Slides: **http://brosz.ca/slides/**

Mar 13, 2025

Data Visualization

JOHN BROSZ, PHD

LAND604

Slides: http://brosz.ca/slides/

Visualization Studio

34.5 million pixels

- Compare that to HD (2.07 million) or even 4K (8.3 million).

16’ (4.9m) wide and 6’ (1.8m) tall

- A standard IMAX screen is 22m X 16.1m.

Bookable by faculty & grad students at https://library.ucalgary.ca/services/visualization

Uses:

Analysis

Presentation

Teaching

Events

Slides at https://brosz.ca/slides

Topics

What is data visualization?

How do we visualize data / data encoding

- Visual variables

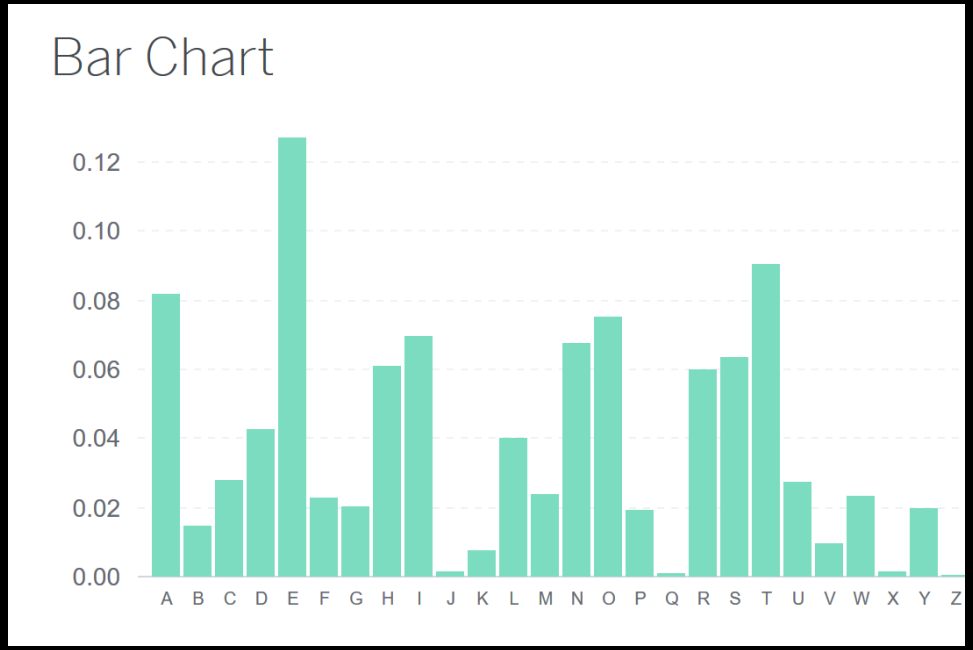
Data visualization tools

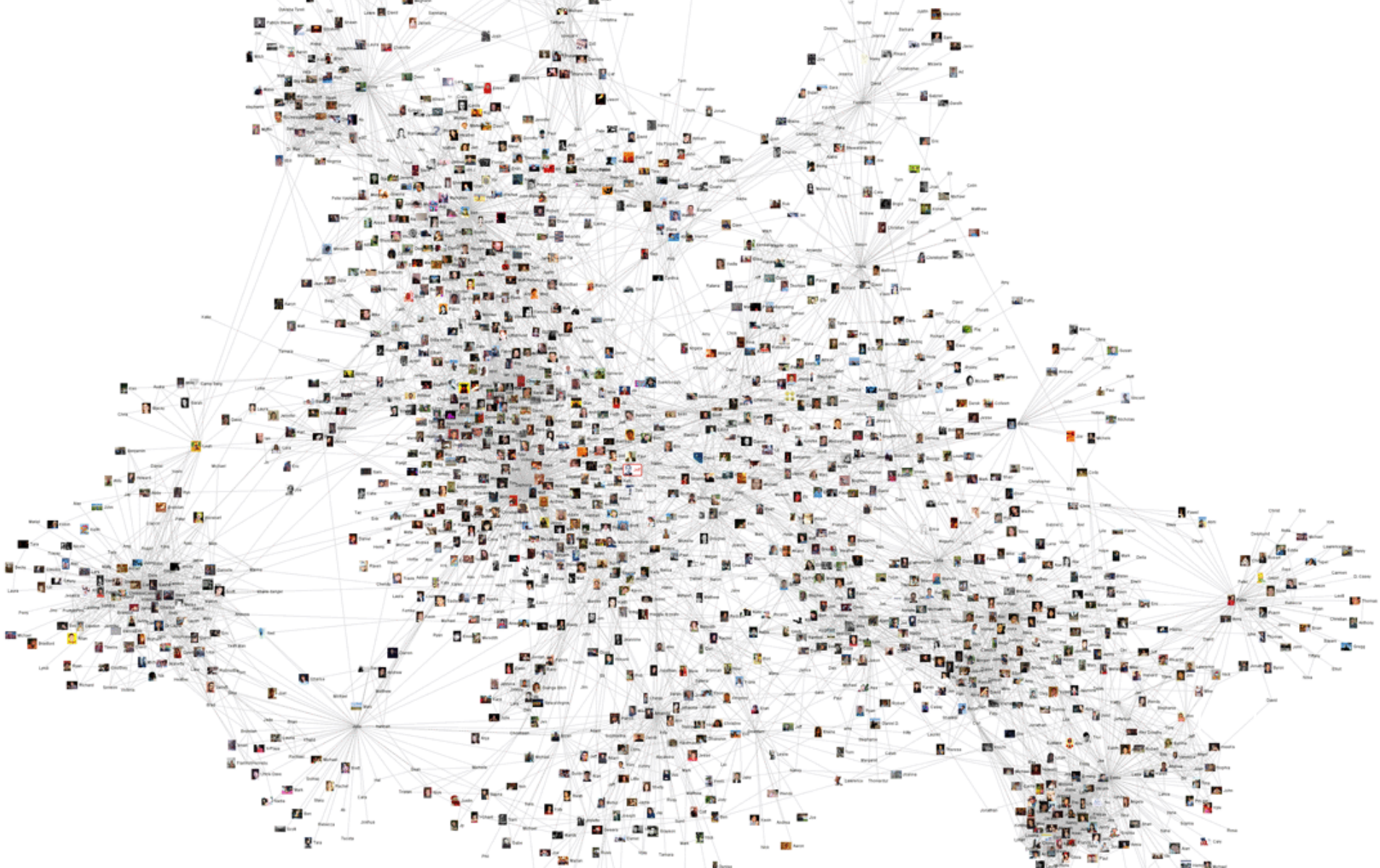
Resources (books & websites)

Slides at https://brosz.ca/slides

What Is Data Visualization?

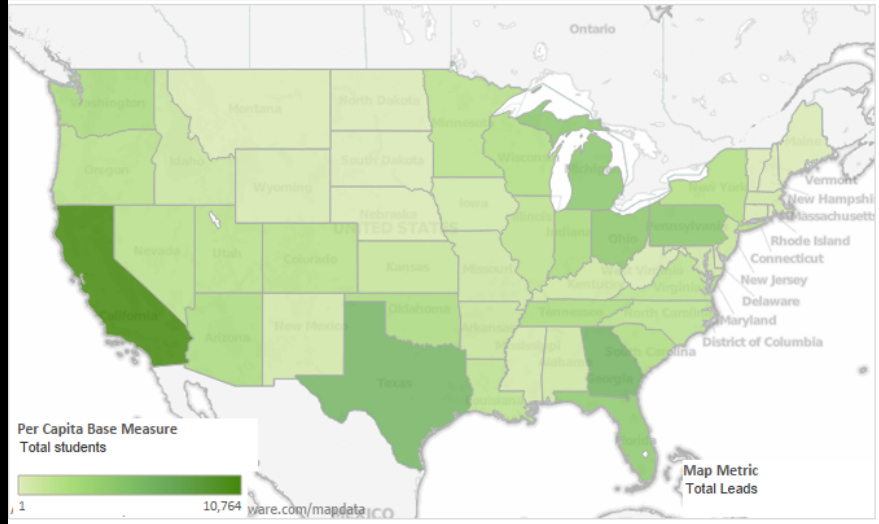
MORE THAN JUST MAKING CHARTS OR PRETTY PICTURES



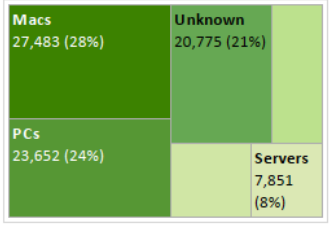


JEFF HEER

Leads by State

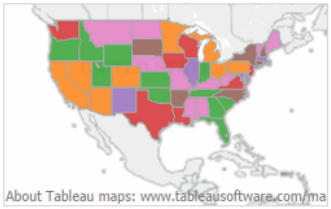


Product Area



Show Classifications Of

State Type



Response Time

Response Time	Convert %	Leads	Converted
< 2 Hrs	6.46%	5,310	343
< 1 Day	4.67%	9,556	446
Later	3.89%	84,134	3,270

Lead Volume Change

	Leads		WoW Change		YoY Change	
	2012	2013	2012	2013	2012	2013
1	4,475	1,933				-57%
2	3,249	1,645	-27%	-15%		-49%
3	1,714	2,035	-47%	24%		19%
4	1,322	4,854	-23%	139%		267%
5	1,476	2,743	12%	-43%		86%
6	5,300	2,643	259%	-4%		-50%
7	3,624	2,420	-32%	-8%		-33%
8	360	1,888	-90%	-22%		424%
9		1,051	-100%	-44%		
10		1,113		6%		
11	1,196	2,639		137%		121%
12	4,418	2,345	269%	-11%		-47%
13	3,990	2,904	-10%	24%		-27%
14	1,155	2,358	-71%	-19%		104%
15		1,809	-100%	-23%		
16		1,086		-40%		
17		1,193		10%		
18		2,941		147%		
19		2,889		-2%		
20		2,616		-9%		
21		3,358		28%		
22		2,554		-24%		
23		1,188		-53%		
24		1,326		12%		
25		2,515		90%		
26		2,411		-4%		
27		2,166		-10%		
28		2,494		15%		
29		1,742		-30%		

Summary

Lead Gen Budget	$3,226,785
Leads	99,000
Budget per Lead	$32.59
Converted	4,059
Budget per Conversion	$794.97
Convert %	4.10%

Filters

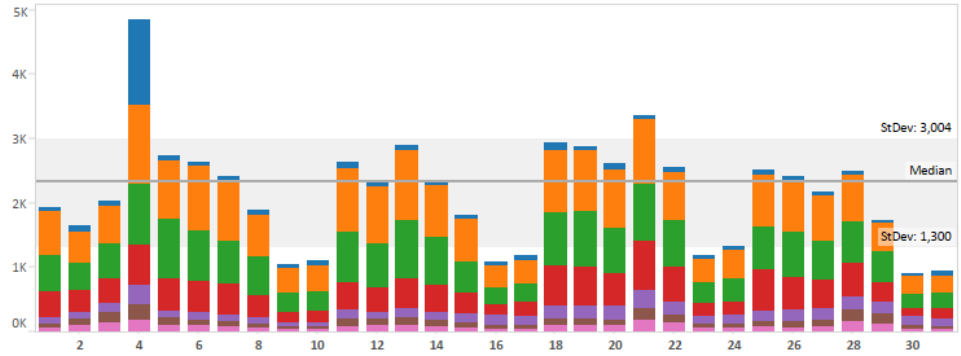
Region

State Type

Lead Source
All

Generated By
All

Breakdown Lead By Day



Measure: Lead

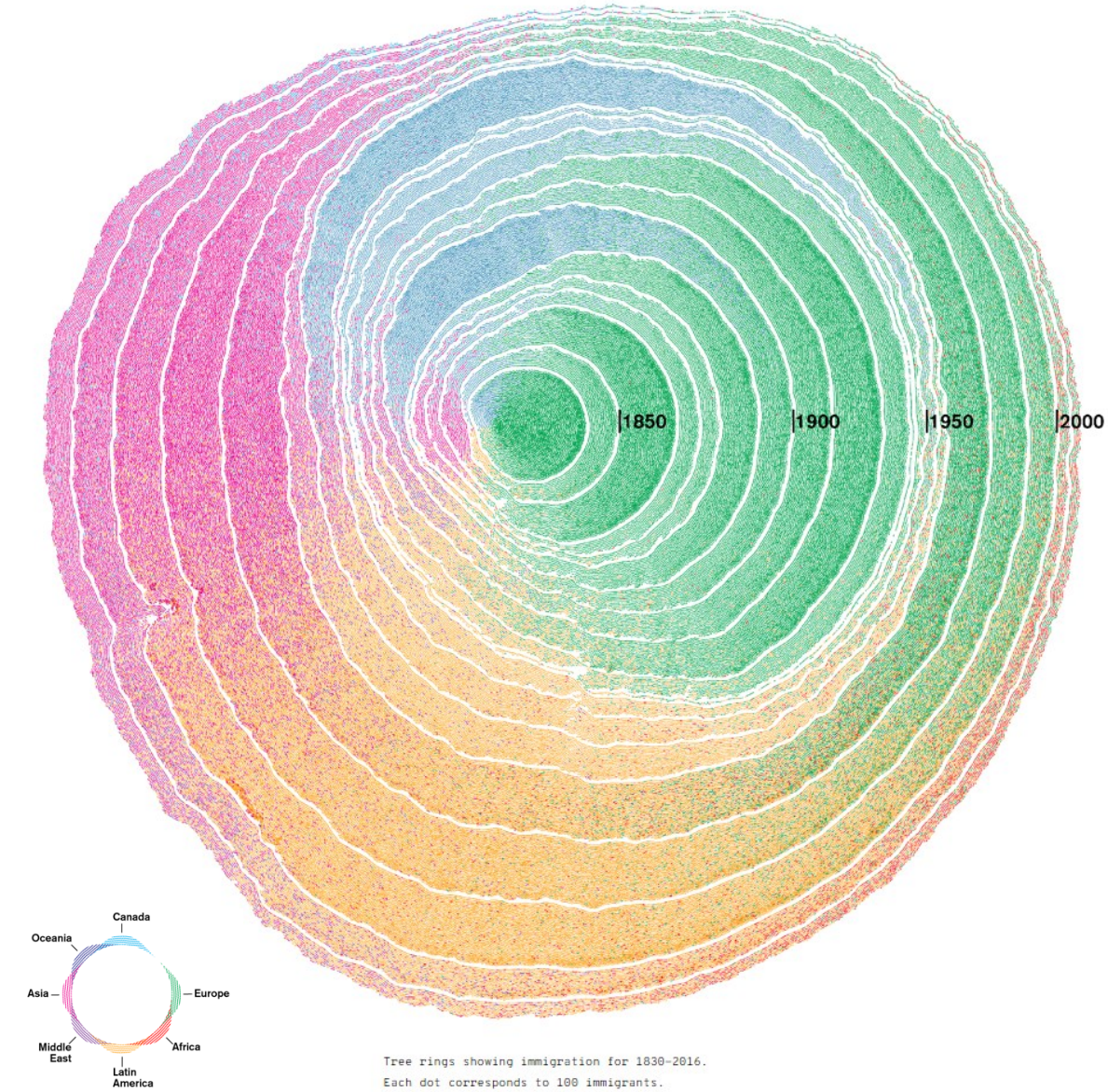
Date Interval: Day

Color By: State Type

- Unspecified
- Type A
- Type B
- Type C
- Type D
- Type E
- Type F

US Immigration Data

https://web.northeastern.edu/naturalizing-immigration-dataviz/



What is Data Visualization ?

Visual representation of data

"Transformation of the symbolic into the geometric" [McCormick et al, 1987]

"… artificial memory that best supports our natural means of perception" [Bertin, 1967]

"Use of computer-generated, interactive, visual representations of data to amplify cognition" [Card, Mackinlay, & Shneidermann, 1999]

Why Data Visualization?

"The ability to take data –

to be able to **understand** it,
to **visualize** it,
to **communicate** it –

that's going to be a hugely important skill in the next decades,

… because now we really do have essentially free and ubiquitous data.

So the complimentary **scarce factor is the ability to understand** the data and extract value from it."

Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

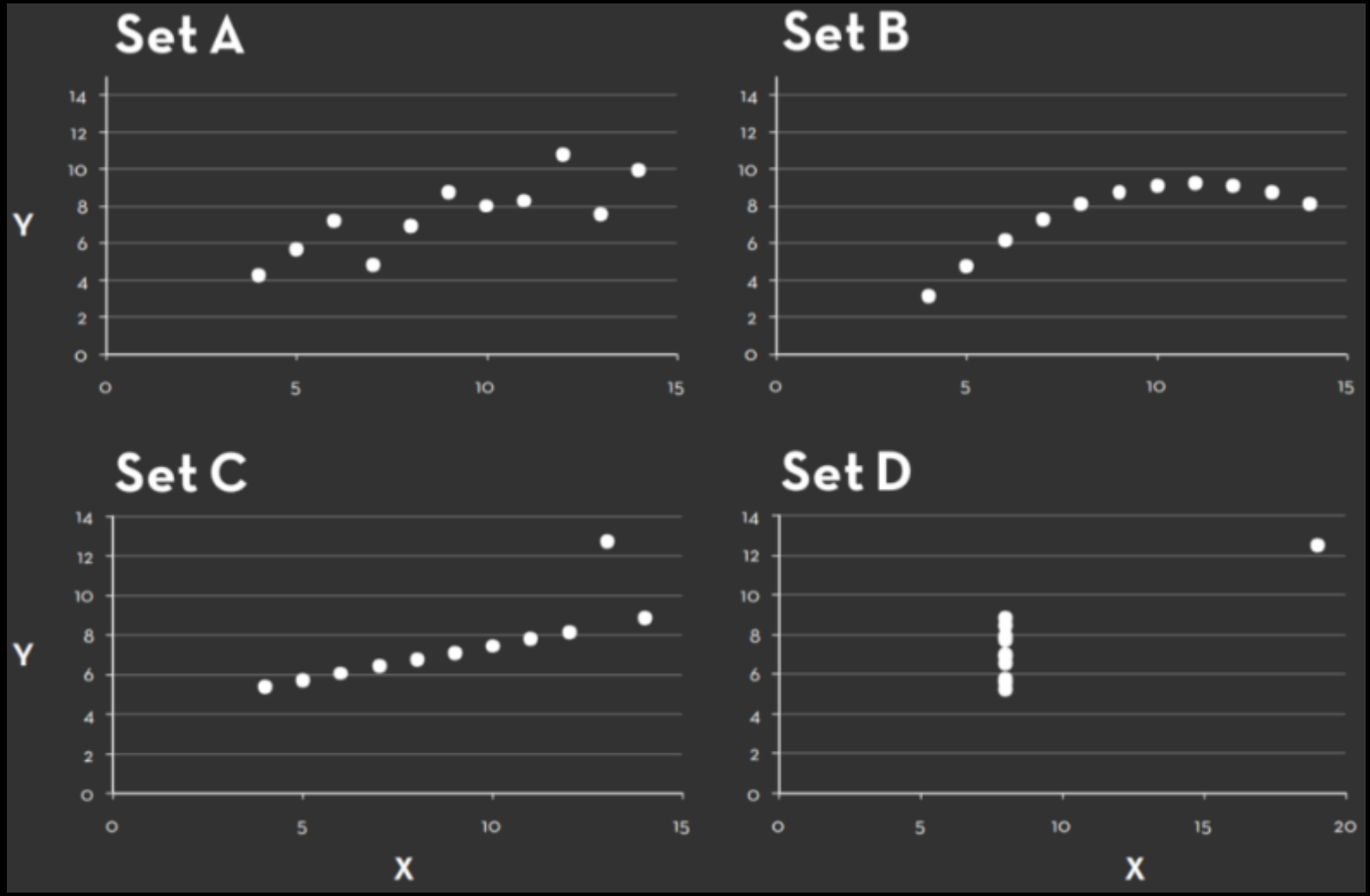
Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.08	10	9.14	10	7.47	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

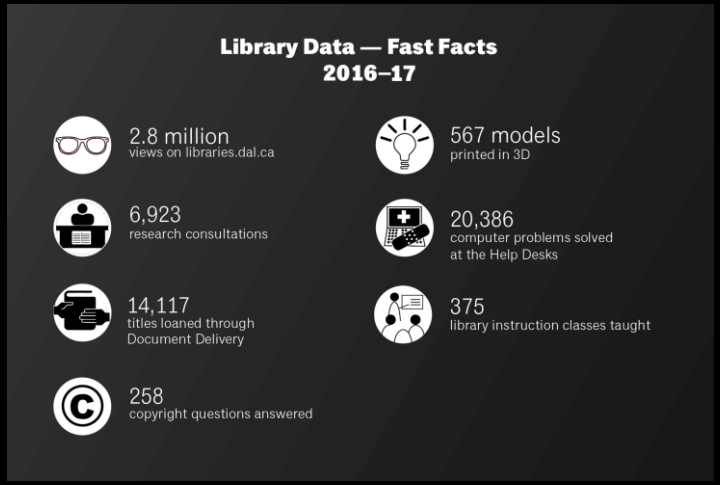
Anscombe's Quartet (Anscombe, Francis J., 1973)


Set A
Y
14
12
10
8
6
4
2
0
0
5
10
15
Set B
14
12
10
8
6
4
2
0
0
5
10
15
Set C
Y
14
12
10
8
6
4
2
0
0
5
10
15
X
Set D
14
12
10
8
6
4
2
0
0
5
10
15
20
X

InfoGraphics vs DataGraphics

DataGraphic

Uses words and numbers to represent quantitative data. Icon-like graphical elements for visual appeal



https://libraries.dal.ca/about/library-assessment/library-data.html

InfoGraphic

Visually encodes quantitative and/or qualitative data into marks, shapes, sizes, colours, etc.

https://library2.lincoln.ac.nz/dashboard/#library

http://understandinggraphics.com/visualizations/infoposters-are-not-infographics/

InfoGraphic vs Data Visualizations

Infographics tell a premeditated story to guide the audience (subjective).

Data Visualizations leave the audience their own conclusions (objective).

InfoGraphic

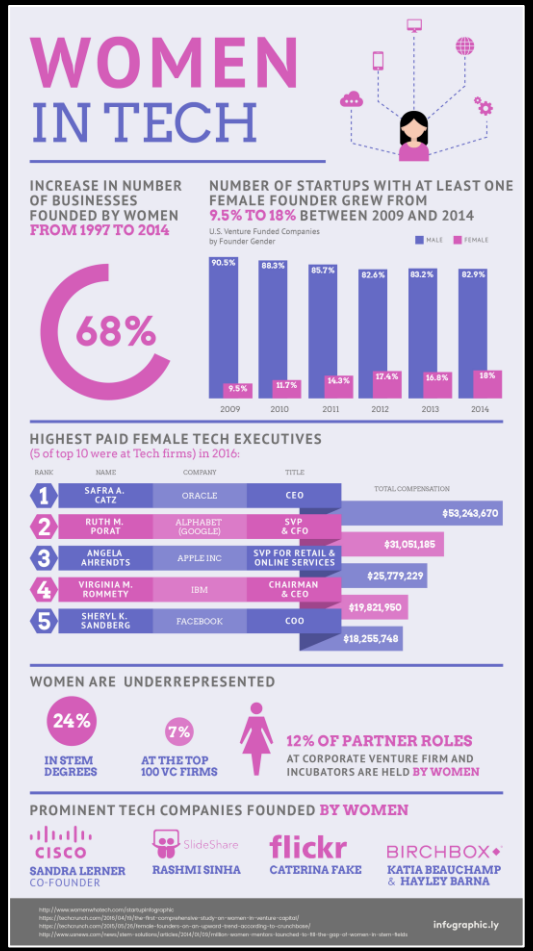
- Best for telling a **premeditated story** and offer **subjectivity**
- Best for **guiding the audience** to conclusions and **point out relationships**
- Created **manually** for one **specific dataset**

Data Visualization

- Best for allowing the audience to **draw their own conclusions**, and offer **objectivity**
- Ideal for understanding **data at a glance**
- **May not** be designed for a **specific data**

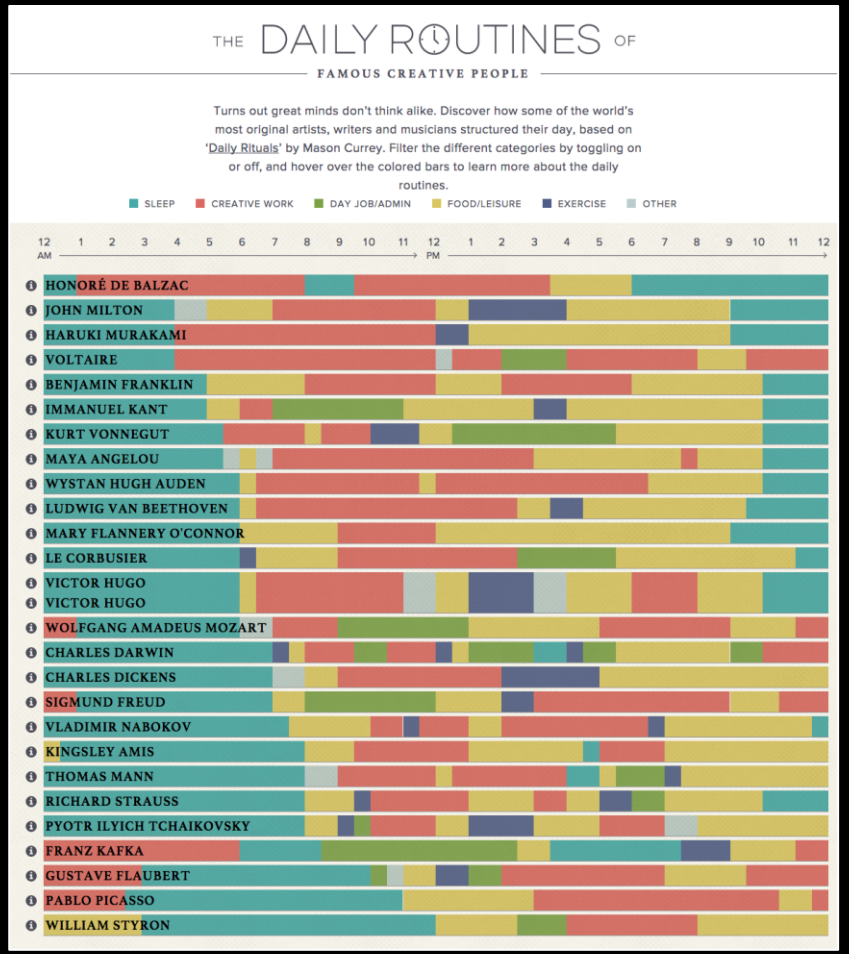
https://blog.prototypr.io/getting-it-right-why-infographics-are-not-the-same-as-data-visualizations-a23da7de745e

InfoGraphic



From https://www.entrepreneur.com/article/289932

Data Visualization



From https://podio.com/site/creative-routines

How Do We Make a Good Data Visualization?

