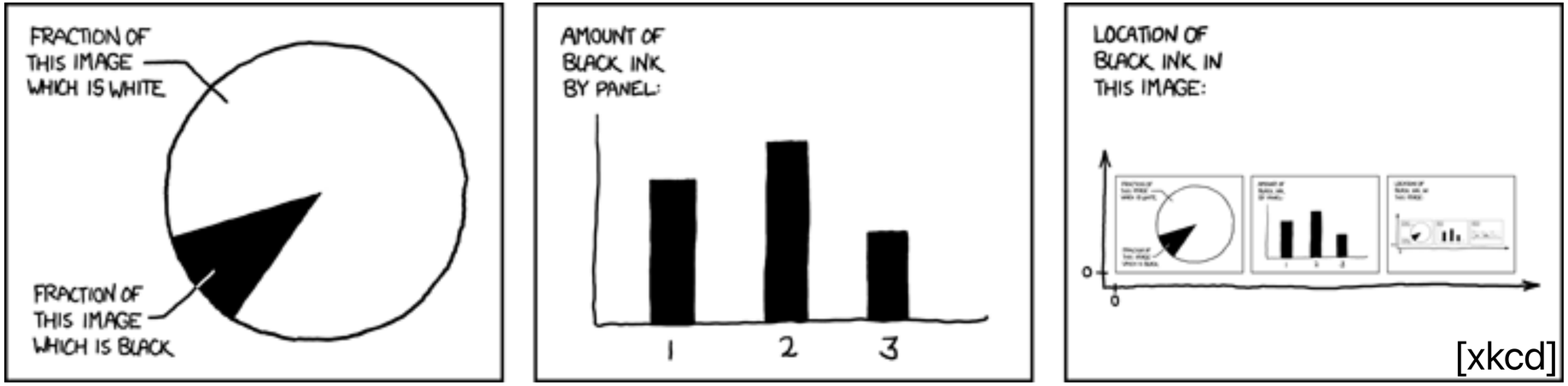


Basics of Data Visualization

DSAIT4000

Thomas Höllt - Computer Graphics & Visualization



"The purpose of computing is insight, not numbers."

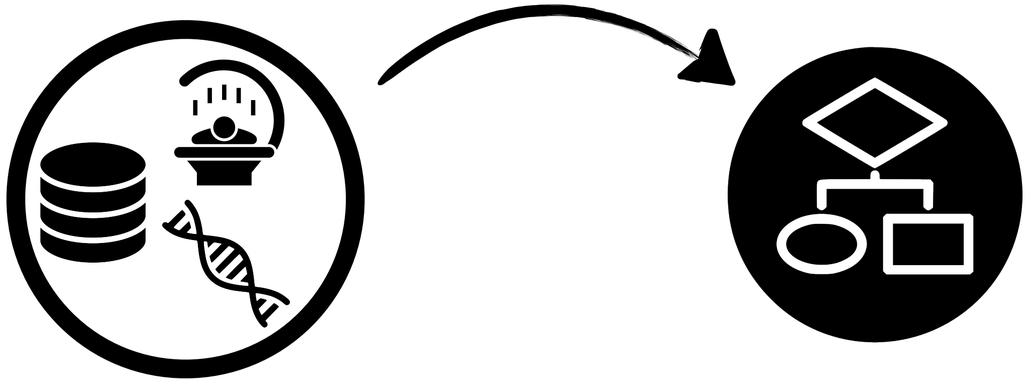
R.W. Hamming, *Numerical Methods for Scientists and Engineers*, 1962

Data Analytics



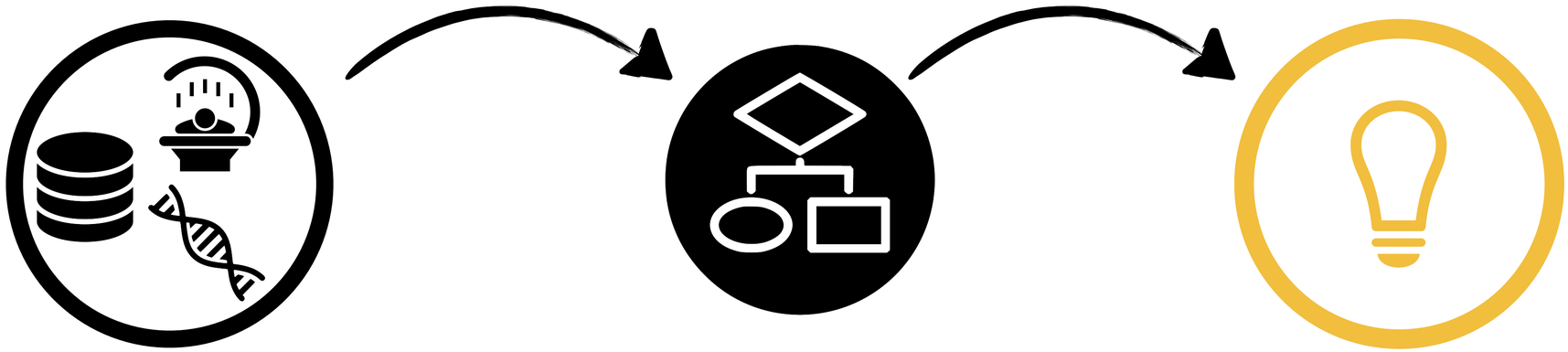
Keim et al., Lecture notes in computer science, 2008

Data Analytics



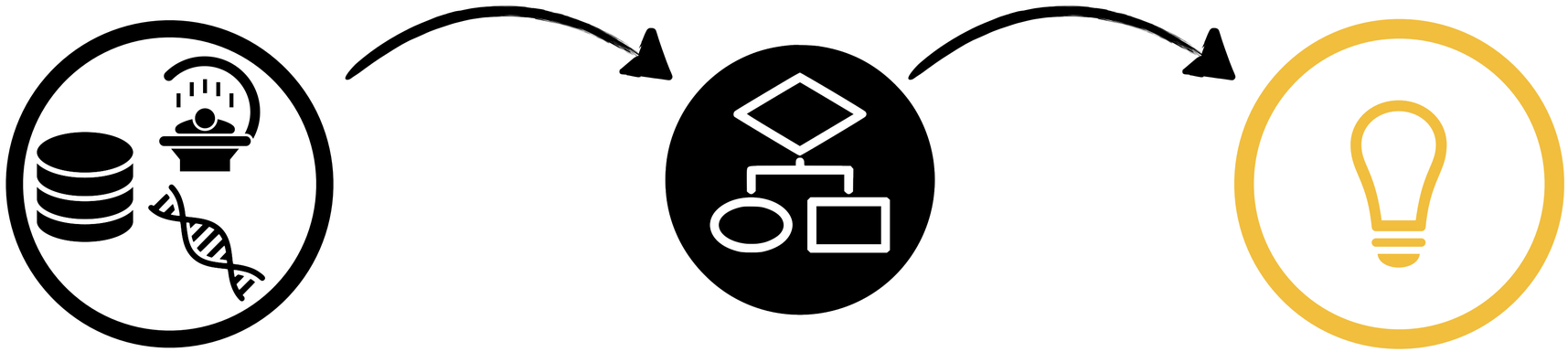
Keim et al., Lecture notes in computer science, 2008

Data Analytics



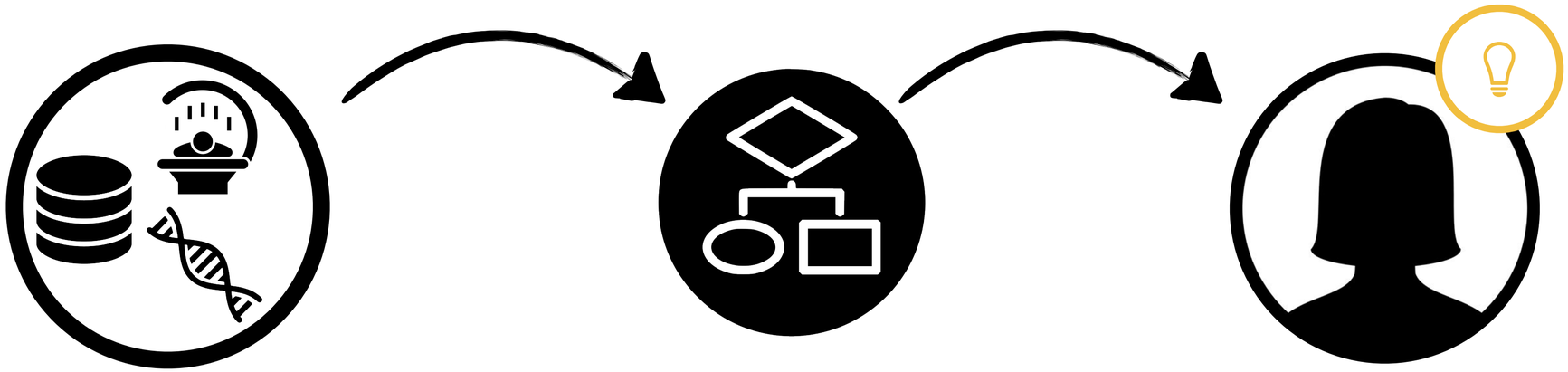
Keim et al., Lecture notes in computer science, 2008

Where is Insight Generated?



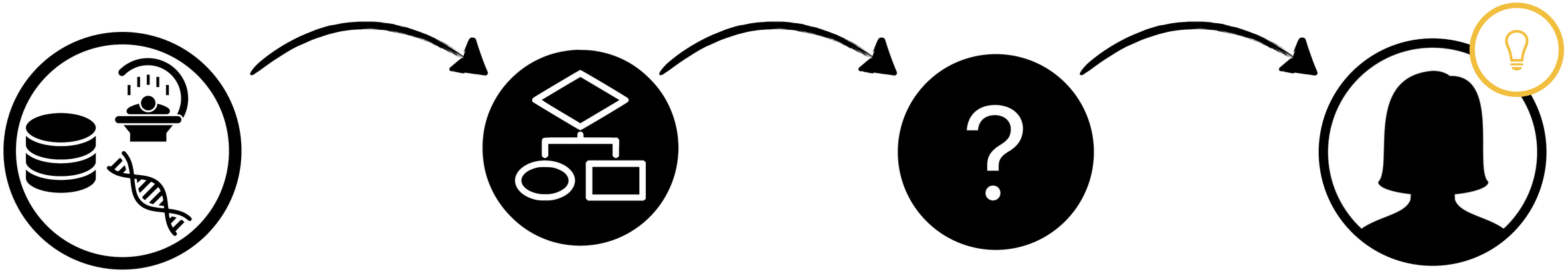
Keim et al., Lecture notes in computer science, 2008

Where is Insight Generated?



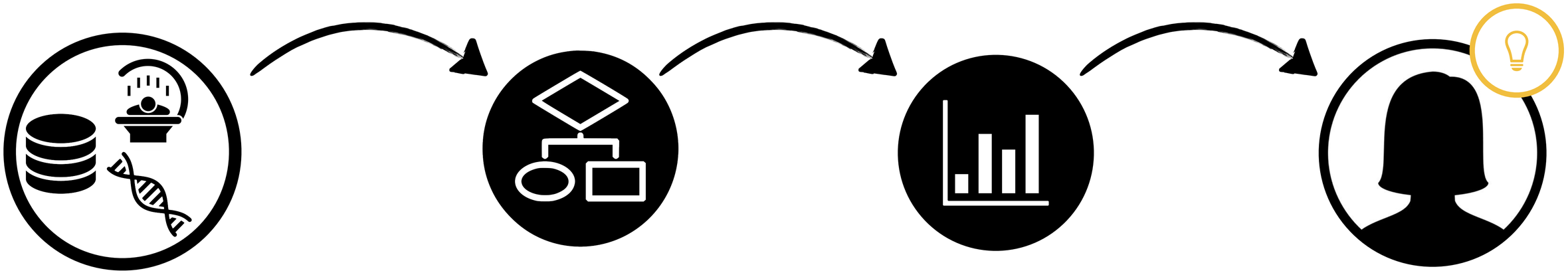
after Keim et al., Lecture notes in computer science, 2008

How is Insight Generated?



after Keim et al., Lecture notes in computer science, 2008

How is Insight Generated?



