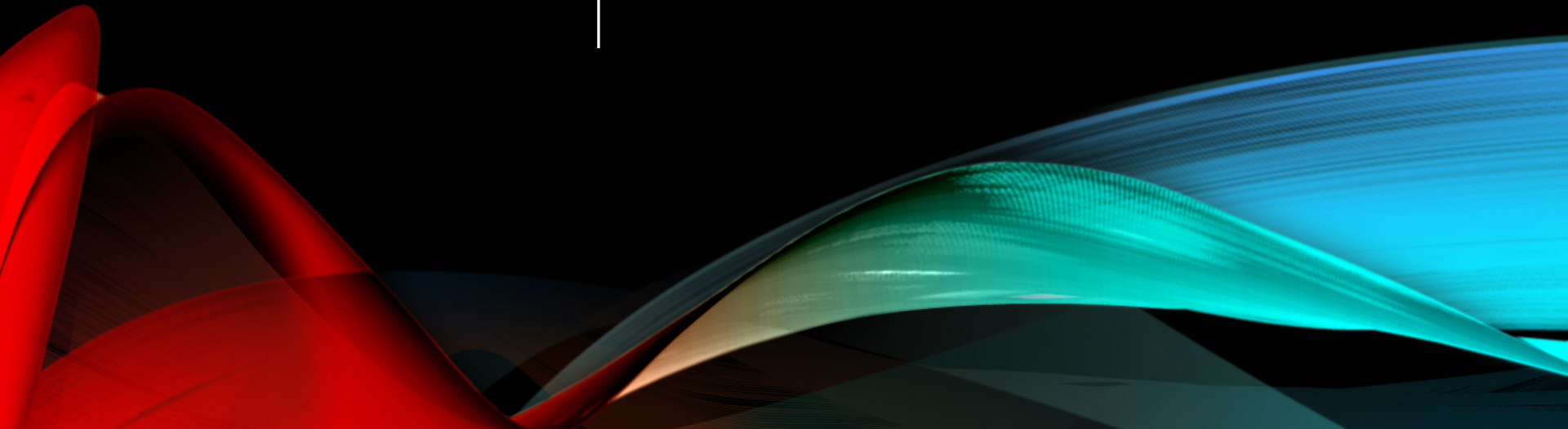


Isara Anantavrasilp

Lecture 11:
Data Visualization

BIG DATA ANALYTICS



DASHBOARD



DASHBOARD

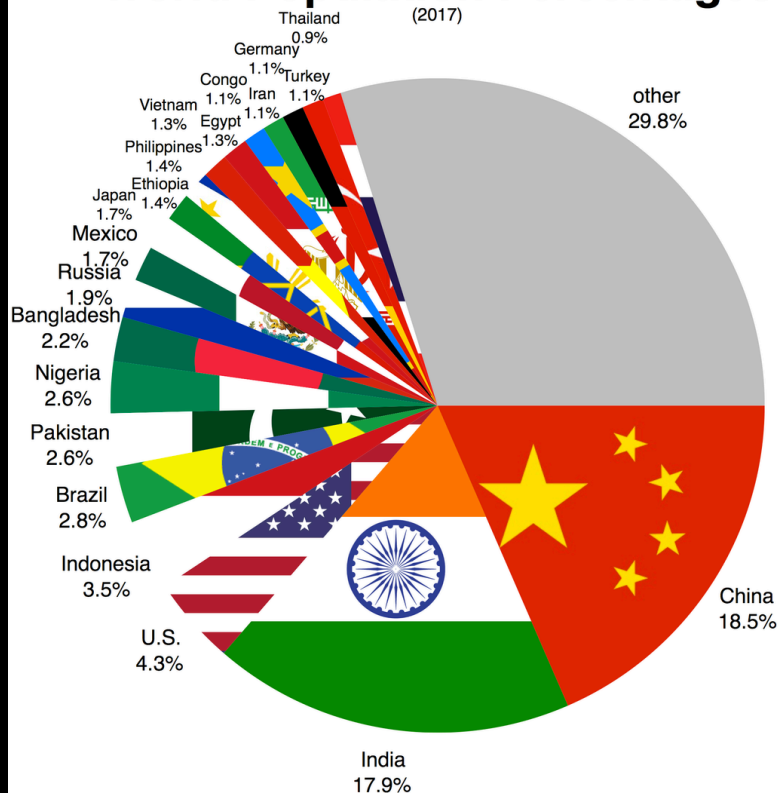


DASHBOARD



WORLD POPULATION

World Population Percentages



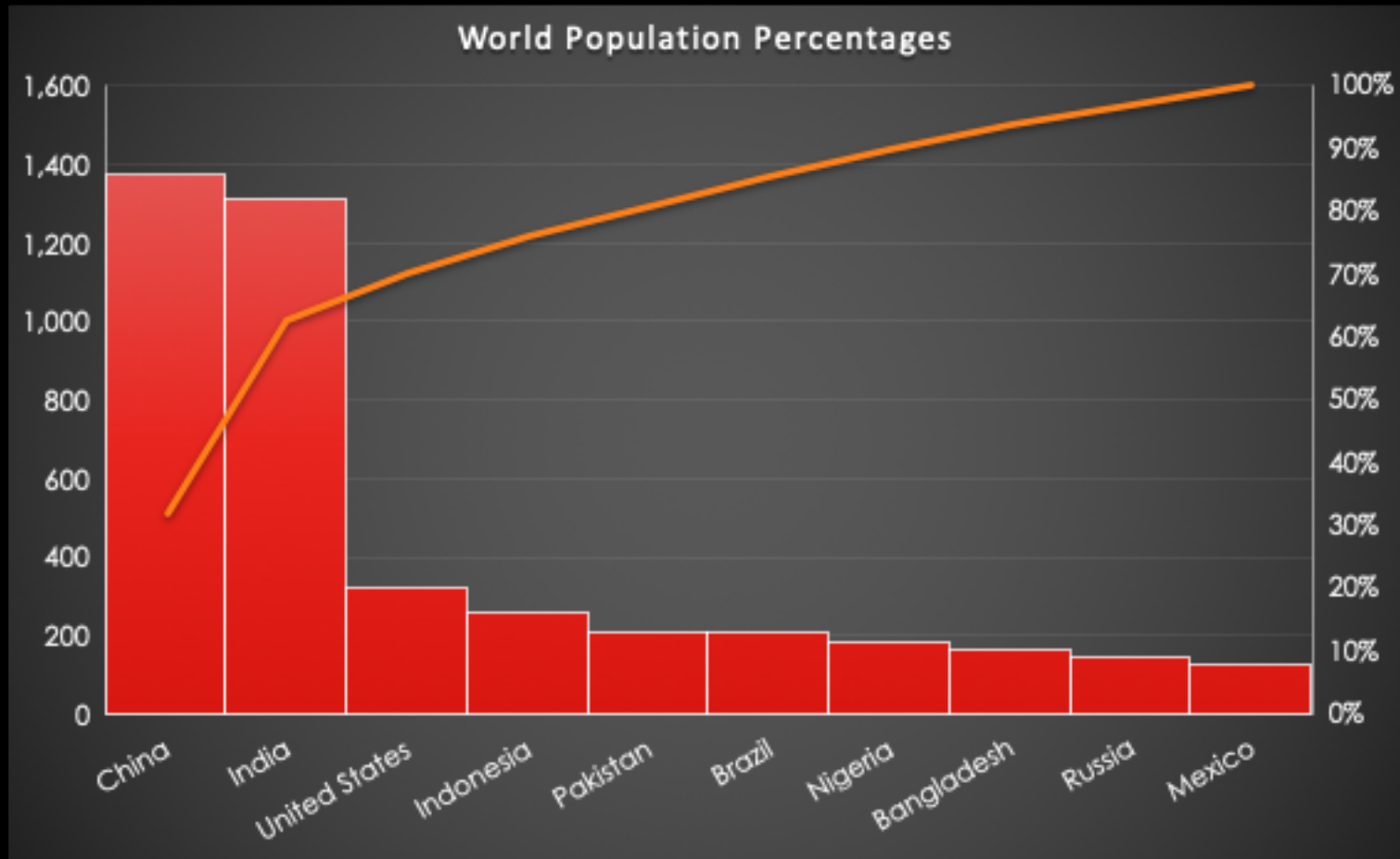
World population (millions, UN estimates)^[13]

#	Top ten most populous countries	2000	2015	2030 ^[A]
1	China ^[B]	1,270	1,376	1,416
2	India	1,053	1,311	1,528
3	United States	283	322	356
4	Indonesia	212	258	295
5	Pakistan	136	208	245
6	Brazil	176	206	228
7	Nigeria	123	182	263
8	Bangladesh	131	161	186
9	Russia	146	146	149
10	Mexico	103	127	148
World total		6,127	7,349	8,501

Notes:

1. ^ China excludes Hong Kong and Macau.
2. ^ 2030 = Medium variant.

WORLD POPULATION



What are the least profitable companies?

