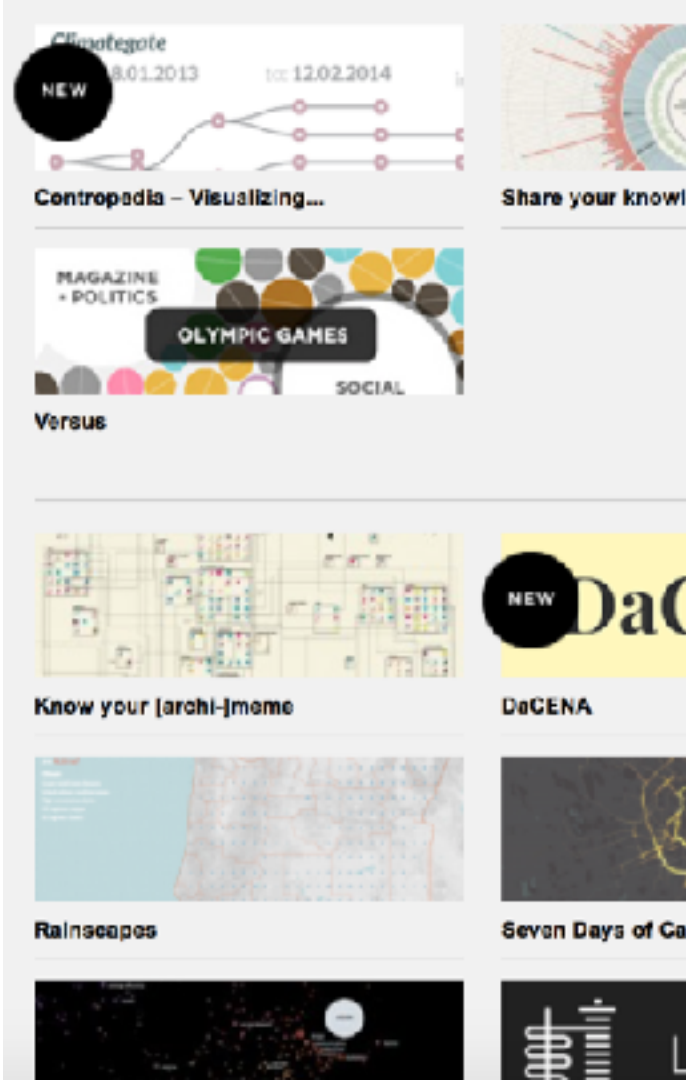
Announcement

- No regular class on Monday Feb20
- Instead, come to the lecture by **Paolo Ciuccarelli** http://www.densitydesign.org/
- ~~Monday (Feb20) at 4:30pm~~
- **Tuesday (Feb21) at 5pm**
- CFA Building Room 111



Design Critique

code_swarm: A Design Study in Organic Software Visualization

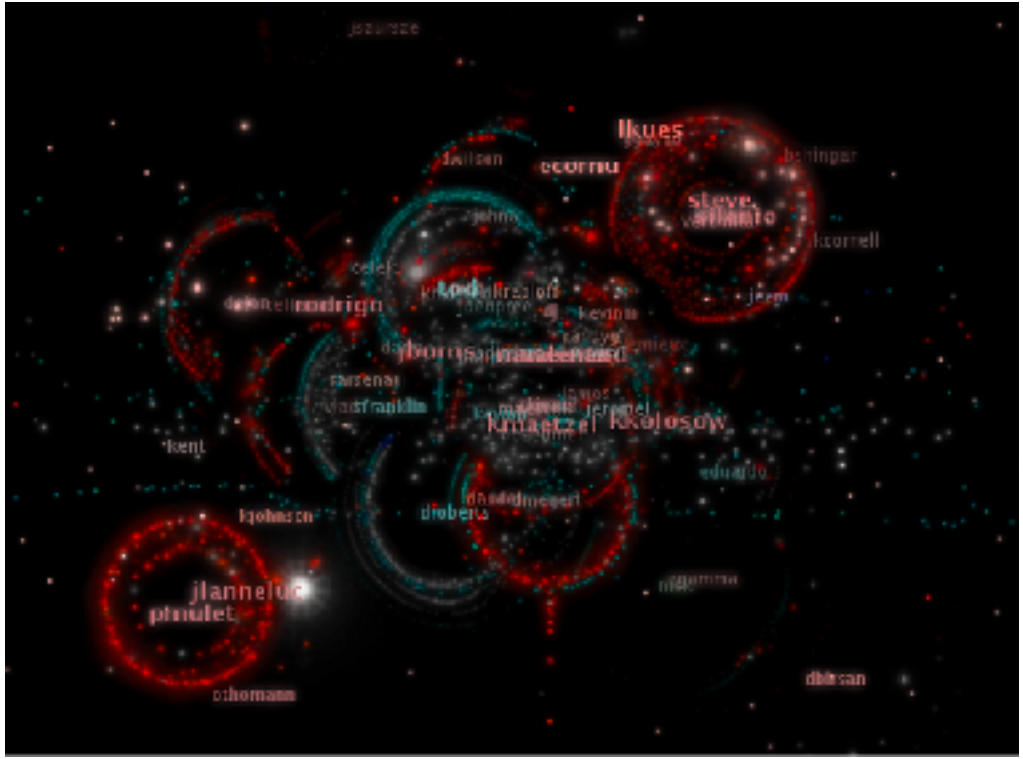
by Michael Ogawa and Kwan-Liu Ma
University of California, Davis

