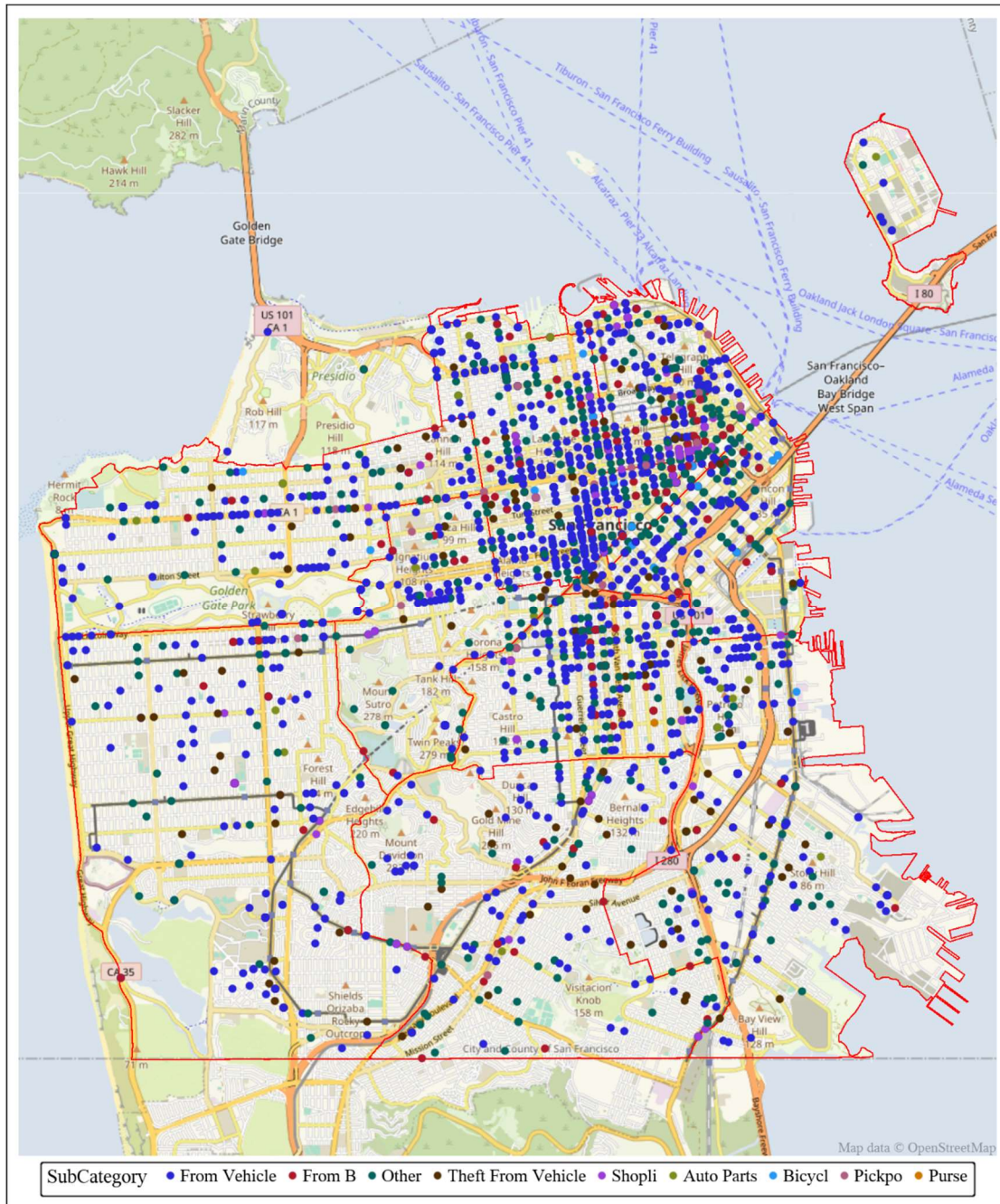


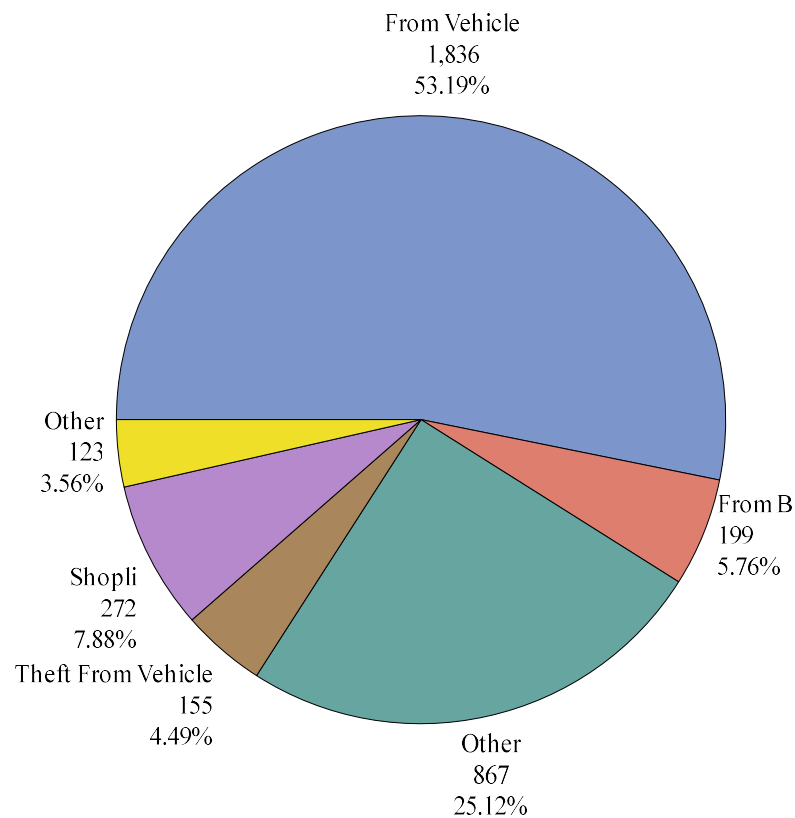
Author: My Nguyen

Class: STA 402 A

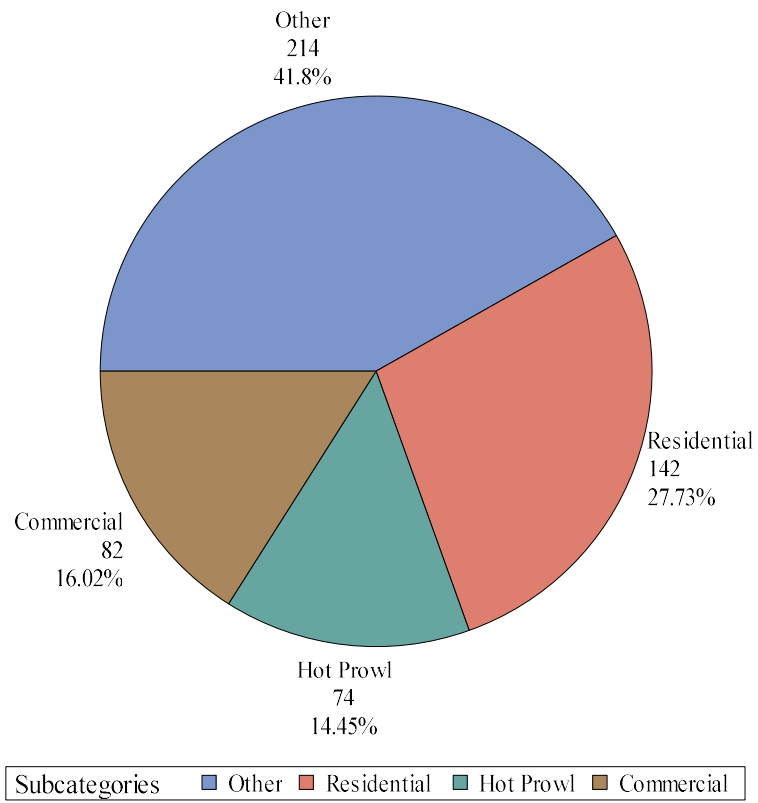
Assignment Statement: Write a SAS macro program that accepts a month and incident type specified by the user, then produces a map of San Francisco with all locations of that type of incident in that given month. The program should work for any month in the given dataset. You will need to learn the SAS Procedure PROC GMAP (there are many SAS Global Forum examples online that should be helpful, but please cite those sources if you use them). In your report, use your program to study whether and how different types of incidents in San Francisco have changed over the years.

