Toronto Metropolitan University | TED ROGERS SCHOOL OF MANAGEMENT | Centre for Digital Enterprise Analytics and Leadership

Business Analytics and Data Driven Decision Making

Session#05: Lecture#09_#10: Visual Analytics: Concepts, Methods & Tools

Ravi Vatrapu

Director, Centre for Digital Enterprise Analytics and Leadership (DEAL)

Loretta Rogers Research Chair in Digital Enterprise

Professor, Department of Information Technology Management

Ted Rogers School of Management

Toronto Metropolitan University, Canada

Founding Director, Centre for Business Data Analytics (bda.cbs.dk)

Professor, Copenhagen Business School, Denmark

Visiting Professor, Indian Institute of Management Visakhapatnam, India

Honorary Visiting Professor, GITAM Deemed University, India

Adjunct Faculty, Indian Institute of Management Rohtak, India

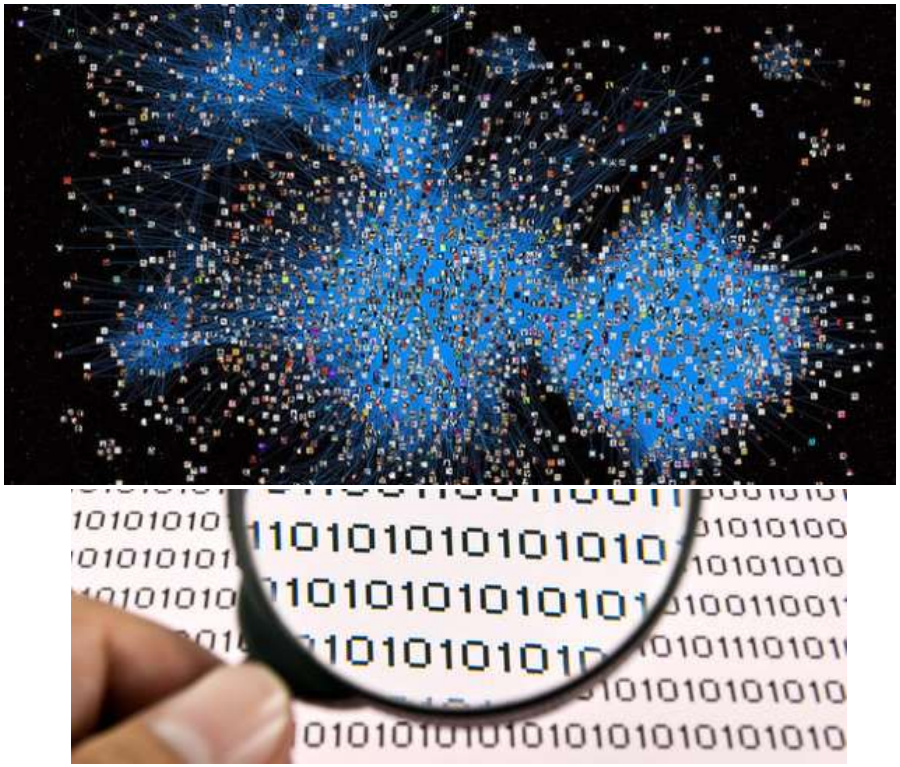
Email: vatrapu@torontomu.ca

DEAL Website: https://www.torontomu.ca/tedrogersschool/digital-enterprise-analytics-and-leadership/

Faculty *Web*: https://www.torontomu.ca/information-technology-management/faculty-research/ravi-vatrapu/

CBS CENTRE FOR BUSINESS DATA ANALYTICS
TRANSFORMING BIG DATA SETS INTO BUSINESS ASSETS

Data ”Explosion”

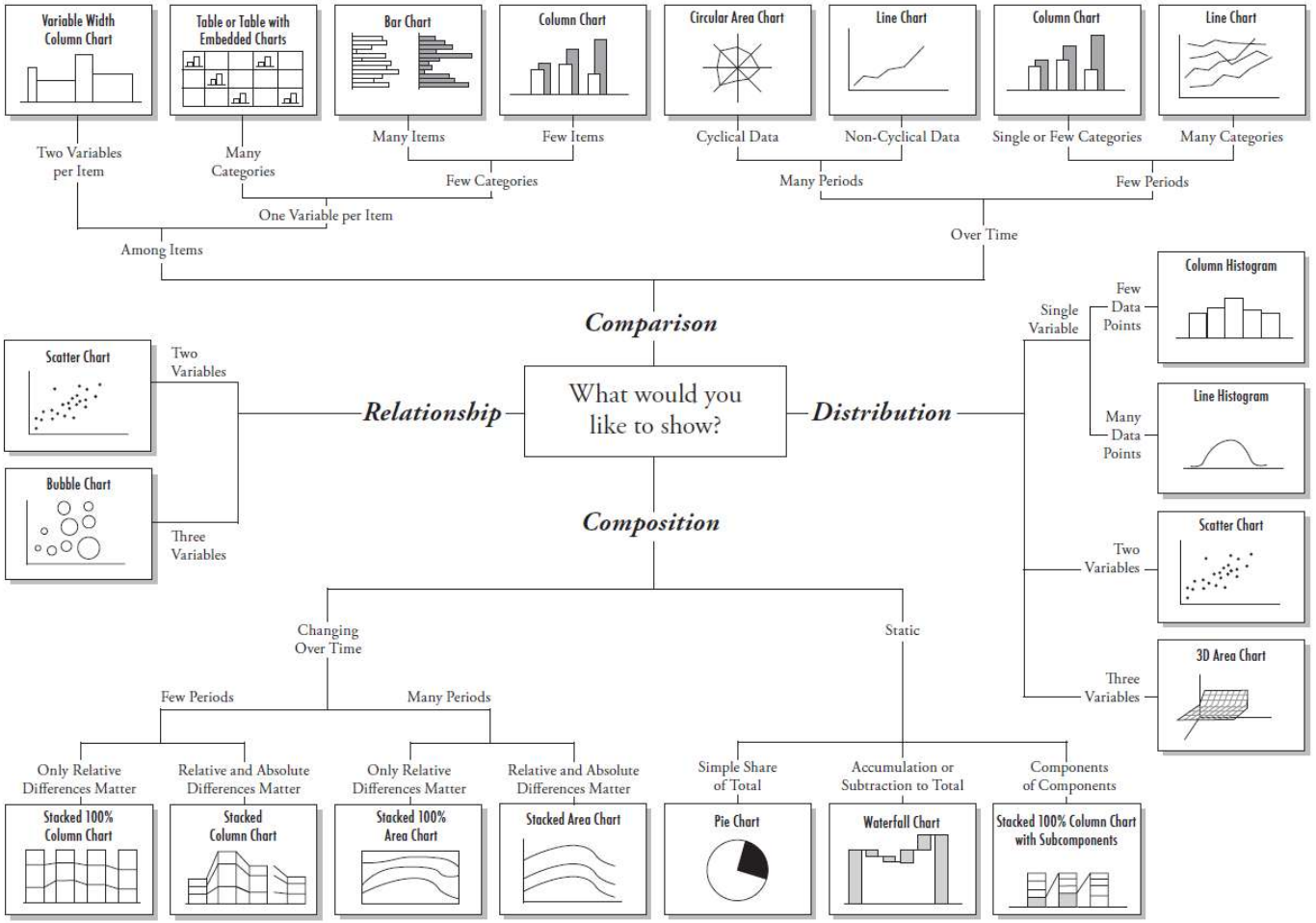


Risk: WORN (write-once, read-never) data

What is Visual Analytics Good For?

- Spotting **Trends**
- Identifying **Patterns**
- Detecting **Anomalies**
- Inferring **Reasons**

What is the Correct Visualization?



Andrew Abela's "Chart Suggestions" diagram

Information Visualization vs. Visual Analytics

- **Information Visualization**
  - 1980's
  - Visual-information seeking mantra (Shneiderman, 1996):
    - Overview first,
    - zoom and filter,
    - then details-on-demand

- **Visual Analytics**
  - ~ 18 years ago
  - Visual Analytics mantra (Keim, 2005):
    - Analyze First –
    - Show the Important - Zoom,
    - Filter and
    - Analyze Further - Details on Demand

Visual analytics is an integrated approach combining visualization, human factors and data analysis (Keim, 2005)

Visual Analytics System

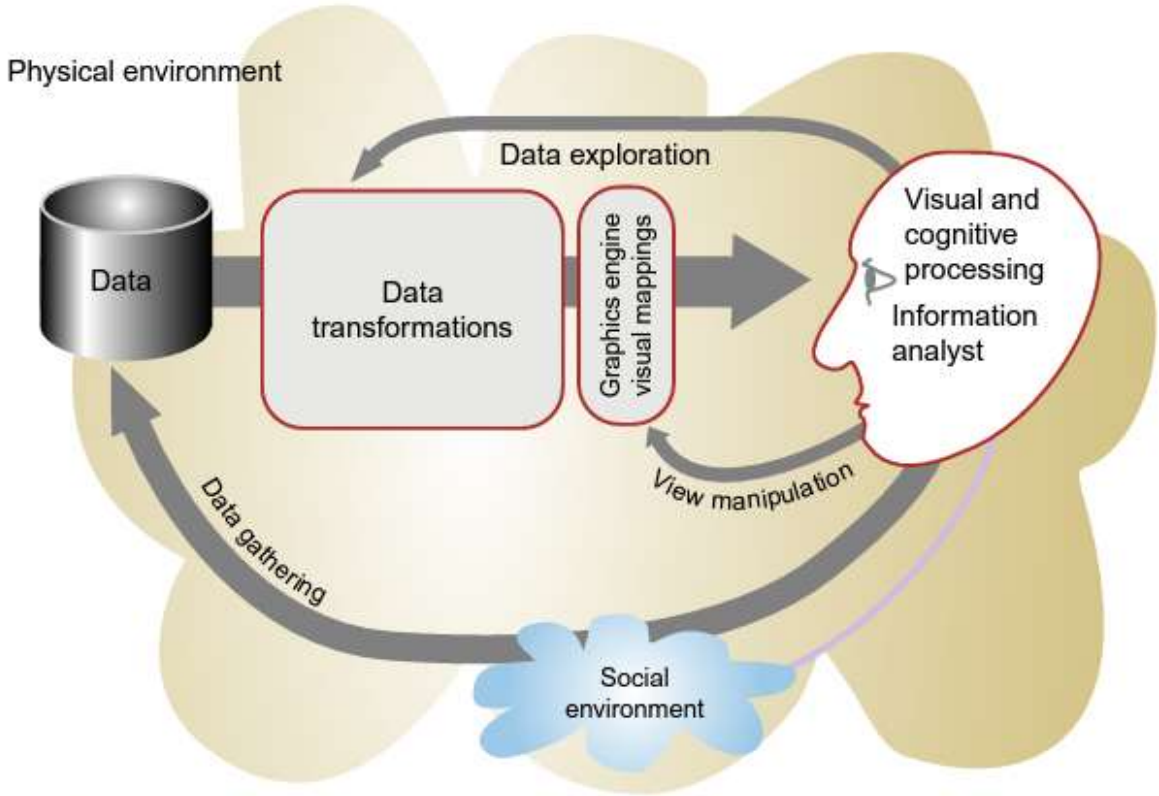


Figure 1.2 The visualization process.

Ware, C. (2013). Information visualization: perception for design (Third ed.): Elsevier.

Visual Search: Symbols: Preattentive Processing: Examples

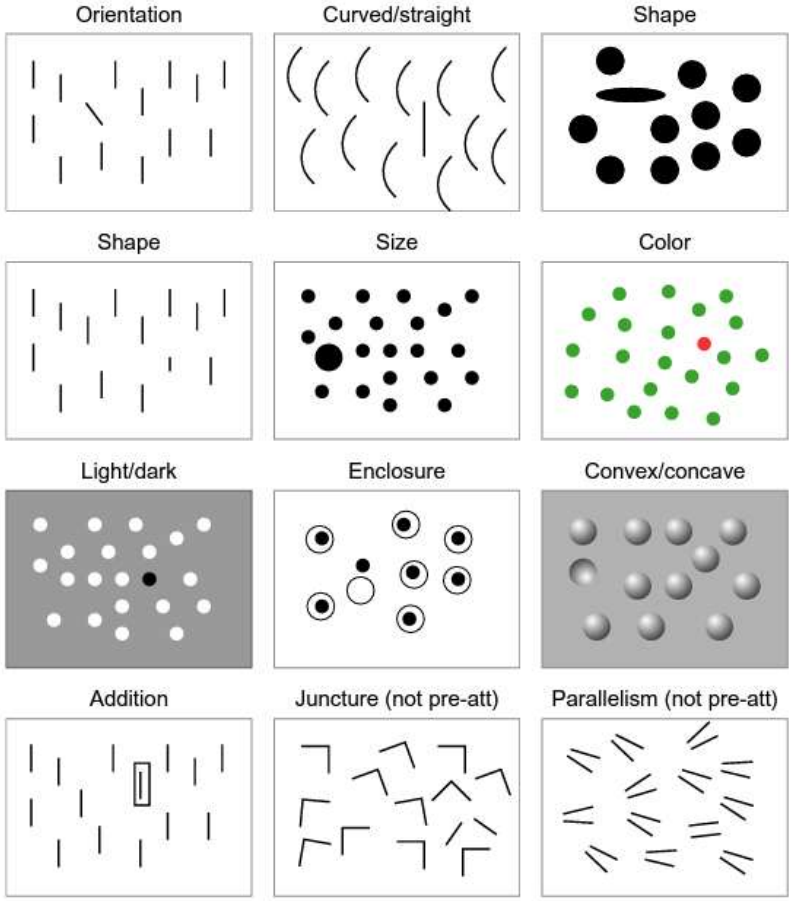


Figure 5.11 Most of the preattentive examples given here can be accounted for by the processing characteristics of neurons in the primary visual cortex.

Ware, C. (2013). Information visualization: perception for design (Third ed.): Elsevier.

Stroop Effect

RED GREEN YELLOW BLUE BLACK GREEN PURPLE BLUE BLACK
ORANGE GREEN RED GREEN YELLOW BLUE BLACK GREEN
PURPLE BLUE BLACK ORANGE BLACK GREEN RED

GREEN RED BLUE YELLOW PURPLE RED BLACK BLUE BLACK
GREEN ORANGE BLUE RED PURPLE YELLOW RED BLACK
YELLOW GREEN ORANGE BLACK GREEN RED GREEN

Figure 11.6 As quickly as you can, try to name the colors in the set of words at the top, and then try to name the colors in the set of words below. Even though they are asked to ignore the meaning of the words, people are slowed down by the mismatch in the second set. This is referred to as the *Stroop effect*, which shows that some processing is automatic.

Ware, C. (2013). Information visualization: perception for design (Third ed.): Elsevier.

Visual Search: Representing Quantity

[G5.16] When designing a set of glyphs to represent quantity, mapping to any of the following glyph attributes will be effective: size, lightness (on a dark background), darkness (on a light background), vividness (higher saturation) of color, or vertical position in the display.

Figure 5.24 The same information is shown using length, area, and volume. Research shows that the quantities shown in the volume display on the right will be mostly judged according to the relative area of the images, not according to volume, resulting in large errors.

Ware, C. (2013). Information visualization: perception for design (Third ed.): Elsevier.

Example #0: **Time Use Survey: US 2008**

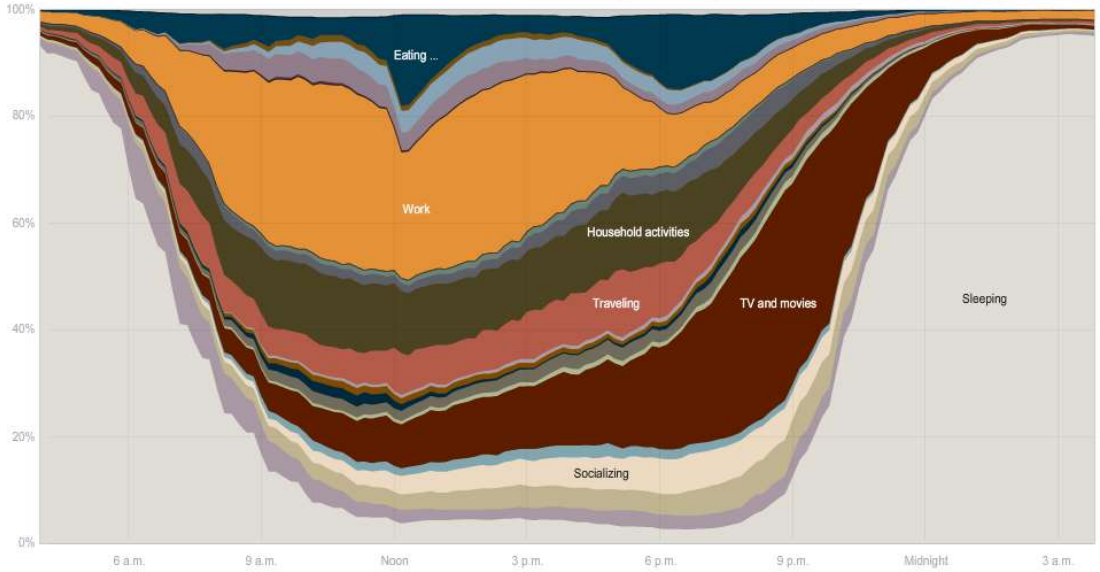
How Different Groups Spend Their Day

The American Time Use Survey asks thousands of American residents to recall every minute of a day. Here is how people over age 15 spent their time in 2008. Related article

Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

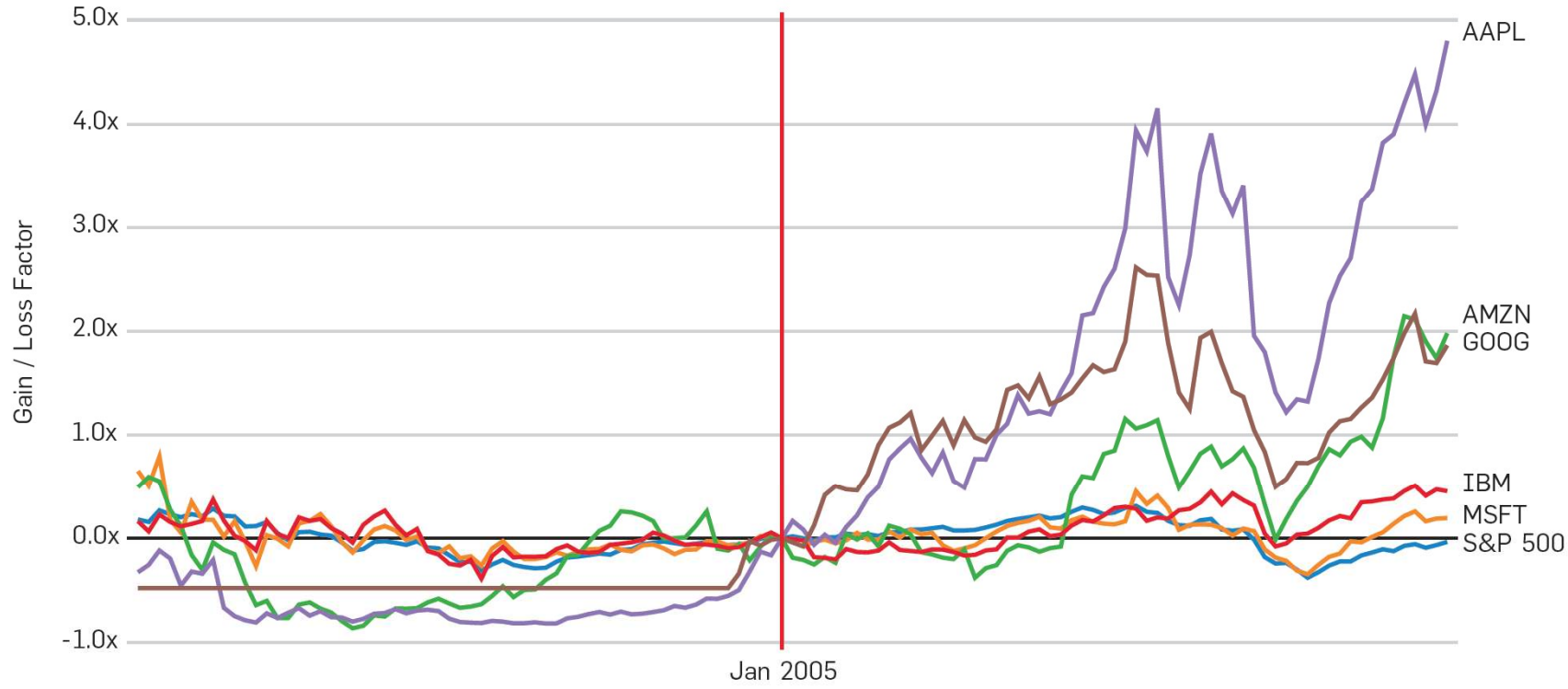
Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html?_r=0

