Getting Started with Data Visualization

Oluwakemi Ola, PhD
Associate Professor of Teaching
Computer Science Department

https://tinyurl.com/ashesi-viz

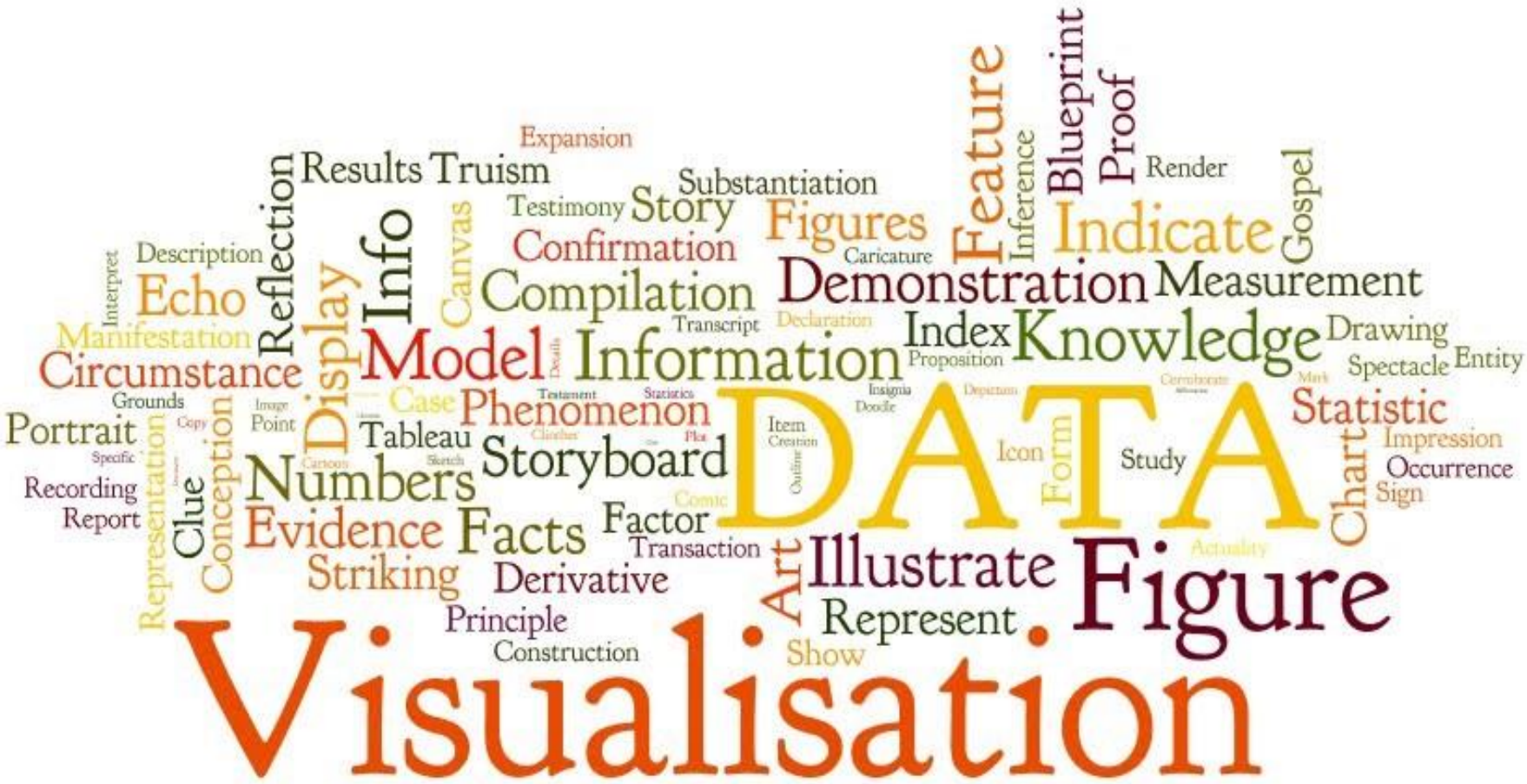
IF YOU
GET TIRED,
LEARN
TO REST,
NOT TO QUIT

Let’s get to coding.

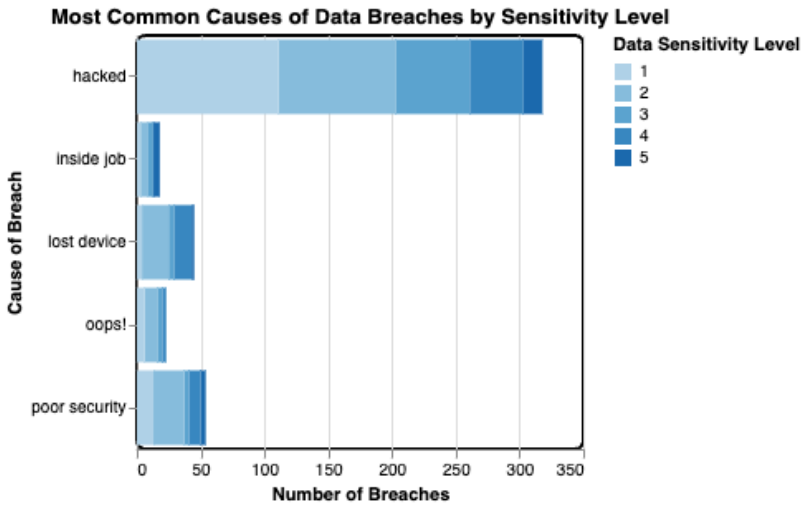
https://tinyurl.com/ashesi-viz

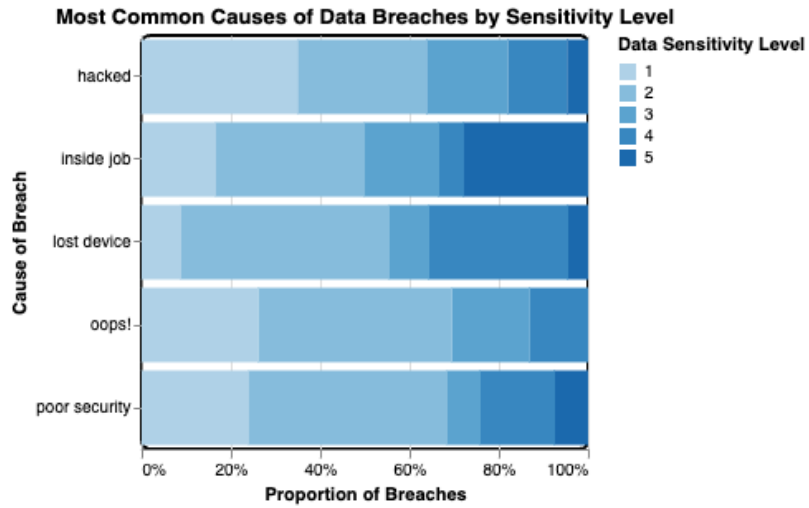
This is the repo for all the files for the workshop

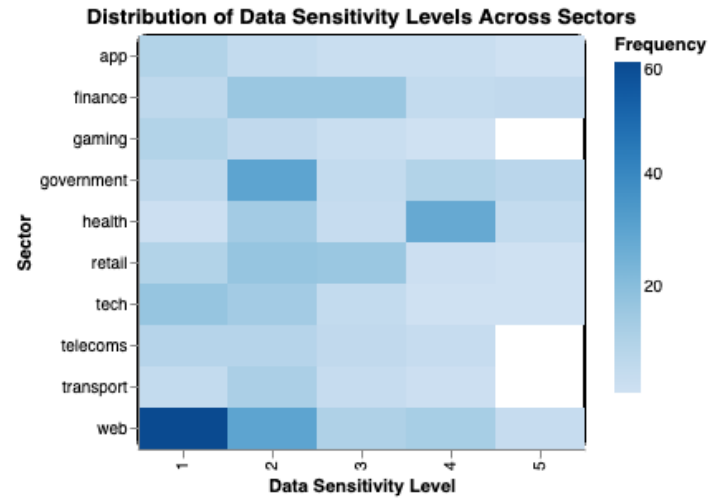
Today’s file is Ashesi_WK2S.ipynb

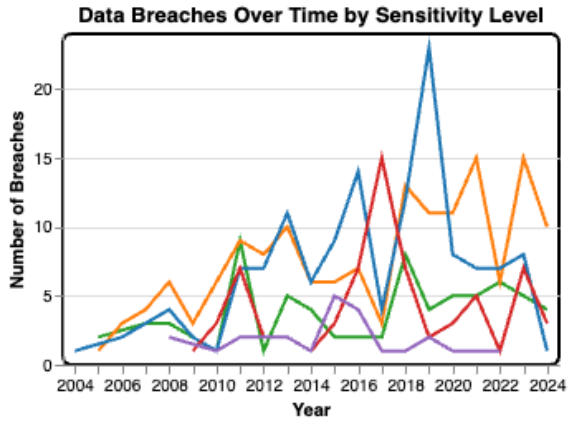

Visualisation
DATA
Figure
Information
Model
Knowledge
Demonstration
Compilation
Storyboard
Illustrate
Evidence
Facts
Numbers
Phenomenon
Indicate
Measurement
Feature
Story
Figures
Confirmation
Results
Truism
Expansion
Substantiation
Testimony
Reflection
Display
Info
Canvas
Echo
Description
Interpret
Manifestation
Circumstance
Portrait
Grounds
Recording
Report
Representation
Clue
Conception
Point
Case
Tableau
Striking
Derivative
Principle
Construction
Factor
Transaction
Art
Represent
Show
Transcript
Declaration
Index
Proposition
Caricature
Inference
Blueprint
Proof
Render
Gospel
Drawing
Spectacle
Entity
Statistic
Impression
Occurrence
Sign
Chart
Study
Form
Icon
Item
Creation
Insignia
Doodle
Depiction
Corroborate
Mark
Actuality
Comic
Outline
Plot
Clinches
Sketch
Cartoon
Image
Copy
Specific
Details
Testament
Statistics