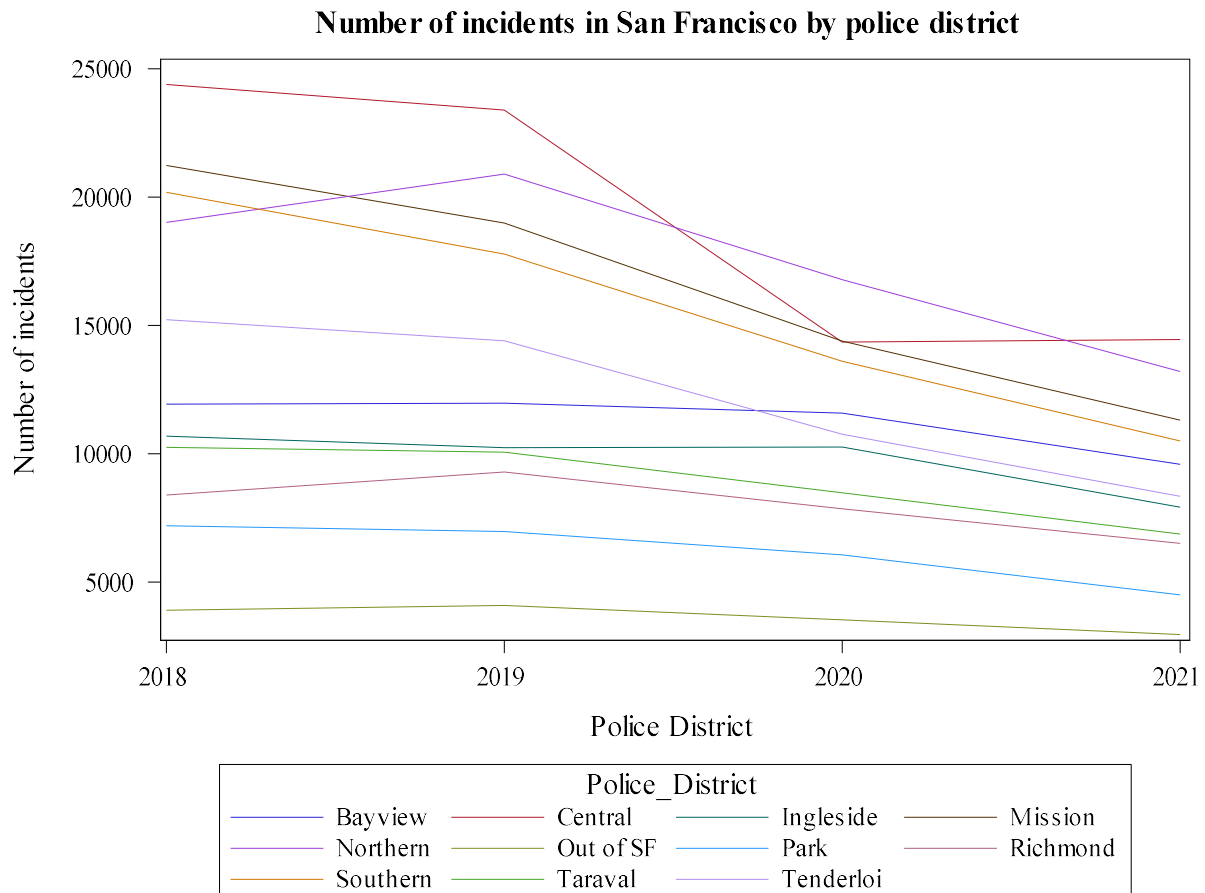


Larceny Theft is the category with greatest total number of reported incidents. From the map above which showing all larceny/theft incident locations in February 2020, it is clear larceny and theft relating vehicles is the most common in San Francisco as shown in the map and the pie chart below (53.19% of all incidents). The locations of incidents scatter around the whole city but we can see a clear incident cluster to the North West of the city. This is also where the main financial district of San Francisco locates. Thus, this shows a relationship between the likelihood of a larceny/theft and the population/GDP of the district.