Example #1: **Index Chart**

Time-Series Data: **Figure 1a. Index chart of selected technology stocks, 2000–2010.**

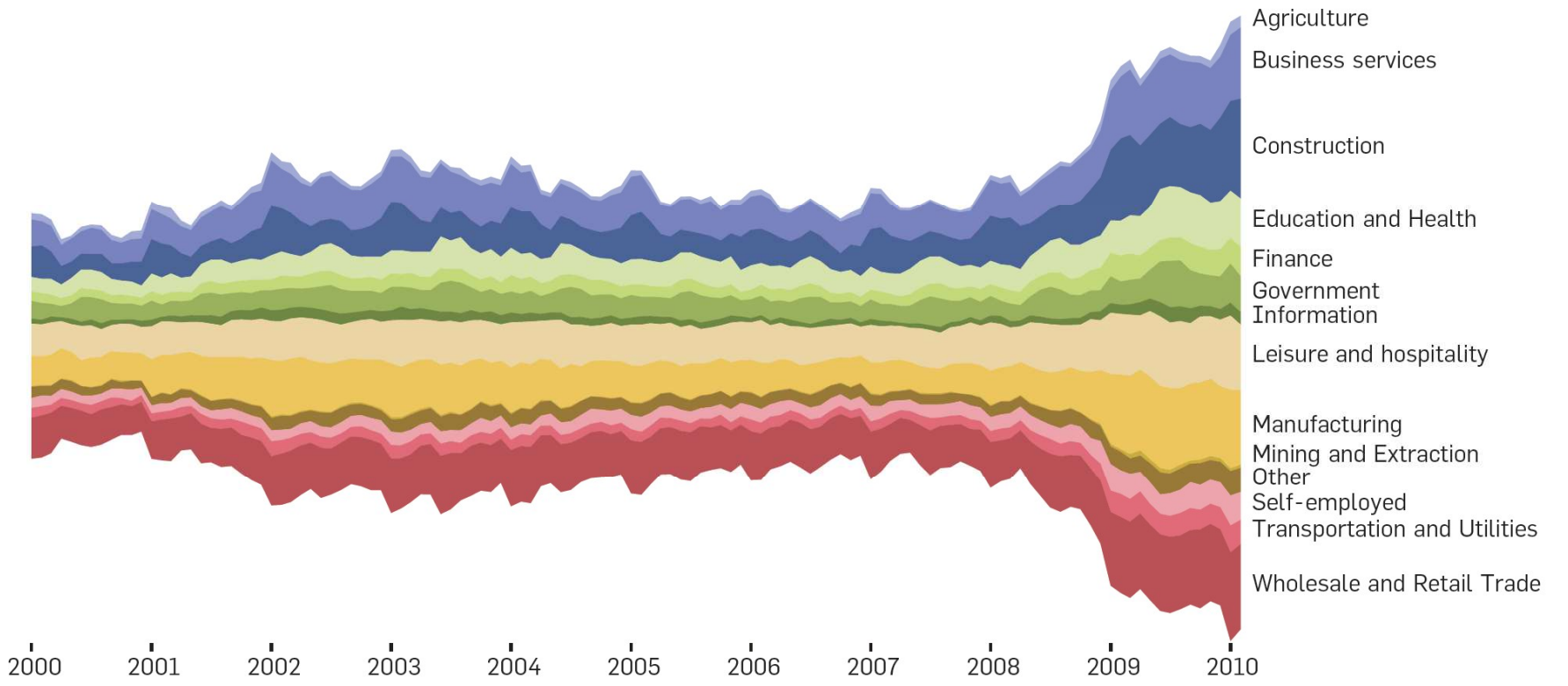


Source: Yahoo! Finance; http://hci.stanford.edu/jheer/files/zoo/ex/time/index-chart.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/time/index-chart.html

Example #2: **Stacked Graph**

Time-Series Data: **Figure 1b. Stacked graph of unemployed U.S. workers by industry, 2000–2010.**

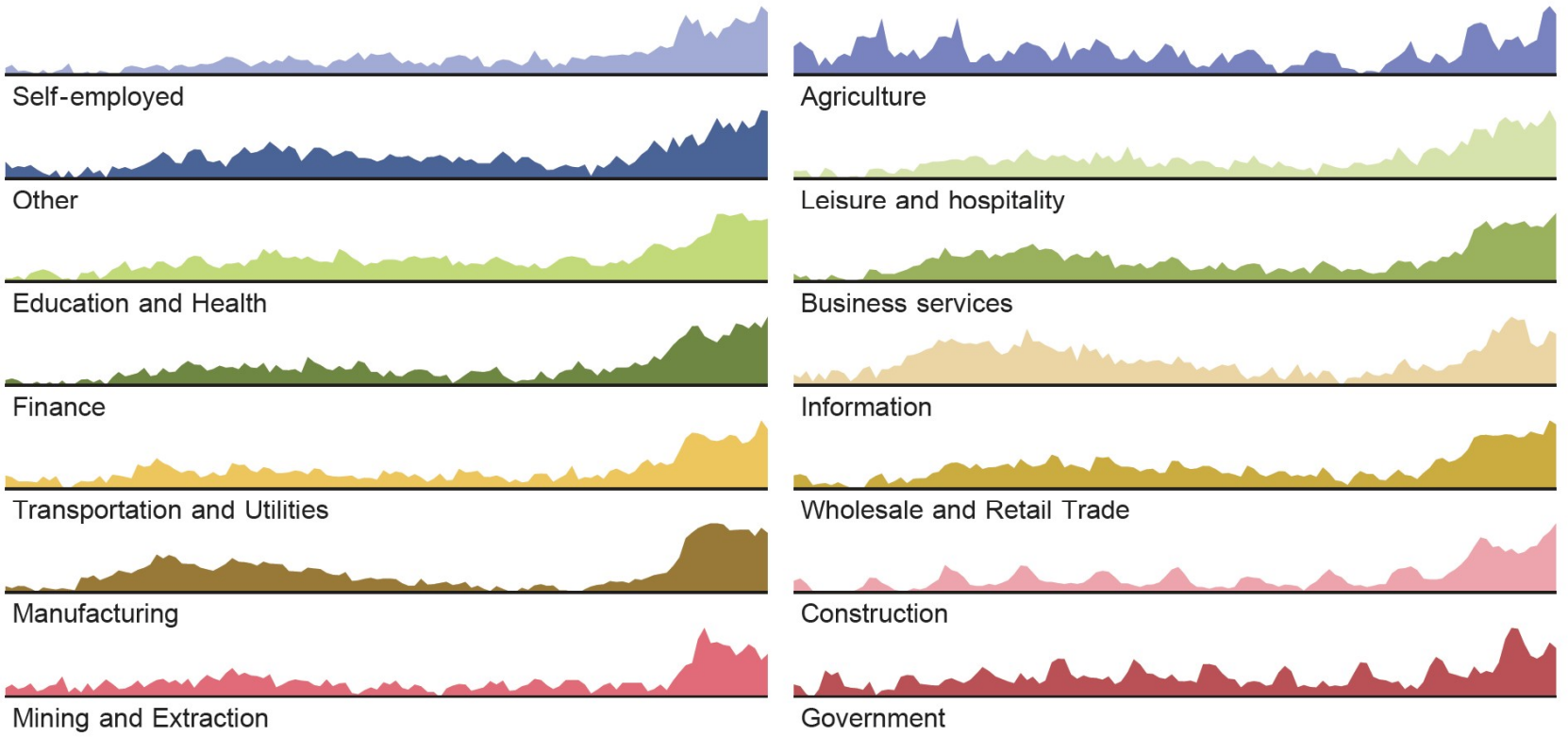


Source: U.S. Bureau of Labor Statistics; http://hci.stanford.edu/jheer/files/zoo/ex/time/stack.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/time/stack.html

Example #3: **Small Multiples**

Time-Series Data: **Figure 1c. Small multiples of unemployed U.S. workers, normalized by industry, 2000–2010.**

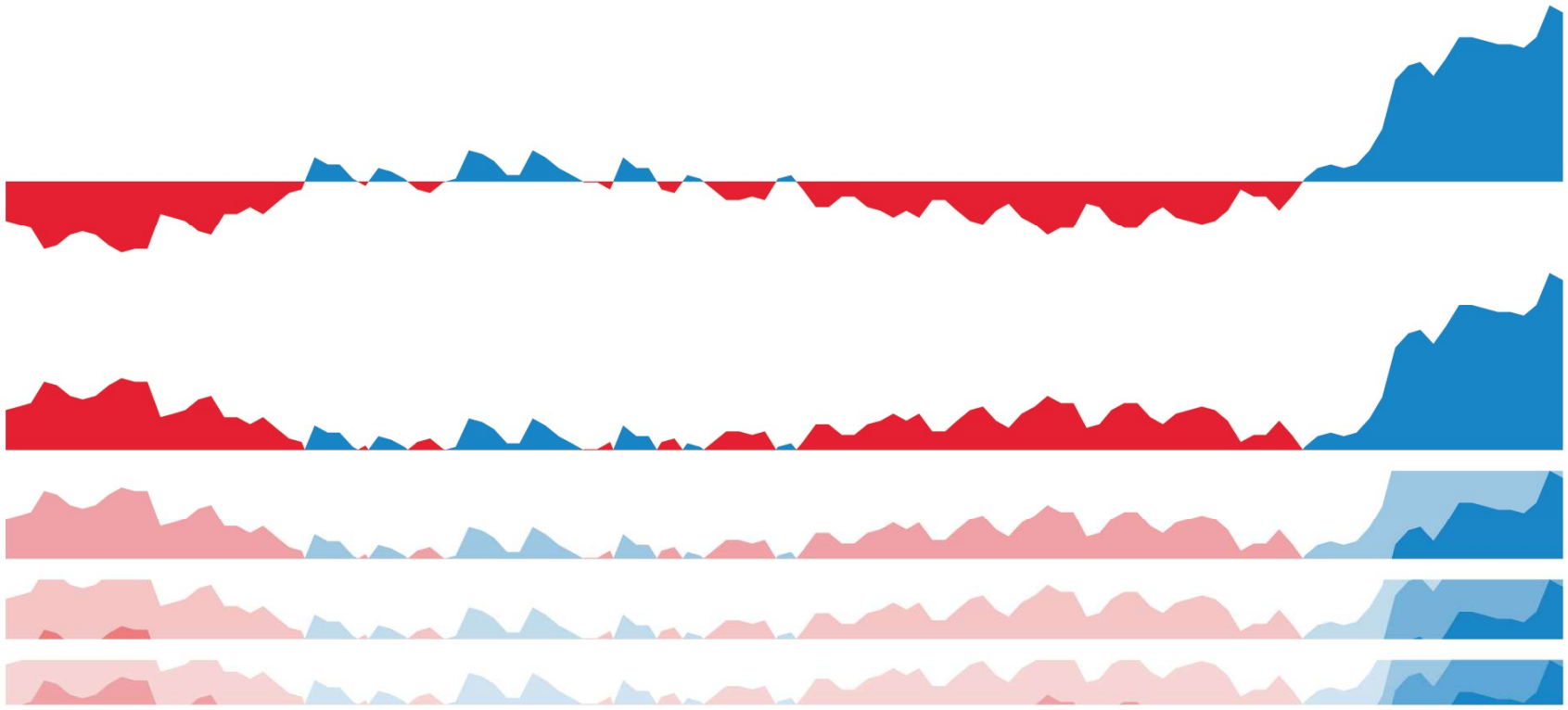


Source: U.S. Bureau of Labor Statistics; http://hci.stanford.edu/jheer/files/zoo/ex/time/multiples.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/time/multiples.html

Example #4: **Horizon Graph**

Time-Series Data: **Figure 1d. Horizon graphs of U.S. unemployment rate, 2000–2010.**



Source: U.S. Bureau of Labor Statistics; http://hci.stanford.edu/jheer/files/zoo/ex/time/horizon.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/time/horizon.html

Example #5: **Stem-and-Leaf Plot**

Statistical Distributions: **Figure 2a. Stem-and-leaf plot of Mechanical Turk participation rates.**

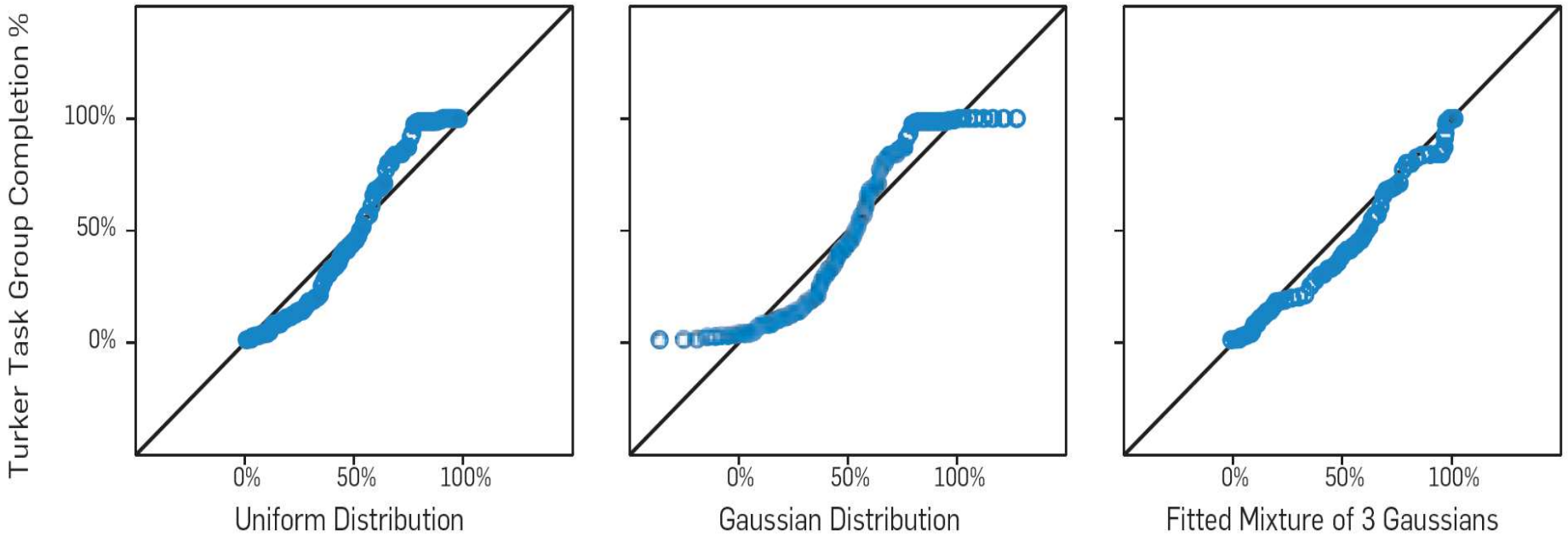
0	1	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	6	7	8	8	8	8	8	8	9
1	0	0	0	0	1	1	1	1	2	2	3	3	3	3	4	4	4	4	5	5	6	7	7	8	9	9	9	9	9			
2	0	0	1	1	1	5	7	8	9																							
3	0	0	1	2	3	3	3	4	6	6	8	8																				
4	0	0	1	1	1	1	3	3	4	5	5	5	6	7	8	9																
5	0	2	3	5	6	7	7	7	9																							
6	1	2	6	7	8	9	9	9																								
7	0	0	0	1	6	7	9																									
8	0	0	1	2	3	4	4	4	4	4	4	4	5	6	7	7	7	9														
9	1	3	3	5	7	8	8	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9					
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0													

Source: Stanford Visualization Group; http://hci.stanford.edu/jheer/files/zoo/ex/stats/stem-and-leaf.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/stats/stem-and-leaf.html

Example #6: **Q-Q Plot**

Statistical Distributions: Figure 2b. Q-Q plots of Mechanical Turk participation rates.

