Visualizations

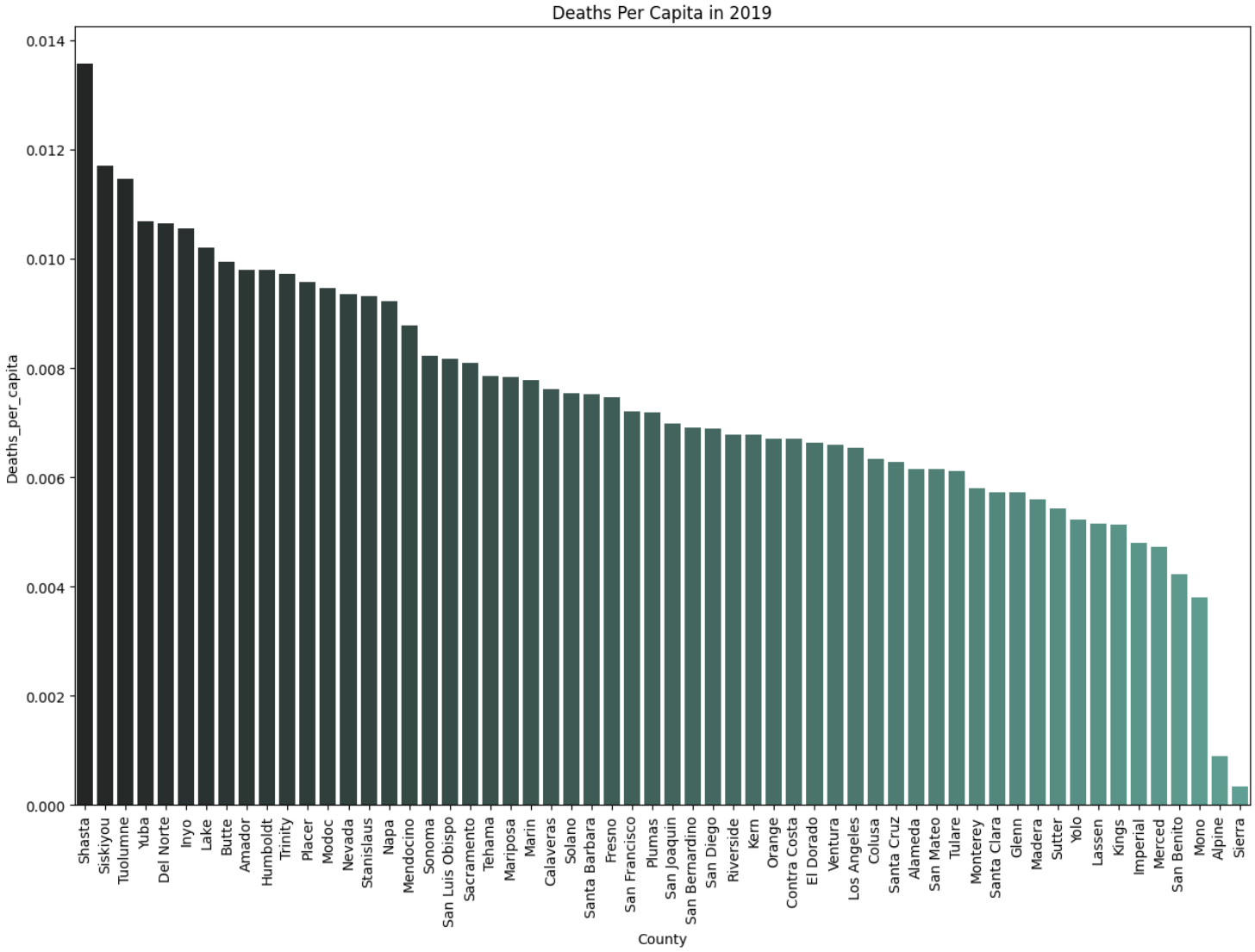
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
main_df = pd.read_csv('main_df.csv')
main_df['Deaths_per_capita'] = main_df['Count']/main_df['Population']
main_df
```

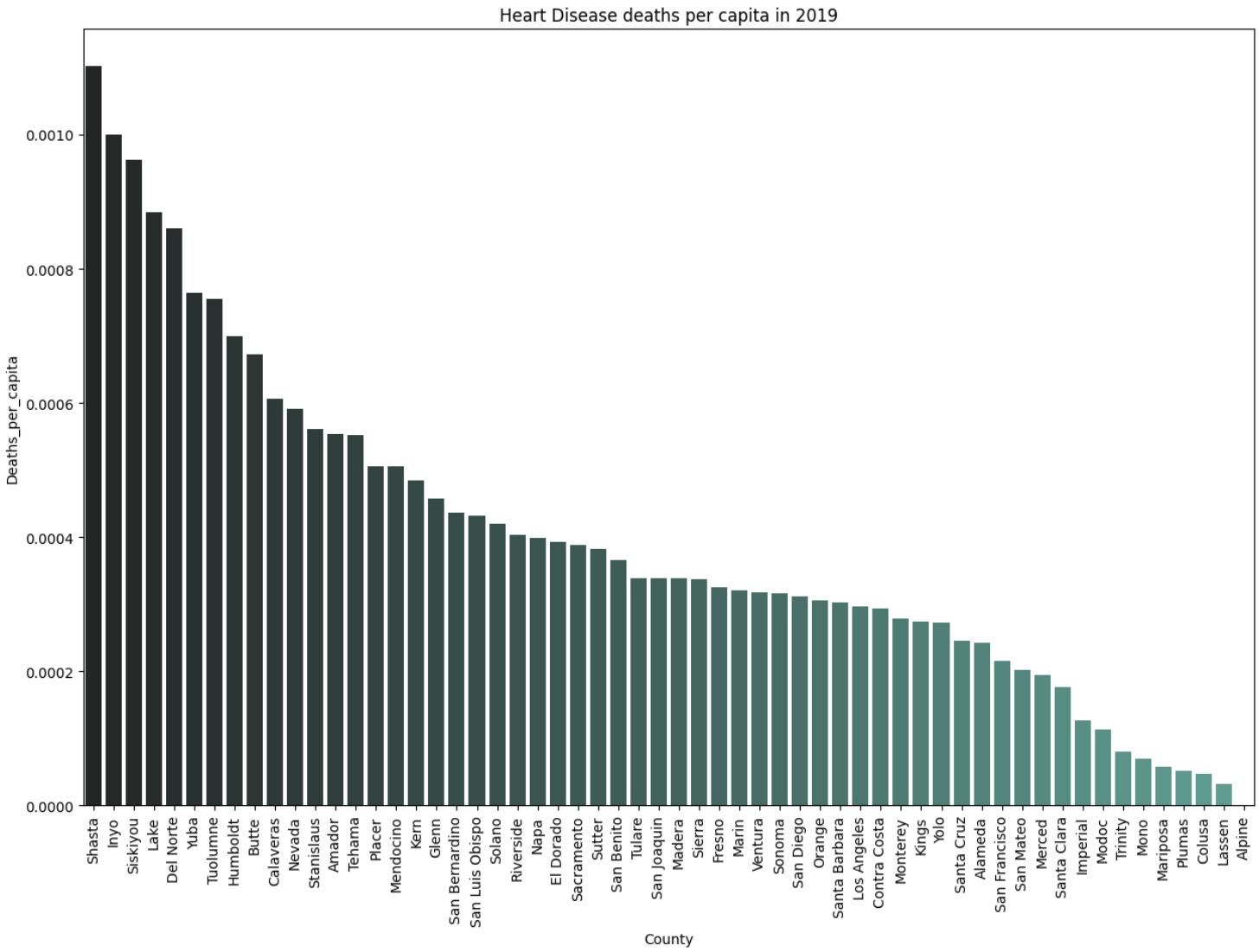
	Year int64	County object	Geography_Type ...	Strata object	Strata_Name obje...	C
170	2014	Alameda	Occurrence	Race-Ethnicity	Other/Unknown	S
171	2014	Alameda	Occurrence	Total Population	Total Population	S
172	2014	Alameda	Occurrence	Gender	Female	S
173	2014	Alameda	Occurrence	Gender	Male	S
174	2014	Alameda	Occurrence	Race-Ethnicity	American Indian/Alaska...	S
175	2014	Alameda	Occurrence	Race-Ethnicity	Asian	S
176	2014	Alameda	Occurrence	Race-Ethnicity	Black	S
177	2014	Alameda	Occurrence	Race-Ethnicity	Hawaiian/Pacific Islander	S
178	2014	Alameda	Occurrence	Race-Ethnicity	Hispanic	S
179	2014	Alameda	Occurrence	Race-Ethnicity	Multi-Race	S

Focus: Heart Disease