Subcategories	From Vehicle	From B	Other	Theft From Vehicle	Shopli	Other
---------------	--------------	--------	-------	--------------------	--------	-------





Overall, we can observe a decreasing trend in the total number of incidents in San Francisco from 2018 to 2021 in all districts in the line plot above. As shown in the map above, Central District is the neighborhood with greatest number of reported incidents from 2018-2021 period, except for the year 2020 which it lost to Tenderloin. While Bayview, Park, Richmond, Ingleside, Taraval district maintain their total number of incidents around and under 10000 cases for the whole period, Central, Northern, Mission, Southern, Tenderloin district experience a sharp drop at over 20000 cases in 2018 to just around 15000 in 2021.

For more detailed analysis, the table in next page provides a thorough insight into how total number of incidents in each subcategory are change over time. The percentage is calculated by taking the change from one year to the next, which is then divided over the total number of incidents in current year and multiplied with 100. With the number of red boxes decrease over time, it is clear that the trend aligns with all of the plots presented above. Due to data unavailability, some categories are left blank. The first row in the table is displaying the total number of reported incidents with no category labeled.

Annual percentage change (%) in number of incidents in SF by category

Obs	Category	2018-2019	2019-2020	2019-2020
1		-400.00	99.35	-123.36
2	Aggravated Assault	1.32	-12.48	-16.00
3	Arrest	50.00	.	33.33
4	Arson	-16.93	27.04	-16.89
5	Bad Checks	-40.00	16.67	-200.00
6	Bribery	33.33	.	.
7	Burglary - Commercial	-15.20	41.82	-54.00
8	Burglary - Hot Prowl	-30.24	46.67	-6.35
9	Burglary - Other	-17.74	30.53	-52.69
10	Burglary - Residential	-14.50	30.75	-20.46
11	Case Closure	-8.49	-28.63	-86.32
12	Courtesy Report	12.55	-64.59	-45.24
13	Disorderly Conduct	37.03	-134.87	-19.18
14	Drug Violation	-11.81	-40.41	-40.48
15	Drunkenness	-11.63	-95.45	-59.04
16	Embezzlement	-11.96	-25.90	-118.42
17	Extortion-Blackmail	-1.92	14.75	-5.17
18	Fire Report	-7.64	25.39	2.53
19	Forgery And Counterfei	-15.96	-61.07	-77.03
20	Fraud	-2.96	-31.03	-54.63
21	Gambling	-133.33	-200.00	0.00
22	Homicide	-433.33	25.00	-33.33
23	Human Trafficking, Com	-266.67	-100.00	0.00
24	Human Trafficking, Inv	.	.	.
25	Intimidation	0.57	-27.79	-17.69
26	Kidnapping	-6.90	-24.29	7.89
27	Larceny - Auto Parts	59.57	52.28	22.59
28	Larceny - From Vehicle	-0.89	-82.41	1.69
29	Larceny Theft - Bicycl	-16.60	-13.01	-34.00

Obs	Category	2018-2019	2019-2020	2019-2020
30	Larceny Theft - From B	-9.40	-42.10	-49.89
31	Larceny Theft - Other	0.75	-42.70	-66.41
32	Larceny Theft - Pickpo	-12.24	-243.48	-48.92
33	Larceny Theft - Purse	18.92	-146.67	-7.14
34	Larceny Theft - Shopli	1.87	-51.07	0.90
35	Liquor Law Violation	-25.00	-28.57	-16.67
36	Loitering	-23.96	-301.28	-81.40
37	Lost Property	-5.67	-120.76	-37.14
38	Manslaughter	33.33	-50.00	0.00
39	Miscellaneous Investig	0.74	-13.72	-13.90
40	Missing Adult	0.12	-25.17	-41.82
41	Missing Person	-22.28	-35.63	-41.98
42	Motor Vehicle Theft	1.75	27.90	-25.51
43	Motor Vehicle Theft (A	4.44	55.88	22.14
44	Non-Criminal	-0.79	-40.37	-4.79
45	Other	-7.47	-29.01	-29.24
46	Other Offenses	26.06	34.82	-29.40
47	Prostitution	1.05	-506.35	-18.87
48	Rape	-4.00	-47.06	-30.77
49	Rape - Attempted	-50.00	-71.43	-133.33
50	Recovered Vehicle	-2.02	24.66	-23.84
51	Robbery - Carjacking	25.62	30.46	-58.18
52	Robbery - Commercial	-11.51	-1.43	-30.68
53	Robbery - Other	0.77	-39.54	-36.46
54	Robbery - Residential	-150.00	-100.00	50.00
55	Robbery - Street	-16.18	-48.41	-36.96
56	Sex Offense	-7.88	-16.20	-51.06
57	Simple Assault	-3.52	-37.43	-17.20
58	Stalking	31.43	-25.00	-47.37
59	Stolen Property	-5.09	-14.40	-39.23
60	Suicide	34.15	-70.83	-4.35

