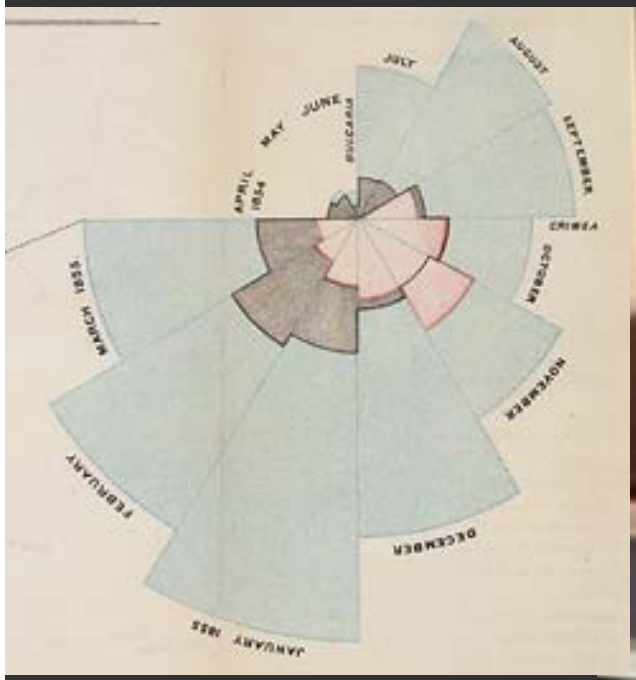
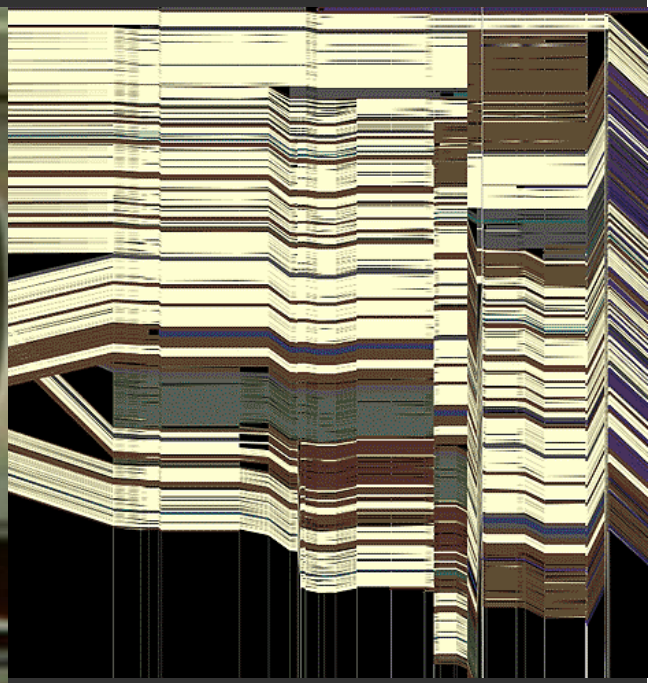
Interactive Visual Analysis for Networks & Text





Jeffrey Heer Stanford University

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

u_X = 9.0 σ_X = 3.317

u_Y = 7.5 σ_Y = 2.03

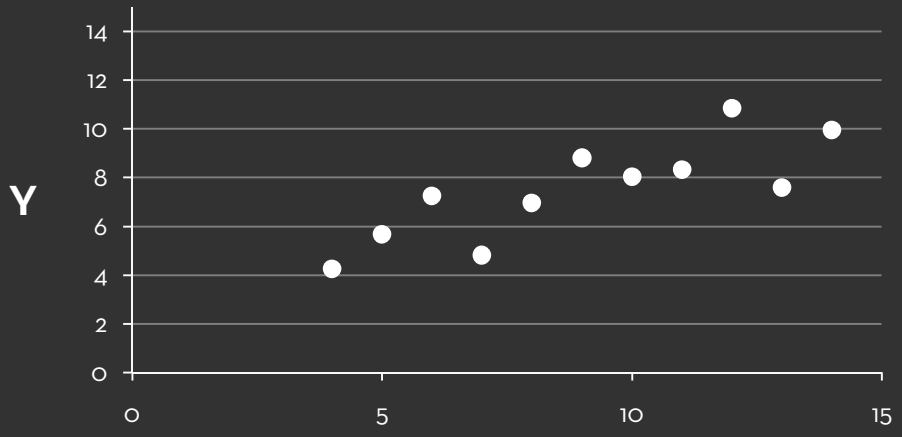
Linear Regression

Y^2 = 3 + 0.5 X

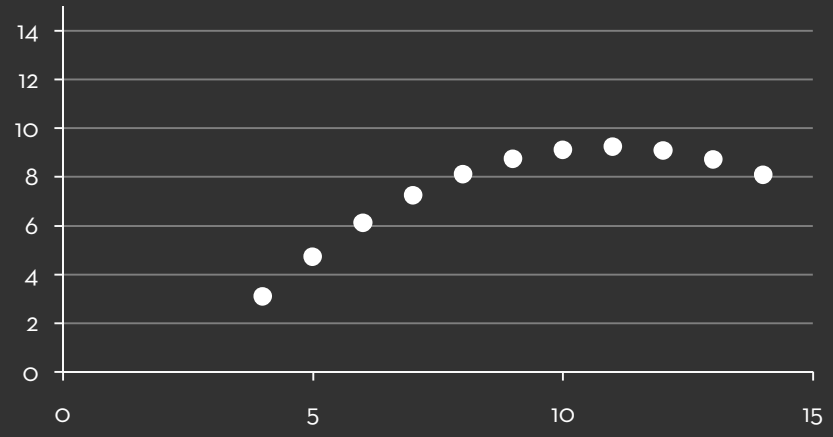
R^2 = 0.67

[Anscombe 73]

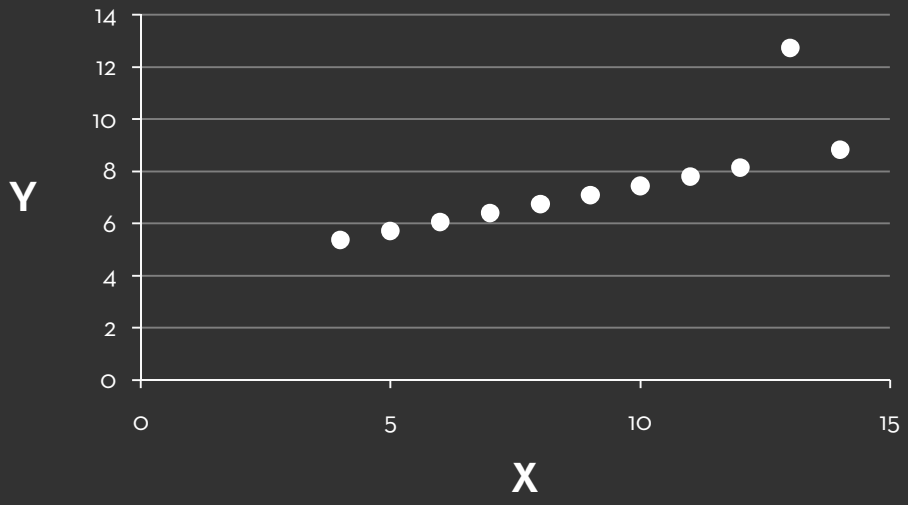
Set A



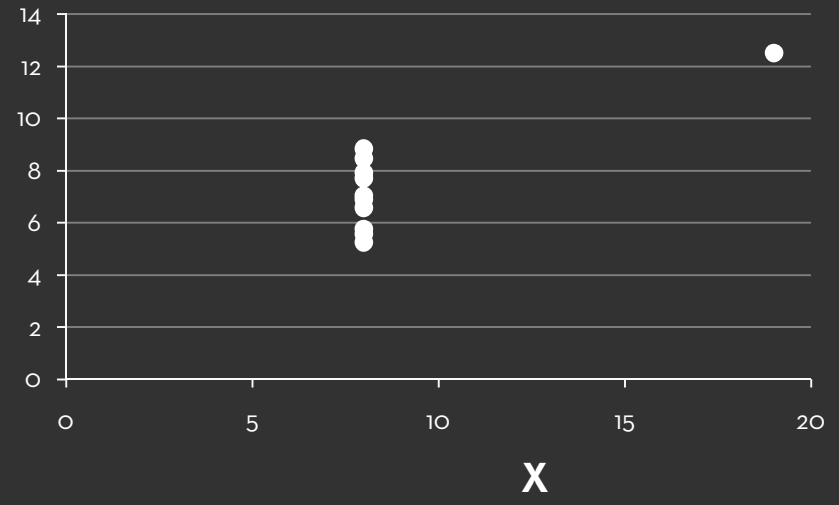
Set B



Set C



Set D



How much data (bytes) will we produce in 2010?

2010: 1,200 exabytes

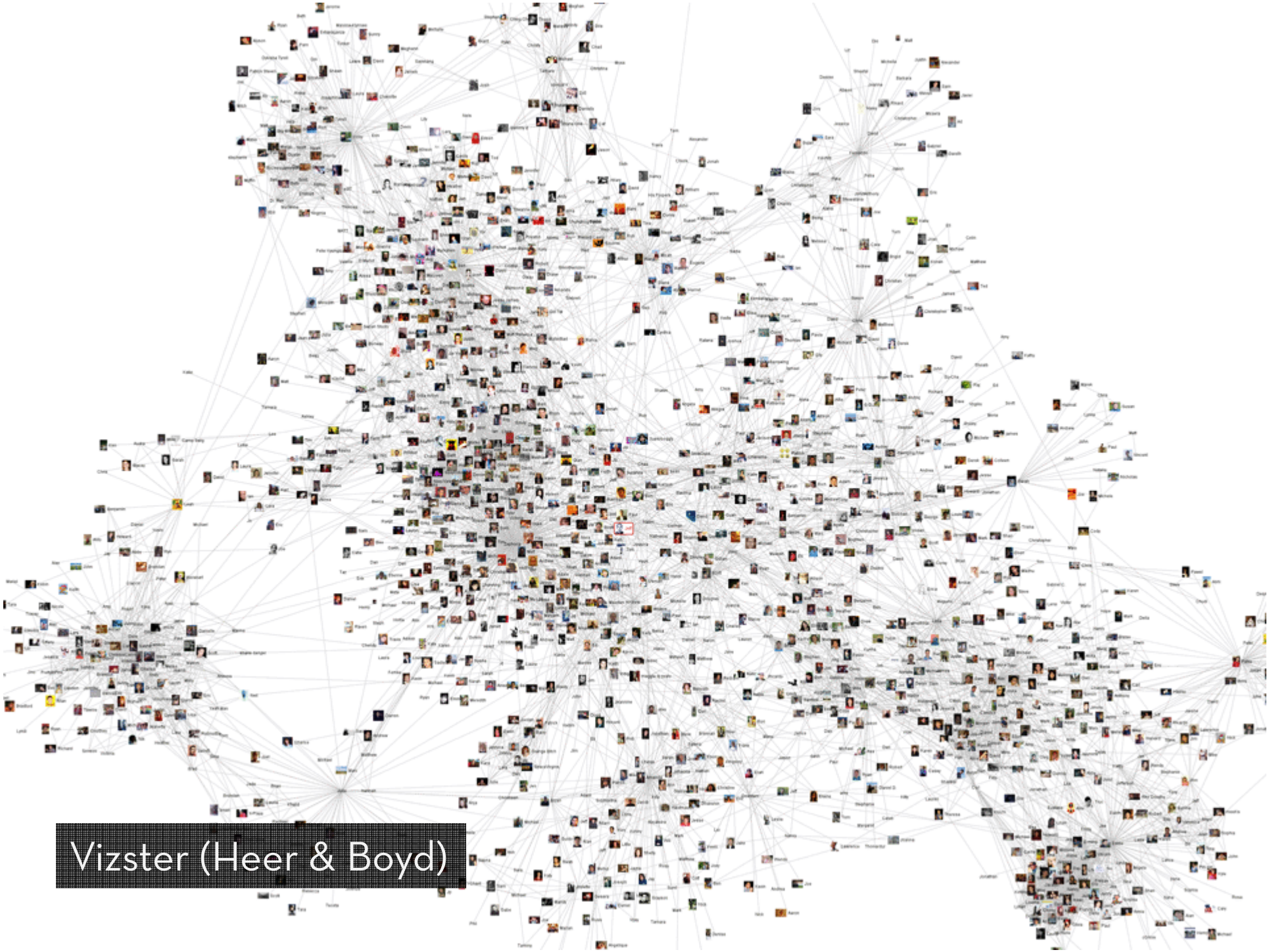
10x increase over 5 years

Gantz et al, 2008, 2010

The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that's going to be a hugely important skill in the next decades, … because now we really do have **essentially free and ubiquitous data**. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, *The McKinsey Quarterly*, Jan 2009


cabspotting.org


Vizster (Heer & Boyd)

1 Visualization

2 Networks

3 Text

1 Visualization

2 Networks

3 Text

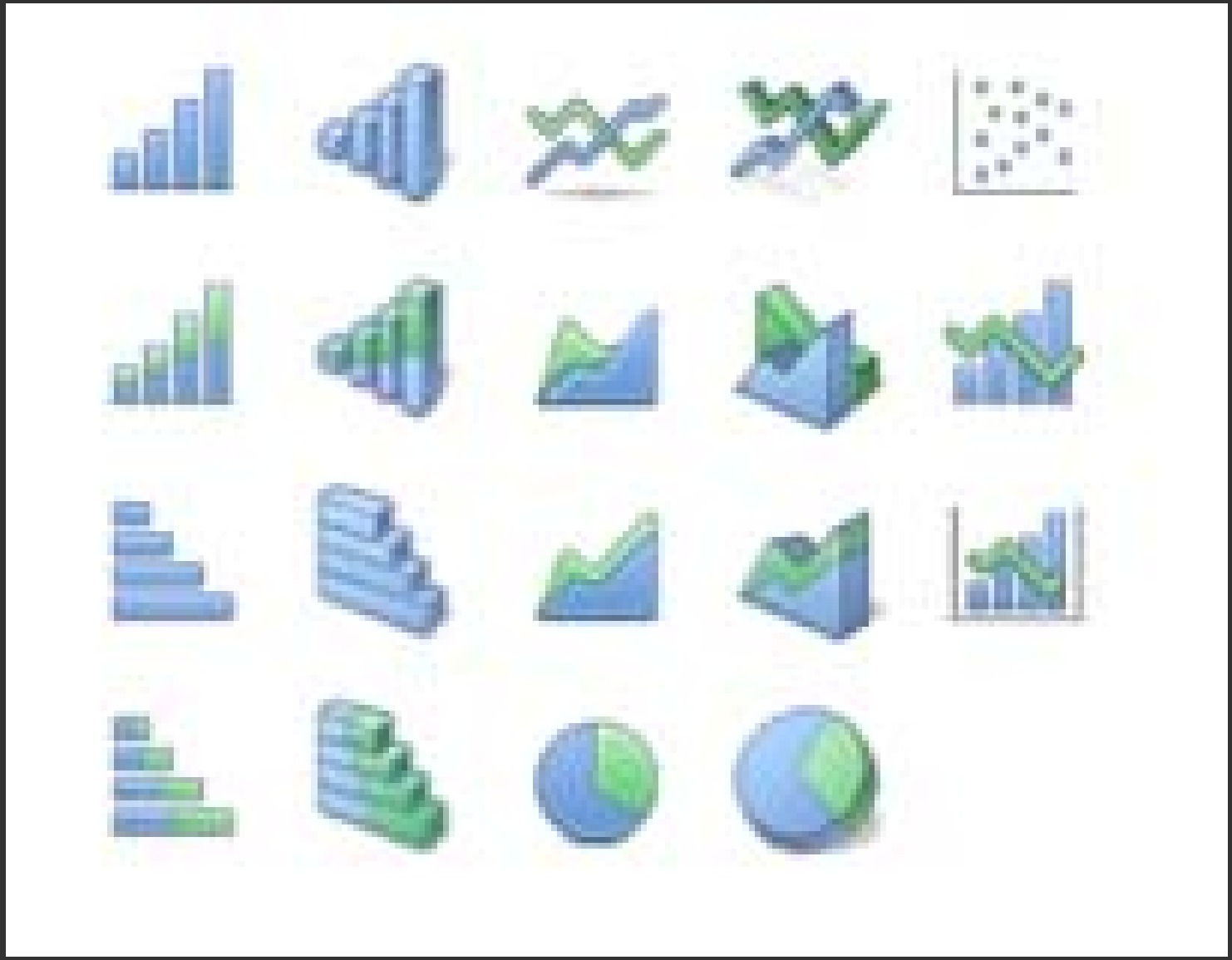


Chart Typology

Data Sets : State Quick Facts

Uploaded By: zinggoat
Data Source: US Census Bureau
Description:
Tags: people census

Created at: Friday May 18, 3:08 PM

view as text | edit data set

	People QuickFacts	Population 2005 estimate	Population percent change April 1 2000 to July 1 2005	Population 2000	Population percent change 1990 to 2000	Persons under 5 years old percent 2004	Persons under 18 years old percent 2004	Persons 65 years old and over percent 2004
1	Alabama	4557808	0.03	4447100	0.1	0.07	0.24	0.13
2	Alaska	663661	0.06	626932	0.14	0.08	0.29	0.06
3	Arizona	5939292	0.16	5130632	0.4	0.08	0.27	0.13
4	Arkansas	2779154	0.04	2673400	0.14	0.07	0.25	0.14
5	California	36132147	0.07	33871648	0.14	0.07	0.27	0.11
6	Colorado	4665177	0.08	4301261	0.31	0.07	0.26	0.1
7	Connecticut	3510297	0.03	3405565	0.04	0.06	0.24	0.14
8	Delaware	843524	0.08	783600	0.18	0.07	0.23	0.13
9	Florida	17789864	0.11	15982378	0.24	0.06	0.23	0.17
10	Georgia	9072576	0.11	8186453	0.26	0.08	0.26	0.1
11	Hawaii	1275194	0.05	1211537	0.09	0.07	0.24	0.14
12	Idaho	1429096	0.1	1293953	0.29	0.07	0.27	0.11
13	Illinois	12763371	0.03	12419293	0.09	0.07	0.26	0.12

Choosing a visualization type for State Quick Facts

Analyze a text

Tag Cloud

How are you using your words? This enhanced tag cloud will show you the words popularity in the given set of text.

