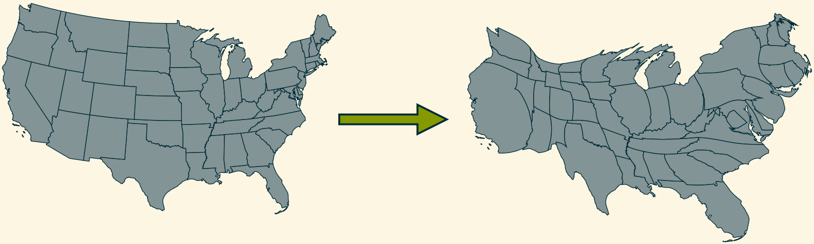
Applications of contiguous area cartograms

Michael T. Gastner

YaleNUSCollege

07 December 2018



Cartograms

Definition

"A cartogram is a map in which some thematic mapping variable—such as travel time, population or GNP—is substituted for land area or distance."

(Wikipedia, 2018)

Types of cartograms

Distance cartogram

Railway time-distances from Tokyo in 1965.

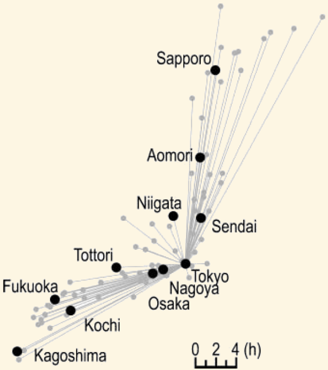
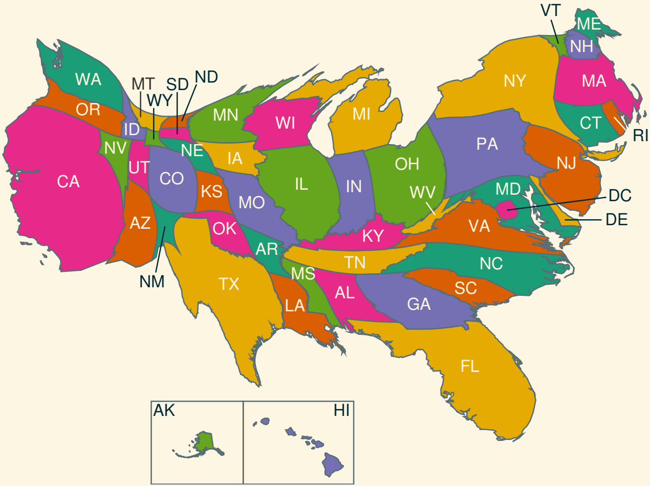


Image from E. Shimizu and R. InoueS, Int. J. Geogr. Inf. Sci. 23(11):1453-1470 (2009).

Types of cartograms

Area cartogram

Population sizes of states in the USA.



Objectives of this talk

Motivate, apply, explain

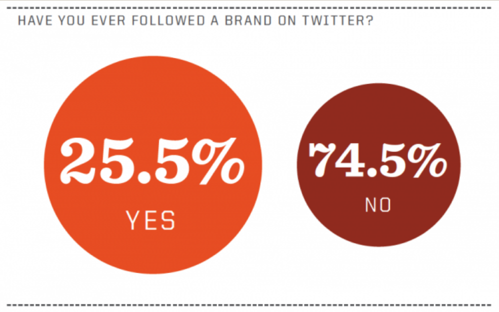
1. Motivate why area cartograms are useful
2. Categorize different types of area cartograms
3. Show two applications of *contiguous* cartograms:
   - Mortality statistics
   - Facility location
4. Explain how to generate contiguous cartograms

The area principle

A guideline for displaying statistical data

Each part of a diagram should have an area in proportion to the number it represents.[1]

Here is a particularly bad example.[2]

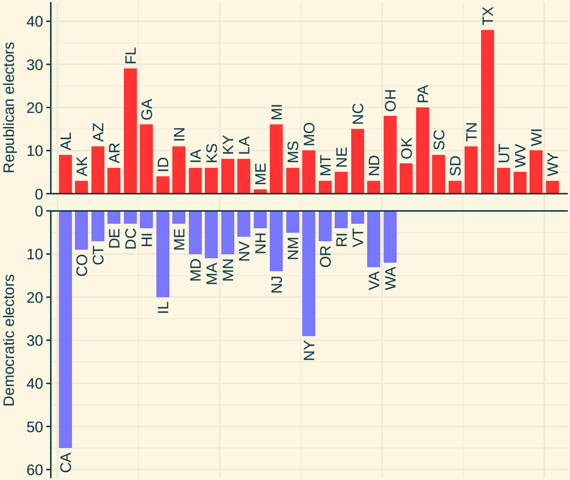


[1] R. D. de Veaux et al., *Stats: Data and Models.* Pearson, 4th ed. (2016).
[2] https://martech.zone/bad-infographics/

Motivating example

2016 US presidential election

Bar charts satisfy the area principle.

