Data Description and Sourcing:

Mortality: Vital Statistics tables that contain aggregate metrics on the mortality (deaths) of Philadelphia residents. Included in these datasets are mortality metrics by planning district. This data was obtained from OpenDataPhilly.
The original table overview is included below in Fig 1A. And Fig 1B.:

gid	geom	objectid	geography_name	year	geography	sex	race_ethnicity	age_category
1	POLYGON ((-75.147912265285 39.9673343870	1	Central	2,018	Planning district	All sexes	All races/ethnicities	All ages
2	POLYGON ((-75.147912265285 39.9673343870	2	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
3	POLYGON ((-75.147912265285 39.9673343870	3	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
4	POLYGON ((-75.147912265285 39.9673343870	4	Central	2,015	Planning district	All sexes	All races/ethnicities	All ages
5	POLYGON ((-75.147912265285 39.9673343870	5	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
6	POLYGON ((-75.147912265285 39.9673343870	6	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
7	POLYGON ((-75.147912265285 39.9673343870	7	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
8	POLYGON ((-75.147912265285 39.9673343870	8	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
9	POLYGON ((-75.147912265285 39.9673343870	9	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
10	POLYGON ((-75.147912265285 39.9673343870	10	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
11	POLYGON ((-75.147912265285 39.9673343870	11	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
12	POLYGON ((-75.147912265285 39.9673343870	12	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
13	POLYGON ((-75.147912265285 39.9673343870	13	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
14	POLYGON ((-75.147912265285 39.9673343870	14	Central	2,014	Planning district	All sexes	All races/ethnicities	35+
15	POLYGON ((-75.147912265285 39.9673343870	15	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
16	POLYGON ((-75.147912265285 39.9673343870	16	Central	2,018	Planning district	All sexes	All races/ethnicities	All ages
17	POLYGON ((-75.147912265285 39.9673343870	17	Central	2,017	Planning district	All sexes	All races/ethnicities	<1
18	POLYGON ((-75.147912265285 39.9673343870	18	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
19	POLYGON ((-75.147912265285 39.9673343870	19	Central	2,017	Planning district	All sexes	All races/ethnicities	<1
20	POLYGON ((-75.147912265285 39.9673343870	20	Central	2,017	Planning district	All sexes	All races/ethnicities	<1
21	POLYGON ((-75.147912265285 39.9673343870	21	Central	2,017	Planning district	Female	All races/ethnicities	At birth
22	POLYGON ((-75.147912265285 39.9673343870	22	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages
23	POLYGON ((-75.147912265285 39.9673343870	23	Central	2,016	Planning district	All sexes	All races/ethnicities	All ages

Fig 1A.

age_category	leading_cause_death	metric_name	metric_value	quality_flag	shape__area	shape__length
All ages	Prostate cancer	age_adjusted_mortality_rate_per_100k	9.30541842	unreliable	28,221,330.1835938	28,401.8230651796
All ages	Intentional self-harm (suici	count_deaths	13	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	HIV/AIDS	count_deaths	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
All ages	Chronic kidney disease	count_deaths	15	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	HIV/AIDS	age_adjusted_mortality_rate_per_100k	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
All ages	Influenza and pneumonia	count_deaths	18	[NULL]	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable canc	age_adjusted_smoking_attributable_m	85.12155938	[NULL]	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable cardi	smoking_attributable_deaths	39.29683745	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	Homicide	age_adjusted_mortality_rate_per_100k	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable canc	smoking_attributable_deaths	50.2549263	[NULL]	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable pulm	age_adjusted_smoking_attributable_m	46.64143828	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	Diabetes	age_adjusted_mortality_rate_per_100k	9.16613196	unreliable	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable cardi	age_adjusted_smoking_attributable_m	69.08884382	[NULL]	28,221,330.1835938	28,401.8230651796
35+	Smoking-attributable pulm	smoking_attributable_deaths	27.24582455	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	Heart disease	age_adjusted_mortality_rate_per_100k	145.26540333	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	Cerebrovascular diseases	age_adjusted_mortality_rate_per_100k	34.41085062	[NULL]	28,221,330.1835938	28,401.8230651796
<1	All causes	count_perinatal_deaths	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
All ages	Homicide	count_deaths	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
<1	All causes	count_infant_deaths	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
<1	All causes	infant_mortality_rate_per_1k_live_births	-99,999	suppressed	28,221,330.1835938	28,401.8230651796
At birth	All causes	life_expectancy_at_birth	83.56699946	[NULL]	28,221,330.1835938	28,401.8230651796
All ages	Influenza and pneumonia	age_adjusted_mortality_rate_per_100k	15.54433756	unreliable	28,221,330.1835938	28,401.8230651796
All ages	Intentional self-harm (suici	age_adjusted_mortality_rate_per_100k	8.93869939	unreliable	28,221,330.1835938	28,401.8230651796

Fig 1B.

