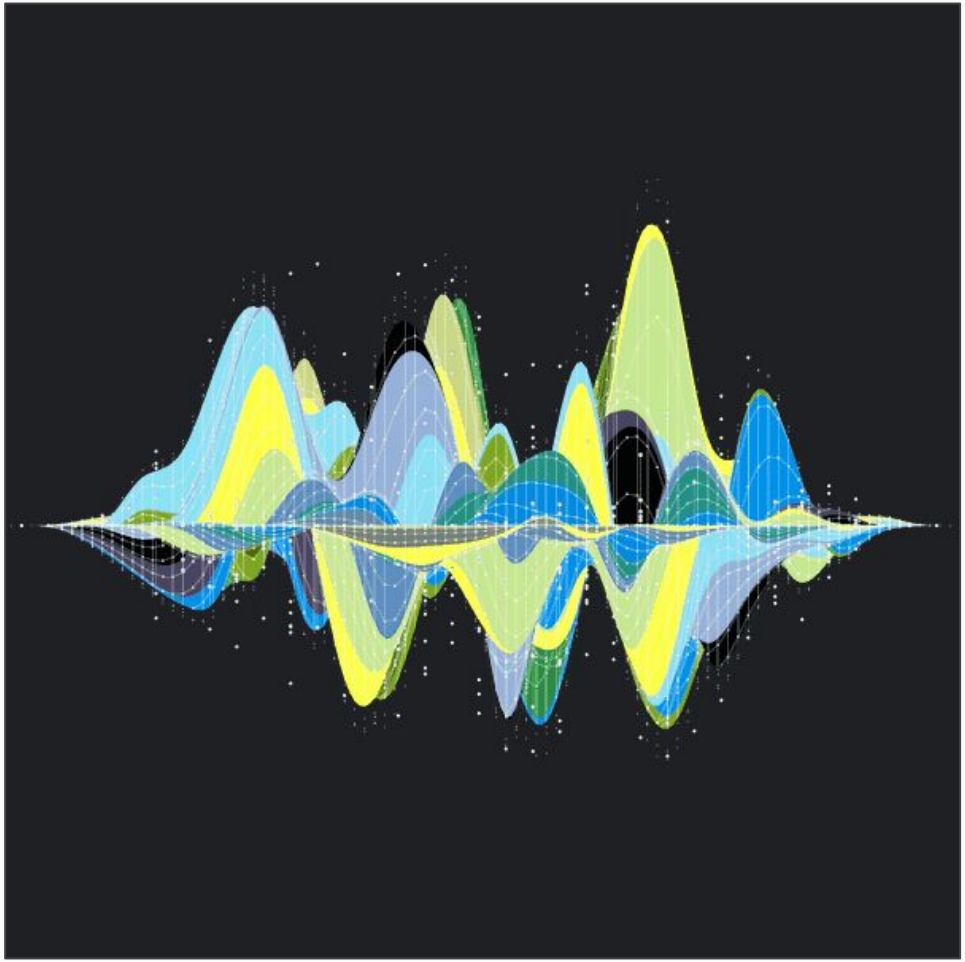
APPLIED DATA SCIENCE I

Data Visualization Theory

Kyle Scot Shank, '14
FA-21

WHAT IS DATA VISUALIZATION?



THIS IS THE VERY SIMPLISTIC VIEW

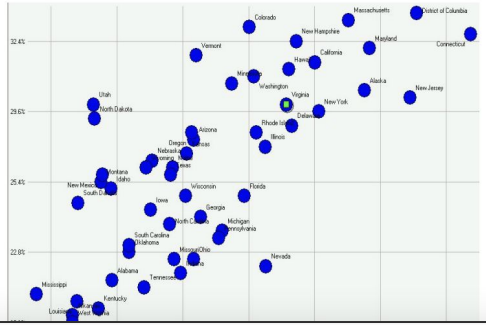
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box		7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	[illegible]	5-Low	Wrap Bag	0.56	1/20/05
69	[illegible]5	4-Not Specified	Small Pack	0.44	6/6/05
69	[illegible]5	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06

attribute

item

cell

Data → Visualization → Information → Action



ENCODING
(DESIGNER)

01 DATA
What is my data?

02 VISUAL MAPPING
What is my visual representation?

03 RENDERING
What is my medium?

VIEW

DECODING
(USER)

04 PERCEIVING
What does it show?

05 INTERPRETING
What does it mean?

06 COMPREHENDING
What does it mean for me?

THIS IS A BETTER: THE DESIGNER /USER FLOW

ENCODING
(DESIGNER)

01 **DATA**
What is my data?

02 **VISUAL MAPPING**
What is my visual representation?

03 **RENDERING**
What is my medium?

VIEW

04 **PERCEIVING**
What does it show?

05 **INTERPRETING**
What does it mean?

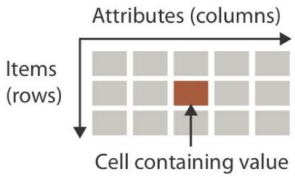
06 **COMPREHENDING**
What does it mean for me?

DECODING
(USER)

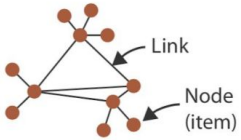
DATA

Dataset Types

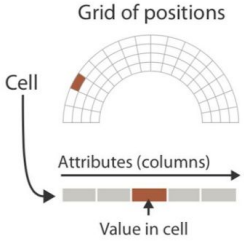
➔ Tables



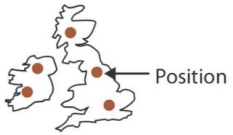
➔ Networks



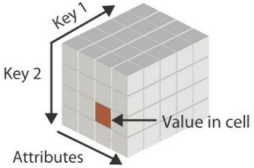
➔ Fields (Continuous)



➔ Geometry (Spatial)

