

Information Visualization – An Introduction and Survey



Last update August 2016

Outline



- How InfoVis relates to big data and data/business analytics
- Compelling InfoVis examples - throughout
- Basic data organizations & data types
- How visualize basic data organizations
- Basic InfoVis interactions
- InfoVis software

Information Visualization



- **Information Presentation** (aka InfoGraphics)

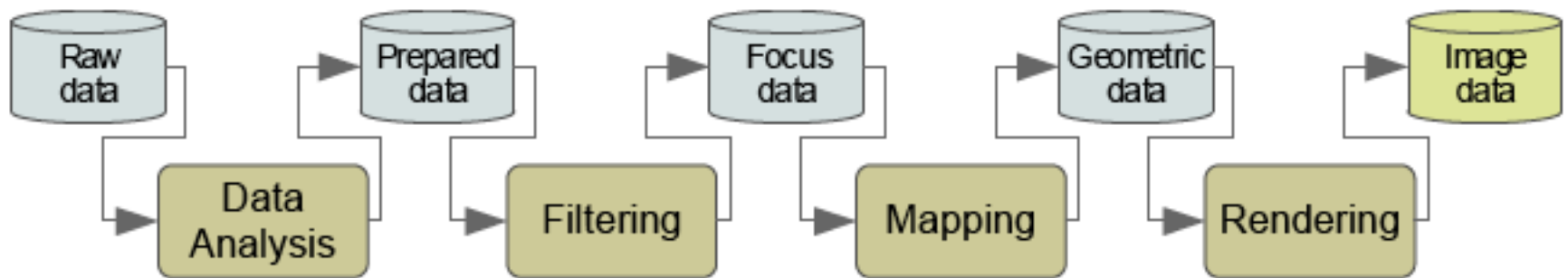
PLUS

- **Interaction** with the Information Presentation

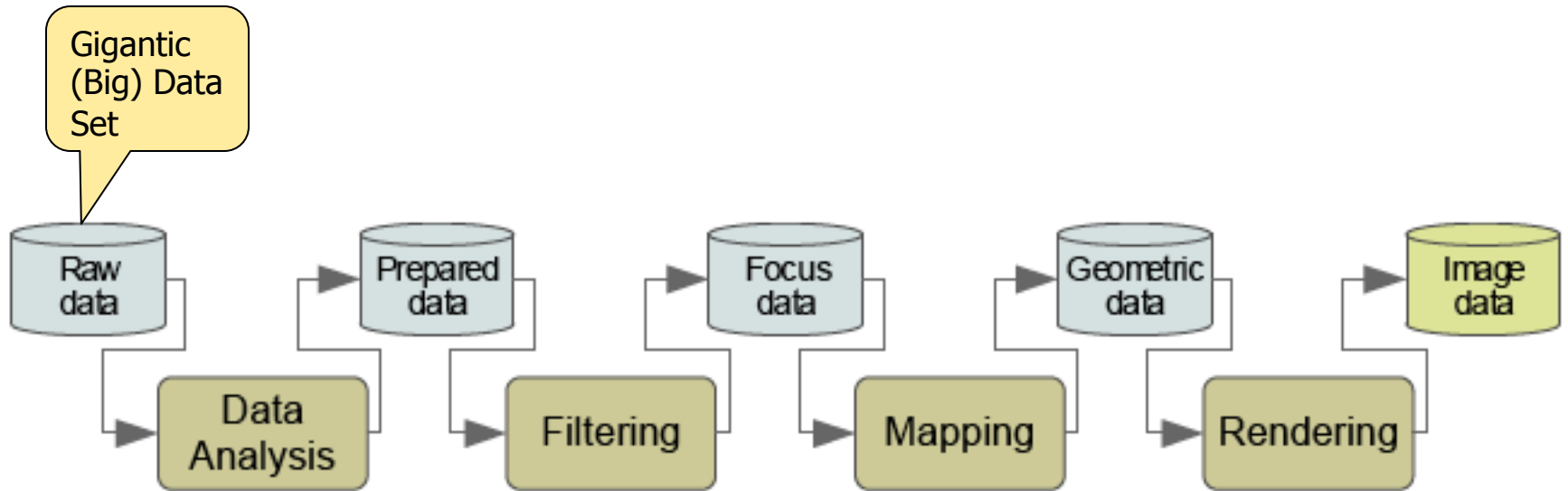
EQUALS

- **Information Visualization**

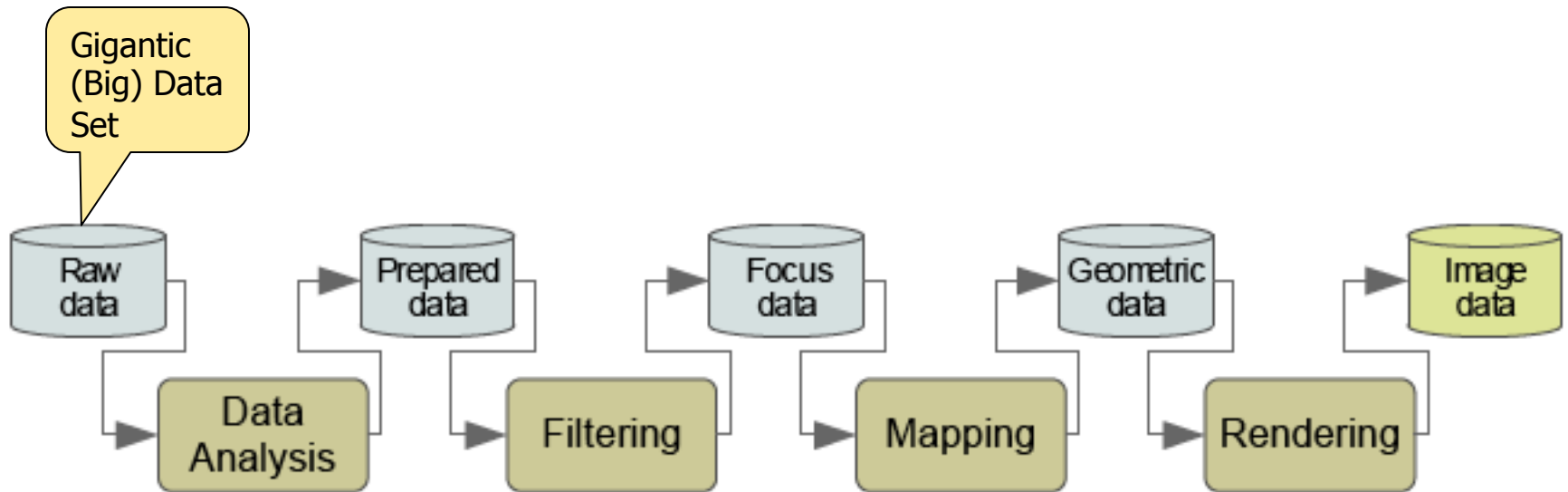
InfoVis and “Big Data”



InfoVis and "Big Data"



InfoVis and "Big Data"

