Geometry of Gerrymandering

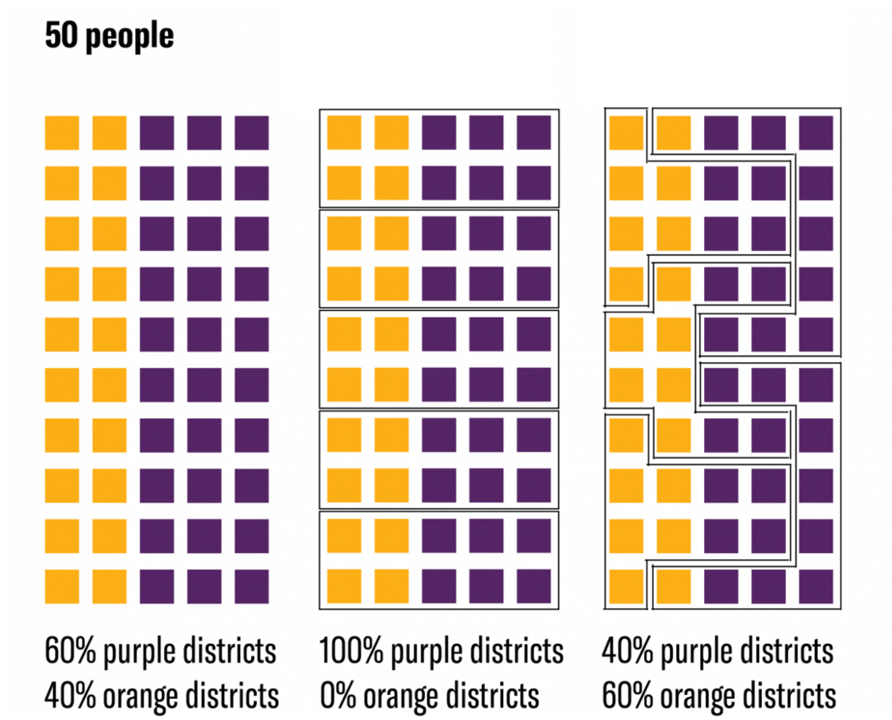
Arturo Ortiz San Miguel

April 22, 2021

1 Introduction

It is very hard for a government to control a whole country because people from different regions of the same country can be drastically different. Thus, governments often partition their territory into separate regions or *districts* and establish local governments for these regions to suit their local needs.

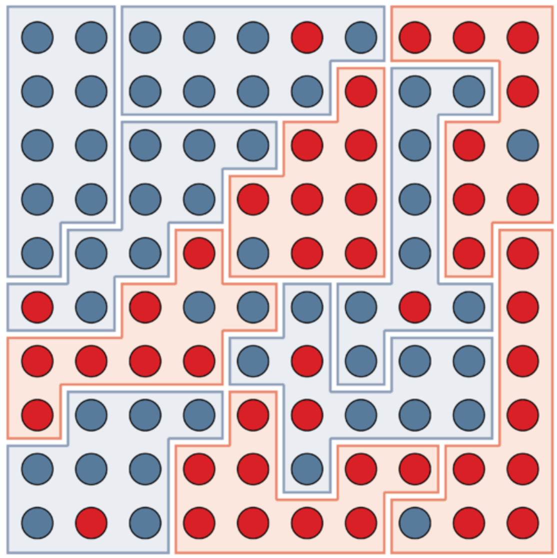
Additionally, these districts are used to place people in voting groups. Then, each district elects a representative whose beliefs align with the majority of the people in that district. This is best illustrated by an example.



Here, we separate the population into five districts with ten people each. We see that the election depends on how the borders are drawn.

Definition 1.1. *Gerrymandering* is the intentional redistricting of a state to gain a political advantage.

Traditionally, the word gerrymandering has a negative connotation. However, it is often necessary to gerrymandering to have a fair representative system. The following example shows how.



This region was partitioned in such a way that around 90% of all voters had a representative that aligned with the way they voted. In the real world, good gerrymandering can give minorities representation that they would not have otherwise. However, gerrymandering is commonly used to make sure the party in power has a better chance of winning the next election. This is considered unfair gerrymandering. This project is focused on using geometry to detect and avoid unfair gerrymandering.

