# Evaluation of Fairness of the Bisection Protocol Algorithm for Political Redistricting by Examining Voter Populations

**Soujanya S. Duggirala**
**Sondra H. Plotnik-Sherry**
Columbia University

## Abstract

The drawing of electoral districts in the United States has long been criticized for being heavily gerrymandered. There have been multiple game theoretic algorithms introduced to approach this problem from a fair allocation perspective. A promising one, the Bisection Protocol, is able to avoid the issue of having a neutral party, is more efficient than its competitor ICYF, and follows a simple majority rule. It, and others, however, assumes a theoretical framework that is far removed from reality. This paper will explore this algorithm and challenge two of its core assumptions. First, that all citizens are voters, and second, its non-geometric setting. This algorithm is tested using the state of Georgia, which is now a swing state and arguably very gerrymandered. It was judged on its ability to fairly allocate votes across districts while being completely agnostic to voter turnout and voting age population. It was found that the Bisection Protocol was in fact successful in reducing voter turnout spread, which is an improvement in fairness compared to the current districting plan. This paper was not successful in creating completely contiguous districts but provides a theoretical framework for exploring this problem further. For real-world applications of political redistricting algorithms, the geometric setting of the data must be considered.

## 1 Introduction

### 1.1 Background

In the United States, elections are conducted on many different geographical levels, such as in federal, state, and local election districts. Different election districts are created on each level, and each election is determined by different "rules." Typically, the United States Congress and state legislature seats are determined by winning a popular vote in the district the seat represents. The election districts are usually drawn by the political party that is in power at the time of redistricting, using census and voting data over the last few election cycles. One exception is California, which has a nonpartisan redistricting committee to attempt to ensure fairness between the two political parties.

Gerrymandering is a technique that has been developed over the years for the political party in power to draw up districts such that they have an advantage in the upcoming election. It involves splitting up areas of a state which have similar political leanings in a manner where their votes overpower the votes of the opposition party in as many districts as possible. This technique is currently not illegal, but some states are moving away from it so that people in districts are better represented, rather than political parties "rigging the game" against the minority in a district.

One interesting case is the state of Georgia, which is heavily gerrymandered. The metropolitan Atlanta area is heavily Democratic, but the state as a whole tends to be more Republican. Four Democratic "safe seat" districts are allocated to the central Atlanta area, but the surrounding metropolitan Democratic areas are part of gerrymandered Republican districts. Thus, in recent U.S. House and Georgia State House and Senate elections in those gerrymandered districts, the Democratic candidates had immense backing from the constituents and were favored to win, but on election day, the Republicans narrowly won. This tended to be because the gerrymandered districts simply had many more Republican voters living in them, and the Democratic voters were split up into various districts.

Unprecedentedly, in the 2020 presidential election, Georgia flipped from being a Republican stronghold state to electing the Democratic presidential candidate and two Democratic United States Senators. This was because the popular vote is used to determine the outcome of the presidential and senate races which confirmed to many political organizations that Georgia had more Democratic voters than previous elections suggested. However, the United States House and state legislatures are still not representative of the political leanings of the state because of gerrymandering.

States such as California have passed laws banning gerrymandering, and they instead use a heavily vetted commission of people unaffiliated with either political party to create the districts.

### 1.2 Project Overview

It is, however, difficult to find and vet people who are able to act as neutral parties, especially in those states that are already heavily gerrymandered. This suggests that an approach that can be done without a neutral party might be more practical. Over the last 60 years, there has been a lot of work done by researchers to build "fair" algorithms to redistrict. The issue with creating an algorithm to create something as potentially impactful as this is that by controlling or maximizing fairness with regard to one aspect, one might see detrimental and unintended consequences in another. A classic example of gerrymandering looks like very oddly shaped districts with the intent to split up the opposing parties vote. Yet, if one were to maximize only compactness to counteract gerrymandering, it leads to a biased - favoring Republican - outcome [2].

