Geographic Data Science - Lecture IV

Mapping Data

Dani Arribas-Bel

Today

- Mapping data
- MAUP
- Choropleths
  - Definition
  - Classes
- Cartograms
- Conditional maps
- Space-Time mapping

Data **maps**

- Abstraction from the purely geographical map
- Representing numerical values within a spatial context

Data **maps**

- Abstraction from the purely geographical map
- Representing numerical values within a spatial context

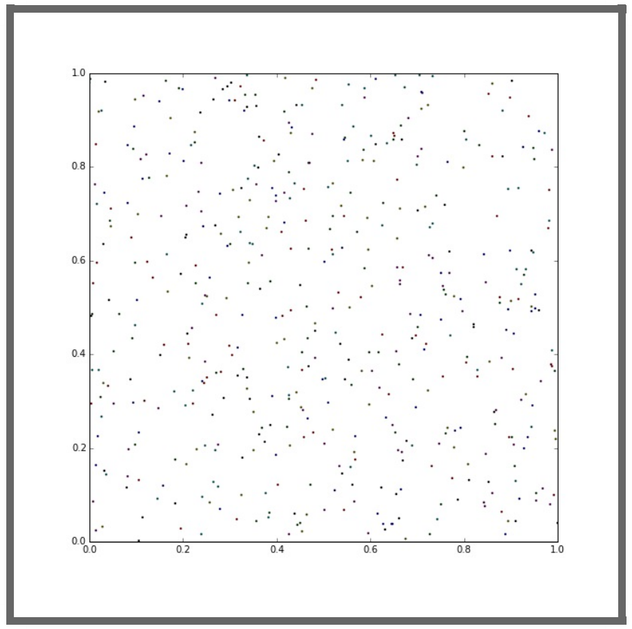
Mapping **data**

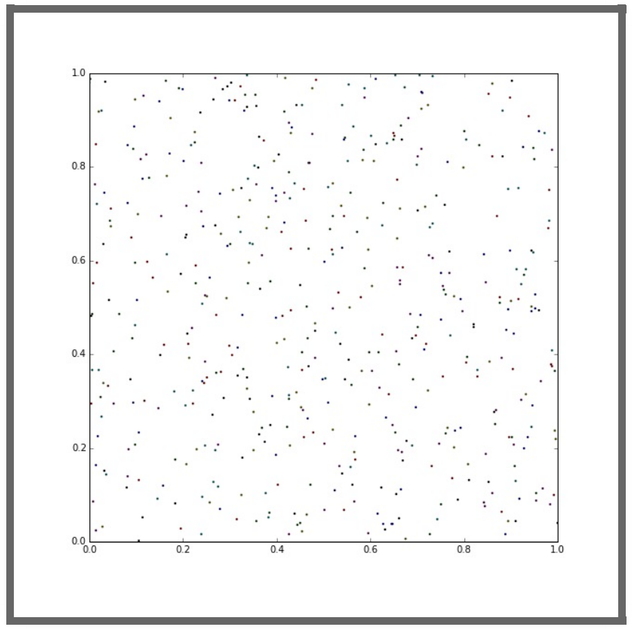
- A geographical approach to statistical visualization
- The spread of data is considered in its geographical dimension

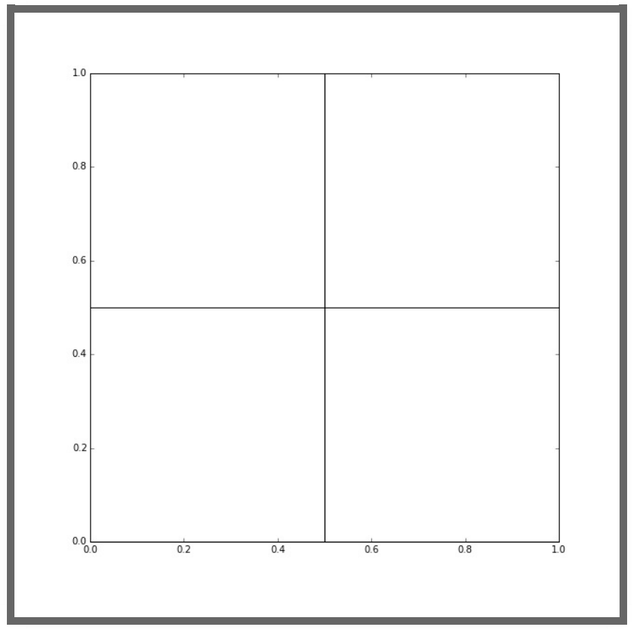
Before we delve into different types of data maps...

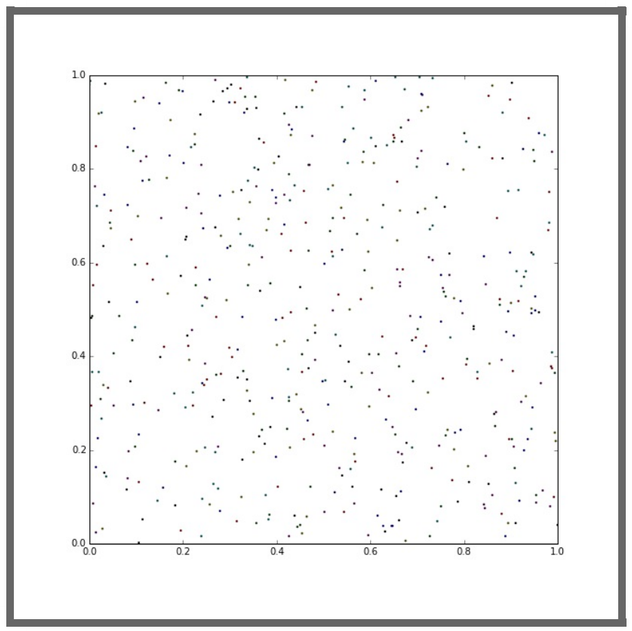
MAUP

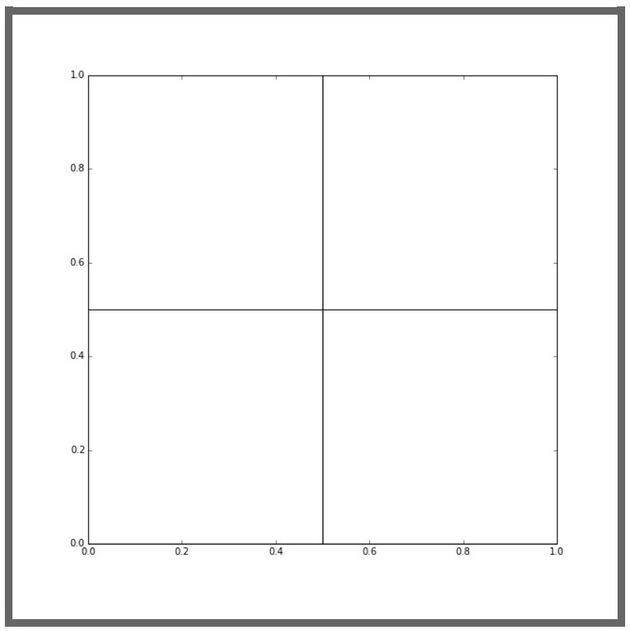
Modifiable **A**real **U**nit **P**roblem