The rest of this paper is structured as follows: in Section 2, we will discuss the case where the population is distributed equally throughout the whole region. Additionally, we will introduce theorems about fair partitioning. In Section 3, we will introduce a measure of compactness and explain why it is useful to detect potentially unfair gerrymandering. Finally, in Section 4, we will use these concepts and apply them to the representative districts of Puerto Rico.

2 The Simple Case

First, we will consider regions with constant population density. For fairness, we want all districts to have the same population. Then, since population is constant, all districts will have equal area. There are many ways to split a polygon into regions of equal area. An interesting (and seemingly fair) way to partition the polygon would be to have regions of equal perimeter.

Definition 2.1. A *fair bisector* is a line that partitions a polygon into two convex regions with equal area and perimeter.

Definition 2.2. Two points are called *opposite* if they are the endpoints of a line segment that partitions a polygon into two regions of equal area.

Thus, if we want to know which regions we can divide by using fair bisectors we must ask: which polygons have fair bisectors?

Theorem 2.1. Every convex polygon has a fair bisector.

Proof:
For any point $P \in \mathbb{R}^2$, on a convex polygon, there exists a point $Q \in \mathbb{R}^2$ opposite to it. Without loss of generality, we can assume that the region on the left, A, has a greater perimeter than the other, B. Then, we can move P by traversing through the polygon clockwise, and move Q such that they remain opposite until P is where Q was originally and vice-versa. Now, the region on the left is B and the one on the right is A. So B has a greater perimeter. Since, there is some pair of points, P, Q where perimeter(A) > perimeter(B), there another pair P', Q', such that perimeter(A) < perimeter(B), and the process moves both points continuously, there must be some pair of points, P, Q, on the polygon such that line going though both of them the polygon into two regions of equal area and perimeter.

□

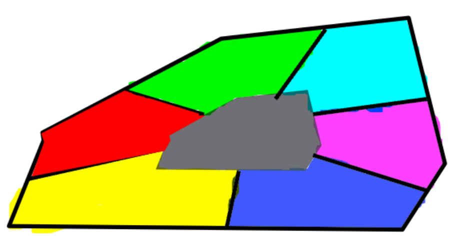
This proof can be generalized to prove the following statement:

Theorem 2.2. Every convex polygon can be fairly partitioned into n^2, $n \in \mathbb{N}$ regions with equal area and perimeter.

This is a good example of a theorem that can help us study how we can partition a country fairly. However, not all countries have a number of districts that is square, have convex boundaries, and have constant population density. We will now consider all polygons.

Lemma 2.1. Every polygon can be partitioned into N regions with equal area.

Proof Sketch:
Create a region of that is the same as the original polygon but scaled by $\frac{1}{N}$ and place its centroid on the centroid of the original polygon. Then split the perimeters of both into $N-1$ paths of equal length. Then the segments joining each path from n regions of equal area.



With this lemma, we can now prove a theorem about fairly partitioning into N regions.

Theorem 2.3. Every polygon can be fairly partitioned into N regions.

Proof:
Split the polygon into N regions of equal area. Now, we will add zigzags on the new edges to keep areas equal but change the perimeter. Adding zigzags will increase the perimeter of

both regions that share the edge. Add zigzags on all of the outer regions with the center except the one with the largest perimeter until they all have the same perimeter. Now there are two cases. If, the middle region has a larger perimeter than the rest, then add zigzags evenly on the edges not touching the middle region until the outer regions have the same perimeter as the inner one. Otherwise, keep adding zigzag evenly along the edges of the inner region until it has the same perimeter as the outer region.

□

We can apply this procedure to redistricting by using a weighted area based on population density to make regions of equal population. These would not have equal perimeter. However, this is not a problem because similarity in perimeter is not a requisite of fairness. Unfortunately, this approach does not take interests minorities, and other subjective issues into account. So, it must be modified.

.

3 Compactness

Since equal population, or weighted area, does not ensure fairness by itself, and perimeter is not a factor that determines fairness, we must use another measurement that is heavily correlated with fairness. This measurement is called *compactness*. It is defined in many ways. We will show the most commonly used.

