# Removing partisan bias in redistricting: computational complexity meets the science of gerrymandering[+]

**Bhaskar DasGupta**[♥]
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
*bdasgup@uic.edu*

[+] Joint result with Tanima Chatterjee, Laura Palmieri, Zainab Al-Qurashi and Anastasios Sidiropoulos

[♥] Partially supported by NSF grants IIS-1160995 and IIS-1814931

# Gerrymandering

## Creation of district plans with *highly asymmetric* electoral outcomes to *disenfranchise* voters

- **Long history starting from as early as 1812**

  **1812 :** shape of South Essex district (Massachusetts) resembling
  a ***salamander*** created to favor selected candidates

- **Extensive legal history too!**

  **1986**: US Supreme Court : gerrymandering *is justiciable*
  **2006***:* US Supreme Court : *some measure* of partisan symmetry
  *may be used* to remedy gerrymandering
  Which measure? Court did not say. Depends case by case.
  **2019:** US Supreme Court : best settled at the legislative and political
  level (ALAS!)

- **Major impediment to removing gerrymandering**

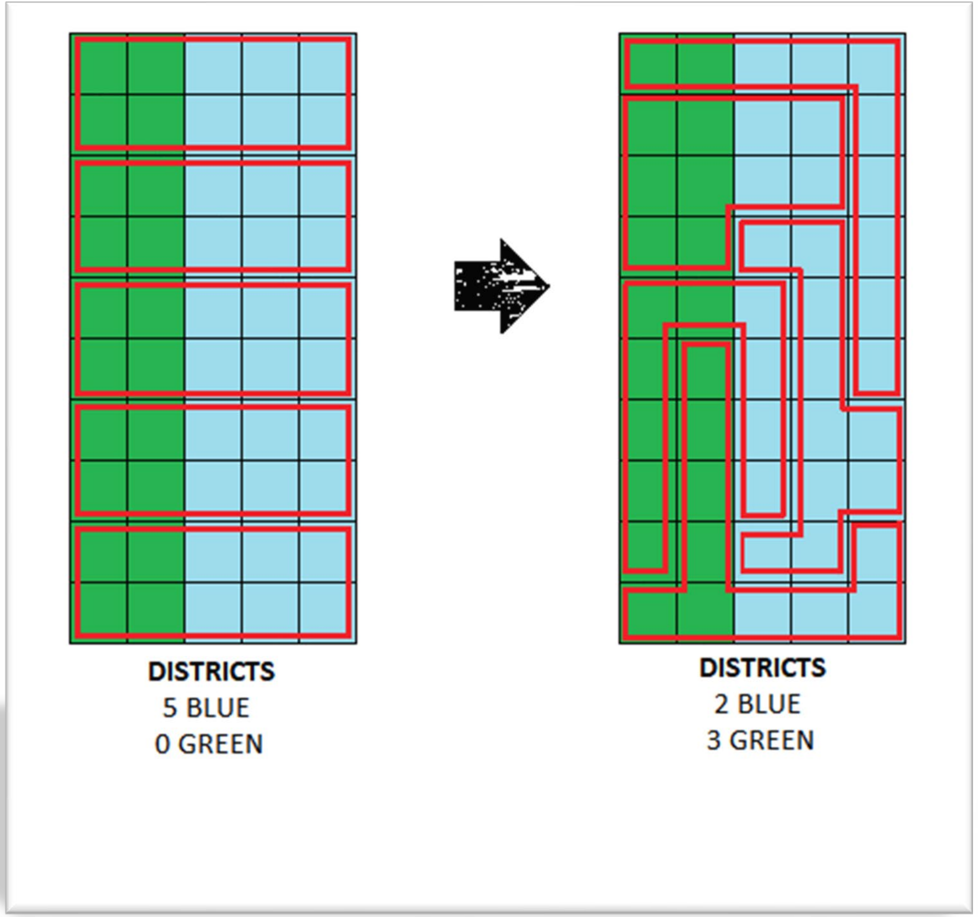
  How to formulate an effective and precise measure for partisan bias
  that will be **acceptable in courts** ?


***"Gerry" and "salamander"***
***1812, State Senate Elections, Massachusetts***

# Some tools politicians use for partisan gerrymandering in 2-party system

- ❑ **Packing** → concentrate voters of opposition party in a single district

- ❑ **Cracking** → spread voters of opposition party across many districts



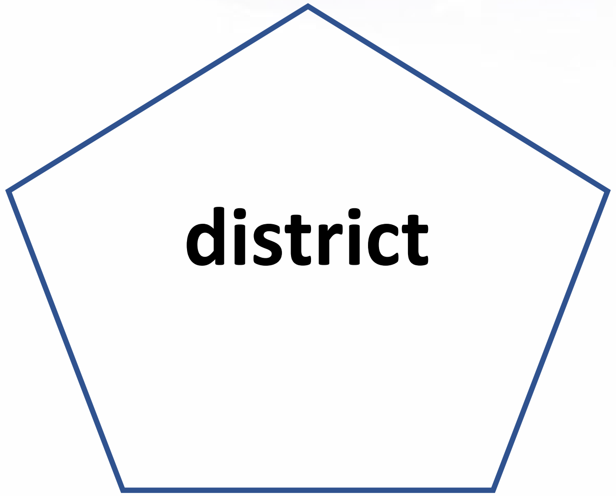

Other methods include

- Hijacking
- Kidnapping etc.

