Information Visualization

Dianna Xu
Department of Computer Science
Bryn Mawr College

Big Data

- The amount of available data is quickly outpacing our ability to understand and use it in meaning full ways.
- The eventual solution to big data is likely really good black boxes.
- Black boxes only give answers if you know right questions to ask.

Visualization and Explorative Data Analysis

- Visualization allows one to view things in global context.
- Increase the likelihood of spotting unexpected trends and perspectives.
- Allows new questions to be asked.
- The life cycle of explorative data analysis with visualization is likely an iterative process

Visualization Types

- Scientific visualization
  - numerical, high precision and has relative simple spatial relationships
  - accurate representation of numerical proportions and realistic rendering of physical properties
- Data visualization
  - social/economical data
  - highly categorical and have strong association with physical locations/coordinates on the map

Information Visualization

- Arbitrary or complex relationships with no clearly prescribed spatial representation choices
- Abstract representation techniques
- Communication is an important goal
- infoviz versus “data journalism”

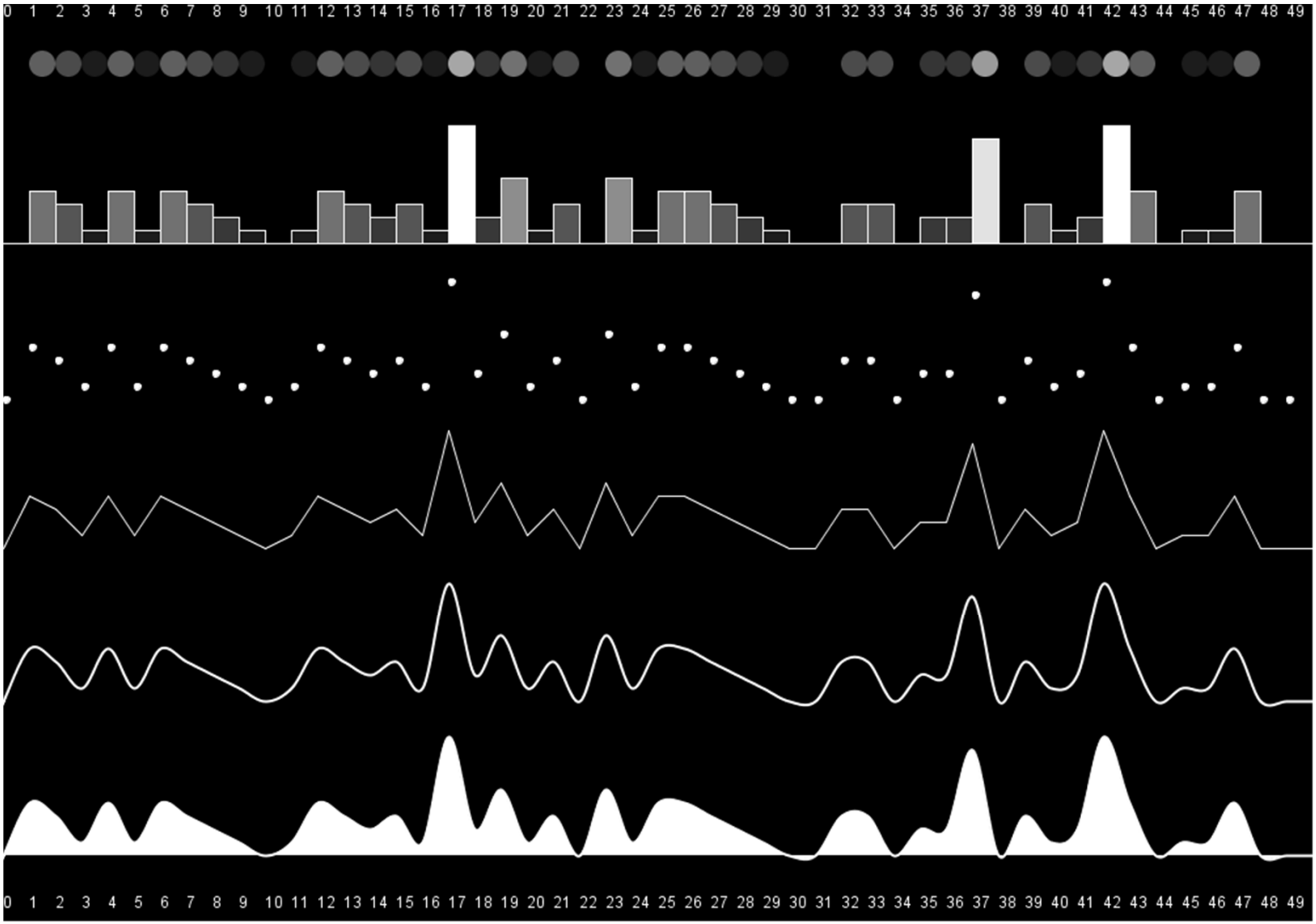
The Visualization Process

- Acquire
- Parse
- Filter
- Mine
- Represent
- Refine
- Interact

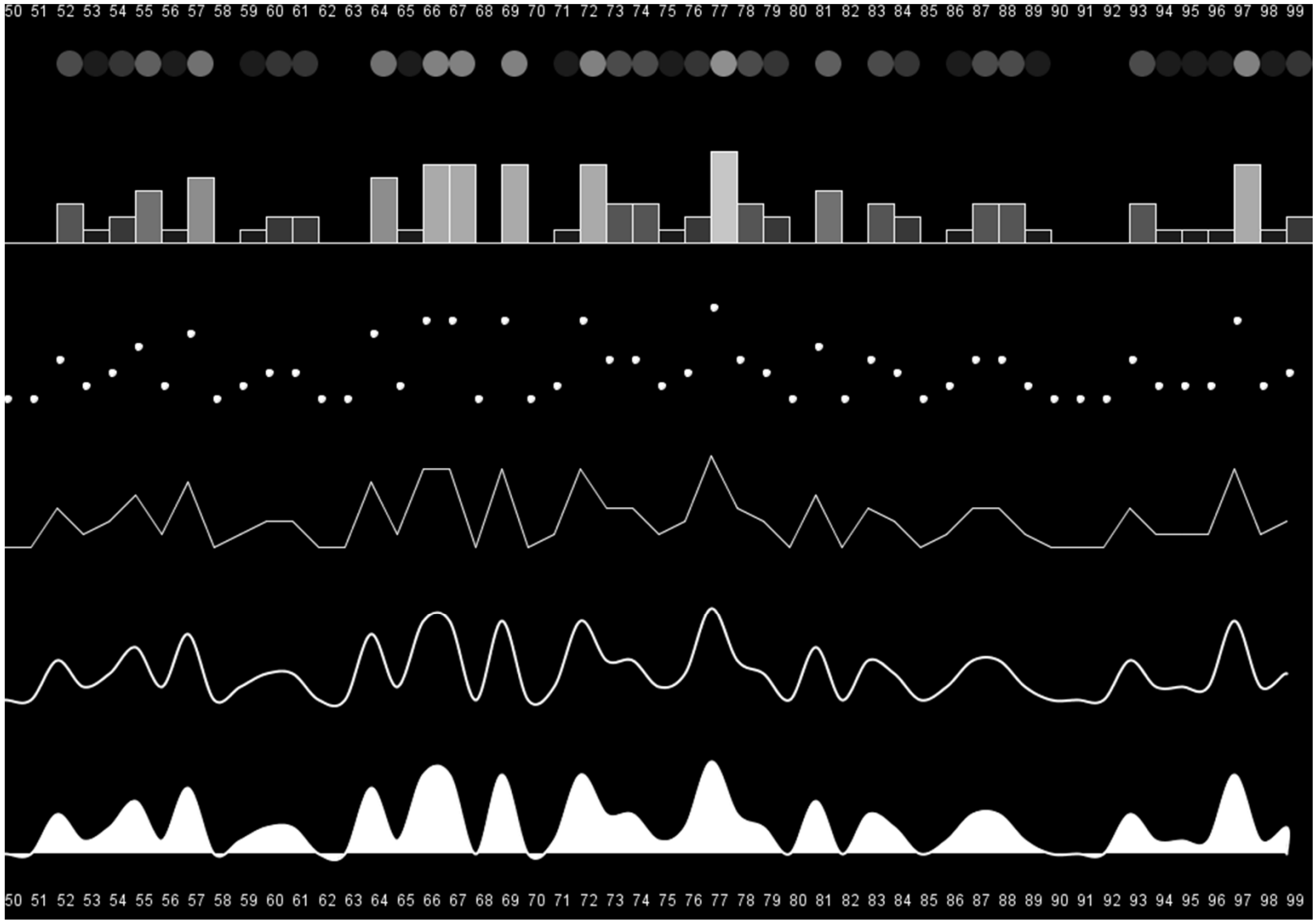
Mapping Numbers

- How to represent numbers with visual primitives so that the relative values of numbers can be quickly deduced from visual cues
- Example with simple data set
  - 255 integers valued 0-99
  - collected on twitter as random numbers

