Data and Information Visualization:

Map Visualization

Semester 2 2018/2019

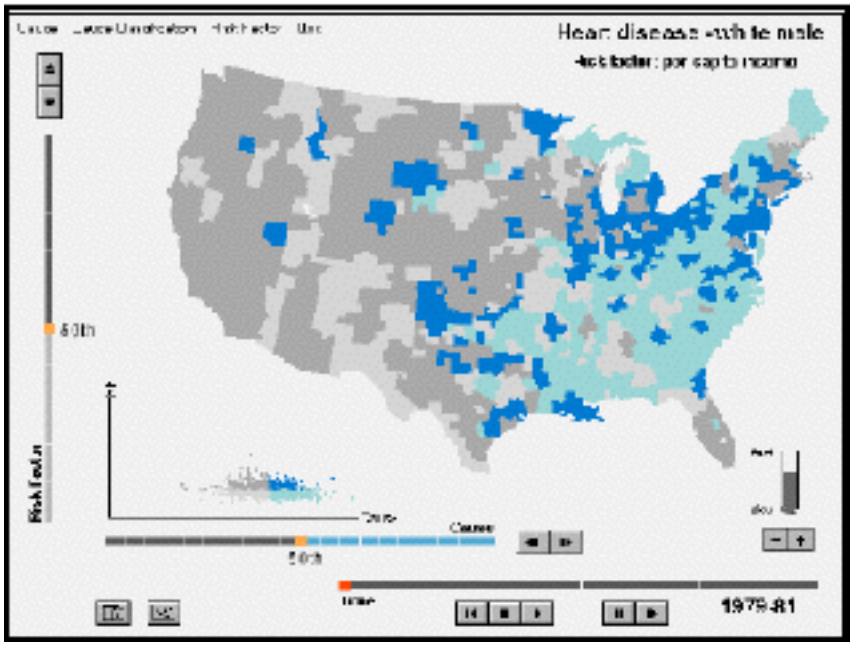
References

- Few Stephen, **Introduction to Geographical Data Visualization**, Perceptual Edge *Visual Business Intelligence Newsletter*, March/April 2009.
- Brewer, Cynthia A., **Basic Mapping Principles for Visualizing Cancer Data Using Geographic Information Systems (GIS)**
- Peterson, Gretchen N., **GIS Cartography: A Guide to Effective Map Design**, CRC Press, 2009.

Contents

- Methods for Encoding Quantitative Data on a Map
- Classing
- Mapping Multiple Variables

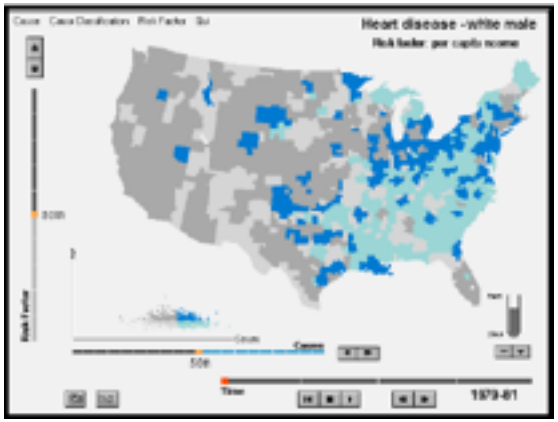
Which One is Cartographic Viz or Map Viz?



Different maps, but same domain

- Applying cartographic principles to visualization of non-geographic information

  (Skupin, 2000)



- Visualization methods for interacting with geographic information

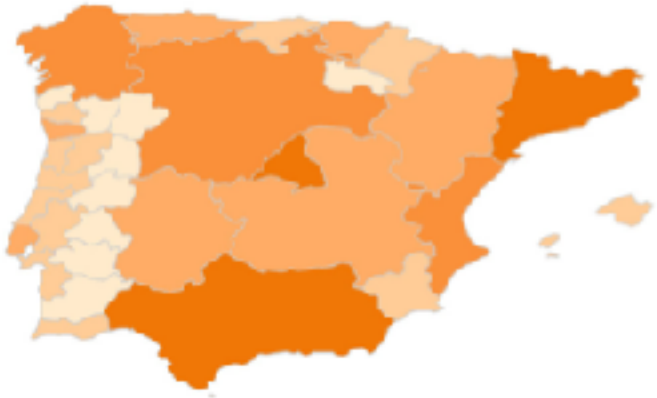
  (MacEachren, 1998)

Why Cartographic/Map Viz?

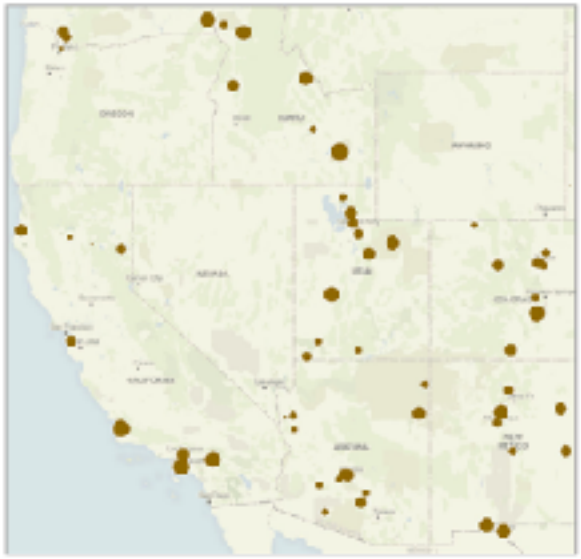
- Dynamic and interactive visualization of geospatial information
  - Focus and Context
  - Linked highlighting
  - Fluid navigation
- Spatial visualization of non-geospatial data
  - Cartographic principles

Two approaches best displaying quantitative information on maps

1. Color Symbols:
   - Variations in color intensity



2. Proportional Symbols:
   - Variations in size



or combination both

Color Symbols

- The main goal is to order lightness (choropleth map)
  - so it parallels ordering in the data.
- Use light-to-dark color for low-to-high values with a constant hue
- A lightness sequence combined with a progression through adjacent hues
  - yellow-green-blue are adjacent in the ordering of hues through the spectrum
- Spectral (rainbow) schemes are not well suited to sequential data because lightness varies through the spectrum
  - Spectral schemes can be adjusted to better order lightness
  - the light yellow hue can be used to emphasize critical values within a data range.

Color Theory

- In painting, the word **COLOR is the general term** for everything we see.
- **Hue, Tint, Tone and Shade** are the four categories of **color**.
- However, the word **HUE** refers to the brightest 6 - 12 pure, **unmixed pigment families on the Color Wheel.**
- **a TINT any Hue with White added.** The color remains the **same only lighter.**
- **a TONE is any pure Hue with Neutral Gray added.** The color **remains the same only less vibrant.** The **Values can range** from very light to very dark.
- **a SHADE is any pure Hue with Black added.** The color remains the same only darker.

https://color-wheel-artist.com/hue/

Color Symbols

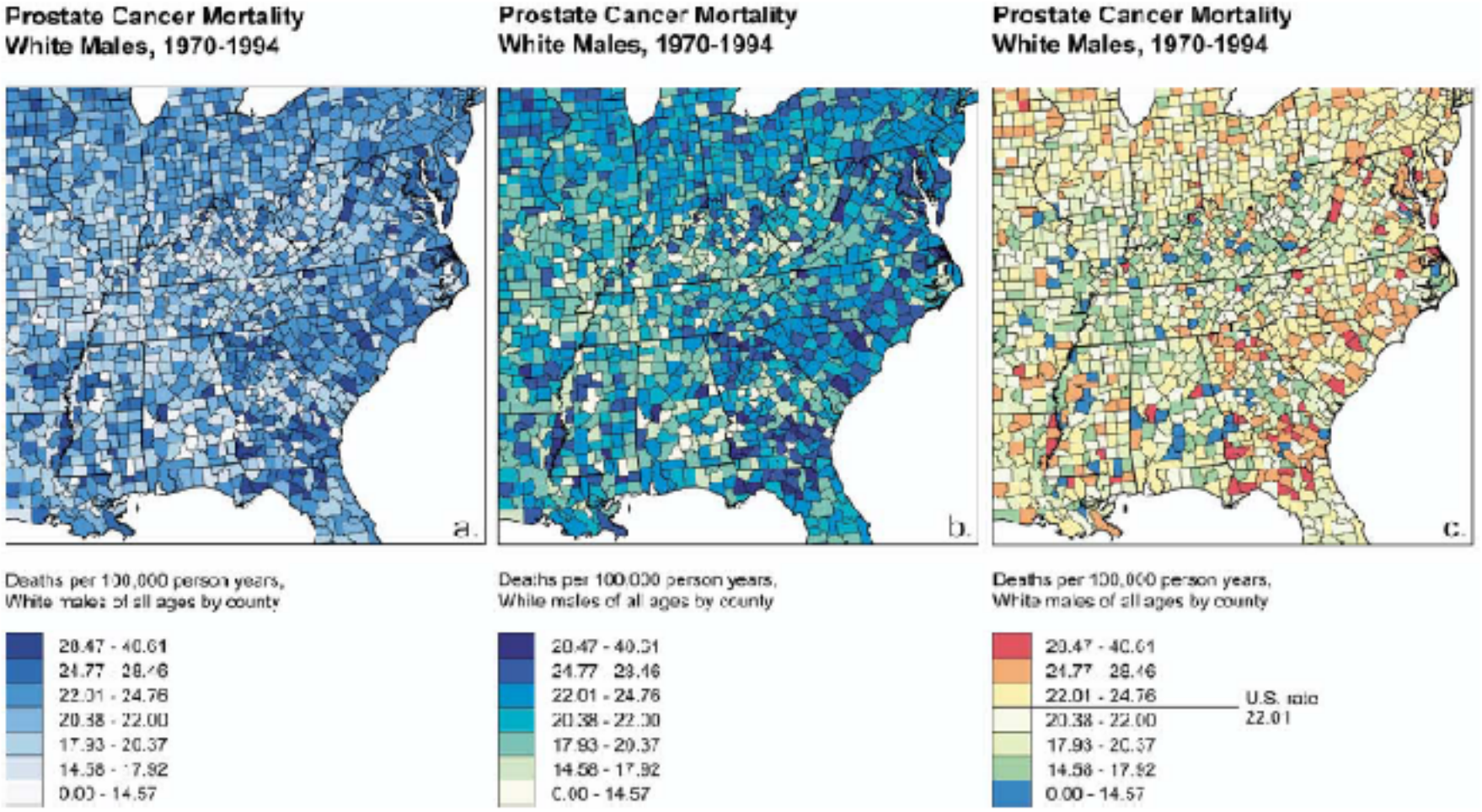
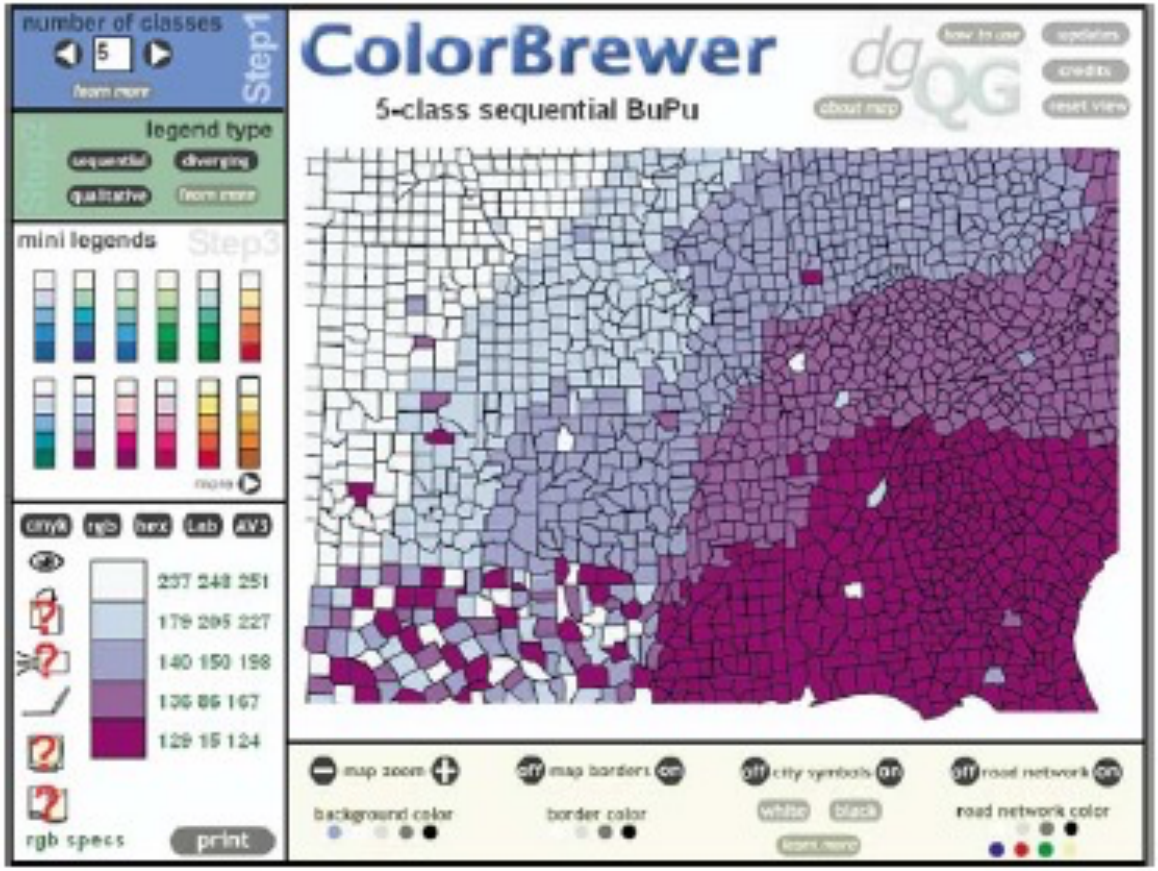


Figure 2. Three color schemes are shown for the same data set. (a) Sequential, single hue scheme (blue). (b) Sequential scheme with hue transition (yellow-green-blue). (c) Spectral scheme. The spectral scheme is used as a diverging scheme with the lightest colors marking the overall U.S. rate.

Sequential Schema: Available Tool



ColorBrewer.org

Diverging Schema

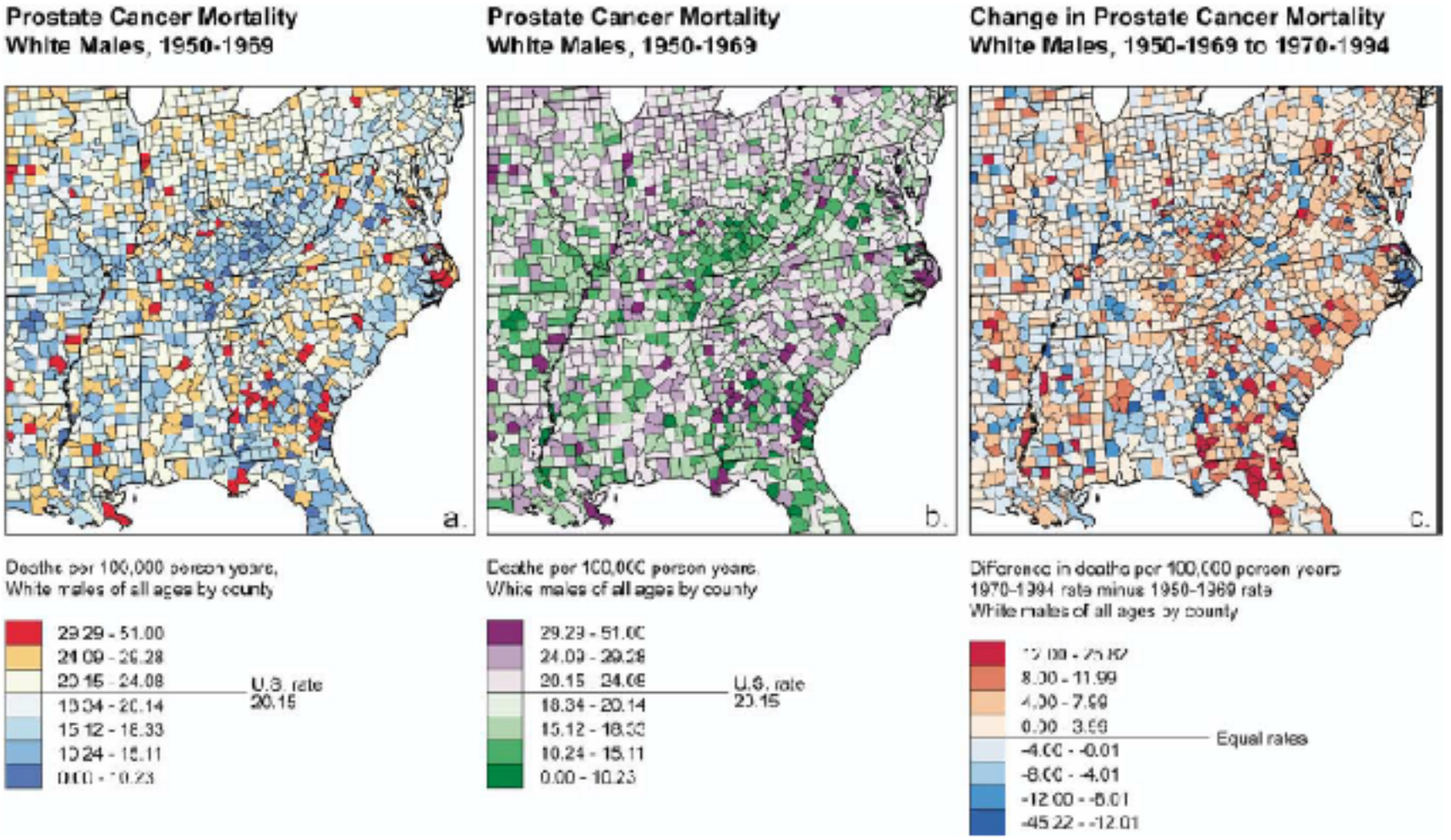


Figure 4. Example diverging schemes. (a) Spectral scheme modified to accommodate color blind map readers by skipping green hues. (b) Two hues (green and purple) diverging from a central light class at the U.S. rate. (c) Change in rates between two time periods with diverging reds (increasing rates) and blues (decreasing).