- **Know the Data**
  - Number of attributes
  - Date types: ordinal vs ordered (ordinal or quantitative)
  - Trustworthiness: bad fields, inaccuracies, missing values
- **Know your purpose (& audience)**
  - What do you/they want to see?
  - What might you/they want to focus on?
- **Decide how encode the data**
  - Ensure information can be decoded accurately
  - Human perceptual system
  - Display capacity
  - Characteristics of data (size, type)
  - Task

Data

Categorical

Ordered Small Medium Large

Quantitative 1 2 5.29 42 101

How Do We Make a Good Data Visualization?

- **Know the Data**
  - Number of attributes
  - Date types: ordinal vs ordered (ordinal or quantitative)
  - Trustworthiness: bad fields, inaccuracies, missing values
- **Know your purpose (& audience)**
  - What do you/they want to see?
  - What might you/they want to focus on?
- **Decide how encode the data**
  - Ensure information can be decoded accurately
  - Human perceptual system
  - Display capacity
  - Characteristics of data (size, type)
  - Task

How do we change data symbols into geometry?

Exercise:

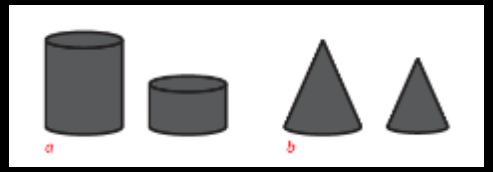
How many ways can you communicate two quantities?

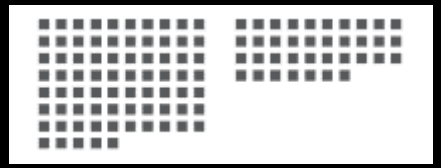
75 and 37

http://blog.visual.ly/45-ways-to-communicate-two-quantities/

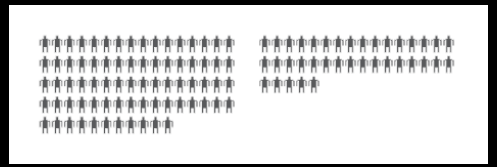

75, 37
a
b
c


a
b


a
b

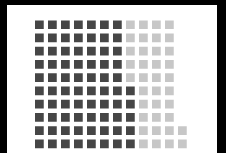


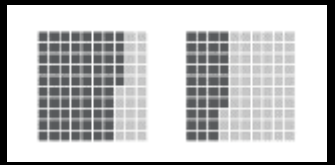

75
37
a
b

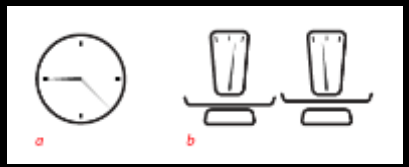



75
37


a
b



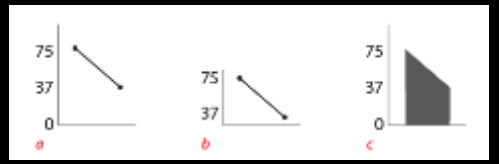



a
b

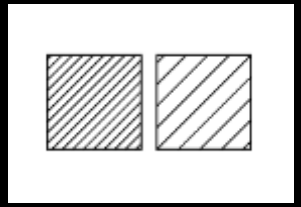

a
b
c

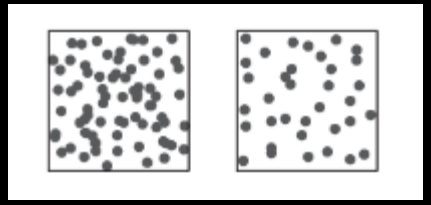
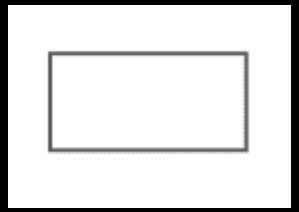
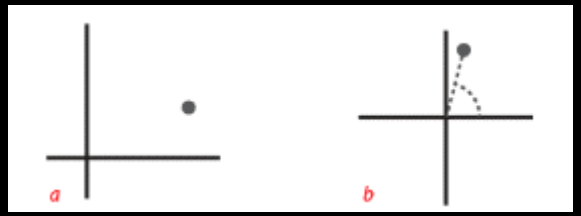

a
b
c


a
b


75
37
0
a
75
37
b
75
37
0
c


a
b
c

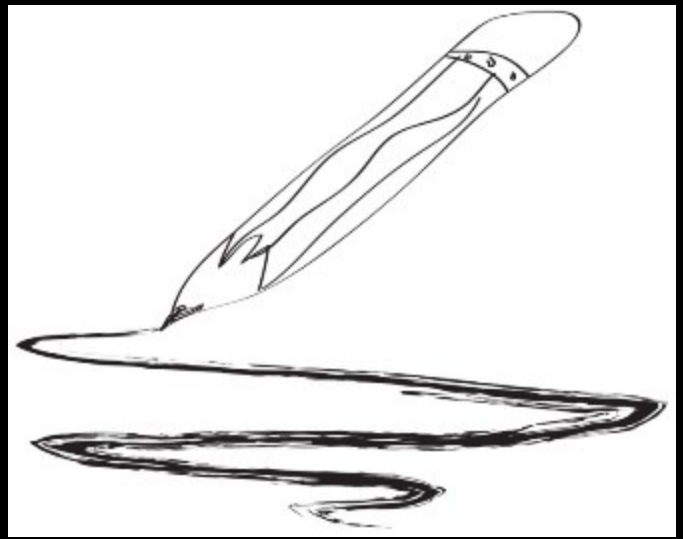




http://blog.visual.ly/45-ways-to-communicate-two-quantities/

Visual Encoding

- Marks



- What can we change about a mark?

Visual Encoding With Visual Variables

Visual Variables:

- *Position*
- *Angle*
- *Slope*
- *Length*
- *Area*
- *Volume*
- *Shape*

Size

- *Lightness*
- *Saturation*
- *Hue*

Colour

- *Texture*
- *Connection*
- *Containment*

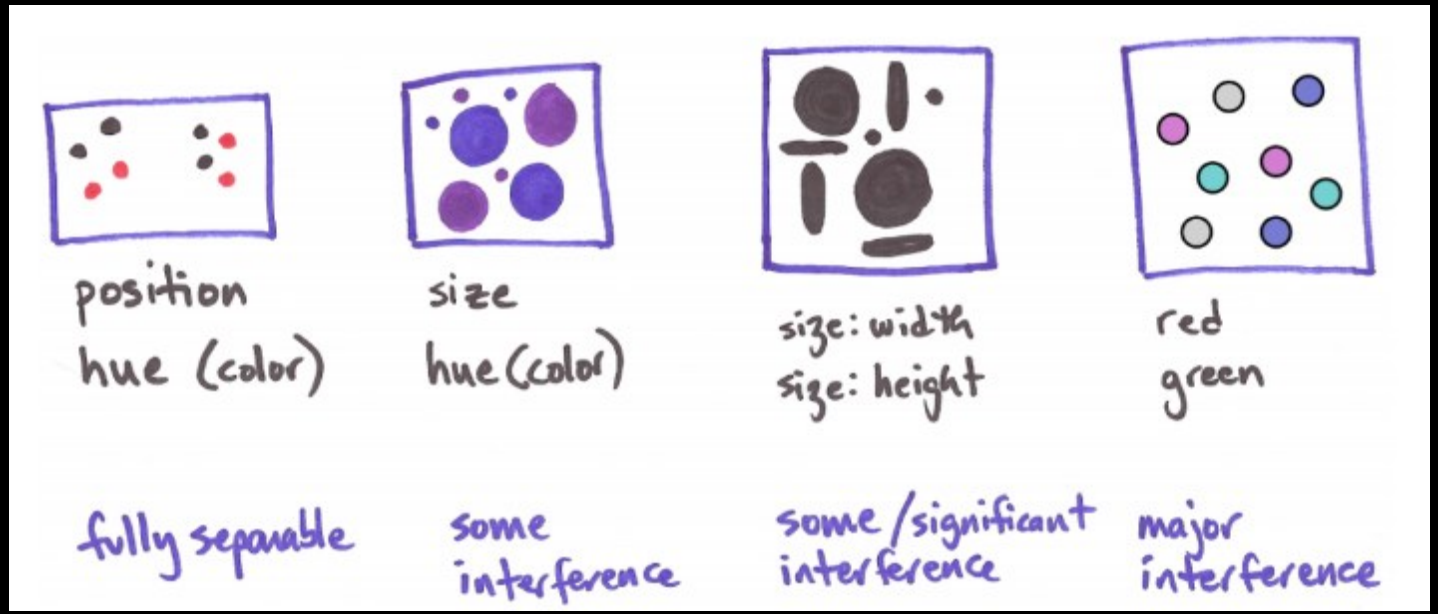
Independent?

? :

- *Opacity*
- *"Sketchiness"*

Visual Encoding

- Are differences perceivable?
- How many bins?
- Ideally all channels would be independent (not so)
  - Interactions between channels



Perception

Brain

Visual Cortex: fast & efficient – pre-attentive processing

Cerebral Cortex: slower, less efficient

Goal: do as much as possible with Visual Cortex as possible

Perception

How many 3s?

1 8 4 7 9 5 3 2 1 2 4 6 7 8 9 5 6 4 3

4 8 0 6 4 8 0 3 2 8 8 7 9 6 2 3 1 0 6

9 9 6 3 4 4 2 6 8 1 5 6 8 7 9 0 3 2 1

1 5 6 8 7 9 6 5 1 2 3 5 9 9 7 8 9 6 5

4 3 2 1 3 2 1 5 4 9 8 3 4 2 5 8 4 8 9

2 2 1 5 6 7 8 6 5 6 3 1 4 5 1 3 4 5 1

Perception

How many 3s?

1 8 4 7 9 5 3 2 1 2 4 6 7 8 9 5 6 4 3

4 8 0 6 4 8 0 3 2 8 8 7 9 6 2 3 1 0 6

9 9 6 3 4 4 2 6 8 1 5 6 8 7 9 0 3 2 1

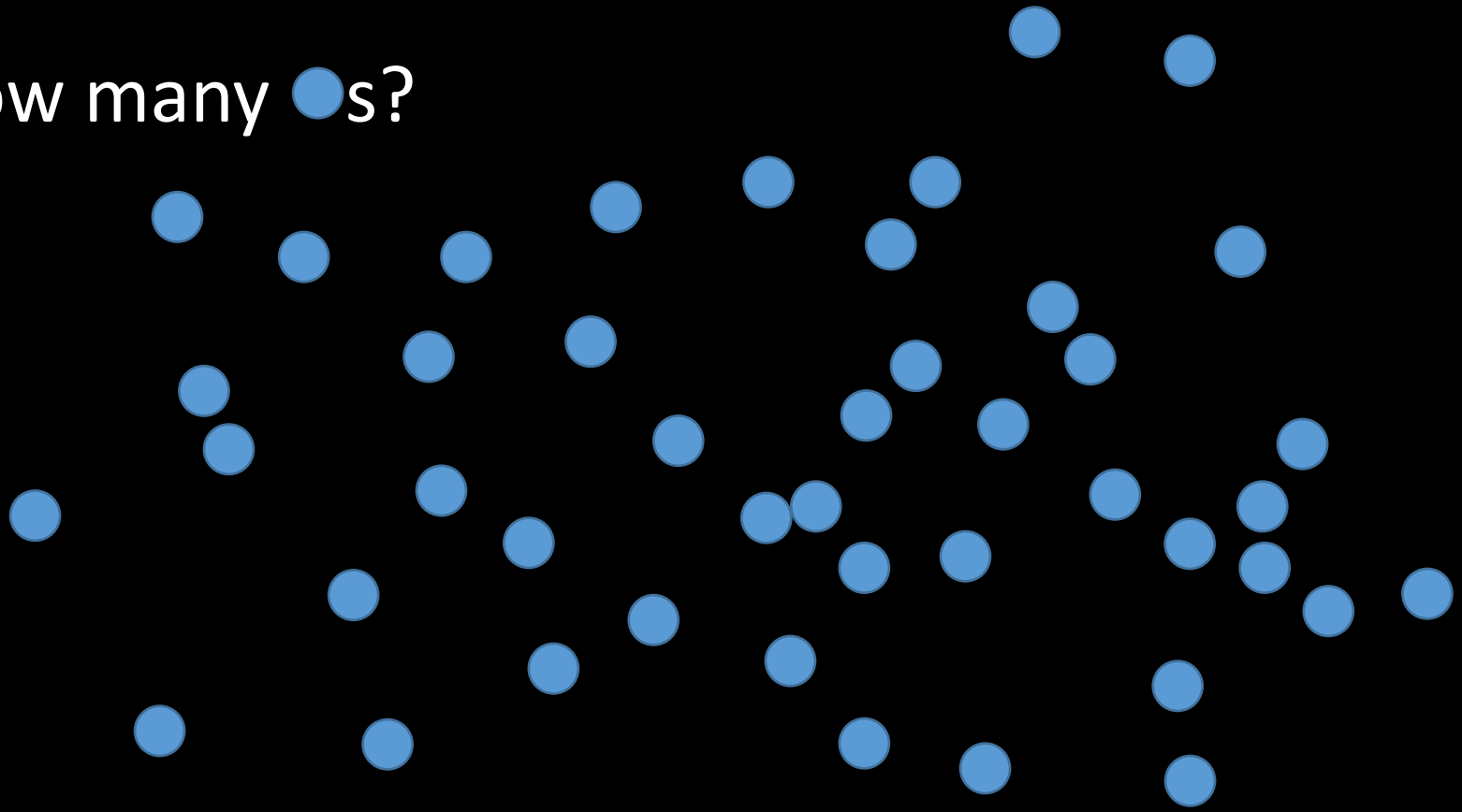
1 5 6 8 7 9 6 5 1 2 3 5 9 9 7 8 9 6 5

4 3 2 1 3 2 1 5 4 9 8 3 4 2 5 8 4 8 9

2 2 1 5 6 7 8 6 5 6 3 1 4 5 1 3 4 5 1

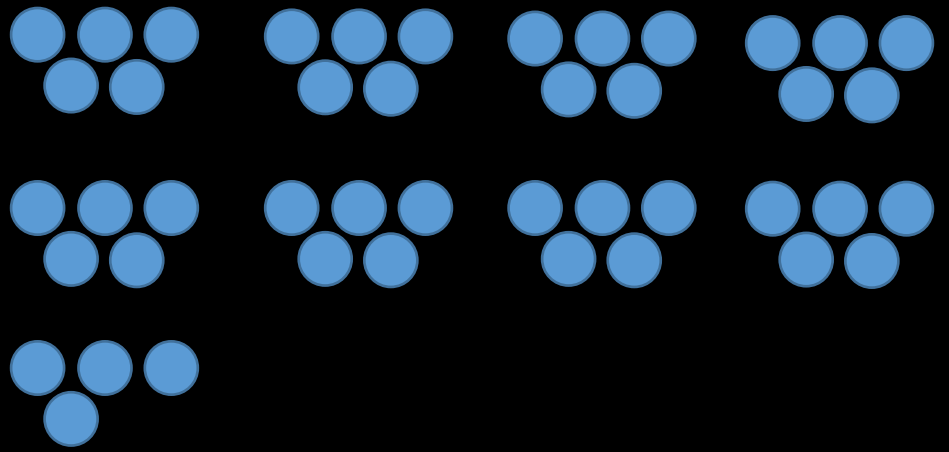
Perception

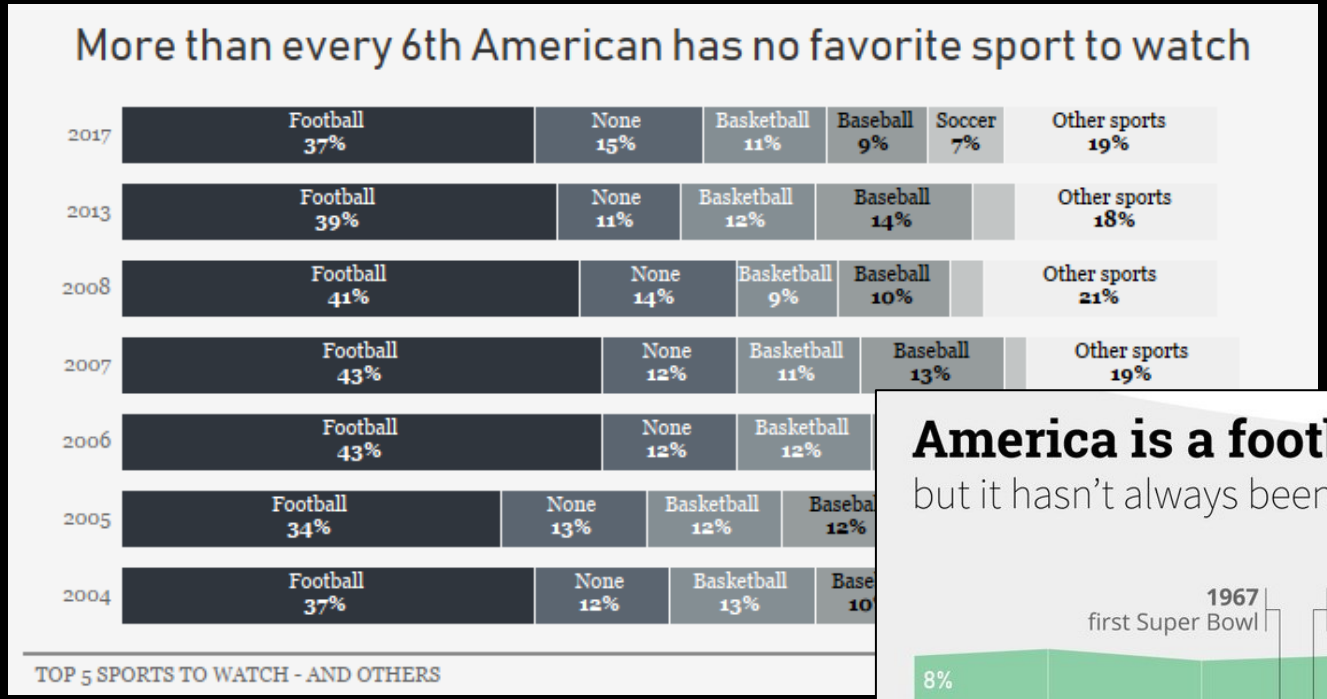
How many s?

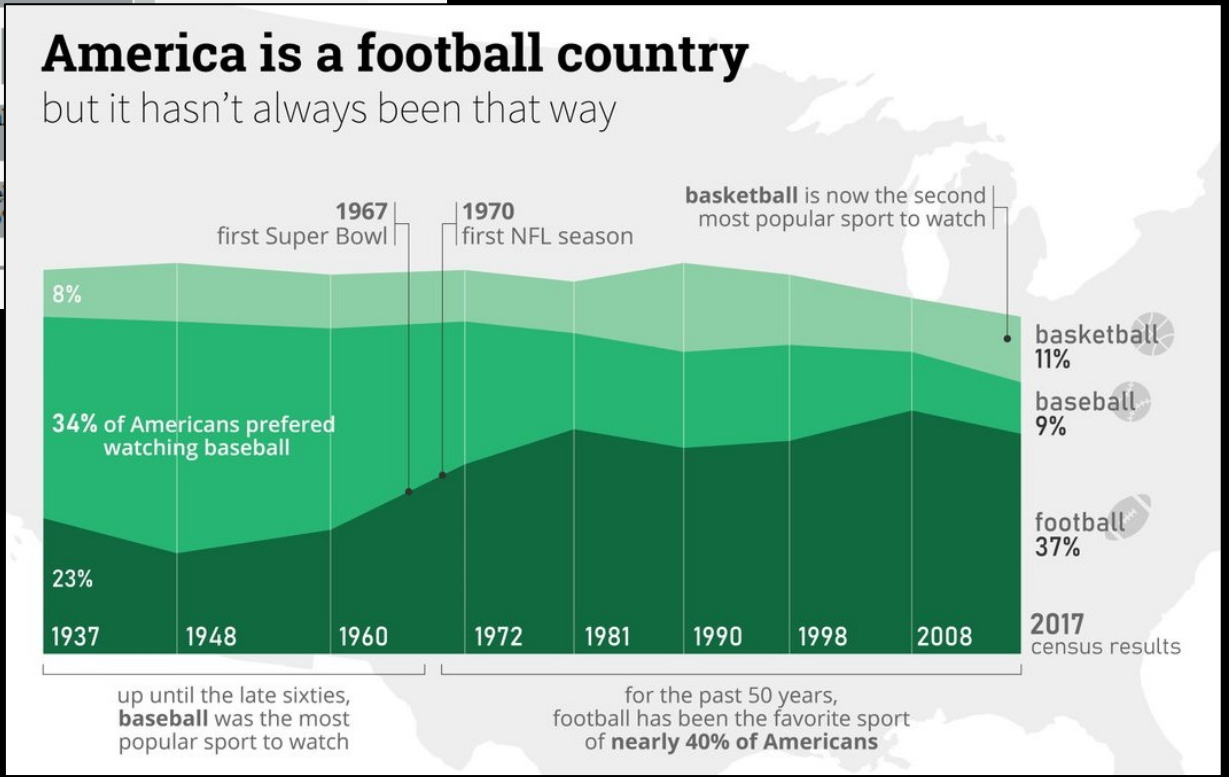


Perception

How many ●s?







Example from
https://www.makeovermonday.co.uk/gallery/

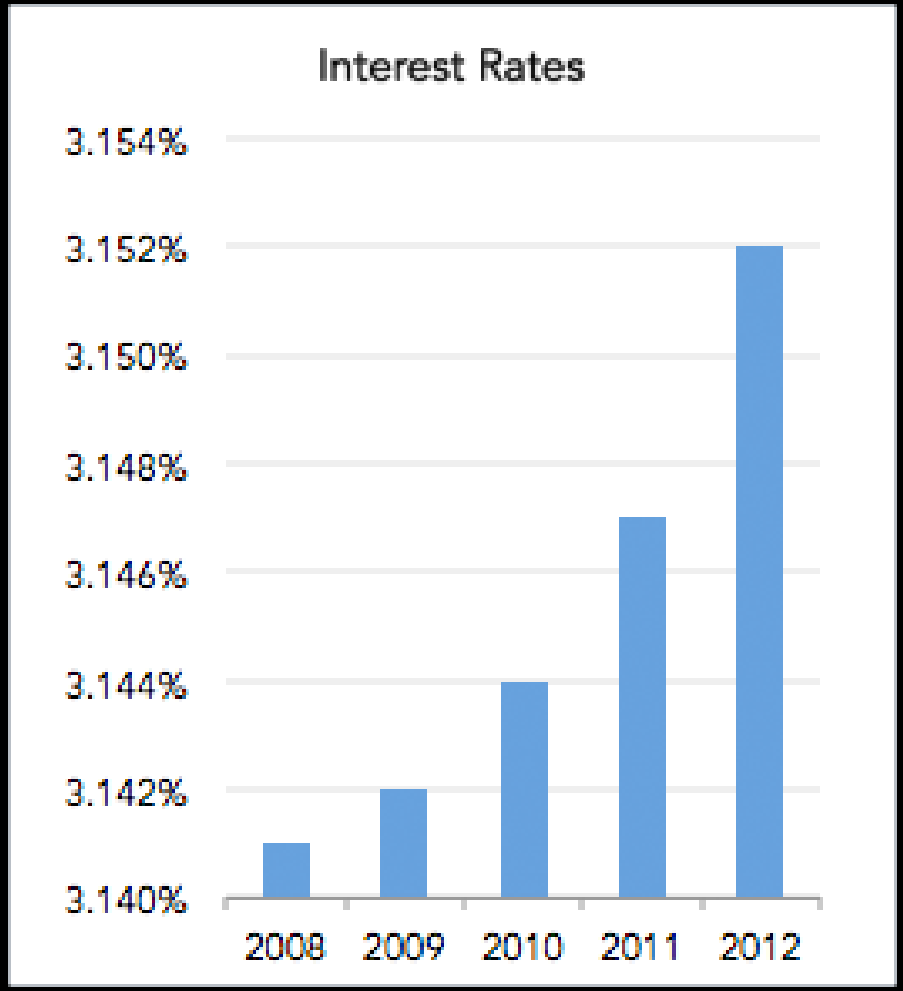
Visual Encoding works with a Sign System

