Dr. Neil Smith

Information Visualization Concepts

Education:

- PhD. in Anthropology, University of California, San Diego, 2009.

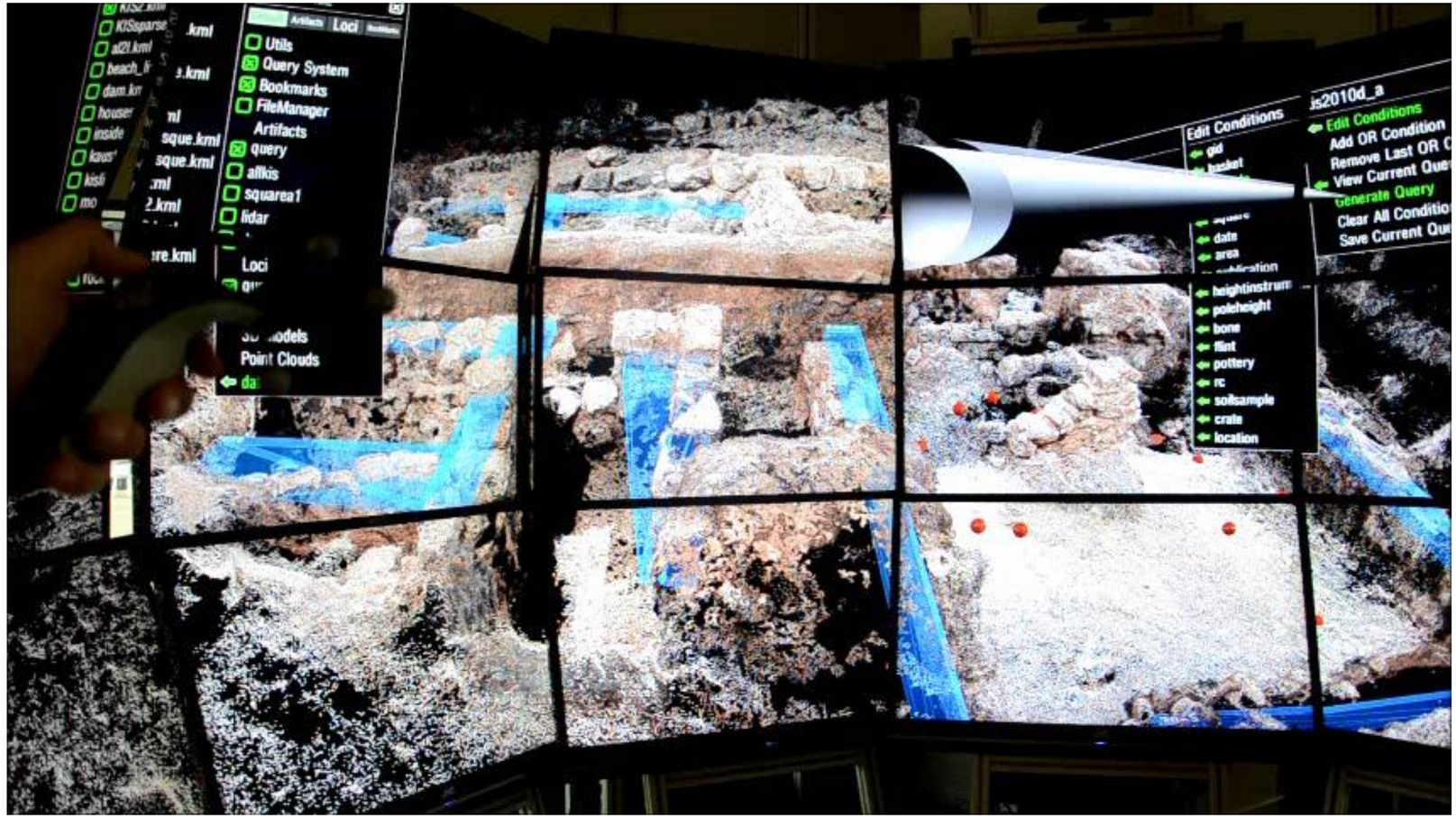
Professional Profile:

- Director Convergent Visualization Center, Qi
- Co-Directory of the Center for Cyber-Archaeology and Sustainability 2022-present
- Chief Information Officer (CIO), FalconViz LLC, 2019-present
- Chief Executive Officer (CEO), FalconViz LLC, 2014-2019
- Research Scientist, VCC, King Abdullah University of Science and Technology (KAUST), 2013-2019.

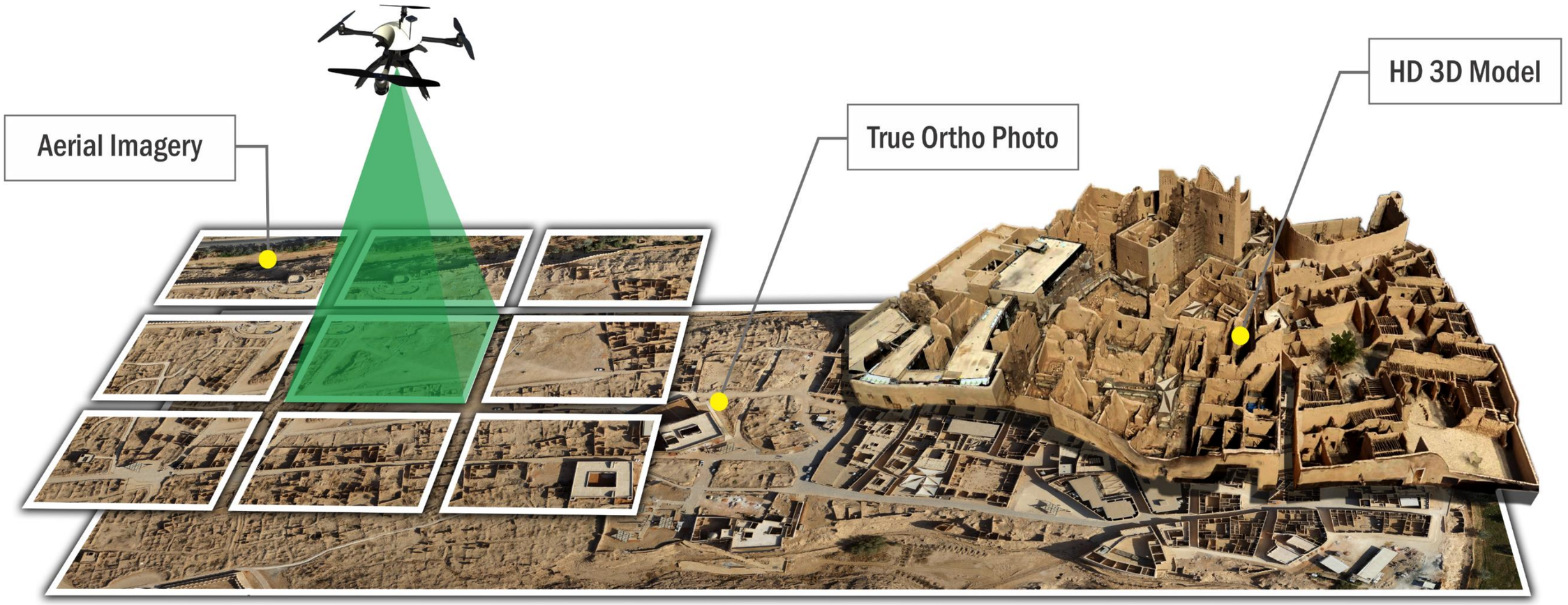

ArtifactVis (2008-2011)
Jurgen Schulze
Kyle Knabb
Connor Defanti
Thomas Levy
ArtifactVis2 (2011-)
Neil Smith
Jurgen Schulze
Connor DeFanti
Marcin Rogowski
Thomas Levy

ArtifactVis 2 (Presented in DHIC 2013

- Fully Immersive 4D GIS
- Fully Integrates all Digital Content recovered from Excavations
- Real-time Query System for conducting Multi-Variate Studies
- Scalable to Various Computer Architectures
- Part of CalVR Middle-ware allowing full exploitation of OpenGL (GLSL), CUDA, PhysX and other libraries



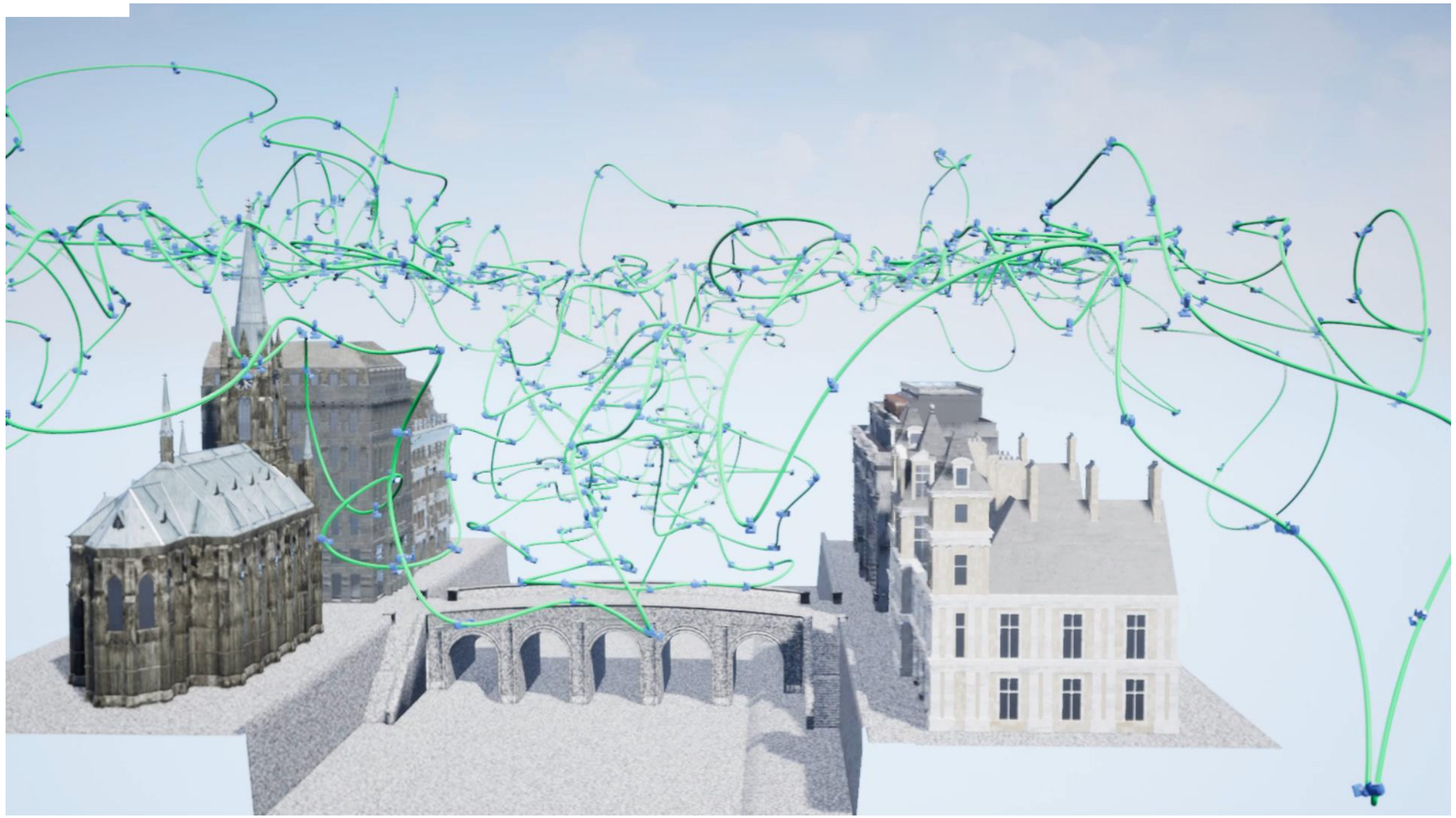
END TO END SOLUTION



From Planning ----------→to Visualizing


FALCONVIZ

AERIAL PATH PLANNING

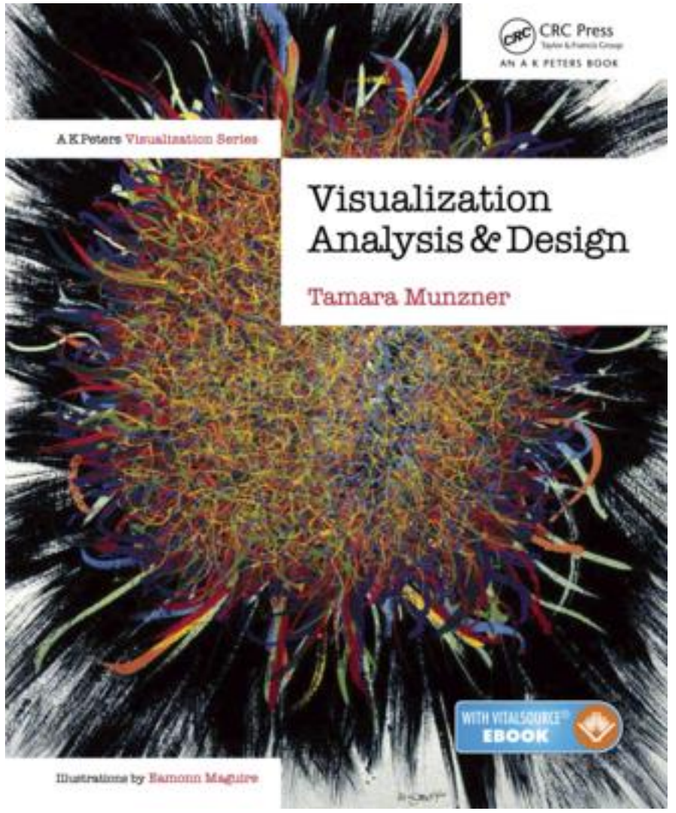


TEACHING AUTONOMOUS DRIVING AND DRONES HOW TO RACE

MetaHumans Preview [NetMode: Standalone 0] (64-bit/D3D Shader Model 5)
{ab561f08-92e1-419c-a182-8509f82992be} CABLE Output (VB-Audio Virtual Cable)
Default Device
2

Visualization Analysis & Design

Tamara Munzner
Department of Computer Science
University of British Columbia



@tamaramunzner

Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?...

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

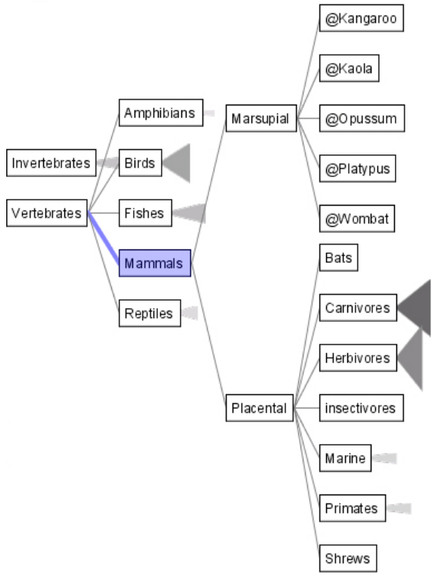
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- don't need vis when fully automatic solution exists and is trusted
- many analysis problems ill-specified
  - don't know exactly what questions to ask in advance
- possibilities
  - long-term use for end users (ex: exploratory analysis of scientific data)
  - presentation of known results (ex: New York Times Upshot)
  - stepping stone to assess requirements before developing models
  - help automatic solution developers refine & determine parameters
  - help end users of automatic solutions verify, build trust

Why analyze?

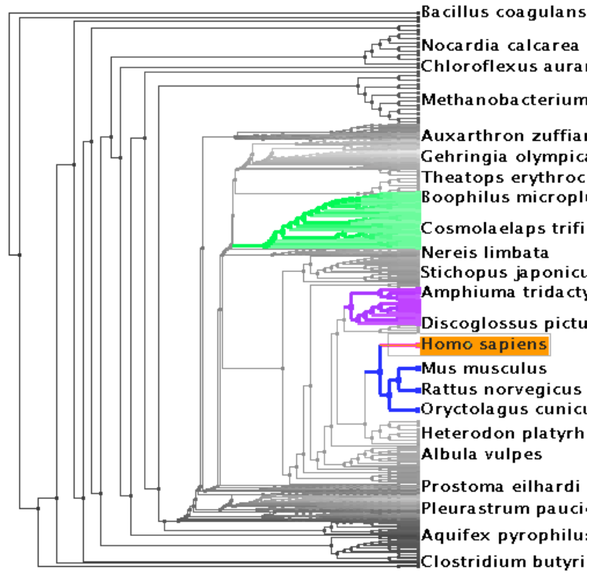
- When you're confronted with a vis problem as a designer, it can be hard to decide what to do?
- Analysis imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

SpaceTree



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

TreeJuxtaposer



[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]

Expressiveness and Effectiveness main focus of analysis

Idiom: choosen visual encoding and interaction

Expressiveness: Does the idiom match the data and task one wants to achieve?

- Out of all “Design Choices” the best expression of the data is ….
- The decision on the Interaction idiom can be as important as the visual encoding idiom.
- The **channel** selected is the correct match to the data (e.g. the color channel is categorical and best when showing two different categories on a scatterplot) (e.g. magnitude channel correct match for ordered attributes that express ordinal/quantitative data)

Why Effectiveness and Expressiveness main focus of analysis

Effectiveness:

Some idiom/channels/interactions are more effective for humans to perform the desired task.

(Bar graphs which use spatial and line channel much more effective then pie charts using area channel and angle)

Channels effectiveness can be measured by their human perceived <u>accuracy, discriminability, separability, and popout</u>

Computer-based visualization systems provide visual representations of datasets designed to help people <u>carry out tasks more effectively</u>. (Effectiveness)

Why Effectiveness and Expressiveness main focus of analysis

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays. (Effectiveness)

Effectiveness addresses three constraints:

- computational limits
  - computation time, system memory
- display limits
  - pixels are precious & most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space
    - find sweet spot between dense and sparse
- human limits
  - human time, human memory, human attention

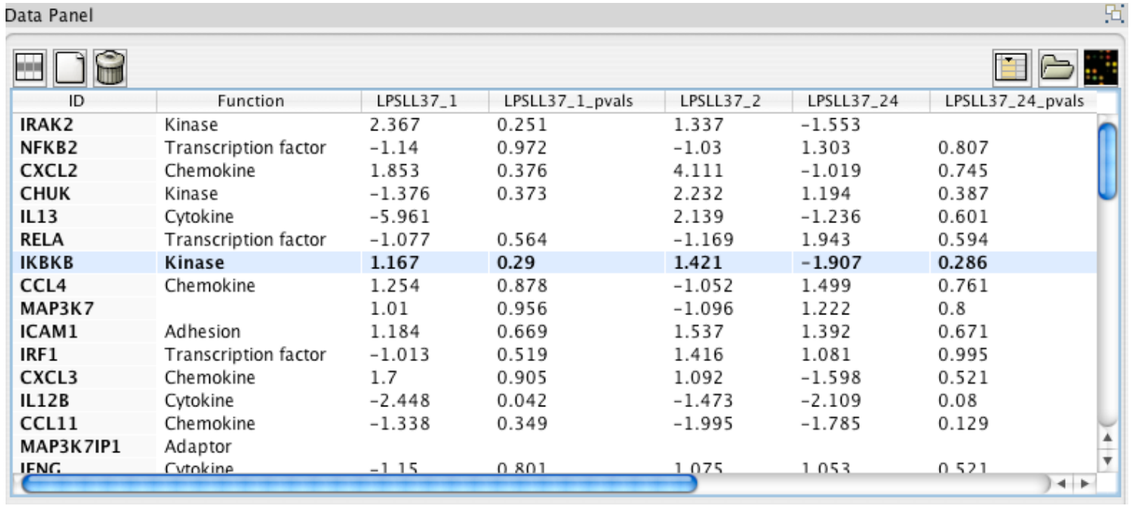
Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception
- Purpose is to speed up:

Search (find groups of importance)

Recognition (group relevant items in same location so don't have to remember)

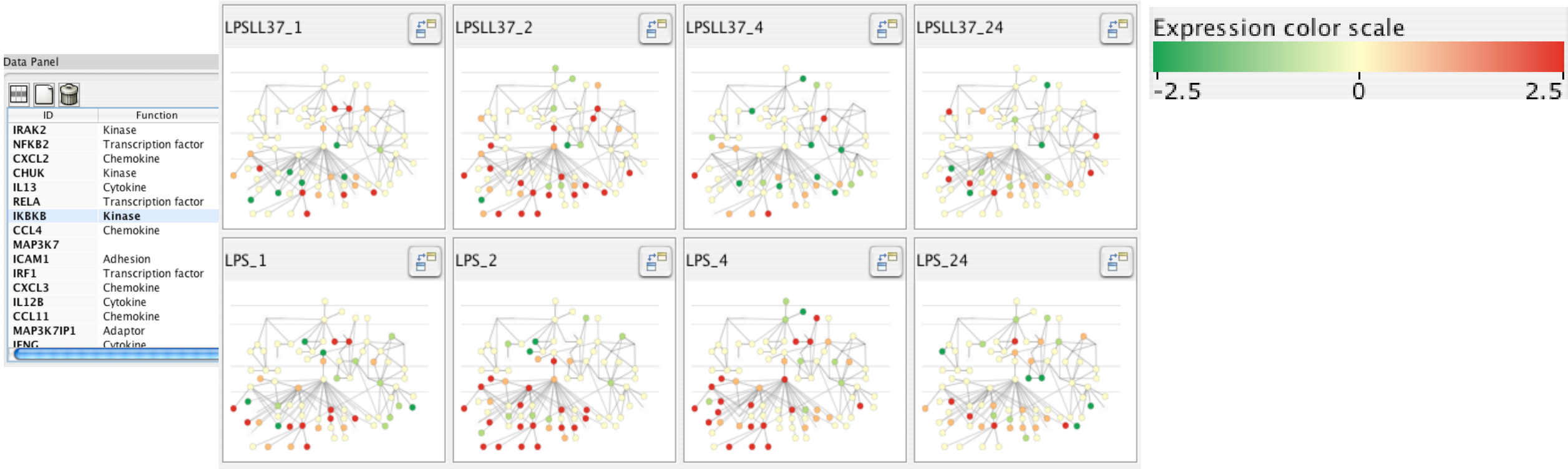


[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception



[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why depend on vision?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- human visual system is high-bandwidth channel to brain
  - overview possible due to background processing
    - subjective experience of seeing everything simultaneously
    - significant processing occurs in parallel and pre-attentively
- sound: lower bandwidth and different semantics
  - overview not supported
    - subjective experience of sequential stream
- touch/haptics: impoverished record/replay capacity
  - only very low-bandwidth communication thus far
- taste, smell: no viable record/replay devices

Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

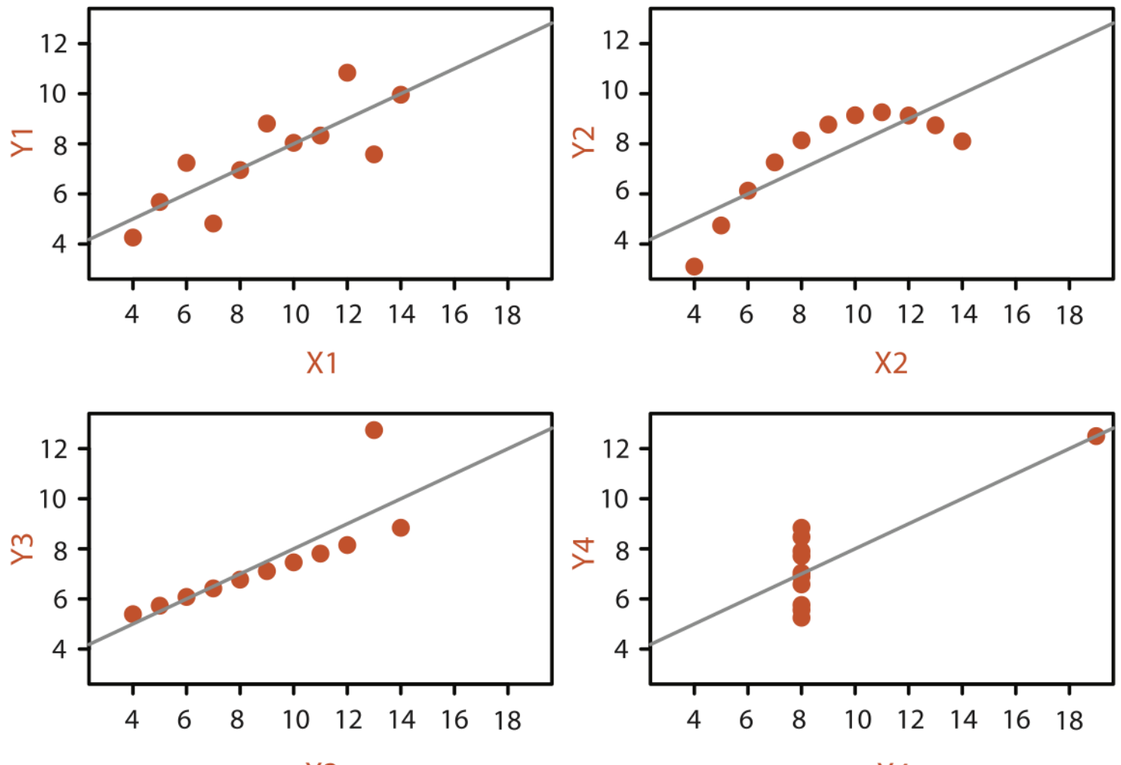
Anscombe's Quartet

Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816

Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model



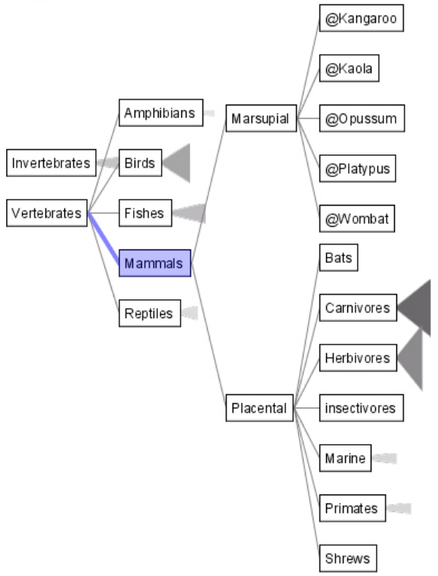
Anscombe's Quartet

Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816

Why analyze?

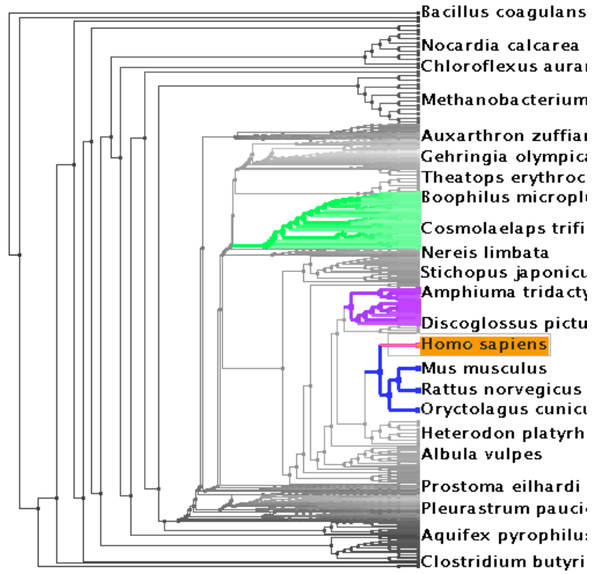
- What? (Data User needs to)
- Why? (Why use this vis tool)
- How? (How they are constructed based on design choices)

SpaceTree



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

TreeJuxtaposer



[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]

What?

- Tree

Why?

- Actions
  - → Present → Locate → Identify



- Targets
  - → Path between two nodes

How?

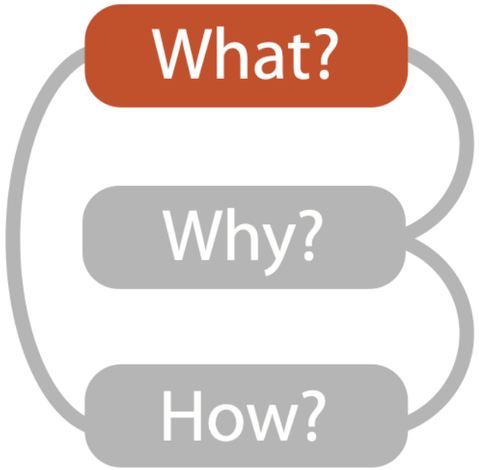
- SpaceTree
  - → Encode → Navigate → Select → Filter → Aggregate





- TreeJuxtaposer
  - → Encode → Navigate → Select → Arrange



What is to be Visualized?

What?

Datasets

Data Types

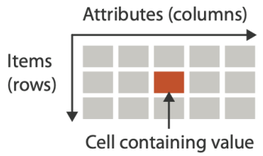
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

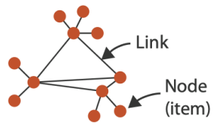
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

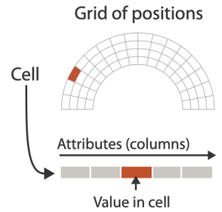
- Tables



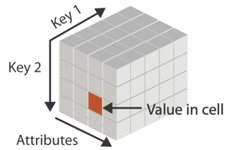
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

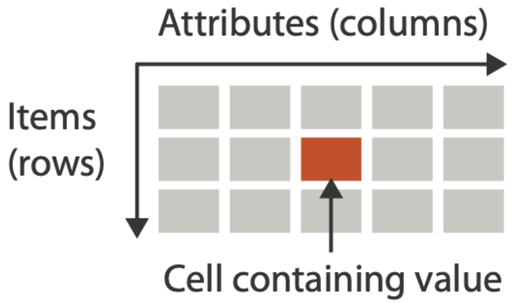
- Cyclic

Four major datatypes

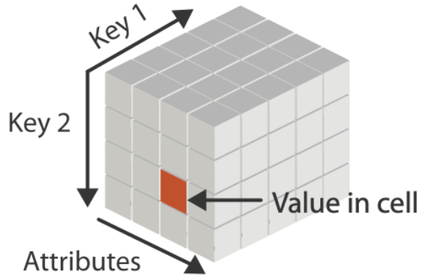
Four major datatypes

Dataset Types

- Tables



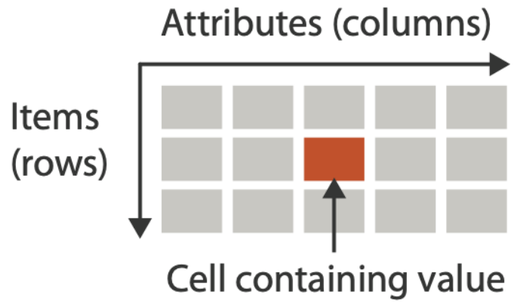
- *Multidimensional Table*



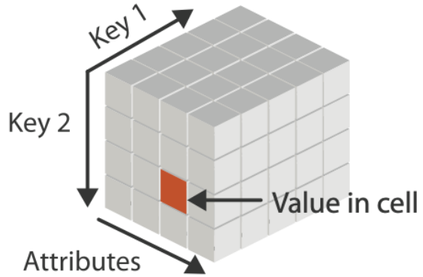
Four major datatypes

Dataset Types

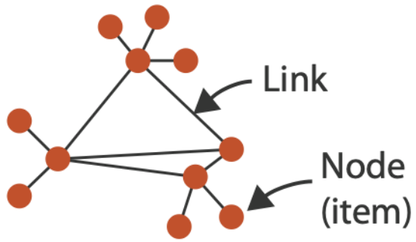
Tables



Multidimensional Table



Networks

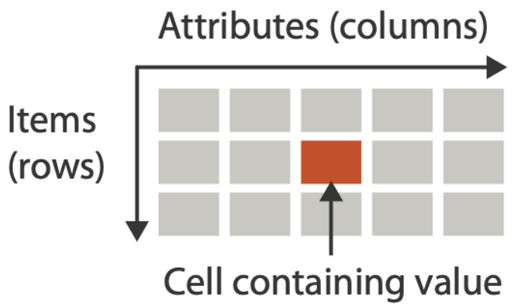


Trees

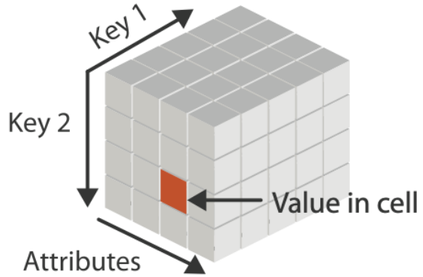
Major datatypes

Dataset Types

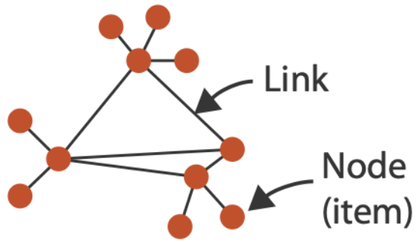
Tables



Multidimensional Table



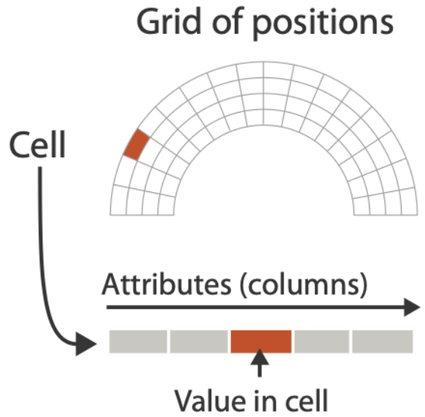
Networks



Trees

Spatial

Fields (Continuous)



Geometry (Spatial)

Attribute types

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

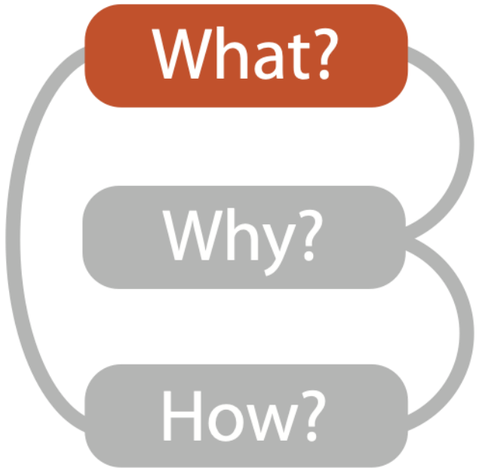
  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic



What?