CodeSwarm

http://bit.ly/vis-codeswarm

In your breakout group, discuss the following questions:

- Who is the intended audience?
- What questions does this visualization answer?
- What data is represented in this visualization? Be specific.
- How is each data type visually encoded?
- What could be the goals of the designer?
- Can you read the data precisely?
- This visualization was much publicized in the open source community. Why do you think that happened?
- Why do you like / dislike this visualization?



Visualization in HCI

05-499/05-899 Section C

Carnegie Mellon University Human-Computer Interaction Institute

Data

February 15, 2017

What is Data?

Where does it come from?

What does it represent?

3 Key Activities …

Data collection/generation.

Data transformation/processing.

Data analysis and visualization.

Data Collection/Generation

Big topic! Not addressed in this course.

Many data collection methods (e.g., sensors, logs, experiments, human-generated data, surveys)

Many many data sources available on the internet

Data Transformation/Processing

Big topic! Not fully addressed in this course.

Data transformed in many ways (aggregated, collated, subsetted, filtered, reshaped, change of scale, etc.)

Data Analysis/Exploration

Being able to familiarize with your data: investigate questions you may have or generating new questions is a crucial skill!

This process is called “Exploratory Data Analysis (EDA)”.

One major goal of visualization is to support EDA.

Data Abstraction

Provide a language to describe data in a way that is meaningful and useful to visualization design.

e.g., *is it a table, a network, or unstructured data?*

e.g., *is it spatial (geographical), temporal, or spatio-temporal?*

Why Data Abstraction?

Data is typically described with *domain language*.

But in order to search for suitable visual representations we need to translate them into more abstract structures we know how to encode.

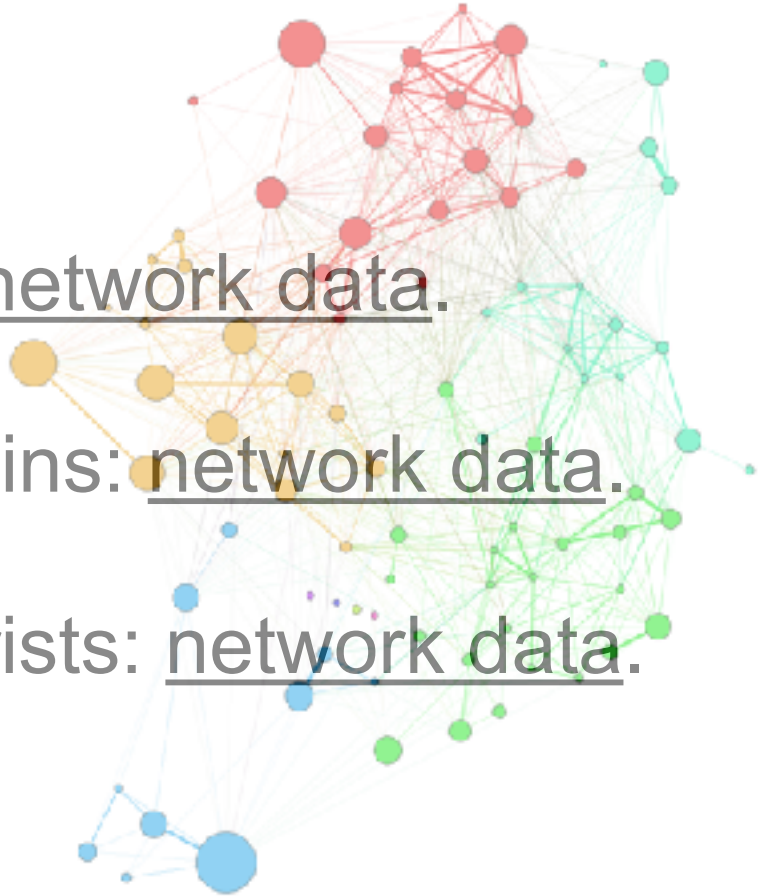
In turn, this narrows down the design space.

