POLITECNICO DI TORINO

Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica

Tesi di Laurea Magistrale

An Algorithmic Approach to Redraw US Gerrymandered District Boundaries by Minimizing Wasted Votes

Relatori:
prof. Silvia Chiusano
prof. Elena Baralis

Candidato:
Laura Palmieri

Anno accademico 2017-2018

Summary

Partisan gerrymandering consists of redrawing the district boundaries to give electoral advantage to a political party.
In 1986, it was declared unconstitutional and justiciable by the US Supreme Court and, since then, many efforts have been done to find a standard that could be adopted by the Court to quantify gerrymandering and eventually reject a redistricting plan. In previous studies, it was concluded that notions such as quantitative measure of shape compactness and other geometric indices had many limitations, as redistricting policies take into account other constraints, and the algorithms that used those indices were highly computationally complex and made the redistricting process infeasible. Recently, Stephanopoulos and McGhee introduced Efficiency Gap, a new measure of partisan gerrymandering, which is defined as the ratio of the difference between the parties' wasted votes (in a two-party electoral system) to the total number of votes cast in the election. This metric was found legally convincing by a US Appeals Court in a case appealed in 2017.
The aim of this project is providing a local search algorithm able to "un-gerrymander" the 2012 congress district maps for Wisconsin, Virginia, Texas and Pennsylvania by bringing their efficiency gaps to acceptable levels.
If the US Supreme Court upholds the decision of lower courts, our work can provide a crucial supporting hand to remove partisan gerrymandering.

Acknowledgements

I would first like to thank my thesis advisor Bhaskar DasGupta, Professor in the Computer Science department at University of Illinois at Chicago (UIC). He was very helpful and his door was always open whenever I needed it.
I would also like to thank two PhD students involved in this research project, Tanima Chatterjee and Zainab Al-Qurashi, the Assistant Professor Anastasios Sidiropoulos in the Computer Science department, Professor Silvia Chiusano and Professor Elena Baralis of the Department of Control and Computer Engineering at Politecnico di Torino (Italy) as, respectively, my Italian advisor and member of my defense committee.
Moreover, this journey would not have been possible without the support of my family and friends. To my mom, my dad and my sister thank you for encouraging me and supporting me.
Thank you Tanima, for being a colleague, a gym partner and especially a friend. You helped me a lot in reaching my goals and become more confident and strong. Thank you Giorgia, for being the best friend ever since my first year at university. Thank you for supporting me every day, for encouraging me, for teaching me life lessons and for making me rise when I was falling down. Without you this journey would have not been the same. You made me grow a lot and you will always have a special place in my heart. Thank you Harshit, for being the nice person that you are and for supporting and encouraging me to take challenges and to never give up. Thank you Andrea, for being almost a roommate and a good friend. Thank you Stefano and Francesca for taking care of me when I was in Turin and for being a second family to me. Thank you Ylenia, for being a wonderful, sweet and caring friend. Thank you Marco, Carlo and Lillo, for our funny and crazy days in Turin. Thank you Alessandro, Fabrizio and Giulia for our special dinners all together.
I am glad I have been surrounded by wonderful people during this years.

LP

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 What is gerrymandering?

1.1.1 Redistricting in the US

Every ten years, in the US, there is a population census to keep track of demography changes. The number of voters can change and the district boundaries must be redefined to make sure that they contain approximately an equal number of people. The redistricting process is very complex and challenging as there are many constraints and requirements to consider in order to accept a certain plan. In most cases, redistricting is controlled by the state legislature and the ruling party may redraw the district boundaries to gain political advantage. Whenever politics is influencing this process we talk about gerrymandering.

1.1.2 Gerrymandering: definition

According to *The American Heritage Dictionary of the English Language*, gerrymandering is "a practice used to divide a geographic area into voting districts in a way that gives one party an unfair advantage in elections" [8]. In this way, new drawn districts can ensure the majority of seats in the House of Representatives even with a minority of the total amount of votes.
Packing and cracking are the two main approaches used to gerrymander a state: packing consists of concentrating voters of one party in a single district so that they may have a majority there but no strength in other districts and cracking makes the opposition voters be spread across many districts as they can be a majority in a single one.
Figure 1.1 represents a state divided into 5 districts and shows how the final number of seats in a two-party system can change depending on the district boundaries of the geographical area. In the initial representation there are compact districts

and, as the blue party has more votes than the green one, the number of seats is proportional to the majority. If district boundaries are redrawn with strange and irregular shapes, such as in the second picture, the final outcome can be different: in this case, even if the green party has a lower number of total votes, it wins 3 seats out of 5.

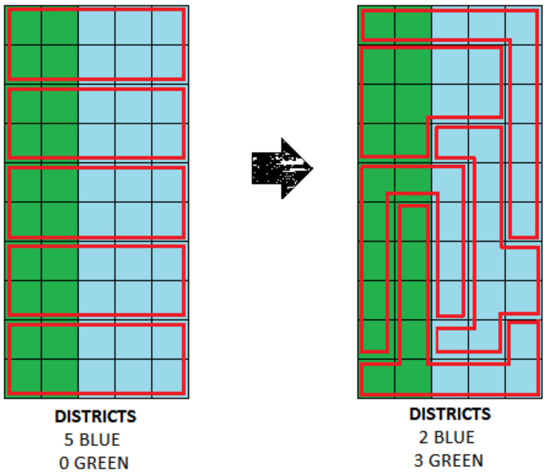


Figure 1.1: Example of redistricting

1.1.3 Gerrymandering: a brief history

The term gerrymander was used for the first time in 1812. During the State Senate elections, the Massachusetts legislature drew the district of South Essex in order to favor the Democratic-Republican candidates of the governor Elbridge Gerry over the Federalists [30]. The map of that district resembled the shape of a salamander and the word "gerrymander" was created by putting together "Gerry" and "salamander". Since then, any unfair district shape manipulation was described as gerrymandered. Gerrymandering has continued to be a curse to fairness of electoral systems in the US for a long time despite the general public disdain for it. In 1986, the US Supreme Court ruled that gerrymandering is justiciable [1], as it violates the Equal Protection Clause, but an agreement on how to effectively estimate it was not found.
In 2004, the perceived partisan gerrymandering in Pennsylvania was not declared unconstitutional and the nine justices in the Court were split: while some of them

believed it was not possible to define an official standard for partisan gerrymandering, others could not agree on which metric to use among the existent ones. In particular, justice Anthony Kennedy did not foreclose the possibility to develop future standards and challenged lower courts to help identify it [2].
In 2006, "partisan bias", defined as an asymmetrical distribution of gained seats and votes for a political party, was found to be an interesting tool to understand and remedy gerrymandering [3] but still was not enough to represent an official standard for the Court.
In 2014, the "efficiency gap" [9, 23], defined as the difference of total wasted votes between parties in an election, has been considered by the court as a valid tool to detect gerrymandering.
After the population census in 2010, there were new redistricting plans for several states. The elections outcomes in the last decade showed the most extreme partisan gerrymandering in the American history and it is very important that the US Supreme Courts adopts a standard to quantify gerrymandering and to have more control over the redistricting process.

1.2 The idea for this project

1.2.1 The case of Wisconsin: Gill v. Whitford

Gill v. Whitford [4] is a recent US Supreme Court case addressing the constitutionality of partisan gerrymandering. After the 2010 census, Wisconsin redistricting plan was created by Republican legislators to maximize their number of seats in the State legislature. Then, in 2016, plaintiffs challenged the plan as unconstitutional partisan gerrymander because the vote strength of the democrats had been diluted statewide.
On October 3rd 2017, the US Supreme Court heard this case. The main issue was that plaintiffs didn't find a workable standard to detect gerrymandering while redistricting and, as computers and data analytics are making gerrymandering become more sophisticated, there is the strong need of a concrete measure that can be accepted and legally used. This case is still pending.
After thirty years from when gerrymandering was declared justiciable, for the first time there may be a positive outcome and the US Supreme Court may decide whether some measures can be officially considered as appropriate standards to quantify gerrymandering. In particular, the efficiency gap was found legally convincing by the court.

1.2.2 Goal of the project

The goal of this research project is to use an algorithmic approach to "un-gerrymander" the 2012 House district maps for the most gerrymandered states of the last decade by bringing their efficiency gap to acceptable levels. If the US Supreme Court upholds the decision of lower courts, this algorithm and its implementation will provide a crucial supporting hand to remove partisan gerrymandering and, moreover, the aim of this work is also demonstrating how mathematics together with computing can represent a useful and powerful tool that can be applied to political science.

Chapter 2

PREVIOUS WORK

2.1 Standard measure for gerrymandering

During the last decades, there have been many theoretical and empirical attempts to quantify gerrymandering and draw "fair" district boundaries using well-known notions such as compactness and symmetry.
Initially, some researches tried to improve geometric indexes, like *compactness* [7, 15, 21, 24] or *convexity* [12], to distinguish strange natural boundaries (eg. shorelines) from artificial complexity due to political manipulation. These studies show that detecting gerrymandering only from the point of view of district shape is not enough: there is a high degree of arbitrariness as some trade-offs are needed to preserve communities of interest and other constraints required in the redistricting process [14]. Despite this, including geometric tools together with other metrics or requiring the districts to be as compact or convex as possible can represent an obstacle for gerrymandering purposes while defining districts boundaries. Other popular measures in the literature are *symmetry* [9, 10], which "identifies whether a party receives a larger share of seats than the other party for the same share of votes", and *responsiveness* [9, 10], which "captures the overall competitiveness of the system as the number of seats that change hands for a given shift in the aggregate vote".
Despite all these efforts, the US Supreme Courts has not been convinced yet to adopt one or more of these measures as possible standards to detect gerrymandering.
Recently, N. Stephanopoulos and E. McGhee in two papers [9, 23] have introduced a new metric called the "efficiency gap" which is based on the difference of total wasted votes between parties in a two-electoral system.
Our research project proposes a tool for redistricting based on the efficiency gap as it provides a mathematically precise measure of gerrymandering and it was found legally convincing by the US Supreme Court case Gill v. Whitford [4].

