

大数据分析实践III

Visualization Design

Qiong Zeng (曾琼)
qiong.zn@sdu.edu.cn

Research is *creative* and systematic work undertaken to increase the stock of *knowledge*.

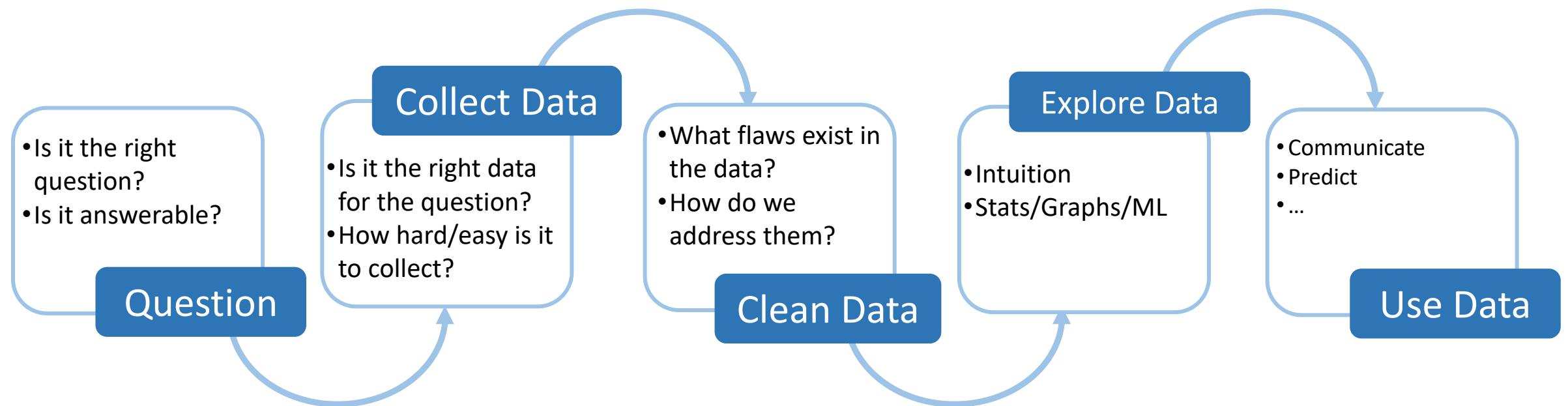




Looking carefully at your data is important:

- to identify mistakes in collection/processing
- to find violations of statistical assumptions
- to observe patterns in the data
- to make hypothesis.

Feeding unvisualized data to a machine learning algorithm is asking for trouble.



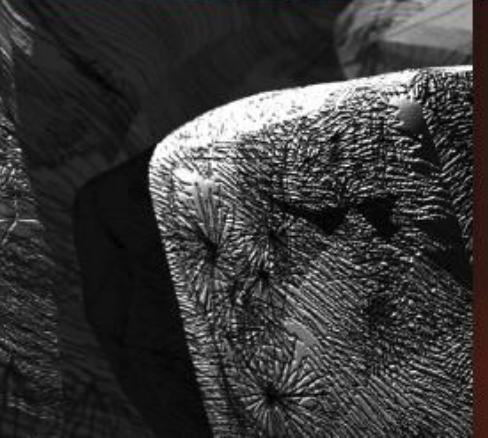
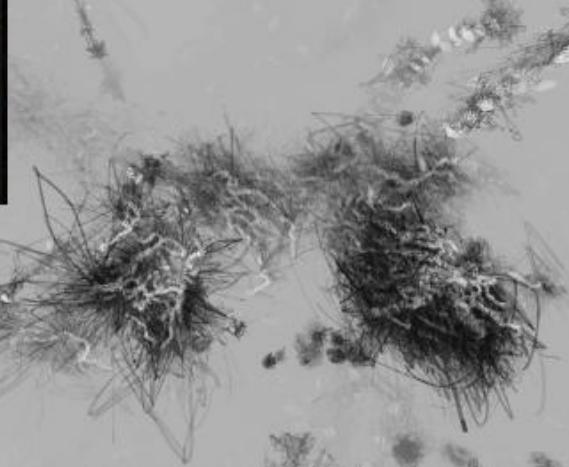
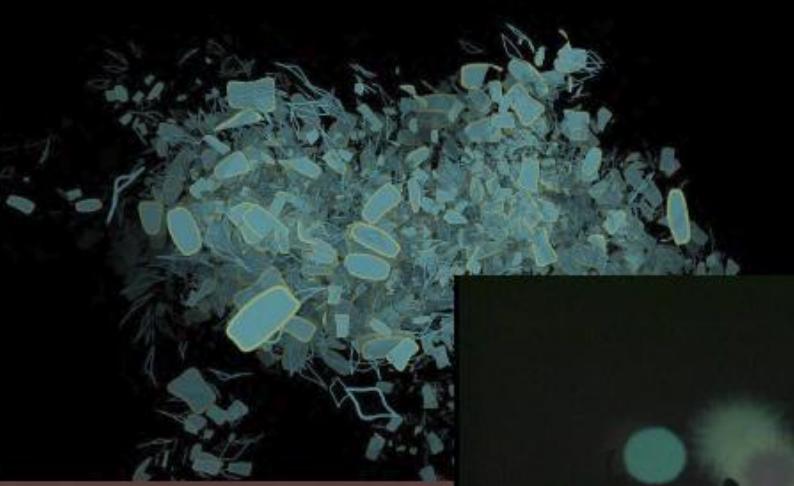


学习目标



**掌握可视化设计的基本流程
并且学会应用**

了解评价可视化好坏的基本方式

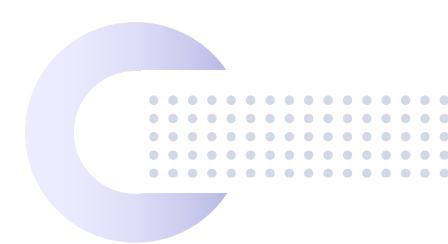


4



On co... to Rebecca Xu

Everything can be
visualized!



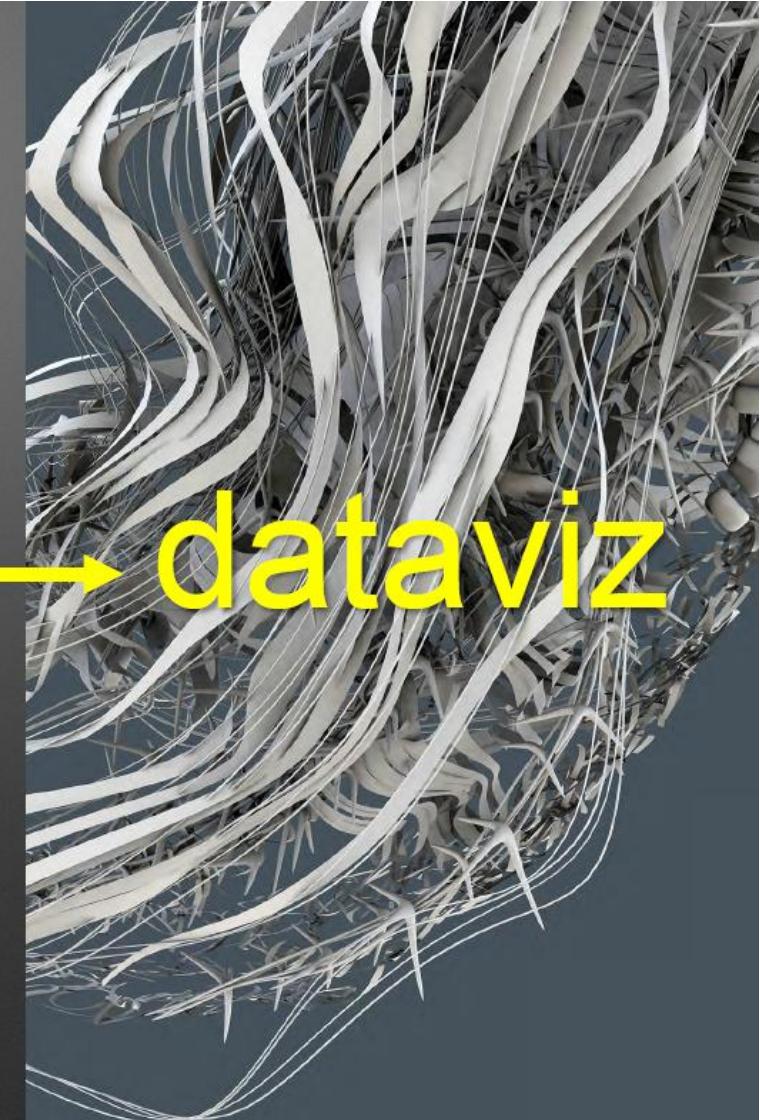
- Books
- Food
- Drinks
- Sounds
- Laughter
- Transportation
- Indecision
- Eavesdropping
- Our
- Thank-yous
- Complaints
- Compliments
- Mirrors
- Emotions
- Workspaces
- Our past
- Doors
- Hours of sleep
- To-do Lists
- Citites
- Apologies
- New things
- Envy
- Desires
- Things we buy
- Love



lifornia

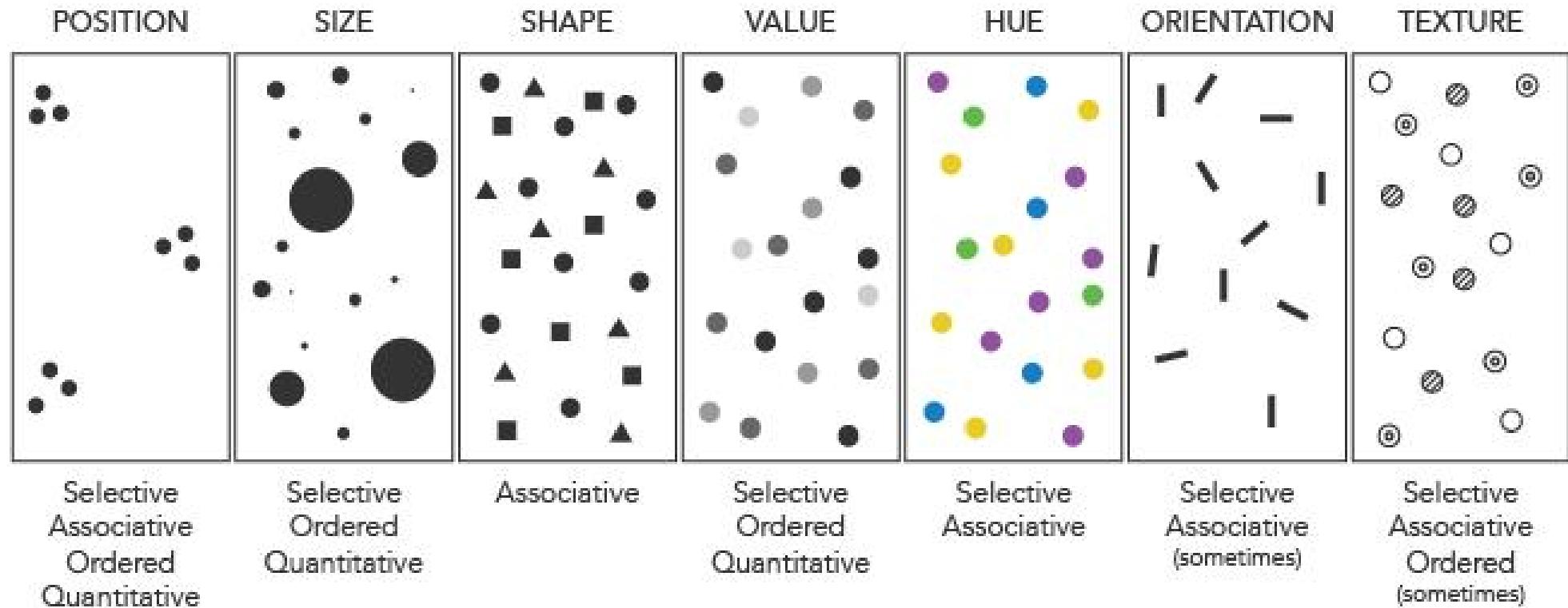
Per Capita Amount (dollars)					
2010	2009	2008	2007	2006	2005
\$9,031.23	\$9,360.24	\$8,199.00	\$7,255.35	\$7,075.64	\$6,742.76
\$321.70	\$271.08	\$230.69	\$206.26	\$213.92	\$194.82
\$22.49	\$13.50	\$8.79	\$10.19	\$7.87	\$6.63
\$446.26	\$512.89	\$435.21	\$473.12	\$473.69	\$430.70
\$291.39	\$321.57	\$232.64	\$210.45	\$202.42	\$192.87
\$495.22	\$513.55	\$448.82	\$516.67	\$456.05	\$471.15
\$275.42	\$257.16	\$216.37	\$121.24	\$155.98	\$162.43
\$301.38	\$394.06	\$215.50	\$200.36	\$212.16	\$180.65
\$123.98	\$128.43	\$93.29	\$80.06	\$77.63	\$74.03
\$2,895.77	\$2,926.15	\$2,883.69	\$2,191.55	\$2,229.75	\$2,069.10
\$80.19	\$92.53	\$85.64	\$71.15	\$68.45	\$64.36
\$165.51	\$194.29	\$133.16	\$135.14	\$123.14	\$129.54
\$52.98	\$43.59	\$37.11	\$9.09	\$31.89	\$29.18
\$46.05	\$5.75	\$3.46	\$9.71	\$15.03	\$45.32
\$338.93	\$40.89	\$2.48	\$1.10	\$48.03	\$154.35
\$213.92	\$203.71	\$231.65	\$250.56	\$217.23	\$203.32
\$4.45	\$14.36	\$3.01	\$2.98	\$2.84	\$2.43
\$194.81	\$249.56	\$200.60	\$192.41	\$172.11	\$135.35
\$211.59	\$174.04	\$164.59	\$145.98	\$148.75	\$142.97
\$175.52	\$291.70	\$181.34	\$170.20	\$145.98	\$144.01
\$0.05	\$0.10	\$0.05	\$0.04	\$0.06	\$0.05
\$16.62	\$28.41	\$11.53	\$12.40	\$12.19	\$12.48
\$5.33	\$4.41	\$3.08	\$3.09	\$3.20	\$3.10
\$38.30	\$30.15	\$23.48	\$20.28	\$23.10	\$24.62
\$106.94	\$113.18	\$111.85	\$91.79	\$60.28	\$98.92
\$26.00	\$30.39	\$22.13	\$22.26	\$20.79	\$21.80
\$208.92	\$204.46	\$196.38	\$193.72	\$185.25	\$173.66
\$1.76	\$5.11	\$1.24	\$1.17	\$1.71	\$1.19
\$1,936.99	\$1,849.88	\$1,724.46	\$1,680.56	\$1,614.39	\$1,547.36
\$0.31	\$0.32	\$0.32	\$0.32	\$0.33	\$0.32
\$32.43	\$27.02	\$21.47	\$23.49	\$21.45	\$26.08

data → rules → dataviz



On courtesy to Rebecca Xu

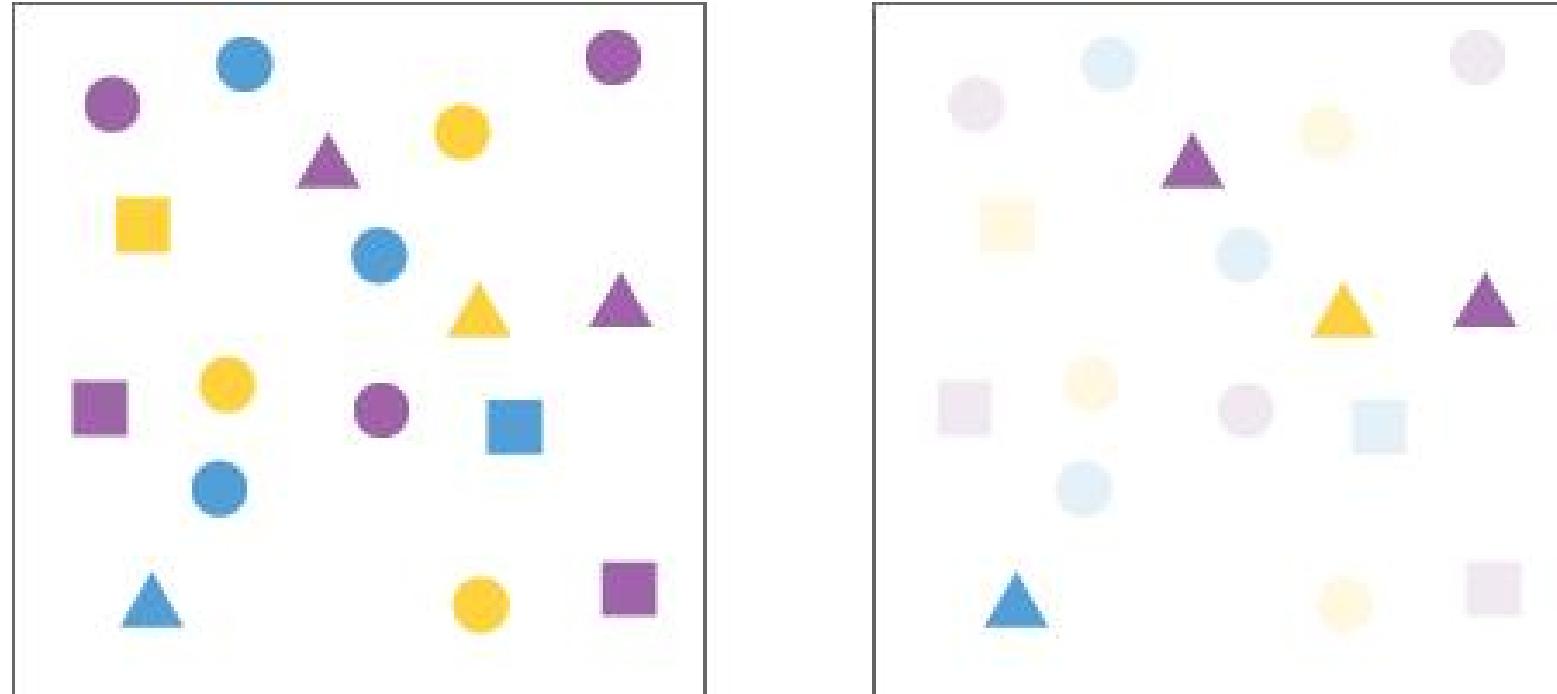
Bertin' s Visual Variables



A **selective** variable allows us to isolate a group of signs based on a change in the variable. A **associative variable** allows grouping across changes in the variable.



