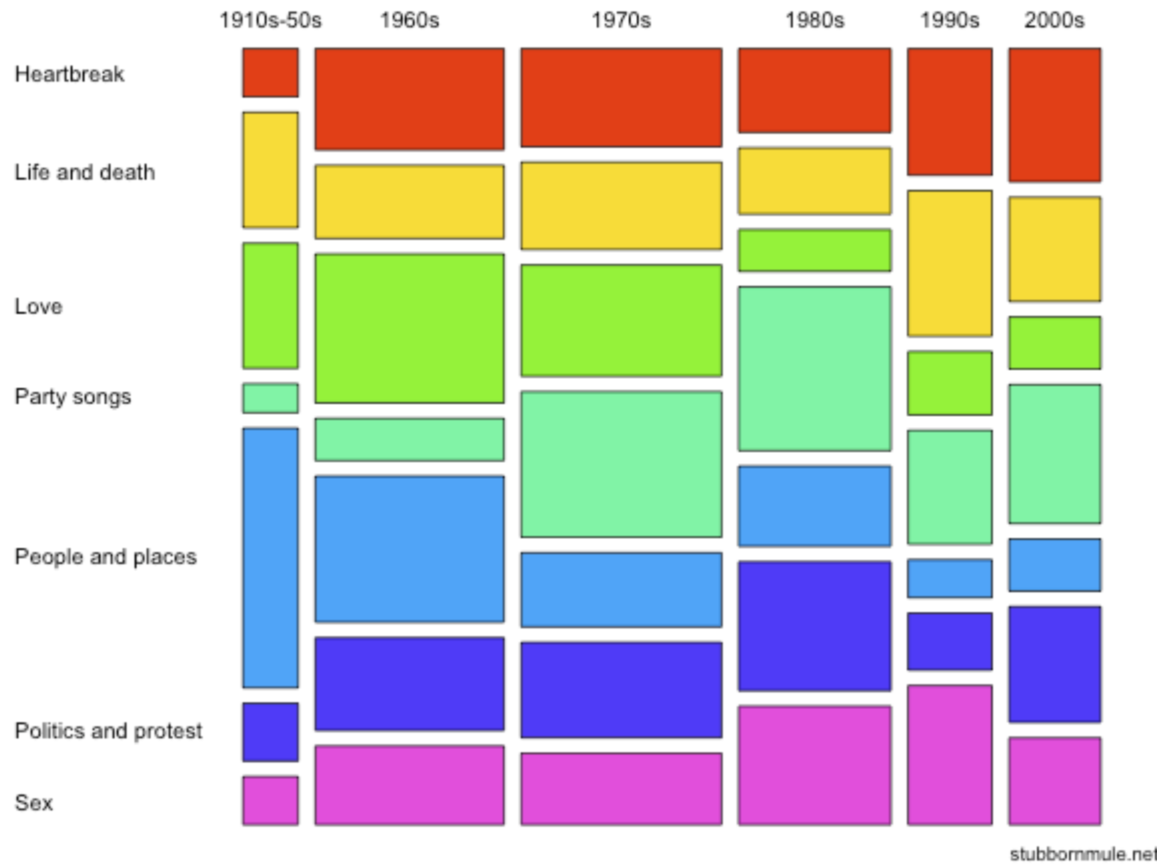


Figure from *Information Visualization Course*, Università Roma 3.

Mosaic Plots

- They give an overview of the data by visualizing the relative proportions.
- Also known as Mekko charts.



Mosaic Plots

- Titanic data (from Wikipedia):

Gender	Survived	1° class	2° class	3° class	Crew
Male	No	118	154	422	670
Male	Yes	62	25	88	192
Female	No	4	13	106	3
Female	Yes	141	93	90	20

Mosaic Plots

- Variable *gender* → vertical axis.

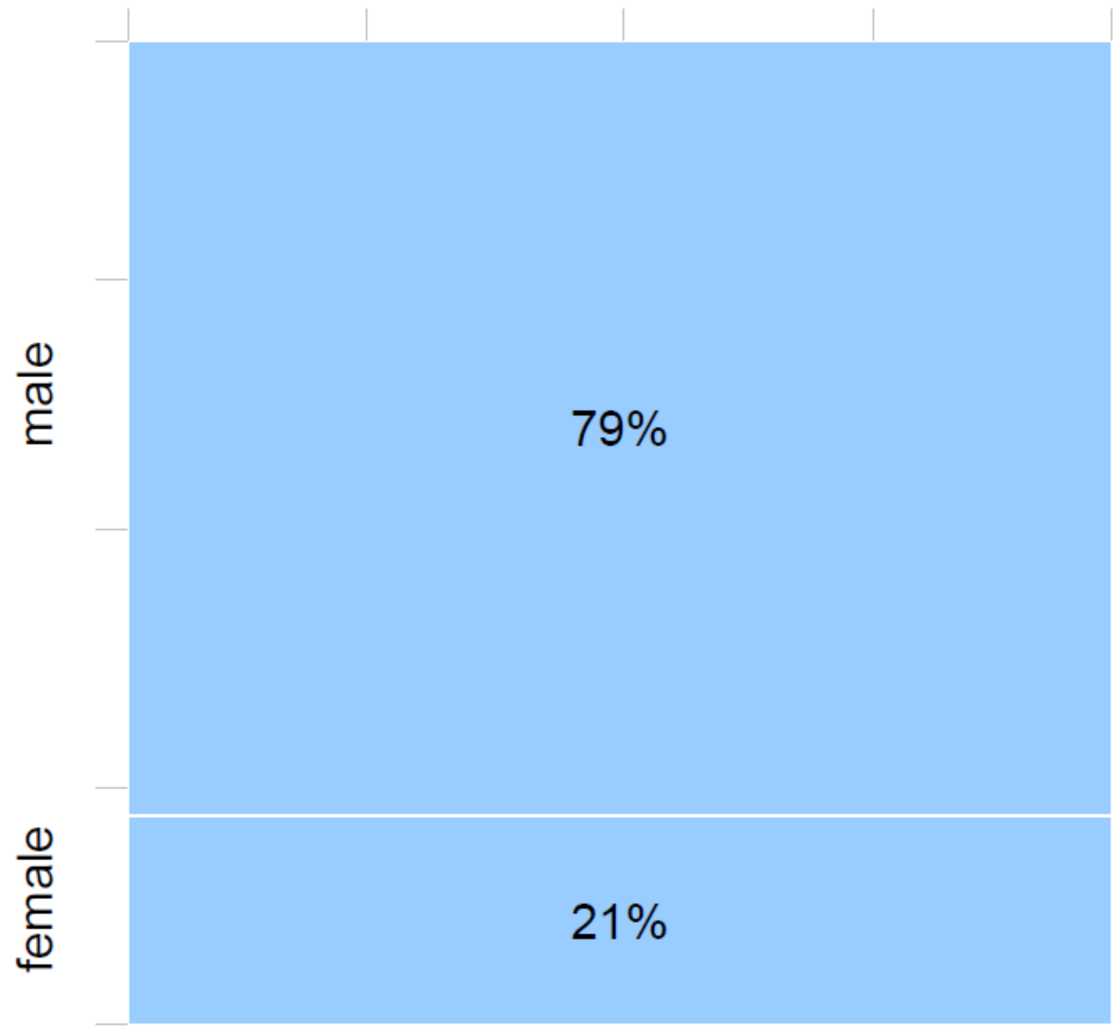


Figure from *Information Visualization Course*, Università Roma 3.

Mosaic Plots

- Add variable *Class* → Horizontal axis.

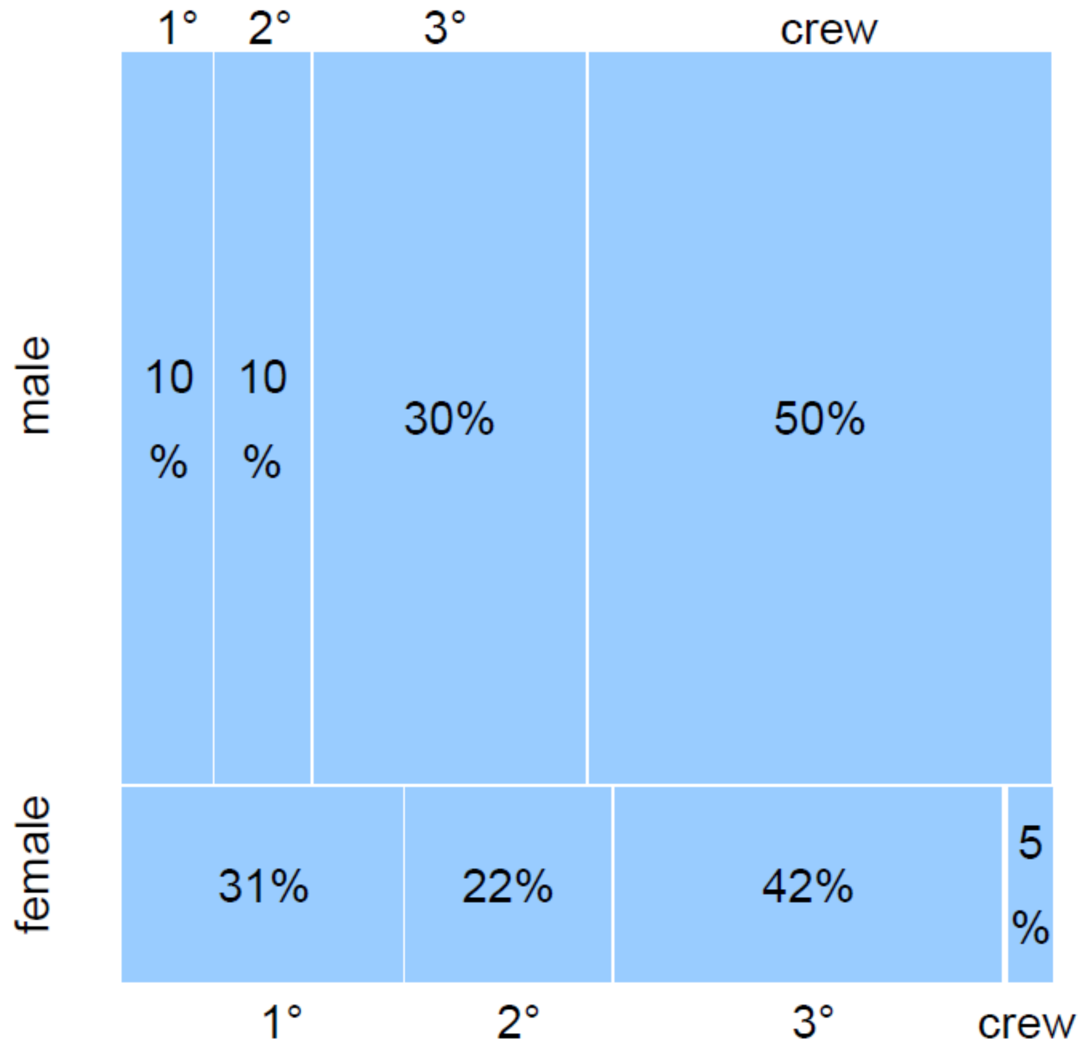


Figure from *Information Visualization Course*, Università Roma 3.

Mosaic Plots

- Add variable *survived* → vertical axis.

