

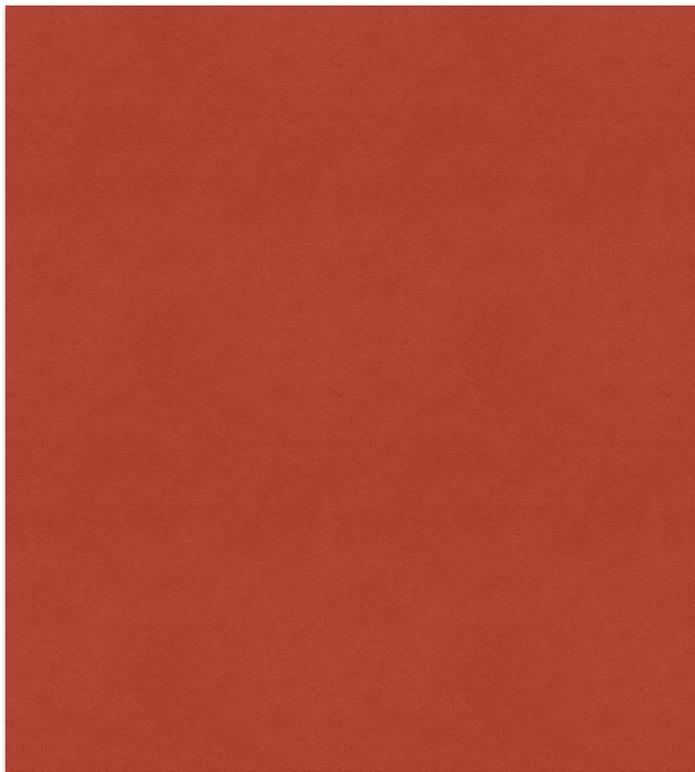
CSE 5544: Introduction to Data Visualization

Raghu Machiraju
machiraju.l@osu.edu



Quiz - which visualization ?

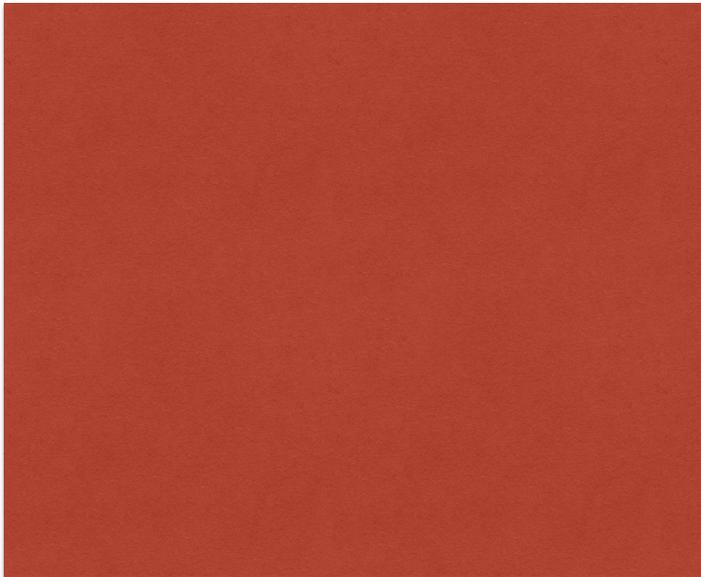
Demographics



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

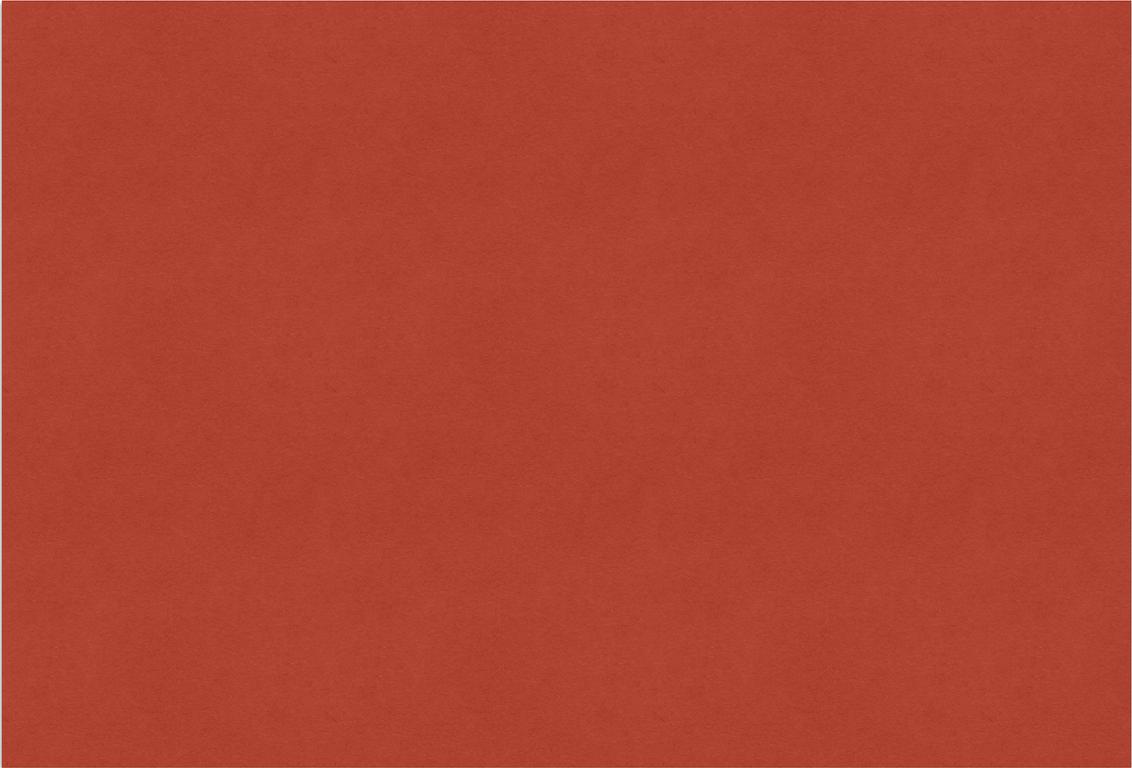
I prefer not to disclose	3	2%
Under 17	2	1%
18 to 24	115	66%
25 to 44	50	29%
45 to 64	5	3%
Over 65	0	0%

Program

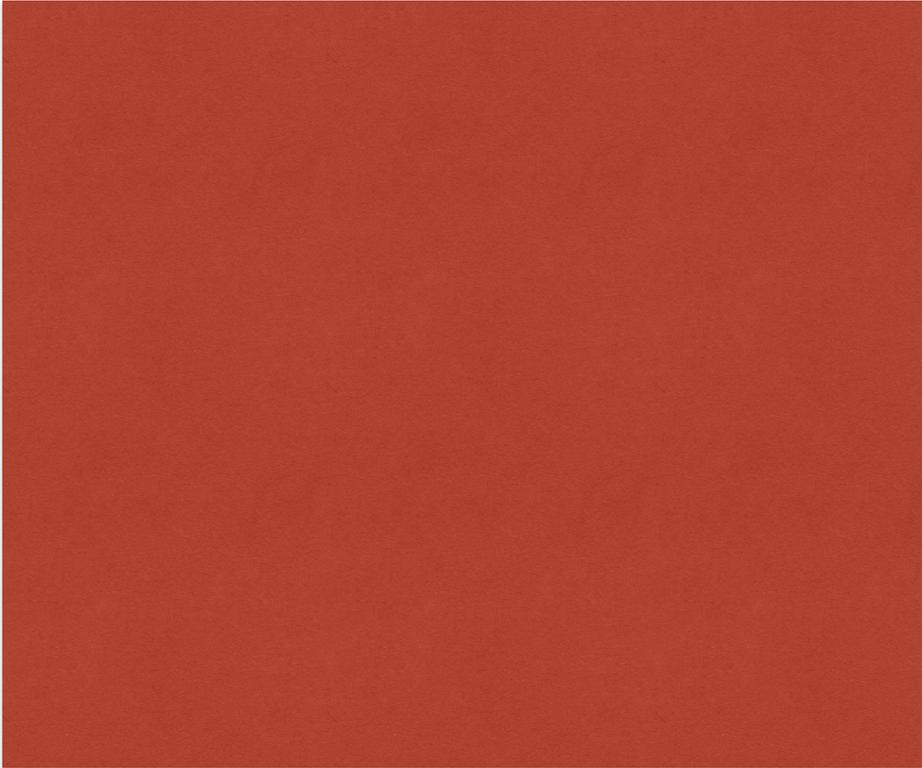


Harvard College	107	61%
Harvard Graduate Student	19	11%
Harvard Division of Continuing Education (DCE)	53	30%
MIT Undergrad	0	0%
MIT Graduate Student	4	2%

Concentrations

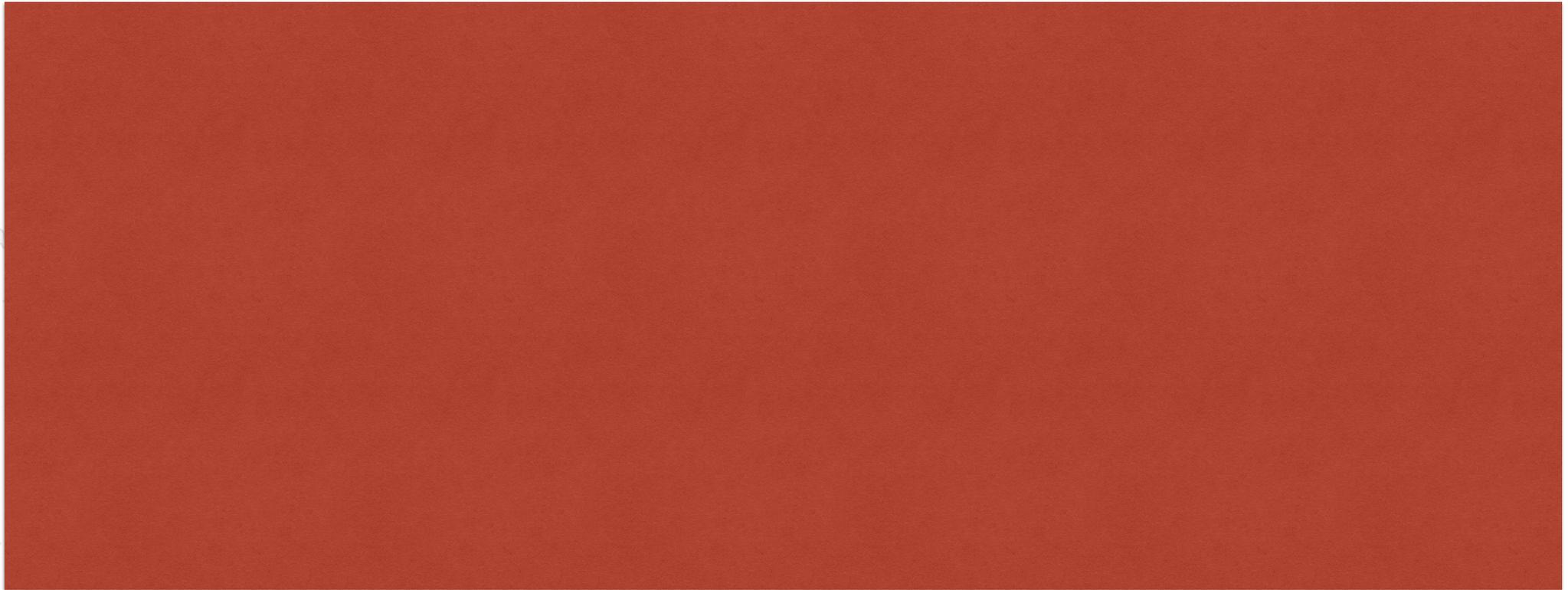


Primary

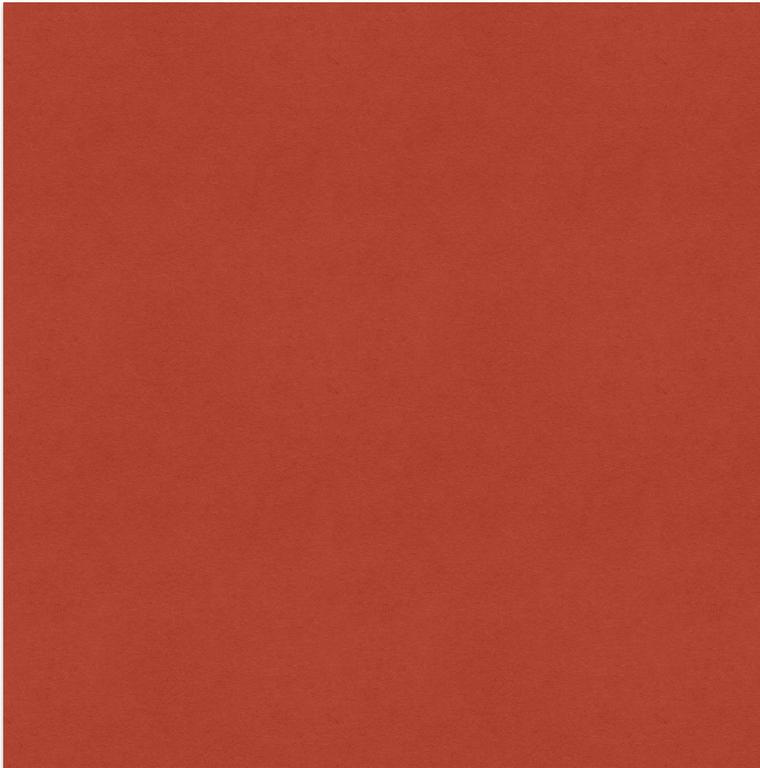


Secondary

Where you're from ?



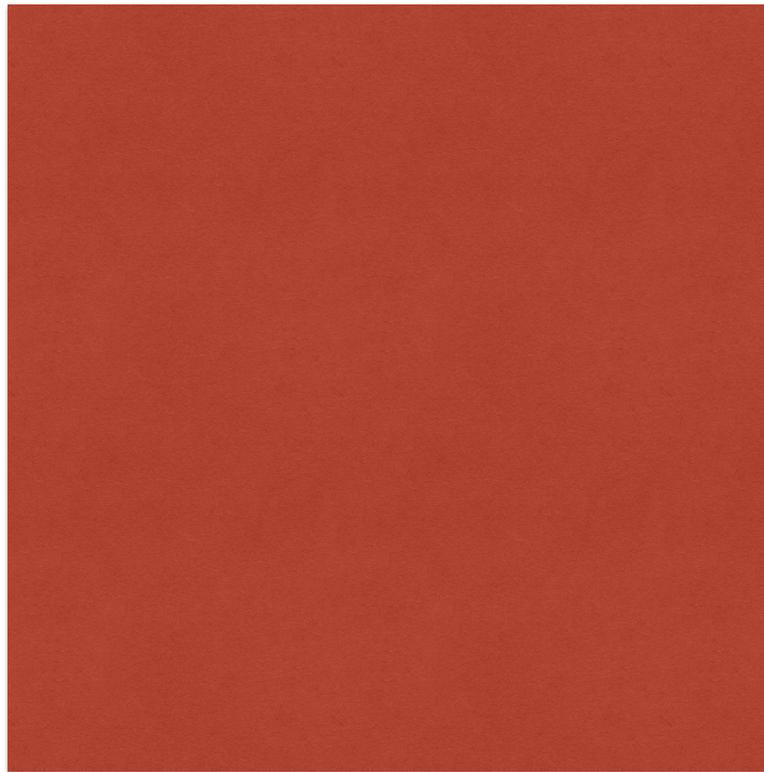
Programming Skills



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%

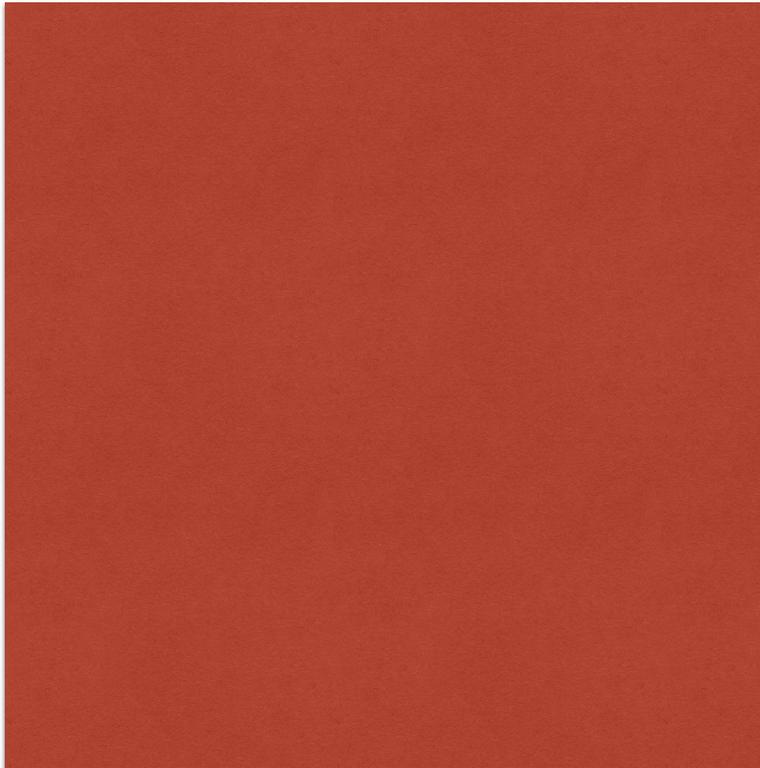
Daily	32	18%
Weekly	63	36%
Two or more times per month	43	25%
Once per month	12	7%
Less than once per month	25	14%