Learn more

Wordle

Wordle is a toy for generating "word clouds" from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.

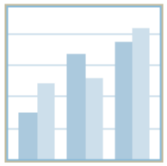
Learn more

Word Tree

See a branching view of how a word or phrase is used in a text. Navigate the text by zooming and clicking.

Learn more

Compare a set of values



Bar Chart

How do the items in your data set stack up? A bar chart is a simple and recognizable way to compare values. You can display several sets of bars for multivariate comparisons.

Learn more

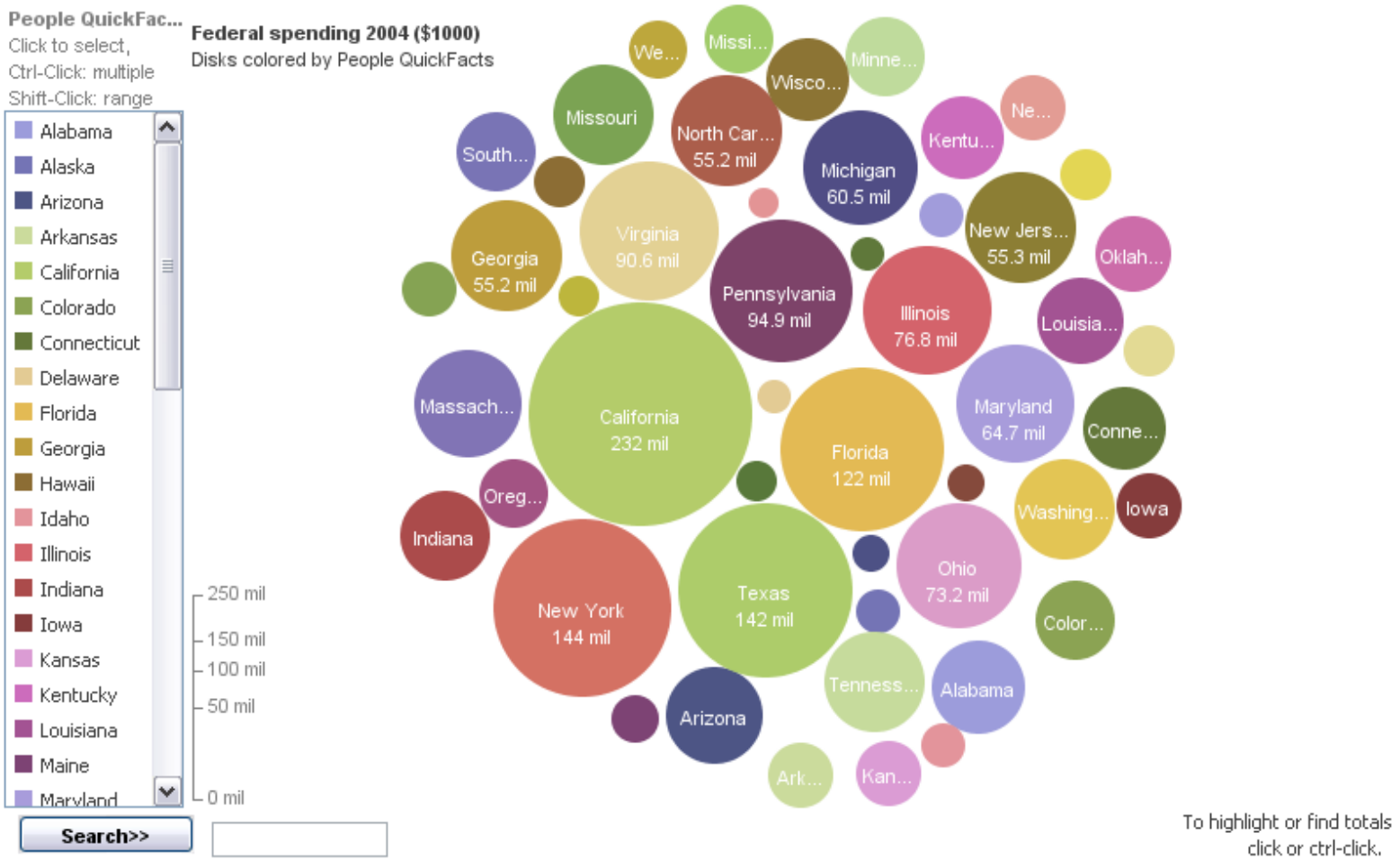
Block Histogram

This versatile chart lets you get a quick sense of how a single set of data is distributed. Each item in the data is an individually identifiable block.

Learn more

Visualizations : Federal Spending by State, 2004

Creator: Anonymous
Tags: census people



Search>>

To highlight or find totals click or ctrl-click.

Bubble Size: Federal spending 2004 ($1000) | Label: People QuickFacts | Color: People QuickFacts

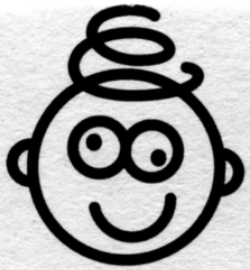
- Retail sales per capita 2002
- Minority-owned firms percent of total 1997
- Women-owned firms percent of total 1997
- Housing units authorized by building permits 2004
- Federal spending 2004 ($1000)
- Land area 2000 (square miles)
- Persons per square mile 2000
- FIPS Code

Data file ... Census Bureau

This data set has not yet been rated

full image

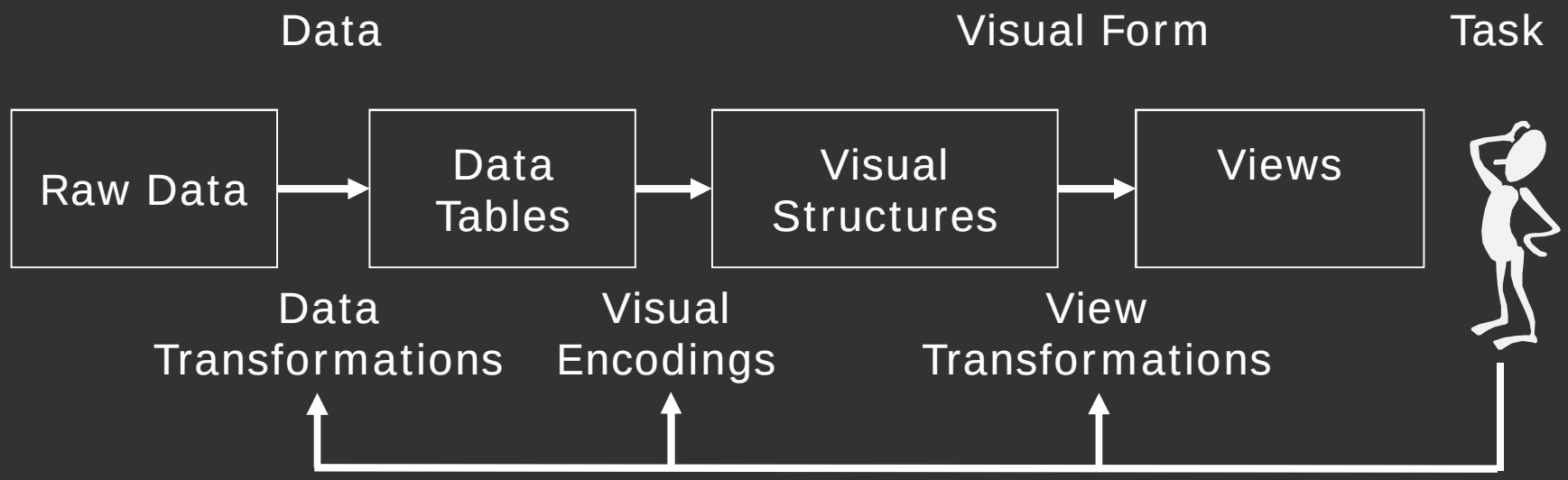
rate this



MAD LIBS®

MY MUSIC LESSON

Every Wednesday, when I get home from school, I have a piano lesson. My teacher is a very strict house (NOUN). Her name is Hillary Clinton (CELEBRITY (FEMALE)). Our piano is a Steinway Concert tree (NOUN) and it has 88 ~~[illegible]~~ cups (PLURAL NOUN). It also has a soft pedal and a/an Smily (ADJECTIVE) pedal. When I have a lesson, I sit down on the piano AIBERTO (NOUN) and play for 16 Minuts (PERIOD OF TIME). I do scales to exercise my cats (PLURAL NOUN), and then I usually play a minuet by Johann Sebastian Washington (CELEBRITY (LAST NAME)). Teacher says I am a natural Haunted House (NOUN) and have a good musical leg (PART OF THE BODY). Perhaps when I get better I will become a concert vet (PROFESSION) and give a recital at Carnegie hospital (TYPE OF BUILDING).


Data
Visual Form
Task
Raw Data
Data Tables
Visual Structures
Views
Data Transformations
Visual Encodings
View Transformations

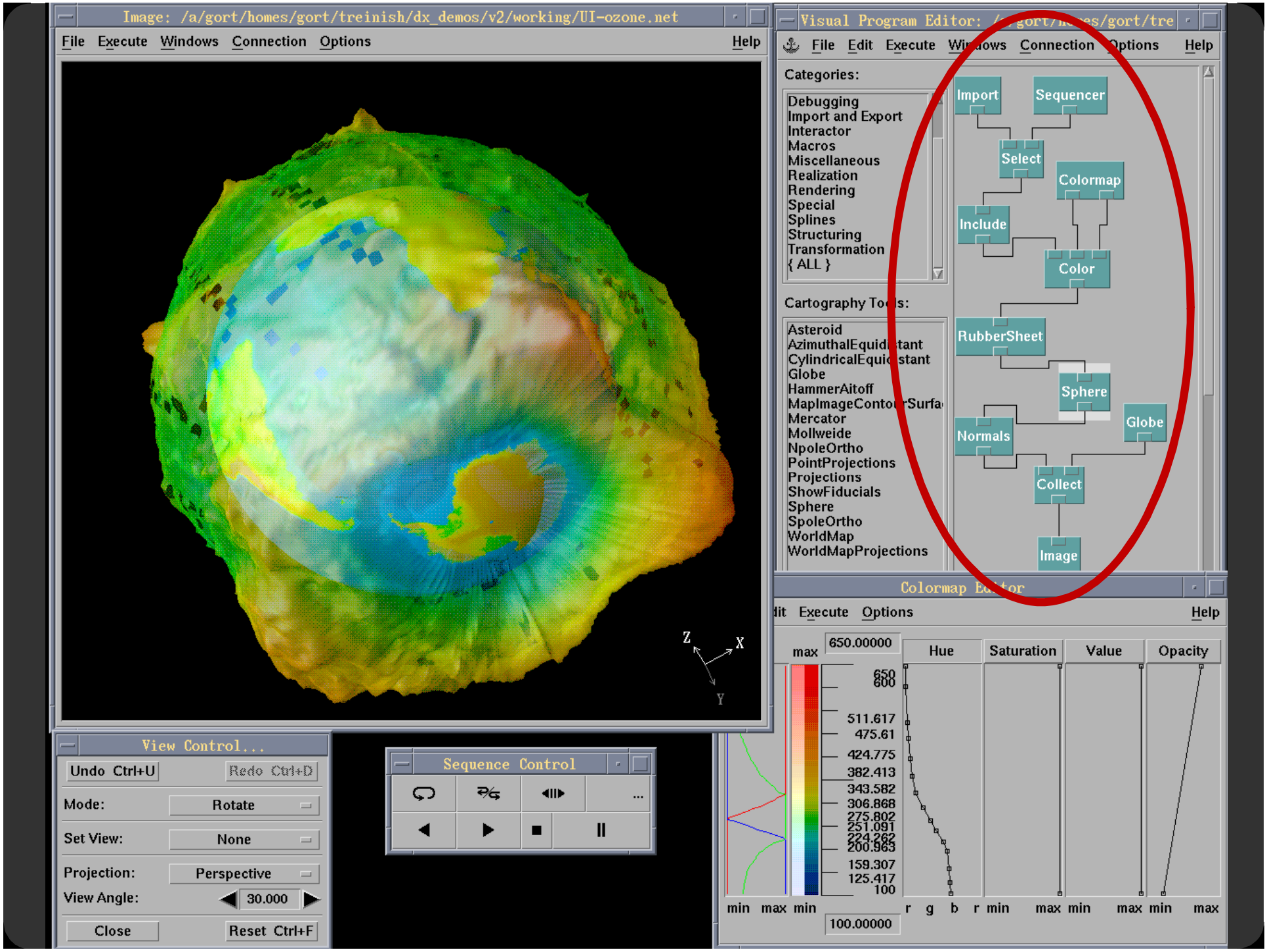
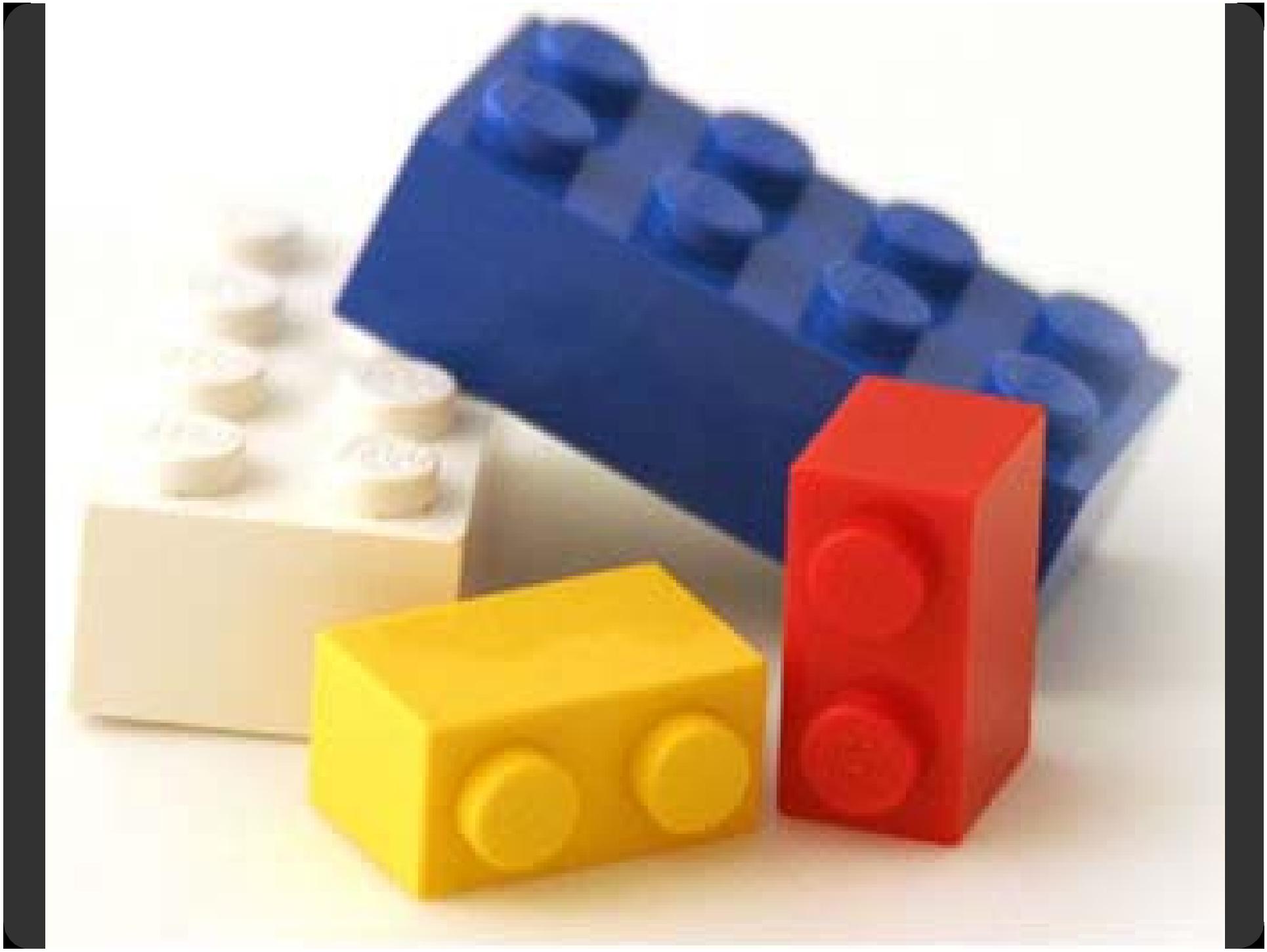


Image: /a/gort/homes/gort/treinish/dx_demos/v2/working/UI-ozone.net
File Execute Windows Connection Options Help
Z
X
Y
Visual Program Editor:
File Edit Execute Windows Connection Options Help
Categories:
Debugging
Import and Export
Interactor
Macros
Miscellaneous
Realization
Rendering
Special
Splines
Structuring
Transformation
{ ALL }
Cartography Tools:
Asteroid
AzimuthalEquidistant
CylindricalEquidistant
Globe
HammerAitoff
MapImageContourSurface
Mercator
Mollweide
NpoleOrtho
PointProjections
Projections
ShowFiducials
Sphere
SpoleOrtho
WorldMap
WorldMapProjections
Import
Sequencer
Select
Colormap
Include
Color
RubberSheet
Sphere
Globe
Normals
Collect
Image
Colormap Editor
Execute Options Help
max
650.00000
650
600
511.617
475.61
424.775
382.413
343.582
306.868
275.802
251.091
224.262
200.363
159.307
125.417
100
100.00000
Hue
Saturation
Value
Opacity
min max min r g b r min max min max min max
View Control...
Undo Ctrl+U
Redo Ctrl+D
Mode:
Rotate
Set View:
None
Projection:
Perspective
View Angle:
30.000
Close
Reset Ctrl+F
Sequence Control

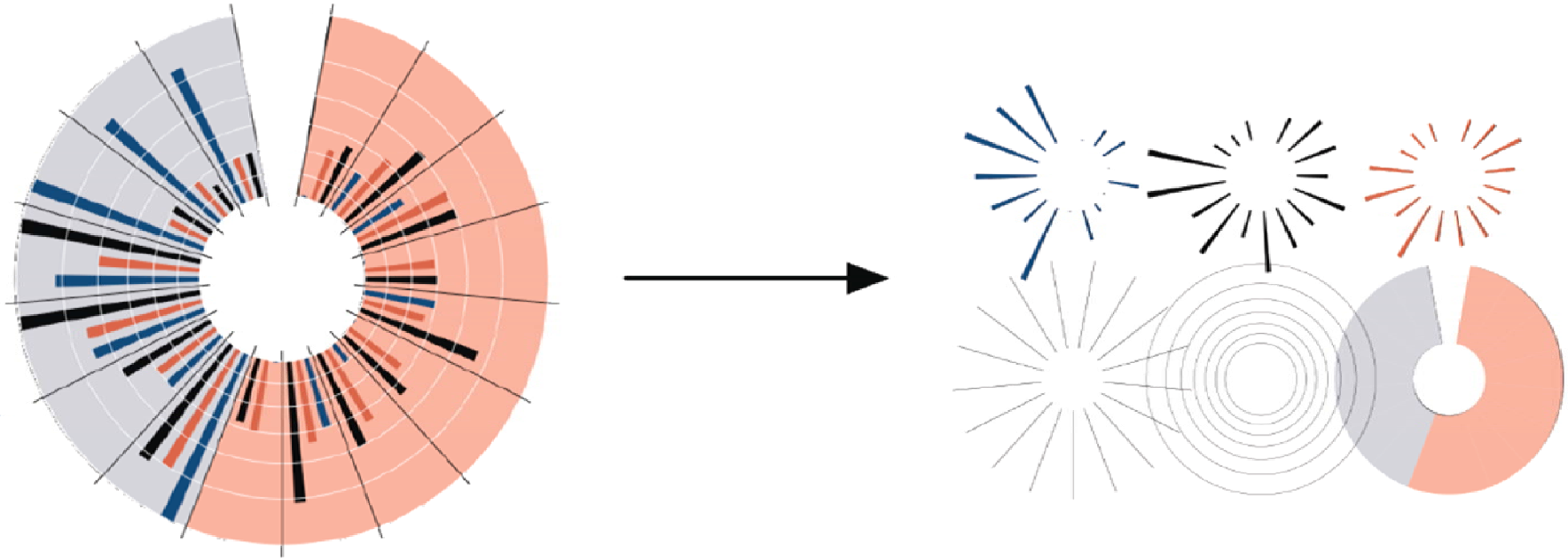


Today's first task is not to invent wholly new [*graphical*] techniques, though these are needed. Rather we need most vitally to recognize and reorganize the essential of old techniques, to make easy their assembly in new ways, and to modify their external appearances to fit the new opportunities.

