Schema description

- Location_Population schema: (varchar pop_code pkey, text pop_desc)
- Divison schema: (int id pkey, text division_name, fkey int state_id INTO state);
- State schema: (int id pkey, varchar abbr, text name, fkey int region_id INTO region);
- Region schema: (int id pkey, text region_name);
- Agency schema: (ORI id pkey, text agency_name, enum type, fkey int div_id INTO division);
- Incident schema: (int incident_id pkey, date date, enum victim_desc, int adult_victim_count, int juvenile_victim_count, int adult_offender_count, int_juvenile_offender_count, fkey (enum, enum) perp_group INTO offender_desc, array text location, fkey varchar pop_code INTO location_population,array text bias_description, varchar bias_type, array int offense_id FKEY into offense, text array bias_type, enum bias_count);

Table Creation Commands

- For region table: CREATE TABLE(region_id int, region_name text, PRIMARY KEY(region_id));
- For district table: CREATE TABLE(div_id int, div_name text, region_id int, PRIMARY KEY(div_id), FOREIGN KEY(region_id) REFERENCES region(region_id));
- For state table : CREATE TABLE(abbr VARCHAR(2), state_name text, division int, PRIMARY KEY(abbr), FOREIGN KEY(division) REFERENCES division(div_id));
- For agencies table : CREATE TABLE(ori VARCHAR(9), pub_agency_name text, agency_type text, state varchar(2), PRIMARY KEY(ori), FOREIGN KEY(state) REFERENCES state(abbr));

- incidents table : This process involves some extra steps as I needed to change data types to create the text arrays, so I created a tmp table whose values I copied into an incidents table once complete.

```
CREATE TABLE tmp(incident_id int, ori varchar(9), state varchar(2), div_id
int, region_id int, pop_code varchar(2),
incident_date date, adult_victim_count int, juvenile_victim_count int,
total_offender_count int, adult_offender_count int,
juvenile_offender_count int,
offender_race offender_race, offender_ethnicity offender_ethnicity,
offense_name text, total_individual_victims int,
location_name text, bias_desc text, victim_type text, mult_offense
varchar(1), multiple_bias varchar(1),
PRIMARY KEY(incident_id), FOREIGN KEY(ori) REFERENCES agencies(ori),
FOREIGN KEY(state) REFERENCES state(abbr),
FOREIGN KEY(div_id) REFERENCES division(div_id), FOREIGN KEY(region_id)
REFERENCES region(region_id),
FOREIGN KEY(pop_code) REFERENCES population_location(pop_code));
UPDATE tmp SET victim_type = string_to_array(victim_type, ’;’);
alter table tmp alter victim_type drop default;
alter table tmp alter victim_type type text[] using array[victim_type];
alter table tmp alter offense_name drop default;
alter table tmp alter offense_name type text[] using array[offense_name];
CREATE TABLE incidents AS TABLE tmp;
```

How was the database created

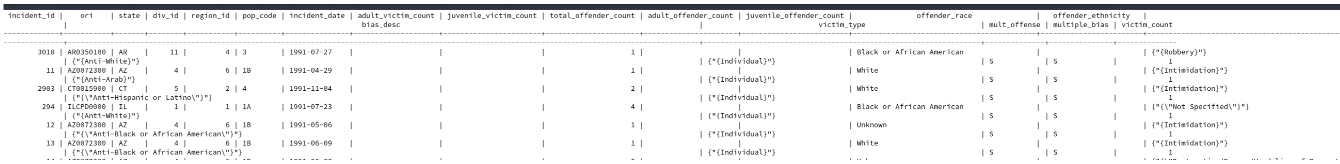
I downloaded the master hate crimes csv from the FBI’s Crime Data Explorer, then I inspected its accompanying README.txt to understand the data better and what relationships or constraints are applied to the data. This includes how data reporting works, changes to the data introduced by law, and an enumeration of options on columns with restricted possible values such as locations, ethnicities, and recognized biases, among a wealth of other information about the data and the caveat to be careful in using the data to make comparisons except against the national due to the data’s complexity and ease of misunderstanding what it means. Then, I set about trying to understand where I had duplicated data so that I could normalize the csv and create additional tables to rely on rather than having a lot of duplicated data. Using a mixture of vlookups and pivot tables, I was able to pull this information out and isolate and create csv's to represent these tables. I set the field I wanted to have treated as dates to a date type and was careful to set Postgres to look for the correct date type as well. I also sanitized nulls in integer columns, as COPY originally wanted to read them as empty strings.