2.1.1 The Efficiency Gap

The Efficiency Gap is defined in [23] as "the difference between the parties' respective wasted votes in an election - where a vote is wasted if it is cast for a losing candidate, or for a winning candidate but in excess of what she needed to prevail". All the district plan choices regarding cracking or packing are grouped together in a single measure: losing candidates may have a large number of votes because they have been spread in many districts and, at the same time, winning candidates may have excessive votes because they have been packed all together in the same district. Table 2.1 shows an example of calculation of the efficiency gap for a two-party system (see a similar example also in [23]). Supposed that there are 10 districts with 100 voters for each, both parties (Party A and Party B) have a certain number of won and wasted votes. Party A wins 540 votes and has 140 wasted votes, while party B wins 460 votes and has 360 wasted votes. The difference between the parties' wasted votes, divided by 1000 total voters in all the districts, gives an efficiency gap of 18%. From a mathematical point of view, this result shows that party A won 18 percent more seats than it would have had if both parties had the same number of wasted votes.

Table 2.1: COMPUTATION OF THE EFFICIENCY GAP

District	WonA	WastedA	WonB	WastedB	Winner
1	60	10	40	40	A
2	70	20	30	30	A
3	70	20	30	30	A
4	54	4	46	46	A
5	54	4	46	46	A
6	54	4	46	46	A
7	54	4	46	46	A
8	54	4	46	46	A
9	35	35	65	15	B
10	35	35	65	15	B
Tot. votes	**540**	**140**	**460**	**360**	A

2.2 Tools for redistricting

It is often not possible to go over every redistricting map to optimize the gerrymandering measure due to rapid combinatorial explosion and researchers tried to find efficient algorithmic solutions for this purpose.
The first explorations for redistricting were based on full enumeration approaches which were computationally expensive and feasible just for small problems. Later, the literature focused on simulation [7, 26] and heuristic approaches: the first one selects a large number of possible redistricting maps and the latter identifies maps with desirable characteristics, even though implementing a good heuristic search may be complex when there is a huge number of feasible maps. Moreover, thanks to supercomputers and technological advances (eg. highly scalable message passing models), parallel computation is another approach that has been used to handle the computational complexity and to make the space search more feasible and optimized [24, 31].

Chapter 3

Experimental Methodology

This chapter is divided into two sections. The first part focuses on the algorithm that has been proposed and a description of the input dataset, while the second part focuses on the method used for visualizing the maps.

3.1 The algorithm

Going over every possible map in a redistricting process to optimize a gerrymandering measure is unfeasible because of its computational complexity. Therefore, minimizing the efficiency gap requires to find a solution to a combinatorial optimization problem.

3.1.1 Computational complexity

The first step in the algorithm implementation was formalizing the optimization problem of minimizing the efficiency gap measure, analyzing its properties, studying the computational complexity and the algorithm design.
Thanks to the theoretical analysis done by Professor B. DasGupta and A. Sidiropoulos at University of Illinois at Chicago (see supplemental material and [25]), we obtained the following theoretical results.
Lemma 1 shows that, depending on the number of districts, there is a finite discrete set of rational values attainable by the efficiency gap. These results are useful to understand the sensitivity of this metric and to design an efficient algorithm. Moreover, even if the problem seems theoretically intractable (see Theorem 1), the empirical results show that, in practice, it is computationally feasible in polynomial time as, in real-life applications, many constraints in the theoretical formulation of this optimization problem are often relaxed. For example:

Redistricting district shapes The state law requires compactness among districts: 37 states require their legislative districts to be reasonably compact and 18 states require congressional districts to be compact as well [14].

Variations in district populations Every district, in the reality, has approximately but not exactly an equal population. For example, standards for congressional districts require an equal population *as nearly as is practicable* while state and local legislative districts have more flexibility but must have *substantially* equal population (no more than 10% between larger and smaller district) [14].

Bounding the Efficiency Gap away from zero The author that originally proposed the efficiency gap measure provided several reasons for not requiring the normalized efficiency gap to be either zero or too close to zero (see [23] pp.886-887).

3.1.2 Input dataset

The algorithm was tested on four congress district maps and related data for the 2012 House of Representatives elections in Wisconsin, Texas, Virginia and Pennsylvania [17–20].
The input map is initially preprocessed to create an undirected unweighted planar graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
Every node in the graph is a county (or a portion of it) that is assigned to a district and two nodes are connected if and only if they share a border on the map. Moreover, every node $\nu \in \mathcal{V}$ has three numbers: PartyA(ν) (total number of voters for PartyA), PartyB(ν) (total number of voters for PartyB) and Pop(ν) = PartyA(ν) + PartyA(ν)[1].
A district $\mathcal{Q}$ is a connected sub-graph of $\mathcal{G}$ with PartyA($\mathcal{Q}$) = $\sum_{\nu \in \mathcal{Q}}$PartyA($\nu$) and PartyB($\mathcal{Q}$) = $\sum_{\nu \in \mathcal{Q}}$PartyB($\nu$).
Every input dataset is an Excel spreadsheet and explanations of various columns of the spreadsheet are as follows:

- *District* (column 1): this column identifies the district the county in column 3 belong to (initially considering the results of the 2012 US House of Representatives elections).
- *County_ID* (column 2): this ID, together with the district number, is a unique identifier for the counties in column 3. Every county is identified by its

[1]The computation of the efficiency gap is here computed for two-party systems. "Third-party" votes other than democratic and republican partied are ignored

County_ID and the District it belongs to. This was specifically needed in the algorithm to handle those counties shared among two or more districts.

- *County* (column 3): name of the county.
- *Republicans* (column 4): this contains the total number of votes in favor of the Republican party (GOP) in the county identified by column 1 and column 2.
- *Democrats* (column 5): this contains the total number of votes in favor of the Democratic party in the county identified by column 1 and column 2.
- *Neighbors* (column 6): this column contains information about the "neighboring counties" of the given county. Neighboring counties represent the counties that share a boundary with the county identified by column 1 and column 2. Individual neighbors are separated by commas.

3.1.3 Local search algorithm

The problem of minimization of the efficiency gap can be solved with a fast semi-randomized algorithm based on the local search paradigm. As producing all the possible maps in the redistricting process is practically unfeasible, the local search allows to keep track of a single current state by moving only to neighbors states and by ignoring paths. Moreover, there is a very low usage of memory and a reasonable solution can be found also in a large search space. The process is also made semi-randomized, as there is a maximum number of counties (set by the user) that can be shifted at every iteration, and this allows the search not to be stuck in a local minimum most of the time.
Initially, the algorithm starts with the existing (possibly gerrymandered) districts and it is executed multiple times to bring the efficiency gap below a certain threshold.
Fixing an appropriate threshold is another important issue to determine whether the considered plan is an example of unconstitutional partisan gerrymander. N. O. Stephanopoulos and E. McGhee in [23] proposed that efficiency gaps of two seats (in congressional delegations) and 8 per cent (in state legislative chambers) can be used as a minimum threshold.
While exploring all the possible redistricting maps, some parameters and constraints are set in order to reduce the search space, by rejecting some solutions, or to randomize the algorithm:

- *Min population factor* and *Max population factor*: they are respectively set to 0.90 and 1.00. Solutions with a population deviation above 10 per cent are rejected. This ensure that every district has almost an equal number of voters.

- *Continuous districts*: one county is shifted to a neighbor district if and only if the neighbors still maintain continuity among the districts without creating holes.
- *Max number of random counties*: at every iteration, a number of counties between 1 and a max parameter can be shifted.

The pseudo-code in the next page (Algorithm 23) describes each step taken in the execution of the algorithm and, in supplemental material, a flow chart is also provided (Figure B.1).
The code was implemented in Python. In each algorithm execution, on average about 100 iterations are carried out. In every iteration, a random number $\boldsymbol{n}$ of counties (or portions of a county if it is shared between two or more districts) is selected from a range $(1, K)$ where K (defined by the user) is the maximum number of counties that can be shifted in other districts and has to be less than the total number of counties in the state.
The next step in the algorithm consists of analyzing each county selected in the random sample. If all the neighbors of the current county belong to the same district, the county is not on a boundary between two districts and its shifting would create a discontinuous map. On the other hand, if the county has neighbors belonging to different district and if it has not been considered and shifted yet, it is reassigned in neighboring districts.
Every time a county is shifted, the efficiency gap is recomputed: if the population deviation inside districts of the new map is below 10 per cent the solution is acceptable and, in case the new efficiency gap is less than the current one, the new district plan with the updated efficiency gap becomes the input for the next iteration.
The algorithm can be run multiple times until the efficiency gap goes below the specified threshold.