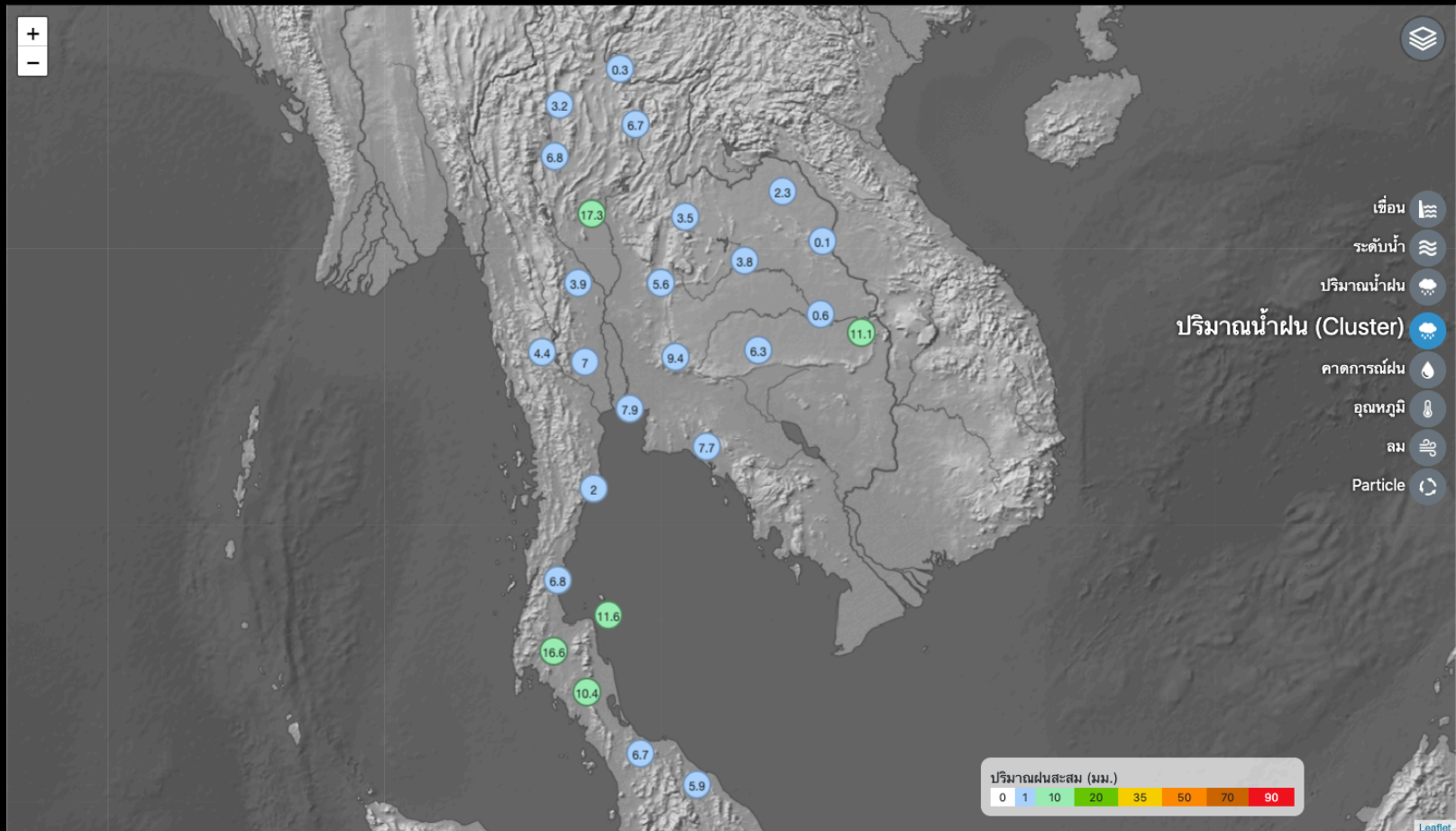
DATE	COMPANY	AMOUNT	PLANNED	COST	REVENUE
23/4/2013	A. Datum Corporation	\$6,400.00	\$6,200.00	\$4,450.00	\$1,950.00
25/4/2013	Contoso Pharmaceuticals	\$8,200.00	\$8,000.00	\$6,400.00	\$1,800.00
7/5/2013	Consolidated Messenger	\$4,400.00	\$4,200.00	\$2,600.00	\$1,800.00
14/5/2013	Proseware, Inc.	\$5,400.00	\$5,500.00	\$4,500.00	\$900.00
14/5/2013	School of Fine Art	\$5,800.00	\$6,000.00	\$4,500.00	\$1,300.00
29/5/2013	Trey Research	\$6,200.00	\$6,000.00	\$4,500.00	\$1,700.00
10/6/2013	A. Datum Corporation	\$6,900.00	\$7,500.00	\$5,400.00	\$1,500.00
21/6/2013	Contoso Pharmaceuticals	\$7,500.00	\$7,200.00	\$6,500.00	\$1,000.00
6/7/2013	Consolidated Messenger	\$8,700.00	\$8,500.00	\$7,250.00	\$1,450.00
5/8/2013	Proseware, Inc.	\$8,500.00	\$8,300.00	\$7,100.00	\$1,400.00
19/8/2013	School of Fine Art	\$7,900.00	\$7,700.00	\$6,600.00	\$1,300.00
4/9/2013	Trey Research	\$9,100.00	\$8,900.00	\$7,900.00	\$1,200.00
20/9/2013	Contoso Pharmaceuticals	\$5,600.00	\$5,800.00	\$4,500.00	\$1,100.00
25/9/2013	Consolidated Messenger	\$9,300.00	\$9,100.00	\$7,500.00	\$1,800.00
15/10/2013	Proseware, Inc.	\$8,800.00	\$9,350.00	\$7,100.00	\$1,700.00
5/11/2013	School of Fine Art	\$9,100.00	\$9,200.00	\$7,850.00	\$1,250.00
26/11/2013	Trey Research	\$9,000.00	\$10,000.00	\$7,575.00	\$1,425.00
30/11/2013	Trey Research	\$7,500.00	\$8,000.00	\$5,850.00	\$1,650.00
11/12/2013	Contoso Pharmaceuticals	\$9,500.00	\$9,200.00	\$8,500.00	\$1,000.00

SALES REPORT

DATE	COMPANY	AMOUNT	PLANNED	COST	REVENUE
23/4/2013	A. Datum Corporation	\$6,400.00	\$6,200.00	\$4,450.00	\$1,950.00
25/4/2013	Contoso Pharmaceuticals	\$8,200.00	\$8,000.00	\$6,400.00	\$1,800.00
7/5/2013	Consolidated Messenger	\$4,400.00	\$4,200.00	\$2,600.00	\$1,800.00
14/5/2013	Proseware, Inc.	\$5,400.00	\$5,500.00	\$4,500.00	\$900.00
14/5/2013	School of Fine Art	\$5,800.00	\$6,000.00	\$4,500.00	\$1,300.00
29/5/2013	Trey Research	\$6,200.00	\$6,000.00	\$4,500.00	\$1,700.00
10/6/2013	A. Datum Corporation	\$6,900.00	\$7,500.00	\$5,400.00	\$1,500.00
21/6/2013	Contoso Pharmaceuticals	\$7,500.00	\$7,200.00	\$6,500.00	\$1,000.00
6/7/2013	Consolidated Messenger	\$8,700.00	\$8,500.00	\$7,250.00	\$1,450.00
5/8/2013	Proseware, Inc.	\$8,500.00	\$8,300.00	\$7,100.00	\$1,400.00
19/8/2013	School of Fine Art	\$7,900.00	\$7,700.00	\$6,600.00	\$1,300.00
4/9/2013	Trey Research	\$9,100.00	\$8,900.00	\$7,900.00	\$1,200.00
20/9/2013	Contoso Pharmaceuticals	\$5,600.00	\$5,800.00	\$4,500.00	\$1,100.00
25/9/2013	Consolidated Messenger	\$9,300.00	\$9,100.00	\$7,500.00	\$1,800.00
15/10/2013	Proseware, Inc.	\$8,800.00	\$9,350.00	\$7,100.00	\$1,700.00
5/11/2013	School of Fine Art	\$9,100.00	\$9,200.00	\$7,850.00	\$1,250.00
26/11/2013	Trey Research	\$9,000.00	\$10,000.00	\$7,575.00	\$1,425.00
30/11/2013	Trey Research	\$7,500.00	\$8,000.00	\$5,850.00	\$1,650.00
11/12/2013	Contoso Pharmaceuticals	\$9,500.00	\$9,200.00	\$8,500.00	\$1,000.00

SALES REPORT

ACCUMULATIVE RAIN



VISUALIZE IT!

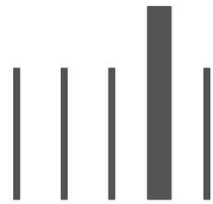
- What we did here was to allow the audience to grasp the important information without / minimal thinking
 - **Seeing = Understanding**
- **Data Visualization:** A process to allow the audiences to understand or capture information from a set of data via visual perception
- The audiences will see only the presented (visualized) the information
 - They might not see every aspect of the data

PREATTENTIVE ATTRIBUTES

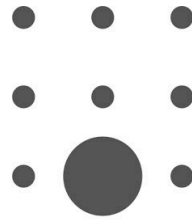
- Human are wired to see visual patterns
 - The eye and visual cortex (at the back of the head) are directly connected
- However, we can see those patterns only in certain ways
- Visual properties that are processed in spatial memory without conscious action is called **Preattentive Attributes** or **Preattentive Visual Properties**
- Colin Ware explains the its concept in his book, **Information Visualization: Perception for Design**



Length



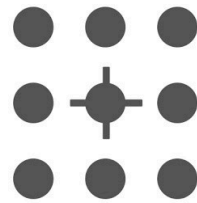
Width



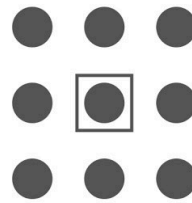
Size



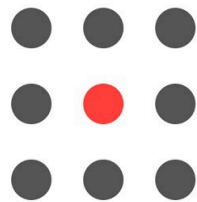
Curvature



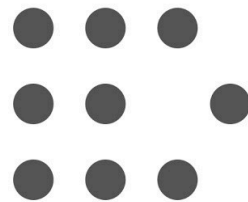
Added Marks



Enclosure



Colour



Position



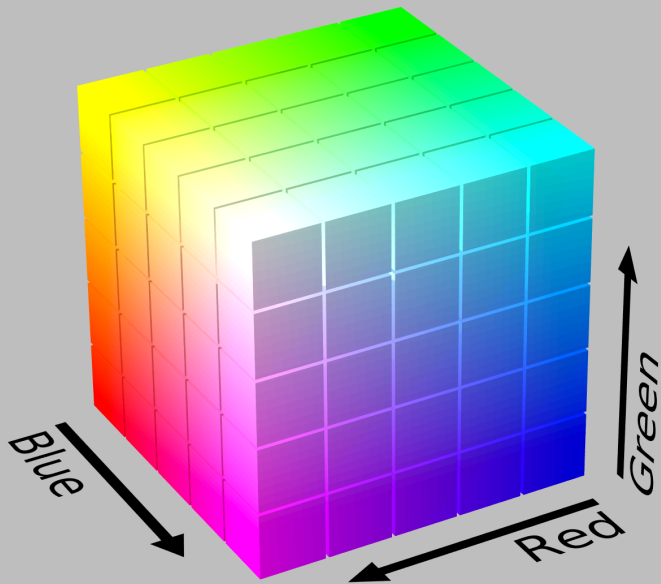
Spatial Grouping

PREATTENTIVE ATTRIBUTES EXAMPLES

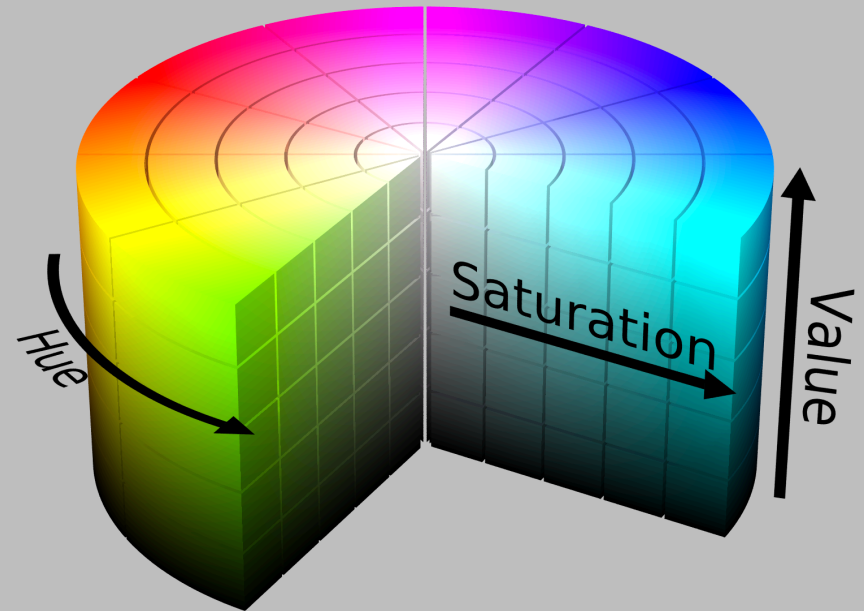
PREATTENTIVE ATTRIBUTES CATEGORIES

1. **Color:** Use the colors to present groups or values of data
2. **Form:** Use shapes, sizes, width, orientations to call attentions to part of visualization
3. **Movement:** User flickers or motion to draw attentions
4. **Spatial Positioning:** Use positions in space to present groups or values of data

