Evolution Strategies for Optimizing Rectangular Cartograms

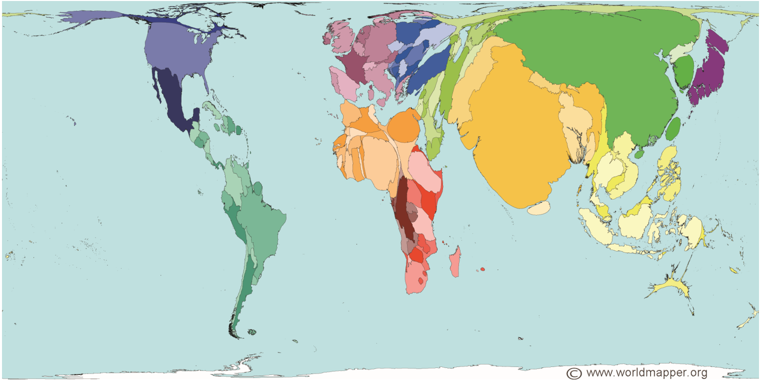
Kevin Buchin[1], Bettina Speckmann[1], and Sander Verdonschot[2]

[1] TU Eindhoven, [2] University of Ottawa

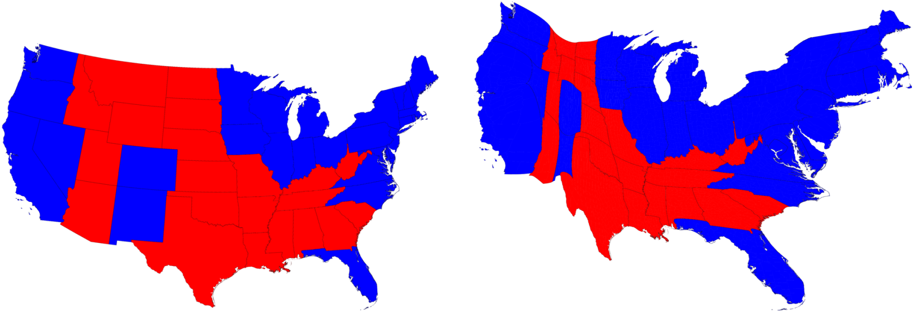
January 15, 2016

Cartograms

- Map with administrative boundaries
- Visualizes a quantitative value per region (e.g. population, GDP)
- Regions are deformed to make their area proportional to this variable



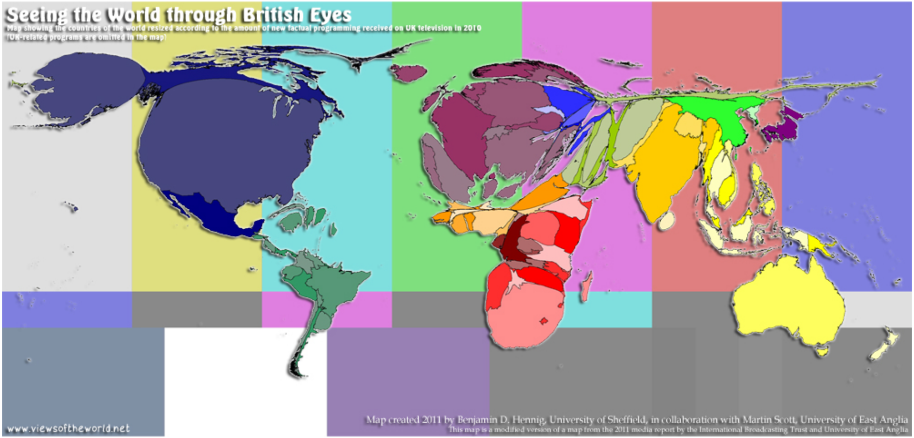
Cartograms



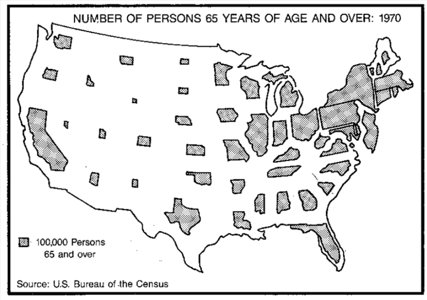
Types of Cartograms

- Contiguous area cartogram
- Non-contiguous area cartogram
- Dorling / Demers cartogram
- Rectangular cartogram

