https://queue.acm.org/detail.cfm?id =1805128

Challenges in Making the Hidden Visible

Based on material in 05-389 Interactive Data Science data.cmubi.org for more info t/th 9-10:30, Spring

Jennifer Mankoff and many other collaborators



Information can change how we act in the world [e.g., СНГО9, ICWSM'09]



stepgreen Penrich your life.

Report Actions Sh	nare Acco	unt Help	About Admin		Logged in	as: jmankof
	Show	ime graphs		Each square is a mo	prewood gardens east tower resi	dent.
Your This week savings 127 lbs	Overall saved 23260	Other person saving	10 6 101/		N. A.	Your saving displayed on
Turn off work screen say	rer 6 lbs			_	with the second second	contains the s
Use CFLs	21 lb	3			a second and	Information
Teach proper hand wash	ing. 8 lbs				and and the	anonymous ; residing in mo
Plant a tree	4 lbs				1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	gardens east
Insulate water heater	4 lbs				and the second of the	
Programmable thermos	tat 27 lb	s			Litt of the property of the second	
Turn off home screen sa	ver 11 lb	5			Statement Statement	
Lower water heater	23 lb	5			Section of the sectio	-
Use sleep mode at home	6 lbs	6 lbs	Use sleep mode at work	- I	The second second	Exit dorm

Information For and About People

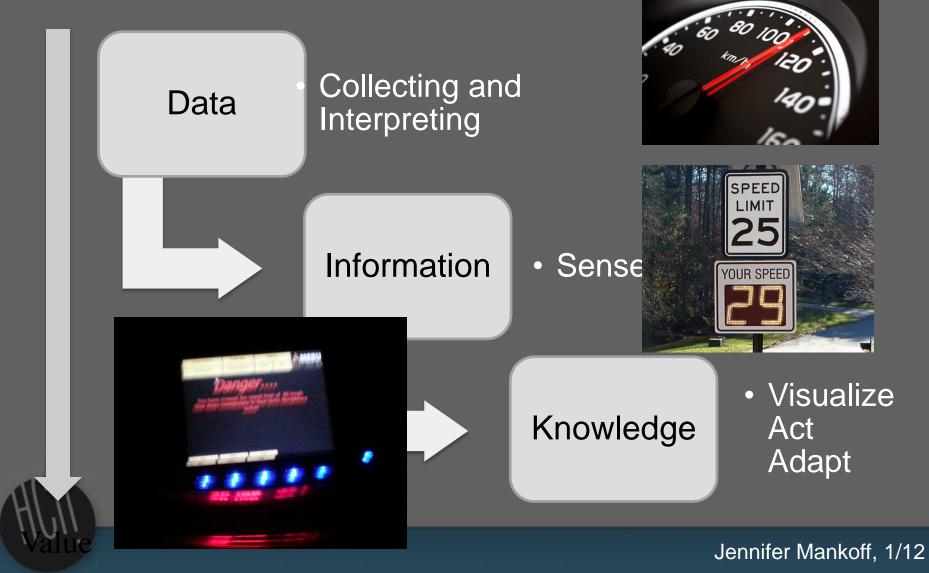
As a prosthetic, changes what we can do in the world

As a motivator, changes how we (or our machines) behave Shared with others, may Change the balance of power [Ubicomp '09, '10] Build new kinds of action and knowledge

Beyond individuals, may support policy, politics, economics [DIS In Submission]



Making Data Actionable



Collecting the right data

What is the problem? What data will solve the problem? How can we get that data?

Techniques needed: Careful analysis & thought

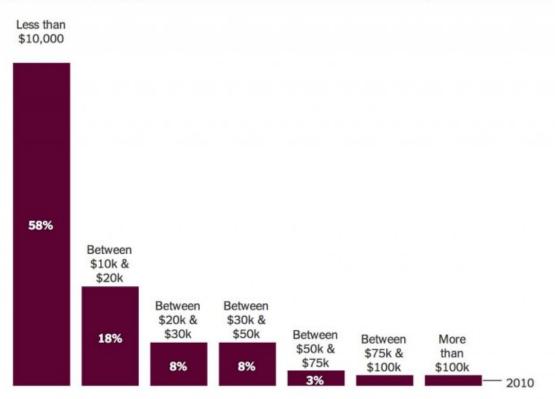
Tools: simulation & prototyping



Can you trust your data?

Large Amounts of Student Debt Are Not Common

Only 7 percent of young-adult households with student debt have more than \$50,000 in such debt.



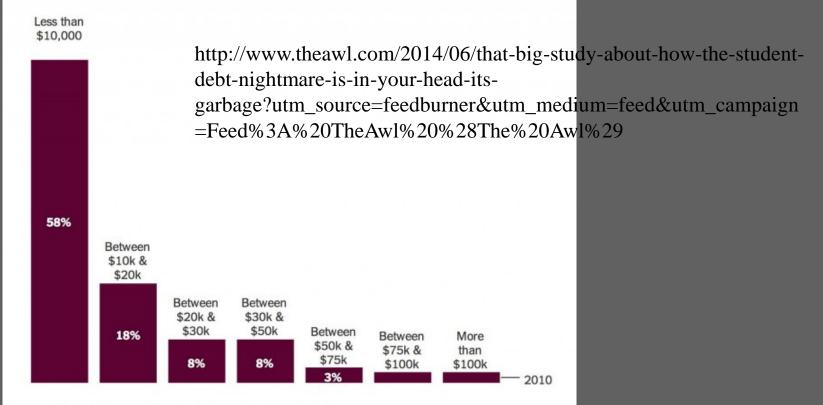
Source: Elizabeth Akers and Matthew Chingos, Brookings Institution

2010 data; based on households with people between 20 to 40 years old with at least some education debt

Can you trust your data?

Large Amounts of Student Debt Are Not Common

Only 7 percent of young-adult households with student debt have more than \$50,000 in such debt.



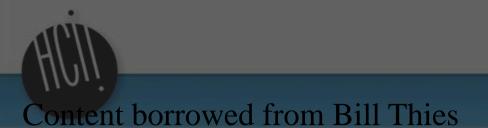
Source: Elizabeth Akers and Matthew Chingos, Brookings Institution

2010 data; based on households with people between 20 to 40 years old with at least some education debt

Can you trust your data?

Just because you have a lot of data, does not mean that it is *good* data

The plural of "anecdote" is not "data".



Sources of Error

Sampling errors

Random Errors due to the sample forming only part of the population Systematic Bias in sampling **Bias During Data Collection Demand Characteristics Illusory Superiority** Data entry / processing errors Data is generated accurately but errors introduced during recording or processing



What Makes a Good Sample?

Representative of the population (Along dimensions that matter to the question being asked)

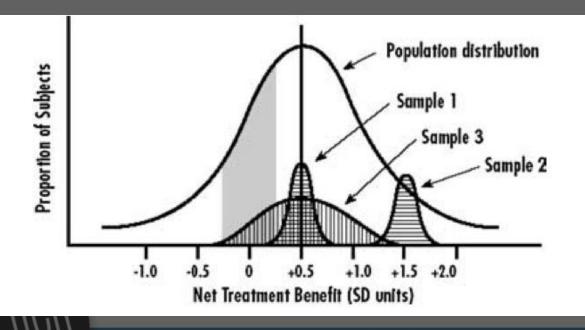


Figure 1: Kravitz *et al*, (2004) Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly* **82**(4):661-687

Sources of Error

Sampling errors *Random Errors* due to the sample forming only part of the population Systematic Bias in sampling **Bias During Data Collection Demand Characteristics Illusory Superiority** Data entry / processing errors Data is generated accurately but errors introduced during recording or processing



Example: Demand Characteristics

An experimental artifact where participants form an interpretation of the experiment's purpose and unconsciously change their behavior to fit that interpretation



Content borrowed from Bill Thies

What can we do to minimize demand characteristics in HCI?

Be aware that response bias affects all studies Dissociate from a particular design or solution Minimize the differences in social status between investigators and participants Use triangulation to validate data collected Tricks for asking sensitive questions

Content borrowed from Bill Thies

Jennifer Mankoff, $1/1_{41}^2$

Measurement Errors

Badly Designed Questions Badly Chosen Sensors Bad Administration of Measurement Instrument Inaccurate Measurements

Jennifer Mankoff, 6/12 See http://www.nchi.nlm.nih.gov/pmc/articles/PMC1323316/

Sources of Error

Sampling errors *Random Errors* due to the sample forming only part of the population Systematic Bias in sampling **Bias During Data Collection Demand Characteristics Illusory Superiority** Data entry / processing errors Data is generated accurately but errors introduced during recording or processing



Four C's of Data Quality

Is your data *Complete*? Is your data *Coherent*? Is your data *Correct*? Is your data *aCcountable*?



Jennifer Mankoff, 4/12

Questions about Completeness

Appropriate Data: Does the data you have match the questions you want to answer? Missing Data: Data does not exist because it was never obtained or was lost

Reporting error: The sensor (or respondent) is incorrect



Jennifer Mankoff, 6/12

Is your Data Coherent?

Does the data "add up"? Does it make sense relative to itself? Are the extreme values? Examples Non number in a numeric field Month field has something other than a month Email has no @ Hourly data adds up to 24 hrs per day Etc.

Is your data Correct?

Itemize aspects of your data that are easy to verify

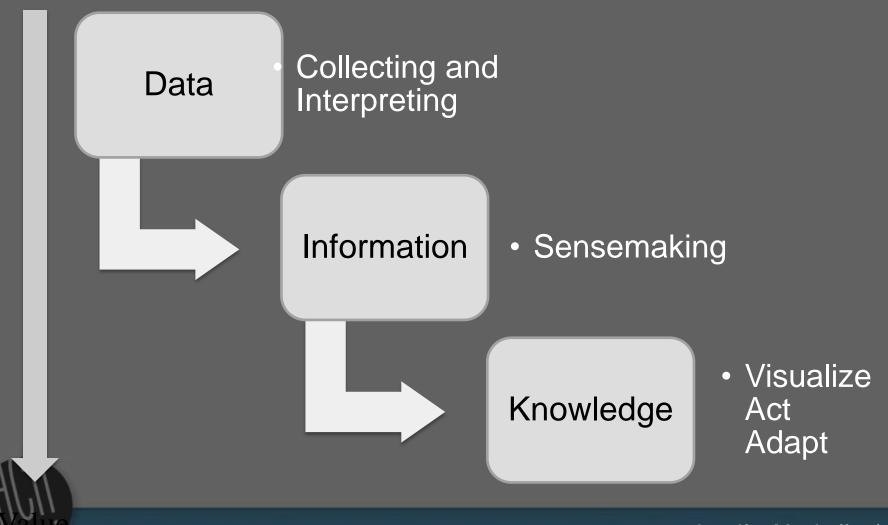
Compare (collect twice or find alt. source) Analyze the data collection strategy and look for sources of bias Within the population across variables (surveys with only round values;

people who report everything in round numbers)

Determine how much is bad



Making Data Actionable



Understanding Humans

Activities Routines Intent Causality



Understanding Humans

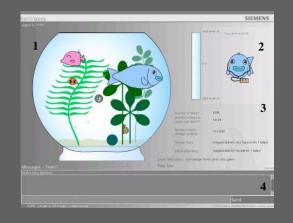
Activities Routines Intent Causality



Visualizing Activities



This Week's Activity



UbiFit Consolvo *et al.* '08



Fish 'n Steps Lin *et al.* '06

Stepgreen/UbiGr een Mankoff et al. '09 Froelich et/anhif@@ankoff, 1/12

Understanding Humans

Activities **Routines** Intent Causality



Why are routines important?

