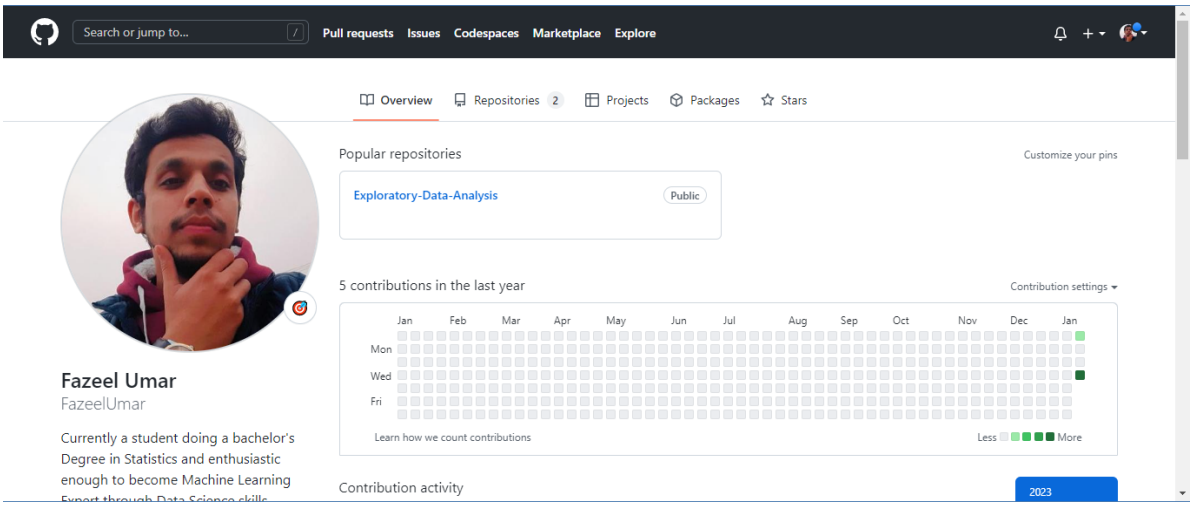
Name - Fazeel Umar

Roll - Fa20-bst-002

Project 2

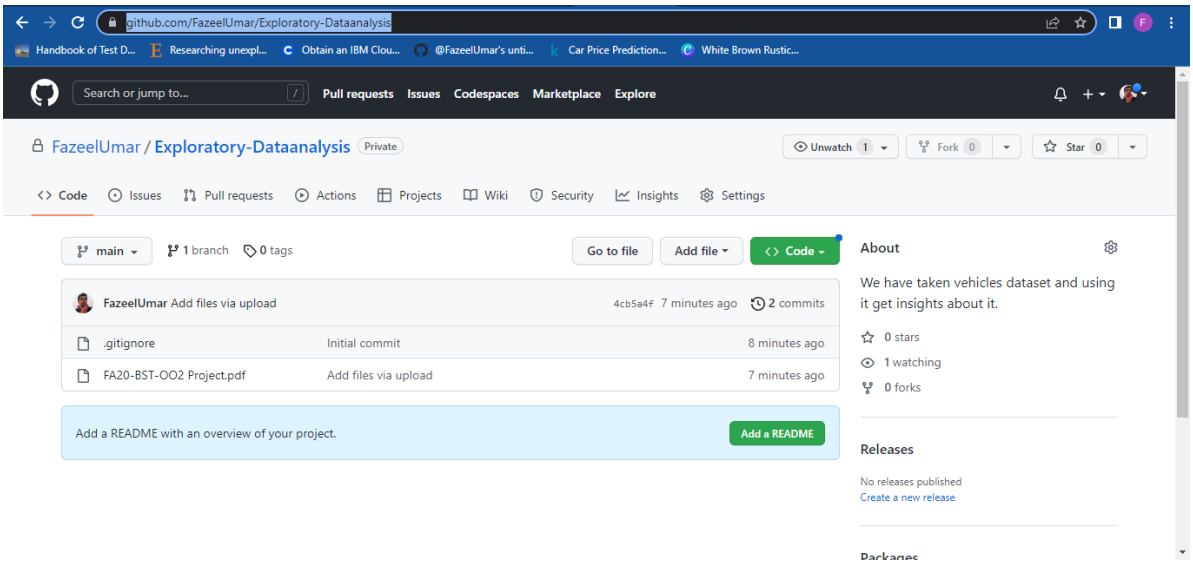
GITHUB ID

https://github.com/FazeelUmar



Proof of Uploading Project attached

https://github.com/FazeelUmar/Exploratory-Dataanalysis



.