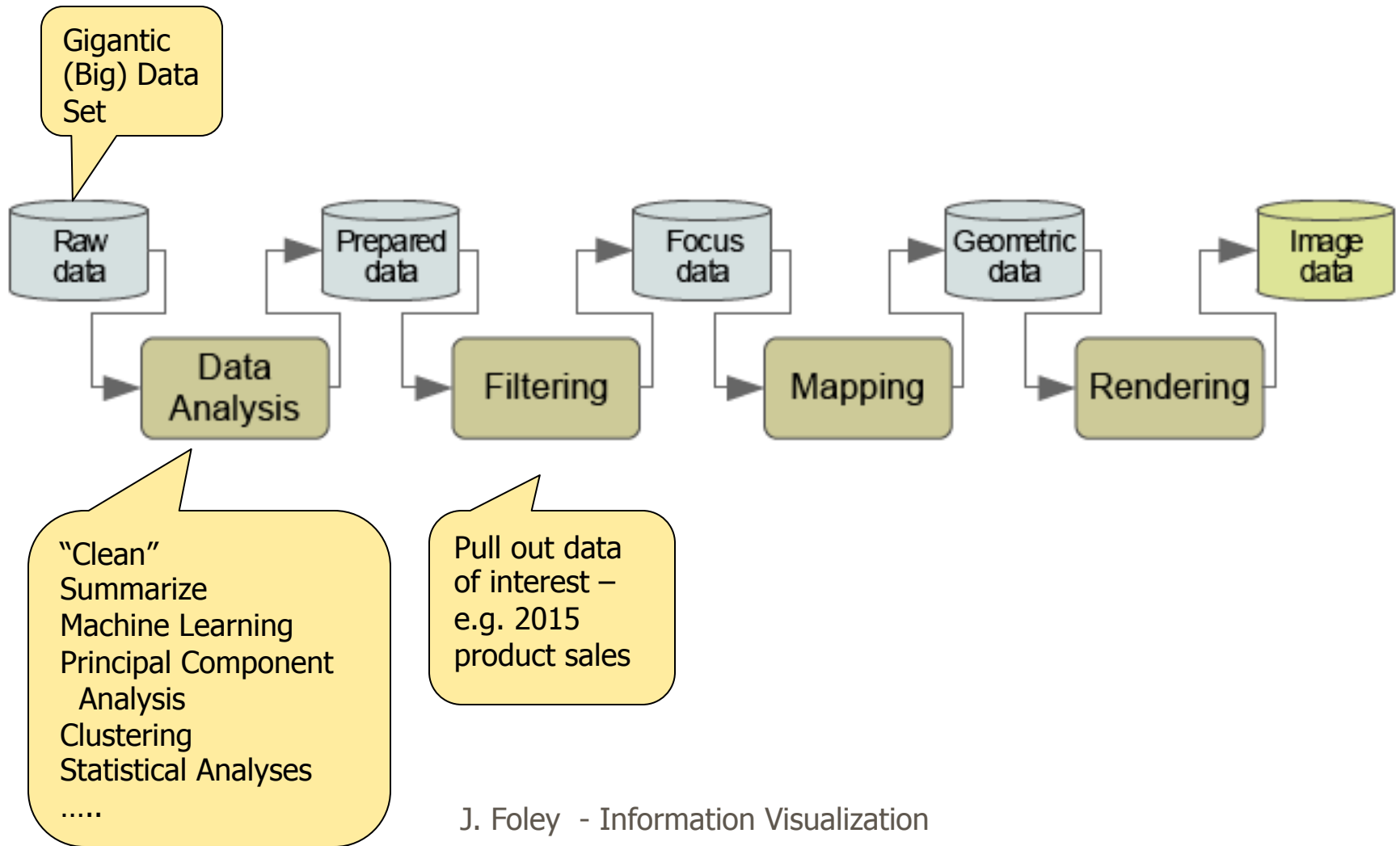


Gigantic
(Big) Data
Set

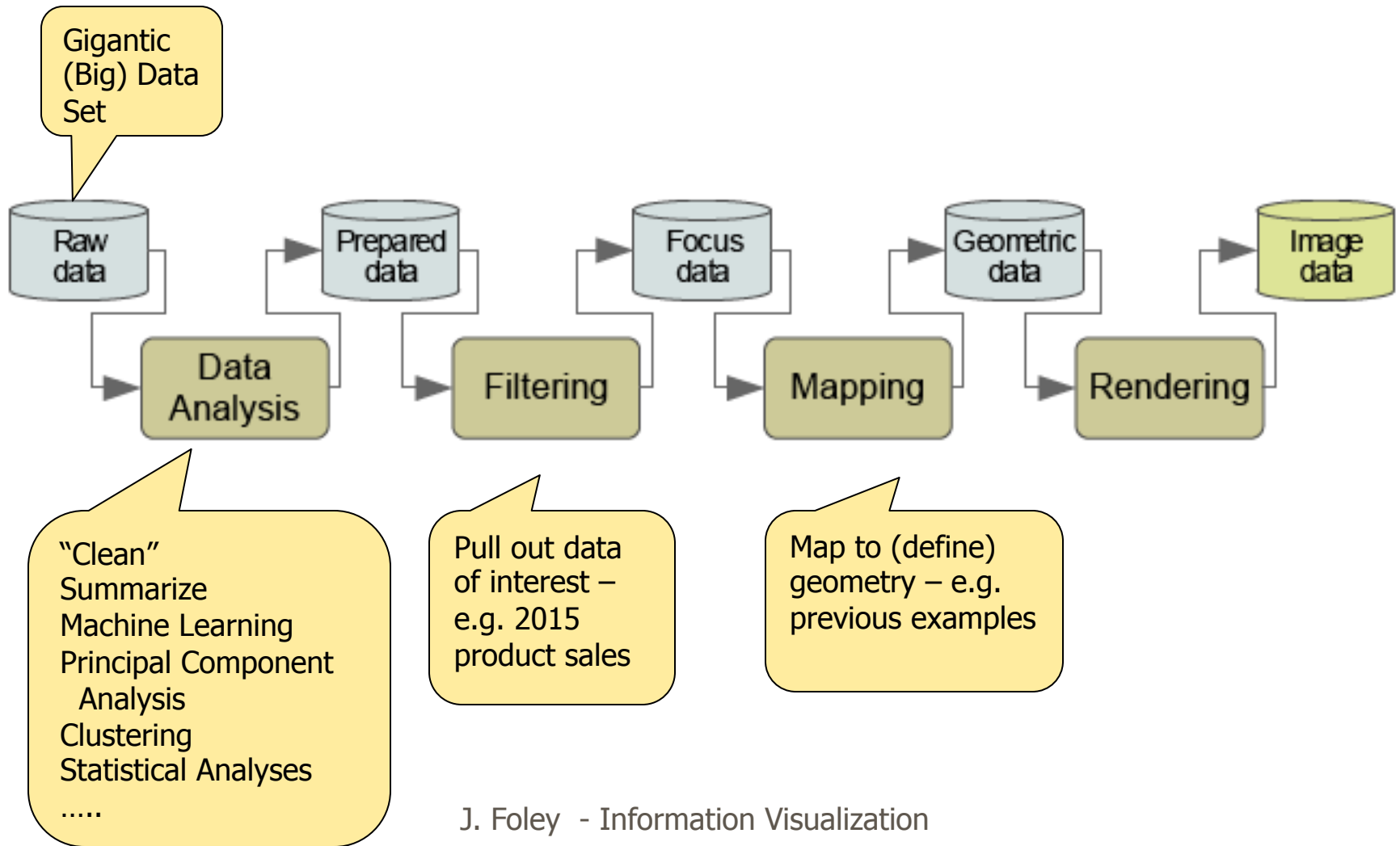
"Clean"
Summarize
Machine Learning
Principal Component
Analysis
Clustering
Statistical Analyses

.....

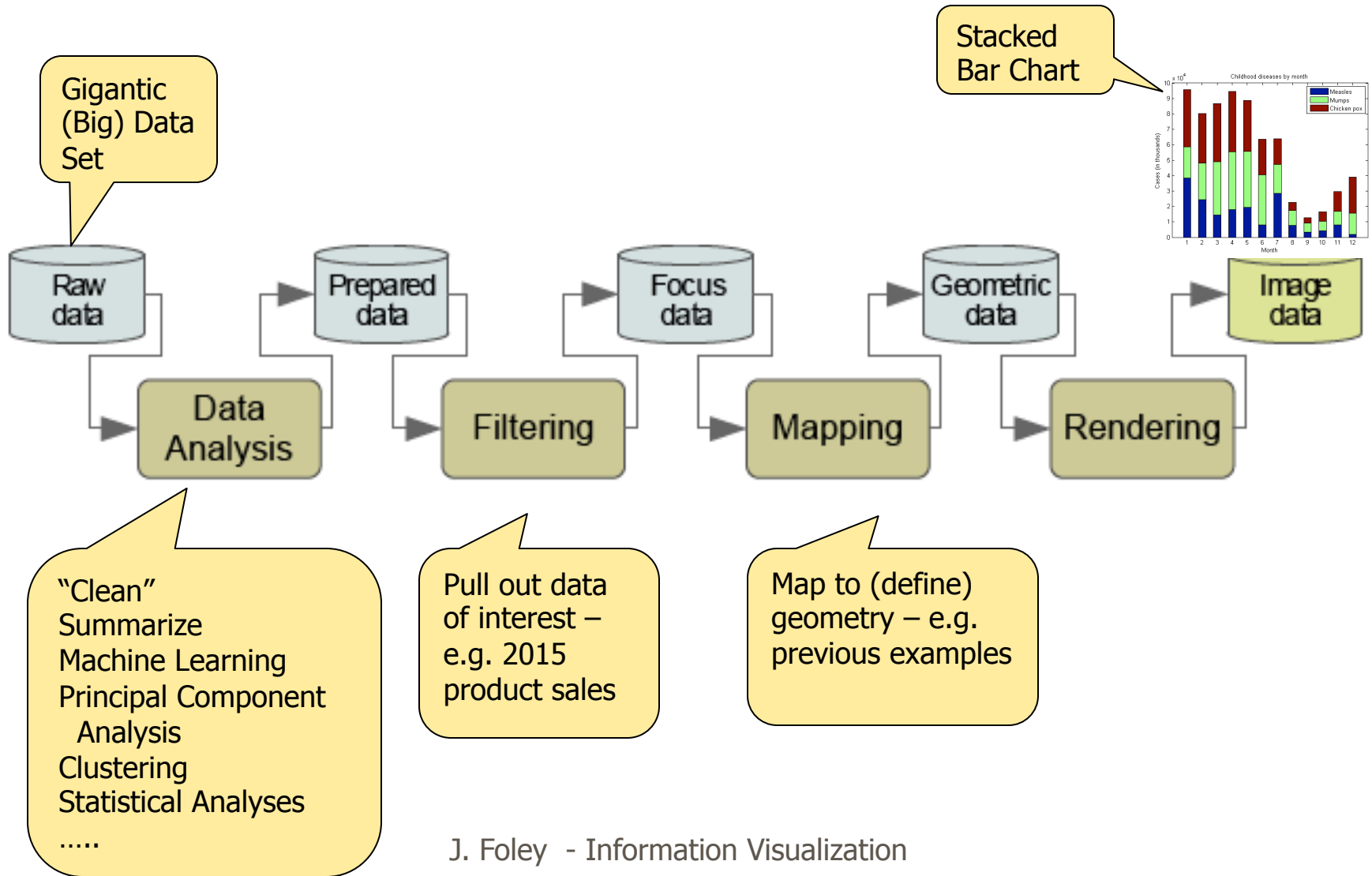
InfoVis and "Big Data"