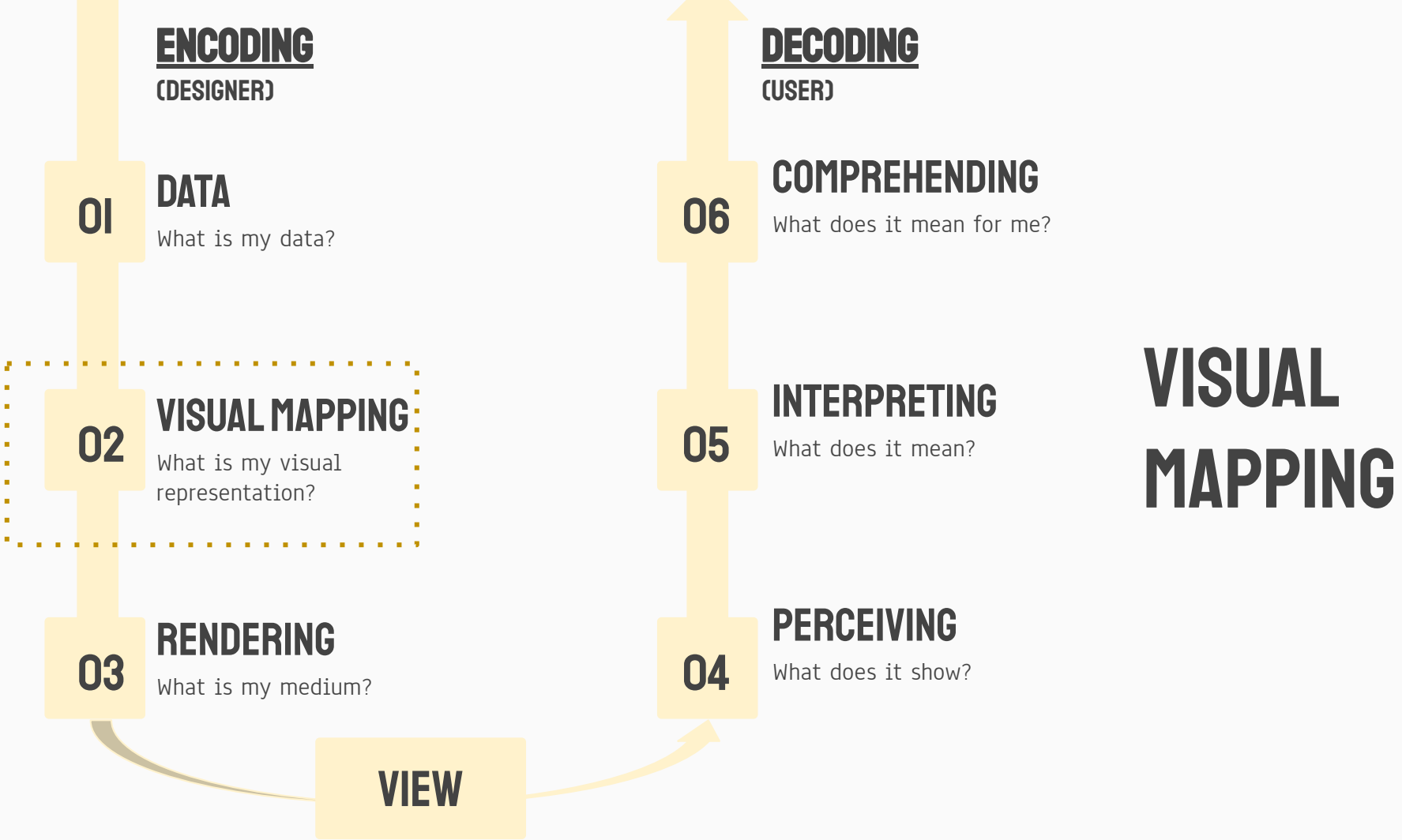


➔ *Multidimensional Table*

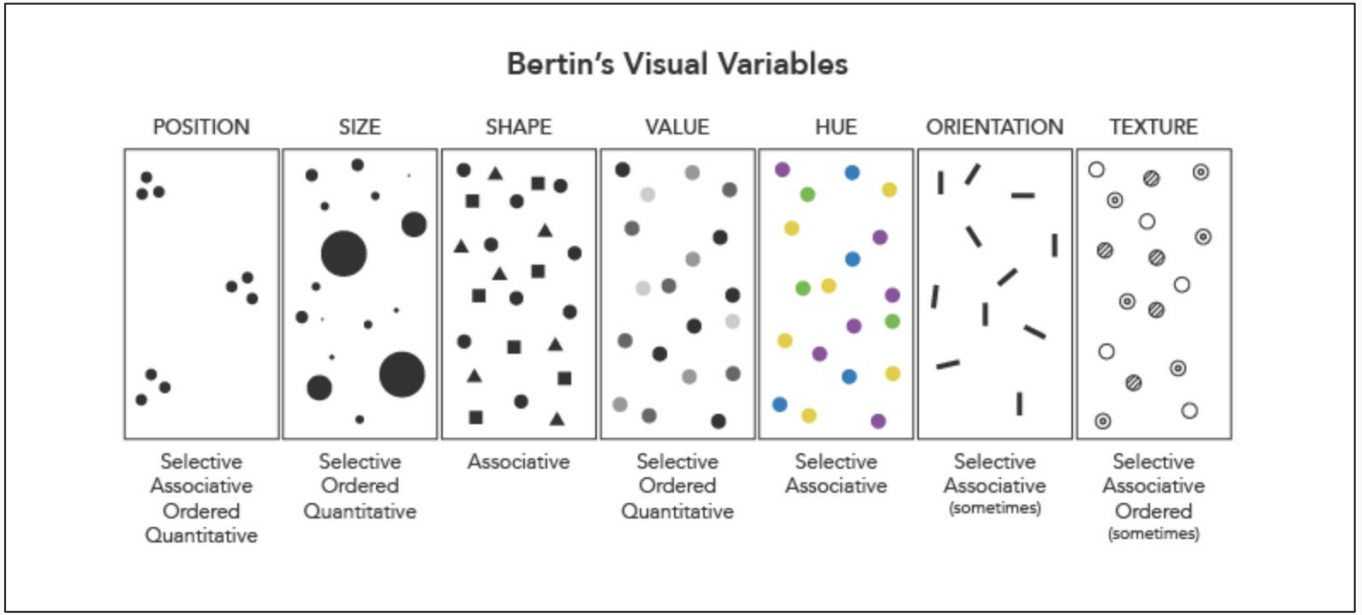


➔ *Trees*

THE STRUCTURE OF YOUR DATA WILL QUITE OFTEN DICTATE THE BEST AND MOST REPRESENTATIVE FORM

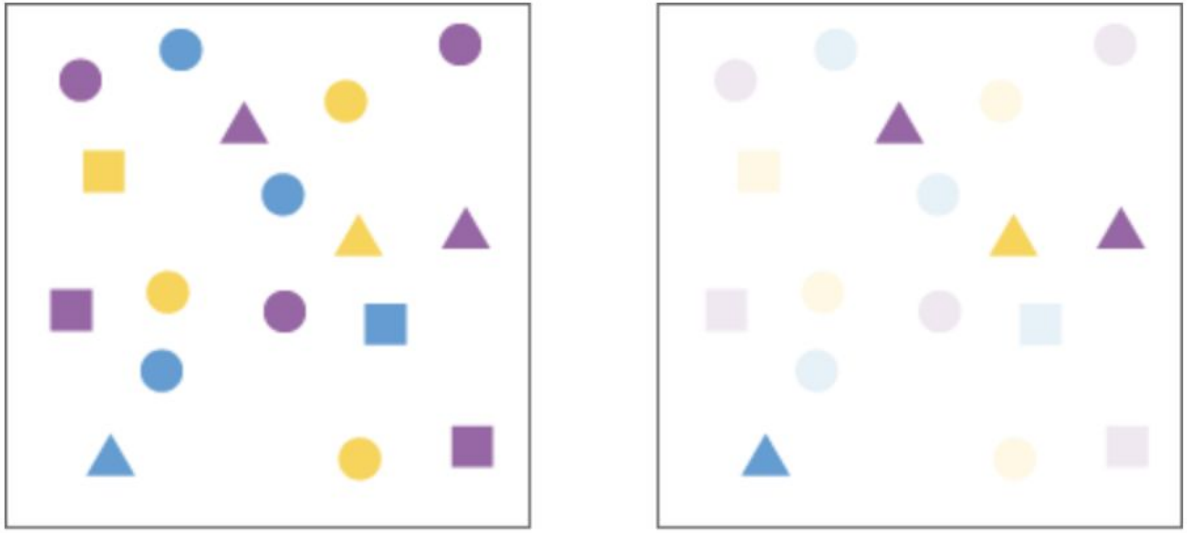

ENCODING
(DESIGNER)
01
DATA
What is my data?
02
VISUAL MAPPING
What is my visual representation?
03
RENDERING
What is my medium?
VIEW
04
PERCEIVING
What does it show?
05
INTERPRETING
What does it mean?
06
COMPREHENDING
What does it mean for me?
DECODING
(USER)
VISUAL MAPPING

JACQUES BERTIN, IN HIS *SEMIOLOGY OF GRAPHICS* (1967), PROPOSED A SET OF "RETINAL VARIABLES" THAT CAN ENCODE INFORMATION EFFECTIVELY FOR THE HUMAN EYE



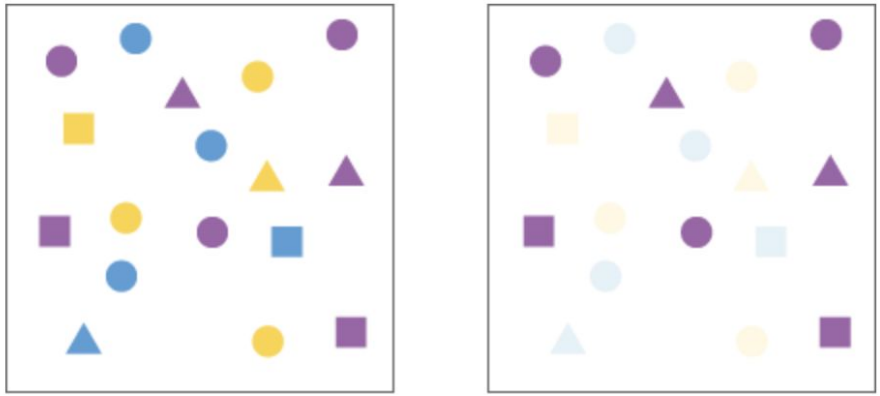
A **selective** variable allows us to immediately isolate a group of signs based on a change in the variable.

We can quickly and easily perceive a group of symbols based on color hue, e.g., the yellow symbols appear as a group. Therefore hue is ***selective***.

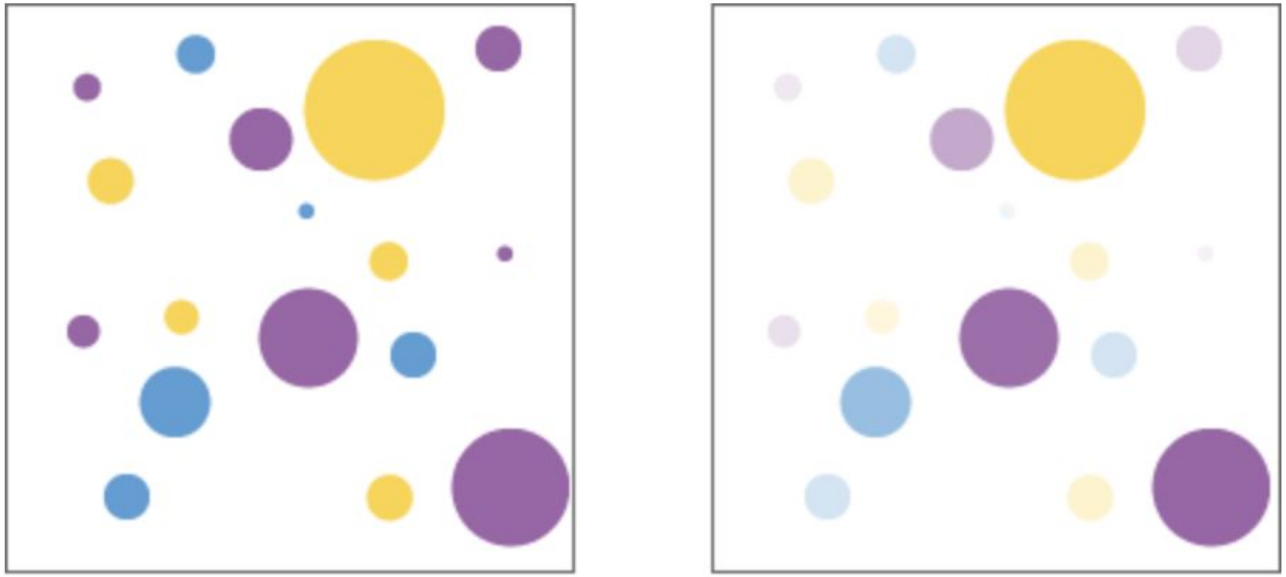


It is not easy to isolate symbols based on shape. Even with all the triangles pulled out at right, they still don't look like a group. Shape is not selective.