Mapping Numbers



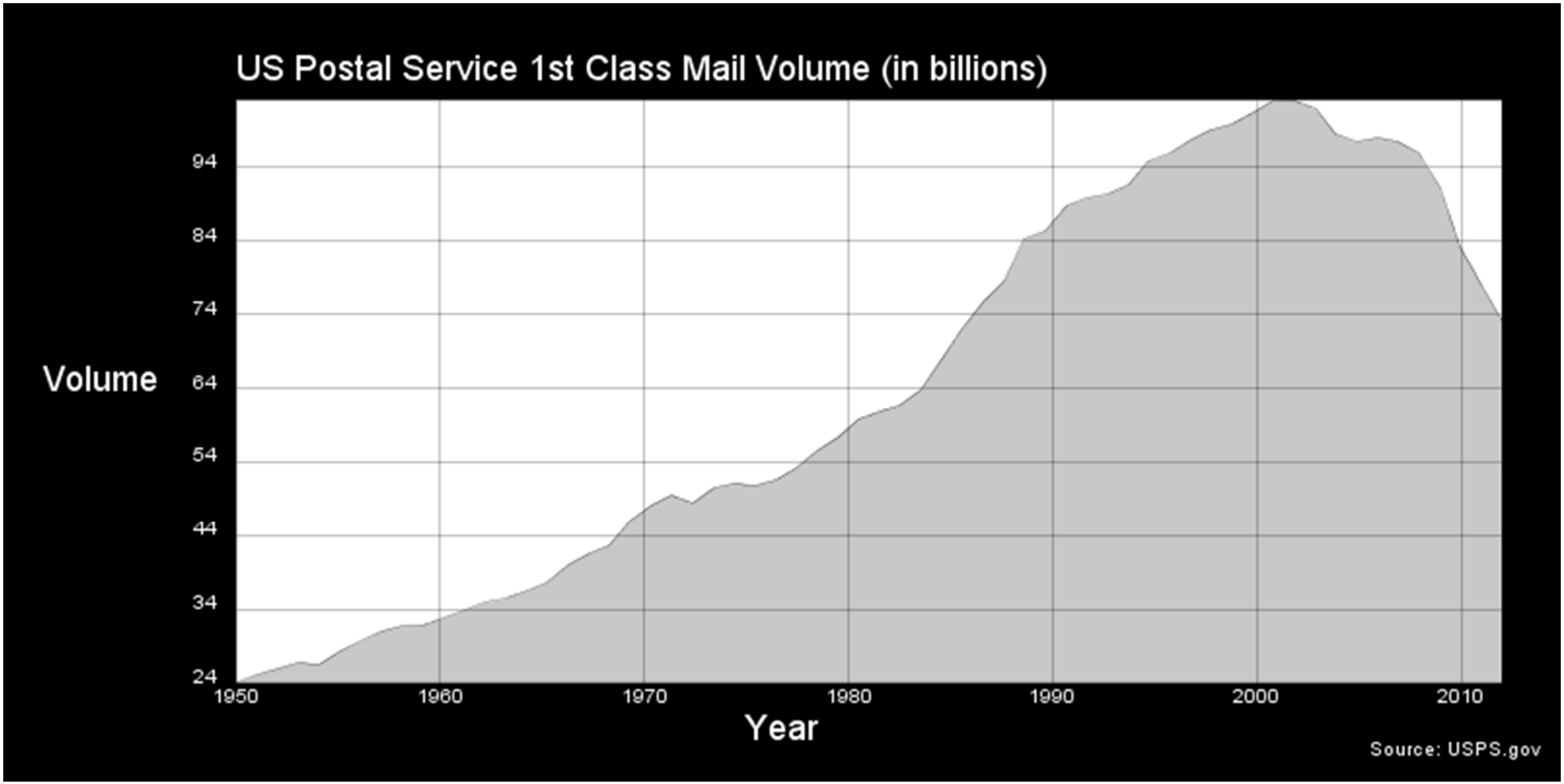
Mapping Numbers



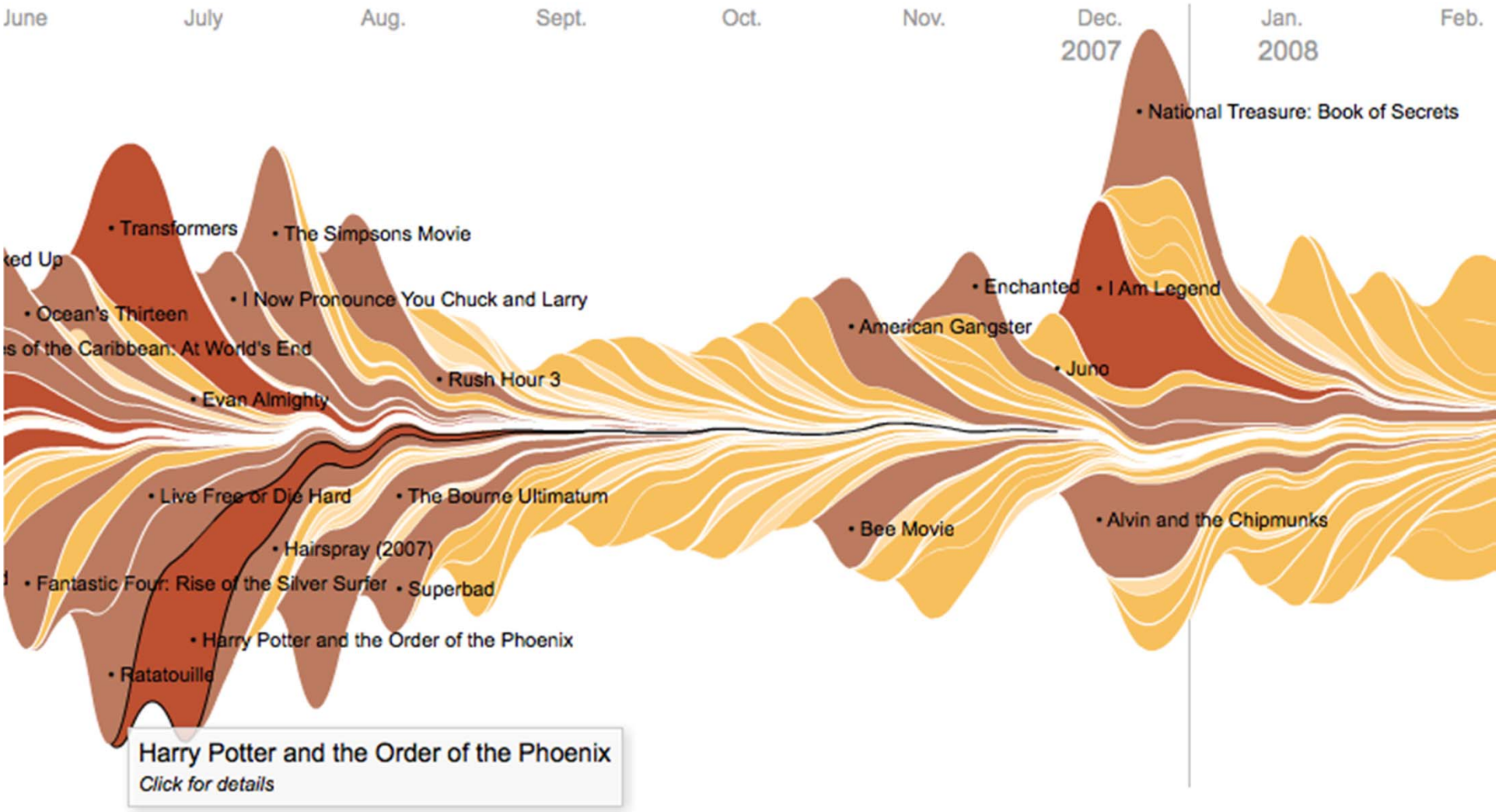
Information Dimensions

- A fundamental challenge of visualization design is to add more information dimensions without introducing clutter
- Display is 2D
- Going 3D is not a solution

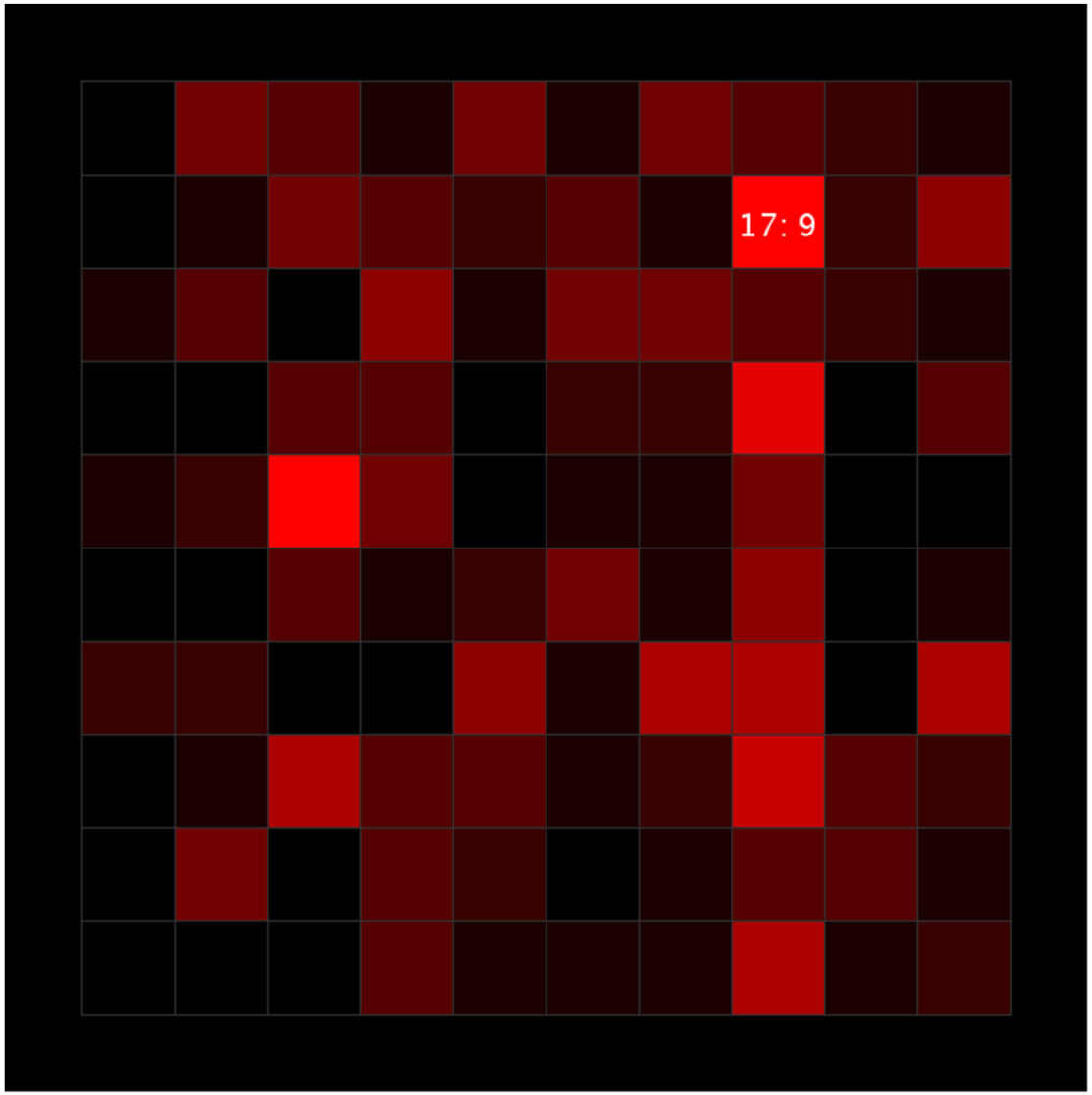
Time Series


US Postal Service 1st Class Mail Volume (in billions)
Volume
94
84
74
64
54
44
34
24
1950
1960
1970
1980
1990
2000
2010
Year
Source: USPS.gov

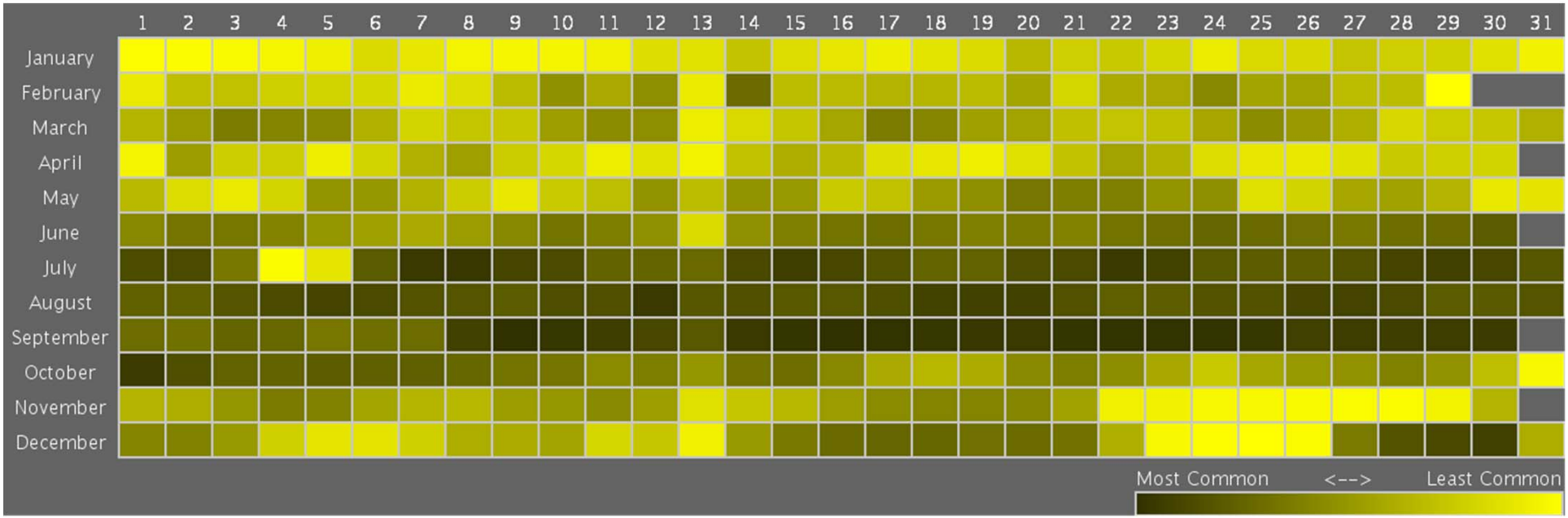
Stacked Graph



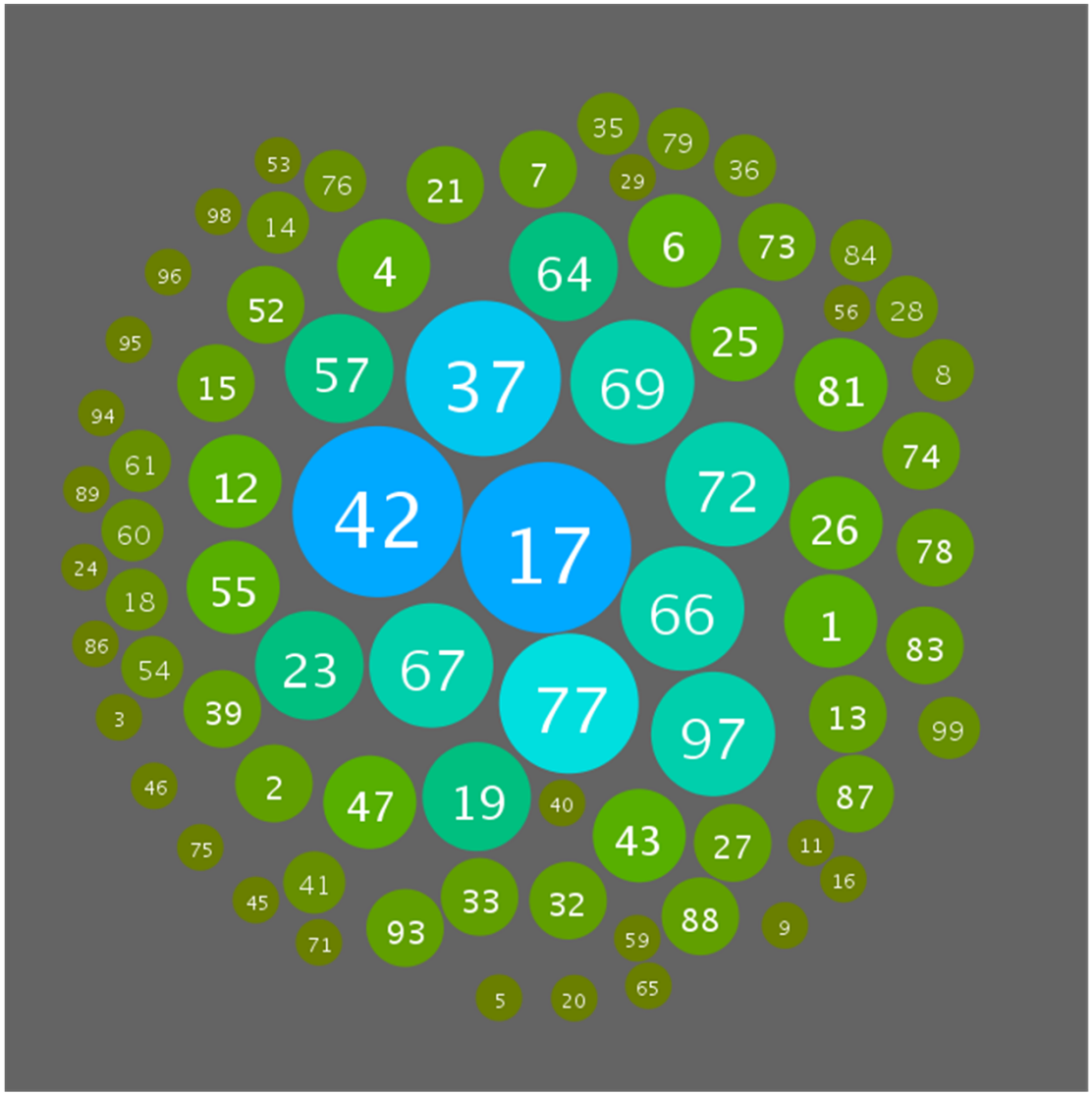
Heat Map



Birthdays


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
January
February
March
April
May
June
July
August
September
October
November
December
Most Common <--> Least Common