```
# Death Per Capita in 2019
plt.figure(figsize=(15,10))
plt.xticks(rotation=90)
plt.title('Deaths Per Capita in 2019')
sns.barplot(data=main_df[(main_df['Strata']=='Total Population')&(main_df["Cause"]=="ALL")
```



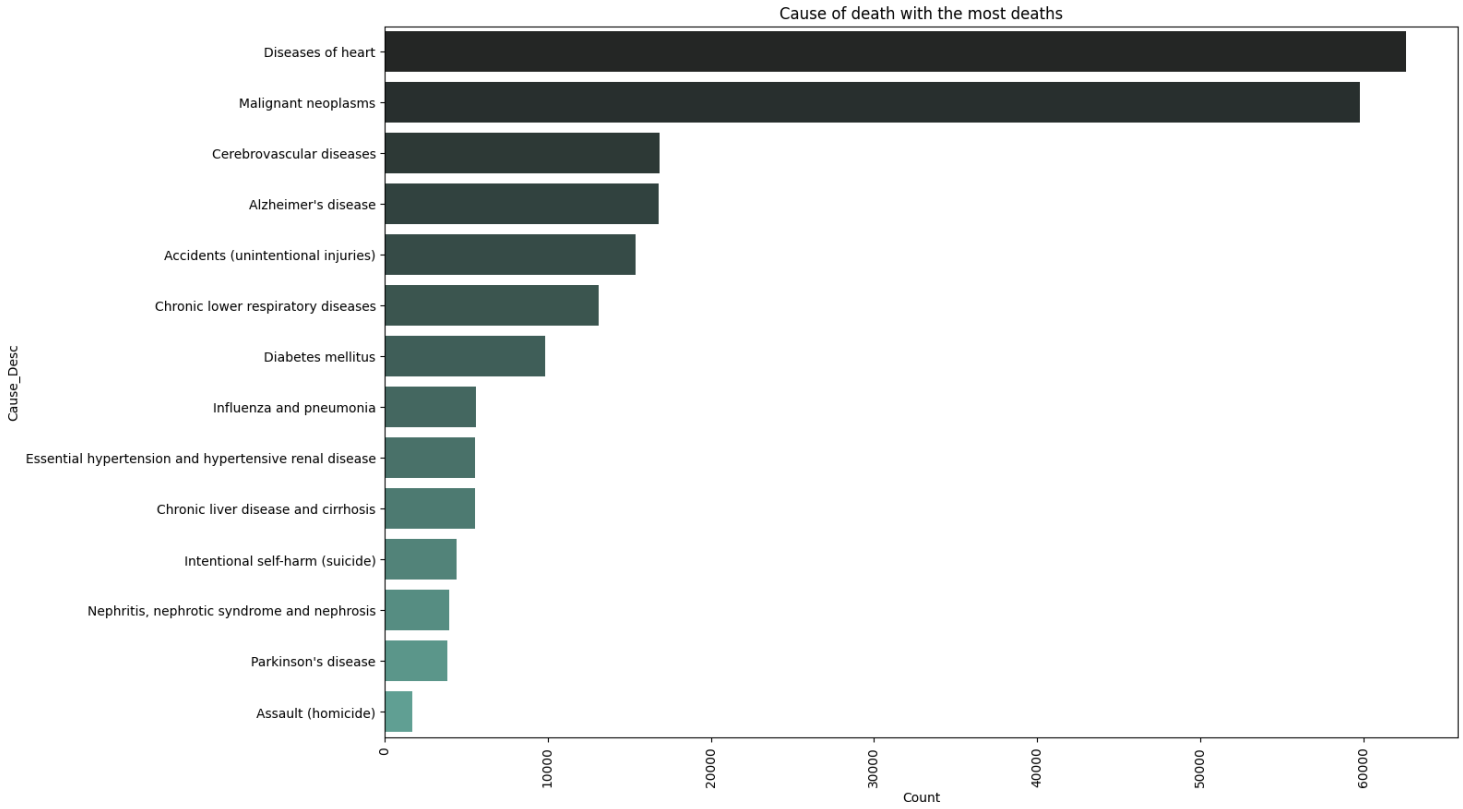
```
#Heart Disease deaths per capita for each county
plt.figure(figsize=(15,10))
plt.xticks(rotation=90)
plt.title('Heart Disease deaths per capita in 2019')
sns.barplot(data=main_df[(main_df['Strata']=='Total Population')&(main_df["Cause"]=="CLD")
```