Develop routines to reduce cognitive effort

Deviations and anomalies cause stress and extra effort

at least as actionable as inferred activities
Longer term, more built-in
Opportunities for change
Barriers for change
Leverage point for understanding human intent
Untapped as a resource: sensing and using

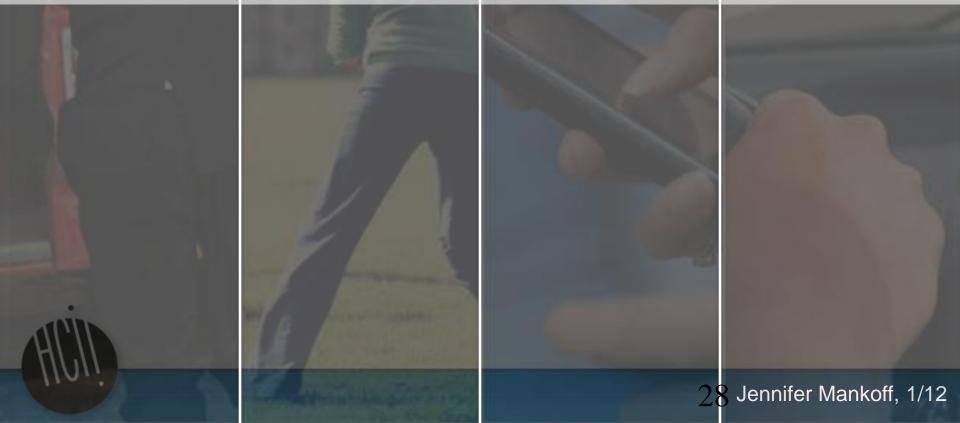


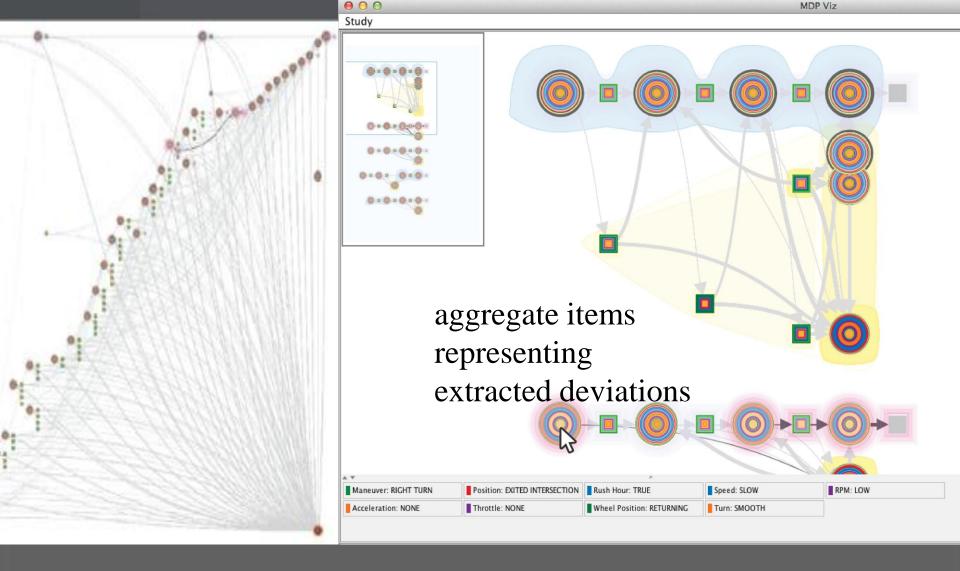
		SLEE	P	CRE	ATIVE	WORK	DAY .	JOB/AD	MIN	F	OOD/L	EISURE	EXER	CISE	DTHER			
12 AM -																	10	

0	HONORÉ DE BALZAC
U	
0	JOHN MILTON
0	HARUKI MURAKAMI
0	VOLTAIRE
0	BENJAMIN FRANKLIN
0	IMMANUEL KANT
0	KURT VONNEGUT Image: Comparison of the comparison of t
0	MAYA ANGELOU
0	WYSTAN HUGH AUDEN
0	LUDWIG VAN BEETHOVEN
0	MARY FLANNERY O'CONNOR
0	LE CORBUSIER
0	VICTOR HUGO
0	WOLFGANG AMADEUS MOZART
0	CHARLES DARWIN
0	CHARLES DICKENS
0	SIGMUND FREUD
0	VLADIMIR NABOKOV

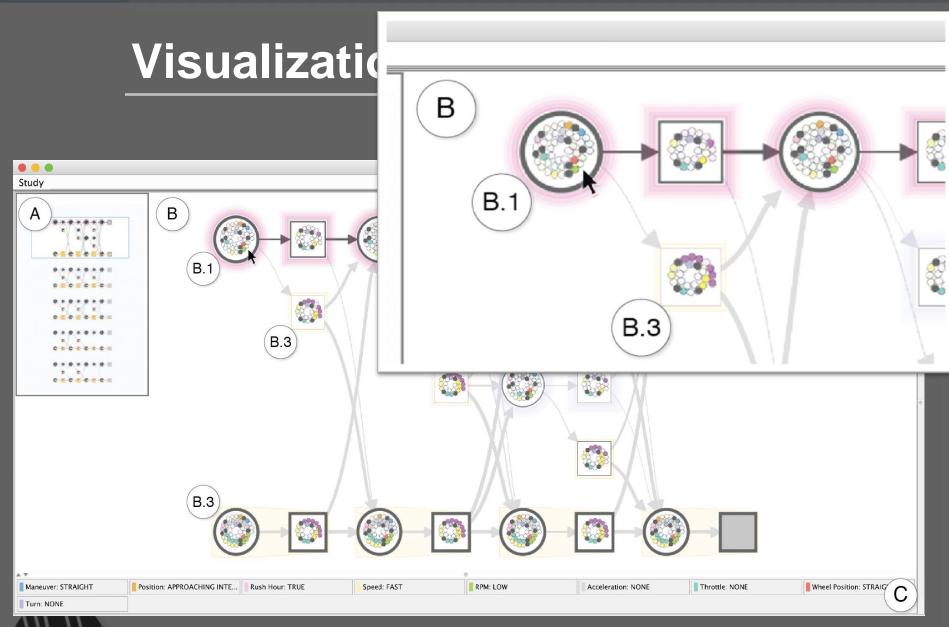
110

automatically extract *patterns* of human behavior that form routines and *deviations* from demonstrated behavior









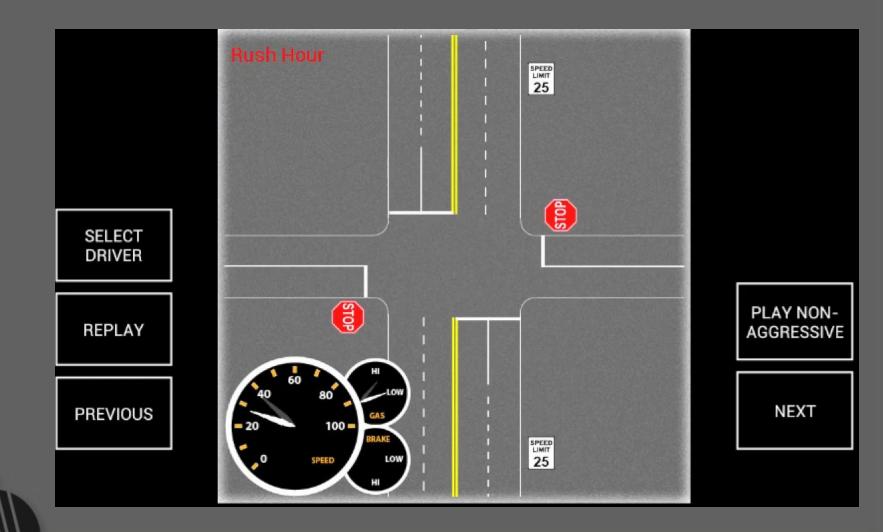
Now automated extraction (CHI 2017?)

Aggressive (Novice) Drivers (CHI 2014, CHI 2016, CHI 2017 submission)

US: 1500 deaths/year Cost of \$40 billion from crashes



Aggressive (Novice) Drivers: Interventions

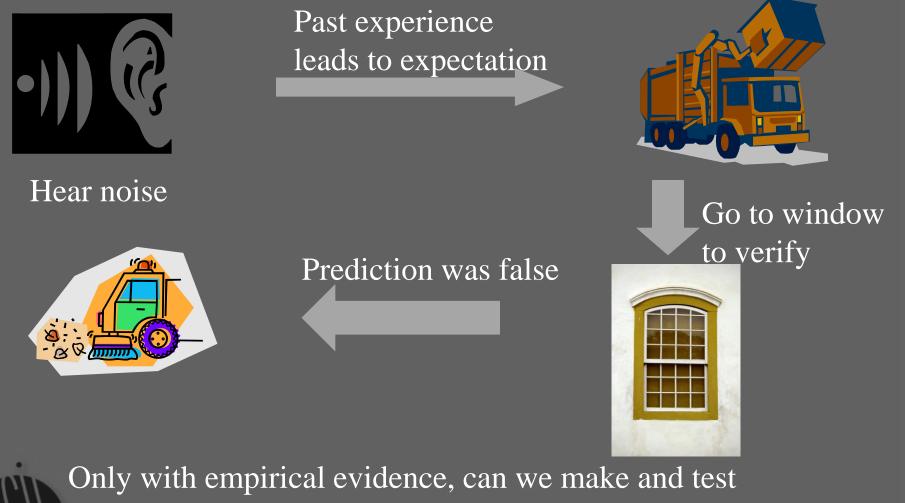


Understanding Humans

Activities Routines Intent **Causality**



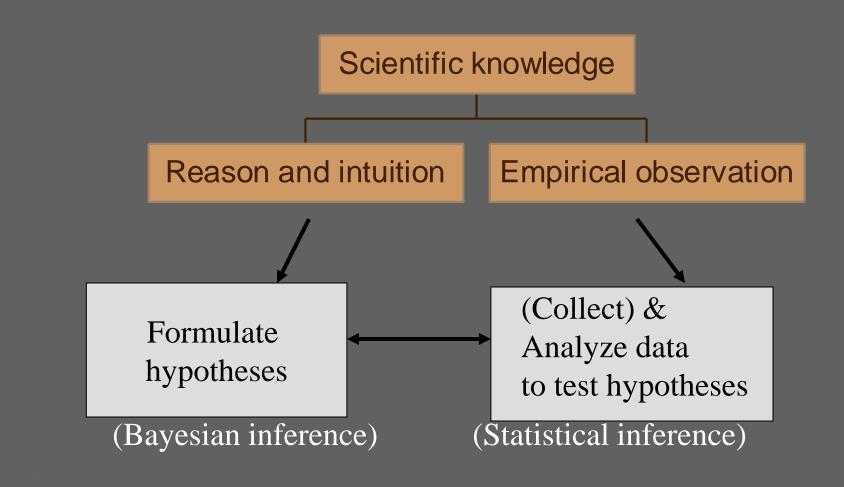
Causality



predictions

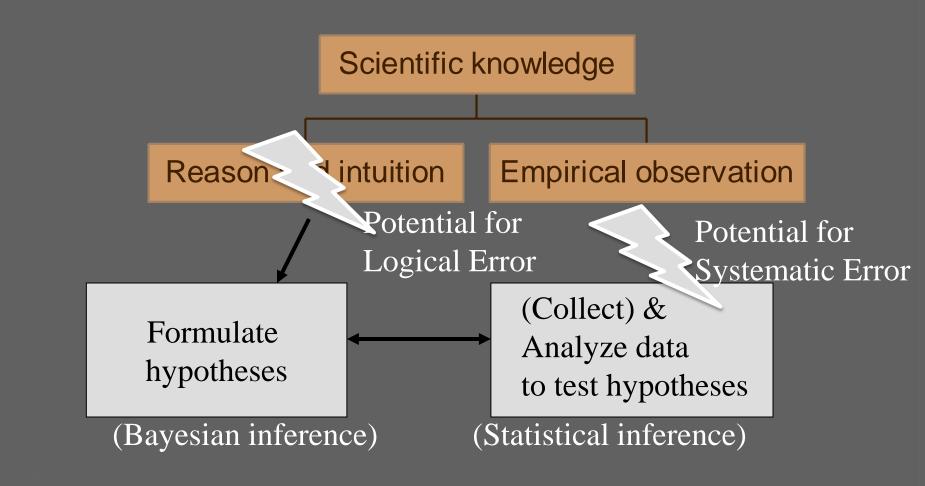
Adapted from class by Samantha Kleinberg

Process vs Frequency data



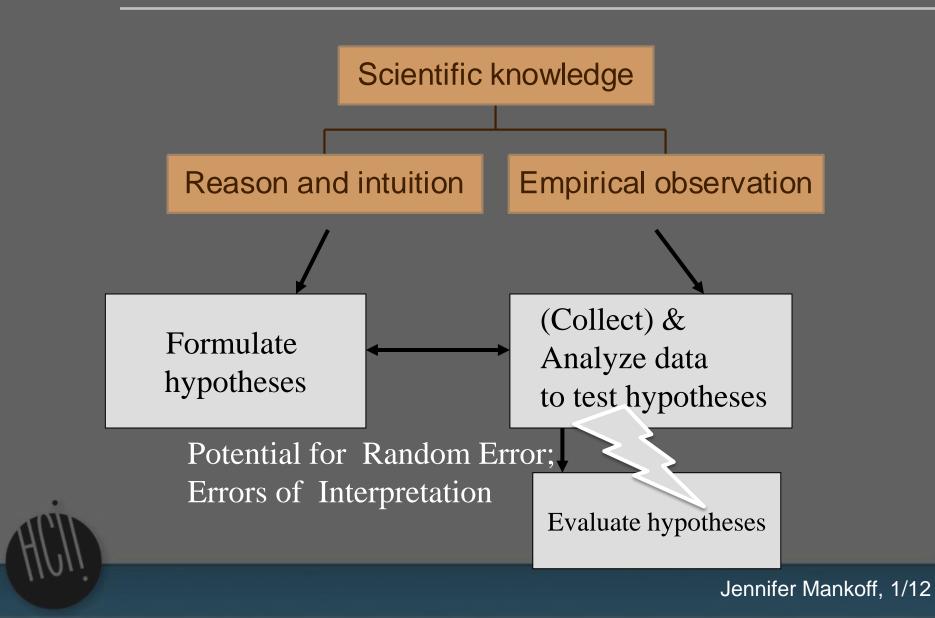


Process vs Frequency data





Process vs Frequency Data



Correlation is not Causation

There is a 0.91 correlation between ice cream consumption and drowning deaths.