# "Efficiency Gap" measure for partisan gerrymandering

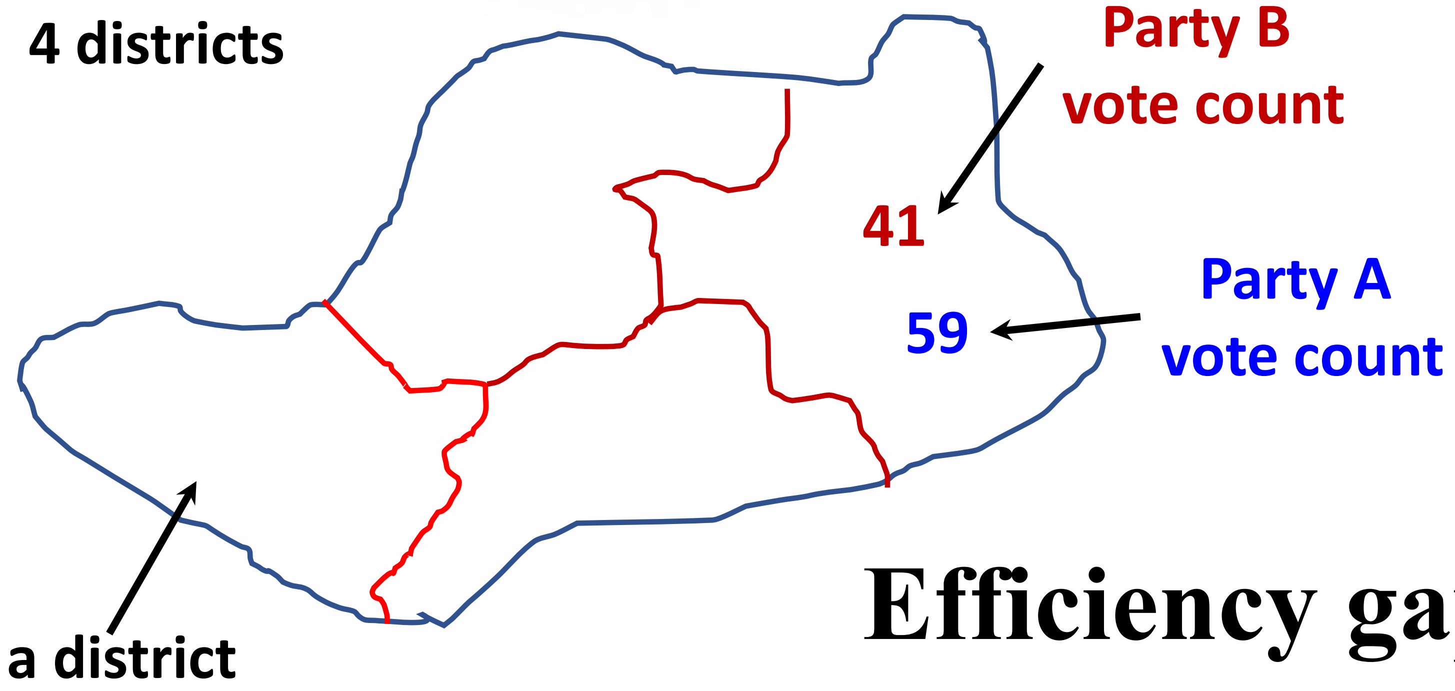
- Introduced by Stephanopoulos and McGhee in 2014 for a 2-party system (such as USA)
- Minimizes absolute difference of total "wasted votes" between the parties
- Very promising in several aspects, e.g.,
  - provides a "mathematically precise" measure of gerrymandering with desirable properties
  - was found legally convincing in a US appeals court case
    - ALAS, Supreme Court overturned the ruling in 2019

# “Wasted votes” for a district




- Total votes 100 (need 51 to win)
  - Party A vote 59
  - Party B vote 41
- Wasted votes for Party A 59-51=8
- Wasted votes for Party B 41

# "Efficiency gap" measure for the whole map




## Efficiency gap

=

$$\frac{\left|\,\text{sum of Party A wasted votes over all districts} - \text{sum of Party B wasted votes over all districts}\,\right|}{\text{Total votes over all districts}}$$

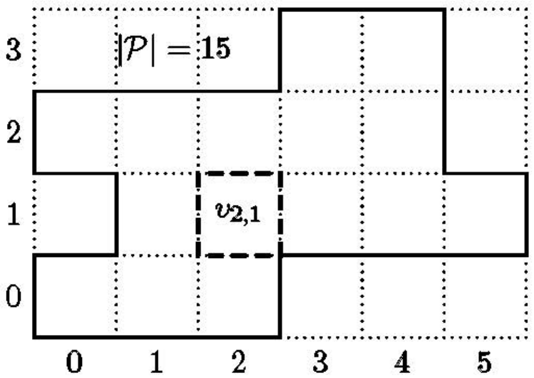
# Formalization of the efficiency gap calculation problem

Basic assumption: *only two* parties: Party A and Party B

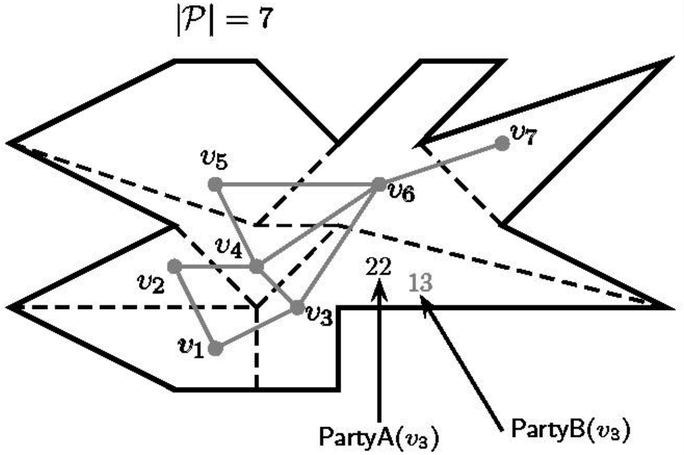
($3^{rd}$ party votes are negligible, like in USA)

Topological part of an input: a "**map**" $\mathcal{P}$

- ▷ partitioned into *atomic elements* or *cells*
  - *e.g.*, , subdivisions of counties



Rectilinear polygon



Arbitrary polygon

Two possible types of maps:

**Rectilinear polygon $\mathcal{P}$ without holes**

- ▷ $\mathcal{P}$ placed on a unit grid of size $m \times n$
- ▷ atomic elements (cells) ⇒ unit squares of grid inside $\mathcal{P}$
- ▷ $v_{i,j}$ : cell on $i^{\text{th}}$ row and $j^{\text{th}}$ column

**Arbitrary polygon $\mathcal{P}$ without holes:**

- ▷ atomic elements (cells) ⇒ sub-polygons (without holes) inside $\mathcal{P}$
- ▷ *Alternate* way of looking: **planar graph $G(\mathcal{P})$**
  - nodes are cells
  - edge connects two cells if they share boundary

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

**Parameters of our gerrymandering problem**

- **Map $\mathcal{P}$:**
  - *size* $|\mathcal{P}|$ : number of cells or nodes in $\mathcal{P}$
- **Cell or node $y$ of $\mathcal{P}$:**
  - $\mathsf{PartyA}(y)$: total number of voters for Party A
  - $\mathsf{PartyB}(y)$: total number of voters for Party B
  - $\mathsf{Pop}(y) = \mathsf{PartyA}(y) + \mathsf{PartyB}(y)$: total number of voters
- **Global:**
  - $\kappa$: *required* (legally mandated) number of districts ($1 < \kappa < |\mathcal{P}|$)
    - Hard constraint: solution with different value of $\kappa$ would be *illegal*
    - precludes designing approximation algorithm in which the value of $\kappa$ changes even by just $\pm 1$
    - computational hardness for a value of $\kappa$ may *not* necessarily imply hardness for another value of $\kappa$

