

Schema

The schema of the raw data is

- year
- state
- state_po
- state_fips
- state_cen
- state_ic
- office
- district
- stage
- special
- candidate
- party_detailed
- writein
- mode
- candidatevotes
- totalvotes
- unofficial
- version
- party_simplified

Since the state and party data are repeated in every row, they were separated out into their own tables, replaced by a unique identifier. As a result, the data was split into 3 relations

electioninfo: - year - stateid - office - district - stage - special - candidate - writein - mode - candidatevotes - totalvotes - unofficial - version

stateinfo: - id - state_name - state_po - state_fips - state_cen - state_ic

party: - id - party_name - party_simplified

The full schema with variable types can be seen in the `create_tables.sql` script.

```
CREATE TABLE stateinfo
(
    id SERIAL PRIMARY KEY,
    state_name varchar(64),
    state_po varchar(16),
    state_fips INT,
    state_cen INT,
    state_ic INT
);
```

```
CREATE TABLE party
```

```

(
    id SERIAL PRIMARY KEY,
    party_name VARCHAR(128),
    party_simplified VARCHAR(128)
);

CREATE TABLE electioninfo
(
    -- There are instances
    -- of "unknown" individuals running for the same party in the same state in the same year
    -- where the name is "NA", therefore we create an index instead of using those values as
    id SERIAL PRIMARY KEY,
    office VARCHAR(64),
    stage VARCHAR(16),
    year INT,
    special BOOLEAN,
    candidatename VARCHAR(128),
    partyid INT,
    stateid INT,
    district VARCHAR(64),
    mode VARCHAR(32),
    candidatevotes INT,
    totalvotes INT,
    unofficial BOOLEAN,
    CONSTRAINT state_fk FOREIGN KEY (stateid)
        REFERENCES stateinfo(id),
    CONSTRAINT party_fk FOREIGN KEY (partyid)
        REFERENCES party(id)
);

```

Import process

Create tables

Run `create_tables.sql` to create tables.

Populate tables

Locally:

Since it was not possible to install 3rd party python libraries on school server (`psycopg2`), the database was populated locally then copied over to the school server with the following steps - connected to the local server with `psql - \i create_tables.sql` - The python script was run `python3 load_data.py`

Remote:

A dump of the tables was done on the local database, and copied to the school servers using `scp` by running `dump_and_copy.sh`

Once on the school server, the data was imported by - connecting to the psql server (`psql -U user022 -h studentdb.csc.uvic.ca db_022`) - schema created using `\i create_tables.sql` - data was imported by running `\i db_022.sql`

Note:

A copy of the dump has been kept within this folder, so it is unnecessary to do the `scp` command.