Data Visualization
– an introduction –

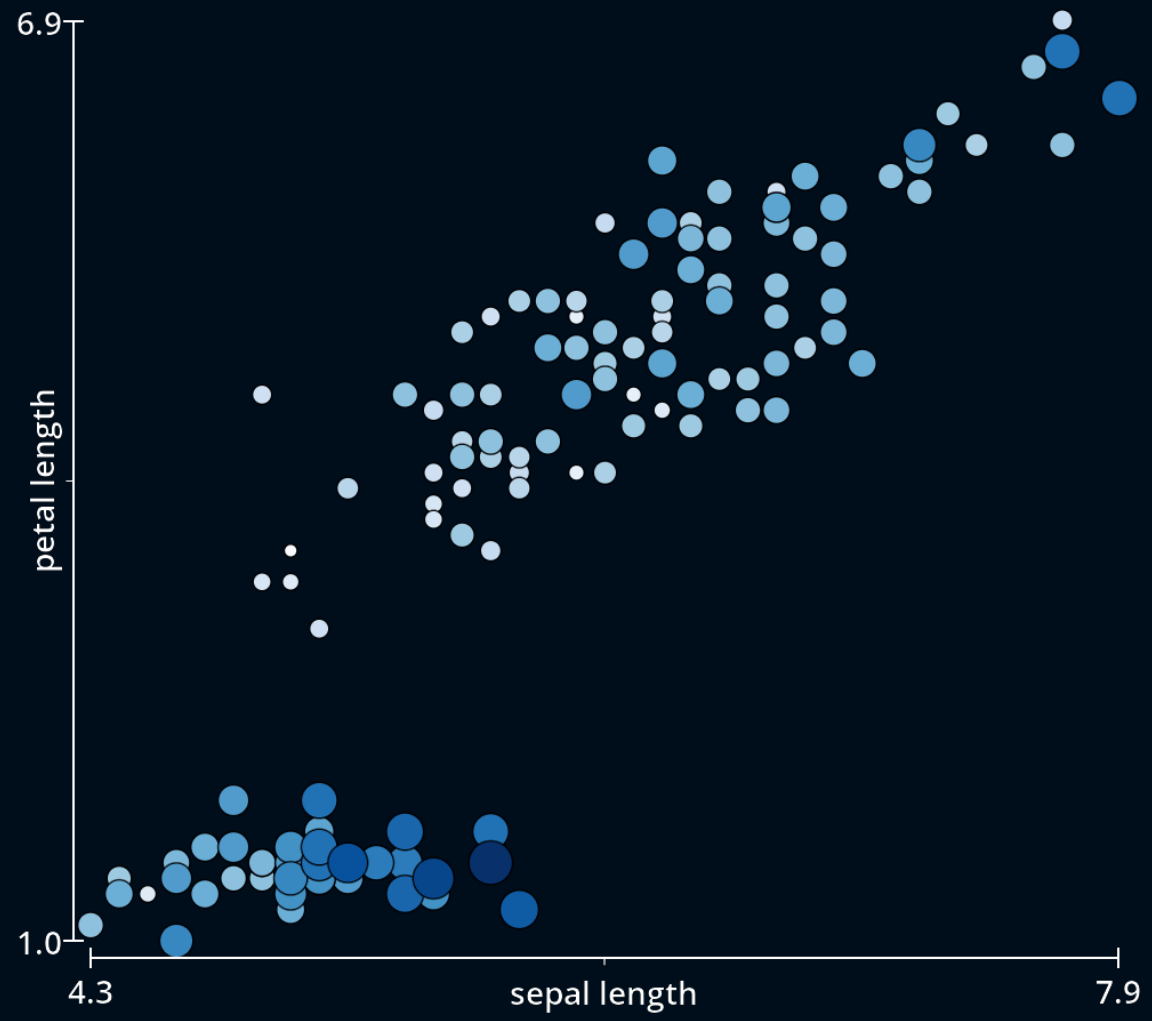
Martin Falk, Daniel Jönsson, Alexander Bock

Scientific Visualization @ LiU

LINKÖPING UNIVERSITY

WASP Community Building Summer School 2022

What is visualization?



Plume from cigar, © Hunter Miller [flowvis.org]

What is visualization?

visualization noun

vi·su·al·i·za·tion | \ ˌvi-zhə-wə-lə-ˈzā-shən, ˌvi-zhə-lə-, ˌvizh-wə-lə- \

Definition of *visualization*

1 : formation of mental visual images

2 : the act or process of interpreting in visual terms or of putting into visible form

3 : the process of making an internal organ or part visible by the introduction (as by swallowing) of a radiopaque substance followed by radiography

[merriam-webster.com]

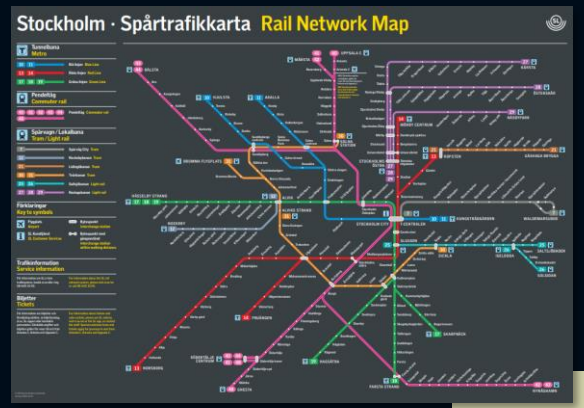
Plume from cigar, © Hunter Miller [flowvis.org]

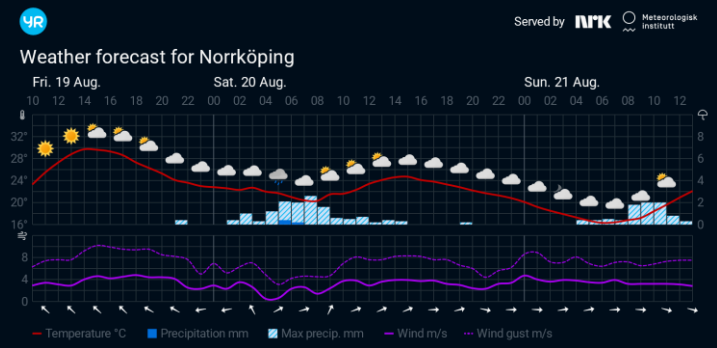
Have you been in touch with visualization?

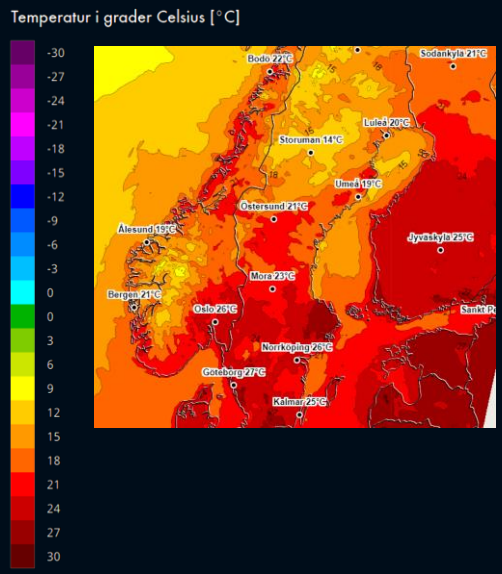
Think of examples in

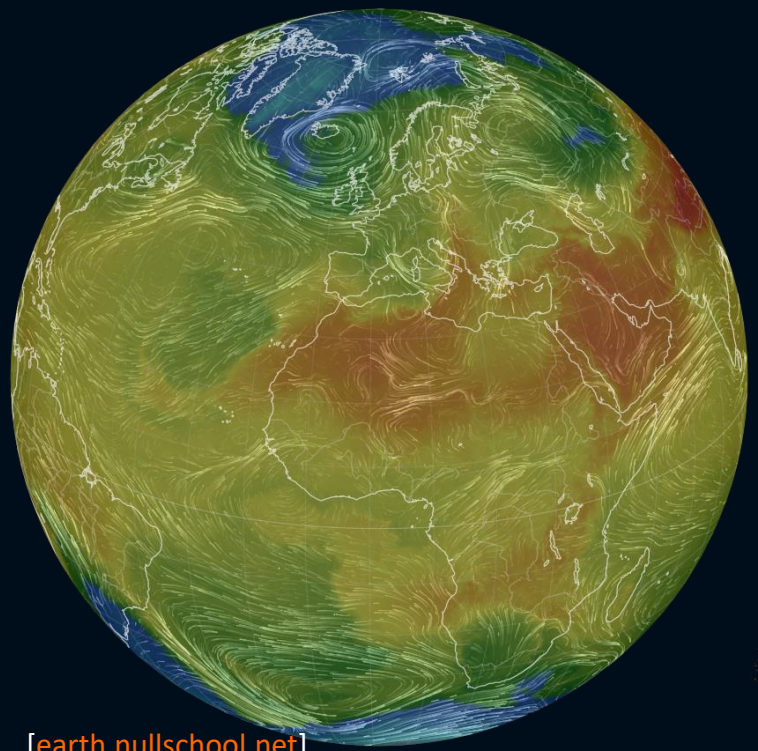
- Your everyday life
- Your work
- Around you

Examples

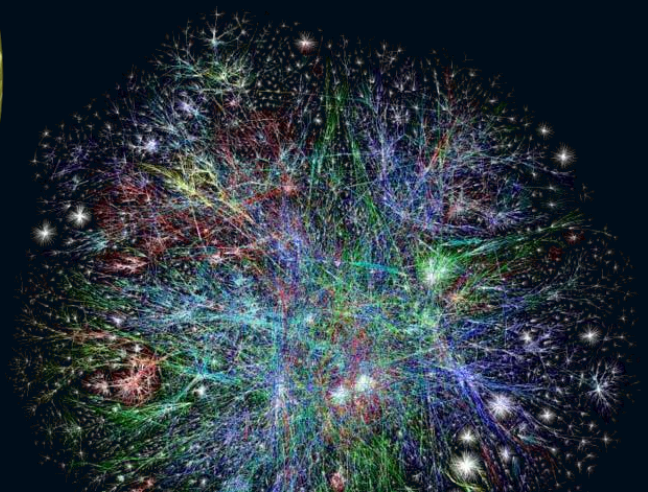




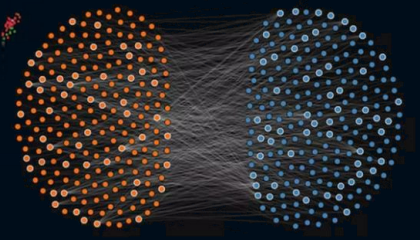




[earth.nullschool.net]

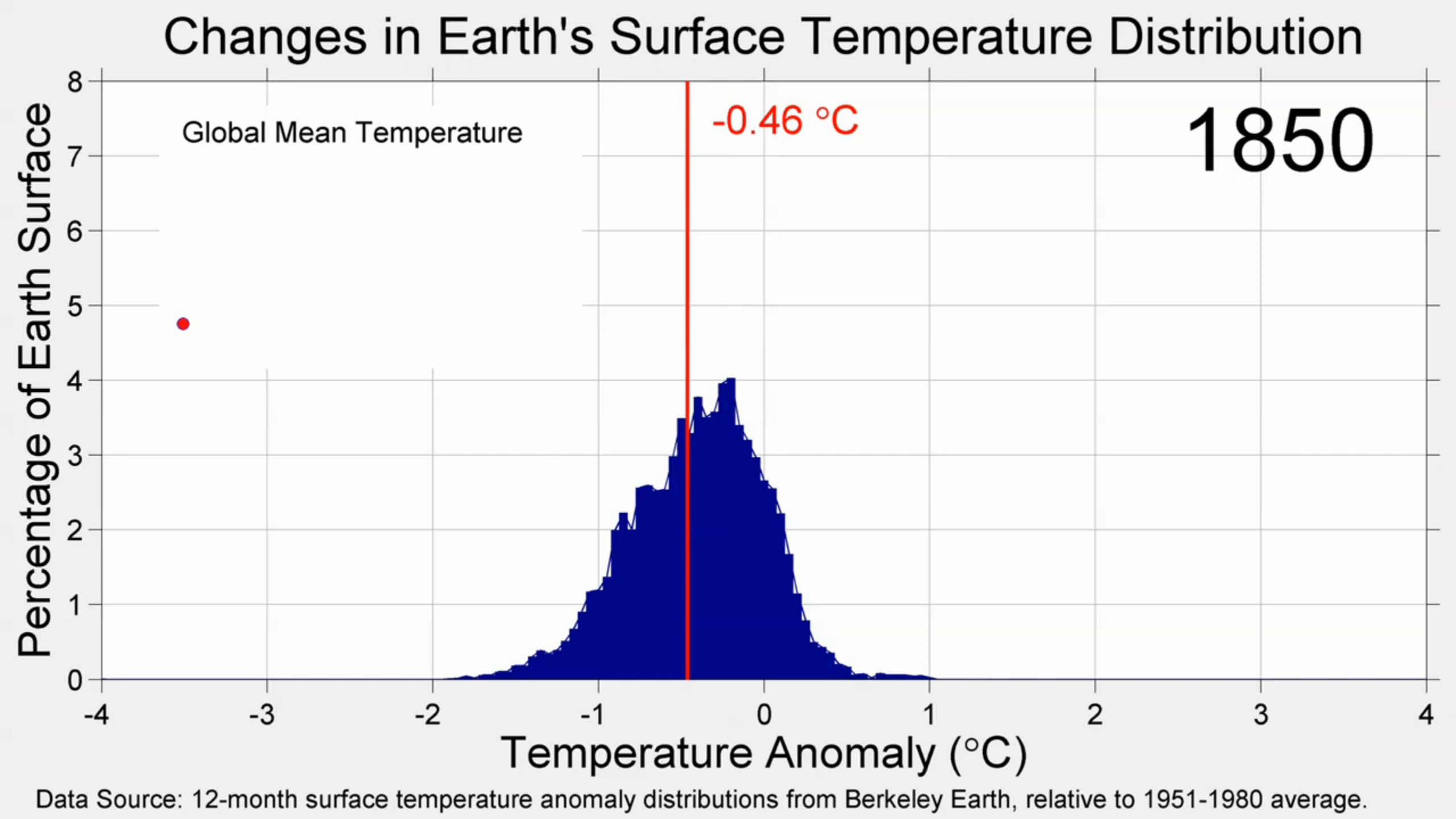


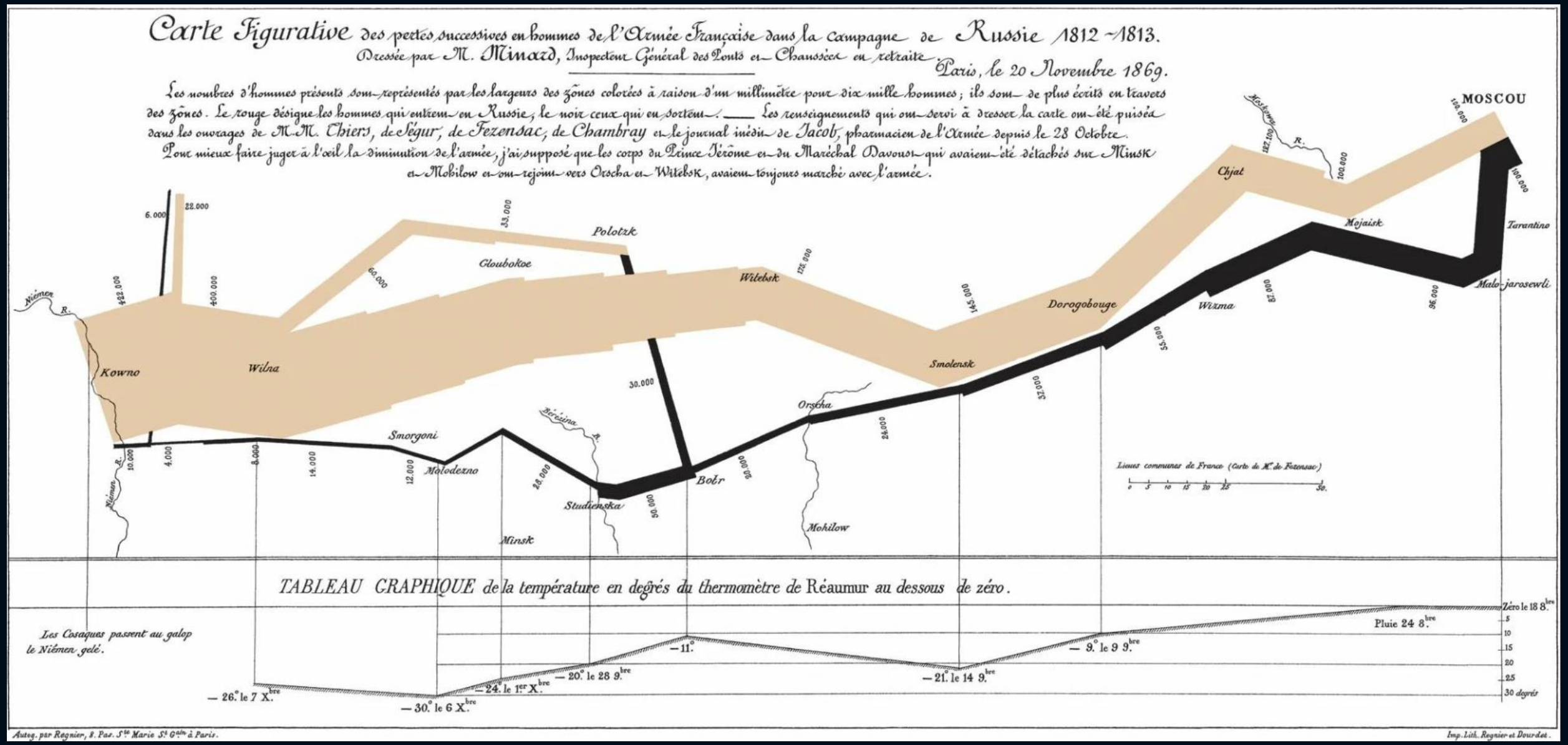
[svt.se/datajournalistik]



SWEDBANK

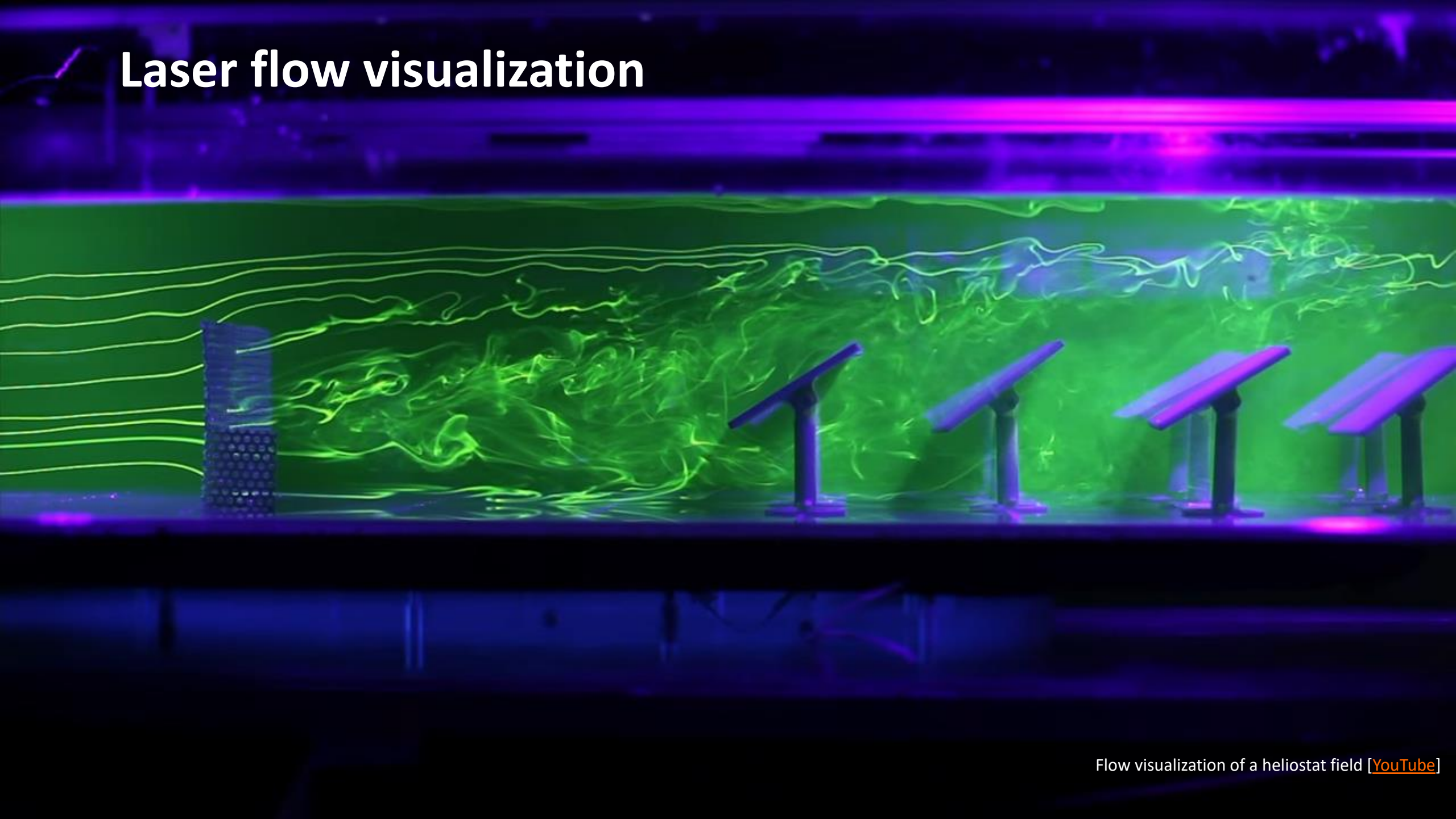
Så har pengarna skickats mellan misstänkta kunder


Changes in Earth's Surface Temperature Distribution
Global Mean Temperature
-0.46 °C
1850
Percentage of Earth Surface
0
1
2
3
4
5
6
7
8
-4
-3
-2
-1
0
1
2
3
4
Temperature Anomaly (°C)
Data Source: 12-month surface temperature anomaly distributions from Berkeley Earth, relative to 1951-1980 average.



[openculture.com]

Laser flow visualization



Flow visualization of a heliostat field [YouTube]

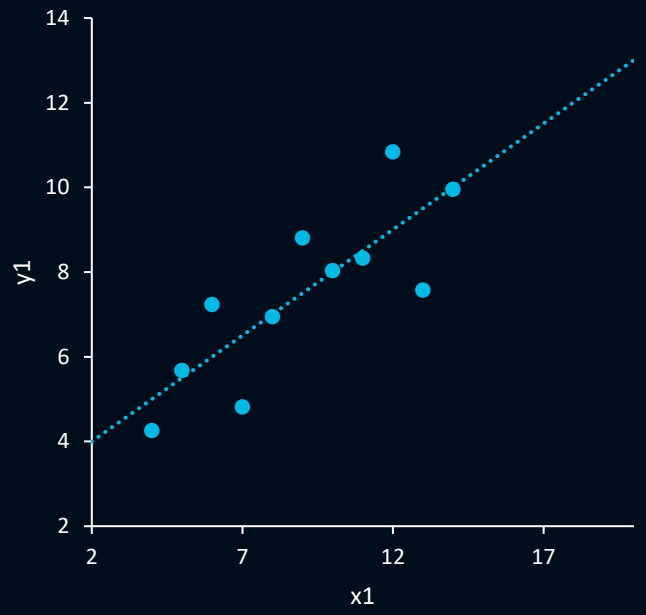
Why visualization?

I		II		III		IV	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

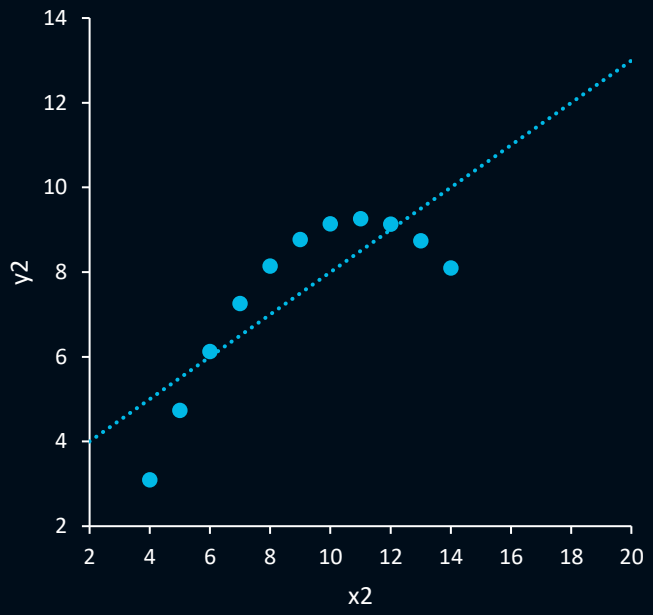
Why visualization?

I		II		III		IV	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean x_1: 9 Mean y_1: 7.5 Var x_1: 11 Var y_1: 4.1 Corr(x_1, y_1): 0.8 Linear regression: 3.0 + 0.5x_1		Mean x_2: 9 Mean y_2: 7.5 Var x_2: 11 Var y_2: 4.1 Corr(x_2, y_2): 0.8 Linear regression: 3.0 + 0.5x_2		Mean x_3: 9 Mean y_3: 7.5 Var x_3: 11 Var y_3: 4.1 Corr(x_3, y_3): 0.8 Linear regression: 3.0 + 0.5x_3		Mean x_4: 9 Mean y_4: 7.5 Var x_4: 11 Var y_4: 4.1 Corr(x_4, y_4): 0.8 Linear regression: 3.0 + 0.5x_4	

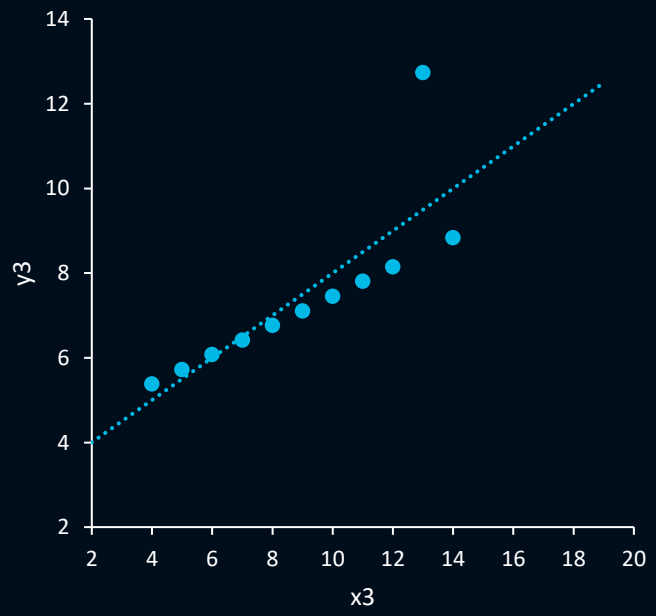
I


14
12
10
8
6
4
2
y1
2
7
12
17
x1

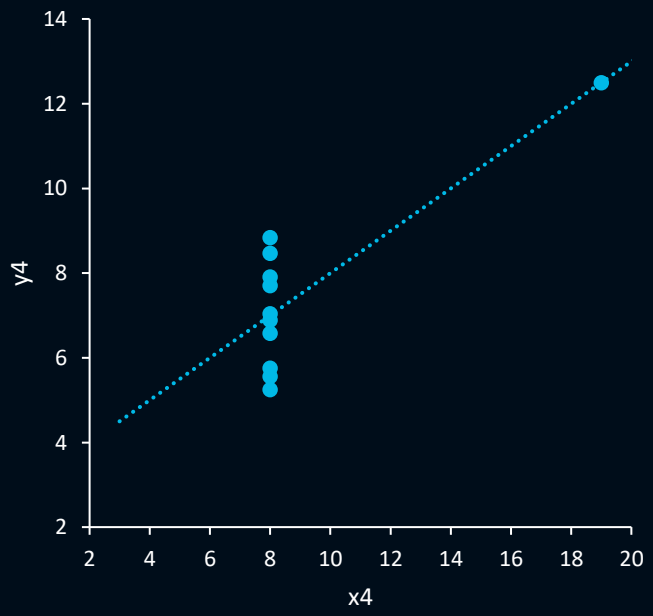
II


14
12
10
8
6
4
2
y2
2
4
6
8
10
12
14
16
18
20
x2

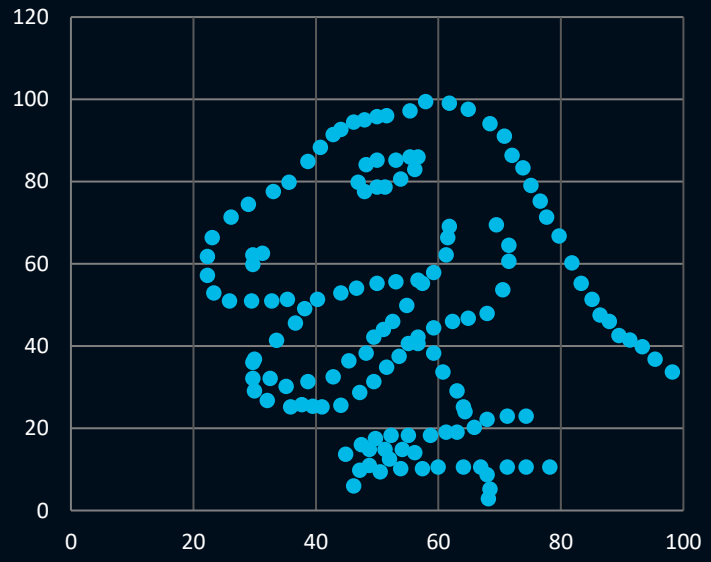
III

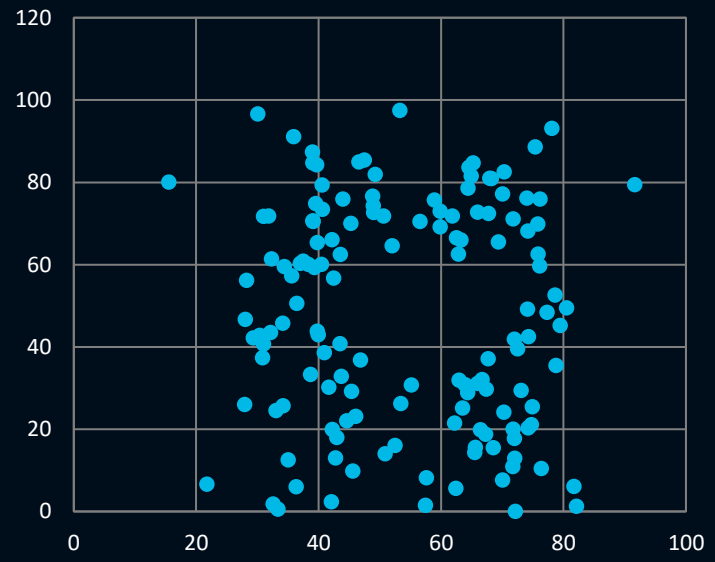

14
12
10
8
6
4
2
y3
2
4
6
8
10
12
14
16
18
20
x3

IV

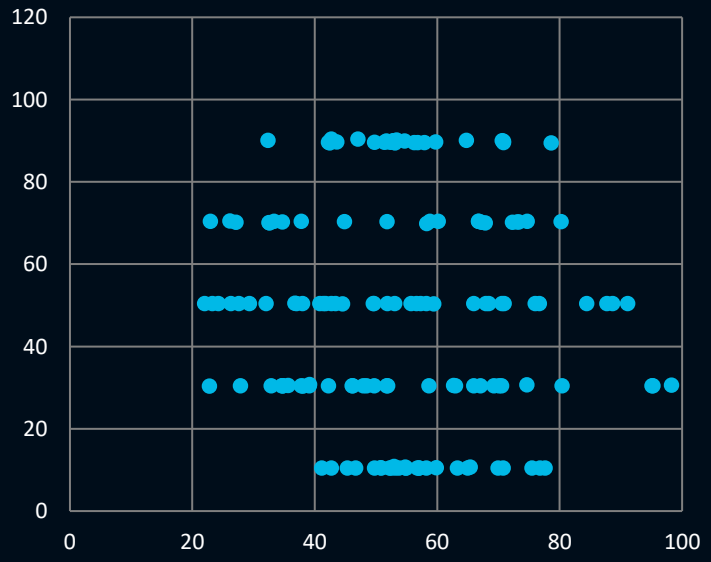

14
12
10
8
6
4
2
y4
2
4
6
8
10
12
14
16
18
20
x4

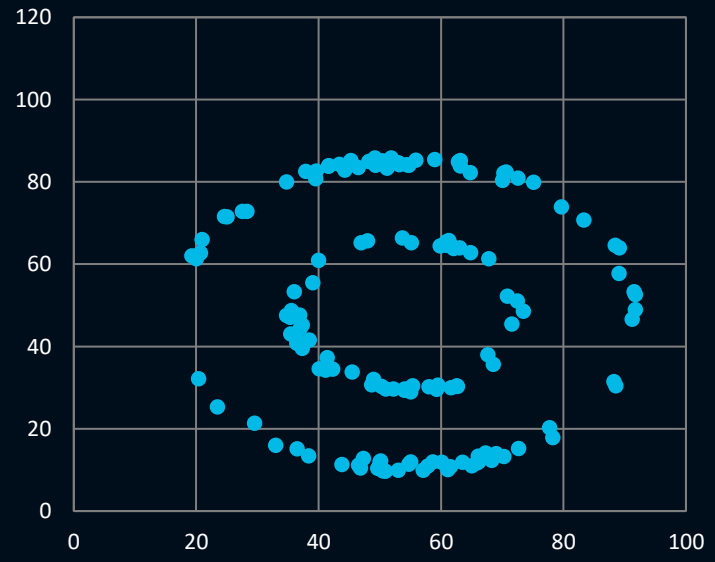
Never trust summary statistics alone; visualize your data!