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

**Granularities of numeric parameters**

- **Course granularity:**
  - $\text{Pop}(y)$'s are numbers of arbitrary size
  - total number of bits contributes to input size
  - data at the "county" level or "census block *group*" level
- **Fine granularity:**
  - $\forall$ cell or node $y$: $0 < \text{Pop}(y) \leq c$ for some *fixed constant c*
  - data at the "Voting Tabulation District" (VTD) level or "census block" level
- **Ultra-fine granularity:**
  - $\forall$ cell or node $y$: $\text{Pop}(y) = c$ for some *fixed constant c*
  - theoretically interesting case, but practically a bit unrealistic

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

| | | |
|---|---|---|
| $\kappa$ | number of districts | |
| $\mathcal{S}$ | set of all cells in given polygonal map $\mathcal{P}$<br>or, set of all nodes in given planar graph $G(\mathcal{P})$ | |
| **districting scheme** | partition of $\mathcal{S}$ into $\kappa$ subsets $\mathcal{S}_1, \ldots, \mathcal{S}_\kappa$ | |
| **Notations for each $\mathcal{S}_j$** | **Party affiliations in $\mathcal{S}_j$** | $\text{PartyA}(\mathcal{S}_j) = \sum_{y \in \mathcal{S}_j} \text{PartyA}(y)$<br>$\text{PartyB}(\mathcal{S}_j) = \sum_{y \in \mathcal{S}_j} \text{PartyB}(y)$ |
| | **Population of $\mathcal{S}_j$** | $\text{Pop}(\mathcal{S}_j) = \text{PartyA}(\mathcal{S}_j) + \text{PartyB}(\mathcal{S}_j)$ |

## Legal requirements for valid re-districting plans

- ☐ Every $\mathcal{S}_j$ must be a connected polygon
- ☐ Populations of different $\mathcal{S}_j$'s must be as equal as possible

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

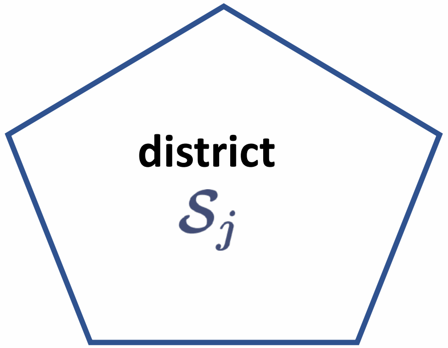
**Legal requirements for valid re-districting plans**

- Every $\mathcal{S}_j$ must be a connected polygon
- Populations of different $\mathcal{S}_j$'s must be as equal as possible
  - **Strict partitioning criteria**
    $\{\mathcal{S}_1, \ldots, \mathcal{S}_\kappa\}$ is an *exact $\kappa$-equipartition* of $\mathcal{S}$, *i.e.*, $\forall\, j : \mathsf{Pop}(\mathcal{S}_j) \in \{\lfloor \mathsf{Pop}(\mathcal{S})/\kappa \rfloor, \lceil \mathsf{Pop}(\mathcal{S})/\kappa \rceil\}$
  - **(Multiplicatively) approximate partitioning criteria**
    $\{\mathcal{S}_1, \ldots, \mathcal{S}_\kappa\}$ is a $\varepsilon$-*approximate* $\kappa$-equipartition of $\mathcal{S}$, *i.e.*, $\dfrac{\max\{\mathsf{Pop}(\mathcal{S}_j)\}}{\min\{\mathsf{Pop}(\mathcal{S}_j)\}} \leq 1 + \varepsilon$
    courts may allow a maximum value of $\varepsilon$ in the range of 0.05 to 0.1
    *e.g.*, , (US Supreme Court ruling in Karcher v. Daggett, 1983)
  - ***Additively* approximate partitioning criteria**
    $\{\mathcal{S}_1, \ldots, \mathcal{S}_\kappa\}$ is an *additive $\varepsilon$-approximate* $\kappa$-equipartition of $\mathcal{S}$, *i.e.*,
    $\max\{\mathsf{Pop}(\mathcal{S}_j)\} \leq \min\{\mathsf{Pop}(\mathcal{S}_j)\} + \varepsilon$

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

## "Wasted votes" for a district




- **Total votes 100 (Party A needs 50 to win)** $\mathsf{Pop}(\mathcal{S}_j)$
  - **Party A vote 59** $\mathsf{PartyA}(\mathcal{S}_j)$
  - **Party B vote 41** $\mathsf{PartyB}(\mathcal{S}_j)$
- **Wasted votes for Party A 59 – 50 = 9** $\mathsf{PartyA}(\mathcal{S}_j) - \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)$
- **Wasted votes for Party B 41** $\mathsf{PartyB}(\mathcal{S}_j)$
- **Efficiency gap for $\mathcal{S}_j$ 9 – 41 = - 32**

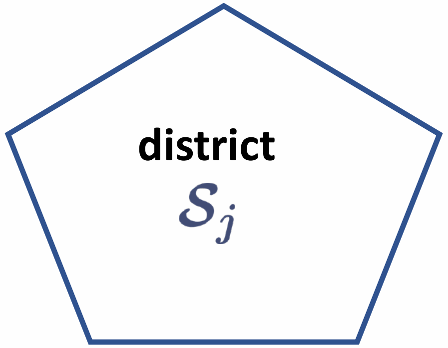
$$\mathsf{Effgap}(\mathcal{S}_j) = \begin{cases} \left(\mathsf{PartyA}(\mathcal{S}_j) - \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)\right) - \mathsf{PartyB}(\mathcal{S}_j) \\ = 2\mathsf{PartyA}(\mathcal{S}_j) - \frac{3}{2}\mathsf{Pop}(\mathcal{S}_j) \end{cases} \quad \text{if } \mathsf{PartyA}(\mathcal{S}_j) \geq \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)$$

**from the point of view of Party A (the victim party of gerrymandering)**

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

## "Wasted votes" for a district




- Total votes 100 (Party A needs 50 to win) — $\mathsf{Pop}(\mathcal{S}_j)$
  - Party A vote 41 — $\mathsf{PartyA}(\mathcal{S}_j)$
  - Party B vote 59 — $\mathsf{PartyB}(\mathcal{S}_j)$
- Wasted votes for Party A 41 — $\mathsf{PartyA}(\mathcal{S}_j)$
- Wasted votes for Party B 59 - 50 = 9 — $\mathsf{PartyB}(\mathcal{S}_j) - \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)$
- Efficiency gap for $\mathcal{S}_j$ 41 - 9 = - 32

$$\mathsf{Effgap}(\mathcal{S}_j) = \begin{cases} \mathsf{PartyA}(\mathcal{S}_j) - \left(\mathsf{PartyB}(\mathcal{S}_j) - \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)\right) \\ = 2\mathsf{PartyA}(\mathcal{S}_j) - \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j) \end{cases} \quad \text{if } \mathsf{PartyA}(\mathcal{S}_j) < \frac{1}{2}\mathsf{Pop}(\mathcal{S}_j)$$

