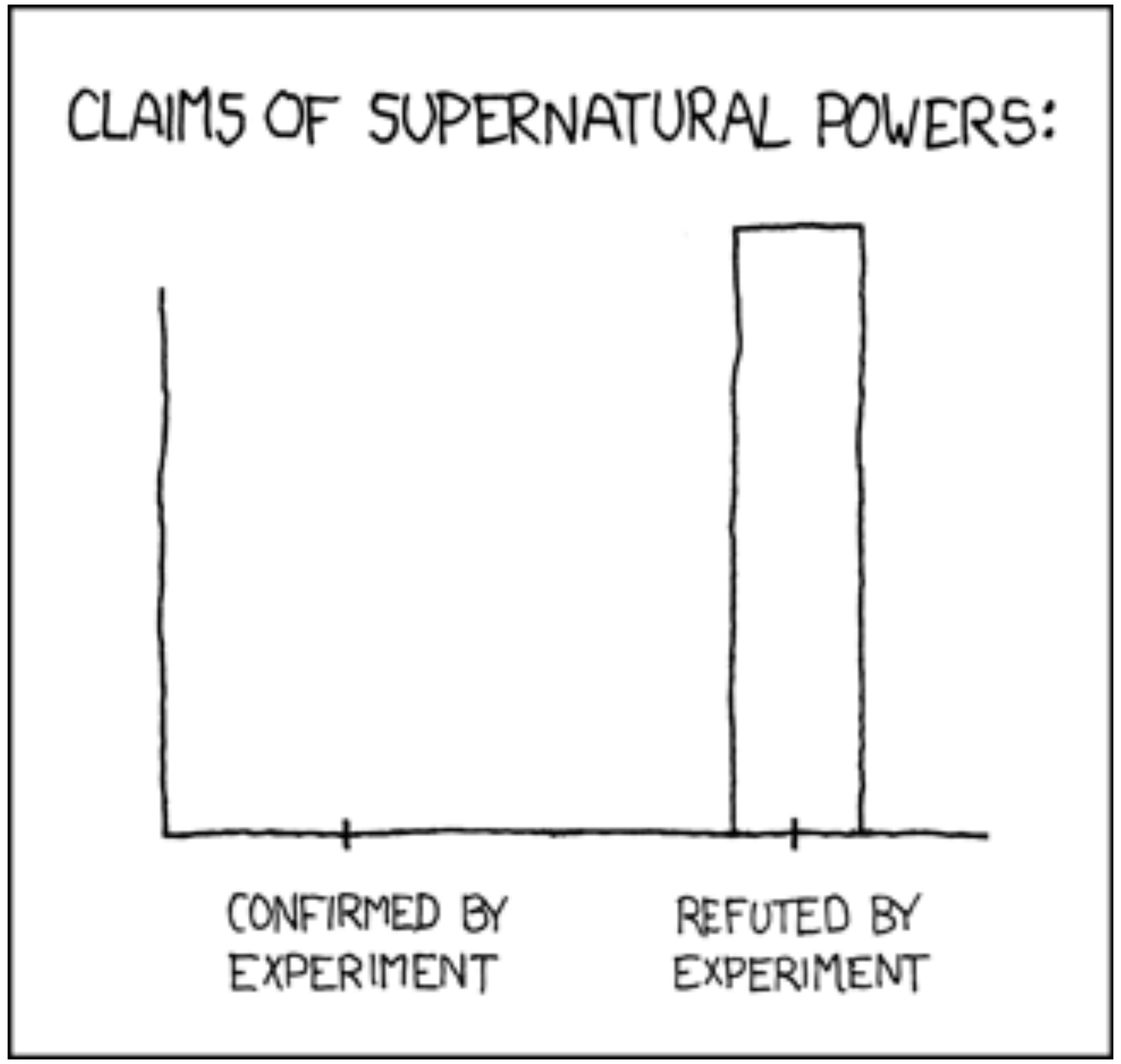
CS 171: Visualization
Data Abstraction &
Data Types

Alexander Lex
alex@seas.harvard.edu



[xkcd]

HARVARD
School of Engineering
and Applied Sciences

This Week

Homework 0:

- due tomorrow!
- **NEW: ANNOUNCE REPOSITORY**
- **& tell us if you don’t have a micro account yet**
- **http://goo.gl/HFVE6h**

Readings:

- D3: Chapters 5-8
- VAD: Chapter 2

Next Week

Lecture 4: The visualization alphabet. Visual Variables. Basic Tasks and Charts.

Introduction to Homework 2

Lecture 5: SKILLS: Sketching and Prototyping I

Reading: D3, Chapters 9-11; VAD, Chapter 3

HW1 Due!

HW 1

Questions?

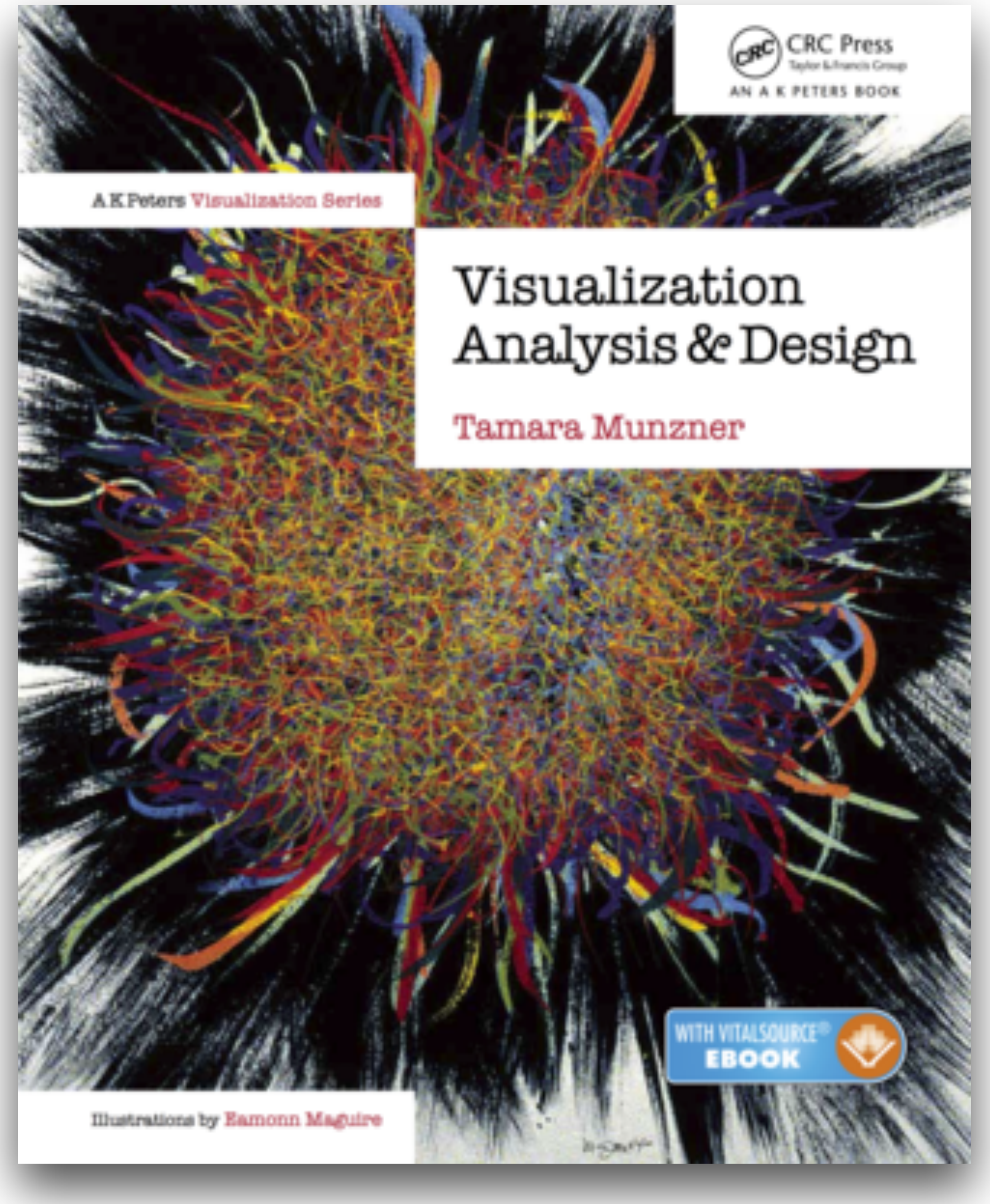
Write clean and general code!

Ask yourself: What would a user expect?

Organizational

Textbook on reserve in
Gordon McKay Library

Image credits, sources & more info on material: see hyperlinks



No Device Policy

- No Computers, Tablets, Phones in lecture hall
    - except when used for exercises
- Switch off, mute, flight mode
- Why?
    - It’s better to take notes by hand
    - Notifications are designed to grab your attention

Survey Results

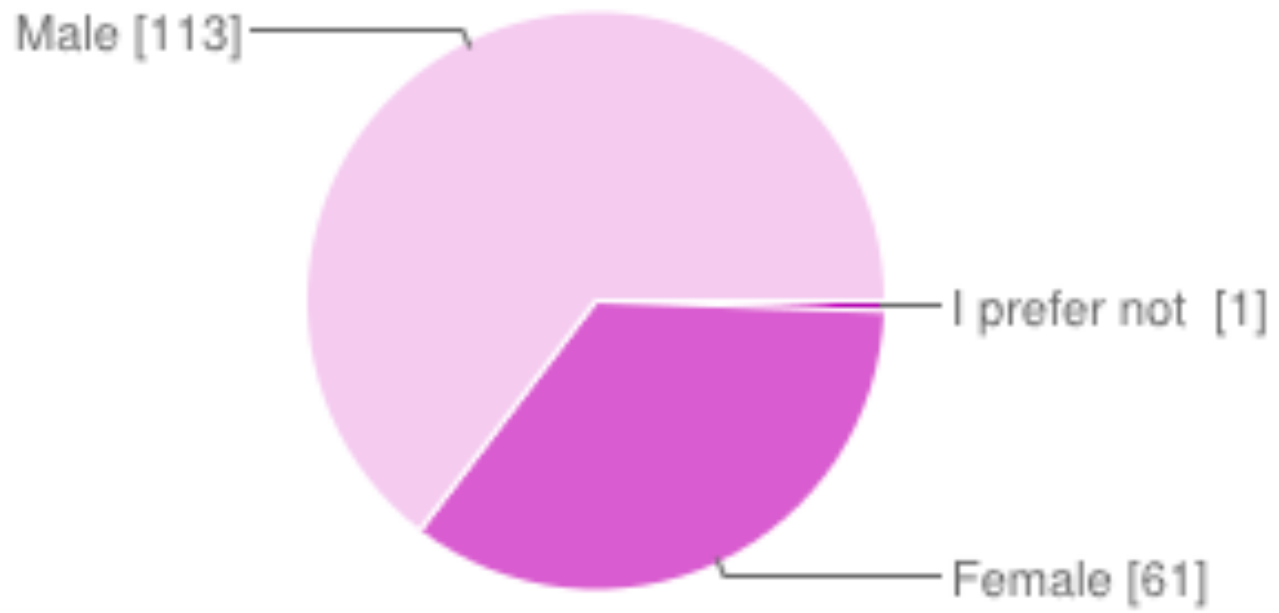
238 registered students (most ever)

+~40 relative to 2014
+~80 relative to 2013

125 College & other, 87 DCE
175 survey responses (Wednesday)

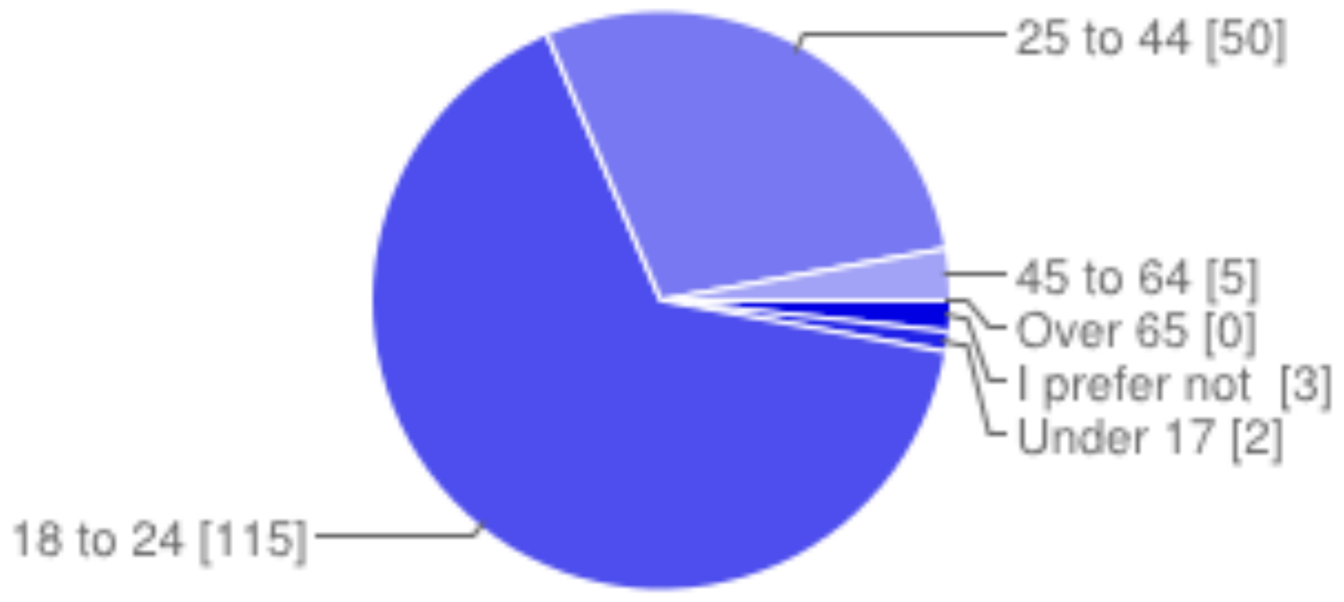
Demographics

Gender



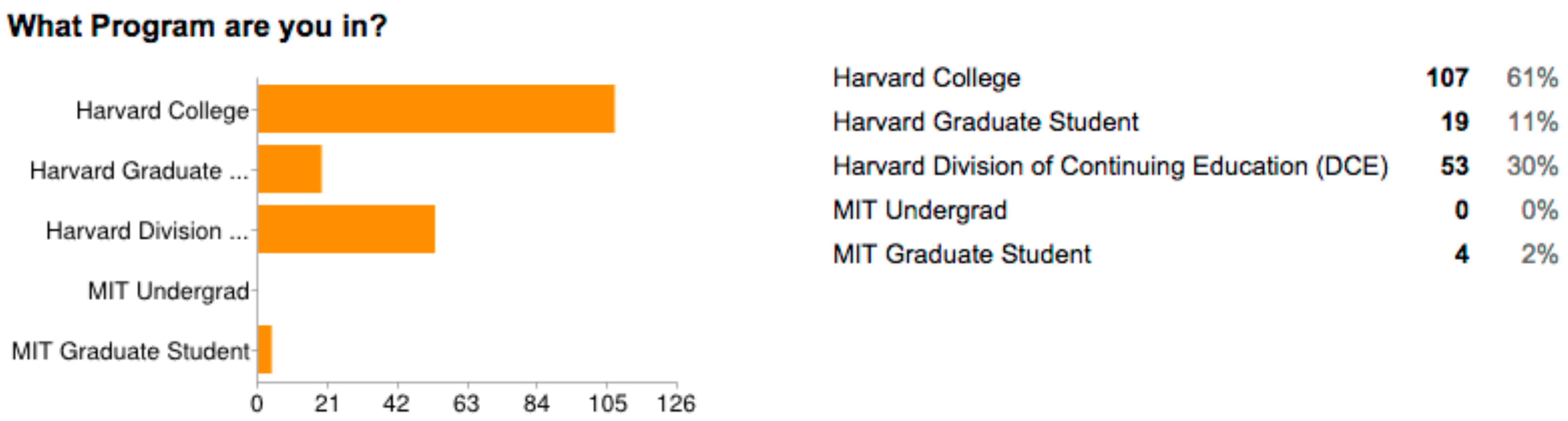
I prefer not to disclose	**1**	1%
Female	**61**	35%
Male	**113**	65%

Age



I prefer not to disclose	**3**	2%
Under 17	**2**	1%
18 to 24	**115**	66%
25 to 44	**50**	29%
45 to 64	**5**	3%
Over 65	**0**	0%

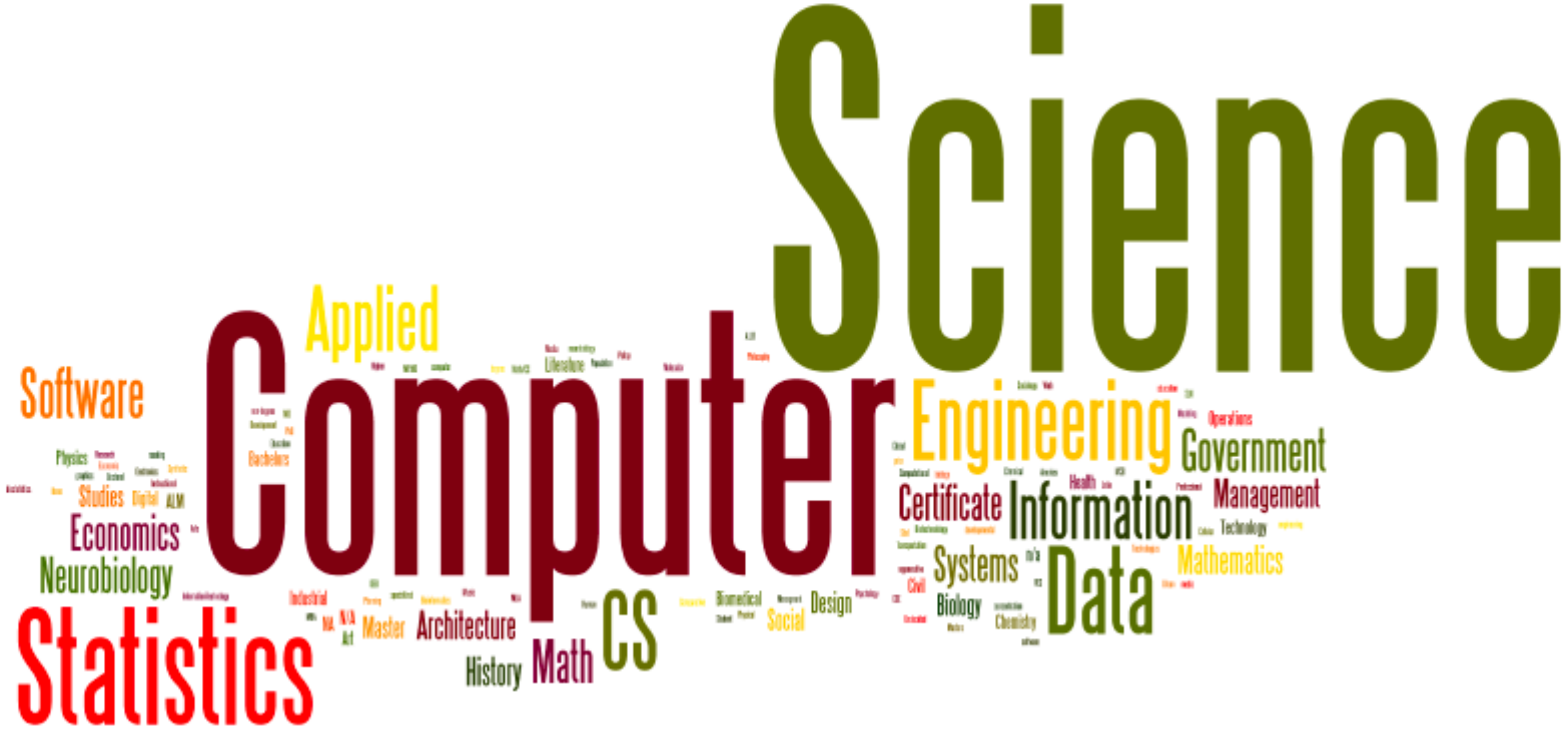
Program



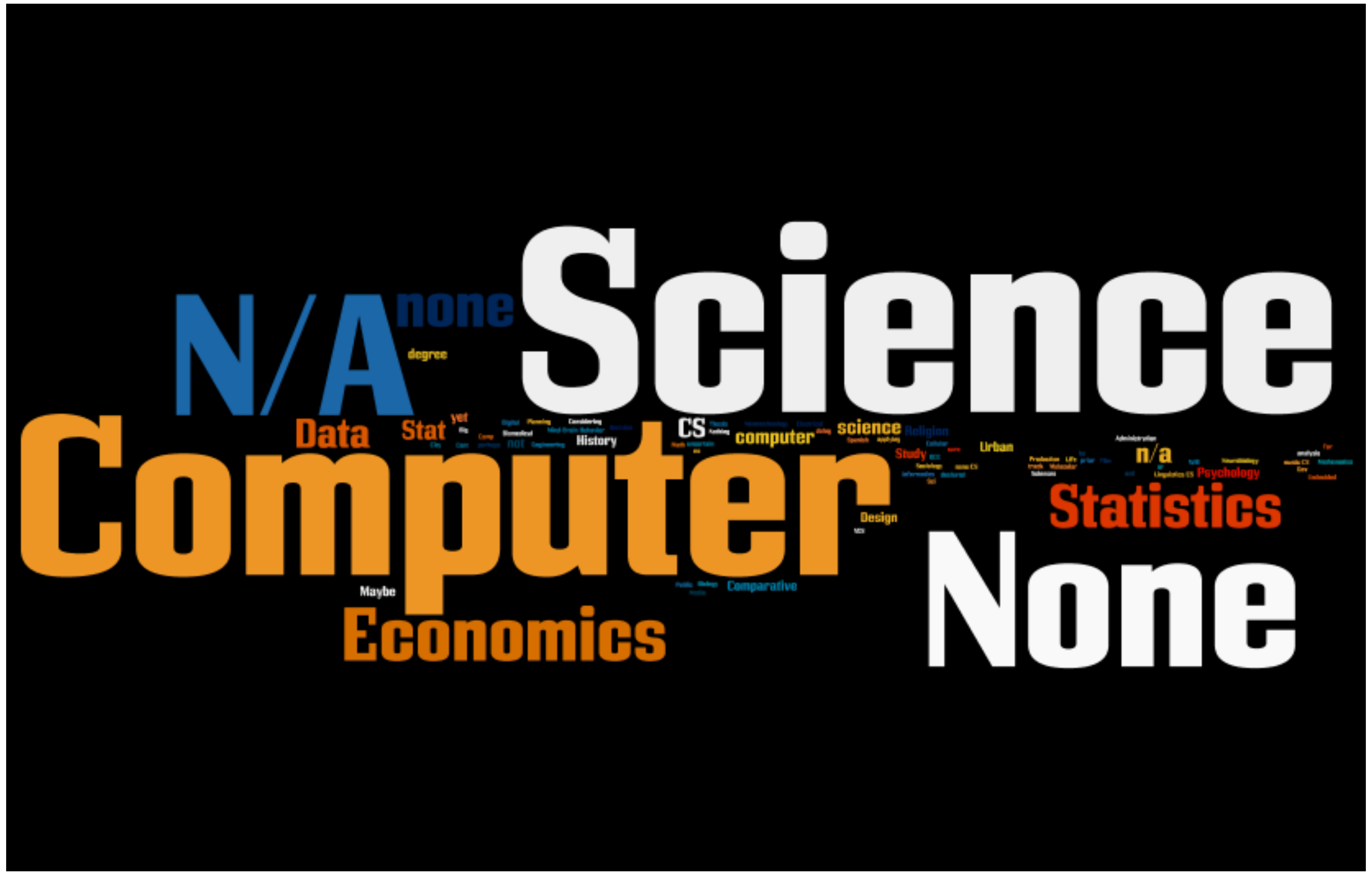
What Program are you in?

Harvard College	**107**	61%
Harvard Graduate Student	**19**	11%
Harvard Division of Continuing Education (DCE)	**53**	30%
MIT Undergrad	**0**	0%
MIT Graduate Student	**4**	2%