Primary Language



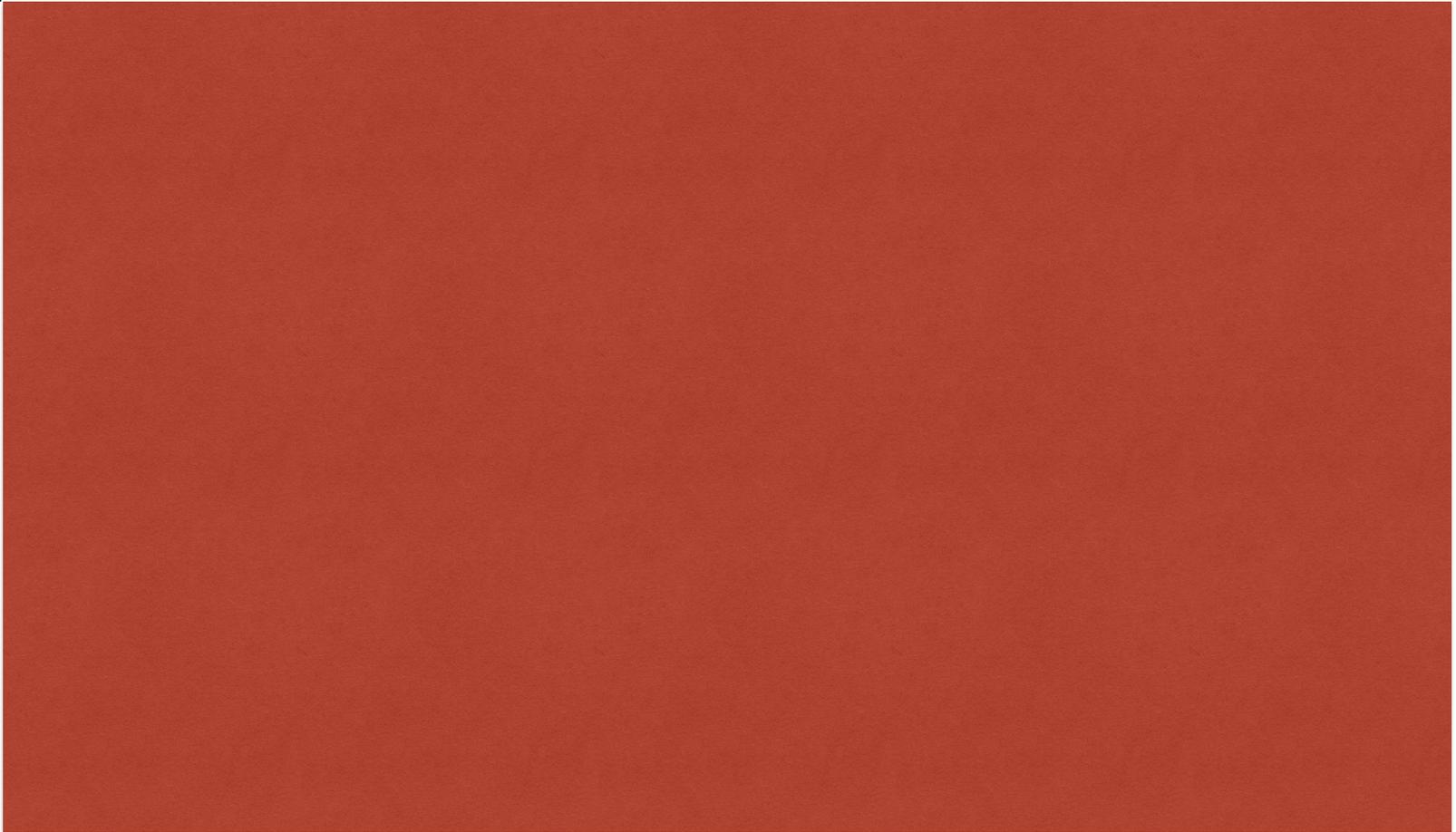
BASIC	1	1%
C	50	29%
C++	4	2%
C#	3	2%
Java	24	14%
JavaScript	15	9%
HTML / CSS	17	10%
LISP	0	0%
Perl	1	1%
PHP	9	5%
Python	42	24%
Ruby	2	1%
SQL	5	3%
VB / VBScript	2	1%

Other Languages



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%

Why take this class?



Data

Terms

Dataset Types

what can be visualized?

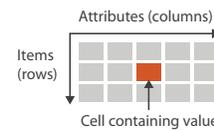
Data Types

fundamental units

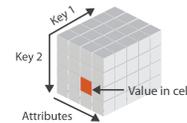
combinations make up Dataset Types

Dataset Types

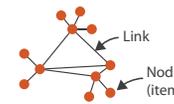
→ Tables



→ Multidimensional Table



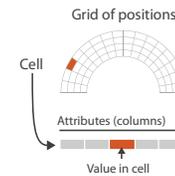
→ Networks



→ Trees



→ Fields (Continuous)



→ Geometry (Spatial)



Data Types

→ Items

→ Attributes

→ Links

→ Positions

→ Grids

Structure

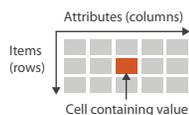
Structured Data

known data types, semantics

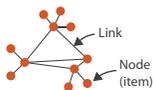
video, images

Dataset Types

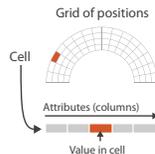
→ Tables



→ Networks



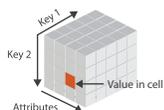
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



Unstructured Data

no predefined data model

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

video, images

Translate into structured data

Similarity processing

Natural Language Processing

Text mining (sentiment, keywords, concepts, categories)

Image Example: Graphie

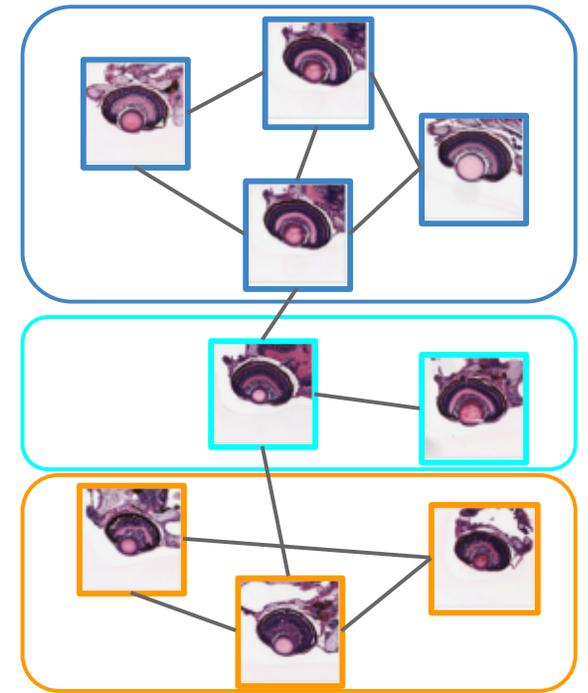
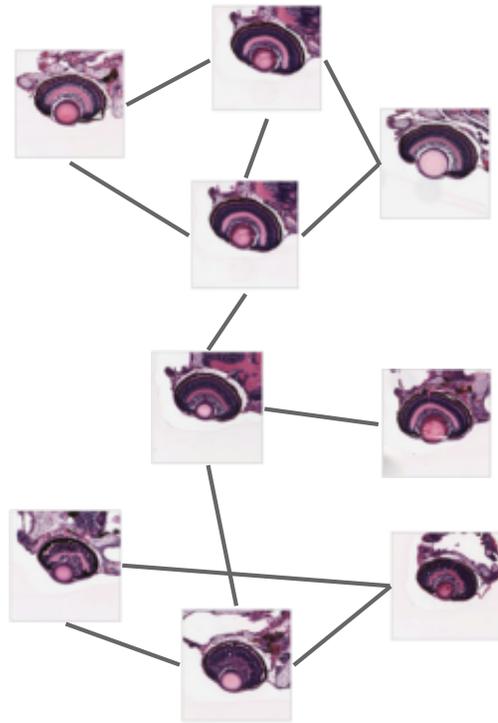
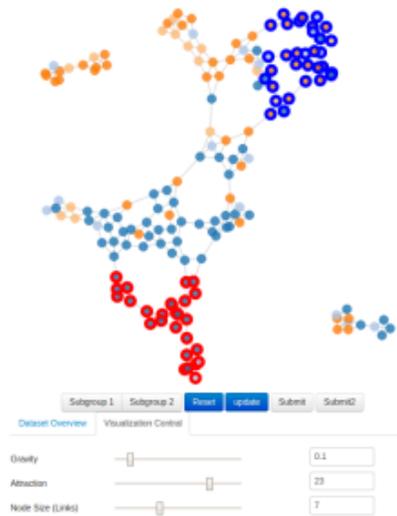


Image Example: Graphie



Example: Phrase Net

Pattern: “X’s Y”

18th & 19th century
novels

Text & Document Vis



Example:

"Is between between is and between or between between and is or between and and or or between or and between or between between and and or between or and and or between or and or?"

	between	
	↓	
is	??	between
between	??	is
and	??	or
or	??	between
between	??	and
or	??	and
or	??	or

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

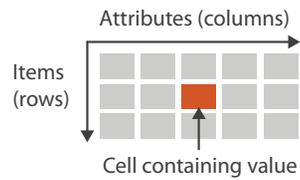
Sampling strategy for continuous data

How many Voxels in MRI scan, positions of weather stations in the US

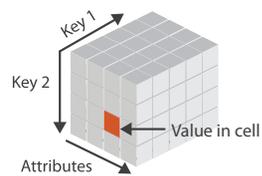
Dataset Types

→ Dataset Types

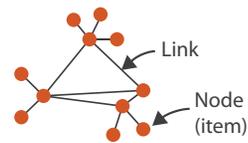
→ Tables



→ Multidimensional Table



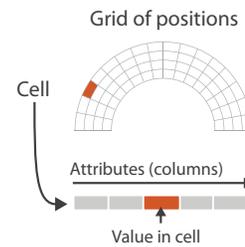
→ Networks



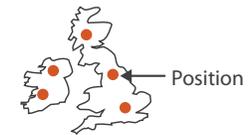
→ Trees



→ Fields (Continuous)



→ Geometry (Spatial)



Tables

Flat Table

one item per row

each column is attribute

unique (implicit) key

no duplicates

Multidimensional Table

indexing based on multiple keys

	Keys	Attributes			Values
	ID	Name	Age	Shirt Size	Favorite Fruit
	1	Amy	8	S	Apple
	2	Basil	7	S	Pear
Item	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: Genes

Keys: Patients

Data - Gene Expression-Structure Matrix

2692

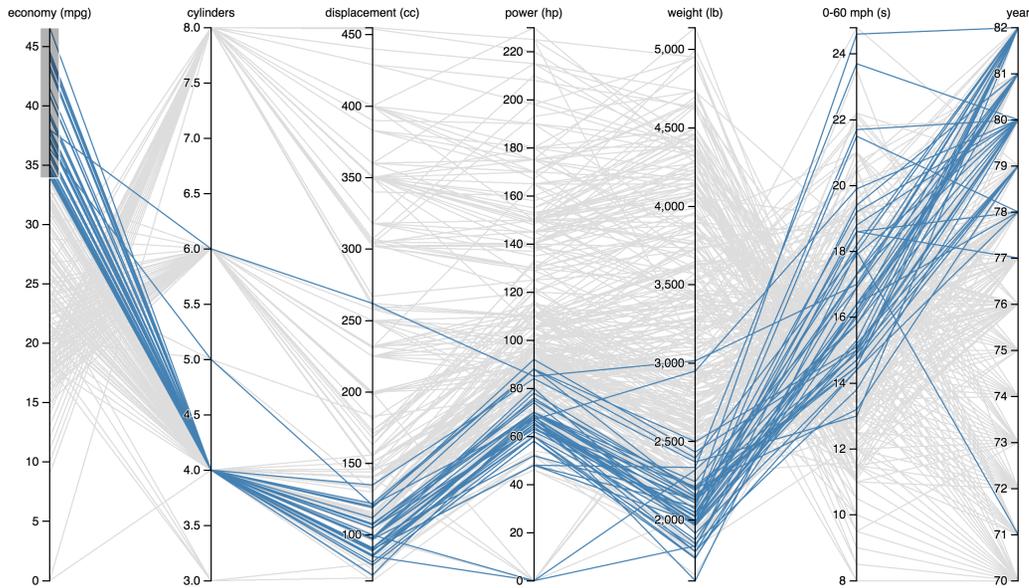
Structures

2 3 5 ... 8 Stages

2105 Genes

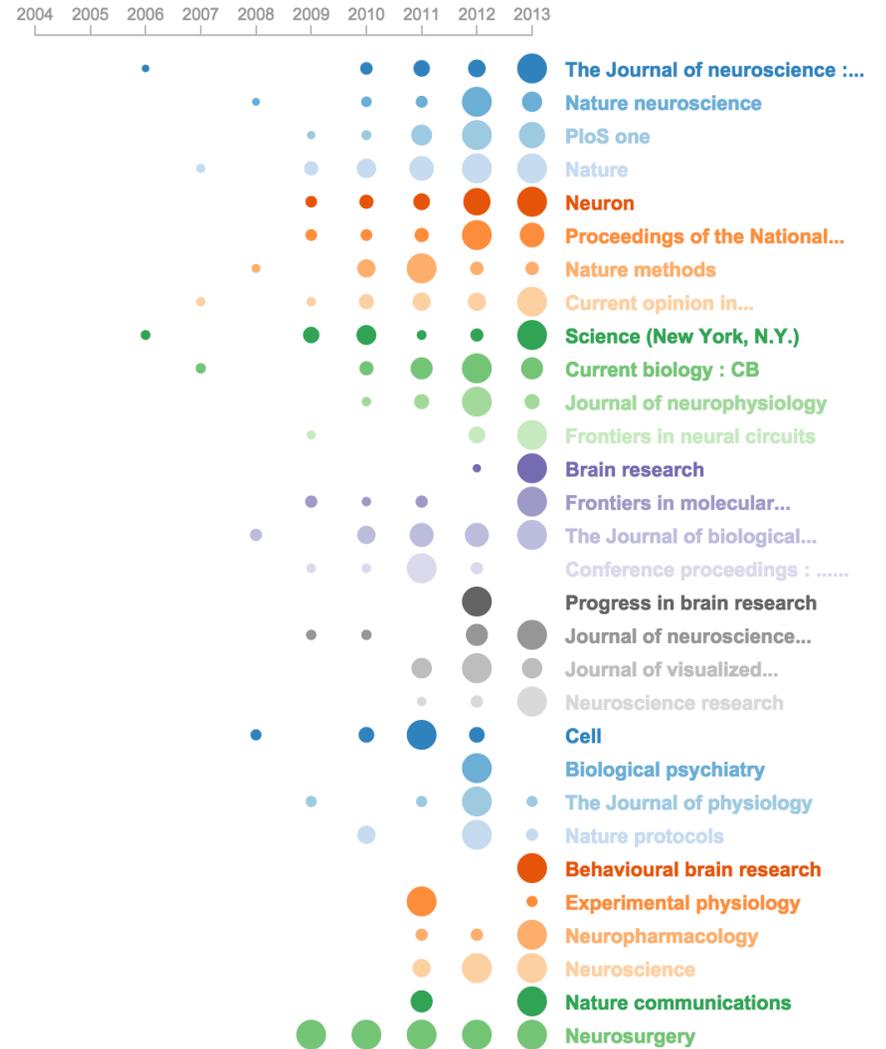
	1	2	3	4	5	6	7
1							
2	1						
3	1	1	2	3	4	5	6
4	2	0.062689800...	0.062689800...	0.077981000...	0.096825800...	0.050108100...	0.047100700...
5	3	0.098959100...	0.098959100...	0.055690400...	0.055093400...	0.051628800...	0.027345400...
6	4	1.620790000...	1.620790000...	1.316000000...	1.316000000...	1.316000000...	1.316000000...
7	5		1	2	3	4	5
8	6	4.662760	0.020330400...	0.020330400...	0.024367600...	0.013878400...	0.020767900...
9	7		2	0.104698000...	0.104698000...	0.107860000...	0.108291000...
10	8	0.694133	0.924802000...	0.924802000...	1.285710000...	1.383380000...	0.192354000...
11	9	0.040772	0.101458000...	0.101458000...	0.094274100...	0.103426000...	0.111858000...
12	10	0.043082	0.139741000...	0.139741000...	0.152459000...	0.164280000...	0.142623000...
13	11	0.090688	6.041510000...	6.041510000...	5.959610000...	6.075560000...	7.313620000...
14	12	0.289733	1.576260000...	1.576260000...	1.525690000...	1.569860000...	1.208210000...
15	13	1.521220	0.774150000...	0.774150000...	0.225078000...	0.225567000...	0.085435300...
16	14	6.547800	0.121122000...	0.121122000...	0.123071000...	0.127634000...	0.017574300...
17	15	0.294734	0.054293300...	0.054293300...	0.053339000...	0.059368100...	0.037003200...
18	16	0.239503	1.947020000...	1.947020000...	2.876400000...	3.288150000...	0.209120000...
19	17		0.185758000...	0.185758000...	0.200427000...	0.205833000...	0.045950900...
20	18	6.098760	6.475020000...	6.475020000...	6.880030000...	7.490330000...	5.194470000...
21	19	1.745570	0.193975000...	0.193975000...	0.277544000...	0.291916000...	0.019118000...
22	20		0.027306300...	0.027306300...	0.029533300...	0.033288600...	0.024705400...
23	21		0.997047000...	0.997047000...	1.271360000...	1.305960000...	1.168350000...
24	22	3.162960	11.43510000...	11.43510000...	15.05490000...	16.01030000...	4.988850000...
	23	3.682790	0	0	0	0	0
	24	1.378740	3.462410000...	3.462410000...	4.107200000...	4.163530000...	1.388010000...
		0.239197	4.678770000...	4.678770000...	6.617860000...	7.555720000...	1.796850000...
			2.809340000...	2.809340000...	3.175210000...	3.275570000...	2.994430000...
			7.861500000...	7.861500000...	10.94040000...	12.06930000...	1.903800000...
			0.516917000...	0.516917000...	0.511336000...	0.517883000...	0.476425000...
			1.296560000...	1.296560000...	1.537690000...	1.612880000...	1.514040000...

