Geographic Data Science

Choropleths

Dani Arribas-Bel

Choropleths

Choropleths

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

- Counterpart of the histogram
- **Values** are **classified** into specific **colors**: value –> bin
- **Information loss** as a trade off for **simplicity**

Classification choices

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

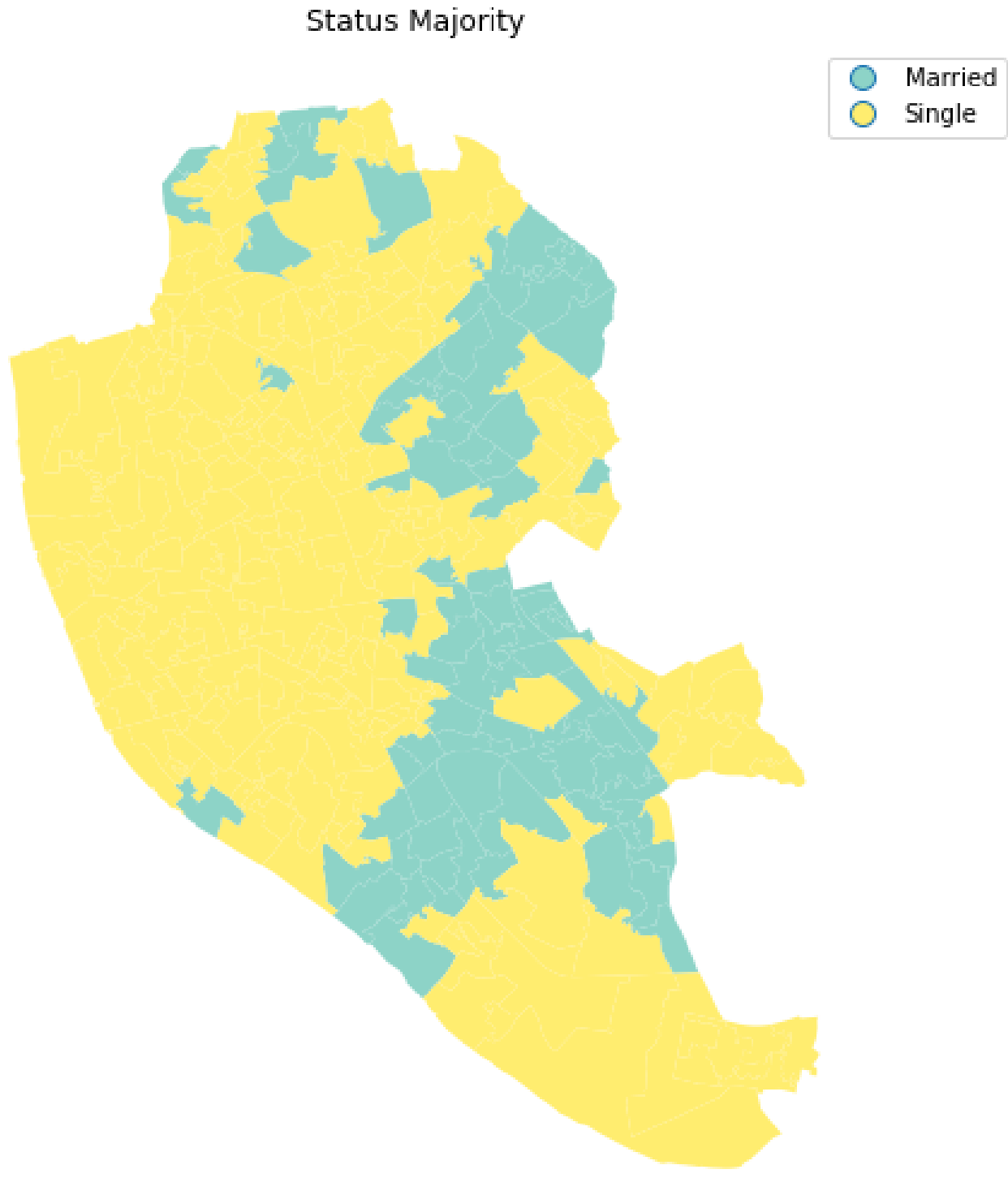
How many bins?

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

How to bin?