The data used for the analytical queries was a general census tracts shape dataset for the city of Philadelphia, and the crime incidents data from the Philadelphia Police Department. Part I crimes include violent offenses such as aggravated assault, rape, arson, among others. The crimes dataset used in the aggregate queries is specifically from the year 2019.

The mortality_metric data in combination with census and crime data yields several interesting queries. It can yield select census tracts with the highest metrics of a specific cause of death such as count or age adjusted mortality rate per 100k for lung cancer. Another question can be a name and crime count per district with the highest age adjusted homicide mortality rate per 100k for all ages and all sexes.

Data Structure and Normalization:

Data Structure: The data itself fits to be structured in an Entity-Attribute-Value model or Object-Attribute-Value model. This means that the entity itself- the planning district existing in a year 2012-2019 has a number of attributes describing it and a single value assigned to it. The attributes describing the entity (redundantly to objectid represented by gid) in the original set are: objectid geom, geography_name, sex, race_ethnicity, age_category, leading_cause_death, metric_name, shape_area, shape_length. Only a few of these attributes actually apply to the entity, which is another characteristic of a dataset fit for EAV modeling. The value describing the entity is the metric_value.

Normalization:

The normalization process for this dataset involved using a pivot table process. This means a transformation into one column per attribute notation, or simply turning the rows with all the attributes into columns representing values for each set of attributes per entity. The entity itself is the planning district and year in which the mortality data about it was collected.

planning_district table: Some of the attributes of the planning district are geometry based- to store these attributes and avoid redundancies in the data (which would violate 2nd NF) I created a planning_district table. This table stores data about the planning district name ‘geography_name’, its geometry ‘geom’, and creates a serial primary key ‘geo_id’ for reference in the main mortality metric table. There are a total of 18 planning districts.

year: The year itself stands alone and is used in the creation of a composite key for the mortality_metric table described below.

mortality_metric table: The primary key for the main mortality metric table is a composite key made up of the year and geo_id of the planning district a specific metric was taken. The metric columns are through an aggregate query using the GROUP BY clause to create one row per entity, and a CASE WHEN for finding the metric for a specific set of attributes. After filtering the attributes for values that added no clearly significant information to the dataset (i.e. the geography attribute values were all ‘planning district’), I create each column to represent metric values for: ‘sex,’ ‘age_category,’ ‘leading_cause_death,’ and ‘metric_name’. In the end there were 50 sets of distinct attribute category combinations.

Mortality_column_reference table: Since I used abbreviations to represent the 50 unique column names, I created an additional table as a way to reference these names. This table is represented below in Fig 2.

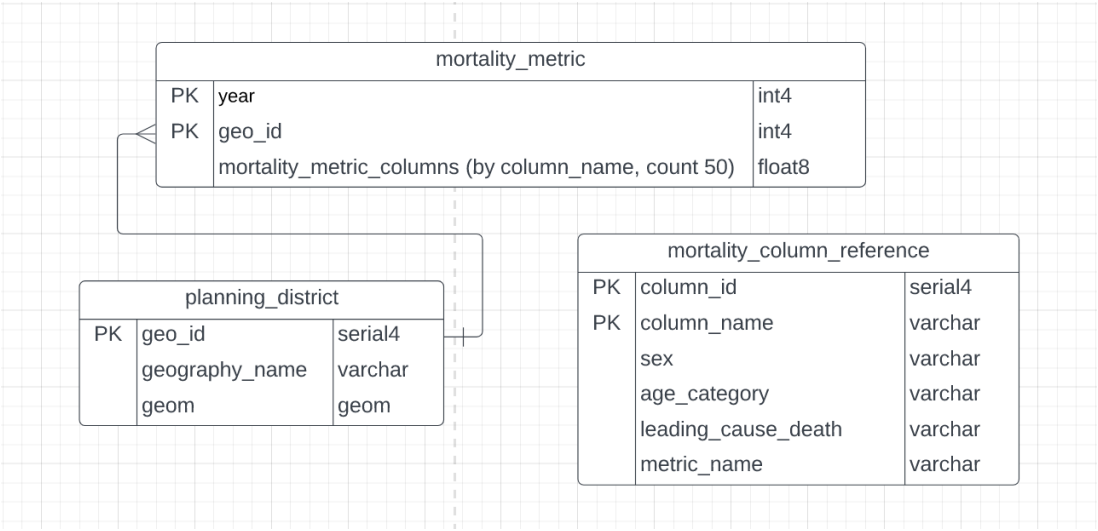

Fig 2. ERD

Optimization:

There is not much room for optimization in this dataset, it's only 144 rows of information. In the optimization milestone I did create indexes for year and geo_id, but it didn't help much with the runtime. I also denormalized the dataset to include 'geom' in the mortality_metric table, to avoid an extra join with the planning_district table but it actually increased the run time, which may have been an issue with my computer after running multiple queries in a short span of time.