Proportional Symbols



Twitter Frequencies

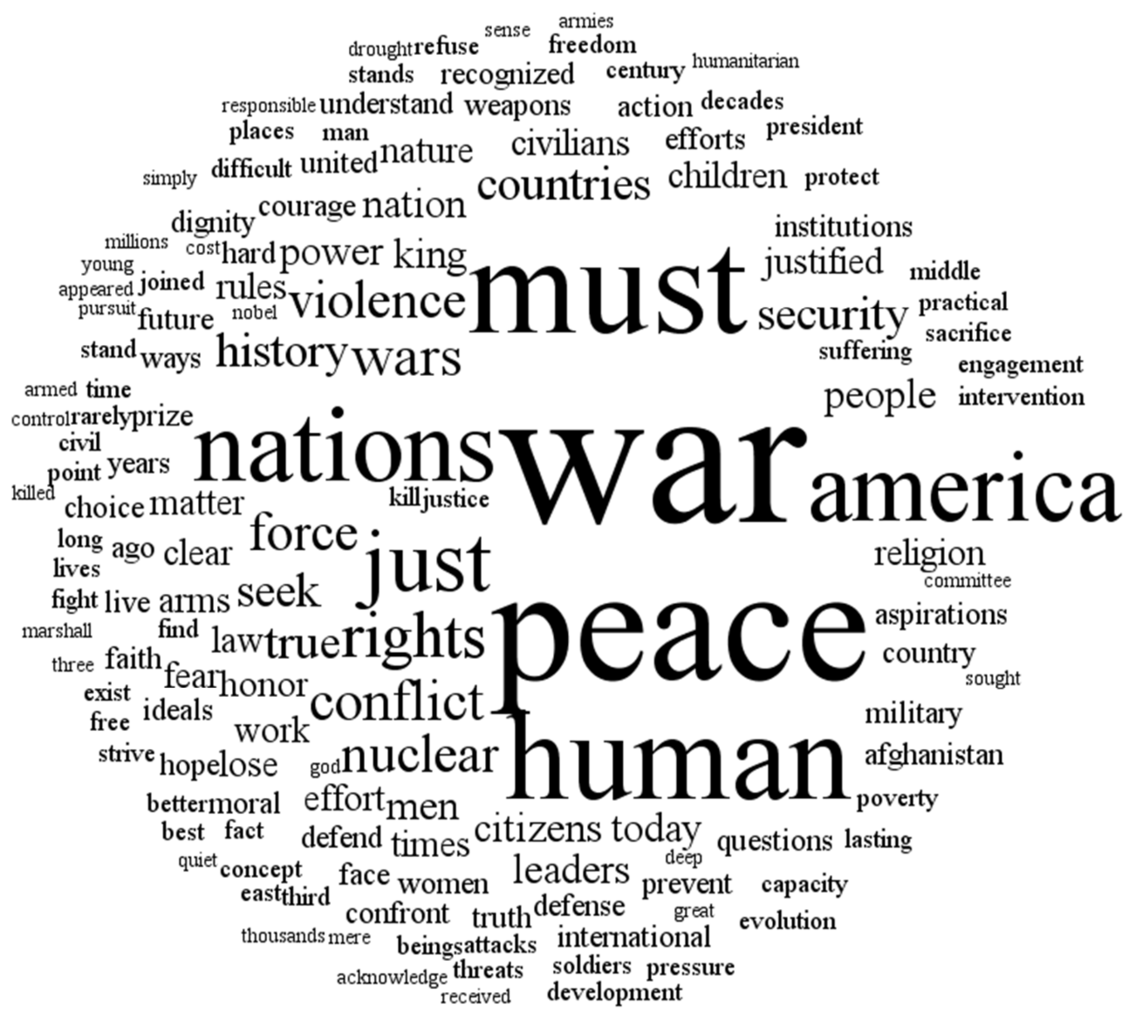
World Cloud


became
prize
refuse
pressure
rules
law
think
another
arms
stands
need
truth
war
efforts
evolution
years
joined
difficult
states
president
violence
many
women
decades
just
children
concept
ideals
yet
clear
king
sacrifice
nuclear
middle
matter
ways
times
protect
power
nothing
country
lose
poverty
cause
let
aspirations
stand
nation
military
sometimes
world
freedom
america
still
even
new
justice
use
history
live
best
may
never
lives
east
free
third
america's
within
fear
conflict
also
without
international
action
religion
one
faith
force
know
weapons
peace
security
citizens
rights
nations
ago
justified
way
better
see
time
courage
future
questions
soldiers
attacks
development
people
believe
countries
moral
leaders
united
confront
places
alone
capacity
different
wars
together
like
make
recognized
human
dignity
seek
necessary
afghanistan
understand
right
choice
work
civilians
enough
nature
come
today
exist
always
institutions
long
true
honor
intervention
engagement
every
prevent
something
first
men
effort
must
century
practical
hope
defend
beings
hard
face

Packing

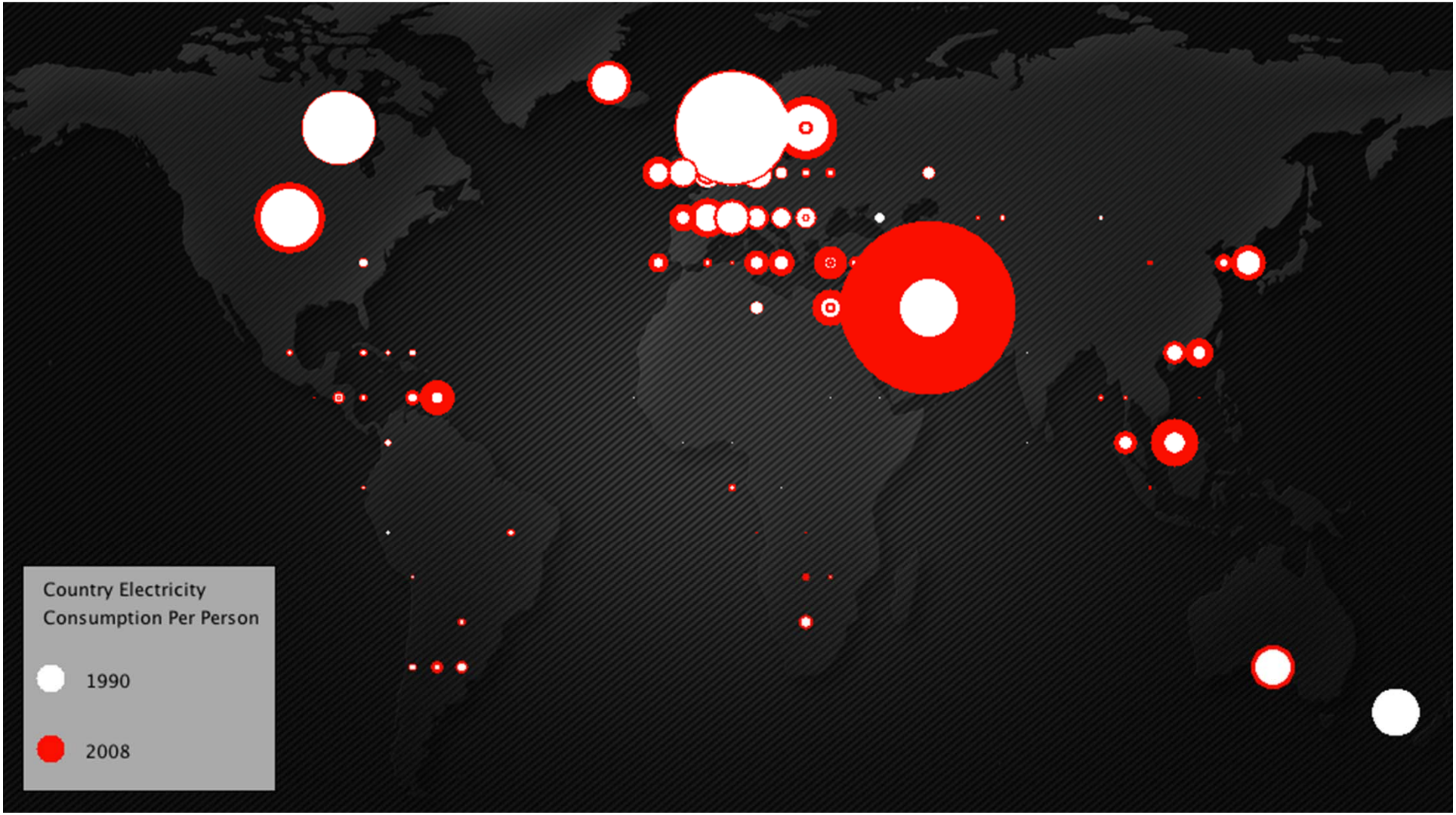
- Randomized placement allowing overlap
- Randomized greedy algorithm
  - fit the largest tiles first at randomly chosen location
  - if there is overlap, try again
- Exhaustive search
- Exhaustive search guided by space filling curve
  - Spiral, Peano, Hilbert, etc

Spiral


war
peace
must
human
nations
america
just
rights
conflict
nuclear
violence
wars
history
security
force
countries
people
religion
seek
true
law
power
king
nation
nature
civilians
children
justified
institutions
leaders
citizens
today
men
effort
times
defense
international
prevent
questions
afghanistan
military
country
aspirations
matter
choice
arms
clear
work
hope
lose
honor
fear
faith
ideals
moral
defend
face
women
confront
truth
rules
future
ways
years
prize
dignity
courage
hard
recognized
understand
weapons
action
efforts
united
freedom
century
decades
president
protect
middle
practical
sacrifice
engagement
intervention
suffering
committee
sought
poverty
lasting
capacity
evolution
pressure
soldiers
development
threats
attacks
beings
third
east
concept
fact
best
better
strive
free
exist
fight
lives
long
point
civil
time
stand
joined
kill
justice
live
find
ago
rarely
armed
control
killed
marshall
three
quiet
thousands
mere
acknowledge
received
great
deep
god
nobel
pursuit
appeared
young
millions
cost
simply
difficult
places
man
responsible
stands
drought
refuse
sense
armies
humanitarian

Intersection via Pixel Buffer

Map-based



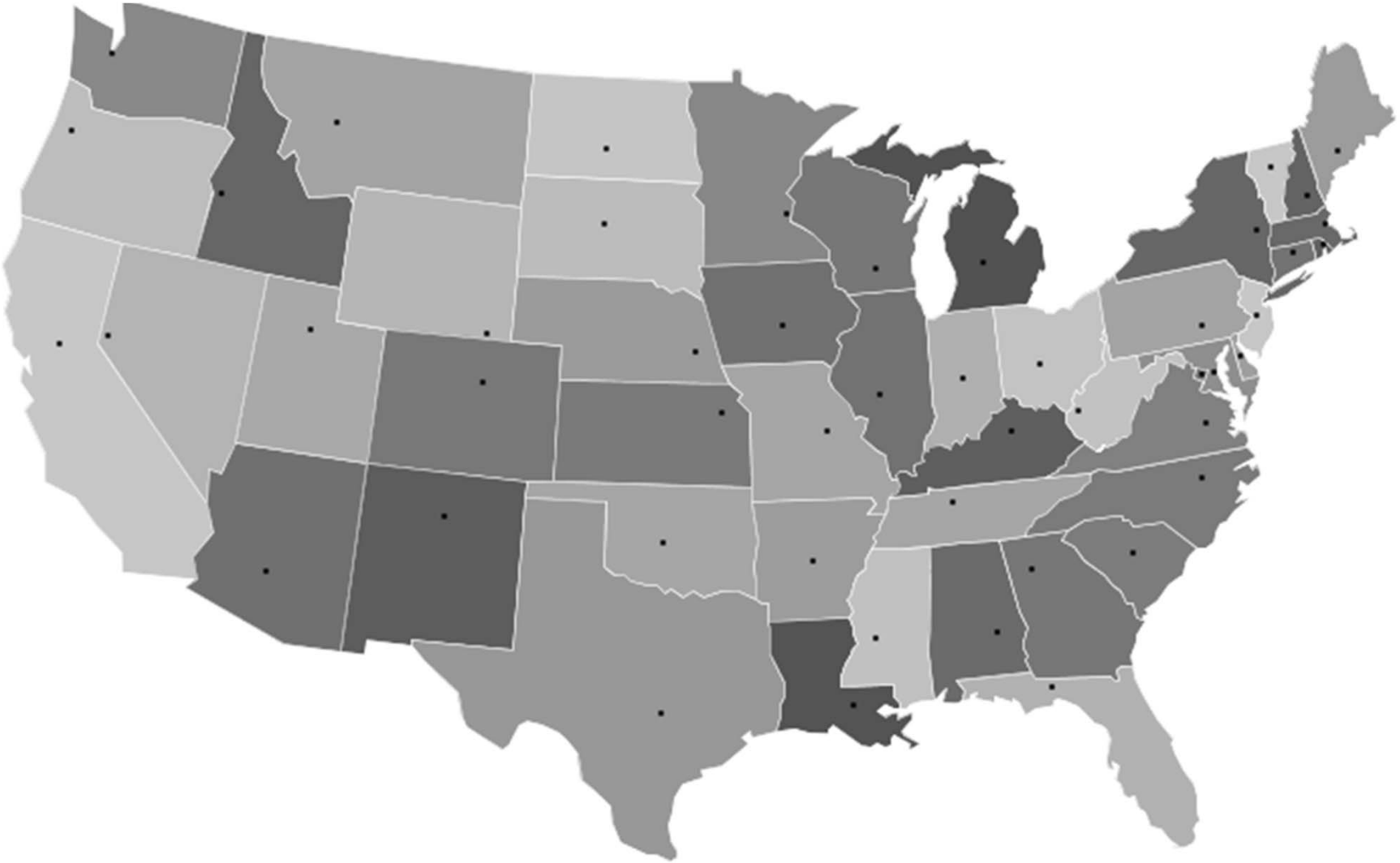
Coordinate Conversions

- Translate physical coordinates to drawing coordinates
- Geographical coordinates
- Modern world map empolying a Universal Transverse Mercator (UTM) projection
- What if your data comes with names of countries or states?

