

Visualization

Strategies and Principles

Statistical Graphics

- to understand our data
- to present conclusions

- convey the data clearly
- focus on key features
- easy to understand

Statistical Graphics

- convey the data clearly
- focus on key features
- easy to understand

- research in perception
- aspects of cognitive science

- must turn 'big data' into small data

Information Visualization

- <http://www.infovis.org>
- a process of transforming information into visual form
- relies on the visual system to perceive and process the information
- <http://ieevis.org/>
- involves the design of visual data representations and interaction techniques

Sheelagh Carpendale

<http://innovis.cpsc.ucalgary.ca/>

- representation
- presentation
- interaction

- “information visualization is the use of computer-supported, interactive, visual representations of data to amplify cognition”
Card et al. 1999
- <http://mariandoerk.de/edgemaps/demo/ - phils;map;:/en/aristotle:>

Scientific Visualization

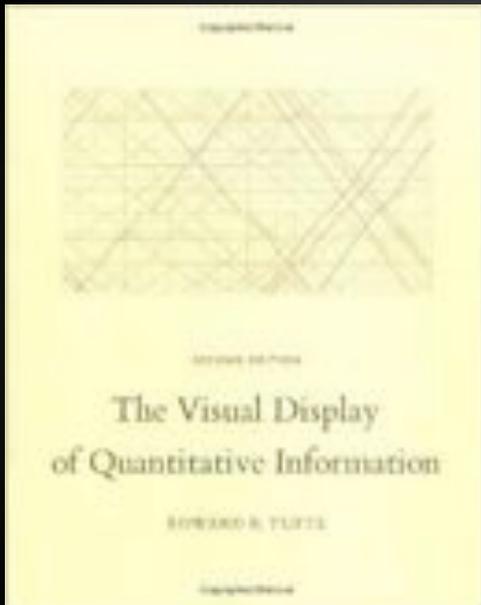
- <http://ieevis.org/>
- advances understanding or provides solutions for real-world problems
- impacts a particular application
- Katy Borner
<http://info.slis.indiana.edu/~katy/index.html>
- http://scimaps.org/maps/map/exploring_the_relati_180/detail

Visualization “for the masses”

[Alex Gonçalves](#)

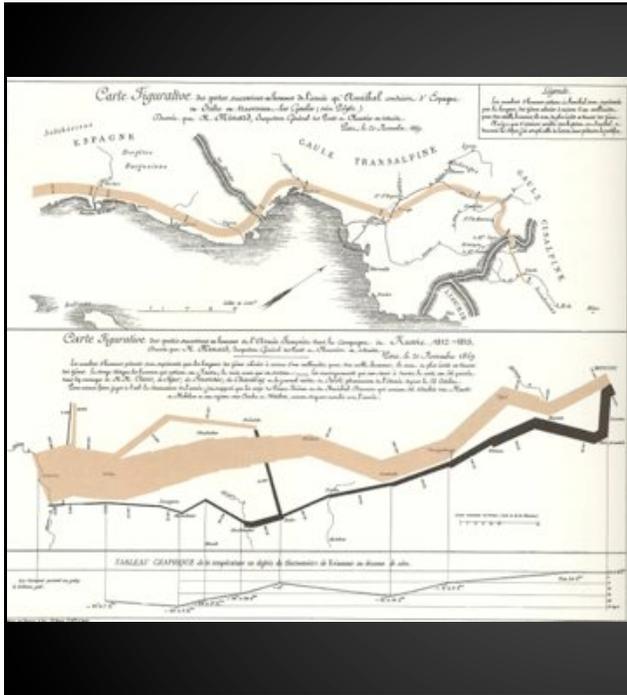
- to build communion
- for social change
- powerful stories
- “duty of beauty”