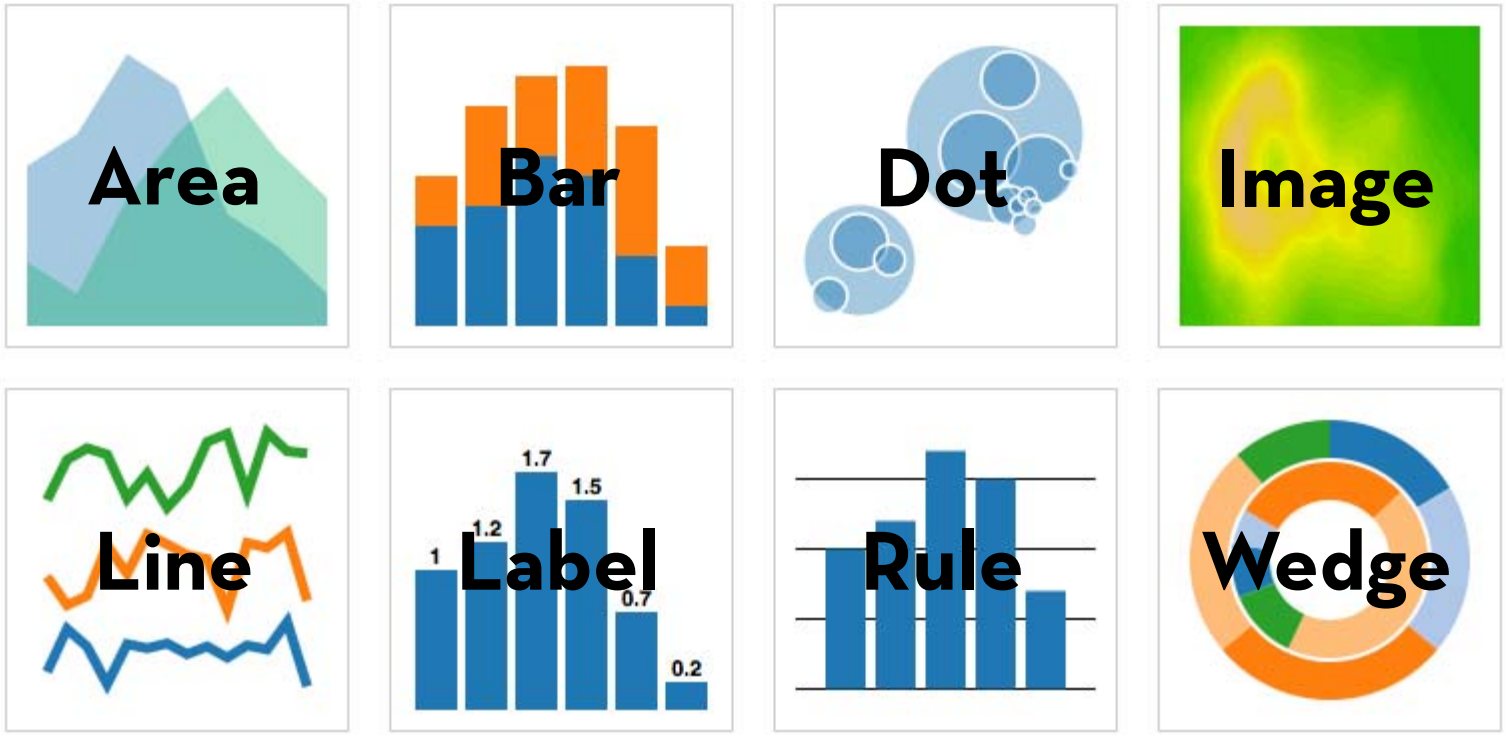
J. W. Tukey, *The Future of Data Analysis*, 1962.

Protovis: A Declarative Language for Visualization

http://protovis.org/

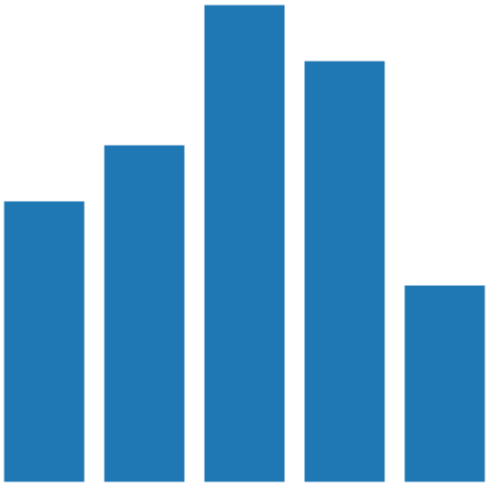


A graphic is a composition of data-representative marks.


Area
Bar
Dot
Image
Line
1
1.2
1.7
1.5
0.7
0.2
Label
Rule
Wedge

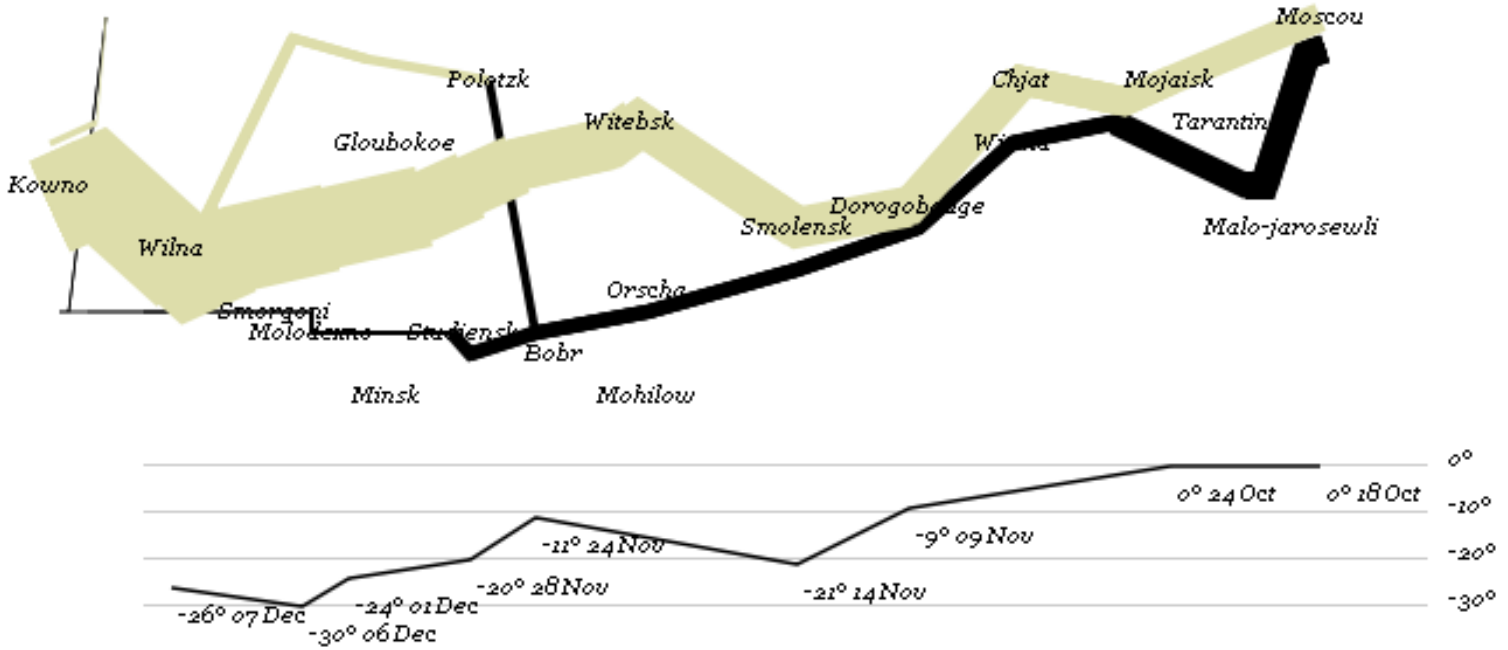
Protovis

Create customized visualizations using a declarative specification language.



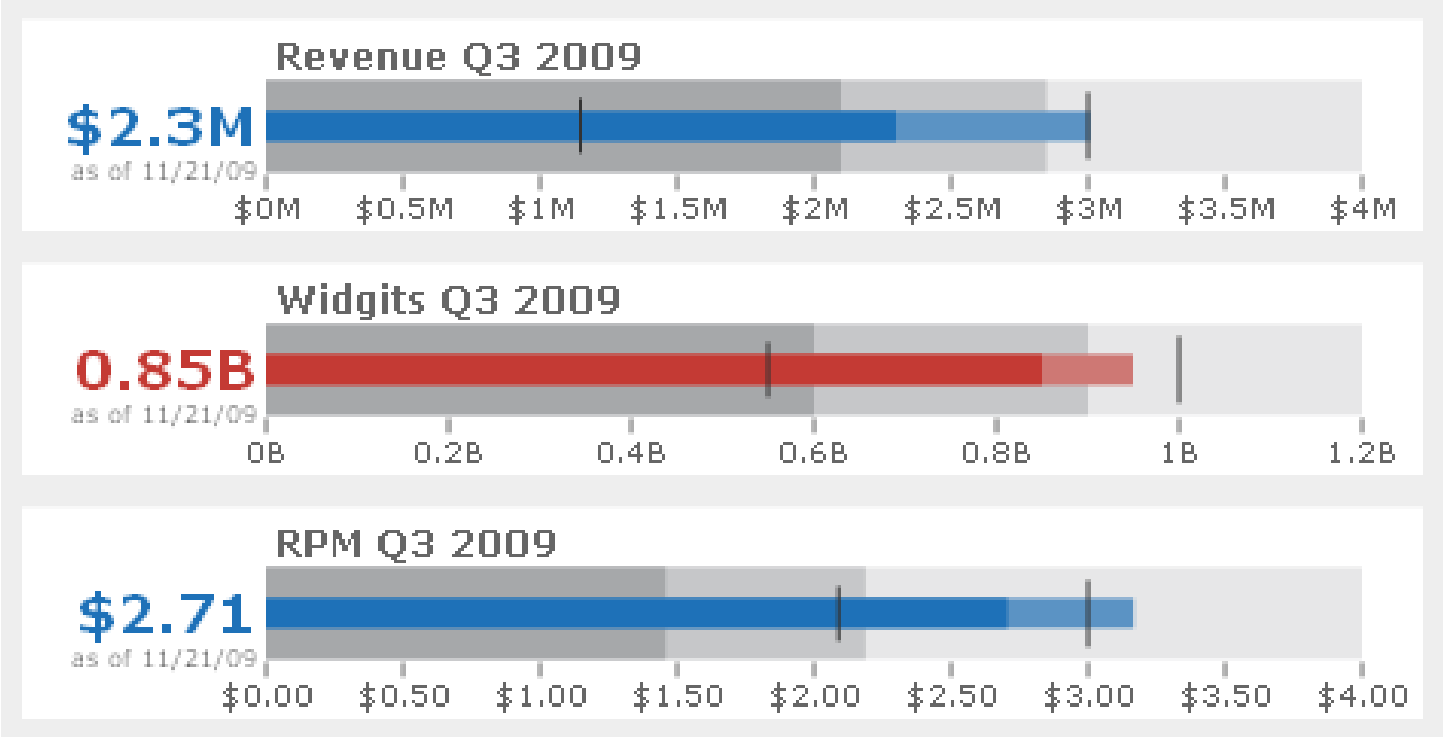
```
var vis = new pv.Panel();
vis.add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, .7])
  .bottom(10).width(20)
  .height(function(d) d * 70)
  .left(function()  this.index * 25 + 20);
vis.render();
```

Protovis (protovis.org) – Declarative Visualization Design

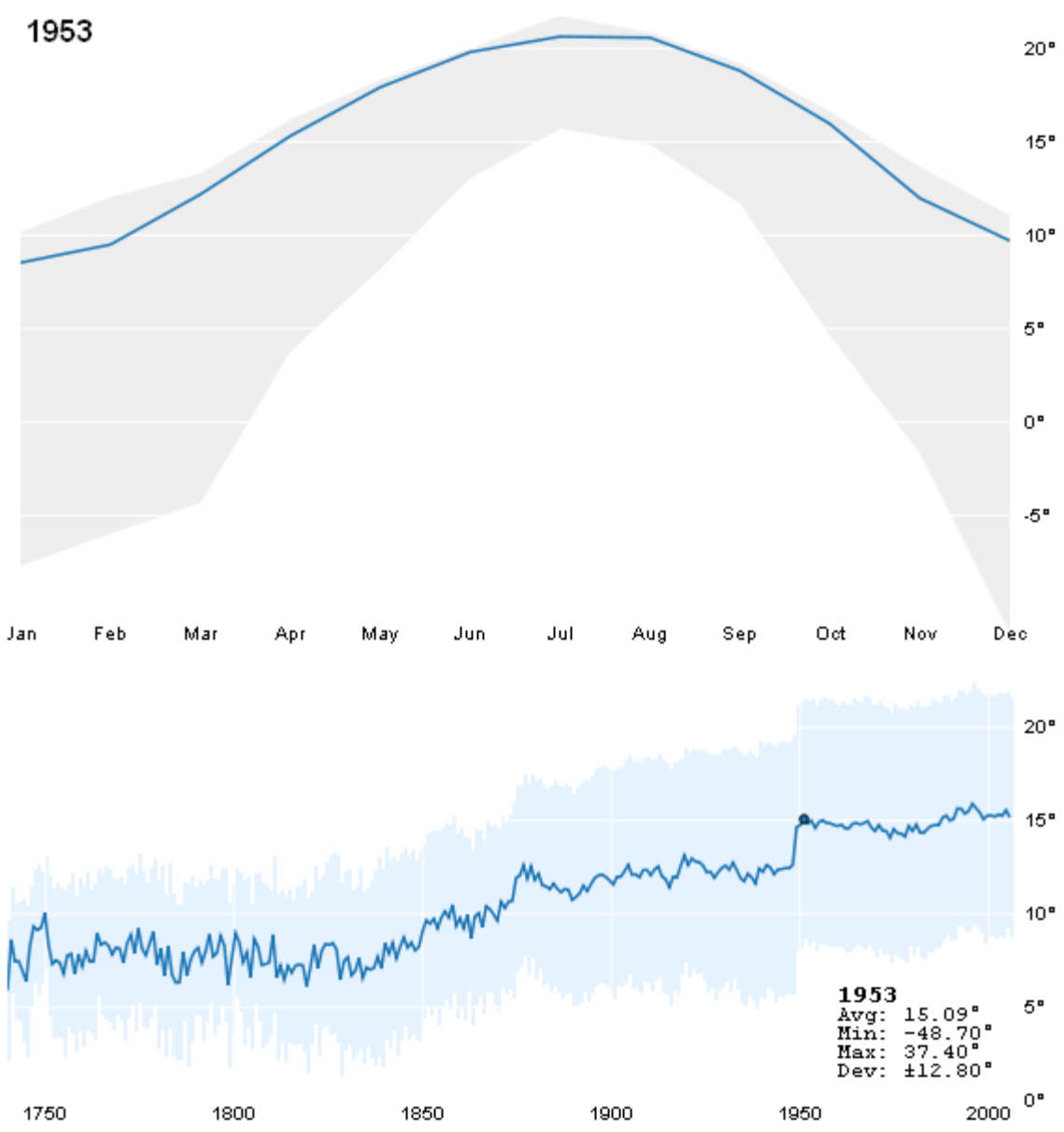



```
var army = pv.nest(napoleon.army, "dir", "group“);
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(function() army[this.idx])
  .left(lon).top(lat).size(function(d) d.size/8000)
  .strokeStyle(function() color[army[paneIndex][0].dir]);

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text(function(d) d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");
```