Visualizing Tables



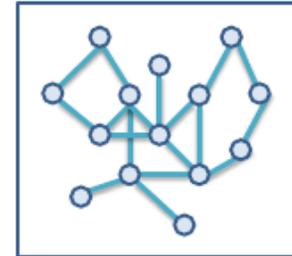
More in Lecture 8: High-Dimensional Data

Optogenetic



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.



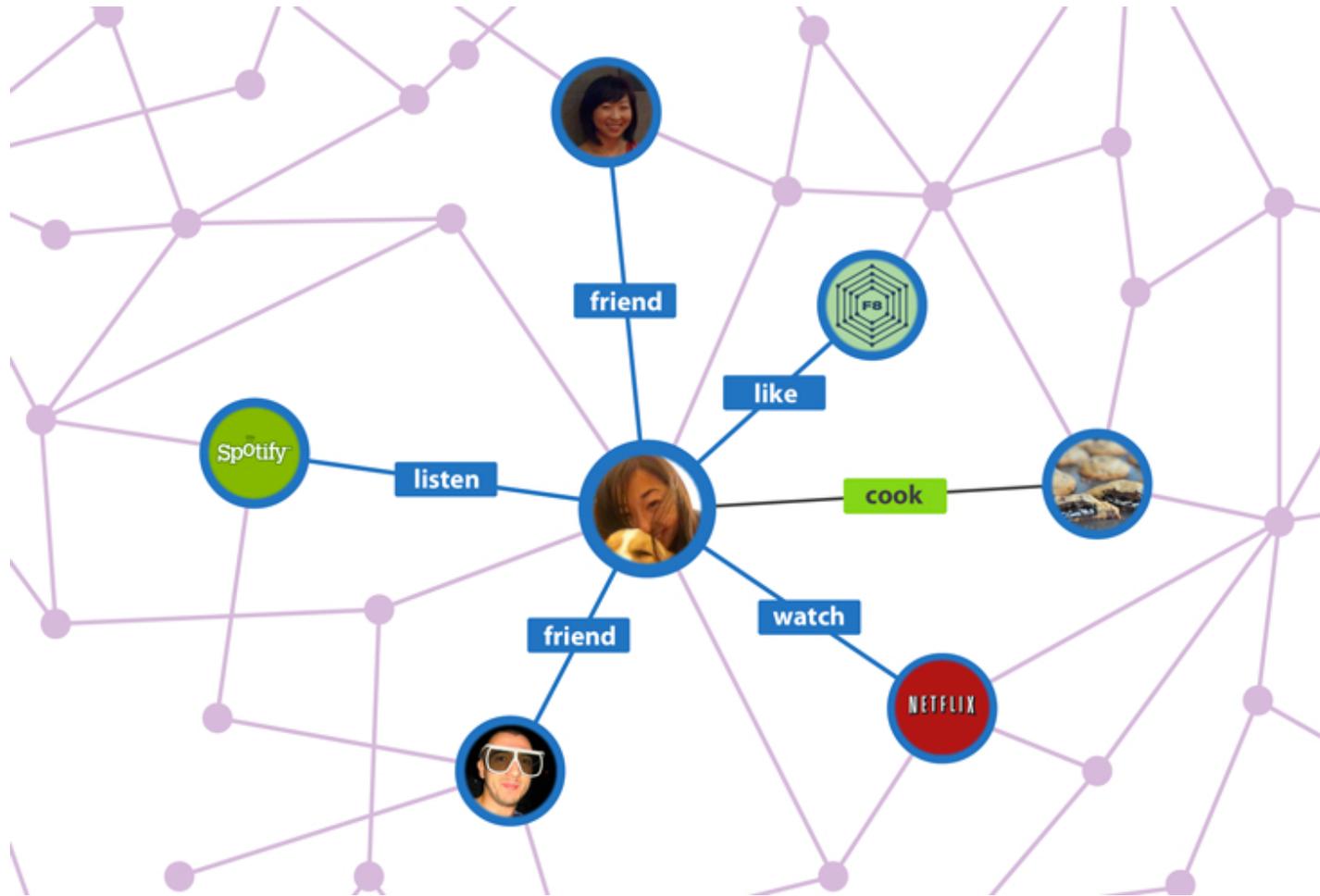
Diagrammatic Example



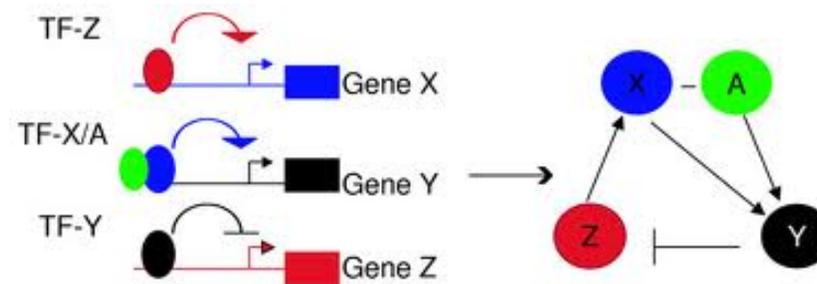
A Source of Hair Balls



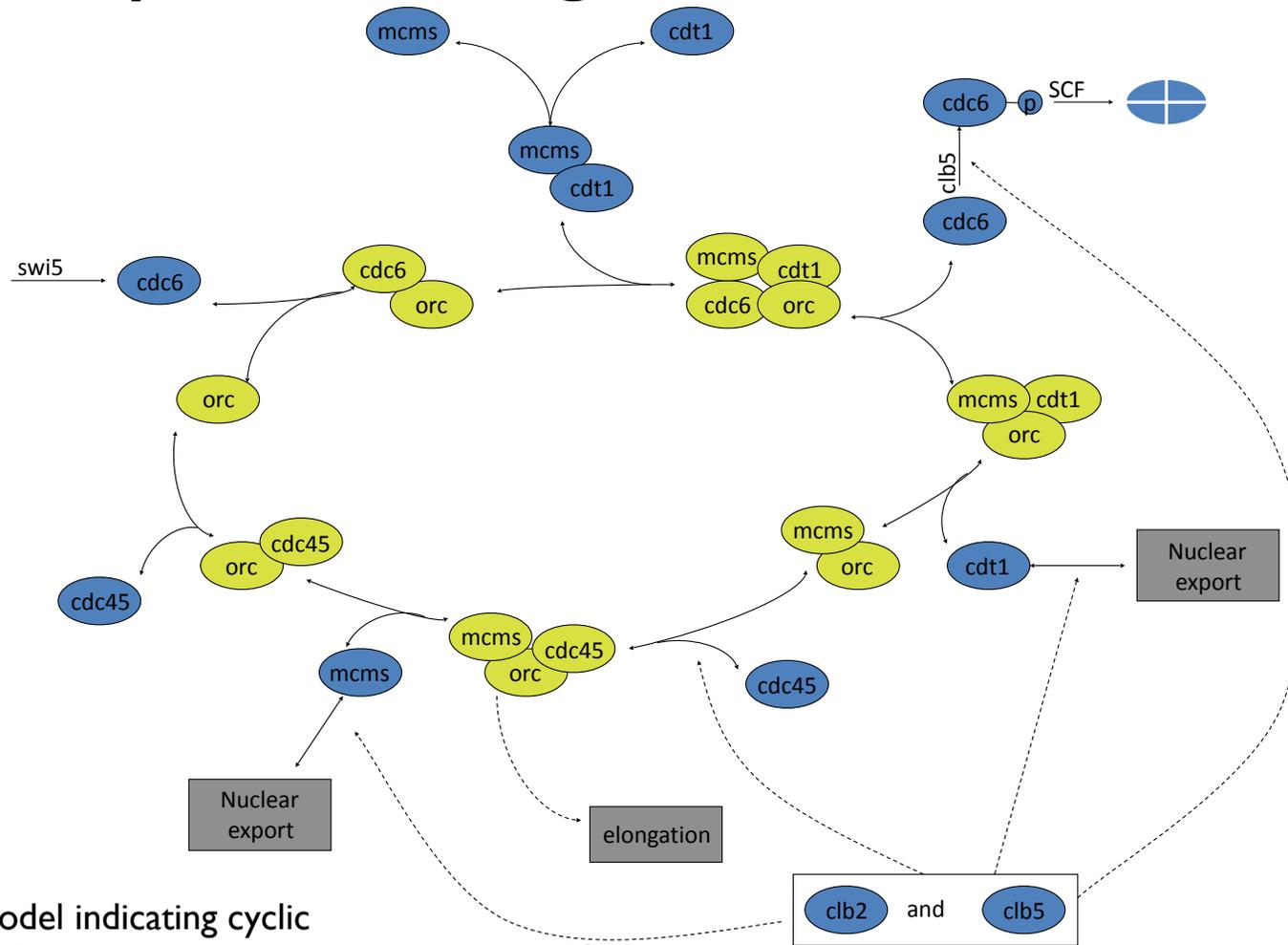
Relationship – “Friend of”



Genes in Homo Sapiens



Relationship – “To Regulate”

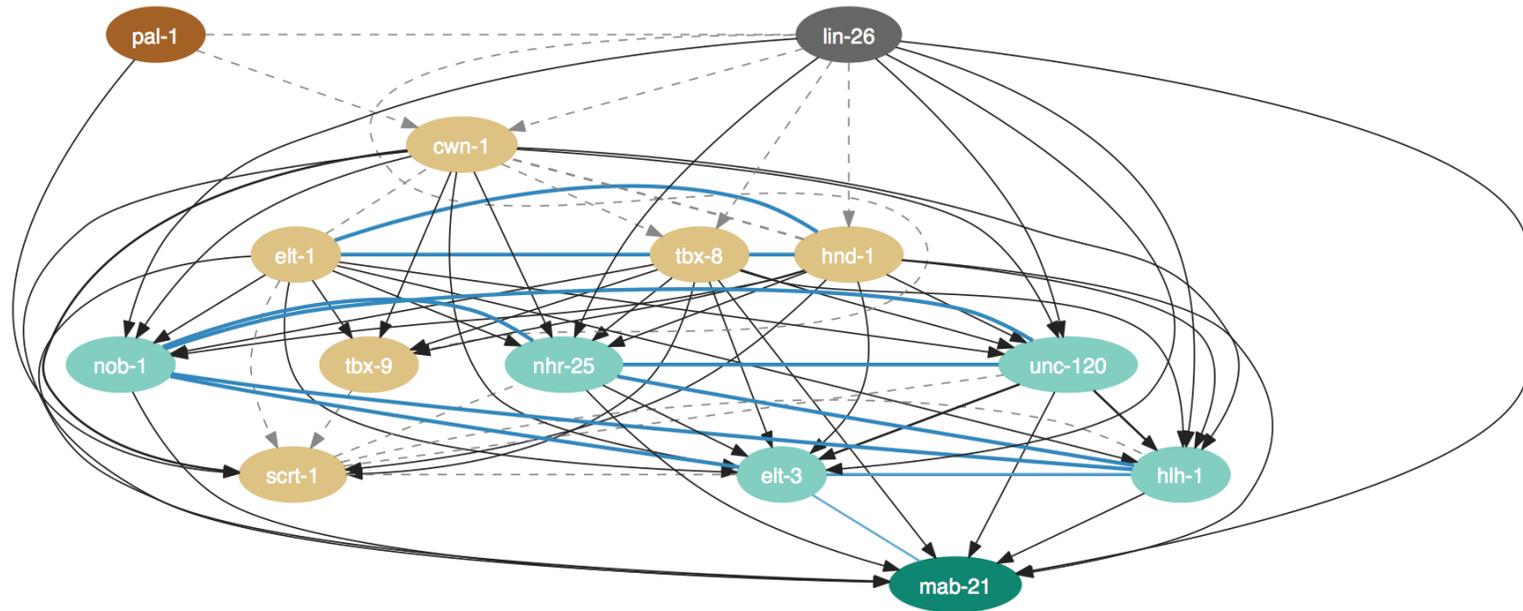


Detailed model indicating cyclic behaviour of pre-replicative machinery

33

1.3

The PAL-I Hairball - Muscle Cells



Graphs/Networks

A simple graph is a graph which contains

No multi-edges

No loops



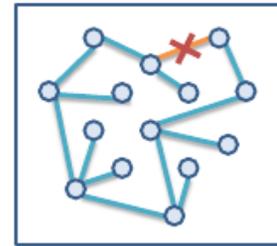
Not a simple graph!
→ A **general graph**

Special Graphs

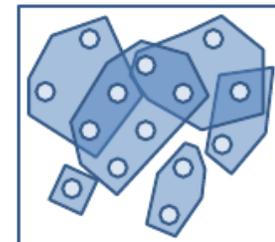
A *tree* is a graph with *no cycles*

A *directed graph* (digraph) is a graph that distinguishes between edges $A \rightarrow B$ and $A \leftarrow B$

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



Tree

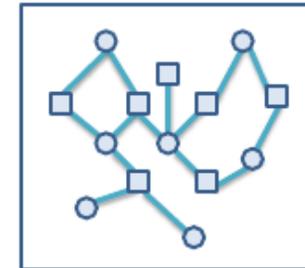


Hypergraph Example

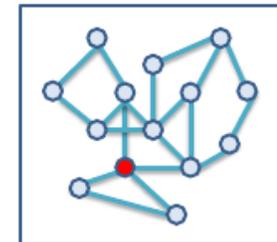
Special Graphs

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

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***

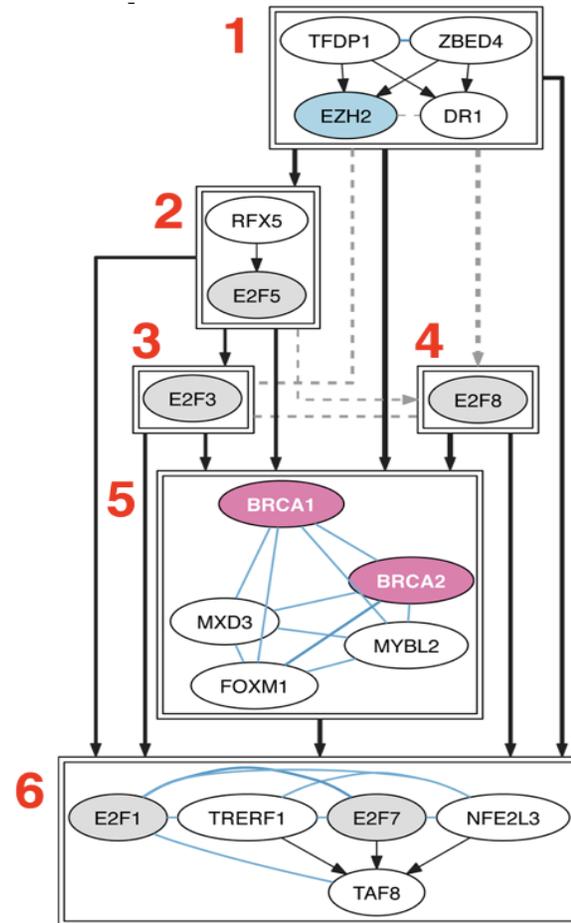
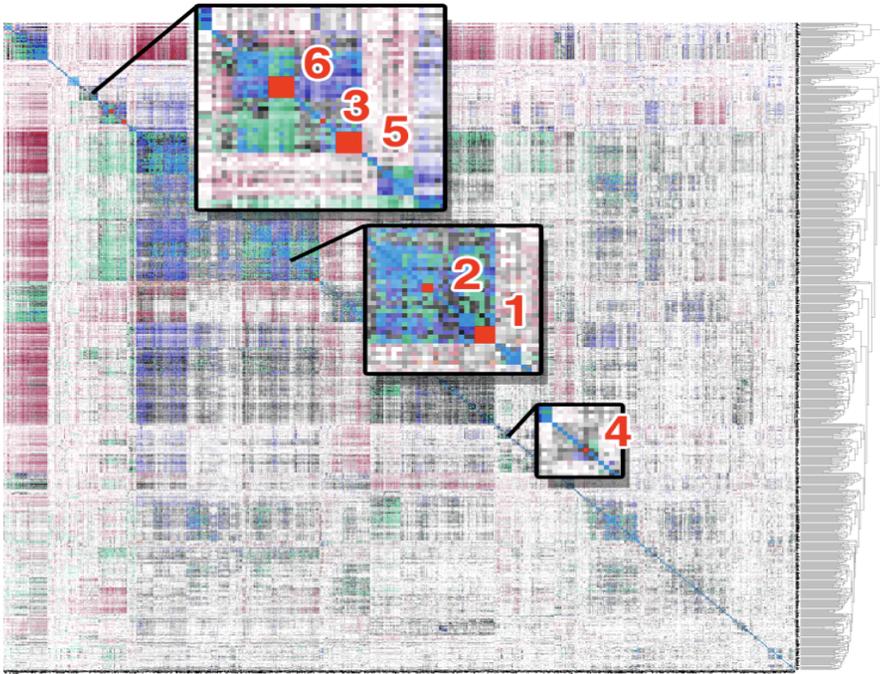