An **associative** variable allows grouping across changes in the variable; in other words, a variable is associative if we can perceive symbols as a group despite differences in this variable. Note that this does not mean the variable is not selective. For example, hue is both: we can easily select symbols of the same color, but a group of differently colored symbols is itself also perceptible as a group.

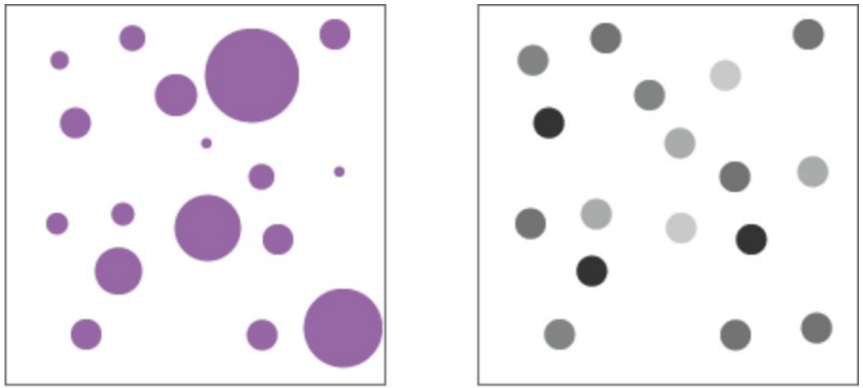


Symbols can be perceived as a group despite variation in shape, e.g., these similarly sized purple symbols appear as a group. Shape is therefore *associative*.



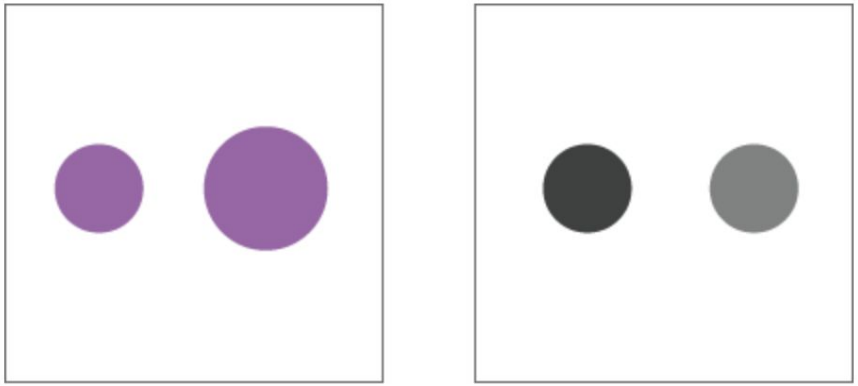
Size variation affects the visibility of symbols and is impossible to ignore. We see larger symbols first, not a single group of circles. Size is *dissociative*.

Ordered variables have an immediately recognizable sequence. In other words, if you can order the symbols by their difference in a variable without having to consult a legend, the variable is ordered. Size and value, for example, have an immediately perceptible order. Position does too: symbols can naturally be seen as ordered from left to right or vice versa, for example. Color hue is an example of a non-ordered variable: there is no clear ordering of, say, red, green, and blue.



With both size and value, it is immediately obvious that there is some sequence to the symbols (small to large, light to dark). Size and value or *ordered*.

Quantitative variables allow an estimation of the actual numerical difference between symbols. Besides position, where we can guess the measurable distance between symbols, Bertin considered only size variation to be quantitative. For example, we can easily see that one symbol is twice the size of another—more or less.



We can see that one purple circle is about twice the size of the other, but can't similarly measure a difference in lightness. Size is ***quantitative***.

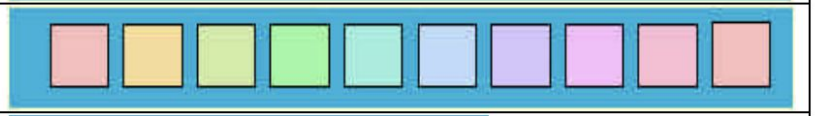
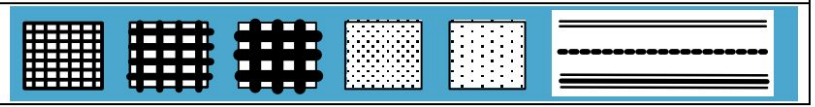
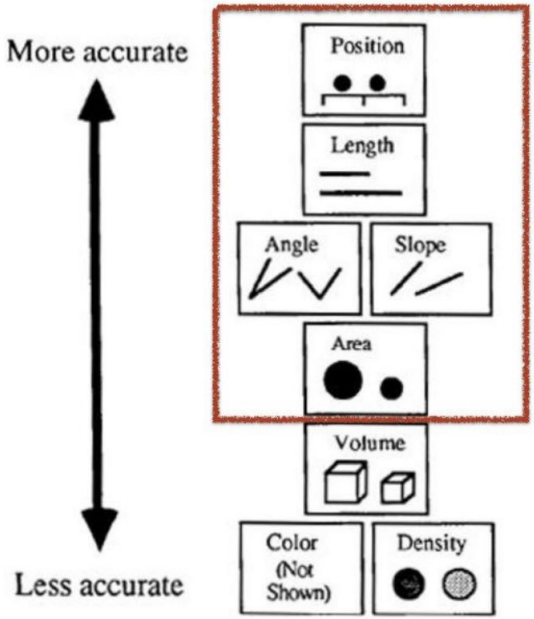
Bertin's Original Visual Variables	
Position changes in the x, y location	
Size change in length, area or repetition	
Shape infinite number of shapes	
Value changes from light to dark	
Colour changes in hue at a given value	
Orientation changes in alignment	
Texture variation in 'grain'	

Table 1: These are Bertin's visual variables



ENCODING
(DESIGNER)

01 DATA
What is my data?

02 VISUAL MAPPING
What is my visual representation?

03 RENDERING
What is my medium?

VIEW

DECODING
(USER)

04 PERCEIVING
What does it show?

05 INTERPRETING
What does it mean?

06 COMPREHENDING
What does it mean for me?

RENDERING IS IMPORTANT BUT A TOPIC FOR ANOTHER DAY!

ENCODING
(DESIGNER)

01 DATA
What is my data?

02 VISUAL MAPPING
What is my visual representation?

03 RENDERING
What is my medium?

VIEW

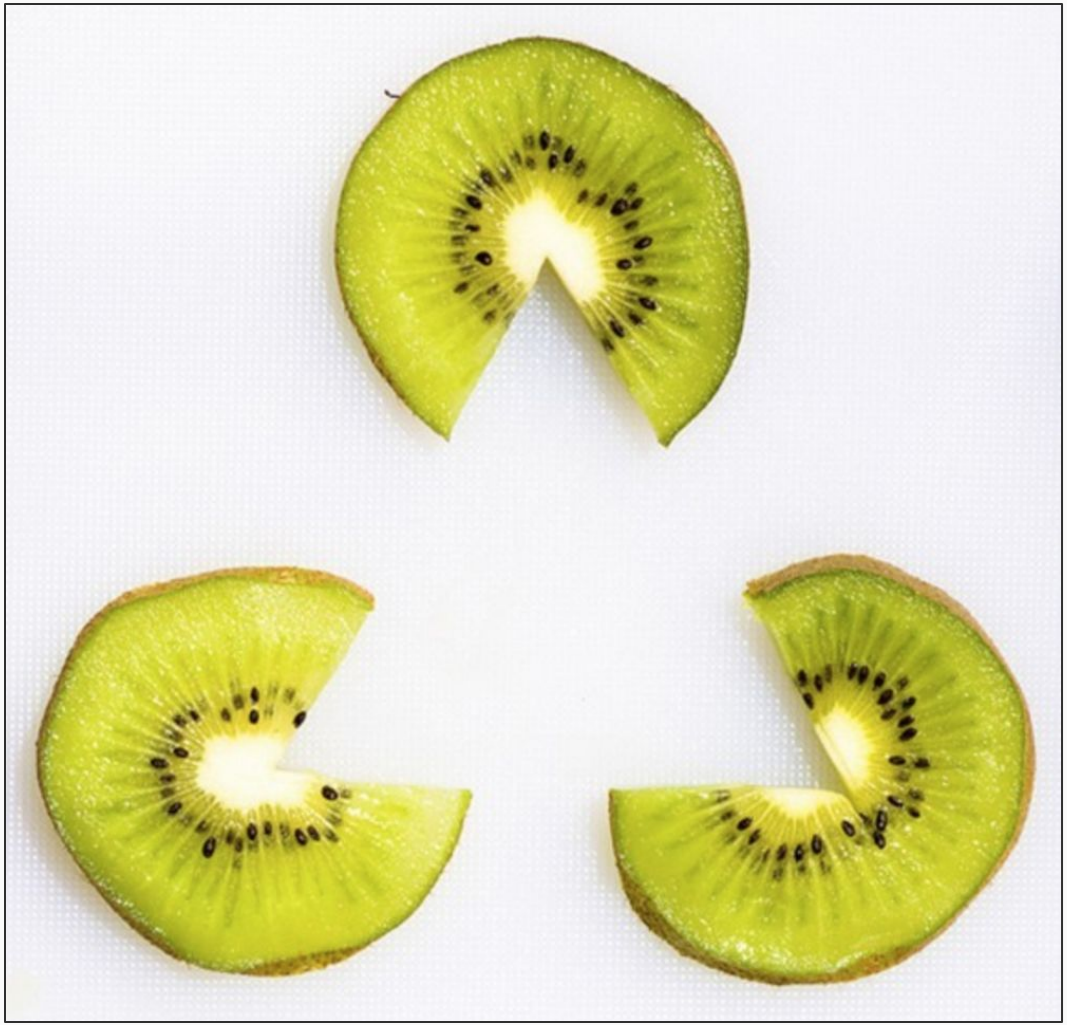
DECODING
(USER)

04 PERCEIVING
What does it show?

05 INTERPRETING
What does it mean?

06 COMPREHENDING
What does it mean for me?

PERCEIVING



WHAT DO YOU SEE?