Spatial Queries:

A. Return the census tracts with the highest count of deaths from lung cancer in descending order: 2.5s

```
select
c.id,
c.namelsad10,
c.geom,
p.geography_name,
max(m.A_AA_LC_count_d),
m.year
from census c
join planning_district p on st_within(st_transform(c.geom, 2272), p.geom)
join mortality_metric m on p.geo_id=m.geo_id
group by c.id, p.geography_name, m.year order by max desc;
```

A simple retrieval query with a spatial join on ST_Within and CRS transformation with ST_Transform. ST_Within returns TRUE if geometry A is completely inside geometry B. St_Transform returns a new geometry with its CRS changed to a different spatial reference system. Results visualized in Fig 3.

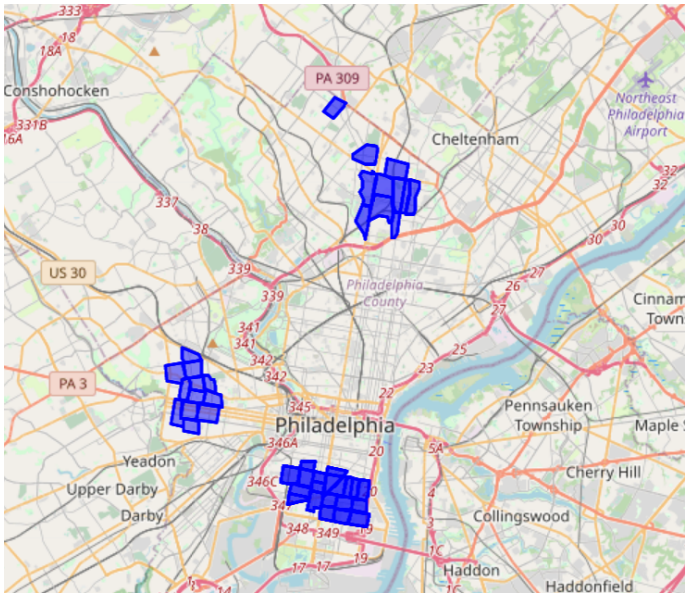

Fig 3.

B. Return the names and crime counts for the 3 planning districts with the highest age adjusted homicide mortality rate per 100k for all ages and all sexes. OG RUN TIME: 14s
NEW RUN TIME WITH : 9s

```
select geography_name, count(c.objectid), year, max(A_AA_H_aamr_100k)
from crime_geom c
join planning_district p on st_within(c.geom, p.geom)
join mortality_metric m on p.geo_id=m.geo_id
group by geography_name, A_AA_H_aamr_100k, year
having year = 2019 order by max desc limit 3;
```

Creating the EAV allows for complex data retrieval by category with less effort. Here again we have a spatial join with ST_Within, which allows for retrieval of data points contained completely within the specified planning districts. In this case the specified planning districts are the 3 (limit 3) with the highest max() age adjusted homicide mortality rate per 100k for all ages and all sexes. There is also an attribute join on geo_id, and a having clause- specifying the target A_AA_H_aamr_100k metric being measured in the year 2019. The crime data is from the year 2019.

C. Return each district selected with its crime count, homicide count, and nearest planning district.
OG RUN TIME: 17s | NEW RUN TIME: 12s

```
select pd.geo_id as district, count(objectid) as district_crime_count,
m.A_AA_H_count_d as homocide_count,
(select p.geo_id from planning_district as p where pd.geo_id != p.geo_id order by
p.geom <-> pd.geom limit 1)
as closest_district
from planning_district as pd
join crime_geom c on st_within(c.geom, pd.geom)
join mortality_metric m on pd.geo_id=m.geo_id
group by m.year, pd.geo_id, m.a_aa_h_count_d
having m.year = 2019;
```

A selection with added aliases. Showcases a nearest neighbor analysis using a KNN operator in a correlated subquery. KNN is a bounding-box centroid distance operator. A <-> B returns the distance between the centroids of the bounding boxes of A and B. So this select statement finds the planning district with the centroid nearest to it (may not be touching). pd.geo_id != p.geo_id ensures that the entry of the nearest neighboring planning district is not itself.