.

Getting 3 different Data sets from Kaggle

.

1) Vehicles Data Set Visualisations

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:
```
data = pd.read_csv('C:\\Users\\Lenovo\\Desktop\\vehicles.csv')
data
```

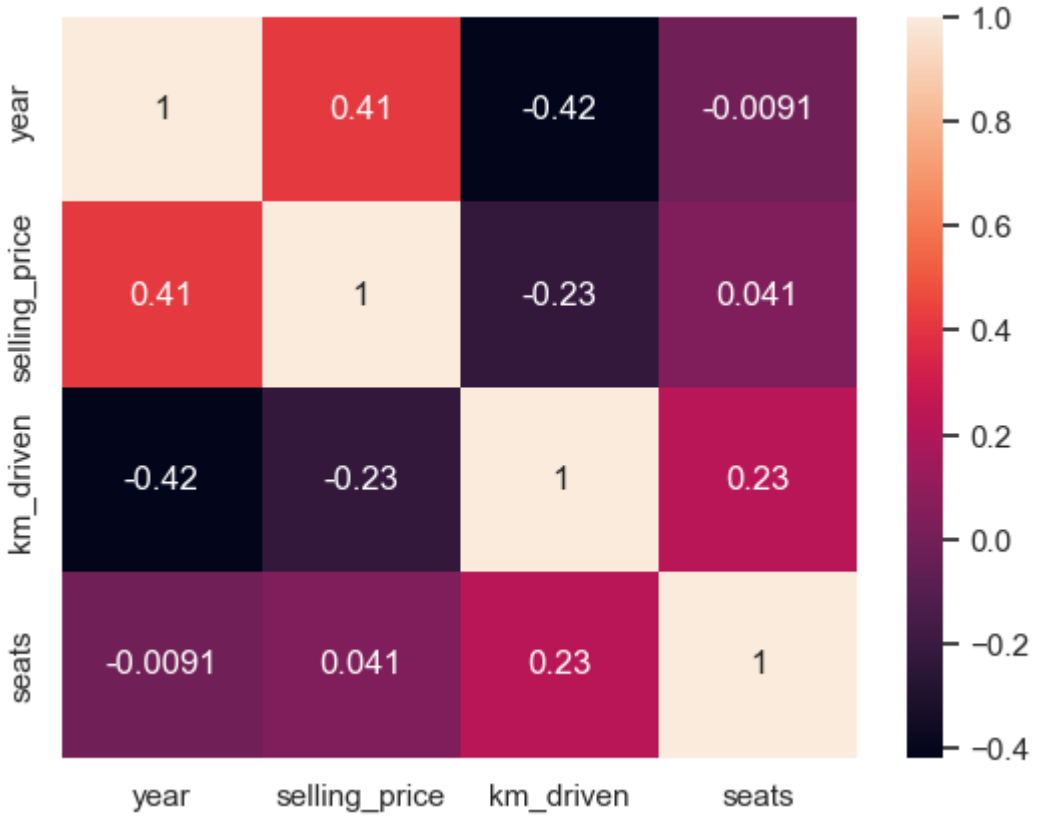
Out[3]:

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	en
0	Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individual	Manual	First Owner	23.4 kmpl	
1	Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	Diesel	Individual	Manual	Second Owner	21.14 kmpl	
2	Honda City 2017-2020 EXi	2006	158000	140000	Petrol	Individual	Manual	Third Owner	17.7 kmpl	
3	Hyundai i20 Sportz Diesel	2010	225000	127000	Diesel	Individual	Manual	First Owner	23.0 kmpl	
4	Maruti Swift VXI BSIII	2007	130000	120000	Petrol	Individual	Manual	First Owner	16.1 kmpl	
...	...	...	...	...	...	...	...	...	...	
8123	Hyundai i20 Magna	2013	320000	110000	Petrol	Individual	Manual	First Owner	18.5 kmpl	
8124	Hyundai Verna CRDi SX	2007	135000	119000	Diesel	Individual	Manual	Fourth & Above Owner	16.8 kmpl	
8125	Maruti Swift Dzire ZDi	2009	382000	120000	Diesel	Individual	Manual	First Owner	19.3 kmpl	
8126	Tata Indigo CR4	2013	290000	25000	Diesel	Individual	Manual	First Owner	23.57 kmpl	
8127	Tata Indigo CR4	2013	290000	25000	Diesel	Individual	Manual	First Owner	23.57 kmpl	