“Visual analysis appears to be functionally divided between an early preattentive level of processing at which simple features are coded spatially in parallel and a later stage at which focused attention is required to conjoin the separate features into coherent objects”.

(Anne Treisman, Computer Vision, Graphics, and Image Processing)

PRE-ATTENTIVE PROCESSING IS INTIMATELY TIED TO VISUAL MAPPING

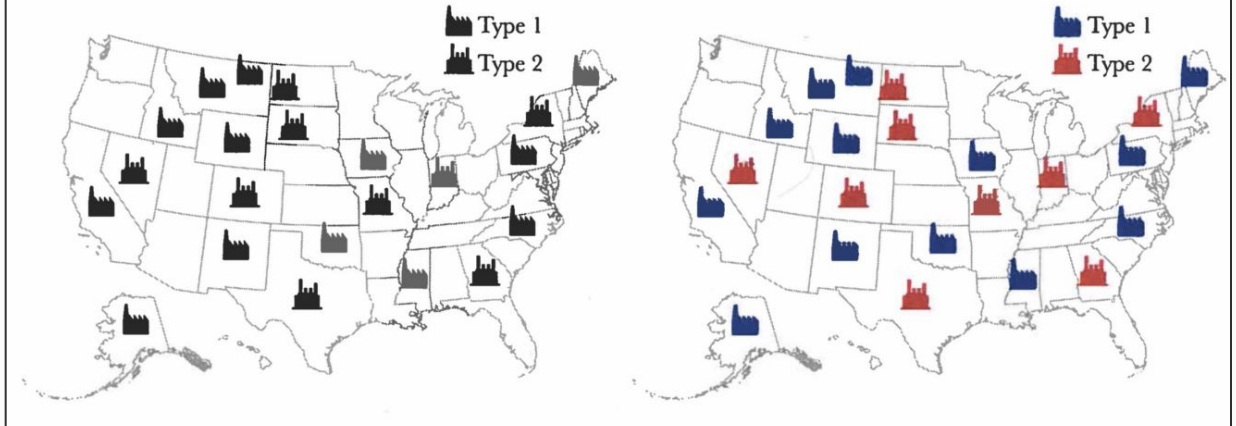
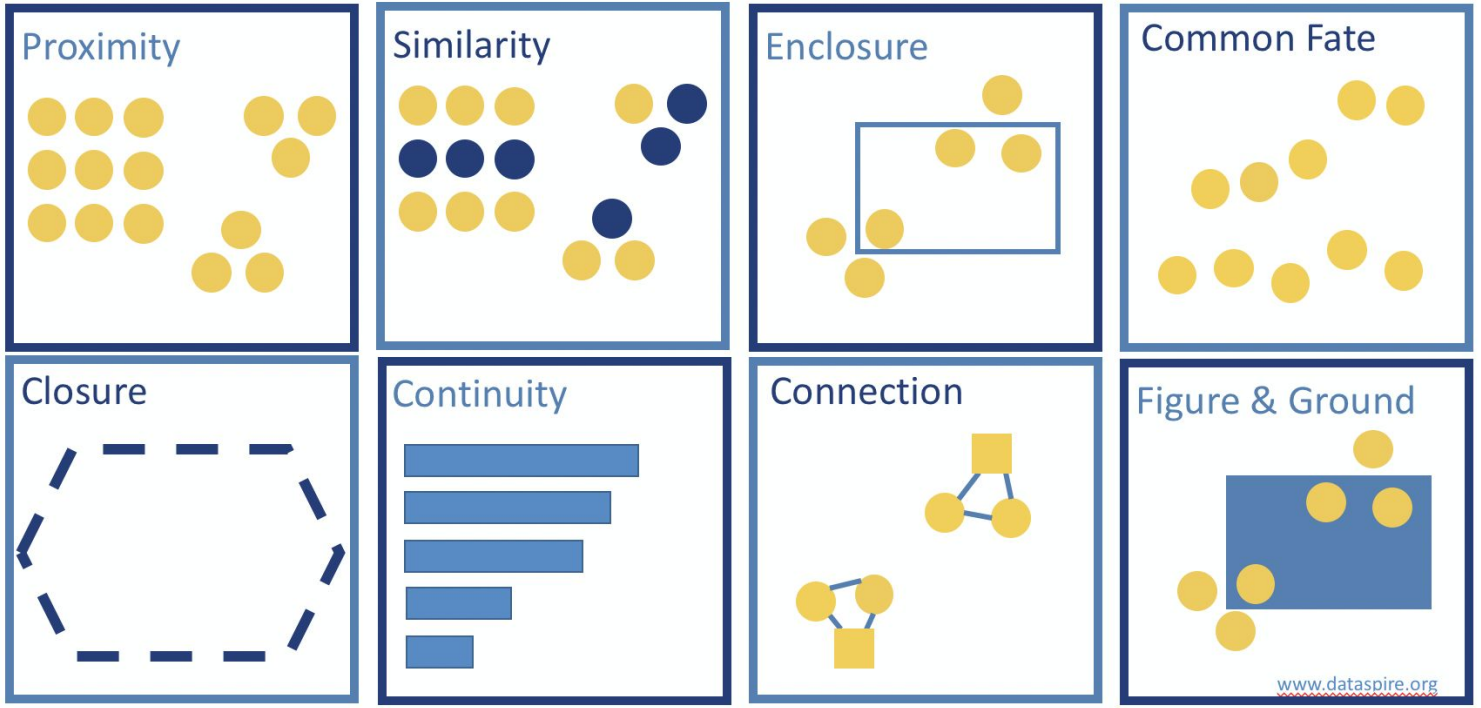


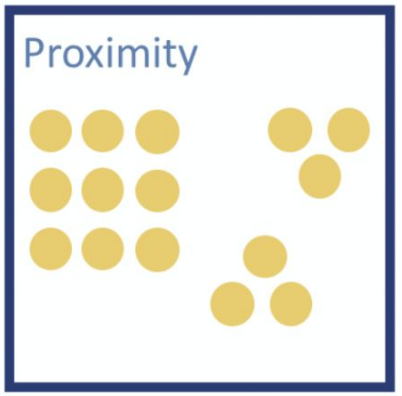
Figure 6.3 On which of these maps is it easier to identify the number of factories of each kind?

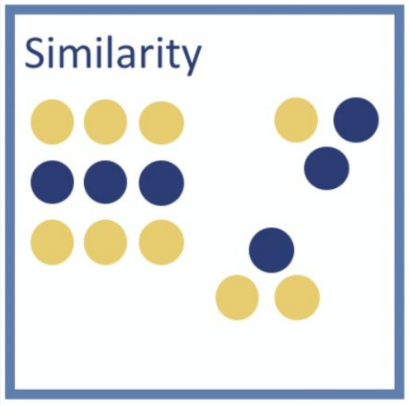
Gestalt's Principles of Visual Perception



PROXIMITY

The **proximity** principle identifies our tendency to perceive objects that are physically close to each other as belonging to part of the same group, whether or not that is true or relevant to the data at hand. For example, our eyes often connect the nine dots in the 3x3 box on the left together, the three dots in the upside-down triangle on the right together, and the three dots in the triangle on the bottom right together. Our eyes group these dots together before we even know what we are looking at in the dots.



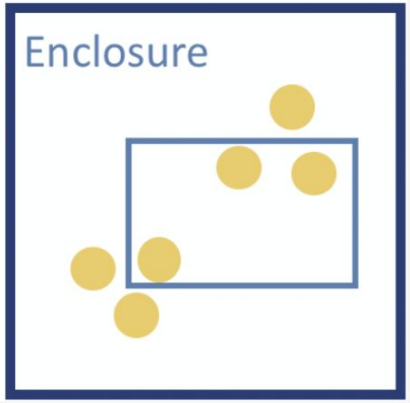


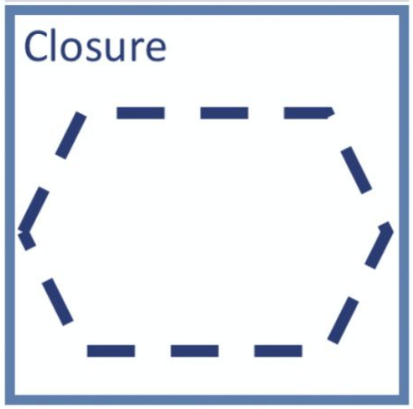
SIMILARITY

The **similarity** principle outlines our tendency to perceive objects of similar color, shape, size, and/or orientation as related or belonging to part of the same group, whether or not that is true or relevant to the data at hand. For example, our eyes group the blue dots together and the yellow dots together as presuming that they have something in common with one another because of their shared color. We do this before we even know what the blue and yellow colors means.

ENCLOSURE

The **enclosure** principle reflects our tendency to perceive objects physically enclosed together in a ways that seems to create a boundary around that which is related or belonging to part of the same group, whether or not that is true or relevant to the data at hand. For example, our eyes see the yellow dots within the blue rectangle as having something in common with one another as they are all "in the box" together. We attribute meaning to the "box" whether that is accurate or not given what we are looking at.



