Database Management Systems Visualization

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

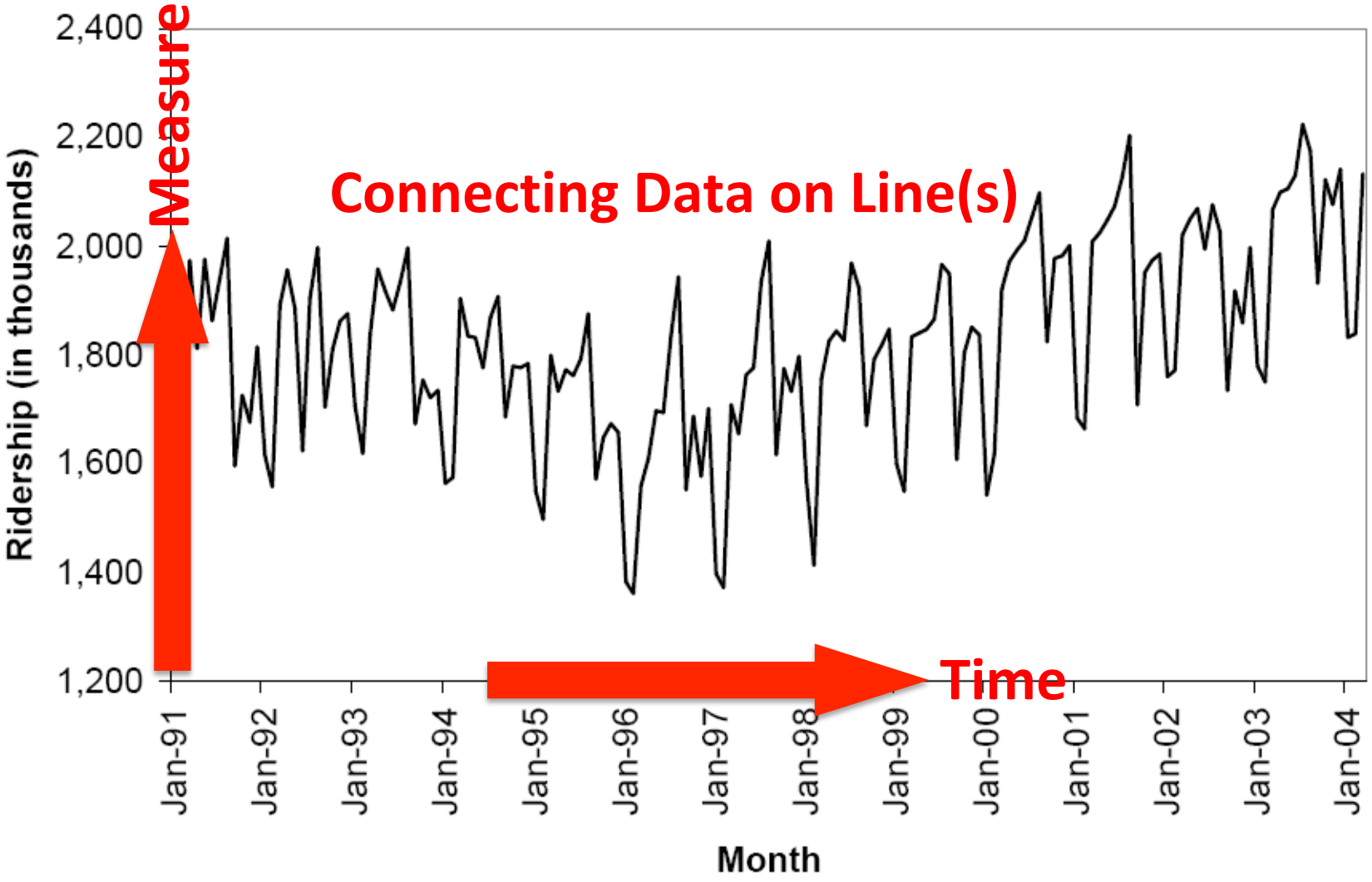
DR. ADAM LEE

Graphs for Data Exploration

- Basic Plots:
  - Line Graphs
  - Bar Charts
  - Scatterplots
- Distribution Plots
  - Boxplots
  - Histograms
- Heatmaps
- Multidimensional Visualization
  - adding variables
  - manipulations

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Line Graph for Time Series



UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Excel: Line and Area Charts

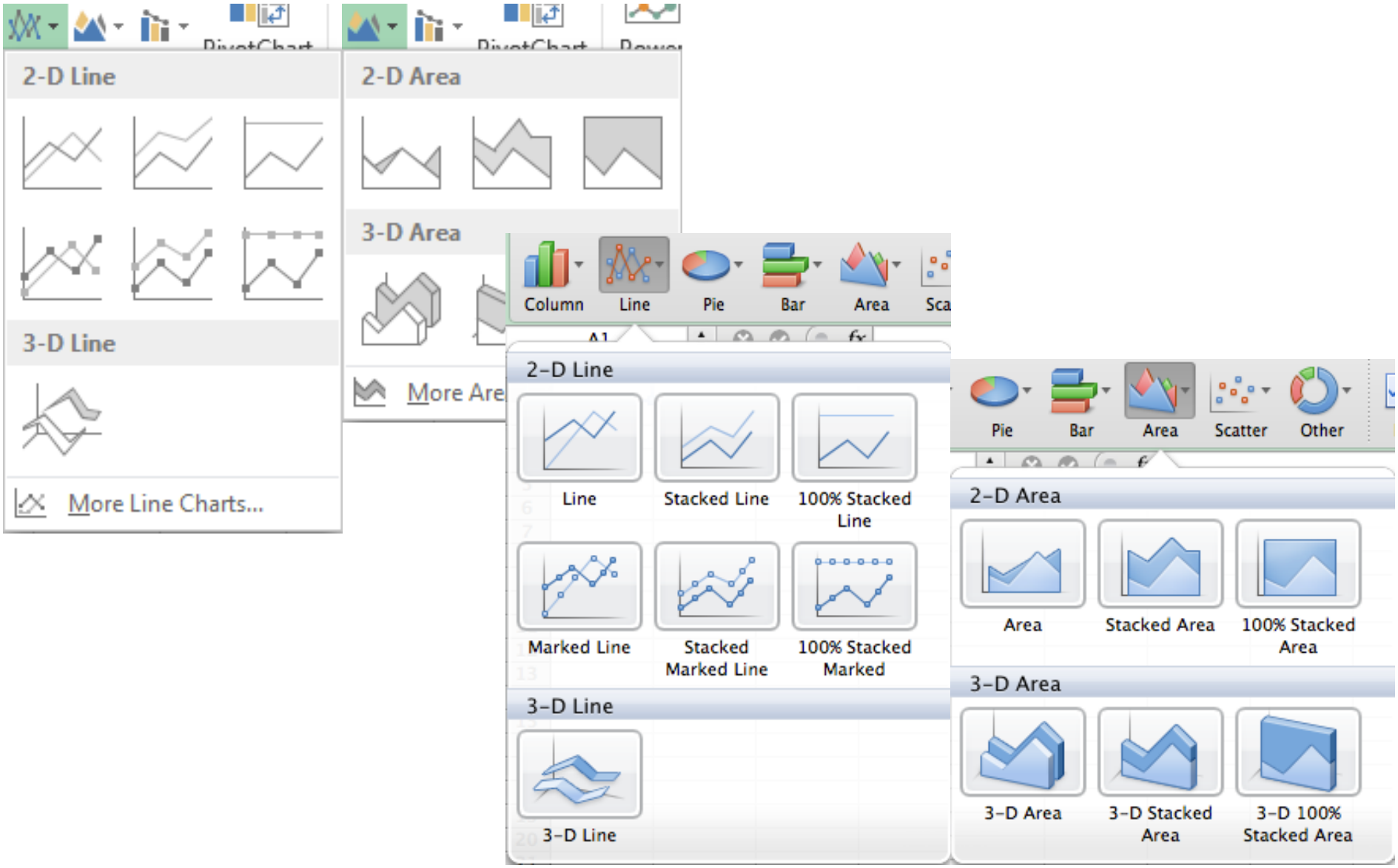
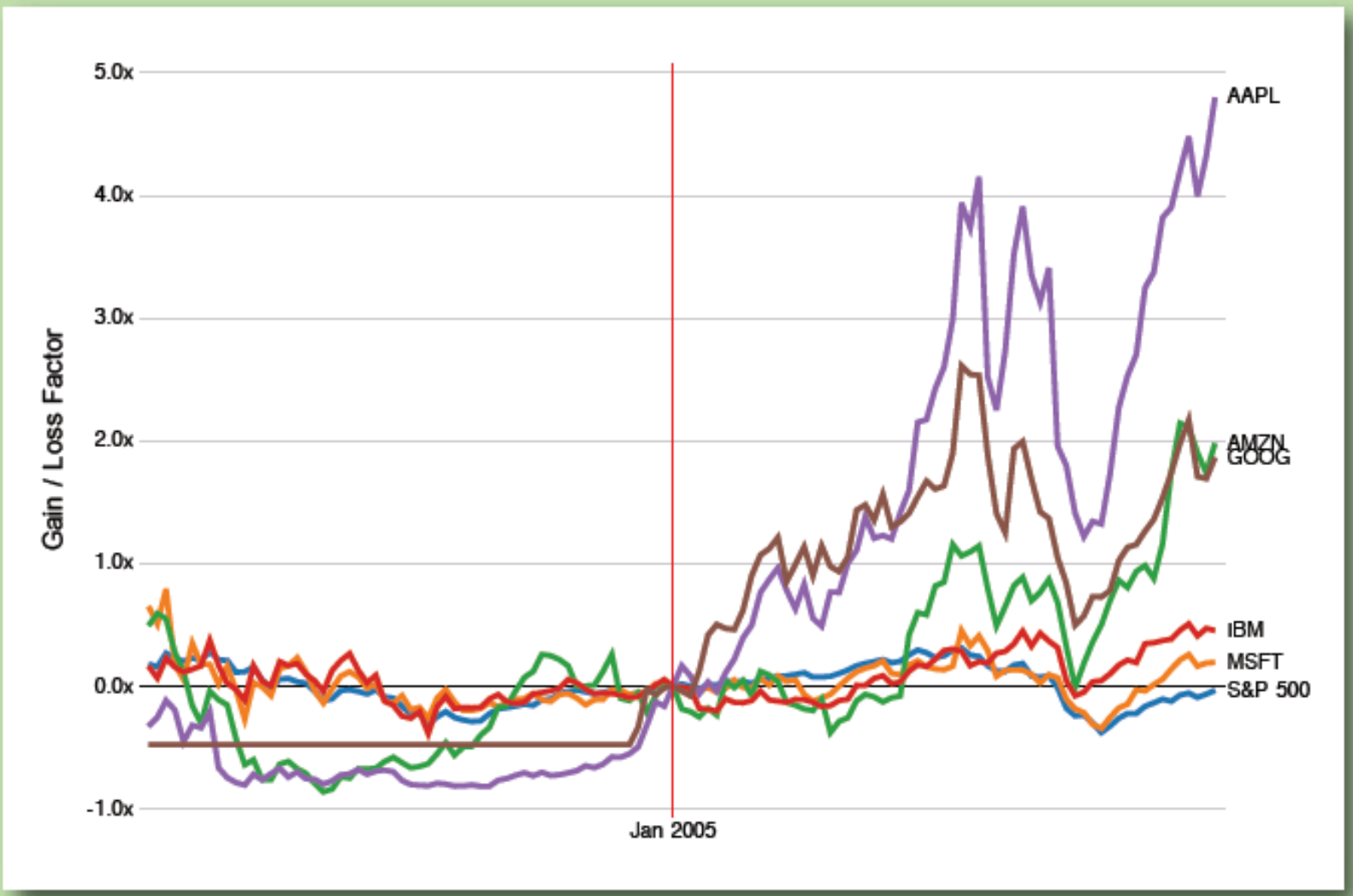


FIGURE 1A

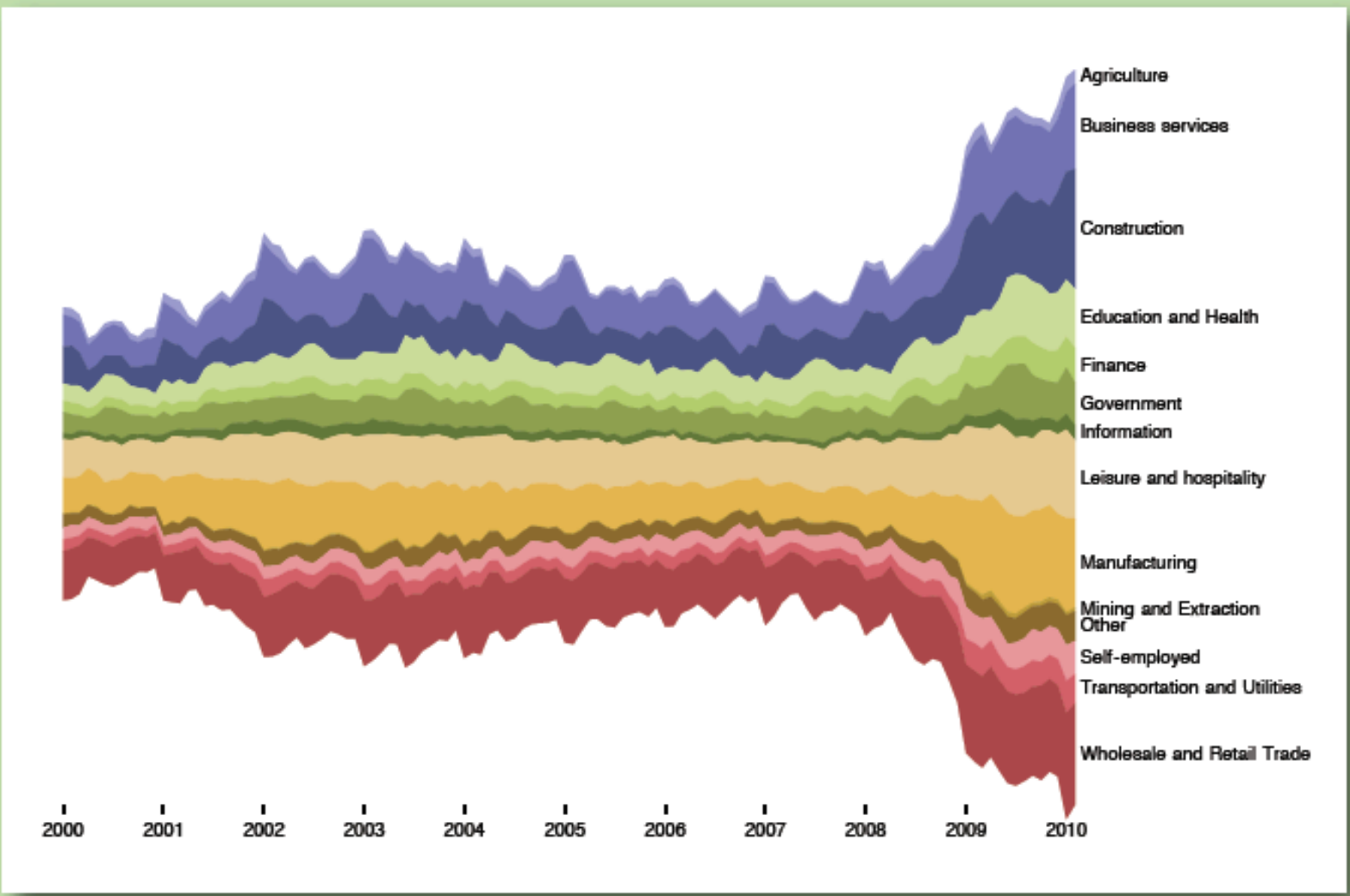
Index Chart of Selected Technology Stocks, 2000-2010



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

FIGURE 1B

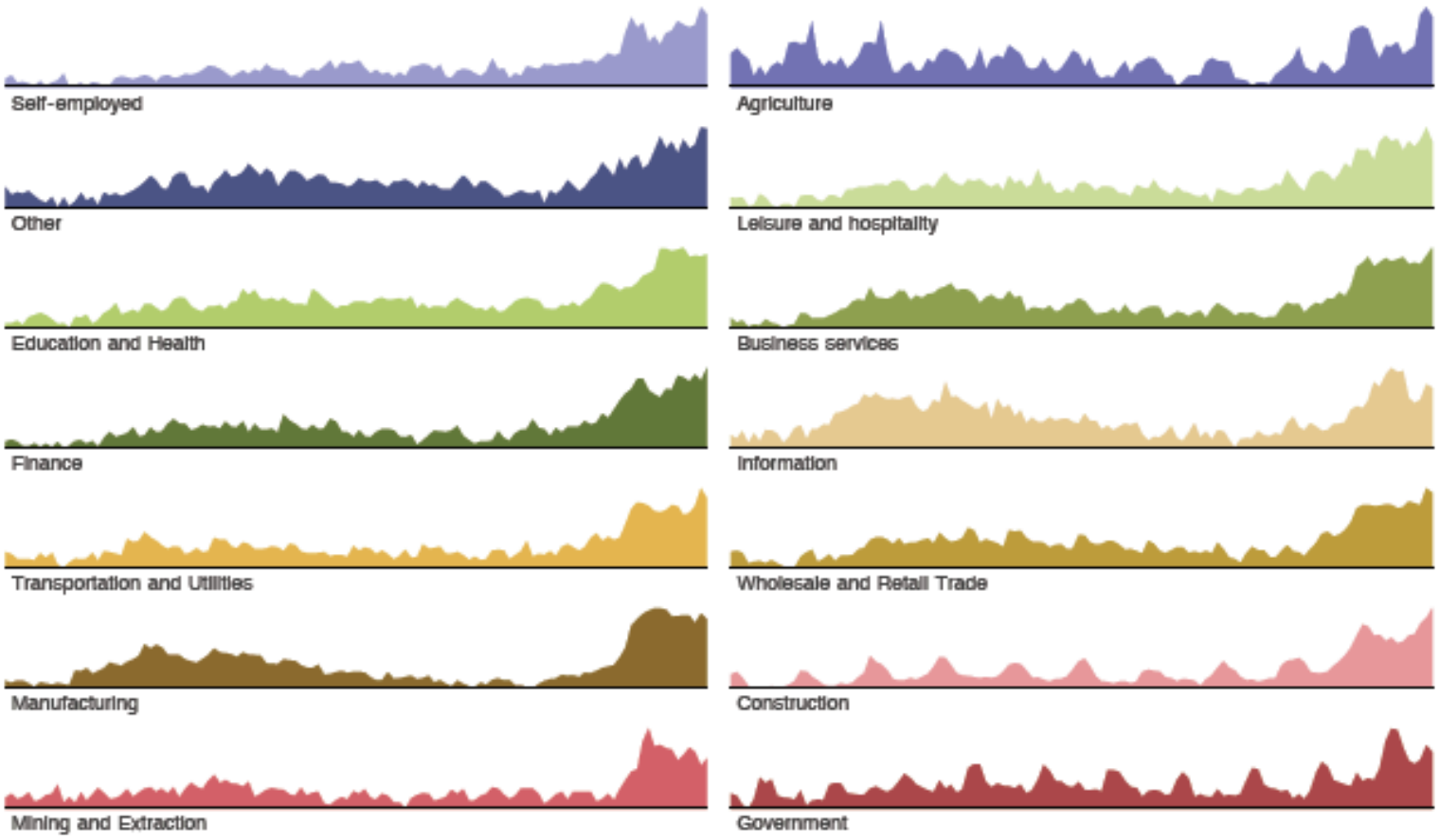
Stacked Graph of Unemployed U.S. Workers by Industry, 2000-2010



