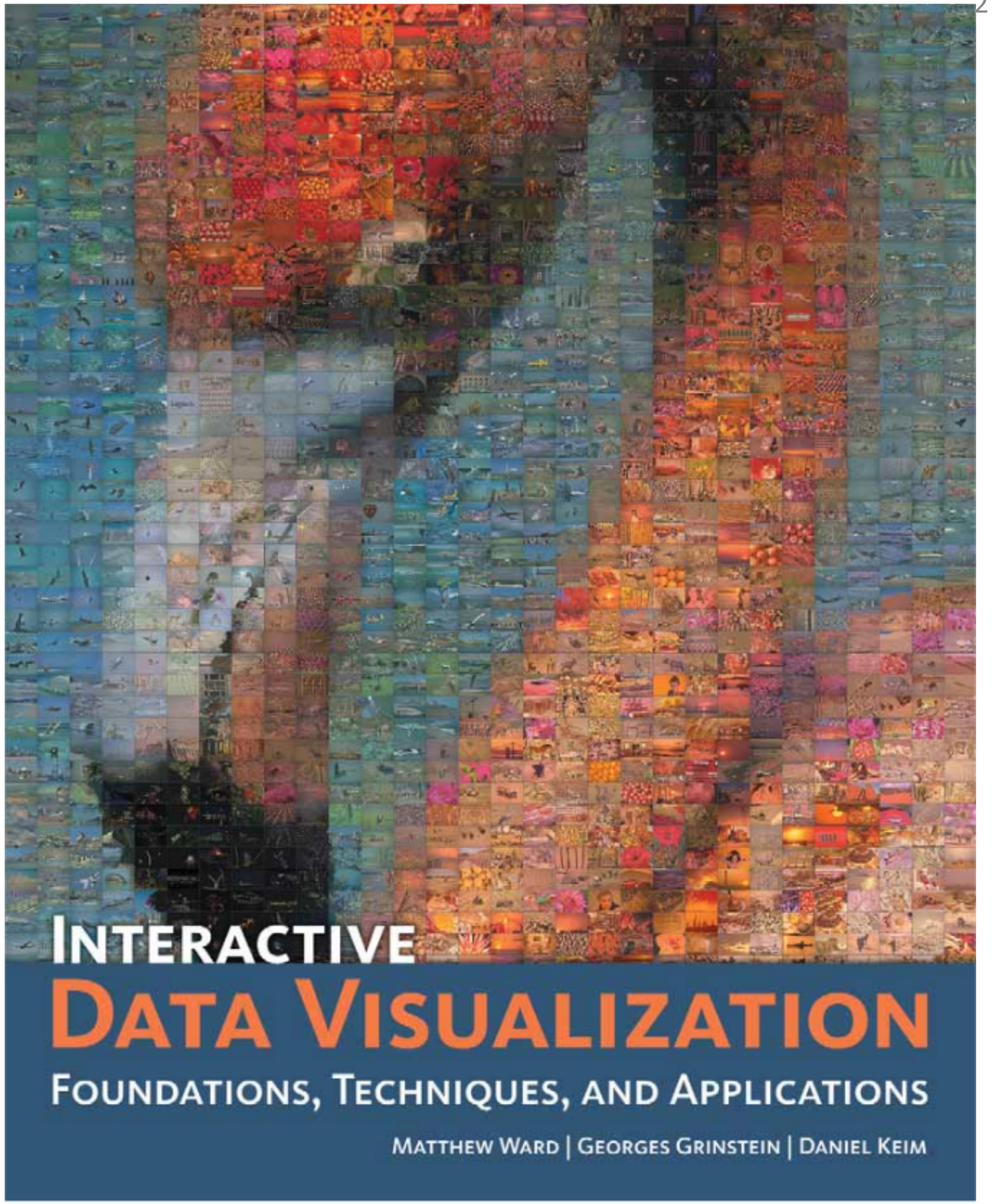
Visualization Theory

Miriam Butt & Dominik Sacha

LingVis: Visual Analytics for Linguistics
DGfS 2016 | 24.-26.2.2016

For More Detailed Information

- IDV book
- Covers all important InfoVis aspects
- Good starting point
- Also used in our InfoVis lectures



Contents

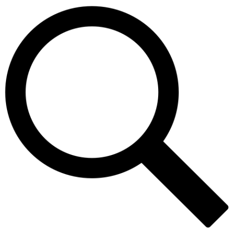
Introduction

Introduction – Goals for Visualization

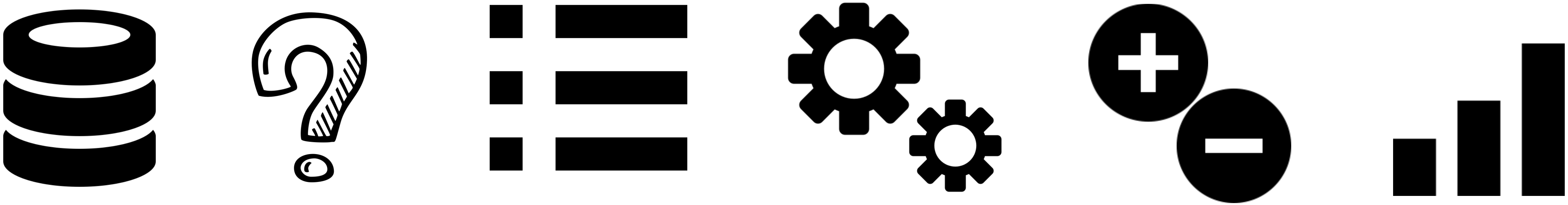
- **Presentation** (Communication)
  - Starting point: facts to be presented are fixed a priori
  - Process: choice of appropriate presentation techniques
  - Result: high-quality visualization of the data to present facts

- **Confirmatory Analysis**
  - Starting point: hypotheses about the data
  - Process: goal-oriented examination of the hypotheses
  - Result: visualization of data to confirm or reject the hypotheses

- **Exploratory Analysis**
  - Starting point: no hypotheses about the data
  - Process: interactive, usually undirected search for structures, trends
  - Result: visualization of data to lead to hypotheses about the data



Introduction – Steps for Visualizing Data



Data Type | **Questions** | **Requirements** | **Preprocessing** | **Pro/Cons** | **Design**

Find out what kind of data you have.

Which questions shall be answered by the data?

Derive requirements from these tasks.

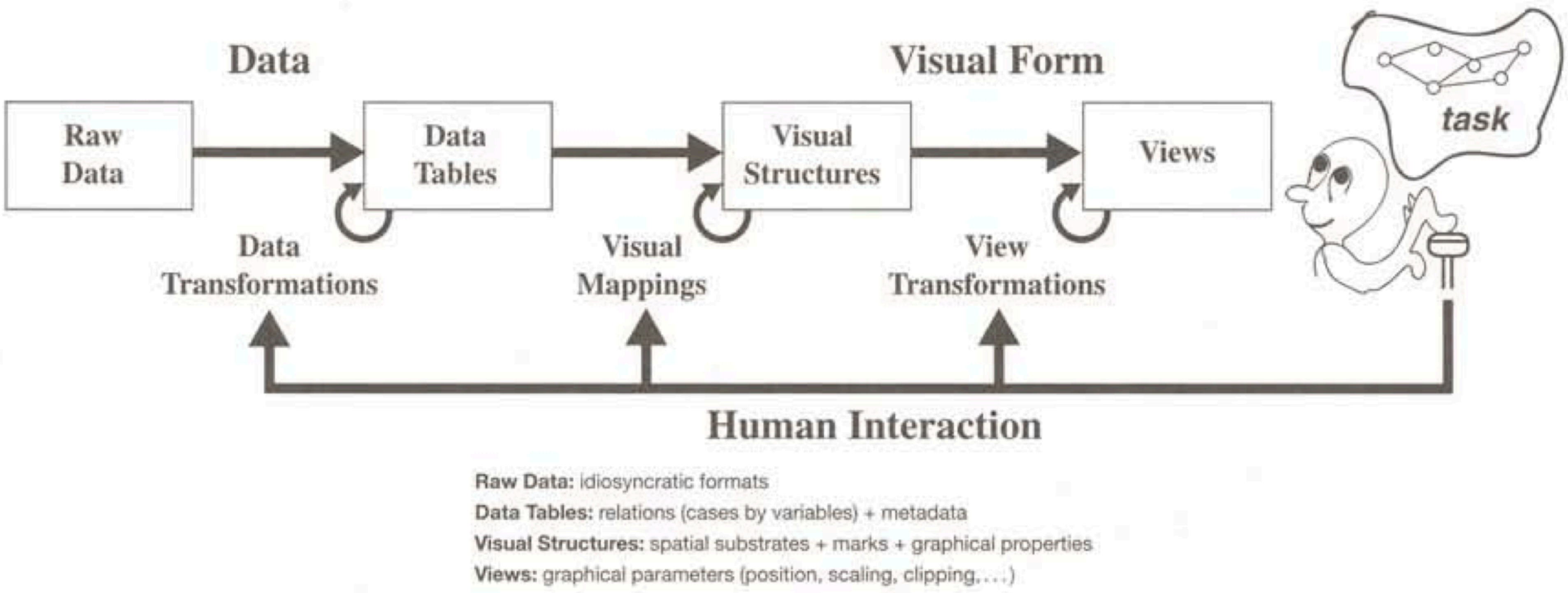
Preprocess data (missing data, outliers, transformations, …).

Consider the pros and cons of different visualization techniques (effectiveness, familiarity …)

Design a visualization system.

Introduction – Steps for Visualizing Data - Pipeline

"Reference Model for Information Visualization"



S. Card, J. Mackinlay, and B. Shneiderman. 1999. Readings in Information Visualization: Using Vision to Think. Chapter 1.

Data Foundations

Data Foundations – Data Types

Nominal

- No quantitative relationship between categories
- Classification without ordering

→ No intrinsic order

e.g. sports, nationality

Ordinal

- Attributes can be rank-ordered
- Distances between values do not have any meaning

→ Ordered in a sequence

e.g. places ordered by noise

Numeric

- Attributes can be rank-ordered
- Distances between values have a meaning
- Mathematical operations are possible

e.g. height of a person

References: C. Ware. 2012. Information Visualization: Perception for Design. Chapter 1.

Data Preprocessing – Real World Data is Dirty, What's Wrong?

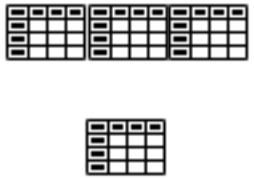
Unique ID	Gender	Age	Smoking Habits	Time
323	0	21	0	1h 40s
435	1	34	2	1h 55s
123	0		3	90 min
352	1	25	0	1h 43s
674	1	25	1	0,4 day
865	0	18	3	1h 50s
341	X	41	2	1h 33s

Data Foundations – Preprocessing

- Data Cleaning
  - Missing Values
  - Noise …

- Normalization
  - Common scale (0-1)
  - E.g., color mapping

- Segmentation
  - Divide data in meaningful parts



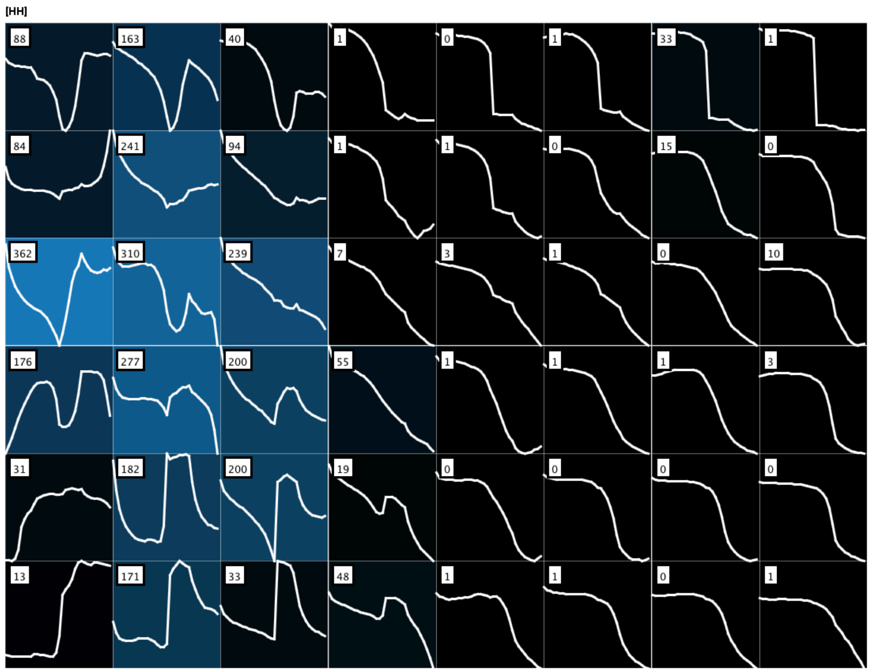
- Data Reduction
  - Sampling
  - Dim. Reduction

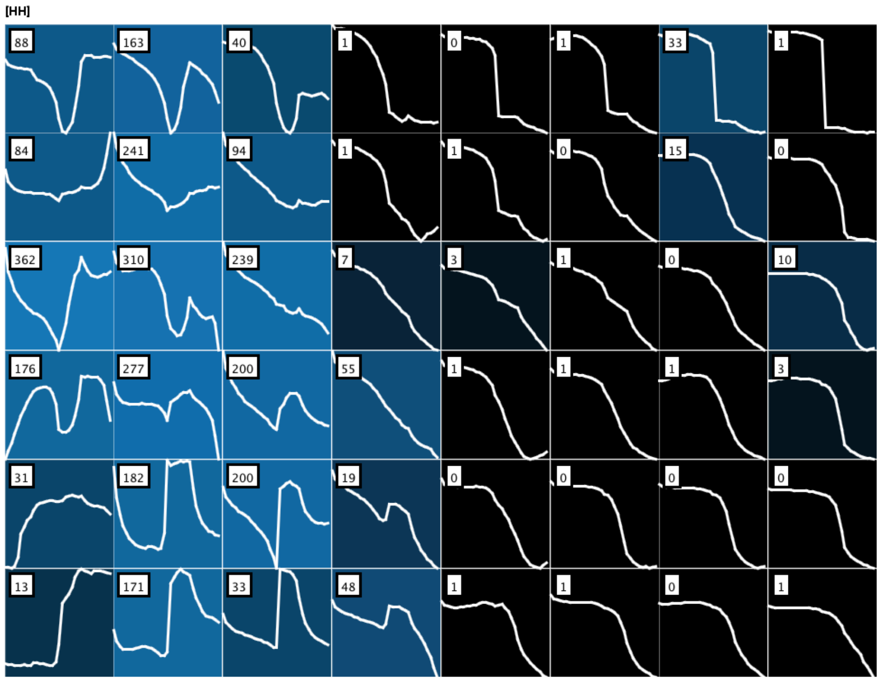
➢ Big part of our daily work!
➢ Methods have to be applied carefully
➢ Analysts needs to know which techniques have been applied! (Loss of Information)

Data Foundations – Linear vs. Logarithmic Normalization

$$f_{lin}(v) = \frac{v - min}{max - min}$$

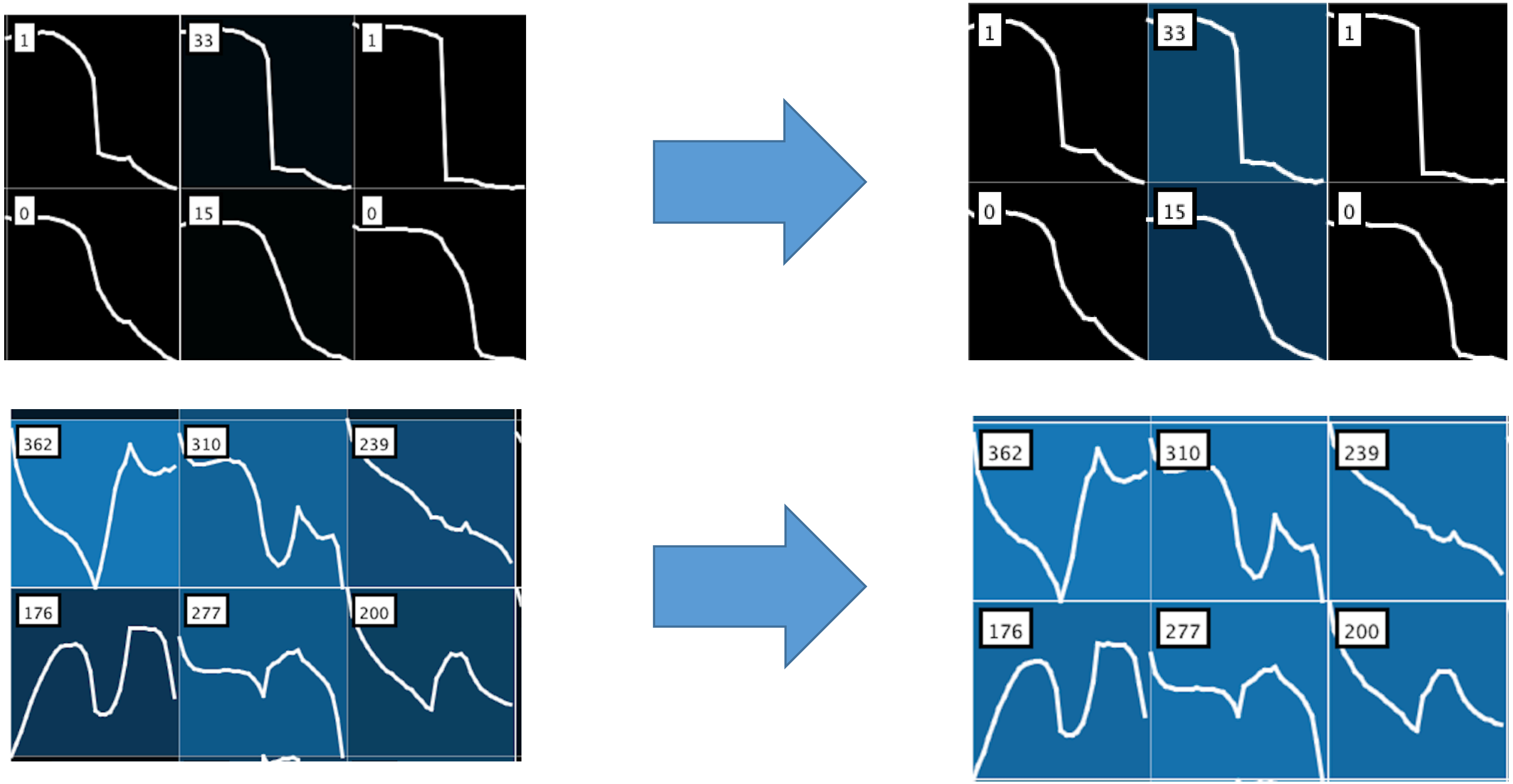
$$f_{\ln}(v) = \frac{\ln(v) - \ln(\min)}{\ln(\max) - \ln(\min)}$$





Data Foundations – Logarithmic Normalization

Advantages and Disadvantages?