Definition 3.1. The compactness, C, of a polygon is defined as

$$C = \frac{4\pi A}{P^2}$$

where A, P are the area and perimeter of the polygon respectively.

It can be visualized as the ratio of the area of the district and a circle.

A downside of this definition is that this circle seems meaningless. Another weakness this measure has is that it punishes polygons that seem compact but have jagged edges. However, it is the fastest to compute, which makes it highly used.

Definition 3.2. The compactness, C, of a polygon is the ratio between its area and the area of its smallest bounding circle.

Here, we can see it visualized.

This definition gives a very intuitive definition of compactness. We can see that if we squeeze the district on its long side, then the compactness lowers because the circle would be much smaller. However finding the smallest bounding circle takes longer than calculating area and perimeter.

Definition 3.3. The compactness, C, of a polygon is the ration of its area and the area of its convex hull.

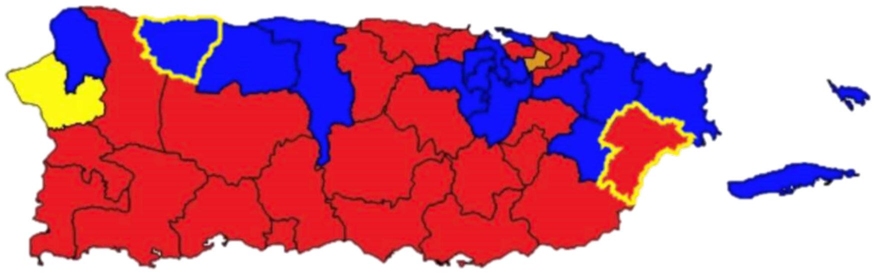
Similarly, this is a good definition intuitively, but takes longer to find than the first definition.

All three definitions give a number in the range $[0, 1]$. Low compactness is correlated with intentional and negative gerrymandering. This is because a lower compactness lowers the chance of someone in that district's neighbors are in the same district. This uses the accurate assumption that people are similar to their neighbors. So, we can highlight districts with low compactness. This dramatically decreases the number of districts people would have to check for unfair gerrymandering.

Additionally, we can make the current bordering of districts more fair by making the regions as compact as possible while still retaining a good amount of its original area. Many times, we would like to retain the core of a district's area because they are there for a reason, and fully changing districts would be a political nightmare.

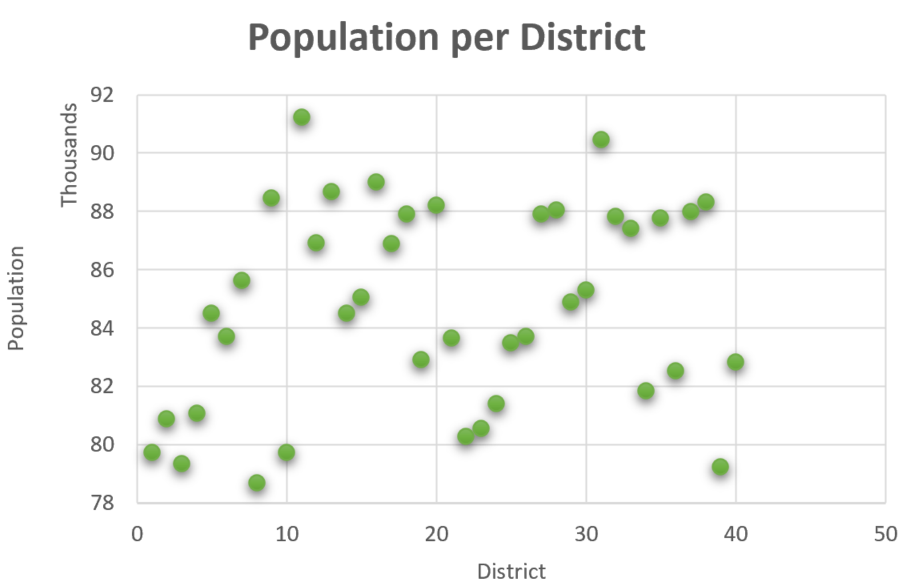
4 Puerto Rico

Puerto Rico is an archipelago in the Caribbean with around 3.2 million inhabitants. It is a colony of the United States of America. Much like the states, Puerto Rico has representative districts, where each district selects a representative to serve in its House of Representatives. There are 40 representative districts and 78 municipalities.