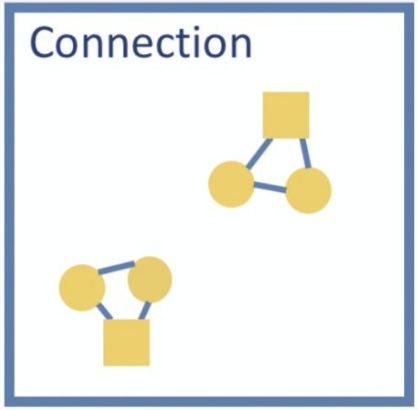


CLOSURE

The **closure** principle points out our tendency to perceive a set of individual components as a single, recognizable shape whenever possible, whether or not that is true or relevant to the data at hand. Meaning rather than perceiving something as open and incomplete, we "fill in the space" to make things closed and complete. For example, our eyes connect the dashed lines to create a hexagon, rather than seeing the unusual nature of different dashes in space. They jump to these closed and complete shapes before we actually know what we are looking at often.

CONTINUITY

The **continuity** principle identifies our tendency to perceive objects that are in line with one another or seem to be a continuation of one another as part of a single whole or group, whether or not that is true or relevant to the data at hand. Our eyes search for the smoothest path to follow, thus looking for continuity among objects. For example, our brains presume that all of these blue bars share a common baseline, our eyes see that the left side of each bars aligns together. Also, our eyes perceive that the blue bars are decreasing as you go down and presume meaning to that.

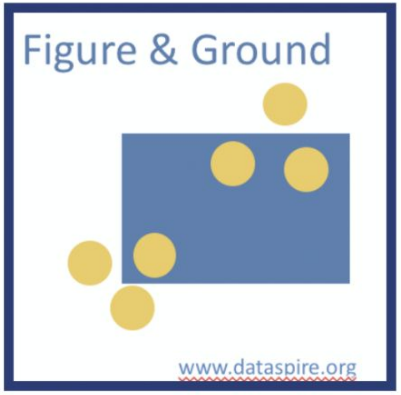


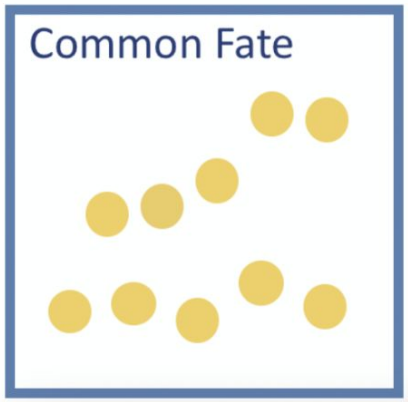
CONNECTION

The **connection** principle points out our tendency to perceive objects that are physically connected are part of the same group, whether or not that is true or relevant to the data at hand. For example, our eyes connect the group of three shapes in the lower left together and the three shapes in the upper right together, before connecting the dots and squares together. In fact, the **connection** principle often is stronger than **proximity** or **similarity** principles, but not as strong as **enclosure**.

FIGURE & GROUND

The **figure & ground** principle outlines our tendency to perceive objects either in the foreground or in the background with the background having less implied meaning, whether or not that is true or relevant to the data at hand. Specifically, the **figure** is considered the object, shape, or person that is in focus of the visual field and where we "should" pay attention, and the **ground** is what is in the background and "should" be ignored. For example, our eyes perceive the yellow dots being "in front of" the blue rectangle and perceive that "in front of" quality as indicating more importance. We either ignore or downplay what is in the background, without knowing what is there.





COMMON FATE

The **common fate** principle reflects our tendency to perceive objects that share a direction of placement together, or seem to be moving together, as a unified group, whether or not that is true or relevant to the data at hand. For example, our eyes see the five top yellow dots from left to right as being a group increasing from left to right, and the bottom five dots as being a group staying the same from left to right. Before we know if the data are related, we begin to group these dots together in our minds.

ENCODING
(DESIGNER)

01 DATA
What is my data?

02 VISUAL MAPPING
What is my visual representation?

03 RENDERING
What is my medium?

VIEW

DECODING
(USER)

04 PERCEIVING
What does it show?

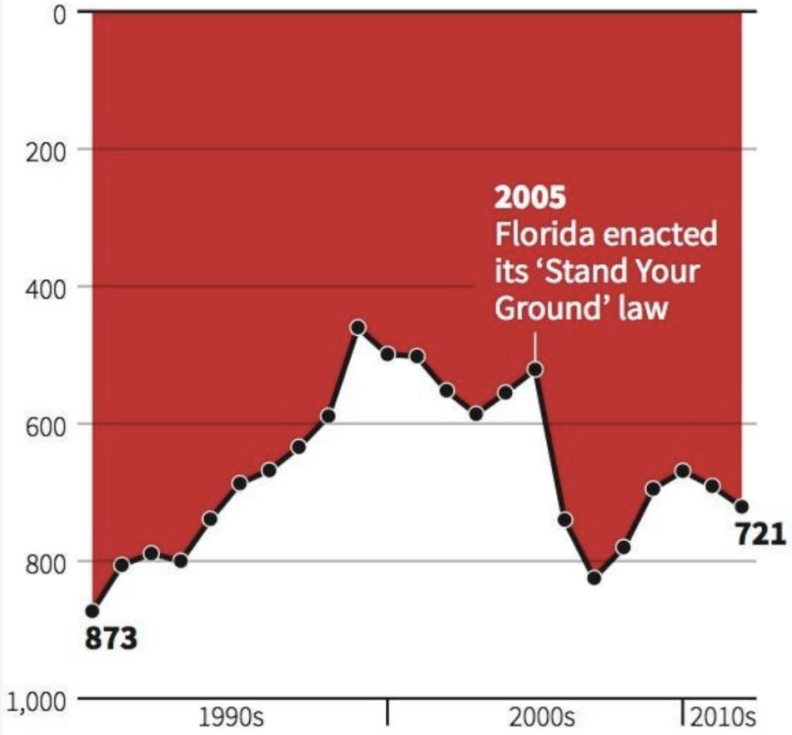
05 INTERPRETING
What does it mean?

06 COMPREHENDING
What does it mean for me?

INTERPRETING

Gun deaths in Florida

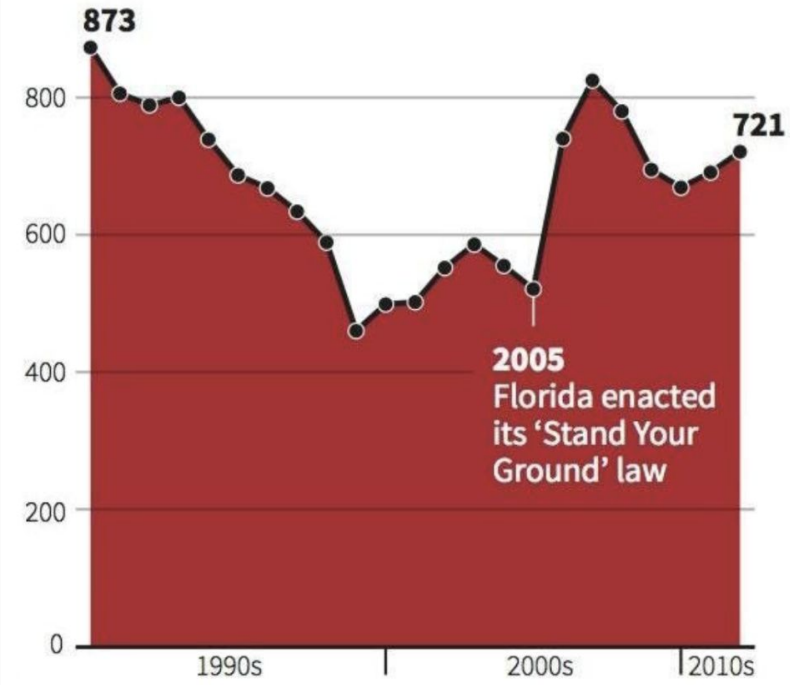
Number of murders committed using firearms



Source: Florida Department of Law Enforcement

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

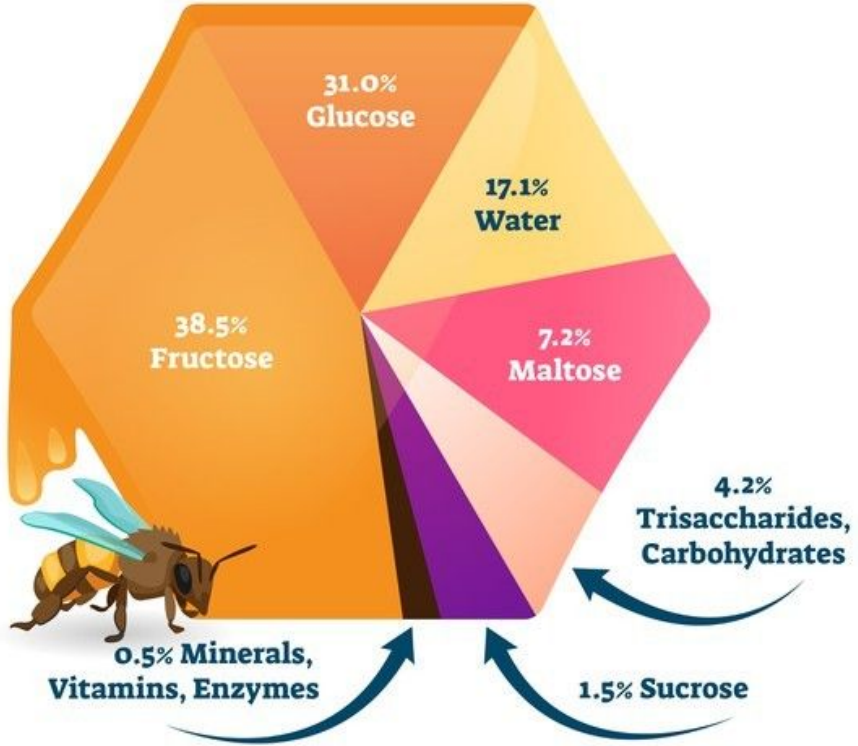
LET'S PLAY A GAME: GOOD OR BAD DATA VIZ?