Data: Input file as described in section 3.1.2 with current κ districts $\{\mathcal{Q}_1, ..., \mathcal{Q}_\kappa\}$

Result: Output file with new κ districts and minimum Eg

repeat
- select a random r from the set $\{0,1,...,k\}$ for some $0 < k < |\mathcal{V}|$ (*k is the max #counties that can be shifted*);
- select r nodes ν_1,...,ν_r from G at random (* Note that a node is a county or part of a county *);
- counties_done $\leftarrow$ 0;
- **foreach** ν_i **do**
  - **if** *all neighbors of* ν_i *do not belong to the same district as* ν_i **then**
    - **if** $\nu_i \notin$ *counties_done* **then**
      - add ν_i to counties_done;
      - **for** *every neighbor* ν_j *of* ν_i **do**
        - **if** *assigning* ν_i *to the district of* ν_j *generates a connected map* **then**
          - assign ν_i to the district of ν_j;
          - recalculate new districts, say $\mathcal{Q}'_1, ..., \mathcal{Q}'_\kappa$;
          - **if** $min_{1\leq i\leq\kappa}(Pop(\mathcal{Q}_i)) \leq Pop(\mathcal{Q}'_j) < max_{1\leq i\leq\kappa}(Pop(\mathcal{Q}_i))$ $\forall\ j$ **then**
            - **if** $Eg_\kappa(\mathcal{P}, \mathcal{Q}'_1, ..., \mathcal{Q}'_\kappa) \leq Eg_\kappa(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ **then**
              - $\mathcal{Q}_1 \leftarrow \mathcal{Q}'_1; \mathcal{Q}_2 \leftarrow \mathcal{Q}'_2; ...; \mathcal{Q}_k \leftarrow \mathcal{Q}'_k$;
            - **end**
          - **end**
        - **end**
      - **end**
    - **end**
  - **end**
- **end**

until *for 100 iterations*;

Algorithm 1: A local search algorithm for computing the efficiency gap

3.2 Map Visualization

The map visualization has been implemented with R [22], a programming language and free software environment for statistical computing and graphics. R has several packages able to create a data frame for a map data suitable for plotting with specific functions:

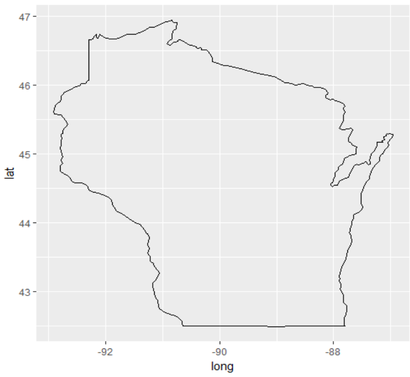
- **maps** [6]: this package contains many outlines to display continents, states and counties.
- **mapdata** [5]: supplement of maps package with a larger and higher resolution databases.
- **ggmap** [13]: this package contains a collection of functions to represent and visualize spatial data and models on the top of static maps from different online sources such as Google Maps.
- **dplyr** [29]: it provides data manipulation functions that can be applied to data frames.
- **ggplot** [28]: once data are provided, this package create graphics by mapping variables to aesthetics, deciding which graphical primitives to use and taking care of details.

Supplementary material shows a portion of R code on how to create a data frame from a map data suitable for plotting with 'ggplot'.
The function *map_data()* returns a sequence of points along an outline into a data frame of those points. Then, the function *subset()* restricts the possible counties selected by the previous function into those ones of the considered state ('state_name'). The generated file ('state_name.csv'), that will be used as input to draw the map, has the following format:

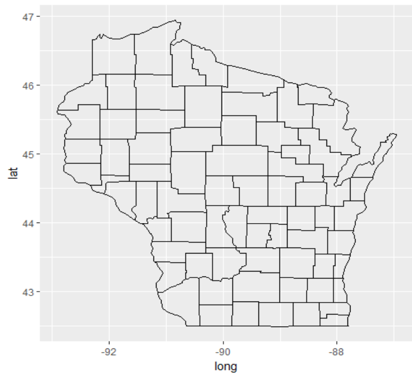
- *coordinate_ID*: this number represents the order in which ggplot should connect the points.
- *long*: longitude. The value is negative if the coordinate is on the west of the prime meridian.
- *lat*: latitude.
- *group*: this argument is essential for ggplot. It controls if adjacent points should be connected by lines. Only points in the same group can be connected.
- *order*: see *coordinate_ID*.

- *region*: region surrounded by a set of points.
- *subregion*: subregion surrounded by a set of points.

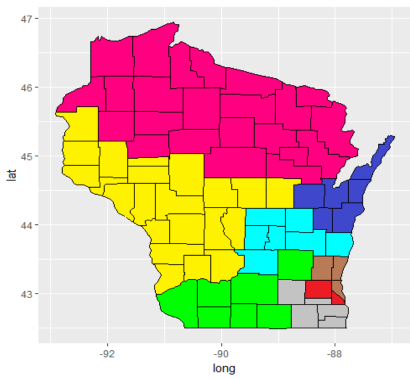
One extra field, *district*, is added to the output file and represents the district number associated with the county (or to the portion of county if it is shared among more districts) and which is assigned to a specific color.
Once the data frame is ready, the function *ggplot()* maps variables in the data to visualize properties of geometry and *geom_polygon()* finds start and end points of the given coordinates, connects and fills them with the right color according to the district number (Figure 3.1).



(a) Map(State)



(b) Map(Counties)



(c) Map(District color)

Figure 3.1: Output Map Visualization

Chapter 4

Results

This chapter presents the results obtained by testing the local search algorithm on four congress maps and related data for the 2012 House of Representatives elections for the US states of Wisconsin [20], Pennsylvania [17], Texas [18] and Virginia [19] N. Stephanopoulos and E. McGhee, in [23], analyzed the State House elections with at least eight congressional districts from 1972 to 2012. They concluded that, especially in 2012, the efficiency gaps turned out to favor the Republican party. Moreover, in that year, there was the first congressional election which used districts drawn in the 2010 US census and those redistricting plans evidence more extreme partisan gerrymandering than any other decade in modern American history. After 2012 elections, according to *Sunday's New York Times*, "Democrats received 1.4 million more votes for the House of Representatives but Republicans won control of the House by a 234 to 201 margin" [27].
Next sections describe, for each state, a table with comparisons between original data and new results in terms of efficiency gap measure and number of seats. In particular, every table has the following structure:

Total votes: total number of votes considering a two-party system.

Current EG: normalized efficiency gap based on the original elections results.

Dem votes: total number of Democratic votes.

Rep votes: total number of Republican votes.

Old D_seats: original number of Democratic seats.

Old R_seats: original number of Republican seats.

New EG: normalized efficiency gap obtained after applying the local search algorithm.

New D_seats: new number of Democratic seats after applying the local search algorithm.

New R_seats: new number of Republican seats after applying the local search algorithm.

Moreover, for every state, the original map is compared to the new one after the algorithm has been tested.

4.1 Wisconsin

As the Wisconsin case, *Gill v. Whitford* [4], addressed the constitutionality of partisan gerrymandering, the first analysis was done on this state. Table 4.1 shows that Wisconsin is highly gerrymandered as the current efficiency gap is close to 15% and, despite the fact that Democrats won almost 51% of total votes, the Republican party holds 5 of the 8 congressional seats.

Table 4.1: WISCONSIN HOUSE ELECTIONS 2012

Total votes	2,841,407
Current EG	0.148
Dem votes 1,441,804	**Rep votes** 1,399,603
Old D_seats 3	**Old R_seats** 5
New EG	0.038
New D_seats 3	**New R_seats** 5

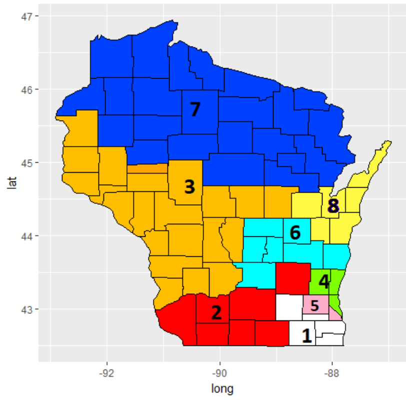
After testing the algorithm on these data, the minimum efficiency gap found was 3.8% (below the threshold) and the number of gained seats did not change.
Figure 4.1 compares the original congress map with the one produced with the new district configuration of the new results.
A closer look at the new data and at the new district map reveals the following interesting insights:

Partisan symmetry vs. efficiency gap. Lowering the efficiency gap from 14.76% to 3.80% did not affect the total seat allocation (3 democrats vs. 5

republicans) between the two parties. This further reinforces the assertion in [23] that the partisan symmetry is a different concept not related to the efficiency gap and thus fewer absolute difference of wasted votes does not necessarily lead to seat gains for loosing party.

Compactness vs. efficiency gap. Even though the value of the efficiency gap is not related to district compactness, the new district map for Wisconsin reveals some interesting aspects. Districts that previously had an irregular and stranger shape became more compact (for example, district 7 and 3 in 4.1(b)). This happened because some counties, which were shared among more districts, were shifted by the algorithm to a single district and were not shared anymore. Moreover, some small districts such as district 4 became bigger and vice versa.

(a) Original congress map



(b) New congress map

Figure 4.1: Wisconsin

4.2 Pennsylvania

Pennsylvania is widely known for its gerrymandered districts. In particular, district 6, 7 and 16 (see the original congress map in Figure 4.2(a)) are the most severely gerrymandered districts according to compactness scores. Moreover, the elections results in the last decade show a high efficiency gap and an asymmetry between gained seats and votes. In February 2018, due to the 2016 Presidential elections results, the local Supreme Court released a new congressional map which comes very close to achieve partisan balance [16]. Due to all these recent updates on Pennsylvania, the local search algorithm was tested on the 2012 House of Representatives elections first and then on the 2016 Presidential ones to compare the new maps and to test the algorithm effectiveness.