after Keim et al., Lecture notes in computer science, 2008

Human in the Loop

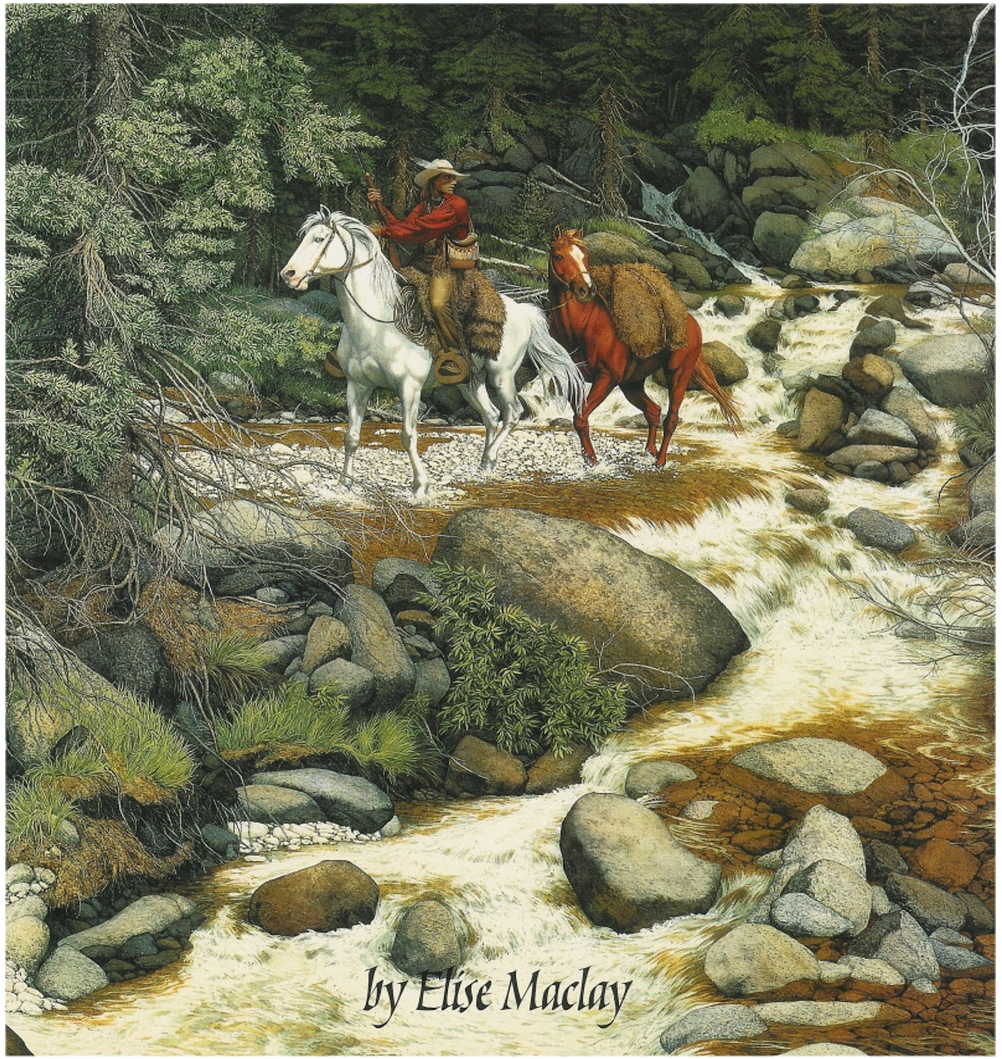
5 Human Senses

5 Human Senses

- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition

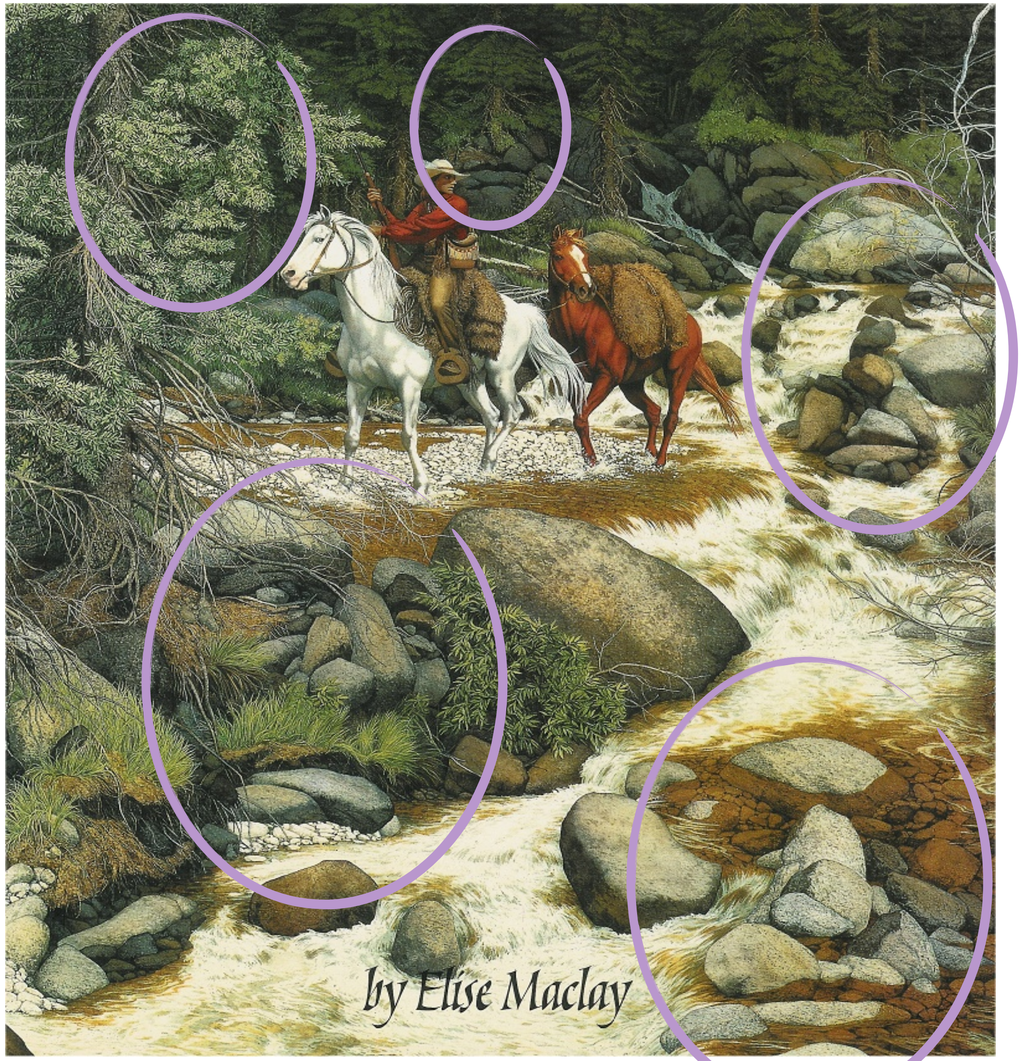
5 Human Senses

- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition



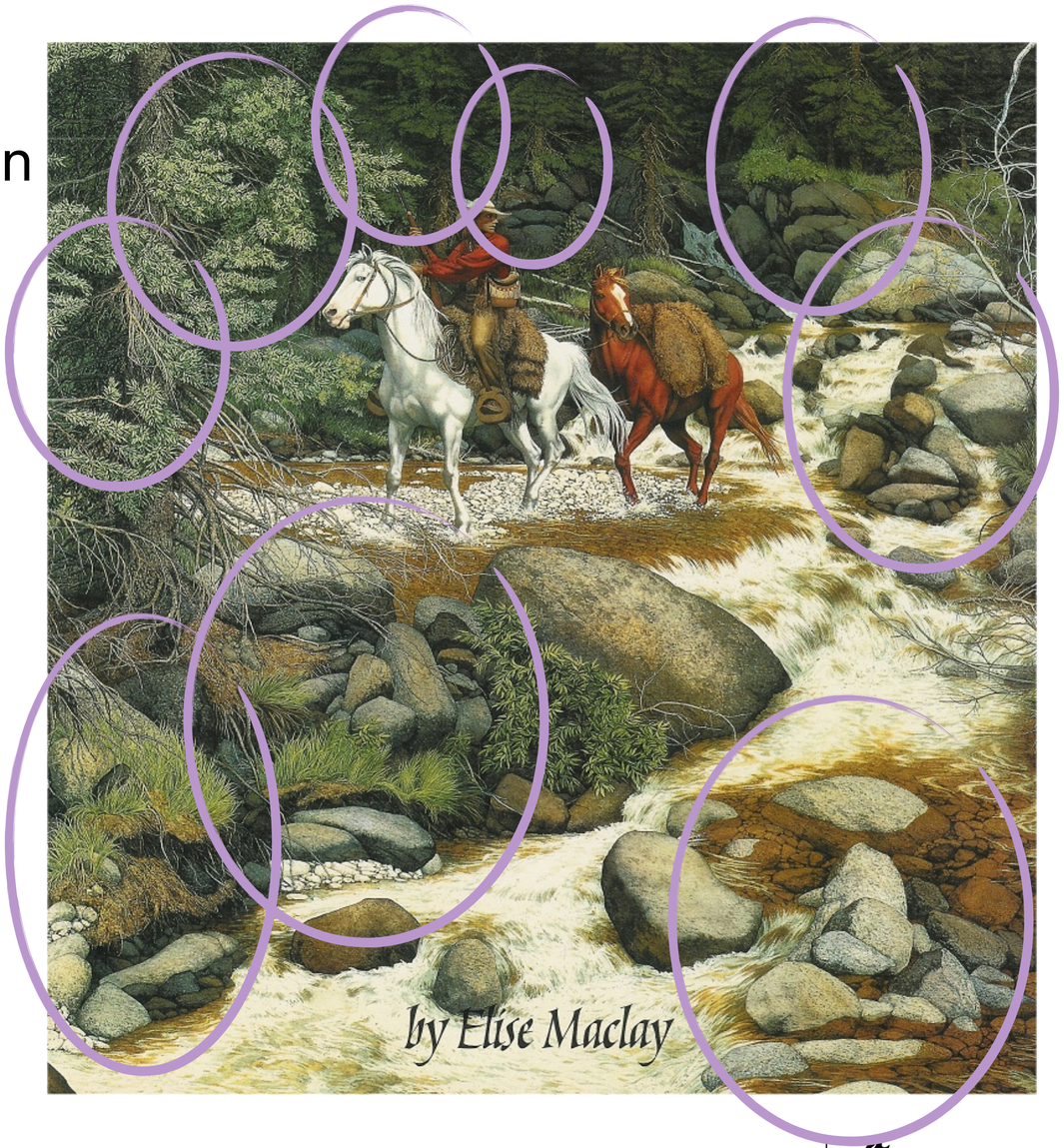
5 Human Senses

- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition



5 Human Senses

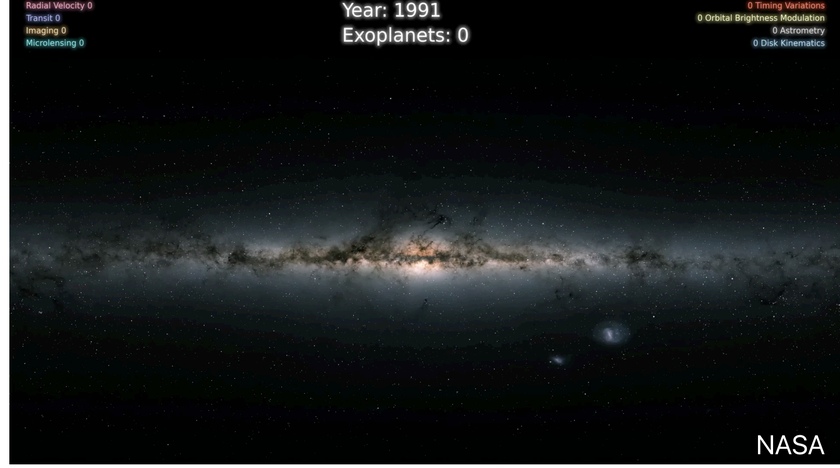
- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition



5 Human Senses

- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition
- **audio**: lower bandwidth and different semantics
  - no overview
  - sequential stream
- **touch**/haptics: impoverished record/replay capacity
  - very low-bandwidth communication thus far
- **taste, smell**: no viable record/replay devices

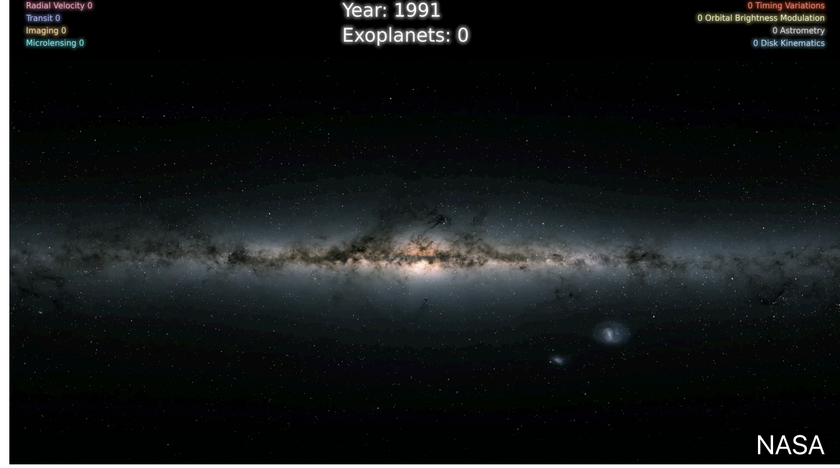
5 Human Senses



- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition
- **audio**: lower bandwidth and different semantics
  - no overview
  - sequential stream
- **touch**/haptics: impoverished record/replay capacity
  - very low-bandwidth communication thus far
- **taste, smell**: no viable record/replay devices

5 Human Senses

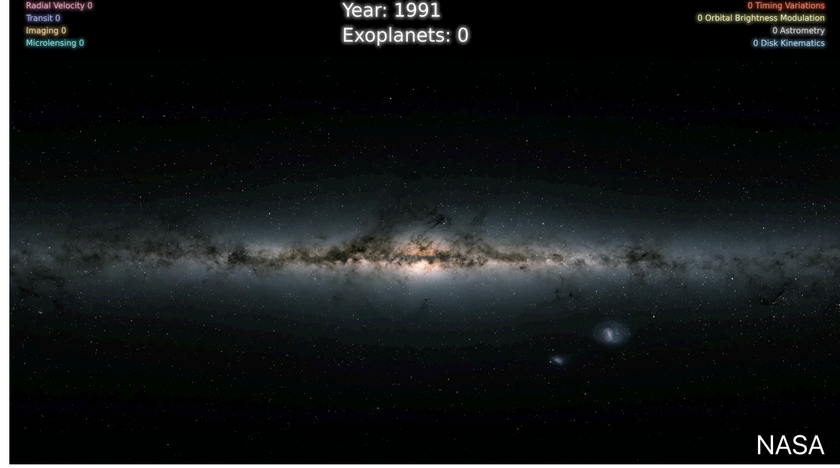
- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition
- **audio**: lower bandwidth and different semantics
  - no overview
  - sequential stream
- **touch**/haptics: impoverished record/replay capacity
  - very low-bandwidth communication thus far
- **taste, smell**: no viable record/replay devices





5 Human Senses

- **human visual system** is high-bandwidth channel to brain
  - overview possible, we see everything simultaneously
  - significant processing occurs in parallel and pre-attentively
  - very strong pattern recognition
- **audio**: lower bandwidth and different semantics
  - no overview
  - sequential stream
- **touch**/haptics: impoverished record/replay capacity
  - very low-bandwidth communication thus far
- **taste, smell**: no viable record/replay devices

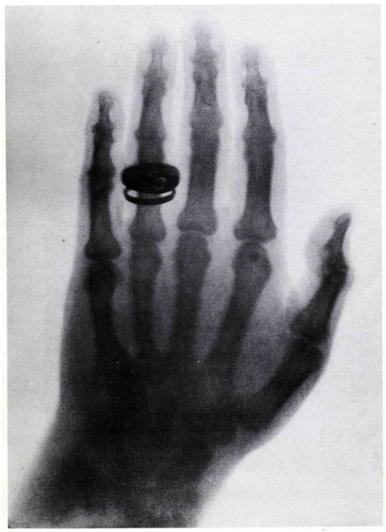


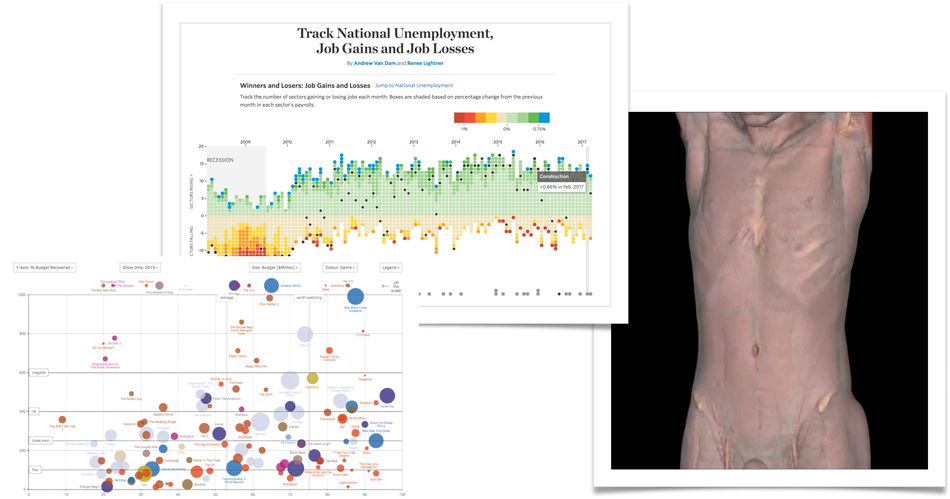




What is Visualization?

A		B		C		D	
x	y	x	y	x	y	x	y
10	8	10	9	10	7	8	7
8	7	8	8	8	7	8	6
13	8	13	9	13	13	8	8
9	9	9	9	9	7	8	9
11	8	11	9	11	8	8	8
14	10	14	8	14	9	8	7
6	7	6	6	6	6	8	5
4	4	4	3	4	5	19	13
12	11	12	9	12	8	8	6
7	5	7	7	7	6	8	8
5	6	5	5	5	6	8	7

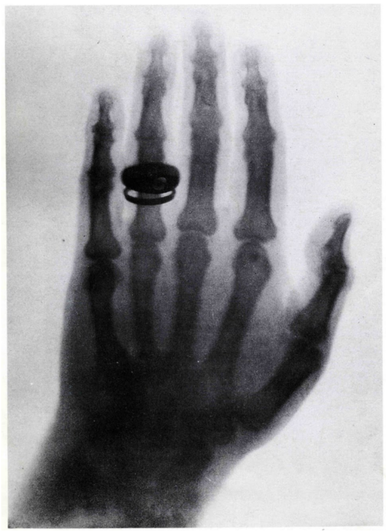


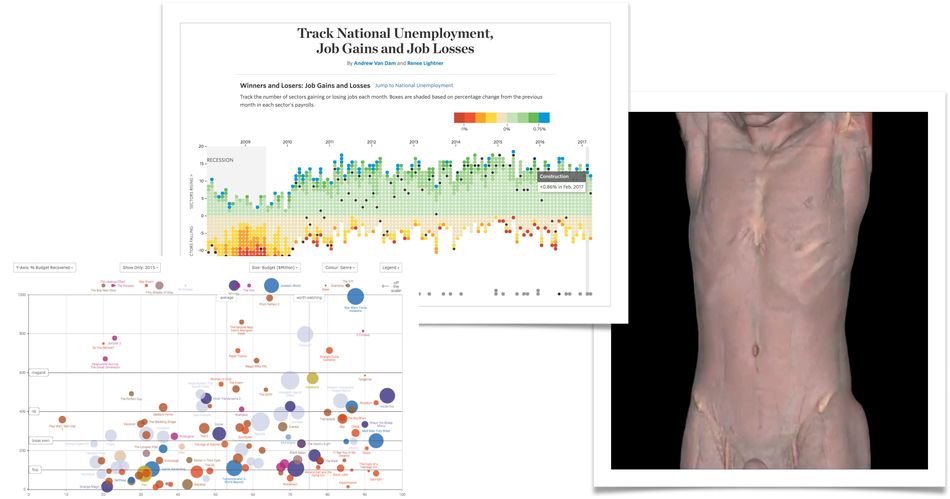


Visualization is the process that transforms

(abstract) data into **graphical representations**

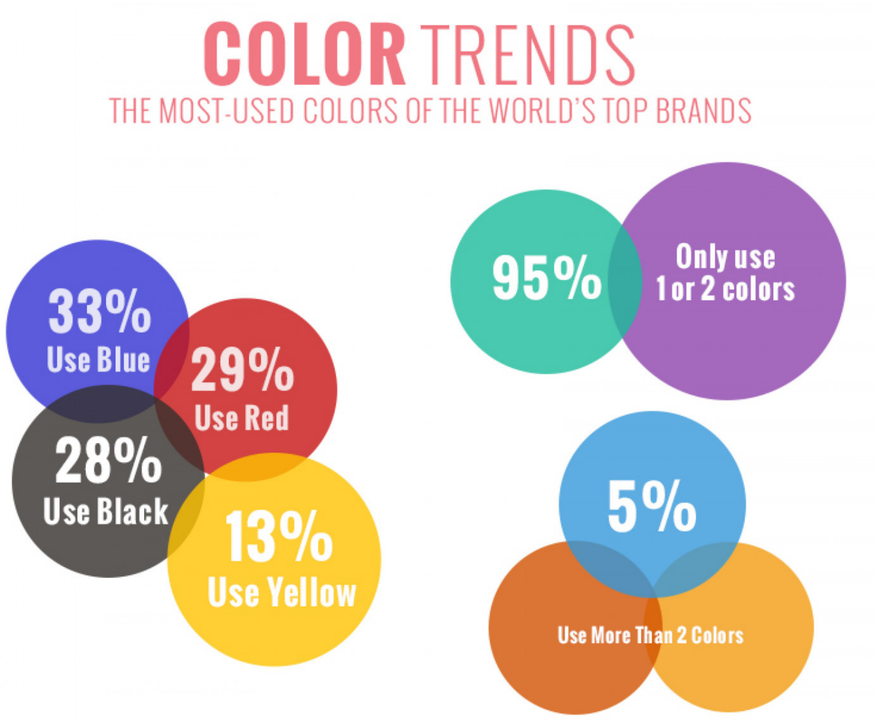
A		B		C		D	
x	y	x	y	x	y	x	y
10	8	10	9	10	7	8	7
8	7	8	8	8	7	8	6
13	8	13	9	13	13	8	8
9	9	9	9	9	7	8	9
11	8	11	9	11	8	8	8
14	10	14	8	14	9	8	7
6	7	6	6	6	6	8	5
4	4	4	3	4	5	19	13
12	11	12	9	12	8	8	6
7	5	7	7	7	6	8	8
5	6	5	5	5	6	8	7

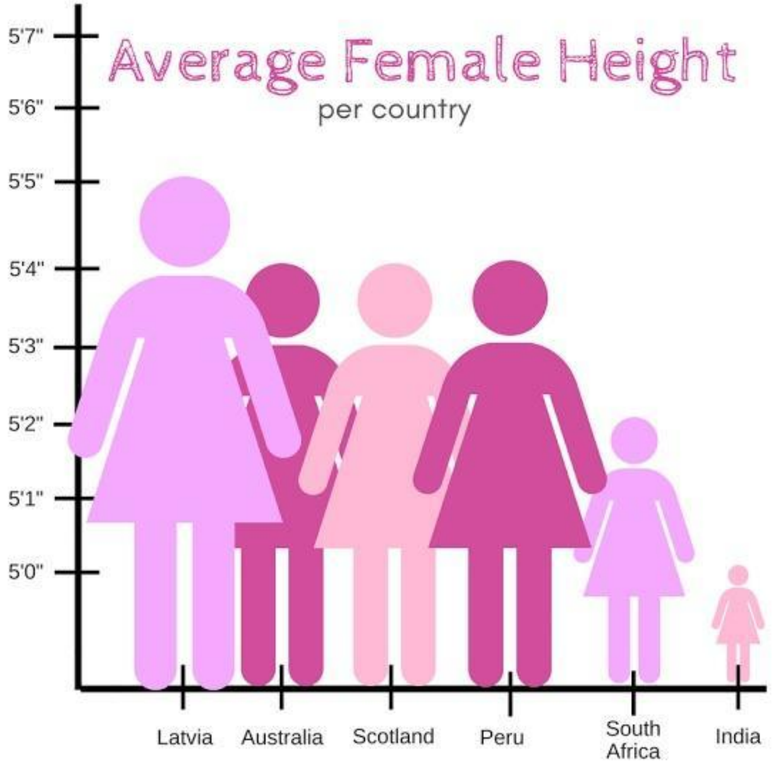


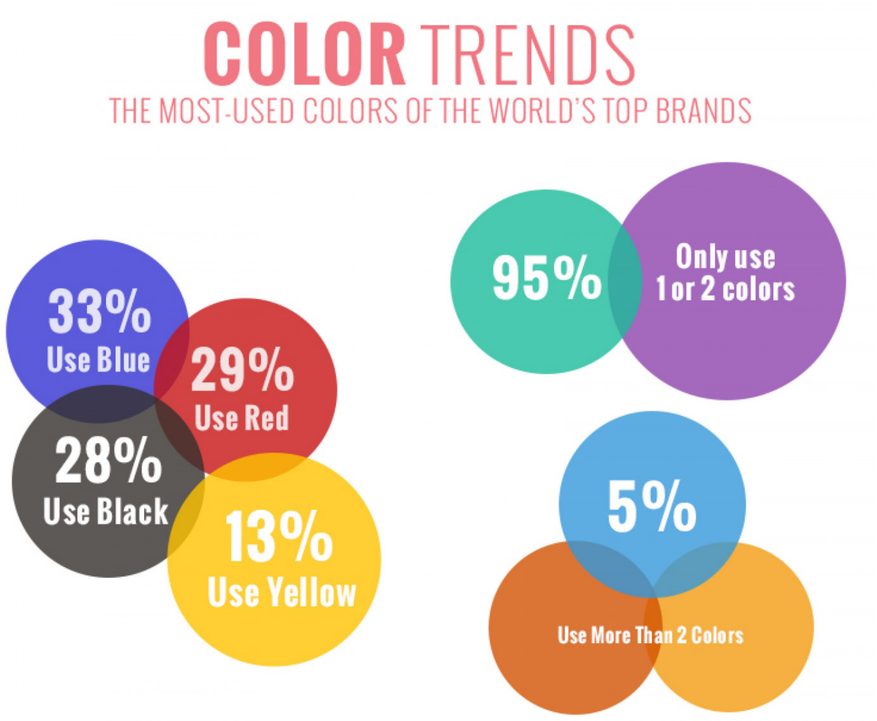


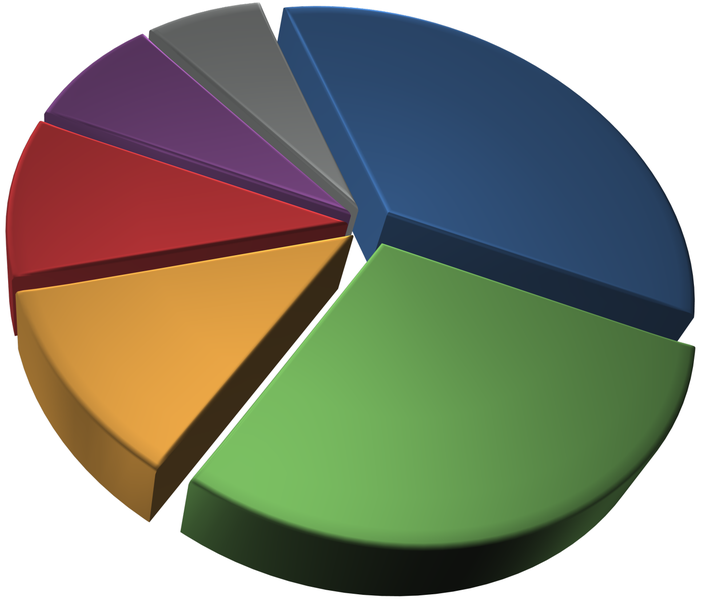
Visualization is the process that transforms