Cities with best

Batting averages

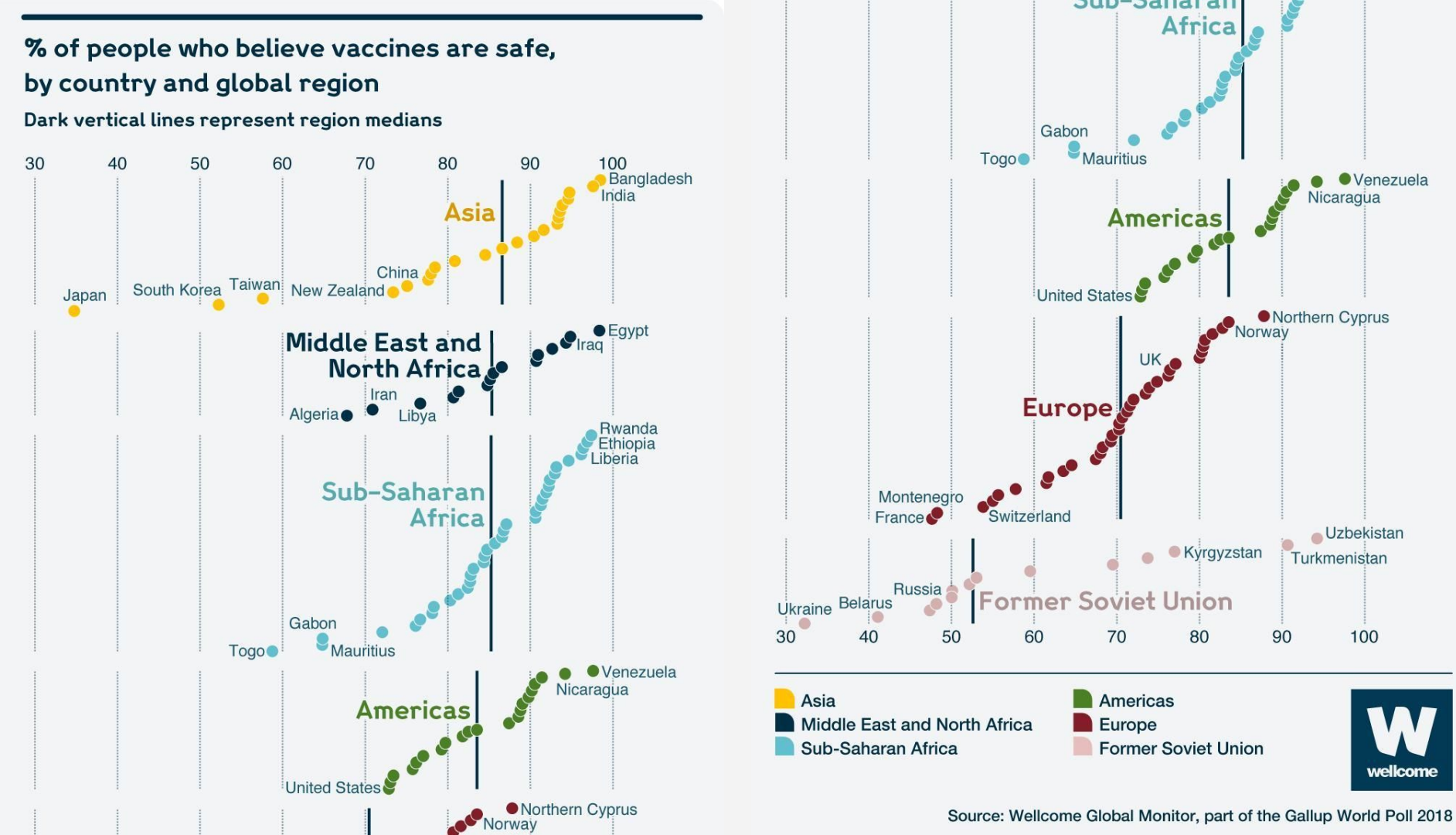
Sydney comes first with 5416 runs at an average of 45.37. It has produced some of the most prolific batsmen like David Warner, Steven Smith, Michael Clarke, Steve Waugh, and Mark Waugh. Next is Launceston with 2657 runs at an average of 44.80. This is mainly due to players like Ricky Ponting, David Boon and George Bailey.

COMPOSITION OF HONEY



% of people who believe vaccines are safe, by country and global region

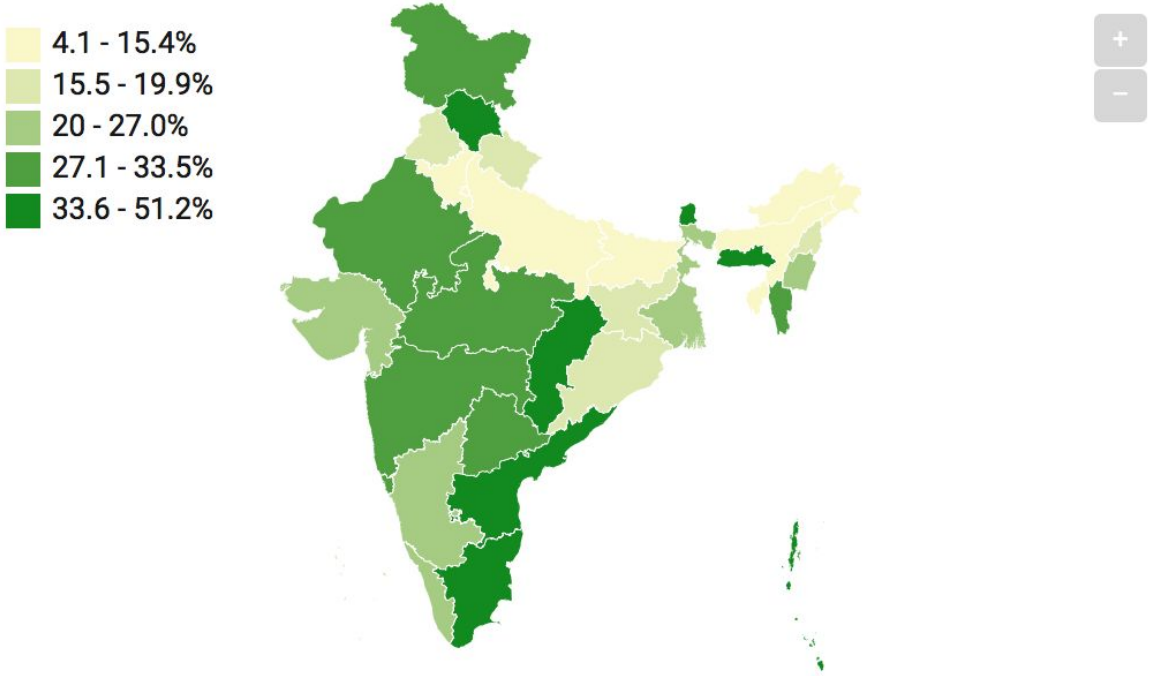
Dark vertical lines represent region medians



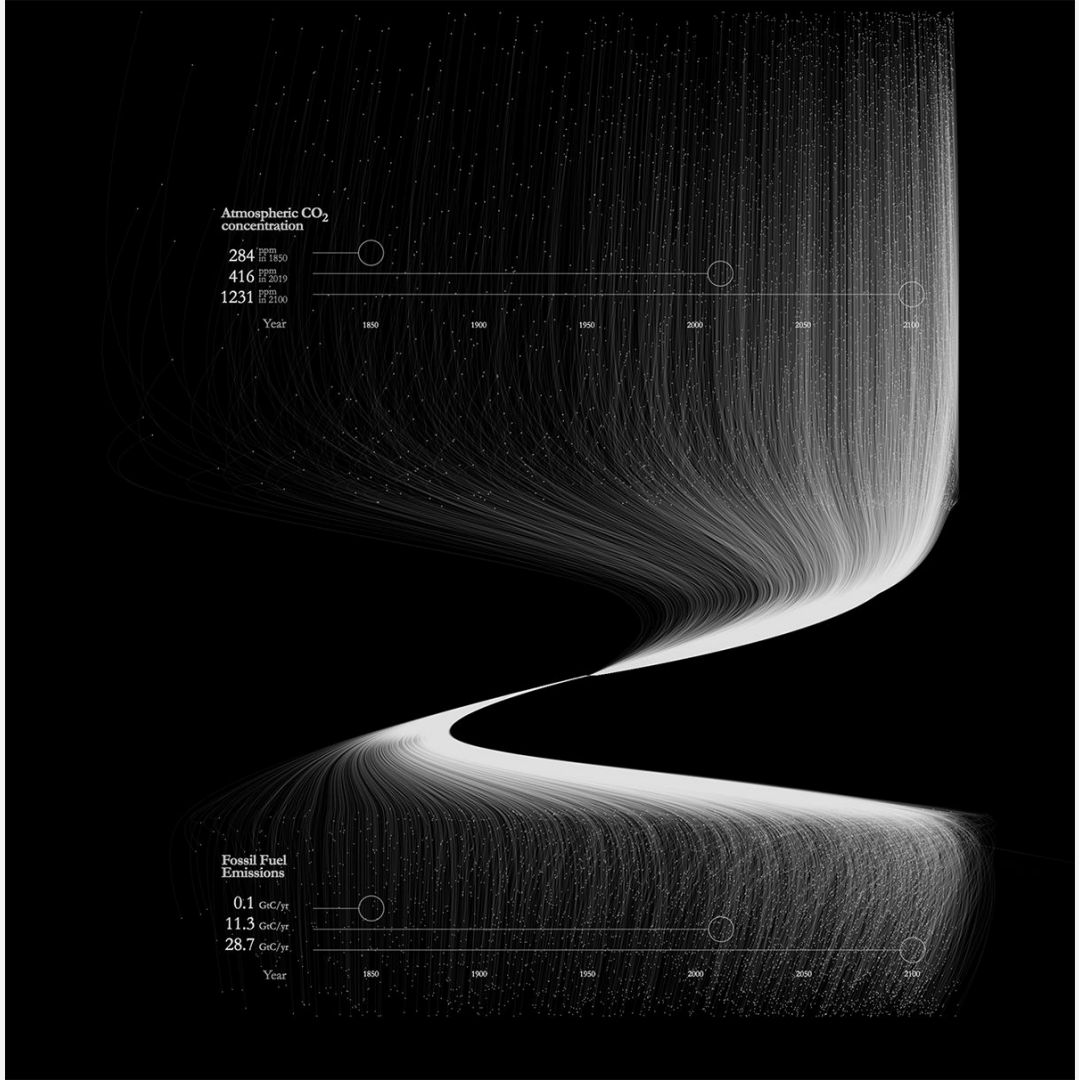
Source: Wellcome Global Monitor, part of the Gallup World Poll 2018