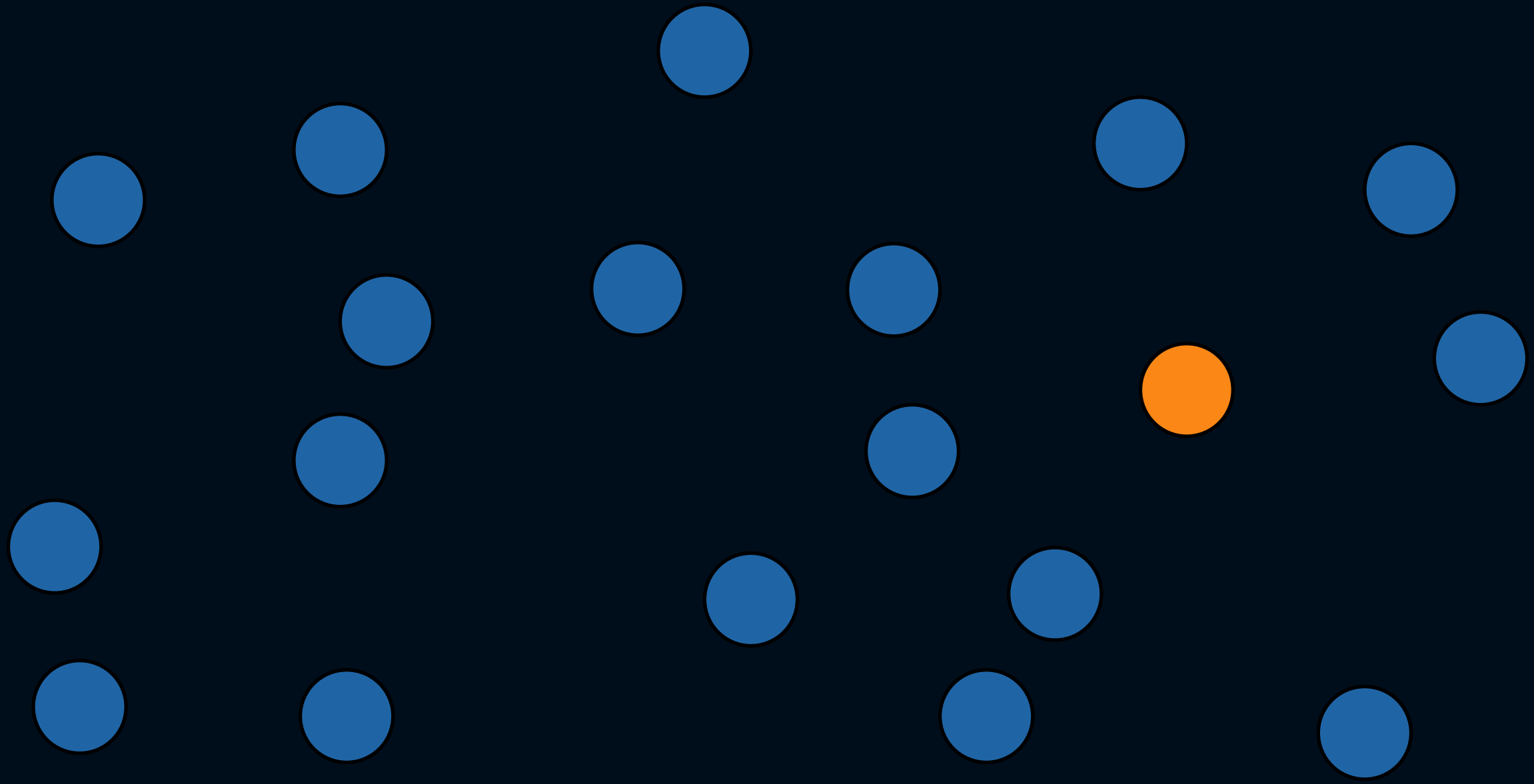


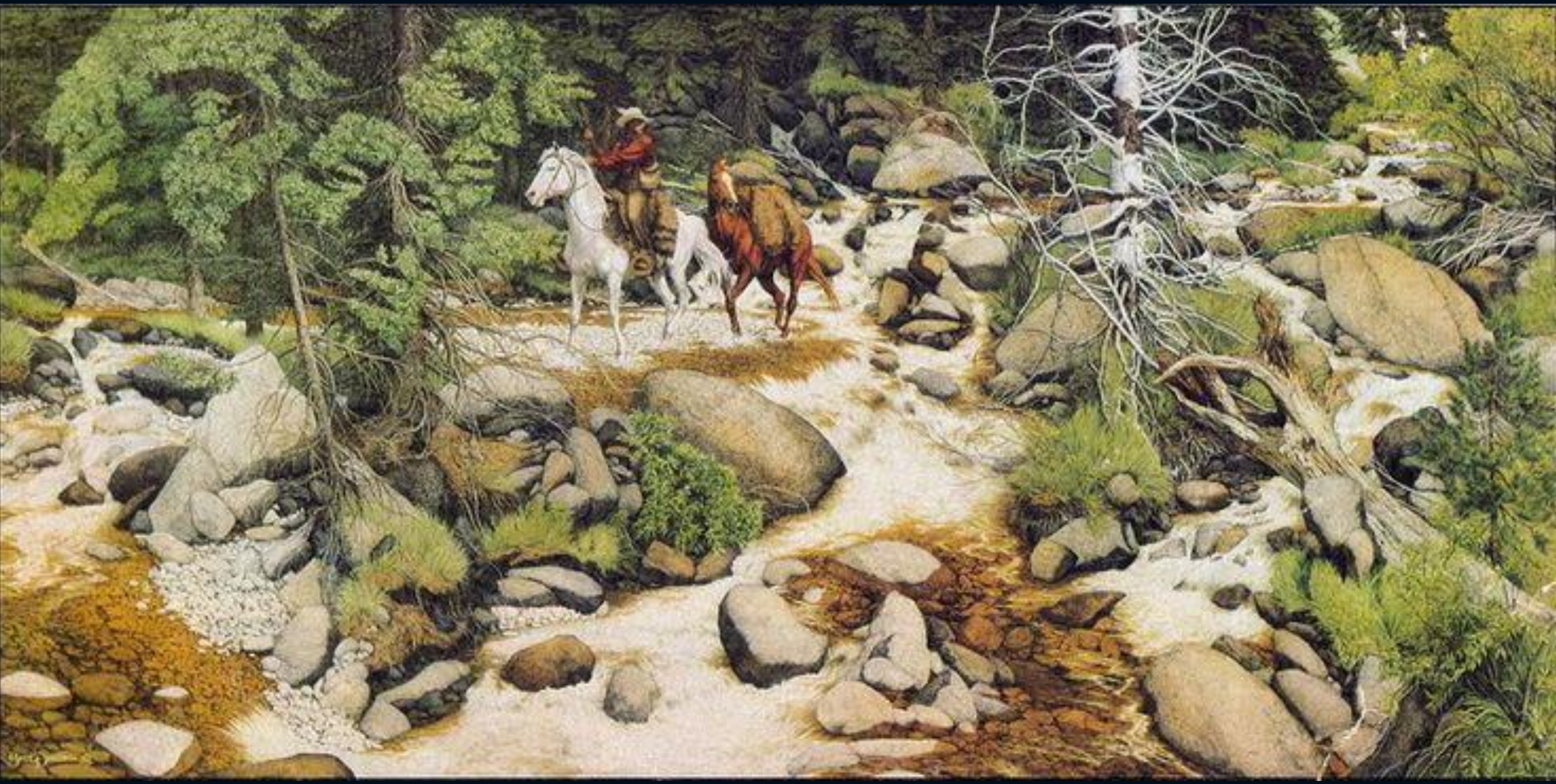
X Mean:	54.26
Y Mean:	47.83
X SD:	16.76
Y SD:	26.93
Corr.:	-0.06





The Datasaurus Dozen [autodesk.com]





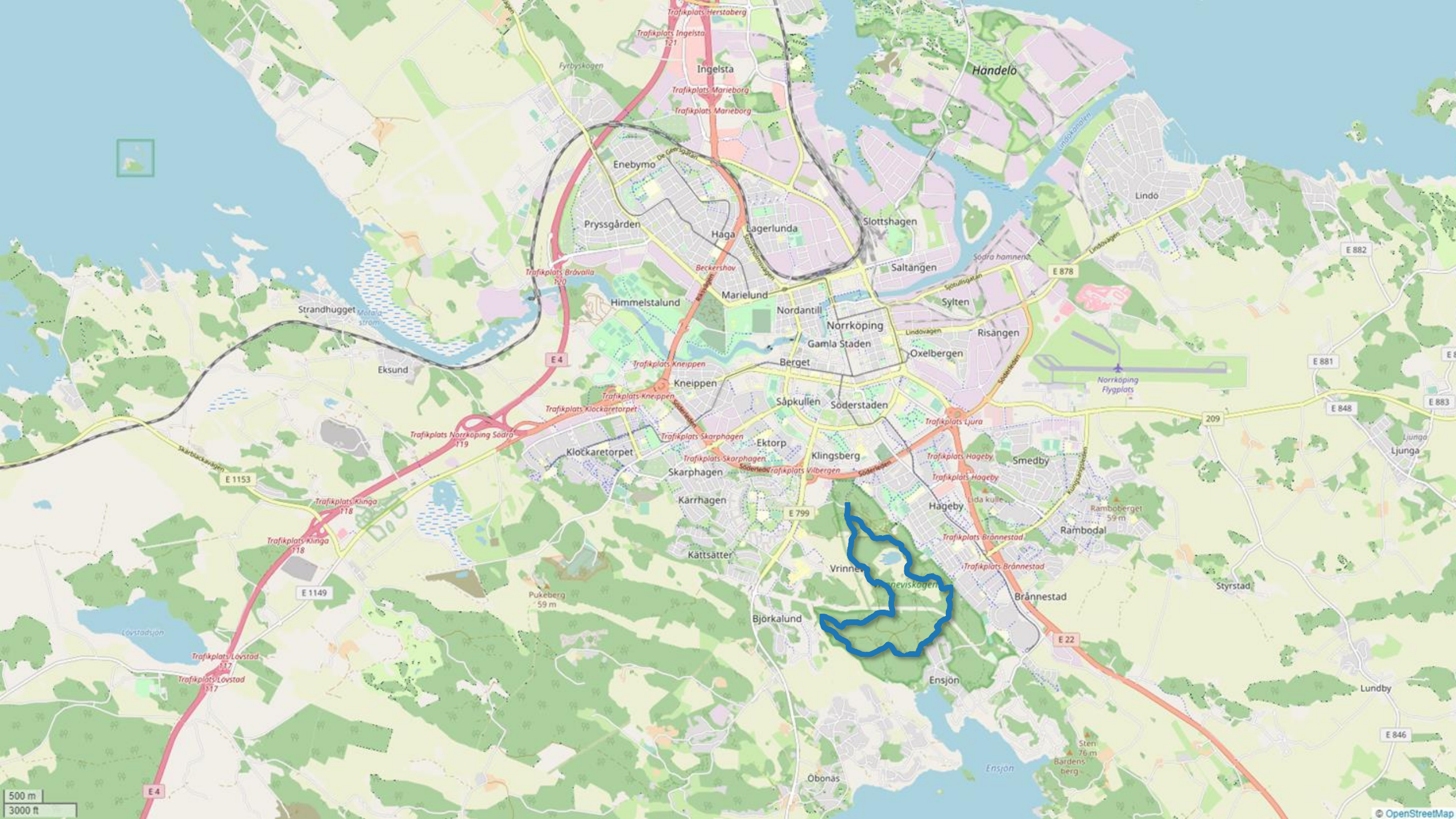
”The Forest Has Eyes” by Bev Doolittle

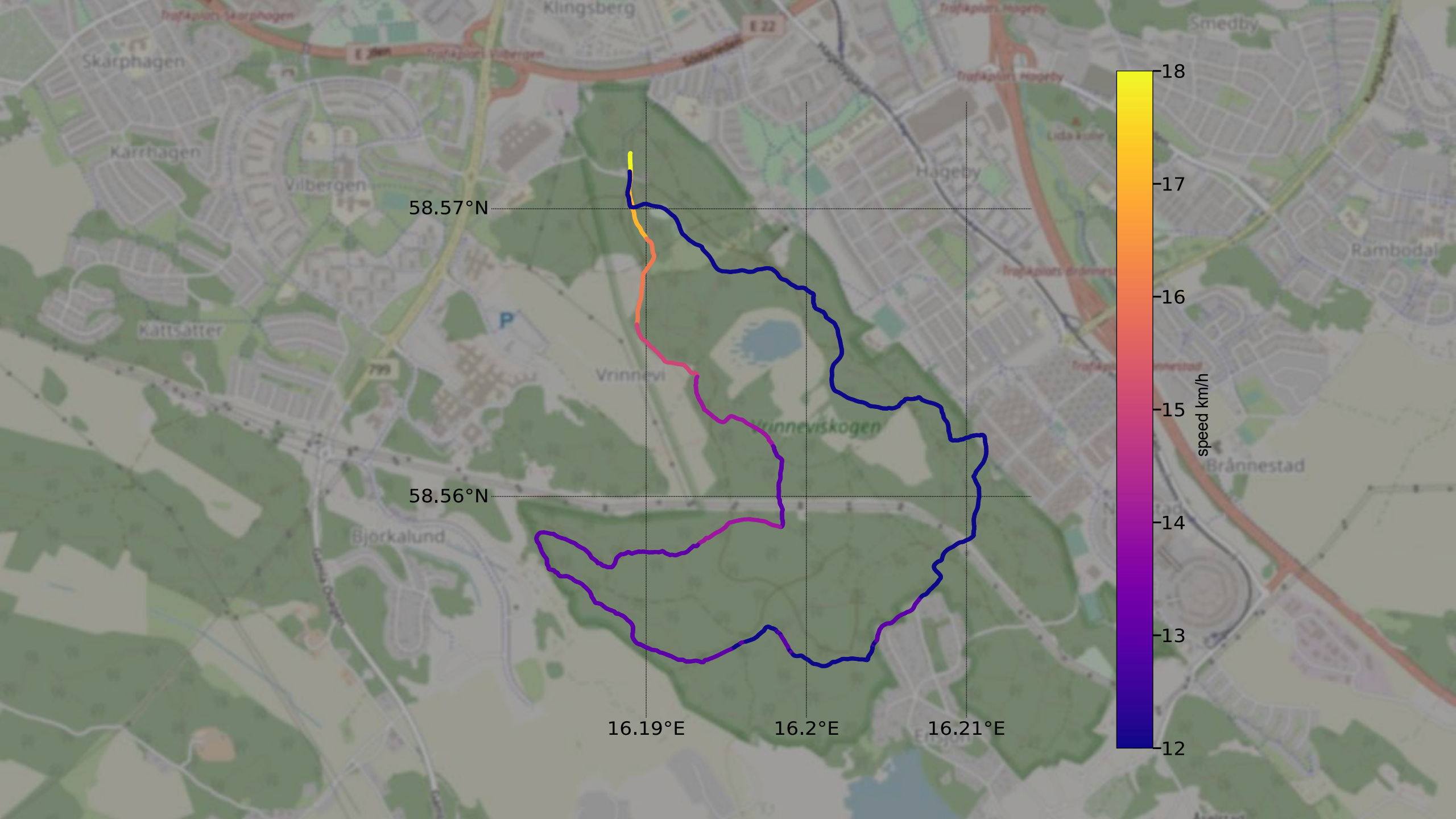
Example: fitness tracker

Type	Local Number	Message	Field 1	Value 1	Units 1	Field 2	Value 2	Units 2	Field 3	Value 3	Units 3	Field 4	Value 4	Units 4	Field 5	Value 5	Units 5	Field 6	Value 6	Units 6	Field 7	Value 7	Units 7	Field 8
Definition	0	file_id	serial_number	1		time_created	1		unknown	1		manufacturer	1		product	1		number	1		type	1		
Data	0	file_id	serial_number	3380794501		time_created	1026748965		manufacturer	1		garmin_product	3122		type	4								
Definition	1	file_creator	unknown	20		software_version	1		hardware_version	1														
Data	1	file_creator	software_version	940																				
Definition	2	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1					
Data	2	unknown	unknown	1026748965		unknown	1																	
Definition	3	event	timestamp	1		data	1		unknown	1		unknown	1		event	1		event_type	1		event_group	1		
Data	3	event	timestamp	1026748965	s	timer_trigger	0		event	0		event_type	0		event_group	0								
Definition	4	device_info	timestamp	1		serial_number	1		cum_operating_time	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown
Data	4	device_info	timestamp	1026748965	s	serial_number	3380794501		manufacturer	1		garmin_product	3122		software_version	94		device_index	0		source_type	5		
Data	4	device_info	timestamp	1026748965	s	manufacturer	1		garmin_product	3122		software_version	94		device_index	1		device_type	4		source_type	5		
Data	4	device_info	timestamp	1026748965	s	serial_number	3338514105		unknown	2		unknown	3514347193		manufacturer	1		garmin_product	255		software_version	254		device_index
Data	4	device_info	timestamp	1026748965	s	serial_number	3380638747		unknown	1		unknown	24867867		manufacturer	1		garmin_product	3192		software_version	40		battery_voltage
Definition	5	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown
Data	5	unknown	unknown	1026748965		unknown	3		unknown	3		unknown	1		unknown	2		unknown	3					
Definition	6	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		
Data	6	unknown	unknown	1026748965		unknown	1025740800		unknown	1026345600		unknown	1											
Definition	7	device_settings	utc_offset	1		time_offset	1		unknown	1		autosync_min_time	1		active_time_zone	1		unknown	1		time_mode	1		time_zone_offset
Data	7	device_settings	utc_offset	0		time_offset	120	s	active_time_zone	0		unknown	0		time_mode	1		time_zone_offset	0	hr	unknown	0		unknown
Definition	9	unknown	unknown	1		unknown	16		unknown	1		unknown	4		unknown	12		unknown	4		unknown	4		unknown
Data	9	unknown	unknown	24867867		unknown	29723		unknown	0		unknown	2270		unknown	3192		unknown	400		unknown	0		unknown
Data	9	unknown	unknown	3514347193		unknown	896697		unknown	1		unknown	255		unknown	2540		unknown	1		unknown	0		unknown
Definition	10	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown
Data	10	unknown	unknown	1026748965		unknown	3917349947		unknown	0		unknown	0		unknown	0		unknown	0		unknown	740		unknown
Definition	11	sport	name	32		unknown	1		sport	1		sub_sport	1		unknown	1		unknown	1		unknown	1		unknown
Data	11	sport	name	MOUNTAIN		unknown	45		sport	2		sub_sport	8		unknown	1		unknown	0		unknown	0		unknown
Definition	12	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown	1		unknown
Data	12	unknown	unknown	500000		unknown	1200000		unknown	18000		unknown	0		unknown	4167		unknown	1389		unknown	60		unknown
Definition	13	zones_target	unknown	1		functional_threshold_power	1		unknown	1		max_heart_rate	1		threshold_heart_rate	1		hr_calc_type	1		pwr_calc_type	1		unknown
Data	13	zones_target	unknown	2		functional_threshold_power	210		threshold_heart_rate	0		hr_calc_type	1		pwr_calc_type	1		unknown	1		unknown	1		unknown
Definition	14	record	timestamp	1		distance	1		altitude	1		speed	1		unknown	1		unknown	1		heart_rate	1		temperature
Data	14	record	timestamp	1026748965	s	distance	4.78	m	altitude	4336	m	speed	4787	m/s	unknown	4574		unknown	659		heart_rate	103	bpm	temperature
Definition	15	unknown	unknown	1		unknown	8		unknown	1		unknown	1											
Data	15	unknown	unknown	1026748968		unknown	121|2|15|0|108|36|18|0		unknown	3														
Data	5	unknown	unknown	1026748970		unknown	3		unknown	3		unknown	1		unknown	2		unknown	3					
Definition	0	record	timestamp	1		distance	1		altitude	1		speed	1		unknown	1		unknown	1		unknown	1		heart_rate
Data	0	record	timestamp	1026748971	s	distance	34.77	m	altitude	4264	m	speed	5067	m/s	unknown	4578		unknown	659		unknown	2598		heart_rate
Data	0	record	timestamp	1026748972	s	distance	39.84	m	altitude	4264	m	speed	5067	m/s	unknown	4578		unknown	659		unknown	2598		heart_rate
Data	0	record	timestamp	1026748973	s	distance	44.92	m	altitude	4266	m	speed	5085	m/s	unknown	4579		unknown	659		unknown	2598		heart_rate
Data	0	record	timestamp	1026748974	s	distance	49.96	m	altitude	4,268E+16	m	speed	5039	m/s	unknown	4580		unknown	659		unknown	2598		heart_rate
Data	0	record	timestamp	1026748978	s	distance	70.2	m	altitude	4270	m	speed	5104	m/s	unknown	4581		unknown	659		unknown	2303		heart_rate
Data	0	record	timestamp	1026748980	s	distance	80.21	m	altitude	4,272E+16	m	speed	4983	m/s	unknown	4582		unknown	659		unknown	2244		heart_rate
Data	0	record	timestamp	1026748983	s	distance	95.19	m	altitude	4276	m	speed	5067	m/s	unknown	4584		unknown	659		unknown	2244		heart_rate
Data	0	record	timestamp	1026748984	s	distance	100.18	m	altitude	4,278E+16	m	speed	4992	m/s	unknown	4585		unknown	659		unknown	2244		heart_rate
Data	0	record	timestamp	1026748988	s	distance	119.74	m	altitude	4,288E+16	m	speed	4927	m/s	unknown	4590		unknown	659		unknown	2165		heart_rate
Data	4	device_info	timestamp	1026749007	s	manufacturer	1		garmin_product	3107		software_version	48		device_index	4		device_type	0		source_type	5		
Data	5	unknown	unknown	1026749007		unknown	3		unknown	3		unknown	1		unknown	2		unknown	3					
Definition	1	record	timestamp	1		position_lat	1		position_long	1		distance	1		grit	1		flow	1		altitude	1		speed
Data	1	record	timestamp	1026749007	s	position_lat	582423992	semicircles	position_long	118853500	semicircles	distance	13991	m	grit	0		flow	0		altitude	4296	m	speed
Definition	2	unknown	unknown	1		unknown	1		unknown	1		unknown	1		unknown	1								

GPS track

Extracted from lat/long columns


Ingelsta
Handelö
Enebymo
Pryssgården
Haga
Lagerlunda
Slottshagen
Lindö
Saltängen
Himmelstalund
Marielund
Nordantill
Norrköping
Sylten
Risängen
Gamla Staden
Oxelbergen
Berget
Strandhugget
Eksund
Kneippen
Söderstaden
Norrköping Flygplats
Klockaretorpet
Ektorp
Klingsberg
Smedby
Skarphagen
Kärrhagen
Hageby
Rambodal
Kättsätter
Vrinnevi
Brånnestad
Styrstad
Björkalund
Ensjön
Lundby
Obonas
Ljunga
Pukeberg
59 m
Rambobergen
59 m
Sten
76 m
Bardens
berg
E 4
E 1153
E 1149
E 799
E 22
E 878
E 882
E 881
E 848
E 883
E 846
209
Trafikplats Herstaberg
Trafikplats Ingelsta
121
Trafikplats Marieborg
Trafikplats Brovalla
120
Trafikplats Kneippen
Trafikplats Klockaretorpet
Trafikplats Norrköping Södra
119
Trafikplats Skarphagen
Trafikplats Vilbergen
Trafikplats Ljura
Trafikplats Hageby
Trafikplats Klinga
118
Trafikplats Lövstad
117
Trafikplats Brånnestad
Fyrbyskogen
Motala ström
Lindöviken
Lindövägen
Södra hamnen
Söderleden
Kungsängsleden
Skärblackavägen
Lövstadsjön
Ensjön
500 m
3000 ft
© OpenStreetMap


58.57°N
58.56°N
16.19°E
16.2°E
16.21°E
18
17
16
15
14
13
12
speed km/h
Skärphagen
Karrhagen
Vilbergen
Klingsberg
Kattsatter
Vrinnevi
Vrinneviskogen
Björkalund
Hageby
Smedby
Rambodal
Brånnestad
E 22
795

What data do you have?

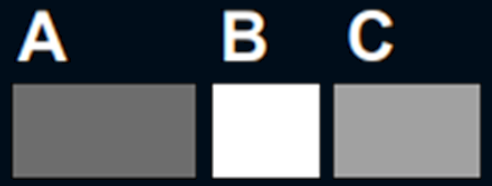
What are the types of your data?

Categories		Ordered Categories		Continuous Metrics		
City	Airline	Class	PriceBracket	Distance	FlightTime	Price
Linköping	KLM	First	$$$	300	90	1600
Nyköping	Ryan Air	Economy	$	500	120	250
...	...	...	...	...	...	...

What is your goal?

Categories		Ordered Categories		Continuous Metrics		
City	Airline	Class	PriceBracket	Distance	FlightTime	Price
Linköping	KLM	First	$$$	300	90	1600
Nyköping	Ryan Air	Economy	$	500	120	250
…	…	…	…	…	…	…

or



?

Types of visualizations

Charts (line chart, bar chart, …)

Scatter plots

Heatmaps

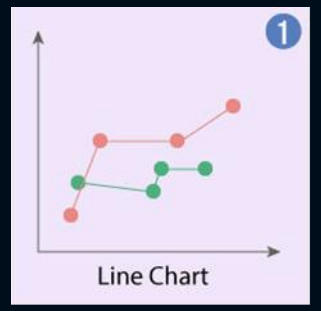
Graphs and diagrams

Glyphs and shapes

Line rendering

Spatial visualization

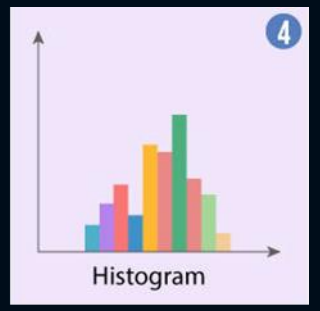
…



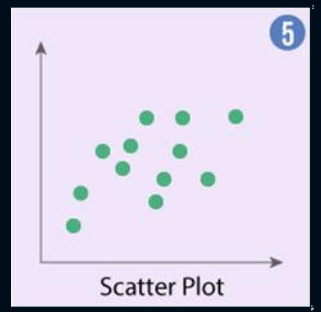
Display trends over time

A line chart with areas below the lines filled with colors

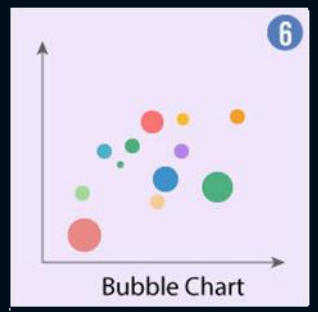
Display trends with multiple variables



Display the shape and spread of continuous dataset samples



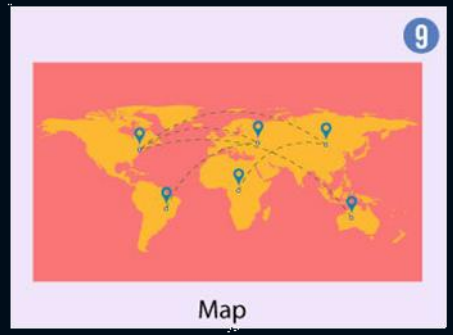
Show correlation in a dataset



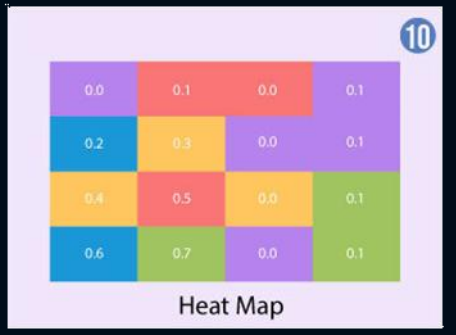
Show and compare the relations between the labelled circles

Show the contribution of data point inside a whole dataset

Visualize the distance between intervals



Show data with location as variable



Show magnitude of a phenomenon

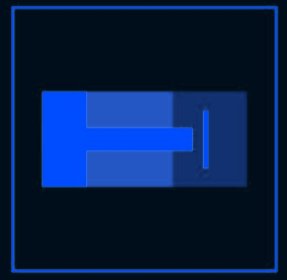
Comparison

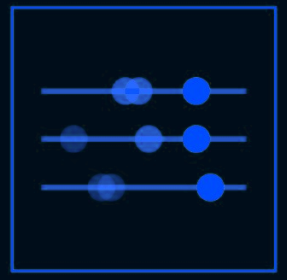

BAR CHART

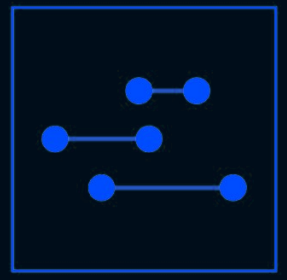

COLUMN CHART


GROUPED BAR/COLUMN CHART


LOLLIPOP CHART


BULLET CHART


DOT PLOT

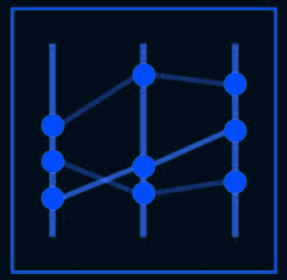

DUMBBELL


PICTOGRAM


ICON CHART


RANGE CHART

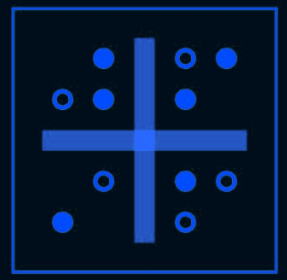

RADIAL BAR CHART


PARALLEL COORDINATES


RADAR CHART

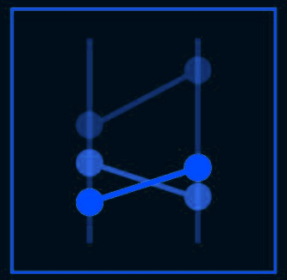

NIGHTINGALE CHART


WATERFALL CHART


MATRIX CHART


SMALL MULTIPLES

WORD CLOUD


SLOPE CHART

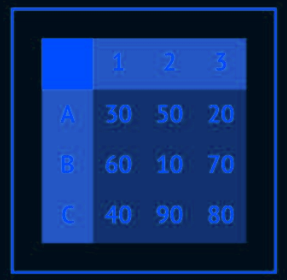
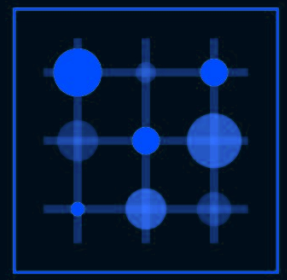
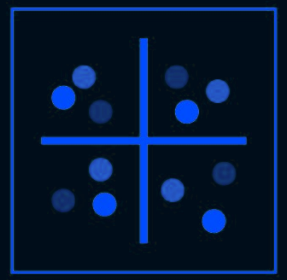


TABLE CHART


CATEGORICAL SCATTER PLOT


QUADRANT CHART

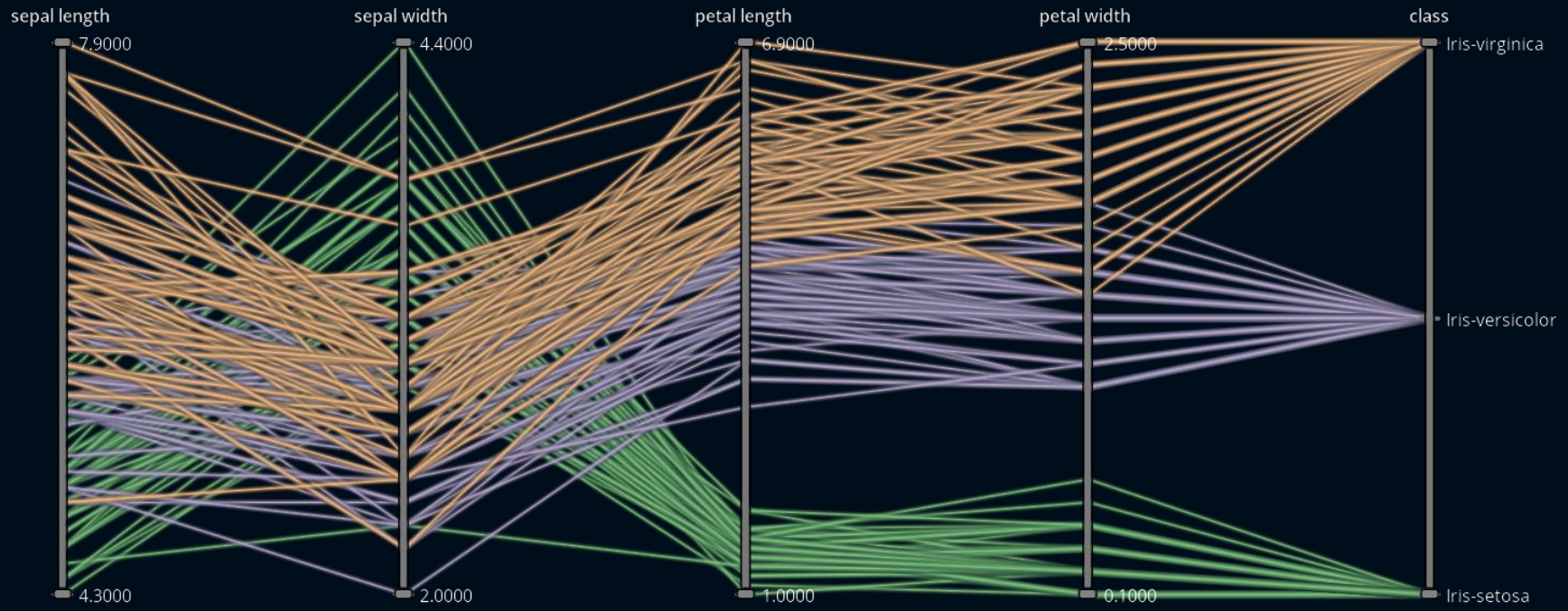
Small multiples

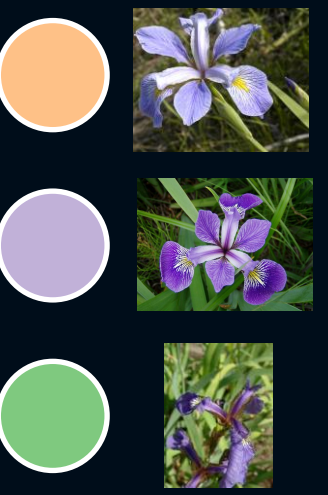
Avoids cluttering in a single plot

[insights.datylon.com/stories]

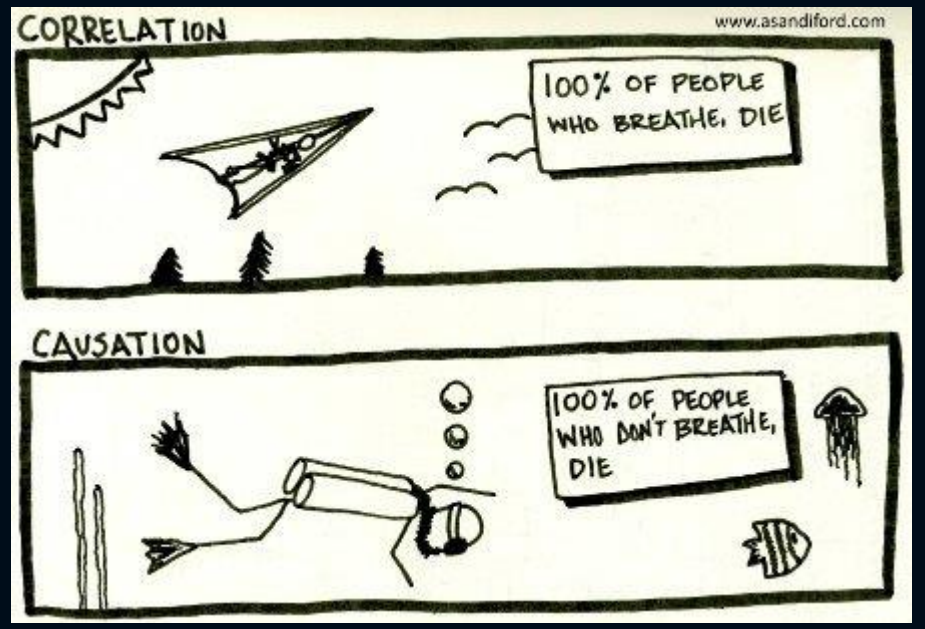
Parallel coordinates

Allows for comparison of many attributes





Correlation



Can you buy happiness for money ?