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

FIGURE 1C

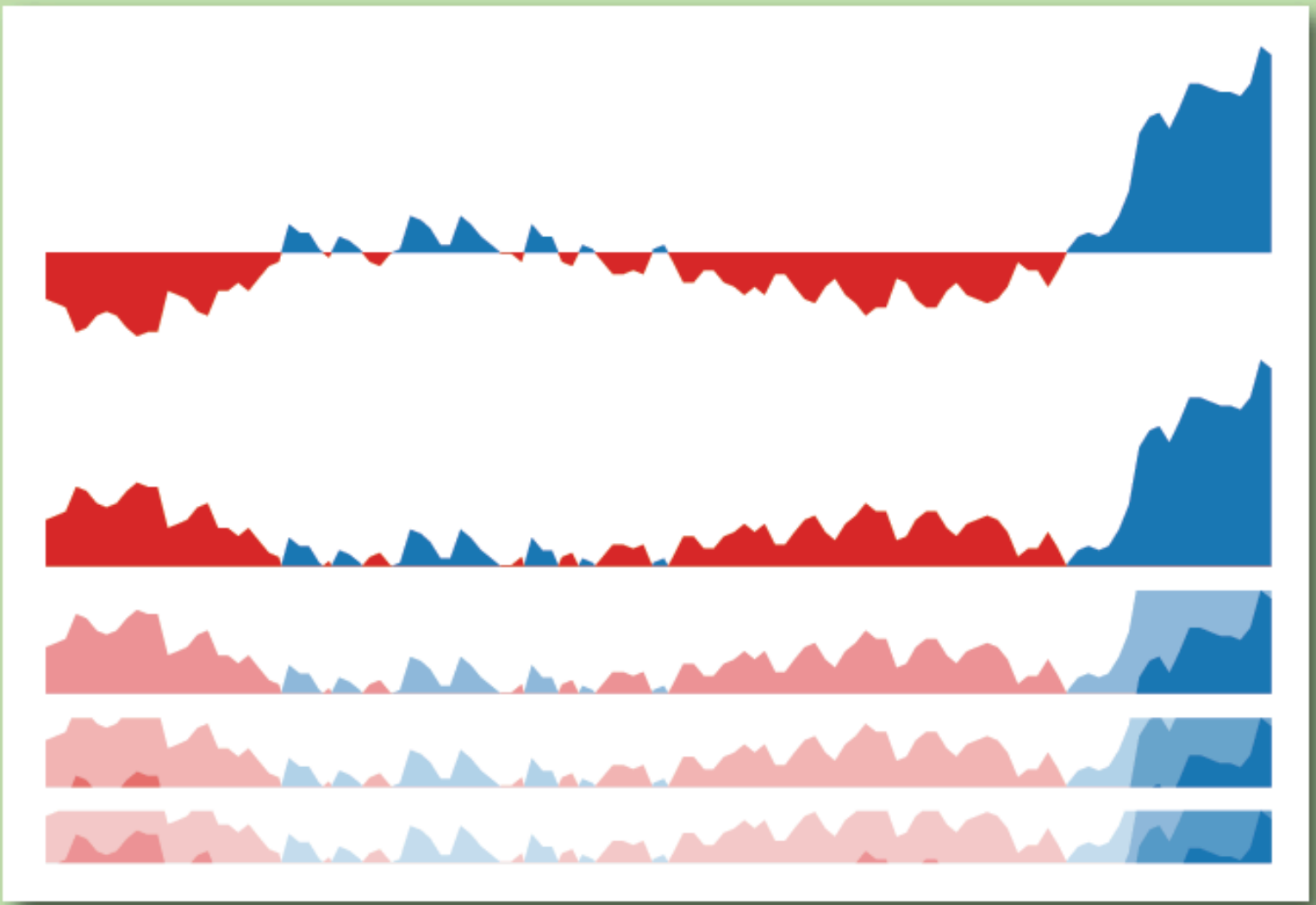
Small Multiples of Unemployed U.S. Workers Normalized by Industry, 2000-2010



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

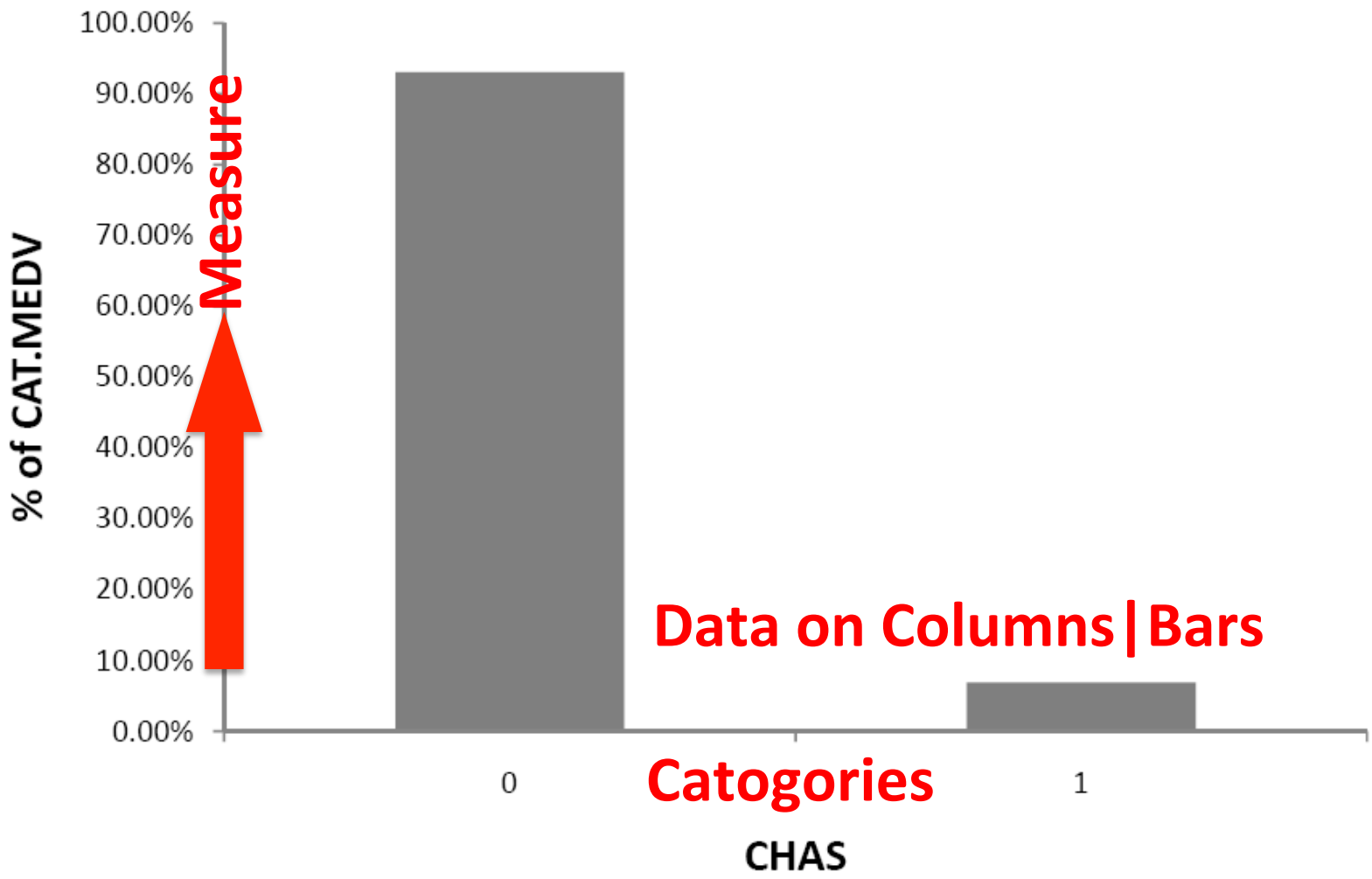
FIGURE 1D

Horizon Graphs of U.S. Unemployment Rate, 2000-2010

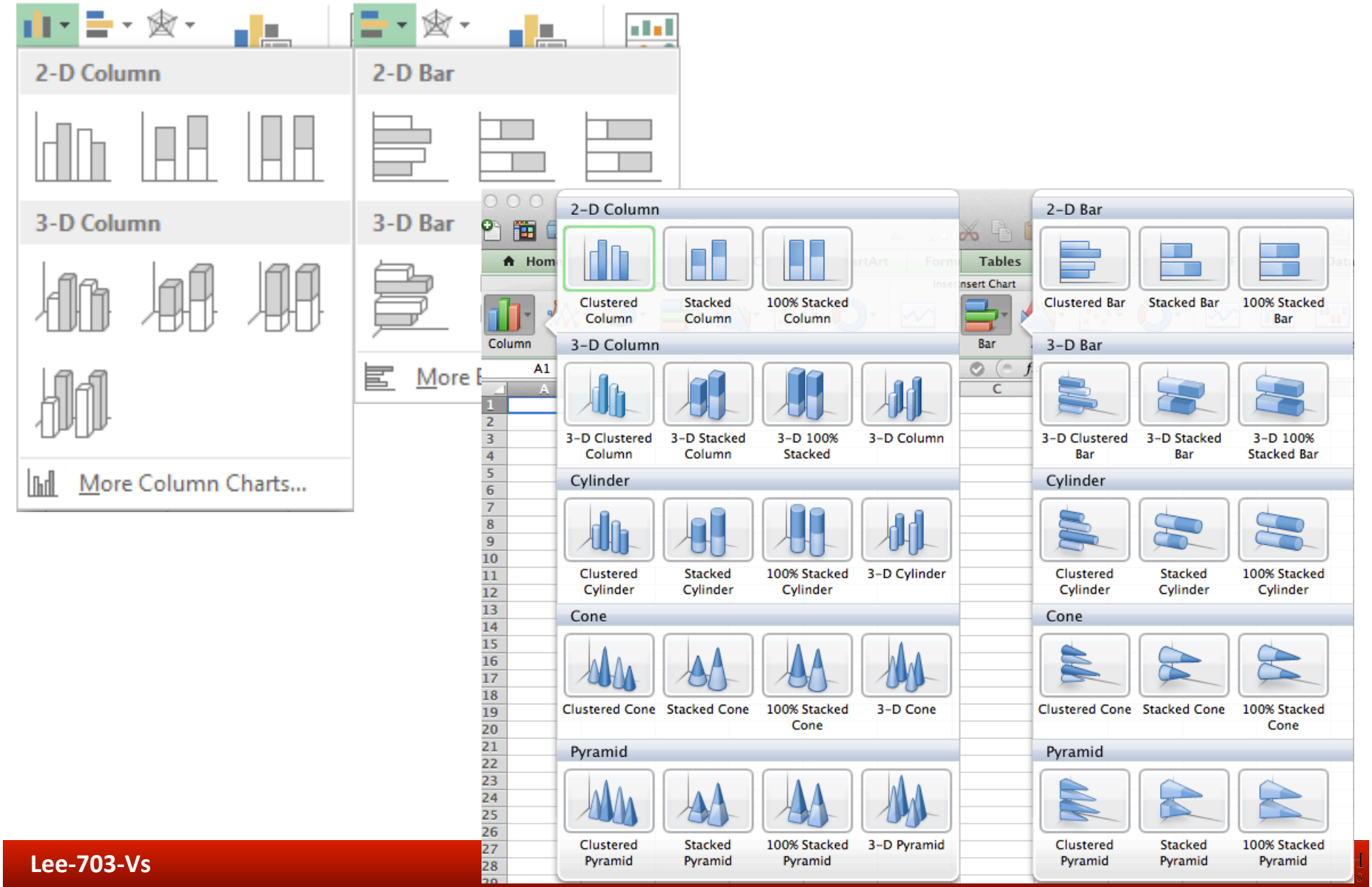


Source: U.S. Bureau of Labor Statistics
http://hci.stanford.edu/jheer/files/zoo/ex/time/horizon.html

Bar Chart for Categorical Variable

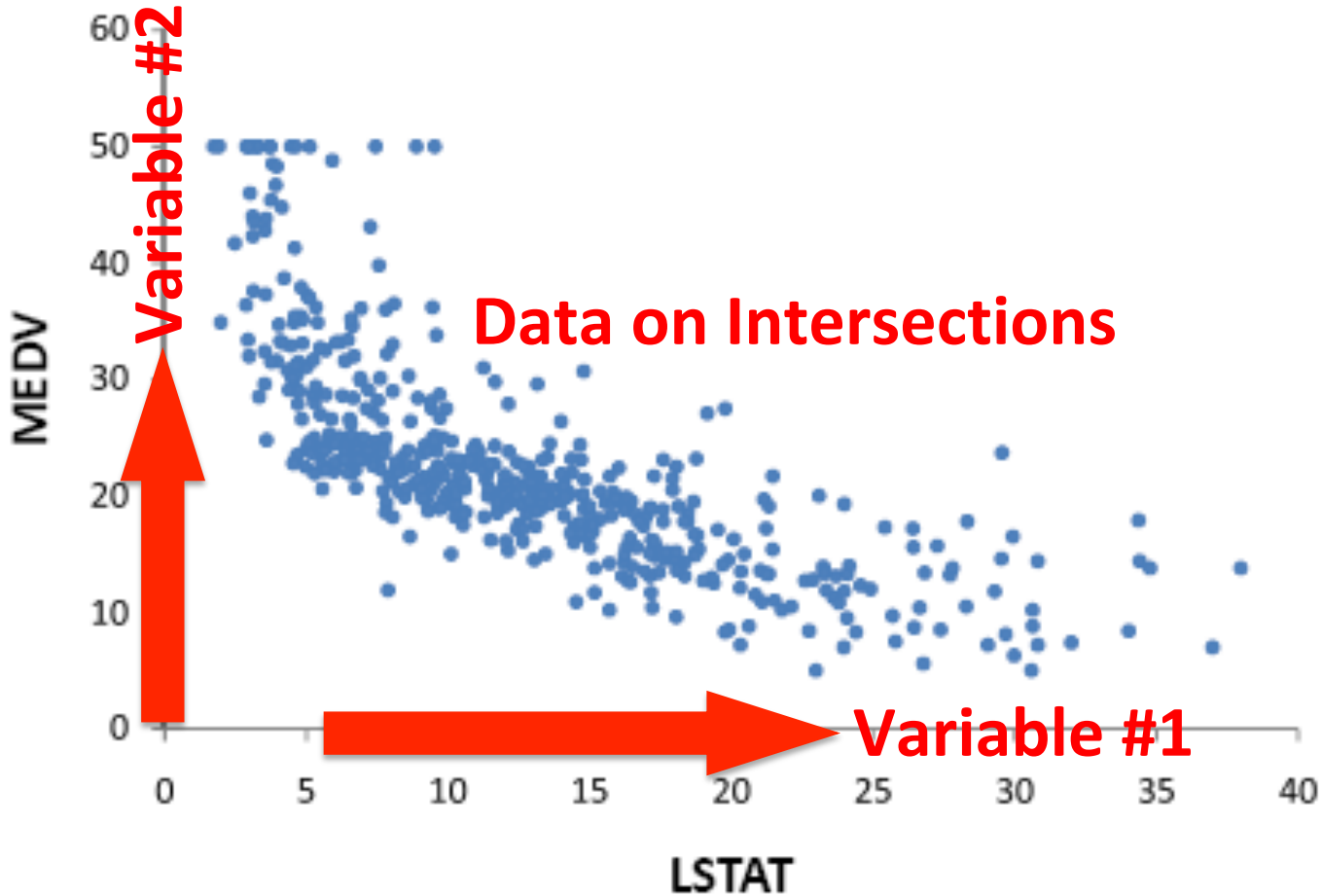


Excel: Column and Bar Charts

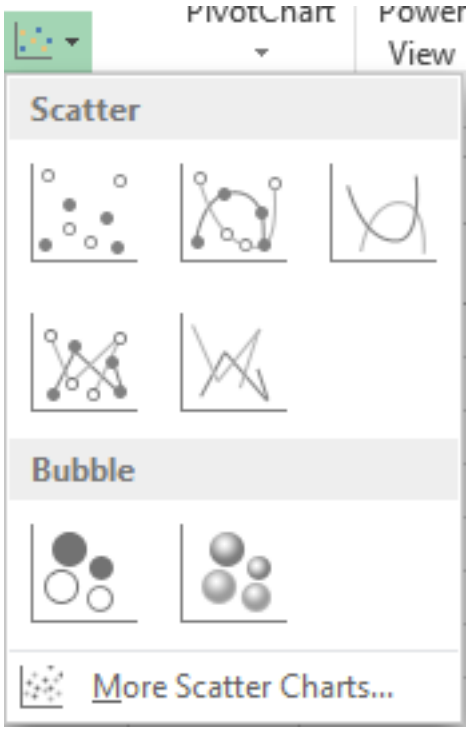


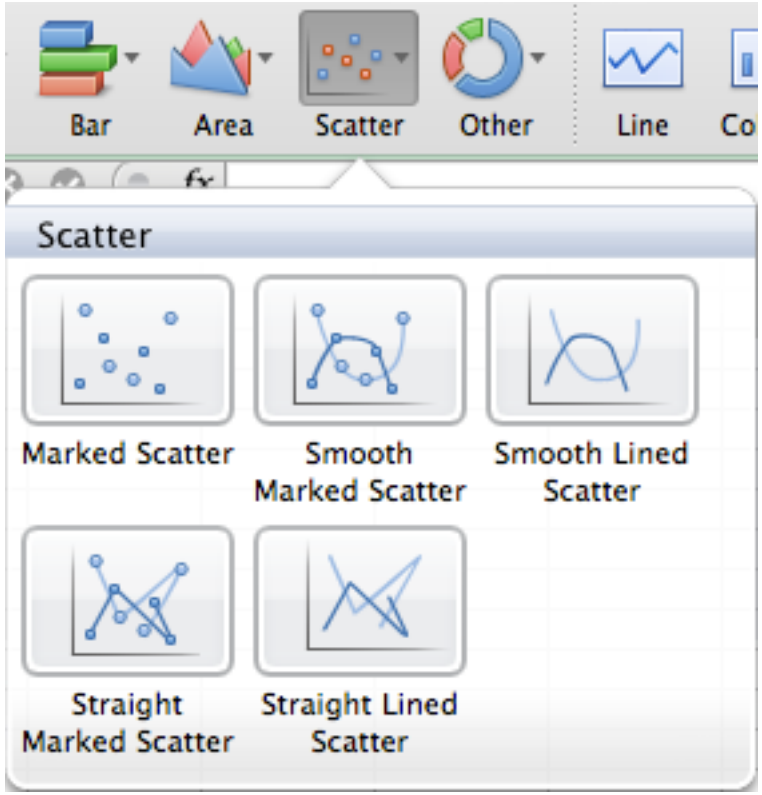
Scatterplot

- Displays relationship between two numerical variables



Excel: Scatter Charts





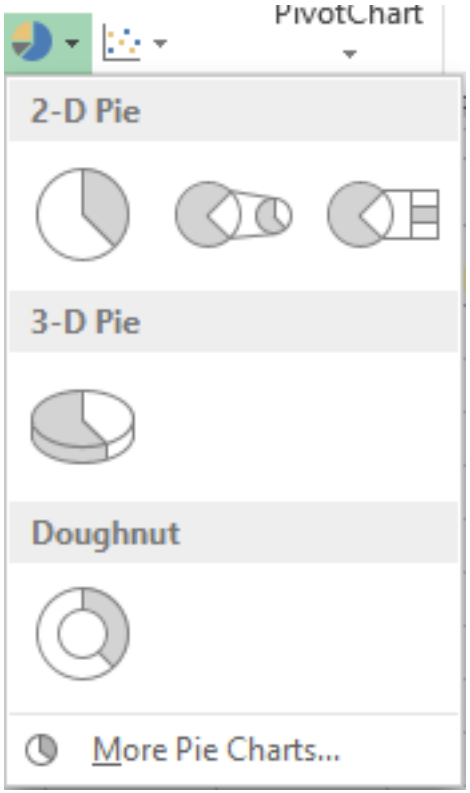
UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