Some tables that I initially thought would make good tables, such as bias description, I eventually decided work better as array types largely as there is no support I could find for having foreign keys as array elements. This is too bad, as bias description is a column with constrained values. Then I uploaded the csv's and copied them into created tables that I had made with the table creations in the previous section of this document. I took additional steps with the incidents table, by far the most complex table in the database. Please see the list of commands I used as I included them along with the table creation information. Attempts to recreate the data will not work without being certain to recreate those text arrays. This is the method I used to do so, although there may be other, more efficient ways. Certainly using more BEGIN and COMMIT statements would have been a good start.

Queries

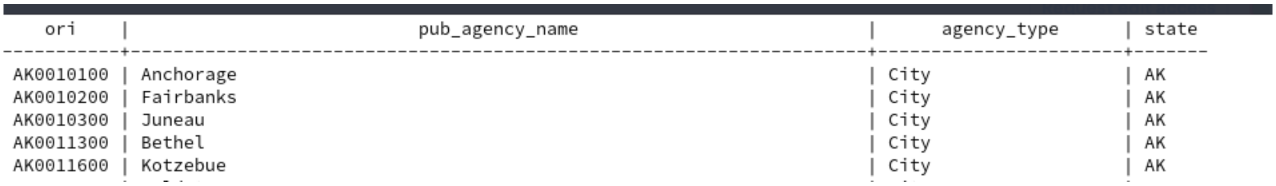
- How many agencies reported hate crimes? select count(distinct ori) from incidents;
- How many hate crimes involved multiple offenses? select count(mult_offense) FROM incidents WHERE mult_offense='M';
- How many hate crimes were committed in Oregon? SELECT COUNT(i.incident_id) FROM incidents i, state s WHERE i.state=s.abbr AND s.state_name='Oregon';

Figures



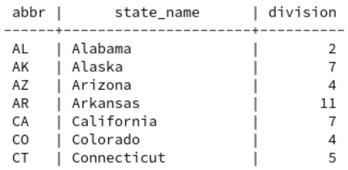
```
incident_id |    ori    | state | div_id | region_id | pop_code | incident_date | adult_victim_count | juvenile_victim_count | total_offender_count | adult_offender_count | juvenile_offender_count |              offender_race              |   offender_ethnicity   |
            |                                                            bias_desc                                                            |                      victim_type                      | mult_offense | multiple_bias | victim_count
------------+-----------+-------+--------+-----------+----------+---------------+--------------------+-----------------------+----------------------+----------------------+-------------------------+-----------------------------------------+------------------------+------------------------------
------------+---------------------------------------------------------------------------------------------------------------------------------+-------------------------------------------------------+--------------+---------------+-------------
       3018 | AR0350100 | AR    |     11 |         4 | 3        | 1991-07-27    |                    |                       |                    1 |                      |                         | Black or African American               |                        | {"{Robbery}"}
            | {"{Anti-White}"}                                                                                                                | {"{Individual}"}                                      | S            | S             |            1
         11 | AZ0072300 | AZ    |      4 |         6 | 1B       | 1991-04-29    |                    |                       |                    1 |                      |                         | White                                   |                        | {"{Intimidation}"}
            | {"{Anti-Arab}"}                                                                                                                 | {"{Individual}"}                                      | S            | S             |            1
       2903 | CT0015900 | CT    |      5 |         2 | 4        | 1991-11-04    |                    |                       |                    2 |                      |                         | White                                   |                        | {"{Intimidation}"}
            | {"{\"Anti-Hispanic or Latino\"}"}                                                                                               | {"{Individual}"}                                      | S            | S             |            1
        294 | ILCPD0000 | IL    |      1 |         1 | 1A       | 1991-07-23    |                    |                       |                    4 |                      |                         | Black or African American               |                        | {"{\"Not Specified\"}"}
            | {"{Anti-White}"}                                                                                                                | {"{Individual}"}                                      | S            | S             |            1
         12 | AZ0072300 | AZ    |      4 |         6 | 1B       | 1991-05-06    |                    |                       |                    1 |                      |                         | Unknown                                 |                        | {"{Intimidation}"}
            | {"{\"Anti-Black or African American\"}"}                                                                                        | {"{Individual}"}                                      | S            | S             |            1
         13 | AZ0072300 | AZ    |      4 |         6 | 1B       | 1991-06-09    |                    |                       |                    1 |                      |                         | White                                   |                        | {"{Intimidation}"}
            | {"{\"Anti-Black or African American\"}"}                                                                                        | {"{Individual}"}                                      | S            | S             |            1
```

Figure 1: Incidents table.