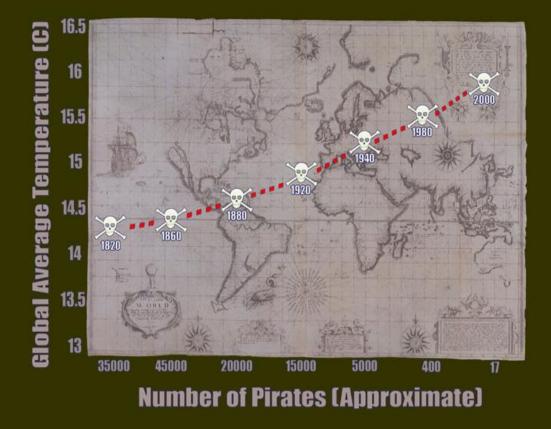
Does eating ice cream cause drowning? Does grief cause us to eat more ice cream?



Jennifer Mankoff, 1/12

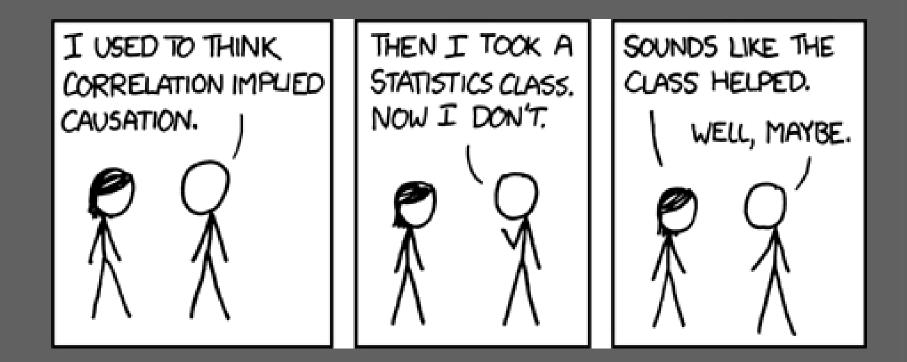
Correlation without causation (1)

Global Temperature Vs. Number of Pirates



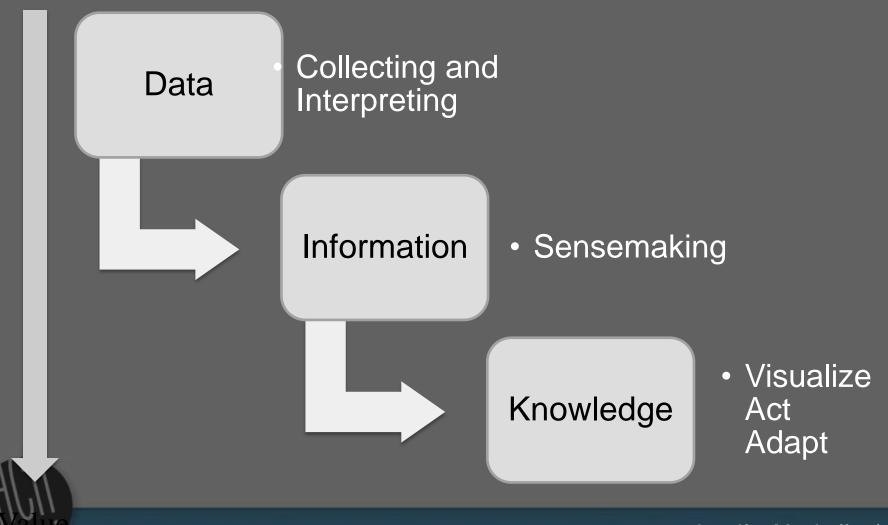
http://bama.ua.edu/~sprentic Jennifer Mankoff, 1/12

Correlation != Causality

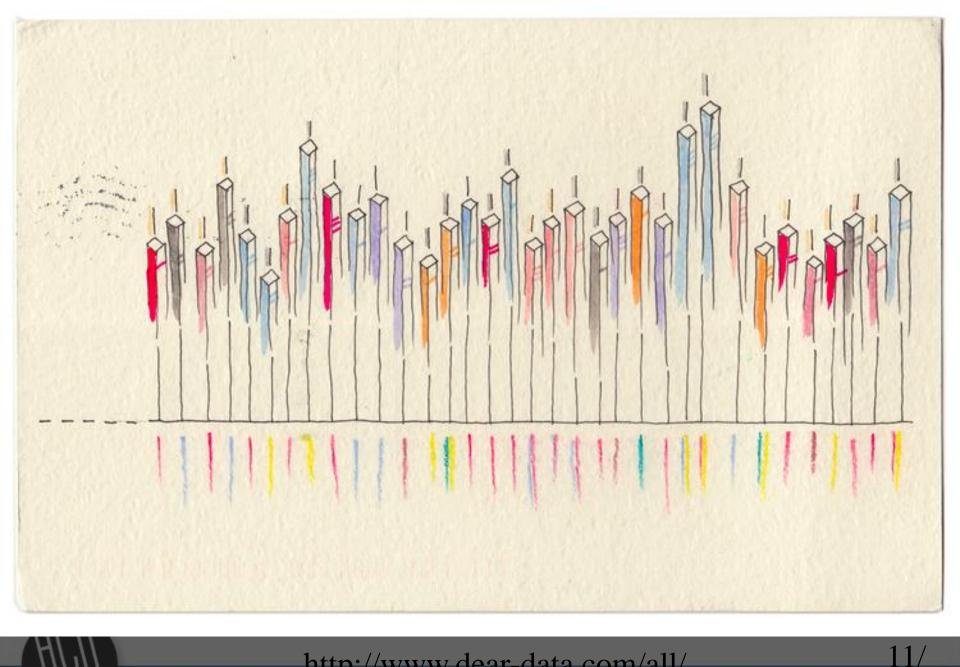


Jennifer Mankoff, 1/12 Adapted from http://www.fordham.edu/economics/Vinod/correl-regr.ppt

Making Data Actionable



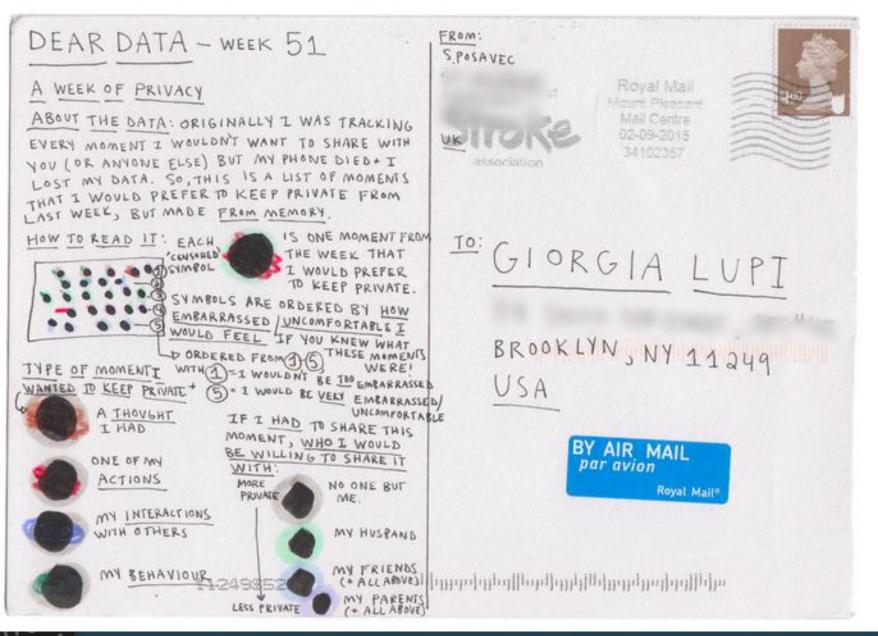
Jennifer Mankoff, 1/12



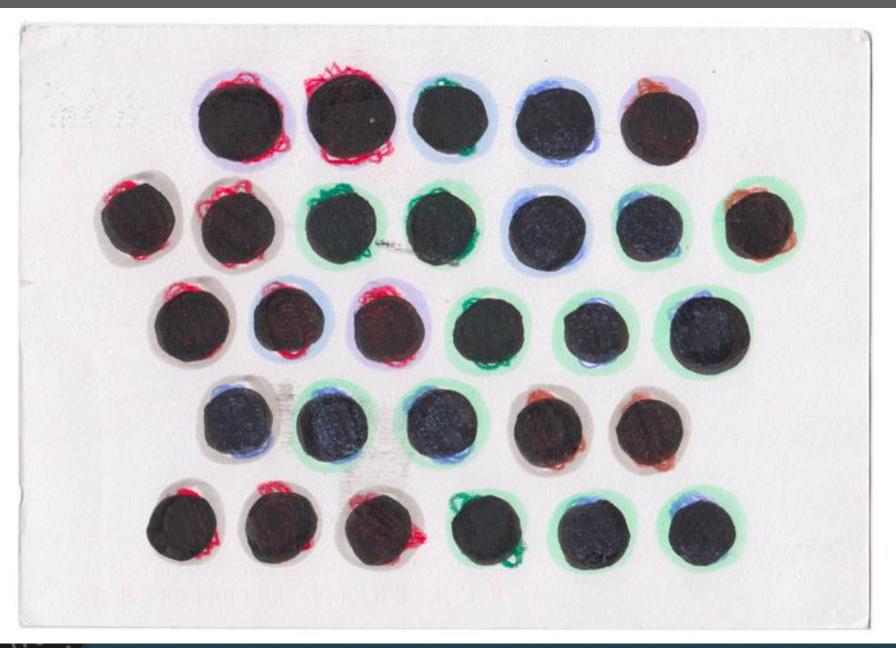
http://www.dear-data.com/all/



66 Deap Data trom WEEK 51: PRIVACY (ppeage!) FLUP something t DIDOCS something Thought HOW TO READ IT : BROOKLYN nponypelf -NY-USA this week I Tracked everything 2 Dio, thought HEIG TH= fie flected inpon that anninnar LEVEL I would' It wanna tell, TRADUCTION OF THE OWNER in chronologic order. PRIVACY COLOR and SEND To: da shed = LINES = lines = documenting the FIRST the Higher Massive DATA VOID in Dear what was The more Data = I forgot to track the STEFANIE POSAVEC I "ashamed Reserved who le Monday morning (about n Bottom lines = why did I want it be sECREt? generally ashamed London my Body/ rak Frunds/ fear people's physical Data my family habits mazement Thingp somebody would be the project's - UKattitude sensations my objessions selfishmess I am a terrible person England boyfrund/ the my drinks am scored what myatinde pantners aspect project nonia have . money - being grumpy my myattitude thonghts Ma as person being critated about the Behavior my thoughts with my fiture being Mean (3) mycoldness Boyfrund



Jennifer Mankoff, 4/12



What is Information Visualization?

Visualize: to form a mental image or vision of

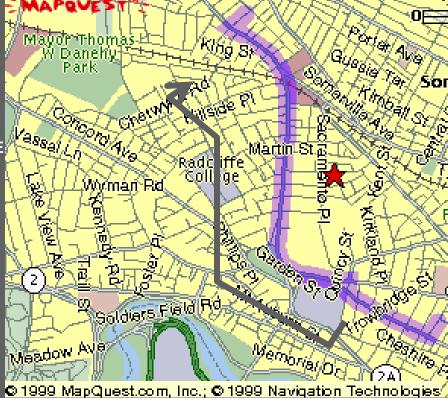
- Visualize: to imagine or remember as if actually seeing.
- American Heritage dictionary, Concise Oxford dictionary



. . .