Determining what measures of fairness to maximize on, while not biasing the results of the election is a very difficult task. This paper seeks to explore a new area that has been neglected in the literature - actual voter turnout of proposed districts. Legally, districts are required to maintain approximately equal population with a certain level of tolerance. There, at least in Georgia, is no requirement to have an approximately equal voting-age population. In fact, there is a notable spread in the voting age population in Georgia, spanning from 489,868 to 541,900 while all district populations are 691,575 ( $\pm 1$) according to 2010 census data [4]. This as well - does not take into consideration the actual voter turnout for these districts in 2020, which varied from 273,017 to 405,882 [1]. At the extreme - some districts have 50

It is relevant to note that there are significant voter suppression acts done in Georgia. High disparities in voter turnout signals a lack of fairness - saying that between districts in the same state, one vote may have different weights. An algorithm that redistricts might therefore want to consider its impact on voter turnout. For the purposes of this paper, the fairness measure introduced will be the standard deviation of voter turnout across districts. A redistricting algorithm can then be judged by its spread, compared to other algorithms and the status quo. It is possible that by leveling other fairness measures, it might indirectly improve upon this measure as well. This paper will do just that - implement a proposed redistricting algorithm - the Bisection Protocol - and judge it on this additional metric.

An additional area this paper builds on is adding a way for the Bisection Protocol Algorithm to take geography of the regions into account. As will be discussed later, the original algorithm works in a "non-continuous geometric" setting, so we try to implement a constraint to the algorithm to ensure that the final election districts are geographically contiguous, as is generally required by law in the United States.

## 2 Prior Work

There are currently four redistricting protocols that approach fairness with a game-theoretic approach. The first, by Landau et al. (2009) does use a neutral party. The second approach by Pegden et al. (2017) is called I-Cut-You-Freeze (ICYF) which does not use a neutral party but does provide asymmetry in outcome based on the first plater. The third by Tucker-Foltz (2019) proposes a Cut-and-Choose algorithm that does not use a simple majority to win a district. The fourth algorithm that this paper aims to build on is the Bisection Protocol [6] which does not require a neutral third party, is more efficient than ICYF and uses a simple majority to win a district. After implementing the algorithm on Georgia's districts, the Bisection Protocol will be assessed on its resulting voter turnout spread using 2020 data compared to the actual spread of the current districts.

### 2.1 The Bisection Protocol

The Bisection Protocol works by taking the entire state and splitting in half (each with a sufficient population for a whole number of districts). When a split results in the population for one district, that becomes the proposed district. The bisections continue until all the districts are created. This analysis is based on multiple assumptions that are not necessarily accurate. The first is that this redistricting algorithm is done in the non-geometric setting - meaning that the proposed districts need not be contiguous at all and their location and relative location is overlooked by the algorithm entirely. This is the same setting that was considered in ICYF. As well, it is assumed that all citizens are voters, there are only two parties each voter selects from.

Population is normalized such that the population for each district is equal to one. Let $n$ be the number of districts. Given a district with population one, if one player has $\geq$1/2 population in that district, they win that district by a simple majority. The minimum vote-share needed by a player to win at least $j$ districts out of $n$ under optimal play is called $t_{n,j}$. The authors then describe a recurrence relationship based on optimal game-theoretic play of both players. When the number of districts is of the form $2n$ then the recurrence relationships are as follows.

LEMMA 1. *When $n = 2^r$ for some $r \in \mathbb{N}$, the thresholds $t_{n,1}$ and $t_{n,n}$ satisfy the system of recurrences*

$$t_{n,1} = t_{n,n} = 1/2 \text{ for } n = 1,$$

$$t_{n,n} = n - 2t_{n/2,1}, \text{ and}$$

$$t_{n,1} = n/2 - t_{n/2,n/2}.$$

*Furthermore, these recurrences have closed-form solutions*

$$t_{n,1} = 2^{\lceil (r-1)/2 \rceil - 1} = \begin{cases} \frac{\sqrt{n}}{2} & \text{if } 2 \mid r \\ \frac{\sqrt{2n}}{4} & \text{otherwise} \end{cases}$$

*and*

$$t_{n,n} = n - 2^{\lceil r/2 \rceil - 1} = \begin{cases} n - \frac{\sqrt{n}}{2} & \text{if } 2 \mid r \\ n - \frac{\sqrt{2n}}{2} & \text{otherwise.} \end{cases}$$

In this case, it was proved that the resulting minimum vote share thresholds defined by this relationship form a subgame perfect Nash equilibrium.

When the number of districts is not of the form $2n$, the authors defined the following recurrence relationship.