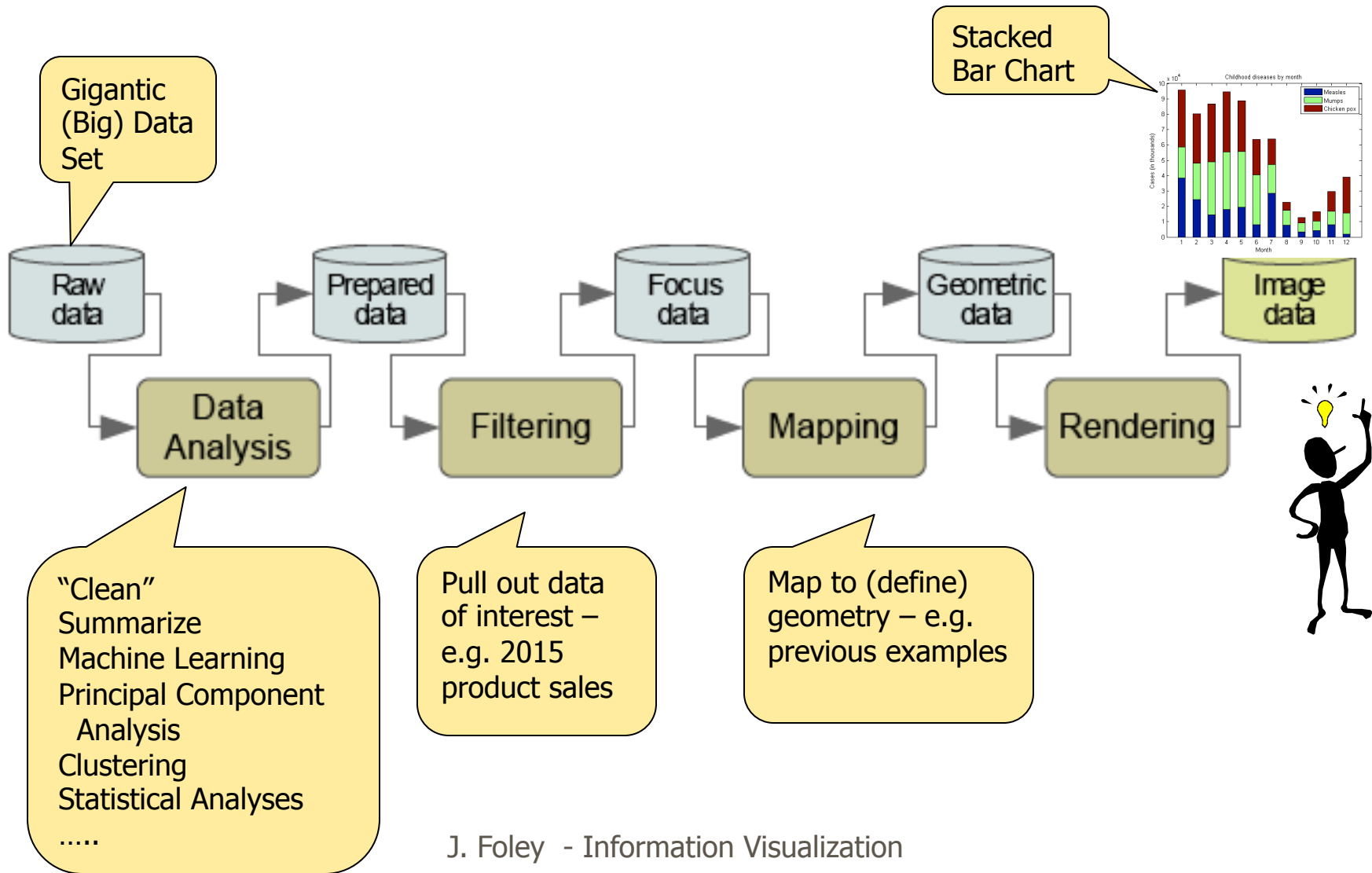
InfoVis and “Big Data”



InfoVis and "Big Data"



InfoVis and "Big Data"



Basic Data Types



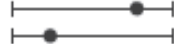
- N-Nominal (categorical)
 - Equal or not equal to other values
 - Example: gender
- O-Ordinal
 - Obeys $<$ relation, ordered set
 - Example: freshman, sophomore, junior, senior
- Q-Quantitative
 - Can do math, equal intervals
 - Examples: distance, weight, temperature, population count, your age


Data Type Implies Mark Type




Data Type: Ordinal & Quantitative

➤ Magnitude Channels: Ordered Attributes

Position on common scale 

Position on unaligned scale 

Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

Effectiveness
Most
Least

Same

Same

Data Type: Nominal

➤ Identity Channels: Categorical Attributes

Spatial region 

Color hue 

Motion 

Shape 

Not an exhaustive list

Useful Data Sub-Types



- Geo-coding – (latitude, longitude)
- Time-coded
 - Time-stamp – (date, time)
 - Time-interval - Δ (date, time)
- What about text (documents)?
 - People, places, things: Nominal
 - Dates, times: Quantitative
 - Text analysis => relationships within/between
 - Typically show attributes and relations

Active Learning Pause



- Discuss previous material with neighbor
- What not clear?
- Most important concept(s)

Data Organizations



- Multivariate (flat file, table, relation)
- Hierarchy (tree)
- Network (graph)
- All can include time and geo-codings
- (Non-exhaustive list)

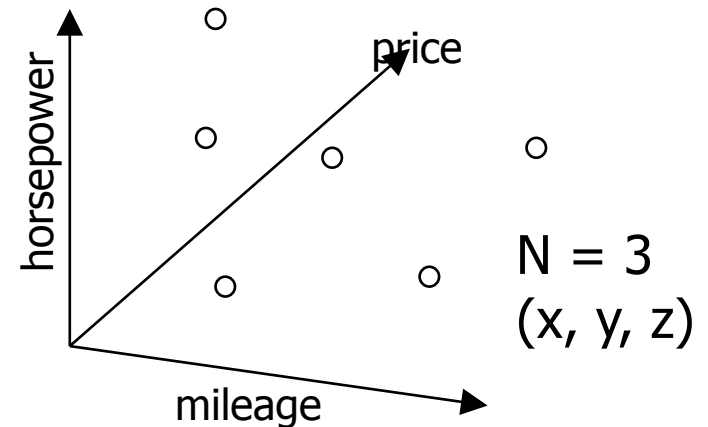
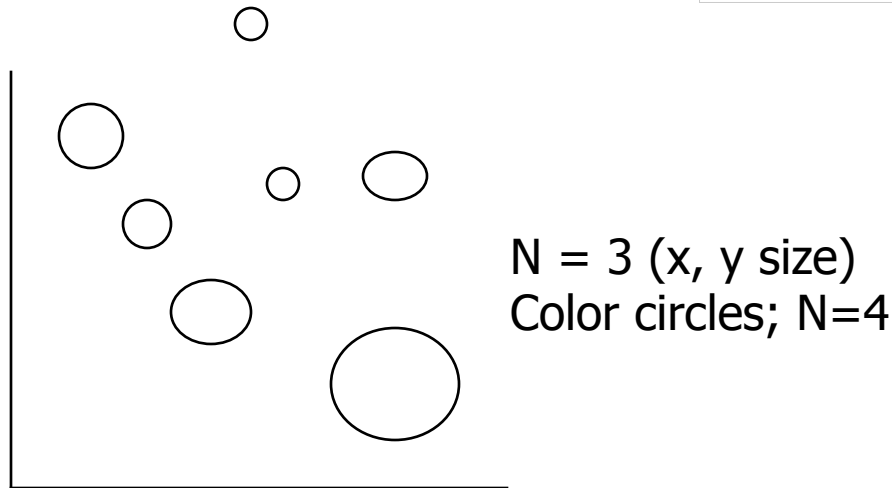
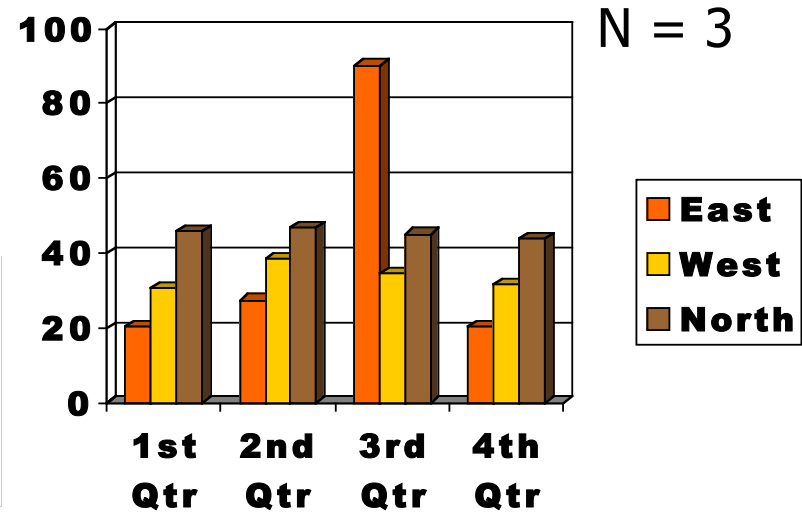
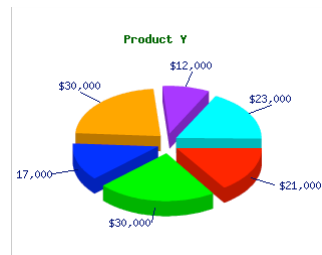
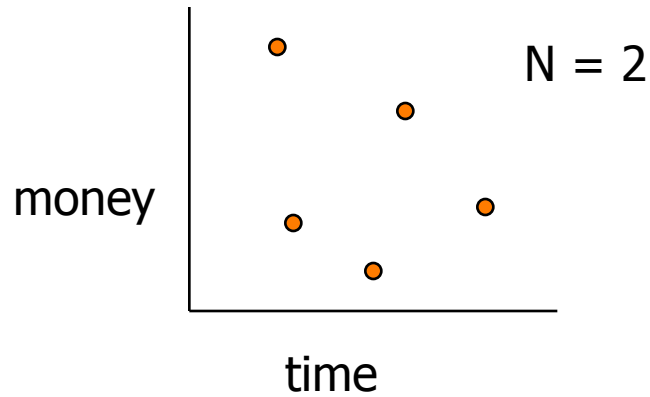
Data Organizations - Multivariate



- Each row is a *case*
- Each column is a *variable*
- Each column has implied or explicit *data type*
- Each column may have explicit *meta data*

	A	B	C	D	E	F	G	H	
1	Name	At Bats	Hits	Home Run	Runs	Rbi	Walks	Years In M	Care
2	STRING	INT	INT	INT	INT	INT	INT	INT	INT
3	Andy Allanson	293	66	1	30	29	14	1	
4	Alan Ashby	315	81	7	24	38	39	14	
5	Alvin Davis	479	130	18	66	72	76	3	
6	Andre Dawson	496	141	20	65	78	37	11	
7	Andres Galarraga	321	87	10	39	42	30	2	
8	Alfredo Griffin	594	169	4	74	51	35	11	
9	Al Newman	185	37	1	23	8	21	2	
10	Argenis Salaza	298	73	0	24	24	7	3	
11	Andres Thomas	323	81	6	26	32	8	2	

Traditional Graphs N = 2 & 3

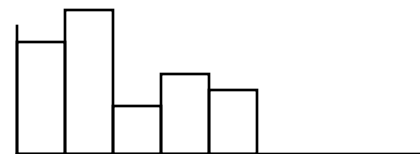
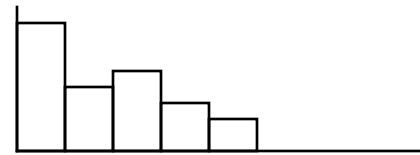
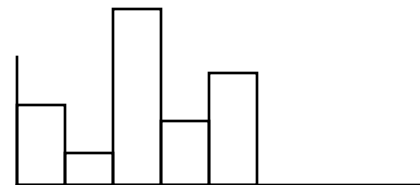