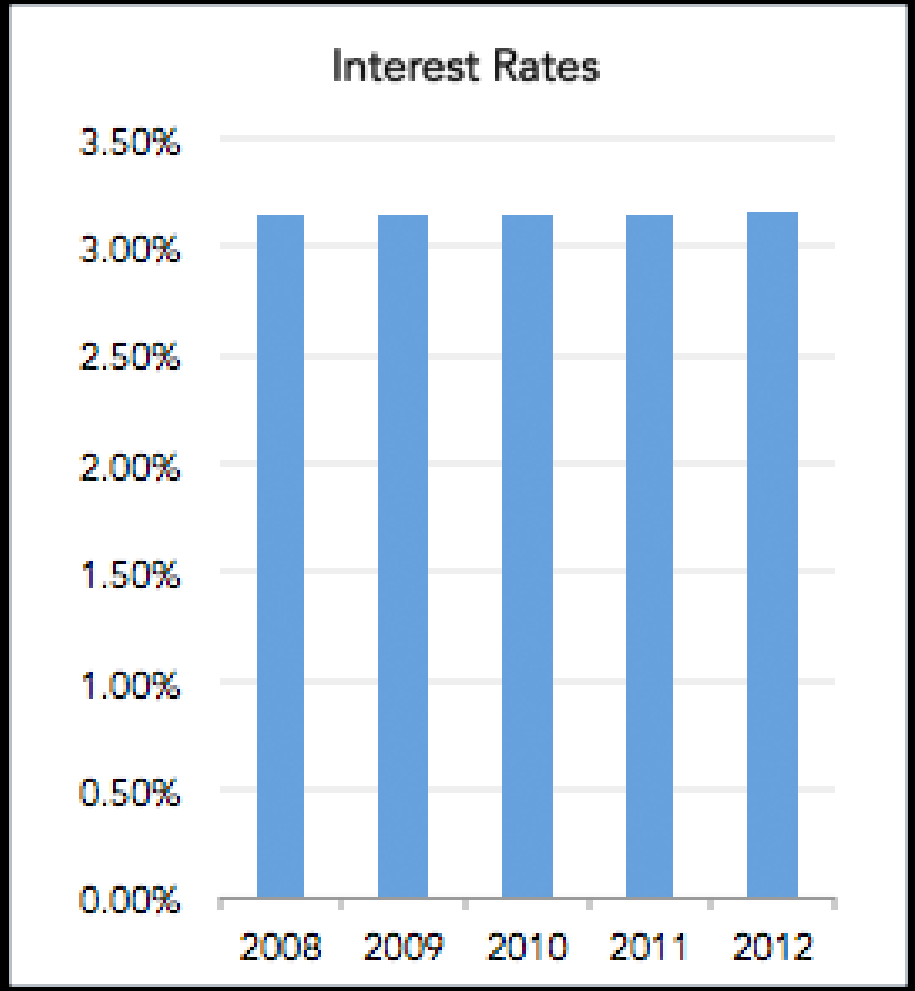
Images are perceived as a set of signs

Designer encodes information in signs

Receiver decodes information from signs

Jacques Bertin, *The Semiology of Graphics*, 1967


Interest Rates
3.154%
3.152%
3.150%
3.148%
3.146%
3.144%
3.142%
3.140%
2008
2009
2010
2011
2012


Interest Rates
3.50%
3.00%
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
2008
2009
2010
2011
2012

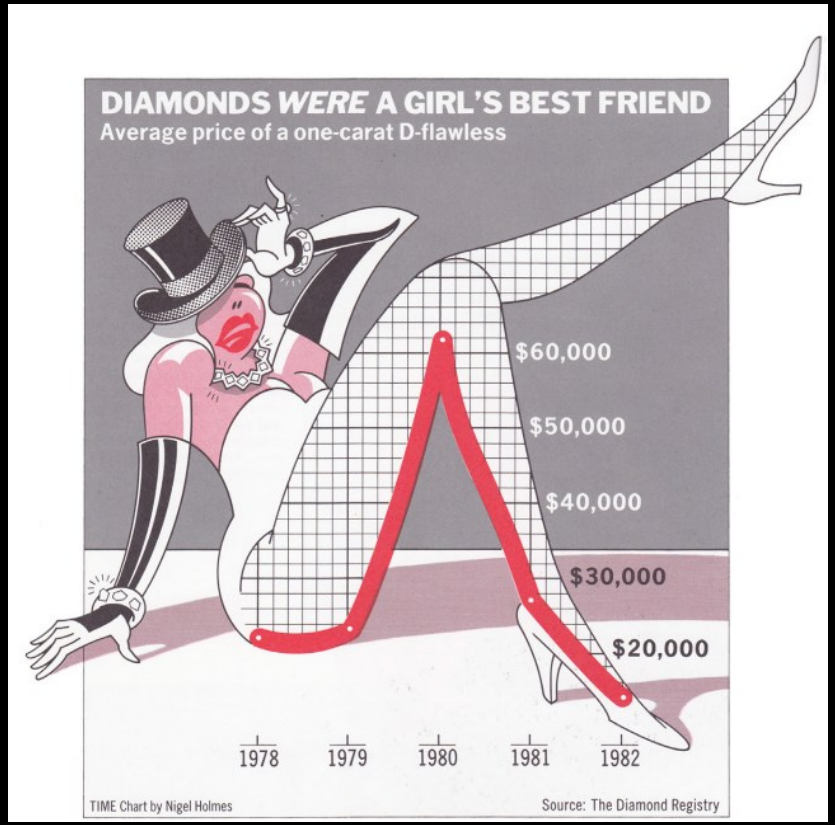
Visualization ~~Rules~~ Advice

"Avoid chart junk at all costs" "Never use pie charts"

"Bright colors don't work"

"No rainbow color-maps"

- These are all very situationally dependent
- Make sure you have reasons for breaking them



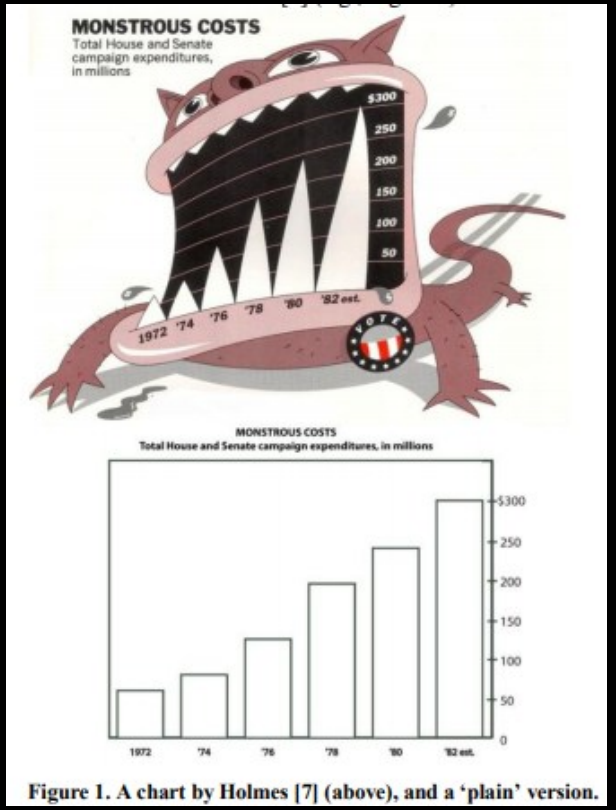
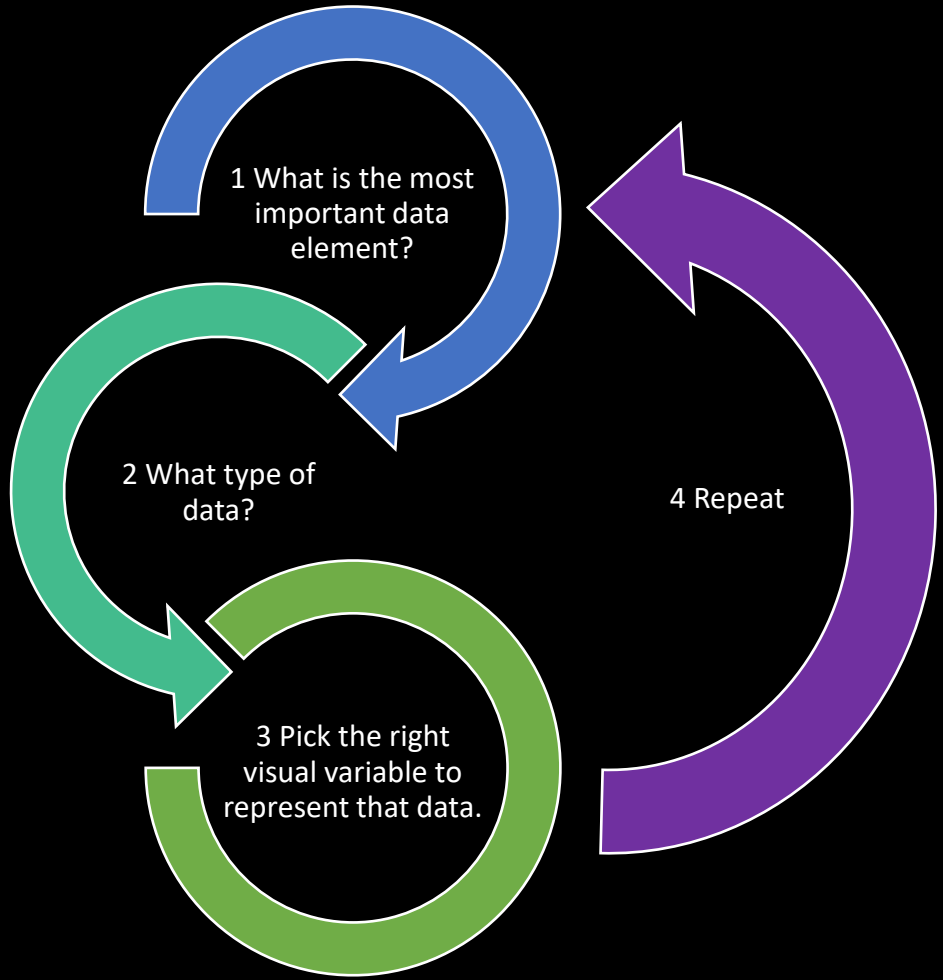


Figure 1. A chart by Holmes [7] (above), and a 'plain' version.

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

Bottom-Up Approach to Designing a Visualization



Visual Variable Properties

1. **Selective**
   - Is a change of a mark in this variable alone enough to allow us to select it from other marks?
2. **Associative**
3. **Quantitative**
4. **Order**
5. **Length**

Is Size Selective?

Can you find the big & small Muppets?

Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Is Size Selective?

Can you find the big & small Muppets?

Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Visual Variable Properties

1. **Selective**
   - Is a change of a mark in this variable alone enough to allow us to select it from other marks?
2. **Associative**
   - Can we identify a group of marks by this variable?
3. **Quantitative**
4. **Order**
5. **Length**

Is Size Associative?

Can you find the small Muppets?

Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Is Size Associative?

Can you find the small Muppets?

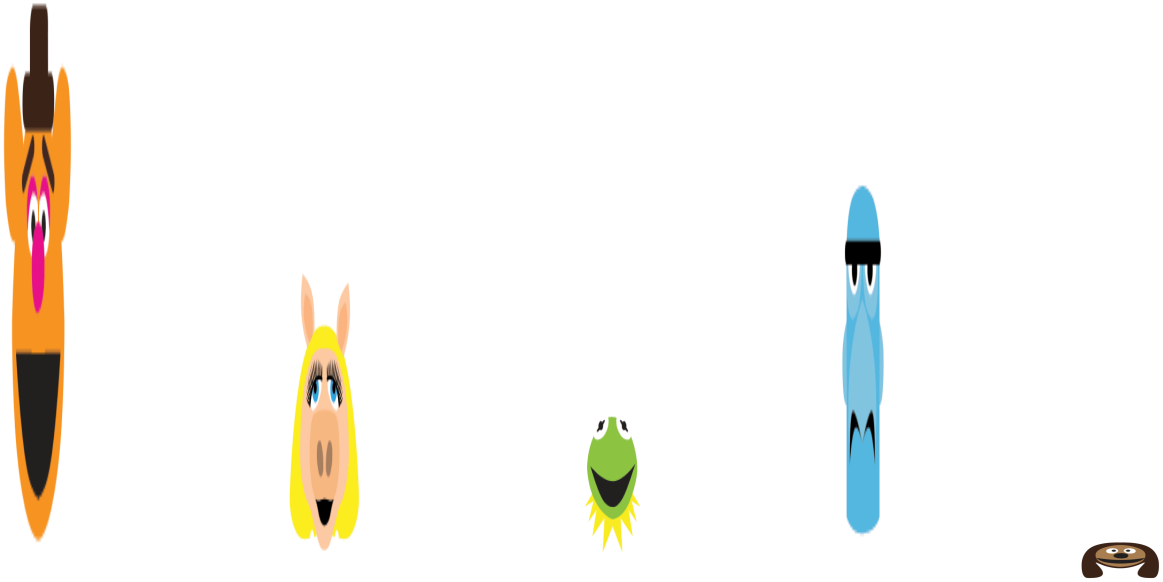
Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Visual Variable Properties

1. **Selective**
   - Is a change of a mark in this variable alone enough to allow us to select it from other marks?
2. **Associative**
   - Can we identify a group of marks in this variable?
3. **Quantitative**
   - Can the relation between two of these marks be seen as numeric? Can we tell if one is 3X another?
4. **Order**
5. **Length**

Is Size Quantitative?

What value is Kermit compared to Fozzie?



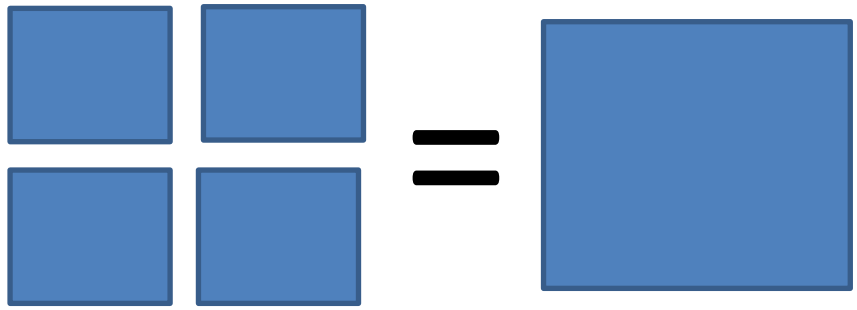
Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

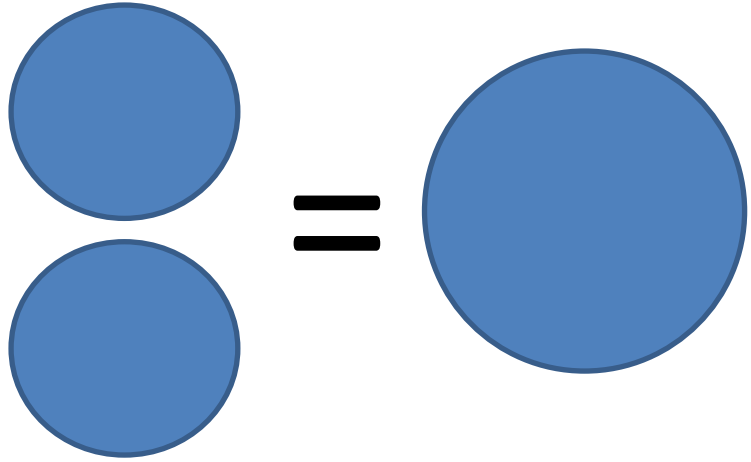
Is Size Quantitative?

What value is Kermit compared to Fozzy?

Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Is Size Quantitative?

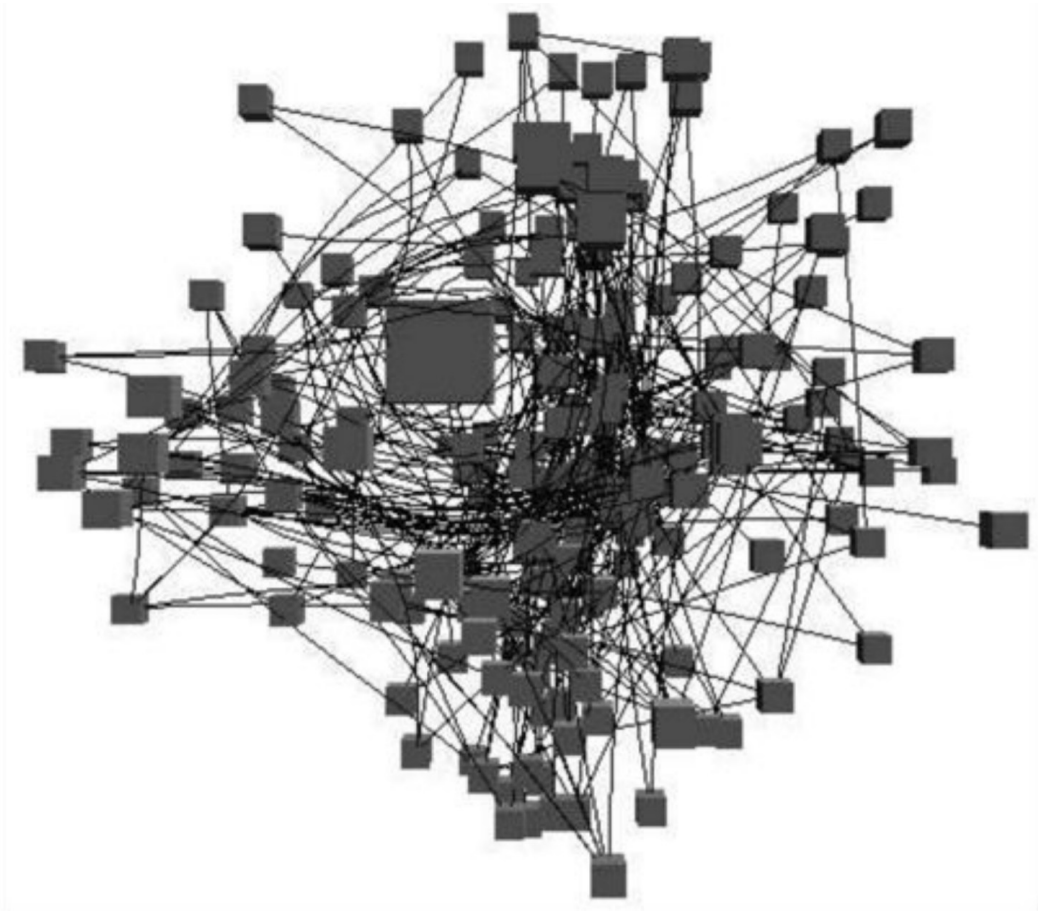




3D

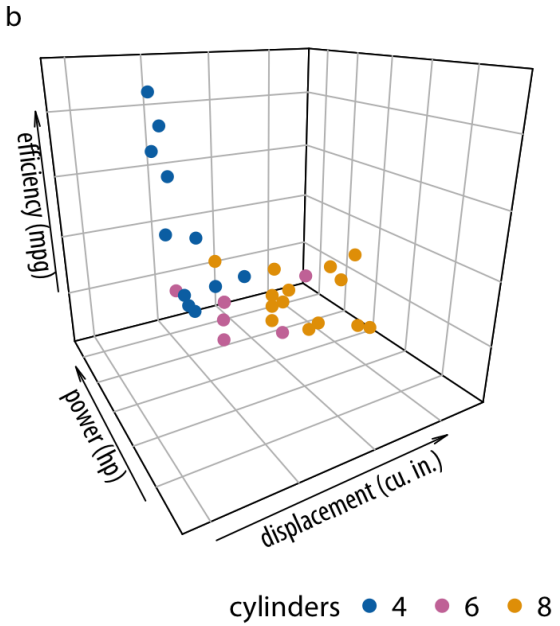
Problems

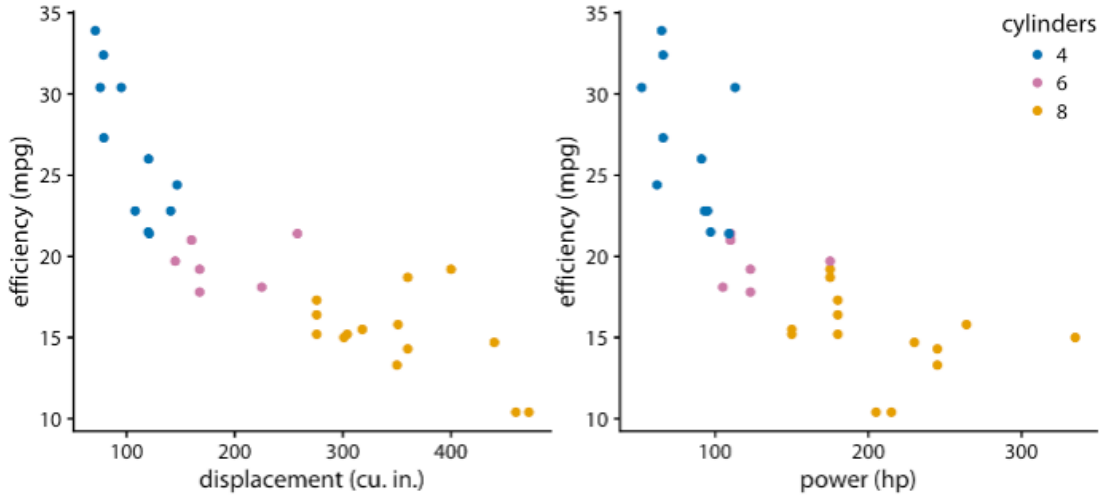
- Comparison
- Occlusion / viewpoint

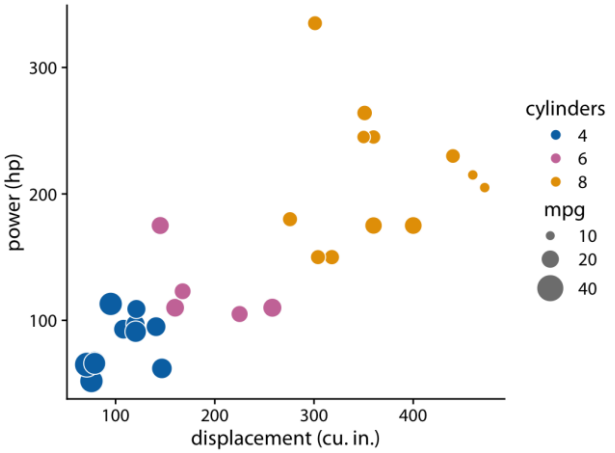


Avoid 3D

Look for a better solution



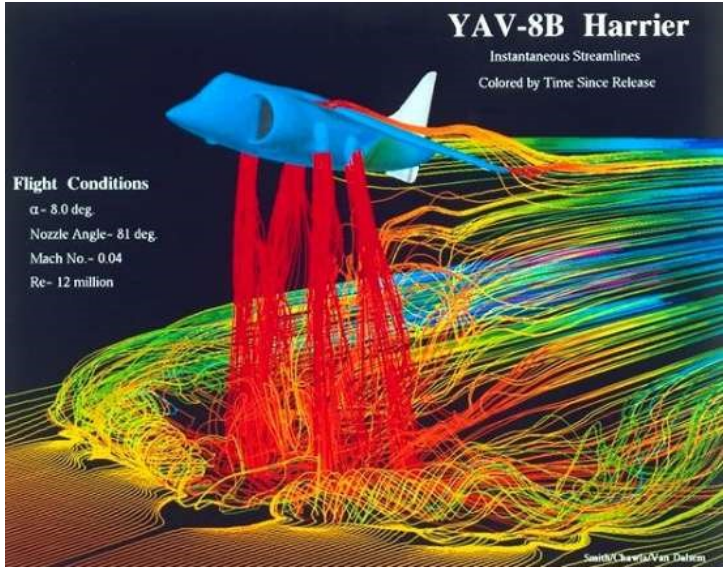




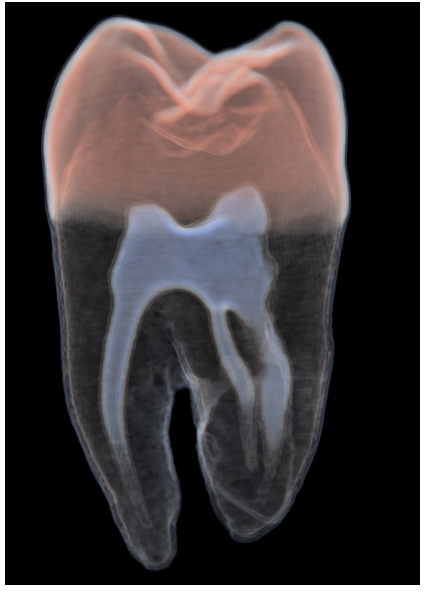
From https://clauswilke.com/dataviz/no-3d.html

When to use 3D?