COLOR MODELS



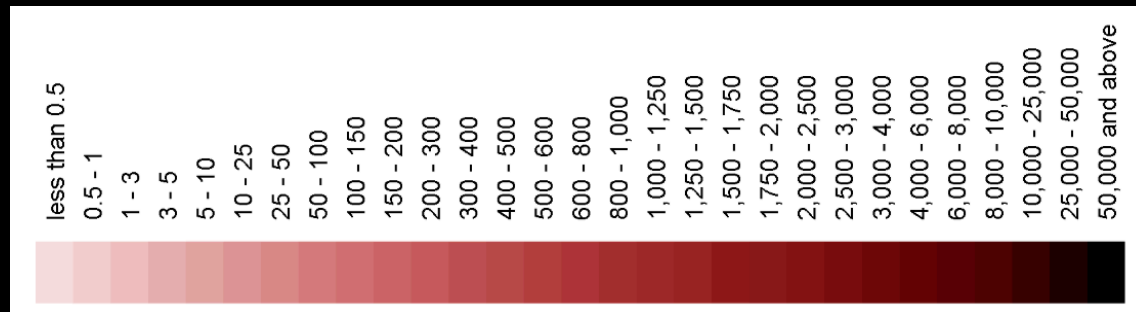
RGB Model



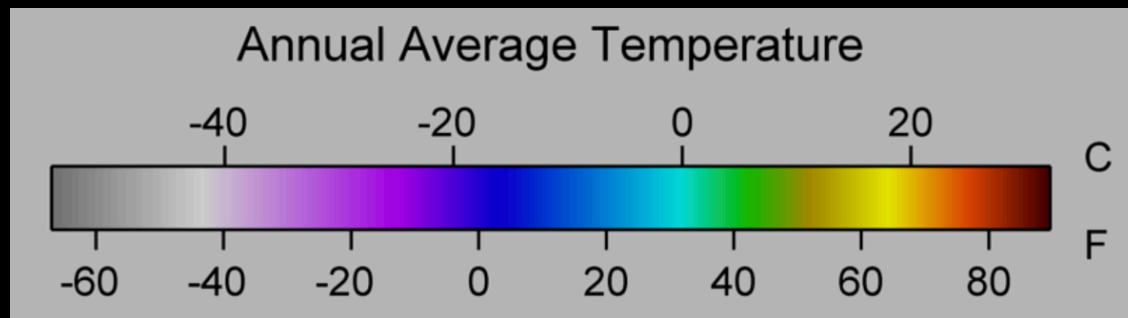
HSV Model

1. COLORS

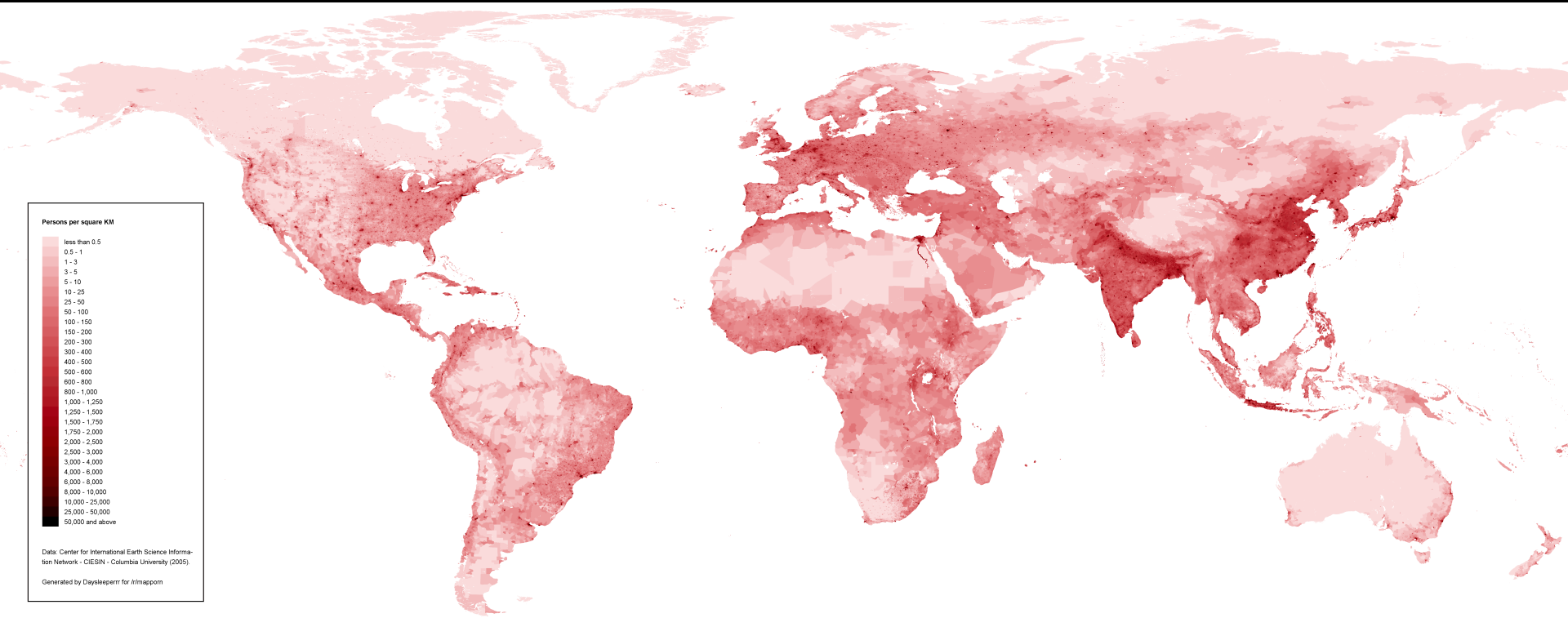
- Colors can be used to present data values
 - Sequential:** Color is ordered from low to high (saturation or lightness/values)



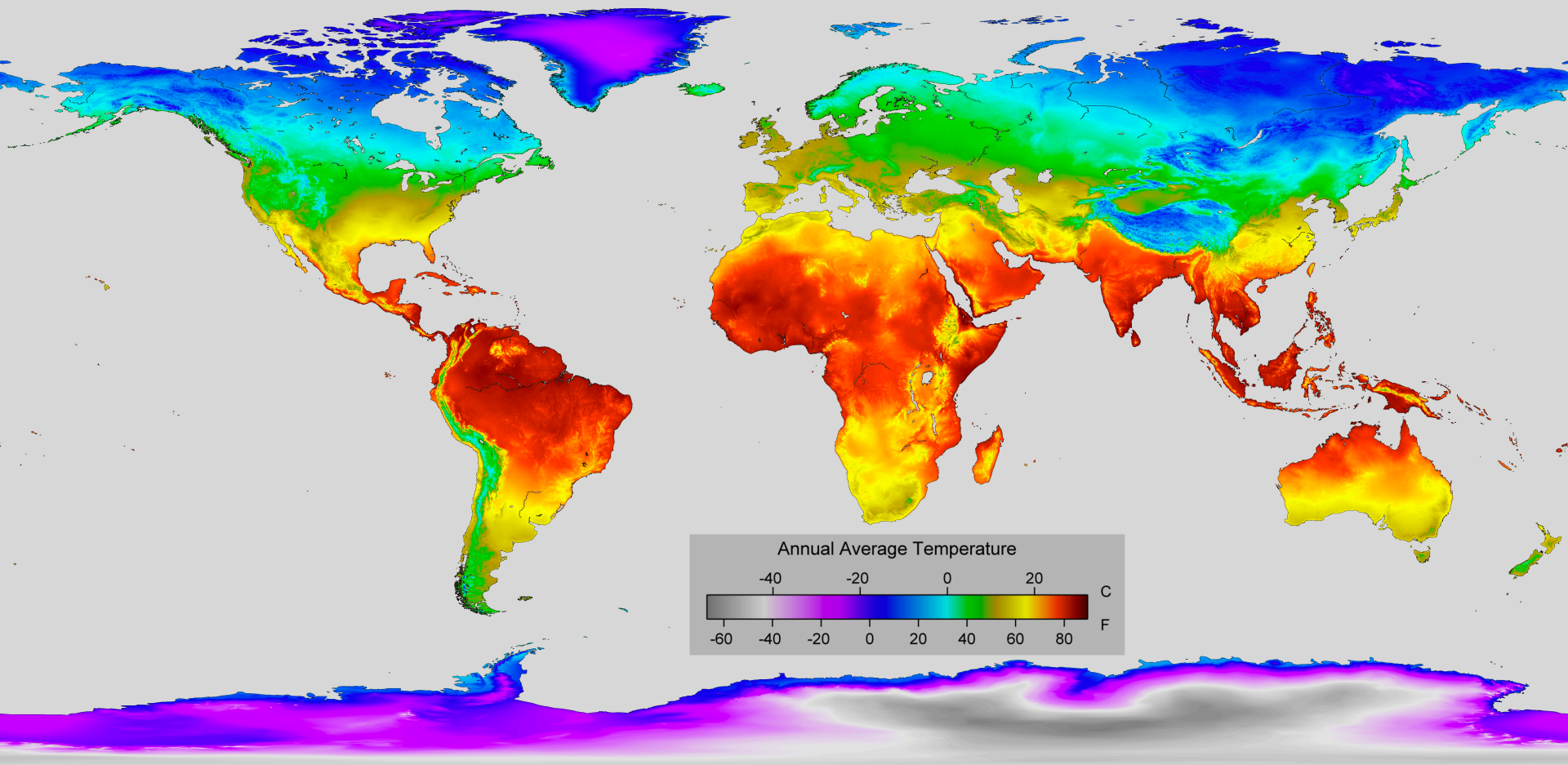
- Diverging:** Two or more sequential colors connected with midpoints