Concentrations

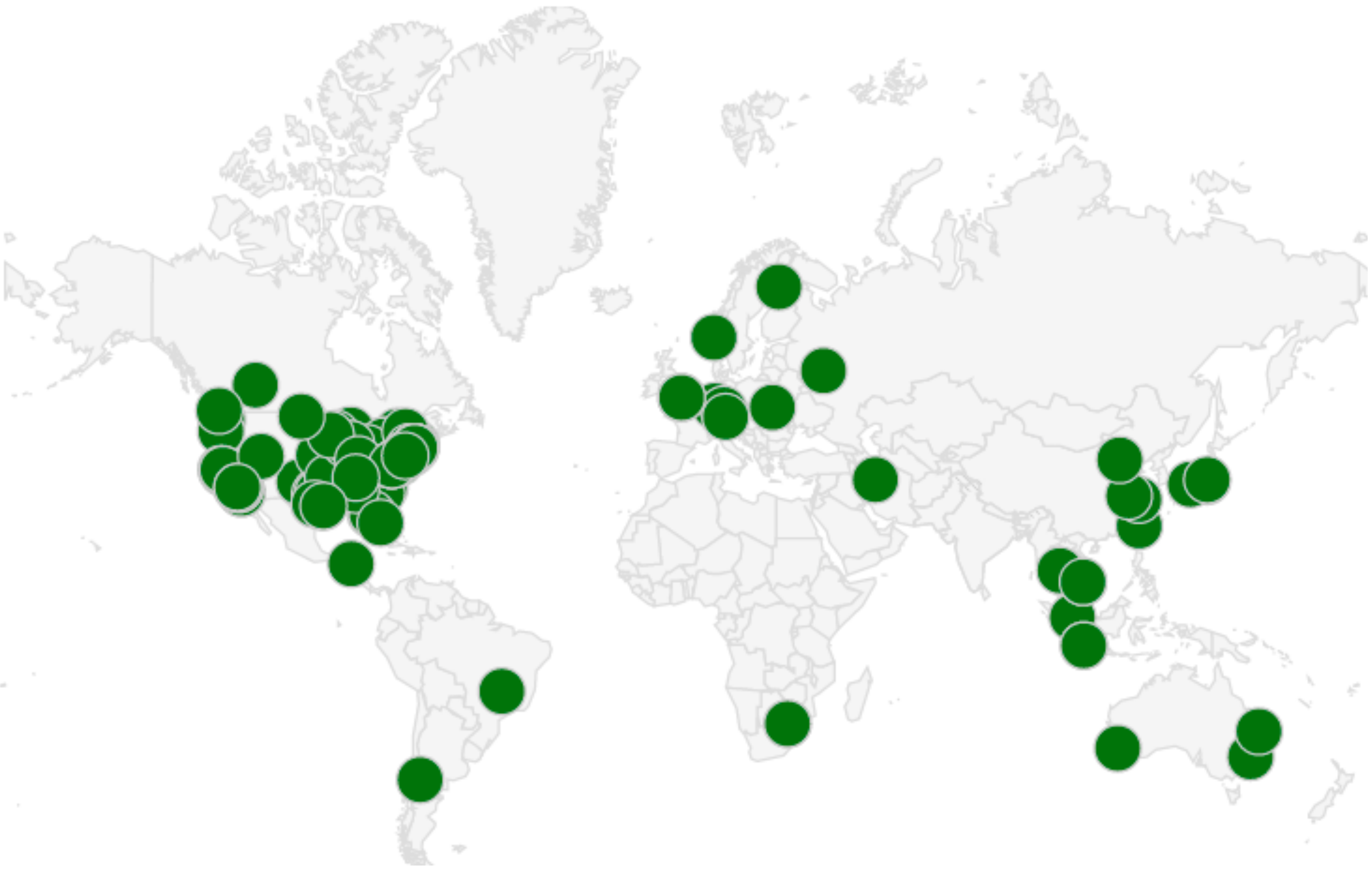


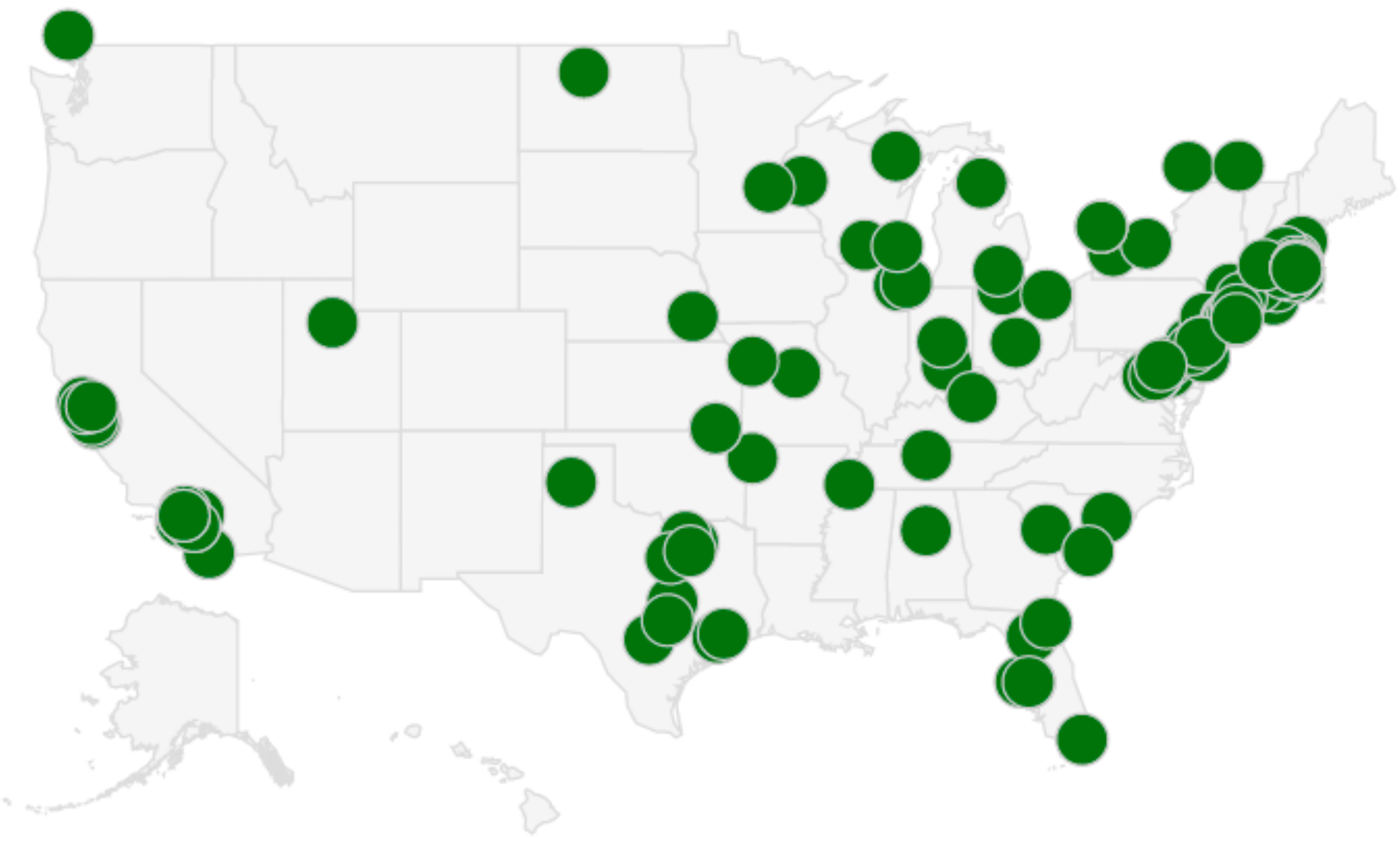
Primary



Secondary

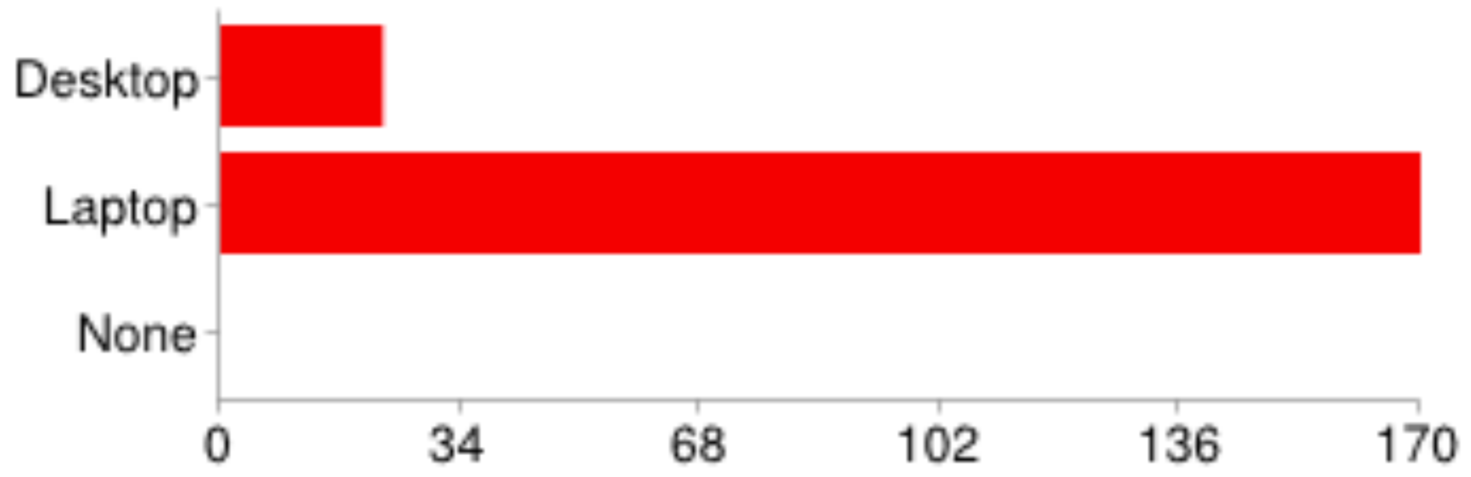
Where you’re from





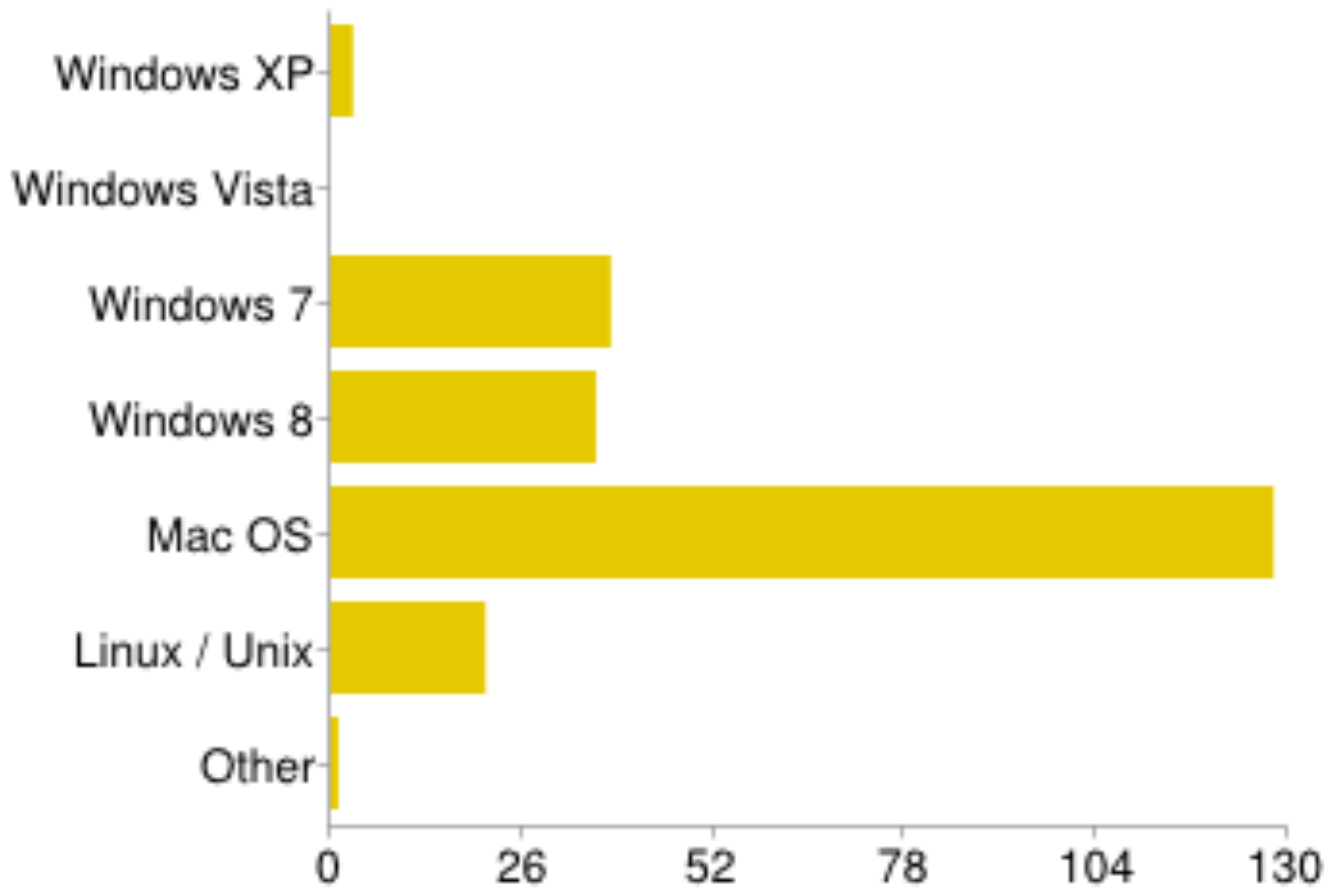
Computer / OS

What kind(s) of computer(s) do you own?



Desktop	**23**	13%
Laptop	**170**	97%
None	**0**	0%

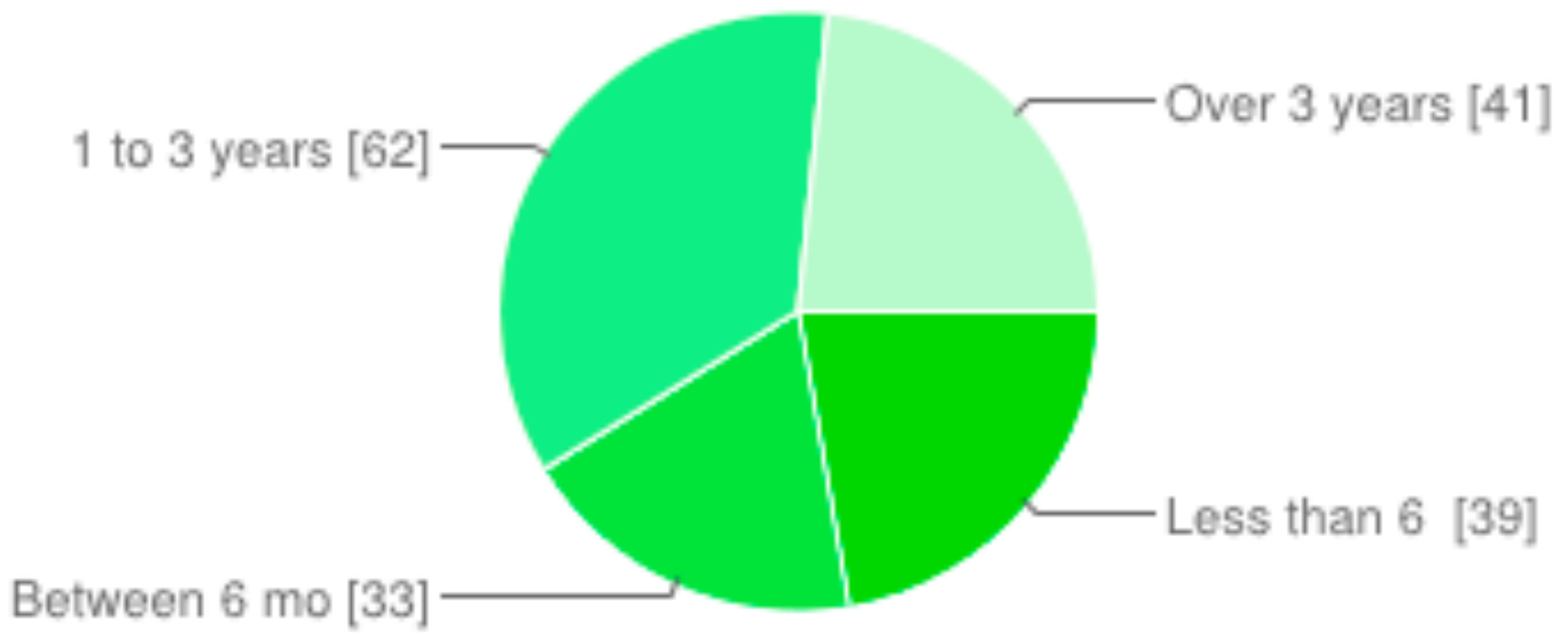
What operating system(s) do you run on your computer(s)?



Windows XP	**3**	2%
Windows Vista	**0**	0%
Windows 7	**38**	22%
Windows 8	**36**	21%
Mac OS	**128**	73%
Linux / Unix	**21**	12%
Other	**1**	1%

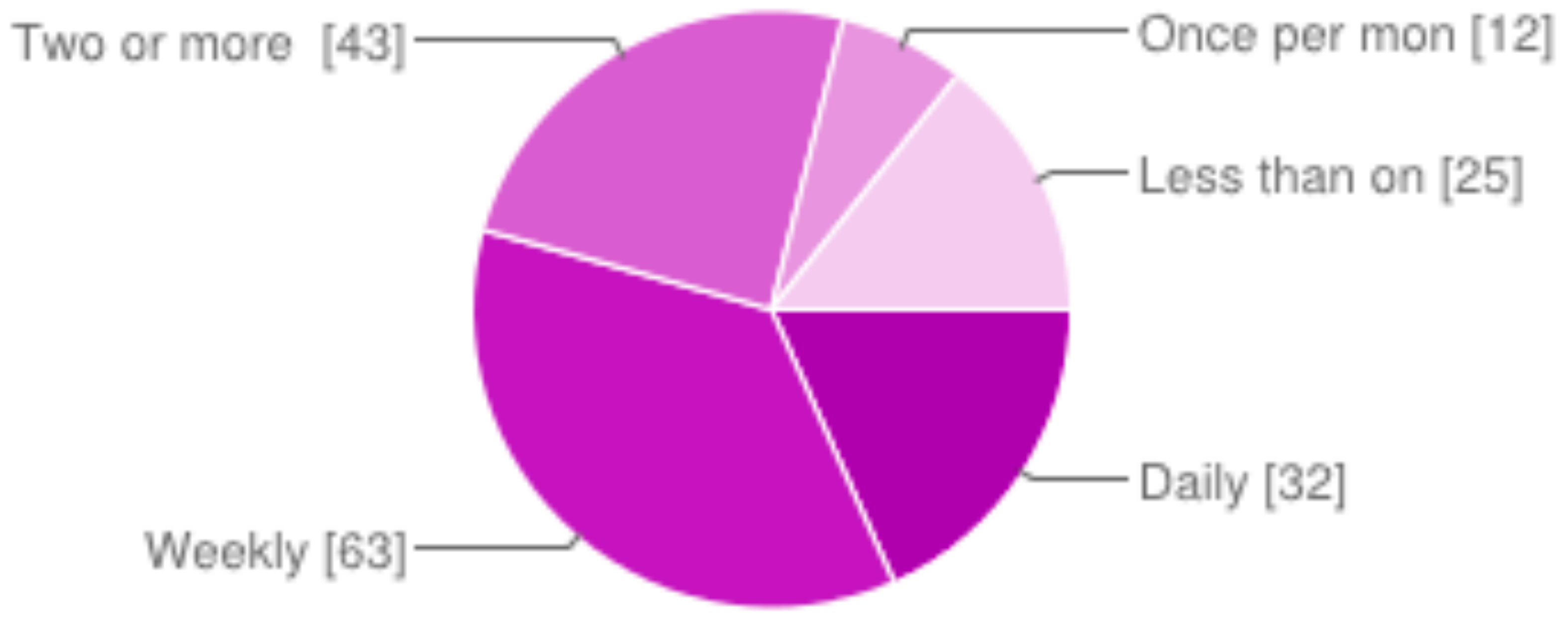
Programming Skills

How long have you been programming?



Less than 6 months	**39**	22%
Between 6 months and one year	**33**	19%
1 to 3 years	**62**	35%
Over 3 years	**41**	23%

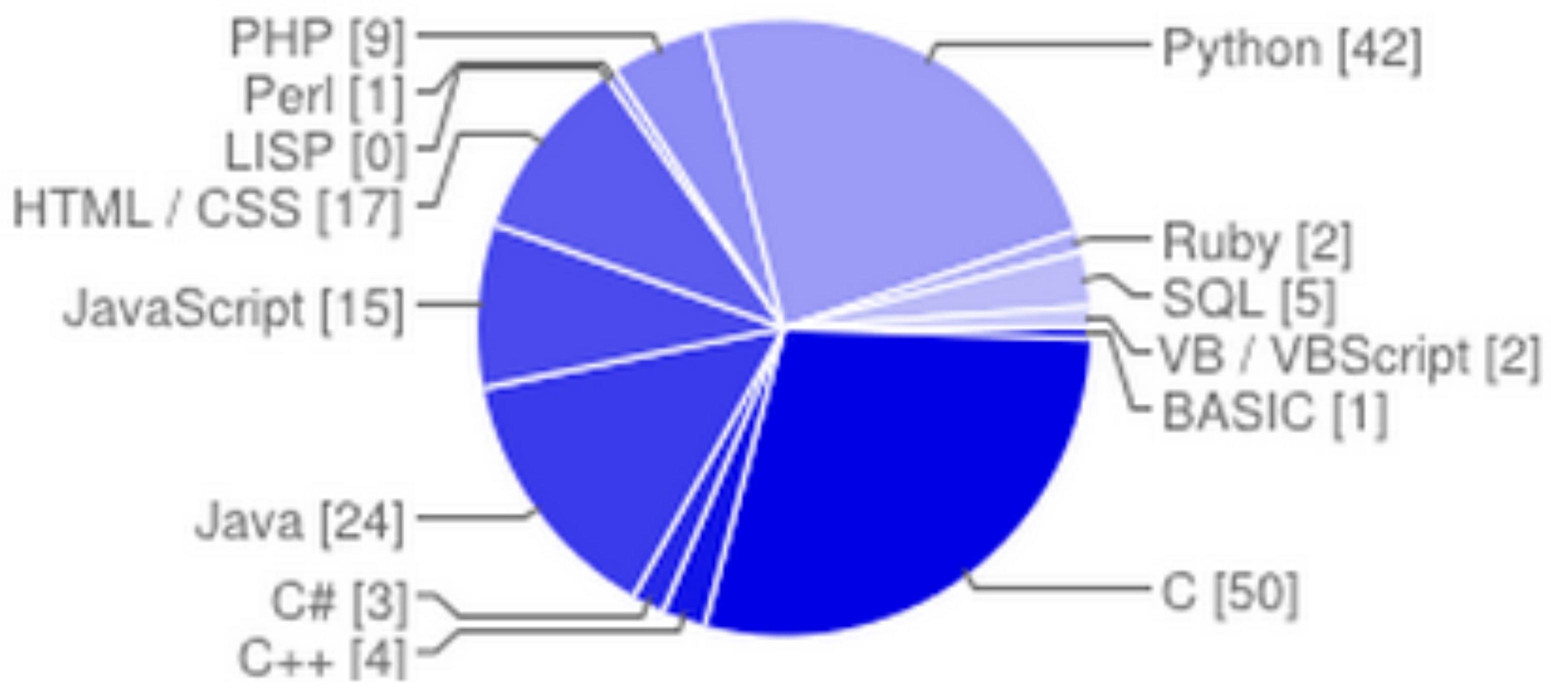
How often do you write code?

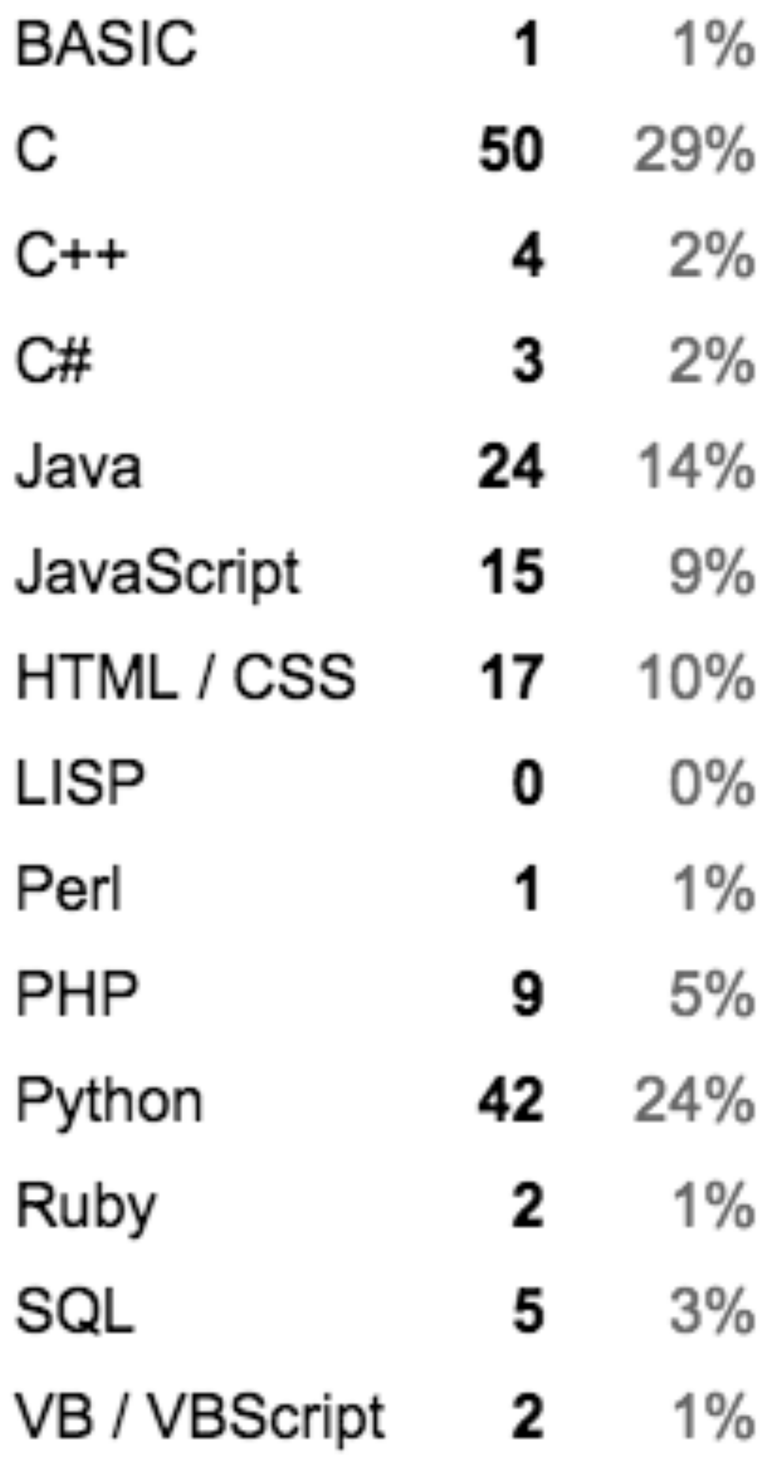


Daily	**32**	18%
Weekly	**63**	36%
Two or more times per month	**43**	25%
Once per month	**12**	7%
Less than once per month	**25**	14%

Primary Language

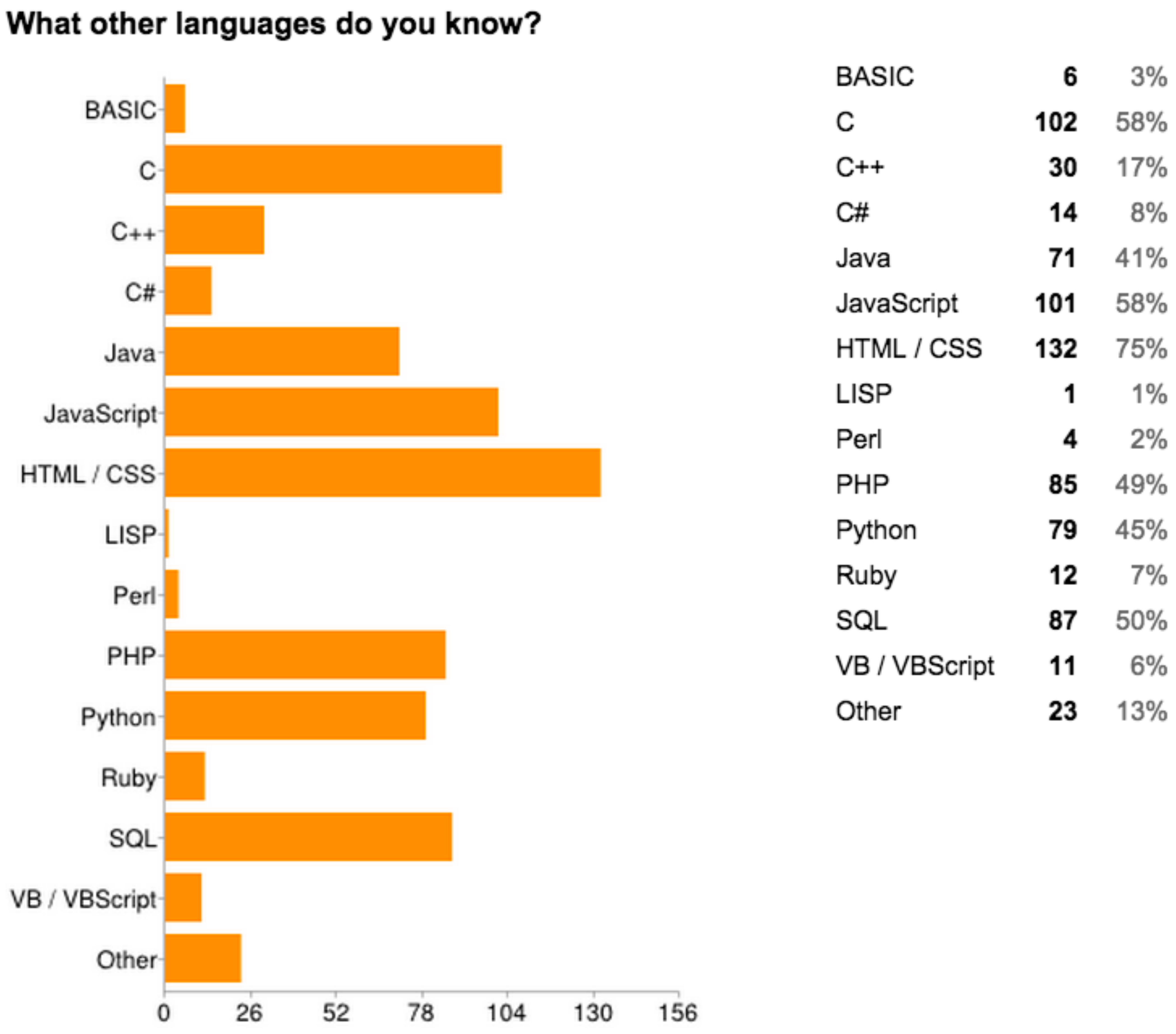
What is your primary programming language?





BASIC	**1**	1%
C	**50**	29%
C++	**4**	2%
C#	**3**	2%
Java	**24**	14%
JavaScript	**15**	9%
HTML / CSS	**17**	10%
LISP	**0**	0%
Perl	**1**	1%
PHP	**9**	5%
Python	**42**	24%
Ruby	**2**	1%
SQL	**5**	3%
VB / VBScript	**2**	1%

Other Languages

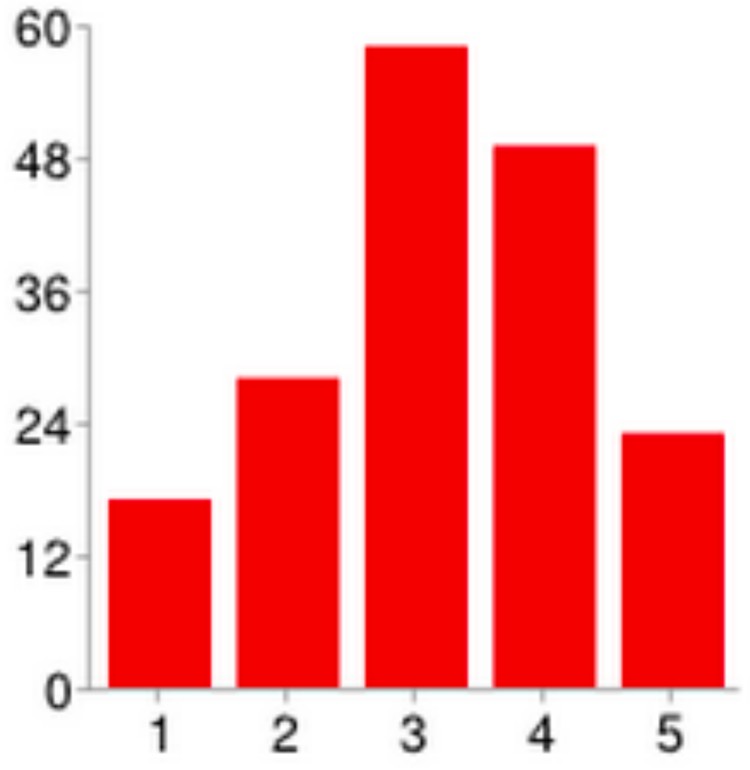


What other languages do you know?

BASIC	**6**	3%
C	**102**	58%
C++	**30**	17%
C#	**14**	8%
Java	**71**	41%
JavaScript	**101**	58%
HTML / CSS	**132**	75%
LISP	**1**	1%
Perl	**4**	2%
PHP	**85**	49%
Python	**79**	45%
Ruby	**12**	7%
SQL	**87**	50%
VB / VBScript	**11**	6%
Other	**23**	13%

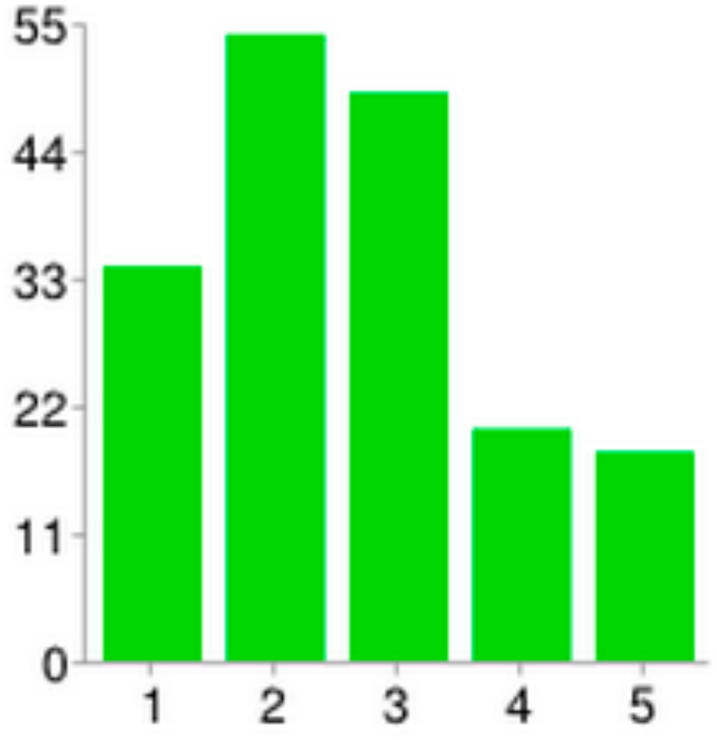
Your Comfort Zone

Overall, how comfortable are you with programming?