8128 rows × 13 columns

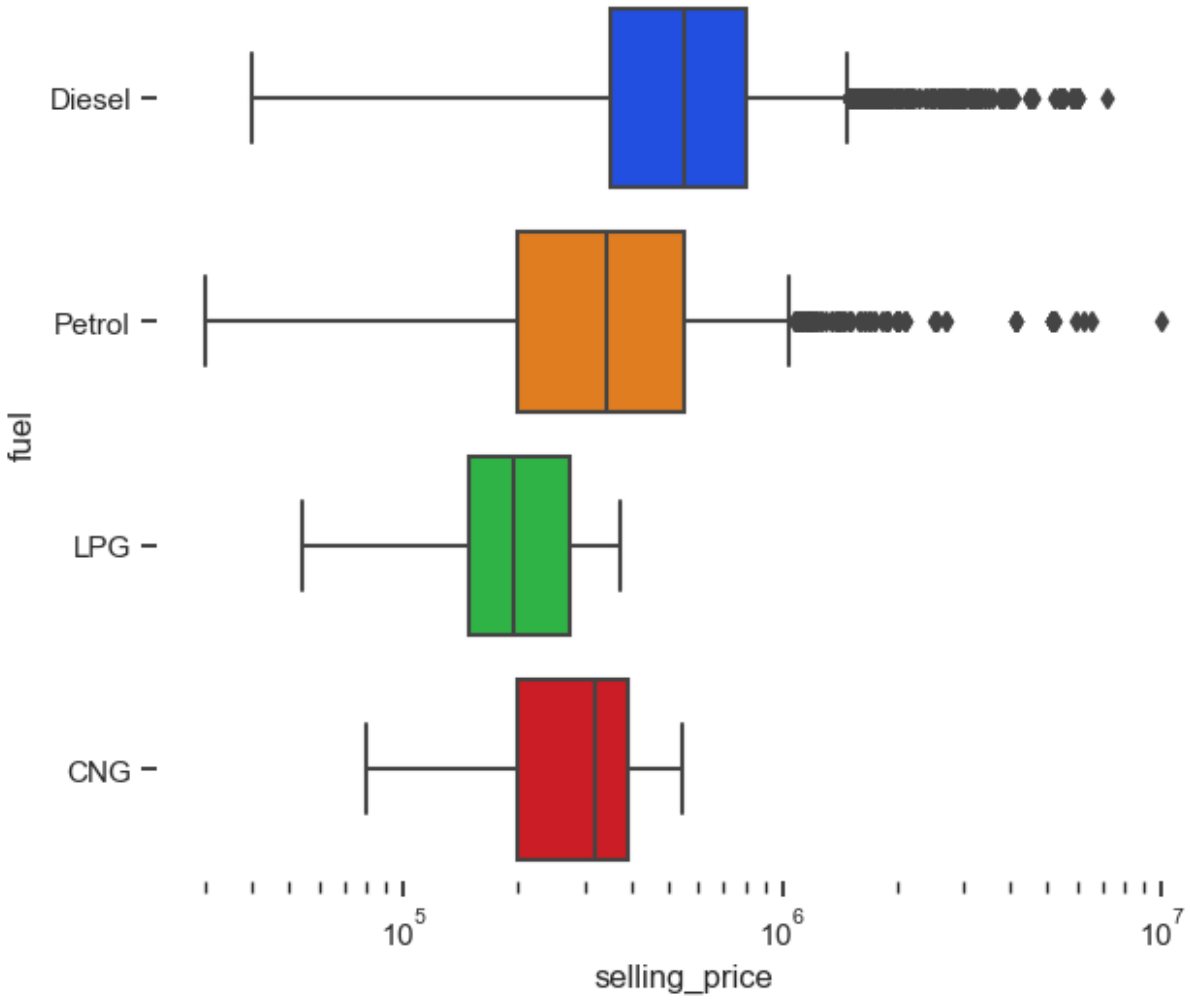
Heat map Graph

In [4]:
```
# Use the heatmap function from the seaborn package
sns.heatmap(data.corr(), annot=True);
# Display the Pharma Sector Heatmap
plt.show()
```