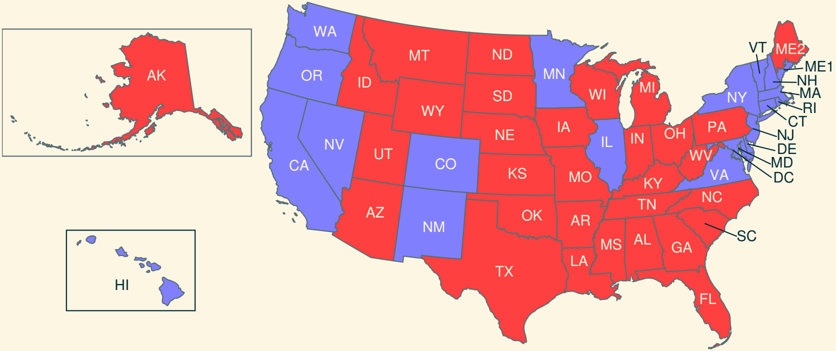


Why are bar charts not perfect for geographic data?

They don't show how regions fit together

Neighbouring bars $\neq$ geographic neighbours

Common alternative: (nearly) equal-area maps



Why are equal-area maps not perfect for statistical data?

They violate the area principle

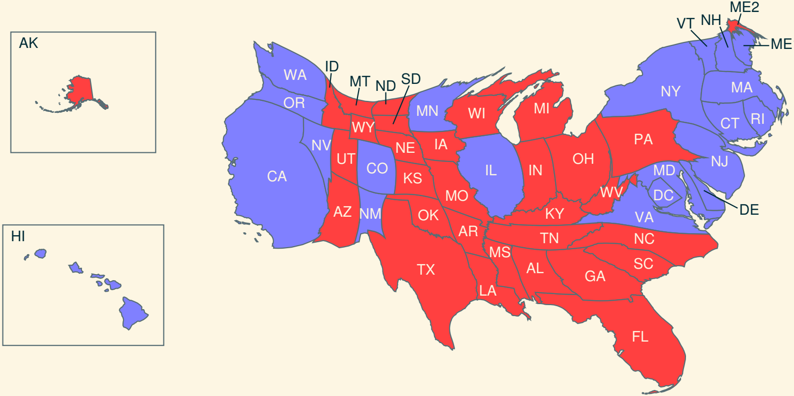
- Montana covers more than 2000 times the area of Washington DC.
- But they have the same number of electors.

A cartogram rescales the areas to be, for example, proportional to the number of electors.

▸ Example

Advantage of cartogram

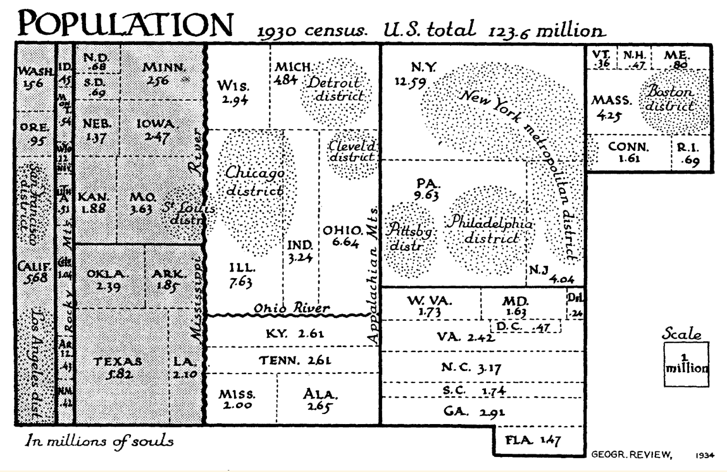
Area principle *and* contiguity



- Similar to a bar chart, we can judge the importance of each region.
- We still see how regions fit together.

Historical example

Rectangular cartogram

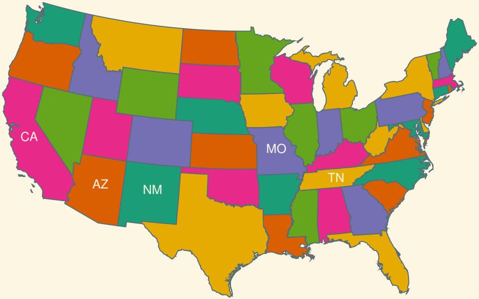


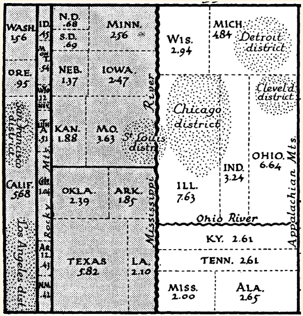
E. Raisz, Geogr. Rev. 24(2):292 (1934)

Problems with Raisz's cartogram

Shapes and topology

- Shapes bear little resemblance with those on conventional maps.
- Topology not preserved:
  - CA and NM should not be neighbours.
  - But MO and TN should share common border.





How to improve shapes and topology?

A solution strategy

1. Divide administrative regions into smaller building blocks (e.g. squares or hexagons).
2. Move blocks into positions that resemble conventional maps.