Bertin' s Visual Variables



It is not easy to isolate symbols based on shape. Even with all the triangles pulled out at right, they still don't look like a group. Shape is not selective.

A **selective** variable allows us to isolate a group of signs based on a **change**. A **associative variable** allows grouping across changes in the variable



Color Scales

Perceived as Ordered

Brightness



Saturation



Hue: not as much

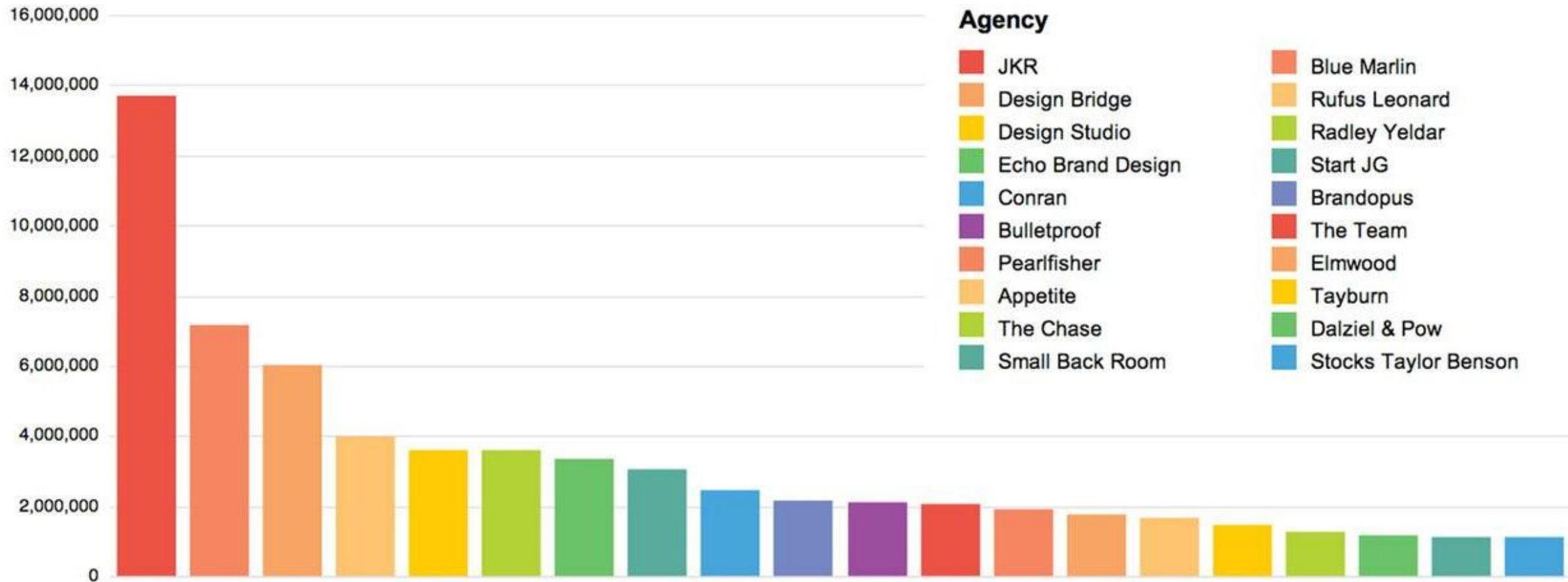




Color Scales



Dramatic Misuse of Color

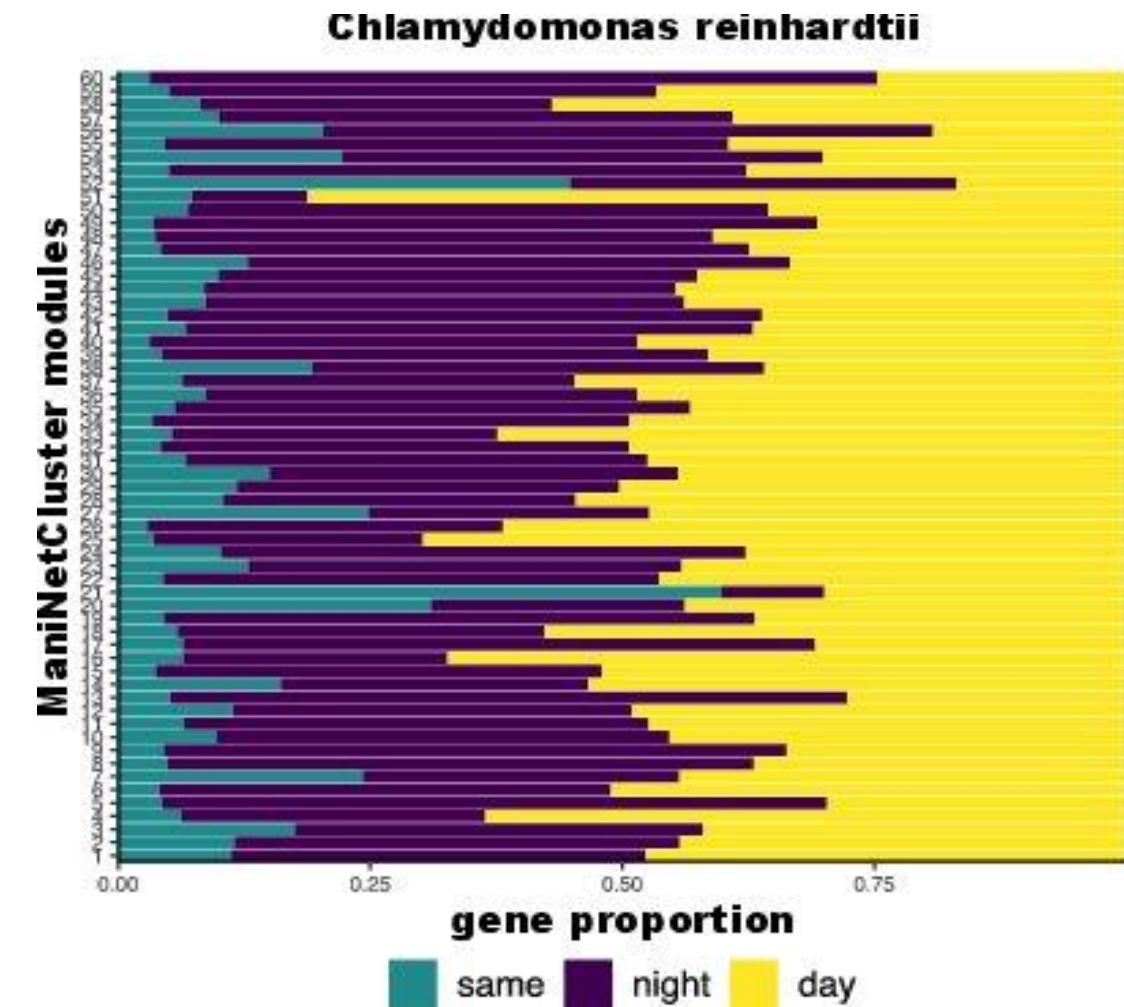
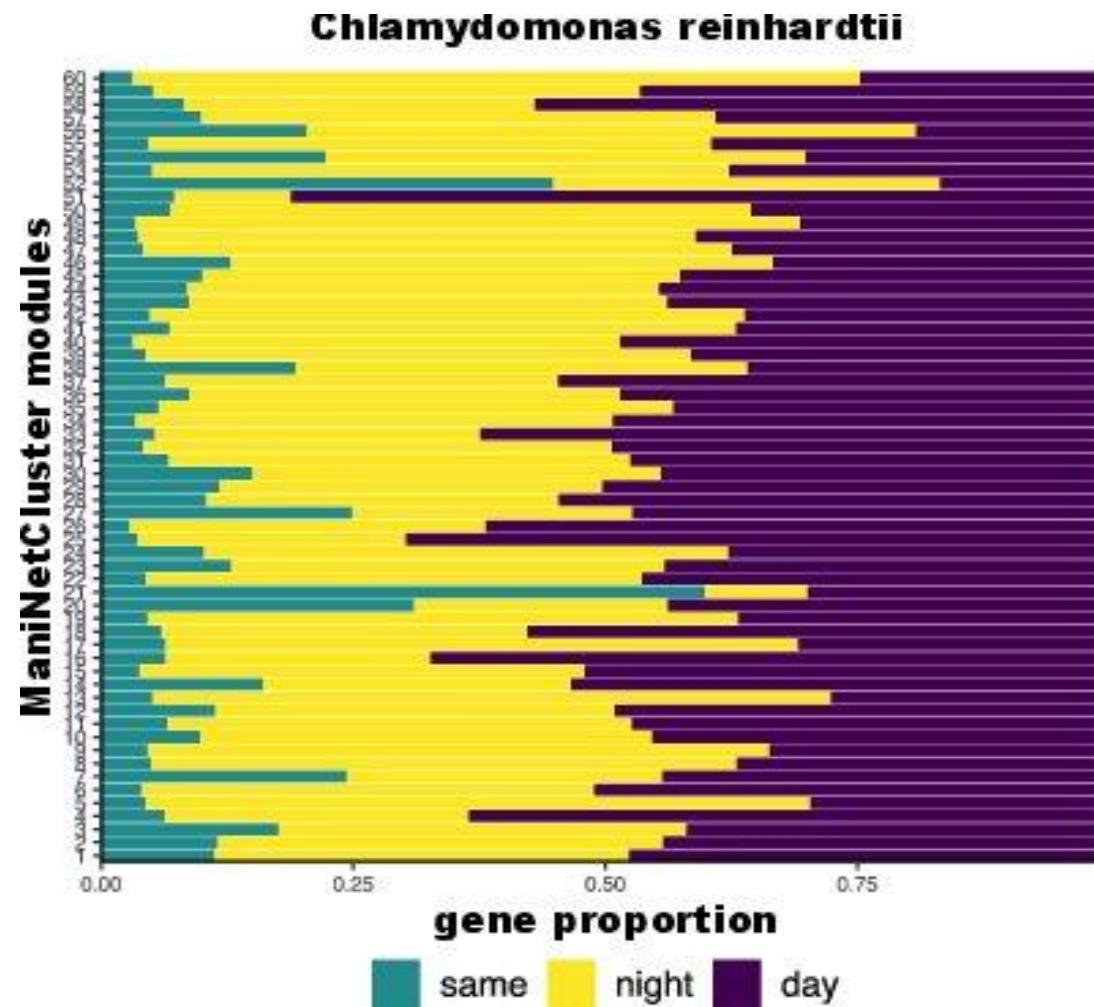




Color Scales



Which coloring is better?



Setting up Rules for visual variables



The higher the amount, the larger the circle.

'The lower the value, the darker the color.

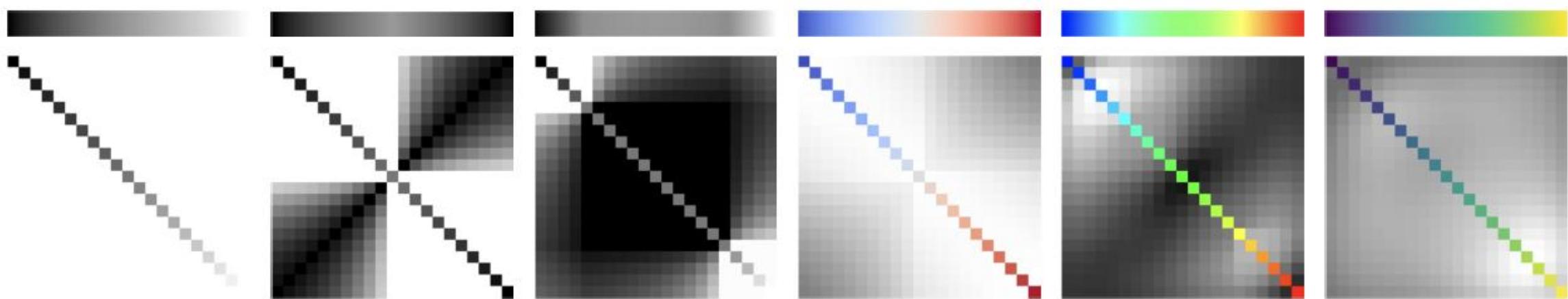
'If it is _____ type of data, I will color it green.

'If it is _____ type of data, I will use a _____ shape'

Color Scales



Rainbows are perceptually non-linear
Distinct positive/negative colors reflected about a center make good scales



(a) Greyscale. Constant speed between each pair of points. (b) Half greyscale. Constant speed in each half, drops to zero across the halves. (c) Flat greyscale. Constant grey in $[0.2, 0.8]$ results in a speed of zero. (d) Cool/warm divergent. Constant speed in each half, but not across halves. (e) Rainbow. High speed around blue, low around green. (f) Viridis. Low variations in speed result in its uniform appearance.

Fig. 1: Our six showcase colormaps and their corresponding matrices of the global speed. The luminance of each entry (i, k) represents how high the speed $V_{i,k}^{\Delta E^{76}} = \Delta E^{76}(x_i, x_k)/(t_k - t_i)$ is from the color $x(t_i)$ in the same row i to the color $x(t_k)$ in the same column k on the diagonal using the ΔE^{76} metric.

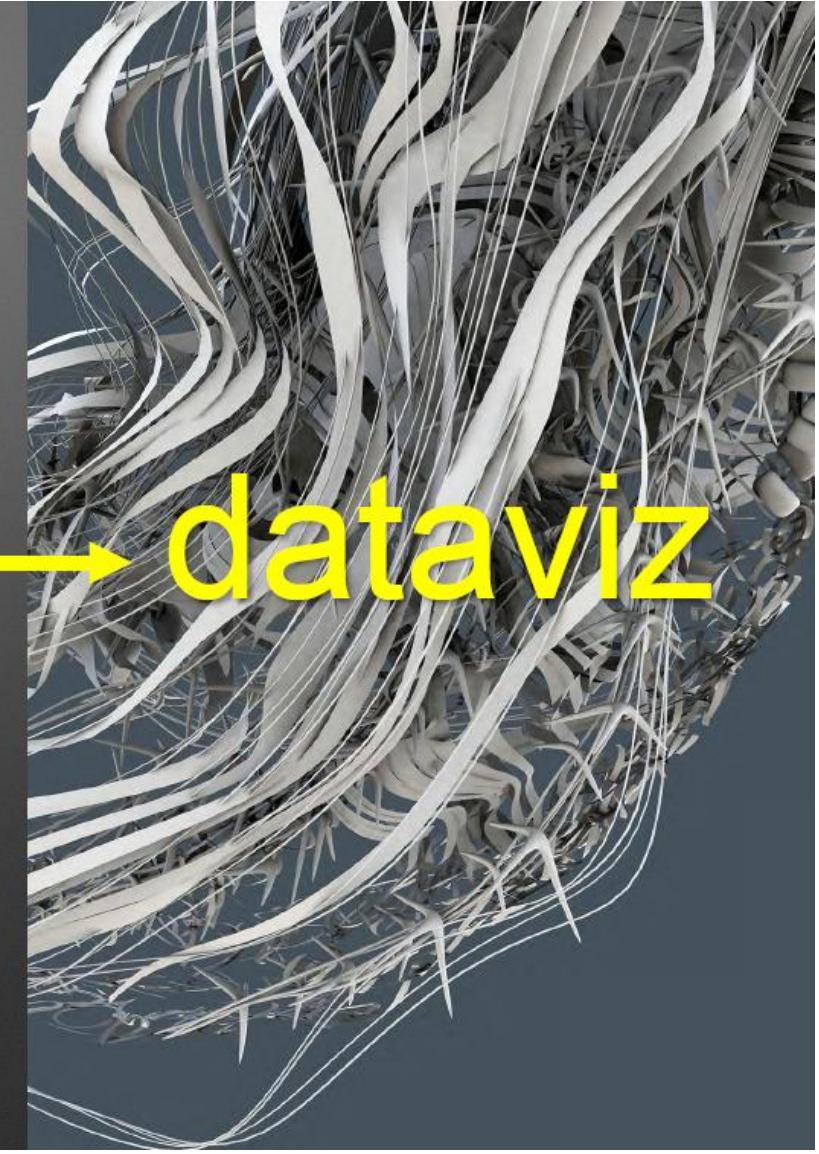
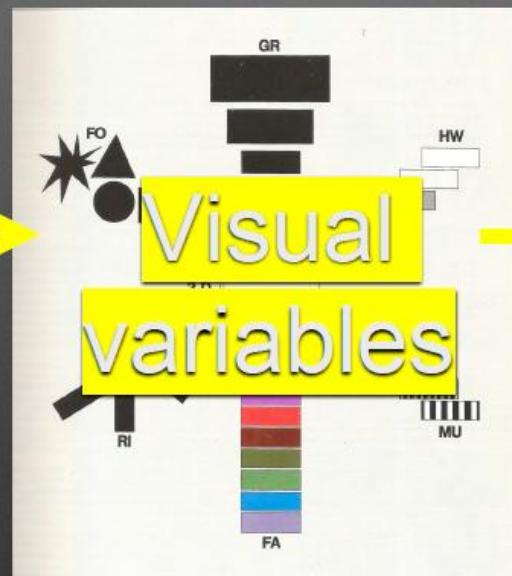