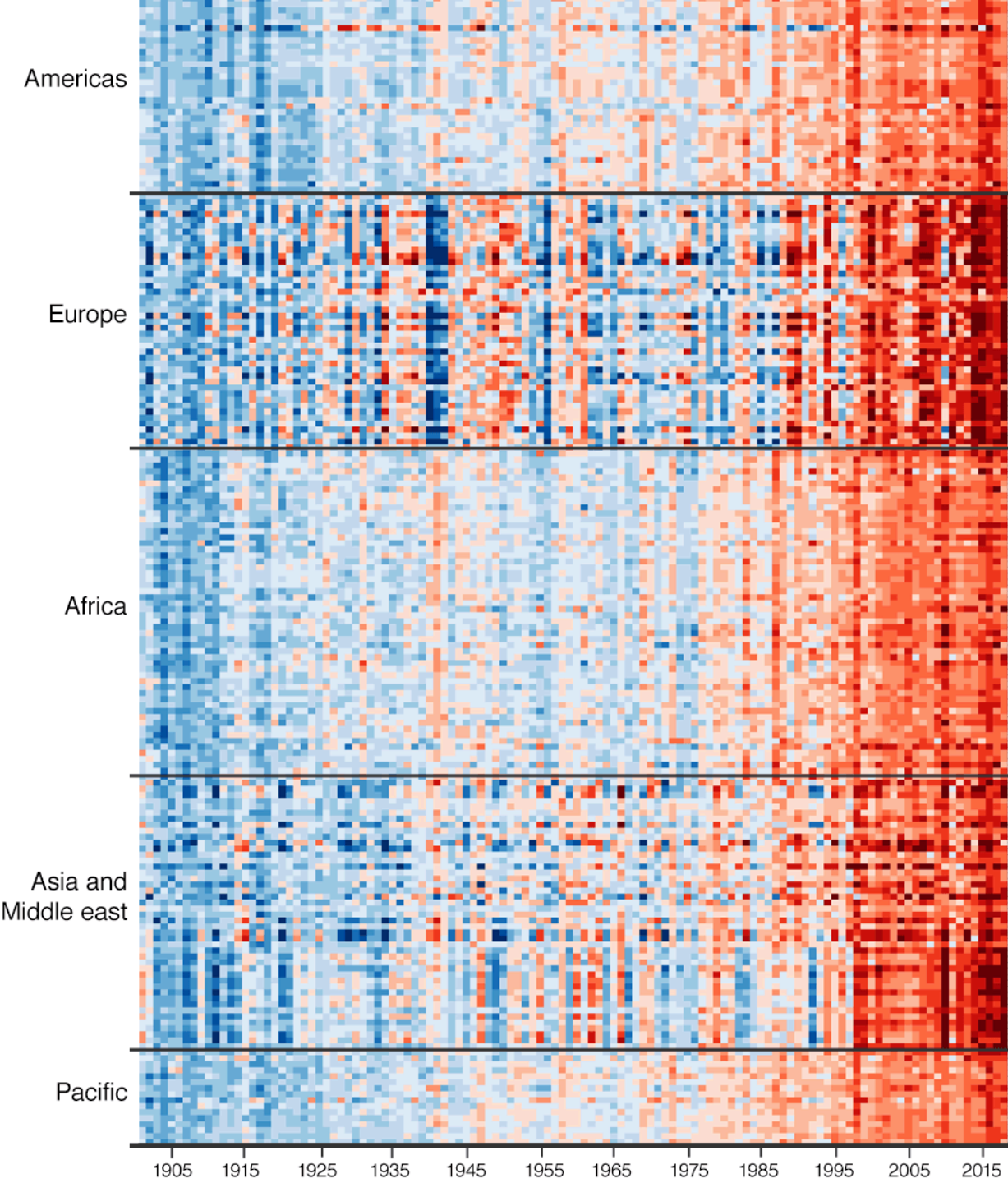


WORLD POPULATION

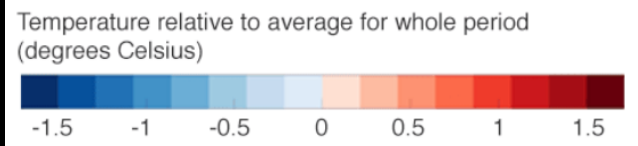


GLOBAL AVERAGE TEMPERATURE





TEMPERATURE CHANGES AROUND THE WORLD

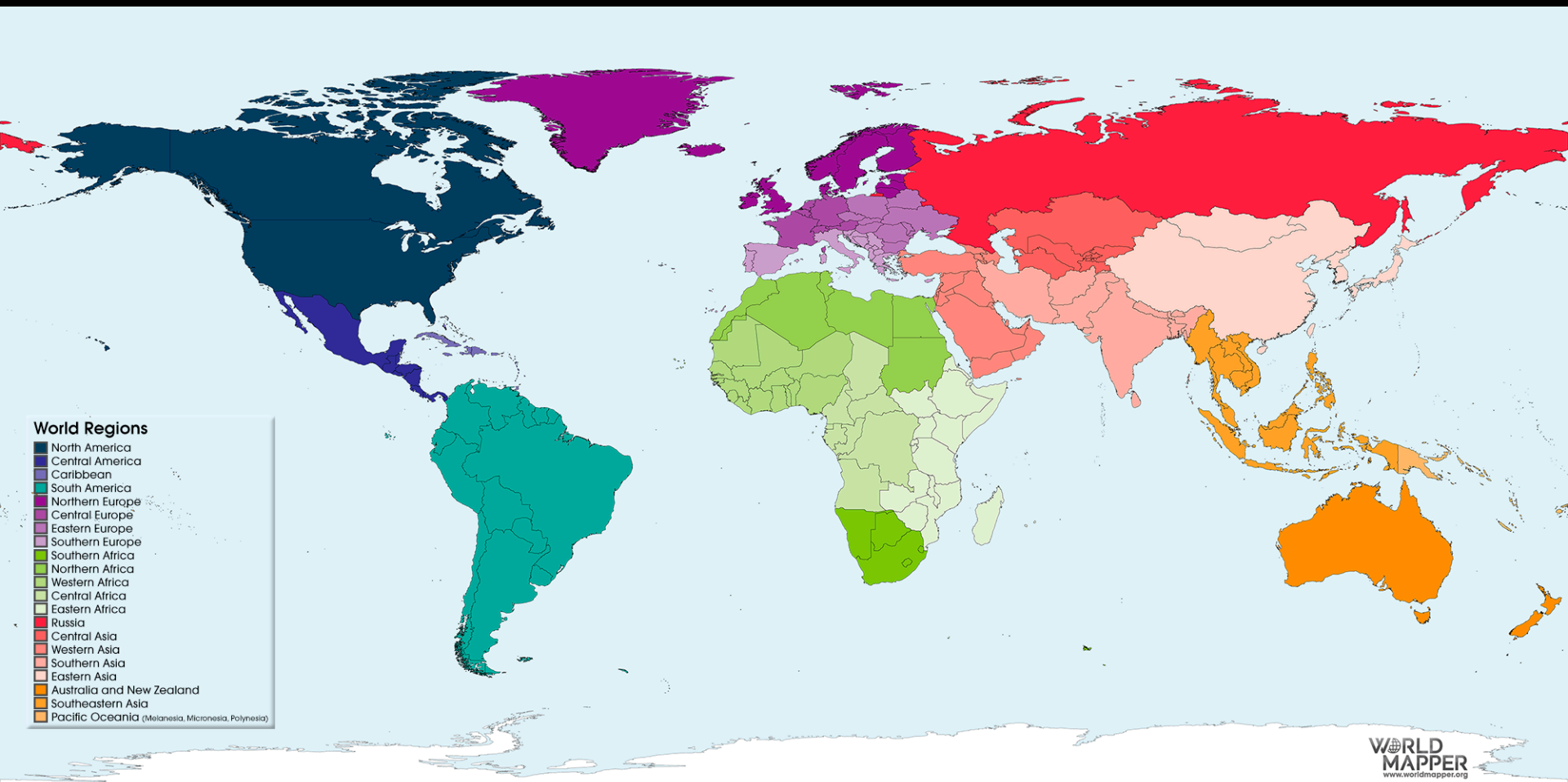


<https://www.bbc.com/news/science-environment-48678196>

1. COLORS

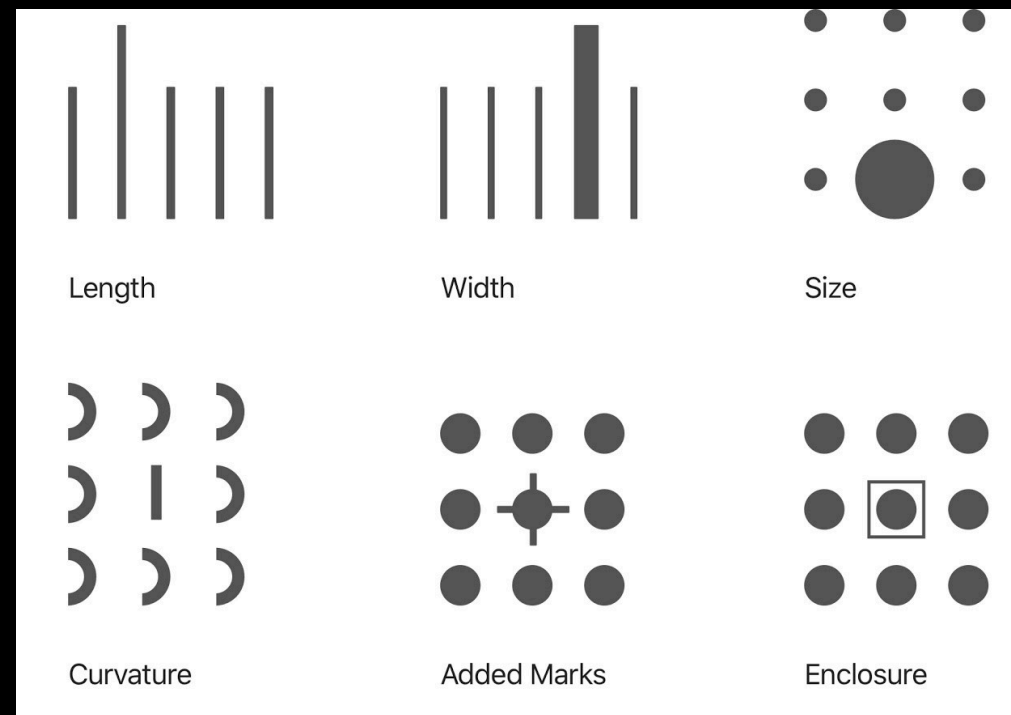
- Colors can be used categorically to present **data group** or **classes**
 - Use contrasting colors for individual comparison
 - Each color represents nominal / distinct values
- We can also use distinguishable colors to **highlights** some values
- Highly distinguishable can also be use to provide **alerts** or **warnings**