Classing Methods

1. Quantiles
   - Assigns the same number of enumeration units to each class (it is a generalized form of percentiles).
   - Four quantiles (quartiles) allocate one quarter of the data values to each class with the median at the middle break

2. Equal intervals
   - Equal interval classing breaks the data range into equal segments for predictable and equal class ranges

3. Jenks optimized method (natural breaks)
   - Minimize variation within classes and maximize variation between classes.

Classing Methods: Quantiles

- Assigns the same number of enumeration units to each class (it is a generalized form of percentiles).
- Four quantiles (quartiles) allocate one quarter of the data values to each class with the median at the middle break
- Each class is equally represented on the map and the classes are easy to compute.
- When using quantile classification gaps can occur between the attribute values: lead to an over-weighting of the outlier in that class division.
- If the number of classes is not correctly created, two areas with the same value can end up in different groups.

Classing Methods: Equal Intervals

- Equal interval classing breaks the data range into equal segments for predictable and equal class ranges

$$\frac{\text{Range of Data}}{\text{Number of Classes}} = \frac{(\text{Highest Value} - \text{Lowest Value})}{\text{Number of Classes}} = \frac{(100 - 10)}{15} = 6$$

- The intervals can easily be computed
- The legend limits contain no missing values or gaps
- The projections are easily interpreted
- It **fails** to consider how data are distributed along the number line.

Classing Methods: Jenks Optimized

- **Minimize** variation **within** classes and **maximize** variation **between** classes.
- **Natural Breaks**: designed to optimize the arrangement of a set of values into "natural" classes.
- The features are divided into classes whose **boundaries** are set where there are relatively big jumps in the data values: identify logical **break** points.

Classing Methods: Jenks Optimized

- **Minimize** variation **within** classes and **maximize** variation **between** classes.
- It identifies **real** classes within the data.
- It classifies data **subjectively**.
- Not recommended for data that have a low variance.
- The algorithm can become very **long** with **large** data sets
  - It is a successful one when attempting to **decrease** the amount of **deceptive** information.

Classing Methods Examples

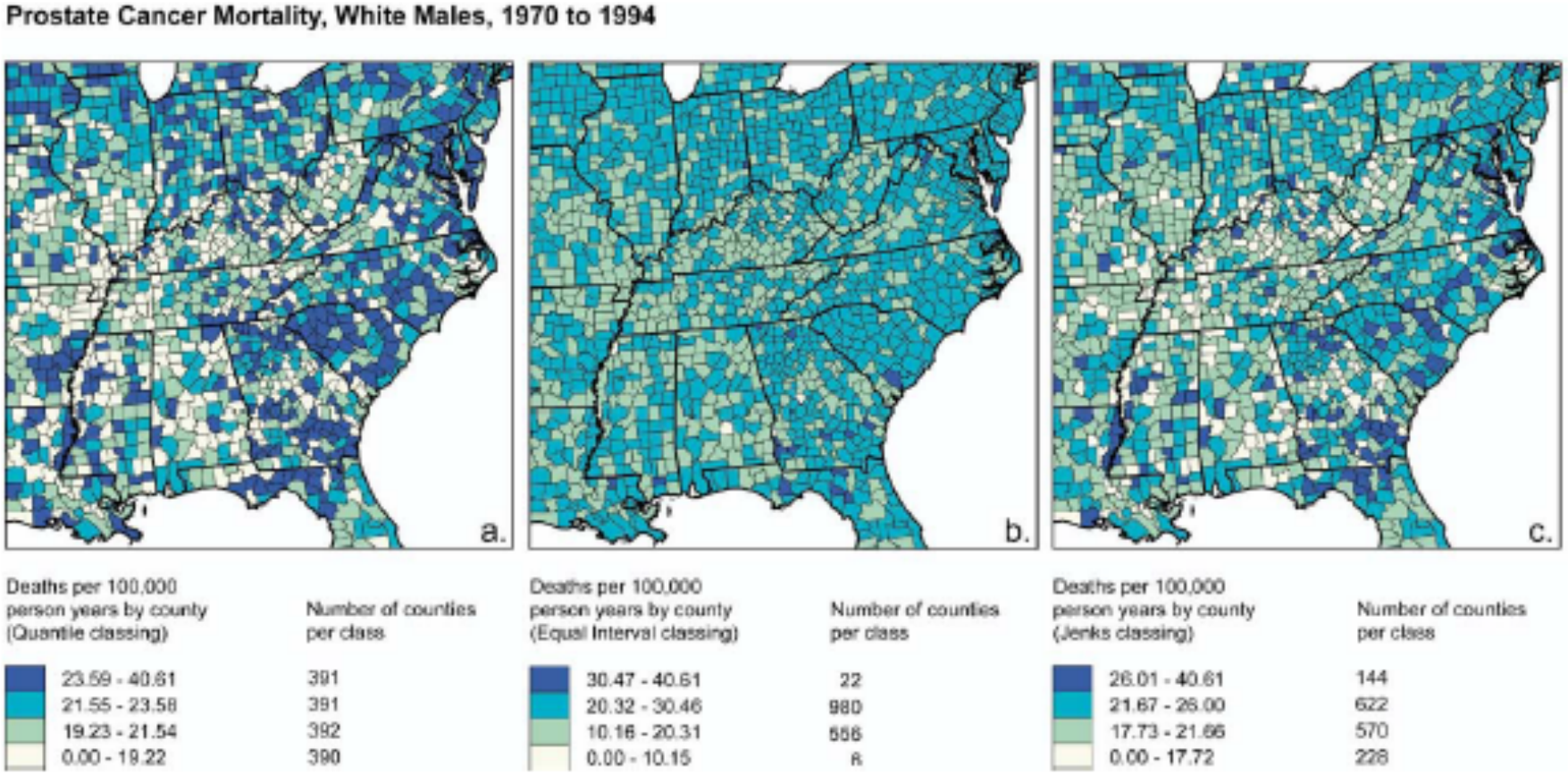


Figure 5. Three classifications of the same data set showing different patterns resulting from different classing methods. (**a**) Quantile. (**b**) Equal interval. (**c**) Jenks optimized classification (natural breaks). The number of counties in each class is shown to the right of each legend.

Mapping Multiple Variables: Time Series

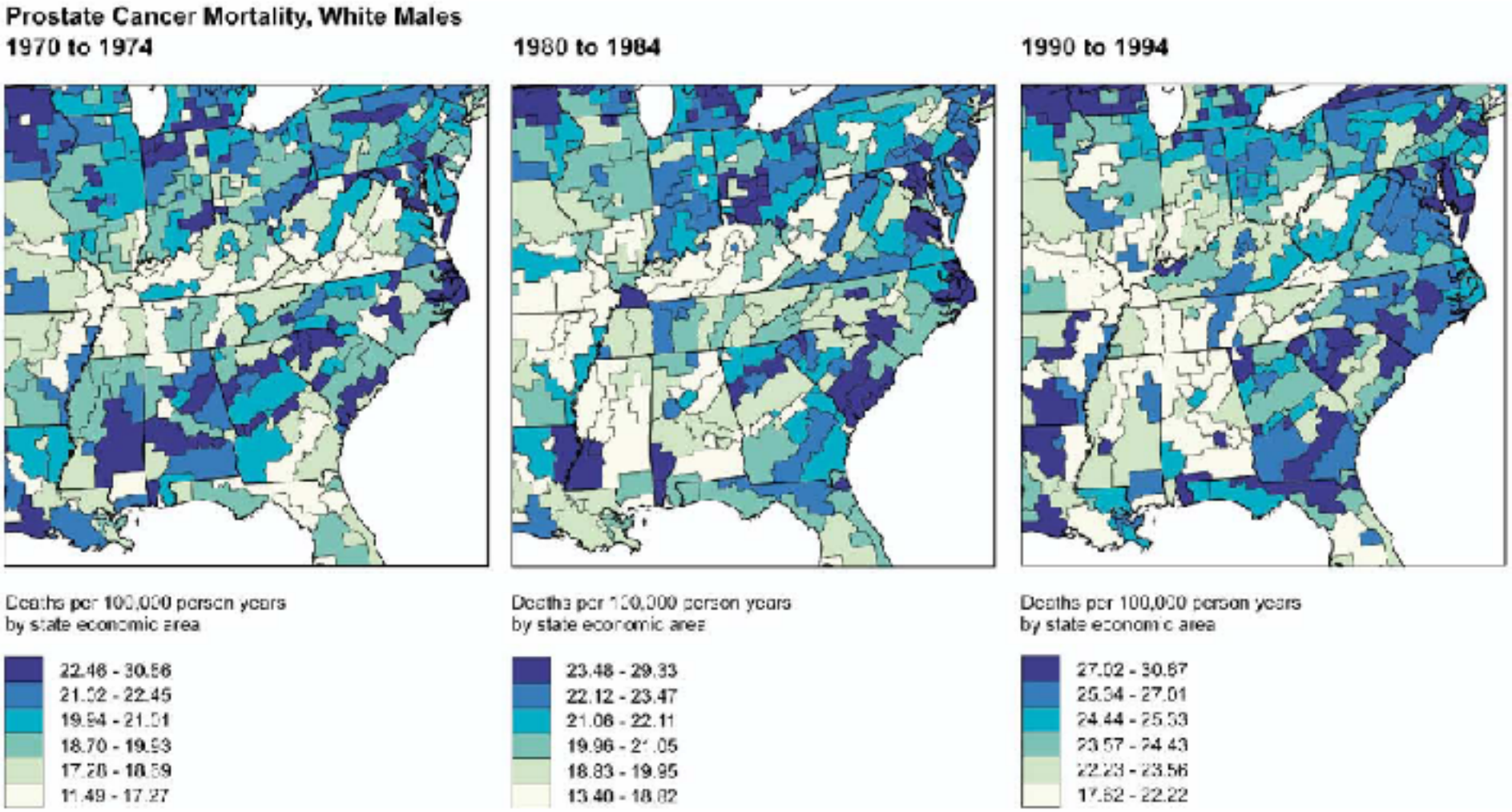


Figure 9. Example of a map series with each map classed separately using quantile classing. The maps are a time series.

Mapping Multiple Variables : Class Breaks

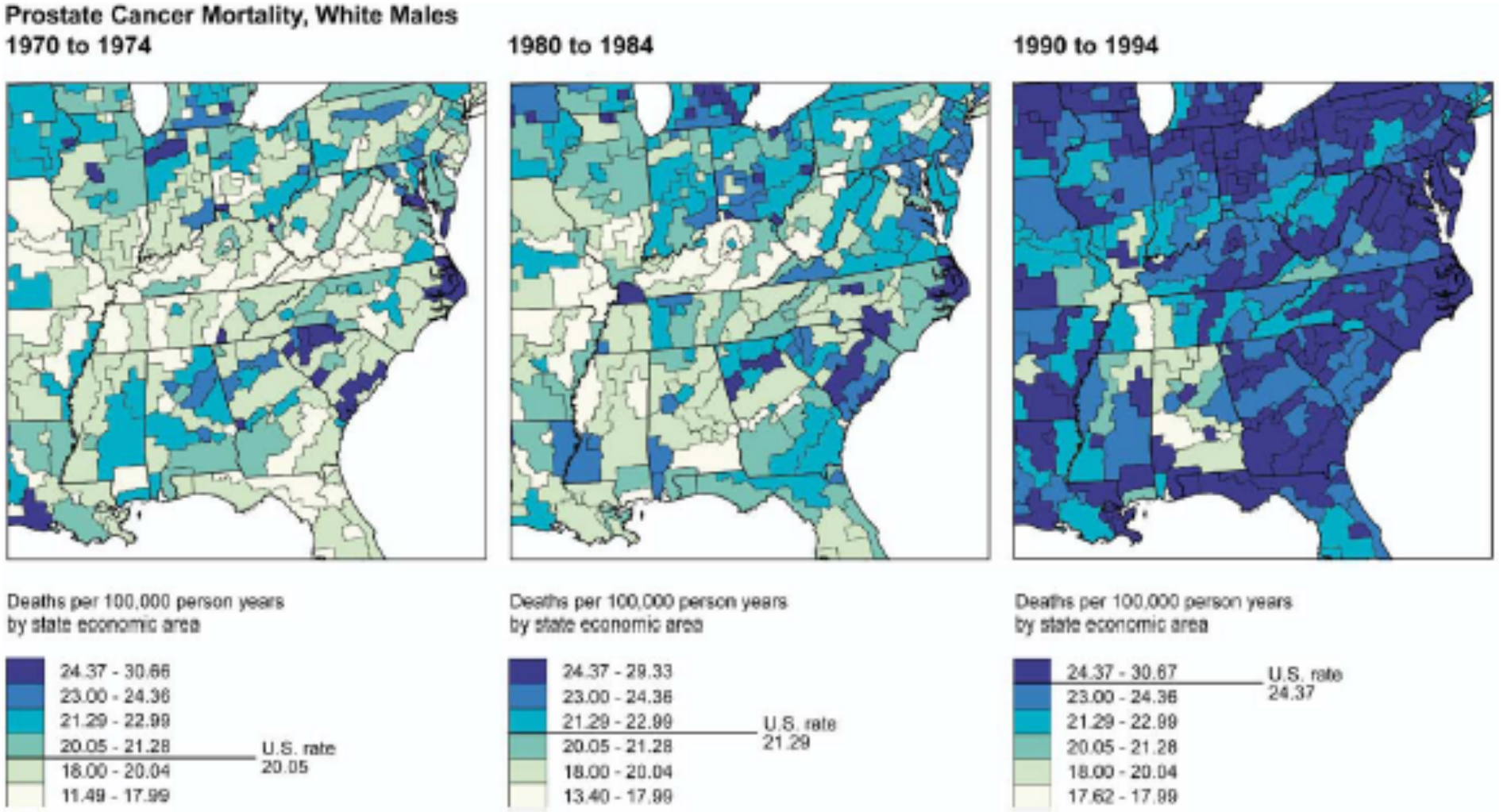
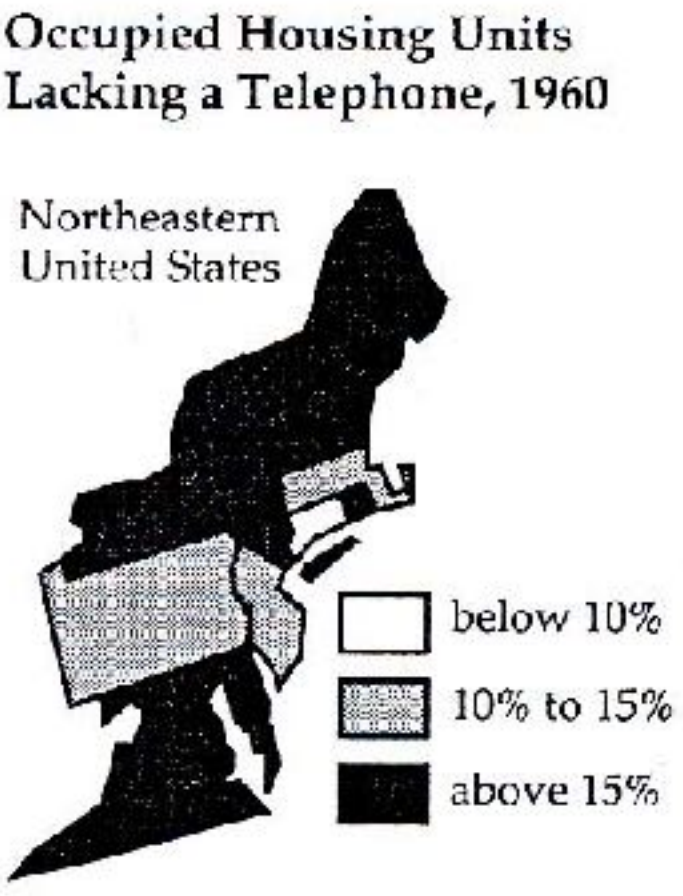
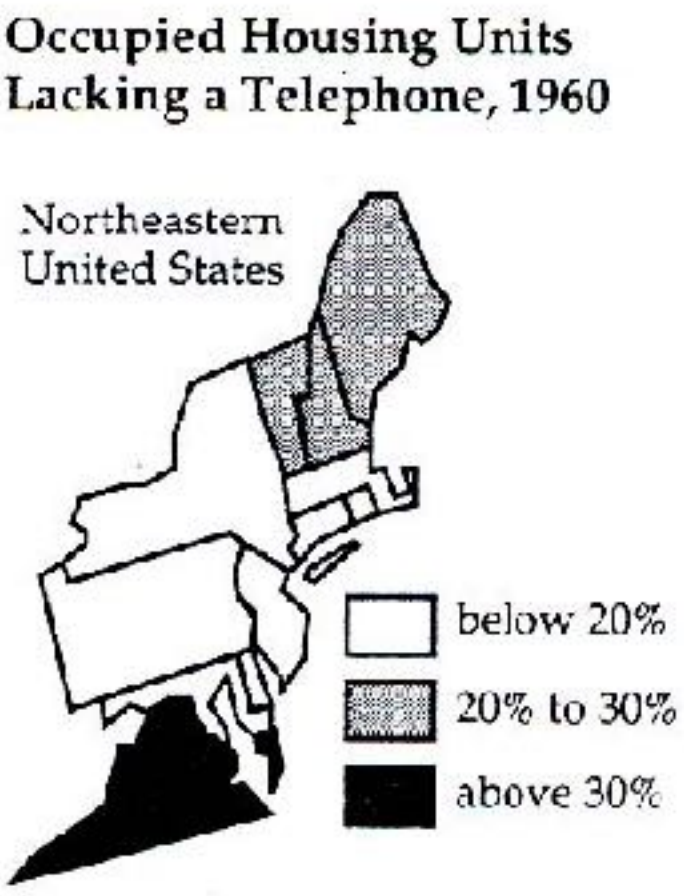


Figure 10. The same map series seen in Figure 9 with all maps sharing the same set of classes to aid map comparison within the time series. Class breaks based on the U.S. rate for each time period are included on all maps; the U.S. rate for the 5-year period mapped is highlighted in each maps' legend.