Per Capita Amount (dollars)					
2010	2009	2008	2007	2006	2005
\$9,031.23	\$9,360.24	\$8,199.00	\$7,235.35	\$7,075.64	\$6,742.76
\$321.70	\$271.08	\$230.69	\$206.26	\$213.92	\$194.82
\$22.49	\$13.50	\$8.79	\$10.19	\$7.87	\$6.63
\$446.26	\$512.89	\$435.21	\$473.12	\$473.69	\$430.70
\$291.39	\$321.57	\$232.64	\$210.45	\$202.42	\$192.87
\$495.22	\$513.55	\$448.82	\$516.67	\$456.05	\$471.15
\$275.42	\$257.16	\$216.37	\$121.24	\$155.98	\$162.43
\$301.38	\$394.06	\$215.50	\$200.36	\$212.16	\$180.65
\$123.98	\$128.43	\$93.29	\$80.06	\$77.63	\$74.03
\$2,895.77	\$2,926.15	\$2,883.69	\$2,191.55	\$2,229.75	\$2,069.10
\$80.19	\$92.53	\$85.64	\$71.15	\$68.45	\$64.36
\$165.51	\$194.29	\$133.16	\$135.14	\$123.14	\$129.54
\$52.98	\$43.59	\$37.11	\$19.09	\$31.89	\$29.18
\$46.05	\$57.75	\$24.46	\$19.72	\$15.03	\$45.32
\$338.93	\$40.89	\$22.48	\$11.00	\$8.03	\$154.35
\$213.92	\$203.71	\$231.63	\$258.56	\$217.23	\$203.32
\$4.45	\$14.36	\$3.01	\$2.98	\$2.84	\$2.43
\$194.81	\$249.56	\$200.60	\$192.41	\$172.11	\$135.35
\$211.59	\$174.04	\$164.59	\$145.98	\$148.75	\$142.97
\$175.52	\$291.70	\$181.34	\$170.20	\$145.98	\$144.01
\$0.05	\$0.10	\$0.05	\$0.04	\$0.06	\$0.05
\$16.62	\$28.41	\$11.53	\$12.40	\$12.19	\$12.48
\$5.33	\$4.41	\$3.08	\$3.09	\$3.20	\$3.10
\$38.30	\$30.15	\$23.48	\$20.28	\$23.10	\$24.62
\$106.94	\$113.18	\$111.85	\$91.79	\$60.28	\$98.92
\$26.00	\$30.39	\$22.13	\$22.26	\$20.79	\$21.80
\$208.92	\$204.46	\$196.38	\$193.72	\$185.25	\$173.66
\$1.76	\$5.11	\$1.24	\$1.17	\$1.71	\$1.19
\$1,936.99	\$1,849.88	\$1,724.46	\$1,680.56	\$1,614.39	\$1,547.36
\$0.31	\$0.32	\$0.32	\$0.32	\$0.33	\$0.32
\$32.43	\$27.02	\$21.47	\$23.49	\$21.45	\$26.08

data

Rules

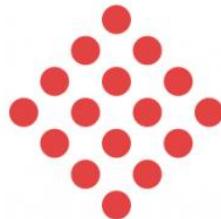
+



dataviz

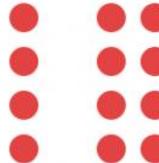


Gestalt Principles



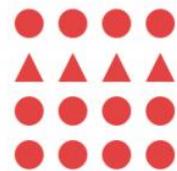
Good Figure

Objects grouped together tend to be perceived as a single figure. Tendency to simplify.



Proximity

Objects tend to be grouped together if they are close to each other.



Similarity

Objects tend to be grouped together if they are similar.



Continuation

When there is an intersection between two or more objects, people tend to perceive each object as a single uninterrupted object.



Closure

Visual connection or continuity between sets of elements which do not actually touch each other in a composition.

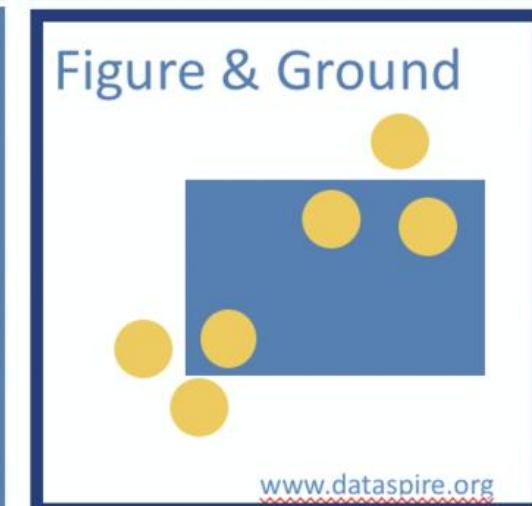
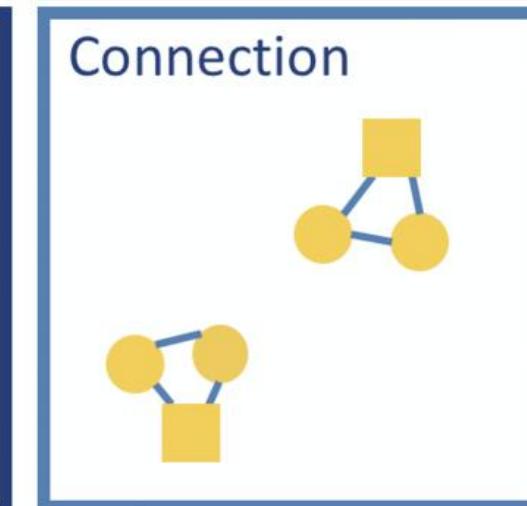
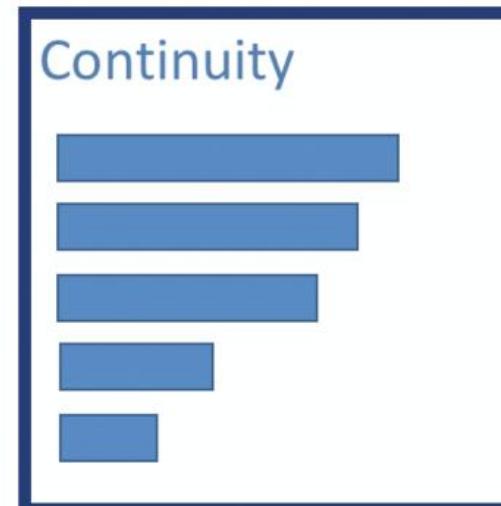
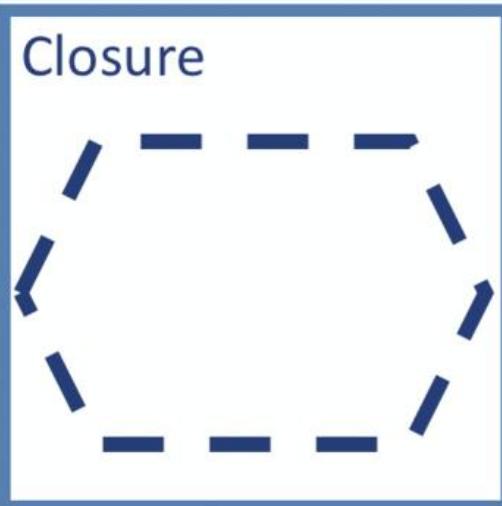
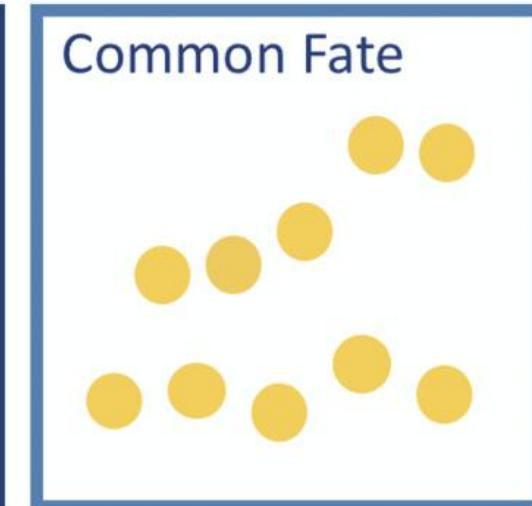
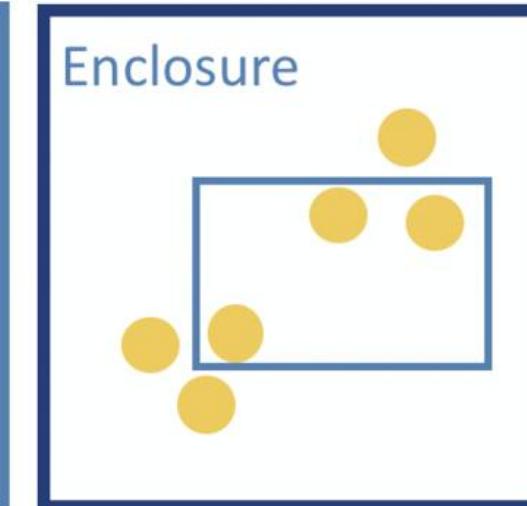
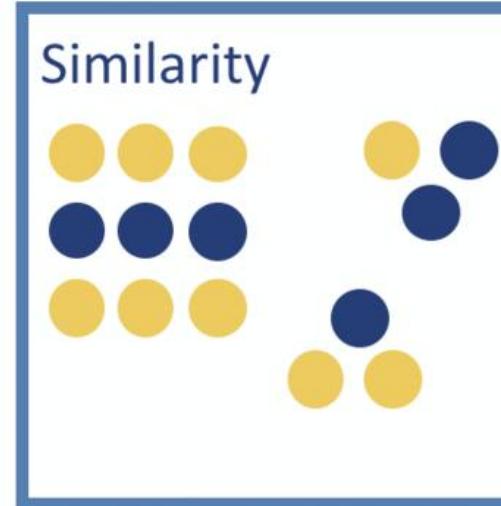
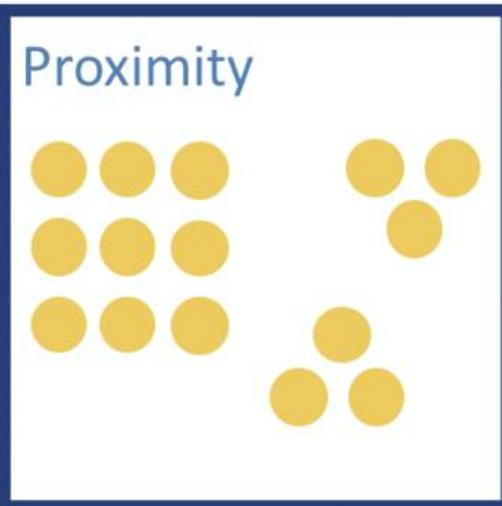


Symmetry

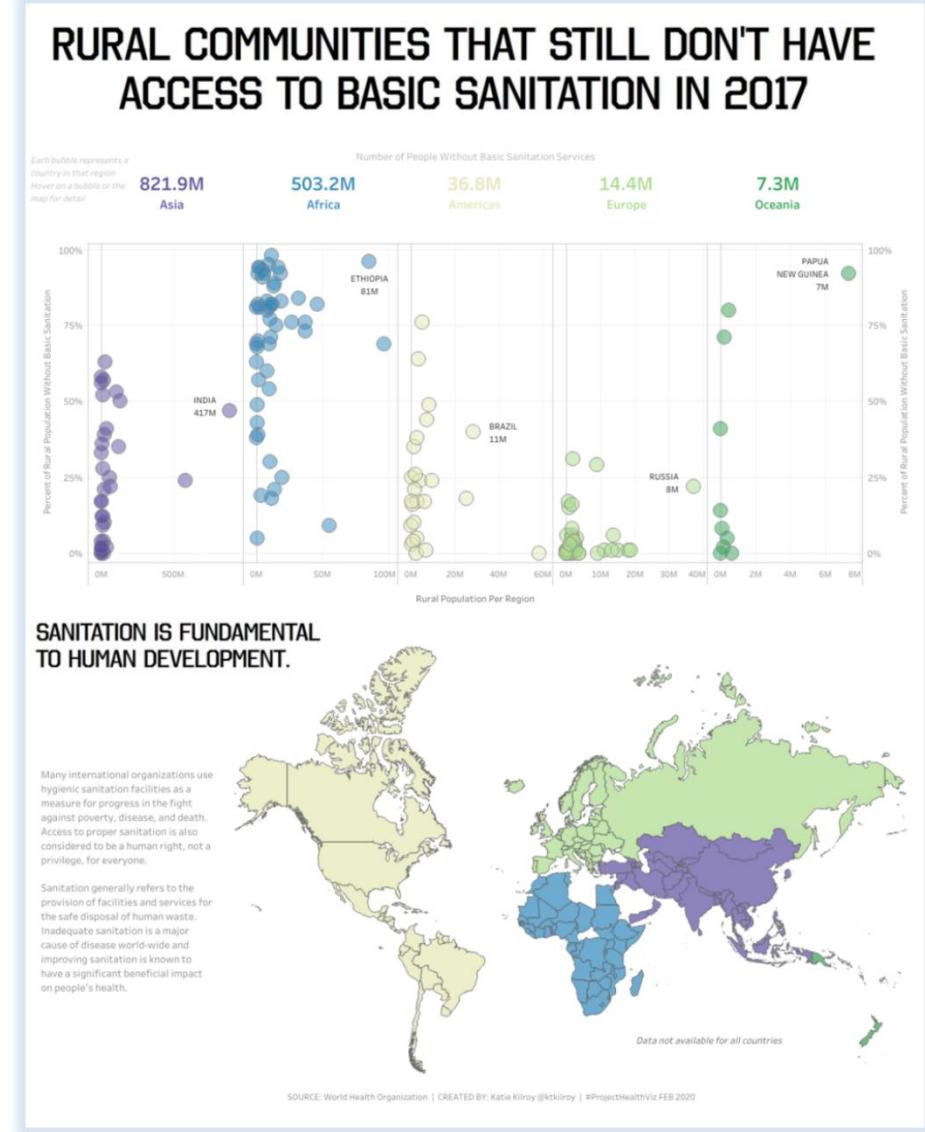
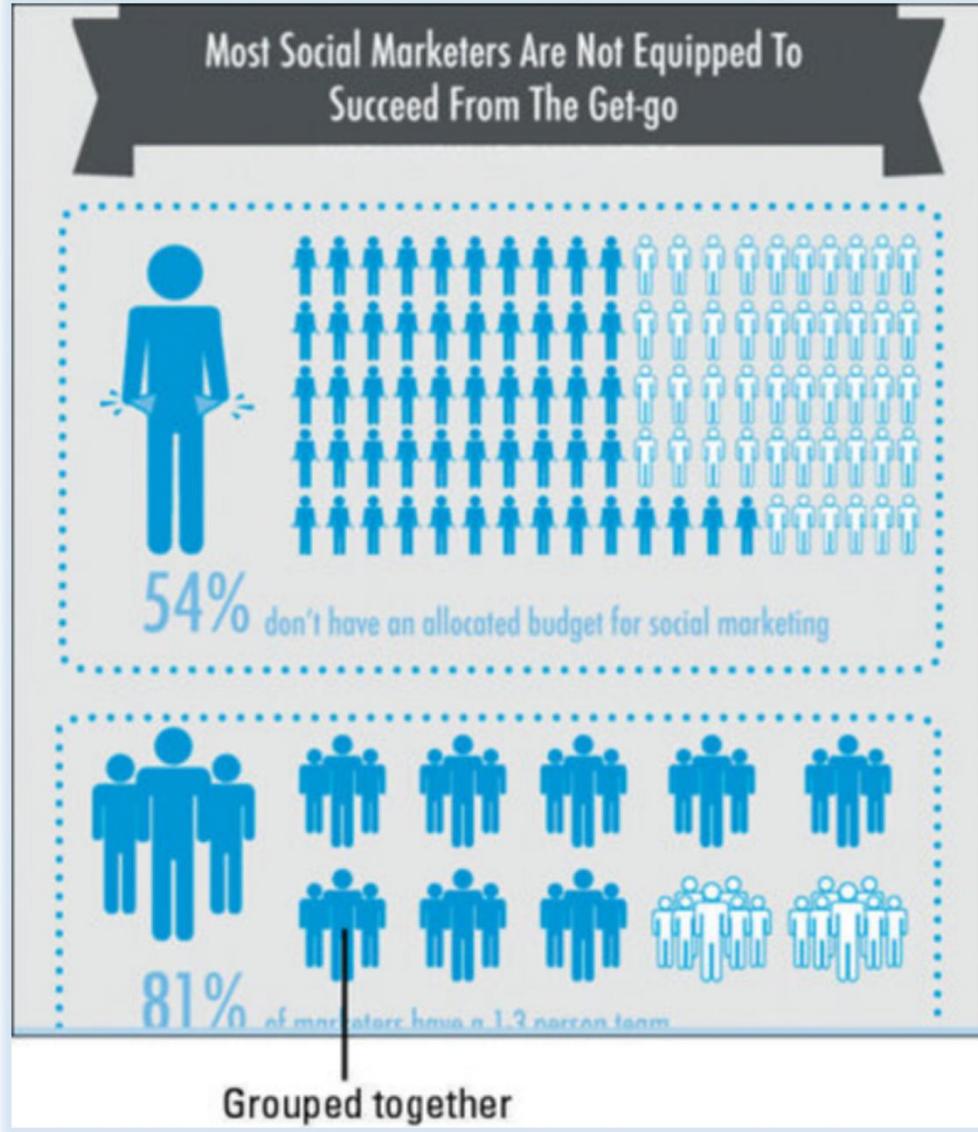
The objects tend to be perceived as symmetrical shapes that form around their center.