Example

Friendships in Facebook: network data.

Interactions between proteins: network data.

Connection between terrorists: network data.



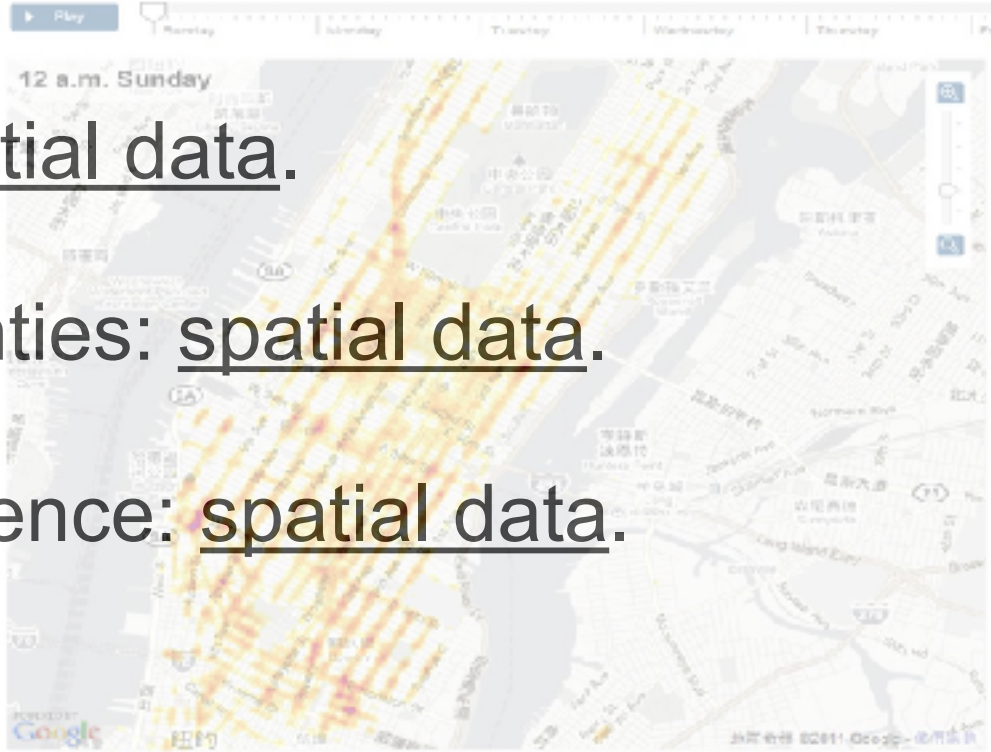
Visit Visual Complexity to see a very large number of examples of network visualizations.

Example

Animal movement: spatial data.

Election results in counties: spatial data.

Simulation of air turbulence: spatial data.



Data Semantics

Basil, 7, S, Pear

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

Data Types

Structural or mathematical interpretation of data

➔ **Data Types**

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

Items & Attributes

Item: individual entity, discrete

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

“independent variable”

Attribute: measured, observed, logged property

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

“dependent variable”

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

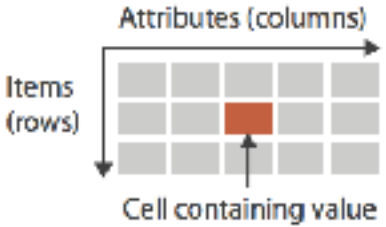
- Spatial data -> location in 2D or 3D
- Pixels in photo, Voxels in MRI scan, latitude/longitude

Grids

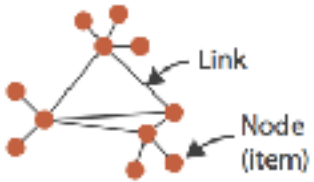
- Sampling strategy for continuous data
- How many Voxels in MRI scan