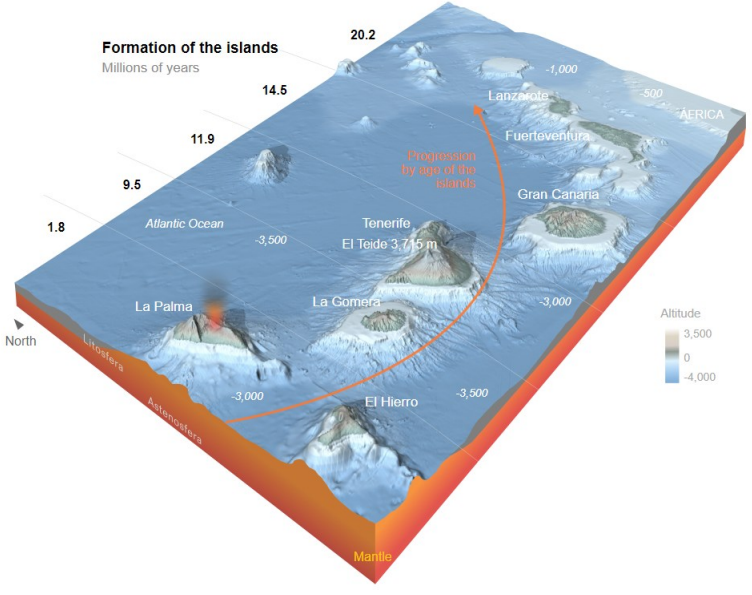
➢When your data has 3D position



https://aerospaceweb.org/question/planes/q0102a.shtml



https://www.kitware.com/new-in-paraview-5-9-volume-rendering-with-a-separate-opacity-array/



https://english.elpais.com/science-tech/2021-10-06/the-underwater-hotspot-feeding-la-palmas-volcano-will-create-new-islands.html

Visual Variable Properties

1. **Selective**
   - Is a change of a mark in this variable alone enough to allow us to select it from other marks?
2. **Associative**
   - Can we identify a group of marks in this variable?
3. **Quantitative**
   - Can the relation between two of these marks be seen as numeric? Can we tell if one is 3X another?
4. **Order**
   - Does this variable support ordered reading (more/less)?
5. **Length**

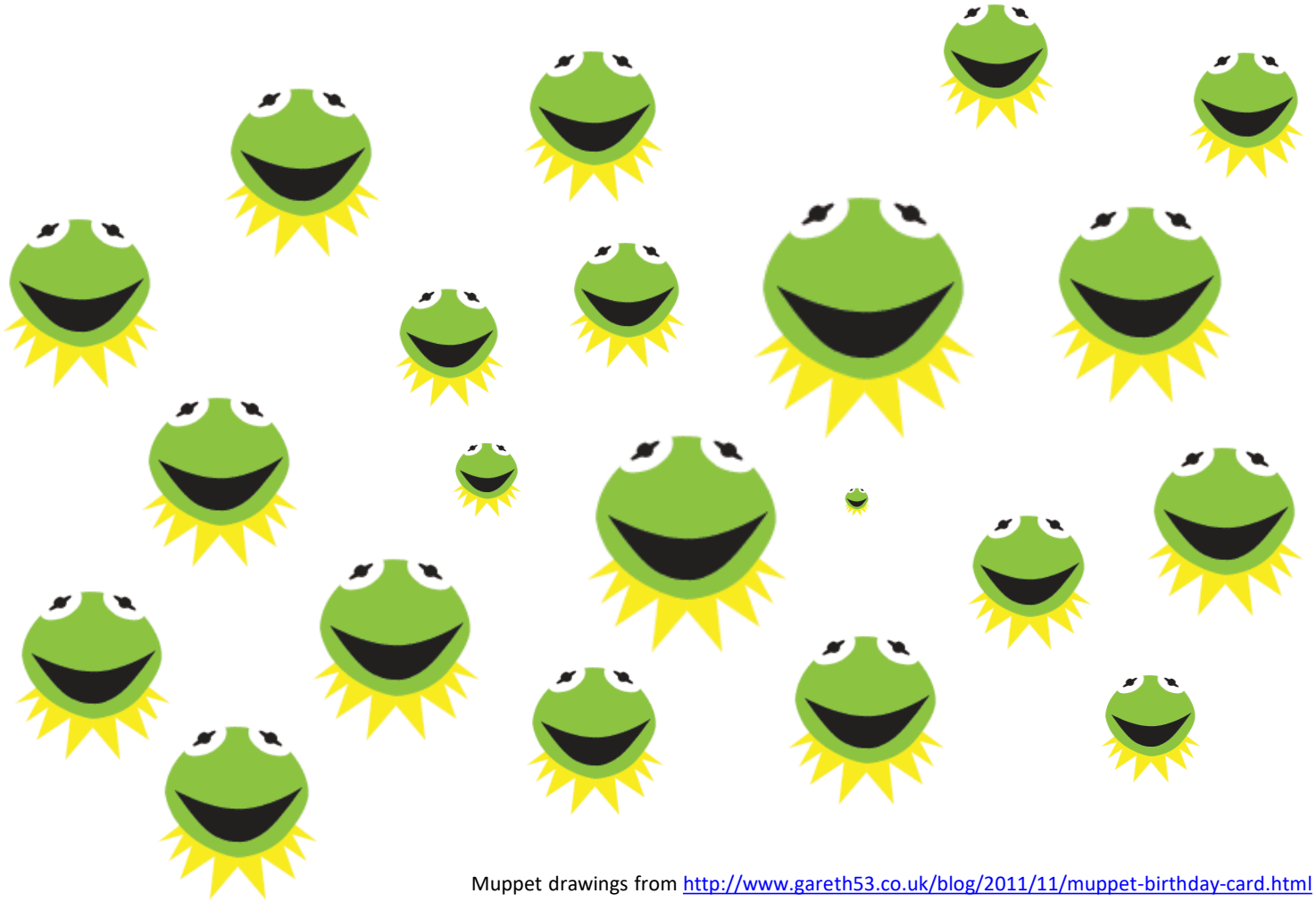
Is Size Ordered?

Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

Visual Variable Properties

1. **Selective**
   - Is a change of a mark in this variable alone enough to allow us to select it from other marks?
2. **Associative**
   - Can we identify a group of marks in this variable?
3. **Quantitative**
   - Can the relation between two of these marks be seen as numeric? Can we tell if one is 3X another?
4. **Order**
   - Does this variable support ordered reading (more/less)?
5. **Length**
   - How many differences in this variable can be discerned?

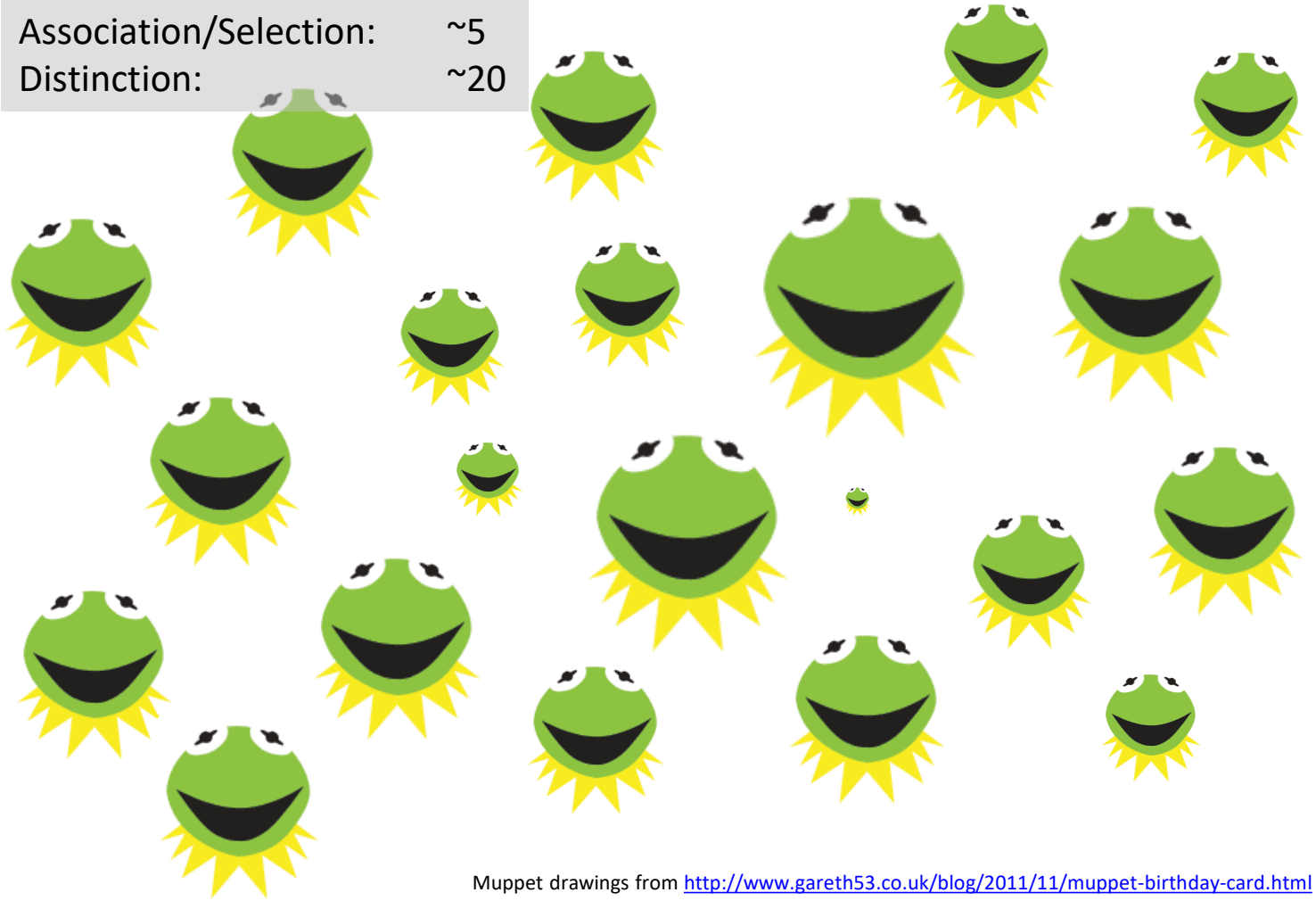
What Length Does Size Have?



Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

What Length Does Size Have?

Association/Selection:	~5
Distinction:	~20



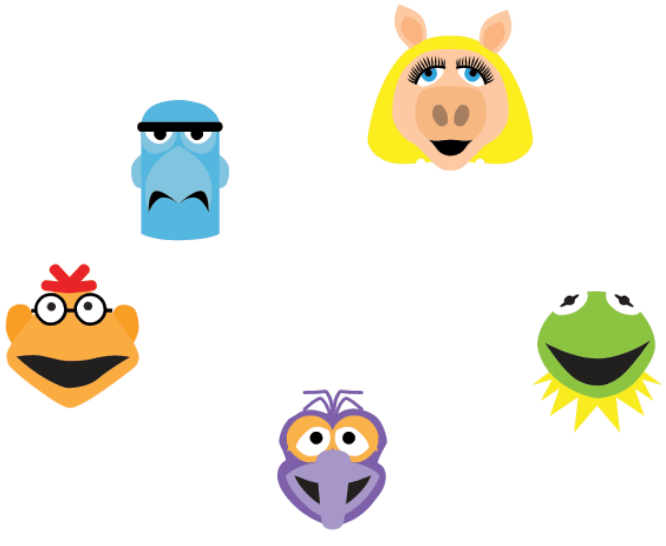
Muppet drawings from http://www.gareth53.co.uk/blog/2011/11/muppet-birthday-card.html

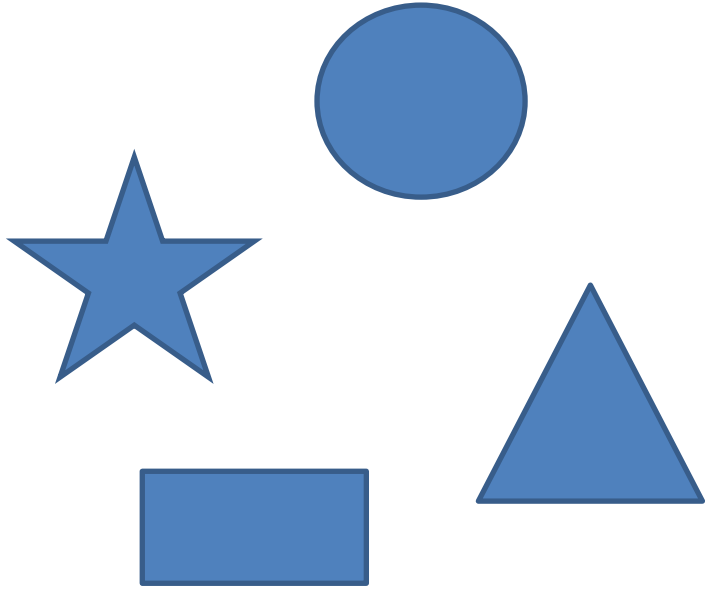
Visual Encoding
Channel Properties

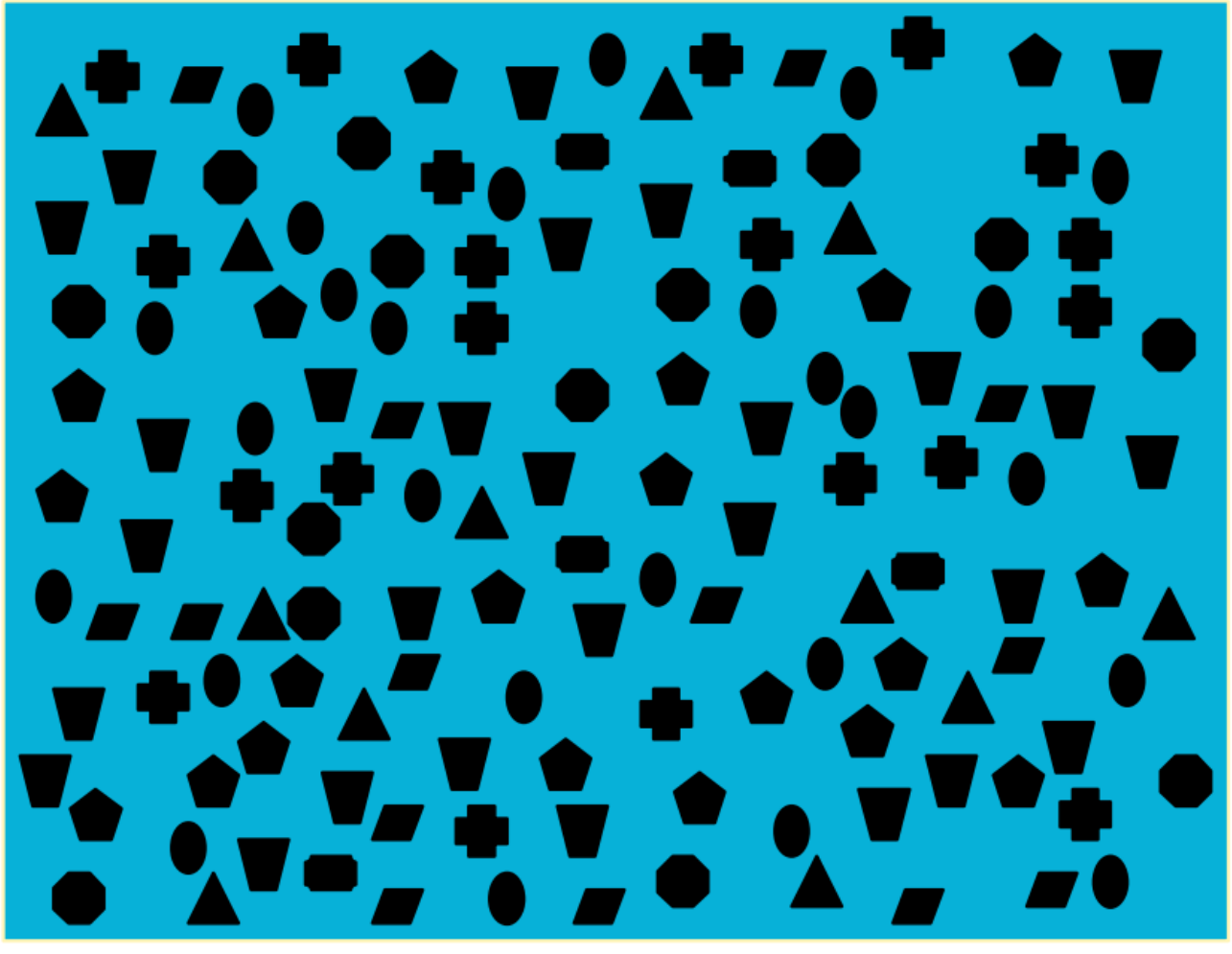
- **Selective**
  - Is a change in this variable alone enough to allow us to select it from a group?
- **Associative**
  - Can we identify a group of these marks?
- **Quantitative**
  - Can the relation between two of these marks be seen as numeric? Can we tell if one is 3X as much as another?
- **Order**
  - Does this variable support ordered reading (more/less)?
- **Length**
  - How many differences in this variable can be discerned?

Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape					
Lightness					
Saturation					
Hue					
Angle					
Texture					

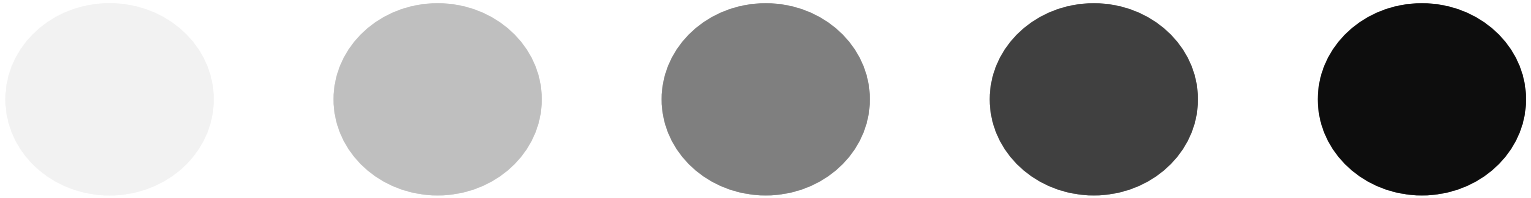
Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness					
Saturation					
Hue					
Angle					
Texture					

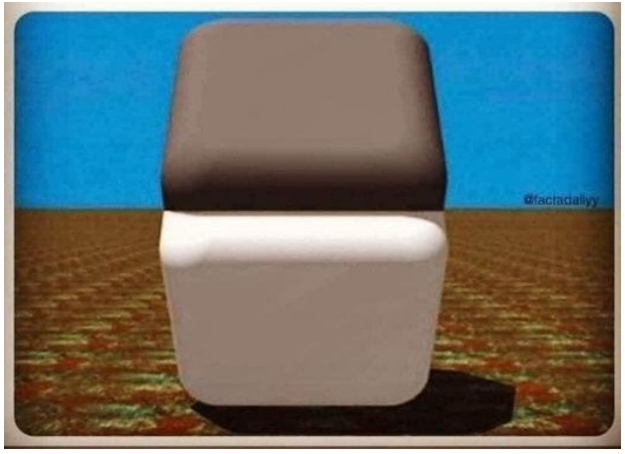


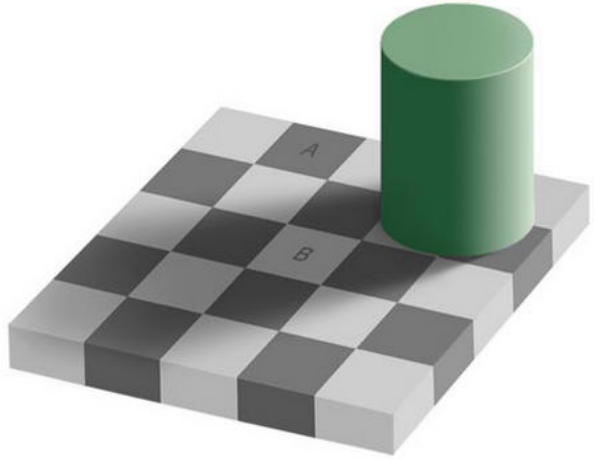


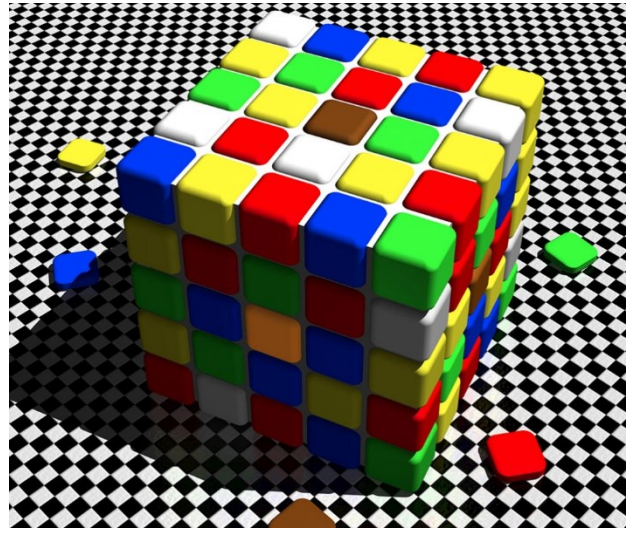


Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue					
Angle					
Texture					



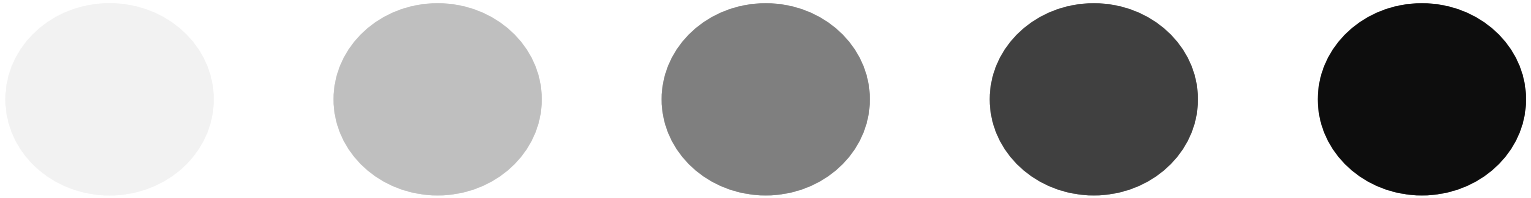






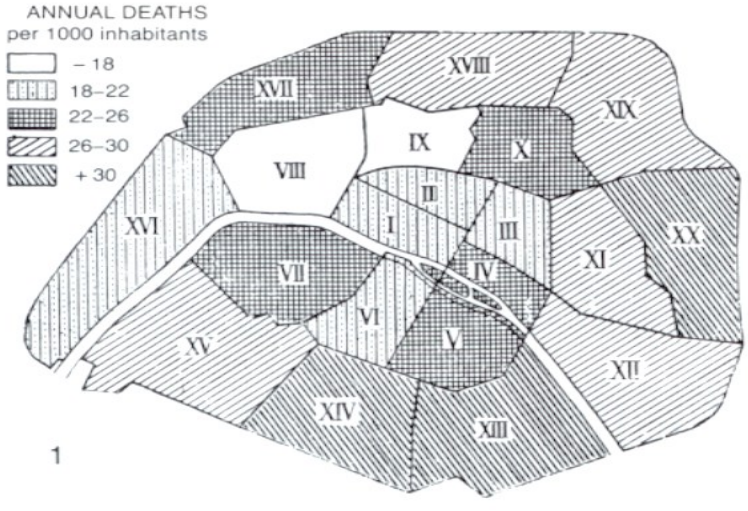
Weber’s Law: human perception is fundamentally based on relative judgments, not absolute values.

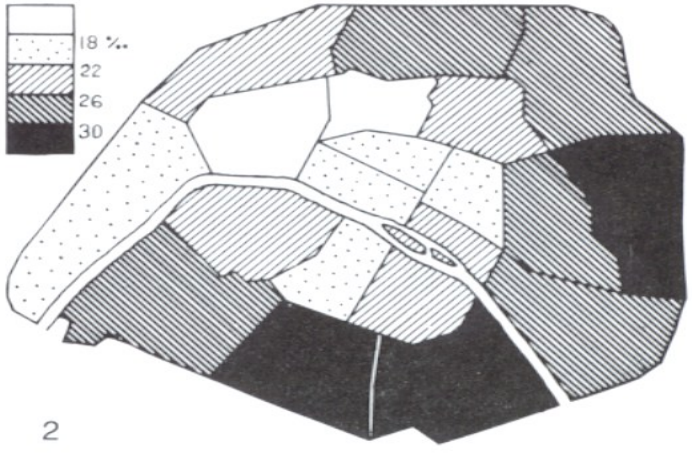
Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue					
Angle					
Texture					



Lightness & Saturation
Are ORDERED, cannot be REORDERED

ANNUAL DEATHS PER 1000 INHABITANTS, PARIS

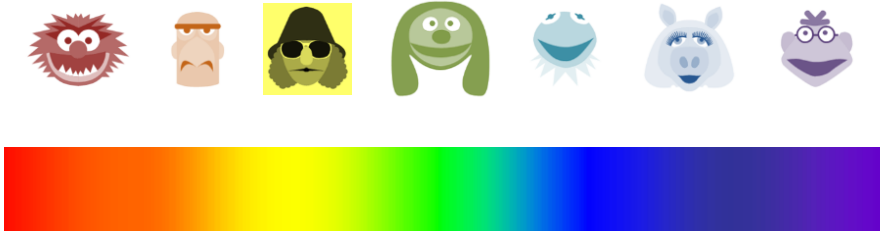




VALUES NOT ORDERED CORRECTLY ACCORDING TO SCALE
INFORMATION HAS TO BE READ POINT BY POINT

VALUES ORDERED CORRECTLY MAKE
THE IMAGE MUCH MORE USEFUL

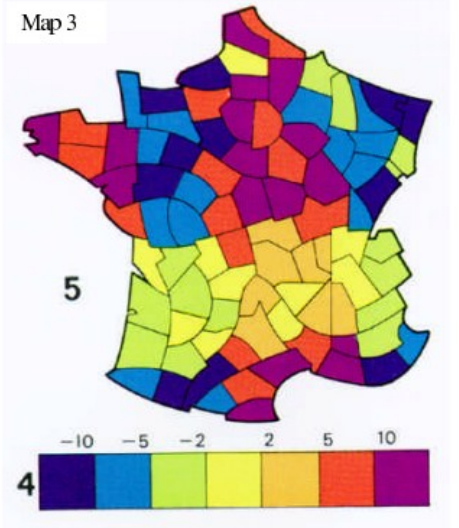
Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue	Yes	Yes	Not advisable	Not advisable	7 / 10
Angle					
Texture					