LEMMA 2. *Let $n, j \in \mathbb{N}$. If $j = 0$, then $t_{n,j} = 0$. When $n = j = 1$, $t_{n,j} = 1/2$. For $n \geq 2$ and $j \geq 1$,*

$$t_{n,j} = \min_{k \in K} \left\{ \left(a - t_{a,a-k+1}\right) + \left(b - t_{b,b-(j-k)+1}\right) \right\}$$
$$= \min_{k \in K} \left\{ n - t_{a,a-k+1} - t_{b,b-(j-k)+1} \right\},$$

*where $a = \lfloor n/2 \rfloor$ is the size of the smaller piece after the cut (side A), $b = \lceil n/2 \rceil$ is the size of the larger piece (side B), and $K = \{k \in \mathbb{N} \colon 0 \leq k \leq j, 0 \leq k \leq a, 0 \leq j - k \leq b\}$ is the set of feasible seat-shares from the part of size $a$.*

As a result of Lemma 2, one can extract an optimal $k^*$ which is the $k$ that results in the minimum. Then, one can also extract an optimal $j^*$ as defined as the maximum districts a player can win given their vote share and the current remaining number of districts. Thus, given a $j^*$ and $k^*$, a player aims to get $k^*$ seats from the smaller side of the bisection and $j^* - k^*$ seats from the larger side getting them to their goal of $j^*$ between the 2 halves.

Given a method to find the optimal threshold given the number of districts remaining and the target number of districts to win, the Bisection Protocol defines a formal algorithm.

### 2.1.1 Bisection Mixed Integer Program

The input to this function is $S$ - the set of population vectors that will be bisected by the algorithm, $m$ - the remaining number of districts the set $S$ comprises of, $j^*$ and $k^*$ derived from Lemma 2, and $d$ - the current player. $\tau$ represents the acceptable level of population imbalance whereas PA Bar represents the ideal total population of side A. $v_{d,i}$ represents the vote share of player $d$ in unit $i$ in $S$. $X_i$ is 1 if $i$ is included in side A of the bisection, and 0 if it is included in side B of the bisection. This is formulated as a feasibility MIP - meaning that there is not an objective and there may be more than one solution for any bisection.

$$\begin{aligned}
(\text{Bisection}(S, m, j^*, k^*, d)) \quad & (1-\tau)\overline{P}_A \leq \sum_{i \in S} p_i x_i \leq (1+\tau)\overline{P}_A, \\
& \sum_{i \in S} v_{d,i} x_i \geq \frac{t_{a,k^*}}{a} \sum_{i \in S} (v_{d,i} + v_{\phi,i}) x_i, \\
& \sum_{i \in S} v_{d,i}(1 - x_i) \geq \frac{t_{b,j^*-k^*}}{b} \sum_{i \in S} (v_{d,i} + v_{\phi,i})(1 - x_i), \\
& x_i \in \{0, 1\} \quad \forall\, i \in S.
\end{aligned}$$

The first constraint ensures the population size between each side is in accordance with the required threshold $\tau$. The second and third constraint ensure that for both side A and side B, players are playing optimally according to Lemma 2 - such that each player is able to get their $j^*$ and $m - j^*$ respective seats in the resulting bisection.

#### 2.1.2 Bisect Function

**Function** *Bisect($S, m, d$)*:

Compute $j^*, k^*$ (via dynamic programming using Lemma 2)

**while** *Bisection($S, m, j^*, k^*, d$) is infeasible* **do**

$j^* \leftarrow j^* - 1$

Re-compute $k^*$

**end**

**return** *Bisection*$(S, m, j^*, k^*, d)$

**Algorithm 2:** Bisection protocol for the discrete nongeometric setting

The Bisect Function with loop through different values of $j^*$ and $k^*$ until a feasible solution is found for the Bisection MIP. $J^*$ must be greater than or equal to zero, so if $j^*$ were to drop below that level, the set $S$ is unable to be bisected.

Once a resulting bisection has scaled population one, meaning that it is the population required for just one district, this piece is finalized and continues a district. Players then switch turns bisecting until all n districts have been created.