The Power of Visualization

- 1. Start out going Southwest on ELLSWORTH AVE Towards BROADWAY by turning right.
- 2: Turn RIGHT onto BROADWAY.
- 3. Turn RIGHT onto QUINCY ST.
- 4. Turn LEFT onto CAMBRIDGE ST.
- 5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE
- 6. Turn RIGHT onto RUSSELL ST.



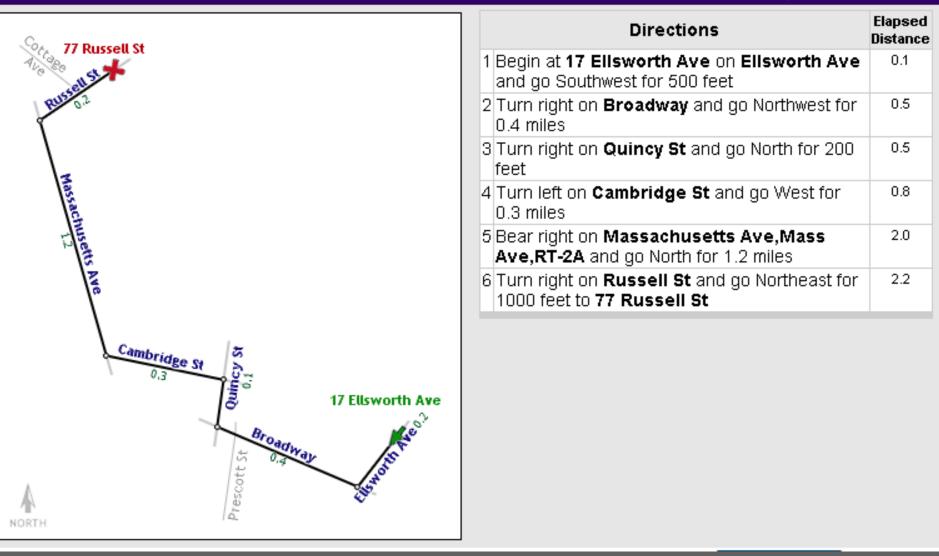


î rom

COM

Jennifer Mankoff, 1/12

The estimated travel time is 5 minutes for 2.16 miles of travel, total of 6 steps.

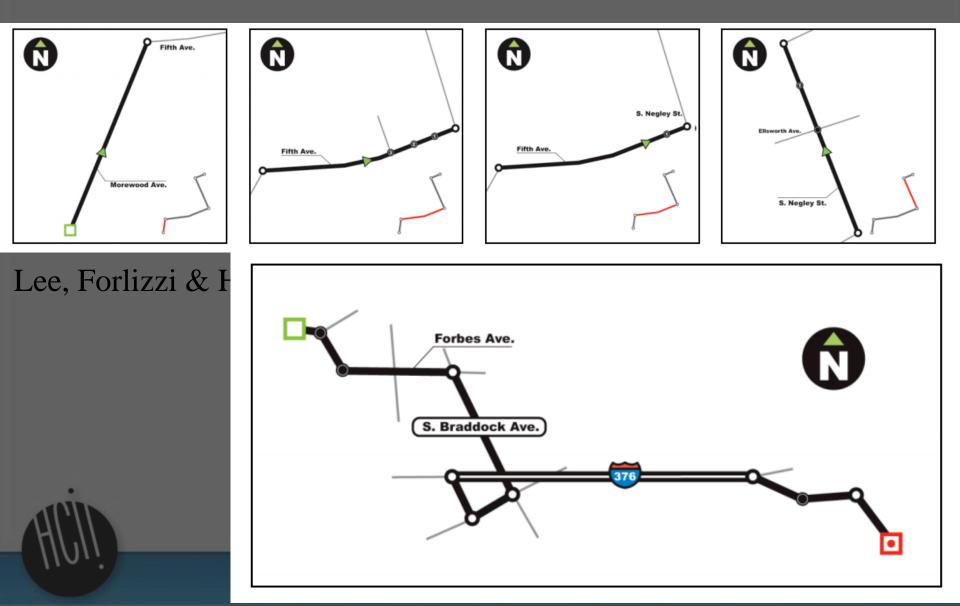




Line drive tool by Maneesh Agrawala http://graphics.stanford.ed

Jennifer Mankoff, 1/12

The Power of Design *and* interaction in Visualization



Planning a Visualization

- 1. What is its goal?
- 2. What visual queries does it support?
- 3. What are some compelling, <u>useful</u> examples? [COPY COPY COPY!]
- 4. Could it have been done more simply?

Making Queries

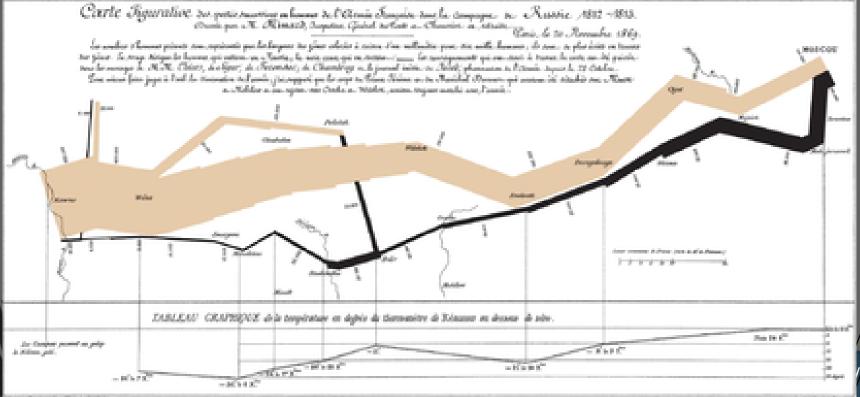
Define the query[ies] you wish to support "The special skill of designers ... [is] the talent to analyze a design in terms of its ability to support the visual queries of others..." Patterns ⇔ visual system Cognitive process prediction A continual fresh eye



Narrative in Visualization

Documenting the question you are answering

Leading the viewer through a story

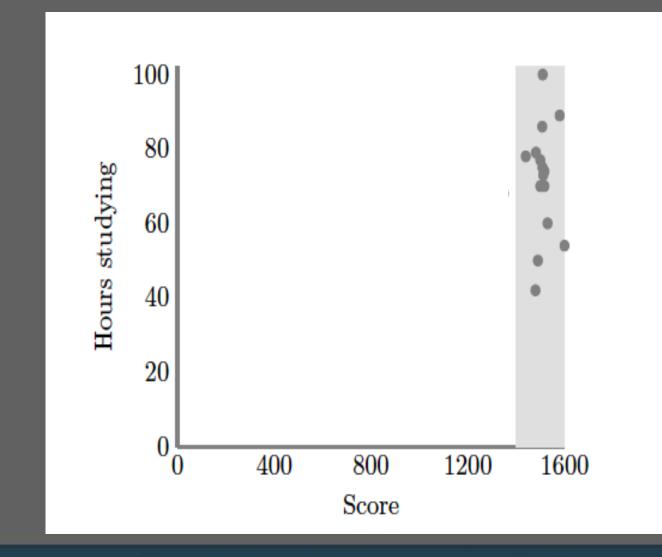


they are beginned for 27 their 2 1990 hours

by 14 April 1 Barris

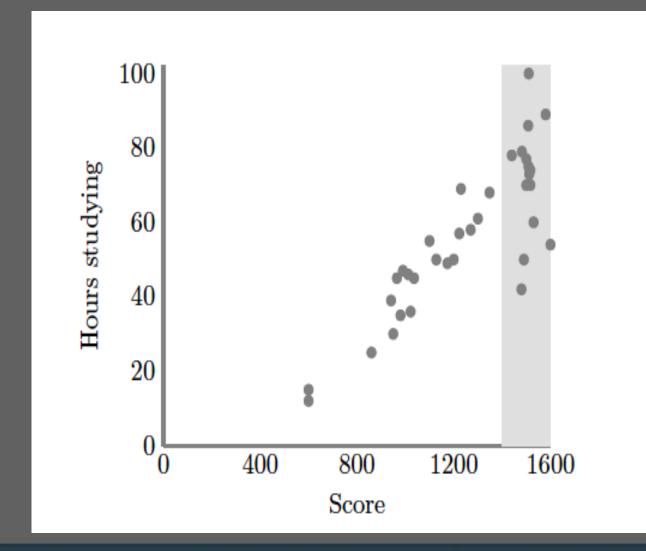
12

Visualizations freequently casue us to draw conclusions



S. Kleinberg (2015) Why: A Guide to Finding and Jersifeg Cates [12]

Which is why visualization choices are so important...



S. Kleinberg (2015) Why: A Guide to Finding and Jensifeg Makes [12]

Planning a Visualization

- 1. What is its goal?
- 2. What visual queries does it support?
- 3. What are some compelling, <u>useful</u> examples? [COPY COPY COPY!]
- 4. Could it have been done more simply?





Chart doesn't match data type

- If no colour appears, there is no such law on file
- 2012 election results Background check law Permit required to purchase Licence required to sell
- Records kept on file
- Firearms banned from workplace

Virginia

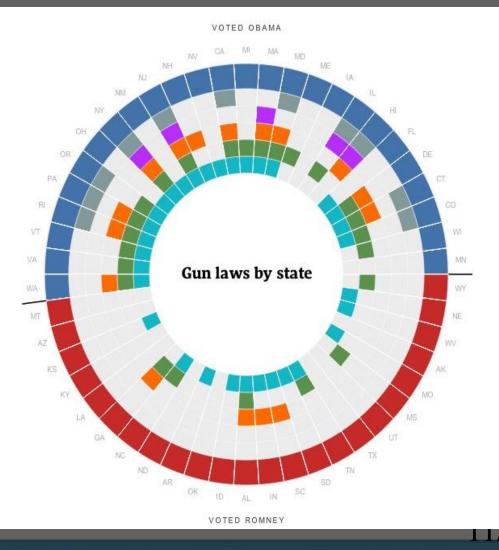
- Voted for Obama in the 2012 election Background check: not required for handguns Permit: not required to buy firearms Licence: not required for dealers Records: kept on file for handgun owners
- Workplace: firearms not allowed in parking lots

Overall gun control score: 12

Virginia has a Brady Campaign score of 12, which is lower than the national average of 16. The score comes from measuring these and other gun laws according to a weighted points system.

Murder rate: 2.58

There were 2.58 firearm murders per 100,000 people in Virginia during 2011, which is lower than the national average of 2.77. Overall, it is ranked #27 in murder rates out of 48 states with this data.



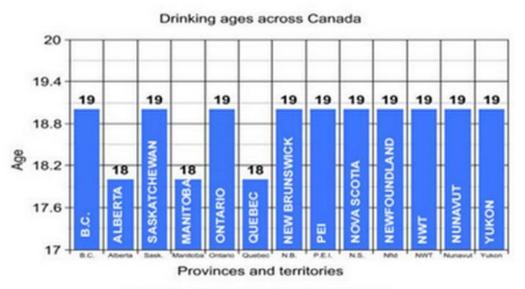
Jennifer Mankoff, 4/12

http://www.businessinsider.com/the_27_worst_charts_of_all_

Confusing Axes

Drinking age will remain 19 in Saskatchewan

CBC News Posted: Mar 4, 2013 11:59 AM CST | Last Updated: Mar 4, 2013 11:55 AM CST 🛄 25







Canadian Centre on Substance Abuse

You have to be 19 in Saskatchewan to have a drink, while in Alberta and Manitoba, the drinking age 18. (CBC)

The Saskatchewan Party government has ruled out lowering the drinking age, four months after party members put the issue in the public eye.