Unique values

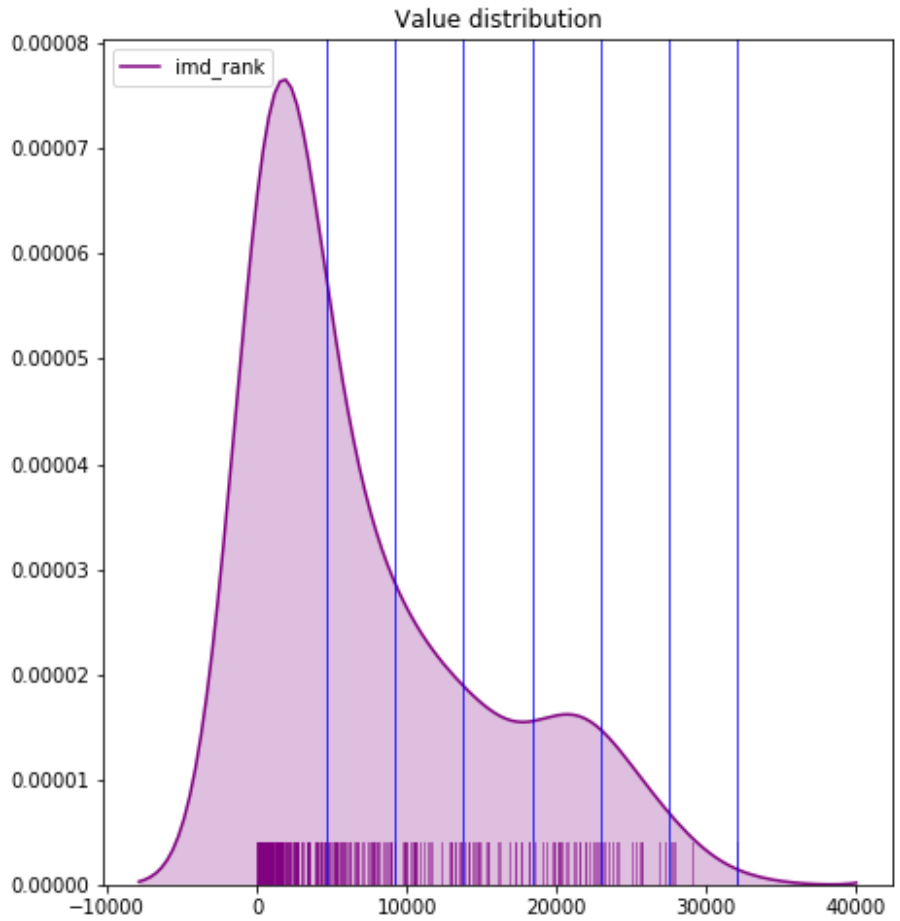
- Categorical data
- No gradient (reflect it with the color scheme!!!)
- Examples: Religion, country of origin…