Data Foundations – Practically Speaking

- Think carefully about appropriate preprocessing methods
  - Consider data type, class, tasks and data quality
  - “Tailor data to the analysis goals”
- Keep track and communicate which preprocessings have been applied and consider their impacts on the final data product
- Poorly designed preprocessings may create artificial patterns in the data leading to wrong conclusions or analysis results

Key Aspects of Visualization

Visualization Basics

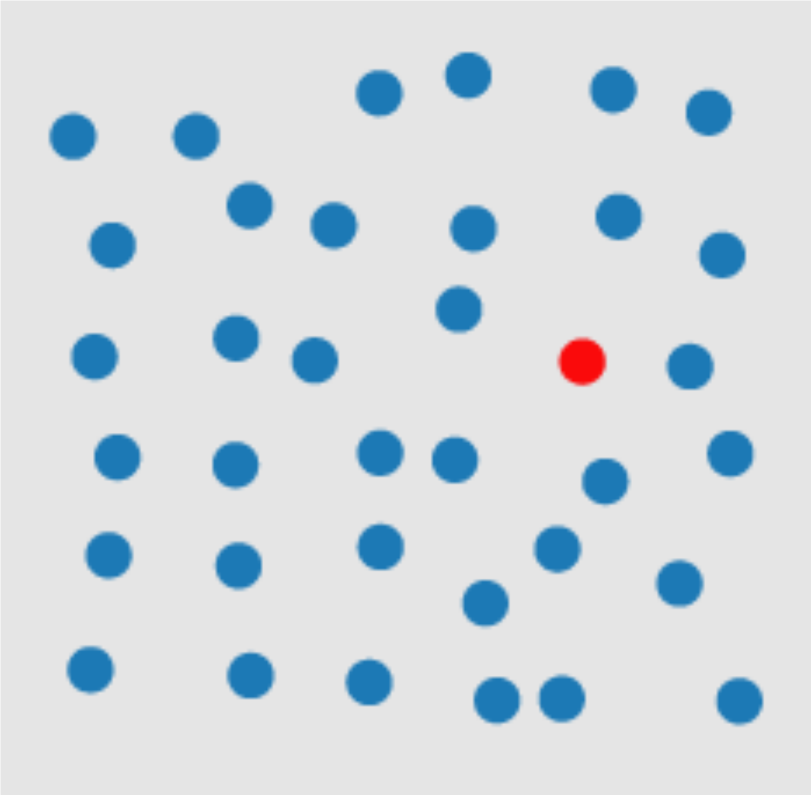
- **Perceptual/Pre-Attentive Processing**
  - Certain low level visual aspects are recognized before *conscious* awareness
- **Gestalt Laws**
  - The *tendency* to **perceive** elements as belonging to a group, based on certain visual properties
- **Visual Variables**
  - The different visual aspects that can be used to **encode** information

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Preattentive Processing

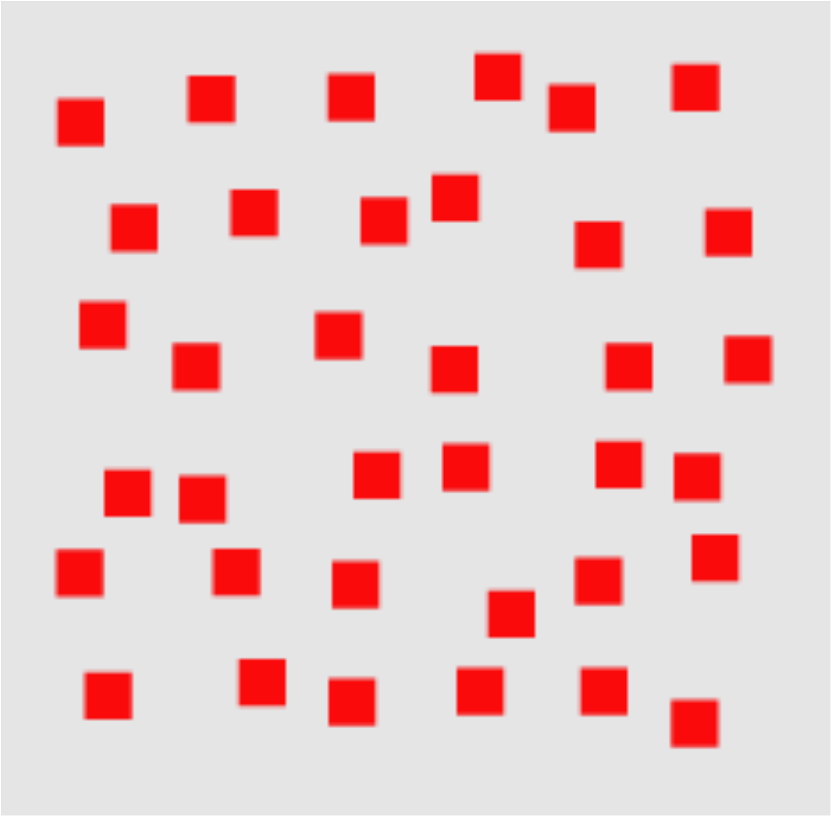
- Perception of visual features managed by the low-level visual machinery
- Extremely fast: < 200 msec (the eyes take > 200 msec to initiate movement) => processed in parallel

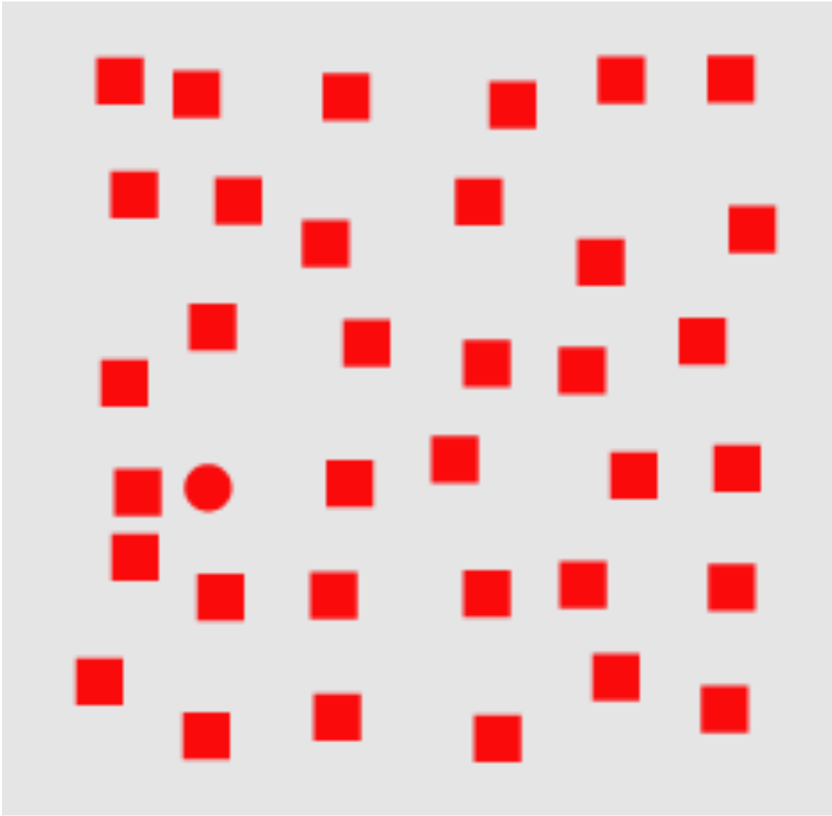
Preattentive Processing (color)



http://www.idvbook.com/

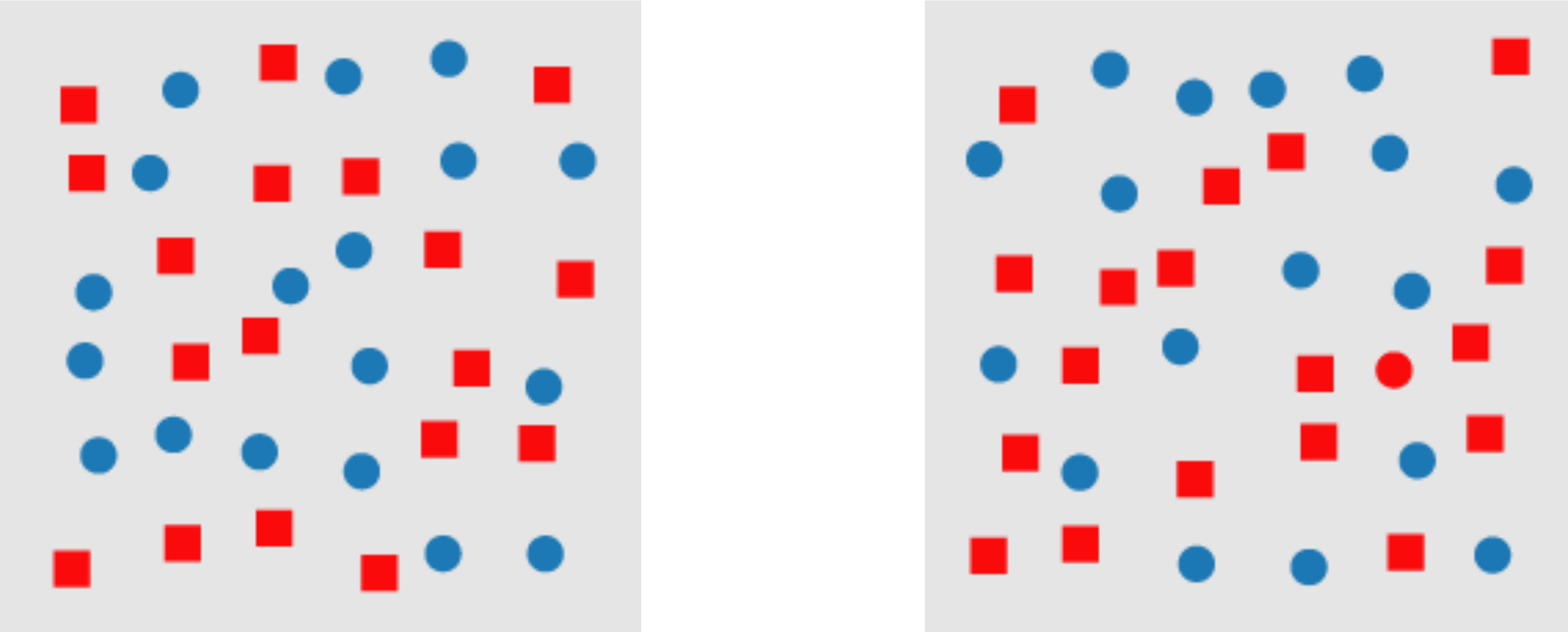
Preattentive Processing (shape)





http://www.idvbook.com/

Serial Processing



http://www.idvbook.com/

- Finding the red circle requires serial processing
- The red circle is a conjuctive target (composed of non-unique features)

- Demo: http://www.csc.ncsu.edu/faculty/healey/PP/index.html

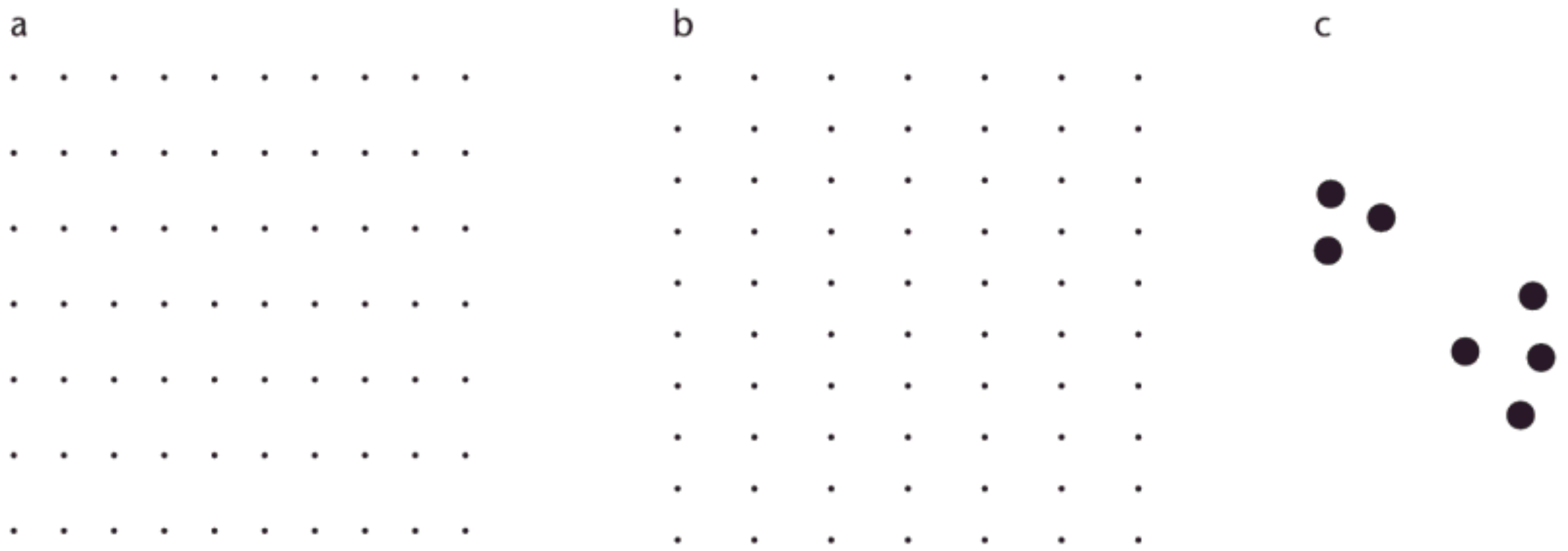
Gestalt Laws

Perceptual laws about how we **group** visual objects together to form visual entities.