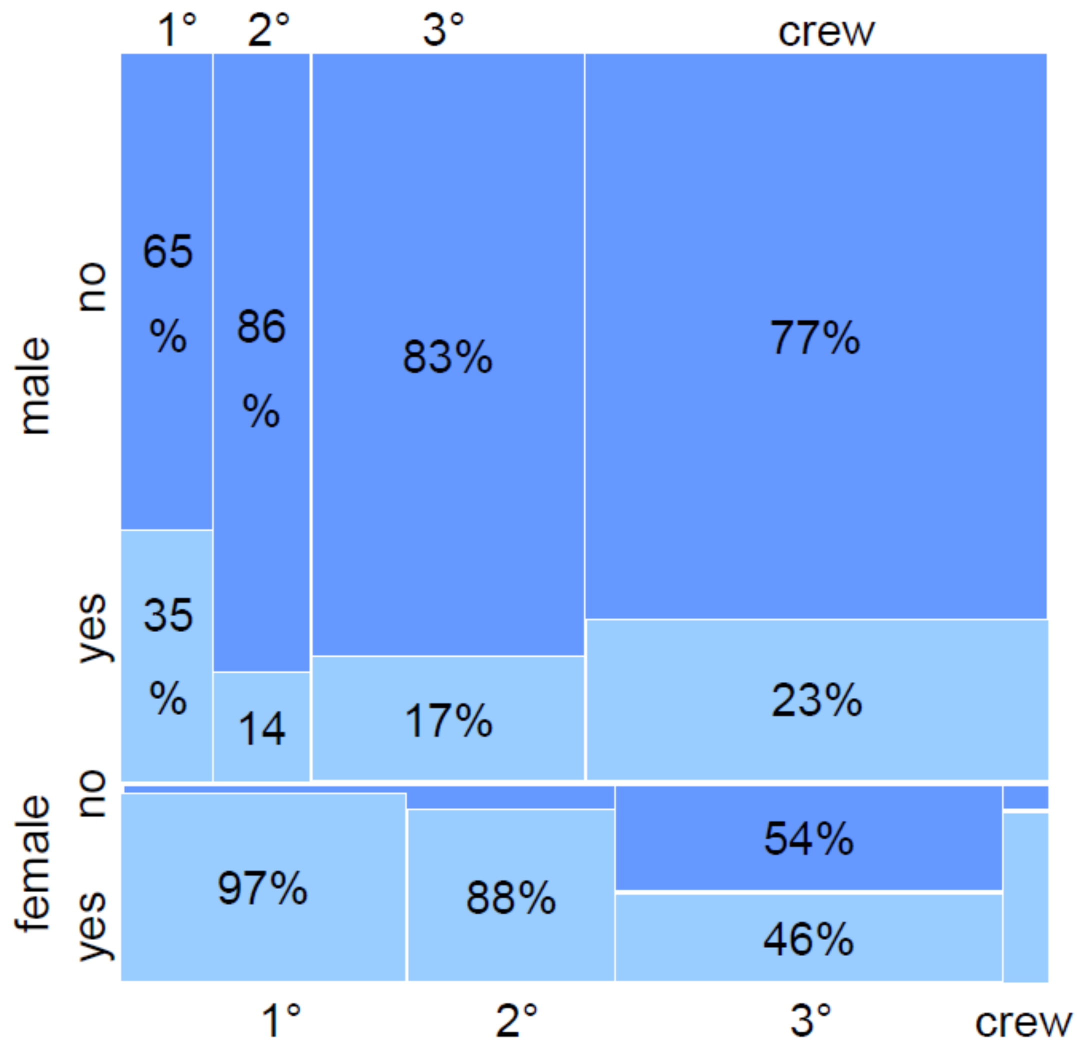


Figure from *Information Visualization Course*, Università Roma 3.

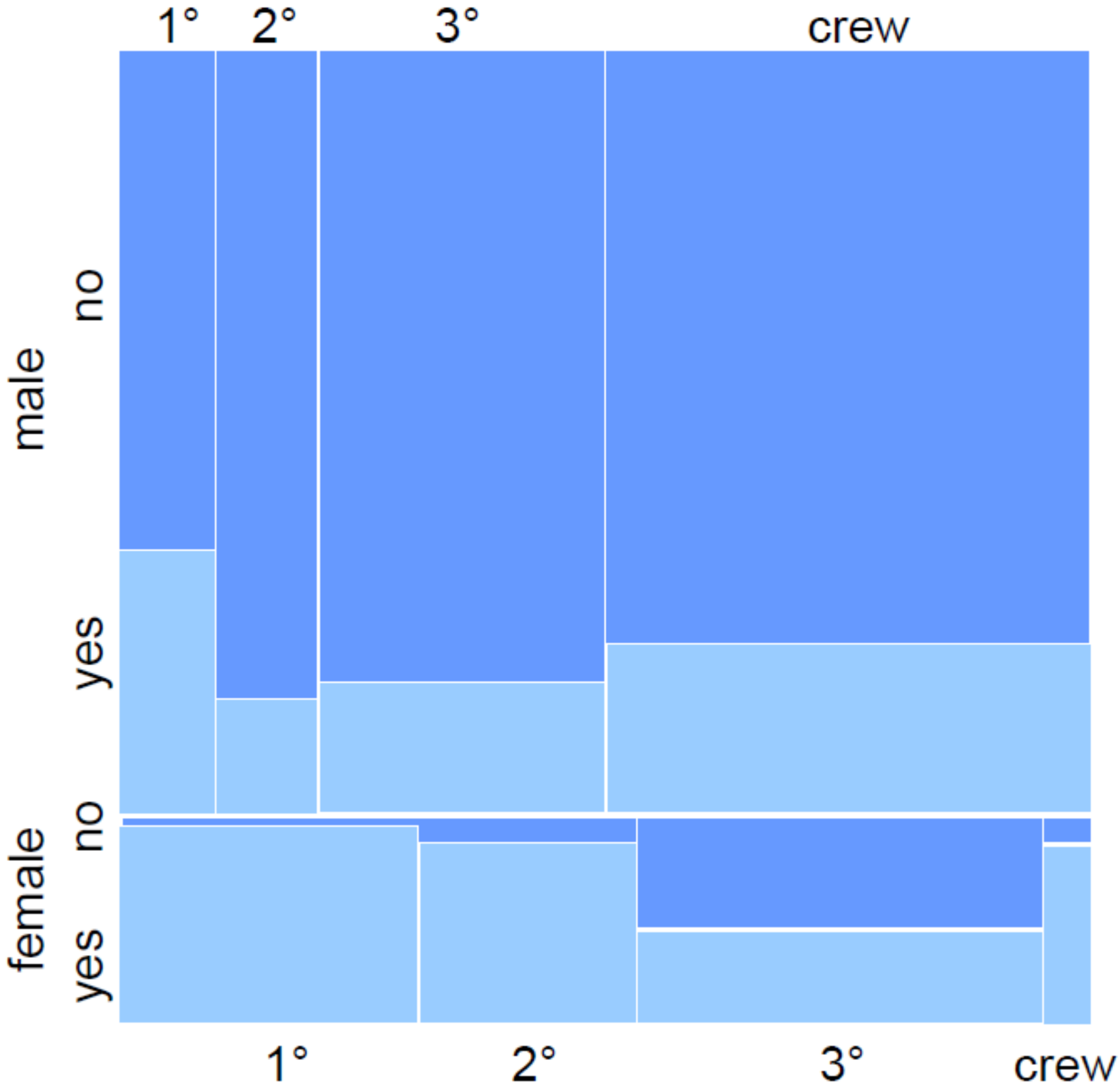


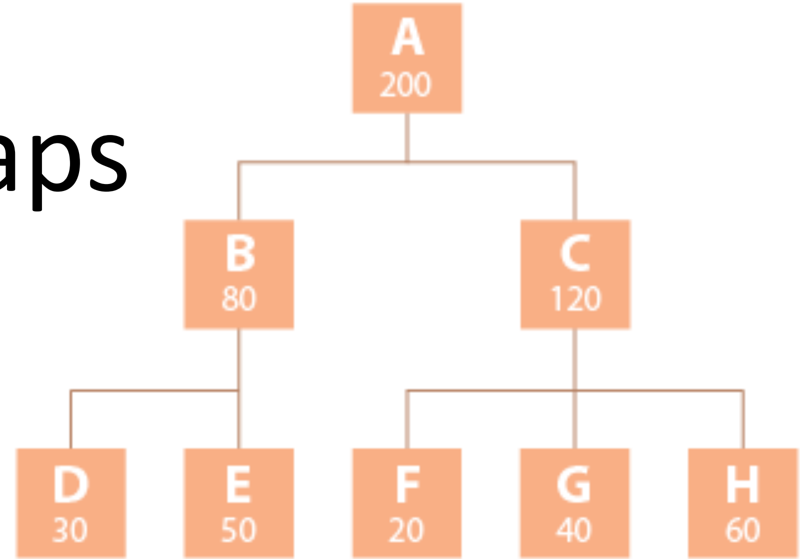
Figure from *Information Visualization Course*, Università Roma 3.

Mosaic Plots

- Advantages:
  - Maximal use of the available space
  - Good overview of the proportions between data
  - Good overview of the variable dependency
- Disadvantages:
  - Extension to many variables is difficult

Tree Maps

- An alternative way to visualize a tree data structure.
- Space-efficient (!)
- The way rectangles are divided and ordered into sub-rectangles is dependent on the tiling algorithm used.



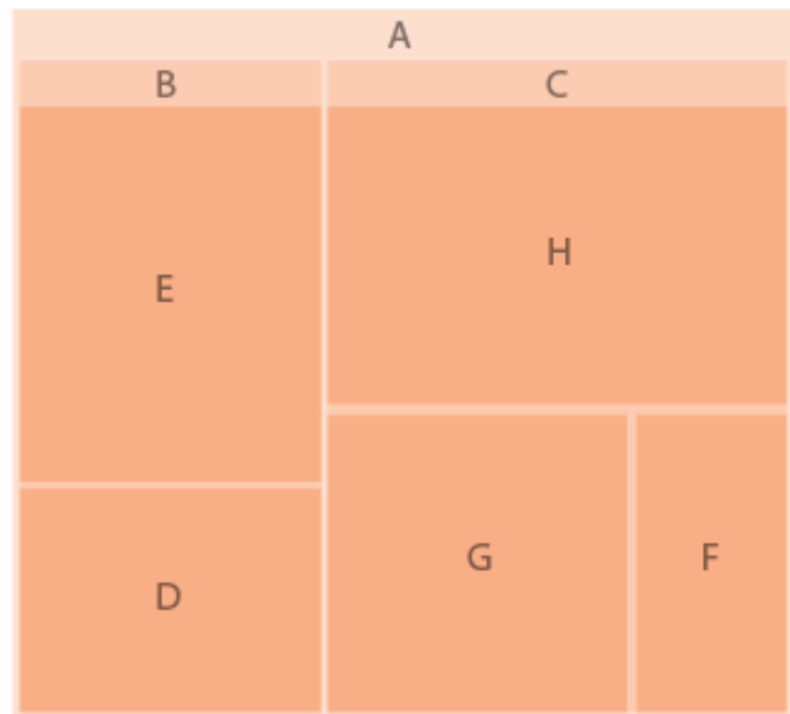


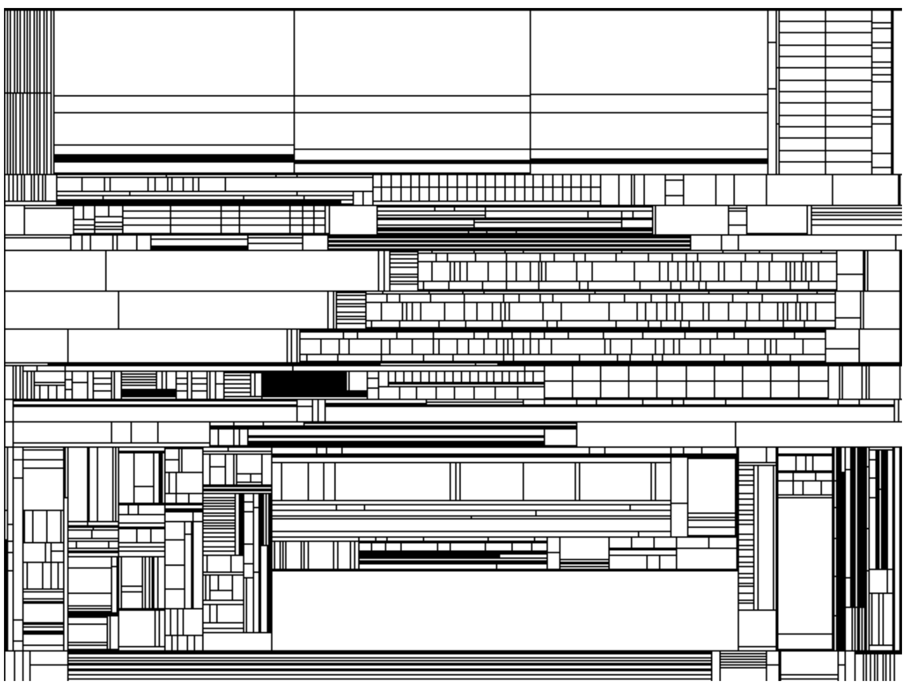
Figure from Data Visualization Catalogue.