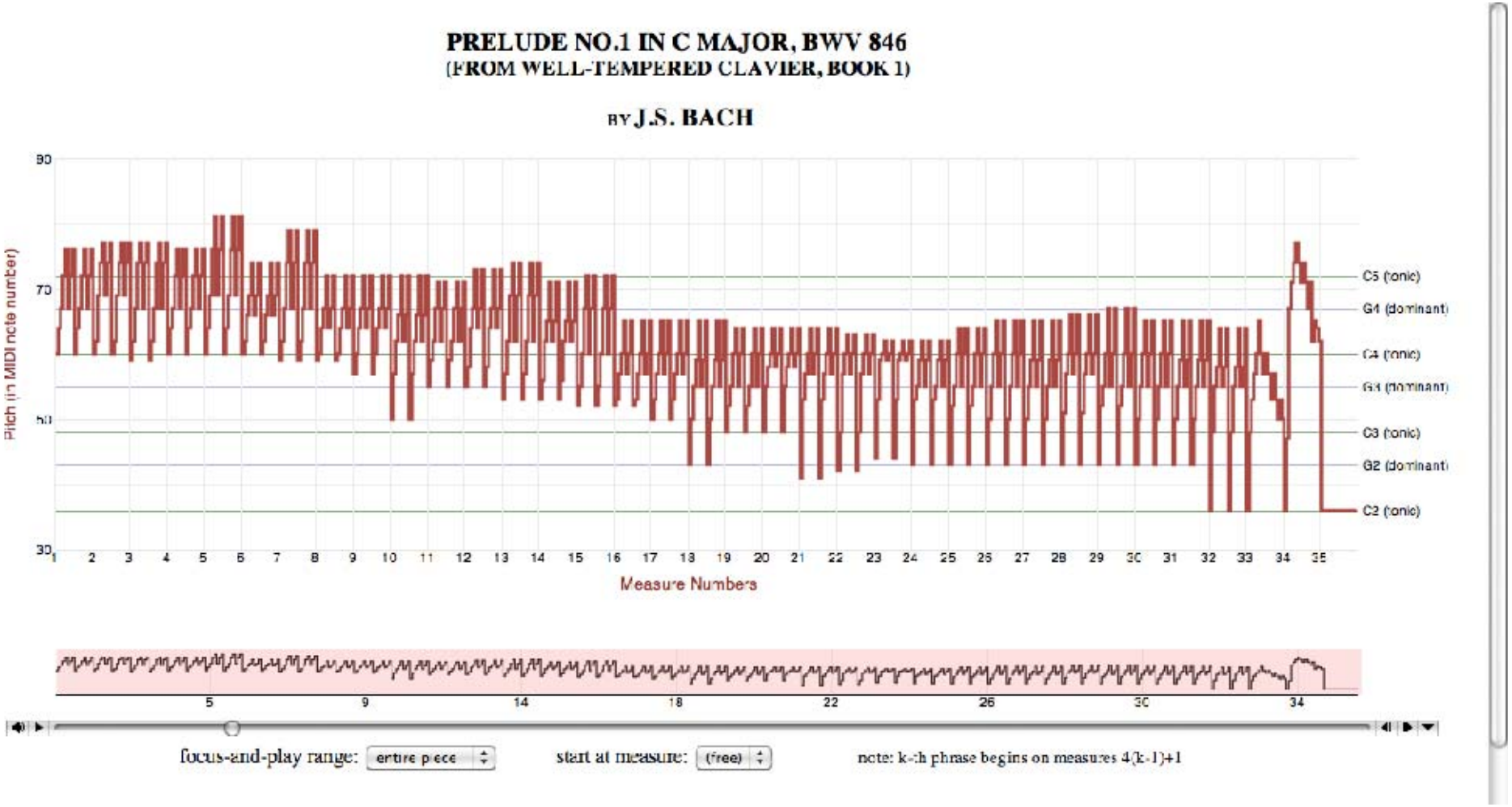
```
vis.add(pv.Rule).data([0,-10,-20,-30])
  .top(function(d) 300 - 2*d - 0.5).left(200).right(150)
  .lineWidth(1).strokeStyle("#ccc")
  .anchor("right").add(pv.Label)
    .font("italic 10px Georgia")
    .text(function(d) d+"°").textBaseline("center");

vis.add(pv.Line).data(napoleon.temp)
  .left(lon).top(tmp) .strokeStyle("#0")
 .add(pv.Label)
  .top(function(d) 5 + tmp(d))
  .text(function(d) d.temp+"° "+d.date.substr(0,6))
  .textBaseline("top").font("italic 10px Georgia");
```



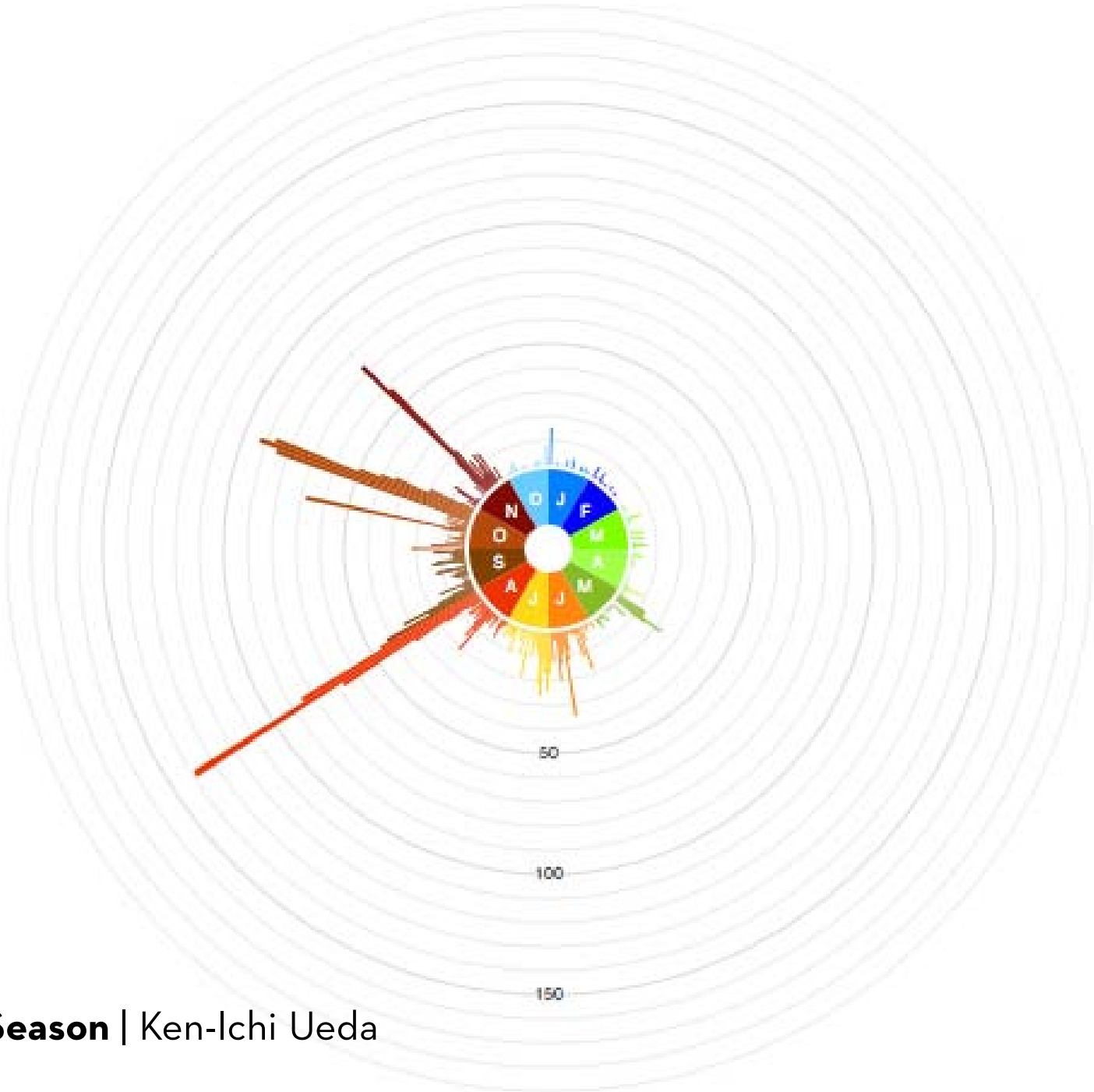
Bullet Charts | Clint Ivy



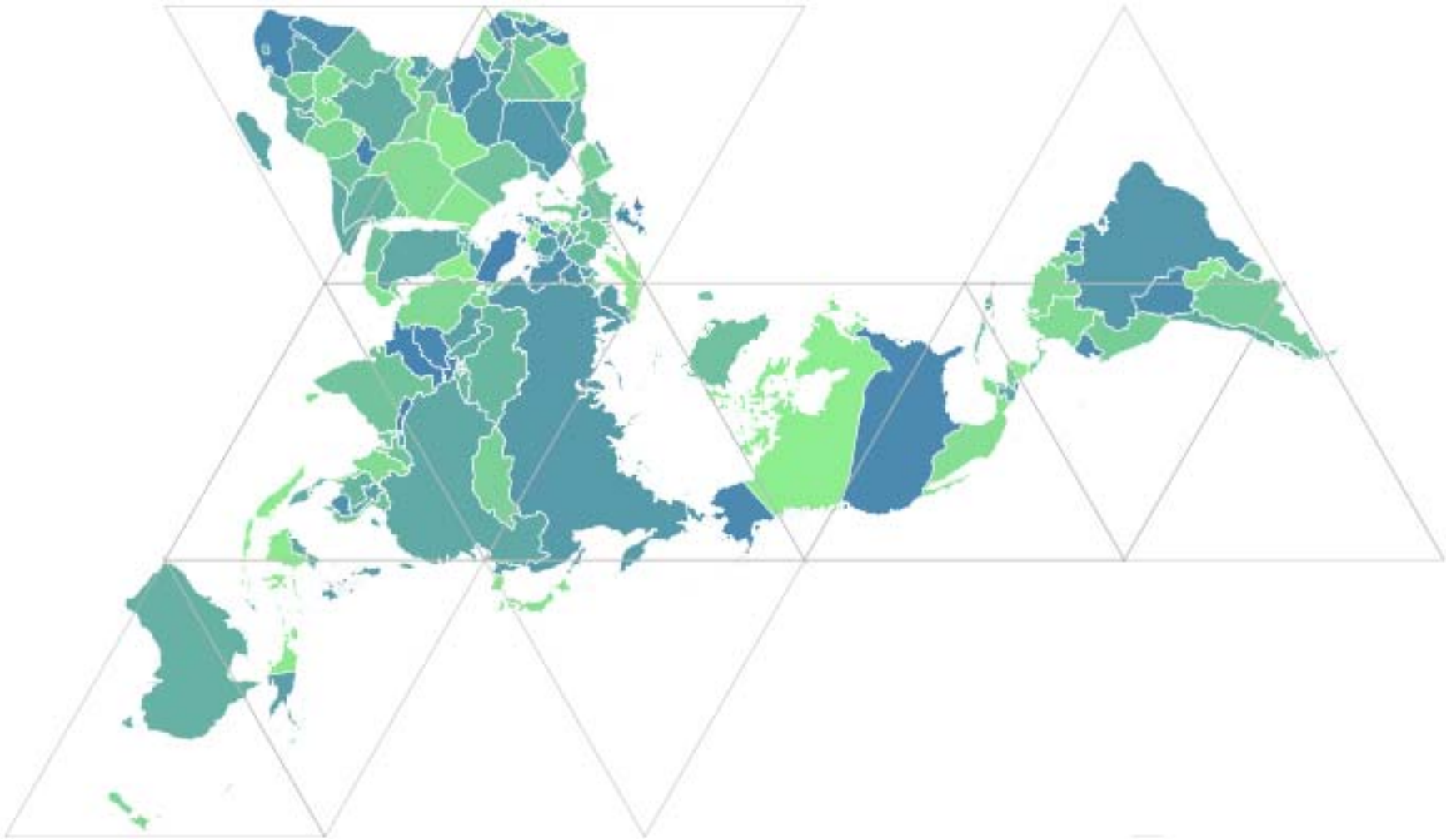
Climate Graph | Robert Kosara



Bach's Prelude #1 in C Major | Jieun Oh



FlickrSeason | Ken-Ichi Ueda



Dymaxion Maps | Vadim Ogievetsky

Current Status & Results

Protovis has led to faster designs, less code

- Job Voyager: 5x less code, 10x less dev time
- Over 20,000 downloads and widely in use

Multiple implementations: JavaScript & Java

Behind-the-scenes optimization & parallelization

- 20x scalability over prior systems (in Java)

http://protovis.org

1 Visualization

2 Networks

3 Text

facebook

Profile edit | Friends | Networks | Inbox | home account privacy logout

Search

Applications edit

- Photos
- Groups
- Events
- Marketplace
- The New York Times News Quiz

more

Bill Gates
is glad he finally joined facebook and hopes you will too!!! :)
Updated 6 minutes ago

Send Bill a Gift

Send Bill a Message

Poke Bill!

▼ Mini-Feed

Displaying 15 stories See All

Update: **Bill** has posted a note:
Friends, I have finally caved and joined facebook, America's fastest-growing social-networking web site! At first I didn't join because you needed a college alumni address, and I never quite got one... Then when the place started opening up to high schools and corporations, with everyone and his grandmother joining, I wanted in. But by then I was mad I didn't have any shares in this $15 billion baby. ... So just now I decided to plunk down $240 million to buy 1.6% of the company from cool kid CEO Mark Zuckerberg. Sure I saw the potential for ad revenue right away -- but this is wild. I've never had my own Facebook page before!

Don't have a lot of friends yet but I've been running into people... Seeing their status updates... Wow, it's a great place to check up on my employees and my kids! They keep saying what their weekend plans are in their status bars. And of course I love how you can add all these little software applications to the page. Or write your very own!

I'm still getting the hang of the whole thing... It seems there are a variety of forums where I can vent to others and display details of my life? O.K.: It wasn't easy being so much smarter than everyone else, pretending to be a grown-up over the telephone so I could get grown-up jobs programming these new things called computers when I was still a child.

At college I led the anti-social group. Never led a social group, or had a social network... Ha, ha, now I've done even better than that: I've bought a piece of the national friend system! Take that cliquesters. Anyone who ever ignored me in the dining hall... Got friends? I own 1.6 % of your friends.

But don't worry about them. Send me a message! Write on my wall!

Bill and Mark Zuckerberg are now friends

Bill and Warren Buffett have joined the group Save the World Now through Creative Capitalism (3 Members)

