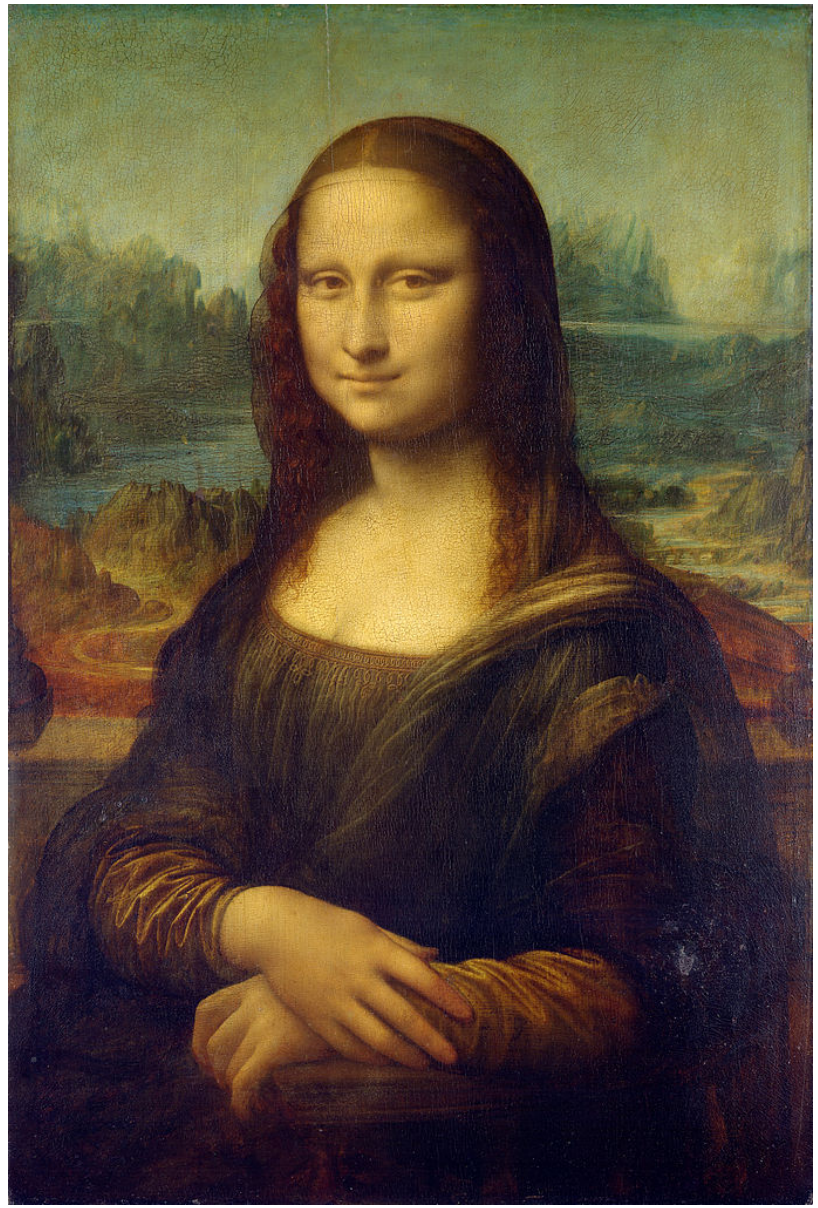
Geographic Data Science

Vector

Elisabetta Pietrostefani & Carmen Cabrera-Arnau

(Geo)visualisation

121 84 93 90 87 76 84 86 82 100 89 84 73 64 79 72 55 66 68 62 63 72 74 67 65 67 64 68 75 72 67 74 74 80 73 77 90 73
107 116 120 72 62 97 101 68 88 101 86 81 77 66 77 78 59 63 64 61 58 58 58 59 61 66 70 73 76 69 67 78 66 83 82 81 86
86 80 93 99 85 86 79 87 90 100 84 82 84 71 72 79 63 61 63 65 59 52 53 59 54 62 69 74 70 58 56 72 54 60 53 54 69 71
86 99 112 113 89 95 76 79 90 95 85 85 92 76 67 78 63 65 67 68 62 58 62 68 61 70 76 78 74 58 55 69 72 74 64 60 76 87
87 107 110 95 78 99 86 89 94 93 86 83 87 78 68 80 67 79 82 72 63 66 69 70 65 67 70 78 76 61 56 69 62 76 78 65 57 58
84 75 80 97 96 85 76 109 103 89 89 81 76 79 77 85 82 97 97 79 67 67 72 69 71 71 72 82 87 75 71 86 80 89 95 87 78 73
96 98 104 93 102 96 81 86 103 86 97 87 72 82 86 82 84 95 96 82 66 62 66 69 84 82 82 94 99 90 89 108 90 83 80 81 80
97 98 100 73 95 94 87 90 98 82 108 98 74 89 89 74 73 76 81 74 60 51 57 69 70 67 69 80 83 73 75 96 82 71 66 60 50 50
83 89 97 88 81 108 80 88 83 85 85 105 80 89 91 71 77 93 78 104 63 68 72 70 64 81 74 76 95 92 64 54 58 44 71 86 98 9
105 75 104 120 77 68 87 112 108 101 91 107 83 80 86 84 81 105 70 81 86 85 71 85 73 61 71 80 71 82 95 87 80 85 93 64
106 72 72 89 97 102 112 76 69 78 79 100 93 87 89 90 89 109 88 88 103 89 70 74 89 77 95 104 83 70 86 89 80 83 91 67
87 139 128 94 95 81 104 92 103 110 91 85 85 89 97 95 102 99 110 102 93 78 83 65 81 89 100 105 114 78 72 101 97 98 1
103 122 89 77 100 71 87 79 102 100 82 87 104 104 112 110 108 95 105 84 83 79 97 88 93 89 90 91 112 79 78 102 76 89
96 103 103 123 113 87 113 100 104 86 83 104 118 91 86 76 100 107 100 74 104 100 87 96 109 79 96 98 102 85 103 91 92
91 109 113 125 105 90 92 75 97 78 77 87 104 104 123 105 94 111 115 96 118 112 84 92 92 74 101 98 110 102 123 103 11
100 122 92 99 109 98 77 99 89 92 98 82 89 106 127 77 100 101 129 118 101 108 104 100 96 106 110 81 112 96 103 105 1
120 104 110 94 91 85 99 85 110 96 113 127 103 124 117 97 109 95 135 89 92 115 91 88 104 105 105 94 97 105 108 127 1
103 96 108 102 103 101 115 103 114 104 116 109 91 118 118 92 105 98 123 94 94 108 98 118 110 111 103 100 101 98 105
83 84 101 100 105 111 127 126 126 95 110 112 101 101 106 102 111 110 106 102 102 103 94 122 106 116 104 120 120 97
89 95 105 100 101 106 120 125 119 88 110 116 115 101 106 106 111 116 89 106 113 108 89 104 104 115 97 120 122 93 10
106 118 120 115 104 102 100 105 103 102 123 103 106 115 117 93 98 113 86 103 115 114 98 104 112 115 86 103 108 84 1
97 118 124 127 116 111 93 96 111 108 120 103 102 110 108 99 97 113 99 99 102 106 108 116 114 116 97 109 113 94 98 9
84 113 122 134 128 123 99 100 126 107 109 118 107 102 94 112 110 108 109 97 100 101 108 117 113 114 110 115 119 110
94 123 128 140 134 130 102 104 121 117 115 120 106 114 98 109 117 94 103 94 106 105 110 112 120 113 105 95 104 112
96 113 118 114 122 127 119 114 117 111 101 118 109 111 96 98 93 105 110 119 109 106 125 109 124 115 107 98 93 102 1
101 111 114 111 118 122 120 119 105 108 104 105 104 115 120 128 92 105 103 106 97 94 111 104 120 116 111 104 98 108
115 121 119 116 121 119 118 125 117 117 110 91 91 98 107 104 100 118 116 115 113 103 111 113 111 113 115 107 102 11
126 129 125 125 124 118 113 121 116 110 110 96 110 106 113 102 98 111 117 120 121 107 100 109 103 110 114 107 104 1
119 127 127 126 125 114 105 111 104 96 106 102 116 103 109 108 91 106 115 118 122 107 94 104 103 109 112 105 101 11
111 122 125 124 124 116 109 111 105 101 107 105 98 85 93 106 110 108 118 123 122 113 99 111 111 112 113 107 102 110
108 124 125 118 117 116 113 113 114 115 109 109 102 107 109 118 114 99 115 122 120 114 101 108 117 112 114 115 108
108 122 120 106 106 109 109 110 130 127 105 105 102 120 107 95 116 95 121 134 131 127 110 110 117 110 115 120 114 1
125 119 103 118 123 119 128 108 117 106 120 101 124 105 116 104 114 102 124 134 104 105 130 121 122 113 107 100 106
110 110 121 116 120 101 115 110 108 109 106 116 116 125 120 121 124 99 126 126 106 113 112 108 125 122 112 117 121
115 101 121 103 120 101 107 105 113 116 105 121 111 128 117 121 114 90 114 110 113 134 118 124 112 115 99 117 126 1
123 106 118 102 114 108 100 111 119 115 113 112 116 115 113 111 127 112 112 108 120 133 119 124 116 127 101 118 129