**from the point of view of Party A (the victim party of gerrymandering)**

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

| | |
|---|---|
| $\kappa$ | number of districts |
| $\mathcal{S}$ | set of all cells in given polygonal map $\mathcal{P}$ <br> or, set of all nodes in given planar graph $G(\mathcal{P})$ |
| districting scheme | partition of $\mathcal{S}$ into $\kappa$ subsets $\mathcal{S}_1, \ldots, \mathcal{S}_\kappa$ |

$$\textsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa) = \left| \sum_{j=1}^{\kappa} \textsf{Effgap}(\mathcal{S}_j) \right|$$

**(to be minimized)**

**from the point of view of Party A (the victim party of gerrymandering)**

# Formalization of the efficiency gap calculation problem

*only two* parties: Party A and Party B

**$\kappa$-district Minimum Wasted Vote Problem (MIN-WVP$_\kappa$)**

---

**Input**
- ▷ map $\mathcal{P}$ with $\mathsf{Pop}(y)$, $\mathsf{PartyA}(y)$, $\mathsf{PartyB}(y)$ for every cell $y \in \mathcal{P}$
- ▷ integer $1 < \kappa \leq |\mathcal{P}|$

**Assumption** $\mathcal{P}$ has at least one $\kappa$-equipartition **why this assumption ?**

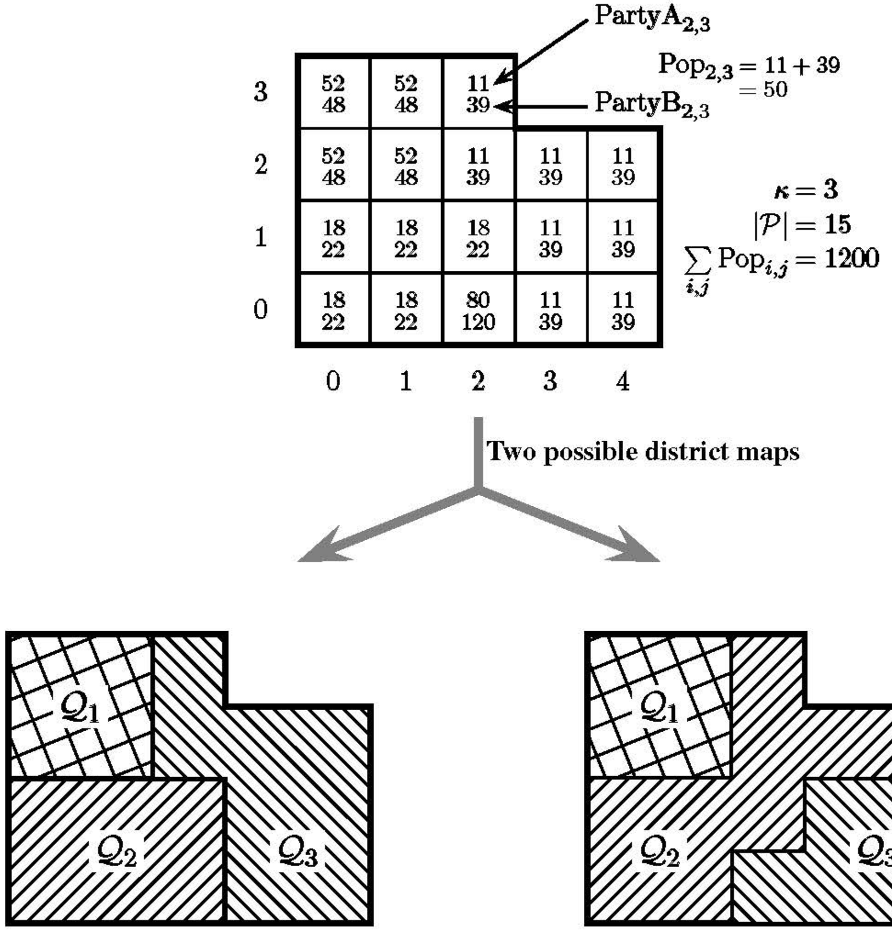
**Valid solution** Any $\kappa$-equipartition $\mathcal{S}_1, \ldots, \mathcal{S}_\kappa$ of $\mathcal{P}$ ‡

**Objective** *minimize* $\mathsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa) = \left| \sum_{j=1}^{\kappa} \mathsf{Effgap}(\mathcal{S}_j) \right|$

**Notation** $\mathsf{OPT}_\kappa(\mathcal{P}) \stackrel{\text{def}}{=} \min \left\{ \mathsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa) \mid \mathcal{S}_1, \ldots, \mathcal{S}_\kappa \text{ is a } \kappa\text{-equipartition of } \mathcal{P} \right\}$

‡ in exact or approximate sense

# A numerical example to illustrate efficiency gap calculation problem




| | PartyA($Q_{...}$) | PartyB($Q_{...}$) | Effgap($Q_{...}$) |
|---|---|---|---|
| $Q_1$ | 208 | 192 | −184 |
| $Q_2$ | 170 | 230 | 140 |
| $Q_3$ | 88 | 312 | −24 |

$\text{Effgap}(\mathcal{P}, Q_1, Q_2, Q_3) = |-184 + 140 - 24| = 68$

| | PartyA($Q_{...}$) | PartyB($Q_{...}$) | Effgap($Q_{...}$) |
|---|---|---|---|
| $Q_1$ | 208 | 192 | −184 |
| $Q_2$ | 134 | 266 | 58 |
| $Q_3$ | 124 | 276 | 48 |

$\text{Effgap}(\mathcal{P}, Q_1, Q_2, Q_3) = |-184 + 58 + 48| = 78$

## Mathematical properties of $\mathsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa)$: set of attainable values

**assume *strict* $\kappa$-equipartition, *i.e.*, $\mathsf{Pop}(\mathcal{S}_1) = \cdots = \mathsf{Pop}(\mathcal{S}_\kappa)$**

### Lemma 1

- $\mathsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa)$ assumes one of the $\kappa + 1$ values:

$$\left| 2 \times \mathsf{PartyA}(\mathcal{P}) - \left(z + \frac{\kappa}{2}\right) \frac{\mathsf{Pop}(\mathcal{P})}{\kappa} \right| \quad \text{for } z = 0, 1, \ldots, \kappa$$

- If $\mathsf{Effgap}_\kappa(\mathcal{P}, \mathcal{S}_1, \ldots, \mathcal{S}_\kappa) = \left| 2\, \mathsf{PartyA}(\mathcal{P}) - \left(z + \frac{\kappa}{2}\right) \frac{\mathsf{Pop}(\mathcal{P})}{\kappa} \right|$ then

$$\frac{\mathsf{Pop}(\mathcal{P})}{2\,\kappa} z \leq \mathsf{PartyA}(\mathcal{P}) \leq \frac{\mathsf{Pop}(\mathcal{P})}{2\,\kappa} z + \frac{1}{2} \mathsf{Pop}(\mathcal{P})$$