1. Law of Proximity
2. Law of Similarity
3. Law of Connectedness
4. Law of Continuity
5. Law of Symmetry
6. Law of Closure and Common Region
7. Law of Figure and Ground

Gestalt Laws - Proximity

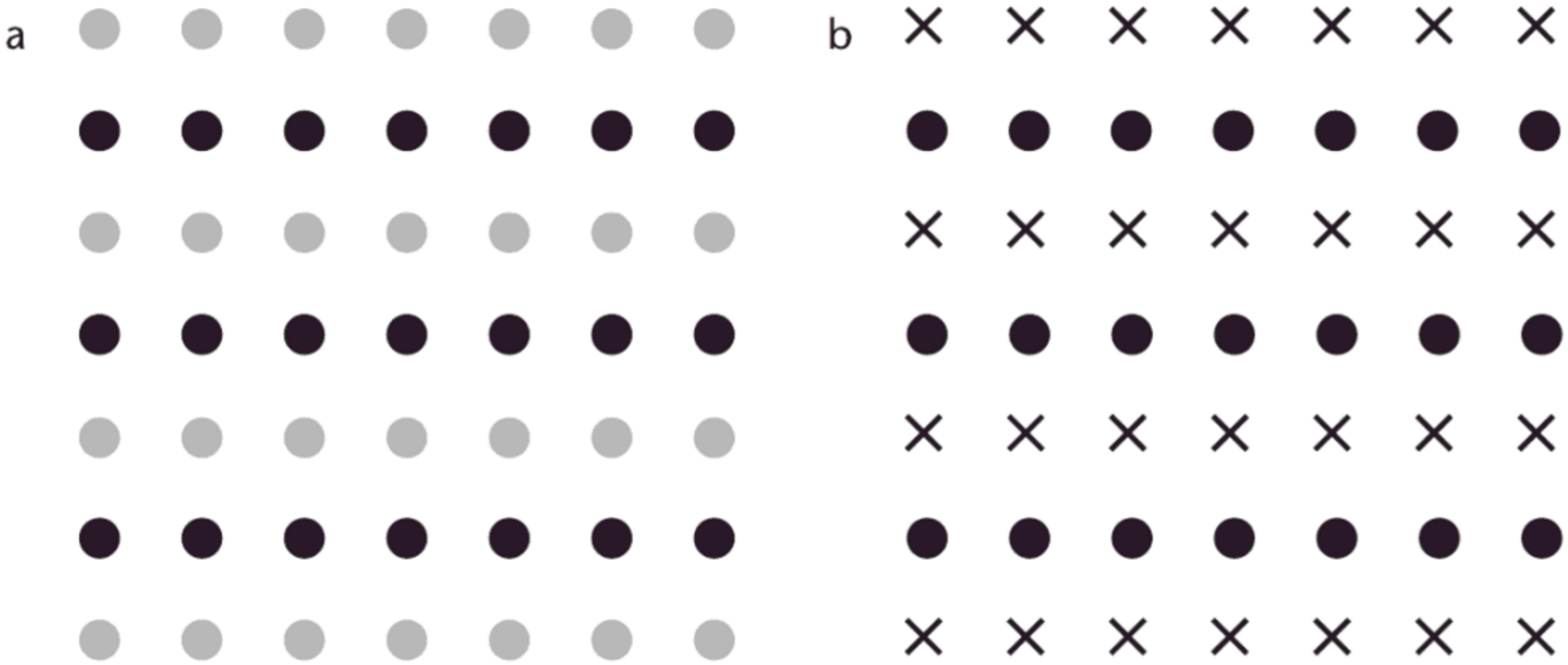
Close objects are perceptually grouped.



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws - Similarity

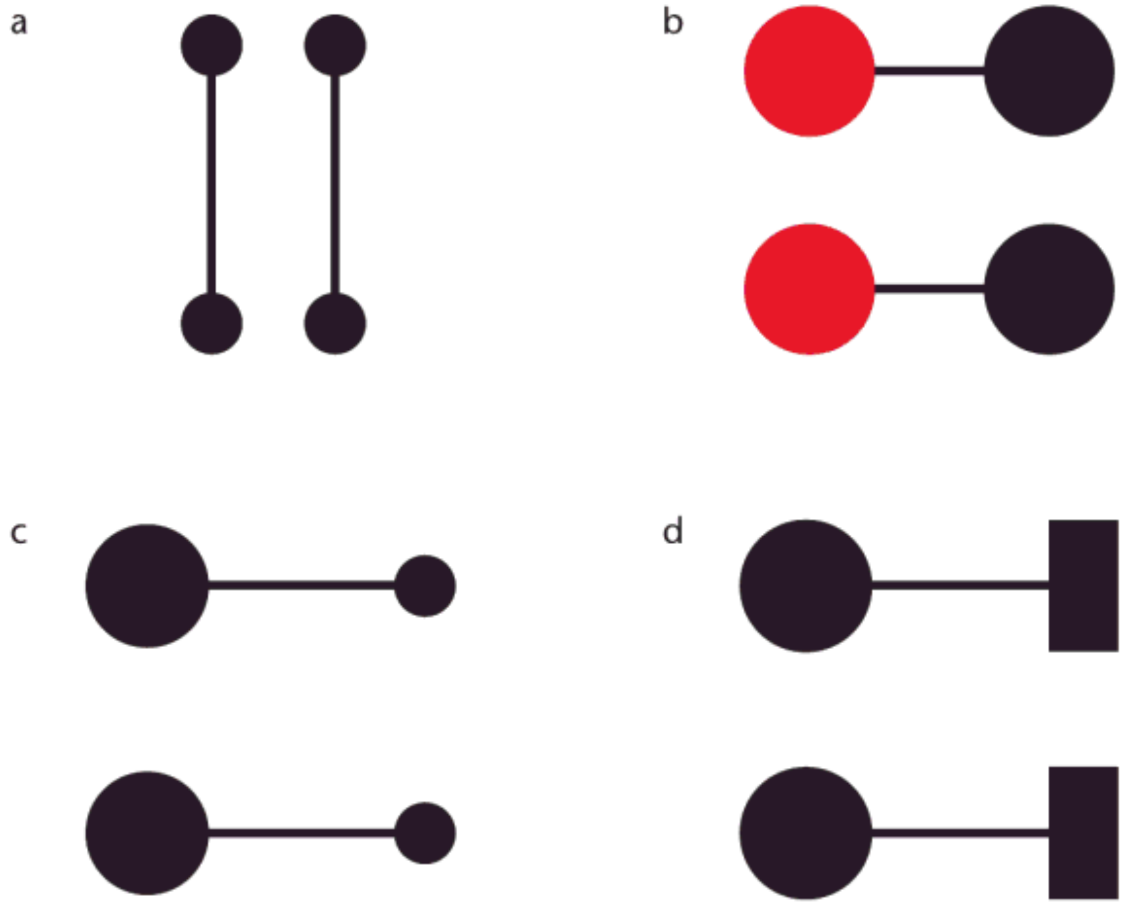
Similar objects are perceptually grouped.



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws - Connectedness

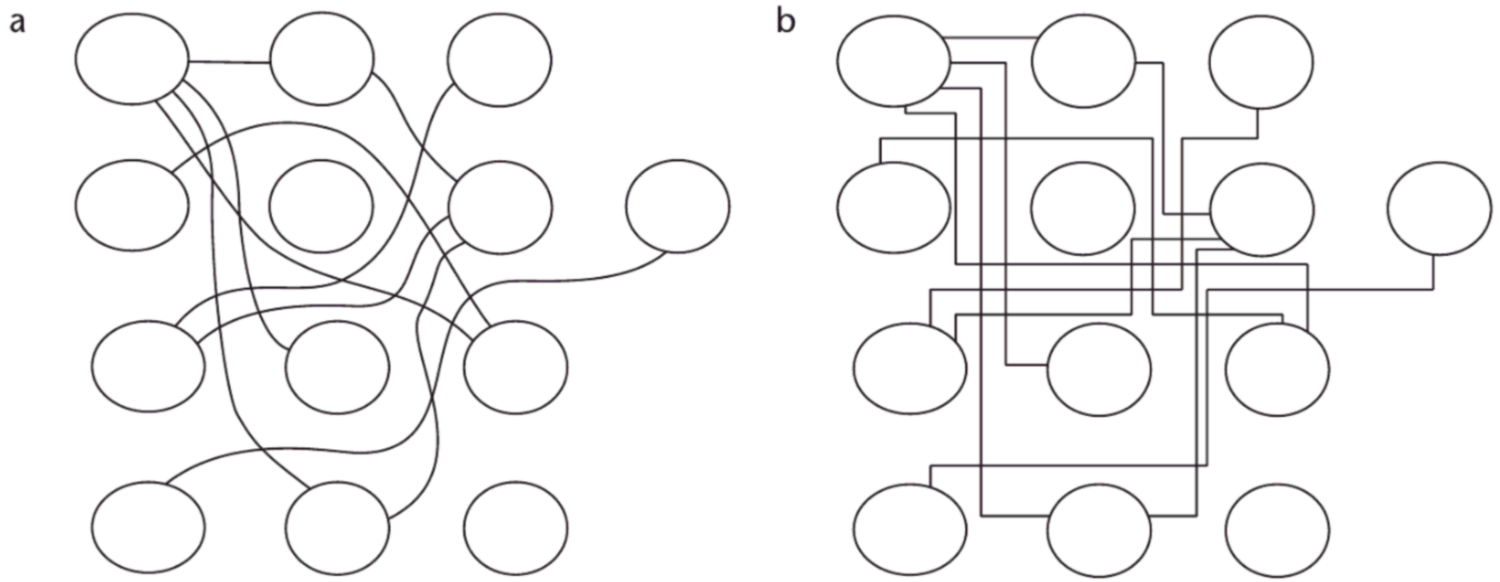
- Connected objects are perceptually grouped.
- Can be more powerful than proximity, color, size or shape.
- Used in node–link diagrams



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws - Continuity

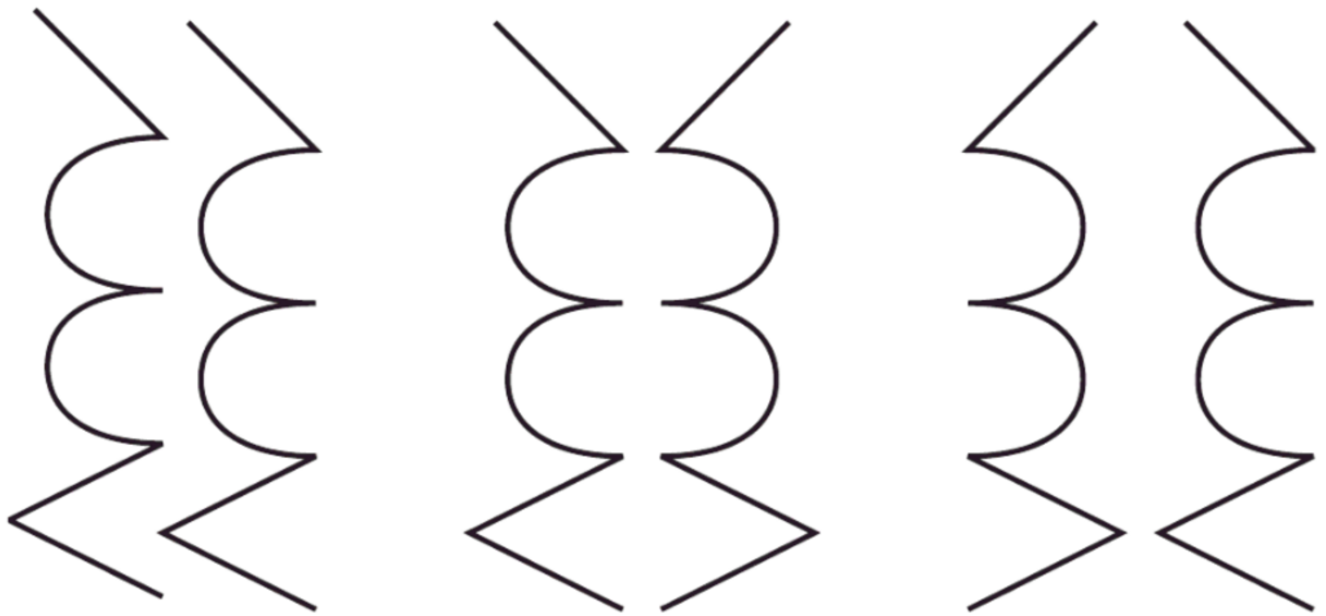
It is easier to construct visual objects out of visual elements if the connection is smooth and continuous. Lines with abrupt changes in direction are harder to read.



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws - Symmetry

- powerful organization principle
- symmetric forms build a visual whole



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

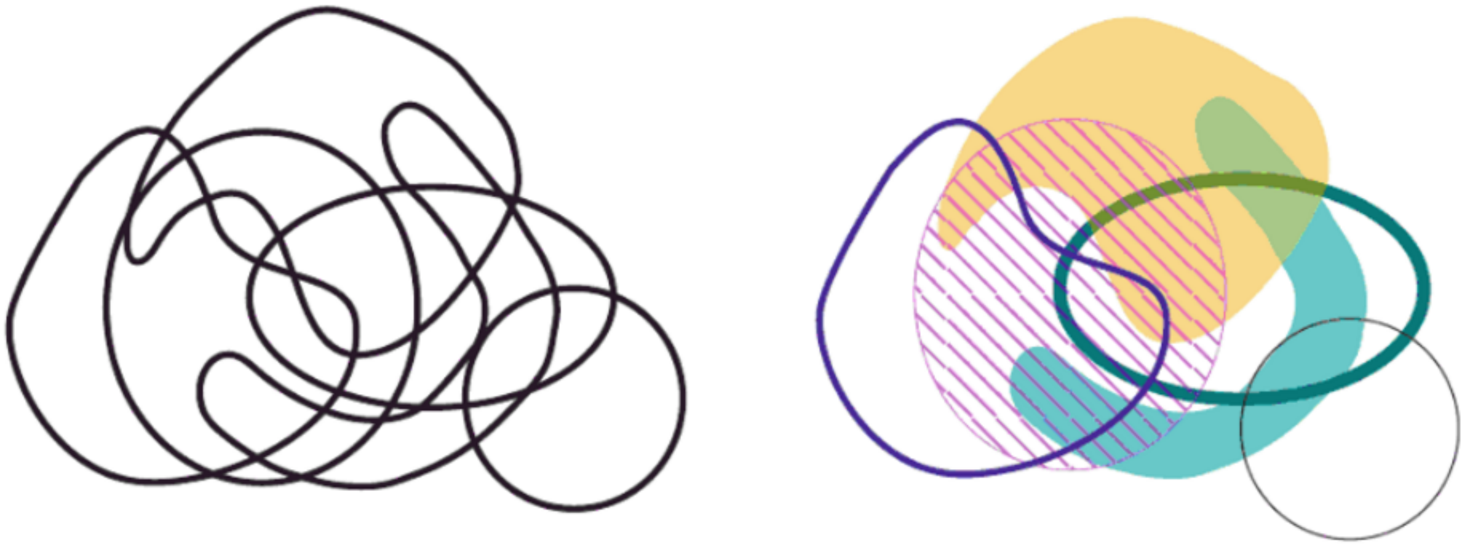
Gestalt Laws – Closure and Common Region

We prefer perceptual solutions with closed contours.
We see a ring behind a rectangle rather than a broken ring.

Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws – Closure and Common Region

Euler diagram to using color and texture to highlight contours.



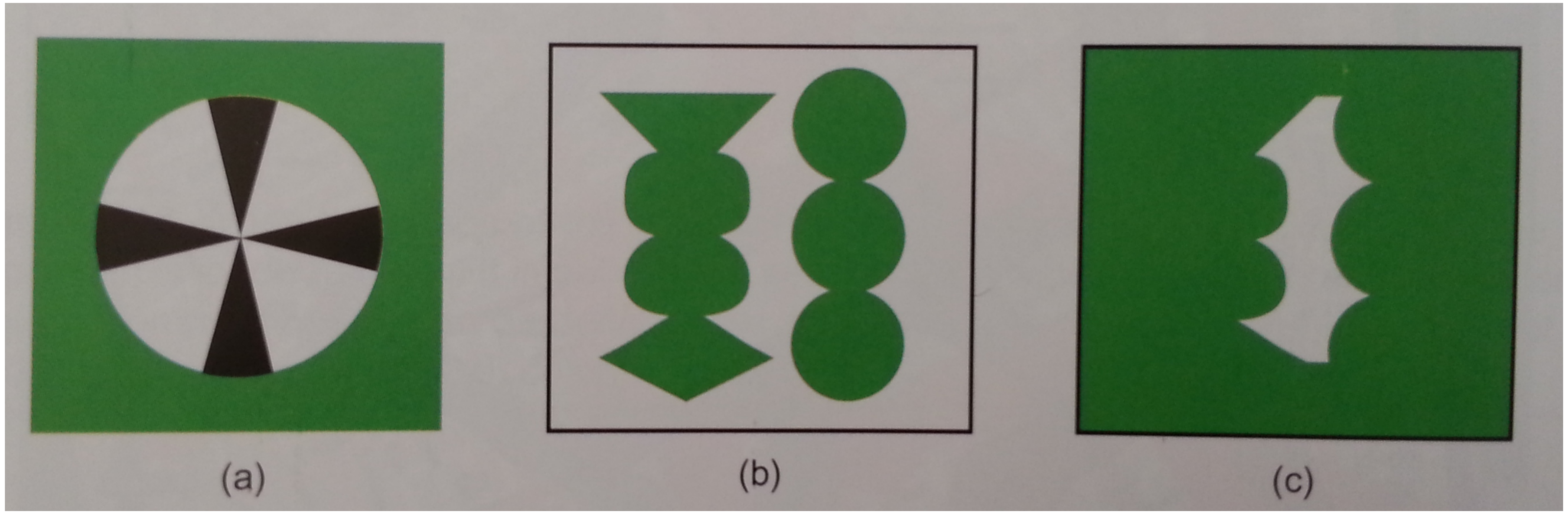
(a) (b)

Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Gestalt Laws – Figure and Ground

- **figure**: object like, perceived to be in the foreground.
- **ground**: whatever lies behind the figure.