Bipartite Graph



Articulation Point (red)

Perceptual Grouping & Order



Fields

Attribute values associated with cells

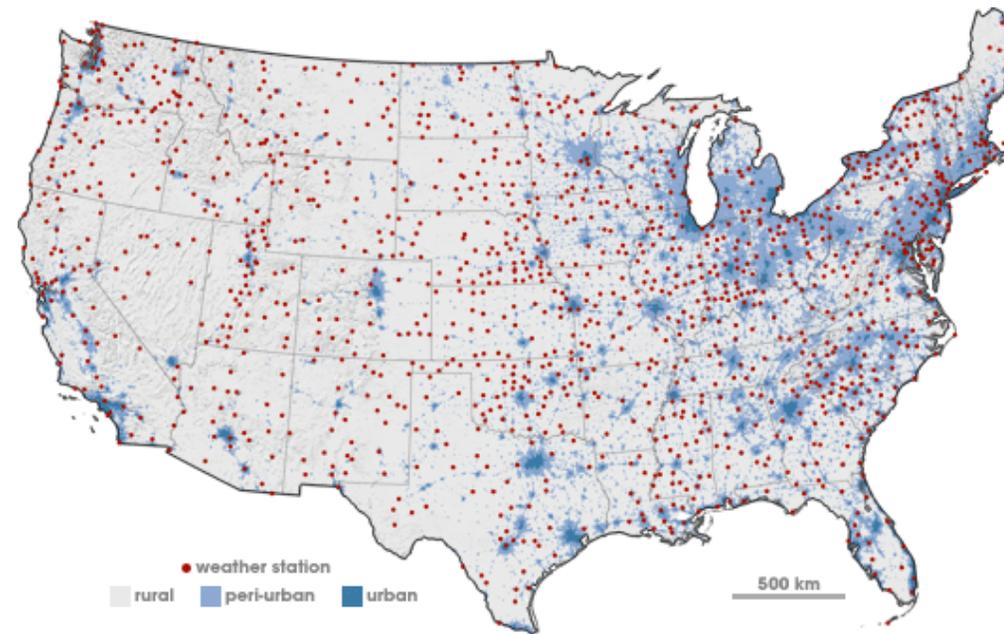
Cell contains data from continuous domain

Temperature, pressure, wind velocity

Measured or simulated

Sampling & Interpolation

Signal processing & stats



Fields: Grid Types

Uniform Grid

Geometry & topology can be computed

Rectilinear Grid

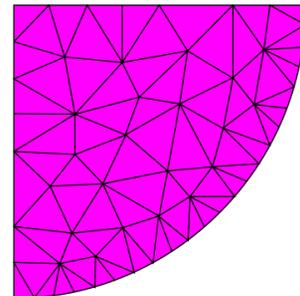
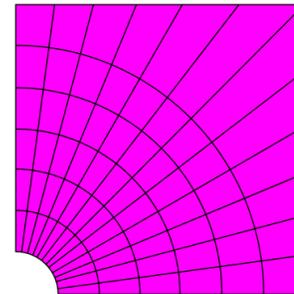
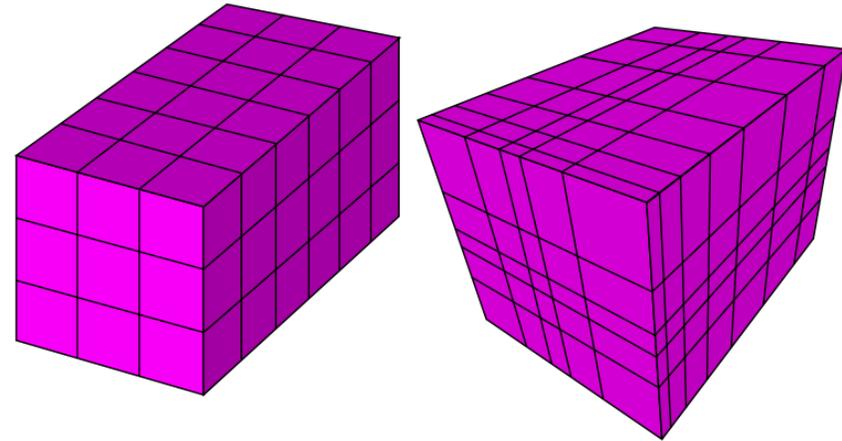
Nonuniform sampling

Structured Grid

allows curvilinear grids

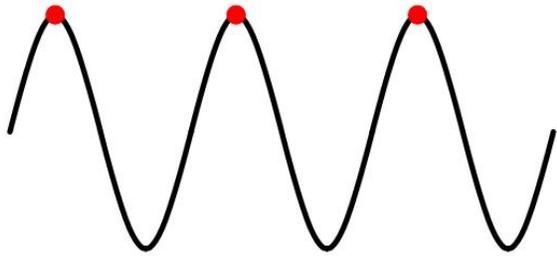
Unstructured Grid

full flexibility, store position and connection

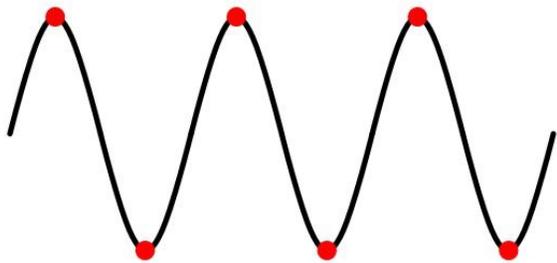


[Wikipedia]

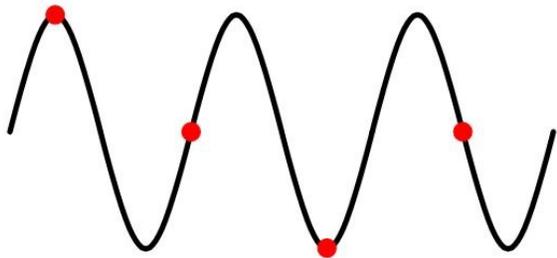
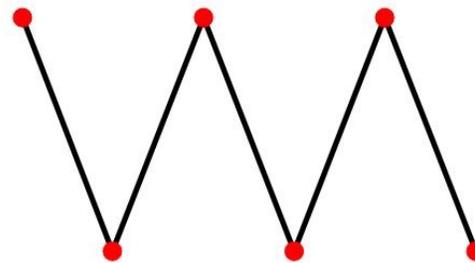
Grids are about sampling



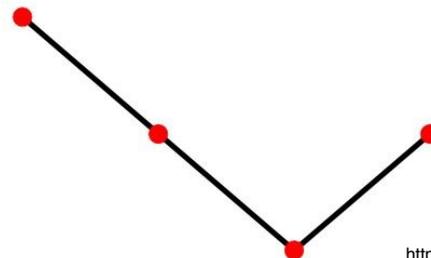
A
→
Sampled at f



B
→
Sampled at $2f$

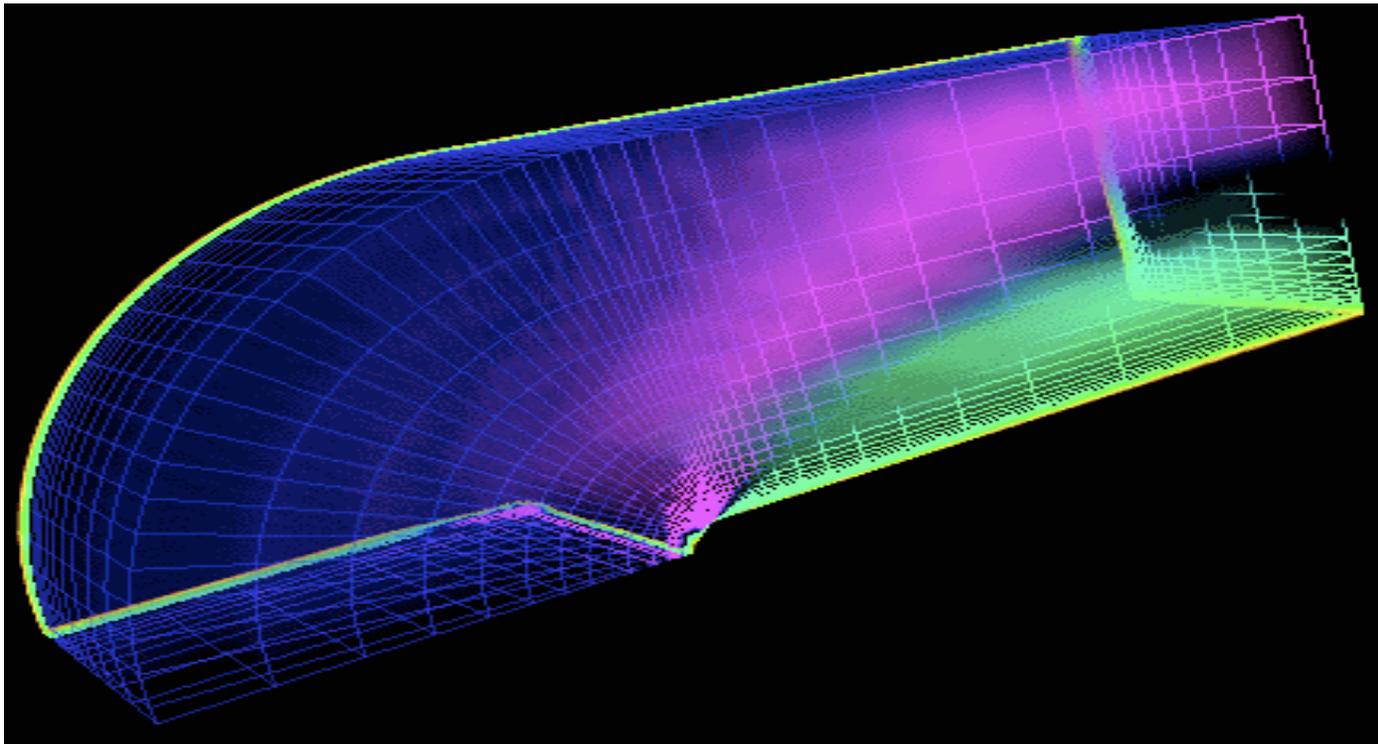


C
→
Sampled at $4f/3$

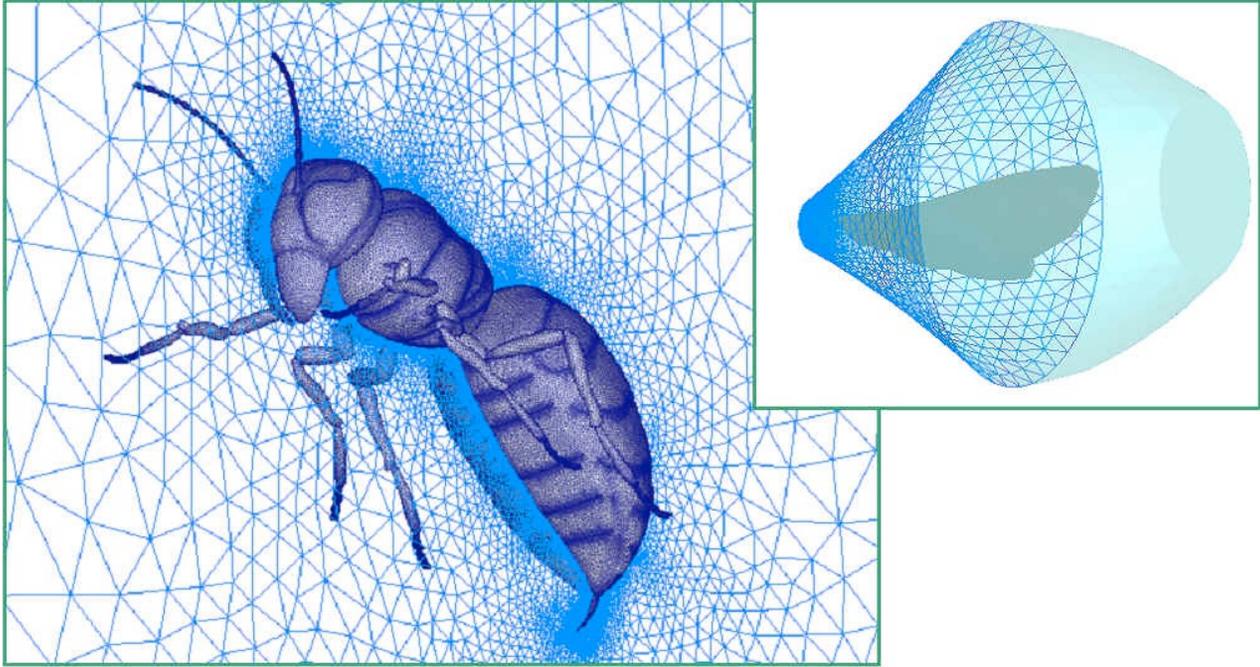


Grids are about geometry

Curvilinear grids



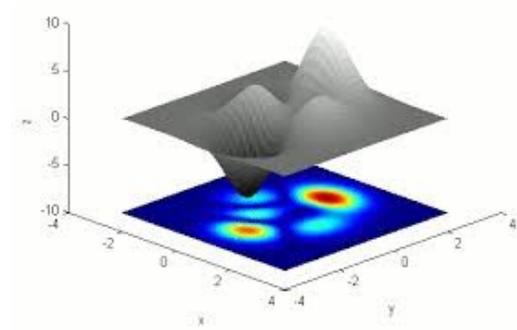
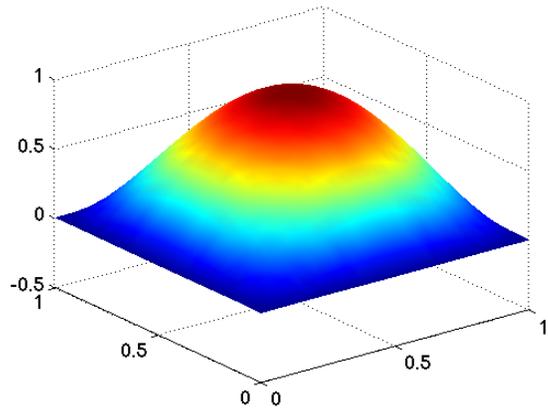
Unstructured grids



Fields

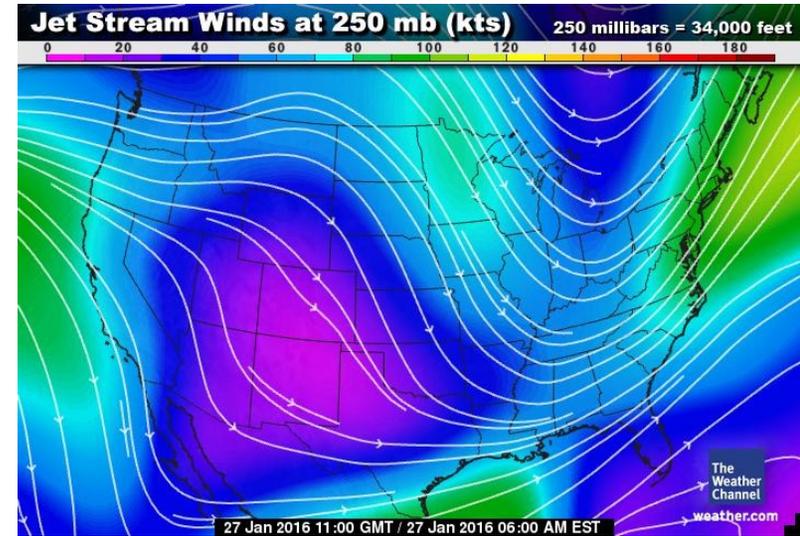
- **Scalar data**
 - mapping $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $(x_1, \dots, x_n) \rightarrow y$
 - n independent variables (keys) x_i (1D, 2D, or 3D, +time)
 - value y is just univariate
- **Examples:**
 - MRI data
 - 2D grey-scale image data

Fields



Vector Fields

- Vector data
 - mapping $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $(x_1, \dots, x_n) \rightarrow (y_1, \dots, y_m)$
 - representing direction and magnitude
 - usually $m=n$
 - Exceptions, e.g., due to projection
- Example:
 - weather map (wind direction)
 - flow around airplane wings