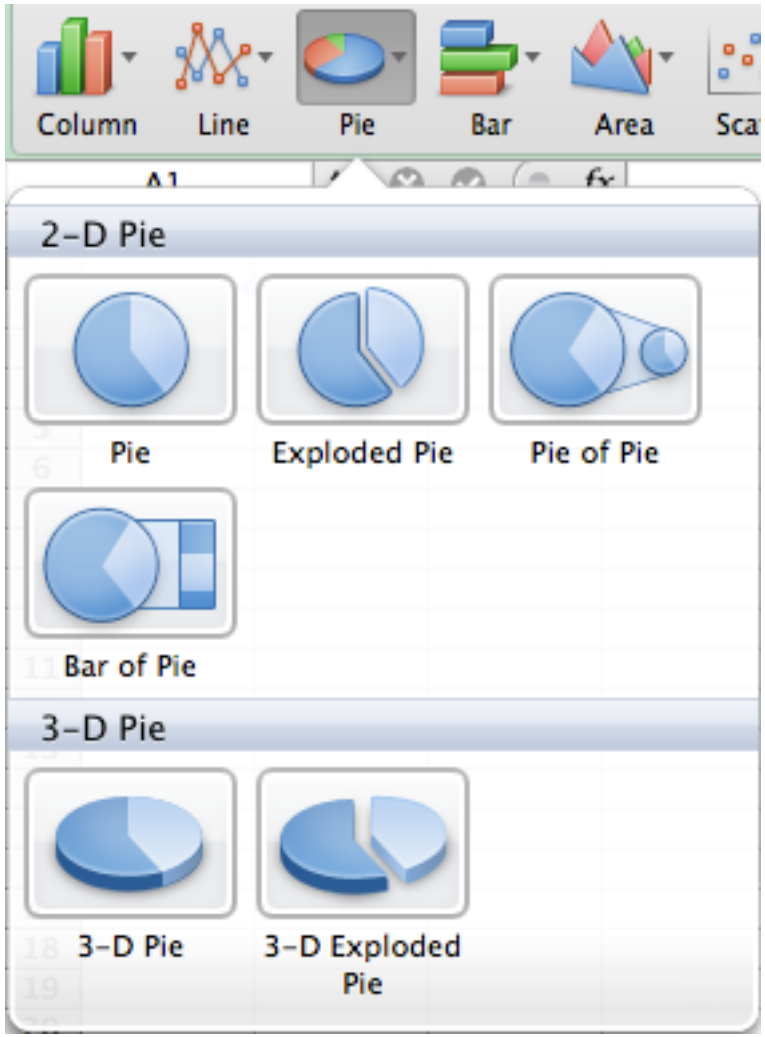
Distribution Plots

- Display “how many” of each value occur in a data set
- Or, for continuous data or data with many possible values, “how many” values are in each of a series of ranges or “bins”

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Excel: Pie Charts





UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Histograms

- Histogram shows the distribution of the outcome variable (e.g., median house value in Boston Housing example)

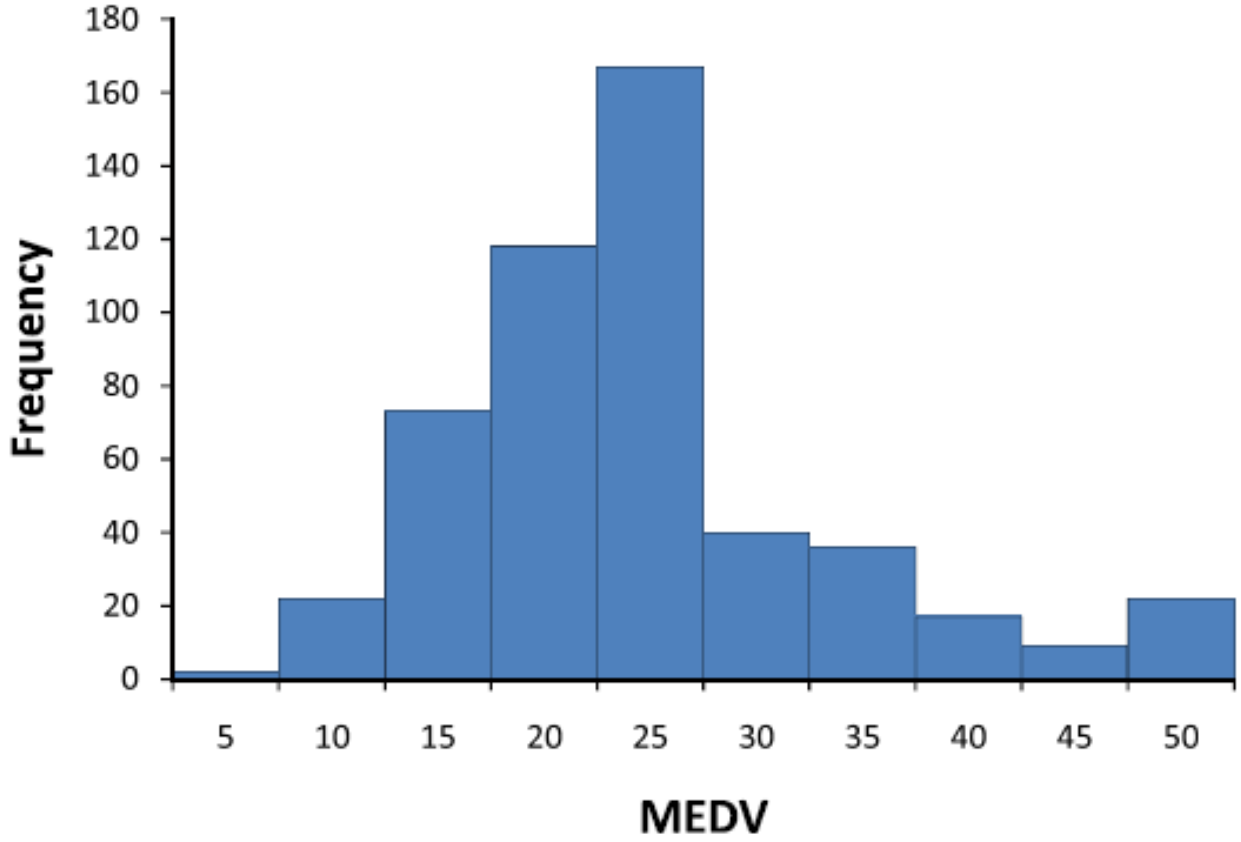


FIGURE 2A

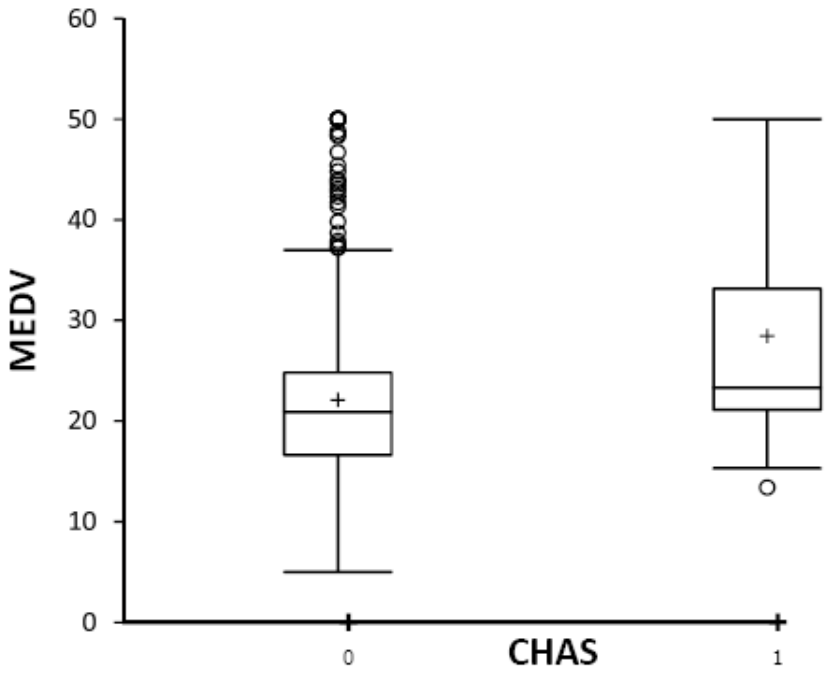
Stem-and-Leaf Plot of Mechanical Turk Participation Rates

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 9
10	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

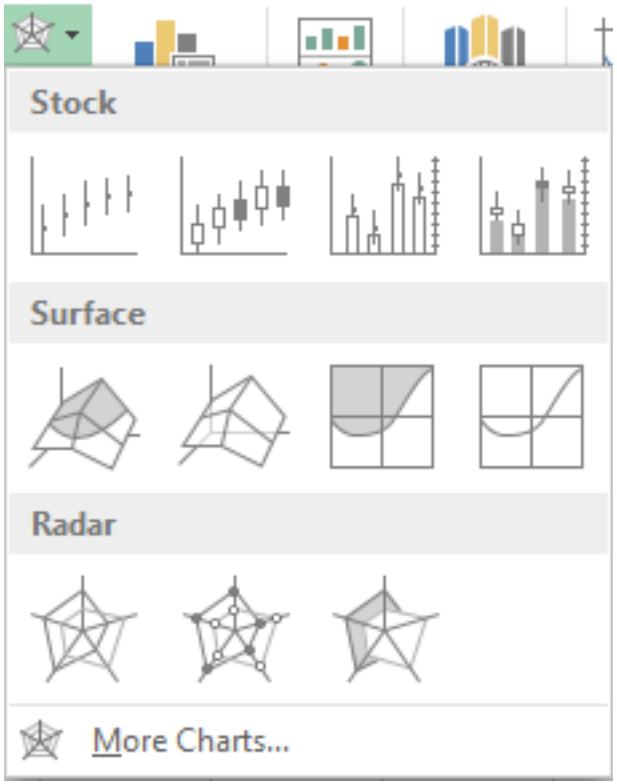
Source: Stanford Visualization Group
http://hci.stanford.edu/jheer/files/zoo/ex/stats/stem-and-leaf.html

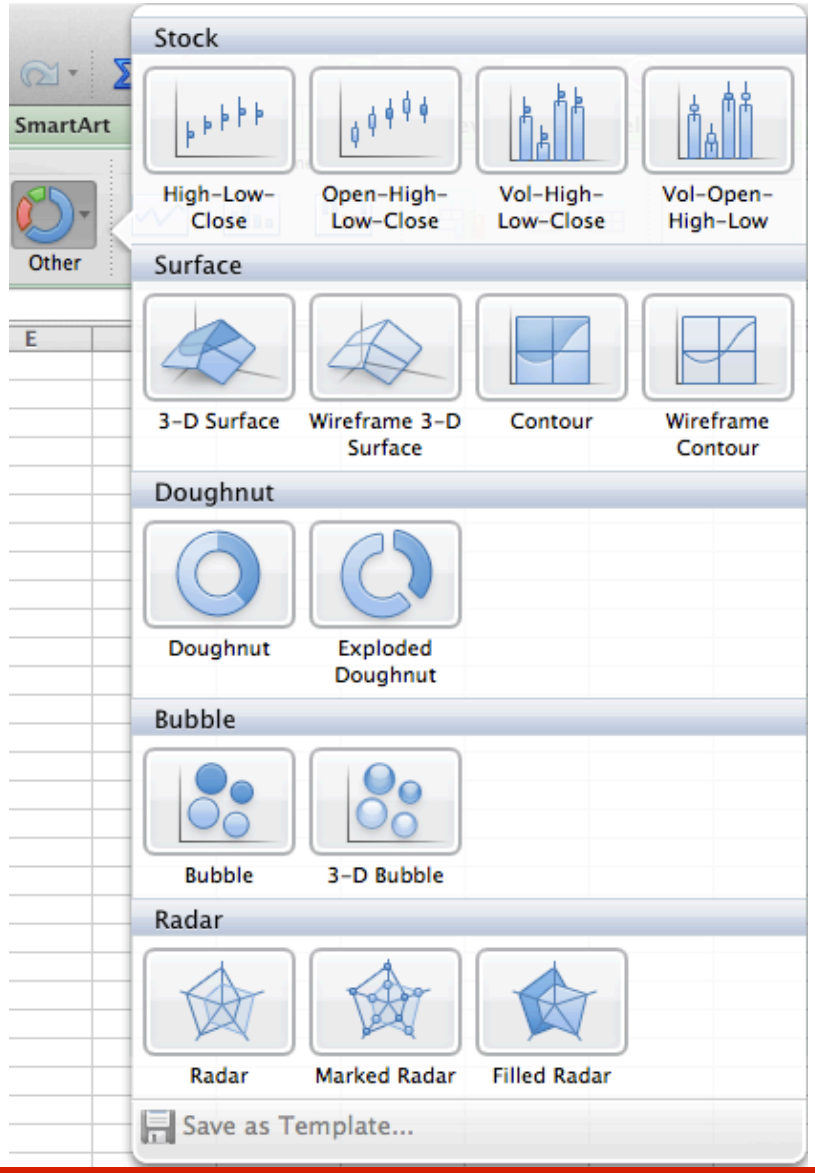
Boxplots

- Side-by-side boxplots are useful for comparing subgroups (e.g., distribution of outcome variable (MEDV) for neighborhoods on Charles River (1) and not on Charles River (0) in Boston Housing example)



Excel: Stock Charts

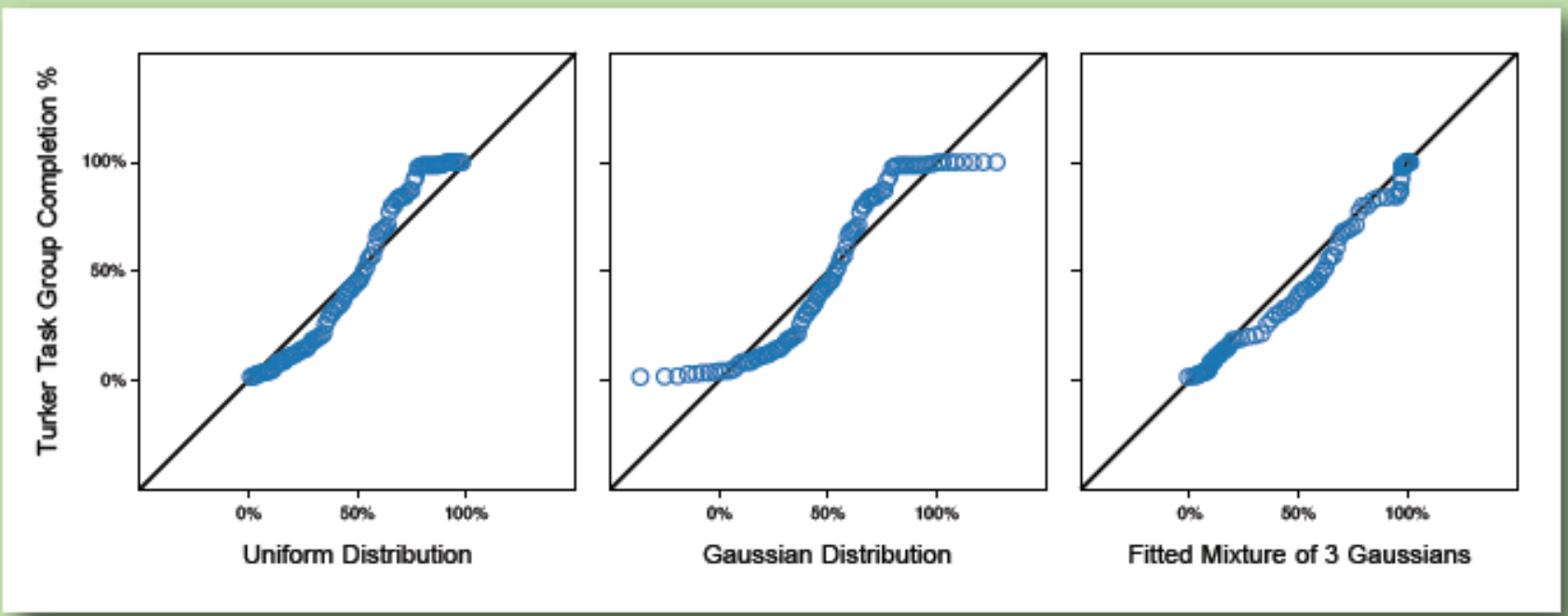




UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

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

UNIVERSITY OF MARYLAND

ROBERT H. SMITH SCHOOL OF BUSINESS

Heatmaps

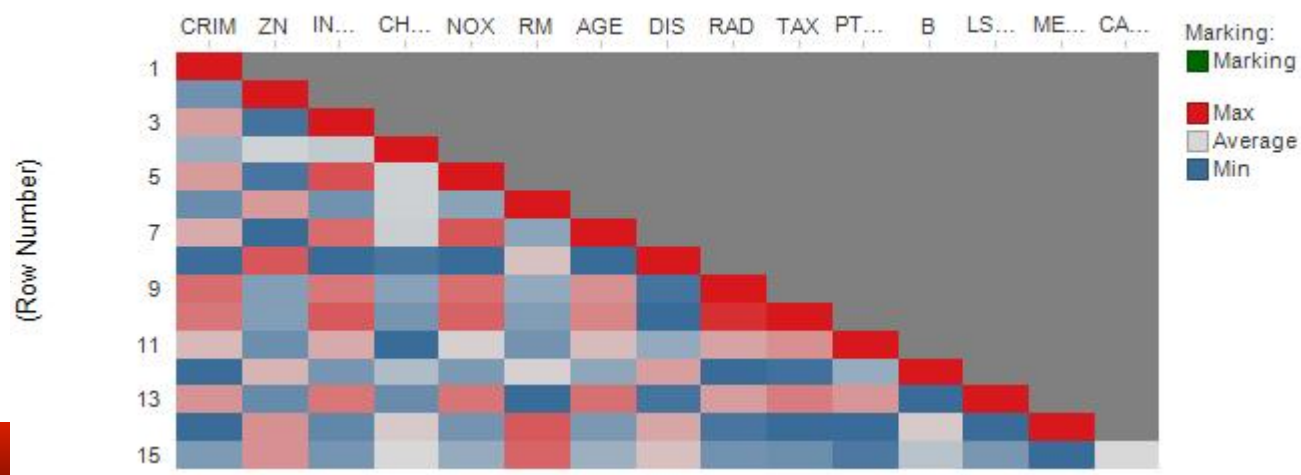
- Color conveys information

To visualize:

- Correlations
- Missing Data

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
CRIM														
ZN	-0.20													
INDUS	0.41	-0.53												
CHAS	-0.06	-0.04	0.06											
NOX	0.42	-0.52	0.76	0.09										
RM	-0.22	0.31	-0.39	0.09	-0.30									
AGE	0.35	-0.57	0.64	0.09	0.73	-0.24								
DIS	-0.38	0.66	-0.71	-0.10	-0.77	0.21	-0.75							
RAD	0.63	-0.31	0.60	-0.01	0.61	-0.21	0.46	-0.49						
TAX	0.58	-0.31	0.72	-0.04	0.67	-0.29	0.51	-0.53	0.91					
PTRATIO	0.29	-0.39	0.38	-0.12	0.19	-0.36	0.26	-0.23	0.46	0.46				
B	-0.39	0.18	-0.36	0.05	-0.38	0.13	-0.27	0.29	-0.44	-0.44	-0.18			
LSTAT	0.46	-0.41	0.60	-0.05	0.59	-0.61	0.60	-0.50	0.49	0.54	0.37	-0.37		
MEDV	-0.39	0.36	-0.48	0.18	-0.43	0.70	-0.38	0.25	-0.38	-0.47	-0.51	0.33	-0.74	

Heat Map



Multidimensional Visualization

Adding variables:

- Add colors
- Matrix plots

Manipulations:

- Aggregation
- Rescaling
- Jittering
- Linked plots
- …

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

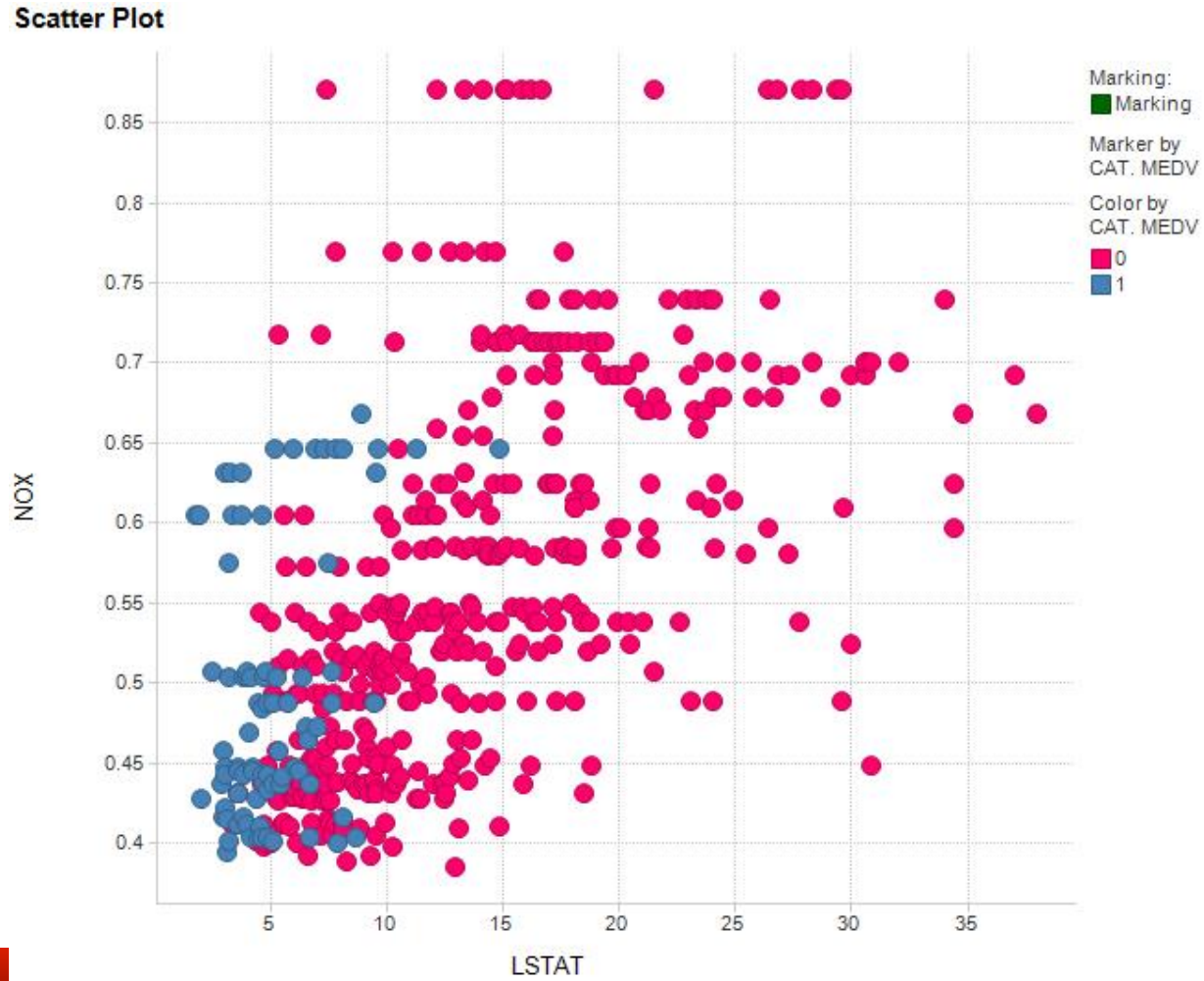
Scatterplot with Color Added

- Boston Housing example:

NOX vs. LSTAT

Red = low MV

Blue = high MV



Matrix Plot

- Shows scatterplots for variable pairs
- Example: scatterplots for 3 Boston Housing variables

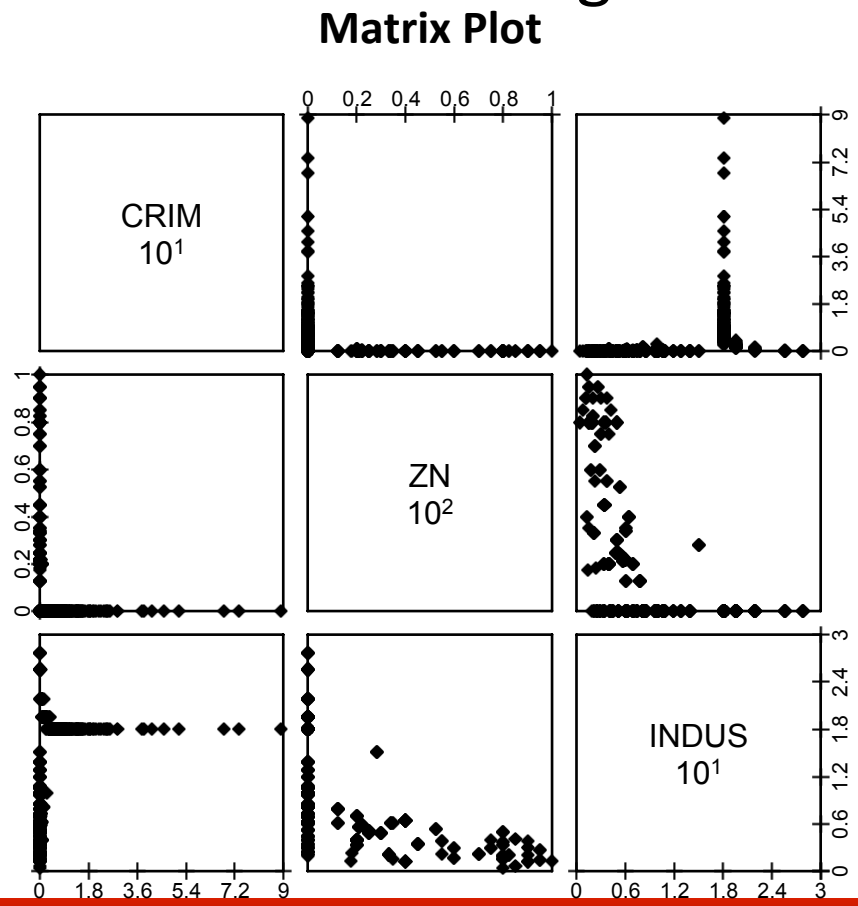
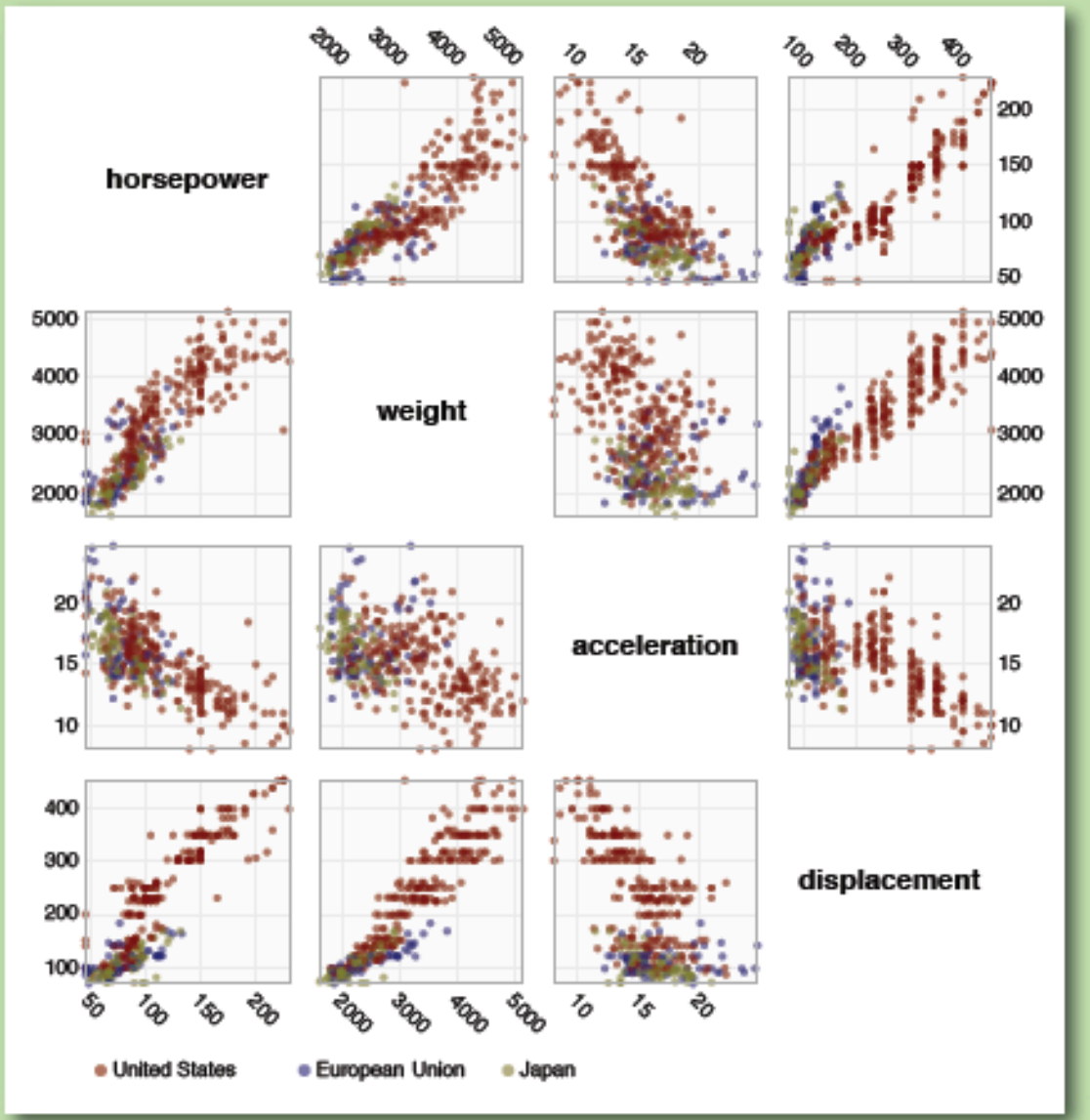


FIGURE 2C

Scatter Plot Matrix of Automobile Data



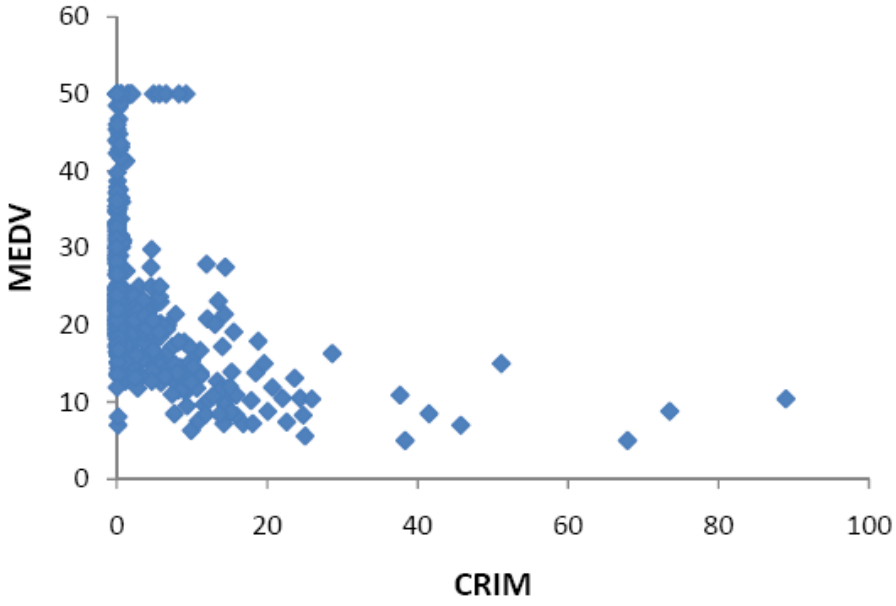
Source: GGobi
http://hc.ssb.stanford.edu/jlee/files/zoo/tx/stats/splom.html

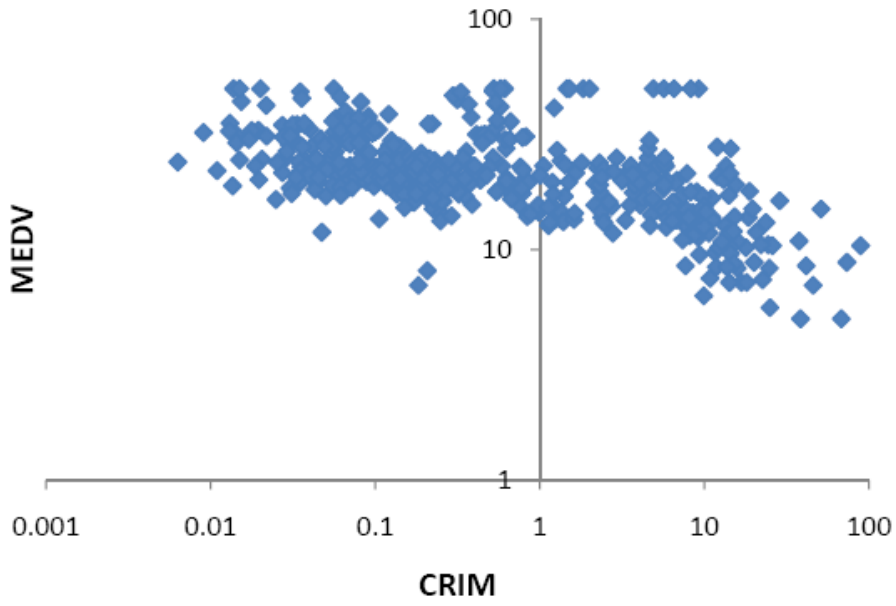
UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Logarithmic Scale

- A nonlinear scale used on large ranges of quantities
  - earthquake strength (Richter magnitude scale)
  - sound loudness (dB: decibel)
  - light intensity (cd: candela)
  - acidity or basicity of solutions (pH)
  - …
- It is based on orders of magnitude, rather than a standard linear scale
- Each mark on the scale is the previous mark multiplied by a value

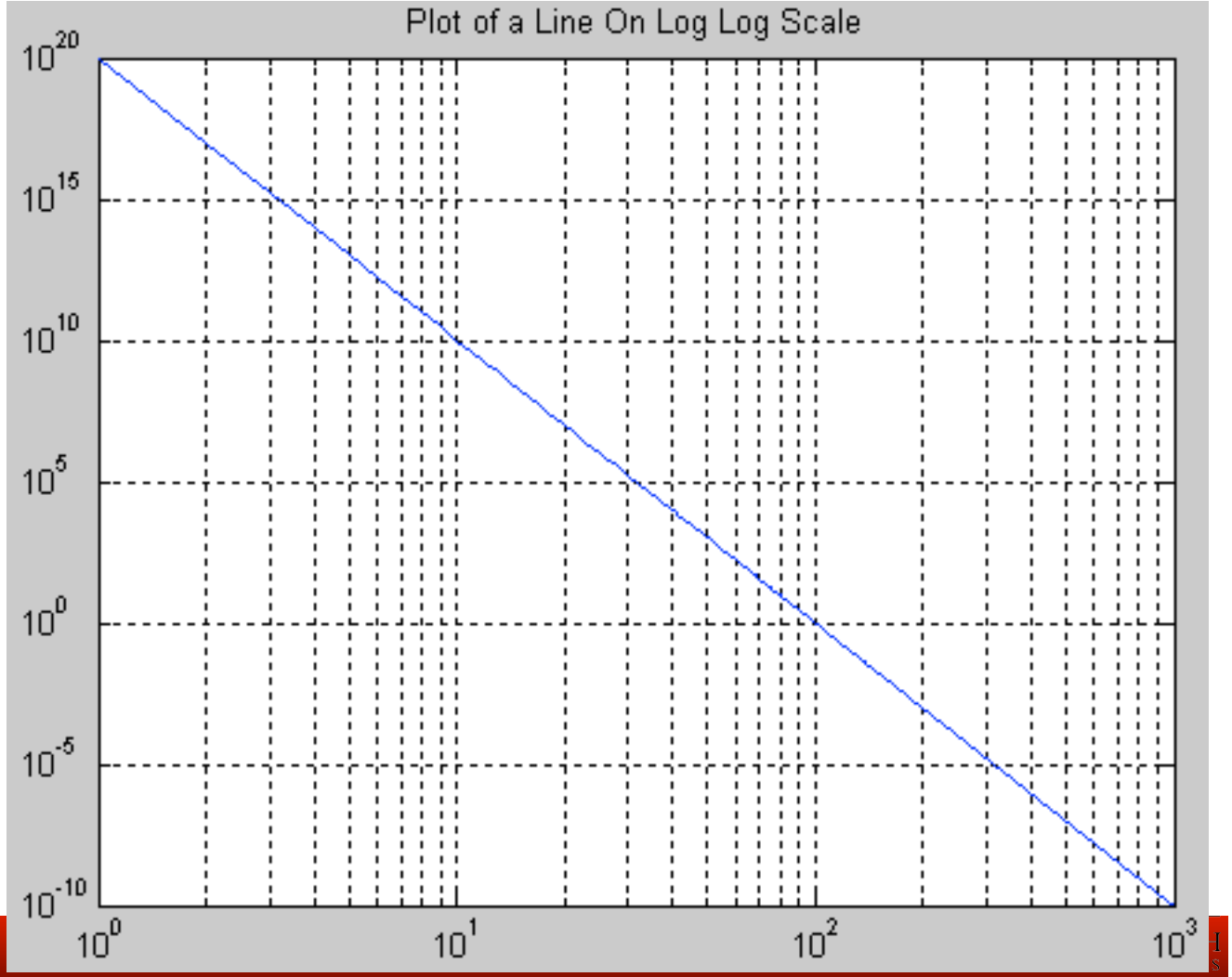
Rescale to log Scale (i.e., “uncrowd” the Data)