Appendix A:

vital_mortality_pd data available for download with this link:
https://opendata.arcgis.com/api/v3/datasets/f2f78582a66e441aa6fa5b900c63a9cb_0/downloads/data?format=csv&spatialRefId=4326

```
C:\Users\tup93308\Documents\Fall 2022\Spatial Database
Design\FINAL>ogr2ogr -f PostgreSQL PG:"host=localhost port=5432
dbname=universe user=postgres password=postgres" -lco
SCHEMA=final_project -lco PRECISION=NO -oo
EMPTY_STRING_AS_NULL=YES vital_mortality_pd.csv
```

crime_incident data available for download with this link:
https://phl.carto.com/api/v2/sql?filename=incidents_part1_part2&format=csv&skipfields=cartodb_id,the_geom,the_geom_webmercator&q=SELECT%20*%20,%20ST_Y(the_geom)%20AS%

20lat,%20ST_X(the_geom)%20AS%20lng%20FROM%20incidents_part1_part2%20WHERE%20dispatch_date_time%20%3E=%20%272019-01-01%27%20AND%20dispatch_date_time%20%3C%20%272020-01-01%27

```
C:\Users\tup93308\Documents\Fall 2022\Spatial Database Design\FINAL>ogr2ogr -f PostgreSQL PG:"host=localhost port=5432 dbname=universe user=postgres password=postgres" -lco SCHEMA=streetpole -lco PRECISION=NO -oo EMPTY_STRING_AS_NULL=YES crime_incident.csv
```

Appendix B:

```
 --general column_category table


CREATE TABLE final_project.mortality_column_category (
    column_id serial PRIMARY KEY,
    sex varchar,
    age_category varchar,
    leading_cause_death varchar,
    metric_name varchar);

INSERT INTO mortality_column_category (sex, age_category,
leading_cause_death, metric_name)
select distinct sex, age_category, leading_cause_death, metric_name
from vital_mortality_pd;

-------------------------------------------------
--column name and code table
DROP TABLE IF EXISTS mortality_column_name CASCADE;

CREATE TABLE final_project.mortality_column_name (
    column_id serial PRIMARY KEY,
    column_name varchar);

INSERT INTO mortality_column_name (column_name)
values ('A_AA_LC_count_d'),
('A_35plus_SAPD_smoking_d'),
('A_35plus_SAPD_smoking_aamr_100k'),
('A_AA_CRVD_count_d'),
('A_AA_HA_aamr_100k'),
('A_less1_ALL_count_id'),
('A_35plus_ALLSA_smoking_aamr_100k'),
('A_AA_Unintentional_aamr_100k'),
('A_less1_ALL_pren_mr_1k_lb'),
('A_35plus_SAC_smoking_d'),
('A_35plus_SAC_smoking_aamr_100k'),
('A_AA_Sep_aamr_100k'),
('A_AA_CC_aamr_100k'),
```

('A_AA_CKD_count_d'),
('A_AA_ALL_aamr_100k6'),
('F_AB_ALL_life_exp_ab'),
('A_AA_ALL_count_d'),
('A_AA_Sep_count_d'),
('A_AA_IP_count_d'),
('A_AA_OD_count_d'),
('A_AA_CRVD_aamr_100k'),
('A_AA_H_aamr_100k'),
('A_AA_CLRD_aamr_100k'),
('A_AA_CLRD_count_d'),
('A_AA_SH_count_d'),
('A_AA_PC_aamr_100k'),
('A_35plus_ALLSA_smoking_d'),
('A_AA_D_aamr_100k'),
('A_AA_HA_count_d'),
('A_AA_IP_aamr_100k'),
('A_AA_PC_count_d'),
('A_less1_ALL_count_d'),
('A_AA_C_count_d'),
('A_35plus_SACMD_smoking_d'),
('M_AB_ALL_life_exp_ab'),
('A_AA_HD_count_d'),
('A_AA_H_count_d'),
('A_AA_OD_aamr_100k'),
('A_AA_HD_aamr_100k'),
('A_AA_BC_aamr_100k'),
('A_AA_CC_count_d'),
('A_AA_Unintentional_count_d'),
('A_less1_ALL_inf_mr_1k_lb'),
('A_35plus_SACMD_smoking_aamr_100k'),
('A_AA_BC_count_d'),
('A_AA_C_aamr_100k'),
('A_AA_CKD_aamr_100k'),
('A_AA_SH_aamr_100k'),
('A_AA_D_count_d'),
('A_AA_LC_aamr_100k');

```
--Final column name reference table-- can be used to reference codified column
--names using specified select queries
--example "select column_name, metric_name from mortality_column_reference
--where leading_cause_death = 'HIV/AIDS';"
--drop table if exists mortality_column_reference cascade;

CREATE TABLE final_project.mortality_column_reference (
    column_id serial PRIMARY KEY,
    column_name varchar,
    sex varchar,
    age_category varchar,
    leading_cause_death varchar,
    metric_name varchar);

insert into mortality_column_reference (column_name, sex, age_category,
leading_cause_death, metric_name)
select column_name, sex, age_category, leading_cause_death, metric_name
from mortality_column_name as n join mortality_column_category as c on
n.column_id=c.column_id;

--GEOMETRY TABLE

CREATE TABLE final_project.planning_district (
    geo_id serial PRIMARY KEY,
    geography_name varchar,
    geom geometry);

INSERT INTO planning_district (geography_name, geom)
select distinct geography_name, ST_Transform(geom, 2272)
from vital_mortality_pd;

-----------------------------------
--MORTALITY TABLE

CREATE TABLE final_project.mortality_metric
(year int,
geo_id int,
A_AA_LC_count_d float8,
A_35plus_SAPD_smoking_d float8,
A_35plus_SAPD_smoking_aamr_100k float8,
```