N > 3 – Small Multiples



Give each case its own bar graph

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5



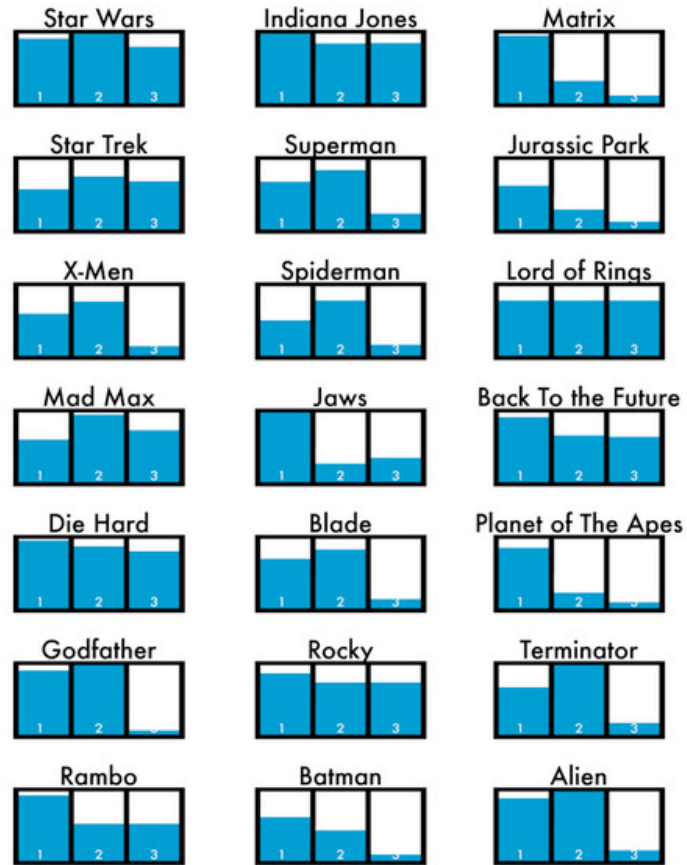
Or could do for each variable
=> 5 small bar graphs 😊

Do we have to use bar graphs?

Small Multiples: Movie Trilogy Successes



THE TRILOGY METER



#1 In A Series of Pop Cultural Charts

DANMETH.COM

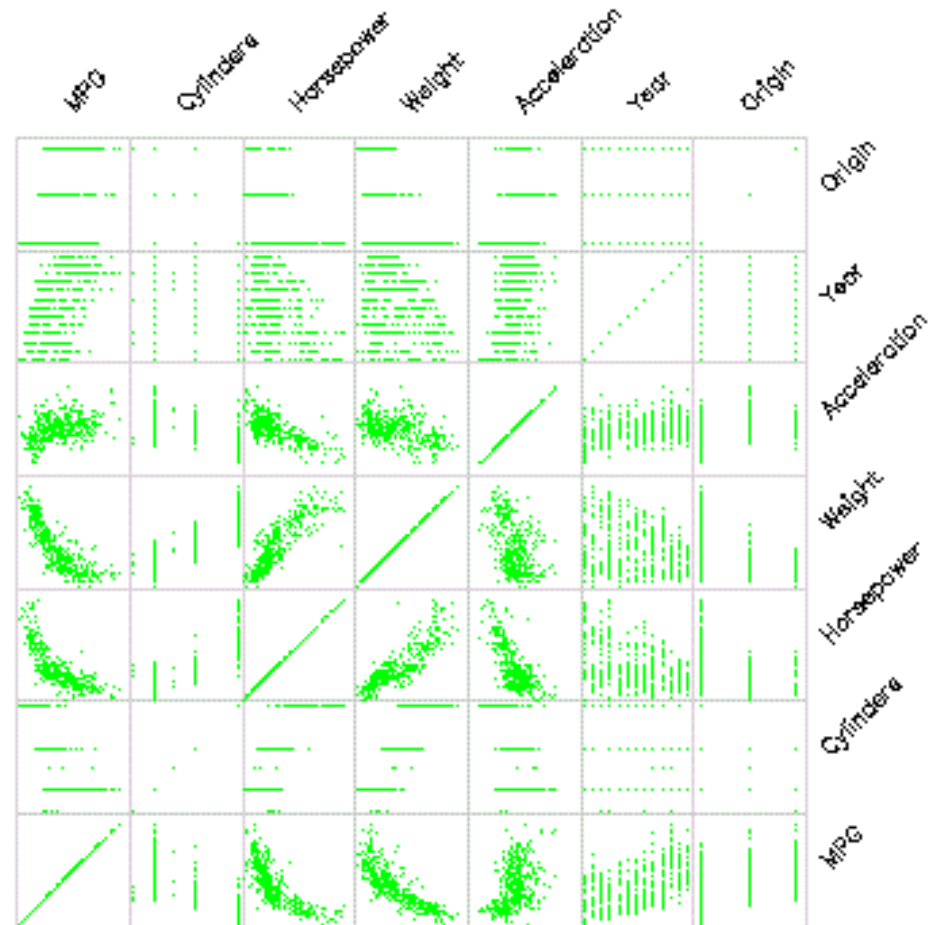
<http://danmeth.com/post/77471620/my-trilogy-meter-1-in-a-series-of-pop-cultural>

$N > 3$ – Scatterplot Matrix



Represent each possible pair of variables in their own 2-D scatterplot

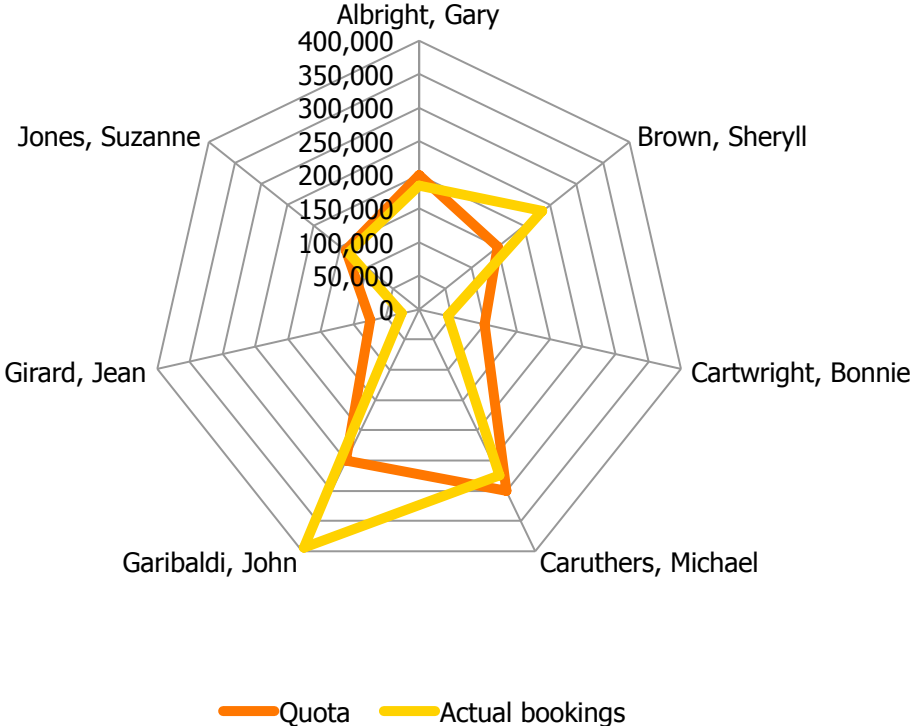
Useful for what?
Misses what?