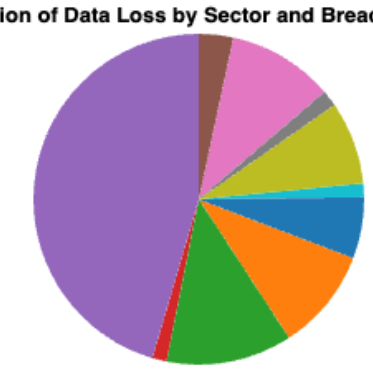
Data Breach Dashboard

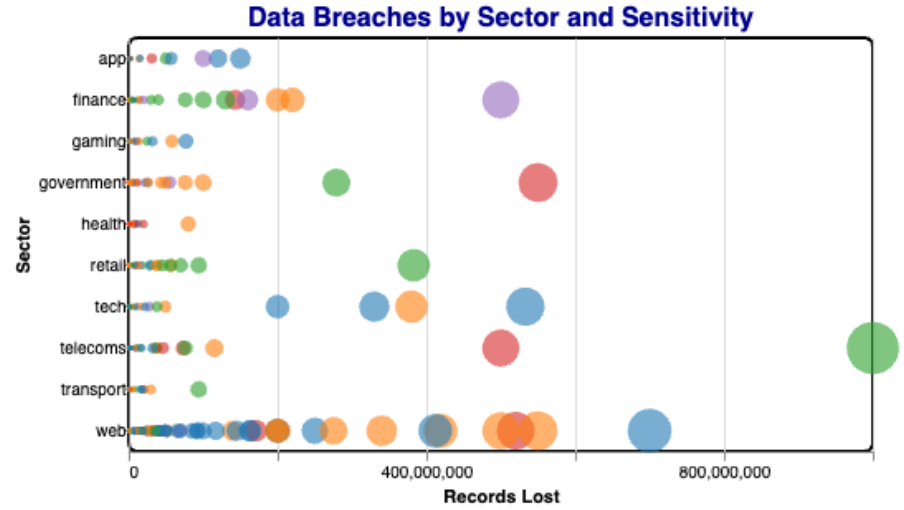












Altair Basics

Create a Chart Object

Attach data to the Chart Object

Specify the mark type

```
alt.Chart(data).mark_bar().encode(
        channel_1 = 'column1',
        channel_2 = 'column2',
)
```

Specify each channel and what data it encodes

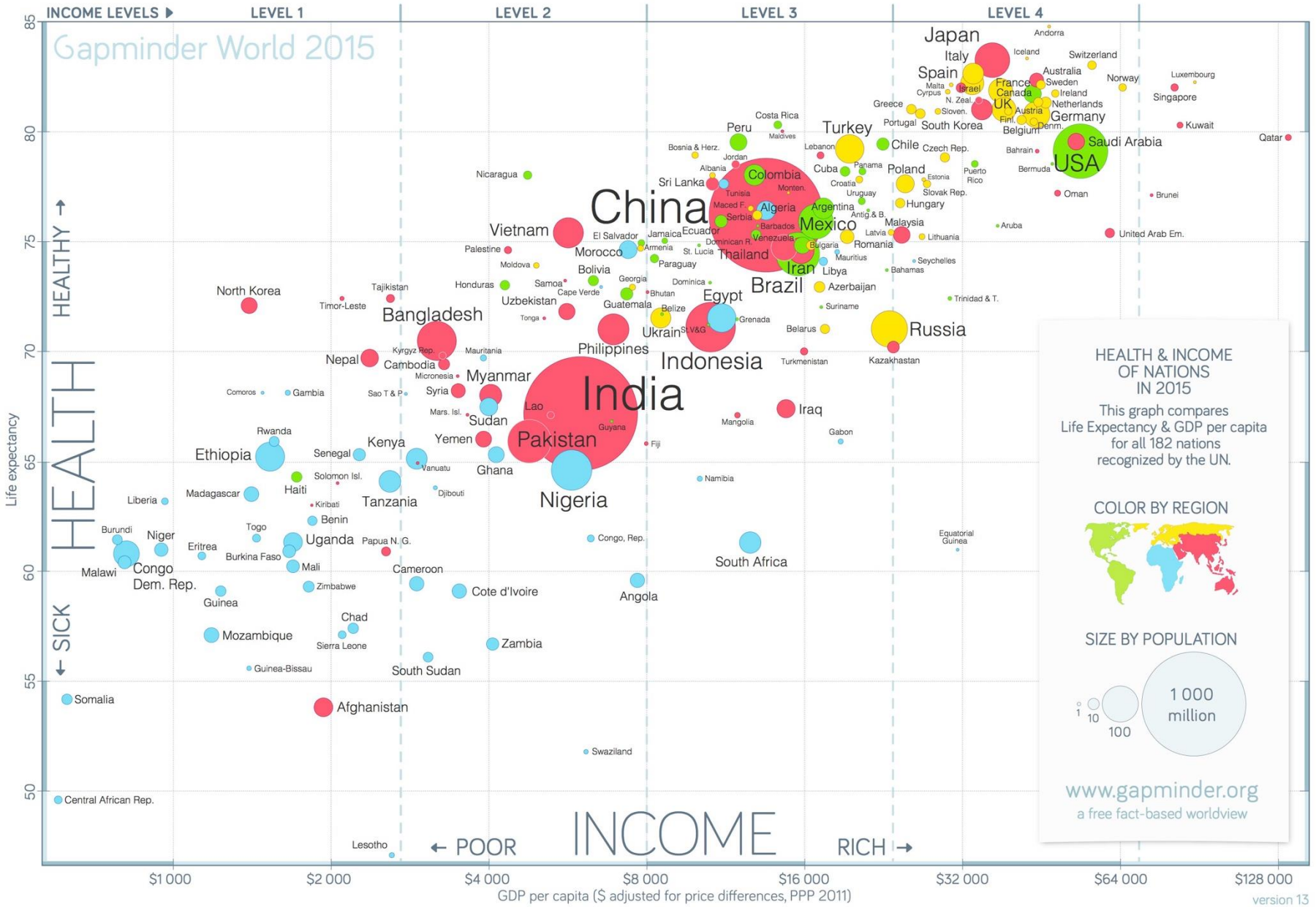
Encoding Data Types

The details of any mapping depend on the *type* of the data. Altair recognizes five main data types:

Data Type	Shorthand Code	Description
quantitative	`Q`	a continuous real-valued quantity
ordinal	`O`	a discrete ordered quantity
nominal	`N`	a discrete unordered category
temporal	`T`	a time or date value
geojson	`G`	a geographic shape