Power Law

- A straight line on a log-log plot is strong evidence for power-laws

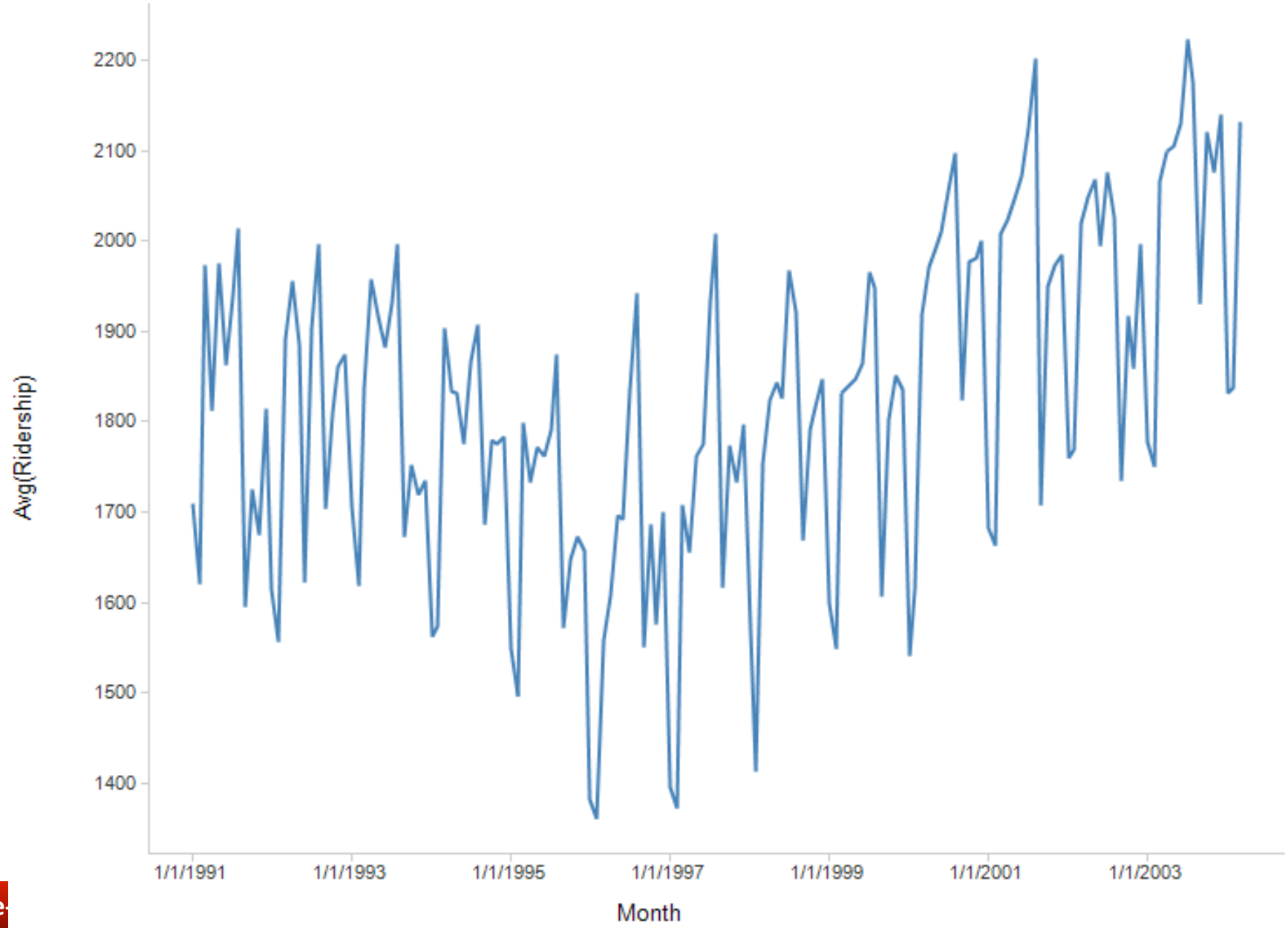


Aggregation

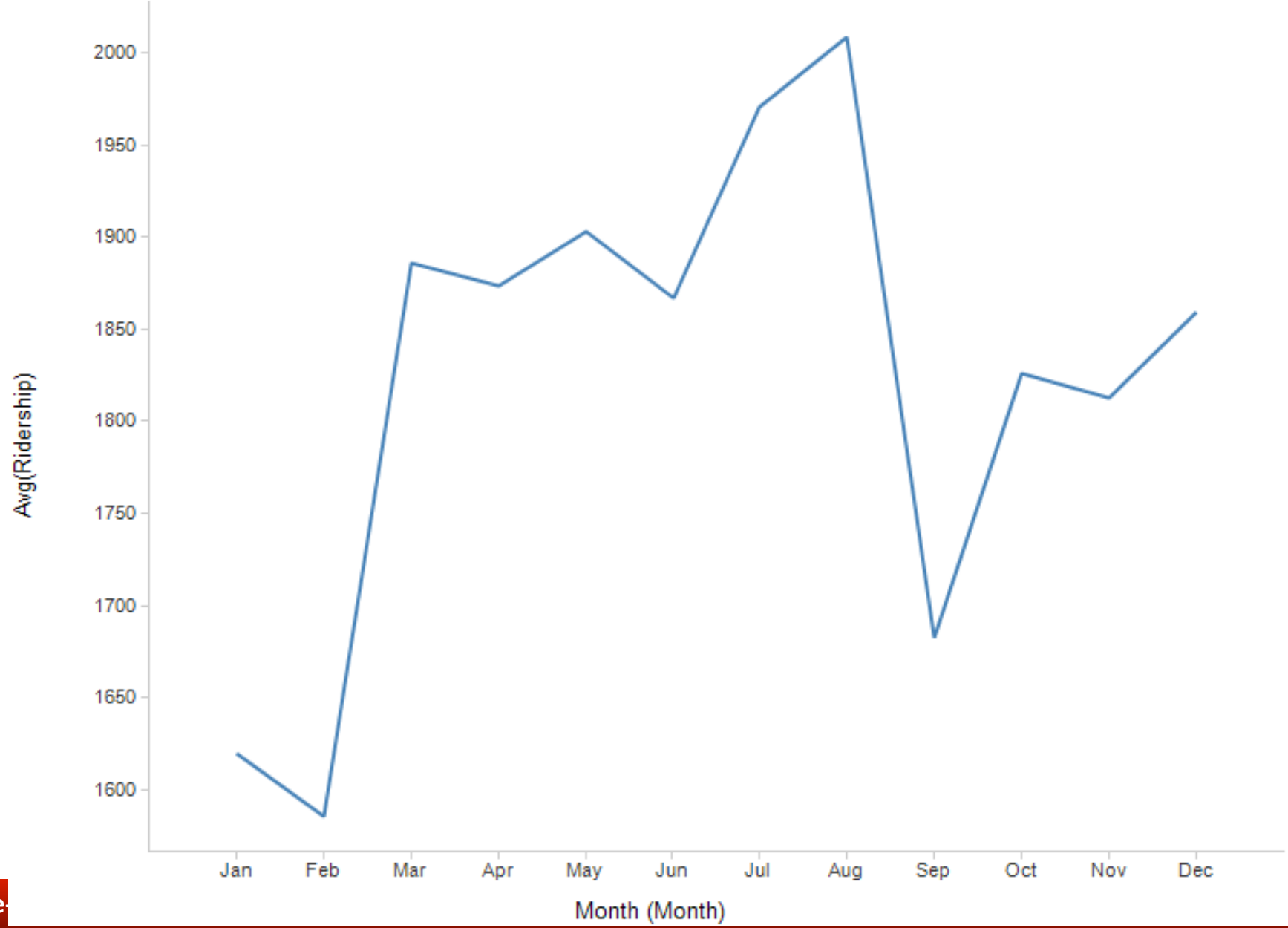
- Look at numerical data in an aggregated form:
  - count
  - summation
  - average (i.e. mean)
  - maximum
  - minimum
  - …
- e.g., 3,000 sales transactions from 50 products, to view the sum of sales for each product
- e.g., to aggregate over longer time periods

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Amtrak Ridership – Monthly Data



Aggregated: Monthly Average



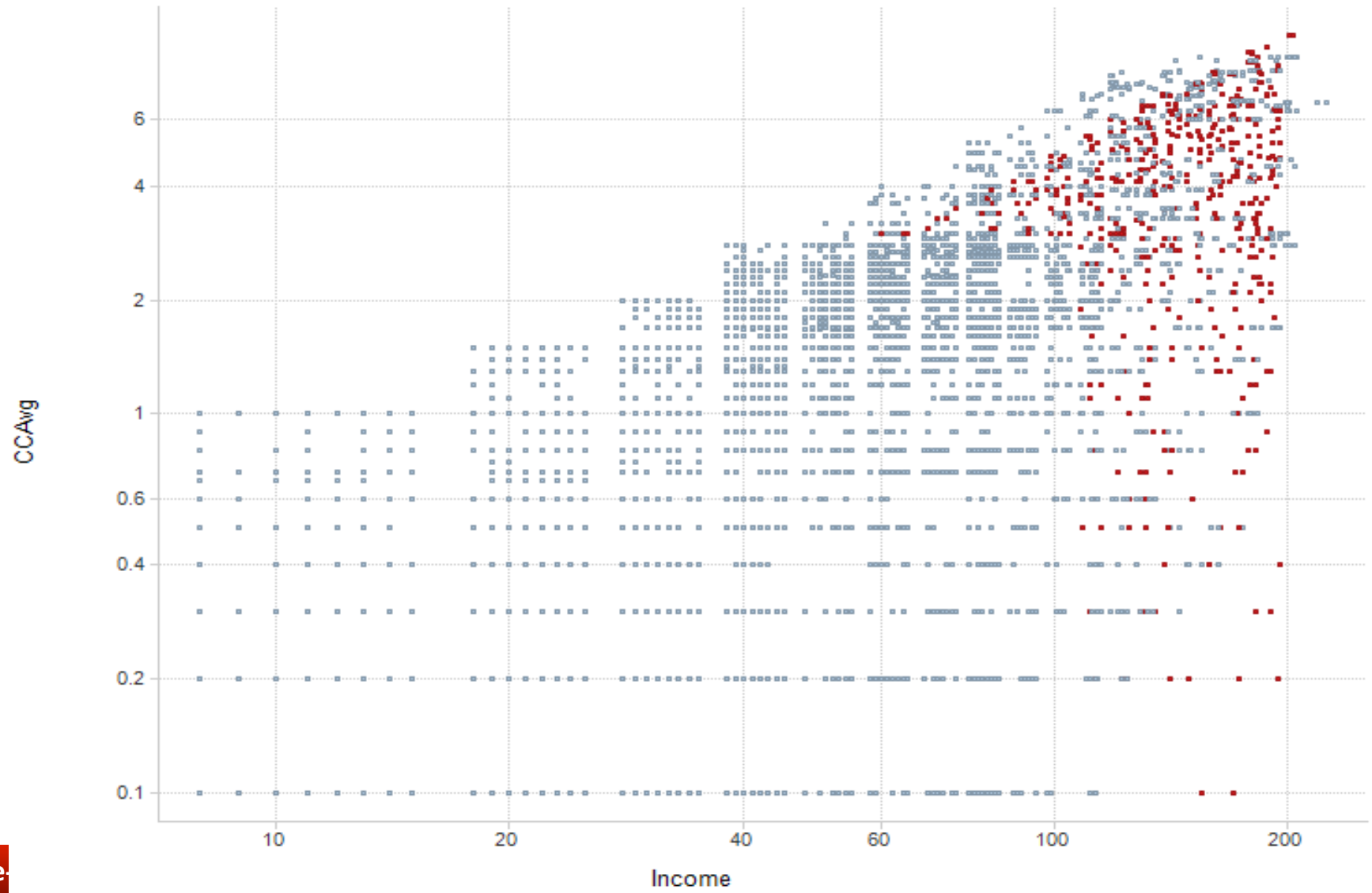
Aggregated: Annual Average

Jittering

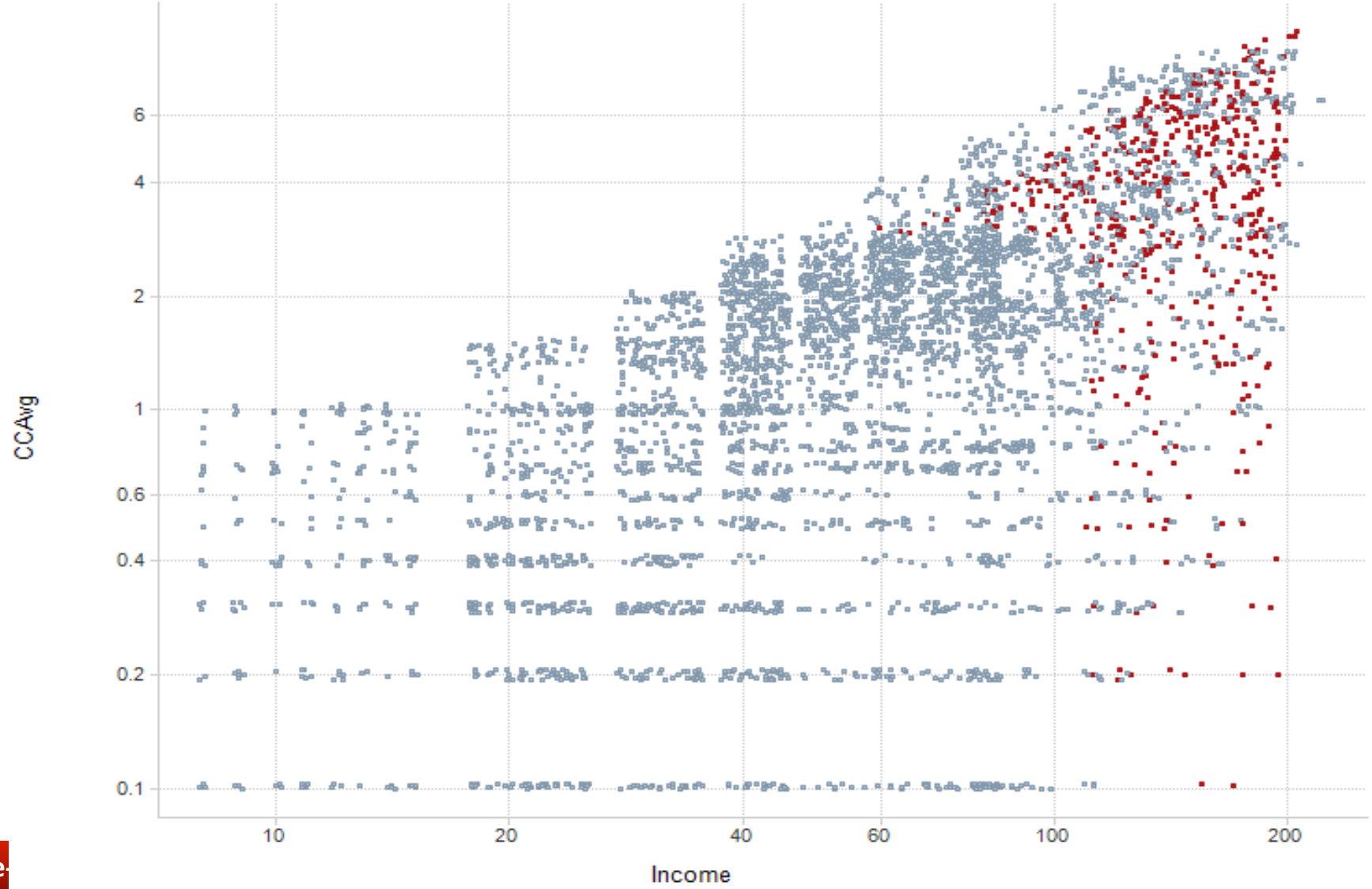
- Move markers by a small random amount
- Uncrowd the data by allowing more markers to be seen

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Without Jittering (for Comparison)



Scaling: Smaller Markers, Jittering, Color Contrast (Universal Bank; red = Accept Loan)



Scatter Plot with Labels (Utilities)

Parallel Coordinate Plot (Boston Housing)

CATMEDV =1

CATMEDV =0

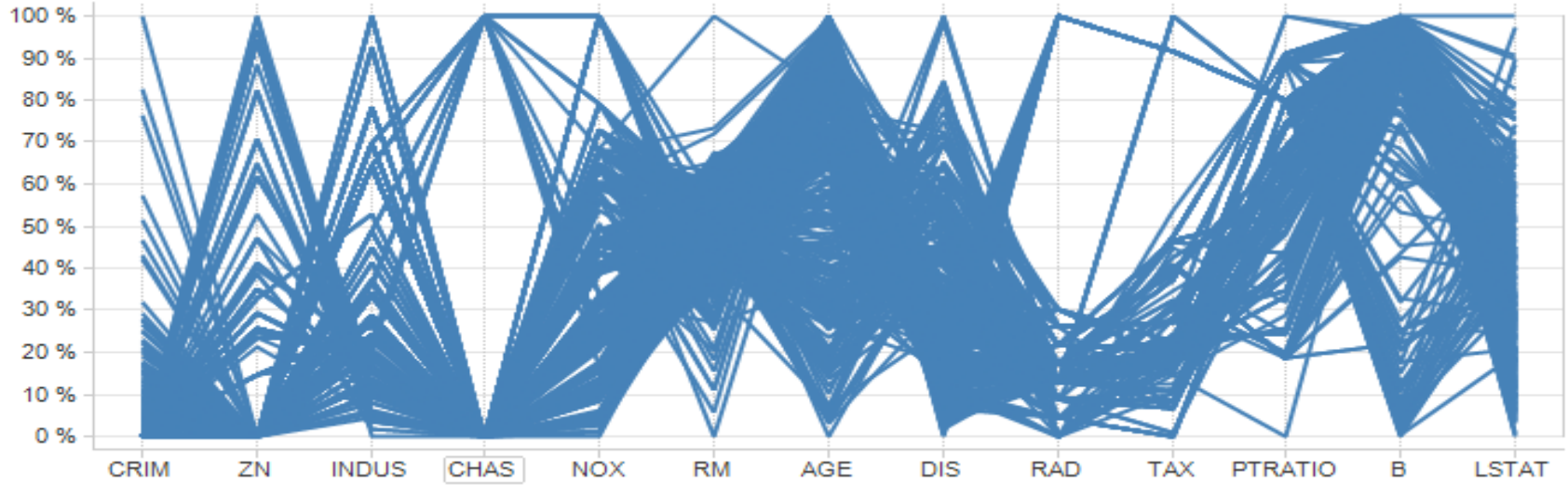
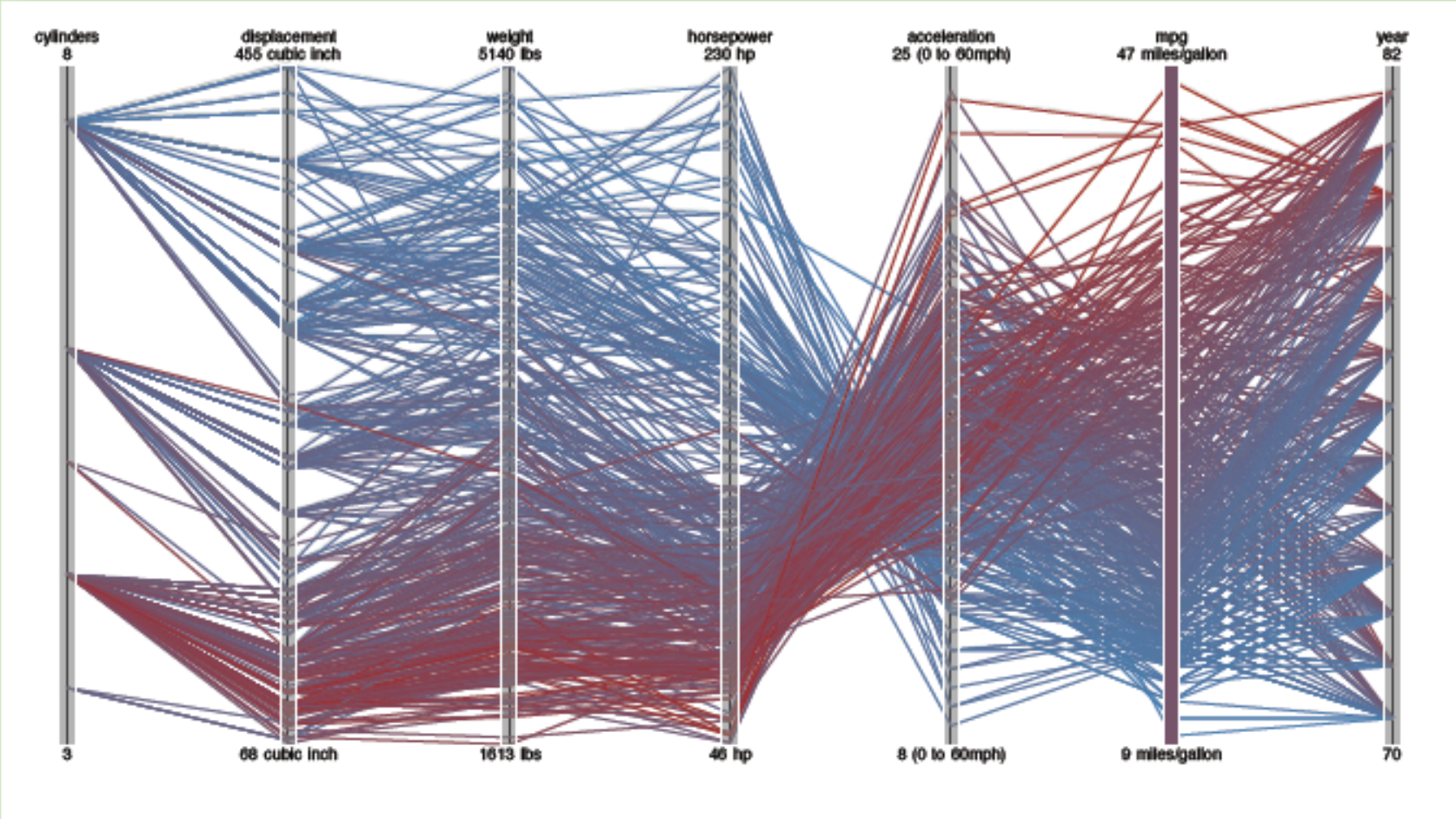