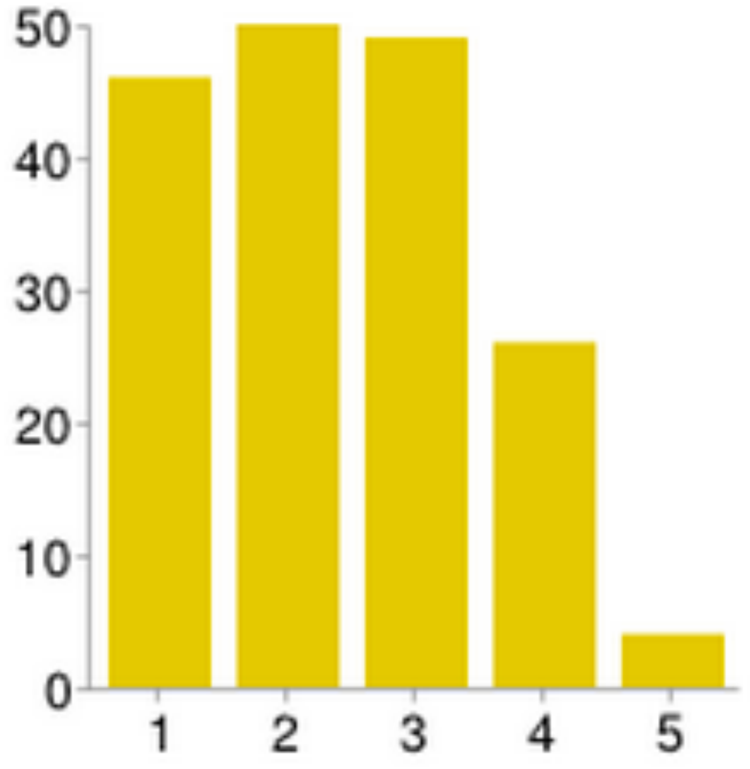
1	**17**	10%
2	**28**	16%
3	**58**	33%
4	**49**	28%
5	**23**	13%

How comfortable are you with design?



1	**34**	19%
2	**54**	31%
3	**49**	28%
4	**20**	11%
5	**18**	10%

Are you familiar with git for version control?



1	**46**	26%
2	**50**	29%
3	**49**	28%
4	**26**	15%
5	**4**	2%

Why take this class?

data
visualization
interested
learn
class
visualizations
want
course
CS
like
skills
design
science
useful
work
Visualization
interesting
working
also
Data
visualize
take
think
important
secondary
web
part
field
love
taking
really
learning
job
year
D3
Harvard
potential
decisions
find
know
information
creating
trying
area
interactive
experience
create
analysis
new
technical
cool
understanding
developing
looking
interest
seems
art
engineering
get
thought
well
others
feel
always
research
develop
career
making
use
semester
javascript
build
large
thesis
topic
Interested
end
computer
give
communicate
good
concentration
took
development
websites
wanted
best
order
requirement
elective
effectively
need
one
easy
combination
programming
last
applications
application
much
knowledge
great
people
done
many
front-end
project
background
Concentration
day
excited
make
way
something
ways
degree
program
projects
skill
better
pursue
effective
used
toward
Science
Also
concentrator
front
CS171
classes
practical
present
help
things
tools
big
datasets
able
fields
professional
lot
applied
opportunity
appealing
future
requirements
decided
going
company
statistics

What do you want to get out?

data
visualization
learn
like
visualizations
D3
design
able
skills
want
understanding
create
experience
Learn
coding
better
theory
knowledge
understand
programming
communicate
best
good
become
Javascript
javascript
web
creating
practices
make
get
use
comfortable
development
also
tools
effective
cool
represent
interested
apply
ability
principles
interactive
present
gain
job
practice
ways
course
really
learning
future
hope
things
sets
useful
end
visual
JavaScript
making
work
well
new
lot
skill
develop
dataset
way
something
Better
techniques
information
effectively
Design
improve
using
Data
know
visualizing
visualize
large
projects
class
looking
necessary
awesome
code
methods
display
just
Proficiency
Learning
Practical
visually
field
solid
graphics
build
project
side
d3
dynamic
pleasing
love
working
maybe
ideas
Become
advanced
familiar
building
possible
front-end

Design Experience

design

experience
sketch
None
sketching
like
web
designer
done
lot
designed
art
work
school
CS179
taking
took
high
classes
much
Photoshop
Nope
graphic
draw
learn
professional
visual
drawing
really
projects
years
CS
urban
Design
Adobe
studio
graphics
N/A
designs
little
taken
courses
knowledge
photoshop
websites
architecture
data
applications
architect
class
good
designing
also
know
use
mostly
formal
love
Architecture
Architect
UI/UX
Harvard
Currently
layouts
product
animation
software
developed
publications
Painting
arts
GIS
Yes
yes
well
computer
principles
enjoy
background
undergraduate
basic
planning
templates
comfortable
although
business
many
people
students
anything
Visual
print
since
artist
website
photography
Taking
college
focus
including
various
figures
sculpture
painting
definitely
trained
new
paint
often
semester
personal
project
nothing
coursework
several
Illustrator
management
fairly
pretty
now
means
part
learning
created
illustrator

Last Week

Visualization Definition

Visualization is the process that **transform**s
(abstract) **data** into
interactive graphical representations for the purpose of
exploration, confirmation, or presentation.

Why Visualize?

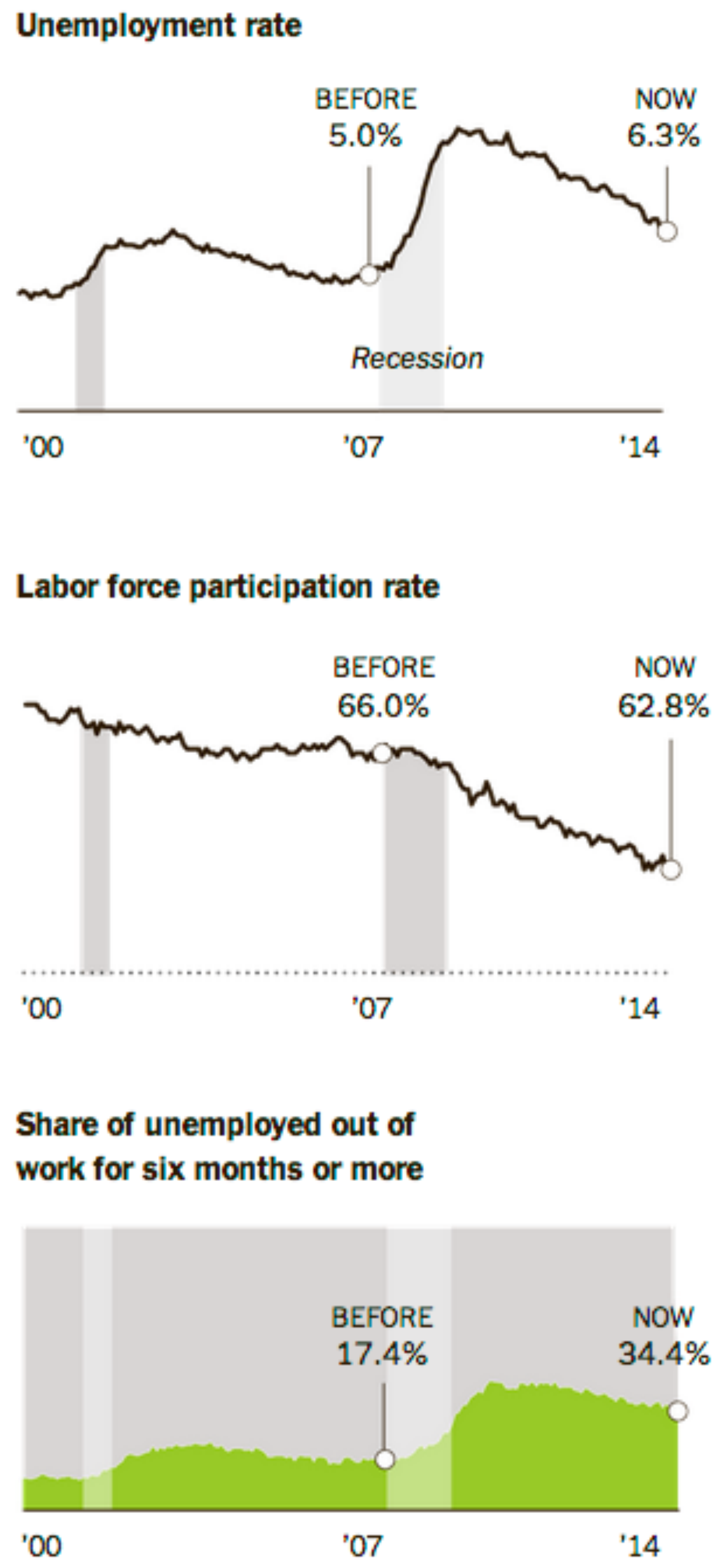
To inform humans: **Communication**

How did the unemployment and labor force develop over the last years?

When questions are not well defined: **Exploration**

Which combination of genes causes cancer?

Which drug can help patient X?



[New York Times]

When not to visualize?
When to automate?

Well defined question on well-defined dataset

Which gene is most frequently mutated in this set of patients?

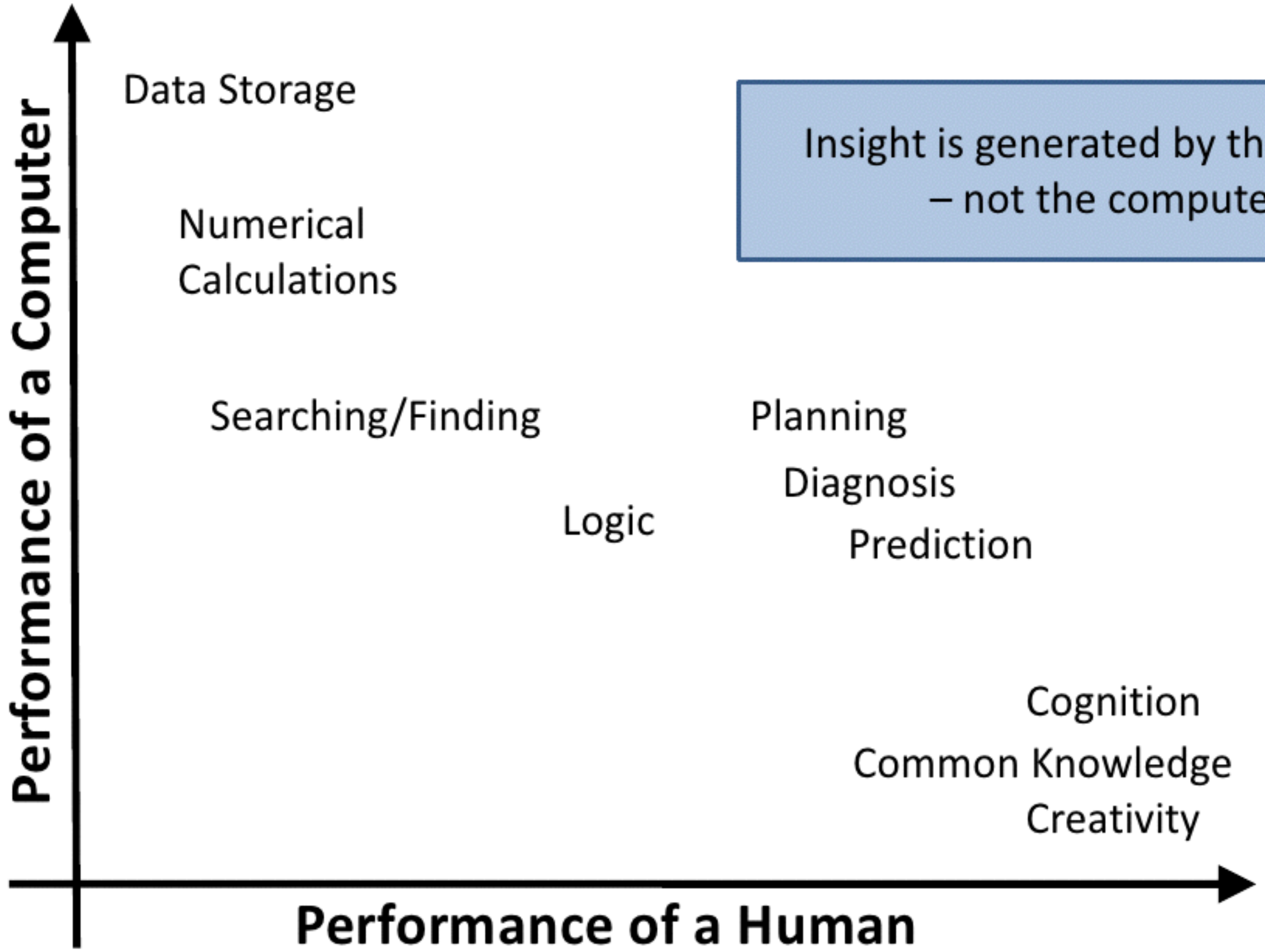
What is the current unemployment rate?

Decisions needed in minimal time

High frequency stock market trading: which stock to buy/sell?

Manufacturing: is bottle broken?

The Ability Matrix



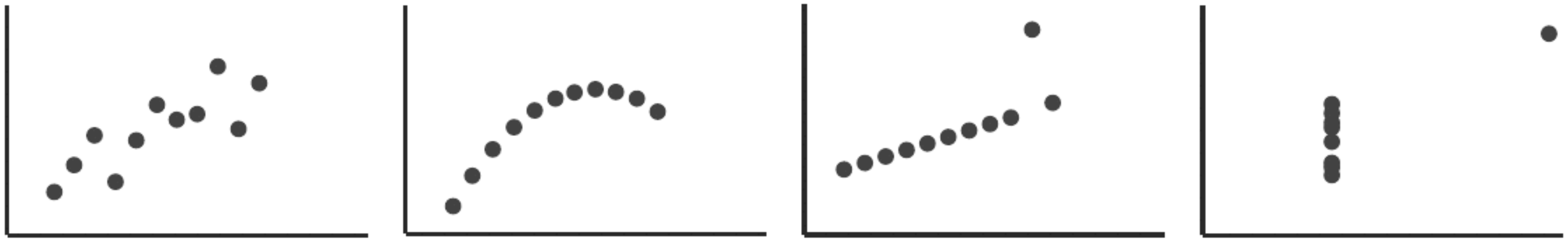
Insight is generated by the human – not the computer!

Why not just use Statistics?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.0	10	9.1	10	7.4	8	6.5
8	6.9	8	8.1	8	6.7	8	5.7
13	7.5	13	8.7	13	12.	8	7.7
9	8.8	9	8.7	9	7.1	8	8.8
11	8.3	11	9.2	11	7.8	8	8.4
14	9.9	14	8.1	14	8.8	8	7.0
6	7.2	6	6.1	6	6.0	8	5.2
4	4.2	4	3.1	4	5.3	19	12.
12	10.	12	9.1	12	8.1	8	5.5
7	4.8	7	7.2	7	6.4	8	7.9
5	5.						6.8

Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x

Anscombe’s Quartett



Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x

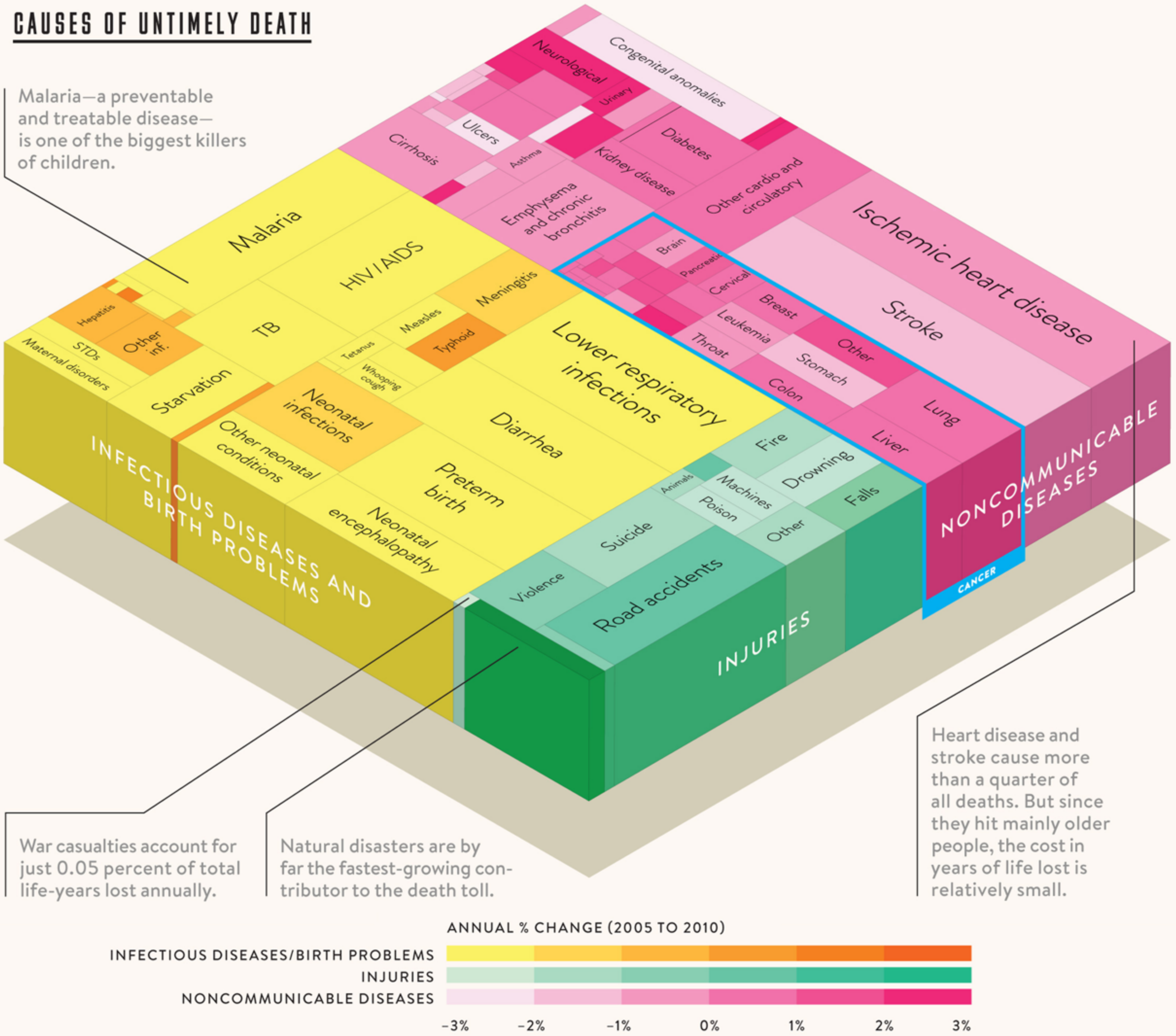
Design Critique

Design Excellence

“Well-designed presentations of interesting data are a matter of substance, of statistics, and of design.”

E. Tufte

CAUSES OF UNTIMELY DEATH



Malaria—a preventable and treatable disease—is one of the biggest killers of children.

War casualties account for just 0.05 percent of total life-years lost annually.

Natural disasters are by far the fastest-growing contributor to the death toll.

Heart disease and stroke cause more than a quarter of all deaths. But since they hit mainly older people, the cost in years of life lost is relatively small.

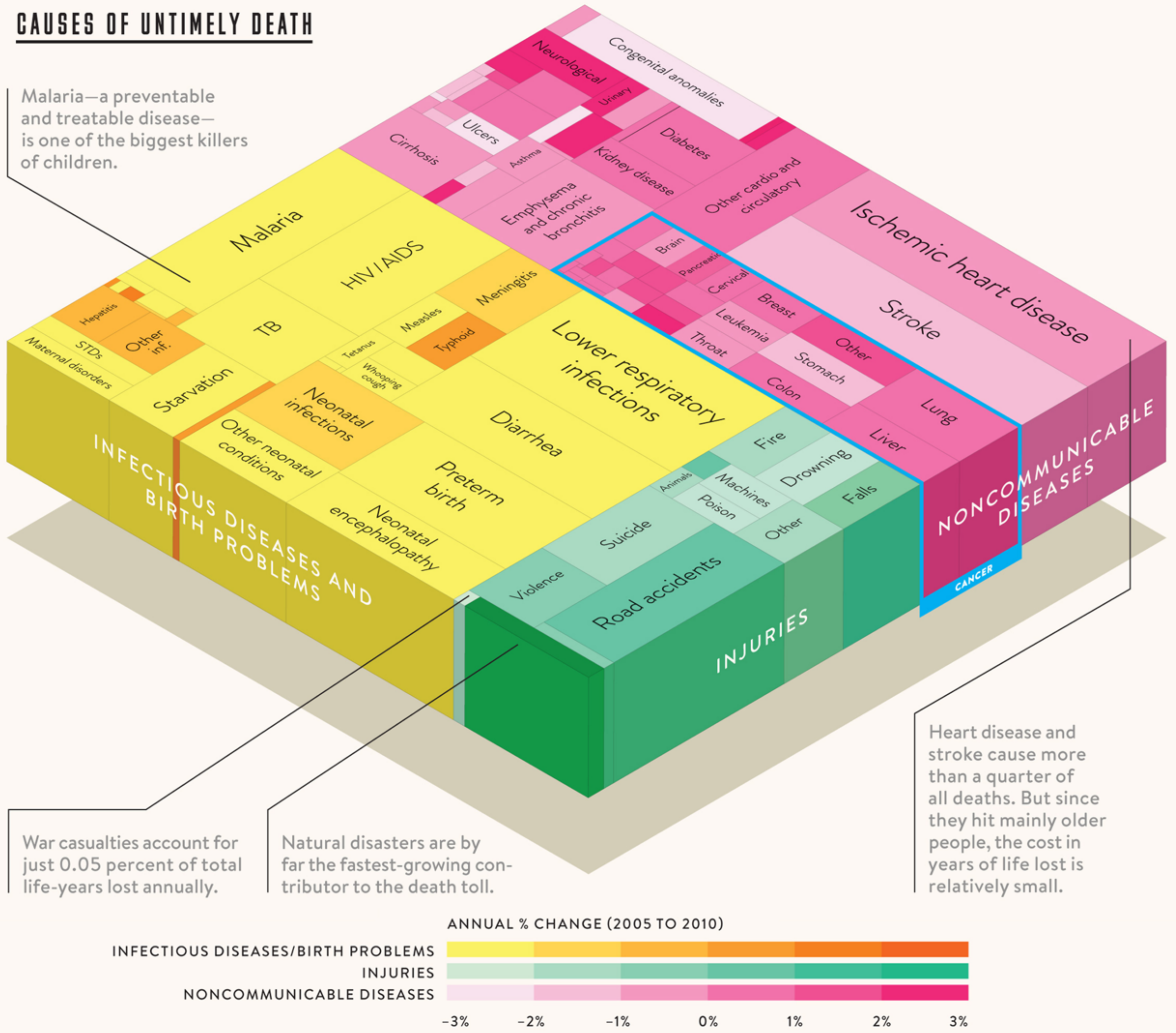
Graph of the Year?

"I love this graph because it shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. […] But there remains much to do to cut down the deaths in that yellow block even more dramatically. We have the solutions. But we need to keep up the support where they're being deployed […]“

-Bill Gates

http://goo.gl/W7ac3m

CAUSES OF UNTIMELY DEATH



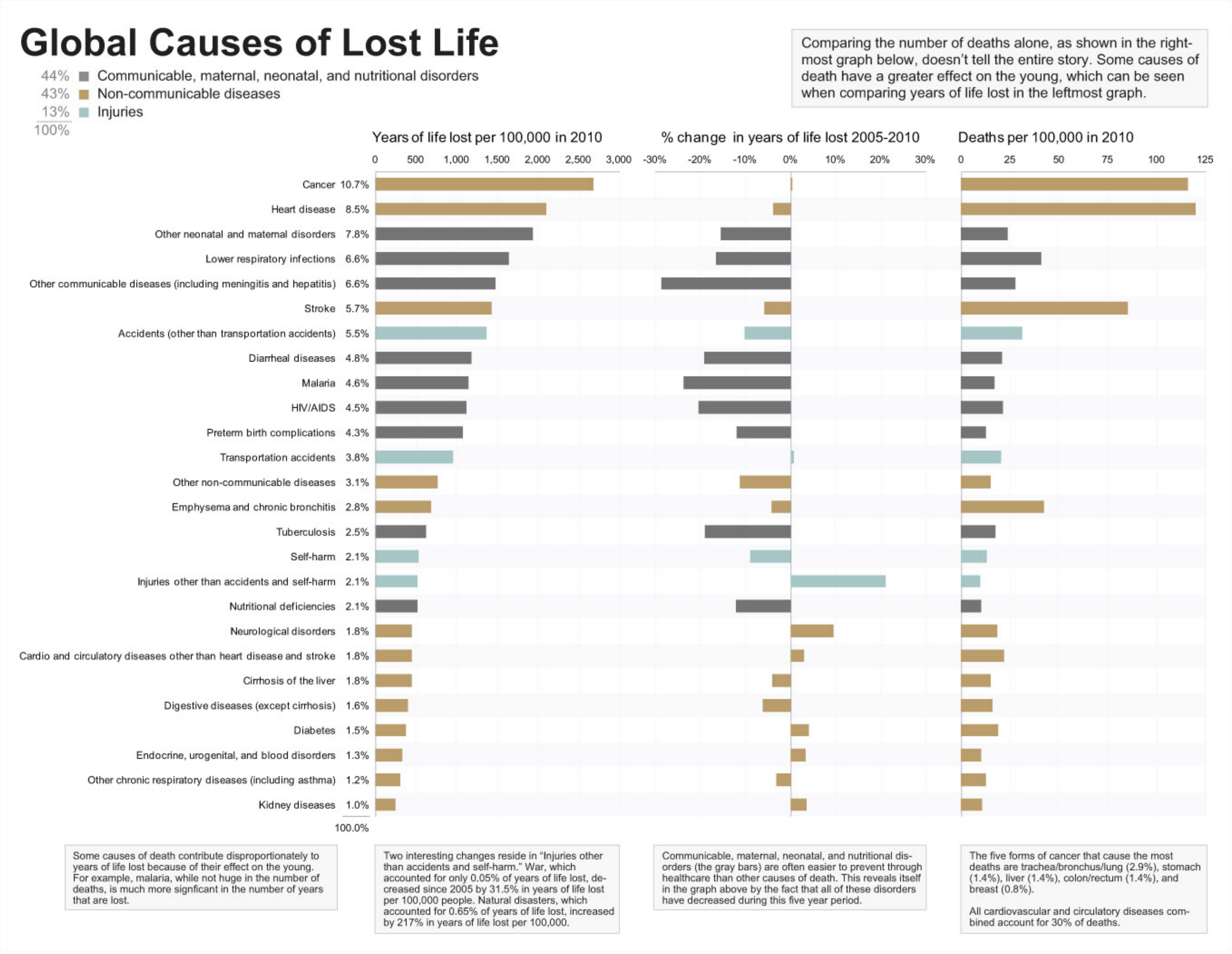
Malaria—a preventable and treatable disease—is one of the biggest killers of children.

War casualties account for just 0.05 percent of total life-years lost annually.

Natural disasters are by far the fastest-growing contributor to the death toll.

Heart disease and stroke cause more than a quarter of all deaths. But since they hit mainly older people, the cost in years of life lost is relatively small.

http://goo.gl/g6iTLb



Global Causes of Lost Life

Comparing the number of deaths alone, as shown in the right-most graph below, doesn't tell the entire story. Some causes of death have a greater effect on the young, which can be seen when comparing years of life lost in the leftmost graph.

Some causes of death contribute disproportionately to years of life lost because of their effect on the young. For example, malaria, while not huge in the number of deaths, is much more signficant in the number of years that are lost.

Two interesting changes reside in "Injuries other than accidents and self-harm." War, which accounted for only 0.05% of years of life lost, decreased since 2005 by 31.5% in years of life lost per 100,000 people. Natural disasters, which accounted for 0.65% of years of life lost, increased by 217% in years of life lost per 100,000.

Communicable, maternal, neonatal, and nutritional disorders (the gray bars) are often easier to prevent through healthcare than other causes of death. This reveals itself in the graph above by the fact that all of these disorders have decreased during this five year period.

The five forms of cancer that cause the most deaths are trachea/bronchus/lung (2.9%), stomach (1.4%), liver (1.4%), colon/rectum (1.4%), and breast (0.8%).

All cardiovascular and circulatory diseases combined account for 30% of deaths.