Exploring a link between countries' happiness score and GDP per capita value.

Source: World Happiness Report 2021, The World Bank

Part-to-whole and hierarchical


STACKED BAR/COLUMN CHART


DIVERGING BAR CHART


POPULATION PYRAMID

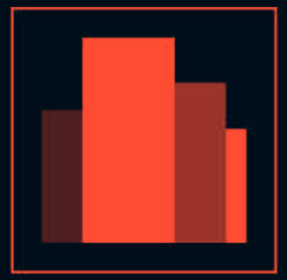

ICON ARRAY


WAFFLE CHART

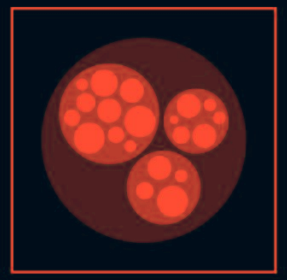

PIE CHART


DONUT CHART


SEMI-CIRCLE DONUT CHART


MARIMEKKO CHART

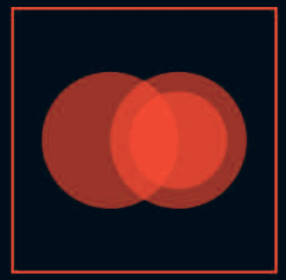

TREEMAP


CIRCULAR TREEMAP


CONVEX TREEMAP


DENDROGRAM


VENN DIAGRAM

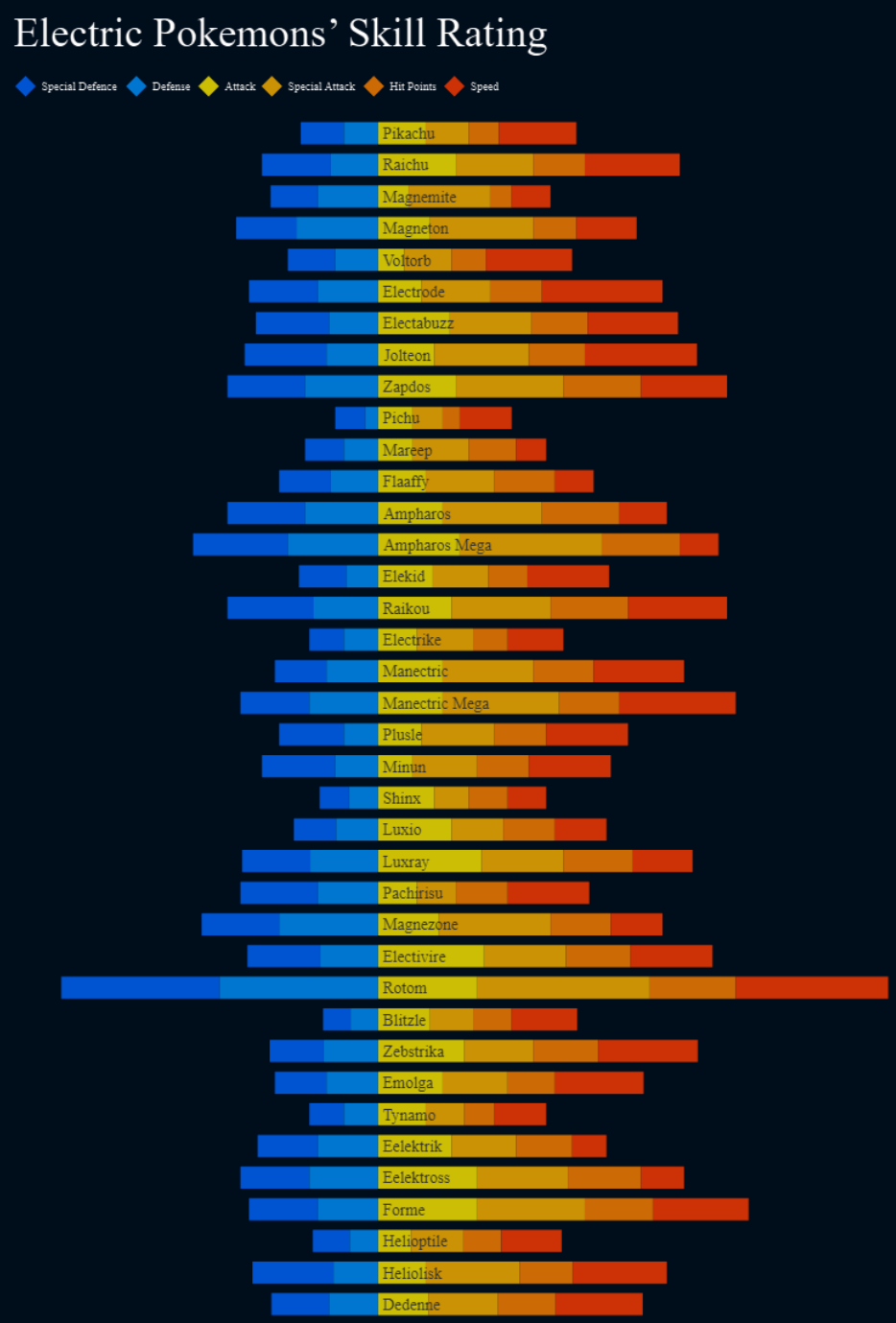

EULER DIAGRAM


CIRCULAR GAUGE


SUNBURST CHART


FUNNEL & PYRAMID CHART

Diverging stacked bar chart

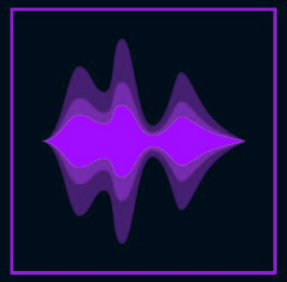


Data source: https://www.kaggle.com/abcsds/pokemon

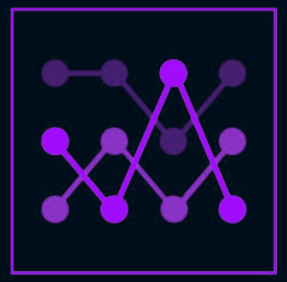
Data over time (temporal)

AREA CHART

STACKED AREA CHART



STREAM GRAPH

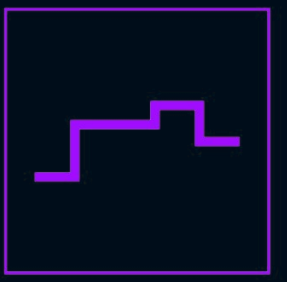


BUMP CHART

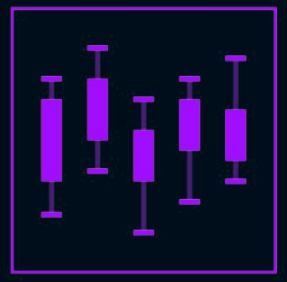
BUMP AREA CHART

LINE CHART

SPLINE CHART



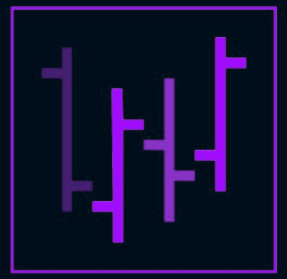
STEP LINE CHART



CANDLESTICK CHART

GANTT CHART

BARCODE CHART

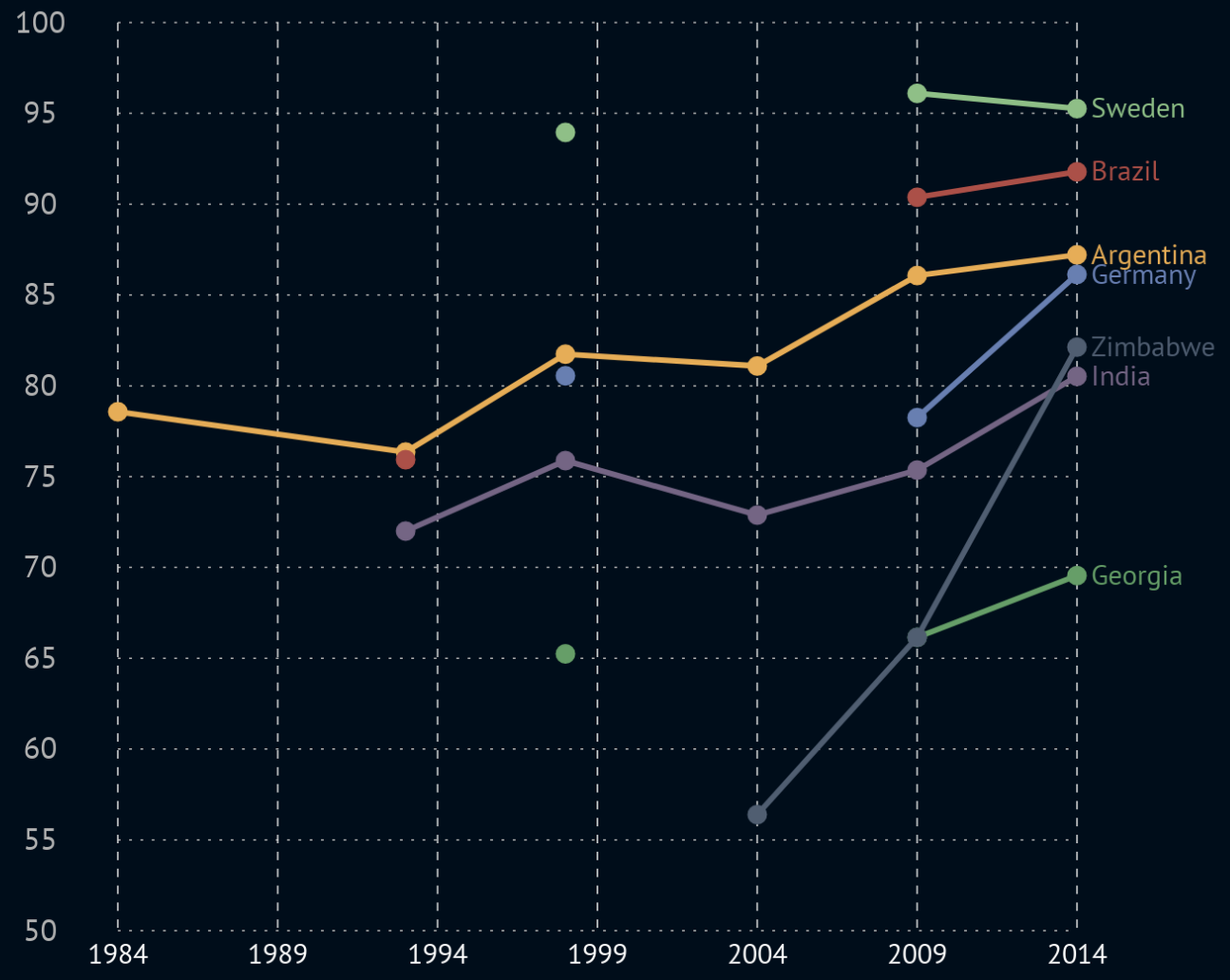


OHLC CHART

Line chart

Share of people who say they are happy

Share of people who say they are “very happy” or “rather happy”.

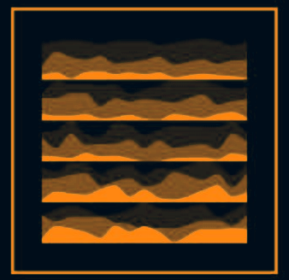


Source: World Value Survey (2014)

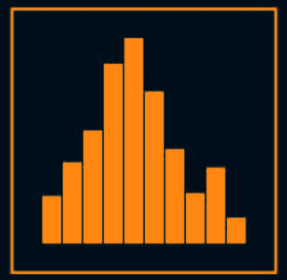
Distributions

DENSITY PLOT

RIDGELINE PLOT

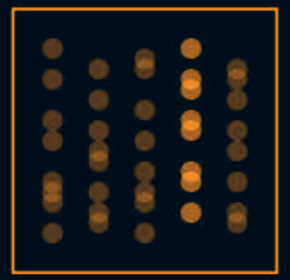


HORIZON CHART



HISTOGRAM

RADIAL HISTOGRAM

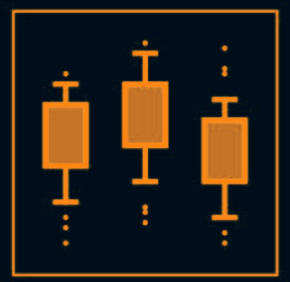


STRIP PLOT

JITTER PLOT

ONE-DIMENSIONAL HEATMAP

BEESWARM CHART

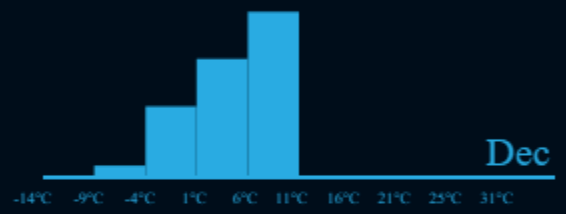


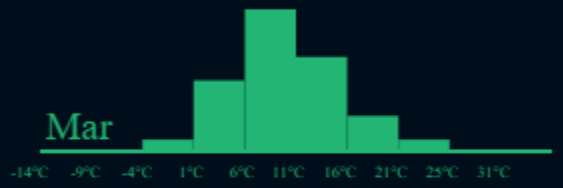
BOX CHART

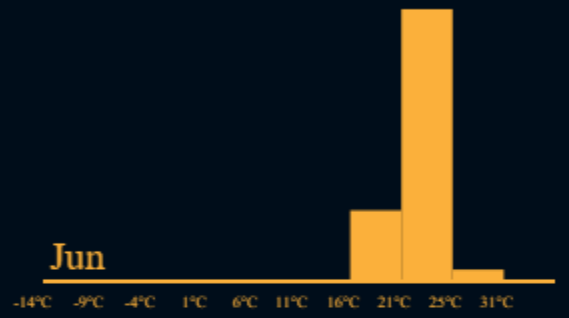
VIOLIN PLOT

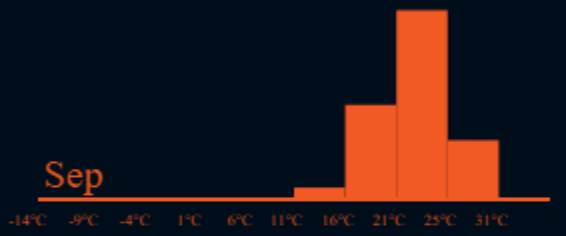
Seasons in New York City

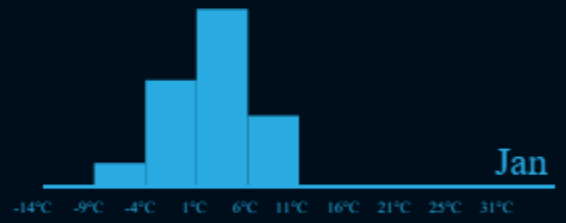
Showing distribution of days with a certain temperature range

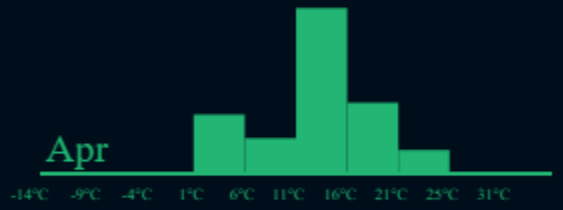


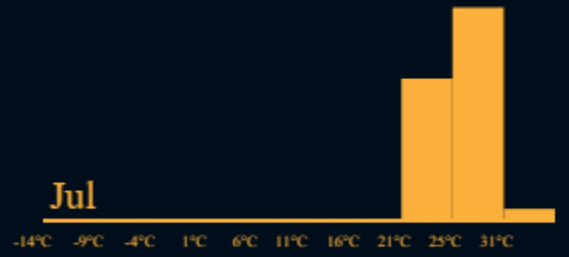


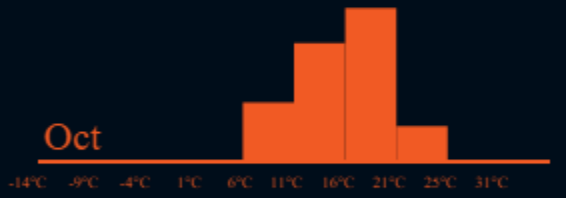


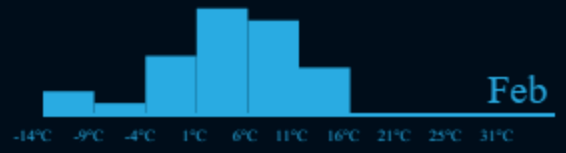


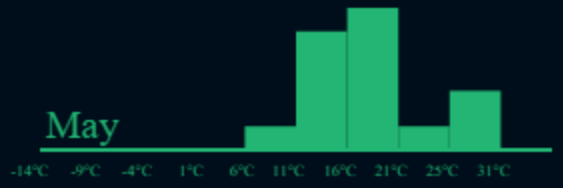


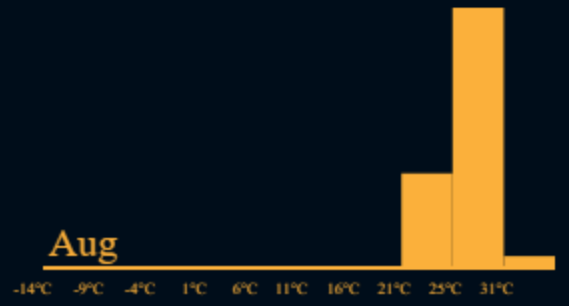


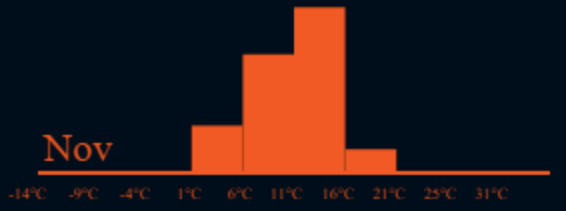












Geospatial and connectivity

GEOGRAPHIC HEATMAP

CHOROPLETH MAP

TILE MAP

CHORD DIAGRAM

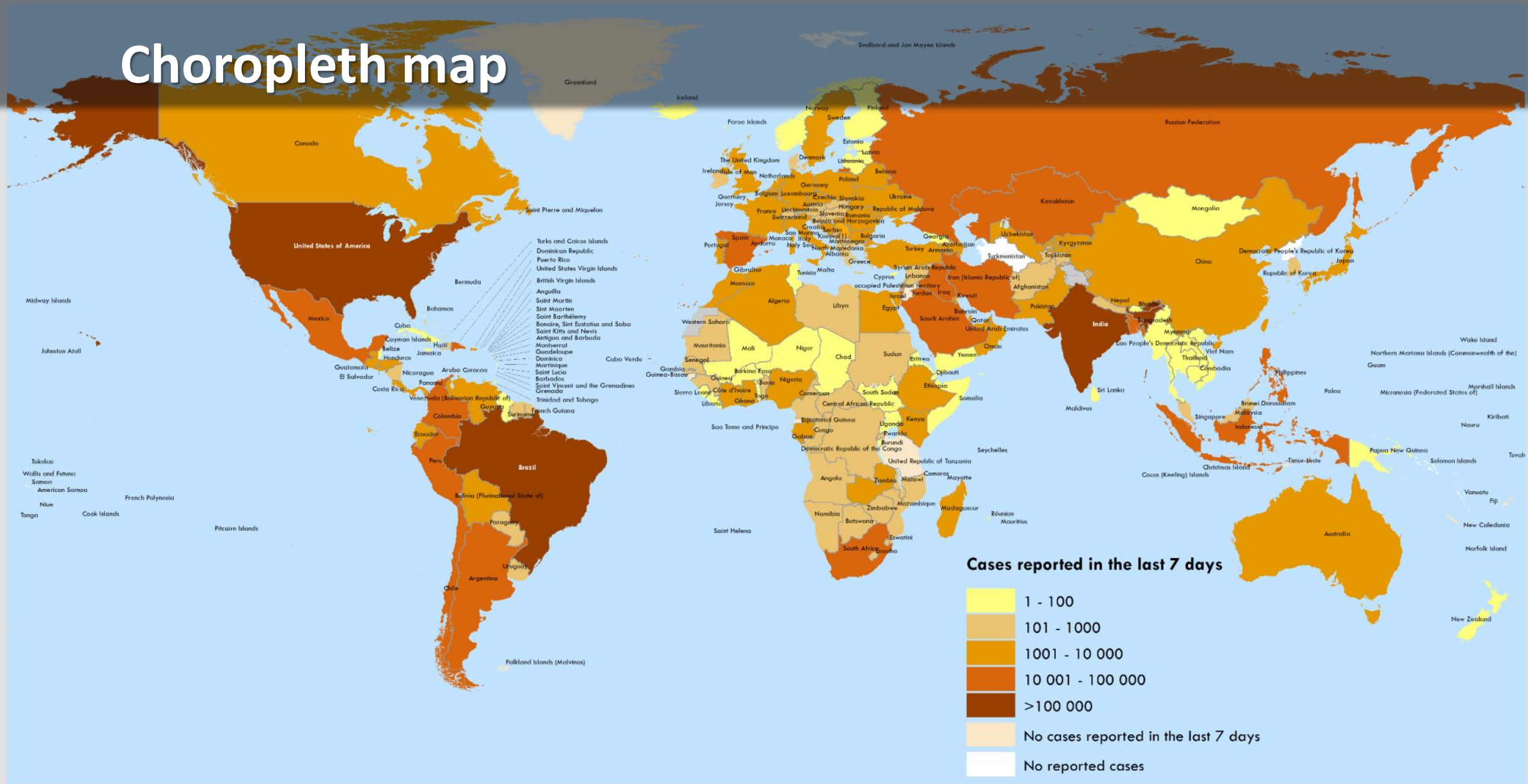
ARC DIAGRAM

SANKEY

NETWORK DIAGRAM

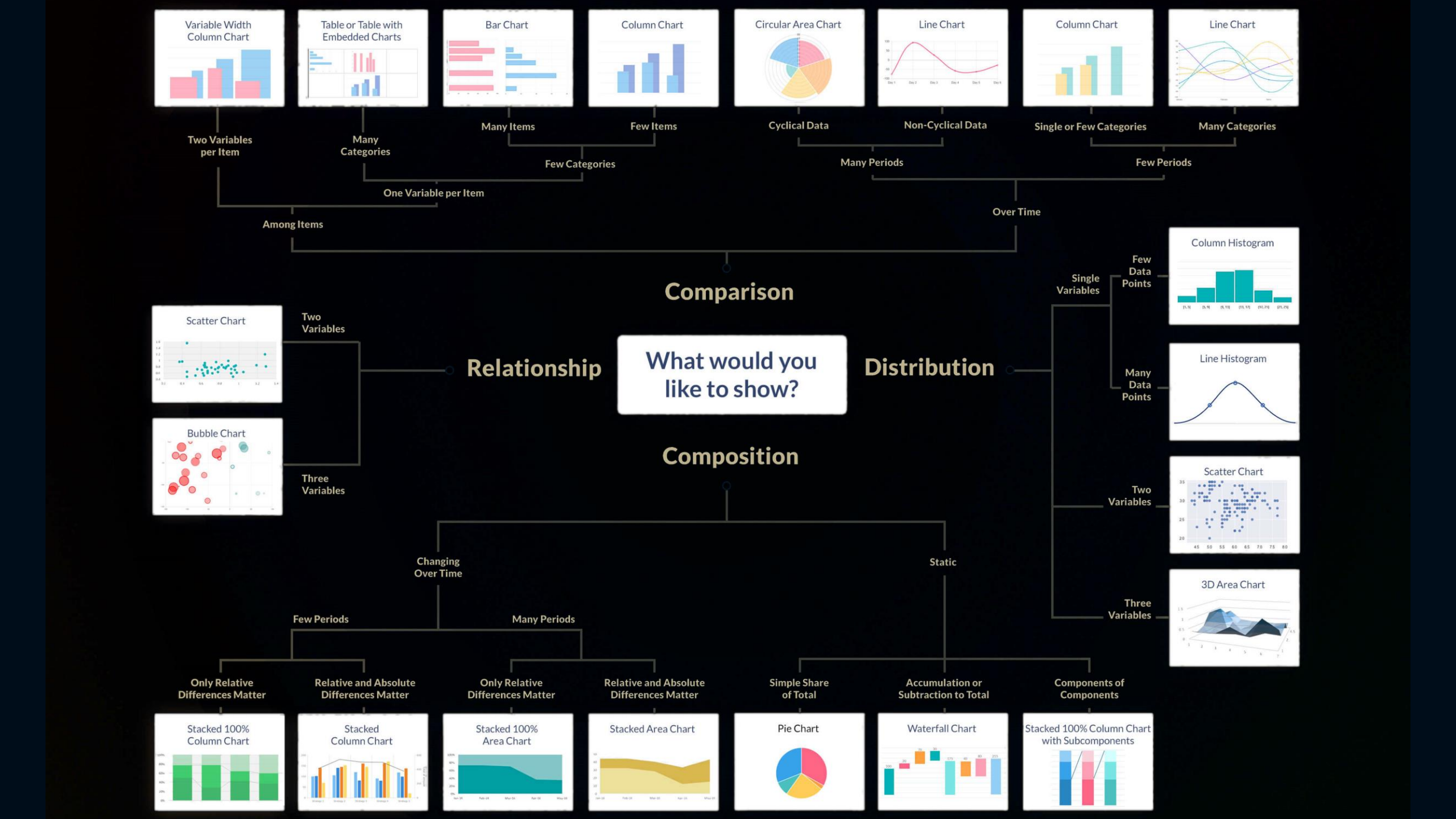
FLOWCHART

Choropleth map



Data Source: *World Health Organization*
Map Production: *WHO Health Emergencies Programme*

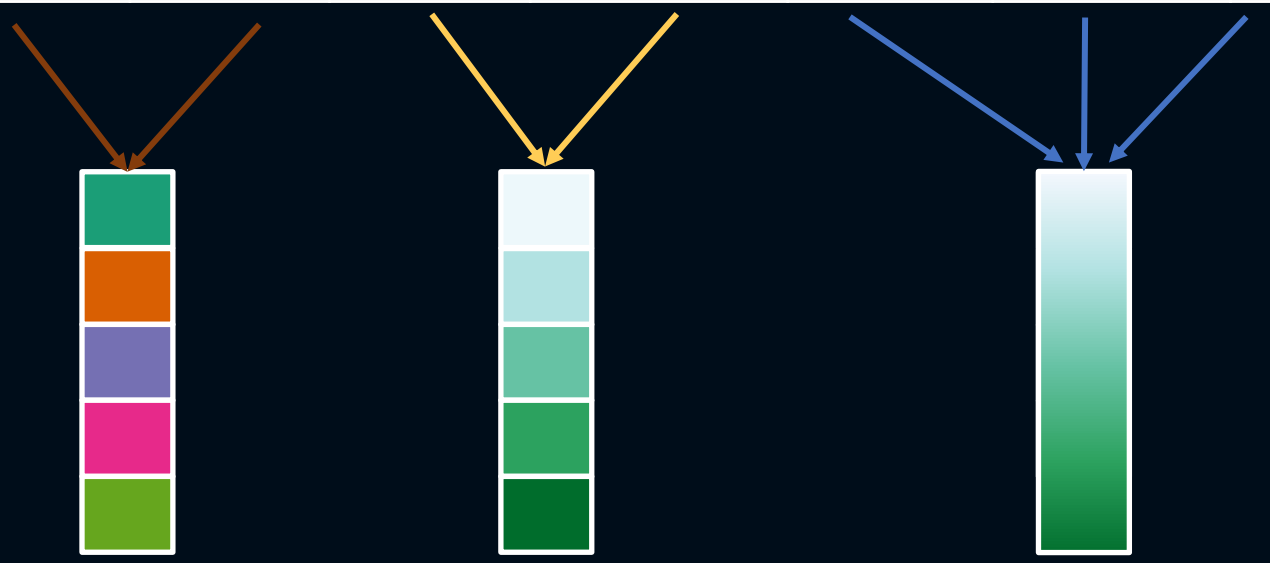
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.


What would you like to show?
Comparison
Relationship
Distribution
Composition
Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Few Categories
Many Periods
Few Periods
One Variable per Item
Over Time
Among Items
Scatter Chart
Bubble Chart
Two Variables
Three Variables
Single Variables
Few Data Points
Many Data Points
Column Histogram
Line Histogram
Two Variables
Scatter Chart
Three Variables
3D Area Chart
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Stacked 100% Column Chart
Stacked Column Chart
Stacked 100% Area Chart
Stacked Area Chart
Pie Chart
Waterfall Chart
Stacked 100% Column Chart with Subcomponents

What do you want to tell with your data?

Categories		Ordered Categories		Continuous Metrics		
City	Airline	Class	PriceBracket	Distance	FlightTime	Price
Linköping	KLM	First	$$$	300	90	1600
Nyköping	Ryan Air	Economy	$	500	120	250
…	…	…	…	…	…	…

Color?