<http://www.nytimes.com/newsgraphics/2014/02/14/fashion-week-editors-picks/>



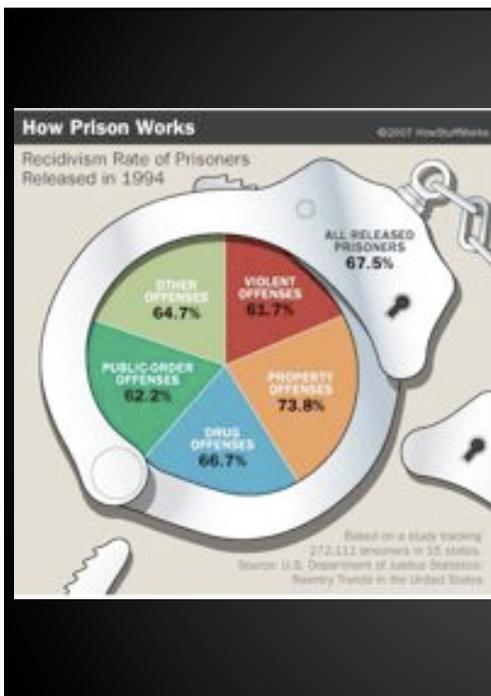
Graphical Excellence

- show the data
- convey the substance
- avoid distortion
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail
- serve a clear purpose
- be closely integrated with statistics



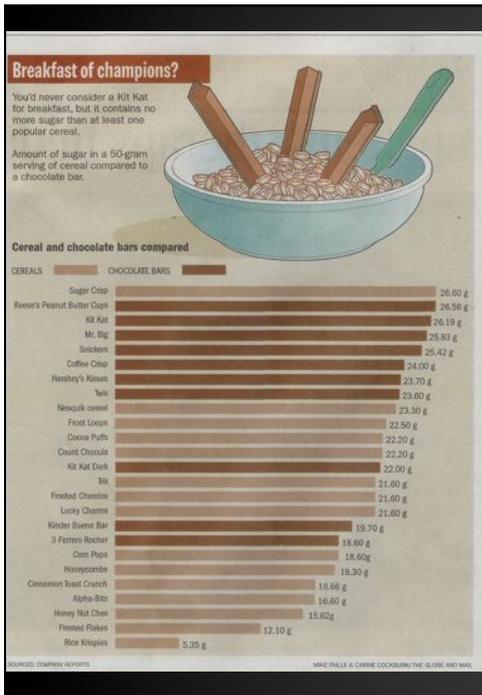
Graphical Excellence

- well-designed presentation
- substance, statistics, design
- clarity, precision, efficiency
- least ink in the smallest space
- telling the truth



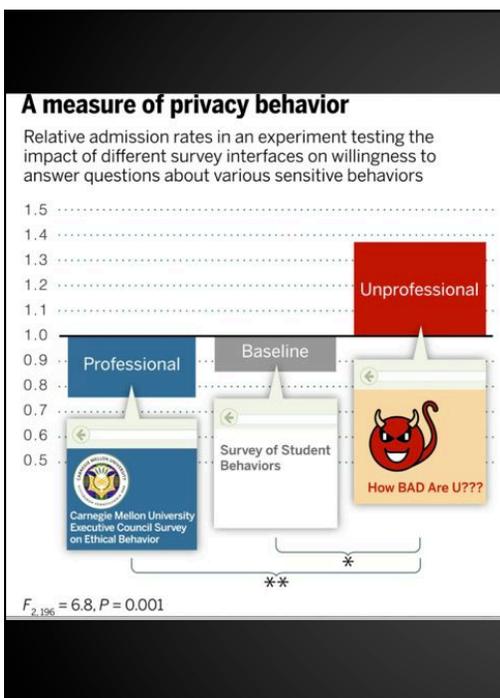
Graphical Integrity

- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context



Graphical Integrity

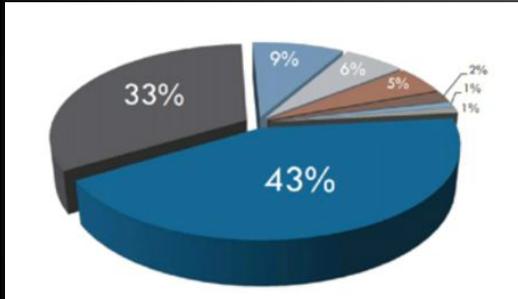
- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context