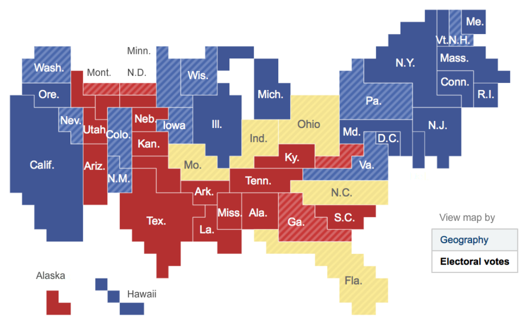
$\Longrightarrow$ "Mosaic" cartograms

▸ Examples

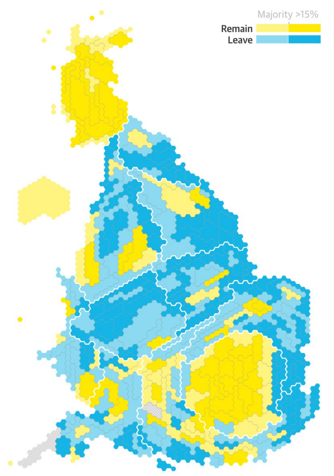
Mosaic cartograms

Examples

New York Times: squares.



Guardian: hexagons



Circular cartograms

Example

Cartogram algorithm by Daniel Dorling (1996):

- each administrative region is represented by a circle,
- circles touch, but don't overlap,
- centre of circle ≈ centroid on conventional map.

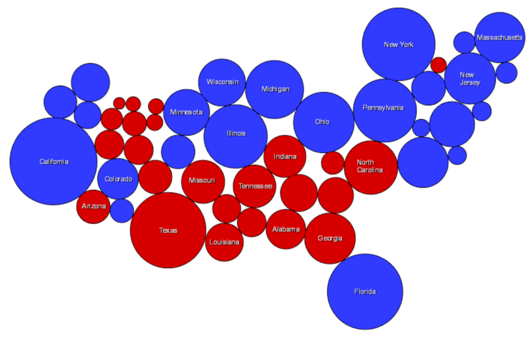


Image by Kenneth Field, politico.com.

Noncontiguous cartograms

Remove topological constraints

Dorling's circular cartograms belong to the class of *noncontiguous* cartograms: neighbours on the cartogram are separated.

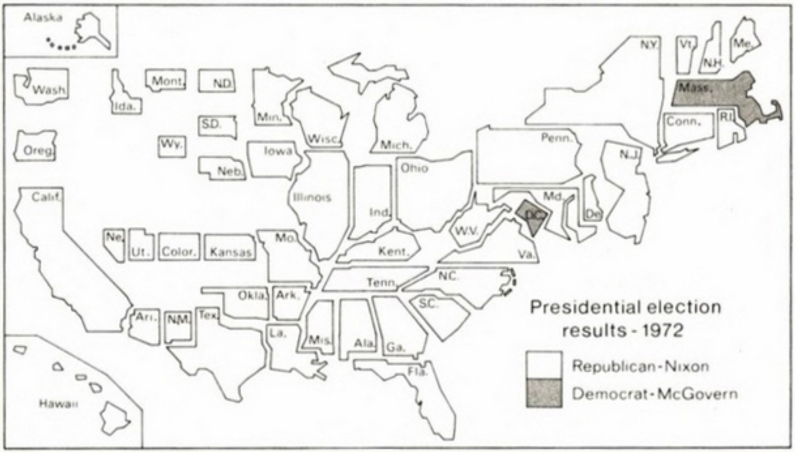
Spinning this idea further, we don't need to use circles at all.

Instead we can perfectly preserve the shapes.

Example

Noncontiguous cartogram with perfect shape preservation

1972 US electoral college



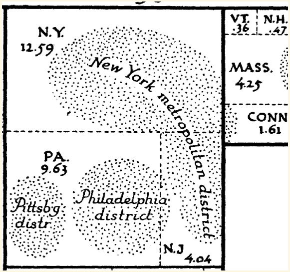
G. Rowley, Geogr. Mag. 45:344 (1973)

Disadvantage of *noncontiguous* cartograms

They may emphasize irrelevant boundaries

Noncontiguous cartograms obfuscate spatial patterns distributed across several regions.

For example, New York City's zone of influence crosses the NY/NJ state border. It would be artificial to insert a gap.

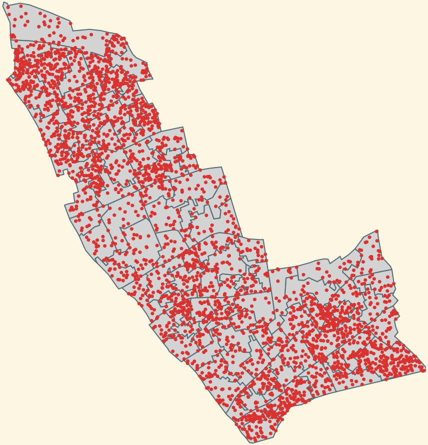


Let's look at an application where we need a *contiguous* cartogram to perform geospatial statistics.