“Data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color.”

The Visual Display of Quantitative Information. Edward R. Tufte.

Visualization

By encoding information visually, they allow to present large amounts of numbers in a meaninful way. If well made, visualizations provide leads into the processes underlying the graphic.

The Visual Display of Quantitative Information. Edward R. Tufte.

Geovisualization

Tufte (1983)

“The most extensive data maps […] place millions of bits of information on a single page before our eyes. No other method for the display of statistical information is so powerful”

MacEachren (1994)

“Geographic visualization can be defined as the use of concrete visual representations –whether on paper or through computer displays or other media–to make spatial contexts and problems visible, so as to engage the most powerful human information processing abilities, those associated with vision.”

Geovisualization

- Not to replace the human in the loop, but to augment her/him.
- Augmentation through engaging the pattern recognition capabilities that our brain inherently has.
- Combines cartography, infovis and statistics

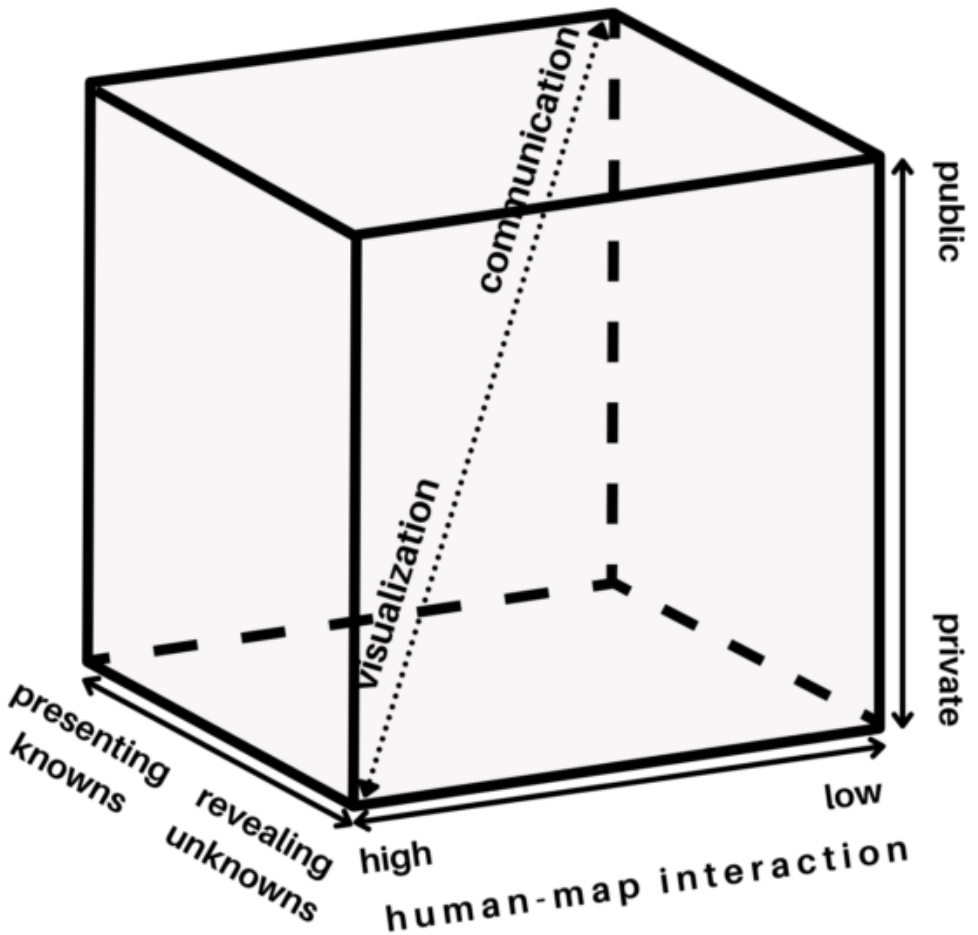
A map for everyone

Maps can fulfill several needs, looking very different depending on the end-goal.