https://altair-viz.github.io/user_guide/encodings/index.html

What would the specifications be for this visualization

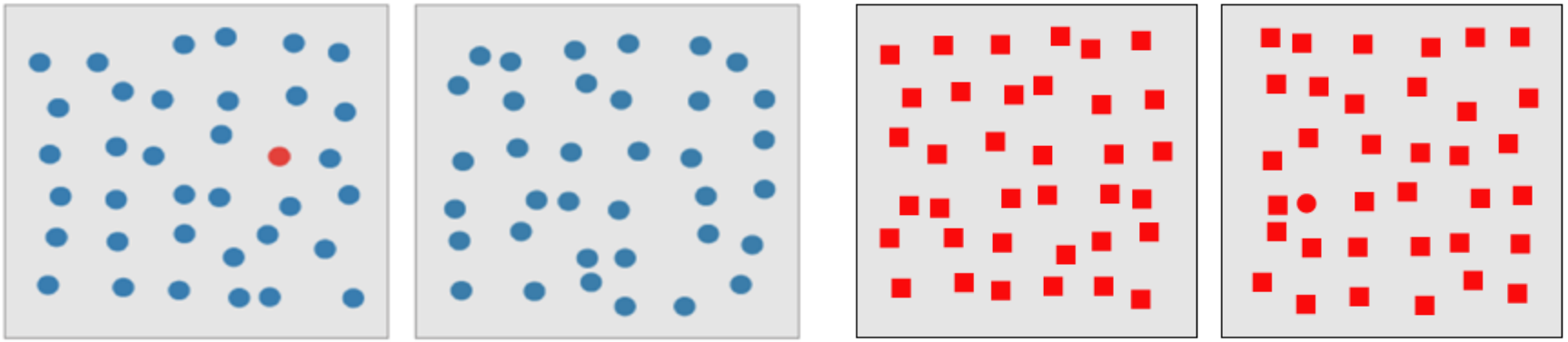

INCOME LEVELS
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4
Gapminder World 2015
HEALTHY →
HEALTH
← SICK
Life expectancy
China
India
Nigeria
Pakistan
Indonesia
Brazil
Russia
USA
Japan
Bangladesh
Ethiopia
Vietnam
Mexico
Thailand
Iran
Egypt
Philippines
Turkey
Germany
France
Spain
Italy
UK
Canada
Australia
Saudi Arabia
South Africa
Colombia
Argentina
Algeria
Iraq
Sudan
Kenya
Tanzania
Uganda
Ghana
Yemen
Myanmar
Nepal
Cambodia
North Korea
Afghanistan
Somalia
Angola
Zambia
Mozambique
Cameroon
Cote d'Ivoire
Malaysia
Poland
Ukraine
Peru
Chile
Netherlands
Switzerland
Sweden
Norway
Singapore
Kuwait
Qatar
United Arab Em.
Oman
Brunei
Lithuania
Romania
Hungary
Czech Rep.
Portugal
South Korea
Belgium
Austria
Ireland
Greece
Israel
Venezuela
Cuba
Panama
Bolivia
Paraguay
Honduras
Nicaragua
Guatemala
Uzbekistan
Azerbaijan
Libya
Kazakhstan
Belarus
Turkmenistan
Suriname
Trinidad & T.
Bahamas
Seychelles
Mauritius
Aruba
Bermuda
Puerto Rico
Slovak Rep.
Uruguay
Ecuador
Jamaica
El Salvador
Morocco
Palestine
Moldova
Tajikistan
Timor-Leste
Comoros
Gambia
Senegal
Haiti
Madagascar
Liberia
Rwanda
Burundi
Niger
Eritrea
Togo
Burkina Faso
Mali
Malawi
Congo Dem. Rep.
Guinea
Zimbabwe
Chad
Sierra Leone
Guinea-Bissau
South Sudan
Lesotho
Central African Rep.
Swaziland
Namibia
Gabon
Equatorial Guinea
Congo, Rep.
Papua N. G.
Solomon Isl.
Vanuatu
Fiji
Tonga
Samoa
Micronesia
Kiribati
Benin
Djibouti
Mauritania
Kyrgyz Rep.
Georgia
Armenia
Albania
Bosnia & Herz.
Jordan
Lebanon
Sri Lanka
Maldives
Costa Rica
Andorra
Iceland
Luxembourg
Malta
Cyprus
N. Zeal.
Finl.
Denm.
Sloven.
Croatia
Estonia
Latvia
Bulgaria
Serbia
Maced F.
Tunisia
Monten.
Barbados
Antig & B.
Dominican R.
St. Lucia
Grenada
St V&G
Belize
Bhutan
Cape Verde
Dominica
Mongolia
Guyana
Lao
Mars. Isl.
Sao T & P
Syria
Micronesia
Cambodia
Nepal
Kyrgyz Rep.
Mauritania
Moldova
Tajikistan
Timor-Leste
Honduras
Samoa
Cape Verde
Bhutan
Guatemala
Uzbekistan
Tonga
Belize
Grenada
Suriname
Trinidad & T.
Bahamas
Seychelles
Mauritius
Aruba
Bermuda
Puerto Rico
Slovak Rep.
Uruguay
Ecuador
Jamaica
El Salvador
Morocco
Palestine
Moldova
Bolivia
Paraguay
Armenia
St. Lucia
Dominican R.
Barbados
Antig & B.
Maced F.
Tunisia
Monten.
Serbia
Albania
Bosnia & Herz.
Jordan
Lebanon
Sri Lanka
Maldives
Costa Rica
Andorra
Iceland
Luxembourg
Malta
Cyprus
N. Zeal.
Finl.
Denm.
Sloven.
Croatia
Estonia
Latvia
Bulgaria
Peru
Chile
Netherlands
Switzerland
Sweden
Norway
Singapore
Kuwait
Qatar
United Arab Em.
Oman
Brunei
Lithuania
Romania
Hungary
Czech Rep.
Portugal
South Korea
Belgium
Austria
Ireland
Greece
Israel
Venezuela
Cuba
Panama
Poland
Colombia
Argentina
Algeria
Iraq
Sudan
Kenya
Tanzania
Uganda
Ghana
Yemen
Myanmar
Nepal
Cambodia
North Korea
Afghanistan
Somalia
Angola
Zambia
Mozambique
Cameroon
Cote d'Ivoire
Malaysia
Poland
Ukraine

Visual Perception

It has been estimated that almost 50% of your brain is involved in visual processing and 70% of all your sensory receptors are in your eyes.

Our visual systems **pre-attentively** process many features within our visual field in less than 250 milliseconds without requiring any conscious cognitive effort.

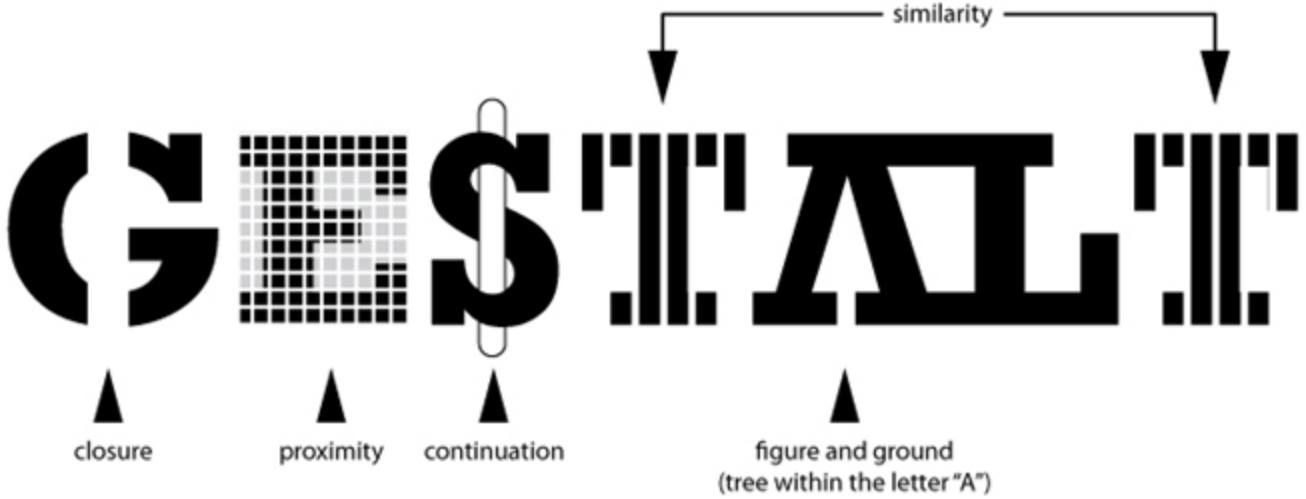
When you spot the red dot say OOOOO



https://www.csc2.ncsu.edu/faculty/healey/PP/

Gestalt Principles

Theories/principles proposed by psychologists in the 1920s to explain how people organize/group information visually, in other words, the ways we visually assemble objects into groups.



Similarity

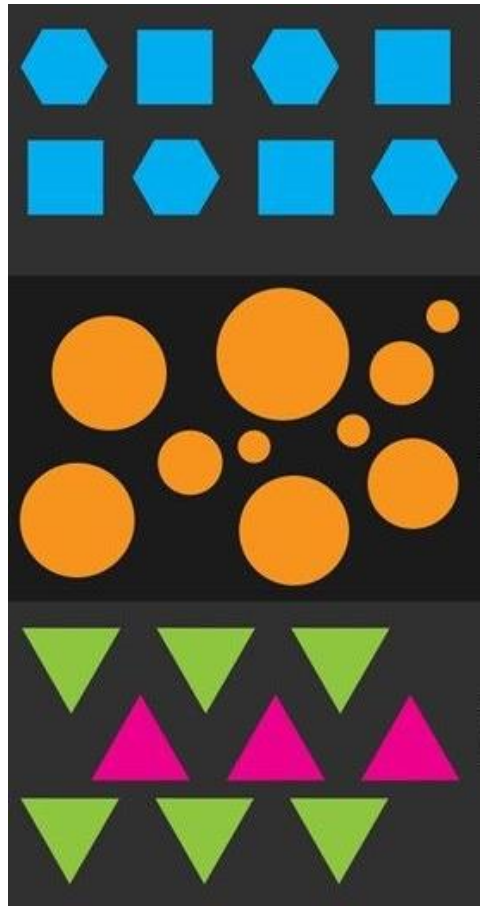
People tend to see things that physically resemble each other as part of the same group.

We typically use

- Shape
- Size
- Color

To organize visual items into groups.