Tree Maps – Example

Figure from Data Visualization Catalogue.

Cushion Treemaps

- Use suitable shading to reveal the fine structure of the hierarchy.



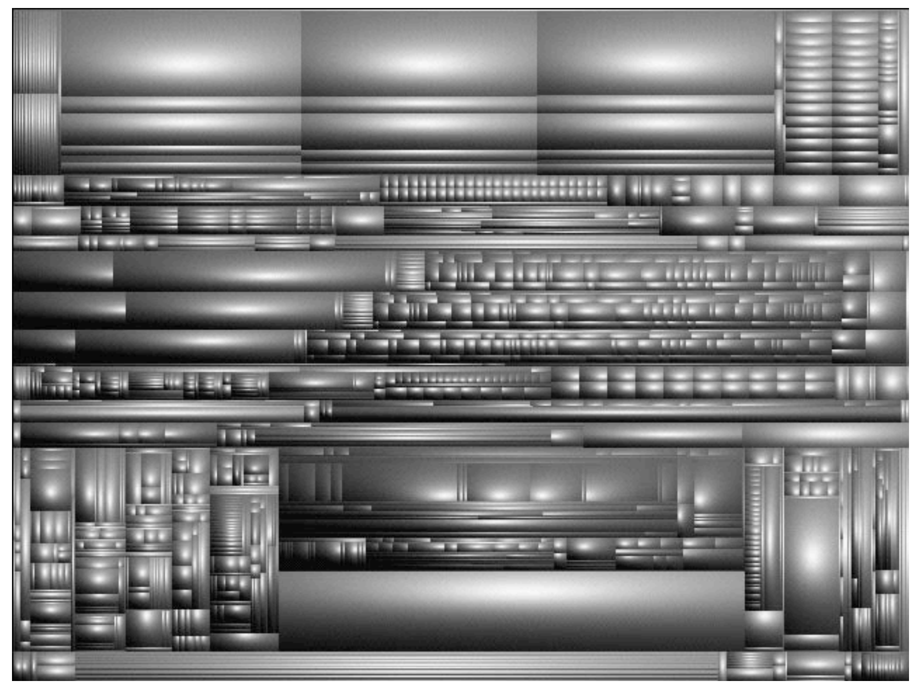


Figure from Jarke J. van Wijk Huub van de Wetering, *"Cushion Treemaps: Visualization of Hierarchical Information"*, InfoVis'99.

Cushion Treemaps

- Shading function – 1D case:

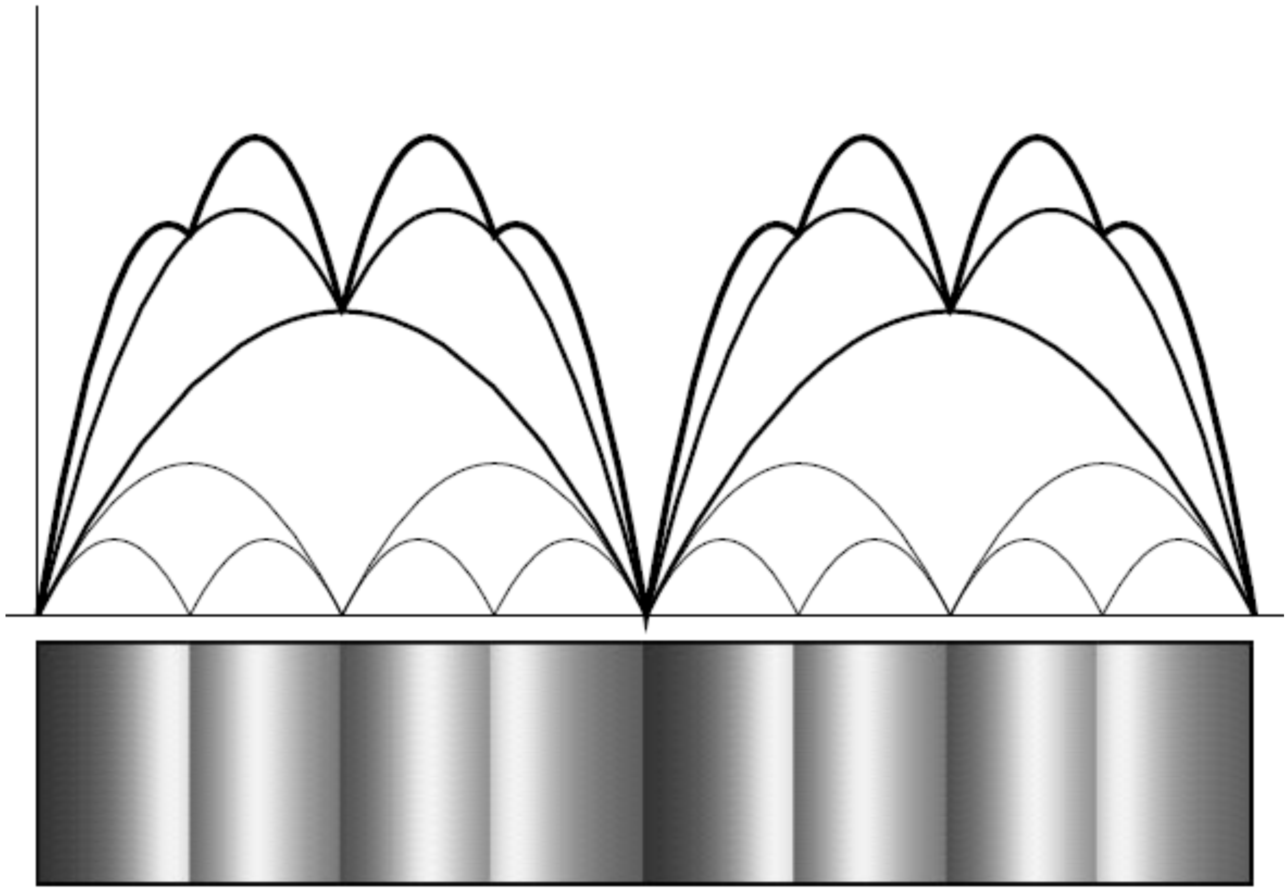
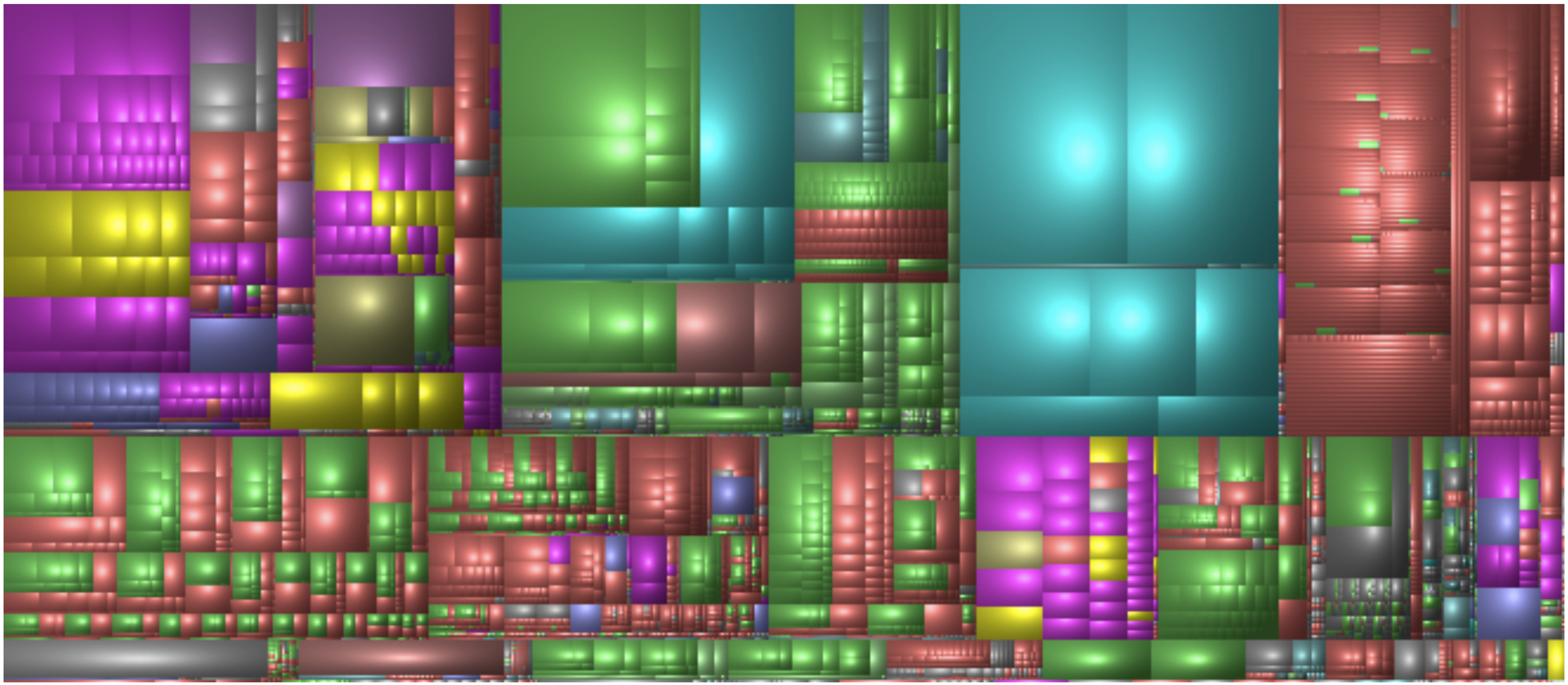


Figure from Jarke J. van Wijk Huub van de Wetering*, "Cushion Treemaps: Visualization of Hierarchical Information"*, InfoVis'99.