4.2.1 2012 House of Representatives elections

Table 4.2 confirms a disparity between votes and seats: almost 52% of votes went to Democrats but they only won 5 of the 18 House seats, while Republicans won 13 seats with 48% of votes. From a high original efficiency gap of 23.76%, by switching a maximum number of counties ranged from 2 to 10, the final efficiency gap was 8.64% and it could not go below this value. In addition, while Wisconsin new results (Table 4.1) did not show any change in seats distribution, in this case Democrats gained one more seat for a total of 6 seats.

4.2.2 2016 Presidential elections

The 2016 Presidential elections results (Table 4.3) show an outcome similar to the previous one. Democrats won almost 50% of votes and only gained 6 seats of 18, while Republicans won 12 seats with slightly more than 50% of votes.
The efficiency gap, from an original value of 14.24% became 8.05% and, in the new district configuration, the Democrats gained one more seat for a total of 7 seats of 18.

Table 4.2: PA HOUSE ELECTIONS 2012

Total votes	5,374,461
Current EG	0.2380
Dem votes	**Rep votes**
2,722,560	2,651,901
Old D_seats	**Old R_seats**
5	13
New EG	0.0864
New D_seats	**New R_seats**
6	12

Table 4.3: PA PRESIDENTIAL ELECTIONS 2016

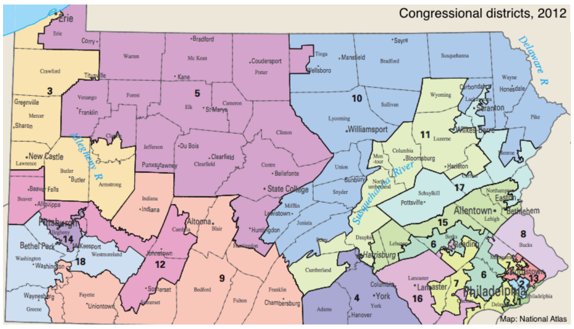
Total votes	5,896,628
Current EG	0.1434
Dem votes	**Rep votes**
2,925,776	2,970,852
Old D_seats	**Old R_seats**
6	12
New EG	0.0805
New D_seats	**New R_seats**
7	11

A deeper look at the results in both elections reveals additional interesting insights:

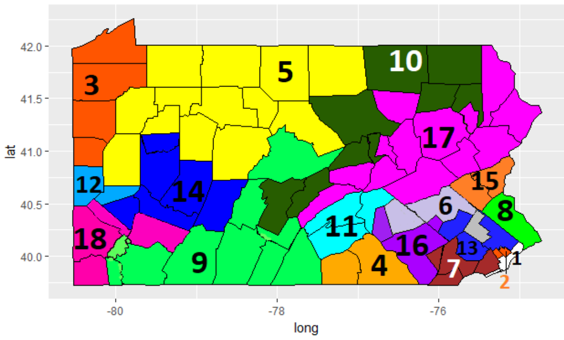
Efficiency Gap: threshold. With Wisconsin data it was easier to find an efficiency gap below the threshold, while with Pennsylvania it was not possible to go below 8.64% and 8.05% (respectively for 2012 and 2016 elections) as the algorithm rejected all solutions below the threshold due to inconsistent population or discontinuous districts. Despite this, as the original efficiency gap was very high and the initial map heavily gerrymandered, they can both be considered good results with a seat distribution a little closer to partisan balance.

Compactness of districts 6, 7, 16. As in the previous map, even if the efficiency gap is not related to compactness, some districts became more compact and many shared portions of county shifted in a single district (Figure 4.2(b) and Figure 4.2(c)).

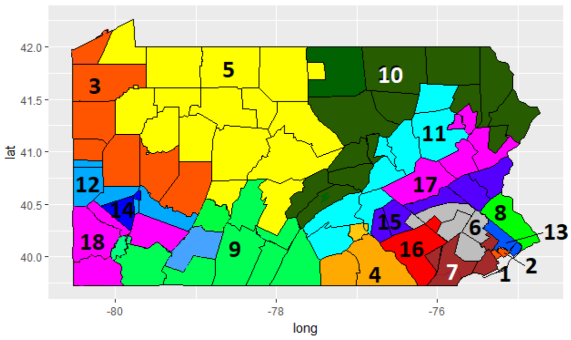
Comparison with the new actual map (2018). In the new map, drawn by Pennsylvania Supreme Court (see [16]) and based on partisan symmetry, some districts totally changed and were moved in different portions of the map. This allowed Democrats to gain 2 more seats, for a total of 8 seats, and the new efficiency gap became 3%.
The algorithm proposed in this project works in a different way, as it considers the efficiency gap as a metric for redistricting rather than partisan symmetry. Moreover, districts can't be totally moved in other portions of the map: their borders can drastically change as they become larger or smaller, but they still remain in the same area in the map. Even though the two algorithms are different, in both cases the Democratic party gains more seats and gets closer to partisan balance. It is also interesting to notice the increasing awareness in trying to defeat partisan gerrymandering in these years.



(a) Original congress map



(b) New congress map



(c) New map (Presidential elections)

Figure 4.2: Pennsylvania

4.3 Virginia

Virginia is ranked as one of the most gerrymandered states based on lack of compactness and contiguity of its districts.
Table 4.4 shows an original efficiency gap of 22%. Democrats won 3 seats of 11 House seats with almost 49% of total votes and Republicans won 8 seats with more or less 51% of total votes.
Testing the local search algorithm on the original Virginia congress map produced outstanding results. The final efficiency gap was lowered to as little as 3.6% and the Democrats totally gained 5 seats instead of 3.
This allowed us to make further considerations related to the efficiency gap metric:

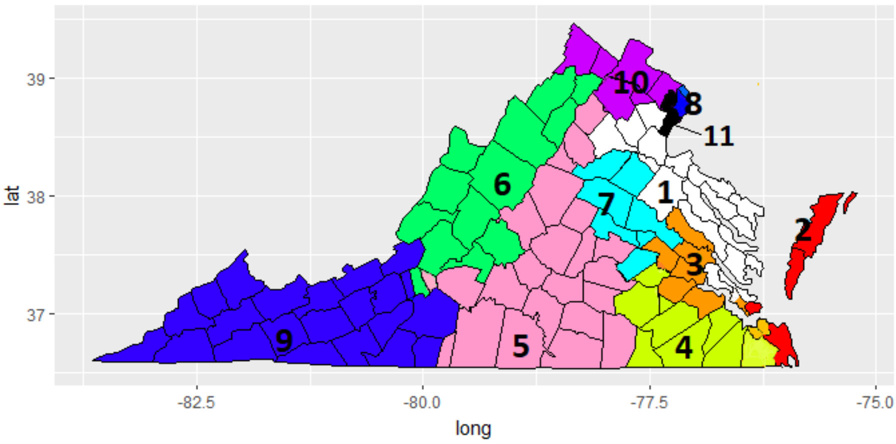
Table 4.4: VIRGINIA HOUSE ELECTIONS 2012

Total votes	3,569,498
Current EG	0.22
Dem votes	**Rep votes**
1,736,164	1,833,334
Old D_seats	**Old R_seats**
3	8
New EG	0.036
New D_seats	**New R_seats**
5	6

Partisan Symmetry e vs. Efficiency gap As previously highlighted there is no relationship between partisan symmetry and efficiency gap. From these experimental results, we can see how minimizing the wasted votes in some cases shows no change in the seat outcome (such as in Wisconsin) while in other there is a quite significant change which makes the new plan more balanced without intentionally looking at the partisan outcome.

Compactness and geographical irregularities . It is already known from the literature that sometimes irregular geographical shapes, such as shorelines, can make district less compacts and a trade-off is needed. Figure 4.3 shows some changes in districts which are inside these geographical irregularities (for example, district 2 and 3). Even if the shape still remains odd and apparently not very compact due to shorelines, it seems that they became less odd in some areas.

(a) Original congress map



(b) New congress map

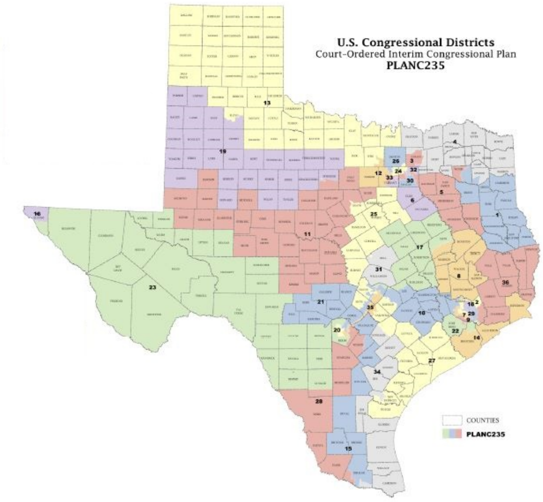
Figure 4.3: Virginia

4.4 Texas

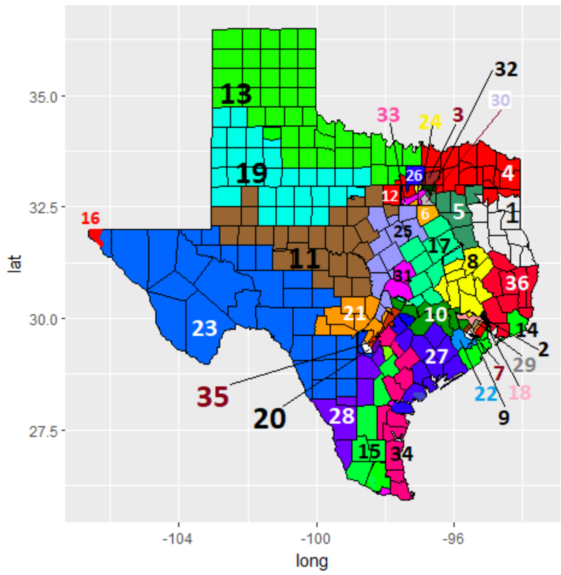
Texas is another state classified as one with the most gerrymandered districts according to compactness score and election results. Its congress district map was partially redrawn by a court prior to the 2012 elections and Table 4.5 shows the election outcome. These results display a seat gap in favor of Republicans between two to three seats. Democrats won 12 seats with almost 40% of votes and Republicans won 24 districts with 60% of votes. From an original efficiency gap of 4%, wasted votes has been reduced until a final efficiency gap of 3% and there was no change in the seat share.