A_AA_CRVD_count_d float8,
A_AA_HA_aamr_100k float8,
A_less1_ALL_count_id float8,
A_35plus_ALLSA_smoking_aamr_100k float8,
A_AA_Unintentional_aamr_100k float8,
A_less1_ALL_pren_mr_1k_lb float8,
A_35plus_SAC_smoking_d float8,
A_35plus_SAC_smoking_aamr_100k float8,
A_AA_Sep_aamr_100k float8,
A_AA_CC_aamr_100k float8,
A_AA_CKD_count_d float8,
A_AA_ALL_aamr_100k float8,
F_AB_ALL_life_exp_ab float8,
A_AA_ALL_count_d float8,
A_AA_Sep_count_d float8,
A_AA_IP_count_d float8,
A_AA_OD_count_d float8,
A_AA_CRVD_aamr_100k float8,
A_AA_H_aamr_100k float8,
A_AA_CLRD_aamr_100k float8,
A_AA_CLRD_count_d float8,
A_AA_SH_count_d float8,
A_AA_PC_aamr_100k float8,
A_35plus_ALLSA_smoking_d float8,
A_AA_D_aamr_100k float8,
A_AA_HA_count_d float8,
A_AA_IP_aamr_100k float8,
A_AA_PC_count_d float8,
A_less1_ALL_count_d float8,
A_AA_C_count_d float8,
A_35plus_SACMD_smoking_d float8,
M_AB_ALL_life_exp_ab float8,
A_AA_HD_count_d float8,
A_AA_H_count_d float8,
A_AA_OD_aamr_100k float8,
A_AA_HD_aamr_100k float8,
A_AA_BC_aamr_100k float8,
A_AA_CC_count_d float8,
A_AA_Unintentional_count_d float8,
A_less1_ALL_inf_mr_1k_lb float8,

```
A_35plus_SACMD_smoking_aamr_100k float8,
A_AA_BC_count_d float8,
A_AA_C_aamr_100k float8,
A_AA_CKD_aamr_100k float8,
A_AA_SH_aamr_100k float8,
A_AA_D_count_d float8,
A_AA_LC_aamr_100k float8,
primary key (year, geo_id));

insert into mortality_metric
select distinct vm.year, pd.geo_id,
        MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Lung cancer' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_LC_count_d,
        MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable pulmonary diseases' and
metric_name = 'smoking_attributable_deaths' THEN metric_value END)
A_35plus_SAPD_smoking_d,
        MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable pulmonary diseases' and
metric_name = 'age_adjusted_smoking_attributable_mortality_rate_per_100k'
THEN metric_value END) A_35plus_SAPD_smoking_aamr_100k,
        MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Cerebrovascular diseases' and metric_name =
'count_deaths' THEN metric_value END) A_AA_CRVD_count_d,
        MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'HIV/AIDS' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_HA_aamr_100k,
        MAX(CASE WHEN sex ='All sexes' and age_category = '<1' and
leading_cause_death = 'All causes' and metric_name = 'count_infant_deaths'
THEN metric_value END) A_less1_ALL_count_id,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'All smoking-attributable causes' and metric_name =
'age_adjusted_smoking_attributable_mortality_rate_per_100k' THEN
metric_value END) A_35plus_ALLSA_smoking_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Unintentional injuries (excluding drug overdoses)' and
metric_name = 'age_adjusted_mortality_rate_per_100k' THEN metric_value
END) A_AA_Unintentional_aamr_100k,
```

```
    MAX(CASE WHEN sex ='All sexes' and age_category = '<1' and
leading_cause_death = 'All causes' and metric_name =
'perinatal_mortality_rate_per_1k_live_births' THEN metric_value END)
A_less1_ALL_pren_mr_1k_lb,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable cancers' and metric_name =
'smoking_attributable_deaths' THEN metric_value END)
A_35plus_SAC_smoking_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable cancers' and metric_name =
'age_adjusted_smoking_attributable_mortality_rate_per_100k' THEN
metric_value END) A_35plus_SAC_smoking_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Septicemia' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_Sep_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Colorectal cancer' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_CC_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Chronic kidney disease' and metric_name =
'count_deaths' THEN metric_value END) A_AA_CKD_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'All causes' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_ALL_aamr_100k,
        MAX(CASE WHEN sex ='Female' and age_category = 'At birth' and
leading_cause_death = 'All causes' and metric_name = 'life_expectancy_at_birth'
THEN metric_value END) F_AB_ALL_life_exp_ab,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'All causes' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_ALL_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Septicemia' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_Sep_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Influenza and pneumonia' and metric_name =
'count_deaths' THEN metric_value END) A_AA_IP_count_d,
```

```
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Drug overdoses' and metric_name = 'count_deaths'