Tensors

Mapping $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, (x_1, \dots, x_n)

tensor of level k

a tensor of level 1 is a vector

a tensor of level 2 is a matrix

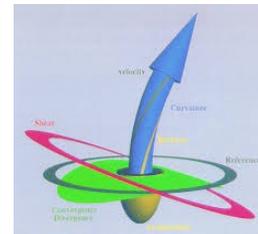


<http://www.cs.rug.nl/svcg/SciVis/COMOTI>

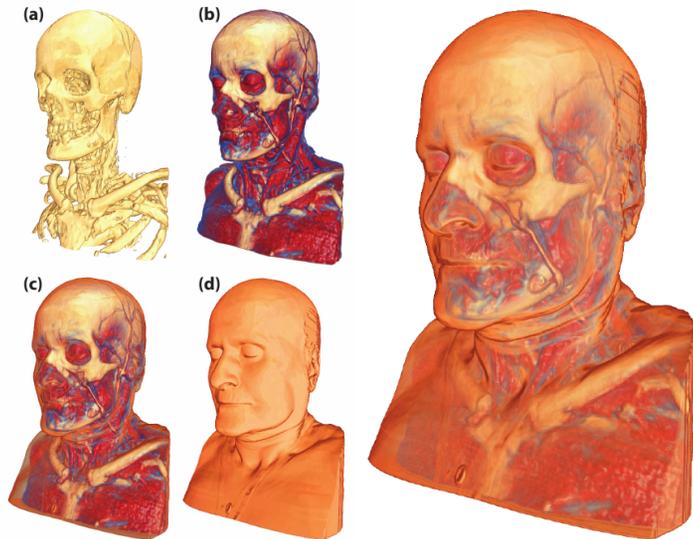
Example:

diffusion-tensor MRI

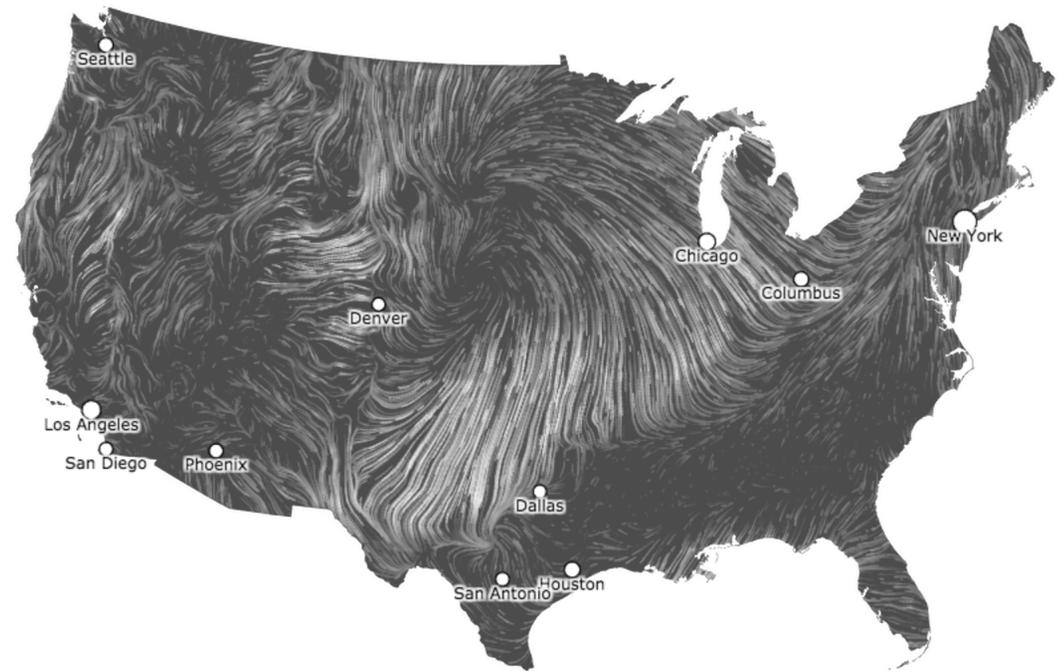
stress-tensor, etc.



Visualizing Fields



[Bruckner 2007]



More later — Visualizing spatial data: Volumes and Flows

Geometry

Shape of items

Explicit spatial positions

Points, lines, curves, surfaces, regions, volumes

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

Scientific Vis

“Spatial
Data” (Fields)

Not free to choose
spatial layout

Find best way to
depict reality

[Johanna, Daniel]

Other Collections

Sets

Unique items, unordered

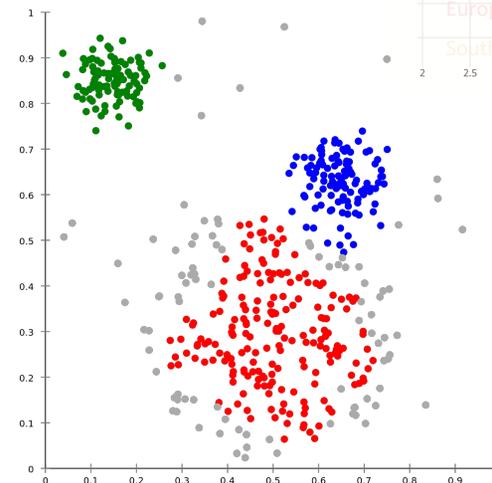
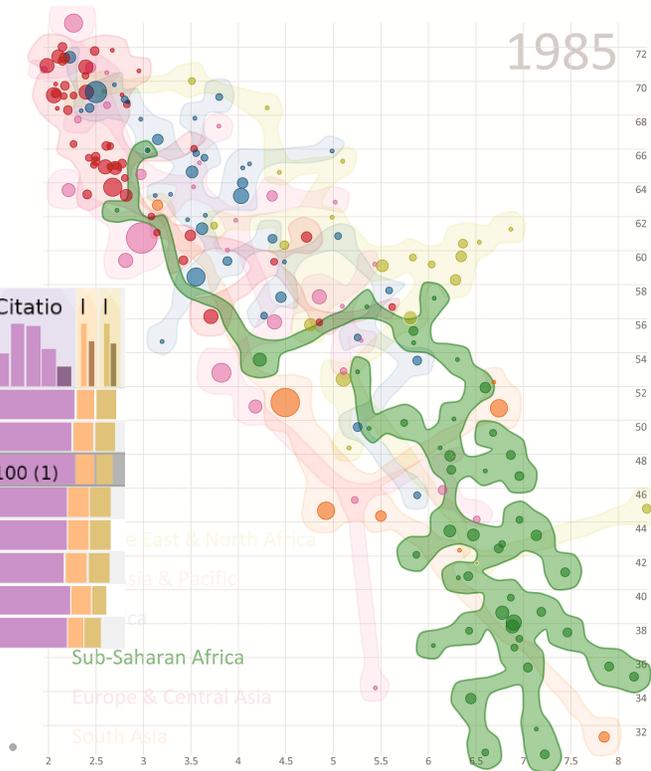
Lists

Ordered, duplicates allowed

Clusters

Groups of similar items

Rank	School Name	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

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



Data Types

- Quantitative (Q)
✓ 10 inches, 23 inches, etc.
- Ordinal (ordered) (O)
✓ Small, medium, large
- Nominal (categorical) (N)
✓ Apples, Oranges, Bananas,...

→ Categorical

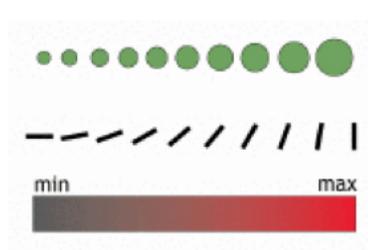
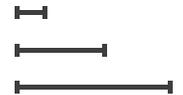


→ Ordered

→ Ordinal



→ Quantitative



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

Hierarchies

- possible for any data type
- sometimes strong implicit hierarchies
- Lab I - building/floor/room/nooks-crannies
- e.g. geography:
 - postal code
 - city district
 - city
 - state
 - country
 - continent

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	<i>Permutation group</i> $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	<i>Isotonic group</i> $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	<i>General linear group</i> $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	<i>Similarity group</i> $x' = ax$	Coefficient of variation

Data Types & Operators

Nominal (labels)

Operations: =, ≠

Ordinal (ordered)

Operations: =, ≠, >, <

Interval (location of zero arbitrary)

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

Ratio (zero fixed)

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

Sequential & Diverging Data

Sequential:

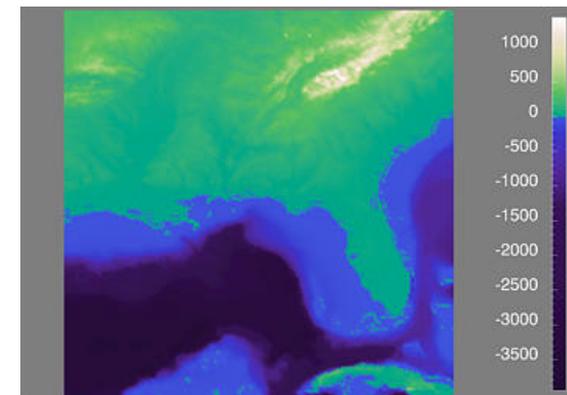
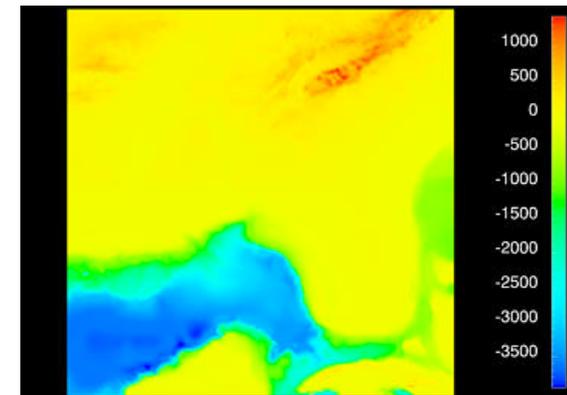
homogeneous from min to max

people in countries

Diverging:

two or multiple sequences that meet

Elevation dataset: above sea level
& below sea level



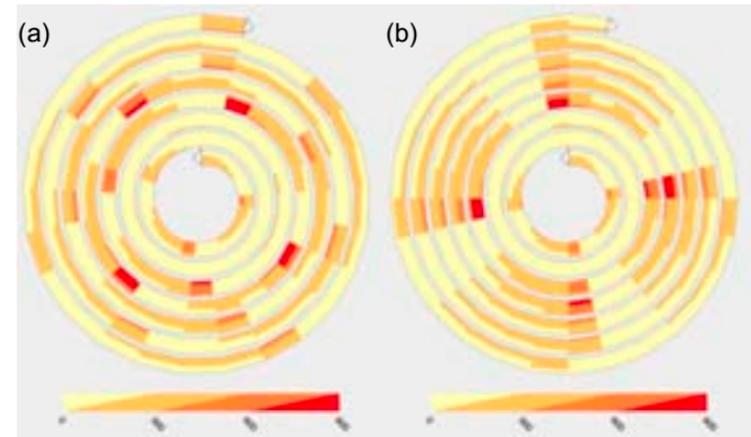
Other Structure

Cyclic data

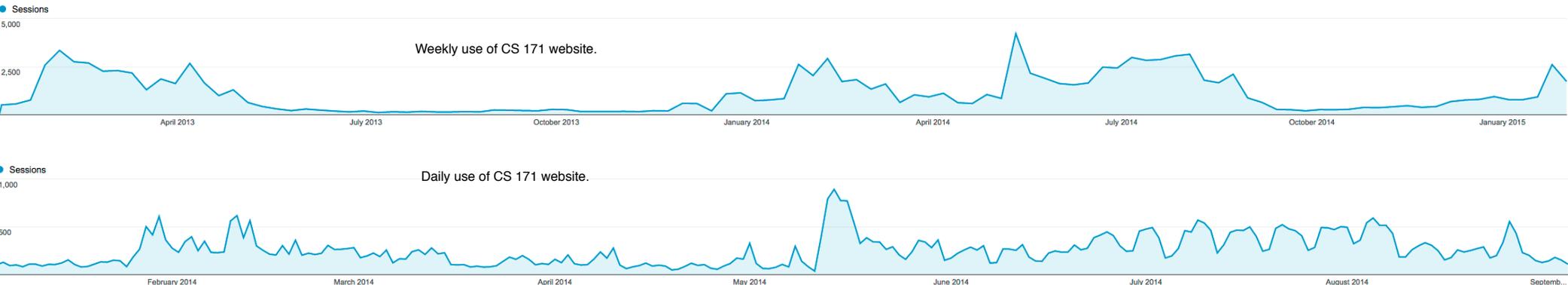
time (hours, week, month, year)

Aggregation

might be patterns on multiple levels



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



Time

- Strong (implicit) hierarchy:
 - minute
 - hour
 - day
 - week
 - month
- ordinal (entries in a diary) ?
- quantitative (timings in a race) ?
- interval vs. ratio — time-stamp vs. duration

Quiz & Recap

	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		10/24/07
9	35	10/23/07	4-Not Specified	Small Box		10/25/07
10	36	11/3/07	1-Urgent	Small Box		11/3/07
11	65	3/18/07	1-Urgent	Small Pack		3/19/07
12	66	1/20/05	5-Low	Wrap Bag		1/20/05
13	69	6/4/05	4-Not Specified	Small Pack		6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag		6/6/05
15	70	12/18/06	5-Low	Small Box		12/23/06
16	70	12/18/06	5-Low	Wrap Bag		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

**Item/Element/
(Independent)
Variable**

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

**Attribute/
Dimension/
(Dependent)
Variable/
Feature**

	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	0.5	2/22/08
4	32	7/16/07	2-High	Small Pack	0.9	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

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

Tasks

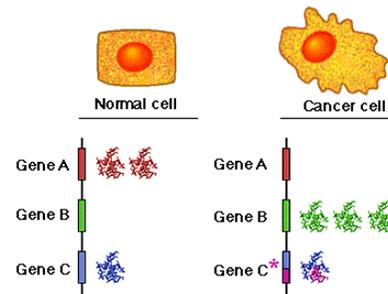
Why are we using Visualization?

Industrial Engineer



Industrial engineering is a branch of engineering which deals with the optimization of complex processes or systems. Industrial engineers work to eliminate waste of time, money, materials, man-hours, machine time, energy and other resources that do not generate value.

The Compare Task



Find the set of genes which are expressed in tumor cells as opposed to non-tumor cells

See if results for tissue samples treated with LL-37 match up with the ones without the peptide

....

**Same Vis Tool Used for Compare
across Use Cases !**

Vis is a Idiom

An idiom (Latin: idioma, "special property", from Greek: ἰδίωμα – idíōma, "special feature, special phrasing, a peculiarity", f. Greek: ἴδιος – ídios, "one's own") is a phrase or a fixed expression that has a figurative, or sometimes literal, meaning. An idiom's figurative meaning is different from the literal meaning.

Vis is a Idiom



Who does Vis ?
Designer or User ?

Domain and Abstract Tasks

Large numbers of domain tasks

Can be broken down into simpler abstract tasks

We know how to address the abstract tasks!

Identify task - data combination: solutions probably exist

Example

LineUp: Visual Analysis of Multi-Attribute Rankings

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit

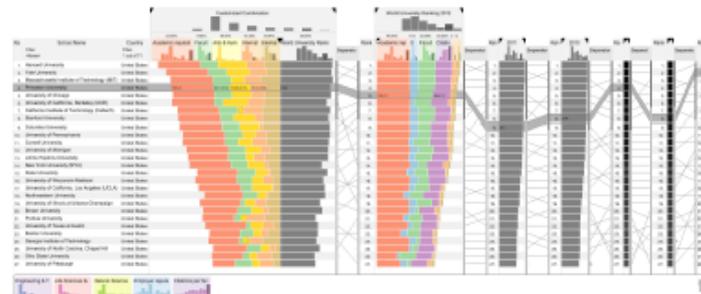


Fig. 1. LineUp showing a ranking of the top Universities according to the QS World University Ranking 2012 dataset with custom attributes and weights, compared to the official ranking.

<http://caleydo.org/projects/lineup/>

Example 1

Find good universities with a high faculty student ratio.