```
types_of_deaths = main_df[(main_df['Strata']=='Total Population')&(main_df['Geography_Type
types_of_deaths
```

	Cause_Desc object All causes (t... 6.7% Diseases of ... 6.7% 13 others 86.7%	Count float64 1717.0 - 270924.0	Deaths_per_capi... 0.001647087892...	Population int64 39437610 - 3943...	
1	All causes (total)	270924.0	0.427518372761 2441	39437610	
8	Diseases of heart	62646.0	0.096422276571 91307	39437610	
12	Malignant neoplasms	59817.0	0.091831096847 98445	39437610	
4	Cerebrovascular diseases	16859.0	0.021066539139 16076	39437610	
2	Alzheimer's disease	16809.0	0.020277018881 113863	39437610	
0	Accidents (unintentional...	15387.0	0.032066035060 87776	39437610	
6	Chronic lower respiratory...	13104.0	0.022973084540 808802	39437610	
7	Diabetes mellitus	9823.0	0.010958122042 520202	39437610	
10	Influenza and pneumonia	5610.0	0.006410299627 920659	39437610	
9	Essential hypertension and...	5537.0	0.005925547996 1371265	39437610	

```
#Heart Disease deaths per capita for each county
plt.figure(figsize=(15,10))
plt.xticks(rotation=90)
plt.title('Cause of death with the most deaths')
sns.barplot(data=types_of_deaths.drop(1), x='Count', y='Cause_Desc', palette="dark:#5A9");
```



Compare with possible cause of heart disease

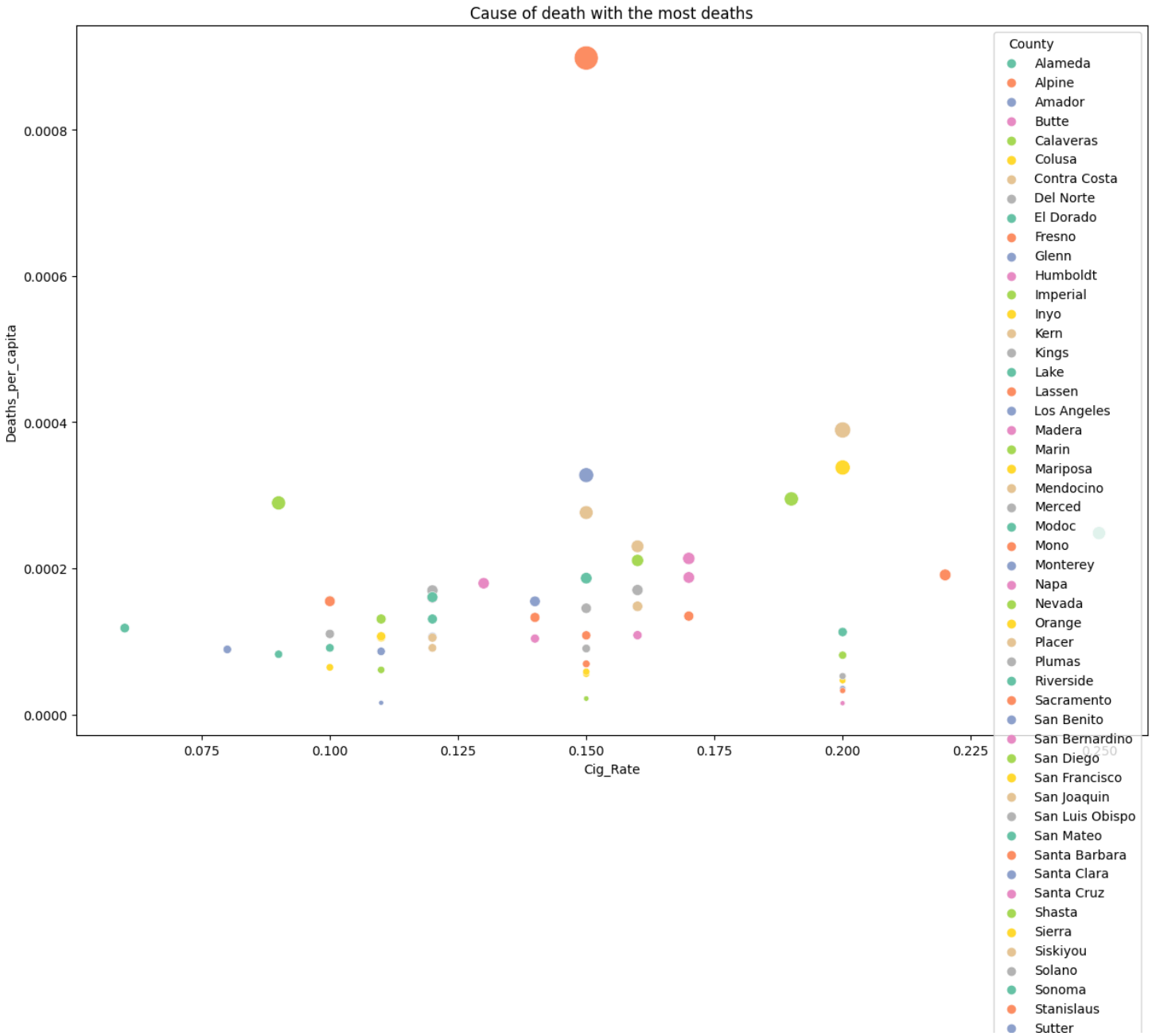
```
cig_vs_htd = main_df[(main_df['Strata']=='Total Population')&(main_df["Cause"]=="SUI")&(ma
cig_vs_htd
```

	Year int64 2019 - 2019	County object Alameda 1.7% Alpine 1.7% 56 others 96.6%	Geography_Type ... Occurrence 100%	Strata object Total Popul... 100%	Strata_Name obje...	C