Redesign by

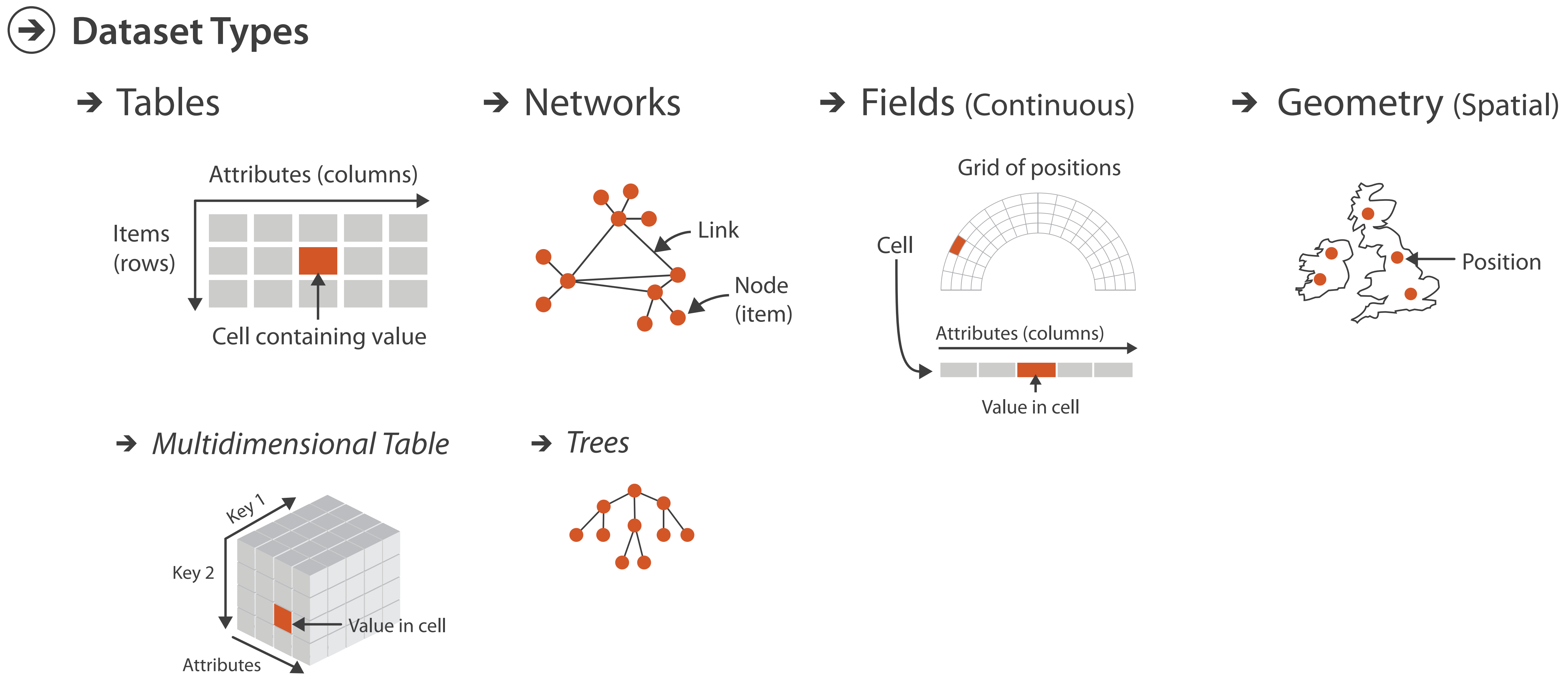
Perceptual Edge

Data

Terms

Dataset Types

what can be visualized?



➔ Dataset Types

➔ Tables ➔ Networks ➔ Fields (Continuous) ➔ Geometry (Spatial)

➔ *Multidimensional Table* ➔ *Trees*

Data Types

fundamental units

combinations make up Dataset Types

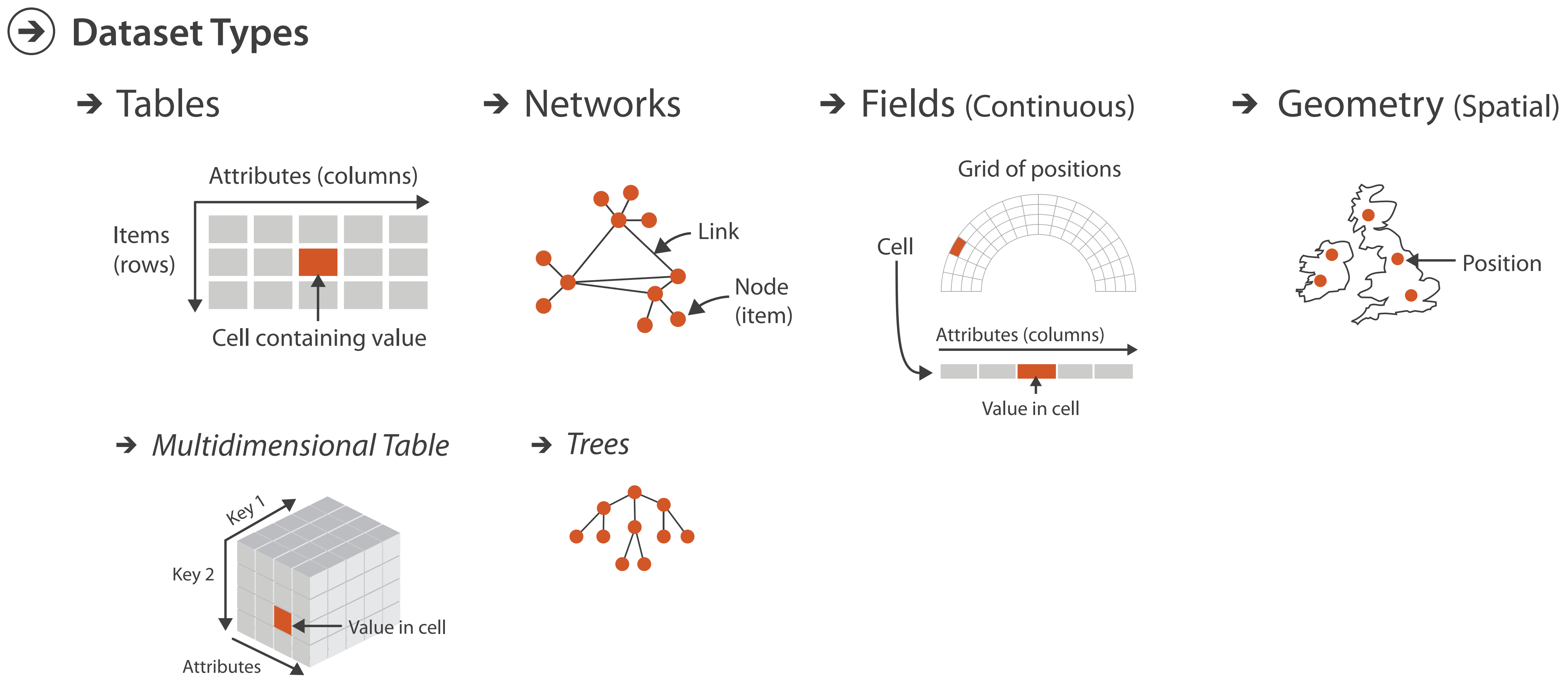
➔ Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

Structure

Structured Data

known data types, semantics



Unstructured Data

no predefined data model

text-heavy, interspersed with facts (dates, times, locations)

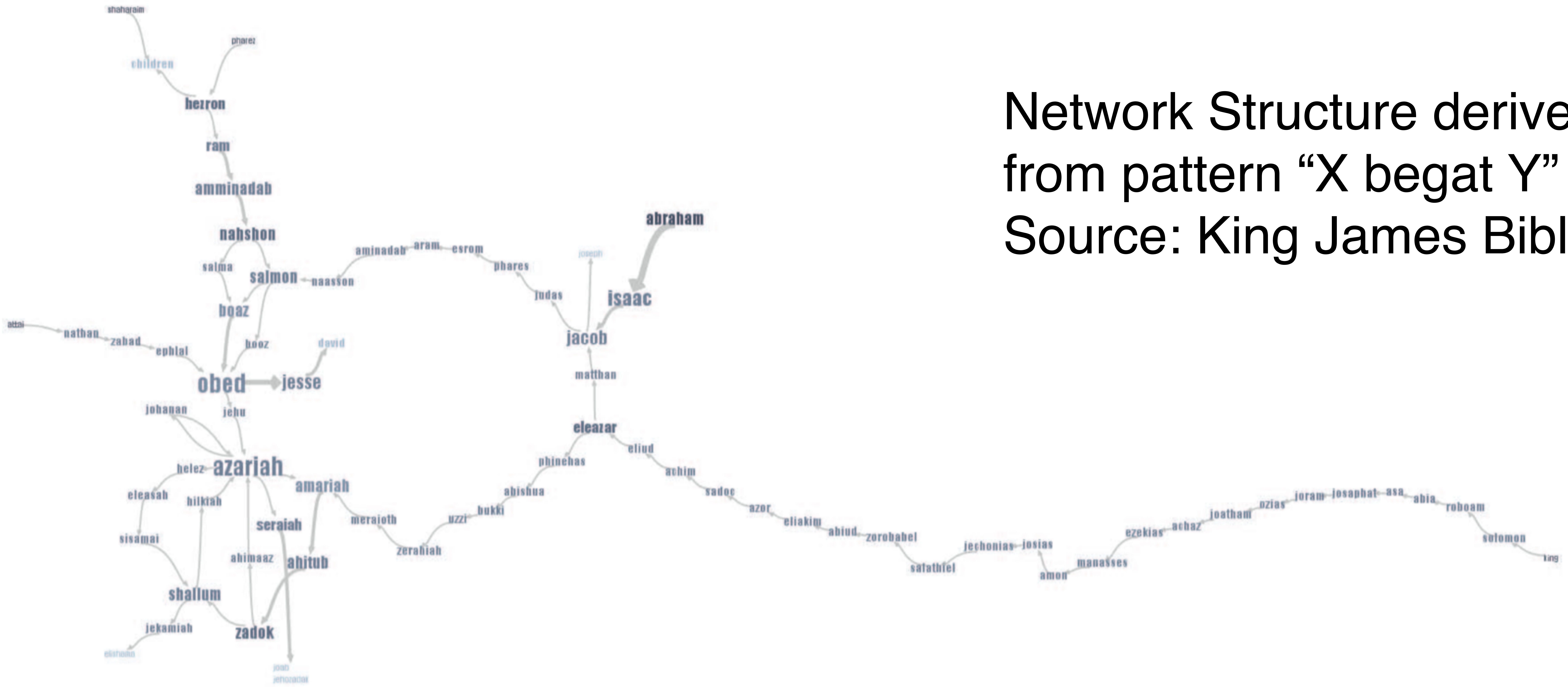
video, images

Translate into structured data

Natural Language Processing

Text mining (sentiment, keywords, concepts, categories)

Text Example: Phrase Net



Network Structure derived from pattern “X begat Y”
Source: King James Bible

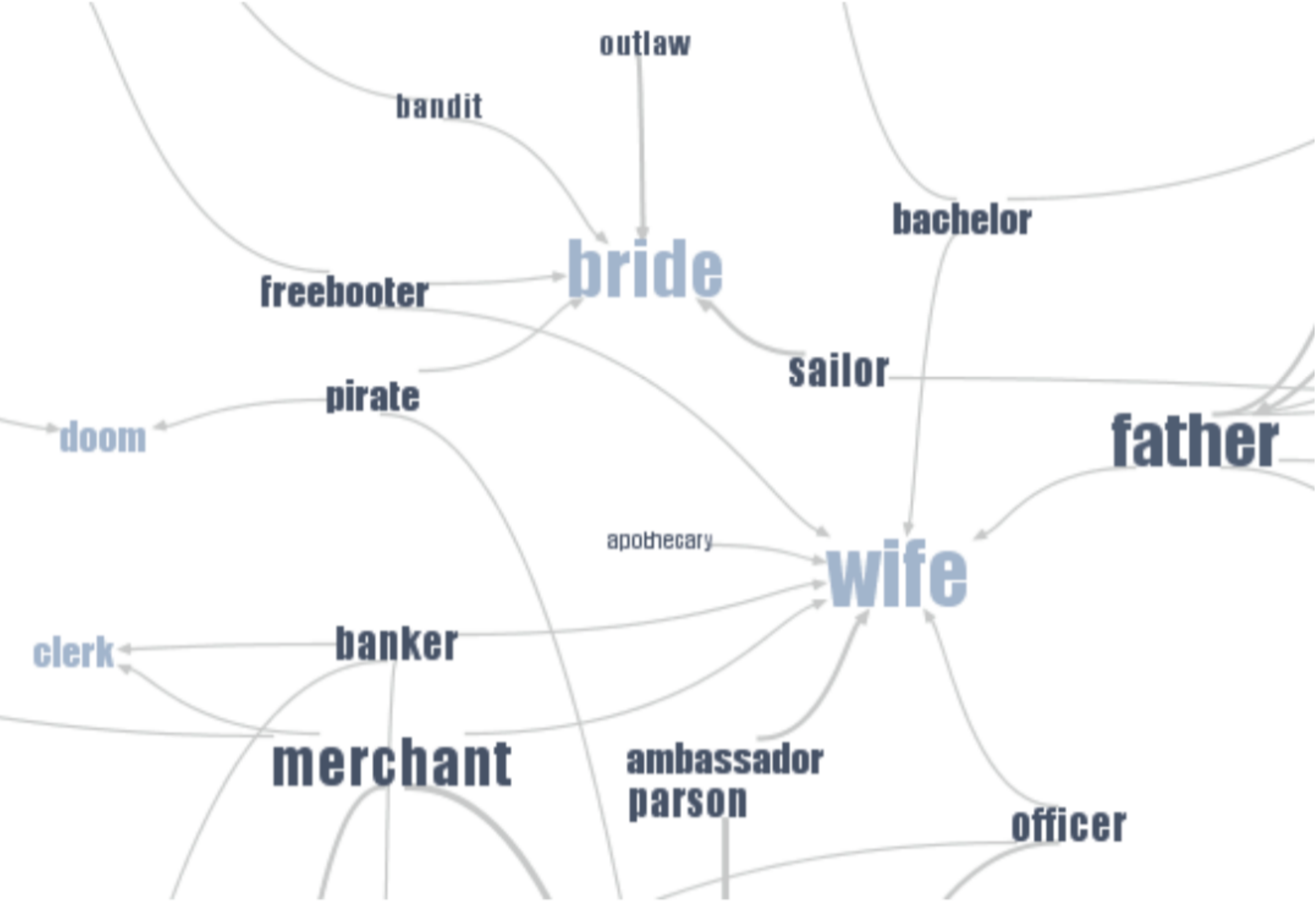
[van Ham, InfoVis 2009]

Example: Phrase Net

Pattern: “X’s Y”

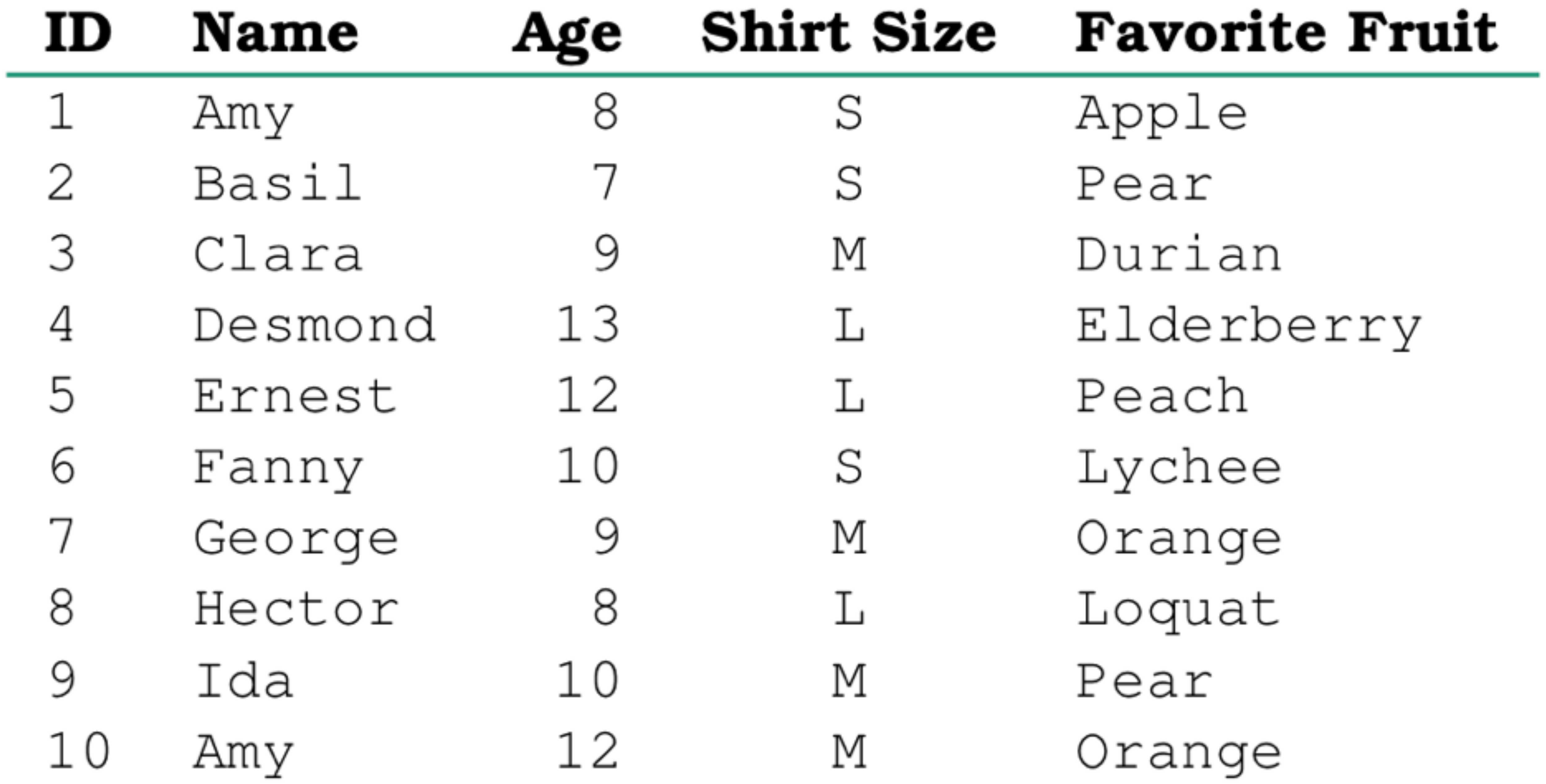
18th & 19th century novels

More in Lecture 13: Text & Document Vis



[van Ham, InfoVis 2009]

Data Semantics



ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Basil, 7, S, Pear

What does it mean?

Semantics: real world meaning

- Name? City? Fruit? Height? Age? Day of Month?
- Metadata

Data Types

structural or mathematical interpretation of data

Item, Link, Attribute, Position, Grid

Different from data types in programming!

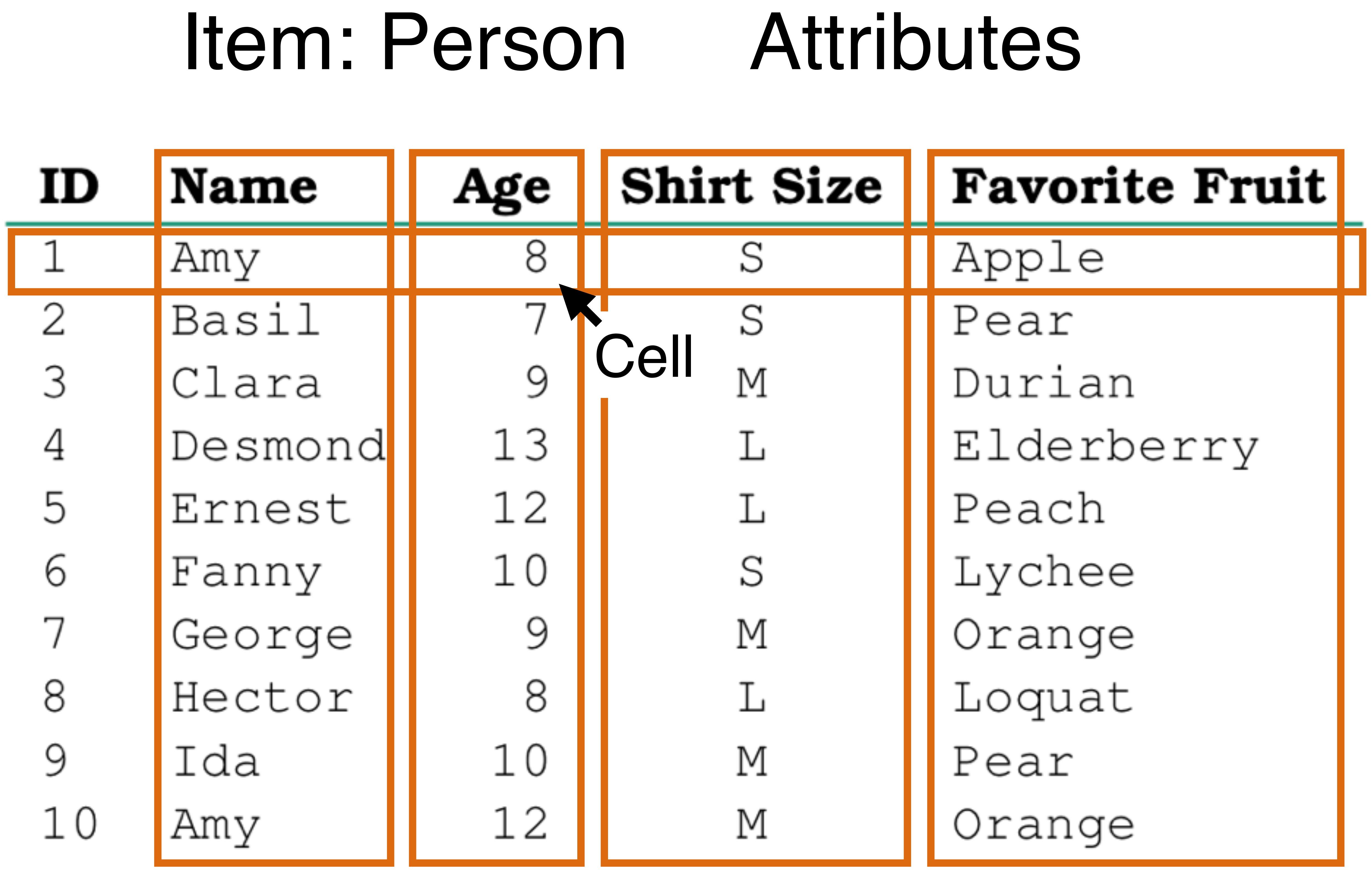
Items & Attributes

Item: individual entity, discrete

e.g., Patient, Car, Stock, City

Attribute: measured, observed, logged property

e.g., Patient: height, blood pressure; Car: horsepower, make



Item: Person

Attributes

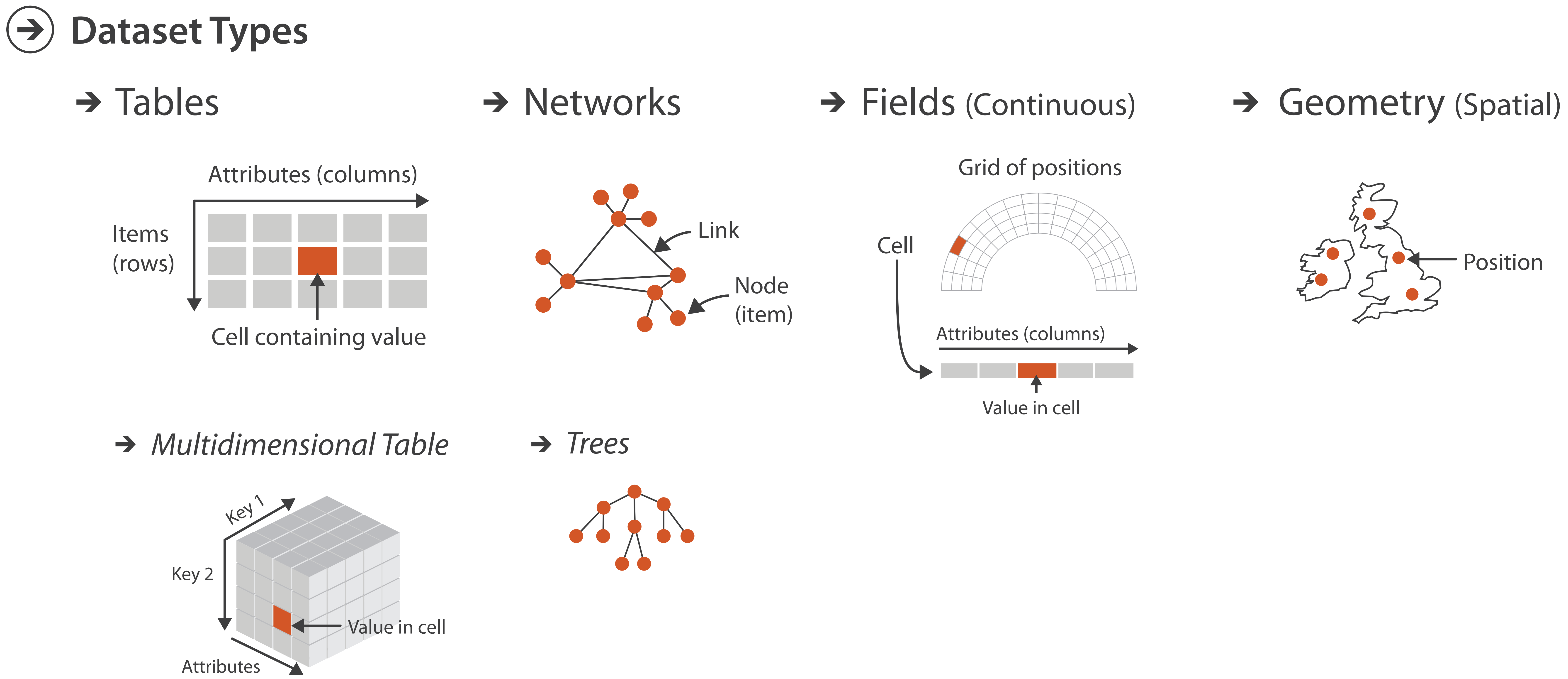
ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Cell

Other Data Types

- Links
  - Express relationship between two items
  - Friendship on Facebook, Interaction between proteins
- Positions
  - Spatial data -> location in 2D or 3D
  - Pixels in photo, Voxels in MRI scan, latitude/longitude
- Grids
  - Sampling strategy for continuous data
  - How many Voxels in MRI scan, positions of weather stations in the US

Dataset Types



Dataset Types

➔ Tables

➔ Networks

➔ Fields (Continuous)

➔ Geometry (Spatial)

➔ *Multidimensional Table*

➔ *Trees*

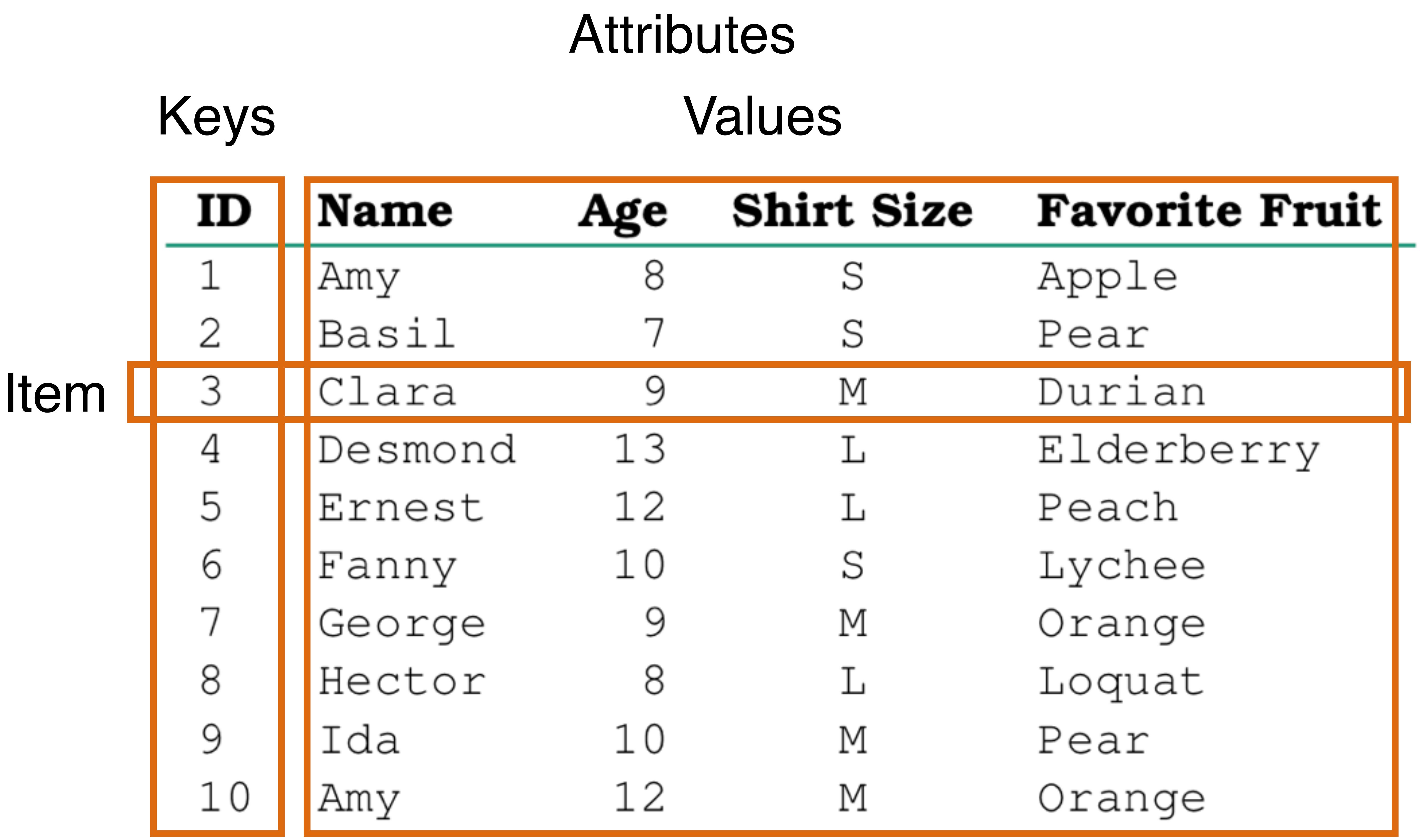
Tables

Flat Table

- one item per row
- each column is attribute
- unique (implicit) **key**
- no duplicates