Dataset Types

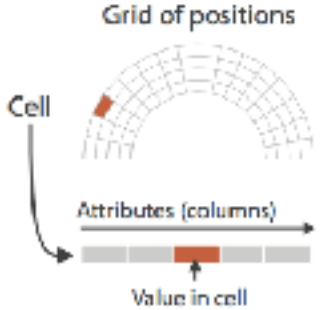
→ Tables



→ Networks

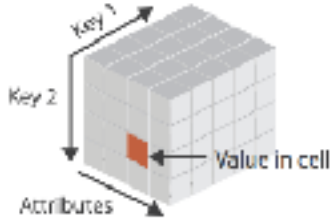


→ Fields (Continuous)



→ Geometry (Spatial)

→ *Multidimensional Table*



→ *Trees*

Tables

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

Flat Table

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

Multidimensional Table

- indexing based on multiple keys

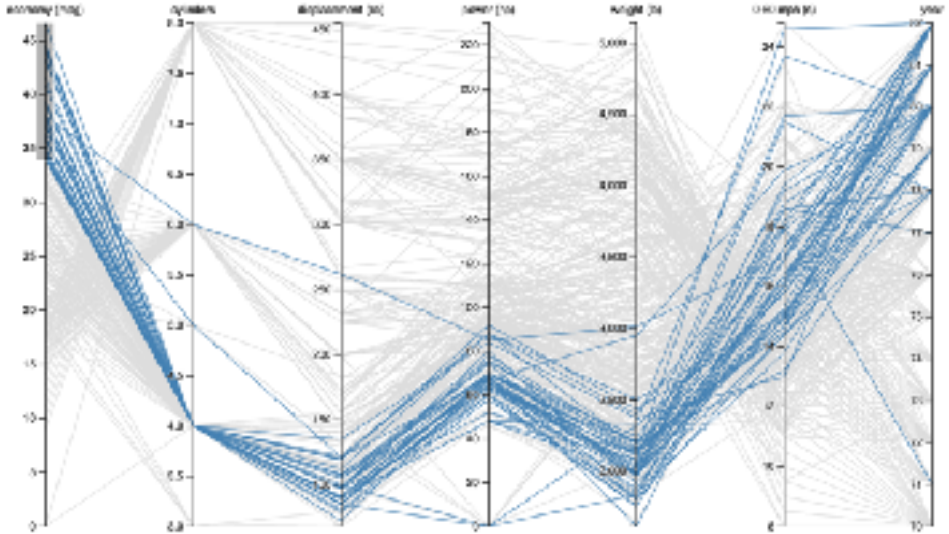
Multidimensional Tables

Keys: Genes

	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	-1.729892883
10	DCX	DCX	-2.2007897G	-2.529795801	-2.844966273
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	1.025827901
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	2.369540659
19	MBP	MBP	-1.906390566	-2.037626447	-2.935744905
20	TF	TF	-4.334123292	-4.680680246	-2.975788865
21	KCND2	KCND2	-1.777692395	-2.100162021	-1.996306032
22	GABRB1	GABRB1	-2.214760175	-3.022654105	-3.185499425