Table 4.5: TEXAS HOUSE ELECTIONS 2012

Total votes	7,379,170
Current EG	0.041
Dem votes	**Rep votes**
2,949,900	4,429,270
Old D_seats	**Old R_seats**
12	24
New EG	0.033
New D_seats	**New R_seats**
12	24



(a) Original congress map



(b) New congress map

Figure 4.4: Texas

Chapter 5

Conclusion

This research project, to the best of our knowledge, is the first algorithmic analysis and implementation of minimization of the efficiency gap measure. It shows that redrawing district maps to remove gerrymandering is practically feasible and can be done in a small amount of time. Moreover, it reveals new aspects of the efficiency gap related to redistricting, such as its relations with partisan symmetry and compactness.

Since this algorithm applies a sequence of carefully chosen semi-random perturbations to the original gerrymandered maps to drastically lower the absolute difference of wasted votes, one could hypothesize that the original gerrymandered districts are far from being the product of random decisions. However, to reach a definitive conclusion regarding this point, one would need to construct a suitable null model, which we do not have yet.

If the US Supreme Court will accept the lower court case in Wisconsin, our work can be considered a valuable and necessary tool to remove partisan gerrymandering in the US. Moreover, some ideas for future researches can be trying to minimize wasted votes by using simulated annealing. In some heavily gerrymandered states bringing down the efficiency gap can be more difficult and the value can be stuck in a local minimum. Simulated Annealing can be useful to avoid these scenarios.

Beyond its scientific implications, we expect this algorithmic analysis and results to have a beneficial impact on the US judicial system. Some justices, whether at the Supreme Court level or in lower courts, seem to have a reluctance to taking mathematics, statistics and computing seriously. Our theoretical and computational results show that the math, whether complicated or not (depending on one's background), can in fact yield fast accurate computational methods that can indeed be applied to un-gerrymander the currently gerrymandered maps.

Appendix A

Formalization of the Optimization Problem

Based on [9, 23], we abstract our problem in the following manner. We are given a rectilinear polygon $\mathcal{P}$ without holes. Placing $\mathcal{P}$ on a unit grid of size m x n, we identify an individual unit square, a *cell*, on the i^{th} row and j^{th} column in $\mathcal{P}$ by $p_{i,j}$ for $0 \leq i \leq m$ and $0 \leq j \leq n$ (see A.1). For each cell $p_{i,j} \epsilon \mathcal{P}$, we are given the following three integers:

- an integer $Pop_{i,j} \geq 0$ (total population inside the cell) and
- two integers $PartyA_{i,j}$, $PartyB_{i,j} \geq 0$, so that $PartyA_{i,j} + PartyB_{i,j} = Pop_{i,j}$.

Note that $PartyA_{i,j}$ and $PartyB_{i,j}$ represent the total number of votes for both parties in $p_{i,j} \epsilon \mathcal{P}$ and the "size" (number of cells) of $\mathcal{P}$ is denoted as $|\mathcal{P}| = |p_{i,j} : p_{i,j} \epsilon \mathcal{P}|$. For any rectilinear polygon $\mathcal{Q}$ included in $\mathcal{P}$ (as a connected subset of the interior of $\mathcal{P}$), we define the following quantities:

- **Party affiliation in $\mathcal{Q}$:**
  $PartyA(\mathcal{Q}) = \sum_{p_{i,j} \epsilon \mathcal{Q}} PartyA_{i,j}$ and $PartyB(\mathcal{Q}) = \sum_{p_{i,j} \epsilon \mathcal{Q}} PartyB_{i,j}$.
- **Population of $\mathcal{Q}$:** $Pop(\mathcal{Q}) = PartyA(\mathcal{Q}) + PartyB(\mathcal{Q})$.
- **Efficiency Gap of $\mathcal{Q}$:**

  $$Eg(\mathcal{Q}) = \begin{cases} (PartyA(\mathcal{Q}) - \frac{1}{2}Pop(\mathcal{Q})) - PartyB(\mathcal{Q}) = 2PartyA(\mathcal{Q}) - \frac{3}{2}Pop(\mathcal{Q}) \\ PartyA(\mathcal{Q}) - (PartyB(\mathcal{Q}) - \frac{1}{2}Pop(\mathcal{Q})) = 2PartyA(\mathcal{Q}) - \frac{1}{2}Pop(\mathcal{Q}) \end{cases}$$

  In the first result, $PartyA(\mathcal{Q}) \geq \frac{1}{2}Pop(\mathcal{Q})$ and vice versa in the second one. Note that if $PartyA(\mathcal{Q}) = PartyB(\mathcal{Q}) = \frac{1}{2}Pop(\mathcal{Q})$, we assume that $PartyA$ is the winner. Also, note that $Eg(\mathcal{Q}) = 0$ if and only if either $PartyA(\mathcal{Q}) = \frac{1}{4}Pop(\mathcal{Q})$ or $PartyB(\mathcal{Q}) = \frac{1}{4}Pop(\mathcal{Q})$.

Table A.1: INPUT POLYGON $\mathcal{P}$ OF SIZE 15 ON A GRID OF SIZE 6x4

0	1	2	3	4	5
1					
2					
3					

Our problem can be now defined as follows.

Problem name: κ-district Minimum Wasted Vote Problem (MIN-WVP$_\kappa$).

Input: a rectangular polygon $\mathcal{P}$ with $Pop_{i,j}$, $PartyA_{i,j}$, $PartyB_{i,j}$ for every cell $p_{i,j} \epsilon \mathcal{P}$, and a positive integer $1 \leq \kappa \leq |\mathcal{P}|$.

Definition: a κ-equipartition of $\mathcal{P}$ is a partition of the interior of $\mathcal{P}$ into exactly κ rectilinear polygons, say $Q_1, ..., Q_\kappa$, such that $Pop(Q_1) = ... = Pop(Q_\kappa)$.

Assumption: $\mathcal{P}$ has at least one κ-equipartition.

Valid solution: any κ-equipartition $Q_1, ..., Q_\kappa$ of $\mathcal{P}$.

Objective: minimize the total absolute efficiency gap $Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa) = |\sum_{j=1}^{\kappa} Eg(\mathcal{Q}_j)|$.

Notation: OPT$_\kappa(\mathcal{P})$=min ($Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa)|Q_1, ..., Q_\kappa$ is a κ-equipartition of $\mathcal{P}$).

Mathematical Properties of the Efficiency Gap

The following lemma defines the set of possible rational number that the efficiency gap of a κ-equipartition can take. Suppose that we partition the polygon $\mathcal{P}$ into κ=2 regions, then (depending on which party won in each district):

$$Eg(\mathcal{P}, \mathcal{Q}_1, \mathcal{Q}_2) = \begin{cases} |2PartyA(\mathcal{P}) - \frac{3}{2}Pop(\mathcal{P})| \\ |2PartyA(\mathcal{P}) - \frac{1}{2}Pop(\mathcal{P})| \\ |2PartyA(\mathcal{P}) - Pop(\mathcal{P})| \end{cases}$$

Lemma 1 *affirms that:*

(a) $\forall$ *κ-equipartition* $(\mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *of* $\mathcal{P}$, $Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *always assumes one of the* κ*+1 values of the form* $|2PartyA(\mathcal{P}) - (z + \frac{\kappa}{2})\frac{Pop(\mathcal{P})}{\kappa}|$ *for* $z = \{0,1, ..., \kappa\}$.

(b) *If* $Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa) = |2PartyA(\mathcal{P}) - (z + \frac{\kappa}{2})\frac{Pop(\mathcal{P})}{\kappa}|$ *for some* $z\epsilon\{0,1, ..., \kappa\}$ *and some κ-equipartition* $(\mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *of* $\mathcal{P}$, *then* $\frac{Pop(\mathcal{P})}{2\kappa}z \leq PartyA(\mathcal{P}) \leq (\frac{Pop(\mathcal{P})}{2\kappa}z + \frac{1}{2}Pop(\mathcal{P}))$.