- fundamental to identify objects.
- all Gestalt laws contribute to create a figure



Colin Ware, Information Visualization: Perception for Design, Morgan Kaufmann

Some more examples you might know

Closure

Figure Ground

Proximity, Closure

https://de.wikipedia.org/wiki/WWF#/media/File:WWF_Logo.svg

https://en.wikipedia.org/wiki/Finder_%28software%29#/media/File:Mac_Finder_icon_%28OS_X_Yosemite%29.png

https://de.wikipedia.org/wiki/IBM#/media/File:IBM_logo.svg

Marks and Visual Variables

Jacques Bertin:

- A **mark** is "something that is visible and can be used ... to show relationships within sets of data"
- **Visual variables** are "the different ways that a mark can be varied"

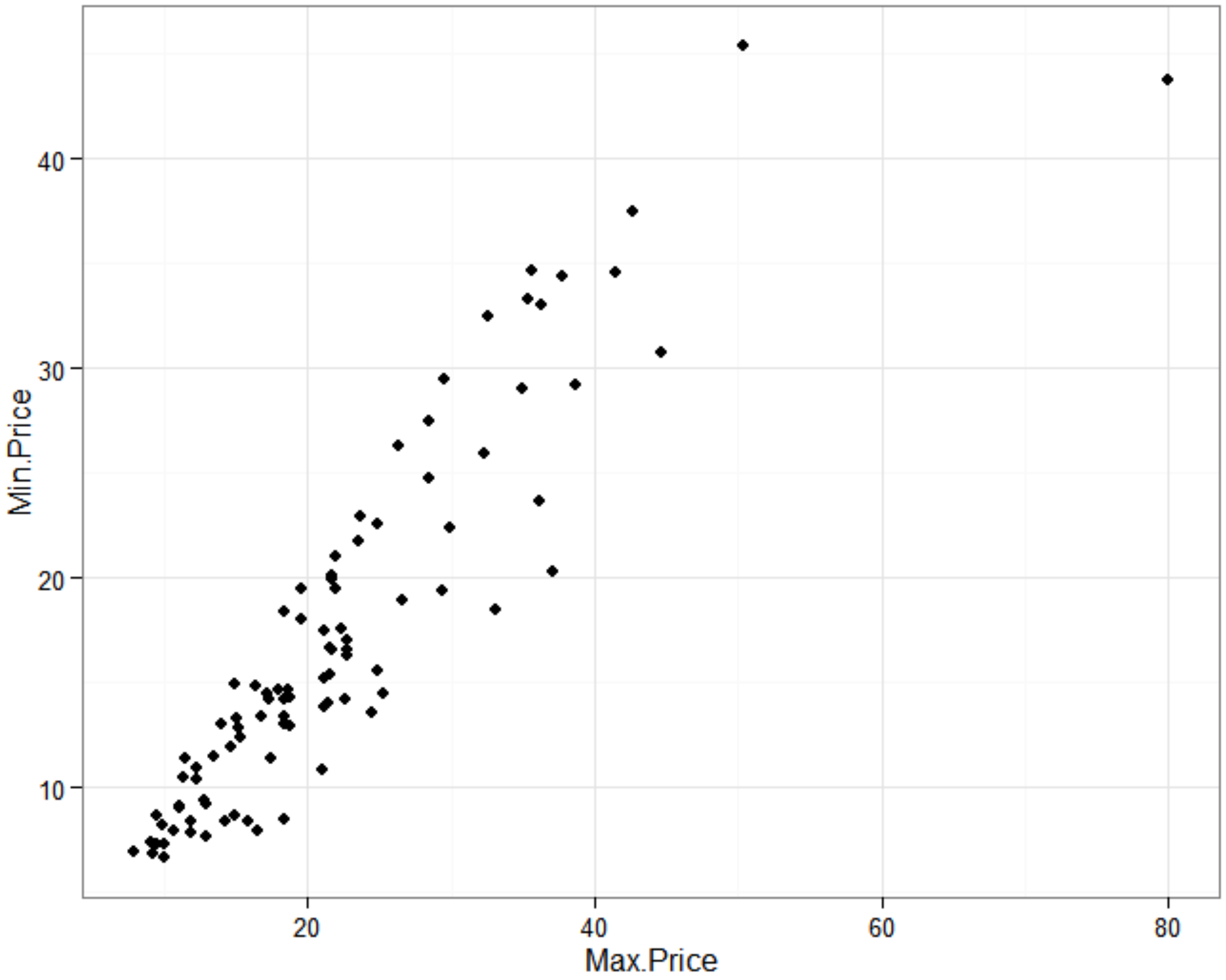
Reference:
S. Carpendale. 2003. Considering Visual Variables as a Basis for Visualization. Research report 2001-693-16, Department of Computer Science, University of Calgary.

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

The Eight Visual Variables

1. Position
2. Mark/Shape
3. Size (Length, Area and Volume)
4. Brightness
5. Color
6. Orientation
7. Texture
8. Motion

The Eight Visual Variables - Position

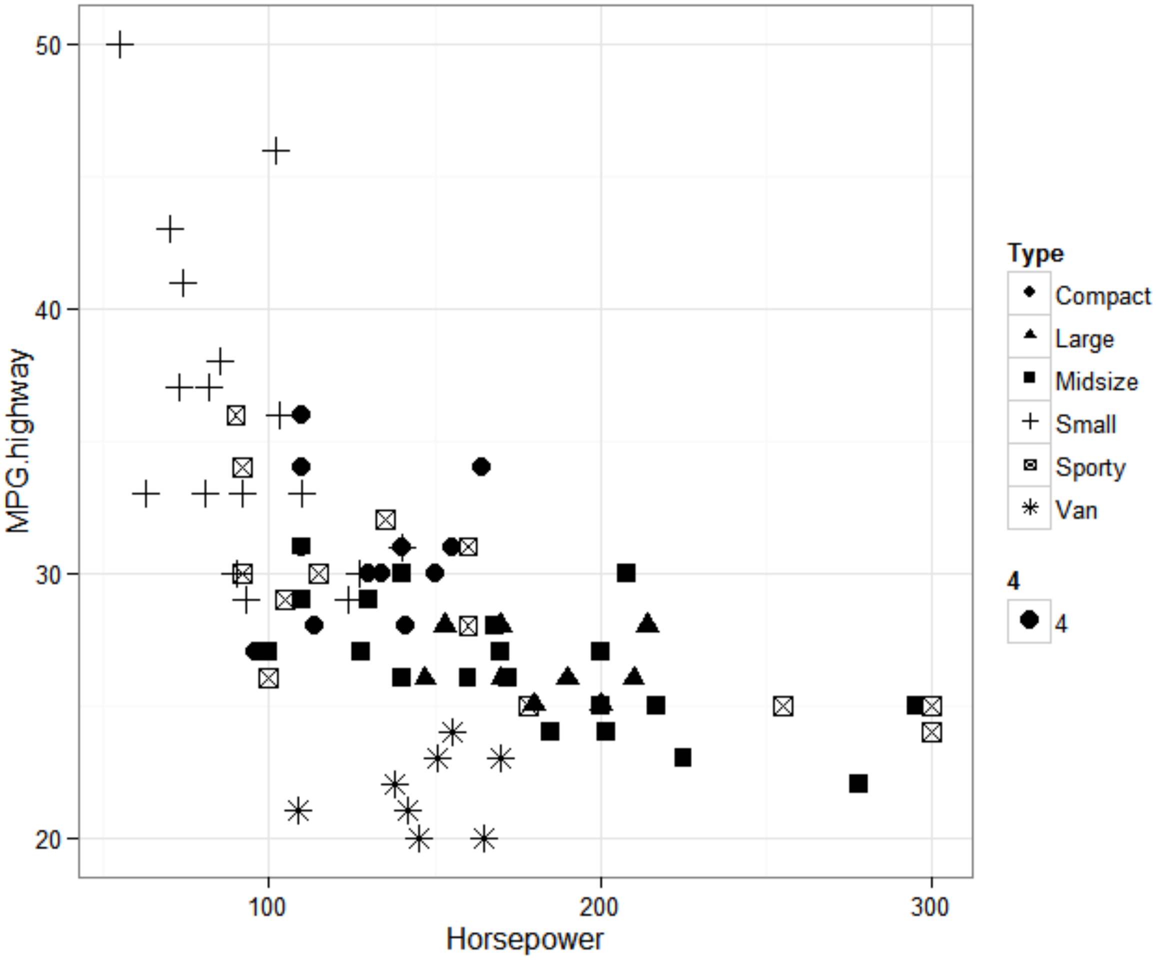


The Eight Visual Variables – Shape/Mark

Several examples of different marks or glyphs that can be used.

http://www.idvbook.com/

The Eight Visual Variables – Shape/Mark



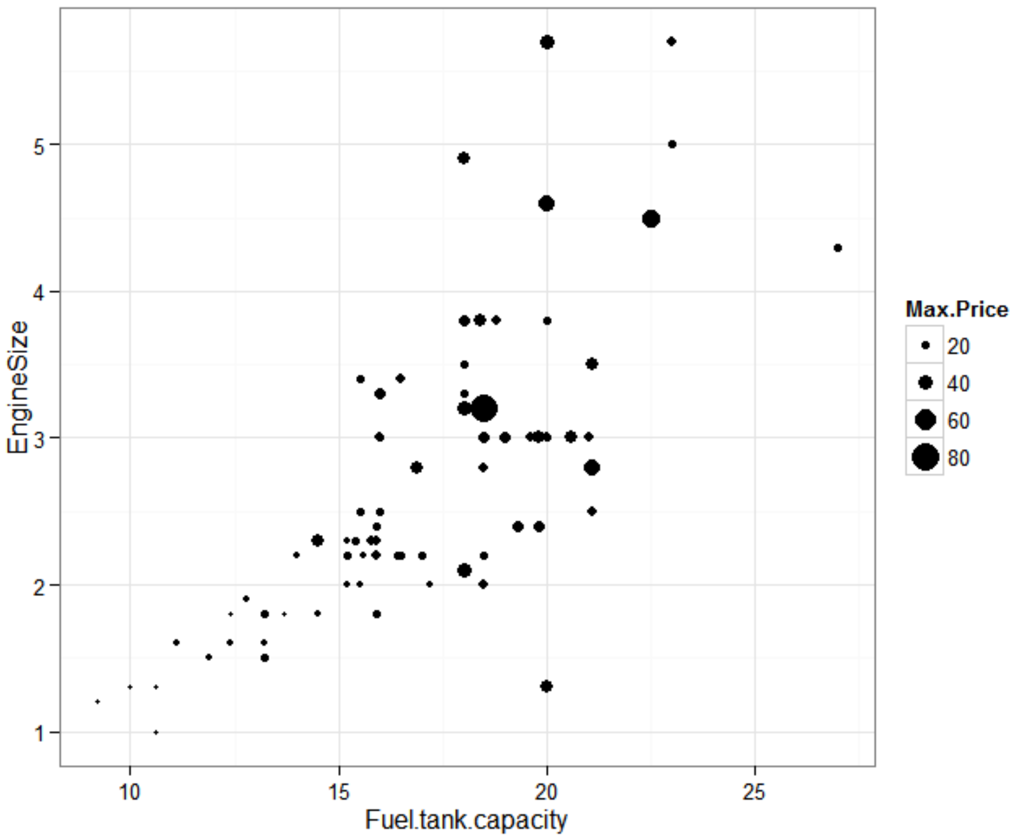
This visualization uses shapes to distinguish between different car types in a plot comparing highway MPG and horsepower. Clusters are clearly visible, as well as some outliers.

The Eight Visual Variables – Size (Length, Area and Volume)

Example sizes to encode data.

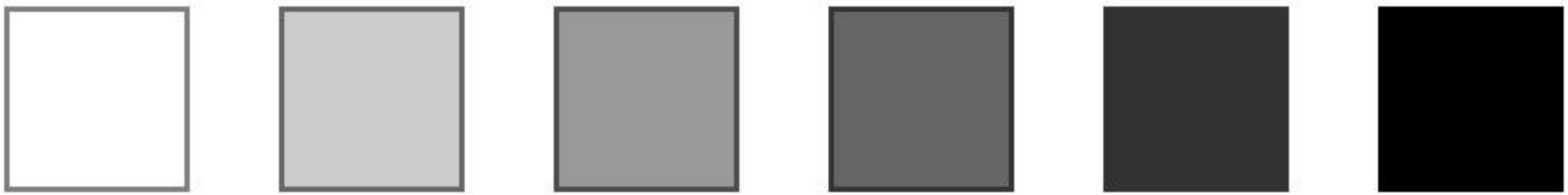
http://www.idvbook.com/

The Eight Visual Variables – Size (Length, Area and Volume)



This is a visualization of the 1993 car models data set, showing engine size versus fuel tank capacity. Size is mapped to maximum price charged.

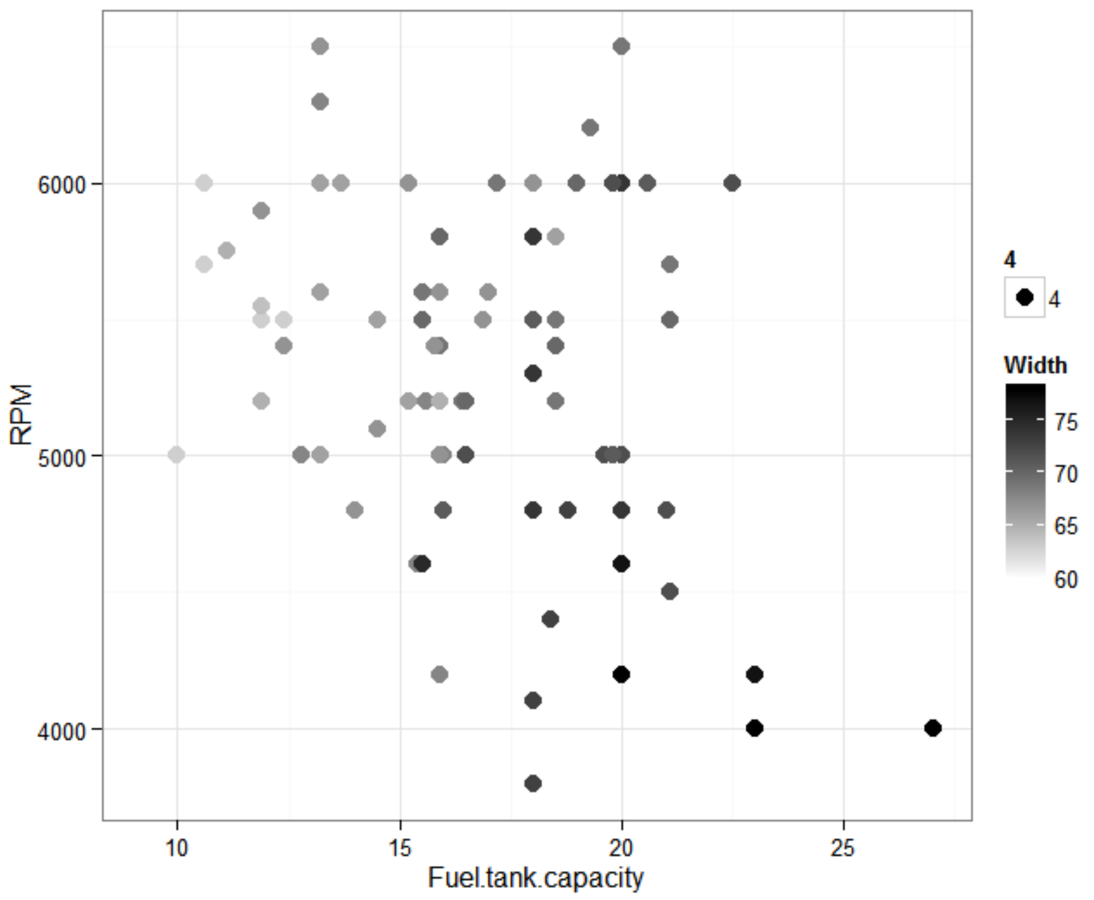
The Eight Visual Variables – Brightness



Brightness scale for mapping values to the display.

http://www.idvbook.com/

The Eight Visual Variables – Brightness

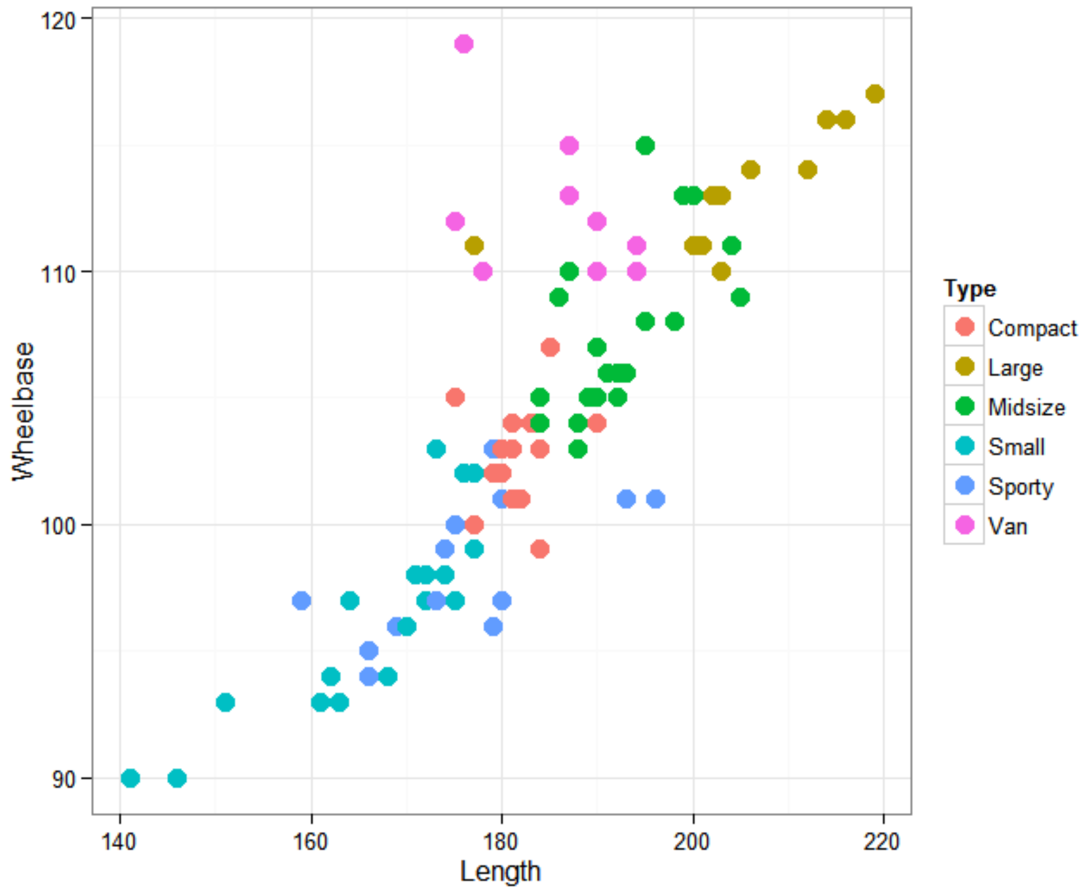


Another visualization of the 1993 car models data set, this time illustrating the use of brightness to convey car width (the darker the points, the wider the vehicle).