Application 1

Death cases in Kensington & Chelsea (London), 2011-2014

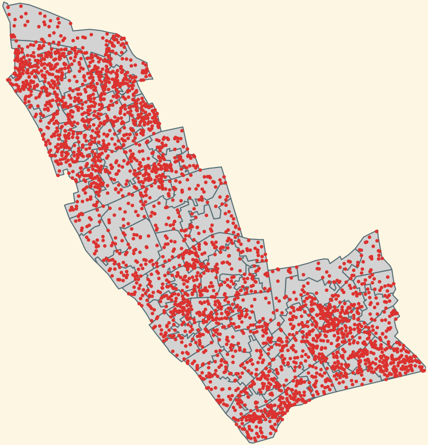
Equal-area scatter map



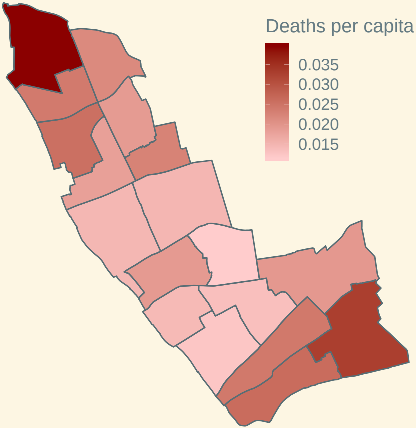
Application 1

Death cases in Kensington & Chelsea (London), 2011-2014

Equal-area scatter map



Choropleth map: rates by ward



Problems with equal-area and choropleth maps

Location information vs. per capita mortality

- Scatter map: can't tell per capita mortality.
- Choropleth map: don't know where the cases actually occur.

Alternative?

Cartogram: scale areas to be proportional to their population.

- Points show the position of each individual case.
- Point density indicates per capita mortality.

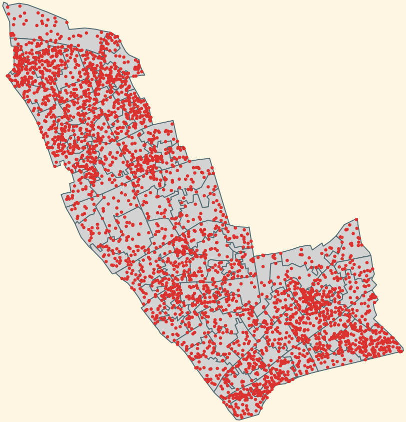
Cartogram should be *contiguous* because boundaries are purely for census purposes.

Example

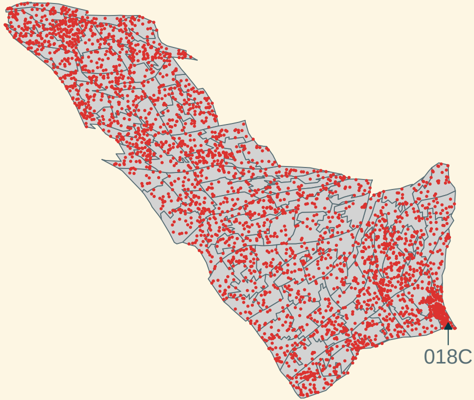
Same death cases

Different map projections

Equal-area projection



Cartogram



Why is there such a high per-capita density in region 018C? Answer

Why are there so many deaths in 018C?

Because of a nursing home?

The technical name for every region on the previous slides is "Lower Layer Super Output Area" (LSOA).

Does a nursing home (St. Wilfrid's) in LSOA 018C explain the high number of deaths?

Age-adjusted mortality

Divide population by age and gender

- Total population is too crude a predictor.
- Instead make a cartogram based on the expected number of death cases given the age distribution in each LSOA.
- Divide the population into age groups (0, 1–4, 5–9, ..., 85–89, $>$90 years), with each age group divided into men and women.

Expected number of deaths

Weighted by age and gender

Define the following quantities.

- p_{ij}: the population that lives in LSOA i and belongs, because of its gender and age, to the demographic group j
- $p_j = \sum_i p_{ij}$: London-wide population size of group j
- d_j: number of London-wide deaths in group j
- $m_j = d_j / p_j$: London-wide mortality in group j

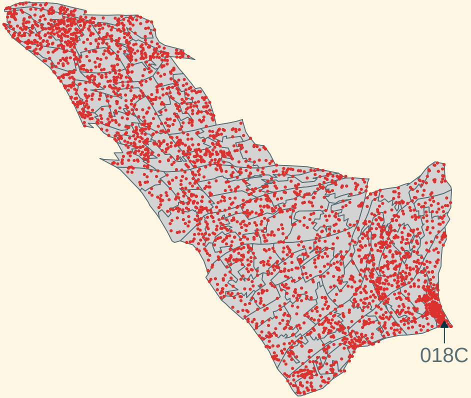
The expected number of deaths in the i-th LSOA is