The rest of this paper will focus on finding each district's population and compactness. We will use the first definition of compactness from Section 3. Therefore, we will also find the perimeter of each district. Then, we will determine whether the current boundaries are fair and point out potentially unfairly gerrymandered areas.

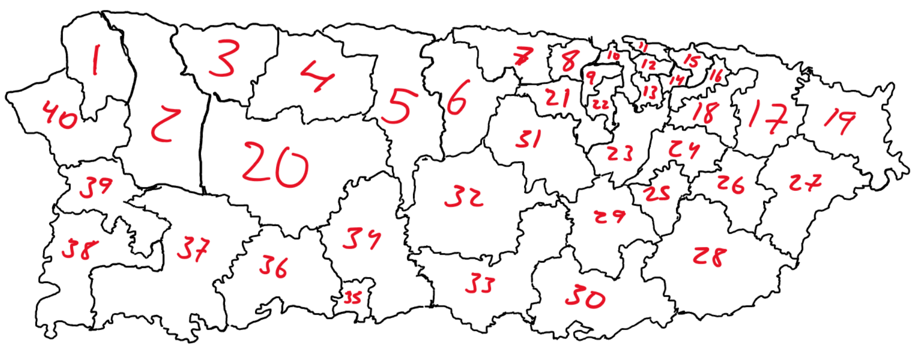
Using raw data from Instituto de Estadísticas de Puerto Rico's socio-economic profile of the representative districts from 2014 to 2018, we get the following information about the districts' populations.



Smallest District: 8 (78676)
Largest District: 11 (91211)
Average population: 84703
Standard Deviation: 3484.0 (4.11% deviation from the average)

The highest deviations from the average of both directions 7.60% and 7.11% are much higher than the standard deviation. We want to get as little deviation as possible so that every vote is worth the same. For example, a vote in District 11 is worth 0.86 votes in district 8. These districts are only 8.88 miles apart. Thus, there can be a lot of improvement on terms of even populations across districts.

There is no easily accessible source with the areas and perimeters of the districts. So, compactness was calculated by using computer assistance. First, an outline of Puerto Rico and its 40 districts was used.[1]



Then, each districts, outline was taken. For example here is district 38:

We find the perimeter by counting the number of black pixels in the image. This is because the black pixels are the outline of the district. Similarly, we can find the area by counting the number of white pixels between two black pixels in a row of pixels. The following is a

[1]Note: Numbers on the map are not the actual district numbers. These were used for easier computing. The later results have the correct district numbers.

script written in Python that will produce the compactness of all 40 districts.

```
import cv2

def rindex(mylist, myvalue):  # Helper function to get last instance of a
value in a list
    return len(mylist) - mylist[::-1].index(myvalue) - 1


for i in range(1, 40):
    # read the image file
    img = cv2.imread('Districts/Districts outline' + str(i) + '.PNG', 2)

    ret, bw_img = cv2.threshold(img, 127, 255, cv2.THRESH_BINARY)

    # converting to its binary form
    bw = cv2.threshold(img, 127, 255, cv2.THRESH_BINARY)

    perim = 0
    area = 0
    pix_count = 0
    # if the pixel is black, add one to the perimeter
    for col in bw_img:
        length = len(col)
        row = 0
        for pix in col:
            if pix == 0:
                perim = perim + 1
            pix_count = pix_count + 1

        if 0 in col: # get first black pixel and last black pixels' indices
            last = rindex(col.tolist(), 0)
            first = col.tolist().index(0)
            area = area + last - first
        else: # if no black pixel in row then area is unchanged
            last = 0
            first = 0
            area = area + last - first

    print(round(100 * area / (perim * perim), 2))
    print(round(area / pix_count, 2))
    print("\n")
```

This code returns two compactness values for each district. They are defined as follows:

Definition 4.1. The first compactness value, C_1, is defined to be

$$C_1 = \frac{100A}{P^2}$$

where A is the number of white pixels between two black pixels for every row. P is the number of black pixels in the image.

