

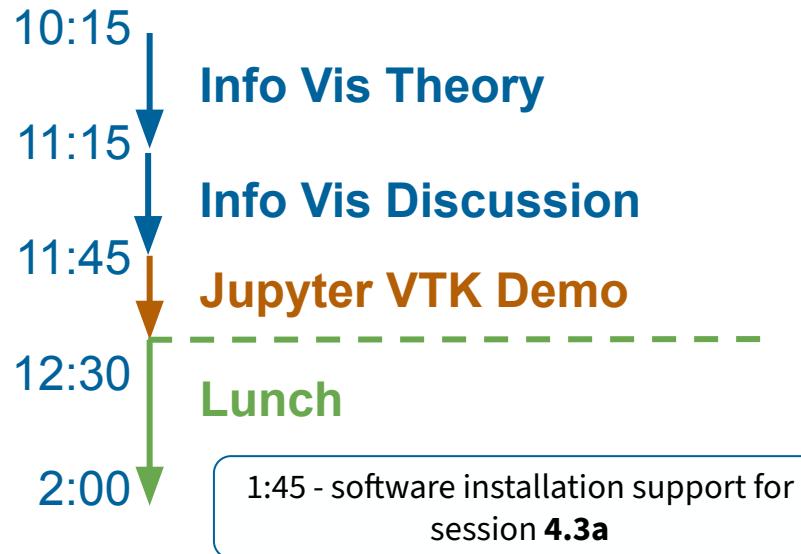
SDSC Summer Institute 2024

4.2b Information Visualization Concepts

Isaac Nealey

computer science and engineering @ UCSD

Agenda:



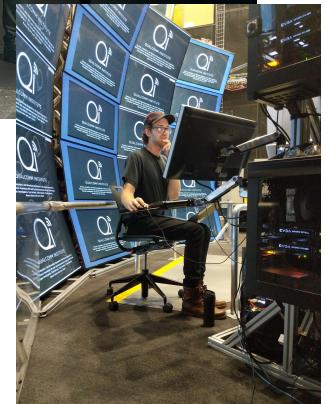
My Background



Prescribed Fire Visualization @ WIFIRE Lab



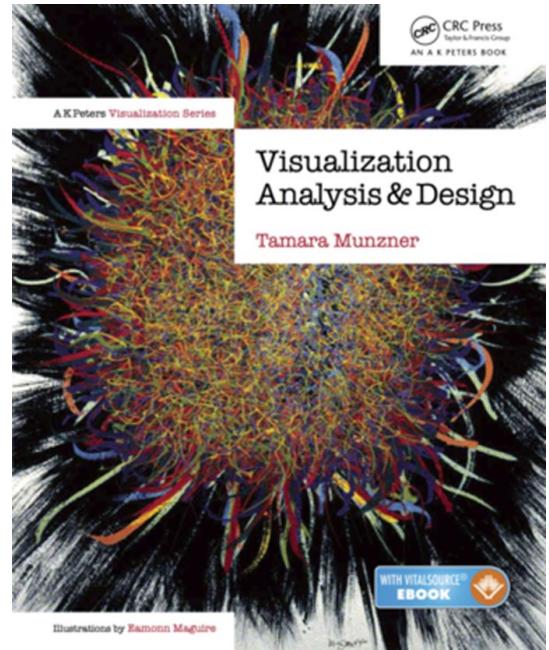
big screens, big dreams



Info Vis Concepts

Resources

- *Visualization Analysis and Design*
 - Tamara Munzner
- *The Visual Display of Quantitative Information*
 - Edward Tufte



Why do data visualization?

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.

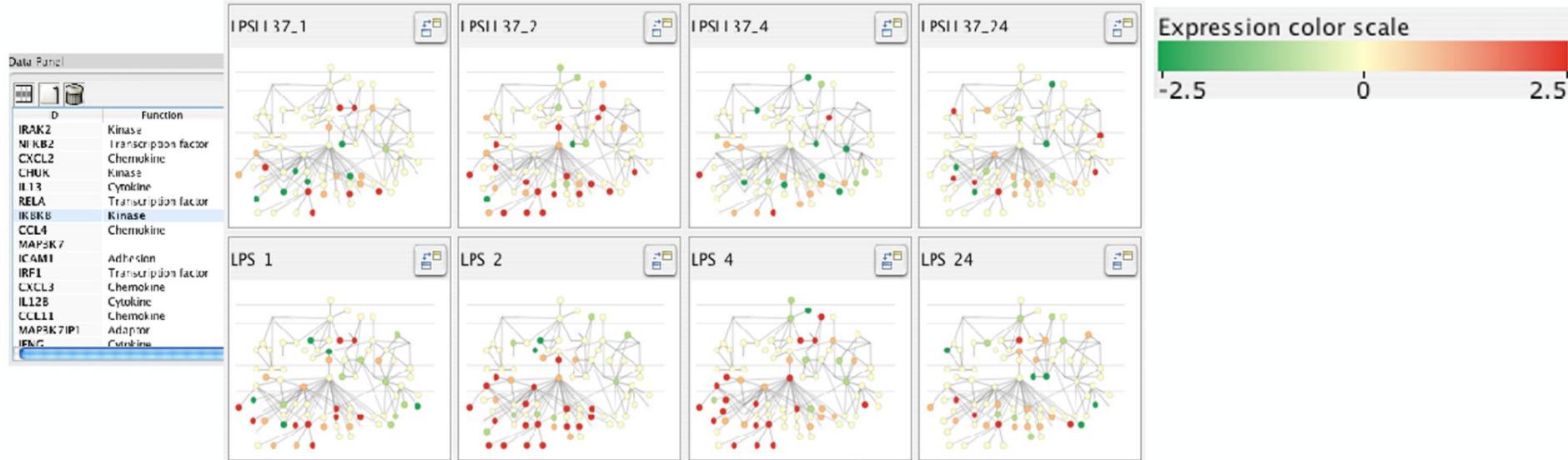
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- 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 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, Gandy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why depend on vision?

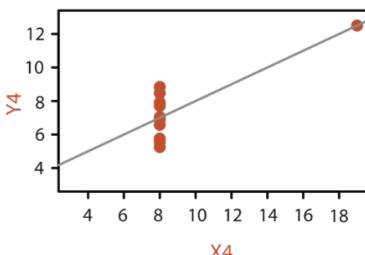
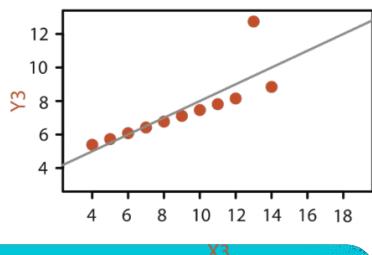
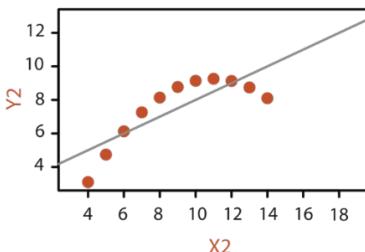
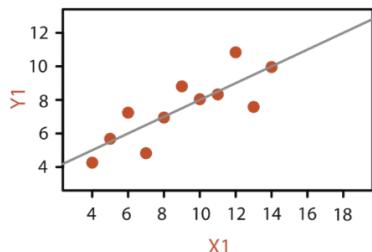
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- 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

What resource limitations are we faced with?

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

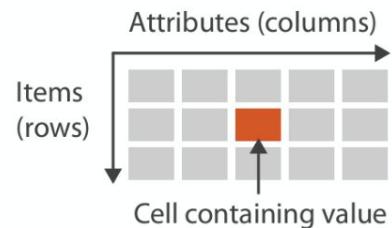
- computational limits
 - computation time, system memory
- display limits
 - pixels are precious & most constrained resource
 - **information density:** ratio of space used to encode info vs unused whitespace
 - tradeoff between clutter and wasting space
 - find sweet spot between dense and sparse
-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

Data Abstraction

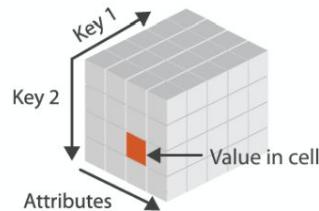
Three major datatypes

→ Dataset Types

→ Tables



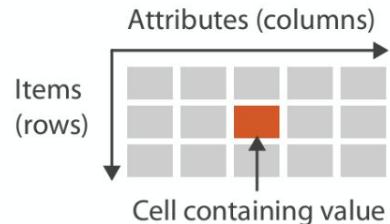
→ Multidimensional Table



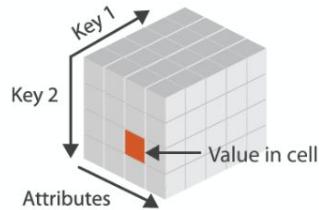
Three major datatypes

→ Dataset Types

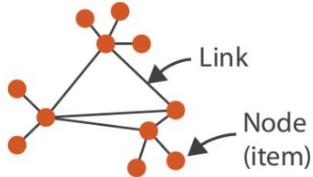
→ Tables



→ Multidimensional Table



→ Networks



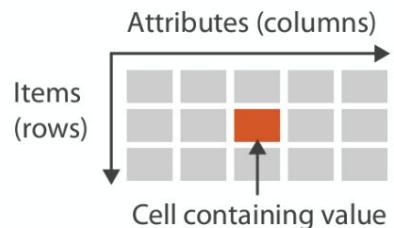
→ Trees



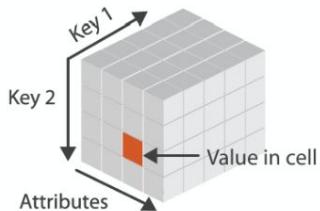
Three major datatypes

→ Dataset Types

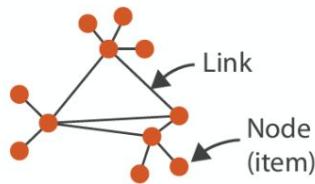
→ Tables



→ Multidimensional Table



→ Networks

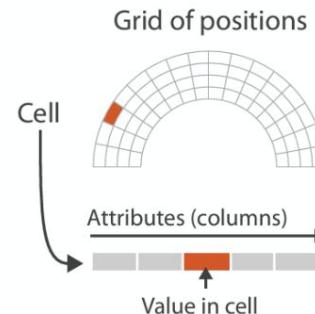


→ Trees



→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)



Now what?

- semantics: real-world meaning
- data types: structural or mathematical interpretation of data
 - item, link, attribute, position, (grid)
 - different from data types in programming!

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"
- 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

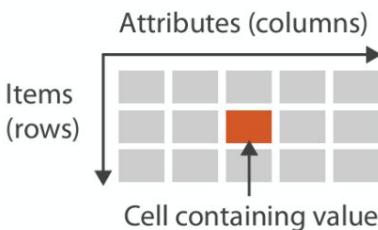
Dataset types

Tables

- flat table

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

→ Tables



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

Dataset types

Tables

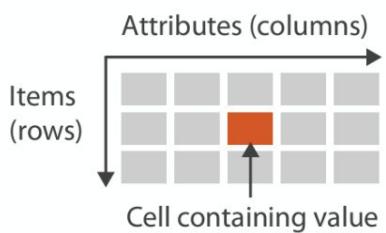
Items

Attributes

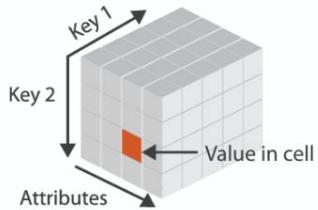
- multidimensional tables
 - indexing based on multiple keys
 - eg genes, patients

A	B	C	D	E	F
1	A	B	C	D	E
2	1	1	1	1	1
3	2	1	1	1	1
4	4	3	2	1	1
5	6	5	4	3	2
6	15	13	12	11	10
7	15	13	12	11	10
8	7	6	5	4	3
9	8	7	6	5	4
10	9	8	7	6	5
11	10	9	8	7	6
12	11	10	9	8	7
13	12	11	10	9	8
14	13	12	11	10	9
15	14	13	12	11	10
16	15	14	13	12	11
17	16	15	14	13	12
18	17	16	15	14	13
19	18	17	16	15	14
20	19	18	17	16	15
21	20	19	18	17	16
22	21	20	19	18	17
23	22	21	20	19	18
24	23	22	21	20	19
25	24	23	22	21	20
26	25	24	23	22	21
27	26	25	24	23	22
28	27	26	25	24	23
29	28	27	26	25	24
30	29	28	27	26	25
31	30	29	28	27	26
32	31	30	29	28	27
33	32	31	30	29	28
34	33	32	31	30	29
35	34	33	32	31	30
36	35	34	33	32	31
37	36	35	34	33	32
38	37	36	35	34	33
39	38	37	36	35	34
40	39	38	37	36	35
41	40	39	38	37	36
42	41	40	39	38	37
43	42	41	40	39	38
44	43	42	41	40	39
45	44	43	42	41	40
46	45	44	43	42	41
47	46	45	44	43	42
48	47	46	45	44	43
49	48	47	46	45	44
50	49	48	47	46	45
51	50	49	48	47	46
52	51	50	49	48	47
53	52	51	50	49	48
54	53	52	51	50	49
55	54	53	52	51	50
56	55	54	53	52	51
57	56	55	54	53	52
58	57	56	55	54	53
59	58	57	56	55	54
60	59	58	57	56	55
61	60	59	58	57	56
62	61	60	59	58	57
63	62	61	60	59	58
64	63	62	61	60	59
65	64	63	62	61	60
66	65	64	63	62	61
67	66	65	64	63	62
68	67	66	65	64	63
69	68	67	66	65	64
70	69	68	67	66	65
71	70	69	68	67	66
72	71	70	69	68	67
73	72	71	70	69	68
74	73	72	71	70	69
75	74	73	72	71	70
76	75	74	73	72	71
77	76	75	74	73	72
78	77	76	75	74	73
79	78	77	76	75	74
80	79	78	77	76	75
81	80	79	78	77	76
82	81	80	79	78	77
83	82	81	80	79	78
84	83	82	81	80	79
85	84	83	82	81	80
86	85	84	83	82	81
87	86	85	84	83	82
88	87	86	85	84	83
89	88	87	86	85	84
90	89	88	87	86	85
91	90	89	88	87	86
92	91	90	89	88	87
93	92	91	90	89	88
94	93	92	91	90	89
95	94	93	92	91	90
96	95	94	93	92	91
97	96	95	94	93	92
98	97	96	95	94	93
99	98	97	96	95	94
100	99	98	97	96	95