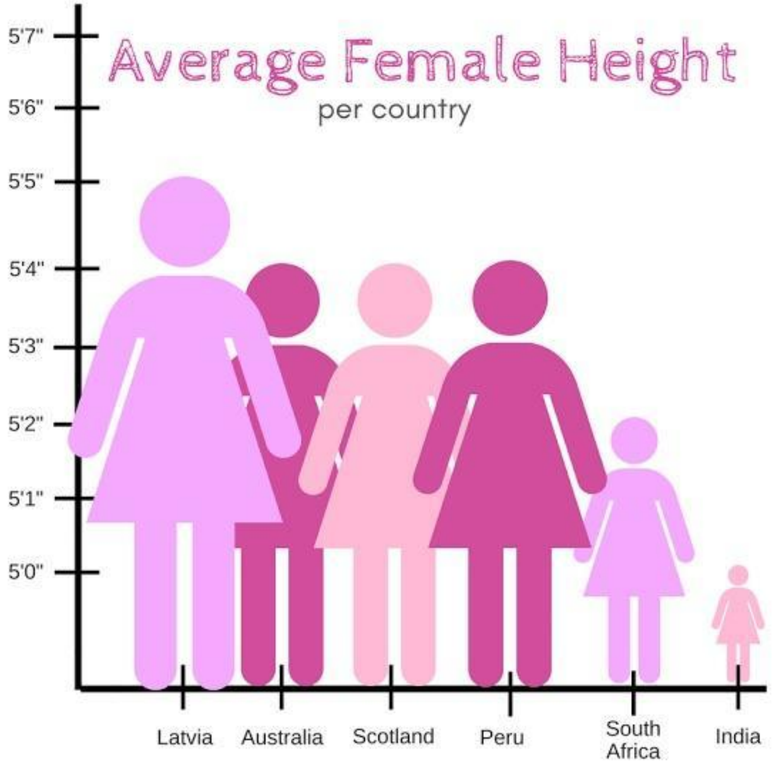
(abstract) data into **graphical representations**


COLOR TRENDS
THE MOST-USED COLORS OF THE WORLD'S TOP BRANDS
33%
Use Blue
29%
Use Red
28%
Use Black
13%
Use Yellow
95%
Only use
1 or 2 colors
5%
Use More Than 2 Colors


Average Female Height
per country
5'7"
5'6"
5'5"
5'4"
5'3"
5'2"
5'1"
5'0"
Latvia
Australia
Scotland
Peru
South Africa
India


COLOR TRENDS
THE MOST-USED COLORS OF THE WORLD'S TOP BRANDS
33%
Use Blue
29%
Use Red
28%
Use Black
13%
Use Yellow
95%
Only use
1 or 2 colors
5%
Use More Than 2 Colors




Average Female Height
per country
5'7"
5'6"
5'5"
5'4"
5'3"
5'2"
5'1"
5'0"
Latvia
Australia
Scotland
Peru
South Africa
India

"The purpose of computing is insight, not numbers."

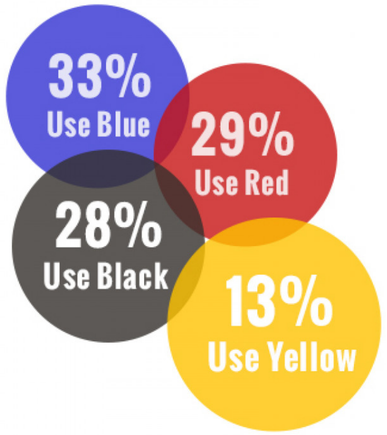
visualization

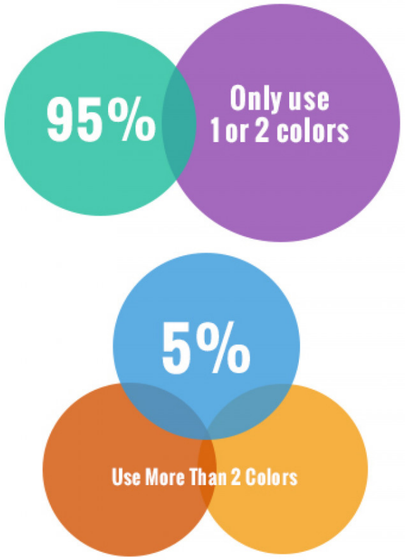
pictures

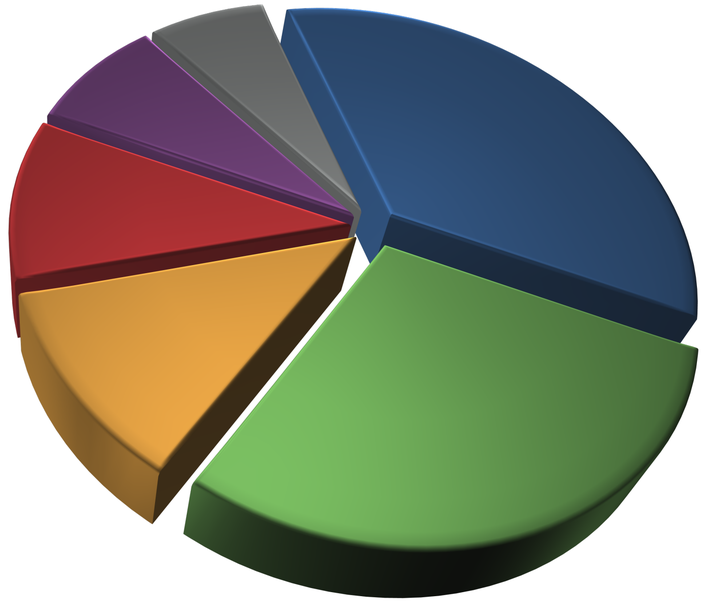
"The purpose of ~~computing~~ is insight, not ~~numbers~~."

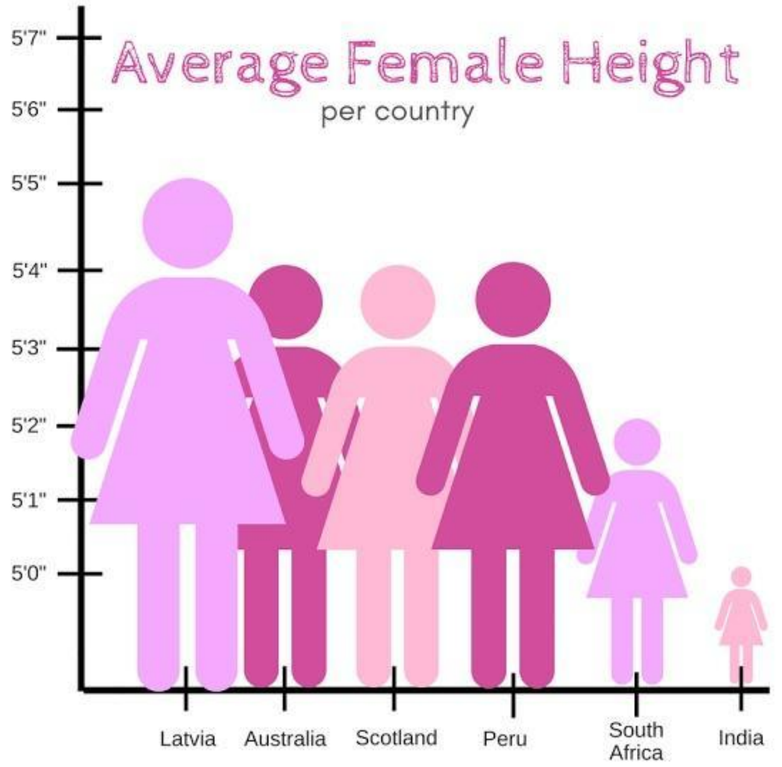
COLOR TRENDS

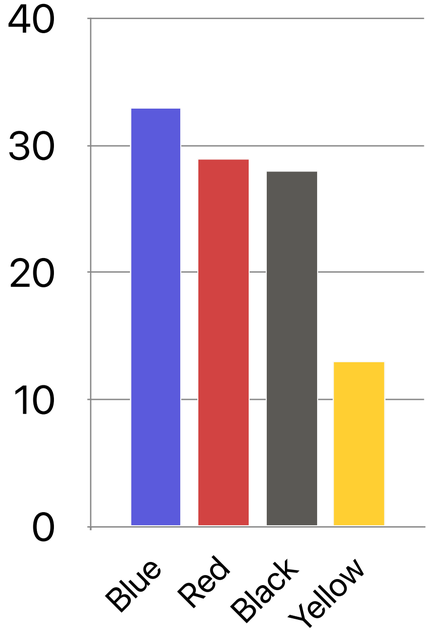
THE MOST-USED COLORS OF THE WORLD'S TOP BRANDS

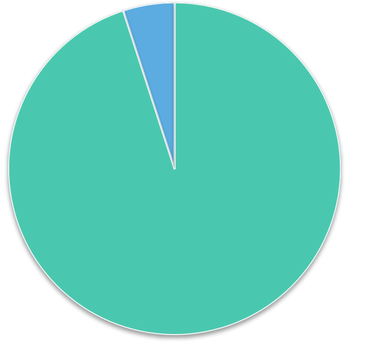


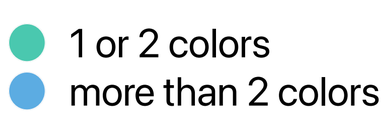


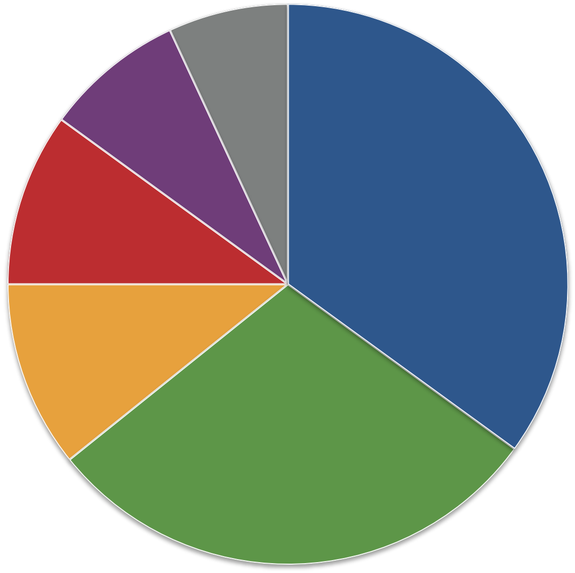


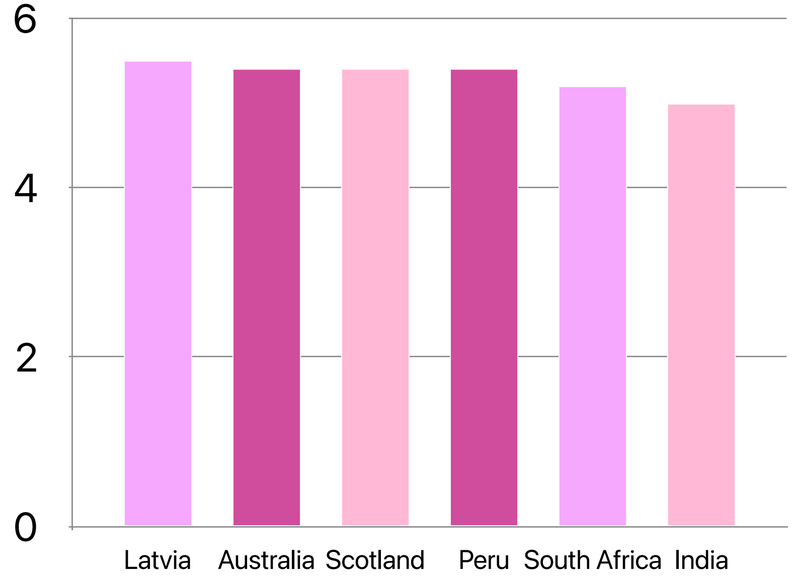




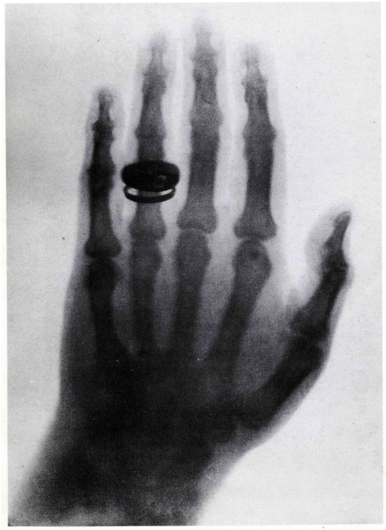


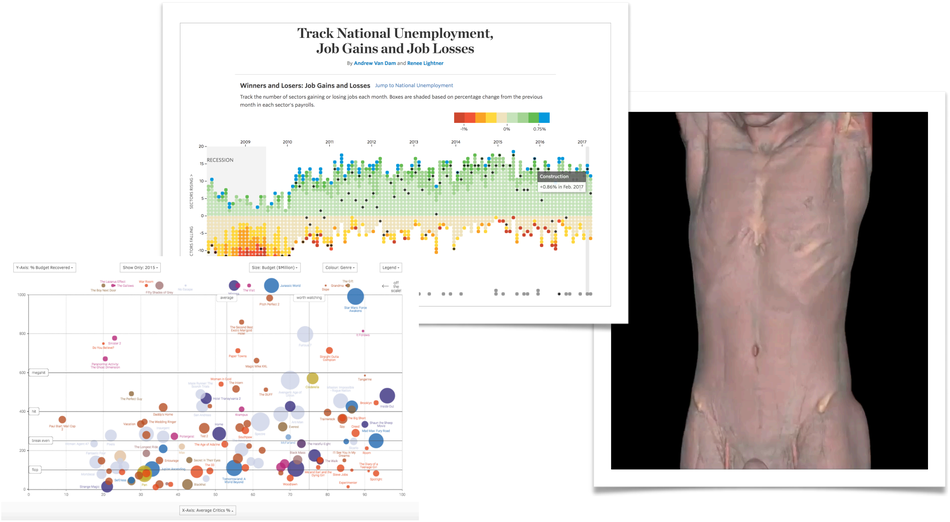






A		B		C		D	
x	y	x	y	x	y	x	y
10	8	10	9	10	7	8	7
8	7	8	8	8	7	8	6
13	8	13	9	13	13	8	8
9	9	9	9	9	7	8	9
11	8	11	9	11	8	8	8
14	10	14	8	14	9	8	7
6	7	6	6	6	6	8	5
4	4	4	3	4	5	19	13
12	11	12	9	12	8	8	6
7	5	7	7	7	6	8	8
5	6	5	5	5	6	8	7

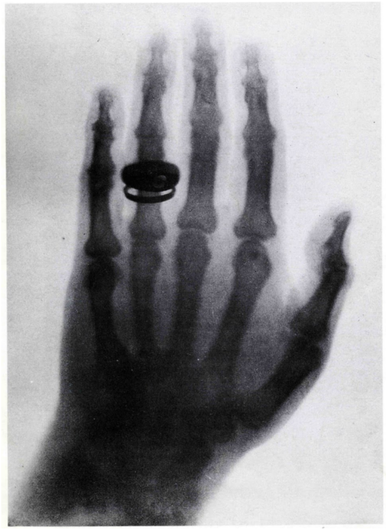


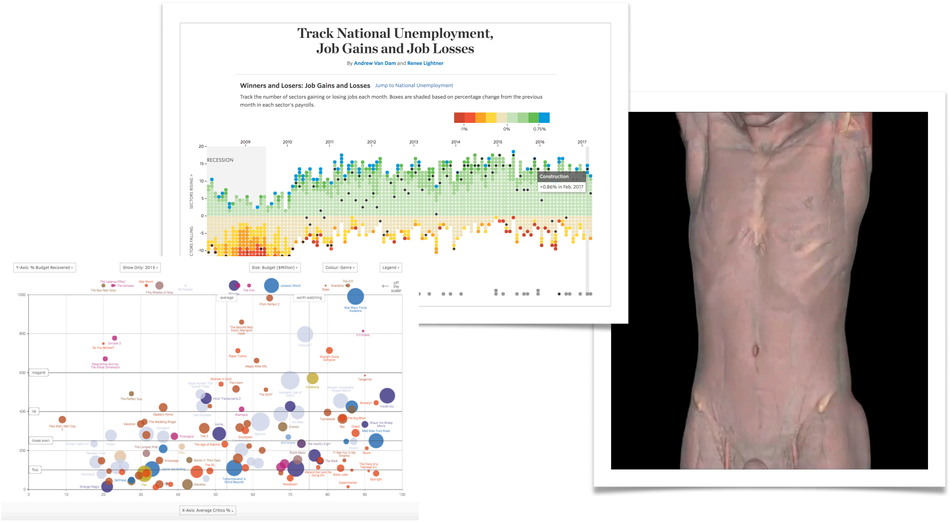


Visualization is the process that transforms

(abstract) data into graphical representations

A		B		C		D	
x	y	x	y	x	y	x	y
10	8	10	9	10	7	8	7
8	7	8	8	8	7	8	6
13	8	13	9	13	13	8	8
9	9	9	9	9	7	8	9
11	8	11	9	11	8	8	8
14	10	14	8	14	9	8	7
6	7	6	6	6	6	8	5
4	4	4	3	4	5	19	13
12	11	12	9	12	8	8	6
7	5	7	7	7	6	8	8
5	6	5	5	5	6	8	7