▼ Friends

See All

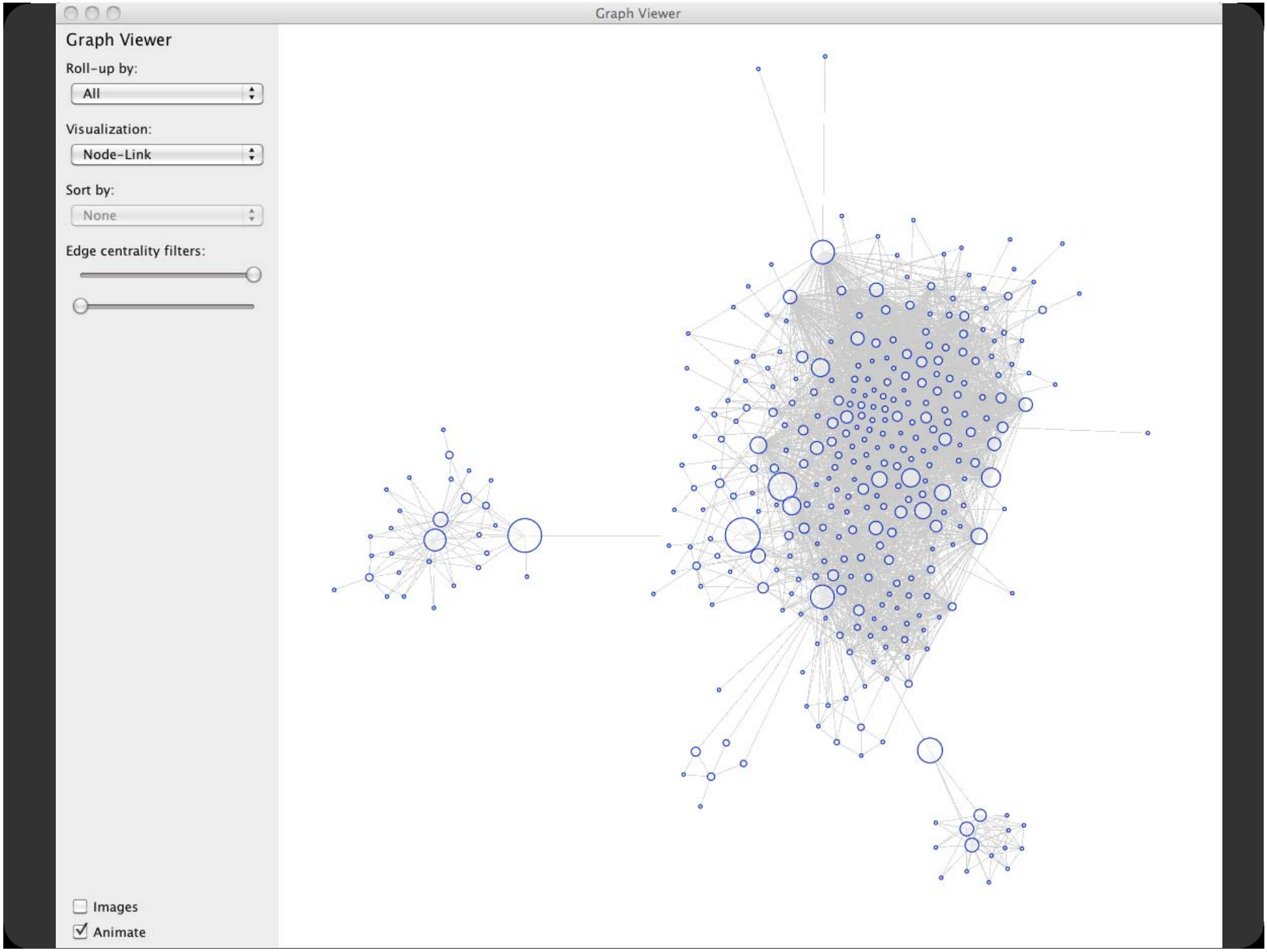

Melinda

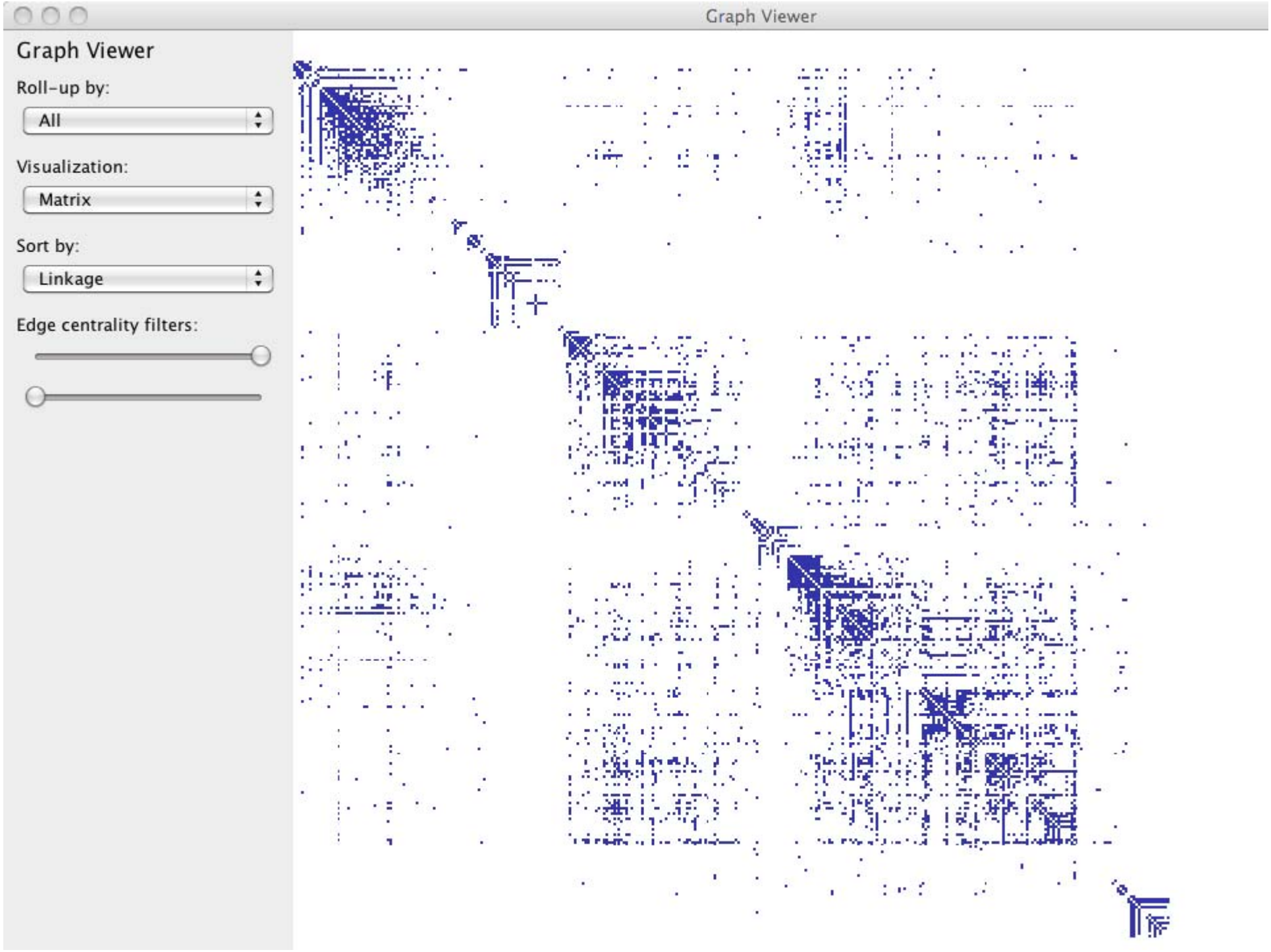

Steve Ballmer

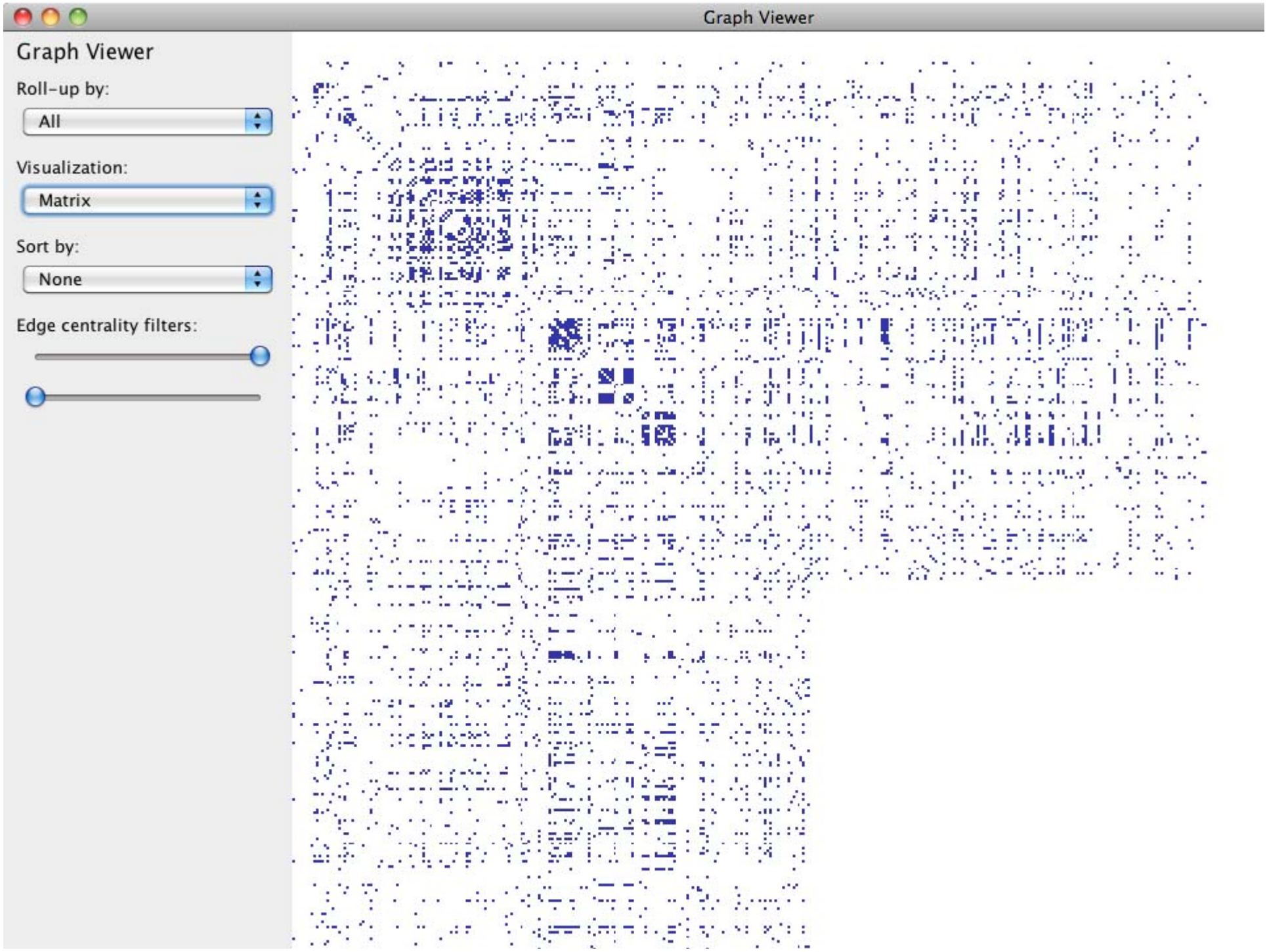

Mark


William Randolph


Bono


Graph Viewer
Graph Viewer
Roll-up by:
All
Visualization:
Node-Link
Sort by:
None
Edge centrality filters:
Images
Animate


Graph Viewer
Graph Viewer
Roll-up by:
All
Visualization:
Matrix
Sort by:
Linkage
Edge centrality filters:


Graph Viewer
Graph Viewer
Roll-up by:
All
Visualization:
Matrix
Sort by:
None
Edge centrality filters:

MedHelp | Forums | Health Tools | Health Information | Drugs | People | Experts | Pets | My MedHelp

Search conditions, symptoms, and treatments

Login | JOIN Free Membership

Finding cures together.

Sign Up Now!

Forums | Tools | Experts | Health Centers

The World's Largest Health Community

...sharing their experiences

Popular Communities

Physical Health	Women's Health	Healthy Living
Dermatology	Breast Cancer	Addiction
Gastroenterology	Fertility	Allergy
Heart Disease	Maternal & Child	Children's Allergies
Lyme Disease	Ovarian Cancer	HIV Prevention
Neurology	Pregnancy	Men's Health

See all communities >>

Member Login

Nickname

Password

Remember login info

Login

Forgot your password or nickname?

Monthly visitors: 10 million

Medical posts: 4,301,267

MedHelp Partners See all

Cleveland Clinic

USF HEALTH

Partner with MedHelp

MedHelp works with doctors, hospitals, and non-profits to find cures together.

Become a partner

Discussions

A A A

Sort by: Last Post

	Replies	
WELCOME to our Asthma Community! by ChitChatNine, Jun 12, 2009 08:52PM Welcome to our Asthma Community! If you or a loved one has experien...	3	Aug 15 helpmyteenag... Jun 18 littlejenny2... Nov 13 loomislady40...
Regina's cough by ladyrgr, 42 minutes ago I've been couging for two years. The cough is very productive. I've see...		
Interstitial Markings by a1cklm1987, 4 hours ago I had a chest xray done to check for TB. TB came back negative but the ...		
elliots asthma powder 1940's by colind, Oct 27, 2007 02:01AM Has anyone info on "Elliots"s Asthma Powder" 1940"s(Elliot's and Austra...	6	Oct 19 flick09 Feb 15 ridleygirl 09/09 Swiftyone 10/08 bladerunner01... 11/07 caregiver222 10/07 jdesouza
Cat or Road? by nautilis, Oct 16, 2010 09:40PM Hi I've recently moved to a new home, and almost straight away I agreed...	4	Oct 19 wisby Oct 18 nautilis Oct 17 bsmsl Oct 16 nautilis
help by squeaker420, Oct 17, 2010 06:29PM i just hit my throat on the corner of my couch and now it feels like ha...	2	Oct 19 wisby Oct 17 bsmsl

Related Communities

Respiratory Disorders

MedHelp Health Answers

Type your medical question here

Submit

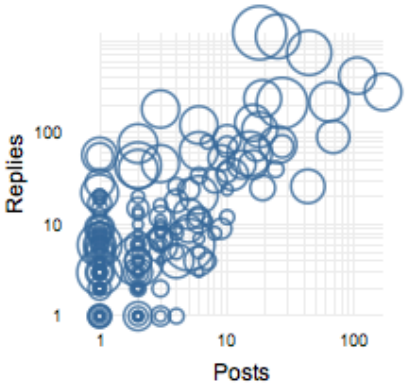
Related Tags

mucus **Chest** **Pain** **Anxiety** eyes loestrin peak flow **Lung** itch smoking **Pain** oxygen Singulair worried help **cough** advair **Asthm**

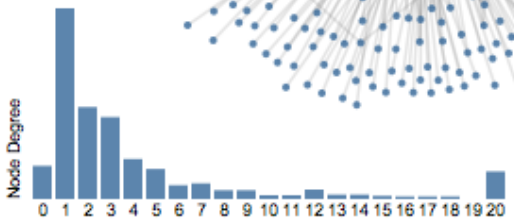
MedHelp > Lyme

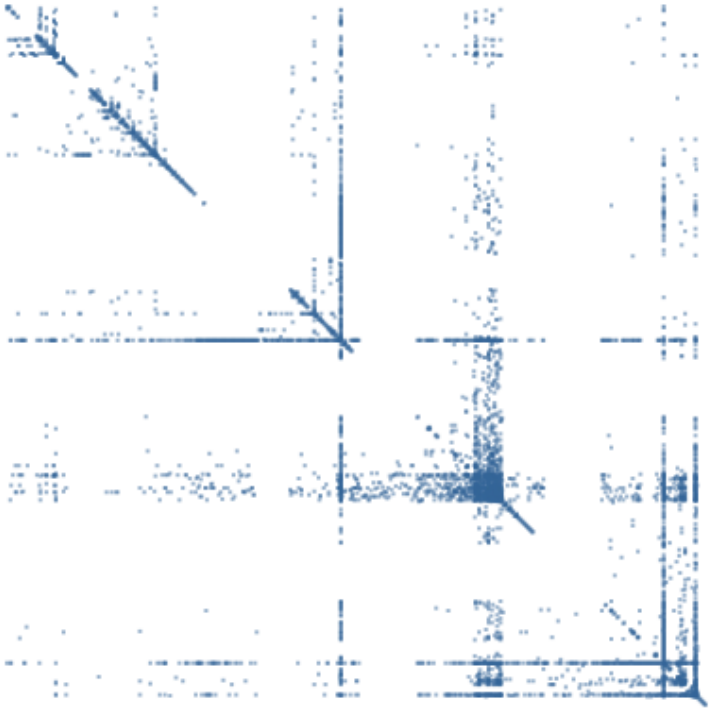
Posting Activity by Month

Community Membership



Social Network





Design Considerations

Multiple representations aid reasoning
- Node-link + Matrix views, linked plots

Networks are **more than just structure**
- Attributes of nodes and edges
- Evolution over time

Transformation & visualization go hand-in-hand
- Filtering, Aggregation, Statistics
- Defining Network Ties, Subdivision

Current Work

ORION: A network processing pipeline

- Extract networks from input data
- Define, merge, and weight links
- Graph manipulation: aggregation, subdivision
- Graph statistics: BC, clustering, community id

Next Steps: design a GUI for interactive specification of transforms & visualizations

1 Visualization

2 Networks

3 Text

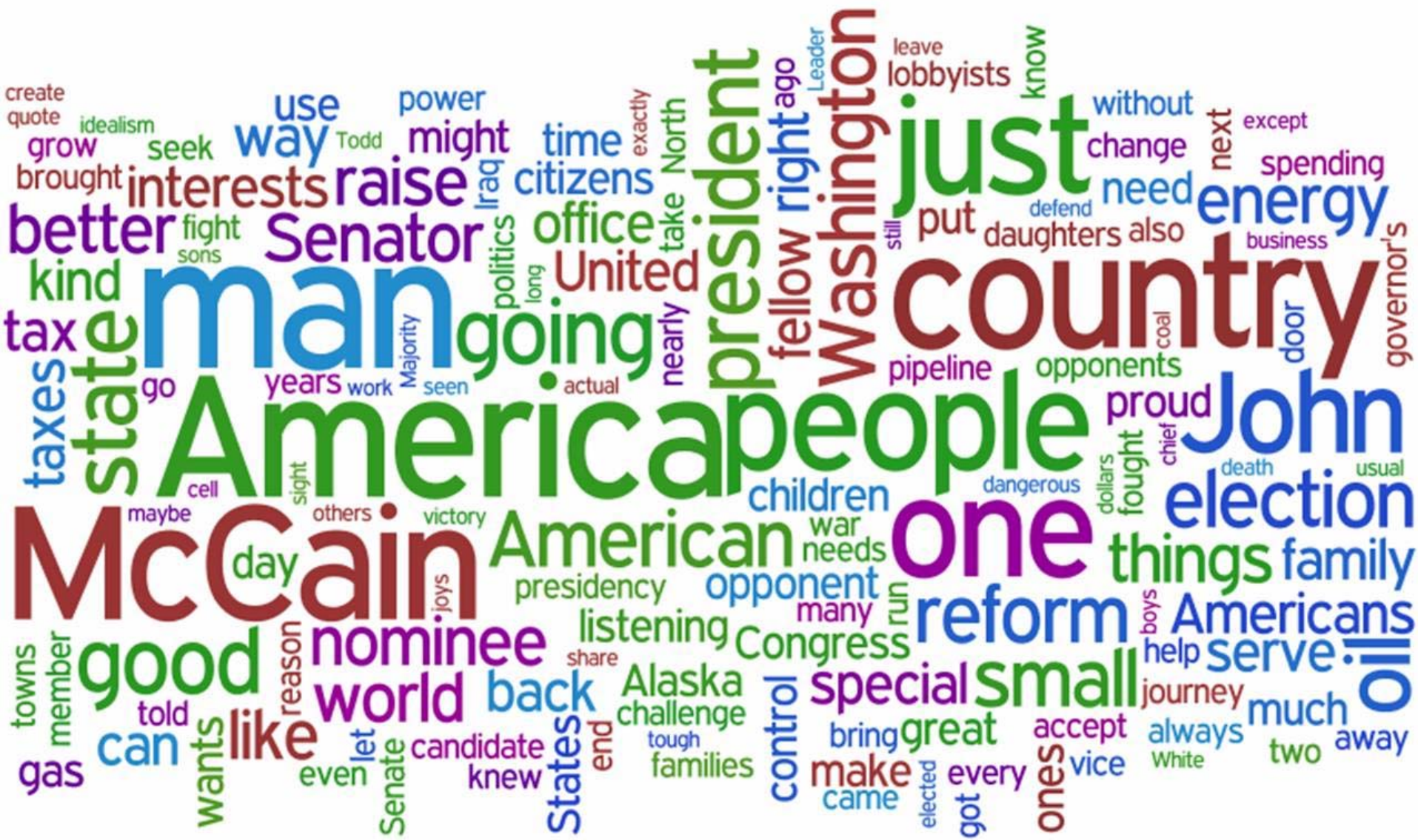
Why Visualize Text?