Datasets

Data Types

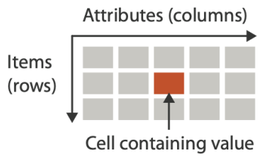
→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

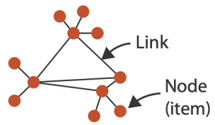
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

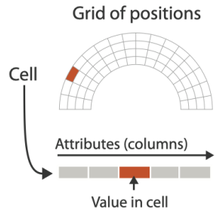
→ Tables



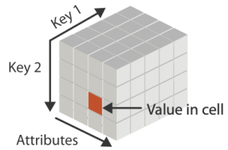
→ Networks



→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*

→ Geometry (Spatial)

Dataset Availability

→ Static

→ Dynamic

Attributes

Attribute Types

→ Categorical

→ Ordered

→ *Ordinal*

→ *Quantitative*

Ordering Direction

→ Sequential

→ Diverging

→ Cyclic

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

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

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

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

item: person

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

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

item: person

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

attributes: name, age, shirt size, fave fruit

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

item: person

Other data types

- links
  - express relationship between two items
  - eg 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

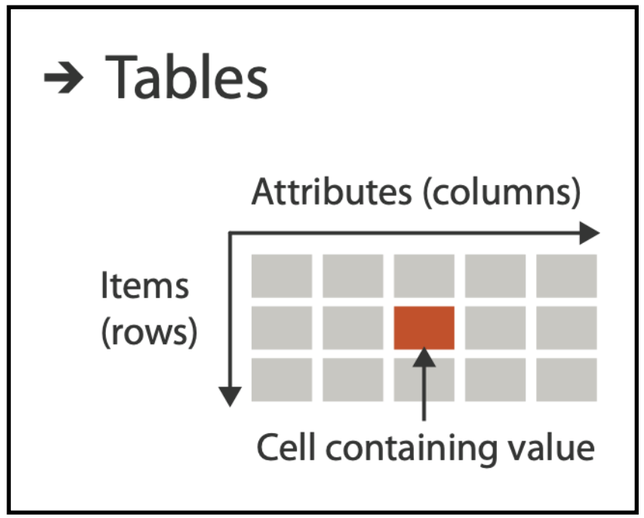
Dataset types

Tables

Items

Attributes

- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair



attributes: name, age, shirt size, fave fruit

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

item: person

Dataset types

Tables

Items

Attributes

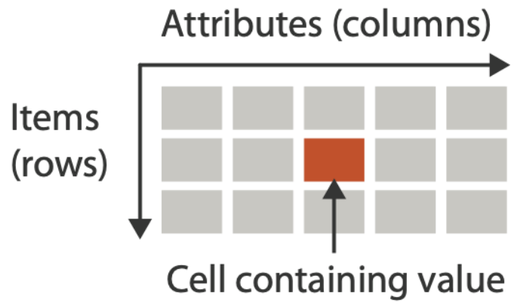
- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair
  - unique key (could be implicit)

attributes: name, age, shirt size, fave fruit

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

item: person

➔ Tables



Table

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

item

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

item

attribute

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

item

cell

attribute

Attribute types

- which classes of values & measurements?
- categorical (nominal)
  - compare equality
  - no implicit ordering
- ordered
  - ordinal
    - less/greater than defined
  - quantitative
    - meaningful magnitude
    - arithmetic possible

categorical

ordinal

quantitative

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06

Other data concerns

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

Dataset Availability

- Static

- Dynamic

Dataset types

Tables

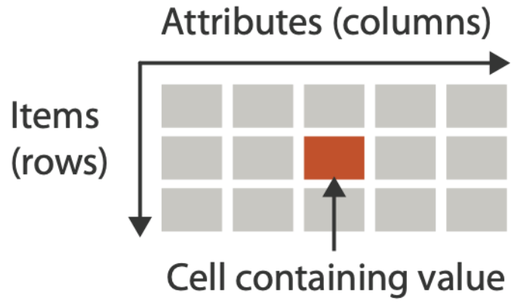
Items

Attributes

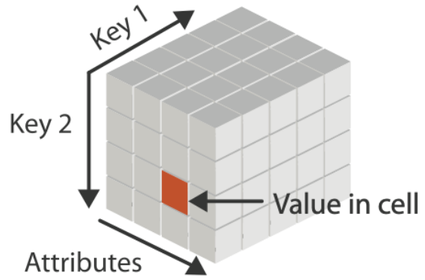
- multidimensional tables
  - –indexing based on multiple keys
    - •eg genes, patients

	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

➔ Tables



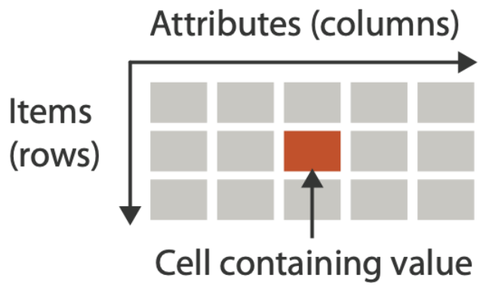
➔ *Multidimensional Table*



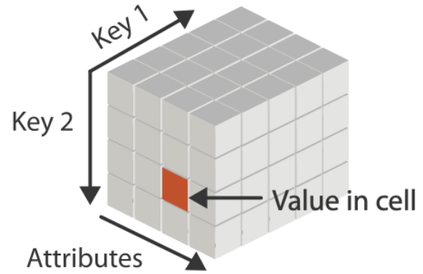
Keys and values

- key
  - independent attribute
  - used as unique index to look up items
  - simple tables: 1 key
  - multidimensional tables: multiple keys
- value
  - dependent attribute, value of cell

➔ Tables



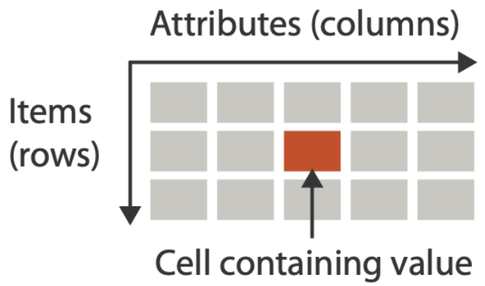
➔ *Multidimensional Table*



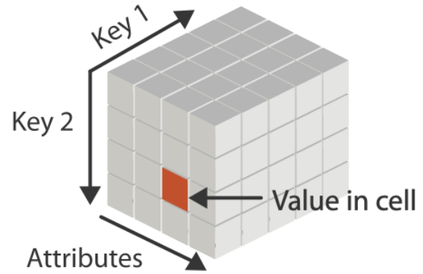
Keys and values

- key
  - independent attribute
  - used as unique index to look up items
  - simple tables: 1 key
  - multidimensional tables: multiple keys
- value
  - dependent attribute, value of cell
- classify arrangements by keys used
  - 0, 1, 2, ...

➔ Tables



➔ *Multidimensional Table*



➔ *0 Keys*

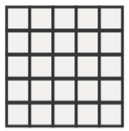
➔ **Express Values**

➔ *1 Key*

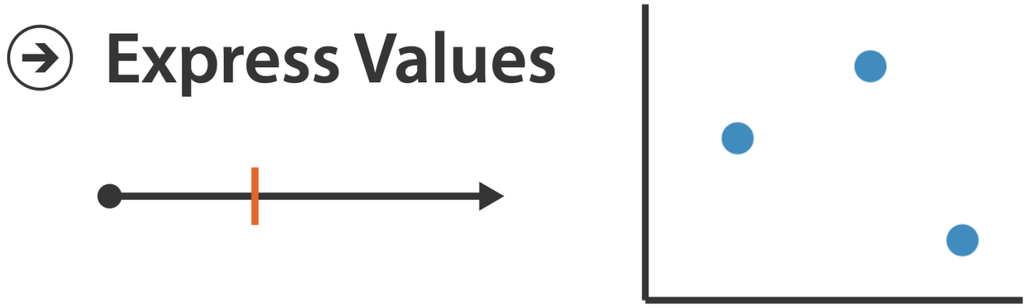
List

➔ *2 Keys*

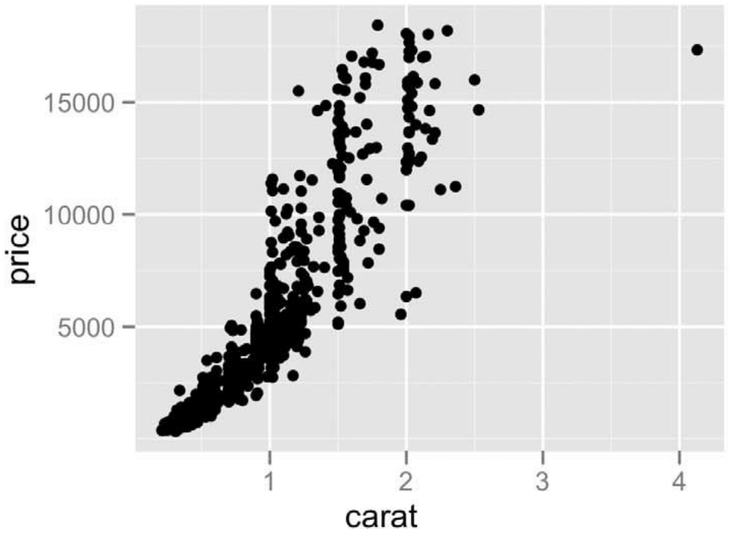
Matrix

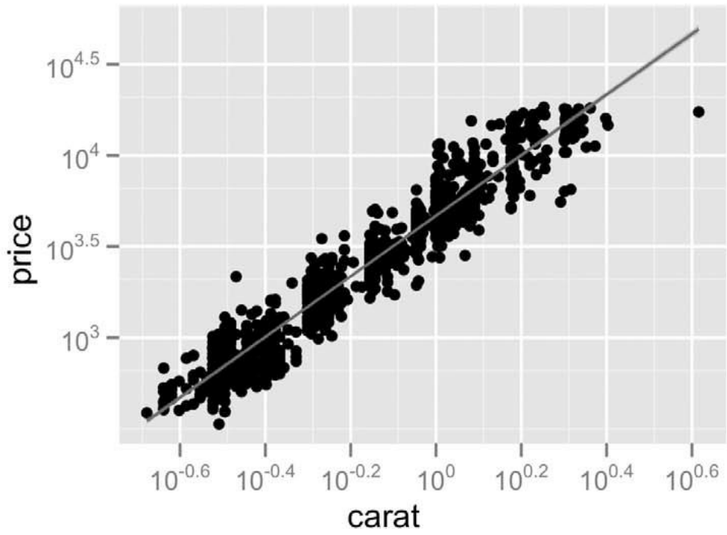


Idiom: **scatterplot**



- **express** values (magnitudes)
  - quantitative attributes
- no keys, only values
  - data
    - 2 quant attribs
  - mark: points
  - channels
    - horiz + vert position
  - tasks
    - find trends, outliers, distribution, correlation, clusters
  - scalability
    - hundreds of items

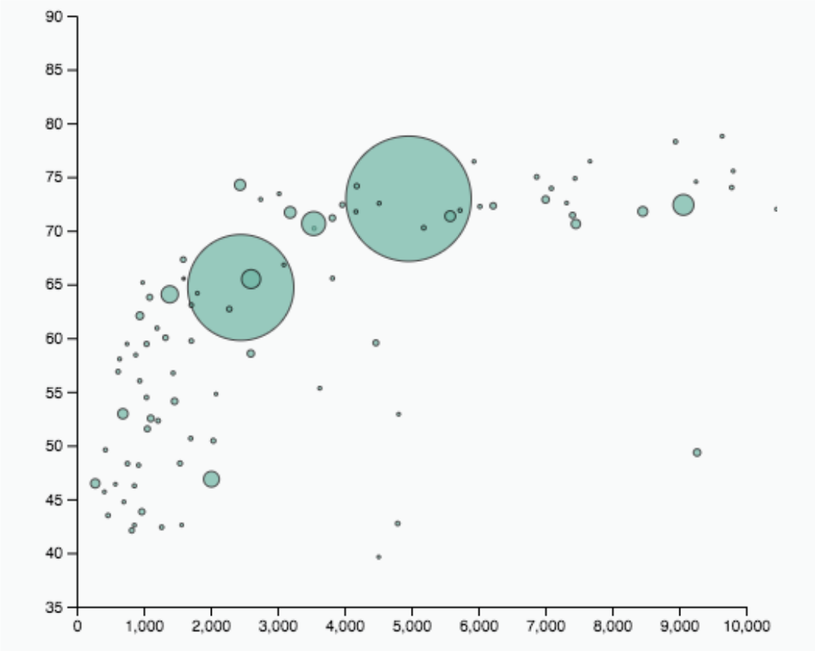


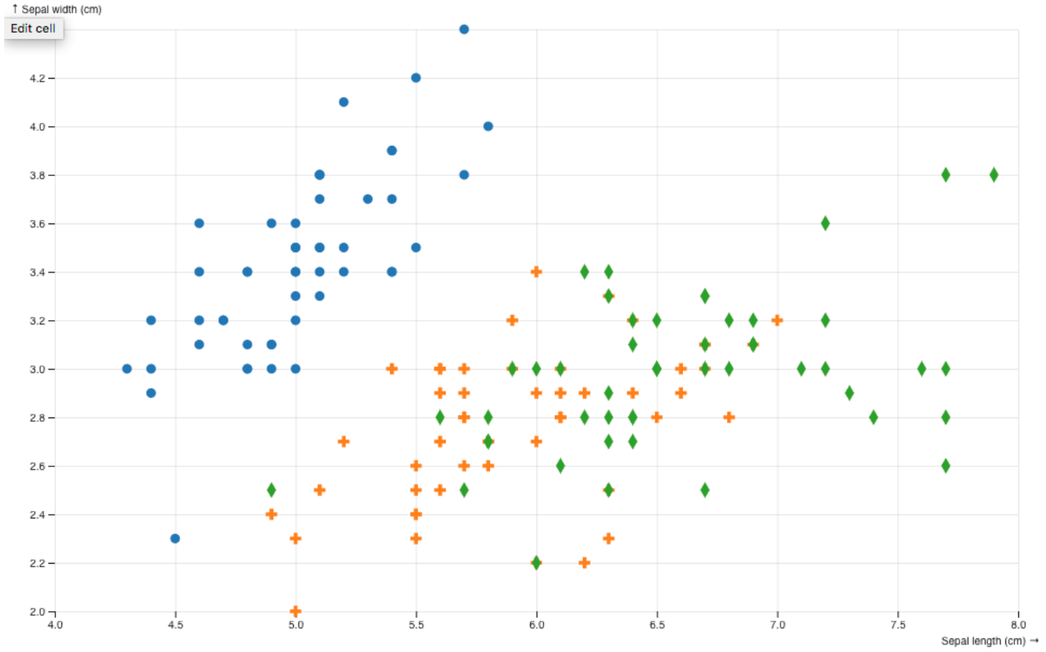


[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Scatterplots: Encoding more channels

- additional channels viable since using point marks
  - color
  - size (1 quant attribute, used to control 2D area)
    - note radius would mislead, take square root since area grows quadratically
  - shape



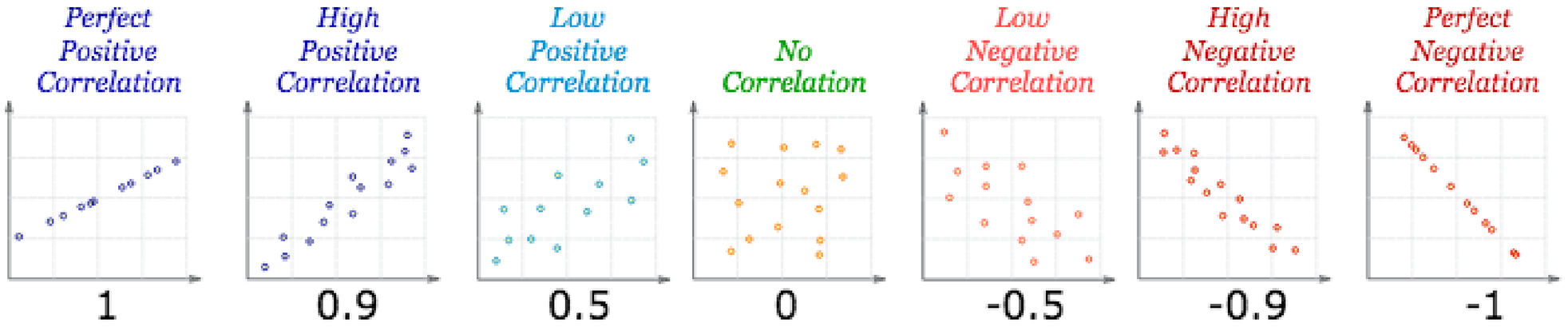


https://www.d3-graph-gallery.com/graph/bubble_basic.html

https://observablehq.com/@d3/scatterplot-with-shapes

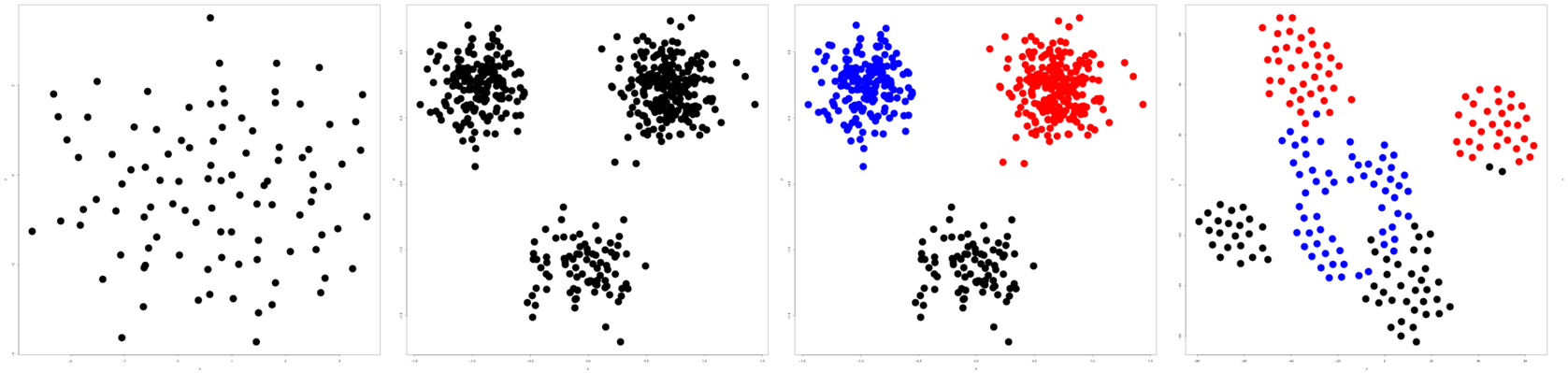
Scatterplot tasks

- correlation



https://www.mathsisfun.com/data/scatter-xy-plots.html

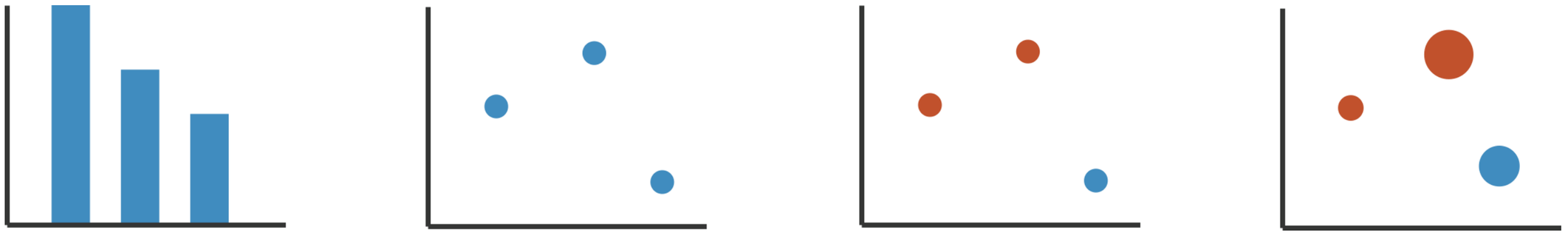
- clusters/groups, and clusters vs classes



https://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/

Visual encoding Idioms

- how to systematically analyze idiom structure?



- marks & channels
  - marks: represent items or links
  - channels: change appearance of marks based on attributes

Marks for items

- basic geometric elements

Points

0D

Lines

1D

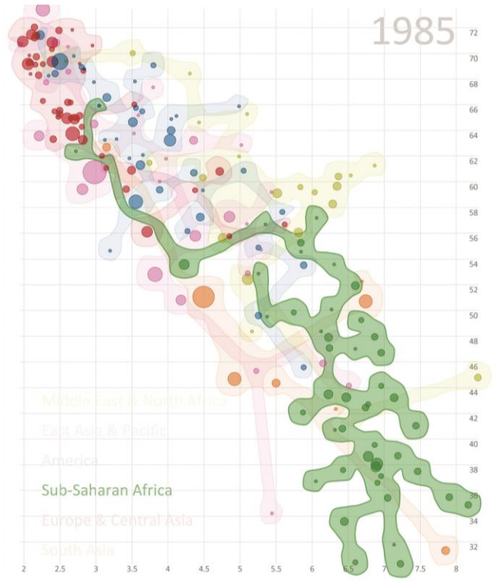
Interlocking Areas

2D

- 3D mark: volume, rarely used

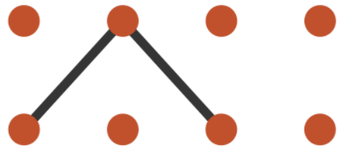
Marks for links

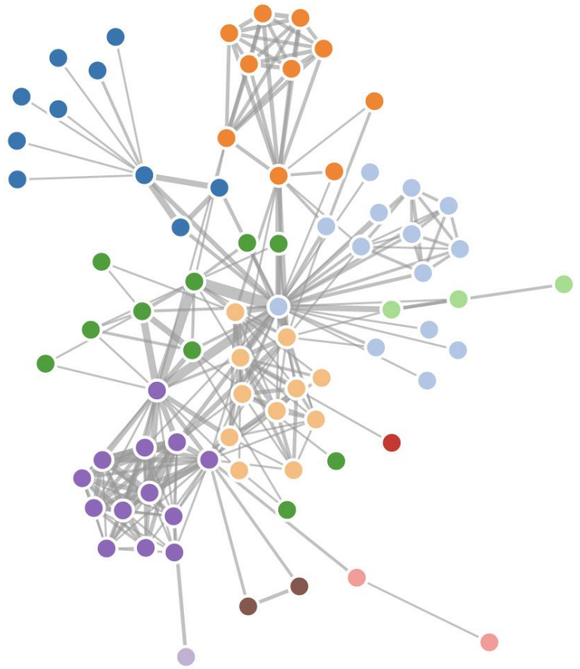
Containment



vialab.science.uoit.ca/portfolio/bubblesets

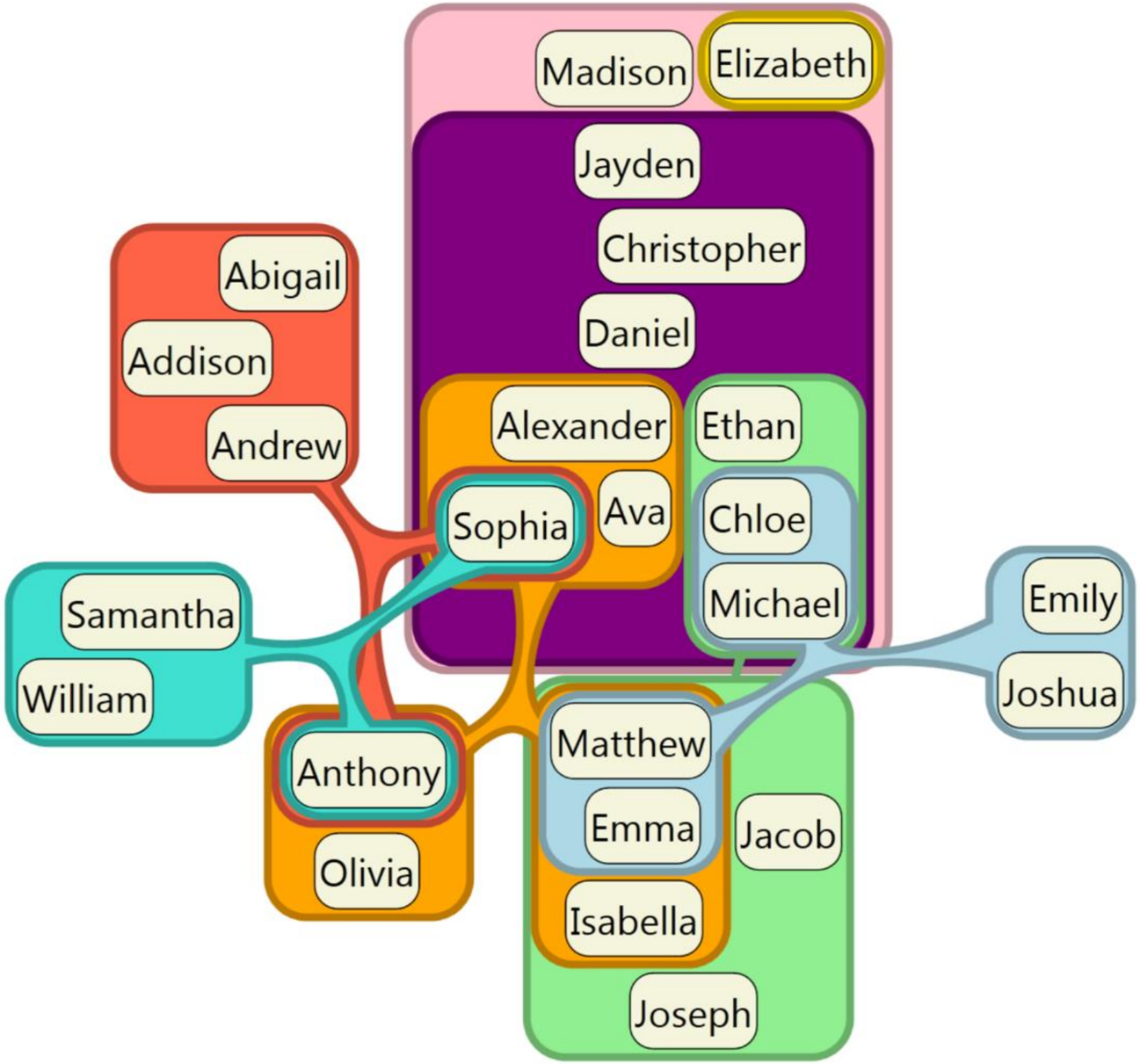
Connection





https://observablehq.com/@d3/force-directed-graph

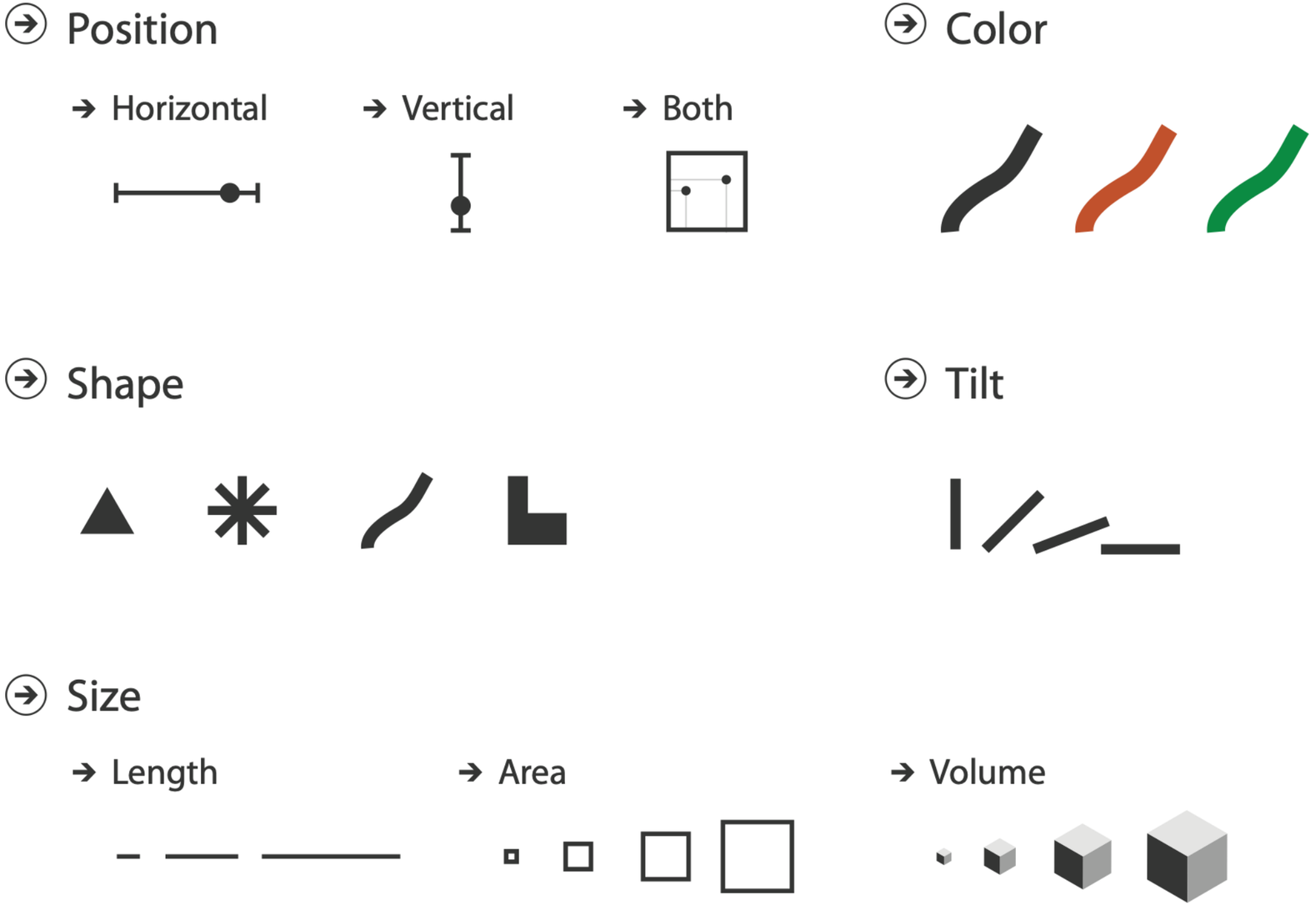
Containment can be nested



[Untangling Euler Diagrams, Riche and Dwyer, 2010]

