Using scores_df.parquet

There are 2,016 scores CSV files in our study containing 48 scores for 6.7 million plans!

- 7 states
- 3 chambers
- 16 ensembles
- 6 categories of scores
- 48 scores per plan
- 20,000 plans per ensemble

Each scores CSV is easy to import into a spreadsheet, but if you want to work with the data, this surface area is daunting.

So we created a single, integrated `pandas` dataframe that contains all of the scores: `scores_df.parquet`. Each set of scores is indexed with three columns:

- `state` -- the state name
- `chamber` -- the chamber name
- `ensemble` -- the ensemble id

For a specific state, chamber, and ensemble combination, all 6 categories of scores—general, partisan, minority, compactness, splitting, and majority-minority (MMD)—are together.

Using `scores_df.parquet`

The information below describes how to load the dataframe from disk and how filter for specific scores.

Loading the Dataframe

To use `scores_df.parquet`, you need to have `pandas` and `pyarrow` installed. You can install them using pip:

```
pip install pandas
pip install pyarrow
```

Then the Python is simple:

```
import os
import pandas as pd

all_scores =
pd.read_parquet(os.path.expanduser("/path/to/scores_df.parquet"))
```

In this example, `all_scores` is a `pandas` dataframe.

Index Columns

You can filter the dataframe using the index columns: `state`, `chamber`, and `ensemble`.

There are 7 states : `FL`, `IL`, `MI`, `NC`, `NY`, `OH`, and `WI`. There are 3 chambers: `congress` and `upper` and `lower` states houses. There are 16 ensembles, the 15 ensembles reported in the paper plus a second reversible ensemble: `A0`, `A1`, `A2`, `A3`, `A4`, `Pop-`, `Pop+`, `B`, `C`, `D`, `Rev⋆`, `Rev`, `R25`, `R50`, `R75`, and `R100`. The `Rev` ensemble corresponds to what is reported in the paper. It has a chain length of 1 billion and a subsampling rate of every 50,000th plan. The `Rev⋆` is the original reversible ensemble that we produced using the same chain length (50 million) and subsampling rate (every 2,500) as the other non-reversible ensembles.