Cushion Treemaps



Squarified Treemaps

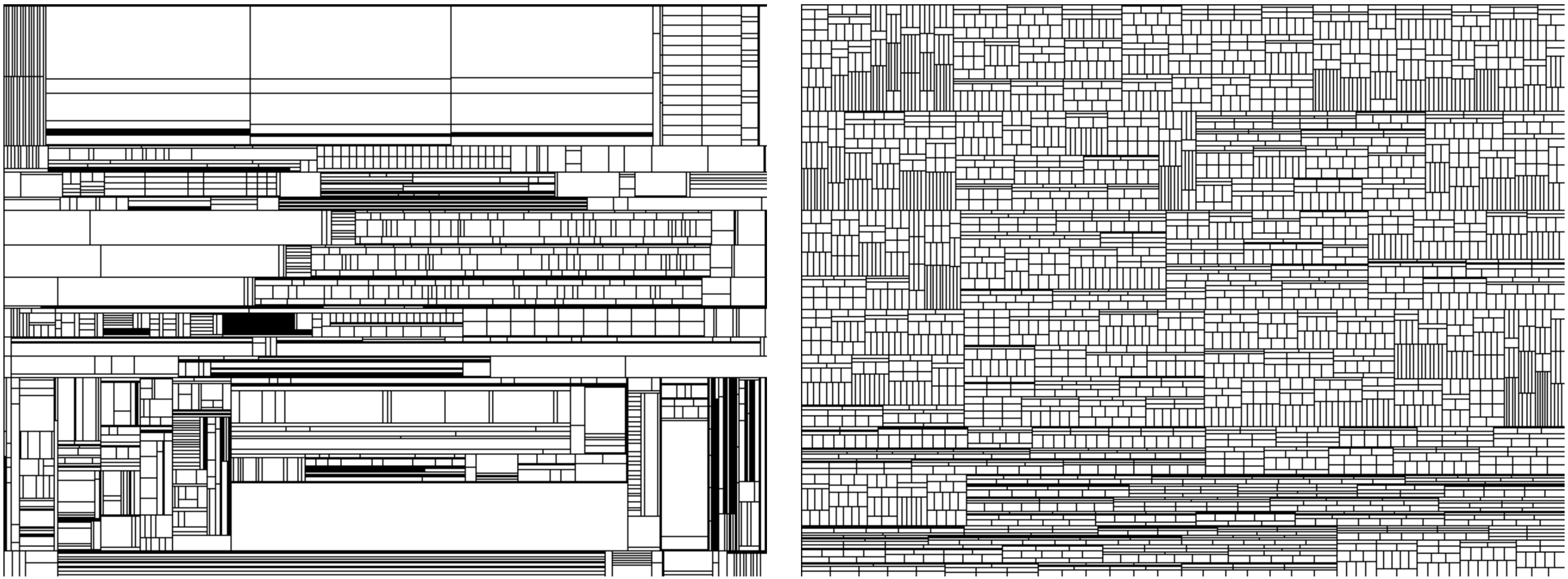


Figure from Mark Bruls, Kees Huizing, and Jarke J. vanWijk, *“Squarified Treemaps”*, Proc. Joint Eurographics and IEEE TCVG Symp. on Visualization.

Squarified Treemaps

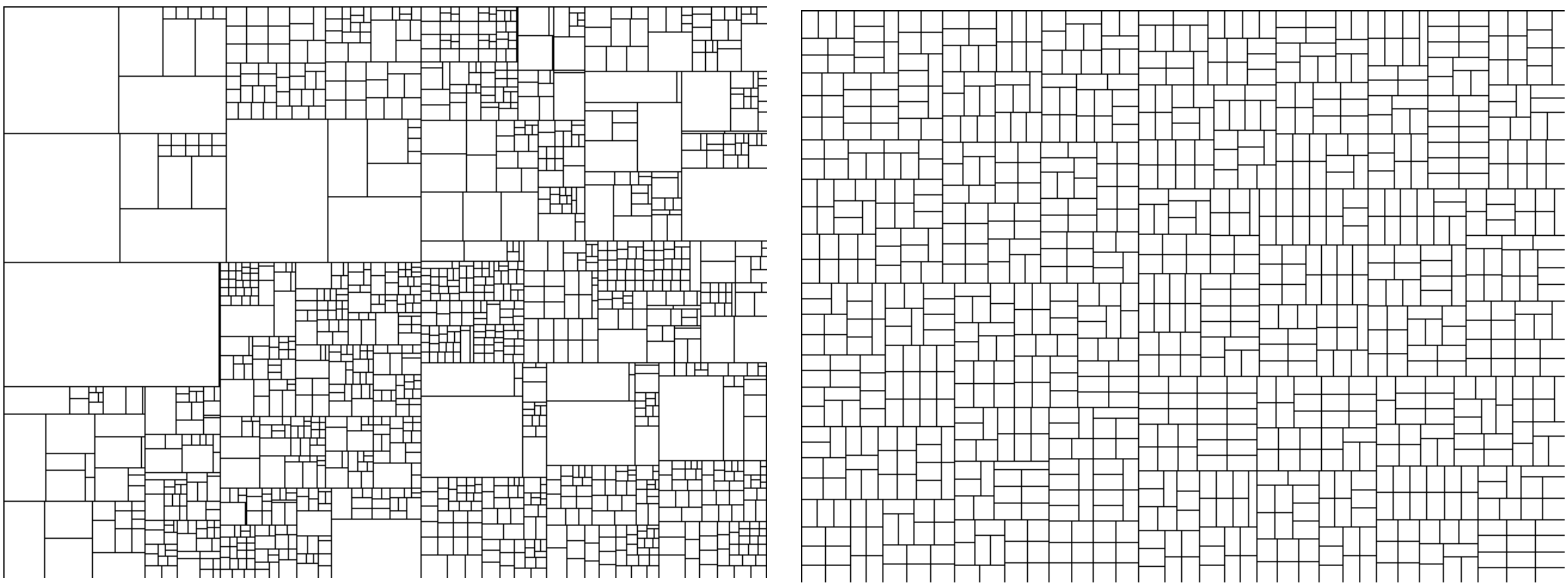
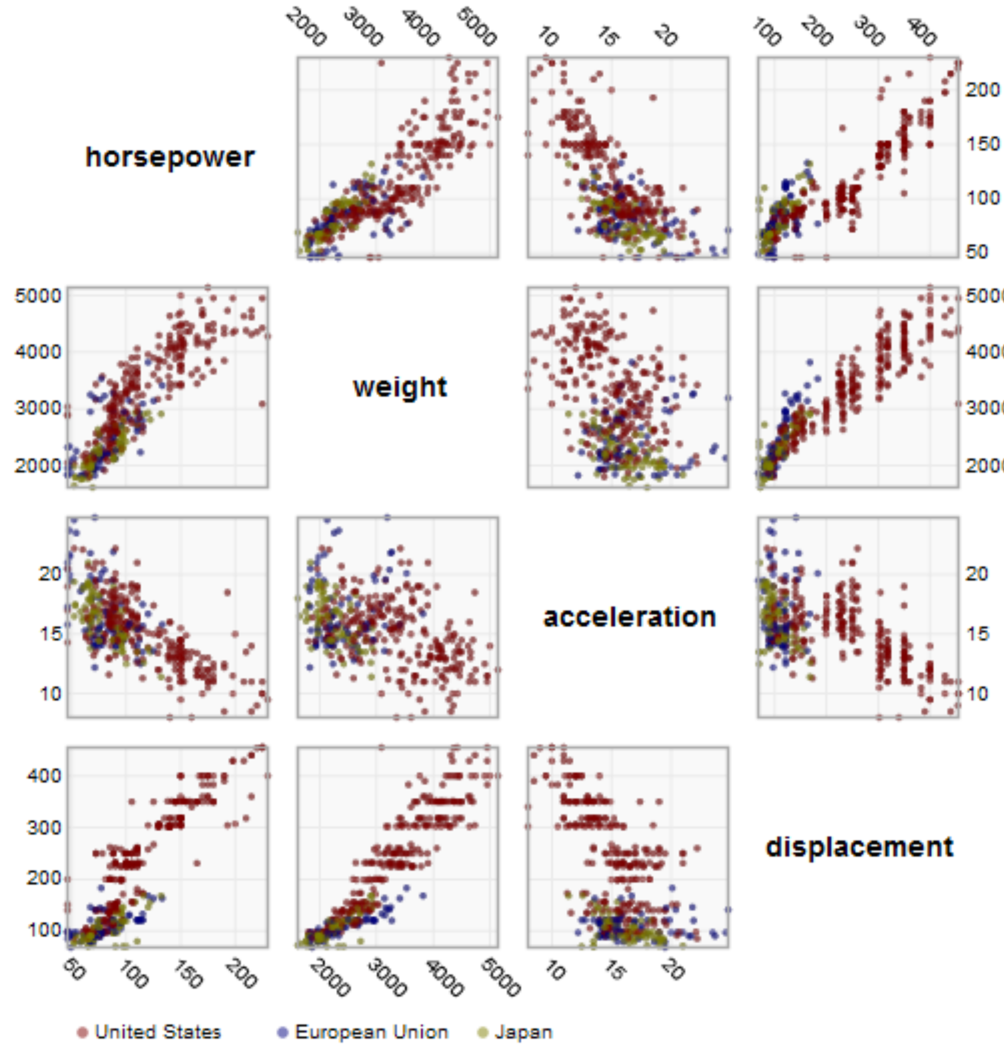


Figure from Mark Bruls, Kees Huizing, and Jarke J. vanWijk, *"Squarified Treemaps"*, Proc. Joint Eurographics and IEEE TCVG Symp. on Visualization.

Scatter Plot Matrix (SPLOM)

- Each possible pair of variables is represented in a standard 2D scatter plot.
- Many works in literature to find efficient way to present them.