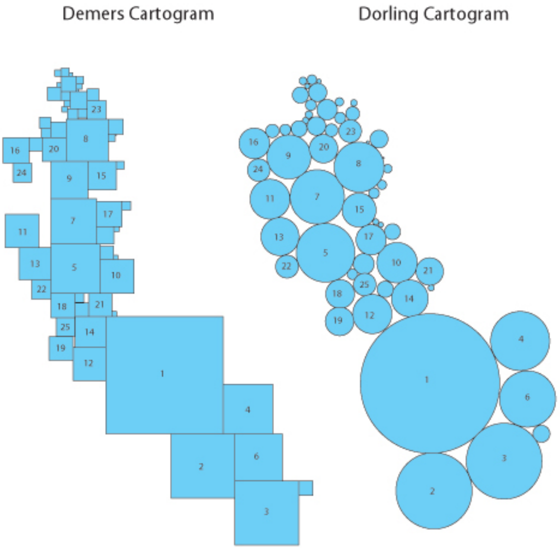
Contiguous area cartogram



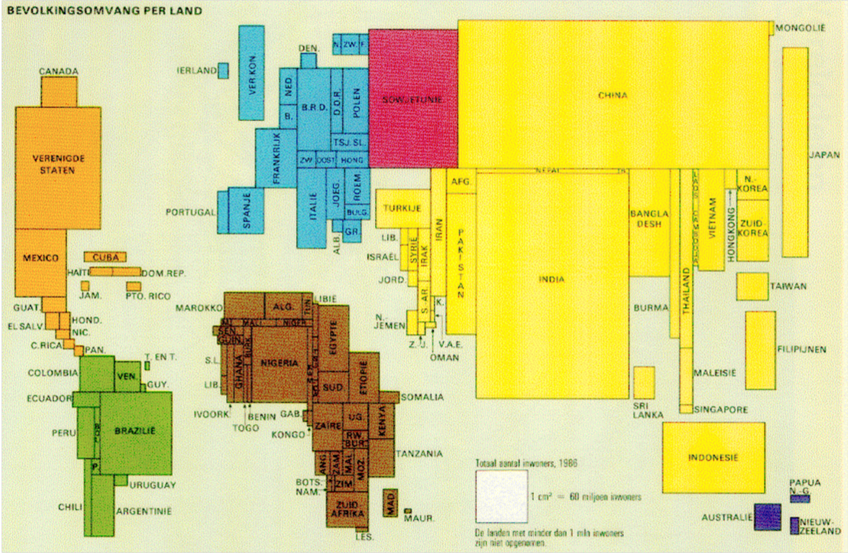
Non-contiguous area cartogram



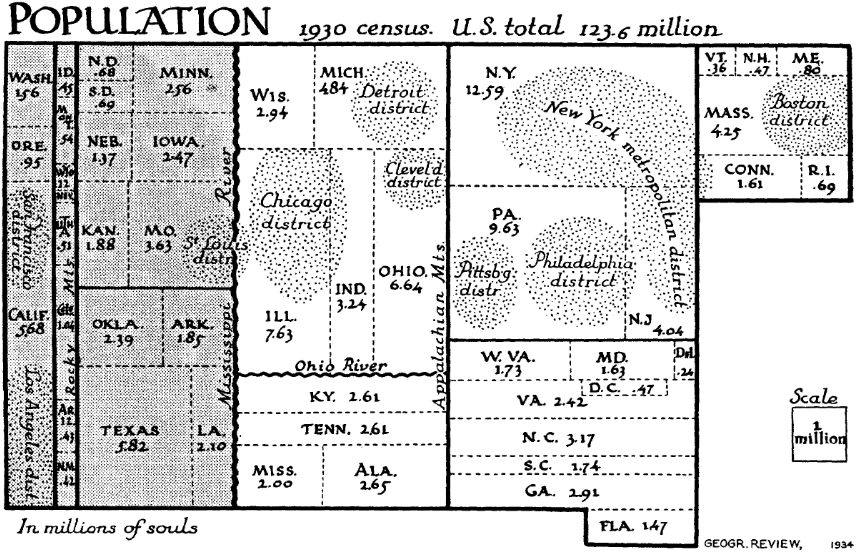
Dorling / Demers cartogram



Rectangular Cartogram

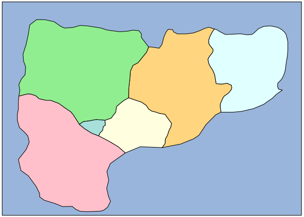


Rectangular Cartogram



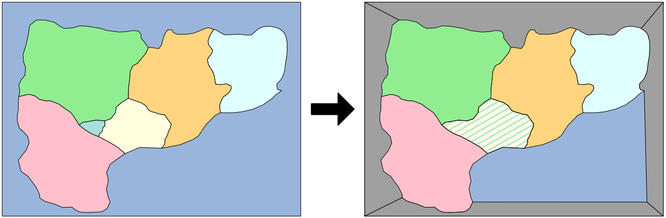
Workflow

- Given a geometric map and a corresponding variable



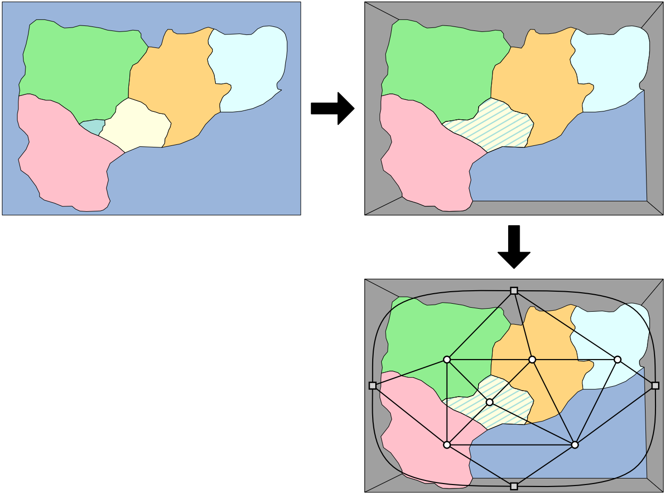
Workflow

- Convert it to the correct format



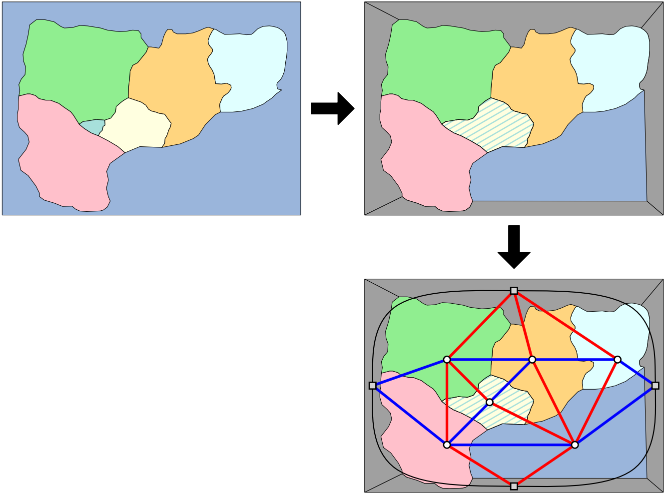
Workflow

- Assign a direction to each adjacency



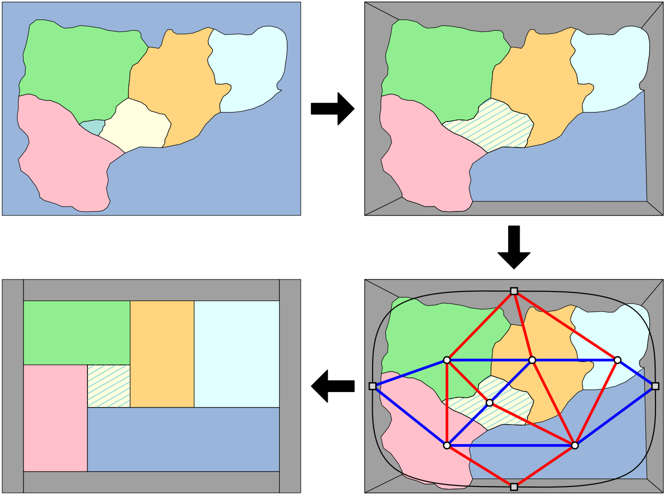
Workflow

- Assign a direction to each adjacency



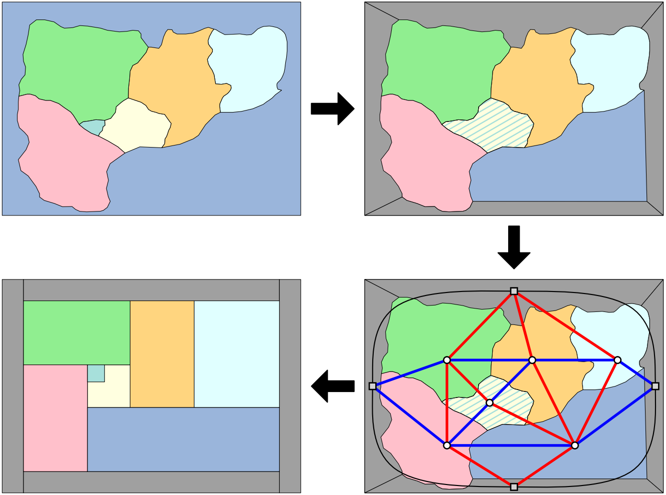
Workflow

- Build the cartogram [Speckmann, van Kreveld, and Florisson, 2006]



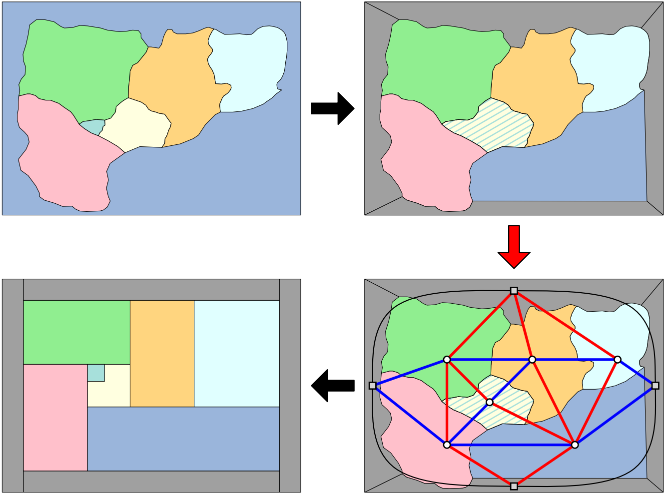
Workflow

- Build the cartogram [Speckmann, van Kreveld, and Florisson, 2006]