Graphical Integrity

- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context

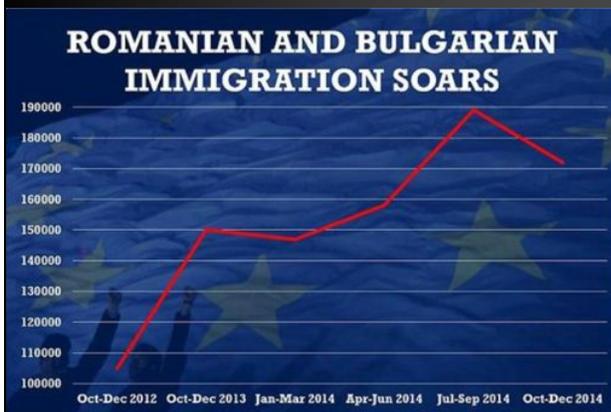
Graphical Integrity



viz.wtf

- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context

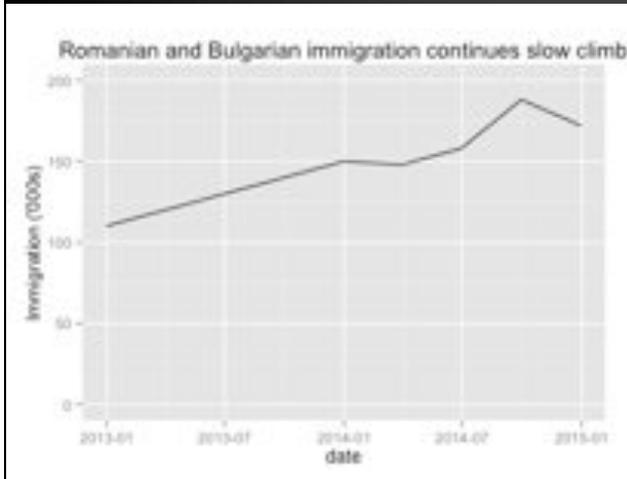
Graphical Integrity



viz.wtf

- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context

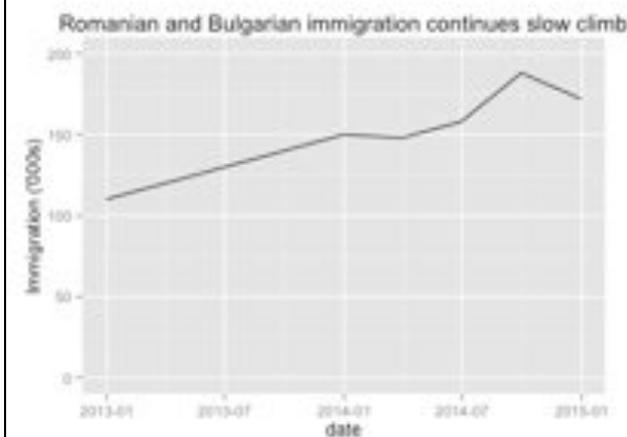
Graphical Integrity



ggplot2

- Physical representation of numbers should be proportional to represented quantities
- Clear and detailed labelling
- Show data variation, not design variation
- Dimension matching
- Context

Graphical Integrity



Physical representation of numbers should be proportional to represented quantities