FIGURE 2D

Parallel Coordinates of Automobile Data



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

Specialized Visualizations

- Geographical data:
  - Map Charts
- Hierarchical data:
  - Treemaps
- Networked data:
  - Network Graphs

- Biomedical data:
  - sequence maps, structure maps, alignments, phylogenetic trees, networks, expressions, bench images, medical images, …

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Map Chart

- World maps comparing countries' well-being with GDP

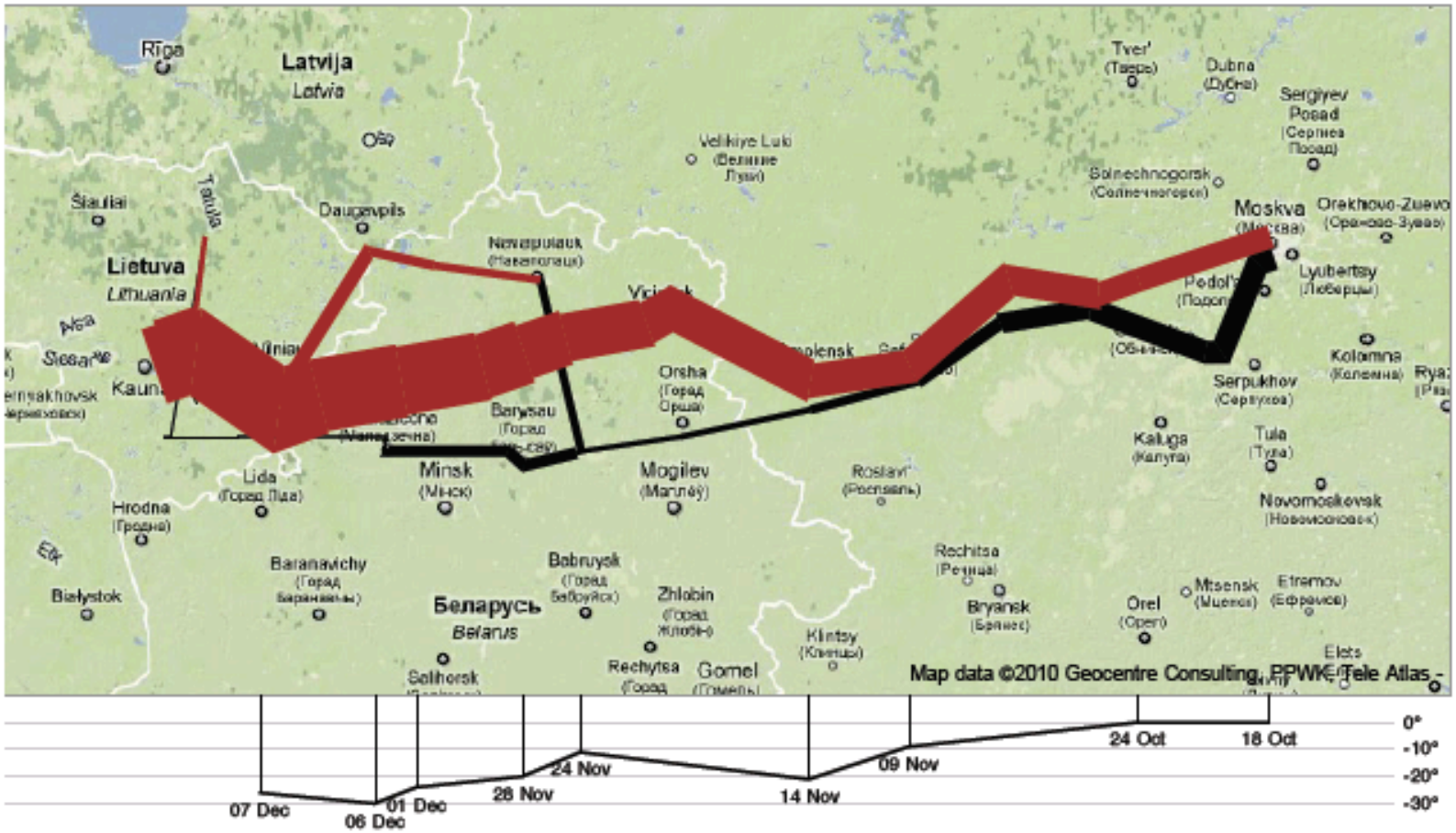
Well-Being Score

darker = higher value

GDP

FIGURE 3A

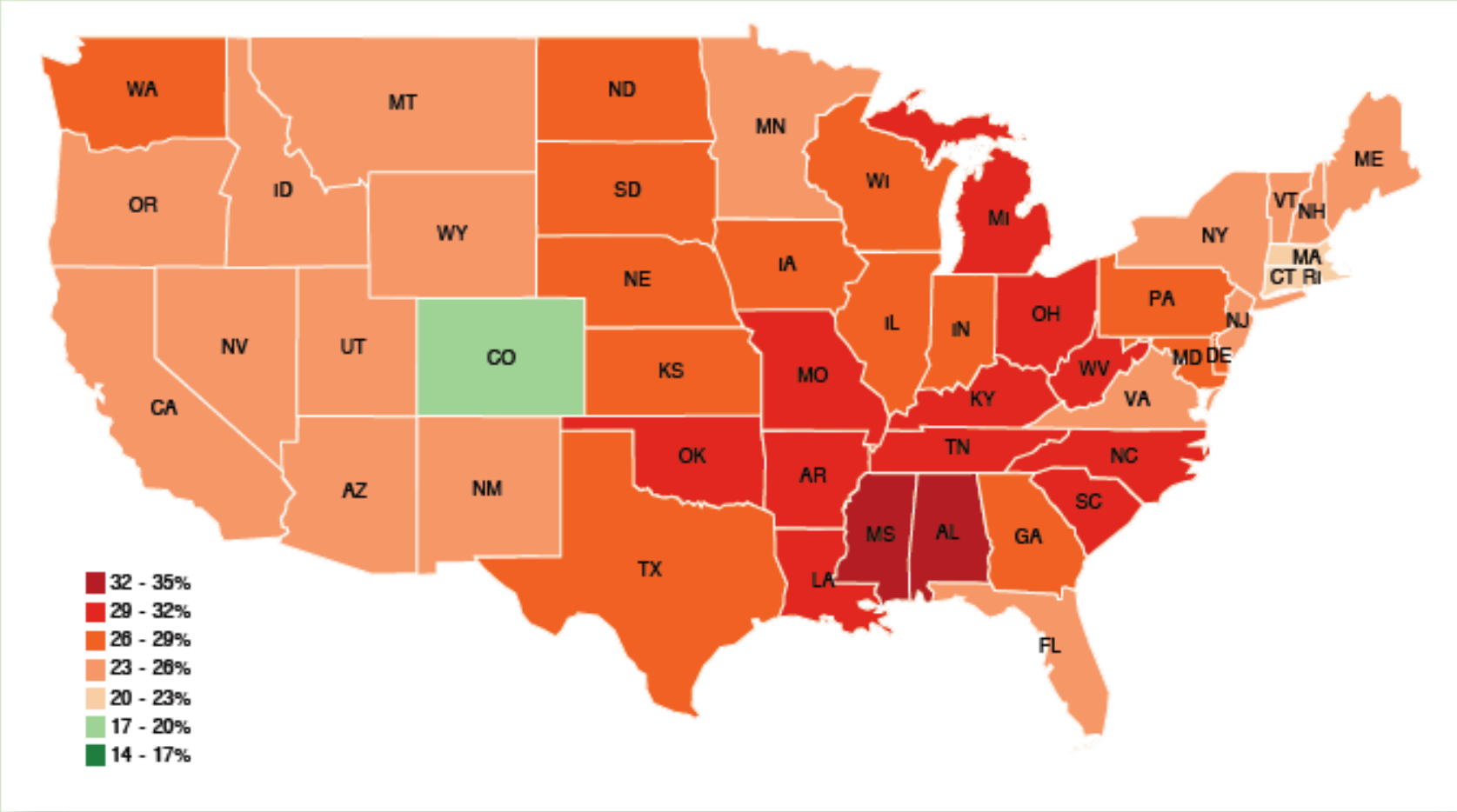
Flow Map of Napoleon's March on Moscow



Based on the Work of Charles Minard
http://hci.stanford.edu/jheer/files/zoo/ex/maps/napoleon.html

FIGURE 3B

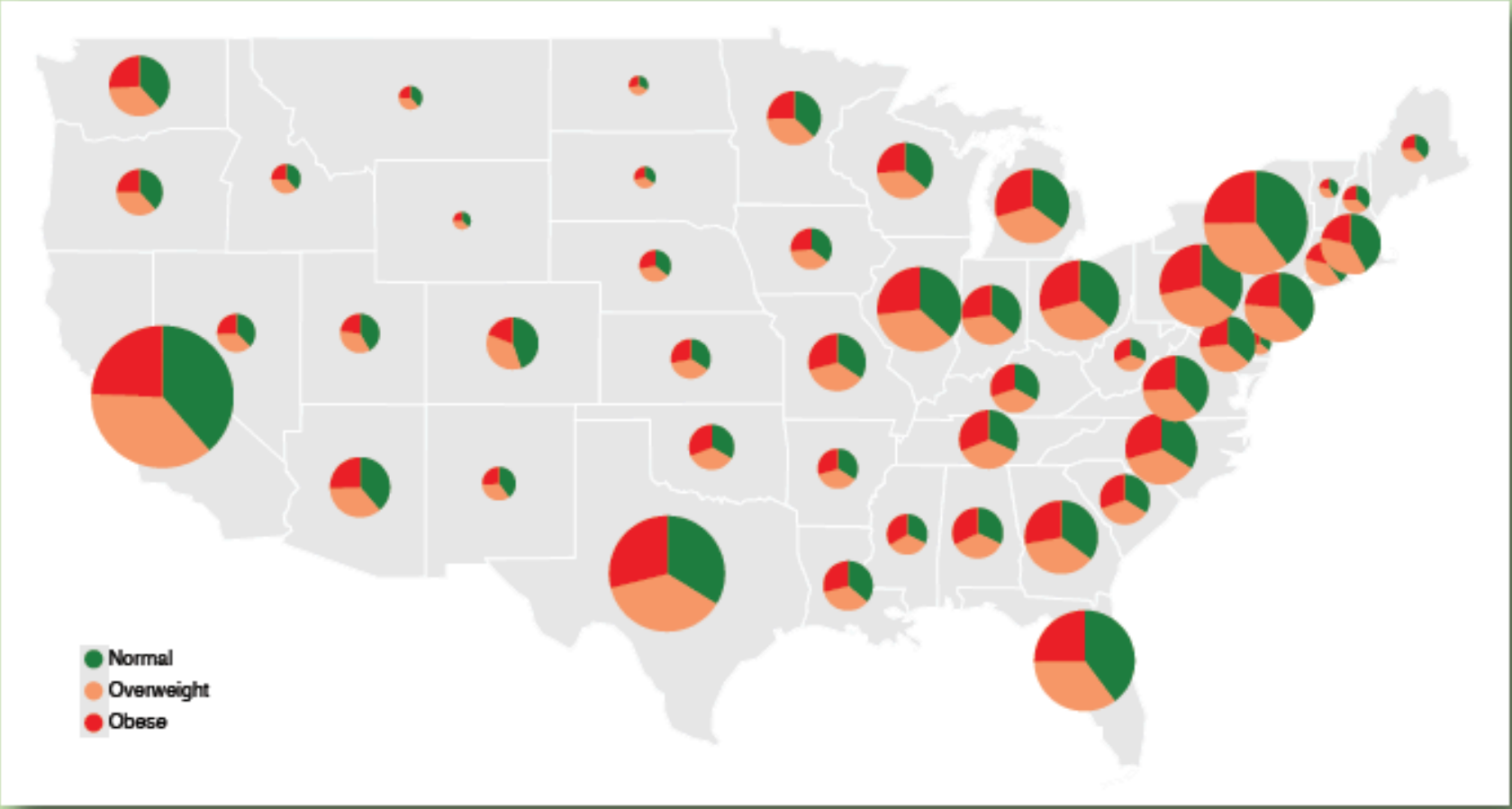
Choropleth Map of Obesity in the U.S., 2008



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

FIGURE 3C

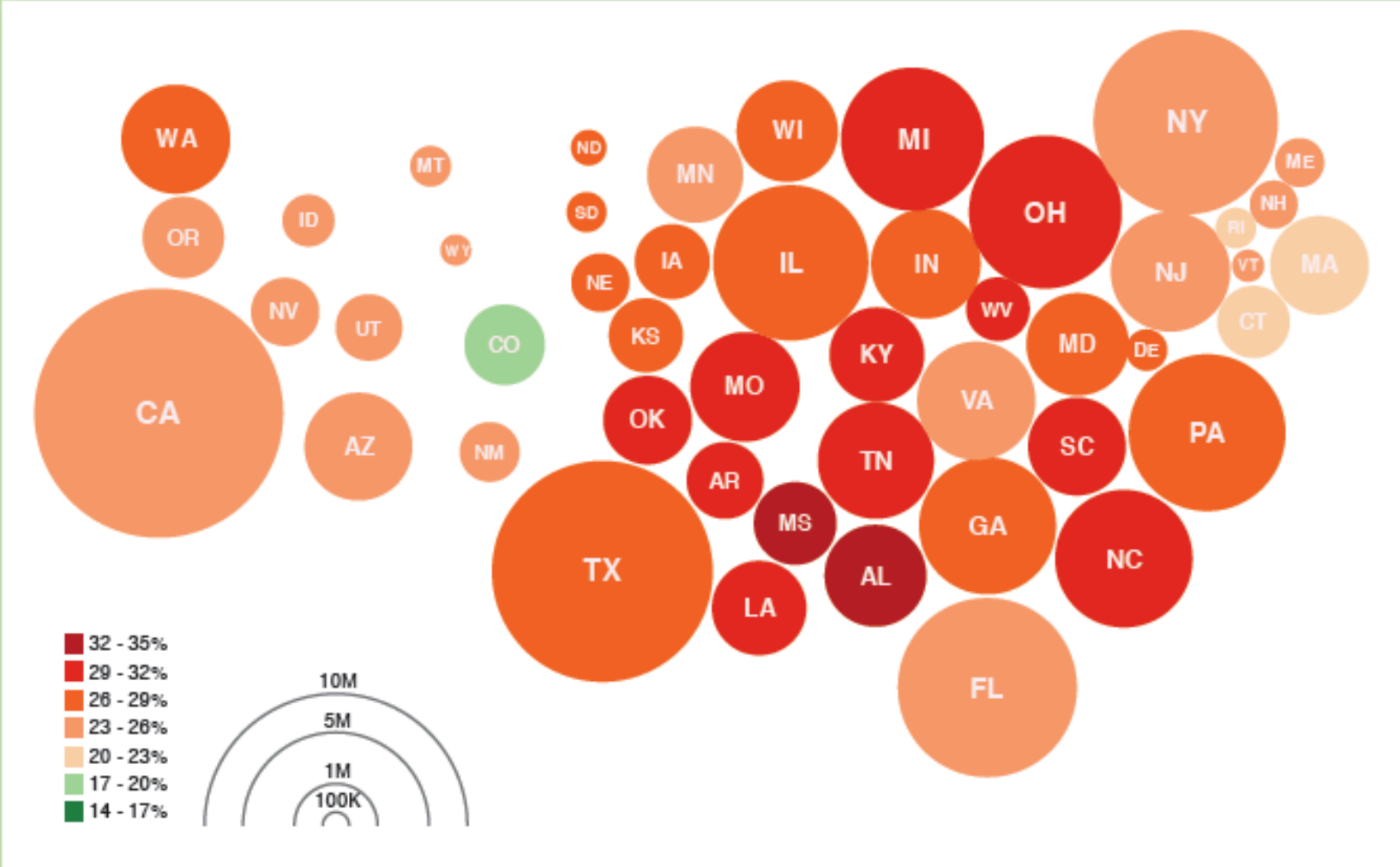
Graduated Symbol Map of Obesity in the U.S., 2008



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

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

Treemap

eBay Auctions:

- Hierarchical eBay data: Category → Sub-category → Brand
- Rectangle size = average closing price
- Color = % sellers w/ negative feedbacks (darker=more)