$$e_i = \sum_j p_{ij} m_j.$$

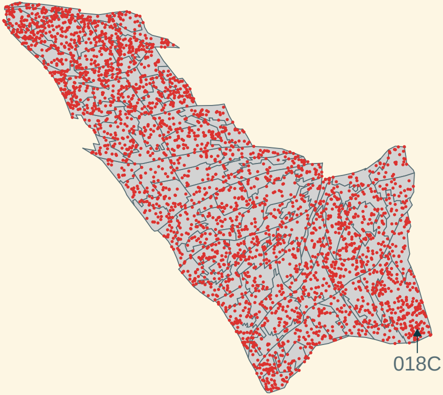
Cartograms

Total vs. age-adjusted population

area = population



area = expected deaths

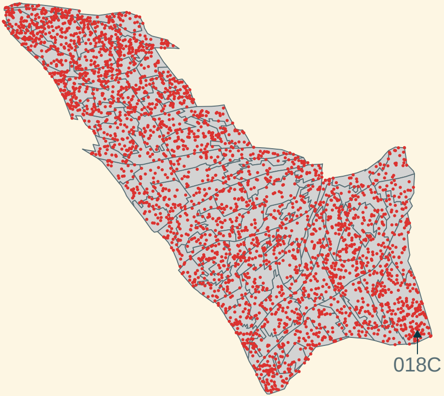


M. T. Gastner et al., Proc. Nat. Acad. Sci. USA 115(10):E2156–E2164 (2018)

Kernel density estimate

Is there a spatial trend?

area = expected deaths

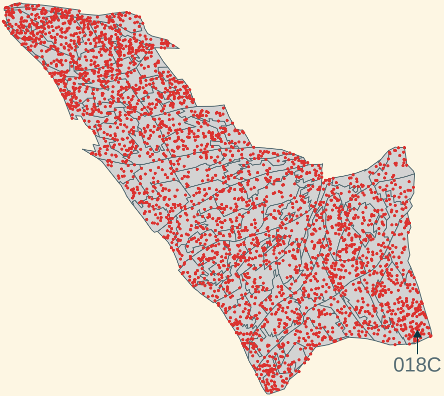


M. T. Gastner et al., Proc. Nat. Acad. Sci. USA 115(10):E2156–E2164 (2018)

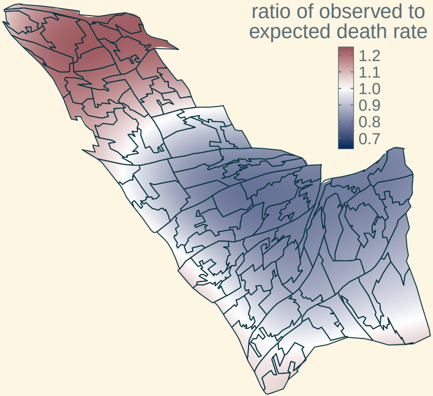
Kernel density estimate

Is there a spatial trend?

area = expected deaths



Kernel density estimation

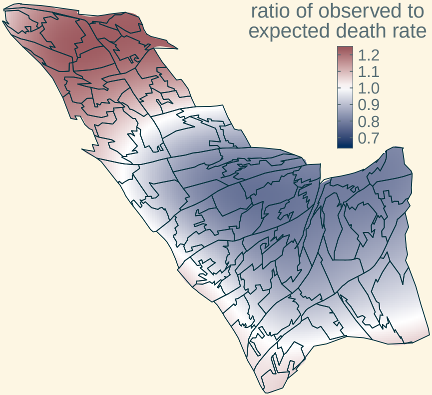


M. T. Gastner et al., Proc. Nat. Acad. Sci. USA 115(10):E2156–E2164 (2018)

The Economist (24 June 2017):

"Kensington and Chelsea: a wealthy but deeply divided borough"

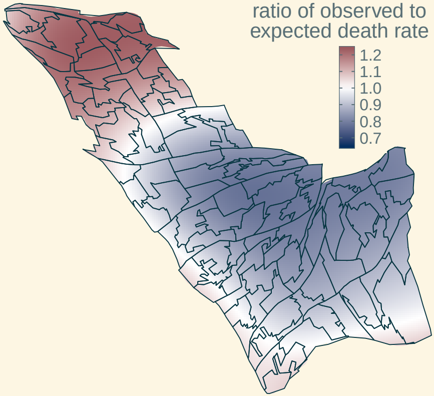
age-adjusted mortality

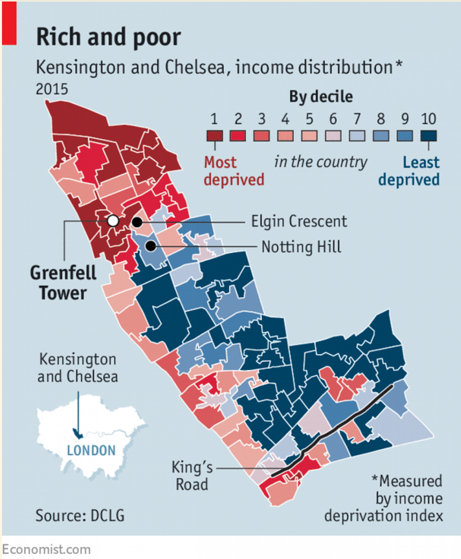


The Economist (24 June 2017):

"Kensington and Chelsea: a wealthy but deeply divided borough"

age-adjusted mortality





Application 2
Facility location

Task:

"Locate churches on a plain disuniformally filled with parishioners in such a manner as to minimize the total number of steps needed by all churchmen to attend services at the nearest church."

William Bunge, "Patterns of location" (1964)[3]

If there are p churches, operations researchers call this task the p-median problem.

[3] Michigan Inter-University Community of Mathematical Geographers, Discussion Paper 3, University of Michigan, Ann Arbor.