→ Tables



→ Multidimensional Table



Dataset types

Tables

Items

Attributes

Networks &
Trees

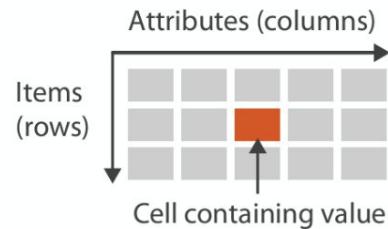
Items (nodes)

Links

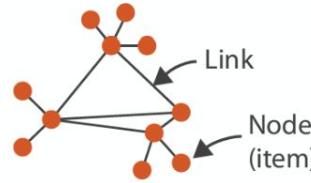
Attributes

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

→ Tables



→ Networks



→ Trees



Dataset types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

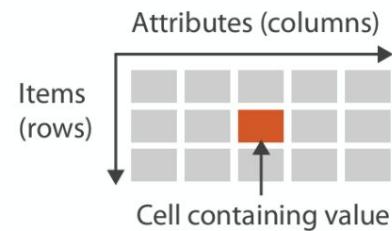
Fields

Grids

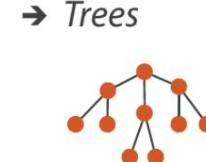
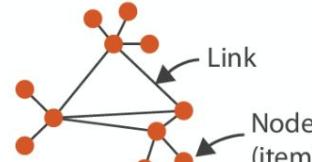
Positions

Attributes

→ Tables

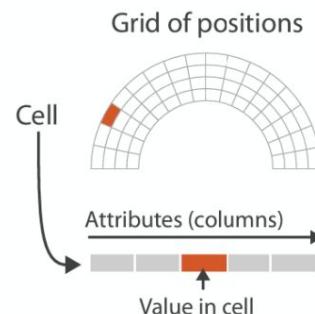


→ Networks



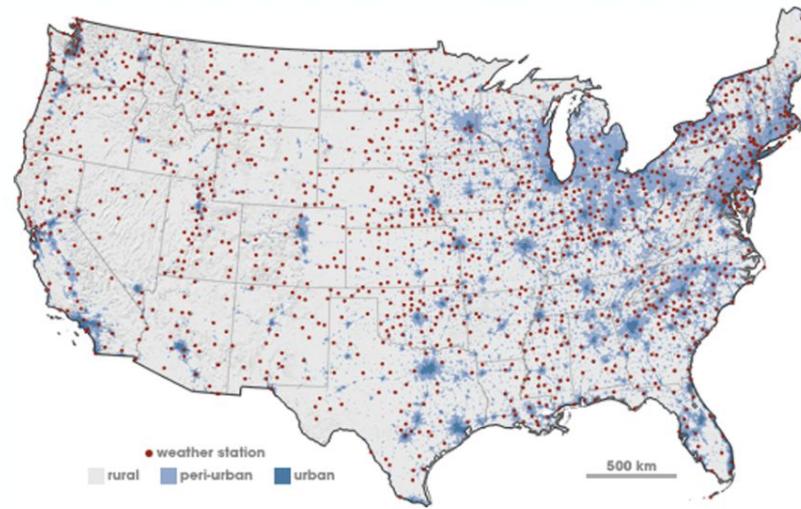
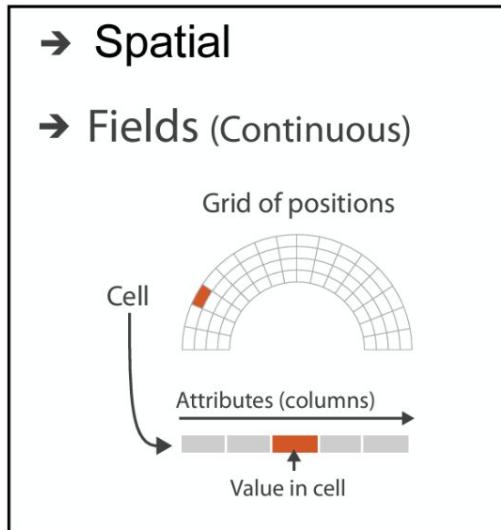
→ Spatial

→ Fields (Continuous)



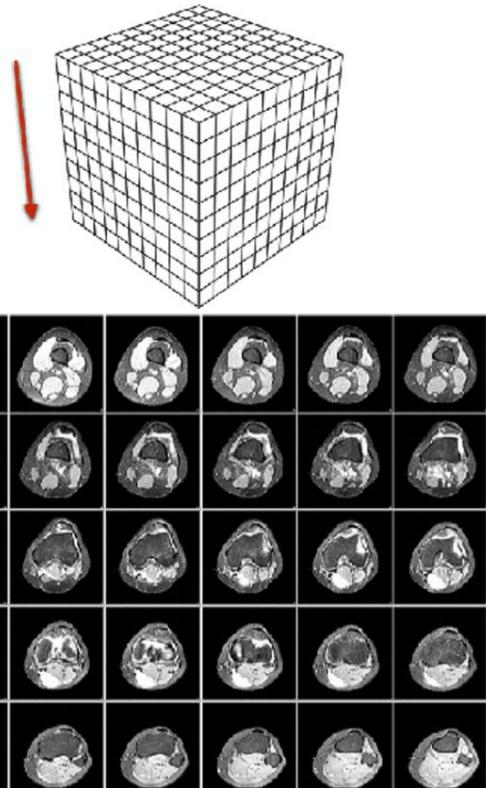
Spatial fields

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



Spatial fields

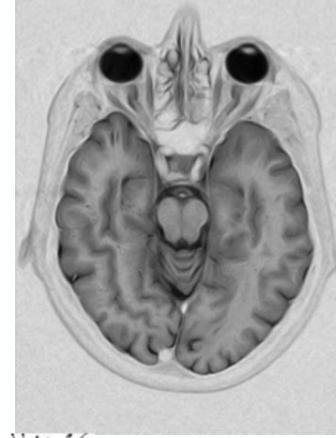
- attribute values associated w/ cells
- cell contains value from continuous domain
 - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
 - sampling:
where attributes are measured
 - interpolation:
how to model attributes elsewhere
 - grid types



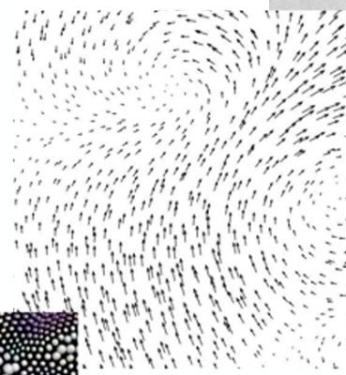
Spatial fields

- 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), tensor (many)

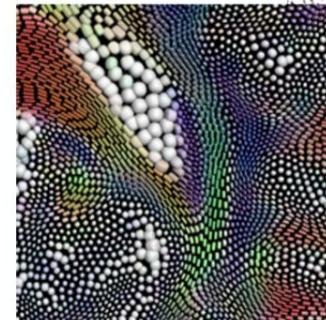
scalar



vector



tensor



Dataset types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Fields

Grids

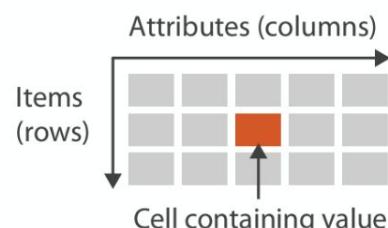
Positions

Geometry

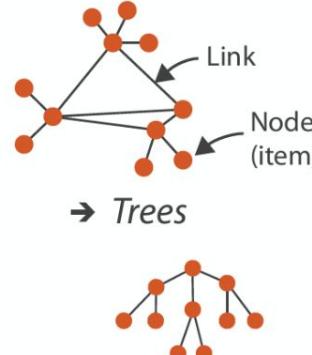
Items

Positions

→ Tables

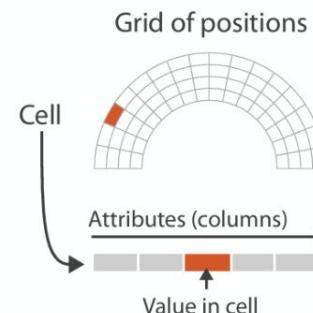


→ Networks



→ Spatial

→ Fields (Continuous)

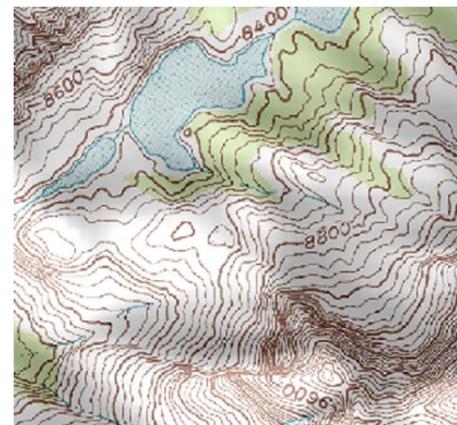


→ Geometry (Spatial)



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

Items

Attributes

Networks & Trees

Items (nodes)

Links

Fields

Grids

Positions

Geometry

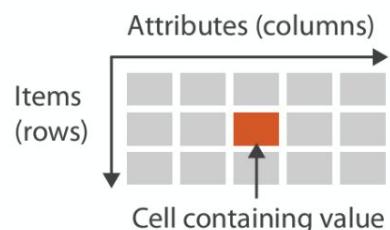
Items

Positions

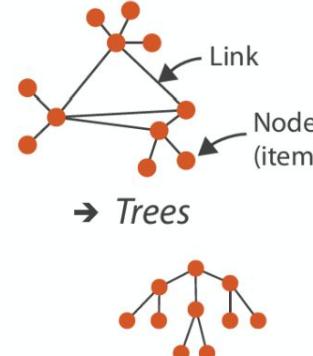
Clusters, Sets, Lists

Items

→ Tables

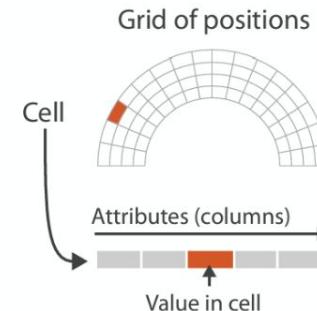


→ Networks



→ Spatial

→ Fields (Continuous)



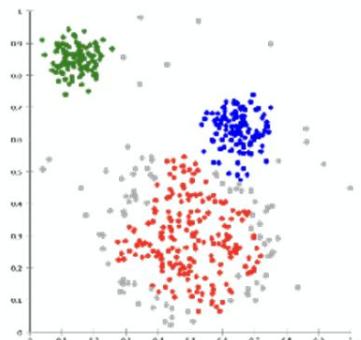
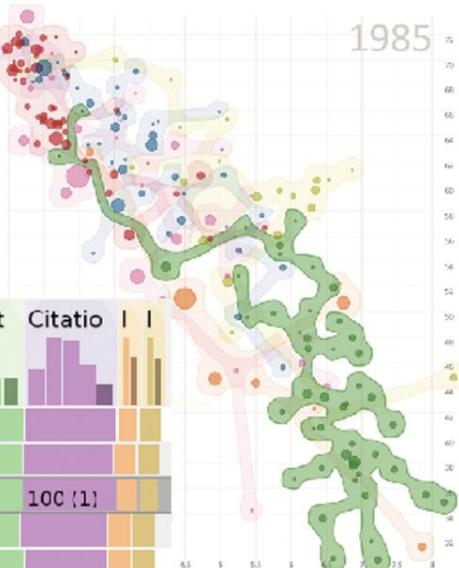
→ Geometry (Spatial)



Collections

- how we group items
- sets
 - unique items, unordered
- lists
 - ordered, duplicates possible
- clusters
 - groups of similar items

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



Attribute types

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

⇒ Attribute Types

→ Categorical



→ Ordered

→ *Ordinal*



→ *Quantitative*



categorical
ordinal
quantitative

A	B	C	D	E	F	G
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*
↑ T-shirt ↑
→ *Quantitative*
— — —

→ Ordering Direction

→ Sequential
— →

→ Diverging
↔ — ↔



→ Dataset Availability

→ Static
A document icon containing several orange dots, representing a static dataset.

→ Dynamic
A series of orange dots forming a path that loops back to the start, with an arrow at the end, representing a dynamic dataset.

Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
 - how many items in the dataset?
 - what is cardinality of each attribute?
 - number of levels for categorical data
 - range for quantitative data
- consider whether to transform data
 - guided by understanding of task

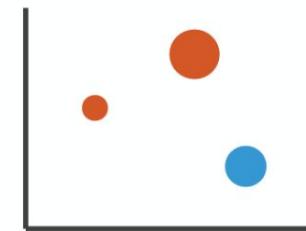
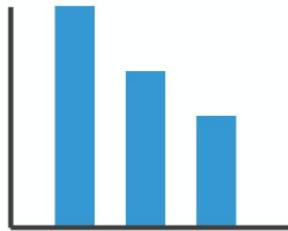
Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
 - how many items in the dataset?
 - what is cardinality of each attribute?
 - number of levels for categorical data
 - range for quantitative data
- consider whether to transform data
 - guided by understanding of task

Marks and Channels

Visual encoding

- how to systematically analyze idiom structure?