SPLOM

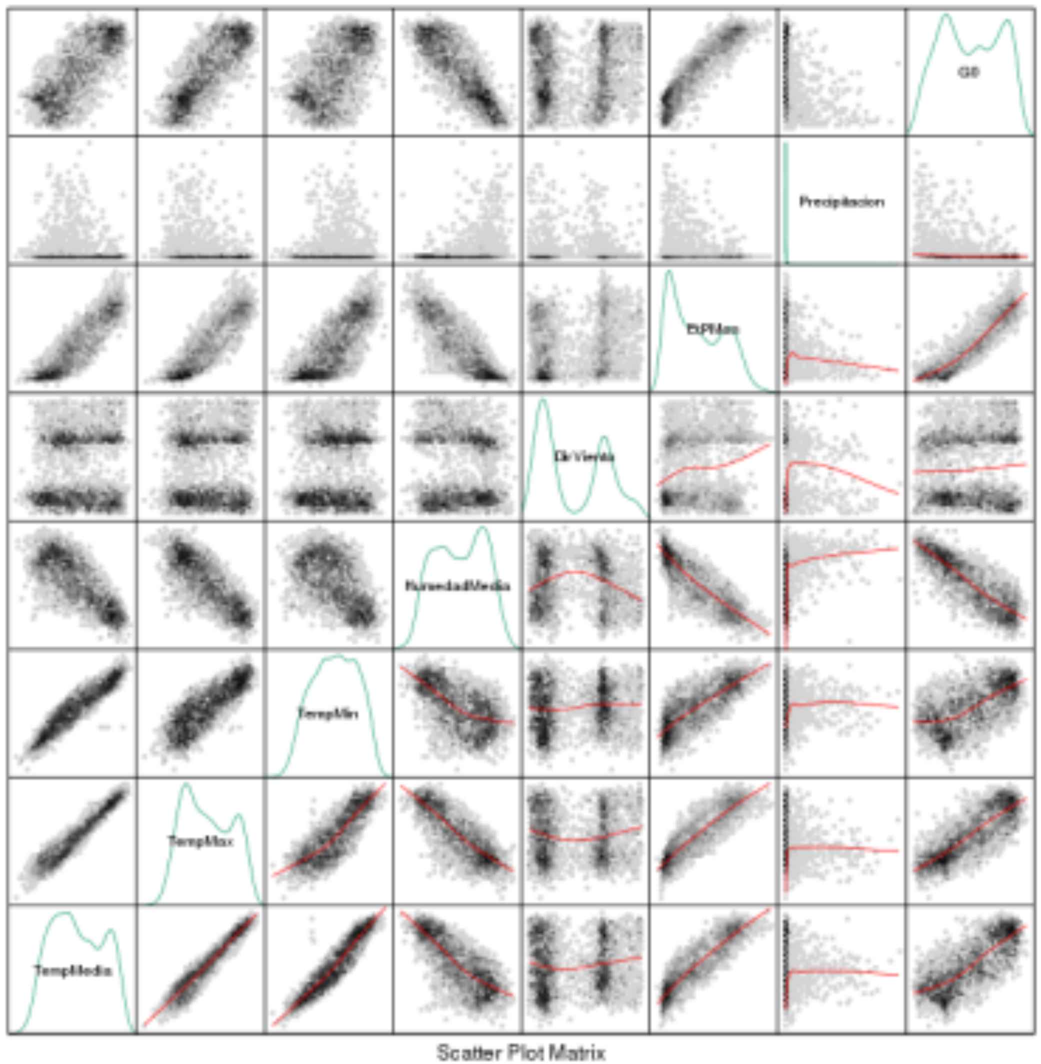
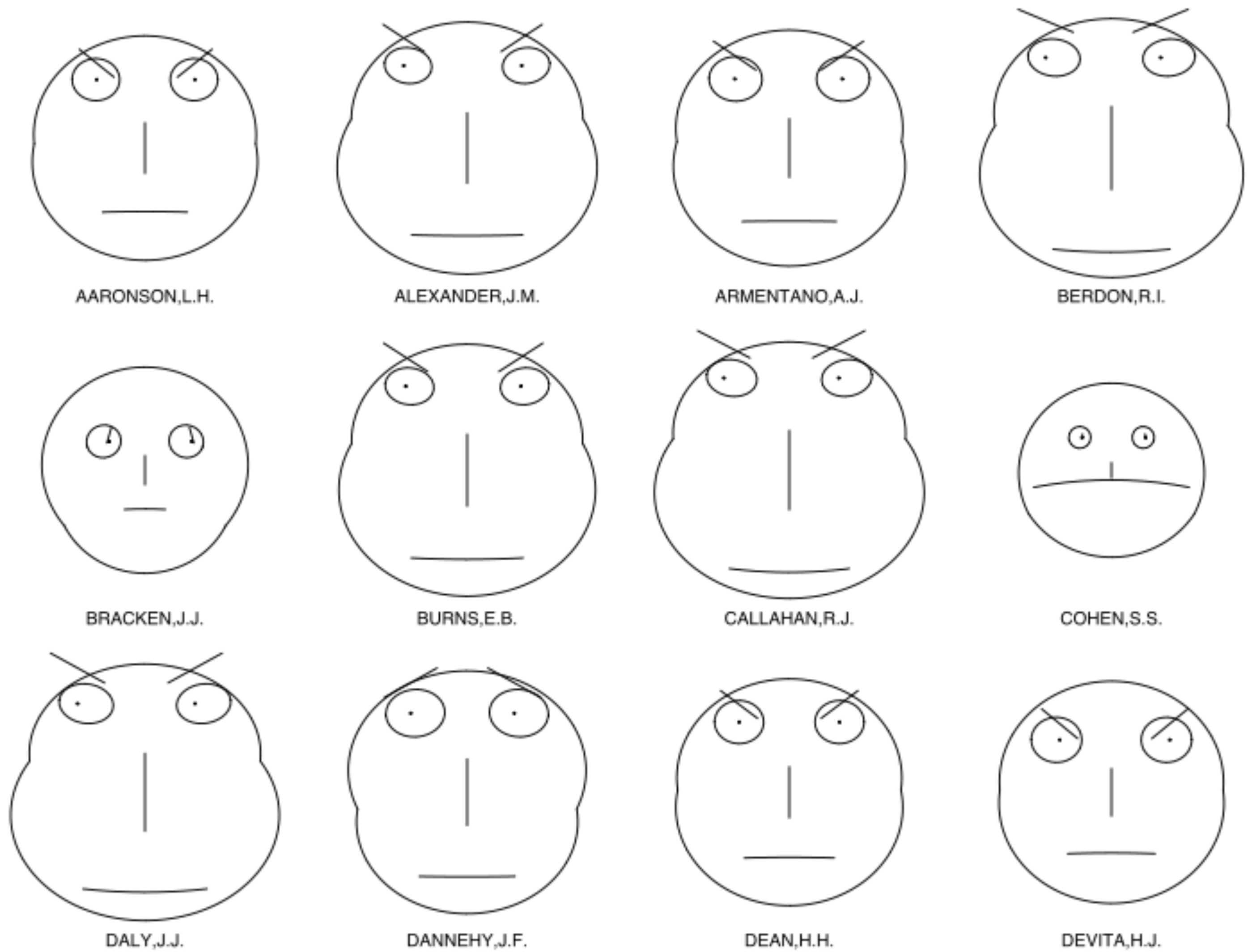


Figure from R-Blogger.

Chernoff Faces

- Introduced by Herman Chernoff in 1973.
- We are very good at recognize faces.
- Variables are mapped on facial features:
  - Width/curvature of mouth
  - Vertical size of the face
  - Size/slant/separation of eyes
  - Size of eyebrows
  - Vertical position of eyebrows

Chernoff Faces (lawyers judge)

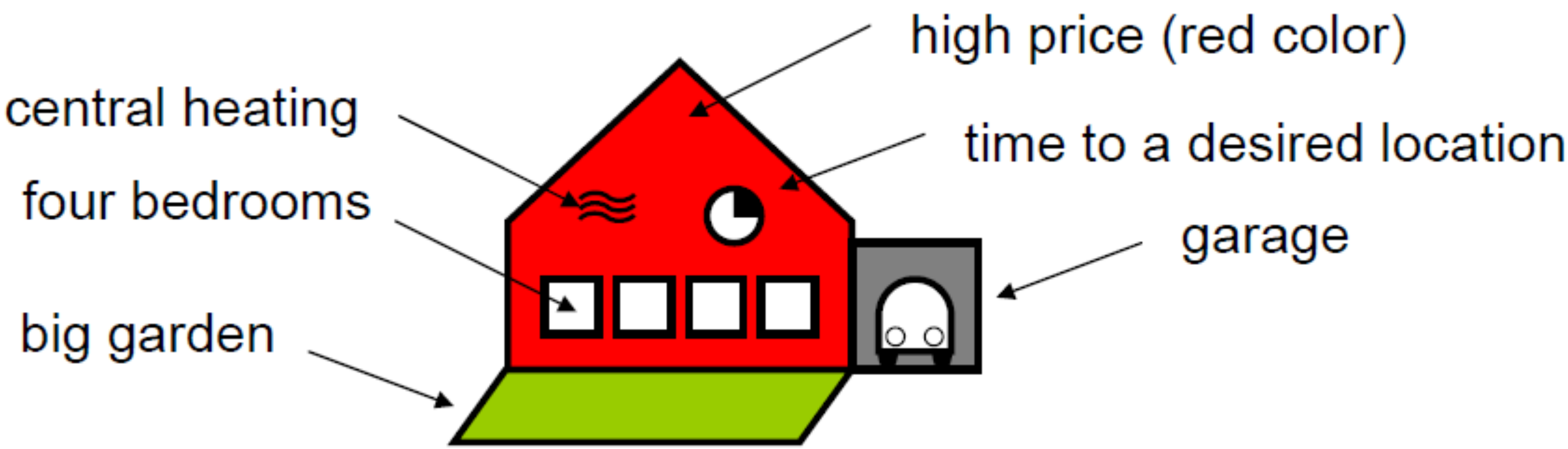


Chernoff Faces

- Chernoff faces received some criticism
  - Difficult to compare.
  - A legend is necessary.
- Moreover, the mapping should be choose carefully to work properly.

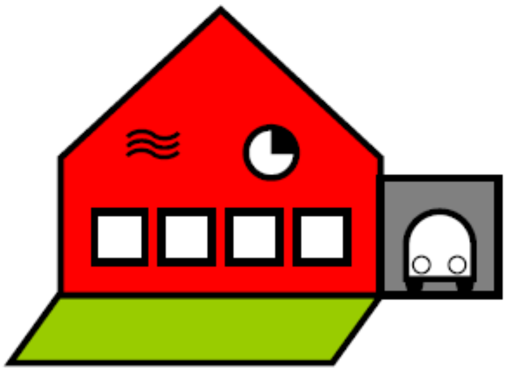
Multi-dimensional Icons

- Spence and Parr (1991) proposed to encode properties of an object in simple iconic representation and assemble them together.
- They applied this approach to check quickly dwell offers.

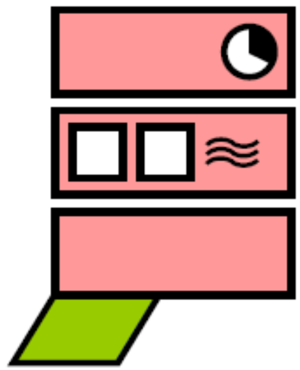


Robert Spence and Maureen Parr, *"Cognitive assessment of alternatives"*, Interacting with Computers 3, 1991.

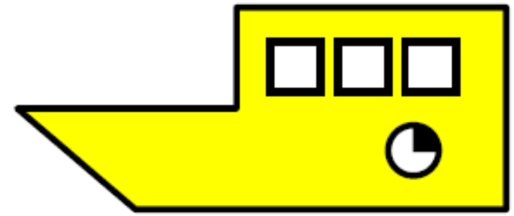
Spence and Parr (1991) (examples)



house
£400,000
garage
central heating
four bedrooms
good repair
large garden
Victoria 15 mins