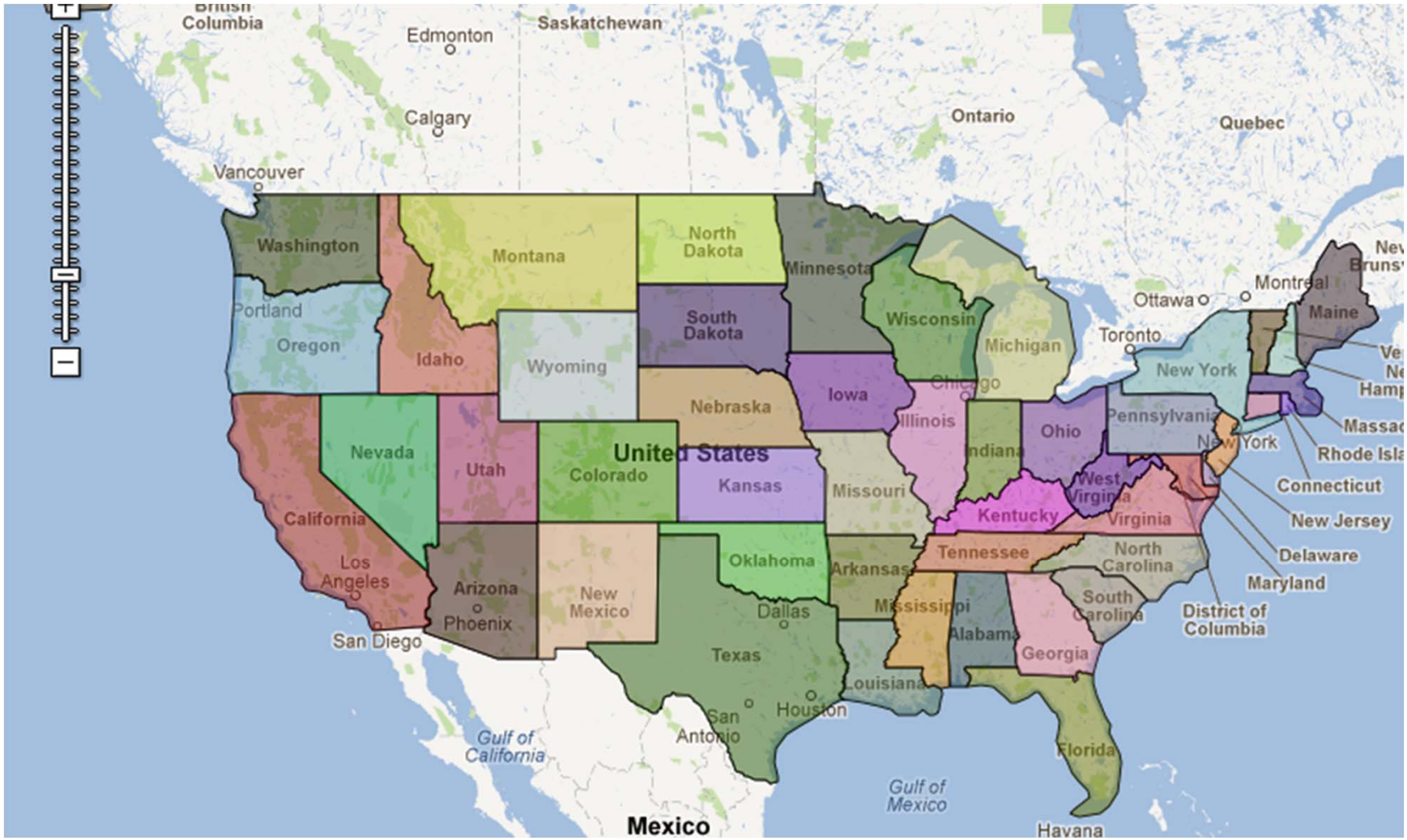
A Choropleth Map

- Thematic maps require that you color an entire area (state) with a particular color
- Need polygonal outlines of a state
- A US political map of SVG format contains coordinates as an XML file
- Combine with Google Maps API to create interactive map overlay