Illustration of facility location

Image from Bunge's "Patterns of location"

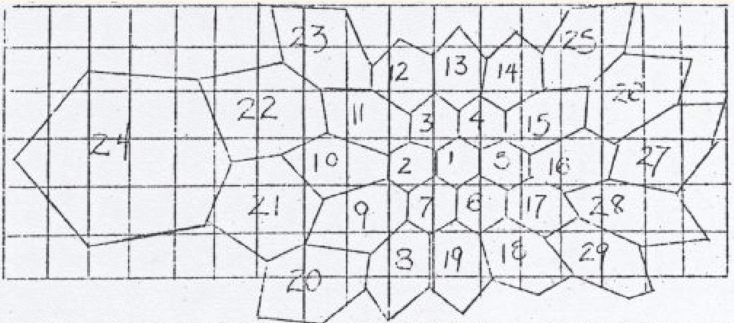


Figure 35. An approximation of a Christaller solution applied to an area of disuniform rural population. (Some errors)

The lines would divide our map in such a way as to minimize the total amount of movement from all the points of the density surface to the nearest central place

Illustration of facility location

Image from Bunge's "Patterns of location"

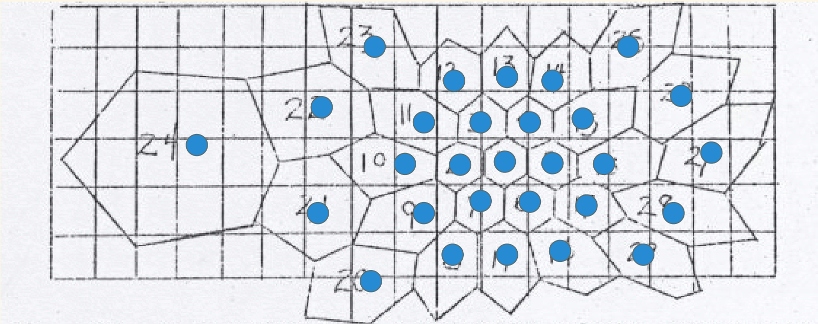


Figure 35. An approximation of a Christaller solution applied to an area of disuniform rural population. (Some errors)

The lines would divide our map in such a way as to minimize the total amount of movement from all the points of the density surface to the nearest central place

Are facilities uniformly distributed on a cartogram?

Another image from Bunge's "Patterns of location"

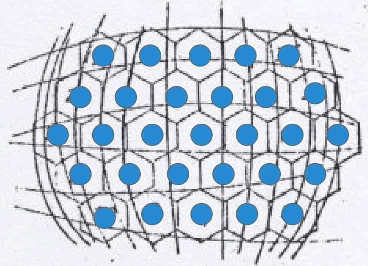


Figure. 36 Map transformed into uniform density.

Bunge's hypothesis

Equal number of parishioners for each church

Let's introduce notation.

- $\rho(\mathbf{r})$: population density at geographic position $\mathbf{r}$
- $s(\mathbf{r})$: "service area" of the parish at $\mathbf{r}$
- Bunge's hypothesis:

$$s(\mathbf{r}) \cdot \rho(\mathbf{r}) \approx \text{const.}$$

$\Longrightarrow$ On a cartogram that assigns equal area to every parishioner, all parish areas look equal.

This argument *isn't* correct.

How to rescue Bunge's idea

Service area increases as $\rho(\mathbf{r})^{-2/3}$ instead of $\rho(\mathbf{r})^{-1}$

One can show that

$$s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} \approx \text{const.}$$

This result has been derived, for example, by Stephan[4] as optimal division of a nation into administrative regions.

- Most facilities are in densely populated areas.
- But sparsely populated areas receive more facilities per capita.

▸ Derivation

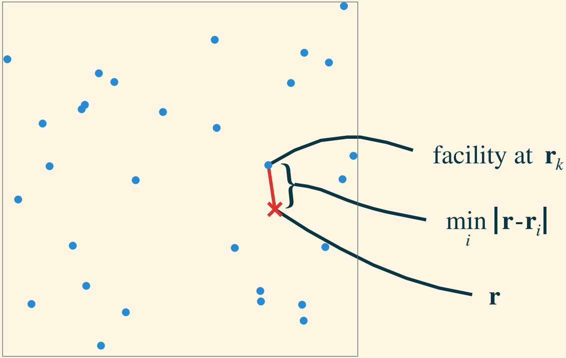
[4] G. E. Stephan, Science 196(4289):523–524 (1977)

Derivation of $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = \text{const.}$

Notation

Facility positions: $\mathbf{r}_1, \ldots, \mathbf{r}_p$

Distance to nearest facility from arbitrary point $\mathbf{r}$: $\min_i |\mathbf{r} - \mathbf{r}_i|$

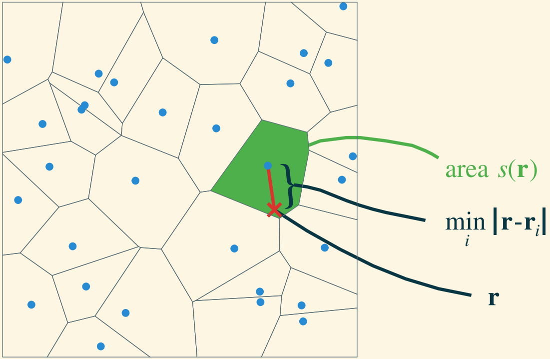


Derivation of $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = \text{const.}$

Distances grow proportional to $\sqrt{s(\mathbf{r})}$

The average distance between a point and its nearest facility scales as $\sqrt{s(\mathbf{r})}$,

$$\min_i |\mathbf{r} - \mathbf{r}_i| \approx g\sqrt{s(\mathbf{r})}\ , \qquad g \approx \text{const.}$$



Derivation of $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = $ const.