The Eight Visual Variables – Color

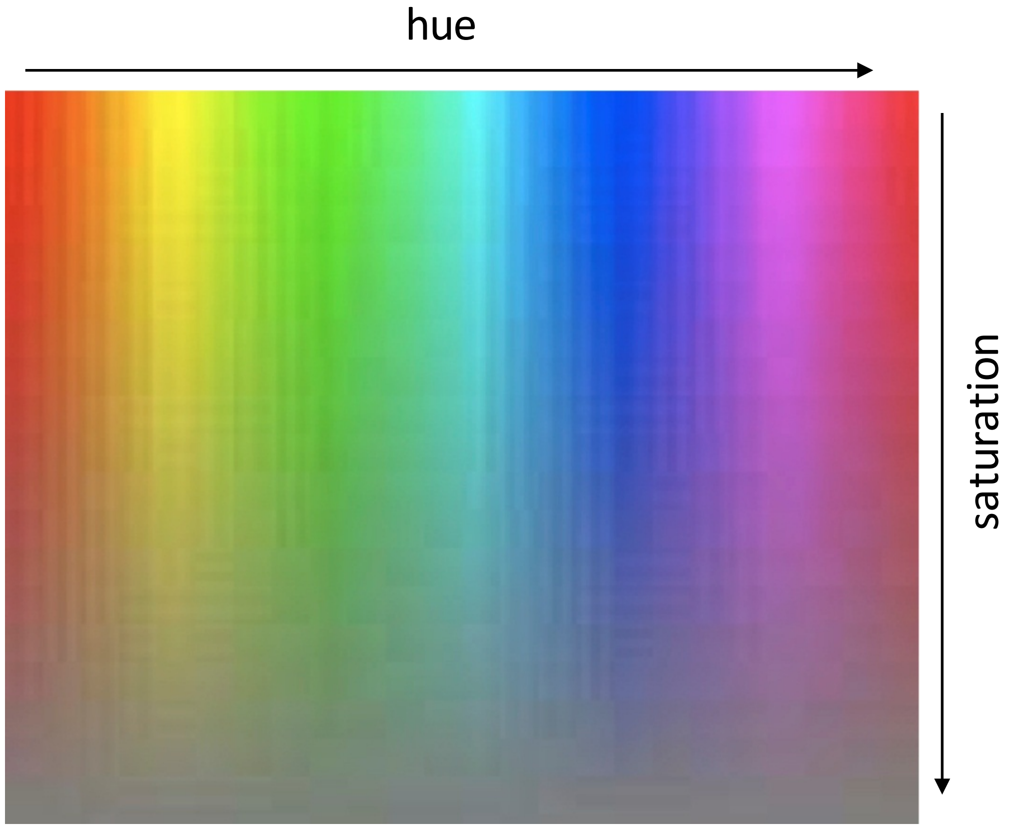
http://www.idvbook.com/

The Eight Visual Variables – Color



A visualization of the 1993 car models, showing the use of color to display the car's length. Here length is also associated with the *y*-axis and is plotted against wheelbase. In this figure, blue indicates a shorter length, while yellow indicates a longer length.

The Eight Visual Variables – Color



Microsoft hue/saturation color selector.

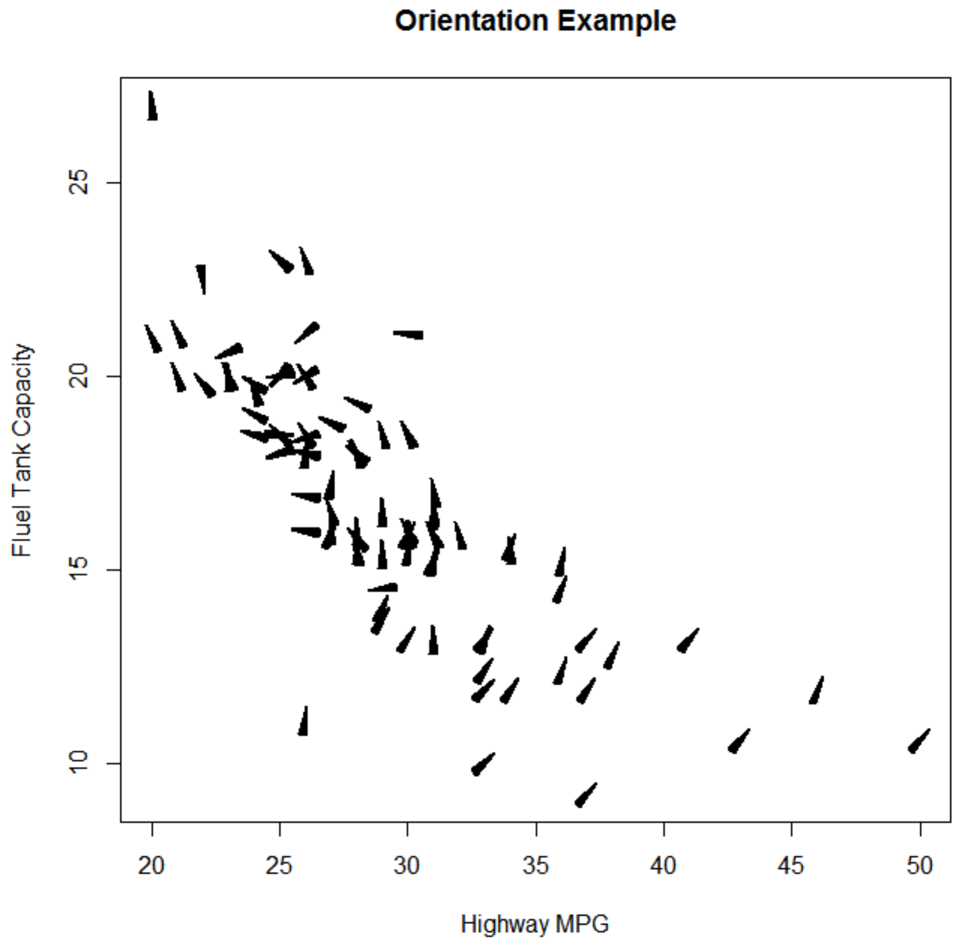
http://www.idvbook.com/

The Eight Visual Variables – Orientation

Example orientations of a representation graphic, where the lowest value maps to the mark pointing upward and increasing values rotate the mark in a clockwise rotation.

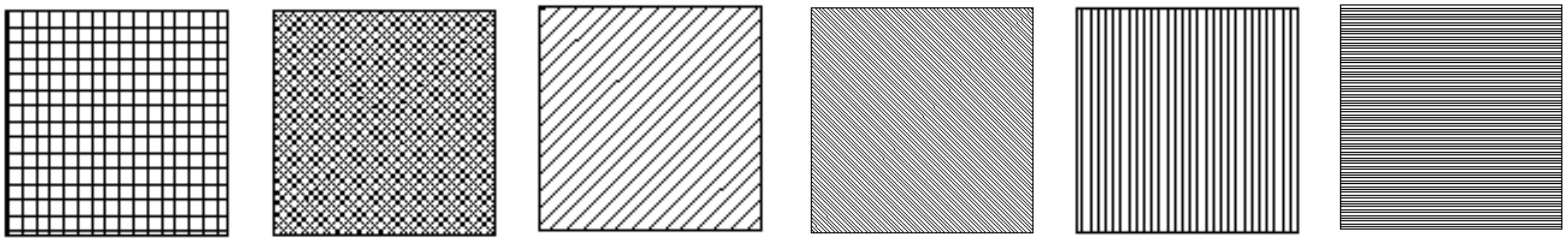
http://www.idvbook.com/

The Eight Visual Variables – Orientation



Sample visualization of the 1993 car models data set depicting using highway miles per-gallon versus fuel tank capacity (position) with the additional data variable, midrange price, used to adjust mark orientation.

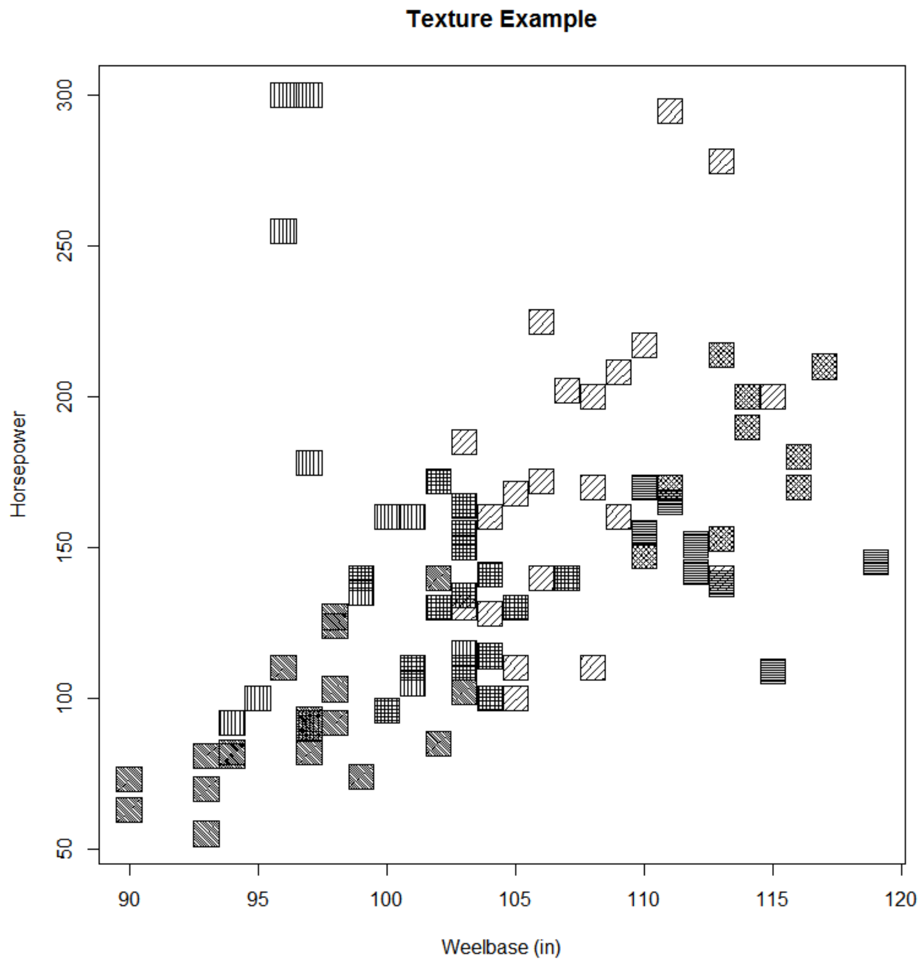
The Eight Visual Variables – Texture



Six possible example textures that could be used to identify different data values.

http://www.idvbook.com/

The Eight Visual Variables – Texture



Example visualization using texture to provide additional information about the 1993 car models data set, showing the relationship between wheelbase versus horsepower (position) as related to car types, depicted by different textures.

The Eight Visual Variables – Motion

- Can be associated with any of the other visual variables

Example : [Gapminder]

Designing Visualizations

Characteristics of Visual Variables

- Selective: Is X different from the others?
- Associative: Is X like the others?
- Order: Is X more/greater/bigger/... than Y?
- Quantitative: How much is the difference between X and Y?
- Length: How many different categories can we represent with this variable for a task?

Reference: Summary of:
S. Carpendale. 2003. Considering Visual Variables as a Basis for Visualization. Research report 2001-693-16, Department of Computer Science, University of Calgary.

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Characteristics of Visual Variables

	Selective	Associative	Quantitative	Order	Length
Position	✔	✔	✔	✔	✔
Size	✔	✔	≈	✔	(✔)
Shape	≈	≈	✗	✗	✔
Value	✔	✔	✗	✔	(✔)
Color (Hue)	✔	✔	✗	✗	(✔)
Orientation	(✔)	(✔)	✗	✗	(✔)
Grain	✔	(✔)	✗	✗	(✔)
Pattern	≈	≈	✗	✗	✔
Texture	✔	✔	✗	✗	✔
Motion	✔	✔	≈	✔	✔

Reference: Summary of:

S. Carpendale. 2003. Considering Visual Variables as a Basis for Visualization. Research report 2001-693-16, Department of Computer Science, University of Calgary.

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Using Visual Variables - Practically Speaking

- Don't use ordered variables for nominal data
- Use position for the most important information
- Many lengths are ~5 — a limiting factor
- For quantity, use (vertical) position, length, size, brightness, saturation
- For many distinctions: use size, shape or brightness
- Redundant encoding *may* make features easier to interpret

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Color– Some Notes

Color is powerful, but there are special considerations

- Hue is nominal, but can be given (cultural) orderings (e.g. temperature, elevation)
- Affective connotations of color vary across cultures
- Depending on who your users are, you may need to design for colorblindness or low-vision
  Allowing users to set colors / color schemes is the most flexible.

Resource:
Sim Daltonism: color blindness simulator for OS X (There other tools, and for other platforms)

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Color - Some Guidelines

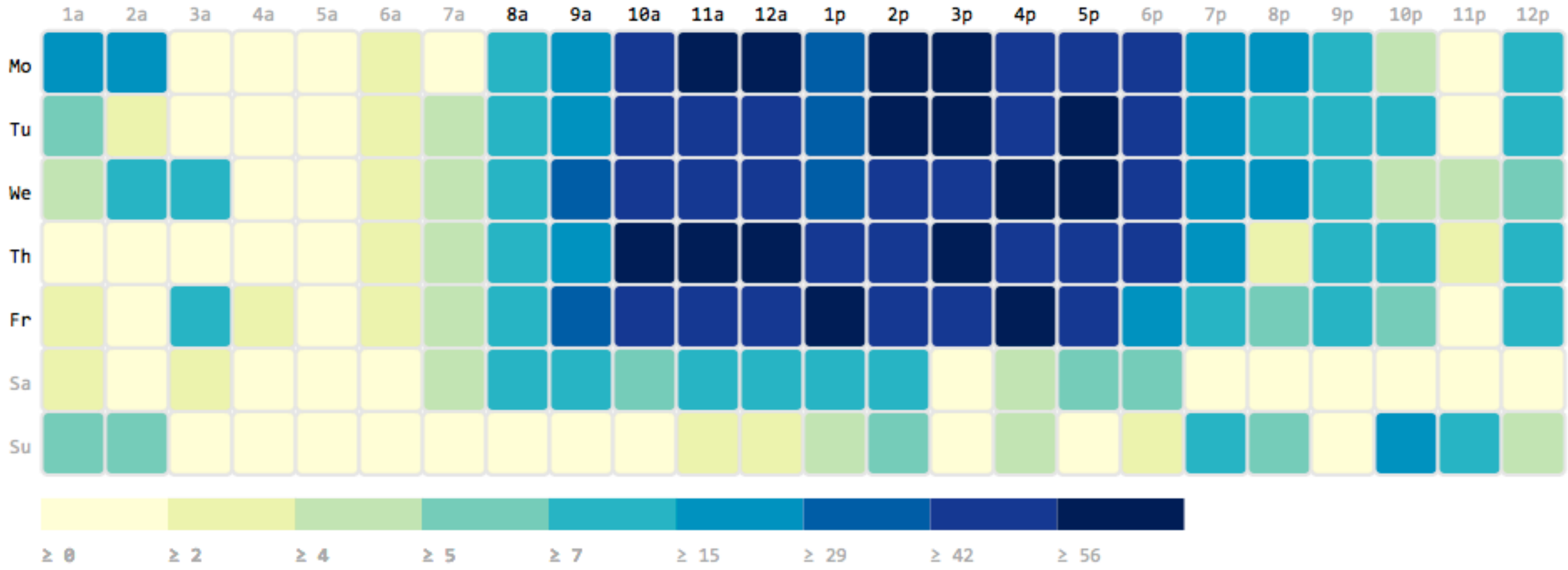
- Use lighter, less saturated colors for larger areas
- Use darker, more saturated colors for smaller areas and lines
- For ordinal data, use darker and more saturated for higher values (e.g. heat map)
- Color (hue) length is ~6-10, so with more categories, color may not be an appropriate visual variable.
- Saturation has a length of ~3
- Ensure a luminance contrast of 3:1 for text.