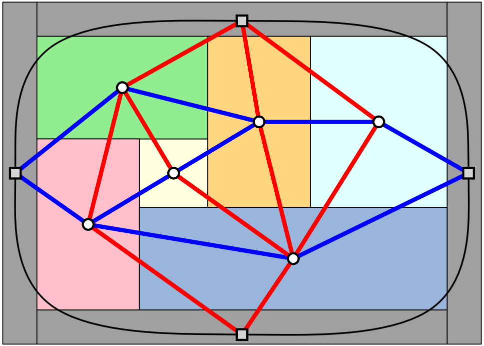
Workflow

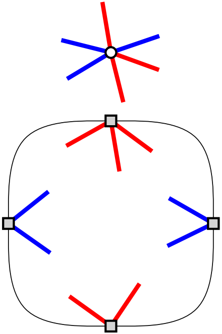
- Our focus is on the direction of the adjacencies



Regular Edge Labelings

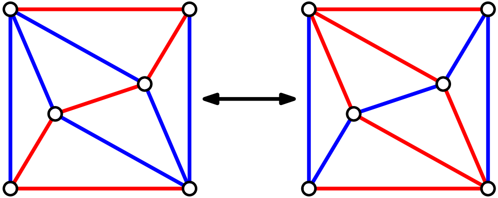
- Colouring of interior edges of the dual graph
- Red: vertical adjacency, Blue: horizontal adjacency
- Must satisfy local constraints





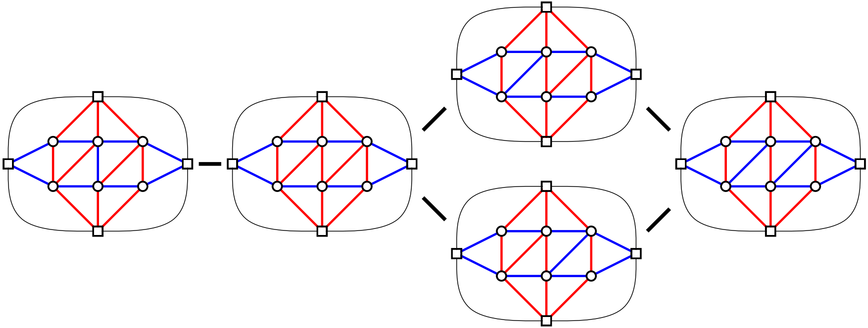
Regular Edge Labelings

- *Flip*: invert colour of all edges inside alternating 4-cycle



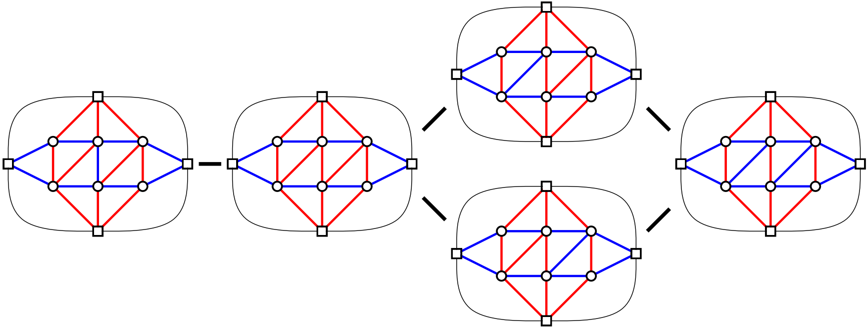
Regular Edge Labelings

- *Flip*: invert colour of all edges inside alternating 4-cycle
- *Flip graph*: Take all labelings, add an edge if two are connected by a flip



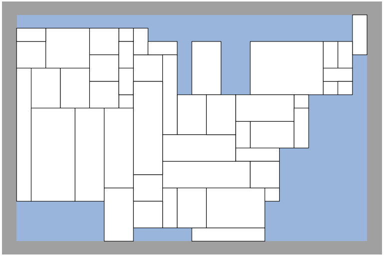
Regular Edge Labelings

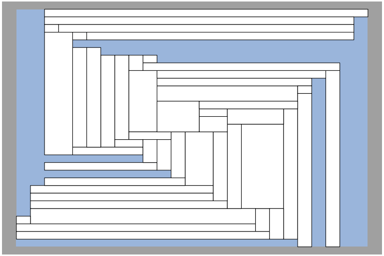
- *Flip*: invert colour of all edges inside alternating 4-cycle
- *Flip graph*: Take all labelings, add an edge if two are connected by a flip
- This graph is a distributive lattice [Fusy, 2009]



Regular Edge Labelings

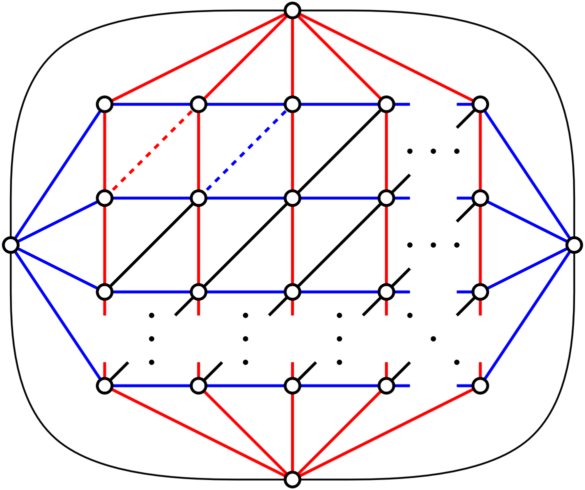
- Different labelings give cartograms of different quality
- The 'best' labeling can vary between data sets
- A local change in the labeling can cause global changes in the cartogram
- Can we try all of them?





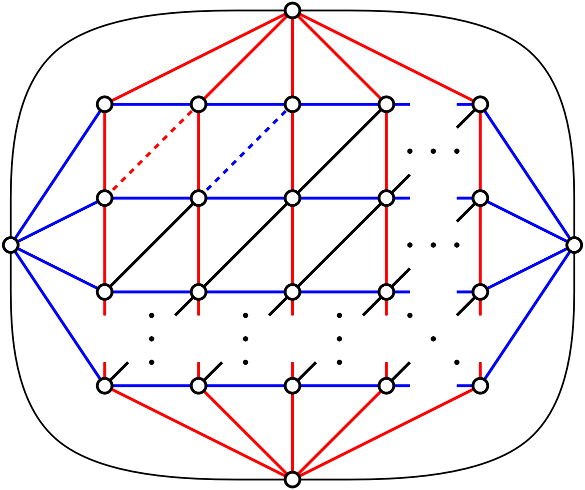
Counting Regular Edge Labelings – Lower Bound

- There are exponentially many



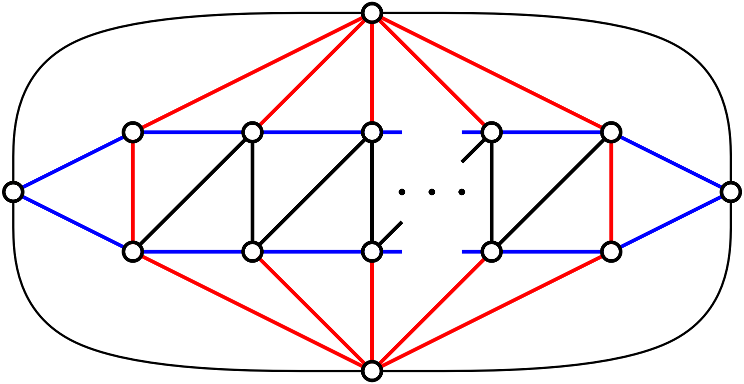
Counting Regular Edge Labelings – Lower Bound

- There are exponentially many
- Very restrictive: all grid edges already colored