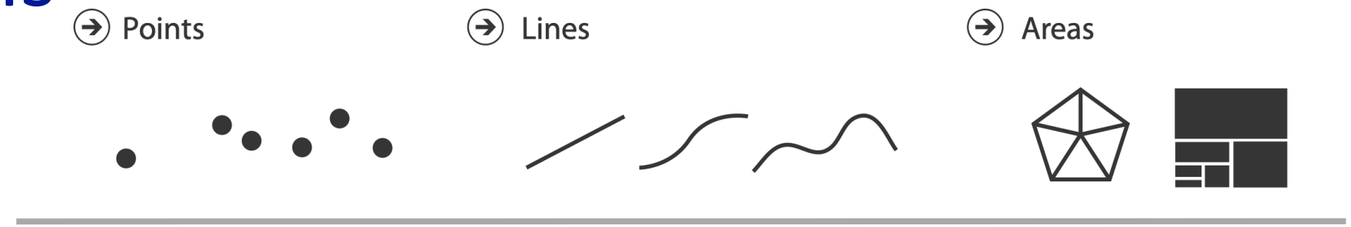
Channels

- control appearance of marks
  - proportional to or based on attributes
- many names
  - **visual channels**
  - visual variables
  - retinal channels
  - visual dimensions
  - …



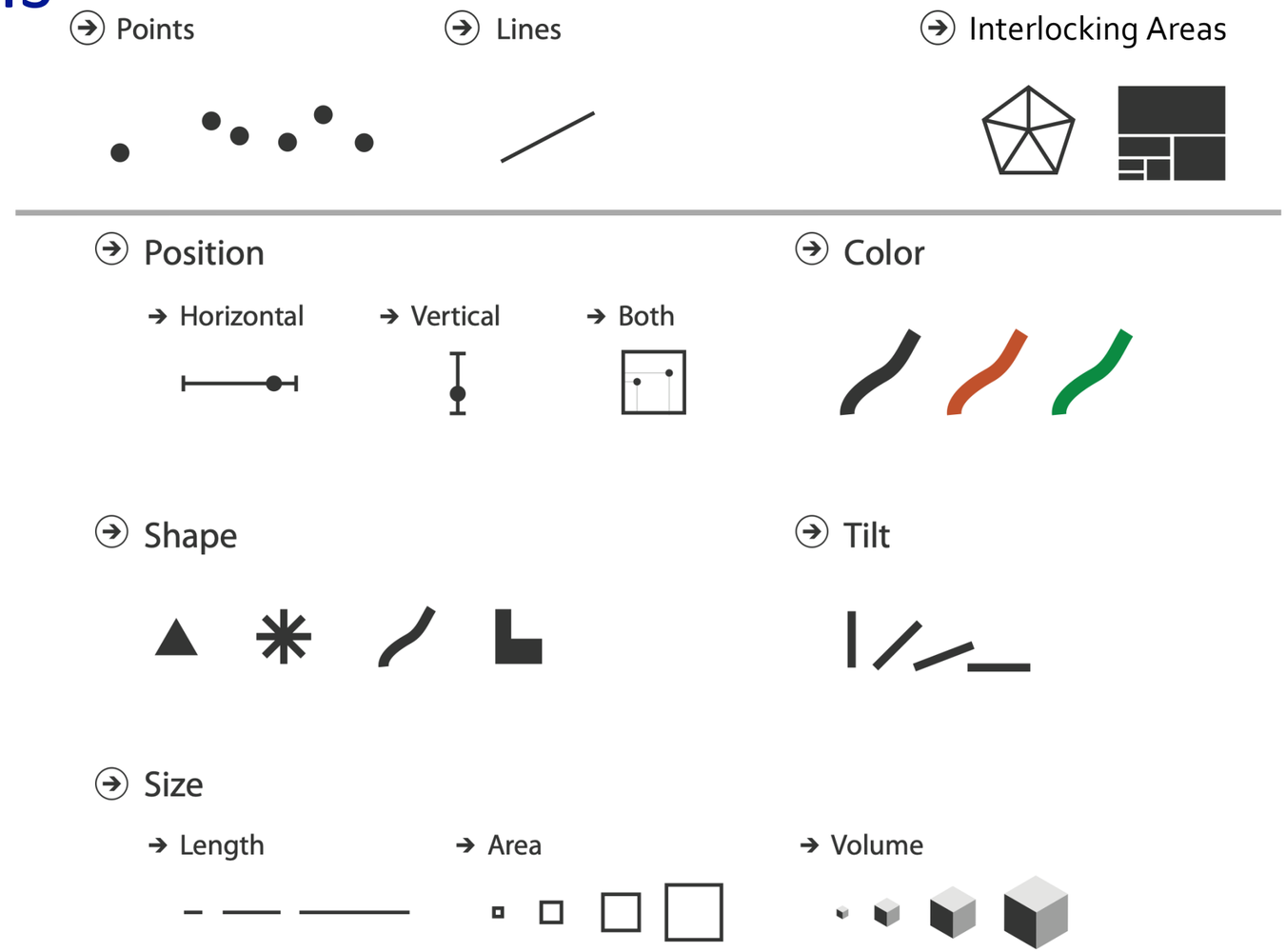
Definitions: Marks and channels

- marks
  - geometric primitives



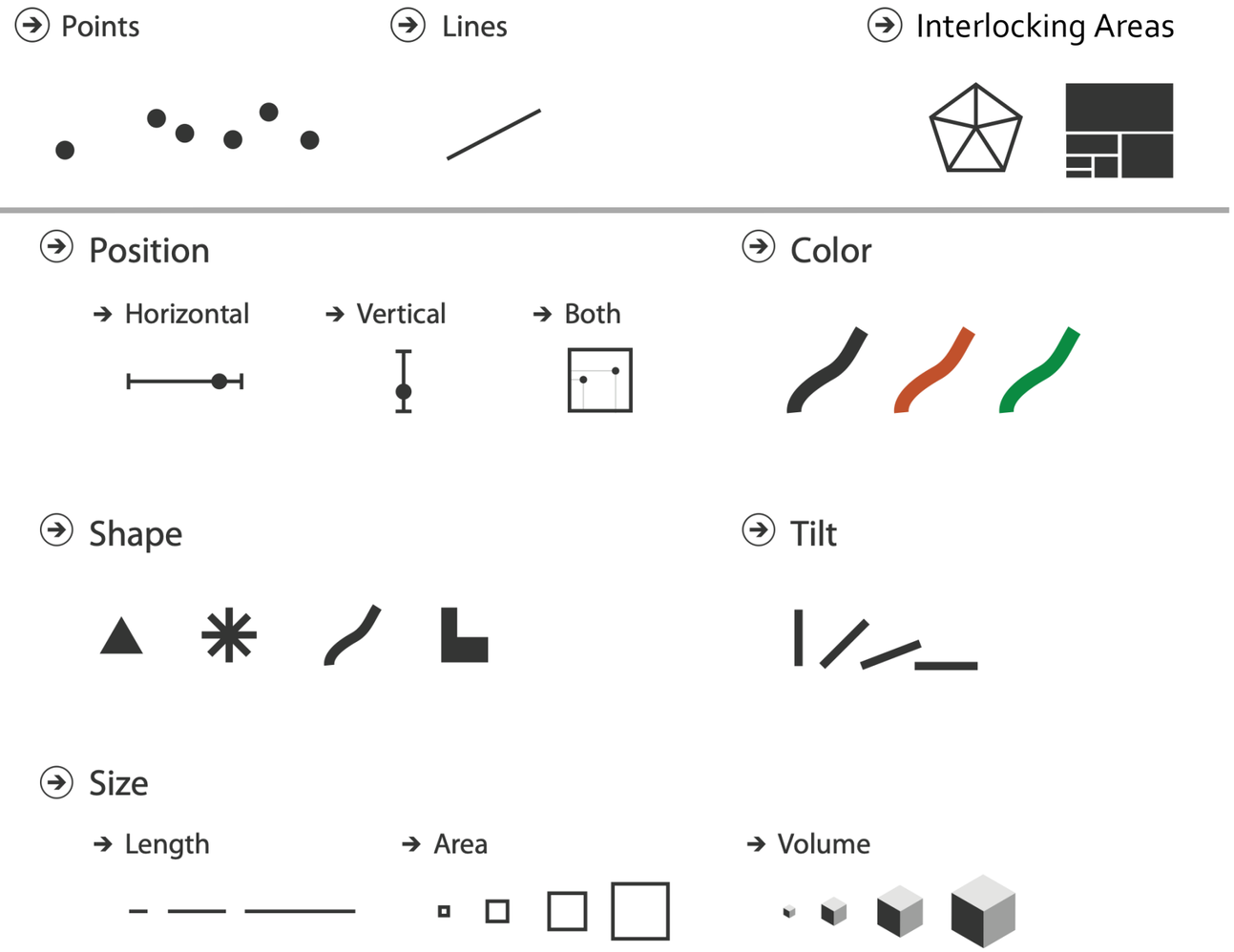
Definitions: Marks and channels

- marks
  - geometric primitives
- channels
  - control appearance of marks



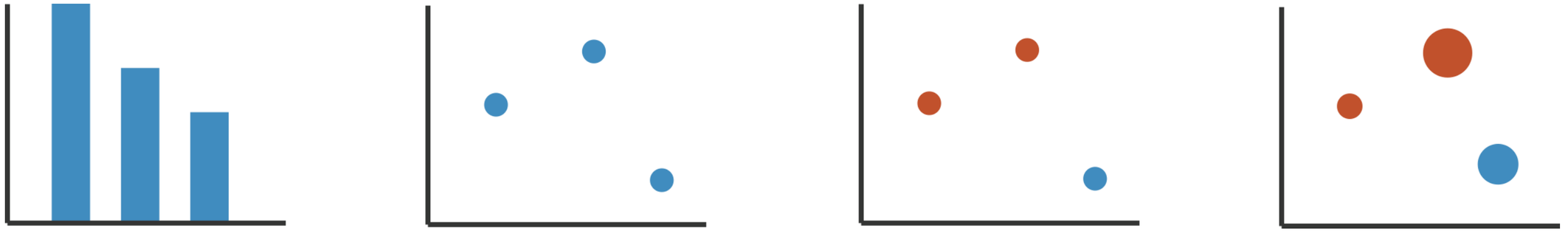
Definitions: Marks and channels

- marks
  - geometric primitives
- channels
  - control appearance of marks
- channel properties differ
    - type & amount of information that can be conveyed to human perceptual system



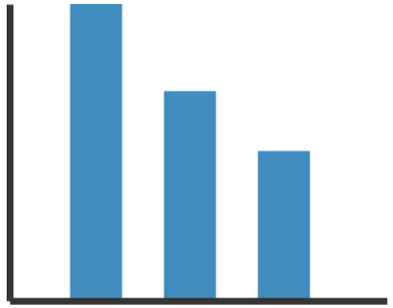
Visual encoding

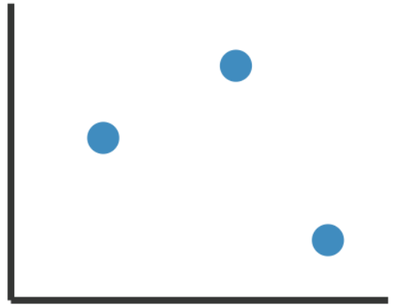
- analyze idiom structure as combination of marks and channels

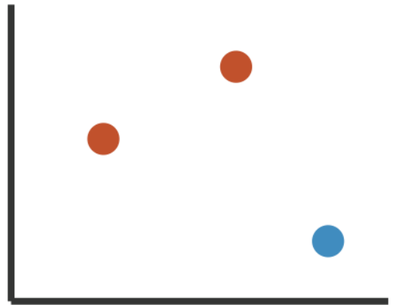


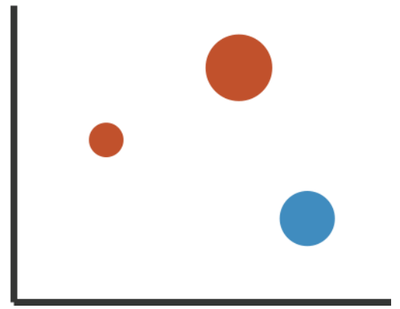
Visual encoding

- analyze idiom structure as combination of marks and channels







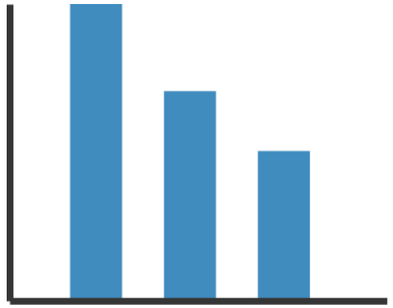


1:
vertical position

mark: line

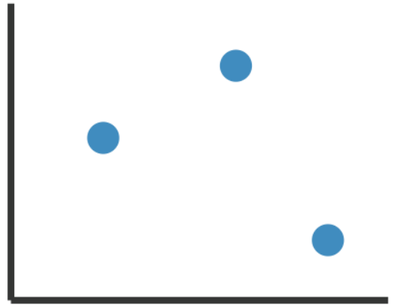
Visual encoding

- analyze idiom structure as combination of marks and channels



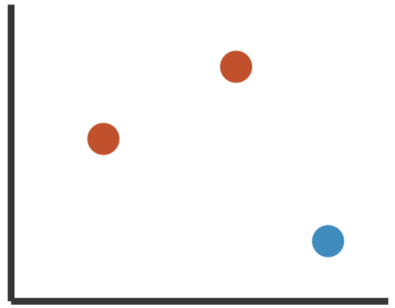
1:
vertical position

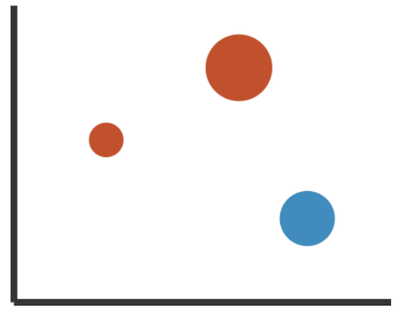
mark: line



2:
vertical position
horizontal position

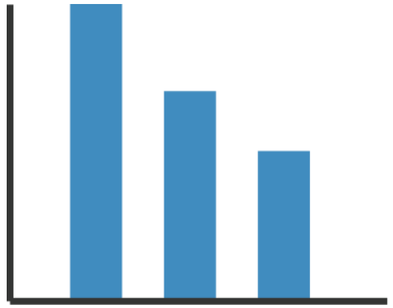
mark: point





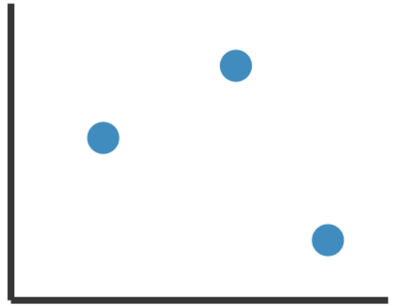
Visual encoding

- analyze idiom structure as combination of marks and channels



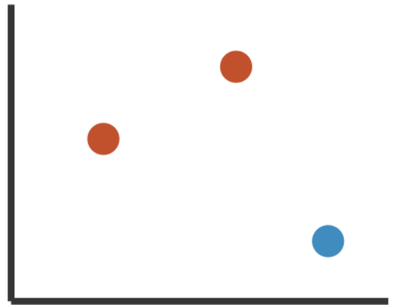
1:
vertical position

mark: line



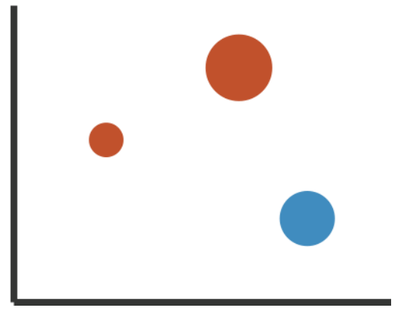
2:
vertical position
horizontal position

mark: point



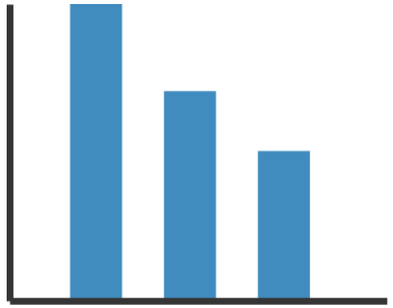
3:
vertical position
horizontal position
color hue

mark: point



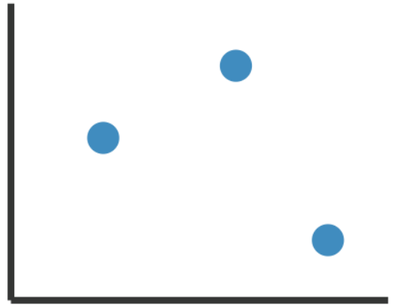
Visual encoding

- analyze idiom structure as combination of marks and channels



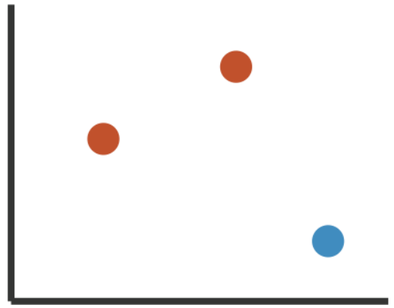
1:
vertical position

mark: line



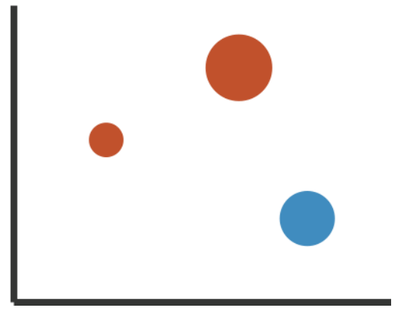
2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

mark: point

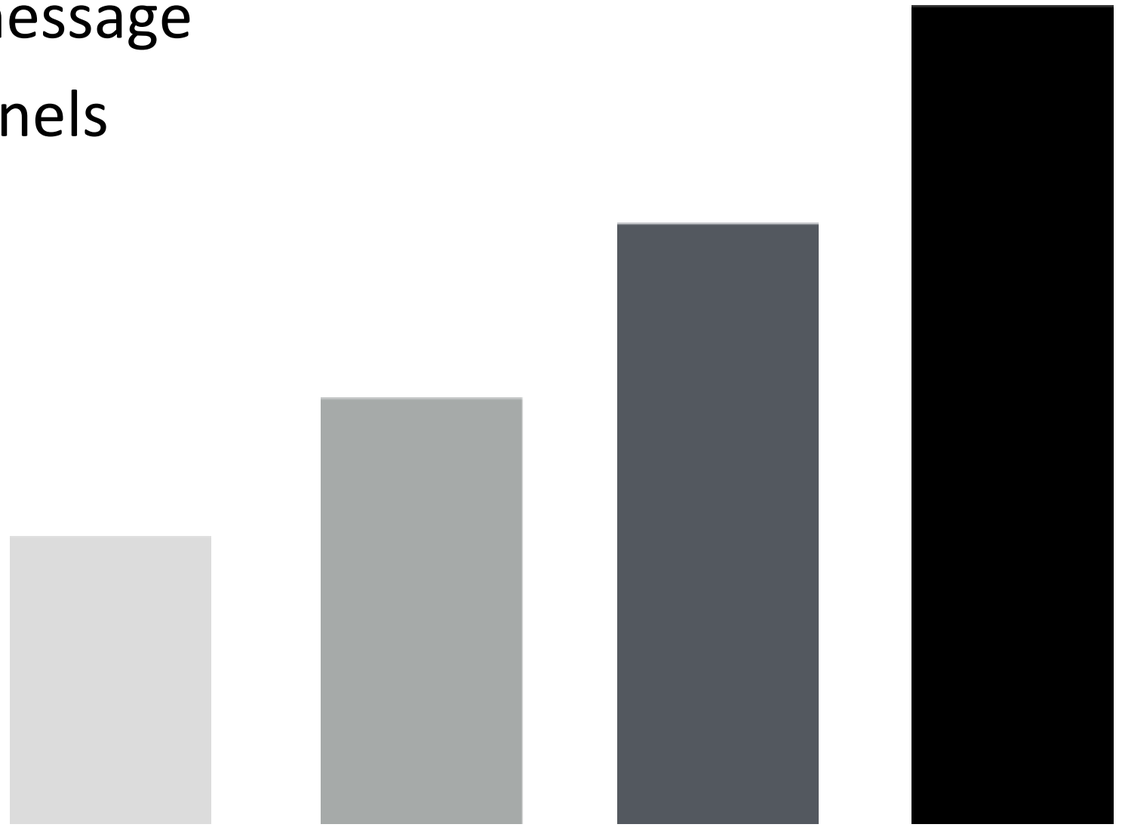


4:
vertical position
horizontal position
color hue
size (area)

mark: point

Redundant encoding

- multiple channels
  - sends stronger message
  - but uses up channels

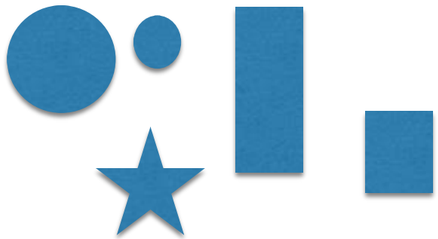


Length and Luminance

Marks as constraints

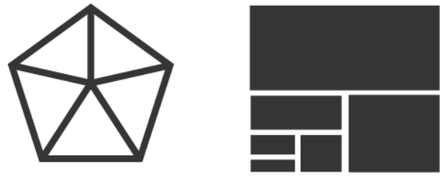
- math view: geometric primitives have dimensions

Points 0D



Lines 1D

Interlocking Areas 2D



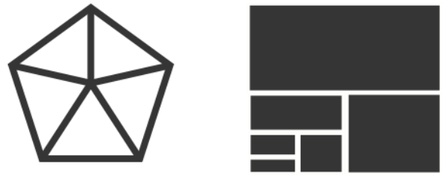
Marks as constraints

- math view: geometric primitives have dimensions

Points 0D

Lines 1D

Interlocking Areas 2D



- constraint view: mark type constrains what else can be encoded
  - points: 0 constraints on size, can encode more attributes w/ size & shape
  - lines: 1 constraint on size (length), can still size code other way (width)
  - interlocking areas: 2 constraints on size (length/width), cannot size or shape code
    - interlocking: size, shape, position

Marks as constraints

- math view: geometric primitives have dimensions

Points 0D Lines 1D Interlocking Areas 2D

- constraint view: mark type constrains what else can be encoded
  - points: 0 constraints on size, can encode more attributes w/ size & shape
  - lines: 1 constraint on size (length), can still size code other way (width)
  - interlocking areas: 2 constraints on size (length/width), cannot size or shape code
    - interlocking: size, shape, position
- quick check: can you size-code another attribute
  - or is size/shape in use?

When to use which channel?

expressiveness

match channel type to data type

effectiveness

some channels are better than others

Channels: Rankings

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Spatial region

Color hue

Motion

Shape

Channels: Rankings

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

- expressiveness
  - match channel and data characteristics

Channels: Rankings

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

Attribute Types

→ Categorical

→ Ordered

→ *Ordinal*

→ *Quantitative*

- expressiveness
  - match channel and data characteristics
    - magnitude for ordered
      - how much? which rank?
    - identity for categorical
      - what?

Channels: Rankings

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Same

Same

Best

Effectiveness

Least

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

- expressiveness
  - match channel and data characteristics
- effectiveness
  - channels differ in accuracy of perception

Channels: Rankings

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Same

Same

Effectiveness

Best

Least

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

- expressiveness
  - match channel and data characteristics
- effectiveness
  - channels differ in accuracy of perception
  - spatial position ranks high for both

Grouping

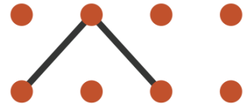
- containment
- connection

- proximity
  - same spatial region
- similarity
  - same values as other categorical channels

Marks as Links

Containment

Connection



Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

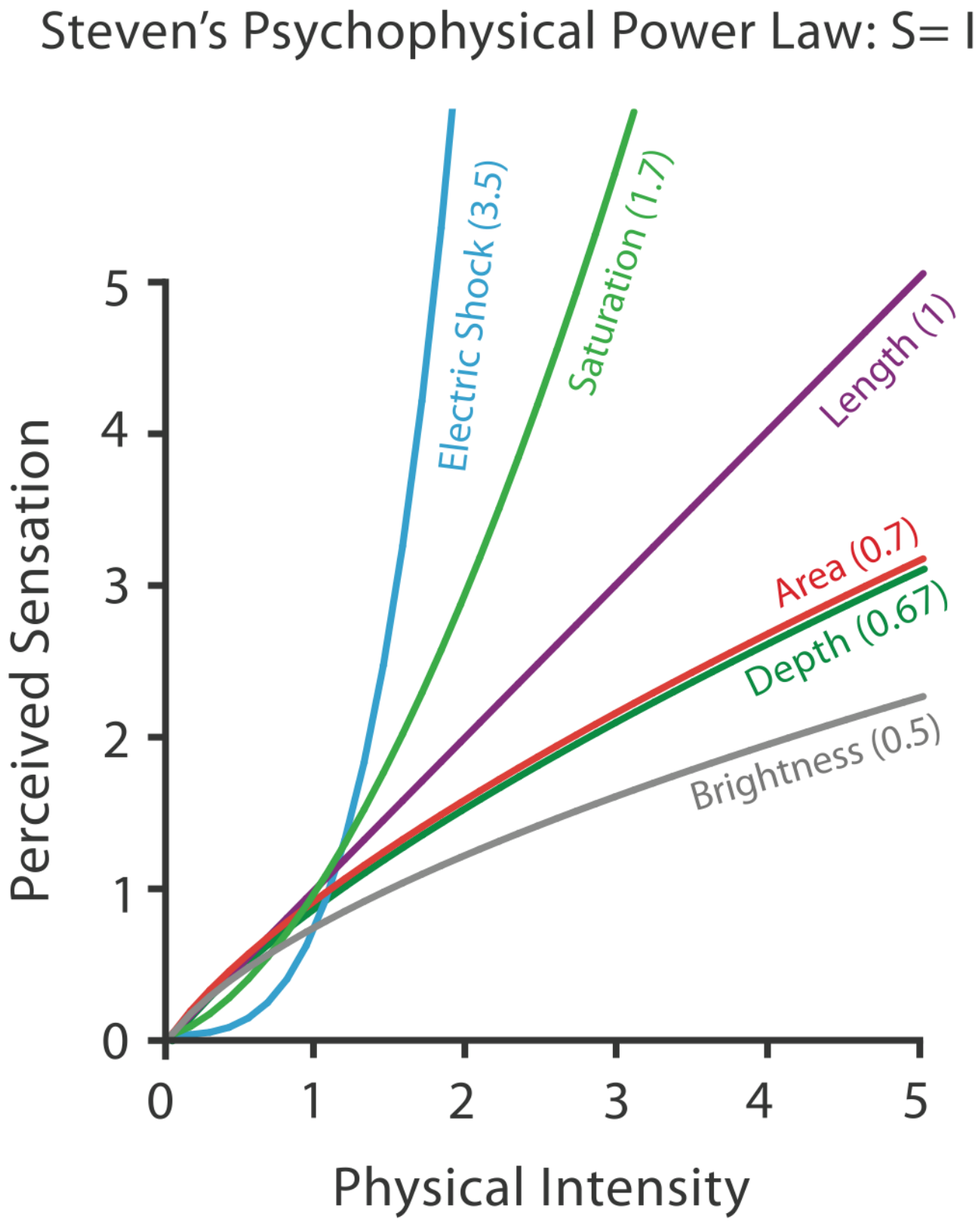
Channel effectiveness

Psychophysics = the subfield of psychology devoted to the systematic measurement of general human perception.

- accuracy: how precisely can we tell the difference between encoded items?
- discriminability: how many unique steps can we perceive?
- separability: is our ability to use this channel affected by another one?
- popout: can things jump out using this channel?