Understanding – get the "gist" of a document

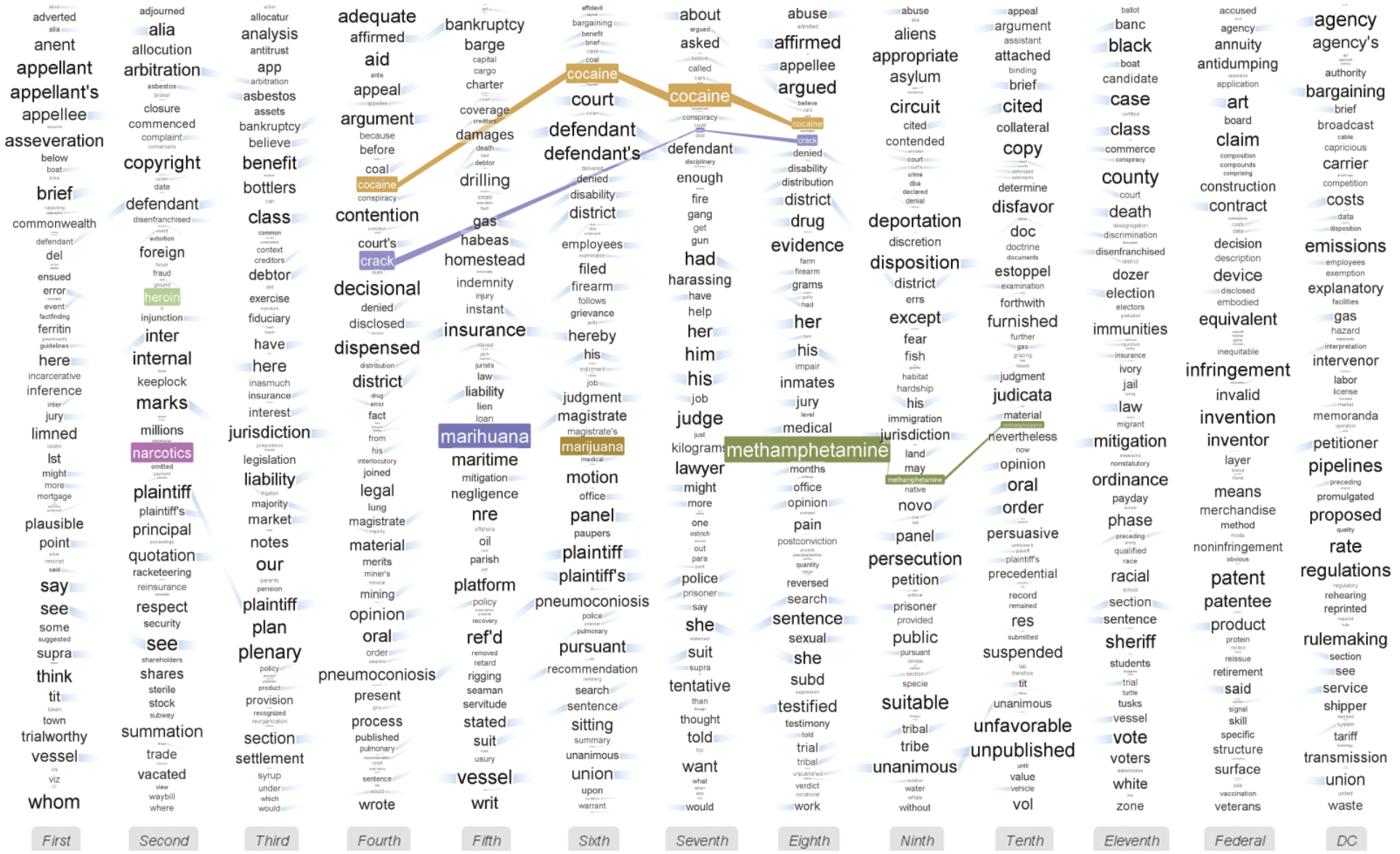
Grouping – cluster for overview or classification

Compare – compare document collections, or inspect evolution of collection over time

Correlate – compare patterns in text to those in other data, e.g., correlate with social network



Wordle Tag Cloud of a Political Speech



Parallel Tag Clouds of U.S. Court Decisions (Collins et al, 2009)

How might we best summarize text documents?

Approach: first observe how *people* read and summarize text

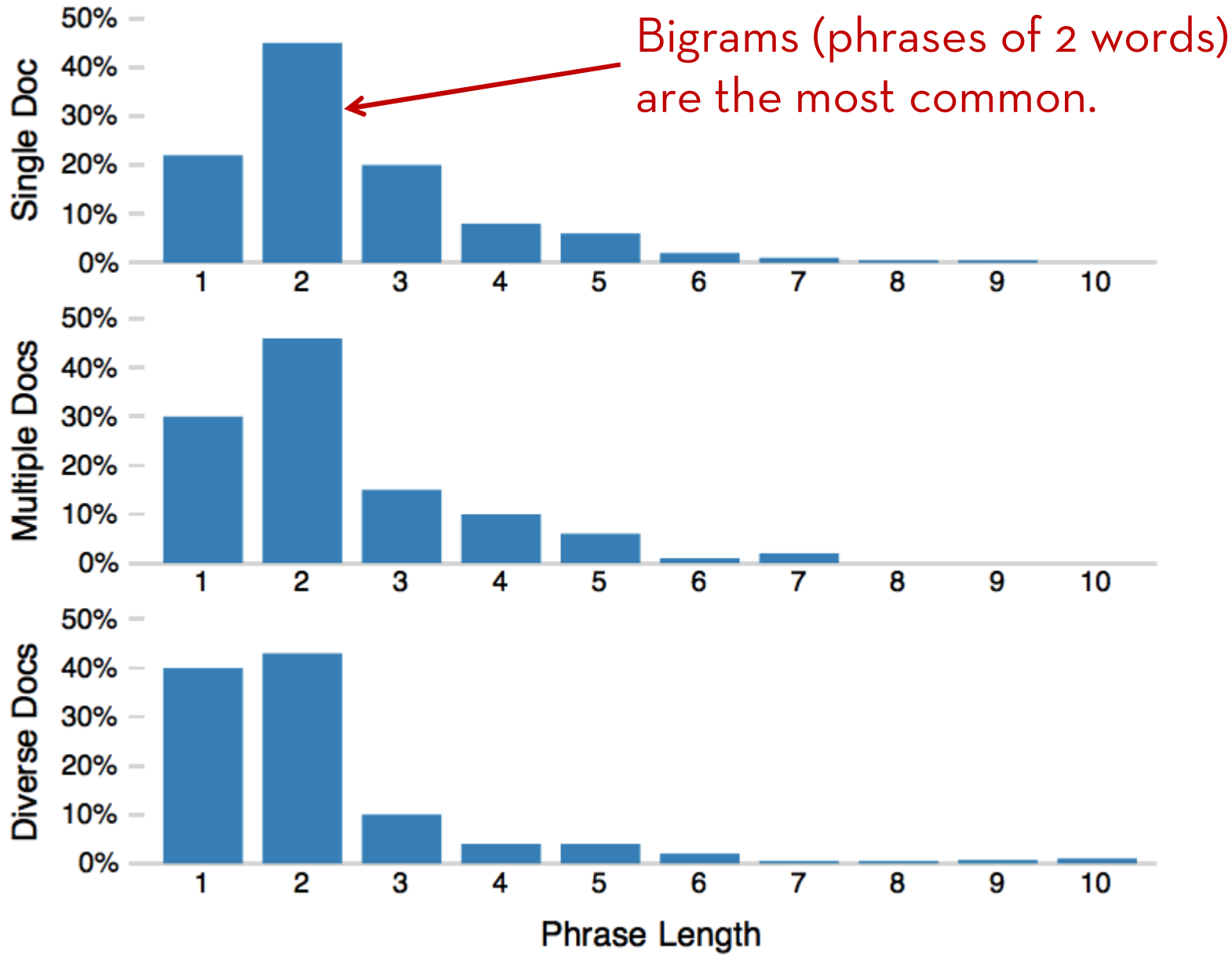
Method – Part 1

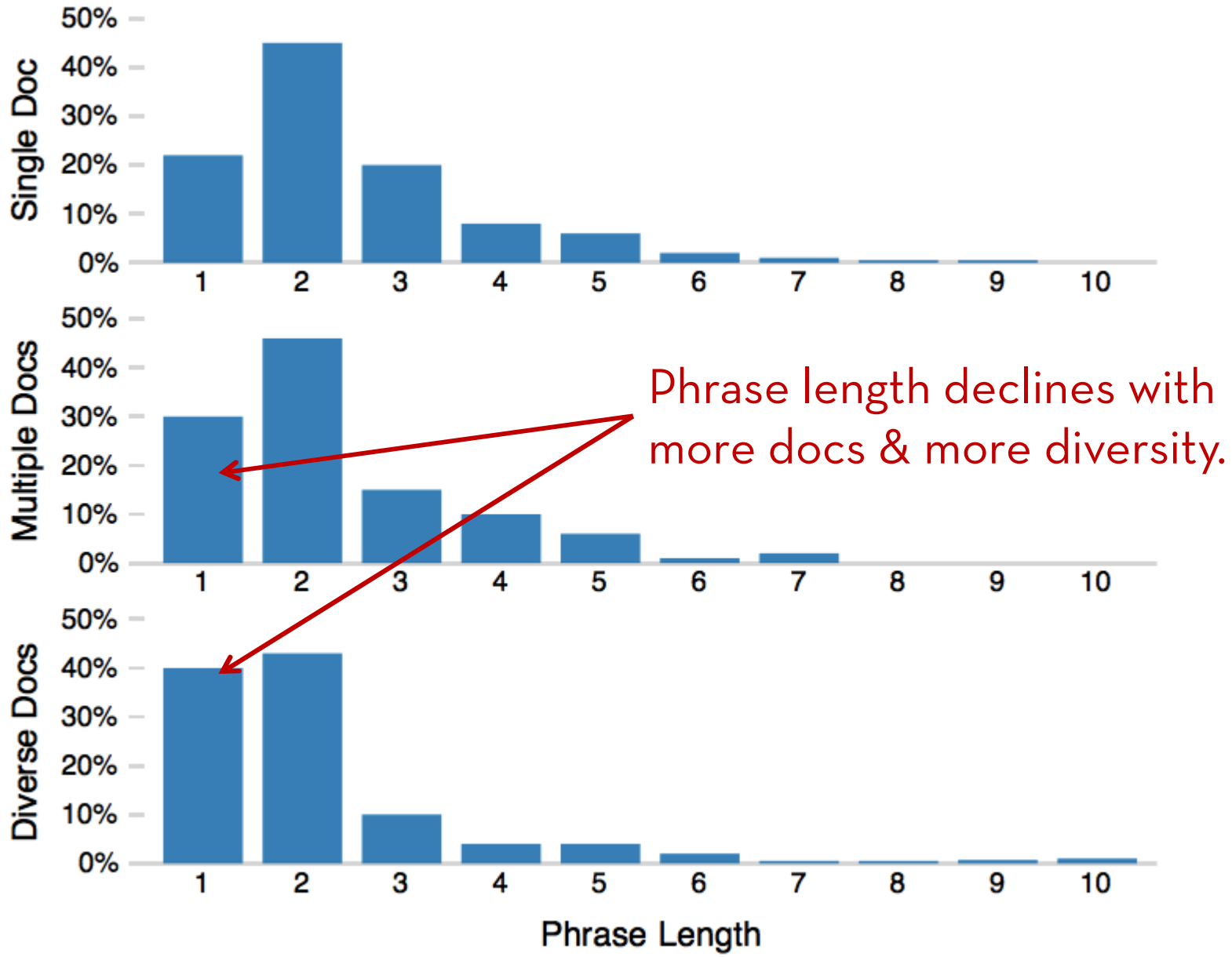
Asked 69 subjects (all Ph.D. students) to read and describe dissertation abstracts.

Students were given 3 documents in sequence, they then described the collection as a whole.

Students were matched to both *familiar* and *unfamiliar* topics; *topical diversity* within a collection was varied systematically.

We then analyzed and modeled the results...


Bigrams (phrases of 2 words)
are the most common.
Single Doc
Multiple Docs
Diverse Docs
50%
40%
30%
20%
10%
0%
1
2
3
4
5
6
7
8
9
10
Phrase Length


Single Doc
Multiple Docs
Diverse Docs
50%
40%
30%
20%
10%
0%
1
2
3
4
5
6
7
8
9
10
Phrase Length
Phrase length declines with more docs & more diversity.

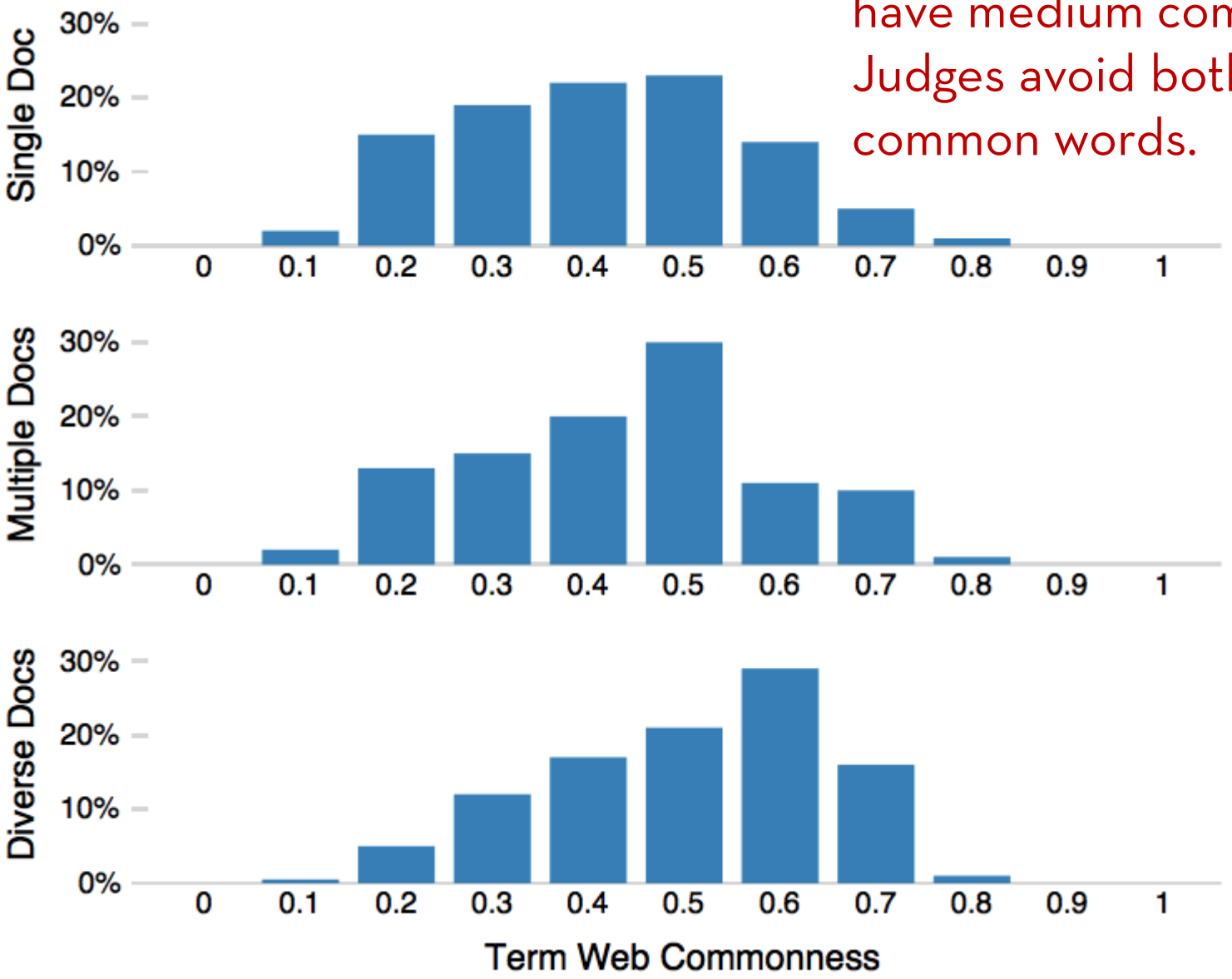
Term Commonness

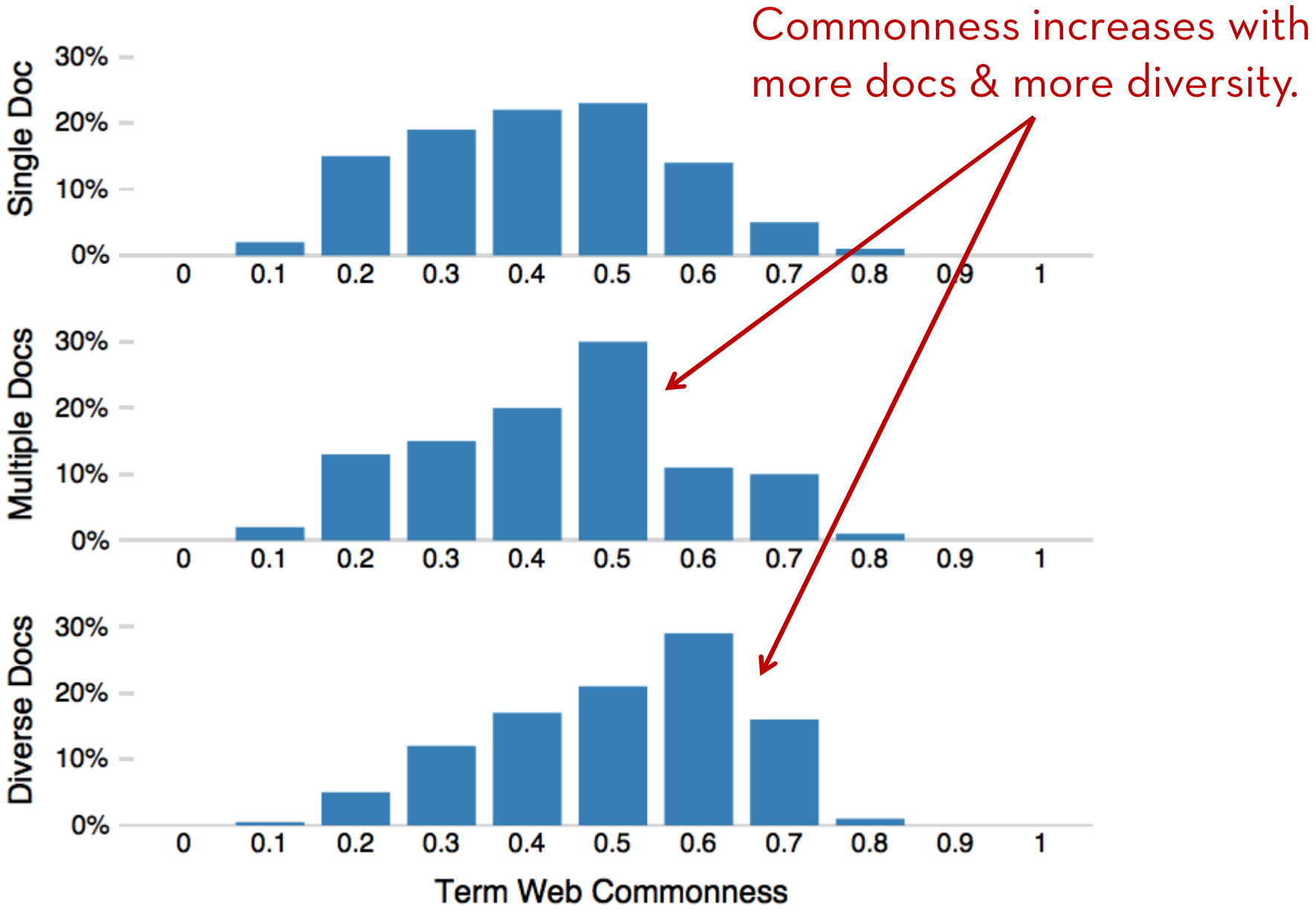
$$\log(tf_w) \ / \ \log(tf_{the})$$

The normalized term frequency relative to the most frequent n-gram, e.g., the word “the”.