Resource:
Color Brewer

Reference: Slides provided by Chris Culy (http://ling.uni-konstanz.de/pages/home/butt/main/material/esslli14-vis/CuC_slides/reveal-based/theory.html#/title)

Heatmap Example

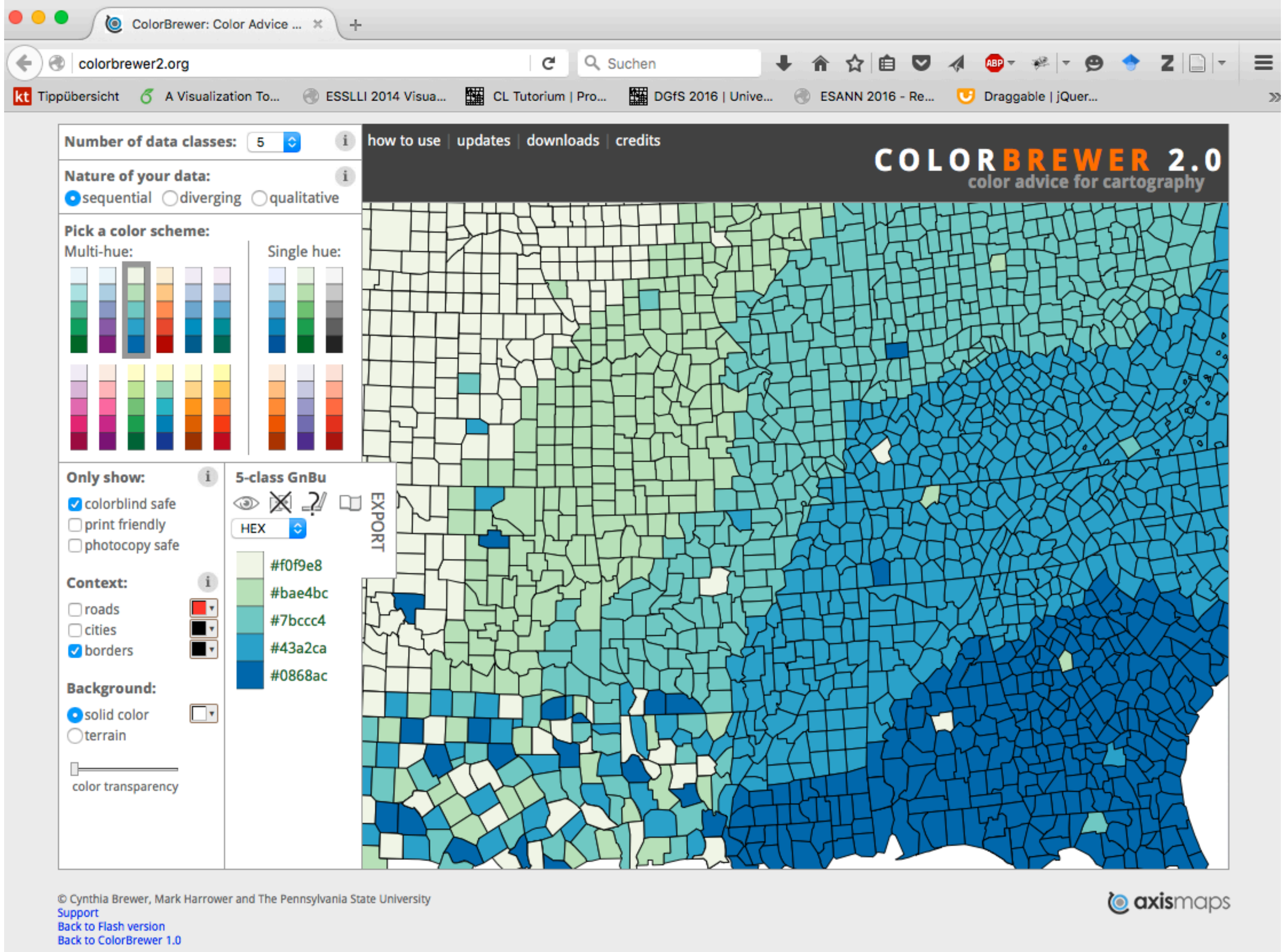
Example data shows concurrent user sessions over time, taken from a development environment.



http://bl.ocks.org/tjdecke/5558084

ColorBrewer

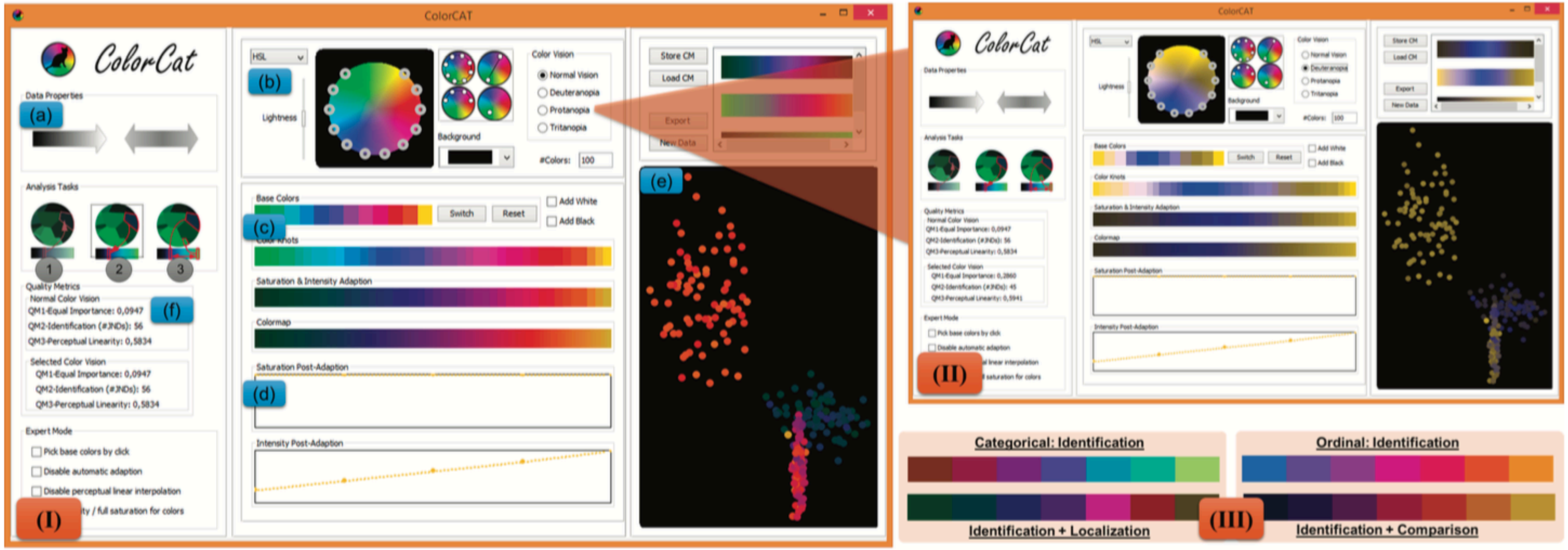
Color Brewer



http://colorbrewer2.org/

More Advanced - ColorCat

Colormaps for **C**ombined **A**nalysis **T**asks.



https://github.com/SebastianMittelstaedt/ColorCAT

S. Mittelstädt, D. Jäckle, F. Stoffel and D. A. Keim. ColorCAT: Guided Design of Colormaps for Combined Analysis Tasks. Eurographics Conference on Visualization (EuroVis) - Short Papers, The Eurographics Association, DOI: 10.2312/eurovisshort.20151135, 2015.

Designing Visualizations

- Visualizations help:
  - To describe some structure, patterns or anomaly in the data.
  - To explore and analyze large datasets.
  - To make effective use of the information overflow.
  - To communicate information to people.

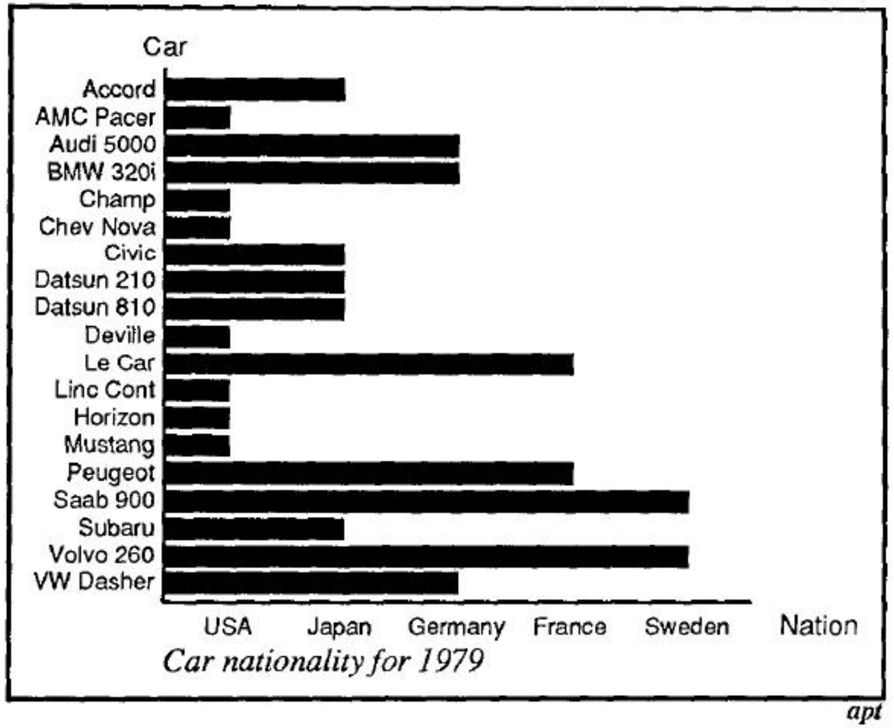
- **Visualization can distort the “truth”!**

Designing Visualizations

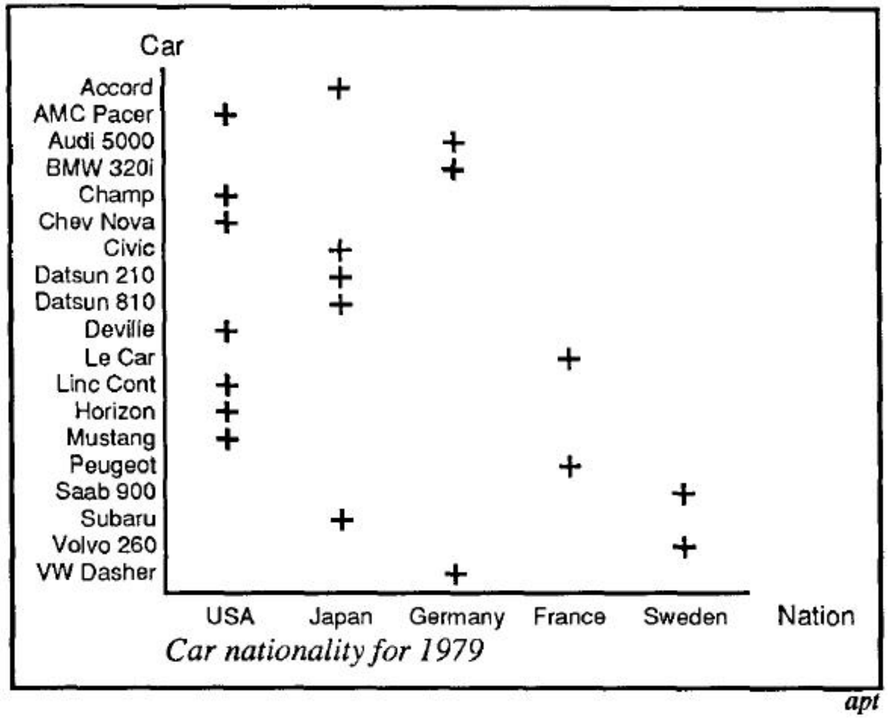
- **Expressiveness**: Visualization presents all the information and only the information.

- **Effectiveness**: Visualization is effective when it can be interpreted accurately and quickly and when it can be rendered in a cost-effective manner.

How to Visualize Badly



Poor use of a bar chart.



Better use of a scatterplot.

http://www.idvbook.com

Interaction

Interactive Analysis Process with Visualizations

- **Shneiderman's "Information Seeking Mantra"**

“Overview first. Zoom and Filter. Details on Demand”

- **Keim's Visual Analytics Mantra**

“Analyze first, show the important, zoom, filter and analyze further, details on demand”

References:
B. Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. Proceedings of the 1996 IEEE Sympoium on Visual Languages, VL '96.
D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. 2006. Challenges in visual data analysis. Proceedings of the conference on Information Visualization, IV '06.

Interaction Types

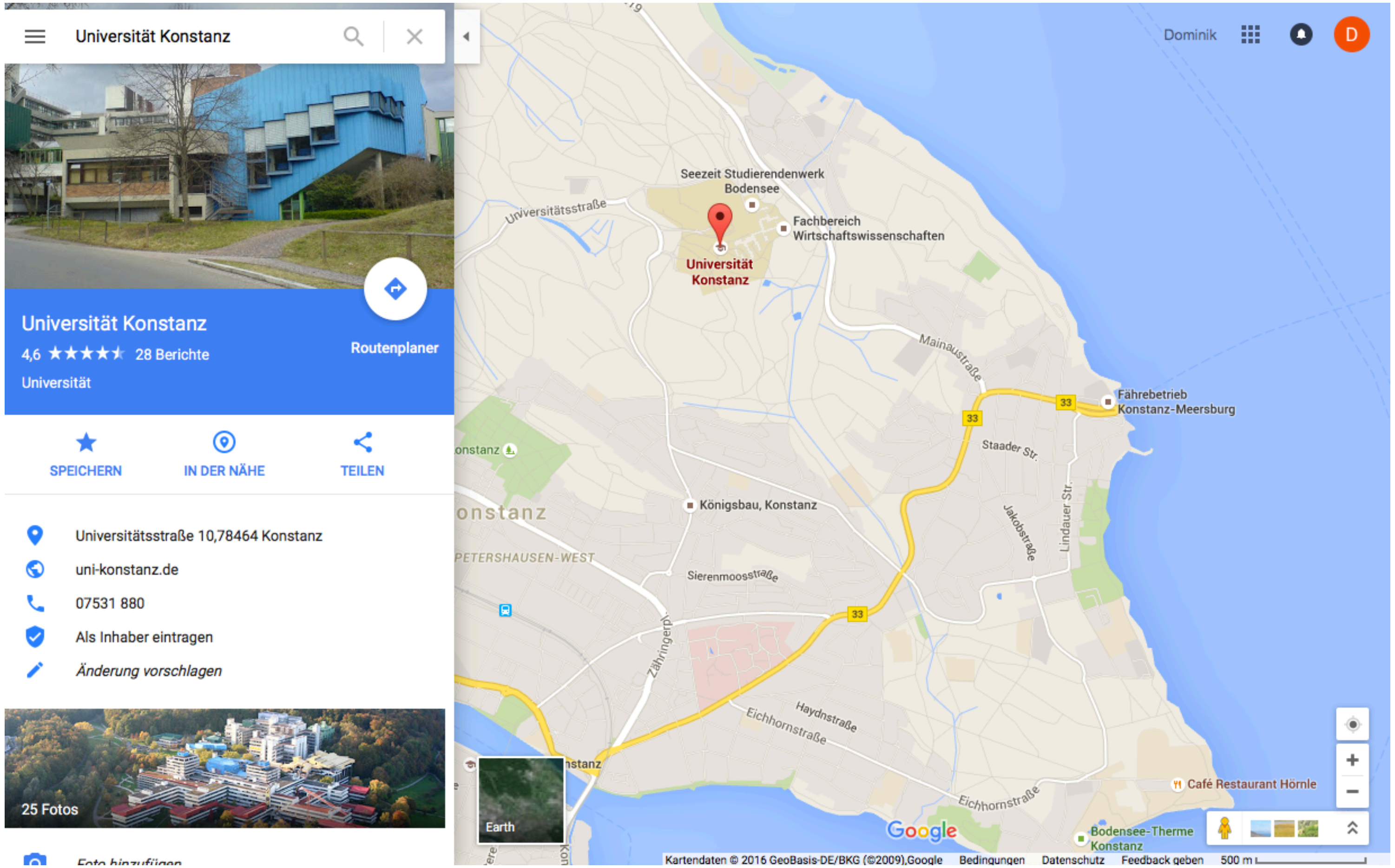
- Select: mark something as interesting
- Explore: show me something else
- Reconfigure: show me a different arrangement [same type]
- Encode: show me a different representation [different type]
- Abstract/Elaborate: show me less or more detail
- Filter: show me something conditionally
- Connect: Show me related items
- Some other things: history, annotate, extract