Identify high-ranked universities

In this subset: **compare** universities & **identify** high faculty student ratio

OR

Derive a ranking with a high weight for faculty student ratio

QS World University Rankings® IREG APPROVED and QS Stars

Filter by region Filter by location reset

Filter by faculty Note: Filtering by subject area will also resort the list by subject-area scores. reset

RANK	UNIVERSITY	LOCATION	COMPARE & MEET	QS STARS
1	100.0	Massachusetts Institute of Technology (MIT)	<input type="checkbox"/>	★★★★★+
2	99.4	University of Cambridge	<input type="checkbox"/>	★★★★★
2	99.4	Imperial College London	<input type="checkbox"/>	
4	99.3	Harvard University	<input type="checkbox"/>	★★★★★
5	99.2	University of Oxford	<input type="checkbox"/>	★★★★★
5	99.2	UCL (University College London)	<input type="checkbox"/>	

Click on a table row to get extended information

Example 2

Contrast Harvard's reputation scores with MIT's

Match up Harvard with Yale

First, find Harvard and Yale, then **compare** their (two) reputation scores

QS World University Rankings® IREG APPROVED and QS Stars

Filter by region Filter by location reset

Filter by faculty Note: Filtering by subject area will also resort the list by subject-area scores. reset

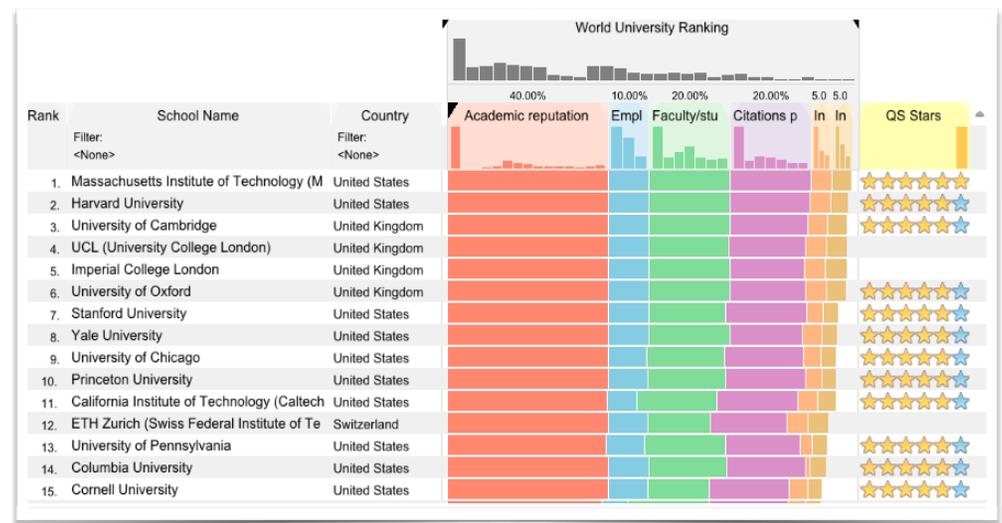
RANK	UNIVERSITY	LOCATION	COMPARE & MEET	QS STARS
Overall Score	Search for universities...			Show only
1	100.0  Massachusetts Institute of Technology (MIT)		<input type="checkbox"/>	
2	99.4  University of Cambridge		<input type="checkbox"/>	
2	99.4  Imperial College London		<input type="checkbox"/>	
4	99.3  Harvard University		<input type="checkbox"/>	
5	99.2  University of Oxford		<input type="checkbox"/>	
5	99.2  UCL (University College London)		<input type="checkbox"/>	

Click on a table row to get extended information

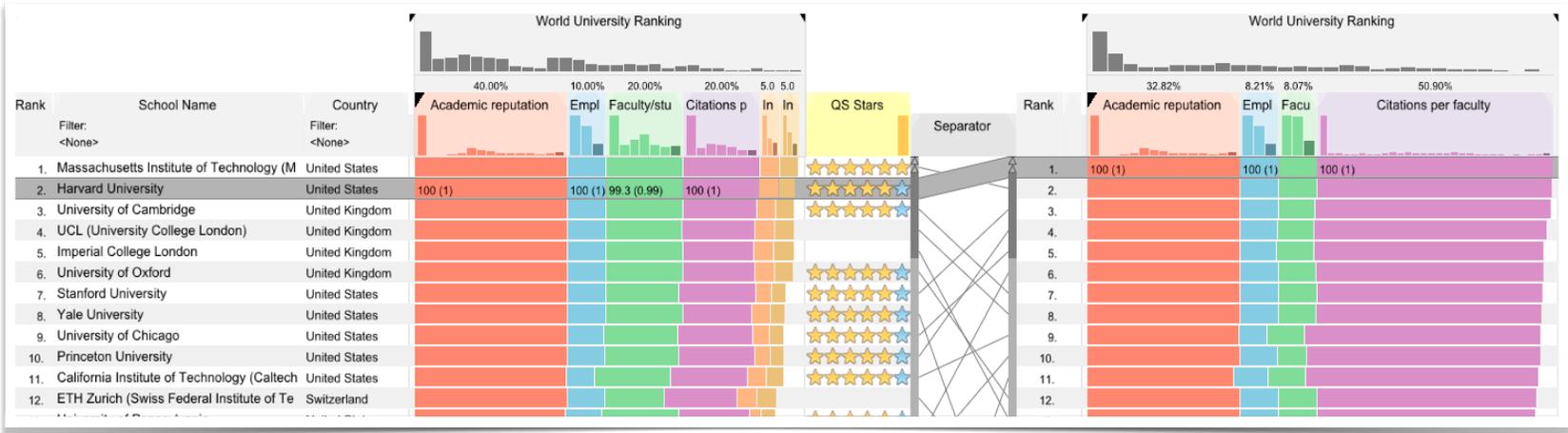
Example 3 - LineUp

Find a combination of weights and parameters where Harvard is better than MIT

Produce a new dataset by deriving from the input parameters



Result



Tasks - Actions

Analyze

- high-level choices
- consume vs produce

Search

- mid-level choices
- find a known/unknown item

Query

- low-level
- find out about characteristics of item
- by itself or compare relative to others

High-level actions: Analyze

→ Analyze

Consume

discover vs present

discover to generate hypotheses

discover to confirm hypotheses

classic split: explore vs explain

present to communicate - planning, forecasting, and instructional

enjoy: casual, social, piece of art

→ Consume

→ Discover



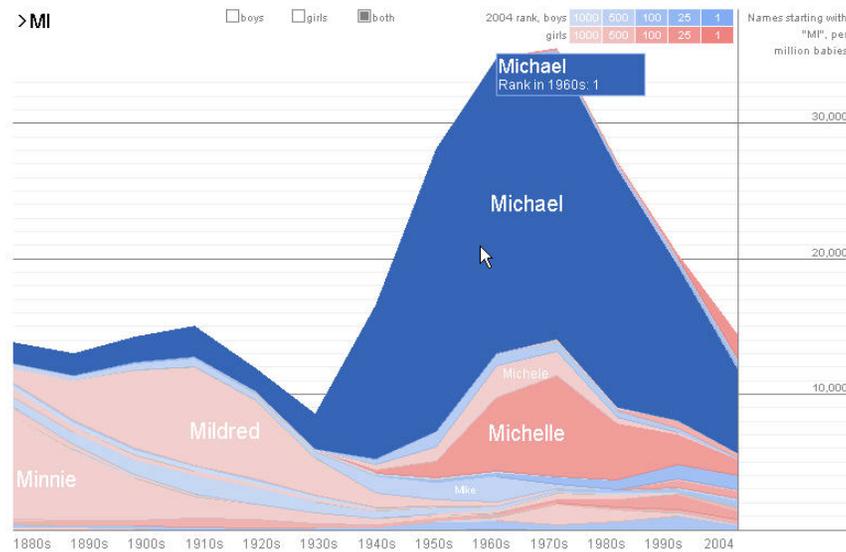
→ Present



→ Enjoy



Just Enjoy - NameVoyager



High-level actions: Analyze

Produce

Annotate,

Record

Derive: crucial design choice

→ Produce

→ Annotate



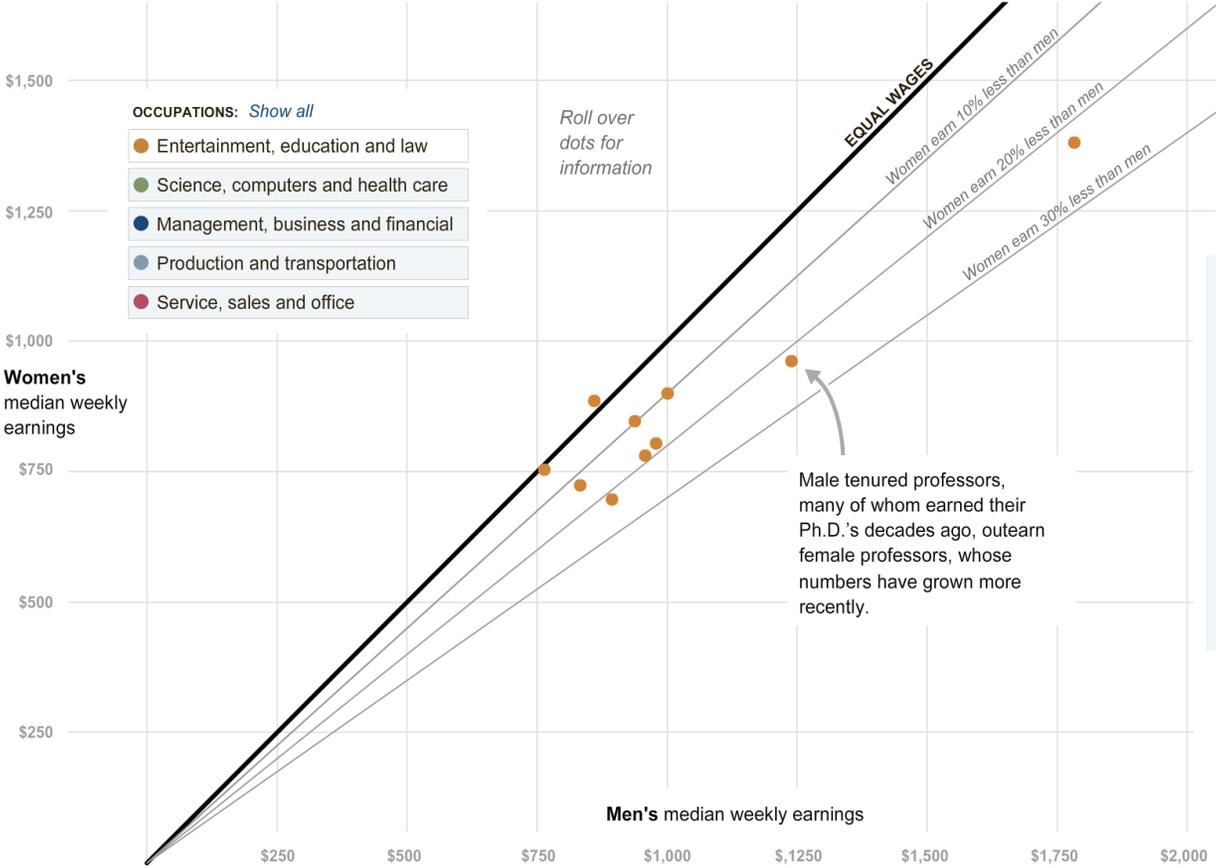
→ Record



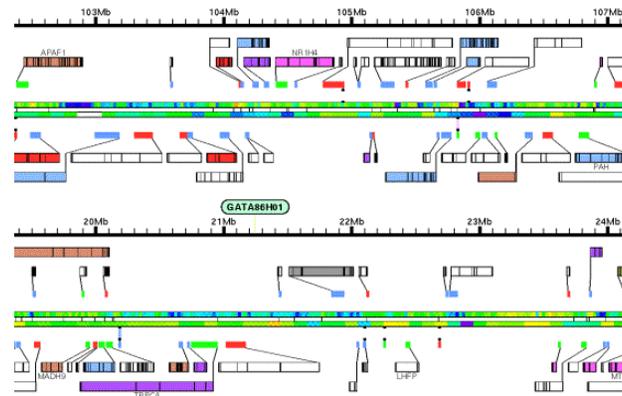
→ Derive



Example: Annotate



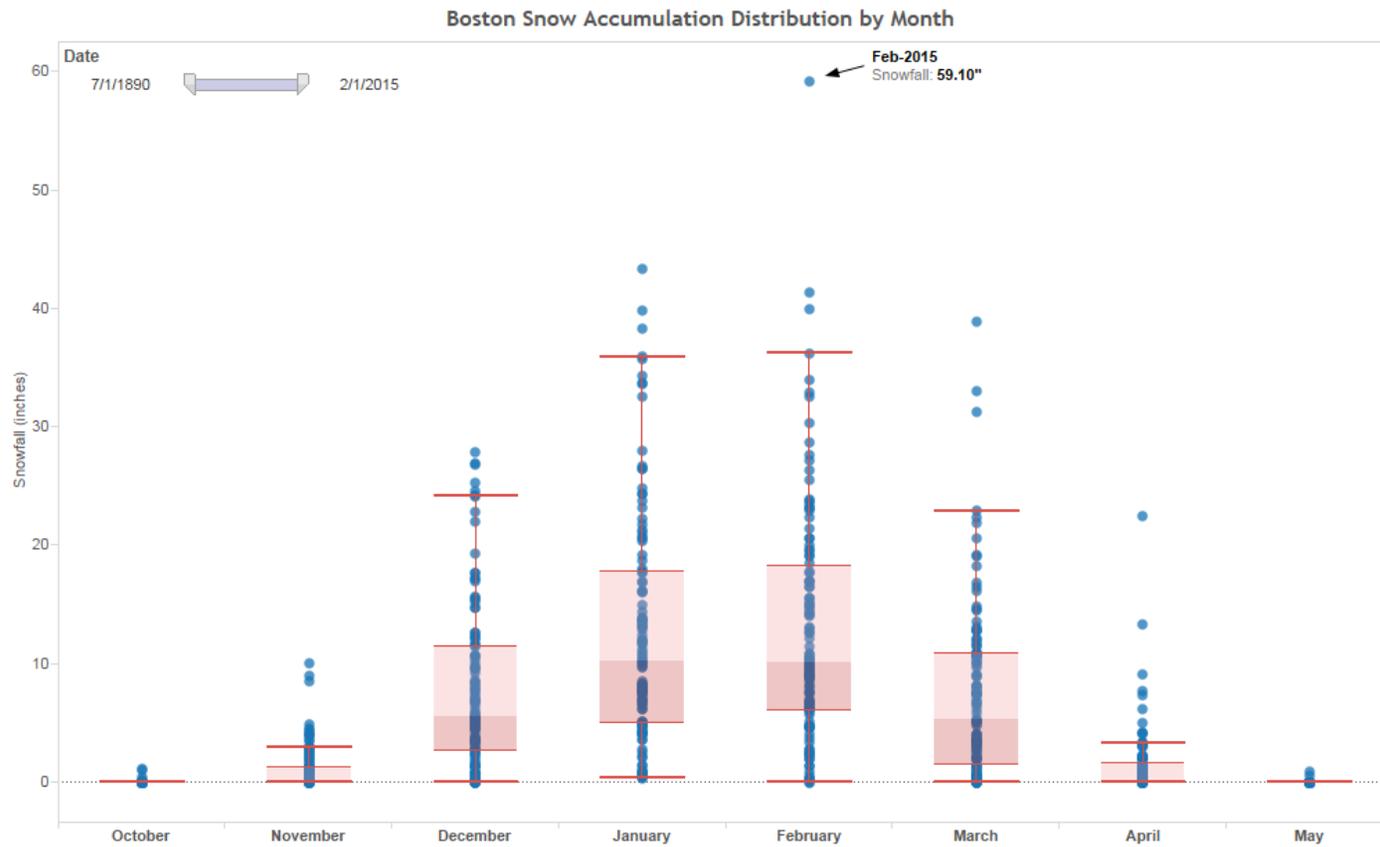
Annotate The Genome



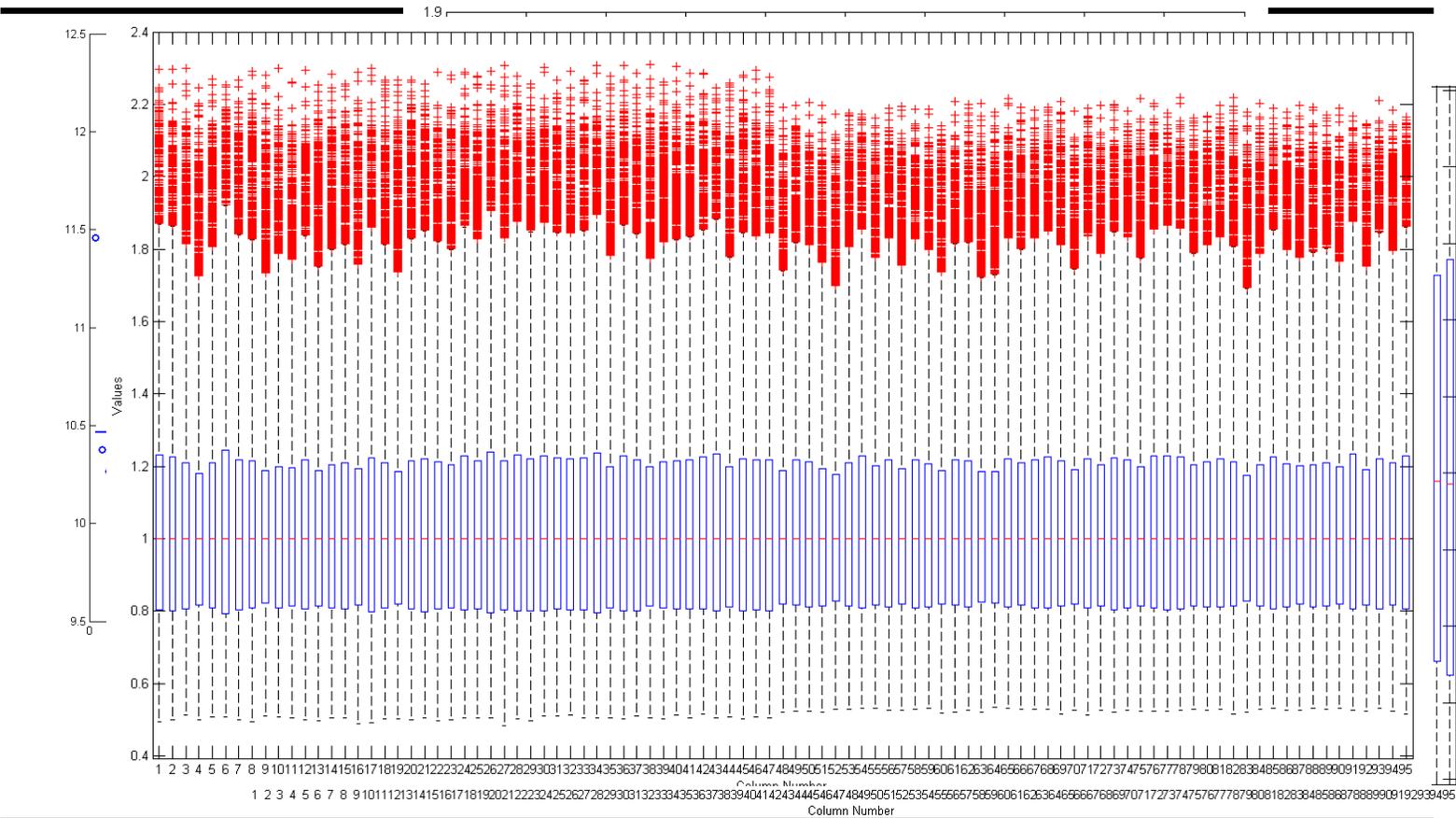
Record - Tableau Dashboards



Example: Derive (Transform)