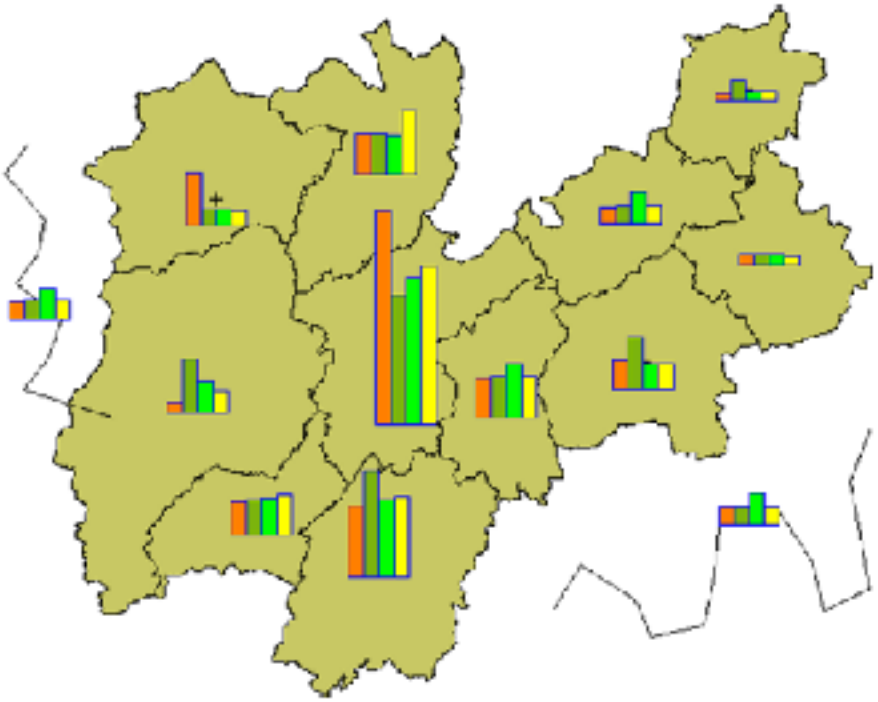


- Be careful not to lie for your interest
- Most of it is bad maps intentionally created to distort or hide data.

Proportional Symbols

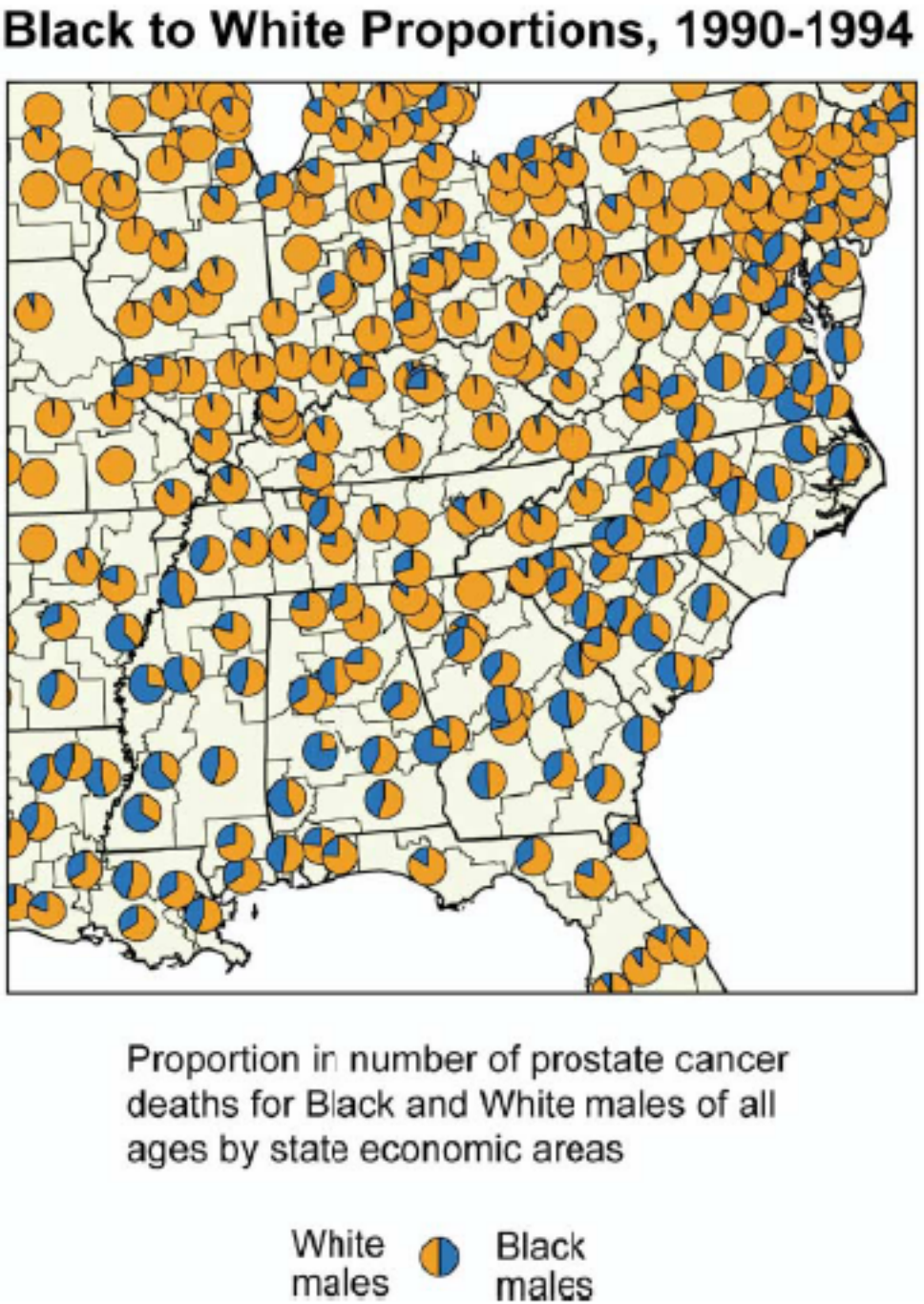
- Symbols such as circles and squares are usually scaled in proportion to the square root of each data value
- A symbol scaled by area, such as a square, is more compact and easier to associate with the location for which it represents data
  - Sizes of linearly scaled symbols, such as bars, are more accurately interpreted by map readers, but they become impractical with large data ranges.
- The order of symbols:
  - the smaller symbols appear above larger ones to aid map reading.

Using Bar Graph



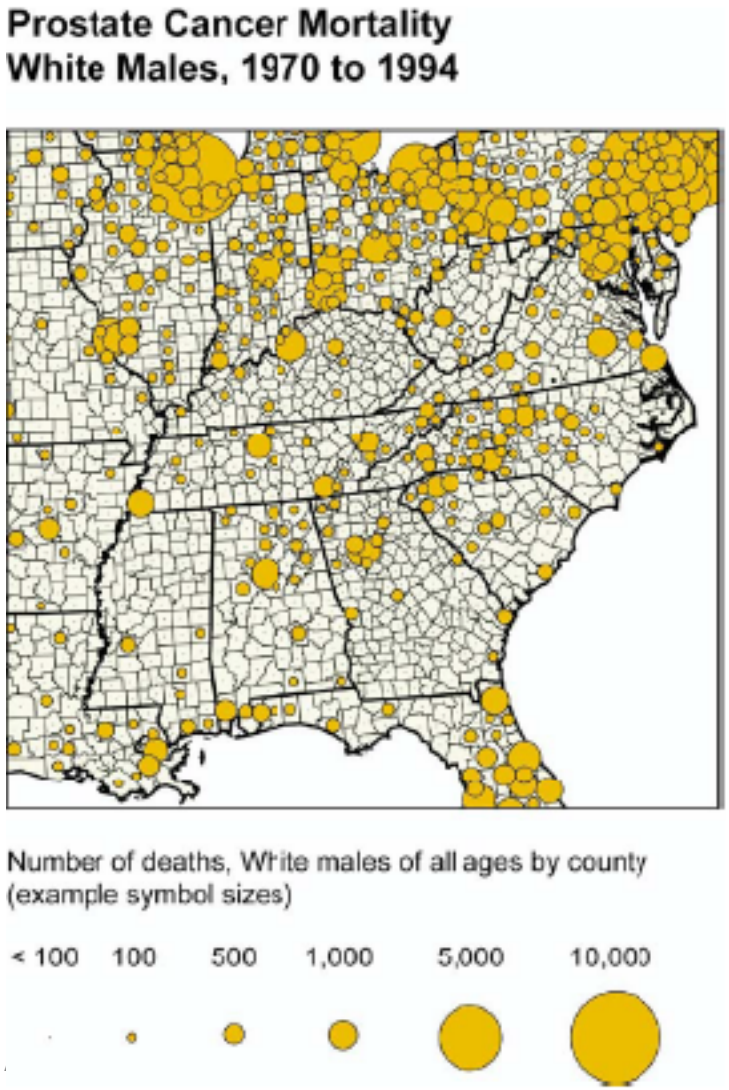
- Problems:
  - The sets of bars don't share a common baseline
  - There wouldn't be enough room to place bar charts everywhere they're needed without overlapping them

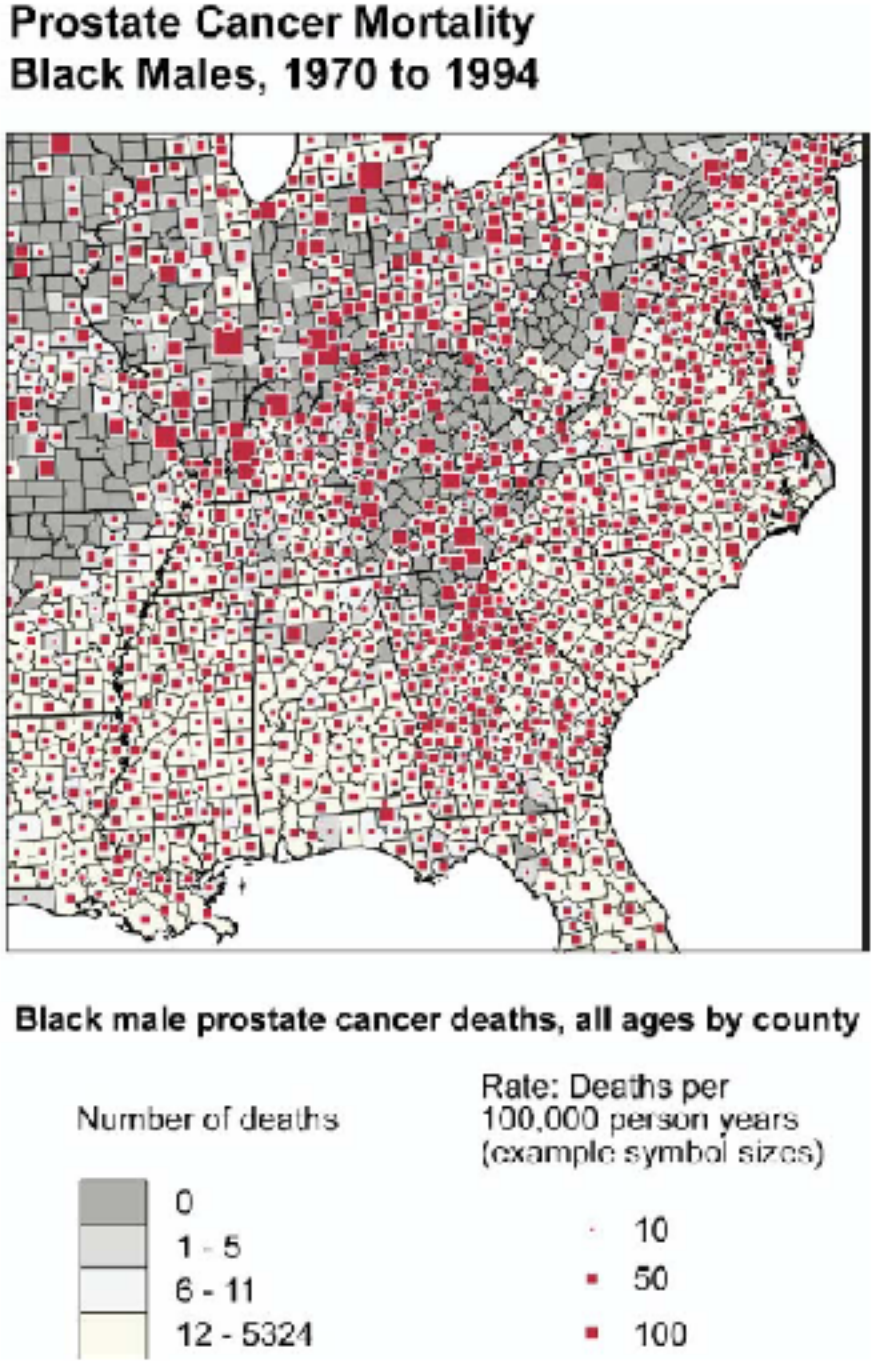
Using Pie Chart

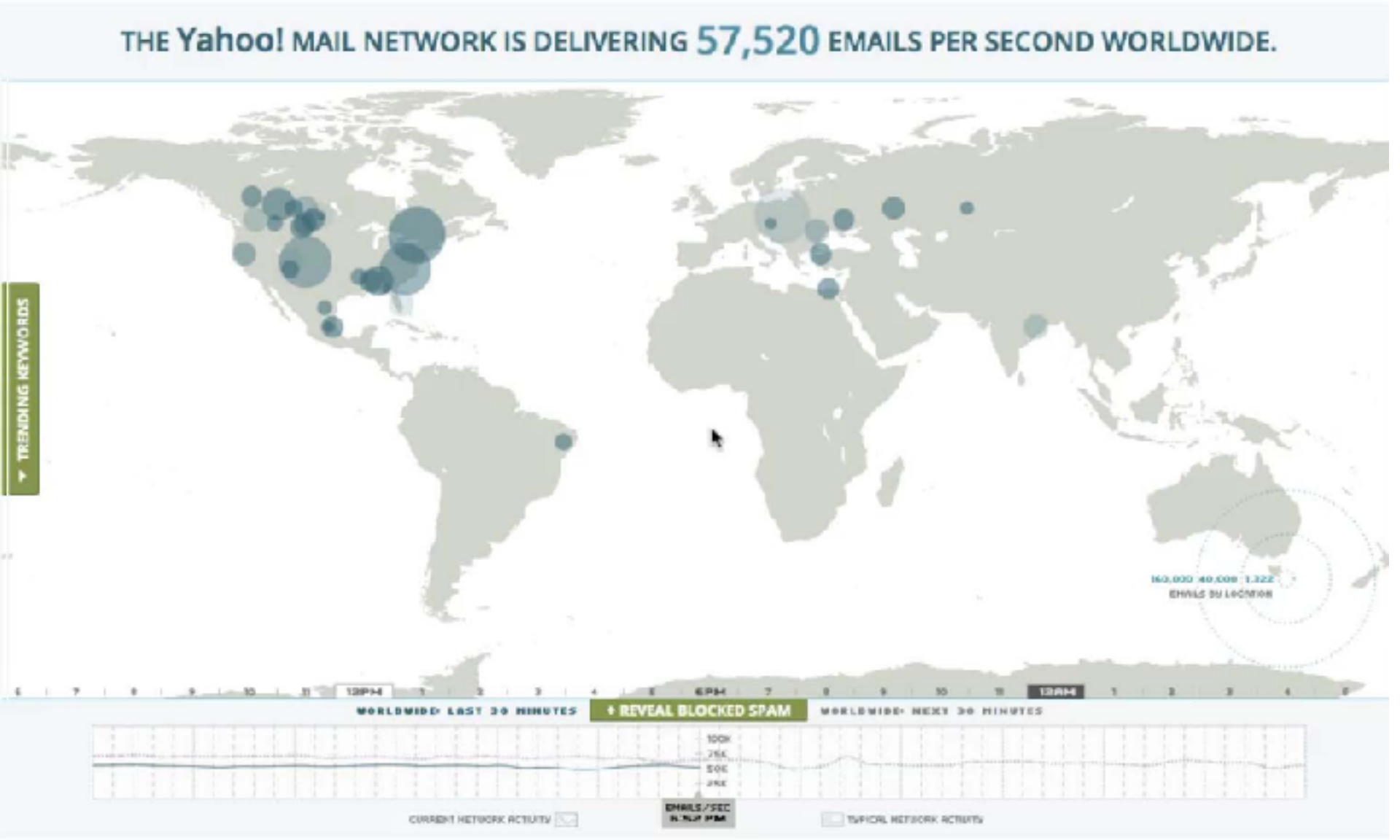



World Population Distribution
Maldives 295,297
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image IBCAO
© 2011 DigitalGlobe
Image © 2011 GeoEye
Google