Accuracy: Fundamental theory

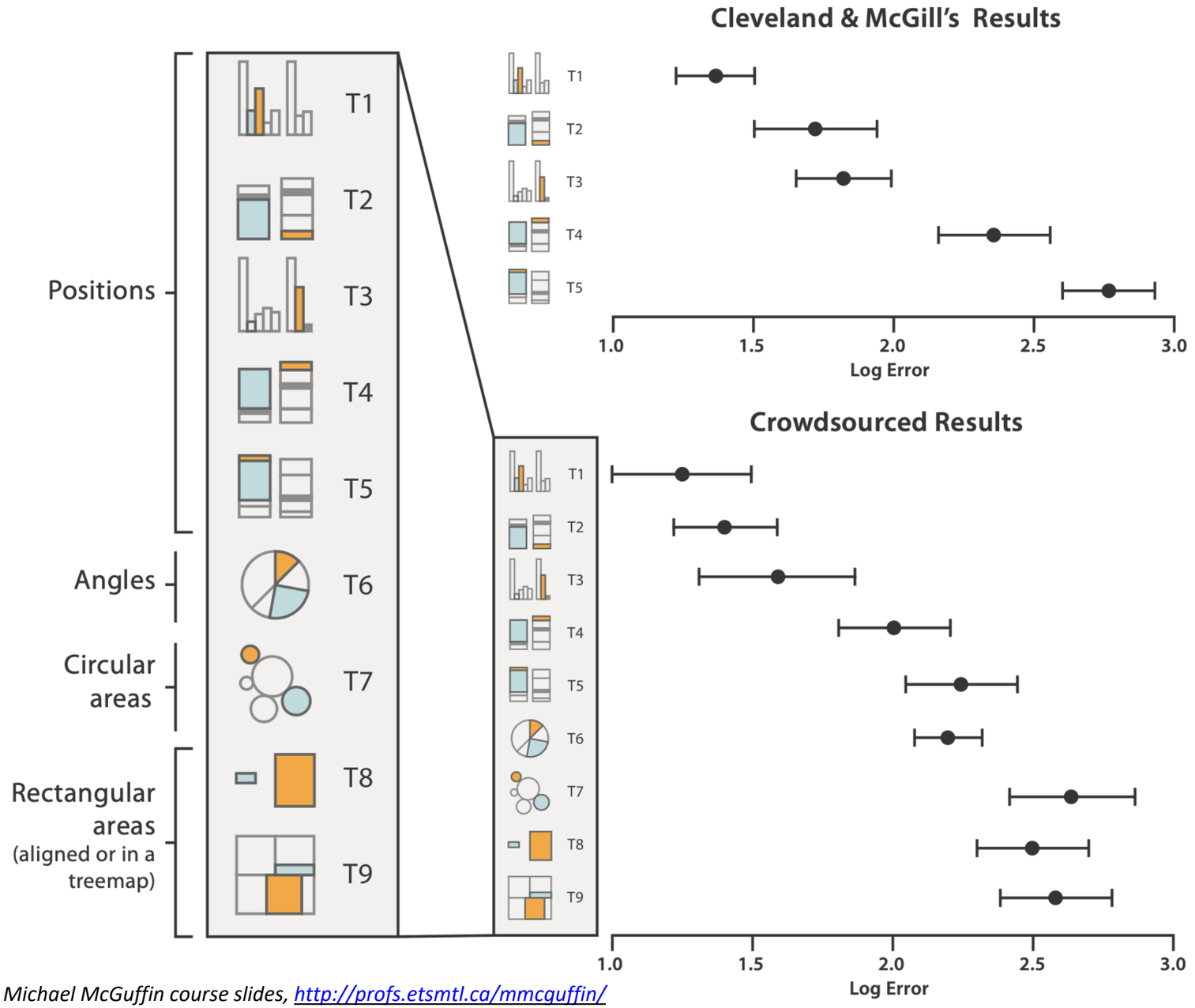
- length is accurate: linear
- others magnified or compressed
  - –exponent characterizes



S = sensation

I = intensity

Accuracy: Vis experiments

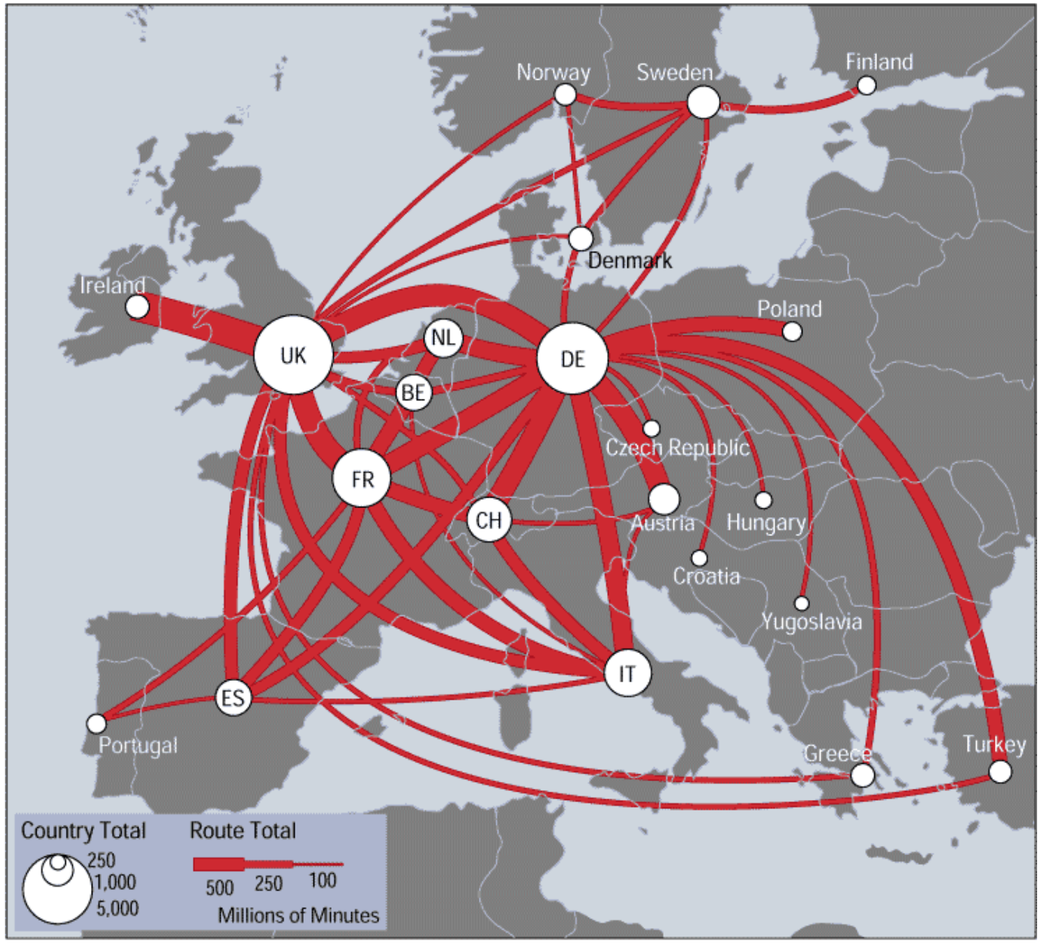


[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

after Michael McGuffin course slides, http://profs.etsmtl.ca/mmcguffin/

Discriminability: How many usable steps?

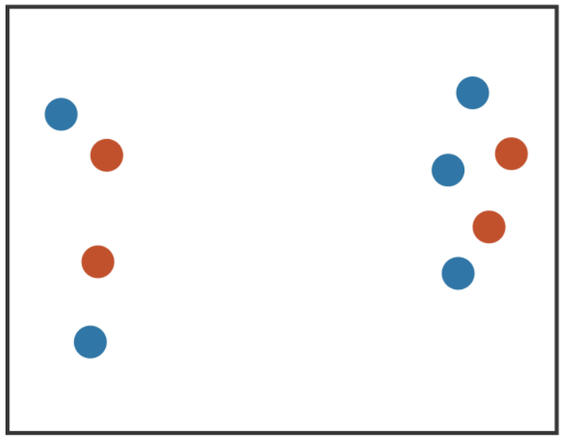
- must be sufficient for number of attribute levels to show
  - linewidth: few bins



[mappa.mundi.net/maps/maps 014/telegeography.html]

Separability vs. Integrality

Position
+ Hue (Color)

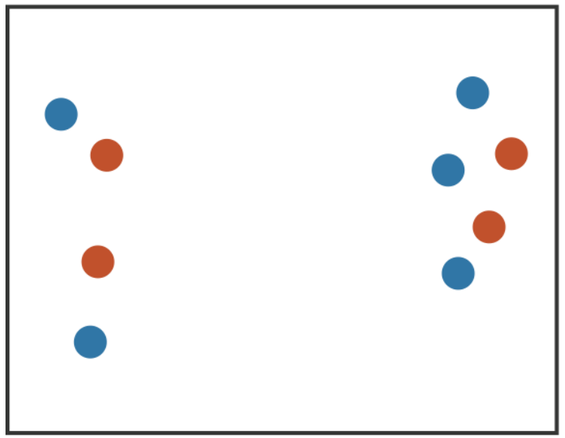


Fully separable

2 groups each

Separability vs. Integrality

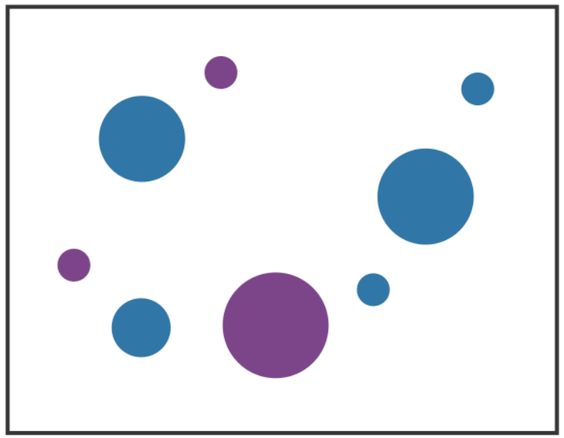
Position
+ Hue (Color)



Fully separable

2 groups each

Size
+ Hue (Color)

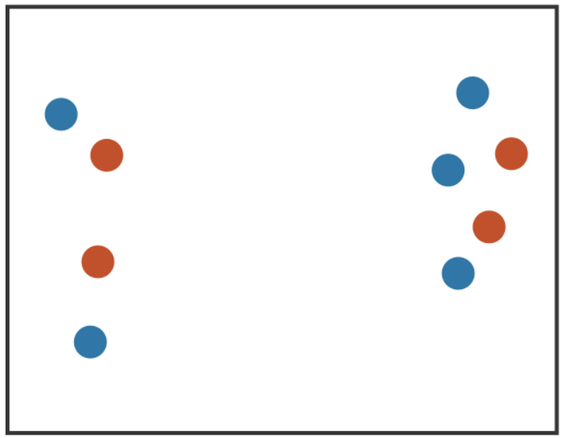


Some interference

2 groups each

Separability vs. Integrality

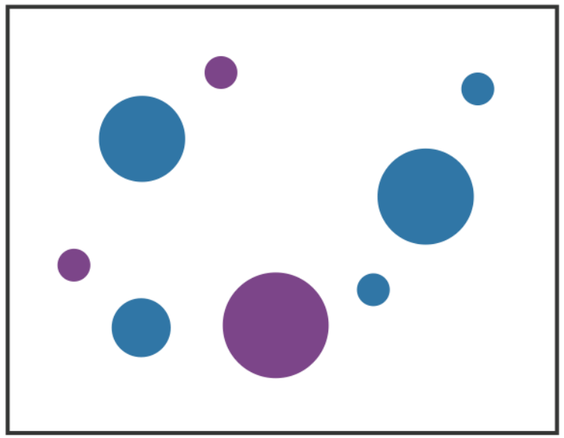
Position
+ Hue (Color)



Fully separable

2 groups each

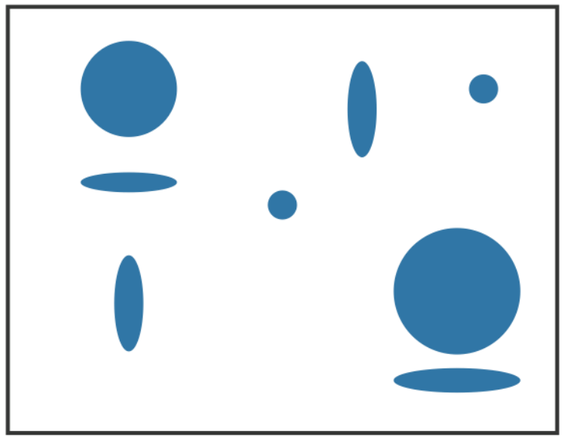
Size
+ Hue (Color)



Some interference

2 groups each

Width
+ Height

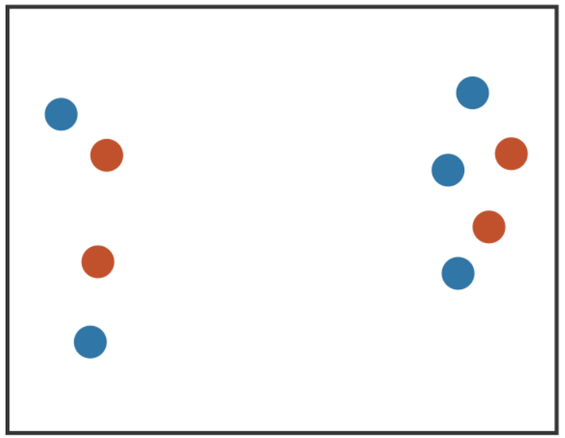


Some/significant interference

3 groups total: integral area

Separability vs. Integrality

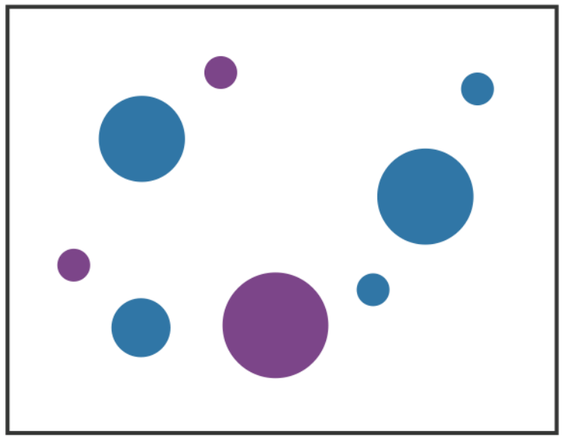
Position
+ Hue (Color)



Fully separable

2 groups each

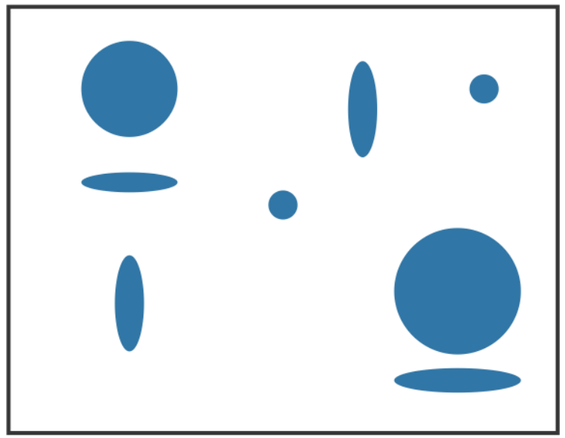
Size
+ Hue (Color)



Some interference

2 groups each

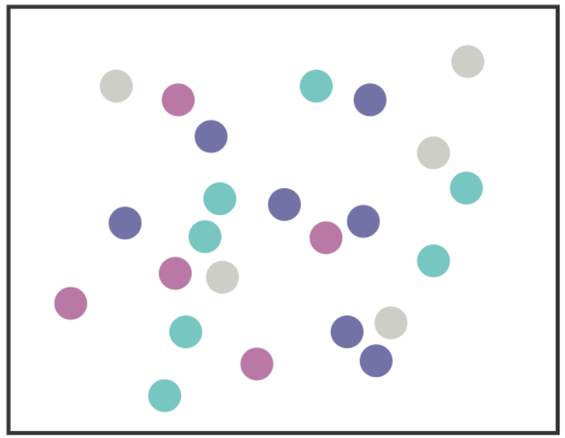
Width
+ Height



Some/significant interference

3 groups total: integral area

Red
+ Green

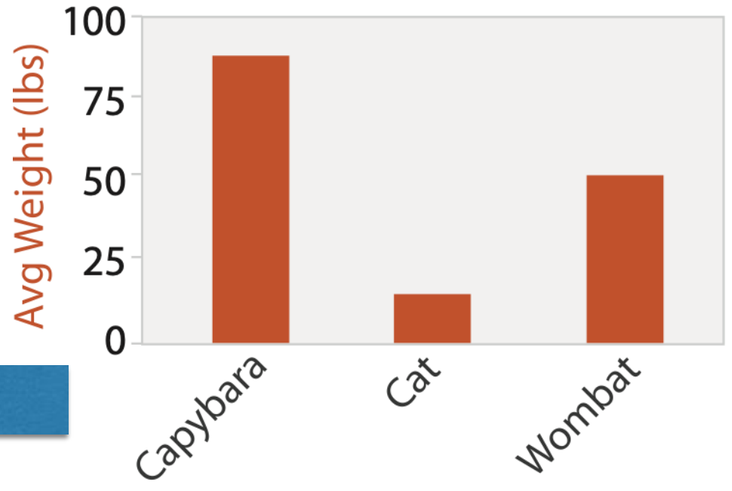


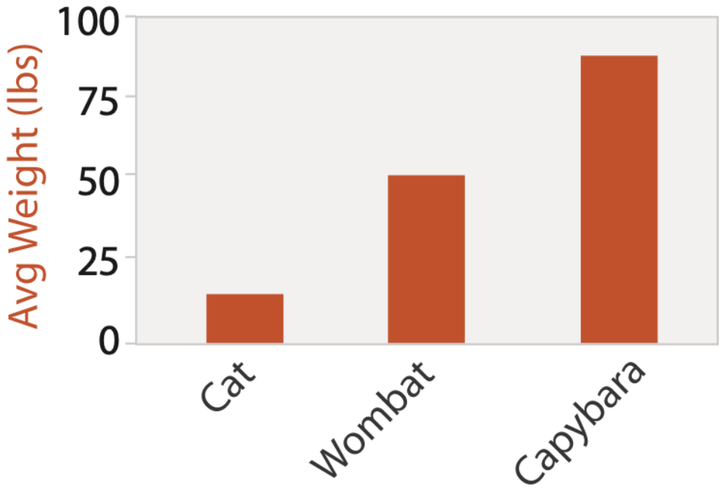
Major interference

4 groups total: integral hue

Idiom: **bar chart**

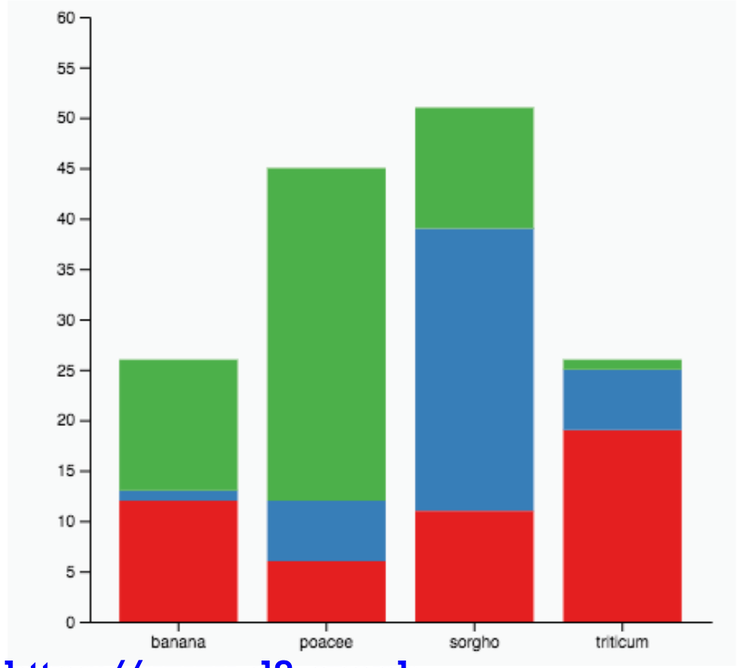
- one key, one value
  - data
  - mark:
  - channels
  - task
    -
  - scalability





Idiom: **stacked bar chart**

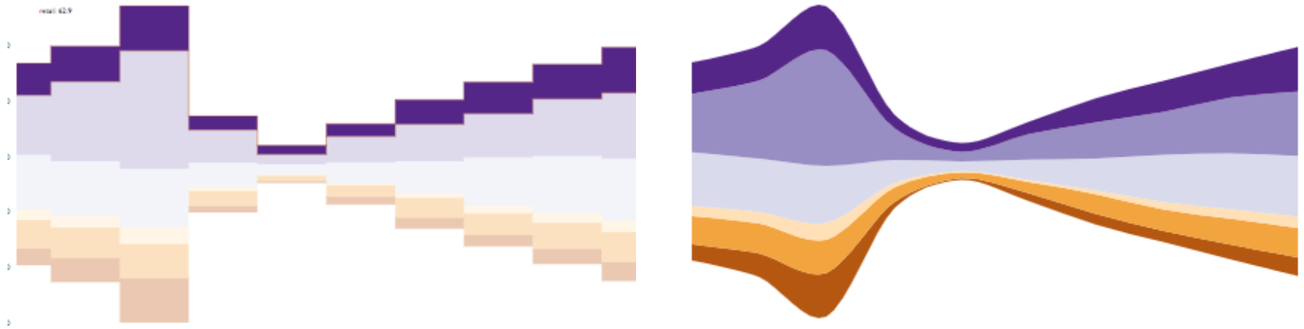
- one more key
  - data
    - 2 categ attrib, 1 quant attrib
  - mark: vertical stack of line marks
    - **glyph**: composite object, internal structure from multiple marks
  - channels
    - length and color hue
    - spatial regions: one per glyph
      - aligned: full glyph, lowest bar component
      - unaligned: other bar components
  - task
    - part-to-whole relationship
  - scalability: asymmetric
    - for *stacked* key attrib, 10-12 levels [segments]
    - for *main* key attrib, dozens to hundreds of levels [bars]



https://www.d3-graph-gallery.com/graph/barplot_stacked_basicW.html

Idiom: **streamgraph**

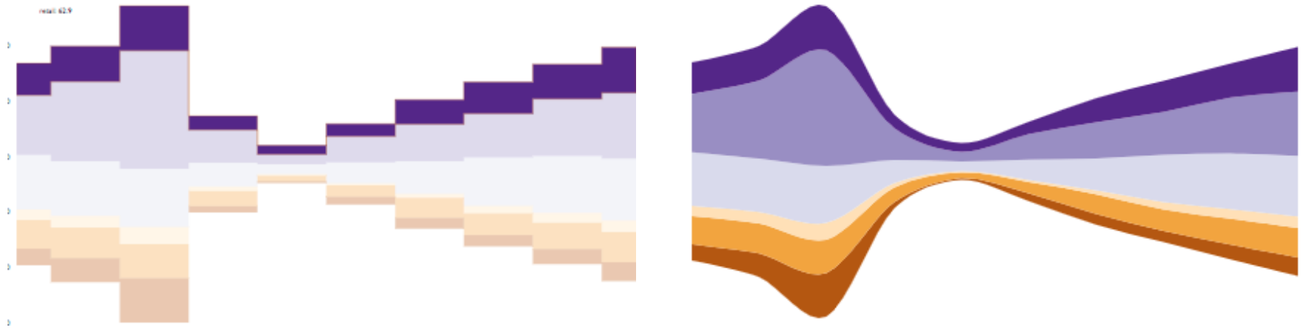
- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 1 ordered key attrib (time)
    - 1 quant value attrib (counts)
  - derived data
    - geometry: layers, where height encodes counts
    - 1 quant attrib (layer ordering)



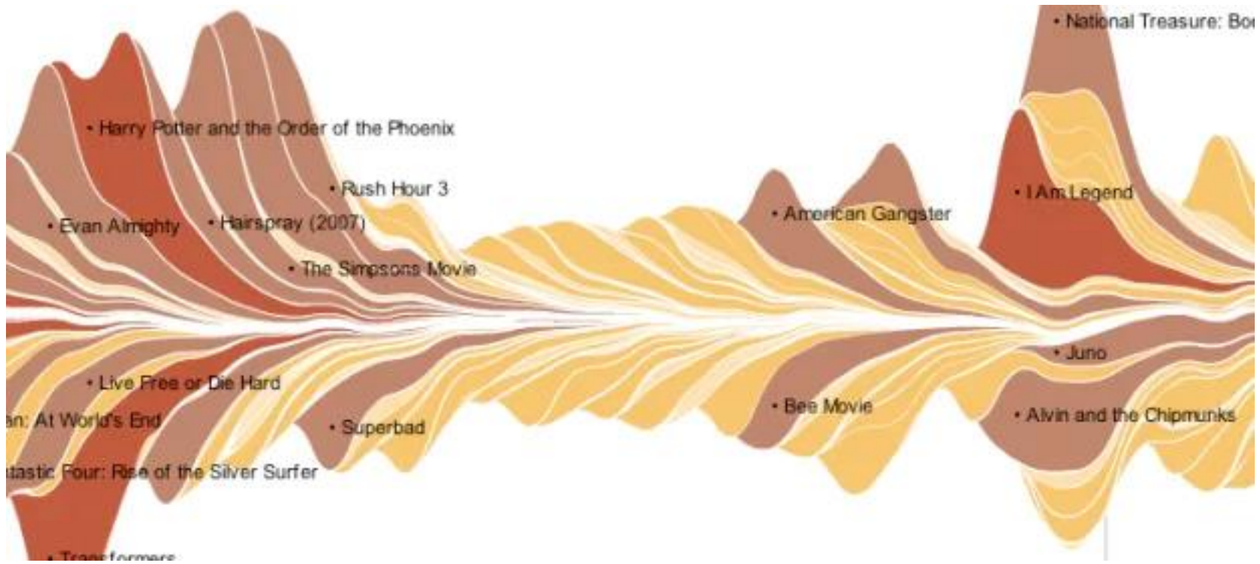
[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

Idiom: **streamgraph**

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 1 ordered key attrib (time)
    - 1 quant value attrib (counts)
  - derived data
    - geometry: layers, where height encodes counts
    - 1 quant attrib (layer ordering)
  - scalability
    - hundreds of time keys
    - dozens to hundreds of movies keys
      - more than stacked bars: most layers don't extend across whole chart



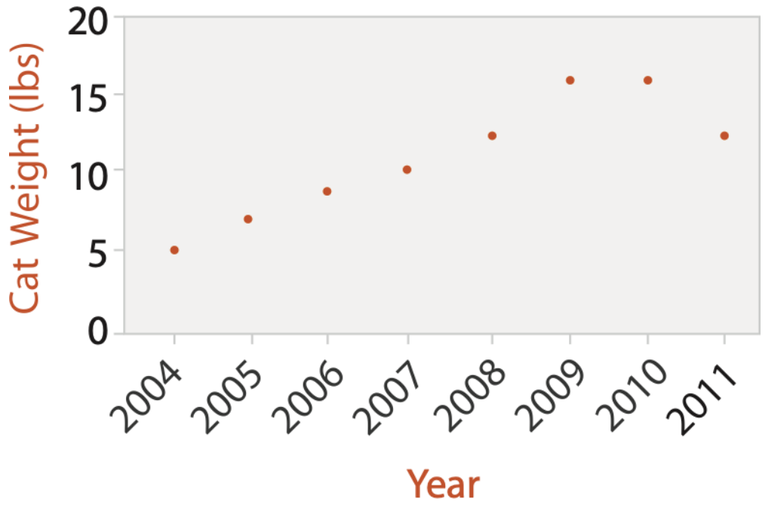
[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

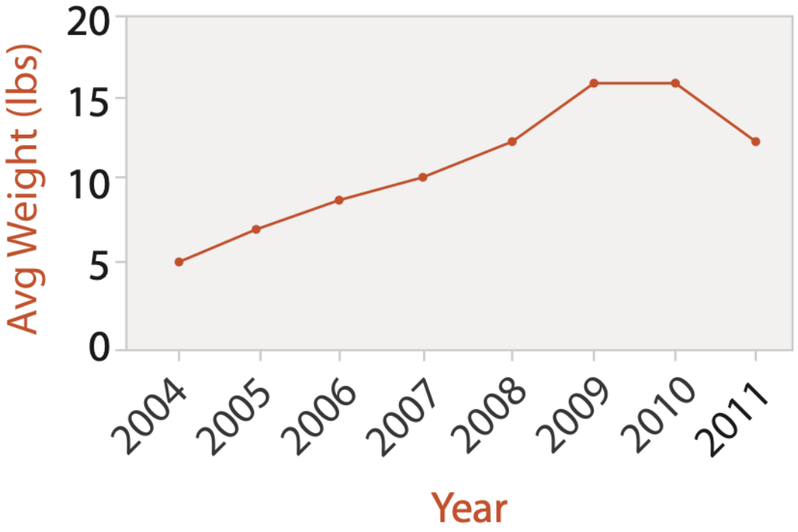


https://flowingdata.com/2008/02/25/ebb-and-flow-of-box-office-receipts-over-past-20-years/

Idiom: **dot / line chart**

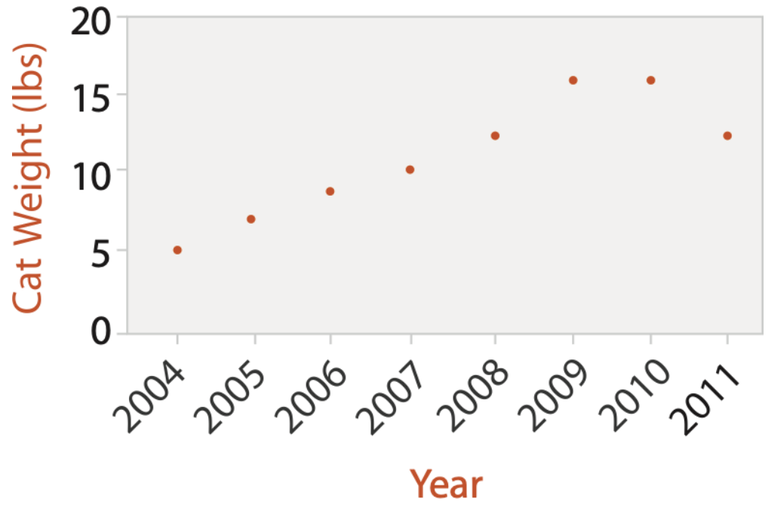
- one key, one value
  - data
    - 2 quant attribs
  - mark: [redacted]
    AND [redacted]
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions

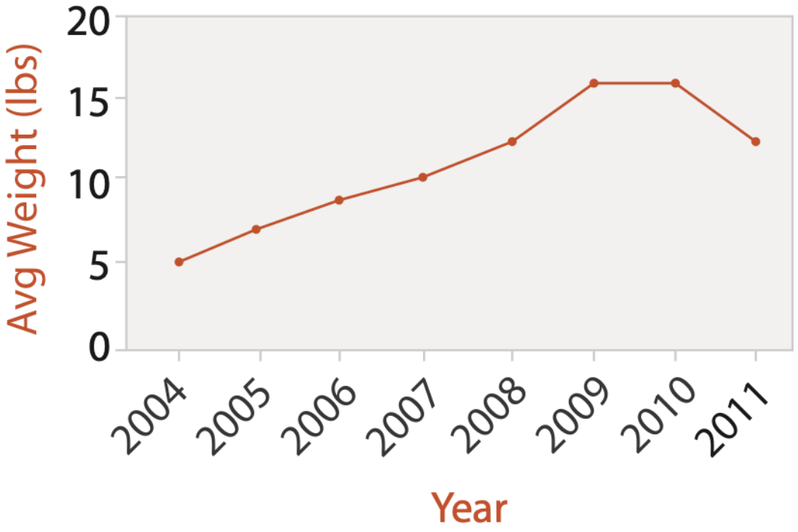




Idiom: **dot / line chart**

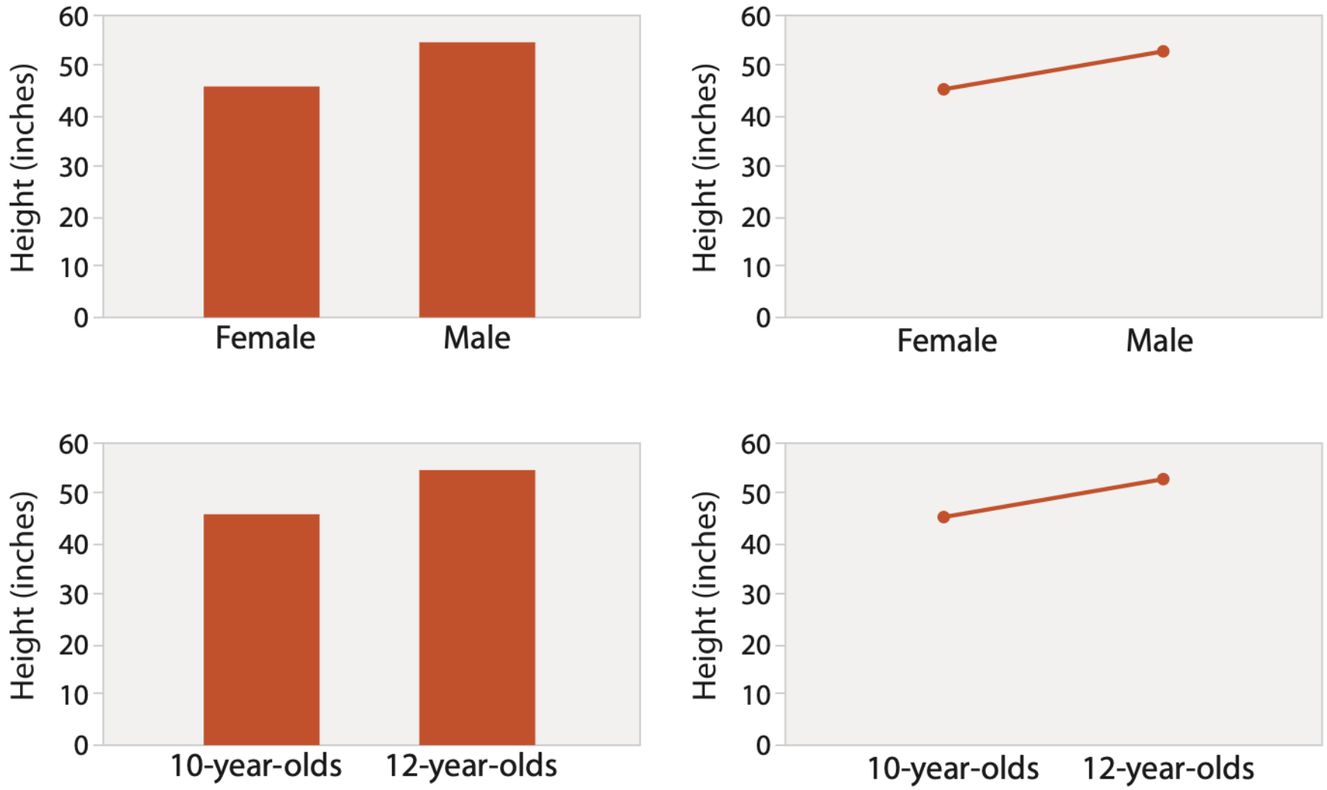
- one key, one value
  - data
    - 2 quant attribs
  - mark: points
    AND line connection marks between them
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions
  - task
    - find trend
      - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
  - scalability
    - hundreds of key levels, hundreds of value levels





Choosing bar vs line charts

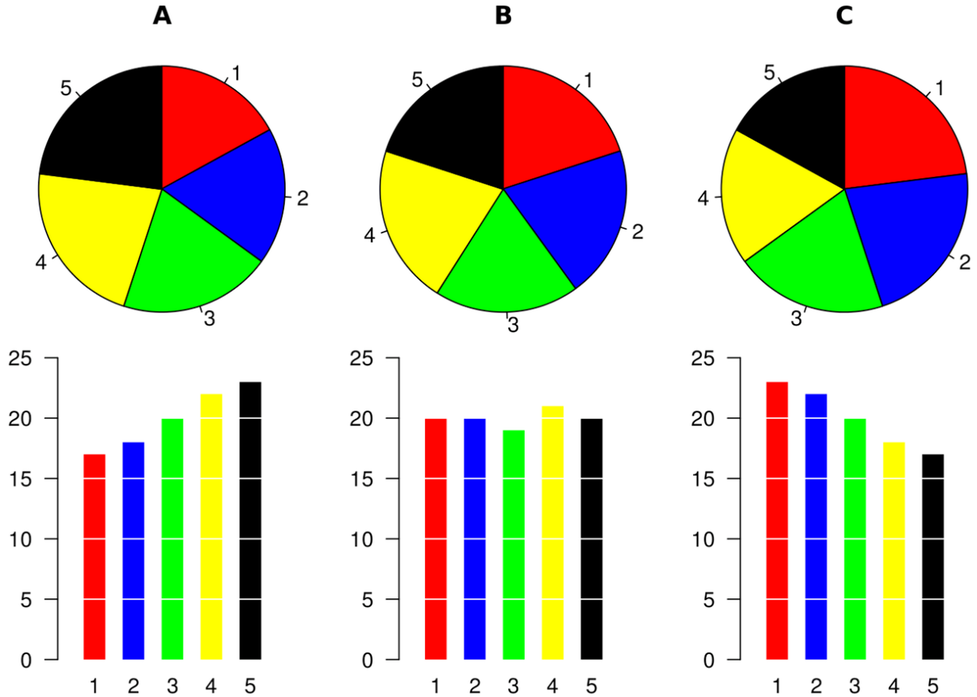
- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attribs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
      - "The more male a person is, the taller he/she is"

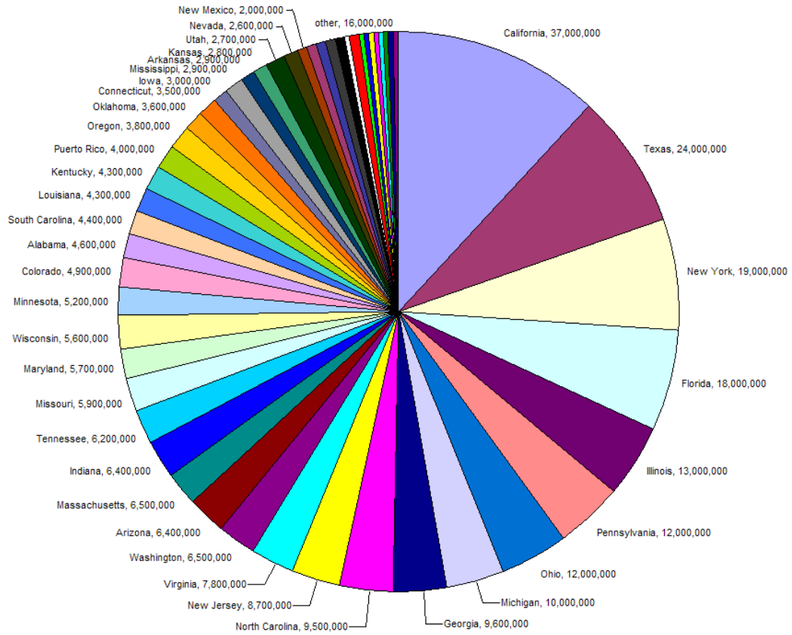


after [Bars and Lines: A Study of Graphic Communication. Zacks and Tversky. Memory and Cognition 27:6 (1999), 1073–1079.]