Obs	Category	2018-2019	2019-2020	2019-2020
61	Suspicious Occ	0.80	-11.08	-12.90
62	Suspicious Package	-34.78	-9.52	-110.00
63	Theft From Vehicle	26.27	1.44	-5.75
64	Traffic Collision	36.65	-30.89	17.45
65	Traffic Collision - Hi	1.92	-10.64	9.03
66	Traffic Violation Arre	-13.09	-68.72	-38.65
67	Trespass	-6.00	-18.77	-7.07
68	Vandalism	2.24	-0.89	-11.76
69	Vehicle Impounded	34.81	-75.32	-42.59
70	Vehicle Misplaced	-8.06	-21.57	-59.38
71	Warrant	-5.28	-76.15	-45.11
72	Weapons Offense	-4.68	0.38	-2.22

```

/* My Nguyen
Final project
Date: May 9, 2022

Please change the folder path on the right-hand side of the statement
below to the path where this SAS file and the Current Police Districts
locate.
*/
%let workPath= M:\STA 402\Project;

/*

This SAS program has 2 macros:
- %incident_map(Month=, Year=, Category=, SubCategoryByPercentages=,
output=):
    This macro accepts a month and incident type specified by the user,
    then produce a map of San Francisco with all locations of that type of
    incident in that given month.

    Month: month(numeric) when the incidents happen
    Year: year when the incident happen
    Category: type of incidents
    SubCategoryByPercentages: 1(=true)/ 0(false), an optional pie chart
    displaying
    subcategories in percentage of the genral category specified above
    output: file path for rtf output of the map and/or pie chart

- % generalAnalysis(output=)
    This macro will generate a line plot showing number of incidents

```


in SF by police district and a color-coded table with annual percentage change
of total number of incidents in SF by category (blue for decrease, red for increase)

output: file path for rtf output of the table and plots

The program is implemented as a SAS macro program so please change the statement

```
%let folder=
```

as indicated above. Next run this SAS file to define the macro.
Maps, tables, and graphs in final report are generated using following macros:

```
%incident_map(Month=2, Year=2020, Category=Larceny Theft,  
SubCategoryByPercentages=1, output=M:\STA 402\Project\Map.rtf);  
%incident_map(Month=2, Year=2020, Category=Burglary,  
SubCategoryByPercentages=1, output=M:\STA 402\Project\Map.rtf);  
%generalAnalysis(output=M:\STA 402\Project\Plots.rtf)
```

This program requires current SAS file to be in the same location as the Current Police District folder.