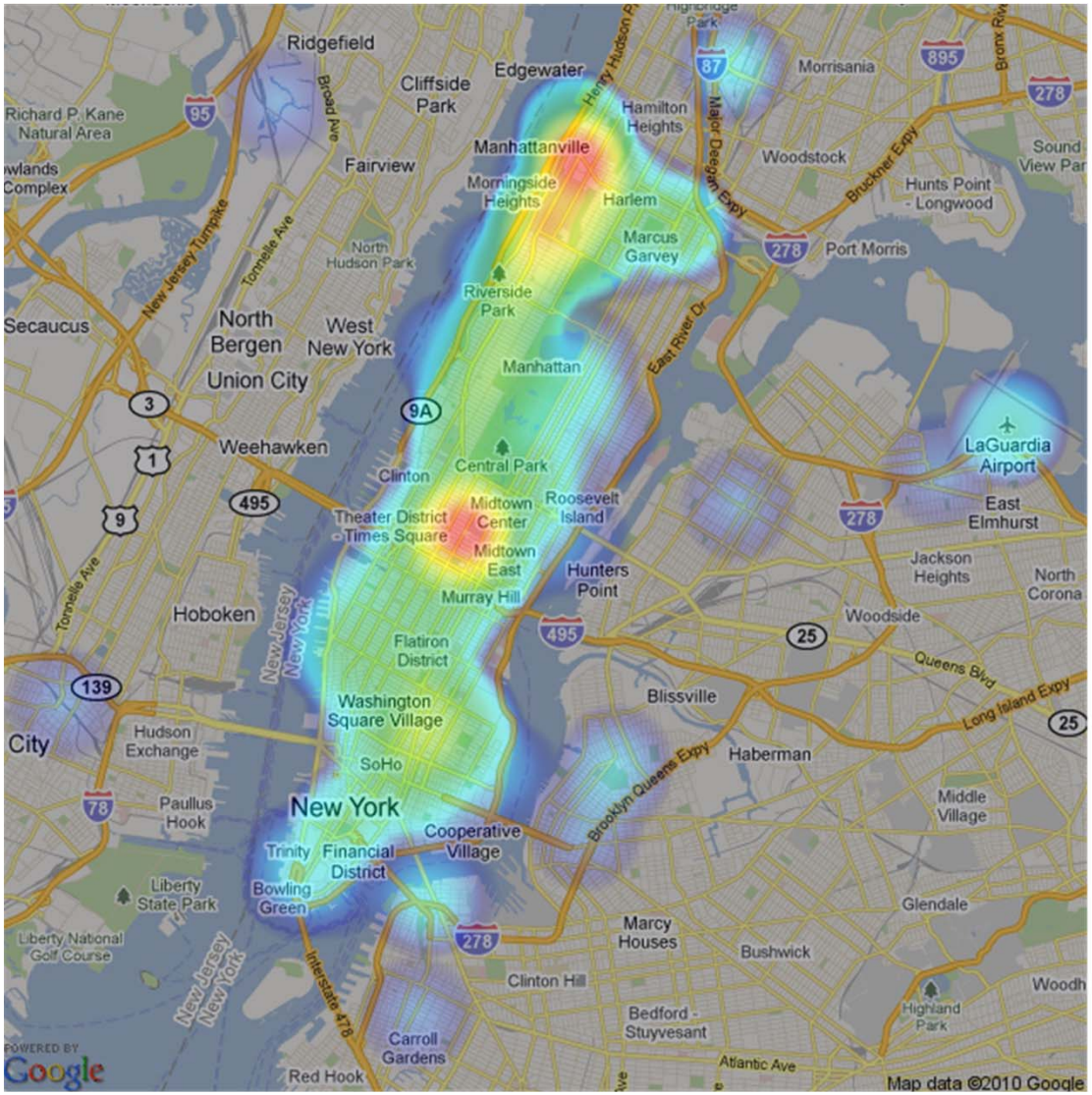
States as Polygons

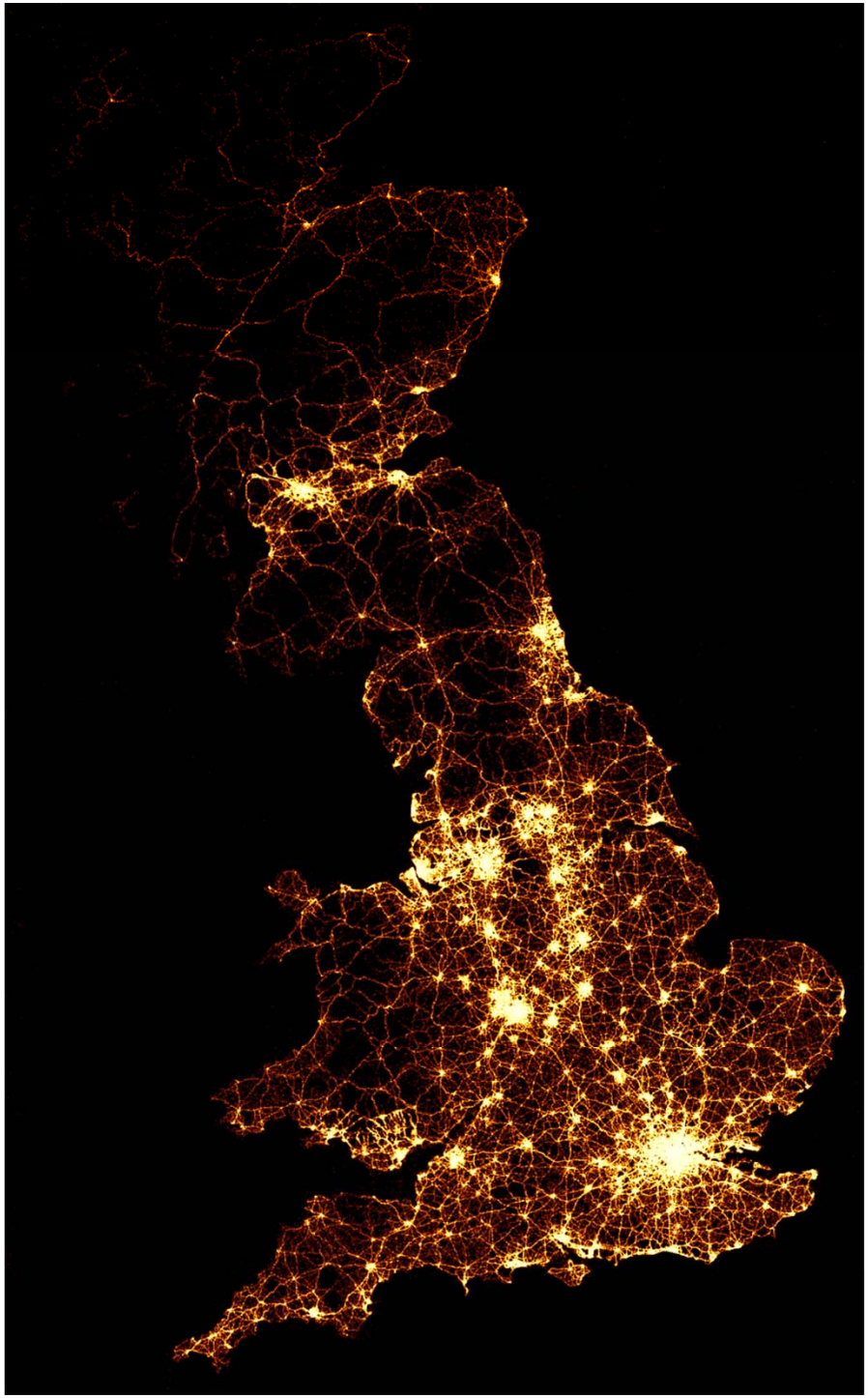


Google Map Overlay



Foursquares Check-ins



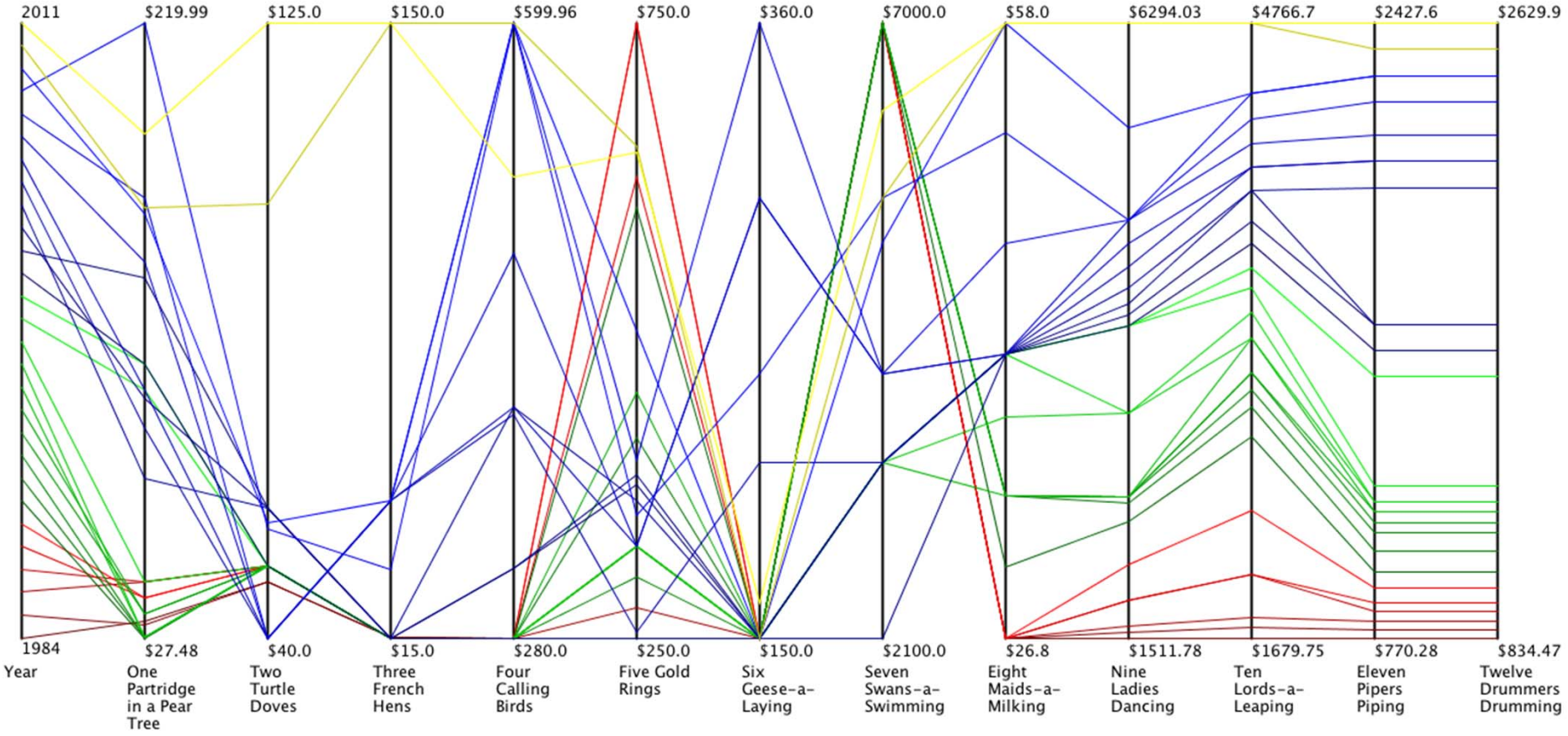


Parallel Coordinates

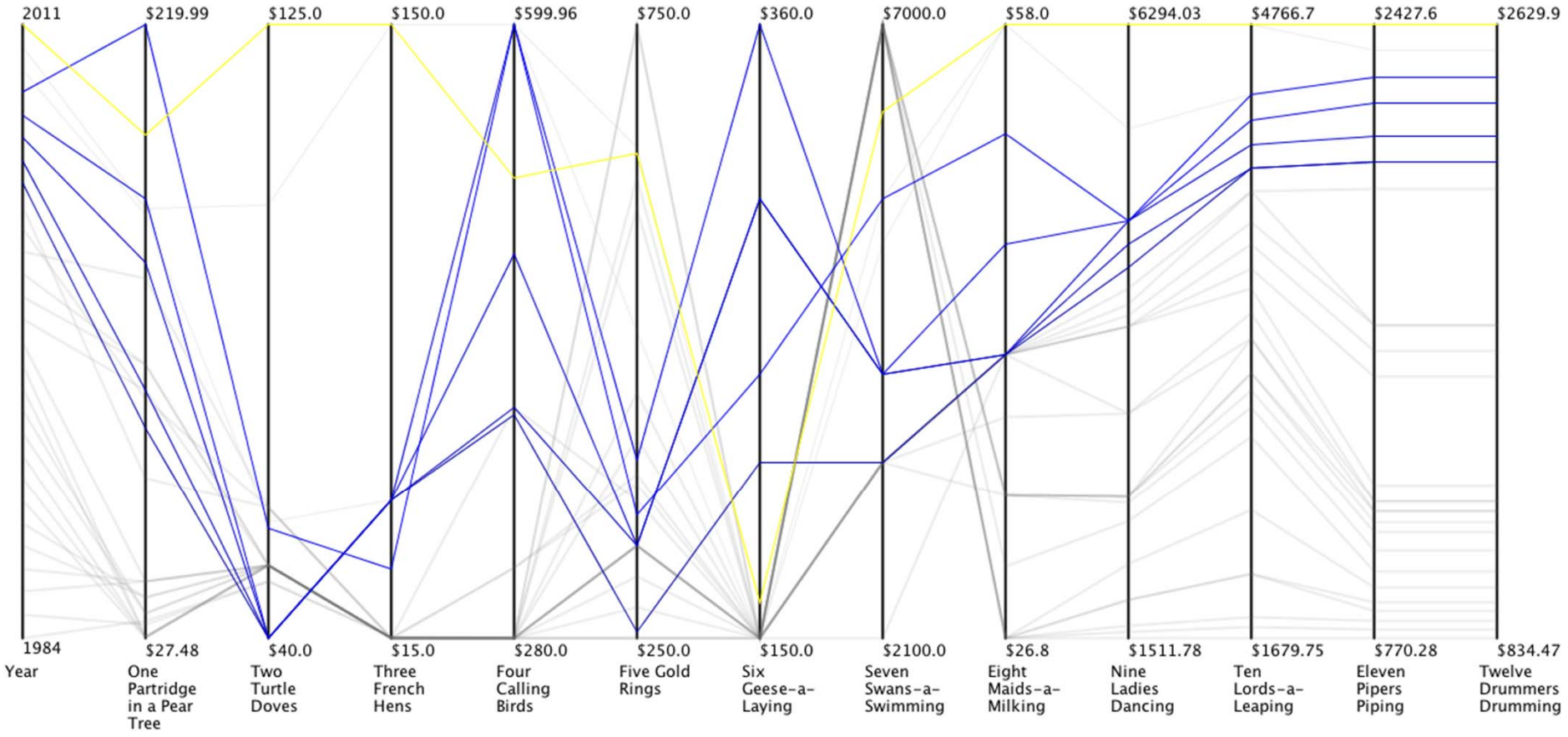
- High dimensional numerical data
- PNC Christmas Price Index

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	One Partridge in a Pear Tree	Two Turtle Doves	Three French Hens	Four Calling Birds	Five Gold Rings	Six Geese-a-Laying	Seven Swans-a-Swimming	Eight Maids-a-Milking	Nine Ladies Dancing	Ten Lords-a-Leaping	Eleven Pipers Piping	Twelve Drummers Drumming
2	1984	32.52	47.73	15.23	280	325	150	7000	26.8	1511.78	1679.75	770.28	834.47
3	1985	31.52	47.73	15.23	280	325	150	7000	26.8	1557	1730	793.32	859.43
4	1986	44.95	50	15	280	275	150	7000	26.8	1602	1780	814.32	882.18
5	1987	44.95	50	15	280	750	150	7000	26.8	1800	2000	842.4	912.6
6	1988	39.95	50	15	280	625	150	7000	26.8	1800	2000	863.46	935.42
7	1989	39.95	50	15	280	750	150	7000	26.8	2084.4	2316	905.58	981.05
8	1990	27.48	50	15	280	600	150	7000	30.4	2417.9	2686.56	947.7	1026.68
9	1991	27.5	50	15	280	412.5	150	7000	34	2555.72	2839.69	1003.86	1087.52
10	1992	27.5	50	15	280	300	150	7000	34	2606.83	2924.88	1053	1140.75
11	1993	34.99	50	15	280	325	150	7000	34	2606.83	3012.63	1081.08	1171.17