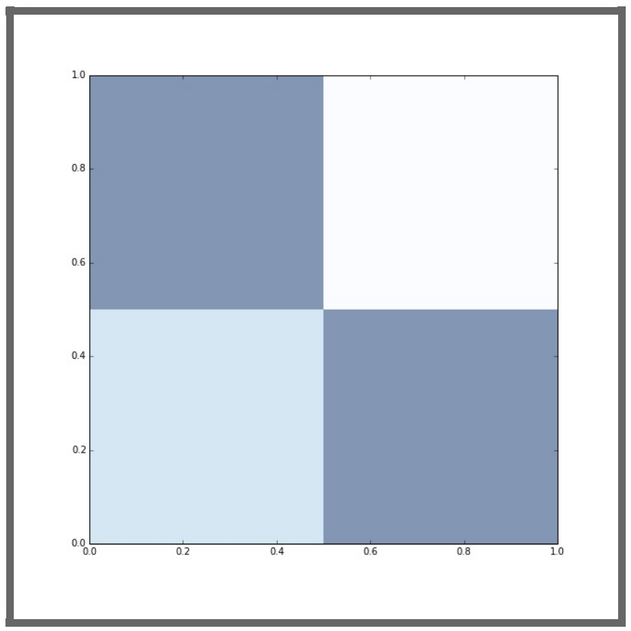

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

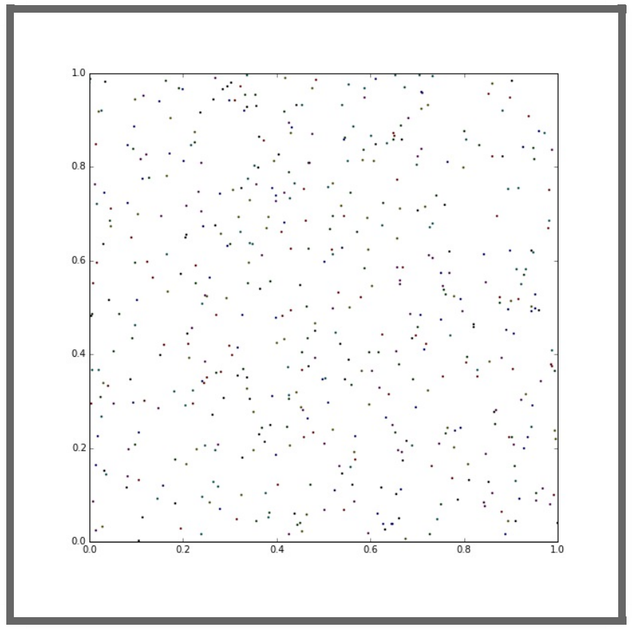

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

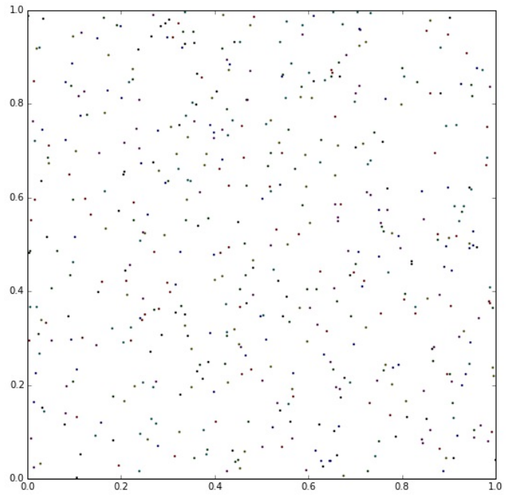

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

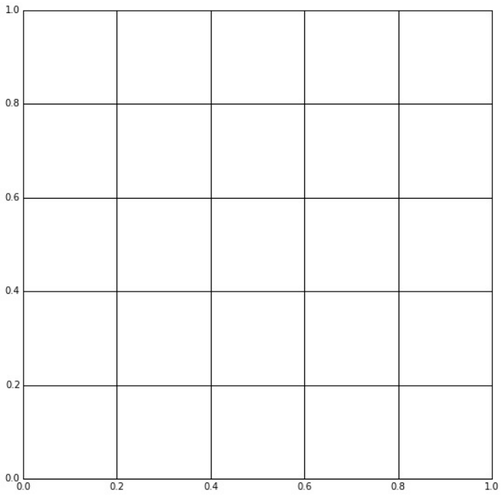

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

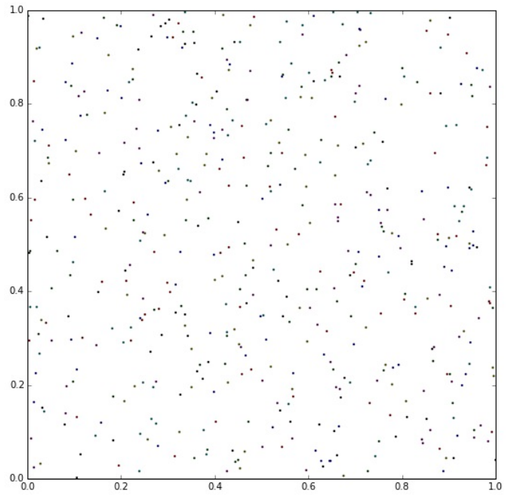

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

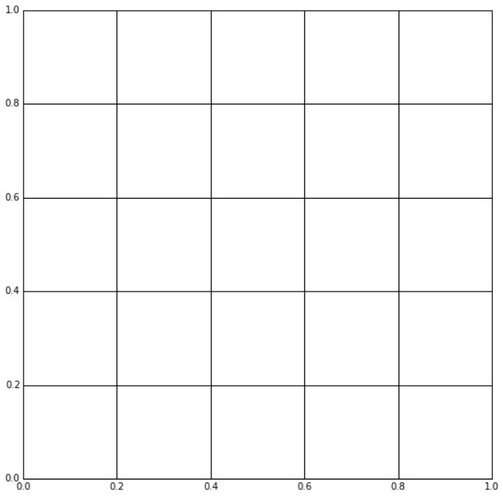

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

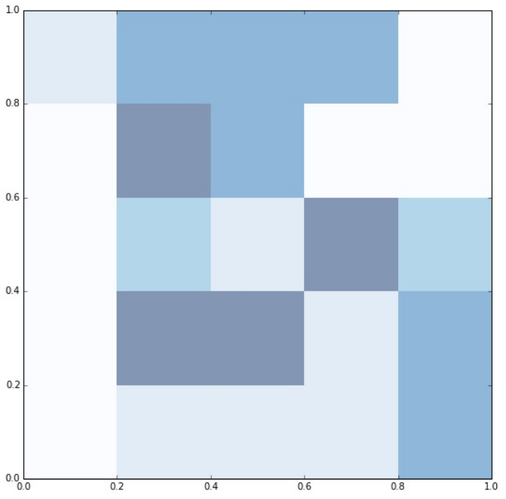

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

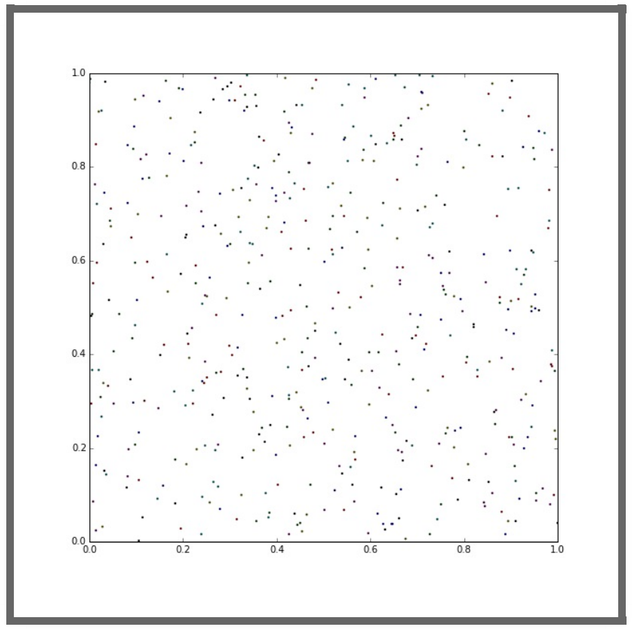

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

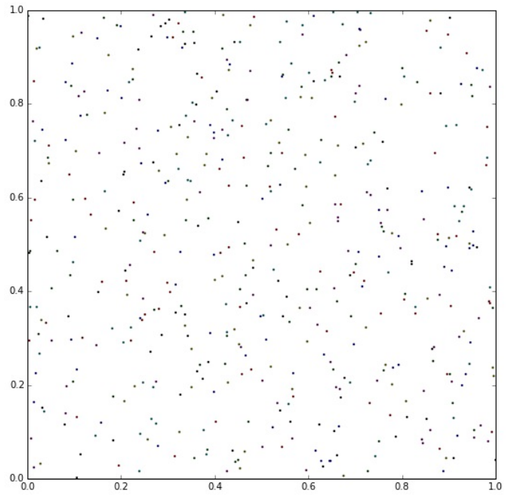

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

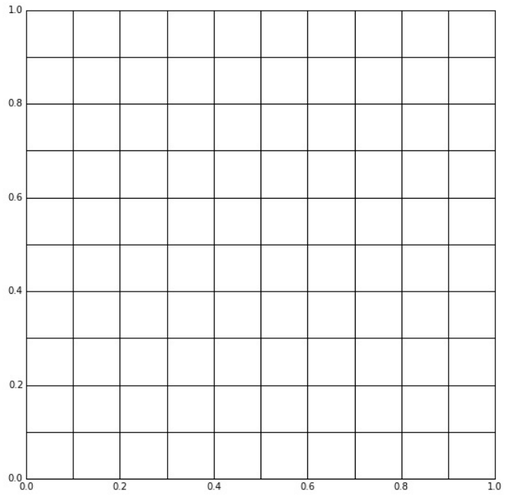

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

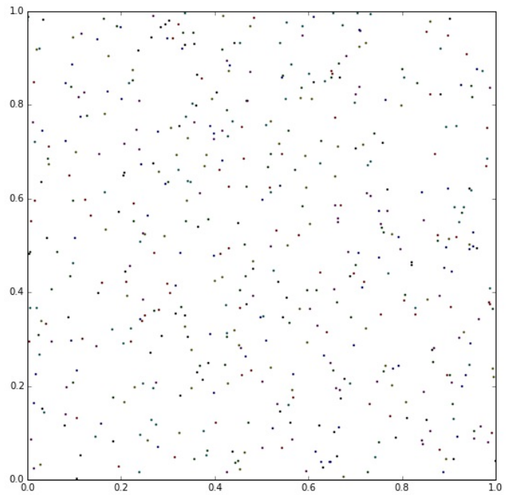

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

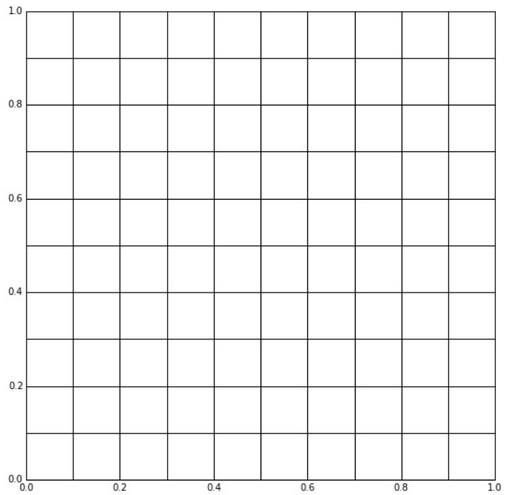