Design Tip: Use the channel that is most effective to highlight the similarities in the data



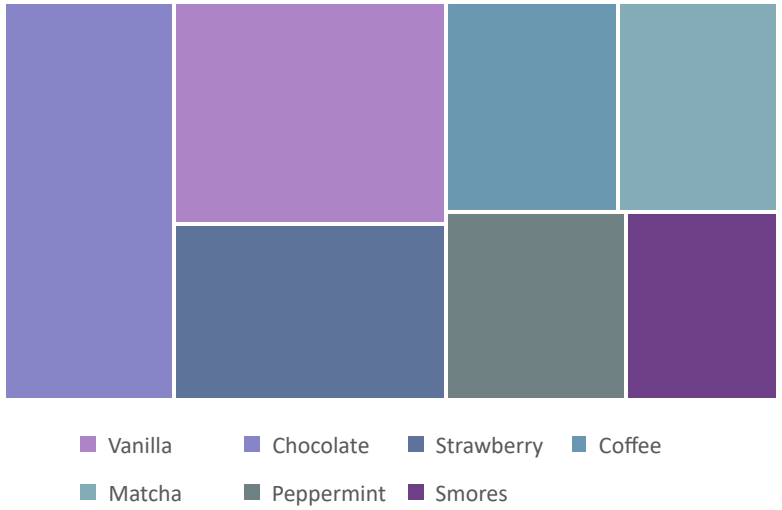
https://www.interaction-design.org/literature/article/the-law-of-similarity-gestalt-principles-1

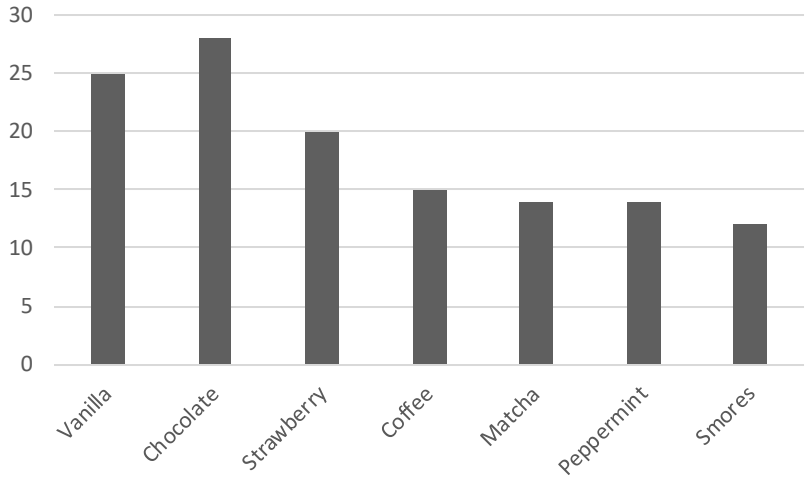
Figure / Ground

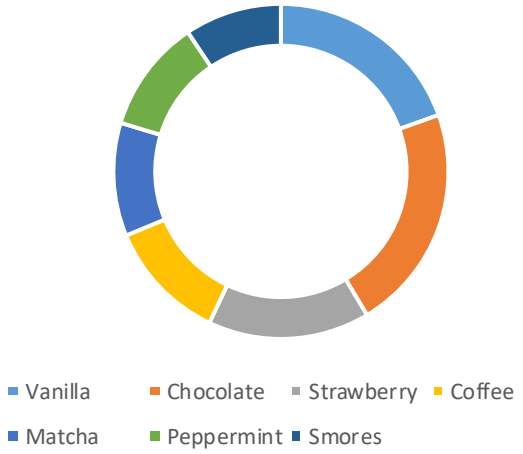
Your brain distinguishes the foreground and the background

So will any visualization work

Visual representations are beneficial when used for appropriate task





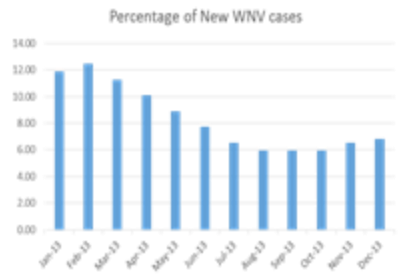


Representation Effect

Which visualization is best suited to answer the question, which month had the highest percentage of new cases?

A

B



C

D

Accuracy: User studies

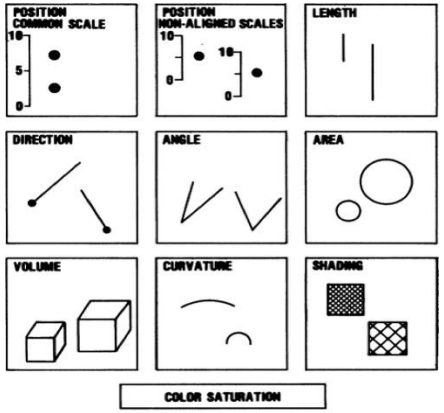
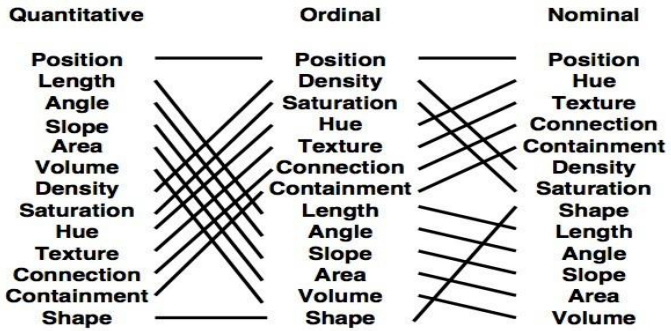


Figure 1. Elementary perceptual tasks.



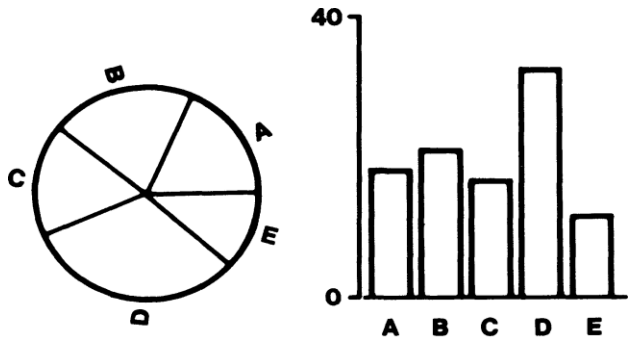


Figure 3. Graphs from position–angle experiment.

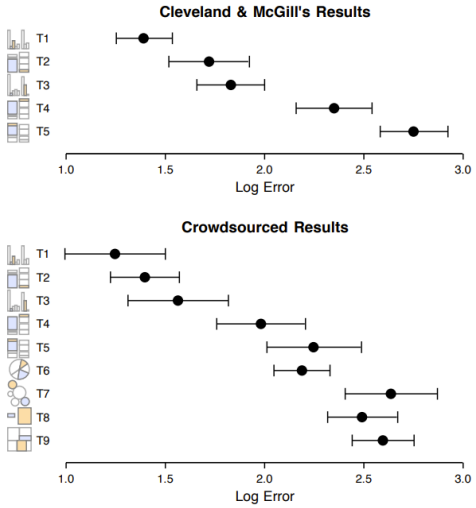
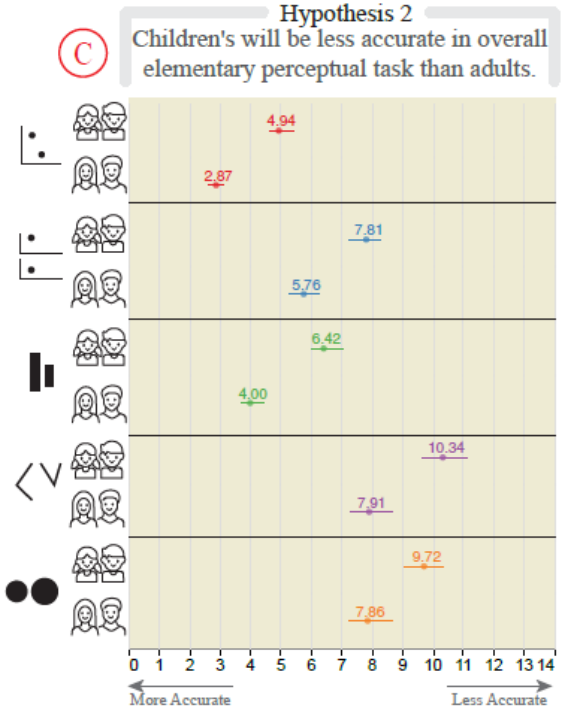


Figure 4: Proportional judgment results (Exp. 1A & B). Top: Cleveland & McGill's [7] lab study. Bottom: MTurk studies. Error bars indicate 95% confidence intervals.