Data encoding

Ordered attributes

1. Position
2. Length
3. Orientation
4. Area
5. Color
   - Color luminance
   - Color saturation
6. Curvature
6. Volume

Same

Same

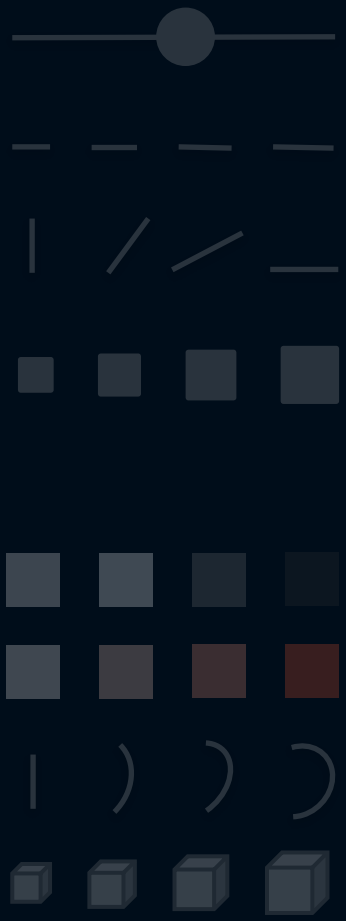
Data encoding

Ordered attributes

1. Position
2. Length
3. Orientation
4. Area
5. Color
   - Color luminance
   - Color saturation
6. Curvature
6. Volume

Same

Same



Best

Effectiveness

Least

Categorical attributes

1. Spatial region
2. Color hue
3. Motion
4. Shape

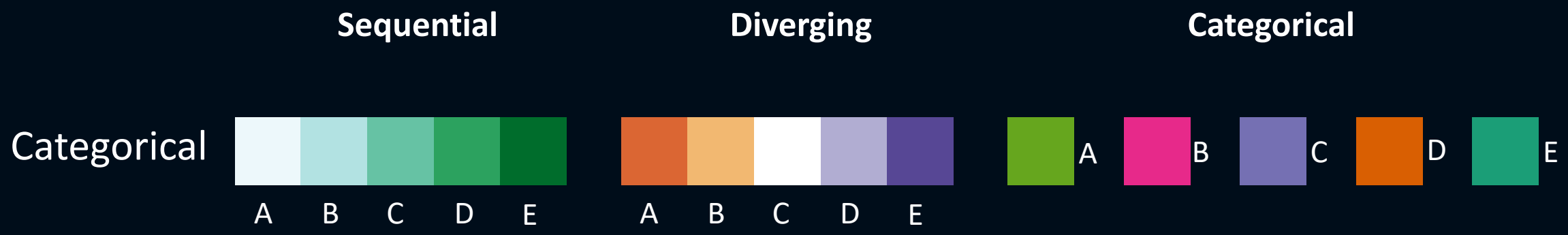
Colormaps

Sequential

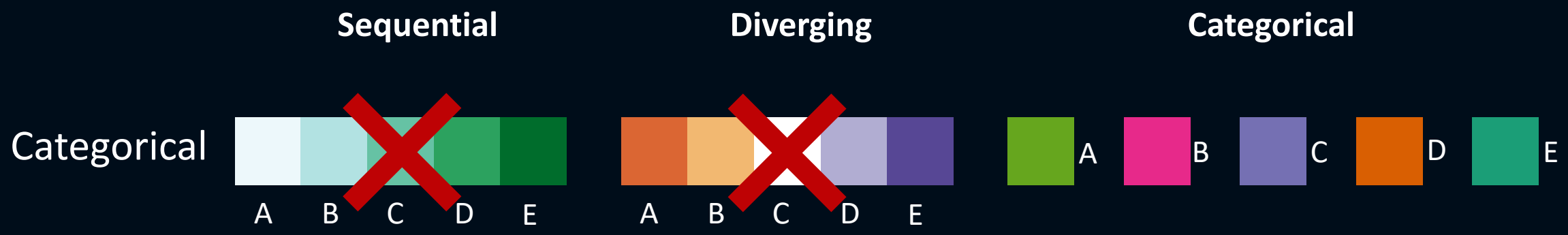
Diverging

Categorical

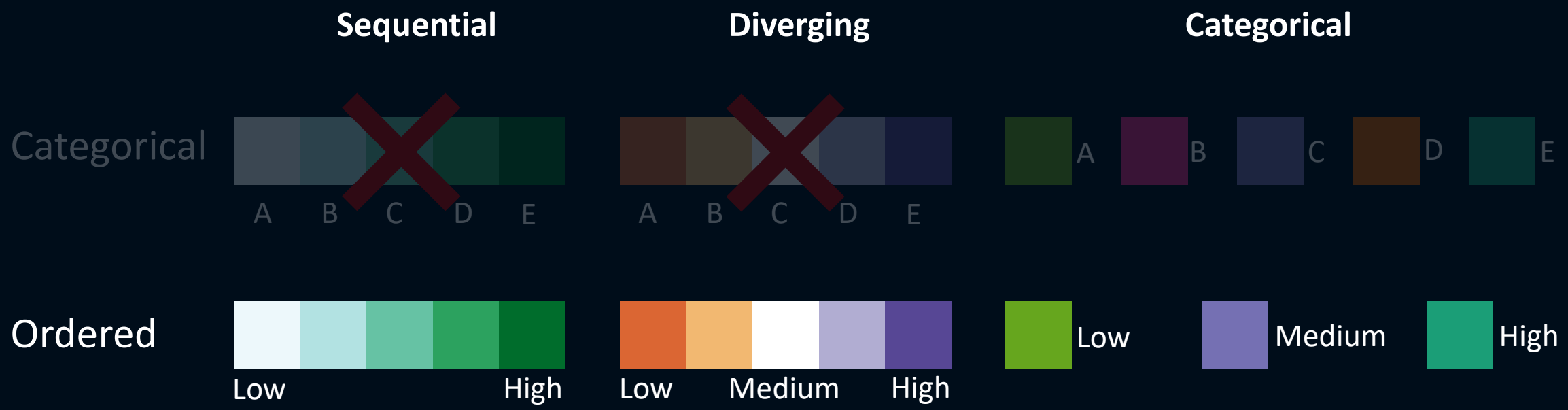
Colormaps


Sequential
Diverging
Categorical
Categorical
A
B
C
D
E
A
B
C
D
E
A
B
C
D
E

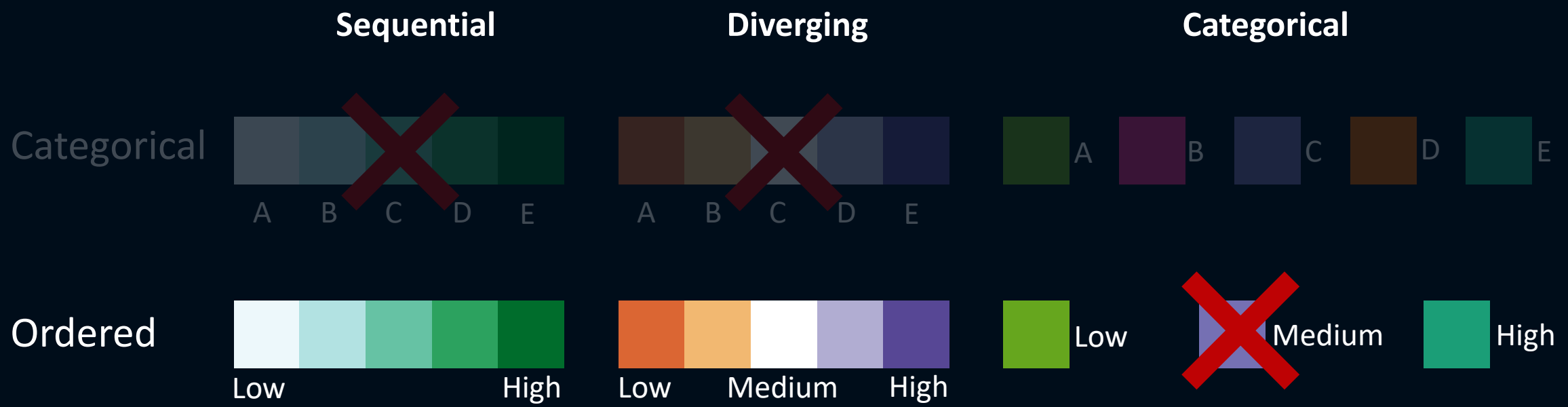
Colormaps


Sequential
Diverging
Categorical
Categorical
A
B
C
D
E
A
B
C
D
E
A
B
C
D
E

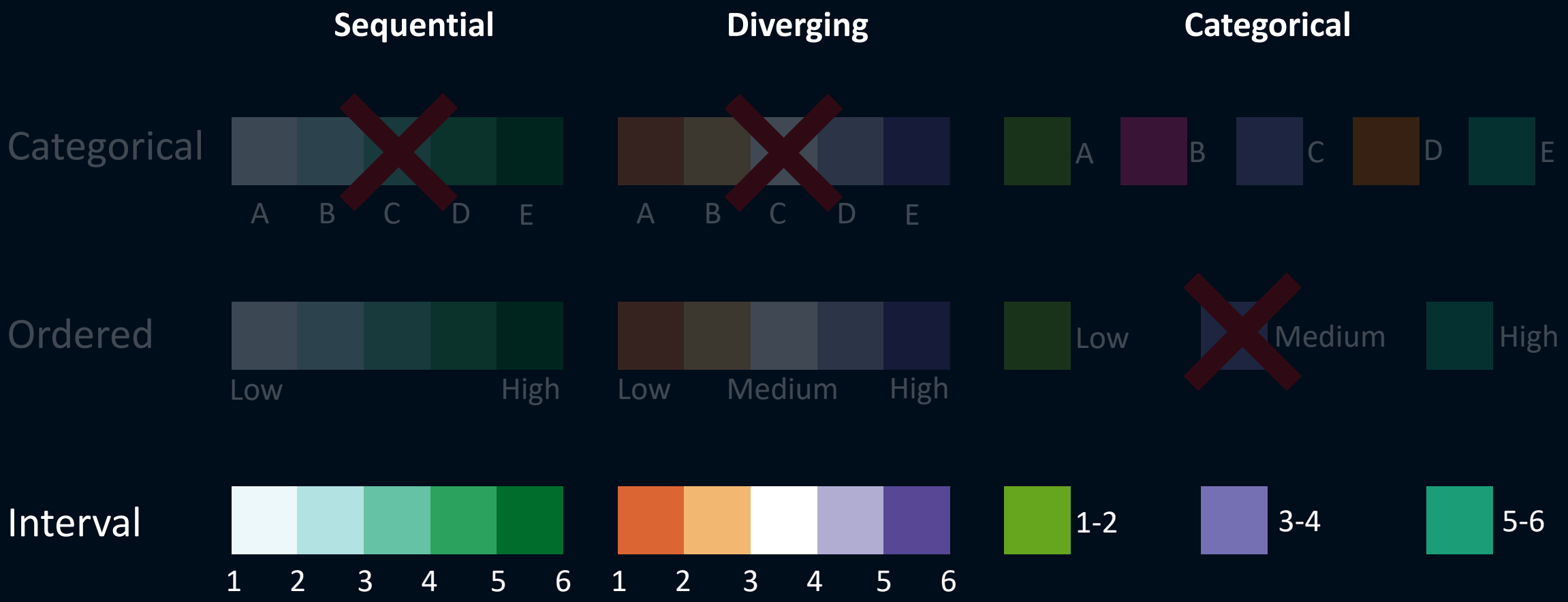
Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low
Medium
High

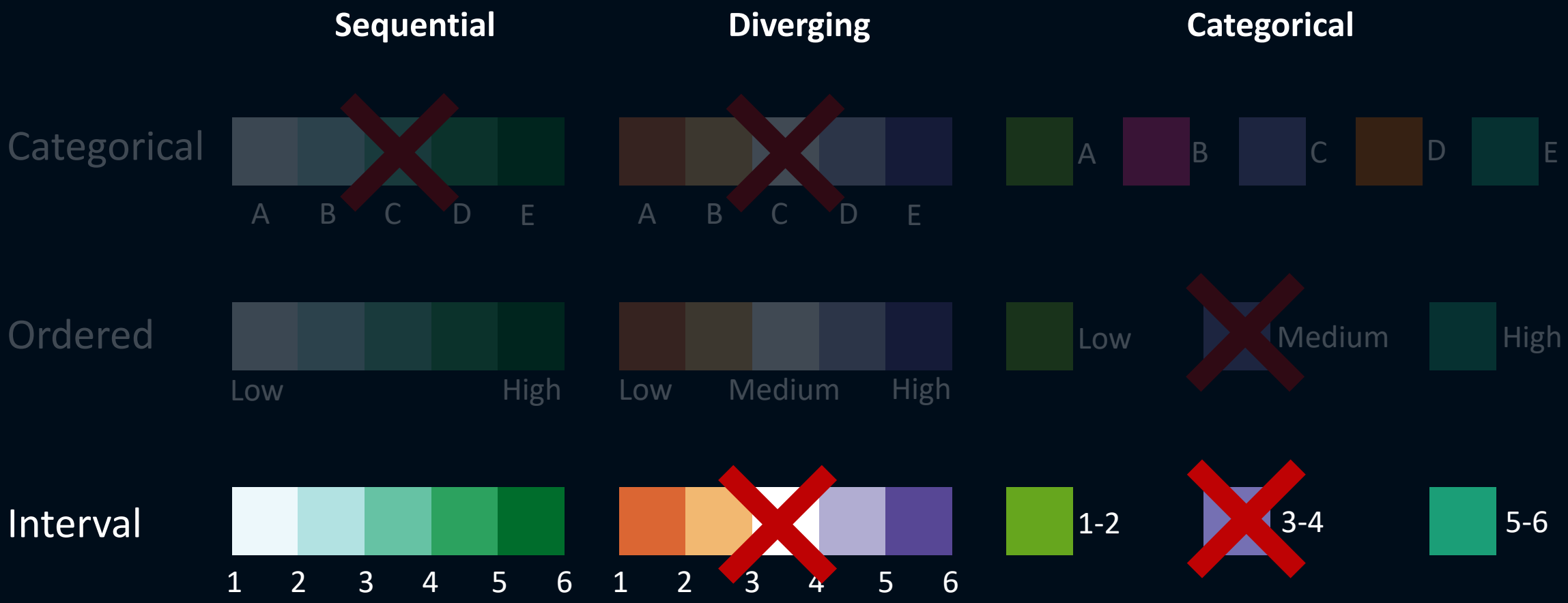
Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low Medium High

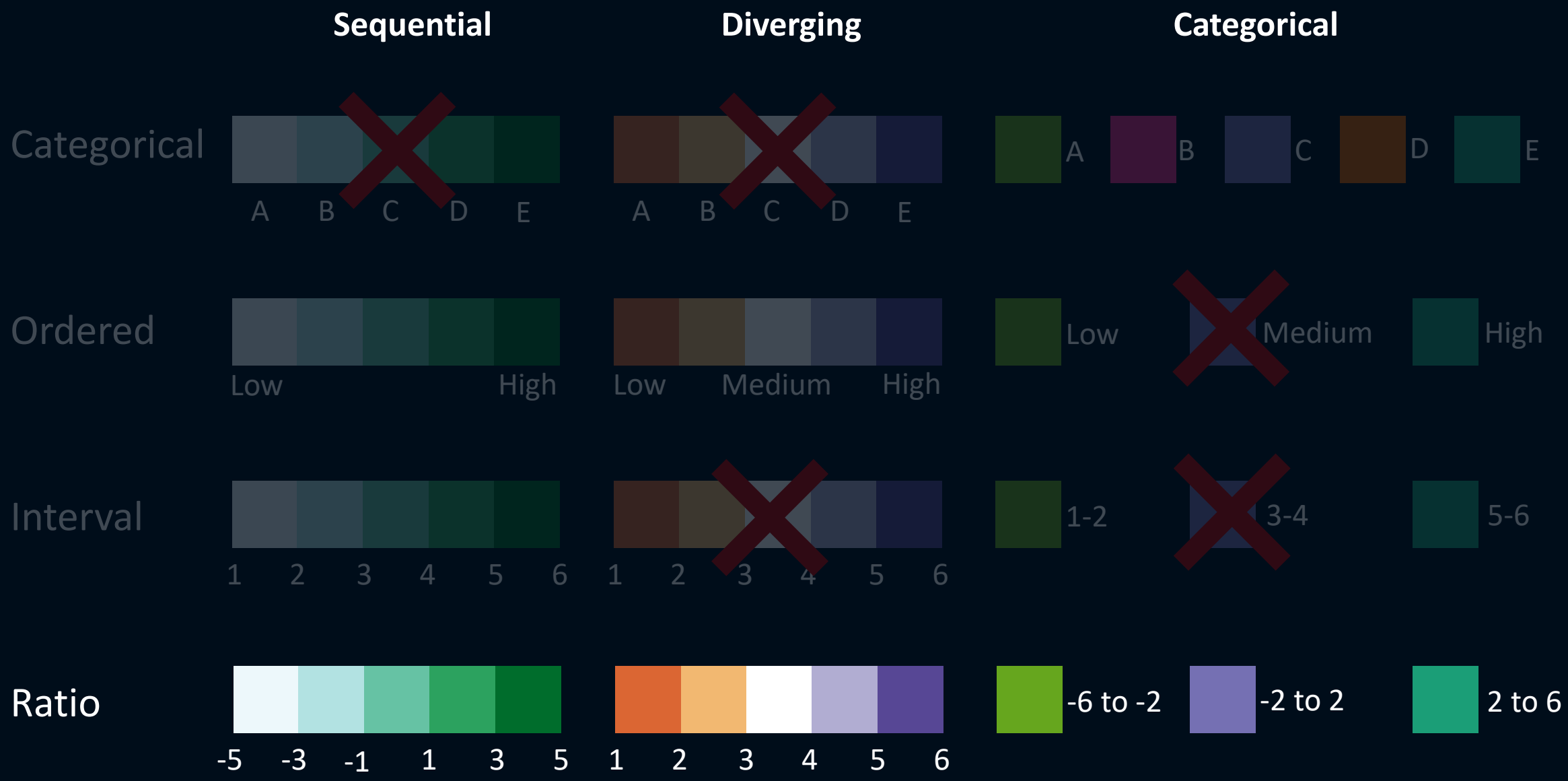
Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low Medium High
Interval
1 2 3 4 5 6
1 2 3 4 5 6
1-2
3-4
5-6

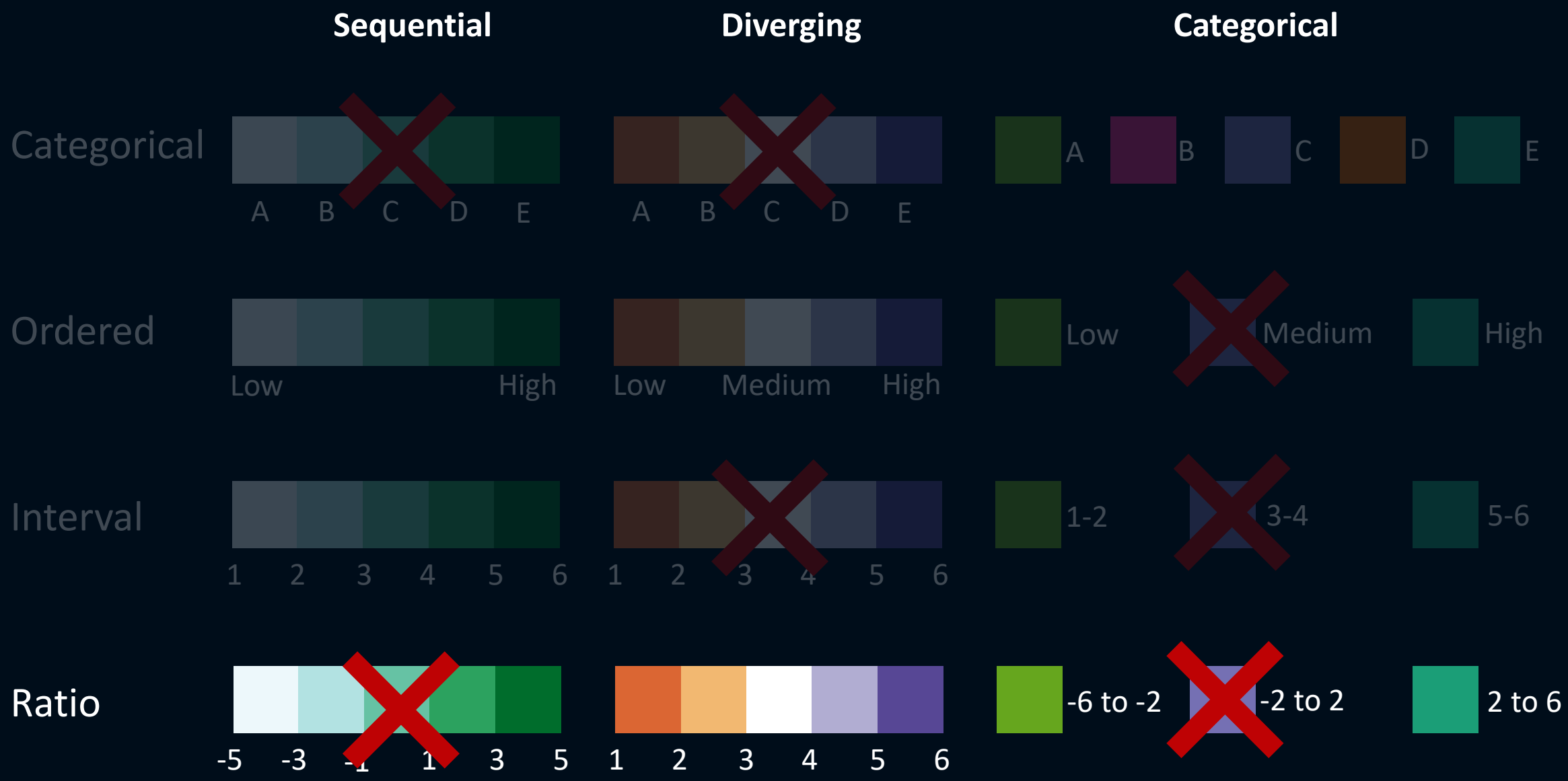
Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low Medium High
Interval
1 2 3 4 5 6
1 2 3 4 5 6
1-2 3-4 5-6

Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low
Medium
High
Interval
1 2 3 4 5 6
1 2 3 4 5 6
1-2
3-4
5-6
Ratio
-5 -3 -1 1 3 5
1 2 3 4 5 6
-6 to -2
-2 to 2
2 to 6

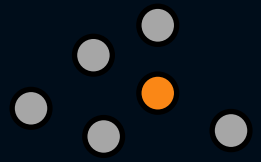
Colormaps


Sequential
Diverging
Categorical
Categorical
A B C D E
A B C D E
A B C D E
Ordered
Low High
Low Medium High
Low Medium High
Interval
1 2 3 4 5 6
1 2 3 4 5 6
1-2 3-4 5-6
Ratio
-5 -3 -1 1 3 5
1 2 3 4 5 6
-6 to -2 -2 to 2 2 to 6

Colormaps

Sequential **Diverging** **Categorical**

Pro tips:



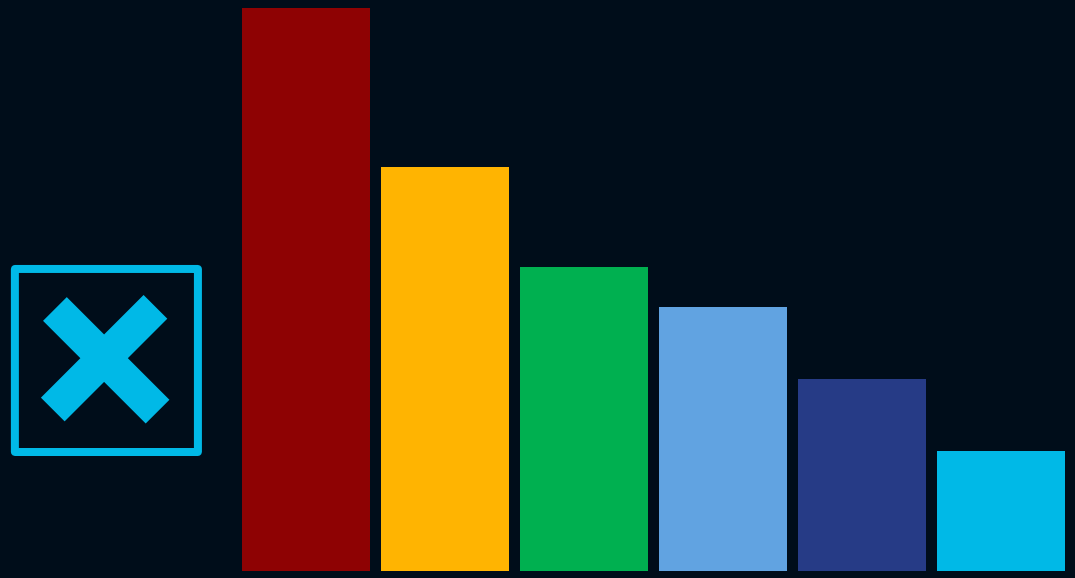
- Gray is one of the most important colors!
- Think about cultural connotations: www.informationisbeautiful.net/visualizations/colours-in-cultures
- Choose colormaps that work for people with color deficiencies (color blind)

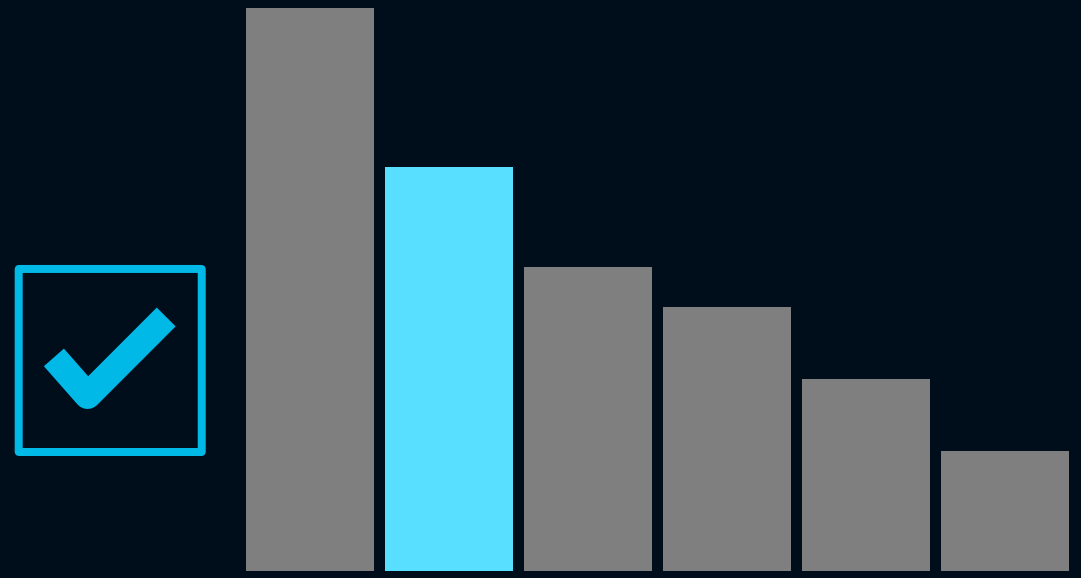
Recommendations & Pitfalls

Intuitiveness

Colors in Data Visualization [twitter.com]

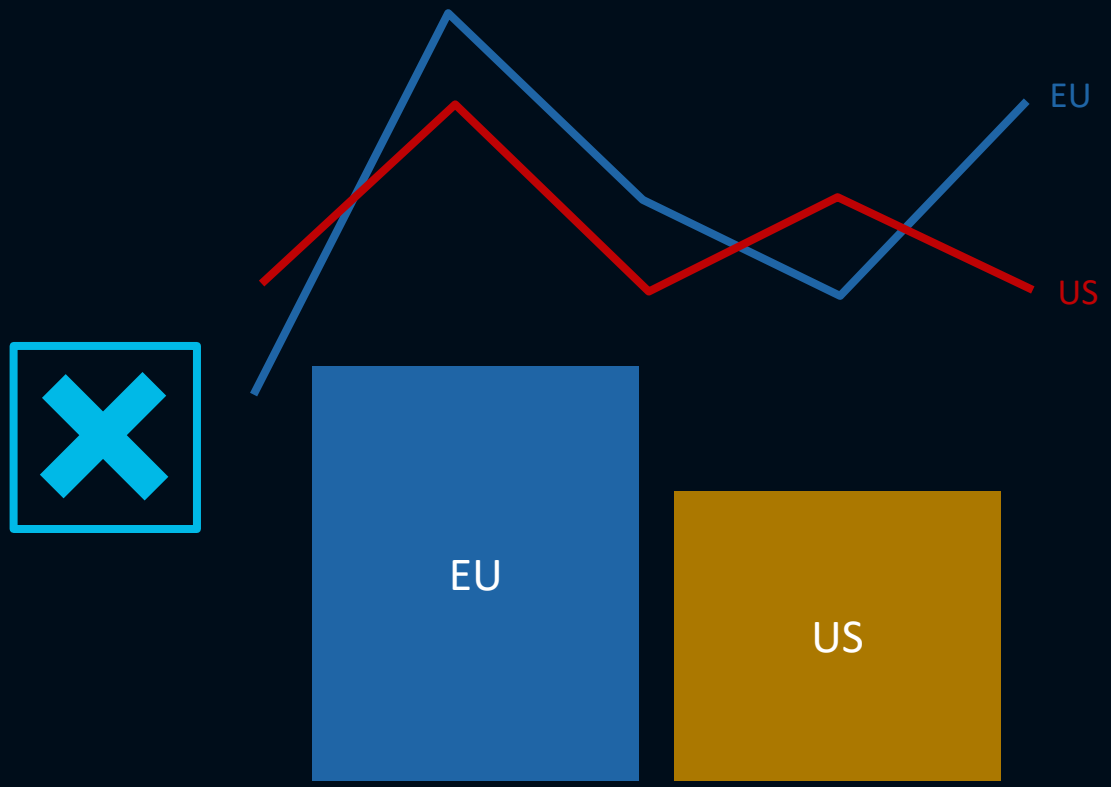
Moderation

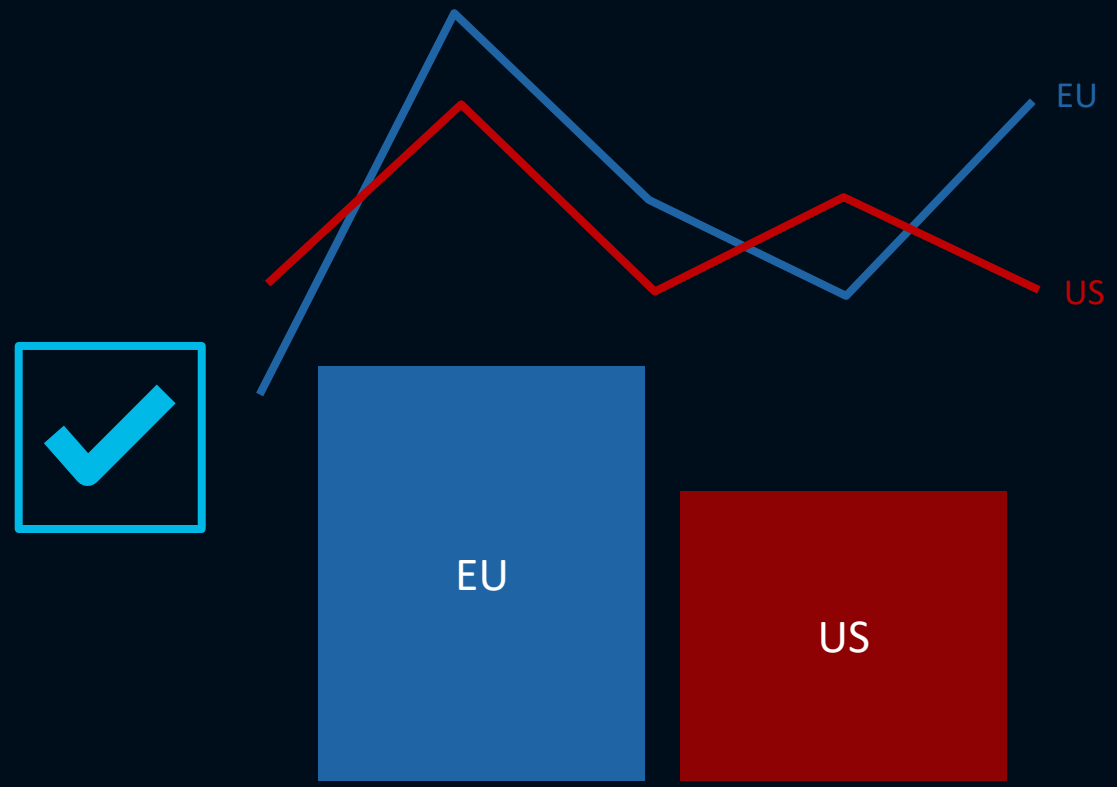




Colors in Data Visualization [twitter.com]