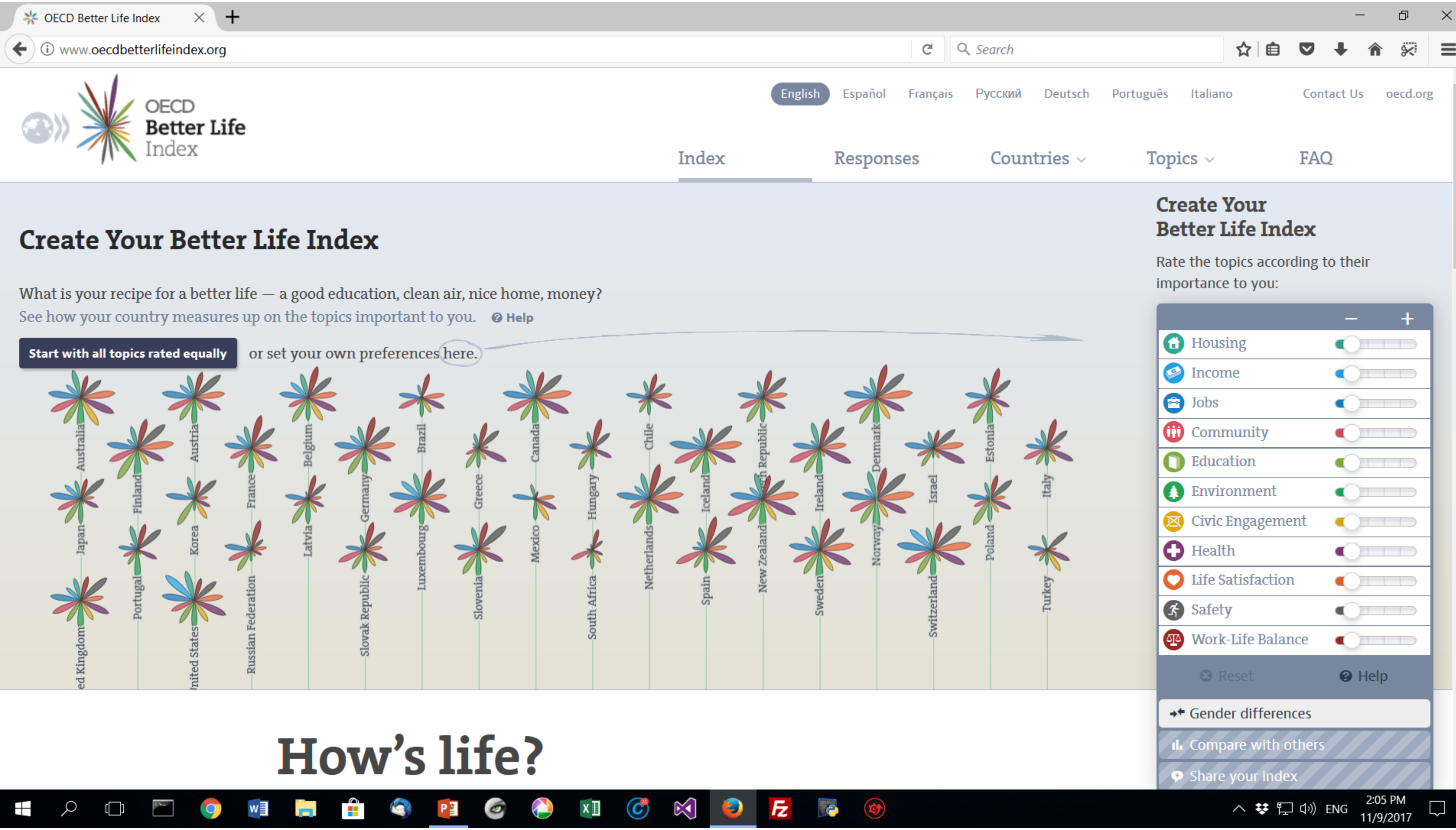
flat
£300,000
no garage
central heating
two bedrooms
poor repair
small garden
Victoria 20 mins



houseboat
£200,000
no garage
no central heating
three bedrooms
good repair
no garden
Victoria 15 mins

Petals as a Glyph

- The idea of Moritz Stefaner to visualize a life index is to map several variables into *petals of different size.*
- Web site: www.oecdbetterlifeindex.org


OECD Better Life Index
www.oecdbetterlifeindex.org
Search
English
Español
Français
Русский
Deutsch
Português
Italiano
Contact Us
oecd.org
OECD Better Life Index
Index
Responses
Countries
Topics
FAQ
Create Your Better Life Index
What is your recipe for a better life — a good education, clean air, nice home, money?
See how your country measures up on the topics important to you.
Help
Start with all topics rated equally
or set your own preferences here.
Australia
Japan
Finland
Portugal
Austria
Korea
France
Russian Federation
Belgium
Latvia
Germany
Slovak Republic
Brazil
Luxembourg
Greece
Slovenia
Canada
Mexico
Hungary
South Africa
Chile
Netherlands
Iceland
Spain
New Zealand
Ireland
Sweden
Denmark
Norway
Israel
Switzerland
Estonia
Poland
Italy
Turkey
Create Your Better Life Index
Rate the topics according to their importance to you:
Housing
Income
Jobs
Community
Education
Environment
Civic Engagement
Health
Life Satisfaction
Safety
Work-Life Balance
Reset
Help
Gender differences
Compare with others
Share your index
How's life?
ENG
2:05 PM
11/9/2017

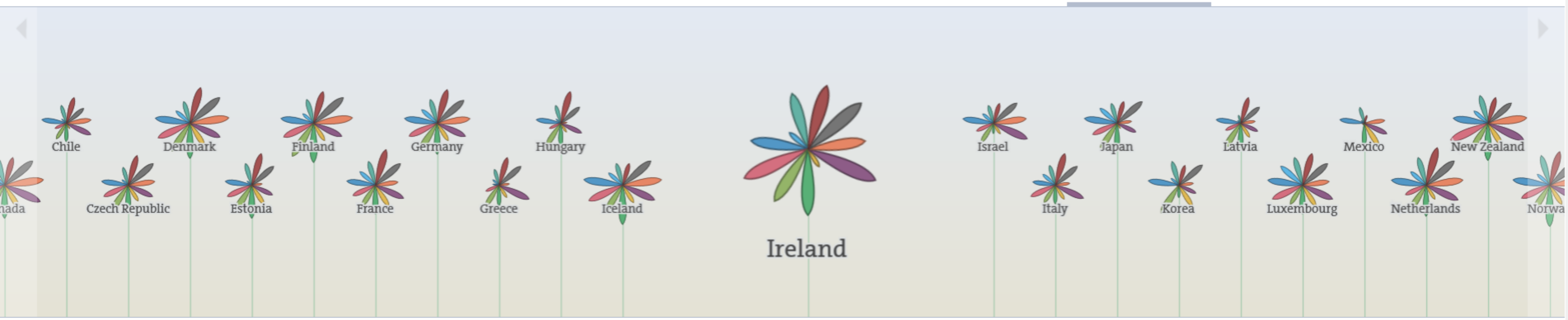
OECD Better Life Index

www.oecdbetterlifeindex.org/countries/ireland/

OECD Better Life Index

English Español Français Русский Deutsch Português Italiano Contact Us oecd.org

Index Responses Countries Topics FAQ



Ireland

→ Learn even more about Ireland at oecd.org

Did you know?

Population	4.6 mil.
Visitors per year	6.0 mil.
Renewable energy	6.1 %

How's Life?

Ireland performs well in many measures of well-being relative to most other countries in the Better Life Index. Ireland ranks above the average in housing, personal security, health status, education and skills, social

Topics

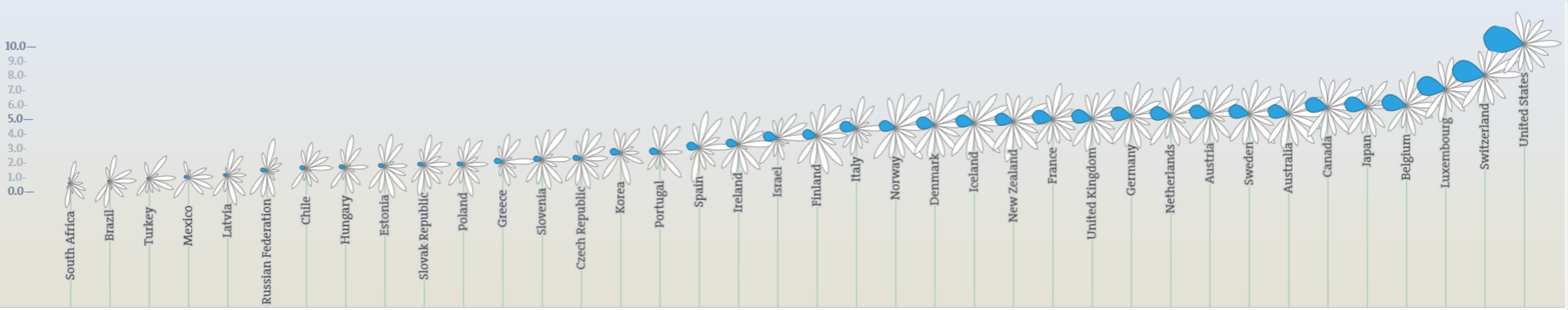
Housing

6.9

OECD in Action

OECD Economic Surveys
IRELAND

ENG 1:59 PM 11/9/2017



Income

Background

While money may not buy happiness, it is an important means to achieving higher living standards and thus greater well-being. Higher economic wealth may also improve access to quality education, health care and housing.

Household net adjusted disposable income

Financial wealth makes up an important part of a household's economic resources, and can protect from economic hardship and vulnerability. For example, a low-income household having above-average wealth will be better off than a low-income household with no wealth at all. Across the

Top Ranking

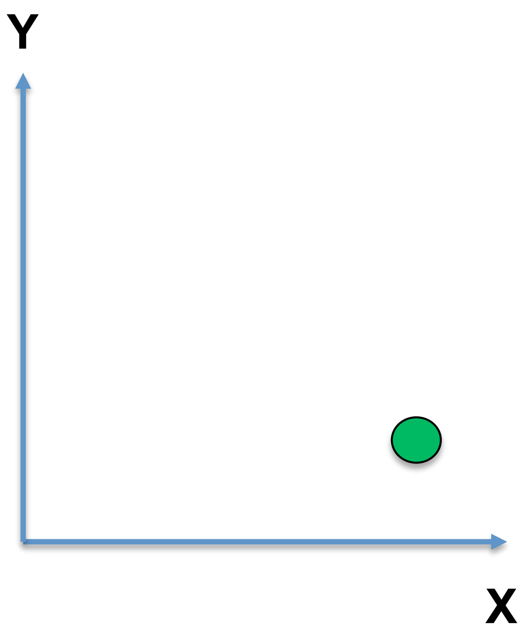
Income

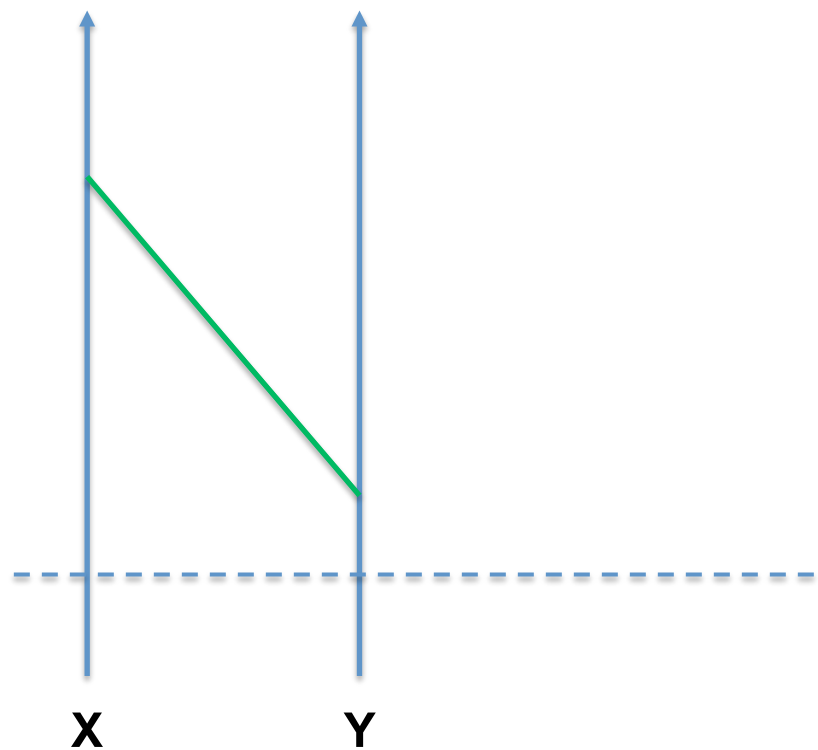
Indicators

Parallel Coordinates

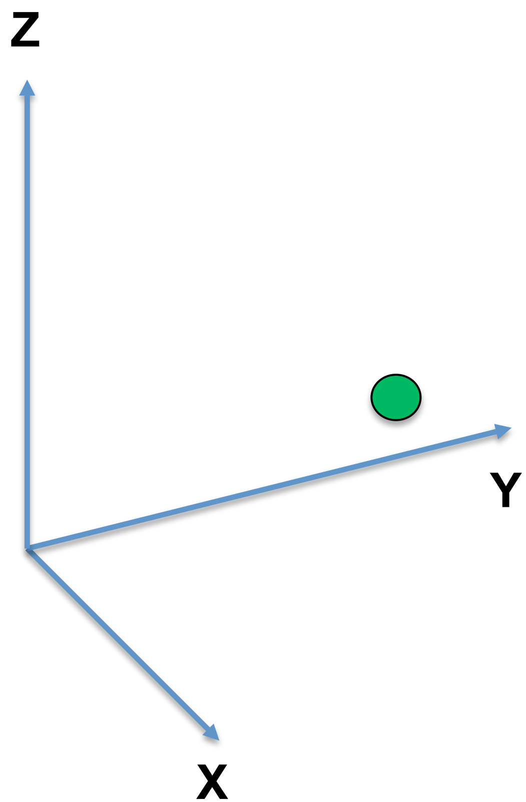
- Originally attributed to Philbert Maurice d'Ocagne (1885).
- Extends classical Cartesian Coordinates System to visualize multivariate data.
- Re-discovered and popularized by Alfred Isenberg in 1970s.