Optimization task

Suppose the country occupies the area A.

Task: minimize total distance travelled by the population in A,

$$\int_A \rho(\mathbf{r}) \min_i |\mathbf{r} - \mathbf{r}_i| \, d^2r \approx g \int_A \rho(\mathbf{r}) \sqrt{s(\mathbf{r})} \, d^2r \ .$$

Constraint: there are exactly p facilities,

$$\int_A \frac{1}{s(\mathbf{r})} \, d^2r = p.$$

Derivation of $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = $ const.

Constrained optimization

Introduce Lagrange multiplier α and take functional derivative,

$$\frac{\delta}{\delta s(\mathbf{r})}\left[g\int_A \rho(\mathbf{r})\sqrt{s(\mathbf{r})}\,d^2r - \alpha\left(p - \int_A \frac{1}{s(\mathbf{r})}\,d^2r\right)\right] = 0,$$

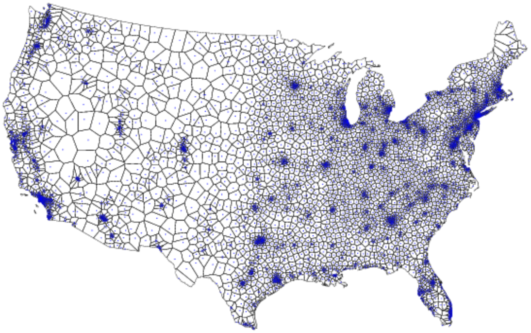
$$\Longrightarrow \qquad s(\mathbf{r}) = \left(\frac{2\alpha}{g\rho(\mathbf{r})}\right)^{2/3} = \text{const.} \cdot \rho(\mathbf{r})^{-2/3} \ .$$

Visualizing the relation $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = \text{const.}$

Cartograms

(1) From US census, obtain block-level population distribution.

(2) Use numerical heuristic to place $p = 5000$ facilities such that the p-median problem is approximately solved.



Visualizing the relation $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = \text{const.}$

Cartograms

(1) From US census, obtain block-level population distribution.

(2) Use numerical heuristic to place $p = 5000$ facilities such that the p-median problem is approximately solved.

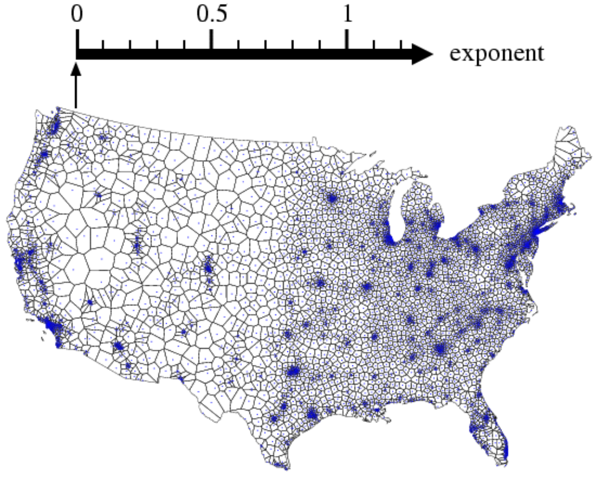
(3) Make cartograms that equalize ρ^x where x is a tunable exponent.

- $x = 0$: equal-area map
- $x = 1$: population cartogram

Theoretical prediction: facility distribution appears most uniform for $x = 2/3$.

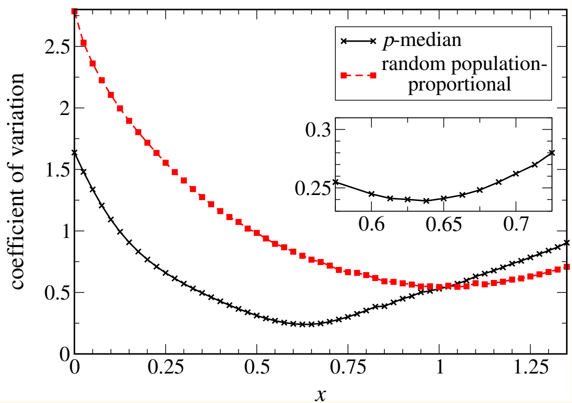
Visualizing the relation $s(\mathbf{r}) \cdot \rho(\mathbf{r})^{2/3} = \text{const.}$

Animation



Which exponent has most uniform facility distribution?

Cartogram as a diagnostic



Coefficient of variation (i.e. the ratio of standard deviation to mean) for service areas as they appear on a cartogram vs. the exponent x of the underlying density ρ^x.