THEN metric_value END) A_AA_OD_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Cerebrovascular diseases' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_CRVD_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Homicide' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_H_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Chronic lower respiratory diseases' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_CLRD_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Chronic lower respiratory diseases' and metric_name =
'count_deaths' THEN metric_value END) A_AA_CLRD_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Intentional self-harm (suicide)' and metric_name =
'count_deaths' THEN metric_value END) A_AA_SH_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Prostate cancer' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_PC_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'All smoking-attributable causes' and metric_name =
'smoking_attributable_deaths' THEN metric_value END)
A_35plus_ALLSA_smoking_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Diabetes' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_D_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'HIV/AIDS' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_HA_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Influenza and pneumonia' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_IP_aamr_100k,
```

```
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Prostate cancer' and metric_name = 'count_deaths'
THEN metric_value END) A_AA_PC_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = '<1' and
leading_cause_death = 'All causes' and metric_name = 'count_perinatal_deaths'
THEN metric_value END) A_less1_ALL_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Cancer' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_C_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable cardiovascular and metabolic
diseases' and metric_name = 'smoking_attributable_deaths' THEN metric_value
END) A_35plus_SACMD_smoking_d,
    MAX(CASE WHEN sex ='Male' and age_category = 'At birth' and
leading_cause_death = 'All causes' and metric_name = 'life_expectancy_at_birth'
THEN metric_value END) M_AB_ALL_life_exp_ab,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Heart disease' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_HD_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Homicide' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_H_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Drug overdoses' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_OD_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Heart disease' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_HD_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Breast cancer' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_BC_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Colorectal cancer' and metric_name = 'count_deaths'
THEN metric_value END) A_AA_CC_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Unintentional injuries (excluding drug overdoses)' and
```

```
metric_name = 'count_deaths' THEN metric_value END)
A_AA_Unintentional_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = '<1' and
leading_cause_death = 'All causes' and metric_name =
'infant_mortality_rate_per_1k_live_births' THEN metric_value END)
A_less1_ALL_inf_mr_1k_lb,
    MAX(CASE WHEN sex ='All sexes' and age_category = '35+' and
leading_cause_death = 'Smoking-attributable cardiovascular and metabolic
diseases' and metric_name =
'age_adjusted_smoking_attributable_mortality_rate_per_100k' THEN
metric_value END) A_35plus_SACMD_smoking_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Breast cancer' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_BC_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Cancer' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_C_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Chronic kidney disease' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_CKD_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Intentional self-harm (suicide)' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_SH_aamr_100k,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Diabetes' and metric_name = 'count_deaths' THEN
metric_value END) A_AA_D_count_d,
    MAX(CASE WHEN sex ='All sexes' and age_category = 'All ages' and
leading_cause_death = 'Lung cancer' and metric_name =
'age_adjusted_mortality_rate_per_100k' THEN metric_value END)
A_AA_LC_aamr_100k
from vital_mortality_pd vm join planning_district pd on
vm.geography_name=pd.geography_name
group by year, geo_id;
```