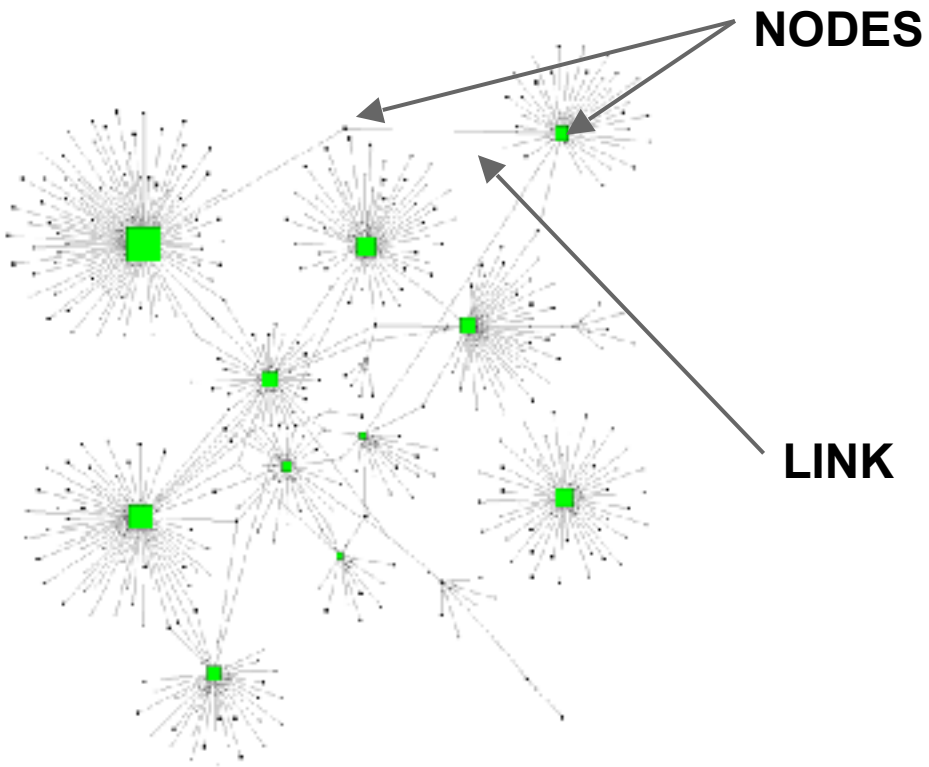
Keys: Patients

Visualizing Tables



More in Lecture on Tables & High-Dimensional Data

NETWORKS AND TREES

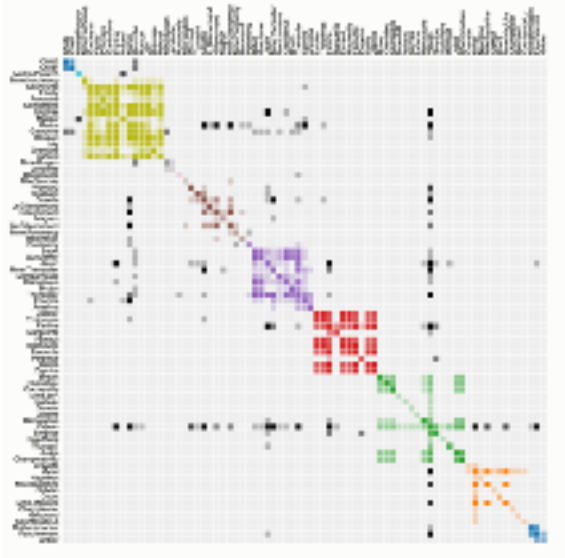


Trees have a hierarchical structure where each node has only one parent.

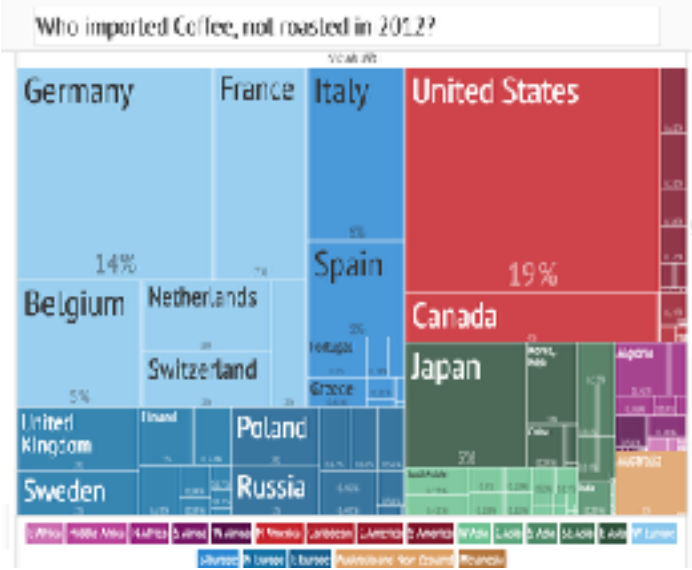
Nodes and links can also have attributes, e.g., in a social network each node can be characterized by *gender*, *age*, *income*.