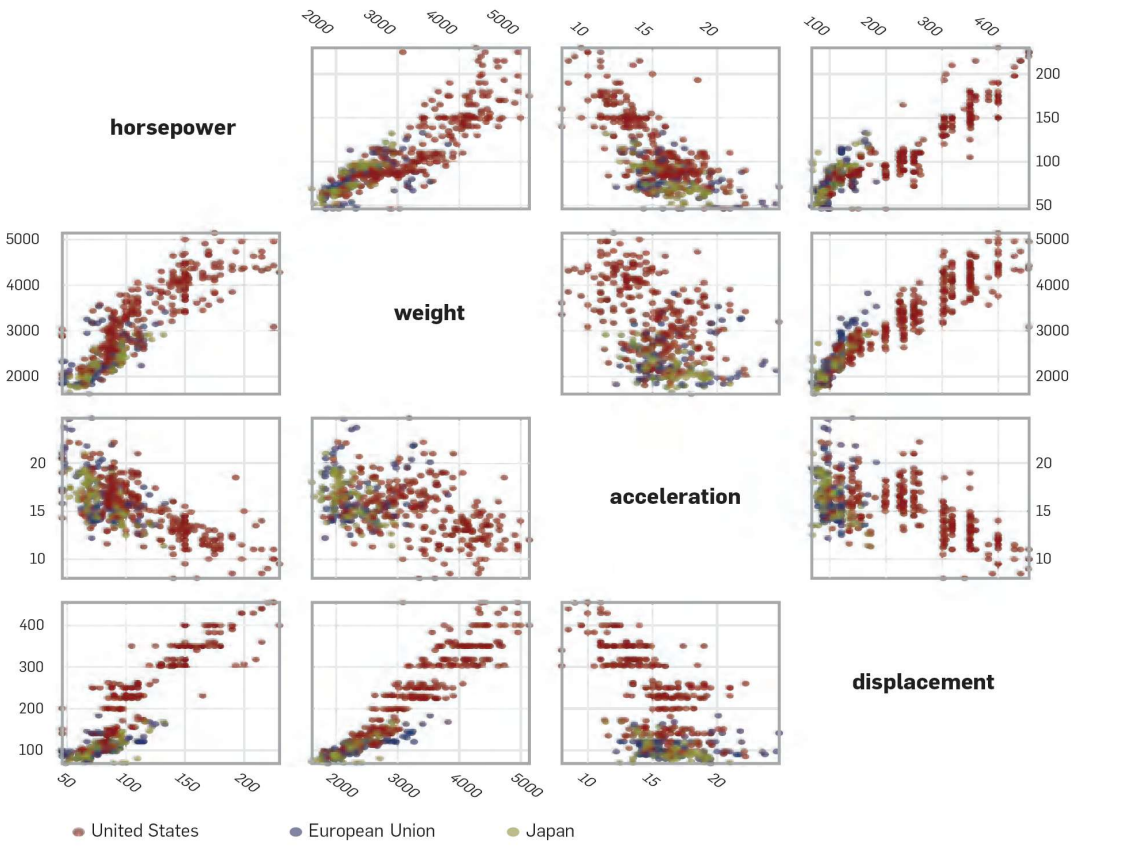


Source: Stanford Visualization Group; http://hci.stanford.edu/jheer/files/zoo/ex/stats/qqplot.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/stats/qqplot.html

Example #7: **Scatter Plot**

Statistical Distributions: **Figure 2c. Scatter plot matrix of automobile data.**

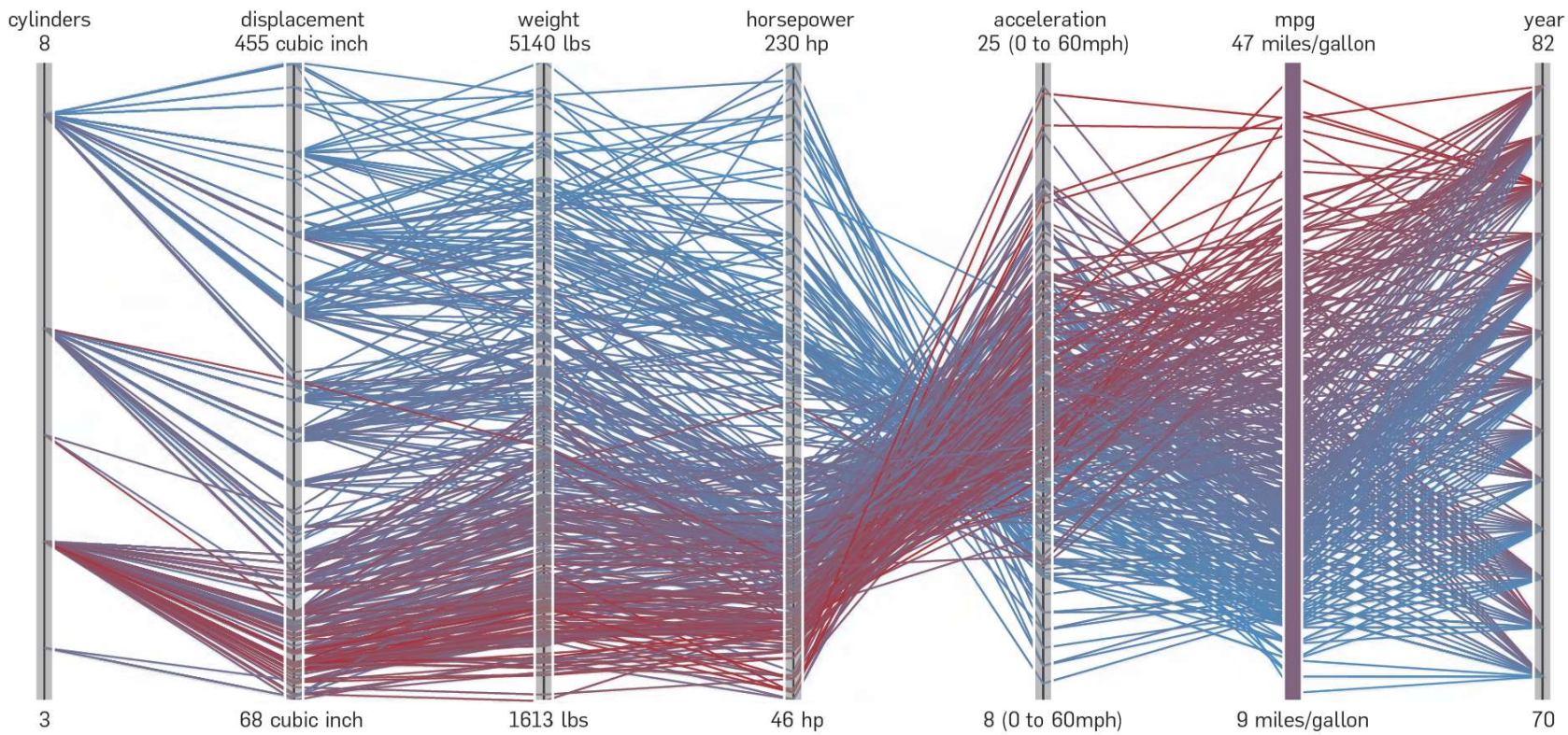


Source: GGobi; http://hci.stanford.edu/jheer/files/zoo/ex/stats/splom.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/stats/splom.html

Example #8: **Parallel Coordinates**

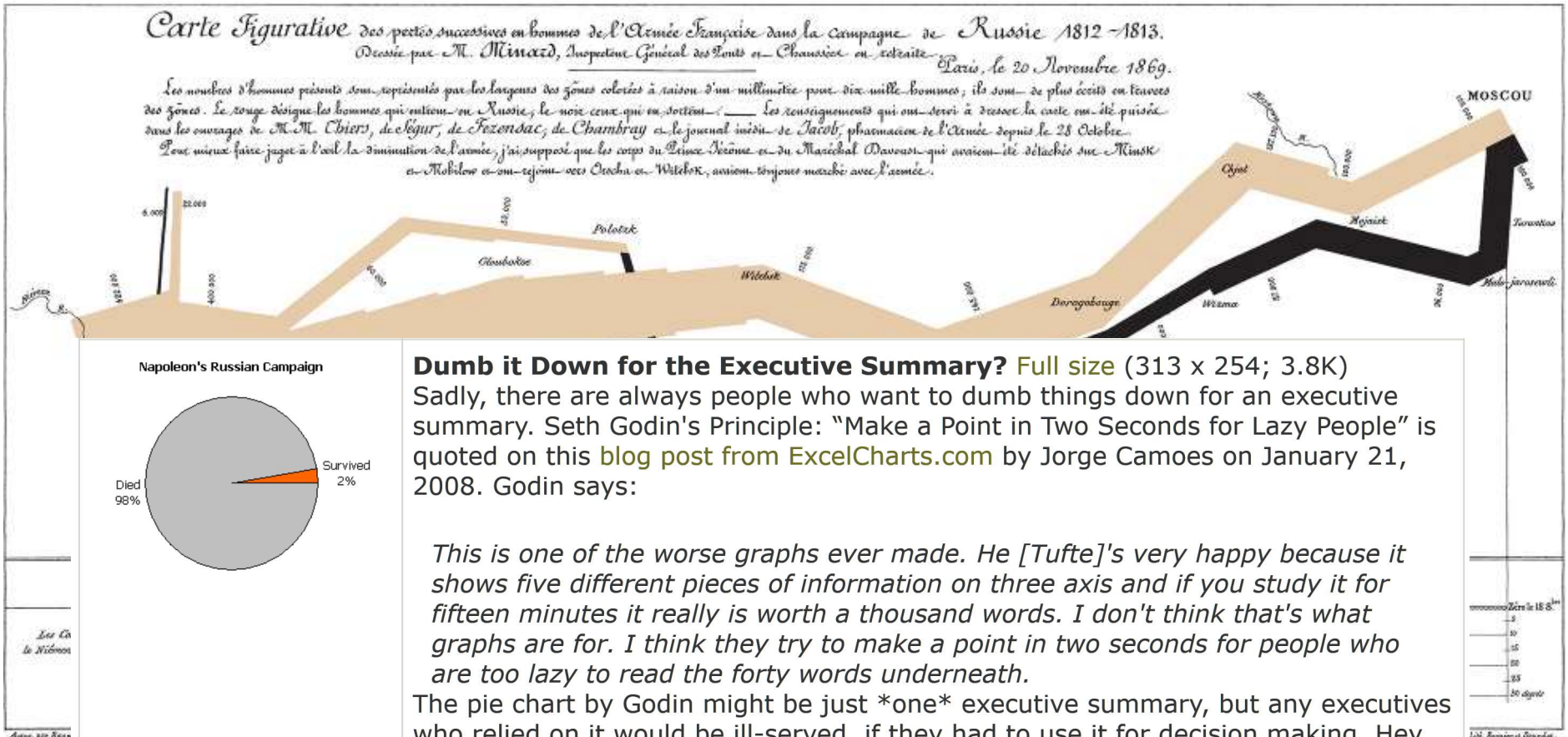
Statistical Distributions: Figure 2d. Parallel coordinates of automobile data.



Source: GGobi; http://hci.stanford.edu/jheer/files/zoo/ex/stats/parallel.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/stats/parallel.html

Example #09: **Flow Map**



Dumb it Down for the Executive Summary? Full size (313 x 254; 3.8K)
Sadly, there are always people who want to dumb things down for an executive summary. Seth Godin's Principle: "Make a Point in Two Seconds for Lazy People" is quoted on this blog post from ExcelCharts.com by Jorge Camoes on January 21, 2008. Godin says:

> *This is one of the worse graphs ever made. He [Tufte]'s very happy because it shows five different pieces of information on three axis and if you study it for fifteen minutes it really is worth a thousand words. I don't think that's what graphs are for. I think they try to make a point in two seconds for people who are too lazy to read the forty words underneath.*

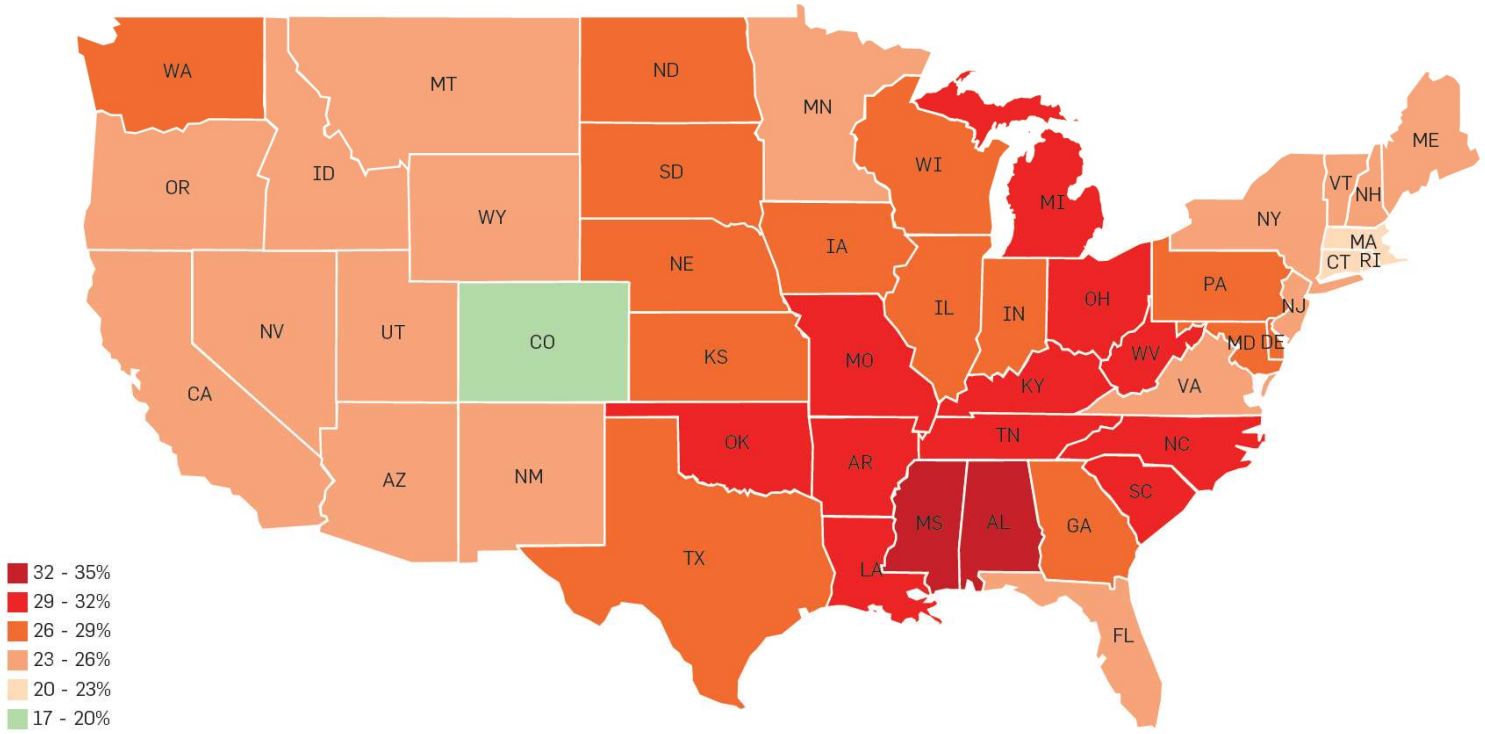
The pie chart by Godin might be just *one* executive summary, but any executives who relied on it would be ill-served, if they had to use it for decision making. Hey -- Napoleon, you shouldn't have done that! Duh? Why not? what went wrong? what could you have done differently?

Minard
The wi
ning.
are

http://www.datavis.ca/gallery/re-minard.php

Example #10: **Choropleth Map**

Maps: Figure 3b. Choropleth map of obesity in the U.S., 2008.

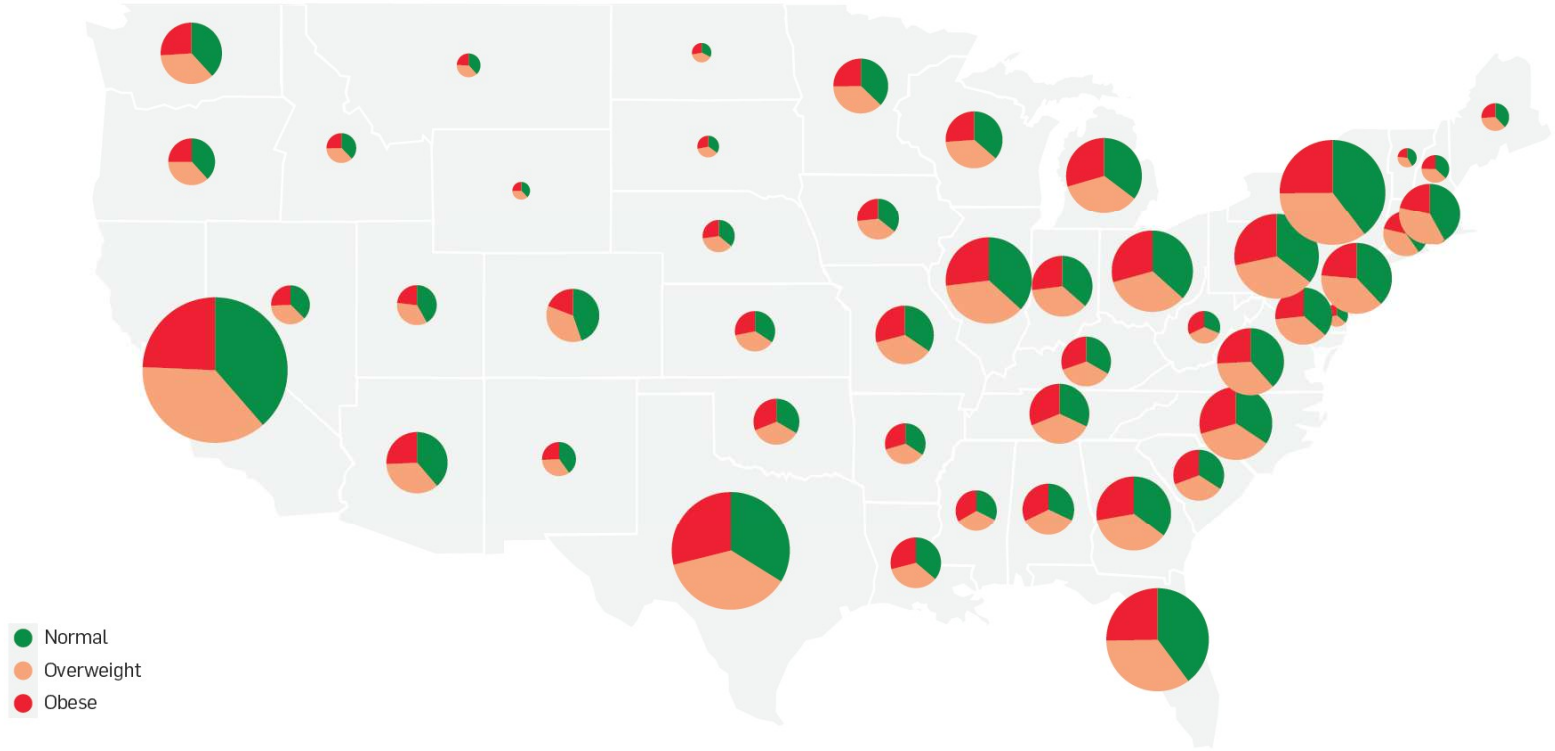


Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/choropleth.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/maps/choropleth.html

Example #11: **Graduated Symbol Map**

Maps: Figure 3c. Graduated symbol map of obesity in the U.S., 2008.