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

Channels

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

⇒ Position

→ Horizontal



→ Vertical



→ Both



⇒ Color



⇒ Shape



⇒ Tilt



⇒ Size

→ Length



→ Area

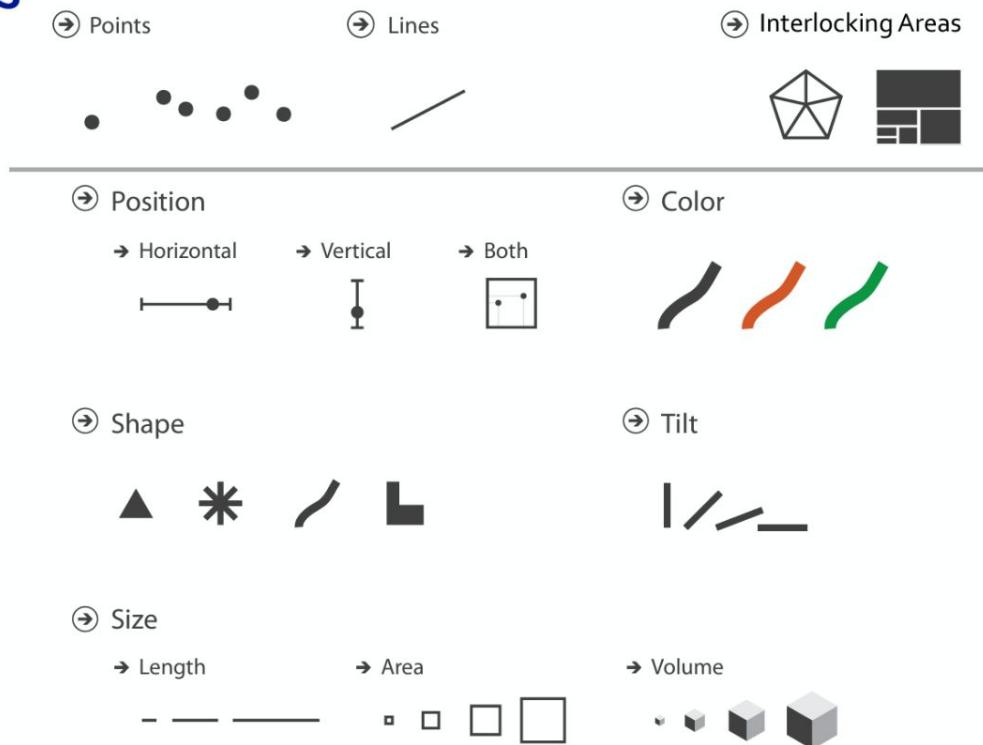


→ Volume



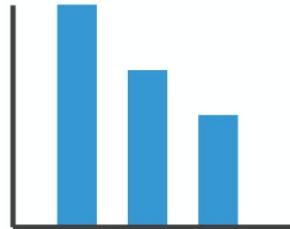
Definitions: Marks and channels

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

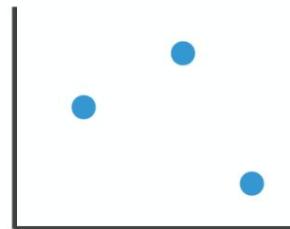


Visual encoding

- analyze idiom structure as combination of marks and channels



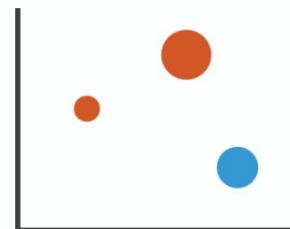
1:
vertical position



2:
vertical position
horizontal position



3:
vertical position
horizontal position
color hue



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

mark: line

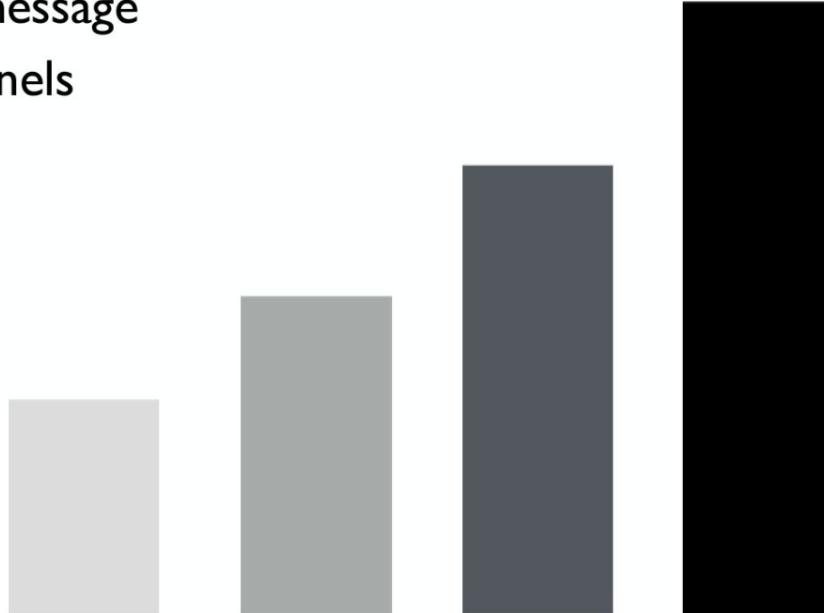
mark: point

mark: point

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



- 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

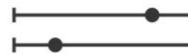
match channel type to data type

effectiveness

some channels are better than others

Channels: Rankings

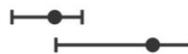
Position on common scale



Spatial region



Position on unaligned scale



Color hue



Length (1D size)



Motion



Tilt/angle



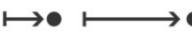
Shape



Area (2D size)



Depth (3D position)



Color luminance



Same

Color saturation



Curvature



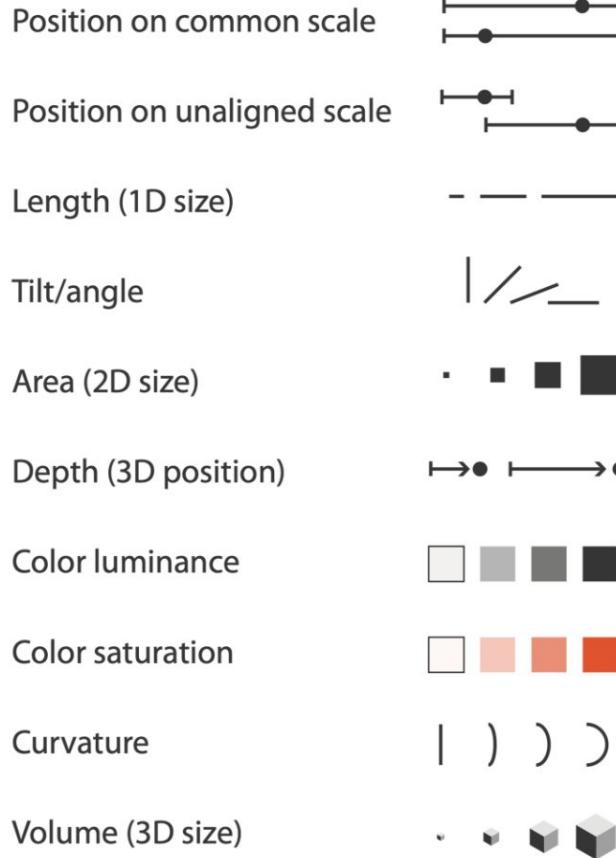
Same

Volume (3D size)



Channels: Rankings

④ Magnitude Channels: Ordered Attributes



⑤ Identity Channels: Categorical Attributes



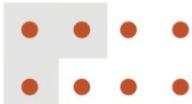
- expressiveness
 - match channel and data characteristics
- effectiveness
 - channels differ in accuracy of perception

Grouping

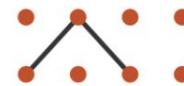
- containment
- connection

Marks as Links

→ Containment



→ Connection



→ Identity Channels: Categorical Attributes

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

Spatial region



Color hue



Motion



Shape

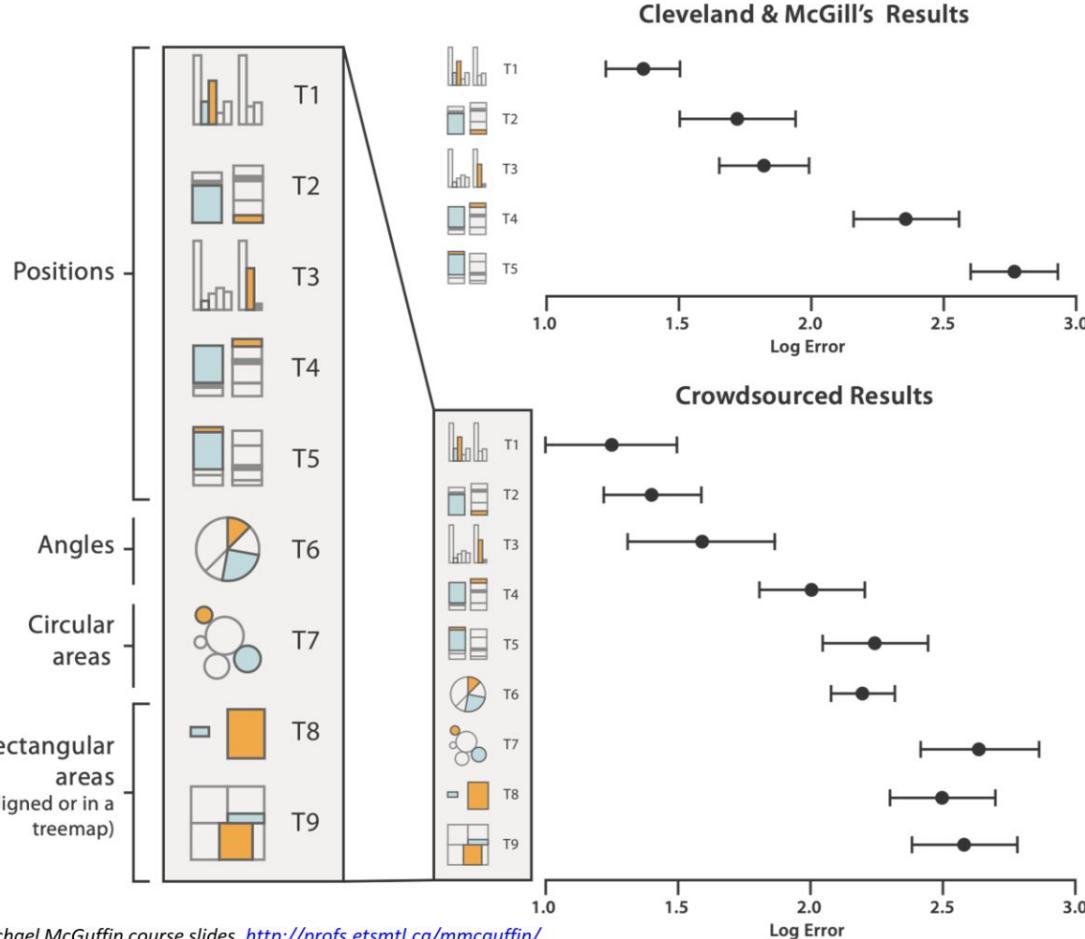


Channel effectiveness

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?

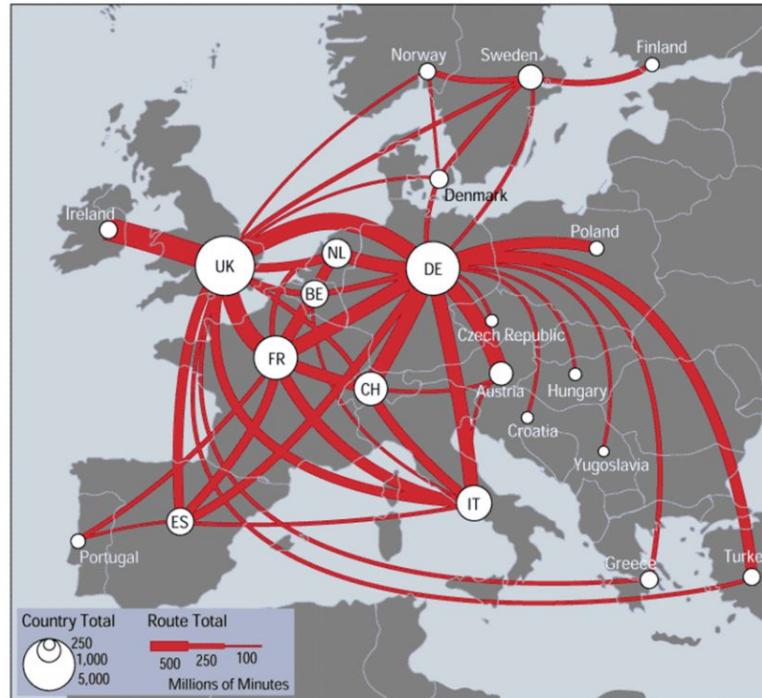
Accuracy: Vis experiments



[*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.]

Discriminability: How many usable steps?

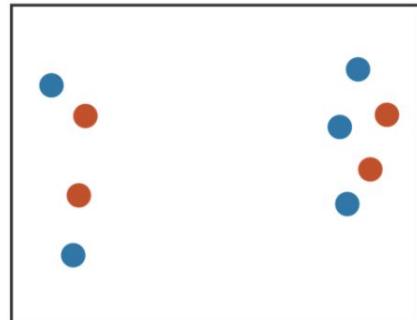
- must be sufficient for number of attribute levels to show
 - linewidth: few bins



[\[mappa.mundi.net/maps/maps_014/telegeography.html\]](http://mappa.mundi.net/maps/maps_014/telegeography.html)

Separability vs. Integrality

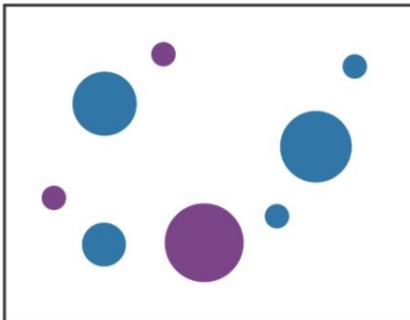
Position
+ Hue (Color)



Fully separable

2 groups each

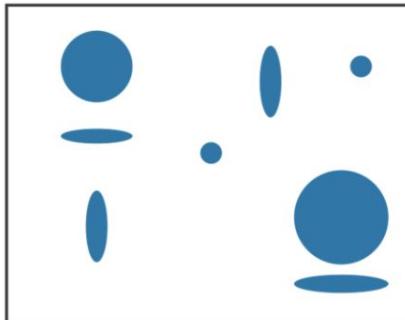
Size
+ Hue (Color)



Some interference

2 groups each

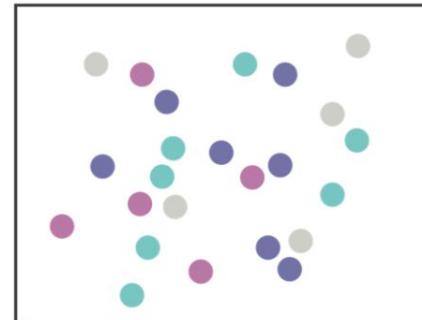
Width
+ Height



Some/significant
interference

3 groups total:
integral area

Red
+ Green



Major interference

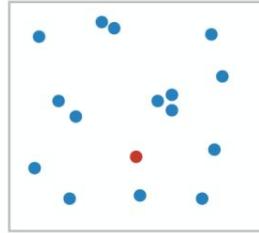
4 groups total:
integral hue

Popout

- find the red dot
 - how long does it take?