M. T. Gastner and M. E. J. Newman, Phys. Rev. E 74(1):016117 (2006)

How to make contiguous cartograms

Methods based on physical analogies

- Rubber sheet methods: relaxation of forces acting on an elastic substrate[5]
- Electrostatics: repulsion of charged particles[6]
- Diffusion: particles undergoing Brownian motion[7]
- Fast flow-based method: similar to diffusion, but equations can be numerically solved more quickly[8]

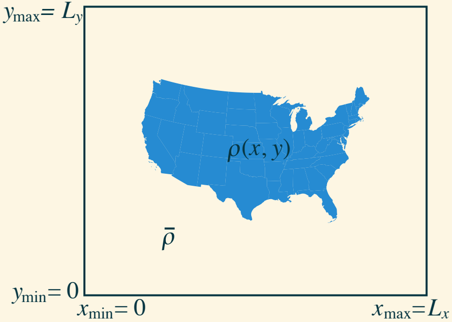
---

[5] S. Sun, Int. J. Geogr. Inf. Sci. 27(3):567–593 (2013)
[6] S. M. Gusein-Zade and V. S. Tikunov, Cartogr. Geogr. Inform. 20(3):167 (1993)
[7] M. T. Gastner and M. E. J. Newman, Proc. Nat. Acad. Sci. USA 101(20):7499 (2004)
[8] M. T. Gastner et al., Proc. Nat. Acad. Sci. USA 115(10):E2156–E2164 (2018)

Flow-based cartogram

Intuition

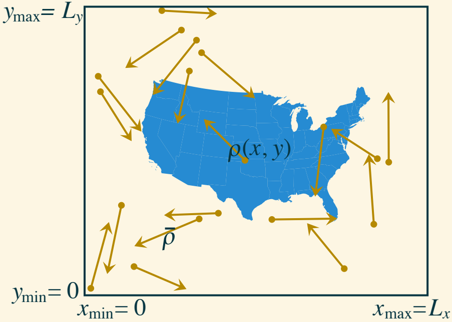
- Think of the map as a rectangular Petri dish covered with a thin layer of water.
- Model the population density $\rho(\mathbf{r})$ by injecting particles into the water.
- Let the particles spread across the entire Petri dish.



Flow-based cartogram

Intuition

- Think of the map as a rectangular Petri dish covered with a thin layer of water.
- Model the population density $\rho(\mathbf{r})$ by injecting particles into the water.
- Let the particles spread across the entire Petri dish.



Flow dynamics

Density and velocity as functions of time

"Fast flow-based method":

$$\rho(x,y,t) = \begin{cases} (1-t)\,\rho(x,y,0) + t\bar{\rho} & \text{if } 0 \le t \le 1, \\ \bar{\rho} & \text{if } t > 1. \end{cases}$$

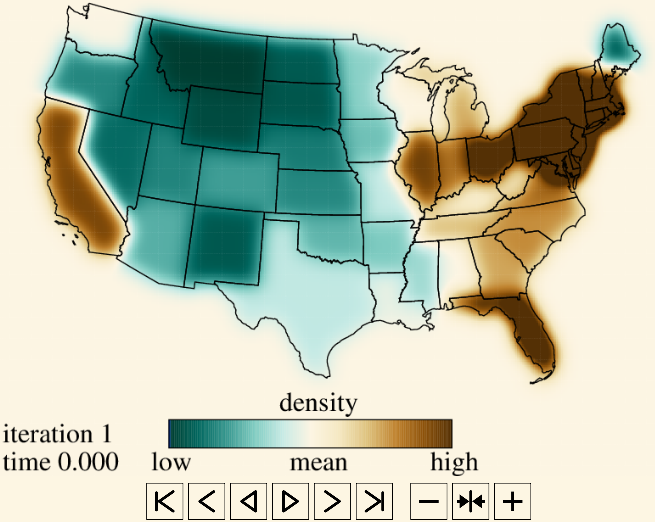
Move the boundaries on the map such that each region always contains the same number of particles.

For details of the implementation, see:

M. T. Gastner, V. Seguy, P. More
Fast flow-based algorithm for creating density-equalizing map
Proc. Nat. Acad. Sci. USA 115(10):E2156–E2164 (2018)

Fast flow-based cartogram: animation

US Electoral College Cartogram



go-cart.io

User-friendly web application for cartograms

The C code for the fast flow-based method is available at:
`https://github.com/Flow-Based-Cartograms/go_cart`

We are working on a user-friendly web application
`https://go-cart.io/`.

Our objectives:

- Neither maths nor programming required.
- No need to find geospatial vector data.
- Calculation finishes within a few seconds.
- Users can explore the cartogram interactively.

Conclusion

Contiguous area cartograms: an underestimated diagnostic tool

- Cartograms satisfy the area principle of statistical data visualization.
- *Contiguous* cartograms also correctly display spatial proximity—unlike noncontiguous cartograms, bar or pie charts.
- Applied to mortality data, contiguous cartograms visualize the quality of statistical models (e.g. effect of age, gender).
- Contiguous cartograms can highlight spatial trends in facility location problems.
- We are building the web application `https://go-cart.io/` to simplify the creation and interpretation of contiguous cartograms.