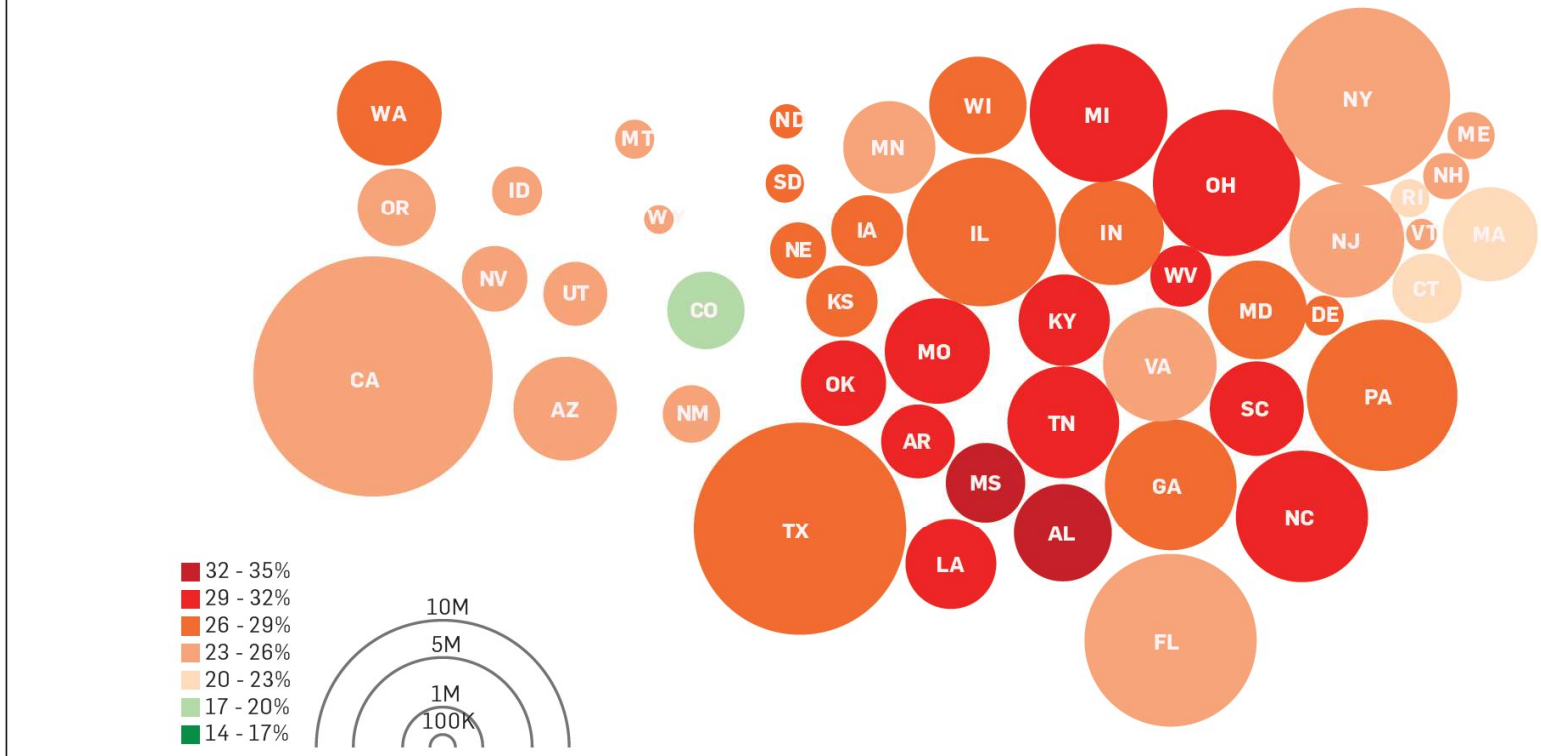


Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/symbol.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/maps/symbol.html

Example #12: **Dorling Cartogram**

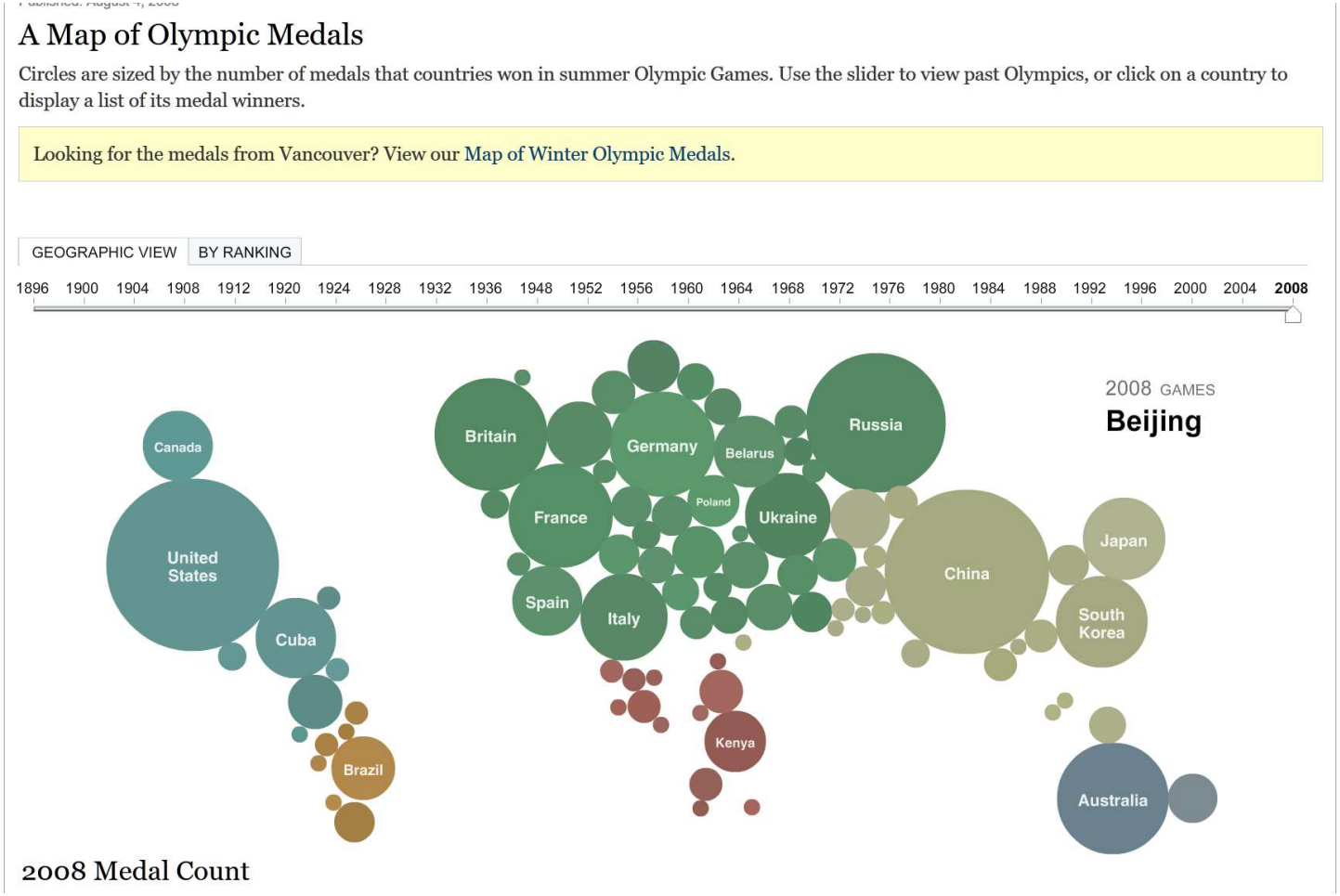
Maps: **Figure 3d. Dorling cartogram of obesity in the U.S., 2008.**



Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/cartogram.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/maps/cartogram.html

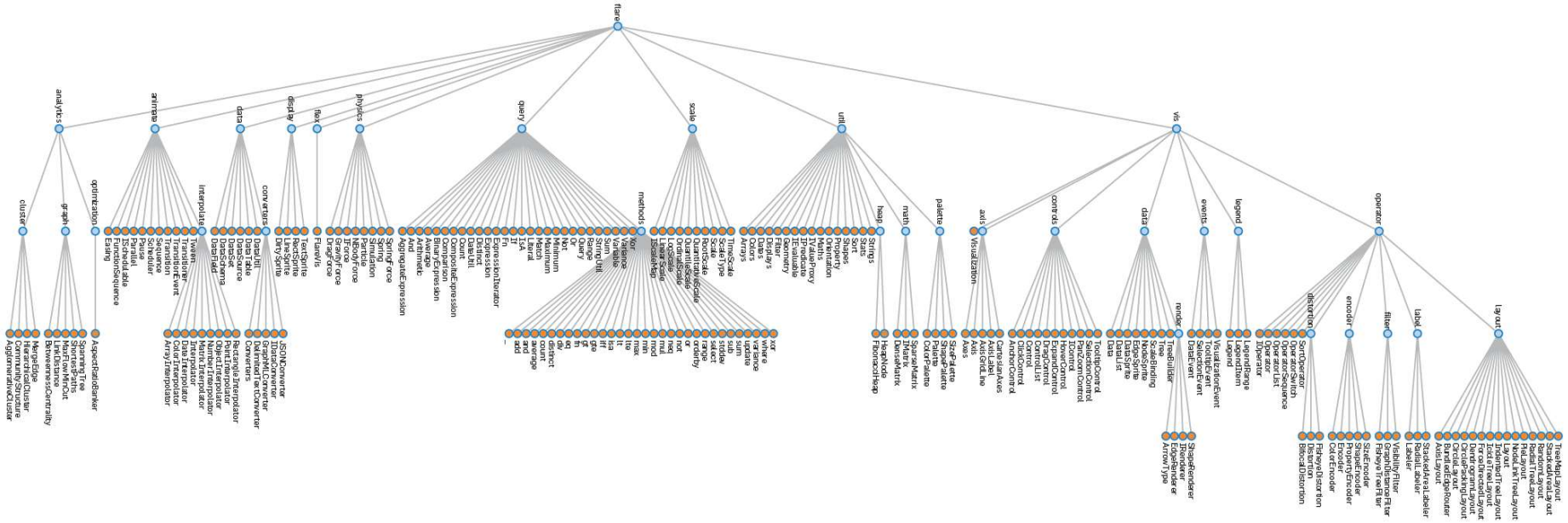
Example #12b: **Dorling Cartogram: Summer Olympics Medals Tally**



https://archive.nytimes.com/www.nytimes.com/interactive/2008/08/04/sports/olympics/20080804_MEDALCOUNT_MAP.html?_r=

Example #13: **Radial Tree Map**

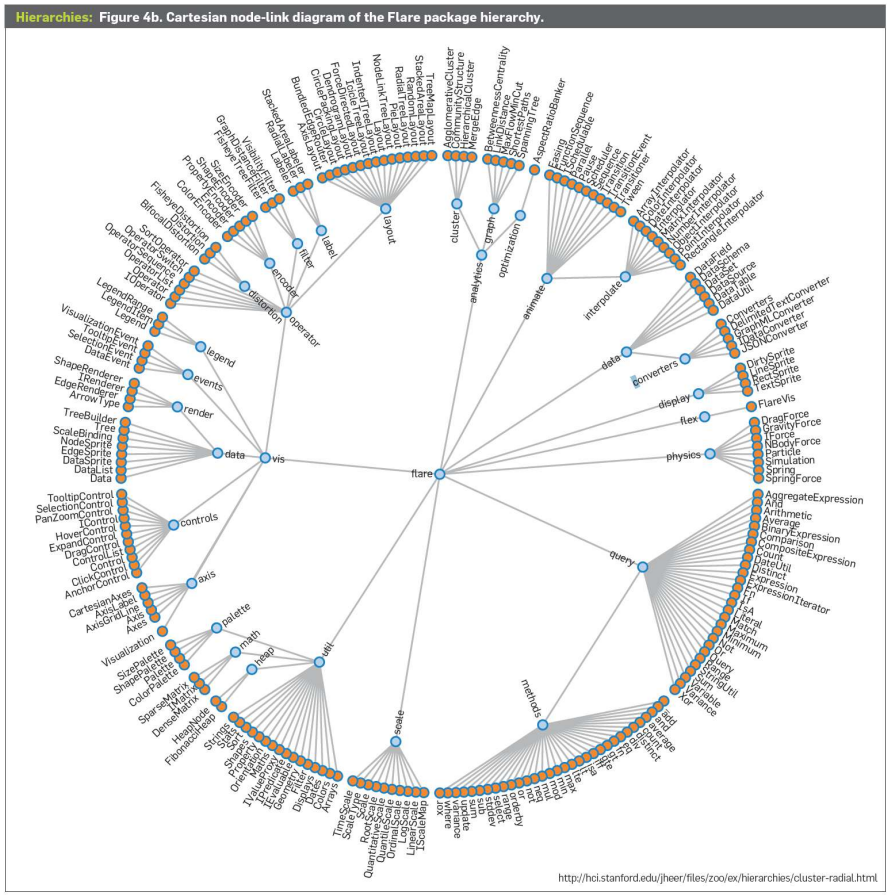
Hierarchies: **Figure 4a. Radial node-link diagram of the Flare package hierarchy.**



http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/tree.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/tree.html

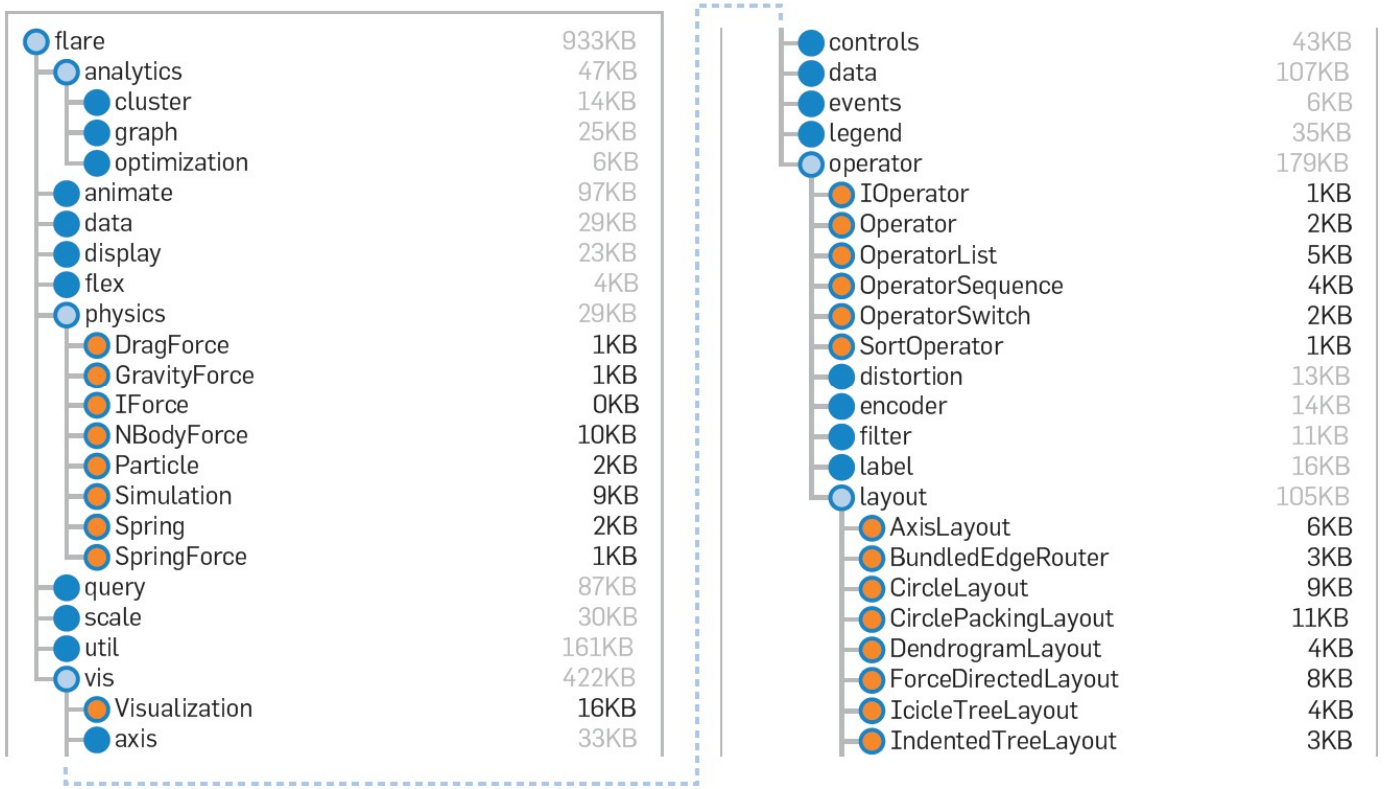
Example #14: **Cartesian Tree Map**



http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/cluster-radial.html

Example #15: **Indented Tree Map**

Hierarchies: **Figure 4c. Indented tree layout of the Flare package hierarchy.**

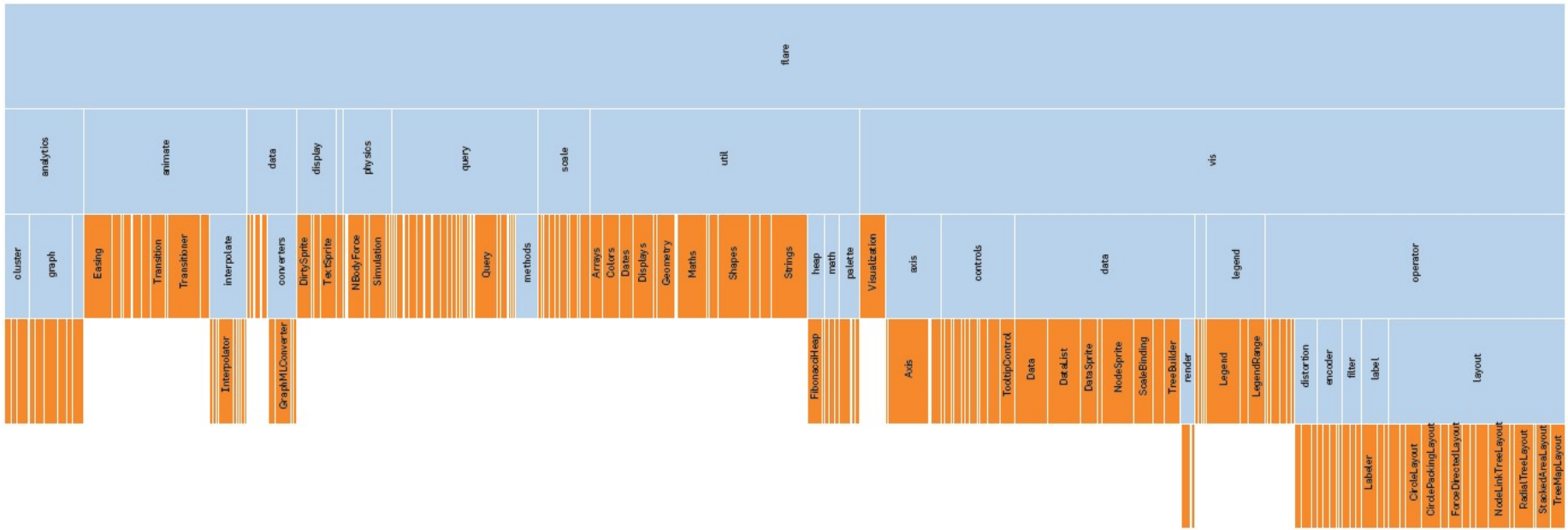


http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/indent.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/indent.html

Example #16: **Icicle Tree Map**

Hierarchies: **Figure 4d. Icicle tree layout of the Flare package hierarchy.**

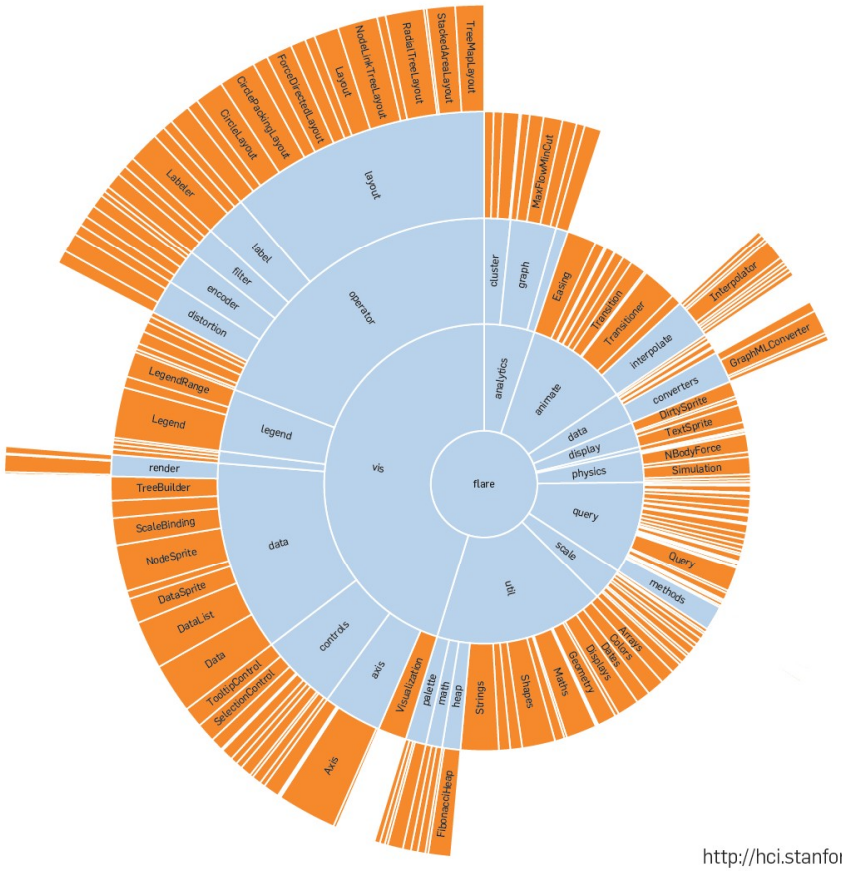


http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/icicle.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/icicle.html