Appendix C:

Optimizing crime_incident data for faster results:

```
drop table if exists crime_geom cascade;
create table final_project.crime_geom
   (geo_id serial PRIMARY KEY,
   objectid int4,
   geom geometry(POINT, 2272));

INSERT INTO crime_geom (objectid, geom)
        select distinct objectid, ST_Transform(ST_SetSRID(ST_Point(point_x,
        point_y), 4326), 2272)
from crime_incident;
```

Creating a geometry table for faster look-up: point creation with st_point from x, and y data, setting SRID for the original data source with ST_SetSRID, Transforming the data into desired coordinate system with ST)Transform.

Indexing:

```
create index geo_id_mm_idx
on mortality_metric (geo_id);

create index year_idx
on mortality_metric (year);

CREATE index district_geom_idx ON planning_district
USING gist (ST_Transform(geom, 2272));

CREATE index crime_geom_idx ON crime_geom
USING gist (ST_Transform(geom, 2272));
```

Appendix D:

Planning district:

geo_id serial4 PK

geography_name varchar name of planning district

geom geometry geometry shape

Mortality metric:

year int4 PK year for the metric

geo_id int4 PK

A_AA_LC_count_d float4

A_35plus_SAPD_smoking_d float4

A_35plus_SAPD_smoking_aamr_100k float4

A_AA_CRVD_count_d float4

A_AA_HA_aamr_100k float4

A_less1_ALL_count_id float4

A_35plus_ALLSA_smoking_aamr_100k float4

A_AA_Unintentional_aamr_100k float4

A_less1_ALL_pren_mr_1k_lb float4

A_35plus_SAC_smoking_d float4

A_35plus_SAC_smoking_aamr_100k float4

A_AA_Sep_aamr_100k float4

A_AA_CC_aamr_100k float4

A_AA_CKD_count_d float4

A_AA_ALL_aamr_100k float4

F_AB_ALL_life_exp_ab float4

A_AA_ALL_count_d float4

A_AA_Sep_count_d float4

A_AA_IP_count_d float4

A_AA_OD_count_d float4

A_AA_CRVD_aamr_100k float4

A_AA_H_aamr_100k float4

A_AA_CLRD_aamr_100k float4

A_AA_CLRD_count_d float4

A_AA_SH_count_d float4

A_AA_PC_aamr_100k float4

A_35plus_ALLSA_smoking_d float4

A_AA_D_aamr_100k float4

A_AA_HA_count_d float4

A_AA_IP_aamr_100k float4

A_AA_PC_count_d float4

A_less1_ALL_count_d float4

A_AA_C_count_d float4

A_35plus_SACMD_smoking_d float4

M_AB_ALL_life_exp_ab float4

A_AA_HD_count_d float4

A_AA_H_count_d float4

A_AA_OD_aamr_100k float4

A_AA_HD_aamr_100k float4

A_AA_BC_aamr_100k float4

A_AA_CC_count_d float4

A_AA_Unintentional_count_d float4

A_less1_ALL_inf_mr_1k_lb float4

A_35plus_SACMD_smoking_aamr_100k float4

A_AA_BC_count_d float4

A_AA_C_aamr_100k float4

A_AA_CKD_aamr_100k float4

A_AA_SH_aamr_100k float4

A_AA_D_count_d float4

A_AA_LC_aamr_100k float4

column_id	column_name	sex	age_category	leading_cause_death	metric_name
1	A_AA_LC_count_d	All sexes	All ages	Lung cancer	count_deaths
2	A_35plus_SAPD_sm	All sexes	35+	Smoking-attributable pulmonary diseases	smoking_attributable_deaths
3	A_35plus_SAPD_sm	All sexes	35+	Smoking-attributable pulmonary diseases	age_adjusted_smoking_attributable_mortality_rate_per_100k
4	A_AA_CRVD_count_	All sexes	All ages	Cerebrovascular diseases	count_deaths
5	A_AA_HA_aamr_100	All sexes	All ages	HIV/AIDS	age_adjusted_mortality_rate_per_100k
6	A_less1_ALL_count_i	All sexes	<1	All causes	count_infant_deaths
7	A_35plus_ALLSA_sm	All sexes	35+	All smoking-attributable causes	age_adjusted_smoking_attributable_mortality_rate_per_100k
8	A_AA_Unintentional_	All sexes	All ages	Unintentional injuries (excluding drug overdoses)	age_adjusted_mortality_rate_per_100k
9	A_less1_ALL_pren_m	All sexes	<1	All causes	perinatal_mortality_rate_per_1k_live_births
10	A_35plus_SAC_smok	All sexes	35+	Smoking-attributable cancers	smoking_attributable_deaths
11	A_35plus_SAC_smok	All sexes	35+	Smoking-attributable cancers	age_adjusted_smoking_attributable_mortality_rate_per_100k
12	A_AA_Sep_aamr_10(	All sexes	All ages	Septicemia	age_adjusted_mortality_rate_per_100k
13	A_AA_CC_aamr_100	All sexes	All ages	Colorectal cancer	age_adjusted_mortality_rate_per_100k
14	A_AA_CKD_count_d	All sexes	All ages	Chronic kidney disease	count_deaths
15	A_AA_ALL_aamr_10(	All sexes	All ages	All causes	age_adjusted_mortality_rate_per_100k
16	F_AB_ALL_life_exp_at	Female	At birth	All causes	life_expectancy_at_birth
17	A_AA_ALL_count_d	All sexes	All ages	All causes	count_deaths
18	A_AA_Sep_count_d	All sexes	All ages	Septicemia	count_deaths
19	A_AA_IP_count_d	All sexes	All ages	Influenza and pneumonia	count_deaths
20	A_AA_OD_count_d	All sexes	All ages	Drug overdoses	count_deaths
21	A_AA_CRVD_aamr_1	All sexes	All ages	Cerebrovascular diseases	age_adjusted_mortality_rate_per_100k
22	A_AA_H_aamr_100k	All sexes	All ages	Homicide	age_adjusted_mortality_rate_per_100k
23	A_AA_CLRD_aamr_1(	All sexes	All ages	Chronic lower respiratory diseases	age_adjusted_mortality_rate_per_100k
24	A_AA_CLRD_count_c	All sexes	All ages	Chronic lower respiratory diseases	count_deaths