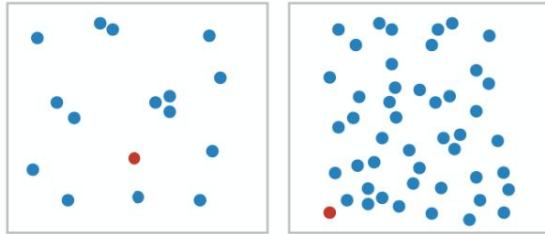
Popout

- find the red dot
 - how long does it take?



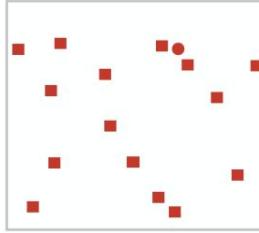
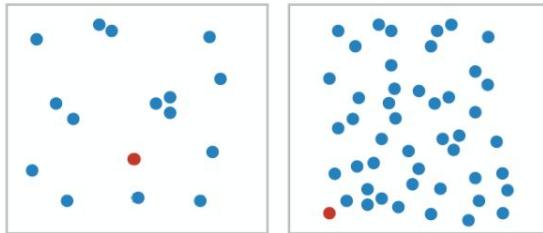
Popout

- find the red dot
 - how long does it take?



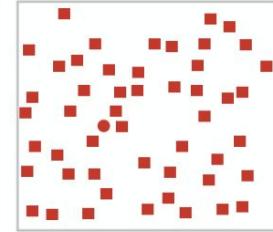
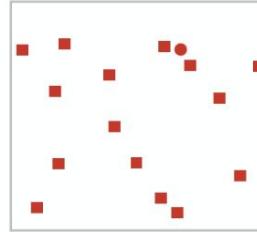
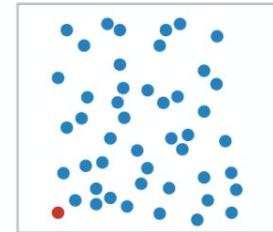
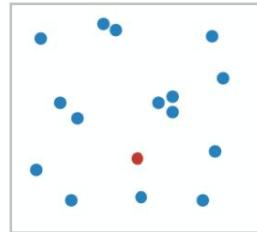
Popout

- find the red dot
 - how long does it take?



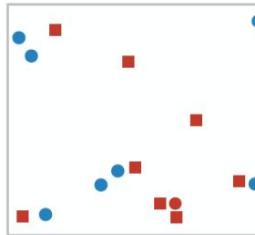
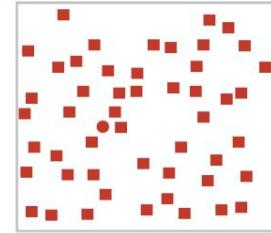
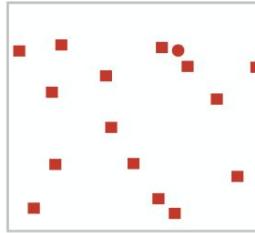
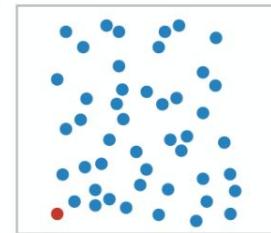
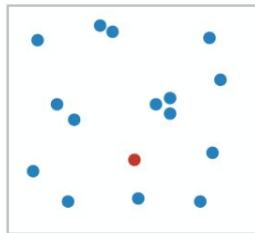
Popout

- find the red dot
 - how long does it take?



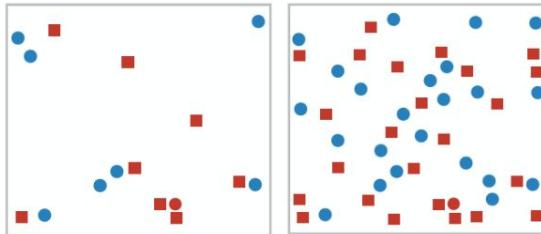
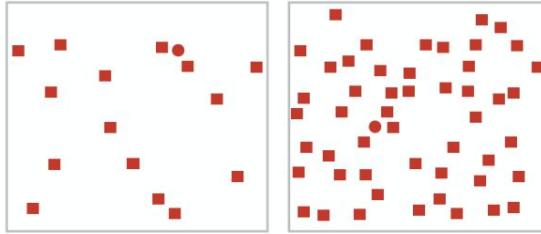
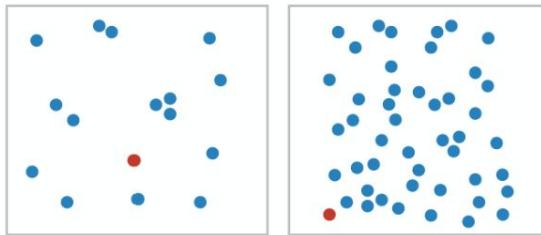
Popout

- find the red dot
 - how long does it take?



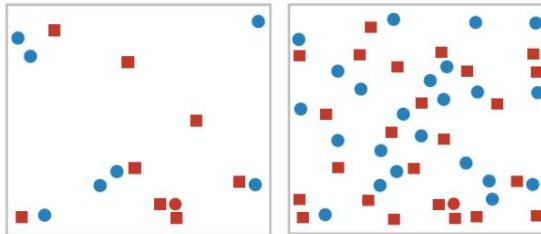
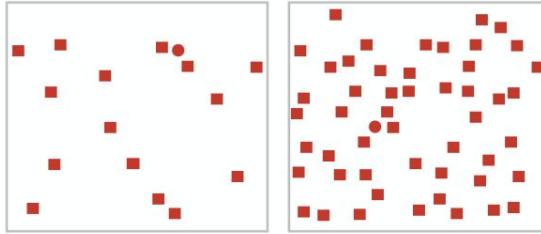
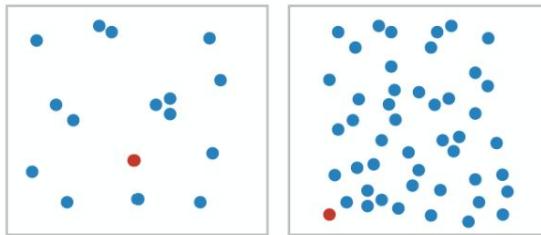
Popout

- find the red dot
 - how long does it take?



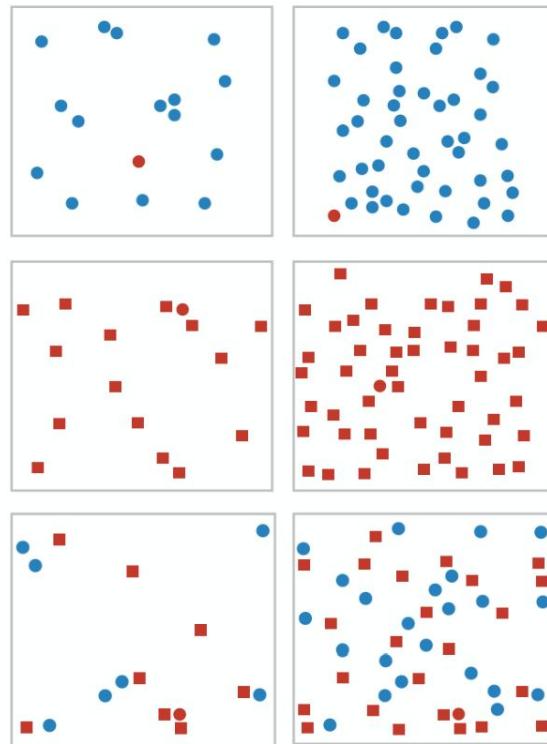
Popout

- find the red dot
 - how long does it take?

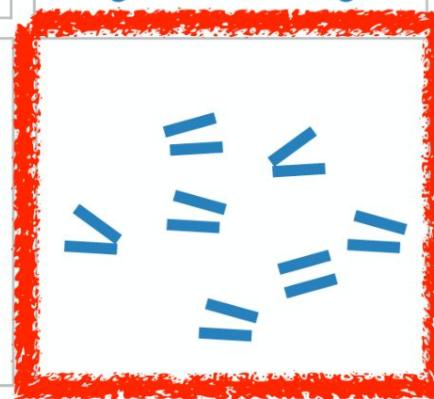
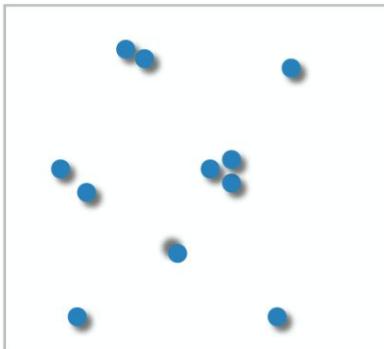
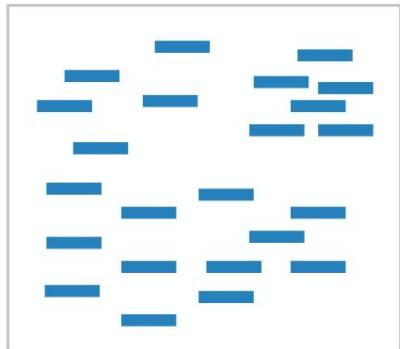
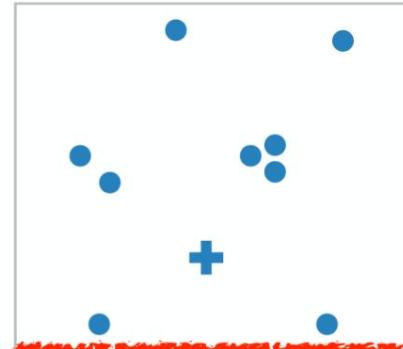
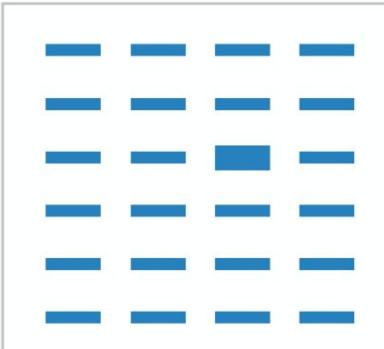
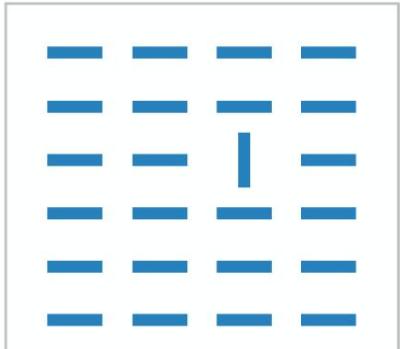


Popout

- find the red dot
 - how long does it take?
- parallel processing on many individual channels
 - speed independent of distractor count
 - speed depends on channel and amount of difference from distractors
- serial search for (almost all) combinations
 - speed depends on number of distractors



Popout



- many channels
 - tilt, size, shape, proximity, shadow direction, ...
- but not all!
 - parallel line pairs do not pop out from tilted pairs

Factors affecting accuracy

- alignment
- distractors
- distance
- common scale / alignment



Relative vs. absolute judgements

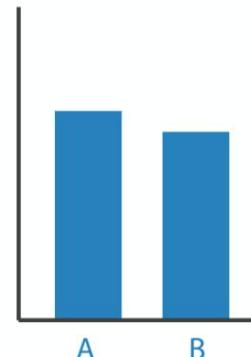
- perceptual system mostly operates with relative judgements, not absolute
 - that's why accuracy increases with common frame/scale and alignment
 - Weber's Law: ratio of increment to background is constant