Example #17: **Sunburst Tree Map**

Hierarchies: **Figure 4e. Sunburst (radial space-filling) layout of the Flare package hierarchy.**

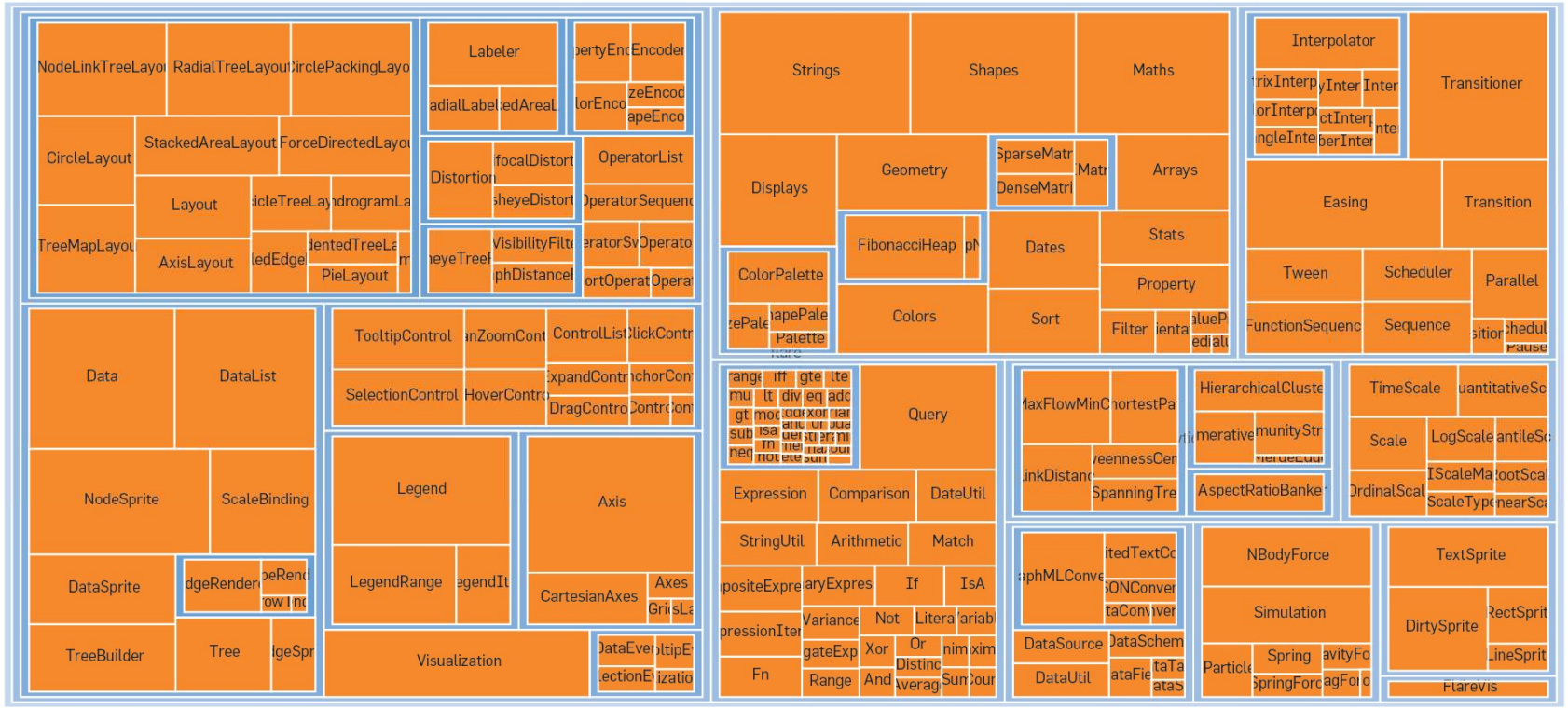


http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/sunburst.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/sunburst.html

Example #18: **Tree Map**

Hierarchies: **Figure 4f. Treemap layout of the Flare package hierarchy.**

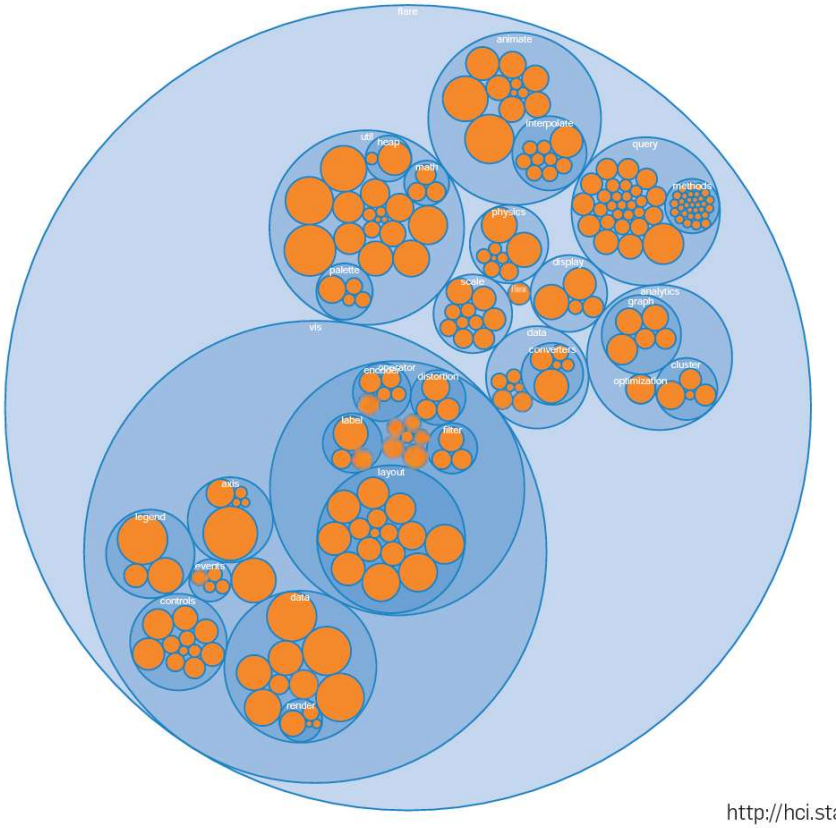


http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/treemap.html

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/sunburst.html

Example #19: **Circle Packing Tree Map**

Hierarchies: Figure 4g. Nested circles layout of the Flare package hierarchy.

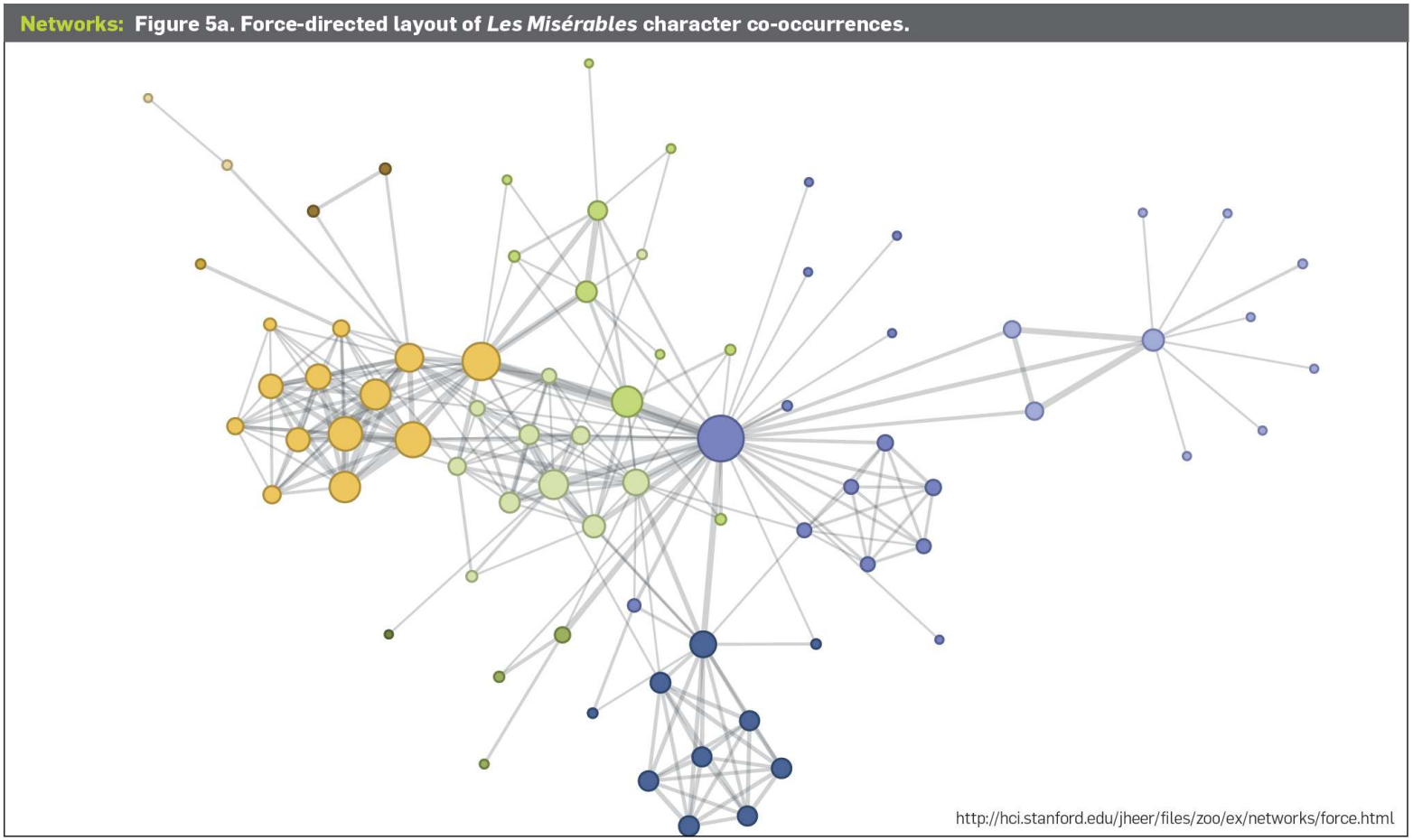


http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/pack.html
Source: The Flare Toolkit http://flare.prefuse.org

http://homes.cs.washington.edu/~jheer//files/zoo/ex/hierarchies/pack.html

Example #20: **Force Directed Graphs**

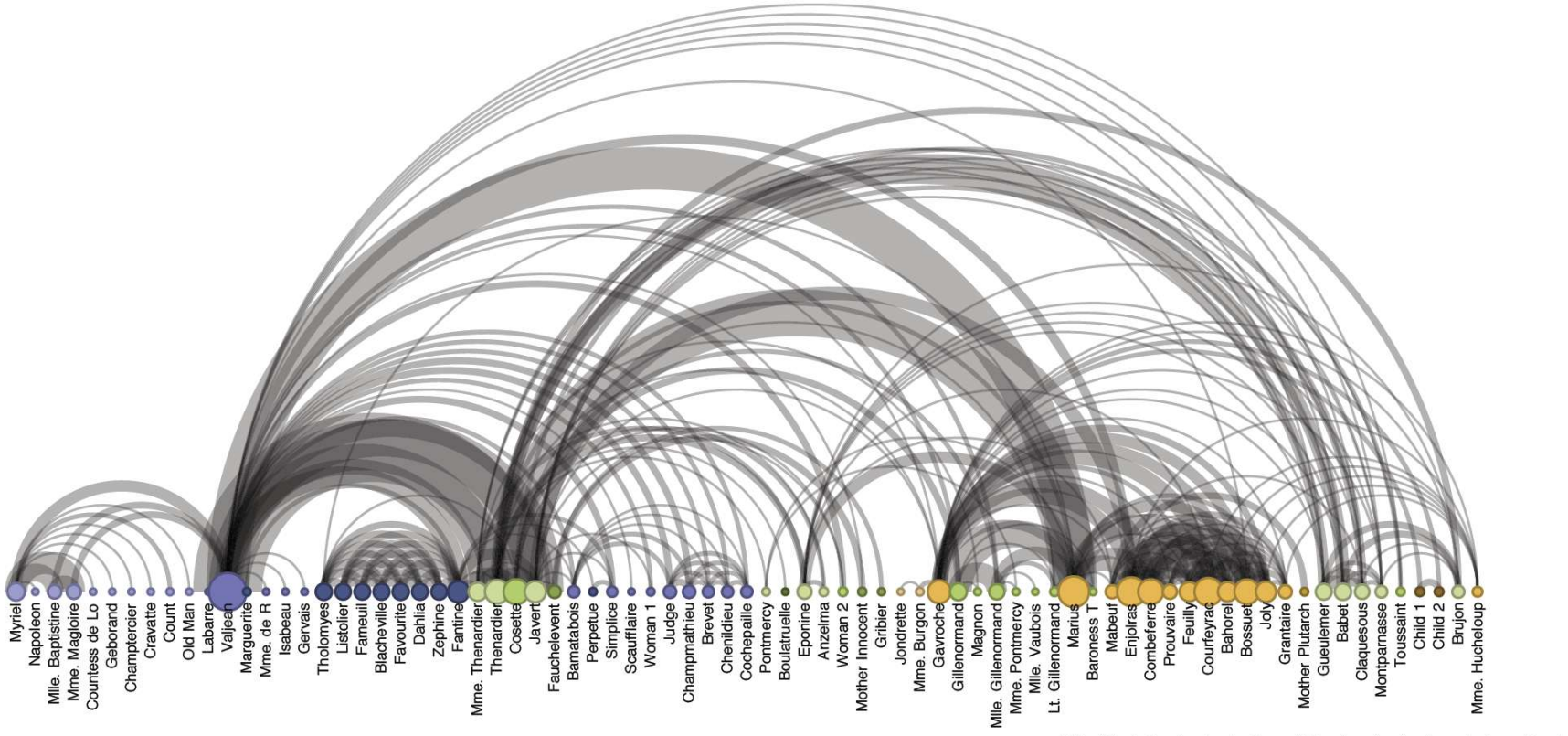
Networks: **Figure 5a. Force-directed layout of *Les Misérables* character co-occurrences.**



http://homes.cs.washington.edu/~jheer//files/zoo/ex/networks/force.html

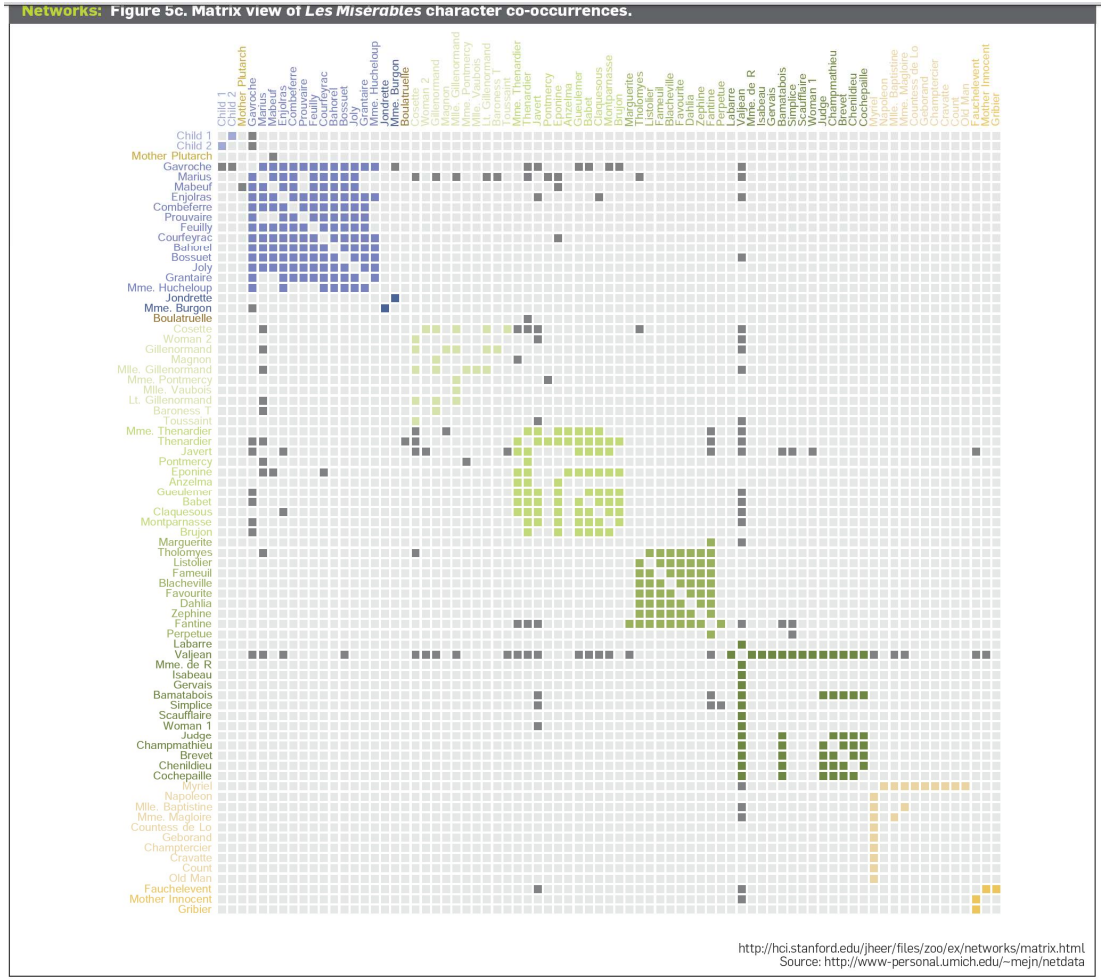
Example #21: **Arc Diagram**

Networks: **Figure 5b. Arc diagram of *Les Misérables* character co-occurrences.**



http://homes.cs.washington.edu/~jheer//files/zoo/ex/networks/arc.html

Example #22: **Matrix Diagram**



http://homes.cs.washington.edu/~jheer//files/zoo/ex/networks/matrix.html

Reflections

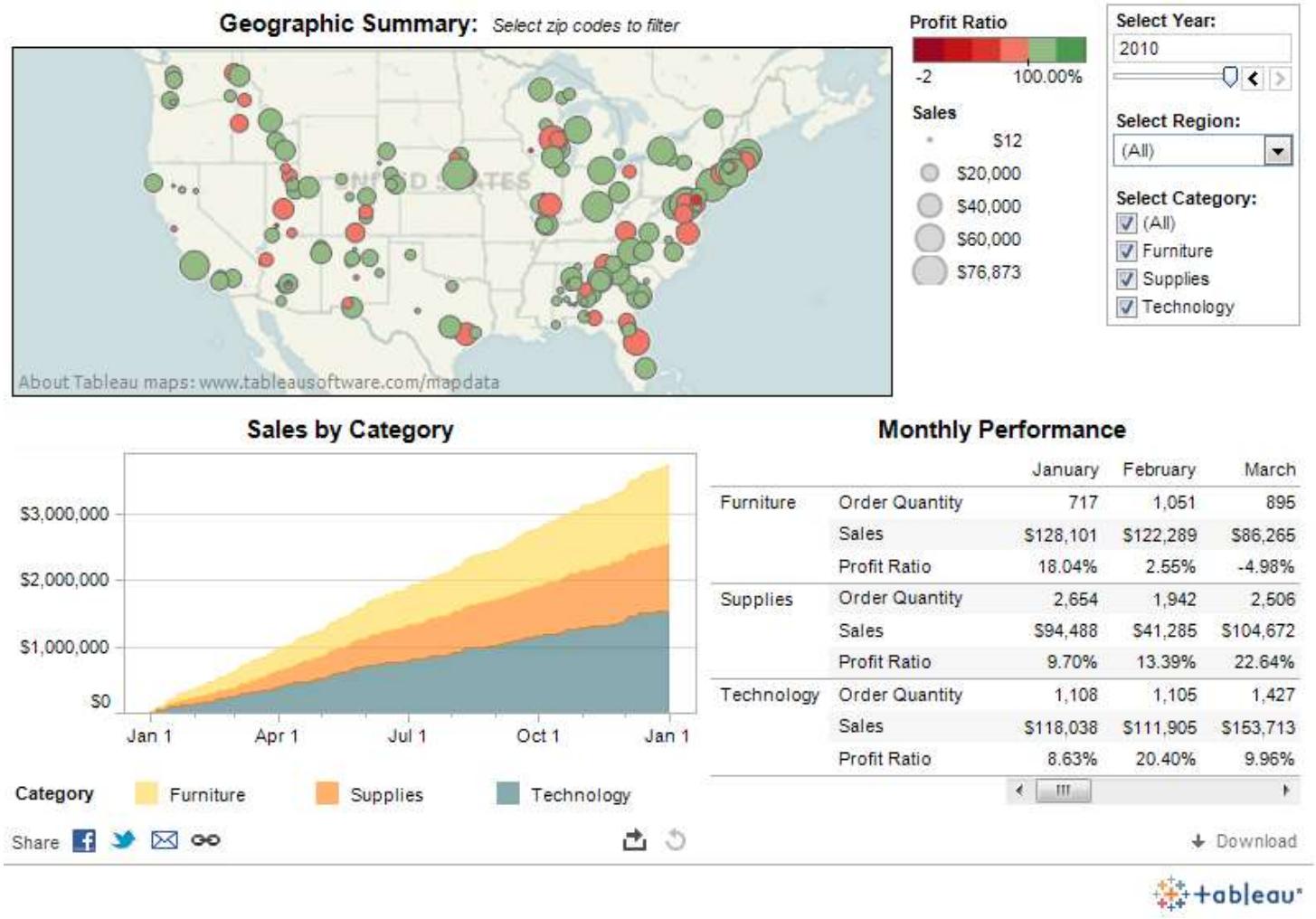
Dashboards

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance." (Few, 2004)