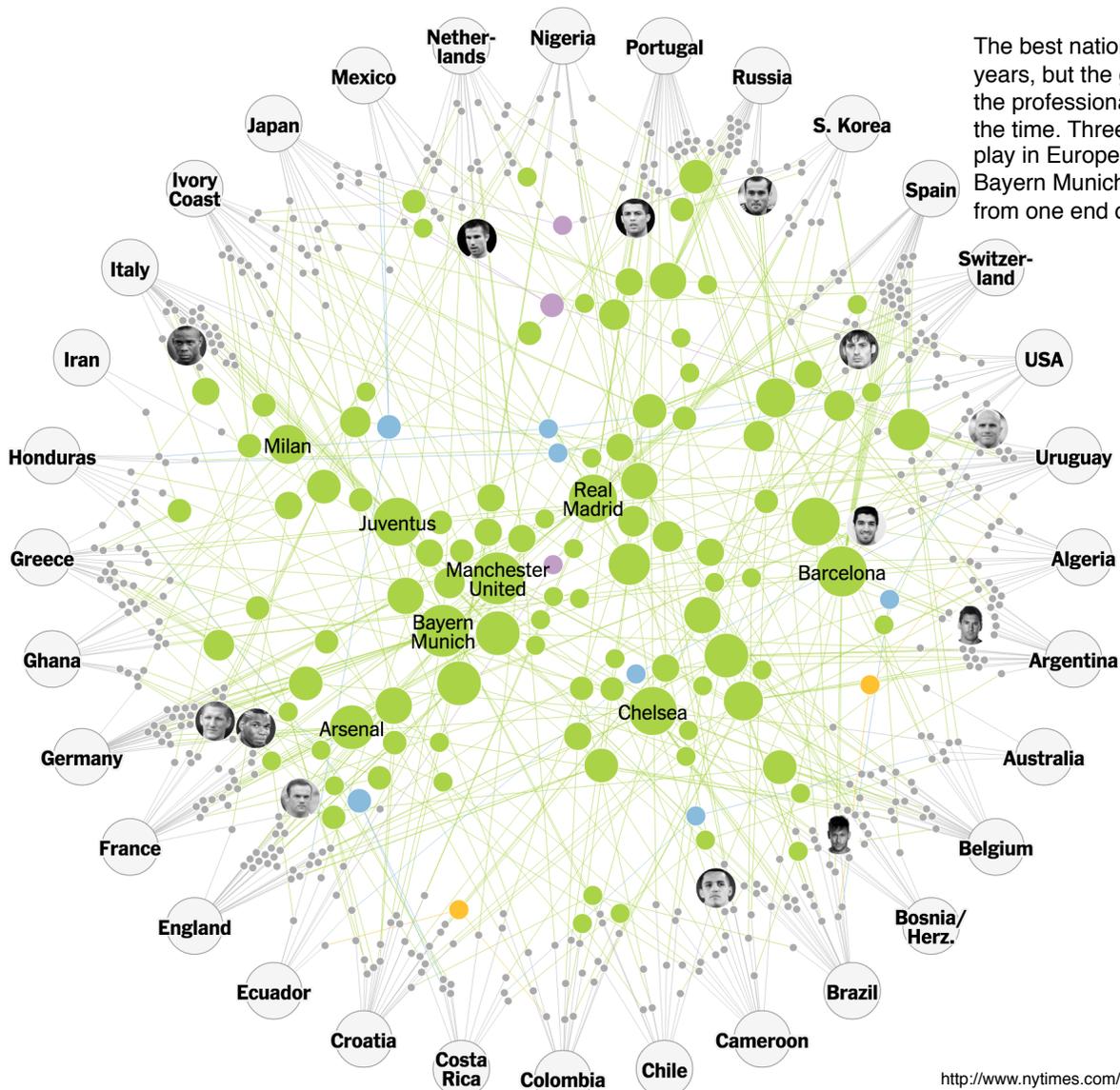
Example MicroArray Normalization



Wexner Medical Center

Example: Derive

	Country	Club	Club Continent
Ronaldo	Portugal	Real Madrid	Europe
Lahm	Germany	Bayern München	Europe
Robben	Netherlands	Bayern München	Europe
Khedira	Germany	Real Madrid	Europe
Phogba	Italy	Juventus	Europe
Messi	Argentina	Barcelona	Europe



The best national teams come together every four years, but the global tournament is mostly a remix of the professional leagues that are in season most of the time. Three out of every four World Cup players play in Europe, and the top clubs like Barcelona, Bayern Munich and Manchester United have players from one end of the globe to the other.

Search

Search

mid-level choices

find a known/unknown item

	<i>target known</i>	<i>target unknown</i>
<i>location known</i>	lookup	browse
<i>location unknown</i>	locate	explore

Known Knowns & Unknowns Unknowns

https://en.wikipedia.org/wiki/There_are_known_knowns

Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns** – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.[1] - D. Rumsfeld



Actions: Mid-level search

→ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

what does user know?

target, location

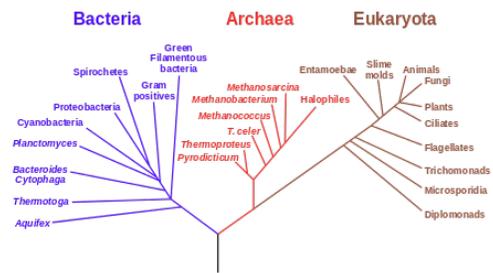
known/known - you know what you are looking for [red M&M]

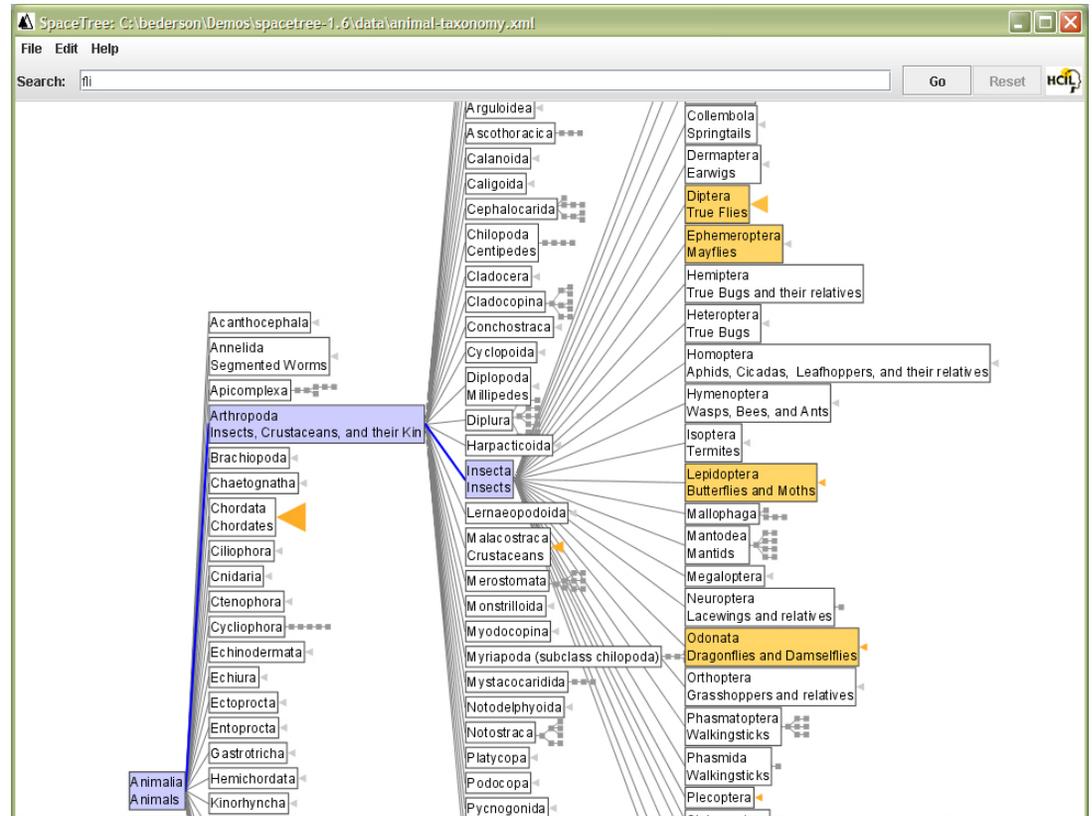
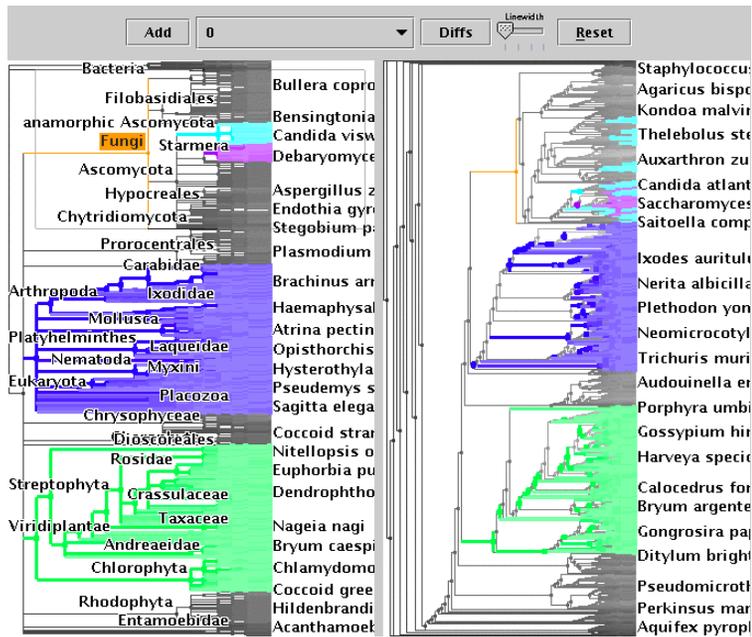
known/unknown - you know characteristics (range, attributes) [Size of red M&M]

unknown/known - just a plain search [no visuals; locate red M&M]

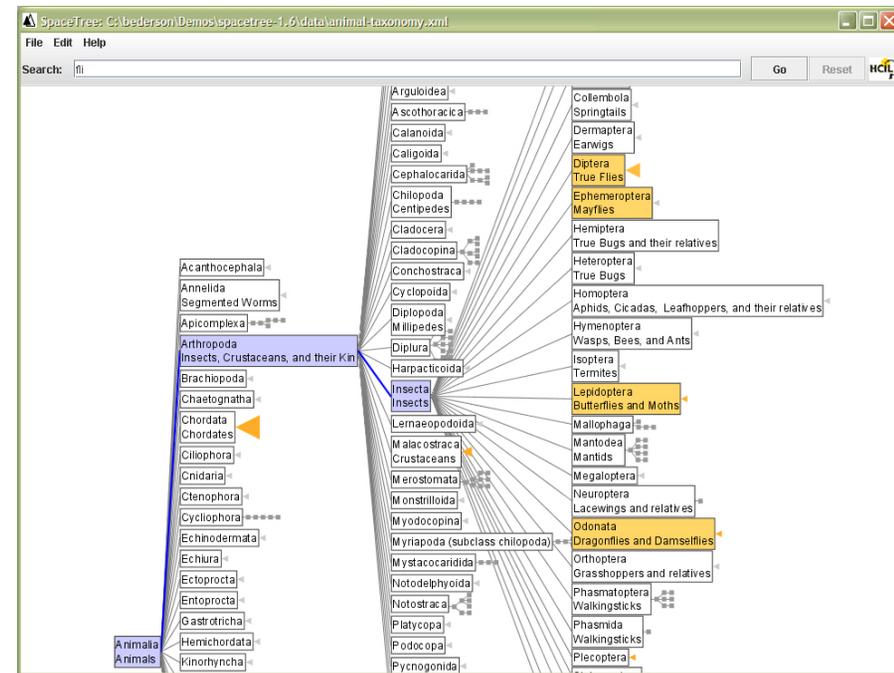
unknown/unknown - outliers, anomalies, unanticipated [no color; just too many]

Phylogenetic Tree of Life





Known/Unknown



known/known - you know what you are looking for [....]

known/unknown - you know characteristics (range, attributes) [...]

unknown/known - just a plain search [...]

unknown/unknown - outliers, anomalies, unanticipated [...]

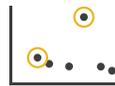
Actions: Low-level query

how much of the data matters?

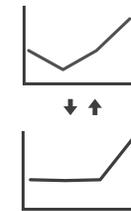
one, some, all

→ Query

→ Identify



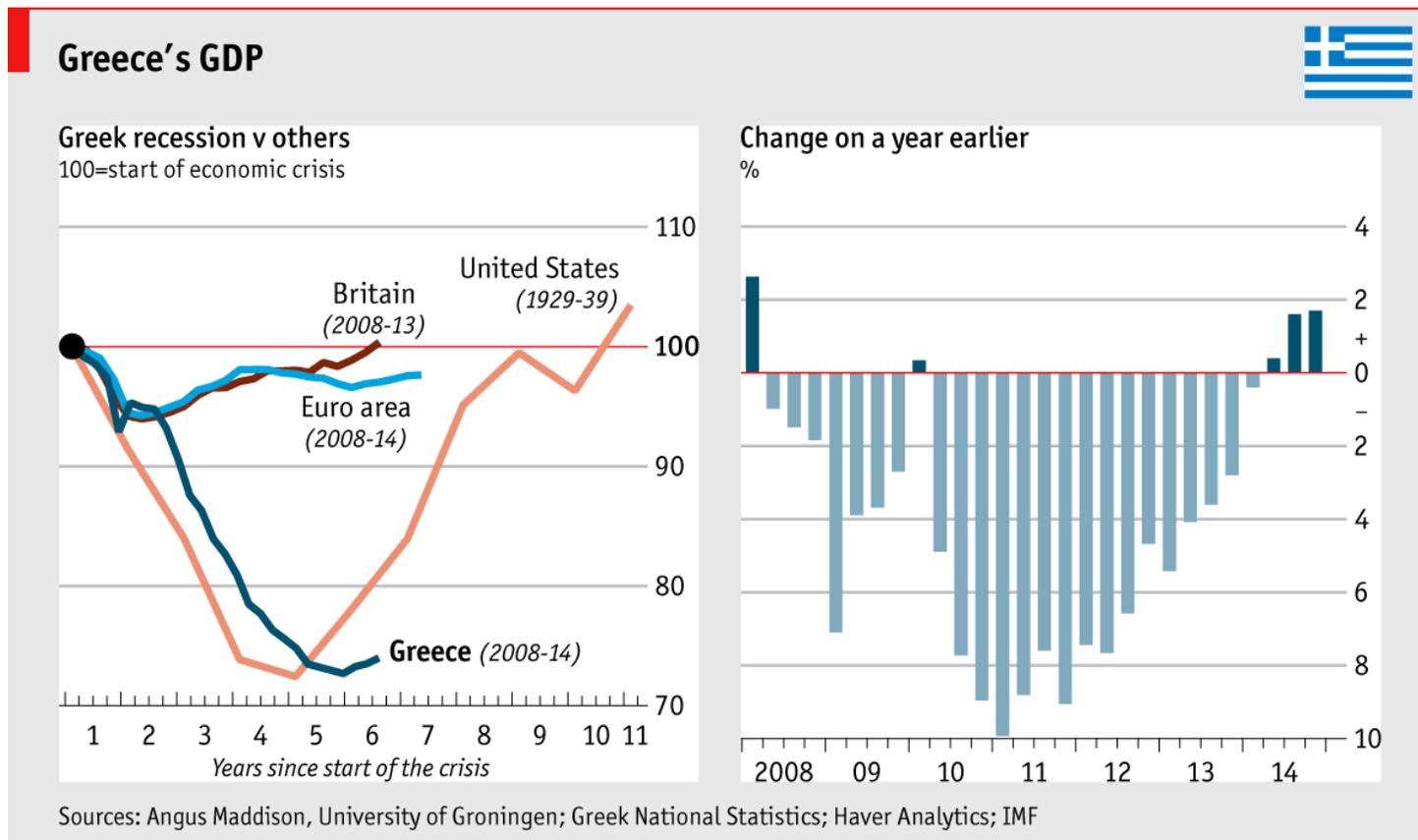
→ Compare



→ Summarize



Example Compare (& Derive)



Targets - Nouns
Actions - Verbs

Data vs Attributes

Data

- ★ Trends
- ★ Outliers
- ★ Features
- ★ Task dependent

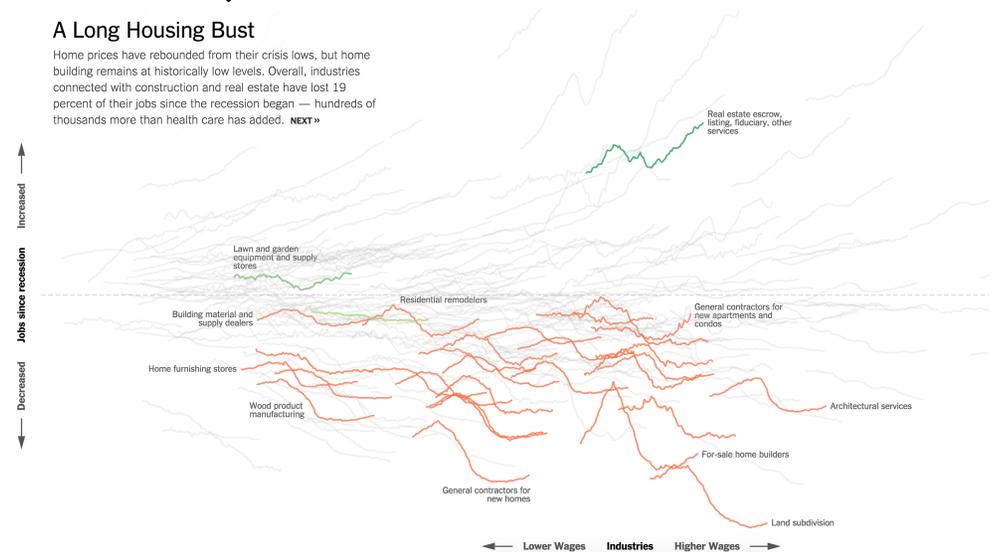
Attributes

- ★ Properties
- ★ Visually encoded
- ★ Multiple attributes
 - dependency
 - correlation
 - similarity

Examples

Trends: How did the job market develop since the recession overall?