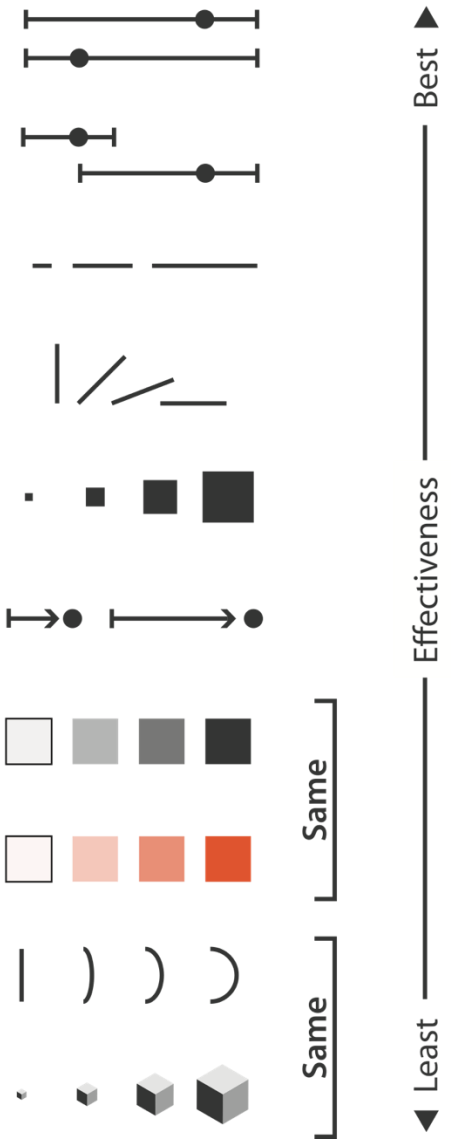


[Cleveland & McGill, 1984]
[Mackinlay, 1986]
[Heer & Bostock, 2010]
[Panavas et al., 2022]

Channels: Rankings

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



Identity Channels: Categorical Attributes

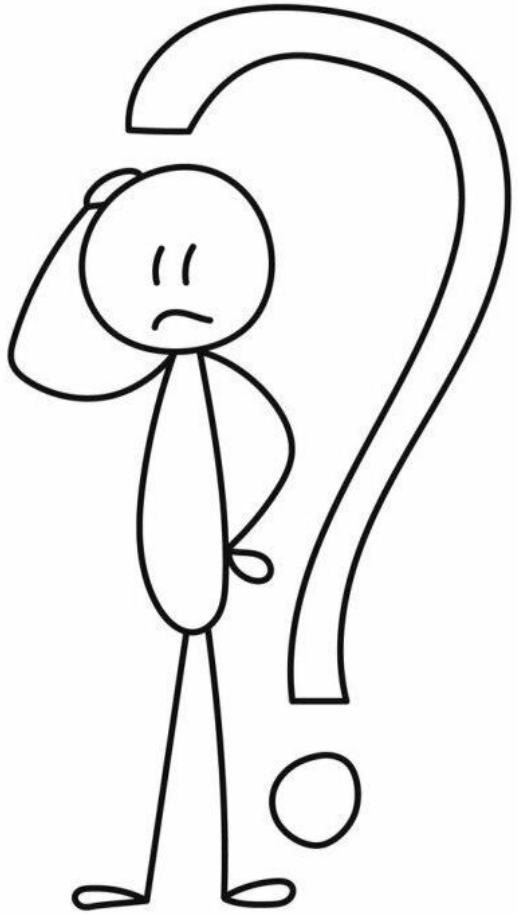
- Spatial region
- Color hue
- Motion
- Shape

Representation Effect

The website below groups charts by their functionality
http://www.datavizcatalogue.com/index.html

Another catalogue

https://datavizproject.com/#



Prioritize choosing the most appropriate channel for each attribute

How do I pick *which* marks or channels to use?

COMPARING CATEGORIES

Compare values across categories

- Bar
- Paired Bar
- Stacked Bar
- Span
- Unit
- Column
- Paired Column
- Stacked Column
- Radial Bar
- Circular Column
- Sankey
- Waterfall
- Triangle Column
- Heatmap
- Chernoff
- Marimekko
- Bubble Comparison
- Gauge
- Slope
- Isotype
- Mosaic
- Nested Bubbles
- Bullet
- Dot Plot
- Word Cloud

TIME

Track changes over time

- Line
- Stream
- Cause-Effect
- Sparklines
- History Flow
- Timeline
- Horizon
- Connected Scatterplot
- Flow Chart
- Calendar
- Arc-Time
- Bump
- Stacked Area
- Gantt
- Cycle

DISTRIBUTION

Representation of the distribution of data

- Histogram
- Fan
- Box & Whisker
- Pyramid
- Violin
- Stem & Leaf
- Stock
- Stripe
- Confidence Interval
- Error Bars

GEOSPATIAL

Relates data to its geography

- Map
- Flow Map
- Icon Map
- Choropleth
- Map with Columns
- Isopleth
- Cartogram
- Map with Pie Charts
- Contour
- Non-Contiguous Cartogram
- Bubble Map
- Dorling Map
- Connection Map
- Point Map
- Subway Map

PART-TO-WHOLE

Relates the part of a variable to its total

- Pie
- Treemap
- Donut
- Square Cloud
- Arc
- Voronoi
- Nightingale
- Triangle Treemap
- Sunburst
- Waffle

RELATIONSHIP

Illustrates correlations or relationships between variables

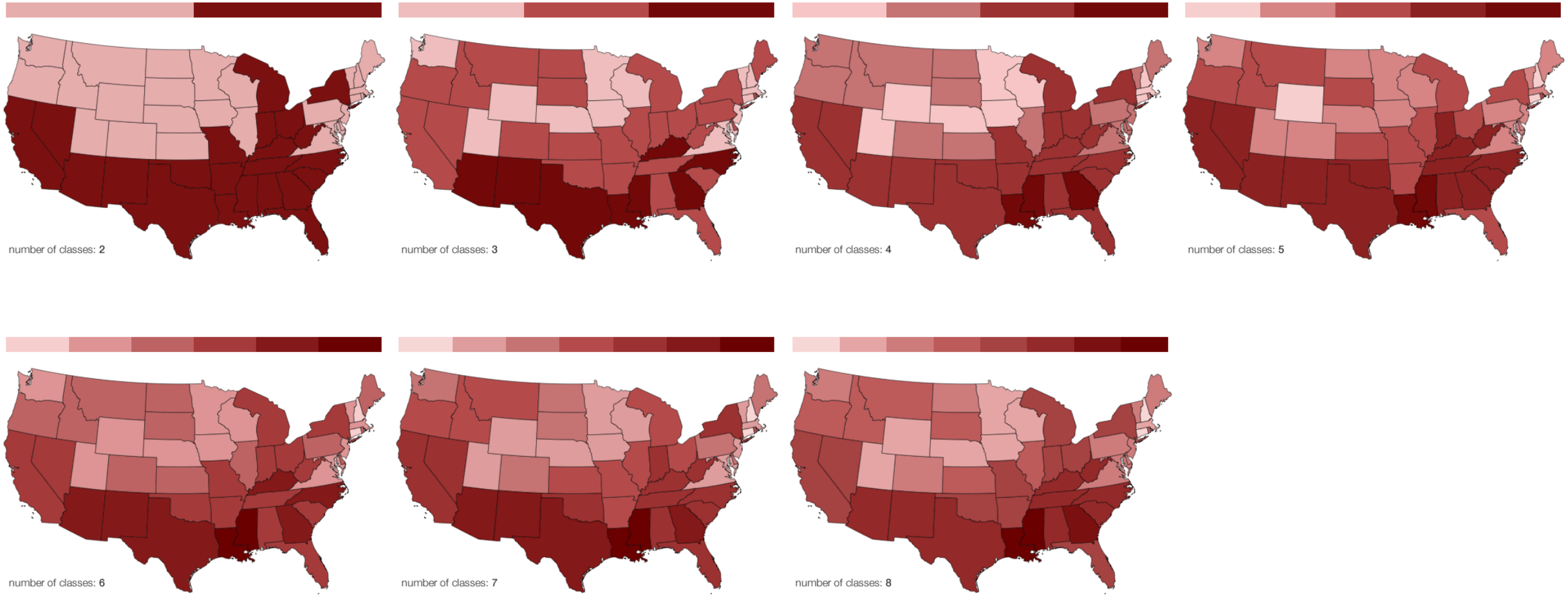
- Scatterplot
- Arc-Connection
- Dendrogram
- Word Tree
- Circle Packing
- Chord
- Tree
- Correlation Matrix
- Bubble
- Hive
- Double Tree
- Radar
- Parallel Coordinates
- Force-Directed
- Network
- Venn Diagram