*/

```
options work("&workPath";
```

```
*Importing in the incident data set;
```

```
proc import  
file="&workPath\Police_Department_Incident_Reports__2018_to_Present.csv"  
out=work.incidents  
dbms=csv;  
run;
```

```
%macro incident_map(Month=, Year=, Category=, SubCategoryByPercentages=,  
output=);
```

```
data work.cleanedIncidents;  
set work.incidents;  
Incident_Month = month(Incident_Date); *Extracting incident month from  
reported date;  
* Stripping repetitive incident category from subcategories for  
better map legend;  
if find(Incident_SubCategory, '-') then SubCategory =  
scan(Incident_SubCategory, -1, '-');  
else SubCategory = Incident_SubCategory;
```

```
*Selecting data given user input;  
if Incident_Month = &Month and  
Incident_Year = &Year and  
Incident_Category = "&Category" and  
Police_District ^= 'Out of SF'; *Select incidents from inside SF  
only;  
keep Incident_Month Incident_Year Incident_Category SubCategory  
Latitude Longitude;
```

```

run;

*Importing in the district boundaries for SF;
proc mapimport datafile="&workPath\Current Police
Districts\geo_export_79d7b3ff-17c4-483e-bf69-766079da4e5f.shp"
out=boundaries;
run;

ods rtf bodytitle file = "&output";
ods graphics on / width=1000px height=1200px;
proc sgmap mapdata=boundaries      /* Map boundaries */
      plotdata=work.cleanedIncidents
      des='San Francisco';
  openstreetmap;
  *Drawing boundary lines for all police districts;
  choromap / mapid=district lineattrs=(color=red) legendlabel='Districts';
  *Drawing locations of selected incidents;
  scatter x=Longitude y=Latitude / group=SubCategory markerattrs=(size=7
symbol=circlefilled);
run;
ods graphics off;

*Creating a pie chart template;
%if &SubCategoryByPercentages = 1 %then %do;
  proc template;
    dynamic title1;
    define statgraph pie;
      begingraph;
      entrytitle title1;
      layout region;
        piechart category = SubCategory /
        datalabellocation = outside
        datalabelcontent = all
        categorydirection = clockwise
        start = 180 name = 'pie';
        discretelegend 'pie' /
        title = 'Subcategories';
      endlayout;
    endgraph;
  end;
run;

  *drawing the pie chart;
  proc sgrender data = work.cleanedIncidents
    template = pie;
    dynamic title1 = "Percentage breakdown of incident by
subcategory";
  run;
%end;
ods rtf close;
%mend incident_map;

%macro generalAnalysis(output=);
*===== Set up for Line Graph;
*Sorting incidents by district for later counting;
proc sort data=work.incidents
  out=work.sortedByDistrict;

```

```

    by Incident_Year Police_District;
run;

*Counting incidents by different districts;
data casesByPoliceDistrict;
    set work.sortedByDistrict;
    by Incident_Year Police_District;

    retain Count;
    if first.Police_District then Count = 1;
    else Count = Count + 1;

    if last.Police_District then output;

    keep Incident_Year Police_District Count;
run;

*===== Set up for Percent table;
*Sorting incidents by subcategories for later counting;
proc sort data=work.incidents
    out=work.sortedBySubCategory;
    by Incident_Year Incident_SubCategory;
run;

*Counting incidents by different subcategories;
data casesBySubCategory;
    set work.sortedBySubCategory;
    by Incident_Year Incident_SubCategory;

    retain Count;
    if first.Incident_SubCategory then Count = 1;
    else Count = Count + 1;

    if last.Incident_SubCategory then output;

    keep Incident_Year Incident_SubCategory Count;
run;

*Sorting to calculate percentage change by rows(years) for each category;
proc sort data=casesBySubCategory
    out=sorted;
    by Incident_SubCategory;
run;

*Calculating the percentage change;
data percentChanges;
    set sorted;
    by Incident_SubCategory notsorted;
    Percent_Change = round((Count - lag(Count)) / Count * 100, .01);
    if first.Incident_SubCategory then Percent_Change = 0;
    drop Count;
run;

*Transposing table;

```

```

proc transpose data=percentChanges out=Tr_percentChanges;
    id Incident_Year;
    by Incident_SubCategory; * transpose the original dataset;
run;

*Creating a color coded format for the data;
proc format;
value changes          -1000 -< 0 = light blue
                        0 <- 1000 = light red
                        other = white ;
run;

*===== Graph and Table output;
ods rtf bodytitle file = "&output";
* Number of incidents analyzed by district;
title 'Number of incidents in San Francisco by police district';
proc sgplot data=casesByPoliceDistrict;
    title 'Number of incidents in San Francisco by police district';
    vline Incident_Year / response=Count group=Police_District
        lineattrs=(pattern=solid thickness=2pt);
    xaxis label="Police District";
    yaxis label="Number of incidents";
run;

* Number of incidents analyzed by categories (percentage change over years);
title "Annual percentage change (%) in number of incidents in SF by
category";
proc print data=Tr_percentChanges label;
    var Incident_Subcategory _2019 _2020 _2021 /
style(data)=[backgroundcolor=changes.];
    label          Incident_SubCategory = "Category"
        _2019 = "2018-2019"
        _2020 = "2019-2020"
        _2021 = "2019-2020";
    *title "Annual percentage change (%) in number of incidents in SF by
category";
run;
ods rtf close;
%mend generalAnalysis;

```

Code Reference for pie chart:

https://www.tutorialspoint.com/sas/sas_pie_charts.htm