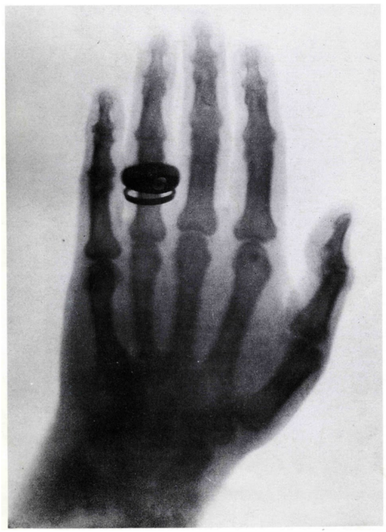


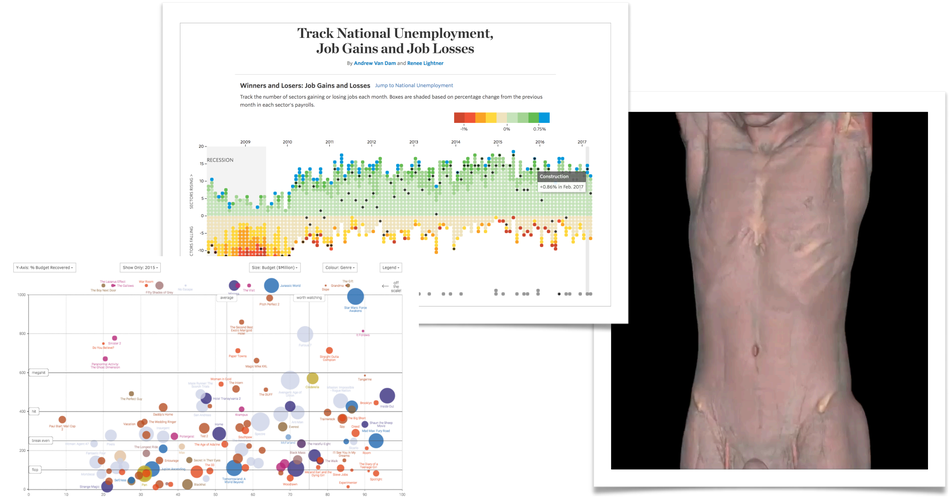


Visualization is the process that transforms

(abstract) data into graphical representations

A		B		C		D	
x	y	x	y	x	y	x	y
10	8	10	9	10	7	8	7
8	7	8	8	8	7	8	6
13	8	13	9	13	13	8	8
9	9	9	9	9	7	8	9
11	8	11	9	11	8	8	8
14	10	14	8	14	9	8	7
6	7	6	6	6	6	8	5
4	4	4	3	4	5	19	13
12	11	12	9	12	8	8	6
7	5	7	7	7	6	8	8
5	6	5	5	5	6	8	7



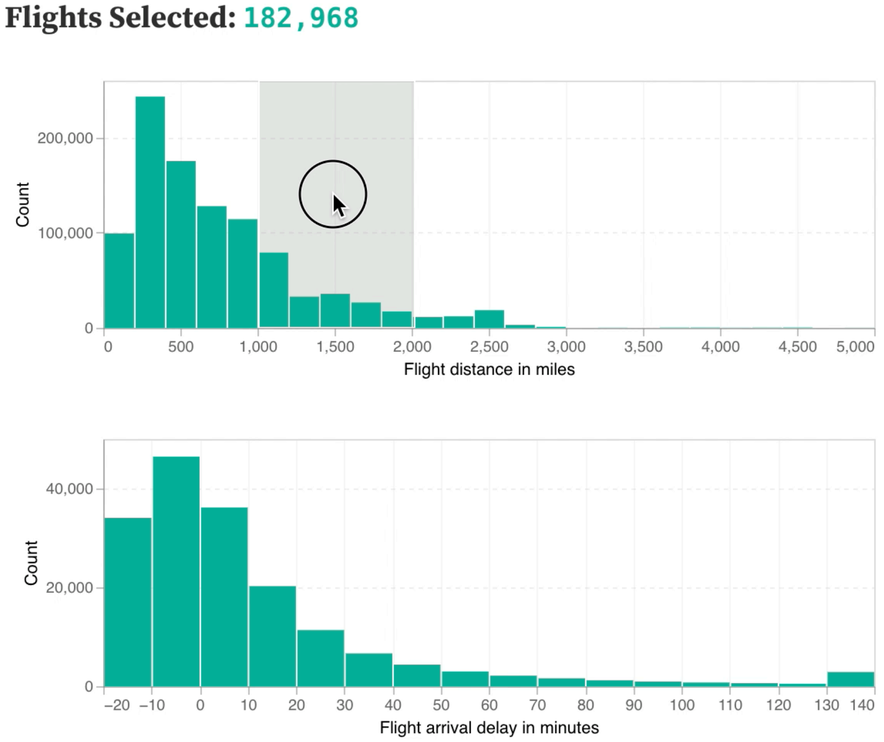


Visualization is the process that transforms

(abstract) data into **interactive** graphical representations

Visualization Is (Often) Interactive

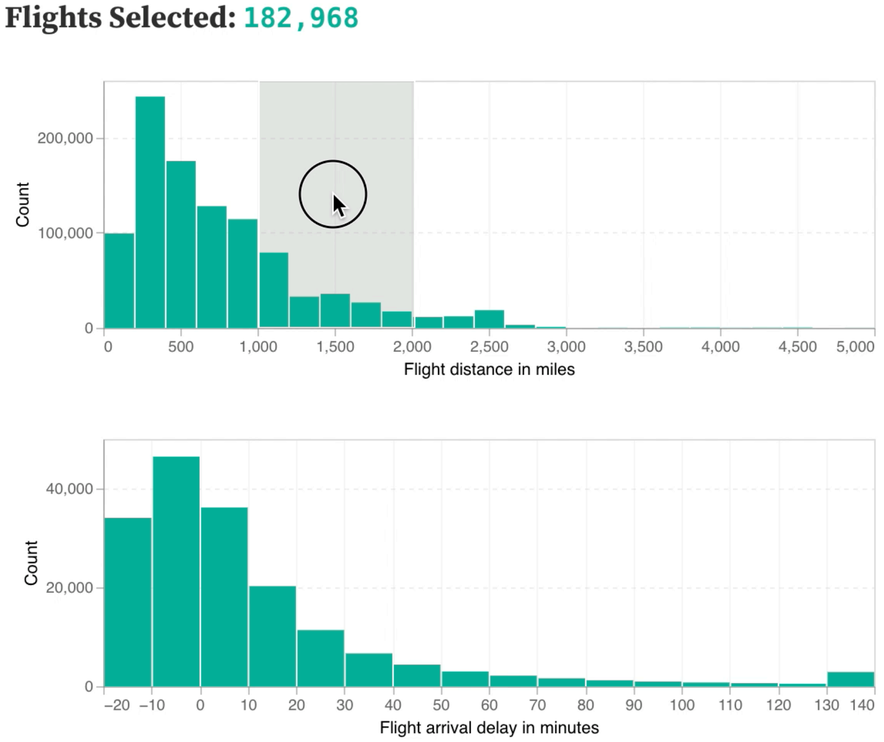






Visualization Is (Often) Interactive







Weekly Danceability Scores Over the Year



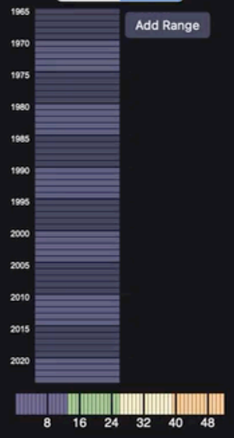
Selector Dashboard

Include The Following Top Positions:

1 3 5 10 20 40

Genres Features

Add Range



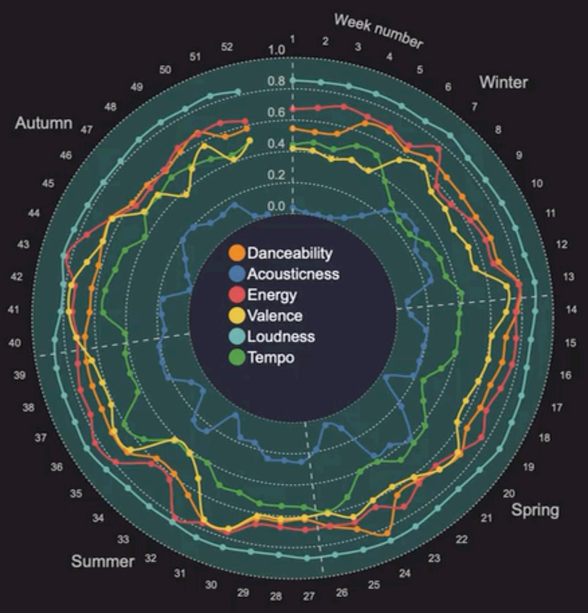
Category-based Longevity of Scores

add standard deviation area

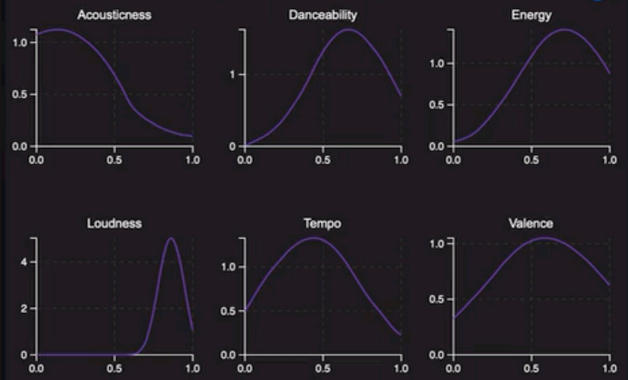


Short Hits Medium Hits Long Hits

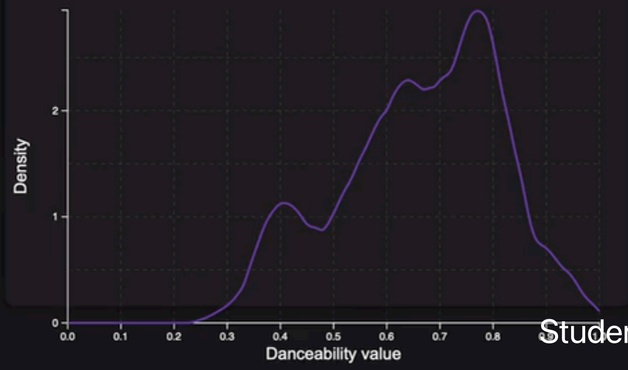
Weekly Scores All Features Averaged over 2000 - 2005



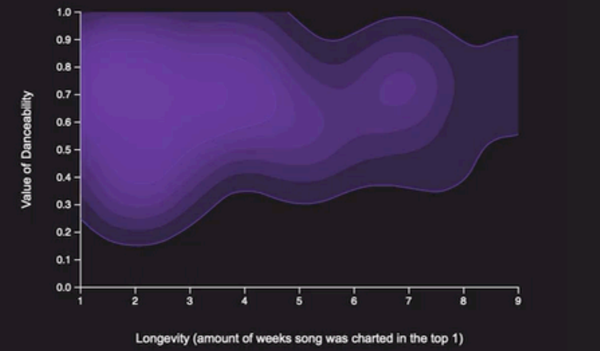
Probability Density Function for Features



Detailed view: Danceability



Danceability Value Distribution by Longevity for S - e



Weekly Danceability Scores Over the Year



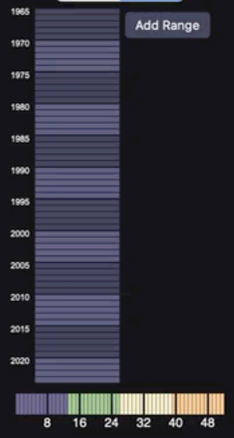
Selector Dashboard

Include The Following Top Positions:

1 3 5 10 20 40

Genres Features

Add Range



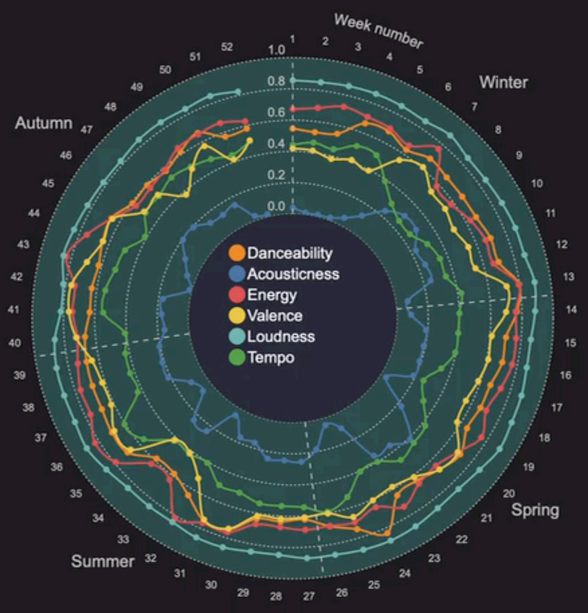
Category-based Longevity of Scores

add standard deviation area

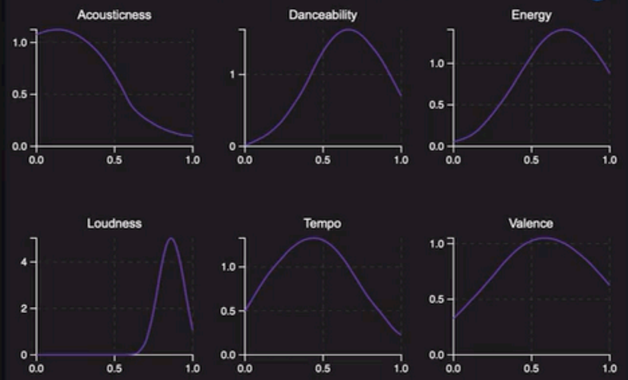


Short Hits Medium Hits Long Hits

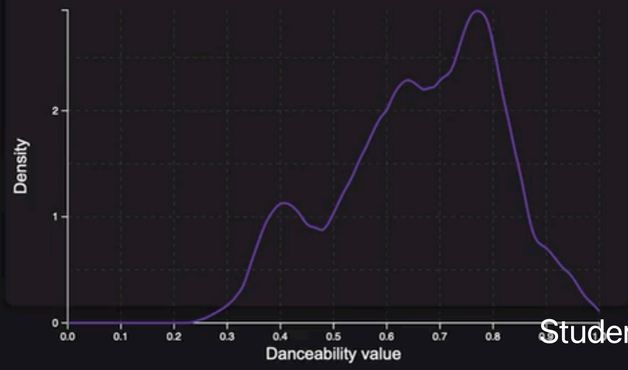
Weekly Scores All Features Averaged over 2000 - 2005



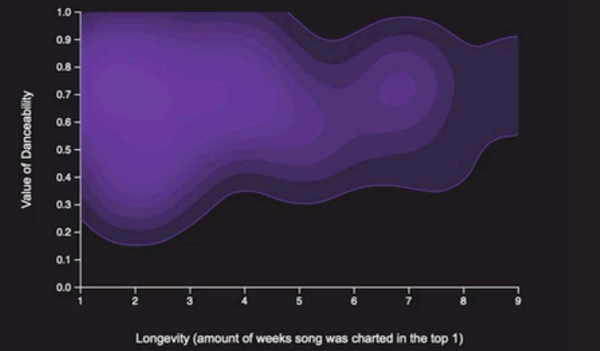
Probability Density Function for Features



Detailed view: Danceability



Danceability Value Distribution by Longevity for S - e





3D Visualization

Load | CPU Raycaster | GPU Raycaster | Transfer function

rendering time(last new frame): 0ms
1316.6267443494 FPS
rendering resolution: (760, 760)

Render Mode:

- (○) MIP
- (○) IsoSurface Rendering
- (●) Compositing

1.000 — Step size

☐ Volume Shading

95.000 — Iso Value

☐ Opacity Modulation

0.000 — Boundary kc
1.000 — Boundary ks
1.000 — Boundary ke

3 — RenderStep

☑ Empty space skipping

8 — Block size

☐ Use volume bricking

32 — Brick size

Interpolation:

- (○) Nearest Neighbour
- (●) Linear



3D Visualization

Load | CPU Raycaster | GPU Raycaster | Transfer function

rendering time(last new frame): 0ms
1316.6267443494 FPS
rendering resolution: (760, 760)

Render Mode:

- (○) MIP
- (○) IsoSurface Rendering
- (●) Compositing

1.000 — Step size

☐ Volume Shading

95.000 — Iso Value

☐ Opacity Modulation

0.000 — Boundary kc
1.000 — Boundary ks
1.000 — Boundary ke

3 — RenderStep

☑ Empty space skipping

8 — Block size

☐ Use volume bricking

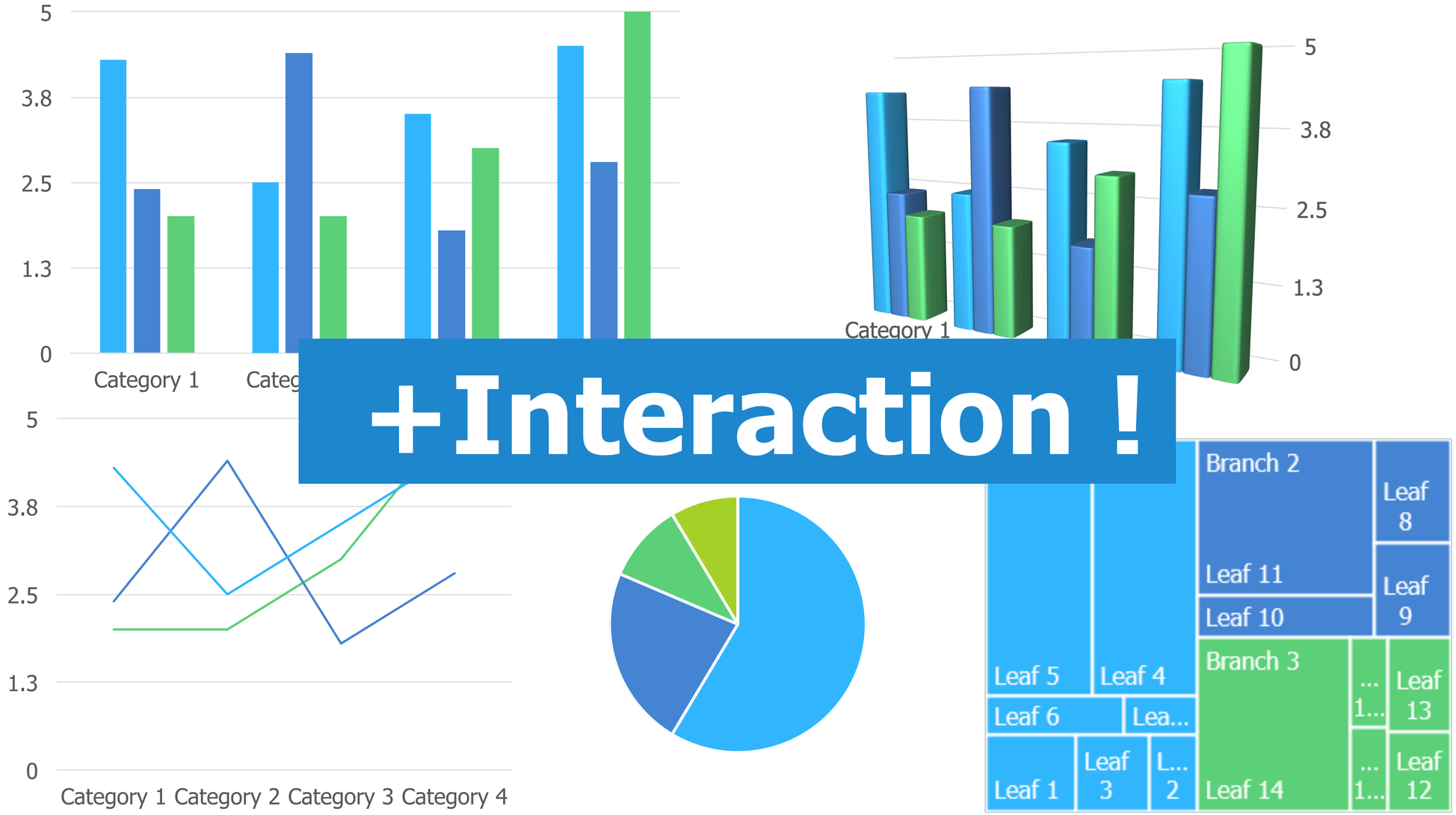
32 — Brick size

Interpolation:

- (○) Nearest Neighbour
- (●) Linear

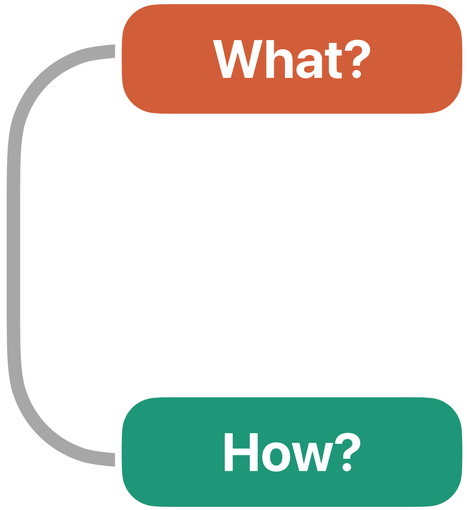
A lot of options – Design space is huge

A lot of options – Design space is huge

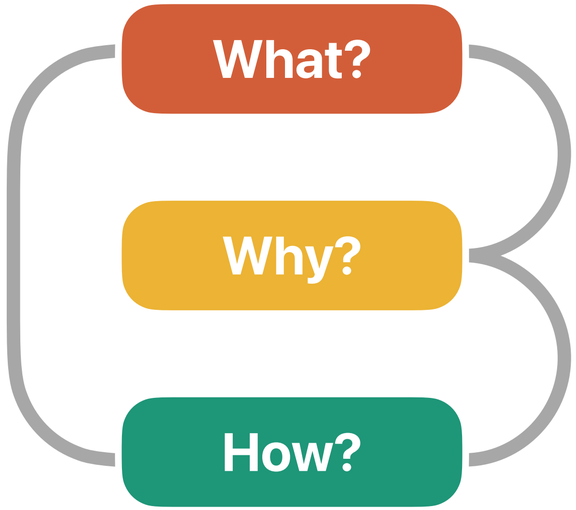


How?