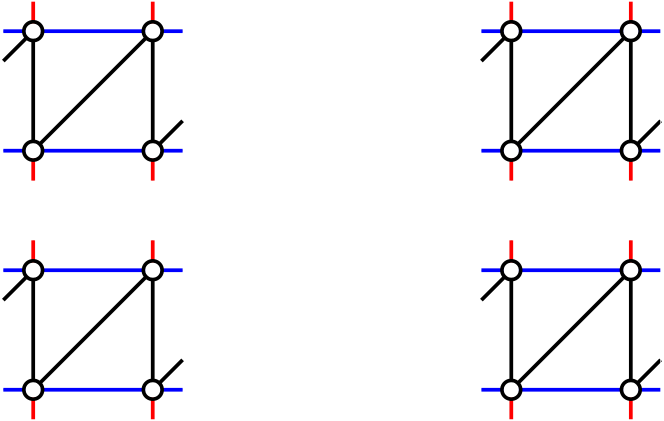
Counting Regular Edge Labelings – Lower Bound

- A simple case:



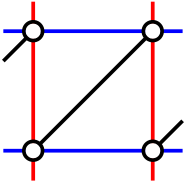
Counting Regular Edge Labelings – Lower Bound

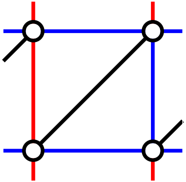
- A simple case:

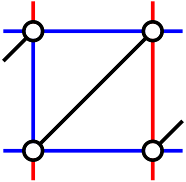


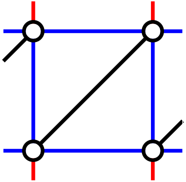
Counting Regular Edge Labelings – Lower Bound

- A simple case:



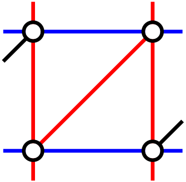
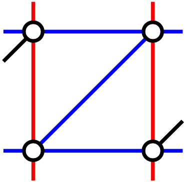
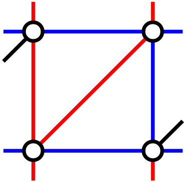
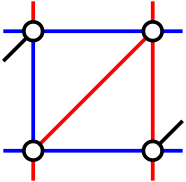
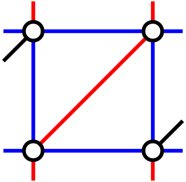






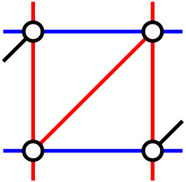
Counting Regular Edge Labelings – Lower Bound

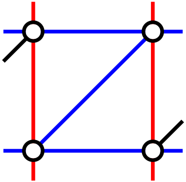
- A simple case:

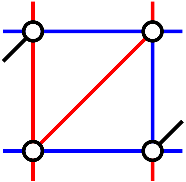






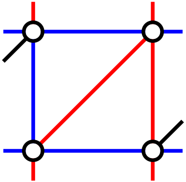
Counting Regular Edge Labelings – Lower Bound

- A simple case:

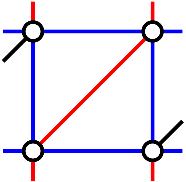






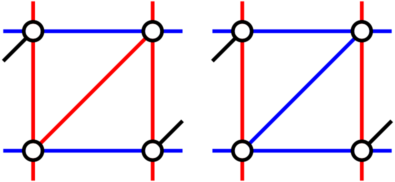


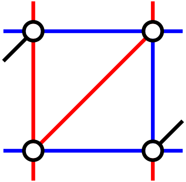
From \ To	R	B
R	2	1
B	1	1



- A simple case:

$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (2\ 1)$$





Counting Regular Edge Labelings – Lower Bound

- A simple case:

$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (2\ 1)$$

$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^2 = (2\ 1)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (5\ 3)$$

Counting Regular Edge Labelings – Lower Bound

- A simple case:

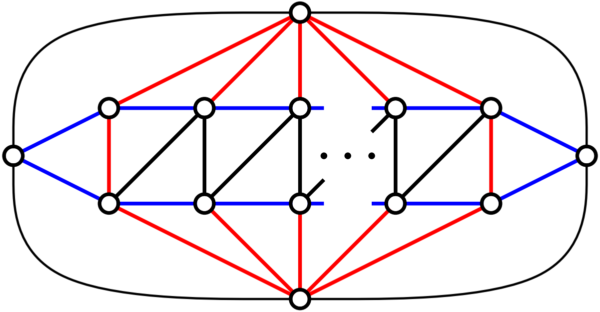
$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (2\ 1)$$

$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^2 = (2\ 1)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = (5\ 3)$$

$$(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^n \approx 2.618\left[(1\ 0)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{n-1}\right]$$

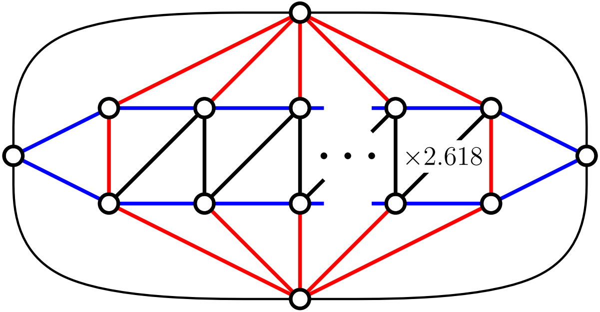
Counting Regular Edge Labelings – Lower Bound

- A simple case:

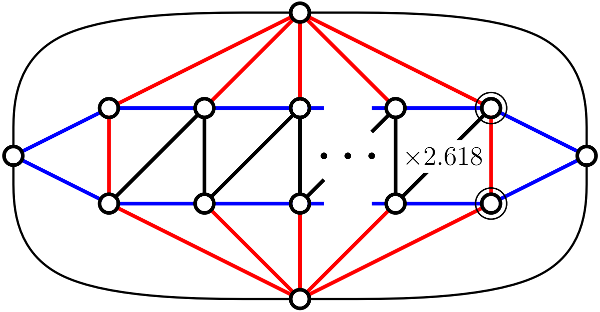


Counting Regular Edge Labelings – Lower Bound

- A simple case:

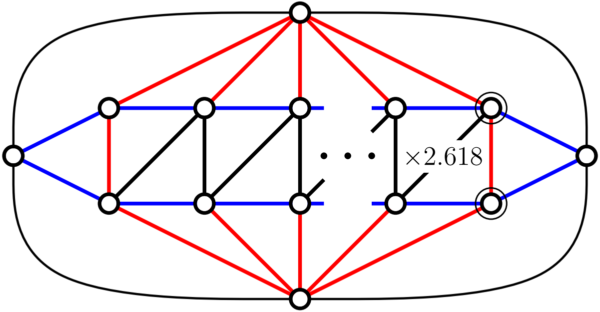


- A simple case:



Counting Regular Edge Labelings – Lower Bound

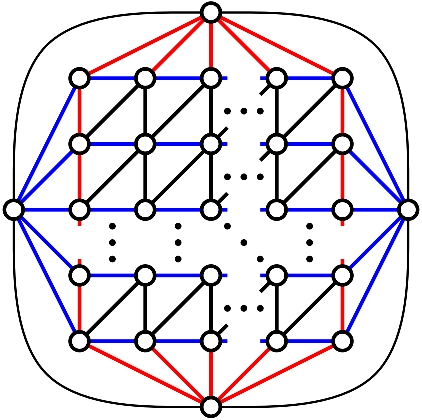
- A simple case:



$$2.618^{n/2} \approx 1.618^{n} \qquad (= \varphi^{n})$$

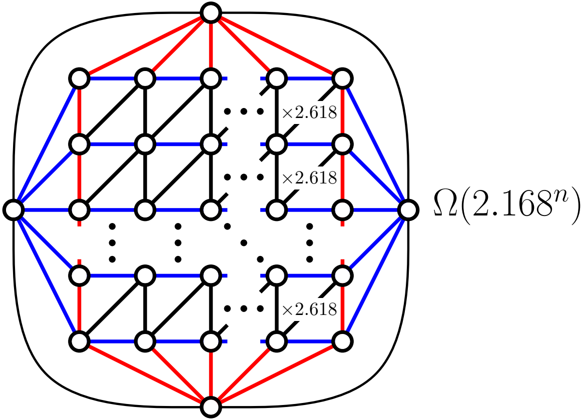
Counting Regular Edge Labelings – Lower Bound