Proportional Symbols



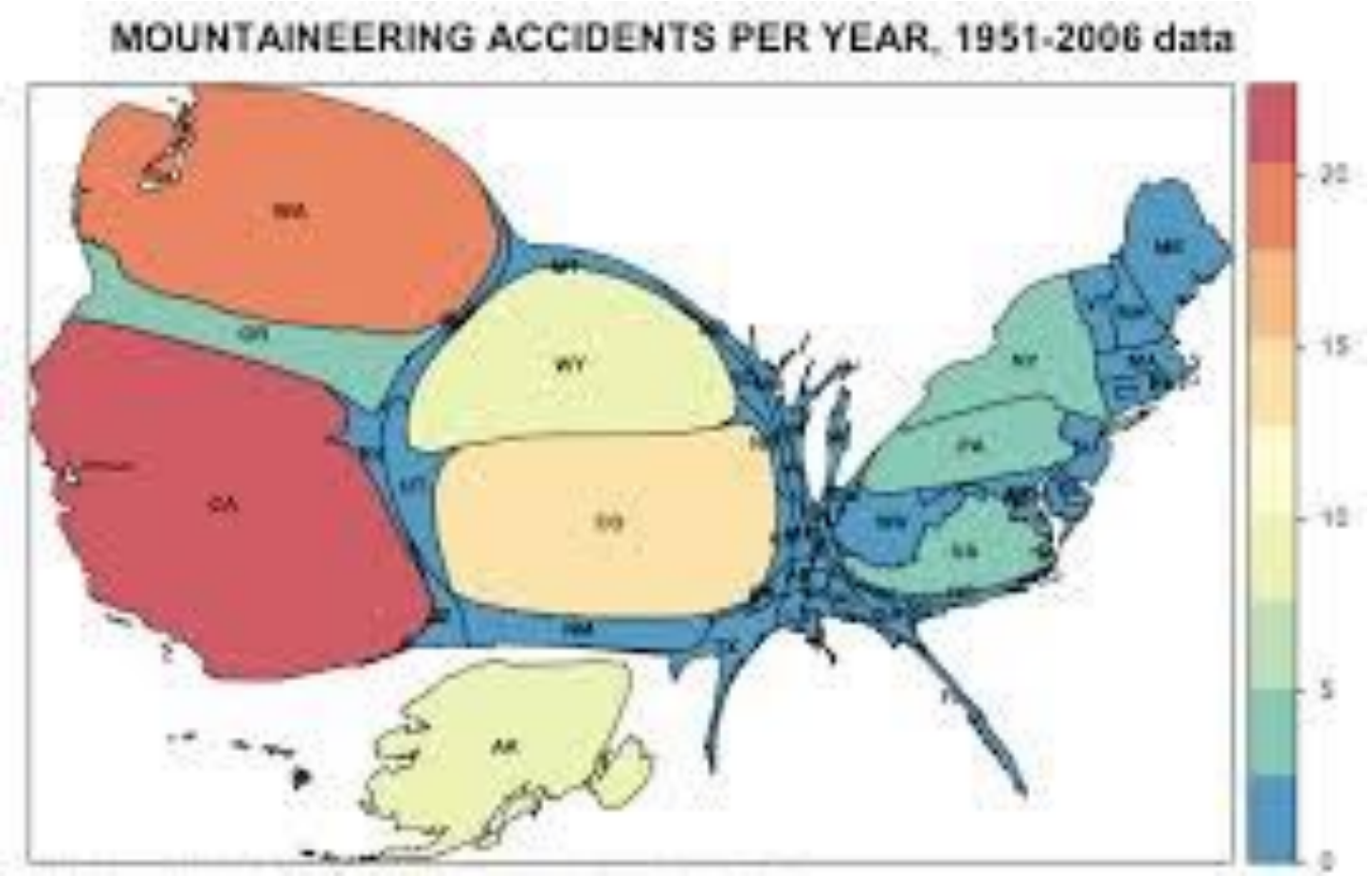


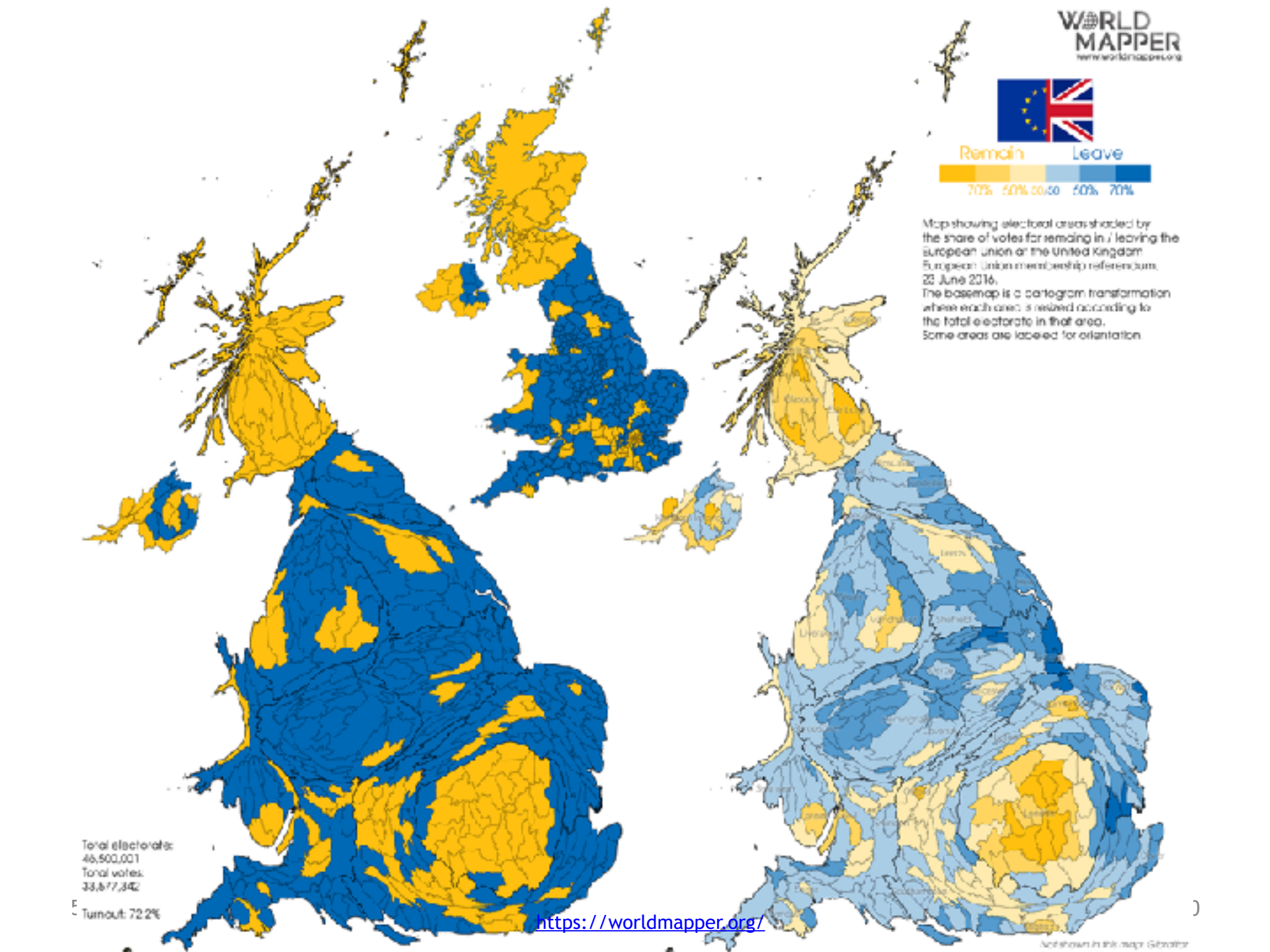

THE Yahoo! MAIL NETWORK IS DELIVERING 57,520 EMAILS PER SECOND WORLDWIDE.
TRENDING KEYWORDS
WORLDWIDE LAST 30 MINUTES
REVEAL BLOCKED SPAM
WORLDWIDE NEXT 30 MINUTES
CURRENT NETWORK ACTIVITY
TYPICAL NETWORK ACTIVITY

Other Symbols

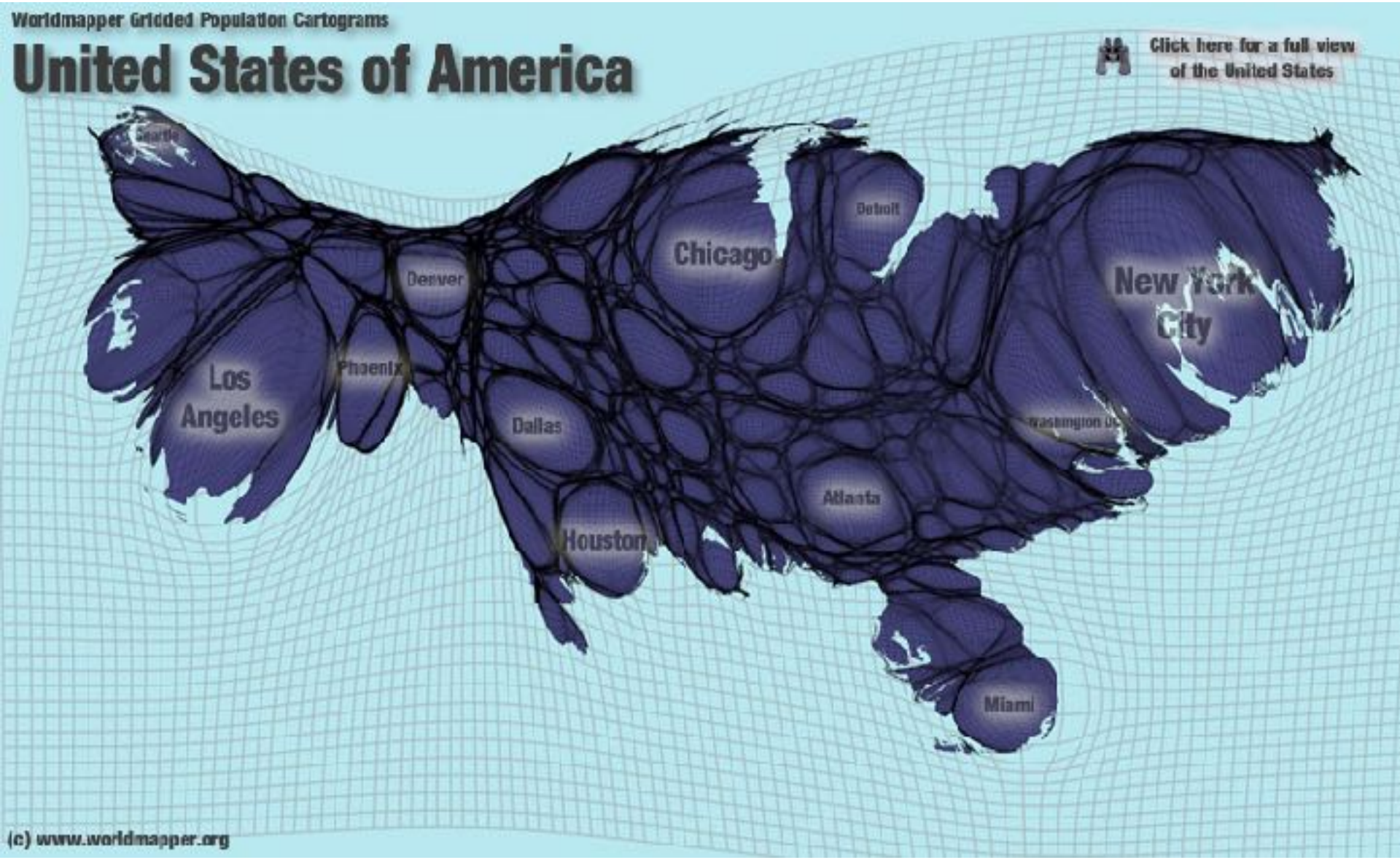
- Cartograms
  - forms of the enumeration units warped or resized to produce areas proportioned to data values
- Dot density representations
  - vary the number of dots in enumeration areas in proportion to data values (for example, one dot represents 100 people)
  - well suited to sparse or discrete phenomena.
- Isolines, filled isolines, and smooth gradation
  - well suited for Continuous surfaces, such as air quality
- Flow lines or networks, and “spider” maps
  - Can be used for linear phenomena
  - offer a version of linear symbol sometimes used for epidemiologic data, such as connecting locations to service points.

Cartograms

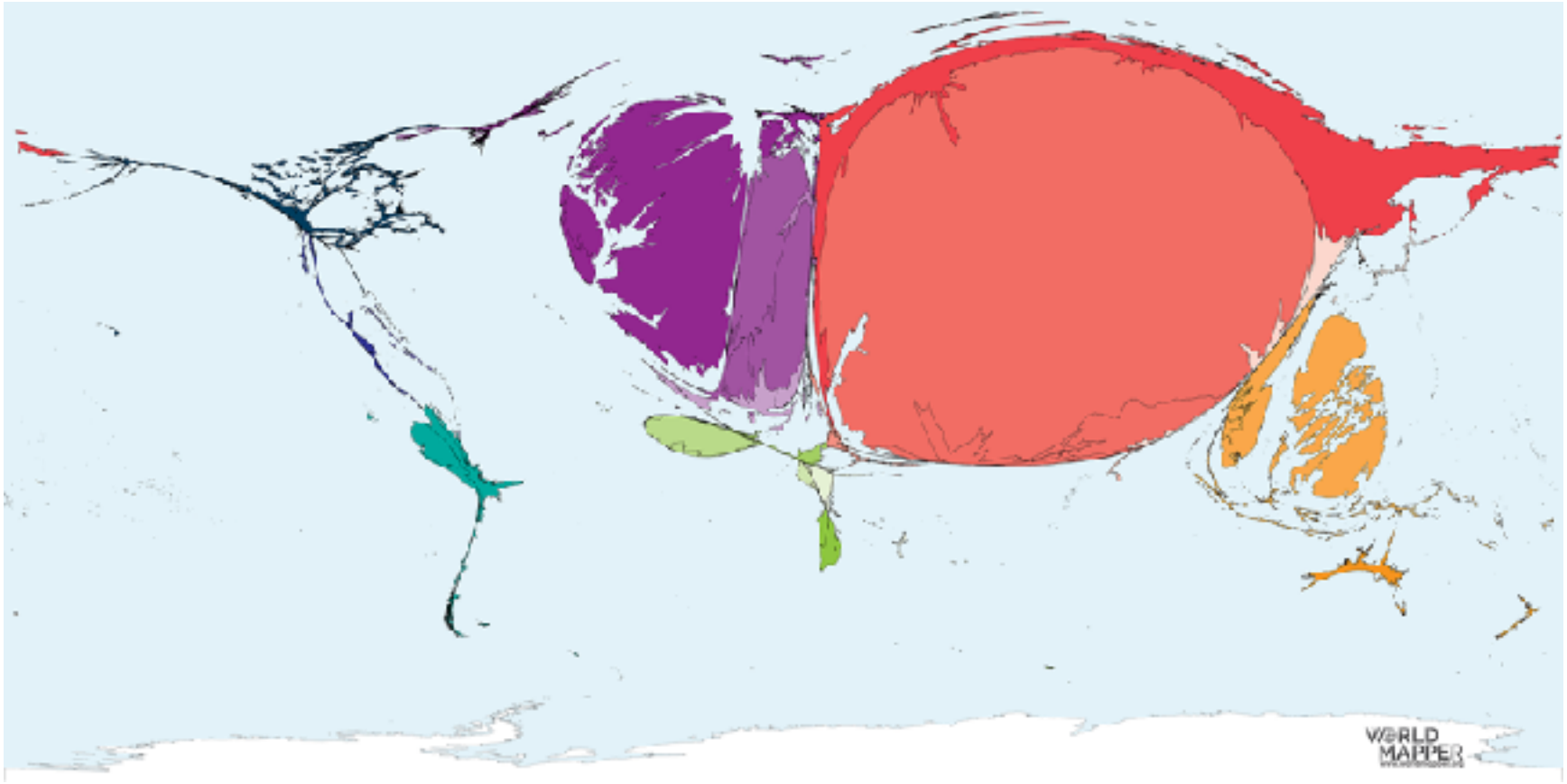

MOUNTAINEERING ACCIDENTS PER YEAR, 1951-2006 data
WA
OR
CA
WY
NM
AK
20
15
10
5
0


WORLD MAPPER
Remain
Leave
70% 60% 50:50 60% 70%
Map showing electoral areas shaded by the share of votes for remaining in / leaving the european union of the United Kingdom European Union membership referendum, 23 June 2016.
The basemap is a cartogram transformation where each area is resized according to the total electorate in that area.
Some areas are labeled for orientation
Total electorate: 46,500,001
Total votes: 33,577,342
Turnout: 72.2%
https://worldmapper.org/