... or building a common vocabulary



... or building a common vocabulary

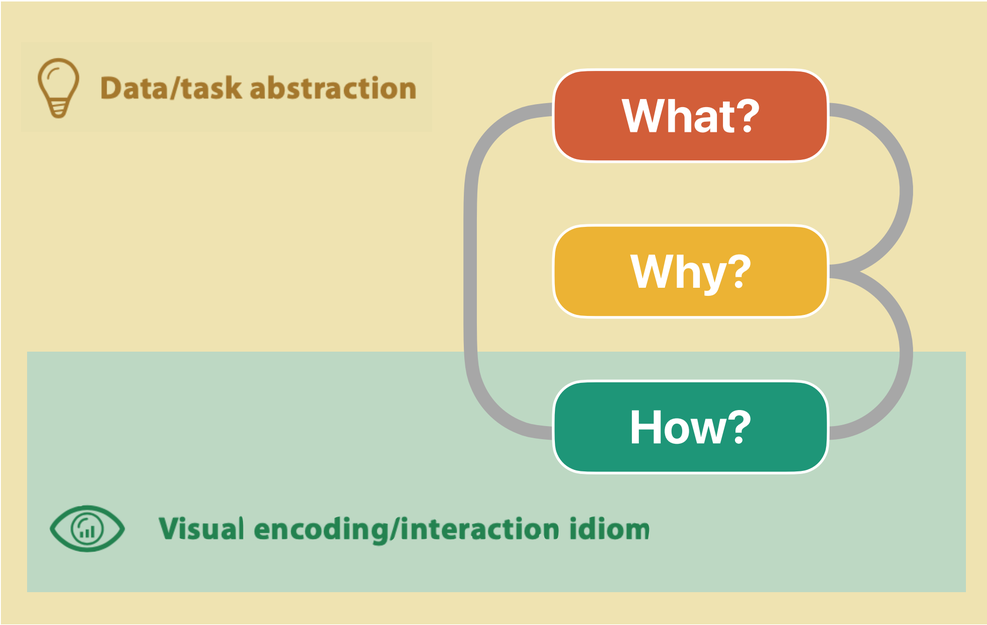


... or building a common vocabulary

Analysis framework

Abstraction

- translate from specifics of domain to vis vocabulary
  - **what** is shown? **data** abstraction
  - **why** is the user looking at it? **task** abstraction



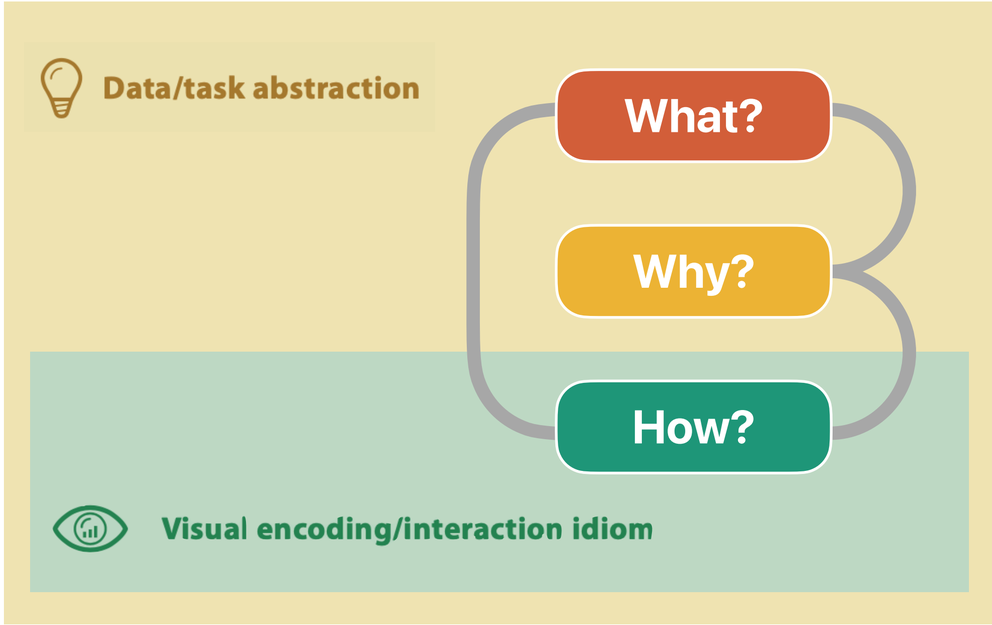
Analysis framework

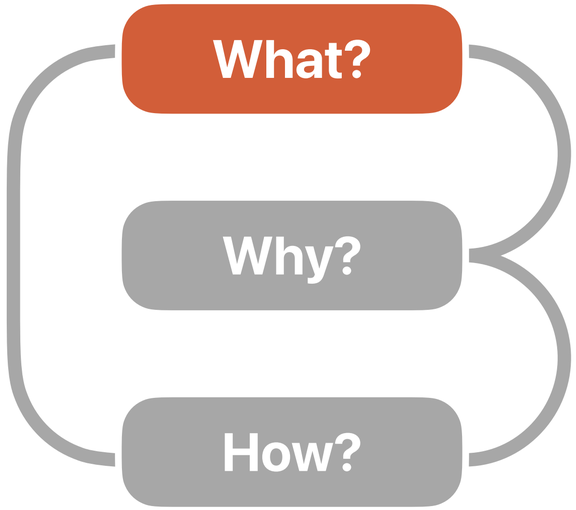
Abstraction

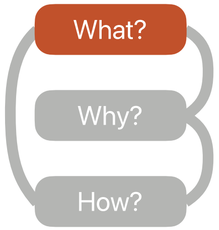
- translate from specifics of domain to vis vocabulary
  - **what** is shown? **data** abstraction
  - **why** is the user looking at it? **task** abstraction

Idiom

- **how** is it shown?
  - **visual encoding** idiom: how to draw
  - **interaction** idiom: how to manipulate

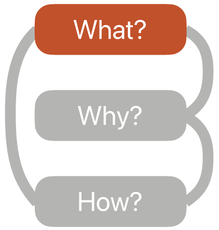



What?
Why?
How?



What Does Data Mean?

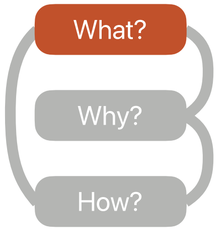
```
14, 2.6, 30, 30, 15, 100001
```



What Does Data Mean?

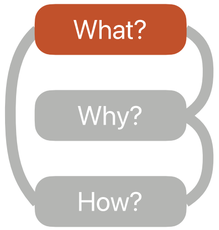


```
Olive, 7, S, Apple
```



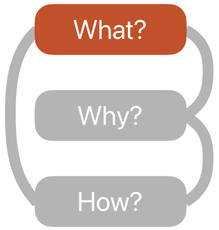
What Does Data Mean?

Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee



What Does Data Mean?

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee

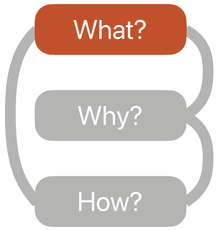


What Does Data Mean?



- semantics:
  - real-world meaning

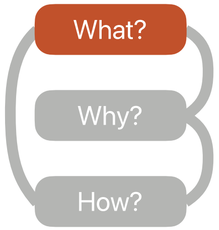
Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee



What Does Data Mean?

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

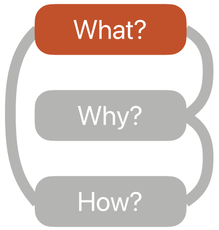
Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee



What Does Data Mean?

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

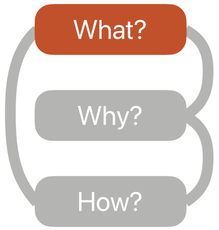
Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee



Items & Attributes

- item:
  - individual entity, discrete
  - “independent variable”

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee

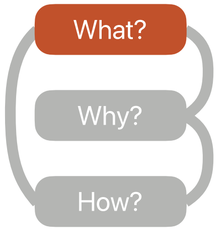


Items & Attributes

- item:
  - individual entity, discrete
  - “independent variable”

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee

item: person

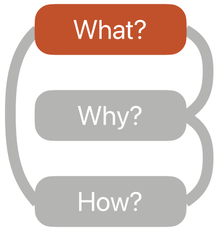


Items & Attributes

- item:
  - individual entity, discrete
  - “independent variable”
- attribute:
  - property that is measured, observed, logged…
  - “dependent variable”

attributes: name, age, shirt size, fave fruit

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee



Dataset Types

Tables

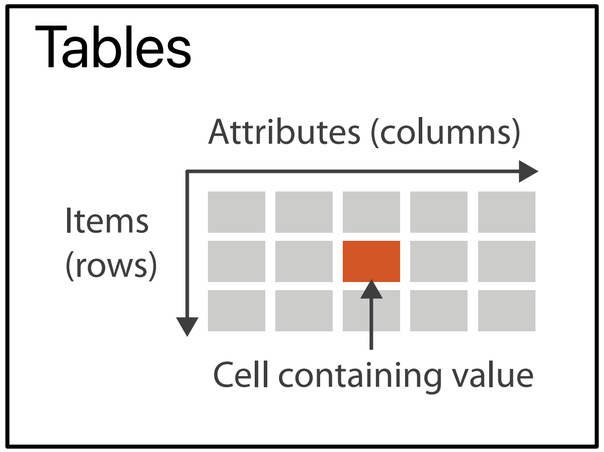
Items

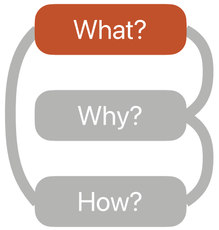
Attributes

Flat table

- one item per row
- each column is an attribute
- cell: value for item-attribute pair

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee





Dataset Types

Tables

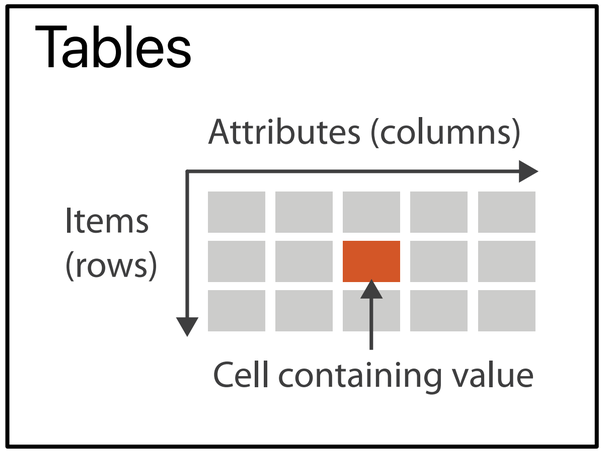
Items

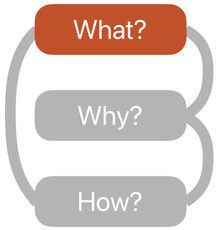
Attributes

Flat table

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

Name	Age	Shirt Size	Favorite Fruit
Pavlov	5	M	Banana
Amy	8	M	Pear
Olive	7	S	Apple
Frans	12	L	Orange
Jette	9	S	Raspberry
Huang	10	M	Orange
Abdul	8	L	Lychee





Dataset Types

Tables

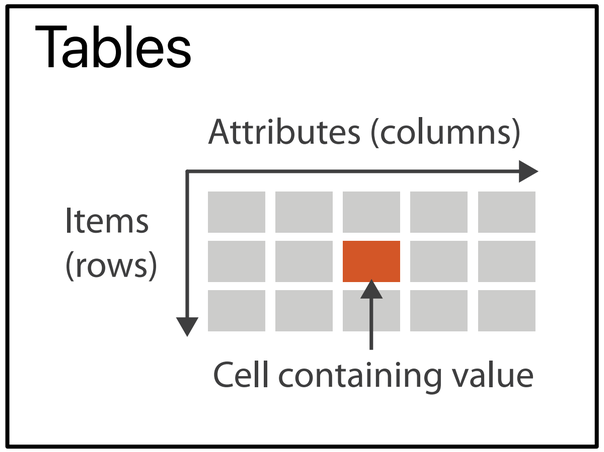
Items

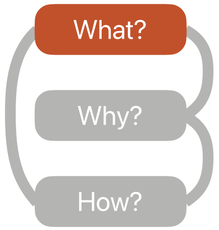
Attributes

Flat table

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

ID	Name	Age	Shirt Size	Favorite Fruit
1	Pavlov	5	M	Banana
2	Amy	8	M	Pear
3	Olive	7	S	Apple
4	Frans	12	L	Orange
5	Jette	9	S	Raspberry
6	Huang	10	M	Orange
7	Abdul	8	L	Lychee





Dataset Types

Tables

Items

Attributes

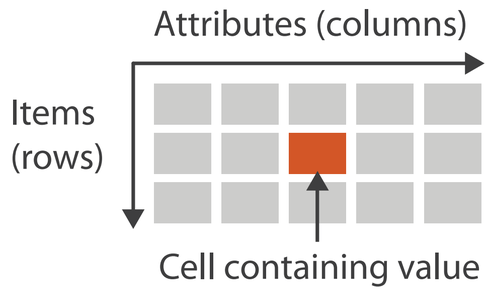
Multi-dimensional Table

- indexing based on multiple keys

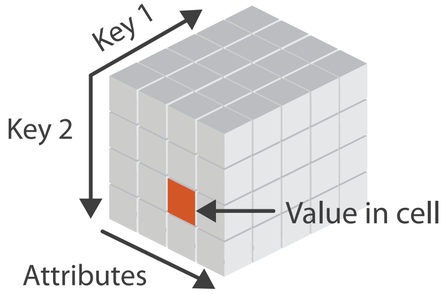
  e.g., genes + patients

	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

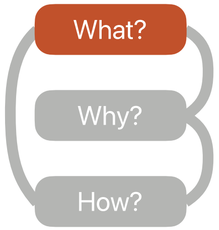
Tables



Multi-dimensional Table



TU Delft



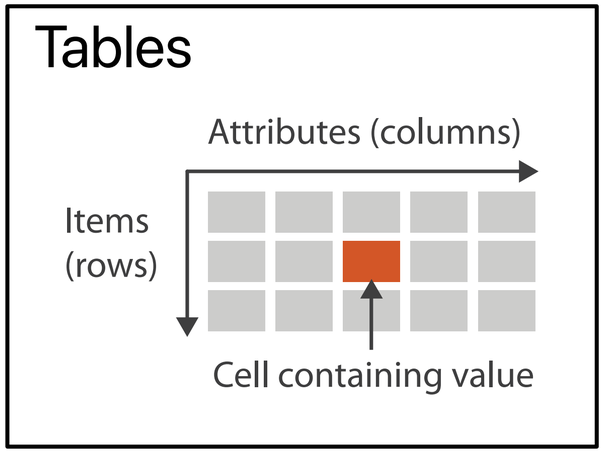
Dataset Types

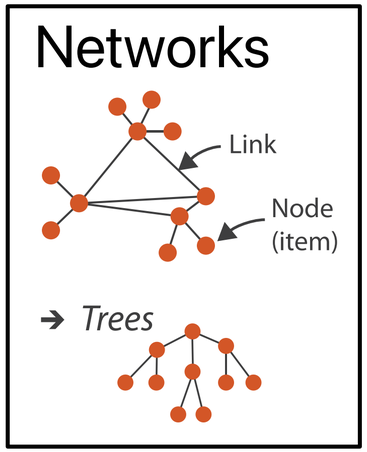
Tables
Items
Attributes

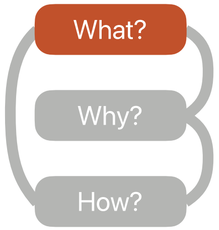
Networks & Trees
Items (nodes)
Links
Attributes

network/graph

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



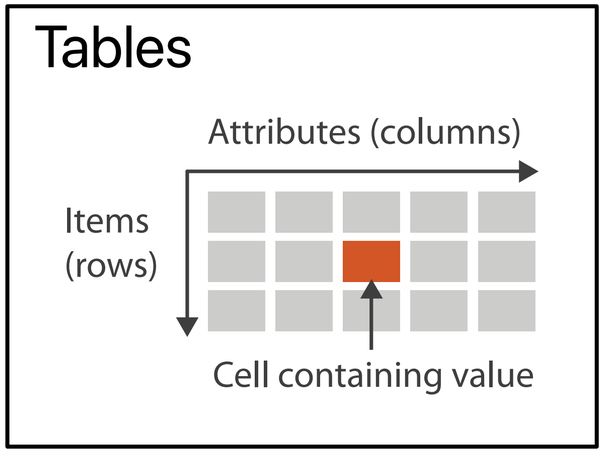


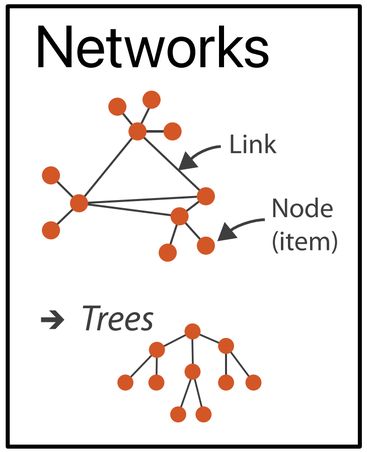


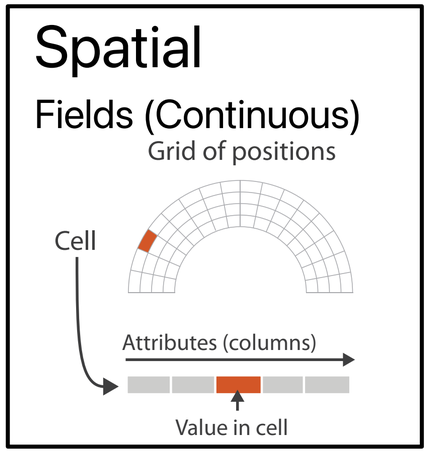
Dataset Types

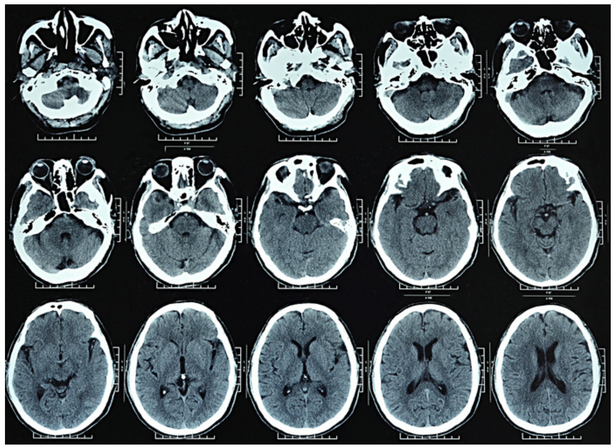
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes