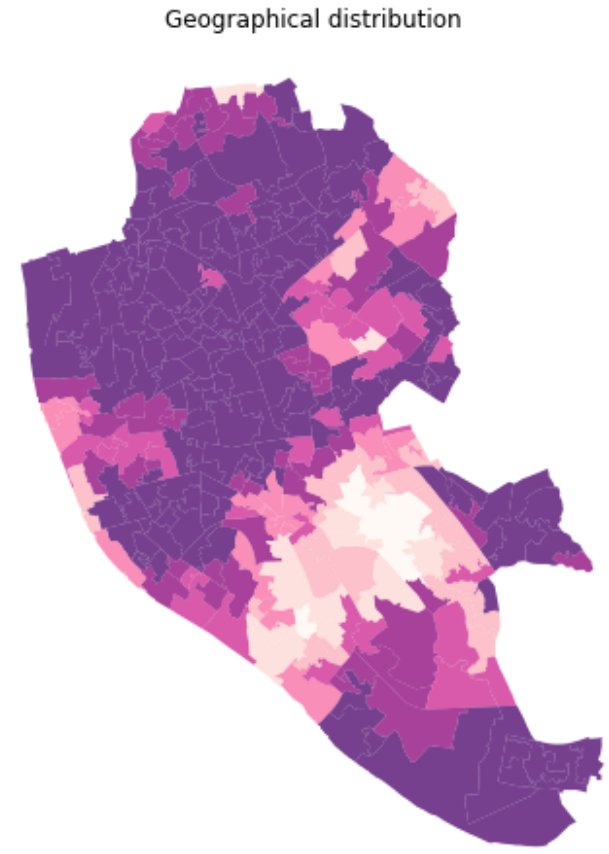

Status Majority
Married
Single

Equal interval

- 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

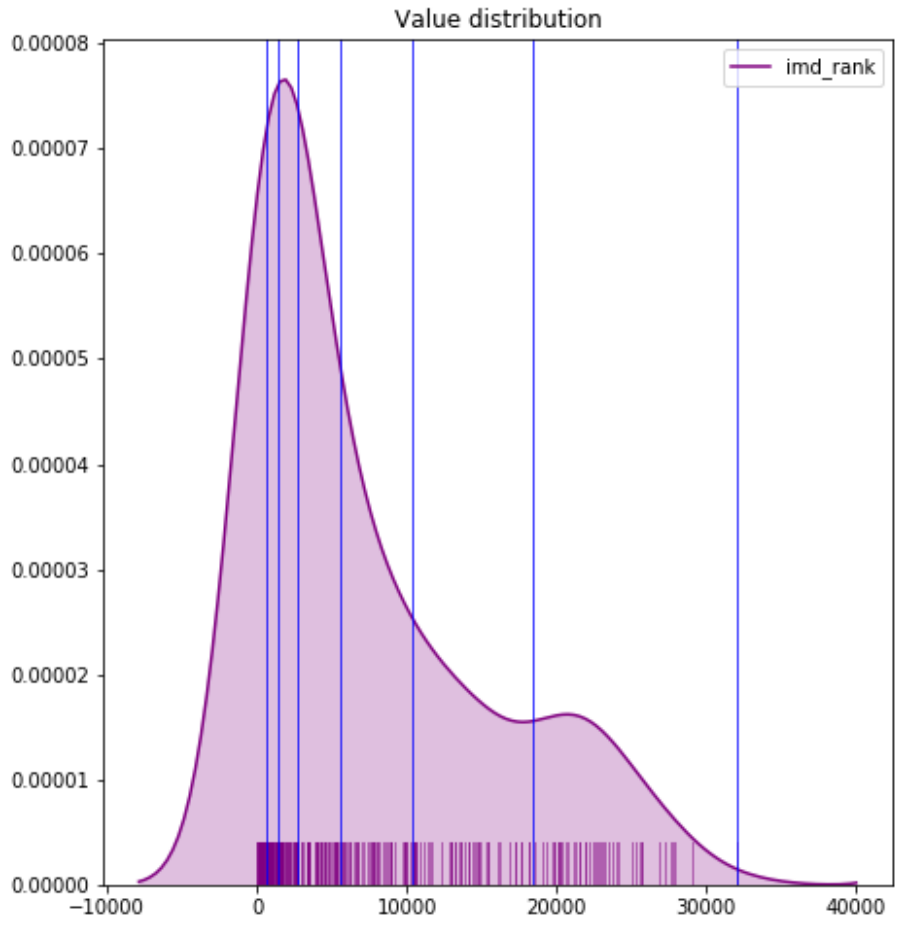


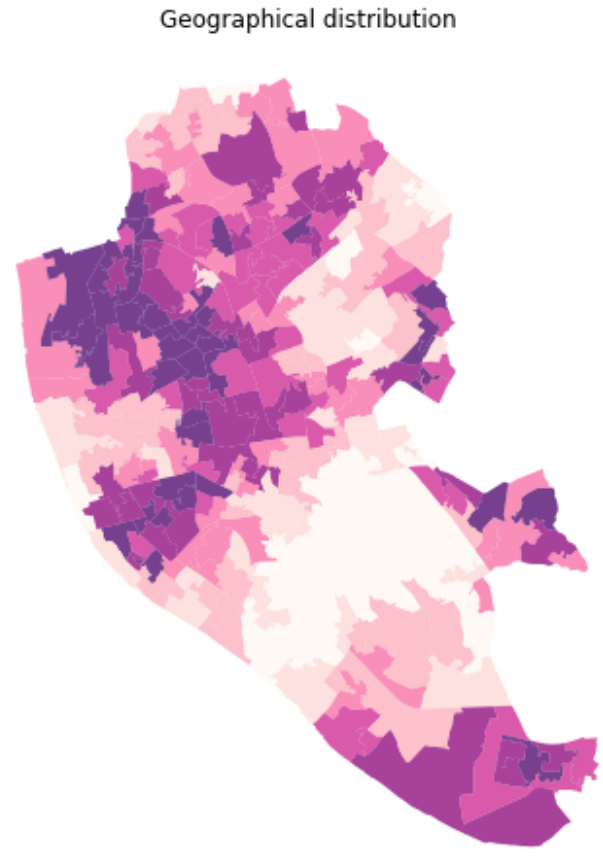


Quantiles

- 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

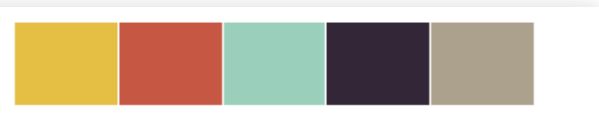
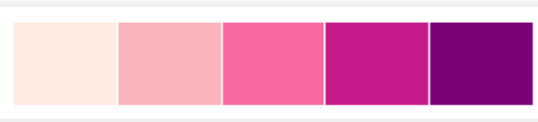
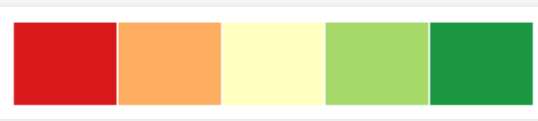

Geographical distribution

Other

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

Color schemes

Align with your purpose

- **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