Rainbow Scale Considerations

Map 1

Map 3

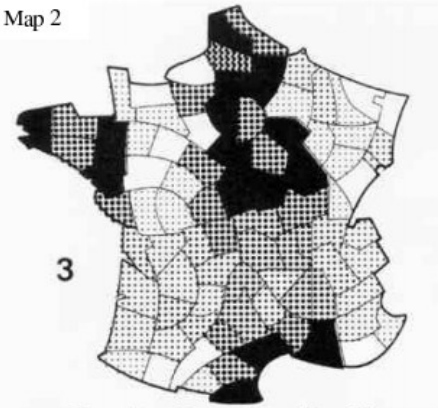


Rainbow Scale Considerations

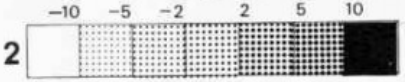
Map 1

1

Map 2

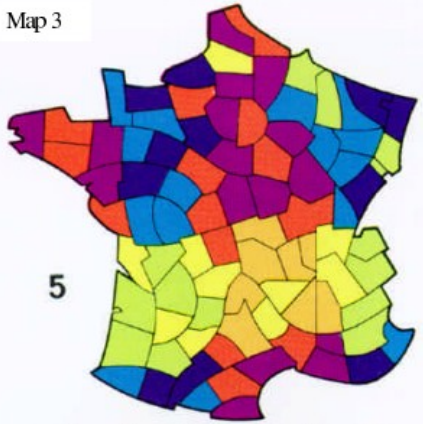


3

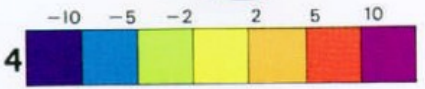


2

Map 3

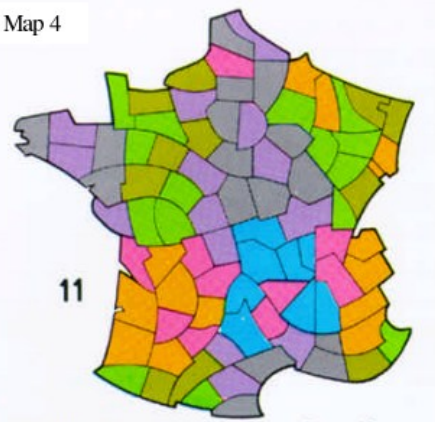


5



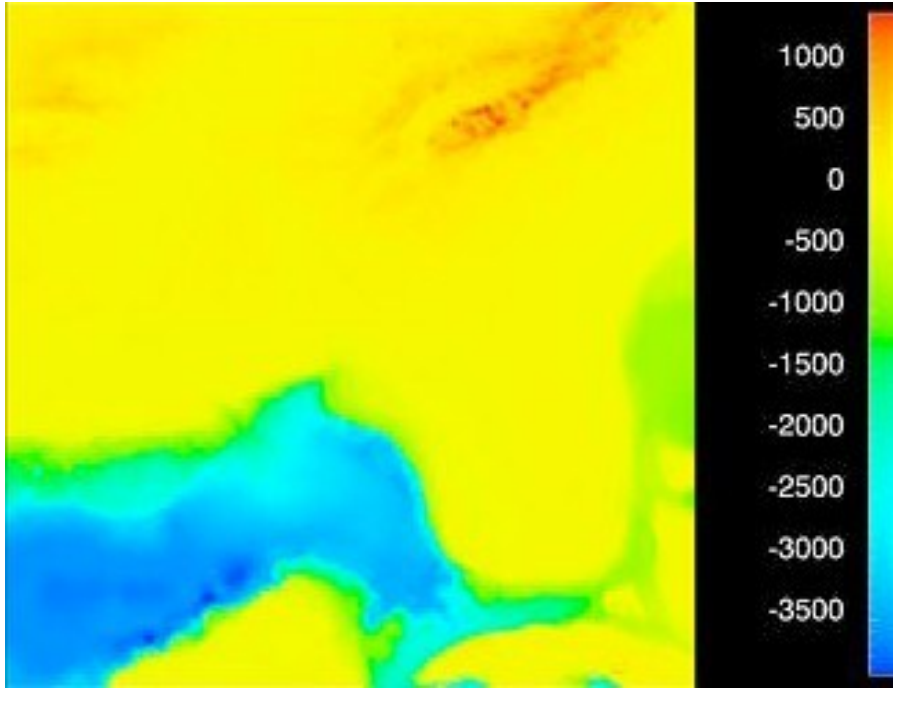
4

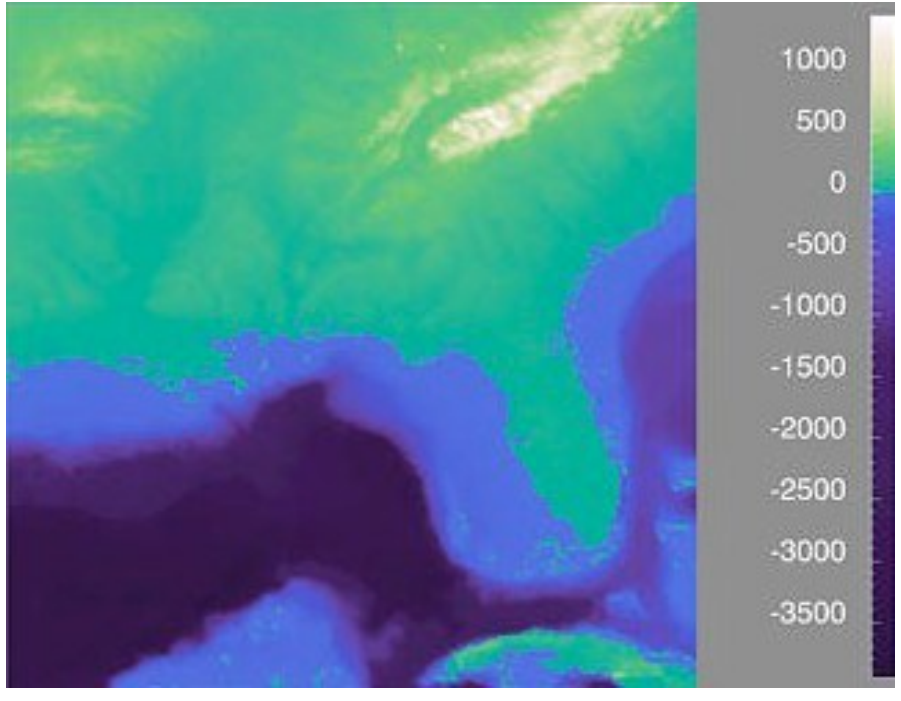
Map 4



11

10


1000
500
0
-500
-1000
-1500
-2000
-2500
-3000
-3500


1000
500
0
-500
-1000
-1500
-2000
-2500
-3000
-3500

Destroys Detail

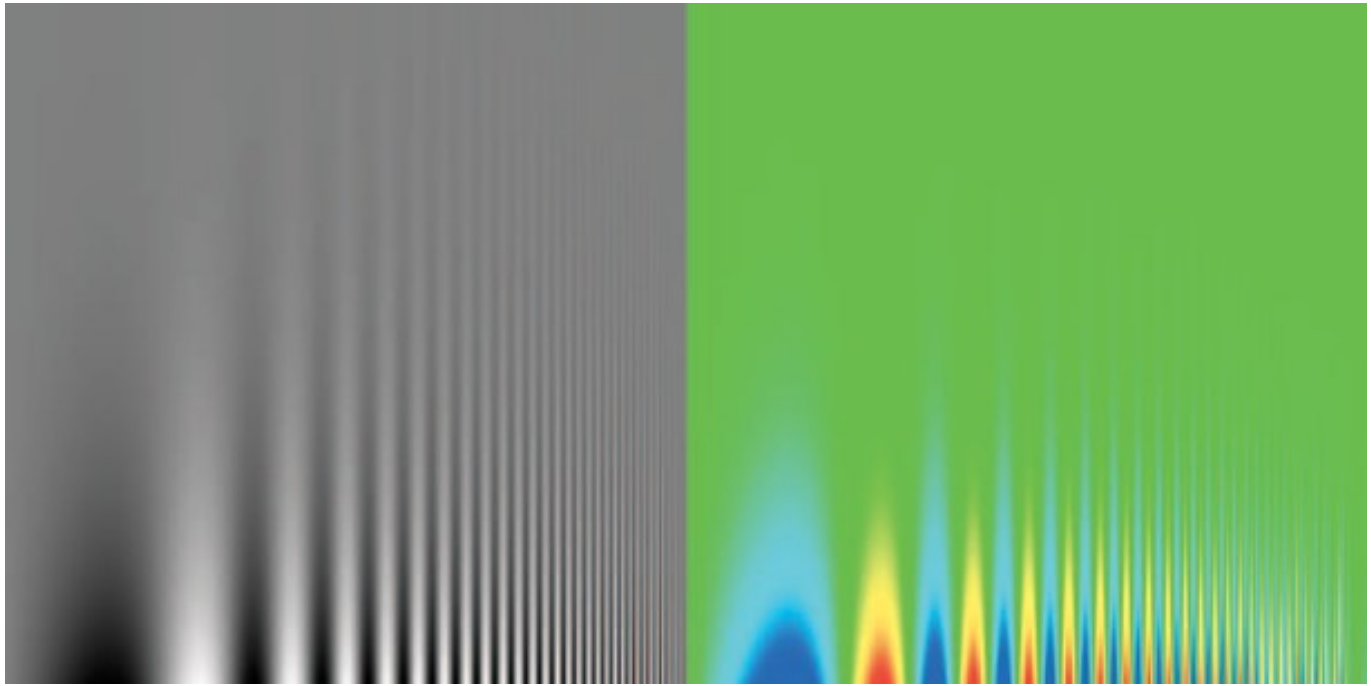
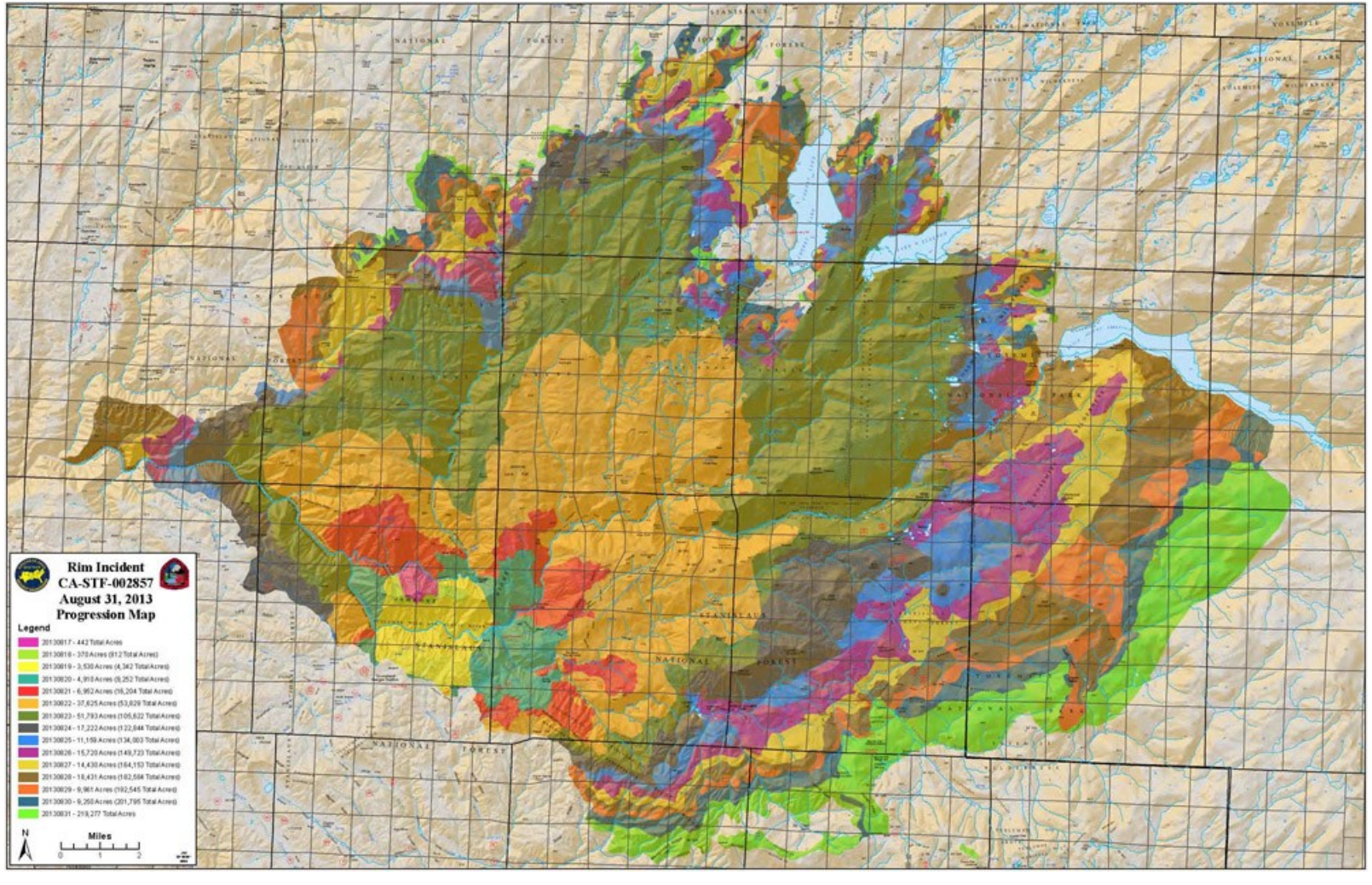
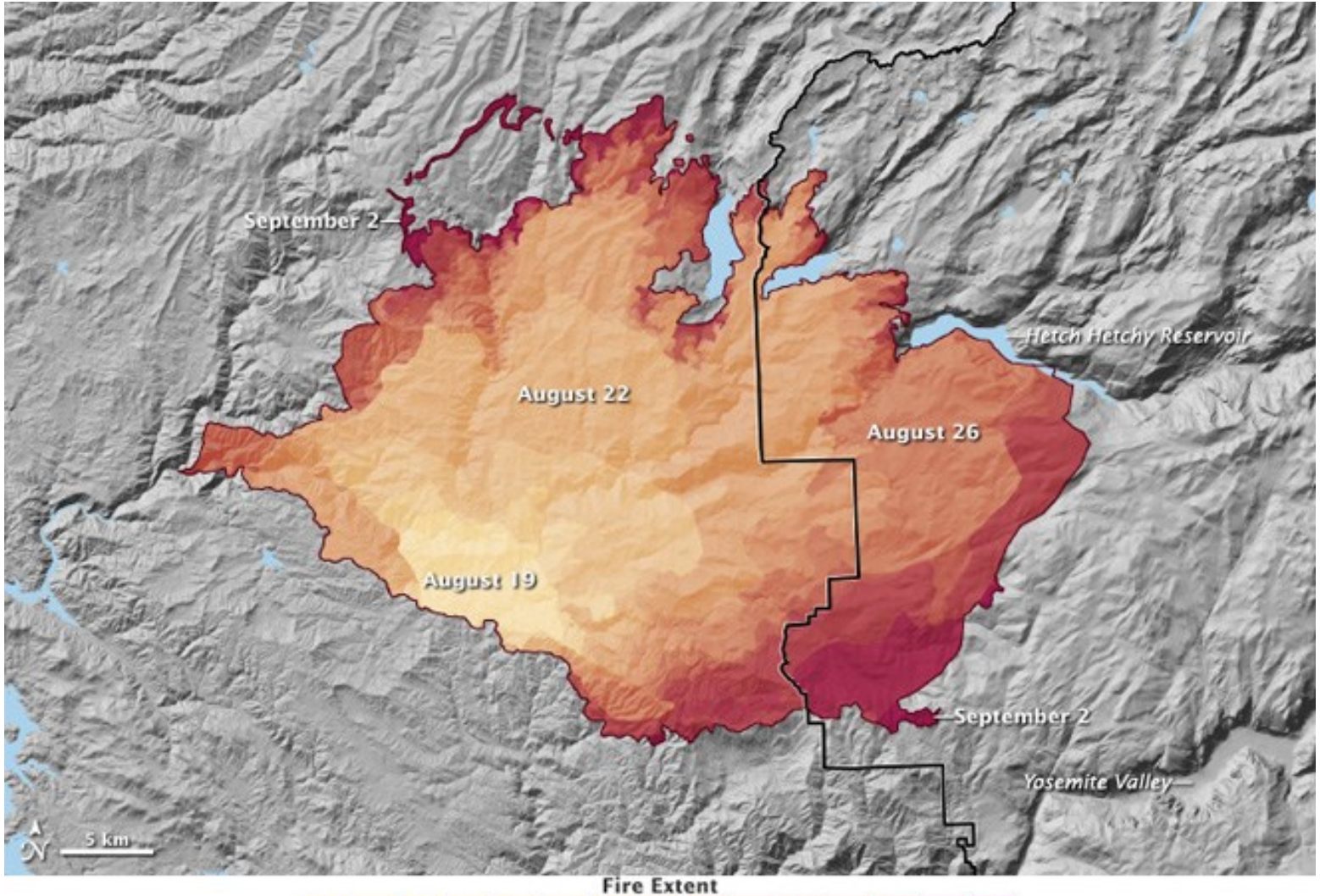


Image from *Borland & Taylor II (2007), Rainbow Color Map (Still) Considered Harmful. IEEE Visualization Viewpoints, pp 14-17.*

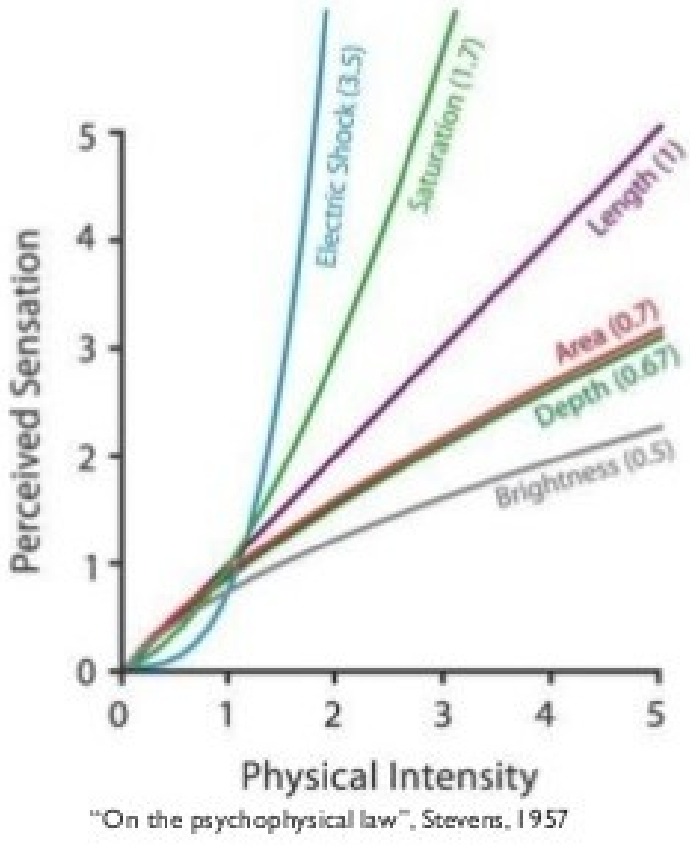

Rim Incident
CA-STF-002857
August 31, 2013
Progression Map
Legend
20130817 - 442 Total Acres
20130818 - 370 Acres (812 Total Acres)
20130819 - 3,530 Acres (4,342 Total Acres)
20130820 - 4,910 Acres (9,252 Total Acres)
20130821 - 6,952 Acres (16,204 Total Acres)
20130822 - 37,625 Acres (53,829 Total Acres)
20130823 - 51,793 Acres (105,622 Total Acres)
20130824 - 17,222 Acres (122,844 Total Acres)
20130825 - 11,159 Acres (134,003 Total Acres)
20130826 - 15,720 Acres (149,723 Total Acres)
20130827 - 14,430 Acres (164,153 Total Acres)
20130828 - 18,431 Acres (182,584 Total Acres)
20130829 - 9,961 Acres (192,545 Total Acres)
20130830 - 9,250 Acres (201,795 Total Acres)
20130831 - 219,277 Total Acres
N
Miles
0
1
2


September 2
August 22
August 26
August 19
Hetch Hetchy Reservoir
September 2
Yosemite Valley
5 km
Fire Extent
August 19
August 26
September 2

Colour

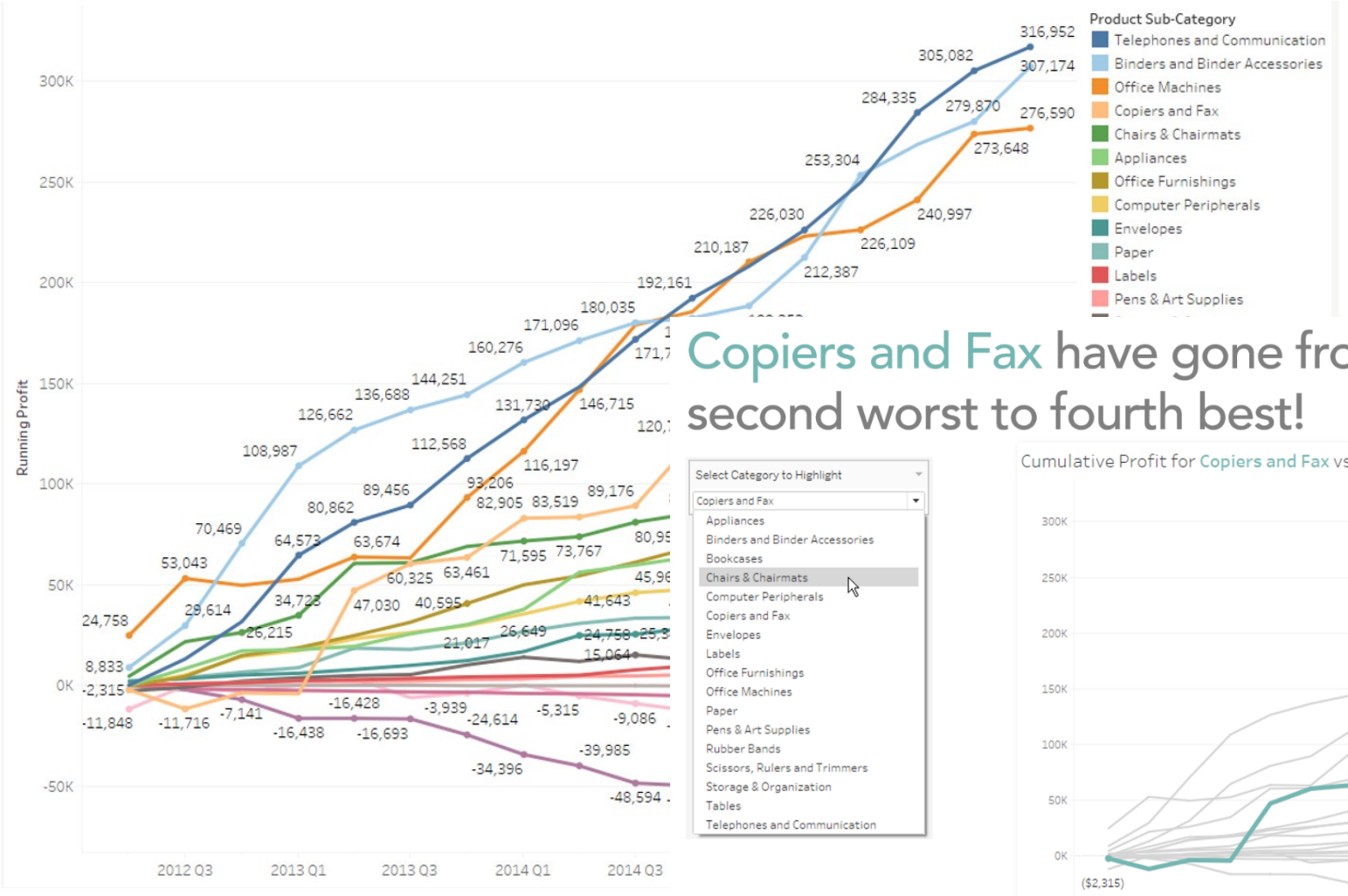
- Use Carefully! Less is more

- Stick to 5-8 colours (less is more)
  - Good contrast
  - Don’t clash
  - Works in grayscale and for colour vision deficiencies (no red & green)

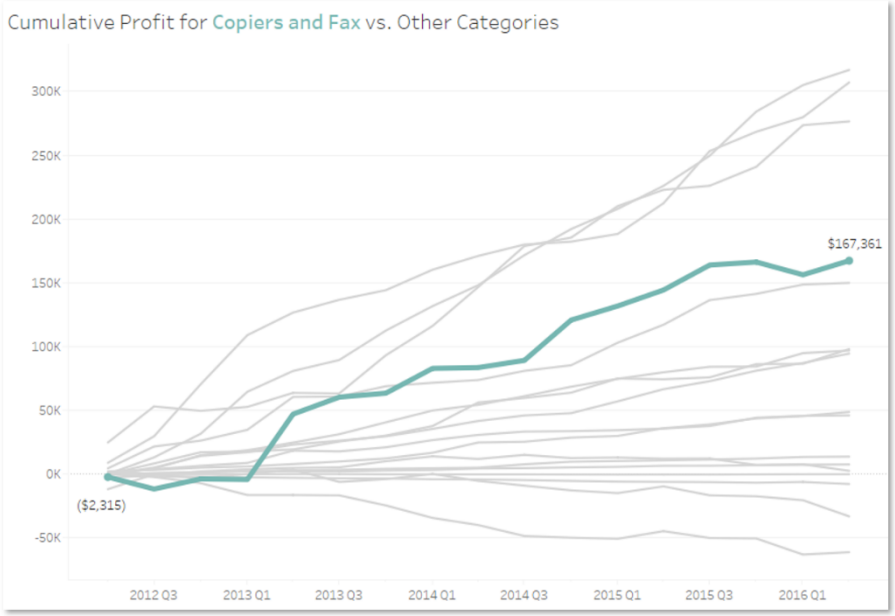


"On the psychophysical law", Stevens, 1957

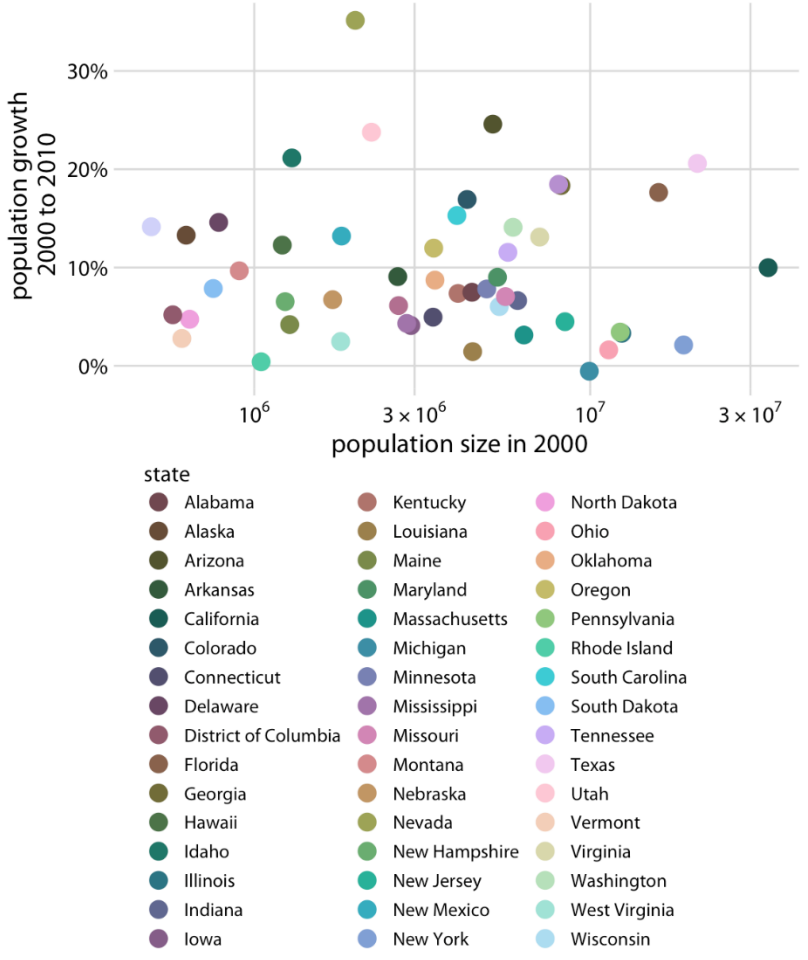
Too much colour



Copiers and Fax have gone from second worst to fourth best!