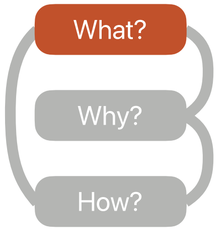
- attribute values associated with cells
- cell contains value from continuous domain
  - e.g., temperature, pressure, wind velocity
- measured or simulated







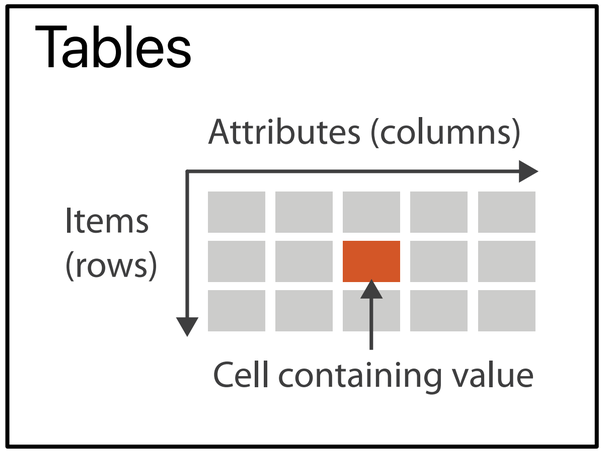


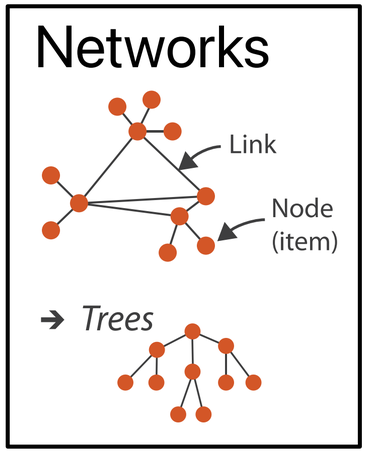


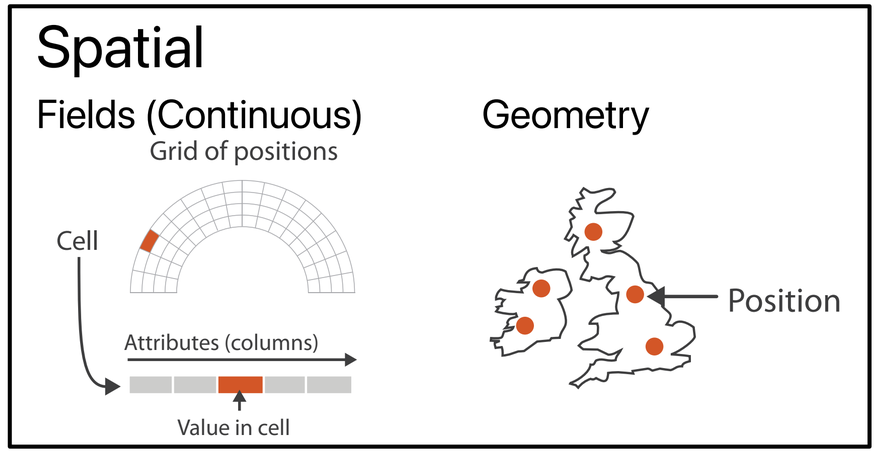
Dataset Types

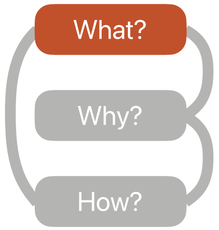
Tables	Networks & Trees	Fields	Geometry
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions
	Attributes	Attributes	

- shape of items
- explicit spatial positions / regions
  - points, lines, curves, surfaces, volumes









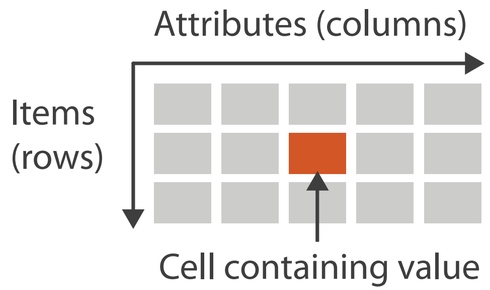
Dataset Types

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

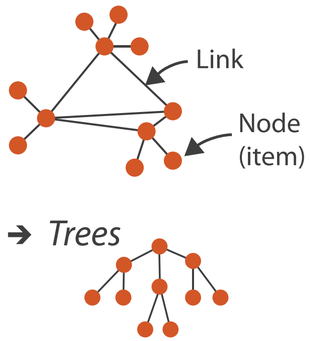
How we group items?

- sets: unique items, unordered
- lists: ordered, duplicates possible
- clusters: groups of similar items

Tables

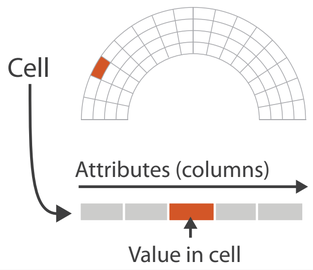


Networks



Spatial

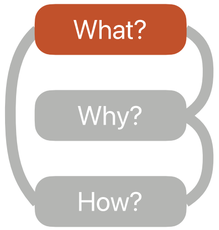
Fields (Continuous)



Geometry



TU Delft



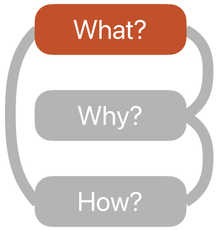
Dataset and Data Types

Data and Dataset Types

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

Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids



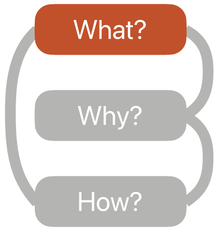
Dataset and Data Types

Data and Dataset Types

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

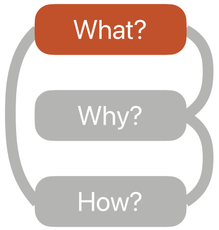
Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids



Attribute Types

Which classes of values & measurements?



Attribute Types

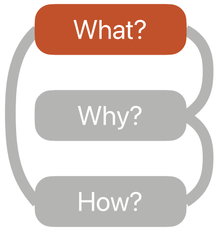
Which classes of values & measurements?

- categorical (nominal)
  - compare equality
  - no implicit ordering

Attribute Types

➔ Categorical





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

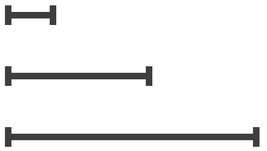


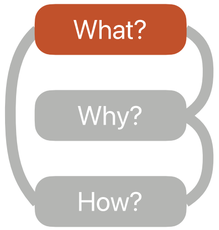
➔ Ordered

➔ *Ordinal*

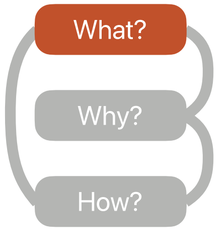


➔ *Quantitative*





ID	Name	Age	Shirt Size	Favorite Fruit
0	Pavlov	5	M	Banana
1	Amy	8	M	Pear
2	Olive	7	S	Apple
3	Frans	12	L	Orange
4	Jette	9	S	Raspberry
5	Huang	10	M	Orange
6	Abdul	8	L	Lychee

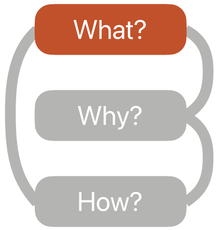


ID	Name	Age	Shirt Size	Favorite Fruit
0	Pavlov	5	M	Banana
1	Amy	8	M	Pear
2	Olive	7	S	Apple
3	Frans	12	L	Orange
4	Jette	9	S	Raspberry
5	Huang	10	M	Orange
6	Abdul	8	L	Lychee

categorical

ordinal

quantitative



Other Data Concerns

Attribute Types

- Categorical



- Ordered
  - *Ordinal*



  - *Quantitative*



Ordering Direction

- Sequential



- Diverging



- Cyclic



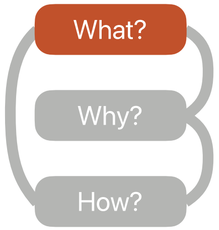
Dataset Availability

- Static



- Dynamic



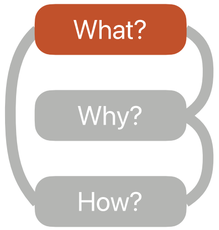


Data Abstraction

- translate domain-specific language to generic visualization language

- identify dataset & attribute type(s)
- identify cardinality
  - dataset: how many items
  - attribute: number of levels for categorical data / range for quantitative data

- consider whether to transform data
  - guided by understanding of task / goal



Data Vs Conceptual Model

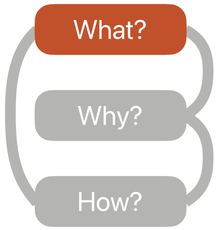
data model

- mathematical abstraction
  - sets with operations, e.g., floats with * / - +
  - variable data types in programming languages

conceptual model

- mental construction (semantics)
- supports reasoning
- typically based on understanding of tasks and goals

data abstraction process relies on conceptual model!



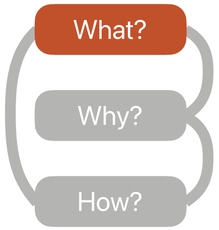
Data vs Conceptual Model, Example

data model: floats

```
32.52, 54.06, −14.35, ...
```

conceptual model

temperature



Data vs Conceptual Model, Example

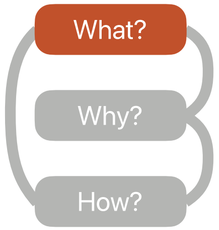
data model: floats

```
32.52, 54.06, −14.35, ...
```

conceptual model

temperature

multiple possible data abstractions



Data vs Conceptual Model, Example

data model: floats

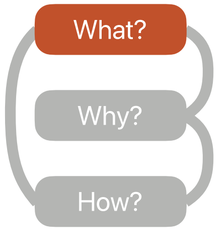
```
32.52, 54.06, −14.35, ...
```

conceptual model

temperature

multiple possible data abstractions

goal: forecasting the weather



Data vs Conceptual Model, Example

data model: floats

```
32.52, 54.06, −14.35, ...
```

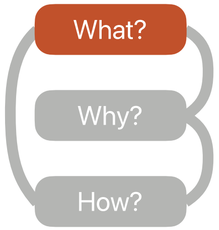
conceptual model

temperature

multiple possible data abstractions

goal: forecasting the weather

➤ continuous to 2 significant digits: *quantitative*



Data vs Conceptual Model, Example

data model: floats

`32.52, 54.06, −14.35, ...`

conceptual model

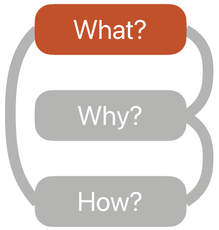
temperature

multiple possible data abstractions

goal: forecasting the weather

goal: decide if bath water is ready

➤ continuous to 2 significant digits: *quantitative*



Data vs Conceptual Model, Example

data model: floats

`32.52, 54.06, −14.35, ...`

conceptual model

temperature

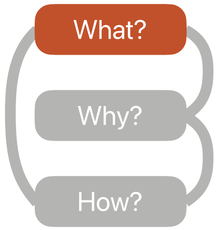
multiple possible data abstractions

goal: forecasting the weather

➤ continuous to 2 significant digits: *quantitative*

goal: decide if bath water is ready

➤ hot, warm, cold: *ordinal*



Data vs Conceptual Model, Example

data model: floats

`32.52, 54.06, −14.35, ...`

conceptual model

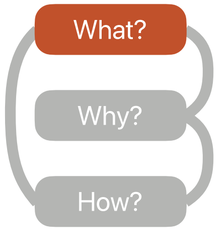
temperature

multiple possible data abstractions

goal: forecasting the weather ➤ continuous to 2 significant digits: *quantitative*

goal: decide if bath water is ready ➤ hot, warm, cold: *ordinal*

goal: decide what to wear today



Data vs Conceptual Model, Example

data model: floats

```
32.52, 54.06, −14.35, ...
```

conceptual model

temperature

multiple possible data abstractions

goal: forecasting the weather ➤ continuous to 2 significant digits: *quantitative*

goal: decide if bath water is ready ➤ hot, warm, cold: *ordinal*

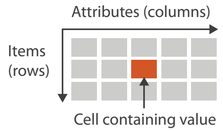
goal: decide what to wear today ➤ sweater temp, shorts temp, ...: *categorical*

What?

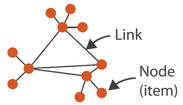
Datasets

Dataset Types

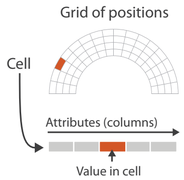
- Tables



- Networks



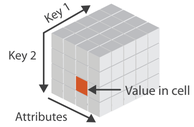
- Fields (Continuous)



- Geometry (Spatial)



- *Multidimensional Table*



- *Trees*



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		

Attributes

Attribute Types

- Categorical



- Ordered
  - *Ordinal*



  - *Quantitative*



Ordering Direction

- Sequential



- Diverging

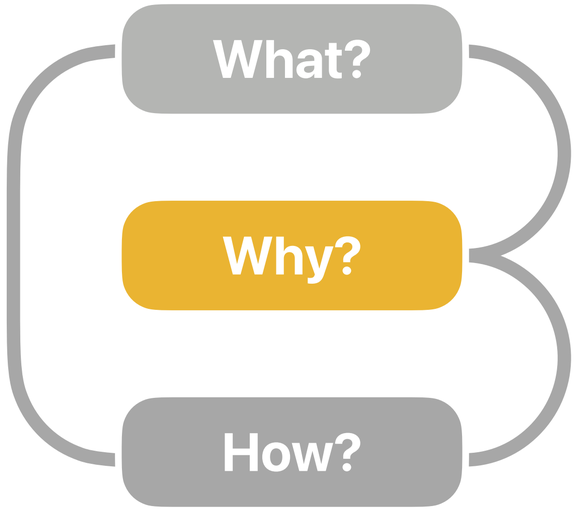


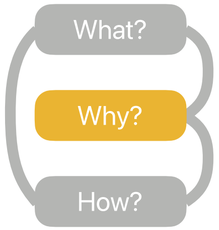
- Cyclic





Questions!


What?
Why?
How?

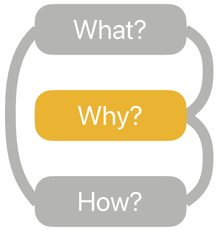


Task Abstraction

Translate domain questions and goals to abstract tasks

very high-level pattern: {action, target} pairs

- actions
  - analyze: high-level choices
  - search: find a known/unknown item
  - query: find out about characteristics of item

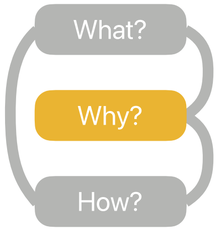


Task Abstraction

Translate domain questions and goals to abstract tasks

very high-level pattern: {action, target} pairs

- actions
  - analyze: high-level choices
  - search: find a known/unknown item
  - query: find out about characteristics of item
- targets
  - what is being acted on



Task Abstraction

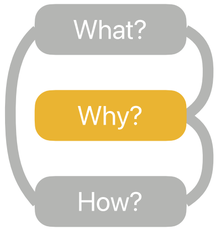
Translate domain questions and goals to abstract tasks

very high-level pattern: {action, target} pairs

- actions
  - analyze: high-level choices
  - search: find a known/unknown item
  - query: find out about characteristics of item
- targets
  - what is being acted on

example {action, target} pairs

- discover distribution
- compare trends
- locate outliers
- browse topology



Actions: Analyze

consume

- discover vs present
  - aka explore vs explain
- enjoy
  - newcomer
  - aka casual, social

Analyze

→ Consume

→ *Discover*

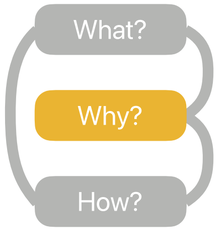


→ *Present*



→ *Enjoy*





Actions: Analyze

consume

- discover vs present
  - aka explore vs explain
- enjoy
  - newcomer
  - aka casual, social

produce

- annotate, record, derive
  - crucial design choice

Analyze

➔ Consume

➔ *Discover*



➔ *Present*

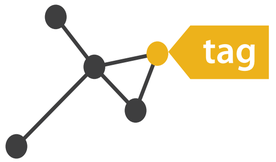


➔ *Enjoy*



➔ Produce

➔ *Annotate*

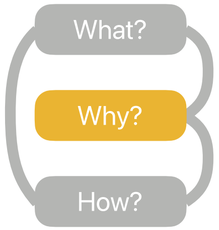


➔ *Record*



➔ *Derive*





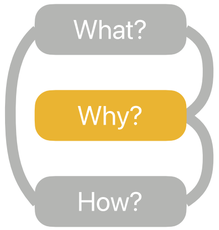
Actions: Search, Query

what does user know?

- target, location

Search

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



Actions: Search, Query

what does user know?

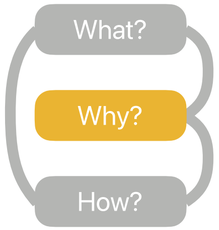
- target, location

how much of the data matters?

- one, some, all

Search

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



Targets

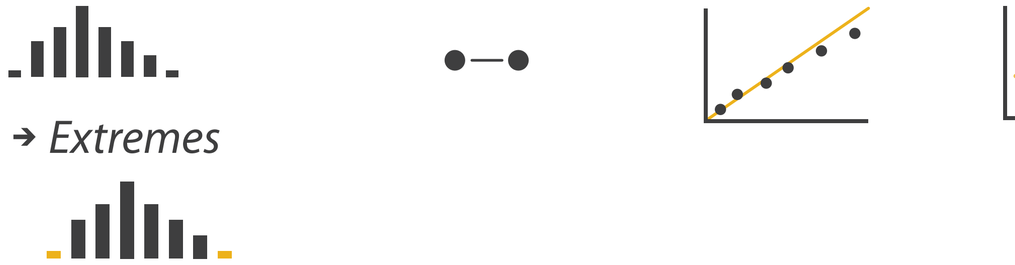
All Data

- Trends
- Outliers
- Features



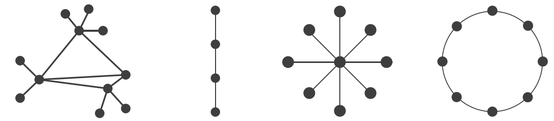
Attributes

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



Network Data

- Topology

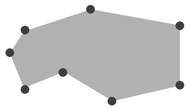


- *Paths*



Spatial Data

- Shape



Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*



- Produce
  - *Annotate*
  - *Record*
  - *Derive*

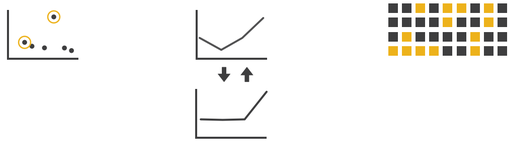


Search

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

Query

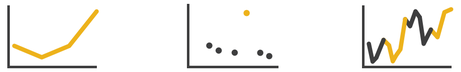
- Identify
- Compare
- Summarize



Targets

All Data

- Trends
- Outliers
- Features



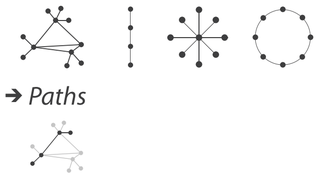
Attributes

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



Network Data

- Topology
- *Paths*



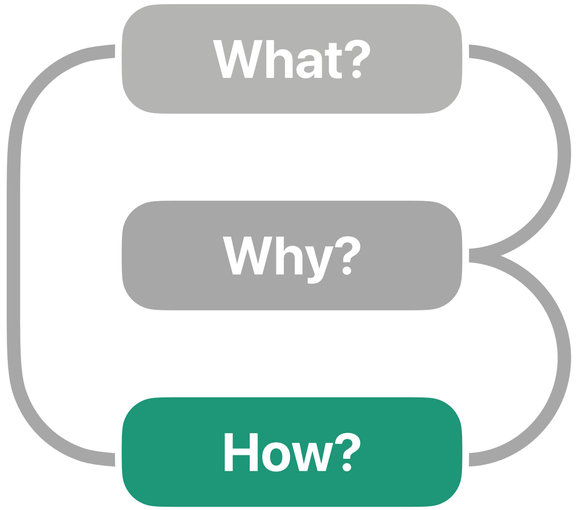
Spatial Data

- Shape





Questions!