Cartograms

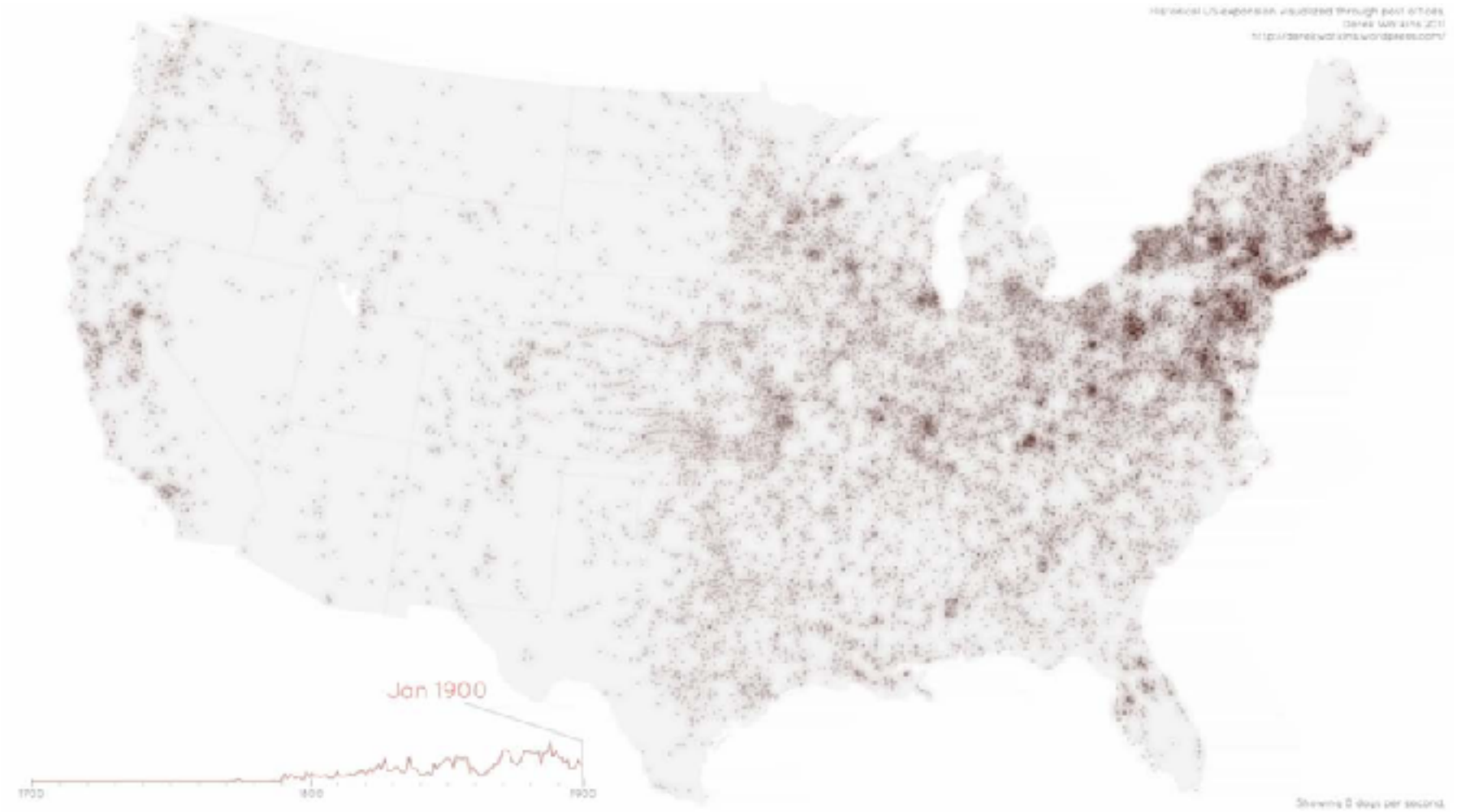

Worldmapper Gridded Population Cartograms
United States of America
Click here for a full view of the United States
Seattle
Denver
Chicago
Detroit
New York City
Los Angeles
Phoenix
Dallas
Atlanta
Houston
Miami
(c) www.worldmapper.org

As long as you look on migration as a problem, as something to solve, you're not going to get anywhere. You have to look at it as a human reality that's as old as humankind. It's mankind's oldest poverty reduction strategy. As citizens, we have to find a way to manage it.
William Lacy Swing, Director General International Organization for Migration (IOM), 2017