length



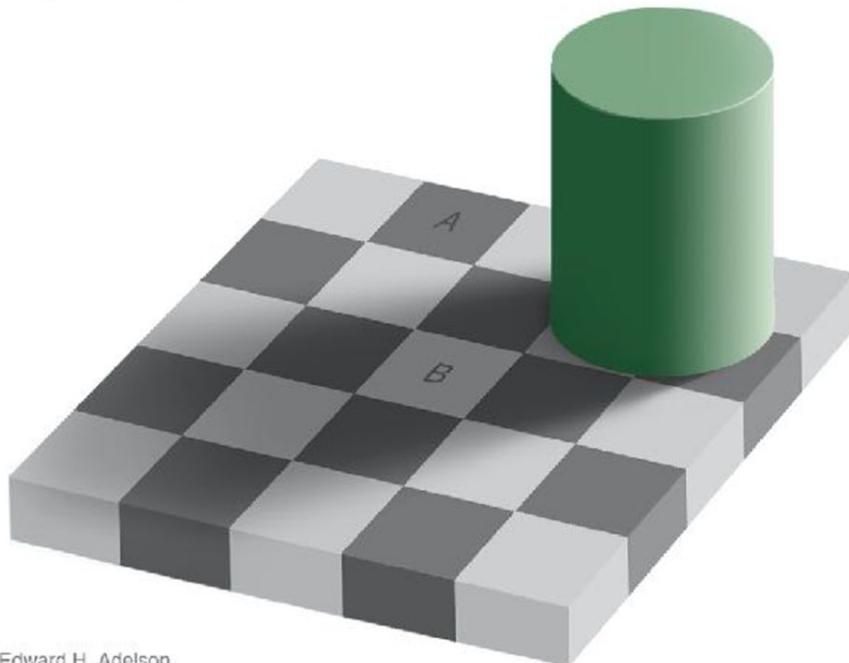
position along
unaligned
common scale



position along
aligned scale

Relative luminance judgements

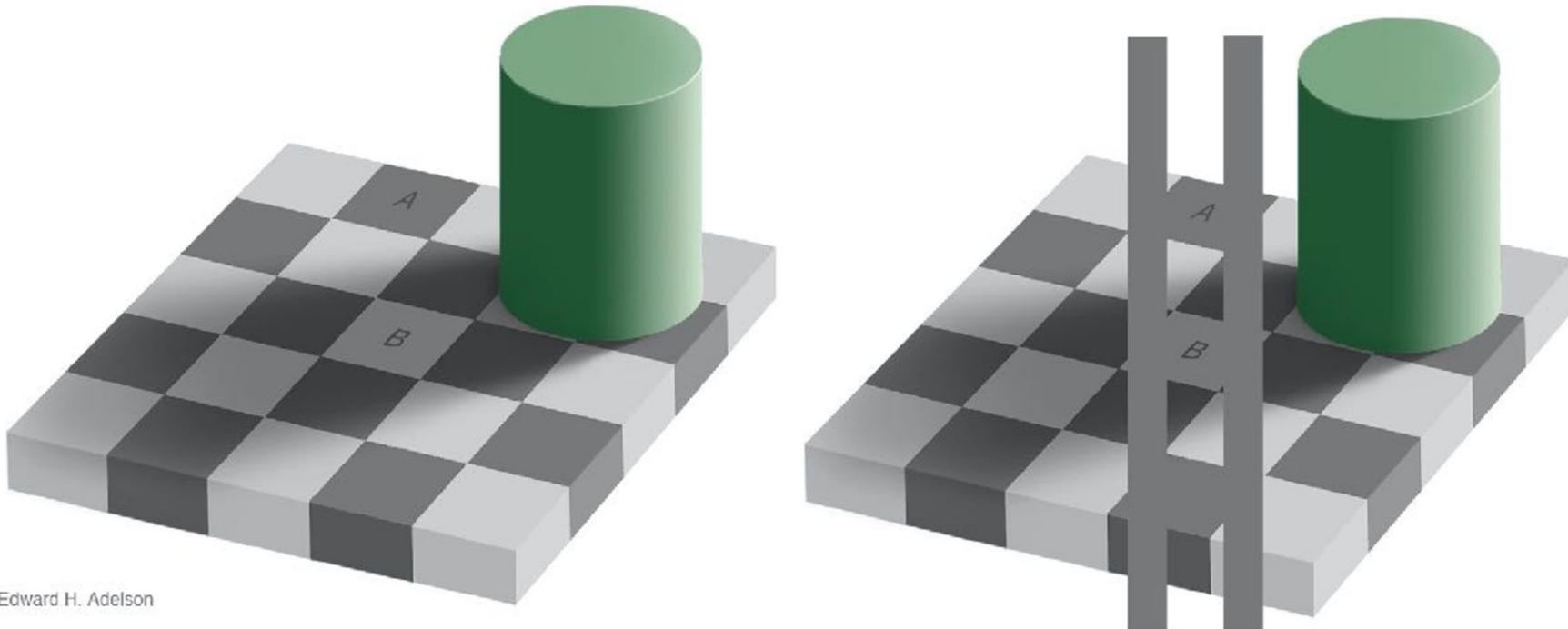
- perception of luminance is contextual based on contrast with surroundings



Edward H. Adelson

Relative luminance judgements

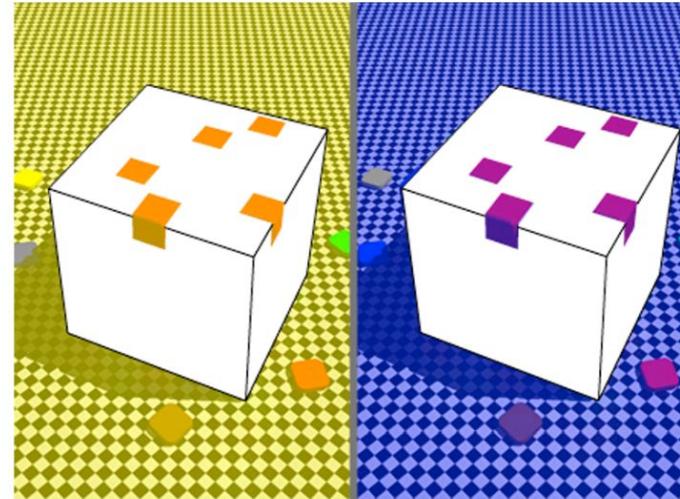
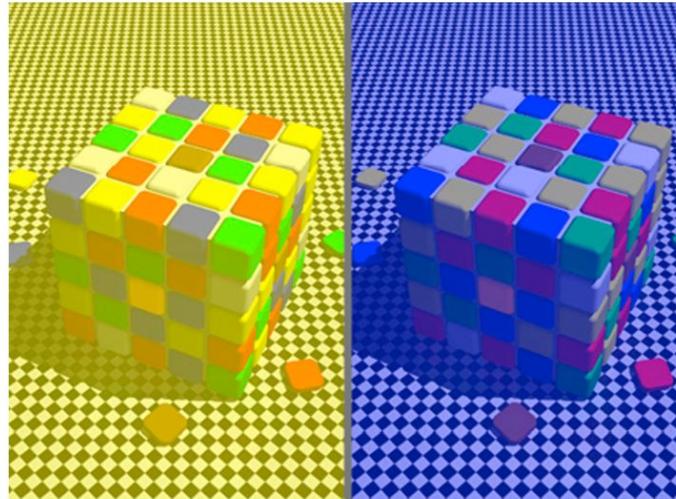
- perception of luminance is contextual based on contrast with surroundings



Edward H. Adelson

Relative color judgements

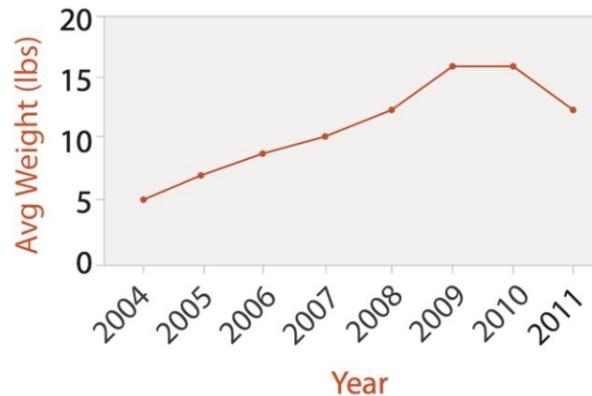
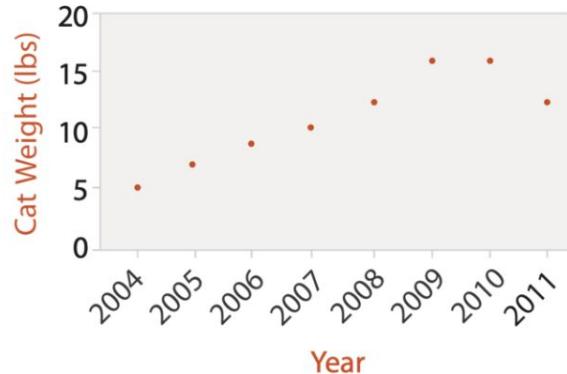
- color constancy across broad range of illumination conditions



Common Info Vis Idioms

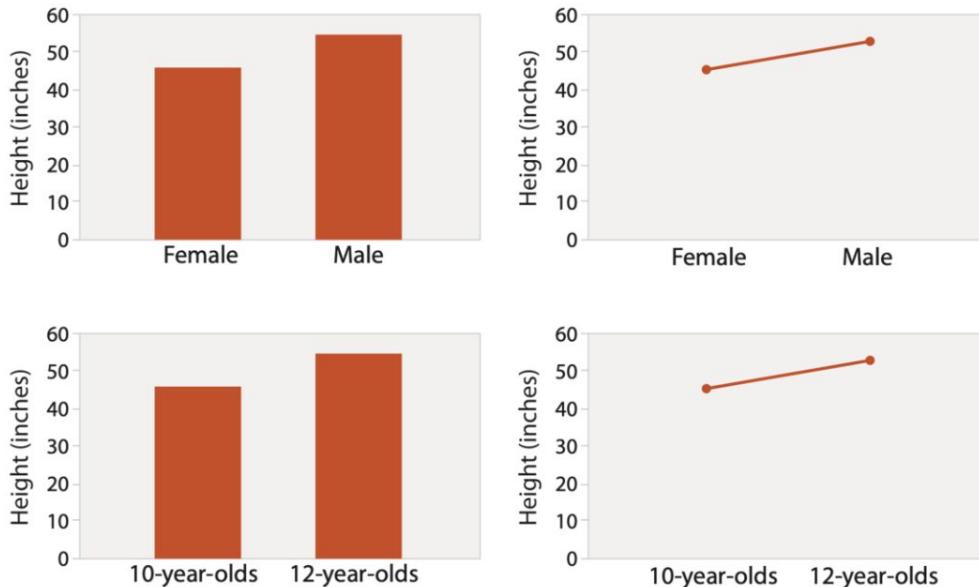
Idiom: dot / line chart

- 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

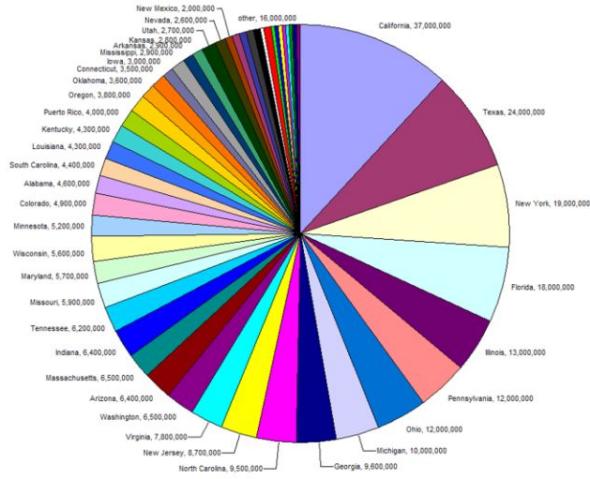
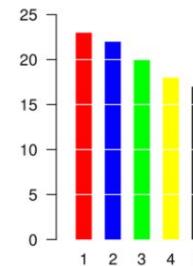
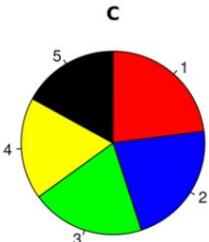
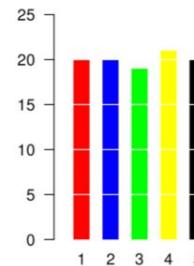
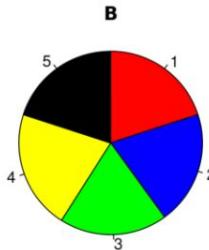
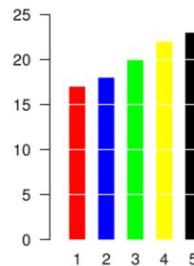
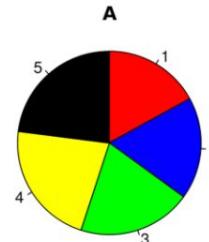
- depends on type of key attrib
 - bar charts if categorical
 - line charts if ordered
- do not use line charts for categorical key attrs
 - violates expressiveness principle
 - implication of trend so strong that it overrides semantics!
 - “The more male a person is, the taller he/she is”



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

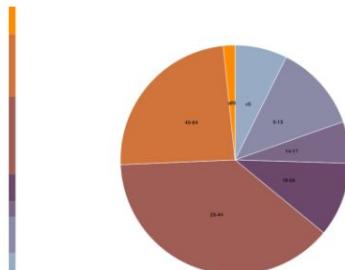
Pie charts: best practices

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

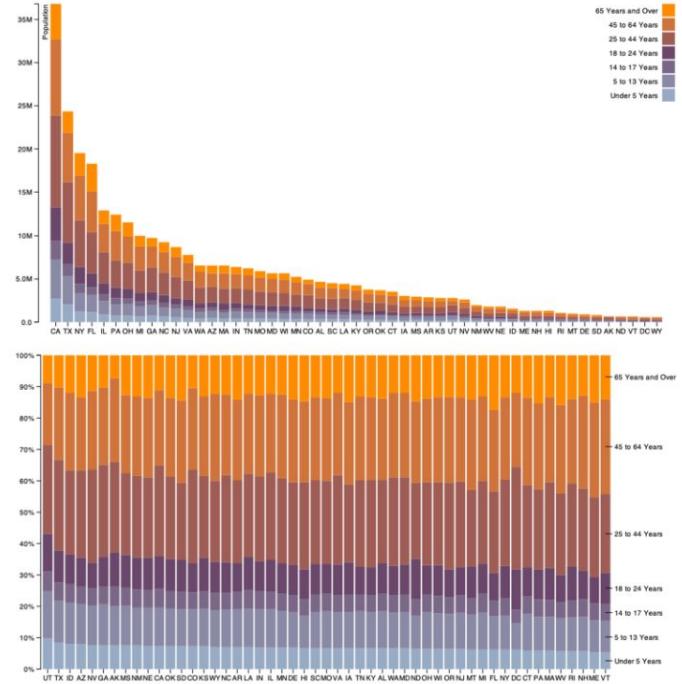


Idioms: normalized stacked bar chart

- 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>

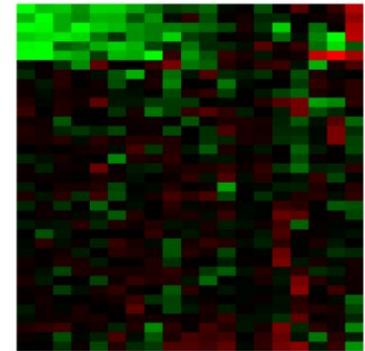
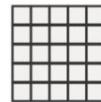