Fig 4 A.

column_id	column_name	sex	age_category	leading_cause_death	metric_name
25	A_AA_SH_count_d	All sexes	All ages	Intentional self-harm (suicide)	count_deaths
26	A_AA_PC_aamr_100k	All sexes	All ages	Prostate cancer	age_adjusted_mortality_rate_per_100k
27	A_35plus_ALLSA_sm	All sexes	35+	All smoking-attributable causes	smoking_attributable_deaths
28	A_AA_D_aamr_100k	All sexes	All ages	Diabetes	age_adjusted_mortality_rate_per_100k
29	A_AA_HA_count_d	All sexes	All ages	HIV/AIDS	count_deaths
30	A_AA_IP_aamr_100k	All sexes	All ages	Influenza and pneumonia	age_adjusted_mortality_rate_per_100k
31	A_AA_PC_count_d	All sexes	All ages	Prostate cancer	count_deaths
32	A_less1_ALL_count_c	All sexes	<1	All causes	count_perinatal_deaths
33	A_AA_C_count_d	All sexes	All ages	Cancer	count_deaths
34	A_35plus_SACMD_sı	All sexes	35+	Smoking-attributable cardiovascular and metaboli	smoking_attributable_deaths
35	M_AB_ALL_life_exp_a	Male	At birth	All causes	life_expectancy_at_birth
36	A_AA_HD_count_d	All sexes	All ages	Heart disease	count_deaths
37	A_AA_H_count_d	All sexes	All ages	Homicide	count_deaths
38	A_AA_OD_aamr_100	All sexes	All ages	Drug overdoses	age_adjusted_mortality_rate_per_100k
39	A_AA_HD_aamr_100	All sexes	All ages	Heart disease	age_adjusted_mortality_rate_per_100k
40	A_AA_BC_aamr_100k	All sexes	All ages	Breast cancer	age_adjusted_mortality_rate_per_100k
41	A_AA_CC_count_d	All sexes	All ages	Colorectal cancer	count_deaths
42	A_AA_Unintentional_	All sexes	All ages	Unintentional injuries (excluding drug overdoses)	count_deaths
43	A_less1_ALL_inf_mr_'	All sexes	<1	All causes	infant_mortality_rate_per_1k_live_births
44	A_35plus_SACMD_sı	All sexes	35+	Smoking-attributable cardiovascular and metaboli	age_adjusted_smoking_attributable_mortality_rate_per_100k
45	A_AA_BC_count_d	All sexes	All ages	Breast cancer	count_deaths
46	A_AA_C_aamr_100k	All sexes	All ages	Cancer	age_adjusted_mortality_rate_per_100k
47	A_AA_CKD_aamr_10	All sexes	All ages	Chronic kidney disease	age_adjusted_mortality_rate_per_100k
48	A_AA_SH_aamr_100l	All sexes	All ages	Intentional self-harm (suicide)	age_adjusted_mortality_rate_per_100k
49	A_AA_D_count_d	All sexes	All ages	Diabetes	count_deaths
50	A_AA_LC_aamr_100k	All sexes	All ages	Lung cancer	age_adjusted_mortality_rate_per_100k

Fig 4 B.