Gestalt Principles



Gestalt Principles



Gestalt Principles



Country	Area	Density	Birthrate	Population	Mortality	GDP
Russia	17075200	8.37	99.6	142893540	15.39	8900.0
Mexico	1972550	54.47	92.2	107449525	20.91	9000.0
Japan	377835	337.35	99.0	127463611	3.26	28200.0
United Kingdom	244820	247.57	99.0	60609153	5.16	27700.0
New Zealand	268680	15.17	99.0	4076140	5.85	21600.0
Afghanistan	647500	47.96	36.0	31056997	163.07	700.0
Israel	20770	305.83	95.4	6352117	7.03	19800.0
United States	9631420	30.99	97.0	298444215	6.5	37800.0
China	9596960	136.92	90.9	1313973713	24.18	5000.0
Tajikistan	143100	51.16	99.4	7320815	110.76	1000.0
Burma	678500	69.83	85.3	47382633	67.24	1800.0
Tanzania	945087	39.62	78.2	37445392	98.54	600.0
Tonga	748	153.33	98.5	114689	12.62	2200.0
Germany	357021	230.86	99.0	82422299	4.16	27600.0
Australia	7686850	2.64	100.0	20264082	4.69	29000.0

Gestalt Principles



Country	Population	Area	Density	Mortality	GDP	Birth Rate
Afghanistan	31,056,997	647,500	47.96	163.07	700	36.0
Australia	20,264,082	7,686,850	2.64	4.69	29,000	100.0
Burma	47,382,633	678,500	69.83	67.24	1,800	85.3
China	1,313,973,713	9,596,960	136.92	24.18	5,000	90.9
Germany	82,422,299	357,021	230.86	4.16	27,600	99.0
Israel	6,352,117	20,770	305.83	7.03	19,800	95.4
Japan	127,463,611	377,835	337.35	3.26	28,200	99.0
Mexico	107,449,525	1,972,550	54.47	20.91	9,000	92.2
New Zealand	4,076,140	268,680	15.17	5.85	21,600	99.0
Russia	142,893,540	17,075,200	8.37	15.39	8,900	99.6
Tajikistan	7,320,815	143,100	51.16	110.76	1,000	99.4
Tanzania	37,445,392	945,087	39.62	98.54	600	78.2
Tonga	114,689	748	153.33	12.62	2,200	98.5
United Kingdom	60,609,153	244,820	247.57	5.16	27,700	99.0
United States	298,444,215	9,631,420	30.99	6.50	37,800	97.0

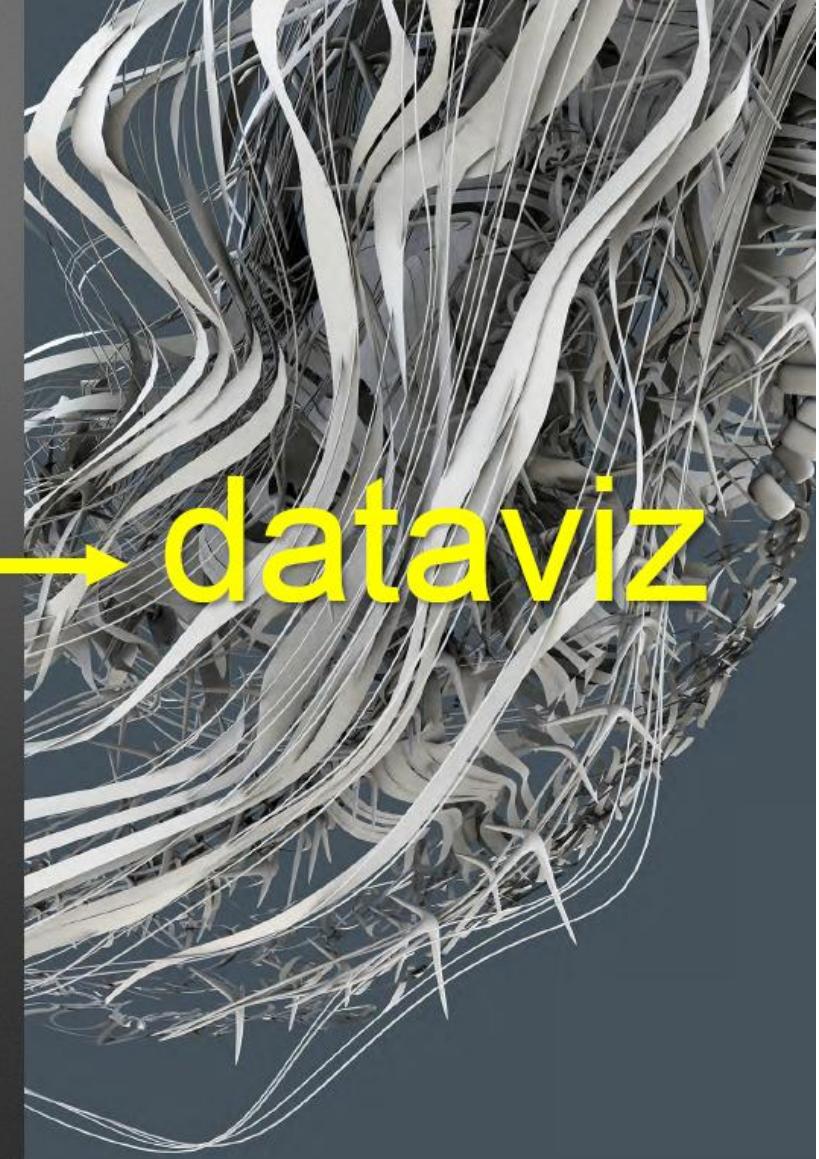


Per Capita Amount (dollars)					
2010	2009	2008	2007	2006	2005
\$9,031.23	\$9,360.24	\$8,199.00	\$7,255.35	\$7,075.64	\$6,742.76
\$321.70	\$271.08	\$230.69	\$206.26	\$213.92	\$194.82
\$22.49	\$13.50	\$8.79	\$10.19	\$7.87	\$6.63
\$446.26	\$512.89	\$435.21	\$473.12	\$473.69	\$430.70
\$291.39	\$321.57	\$232.64	\$210.45	\$202.42	\$192.87
\$495.22	\$513.55	\$448.82	\$516.67	\$456.05	\$471.15
\$275.42	\$257.16	\$216.37	\$121.24	\$155.98	\$162.43
\$301.38	\$394.06	\$215.50	\$200.36	\$212.16	\$180.65
\$123.98	\$128.43	\$93.29	\$80.06	\$77.63	\$74.03
\$2,895.77	\$2,926.15	\$2,883.69	\$2,191.55	\$2,229.75	\$2,069.10
\$80.19	\$92.53	\$85.64	\$71.15	\$68.45	\$64.36
\$165.51	\$194.29	\$133.16	\$135.14	\$123.14	\$129.54
\$52.98	\$43.59	\$37.11	\$9.09	\$31.89	\$29.18
\$46.05	\$5.75	\$4.46	\$9.72	\$5.03	\$45.32
\$338.93	\$40.89	\$2,048	\$1,10	\$48.03	\$154.35
\$213.92	\$203.71	\$231.65	\$238.56	\$217.23	\$203.32
\$4.45	\$14.36	\$3.01	\$2.98	\$2.84	\$2.43
\$194.81	\$249.56	\$200.60	\$192.41	\$172.11	\$135.35
\$211.59	\$174.04	\$164.59	\$145.98	\$148.75	\$142.97
\$175.52	\$291.70	\$181.34	\$170.20	\$145.98	\$144.01
\$0.05	\$0.10	\$0.05	\$0.04	\$0.06	\$0.05
\$16.62	\$28.41	\$11.53	\$12.40	\$12.19	\$12.48
\$5.33	\$4.41	\$3.08	\$3.09	\$3.20	\$3.10
\$38.30	\$30.15	\$23.48	\$20.28	\$23.10	\$24.62
\$106.94	\$113.18	\$111.85	\$91.79	\$60.28	\$98.92
\$26.00	\$30.39	\$22.13	\$22.26	\$20.79	\$21.80
\$208.92	\$204.46	\$196.38	\$193.72	\$185.25	\$173.66
\$1.76	\$5.11	\$1.24	\$1.17	\$1.71	\$1.19
\$1,936.99	\$1,849.88	\$1,724.46	\$1,680.56	\$1,614.39	\$1,547.36
\$0.31	\$0.32	\$0.32	\$0.32	\$0.33	\$0.32
\$32.43	\$27.02	\$21.47	\$23.49	\$21.45	\$26.08

data

Rules
+
Visual
variables
+
Visual
Perception

dataviz



GAME: Draw your profile





Draw your profile



Where were you born?

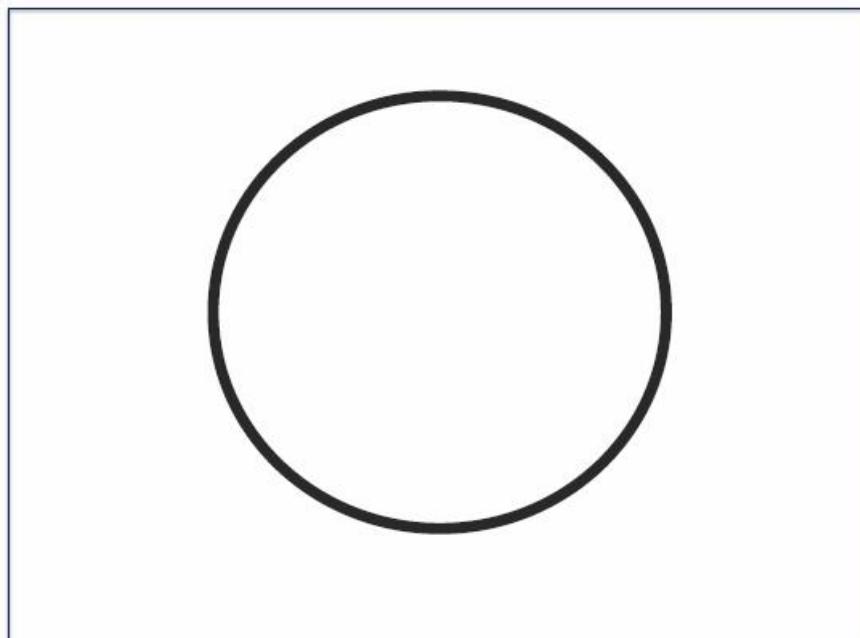




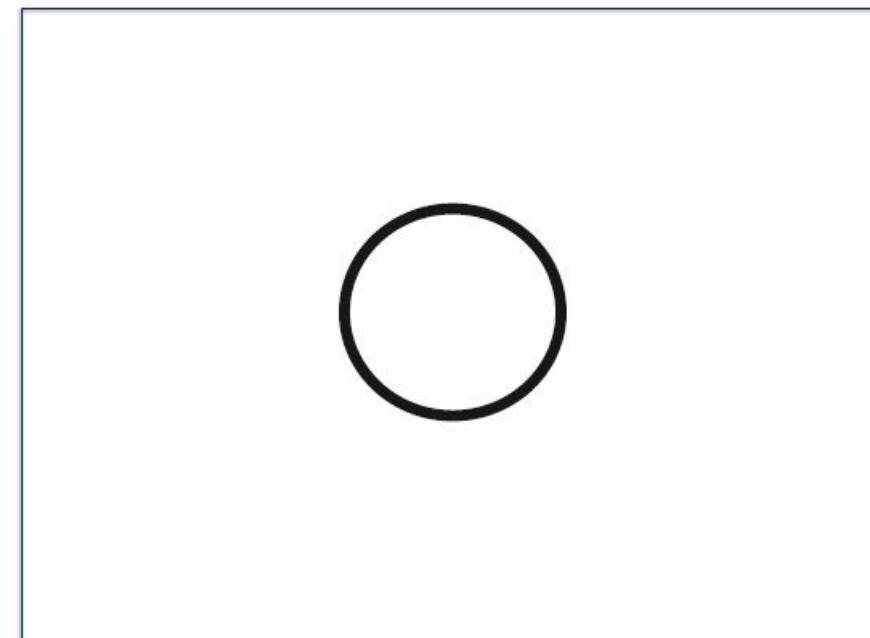
Draw your profile



Which best describes you?



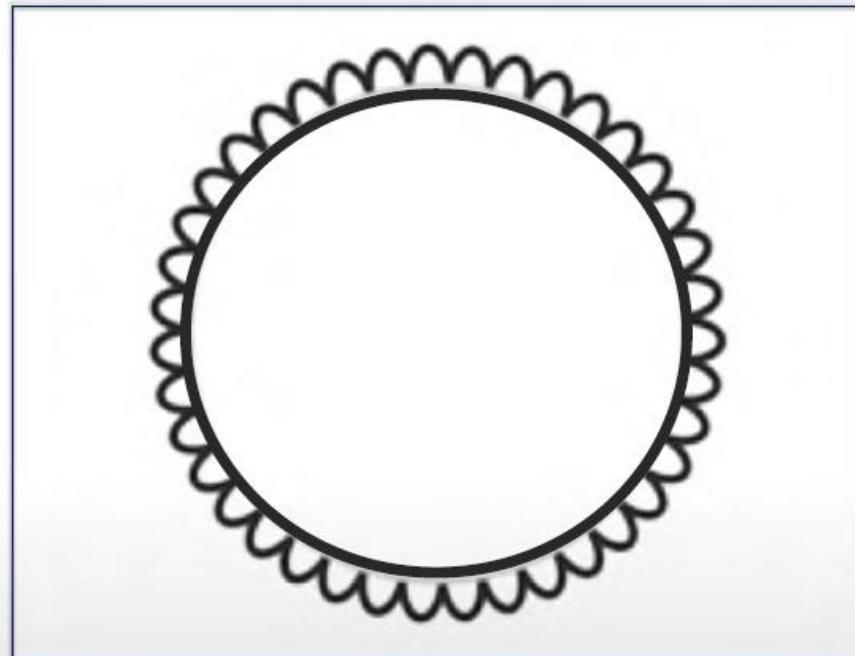
Extrovert



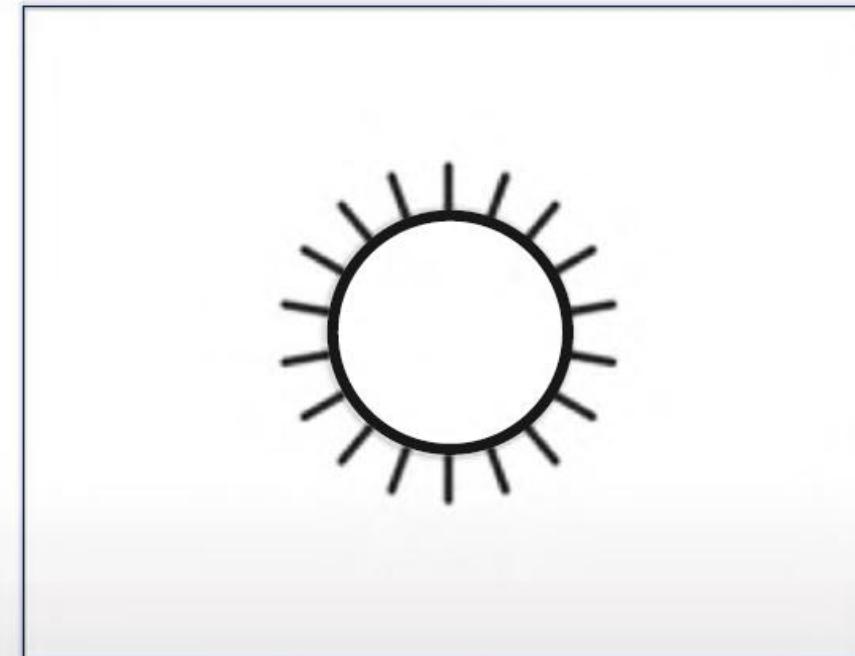
Introvert



Draw your profile



Team Dog



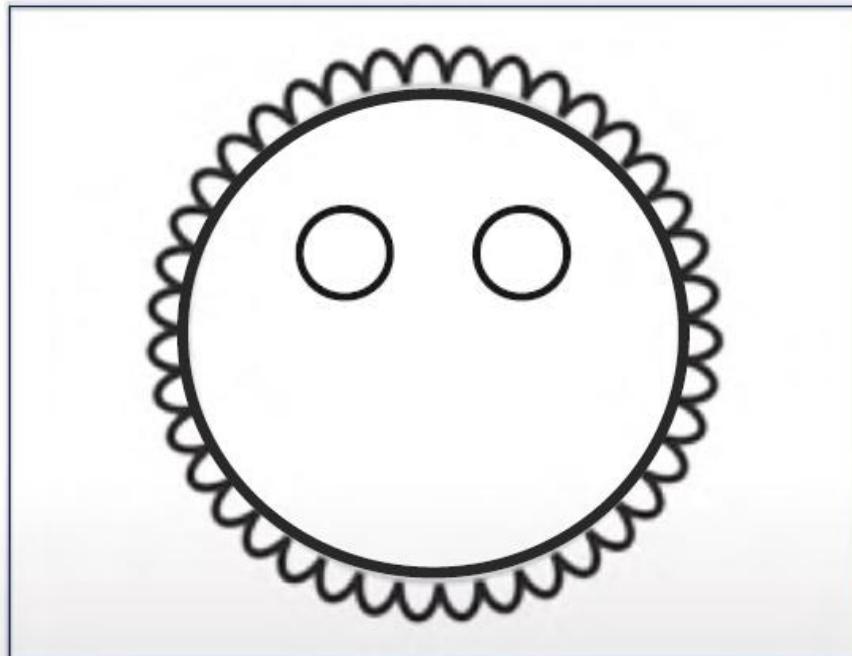
Team Cat



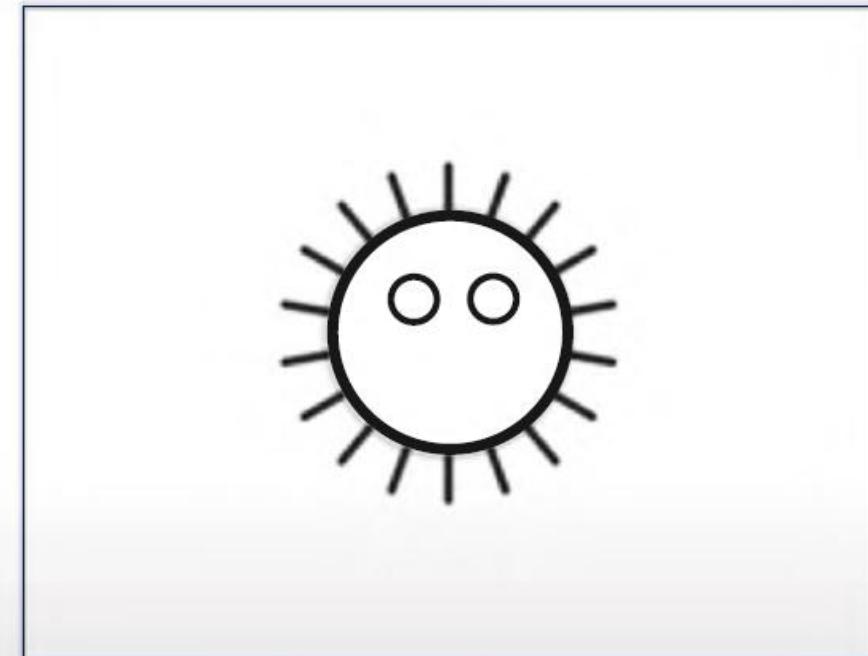
Draw your profile



Add some eyes!