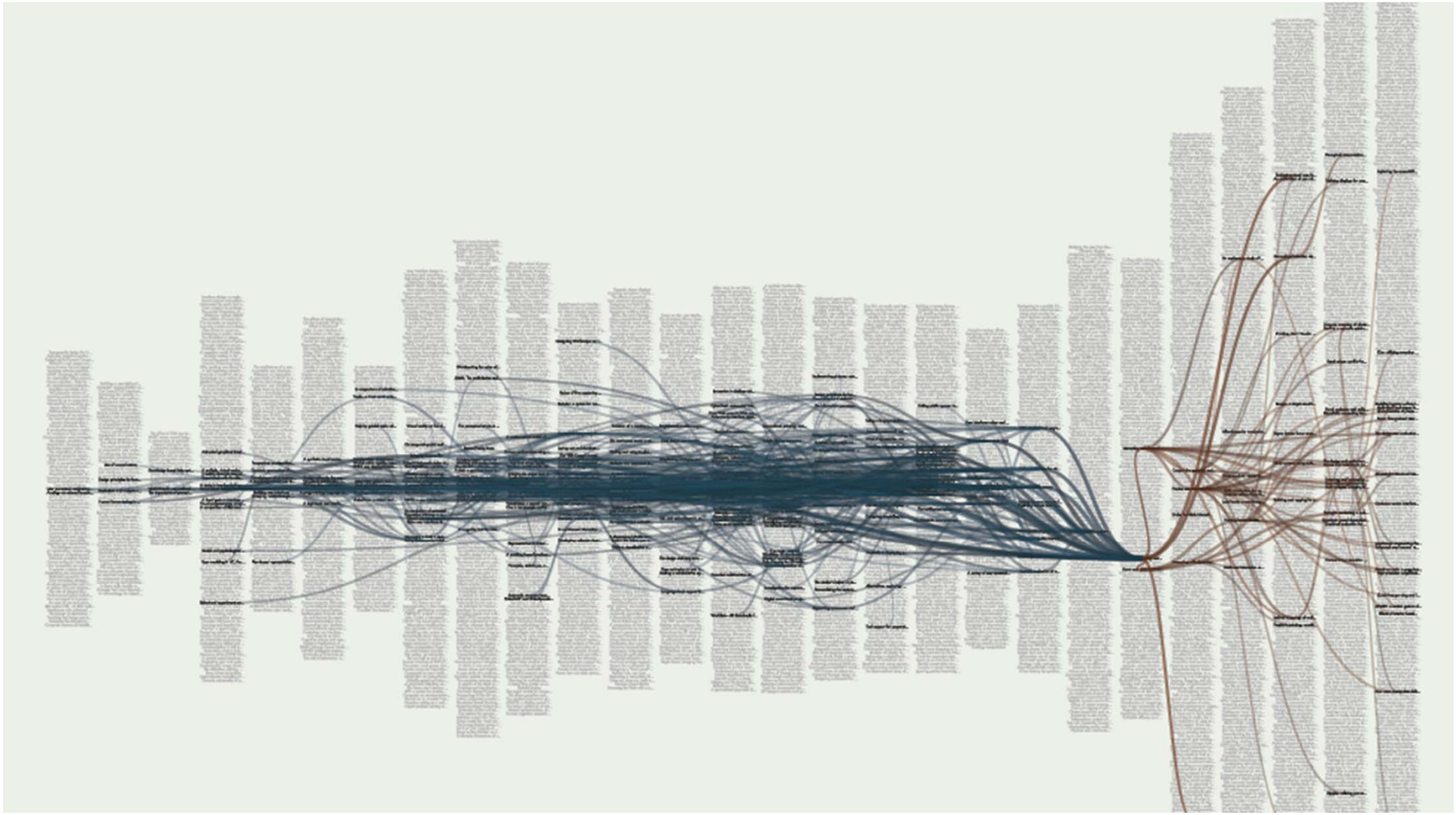
Parallel Coordinates



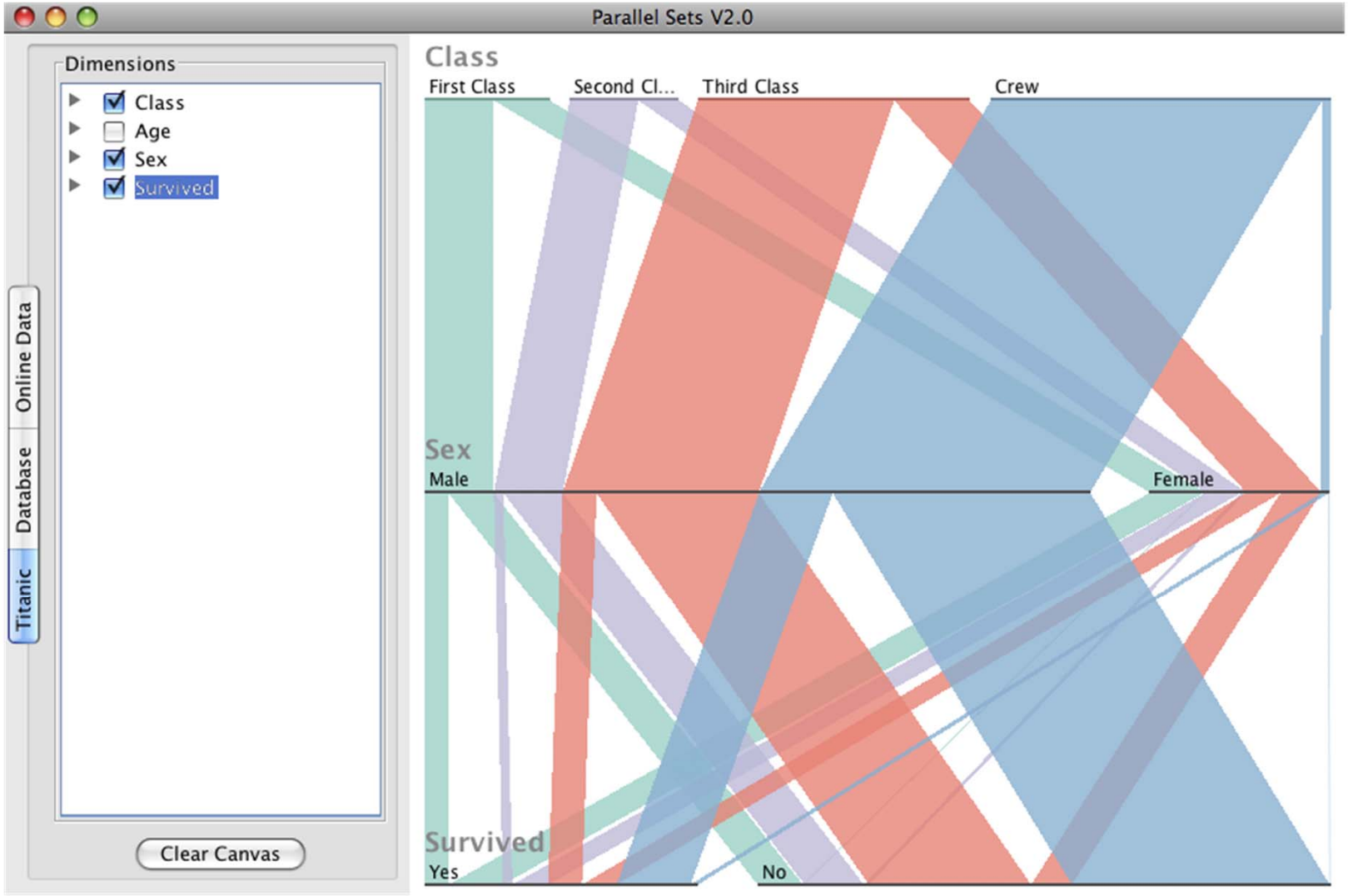
Brushing


2011
$219.99
$125.0
$150.0
$599.96
$750.0
$360.0
$7000.0
$58.0
$6294.03
$4766.7
$2427.6
$2629.9
1984
$27.48
$40.0
$15.0
$280.0
$250.0
$150.0
$2100.0
$26.8
$1511.78
$1679.75
$770.28
$834.47
Year
One Partridge in a Pear Tree
Two Turtle Doves
Three French Hens
Four Calling Birds
Five Gold Rings
Six Geese-a-Laying
Seven Swans-a-Swimming
Eight Maids-a-Milking
Nine Ladies Dancing
Ten Lords-a-Leaping
Eleven Pipers Piping
Twelve Drummers Drumming

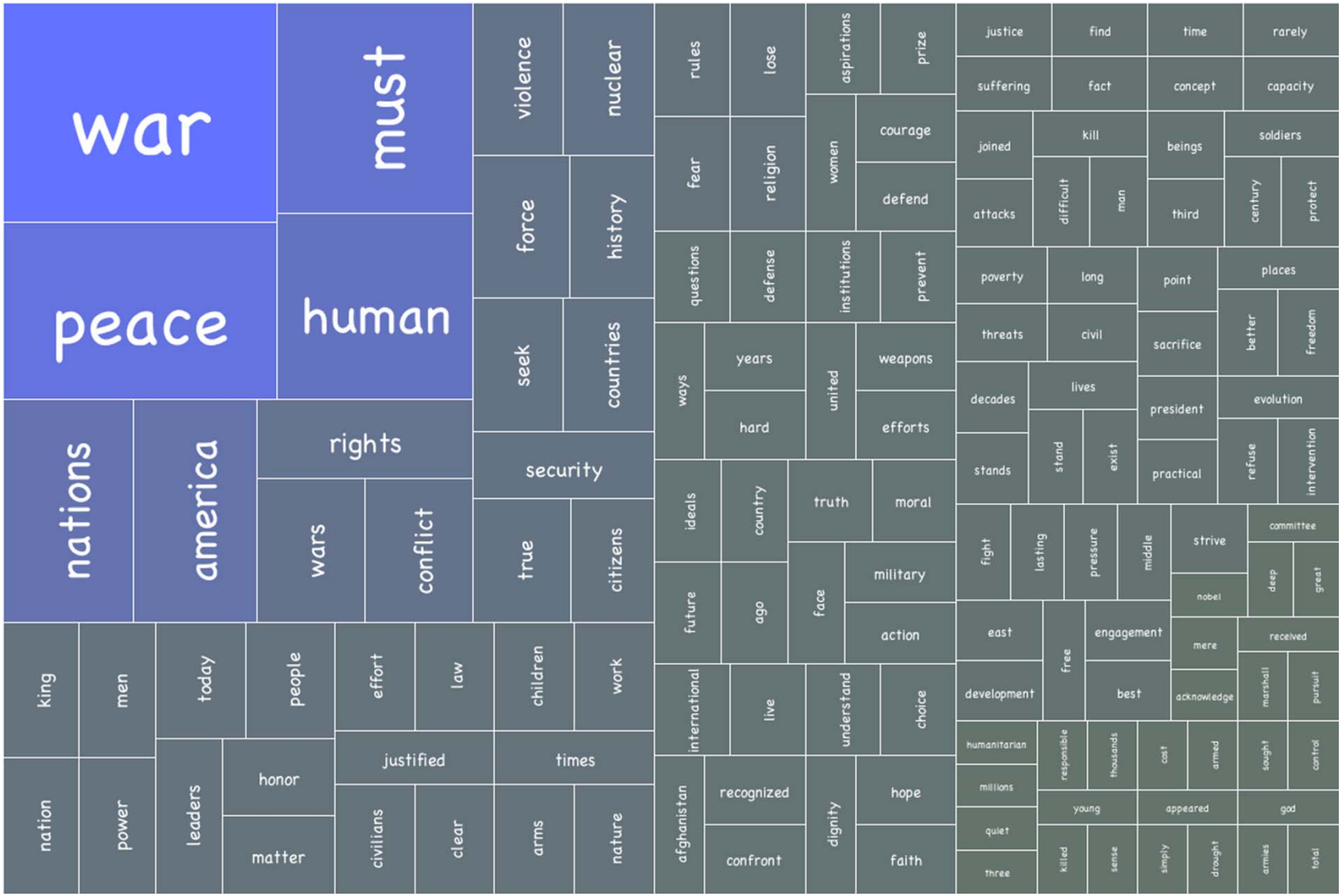
Citeology



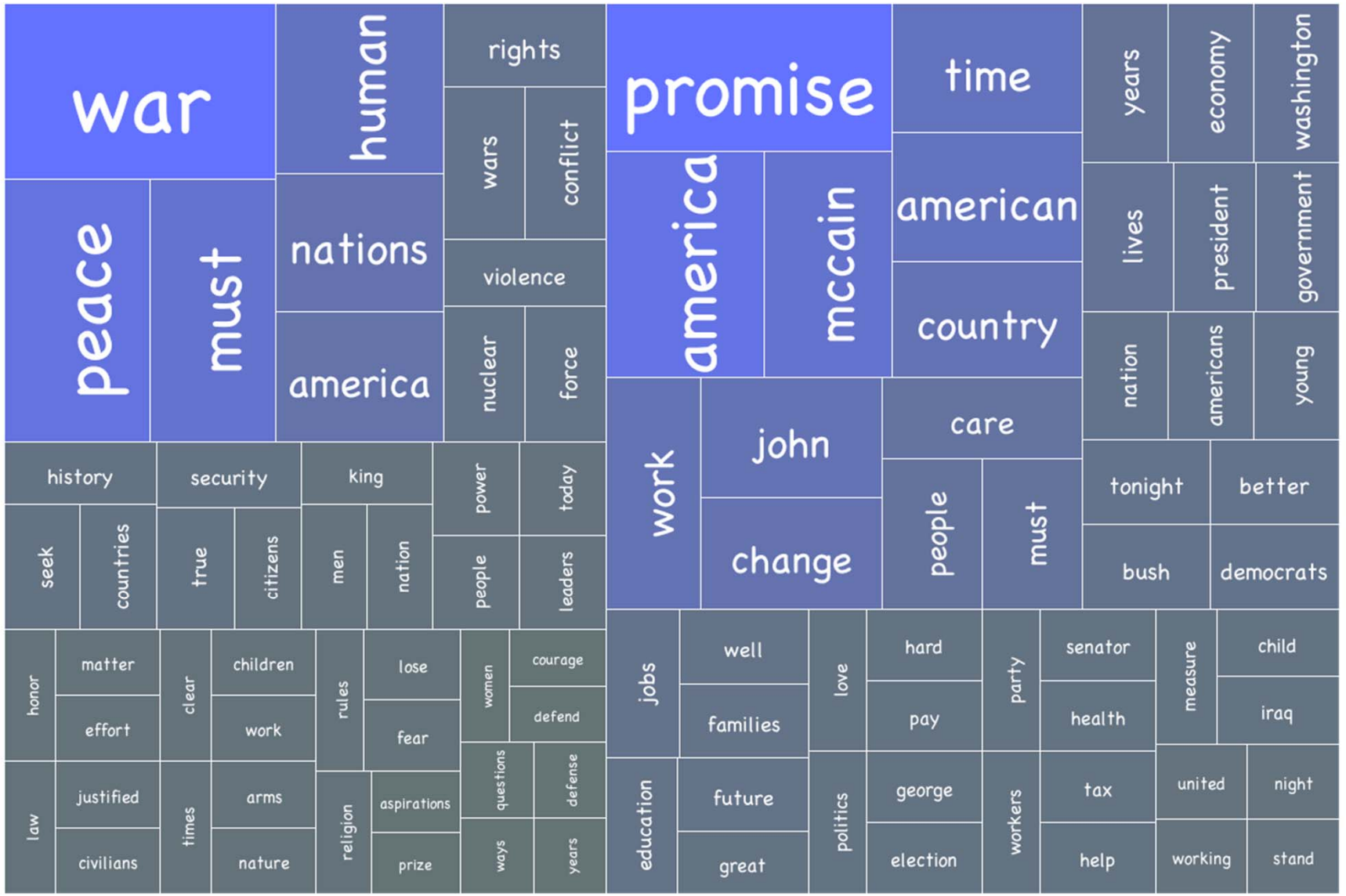
Parallel Sets



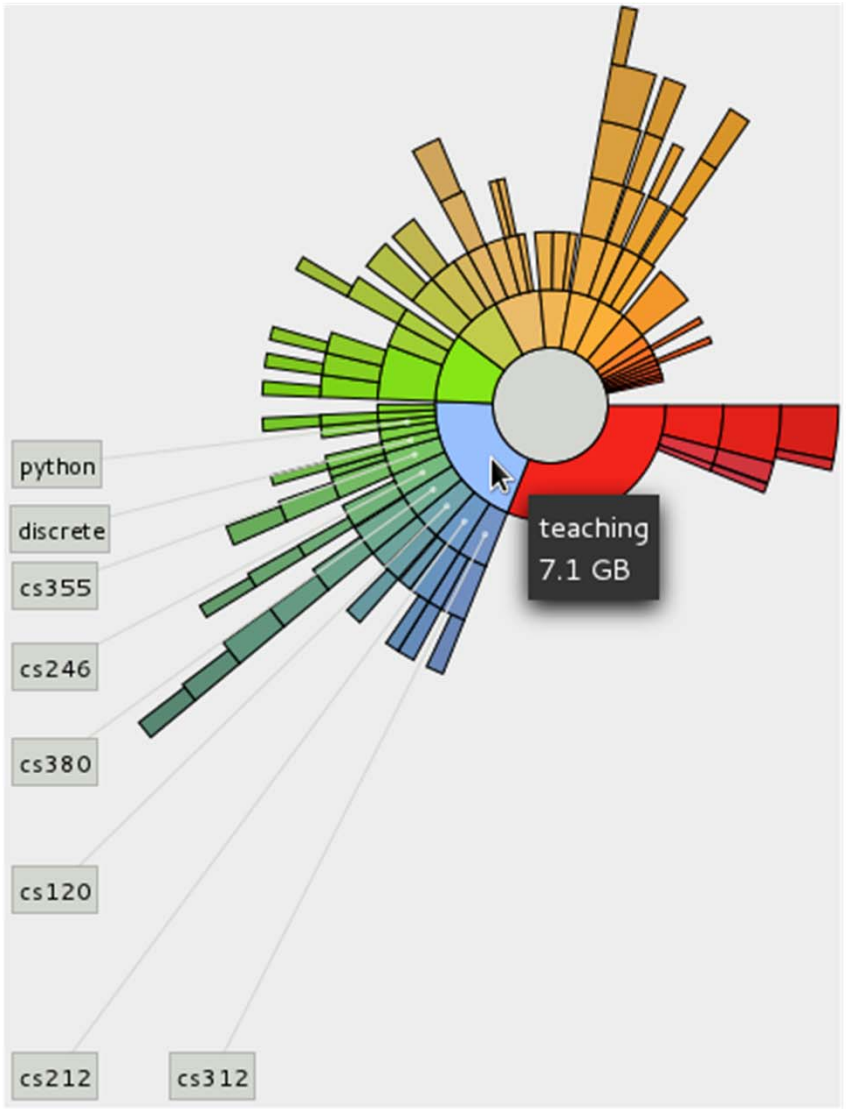
TreeMap


war
must
peace
human
violence
nuclear
force
history
seek
countries
rules
lose
fear
religion
questions
defense
aspirations
prize
women
courage
defend
institutions
prevent
justice
find
time
rarely
suffering
fact
concept
capacity
joined
kill
beings
soldiers
attacks
difficult
man
third
century
protect
poverty
long
point
places
threats
civil
sacrifice
better
freedom
decades
lives
president
evolution
stands
stand
exist
practical
refuse
intervention
nations
america
rights
wars
conflict
security
true
citizens
ways
years
hard
united
weapons
efforts
ideals
country
truth
moral
future
ago
face
military
action
international
live
understand
choice
afghanistan
recognized
confront
dignity
hope
faith
king
men
today
people
effort
law
children
work
nation
power
leaders
honor
matter
justified
times
civilians
clear
arms
nature
fight
lasting
pressure
middle
strive
committee
nobel
deep
great
east
free
engagement
mere
received
development
best
acknowledge
marshall
pursuit
humanitarian
responsible
thousands
cast
armed
sought
control
millions
young
appeared
god
quiet
killed
sense
simply
drought
armies
total
three