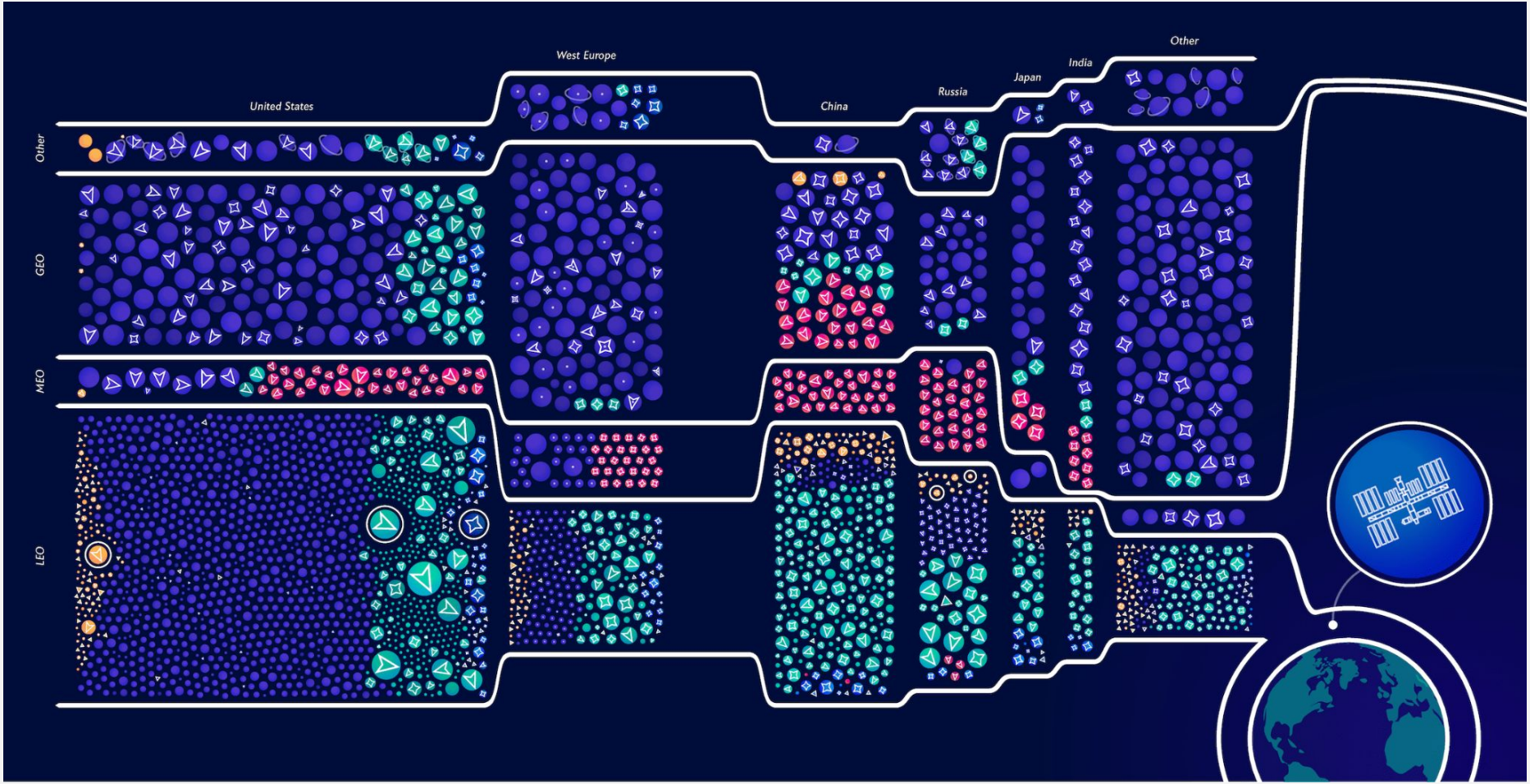
The Hindi belt scores low, while the south does better

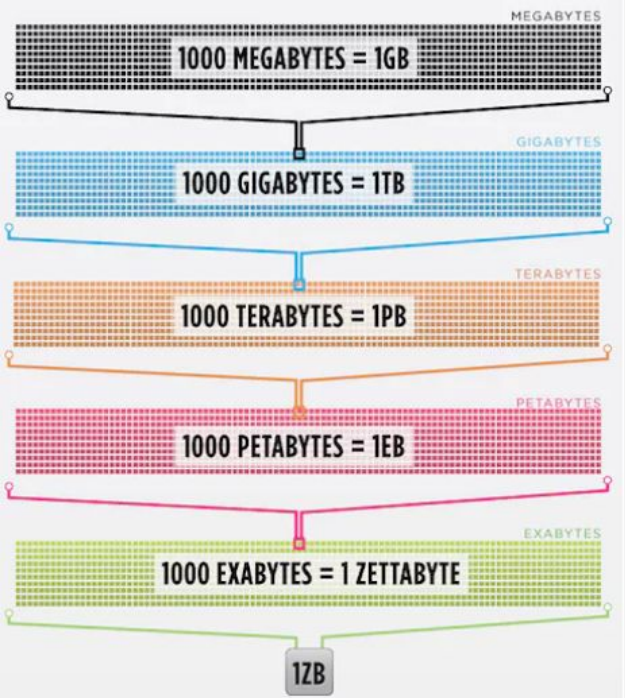
Female labour force participation rate (%)



Source: NSSO • Get the data • Created with Datawrapper

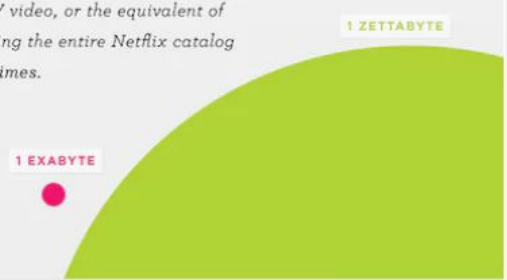

Atmospheric CO_2 concentration
284 ppm in 1850
416 ppm in 2019
1231 ppm in 2100
Year
1850
1900
1950
2000
2050
2100
Fossil Fuel Emissions
0.1 GtC/yr
11.3 GtC/yr
28.7 GtC/yr
Year
1850
1900
1950
2000
2050
2100