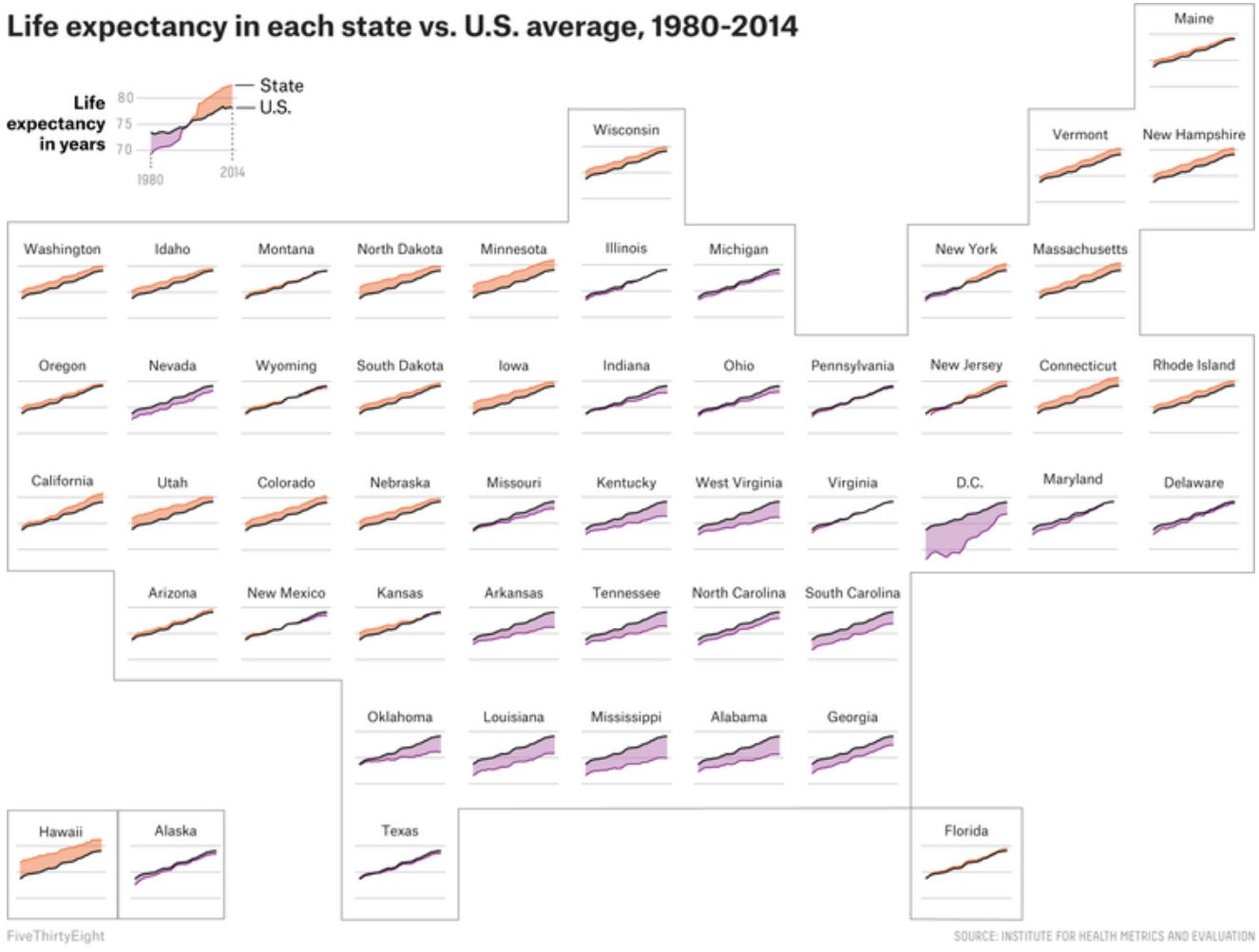


Too many colours



from https://clauswilke.com/dataviz/color-pitfalls.html

Life expectancy in each state vs. U.S. average, 1980-2014



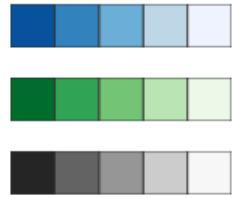
FiveThirtyEight

SOURCE: INSTITUTE FOR HEALTH METRICS AND EVALUATION

Colour Scales

Sequential (dark to light or light to dark)

- Quantitative data or ordered qualitative data
- Single or multiple hues

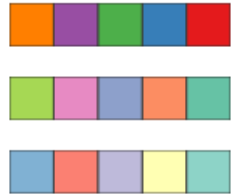


Diverging (dark in 1 hue to light to dark in a different hue)

- Quantitative data or ordered qualitative data
- Use if there is a meaningful middle point

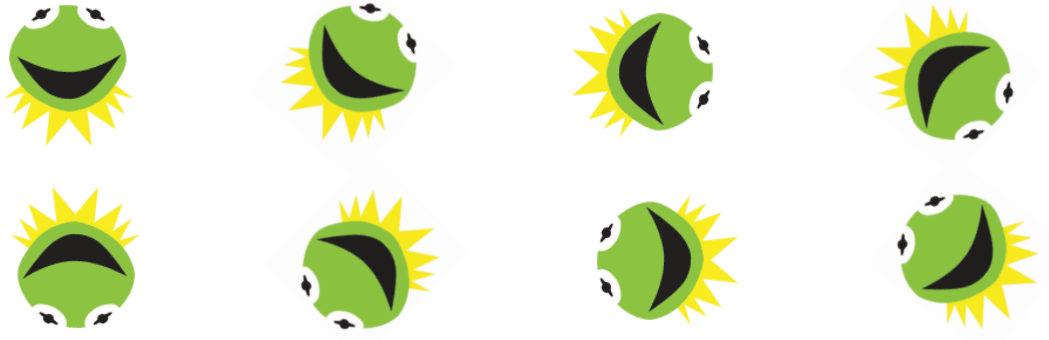
Categorical

- Qualitative data
- Give hues different brightness so that they appear distinct in grayscale
- Be careful with red & green

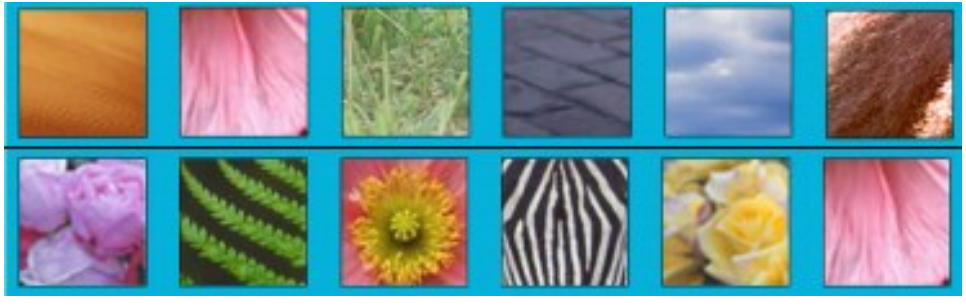


Color scales from ColorBrewer 2.0

Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue	Yes	Yes	Not advisable	Not advisable	7 / 10
Angle	Yes	Yes	No	No	4 / 8
Texture					



Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue	Yes	Yes	Not advisable	Not advisable	7 / 10
Angle	Yes	Yes	No	No	4 / 8
Texture	Yes	Yes	No	No	Infinite

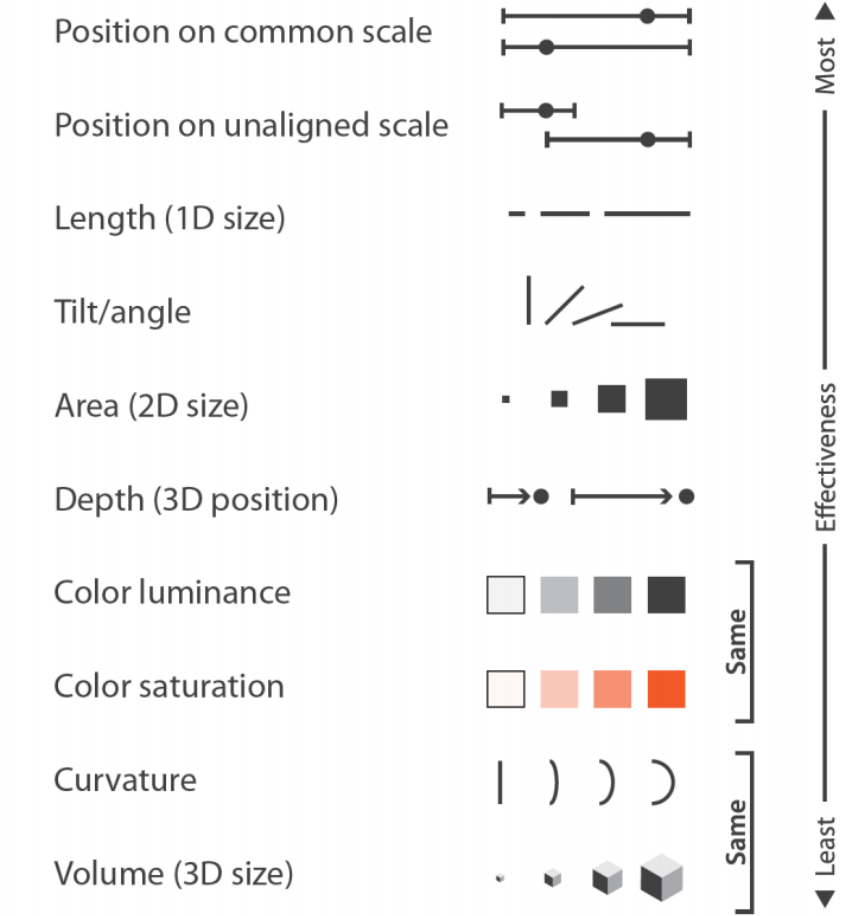


Variable	Selective	Associative	Quantitative	Order	Length
Size	Yes	Yes	Yes (1D), Mostly (2D), Not likely (3D)	Yes	5/20
Position	Yes	Yes	Yes	Yes	Infinite
Shape	< 5	< 5	No	No	5 / Infinite
Lightness	Yes	Yes	No	Yes	7 / 10
Saturation	Yes	Yes	No	Yes	7 / 10
Hue	Yes	Yes	Not advisable	Not advisable	7 / 10
Angle	Yes	Yes	No	No	4 / 8
Texture	Yes	Yes	No	No	Infinite

Semiology of Graphics by Jacques Bertin. 1967.

M.S.T. Carpendale. **Considering Visual Variables as a Basis for Information Visualisation**. Research report 2001-693-16, Department of Computer science, University of Calgary, 2003.

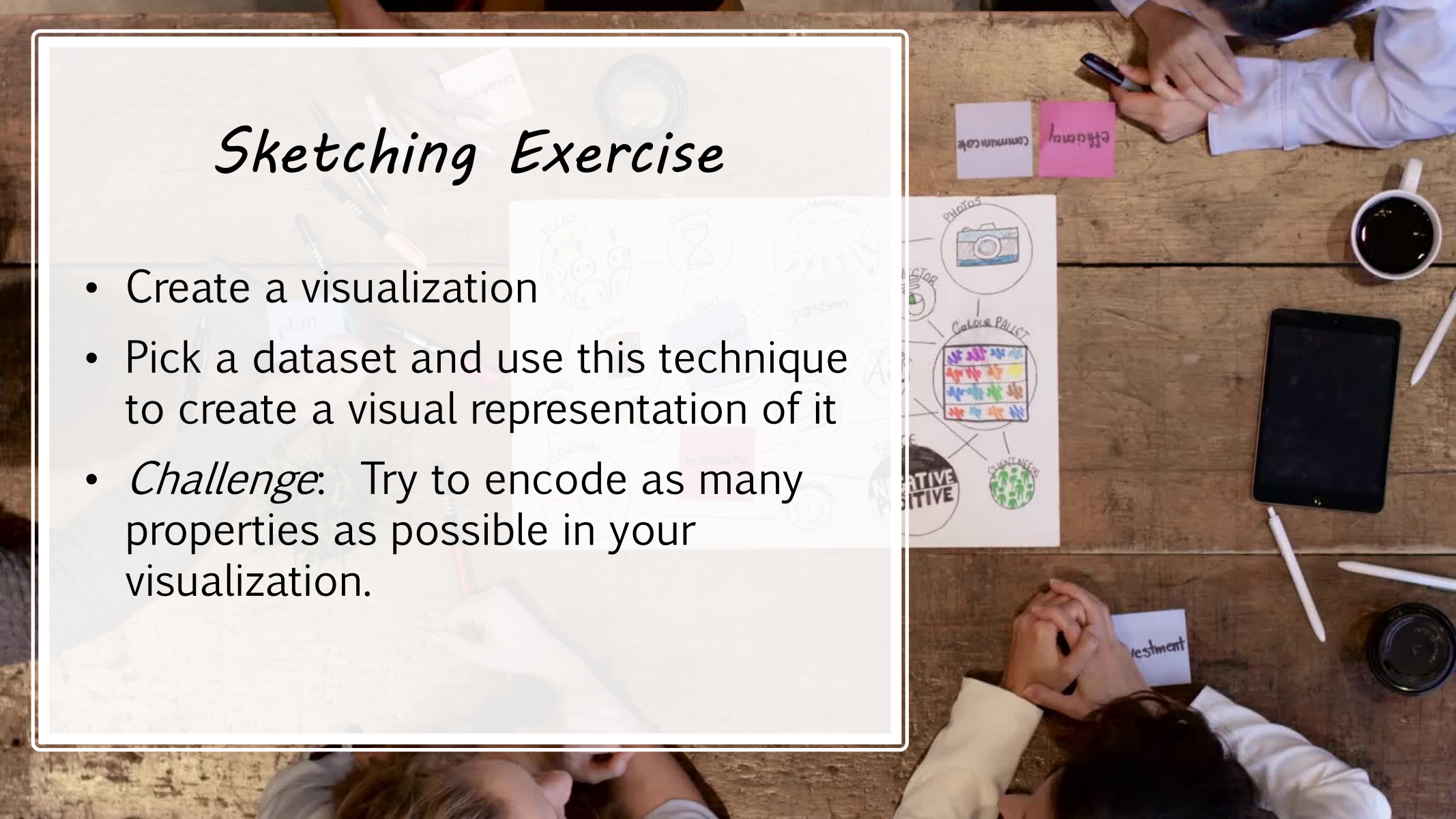
Magnitude Channels: Ordered Attributes



Identity Channels: Categorical Attributes

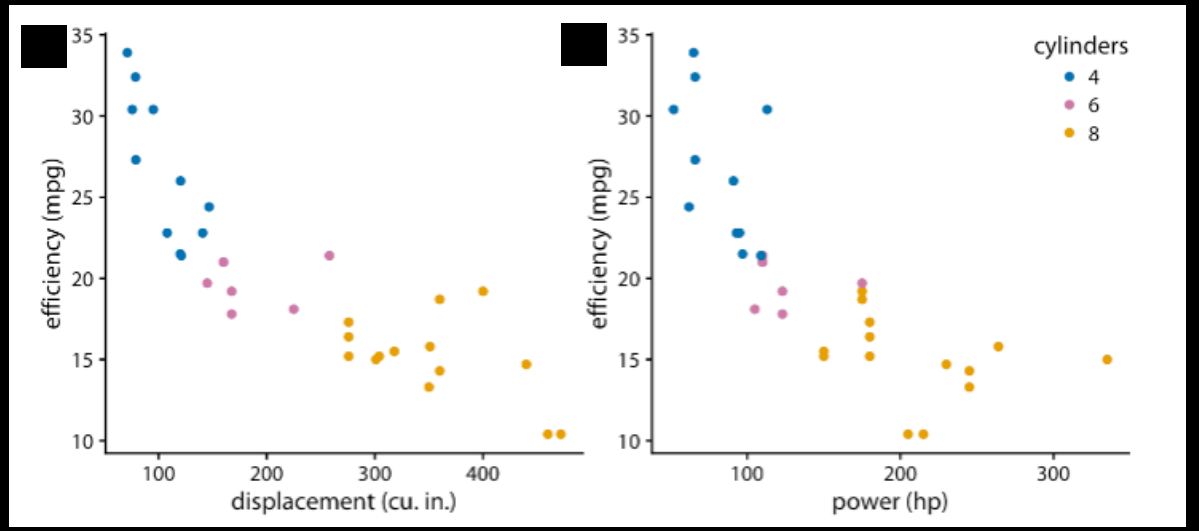
Sketching Exercise

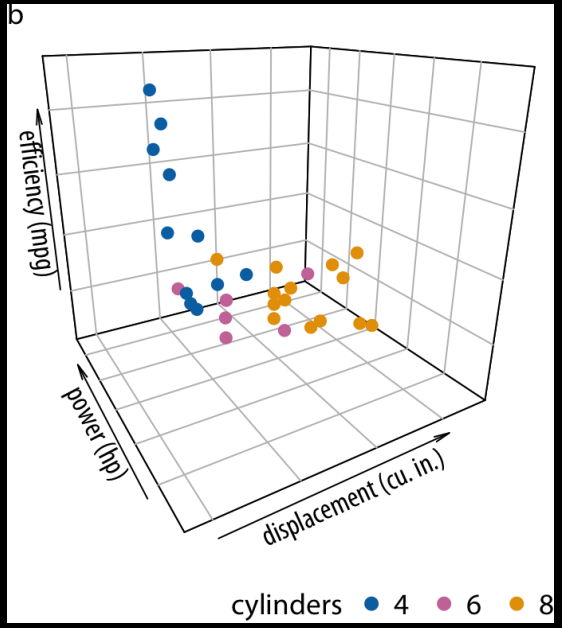
- Create a visualization
- Pick a dataset and use this technique to create a visual representation of it
- *Challenge*: Try to encode as many properties as possible in your visualization.

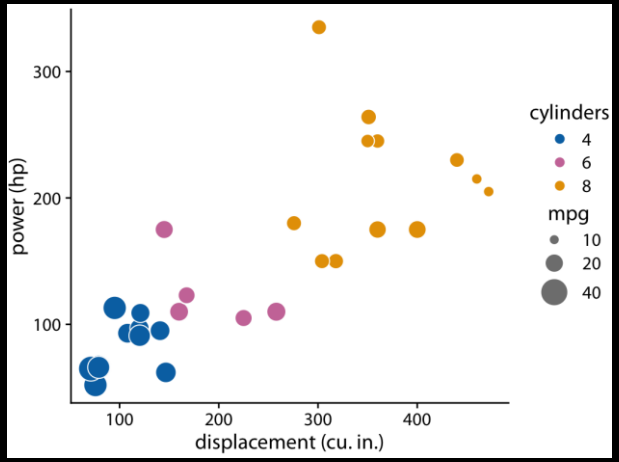


Quad.	Neighbourhood	Type	Rooms Available	# People at Residence	Rent	Dam. Dep.	Transit	Walking Time to Uni	Date Available	Pets	Furn-ished	Smok-ing	Gender	Private Entr.	Private Bath	Includes Utilities
NW	Dalhousie	Room	1	2	500	500	LRT	NA	01-May	No	Yes	No	Any	Yes	No	No
NW	Brentwood	House	4	4	1750	1750	LRT		01-Jul	No	Yes	No	Any	Yes	Yes	Yes
NW	Capitol Hill	Basemt	2	2	1150	1150	LRT	30	01-Apr	No	No	No	Any	Yes	Yes	Shared
NW	Varsity	Room	2	4	370	370	Bus	35	28-Mar	No	Yes	No	Any	Yes	No	Yes
NW	St Andrews	Condo	3	5	575	575	Bus	20	01-May	No	Yes	No	Male	No	No	Shared
NW	Ranchlands	Room	1	3	600	600	Bus	NA	01-May	No	Yes	No	Male	Yes	No	Yes
NW	Uni. Heights	Apt	1	1	1100	1100	Bus	10	01-May	No	No	No	Male	Yes	Yes	Heat+Water
NW	Varsity	Apt	1	1	919	699	Bus	40	01-Apr	Cats	No	No	Any	Yes	Yes	Heat+Water
NW	Varsity	Room	1	4	440	100	LRT	NA	01-May	No	Yes	No	Female	No	Yes	Yes
NW	Citadel	Room	1	3	550	550	Bus	NA	01-May	Other	No	Neg.	Any	No	No	Shared
NW	Brentwood	Basemt	1	1	500	500	LRT	7	01-Apr	No	Yes	No	Male	Yes	Yes	Yes
NW	Capitol Hill	Room	1	4	500	350	LRT	12	01-May	Cats	Yes	No	Female	Yes	Yes	Yes
NW	Briar Hill	House	1	3	500	200	LRT		01-May	No	Yes	No	Any	No	Yes	Yes
NW	Banff Trail	Room	1	2	450	460	LRT	14	01-May	No	Partial	No	Female	No	No	Yes
SW	Downtown	Apt	3	3	1989	699	LRT	60	18-Mar	Yes	No	No	Any	Yes	Yes	Yes
SW	Downtown	Apt	1	1	1209	499	LRT	60	18-Mar	No	No	No	Any	Yes	Yes	Yes
SW	Glamorgan	Apt	1	2	700	350	Bus	NA	01-Jun	No	No	No	Any	Yes	No	Yes

Avoid 3D
Better solution



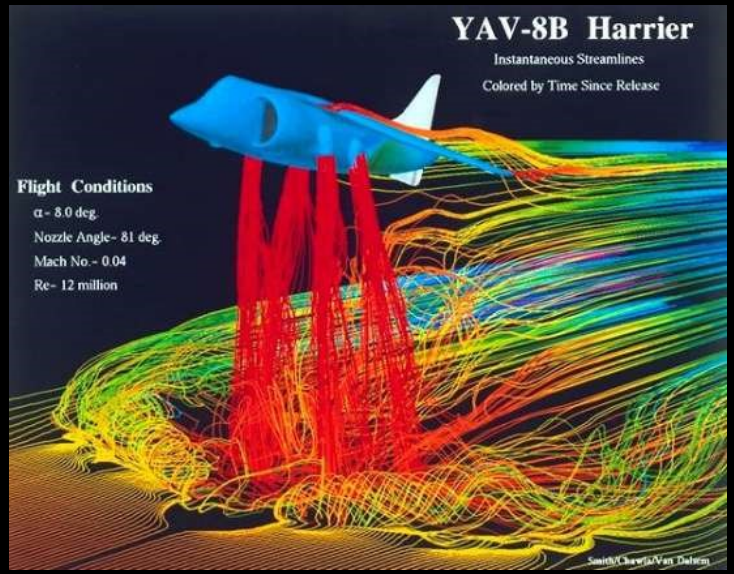


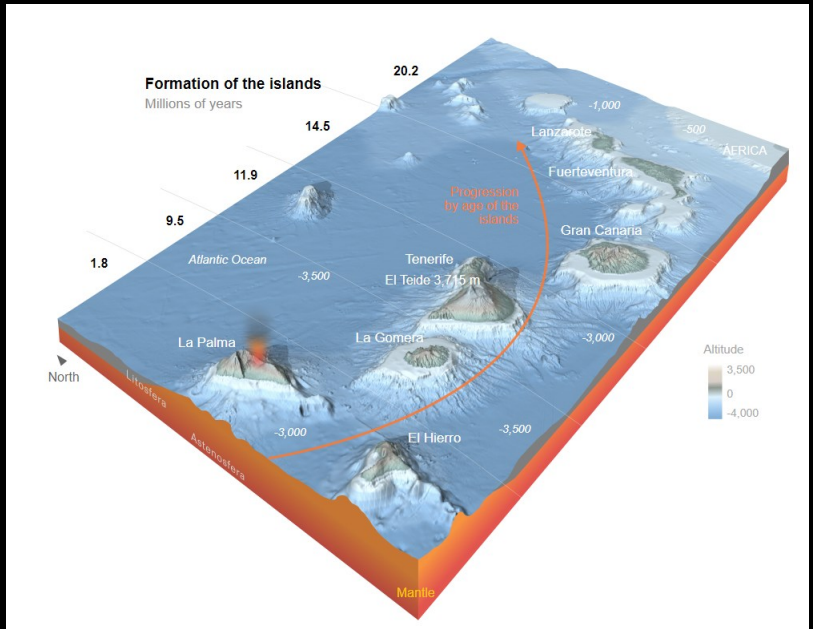


From https://clauswilke.com/dataviz/no-3d.html

When to use 3D?