WORLD REGION

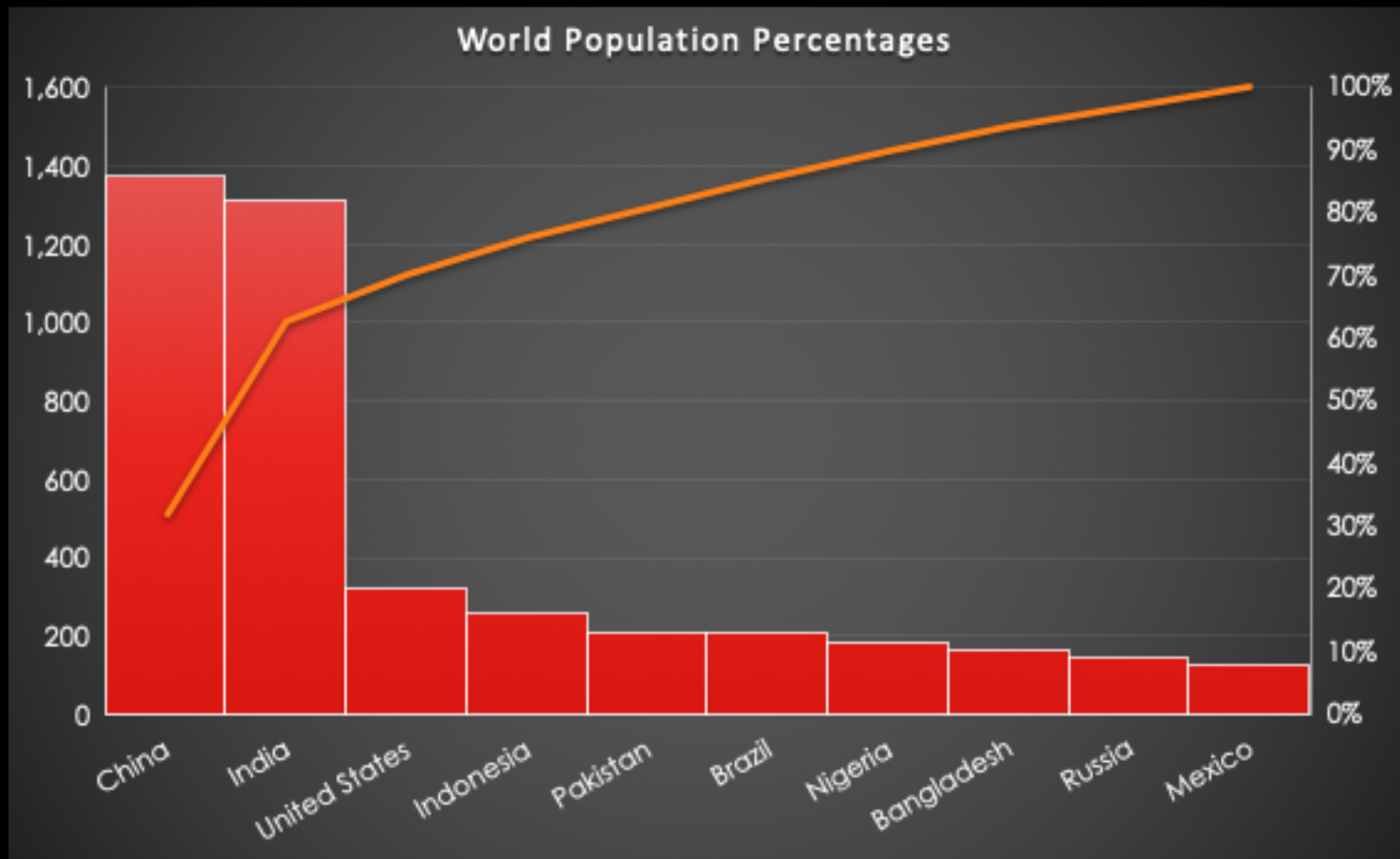


2. FORM

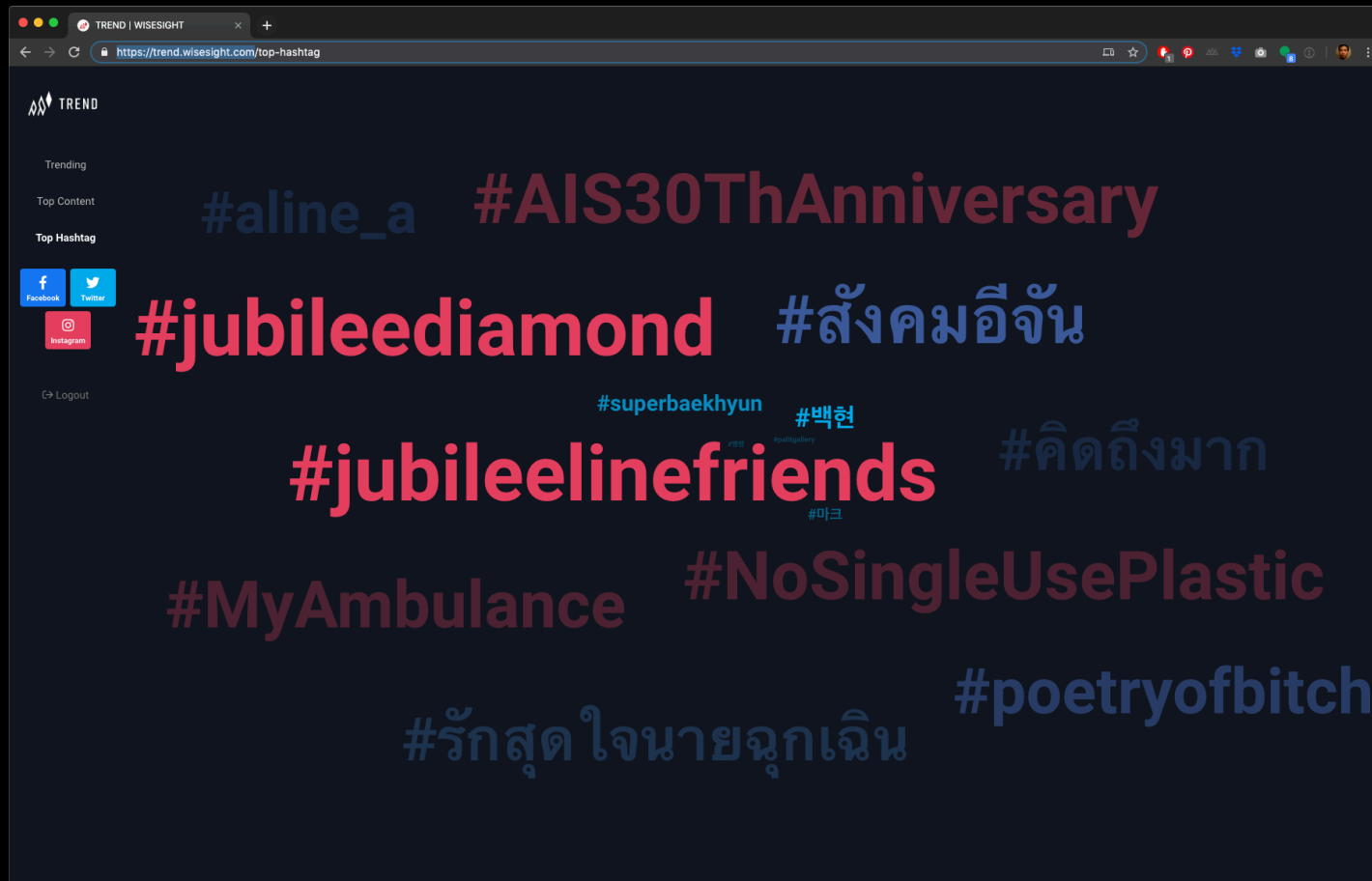
- Distinctive forms, shapes or similar visual attributes can help us differentiate information
 - Length, width, size
 - Added marks
 - Spatial groupings (clusters)