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

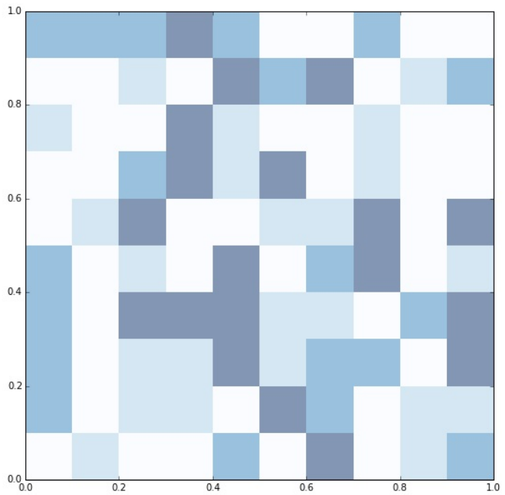

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

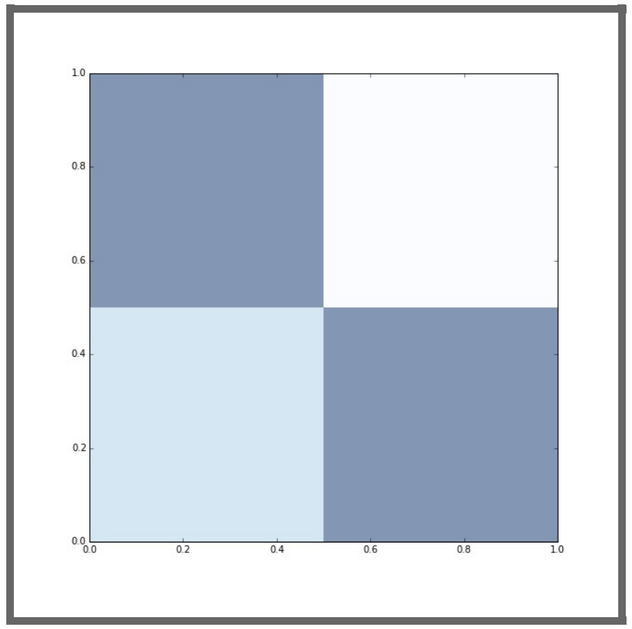

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

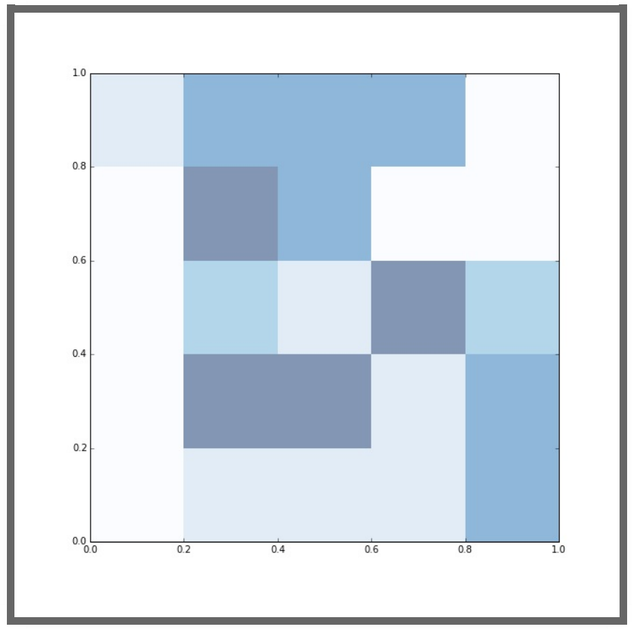

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

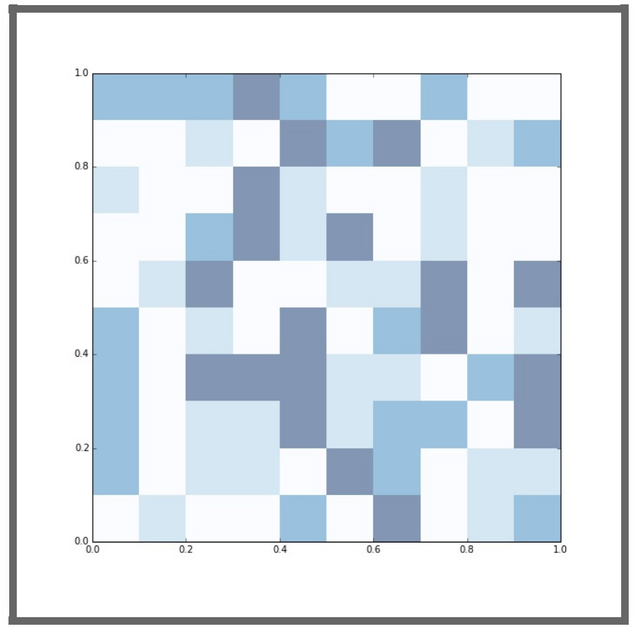

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


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

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

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

Classification choices

- Colors
- Bins
- Algorithm:

Classification choices

- Colors --> in alignment with the goal of the map
- Bins
- Algorithm:

Classification choices

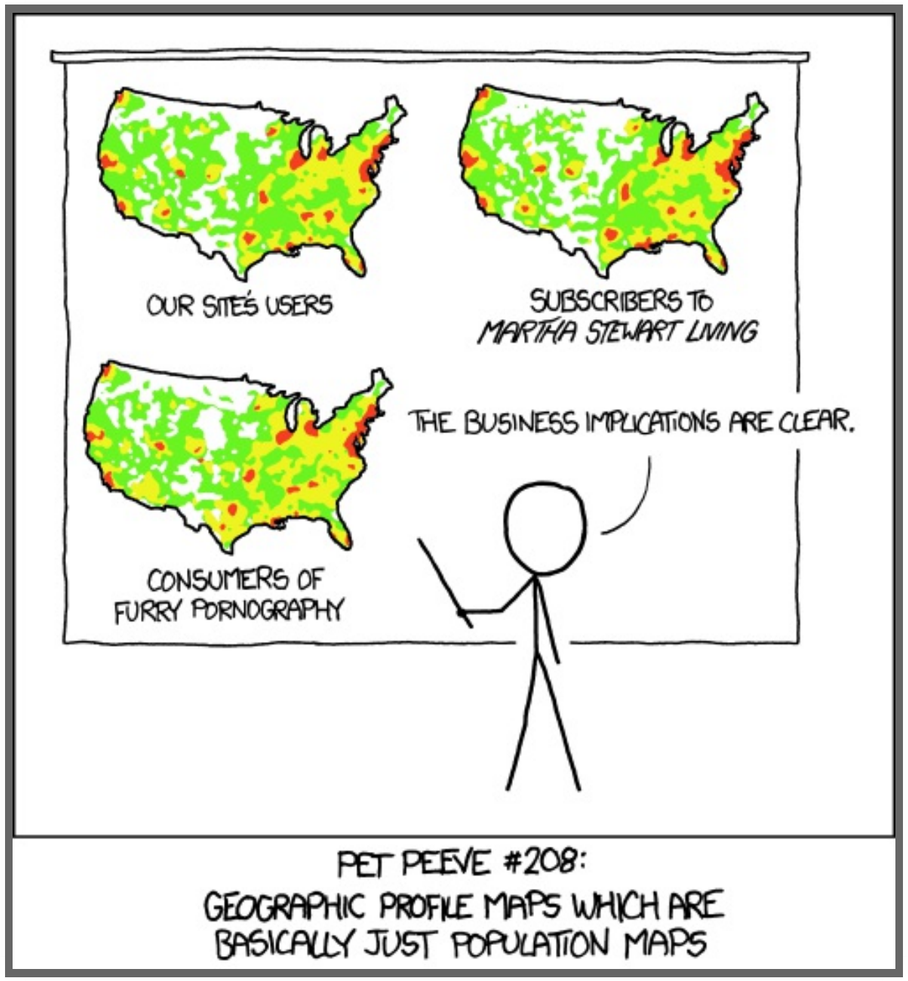
- Colors --> in alignment with the goal of the map
- Bins --> How many?
- Algorithm:

Classification choices

- Colors --> in alignment with the goal of the map
- Bins --> How many?
- Algorithm:
  - Unique values
  - Equal interval
  - Qua/Quintiles (equal count)
  - Fisher-Jenks
  - ...

Beware standarization!!!