MacEachren & Kraak (1997) identify three main dimensions:

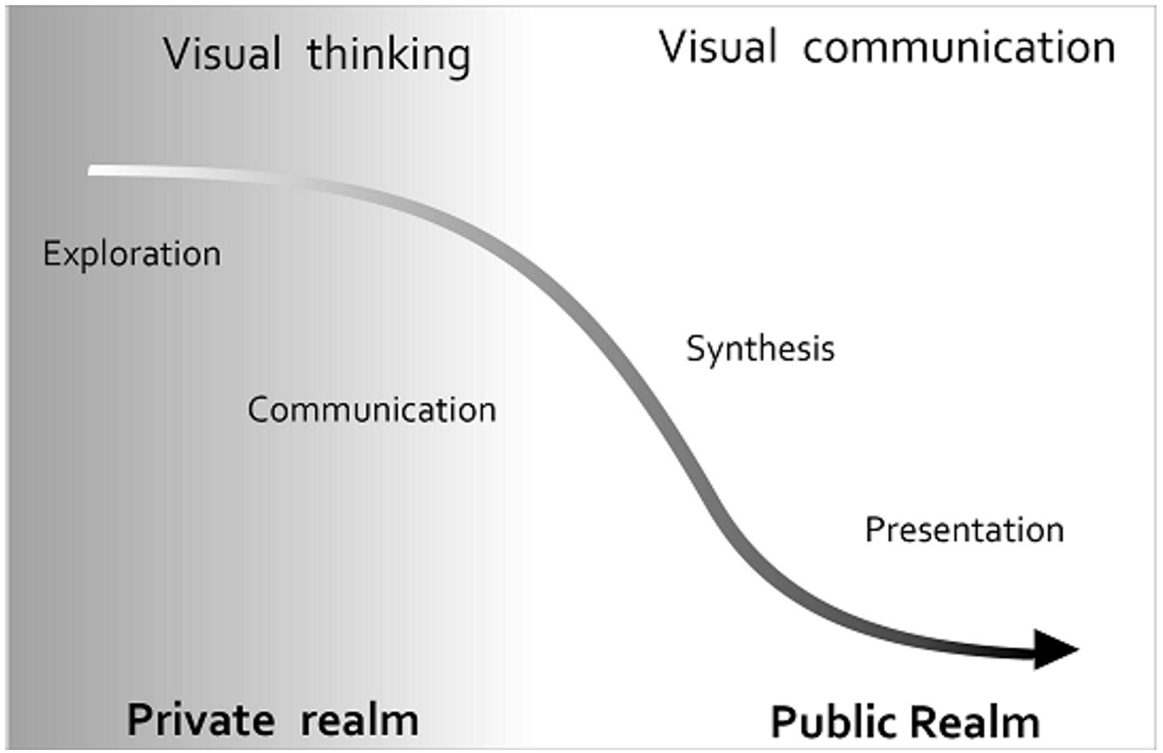
- Knowledge of what is being plotted
- Target audience
- Degree of interactivity

MacEachren & Kraak (1997)



DiBiase's (1990) "Swoopy"

Translating numbers into a (visual) language that the human brain "speaks better"



Exploratory Visualization

“forces us to notice what we never expected to see” (Tukey 1977: vi)

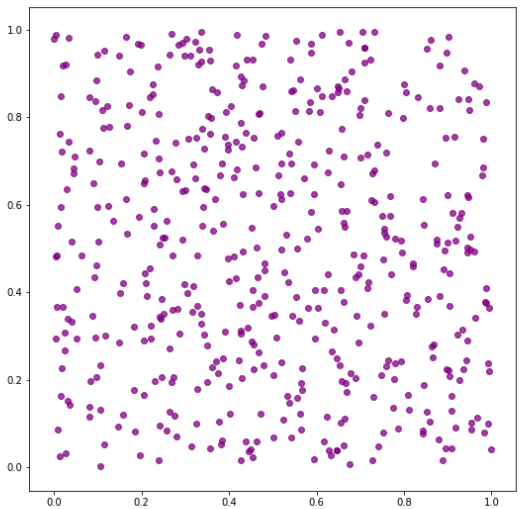
- Mostly for ourselves in the course of the research process.
- Many, quick and dirty, and rather unattractive graphs.

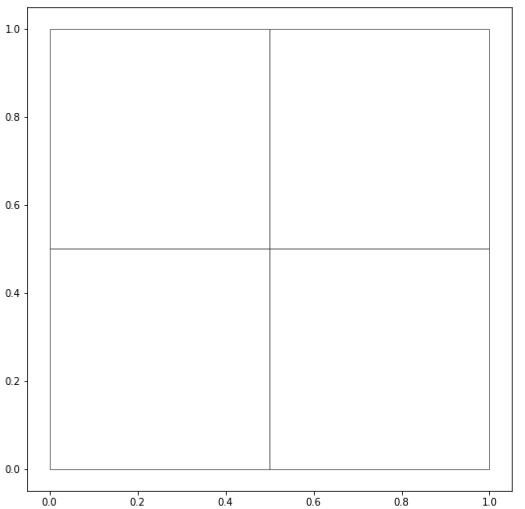
Explanatory Visualization

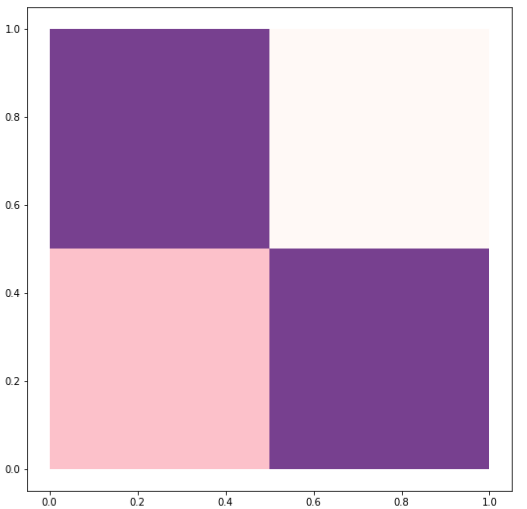
“forces readers to see the information the designer wanted to convey” (Kosslyn 1994: 271)

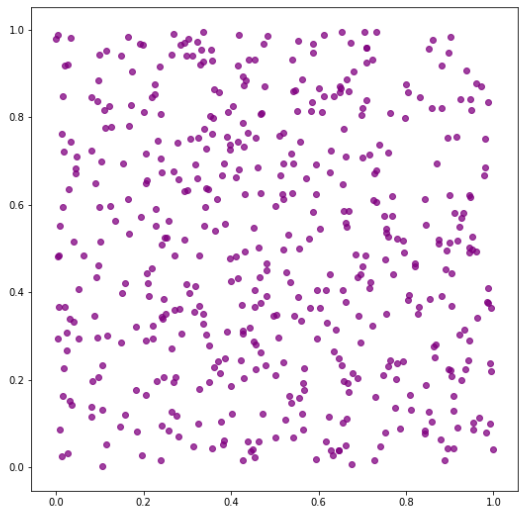
- Mostly for others after the research is completed.
- Few, carefully crafted, and attractive graphs.