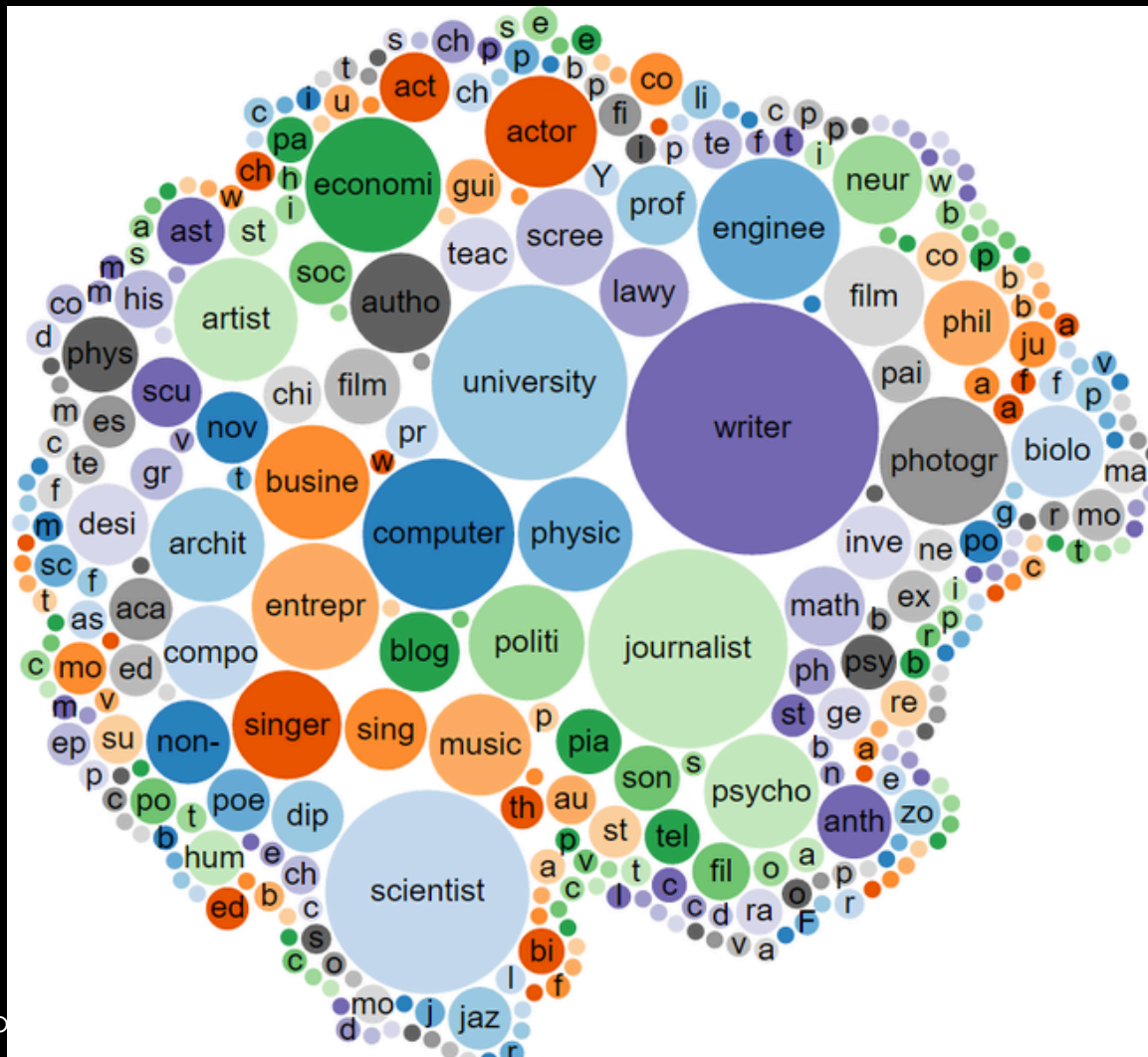
WORLD POPULATION



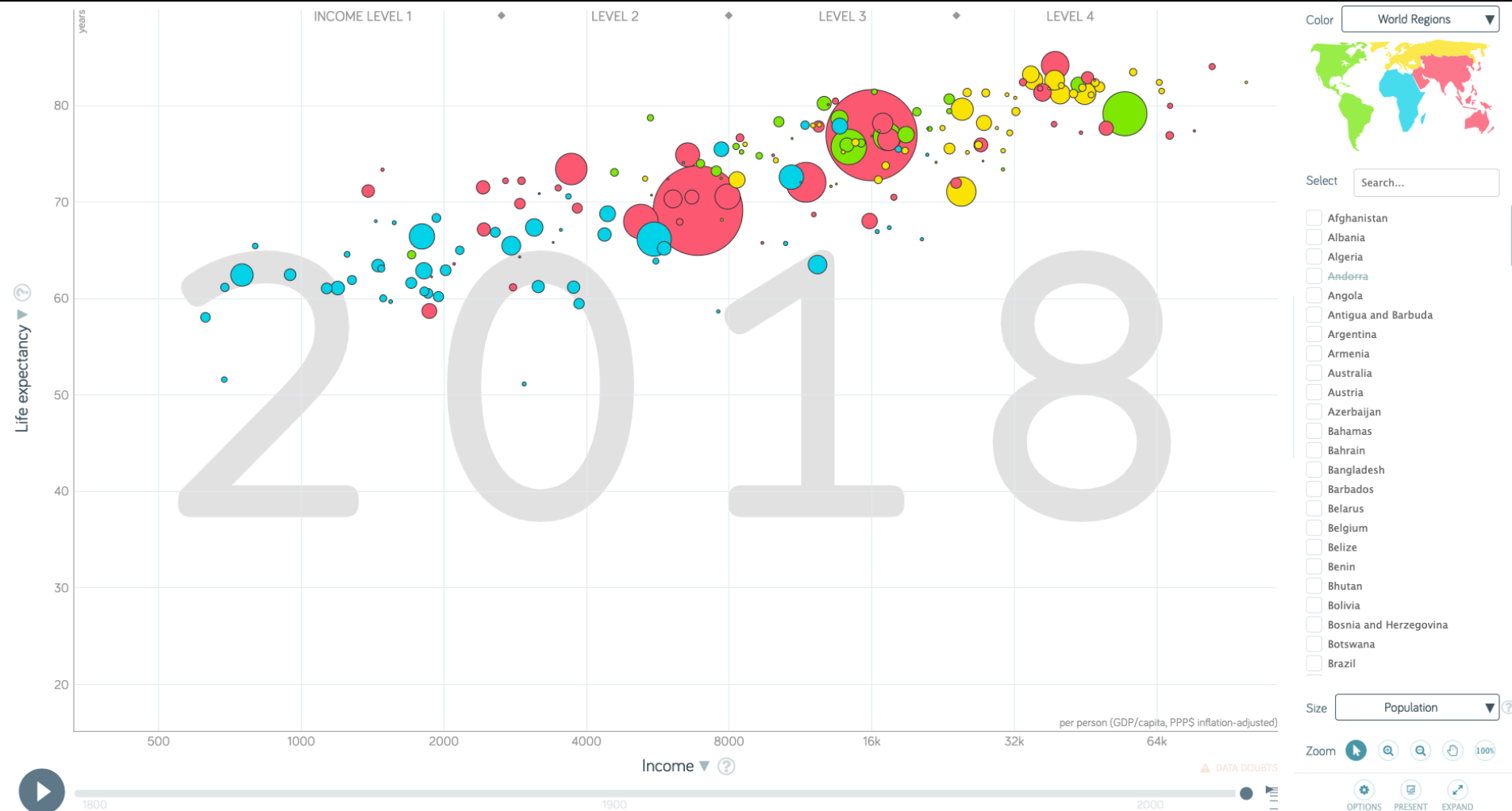
WISESIGHT'S TRENDS



TED SPEAKERS OCCUPATIONS



GAPMINDER



[https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)

3. MOVEMENT

- Instinctively, we are drawn to movements
 - Incoming danger
- Movement can be categorized into
 - Flickers: Flashing or highlighting to draw attention
 - Motion: Moving of objects
- Because we are hard-wired to notice movements, movement can be used to (loudly) called for attention
- It can also be easily over-used
 - Ads
 - Banners
 - Similar effects as the TV in the waiting room



BANNERS



<http://c2sweb.com/welcome-to-c2s-web-planning-your-business-online/>

<https://www.saadsubzwari.com/2017/11/22/alkaram-brings-in-big-savings-on-big-friday/>

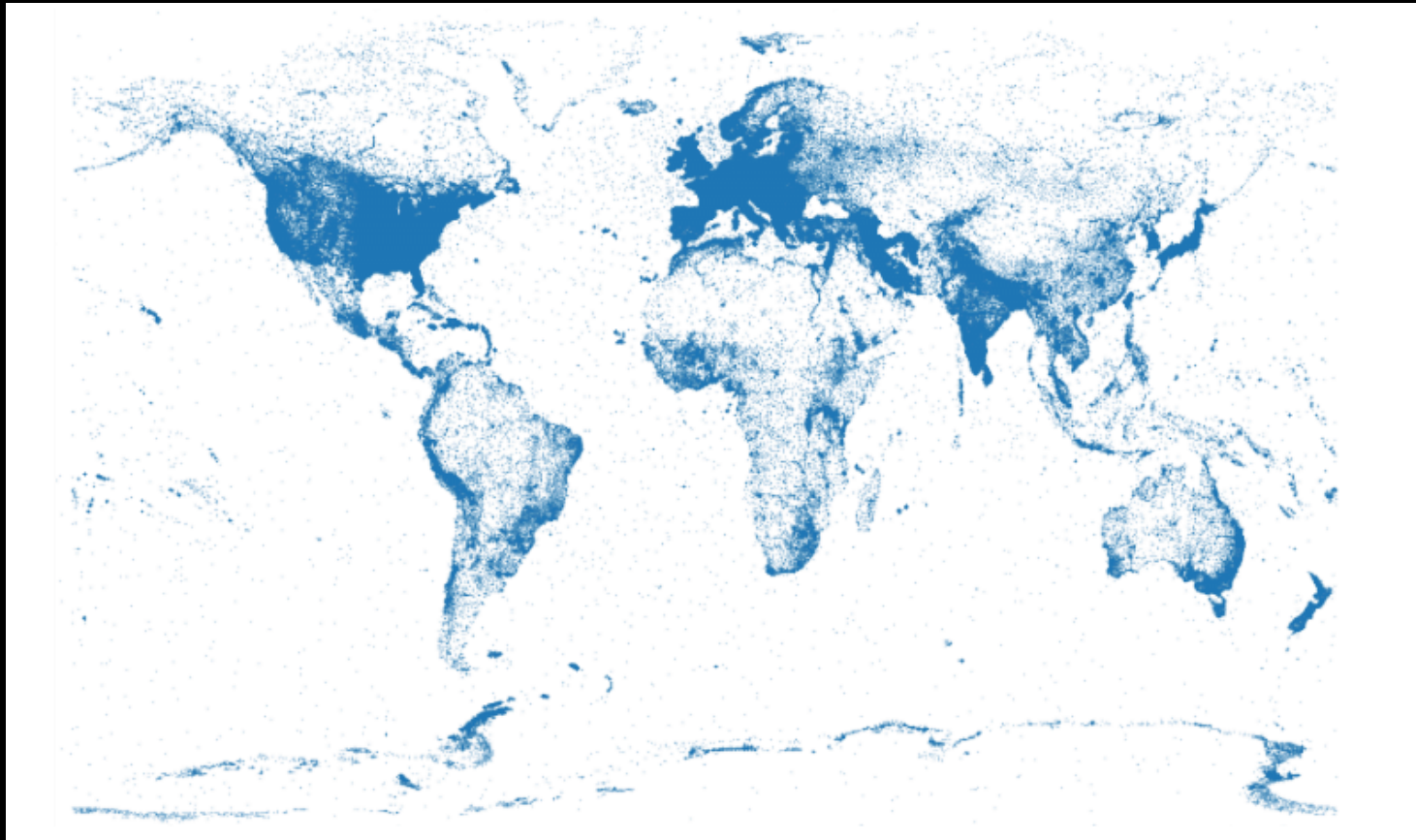
4. SPATIAL POSITIONS

- Spatial positions are positions of objects in space
- Visual positions can convey information
 - **Quantitative values**
 - Groups
- It can also be used to
 - Understand the data or **trends**
 - Detect **outliers** / anomalies
- Examples: Scatter plots, heatmaps, clusters