Measured across an entire corpus or across the entire English language (using Google n-grams)

Selected descriptive terms have medium commonness. Judges avoid both rare and common words.




Commonness increases with more docs & more diversity.
Single Doc
Multiple Docs
Diverse Docs
30%
20%
10%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Term Web Commonness

Grammar: Technical Term Patterns

Technical Term	$T = (A\|N)+ (N\|C) \mid N$
Compound Tech. Term	$X = (A\|N)* N$ *of* T

Regular expressions over part-of-speech tags.

A = adjective, *N* = noun, C = cardinal number.

Prior work suggests these patterns can be used to identify important terms in text.

Over 4/5 of selected terms match pattern!

Method – Part 2

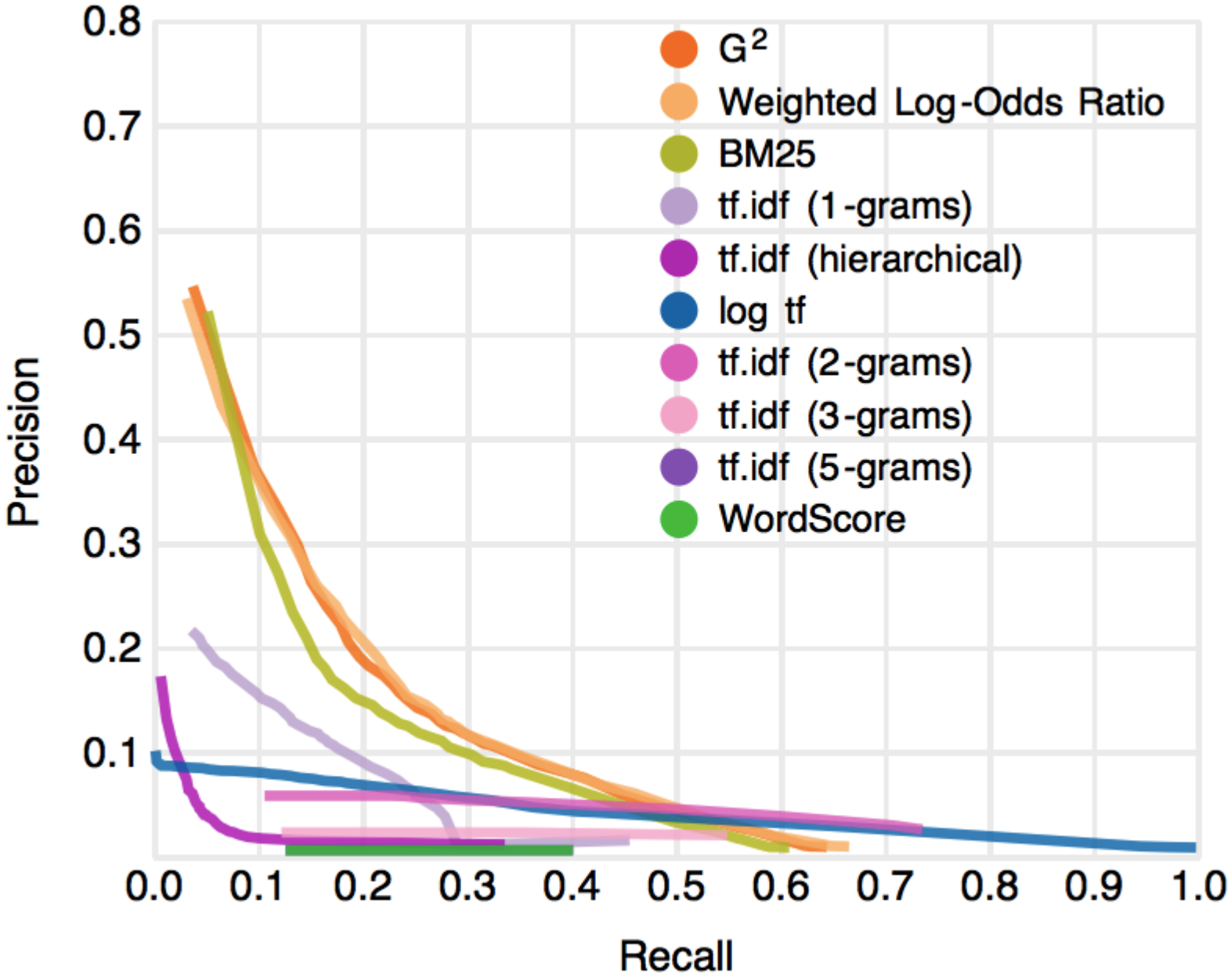
Build a statistical model of keyphrase quality

Train a logistic regression model
- Positive examples: selected phrases
- Negative examples: randomly sampled phrases

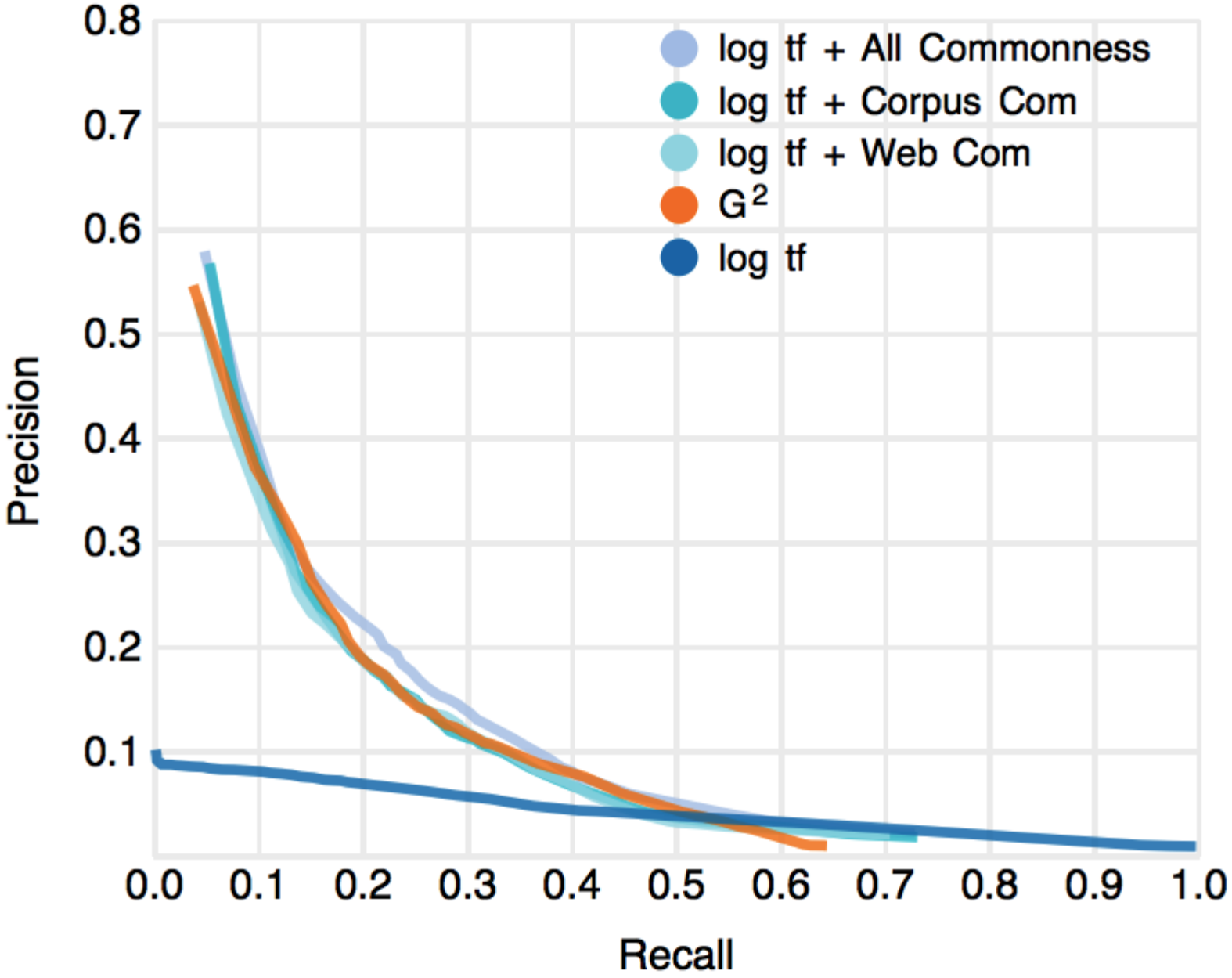
Assess contributions of four classes of features:
- Freq stats, commonness, grammar & position