Multidimensional Table

- indexing based on multiple keys



Attributes

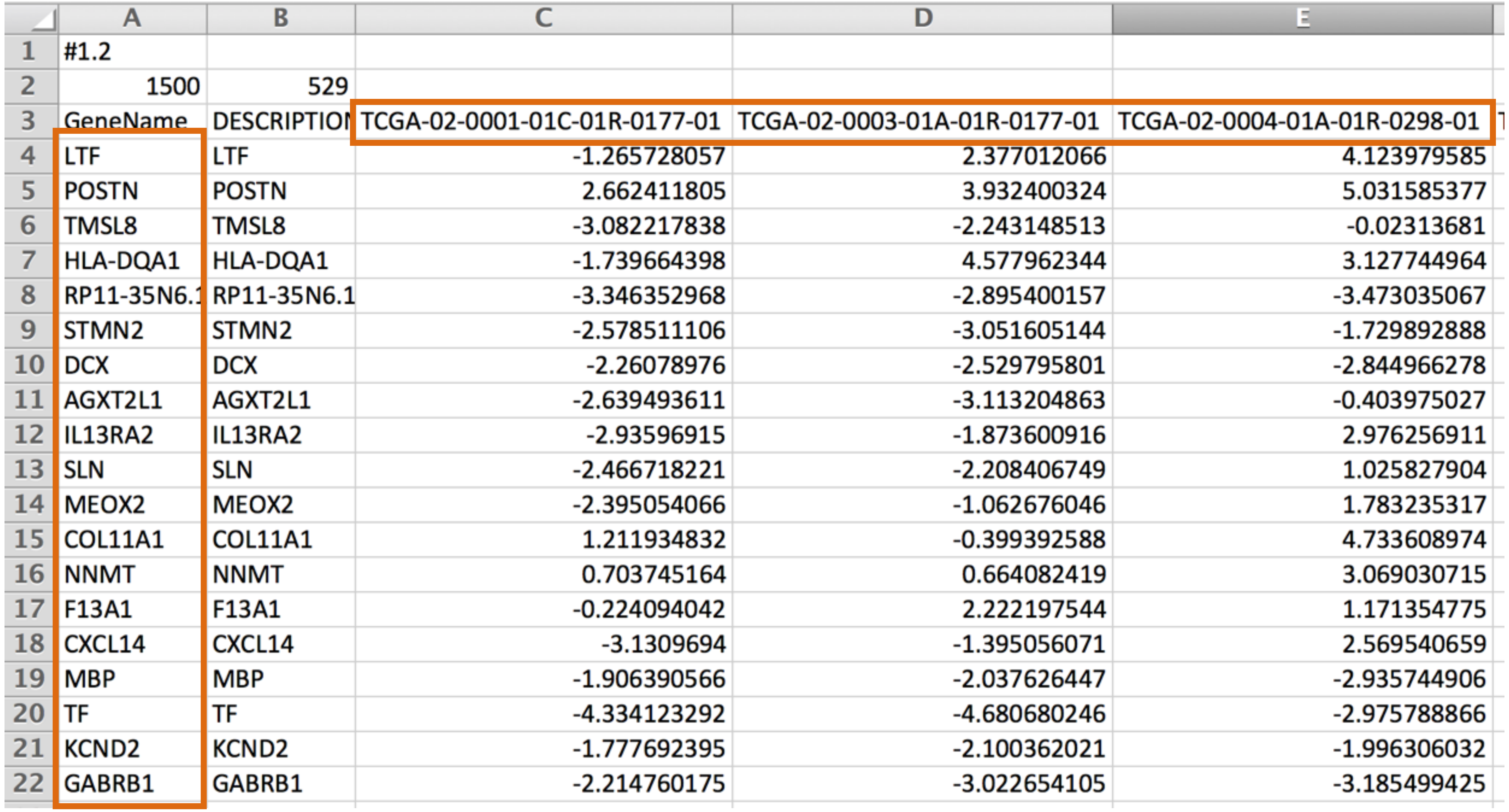
Keys

Values

Item

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Multidimensional Tables

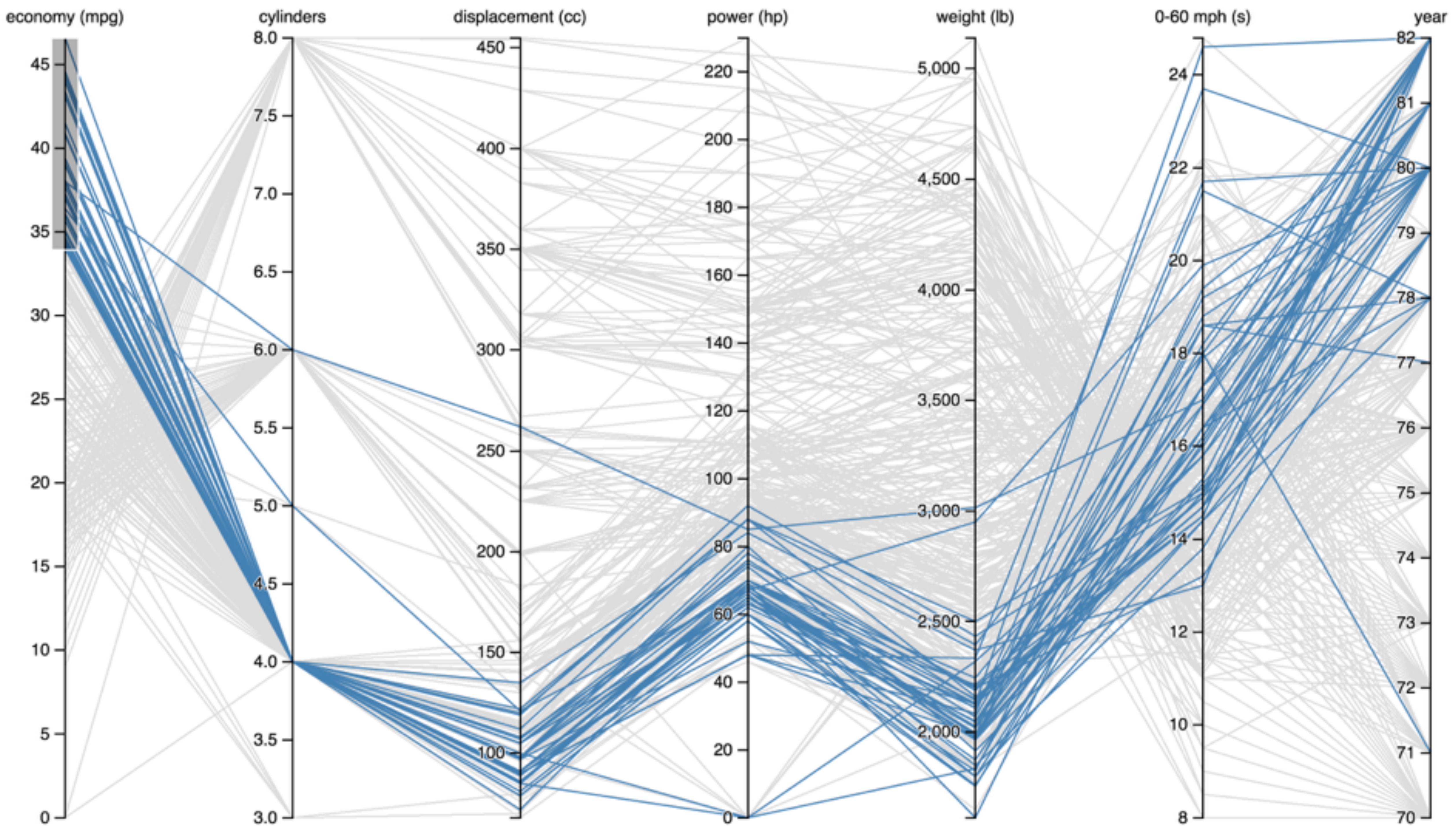


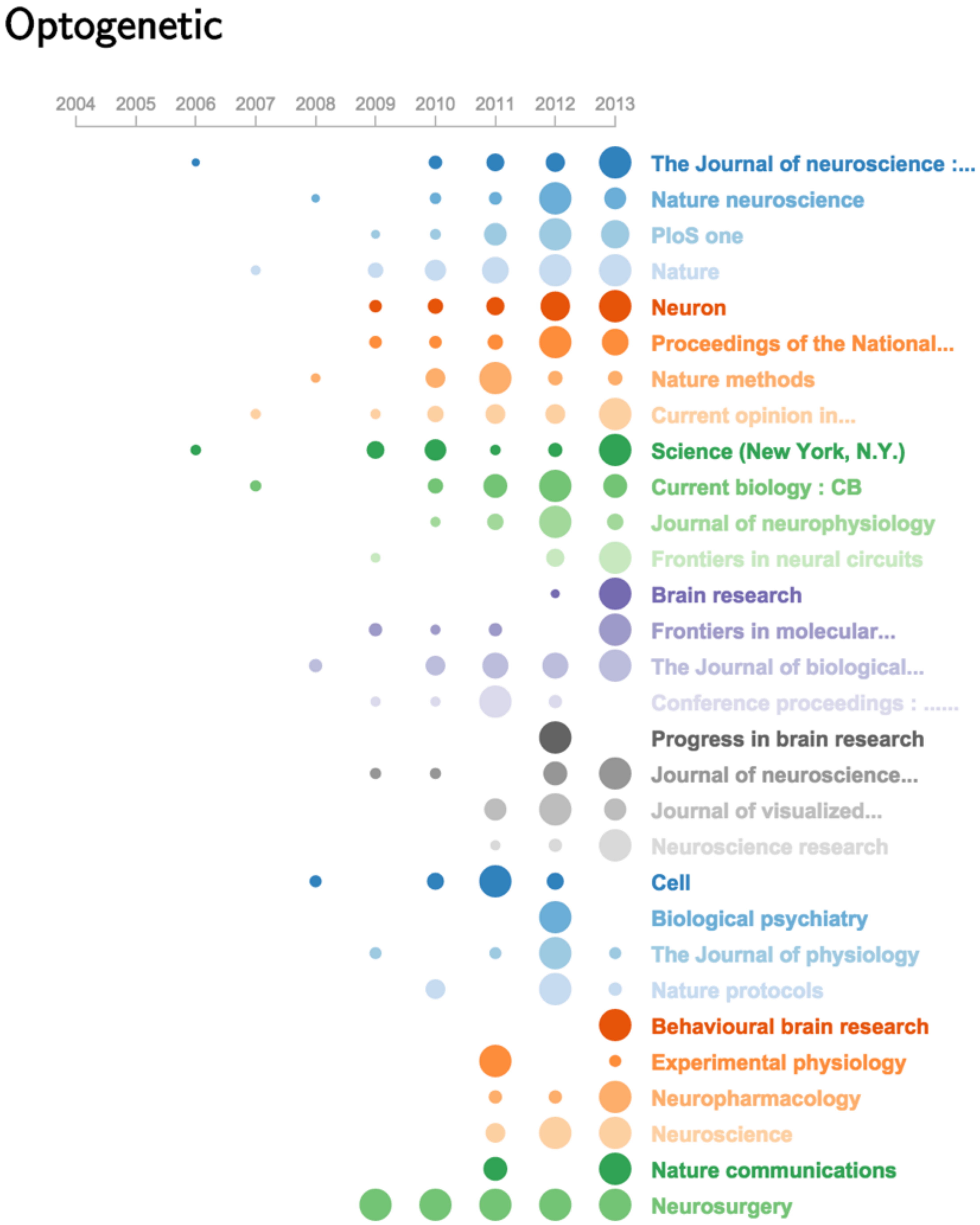
	A	B	C	D	E
1	#1.2				
2	1500	529			
3	GeneName	DESCRIPTION	TCGA-02-0001-01C-01R-0177-01	TCGA-02-0003-01A-01R-0177-01	TCGA-02-0004-01A-01R-0298-01
4	LTF	LTF	-1.265728057	2.377012066	4.123979585
5	POSTN	POSTN	2.662411805	3.932400324	5.031585377
6	TMSL8	TMSL8	-3.082217838	-2.243148513	-0.02313681
7	HLA-DQA1	HLA-DQA1	-1.739664398	4.577962344	3.127744964
8	RP11-35N6.1	RP11-35N6.1	-3.346352968	-2.895400157	-3.473035067
9	STMN2	STMN2	-2.578511106	-3.051605144	-1.729892888
10	DCX	DCX	-2.26078976	-2.529795801	-2.844966278
11	AGXT2L1	AGXT2L1	-2.639493611	-3.113204863	-0.403975027
12	IL13RA2	IL13RA2	-2.93596915	-1.873600916	2.976256911
13	SLN	SLN	-2.466718221	-2.208406749	1.025827904
14	MEOX2	MEOX2	-2.395054066	-1.062676046	1.783235317
15	COL11A1	COL11A1	1.211934832	-0.399392588	4.733608974
16	NNMT	NNMT	0.703745164	0.664082419	3.069030715
17	F13A1	F13A1	-0.224094042	2.222197544	1.171354775
18	CXCL14	CXCL14	-3.1309694	-1.395056071	2.569540659
19	MBP	MBP	-1.906390566	-2.037626447	-2.935744906
20	TF	TF	-4.334123292	-4.680680246	-2.975788866
21	KCND2	KCND2	-1.777692395	-2.100362021	-1.996306032
22	GABRB1	GABRB1	-2.214760175	-3.022654105	-3.185499425

Keys: Patients

Keys: Genes

Visualizing Tables

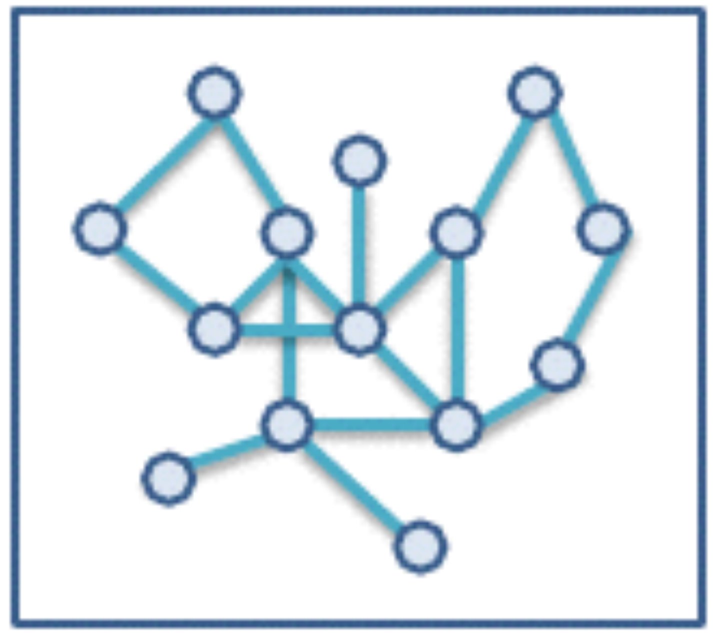




More in Lecture 8: High-Dimensional Data

Graphs/Networks

A graph G(V,E) consists of a set of **vertices (nodes)** V and a set of **edges (links)** E connecting these vertices.

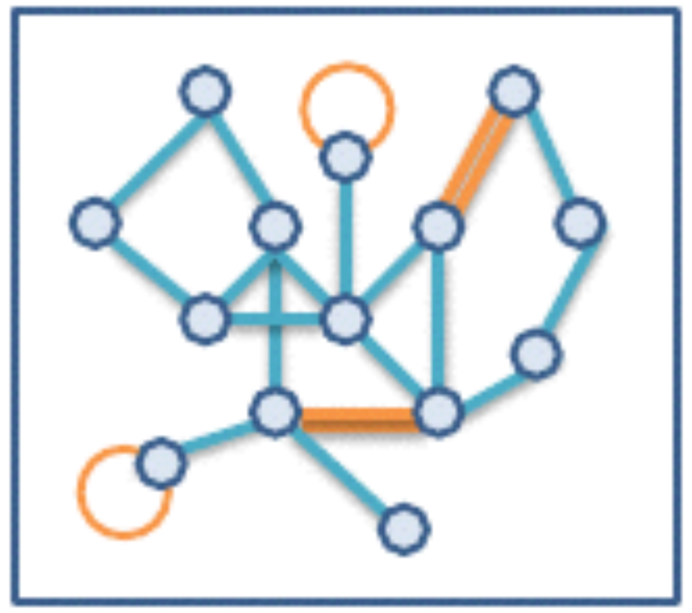


Diagrammatic Example

Graphs/Networks

A simple graph is a graph which contains

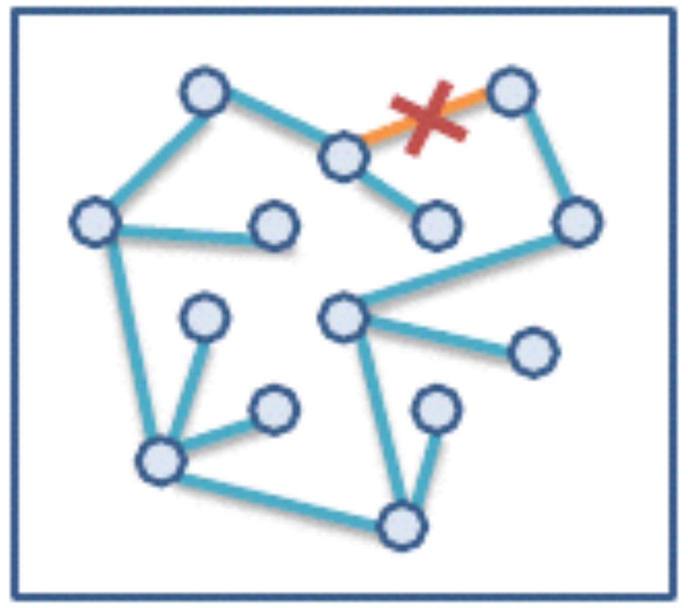
- No multi-edges
- No loops



Not a simple graph!
→ A ***general graph***

Special Graphs

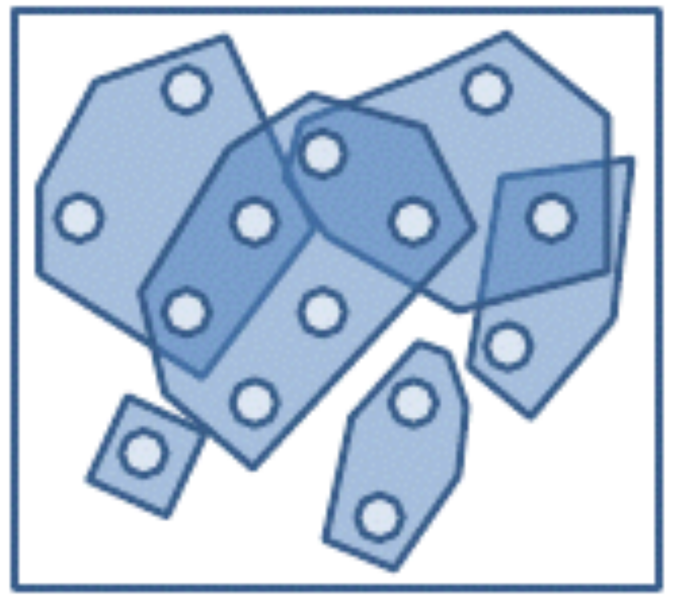
A ***tree*** is a graph with *no cycles*



Tree

A ***directed graph*** (digraph) is a graph that distinguishes between edges A-> B and A <- B

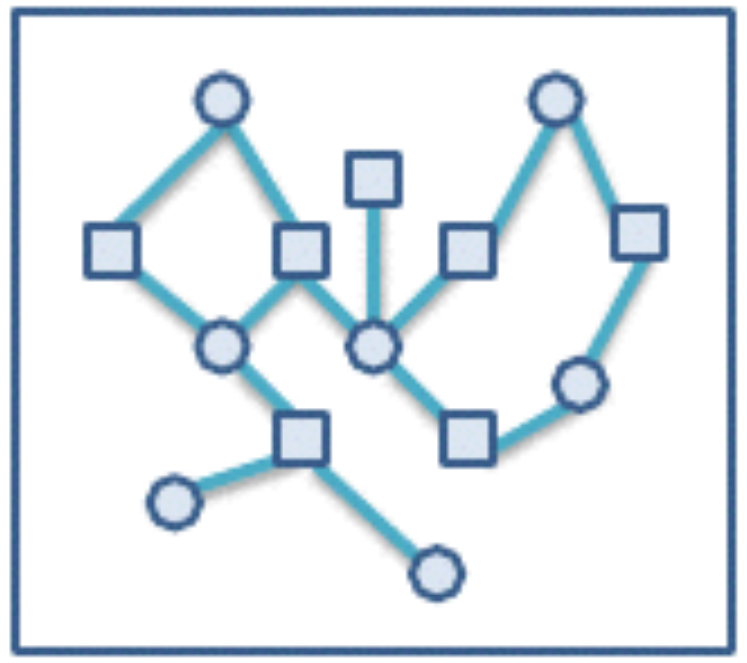
A ***hypergraph*** is a graph with edges connecting any number of vertices



Hypergraph Example

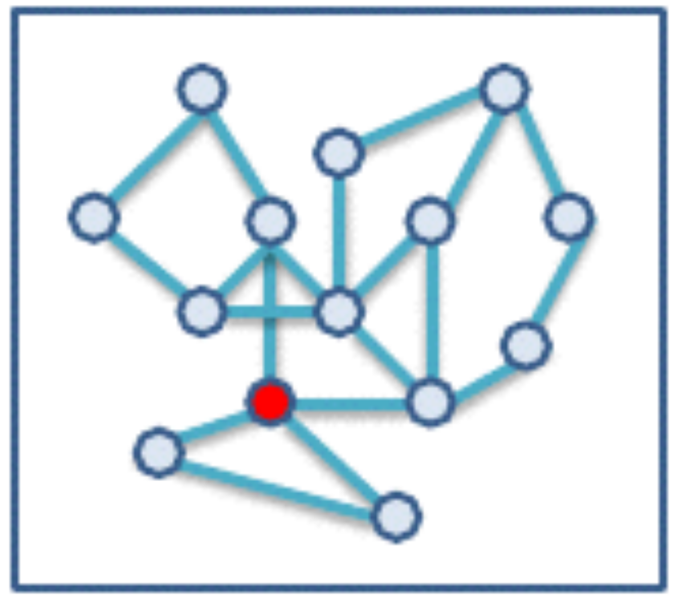
Special Graphs

A ***bipartite graph*** has vertices that can be partitioned into two independent sets



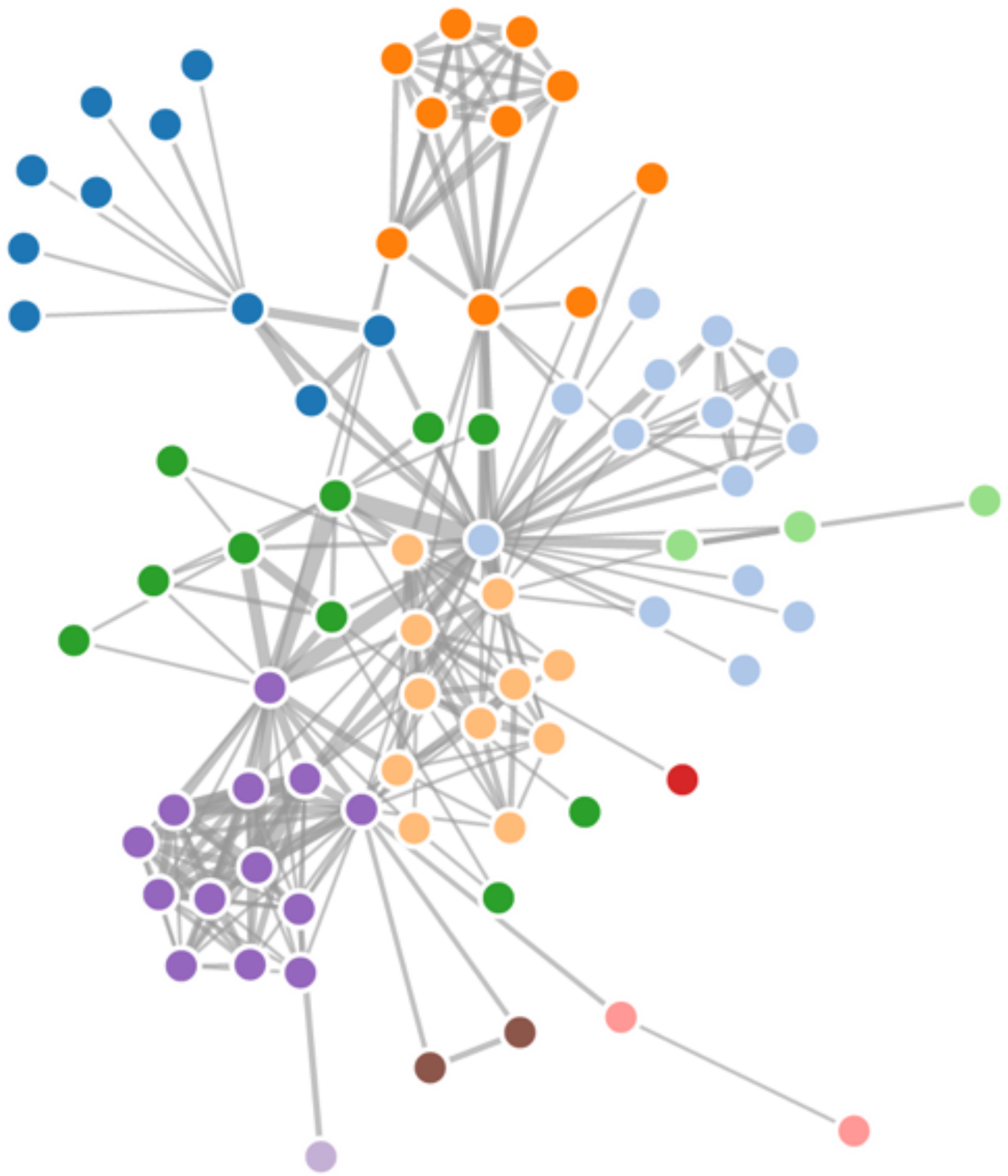
Bipartite Graph

An ***articulation point*** is a Vertex, which if deleted from the graph would break up a ***connected graph*** into multiple graphs,or an ***unconnected graph***

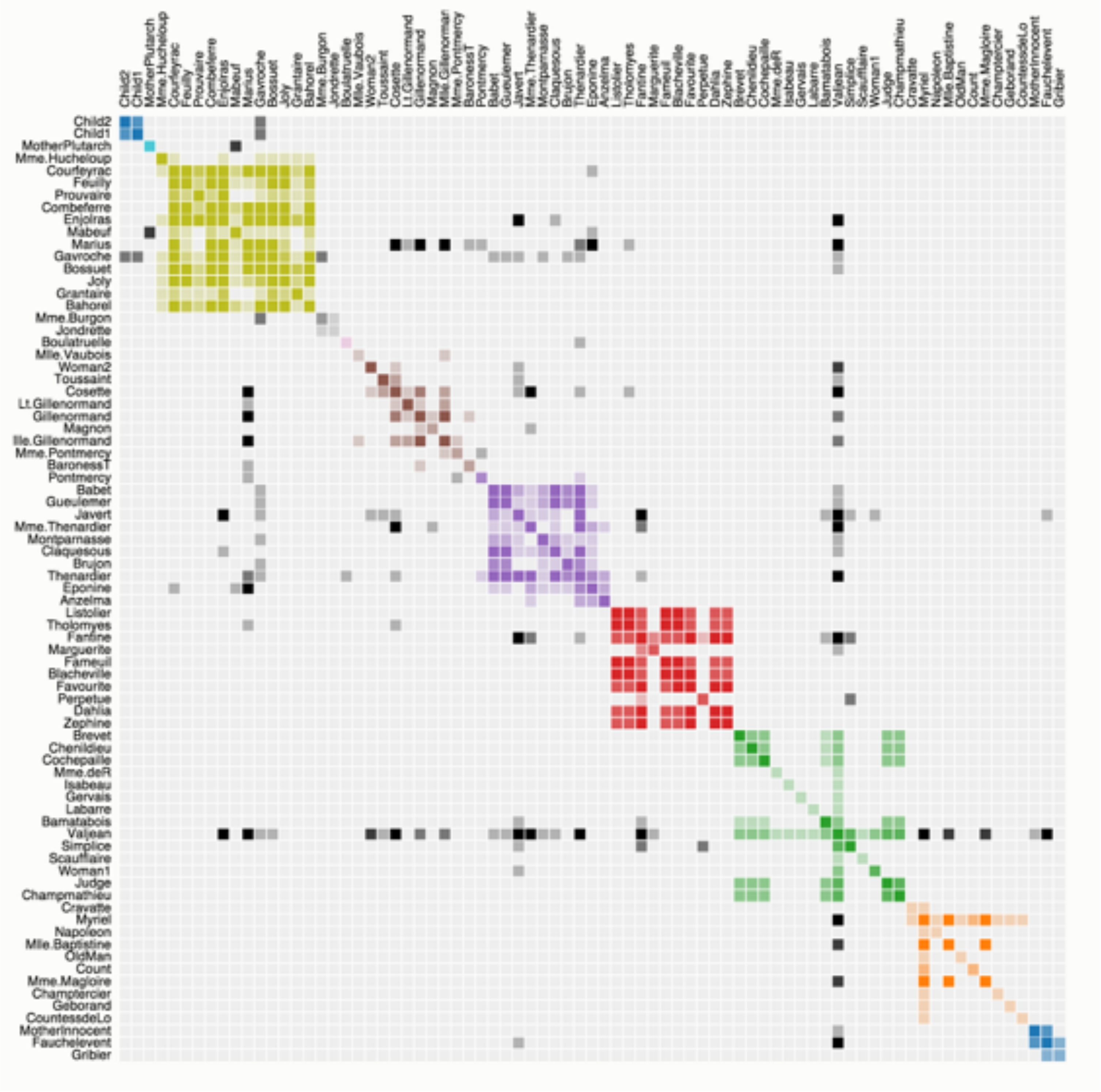


Articulation Point (red)