- Gluing a lot of these together:



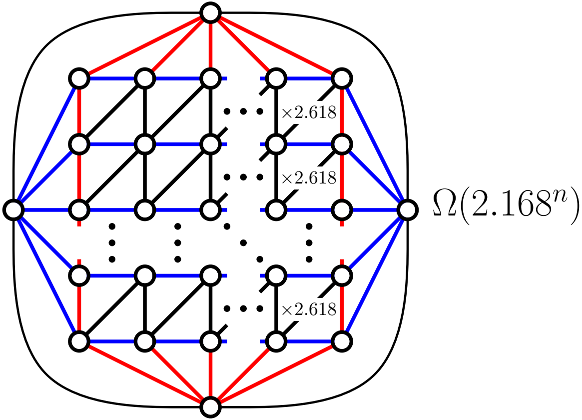
Counting Regular Edge Labelings – Lower Bound

- Gluing a lot of these together:



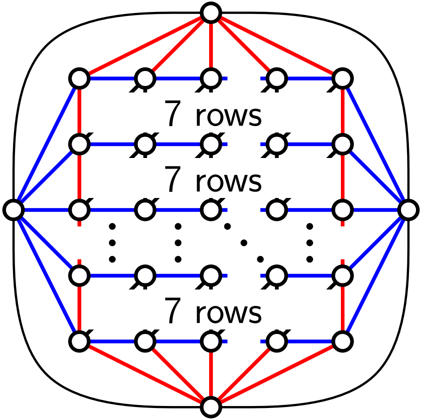
Counting Regular Edge Labelings – Lower Bound

- Still restrictive – all horizontal edges are blue



Counting Regular Edge Labelings – Lower Bound

- We get better bounds for multiple rows



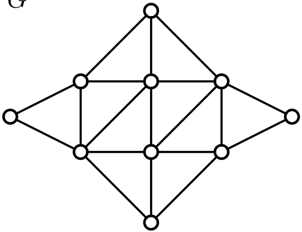
$\Omega(3.042^n)$

Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set

$S =$ Interior edges of G



Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

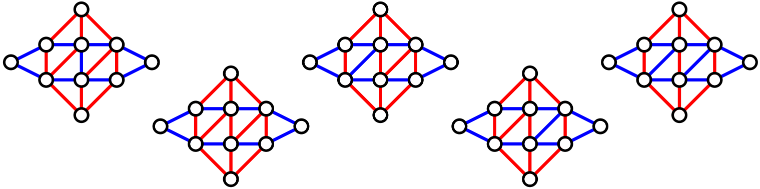
Let S be a finite set

Let $\mathcal{F}$ be a

collection of subsets of S

S = Interior edges of G

$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$



Counting Regular Edge Labelings – Upper Bound

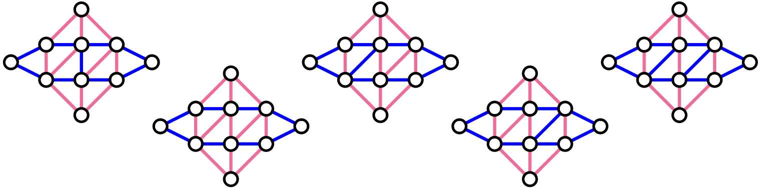
Lemma (Shearer's Entropy Lemma)

Let S be a finite set

Let $\mathcal{F}$ be a collection of subsets of S

S = Interior edges of G

$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$



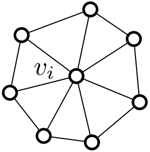
Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S

Let $\mathcal{F}$ be a collection of subsets of S

S = Interior edges of G
$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$
A_i = triangles incident to vertex i

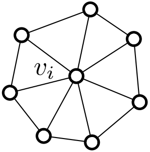


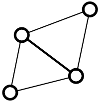
Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S

S = Interior edges of G
$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$
A_i = triangles incident to vertex i
$k = 4$



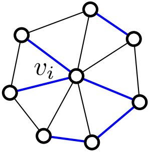


Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \leq i \leq m$.

$S =$ Interior edges of G
$\mathcal{F} = \{$Blue edges of $\mathcal{L} \mid \mathcal{L} \in$ regular edge labelings of $G\}$
$A_i =$ triangles incident to vertex i
$k = 4$
$\mathcal{F}_i =$



Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \leq i \leq m$. Then we have

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

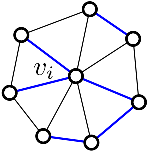
S = Interior edges of G

$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$

A_i = triangles incident to vertex i

$k = 4$

$\mathcal{F}_i =$



$|\mathcal{F}_i| \leq 2\binom{d_i}{4}2^4$

Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \leq i \leq m$. Then we have

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

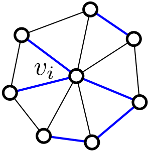
S = Interior edges of G

$\mathcal{F}$ = {Blue edges of $\mathcal{L}$ | $\mathcal{L} \in$ regular edge labelings of G}

A_i = triangles incident to vertex i

$k = 4$

$\mathcal{F}_i =$



$|\mathcal{F}_i| \leq 2\binom{d_i}{4}2^4$

$$|\mathcal{F}| \leq \prod_{i=1}^{n} |\mathcal{F}_i|^{1/4}$$

Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \leq i \leq m$. Then we have

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

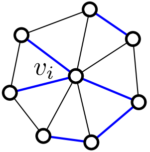
$S =$ Interior edges of G

$\mathcal{F} = \{$Blue edges of $\mathcal{L} \mid \mathcal{L} \in$ regular edge labelings of $G\}$

$A_i =$ triangles incident to vertex i

$k = 4$

$\mathcal{F}_i =$



$|\mathcal{F}_i| \leq 2\binom{d_i}{4}2^4$

$$|\mathcal{F}| \leq \prod_{i=1}^{n} \left(2\binom{d_i}{4}2^4\right)^{1/4}$$

Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \leq i \leq m$. Then we have

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

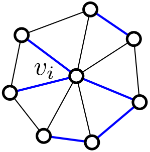
S = Interior edges of G

$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$

A_i = triangles incident to vertex i

$k = 4$

$\mathcal{F}_i =$



$|\mathcal{F}_i| \leq 2\binom{d_i}{4}2^4$

$$|\mathcal{F}| \leq \prod_{i=1}^{n} \left(2\binom{d_i}{4}2^4\right)^{1/4}$$

$$|\mathcal{F}| \leq \left(2\binom{6}{4}2^4\right)^{n/4}$$

Counting Regular Edge Labelings – Upper Bound

Lemma (Shearer's Entropy Lemma)

Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the $A_1, \ldots, A_m$. Let $\mathcal{F}$ be a collection of subsets of S and let $\mathcal{F}_i = \{F \cap A_i : F \in \mathcal{F}\}$ for $1 \le i \le m$. Then we have

$$|\mathcal{F}|^k \le \prod_{i=1}^{m} |\mathcal{F}_i|.$$

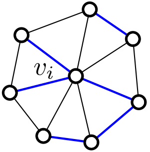
S = Interior edges of G

$\mathcal{F} = \{\text{Blue edges of } \mathcal{L} \mid \mathcal{L} \in \text{regular edge labelings of } G\}$