Dashboard characteristics:

1. Small, concise, clear and intiutive dispaly mechanisms
2. Customized

Executive Dashboard



		January	February	March
Furniture	Order Quantity	717	1,051	895
	Sales	$128,101	$122,289	$86,265
	Profit Ratio	18.04%	2.55%	-4.98%
Supplies	Order Quantity	2,654	1,942	2,506
	Sales	$94,488	$41,285	$104,672
	Profit Ratio	9.70%	13.39%	22.64%
Technology	Order Quantity	1,108	1,105	1,427
	Sales	$118,038	$111,905	$153,713
	Profit Ratio	8.63%	20.40%	9.96%

Is this a **Good** Dashboard?

Business Development and Dashboards: Three Aspects

1. Audience of the dashboard

2. Value added by the dashboard

3. Type of dashboard created

Dashboards: Audience

	Questions	Implication
Role	What decisions do they make? What questions do they need answered?	Structure the information to make it super easy to answer high priority questions.
Work flow	In what context will they be reviewing the dashboard? What information are they using on a daily basis? How much time do they have to review the numbers?	The form and information display needs to fit into an existing work flow. For example, an on-the-road sales person may need information delivered to her BlackBerry, not designed for an online wide-screen monitor.
Data comfort and skills	How sophisticated are they with using data? Are they proficient in Excel? Do they enjoy digging into the numbers?	The dashboard's level of detail and analytical capabilities should match the audiences' comfort zone.
Business and data expertise	How familiar are they with the key performance metrics? Do they understand where the data comes from? Are they familiar with internal company or industry terminology?	This determines the need for embedded explanations and use of natural language.

(Juice, 2009)

Dashboards: Value Created

- ☐ Help management define what is important
- ☐ Educate people in the organization about the things that matter
- ☐ Set goals and expectations for specific individuals or groups
- ☐ Help executives sleep at night because they know what's going on
- ☐ Encourage specific actions in a timely manner
- ☐ Highlight exceptions and provide alerts when problems occur
- ☐ Communicate progress and success
- ☐ Provide a common interface for interacting with and analyzing important business data

(Juice, Inc, 2009)

Dashboards: Categorization

Categorized by:

Variable	Values
Role	Strategic Analytical Operational
Type of data	Quantitative Non-quantitative
Data domain	Sales Finance Marketing Manufacturing Human Resources
Type of measures	Balanced Scorecard (for example, KPIs) Six Sigma Non-performance
Span of data	Enterprise-wide Departmental Individual
Update frequency	Monthly Weekly Daily Hourly Real time or near real time
Interactivity	Static display Interactive display (drill-down, filters, etc.)
Mechanisms of display	Primarily graphical Primarily text Integration of graphics and text
Portal functionality	Conduit to additional data No portal functionality

(Few, 2006)

Data Diversity:

Category	Measures
Sales	Bookings Billings Sales pipeline (anticipated sales) Number of orders Order amounts Selling prices
Marketing	Market share Campaign success Customer demographics
Finance	Revenues Expenses Profits
Technical Support	Number of support calls Resolved cases Customer satisfaction Call durations
Fulfillment	Number of days to ship Backlog Inventory levels
Manufacturing	Number of units manufactured Manufacturing times Number of defects
Human Resources	Employee satisfaction Employee turnover Count of open positions Count of late performance reviews
Information Technology	Network downtime System usage Fixed application bugs
Web Services	Number of visitors Number of page hits Visit durations

Dashboards: Types

Scope	☐ **Broad**: Displaying information about the entire organization		☐ **Specific**: Focusing on a specific function, process, product, etc.	
Business role	☐ **Strategic**: Provides a high-level, broad, and long-term view of performance		☐ **Operational**: Provides a focused, near-term, and tactical view of performance	
Time horizon	☐ **Historical**: Looking backwards to track trends	☐ **Snapshot**: Showing performance at a single point in time	☐ **Real-time**: Monitoring activity as it happens	☐ **Predictive**: Using past performance to predict future performance
Customization	☐ **One-size-fits-all**: Presented as a single view for all users		☐ **Customizable**: Functionality to let users create a view that reflects their needs	
Level of detail	☐ **High**: Presenting only the most critical top-level numbers		☐ **Drill-able**: Providing the ability to drill drill down to detailed numbers to gain more context	
Point of view	☐ **Prescriptive**: The dashboard explicitly tells the user what the data means and what to do about it		☐ **Exploratory**: User has latitude to interpret the results as they see fit	

(Juice, Inc, 2009)

Dashboard Design

”The need to **squeeze** a great deal of **information** into **small amount of space**, resulting in a display that is easily and immediatly understandable” (Few, 2006)

Few identified **6 common mistakes** to avoid. A well designed dashboard may lead to insights - ”Aha” moments

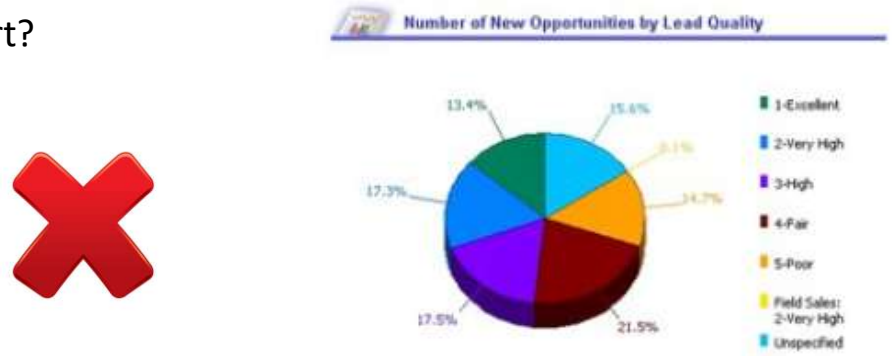
Dashboard Design Mistake #1: **Exceeding the boundaries of a single screen**

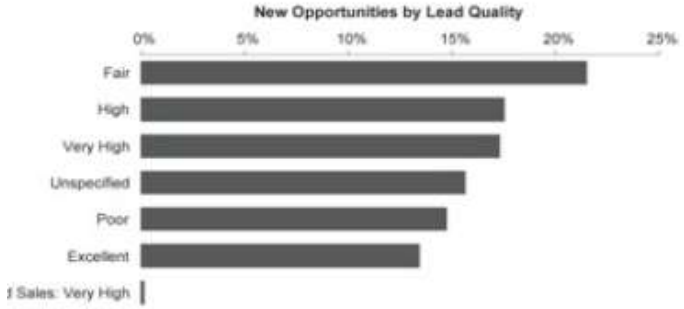
Website of Business Objects.)

Dashboard Design Mistake #2: **Supplying inadequate context for the data**

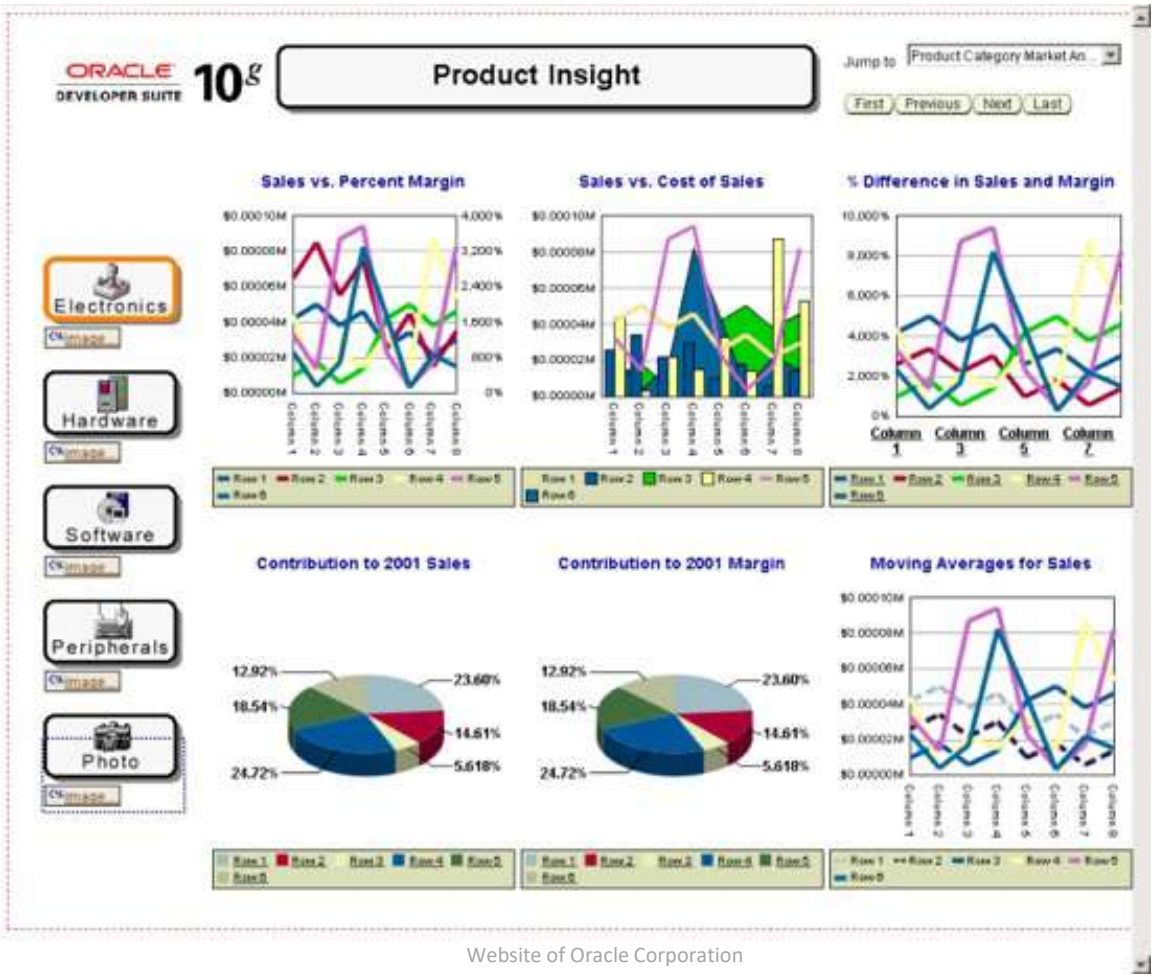
Dashboard Design Mistake #3: **Choosing inappropriate display media**

- Which is the right chart?



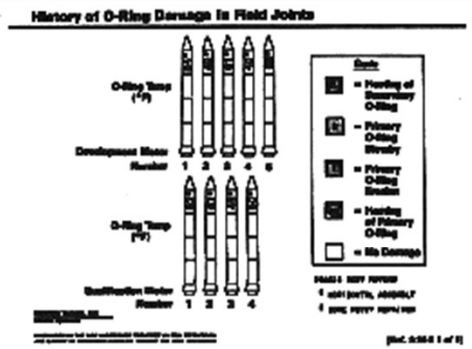


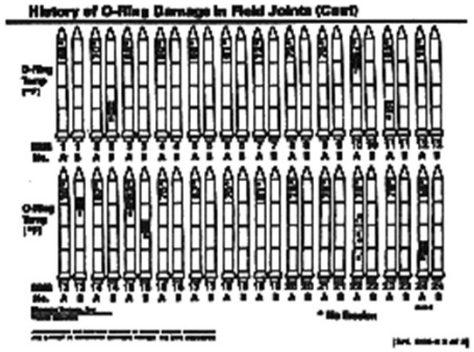
Dashboard Design Mistake #4: **Ineffectively highlighting what's important**

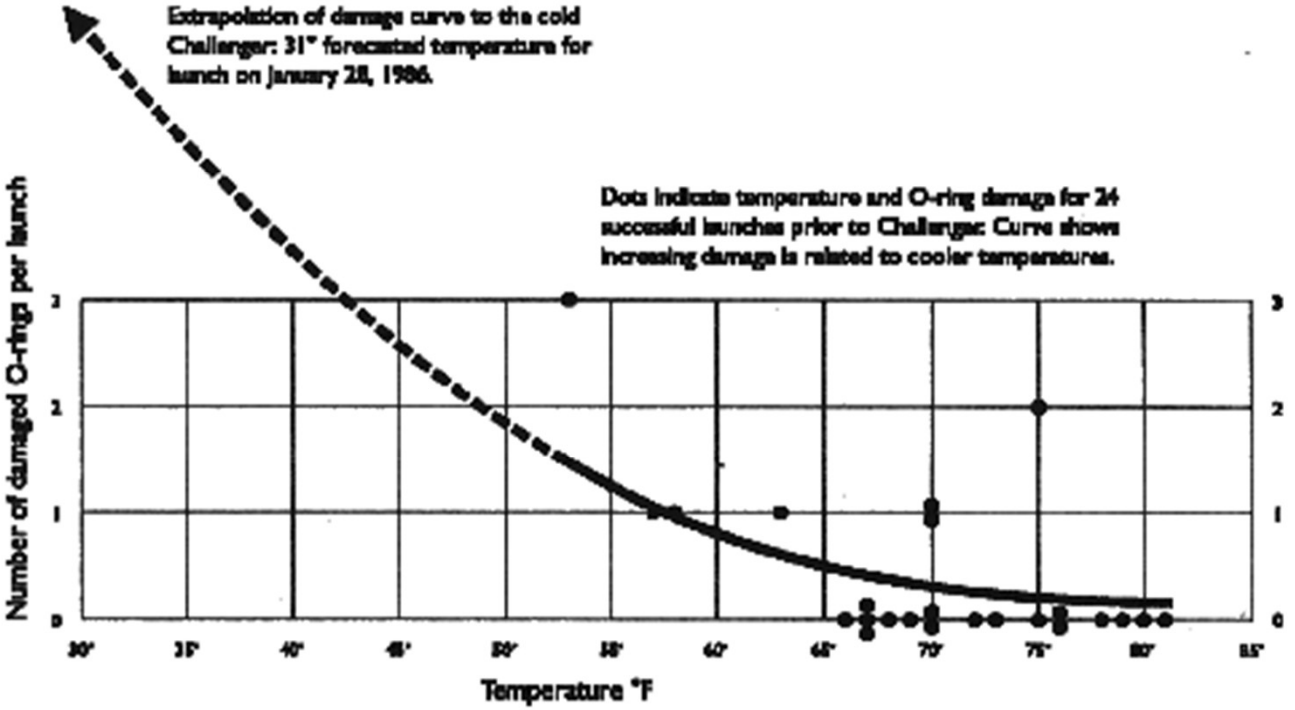


Website of Oracle Corporation

Dashboard Design Mistake #5: **Cluttering it with useless decoration**





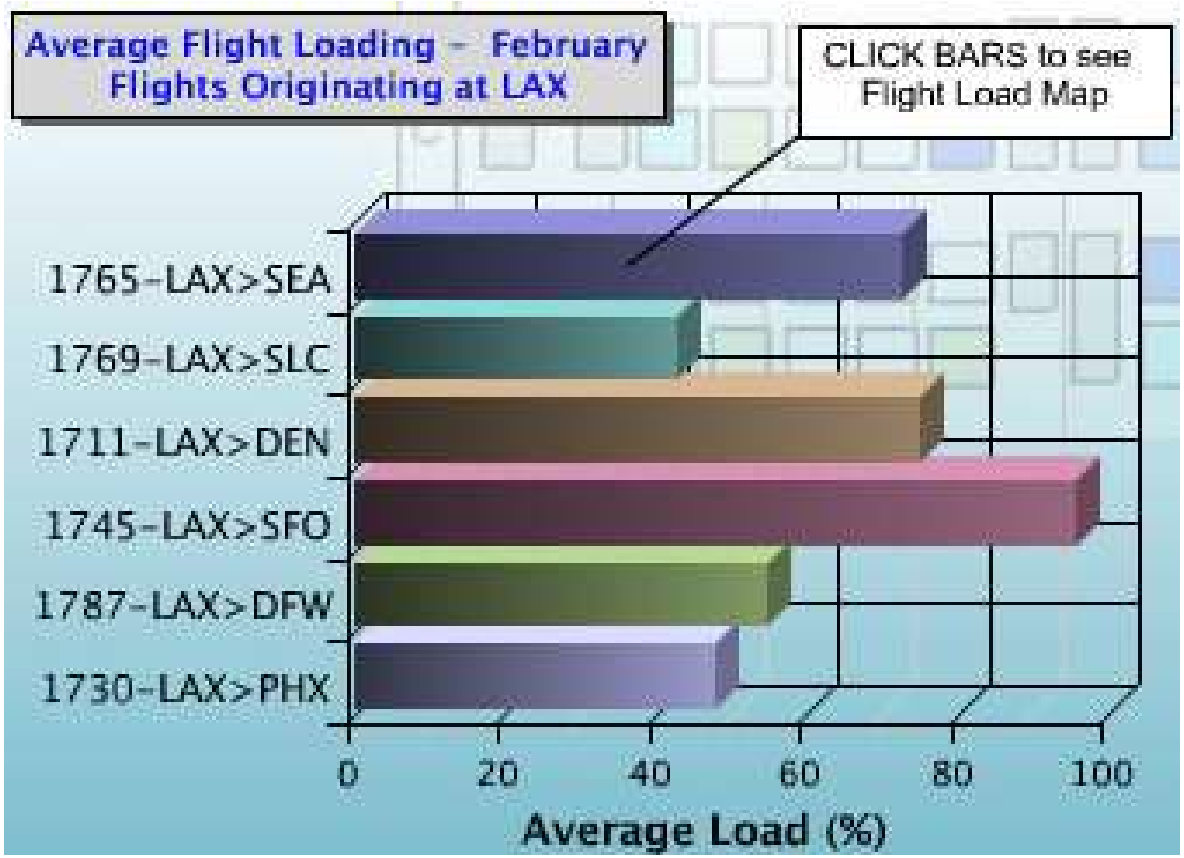


https://en.wikipedia.org/wiki/Space_Shuttle_Challenger_disaster

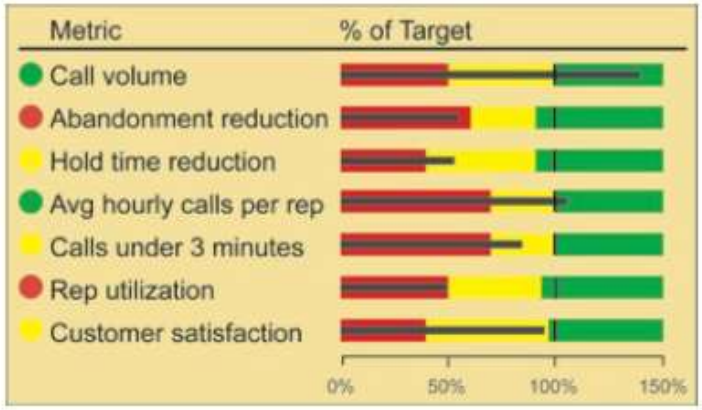
https://www.asktog.com/books/challengerExerpt.html