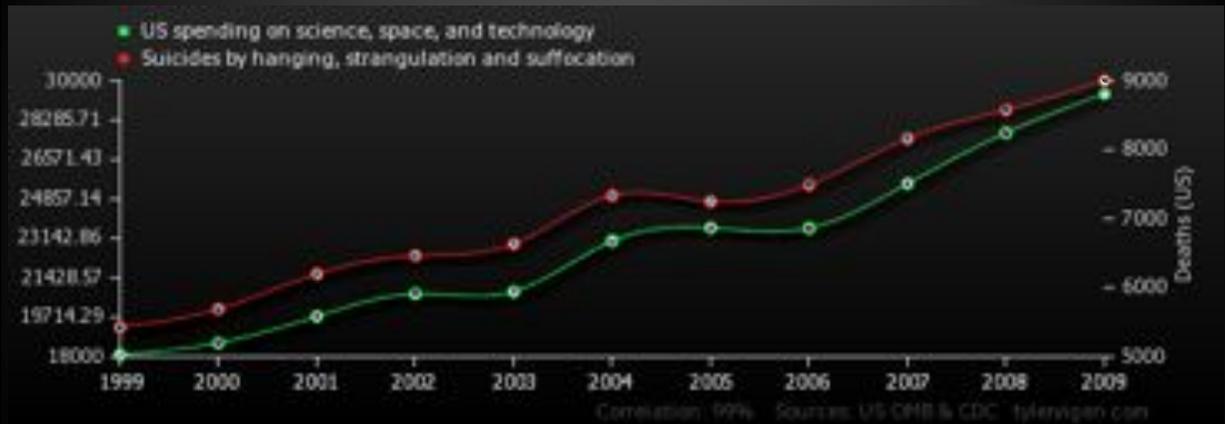
Clear and detailed labelling

Show data variation, not design variation

Dimension matching

Context

Graphical Integrity

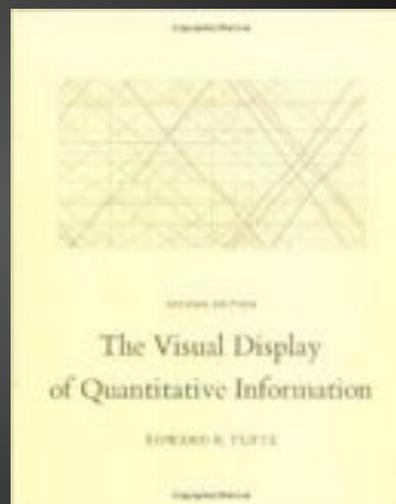


- Context

tylervigen.com

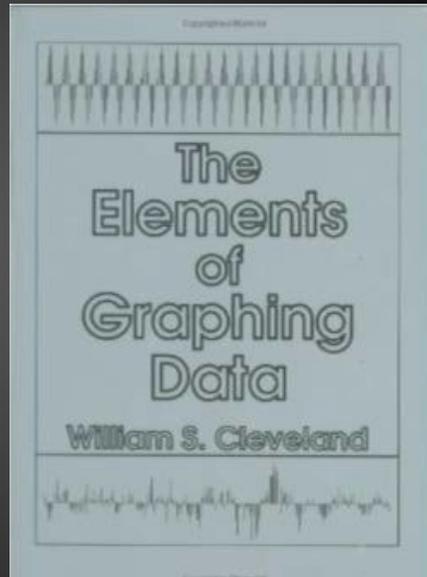
Some Design Concepts

- Increase data ink ratio
- Eliminate chart junk
 - grid lines, shading, artificial dimensions
- Use multifunctioning elements
 - rug plots, stem-and-leaf plots
- Increase data density
- The value of small multiples
- Aesthetics and technique



Statistical Graphics

1. Principles of Graph Construction
2. Graphical Methods
3. Graphical Perception



1. Principles of Graph Construction

Terminology

tick mark, tick mark label, plotting symbol, panel, axis, line type, data label, scale label

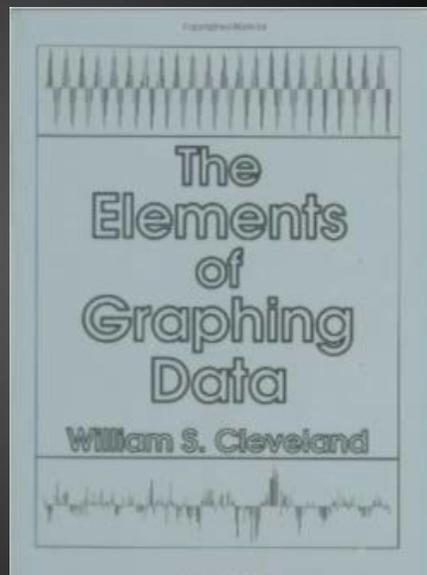
Clear Vision

Make the data stand out

Clear Understanding

Scales

Strategy



2. Graphical Methods

Distributions

Dot plots, histograms, densities, ...