A_i = triangles incident to vertex i

$k = 4$

$\mathcal{F}_i =$



$|\mathcal{F}_i| \le 2\binom{d_i}{4}2^4$

$$|\mathcal{F}| \le \prod_{i=1}^{n} \left(2\binom{d_i}{4}2^4\right)^{1/4}$$

$$|\mathcal{F}| \le \left(2\binom{6}{4}2^4\right)^{n/4}$$

$$|\mathcal{F}| \le 4.68^n$$

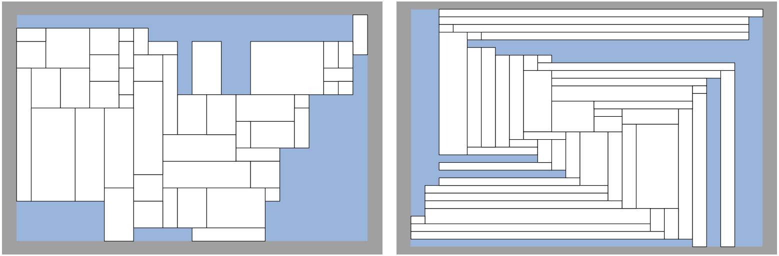
Counting Regular Edge Labelings – Summary

Theorem

A graph on n vertices has no more than 4.68^n regular edge labelings.
There are graphs with more than 3.042^n regular edge labelings.

In Practice

- Maps of EU and US have over 1300 million labelings
- Good case for heuristic search



Evolution Strategy

1. Initialize a random population of labelings
   - Start at the minimum labeling
   - Follow a random path towards the maximal labeling
   - Stop at a random height

Evolution Strategy

1. Initialize a random population of labelings
    - Start at the minimum labeling
    - Follow a random path towards the maximal labeling
    - Stop at a random height
2. Copy a few of the best labelings to the next generation

Evolution Strategy

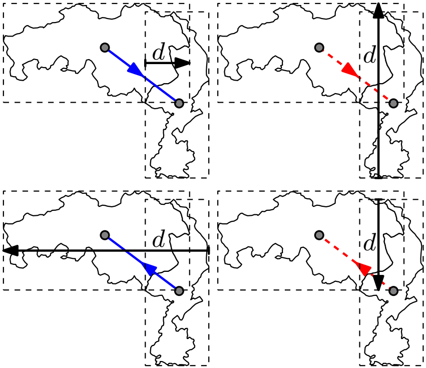
1. Initialize a random population of labelings
   - Start at the minimum labeling
   - Follow a random path towards the maximal labeling
   - Stop at a random height
2. Copy a few of the best labelings to the next generation
3. Use selection and mutation to fill the rest of the population
   - Select a random labeling, favouring higher quality
   - With low probability, take a long random walk in the lattice
   - With high probability, take one random step in the lattice

Evolution Strategy

1. Initialize a random population of labelings
    - Start at the minimum labeling
    - Follow a random path towards the maximal labeling
    - Stop at a random height
2. Copy a few of the best labelings to the next generation
3. Use selection and mutation to fill the rest of the population
    - Select a random labeling, favouring higher quality
    - With low probability, take a long random walk in the lattice
    - With high probability, take one random step in the lattice
4. Repeat steps 2 and 3 a fixed number of times, keeping track of the best labeling found

Quality Criteria

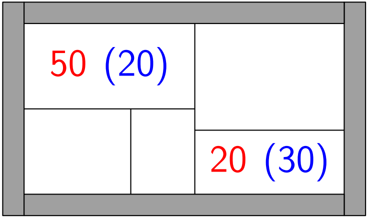
- Correct adjacencies
- Relative positions of the regions



Quality Criteria

- Correct adjacencies
- Relative positions of the regions
- Cartographic error of the regions

$$\frac{|A_c - A_s|}{A_s} \qquad \begin{array}{l} A_c\text{: Area in cartogram} \\ A_s\text{: Area specified} \end{array}$$



Error: 1.5

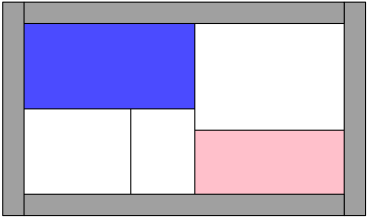
Error: 0.33

Quality Criteria

- Correct adjacencies
- Relative positions of the regions
- Cartographic error of the regions

$$\frac{|A_c - A_s|}{A_s}$$

A_c: Area in cartogram
A_s: Area specified



Error: 1.5

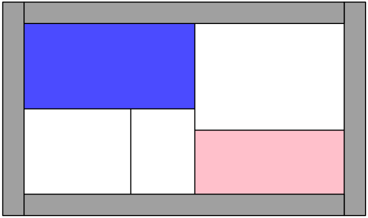
Error: 0.33

Quality Criteria

- Correct adjacencies
- Relative positions of the regions (30 %)
- Cartographic error of the regions (70 %)

$$\frac{|A_c - A_s|}{A_s}$$

A_c: Area in cartogram
A_s: Area specified



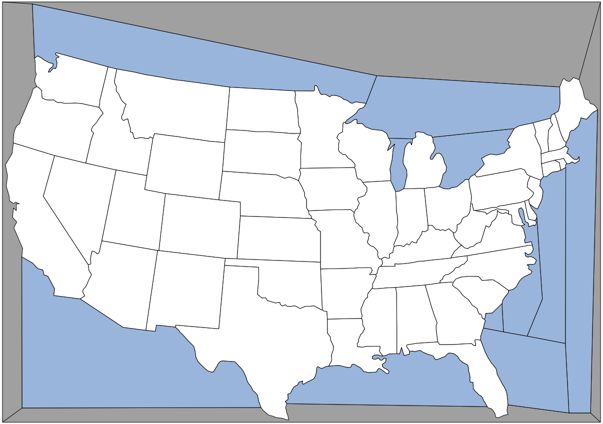
Error: 1.5

Error: 0.33

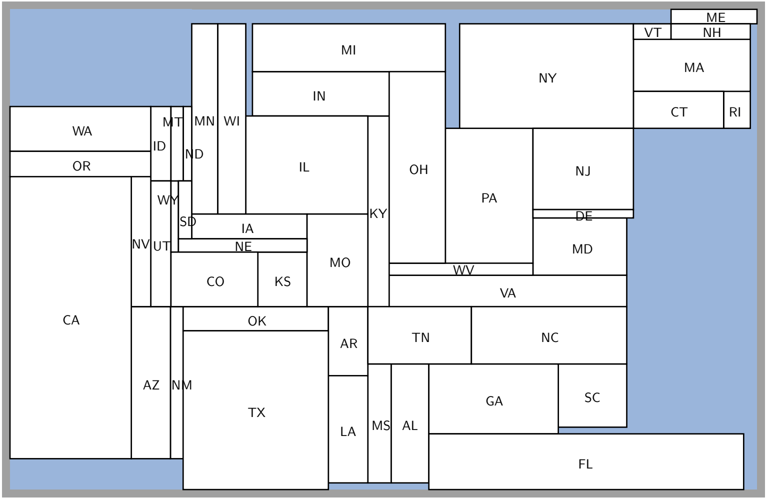
Experimental Results

United States (63 regions)

- Made New Mexico adjacent to Utah to resolve 4-state point
- All US Census data sets with a positive value for each state
- Generating one cartogram took ∼8 minutes