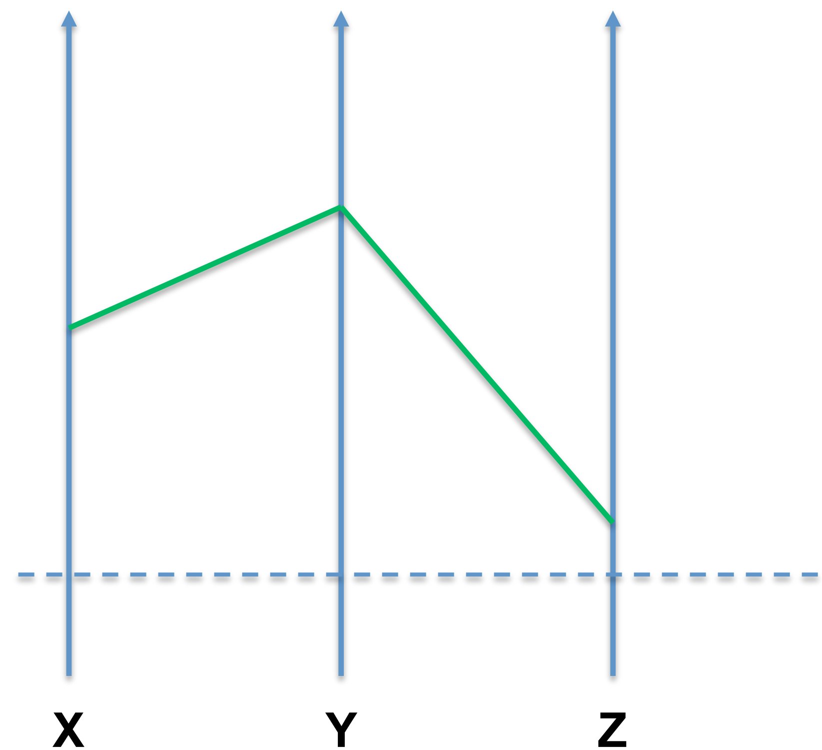
Parallel Coordinates
Two variables



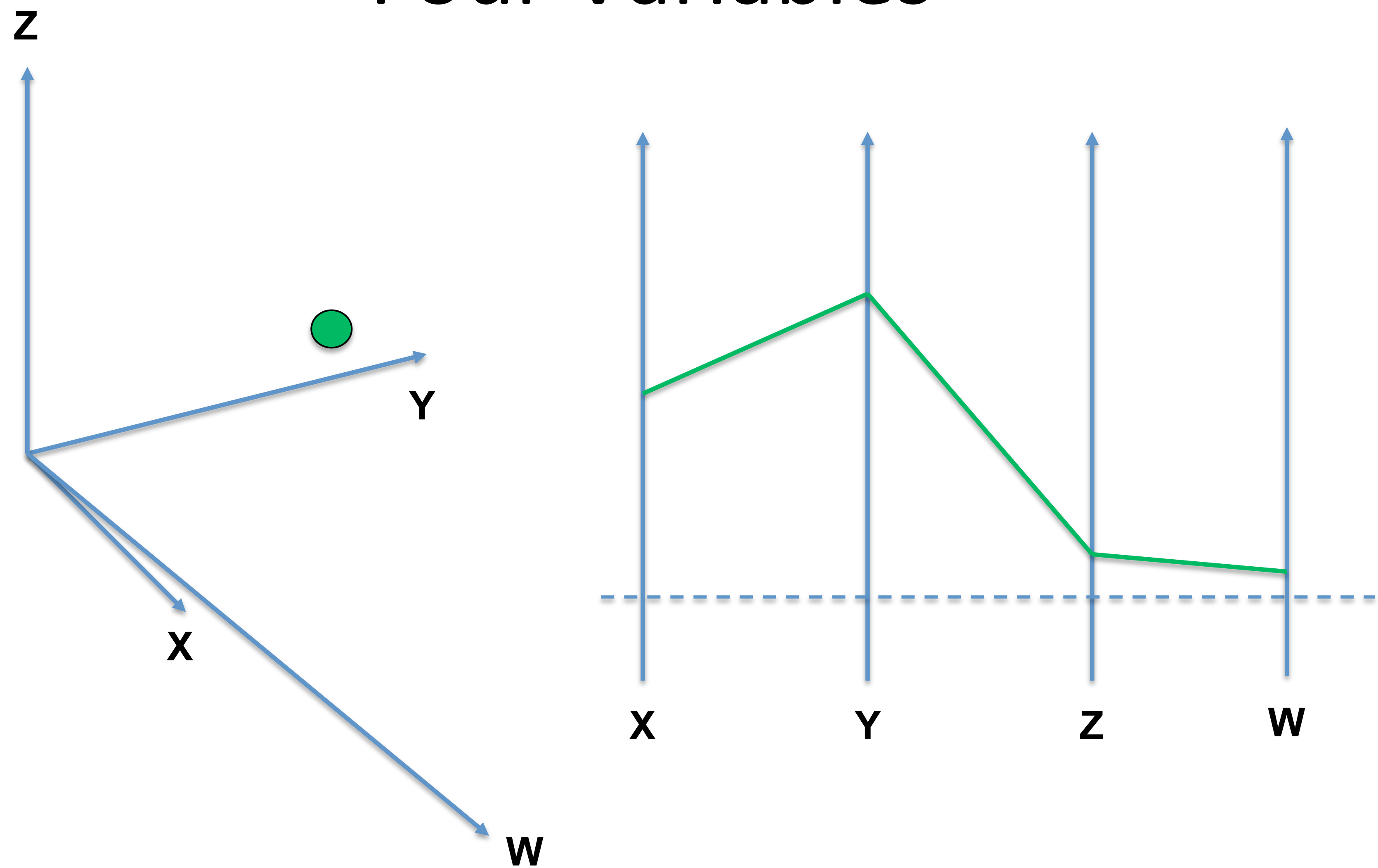


Parallel Coordinates
Three variables





Parallel Coordinates
Four Variables



Parallel Coordinates
N variables

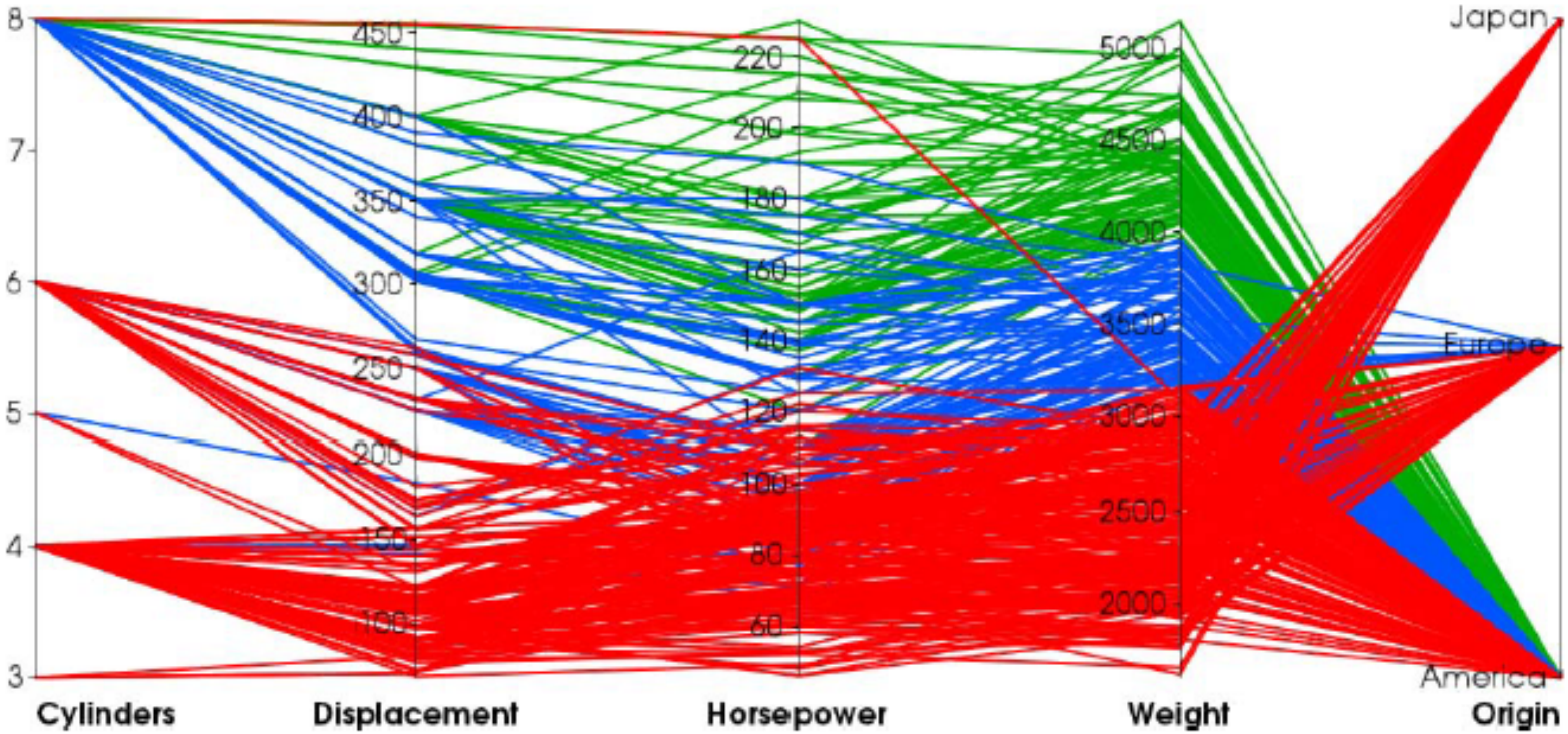


Figure from G. Palmas, M. Bachynsky, A. Oulasvirta, Hans Peter Seidel, T. Weinkauf, *"An Edge-Bundling Layout for Interactive Parallel Coordinates"*, in PacificVis 2014.

Axis Order

- Order of the axes play a fundamental role in readability.
- But we have *n!* combinations.