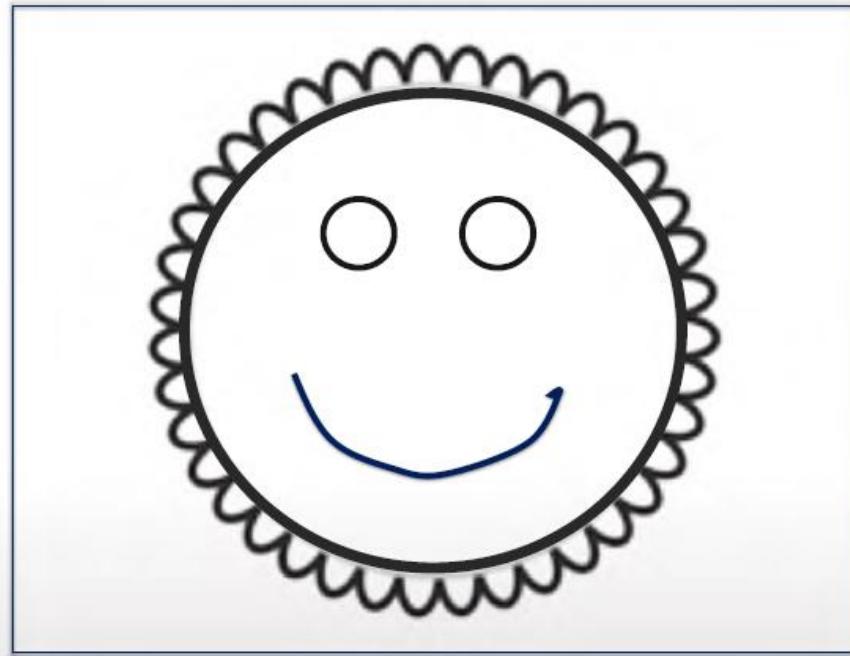
Team Dog



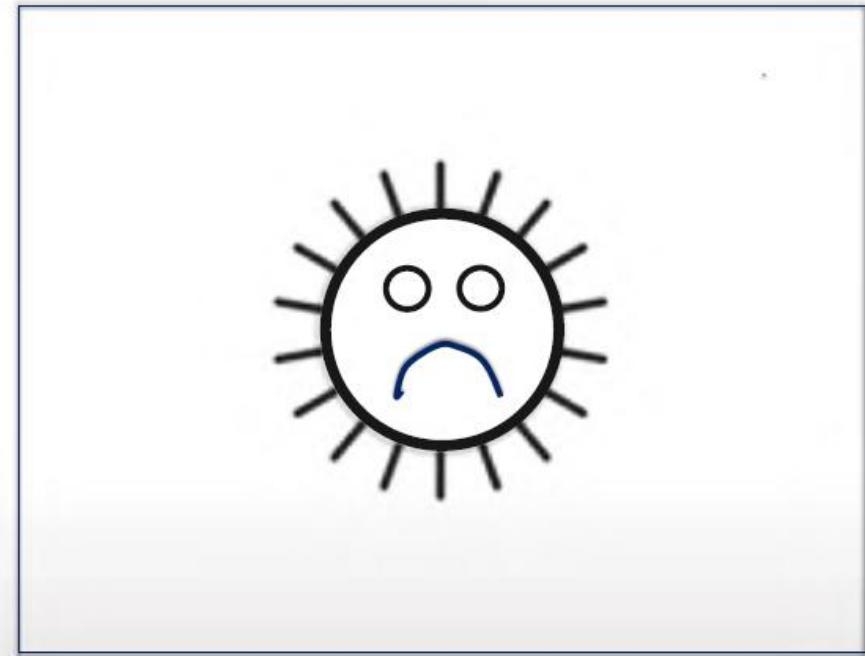
Team Cat



Draw your profile



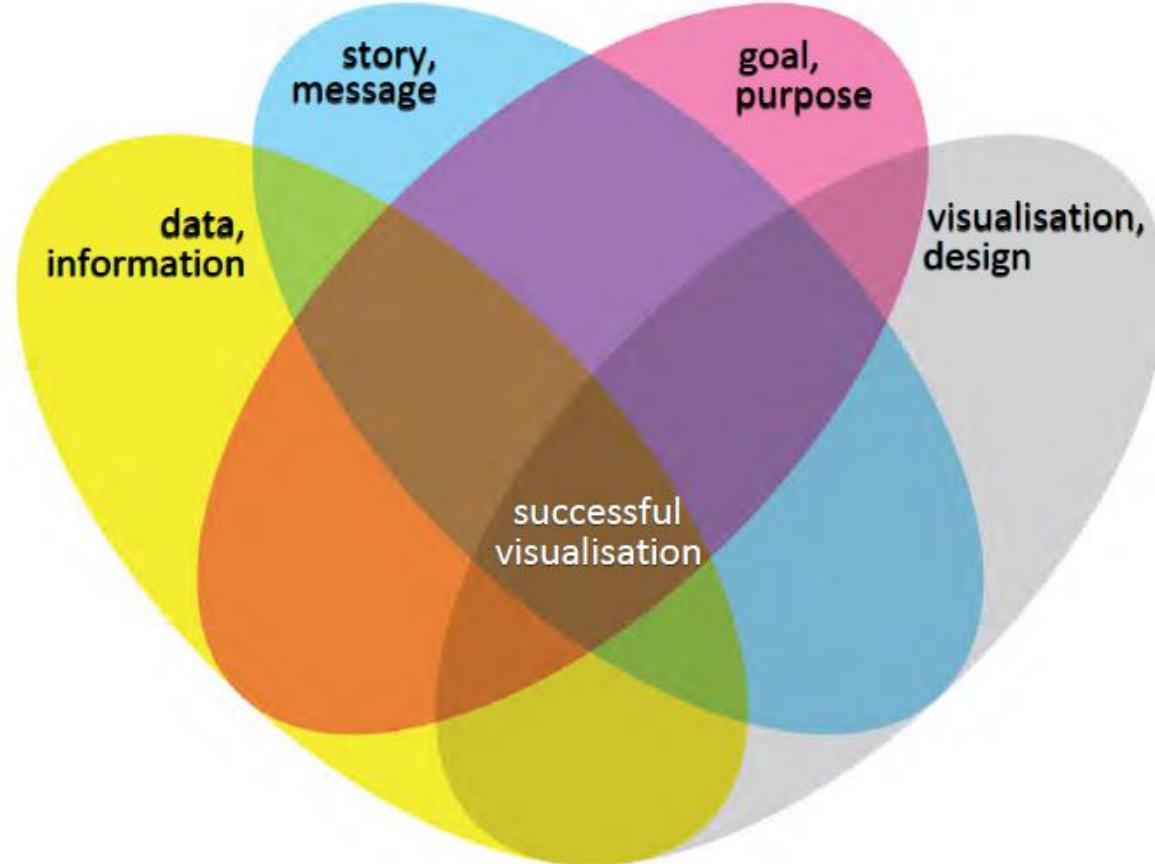
Half empty



Half full



What makes a good visualization?



by David McCandless



What makes a good visualization?

Distinguishing good/bad visualizations requires a design aesthetic, and a vocabulary to talk about data representations:

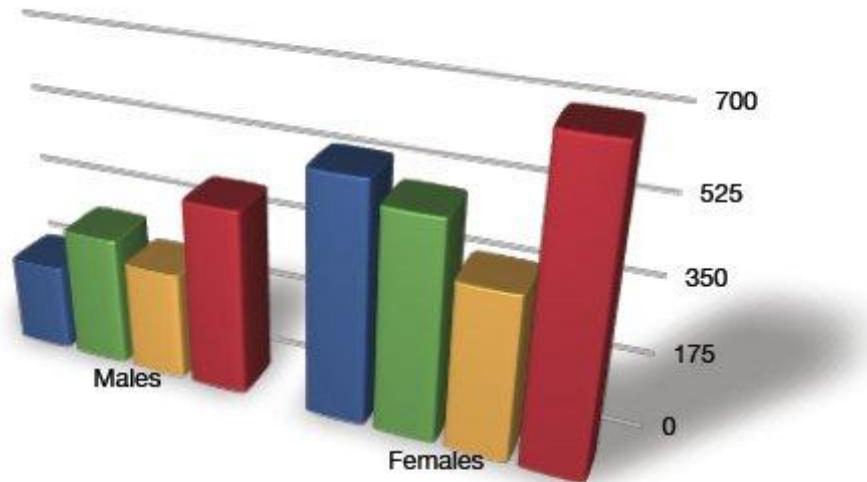
- Maximize data ink-ratio
- Minimize lie factor
- Minimize chartjunk
- Use proper scales and clear labeling



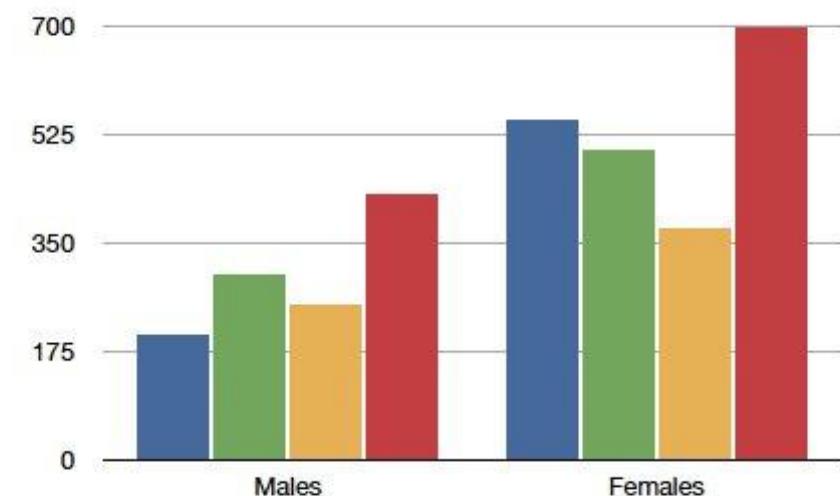
What makes a good visualization?

Maximize Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



■ 0-\$24,999 ■ \$25,000+ ■ 0-\$24,999 ■ \$25,000+



■ 0-\$24,999 ■ \$25,000+ ■ 0-\$24,999 ■ \$25,000+

What makes a good visualization?



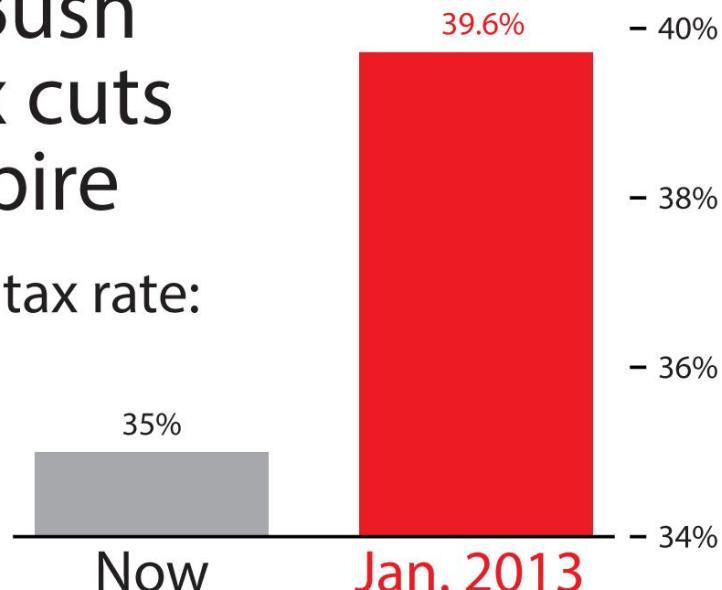
Minimize Lie Factor

(size of effect in graphic) / (size of effect in data)

Misleading

If Bush
tax cuts
expire

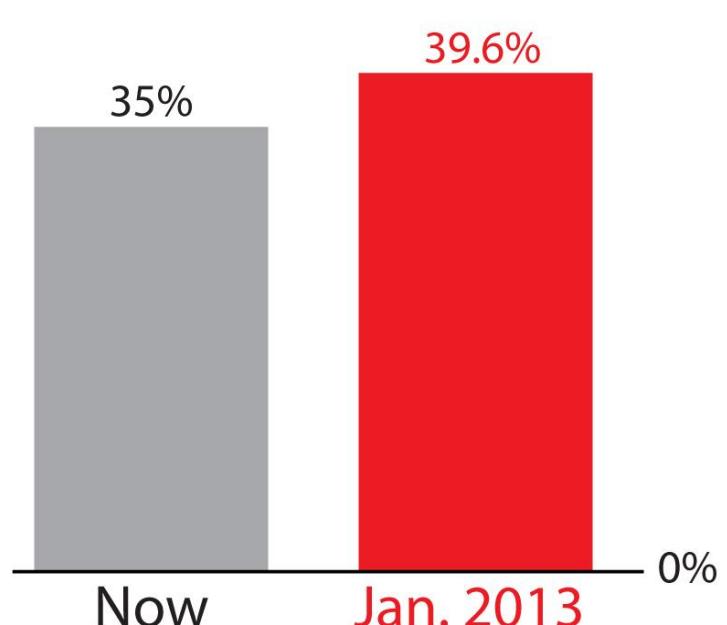
Top tax rate:



More accurate

If Bush
tax cuts
expire

Top tax rate:





What makes a good visualization?

Minimize Lie Factor





What makes a good visualization?

Minimize Chartjunk

Extraneous visual elements distract from the message the data is trying to tell.

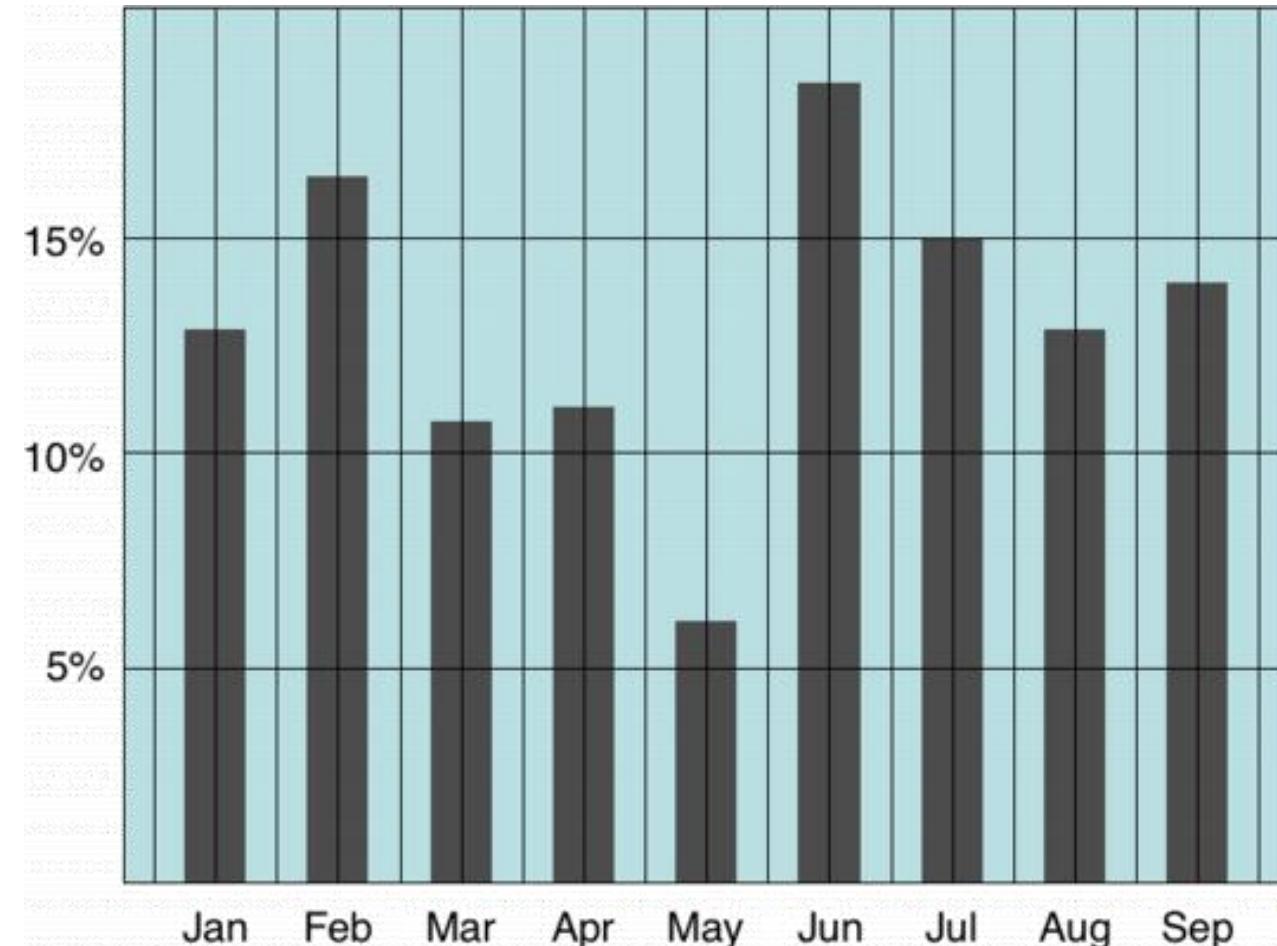
- Extra dimensionality
- Uninformative coloring
- Excessive grids and figurative decoration

In an exciting graphic, the data tells the story, not the chartjunk.



What makes a good visualization?