Box Plots

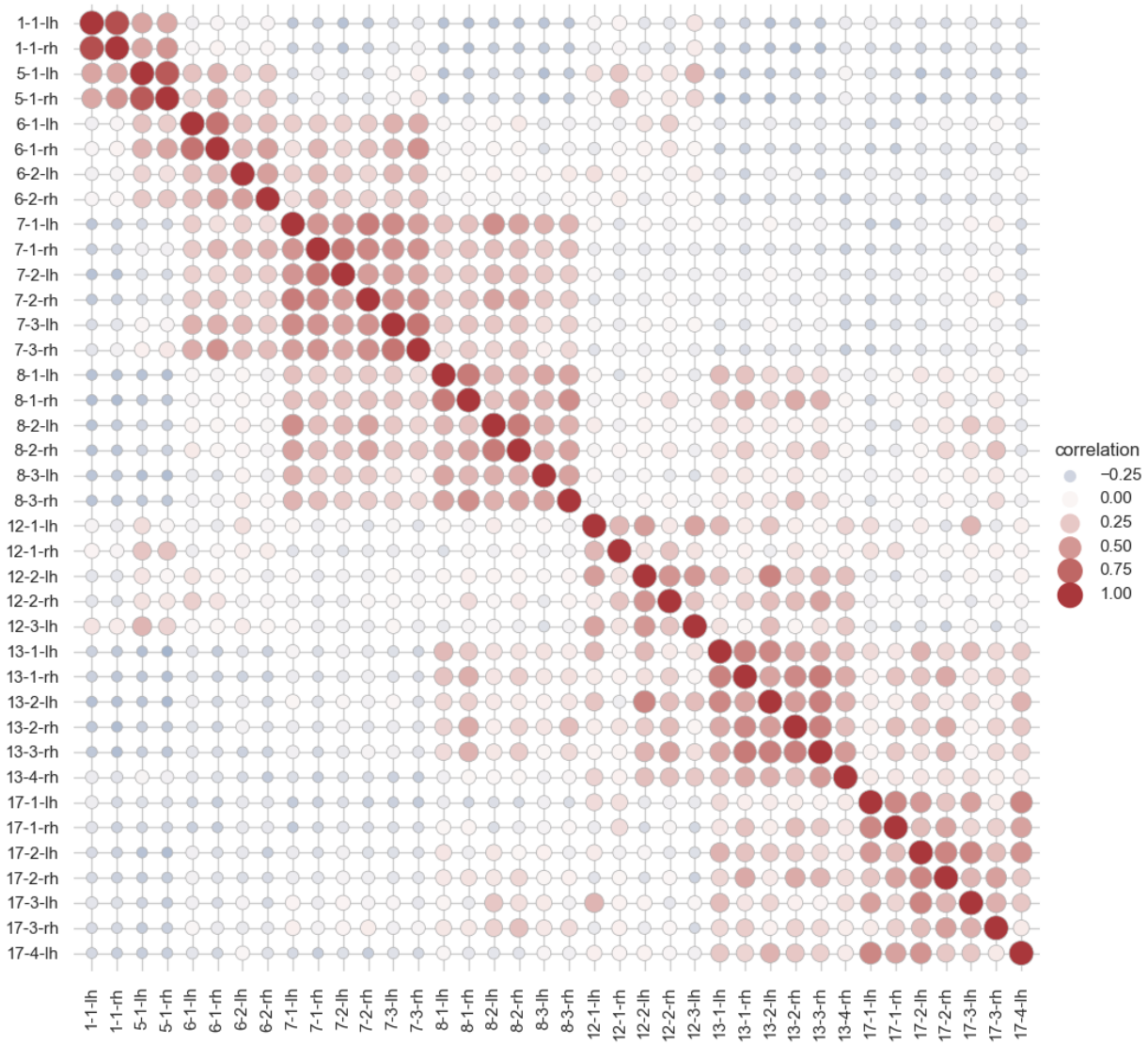
In [5]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv('C:\\Users\\Lenovo\\Desktop\\vehicles.csv')
sns.set_theme(style="ticks")
#Logarithmic x-axis
f, ax = plt.subplots(figsize=(7, 6))
ax.set_xscale("log")