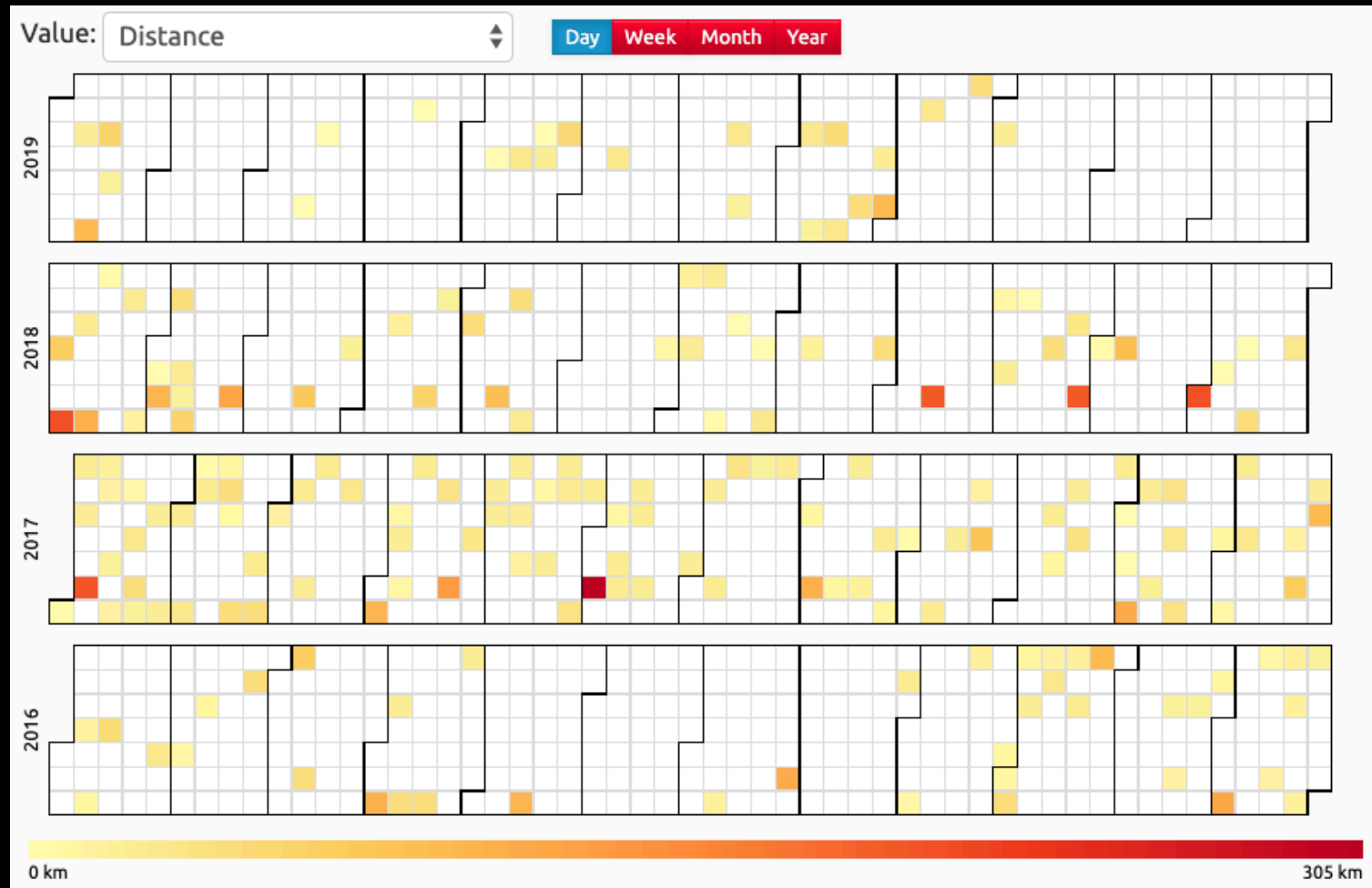
BIKE-SHARING PATHS



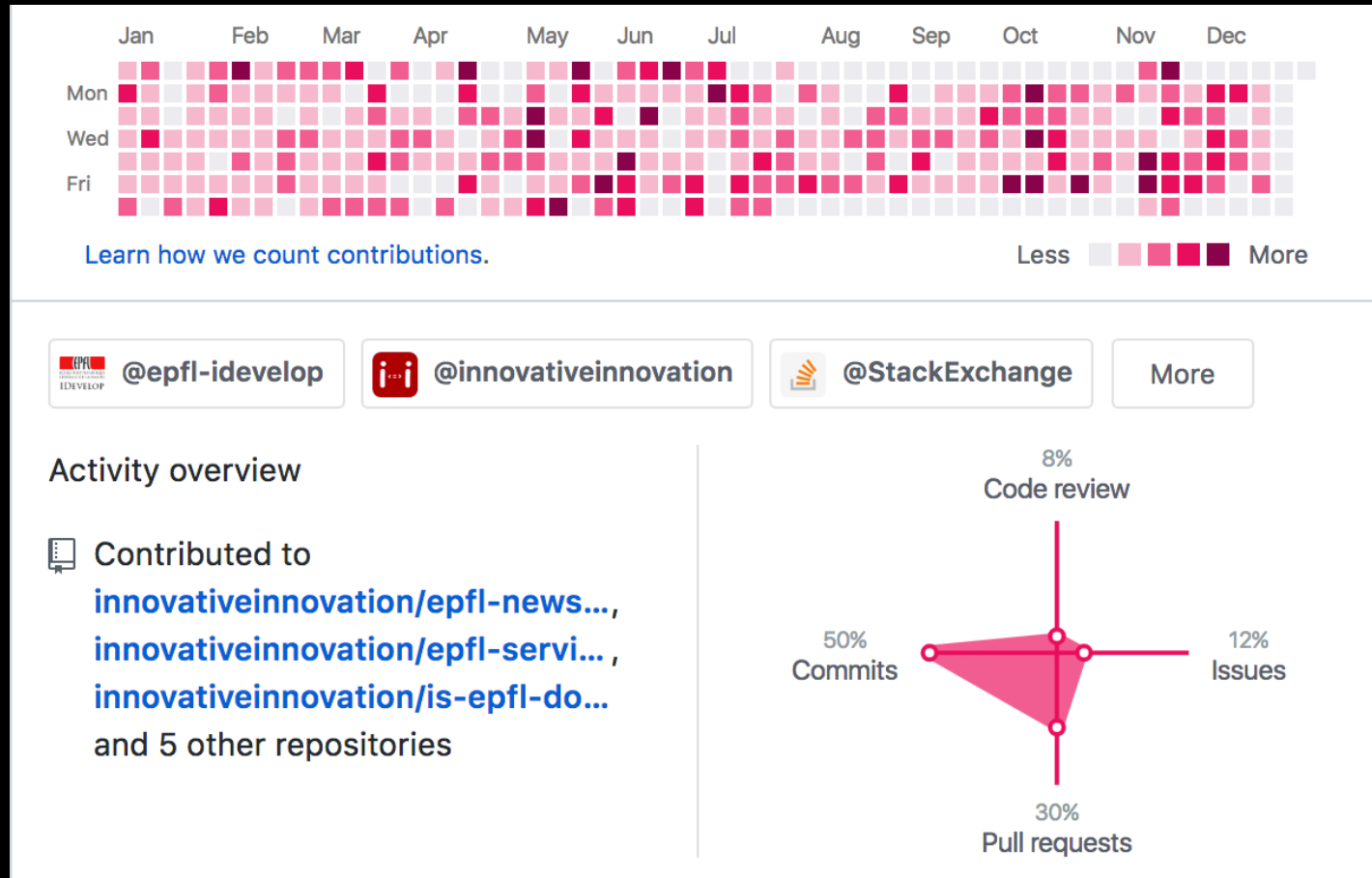
ALL GEOLOCATED WIKIPEDIA ARTICLES



VELOVIEWER ACTIVITY GRAPH



GITHUB ACTIVITY GRAPH

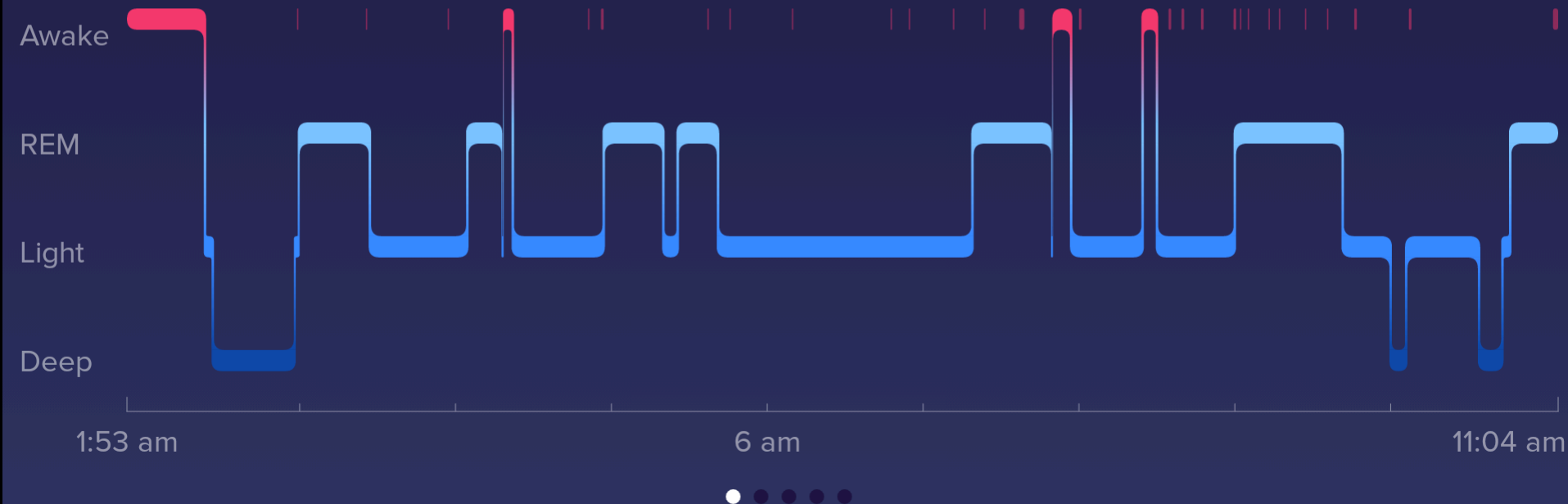


FITBIT ACTIVITY GRAPH

Sleep Stages

Today

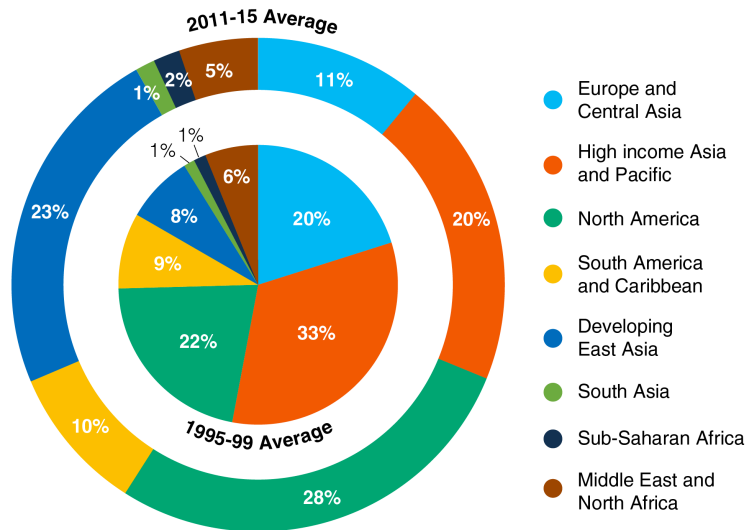
Awake	REM	Light	Deep
12%	29%	50%	8%
1h 8m	2h 40m	4h 36m	47m