105731	2019	Alameda	Occurrence	Total Population	Total Population	S
105913	2019	Alpine	Occurrence	Total Population	Total Population	S
106095	2019	Amador	Occurrence	Total Population	Total Population	S
106277	2019	Butte	Occurrence	Total Population	Total Population	S
106459	2019	Calaveras	Occurrence	Total Population	Total Population	S
106641	2019	Colusa	Occurrence	Total Population	Total Population	S
106823	2019	Contra Costa	Occurrence	Total Population	Total Population	S
107005	2019	Del Norte	Occurrence	Total Population	Total Population	S
107187	2019	El Dorado	Occurrence	Total Population	Total Population	S
107369	2019	Fresno	Occurrence	Total Population	Total Population	S

```
len(cig_vs_htd), len(cig_vs_htd['County'].unique())
```

```
(58, 58)
```

```
plt.figure(figsize=(15,10))
plt.title('Cause of death with the most deaths')
sns.scatterplot(data=cig_vs_htd, x='Cig_Rate', y='Deaths_per_capita', hue='County', palett
```



Statistics

```
#how many people die on avg each year
death_years_count = {}
for i in range(2014,2020):
    year = main_df[(main_df['Strata']=='Total Population')&(main_df['Geography_Type']=='Oc
    pop = sum(main_df[(main_df['Strata']=='Total Population')&(main_df['Geography_Type']==
    death_years_count[i] = year[year['Cause_Desc']=="All causes (total)"]['Count'][1]

death_years_count
```

```
{2014: 246781.0,
 2015: 260196.0,
 2016: 263208.0,
 2017: 269377.0,
 2018: 270129.0,
 2019: 270924.0}
```

```
deaths_count_2019 = sum(main_df[(main_df['Strata']=='Total Population')&(main_df['Geograph
```