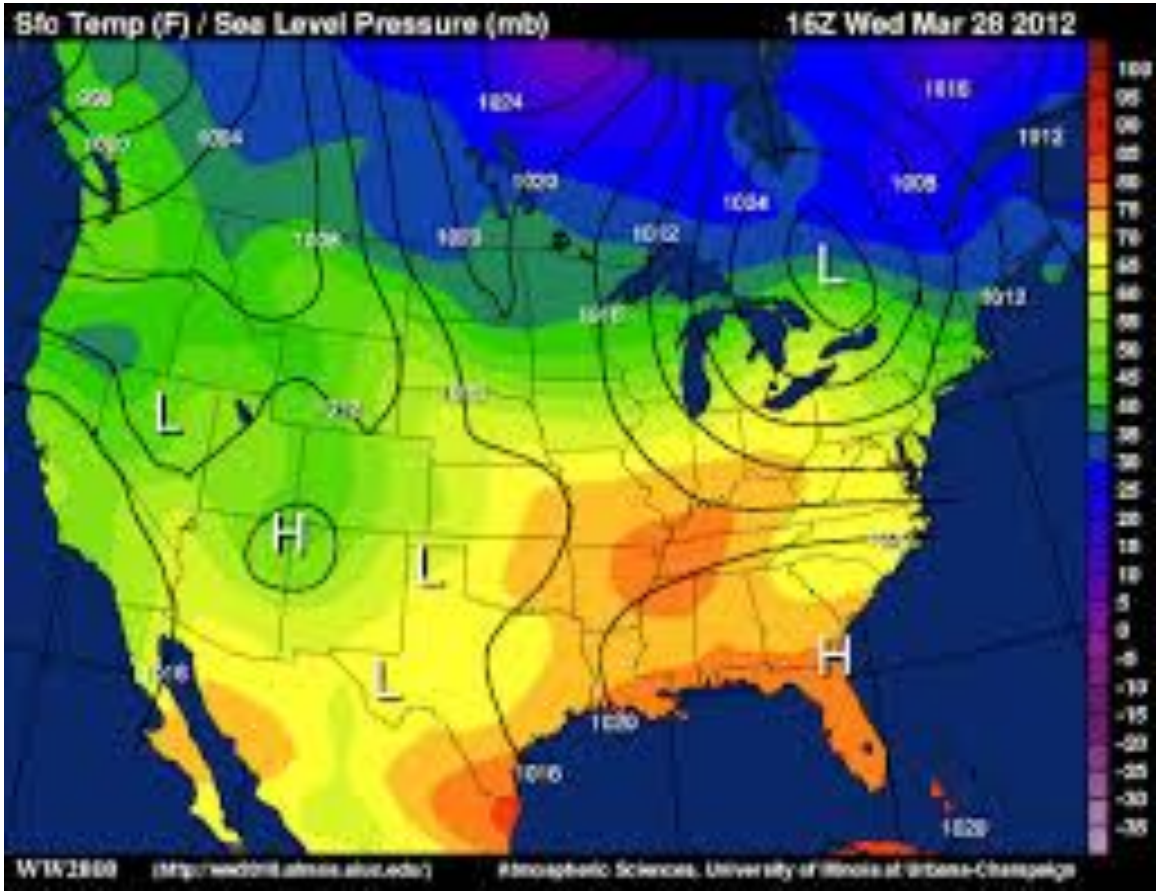


Migration from North Korea 1990-2017

Dot Density Representation

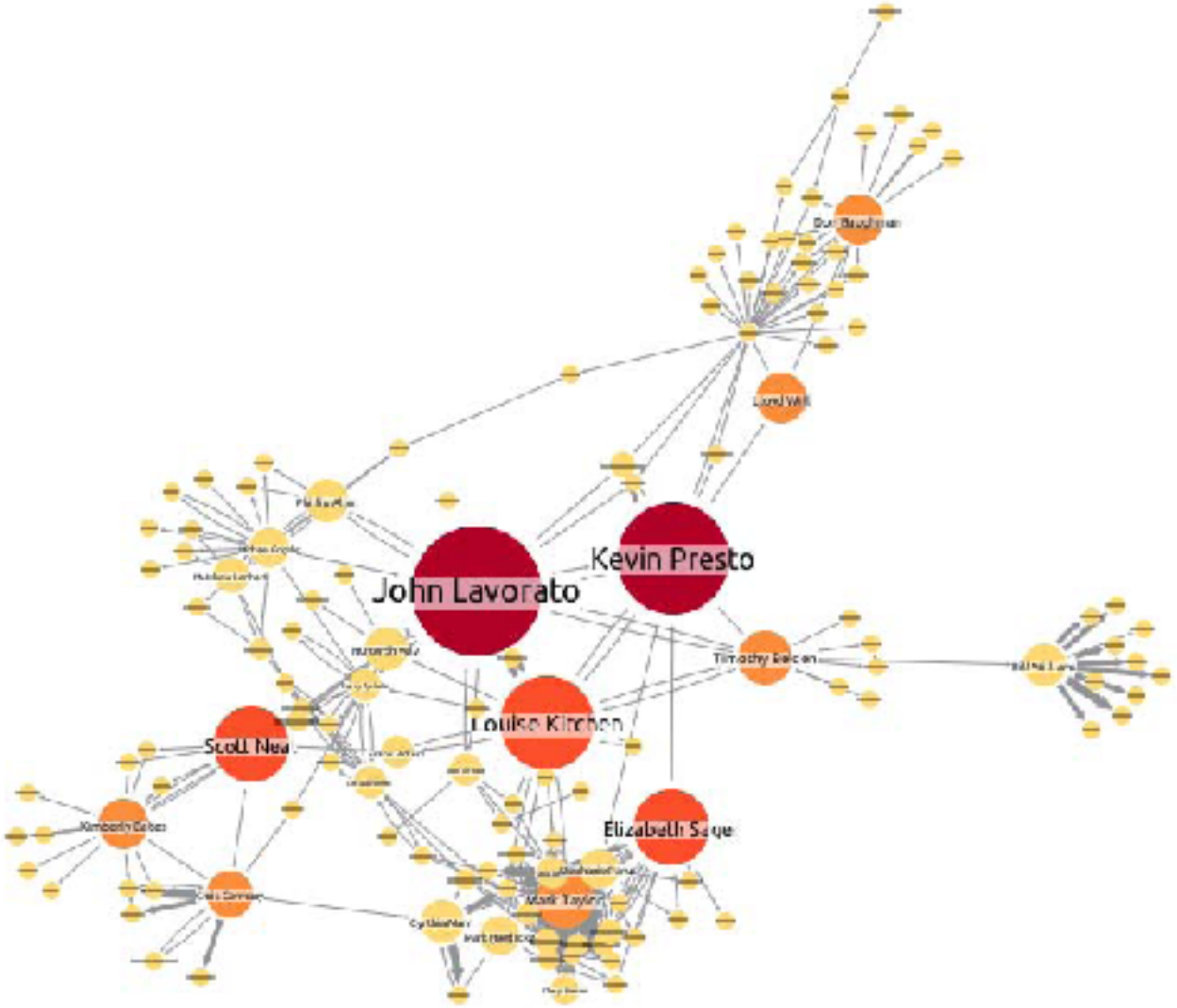


Isolines



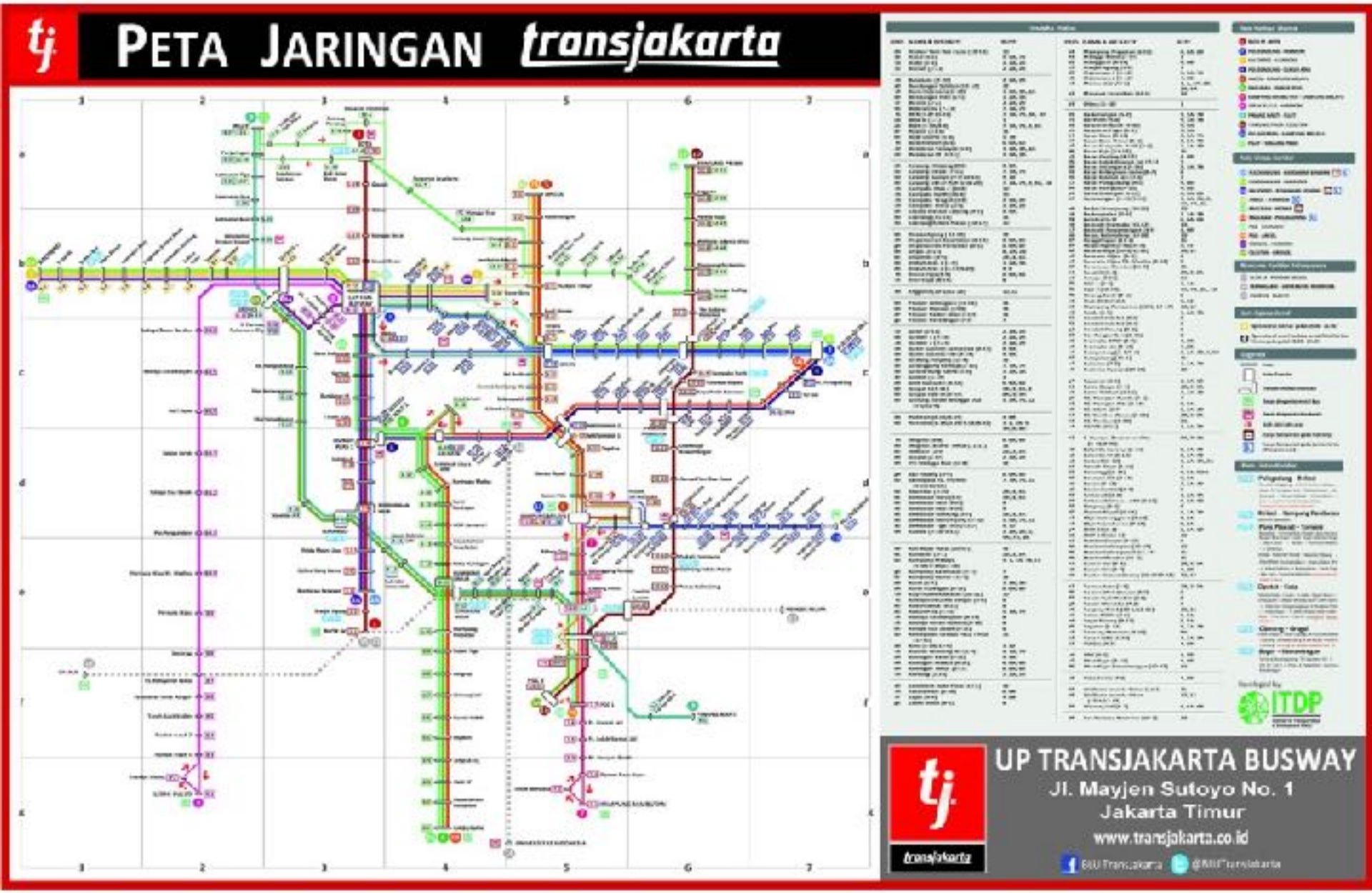
Flow line

Network Diagram



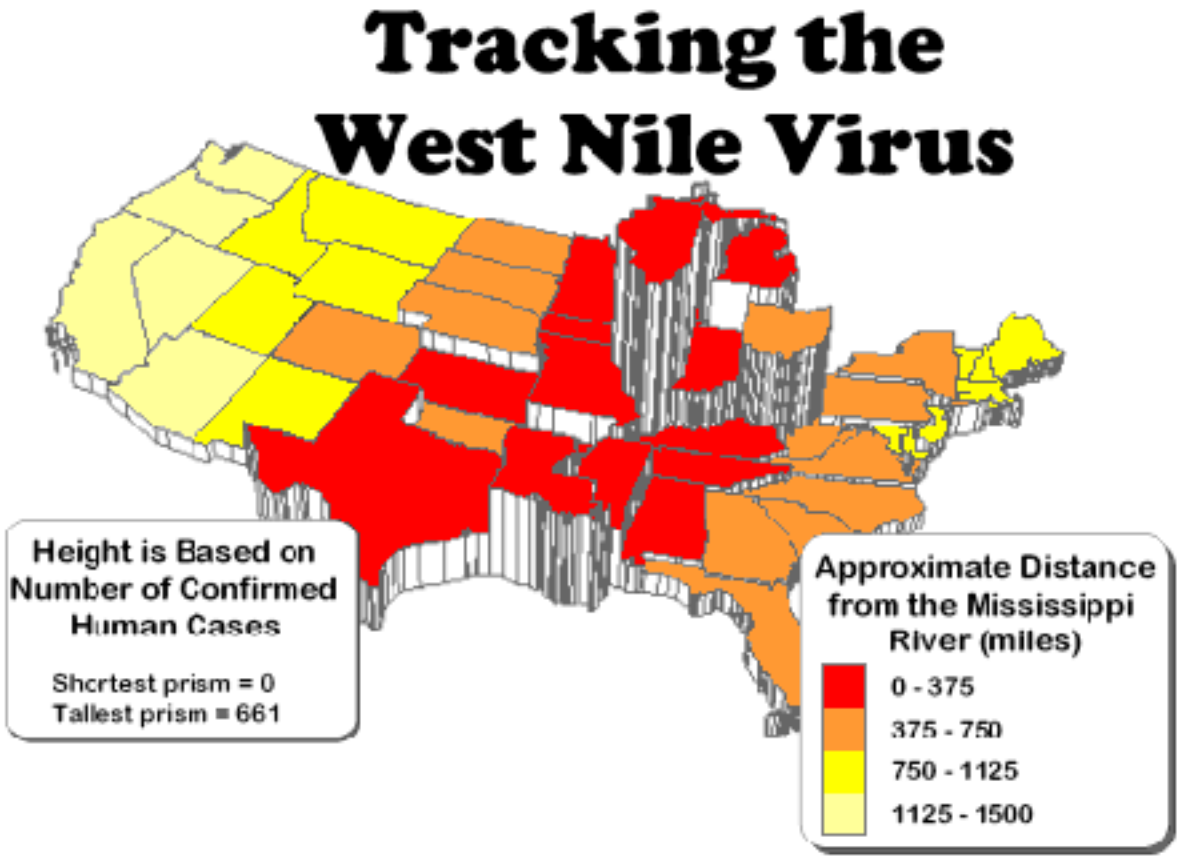
Subway Map


Tube map
MAYOR OF LONDON
tfl.gov.uk
0843 222 1234
Help points
Transport for London
UNDERGROUND


PETA JARINGAN transjakarta
UP TRANSJAKARTA BUSWAY
Jl. Mayjen Sutoyo No. 1
Jakarta Timur
www.transjakarta.co.id
ITDP

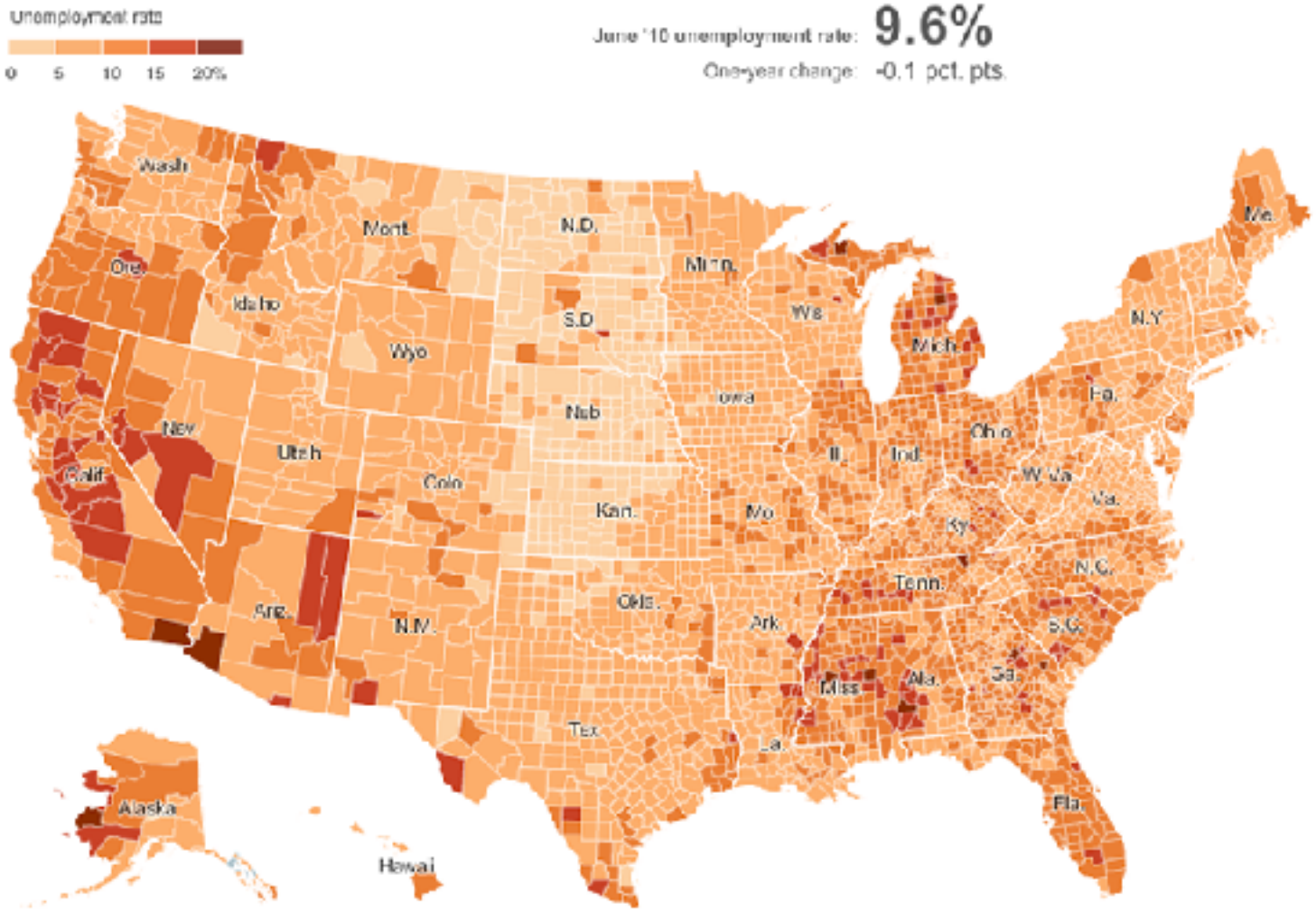

Tangerang
Tanah Tinggi
Batu Ceper
Poris
Kali Deres
Rawa Buaya
Bojong Indah
Taman Kota
Pesing
Grogol
Angke
Kampung Bandan
Ancol
Tanjung Priok
Jakarta Kota
Duri
Tanah Abang
Jayakarta
Mangga Besar
Sawah Besar
Juanda
Gambir
Gondangdia
Cikini
Manggarai
Rajawali
Kemayoran
Pasar Senen
Gang Sentiong
Kramat
Pondok Jati
Jatinegara
Palmerah
Kebayoran
Pondok Ranji
Jurangmangu
Sudimara
Rawa Buntu
Serpong
Cisauk
Cicayur
Karet
Sudirman
Mampang
Tebet
Cawang
Duren Kalibata
Pasar Minggu Baru
Pasar Minggu
Tanjung Barat
Lenteng Agung
Cipinang
Klender
Buaran
Klender Baru
Cakung
Kranji
Bekasi

Prism Map


Tracking the
West Nile Virus
Height is Based on
Number of Confirmed
Human Cases
Shortest prism = 0
Tallest prism = 661
Approximate Distance
from the Mississippi
River (miles)
0 - 375
375 - 750
750 - 1125
1125 - 1500

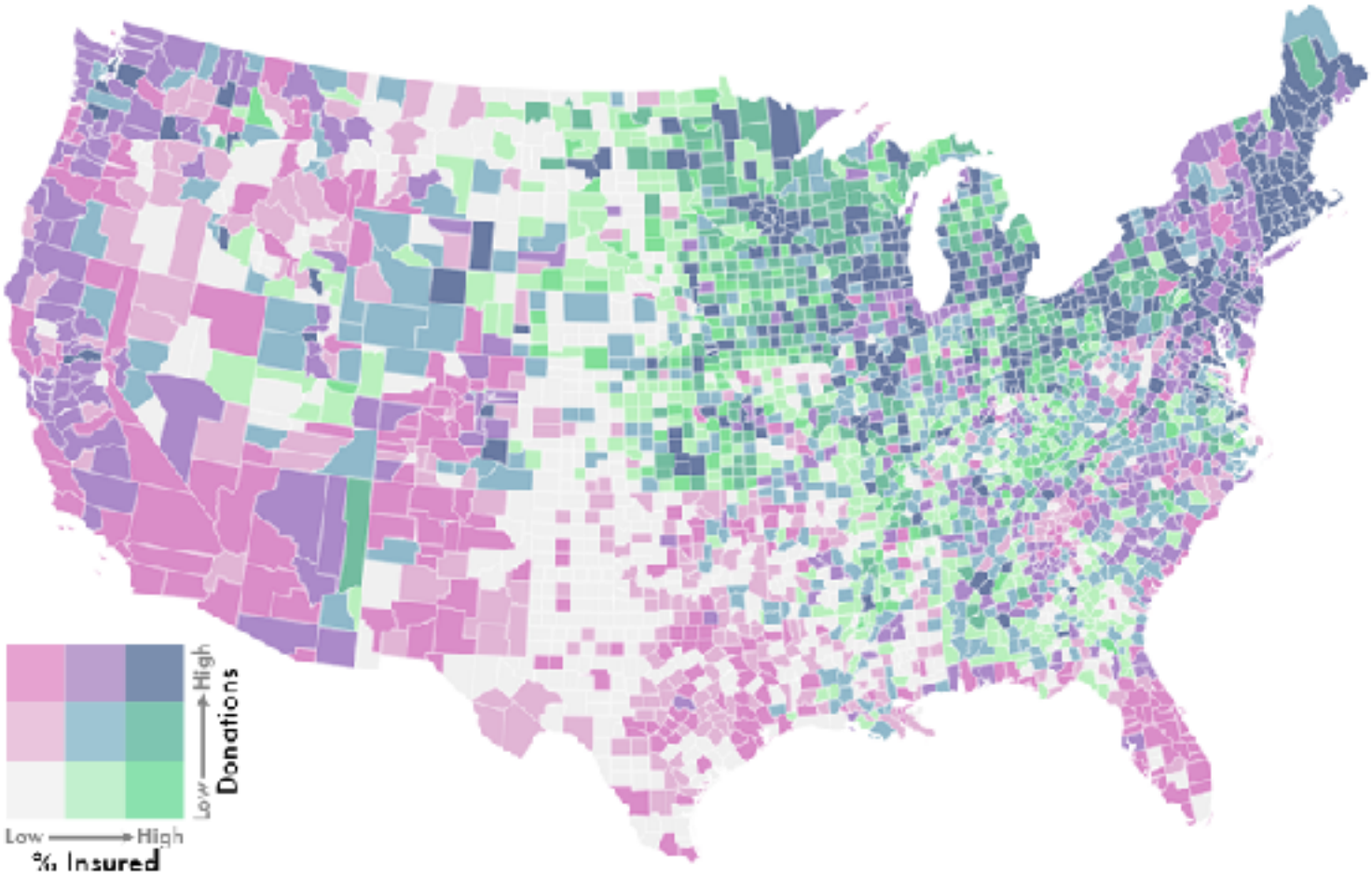
Other Aspects

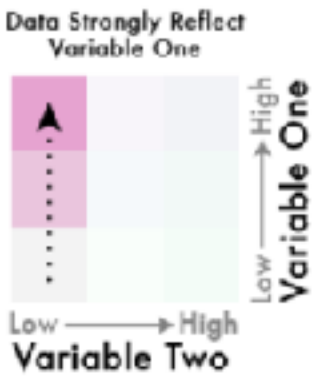
- Normalize, normalize, normalize
- Featuring the data
  - Stand out the data
- Good uses of geographical data displays
- Collaboration between Maps and other Visual Displays

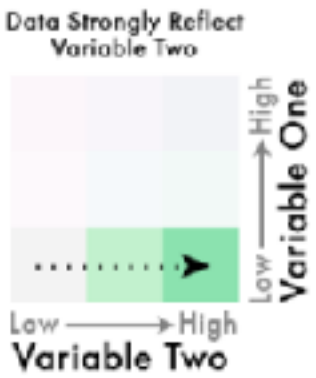


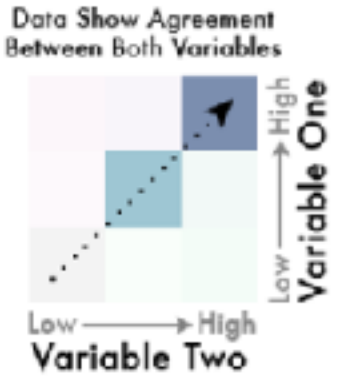
- Use unemployment *rates*, and not the total number of unemployed people

Bivariate

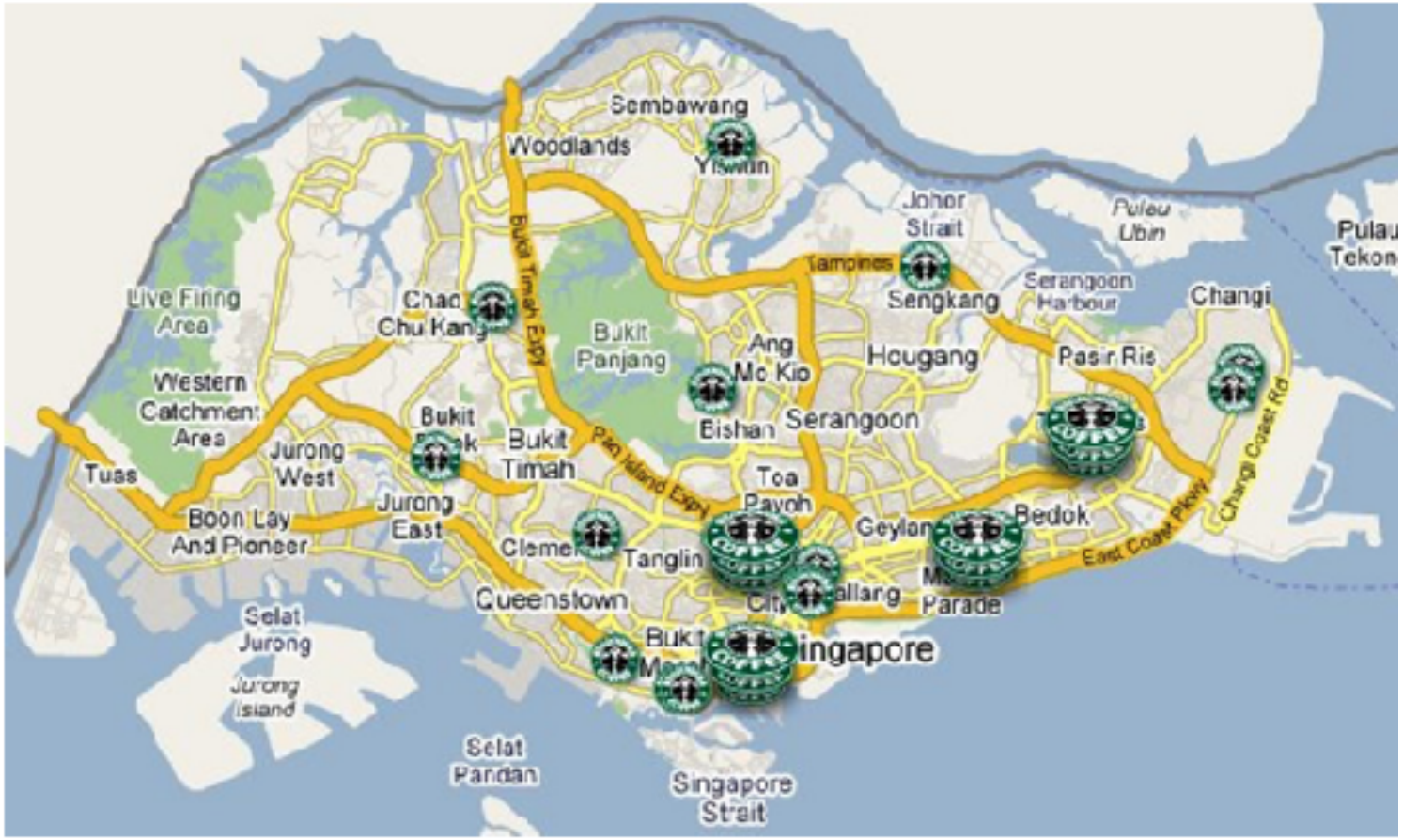

Low
High
% Insured
Low
High
Donations


Data Strongly Reflect
Variable One
Low
High
Variable Two
Low
High
Variable One


Data Strongly Reflect
Variable Two
Low
High
Variable Two
Low
High
Variable One


Data Show Agreement
Between Both Variables
Low
High
Variable Two
Low
High
Variable One