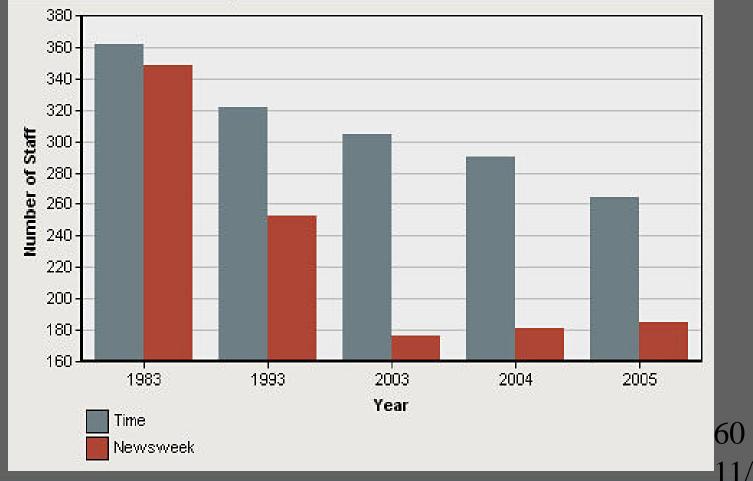
Jennifer Mankoff, A/12

http://www.businessinsider.com/the_27_worst_charts_of_all

Misleading Axis

NEWS MAGAZINE STAFF SIZE OVER TIME

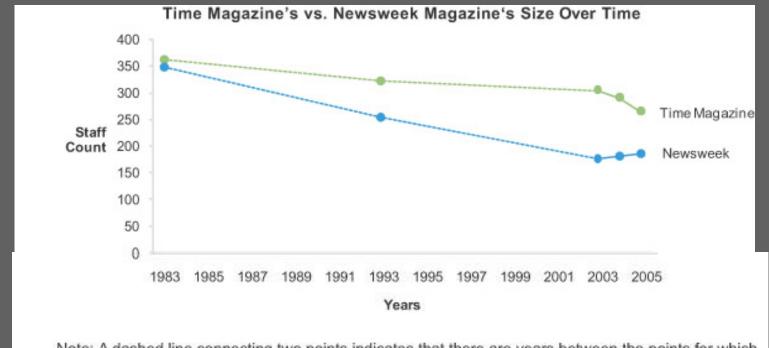
Time and Newsweek select years 1983 - 2005



Jennifer Mankoff, 4/12

https://www.percentualedge.com/example1/1.nhp

Improved



Note: A dashed line connecting two points indicates that there are years between the points for which values were not available. If the values were available, the shape of the lines might vary significantly.

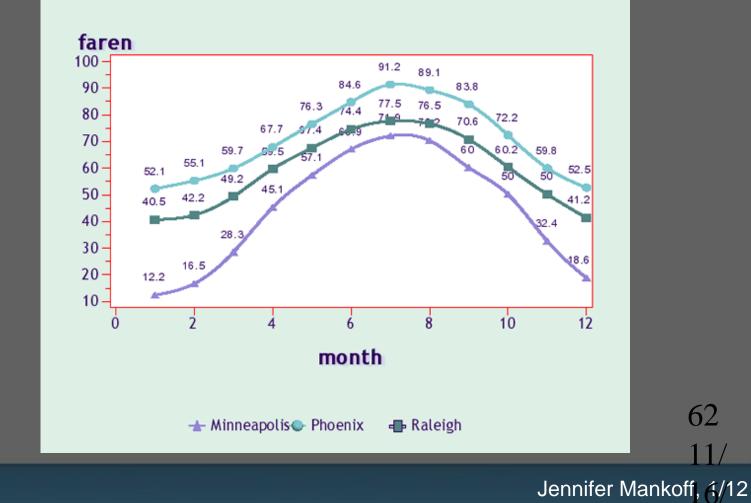
61

Jennifer Mankoff, A/12

https://www.percentualedge.com/example1/1.php

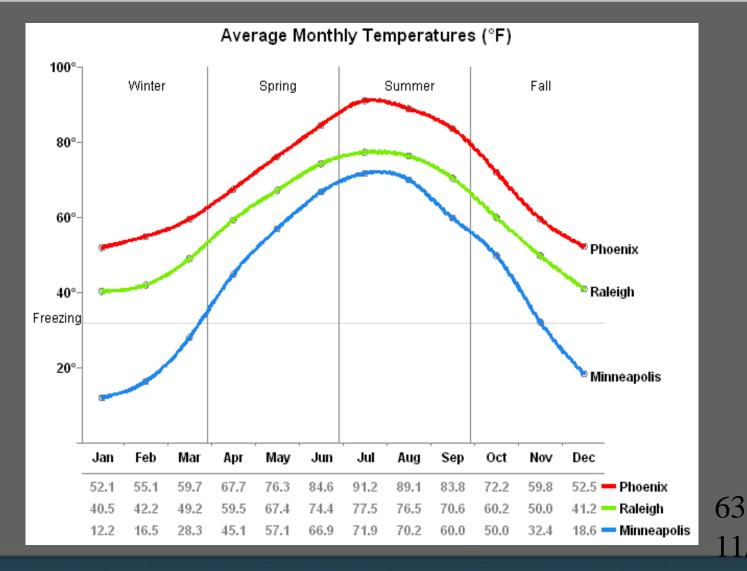
Overly Complex

Average Monthly Temperature



https://www.percentualedge.com/example?.nhp

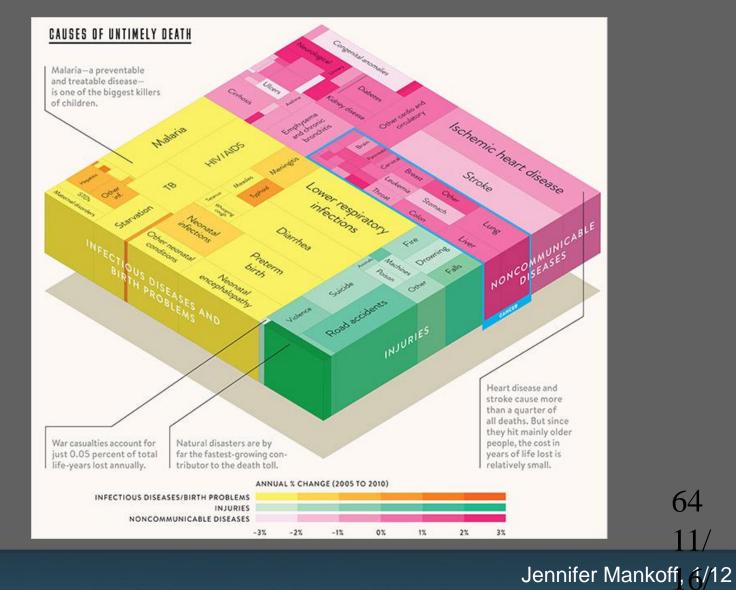
Improved



Jennifer Mankoff, 4/12

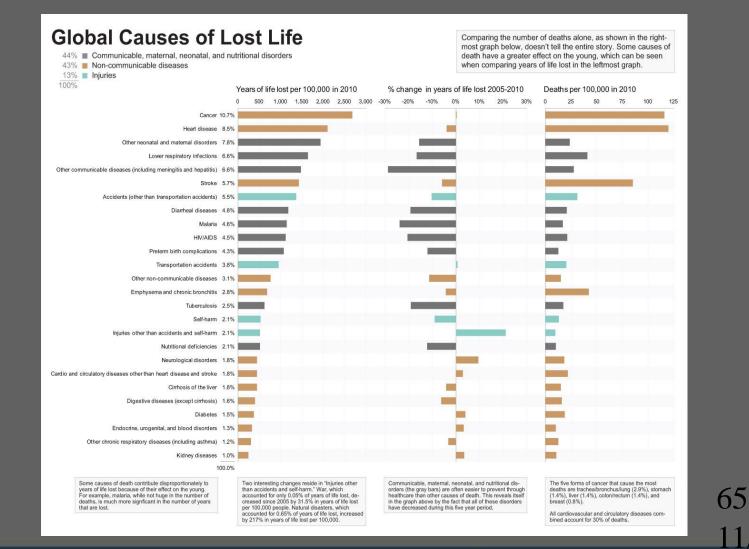
https://www.percentualedge.com/example? nhp

Hard to read



https://www.percentualedge.com/example?0.php

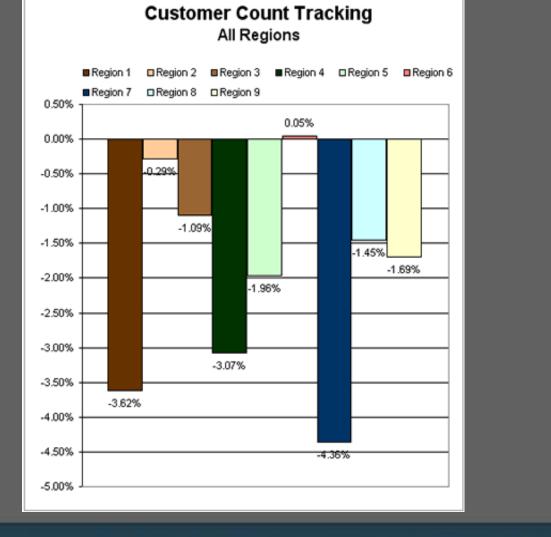
Improved



Jennifer Mankoff, 4/12

https://www.perceptualedge.com/example?0.php

Comparisons Difficult

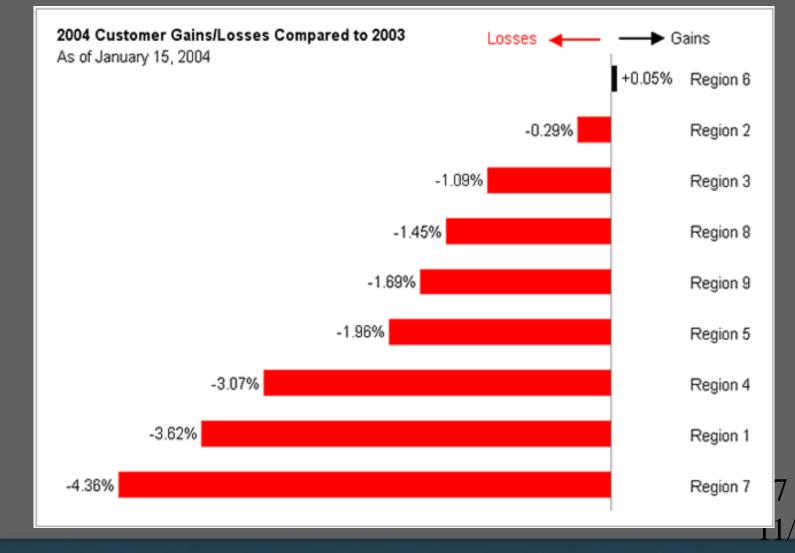


Jennifer Mankoff, 4/12

66

https://www.percentualedge.com/example1.php

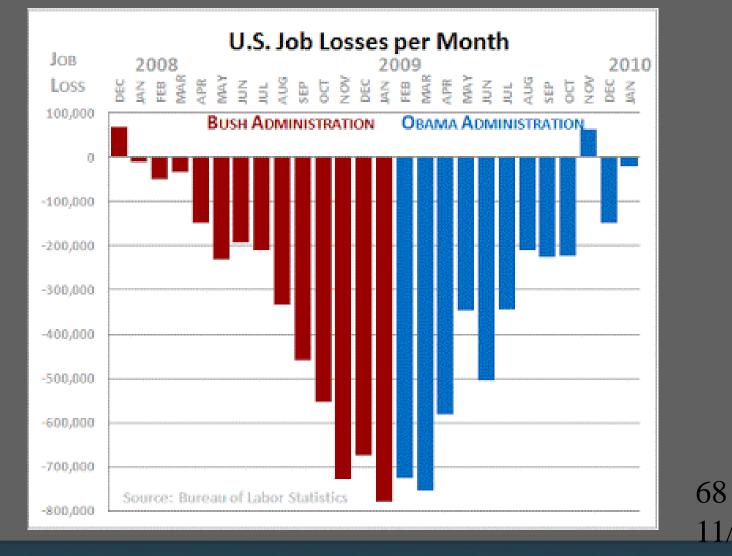
Improved



Jennifer Mankoff, 4/12

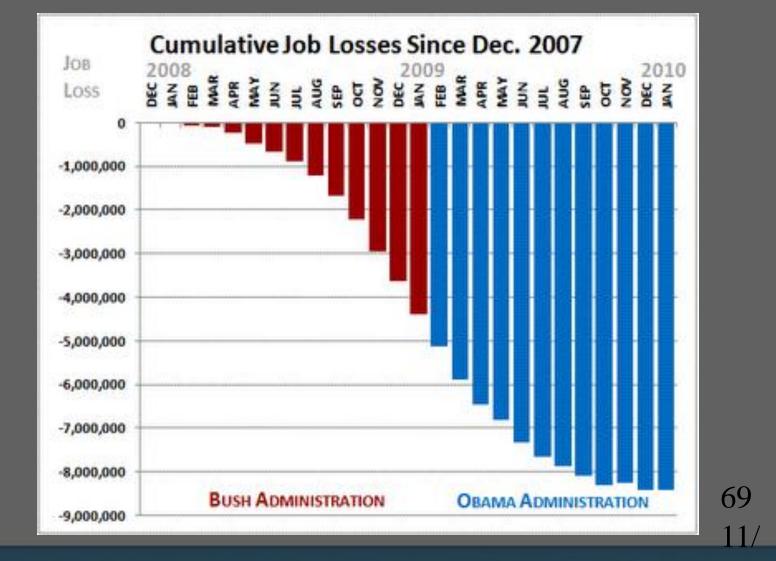
https://www.percentualedge.com/example1.php

Highlighting wrong aspect of data?