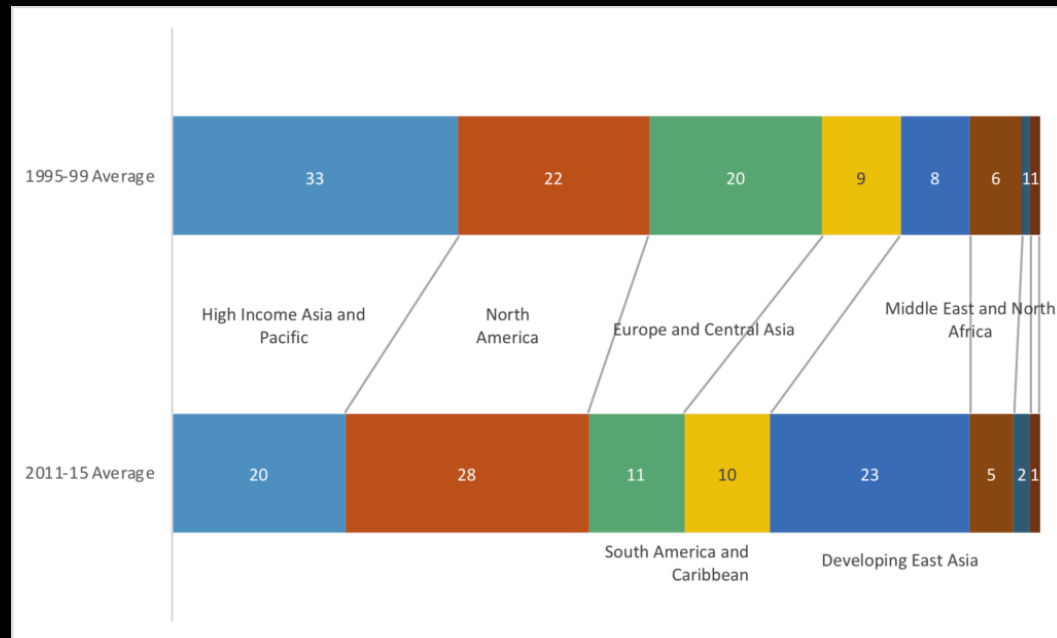
BEST PRACTICE

- Use familiar charts or diagrams that people understand
- Avoid too much details / dimensions

Destinations for U.S. agricultural exports by share of value



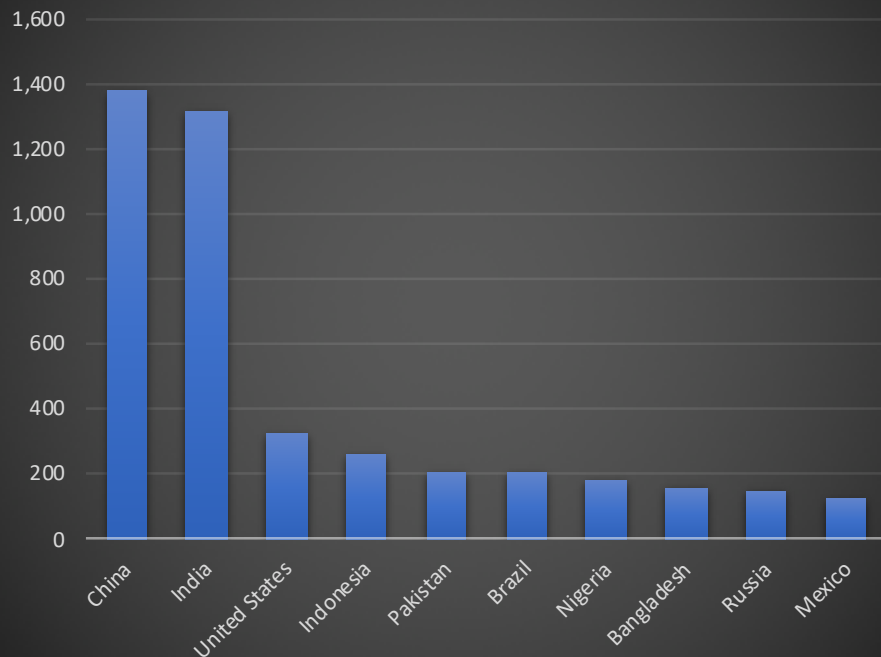
Note: Shares are based on the nominal value of agricultural trade.
Source: USDA, Economic Research Service using United Nations Comtrade Database.



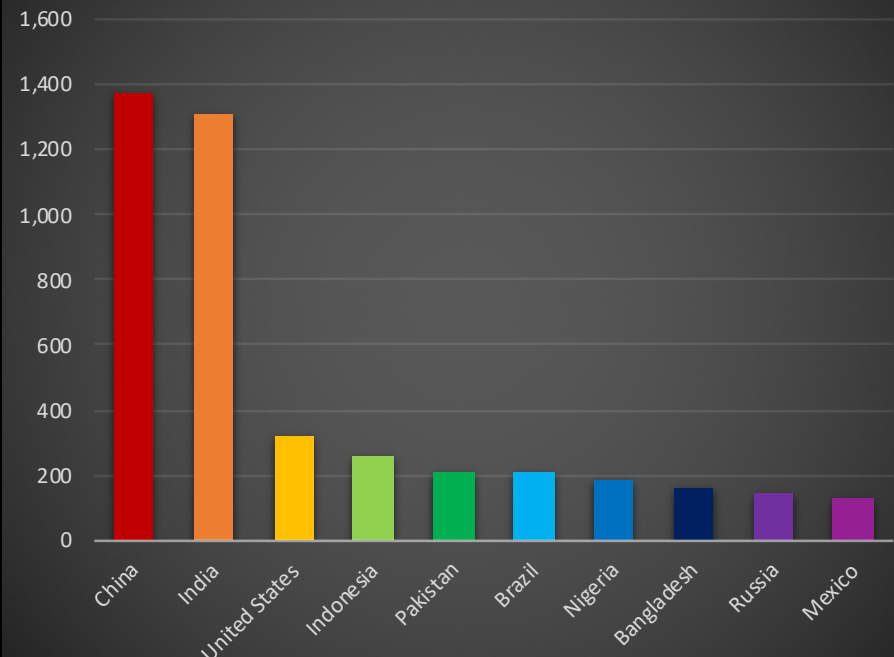
BEST PRACTICE

- Avoid **double encoding**: Use two or more types of visual cues *unintentionally*

World Population (2015)

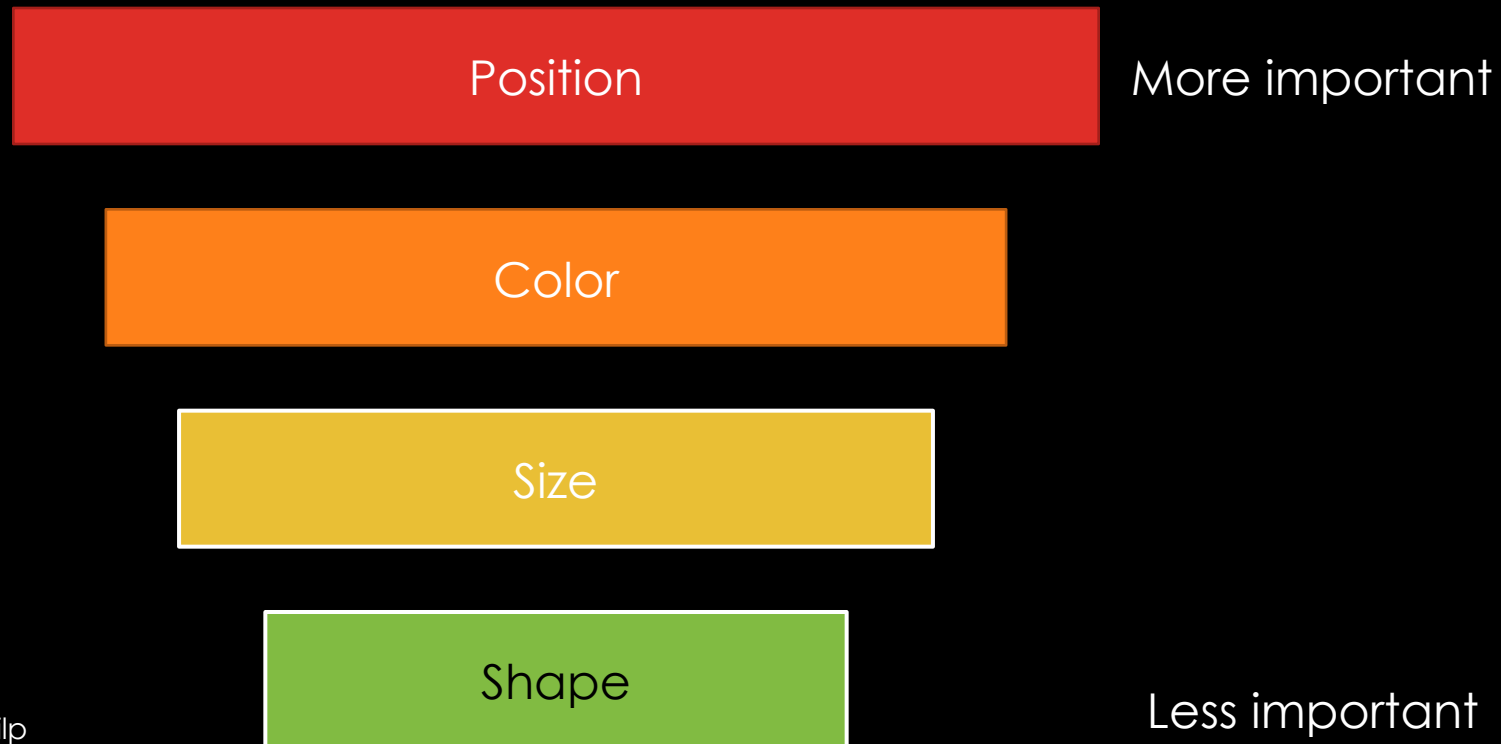


World Population (2015)



BEST PRACTICE

- There are some features that we notice before others
- Hierarchy of visual perception



VISUALIZATION TOOLS

- MS Excel
- Google Sheets
- Tableau
- Apache Superset: <https://superset.incubator.apache.org/>
- RAWGraphs: <https://app.rawgraphs.io/>
- Gapminder: <https://www.gapminder.org/tools-offline/>
- Google Charts: <https://developers.google.com/chart>
- Hicharts: <https://www.highcharts.com/>
- D3.js: <https://www.d3-graph-gallery.com/>

PYTHON VISUALIZATION LIBRARIES

- Matplotlib: <https://matplotlib.org/>
- Pandas: <https://pandas.pydata.org>
- Seaborn: <https://seaborn.pydata.org/>
- Datashader: <http://datashader.org/>
- Plotly: <https://plot.ly/graphing-libraries/>

DATA SOURCES

- Data.gov: <https://catalog.data.gov/dataset>
- US Census Bureau: <http://www.census.gov/data.html>
- Data.gov.sg: <https://data.gov.sg/>
- AWS Public Datasets: <https://registry.opendata.aws/>
- Gapminder: <https://www.gapminder.org/data/>
- Google Public Data:
<https://www.google.com/publicdata/directory>
- Google Ngram Viewer:
<http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>
- EarthData: <https://earthdata.nasa.gov/>
- Data.go.th: <https://data.go.th>