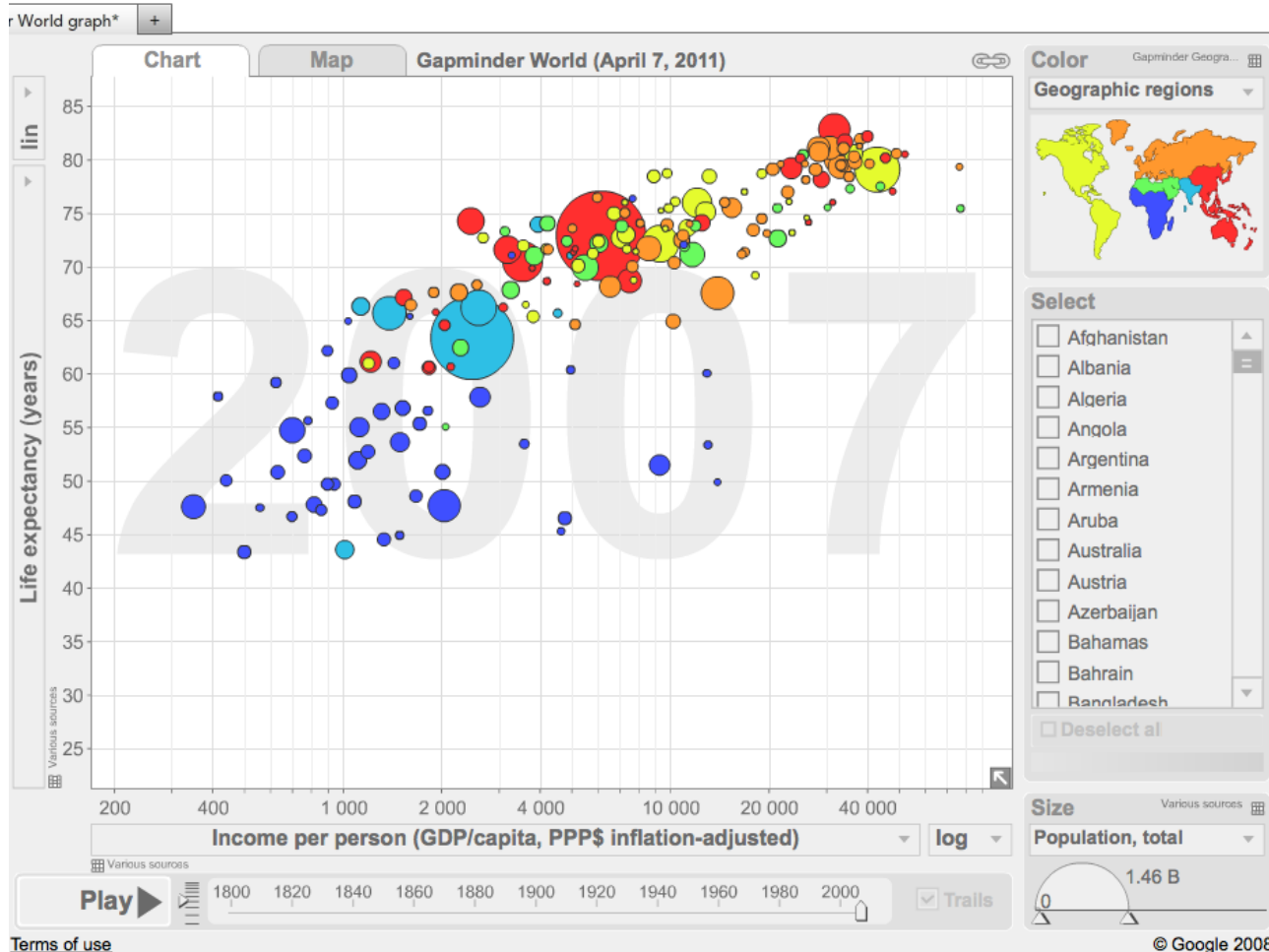
Star Plot (aka Radar Plot) $N > 3$



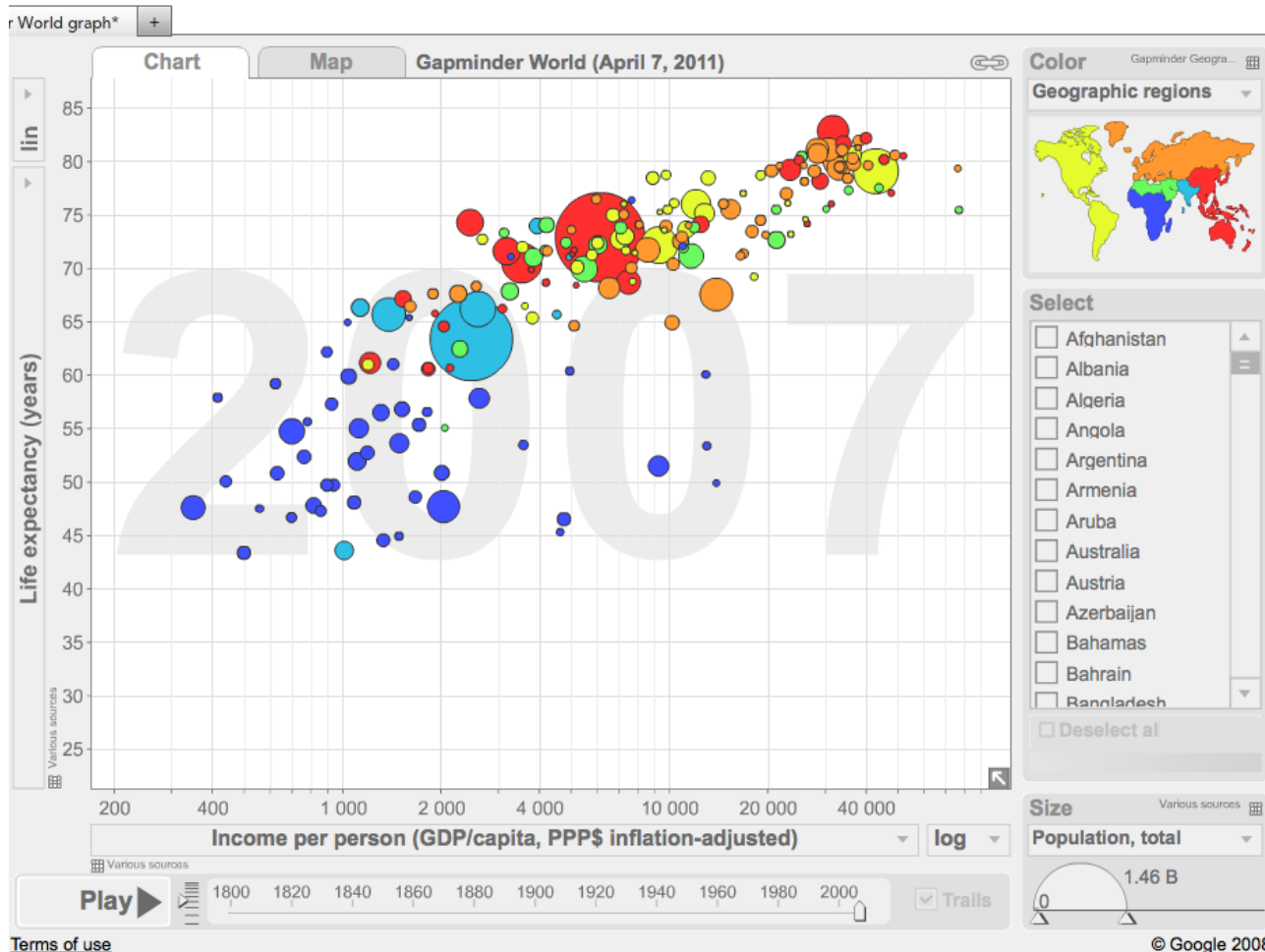
Actual Sales vs. Quota



N > 3 GapMinder – How Many Variables?



N > 3 GapMinder – How Many Variables?



- Life expectancy
- Income
- Population
- Region
- Year
- Country name

Parallel Coordinates



- Infoscope demonstration

Multivariate Data with Time: Electronic Medical Records

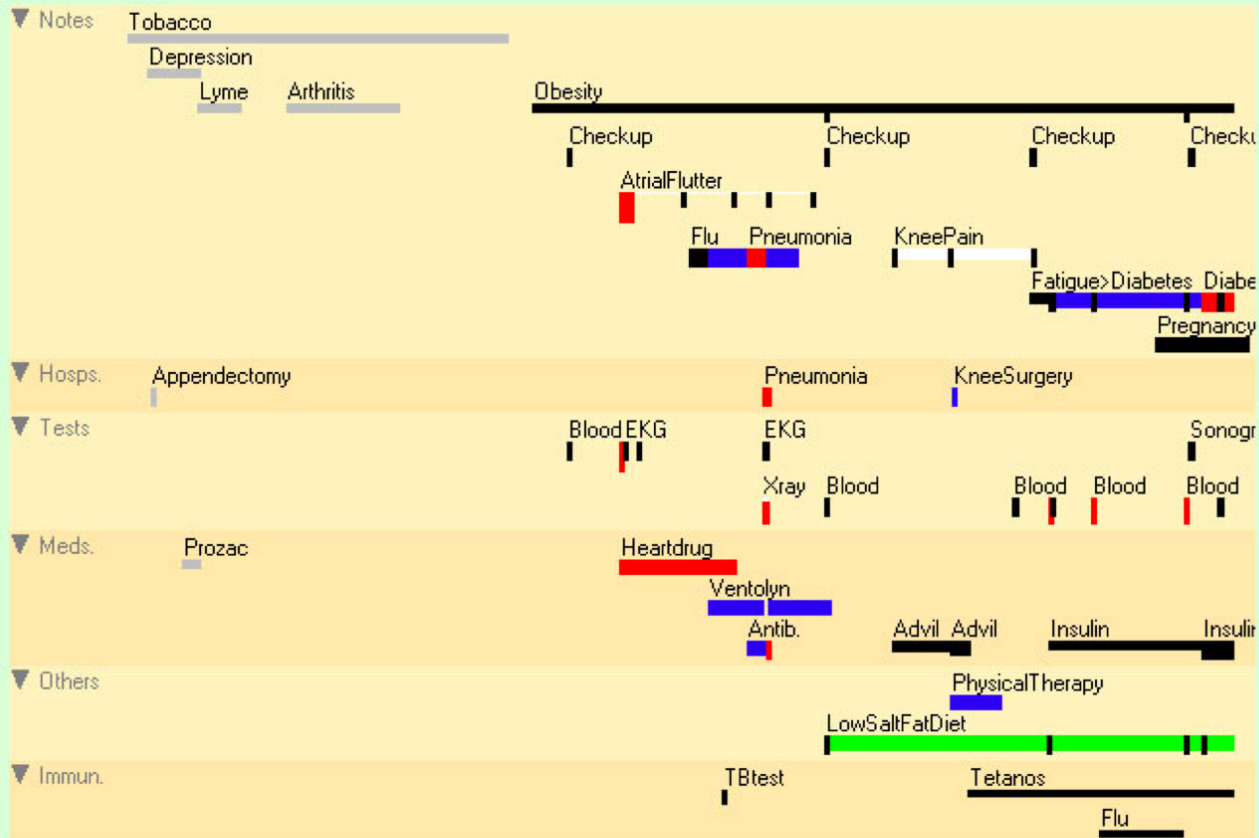


Linda Simpson
Female 40

Line from input file: %-,3-10-1997,3-12-1997,black,p10,Sonogram,images/babysonogra