Idiom: heatmap

- 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

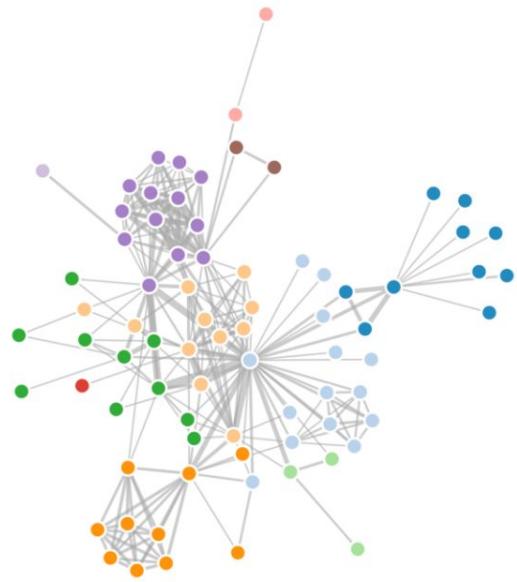
→ 2 Keys

Matrix



Idiom: force-directed placement

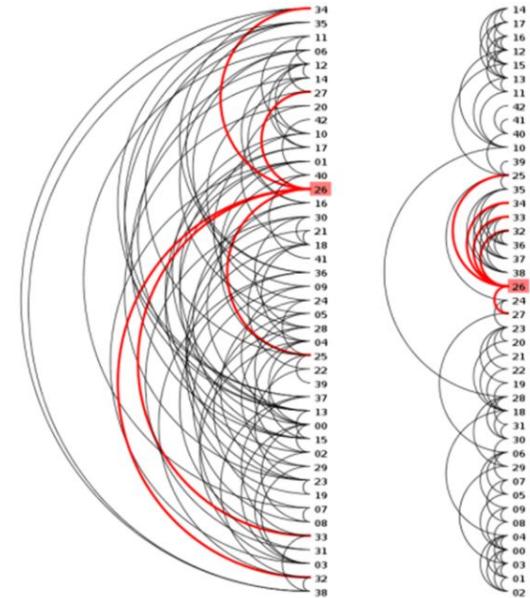
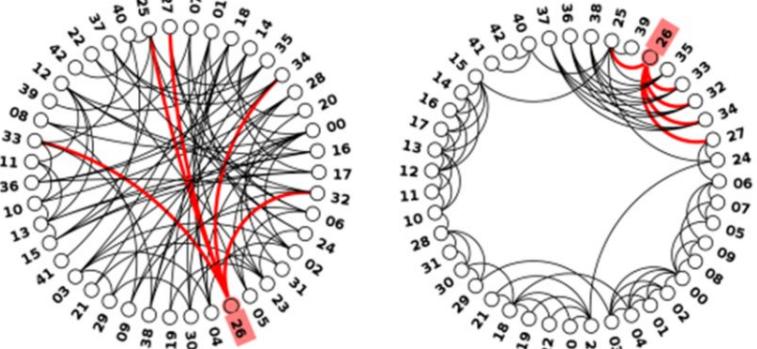
- visual encoding
 - link connection marks, node point marks
- considerations
 - spatial position: no meaning directly encoded
 - left free to minimize crossings
 - proximity semantics?
 - sometimes meaningful
 - sometimes arbitrary, artifact of layout algorithm
 - tension with length
 - long edges more 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

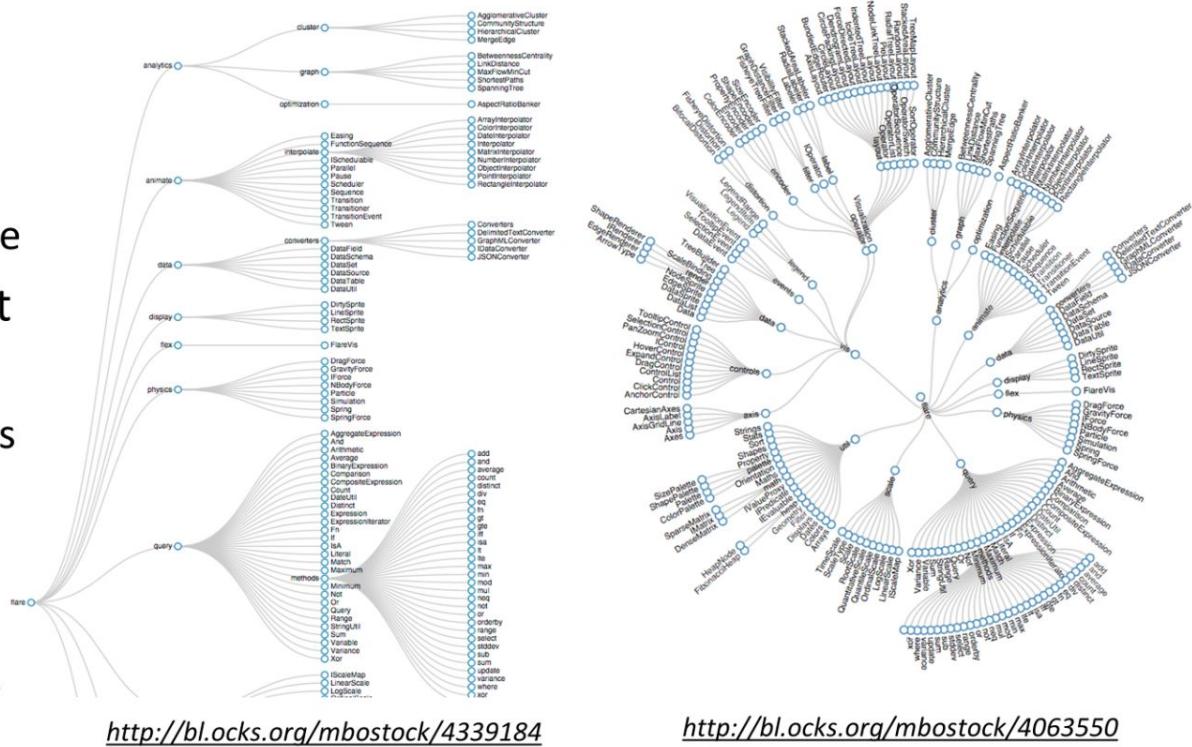


Node-link trees

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

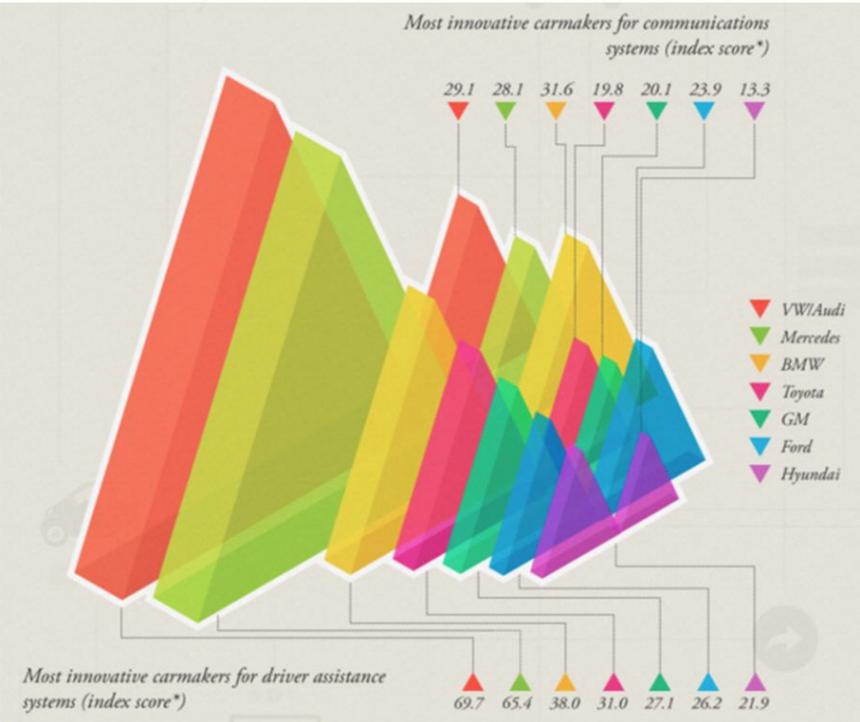
[Tidier drawing of trees. Reingold and Tilford. IEEE Trans. Software Eng., SE-7(2):223–228, 1981.]

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

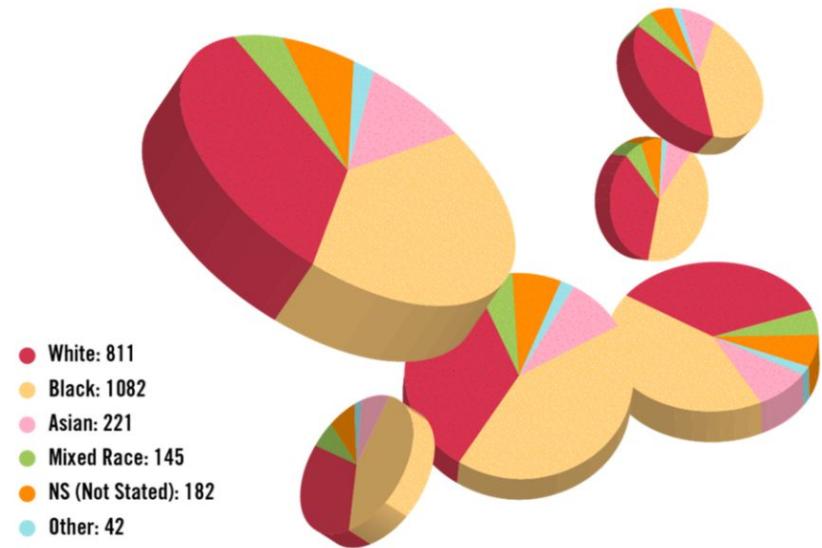


Pitfalls

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



Convictions in London for class A drug supply.

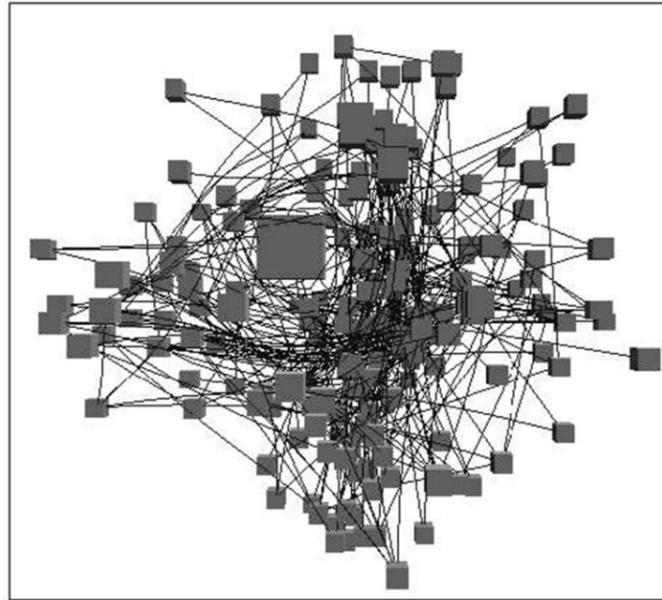


<http://viz.wtf/post/137826497077/eye-popping-3d-triangles>

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

Occlusion hides information

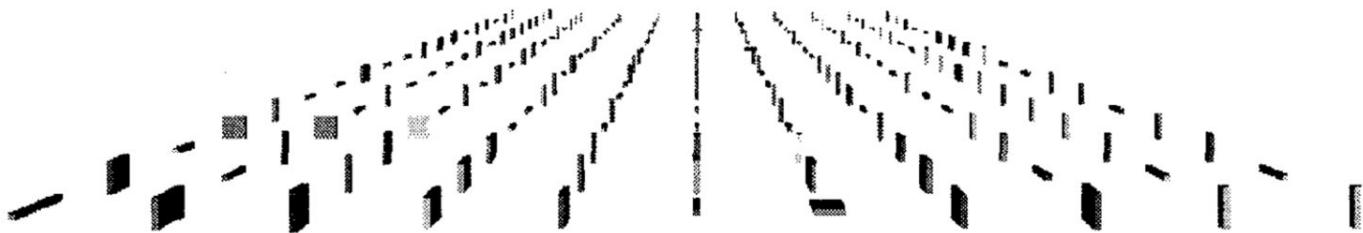
- occlusion
- interaction can resolve, but at cost of time and cognitive load



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

Perspective distortion loses information

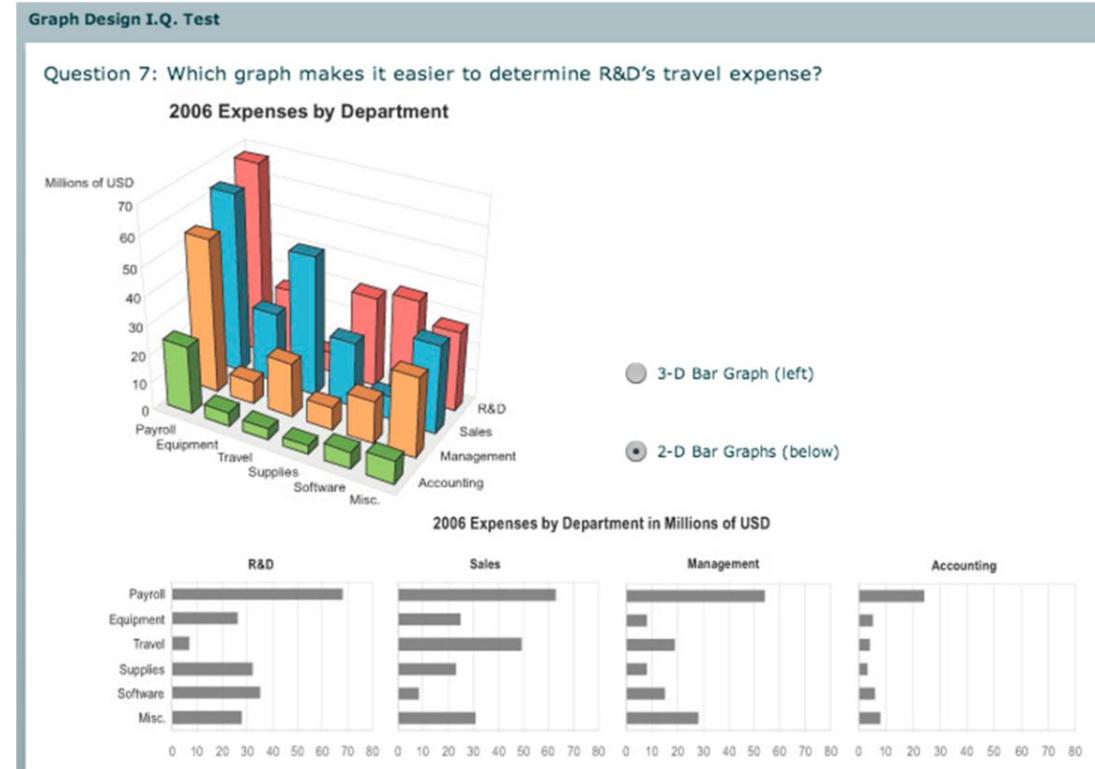
- perspective distortion
 - interferes with all size channel encodings
 - power of the plane is lost!