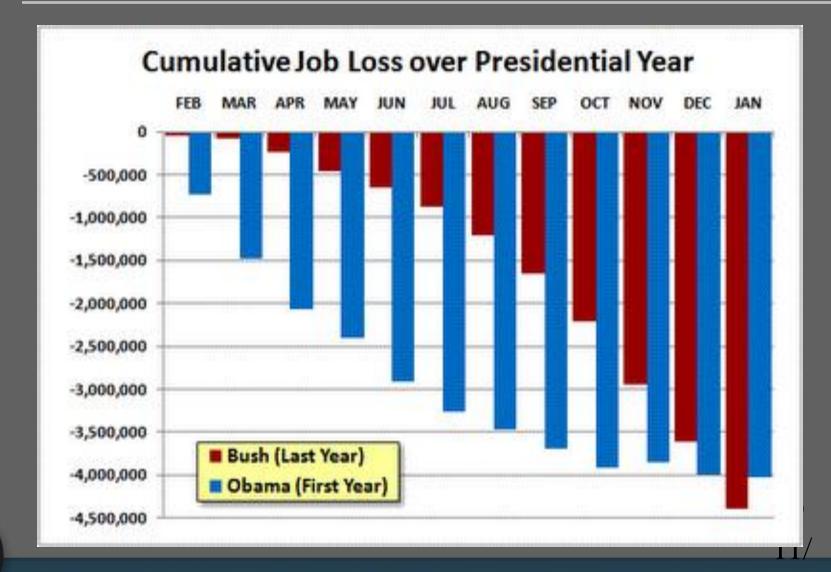
Jennifer Mankoff, A/12

http://socuelbythecreek.blogspot.com/2010/02/what_does_



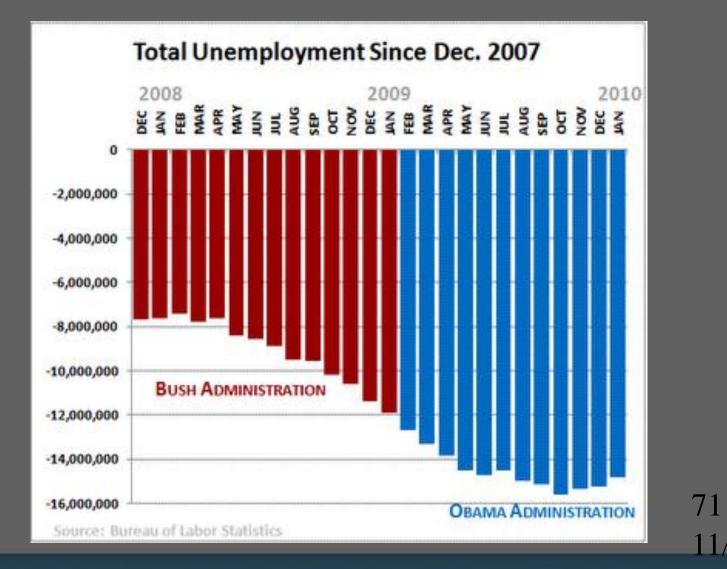
Jennifer Mankoff, 4/12

http://socuelbythecreek.blogspot.com/2010/02/what_does_



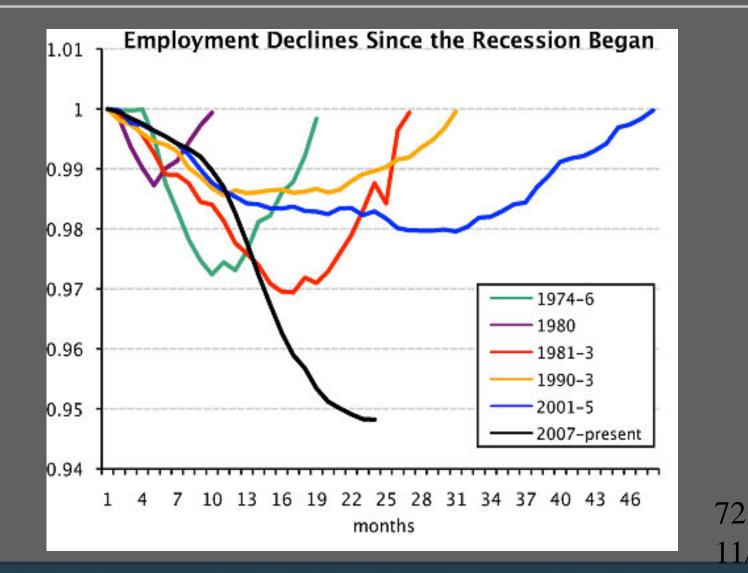
Jennifer Mankoff, A/12

http://soquelbythecreek_blogspot_com/2010/02/what_does_



Jennifer Mankoff, 4/12

http://soquelbythecreek.blogspot.com/2010/02/what-does-



Jennifer Mankoff, 4/12

11/

http://socuelbythecreek blogspot.com/2010/02/what.does

Interactive Visualization



Jeffrey M. Rzeszotarski and Aniket Kittur. 2014. Kinetica: 73 naturalistic multi-touch data visualization. In Proceedings of 11/ the 32nd annual ACM conference on Human factors In^{Jennifer Mankoff, 6/12}

Interactive Visualization

http://queue.acm.org/detail.cfm?id=2146416



Jennifer Mankoff, 4/12

74

Visualizing Big Data

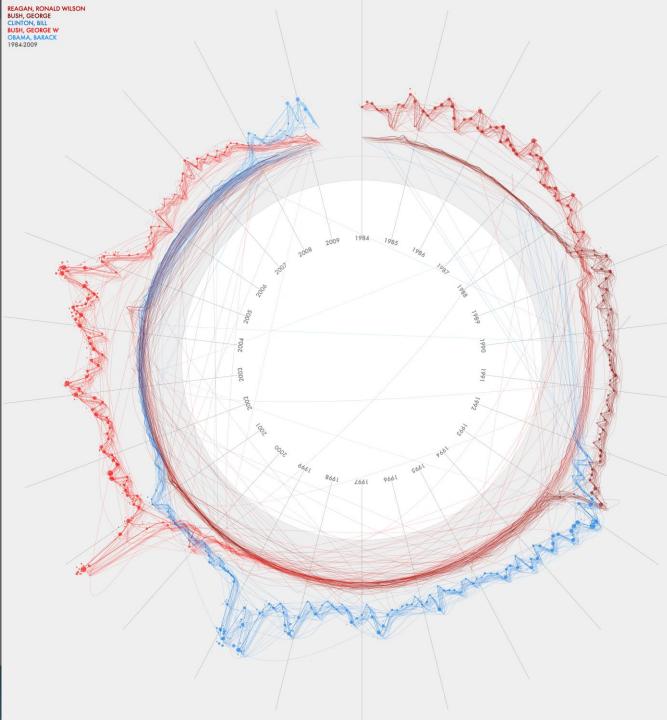
First a tour Then some techniques

http://www.smashingmagazine.com/2007/08/0 2/data-visualization-modern-approaches



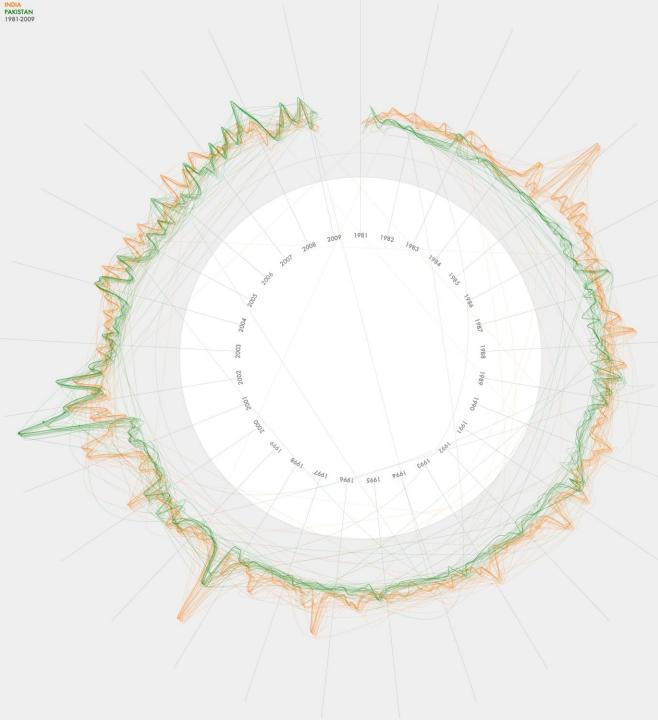
Jennifer Mankoff, 4/12

75



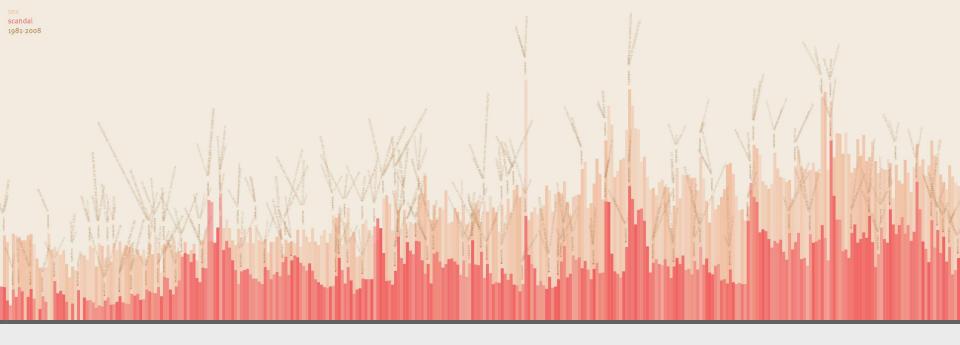
Jer Thorp Data Artist

This graph charts the frequency of mention of the five **US** Presidents between 1984 and 2009. It also indicates weighting of stories - the darkest line shows front page stories while the lighter lines indicate stories buried deepersin, the



Jer Thorp Data Artist

NYTimes Threads -India & Pakistan This graph charts the frequency of articles mentioning India and Pakistan in the NYT between 1981 and 2009. It also indicates weighting of stories - the darkest line shows front page stories while the lighter lines indicate stories buried deeper in the naner



Jer Thorp Data Artist

NYTimes: Sex & Scandal since 1981 This is a visualization of the frequency of occurrence of the words 'sex'



Race in LA (Eric Fi

Nice narrative @lamag.com

http://www.lamag.com/features-hidden/race-in-la-see-how-weve-grown/#0412infographic