Evaluate the phrases selected by our model using precision & recall measures

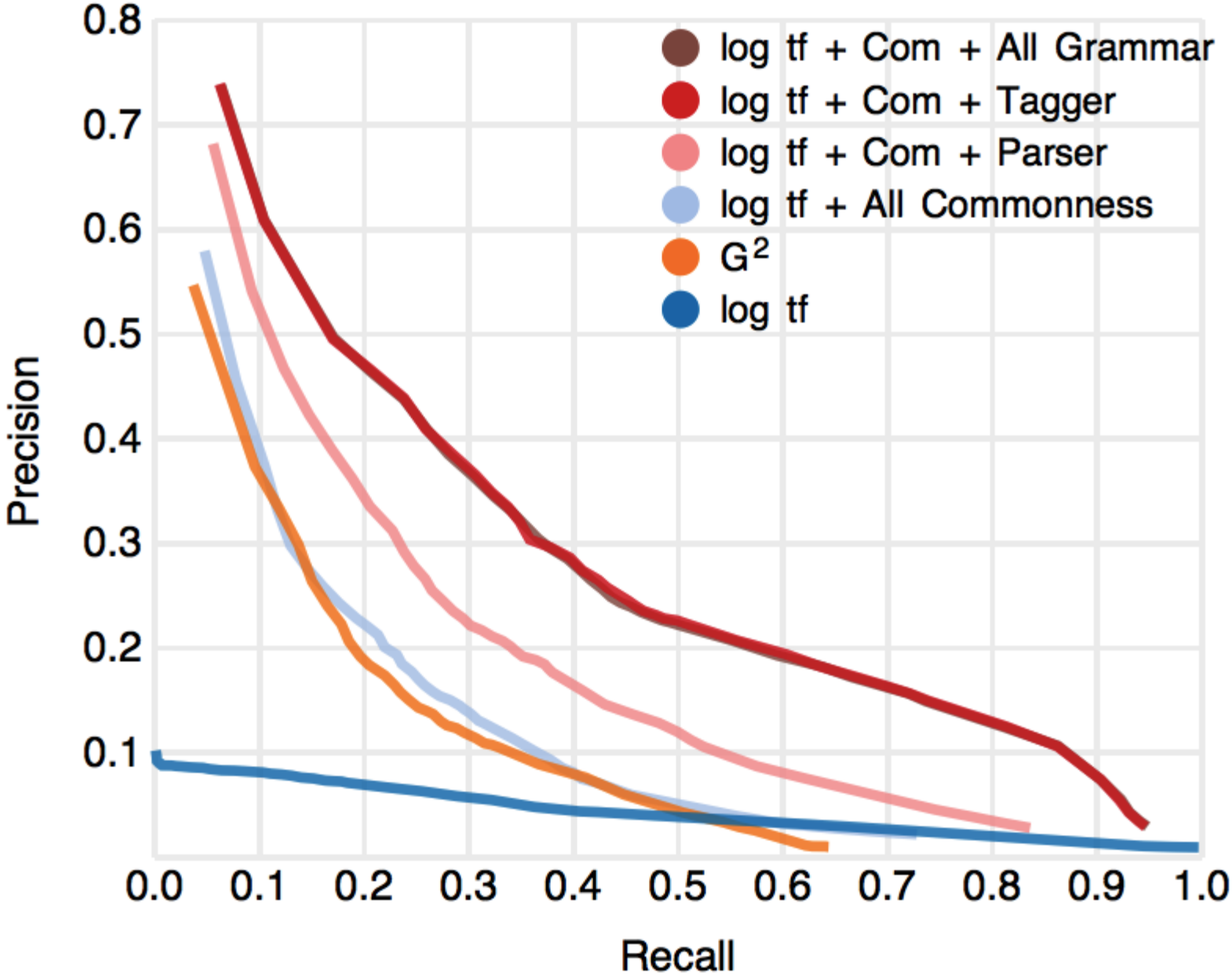
Frequency Statistics Only



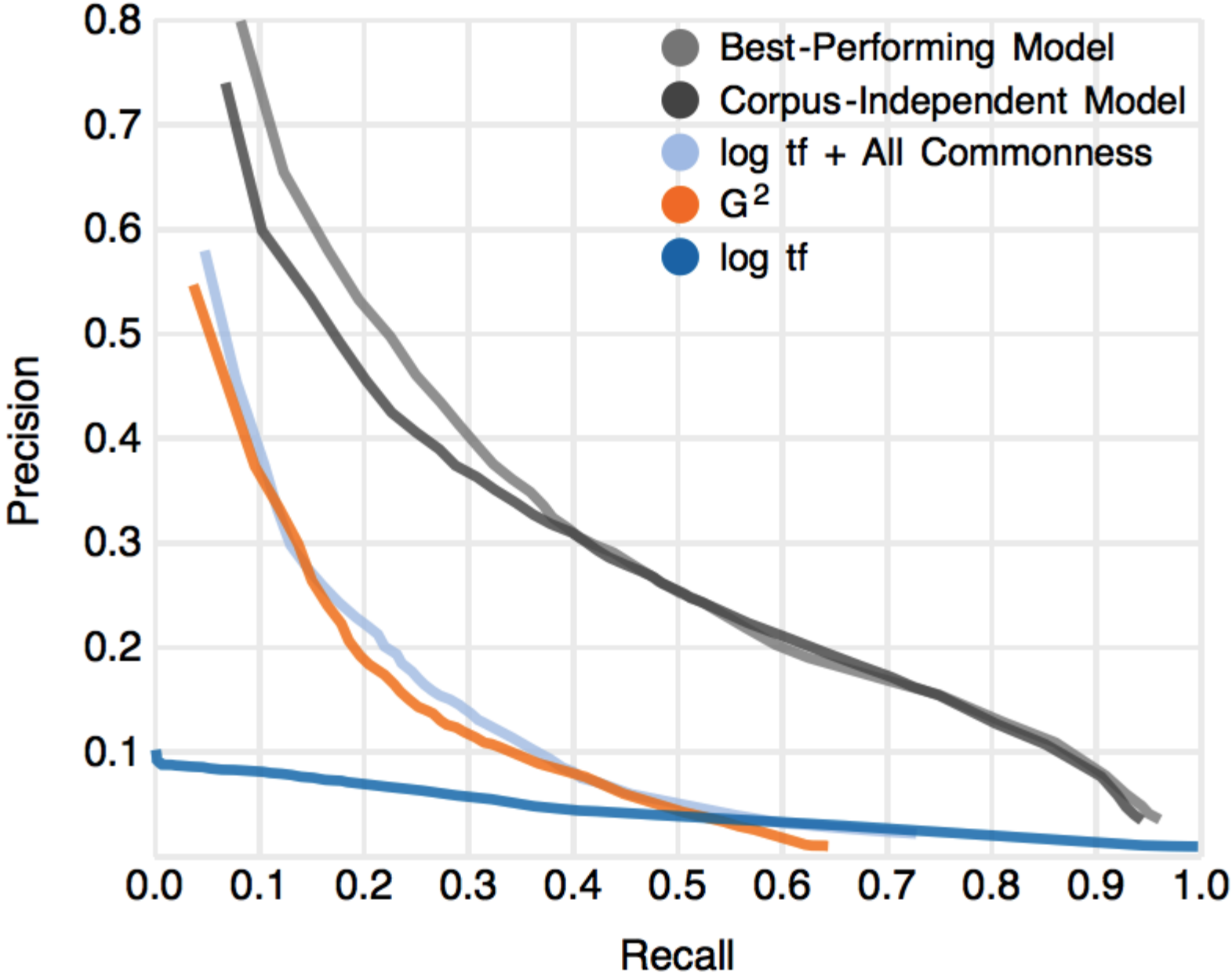
Adding Commonness



Adding Grammatical Features



Adding Positional Features

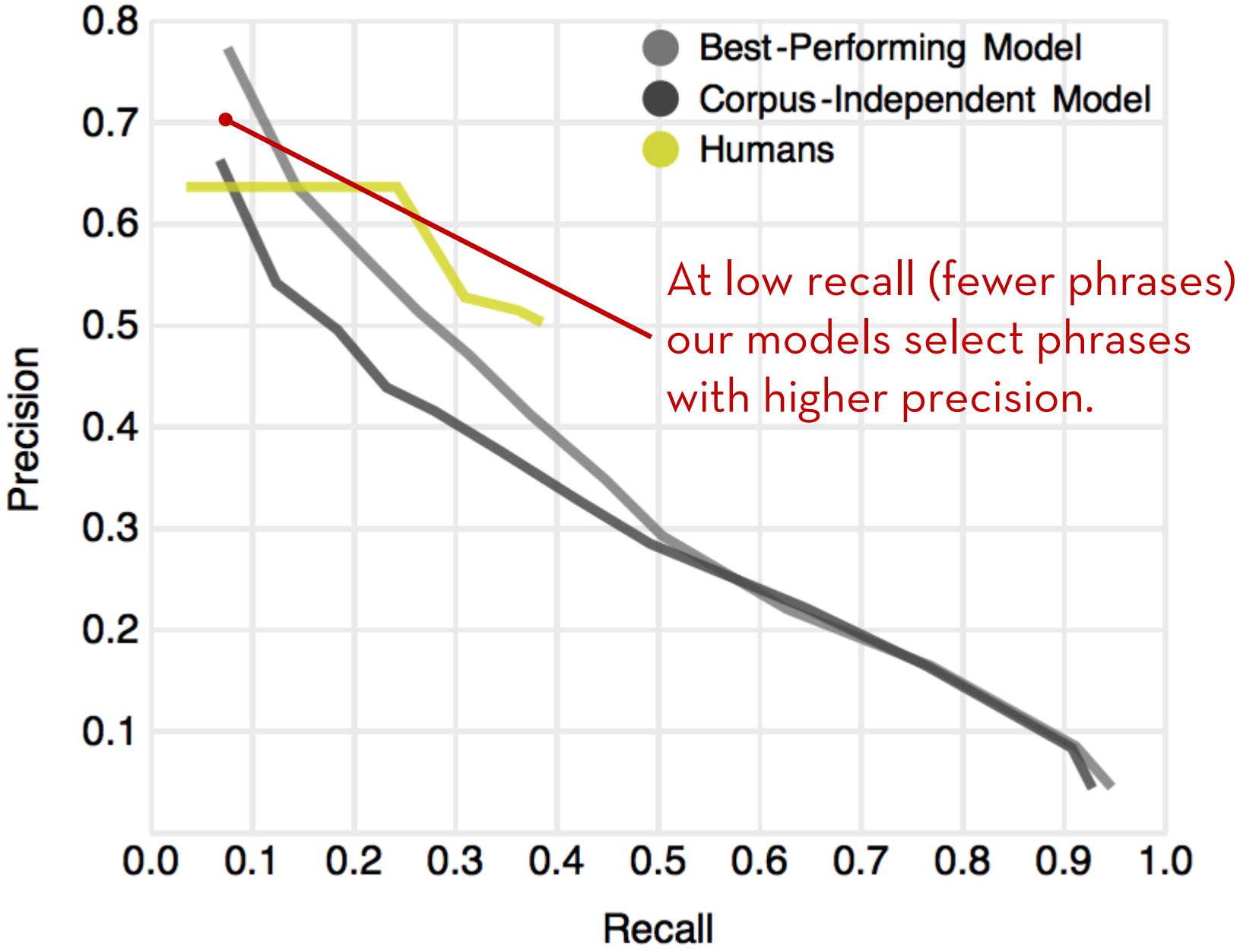


Model Feature	Coefficients
constant	-2.3550***
log(tf)	0.9390***
WC $\in (0\%, 20\%]$	0.1770
WC $\in (20\%, 40\%]$	0.2304*
WC $\in (40\%, 60\%]$	0.0158
WC $\in (60\%, 80\%]$	-0.6205***
WC $\in (80\%, 100\%]$	-1.9081***
relative first occurrence	0.4800**
first sentence	0.9386***
full tech. term	-0.5015
partial tech. term	1.4461**
full compound tech. term	1.1373
partial compound tech. term	1.1806*

Fitted Parameters for a Corpus-Independent Model

WC: web commonness bins, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Compare to User-Selected Phrases



Automatic Keyphrase Extraction

Phase 1 Score candidate terms using our keyphrase quality regression model

Phase 2 Eliminate redundancy by grouping similar terms based on word overlap plus entity and acronym resolution.

- *“analysis”, “data analysis”, ...*
- *“Barack Obama”, “Obama”, ...*

A fighter jet rain check

Story and video by Chamila Jayaweera

Have you ever thought about what it takes to make sure that sea-based fighter jets stay dry?

When it comes to the F/A-18 Super Hornet, Boeing engineers in St. Louis use a special process called the Water Check Test to rule out areas where moisture could seep into the aircraft and its electronics suite.

Program experts douse the jet with simulated rain at a 15-inch-per-hour rate for about 20 minutes inside an enormous hangar in St. Louis.

CHAMILA JAYAWEERA/BOEING

The Water Check team rolls in a large metal frame, which they affectionately call their "spray tree," over a Super Hornet inside a St. Louis hangar.

"Our ultimate customers are U.S. Navy fighter pilots, and we want to ensure their safety in flight and on the ground, and water-tight integrity of the aircraft also helps increase their effectiveness," said Boeing's Rich Baxter, F/A-18 Super Hornet final assembly manager.

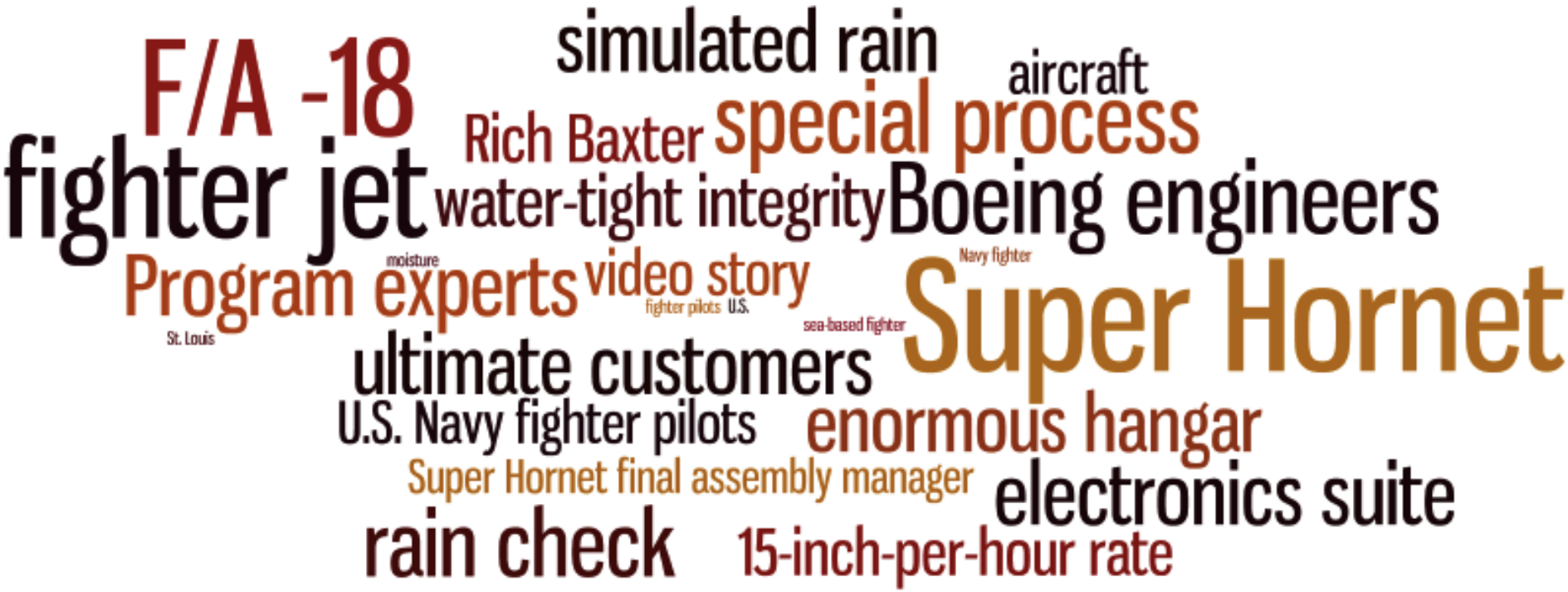
To find out moreabout how the process works and watch the action unfold, click above to see the video story.

G^2

fighter
F/A
Hornet
Super
Boeing
-18
rain
St.
jet
Louis
15-inch-per-hour
douse
hangar
water-tight
Check
Baxter
sea-based
aircraft
Rich
seep
click
Navy
sure
Water
moisture
watch
enormous
stay
want

Our Technique

Super Hornet
F/A -18
fighter jet
Boeing engineers
special process
rain check
electronics suite
Program experts
simulated rain
ultimate customers
enormous hangar
water-tight integrity
Rich Baxter
15-inch-per-hour rate
video story
aircraft
U.S. Navy fighter pilots
Super Hornet final assembly manager
U.S.
Navy fighter
fighter pilots
sea-based fighter


F/A -18
simulated rain
aircraft
fighter jet
Rich Baxter
special process
water-tight integrity
Boeing engineers
Program experts
moisture
video story
fighter pilots
U.S.
Navy fighter
Super Hornet
St. Louis
sea-based fighter
ultimate customers
U.S. Navy fighter pilots
enormous hangar
Super Hornet final assembly manager
electronics suite
rain check
15-inch-per-hour rate

Pilots push 787 Dreamliner to the limit

By Bernard Choi

Captain Mike Carriker walks into his office in Everett, Wash for the first time in weeks and turns on his laptop. He has a lot of catching up to do.

On the way to his desk, he walked by a sign that describes exactly why the 787 Chief Test Pilot had been out of the office: 'Gone Flying'

> "It's tiring because we have to have very specific conditions. We need dry runways, then wet runways. We want wind, then no wind and so on."

In the past month, Capt. Carriker and a team of Boeing pilots have logged thousands of miles chasing the perfect conditions as they continue to test the all-new 787 Dreamliner.

"It's tiring because we have to have very specific conditions. We need dry runways, then wet runways. We want wind, then no wind and so on."

Captain Carriker sat down to describe some of the recent tests he and the team have conducted.

BOEING PHOTO

To test the 787 Dreamliner's takeoff performance, pilots perform multiple takeoffs at varying speeds and conditions.

Takeoff Performance:

"You start with regular takeoff like we recommend the airline crews do all the time and then what we do is we go around and look at how much you can vary from it. You rotate before the predicted air speed. You rotate after the air speed. You rotate too fast. You rotate too slow. It's all to define there's a tolerance for error."

Velocity Minimum Unstick:

This test establishes the lowest speed the airplane can leave the ground and requires putting the tail on the runway. "You don't want to hold the tail on the

G^2

runway
Carriker
787
airplane
brake
rotate
pilots
Capt.
Dreamliner
tail
Captain
office
trucks
ground
hold
get
speed
just
catch
team
Mike
Runway
Takeoff
Unstick
Wash
all-new
anti-skid
gallons
grueling
similate
tiring
ungrooved
â€¦
want

Our Technique

Dreamliner
Capt. Mike Carriker
runway
pilots
airplane
air speed
tail
Chief Test Pilot
perfect conditions
recent tests
airline crews
Takeoff Performance
big brake
take-off speed
specific conditions
all-new 787
good covering
Wet Runway
regular takeoff
limit
wind
grueling schedule
manual brakes


individual parts
big brake
take-off speed
regular takeoff
wind
altitude
grueling schedule
Velocity Minimum Unstick
good covering
gallons
flight deck
all-new 787
Takeoff Performance
Capt. Mike Carriker
landings
Wash
other office
Dreamliner
desk Flying
thousands
Unstick
past month
laptop
trucks
Everett
air speed
Chief Test Pilot
airplane
specific conditions
manual brakes
perfect conditions
limit
airline crews
recent tests
runway
normal technique
pilots
brief break
anti-skid system
tail
Wet Runway

Why Visualize Text?

Understanding – get the "gist" of a document

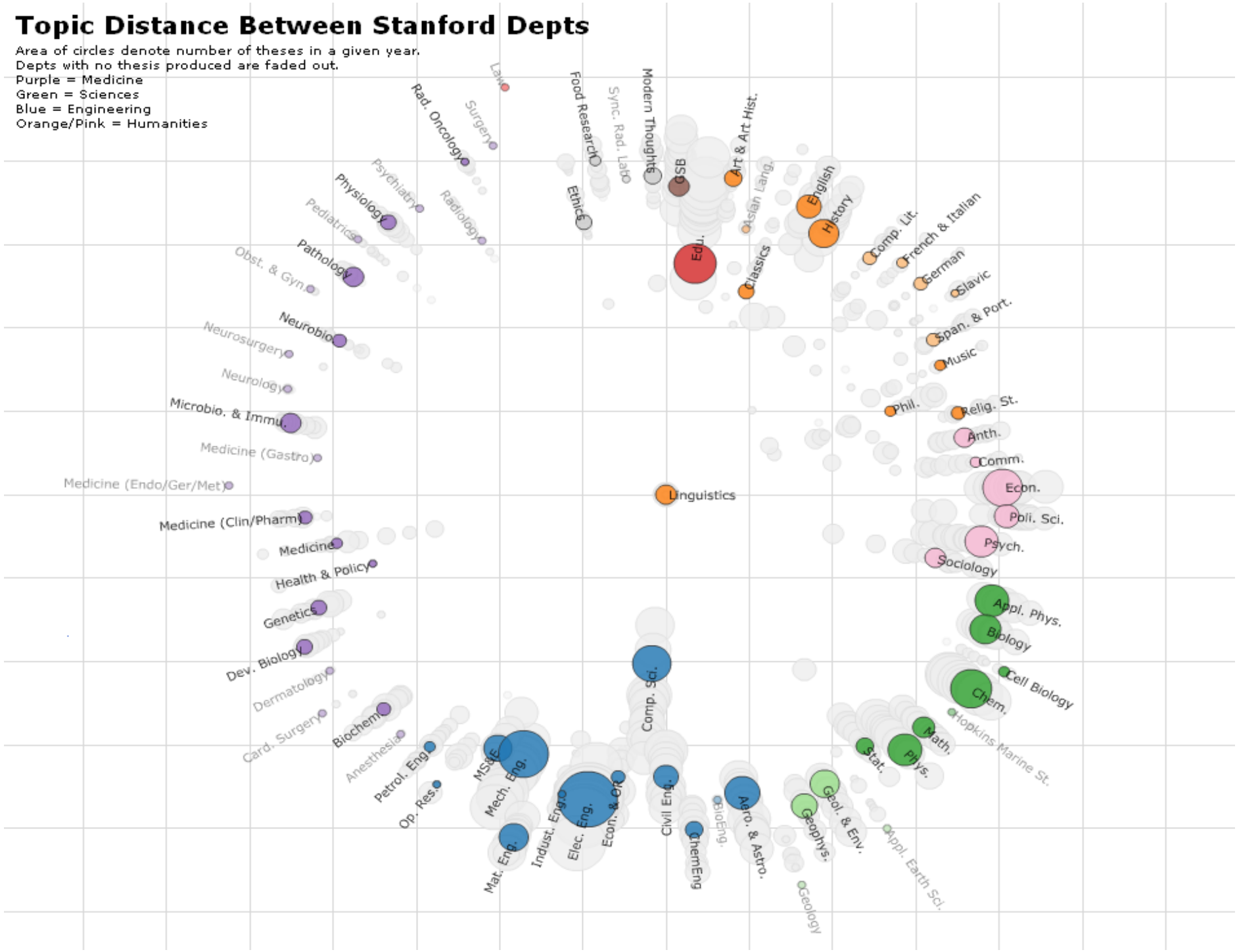
Grouping – cluster for overview or classification

Compare – compare document collections, or inspect evolution of collection over time

Correlate – compare patterns in text to those in other data, e.g., correlate with social network

Topic Distance Between Stanford Depts

Area of circles denote number of theses in a given year.
Depts with no thesis produced are faded out.
Purple = Medicine
Green = Sciences
Blue = Engineering
Orange/Pink = Humanities



1 Visualization

2 Networks

3 Text

Student Collaborators

Mike Bostock

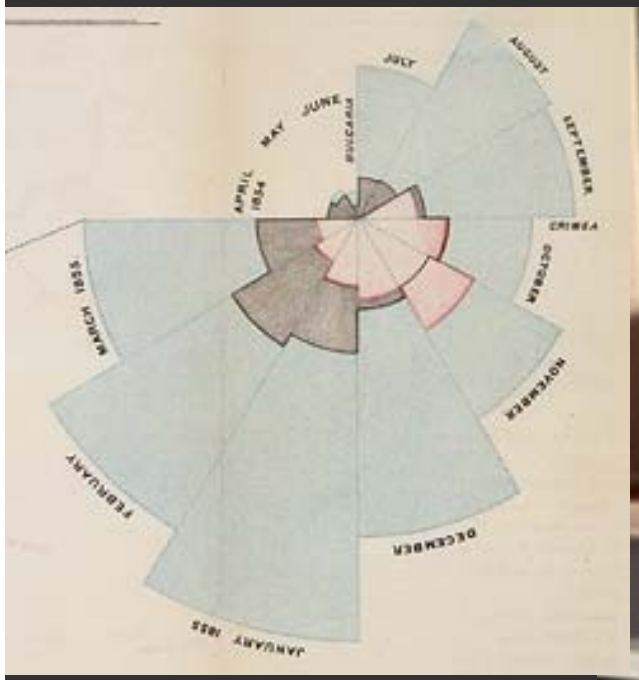
Jason Chuang

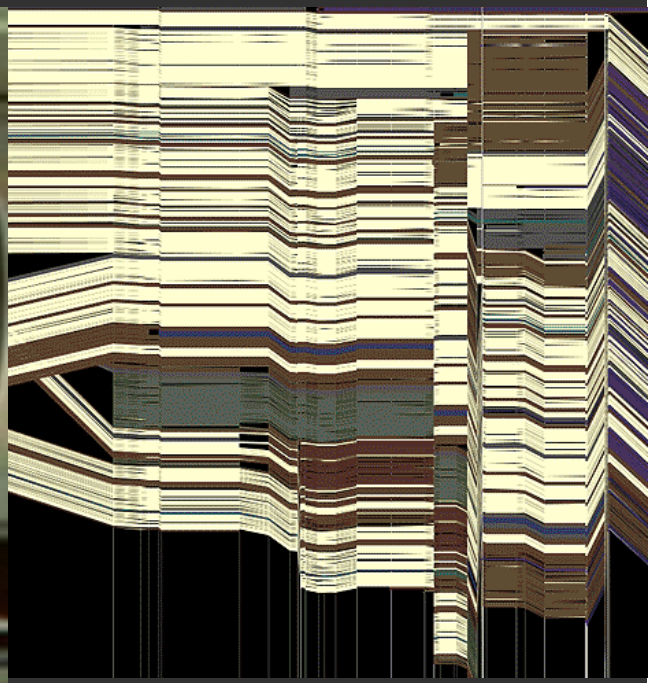
Sean Kandel

Diana MacLean

Vadim Ogievetsky

Interactive Visual Analysis for Networks & Text





Jeffrey Heer http://vis.stanford.edu