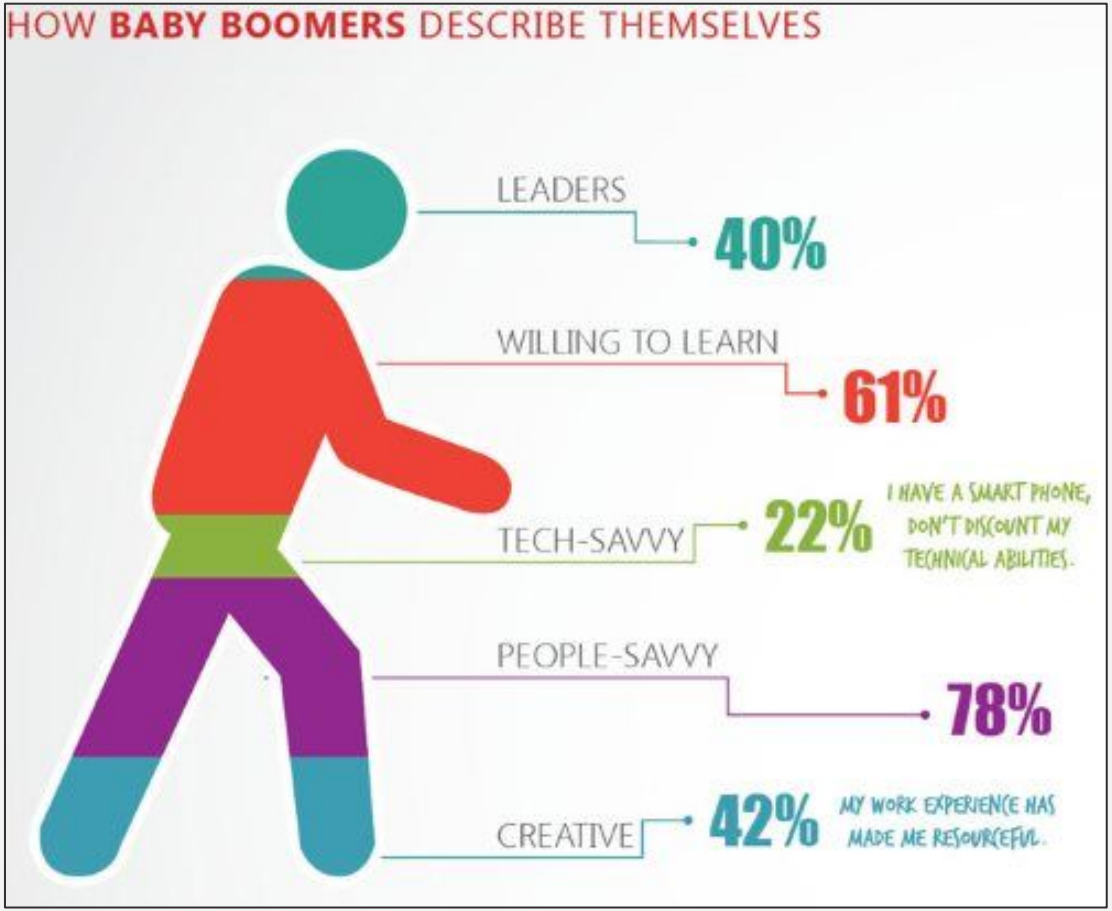

United States
West Europe
China
Russia
Japan
India
Other
Other
GEO
MEO
LEO

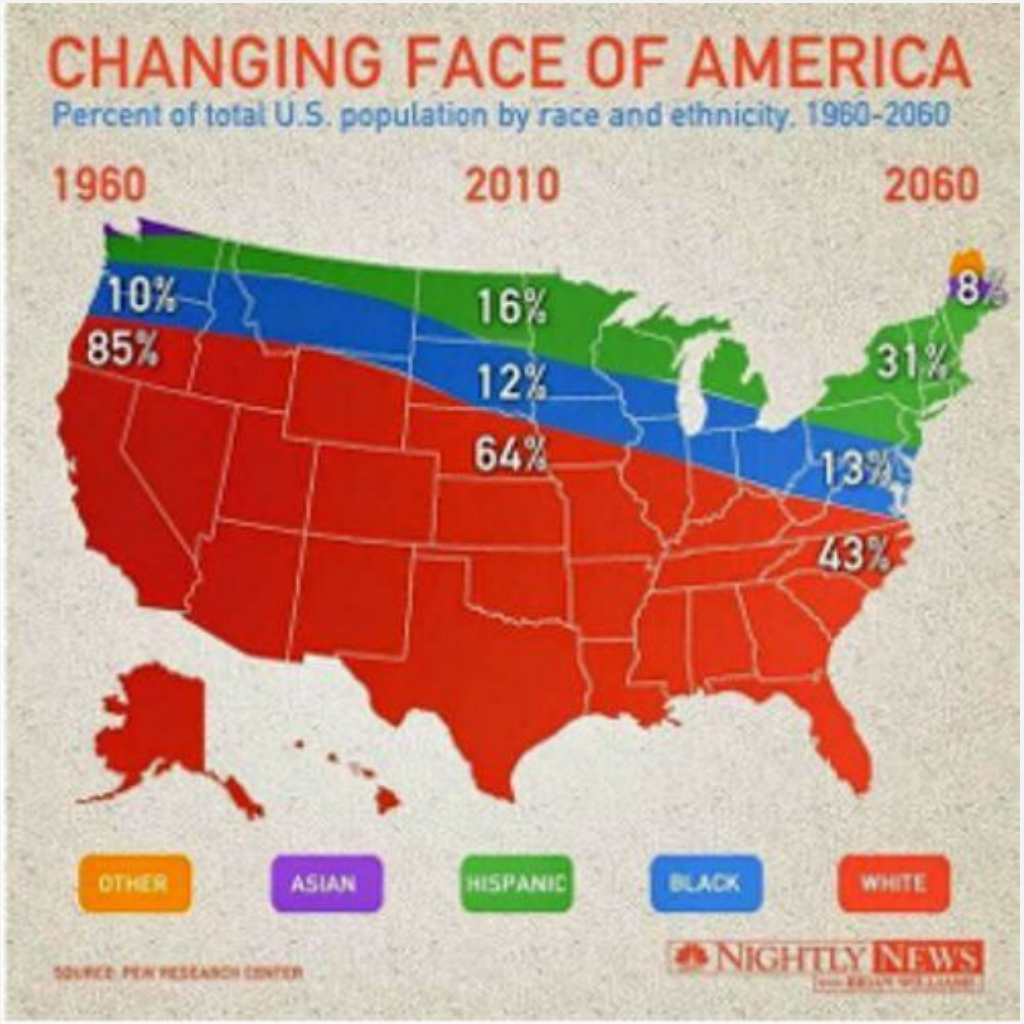


How much is an exabyte really?

1 exabyte amounts to 36,000 years of HD-TV video, or the equivalent of streaming the entire Netflix catalog 3,177 times.




HOW BABY BOOMERS DESCRIBE THEMSELVES
LEADERS
40%
WILLING TO LEARN
61%
TECH-SAVVY
22%
I HAVE A SMART PHONE, DON'T DISCOUNT MY TECHNICAL ABILITIES.
PEOPLE-SAVVY
78%
CREATIVE
42%
MY WORK EXPERIENCE HAS MADE ME RESOURCEFUL.


CHANGING FACE OF AMERICA
Percent of total U.S. population by race and ethnicity, 1960-2060
1960
2010
2060
10%
85%
16%
12%
64%
8%
31%
13%
43%
OTHER
ASIAN
HISPANIC
BLACK
WHITE
SOURCE: PEW RESEARCH CENTER
NIGHTLY NEWS

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par des largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de MM. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

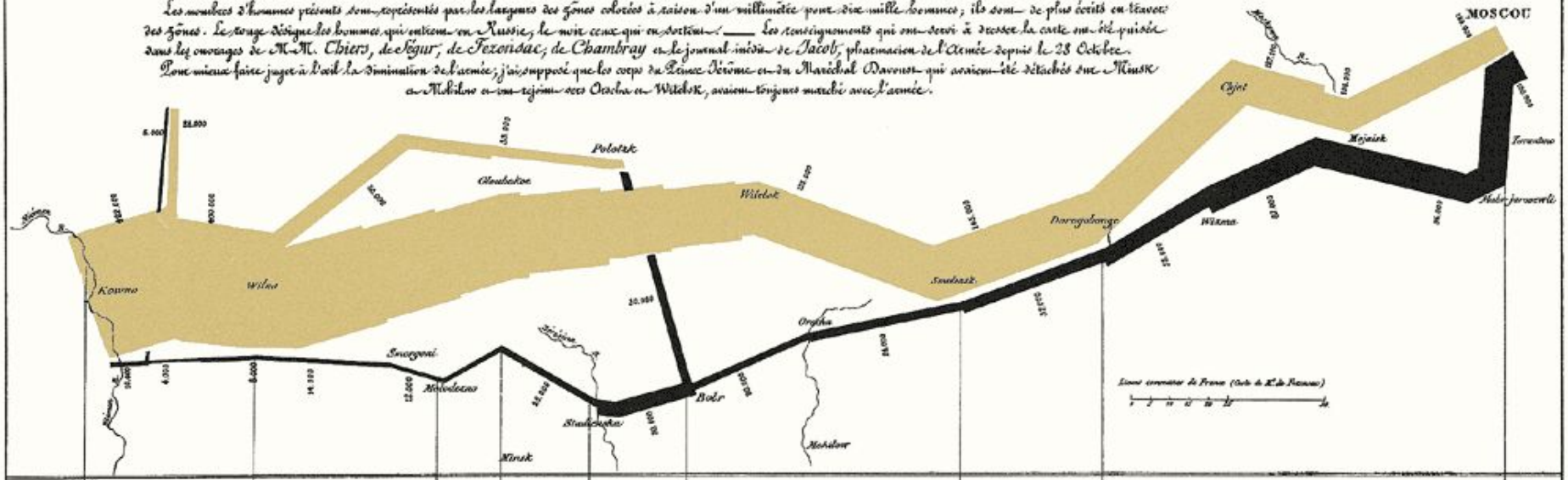
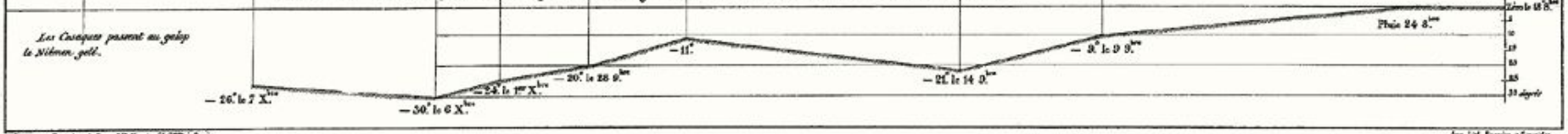


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Imp. par Regnier, 1 Pas. St.e Marie St. Avoye à Paris.

Imp. Lith. Regnier et Dourdet.