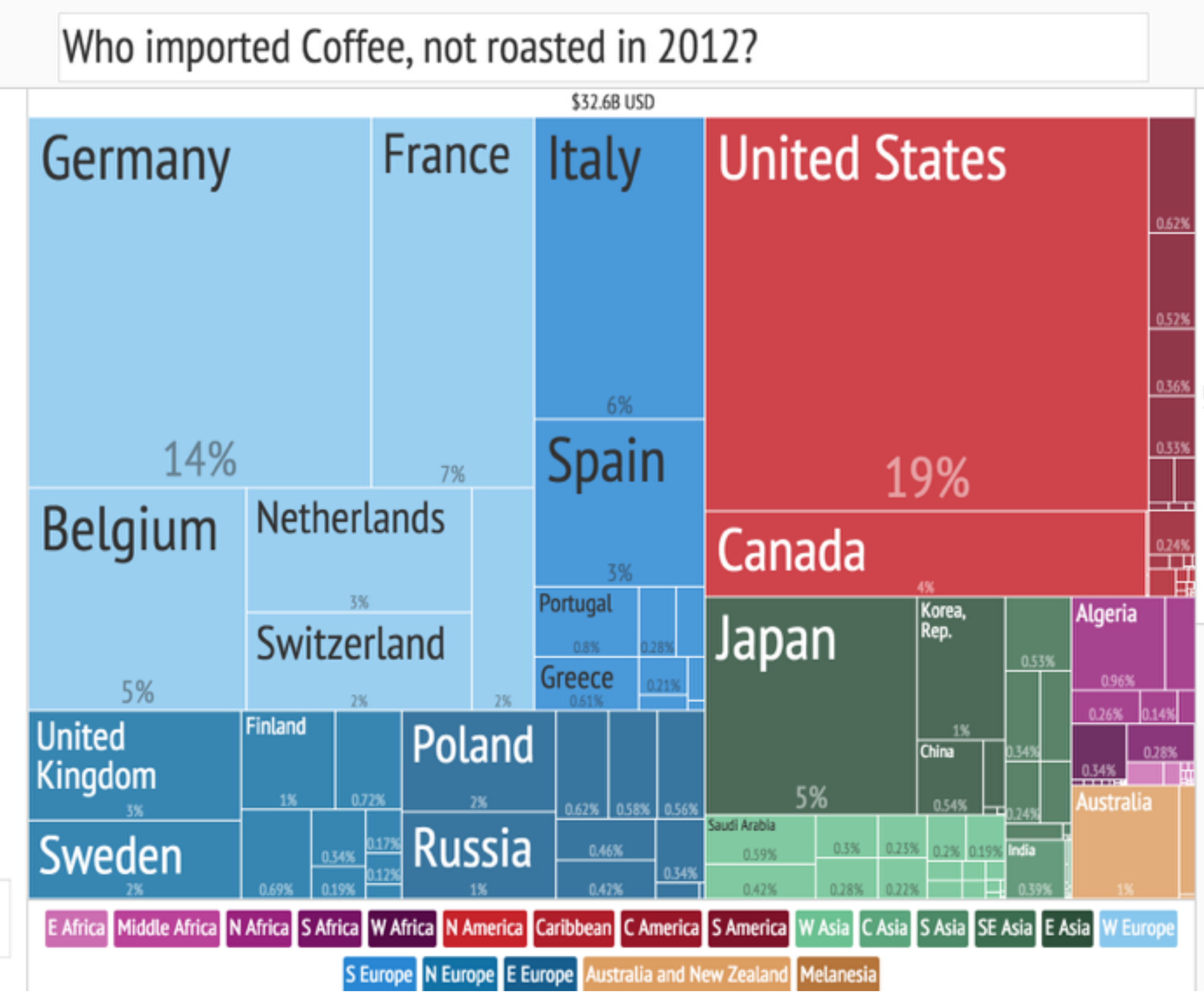
Visualizing Graphs



Node-Link Diagram



Matrix

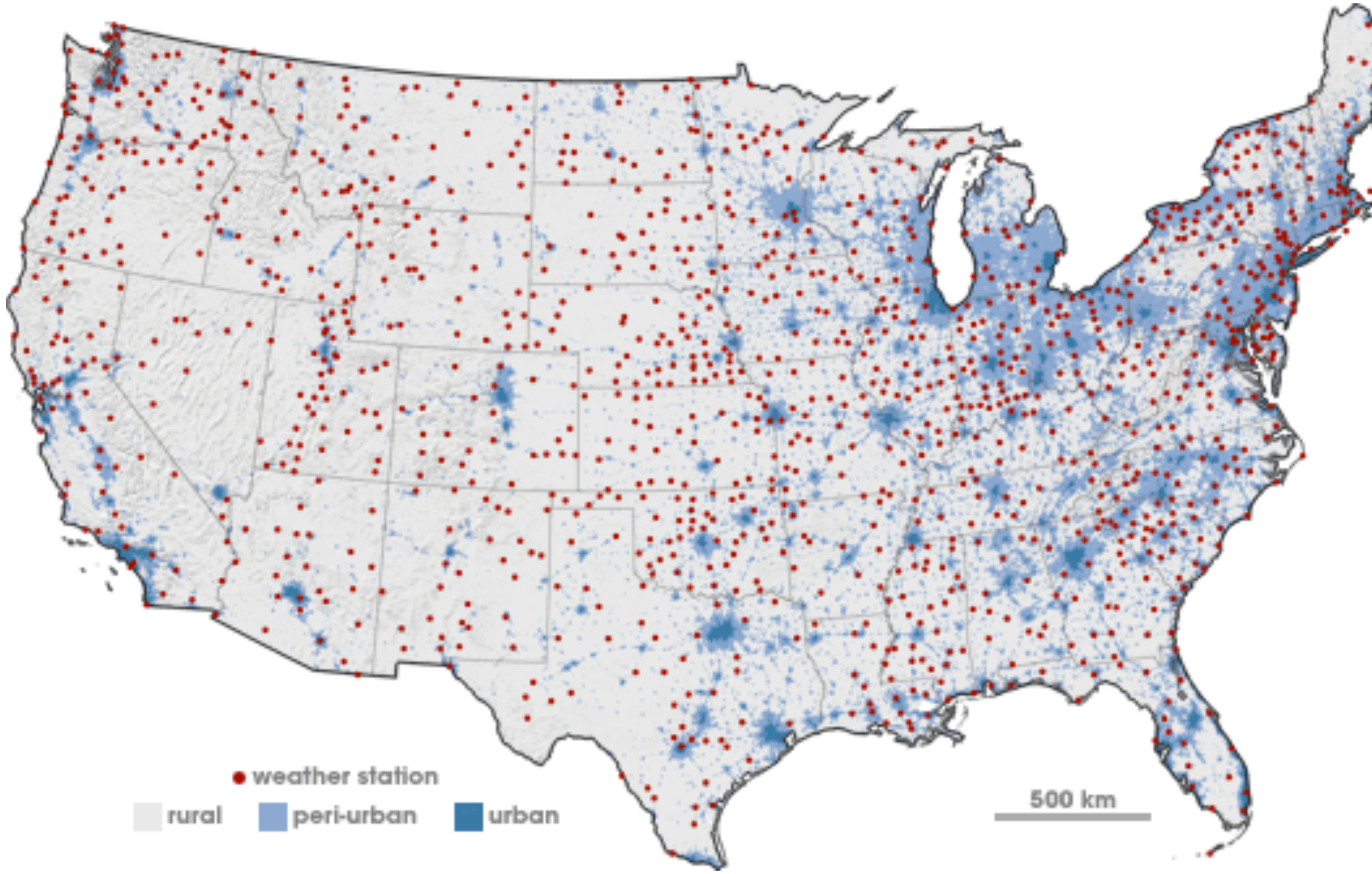


Treemap (Implicit Tree Visualization)

More in Lecture 10: Trees & Networks

Fields

- Attribute values associated with cells
- Cell contains data from continuous domain
  - Temperature, pressure, wind velocity
- Measured or simulated
- Sampling & Interpolation
  - Signal processing & stats



Fields: Grid Types

Uniform Grid

- Geometry & topology can be computed

Rectilinear Grid

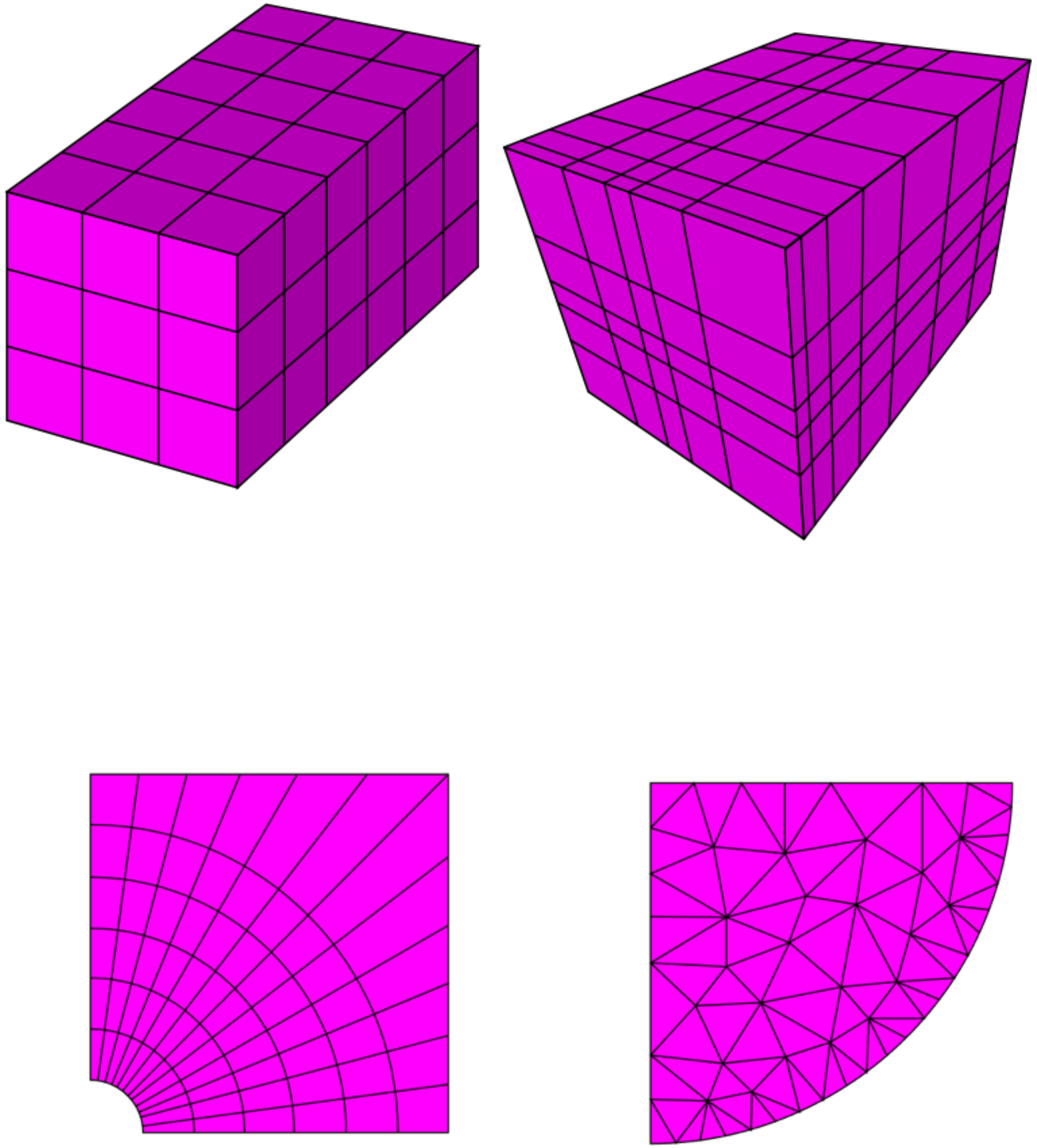
- Nonuniform sampling

Structured Grid

- allows curvilinear grids

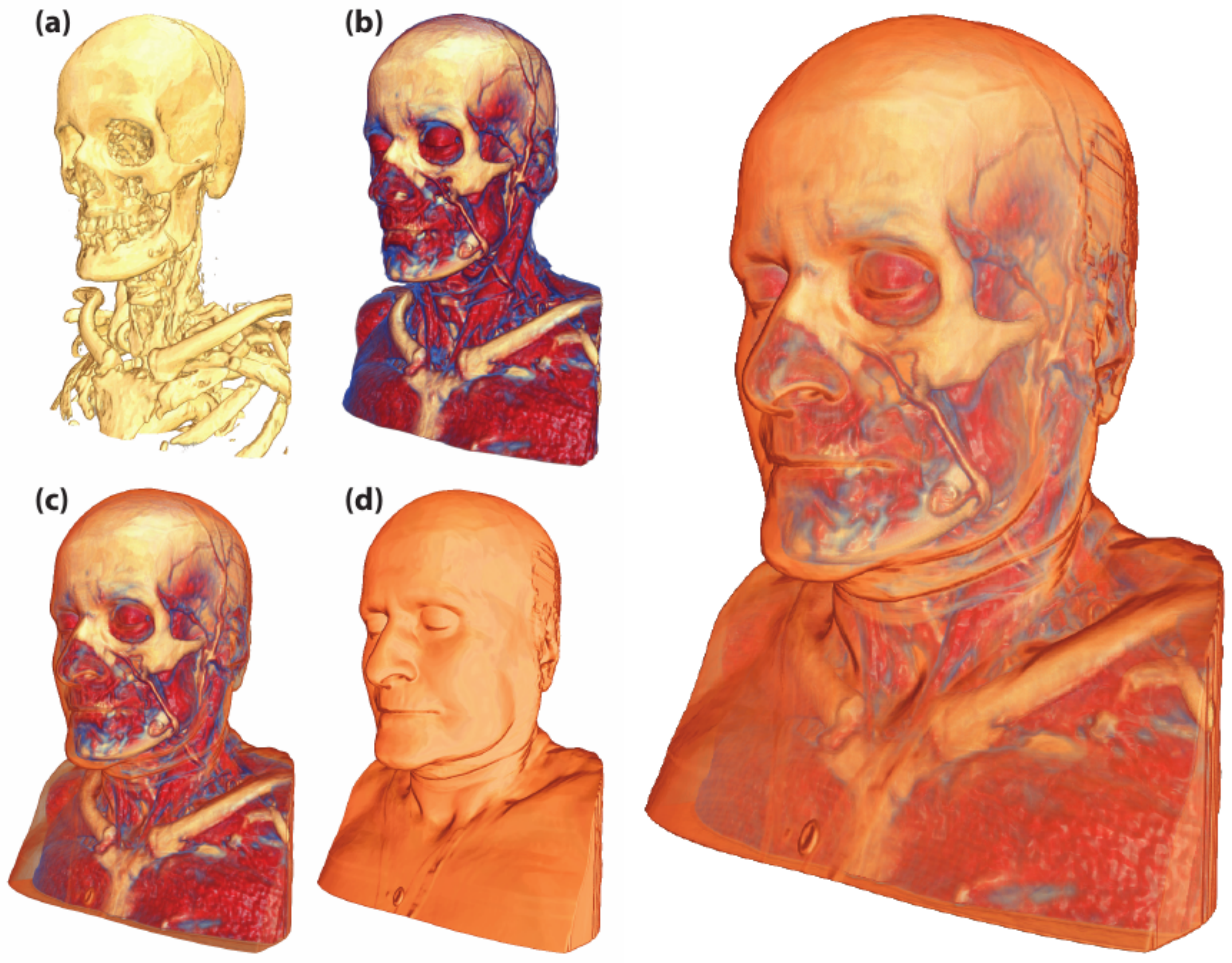
Unstructured Grid

- full flexibility, store position and connection

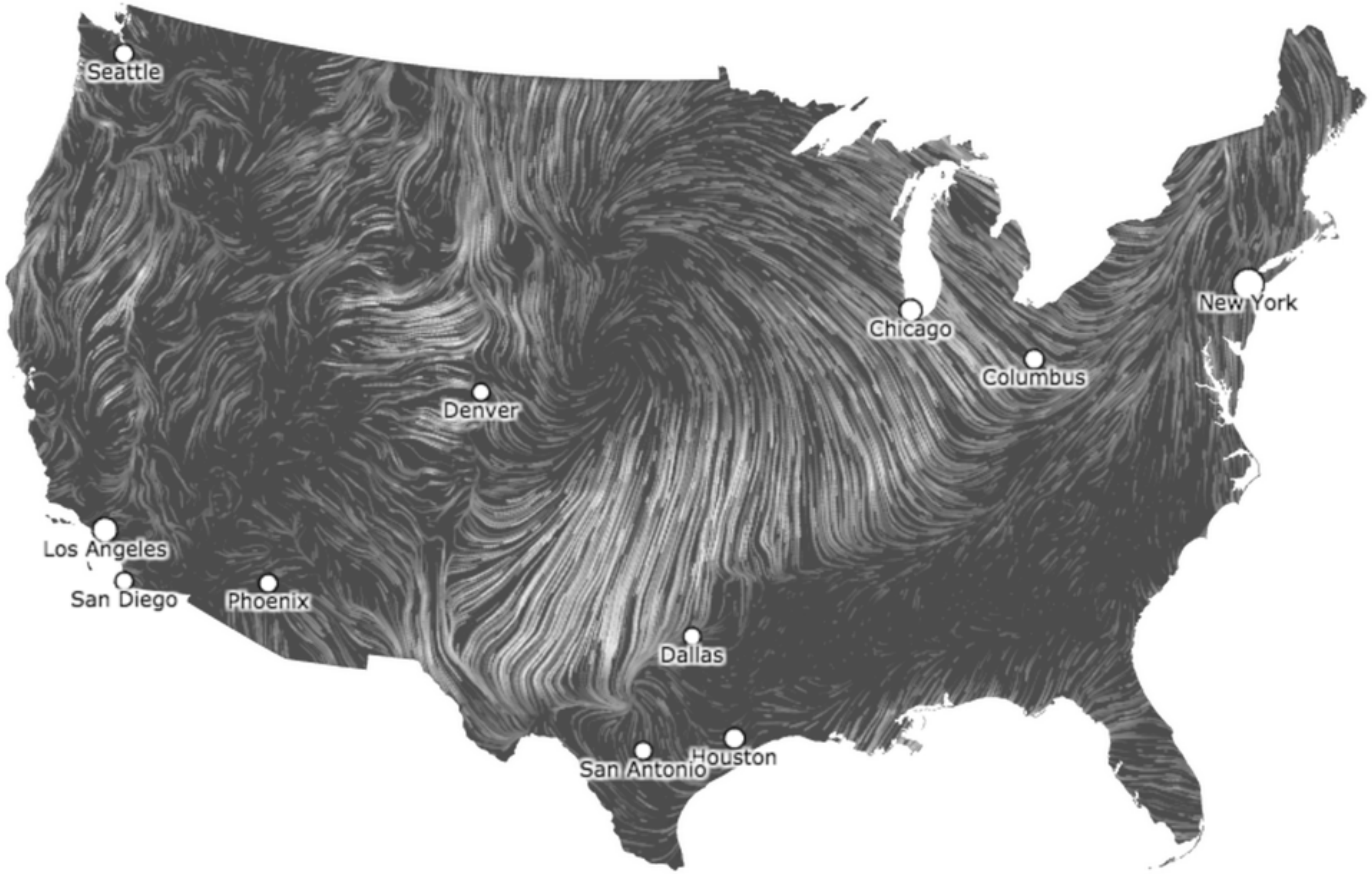


[Wikipedia]

Visualizing Fields



[Bruckner 2007]



More in Lecture 12: Maps & Lecture 15: Visualizing spatial data: Volumes and Flows

Geometry

Shape of items

Explicit spatial positions

Points, lines, curves, surfaces, regions, volumes

Important in Computer Graphics, CAD, …

Not a core Vis topic

Side Note: Academic Trenches

Information Vis

“Abstract Data”

Tables, Graphs

Free to choose spatial layout

[Alex, Hendrik, Romain, Sam]

Visual Analytics

InfoVis + Stats + Machine learning

Applied Work

Funding buzzword

Scientific Vis

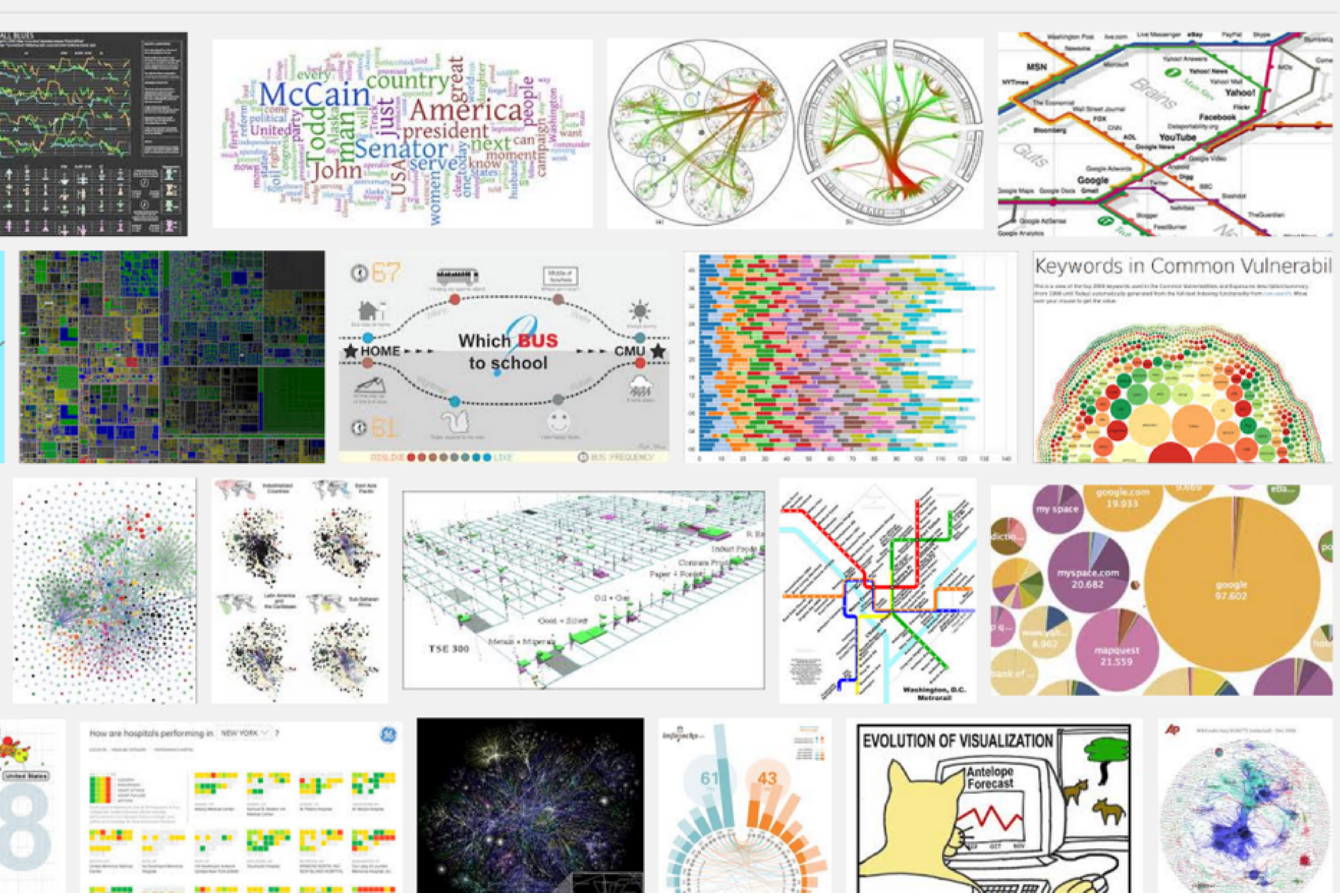
“Spatial Data” (Fields)

Not free to choose spatial layout

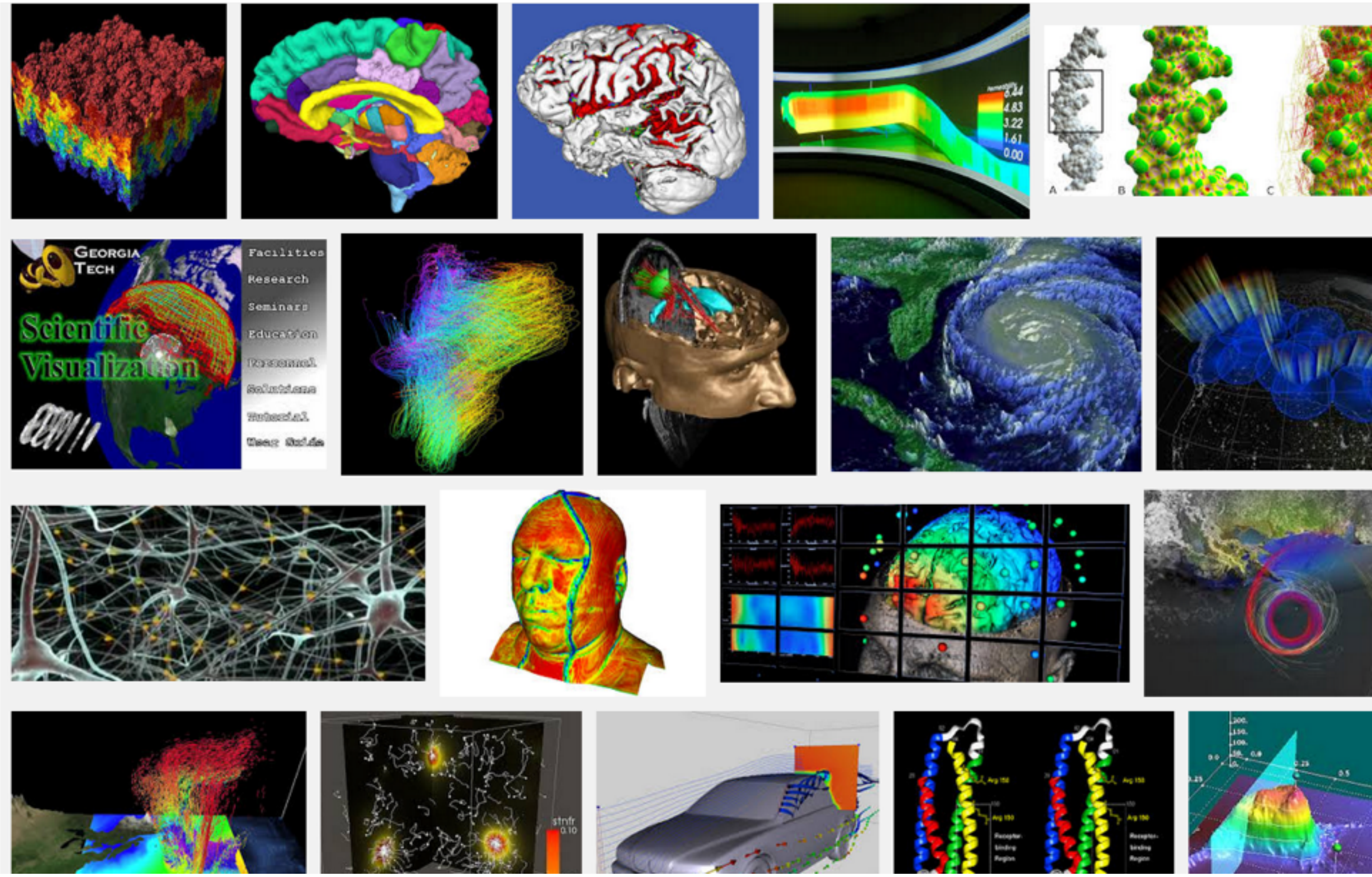
Find best way to depict reality

[Johanna, Daniel]

InfoVis or SciVis?