## 3 Methods

### 3.1 Data Sourcing

Because the goal of this project is to assess the voter turnout for Georgia, the granularity of the data was chosen to be at the county level - except for a small handful of counties (three) that needed to be split in thirds because of their population size, resulting in about 175 counties to be split between 14 districts. For those split counties - it was assumed that voter turnout would be split evenly. The dataset used contains for each county: total population, total vote counts, vote counts for Democrats and Republicans. Because the Bisection Protocol assumes only two players - Republicans and Democrats, and that each voter is a player, additional columns were created as "Democrat Seats" and "Republican Seats." Democrat seats measured the percent of votes that went to the Democrat party in that county over the sum of only Democrat and Republican votes multiplied by the scaled population. The scaled population in a county is that county's population multiplied by 14 and divided by the total Georgian population. This is all the required data sourcing for the Bisection Protocol.

For this project, three additional data sources were required: First, voter-turnout information by current Georgia districts to compare against those proposed by the Bisection Protocol. This data was sourced separately from the original data set and the values are slightly different in terms of total votes. There are fewer recorded votes in this dataset signaling that the spread of votes might be slightly underestimated. Second, locational data that provides the neighboring counties to each county to be used when considering the geometric setting. This neighbors dataset was used to create a contiguity matrix, which was later used in the geographical constraint. Last, FIPS data that connects each county to a physical location on a map to be used for visualizations.

### 3.2 Implementation of the Bisection Protocol Algorithm

This project was run on Google Colab with Gurobi as the optimization engine for the MIP. For the first run of the project, once the data was sourced, the Bisection Protocol was implemented exactly as outlined.

The first split by Bisection resulted in a side A and B both of size 7 (in units of districts). The second level of splits resulted in size 3 and 4. The third level resulted in size 2 and 1 for the 3 unit group and

2 and 2 for the 4. Any split of size one was finalized as a district. The remaining groups were all split in half one more time to get all 14 districts.

After each iteration of Bisect, the result from the MIP program was the binary variable $x$ representing inclusion into side A. This data was then joined with the previous identifiers like FIPS county code, county name, population, etc. for side A, and $1-x$ was joined to get the identifiers for side B to be fed back into the function for the next level of split.

For each bisection, there is a possibility that the $j^*$ derived by Lemma 2 might not produce a feasible solution for the MIP. About one third of the bisections required $j^* = j^* - 1$ but none required $j^* = j^* - i \ \forall \ i \geq 2$.

Each subsequent bisection has fewer counties to allocate between side A and B, resulting in later levels running significantly faster. Level 1, splitting from 14 districts (comprising about 175 counties = 175 binary variables) to 7 and 7 took 2 attempts by Bisect, taking approximately 0.04 and 0.03 seconds. The final level, splitting 2 into one and one approximately one fifth of the time.

After all 14 districts were finalized, a choropleth mapbox was used to visualize the final bisections. Within each district, each county that was included in a district had a corresponding FIPS code that was used for its locational mapping.

### 3.3 Continuous Geometric Setting

As aforementioned, the original Bisection Protocol algorithm assumed a non-continuous geometric setting, meaning that the final districts drawn by the algorithm did not have to be contiguous. To rectify this, constraints were added to the Bisection mixed integer program that checked for contiguity of the resulting district.

A graph theory approach was taken, where each district after a bisection is treated as a connected graph. A connected graph is a graph where there exists a path from any point to any other point in the graph. In our context, if we treat each county as a vertex, we can treat a district as contiguous if it is a connected graph, where there exists a path from any county in the district to any other county. In a connected graph, the minimum number of edges that must exist for $N$ vertices are $N-1$ edges, where each vertex is only connected to one other vertex.

We can implement this as a constraint by using our contiguity matrix, where $C_{ij} = 1$, if county $i$ neighbors/touches county $j$, and 0 otherwise. $C_{ii} = 0$, to make calculations easier. Then, this means that for the smallest case, a graph with two vertices, we will have 2 edges, one saying County 1 touches County 2 and another saying County 2 touches County 1. We can use the following constraint.

$$\sum_{k=1}^{N} \sum_{k=1}^{N} C_{ij} x_i x_j \geq 2(\sum_{k=1}^{N} x_k - 1) \ \forall \ i, j \qquad (1)$$

Equation (1) says that when you sum up the edges across the district between all of the counties, they need to be greater than the minimum number of edges needed for a connected graph, $N-1$. We include the multiplier of 2 since $C_{ij} = C_{ji} \ \forall \ i, j$, so we are double counting each pair of counties.