LifeLine

92 93 94 95 96 97



92

93

94

95

96

97

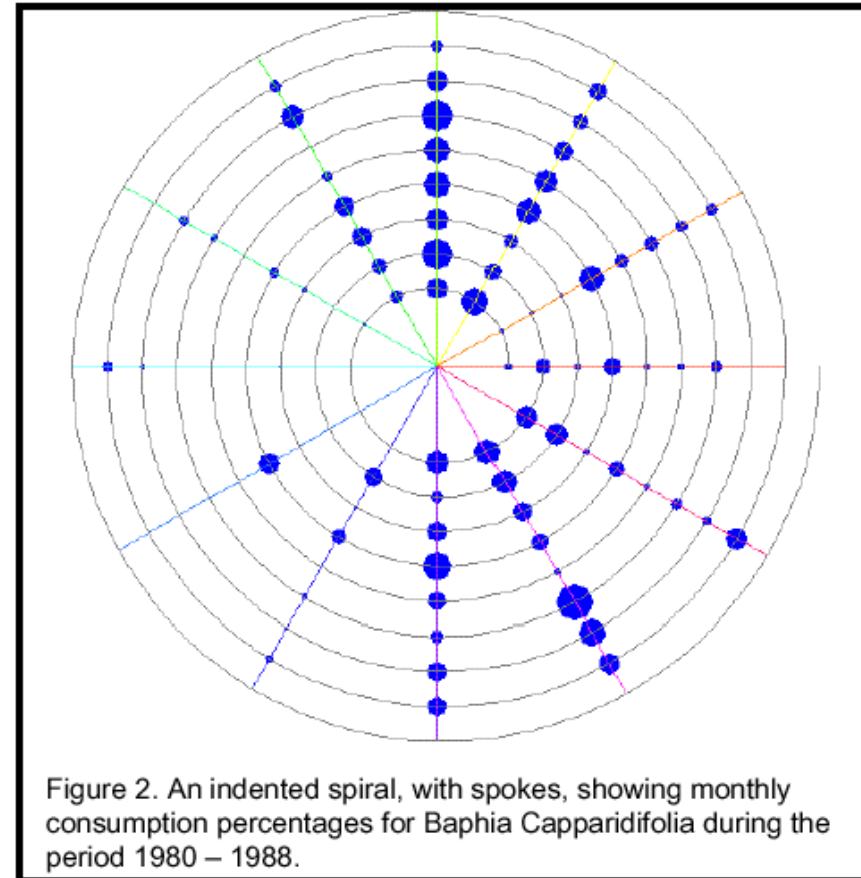


Multivariate with Monthly Time Data



- Monthly food consumption
 - Spoke => month
 - Dot size => consumption
 - Extra spiral on outside provides some closure

Konstan and Carlis, *Interactive Visualization of Serial Periodic Data*, UIST '98; data set is chimpanzee food consumption.



Time-based Multivariate with Events



- A [Nationally Known Business Author Profiles Michigan Company](#)
Trading Markets - 30 minutes ago
- B [Earnings Preview: EMC Corp.](#)
Daily Markets - 1 hour ago
- C [Bank of America - Worst Performing Dow Component](#)
TradersHuddle.com - 1 hour ago
- D [Computer Point takes on IBM firms in Uganda](#)
East African Business Week - 2 hours ago
- E [IBM rolls out virtual desktop offering](#)

[All news for International Business Machines Corp. »](#) [Subscribe](#)

Advertisement

**FREE PENNY
STOCK ALERTS**

- <http://finance.yahoo.com/echarts?s=GOOG#symbol=goog;range=20120203,20130828;compare=;indicator=ud+volume;charttype=candlestick;crosshair=on;ohlcvvalues=0;logscale=off;source=undefined;>
- How many variables; what are they, what are data types?

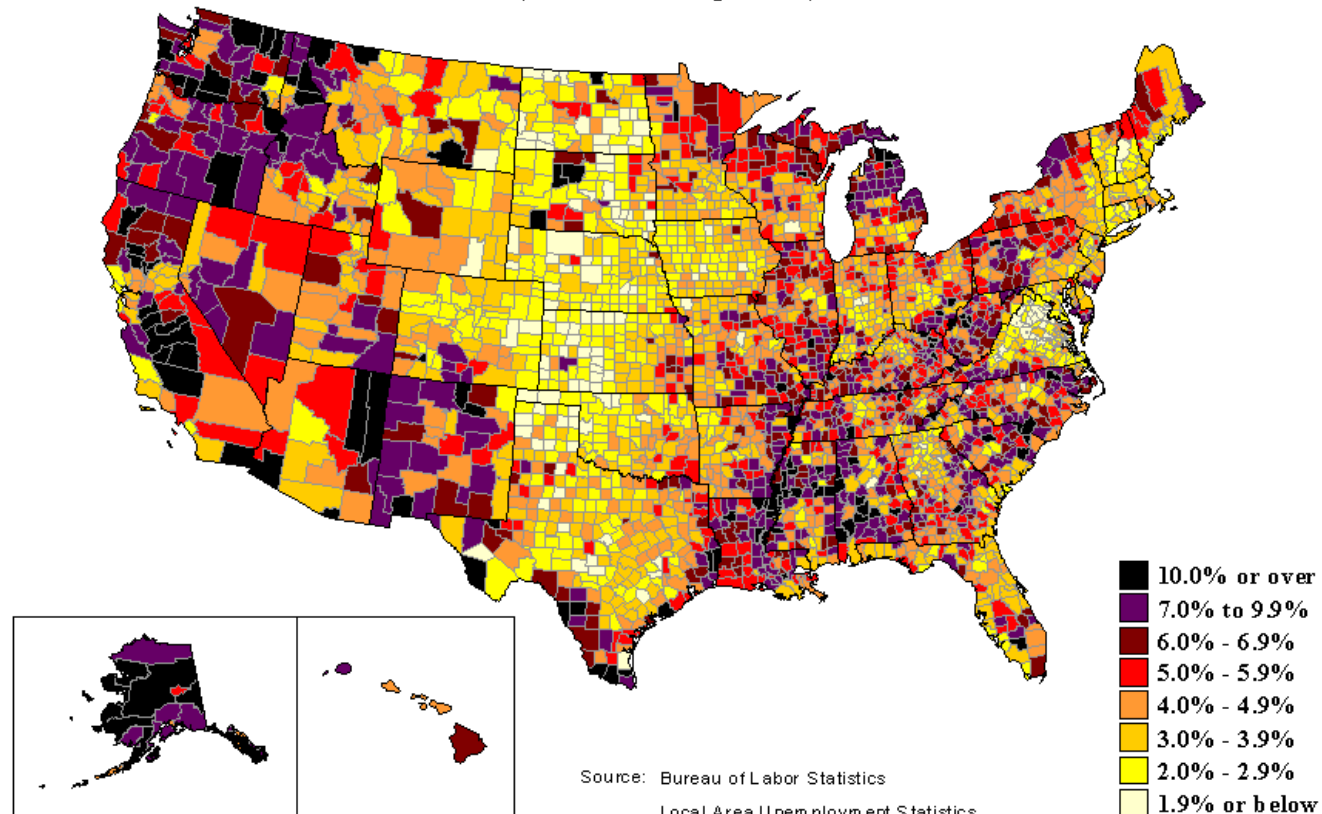
Multivariate with Geo Data



Political,
demographic
data

**Unemployment rates by county,
December 2000 - November 2001 averages**

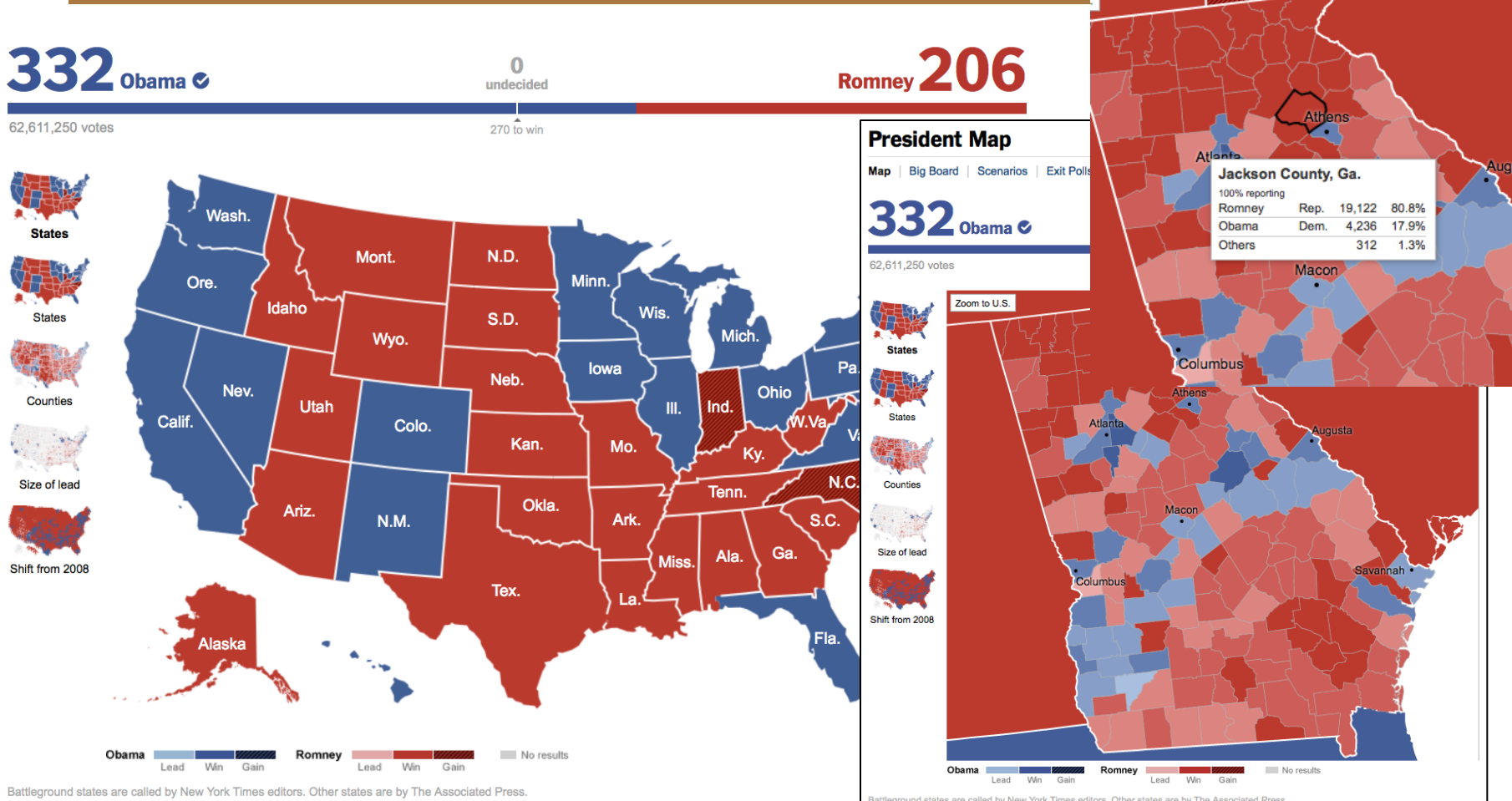
(U.S. rate = 4.6 percent)



New York Times Election Map



- NYT has big InfoViz team
- <http://elections.nytimes.com/2012/results/president>