InfoVis: White Background



SciVis: Black Background

Other Collections

Sets

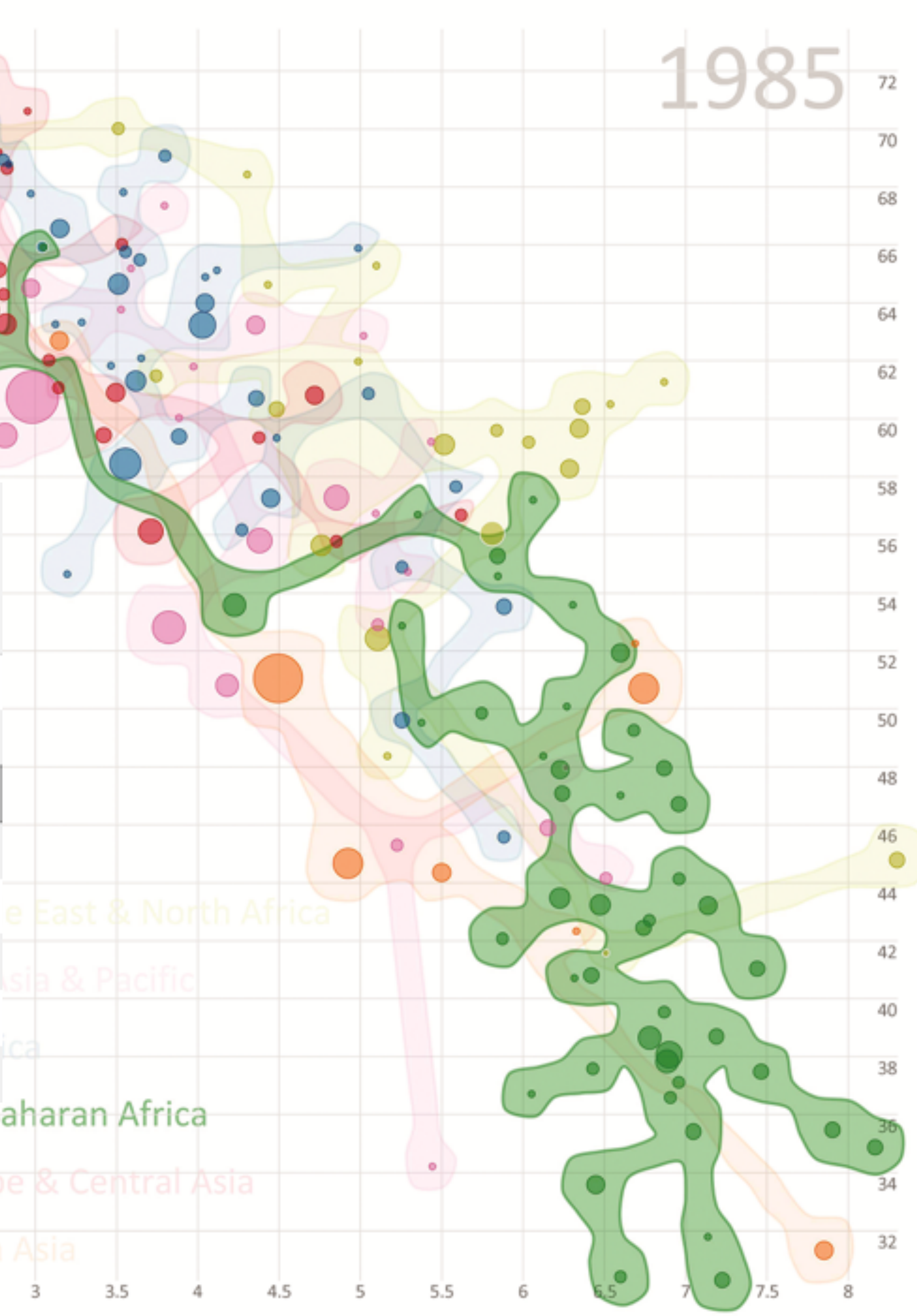
- Unique items, unordered

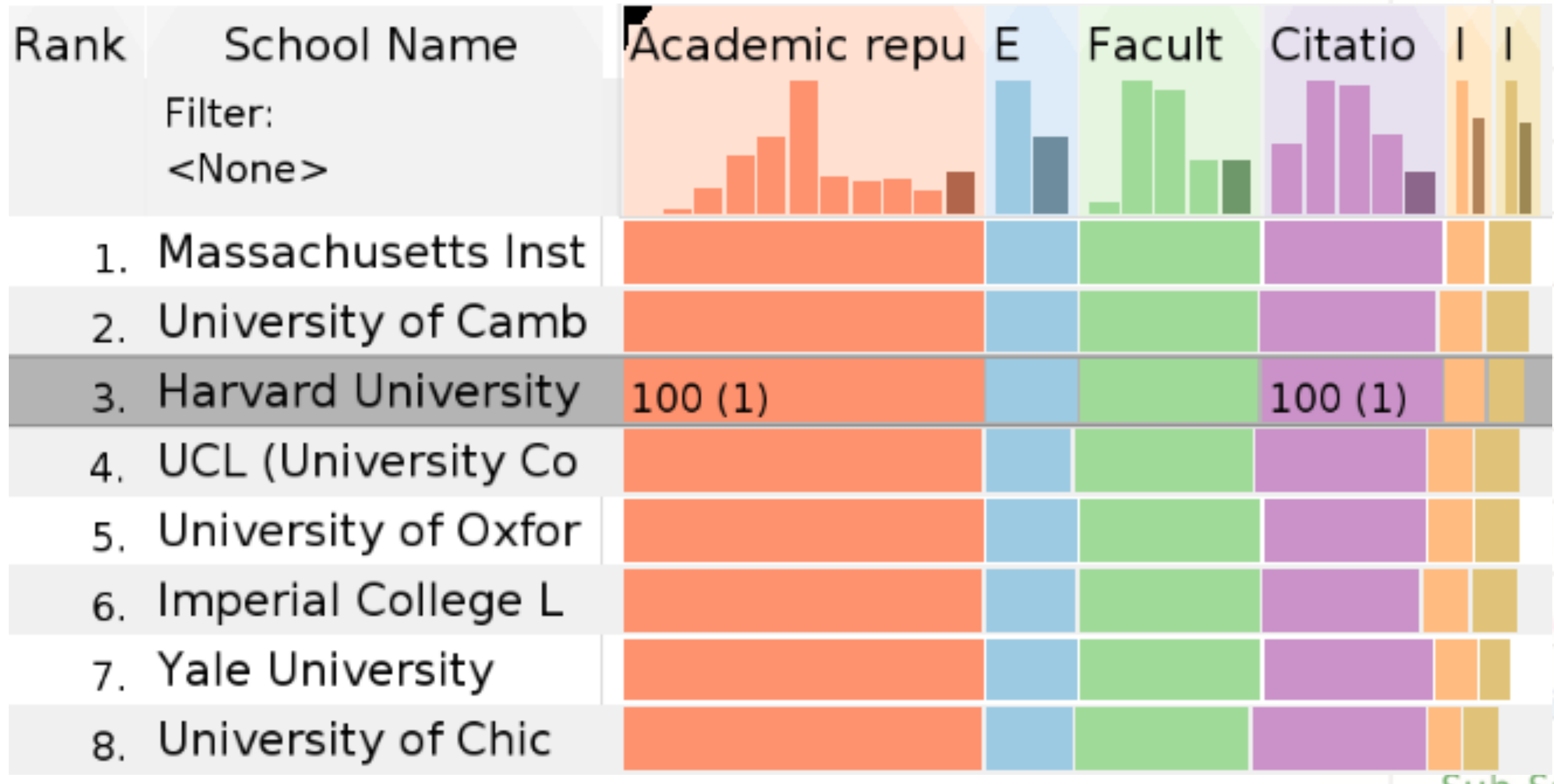
Lists

- Ordered, duplicates allowed

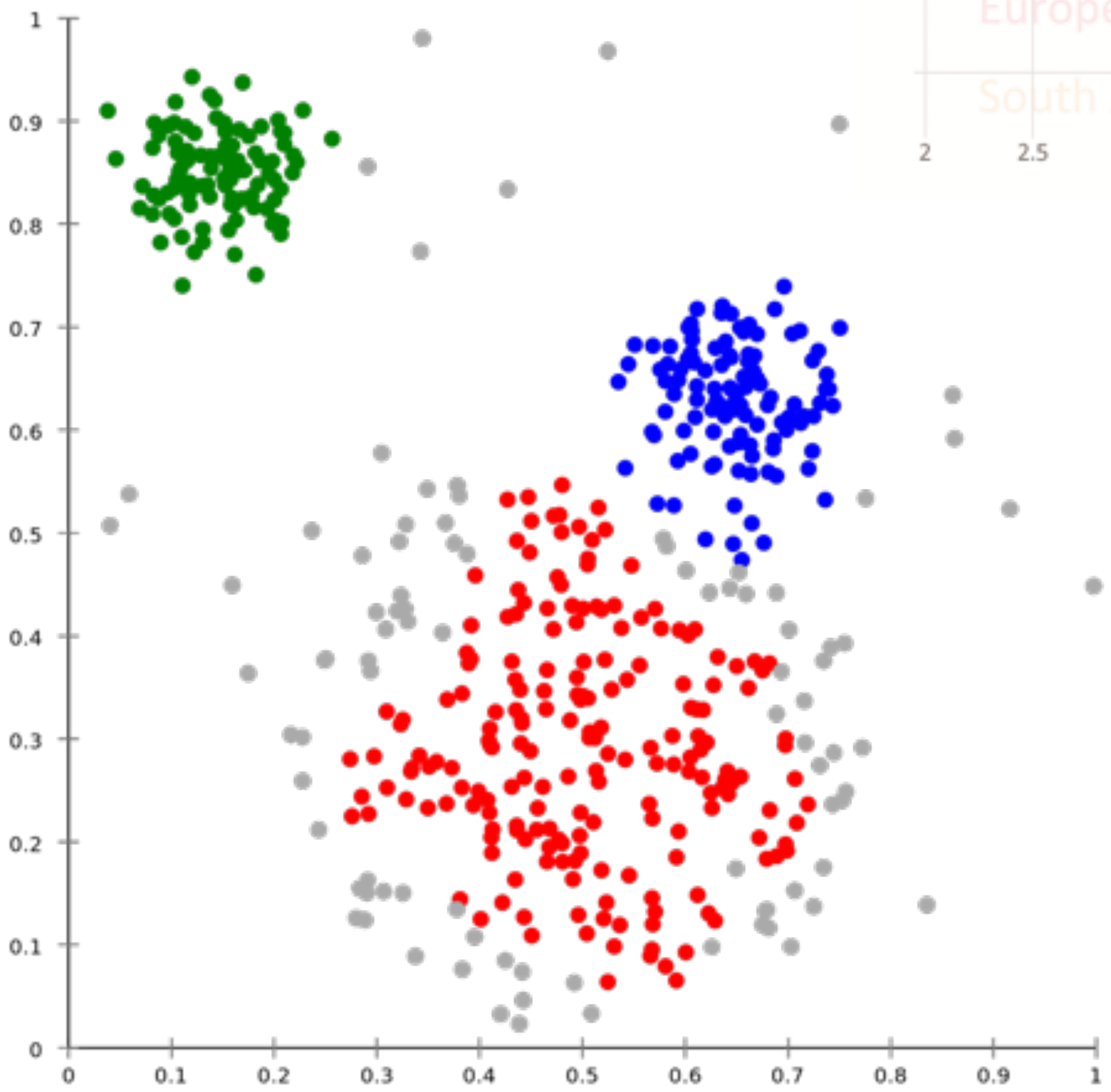
Clusters

- Groups of similar items





Rank	School Name Filter: <None>	Academic repu	E	Facult	Citatio	I	I
1.	Massachusetts Inst						
2.	University of Camb						
3.	Harvard University	100 (1)			100 (1)		
4.	UCL (University Co						
5.	University of Oxfor						
6.	Imperial College L						
7.	Yale University						
8.	University of Chic						



Attribute Types

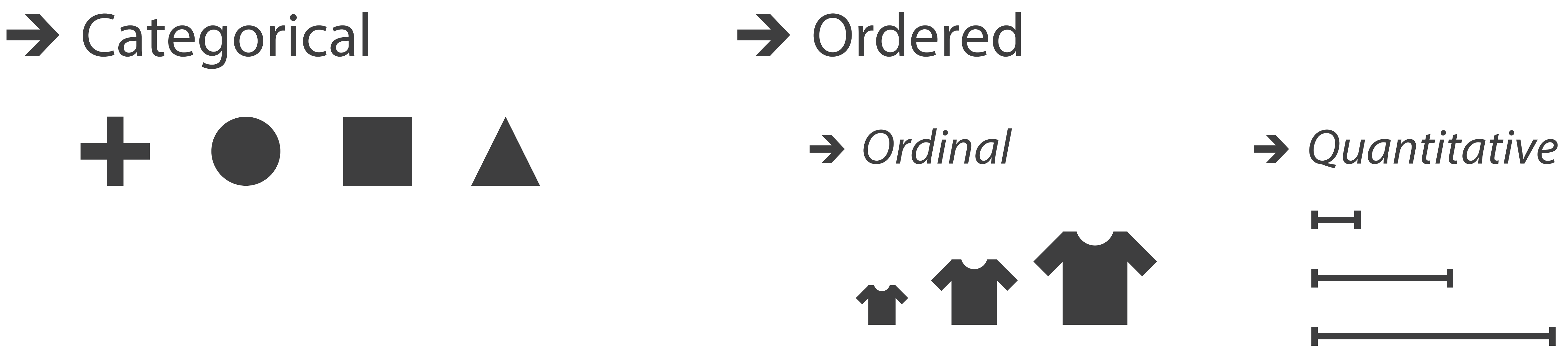
Which classes of values & measurements are there?

Categorical (nominal)

- Compare equality
- *Fruit, Gender, Movie Genres, File Types*

Ordered

- Ordinal
  - Great/Less than defined
  - *Shirt size, Rankings*
- Quantitative
  - Arithmetic possible
  - *Length, Weight, Count*



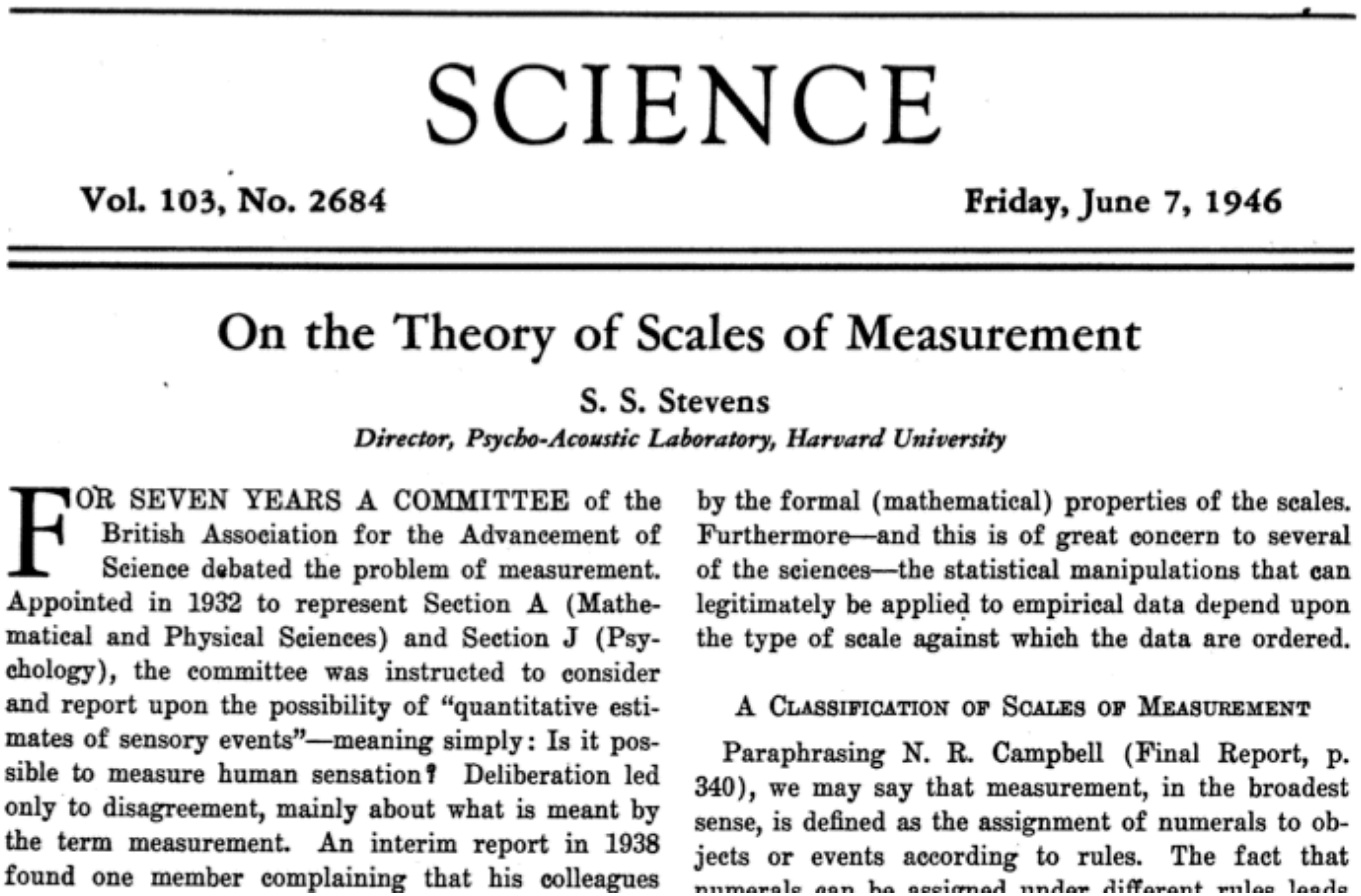
Quantitative Data Types

Interval (arbitrary zero)

- Dates: Jan 19; Location: (Lat, Long)
- Cannot compare directly. Temp in C & F
- Only differences (i.e., intervals) can be compared

Ratio (true zero)

- zero: there is nothing of the measured entity observed
- Measurements: Length, Mass
- Can measure ratios & proportions



SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

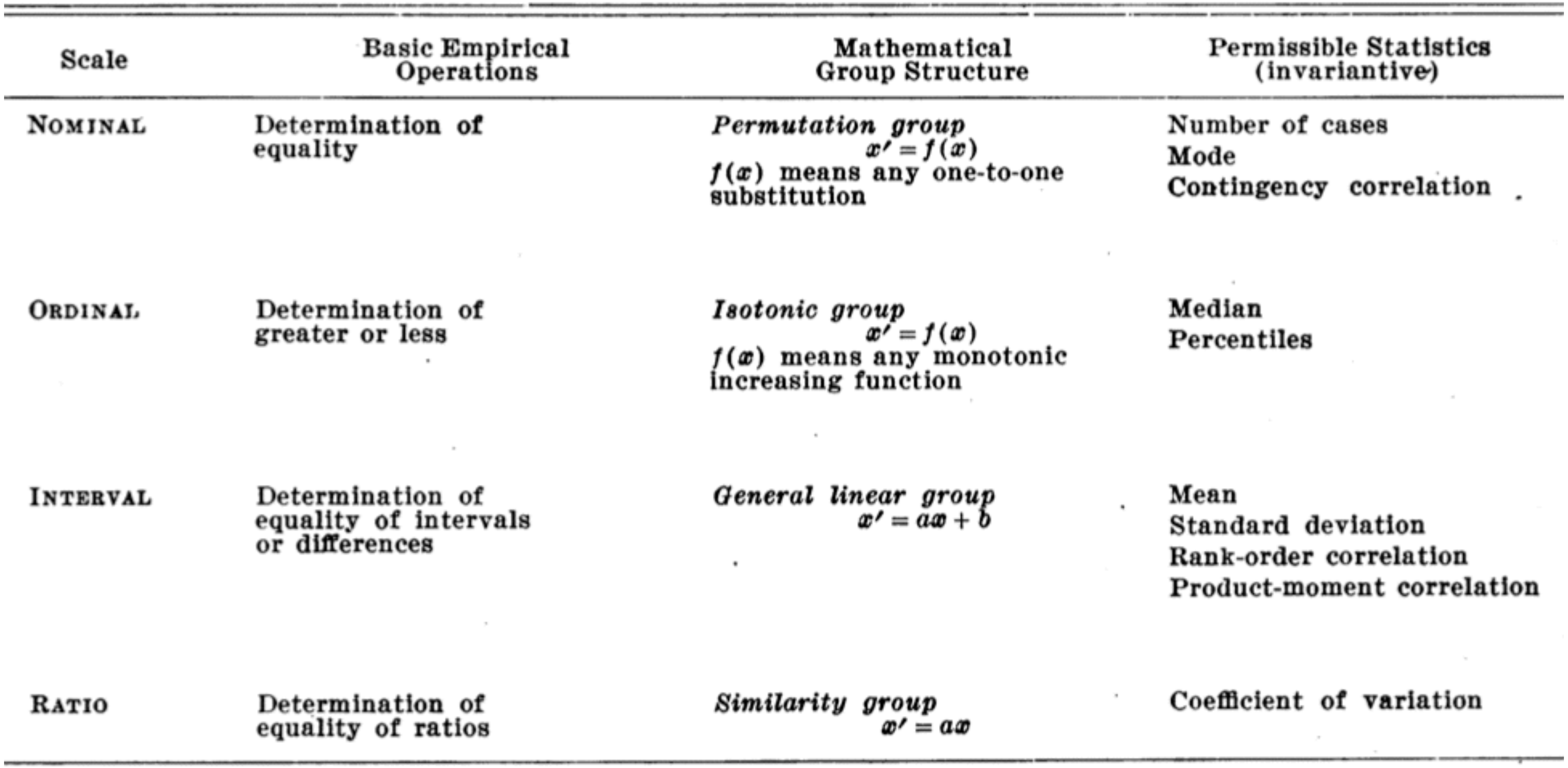
S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads



Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	*Permutation group* $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	*Isotonic group* $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	*General linear group* $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	*Similarity group* $x' = ax$	Coefficient of variation

On the theory of scales and measurements [S. Stevens, 46]

Data Types

Nominal (labels)

Operations: =, ≠

Ordinal (ordered)

Operations: =, ≠, >, <

Interval (location of zero arbitrary)

Operations: =, ≠, >, <, +, − (distance)

Ratio (zero fixed)

Operations: =, ≠, >, <, +, −,×, ÷ (proportions)

On the theory of scales and measurements [S. Stevens, 46]

Sequential & Diverging Data

Sequential:

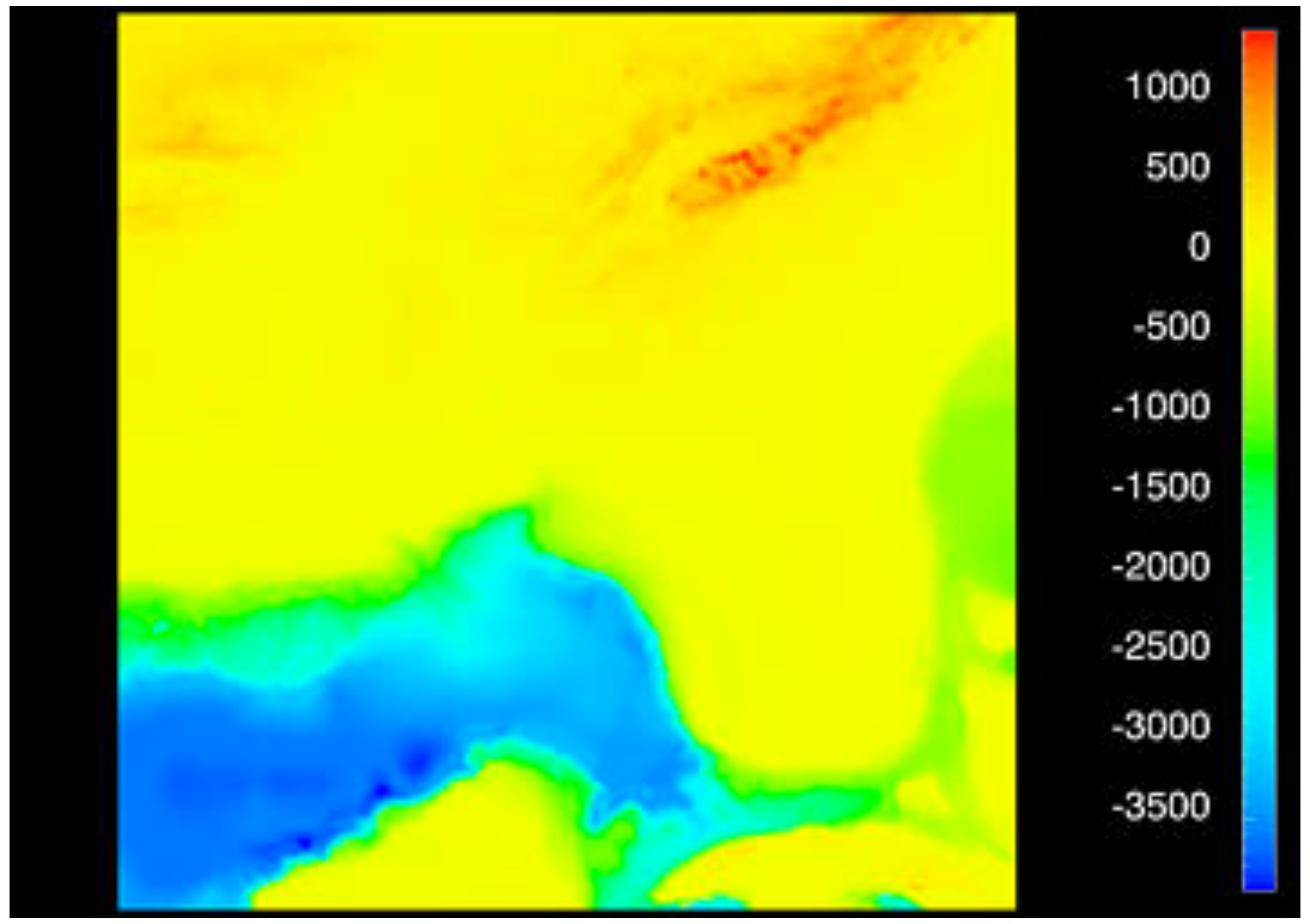
homogeneous from min to max

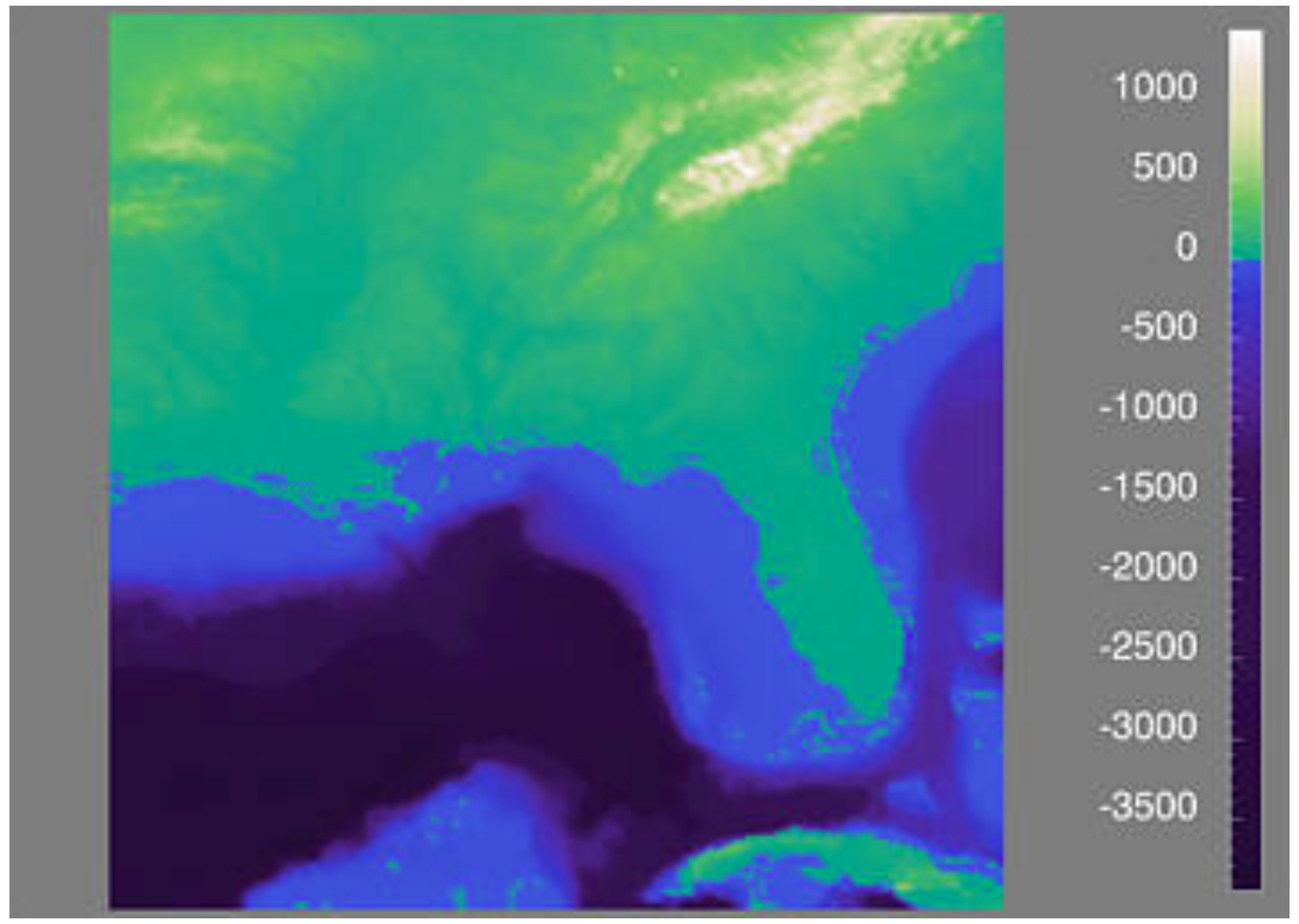
# people in countries

Diverging:

two or multiple sequences that meet

Elevation dataset: above sea level & below sea level





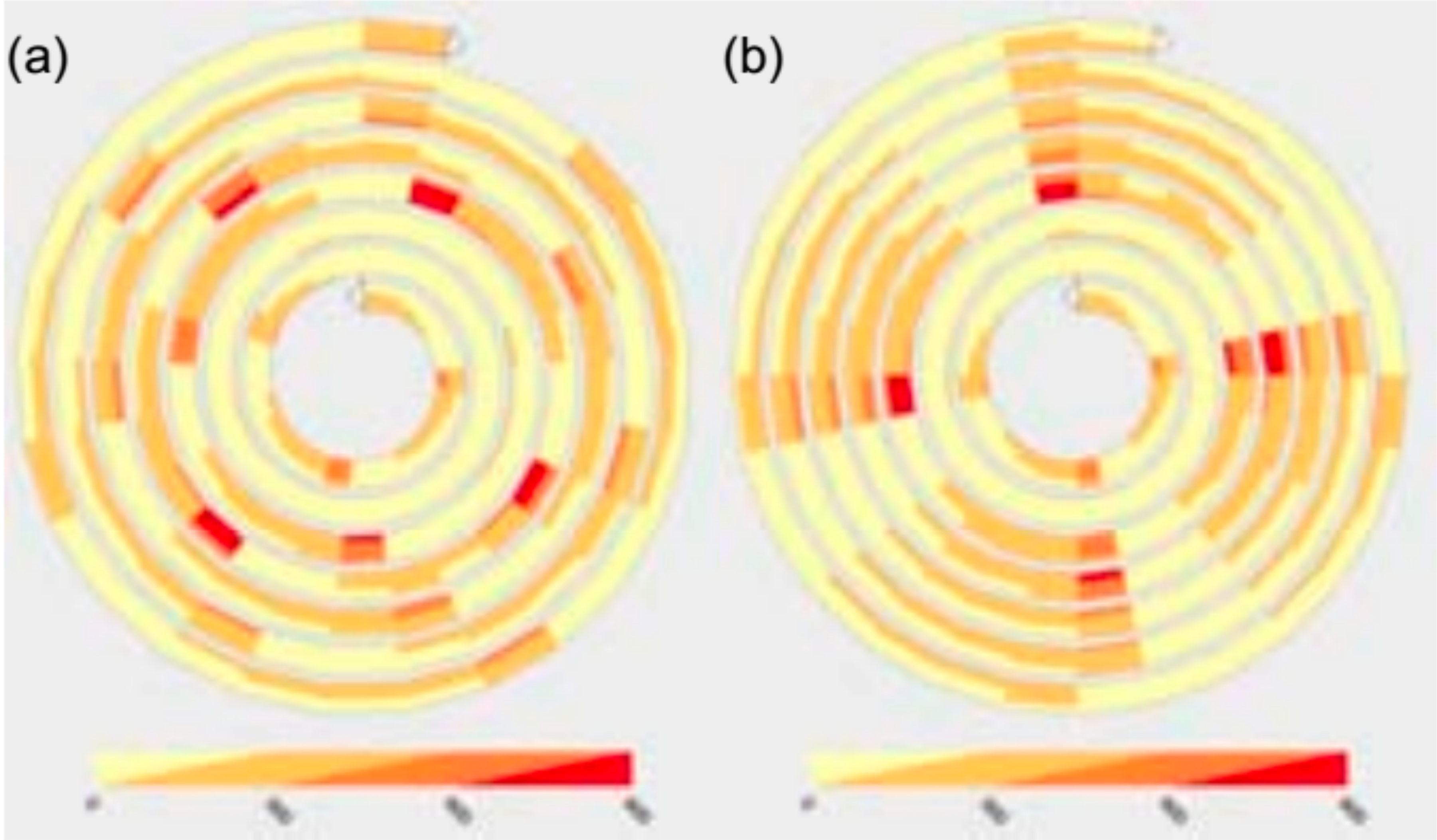
Other Structure

Cyclic data

- time (hours, week, month, year)