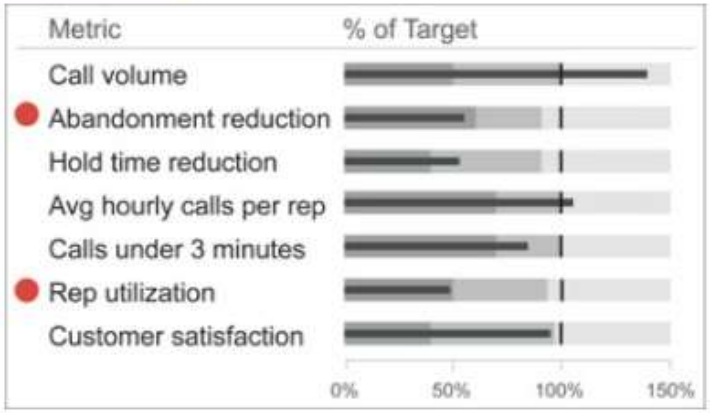
https://eagereyes.org/criticism/tufte-and-the-truth-about-the-challenger

Dashboard Design Mistake #6: **Misusing or overusing colour**



Remember! Simple is Better





"Simplify, simplify, simplify."
Henry David Thoreau

Yes, it's well-designed Dashboard!

Colour has been used carefully

Small, concise display media in a small amount of space.

Graph and Text well balanced

Subtle means to define and group data

Not been cluttered with instructions and descriptions

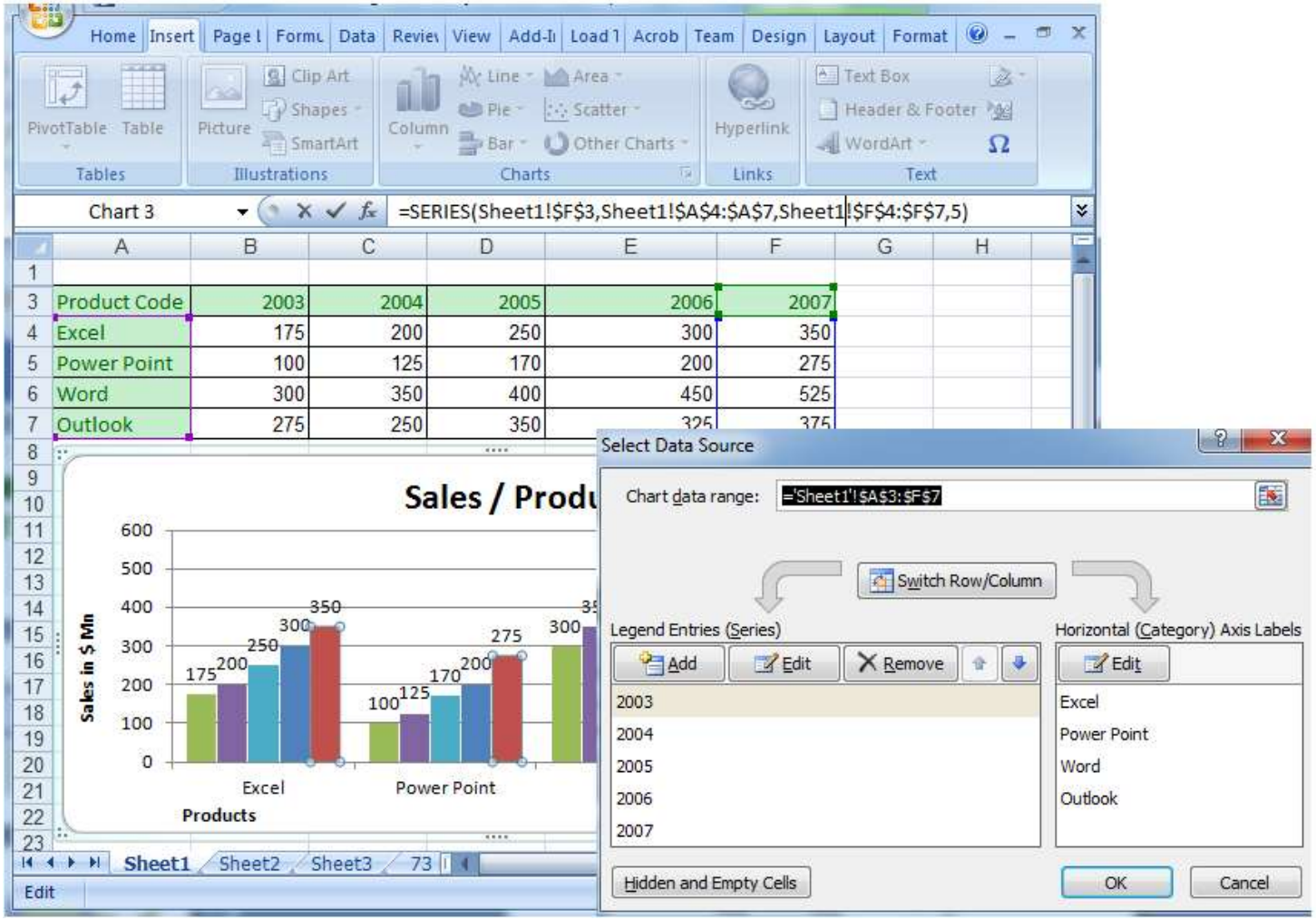
Dashboard Design: Tools

1. Pen and Paper

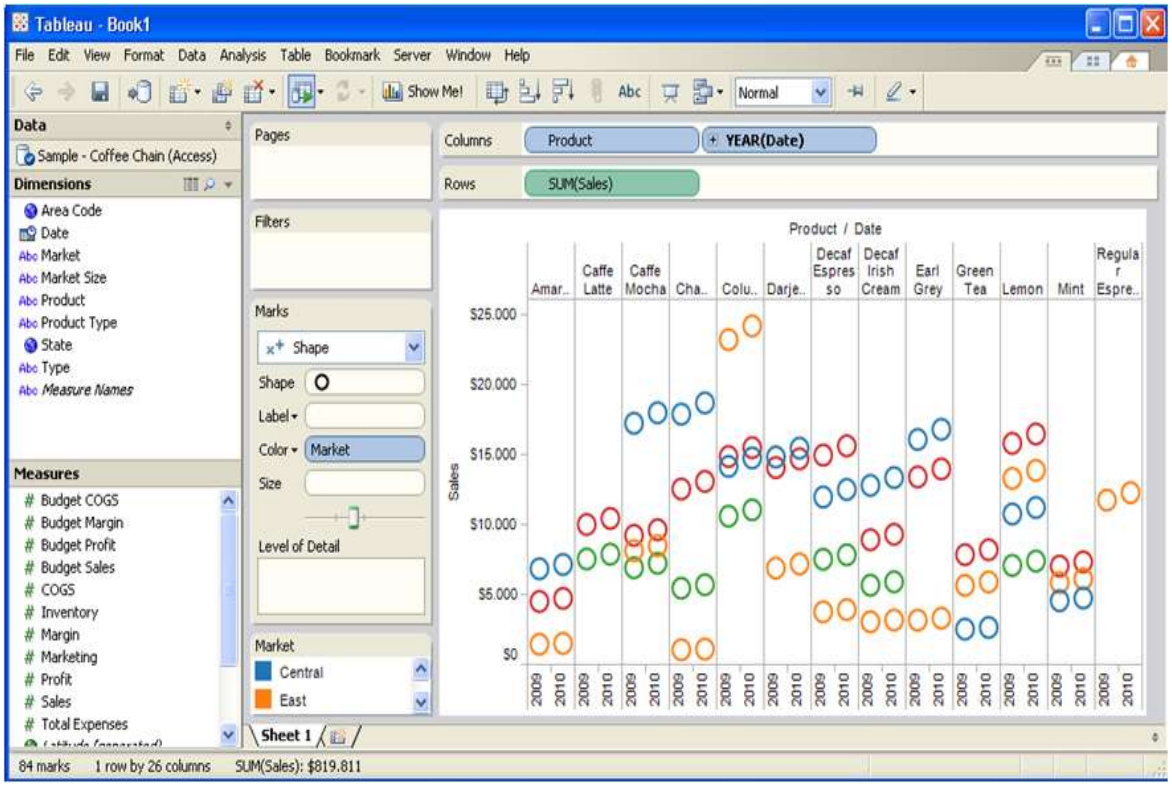
2. Powerpoint, etc.

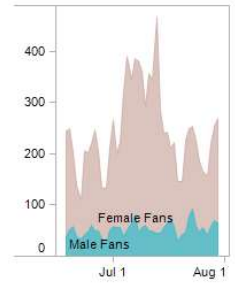
3. InfoVis Tools

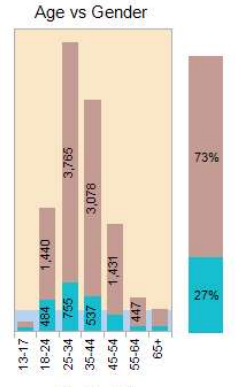
Dashboard Tool #1: **MS Excel**

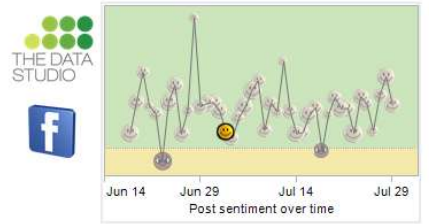


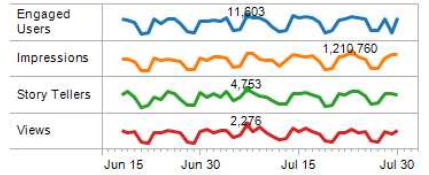
Dashboard Tool #2: **Tableau**



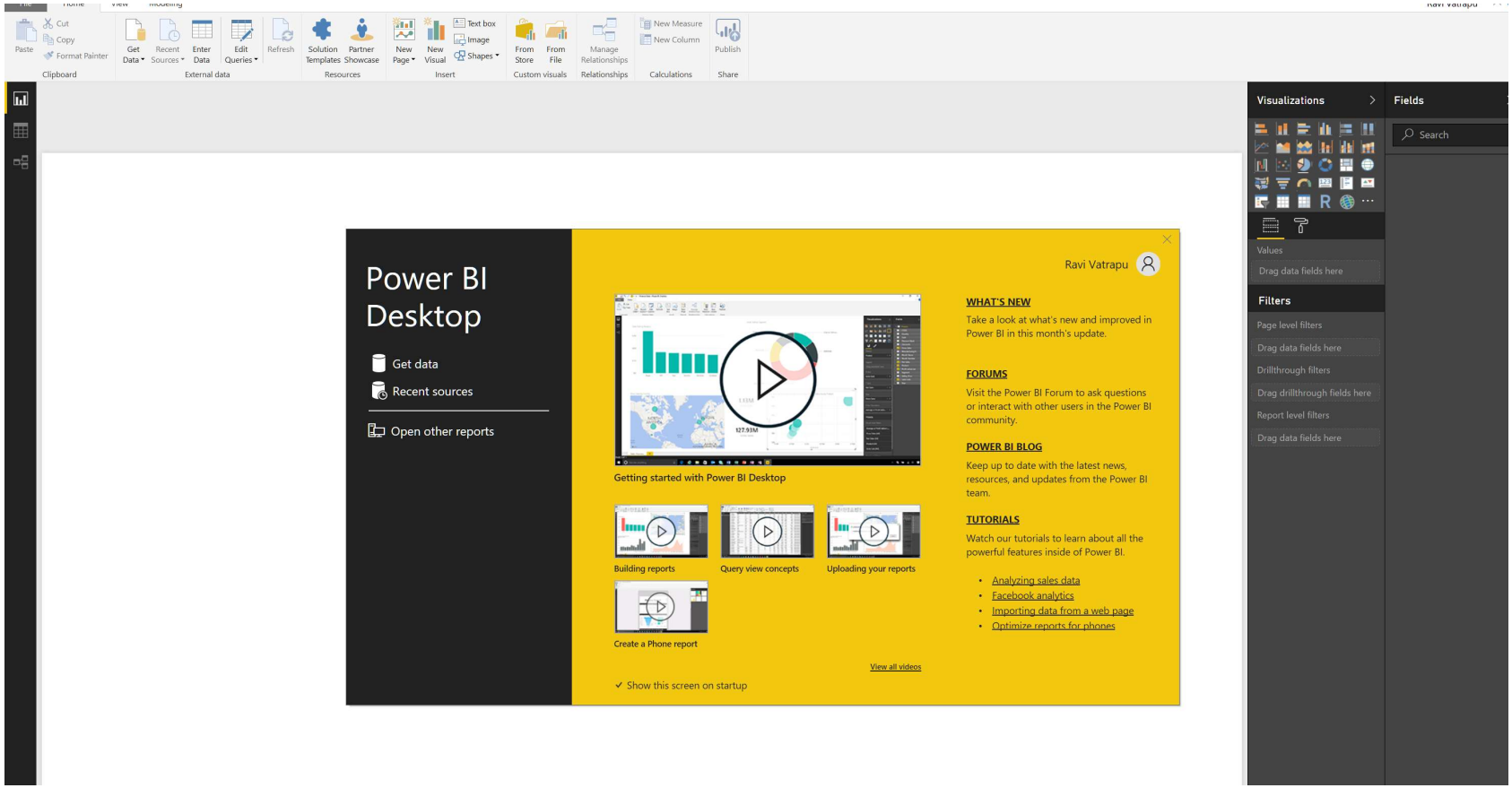




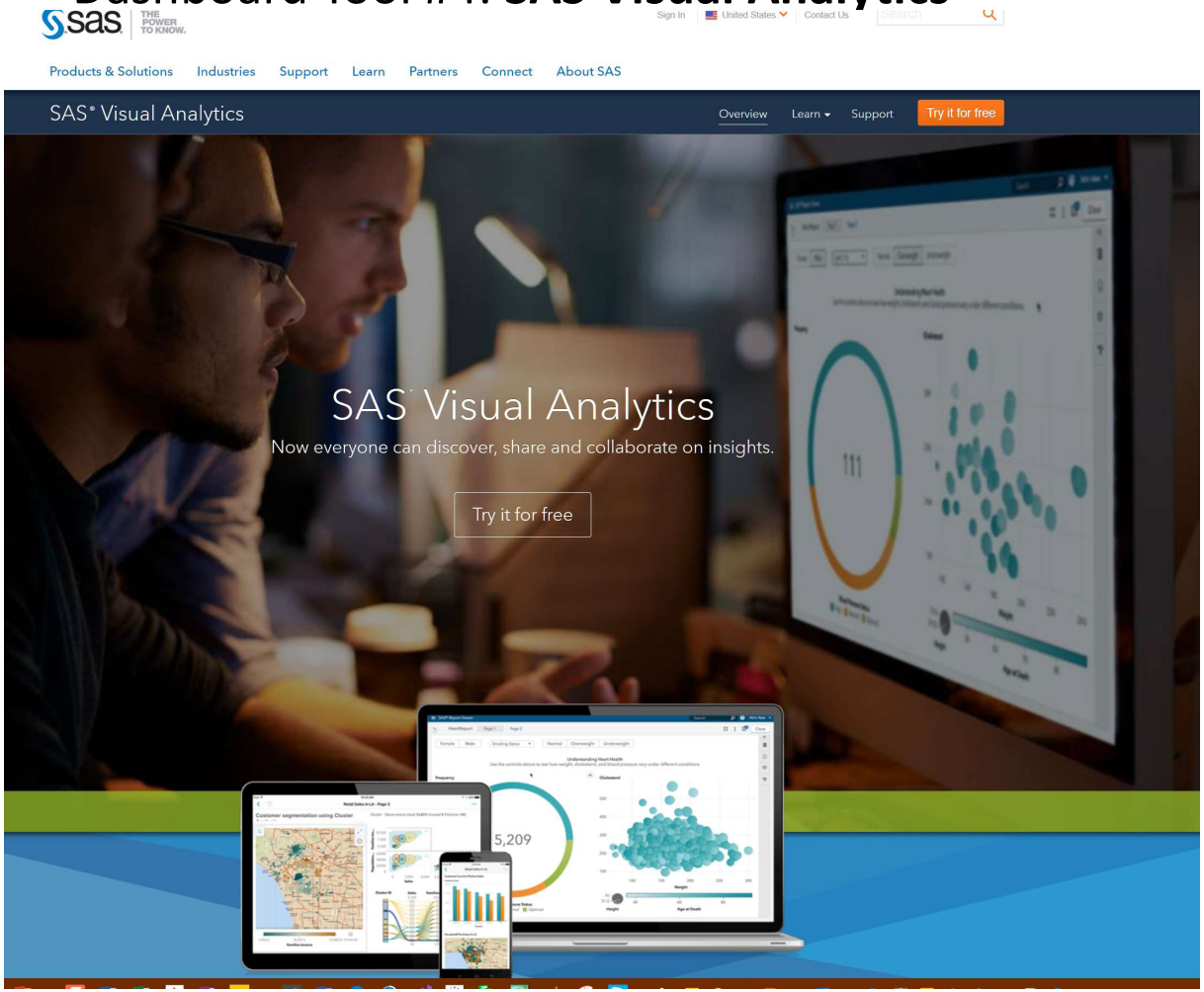




Dashboard Tool #3: **Microsoft Power BI**



Dashboard Tool #4: **SAS Visual Analytics**



Dashboard Tool #5: **IBM Watson**

Overview Details Purchase Resources Sign up for free

Join us at IBM Analytics University! Oct. 10-13, Berlin | Oct. 17-19, New Orleans Register today

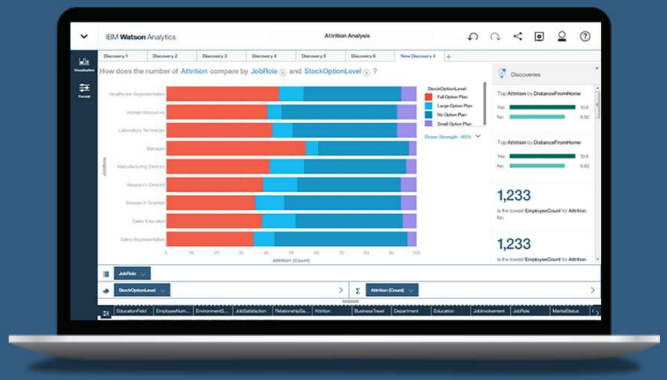
IBM Watson Analytics

IBM Watson Analytics is an advanced data analysis and visualization solution in the cloud that guides you through analysis and discovery of your data.