[Source]



Color schemes

ColorBrewer: Color A ×

colorbrewer2.org

Number of data classes: 3

Nature of your data:
sequential | diverging | qualitative

Pick a color scheme:
Multi-hue:
Single hue:

Only show:
colorblind safe
print friendly
photocopy safe

Context:
roads
cities
borders

Background:
solid color
terrain

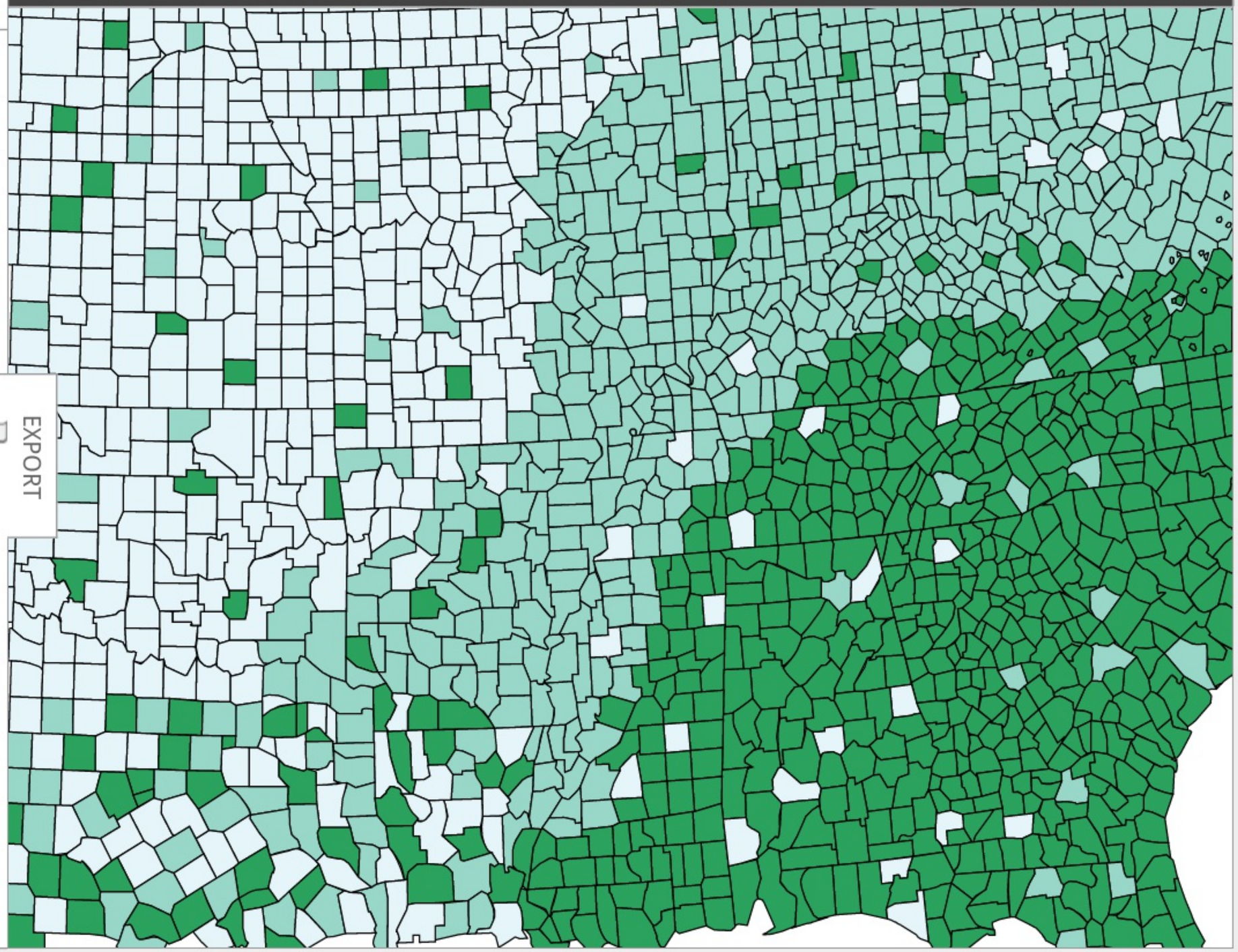
color transparency

3-class BuGn
HEX
#e5f5f9
#99d8c9
#2ca25f
EXPORT

[Source]

how to use | updates | downloads | credits

COLORBREWER 2.0
color advice for cartography



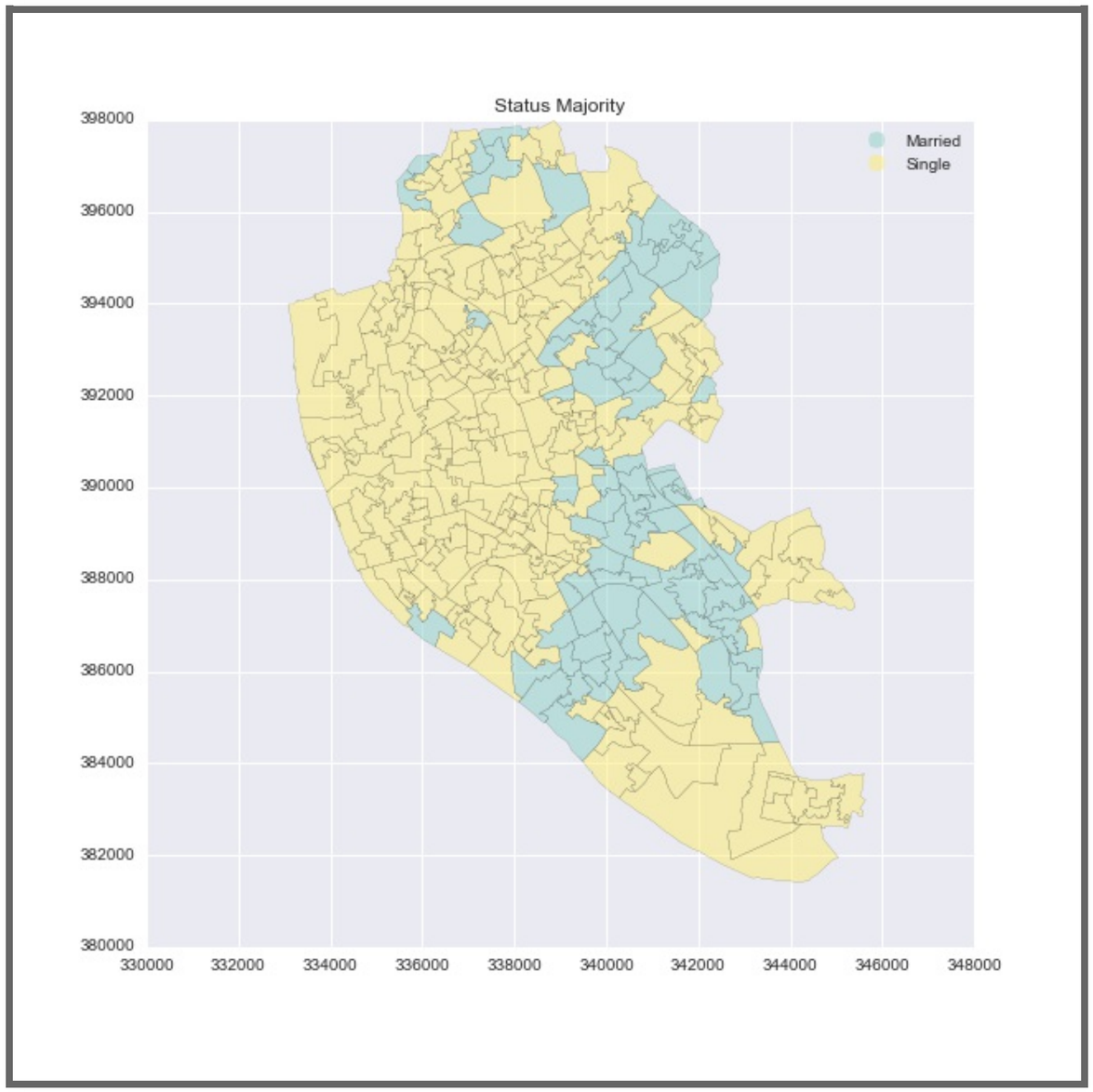
Support
Back to Flash version
Back to ColorBrewer 1.0