**Illustrative example: $\kappa = 2$**

*only* 3 possible values of $\mathsf{Effgap}_2(\mathcal{P}, \mathcal{S}_1, \mathcal{S}_2)$

$$\left| 2 \times \mathsf{PartyA}(\mathcal{P}) - \tfrac{1}{2}\, \mathsf{Pop}(\mathcal{P}) \right| \quad \text{or} \quad \left| 2 \times \mathsf{PartyA}(\mathcal{P}) - \mathsf{Pop}(\mathcal{P}) \right| \quad \text{or} \quad \left| 2 \times \mathsf{PartyA}(\mathcal{P}) - \tfrac{3}{2}\, \mathsf{Pop}(\mathcal{P}) \right|$$

# First a “somewhat” bad news
# (worst-case computational complexity meets gerrymandering)

## Theorem (informal description)

**Not only calculation of efficiency gap is NP-complete, but**

**assuming P $\neq$ NP, no non-trivial approximation is possible in polynomial time**

**But, have no fear !**
**We have only shown hardness**
**in theoretical *worst-case***

# Worst-case computational complexity meets gerrymandering

## Assumptions

- Map $\mathcal{P}$: rectilinear polygon without holes
- Strict partitioning criteria: $\{\mathcal{S}_1, \ldots, \mathcal{S}_\kappa\}$ is exact $\kappa$-*equipartition* of $\mathcal{S}$
- Course granularity: $\mathsf{Pop}(y)$'s are numbers of arbitrary size
- $\mathrm{P} \neq \mathsf{NP}$

**Theorem 1**

For any rational constant $\varepsilon \in (0, 1)$, for any $\rho$ and all $2 \leq \kappa \leq \varepsilon|\mathcal{P}|$,
$\textsc{Min-wvp}_\kappa$ problem for rectilinear polygon $\mathcal{P}$
does not admit a $\rho$-approximation algorithm

**Reduction: from PARTITION problem**

# Worst-case computational complexity meets gerrymandering

## Assumptions

- Map $\mathcal{P}$: planar graph $G = (V, E)$
- (Multiplicatively) approximate partitioning criteria:

  $\{\mathcal{S}_1, \ldots, \mathcal{S}_\kappa\}$ is a $\varepsilon$-*approximate* $\kappa$-equipartition of $\mathcal{S}$, *i.e.*, $\frac{\max\{\mathsf{Pop}(\mathcal{S}_j)\}}{\min\{\mathsf{Pop}(\mathcal{S}_j)\}} \leq 1 + \varepsilon$
- Fine granularity: $\forall$ node $y$: $0 < \mathsf{Pop}(y) \leq c$ for some *fixed constant* $c$

**Theorem 2**

For any constant $0 < \varepsilon < 1/2$,
computing an exact solution of the $\textsc{Min-wvp}_\kappa$ problem
is NP-complete

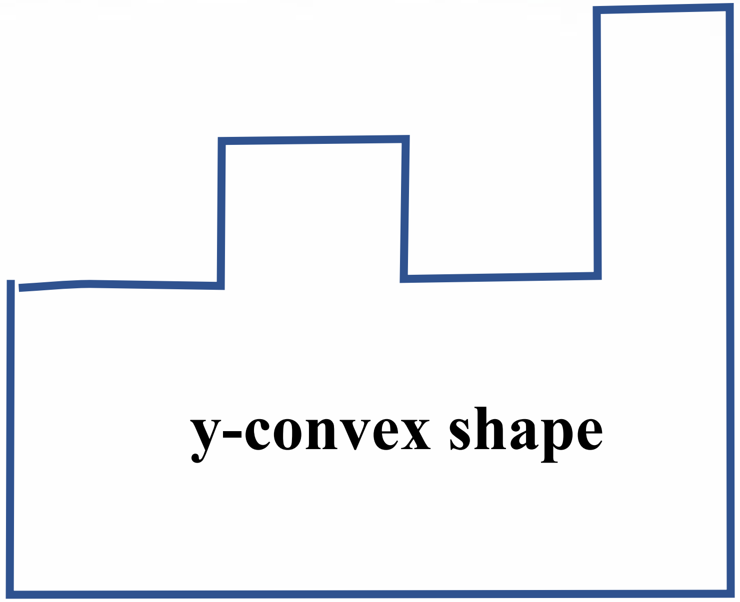
**Proof does *not* provide any non-trivial inapproximability ratio**

**Reduction: from maximum independent set for planar cubic graphs**

# However, even in theory, we can efficiently compute efficiency gap under "*reasonable*" assumptions

*e.g.*, with these assumptions:

- Input map: a rectilinear polygon $\mathcal{P}$ (without holes)
- Every district must have a "nice" shape (*y-convex* shape)
- κ (number of districts) is *constant*
- Total population Pop($\mathcal{P}$) is *polynomial* in number of cells $|\mathcal{P}|$