Proof 1 *To prove **(a)** and **(b)**, consider any κ-equipartition* $(\mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *of* $\mathcal{P}$ *with* $Pop(Q_1) = ... = Pop(Q_\kappa) = \frac{1}{k}Pop(\mathcal{P})$. *Note that* $\forall$ $\mathcal{Q}_j$ *we have* $Eg(\mathcal{Q}_j) = 2PartyA\mathcal{Q}_j - r_jPop(\mathcal{Q})$ *where:*

$$r_j = \begin{cases} \frac{3}{2} & \text{if } PartyA(\mathcal{Q}) \geq \frac{Pop(\mathcal{P})}{2\kappa} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

Consider z as the number of r_j 's that are equal to $\frac{3}{2}$, which means the number of regions where PartyA has the majority, it follows that:
$Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa) = |\sum_{j=1}^{\kappa} Eg(\mathcal{Q}_j)| = |2PartyA(\mathcal{P}) - (\frac{3}{2}z + \frac{1}{2}(\kappa - z))\frac{Pop(\mathcal{P})}{\kappa}| = |2Party(\mathcal{P}) - (z + \frac{k}{2})\frac{Pop(\mathcal{P})}{\kappa}| = |2Party(\mathcal{P}) - (\frac{z}{\kappa} + \frac{1}{2})Pop(\mathcal{P})|$.
Moreover, since $0 \leq PartyA(\mathcal{Q}_j) \leq \frac{Pop(\mathcal{P})}{\kappa}$ $\forall j$, *we have that*

$$PartyA(\mathcal{P}) {=} \sum_{j=1}^{\kappa} PartyA\mathcal{Q}_j \geq \sum_{j:r_j=\frac{3}{2}} \frac{Pop(\mathcal{P})}{2\kappa} = \frac{Pop(\mathcal{P})}{2\kappa}z$$

and

$$PartyA(\mathcal{P}) {=} \sum_{j=1}^{\kappa} PartyA\mathcal{Q}_j \leq \sum_{j:r_j=\frac{3}{2}} \frac{Pop(\mathcal{P})}{\kappa} + \sum_{j:r_j=\frac{1}{2}} \frac{Pop(\mathcal{P})}{2\kappa}(\kappa - z) = \frac{Pop(\mathcal{P})}{2\kappa}z + \frac{1}{2}Pop(\mathcal{P})$$

Corollary 1 *Using the reverse triangle inequality of norms, the absolute difference between two successive values of* $Eg(\mathcal{P}, \mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *is given by:*
$||2Party(\mathcal{P}) - (\frac{z}{\kappa} + \frac{1}{2})Pop(\mathcal{P})| - |2Party(\mathcal{P}) - (\frac{z+1}{\kappa} + \frac{1}{2})Pop(\mathcal{P})|| \leq |(2Party(\mathcal{P}) - (\frac{z}{\kappa} + \frac{1}{2})Pop(\mathcal{P})) - (2Party(\mathcal{P}) - (\frac{z+1}{\kappa} + \frac{1}{2})Pop(\mathcal{P}))| = \frac{Pop(\mathcal{P})}{\kappa}$

Corollary 2 $\forall$ *κ-equipartition* $(\mathcal{Q}_1, ..., \mathcal{Q}_\kappa)$ *of* $\mathcal{P}$, *consider the following quantities as defined in [23] (p.853).*

Normilized seat margin of PartyA: $\frac{|\mathcal{Q}_j : PartyA(\mathcal{Q}_j) \geq \frac{1}{2}Pop(\mathcal{Q}_j)|}{\kappa} - \frac{1}{2}$

Normilized vote margin of PartyA: $\frac{PartyA(\mathcal{P})}{Pop(\mathcal{P})} - \frac{1}{2}$

Then we can write the normalized efficiency gap as

$$\frac{|2Party(\mathcal{P})-(z+\frac{k}{2})\frac{Pop(\mathcal{P})}{\kappa}|}{Pop(\mathcal{P})}$$

which becomes

$$2(\frac{PartyA(\mathcal{P})}{Pop(\mathcal{P})}-\frac{1}{2})-(\frac{z}{k}-\frac{1}{2})$$

This allows us to identify the measure with the previous definitions:

$$\frac{Eg(\mathcal{P},\mathcal{Q}_1,\ldots,\mathcal{Q}_\kappa)}{Pop(\mathcal{P})}=|2\times(\text{ Vote margin of } PartyA)-(\text{ Seat margin of } PartyA)|$$

Approximation Hardness Result for MIN-WVP$_\kappa$

Recall that, for any $\rho \geq 1$, an approximation algorithm with an approximation ratio of ρ (or, simply an ρ-approximation) is a polynomial-time solution of value at most ρ times the value of an optimal solution [11].

Theorem 1 *Assuming $P \neq NP$, for any rational constant $\epsilon \in (0,1)$, the MIN-WVP$_\kappa$ problem for a rectilinear polygon $\mathcal{P}$ does not admit a ρ-approximation algorithm for any ρ and all $2 \leq \kappa \leq \epsilon|\mathcal{P}|$.*

Remark 1 *Since the PARTITION problem is not a strongly NP-complete problem (*i.e.*, admits a pseudo-polynomial time solution), the approximation-hardness result in Theorem 1 does not hold if the total population $Pop(\mathcal{P})$ is polynomial in $|\mathcal{P}|$ (* i.e.*, if $Pop(\mathcal{P}) = O(|Pop(\mathcal{P})|^c)$ for some positive constant c). Indeed, if $Pop(\mathcal{P})$ is polynomial in $|\mathcal{P}|$ then it is easy to design a polynomial-time exact solution via dynamic programming for those instances of MIN-WVP$_\kappa$ problem that appear in the proof of Theorem 1.*

Proof 2 *The NP-complete PARTITION problem [11] is defined as follows: given a set of n positive integers $\mathcal{A} = a_0, \ldots, a_{n-1}$, decide if there exists a subset $\mathcal{A}' \subset \mathcal{A}$ such that $\sum_{a_i \in \mathcal{A}'} a_i = \sum_{a_j \notin \mathcal{A}'} a_j$. We can assume without loss of generality that n is sufficiently large and that each of $a_0, \ldots, a_{n-1}$ is a multiple of any fixed positive integer. For notational convenience, let $\Delta = \sum_{j=0}^{n-1} a_j$.*
Let $\mu \geq 0$ such that $\kappa = 2 + \mu n$ (we will later clarify the maximum upper bound for μ). Our rectangular polygon $\mathcal{P}$ (as shown in A.1(a), consists of a rectangle $\mathcal{C} = \{p_{i,j} | 0 \leq i \leq n, 0 \leq j \leq 2\}$ of size $3 \times (n+1)$ with additional μn cells attached in an arbitrary manner so that the whole polygon $\mathcal{P}$ is fully connected with no holes. For convenience, let $\mathcal{D} = \{p_{i,j} | p_{i,j} \notin \mathcal{C}\}$ be the set of those additional cells. The relevant numbers for each cell are as follows:

$$Pop_{i,j} = \begin{cases} \frac{\Delta}{2} & \textit{if } i = j = 0 \textit{ or if } i = 0, j = 2 \\ a_{i,j} & \textit{if } i = 1 \textit{ and } j \leq n \\ \Delta & \textit{if } p_{i,j} \in \mathcal{D} \\ 0 & \textit{otherwise} \end{cases}$$

$$PartyA_{i,j} = \begin{cases} \frac{\Delta}{2} & if\ i = 0\ \ and\ j = 2 \\ \frac{a_{i,j}}{2} & if\ i = 1\ \ and\ j \leq n \\ \frac{3\Delta}{4} & if\ p_{i,j} \in \mathcal{D} \\ 0 & otherwise \end{cases}$$

First, we show how to select a rational constant μ such that any integer κ in the range $[\,2, \epsilon|\mathcal{P}|\,]$ can be realized. Assume that $\kappa = \epsilon'|\mathcal{P}| \in [\,2, \epsilon|\mathcal{P}|\,]$ for some ϵ'. Since $|\mathcal{P}| = 3(n+1) + \mu n$ the following calculations hold.

$$\kappa = 2 + \mu n = \epsilon'|\mathcal{P}| = \epsilon'(3(n+1) + \mu n) \;\equiv\; \mu = \frac{3\epsilon'(n+1) - 2}{(1-\epsilon')n} < \frac{4\epsilon'}{1-\epsilon'} < \frac{4\epsilon}{1-\epsilon}$$

Claim 1.1 *$Eg(p_{i,j}) = 0$ for each $p_{i,j} \in \mathcal{D}$, and moreover each $p_{i,j} \in \mathcal{D}$ must be a separate partition by itself in any κ-equipartition of $\mathcal{P}$.*

Proof 3 *By straightforward calculation, $Eg(p_{i,j}) = 2 \times \frac{3\Delta}{4} - \frac{3\Delta}{2} = 0$. Since $\kappa = 2 + \mu n$ and $Pop(\mathcal{P}) = \sum_{p_{i,j} \in \mathcal{P}} Pop_{i,j} = \Delta + \Delta + \mu n \Delta = (2 + \mu n)\Delta$, each partition in any κ-equipartition of $\mathcal{P}$ must have a population of $\frac{Pop(\mathcal{P})}{\kappa} = \Delta$ and thus each $p_{i,j} \in \mathcal{D}$ of population Δ must be a separate partition by itself.*

Using Claim 1.1 we can simply ignore all $p_{i,j} \in \mathcal{D}$ in the calculation of the efficiency gap of a valid solution of $\mathcal{P}$ and it follows that the total efficiency gap of a κ-equipartition of $\mathcal{P}$ is identical to that of a 2-equipartition of $\mathcal{C}$. A proof of the theorem then follows provided we prove the following two claims:

(soundness) *If the PARTITION problem does not have a solution then $OPT_2(\mathcal{C}) = \Delta$.*

(completeness) *If the PARTITION problem has a solution then $OPT_2(\mathcal{C}) = 0$.*