Aggregation

- might be patterns on multiple levels



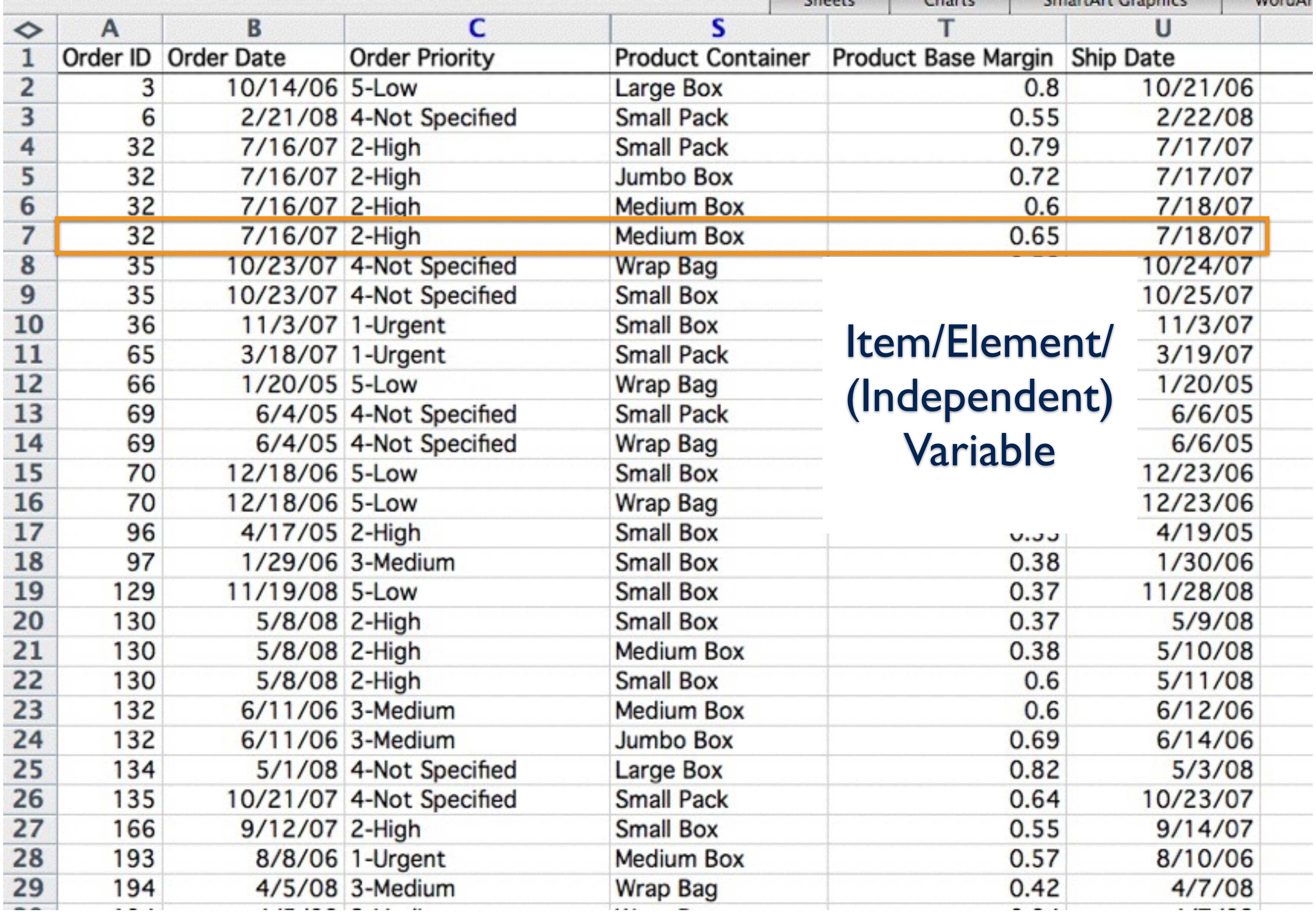
Respiratory disease cases.
Left: 25 day pattern
Right: 28 day pattern
[Tominski 2008]

Sessions
5,000
2,500
April 2013
July 2013
October 2013
January 2014
April 2014
July 2014
October 2014
January 2015

Weekly use of CS 171 website.

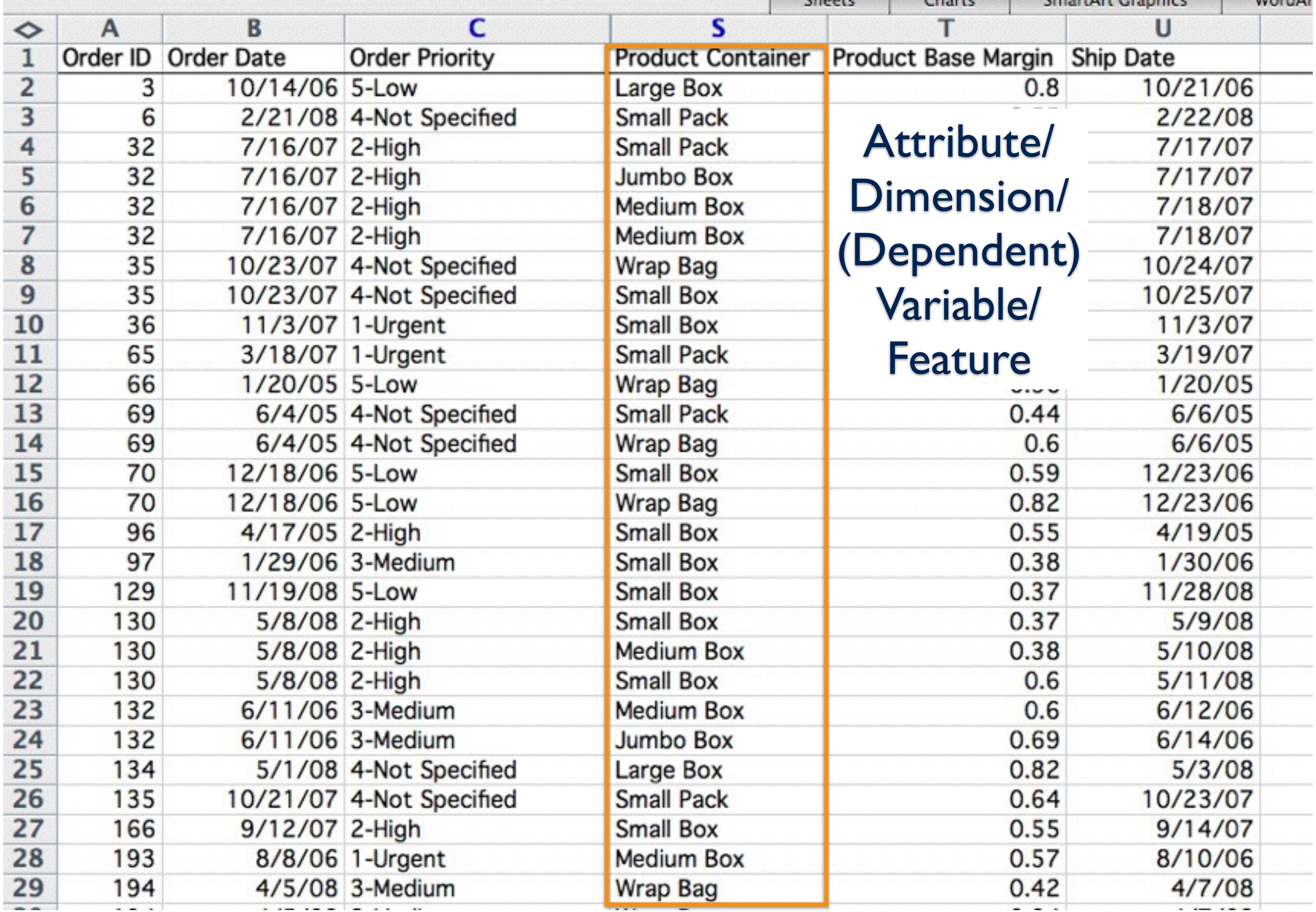
Sessions
1,000
500
February 2014
March 2014
April 2014
May 2014
June 2014
July 2014
August 2014
Septemb...

Daily use of CS 171 website.



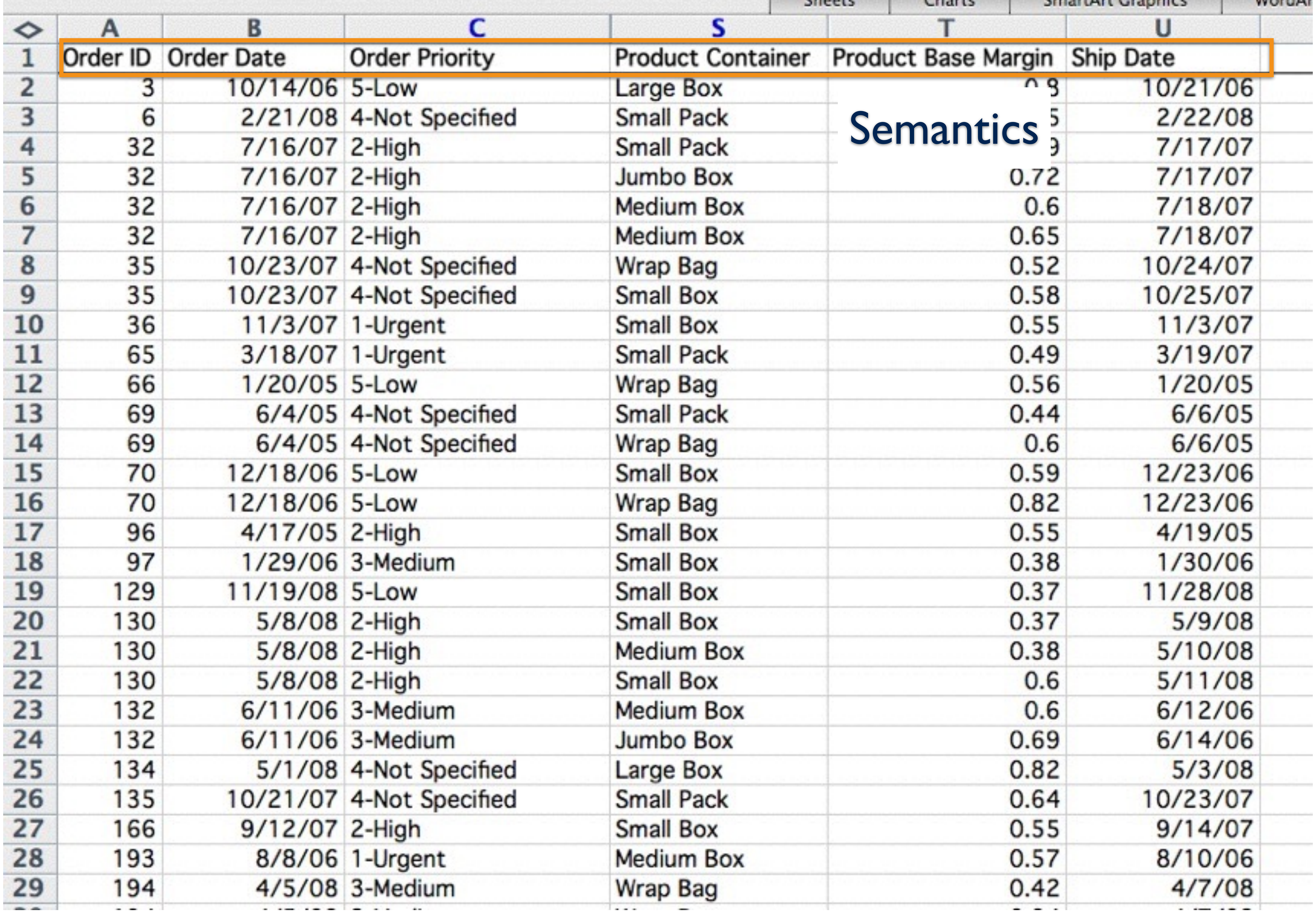
	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	[illegible]	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	[illegible]	10/25/07
10	36	11/3/07	1-Urgent	Small Box	[illegible]	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	[illegible]	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	[illegible]	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	[illegible]	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	[illegible]	6/6/05
15	70	12/18/06	5-Low	Small Box	[illegible]	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	[illegible]	12/23/06
17	96	4/17/05	2-High	Small Box	[illegible]	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Item/Element/
(Independent)
Variable



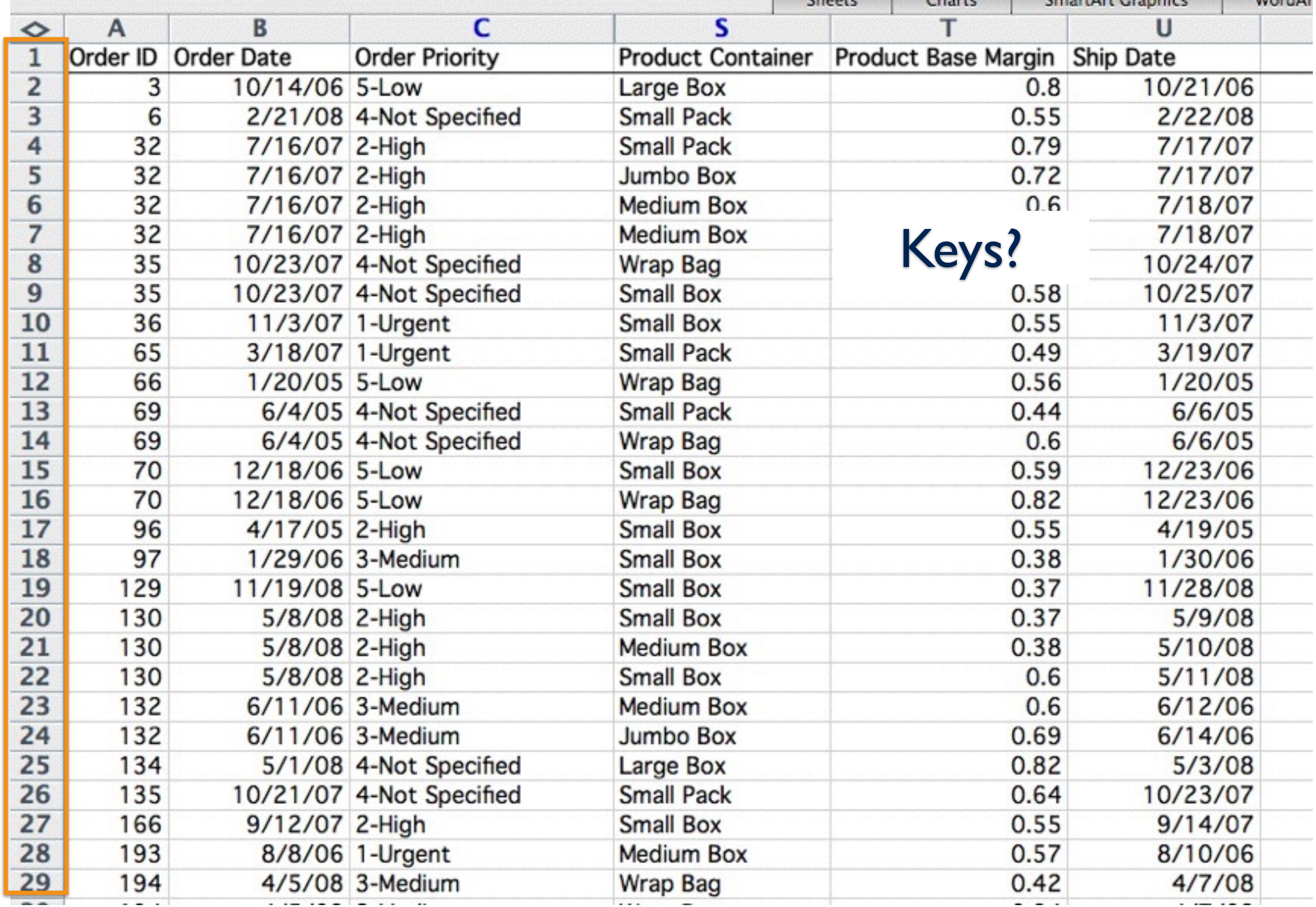
	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	[illegible]	2/22/08
4	32	7/16/07	2-High	Small Pack	[illegible]	7/17/07
5	32	7/16/07	2-High	Jumbo Box	[illegible]	7/17/07
6	32	7/16/07	2-High	Medium Box	[illegible]	7/18/07
7	32	7/16/07	2-High	Medium Box	[illegible]	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	[illegible]	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	[illegible]	10/25/07
10	36	11/3/07	1-Urgent	Small Box	[illegible]	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	[illegible]	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	[illegible]	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Attribute/
Dimension/
(Dependent)
Variable/
Feature



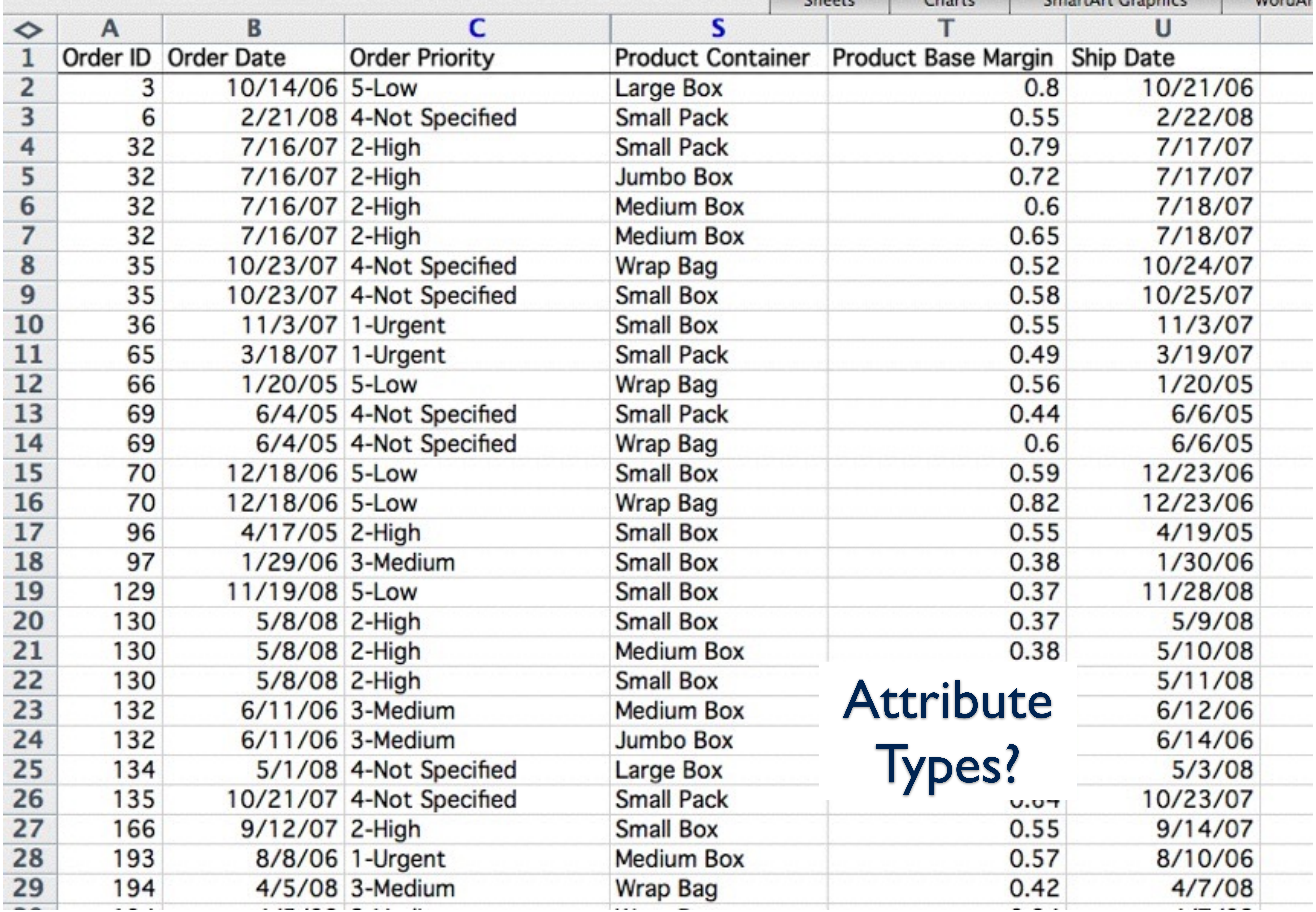
Semantics

◇	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	[illegible]	2/22/08
4	32	7/16/07	2-High	Small Pack	[illegible]	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08



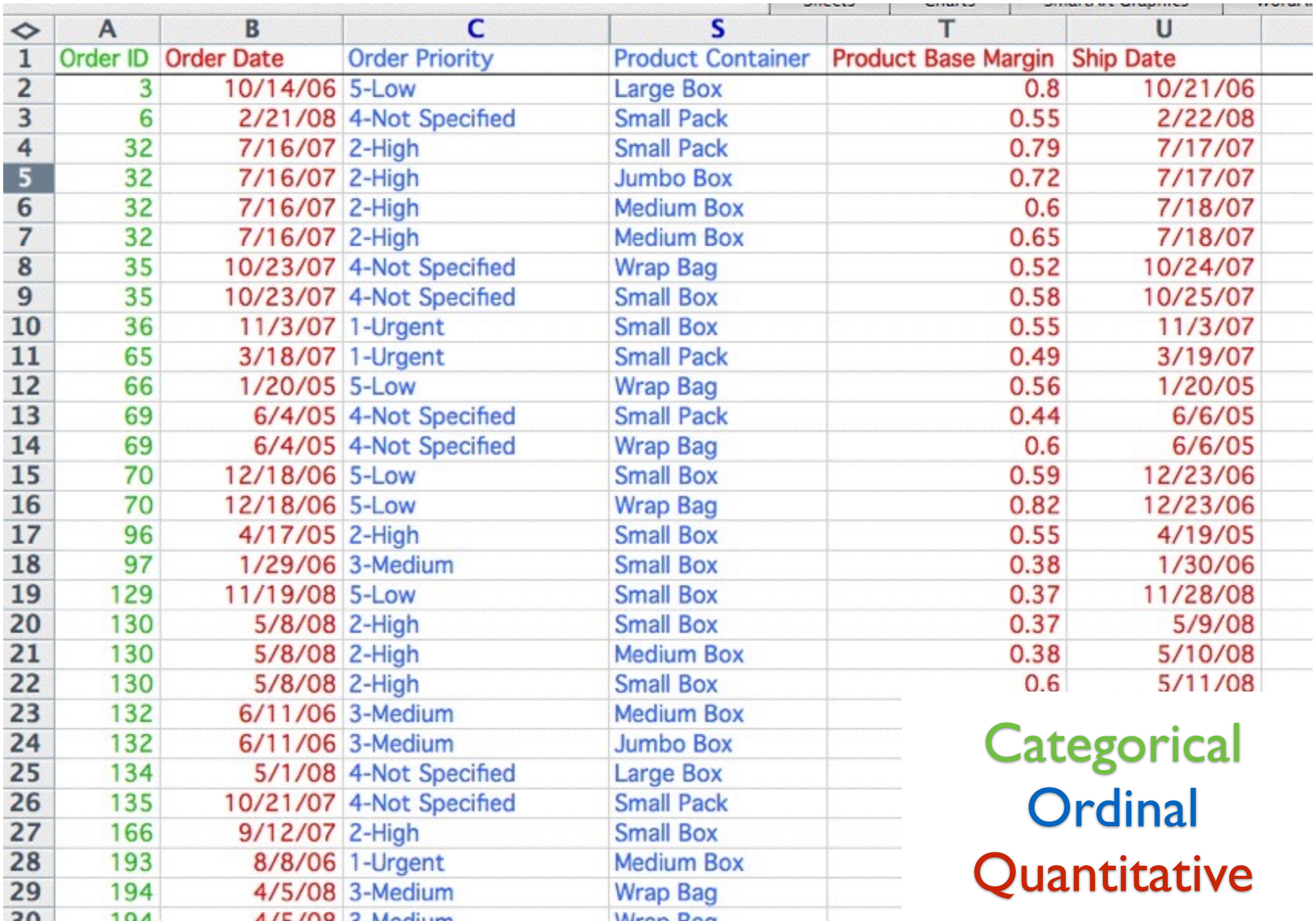
◇	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box		7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag		10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Keys?



◇	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	[illegible]	5/11/08
23	132	6/11/06	3-Medium	Medium Box	[illegible]	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	[illegible]	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	[illegible]	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	[illegible]	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Attribute Types?



◇	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box		
24	132	6/11/06	3-Medium	Jumbo Box		
25	134	5/1/08	4-Not Specified	Large Box		
26	135	10/21/07	4-Not Specified	Small Pack		
27	166	9/12/07	2-High	Small Box		
28	193	8/8/06	1-Urgent	Medium Box		
29	194	4/5/08	3-Medium	Wrap Bag		
30	194	4/5/08	3-Medium	Wrap Bag		

Categorical

Ordinal

Quantitative

Data vs. Conceptual Model

Data Model: Low-level description of the data

- Set with operations, e.g., floats with +, -, /, *

Conceptual Model: Mental construction

- Includes semantics, supports reasoning

Data	Conceptual
1D floats	temperature
3D vector of floats	space

Data vs. Conceptual Model

From data model...

- 32.5, 54.0, -17.3, … (floats)

using conceptual model...

- Temperature

to data type

- Continuous to 4 significant digits (Q)
- Hot, warm, cold (O)
- Burned vs. Not burned (N)

Combinations, Derived Data

Networks can have attributes

Attributes have hierarchies

Data types can be transformed

Real life is complicated…