# We developed and implemented a simple heuristic algorithm based on "local search" method

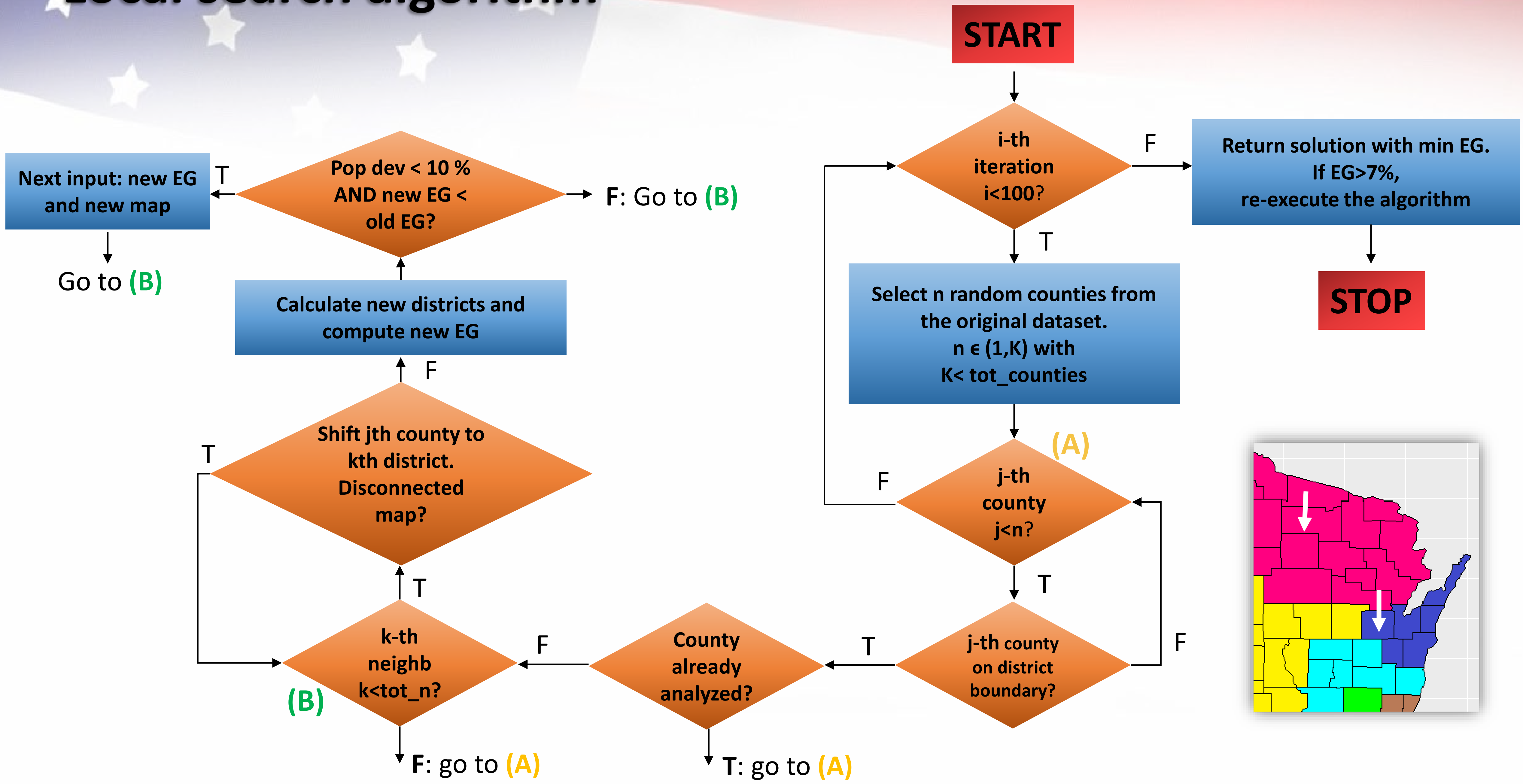
- Start with some existing or random valid solution

- Search for nearby valid solutions by randomly "swapping" local regions among various districts

  - Pitfall: can get stuck with far-away local optima
    but, does not seem to often occur for real maps