Active Learning Pause

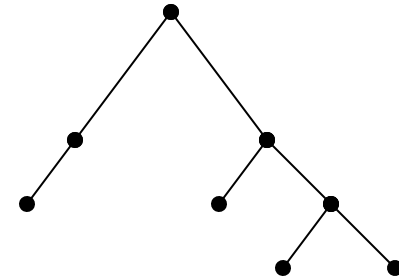


- Discuss previous material with neighbor
- What not clear?
- Most important concept(s)

Data Organizations - Hierarchy



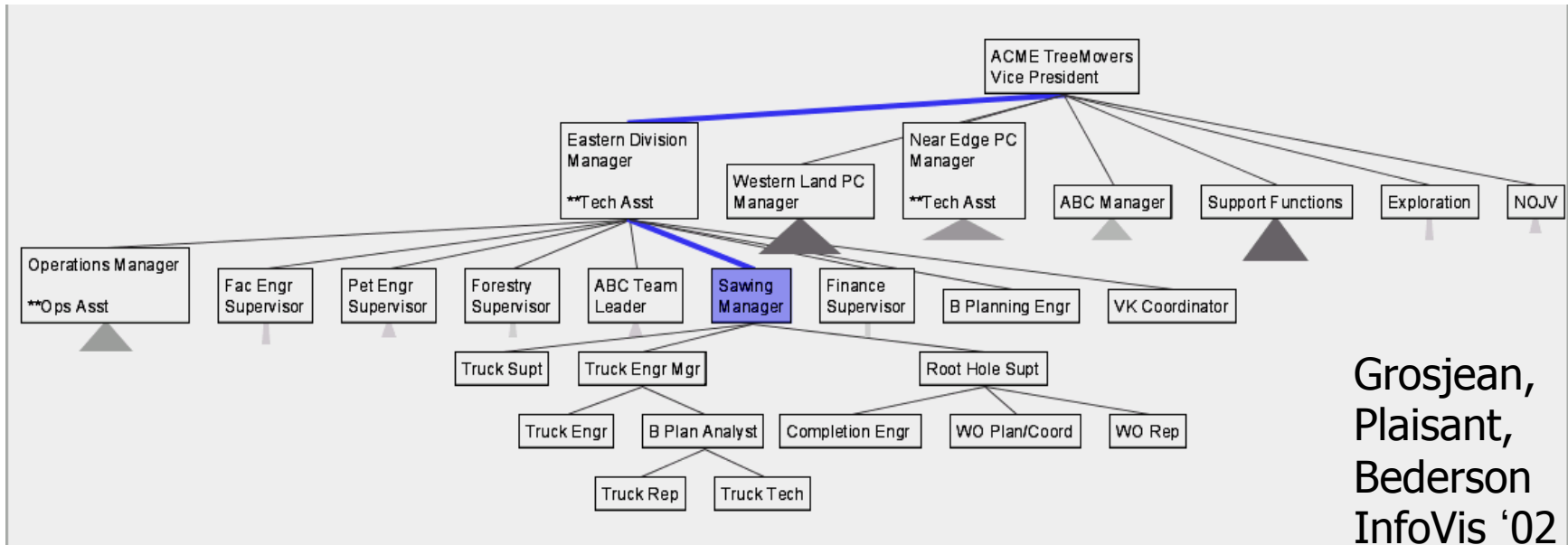
- Tree, as in organization chart
- Single root node
 - In-degree = 0
- Leaf nodes
 - Out-degree = 0
- For all nodes
 - In-degree = 1
 - Except for root node = 0
- For all non-leaf nodes
 - Out-degree > 0



SpaceTree



- Conventional 2D layout techniques with clever interactions



Live Demo at <http://www.cs.umd.edu/hcil/spacetree/applet/applet.shtml>

Treemap



- SmartMoney.com Map of the Market
 - Illustrates stock movements
 - <http://www.smartmoney.com/map-of-the-market/>
- Demo shows lots of interaction methods

Image on next slide

J. Foley - Information Visualization

Wattenberg
CHI '99

Active Learning Pause



- Discuss previous material with neighbor
- What not clear?
- Most important concept(s)

Data Organization - Network



- (Avoid using term *graph* – ambiguous)
- Highway network, Internet, Social network
- Removes tree's limits on in and out-degrees

Social Networks: Facebook TouchGraph

Profiles Networks

Show top 80 Friends Upload Advanced Restart

Zoom: Spacing:

<http://www.facebook.com/home.php>

Jim Foley
Networks: [Georgia Tech](#)
Mutual Friends: 284

[Facebook Profile](#)

Network All All List Photo

Name	Rank #	Friend #
Jim Foley	1	284
Mark Guzdial	2	112
Amy Bruckman	3	106
Irfan Essa	4	84
Beki Grinter	5	99
Annie Antón	6	81
Randy Howard Katz	7	49
Peter Freeman	8	69
James Landay	9	49
Jeff Vitter	10	47
Gene Spafford	11	53
Dick Lipton	12	52
Desney Tan	13	48
Scott McCrickard	14	48
Ed H. Chi	15	47
Dan Reed	16	42
Eugene Medynskiy	17	57
Eileen Kraemer	18	38
David White	19	73
Erika Shehan Poole	20	57
Peter N Wan	21	42
Anind Dey	22	47
Tony Baylis	23	25
Andrea Forte	24	60
Jill Dimond	25	43
Kurt Luther	26	47
Mario Romero	27	55
Sarita Yardi	28	58
Mary Czerwinski	29	43
Nicole Yankelovich	30	25
Robin Jeffries	31	34
Lena Mamykina	32	55
Ashwin Ram	33	46

Georgia T

powered by TouchGraph

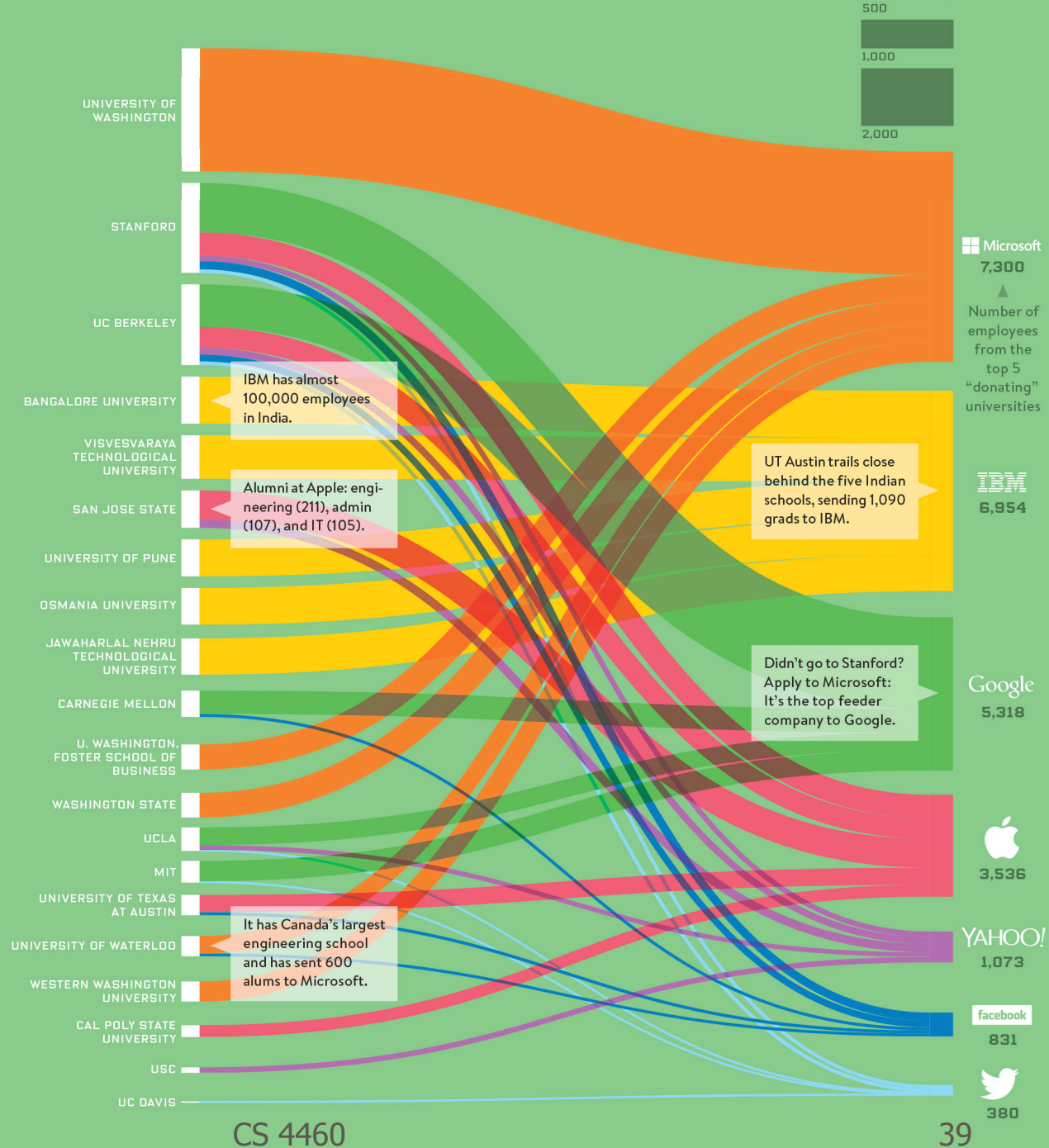
Flows from Schools to Companies

Where's GT?

Two types of nodes – schools and companies – is that OK?

Could this be extended to show flows from one company to another?

