

# 2020 Pennsylvania Congressional Districts

## Redistricting requirements

In Pennsylvania, there are few formal districting requirements, but districts must generally:

1. be contiguous
2. have equal populations
3. be geographically compact
4. preserve county and municipality boundaries as much as possible

## Algorithmic Constraints

We enforce a maximum population deviation of 0.5%. We apply a county/municipality constraint, as described below.

## Data Sources

Data for Pennsylvania comes from the ALARM Project's [2020 Redistricting Data Files](#).

## Pre-processing Notes

No manual pre-processing decisions were necessary.

## Simulation Notes

We sample 10,000 districting plans for Pennsylvania across two runs of the SMC algorithm, then filter down to 5,000 total plans. To balance county and municipality splits, we create pseudocounties for use in the county constraint. These are counties, outside of Allegheny County, Montgomery County, and Philadelphia County. Within Allegheny County, Montgomery County, and Philadelphia County, each municipality is its own pseudocounty as well. These counties were chosen since they are necessarily split by congressional districts. We also apply an additional Gibbs constraint to further avoid splitting municipalities.

## Contents

- `PA_cd_2020_stats.csv` contains summary statistics on the sampled redistricting plans
- `PA_cd_2020_plans.rds` is a compressed `redist_plans` object, which contains the matrix of precinct/block assignments and may be used for further analysis.
- `PA_cd_2020_map.rds` is a compressed `redist_map` object, which contains the precinct/block shapefile and demographic data.

Both the `redist_plans` and `redist_map` object are intended to be used with the [redist package](#).

## Codebook for summary statistics

- `draw`: unique identifier for each sample. Non-numeric draw names are real-world plans, e.g., `cd_2010` for an enacted 2010 plan.

- **district**: a district identifier. District numbers roughly match those in the enacted plan, but the correspondence is not perfect.
- **chain**: a number identifying the run of the redistricting algorithm used to produce this draw. Used for diagnostic purposes.
- **pop\_overlap**: a number indicating the fraction of people in this plan who reside in the same-numbered district in the enacted plan.
- **total\_pop**: the total population of each district.
- **total\_vap**: the total voting-aged population of each district.
- **pop\_\***, **vap\_\***: total (voting-aged) population within racial and ethnic groups for each district. Variable codes documented [here](#).
- **plan\_dev**: the maximum population deviation among districts in the plan. Computed as  $\max(\text{abs}(\text{distr\_pop} - \text{target\_pop}) / \text{target\_pop})$ .
- **comp\_edge**: compactness, as measured by the fraction of internal edges kept. Higher values indicate more compactness.
- **comp\_polsby**: compactness, as measured by the Polsby-Popper score. Higher values indicate more compactness.
- **county\_splits**: the number of counties which belong to more than one district.
- **mun\_splits**: the number of Census Designated Places which belong to more than one district.
- **\*\_##\_dem\_\***, **\*\_##\_rep\_\***: vote counts for statewide Democratic and Republican candidates in a certain election. More information [here](#).
- **adv\_##**, **arv\_##**: average vote counts for statewide Democratic and Republican candidates in a certain year. More information [here](#).
- **ndv**, **nrv**: averages of the **adv\_##** and **arv\_##** variables across all available elections.
- **ndshare**: normal Democratic share, computed as  $\text{ndv} / (\text{ndv} + \text{nrv})$
- **e\_dvs**: average Democratic vote share, computed as the average of the Democratic vote share when first scored under each statewide election.
- **pr\_dem**: probability seat is represented by a Democrat; calculated as the fraction of statewide elections under which the district had a majority Democratic share.
- **e\_dem**: expected number of Democratic seats for the plan; equivalent to summing the **pr\_dem** values across districts
- **pbias**: partisan bias at 50% vote share, averaged across all available elections. Positive values indicate Republican bias.
- **egap**: the efficiency gap, averaged across all available elections. Positive values indicate Republican bias.