Pie charts: best practices

- not so bad for two (or few) levels, for part-to-whole task
- dubious for several levels if details matter
- terrible for many levels

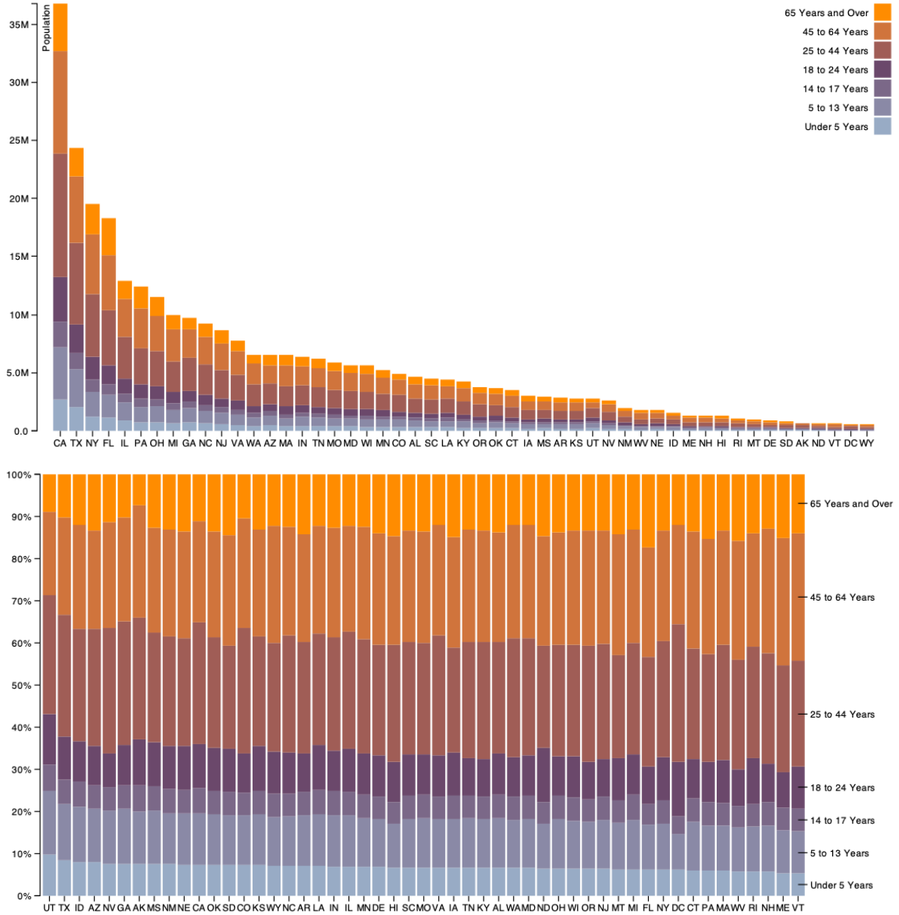


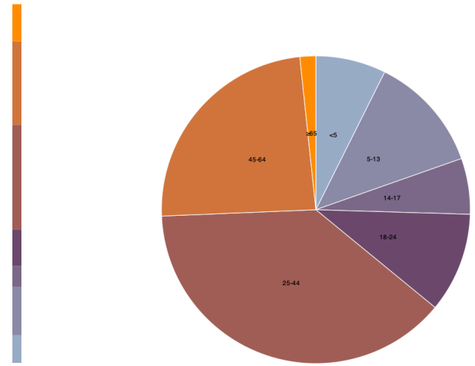


https://eagereyes.org/pie-charts

Idioms: **normalized stacked bar chart**

- task
  - part-to-whole judgements
- normalized stacked bar chart
  - stacked bar chart, normalized to full vert height
  - single stacked bar equivalent to full pie
    - high information density: requires narrow rectangle
- pie chart
  - information density: requires large circle





http://bl.ocks.org/mbostock/3886208,

http://bl.ocks.org/mbostock/3887235

2 Keys

➔ *0 Keys*

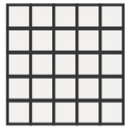
➔ **Express Values**

➔ *1 Key*

List

➔ *2 Keys*

Matrix

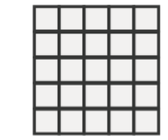


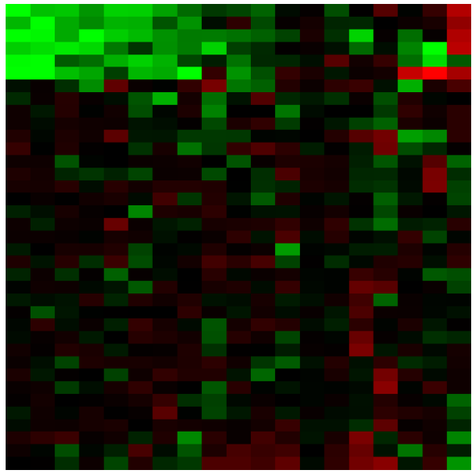
Idiom: **heatmap**

- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: area
    - separate and align in 2D matrix
      - indexed by 2 categorical attributes
  - channels
    - color by quant attrib
      - (ordered diverging colormap)
  - task
    - find clusters, outliers
  - scalability
    - 1M items, 100s of categ levels, ~10 quant attrib levels

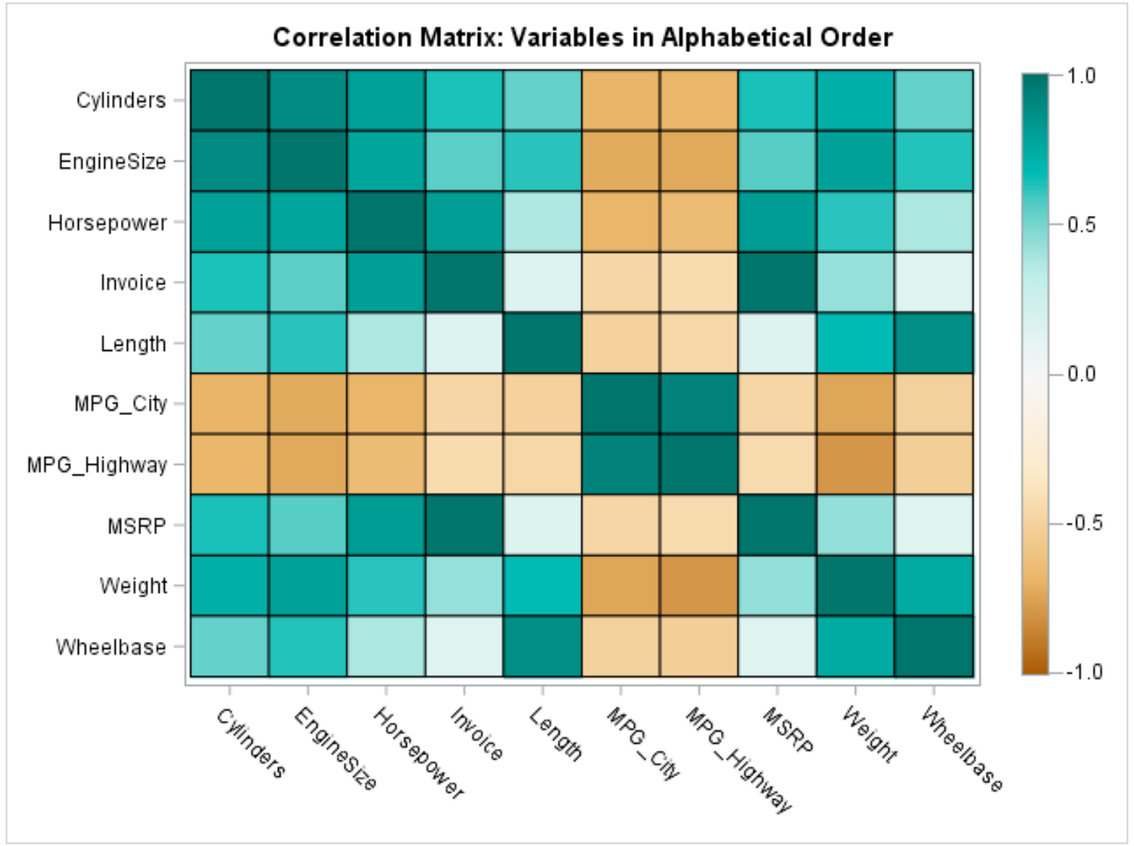
➔ *2 Keys*

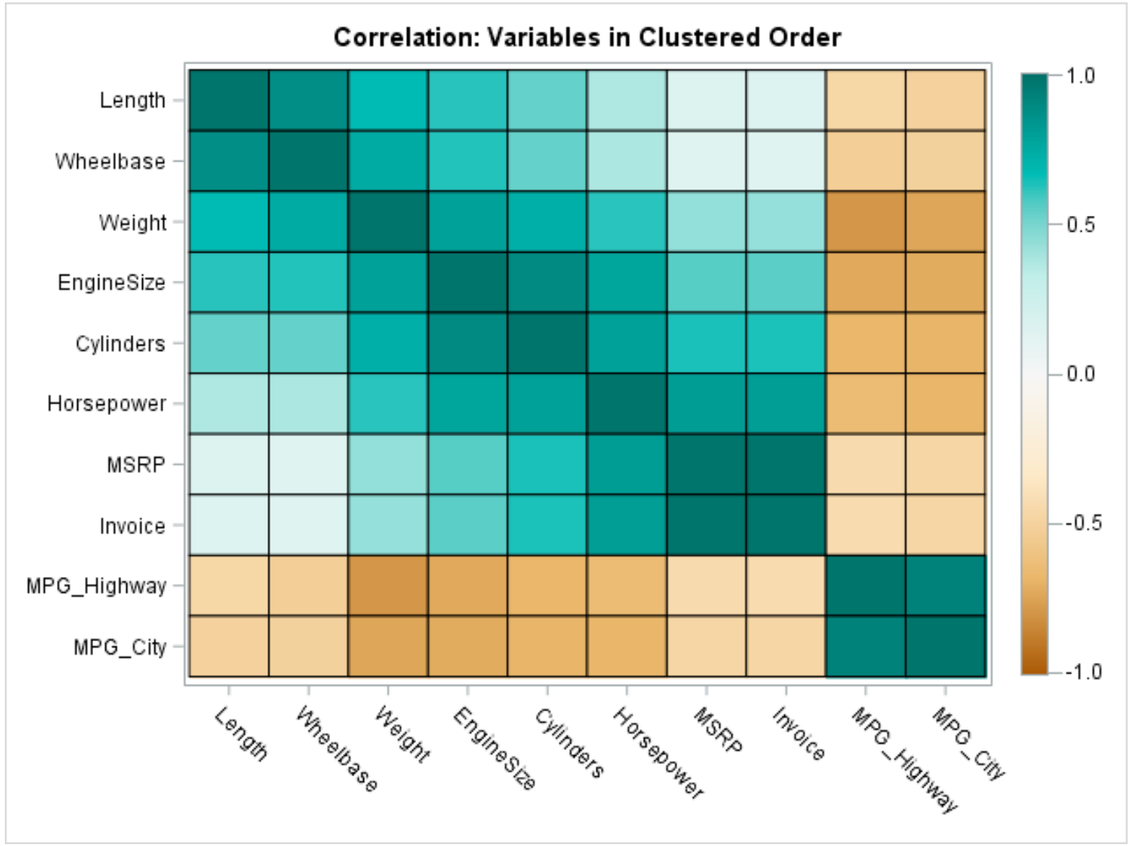
Matrix





Heatmap reordering

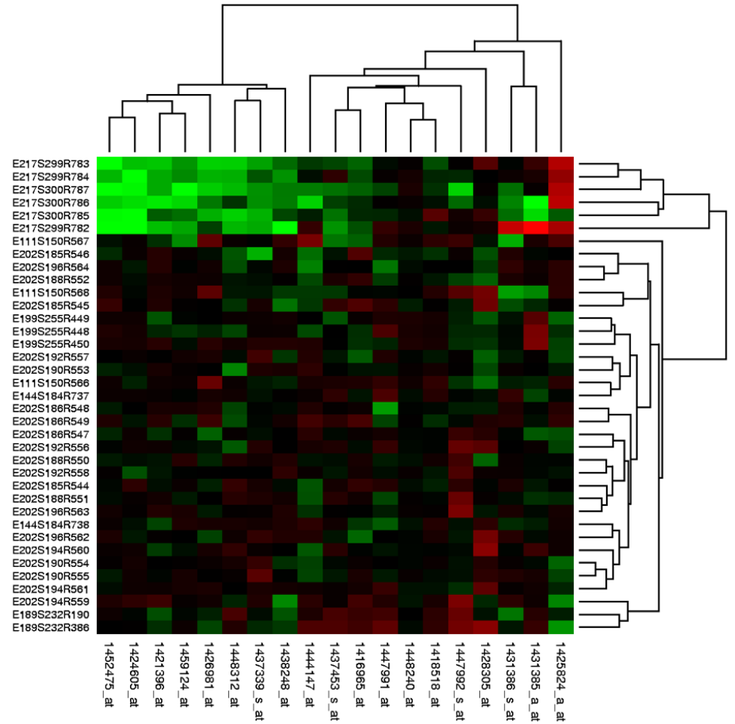




https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html

Idiom: **cluster heatmap**

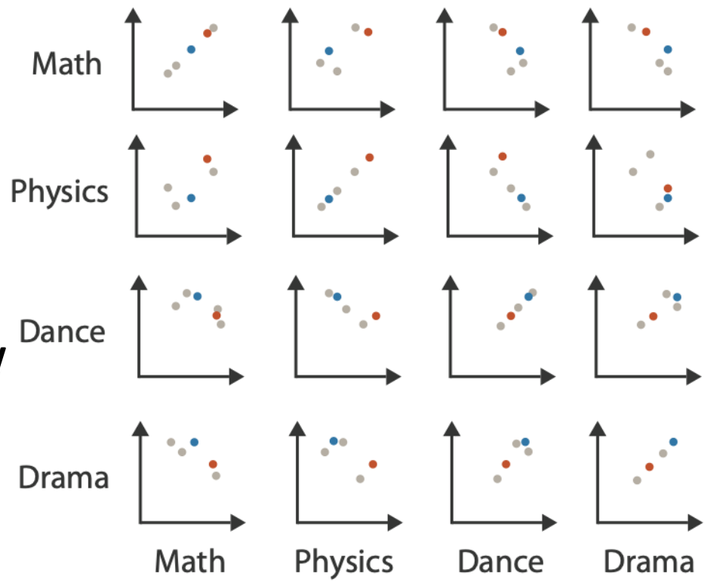
- in addition
  - derived data
    - 2 cluster hierarchies
  - dendrogram
    - parent-child relationships in tree with connection line marks
    - leaves aligned so interior branch heights easy to compare
  - heatmap
    - marks (re-)ordered by cluster hierarchy traversal
    - task: assess quality of clusters found by automatic methods



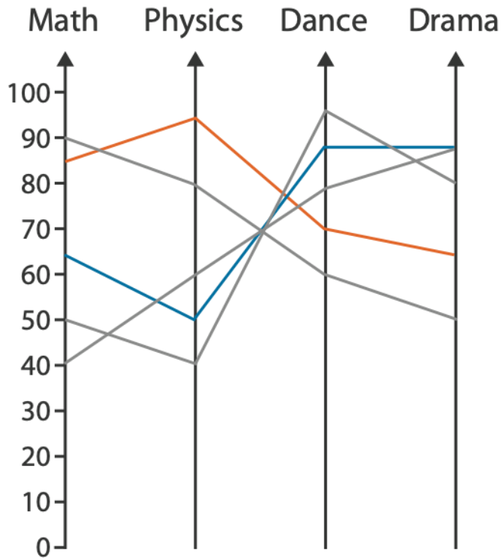
Idioms: **parallel coordinates**

- scatterplot limitation
  - visual representation with orthogonal axes
  - can show only two attributes with spatial position channel
- alternative: line up axes in parallel to show many attributes with position
  - item encoded with a line with n segments
  - n is the number of attributes shown
- parallel coordinates
  - parallel axes, jagged line for item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
    - hundreds of items

Scatterplot Matrix



Parallel Coordinates



Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

after [Visualization Course Figures. McGuffin, 2014. http://www.michaelmcguffin.com/courses/vis/]

How?

Encode

- **Arrange**
  - Express
  - Separate
  - Order
  - Align
- Use

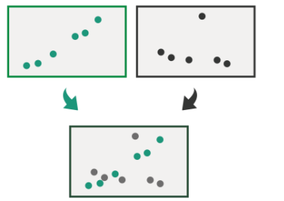
- **Map**
  from categorical and ordered attributes
  - Color
    - *Hue*
    - *Saturation*
    - *Luminance*
  - Size, Angle, Curvature, ...
  - Shape
  - Motion
    Direction, Rate, Frequency, ...

Manipulate

- **Change**
- **Select**
- **Navigate**

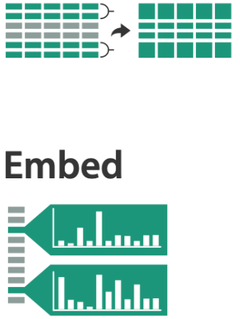
Facet

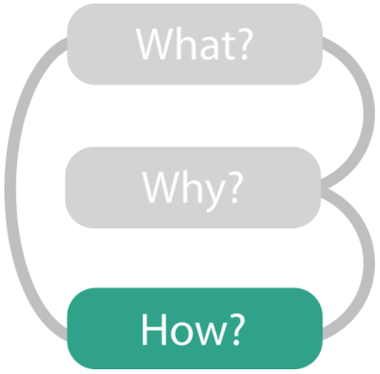
- **Juxtapose**
- **Partition**
- **Superimpose**



Reduce

- **Filter**
- **Aggregate**
- **Embed**

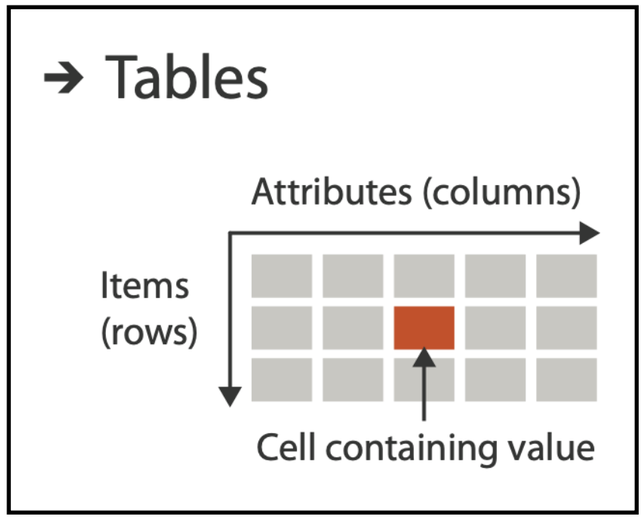


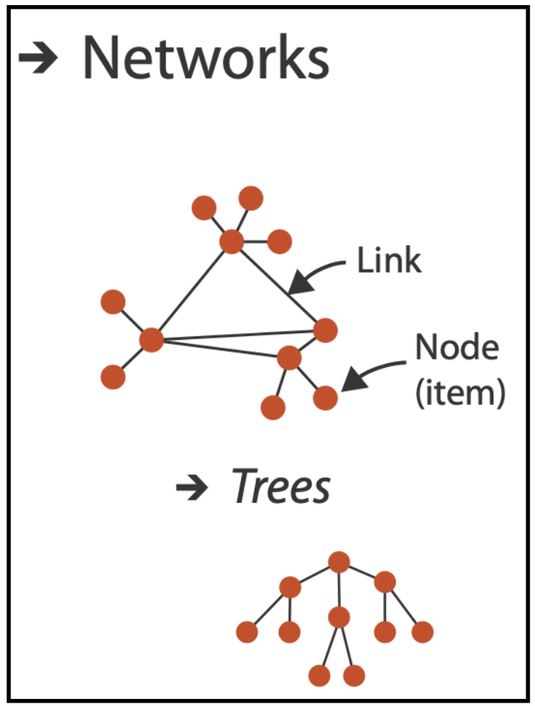


Dataset types

Tables	Networks & Trees
Items	Items (nodes)
Attributes	Links
	Attributes

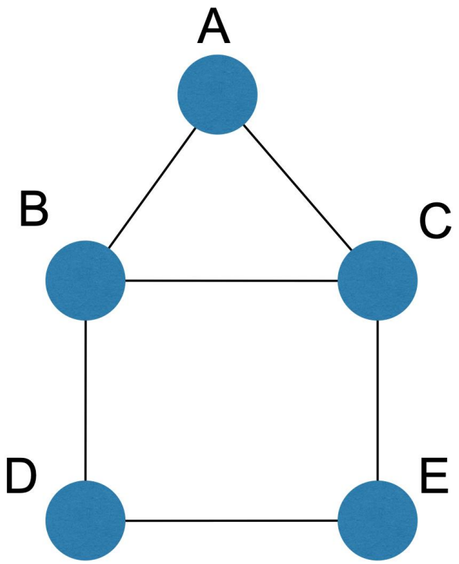
- network/graph
  - –nodes (vertices) connected by links (edges)
  - –tree is special case: no cycles
    - •often have roots and are directed

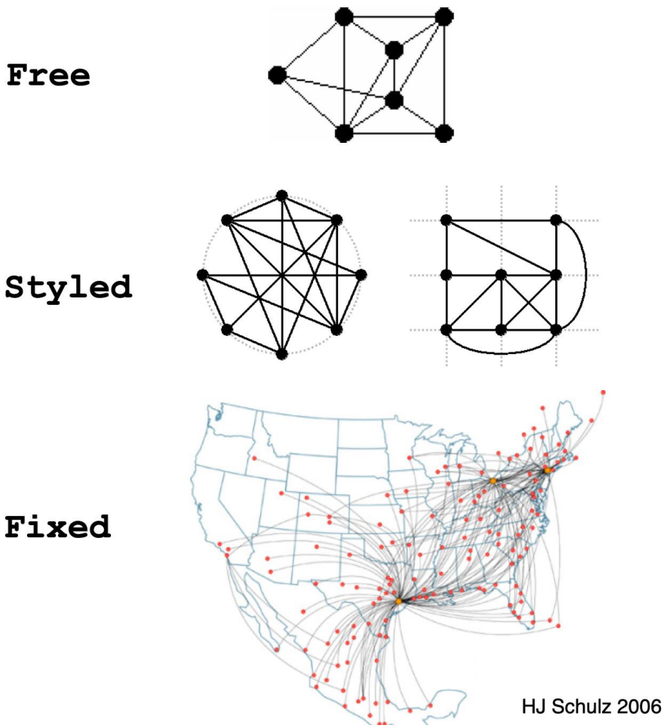




Node-link diagrams

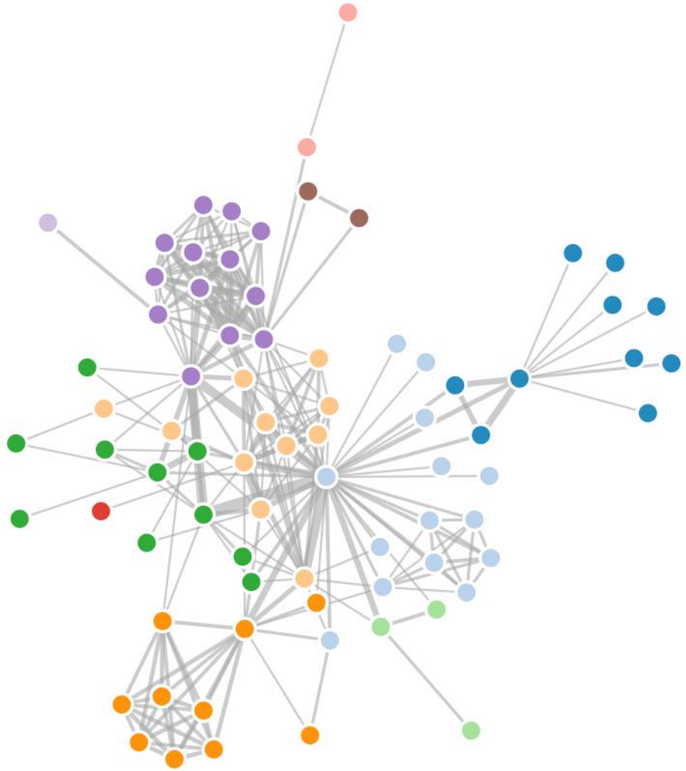
- nodes: point marks
- links: line marks
  - straight lines or arcs
  - connections between nodes
- intuitive & familiar
  - most common
  - many, many variants





Idiom: **force-directed placement**

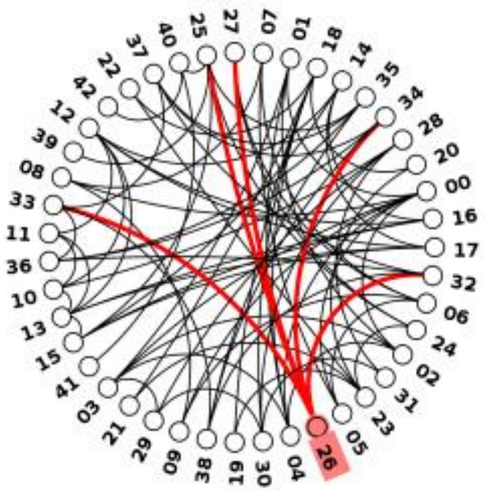
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short
- tasks
  - explore topology; locate paths, clusters
- scalability
  - node/edge density E < 4N

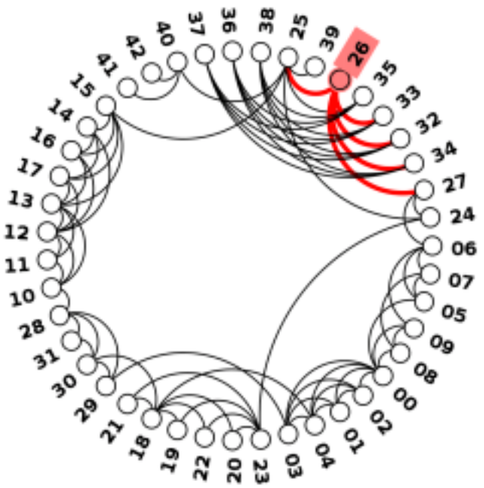


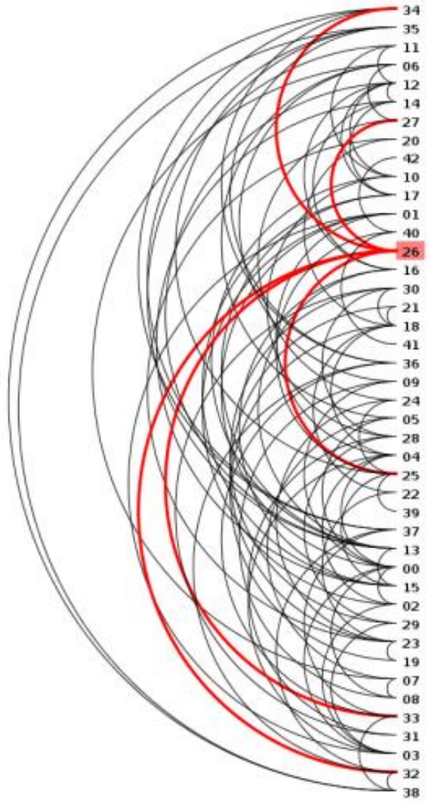
http://mbostock.github.com/d3/ex/force.html

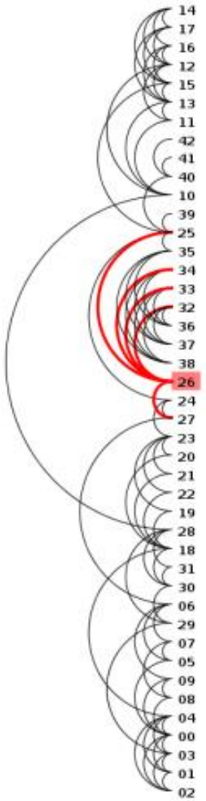
Idiom: **circular layouts / arc diagrams (node-link)**