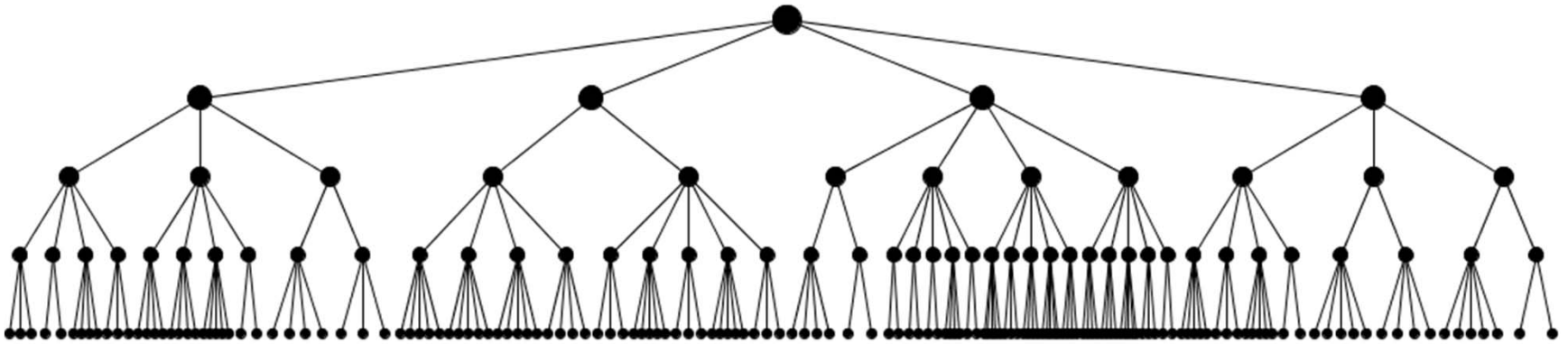
More Levels


war
peace
must
human
nations
america
rights
wars
conflict
violence
nuclear
force
promise
america
mccain
work
john
change
time
american
country
care
people
must
years
economy
washington
lives
president
government
nation
americans
young
tonight
better
bush
democrats
history
seek
countries
security
true
citizens
king
men
nation
power
today
people
leaders
honor
matter
effort
clear
children
work
rules
lose
fear
women
courage
defend
law
justified
civilians
times
arms
nature
religion
aspirations
prize
questions
defense
ways
years
jobs
well
families
future
great
love
hard
pay
george
election
party
senator
health
tax
help
measure
child
iraq
united
night
working
stand
education
politics
workers