Proof of soundness *(refer to A.1(b)) Suppose that there exists a valid solution (for example, a 2-equipartition) $\mathcal{C}_1, \mathcal{C}_2$ of MIN-WVP$_2$ for $\mathcal{C}$ with $p_{0,0} \in \mathcal{C}_1, p_{0,2} \in \mathcal{C}_2$, and let $\mathcal{A}' = \{a_j \,|\, p_{1,j} \in \mathcal{C}_1\}$. Then,*

$$\Delta = \frac{Pop(\mathcal{C})}{2} = Pop_{0,0} + \sum_{p_{1,j} \in \mathcal{C}_1} Pop_{1,j} = \frac{\Delta}{2} + \sum_{a_j \in \mathcal{A}'} a_j \;\equiv\; \sum_{a_j \in \mathcal{A}'} a_j = \frac{\Delta}{2}$$

and thus $\mathcal{A}'$ is a valid solution of PARTITION, a contradiction!
Thus, assume that both $p_{0,0}$ and $p_{0,2}$ belong to the same partition, say $\mathcal{C}_1$. Then, since $Pop_{0,0} + Pop_{0,2} = \Delta = \frac{Pop(\mathcal{C})}{2}$, every $p_{1,j}$ must belong to $\mathcal{C}_2$. Moreover, every $p_{i,j} \in \mathcal{C}$ with $Pop_{i,j} = 0$ must belong to $\mathcal{C}_1$ since otherwise $\mathcal{C}_1$ will not *be a connected region. This provides $Pop(\mathcal{C}_1) = Pop(\mathcal{C}_2) = \Delta$, showing that $\mathcal{C}_1, \mathcal{C}_2$ is indeed a valid solution (for example, a 2-equipartition) of MIN-WVP$_2$ for $\mathcal{C}$.*
The total efficiency gap of this solution can be calculated as

$$
\begin{aligned}
&Eg_2(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2) \\
&\qquad = |\, Eg(\mathcal{C}_1) + Eg(\mathcal{C}_2)\,| \\
&\qquad = \left| 2\, PartyA(\mathcal{C}_1) - \frac{3}{2} PartyA(\mathcal{C}_1) + 2\, PartyA(\mathcal{C}_1) - \frac{3}{2} PartyA(\mathcal{C}_1) \right| \\
&\qquad = \left| 2\frac{\Delta}{2} - \frac{3}{2}\Delta + 2\frac{\Delta}{2} - \frac{3}{2}\Delta \right| \\
&\qquad = \Delta
\end{aligned}
$$

Proof of completeness *(refer to A.1(c)) Suppose that there is a valid solution of $\mathcal{A}' \subset \mathcal{A}$ of PARTITION and consider the two polygons*

$$
\mathcal{C}_1 = \{p_{2,j} \,|\, 0 \le j \le n\} \cup \{p_{1,j} \,|\, a_j \in \mathcal{A}'\}\,, \quad \mathcal{C}_2 = \mathcal{C} \setminus \mathcal{C}_1
$$

By straightforward calculation, it is easy to verify the following:

- $Pop(\mathcal{C}_1) = \sum_{a_j \in \mathcal{A}'} a_j + \sum_{j=0}^{n} Pop_{2,j} = \Delta$, $Pop(\mathcal{C}_2) = \sum_{a_j \notin \mathcal{A}'} a_j + \sum_{j=0}^{n} Pop_{2,j} = \Delta$, *and thus $\mathcal{C}_1, \mathcal{C}_2$ is a valid solution (for example, a 2-equipartition) of MIN-WVP_2 for $\mathcal{C}$.*

- $Eg_2(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2) = OPT_2(\mathcal{C}) = 0$ *since*

$$
\begin{aligned}
Eg_2(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2) &= |\, Eg(\mathcal{C}_1) + Eg(\mathcal{C}_2)\,| \\
&= \left| 2\, PartyA(\mathcal{C}_1) - \frac{3}{2} Pop(\mathcal{C}_1) + 2\, PartyA(\mathcal{C}_1) - \frac{1}{2} Pop(\mathcal{C}_1) \right| \\
&\qquad = \left| 2\left(\frac{\Delta}{2} + \frac{\Delta}{4}\right) - \frac{3}{2}\Delta + 2\frac{\Delta}{4} - \frac{1}{2}\Delta \right| = 0
\end{aligned}
$$

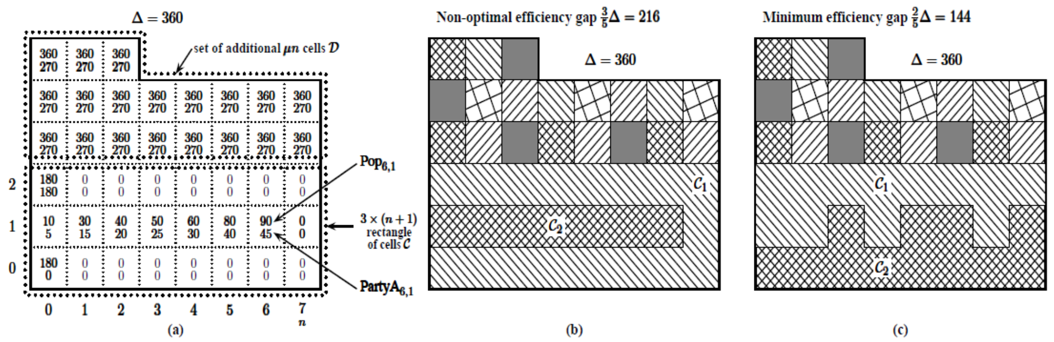


Figure A.1: An illustration of proof of Theorem 1

Appendix B

Algorithm and Map Visualization

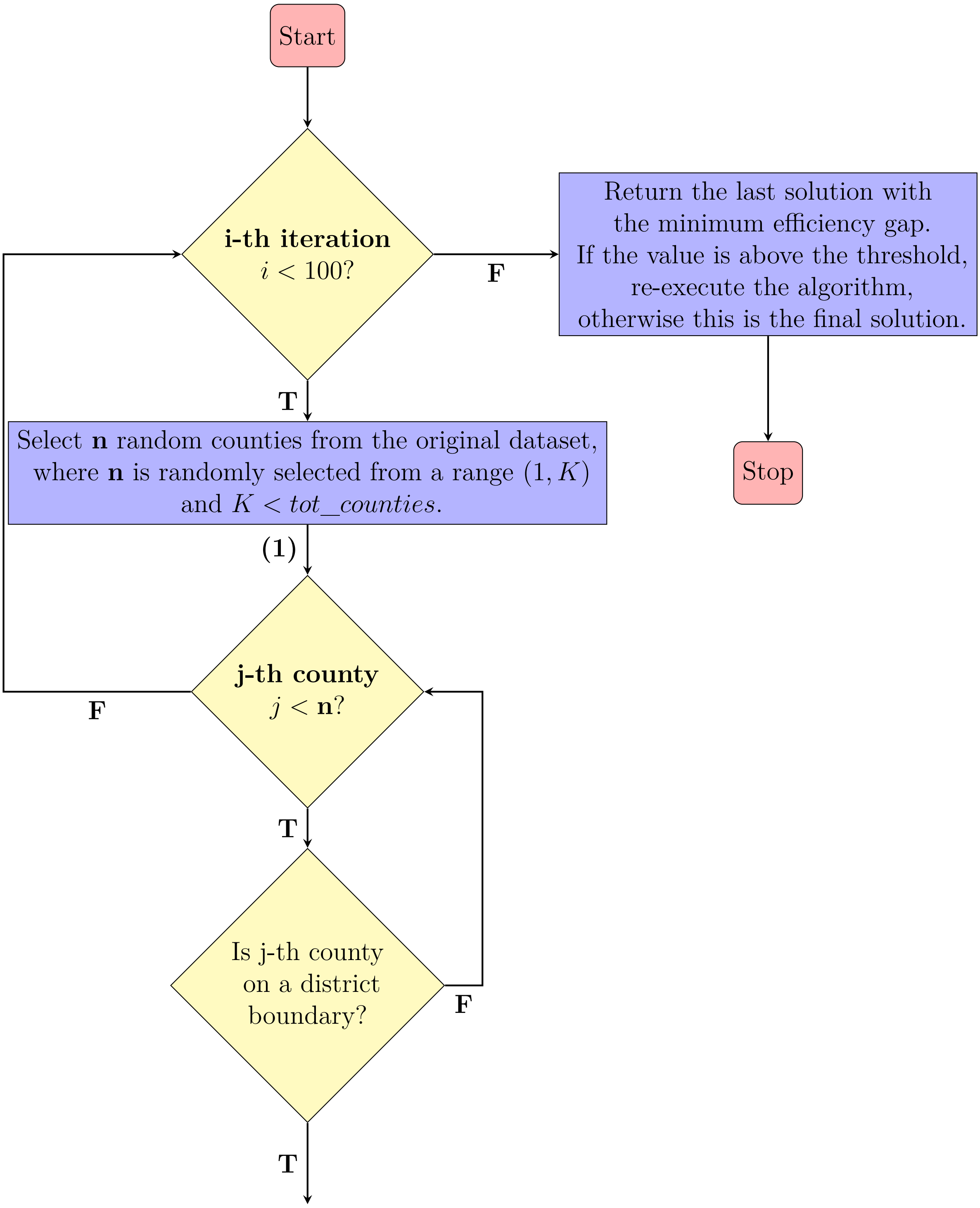
Creation of a data frame to plot map data