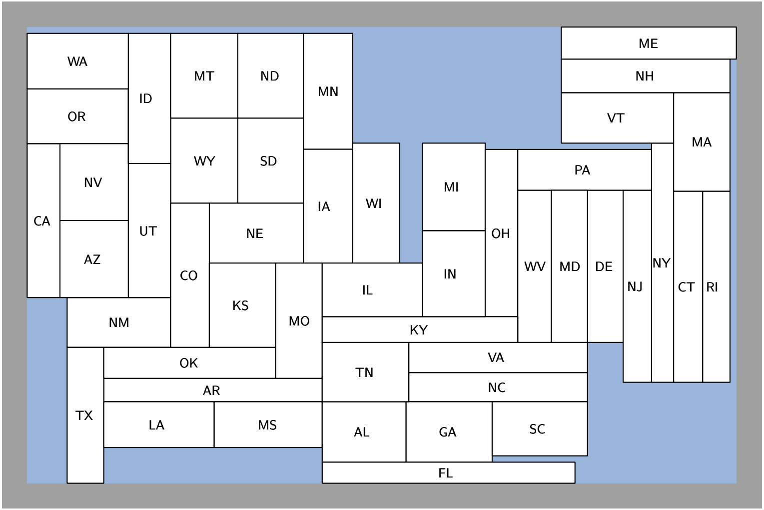


United States - Population



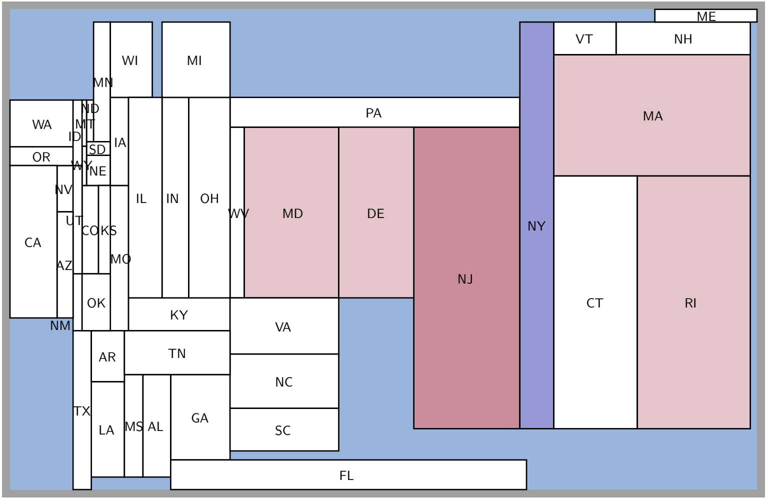
Average error: 0.5 %   Maximum error: 2.2 %

United States - High School Graduates (%)



Average error: 0.0 %　　Maximum error: 0.0 %

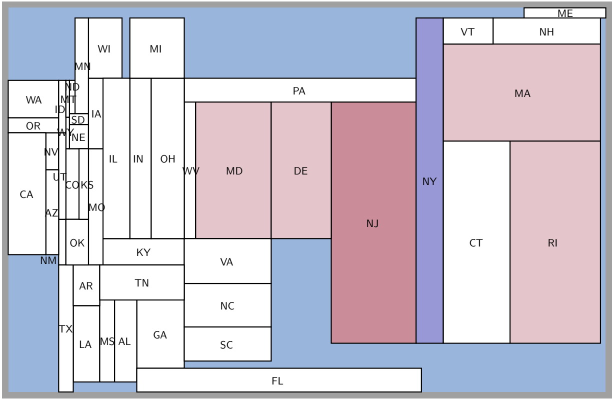
United States - Population per square mile



Average error: 2.0 % Maximum error: 11.3 %

United States - Population per square mile

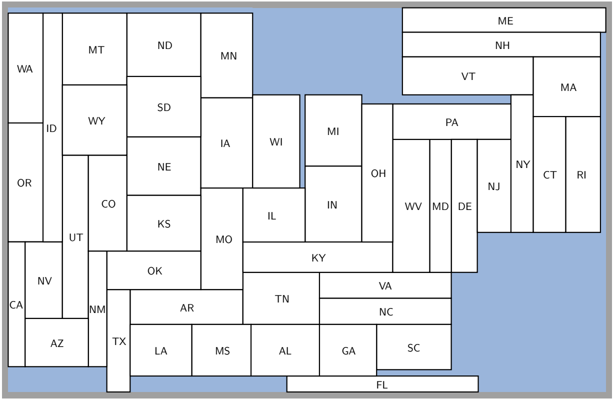
- Negative correlation with land area
- Large variation



Average error: 2.0 % Maximum error: 11.3 %

United States - Non-Hispanic White Population (%)

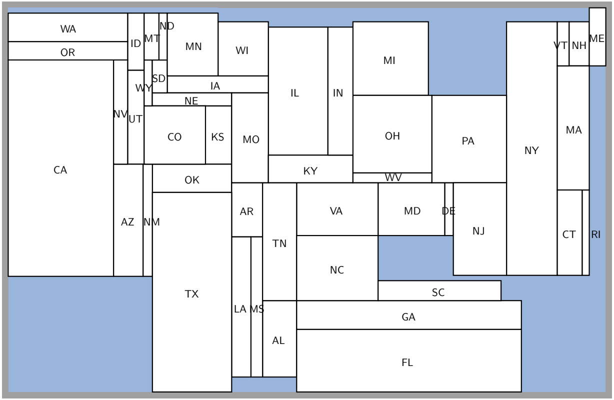
- Negative correlation with land area
- Large variation



Average error: 0.0 % Maximum error: 0.0 %

United States - Businesses w/o Paid Employees

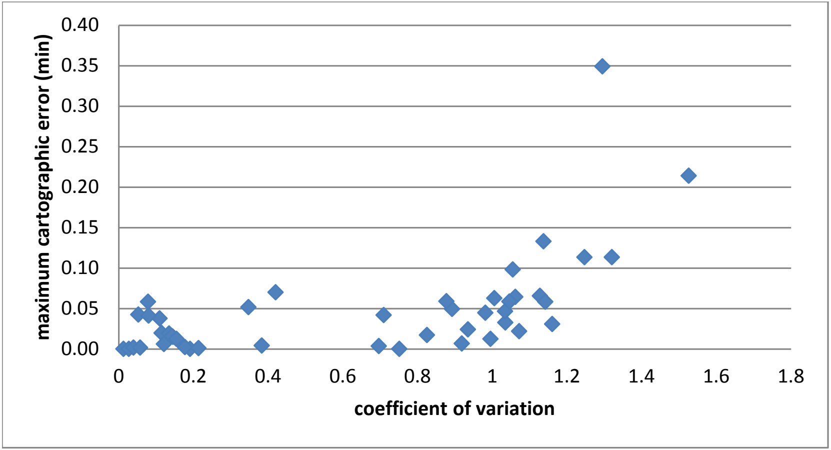
- Negative correlation with land area
- Large variation



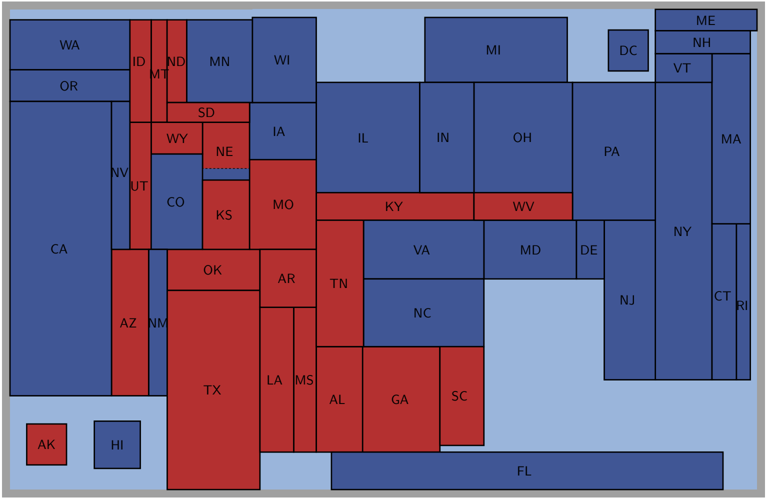
Average error: 0.7 % Maximum error: 3.1 %