Since the thickness of the outlines in the images are constant, then is will give a ratio of area and the square of the perimeter. However, the units do not necessarily align because the thickness of the edges is unknown.

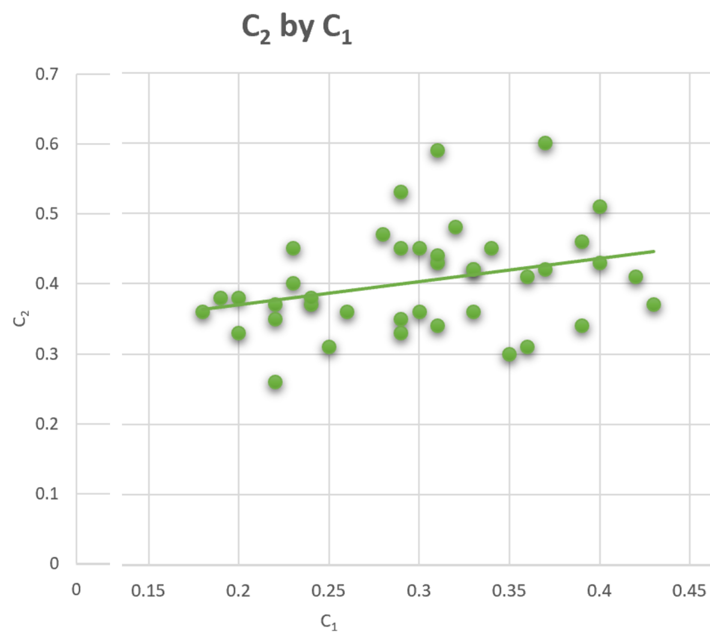
Definition 4.2. The second compactness value, C_2, is defined as the ratio of the area from the first compactness and the total number of pixels in the image.

Essentially, the number of pixels is equal to the area of the smallest bounding rectangle of the district.

The table below shows C_1, C_2 for all districts.

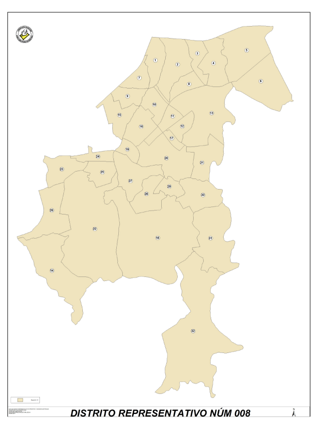
District	C_1	C_2
1	0.19	0.38
2	0.22	0.26
3	0.31	0.34
4	0.39	0.34
5	0.3	0.45
6	0.37	0.6
7	0.35	0.3
8	0.18	0.36
9	0.29	0.33
10	0.31	0.43
11	0.36	0.31
12	0.23	0.45
13	0.42	0.41
14	0.39	0.46
15	0.36	0.41
16	0.24	0.37
17	0.2	0.38
18	0.22	0.35
19	0.32	0.48
20	0.33	0.36
21	0.26	0.36
22	0.4	0.43
23	0.2	0.33
24	0.4	0.51
25	0.33	0.42
26	0.24	0.38
27	0.43	0.37
28	0.29	0.45
29	0.31	0.44
30	0.29	0.53
31	0.34	0.45
32	0.37	0.42
33	0.3	0.36
34	0.33	0.42
35	0.28	0.47
36	0.23	0.4
37	0.25	0.31
38	0.31	0.59
39	0.22	0.37
40	0.29	0.35

The correlation coefficient between C_1 and C_2 is 0.30, which means that there is a weak correlation between them.



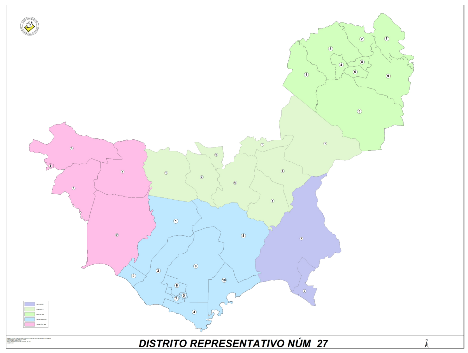
Now, we will qualitatively check the shape of some districts to check if the compactness measures are useful.