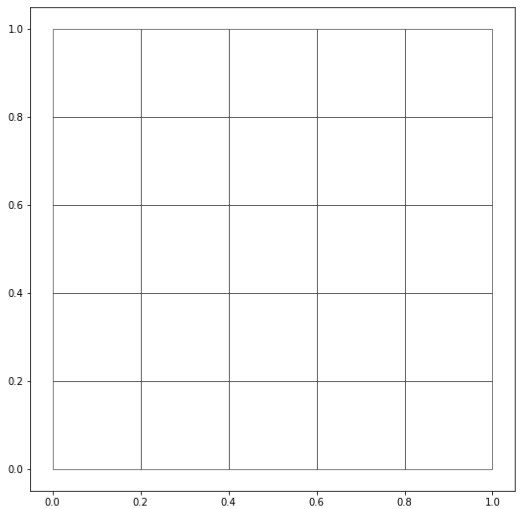
Modifiable Areal Unit Problem (Openshaw, 1984)

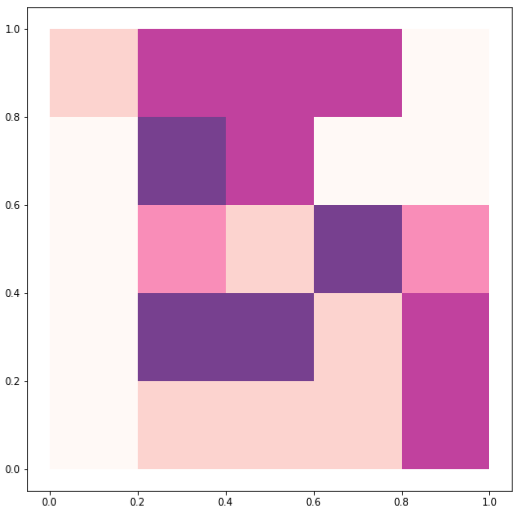

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

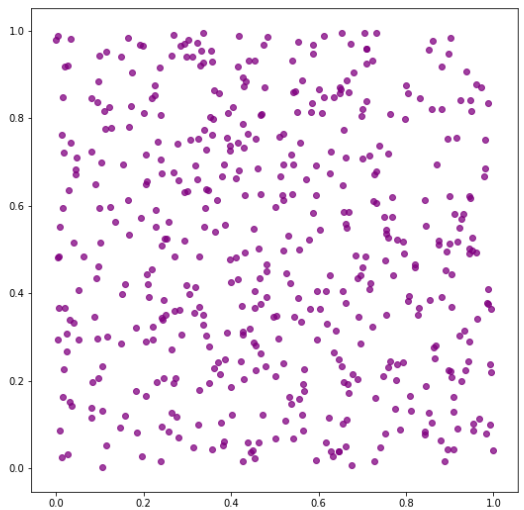

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

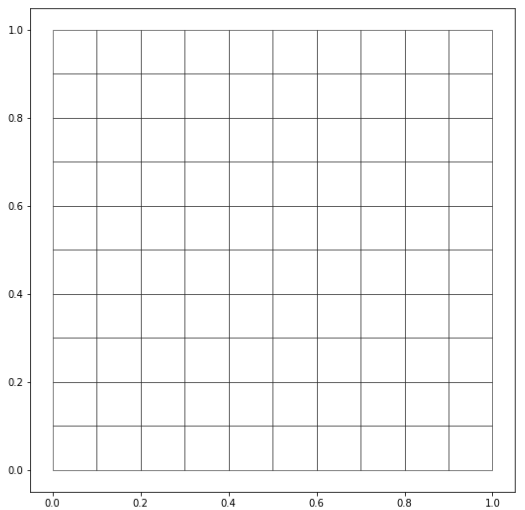

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

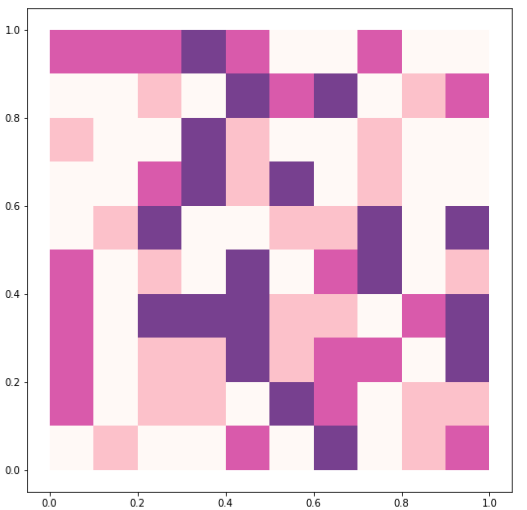

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

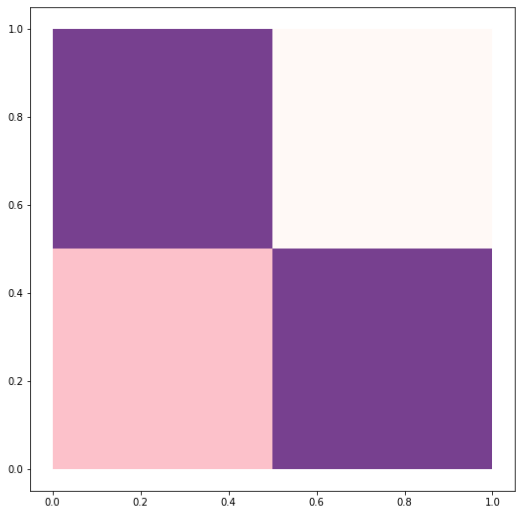

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

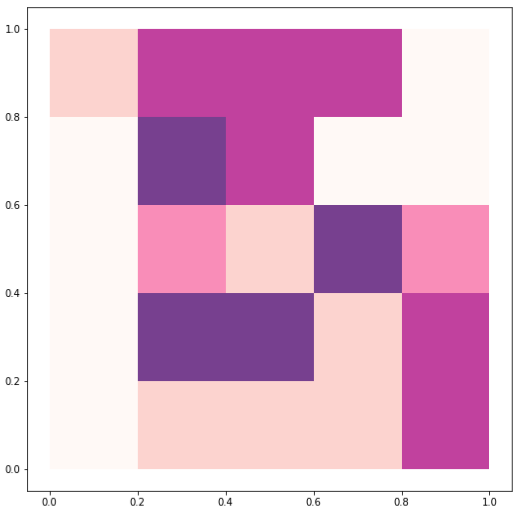

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

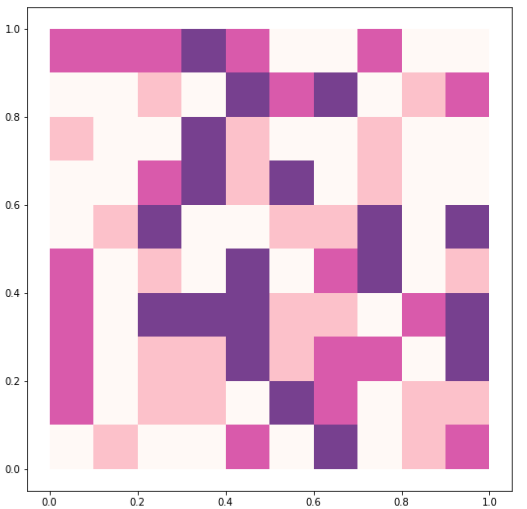

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

MAUP

Scale and delineation mismatch between:

- Underlying process (e.g. individuals, firms, shops)