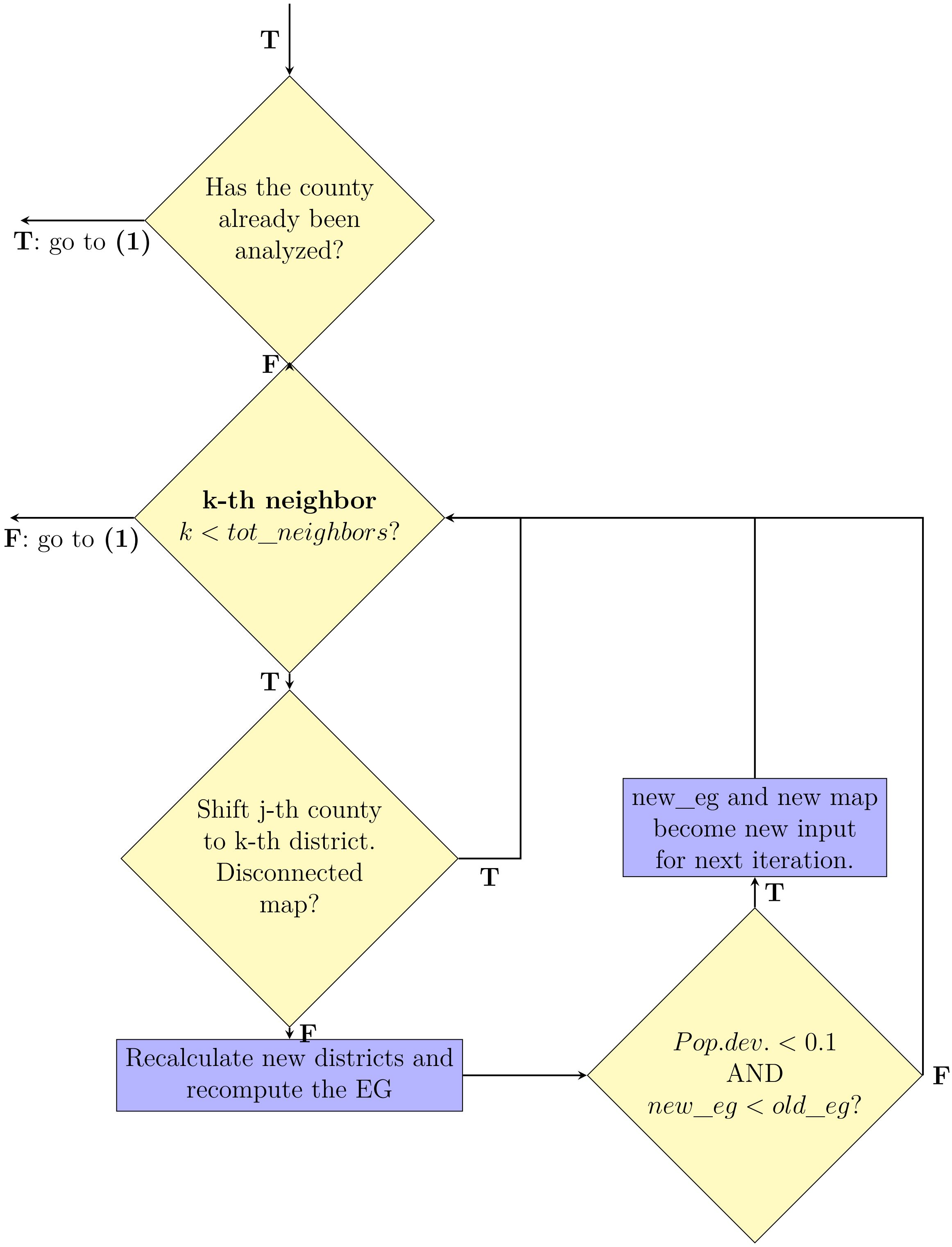
```
library(maps)
library(mapdata)
library(ggmap)
library(dplyr)

counties ← map_data("county");
state_county ← subset(counties, region=="state_name");
write.csv(state_county, "state_name.csv");
state_county ← read.csv("state_name.csv", header=T);
colors=rainbow(#districts);

%Creation of input files for shared counties
%Associate a color to each district number
state_map ← ggplot(data=state_county,mapping=aes(x=long,y=lat,group=
    group)) +
geom_polygon(color="black", fill=colors[state_county$colorBuckets]) +
geom_polygon(color="black", fill=NA) +
geom_polygon(data=shared_county1, aes(x=long, y=lat, group=group),
fill="district_color", color="black") +
geom_polygon(data=shared_county2, aes(x=long, y=lat, group=group),
fill="district_color", color="black") +
... +
coord_fixed(1.3);
```

Figure B.1: Local search algorithm




T
Has the county already been analyzed?
T: go to (1)
F
k-th neighbor
$k < tot\_neighbors$?
F: go to (1)
T
Shift j-th county to k-th district. Disconnected map?
T
F
Recalculate new districts and recompute the EG
new_eg and new map become new input for next iteration.
T
$Pop.dev. < 0.1$
AND
$new\_eg < old\_eg$?
F

Bibliography

[1] *Davis vs. Bandemer*. 478 US 109, 1986.

[2] *Vieth v. Jubelirer*. 541 US 267, 279-80, 2004.

[3] *League of United Latin American Citiziens v. Perry*. 548 US 399, 2006.

[4] *Gill v. Whitford*. US Supreme Court docket no 16-1161, decision pending, 2017.

[5] Richard A. Becker, Allan R. Wilks, and Ray Brownrigg. *mapdata: Extra Map Databases*, r package version 2.2-6 edition, 2016. https://CRAN.R-project.org/package=mapdata.

[6] Richard A. Becker, Allan R. Wilks, Ray Brownrigg, Thomas P Minka, and Alex Deckmyn. *maps: Draw Geographical Maps*, r package version 3.1.1 edition, 2016. https://CRAN.R-project.org/package=maps.

[7] Jowei Chen and Jonathan Rodden. Cutting through the thicket: redistricting simulations and the detection of partisan gerrymanders. *Election Law Journal: Rules, Politics, and Policy*, 14(4):331–345, 2015.

[8] Editors Of The American Heritage Dictionaries and American Heritage Dictionary Editors. *The American Heritage Dictionary of the English Language*. . Series. Houghton Mifflin, 2006. ISBN 9780618701735, https://books.google.com/books?id=0VJPGgAACAAJ.

[9] McGhee Eric. Measuring partisan bias in single member district electoral systems. *Legislative Studies Quarterly*, 39:55–85, 2014.

[10] Nagle John F. Measures of partisan bias for legislating fair elections. *Election Law Journal: Rules, Politics, and Policy*, 14(4):346–360, 2015. https://www.liebertpub.com/doi/abs/10.1089/elj.2015.0311.

[11] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990. ISBN 0716710455.

[12] Emily Marshall Jonathan K. Hodge and Geoff Patterson. Gerrymandering and convexity. *The College Mathematics Journal*, 41(4):312–324, 2010. http://www.jstor.org/stable/10.4169/074683410x510317.

[13] David Kahle and Hadley Wickham. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

[14] Justin Levitt. All about redistricting: Professor justin levitt's guide to drawing the electoral lines. http://redistricting.lls.edu/where.php [Online; accessed 02/20/2018].

[15] John Mackenzie. Gerrymandering and legislator efficiency. Technical report, College of Agricultural Sciences University of Delaware, Newark, DE 19717, 2009.

[16] Matthew Bloch Nate Cohn and Kevin Quealy. The new pennsylvania congress map, district by district, 02 2018. https://www.nytimes.com/interactive/2018/02/19/upshot/pennsylvania-new-house-districts-gerrymandering.html [Online; accessed 03/01/2018].

[17] NBC NEWS.COM. Pennsylvania election results. http://elections.nbcnews.com/ns/politics/2012/Pennsylvania [Online; accessed 02/20/2018].

[18] NBC NEWS.COM. Texas election results. http://elections.sos.state.tx.us [Online; accessed 02/20/2018].

[19] NBC NEWS.COM. Virginia election results. http://elections.nbcnews.com/ns/politics/2012/Virginia [Online; accessed 02/20/2018].

[20] NBC NEWS.COM. Wisconsin election results. http://elections.nbcnews.com/ns/politics/2012/Wisconsin [Online; accessed 01/20/2018].

[21] Richard G. Niemi, Bernard Grofman, Carl Carlucci, and Thomas Hofeller. Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. *The Journal of Politics*, 52(4):1155–1181, 1990. https://doi.org/10.2307/2131686.

[22] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. ISBN 3-900051-07-0, https://www.R-project.org.

[23] Nicholas O. Stephanopoulos and Eric McGhee. Partisan gerrymandering and the efficiency gap. *University of Chicago Law Review*, 82:831–900, 03 2015.

[24] Wendy K. Tam Cho and Yan Y. Liu. Toward a talismanic redistricting tool: A computational method for identifying extreme redistricting plans. *Election Law Journal: Rules, Politics, and Policy*, 15(4):351–366, 12 2016.

[25] Z. Al-Quirashi L. Palmieri T.Chatterjee, B.DasGupta and A.Sidiropoulos. Alleviating partisan gerrymandering: can math and computers help to eliminate wasted votes? 2018.

[26] James D. Thoreson and John M. Liittschwager. Computers in behavioral science. legislative districting by computer simulation. *Behavioral Science*, 12(3):237–247, 1967. https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830120309.

[27] Sam Wang. The great gerrymander of 2012.

/urlhttps://archive.nytimes.com/www.nytimes.com/2013/02/03/opinion/sunday/the-great-gerrymander-of-2012.html[Online; accessed 18-April-2018].

[28] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2009. ISBN 978-0-387-98140-6, http://ggplot2.org.

[29] Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation.* R package version 0.5.0.

[30] Wikipedia contributors. Gerrymandering — Wikipedia, the free encyclopedia, 2018. https://en.wikipedia.org/w/index.php?title=Gerrymandering&oldid=836369686 [Online; accessed 18-April-2018].

[31] Wendy K. Tam Cho Yan Y. Liu and Shaowen Wang. Pear: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30:78 – 92, 2016. http://www.sciencedirect.com/science/article/pii/S2210650216300220.