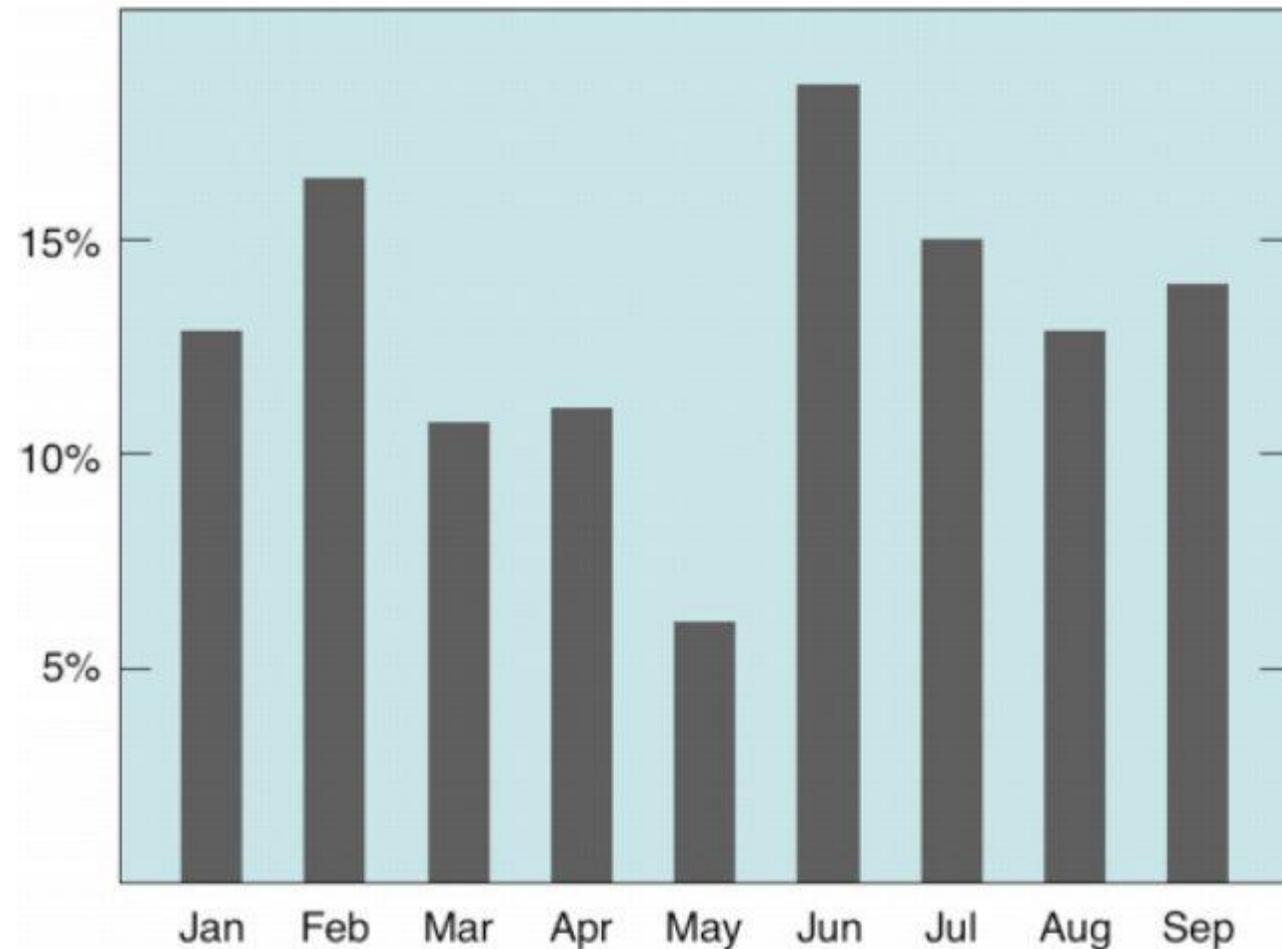
Minimize Chartjunk





What makes a good visualization?

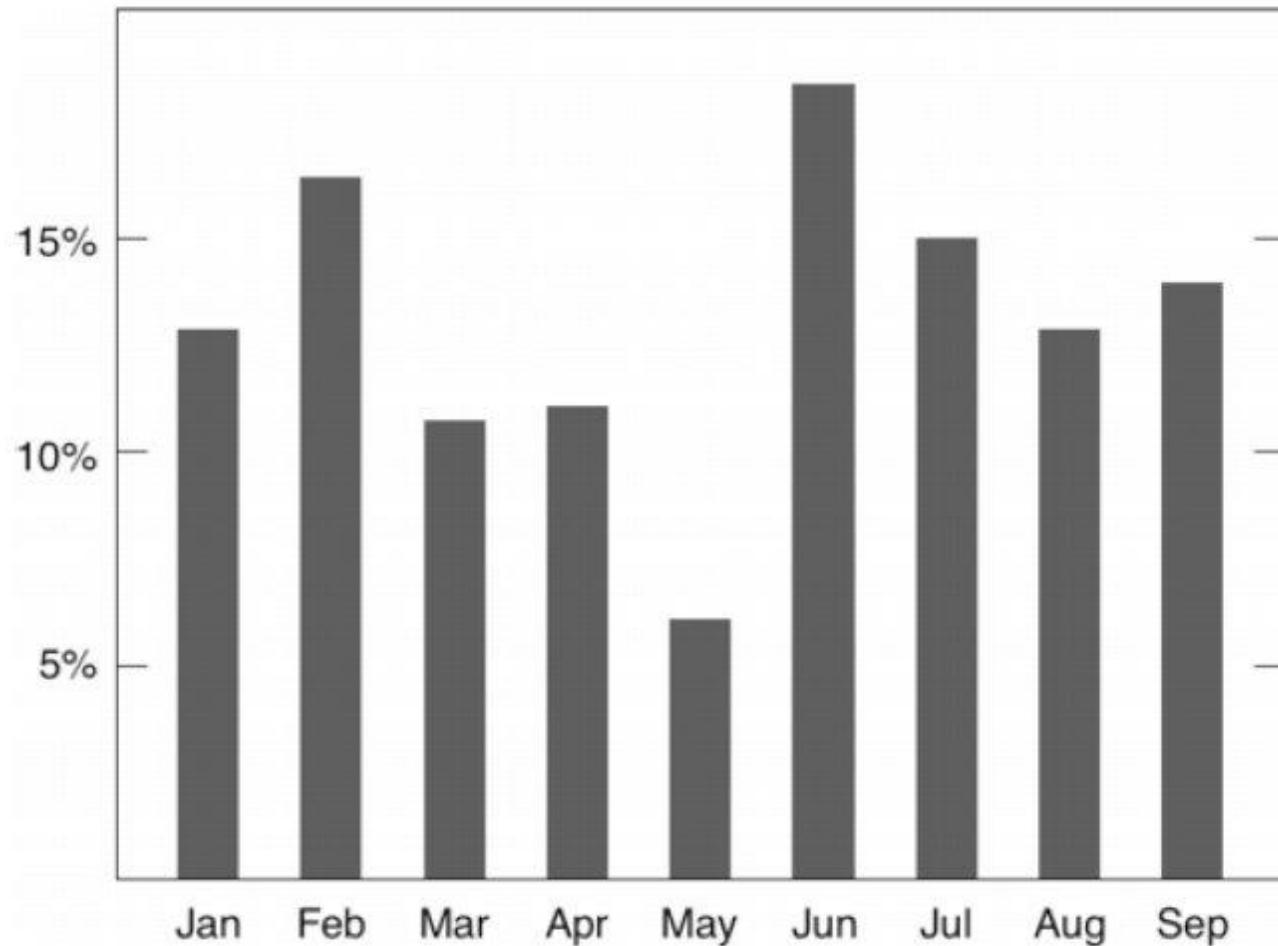
Minimize Chartjunk





What makes a good visualization?

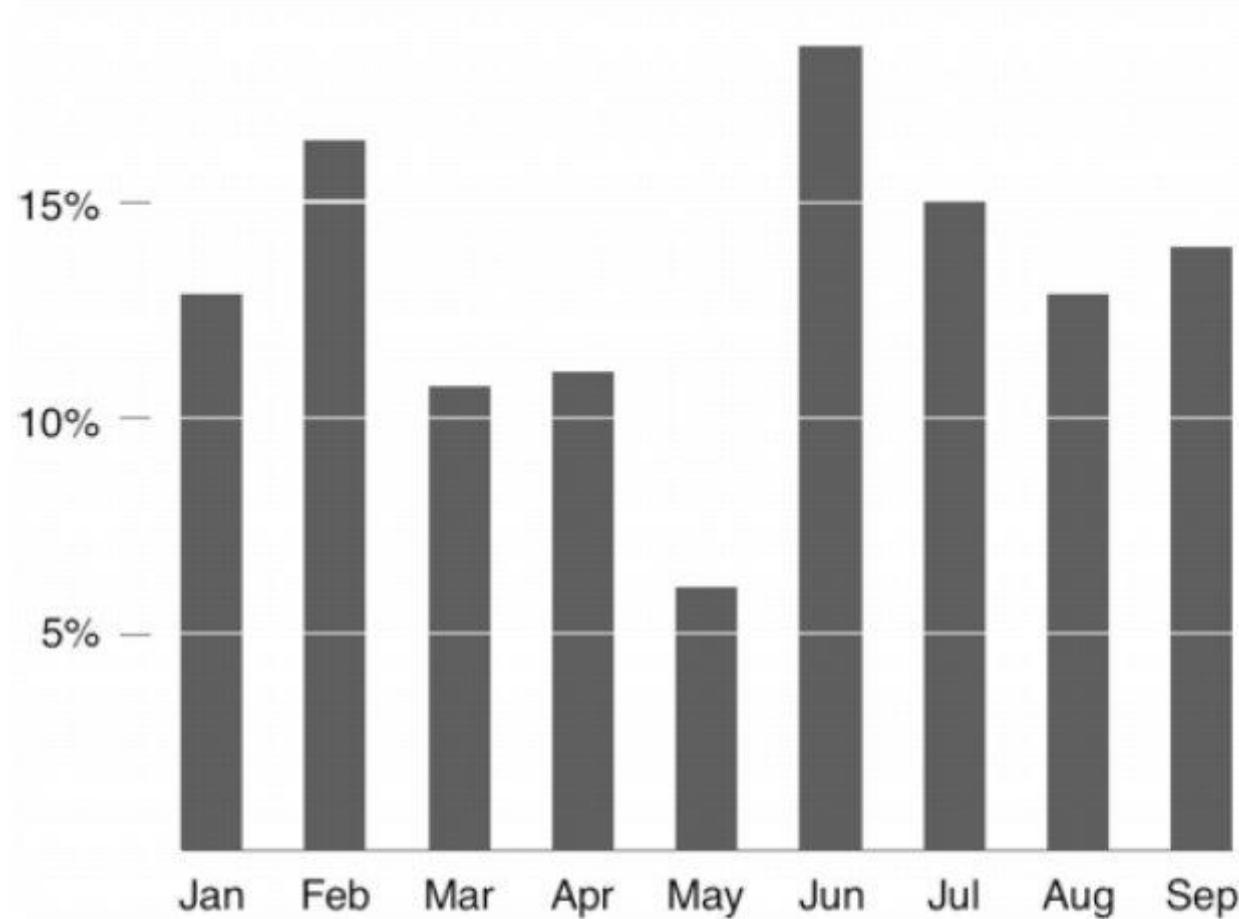
Minimize Chartjunk





What makes a good visualization?

Minimize Chartjunk





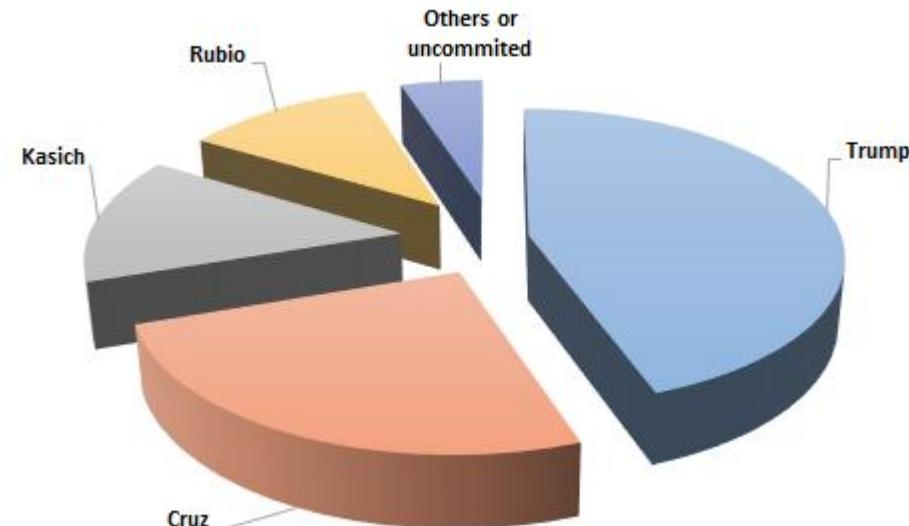
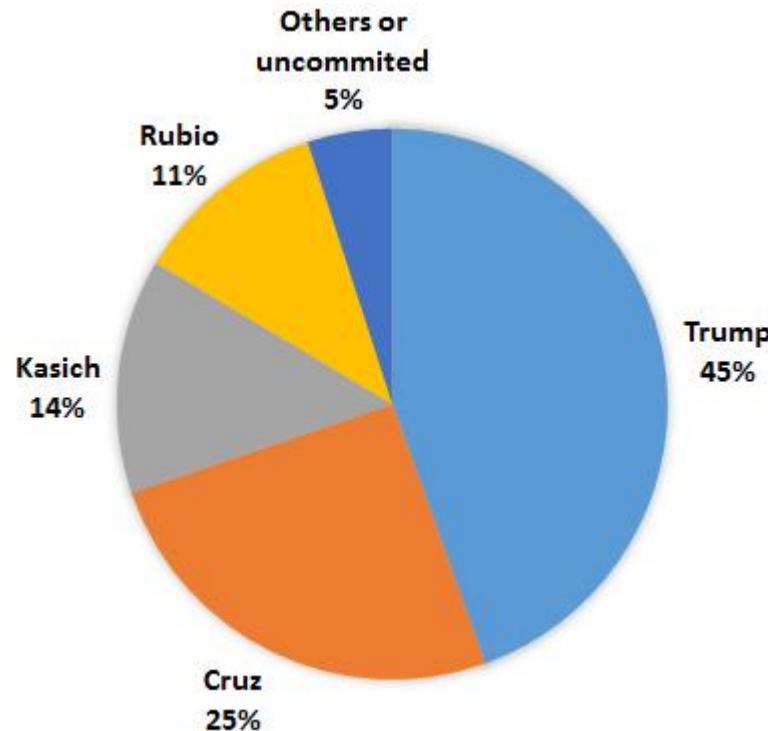
What makes a good visualization?

Minimize Chartjunk: less is more



What makes a good visualization?

Minimize Chartjunk: less is more





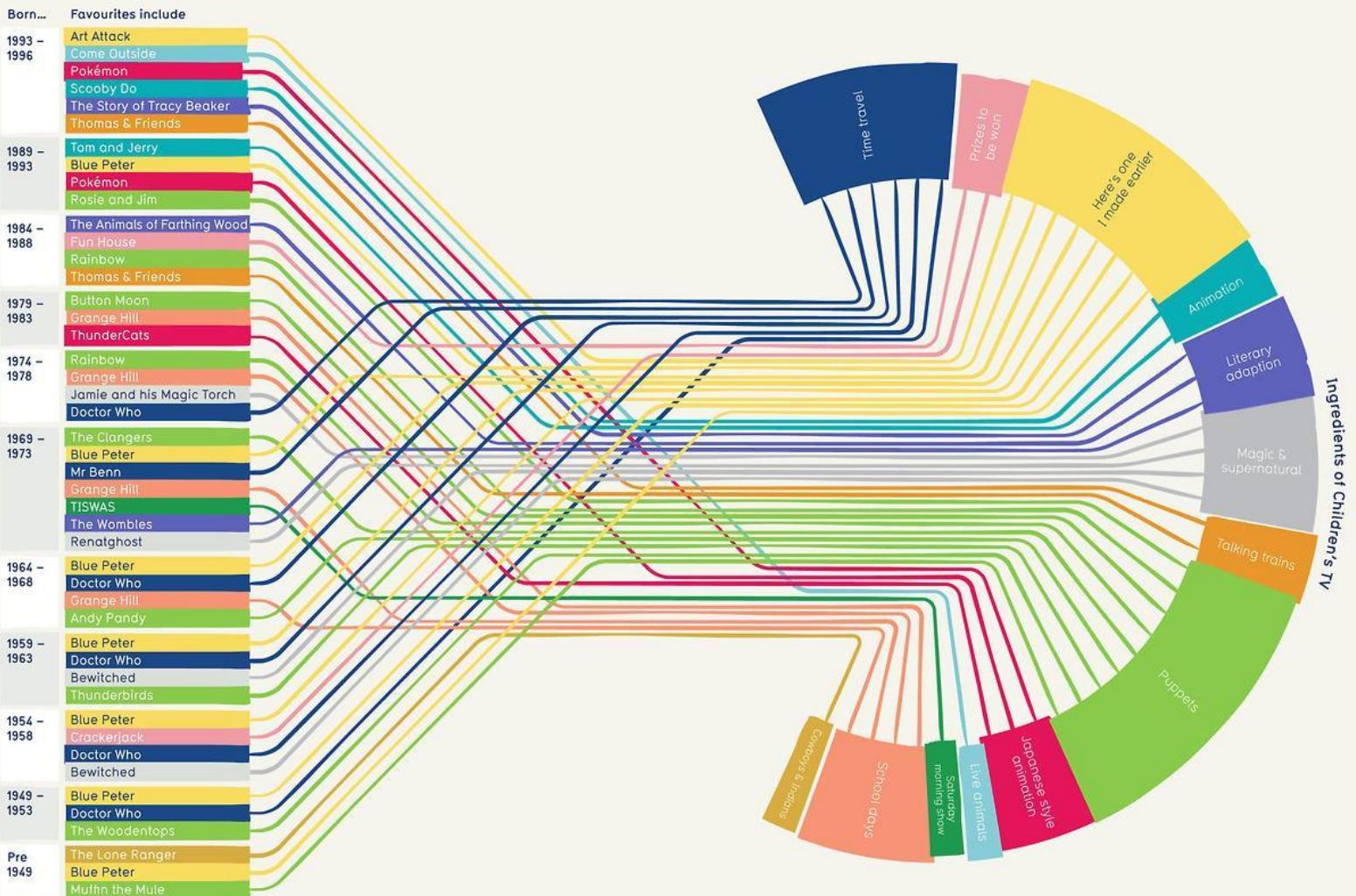
What

Recipe for Children's TV

We asked adults across the UK about their favourite TV programme as a child. The magic of puppetry and the creativity of arts and crafts were favourite themes for every age group. BBC standard, **Blue Peter**, was popular across all age groups and was nominated by almost one in ten of all respondents as their favourite programme.