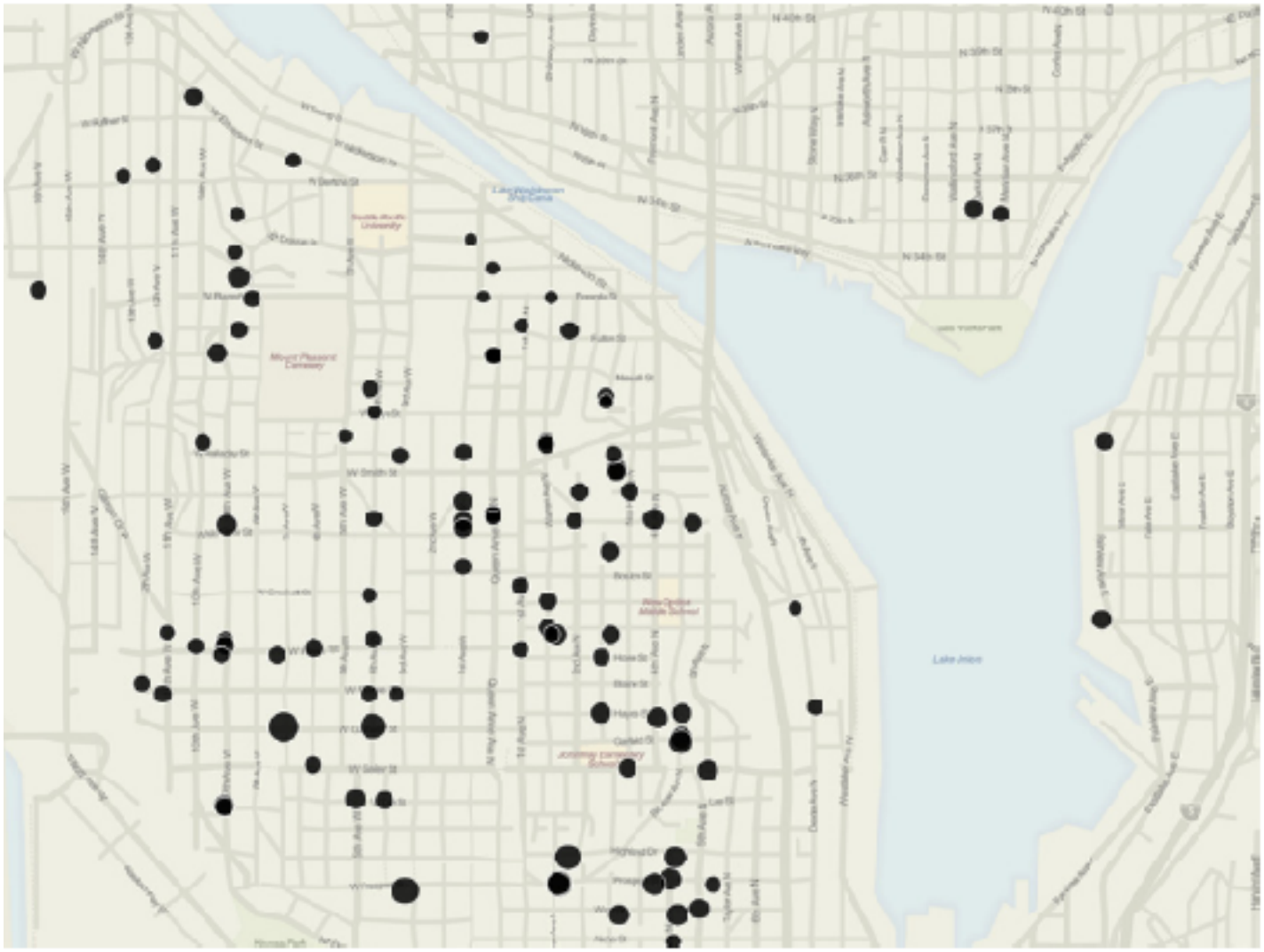
Featuring the Map, not the Data



Featuring the Data, not the Map

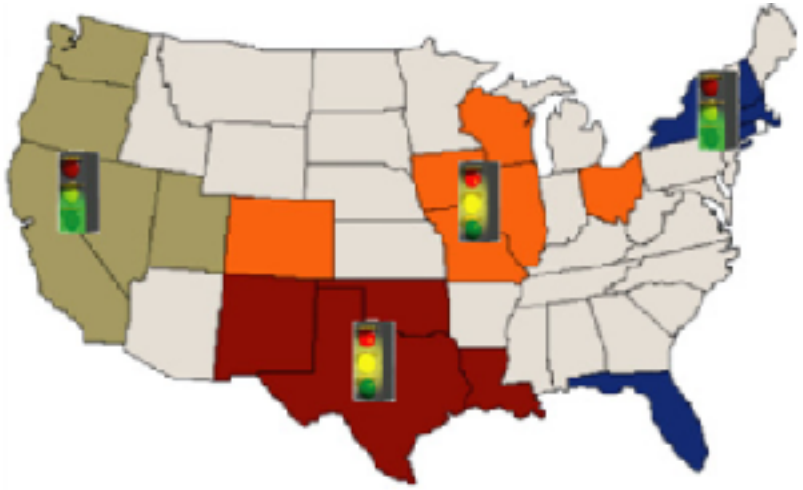
Bad Design Choices

Stand Out the Data



Good Uses of Geographical Data Displays

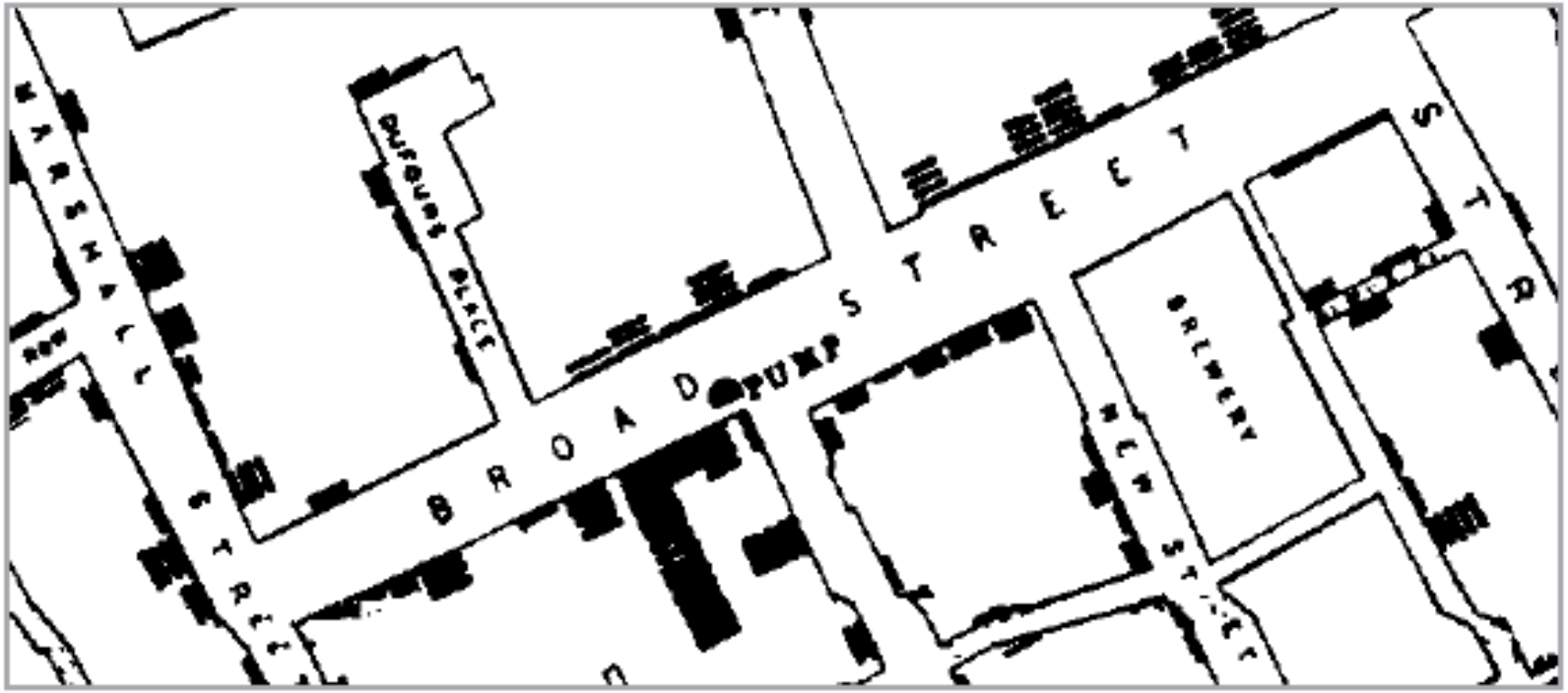
- Just because we can display data on a map doesn't mean we should.
- When to use map:
  - location is an important part of the meaning you're trying to discover or the story you're trying to tell



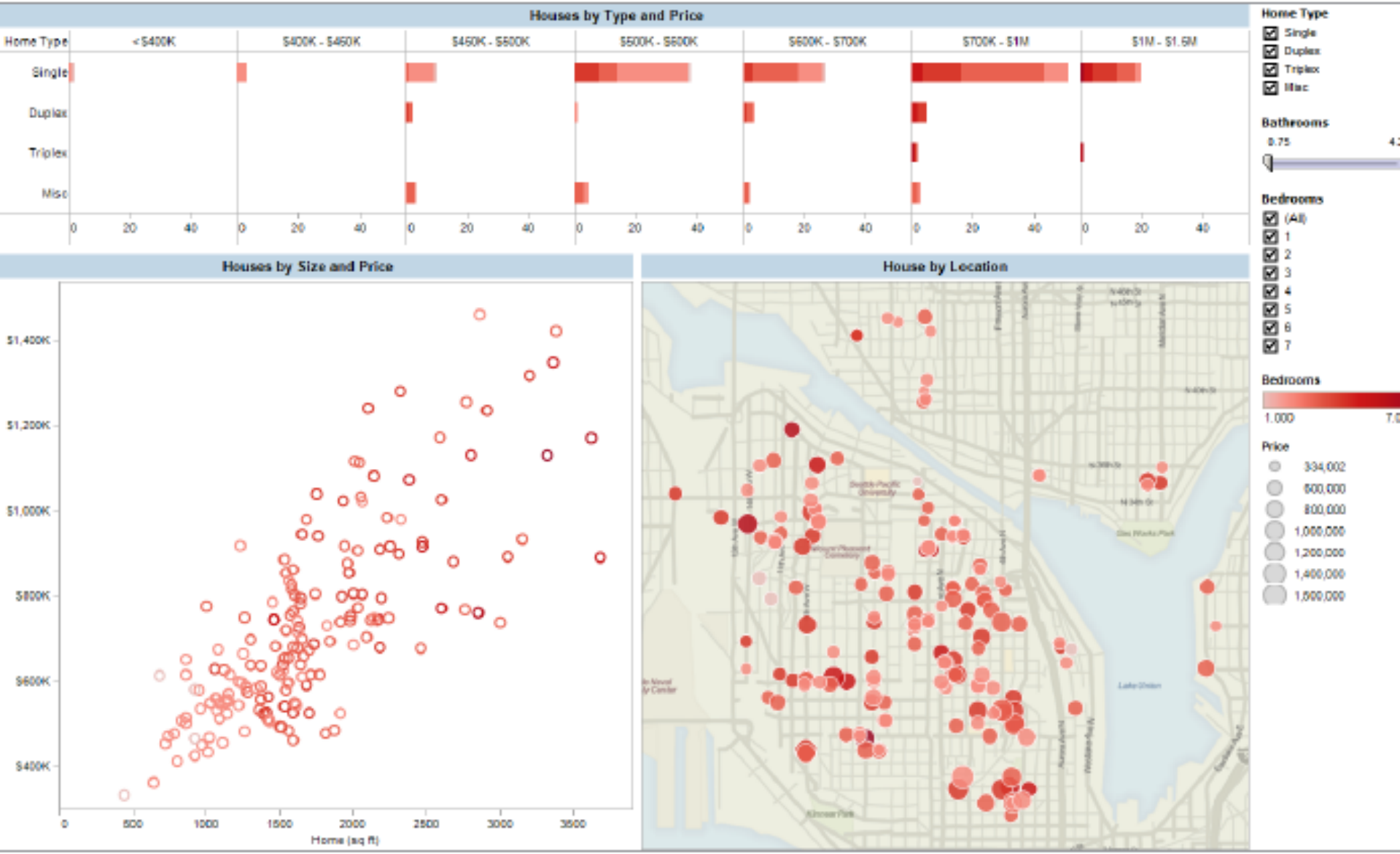
Less effective than a simple table or bar graph
Regional label : "West", "North Central", "South Central" and "East."

Good Uses of Geographical Data Displays

Good Uses of Geographical Data Displays



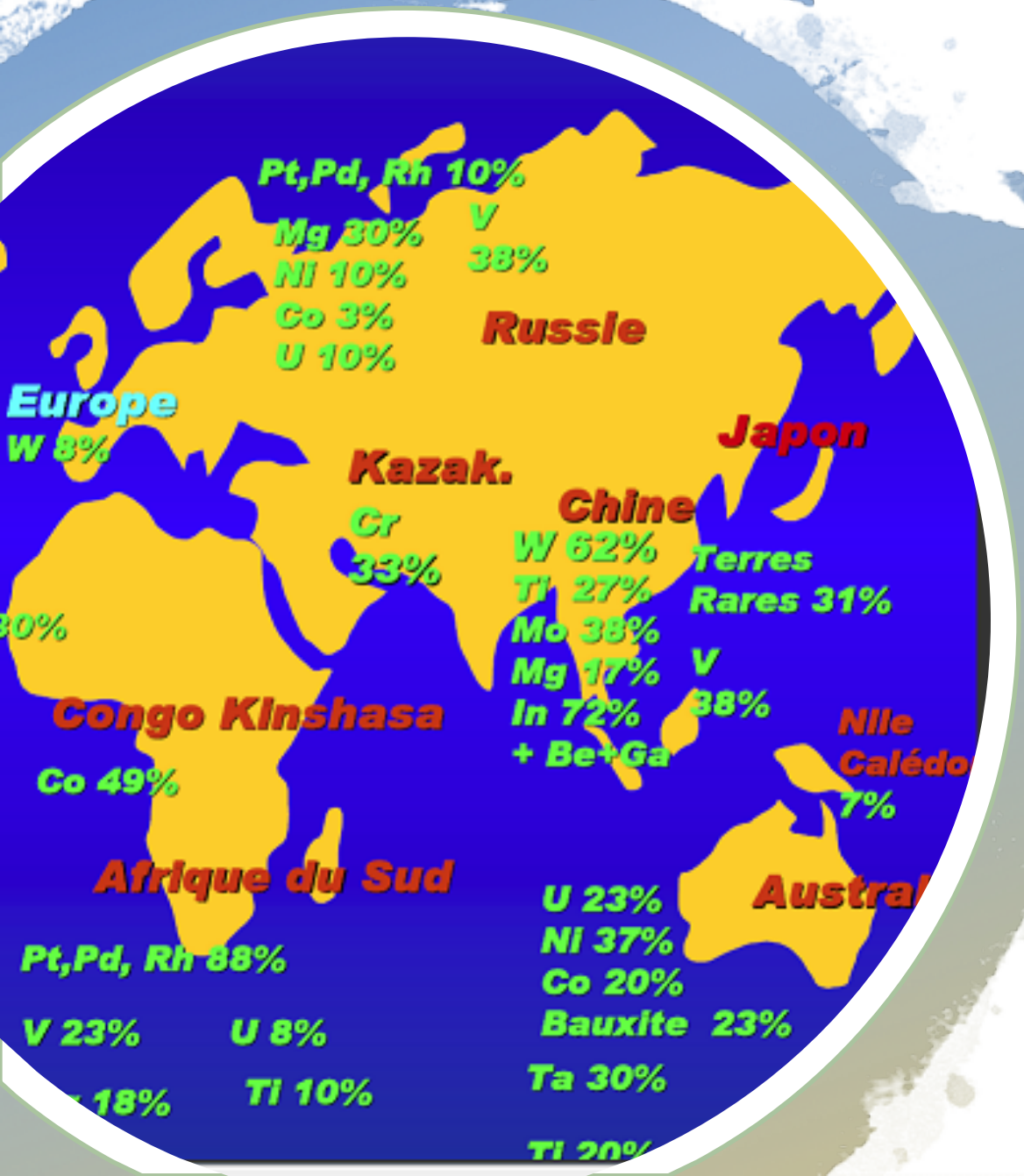
Collaboration between Maps and other Visual Displays