- When your data has 3D position



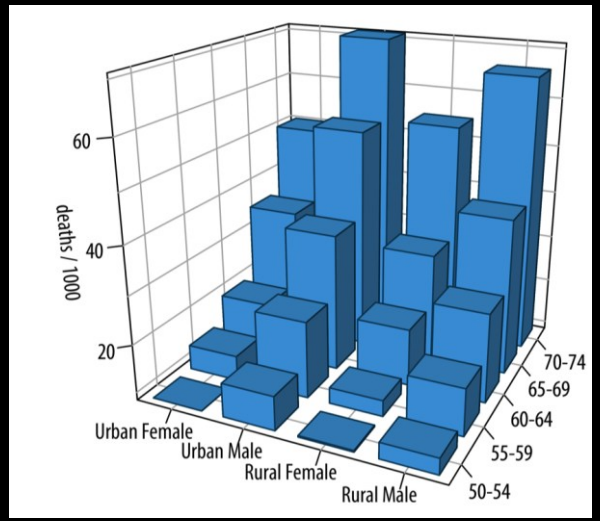


Small Multiples

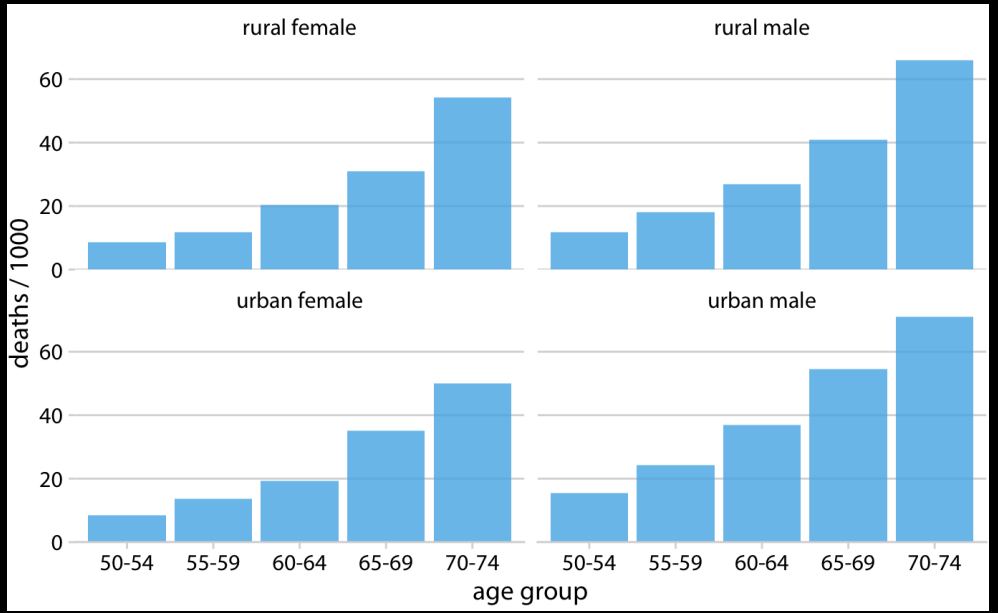
Break complicated charts into smaller, simplified charts

Use alignment and repetition to highlight differences

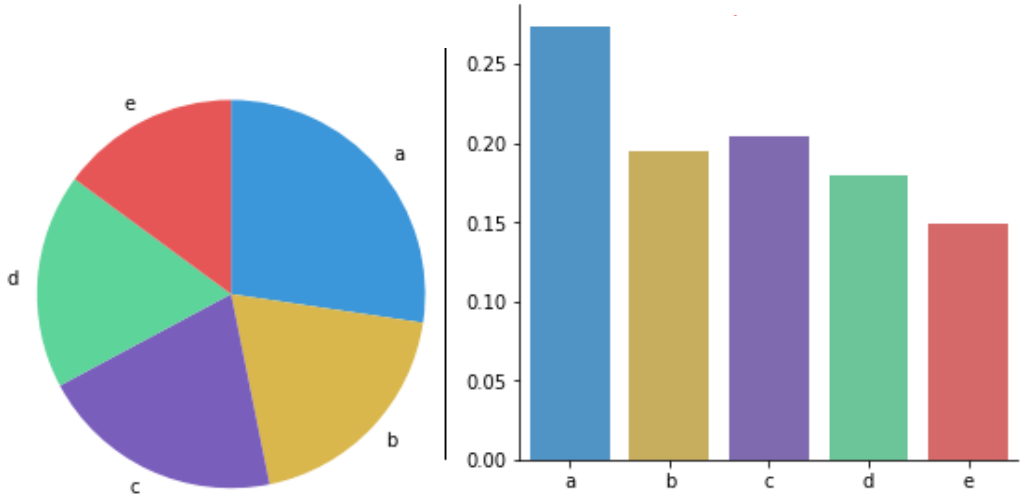
Needs appear in order and same scales, sizes, & shape



From https://clauswilke.com/dataviz/no-3d.html



Pie Charts – Use Carefully



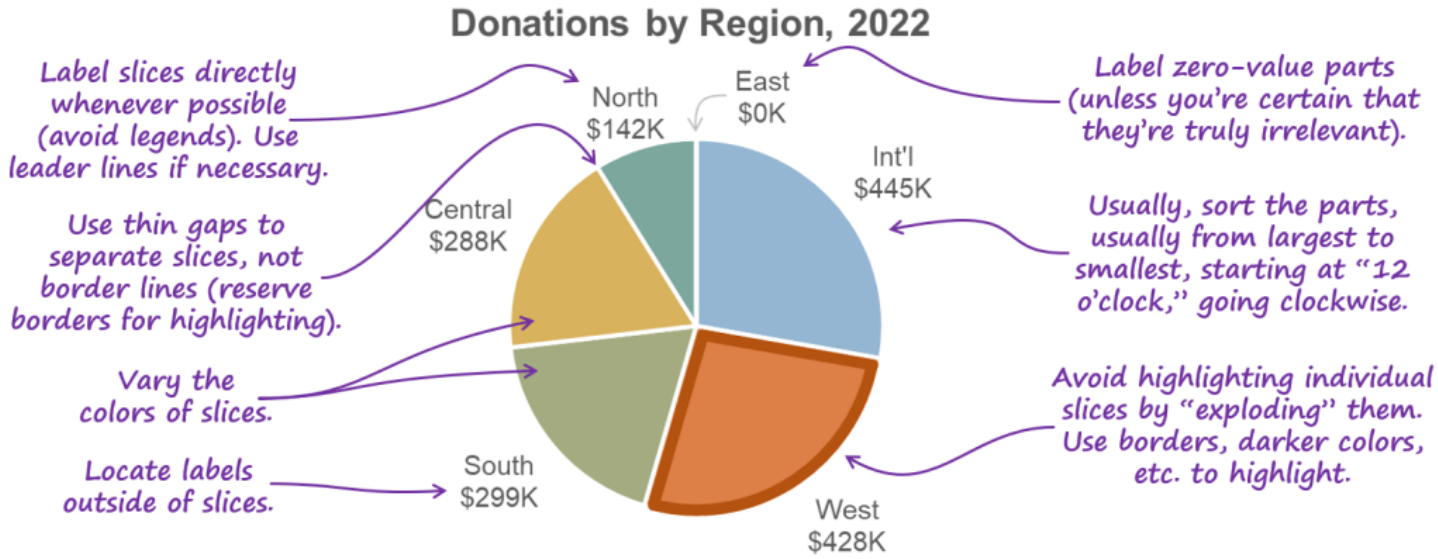


Chart Types

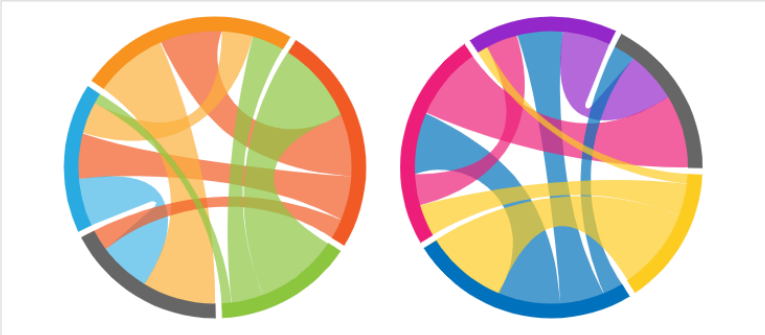
https://datavizcatalogue.com/

https://flowingdata.com/chart-types/

Image from https://datavizcatalogue.com/

Data Vis Catalog

Chord Diagram



Description

This type of diagram visualises the inter-relationships between entities. The connections between entities are used to display that they share something in common. This makes Chord Diagrams ideal for comparing the similarities within a dataset or between different groups of data.

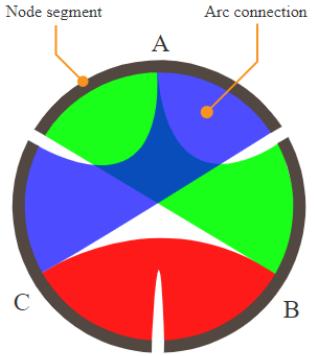
Nodes are arranged along a circle, with the relationships between points connected to each other either through the use of arcs or Bézier curves. Values are assigned to each connection, which is represented proportionally by the size of each arc. Colour can be used to group the data into different categories, which aids in making comparisons and distinguishing groups.

Over-cluttering becomes an issue with Chord Diagrams when there are too many connections displayed.

Functions

Comparisons Relationships

Anatomy



	A	B	C
A		10	10
B	10		10
C	10	10	

A —10→ B
B —10→ C
C —10→ A

Data Visualization Tools

Mostly free

1. Preparing Data
2. Visualization

Data Tools:
Data Wrangler / Trifacta

DataWrangler alpha

TRIFACTA Wrangler

http://vis.stanford.edu/wrangler

https://www.trifacta.com

- Interactive tool for cleaning & rearranging
- Suggests changes
- Wrangler: web tool – data to external site (1000 lines)
- Import: text, CSV, JSON
- Export: CSV, JSON, TDE (Tableau)

Data Tools: Open Refine

http://openrefine.org/

- Consolidate spelling
- Auto-detect outliers
- Sorting & filtering
- Auto-suggests changes
- Import: Excel, XML, JSON, RDF, CSV
- Export: Excel, CSV, ODF, HTML

Data Tools: Tabula

http://tabula.nerdpower.org/

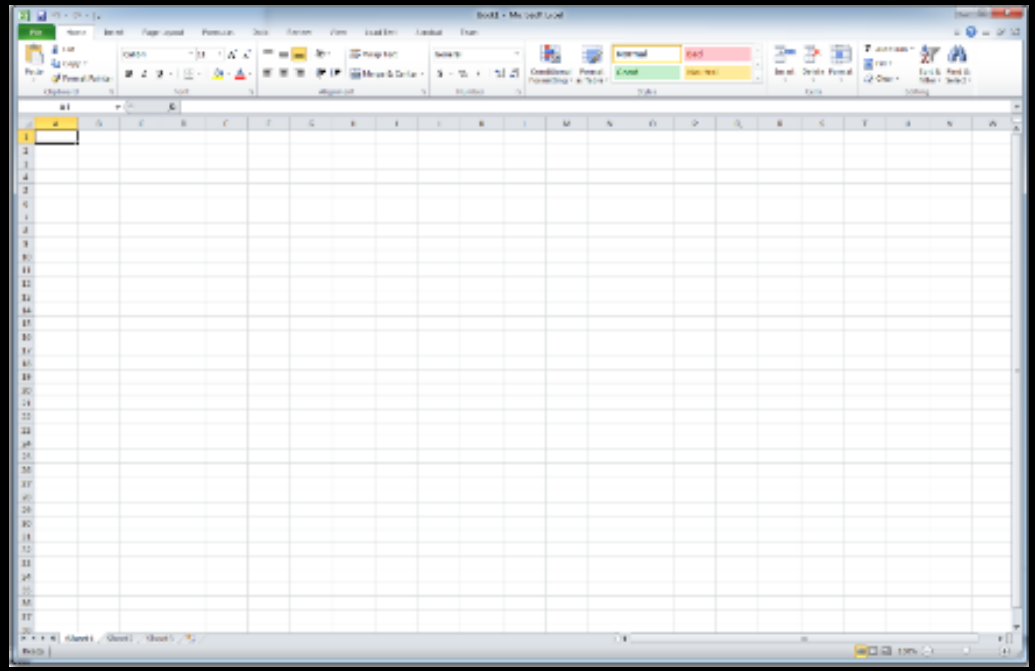
- Extract data from PDFs
- Stand-alone app for Windows/Mac
- Interactively select table
- Output: CSV, Excel

Visualization Tools

- General Purpose
- Special Purpose
  - Text Analysis
  - Sets
  - Maps
  - Networks / Graphs
- Bespoke
- Colour

Excel

- Simple charts
- Poor defaults (getting better)
- Hard to customize

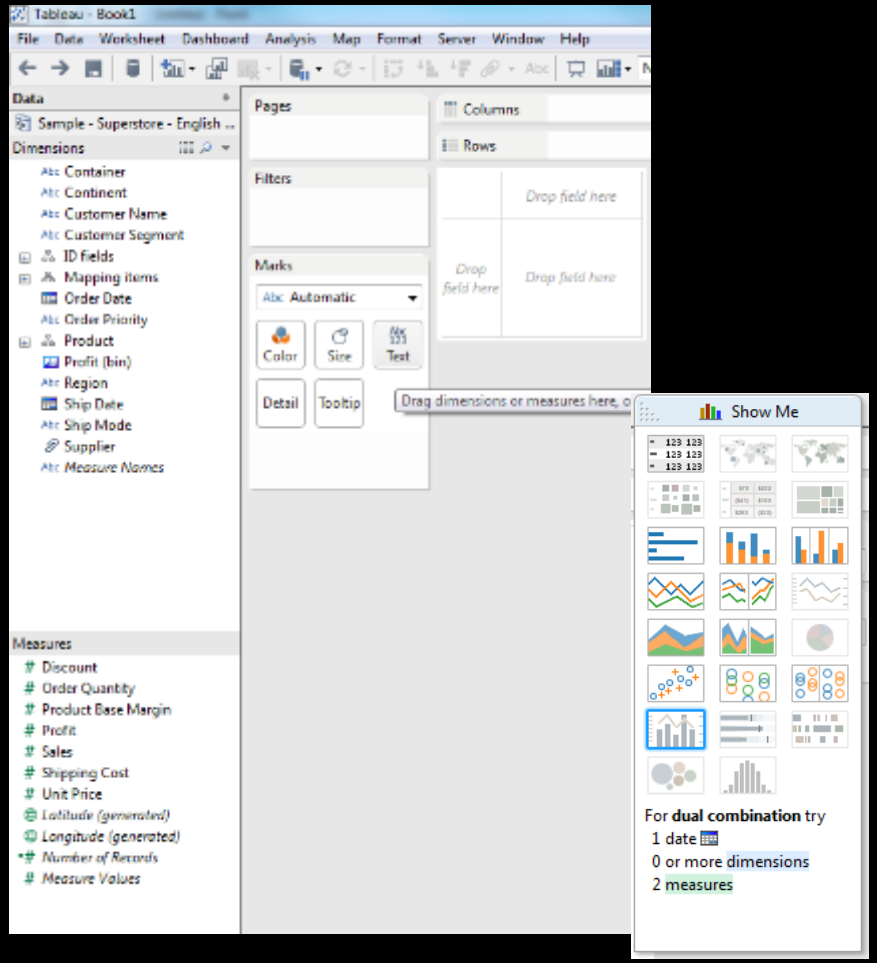


VIS Tools: Tableau

http://www.tableau.com/

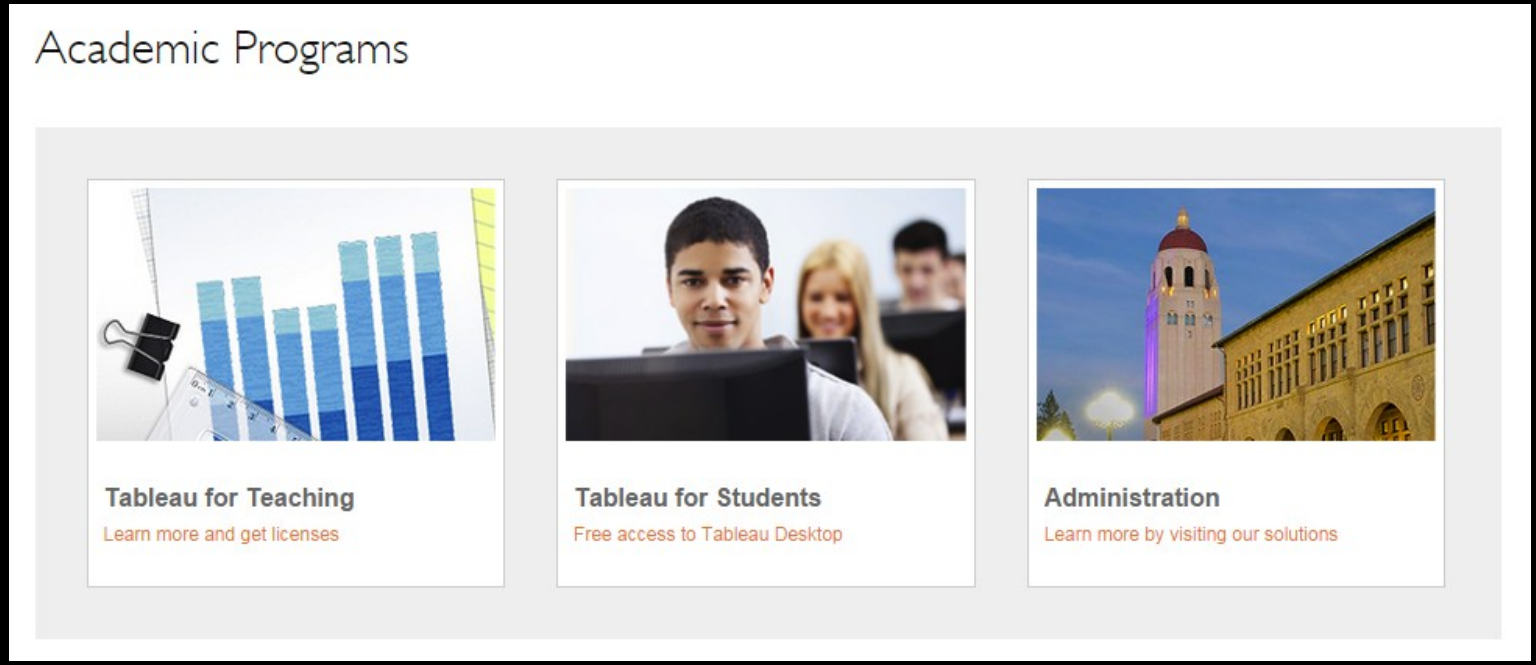
Strengths:

- Many chart types
- Interactive web output
- Access to underlying data
- Many data sources (live)
- Drag & drop – easy to experiment
- Maps
- Good defaults
- Link visualizations
- R can plugin
- Academic Program - Free for students



Student Program

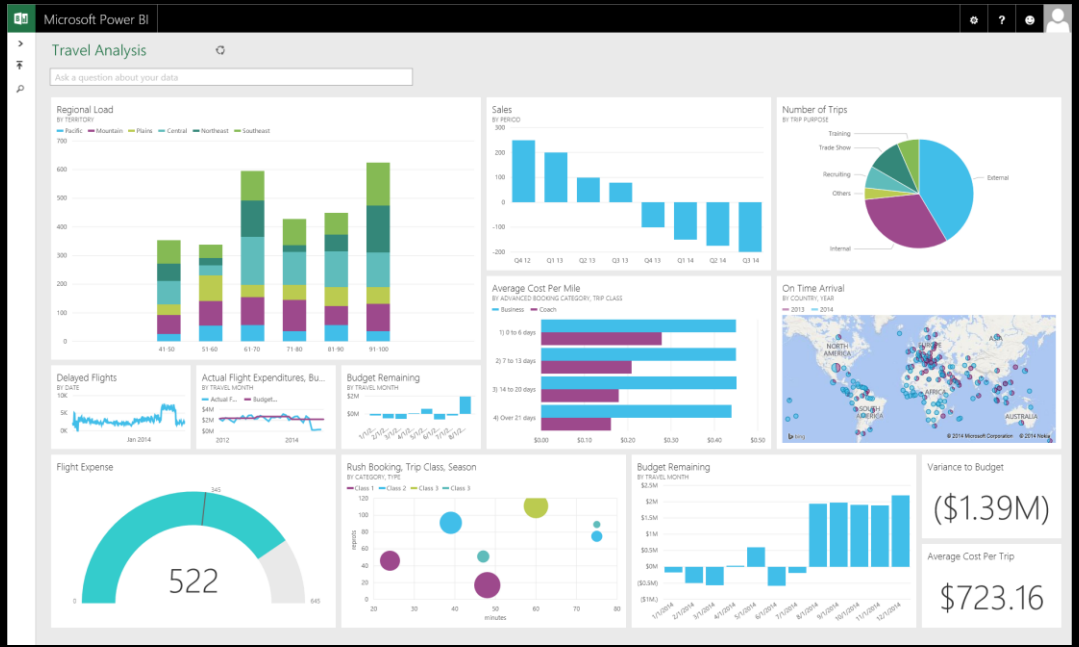
- Tableau desktop **free** for post-secondary students.
- http://www.tableau.com/academic



VIS Tools : POWER BI

https://powerbi.microsoft.com/en-us/

- Similar capabilities as tableau
- Can build plugins
- Better data modeling
- Not as customizable
- Exploration not as easy
- Lots of menus
- Less data capacity than Tableau



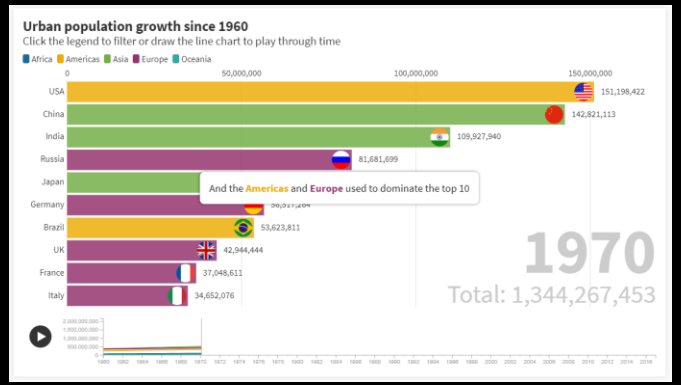
https://www.em360tech.com/microsoft-power-by-dashboard/

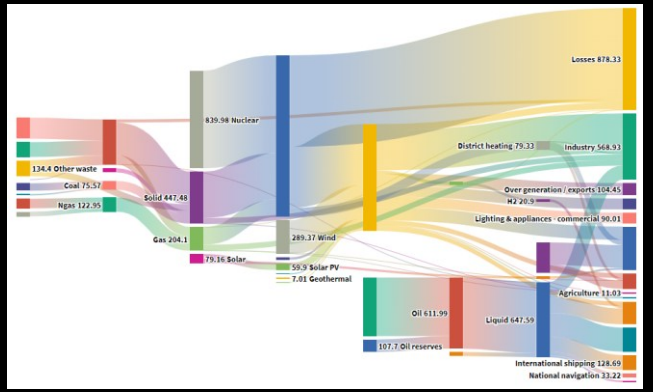
VIS Tools: Flourish

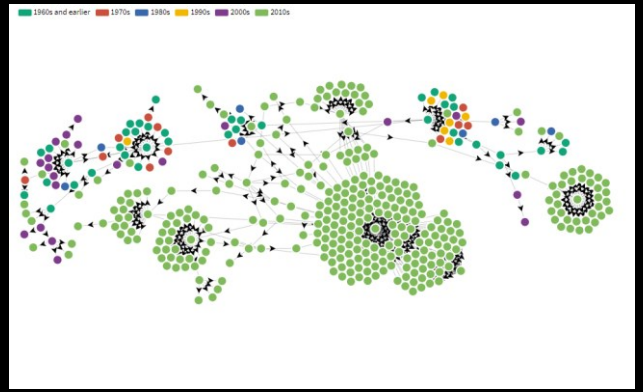
https://flourish.studio/

- From the data journalism community; focus on storytelling
- Interactive visualizations that can be embedded in website

- Free (data shared publicly), paid otherwise
- Stick to relatively small datasets



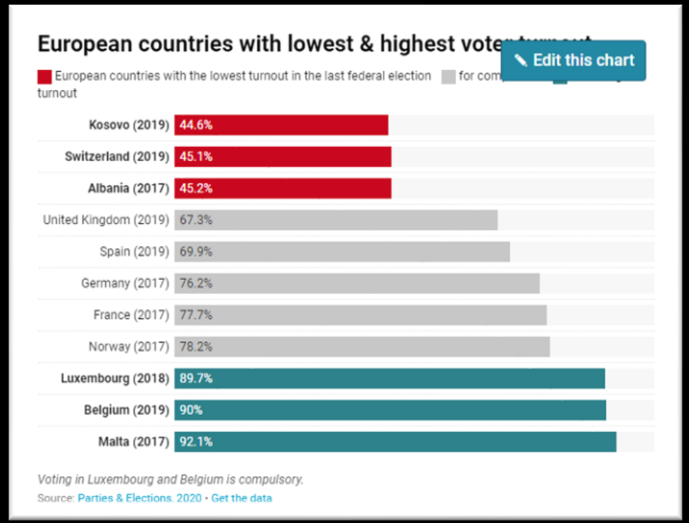


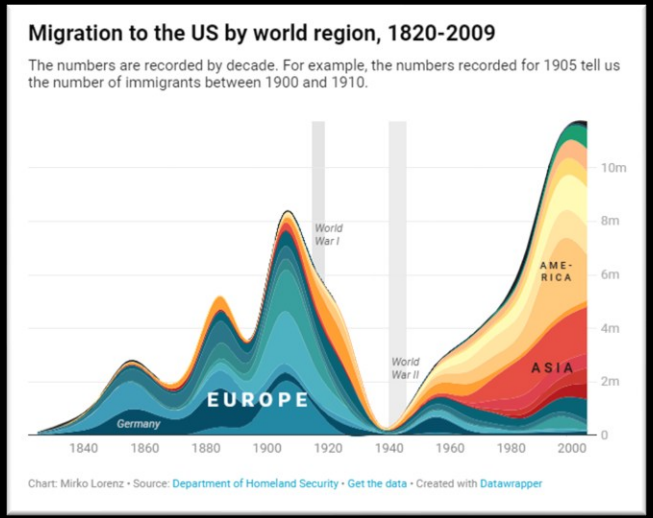


VIS Tools: Datawrapper

https://www.datawrapper.de/

- 20+ chart types
- Variety of mapping types (choropleth, symbols, locator)
- Free, but results have “created with datawrapper watermark”.