© Jonathan Schwabish & Severino Ribecca

@jschwabish @SR_Visual_Info

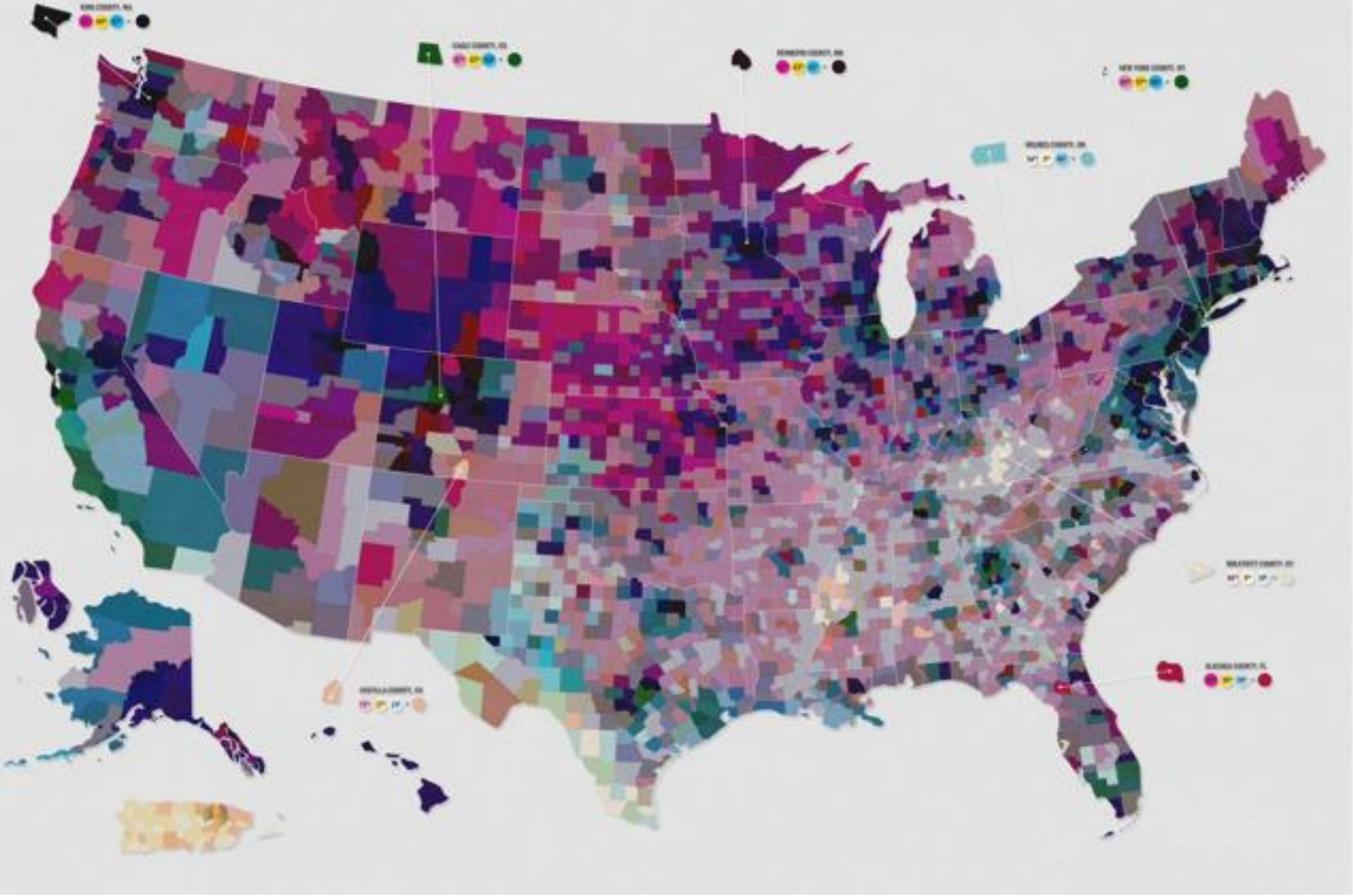
“HARD WORK BEATS TALENT WHEN TALENT DOESN'T WORK HARD”
-TIM NOTKE