- Unit of measurement (e.g. neighborhoods, regions, etc.)
- In some cases, it can seriously mislead analysis on aggregated data (e.g. Flint)

Always keep MAUP in mind when exploring aggregated data!!!

Choropleths

Choropleths

Thematic map in which values of a variable are encoded using a color gradient of some sort

- Counterpart of the histogram

  Both allows us to gage the distribution of a variable

- **Values** are classified into specific **colours**: value –> bin
- **Information loss** as a trade off for simplicity

Key decision to be made why a given value is a specific colour!

Classification choices

- N. of bins
- How to bin?
- Colours

How many bins?

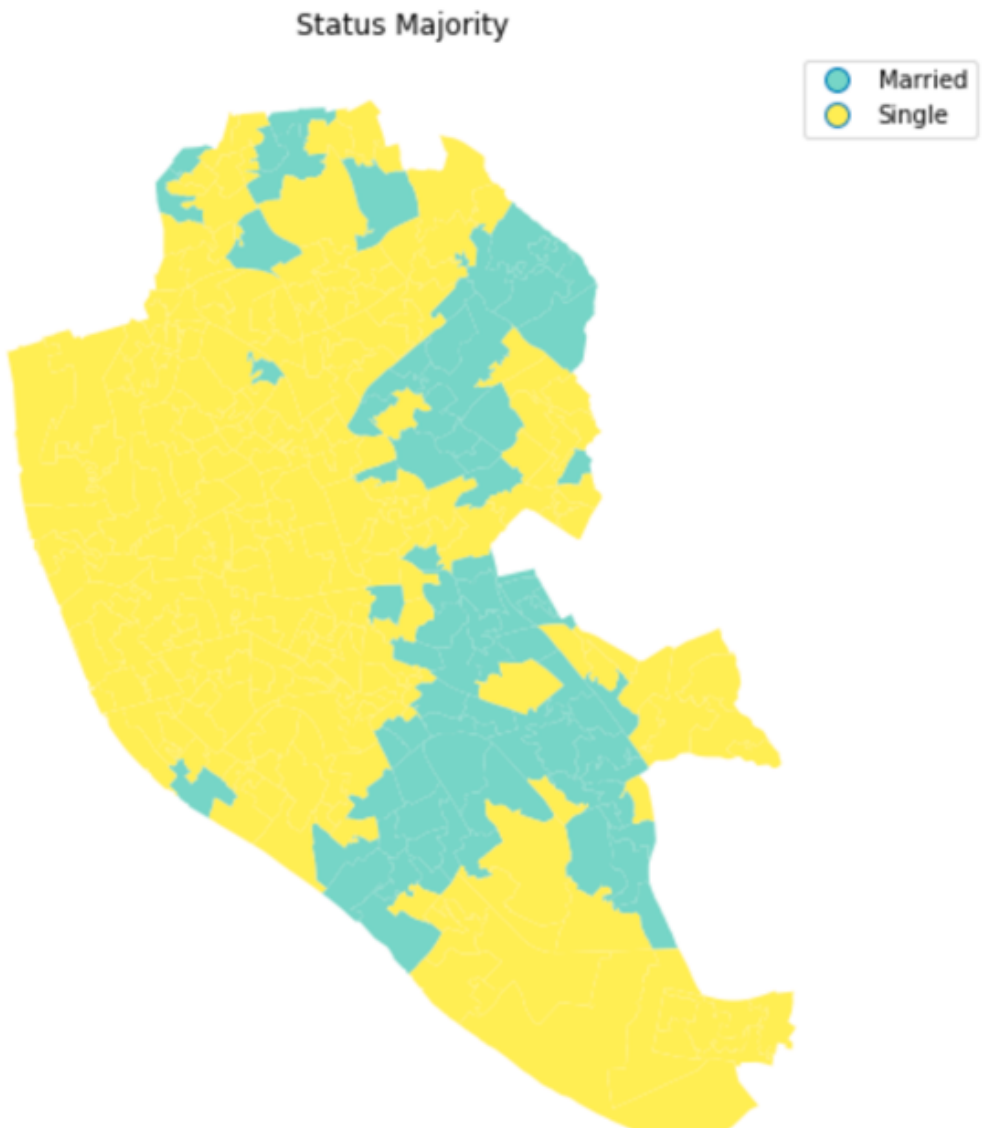
- Trade-off: detail vs cognitive load
- Exact number depends on purpose of the map
- Usually not more than 12

How do we bin?