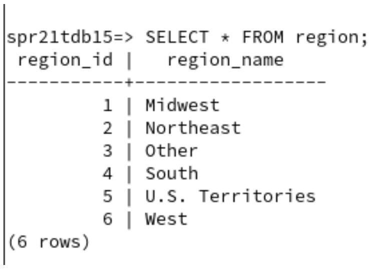
```
    ori    |                        pub_agency_name                        |      agency_type       | state
-----------+---------------------------------------------------------------+------------------------+-------
 AK0010100 | Anchorage                                                     | City                   | AK
 AK0010200 | Fairbanks                                                     | City                   | AK
 AK0010300 | Juneau                                                        | City                   | AK
 AK0011300 | Bethel                                                        | City                   | AK
 AK0011600 | Kotzebue                                                      | City                   | AK
```

Figure 2: Agencies table.



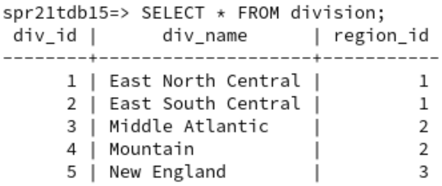
```
 abbr |      state_name      | division
------+----------------------+----------
 AL   | Alabama              |        2
 AK   | Alaska               |        7
 AZ   | Arizona              |        4
 AR   | Arkansas             |       11
 CA   | California           |        7
 CO   | Colorado             |        4
 CT   | Connecticut          |        5
```

Figure 3: State table.



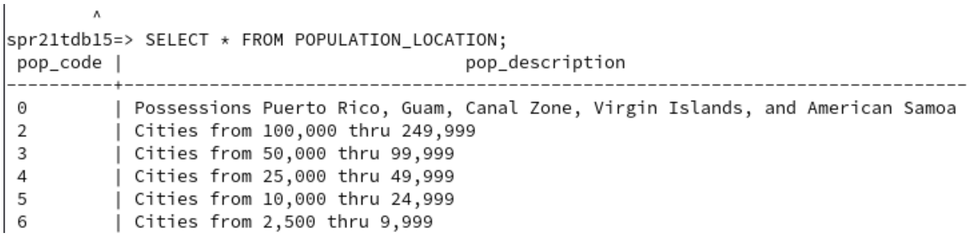
```
spr21tdb15=> SELECT * FROM region;
 region_id |   region_name
-----------+------------------
         1 | Midwest
         2 | Northeast
         3 | Other
         4 | South
         5 | U.S. Territories
         6 | West
(6 rows)
```

Figure 4: Region table.



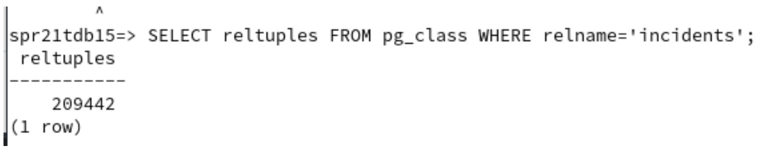
```
spr21tdb15=> SELECT * FROM division;
 div_id |      div_name      | region_id
--------+--------------------+-----------
      1 | East North Central |         1
      2 | East South Central |         1
      3 | Middle Atlantic    |         2
      4 | Mountain           |         2
      5 | New England        |         3
```

Figure 5: Division table.