axismaps

Unique values

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

Unique values

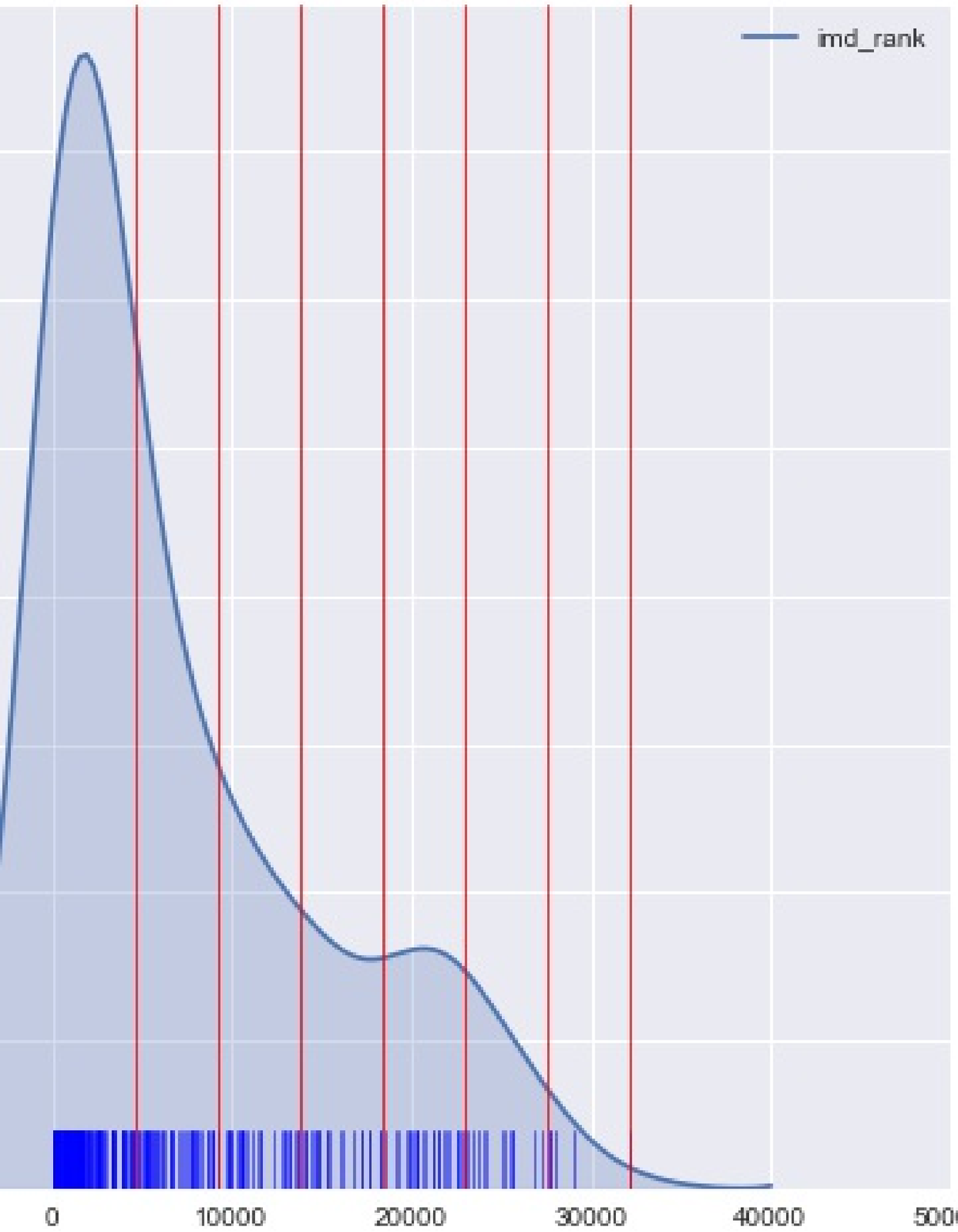


Equal interval

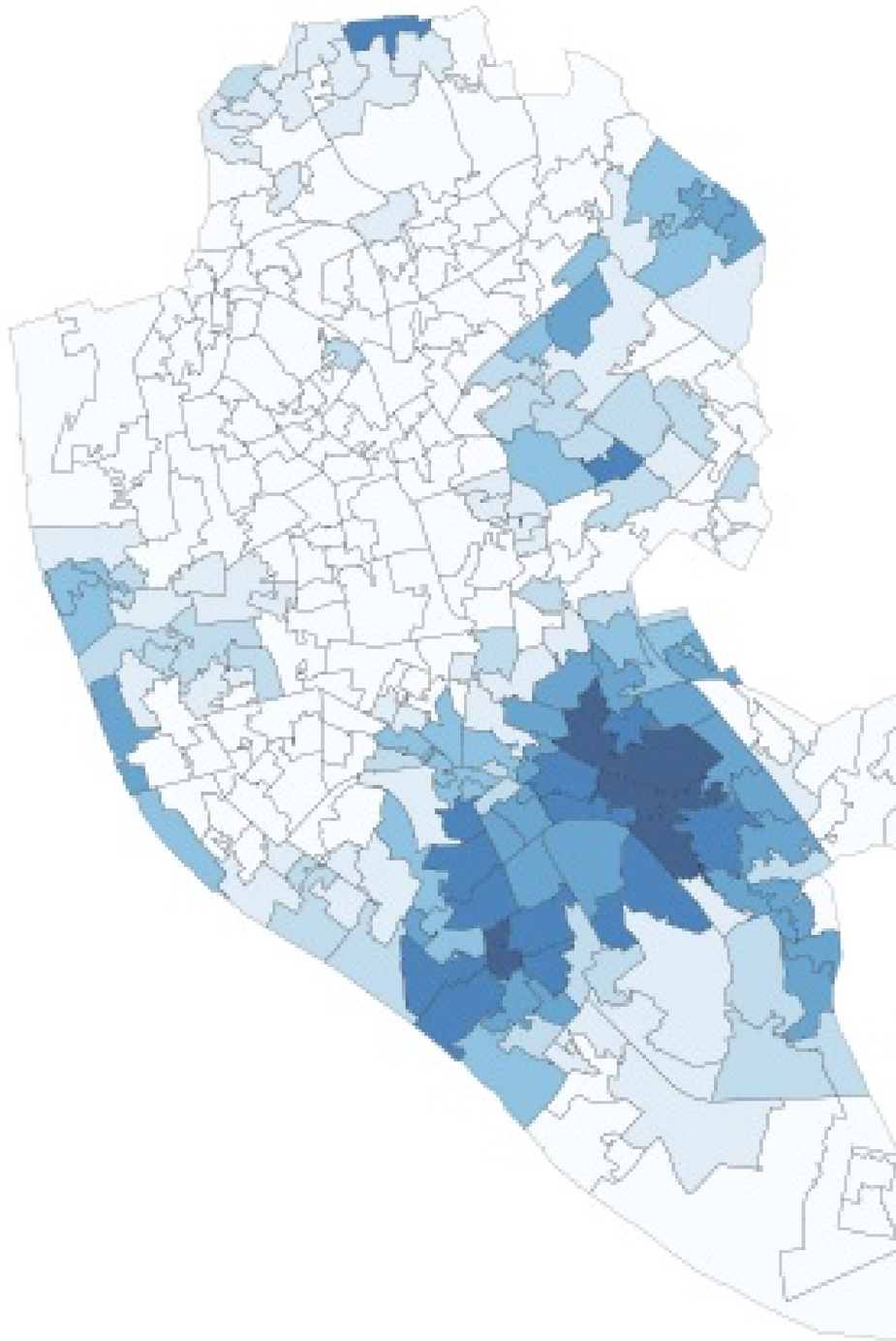
- Take the **value** span of the data to represent and split it equally
- **Splitting** happens based on the **numerical value**
- Gives more weight to outliers if the distribution is skewed

equal_interval

Value distribution



Geographical distribution



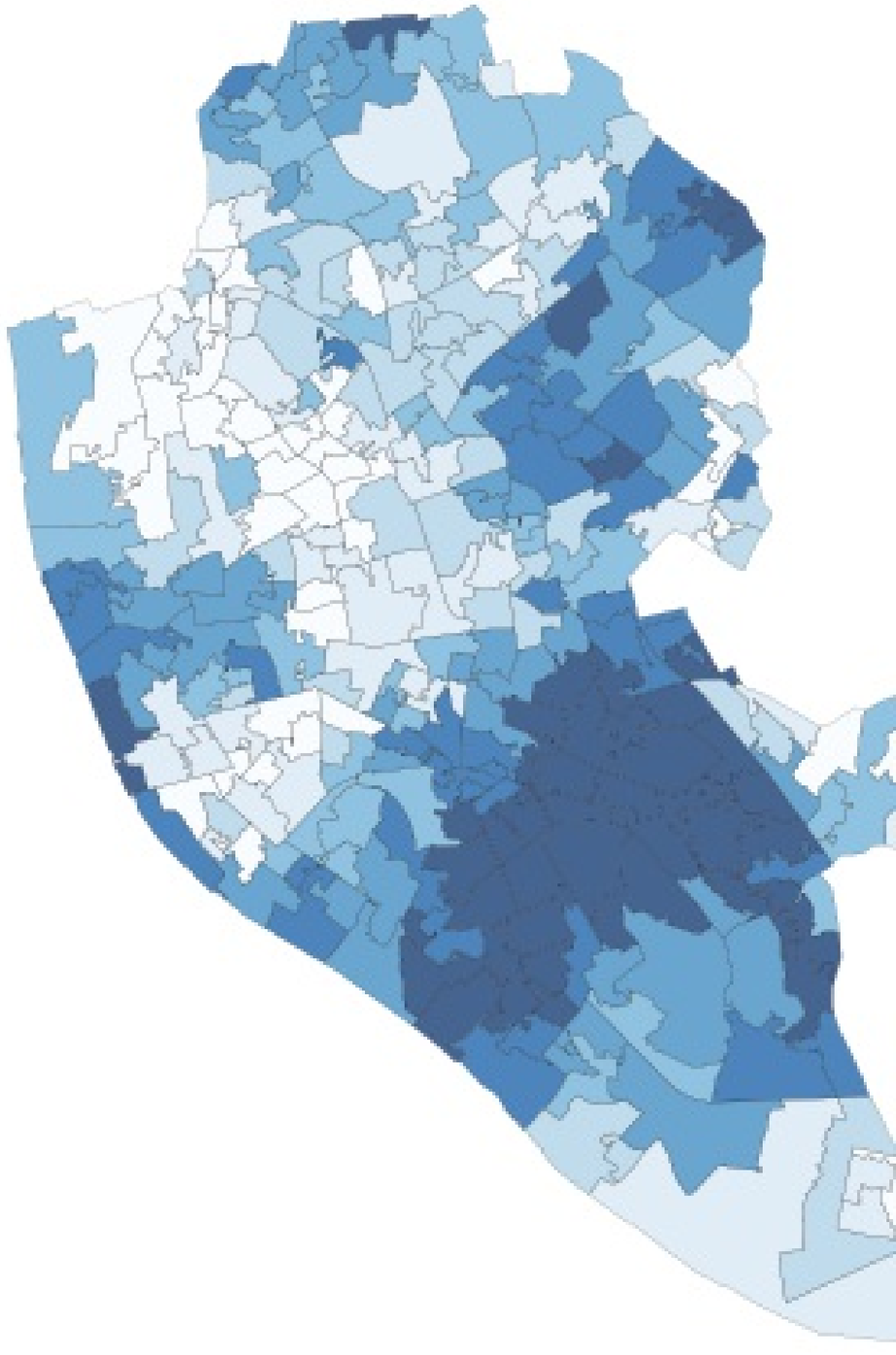
Quantiles

- 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

Geographical distribution



Other

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

Tips

Different classification schemes can produce widely different maps as a result of:

- The distribution of the values
- The inherent simplification that a choropleth implies

Best advice is to **explore** different ones and **combine** choropleths with other graphical devices like histograms or density plots