Essentially a statistical problem

Unique values

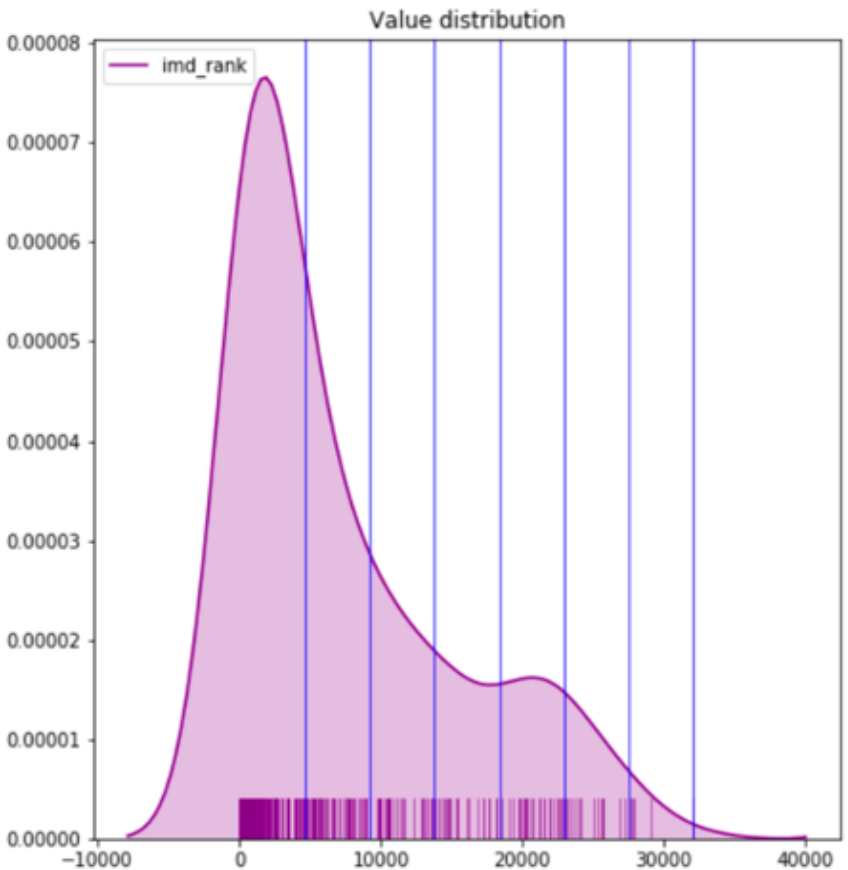
- Categorical data
- No gradient (reflect it with the colour scheme!!!)
- Examples: Religion, country of origin…

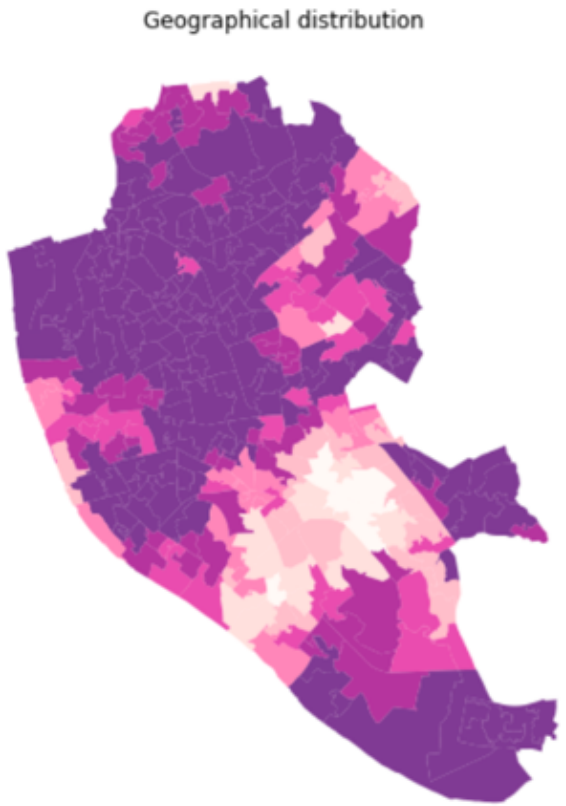

Status Majority
Married
Single

Equal interval (continuous)

- Take the value span of the data to represent and split it equally
- Splitting happens based on the numerical value
- Gives more weight to outliers if the distribution is skewed