This is, of course, not a linear constraint, which is necessary for a mixed integer programming. However, using basic principles of integer programming, it can be reformulated into multiple linear constraints using dummy variables. The constraint has not been reformulated for the purposes of this project, as the implementation using Gurobi handles the reformulation on its own when solving the problem.

If this constraint holds for both sides of the bisection then this connected graph constraint ensures contiguity of the resulting districts.

When implementing this constraint, an issue arose in how the districts were created. The geographical constraint did not create contiguous districts on both sides of the bisection. This paper's assumption is that more granular data is needed to ensure contiguity while meeting the population constraints for each district. Currently, county data is used, and some of the largest counties in the state have populations greater than the maximum population threshold for the final districts. By using more granular data, such as census tracts or zip codes, the algorithm may be able to find contiguous districts.

### 3.4 Evaluating Fairness of the New Algorithm

After the 14 districts were finalized, the scaled population and total votes for each district were extracted. Because the algorithm introduces a level of intolerance that builds upon itself every subsequent split, the variation in population size in each proposed district was much greater than the actual variation population for each of Georgia's districts. Therefore, the bisection algorithm will be scored against both the scaled spread and the unscaled spread versus the current spread of Georgia's district voter-turnout in 2020. For the proposed districts by the bisection algorithm, a scaled population vector was created representing how far off from a perfect split each district population was.

This Bisection Algorithm was run twice - once for the Democrat player bisecting first, and once for the Republican player bisecting first.

The actual standard deviation of voter turnout in the 2020 election for Georgia's 14 districts is 37234.84.

For the Democrat player going first: The standard deviation of the scaled voter turnout for the proposed districts by the Bisection Protocol is 17945.12. The standard deviation of the unscaled voter turnout for the proposed districts is 22581.29.

For the Republican player going first: The standard deviation of the scaled voter turnout for the proposed districts by the Bisection Protocol is 18320.02. The standard deviation of the unscaled voter turnout for the proposed districts is 26649.19.

Thus, for both players starting and the scaled and unscaled case, even though it is an assumption of the Bisection Protocol that all citizens are voters which is historically not true in the United States, this algorithm is still able to improve upon the voter turnout spread versus the current districts by a meaningful amount.

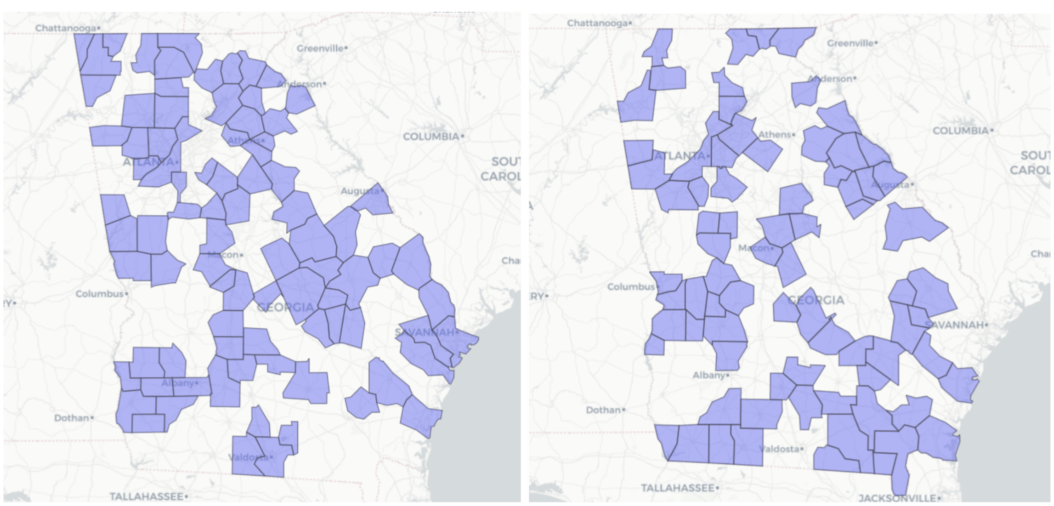


Figure 1: Side A and B after first Bisection - Surprisingly the same if Republican or Democrat starts.