Ordered color: limited number of discriminable bins

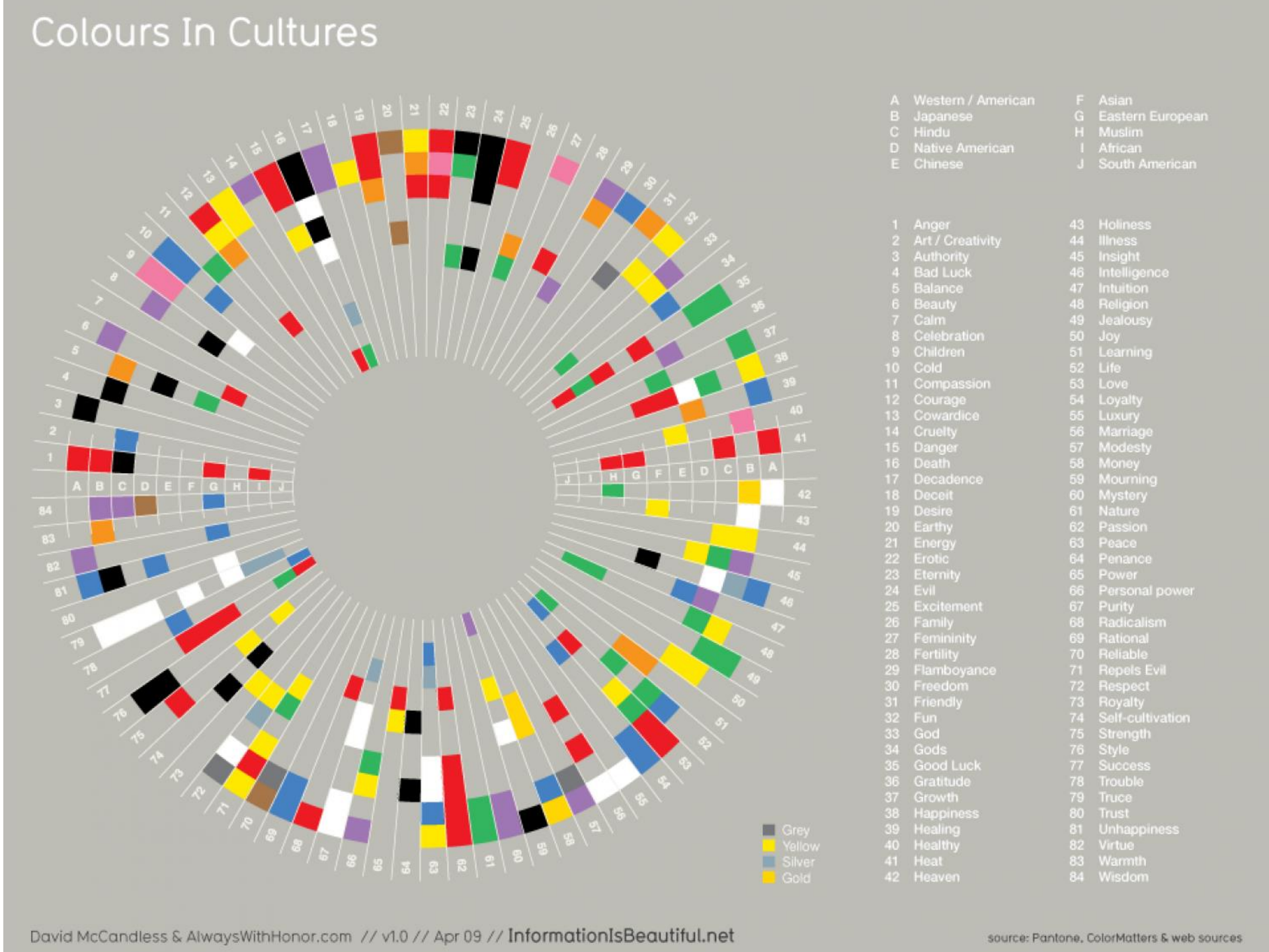


Gregor Aisch, vis4.net/blog/posts/choropleth-maps/

READING, WRITING, AND EARNING MONEY

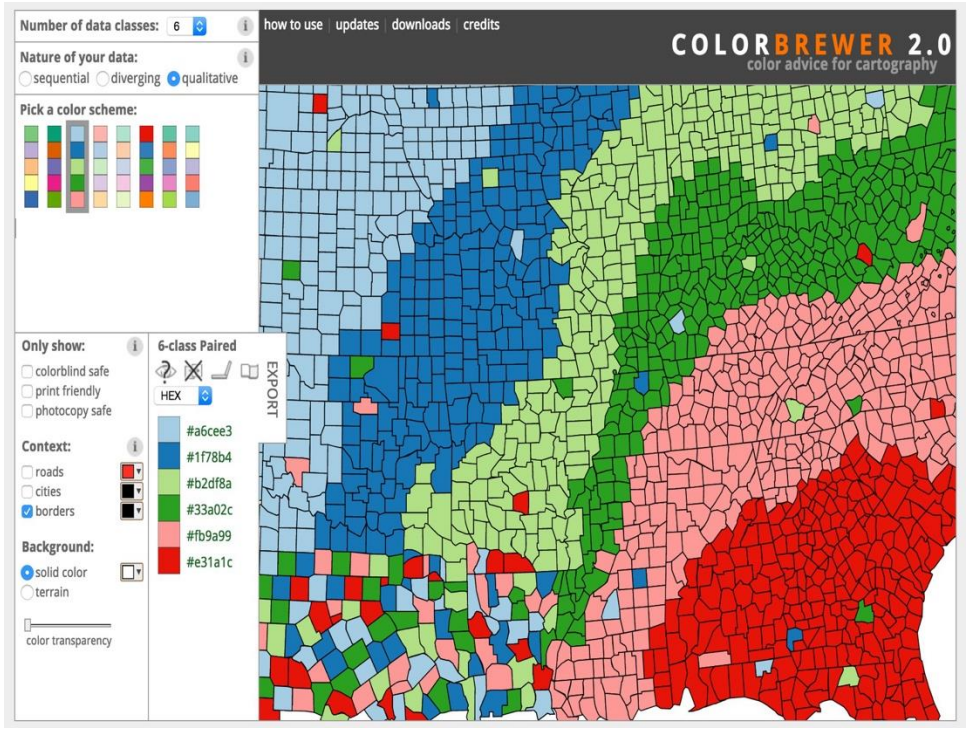


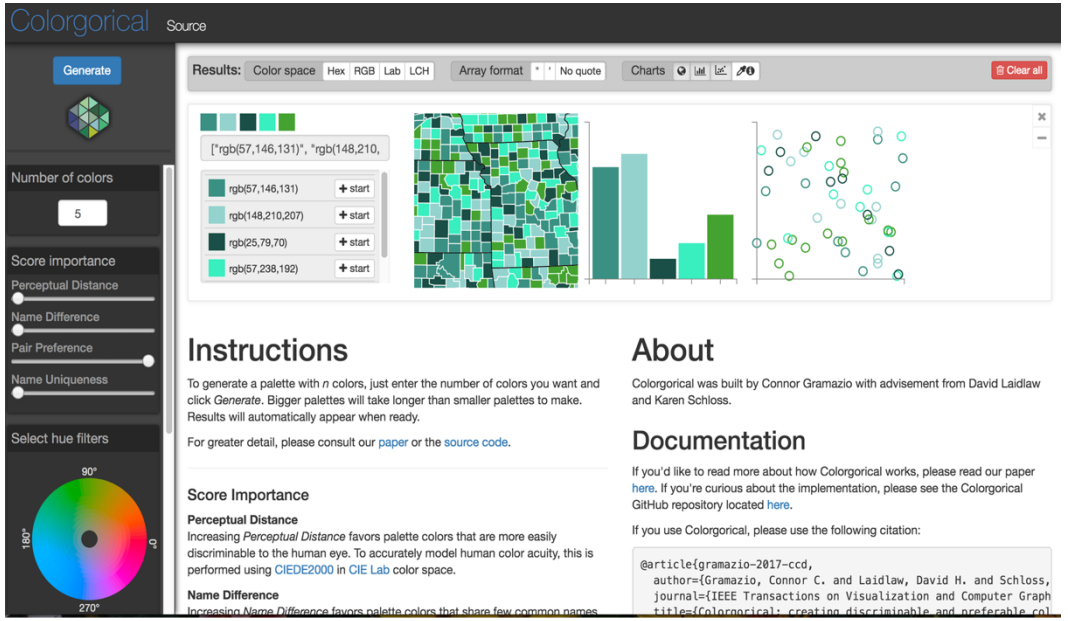
What does the color mean to you or to them?



Tools for picking Color

- Color Brewer
- Colorgorical
- Programming Libraries





Color/Lightness constancy: Illumination conditions

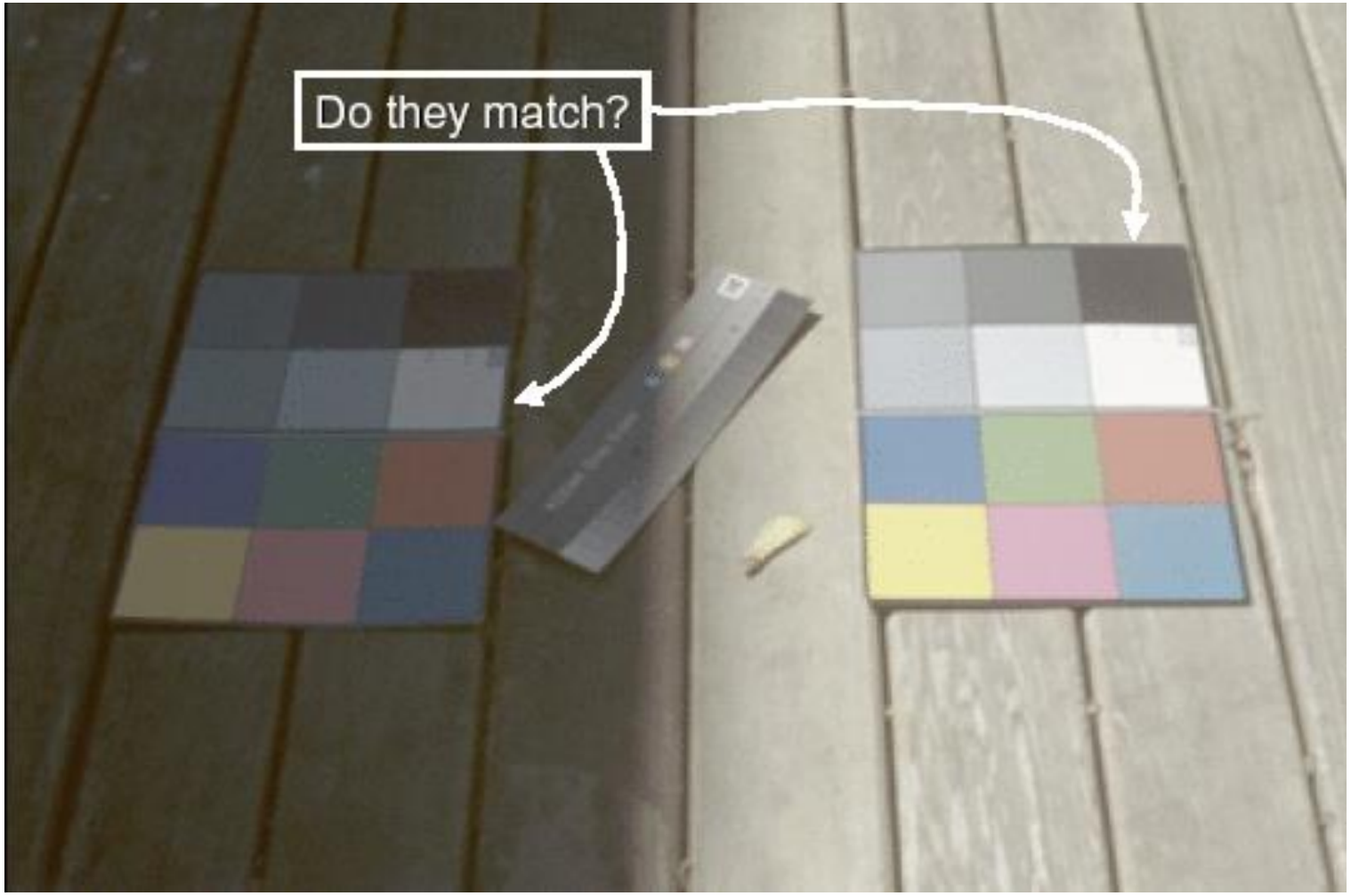
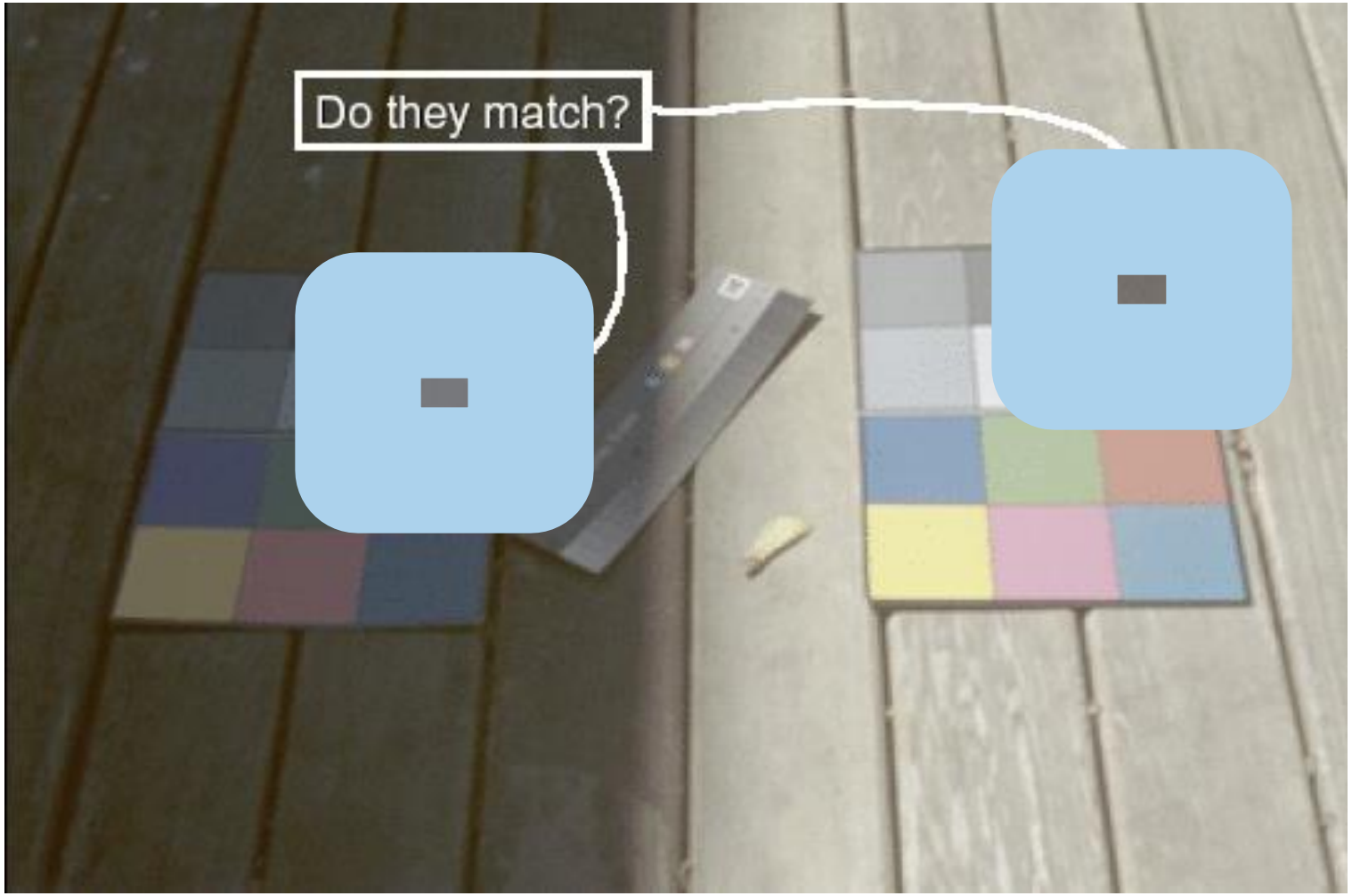