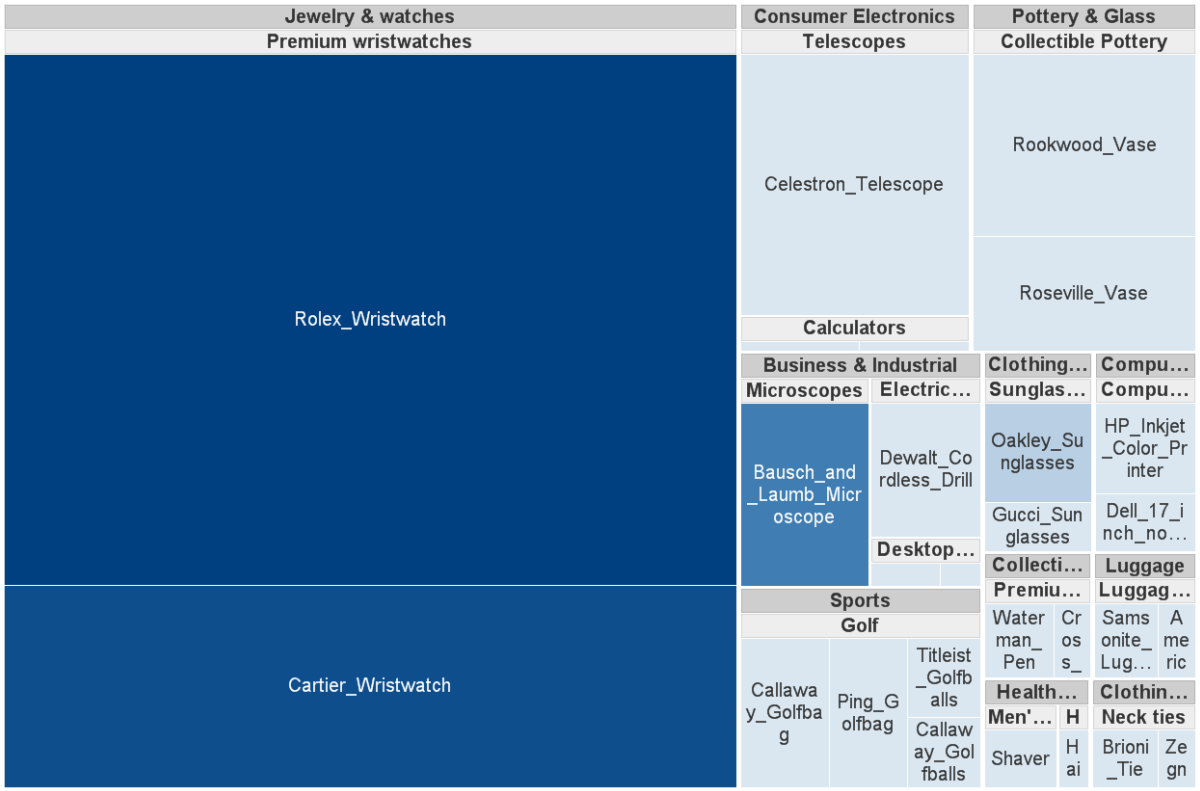
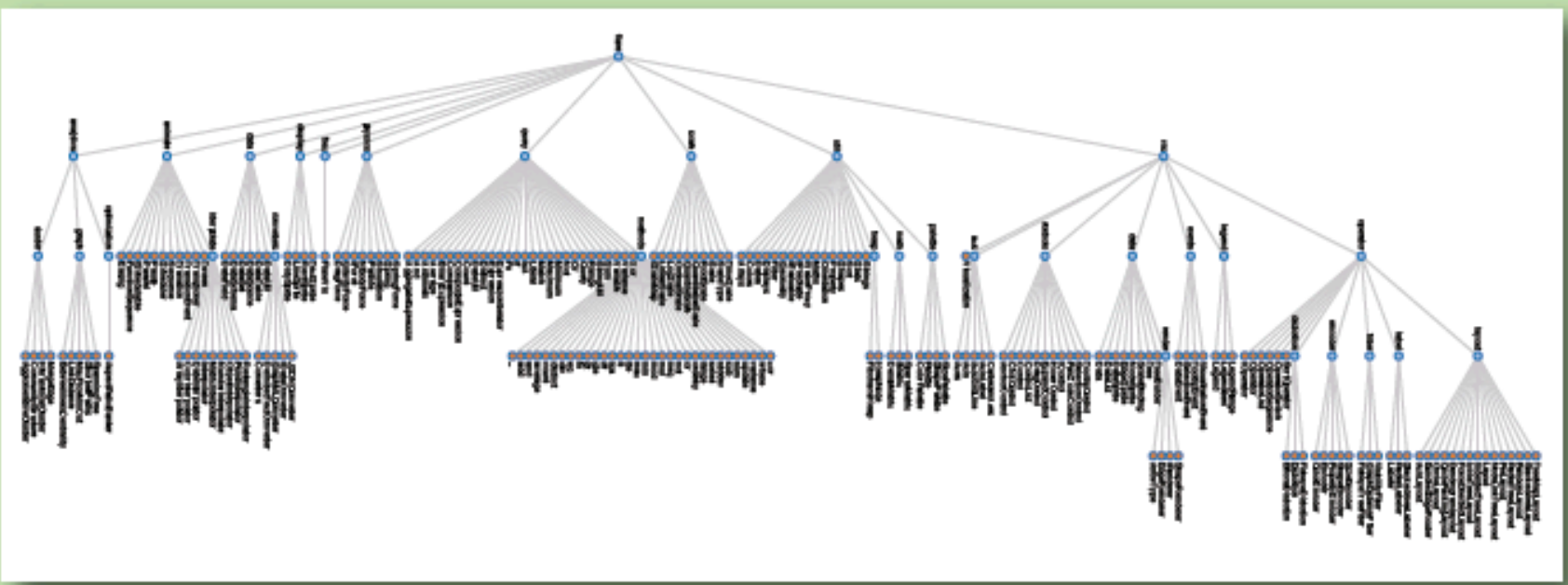


FIGURE 4A

Radial Node-link Diagram of the Flare Package Hierarchy



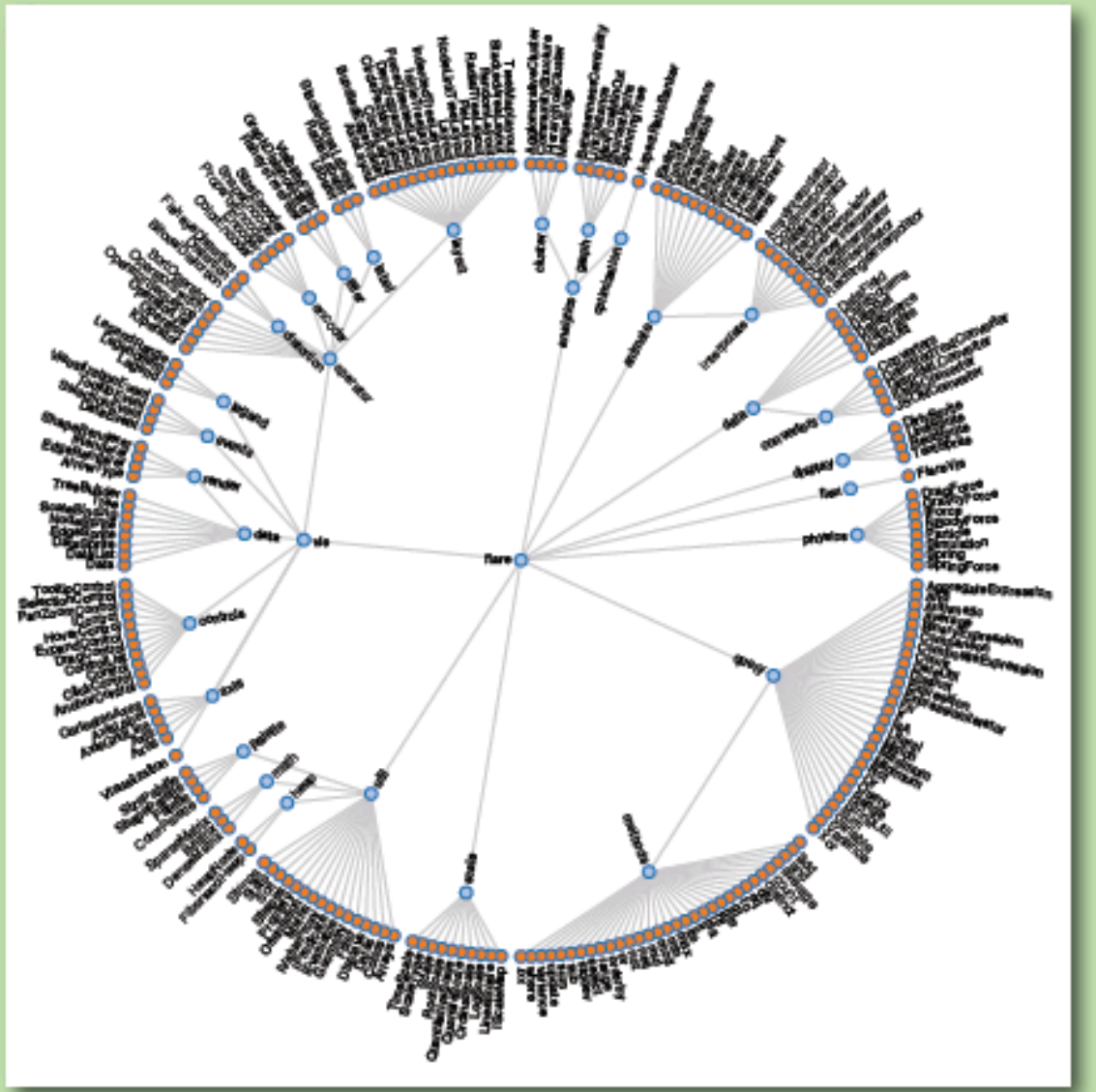
Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/tree.html

UNIVERSITY OF MARYLAND

ROBERT H. SMITH SCHOOL OF BUSINESS

FIGURE 4B

Cartesian Node-link Diagram of the Flare Package Hierarchy



Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/cluster-radial.html

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

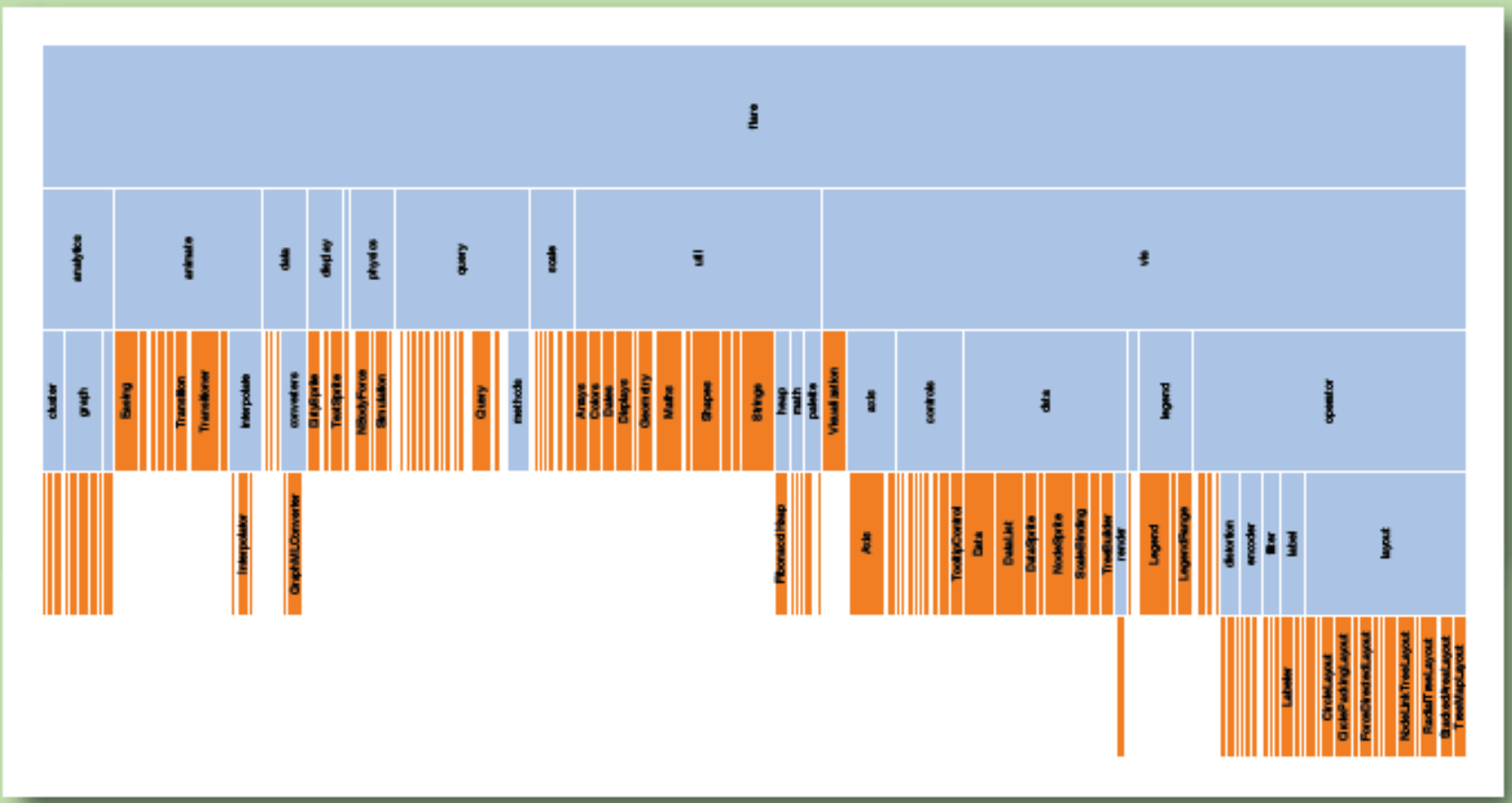
FIGURE 4C

Indented Tree Layout of the Flare Package Hierarchy

flare	933KB
analytics	47KB
cluster	14KB
graph	25KB
optimization	6KB
animate	97KB
data	29KB
display	23KB
flex	4KB
physics	29KB
DragForce	1KB
GravityForce	1KB
IForce	0KB
NBodyForce	10KB
Particle	2KB
Simulation	9KB
Spring	2KB
SpringForce	1KB
query	87KB
scale	30KB
util	161KB
vis	422KB
Visualization	16KB
axis	33KB
controls	43KB
data	107KB
events	6KB
legend	35KB

FIGURE 4D

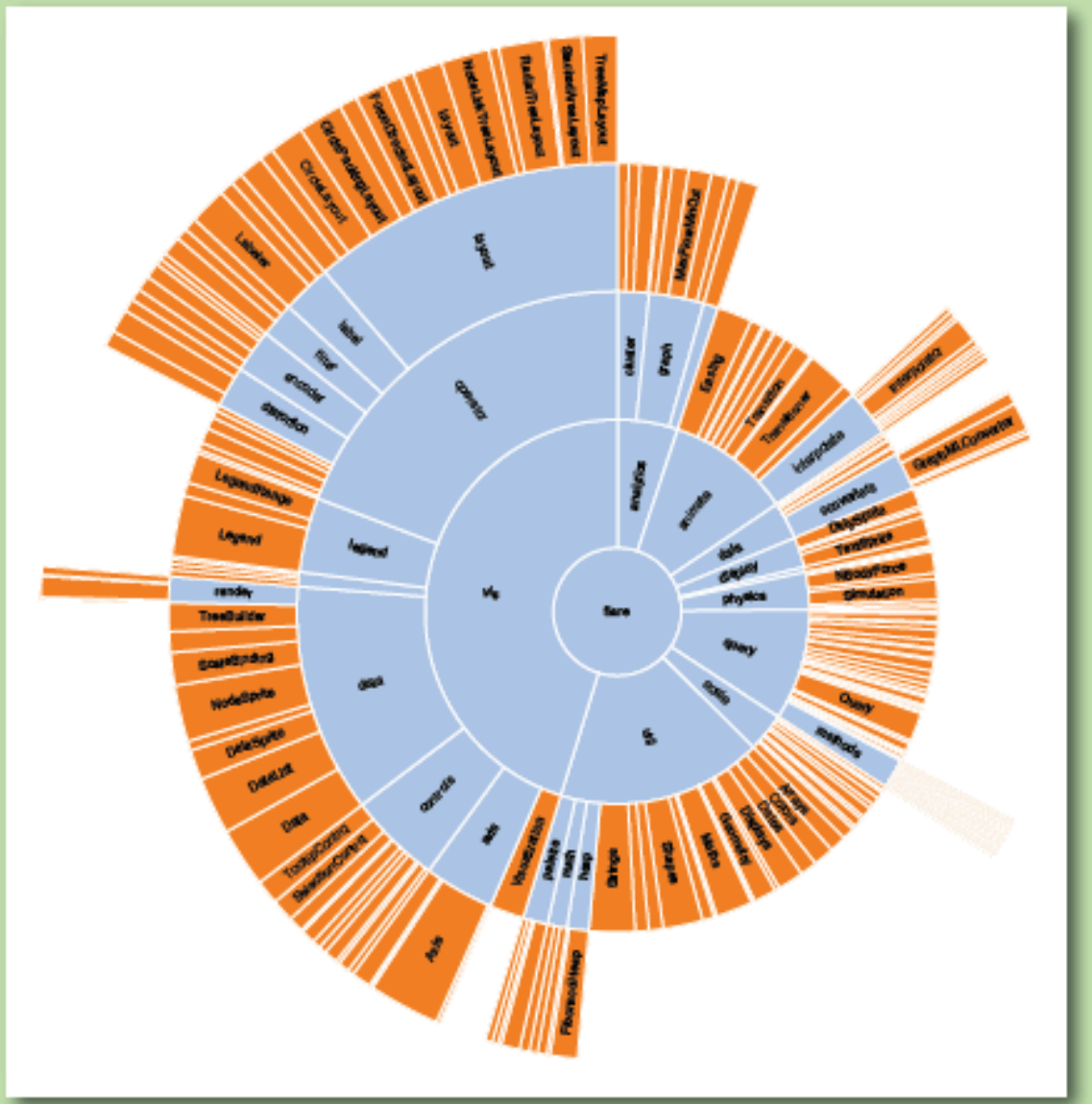
Icicle Tree Layout of the Flare Package Hierarchy



Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/icicle.html

FIGURE 4E

Sunburst (Radial Space-filling) Layout of the Flare Package Hierarchy

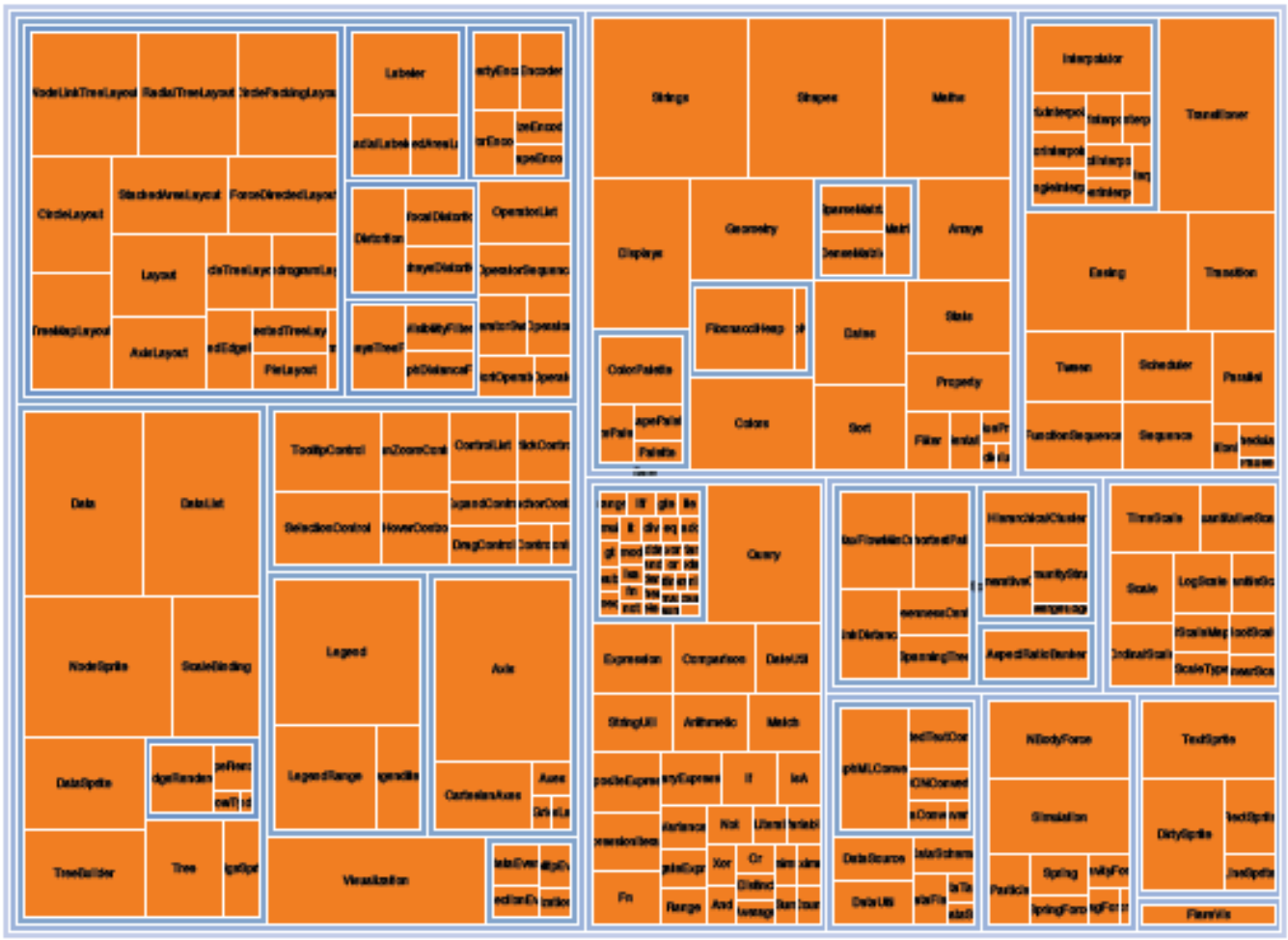


Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/sunburst.html

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

FIGURE 4F

Treemap Layout of the Flare Package Hierarchy

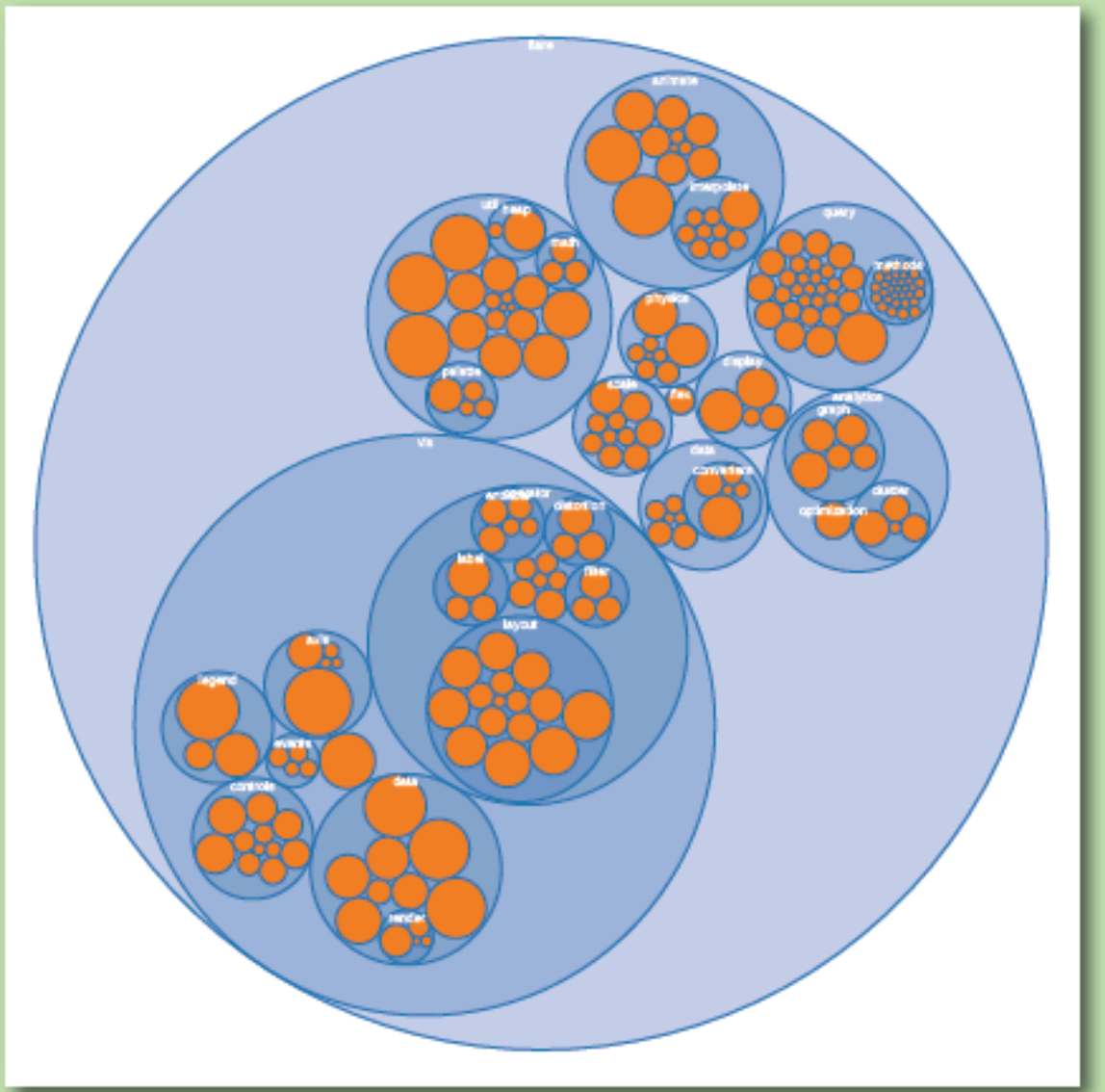


Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/treemap.html

UNIVERSITY OF MARYLAND
R.H. SMITH SCHOOL OF BUSINESS

FIGURE 4G

Nested Circles Layout of the Flare Package Hierarchy



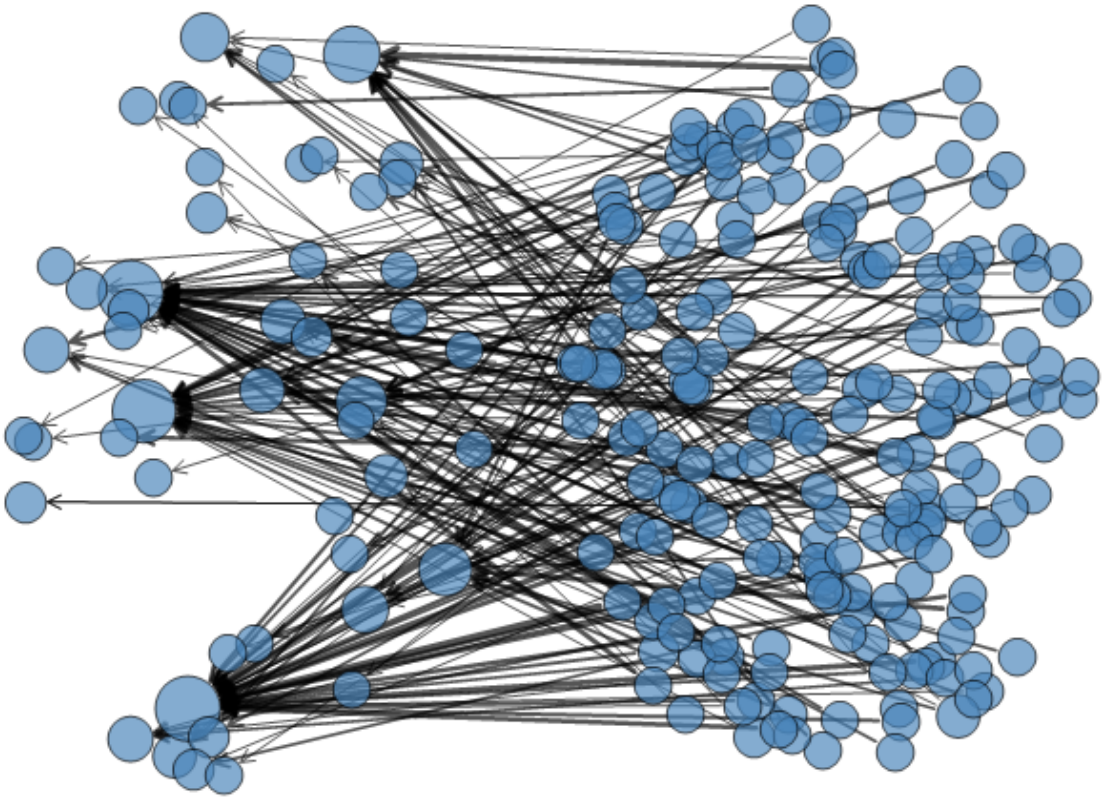
Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/pack.html

UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Network Graph

eBay Auctions:

- sellers on left
- buyers on right
- Circle size = # of transactions
- Line width = # of auctions
- Arrows point from buyer to seller



UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

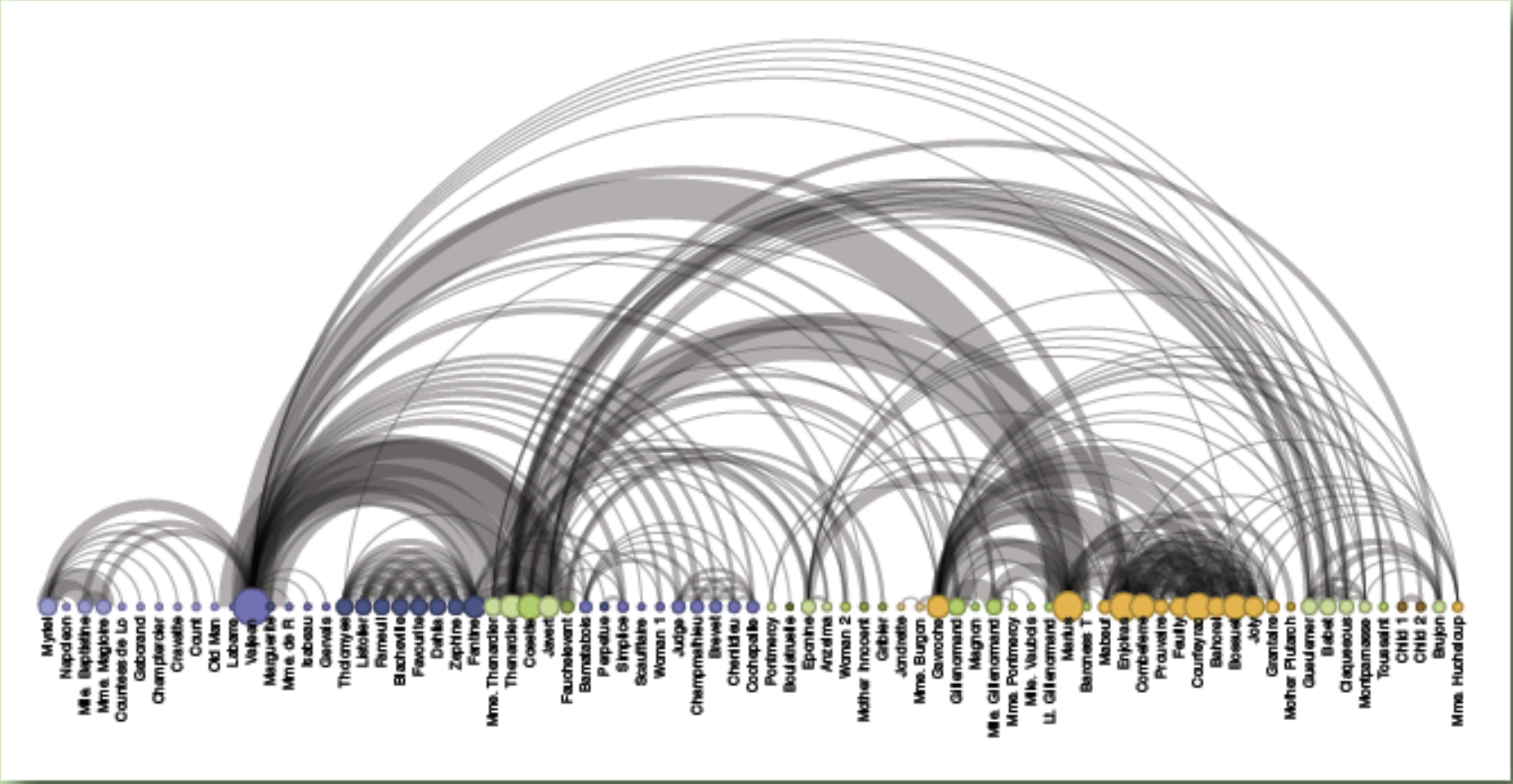
FIGURE 5A

Force-directed Layout of *Les Misérables* Character Co-occurences

Source: Knuth, D. E. 1993. *The Stanford GraphBase: A Platform for Combinatorial Computing*, Addison-Wesley.
http://hci.stanford.edu/jheer/files/zoo/ex/networks/force.html

FIGURE 5B

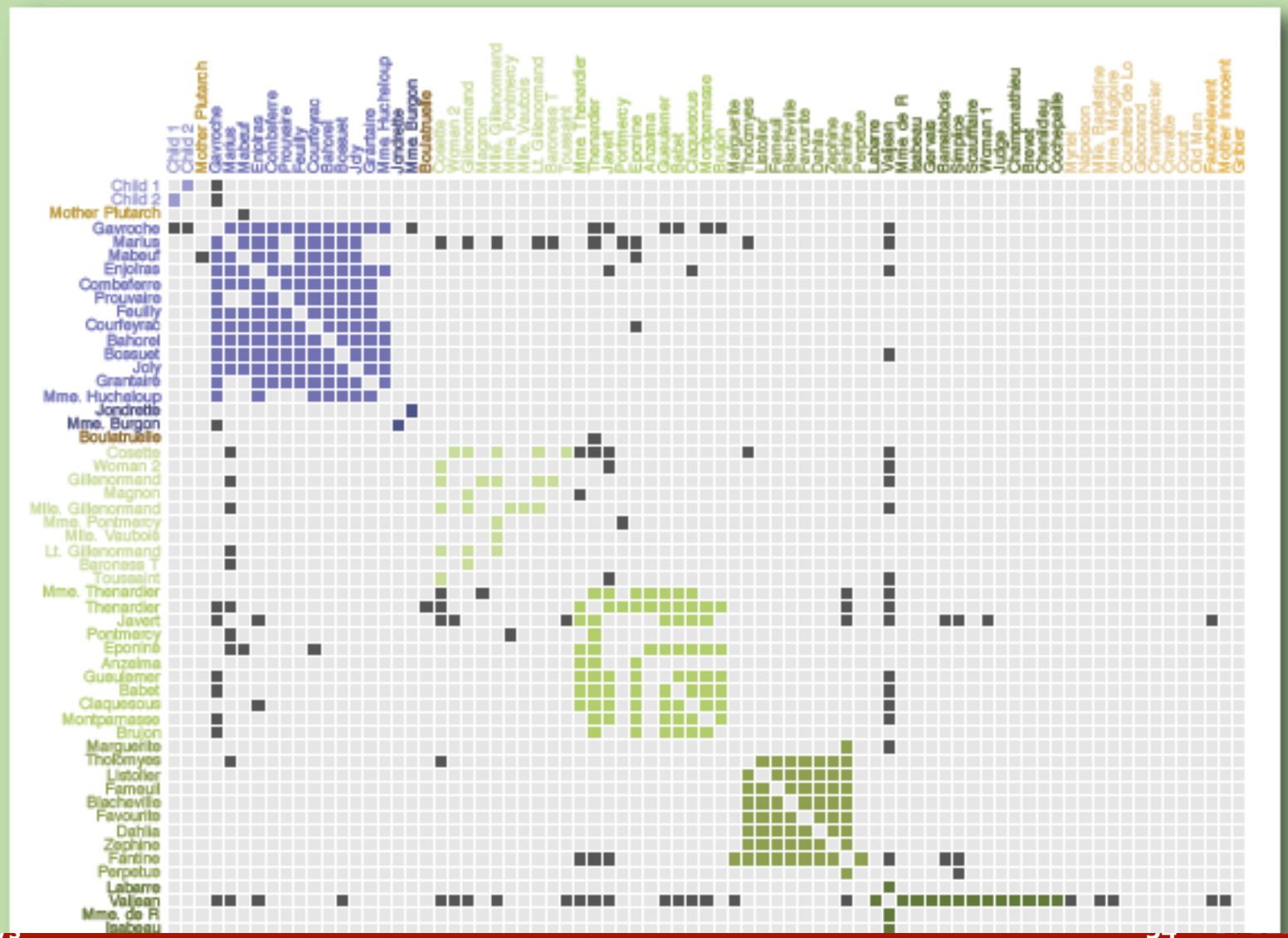
Arc Diagram of *Les Misérables* Character Co-occurences



Source: Knuth, D. E. 1993. *The Stanford GraphBase: A Platform for Combinatorial Computing*, Addison-Wesley.
http://hci.stanford.edu/jheer/files/zoo/ex/networks/arc.html

FIGURE 5C

Matrix View of *Les Misérables* Character Co-occurences



Graph/Network Data

- Lots of data/dataset has a graph/network structure:
  - Social networks
  - Communication networks
  - Computer networks
  - Road networks
  - Citation networks
  - Collaborations/Relationships
  - …

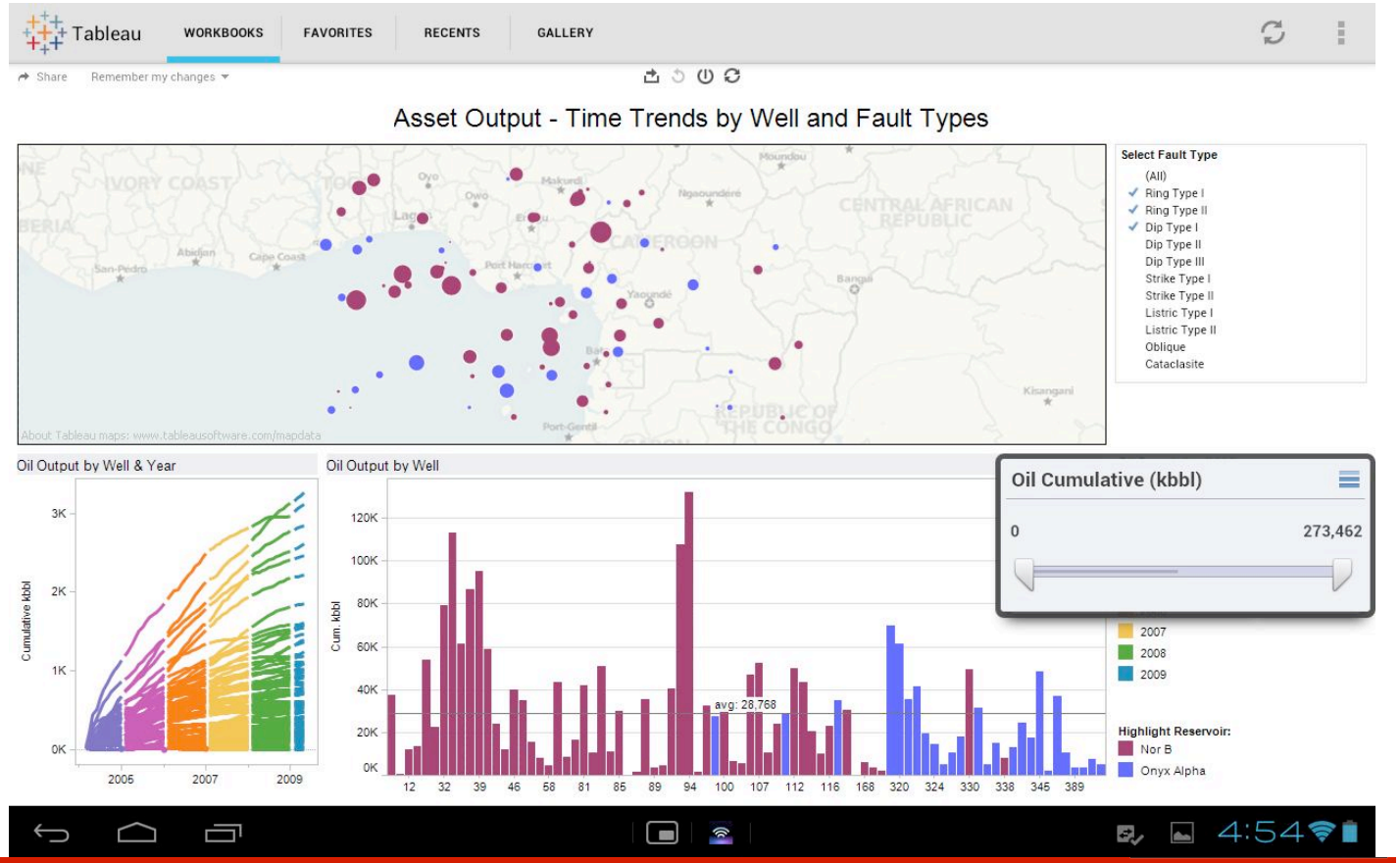
UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

Data Visualization Tools

Tool	Vendor	Platform	First release	Latest release
BusinessObjects	SAP	Cloud	1990	2009
Cognos Analytics	IBM	Cloud	2005	2020
Dash	Plotly	Python / JavaScript / Cloud	2012	2018
FusionCharts	Idera	JavaScript / Cloud	2002	2014
MicroStrategy	MicroStrategy	Cloud	1989	2019
Power BI	Microsoft	Cloud	2013	2019
Qlik View	Qlik	Cloud	1993	2018
Sisense	Sisense	Cloud	2004	2018
Tableau	Tableau	Desktop / Server / Cloud	2003	2020
Visual Analytics	SAS	Windows / Linux	2002	2013

Tableau

- Line Chart
- Bar Chart
- Scatter Plot
- Pie Chart
- Histogram
- Box Plot
- Heat Map
- Map View
- Treemap
- Gantt Chart
- Packed Bubble Chart
- Table



ROBERT H. SMITH SCHOOL OF BUSINESS