Starting at kr 249,00 per user per month

Sign up for free | View pricing and buy

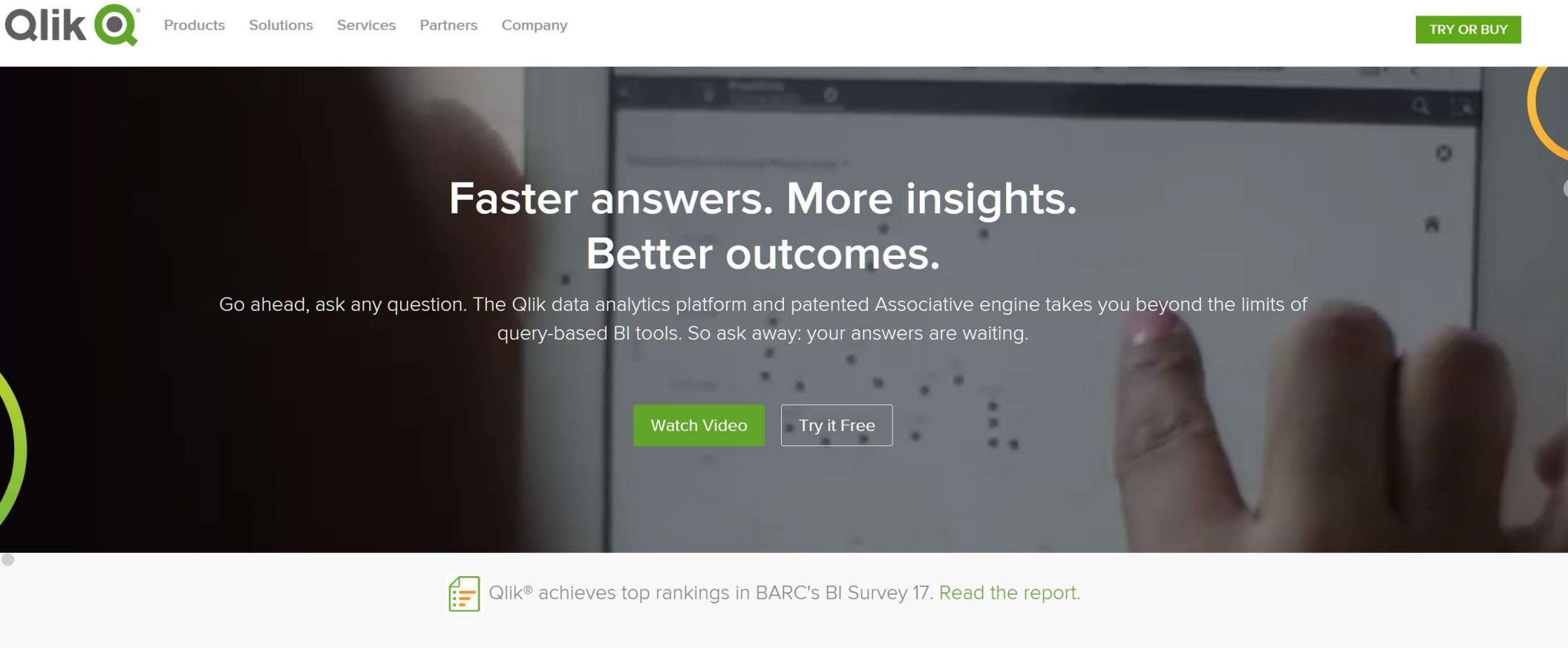
Already a subscriber? **Sign in** >



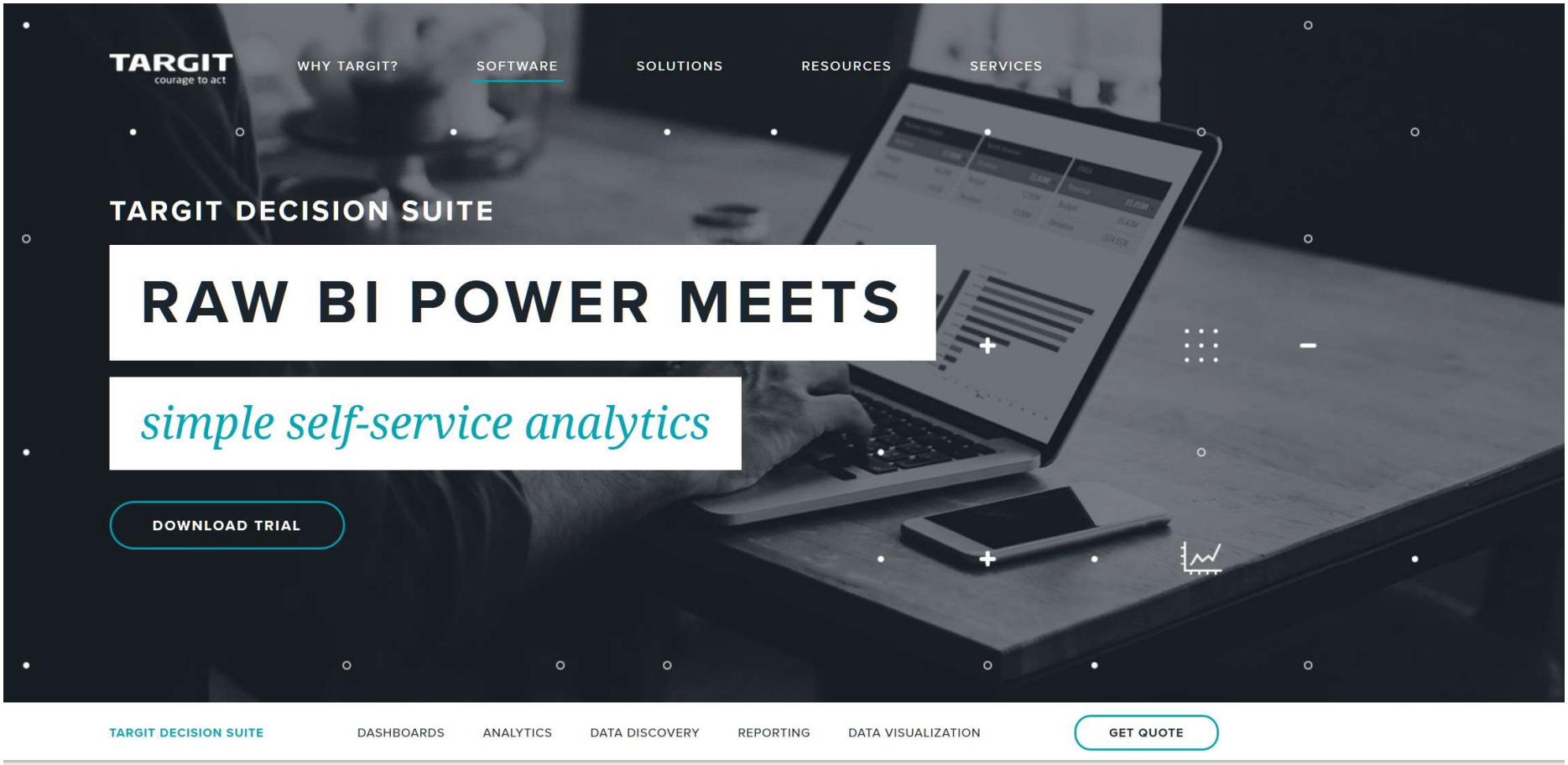
What it can do for your business

Watson Analytics is a smart data analysis and visualization service you can use to quickly discover patterns and meaning in your data – all on your own. With guided data discovery, automated predictive analytics and cognitive capabilities such as natural language dialogue, you can interact with data conversationally to get answers you understand. Whether you need to quickly spot a trend or you have a team that needs to visualize report data in a dashboard, Watson Analytics has you covered.

Dashboard Tool #6: **Qlik**

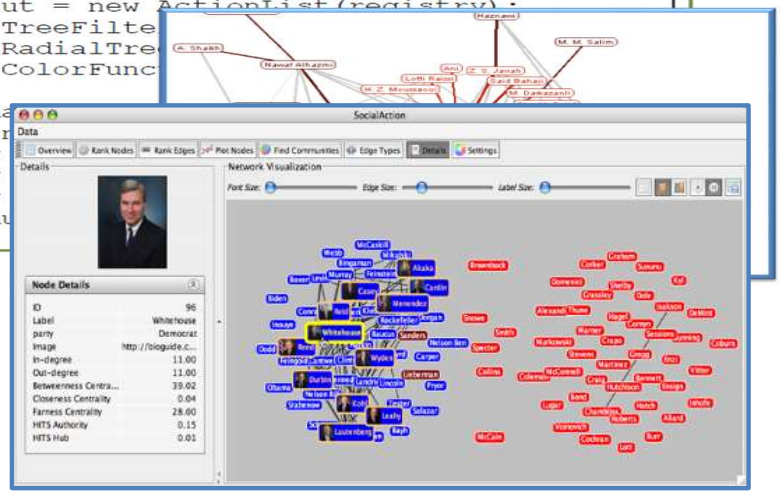


Dashboard Tool #7: **Targit**

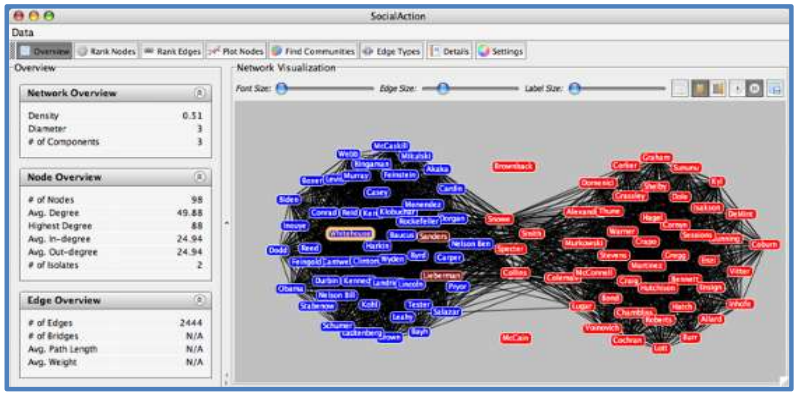


Dashboard Tool #8: **Prefuse**

```
// create graph and registry
Graph g = new XMLGraphReader().loadGraph(datafile);
ItemRegistry registry = new ItemRegistry(g);

// intialize renderers
Renderer nodeR = new TextItemRenderer();
Renderer edgeR = new DefaultEdgeRenderer();
registry.setRendererFactory(
  new DefaultRendererFactory(nodeR, edgeR));

// initialize action lists
ActionList layout = new ActionList(registry);
layout.add(new TreeFilte
layout.add(new RadialTre
layout.add(new ColorFunc

ActionList anima
animate.setPacin
animate.add(new
animate.add(new
animate.add(new
animate.alwaysRu
```



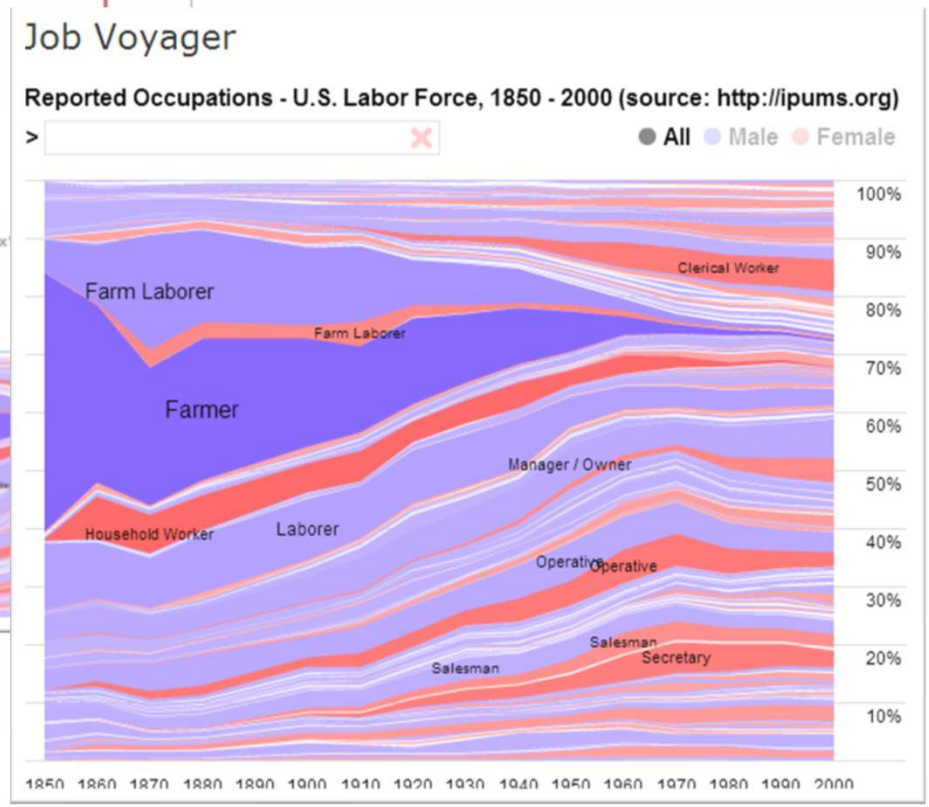
The social network of the U.S. Senators voting patterns. (Perer, 2008)



The social network of the U.S. Senators voting patterns in 2007, after Democrats took control. (Perer, 2008)

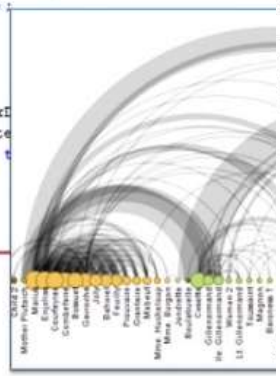
Dashboard Tool #9: **Flare**

```
public class JobVoyager extends App
    {
        private var _bar:ProgressBar;
        private var _bounds:Rectangle;

        protected override function init():void
        {
            addChild(_bar = new ProgressBar());
            _bar.bar.filters = [new DropShadowFilter(1)];

            var ds:DataSource = new DataSource(_url, "tab");
            var ldr:URLLoader = ds.load();
            _bar.loadURL(ldr, function():void {
                // get loaded data, reshape for stacked columns
                var ds:DataSet = ldr.data as DataSet;
                var dr:Array = reshape(ds.nodes.data, ["occupation","sex
                    "year", "people", _cols);
                visualize(Data.fromArray(dr));
                _bar = null;
            });
        }
        private function visualize(data:Data):v
        {
            // prepare data with default settin
            data.nodes.sortBy("data.occupation"
            data.nodes.setProperties({
                shape: Shapes.POLYGON,
                lineColor: 0,
                fillValue: 1,
                fillSaturation: 0.5
            });
            // expression sets male -> blue, fe
            data.nodes.setProperty("fillHue", i
```

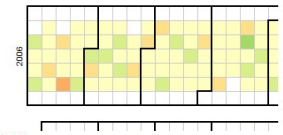


Dashboard Tool #10: **Protovis & D3**

```
var vis = new pv.Panel()
    .width(880)
    .height(310)
    .bottom(90);

var arc = vis.add(pv.Layout.Arc)
    .nodes(miserables.nodes)
    .links(miserables.links)
    .sort(function(a, b) a.group == b.group
        ? b.linkDegree - a.linkDegree
        : b.group - a.group);

arc.link.add(pv.Line);

arc.node.add(pv.Dot)
    .size(function(d) d.linkD
    .fillStyle(pv.Colors.cate
    .strokeStyle(function() t

arc.label.add(pv.Label)

vis.render();
```



Calendar View




```
    .data(d3.range(1990, 2011))
  .enter().append("svg")
    .attr("width", width + margin.right + margin.left)
    .attr("height", height + margin.top + margin.bottom)
    .attr("class", "RdYlGn")
  .append("g")
    .attr("transform", "translate(" + (margin.left + (width - cellSize * 53) / 2) + "," +

svg.append("text")
    .attr("transform", "translate(-6," + cellSize * 3.5 + ")rotate(-90)")
    .attr("text-anchor", "middle")
    .text(String);

var rect = svg.selectAll("rect.day")
```

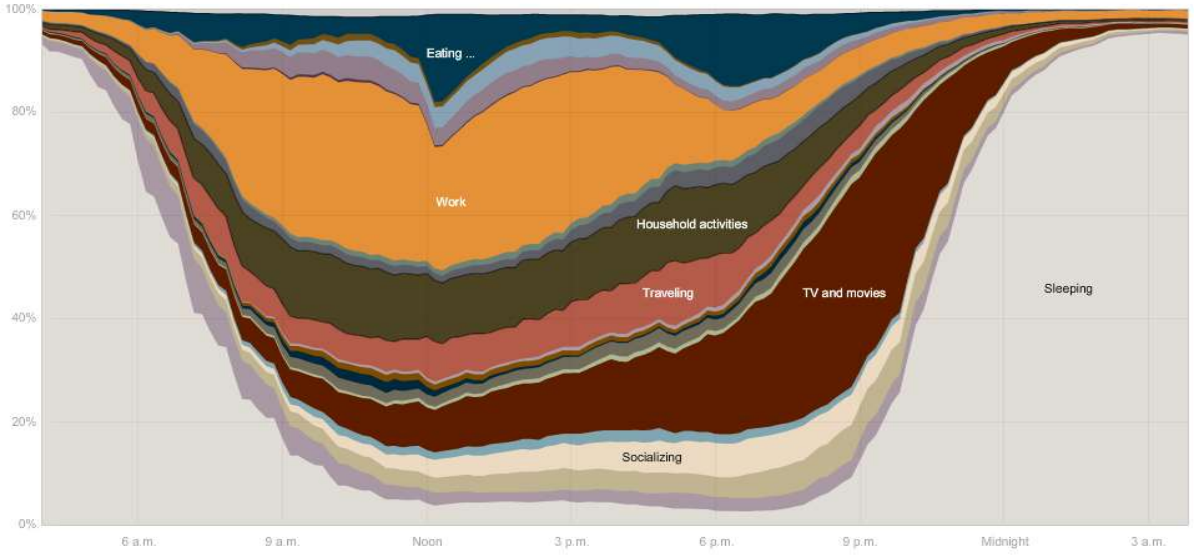
How Different Groups Spend Their Day

The American Time Use Survey asks thousands of American residents to recall every minute of a day. Here is how people over age 15 spent their time in 2008. Related article

Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



Dashboard Tool #11: **Python**

Reflections