```
************************************;
* City of San Diego's Data Science     ;
* Candidate Assignment                 ;
* Assessing Streetlight Repair Services;
* Chhandara Pech | 10/7/2022           ;
************************************;


*Connect to data spreadsheet;
libname SD pcfiles path="C:\Users\chhan\Box\Chhandara Pech\City of
SD\Data\StreetLight_Resource_Assignment_SDCity.xlsx" scantime=yes
stringdates=no dbmax_text=2000;


*Bring in data;
data lights; set SD.'street_light_open$'n; *street lights;
      in_lights=1; *dummy variable;
data crime; set SD.'ARJISPublicCrime091422$'n; *crime;
      in_crime=1; *dummy variable;
data acs; set SD.'zcta_acs1519$'n; *ACS;
      in_acs=1; *dummy variable;
data xwalk; set SD.'zcta_to_SDcity_xwalk$'n; *crosswalk;
      in_xwalk=1; *dummy variable;
      run;


*1st - work with crime data;

*Freq of crime type;
proc freq data=crime;
      tables CM_LEGEND;
      title1 "Public Crime Data";
      title2 "Type of Crime Frequency";
      run;

*Categorize crime into property, violent, property + violent categories;
data crime; set crime;
      if CM_LEGEND in ("ARSON", "BURGLARY", "MOTOR VEHICLE THEFT",
"THEFT/LARCENY", "VANDALISM", "VEHICLE BREAK-IN/THEFT") then prop_crime=1;
            else prop_crime=0;
      if CM_LEGEND in ("ASSAULT", "HOMICIDE", "ROBBERY") then viol_crime=1;
            else viol_crime=0;
      viol_prop = viol_crime+prop_crime;
      run;


*Summarize counts of crime by type and by zipcode;
proc sort data=crime; by zipcode;
      run;
proc summary data=crime; by zipcode;
      var in_crime prop_crime viol_crime viol_prop;
      output out = crimezip sum=;
      run;
*Rename variables, drop misc variables, and remove obs with missing values;
data crimezip; set crimezip;
      if zipcode NE .;
```

```
      rename in_crime = all_crime;
      label in_crime = all_crime;
      drop _type_ _freq_;
      run;
*Calculate all other type of crime;
data crimezip; set crimezip;
      oth_crime=(all_crime - viol_prop);
      in_crime =1;
      run;

*2nd - Use geographic crosswalk to identify ZIP Codes/ZCTA in SD City;
      *Keep only ZIP Codes/ZCTA in SD City;

*Sort and merge by zipcode;
proc sort data=acs; by zipcode;
proc sort data=xwalk; by zipcode;
      run;
data acs; merge acs xwalk; by zipcode;
      run;
data acs; set acs;
      if afact>.50; *where ZCTA is at least 50% in SD City;
      run;


*Merge ACS to summarized crime data by ZCTA/ZIP Code;
proc sort data=crimezip; by zipcode;
proc sort data=acs2; by zipcode;
      run;
data crimezip; merge crimezip acs; by zipcode;
      run;
data crimezip; set crimezip;
      if afact>.50;
      run;

*Calculate ZIP Code/ZCTA crime rate and poverty rate;
*Normalize by 1,000;
data crimezip; set crimezip;
      all_crimerate = (all_crime/(pop/1000)); *all crime rate;
      viol_crimerate=(viol_crime/(pop/1000)); *violent crime;
      prop_crimerate=(prop_crime/(pop/1000)); *property crime;
      violprop_crimerate=(viol_prop/(pop/1000)); *violent + property crime
rate;
      pct_povt = npov/dpov; *% poverty;
      pct_poc=poc/pop; *% people of color;
      run;


*Check distribution of crime rate;
proc univariate data=crimezip;
      var all_crimerate viol_crimerate prop_crimerate violprop_crimerate;
      run;

proc corr data=crimezip;
      var pct_povt all_crime all_crimerate viol_crimerate prop_crimerate
violprop_crimerate oth_crime;
      title2 "Correlation between poverty and crime rates";
      run;
```

```
*Rank ZIP Code/ZCTA by crime rates into quintiles;
proc rank data=crimezip out=crimezip group=5;
      var pct_povt all_crimerate violprop_crimerate;
      ranks rank_povt rank_allcrime rank_violprop;
      run;

*Rank ZIP Code/ZCTA by violent + prop crime rates and poverty rates;
proc rank data=crimezip out=crimezip;
      var pct_povt violprop_crimerate;