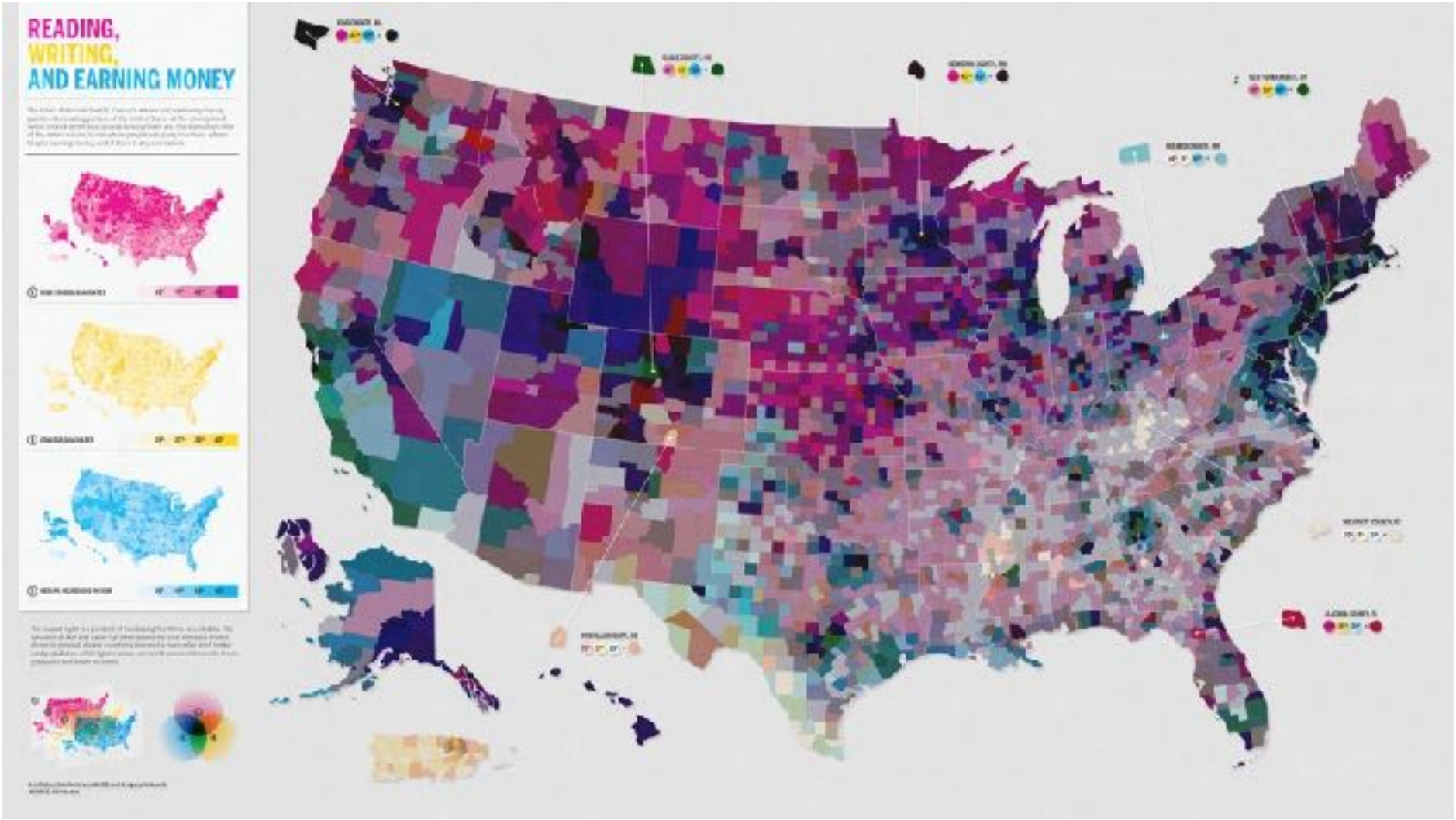


Example


Secaucus
North Bergen
West New York
Union City
Hoboken
Jersey City
New York
Hunters Point
Roosevelt Island
Cooperative Village
Middle Village
Midtown Expy
A
B
C
D
E
F
G
I
J

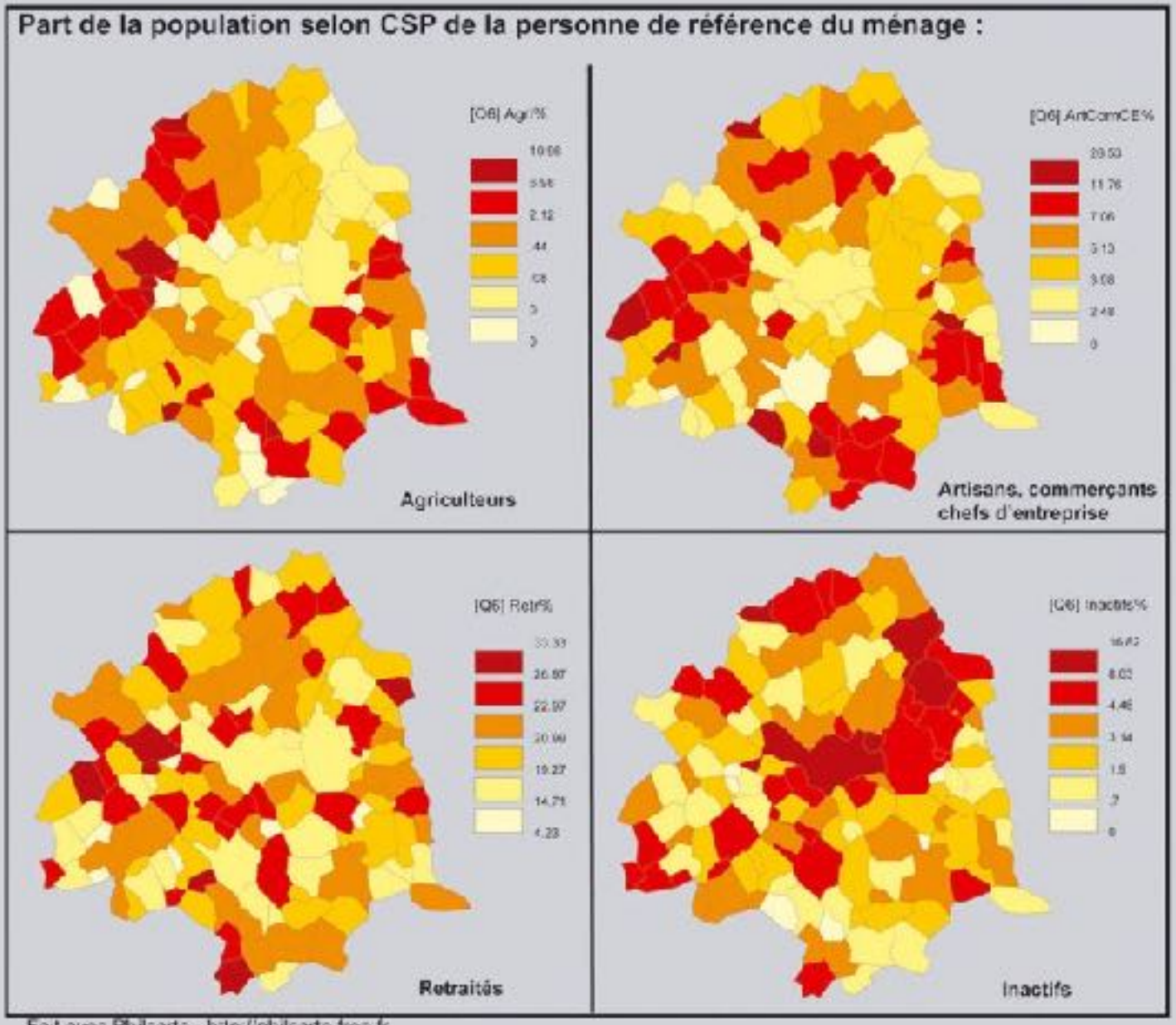
Example


Pt,Pd, Rh 10%
Mg 30%
Ni 10%
Co 3%
U 10%
V
38%
Russie
Europe
W 8%
Japon
Kazak.
Cr
33%
Chine
W 62%
Ti 27%
Mo 38%
Mg 17%
In 72%
+ Be+Ga
Terres
Rares 31%
V
38%
Congo Kinshasa
Co 49%
Nlle
Calédo
7%
Afrique du Sud
Pt,Pd, Rh 88%
V 23%
U 8%
Ti 10%
U 23%
Ni 37%
Co 20%
Bauxite 23%
Ta 30%


READING,
WRITING,
AND EARNING MONEY

Péri-urbanisation, déplacements domicile-travail et dépendance automobile dans l'arrondissement de Lille

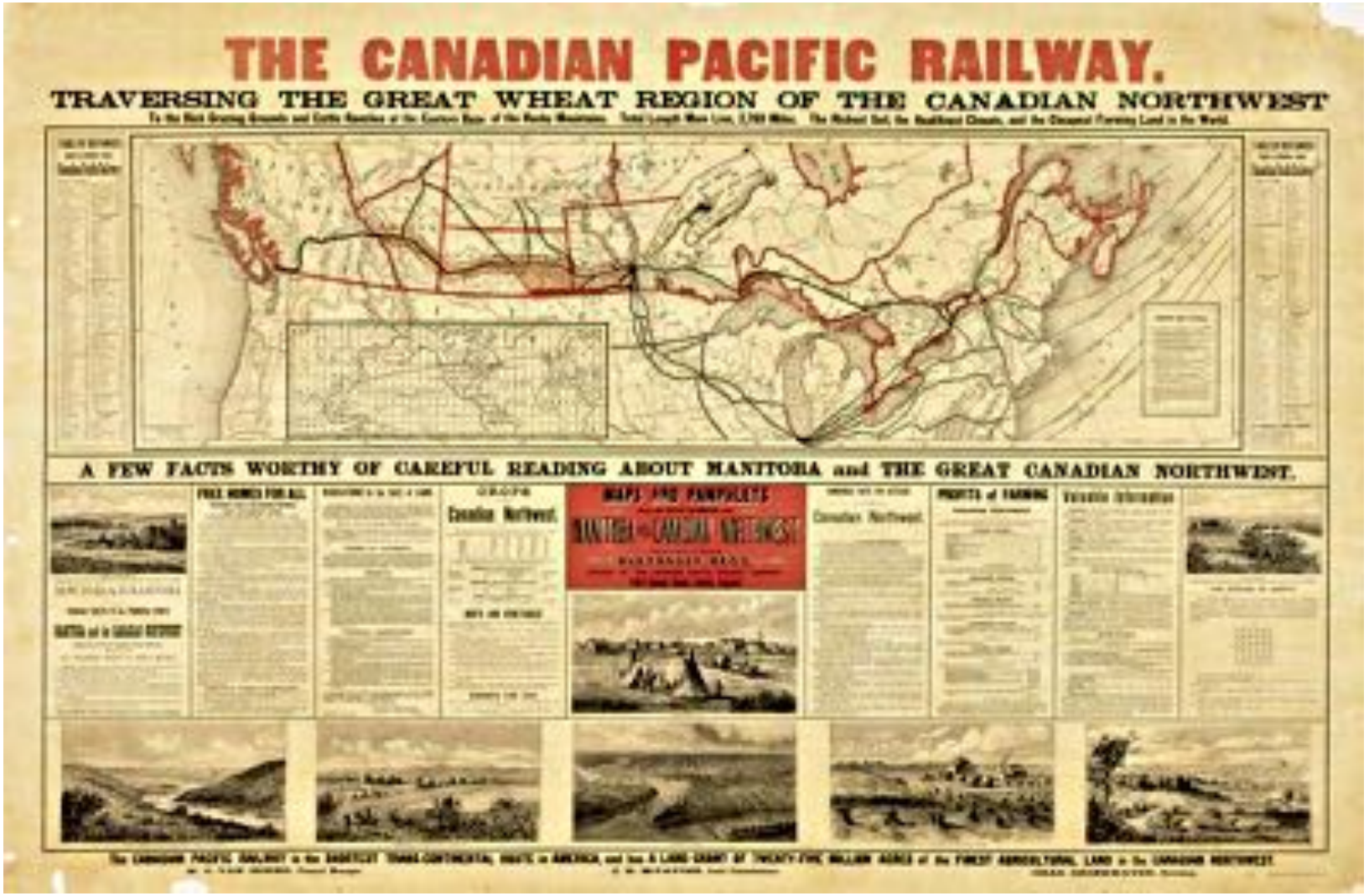
Part de la population selon CSP de la personne de référence du ménage :



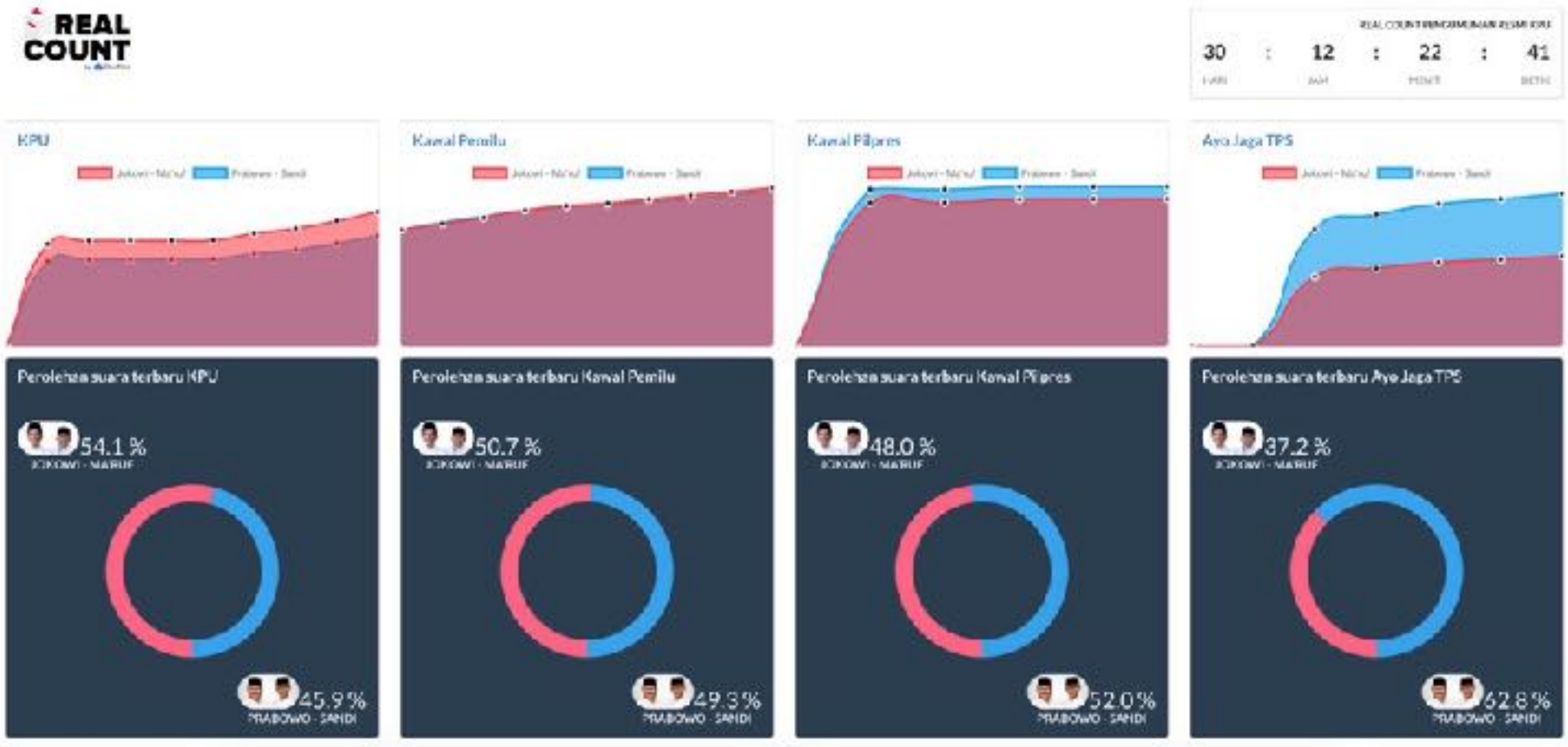
Fait avec Philcarto - http://philcarto.free.fr

Données INSEE recensement 2007

C.Korgomard 2010



In this 1881 map, the Canadian Pacific Railway tried to make western lands seem more inviting for settlement by suggesting that it was adequately serviced by rail. The map draws the viewer's attention to a completed transcontinental rail line, but the lack of station stops west of Swift Current suggests otherwise. The map also shows branch railways feeding into the main CPR line that were not opened until years later.
(Library and Archives Canada, NMC 11868)


REAL COUNT
30 : 12 : 22 : 41
KPU
Kawal Pemilu
Kawal Pilpres
Ayo Jaga TPS
Perolehan suara terbaru KPU
54.1 %
45.9 %
Perolehan suara terbaru Kawal Pemilu
50.7 %
49.3 %
Perolehan suara terbaru Kawal Pilpres
48.0 %
52.0 %
Perolehan suara terbaru Ayo Jaga TPS
37.2 %
62.8 %