Cartograms

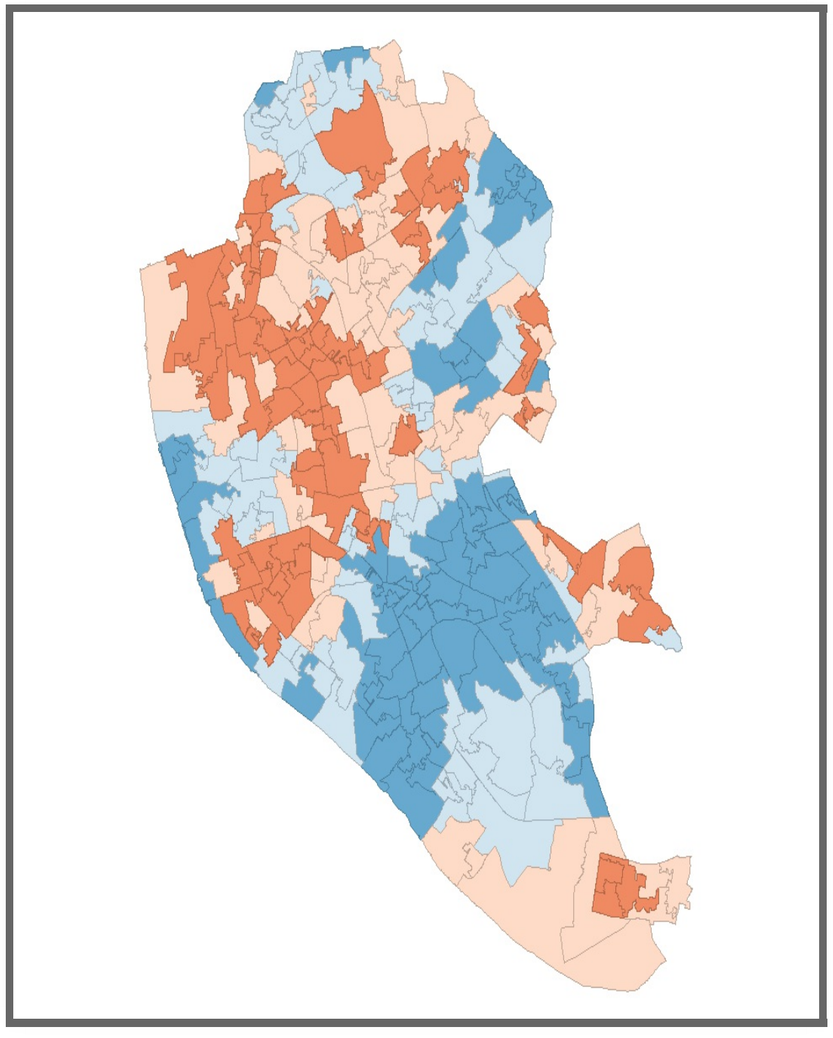
*"**Data maps** where the variable is encoded, not by a color gradient, but by **distorting the shape/size** of the geographical objects"*

Cartograms

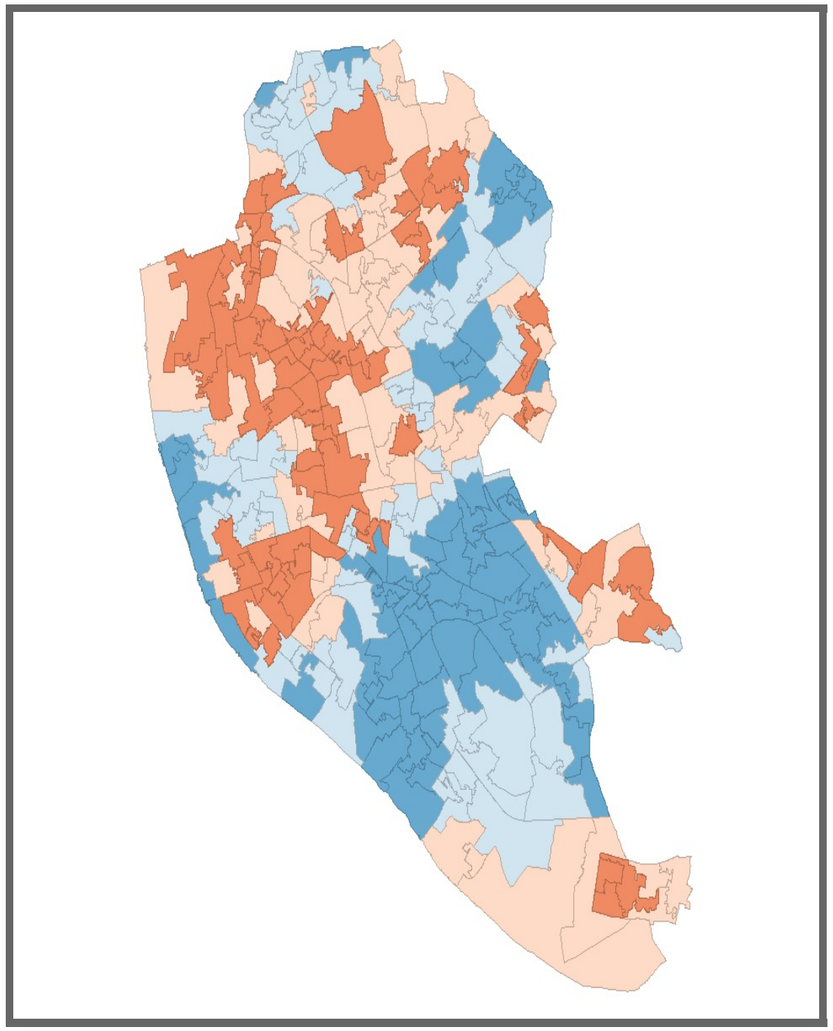
*"**Data maps** where the variable is encoded, not by a color gradient, but by **distorting the shape/size** of the geographical objects"*

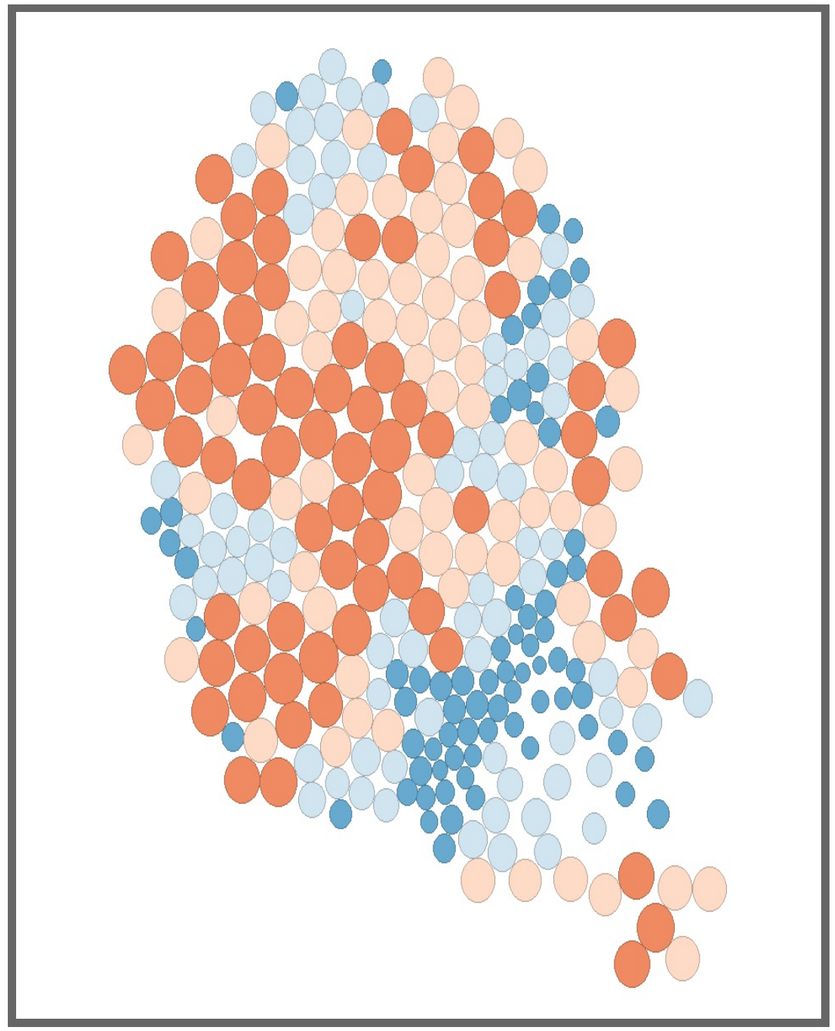
- Useful in cases where the natural size/shape induces to wrong interpretation, or obscures the intended representation.
- If not done carefully, it can distort the message in unintended ways