What?
Why?
How?

Definitions: Marks and channels

- marks
  - geometric primitives

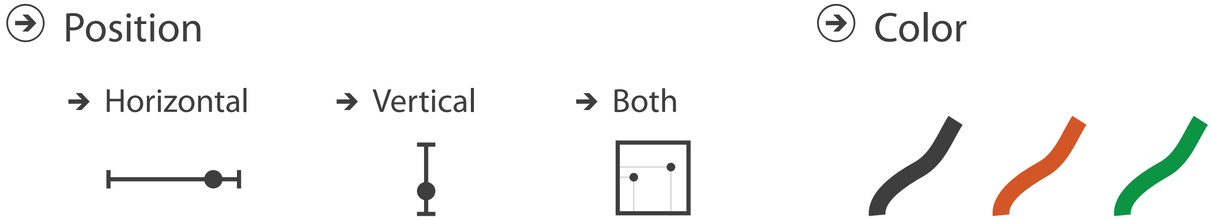


Definitions: Marks and channels

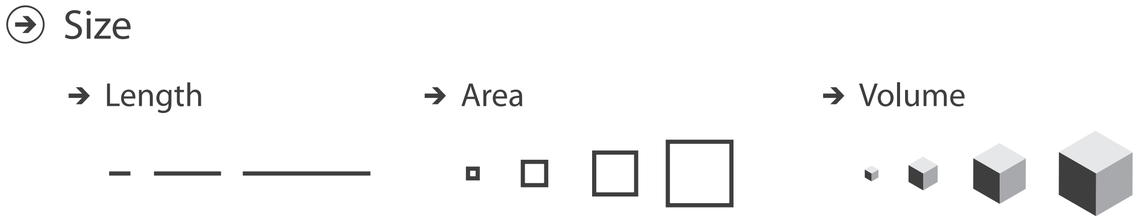
- marks
  - geometric primitives



- channels
  - control appearance of marks
  - can redundantly code with multiple channels

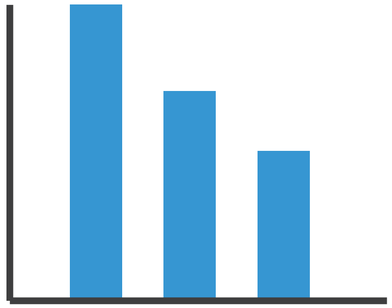






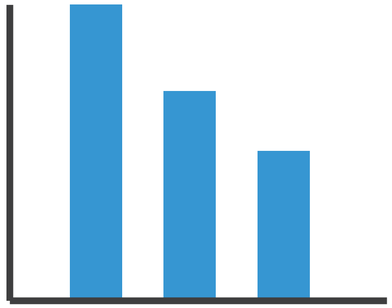
Visual encoding

analyze idiom structure ⇢ as combination of marks and channels



Visual encoding

analyze idiom structure ⇢ as combination of marks and channels

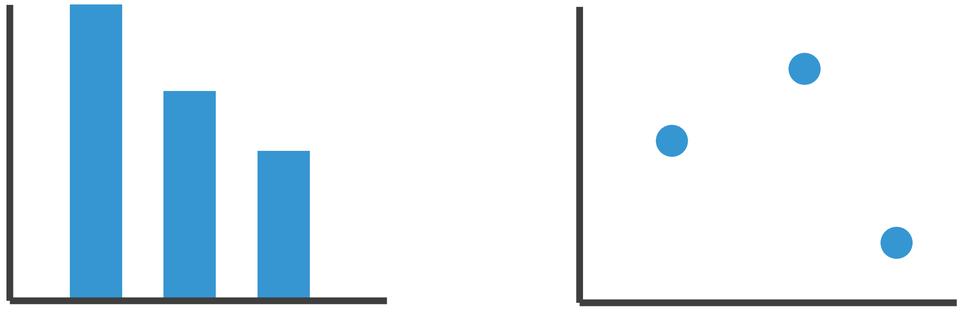


length

mark: line

Visual encoding

analyze idiom structure ⇢ as combination of marks and channels

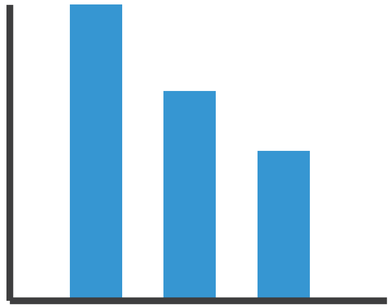


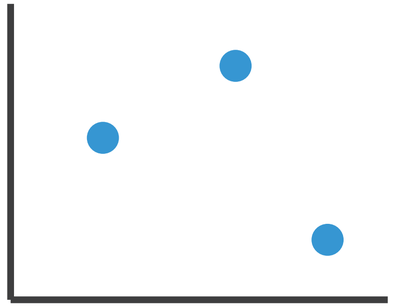
length

mark: line

Visual encoding

analyze idiom structure ⇢ as combination of marks and channels





length

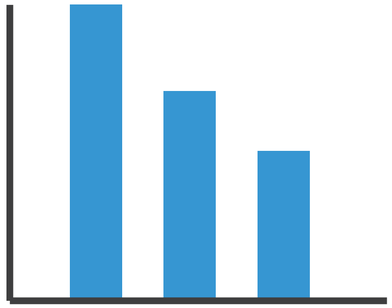
vertical position
horizontal position

mark: line

mark: point

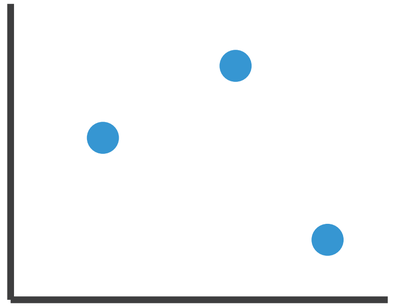
Visual encoding

analyze idiom structure ⇢ as combination of marks and channels



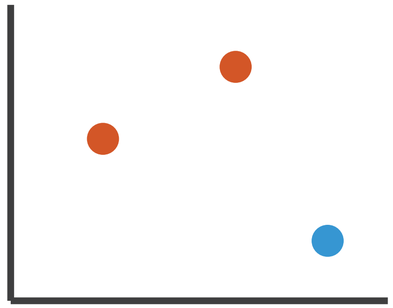
length

mark: line



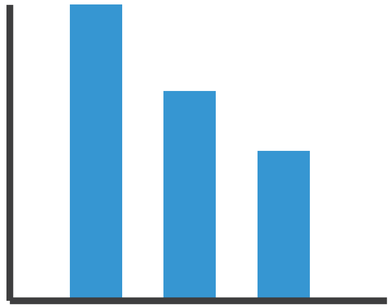
vertical position
horizontal position

mark: point



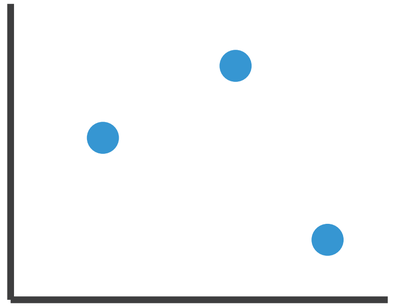
Visual encoding

analyze idiom structure ⇢ as combination of marks and channels



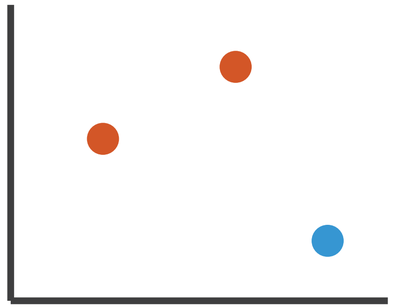
length

mark: line



vertical position
horizontal position

mark: point

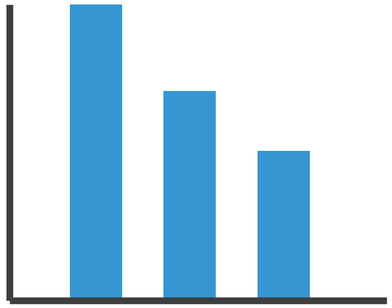


vertical position
horizontal position
color hue

mark: point

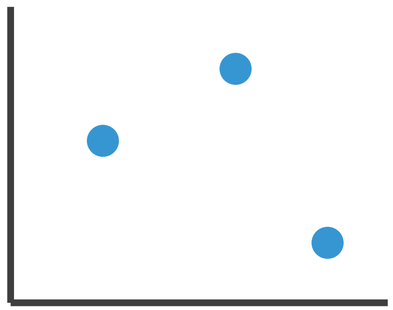
Visual encoding

analyze idiom structure ⇢ as combination of marks and channels



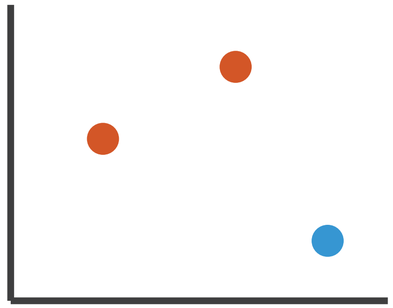
length

mark: line



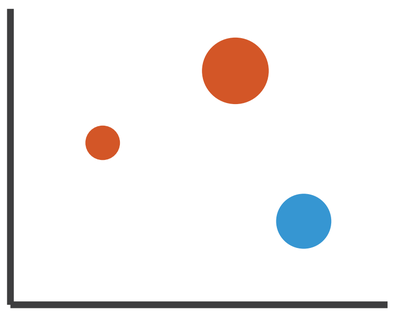
vertical position
horizontal position

mark: point



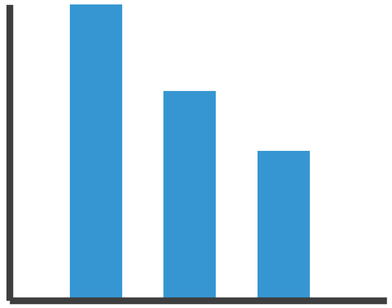
vertical position
horizontal position
color hue

mark: point



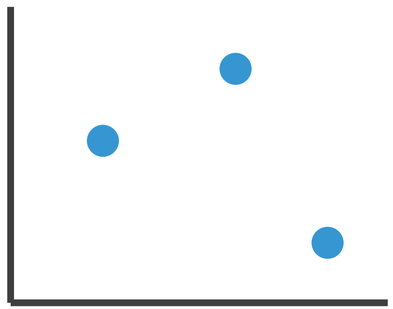
Visual encoding

analyze idiom structure ⇢ as combination of marks and channels



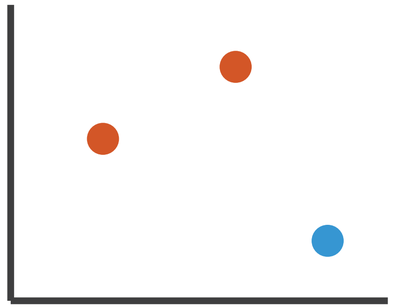
length

mark: line



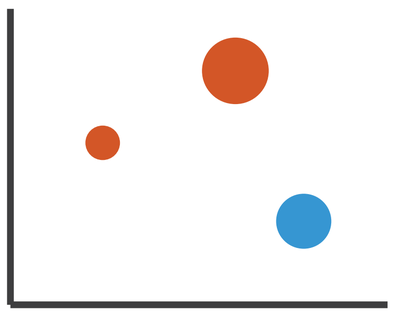
vertical position
horizontal position

mark: point



vertical position
horizontal position
color hue

mark: point



vertical position
horizontal position
color hue
size (area)

mark: point

Channels

Position on common scale

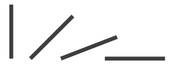


Position on unaligned scale



Length (1D size)

Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



Spatial region



Color hue



Motion



Shape



Channels: Matching Types

Magnitude Channels: Ordered Attributes

Position on common scale

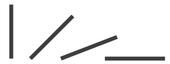


Position on unaligned scale



Length (1D size)

Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



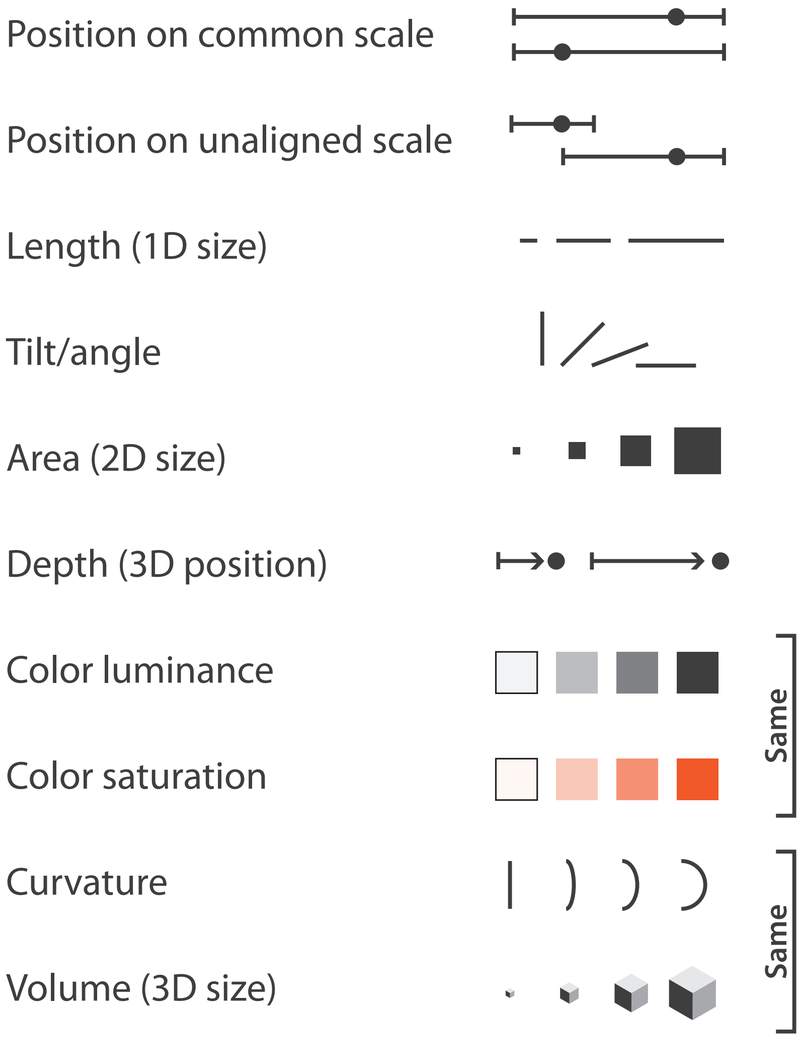
Shape



- expressiveness principle
  - –match channel and data characteristics

Channels: Rankings

Magnitude Channels: Ordered Attributes



Identity Channels: Categorical Attributes



- expressiveness principle
  - –match channel and data characteristics
- effectiveness principle
  - –encode most important attributes with highest ranked channels

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 principle
  - –match channel and data characteristics
- effectiveness principle
  - –encode most important attributes with highest ranked channels

✦ spatial position ranks high for both

Channels: Expressiveness types and effectiveness rankings

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D p

Color luminance

Color saturation

Curvature

Volume (3D size)

Best

Same

Same

Least

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

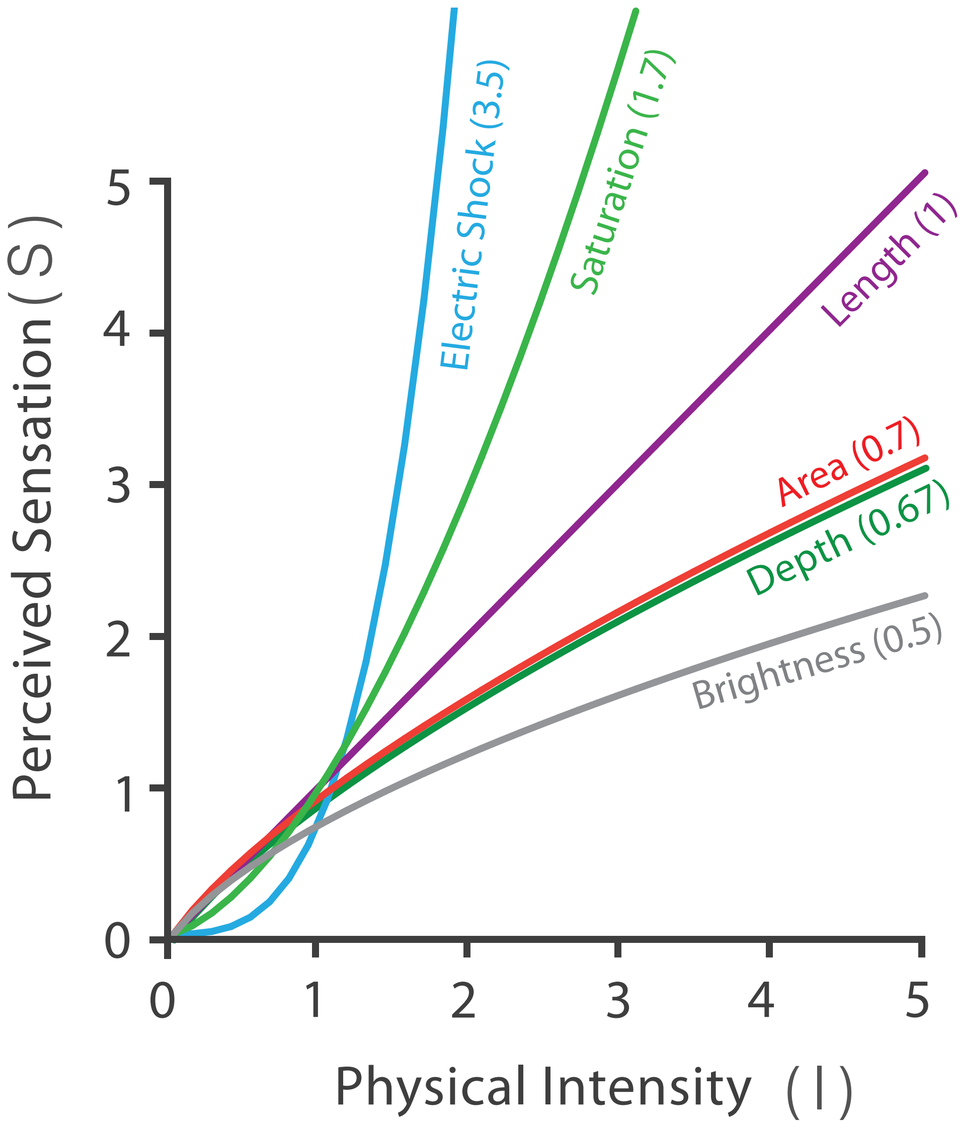
Where does this order come from?