equal_interval

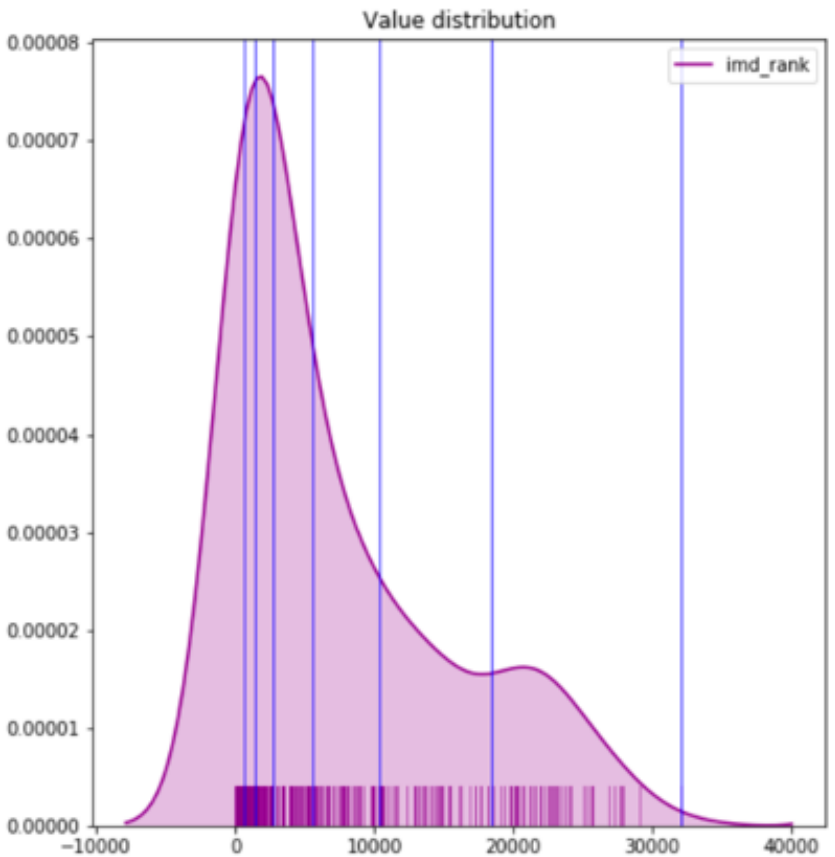


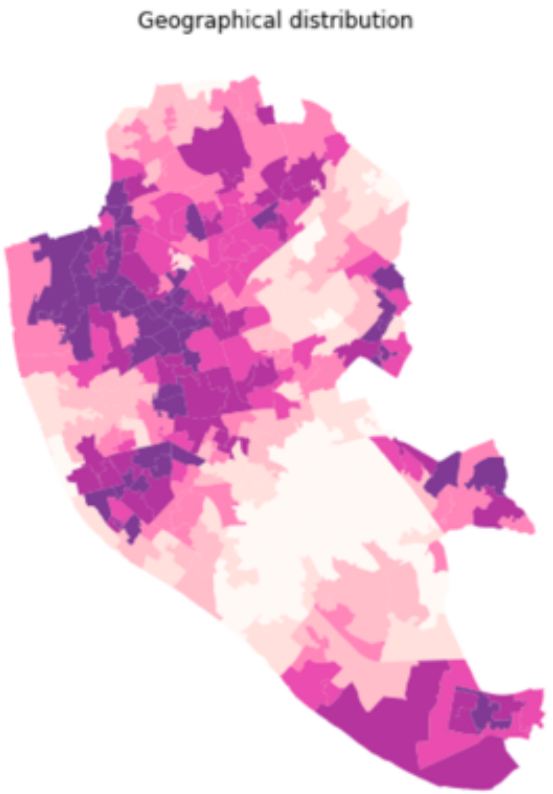


Quantile

- Regardless of numerical values, split the distribution keeping the same amount of values in each bin
- Splitting based on the rank of the value
- If distribution is skewed, it can put very different values in the same bin

quantiles


Value distribution
imd_rank
0.00008
0.00007
0.00006
0.00005
0.00004
0.00003
0.00002
0.00001
0.00000
−10000
0
10000
20000
30000
40000


Geographical distribution

Different type of algorithms will optimize for different types of splits

- Fisher-Jenks
- Natural breaks
- Outlier maps: box maps, std. maps…

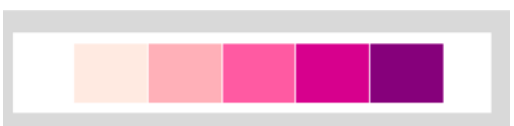
Some involve some fairly fancy statistics.