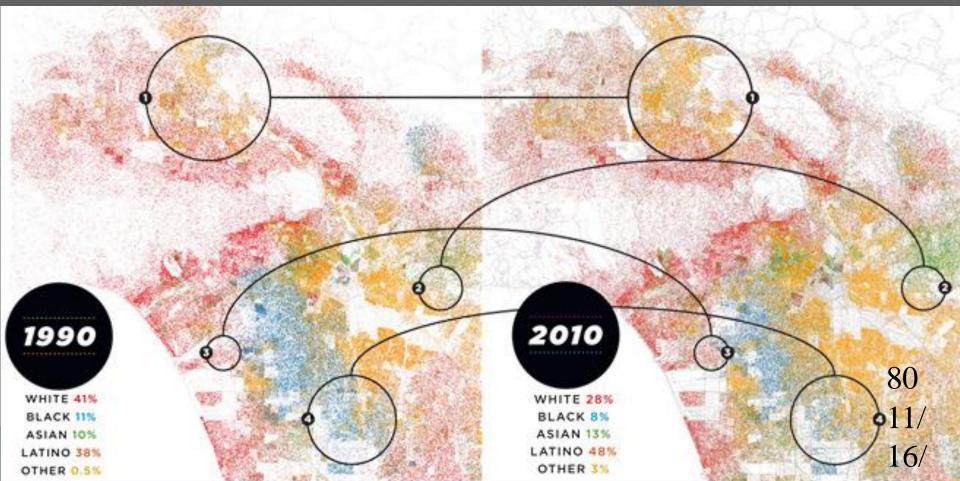
WHITE

BLACK

ASIAN

LATINO

OTHER



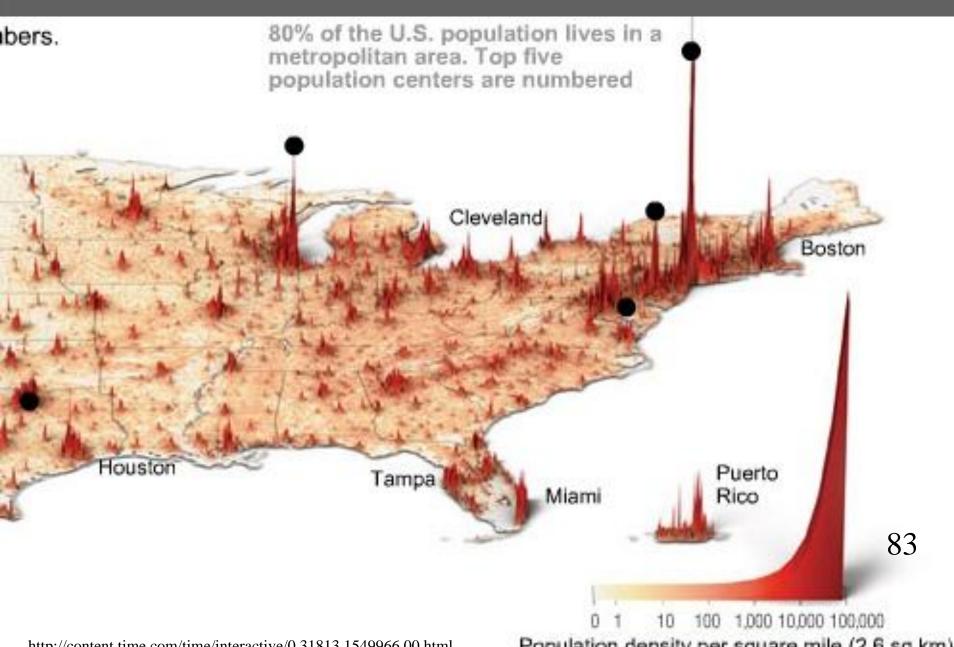
World travel and communications recorded on Twitter Green is physical movement from place to place; purple is @replies from someone in one location to someone in another; combining to white where there is both.

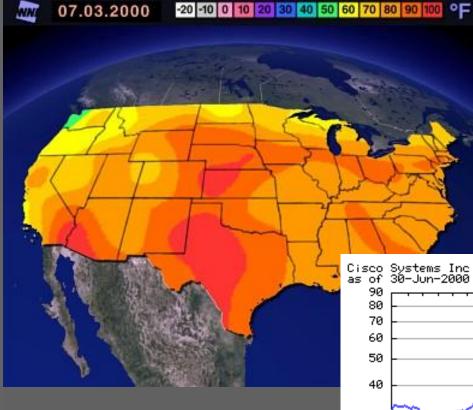
Reported trips to Null Island excluded; all other geotags trusted. Endpoints of trips are real data; routes in between are fabricated. Brightness is logarithmic.

Data from the Twitter streaming API through September 1, 2011. Continent shapes from Natural Earth. Author: Eric Fischer https://www.flickr.com/photos/walkingsf/6635655755/in/photostream/

Tourist vs locals in Barcelona. Shows. Blue photos (Flickr) live in the city, red are tourists, yellow are unknown. Eric Fischer again 82 More_{koff, (}/12 dataile

Time Magazine





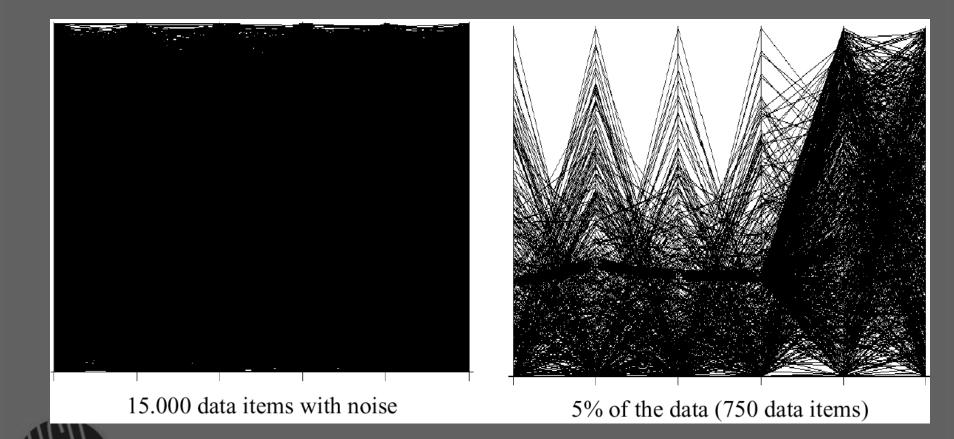
Small Data?



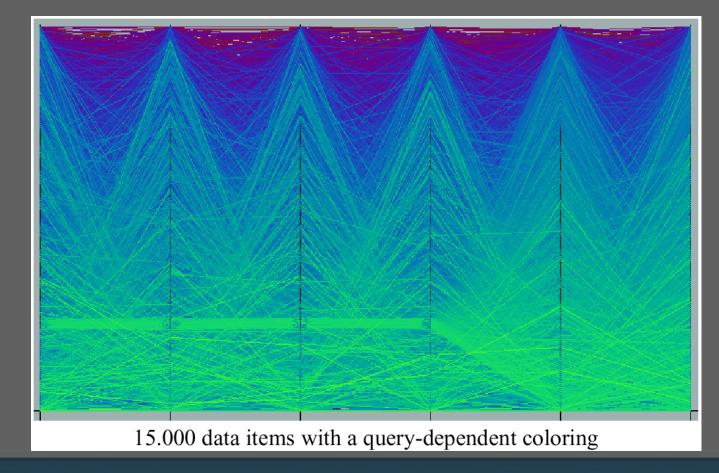
Jennifer Mankoff, 1/12

Images

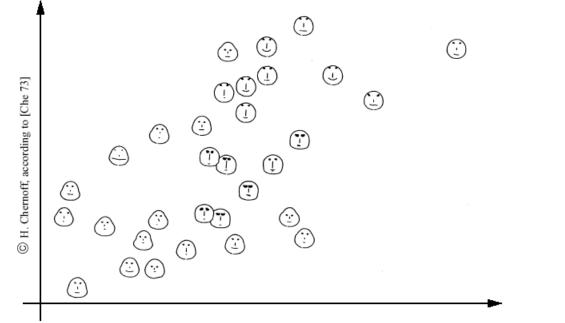
Visualization Techniques for Big Data



Query Dependent Coloring

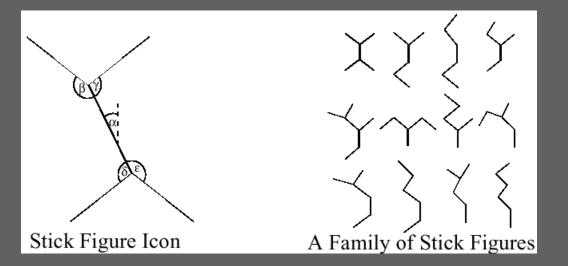


Chernoff-faces



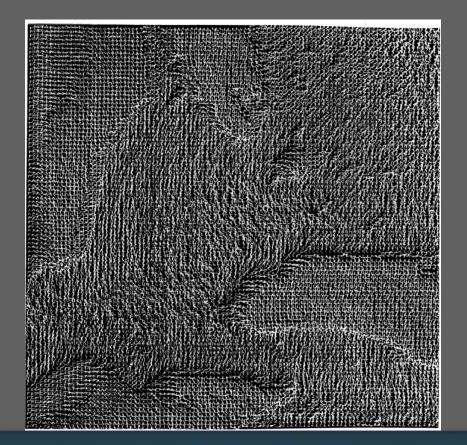
visualization of the multidim. data using the properties of a face icon (shape of nose, mouth, eyes, and the shape of the face itself)

Stick Figures



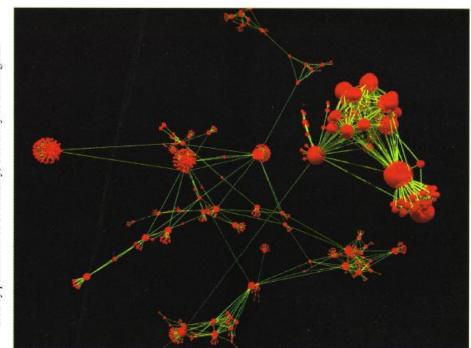






Graph-based Techniques

Narcissus [HDWB 95]



visualization of a large number of web pages

used by permission of B. Hendley, University of Birmingham

visualization of complex highly interconnected data (e.g., graphs such as the web)

Zoomable Ulsttp://gigapan.com/gigapans/4304

560

AN Emergent Mosaic OF Wikipedian Activity

Reserved Hammable Memium in 2007 Notici Visualizing Network Dynamics Comp https://www.indiana.edu/97/petrof/

AUTHOR

Bruce W. Herr (Vitualization Expert) Tada M. Hollown, (Dan Mining Expert). Etilda Handy (Corphic Design) Kayi Bitaru (Ashiwa & Spannor) Kevin Boynda, Sandia Lafo (Node Layrou) School of Library and Taformation Science: Deparament at Company. Cristication Indiana University Biomington, 10: 47405, USA (bhrrachollow,chiang)katry (giradiana,edu Hoyackij-andia gow

INTRODUCTION

This energetit massic supplies a macroview of all of the English Wikipedia (http://n.wikipedia.org) and reveals those areas that are currently boe, meaning, of lase, they are being forquently revised.

ARTICLE NETWOR

The provided dataset [1,2] comprises 659,388 interconnected Wikipedia articles. One article (node) is connected in another if it links to it. There are 16,582,425 links.

POSITIONIN

Articles are shown close to one another if they are similar, and far apart if they are different. Two articles are said to be similar if articles that link so one often link to the other.

MCISAL

The L369 image were taken directly from Wikipedia. This represents approximately one image for every 300 articles. The images were selected automatically in a three way process. First, the layout was cut into ball inch first. Next, the heath of the were tanked by their indegree, i.e., the number of articles that link to them. Finally, the first image of the highest ranked article that commiting a non-icon image was selected for display. All images of controversial subject matter were lept making it appropriate for matter audiretus.

The means in the local agent when it would get the theory of the scher it would get the theory agent agent agent agent theory of the schere agent the schere agent theory of the schere







KYT AREAS

From February 6th, 2001 to April 6th, 2007, articles were allisted 52,300,922 times. The red, larger nodes are those articles that have been revised more frequently than those shown as small, yillow nodes, We gave more consideration is current and major revisions. The result is a map sevenling articles which have been farinarily service lately.