Image courtesy of John McCann via Maureen Stone

Color/Lightness constancy: Illumination conditions



- When we mask out the context
- It's the same color
- One in shadow, one in sunlight
- Automatic white balancing, we are taking into consideration the lamination in the room
- What is around me influences me

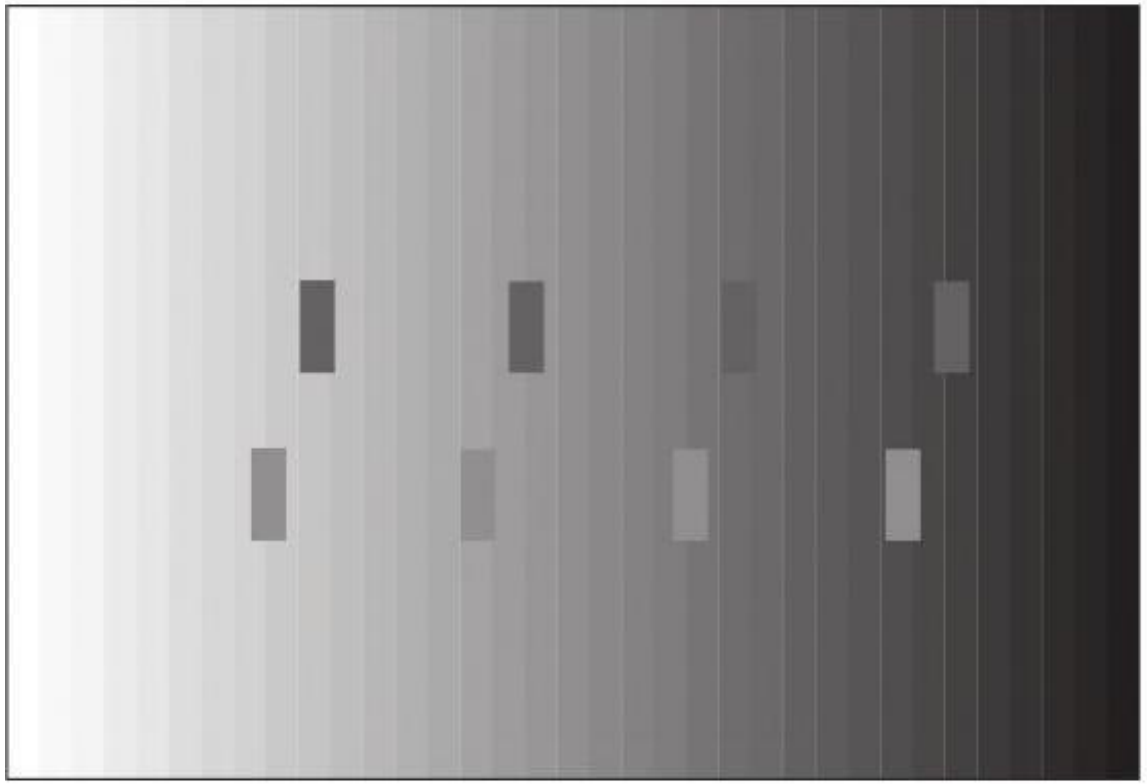
Image courtesy of John McCann via Maureen Stone

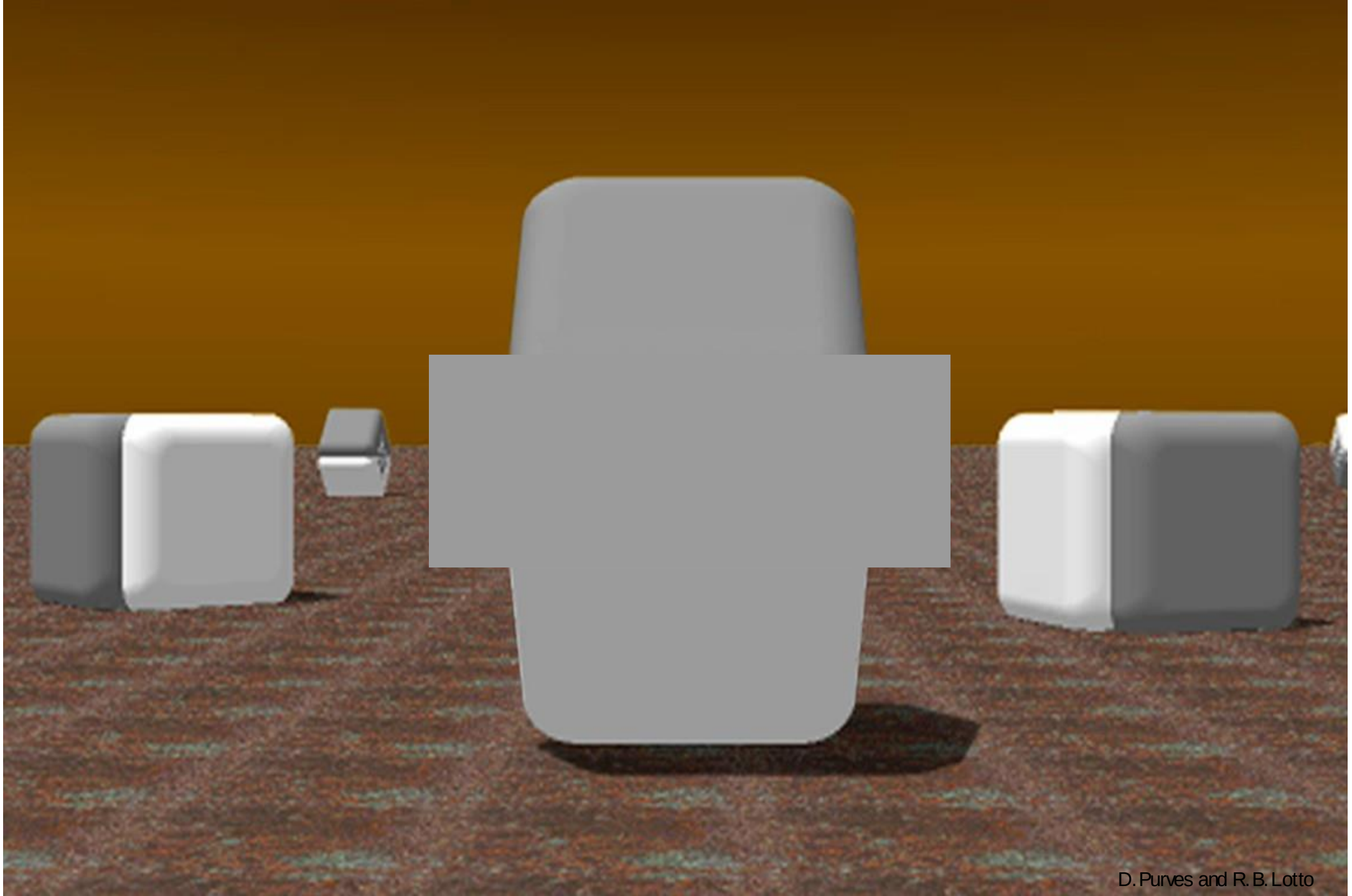
Simultaneous Brightness Contrast

The perceived brightness of an object is relative to it's background

Which bar is the darkest
Which bar is the lightest

Top Row A B C D
Bottom Row E F G H




D. Purves and R. B. Lotto