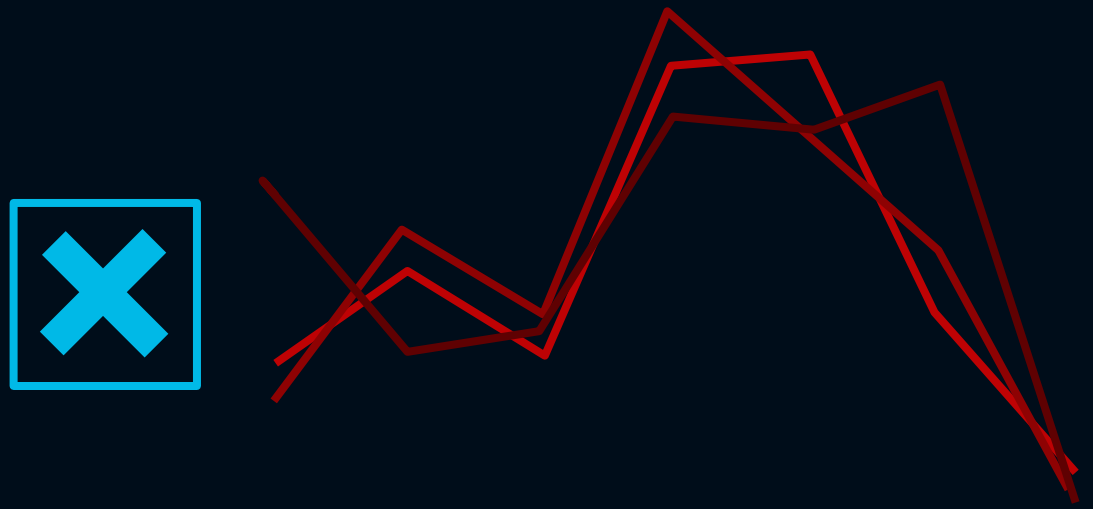
Consistency

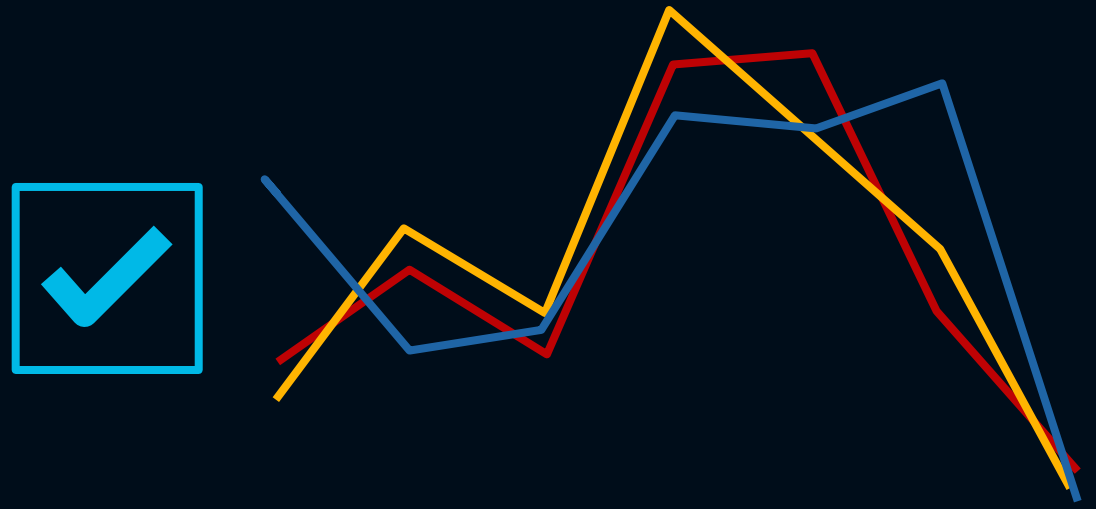




Colors in Data Visualization [twitter.com]

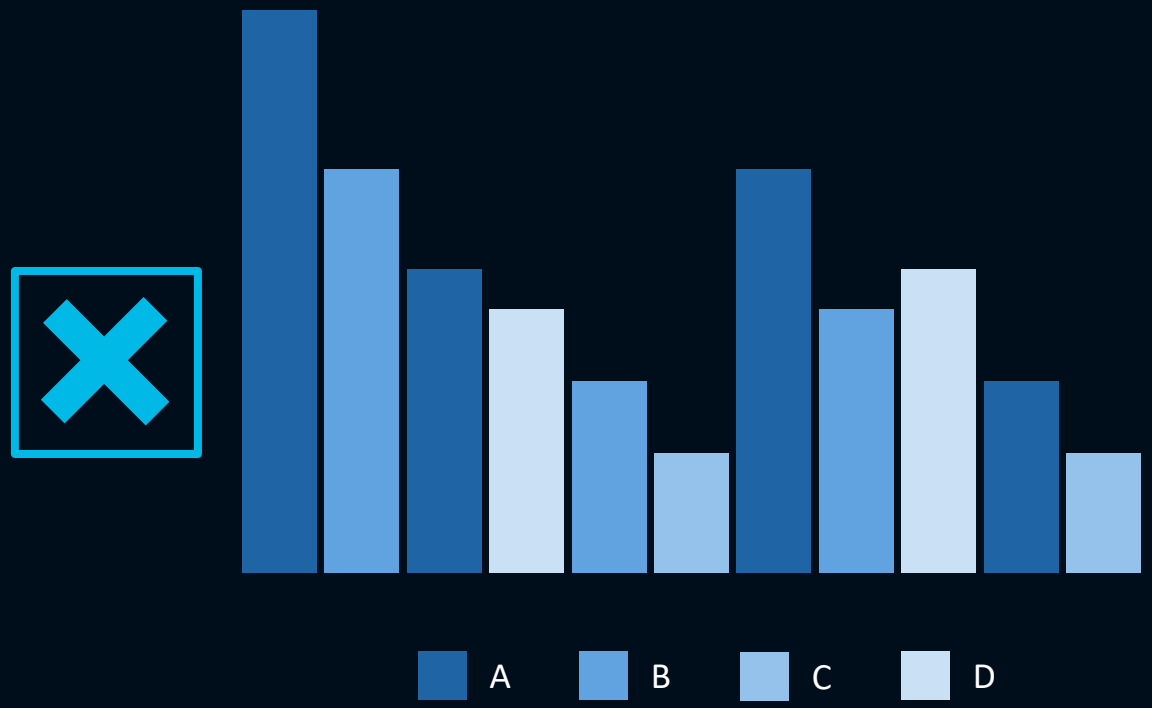
Clarity

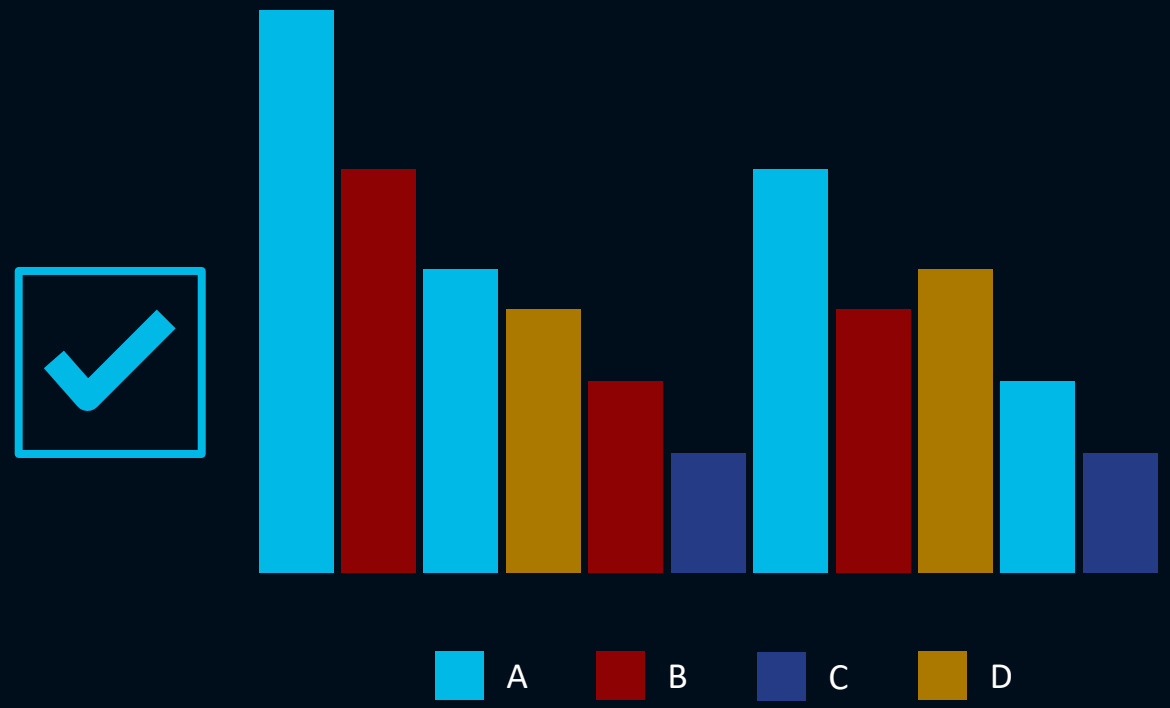




Colors in Data Visualization [twitter.com]

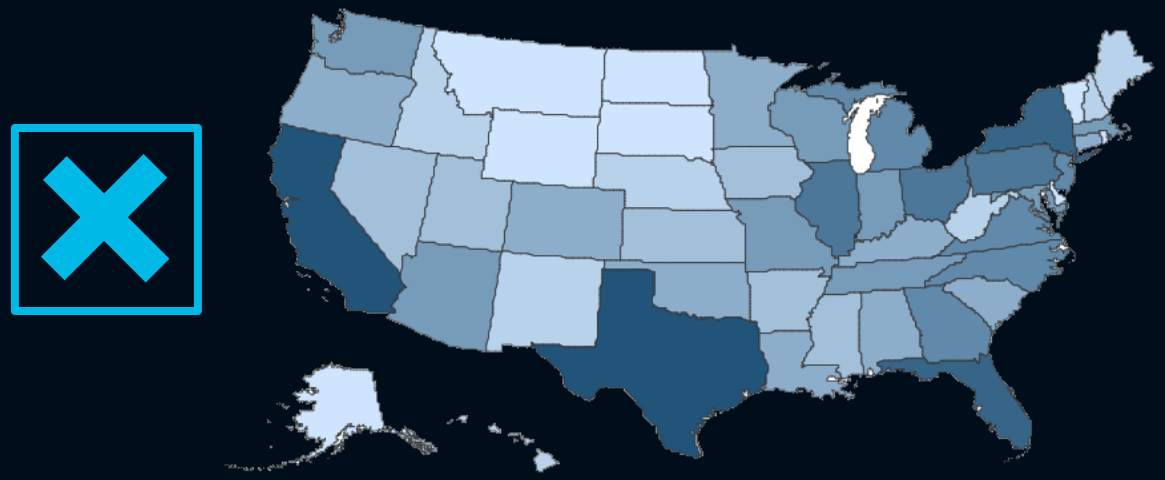
Classification



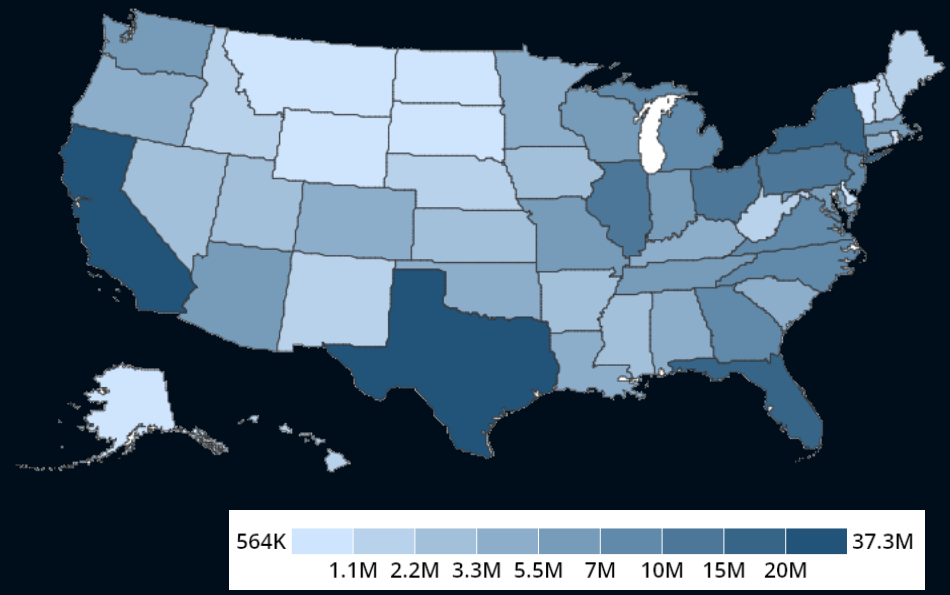


Colors in Data Visualization [twitter.com]

Explainability



US Population 2010



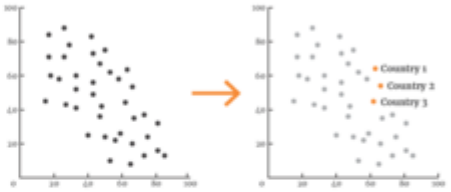
Colors in Data Visualization [twitter.com]

[Map from chartio.com]

Core Principles of Data Visualization

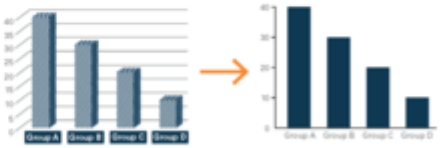
Show the data

People read graphs in a research report, article, or blog to understand the story being told. The data is the most important part of the graph and should be presented in the clearest way possible. But that does not mean that all of the data must be shown—indeed, many graphs show too much.



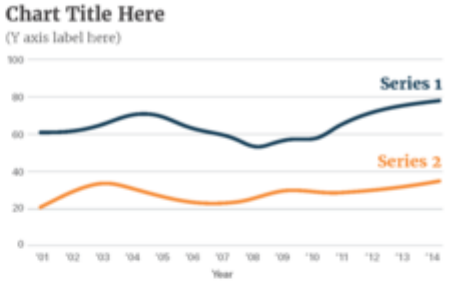
Reduce the clutter

Chart clutter, those unnecessary or distracting visual elements, will tend to reduce effectiveness. Clutter comes in the form of dark or heavy gridlines; unnecessary tick marks, labels, or text; unnecessary icons or pictures; ornamental shading and gradients; and unnecessary dimensions. Too often graphs use textured or filled gradients.



Integrate the text and the graph

Standard research reports often suffer from the **slideshow effect**, in which the writer narrates the text elements that appear in the graph. A better model is one in which visualizations are constructed to complement the text and at the same time to contain enough information to stand alone. As a simple example, legends that define or explain a line, bar, or point are often placed far from the content of the graph—off to the right or below the graph. Integrated legends—right below the title, directly on the chart, or at the end of a line—are more accessible.



Preattentive Processing

Effective data visualization taps into the brain's **preattentive visual processing.** Because our eyes detect a limited set of visual characteristics (such as shape and contrast), we combine various characteristics of an object and unconsciously perceive them as comprising an image. Preattentive processing refers to the cognitive operations that can be performed prior to focusing attention on any particular region of an image. In other words, it's the stuff you notice right away.

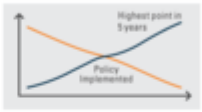
PolicyViz

Core Principles of Data Visualization

Audience

Always consider your audience—whether they need a short, written report, a more in-depth paper, or an online exploratory data tool.

Include annotation



Add explanatory text to help the reader understand how to read or use the visualization (if necessary) and also to guide them through the content.

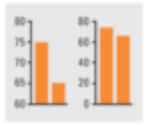
Use pie charts with care

We are not very good at discerning quantities from the slices of the pie chart. Other chart types—for example, bars, stacked bars, treemaps, or slope charts—may be a better choice.

Avoid 3D

Using 3D when you don't have a third variable will usually distort the perception of the data and should thus be avoided.

Start bar and column charts at zero



Bar and column charts that do not start at zero overemphasize the differences between the values. For small changes in quantities, consider visualizing the difference or the change in the values.

Make labels easy to read

When applicable, rotate bar and column charts to make the labels horizontal. If possible, make vertical axis labels horizontal, possibly below the title. In general, make labels clear, concise, and easy for your reader to understand.

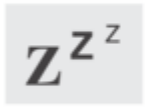
Try small multiples

Breaking up a complicated chart into smaller chunks can be an effective way to visualize your data.

Use maps carefully

Use maps carefully, always being sure it is the geographic point you are trying to make. Column and bar charts, for example, are often better at enabling comparisons between geographic units.

Color and font considerations

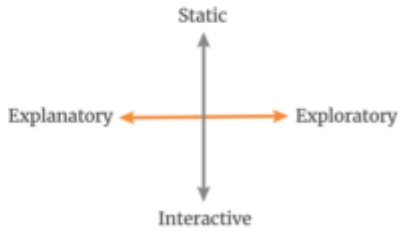


Avoid default colors and fonts—they all look the same and don't stand out.

Consider color blindness—about 10% of people (mostly men) have some form of color blindness.

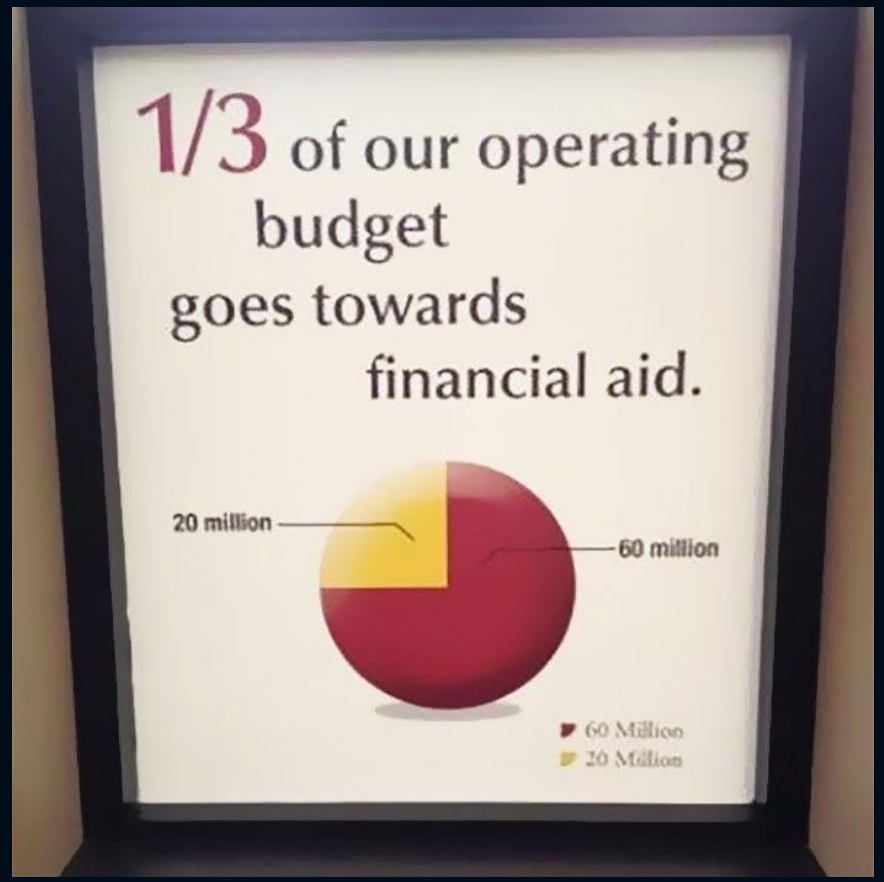
Avoid the rainbow color palette—it doesn't map to our number system and there is no logical ordering.

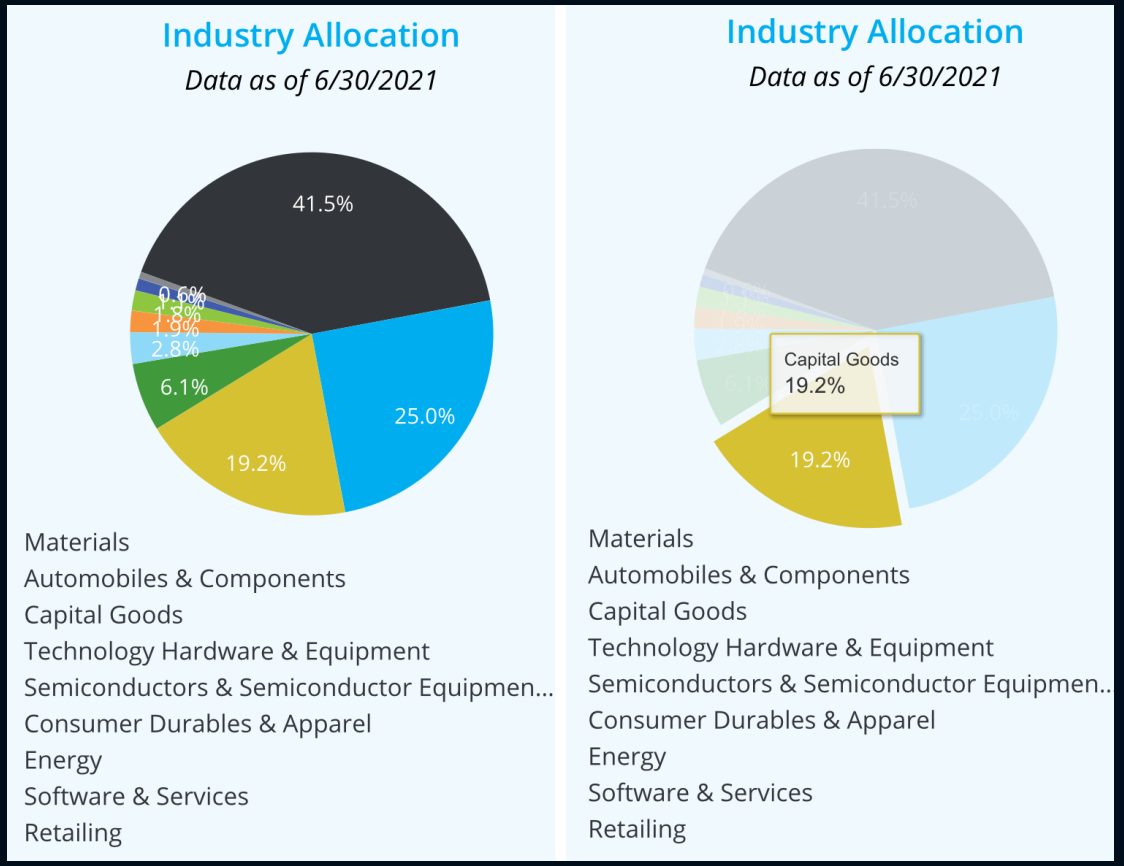
Visualization Mapping: Form and Function



PolicyViz

Pitfalls: missing/wrong labels





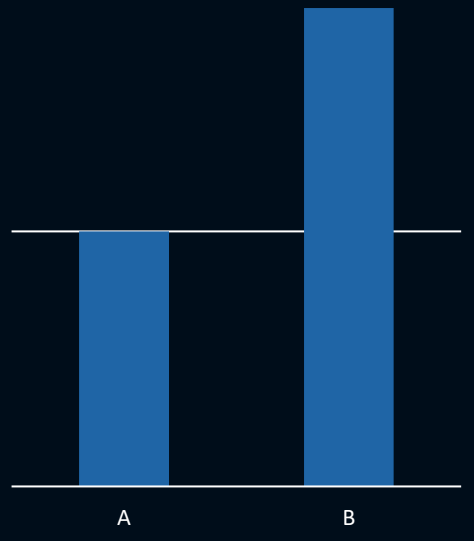
Pitfalls: axis labels

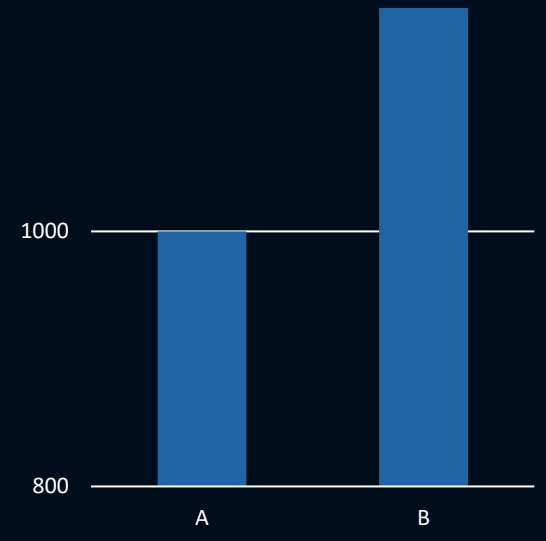
Missing axis labels

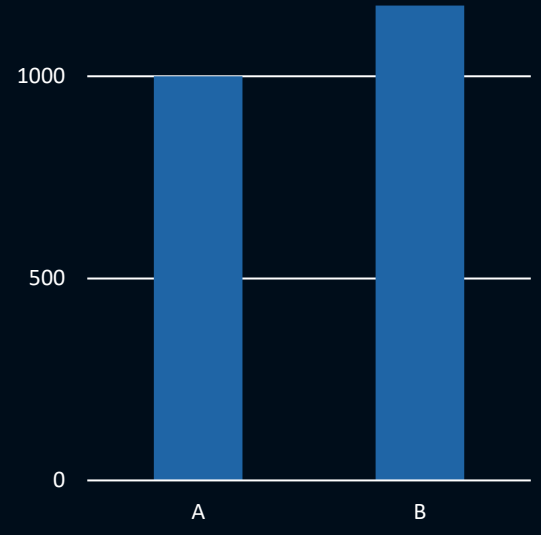
Misleading labels

- Not zero-based
- Using different scales in multiple plots

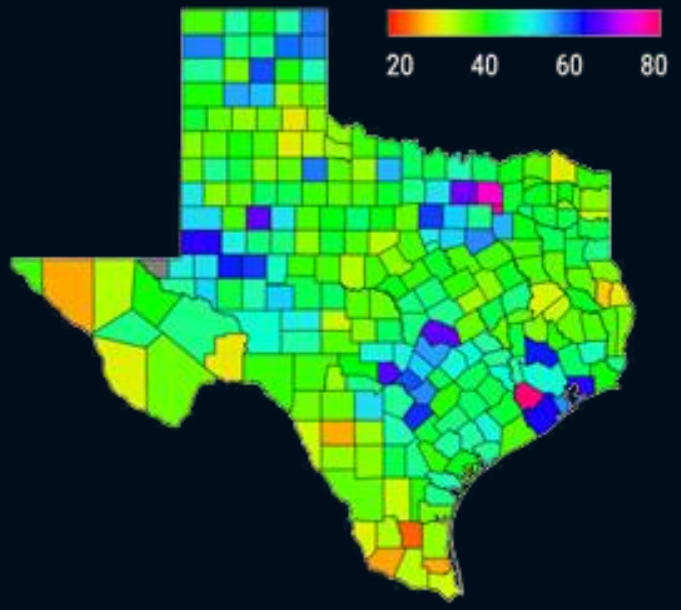
Missing units



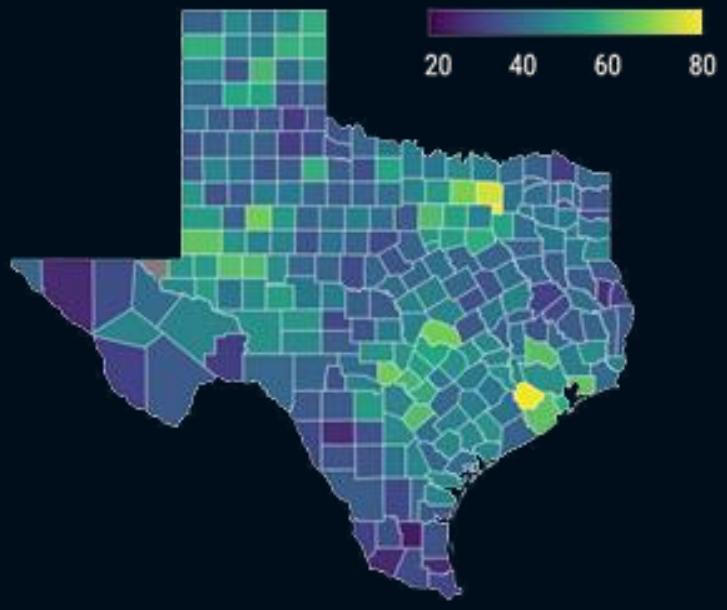




Pitfalls: Colormaps



Rainbow color map

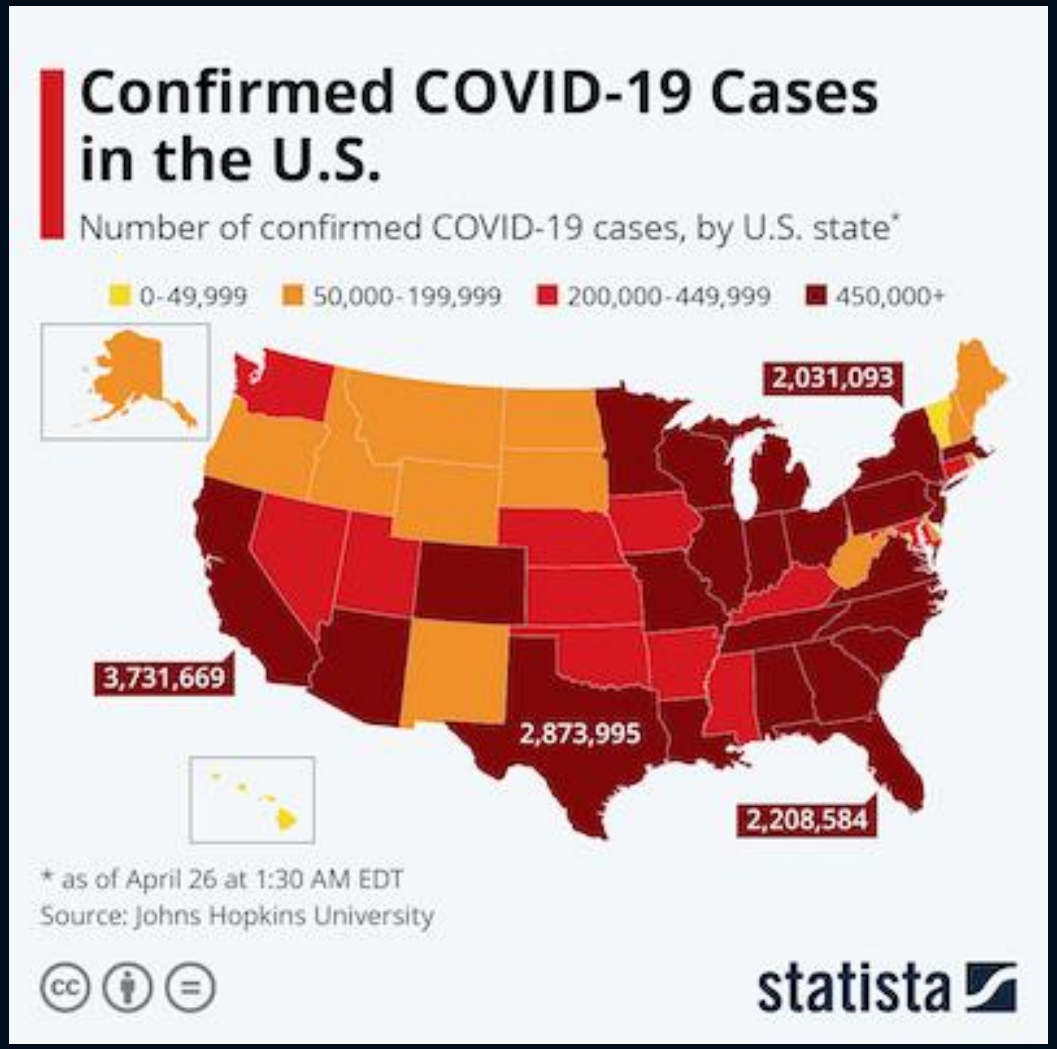


Perceptually uniform color map

Perception distorted by colorful graphics [phys.org]

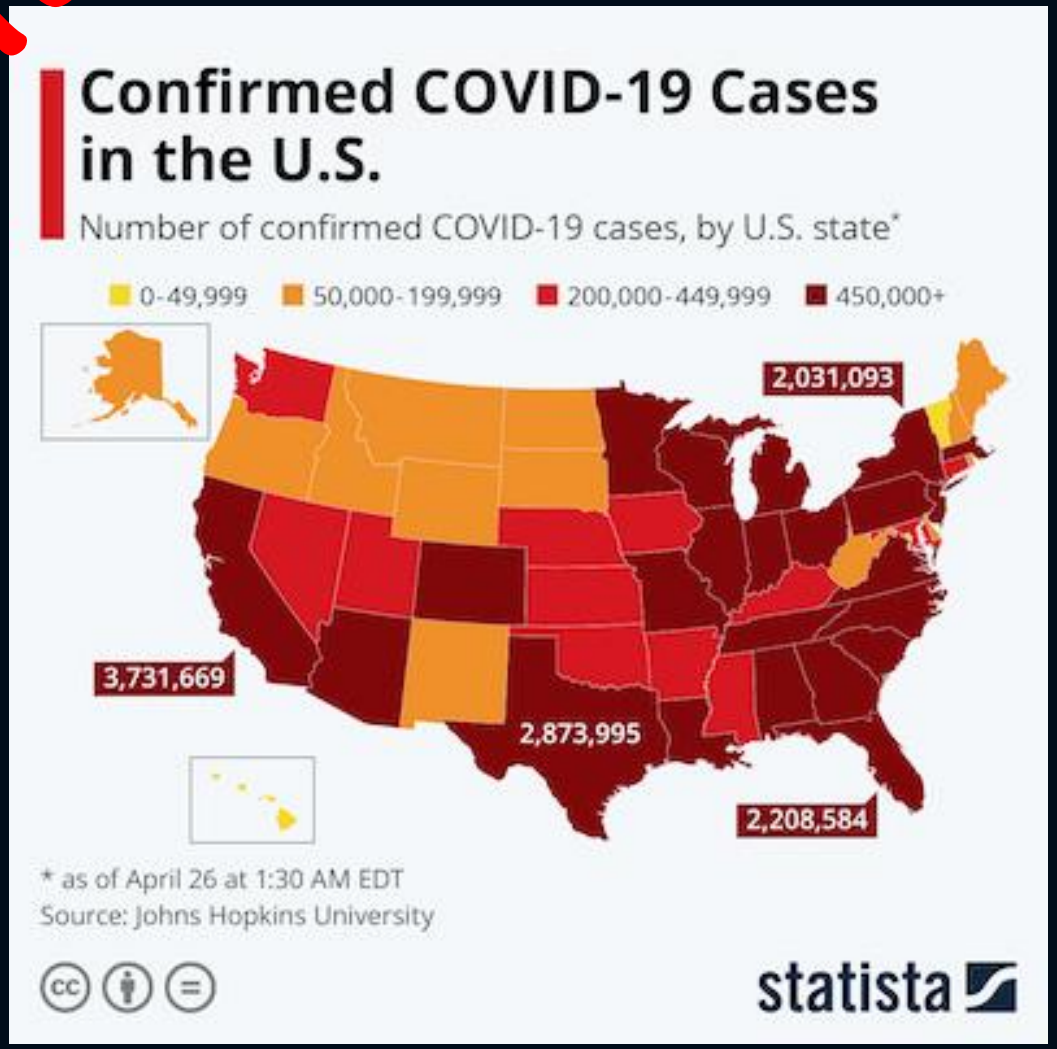
Choice of colormap can distort perception of data values