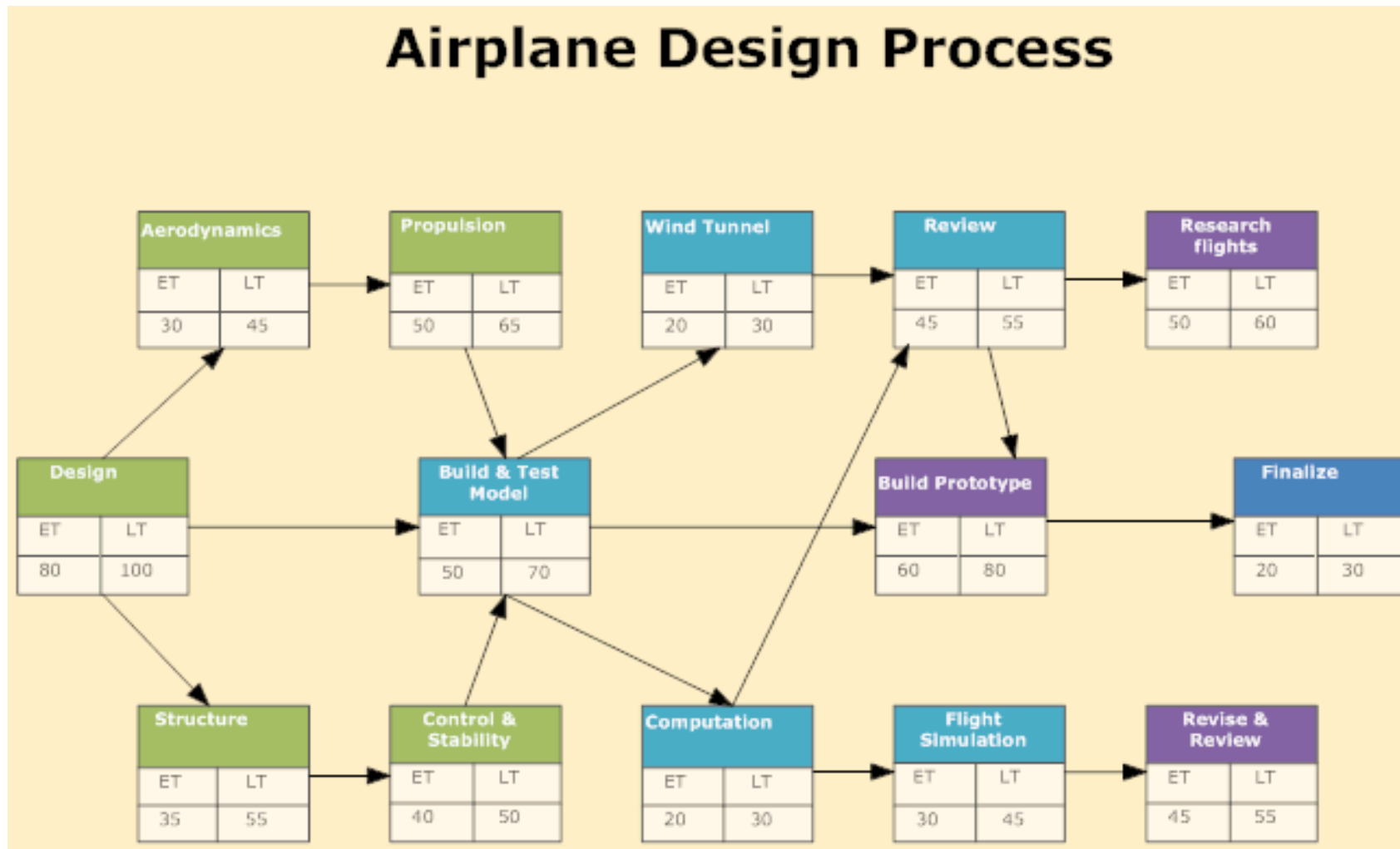
Source: Wired http://www.wired.com/wp-content/uploads/2014/05/in_schools_f.jpg



PERT Chart - Time Dependencies



- PERT = Project Evaluation and Review Technique

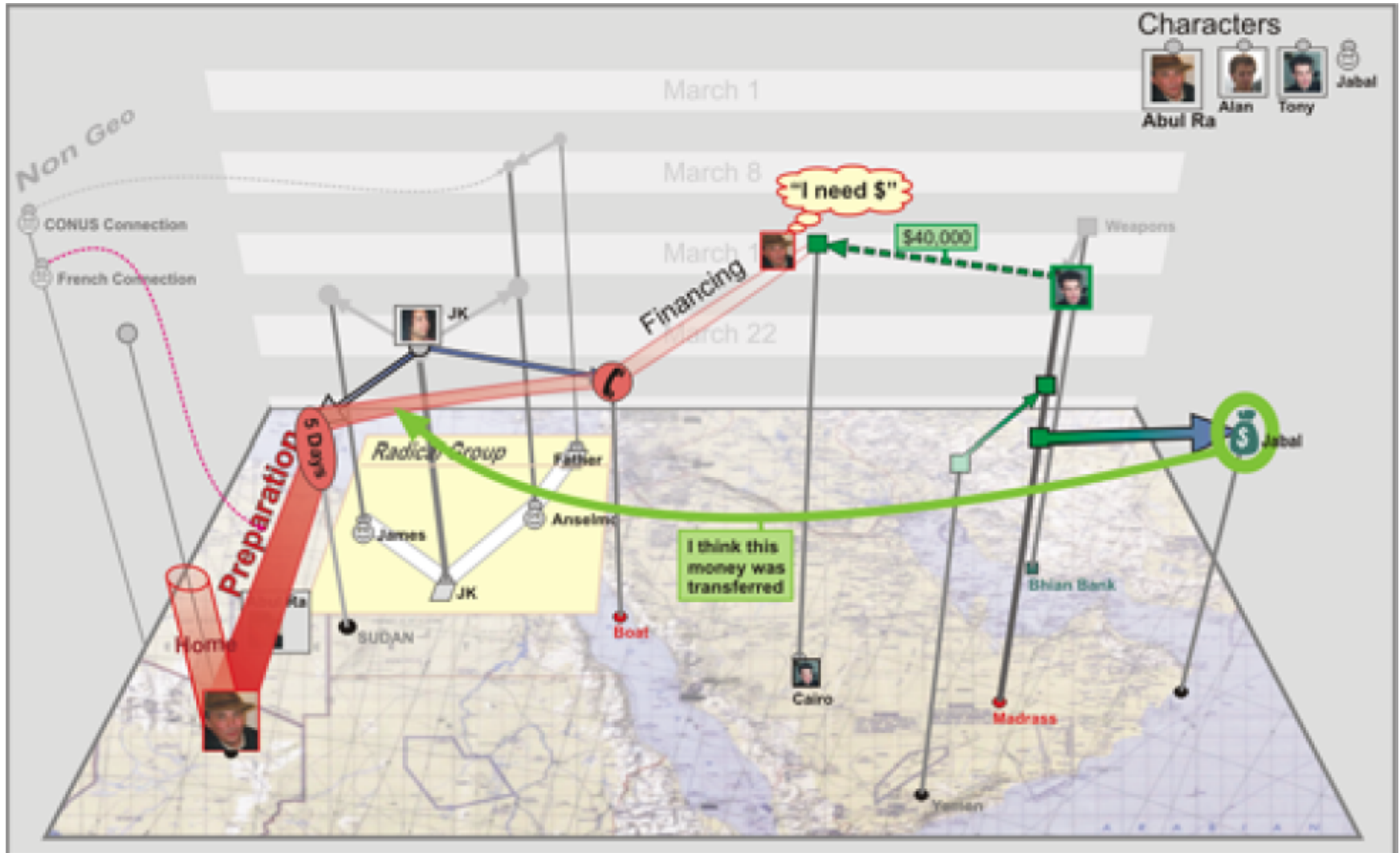


Network with Time- and Geo-data



Or an intelligence analysis

GeoTime



Active Learning Pause



- Discuss previous material with neighbor
- What not clear?
- Most important concept(s)

Interaction – Where the Action is!



- Two main components in an infovis
 - Representation
 - Interaction
- Representation gets all the attention
BUT REMEMBER
- Interaction separates Information Presentation (aka InfoGraphics) from Information Visualization

Types of Interactions



- Context & Detail aka Overview & Detail
 - Details on Demand – details on one specific case
 - Focus & Context – focus on multiple cases
- Dynamic query/filter
- Brushing, aka Linking
- Change Representation, aka Re-encode
- Animation
- Explore
- Reconfigure/rearrange
- Abstract/Elaborate
- Zoom and Pan
- (not an exhaustive list but a good start :-))

- We saw some of these in Gapminder – more examples coming

How to do InfoViz



- Interactive GUI
 - Tableau
 - Spotfire
 - Many more
 - Dashboard
 - Business intelligence
- Great place to start
 - May not get you what you want
 - Some are expensive
- Programming
 - Processing
 - D3
 - Google API
- Customize!
- Takes time!
- Learning curve!

Resources - Books



- The Visual Display of Quantitative Information. Edward Tufte, Graphics Press, 1983.
- Envisioning Information. Edward Tufte, Graphics Press, 1990.
- Visual Explanations. Edward Tufte, Graphics Press, 1997.
- Show Me the Numbers; Designing Tables and Graphs to Enlighten. Stephen Few, Analytics Press, 2004.
- *Now You See It*, Stephen Few, Analytics Press, 2009.
- *Interactive Data Visualization for the Web*, Scott Murray, O'Reilly Media. All about D3
- *Design for Information*, Isabel Meirelles, Hachette Books
- *Search User Interfaces*, Marti Hearst, free download at <http://searchuserinterfaces.com/book/>.
For business intelligence and business dashboards: Wayne Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*, Wiley, 2005
- For Network Visualization, particularly Social Networks: Hansen, Shneiderman and Smith, *Analyzing Social Media Networks with NodeXL*, Morgan Kaufman, 2011.
- For psychological/perceptual factors affecting information visualization: Colin Ware, *Information Visualization: Perception for Design*, Morgan Kaufman 2004.
- For a deeper treatment of many aspects of InfoViz: *Visualization Analysis and Design*, Tamara Munzer, CRC Press, 2015.

Resources – Web Sites - Many



- GapMinder
 - <http://www.gapminder.org/world>
- John Stasko's resource page
 - <http://www.cc.gatech.edu/~john.stasko/7450/resources.html>
- My Course:
 - <https://cs4460infovis.wordpress.com/>
- Google terms like
 - information visualization software
 - information visualization course
 - information visualization book
 - information visualization examples
 - information visualization tools
 - information visualization jobs
 - information visualization journal
 - information visualization D3
- A Gallery of Galleries of InfoVis examples
 - <http://visualoop.com/blog/11044/30-amazing-data-viz-galleries-everyone-should-follow>

The End