References:

J.S. Yi, Y.A. Kange, J.T. Sasko, J.A. Jacko. 2007. Toward a deeper understanding of the role of interaction in information visualization. IEEE Transactions on Visualization and Computer Graphics. 13:6

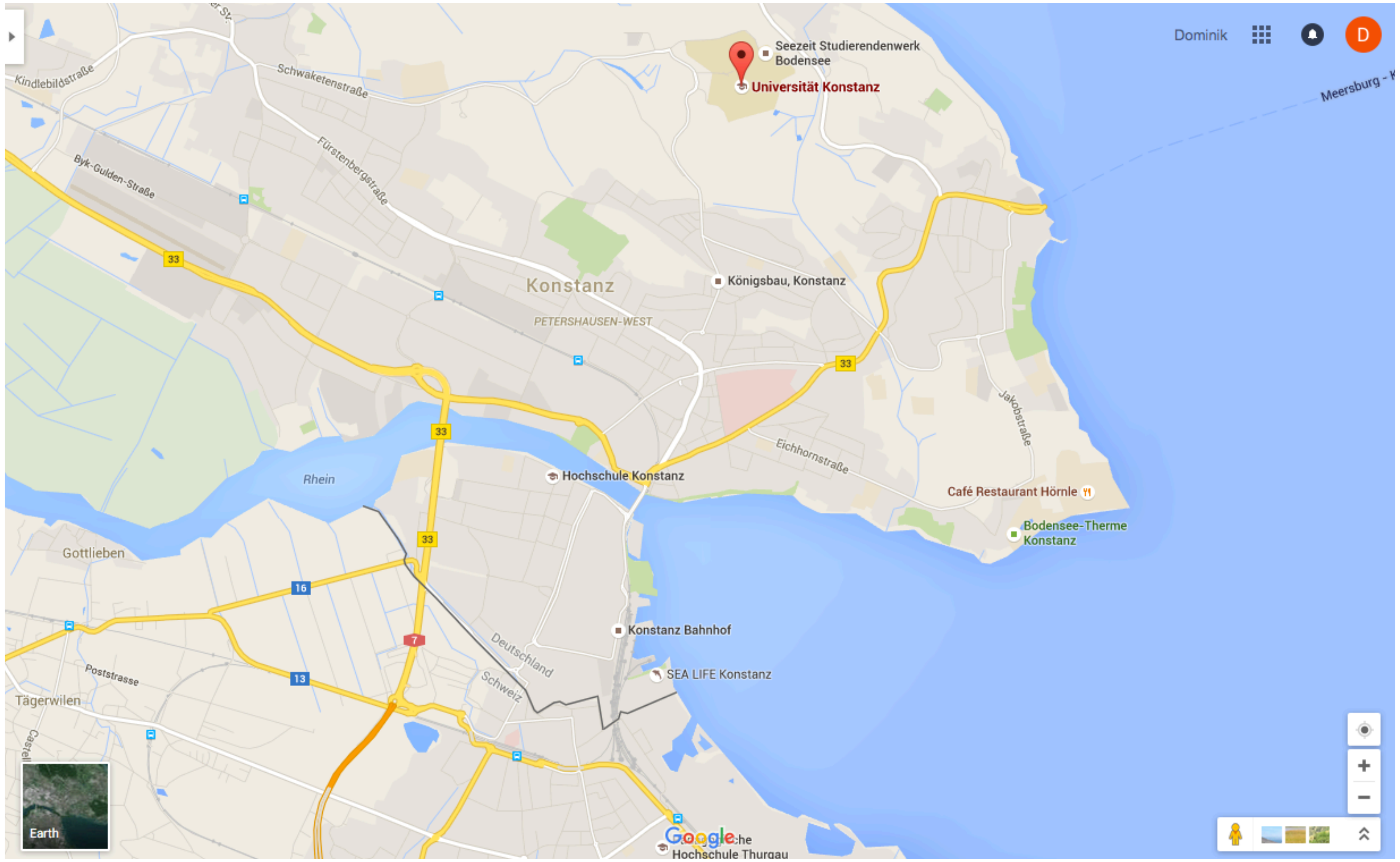
B. Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. Proceedings of the 1996 IEEE Sympoium on Visual Languages, VL '96.

Interaction – Select   Example



https://www.google.de/maps/

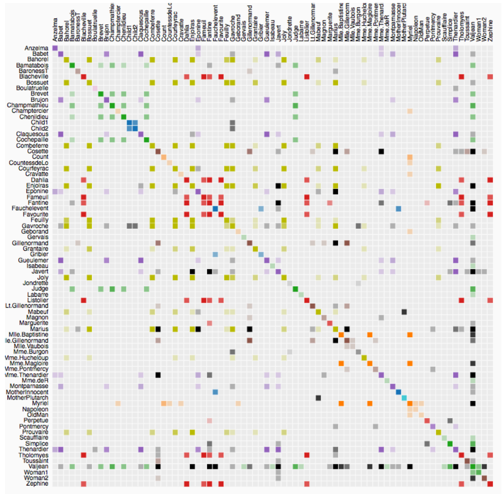
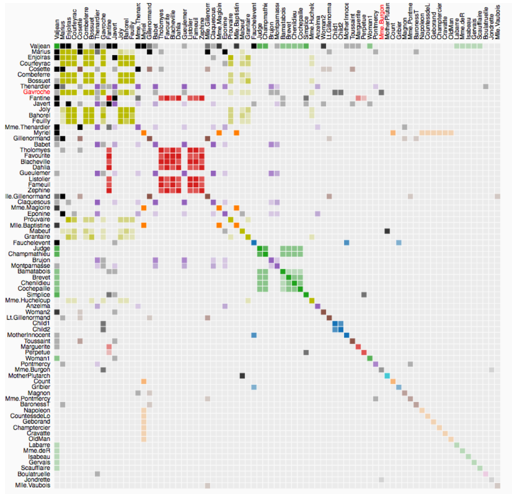
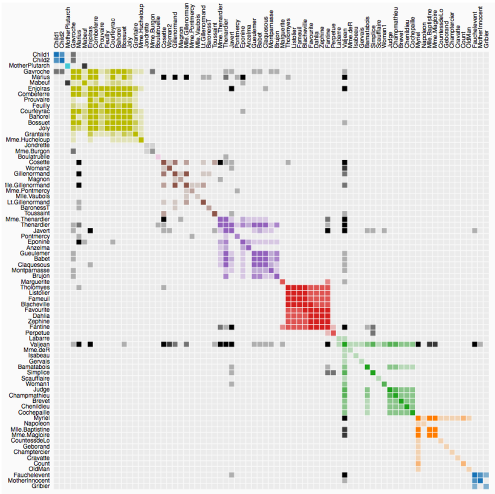
Interaction - Explore [Example]



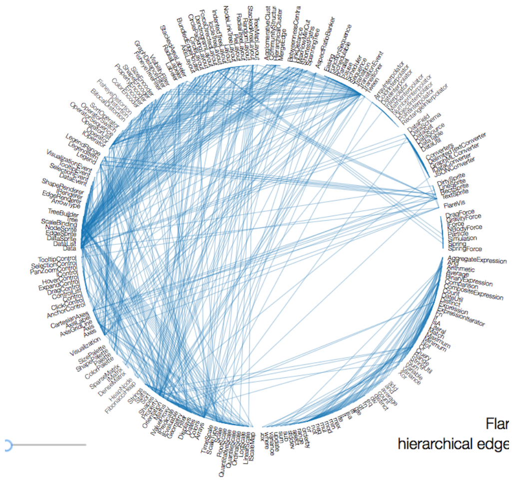
https://www.google.de/maps/

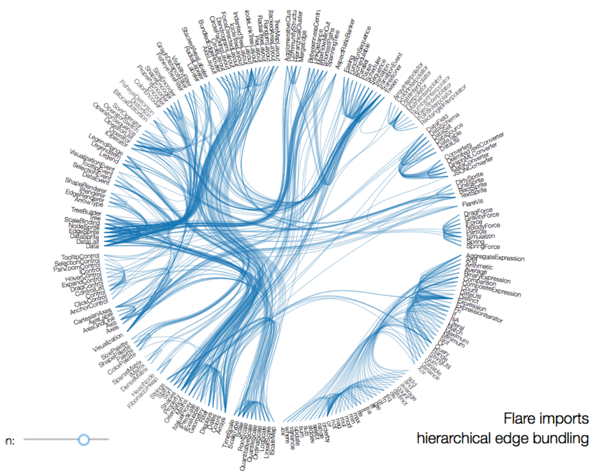
Interaction - Reconfigure

Example Example

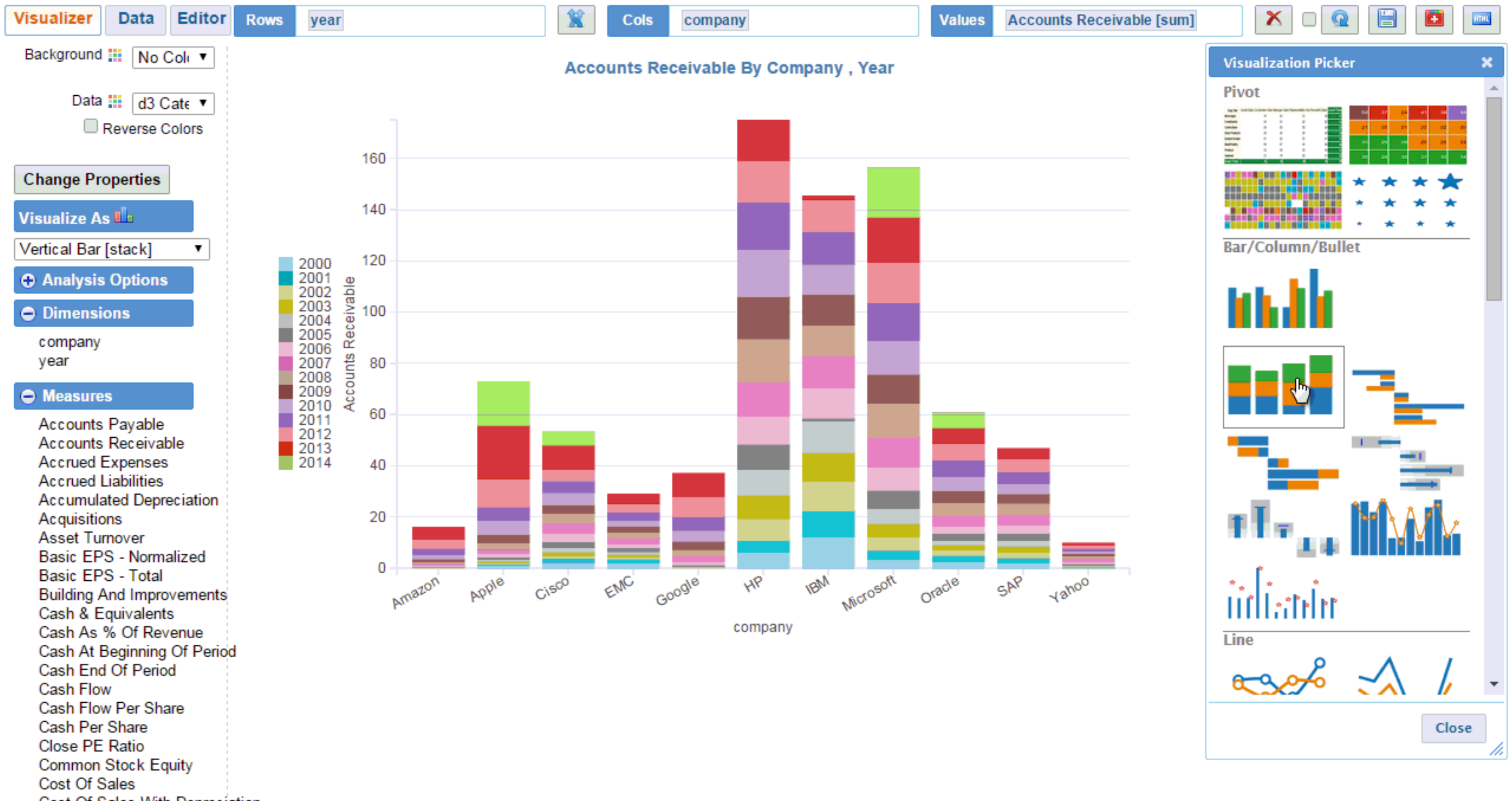
https://bost.ocks.org/mike/miserables/





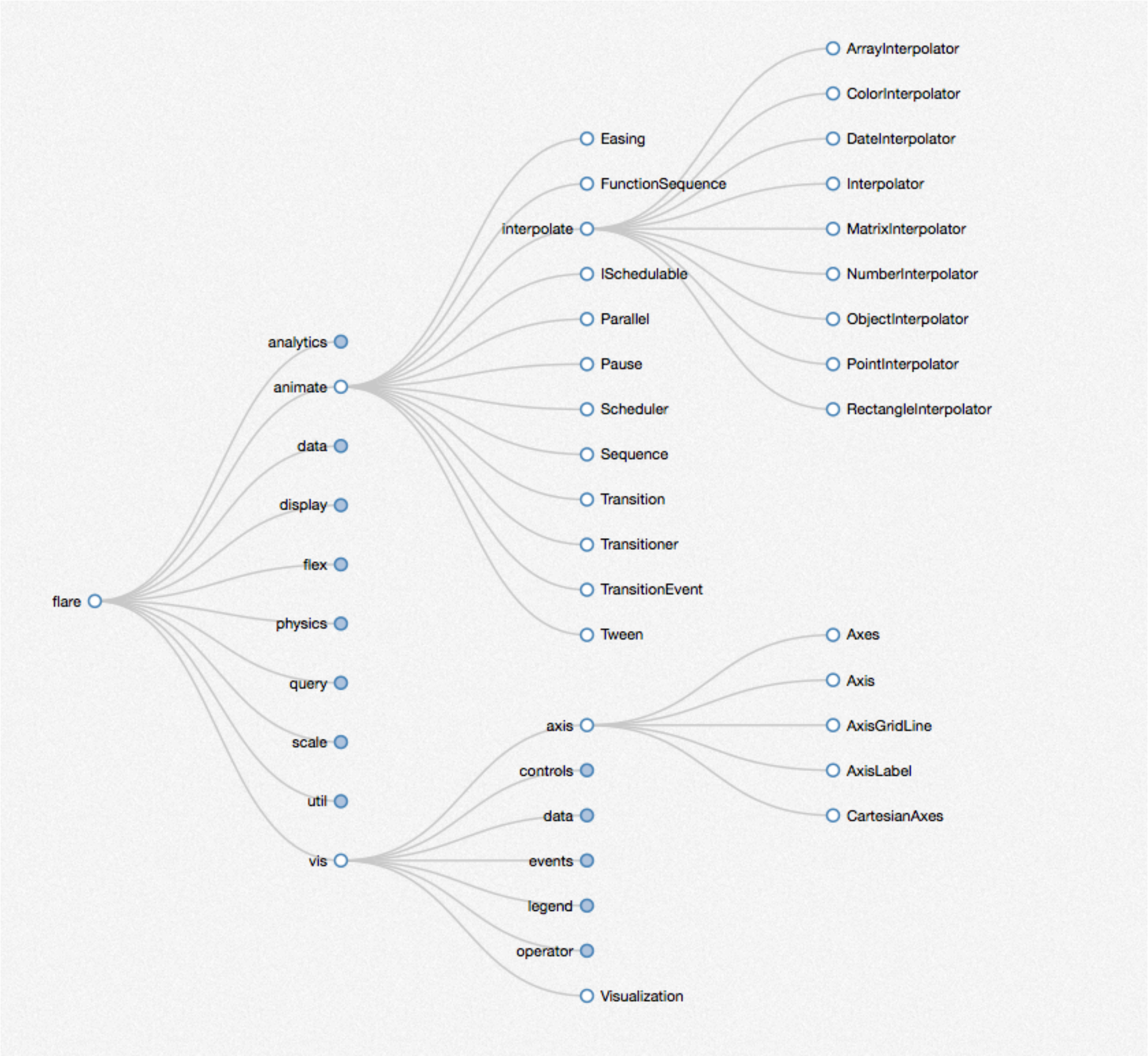
http://mbostock.github.io/d3/talk/20111116/bundle.html