## 4 Conclusions

This project built upon previous work by creating a theoretical continuous geometric setting and introducing a new fairness measure of voter turnout spread to evaluating the Bisection Protocol

Algorithm. The voter turnout spread was reduced by using this algorithm when compared to the current gerrymandered districts, indicating that the algorithm approach equalizes voter turnout across districts. The theoretical geometric setting was implemented but mostly failed to create contiguous districts due to both the granularity of the data used and the geometric constraint combined with the other constraints of the Bisection Protocol. Future work can be done in implementing this algorithm with the geometric constraint and more granular geographical regions, such as census tracts or zip codes. As well, further work into allocating districts based on voting age population and voter turnout should be further considered when exploring new districting approaches.

# 5 Visualizations

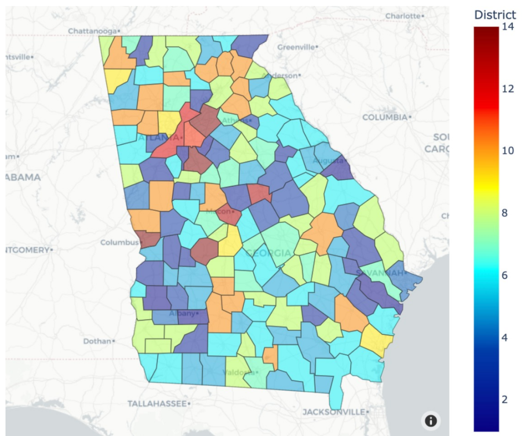


Figure 2: Finalized District Map for Georgia: Player 0 starts (Democrat)

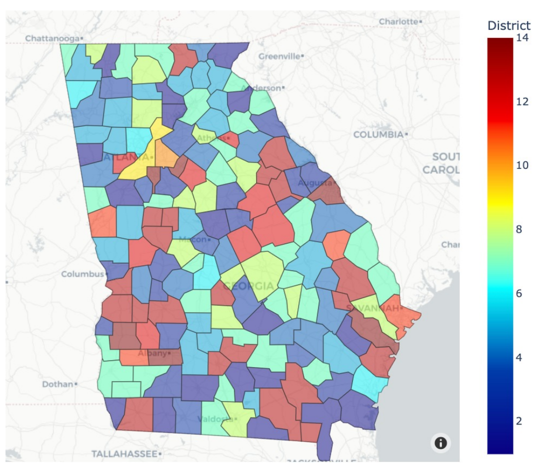


Figure 3: Finalized District Map for Georgia: Player 1 starts (Republican)

## 6 References


[1] CBS News. (n.d.). 2020 Georgia election results. CBS News. Retrieved April 1, 2022, from https://www.cbsnews.com/elections/2020/georgia/

[2] Chen J, Rodden J (2013) Unintentional gerrymandering: Political geography and electoral bias in legislatures. Quarterly Journal of Political Science 8:239–269

[3] Garbus, R. (2022, January 27). Here's how redistricting changed Georgia's 2022 maps for the House, state Senate, and U.S. House. Atlanta Magazine. Retrieved April 1, 2022, from https://www.atlantamagazine.com/news-culture-articles/heres-how-redistricting-changed-georgias-2022-maps-for-the-house-state-senate-and-u-s-house/

[4] Georgia Congressional Districts. (2020, October 16). Georgia General Assembly. Retrieved 1 April, 2022, from https://web.archive.org/web/20201016162356/http://www.legis.ga.gov/Joint/reapportionment/Documents/SEPT

[5] Georgia Secretary of State. (2020, November 20). Election Night Reporting. Election Night Reporting. Retrieved April 4, 2022, from https://results.enr.clarityelections.com/GA/105369/web.264614//summary

[6] Ludden, Ian G., Rahul Swamy, Douglas M. King, and Sheldon H. Jacobson. n.d. "A Bisection Protocol for Political Redistricting." Operations Research, 1-51. Accessed April 1, 2022.

[7] NBC News. (2022, April 12). Georgia presidential election results 2020: Live results and polls. NBC News. Retrieved April 1, 2022, from https://www.nbcnews.com/politics/2020-elections/georgia-president-results

[8] United States Census Bureau. (2021, October 8). County Adjacency File. U.S. Census Bureau. Retrieved April 1, 2022, from https://www.census.gov/programs-surveys/geography/library/reference/county-adjacency-file.html

[9] USDA Natural Resources Conservation Service. (n.d.). County FIPS Codes | NRCS. Natural Resources Conservation Service. Retrieved April 1, 2022, from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143$_0$13697