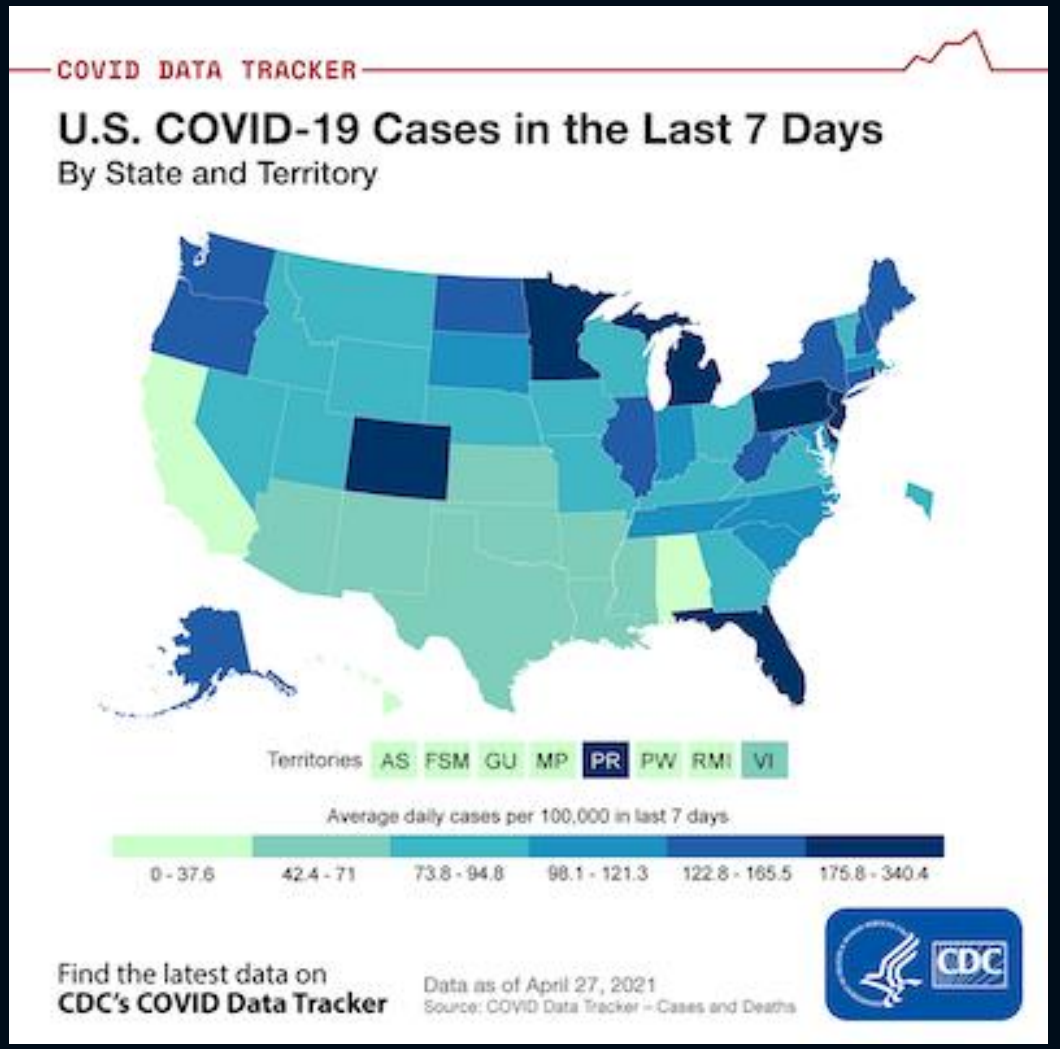
Pitfalls: absolute numbers



Pitfalls: biased visualization

LIE

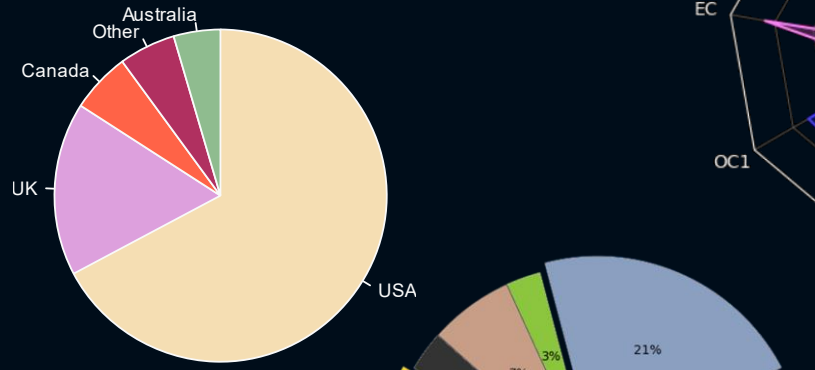


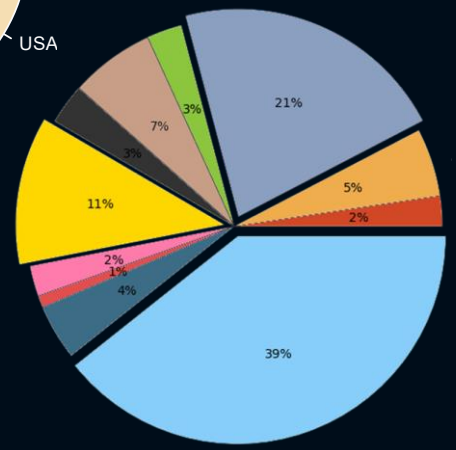


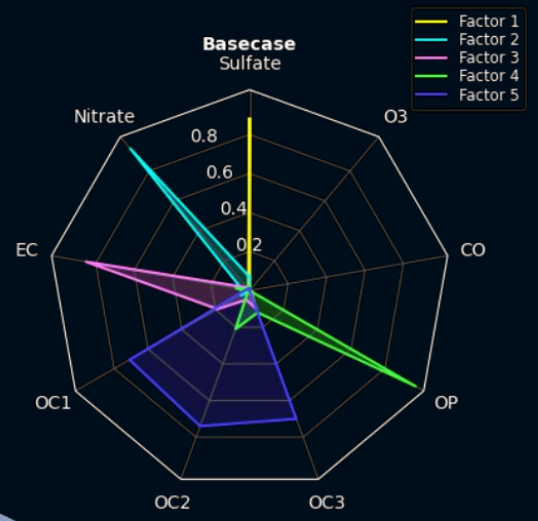
Pitfalls: area effects

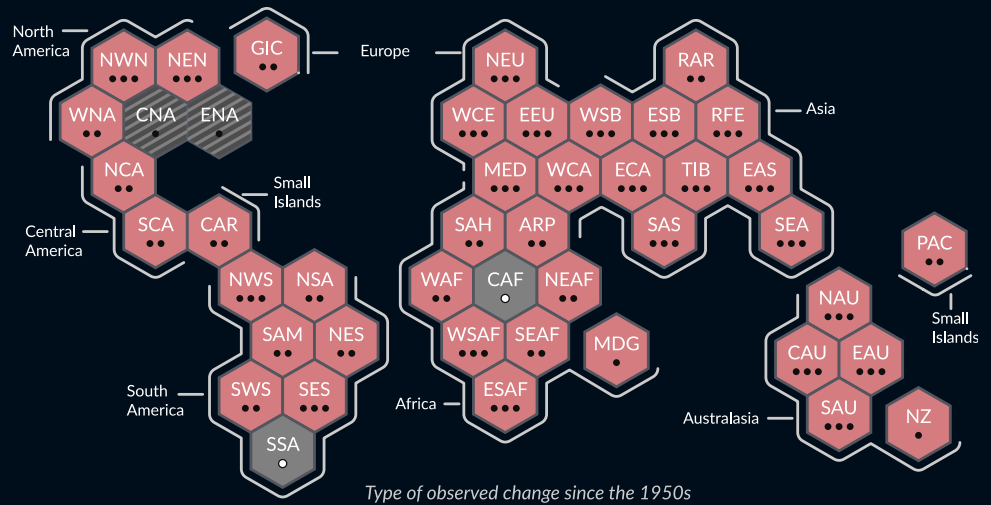
Data often not mapped to area

- Intrinsic part of the visualization
- Can mislead
- Affects comparison



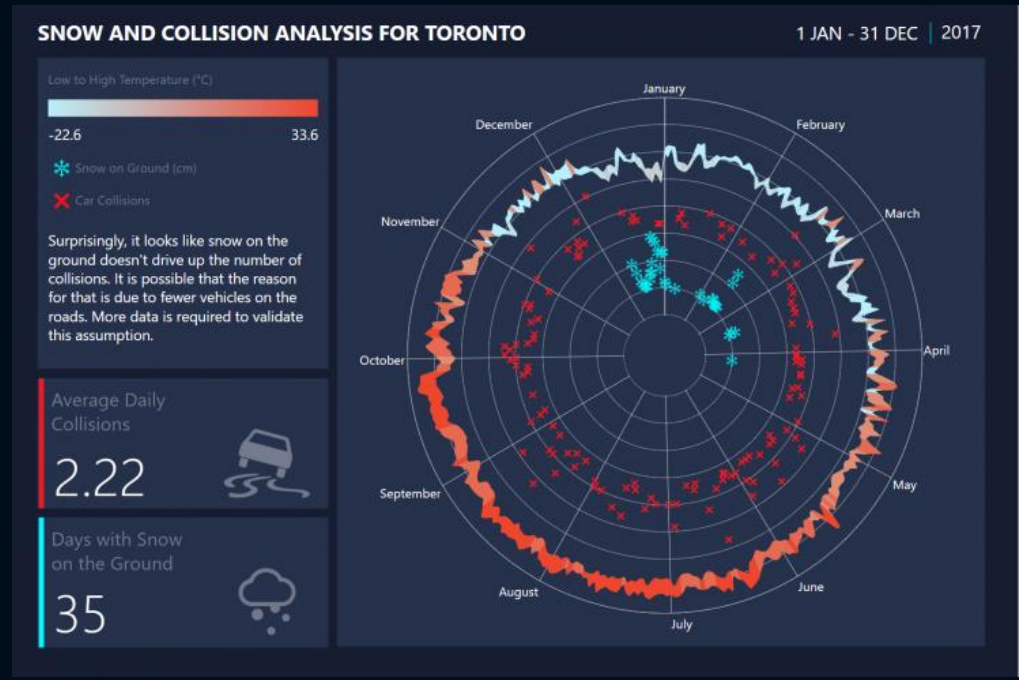






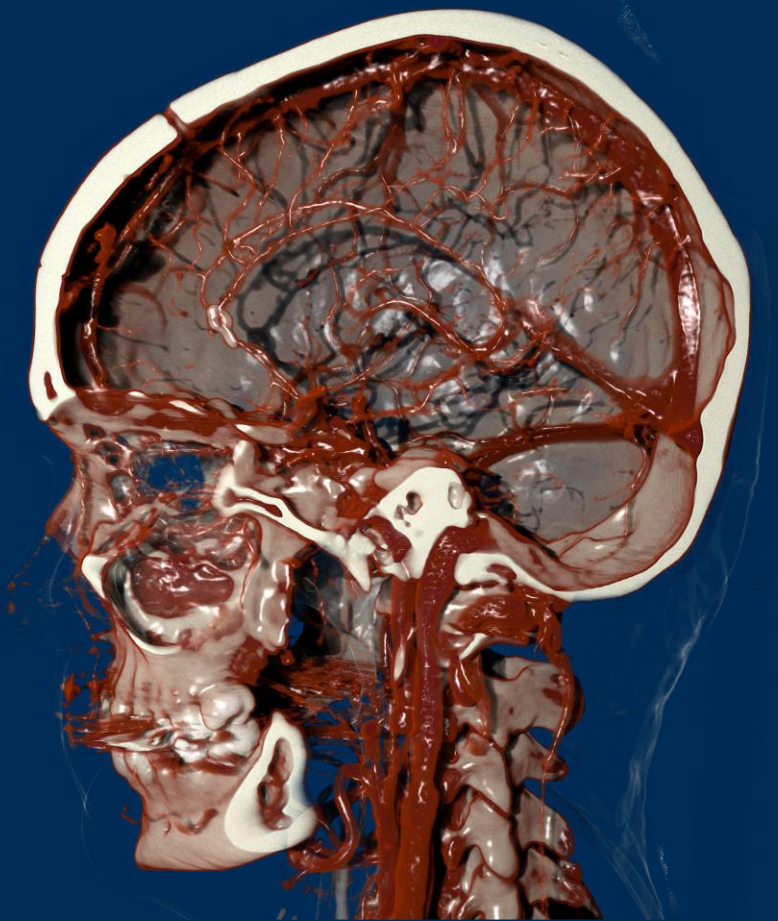
Type of observed change since the 1950s

IPCC report summary [ipcc.ch]



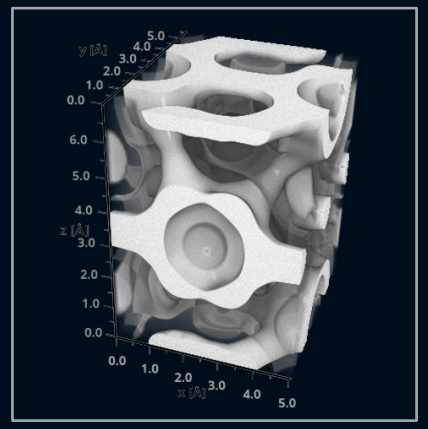
[Perception distorted by colorful graphics, phys.org]

Spatial data

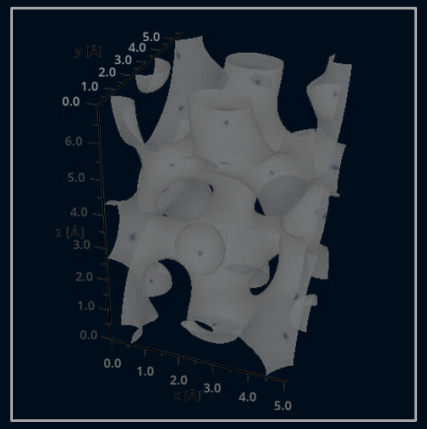


Spatial data (scalar and vector fields)

SCALAR FIELDS



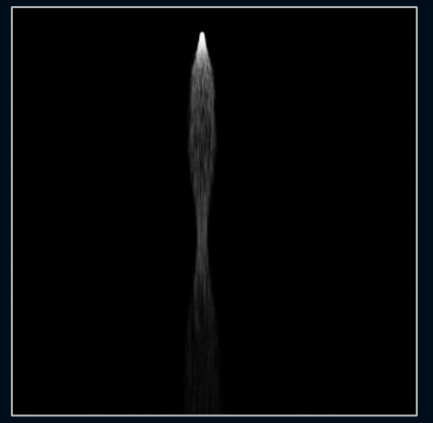
VOLUME RENDERING



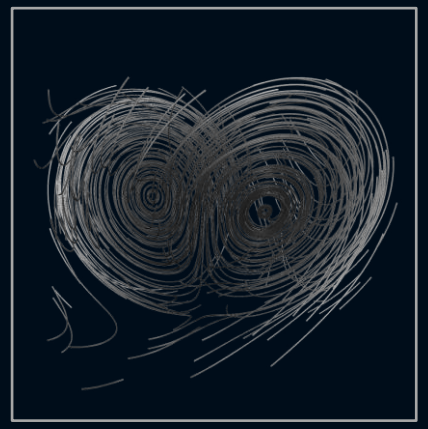
ISO-SURFACE

VECTOR FIELDS

LINE INTEGRAL CONVOLUTION



STREAM PARTICLES



STREAM LINES

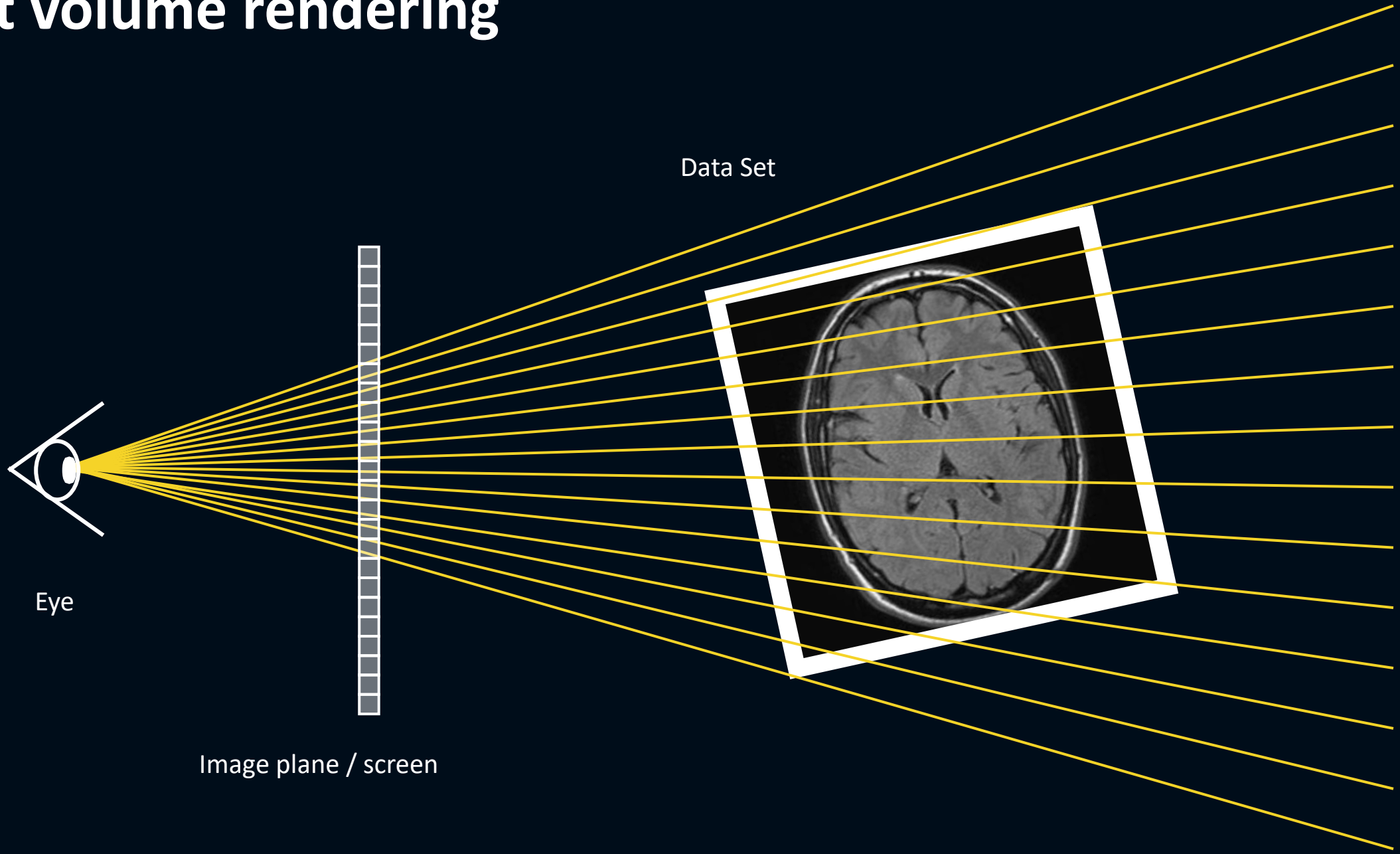
Volume rendering

“Looking into a volumetric dataset”

Requires transparency

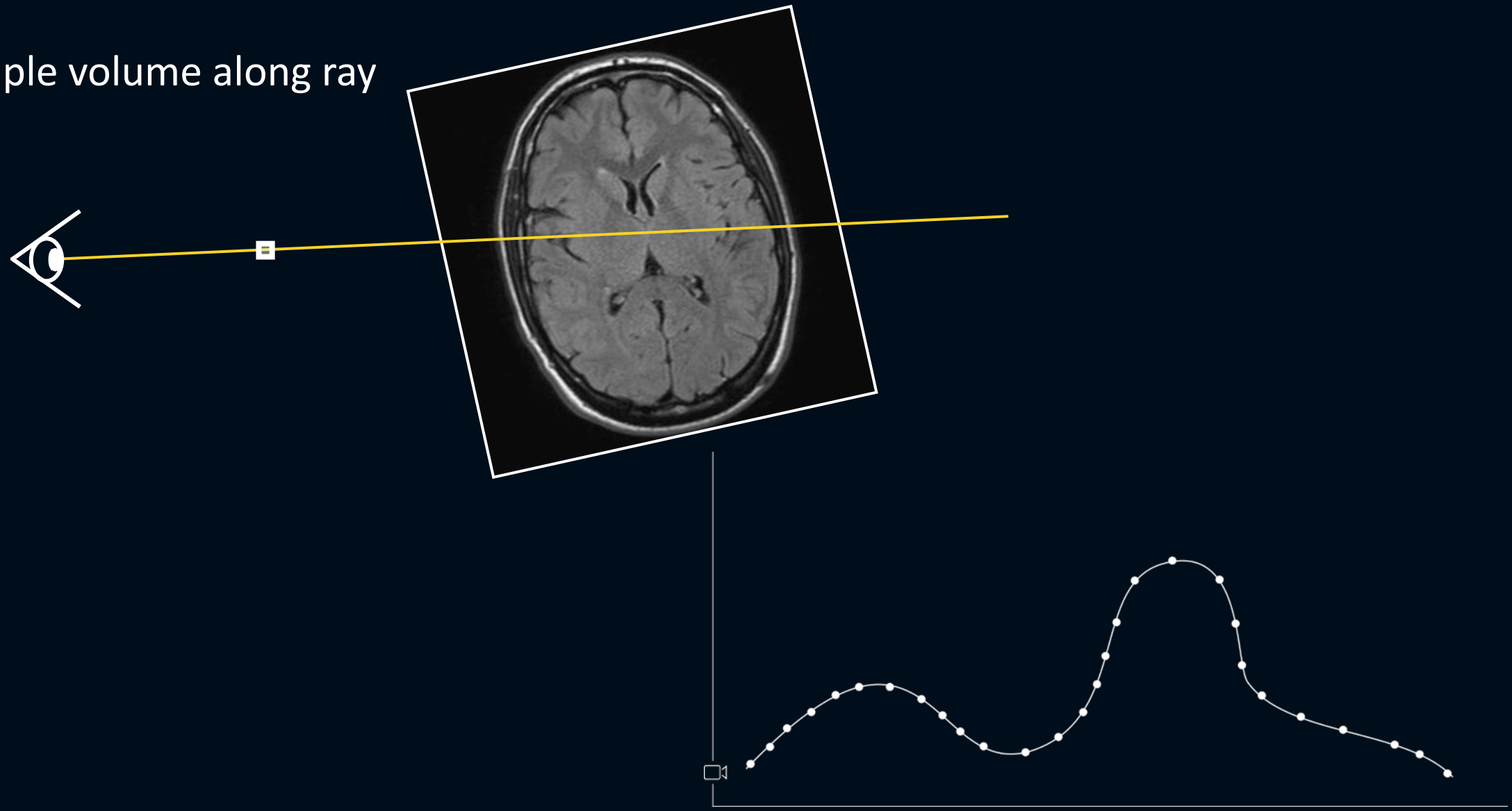
Scalar values describe optical properties