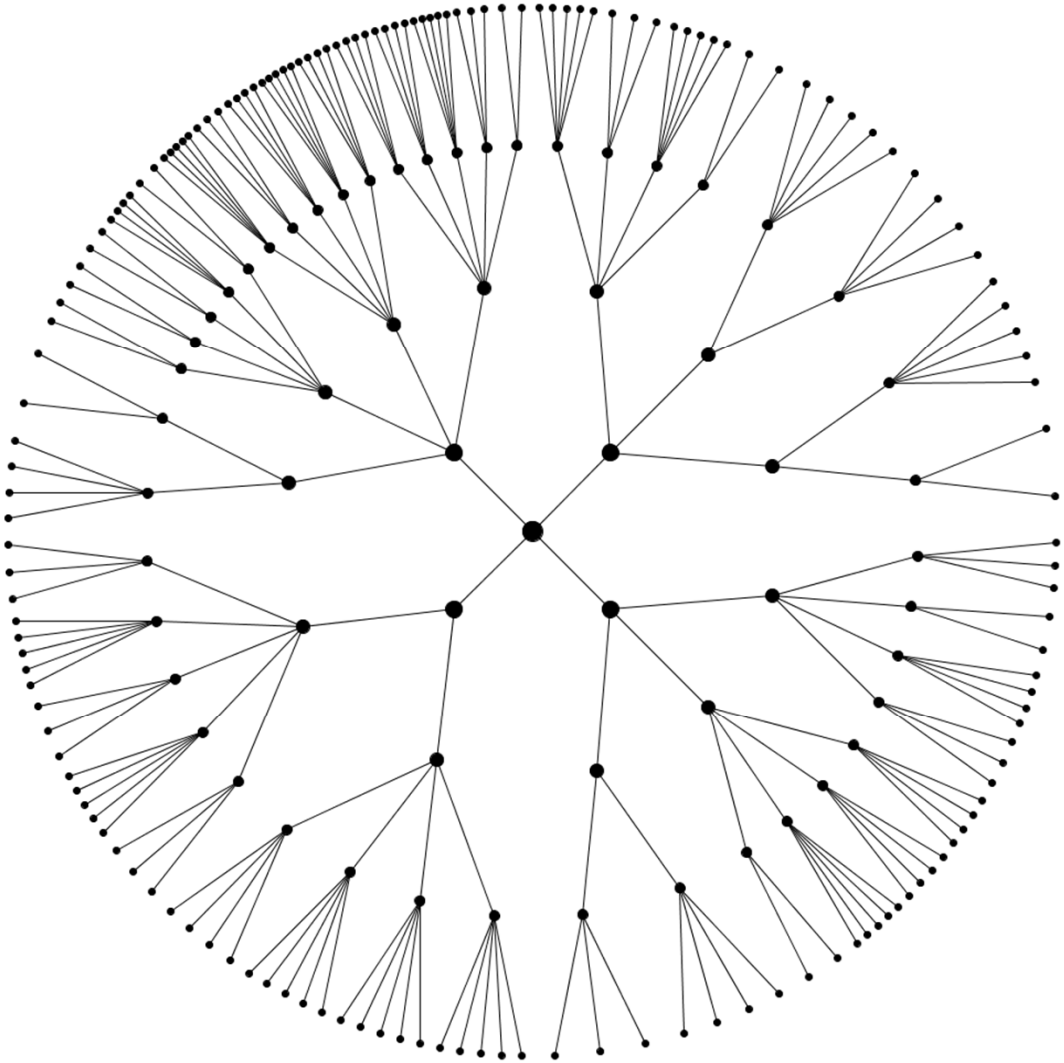
Radial


python
discrete
cs355
cs246
cs380
cs120
cs212
cs312
teaching
7.1 GB

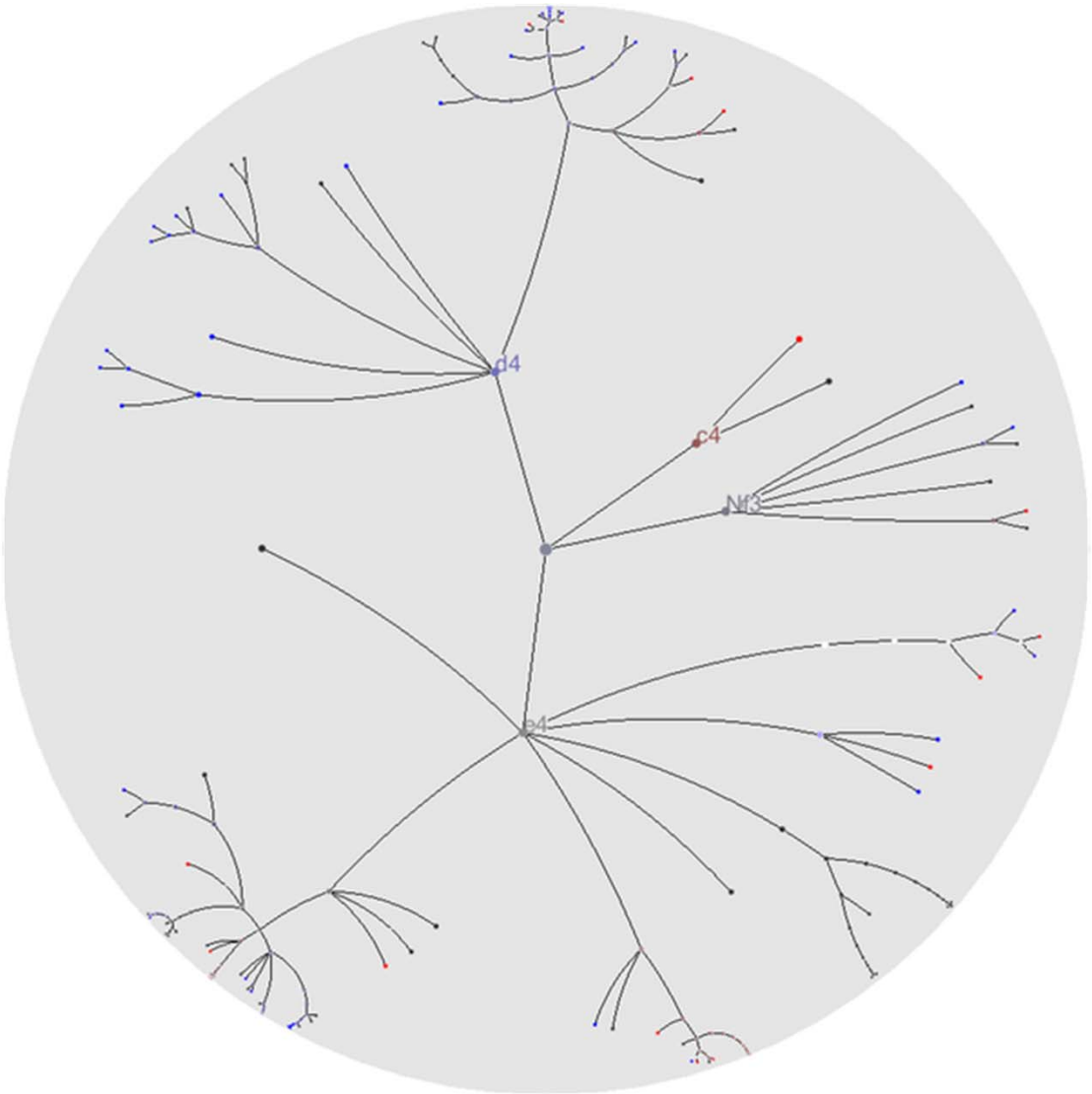
Tree



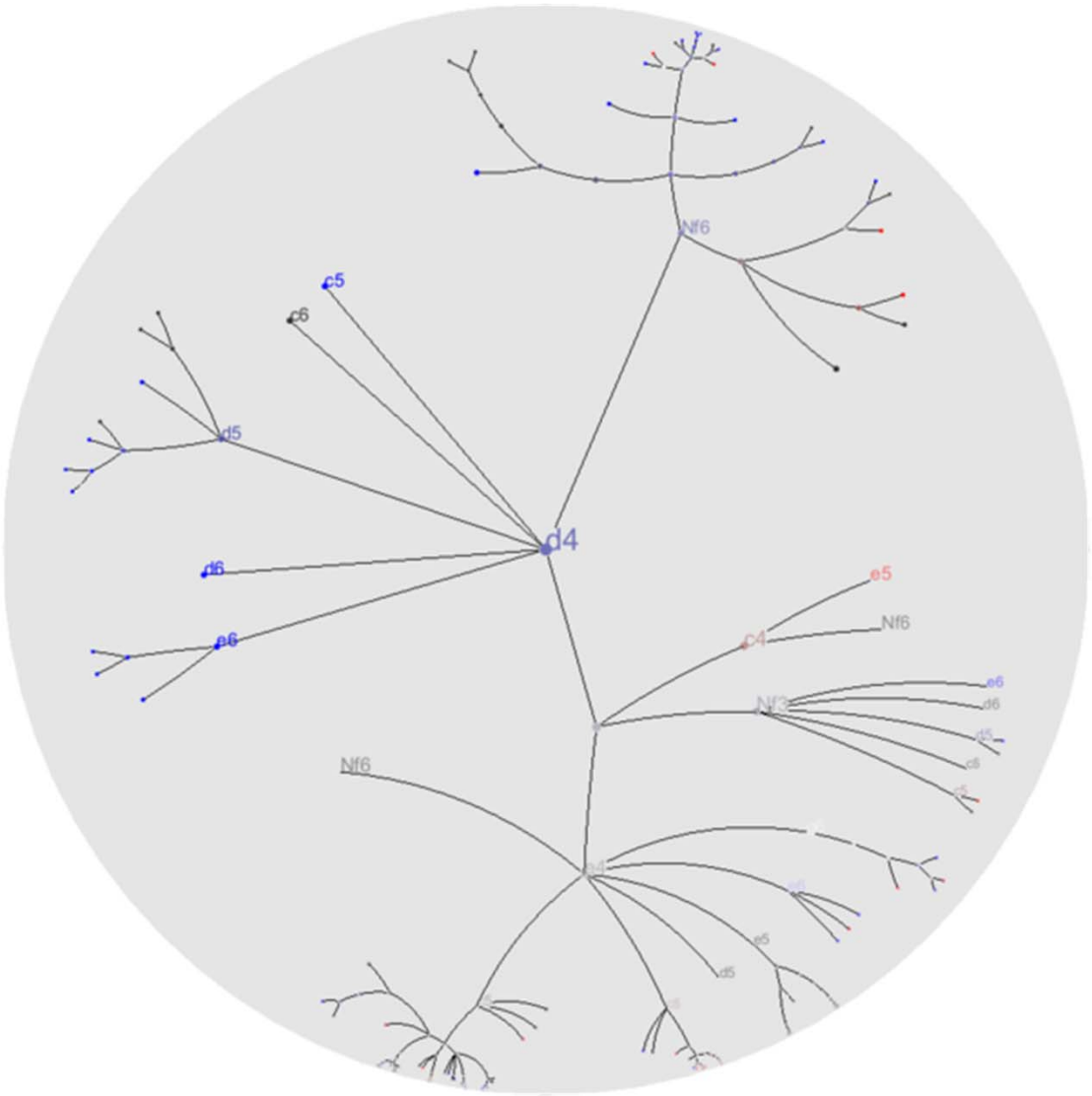
Radial Tree



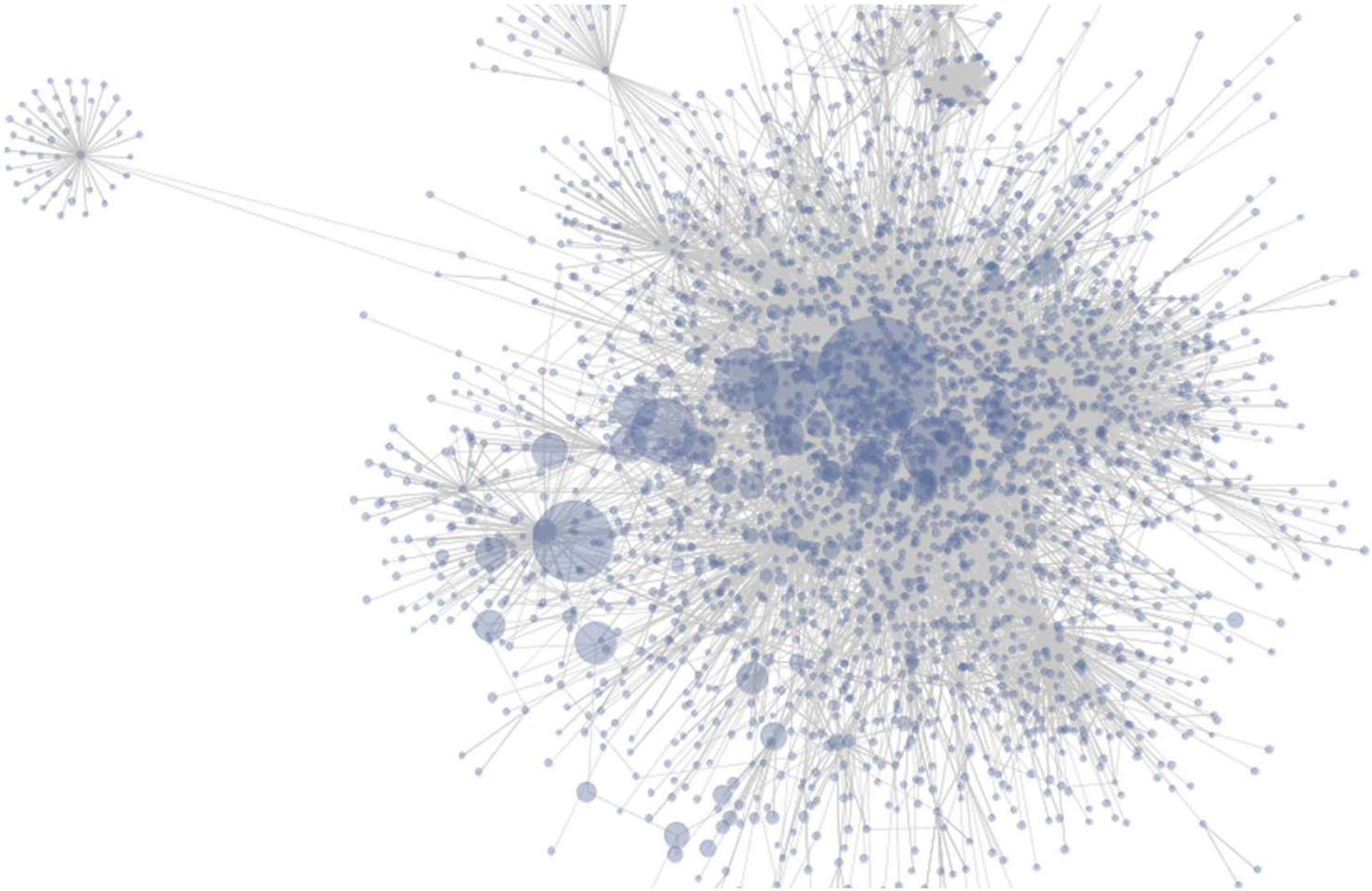
Hyperbolic Tree



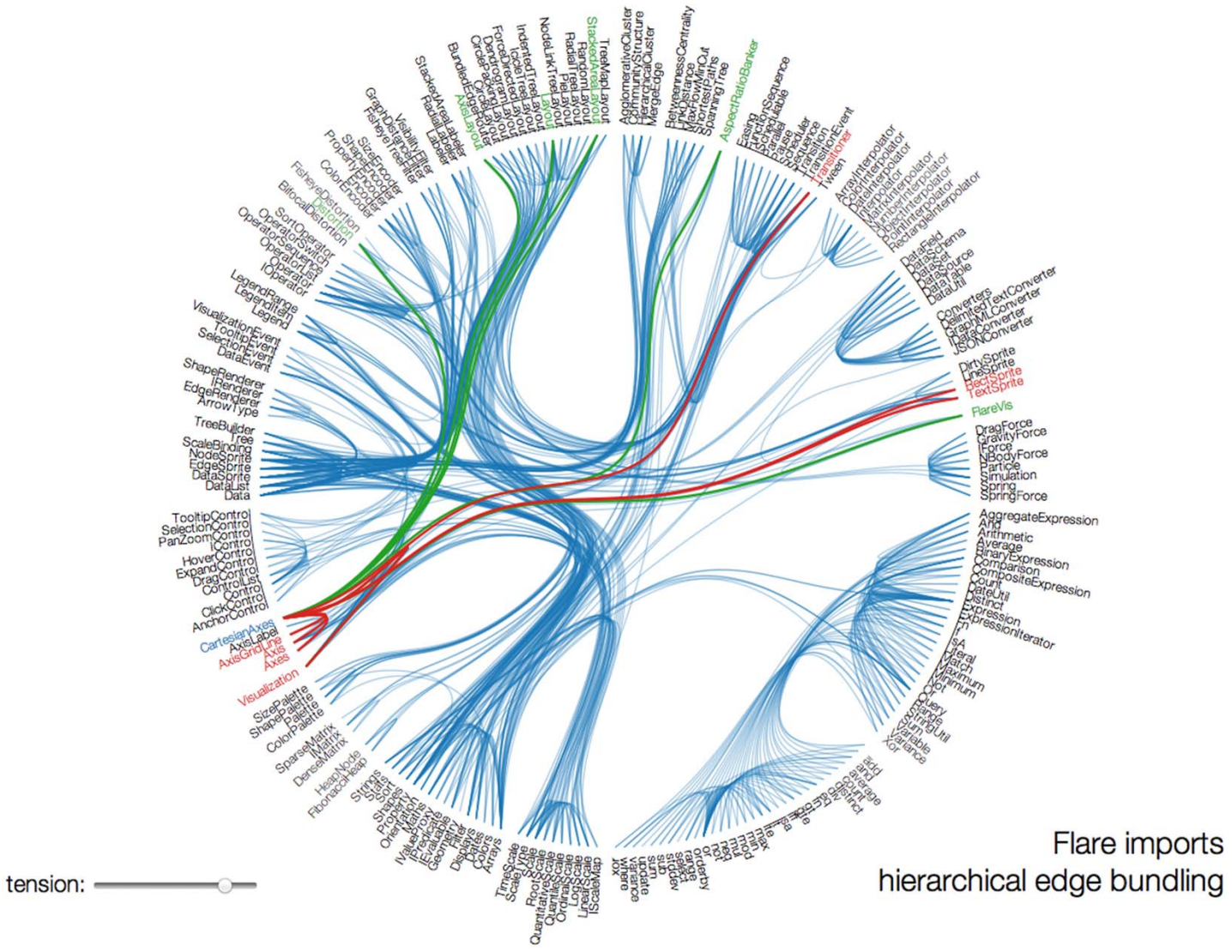
Focus+Context



Network/Graph-based



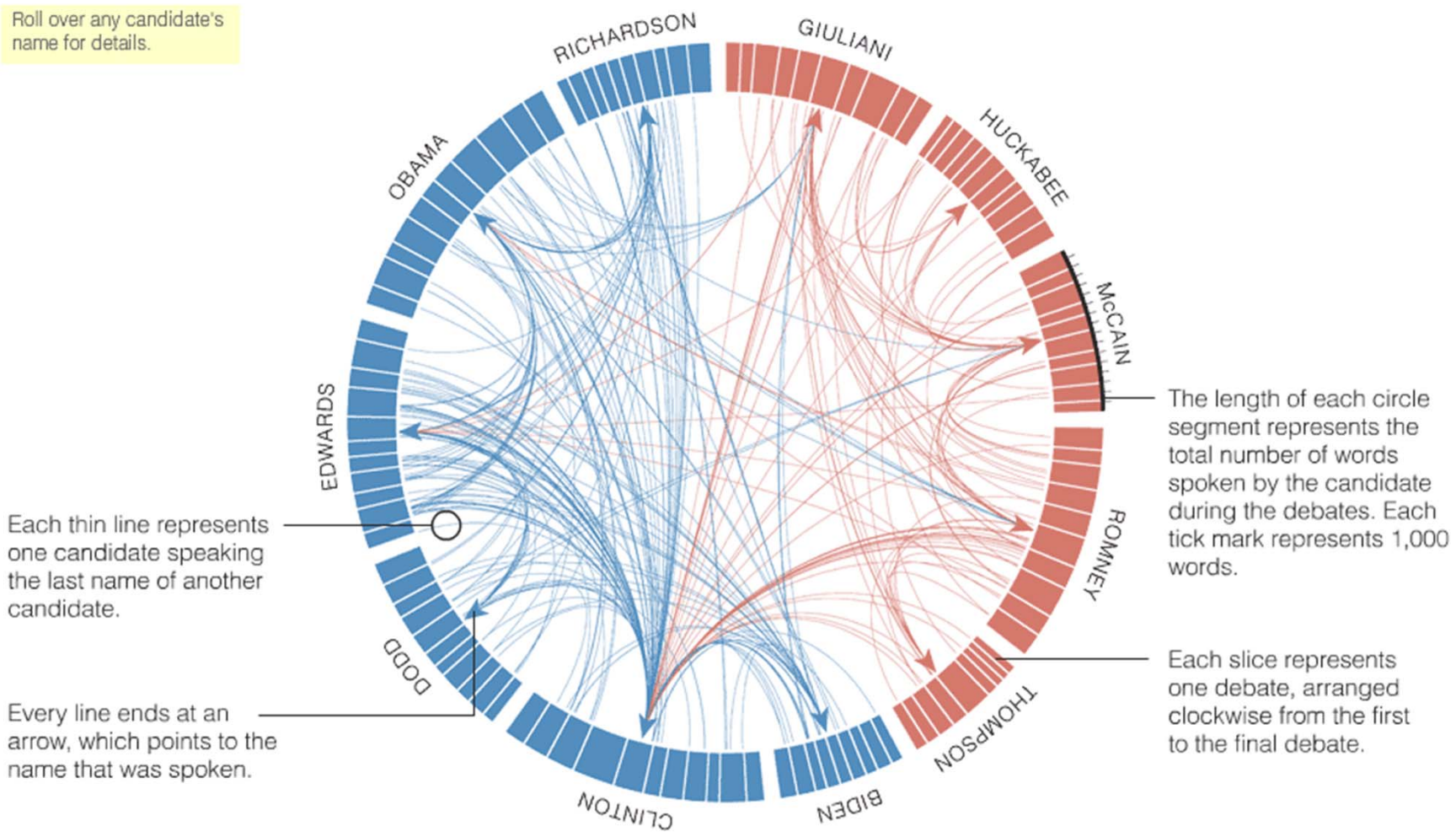
Chord Diagrams



Circular Layouts

- Enjoyed great popularity lately
- We are drawn to circles
  - symmetry
  - aesthetics
  - perceived novelty
- Not always justified by increased readability
- Human visual perception supports only rough comparison of areas and angles

Naming Names



Data Analysis

- Unstructured data
- A lot of it

- Dimension reduction
- Cluster analysis
- Statistical methods
- Network models and optimization
- Graph theory