Visualizing Graphs

Node-Link Diagram



Matrix

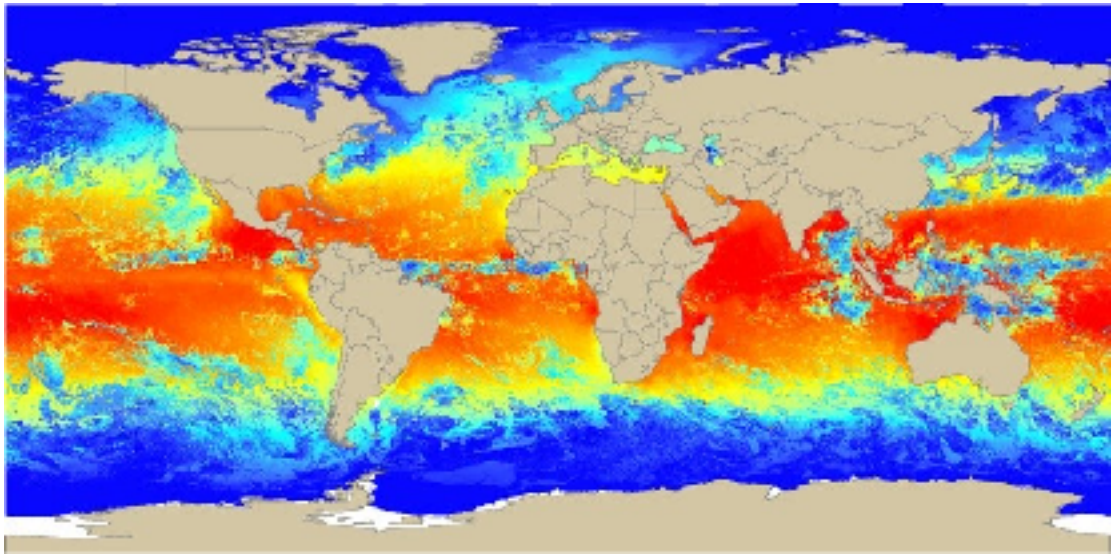


Treemap (Implicit Tree Visualization)

More in Lecture on Graphs & Trees

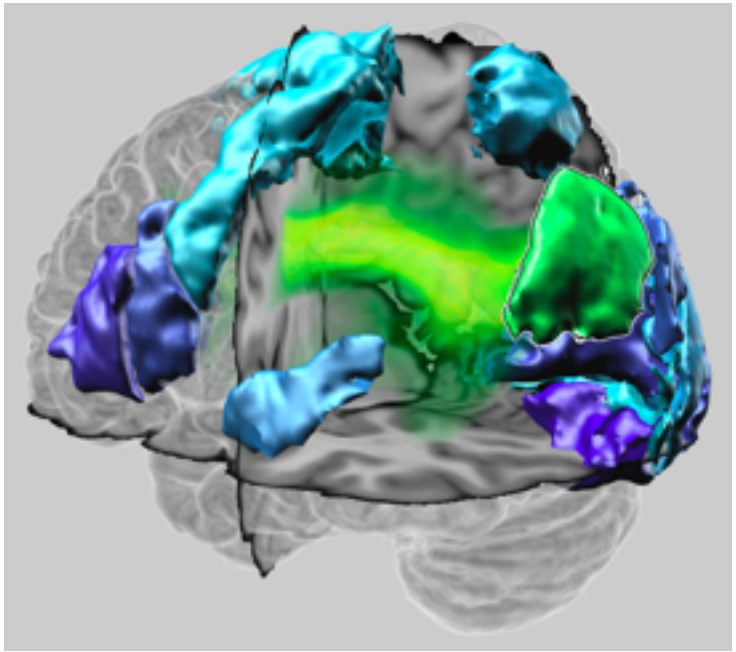
CONTINUOUS FIELDS

Measurements from a CONTINUOUS DOMAIN



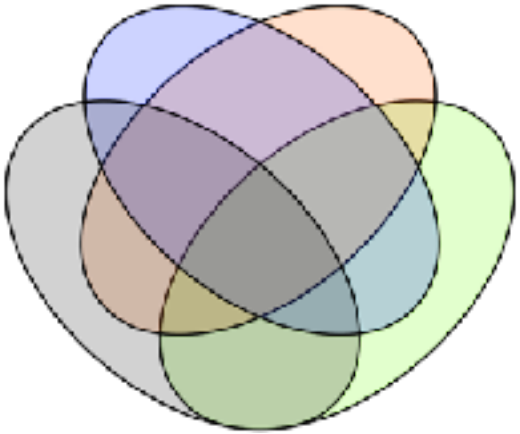
Being a SPATIAL FIELD has far-reaching implications on what visualization *idioms* are available and suitable.