Outliers: Looking at real estate related jobs



Why: Targets

→ ALL DATA

→ Trends



→ Outliers



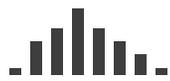
→ Features



→ ATTRIBUTES

→ One

→ *Distribution*



↓ *Extremes*



→ Many

→ *Dependency*



Minimum, Maximum

→ *Correlation*

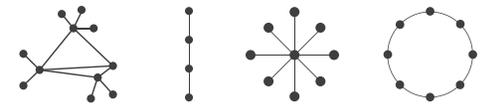


→ *Similarity*



→ NETWORK DATA

→ Topology



→ *Paths*



→ SPATIAL DATA

→ Shape



Yet Another Taxonomy
How ?

what interactions enable the tasks ?

Categories of how to manipulate visualizations:

encode

manipulate

facet

reduce

First Introduce (covered)

import

- new data items to be loaded

derive

annotate

- with text label etc. (classification)

- acts as a new attribute

record

- screenshots, bookmarks, parameter settings, logs, etc.

- graphical / use history

- analytical provenance!

How? A Preview

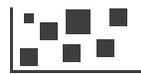
Encode

➔ Arrange

➔ Express



➔ Separate



➔ Order



➔ Align



➔ Use



➔ Map

from **categorical** and **ordered**

Manipulate

➔ Change



➔ Select



➔ Navigate



Facet

➔ Juxtapose



➔ Partition



➔ Superimpose



Reduce

➔ Filter



➔ Aggregate



➔ Embed

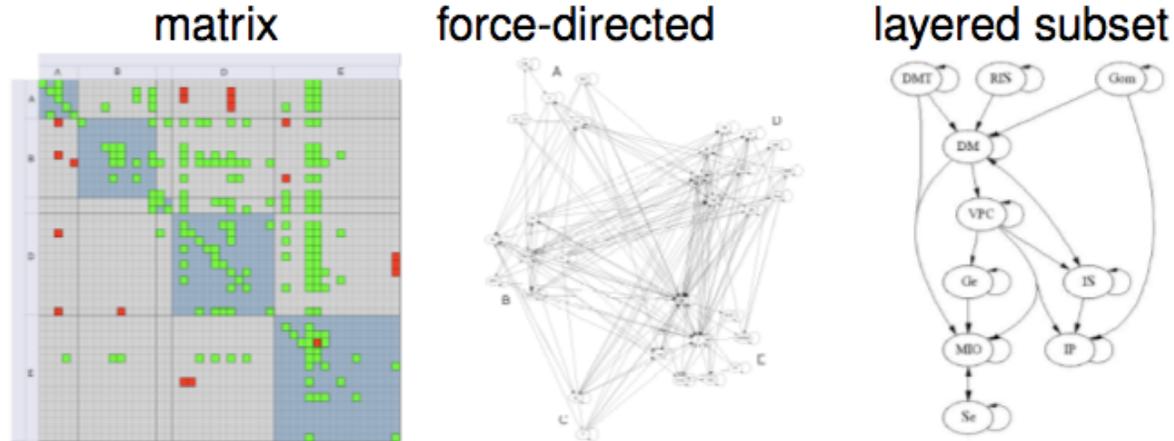


Encode

through channels and marks

e.g. color, shapes, size, position etc.

e.g. different visual encodings of a graph:



[van Ham, Using Multilevel Call Matrices in Large Software Projects. InfoVis03
<http://www.win.tue.nl/~fvham/DL/callmatrix.pdf>]

Facet

how to use views:

partition (side-by-side, simultaneously)

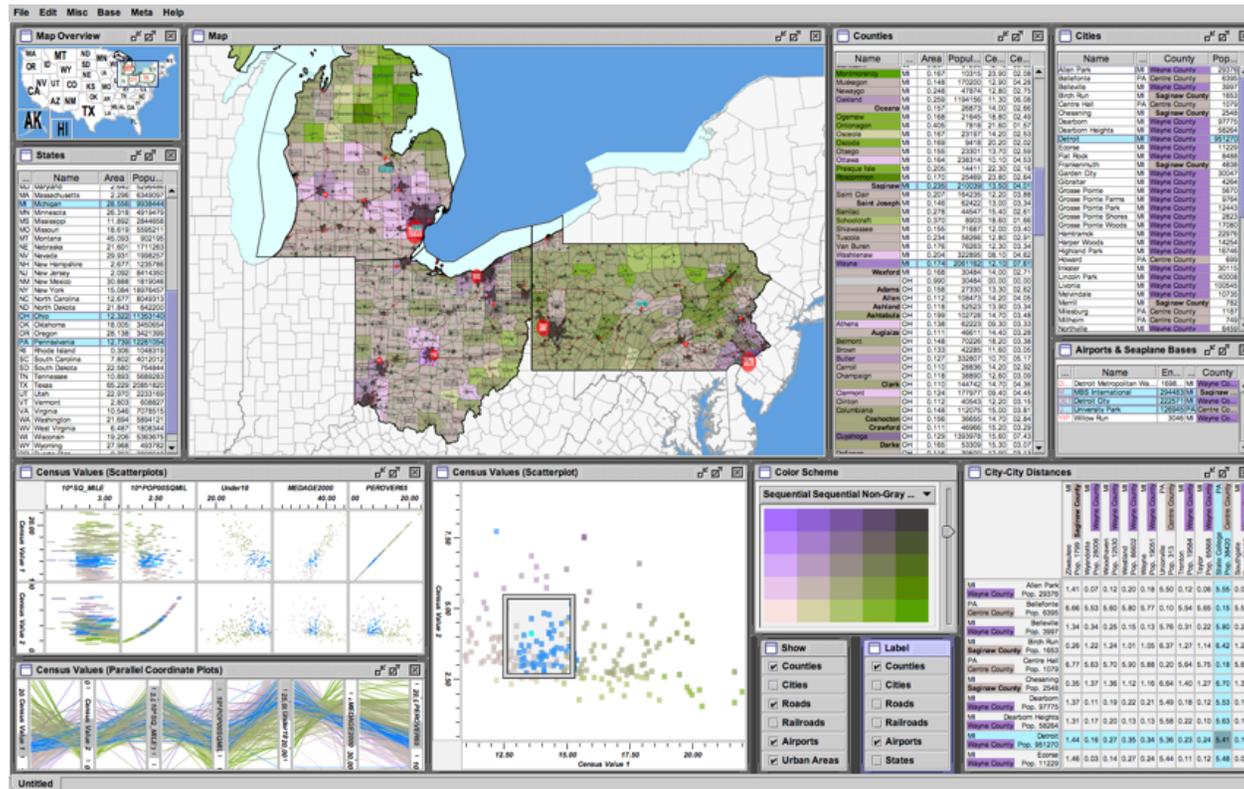
superimpose (multiple layers)

change (layout, encoding --> interaction)

select (demarcation, highlighting)

coordinate (brushing+linking, linking views)

Partition



How? A Preview

→ Map
from **categorical** and **ordered**
attributes

→ Color

→ Hue → Saturation → Luminance



→ Size, Angle, Curvature, ...

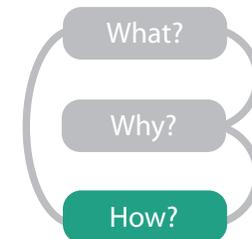


→ Shape



→ Motion

Direction, Rate, Frequency, ...



Linking (coordinate)



Reduce

reduce (increase) number of elements shown

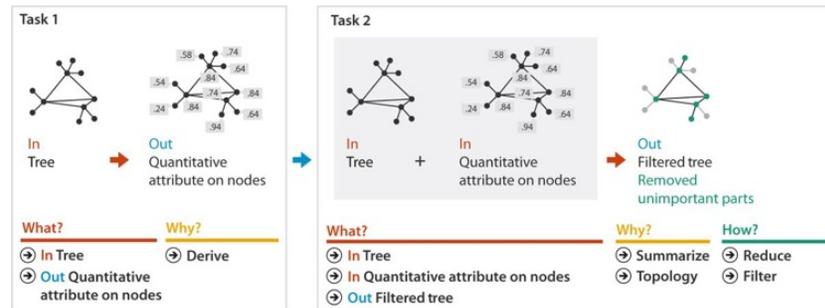
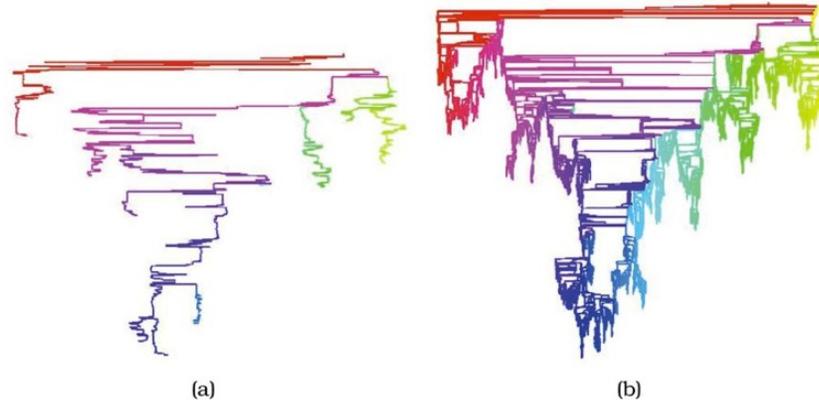
filter

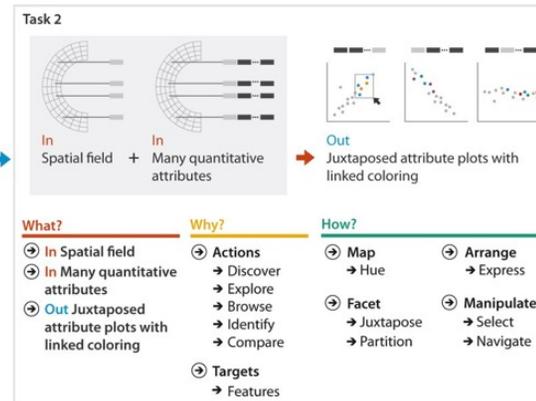
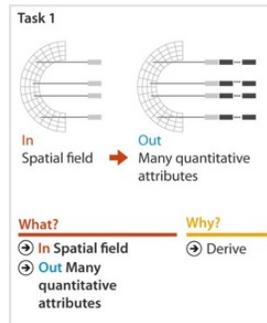
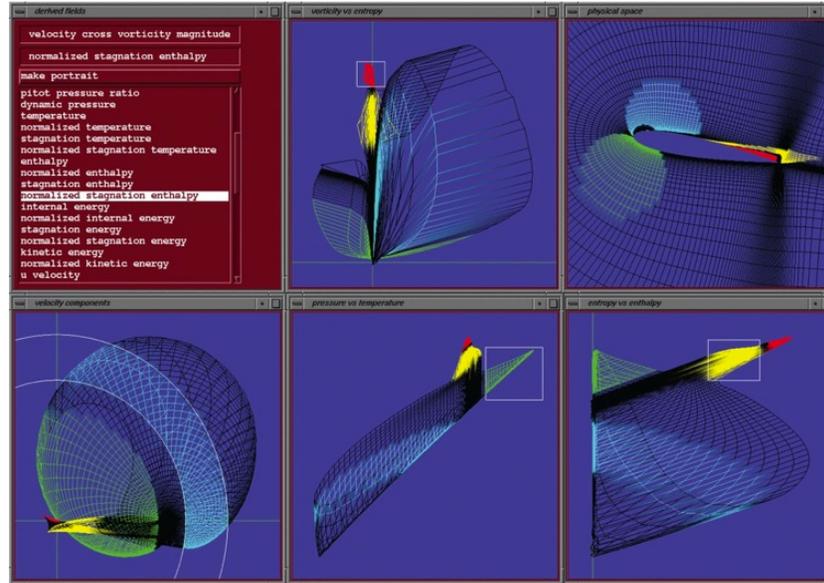
aggregate

navigate (alter viewpoint, e.g. zooming, detail-on-demand)

embed (focus+context)

compression





What?

Tree



Why?

Actions

→ Present → Locate → Identify



Targets

→ Path between two nodes



How?

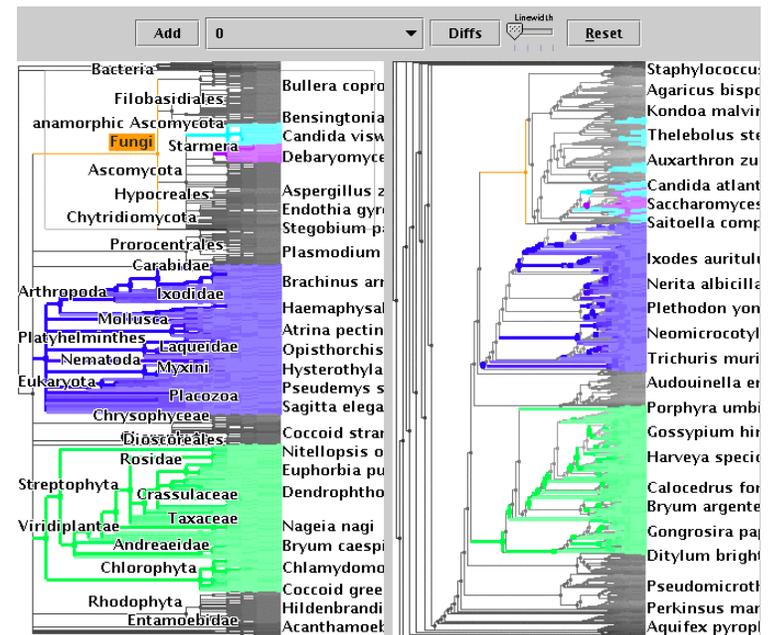
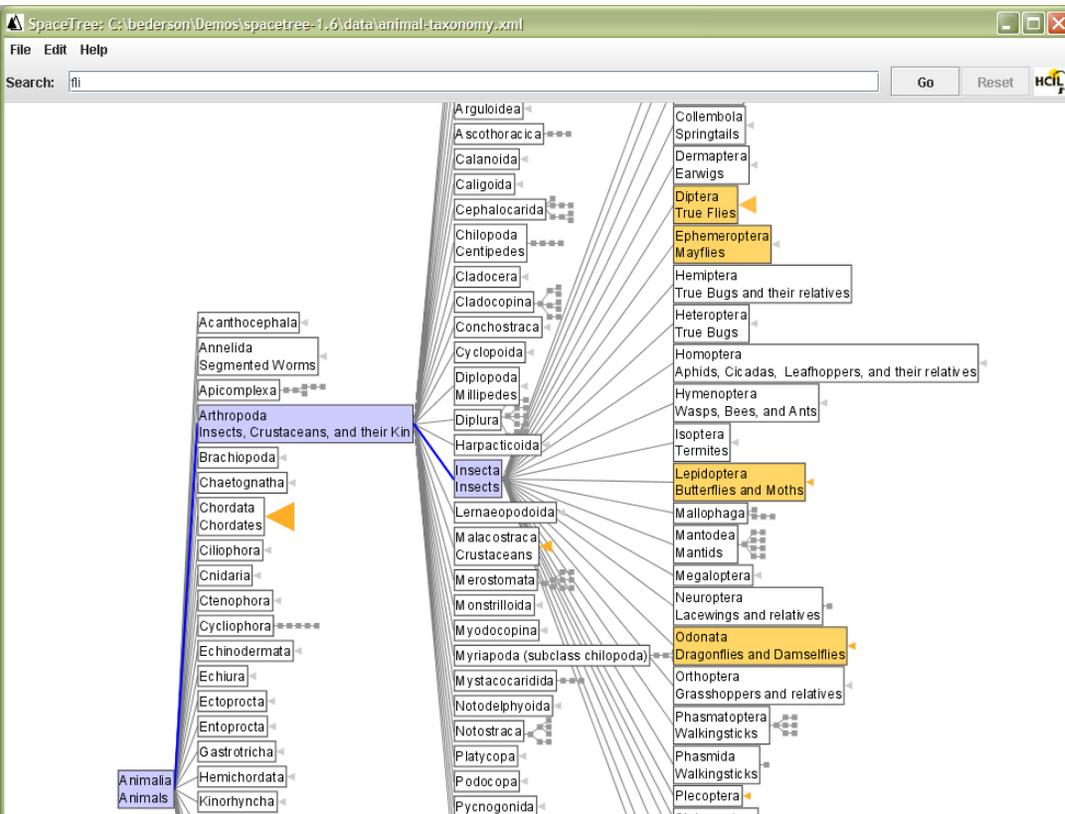
SpaceTree

→ Encode → Navigate → Select → Filter → Aggregate



TreeJuxtaposer

→ Encode → Navigate → Select → Arrange



Next Analysis