- restricted node-link layouts: lay out nodes around circle or along line
- data
  - original: network
  - derived: node ordering attribute (global computation)
- considerations: node ordering crucial to avoid excessive clutter from edge crossings
  - examples: before & after barycentric ordering









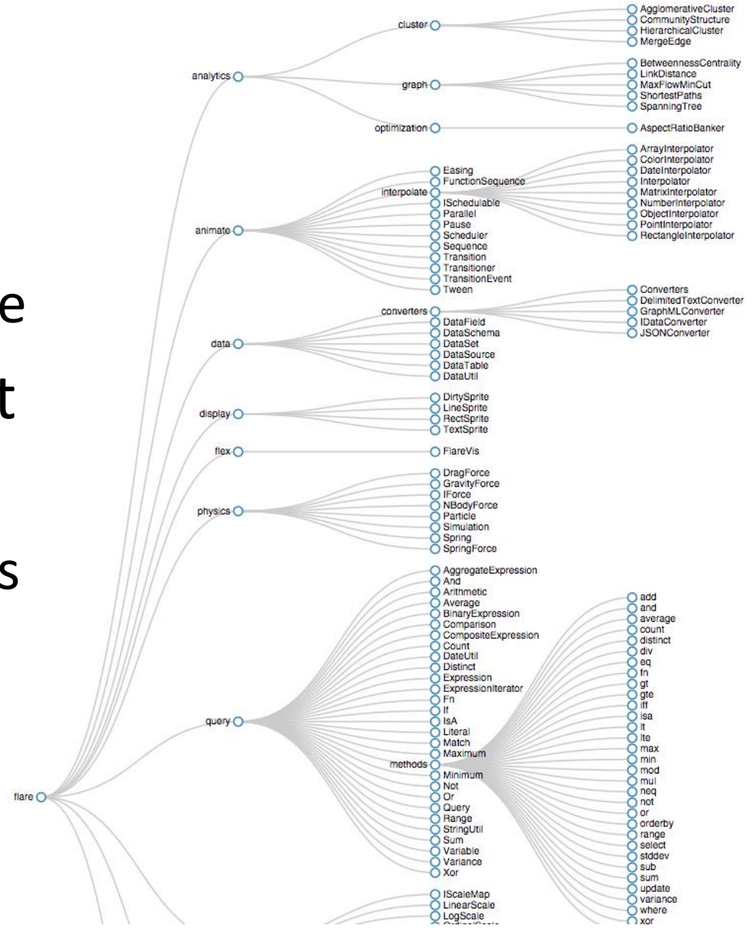
http://profs.etsmtl.ca/mmcguffin/research/2012-mcguffin-simpleNetVis/mcguffin-2012-simpleNetVis.pdf

Node-link trees

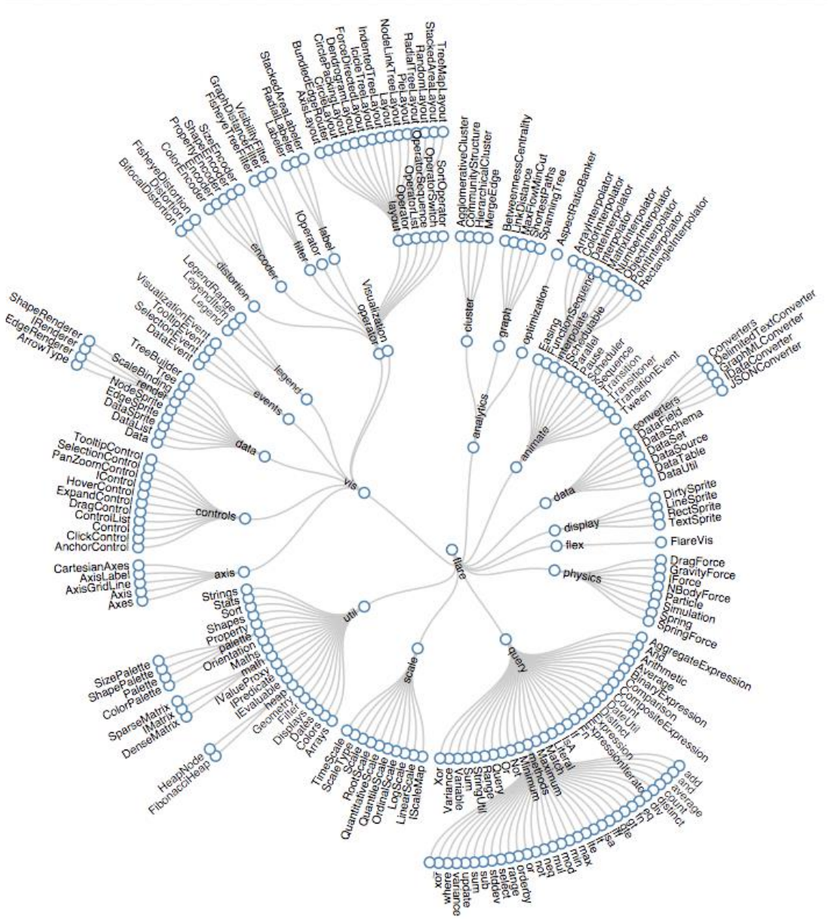
- Reingold-Tilford
  - tidy drawings of trees
    - exploit parent/child structure
  - allocate space: compact but without overlap
    - rectilinear and radial variants

[Tidier drawing of trees. Reingold and Tilford. IEEE Trans. Software Eng., SE-7(2):223–228, 1981.]

  - nice algorithm writeup
    http://billmill.org/pymag-trees/



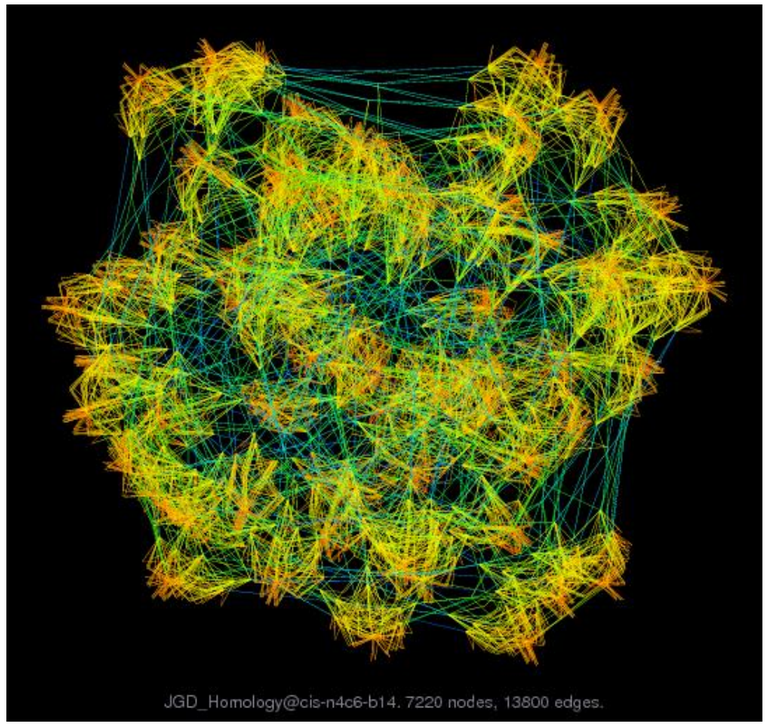
http://bl.ocks.org/mbostock/4339184



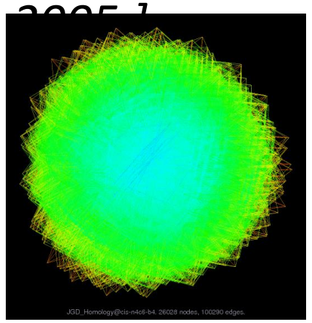
http://bl.ocks.org/mbostock/4063550

Idiom: **sfdp** (multi-level force-directed placement)

- data: compound graph
  - original: network
  - derived: cluster hierarchy atop it
- considerations
  - better algorithm for same encoding technique
    - same: fundamental use of space
    - hierarchy used for algorithm speed/quality but not shown explicitly
- scalability
  - nodes, edges: 1K-10K
  - hairball problem eventually hits



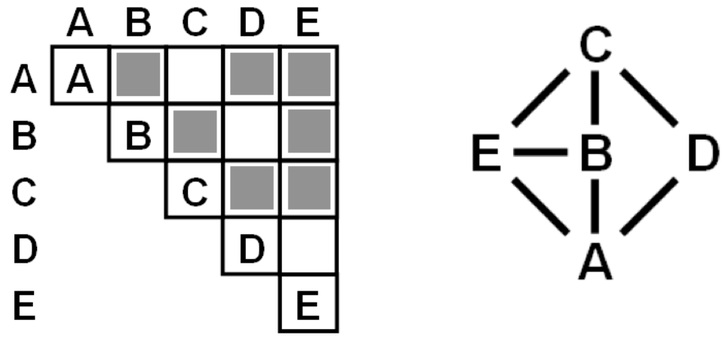
[Efficient and high quality force-directed graph drawing. Hu. The Mathematica Journal 10:37–71, 2005.]



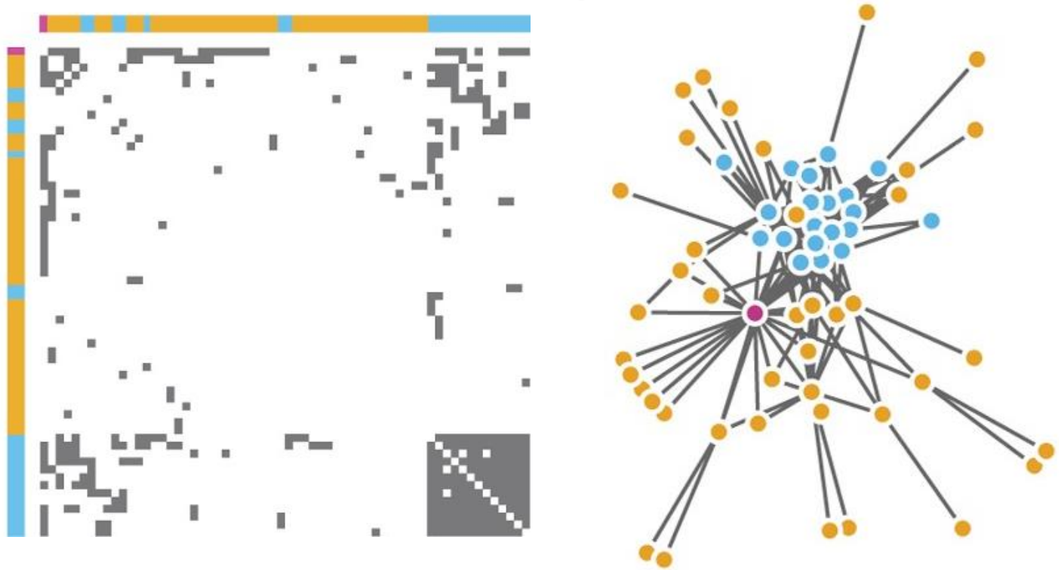
http://www.research.att.com/yifanhu/GALLERY/GRAPHS/index1.h

Idiom: **adjacency matrix view**

- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quant attrib
    - weighted edge between nodes
  - 2 categ attribs: node list x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - 1K nodes, 1M edges



[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]

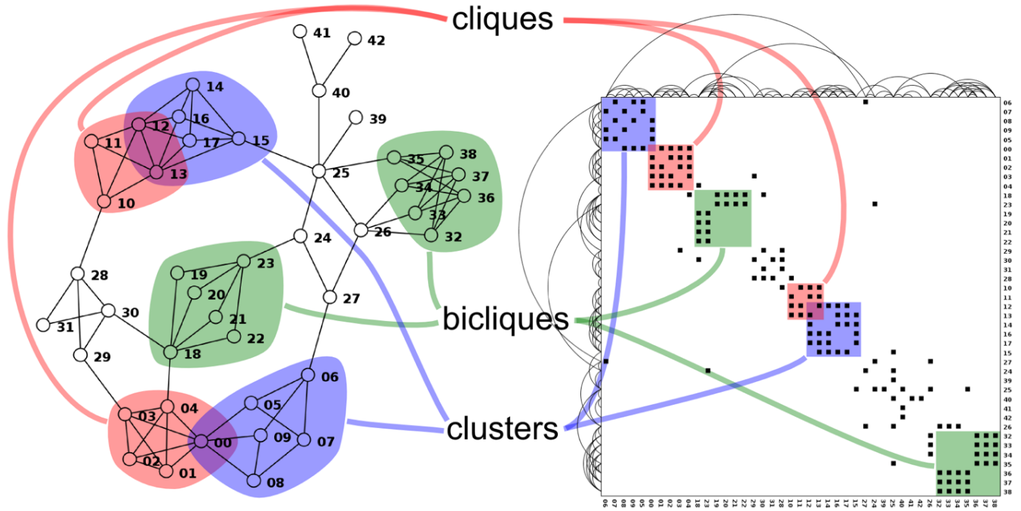


[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]

Node-link vs. matrix comparison

- node-link diagram strengths
  - topology understanding, path tracing
  - intuitive, flexible, no training needed
- adjacency matrix strengths
  - focus on edges rather than nodes
  - layout straightforward (reordering needed)
  - predictability, scalability
  - some topology tasks trainable
- empirical study
  - node-link best for small networks
  - matrix best for large networks
    - if tasks don't involve path tracing!

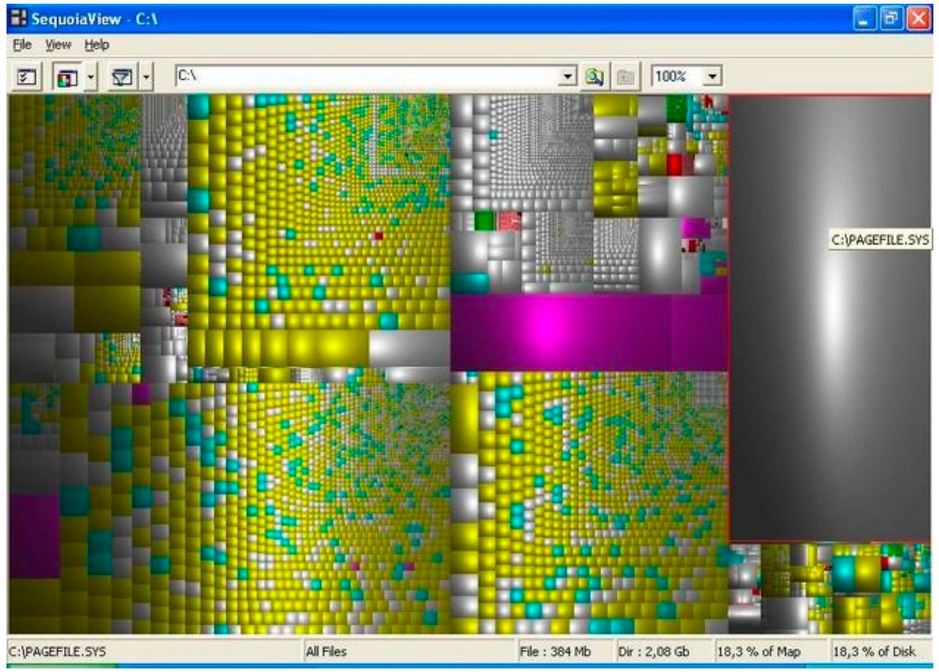
[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Ghoniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114–135.]



http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png

Idiom: **treemap**

- data
  - tree
  - 1 quant attrib at leaf nodes
- encoding
  - area containment marks for hierarchical structure
  - rectilinear orientation
  - size encodes quant attrib
- tasks
  - query attribute at leaf nodes
  - ex: disk space usage within filesystem
- scalability
  - 1M leaf nodes



https://www.win.tue.nl/sequoiaview/

[Cushion Treemaps. van Wijk and van de Wetering. Proc. Symp. InfoVis 1999, 73-78.]

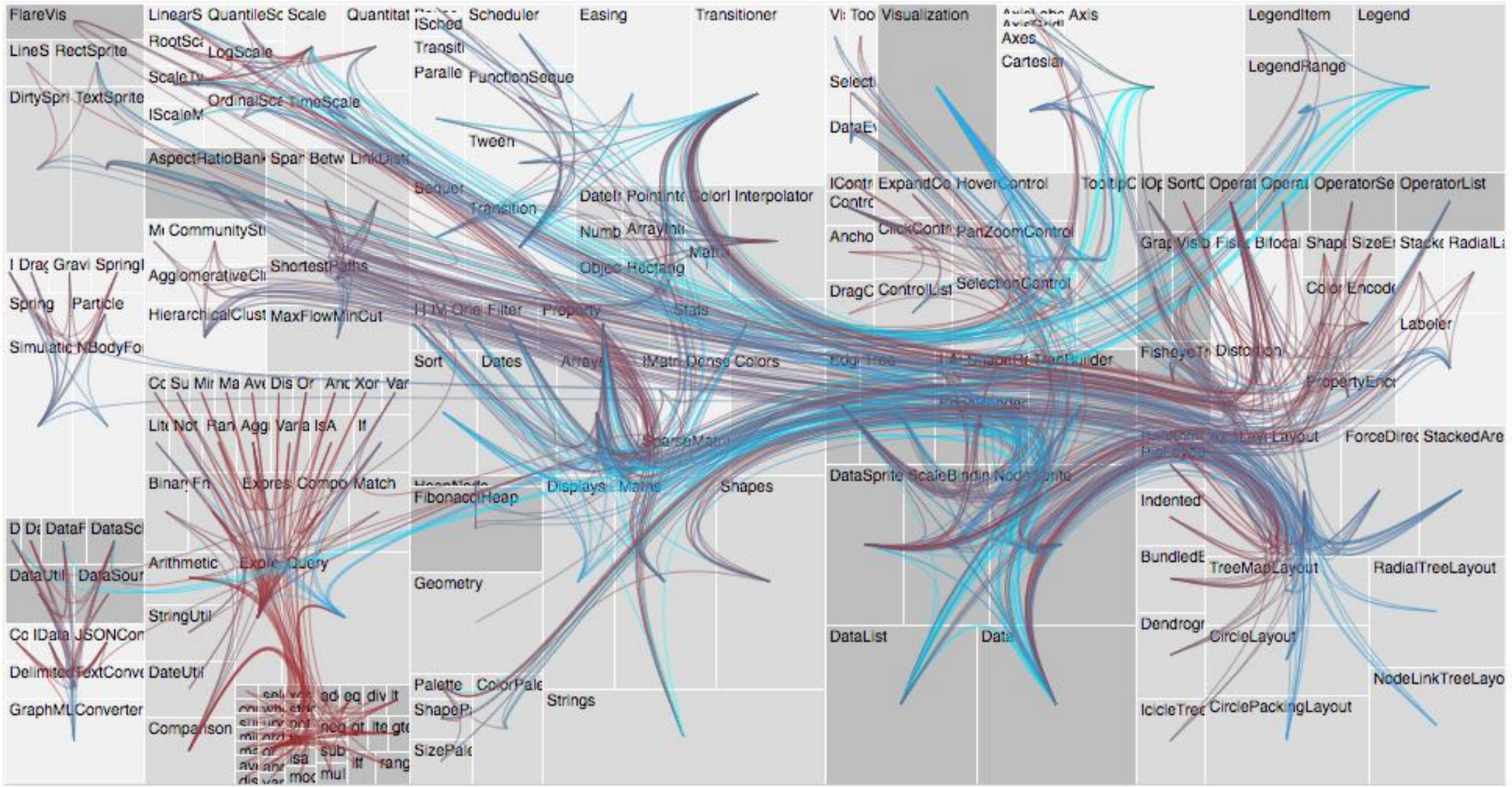
Enclosure

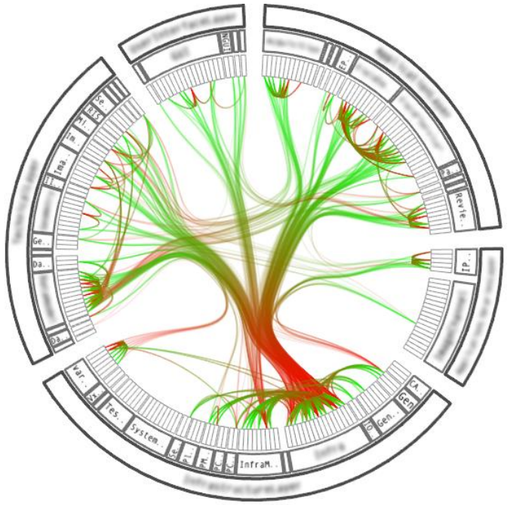
Containment Marks

✕ NETWORKS ✔ TREES

Hierarchical edge bundling

- works for any layout: treemap vs radial



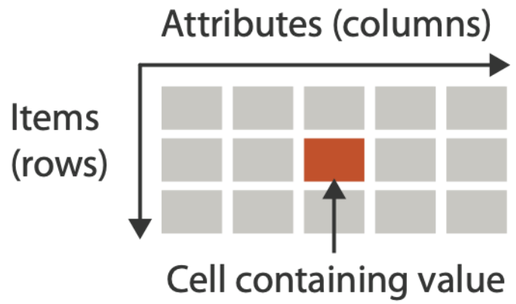


[Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. Danny Holten. TVCG 12(5):741-748 2006]

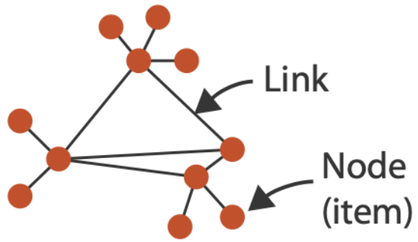
Dataset types

Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes

➔ Tables



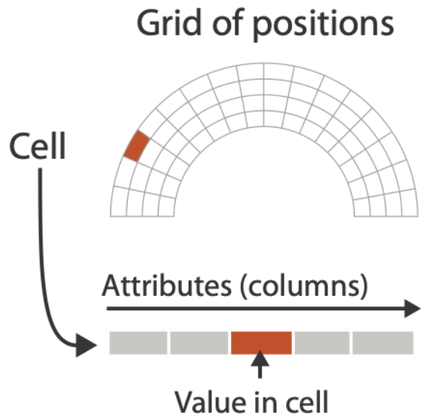
➔ Networks



➔ *Trees*

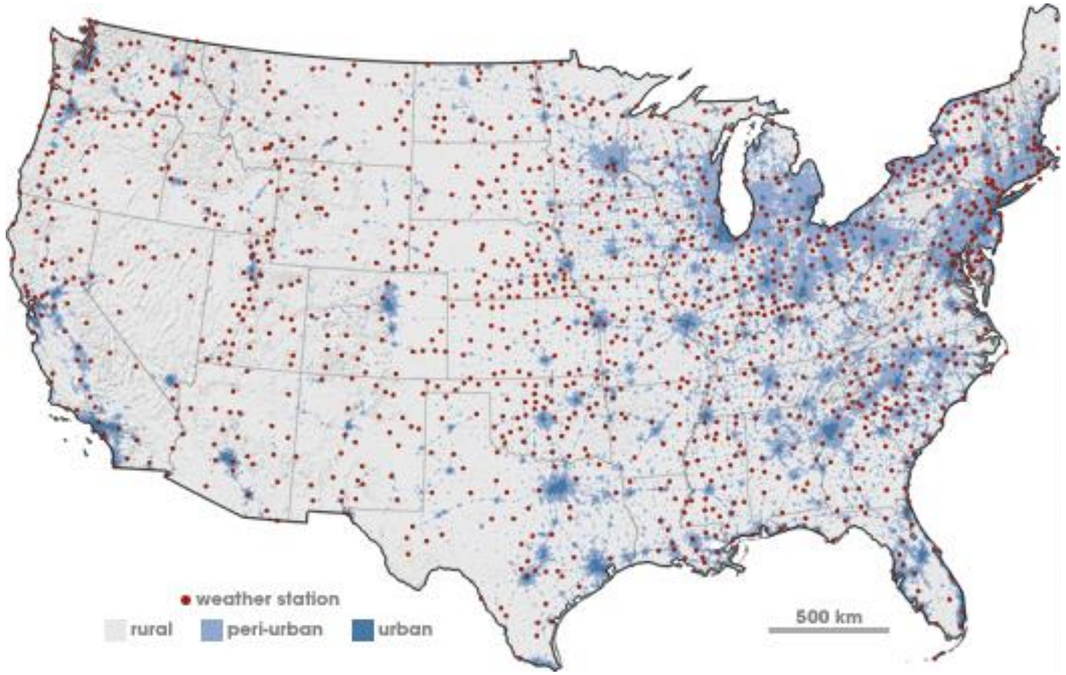
➔ Spatial

➔ Fields (Continuous)



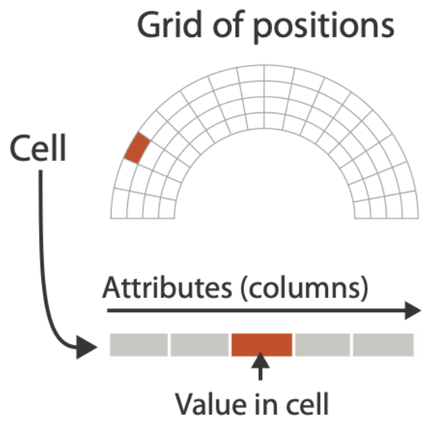
Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated



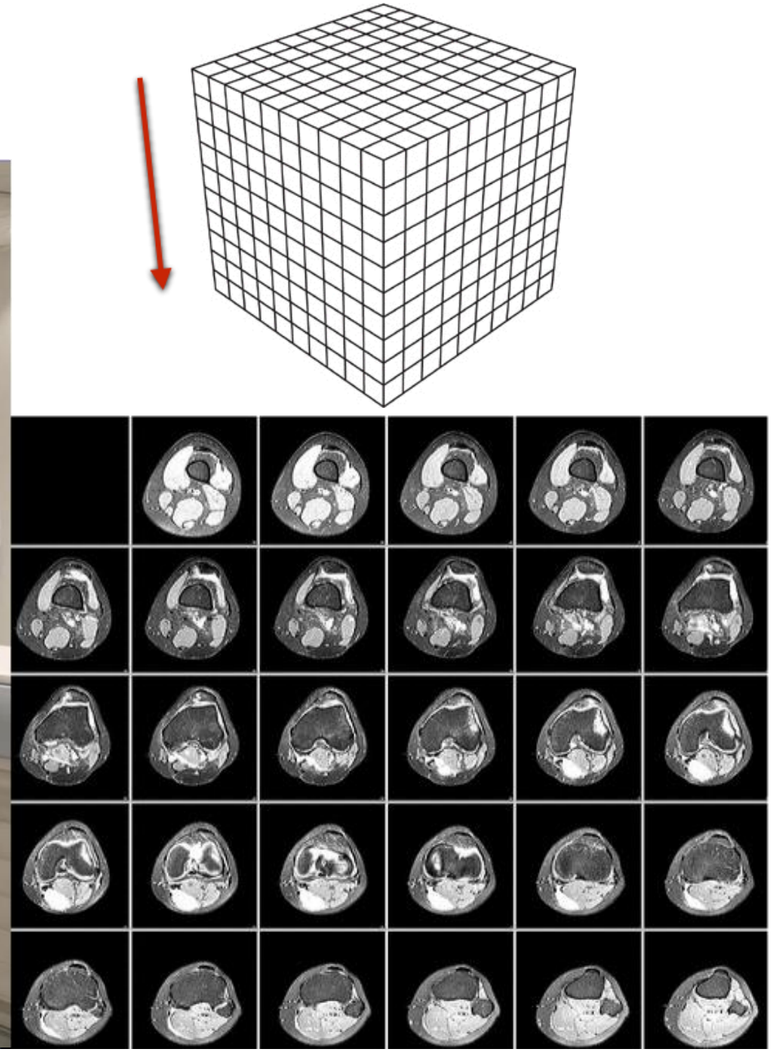
➔ Spatial

➔ Fields (Continuous)



Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling:
    where attributes are measured
  - interpolation:
    how to model attributes elsewhere
  - grid types



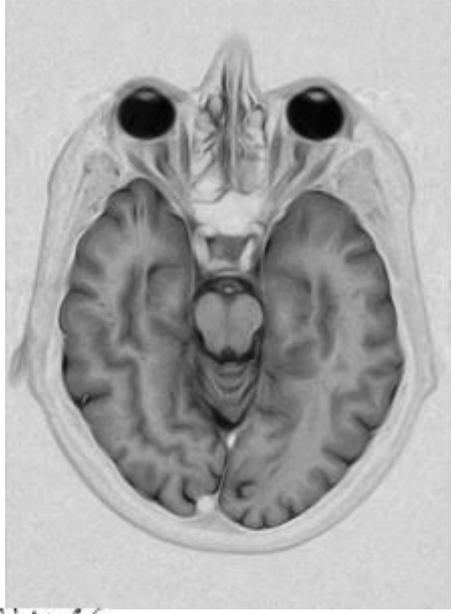
Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling:
    where attributes are measured
  - interpolation:
    how to model attributes elsewhere
  - grid types
- major divisions
  - attributes per cell:
    scalar (1), vector (2-4), tensor (many, 9)

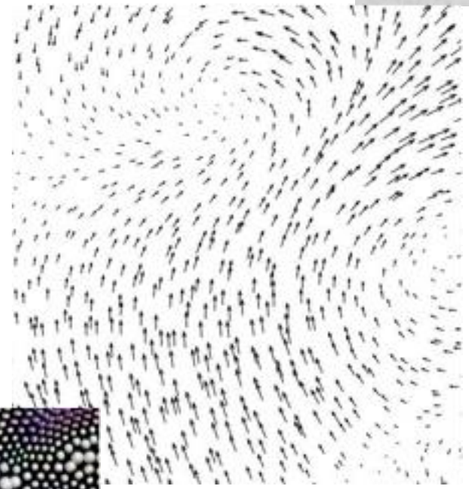
  Vector: 2 (Direction, Magnitude)

  Tensor: 9 (Sphere pulled in 3x3 directions)

scalar



vector

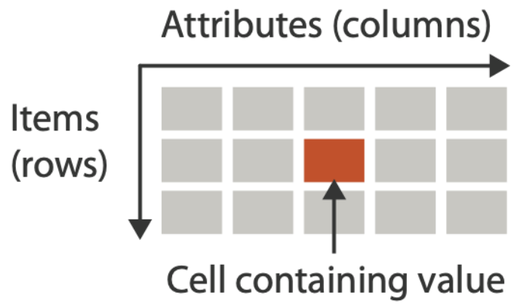


tensor

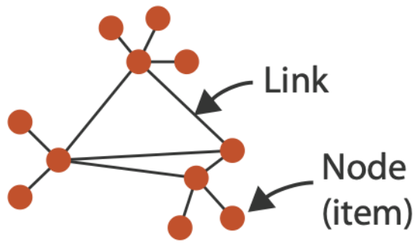
Dataset types

Tables	Networks & Trees	Fields	Geometry
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions
	Attributes	Attributes	