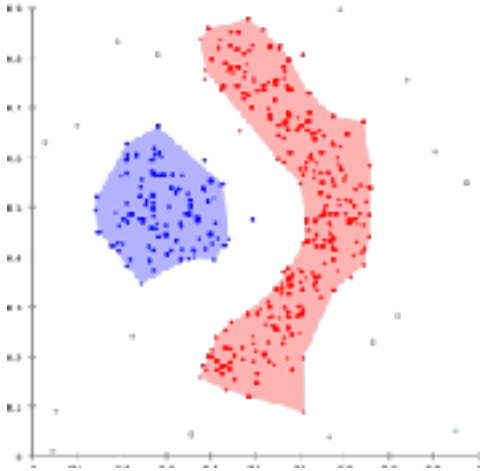
GEOMETRY




Beijing
The Great Wall
Yellow Sea
Daning River
Nanjing
Suzhou
Shanghai
CHINA
Xi'an
Three Lesser Gorges
Wuhan
Yangtze
Tongli
Yichang
East China Sea
Chengdu
Wushan
Huangshan
Daizu
Three Gorges Area
Chongqing
Yangtze
Tibet
Lhasa
Guilin
Hong Kong

CLUSTERS, SETS, LISTS (Collection of items)





Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Availability

- Static

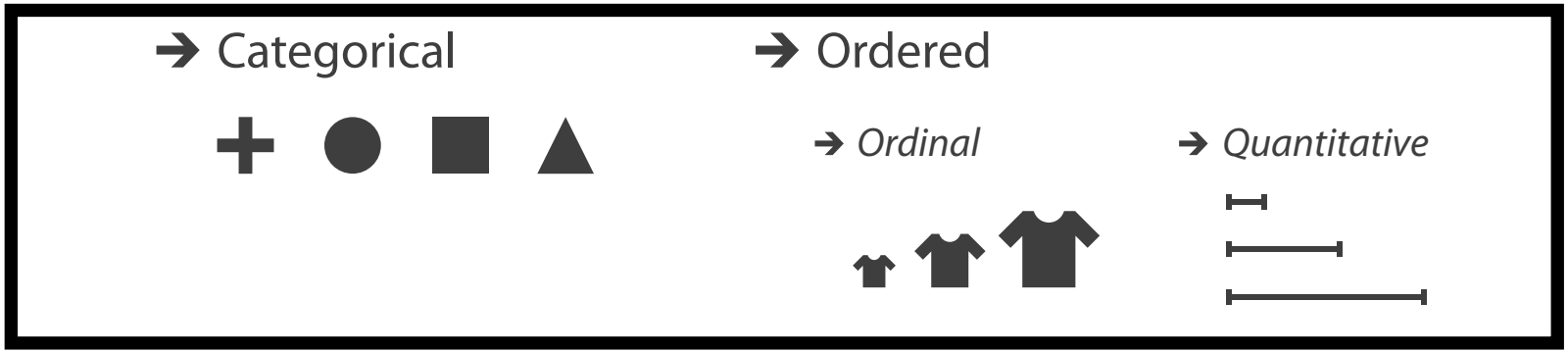
- Dynamic

Attribute Types

Categorical (e.g., gender, race, eye color)

Ordinal (e.g., edu level, position in a race)

Quantitative (e.g., age, height, weight)



	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		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		2/22/08
4	32	7/16/07	2-High	Small Pack		7/17/07
5	32	7/16/07	2-High	Jumbo Box		7/17/07
6	32	7/16/07	2-High	Medium Box		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		10/25/07
10	36	11/3/07	1-Urgent	Small Box		11/3/07
11	65	3/16/07	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	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

	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	.5	2/22/08
4	32	7/16/07	2-High	Small Pack	.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
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	3/16/07	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		5/11/08
23	132	6/11/06	3-Medium	Medium Box		6/12/06
24	132	6/11/06	3-Medium	Jumbo Box		6/14/06
25	134	5/1/08	4-Not Specified	Large Box		5/3/08
26	135	10/21/07	4-Not Specified	Small Pack		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		

Categorical

Ordinal

Quantitative

Ordering Direction

- Sequential

- Diverging

- Cyclic

Sequential: e.g., age, height, weight.

Diverging: e.g., temperature, altitude.

Cyclic: e.g., hour, week, year.

Hierarchical Attributes

Some attributes may have an internal hierarchical structure

For instance: dates, spatial regions, taxonomies.

Derived Data

It is almost never the case that the data you receive will be visualized without manipulation.

There are many ways data can be transformed to better fit the task at hand.

And it the designer’s responsibility to choose which data structures and segments work best.