- Try by yourself at this website.
- Many strategies have been investigating for axes re-ordering.

Re-ordering and Edge-bundling

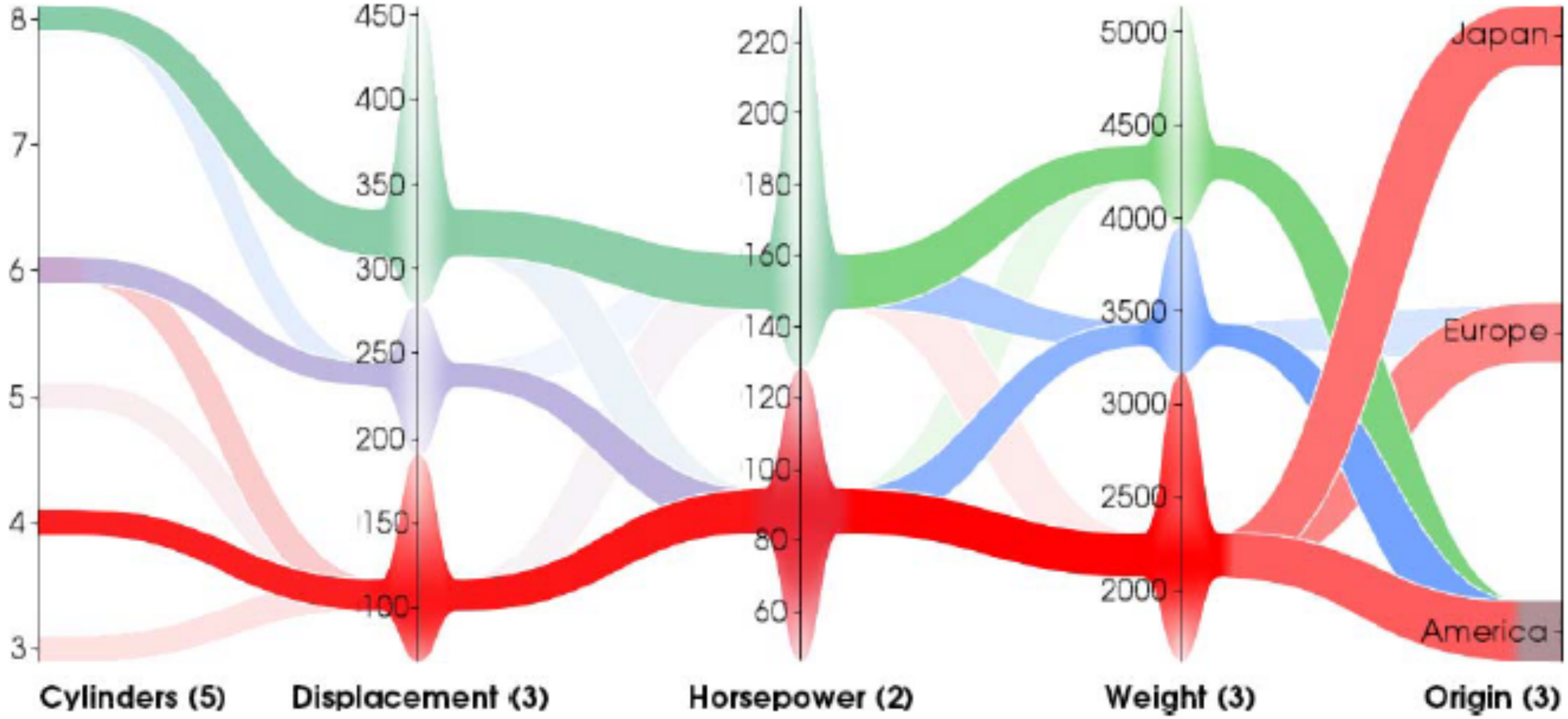


Figure from G. Palmas, M. Bachynsky, A. Oulasvirta, Hans Peter Seidel, T. Weinkauf, *"An Edge-Bundling Layout for Interactive Parallel Coordinates"*, in PacificVis 2014.

Illustrative Rendering and Parallel Coordinates

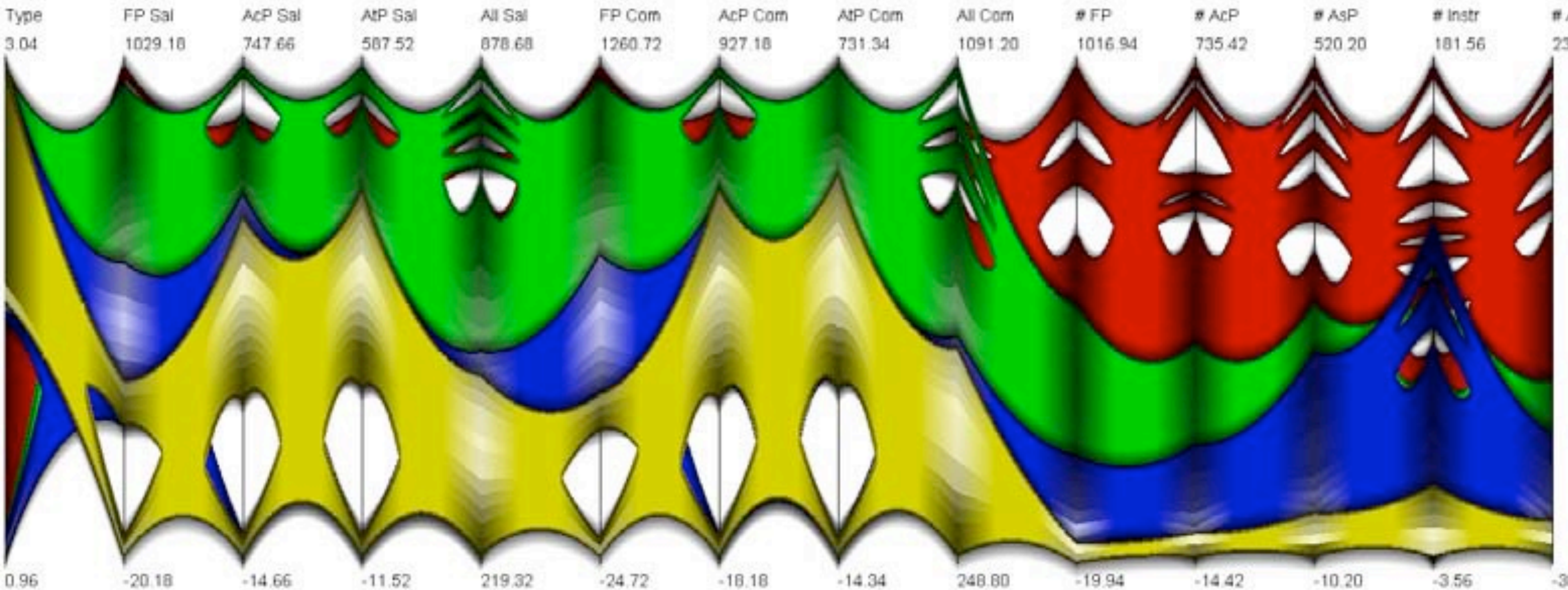
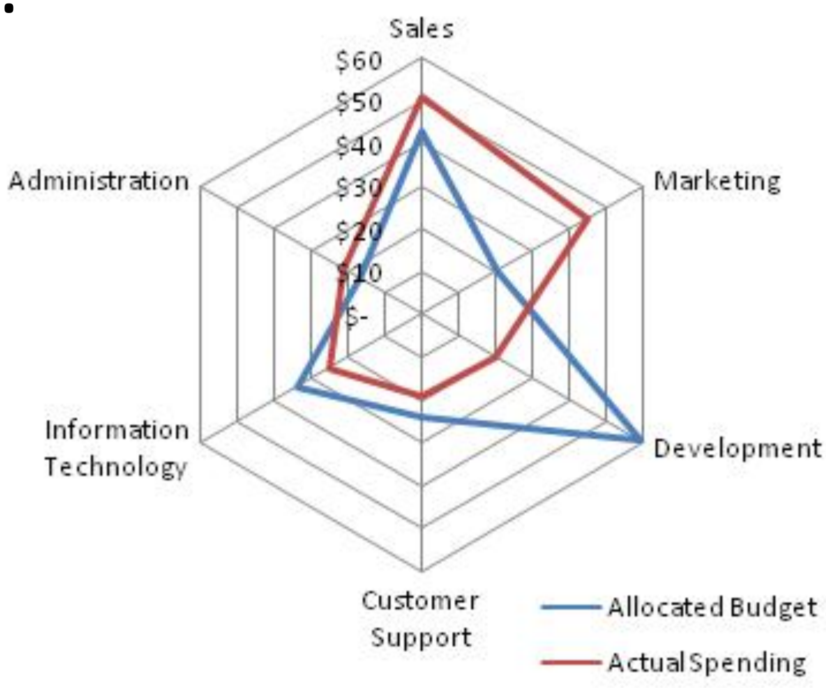


Figure from K. T. McDonnell and K. Mueller, *“Illustrative Parallel Coordinates“*, in Computer Graphics Forum, vol. 27, no. 3, pp. 1031-1038, 2008.

Star Plot

- Known with many names: radar chart, spider chart, web chart, etc.
- Analogous to parallel coordinates, but the axes are positioned in polar coordinates (equi-angular).
- Position of the first axis is uninformative.

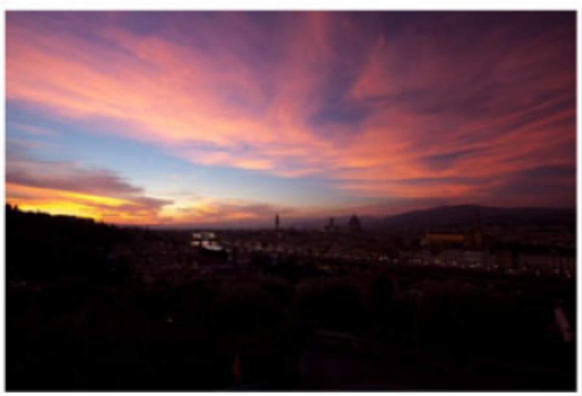


Star Plot

- Easy to compare properties of a class of objects or a category.
- Not easy to understand trade-off between different variables.
- Not suitable for many variables or many data.

Star Plot

- Example: measure the quality of a photograph.


Original


Edited

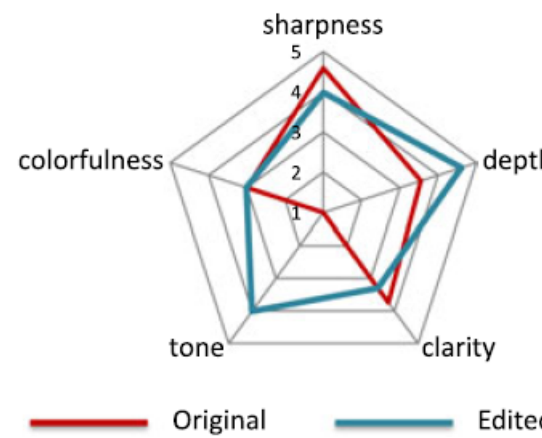


Figure from T. O. Aydın, A. Smolic and M. Gross, *"Automated Aesthetic Analysis of Photographic Images“*, in IEEE Transactions on Visualization and Computer Graphics, vol. 21, no. 1, pp. 31-42.

Summary

- Many solutions exists (a recent survey cites more than 250 papers).
- We have seen some of the most famous methods (e.g. SPLOM, parallel coordinates).
- Next lesson we will focus on dimensionality reduction.

Questions ?