United States - Variation vs Maximum Error

- Negative correlation with land area
- Large variation



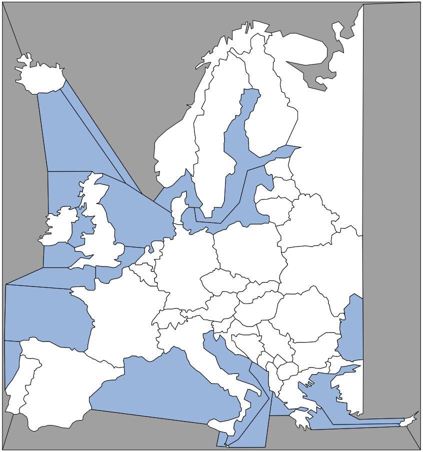
United States - 2008 Presidential Elections



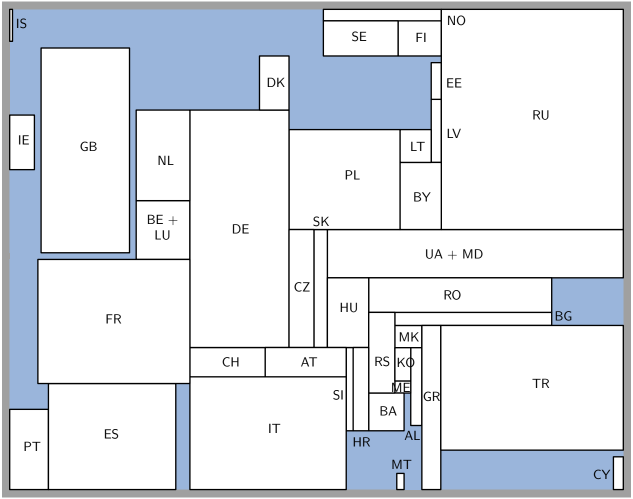
Average error: 0.0 %   Maximum error: 0.0 %

Europe (65 regions)

- Merged Luxembourg with Belgium, and Moldova with Ukraine
- All CIA World Fact Book ranked data sets with data for each country
- Generating one cartogram took ~6 minutes



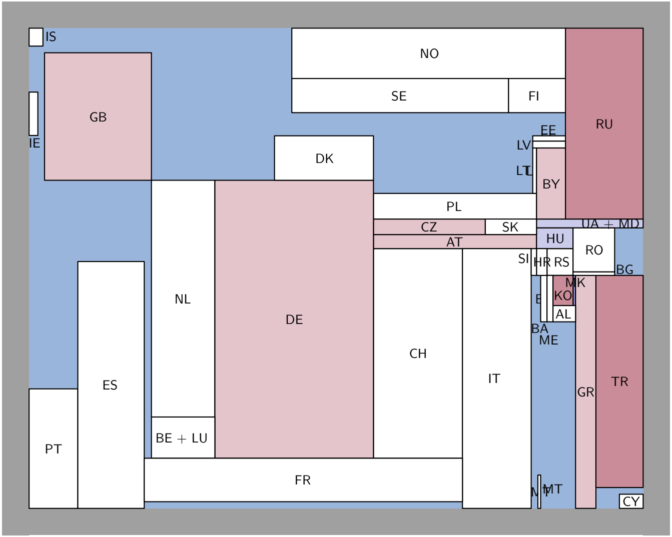
Europe - Population



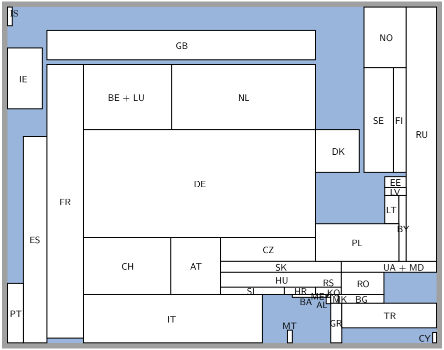
Average error: 0.0 %

Maximum error: 0.0 %

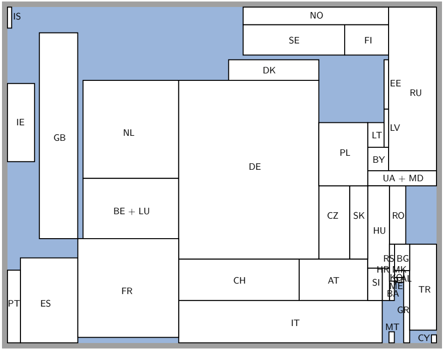
Europe - Account Balance



Average error: 4.0 % Maximum error: 19.7 %



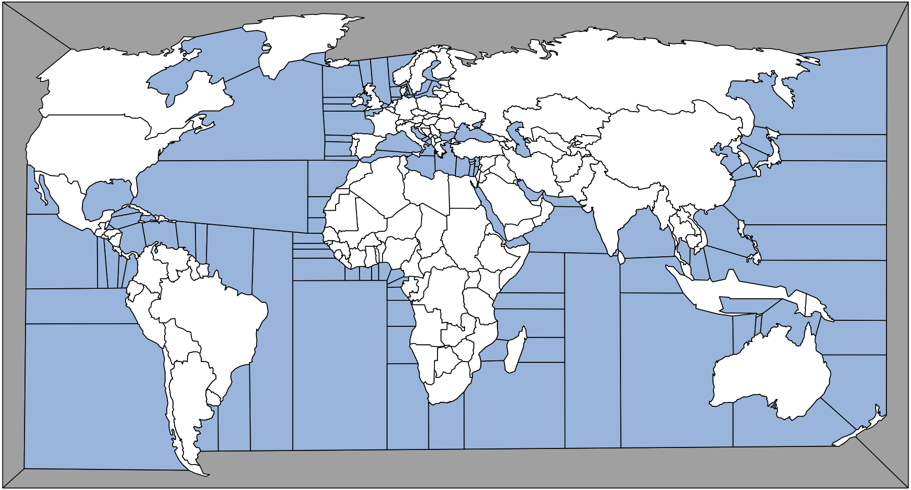
Max error: 0.0 %   BBSD: 0.088



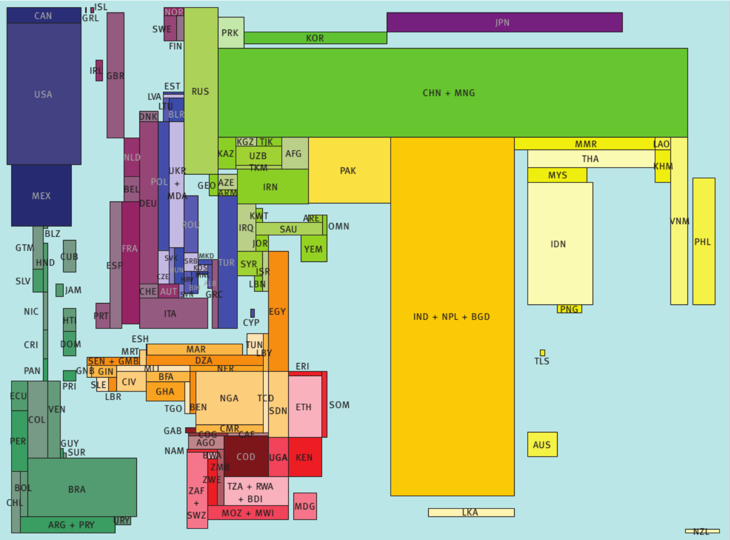
Max error: 1.6 %   BBSD: 0.078

World (275 regions)

- All countries with population over 1 million
- Generating one cartogram took ~3.5 hours



World - Population



Average error: 1.17 %   Maximum error: 18.5 %