Cartograms



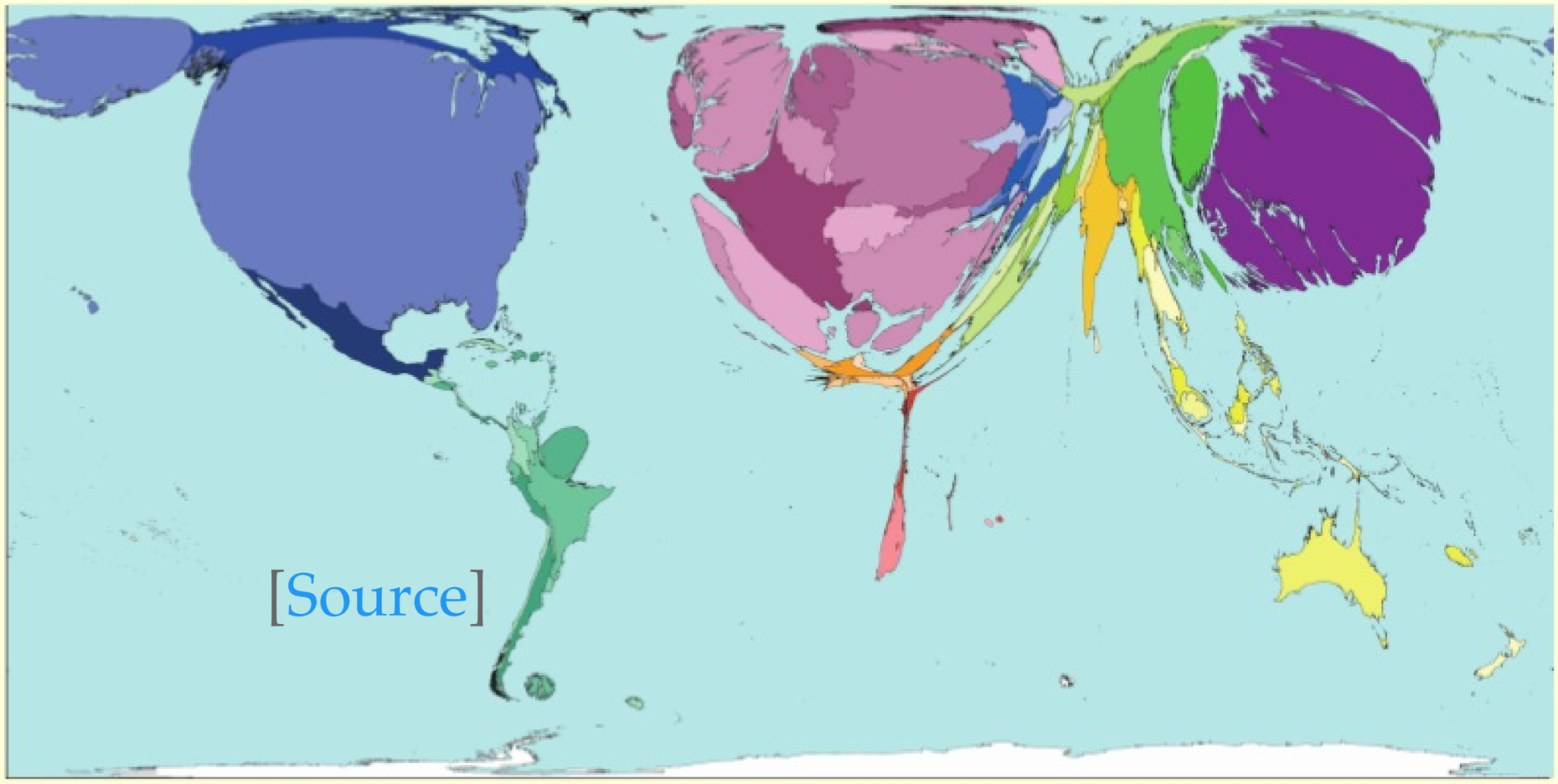
Cartograms





Capital Consumption

The University Of Sheffield. The Leverhulme Trust Geographical Association
Produced by the SASI group (Sheffield) and Mark Newman (Michigan)



Shown here is the cost of maintaining fixed assets and replacing them when necessary. Fixed assets include buildings, vehicles, communication and transport infrastructures. Deterioration occurs due to use, time, and becoming outdated. Fixed capital consumption is the cost of preventing this deterioration.

Large, technologically advanced infrastructures are likely to cost more to maintain than those that are small and simple. The United States, Japan, Germany and France have the highest absolute costs of fixed capital consumption. Costs per person are high in Norway, Switzerland and Denmark - roughly one thousand times higher than in Nepal, Ethiopia and Burundi.

Territory size shows the proportion of all fixed capital consumption that occurred there.

Land area

Technical notes
- Data are from the World Bank's 2005 World Development Indicators.
- See website for further information.

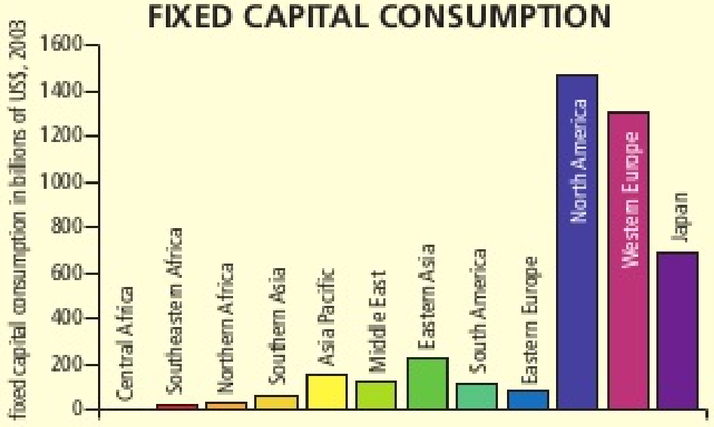
HIGHEST AND LOWEST FIXED CAPITAL CONSUMPTION

Rank	Territory	Value
1	Norway	7.0
2	Switzerland	6.2
3	Japan	5.4
4	Denmark	5.1
5	United States	4.5
6	Finland	4.4
7	Sweden	4.0
8	Netherlands	4.0
9	Austria	3.9
10	Germany	3.8

Rank	Territory	Value
191	Eritrea	0.0126
192	Sierra Leone	0.0112
193	Malawi	0.0101
194	Guinea-Bissau	0.0099
195	Liberia	0.0091
196	Haiti	0.0072
197	Democratic Republic Congo	0.0070
198	Nepal	0.0058
199	Ethiopia	0.0055
200	Burundi	0.0054

Fixed Capital Consumption in thousands of US$ per person per year, 2003

FIXED CAPITAL CONSUMPTION



"... it will be more expensive in the long run if we let existing infrastructure become so run down, that it must eventually be replaced at great costs."

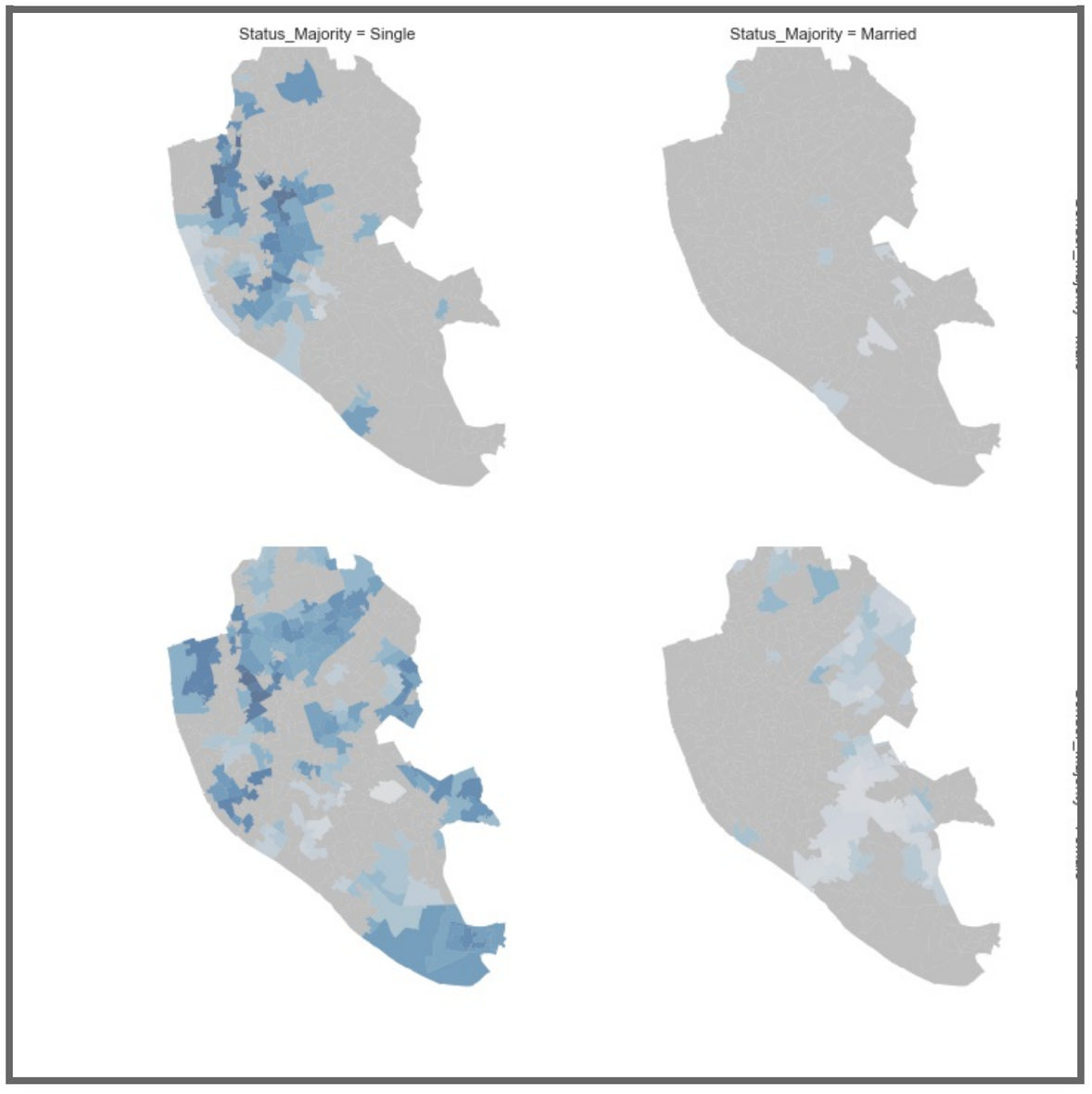
Education and Training Unit (South Africa), 2007