Colour palette

Categories, non-ordered

Graduated, sequential

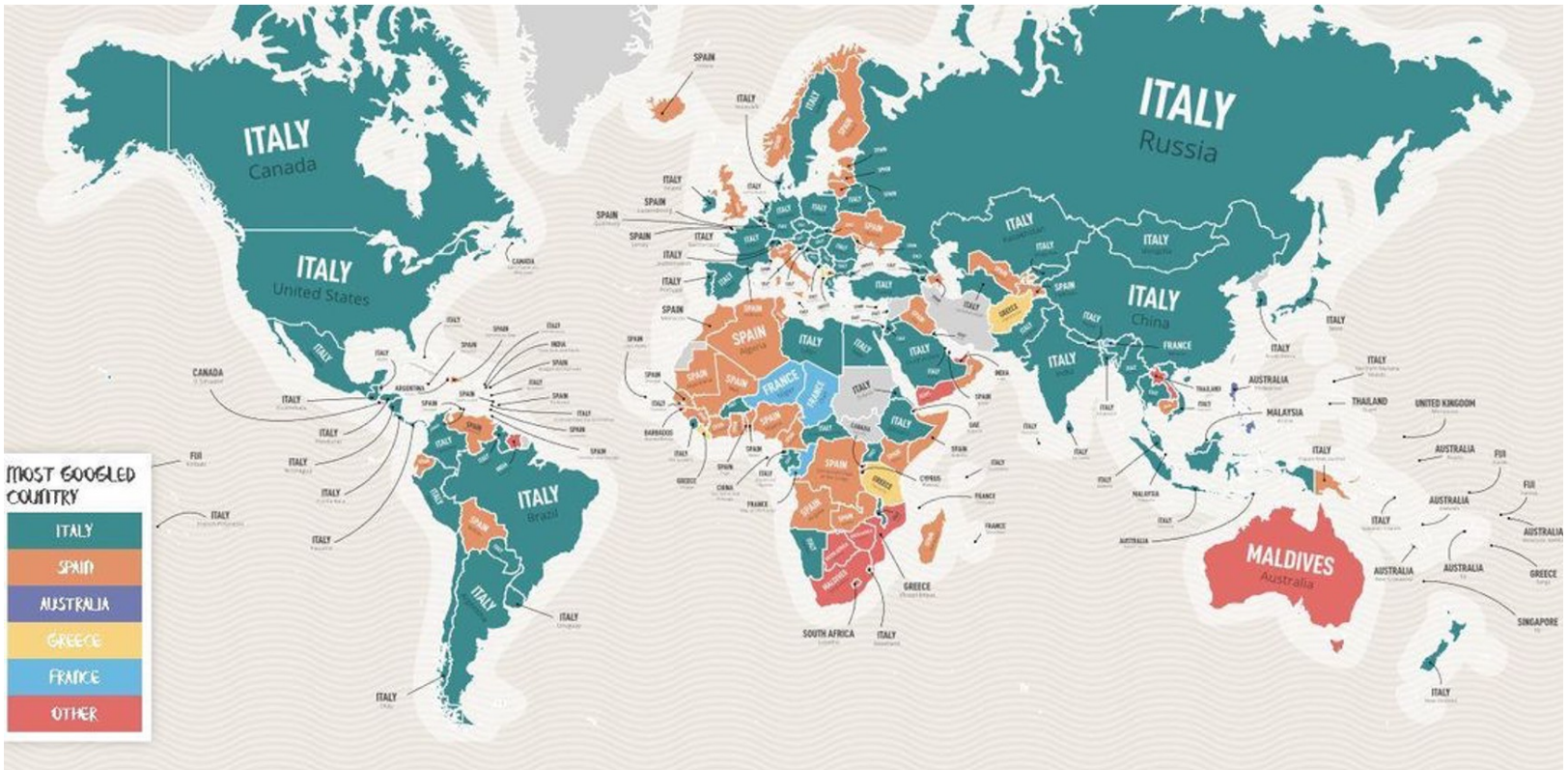
Graduated, divergent



TIP: check ColorBrewer for guidance

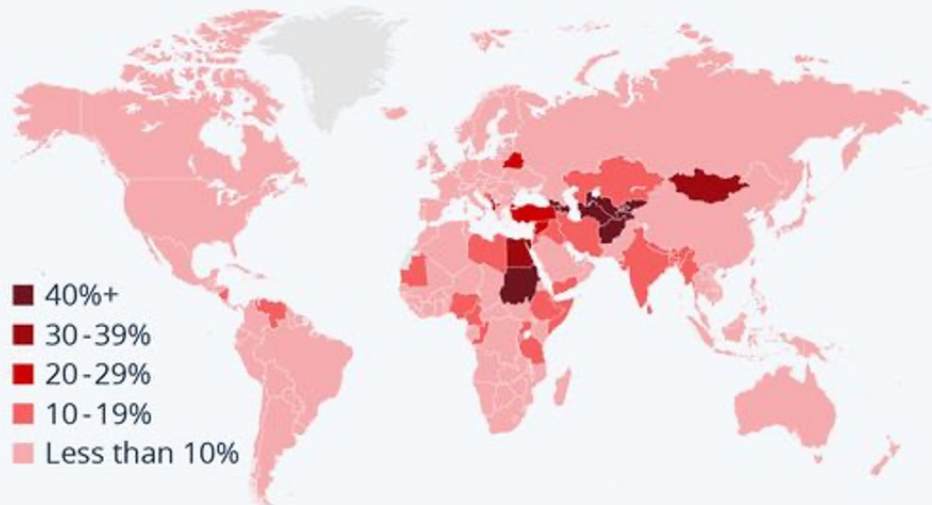
Tips

- Think of the purpose of the map
- Explore by trying different classification alternatives
- Combine (geo)visualisation with other statistical devices

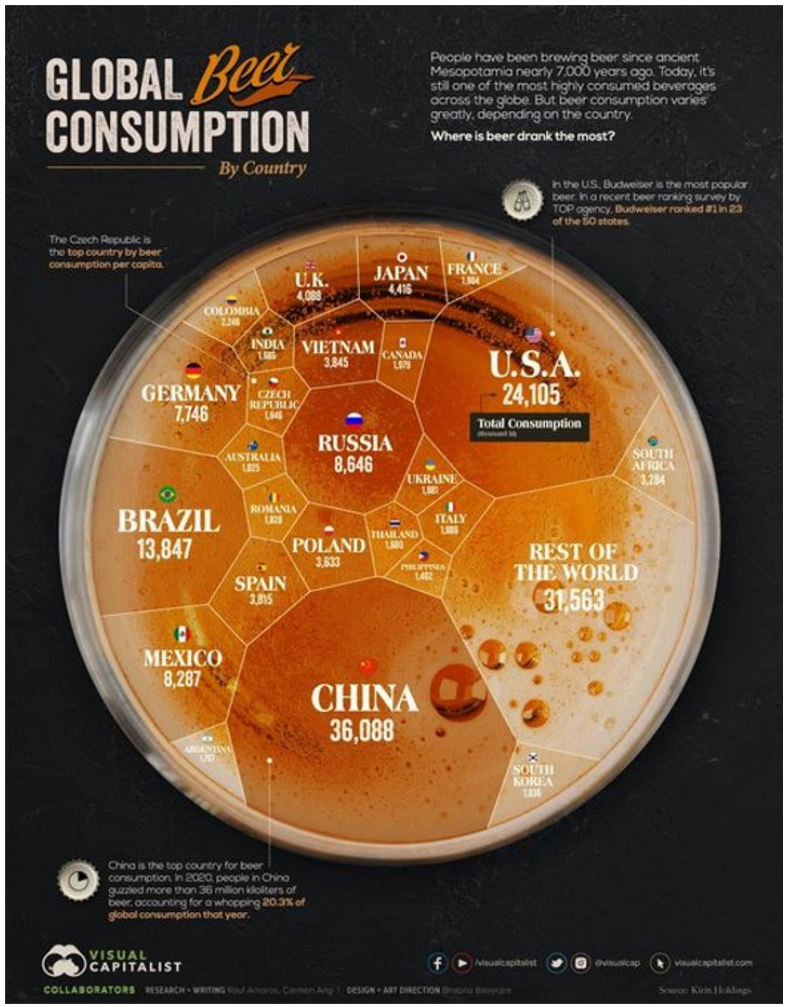

MOST GOOGLED COUNTRY
ITALY
SPAIN
AUSTRALIA
GREECE
FRANCE
OTHER
ITALY
Canada
ITALY
United States
ITALY
Russia
ITALY
China
ITALY
Brazil
MALDIVES
Australia
SPAIN
FRANCE
GREECE
SOUTH AFRICA
AUSTRALIA
SINGAPORE
UNITED KINGDOM
THAILAND
MALAYSIA
FIJI
CANADA

Where Food Imports Are Affected by the Ukraine Crisis

Share of food imports per country affected by export restrictions elsewhere (in percent of calories)



As of August 1, 2022. Excludes exports held back from Ukraine.
Source: International Food Policy Research Institute


GLOBAL Beer CONSUMPTION
By Country
People have been brewing beer since ancient Mesopotamia nearly 7,000 years ago. Today, it's still one of the most highly consumed beverages across the globe. But beer consumption varies greatly, depending on the country.
Where is beer drank the most?
In the U.S., Budweiser is the most popular beer. In a recent beer ranking survey by TOP agency, Budweiser ranked #1 in 23 of the 50 states.
The Czech Republic is the top country by beer consumption per capita.
U.K. 4,088
JAPAN 4,416
FRANCE 1,984
COLOMBIA 2,248
INDIA 1,665
VIETNAM 3,845
CANADA 1,979
U.S.A. 24,105
Total Consumption
GERMANY 7,746
CZECH REPUBLIC 1,946
RUSSIA 8,646
AUSTRALIA 1,825
UKRAINE 1,881
SOUTH AFRICA 3,284
ROMANIA 1,828
ITALY 1,888
BRAZIL 13,847
THAILAND 1,680
POLAND 3,833
PHILIPPINES 1,482
REST OF THE WORLD 31,563
SPAIN 3,815
MEXICO 8,287
CHINA 36,088
ARGENTINA 1,787
SOUTH KOREA 1,838
China is the top country for beer consumption. In 2020, people in China guzzled more than 36 million kiloliters of beer, accounting for a whopping 20.3% of global consumption that year.
VISUAL CAPITALIST
/visualcapitalist
@visualcap
visualcapitalist.com
COLLABORATORS
RESEARCH + WRITING Raul Amoros, Carmen Ang
DESIGN + ART DIRECTION Brabna Banerjee
Source: Kirin Holdings

Questions