District 8 has the lowest C_1, which is 0.18. It is also the district with the lowest population.



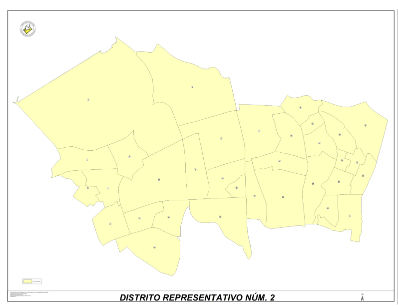
This district is thin, so it should have a low compactness number. Additionally, this region of district bordering splits the municipality of Guaynabo in an awkward way. Thus, this measure is promising so far.

Contrarily, District 27 has the highest C_1, 0.43.

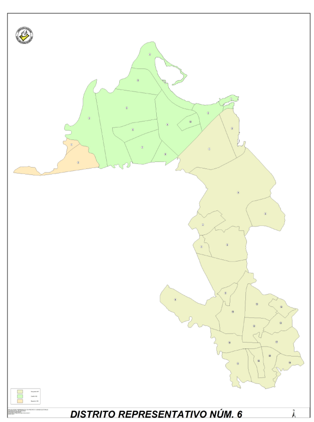


This district has a fairly big enclave in the northeast. However, it is quite thick. This should give a higher compactness value than a district with thinner enclaves. Although this does not look like the most compact district, most (if not all) districts that do not seem compact have a significantly lower C_1.

The district with the smallest C_2 is District 2 with 0.26.

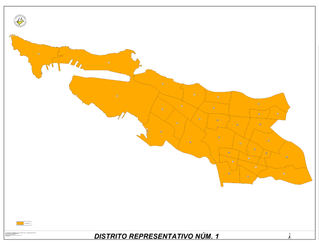


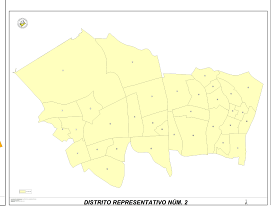
The largest C_2 was from District 6.

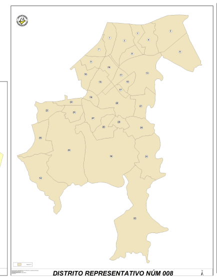


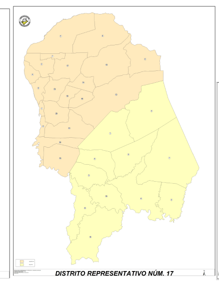
We see that District 2 should have a high compactness. Conversely, District 6 should have a low compactness. Furthermore, there seems to be no meaning between C_2 and the qualitative compactness of a district. This is possibly because the approximation that the number of pixels is equal to the area of the smallest bounding rectangle is bad. In reality, the number of pixels more accurately represents the area of the smallest bounding rectangle whose edges are horizontal and vertical. Therefore, using this version of C_2 to measure compactness is incorrect. However, a similar idea that uses the number of pixels in the smallest rectangle (not necessarily horizontal) should give good results.

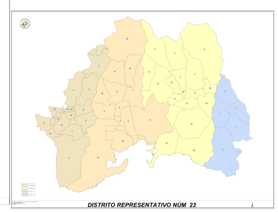
For these reasons we will use C_1 as a measure of compactness. This measure points out Districts 1, 2, 8, 17, and 23 as the least compact. Thus, these five districts have a higher chance of being unfairly divided and should be studied.











Although this data is not highly accurate, it is the only study of its kind and cannot be significantly improved without accessible measurements and data of the districts.

References

Bobonis, Luis E, et al. "2014-2018 Perfil Socio-Demográfico De Los Distritos Electorales De Puerto Rico Autorizado Por La Oficina Del Contralor Electoral." Instituto De Estadisticas De Puerto Rico, 17 Nov. 2020.

Mapas De Los Distritos Representativos Según La Redistribución Electoral De 2010., 2010.

Nandakumar, R., and N Ramana Rao. 'Fair' Partitions of Polygons - an Introduction. 2010.