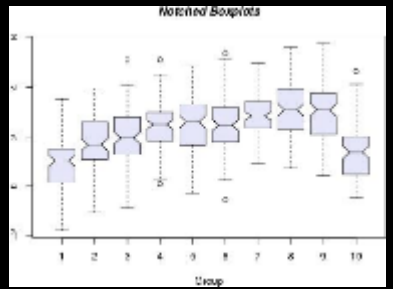


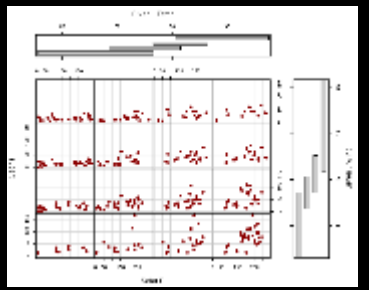


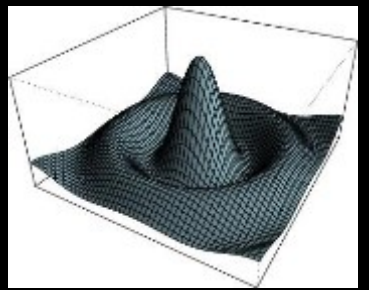
VIS Tools: R

http://www.r-project.org/

- Open-source software for statistics & graphics
- All sorts of advanced stats
  - Regression, linear/nonlinear models, time series analysis, clustering, nonparametric tests
- Data wrangling
- Charts & Plots
- Command line*
- Many add-ons (> 4400)
- 60+ Resources for R
  http://www.computerworld.com/article/2497464/business-intelligence/60-r-resources-to-improve-your-data-skills.html







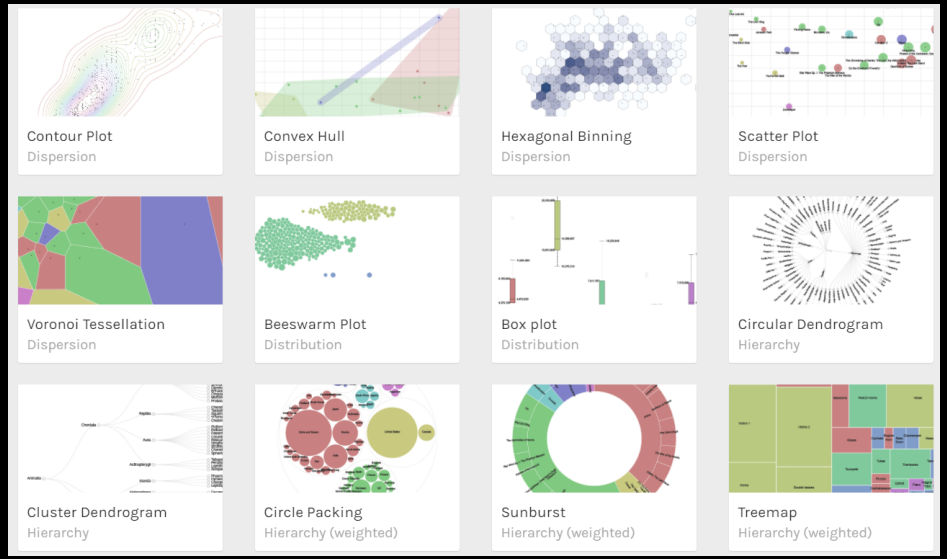
VIS Tools: RAWGraphs

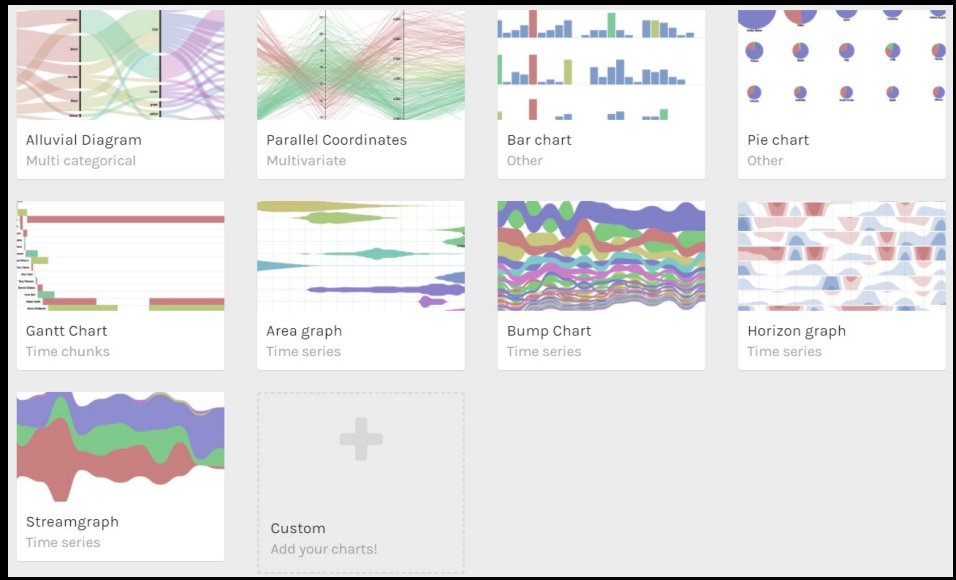
RAWGraphs

The missing link between spreadsheets and data visualization.

https://rawgraphs.io/

- Create SVG graphics
- Data not uploaded (so remains private)
- 20+ chart types

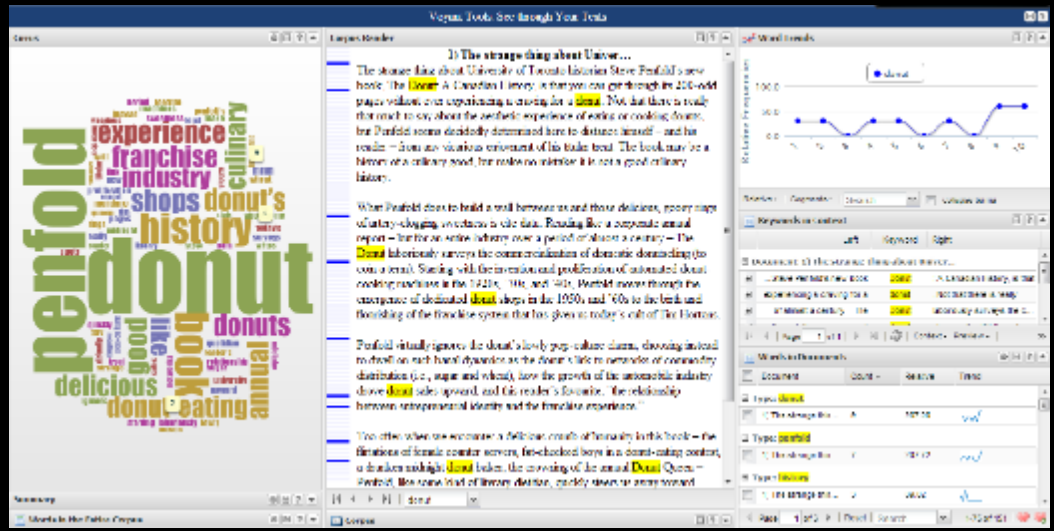




Text Analysis: Voyant

http://voyant-tools.org/

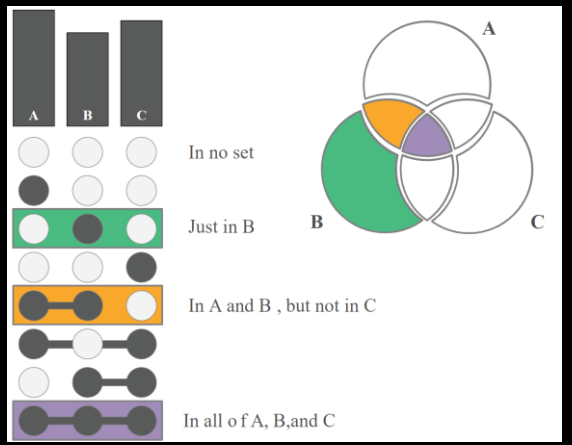
- Import: txt, HTML, XML, PDF, RTF, & Word
- Lexical analysis
  - frequency and distribution
- Export: XML, tsv, html widgets

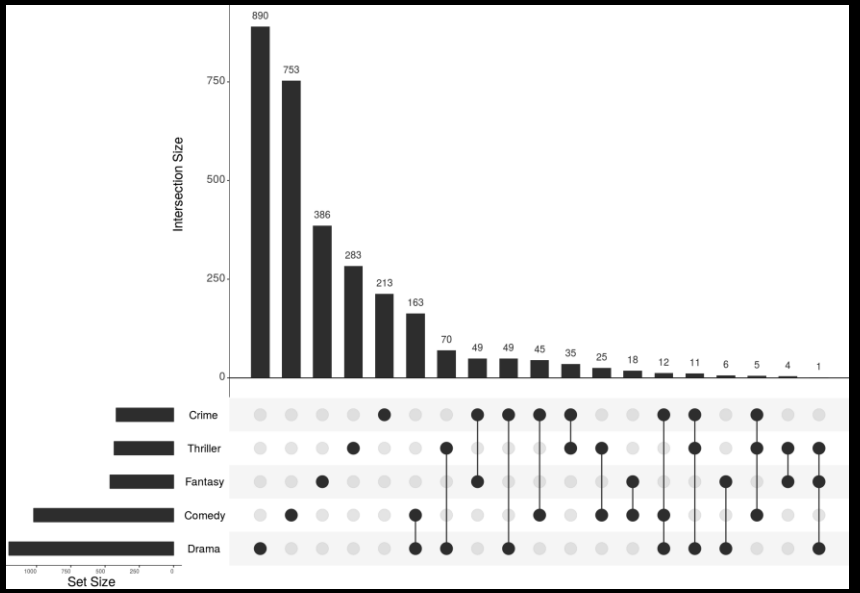


Sets: UpSet

https://upset.app/implementations/

- Alternative to Venn diagram when you have 4-30 sets.
  - With fewer than 4, use a Venn diagram (e.g., https://bioinfogp.cnb.csic.es/tools/venny/index.html).
- Many implementations
  - Web tool, R library, javascript library, etc



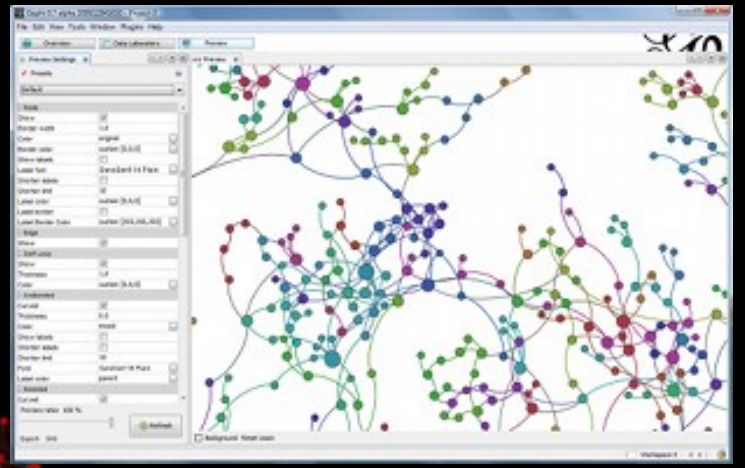


Images from https://upset.app/

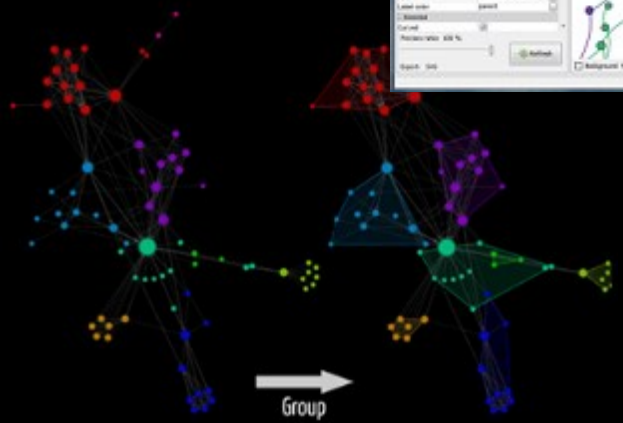
Graphs/Networks: Gephi

Gephi
makes graphs handy

http://gephi.github.io



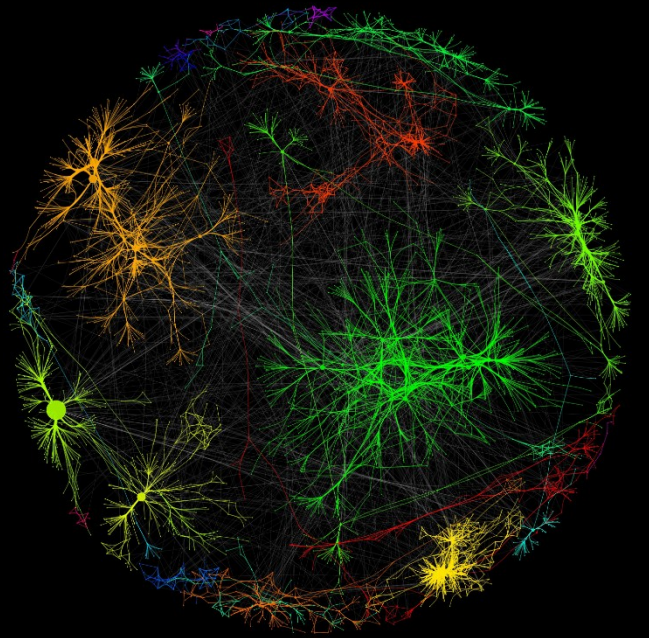
- Windows/Linux/OS X
- Can handle 50K nodes & 1000K edges
- Interactive
  - Filter
  - Dynamic layout
  - Clustering/hierarchies

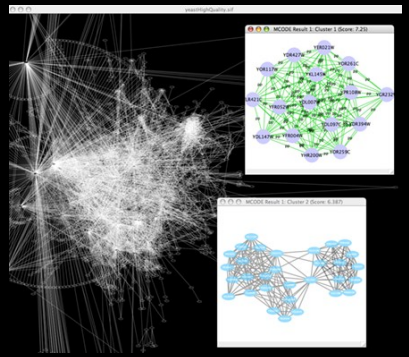


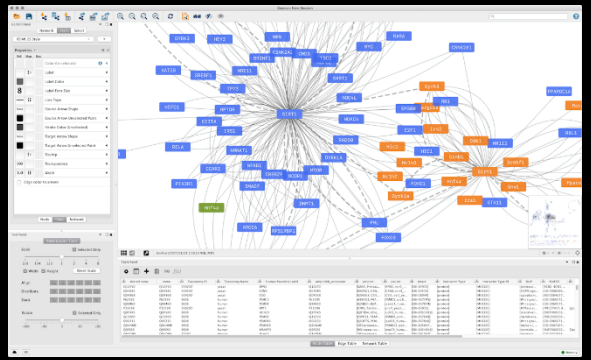
Graphs/Networks: Cytoscape

https://cytoscape.org/

- Windows/Linux/Mac OS/Unix
- Designed for biological research & molecular networks
- Many apps (aka plugins)
- Javascript library - cytoscape.js
- Export: PDF, PS, SVG, PNG, JPG



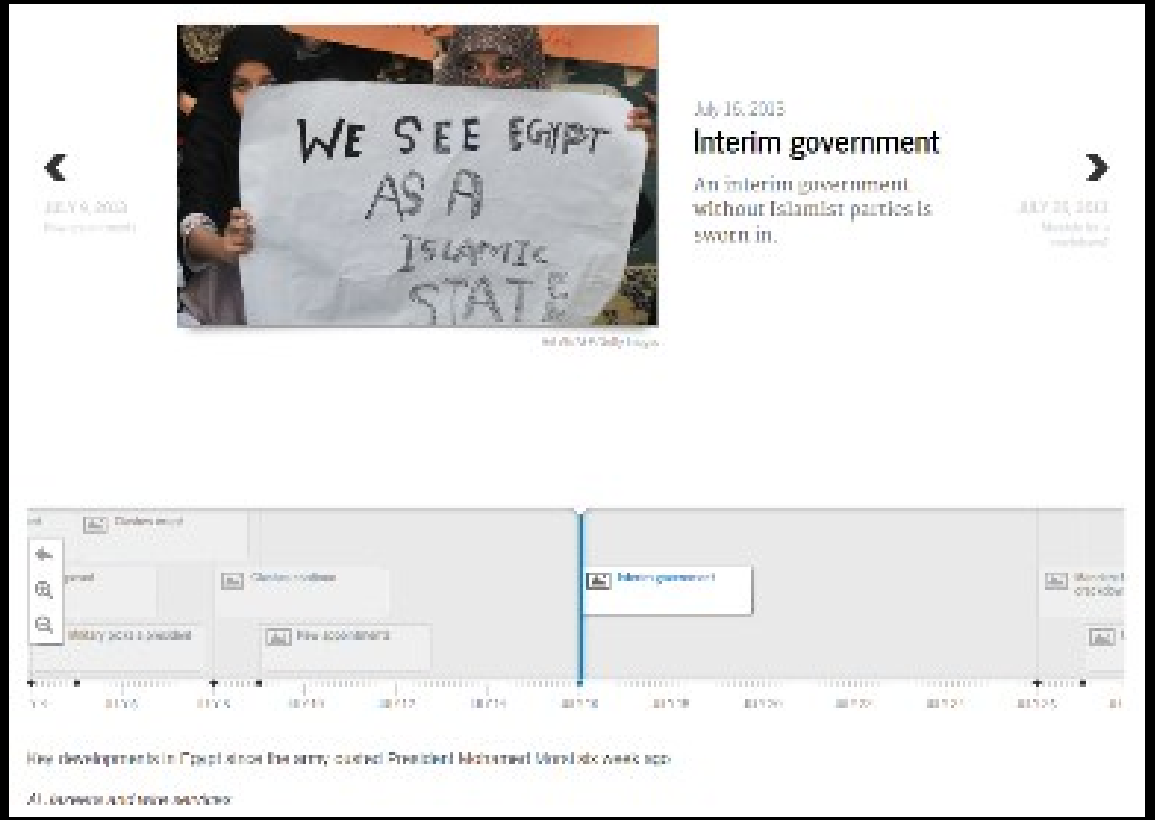




Time: Timeline JS

http://timeline.knightlab.com

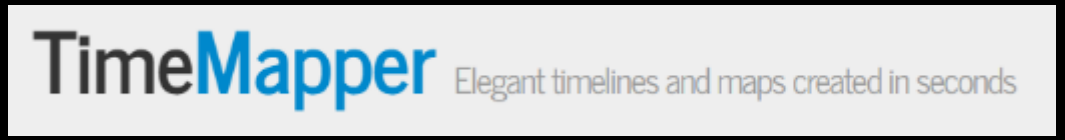
- Interactive, web timelines
- Link in URLs and web resources
- Built upon Google spreadsheet
- Produces embeddable iframe widget

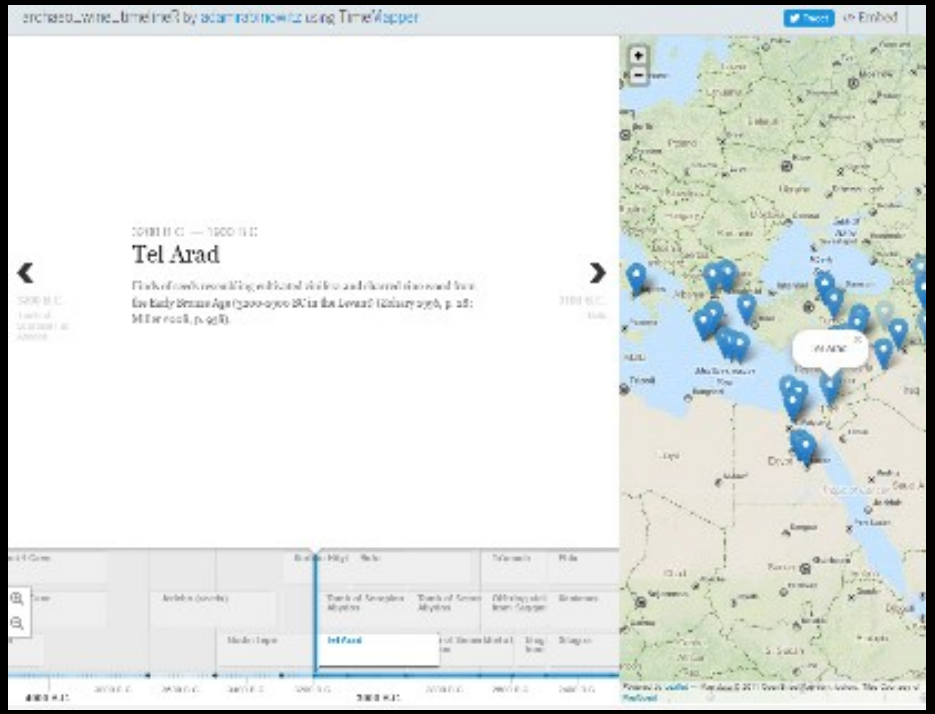


Time: TimeMapper

http://timemapper.okfnlabs.org

- Interactive, web timelines
- Built on Google spreadsheet data
- Adds a map with identified locations
- Creates web page that can be embedded into other sites





Colour

Adobe Color

- https://color.adobe.com/
- Pick great colour palettes
- Given this colour pick complementary colours

ColorBrewer

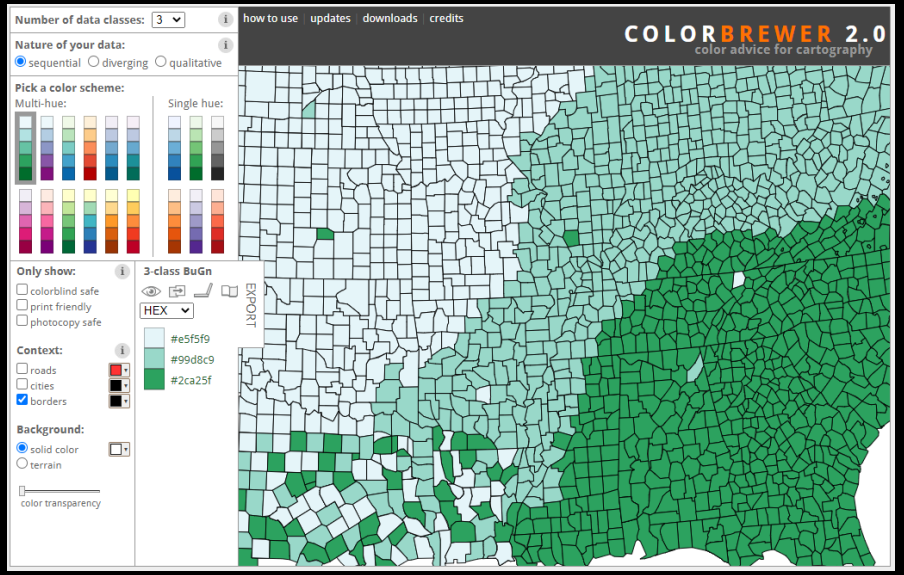
- http://colorbrewer2.org/
- Help in choosing colours for maps
- Colourblind, printing, etc.

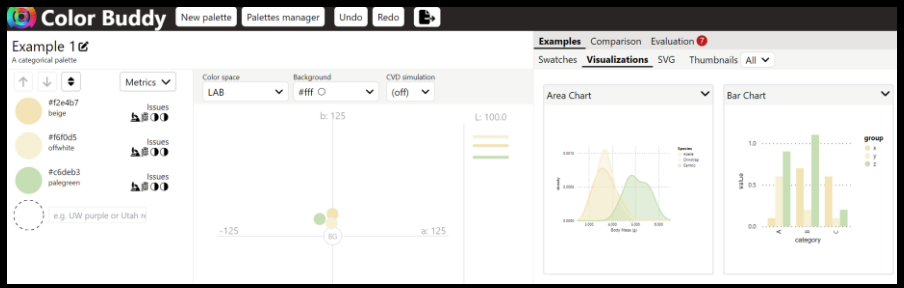
Color Buddy

- https://color-buddy.netlify.app/
- More options, shows several different types of visualization

Magic Color Picker

- https://text2color.com/picker.html
- AI that changes text to RGB colour codes



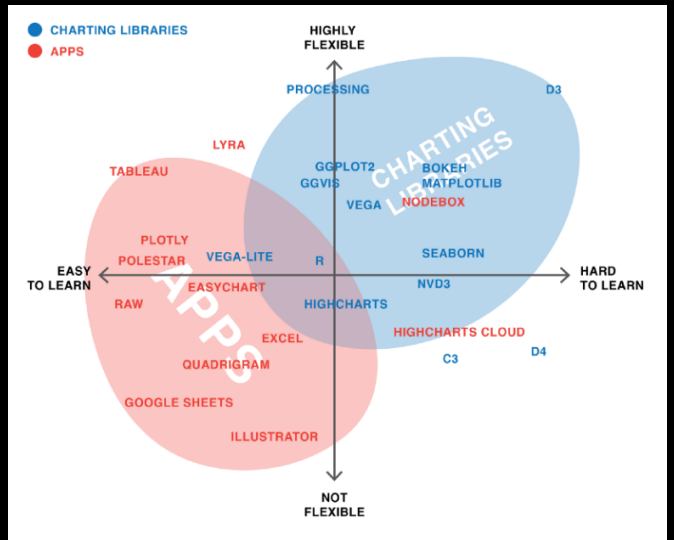


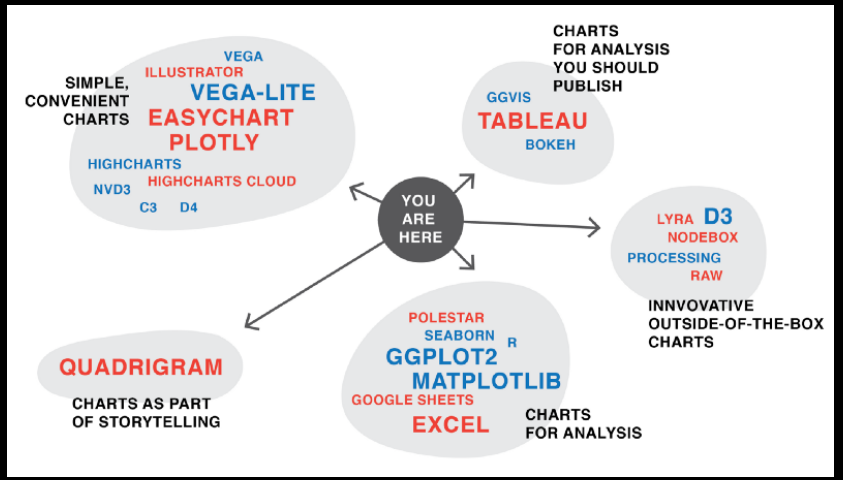
Resources – Data Vis Tools

Visualizing Data - https://www.visualisingdata.com/resources/

120+ visualization tools

Article on picking a vis tool –
https://source.opennews.org/articles/what-i-learned-recreating-one-chart-using-24-tools/





Visualization Resources: Books

Semiology of Graphics by Jacques Bertin

- *The book on visual variables*

Visual Display of Quantitative Information, Beautiful Evidence, Visual Explanations, *or* ***Envisioning Information*** by Edward Tufte

- *Beautiful examples of historic visualizations*

Visual Thinking for Design by Colin Ware

- Ties perception theory and design processes to visualization practices.

Beautiful Visualization by Steele & Iliinsky

- Combines techniques from artists, designers, scientists, and others.

Visual Analysis & Design by Tamara Munzner

- Good text book by long-time practitioner & prof

Visualization Resources: Websites

New York Times https://www.nytimes.com/interactive/2024/12/20/us/2024-year-in-graphics.html

- Dedicated team producing exceptional work.

Gapminder https://www.gapminder.org/

- Hans Rosling's stat software & data.

Visualizing Data https://www.visualisingdata.com/blog/

- Quarterly best visualizations posts, the little of visualization design

Flowing Data https://flowingdata.com/

- Daily posts showing various visualizations created by the site author and elsewhere.

Questions?

John Brosz, PhD
Data and Visualization Curator
JDLBrosz@UCalgary.ca

Slides: http://brosz.ca/slides/