Direct volume rendering



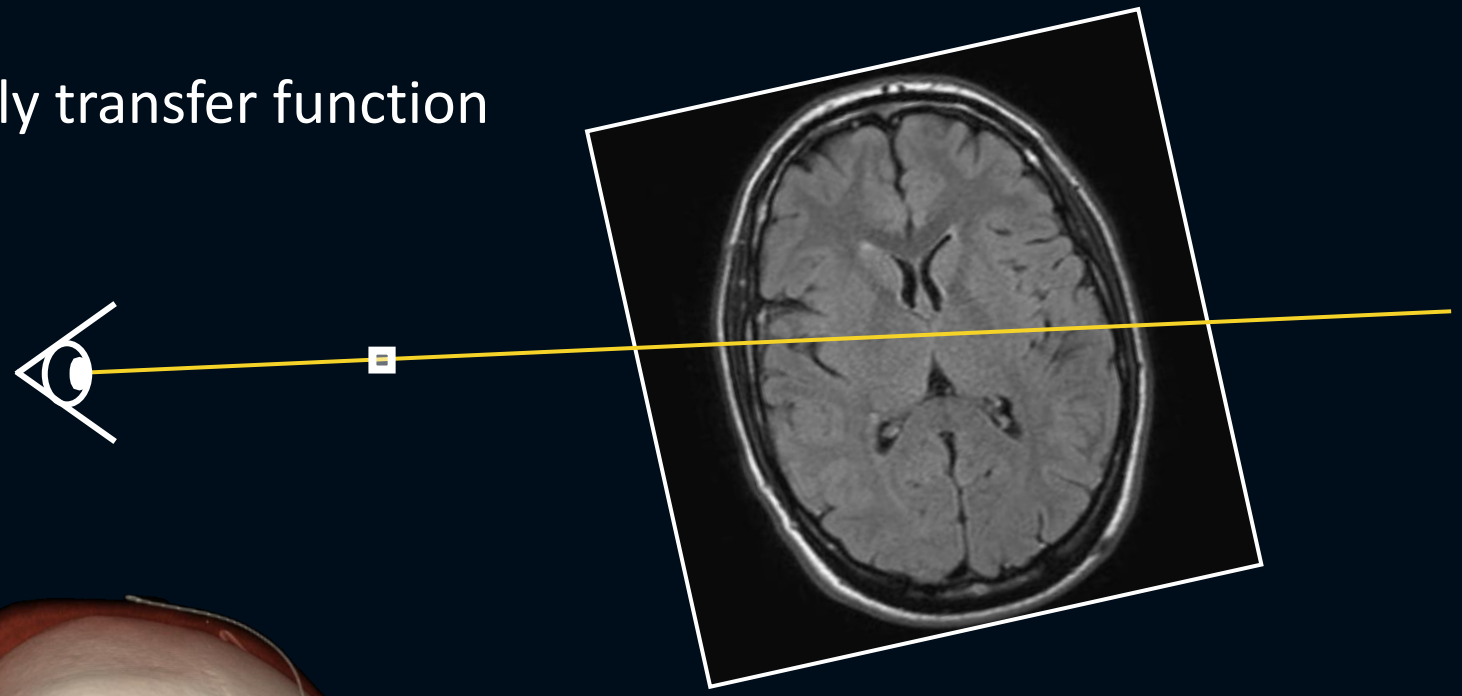
Ray casting

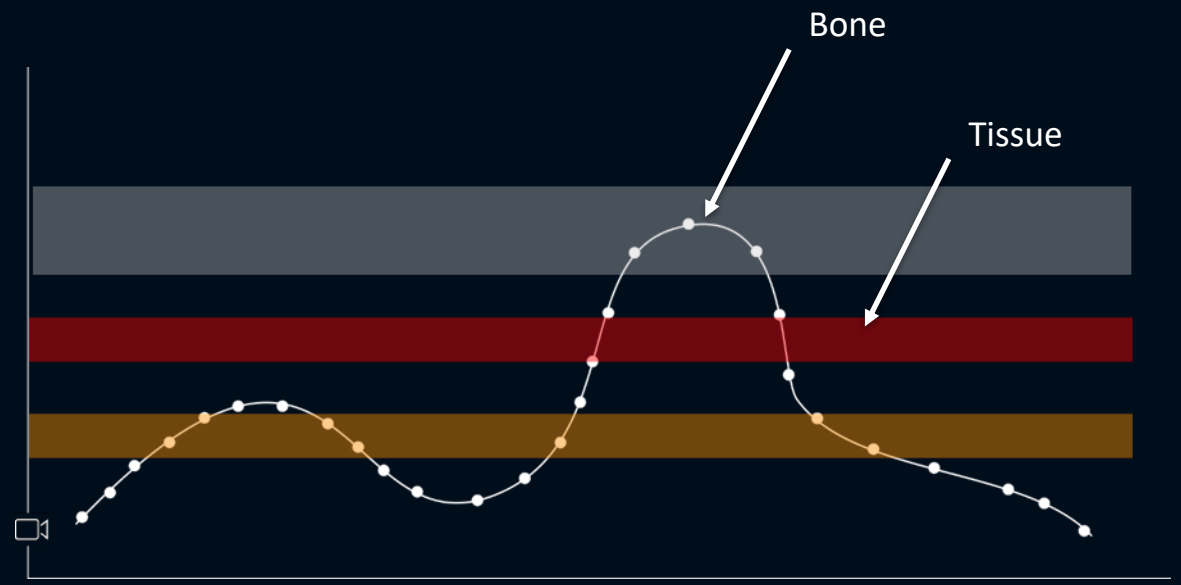
Sample volume along ray

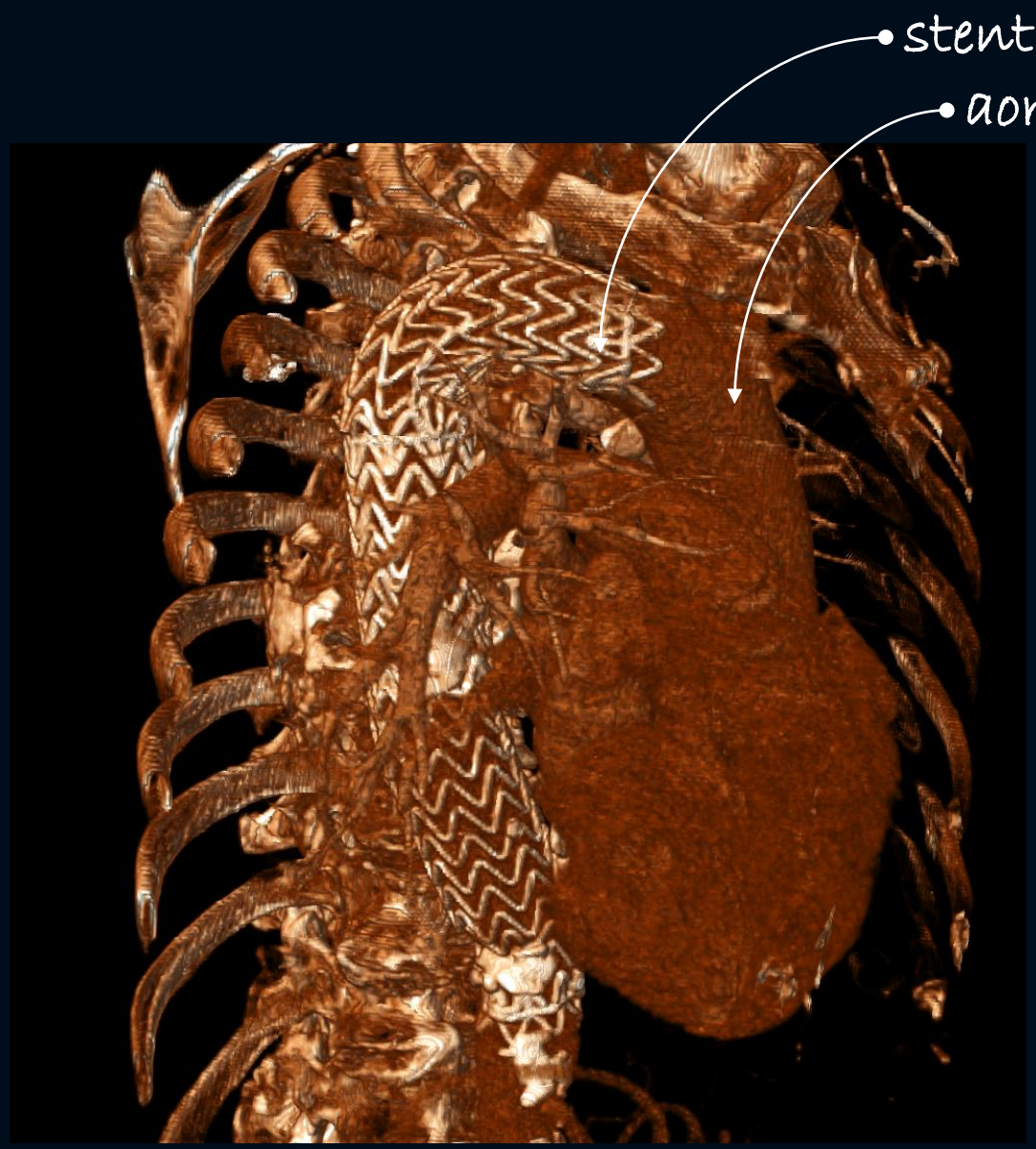


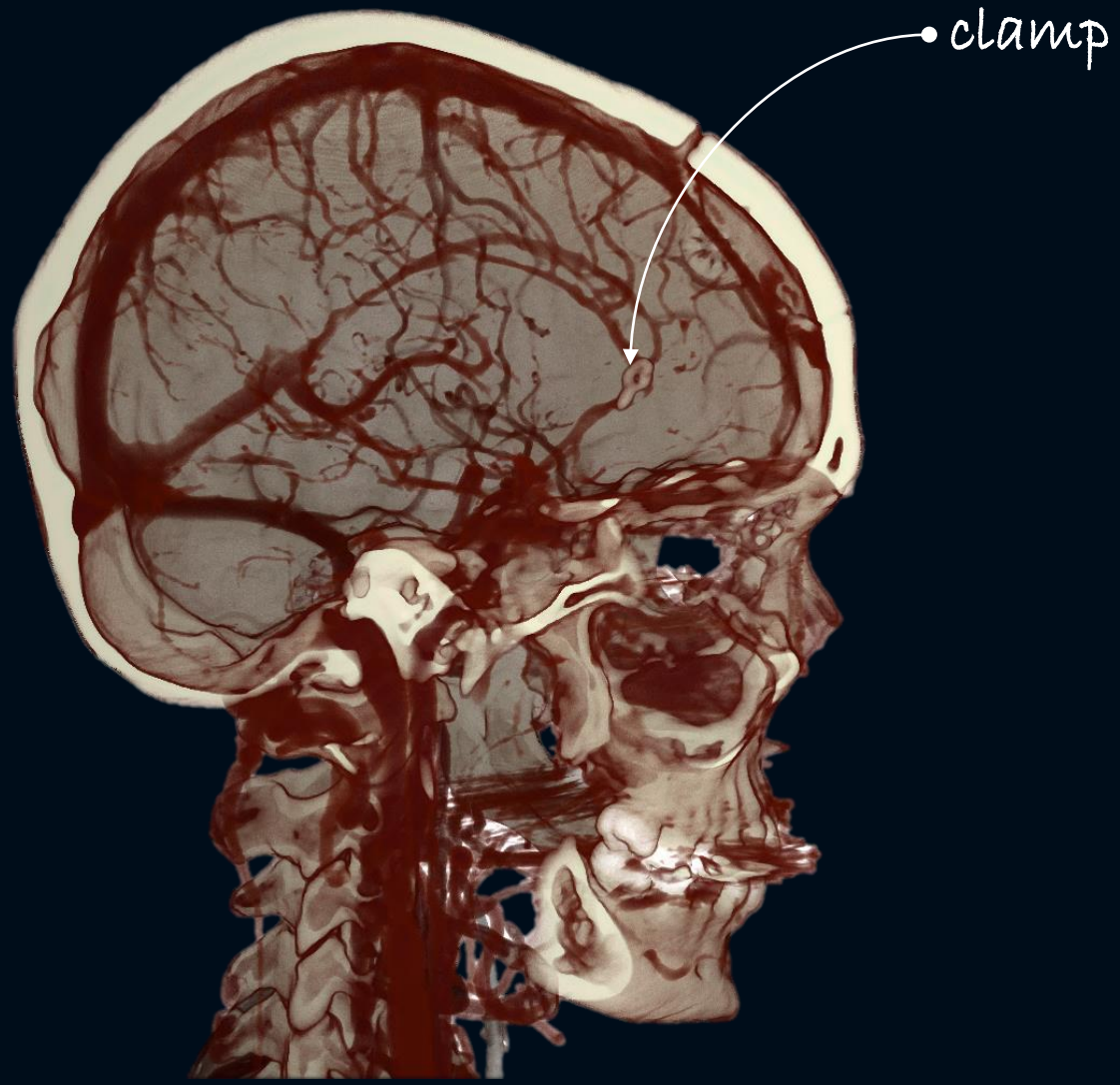
Ray casting: color mapping

Apply transfer function

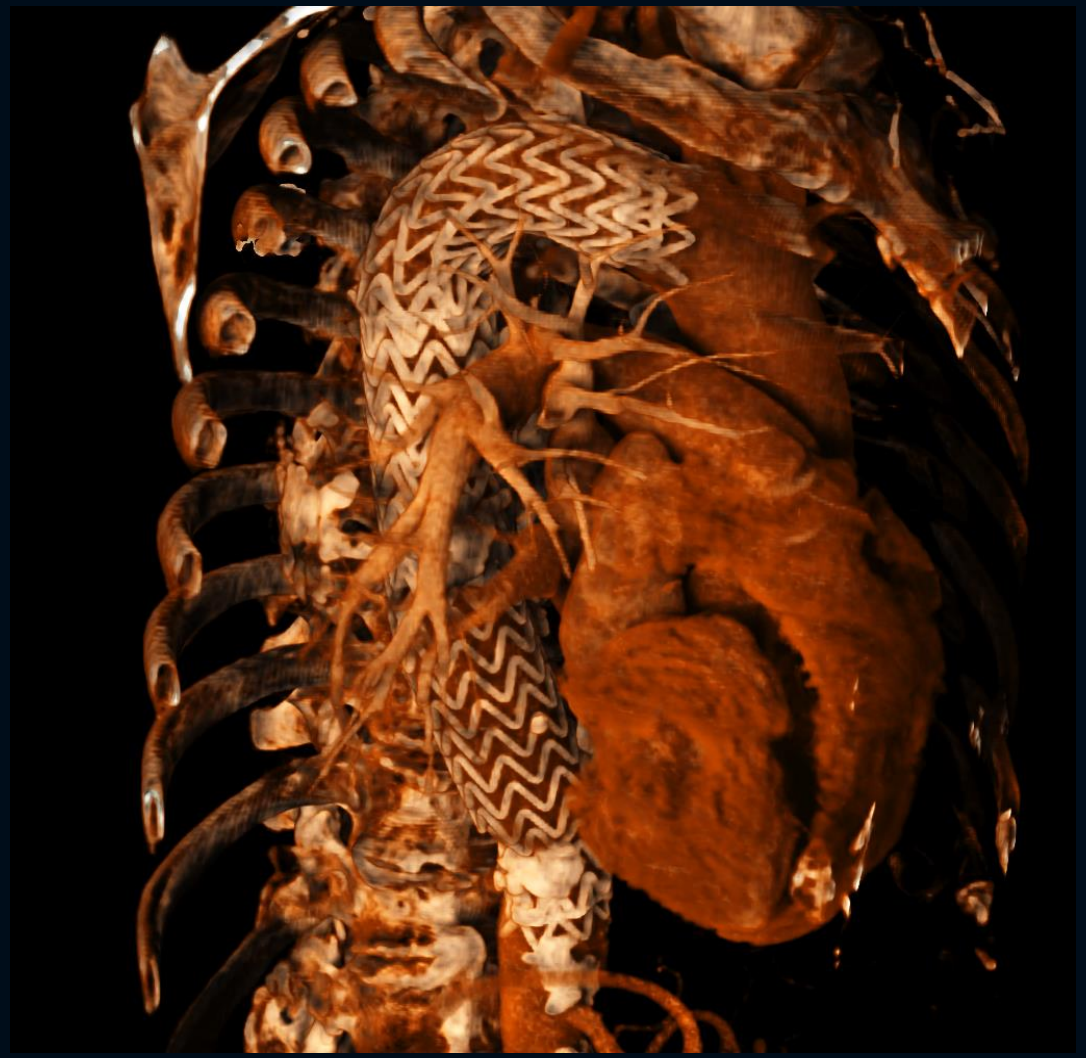


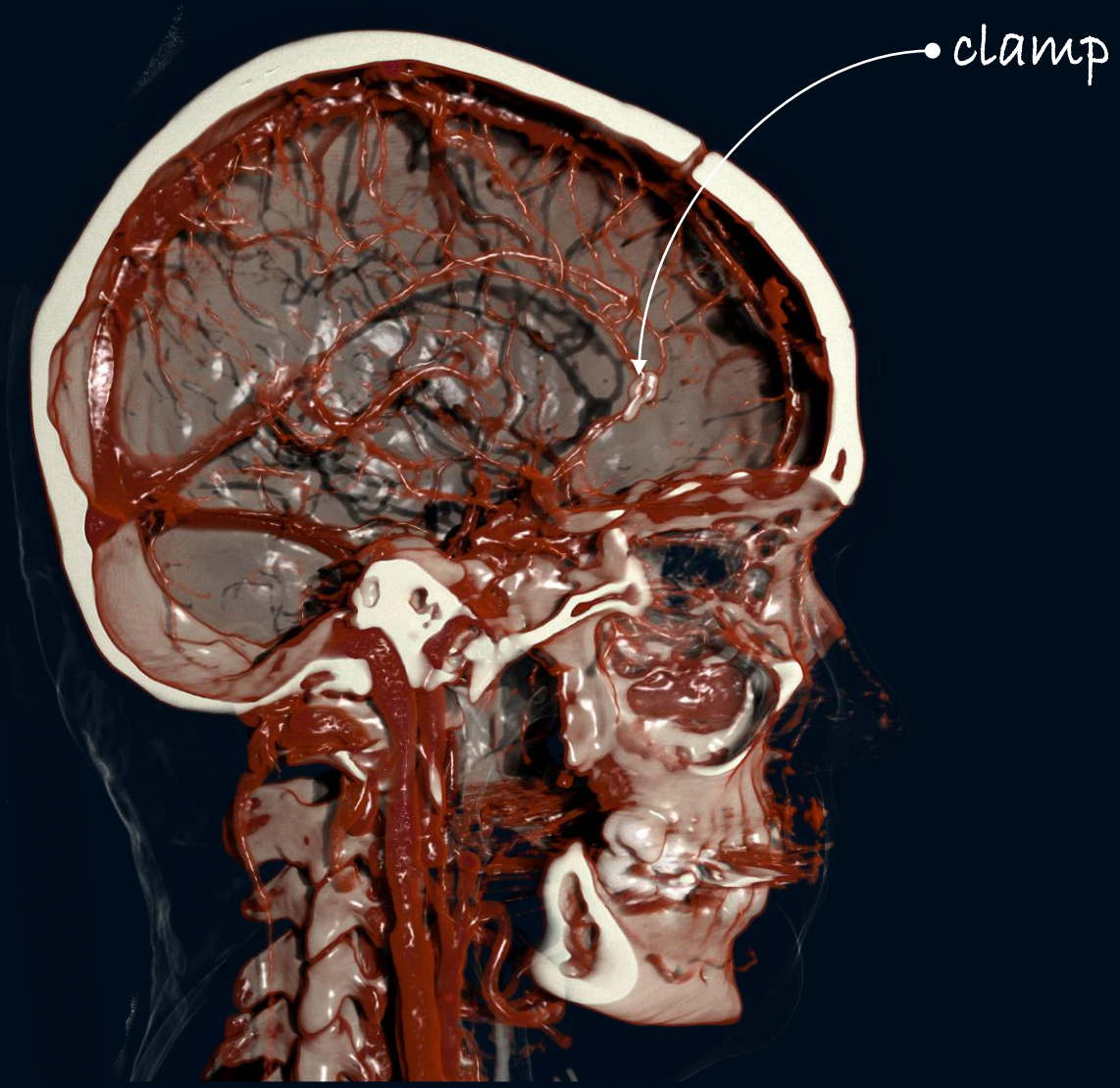



stent
aorta

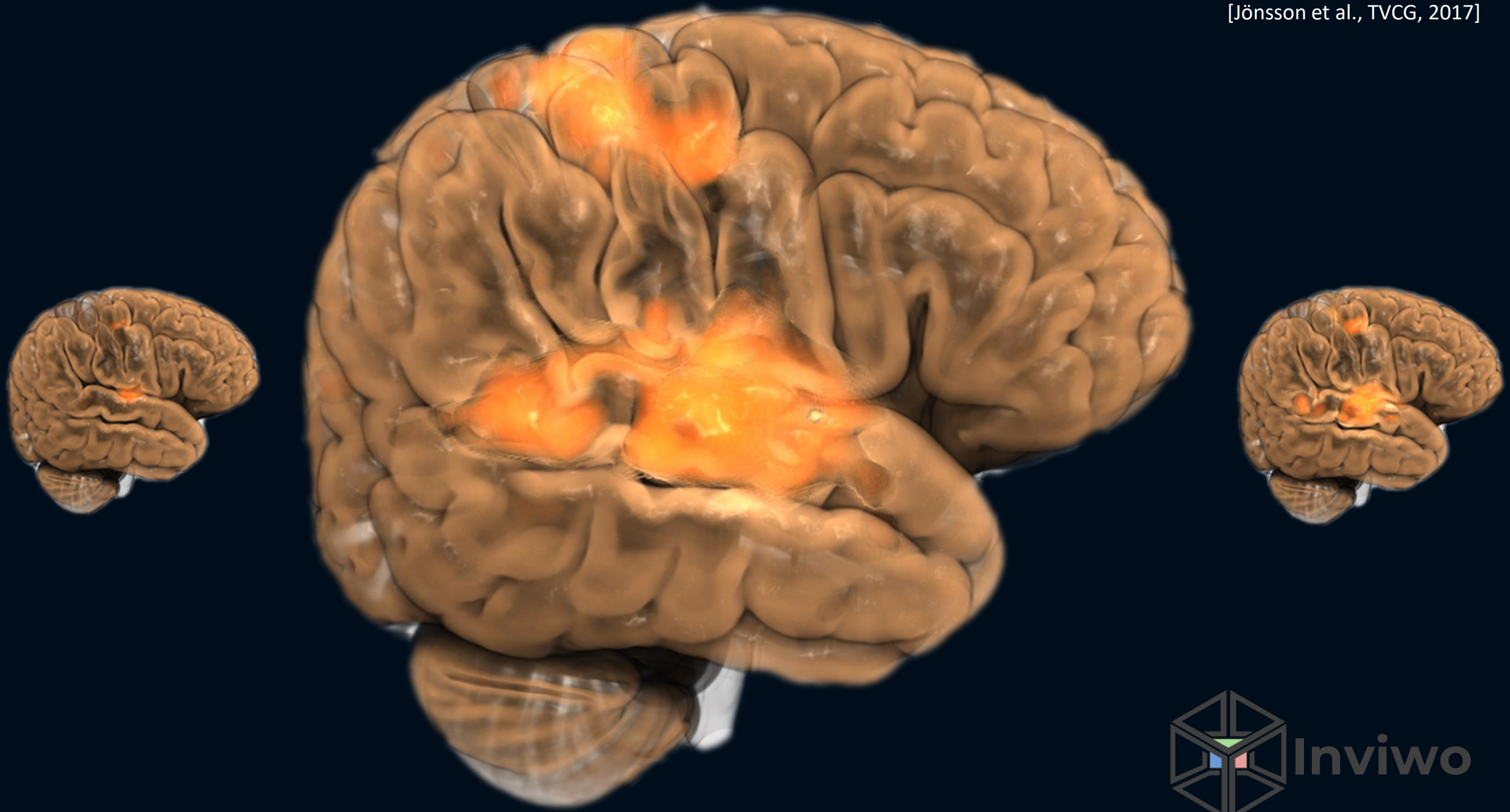

clamp

Depth cue through lighting

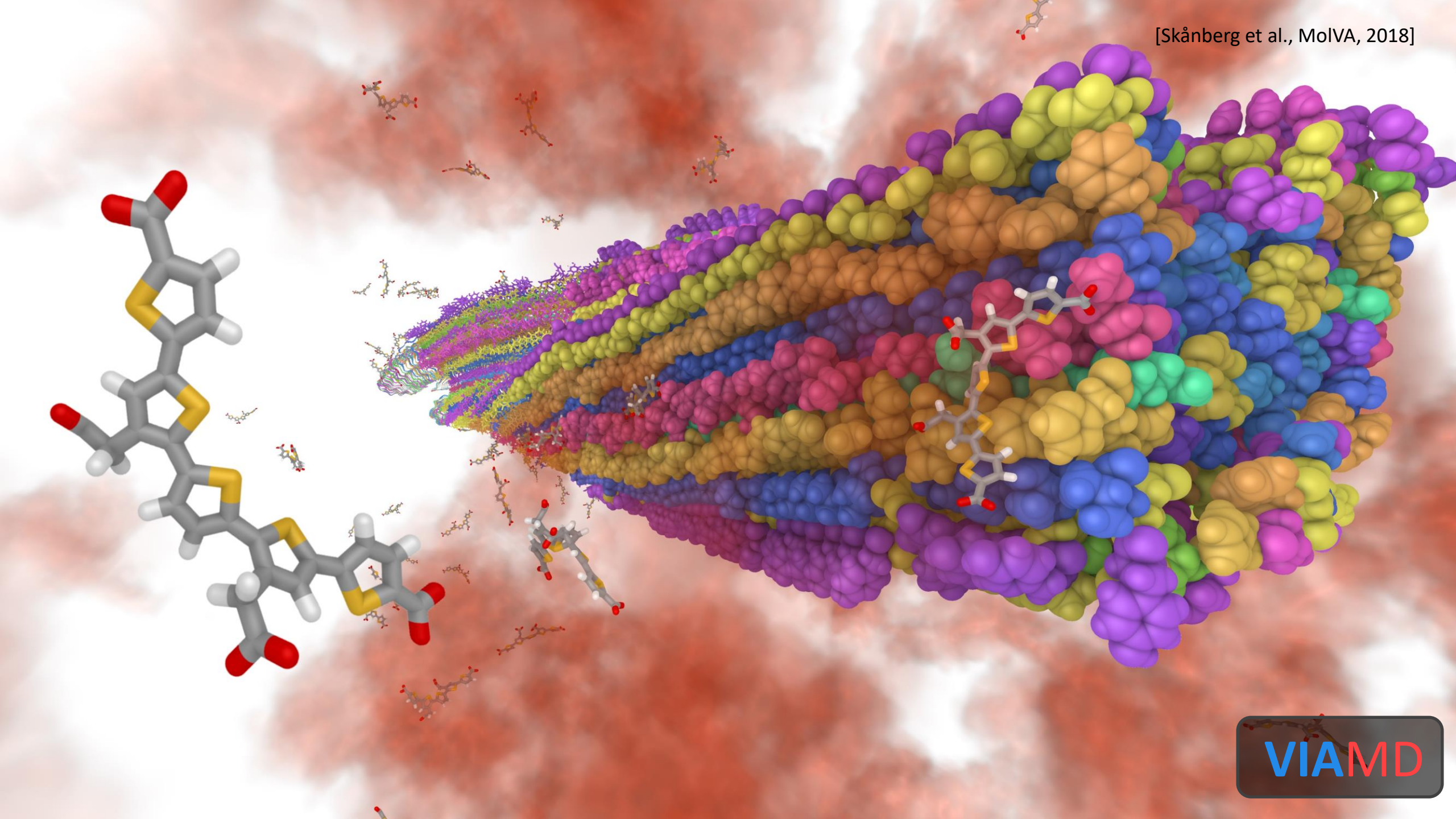




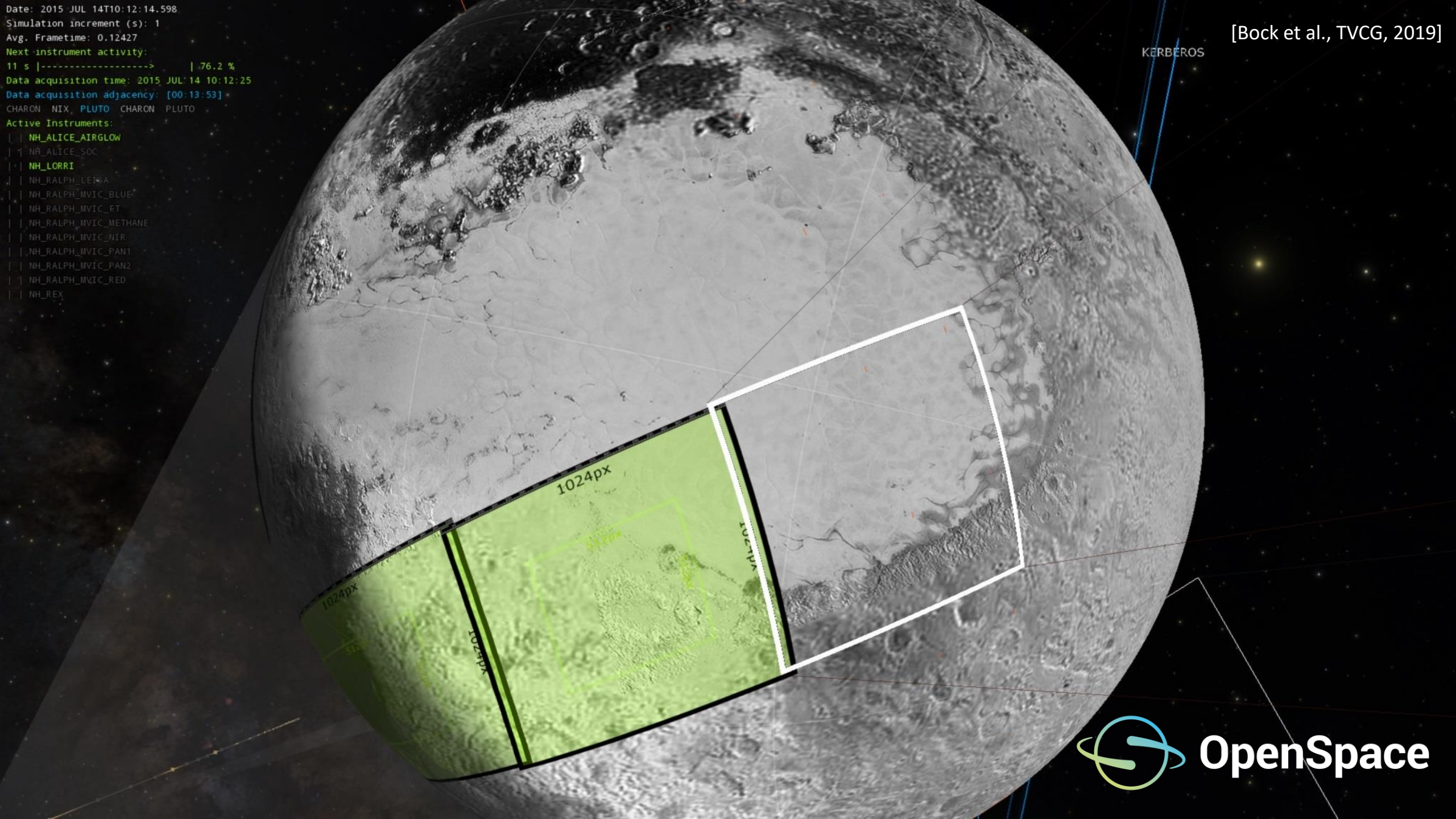
[Jönsson et al., TVCG, 2017]



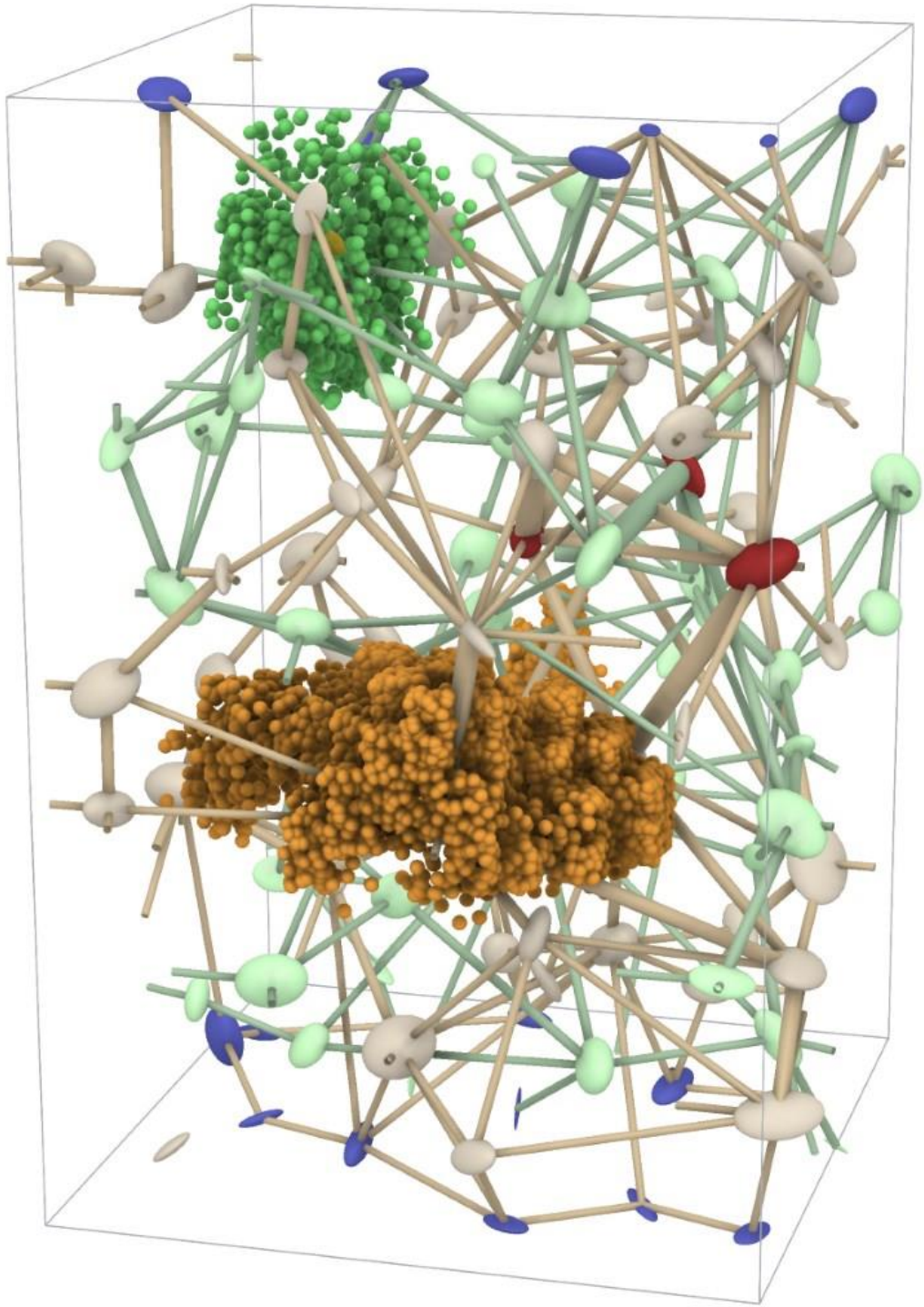
Inviwo

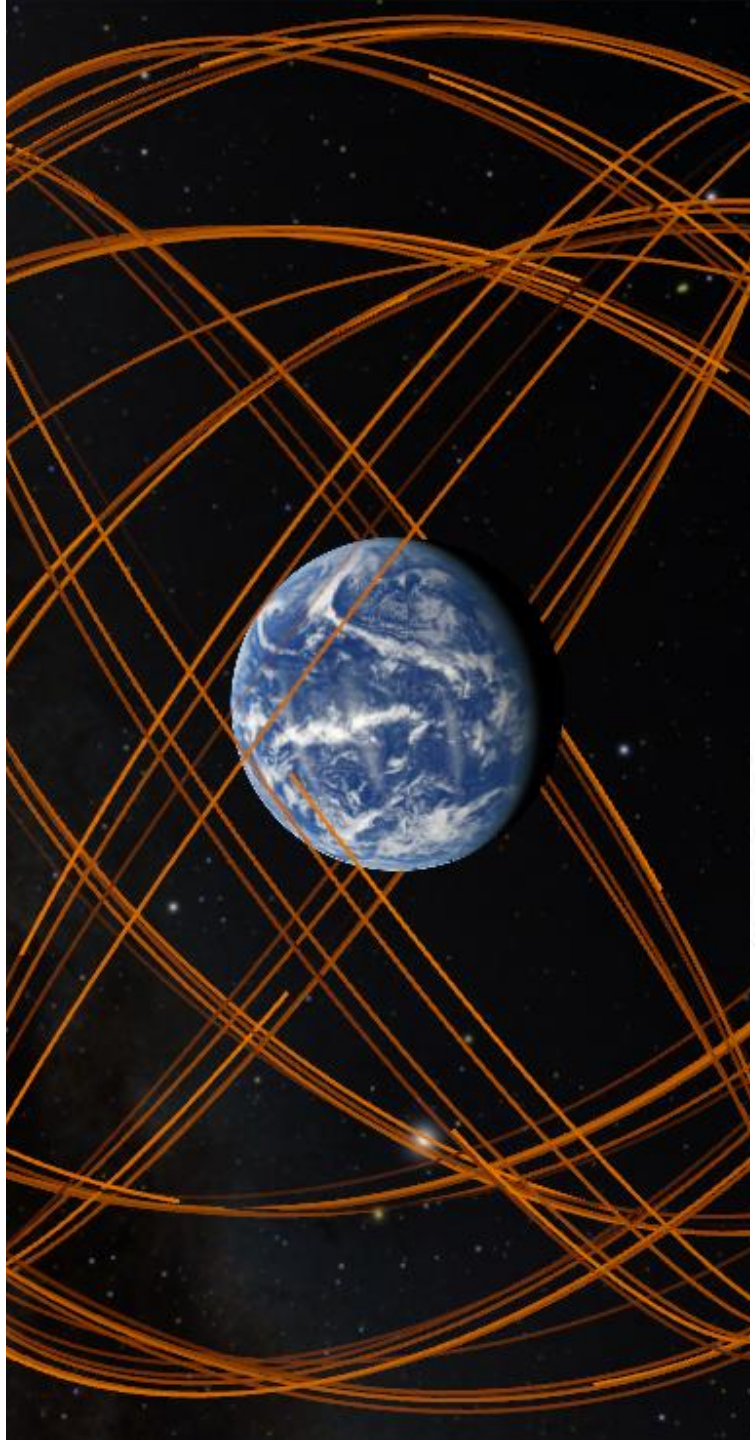

[Skånberg et al., MolVA, 2018]
VIAMD

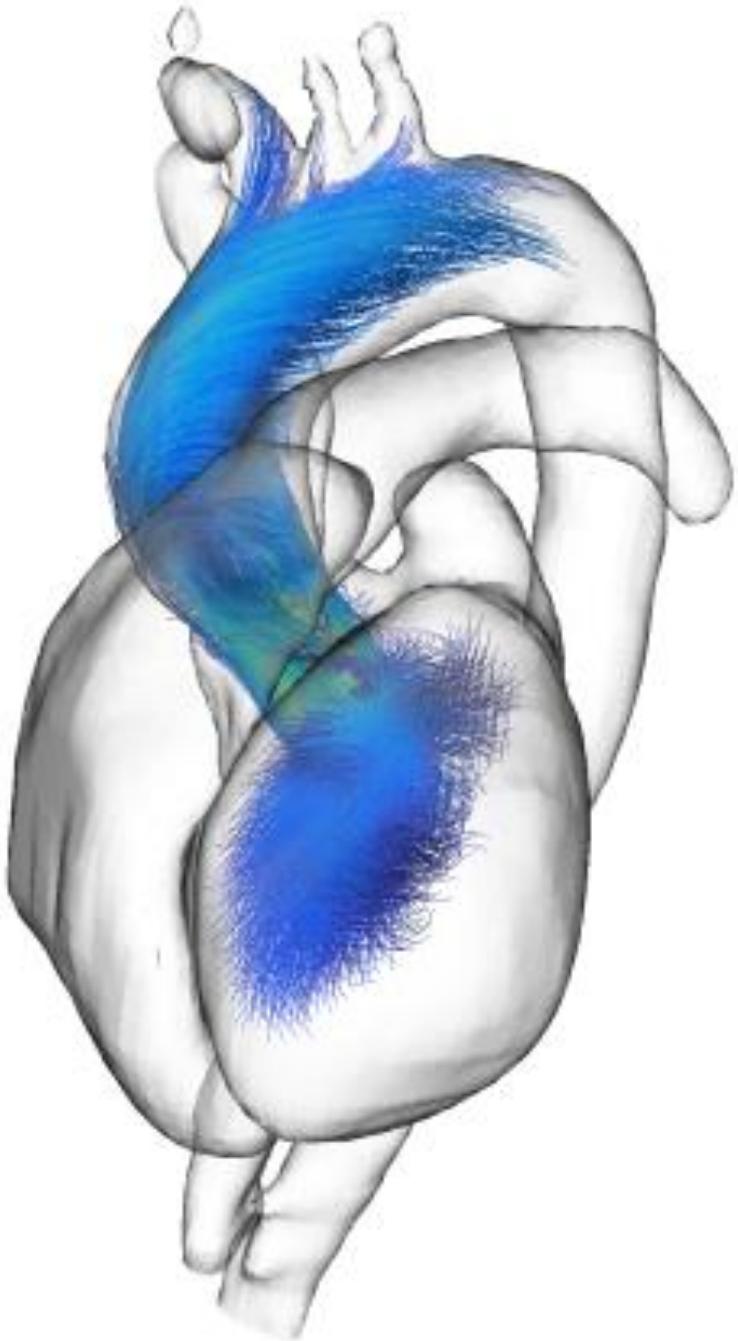
[Engelke et al., CGF, 2019]