➔ Tables



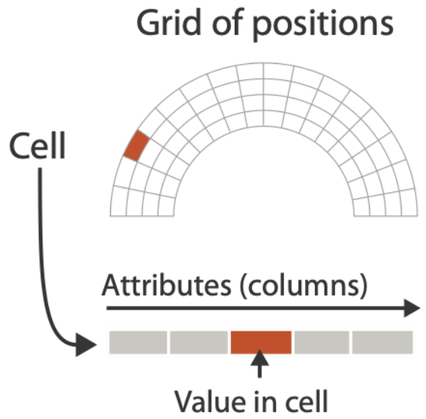
➔ Networks



➔ *Trees*

➔ Spatial

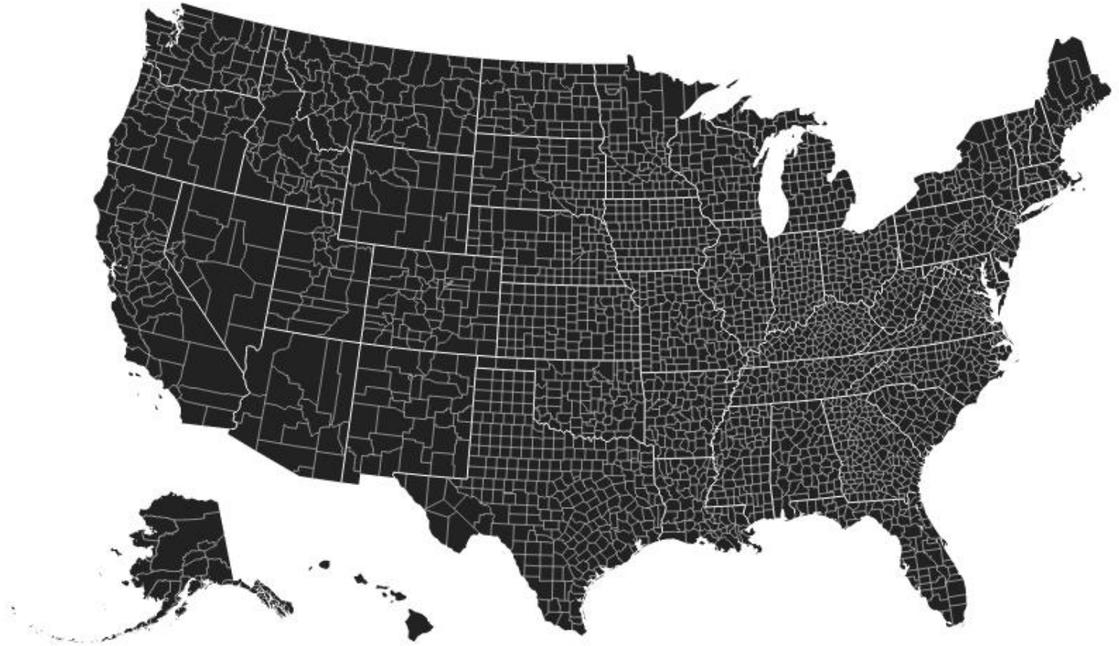
➔ Fields (Continuous)



➔ Geometry (Spatial)

Geometry

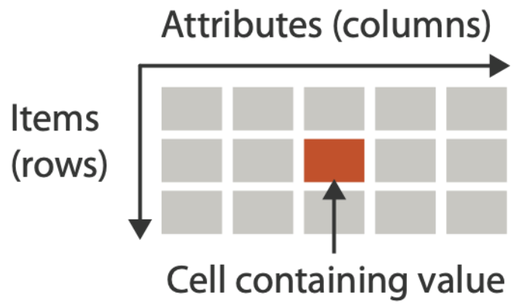
- shape of items
- explicit spatial positions / regions
  - points, lines, curves, surfaces, volumes
- boundary between computer graphics and visualization
  - graphics: geometry taken as given
  - vis: geometry is result of a design decision



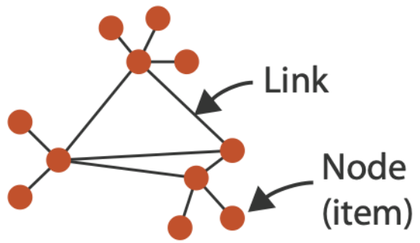
Dataset types

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

➔ Tables



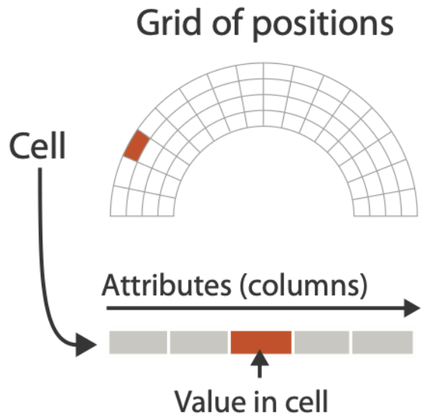
➔ Networks



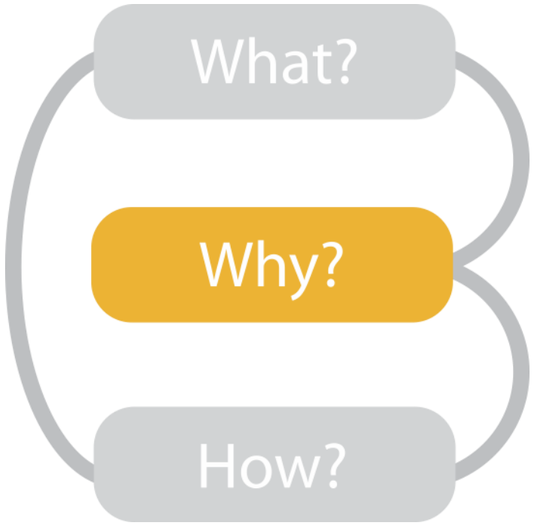
➔ *Trees*

➔ Spatial

➔ Fields (Continuous)



➔ Geometry (Spatial)



Why?

Actions

Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*



→ Produce

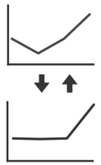
→ *Annotate* → *Record* → *Derive*


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

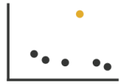
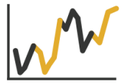
→ Identify → Compare → Summarize


Targets

All Data

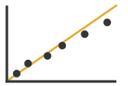
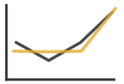
→ Trends → Outliers → Features

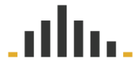
Attributes

→ One → Many

→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

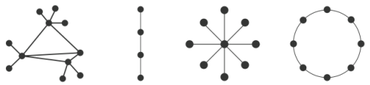



→ *Extremes*



Network Data

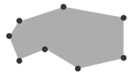
→ Topology

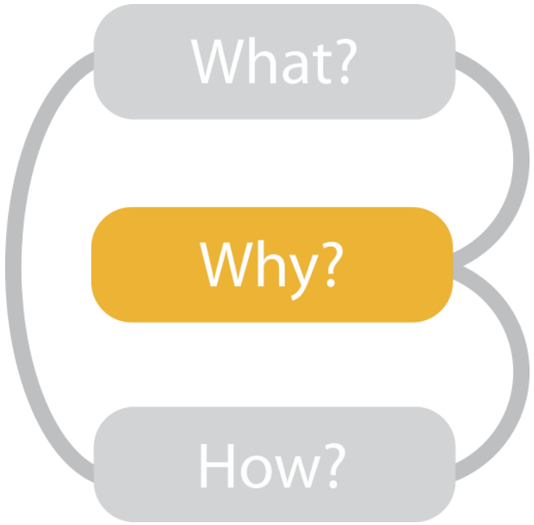


→ *Paths*

Spatial Data

→ Shape





Why *seeks to understand what* ***Task*** *you must take to achieve some* <u>*understanding*</u> *of the data!*

Why?

Actions

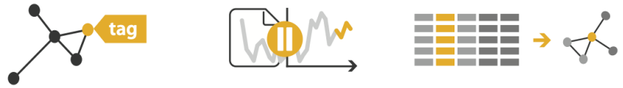
Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

→ *Annotate* → *Record* → *Derive*

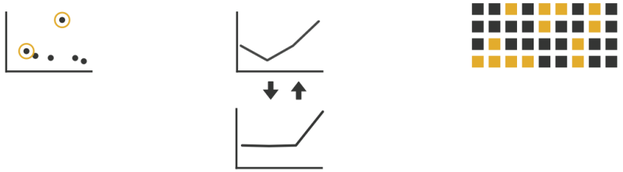


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

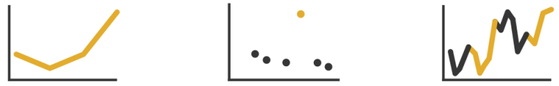
→ Identify → Compare → Summarize



Targets

All Data

→ Trends → Outliers → Features

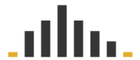


Attributes

→ One → Many

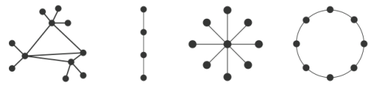
→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

→ *Extremes*



Network Data

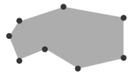
→ Topology



→ *Paths*

Spatial Data

→ Shape



Why?

Actions

- **Analyze**
  - → Consume
    - → *Discover*
    - → *Present*
    - → *Enjoy*
  - → Produce

Targets

- **All Data**
  - → Trends
  - → Outliers
  - → Features
- **Attributes**
  - → One
    - → *Distribution*
    - → *Extremes*
  - → Many
    - → *Dependency*
    - → *Correlation*
    - → *Similarity*
- **Network Data**
  - → Topology
  - → *Paths*
- **Spatial Data**
  - → Shape

Tasks:

- Compare/Similarity-Barchart/Lineplot
- Lookup-Barchart/Scatterplot
- Outliers-Scatterplot
- Part-to-Whole (Stacked bar chart/pie chart)
- Change over time/value-Streamgraph/lineplot/SlopeGraph
- Find Trends-Scatterplot/Barchart/LinePlot
- Find clusters-Scatterplot
- Correlation/Dependency-Scatterplot/Heatmap/Parallel Coords
- Emphasize Overlap/Dependencies-Gantt Charts/LinePlots

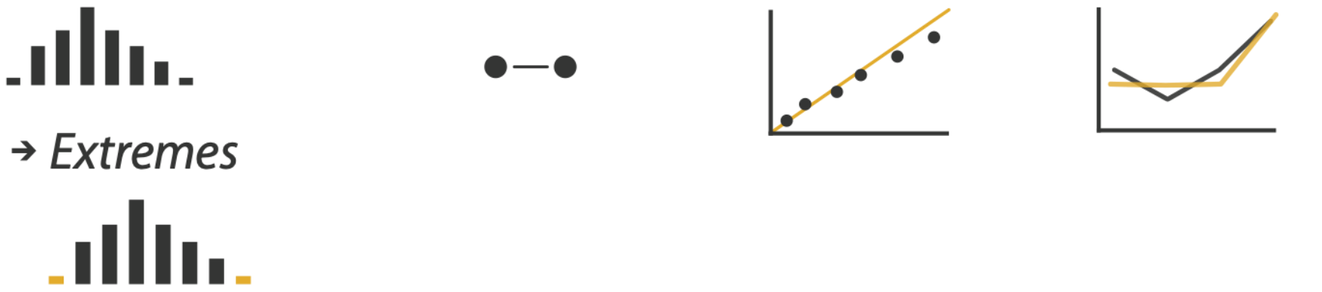
Task abstraction: Targets (data of interest to the user)

All Data

- Trends
- Outliers
- Features

Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*



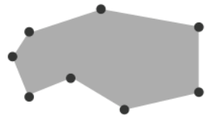
Network Data

- Topology

- *Paths*

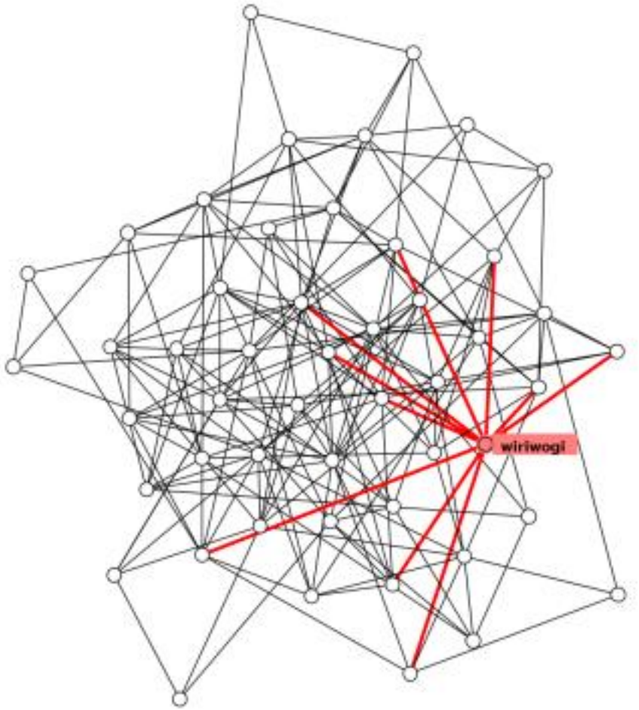
Spatial Data

- Shape

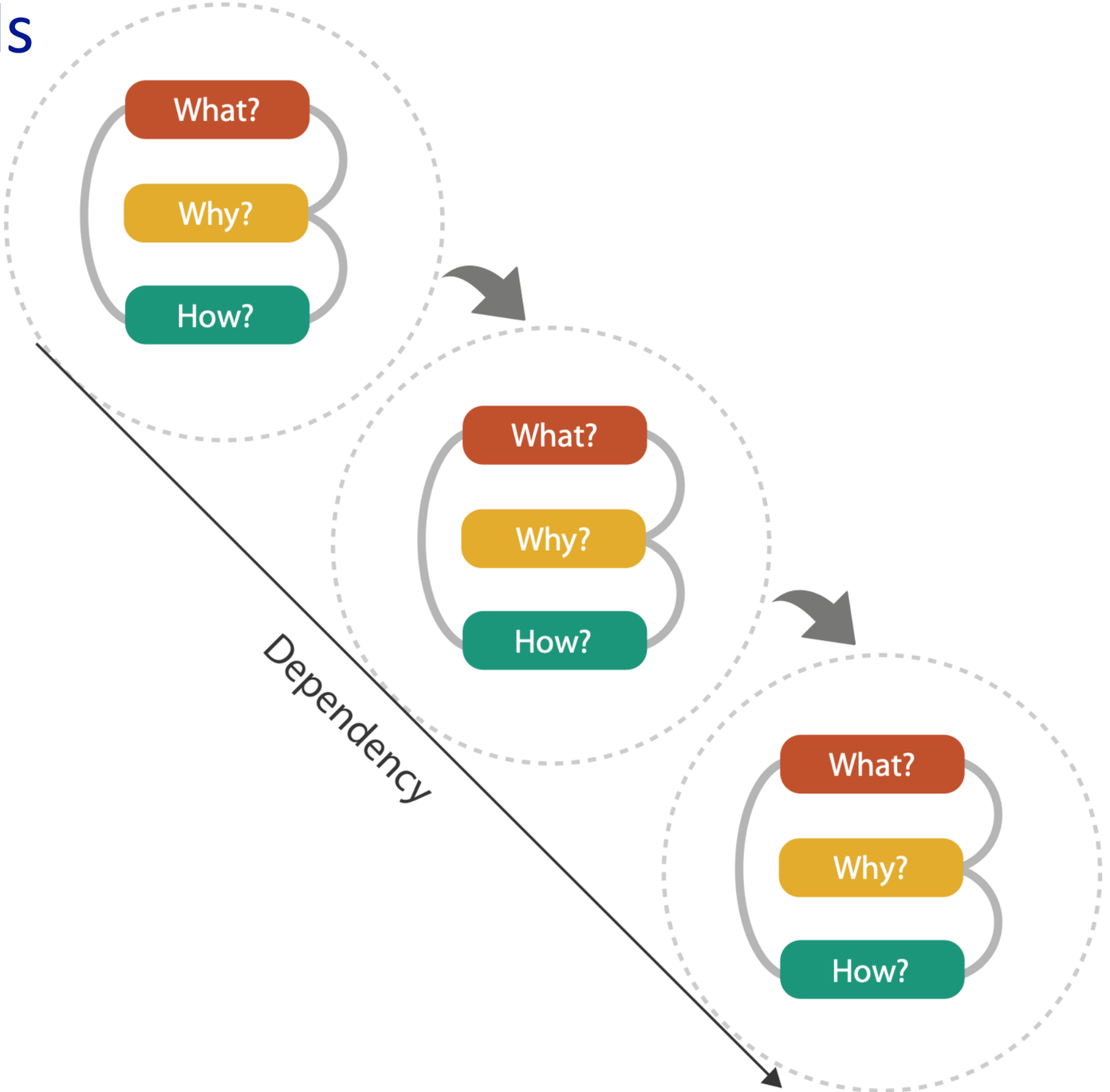


Network tasks: topology-based and attribute-based

- topology based tasks
  - find paths
  - find (topological) neighbors
  - compare centrality/importance measures
  - identify clusters / communities
- attribute based tasks (similar to table data)
  - find distributions, ...
- combination tasks, incorporating both
  - example: find friends-of-friends who like cats
    - topology: find all adjacent nodes of given node
    - attributes: check if has-pet (node attribute) == cat



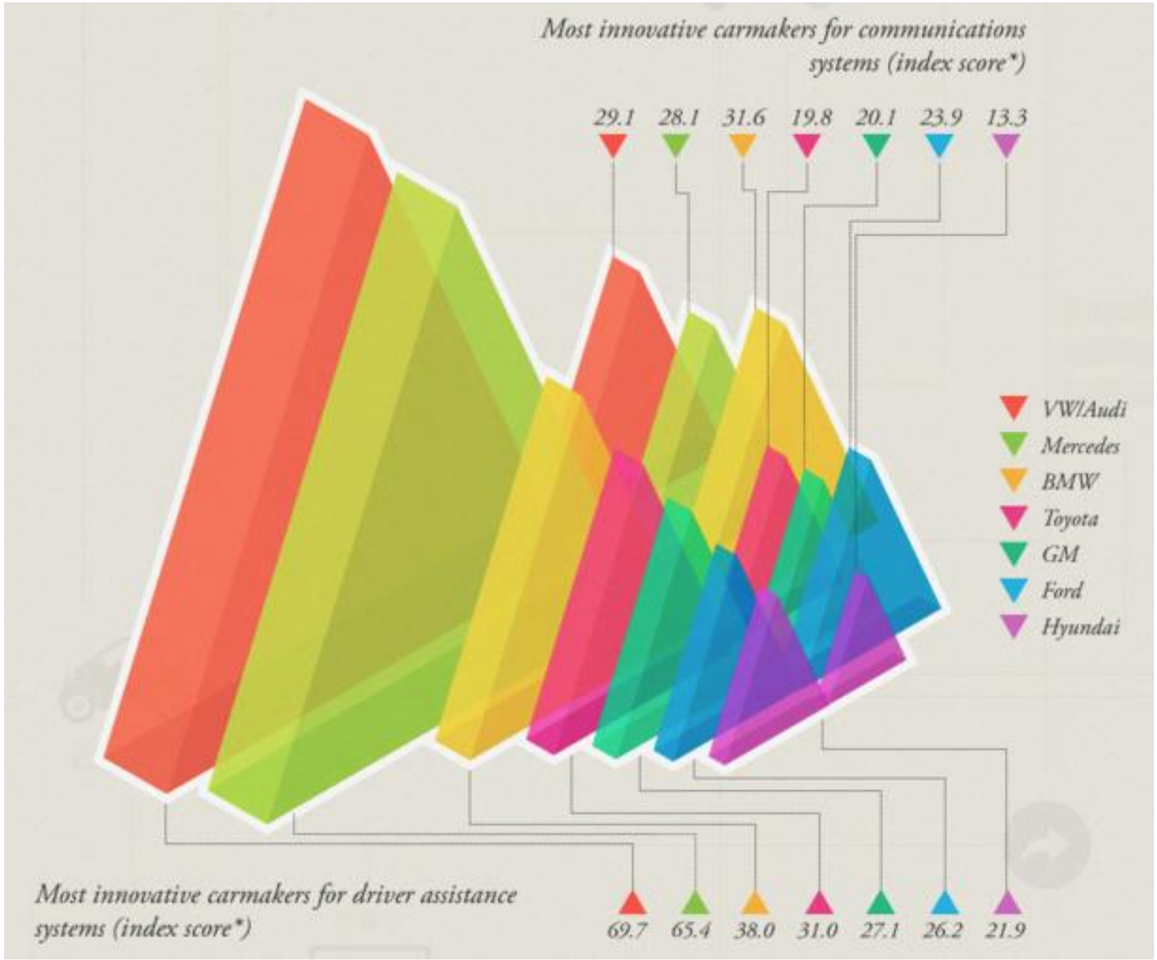
Means and ends

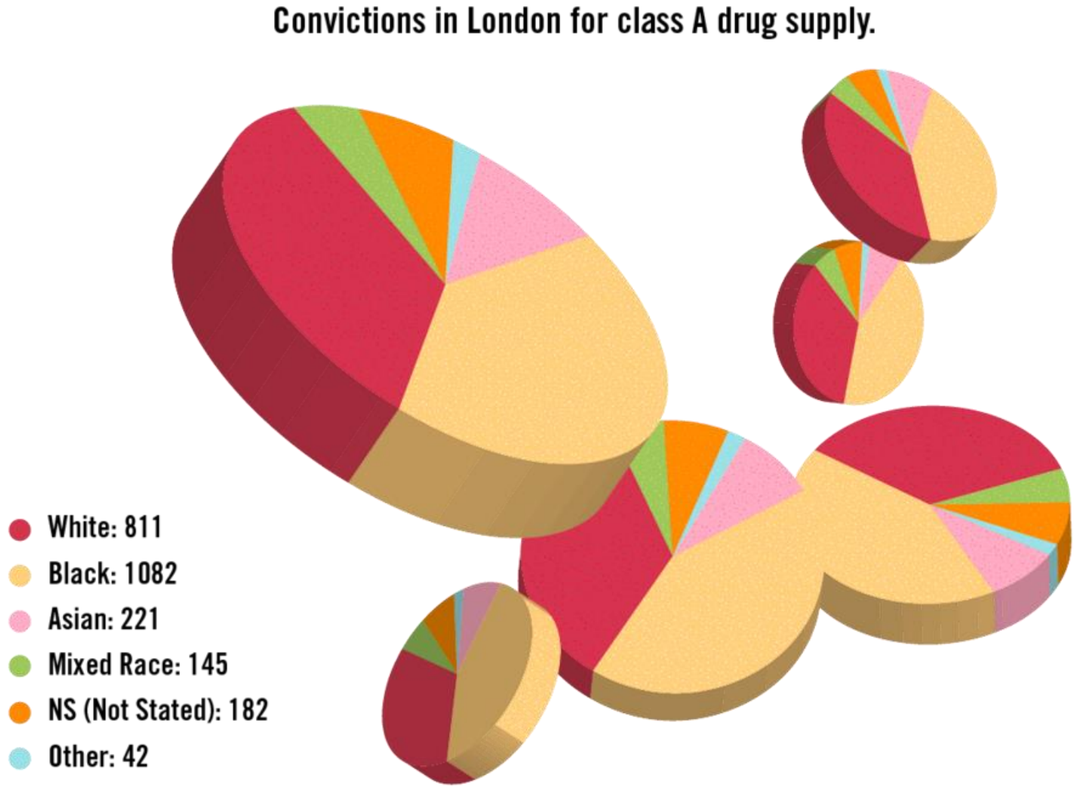


Rules of Thumb

- Guidelines and considerations, not absolute rules
  - when to use 3D? when to use 2D?
  - when to use eyes instead of memory?
  - when does immersion help?
  - when to use overviews?
  - how long is too long?
  - which comes first, form or function?

Unjustified 3D all too common, in the news and elsewhere



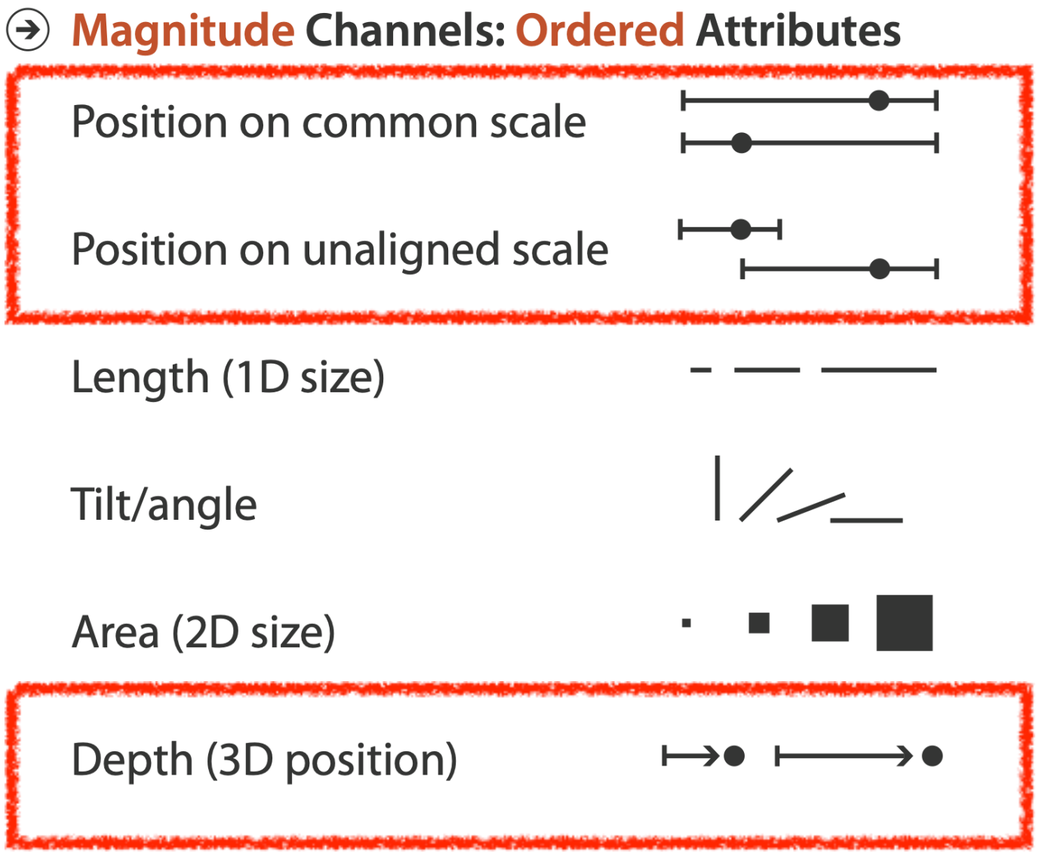


http://viz.wtf/post/137826497077/eye-popping-3d-triangles

http://viz.wtf/post/139002022202/designer-drugs-ht-ducqn

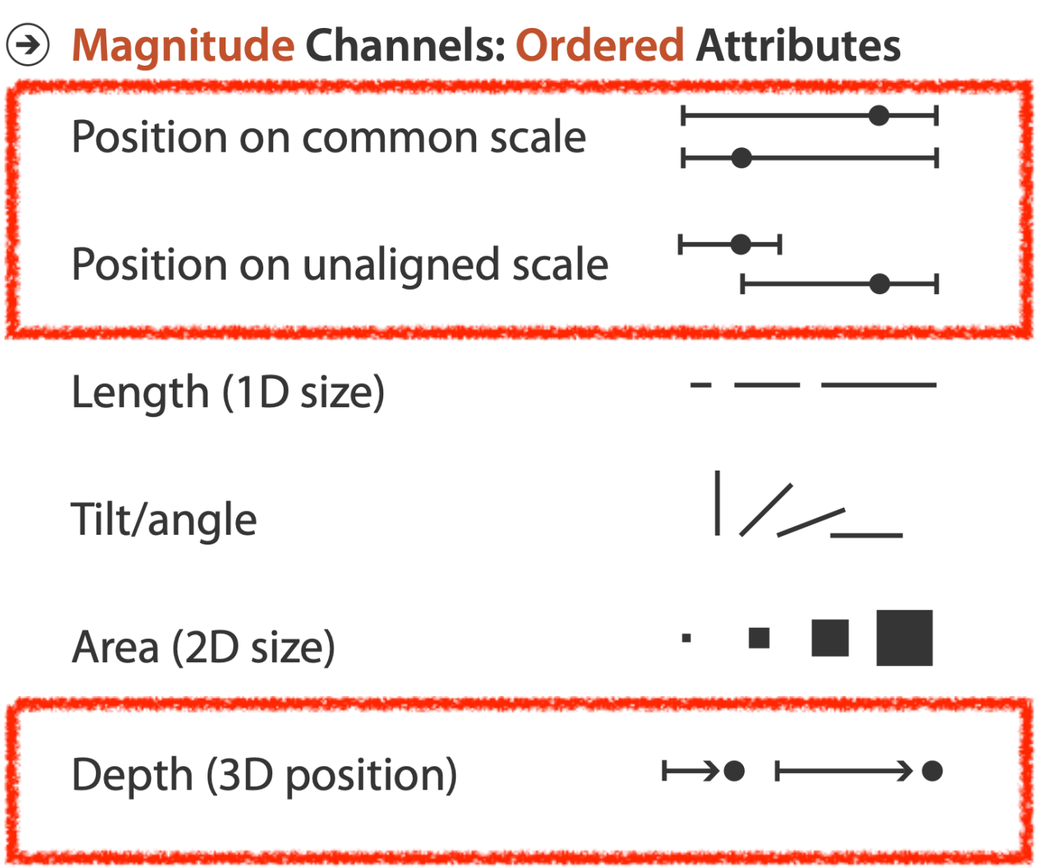
Depth vs power of the plane

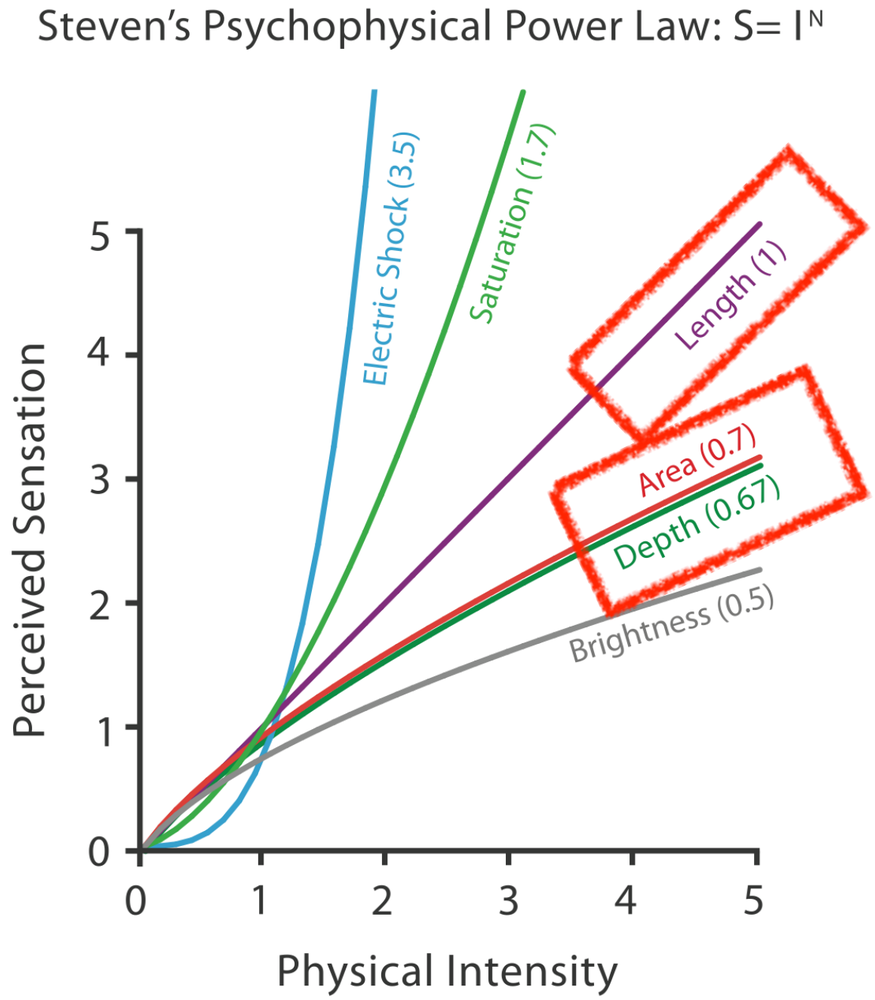
- high-ranked spatial position channels: **planar** spatial position
  - not depth!