Top 20 Most Animity Revised Article

1. lesus 2. Adolf Hitler 3. October 2003 4. Nintendo Revolution 5. Hurricane Katrina 6. India 7. RuneScap 8. Anarchism 9. Britney Spears 10. PlayStation 3 11, Saddam Hussein 12. Japan 13. Albert Einstein 14. 2004 Indian Ocean Earthquake 15. New York City 16. Germany 17. Mahammad 18. Pope Benedict XVI 19. Ronald Reagan 20. Hinduism

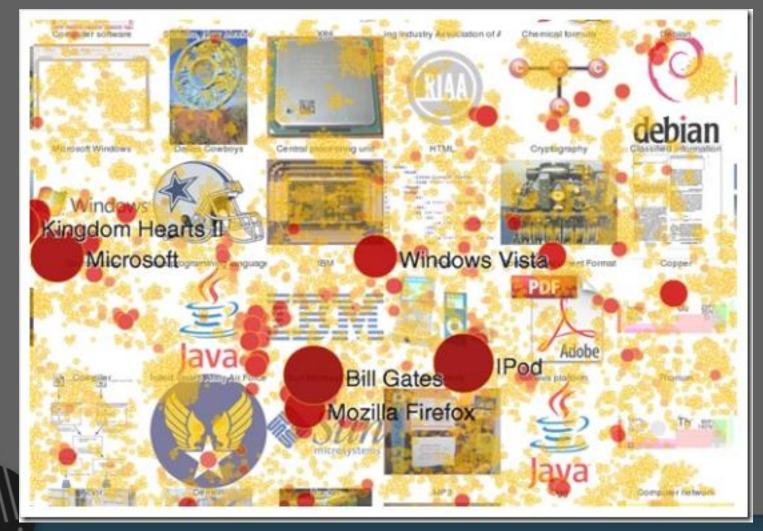
Jennifer Mankoff, 1/12

NONDOT USE OF THE ADDRESS OF THE



91

Zoomable UIsttp://gigapan.com/gigapans/4304



92

Distortion Techniques

Basic Idea: Distortion of the image to allow a visualization of larger amounts of dataAn alternative to zoomable Uis (or complement)



Distortion Techniques

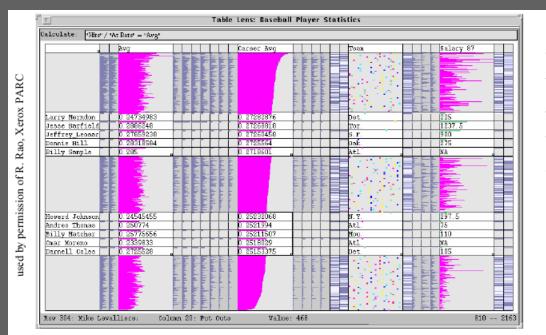
Basic Idea: Distortion of the image to allow a visualization of larger amounts of data

Simple:

Perspective Wall [MRC91] Bifocal Displays [SA 82] TableLens [RC94] Graph. Fisheye Views [Fur 86, SB94] Hyperbolic Repr. [LR94, LRP95] Complex: Hyperbolic Repr. [LR94, LRP95] 3D-Hyperbolic Repr. [MB95] Hyperbox [AC91]



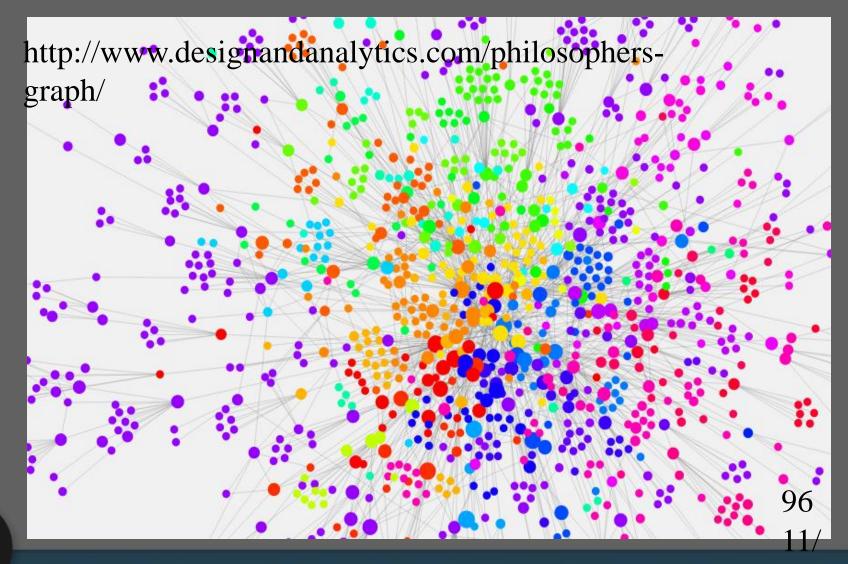




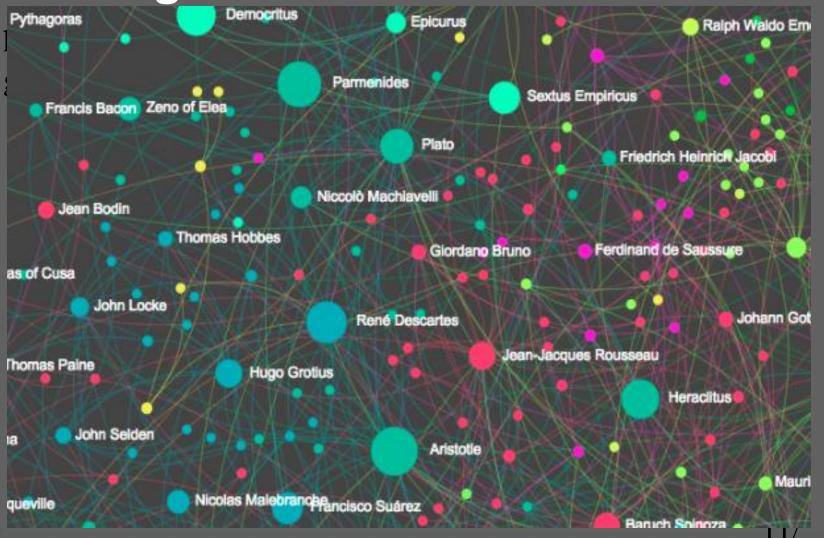
visualization of a baseball database with a few rows being selected in full detail

Compact visualization of a table (spreadsheet / database) with the possibility of viewing portions of the table in more detail

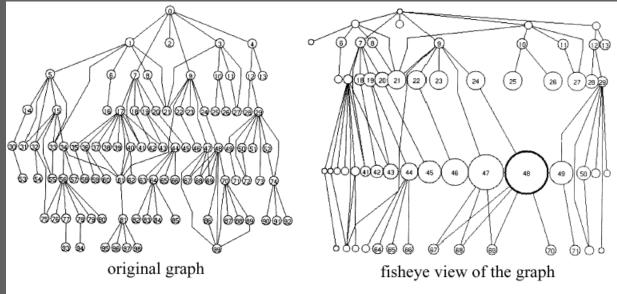
Networks of Information



Networks of Information – with zooming

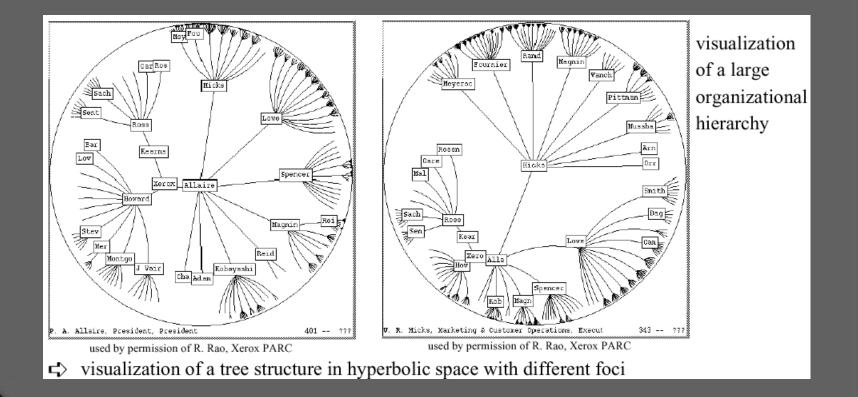


Fisheye View

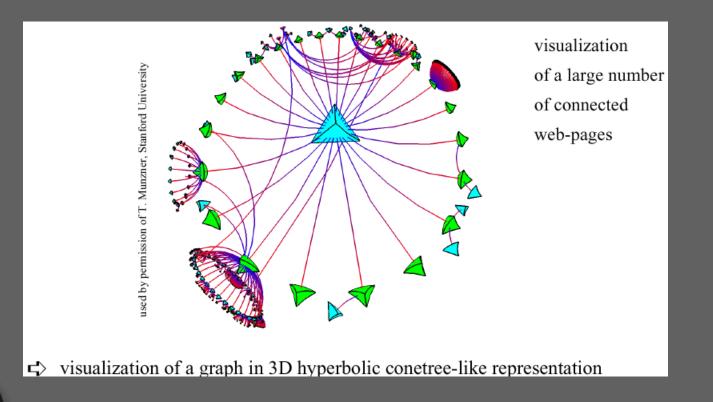


- r graph visualization using a fisheye perspective
- Shows an area of interest quite large and with detail and the other areas successively smaller and in less detail

Hyperbolic Tree



3D Hyperbolic Representation



Summary

Make use of every pixel

Use color and location to break the data up and allow the viewer to easily filter

Give a sense of things and use zooming for detail

Add a dimension such as time or height for a key variable

Allow exploration through distortion, filtering, highlighting, and linking

Exploit hierarchy and connectivity

Product Tree Man of Renin (Exports) 2009 by Mostol from

101

Applied Data Use: StepGreen

Capture Self reported data on behavior Motion & GPS Temperature & Energy Know **Green Actions Transportation choices Appliances in Use** Adapt and Act Visualize Expose

. . .

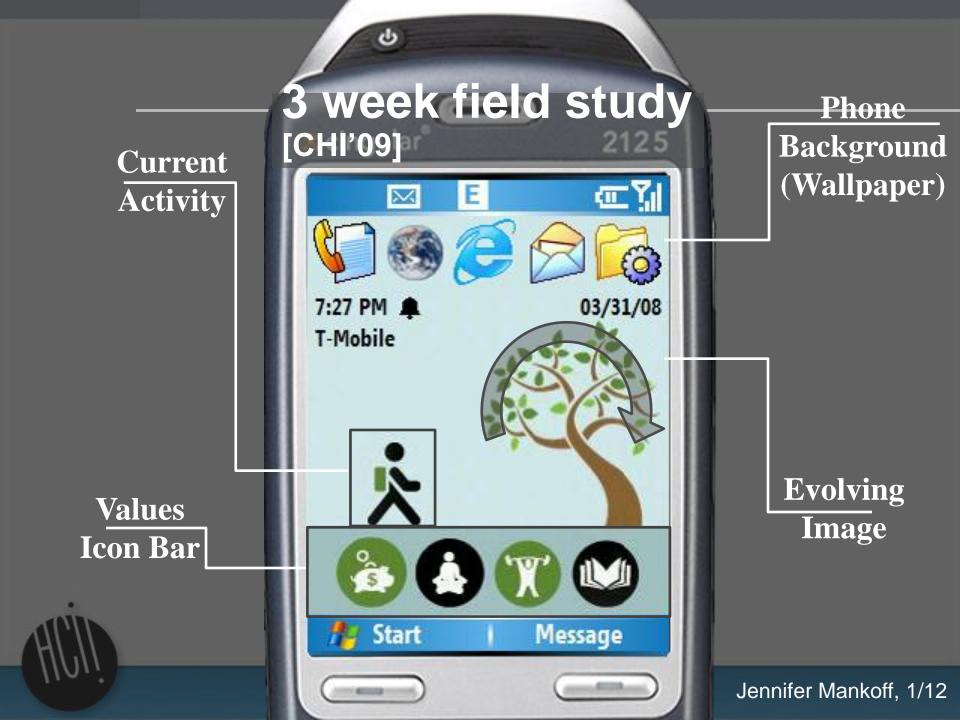


Applied Example: Ubigreen

Capture Self reported data on behavior **Motion & GPS Temperature & Energy** Know **Green Actions Transportation choices Appliances in Use** Adapt and Act Visualize **Expose**

. . .





Engagement

"It's omnipresent"

cingula

00

T-Mobile

Start

-

2125

Con Ya

Contacts

-

- Participant 9

- Participant 8

"I want to have different stories every week ... to maintain curiosity in the app"

Real-life game

One participant complained that when a trip hadn't been automatically recorded, "I felt like I was being cheated out of my 'points" - Participant 15

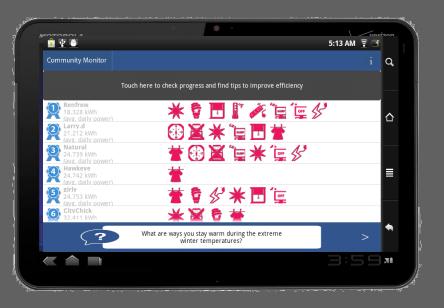
"Some people at work knew about the polar bear and every day they asked me about it. 'Did you get a seal today?"

- Participant 14

Longer Term, Real-World Deployments







Other Application Areas

Health

Self reported data on symptoms and conditions

Internet data: Extract argument Most people who report any of these symptoms report all three. Very few people have some, but not all at once.

structures & enhance search

Forums: predict expertise, highlight time on site, etc.

hxx Results

The Combination of Weight loss, Abdominal pain, and Dizzyness Has Been Reported

60 people reported experiencing at least one of the symptoms you searched for

46 people experienced all three

A Combination of Weight Loss and Abdominal Pain s Most Common in Crohn's Disease

Of the four conditions matching these symptoms, Crohn's Disease, Celiac, Cirrhosis, and Lymphoma, Crohn's is the most common and is the most often reported with both weight loss and abdominal pain, rather than one or the other. Dizzyness is only reported in Cirrhosis and Lymphoma

Summary: Making Data Actionable