www.tvlicensing.co.uk



tvlicensing.co.uk/telescope2014

@tvlicensingnews

youtube.com/tvlicensing

flickr.com/photos/tvlicensing

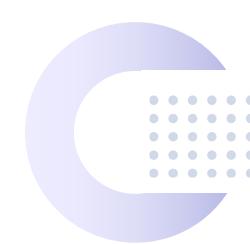


See bonus 'Trip Down Memory Lane' footage at youtube.com/tvlicensing

What

% Change in 5 Year Attendance of Top 25 Theme Parks



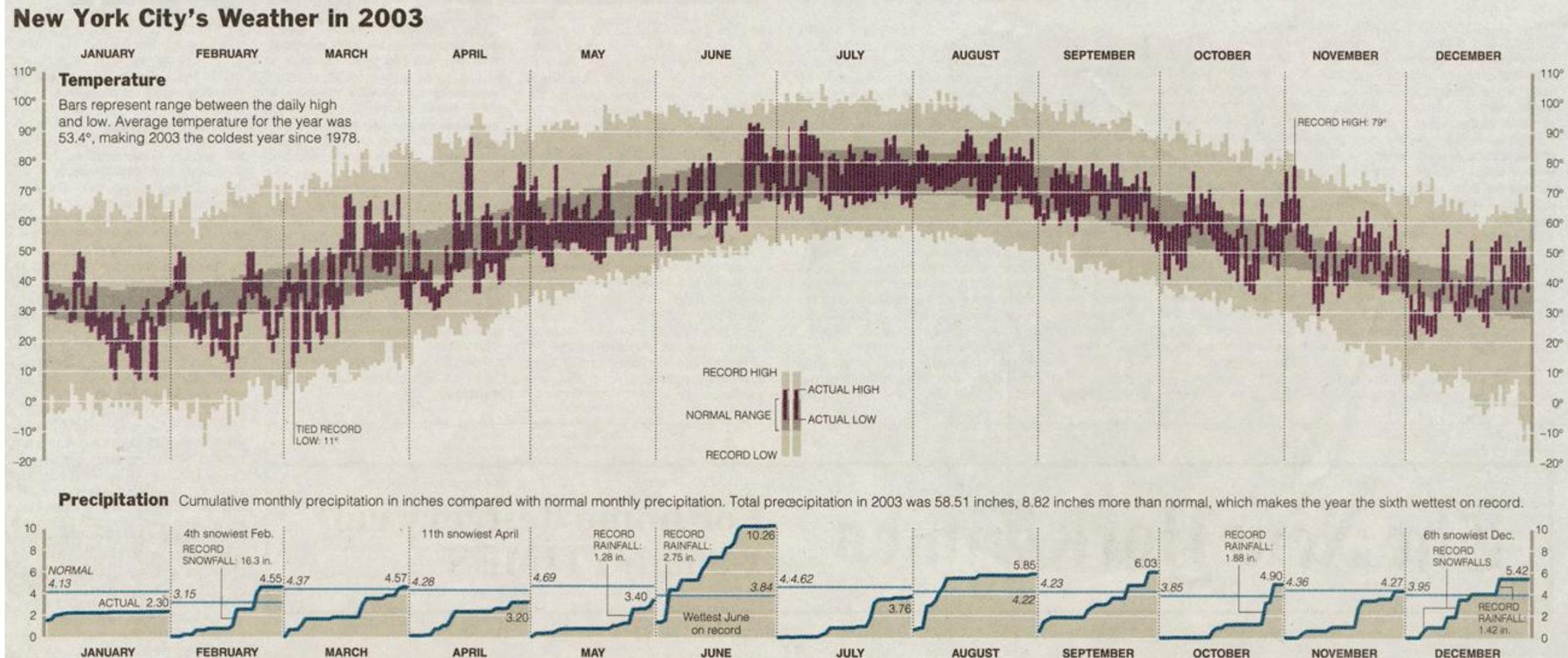


Most Common Break Up Times

according to Facebook status updates

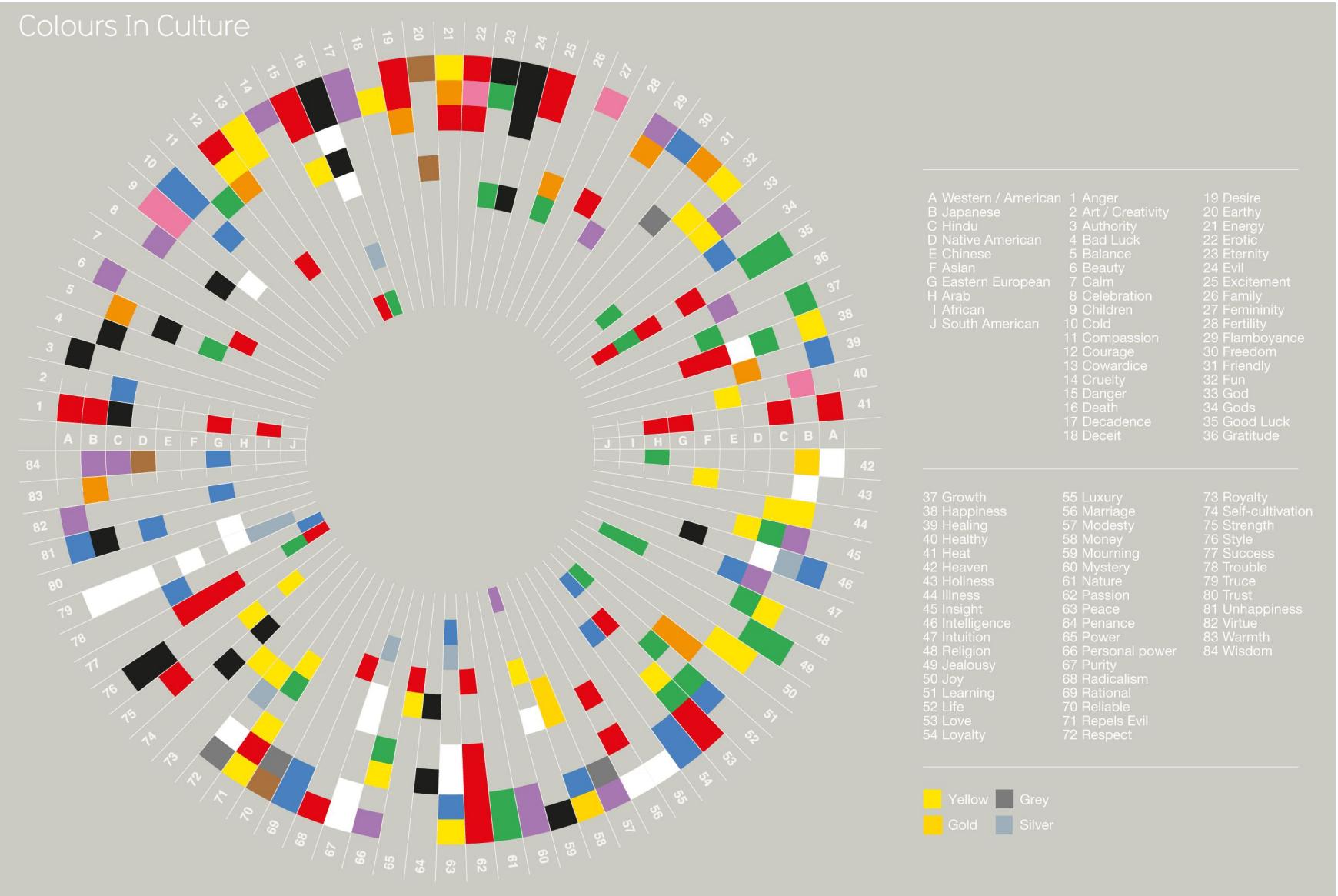


New York's Weather Year in Review



Visualization Case

David McCandless
Information is beautiful

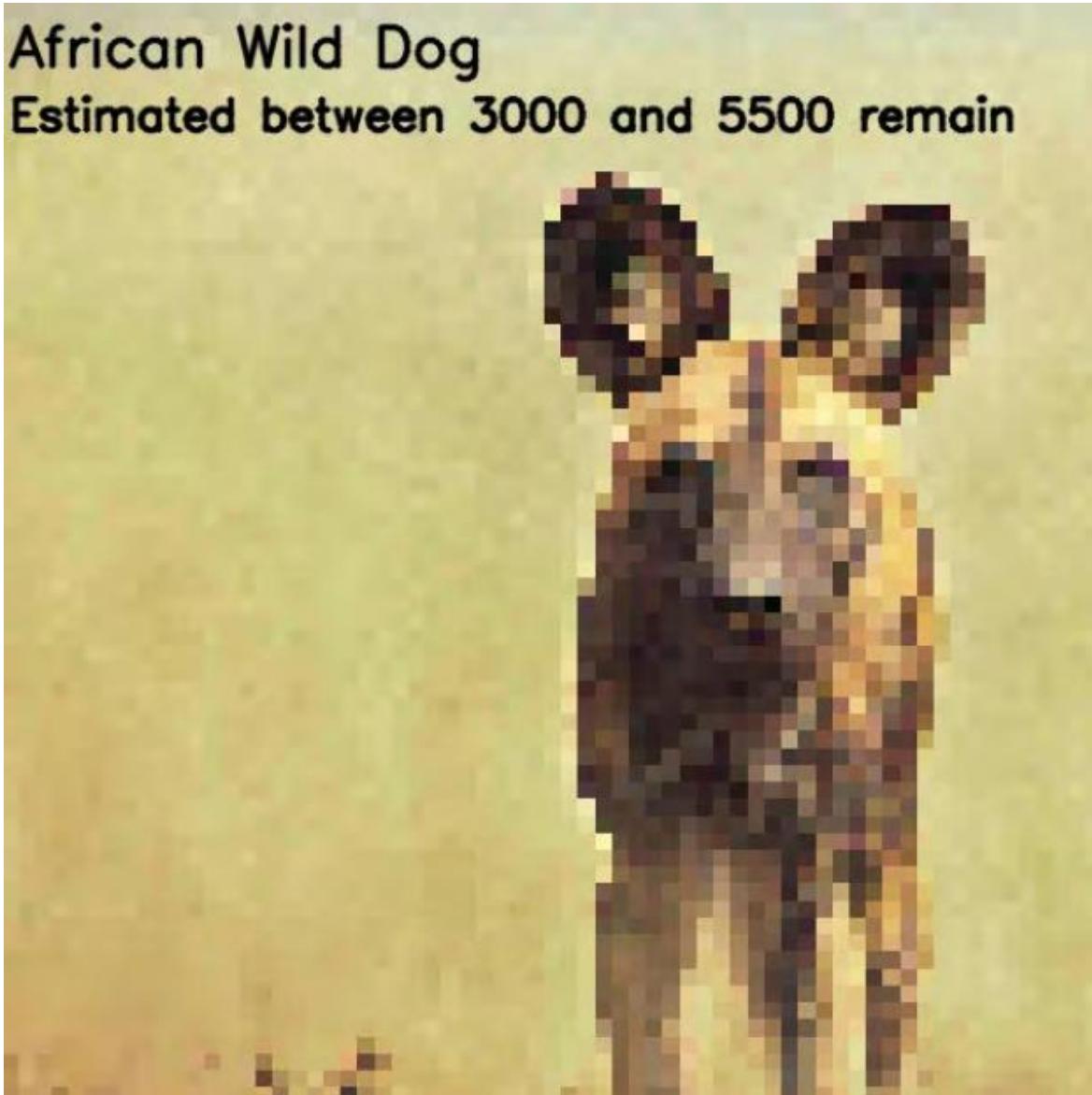




Visualization Case

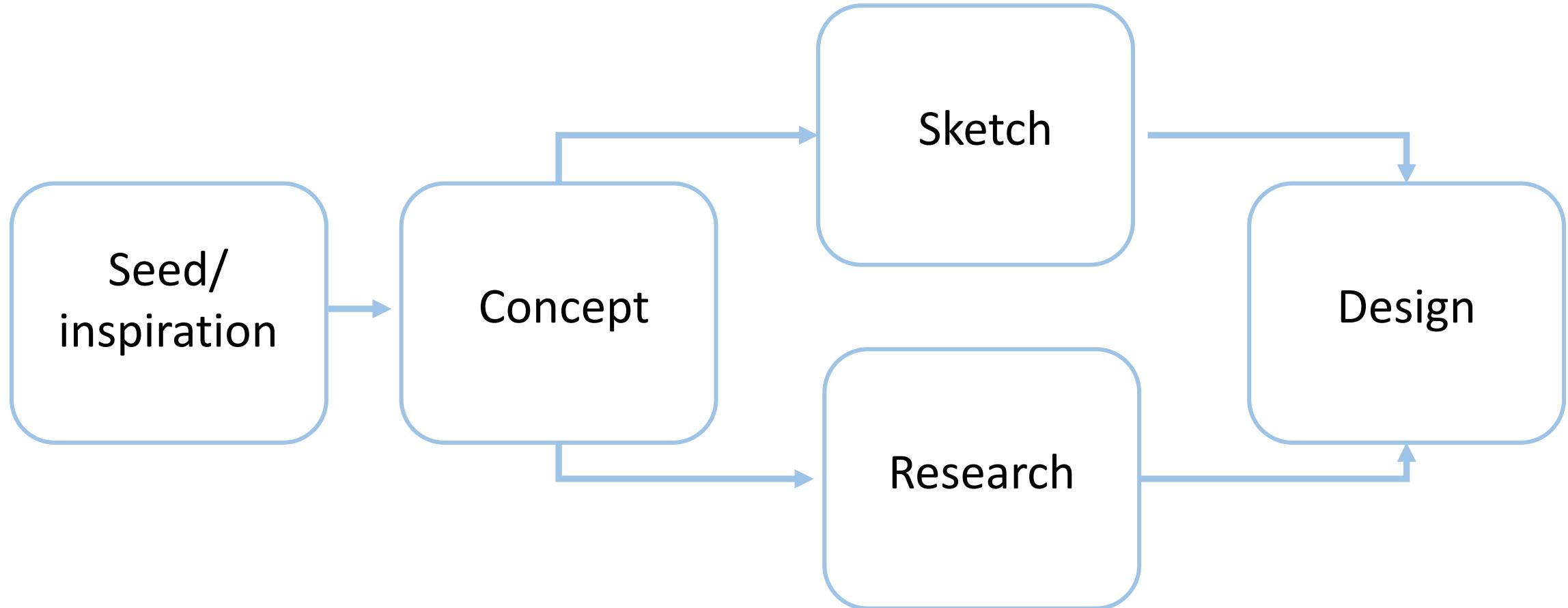
African Wild Dog

Estimated between 3000 and 5500 remain



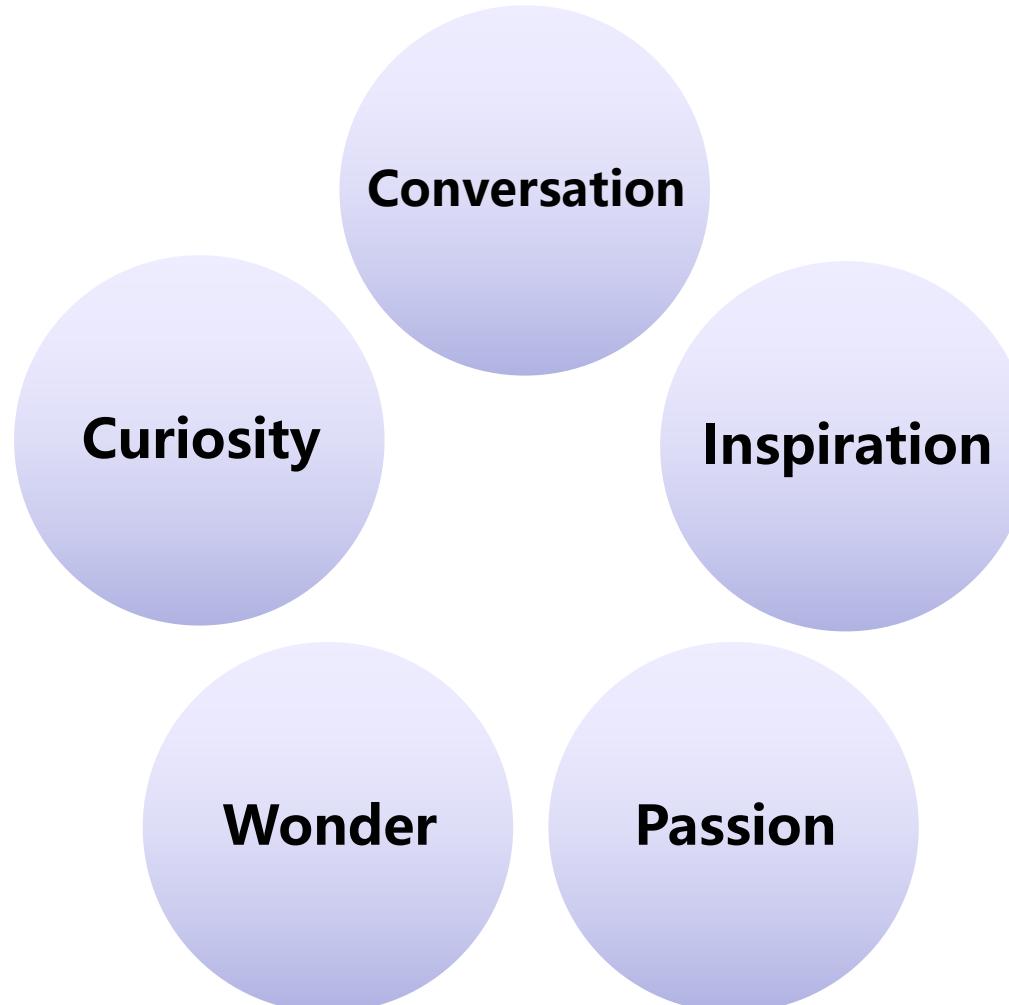


Visualization Design Process





Visualization Design Process: Inspiration



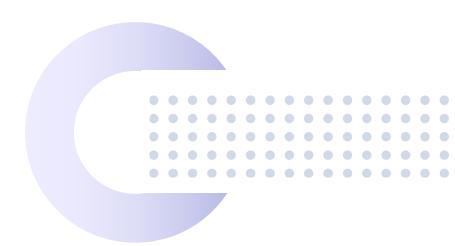


Visualization Design Process

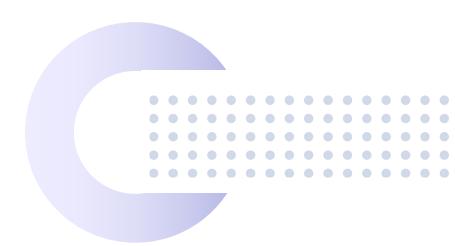


systematic analytic communication

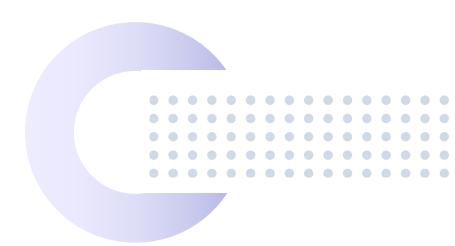
by David McCandless



10 min group challenge
Create 10 concepts for infographics
& visualizations



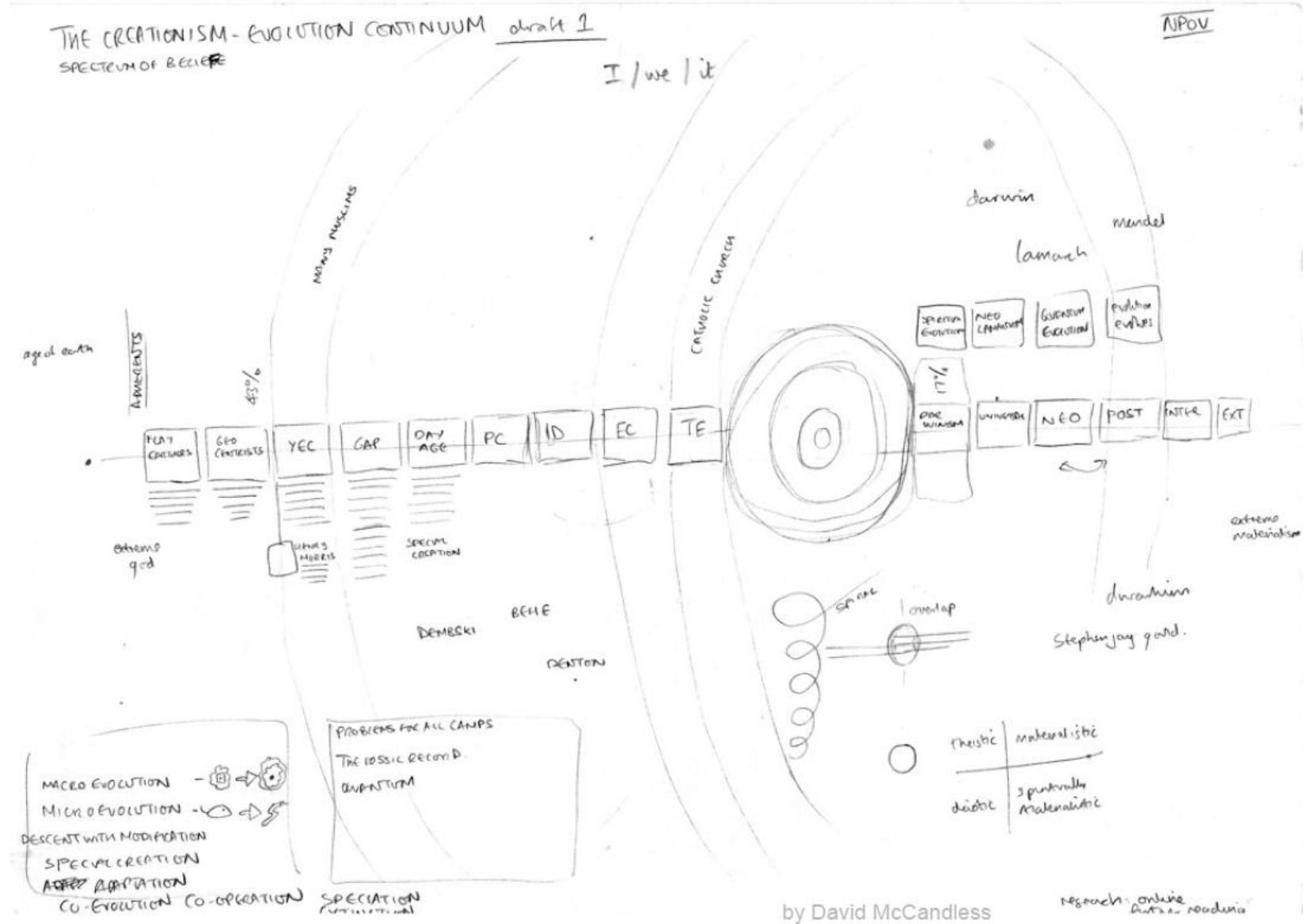
**Pick out 3 best concept
2 min**



**Choose 1 best concept
1 min**



Visualization Design Process: sketch





Visualization Design Process: sketch

Creationism vs. Evolutionism

Where did all this come from?

Circle size = approximate volume of believers

FLAT EARTHERS
This Earth is flat because the Bible says it is. Flat like a coin. Or perhaps a plate.

GEOCENTRICISM
Earth is centre of the universe. All modern science is bunk.

YOUNG EARTH CREATIONISM
The Earth is 6000 years old. Humans are not related to animals. And you will burn in Hell if you don't believe this.

GAP CREATIONISM
The world and mankind created in 7 days, but then there was a gap of a billion years or so.

DEISM
God created Earth and man but has been absent since.

SCIENTOLOGY
Galactic ruler Xenu gathered the disembodied spirits of alien beings called Thoptans into human form.

MOST IRRELIGIOUS

PROGRESSIVE CREATIONISM

God created man and the animals.
Everything else evolved.

INTELLIGENT DESIGN
Because things are so well designed, they could not have evolved.
Therefore God must've created them.

THEISTIC EVOLUTION

God creates through evolution
Evolution is fine. But God intervenes at critical moments in the history of life.
The Hand of God is also necessary for the creation of the human soul.

HINDU EVOLUTION

God is evolution
Evolution is a spiritual process.
Humanity is God incarnate, evolving towards a realization of this living truth.

POSITION OF THE CATHOLIC CHURCH

BUDDHISM
The creator of...
unbelievable... Belief in... What is...
Belief in...

UNIVERSAL DARWINISM

All universal processes, yes!...