- effectiveness principle
  - –encode most important attributes with highest ranked channels

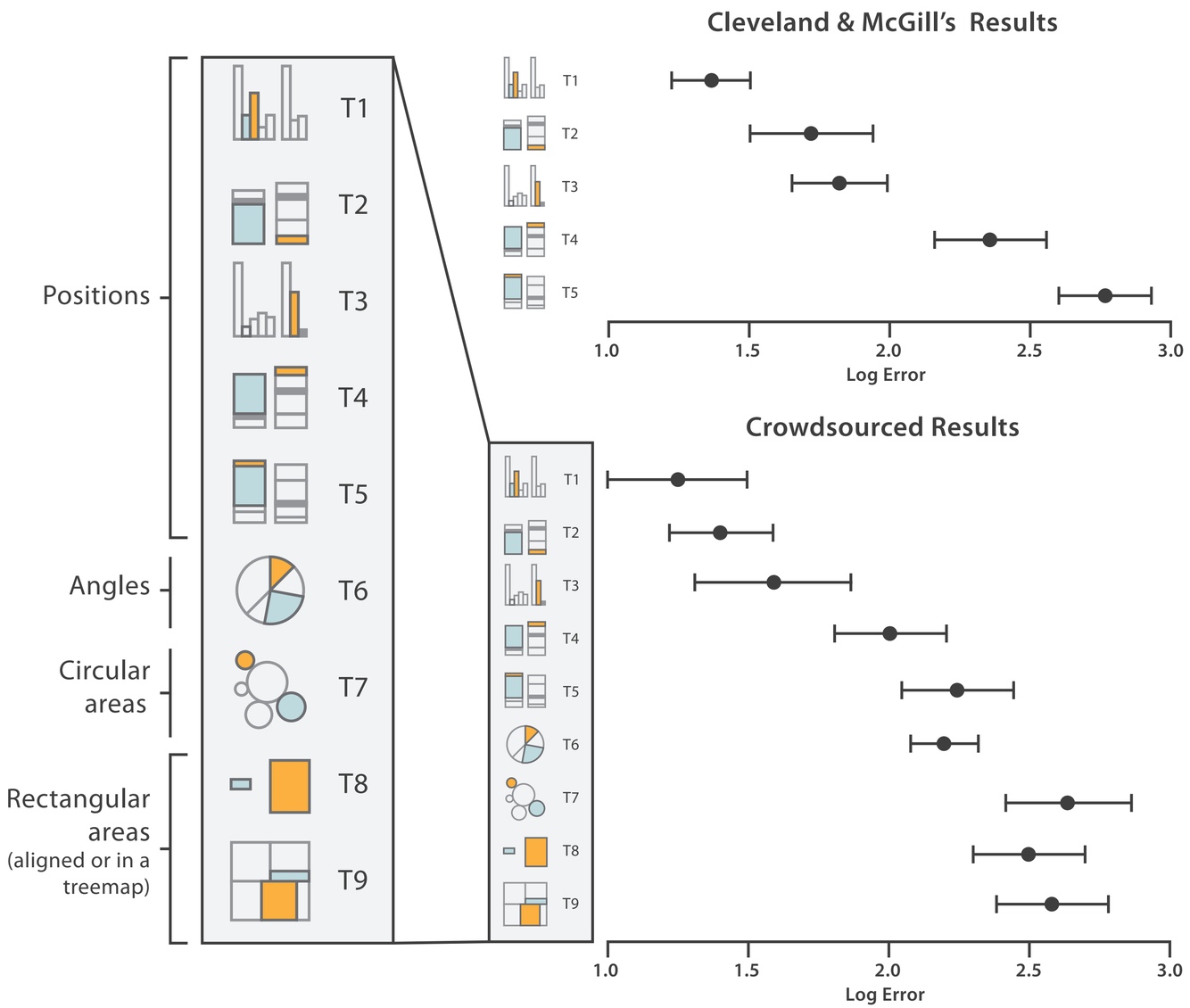
✦ spatial position ranks high for both

Accuracy: Fundamental Theory

Steven's Psychophysical Power Law: $S = I^N$



Accuracy: Vis experiments

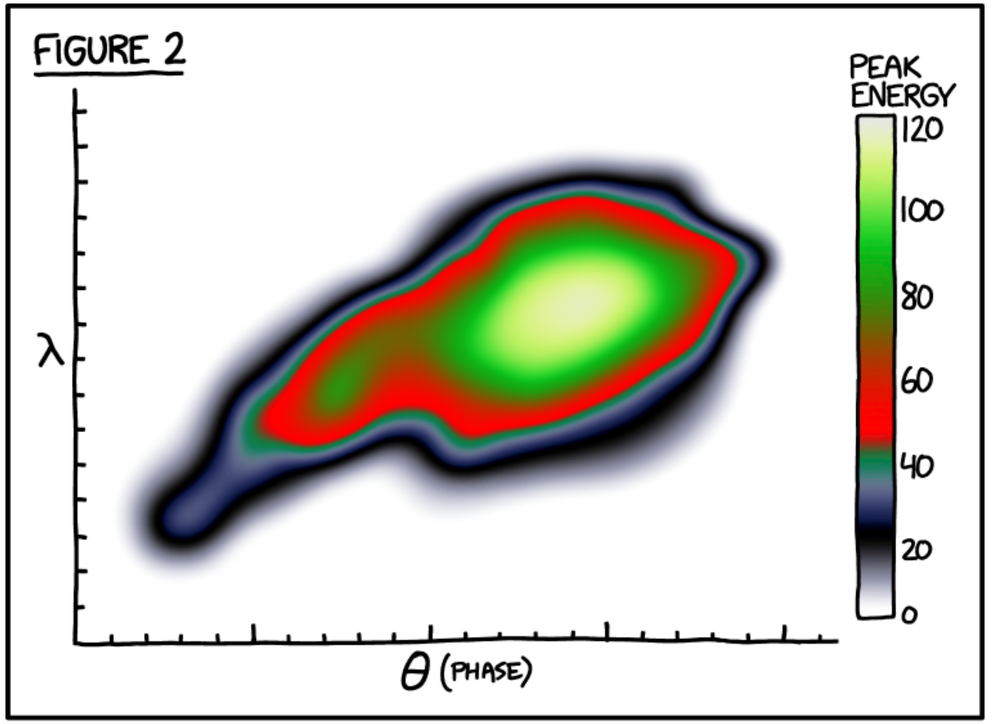


Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Cleveland and McGill, Journal of the American Statistical Association, 1984

Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010

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

Color as a Channel



EVERY YEAR, DISGRUNTLED SCIENTISTS COMPETE FOR THE PAINBOW AWARD FOR WORST COLOR SCALE.



first rule of color: do not talk about color!

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

Same

Same

Best

Effectiveness

least

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

Decomposing color

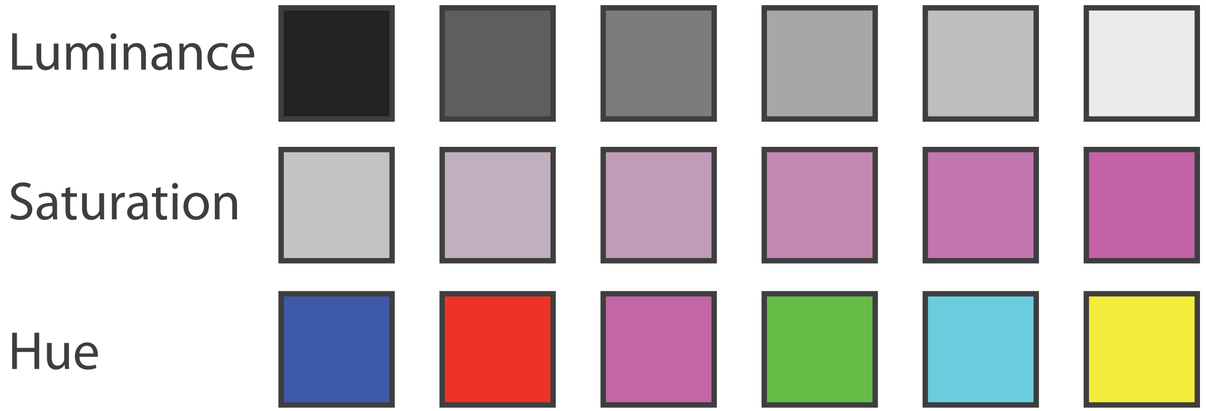


- color is confusing if treated as monolithic

Decomposing color

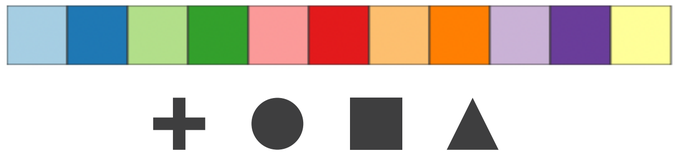
- color is confusing if treated as monolithic

- ordered can show magnitude
  - luminance: how bright
  - saturation: how colorful
- categorical can show identity
  - hue: what color

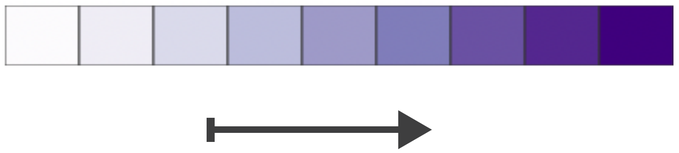


Attribute Types

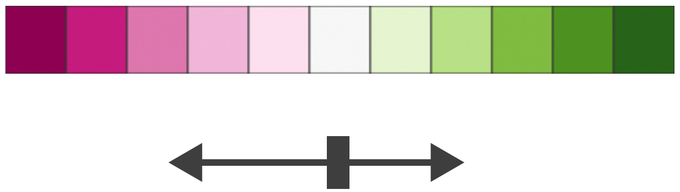
- Categorical



- Ordered
  - Sequential

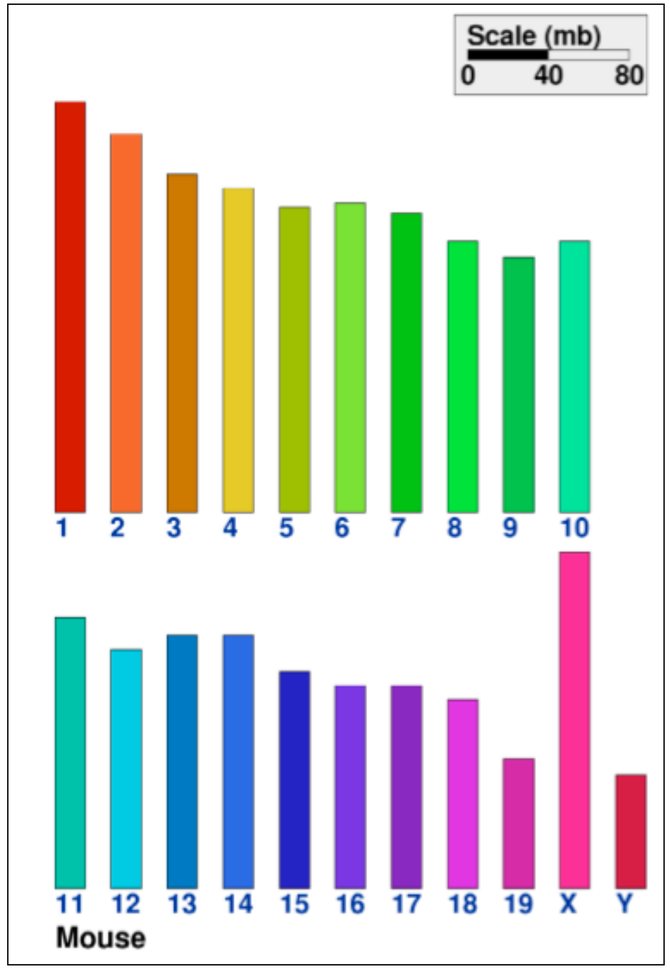


  - Diverging



Categorical color: limited number of discriminable bins

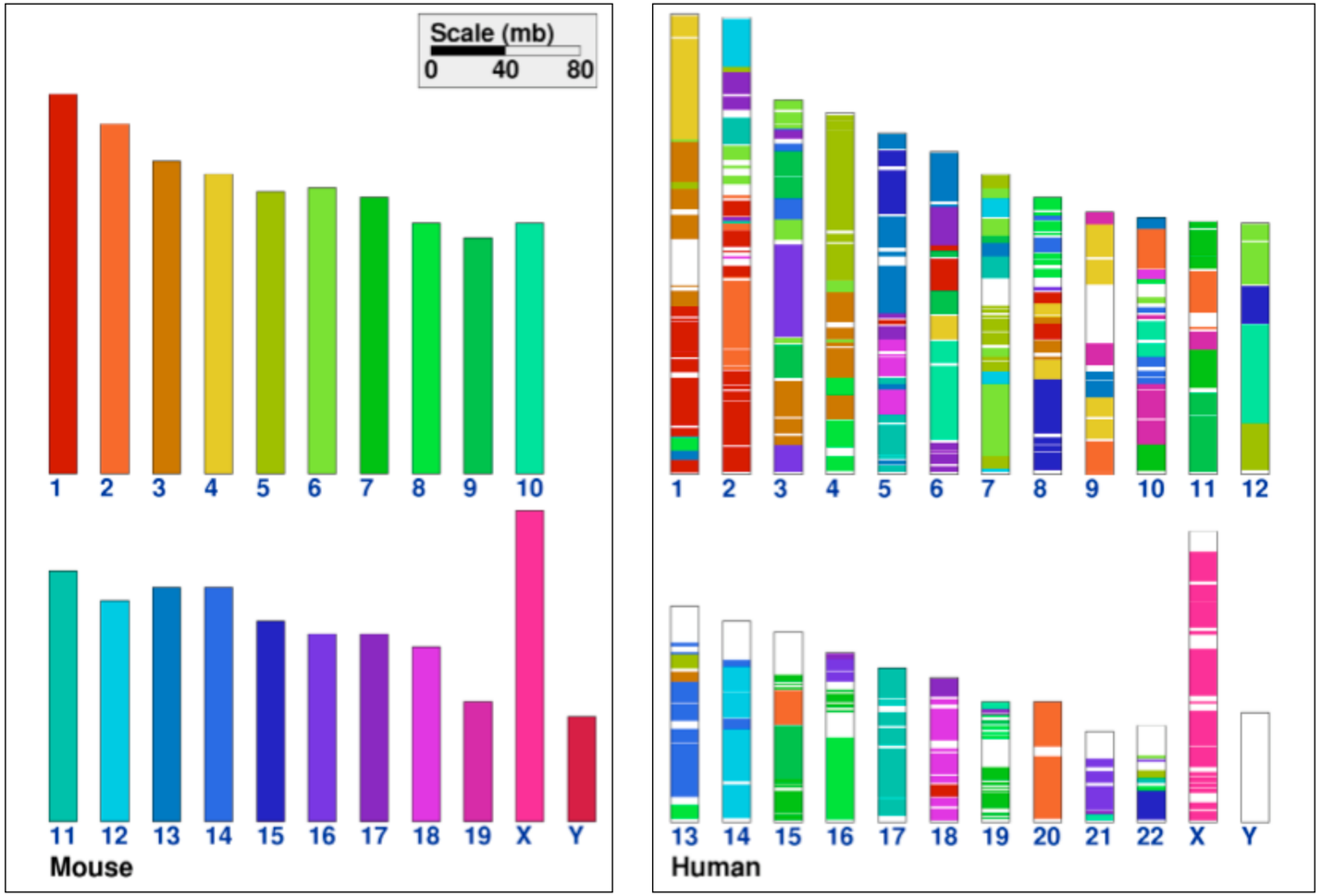
- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons



Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics

Categorical color: limited number of discriminable bins

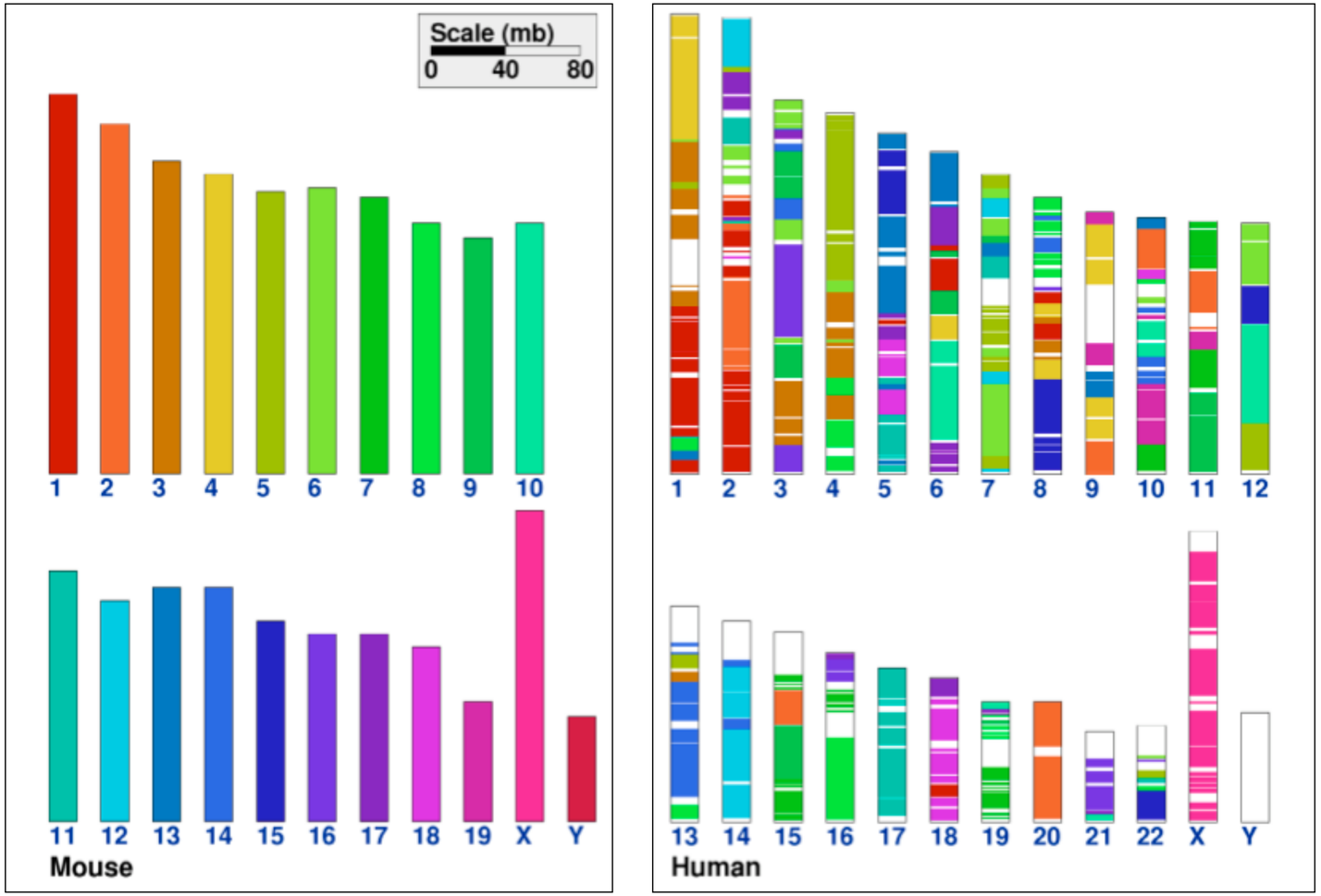
- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons



Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics

Categorical color: limited number of discriminable bins

- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
  - fewer bins than you want
  - rule of thumb: 6-12 bins, including background and highlights



Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics



Questions!

Slides (partially) based on lectures by

Anna Vilanova (TU Eindhoven)

Tamara Munzner (University of British Columbia)