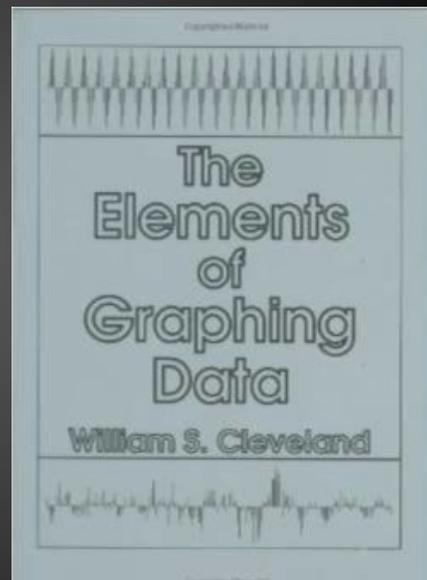
Scatterplots

Reference grids, log transformations, plotting symbols, curve types

Coplots

Brushing

Colour



3. Graphical Perception

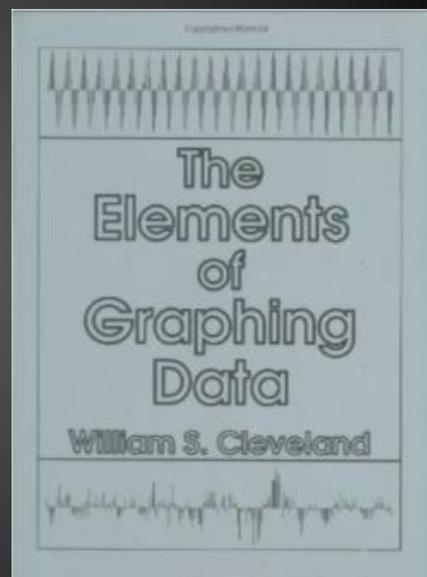
Models

Superimposed curves

Colour encoding

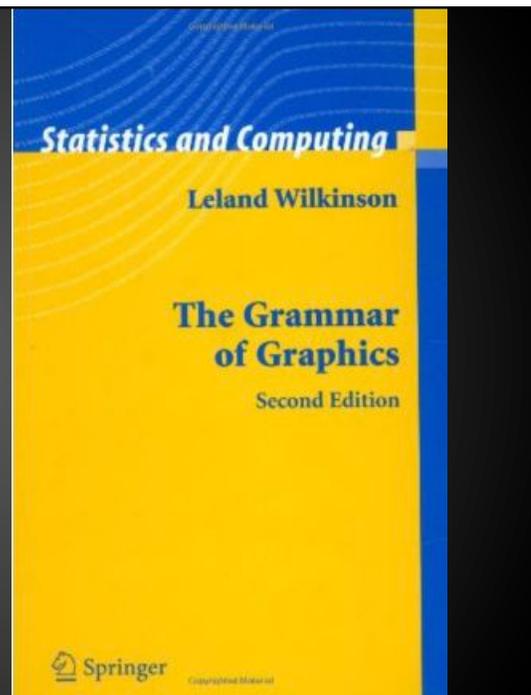
Texture symbols

Common scale



“rules for constructing
graphs mathematically
and representing them
as graphs
aesthetically”

1. specification
2. assembly
3. display



1. Specification

Data variables from datasets

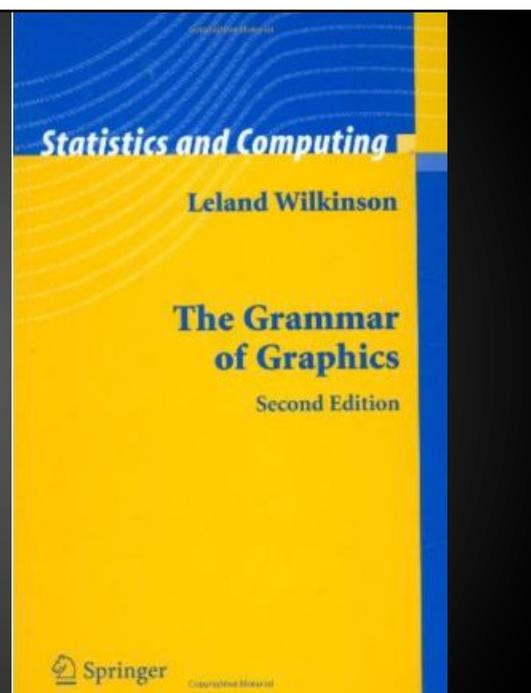
Trans Operations on variables,
e.g. *rank*

Scale scale transformations,
e.g. *log*

Coord Coordinate System,
e.g. *Cartesian, polar*

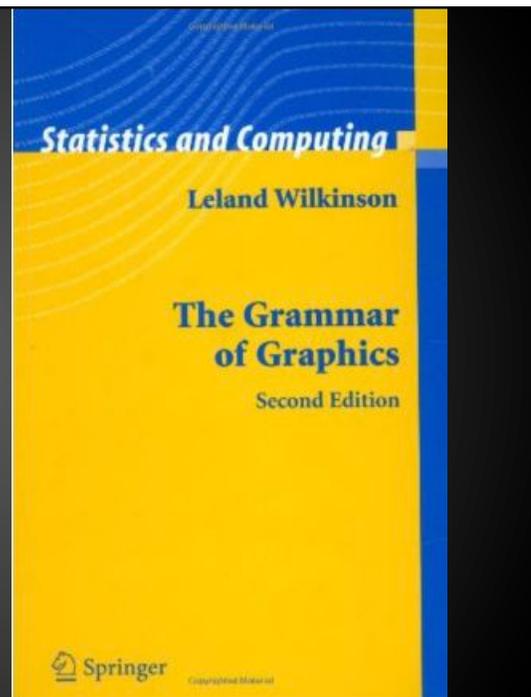
Element graphs e.g. *points*
and their attributes, e.g. *colour*