Interaction - Encode Example



http://www.infocaptor.com/

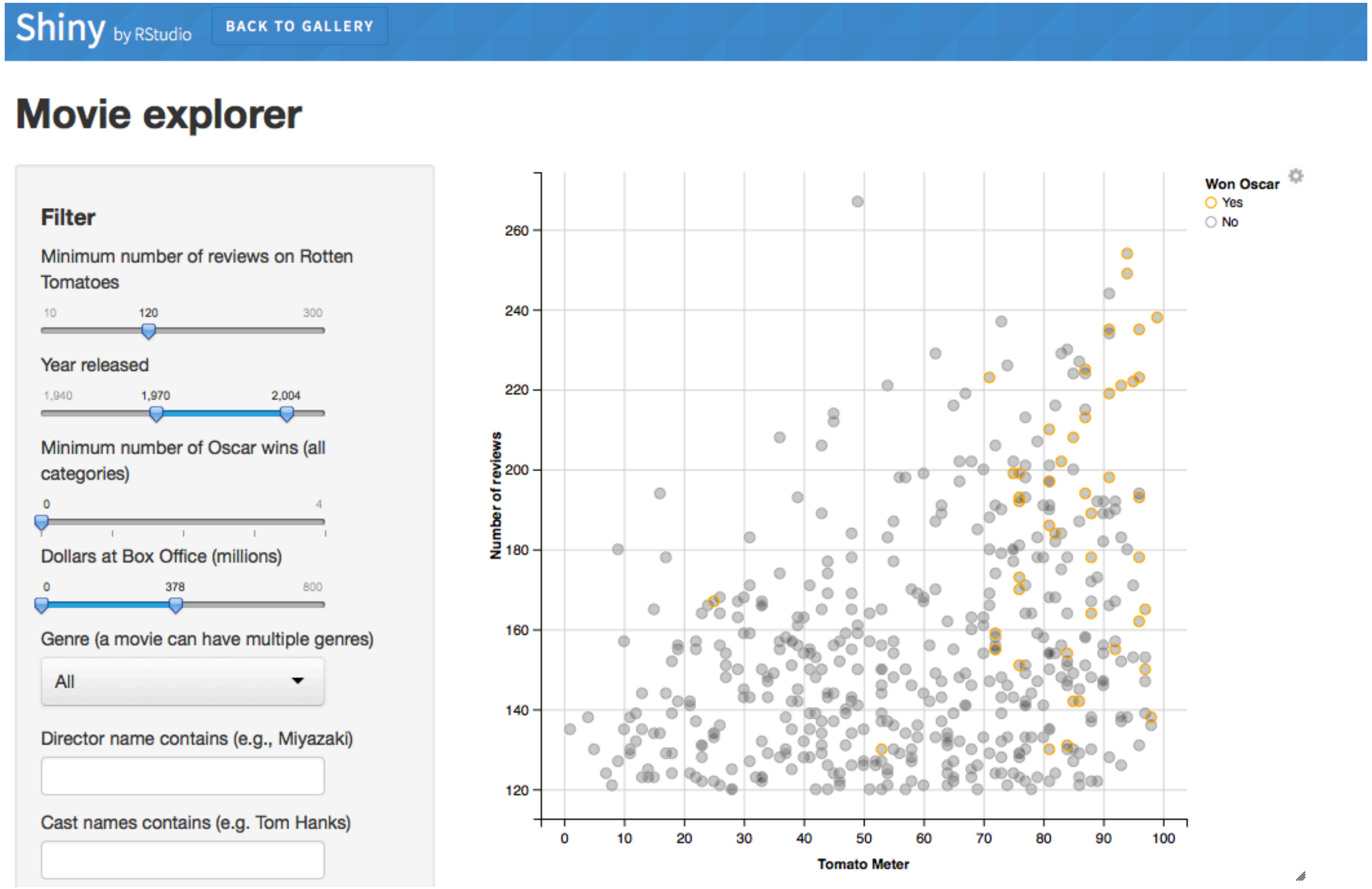
Interaction – Abstract/Elaborate   Example



http://mbostock.github.io/d3/talk/20111018/tree.html

Interaction – Filter

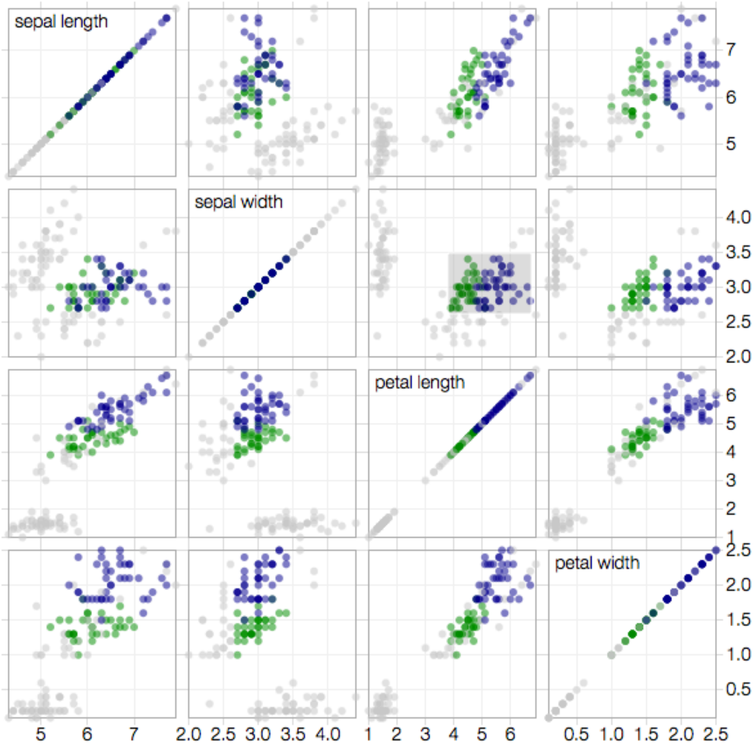
Example



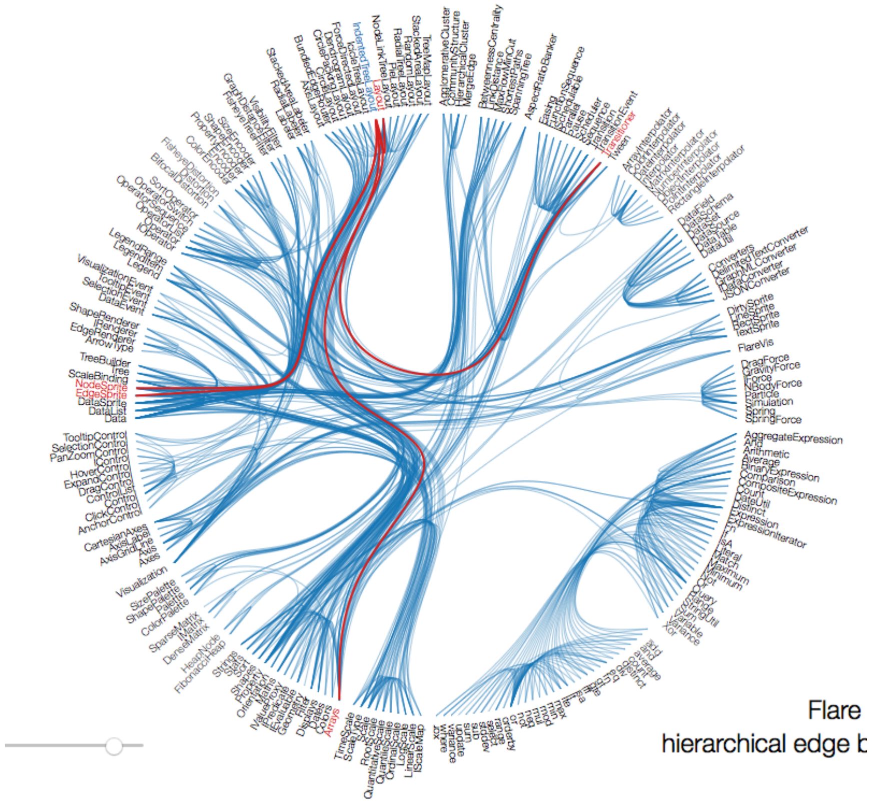
http://shiny.rstudio.com/gallery/movie-explorer.html

Interaction – Connect

[Example](http://mbostock.github.io/d3/talk/20111116/iris-splom.html) [Example](http://mbostock.github.io/d3/talk/20111116/bundle.html)



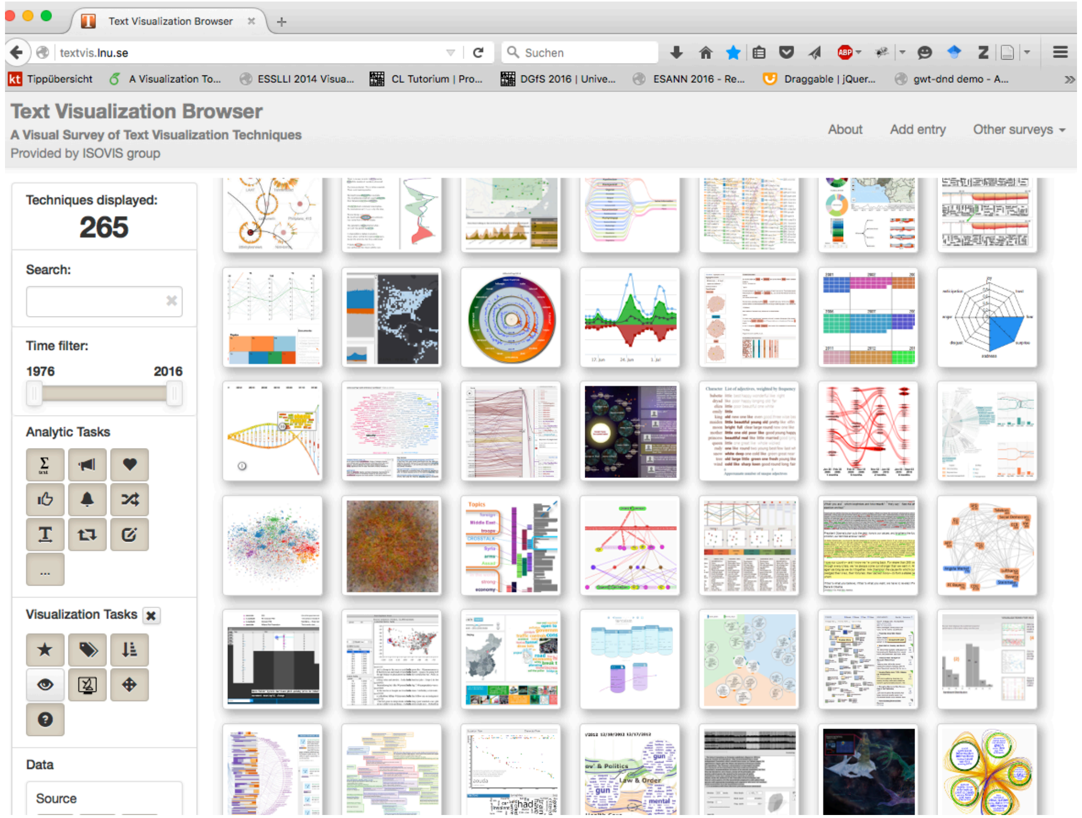
http://mbostock.github.io/d3/talk/20111116/iris-splom.html

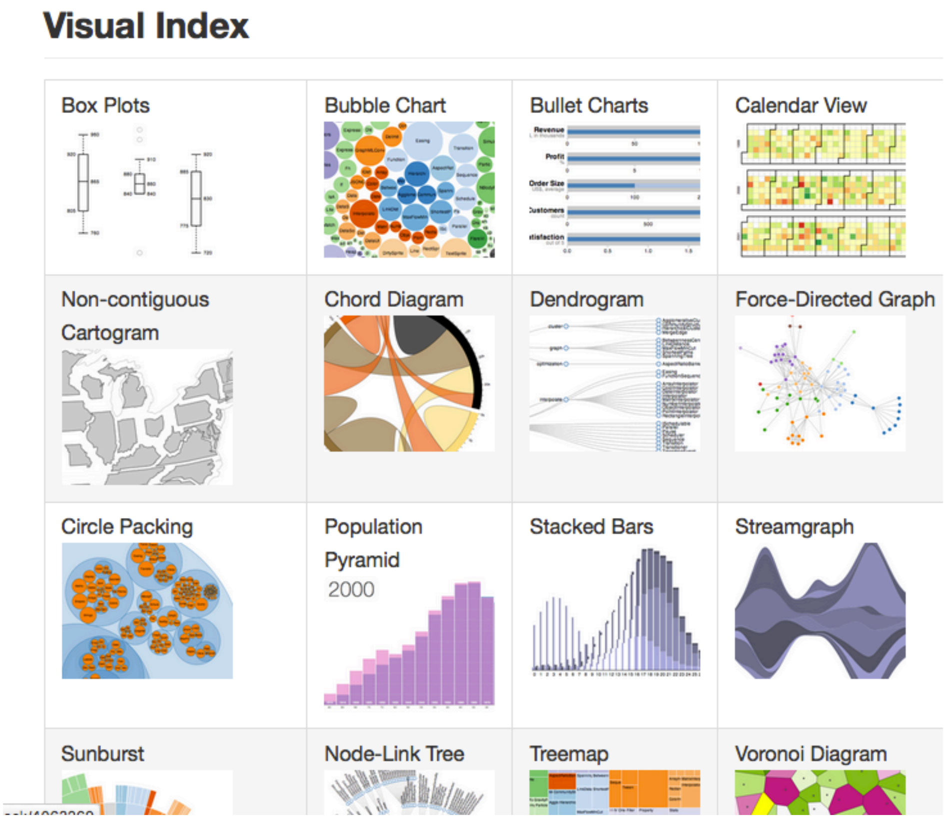


http://mbostock.github.io/d3/talk/20111116/bundle.html

Many Design Alternatives! – Checkout Existing Visualizations

- http://textvis.lnu.se/
- https://github.com/mbostock/d3/wiki/Gallery



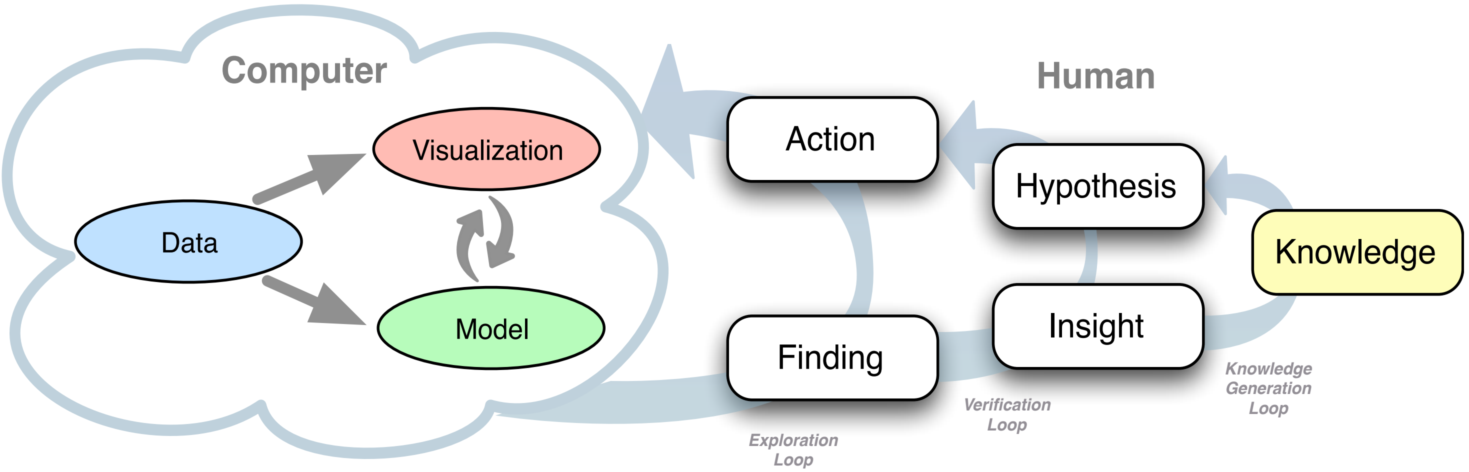


THANK YOU!

Questions?

For more information
about our work
please contact

Dominik Sacha
Tel. +49 7531 88-4046
dominik.sacha@uni-konstanz.de



http://vis.uni-konstanz.de/