Conditional maps

Split a dataset in *buckets* by *conditioning* on additional variables, then create a map for each *bucket*

- If no association, maps should look the same
- But, if the conditioning variables are somewhat related to the outcome we are mapping, the spatial distribution can vary substantially
- **Exploration** of **multivariate** relationships

Conditional maps



Space-Time mapping

Space-Time mapping

- Bringing time into a spatial 2D context is "tricky" (it's really 3D!)
- Traditionally
- More recently

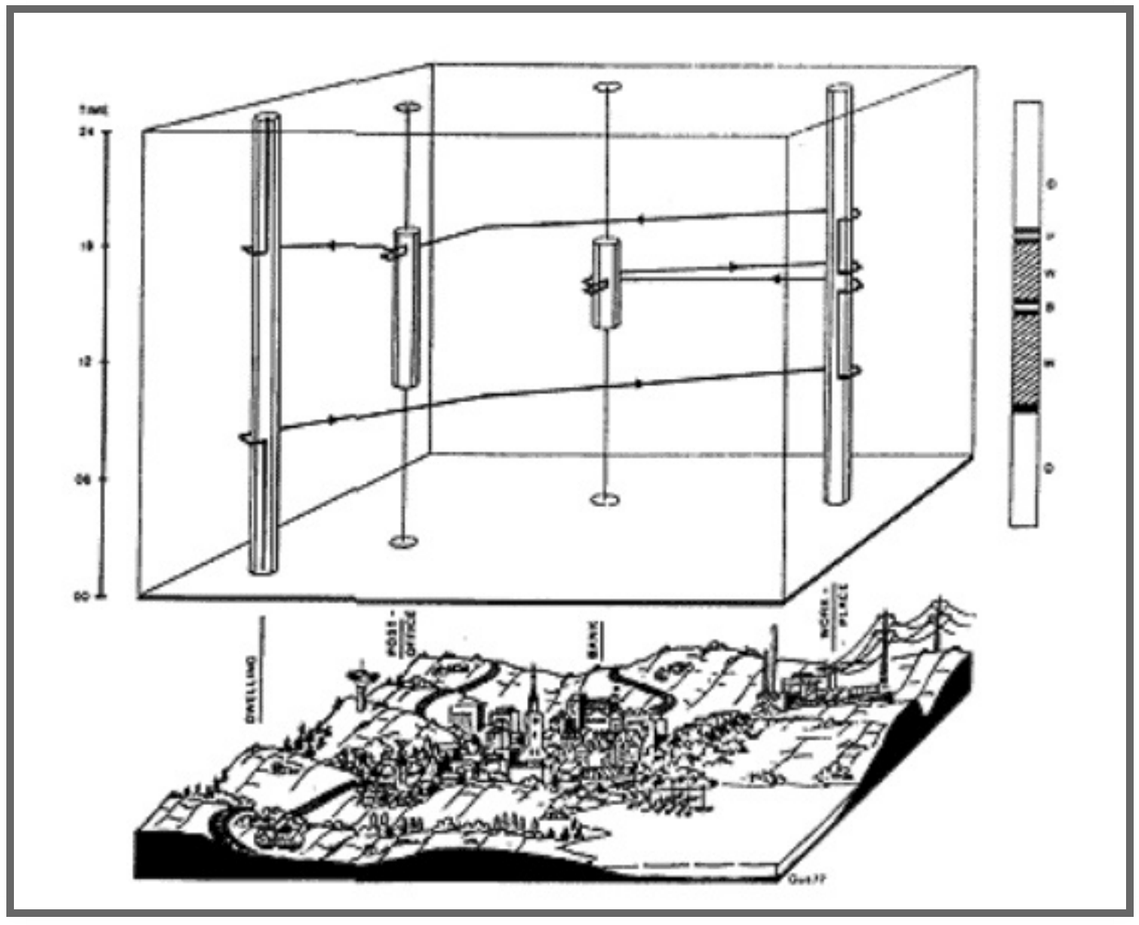
Space-Time mapping

- Bringing time into a spatial 2D context is "tricky" (it's really 3D!)
- Traditionally --> sequence of time periods, 3D plots
- More recently

Space-Time mapping

- Bringing time into a spatial 2D context is "tricky" (it's really 3D!)
- Traditionally --> sequence of time periods, 3D plots
- More recently --> animation and interactivity

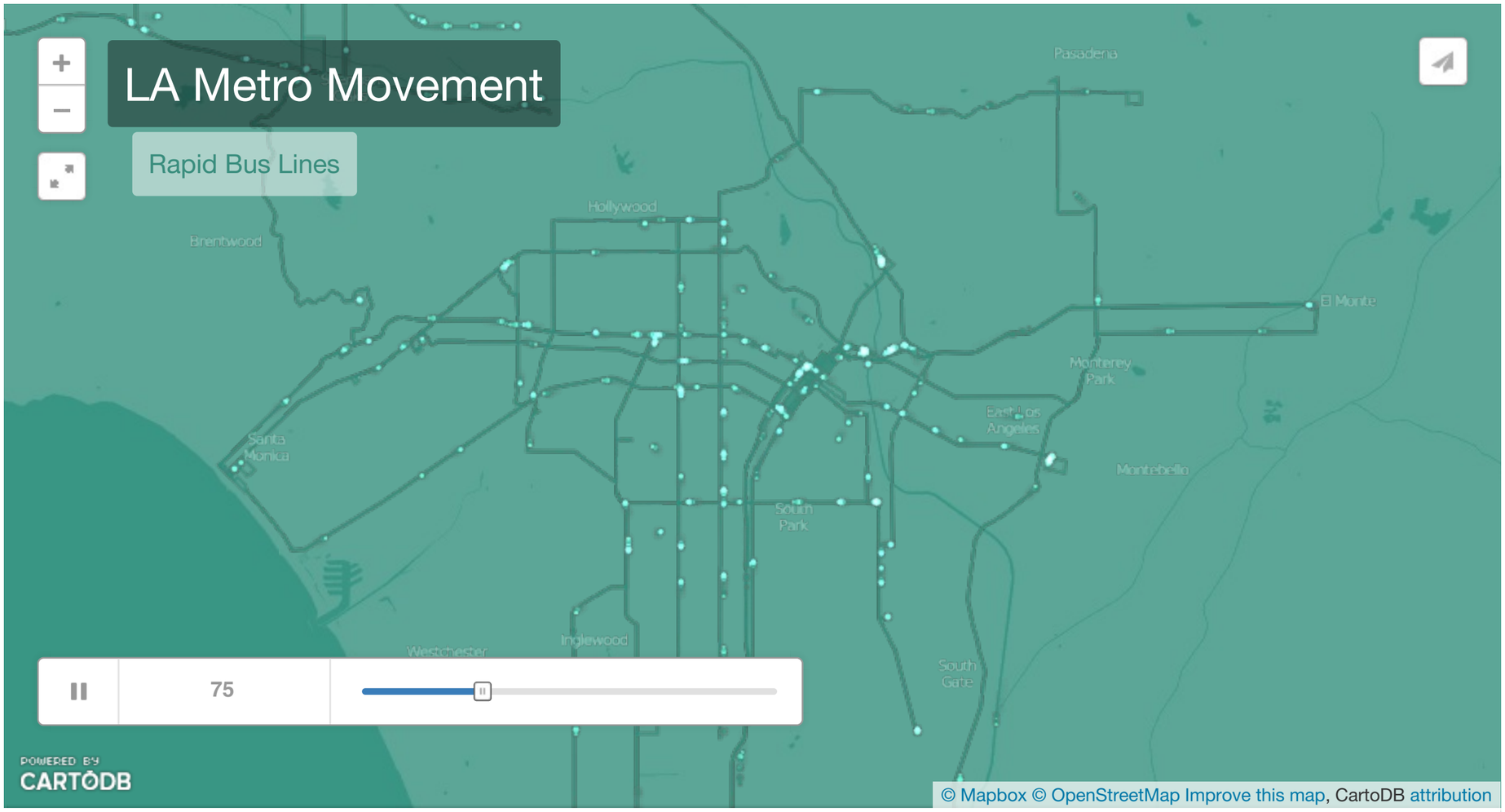
[Source]



[Source]


+
-
Iceland
Norway
Finland
Sweden
Estonia
Latvia
Lithuania
Belarus
United Kingdom
Poland
Germany
Czech Rep.
Ukraine
Kazakhstan
Austria
France
Croatia
Romania
Italy
Bulgaria
Georgia
Uzbekistan
Kyrgyzstan
Azerbaijan
Spain
Turkey
Turkmenistan
Tajikistan
Syria
Iraq
Iran
Afghanistan
Pakistan
Morocco
Algeria
Libya
Egypt
W. Sahara
United Arab Emirates
Oman
Mauritania
Mali
Niger
Chad
Sudan
Eritrea
Yemen
Senegal
Burkina Faso
Guinea
Nigeria
Côte d'Ivoire
Ethiopia
S. Sudan
Cameroon
Somalia
Venezuela
Guyana
Colombia
Kenya
Congo
Leaflet | CartoDB

[Source]



Map created by d9a

4