Guide *axes, legends*



2. Assembly

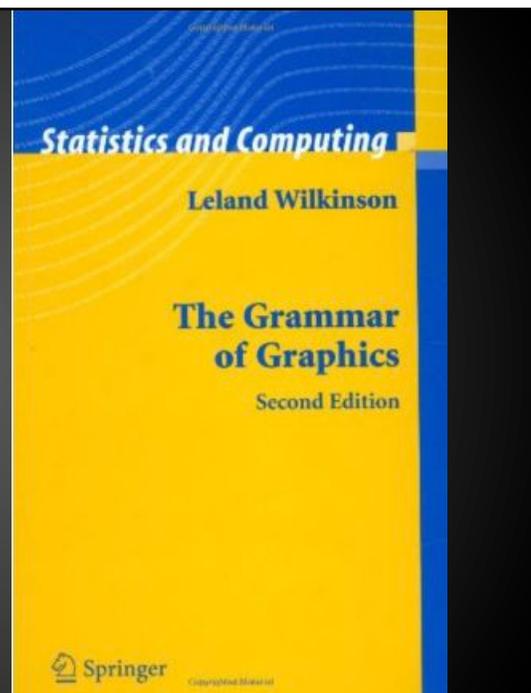
“In order to portray a scene, we must coordinate its geometry, layout, and aesthetics”



3. Display

“render the graph using aesthetic attributes and a display system”

- static, dynamic,
scientific visualization



Review Paper

Advanced Review

The grammar of graphics

Leland Wilkinson*



The grammar of graphics (GoG) denotes a system with seven classes embedded in a data flow. This data flow specifies a strict order in which data are transformed from a raw dataset to a statistical graphic. Each class contains multiple methods, each of which is a function executed at the step in the data flow corresponding to that class. The classes are orthogonal, in the sense that the product set of all classes (every possible sequence of class methods) defines a space of graphics which is meaningful at every point. The meaning of a statistical graphic is thus determined by the mapping produced by the function chain linking data and graphic. © 2010 John Wiley & Sons, Inc. *WIREs Comp Stat* 2010 2 673-677 DOI: 10.1002/wics.118

Keywords: visualization; statistical graphics

The grammar of graphics

Object oriented graphics system, for creating and automating statistical graphics

A system with seven (orthogonal) classes

A description of what it means to construct a statistical graphic

Not

- a command language;
- taxonomy;
- drafting package;
- "book of virtues";
- heuristic system;
- GIS;
- visualization system

Seven Orthogonal Classes

- Variable: mapping of object to set of values, *varset*: variable set
- Algebra: operation on varset to produce combinations; *cross*, *nest*, *blend*
- Scales: *log*, *category*, *order*, *probability*
- Statistics: *identity*, *summaries*, *smoothers*, *confidence intervals*
- Geometry: *point*, *line*, *area*, *path*
- Coordinates: *Cartesian*, *polar*
- Aesthetics: *position*, *size*, *shape*, *orientation*, *brightness*, *colour*, *granularity*

Examples of implementation

- Java Graphics Production Language GPL <http://www.graphics.stanford.edu/projects/polaris/>
- marketed by Illumitek Inc as nViZn <http://mbostock.github.io/protovis/>
- Rewritten, with XML interface and refinements at SPSS <http://d3js.org>
- Polaris system for producing visualizations <http://d3js.org>
- evolved into Tableau visualization system [A survey of visualization construction user interfaces](#) L Grammel, C Bennett, M Tory, MA Storey - EuroVis-Short Papers, 2013
- ggplot2
- Protovis, now D3.js

ggplot2's grammar of graphics

There are 5 components to a layer: Data, Aesthetics, Statistics, Geometry, and Positional Adjustment.

Carson Sievert, Iowa State

Wickham, H. (2009)
Elegant Graphics for
Data Analysis



ggplot2

- statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars)
- also requires specification of: transformations, coordinate system, and possible faceting/conditioning
- shorthand: `data`, `geom`, `stat`, `scale`, `coord`, `facet`
- `ggplot`: implements `ggplot2`, using syntax similar to the `plot` function in base R

Wickham, H. (2009)
Elegant Graphics for
Data Analysis