[Bock et al., TVCG, 2019]







Use cases

Inviwo

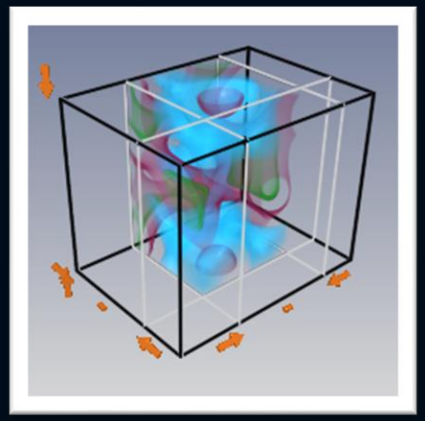
Interactive Visualization Workshop

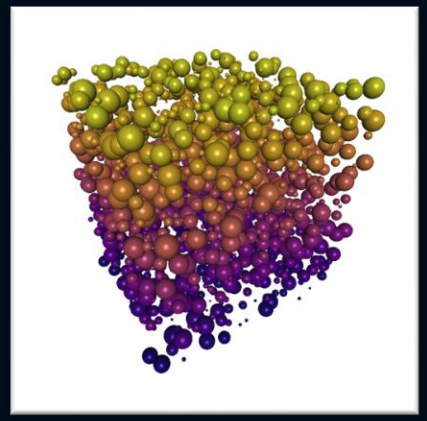
Rapid prototyping

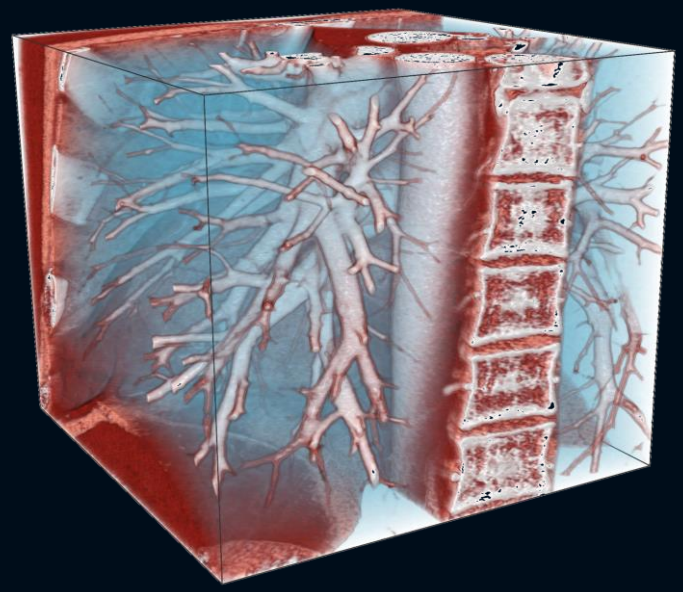
Developed in-house (LiU)

Available on github.com/inviwo

inviwo.org







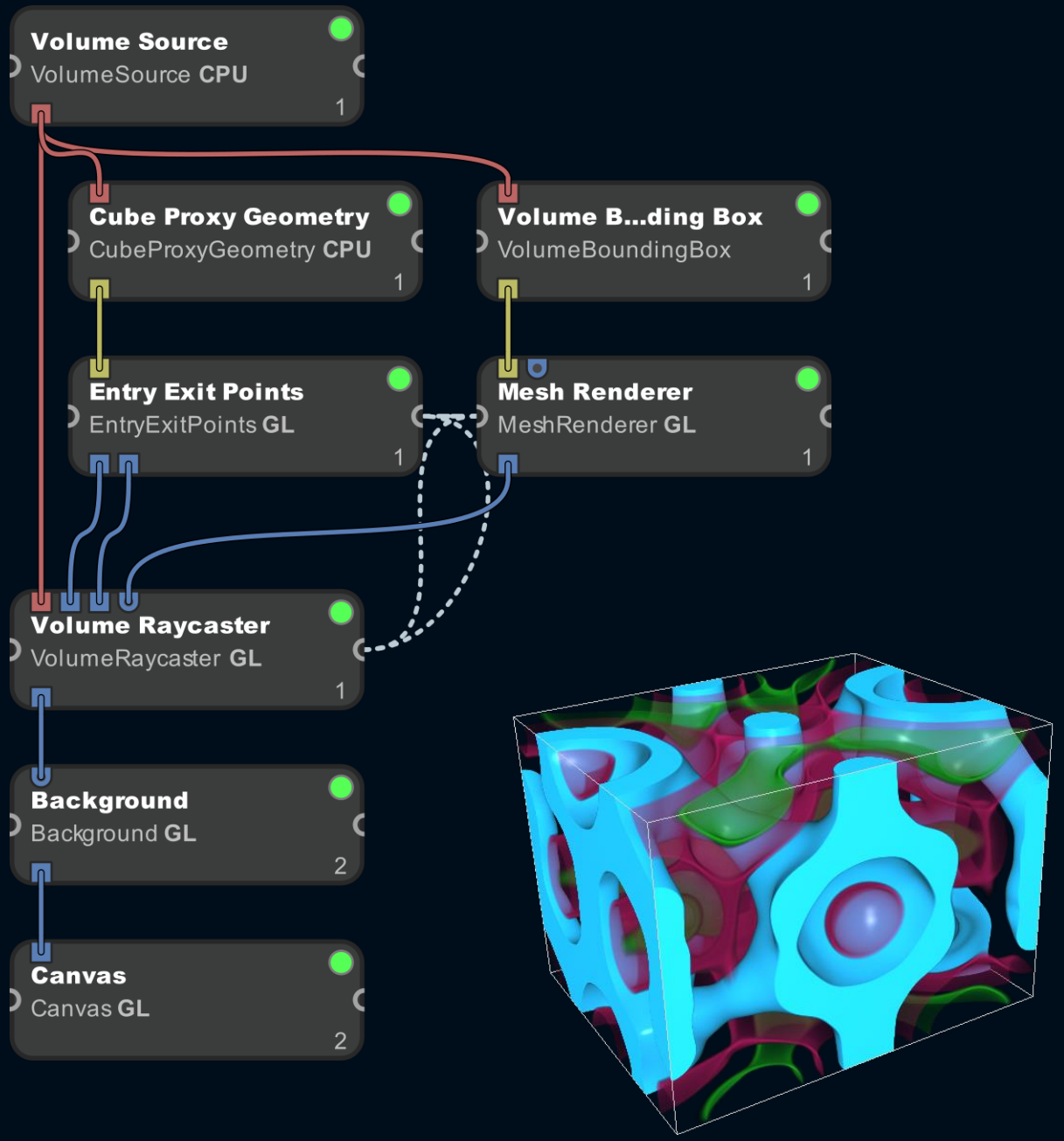
Inviwo

Models a visualization pipeline

- Data flows through graph

Result at the bottom

Blocks can be linked



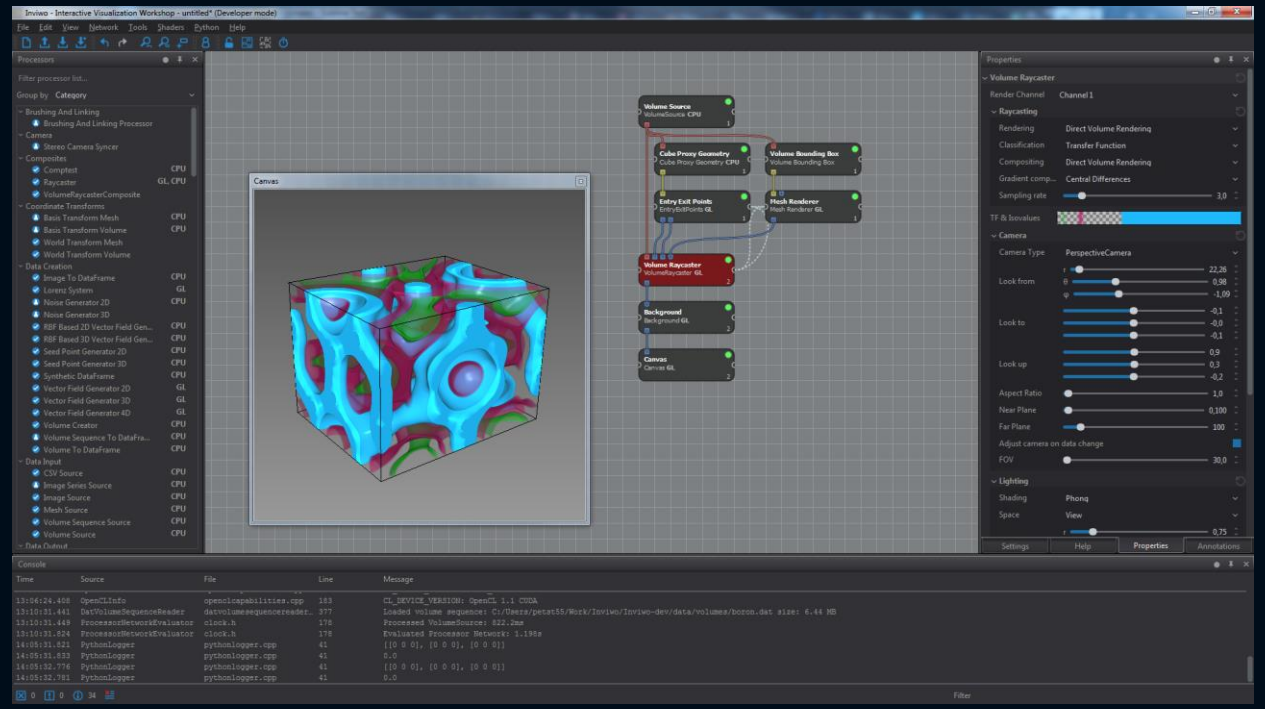
Visualizing data with Inviwo

Build visualization networks

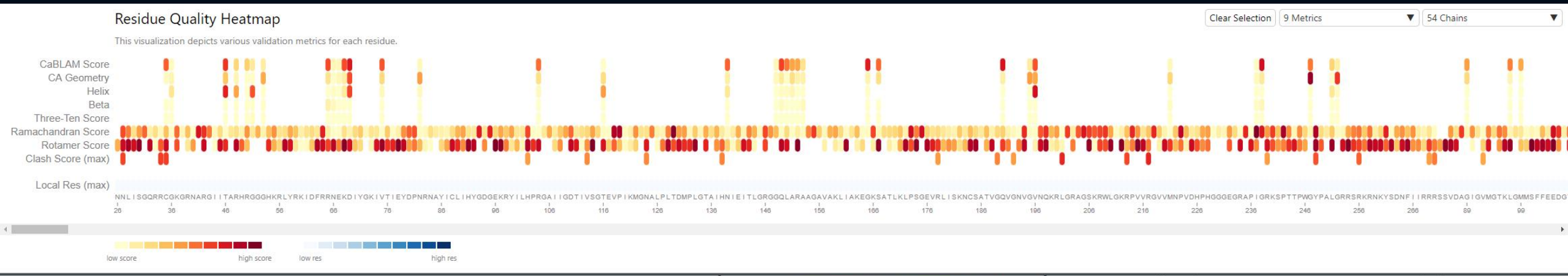
Add your own Python processor

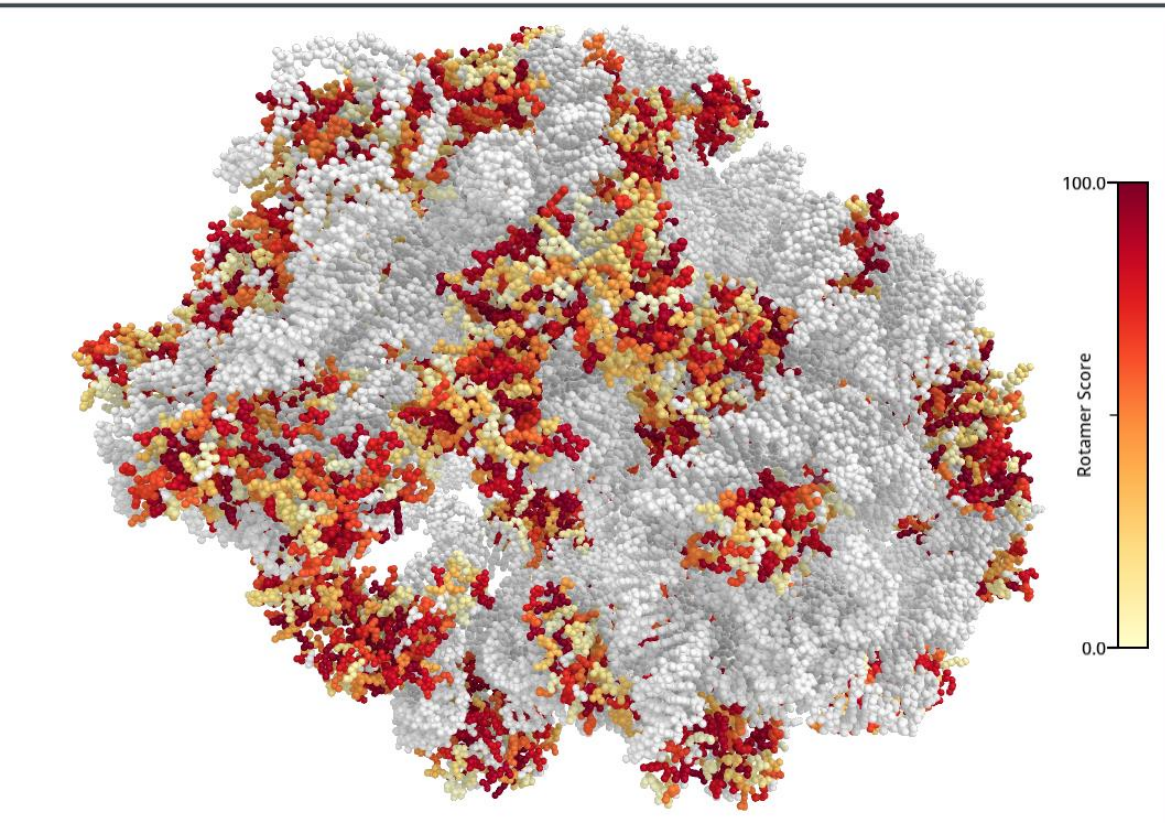
Add a C++ processor

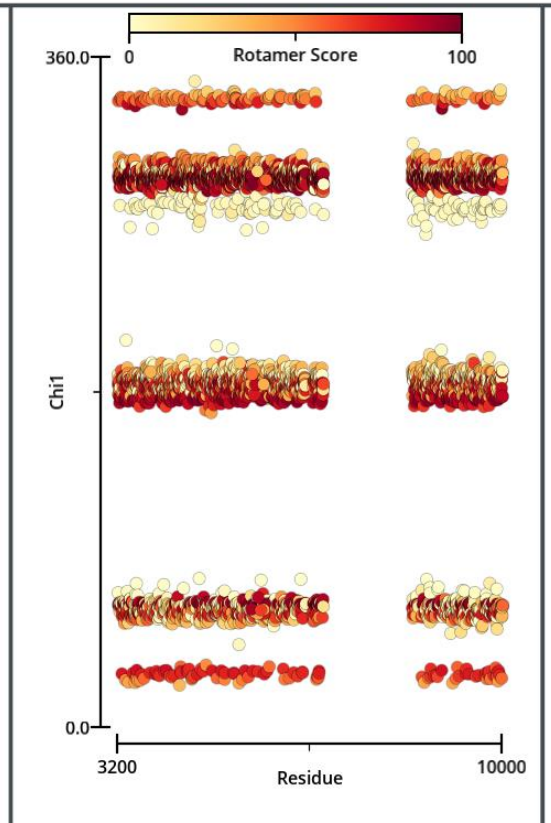
Include web pages via HTML5/WebGL

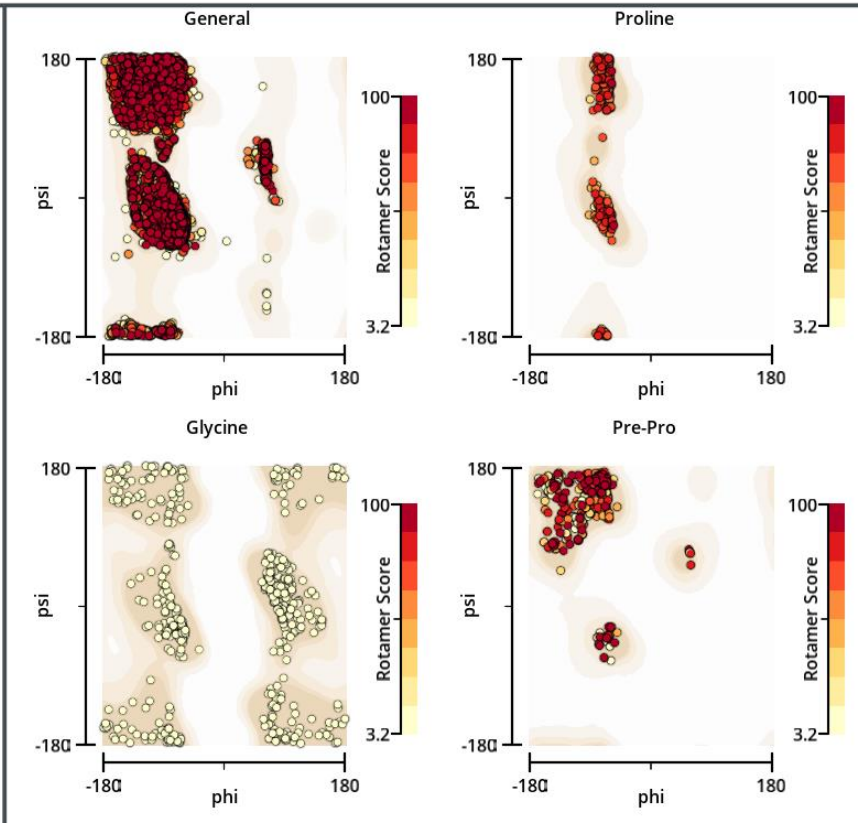


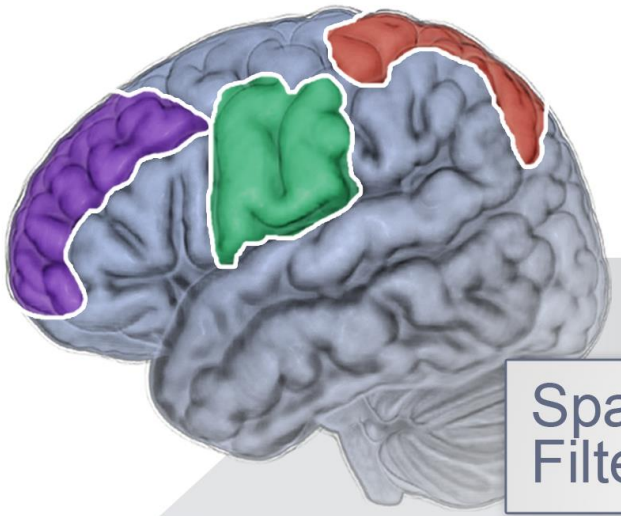
Cryo-EM map validation





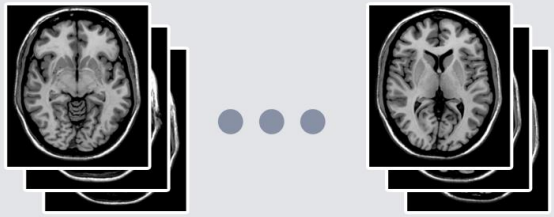






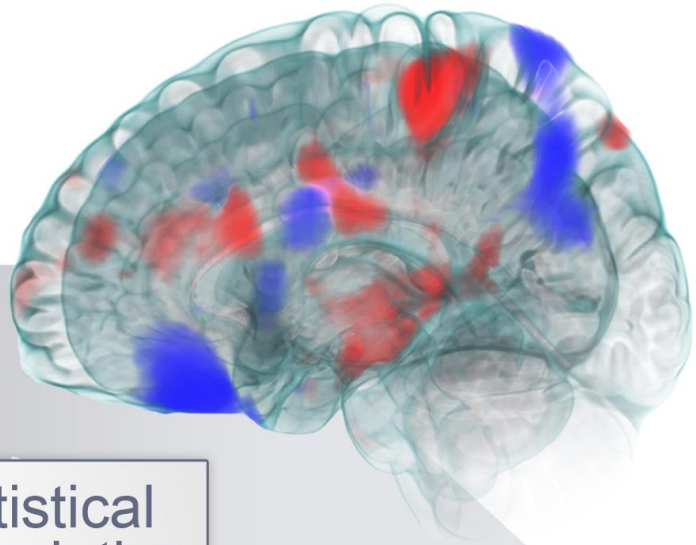
Spatial Filtering

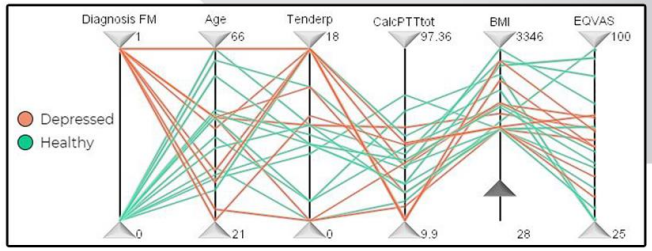
Multivariate Filtering



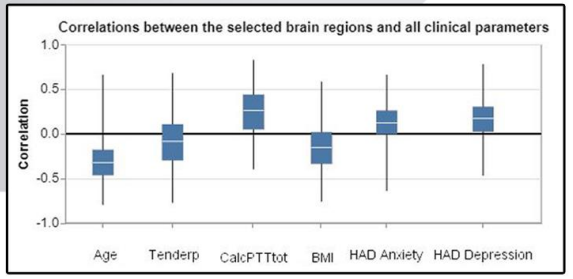
Hypothesis Formation and Reasoning

Statistical Correlation





	D	E	F	G	H	I	J	K	L	M	N
1	BLTSYS	BLTDIAS	TENDERP	PVA1	PVA2	VAS1	VAS2	PTL1	PTL2	TVA1	TVA2
2	130	85	0	17	14.44	10	10	10	9.99	170	144.68
3	90	55	0	96.54	21.1	10.02	10	9.99	9.99	337	211
4	105	75	0	37.23	10.12	8.08	10.02	8.06	9.99	81.91	101.18
5	105	65	0	6.08	35.92	10.02	10.01	10.01	9.99	60.77	135.31
6	145	80	0	25.82	26.68	10	10	9.97	10	88.7	107.76
7	115	90	0	64.72	36.79	10.01	10.01	10.01	10	255.62	158.94
8	130	90	0	17.4	22.34	10	10	10	9.99	86.71	78.39
9	140	80	0	54.79	12.67	10	10	9.99	9.99	185.07	127.09
10	145	90	0	9.09	47.47	9.97	9.97	9.78	9.92	90.71	140.51

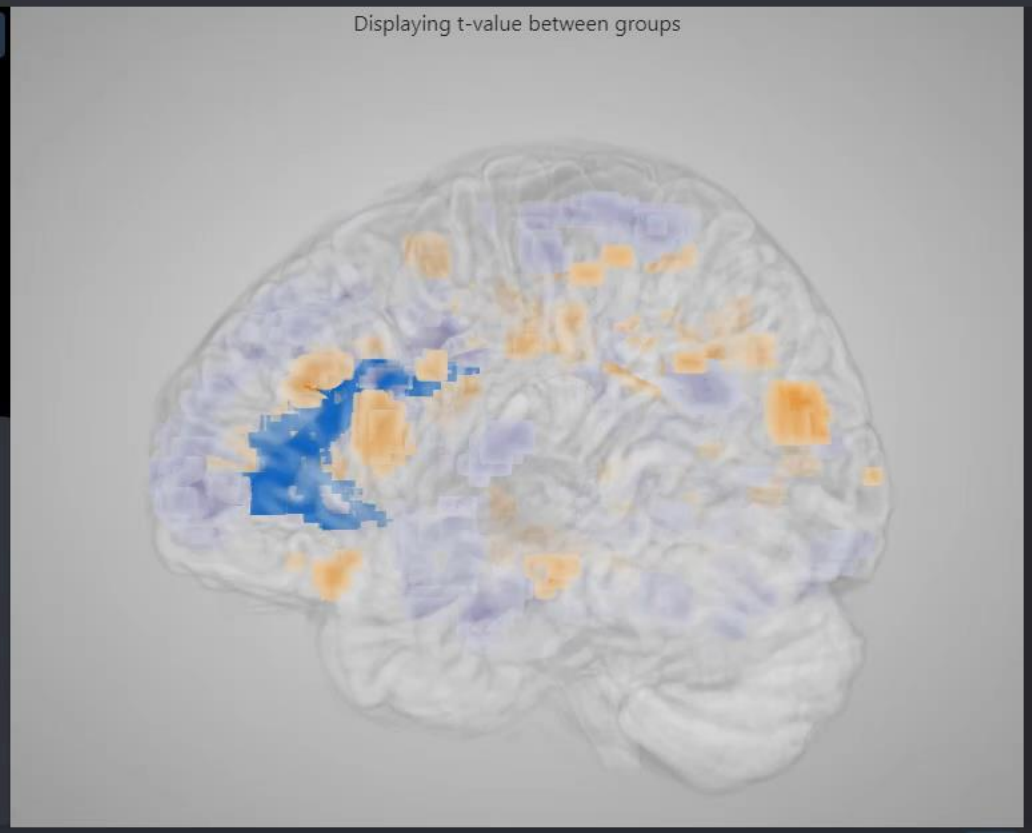


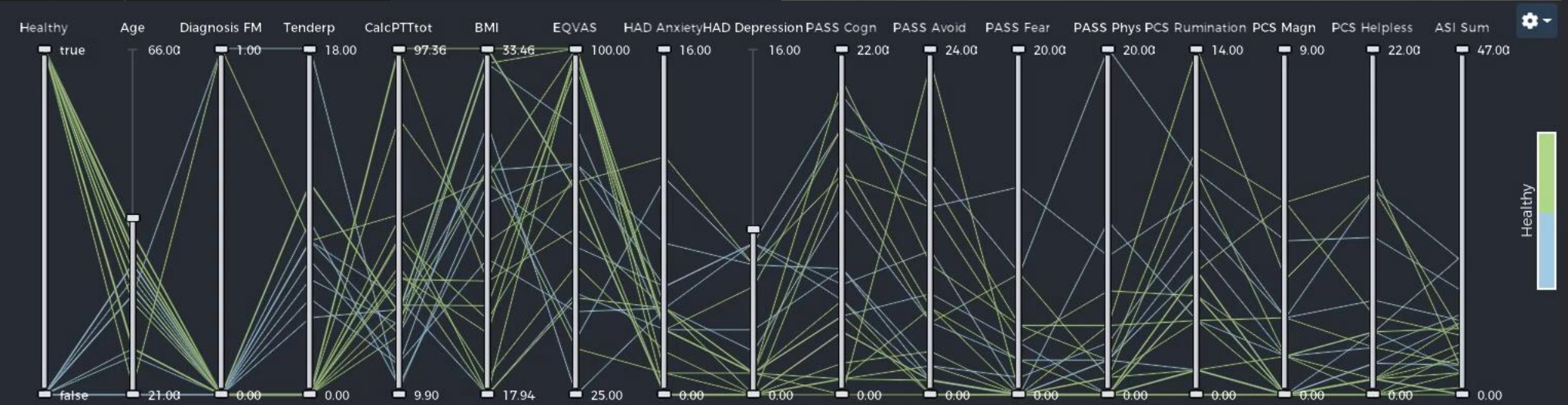
Coordinate: *-4.0, 31.2, 6.6*
Value: *0.00*

Brain atlas regions

Broadmann: *Corpus Callosum*

AAL:





Group selection

A B

Reset selection

Comparison

Show: Group t-Test

p-Value: 0.05

Visual properties

Threshold: 0.00

Opacity: 0.50

Camera: Side Front Top

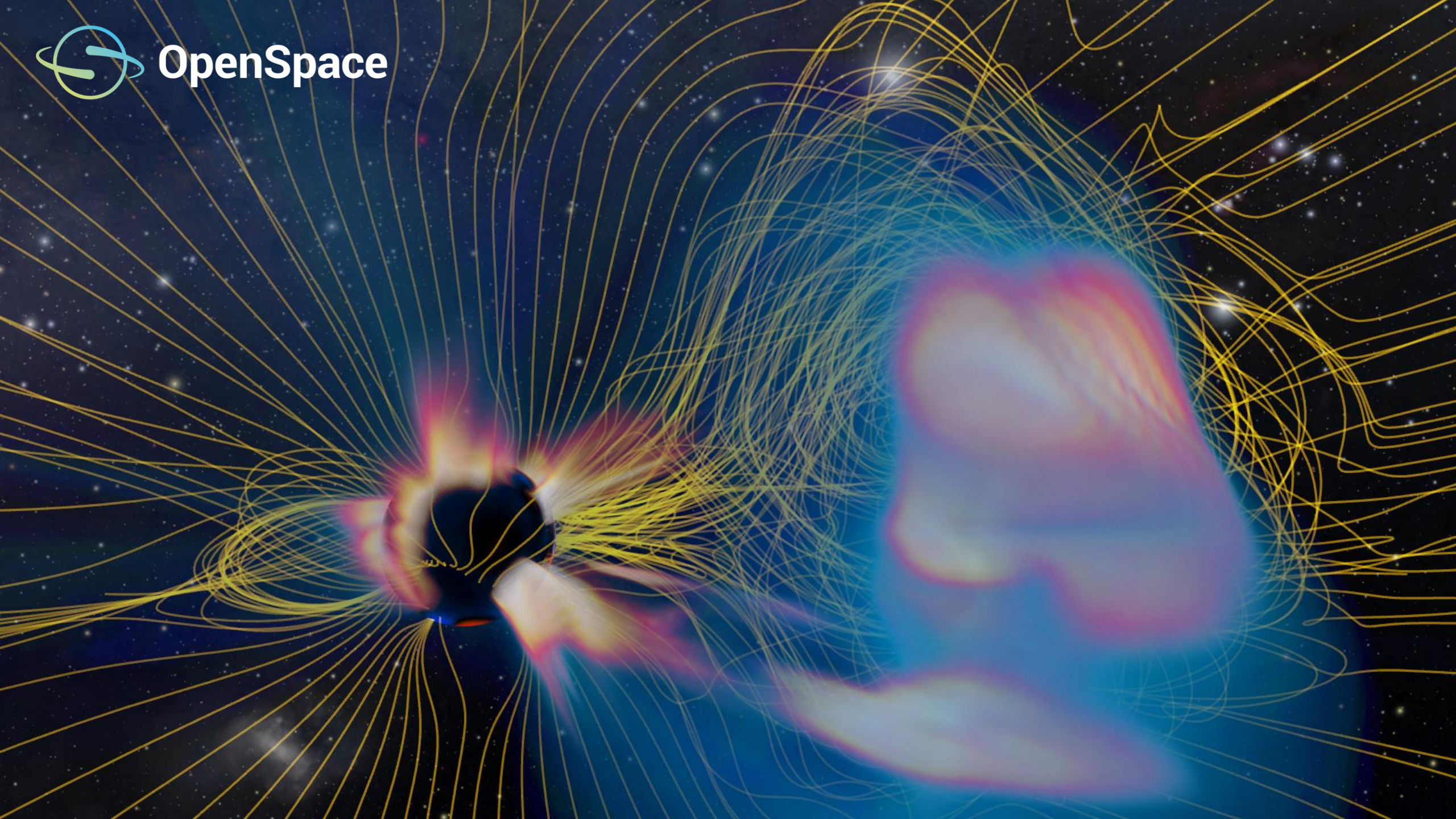
Brain atlas

Show atlas

Id	Region
Sea	left anterior cin
31	Left anterior cingulate gyrus

Region analysis

Deselect all


OpenSpace

Conclusion

12 Basic Principles Of Data Visualization

Keep your Audience in Mind

Choose the Chart Wisely

Think Beyond the PowerPoint Templates

Form follows Function

Direct Attention to the Important Details

Use Tables and Graphs Appropriately

Provide Context

Align the Data and the Displays Right

Choose the Right Colors

Pay Careful Attention to Titles

Use Clear Axis Labels and Numbers

Leverage Interactivity When Appropriate

[dataflog.com]

Tools & Libraries

General visualization

- Libraries:
  - Vega, plotly (high-level)
  - matplotlib, d3.js (low-level)
  - Voyager
- Matlab, Excel
- Tableau, Spotfire, MS Power BI, …

Spatial Data

- Inviwo, inviwo.org
- OpenSpace, openspaceproject.com
- VTK / ParaView, paraview.org

Links

Colormaps colorbrewer2.org

Interactive vis with explanations svt.se/datajournalistik

Examples of visual misinformation vislies.org

More examples visme.co, insights.datylon.com/stories

Data visualization cheat sheet policyviz.com

More inspirations informationisbeautiful.net