all are governed by Darwinian evolution.

DARWINIAN EVOLUTION

Survival of the fittest

Random mutations give certain properties a survival advantage in their environment. This advantage is passed on to their offspring via reproduction of some other mechanism... Heredit... heredity.

CO-EVOLUTION
Co-operation, not competition, between organisms accelerates evolution.
"Survival of the nicest."

GROUP SELECTIONISM

Entire species are actually higher-level organisms in their own right. Natural selection affects species differently.

CONVERGENT EVOLUTION

Complex elements like eyes have evolved more than teeth. Huh?
Exactly. Deeper structures are at play.

NEO LAMARCKISM

Mutation of DNA creates...
changes. Called "Lamarckism".

PUNCTUATED EQUILIBRIUM

Evolution is essentially static until sudden, dramatic events shift it into a higher gear.

LAMARCKISM

Traits acquired by parents during lifetime can be passed on to their offspring. Hmmm... interesting.

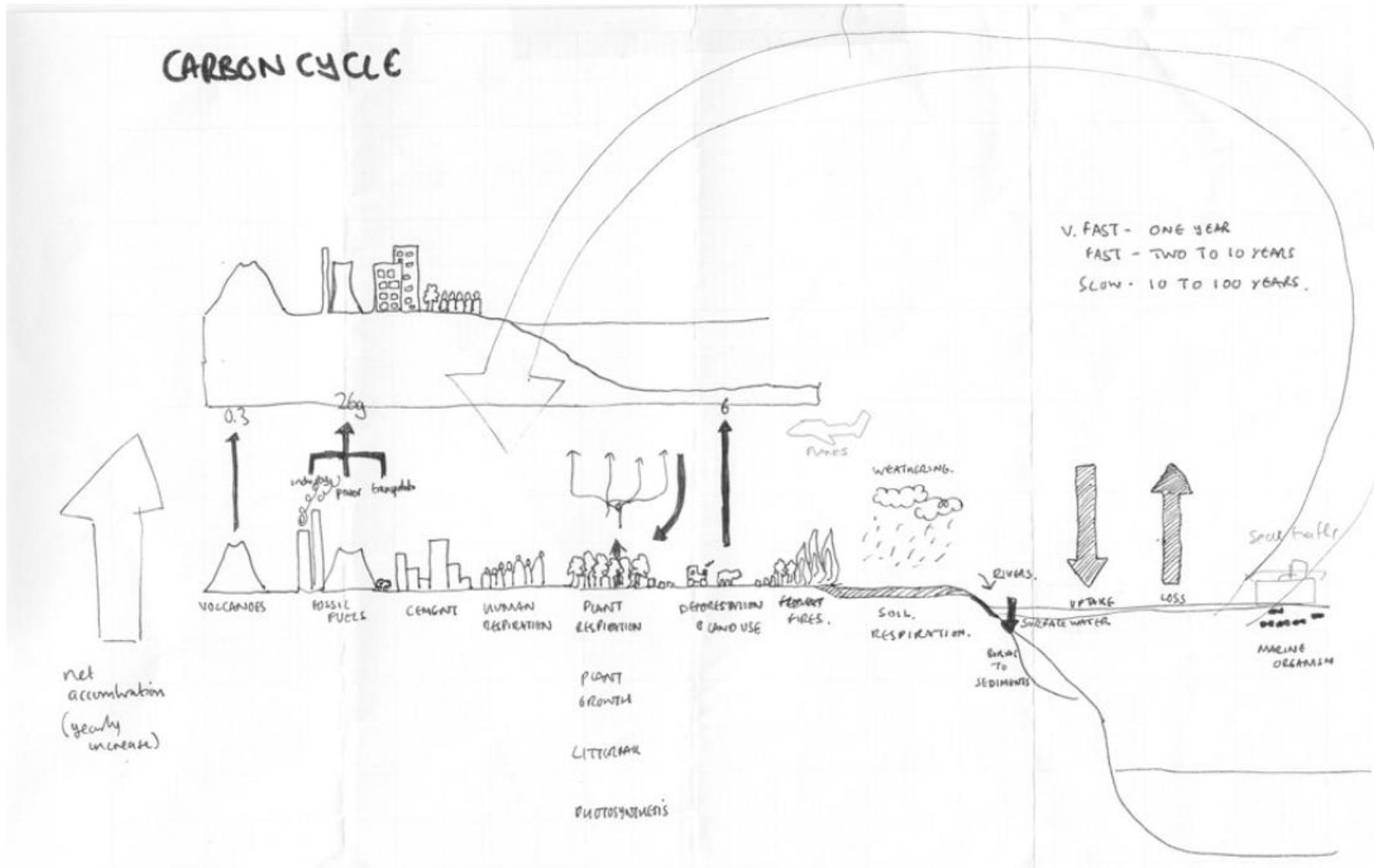
ONE EXTREME

THE OTHER

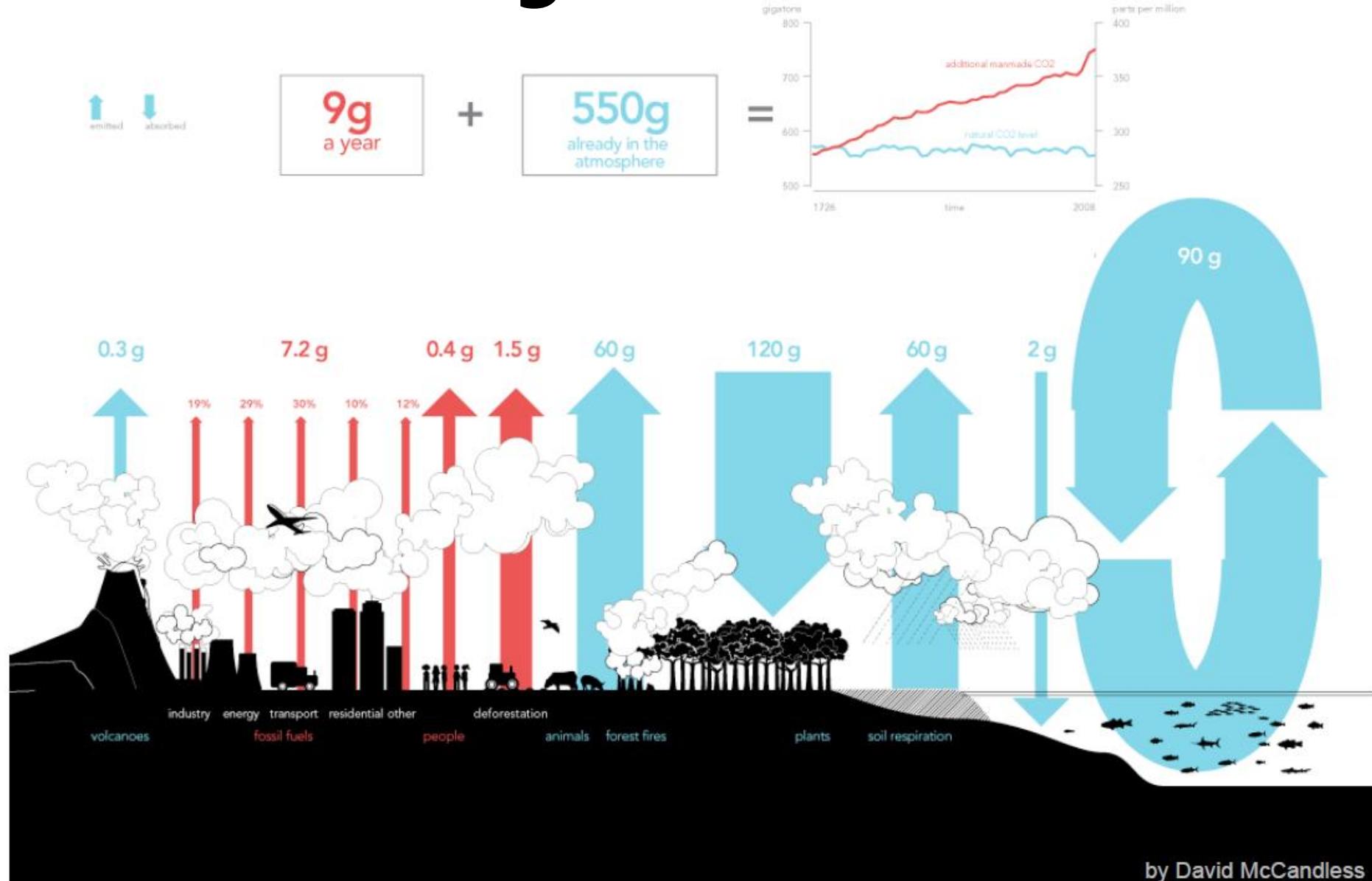
by David McCandless

On courtesy to Rebecca Xu

Visualization Design Process: sketch



Visualization Design Process: sketch



Visualization Design Process: research



Simple Solar System	Concept Fan	World Map with Bubbles	Bubble World Map	Tree Map	Bubble Tree	Pyramidal Hierarchy	Icicle Tree	Charticle	Dunno what to call this	Concentric Circles	Scatterplot Bubble	Scatterplot	Arc Diagram	Balloon Race	Word Cloud
Sliced with Annotations	Circle Packing	Fluid Process	Line Chart	Thermometer	Bubble Comparison	Bubble Comparison - Nested	Cycle Diagram	Pie Chart	Cox Comb	Multiple Incomplete Pie Charts	Hollow Pie Chart	Radial Icicle Tree	Radial Segment	Spiral	Polar Chart
Metro Map	Organic Mind Map	Cone Tree	Family Tree Vertical	Stacked Area Chart	Landscape - Diagram	Area Chart - Squared	Stratigraphic - Column	Sunburst	Radar Chart Cobweb	Venn Diagram	Swim Lane	Stacked Bar Chart	Histogram	Bar Chart	Periodic Table
Area Histogram	Area Chart	Bubble Clusters	Bubble Chart	Bubble Connection - Stacks	Bubble Connection - Venn Style	Bubble Network	Sankey	Matrix Bubble	Matrix Shape	Matrix Pattern	Comparison Table	Bubble Table	Matrix	Matrix Table	Abacus
Star Ring	Dendrites	Mind Map - Hyperbolic	Runway	Clustering	Semantic Radial	Semantic Regimental	Atmospheric Timeline	Table	Box Plot	Triangle	Cartesian Grid	Cartesian Coordinates	Small Multiples	Mass - Scatter	Heatmap Table

Visualization Design Process: research



15 min challenge Sketch out an infographic

Choose a form, consider what visual channels it has (and how many) and then consider how you might map your metrics to those channels

A 做的都对

B 超级侦探认真干饭

C 百分百暴击的水云天苍云海...

D 扫黑除恶专项调查

E !_!

F 啊对对队

G SkyNet

H 你说对不队

I 光宗耀组

J -

K 解封队

Thank You