➢ Next few slides: results for real maps

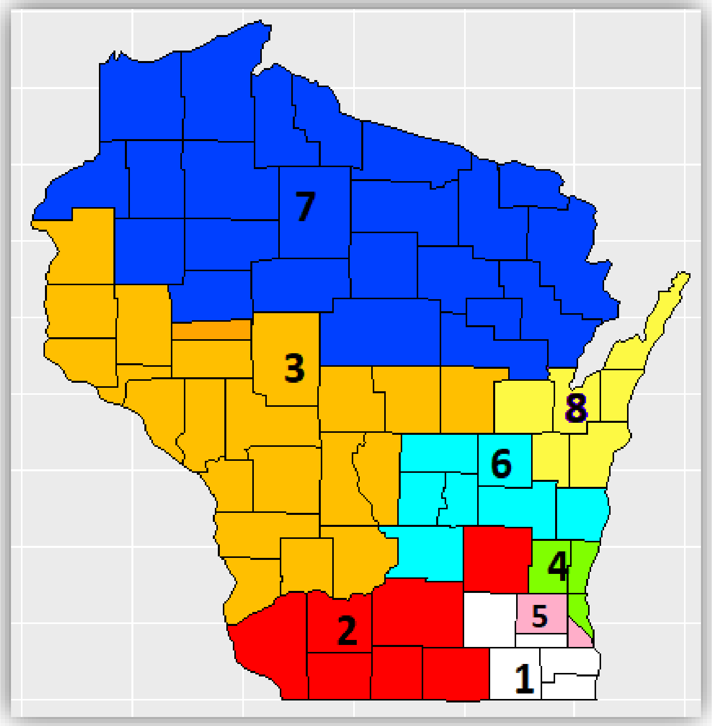
# Local search algorithm



# Wisconsin

**Total votes:** 2,841,407
**Dem votes:** 1,441,804 ~ *51%*
**Rep votes:** 1,399,603 ~ *49%*


**Current EG:** 14.8%
**Dem #seats:** 3
**Rep #seats:** 5

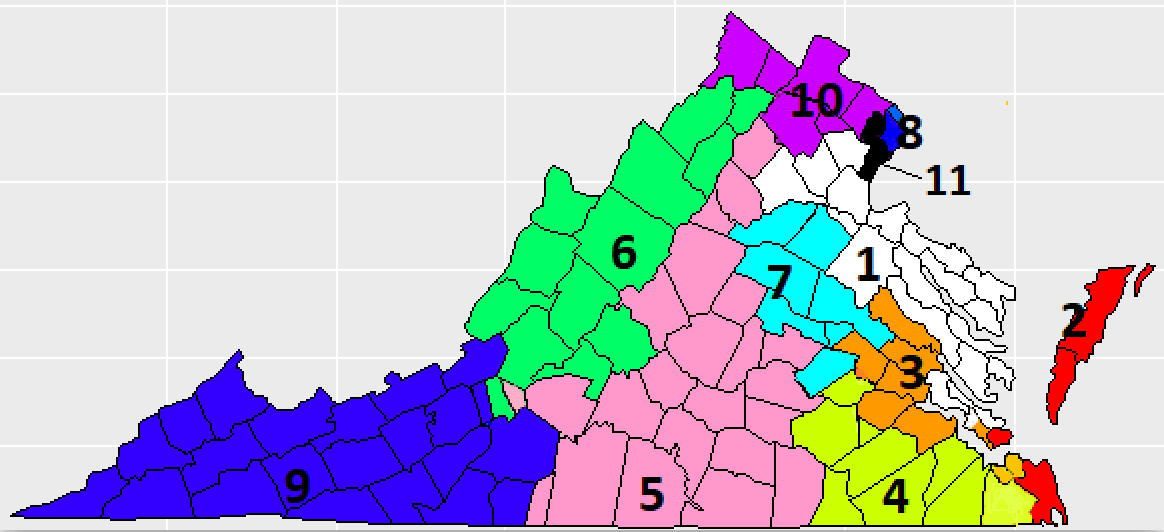



**New EG:** 3.8%
**Dem #seats:** 3
**Rep #seats:** 5

# Virginia

**Total votes:** 3,569,498
**Dem votes:** 1,736,164 ~ *49%*
**Rep votes:** 1,833,334 ~ *51%*

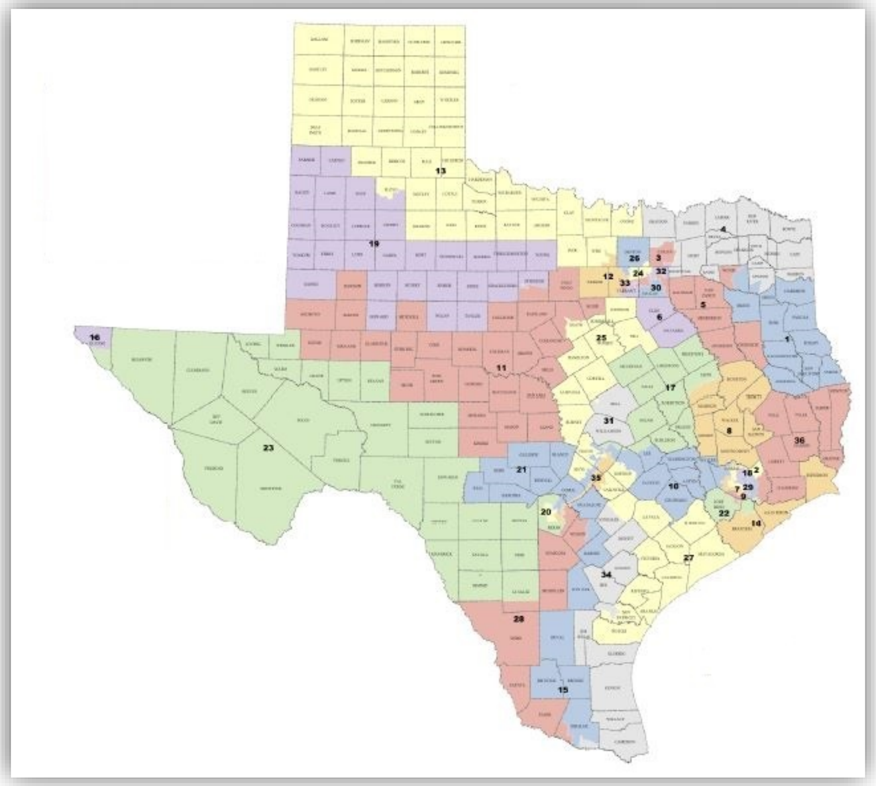

**Current EG:** 22%
**Dem #seats:** 3
**Rep #seats:** 8



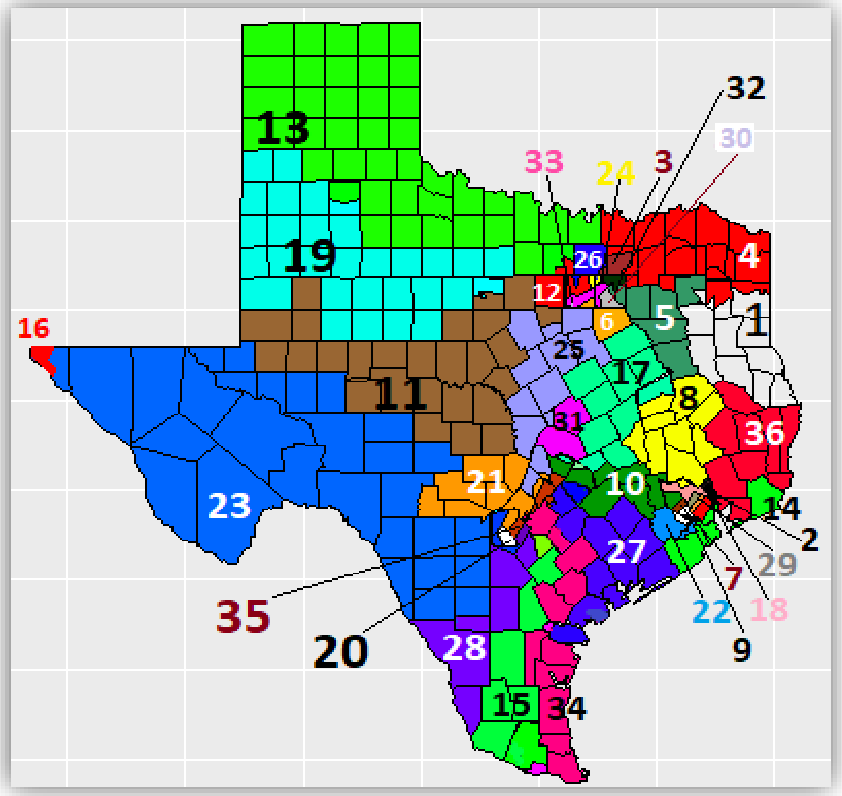
**New EG:** 3.6%
**Dem #seats:** 5
**Rep #seats:** 6