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

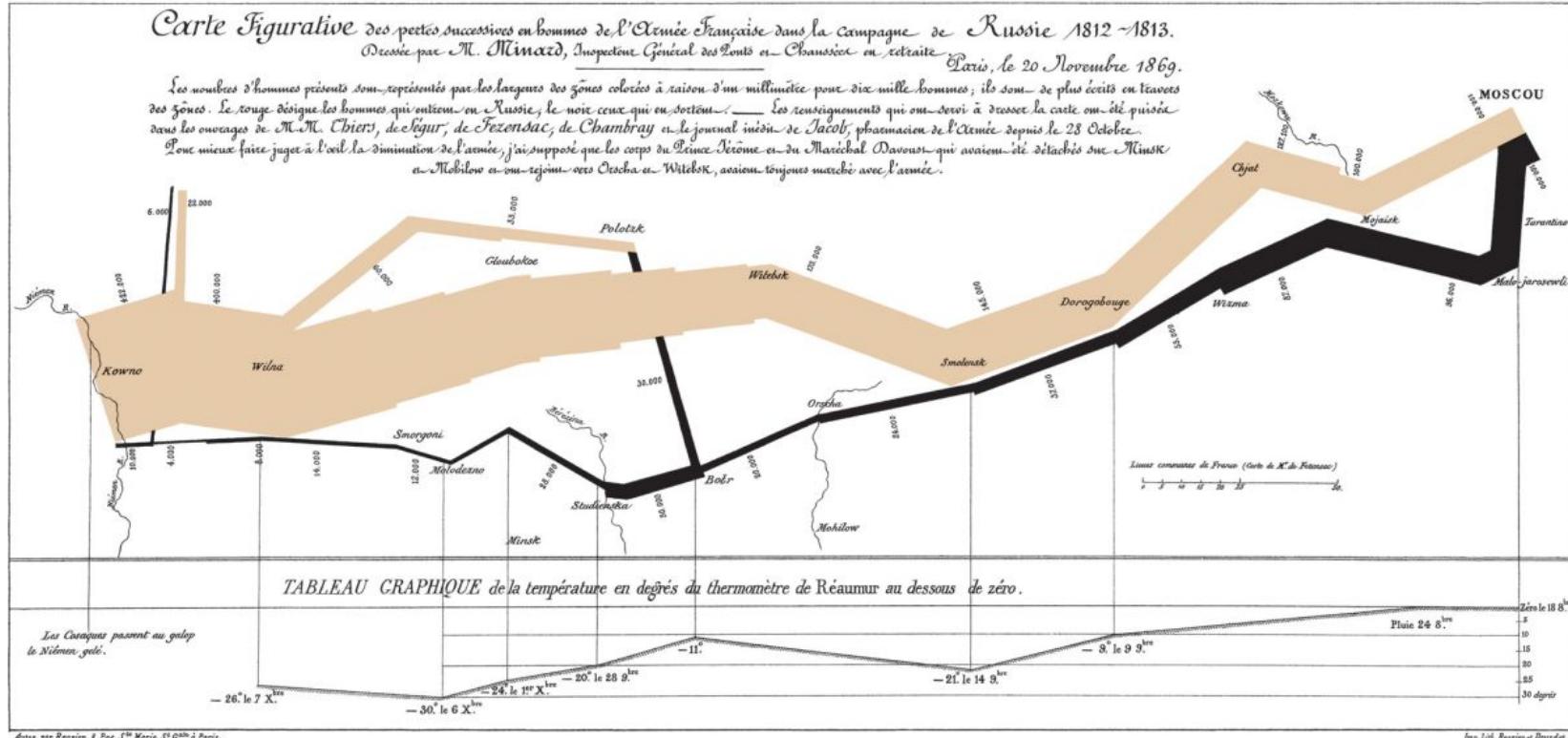
3D vs 2D bar charts

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



Discussion

The Minard Map



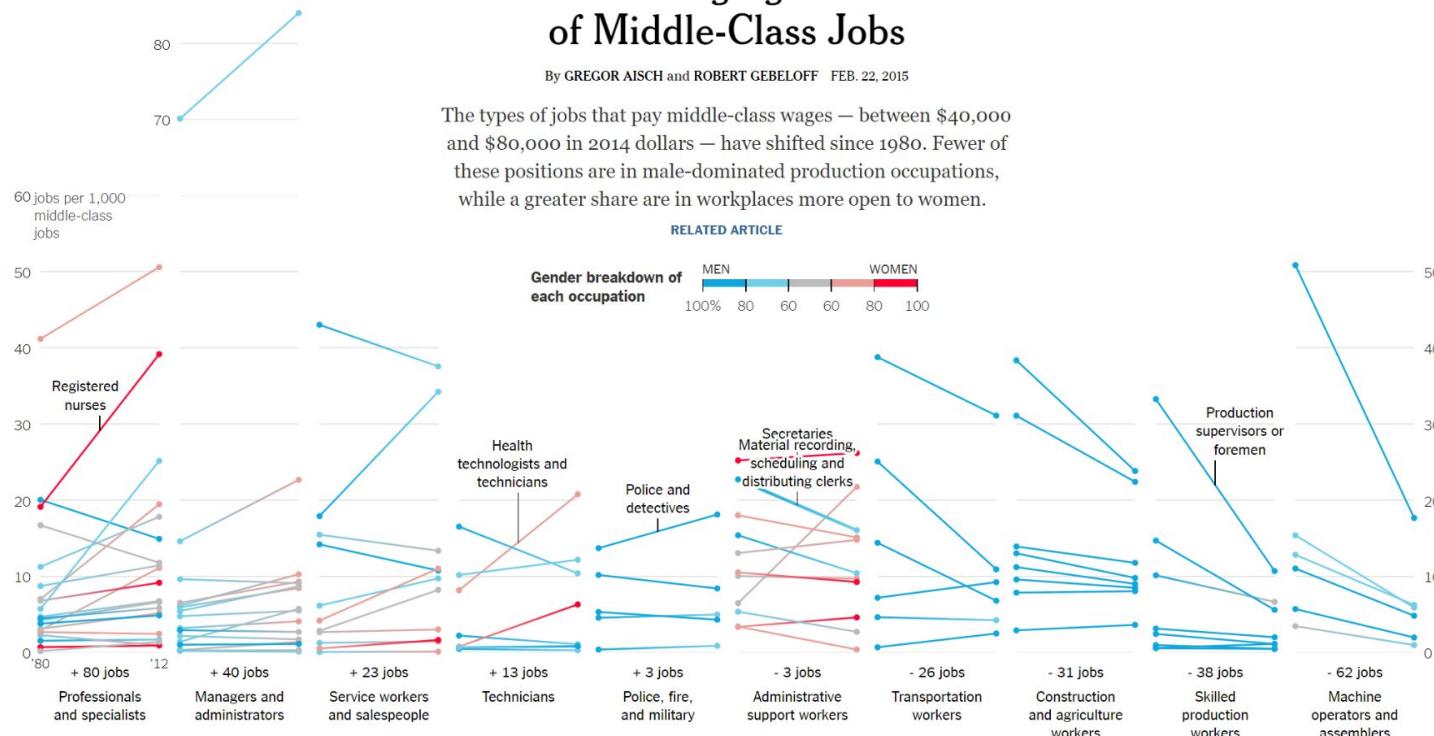
Napoleon's march to Moscow. Depicts geography, army size, march direction, temperature, and time.

The Changing Nature of Middle-Class Jobs

By GREGOR AISCH and ROBERT GEBELOFF FEB. 22, 2015

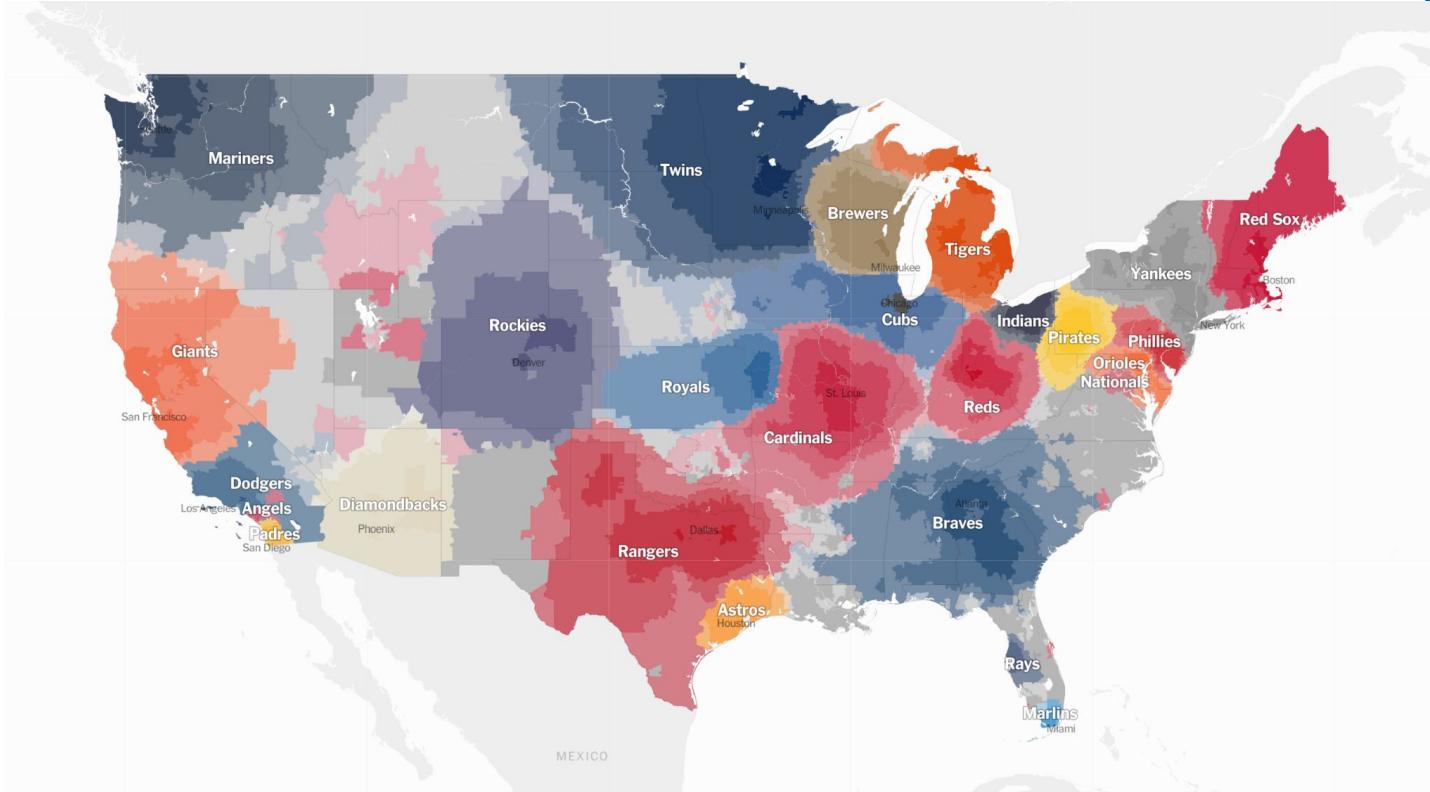
The types of jobs that pay middle-class wages — between \$40,000 and \$80,000 in 2014 dollars — have shifted since 1980. Fewer of these positions are in male-dominated production occupations, while a greater share are in workplaces more open to women.

RELATED ARTICLE



Interactive site:

<https://www.nytimes.com/interactive/2015/02/23/business/economy/the-changing-nature-of-middle-class-jobs.html>



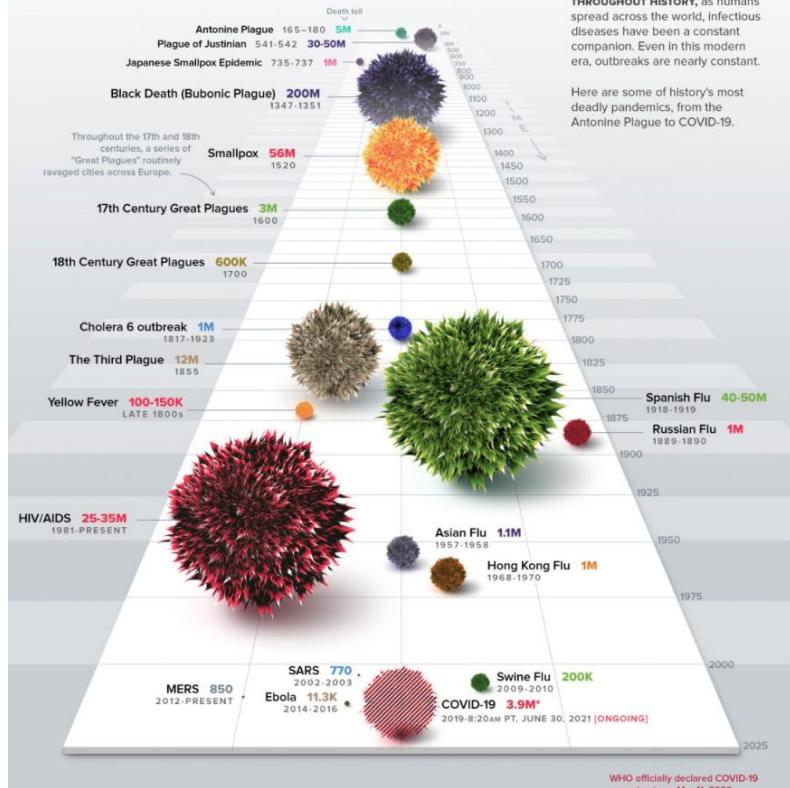
Baseball Fans by zip code (Facebook) - 2014

Source:

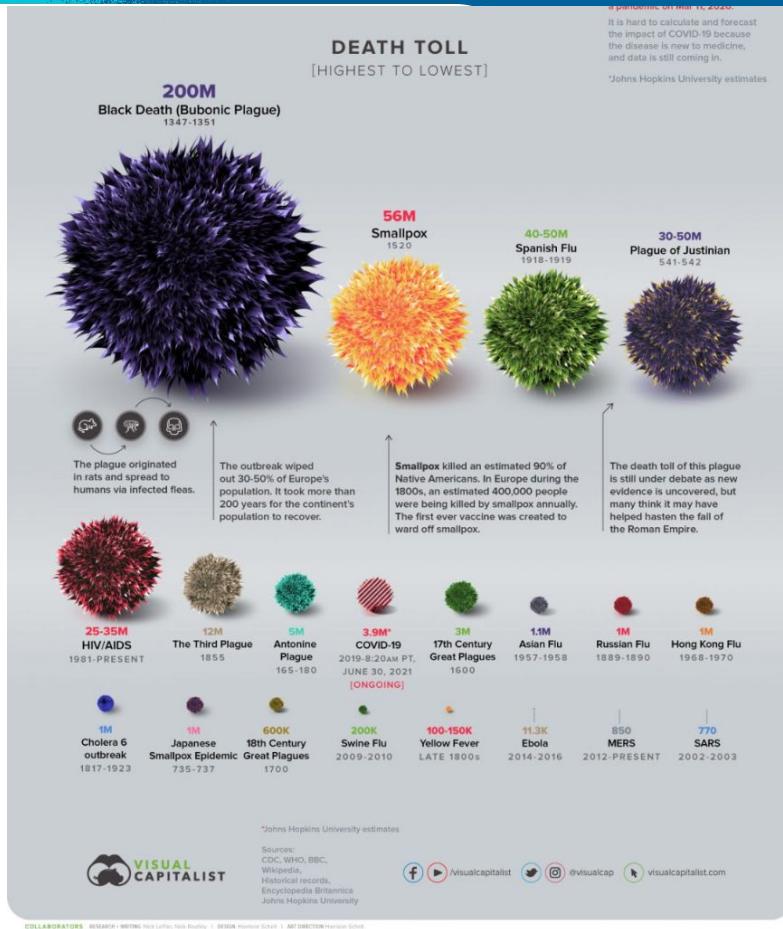
<https://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html>

HISTORY OF PANDEMICS

PAN-DEM-IC (of a disease) prevalent over a whole country or the world.