      ranks rank_povt2 rank_violprop2;
      run;

*Calculate composite score;
*Sum crime and poverty rank scores;
data crimezip; set crimezip;
      PovtCrimeScore=(rank_povt2+rank_violprop2);
      run;

*Rank composite score into quintiles;
proc rank data=crimezip out=crimezip group=5;
      var PovtCrimeScore;
      ranks rank_PovtCrimeScore;
      run;

*Summarize open street light cases by ZCTA/ZIP Code;
proc sort data=lights; by zipcode;
      run;
proc summary data=lights; by zipcode;
      var in_lights;
      output out=lightszip sum=;
      run;
data lightszip; set lightszip;
      if zipcode ne .;
      drop _type_ _freq_;
      rename in_lights = streetlight_cases;
      label in_lights = streetlight_cases;
      in_streetlight=1;
      run;

*Merge summarized streetlight open cases data to ZCTA/ZIP code dataset;
proc sort data=lightszip; by zipcode;
proc sort data=crimezip; by zipcode;
      run;
data crimezip; merge crimezip lightszip; by zipcode;
      run;
data crimezip; set crimezip;
      if afact>.50; *Subset for ZCTA/ZIP codes that are at least 50% in City
of San Diego;
      run;

*Remove observations with no ZIP codes;
data lights; set lights;
      if zipcode ne .;
      run;

*Merge (individual) streetlight open cases data to ZCTA/ZIP code dataset;
```

```
proc sort data=lights; by zipcode;
proc sort data=crimezip; by zipcode;
      run;
data lights; merge lights crimezip; by zipcode;
      run;
data lights; set lights;
      if in_lights=1;
      if afact>.50;
      run;


*Generate some tabulations;

proc freq data=lights;
      tables year_req;
      title1 "Streetlight Repairs in San Diego";
      title2 "Freq of Open Cases by Year";
      run;
proc freq data=lights;
      tables council_district;
      title1 "Streetlight Repairs in San Diego";
      title2 "Freq of Open Cases by Year";
      run;

proc means data=crimezip; class rank_PovtCrimeScore;
      var PovtCrimeScore violprop_crimerate pct_povt pct_POC;
      title2 "Neighborhood Characteristics by Crime + Poverty Score";
      run;

proc means data=crimezip sum; class rank_povt;
      var streetlight_cases;
      title2 "Number of Open cases by Neighborhood Poverty Rate";
      run;
proc means data=lights; class rank_povt;
      var case_age_days;
      title2 "Average Number of Days Since Submission by Neighborhood Poverty
Rate";
      run;
proc means data=lights; class rank_allcrime;
      var case_age_days;
      title2 "Average Number of Days Since Submission by Neighborhood Overall
Crime Rate";
      run;
proc means data=lights; class rank_violprop;
      var case_age_days;
      title2 "Average Number of Days Since Submission by Neighborhood
Violent/Property Crime Rate";
      run;
proc means data=lights; class rank_PovtCrimeScore;
      var case_age_days;
      title2 "Average Number of Days Since Submission by Neighborhood Crime +
Poverty Rate";
      run;

proc means data=crimezip n min max mean median; class rank_povt;
      var pct_povt;
      title2 "Poverty Rate Ranges by Povt Ranking";
```

```
      run;
proc means data=crimezip n min max mean median; class rank_violprop;
      var violprop_crimerate;
      title2 "Crime Rate Ranges by Crime Ranking";
      run;
proc means data=crimezip n min max mean median; class rank_PovtCrimeScore;
      var PovtCrimeScore;
      title2 "Composite Score Ranges by C+P Ranking";
      run;


libname sd "C:\Users\chhan\Box\Chhandara Pech\City of SD\Data\SAS";

*Save SAS datasets;
data SD.crimezip; set crimezip;
data SD.lights; set lights;
data SD.lightszip; set lightszip;
      run;

*Export data as CSV file, use data to map;
PROC EXPORT DATA= WORK.CRIMEZIP
            OUTFILE= "C:\Users\chhan\Box\Chhandara Pech\City of
SD\Maps\SDCity_Resource_ZCTAs_GIS.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;
```