# Texas

**Total votes:** 7,379,170
**Dem votes:** 2,949,900 ~ *40%*
**Rep votes:** 4,429,270 ~ *60%*



**Current EG:** 4.01%
**Dem #seats:** 12
**Rep #seats:** 24

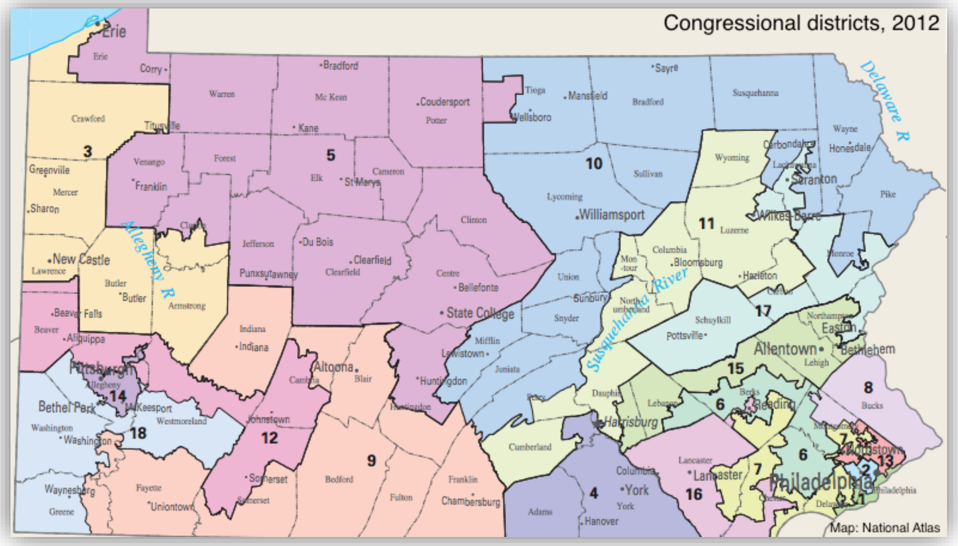


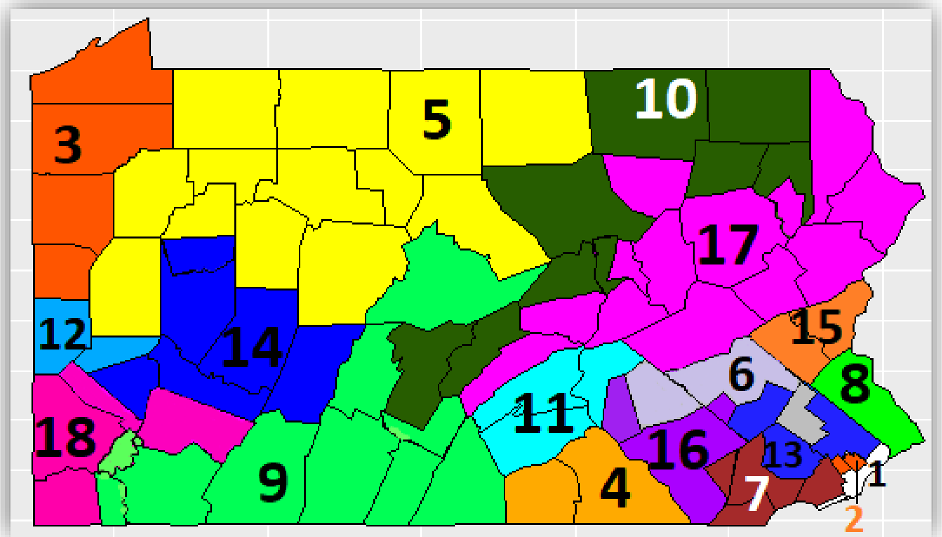

**New EG:** 3.3%
**Dem #seats:** 12
**Rep #seats:** 24

# Pennsylvania

## 2012 House Elections

**Total votes:** 5,374,461
**Dem votes:** 2,722,560 ~ *51%*
**Rep votes:** 2,651,901 ~ *49%*




**Current EG:** 23.8%
**Dem #seats:** 5
**Rep #seats:** 13

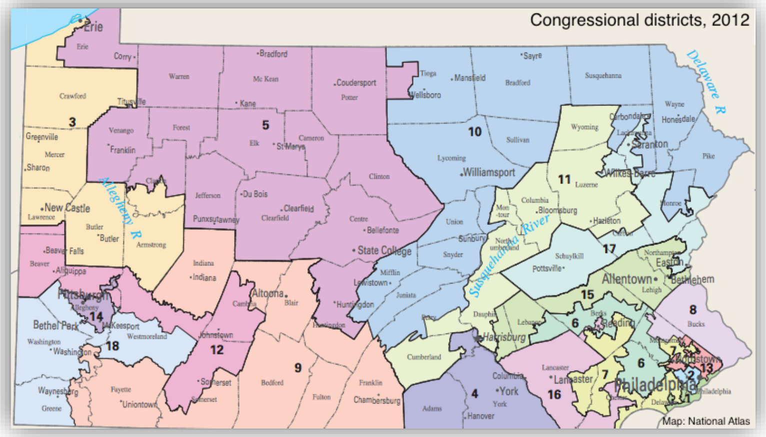
**New EG:** 8.64%
**Dem #seats:** 6
**Rep #seats:** 12

# Pennsylvania

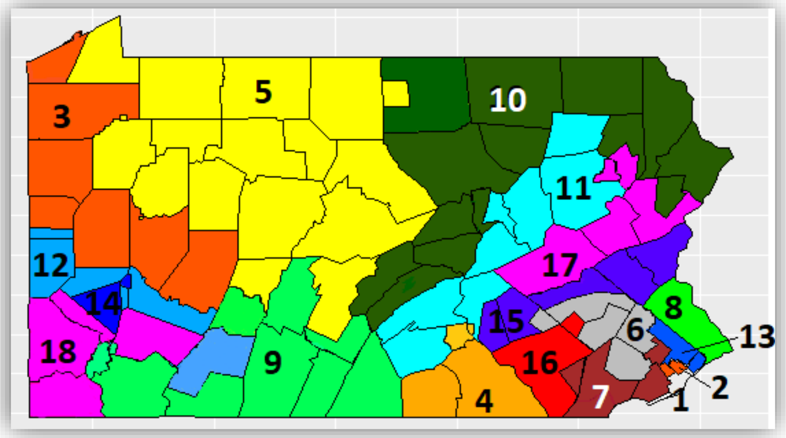
## 2016 Presidential Elections

**Total votes:** 5,896,628
**Dem votes:** 2,925,776 ~ *50%*
**Rep votes:** 2,970,852 ~ *50%*

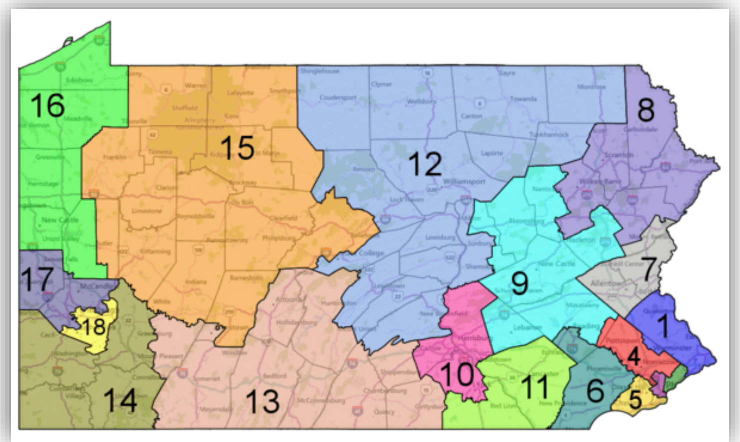
*Created on Feb. 2018 (by a local court in PA) and based on symmetry between seat share and vote share.*




**Current EG:** 14.34%
**Dem #seats:** 6
**Rep #seats:** 12




**New EG:** 8.05%
**Dem #seats:** 7
**Rep #seats:** 11




**New EG:** 3%
**Dem #seats:** 8
**Rep #seats:** 10

# Some Interesting Insights based on simulation results

- ❑ **Seat gain vs. efficiency gap**
  - ❑ lower efficiency gap does not necessarily lead to seat gains for the loosing party
- ❑ **Compactness vs. efficiency gap**
  - ❑ Our new district map have fewer districts that are oddly shaped compared to the current gerrymandered maps
- ❑ **How natural are current gerrymandered districts?**
  - ❑ It seems that original gerrymandered districts are far from being a product of arbitrarily random decisions

# Future research

**Science of gerrymandering is a huge garden with so many unknown fruits for hungry theoretical computer scientists !**

**So many questions, so few answers**

- **Define and analyze new quantitative measures of gerrymandering**
  - **What about 3 or more party systems ?**

- **Analyze computational complexities of existing measures of gerrymandering**

- **Join court cases as an expert witness and convince judges that computational complexity matters**

**Journal paper:** https://link.springer.com/article/10.1007/s10878-020-00589-x

**Data files:** https://www.cs.uic.edu/~dasgupta/gerrymander/

# Thank you for your attention!

# Questions?