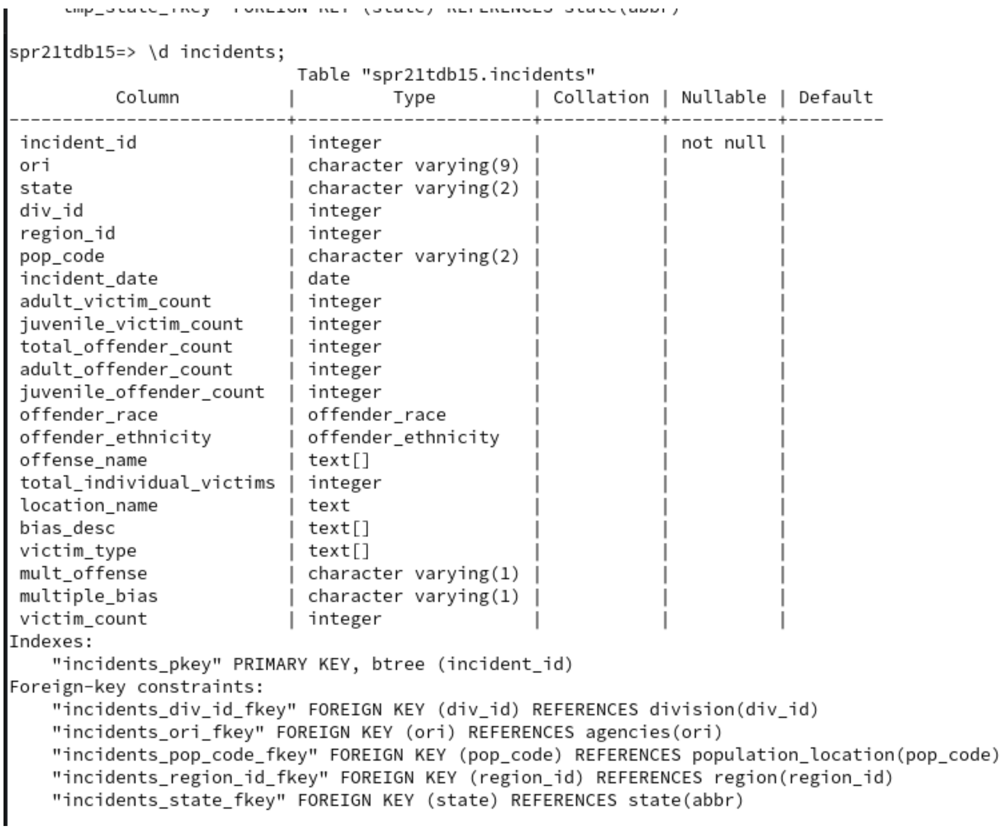
```
        ^
spr21tdb15=> SELECT * FROM POPULATION_LOCATION;
 pop_code |                              pop_description
----------+--------------------------------------------------------------------------------
 0        | Possessions Puerto Rico, Guam, Canal Zone, Virgin Islands, and American Samoa
 2        | Cities from 100,000 thru 249,999
 3        | Cities from 50,000 thru 99,999
 4        | Cities from 25,000 thru 49,999
 5        | Cities from 10,000 thru 24,999
 6        | Cities from 2,500 thru 9,999
```

Figure 6: Population location table.



```
        ^
spr21tdb15=> SELECT reltuples FROM pg_class WHERE relname='incidents';
 reltuples
-----------
    209442
(1 row)
```

Figure 7: Number of tuples in incidents.



```
    tmp_state_fkey" FOREIGN KEY (state) REFERENCES state(abbr)

spr21tdb15=> \d incidents;
                           Table "spr21tdb15.incidents"
         Column          |         Type          | Collation | Nullable | Default
-------------------------+-----------------------+-----------+----------+---------
 incident_id             | integer               |           | not null |
 ori                     | character varying(9)  |           |          |
 state                   | character varying(2)  |           |          |
 div_id                  | integer               |           |          |
 region_id               | integer               |           |          |
 pop_code                | character varying(2)  |           |          |
 incident_date           | date                  |           |          |
 adult_victim_count      | integer               |           |          |
 juvenile_victim_count   | integer               |           |          |
 total_offender_count    | integer               |           |          |
 adult_offender_count    | integer               |           |          |
 juvenile_offender_count | integer               |           |          |
 offender_race           | offender_race         |           |          |
 offender_ethnicity      | offender_ethnicity    |           |          |
 offense_name            | text[]                |           |          |
 total_individual_victims | integer              |           |          |
 location_name           | text                  |           |          |
 bias_desc               | text[]                |           |          |
 victim_type             | text[]                |           |          |
 mult_offense            | character varying(1)  |           |          |
 multiple_bias           | character varying(1)  |           |          |
 victim_count            | integer               |           |          |
Indexes:
    "incidents_pkey" PRIMARY KEY, btree (incident_id)
Foreign-key constraints:
    "incidents_div_id_fkey" FOREIGN KEY (div_id) REFERENCES division(div_id)
    "incidents_ori_fkey" FOREIGN KEY (ori) REFERENCES agencies(ori)
    "incidents_pop_code_fkey" FOREIGN KEY (pop_code) REFERENCES population_location(pop_code)
    "incidents_region_id_fkey" FOREIGN KEY (region_id) REFERENCES region(region_id)
    "incidents_state_fkey" FOREIGN KEY (state) REFERENCES state(abbr)
```

Figure 8: Describe incidents.