Depth vs power of the plane

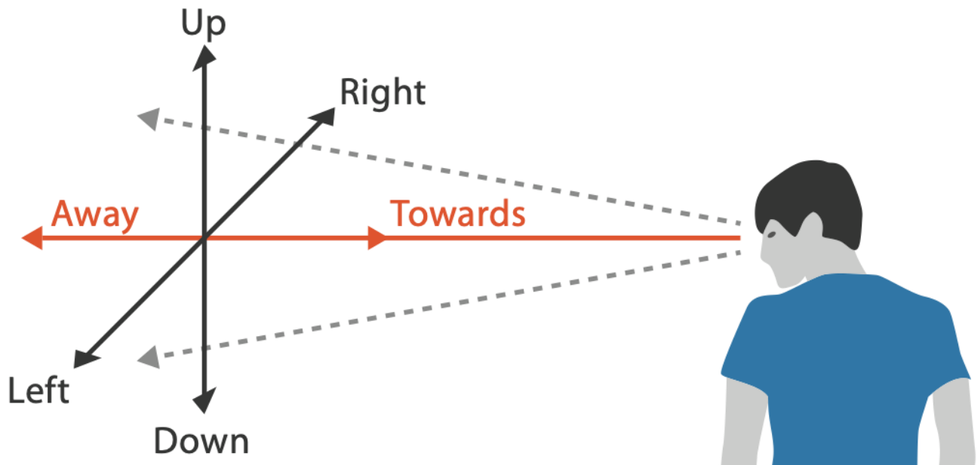
- high-ranked spatial position channels: **planar** spatial position
  - not depth!

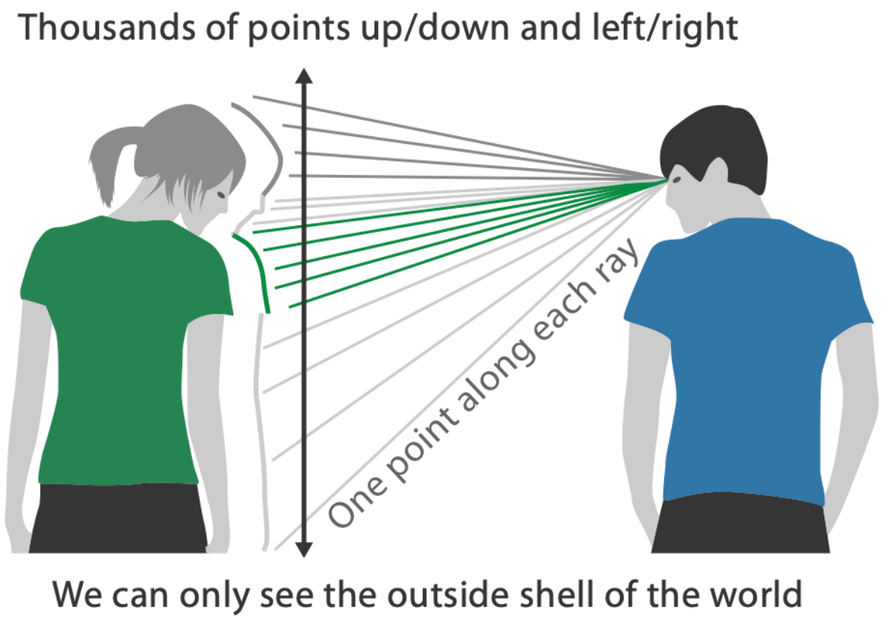




No unjustified 3D: Danger of depth

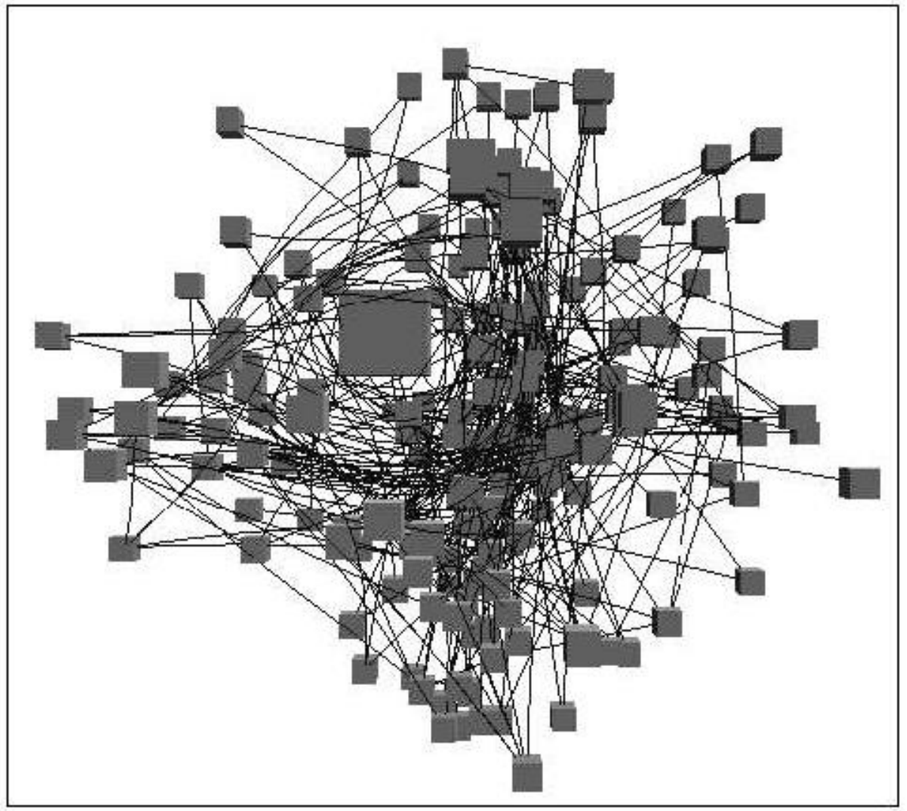
- we don't really live in 3D: we ***see*** in 2.05D
  - acquire more info on image plane quickly from eye movements
  - acquire more info for depth slower, from head/body motion





Occlusion hides information

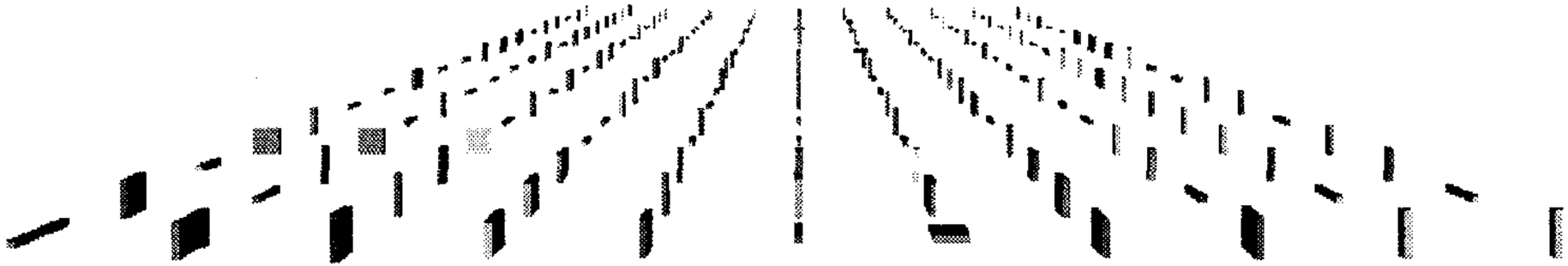
- occlusion
- interaction can resolve, but at cost of time and cognitive load



[Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.]

Perspective distortion loses information

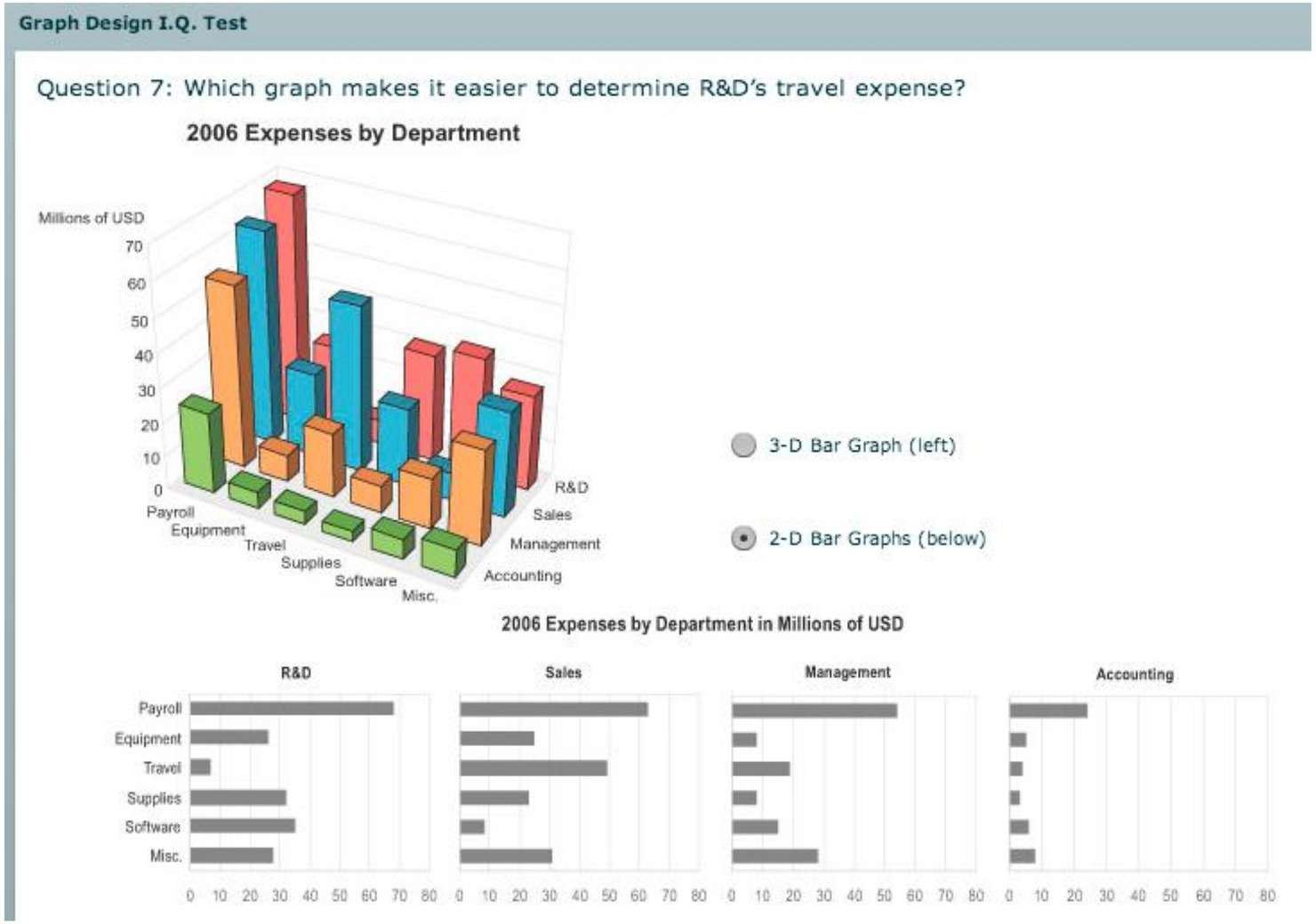
- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!



[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

3D vs 2D bar charts

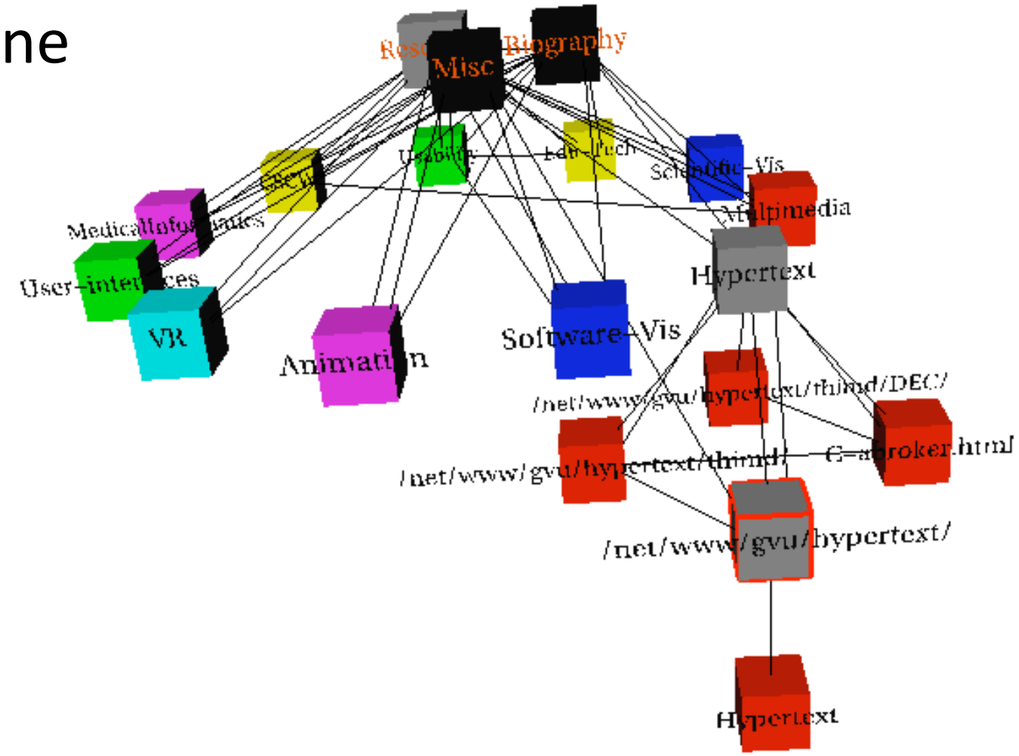
- 3D bars:
  very difficult to justify!
  - perspective distortion
  - occlusion
- faceting into 2D almost always better choice



[http://perceptualedge.com/files/GraphDesignIQ.html]

Tilted text isn’t legible

- text legibility
  - far worse when tilted from image plane



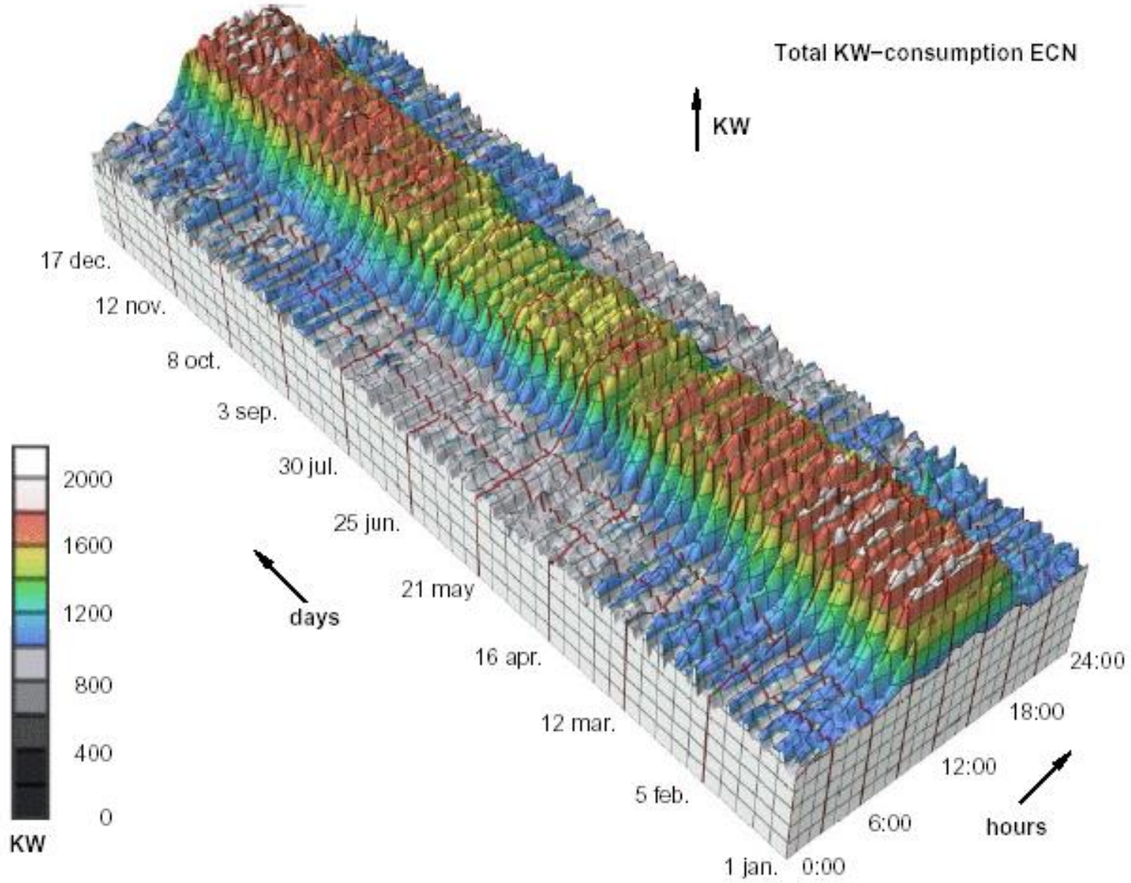
- further reading

Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007

[*Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.]*

No unjustified 3D example: Time-series data

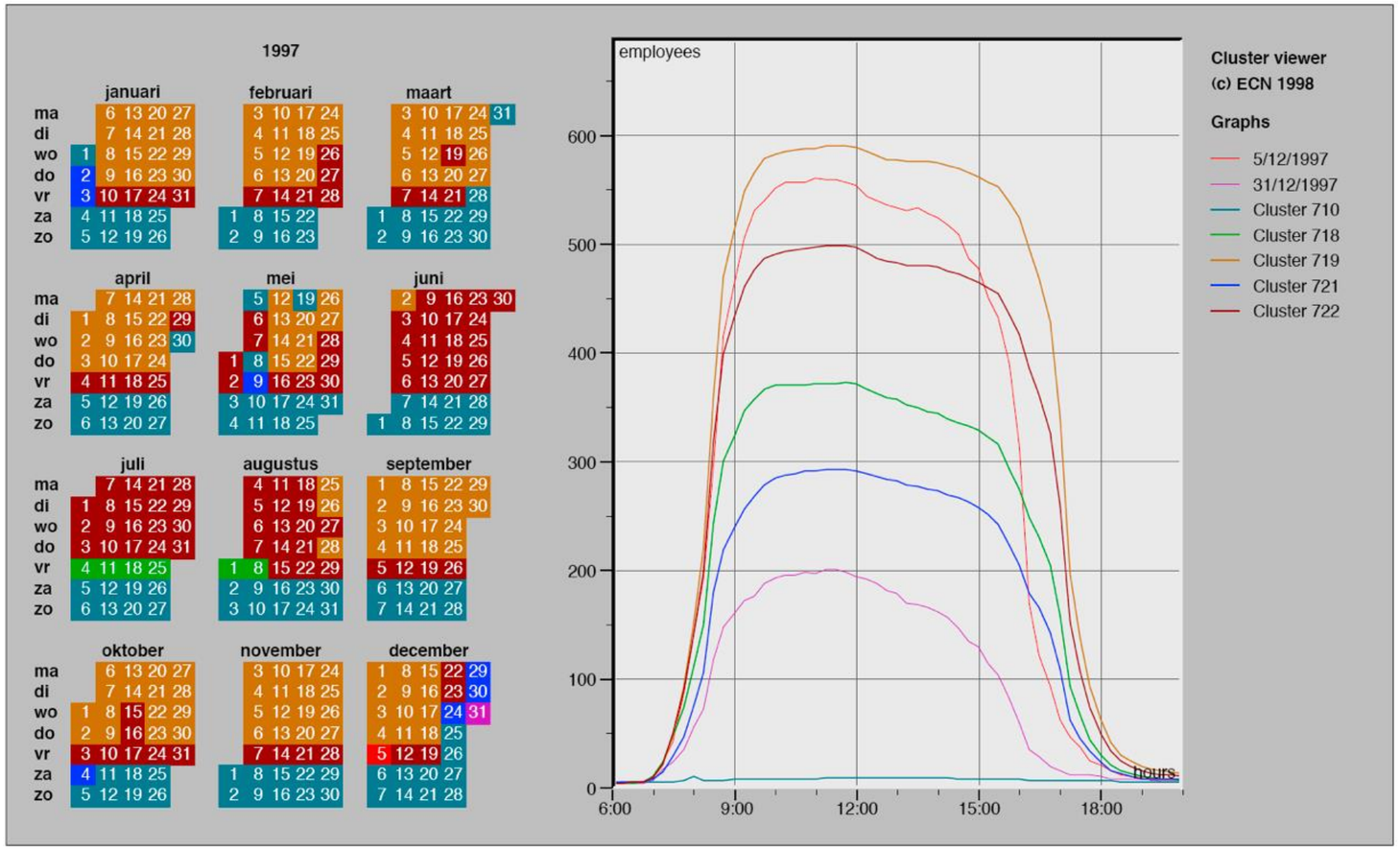
- extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

No unjustified 3D example: Transform for new data abstraction

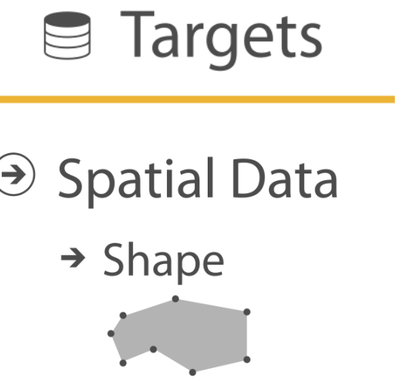
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

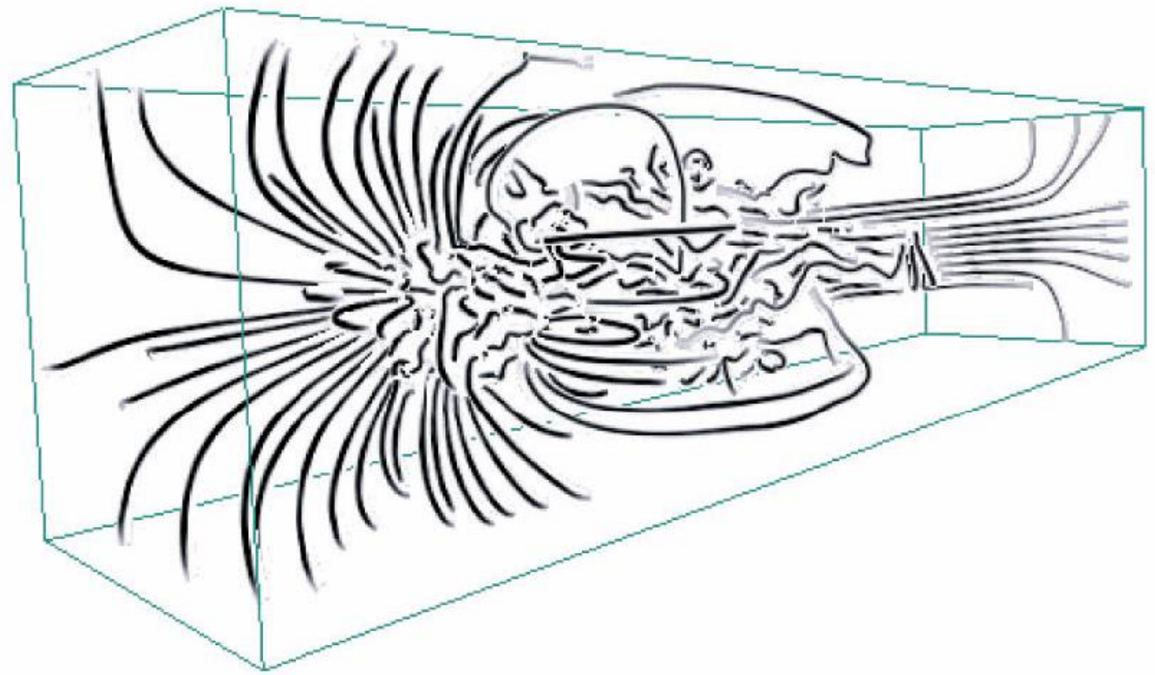


[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Justified 3D: shape perception

- benefits outweigh costs when task is shape perception for 3D spatial data
  - interactive navigation supports synthesis across many viewpoints

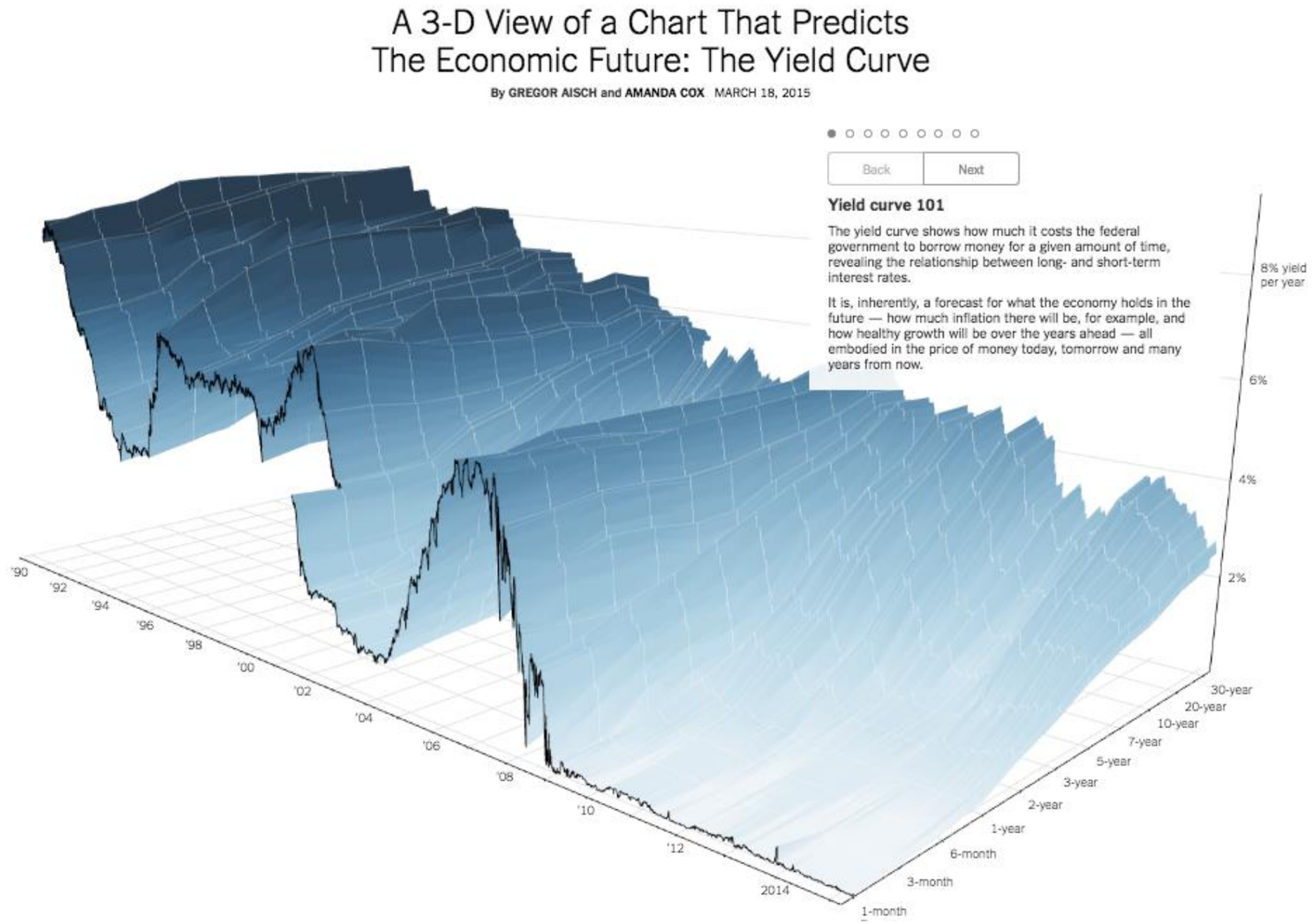




[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630–640.]

Justified 3D: Economic growth curve

- constrained navigation steps through carefully designed viewpoints



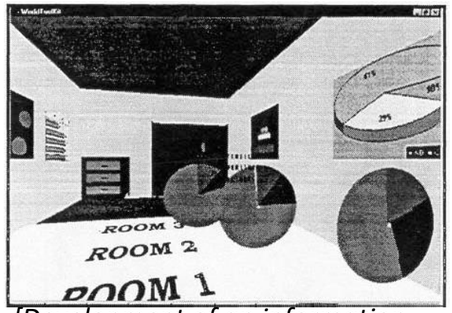
http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

Eyes beat memory

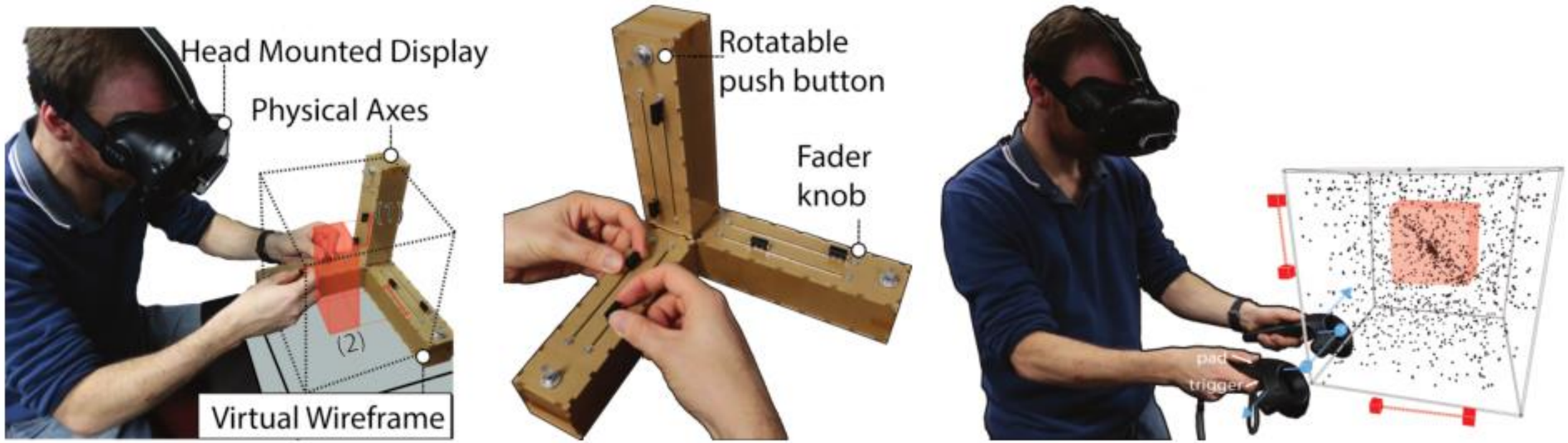
- principle: external cognition vs. internal memory
  - easy to compare by moving eyes between side-by-side views
  - harder to compare visible item to memory of what you saw
- implications for animation
  - great for choreographed storytelling
  - great for transitions between two states
  - poor for many states with changes everywhere
    - consider small multiples instead

Resolution beats immersion

- immersion typically not helpful for abstract data
  - do not need sense of presence or stereoscopic 3D
  - desktop also better for workflow integration
- resolution much more important: pixels are the scarcest resource
- first wave: virtual reality for abstract data difficult to justify
- second wave: AR/MR (augmented/mixed reality) has more promise



[Development of an information visualization tool using virtual reality. Kirn and Martins. Proc. Symp. Applied Computi 2000]



[A Design Space for Spatio-Data Coordination: Tangible Interaction Devices for Immersive Information Visualisation. Cordeil, Bach, Li, Elliott, and Dwyer. Proc. PacificVis 2017 Notes.]

Rule of thumb: **Responsiveness is required**

- *visual feedback: three rough categories*
  - *0.1 seconds: perceptual processing*
    - subsecond response for mouseover highlighting - ballistic motion
  - *1 second: immediate response*
    - fast response after mouseclick, button press - Fitts' Law limits on motor control
  - *10 seconds: brief tasks*
    - bounded response after dialog box - mental model of heavyweight operation (file load)
- scalability considerations
  - highlight selection without complete redraw of view (graphics frontbuffer)
  - show hourglass for multi-second operations (check for cancel/undo)
  - show progress bar for long operations (process in background thread)
  - rendering speed when item count is large (guaranteed frame rate)