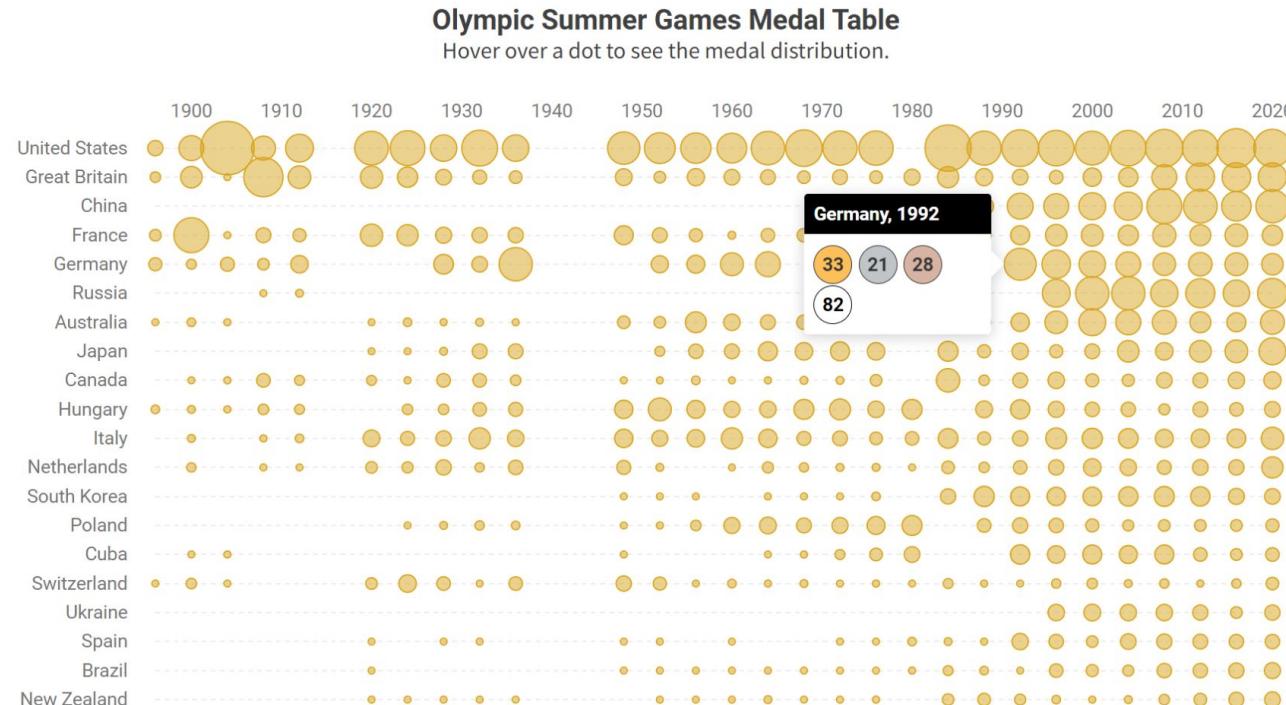


a pandemic on Mar 11, 2020.
It is hard to calculate and forecast the impact of COVID-19 because the disease is new to medicine, and data is still coming in.
*Johns Hopkins University estimates



<https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

Diego

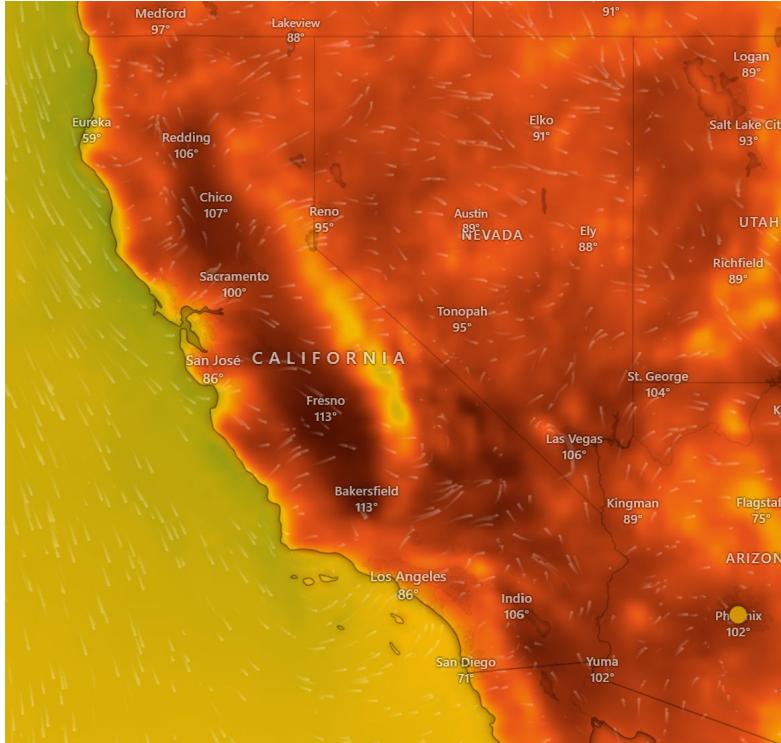


Source: [Bloomberg](#) • Created with the Flourish [Scatter](#) template



Do the real-time updates help? How?

Weather Visualization



Heatwave this week in CA. This image came from the HRRR model predicting temperature at 5pm today

Heatmaps convey variance over space.

Animated heatmaps can convey spatial behavior, such as a storm's path

Source: <https://windy.com>

Data Vis for artistic purposes



Lev Manovich with Jeremy Douglass.

TimeLine: High resolution stitching of all 4535 covers of Time magazine (1923-2009)

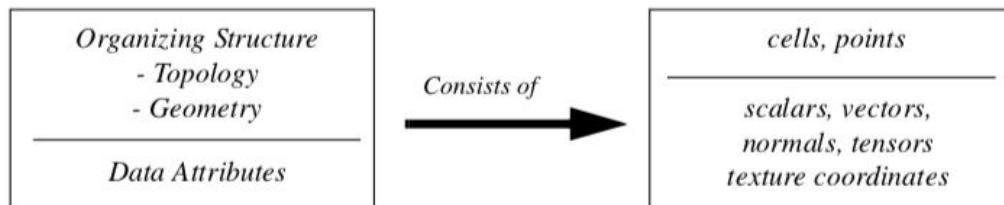
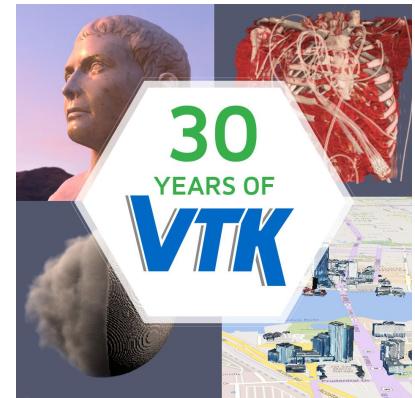


Dr. Thorne did indeed publish his paper: <https://arxiv.org/abs/1502.03808>

VTK Demo

The Visualization Toolkit (VTK)

VTK was originally published as part of the textbook [The Visualization Toolkit, An Object-Oriented Approach to 3D Graphics](#) by Will Schroeder, Ken Martin, and Bill Lorensen—three graphics and visualization researchers at General Electric (1993)



VTK Cell Types

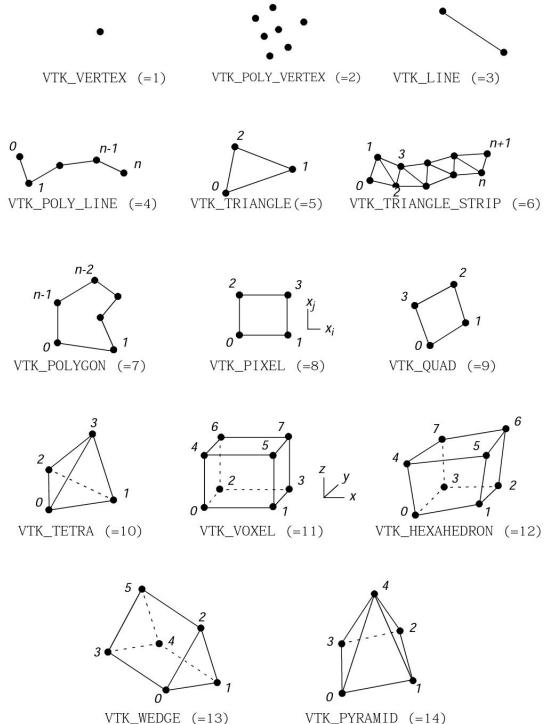


Figure 2 Linear cell types found in VTK. Use the include file `CellType.h` to manipulate cell types.

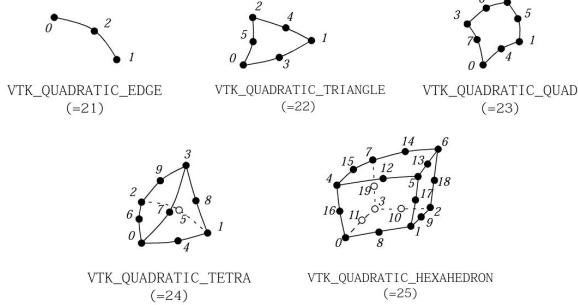
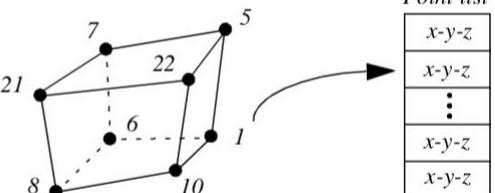


Figure 3 Non-linear cell types found in VTK.

Data cells are simple topological elements like points, lines, and polygons of which visualization datasets are composed.

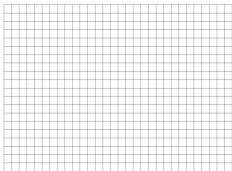
*Definition:
Type: hexahedron
Connectivity: (8,10,1,6,21,22,5,7)*



Hexahedron cell example

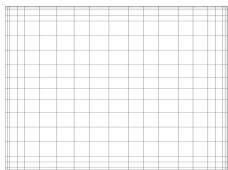
VTK Datasets

A dataset consists of an organizing structure plus associated attribute data.
Five types of datasets are implemented:



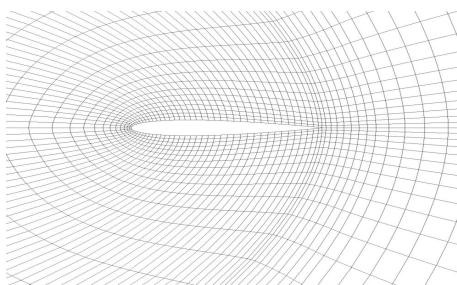
UNIFORM
RECTILINEAR GRID
(image data)

implicit topology and point
coordinates



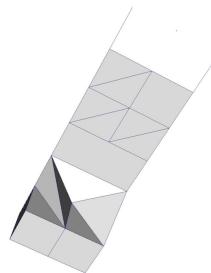
RECTILINEAR GRID

implicit topology and semi-implicit
point coordinates



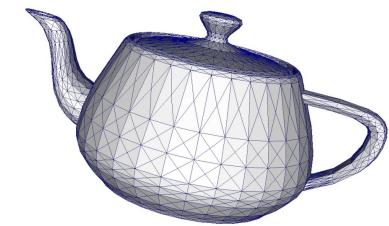
STRUCTURED GRID

implicit topology, explicit point
coordinates



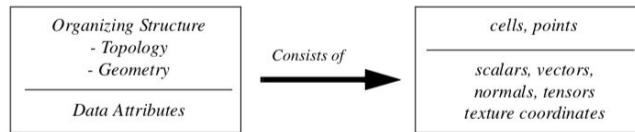
UNSTRUCTURED GRID

explicit topology, explicit point
coordinates



POLYDATA

explicit topology, explicit point
coordinates



Jupyter Tutorial:

Please download the notebooks available in the summer institute github:

[https://github.com/sdsc/sdsc-summer-institute-2024/tree/main/4.3a scientific vis vtk engine5](https://github.com/sdsc/sdsc-summer-institute-2024/tree/main/4.3a%20scientific%20vis%20vtk%20engine5)

ParaView is recommended:

<https://www.paraview.org/download/>