sns.boxplot(x="selling_price",y="fuel",data=data, palette="bright")

sns.despine(f, left=True, bottom=True)
```



Correlation Graph

In [1]:
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.set_theme(style="whitegrid")

df=sns.load_dataset("brain_networks",header=[0,1,2],index_col=0)

used_networks=[1,5,6,7,8,12,13,17]
used_columns=(df.columns
              .get_level_values("network")
              .astype(int)
              .isin(used_networks))
df= df.loc[:, used_columns]
df.columns=df.columns.map("-".join)

#compute a correlation matrix and convert to long-form
corr_mat=df.corr().stack().reset_index(name="correlation")

#draw each cell as a scatter point with varying size and color

g=sns.relplot(data=corr_mat,x="level_0",y="level_1",hue="correlation",size="correlatio
           palette="vlag",hue_norm=(-1,1),edgecolor=".7",height=10,sizes=(50,250),size
#tweak the figure to finalize
g.set(xlabel="",ylabel="",aspect="equal")
```

```
g.despine(left=True,bottom=True)
g.ax.margins(.02)

for label in g.ax.get_xticklabels():
    label.set_rotation(90)
```



2) Causes of Death Dataset Visualisations

In [9]:

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

data2 = pd.read_csv('C:\\Users\\Lenovo\\Desktop\\death.csv')
data2
```

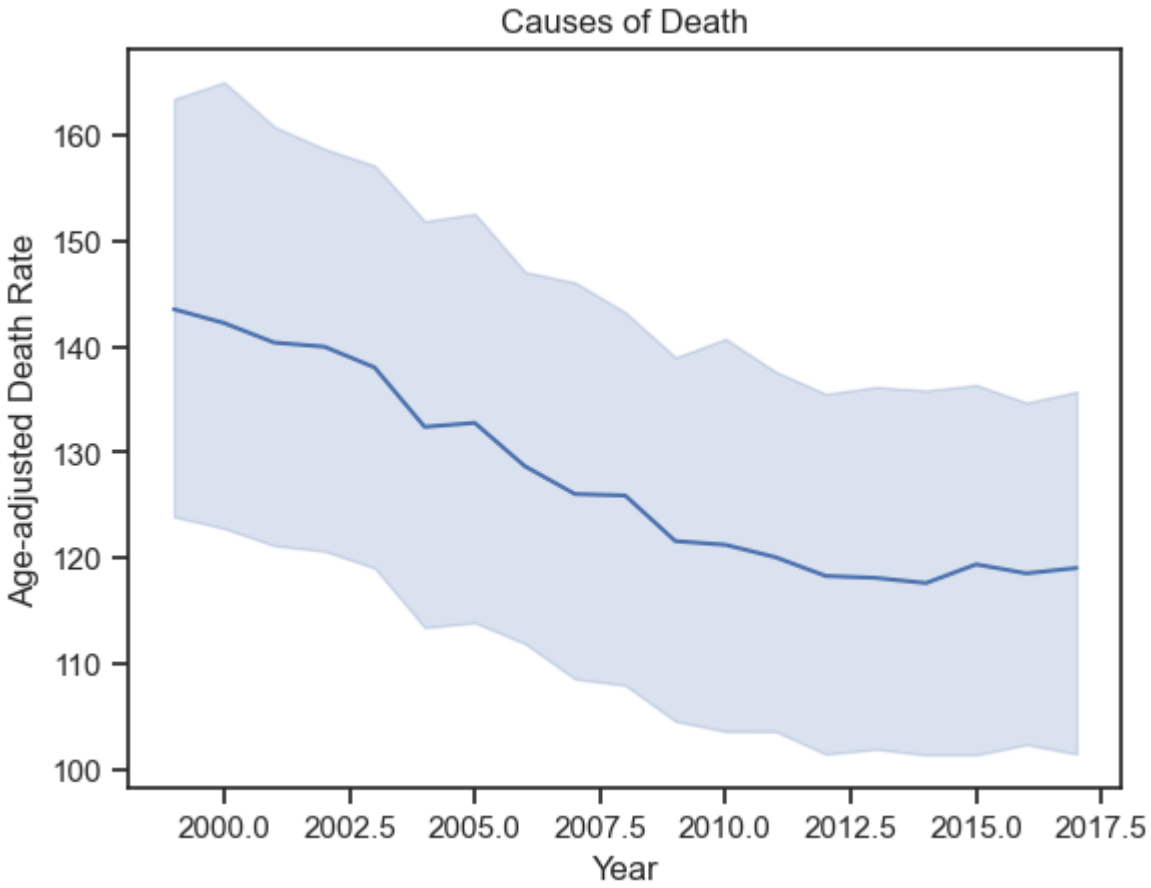
Out[9]:

	Year	113 Cause Name	Cause Name	State	Deaths	Age-adjusted Death Rate
0	2017	Accidents (unintentional injuries) (V01-X59,Y8...	Unintentional injuries	United States	169936	49.4
1	2017	Accidents (unintentional injuries) (V01-X59,Y8...	Unintentional injuries	Alabama	2703	53.8
2	2017	Accidents (unintentional injuries) (V01-X59,Y8...	Unintentional injuries	Alaska	436	63.7
3	2017	Accidents (unintentional injuries) (V01-X59,Y8...	Unintentional injuries	Arizona	4184	56.2
4	2017	Accidents (unintentional injuries) (V01-X59,Y8...	Unintentional injuries	Arkansas	1625	51.8
...	...	...	...	...	...	...
10863	1999	Nephritis, nephrotic syndrome and nephrosis (N...	Kidney disease	Virginia	1035	16.9
10864	1999	Nephritis, nephrotic syndrome and nephrosis (N...	Kidney disease	Washington	278	5.2
10865	1999	Nephritis, nephrotic syndrome and nephrosis (N...	Kidney disease	West Virginia	345	16.4
10866	1999	Nephritis, nephrotic syndrome and nephrosis (N...	Kidney disease	Wisconsin	677	11.9
10867	1999	Nephritis, nephrotic syndrome and nephrosis (N...	Kidney disease	Wyoming	30	6.8

10868 rows × 6 columns

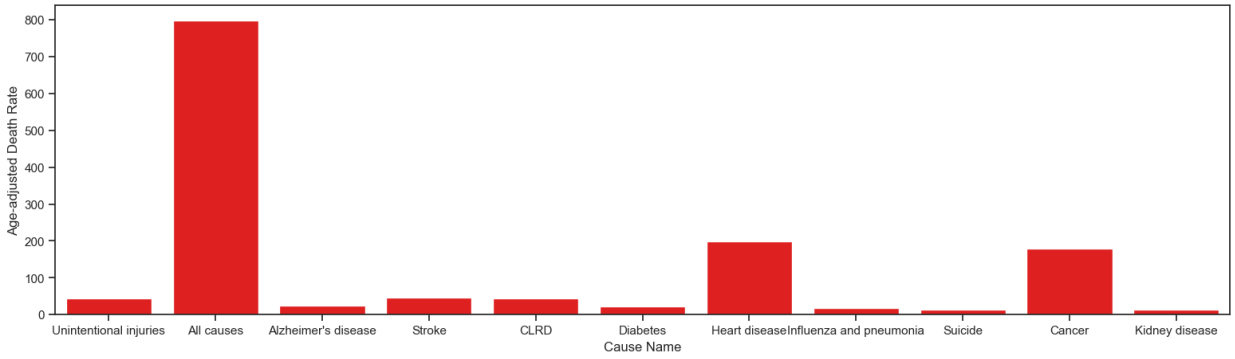
Line Plot

In [12]:
```
sns.lineplot(x="Year", y="Age-adjusted Death Rate", data=data2)
plt.title("Causes of Death")
plt.show()
```



Bar Plot

In [24]:
```
plt.figure(figsize=(19, 5))
sns.barplot(x="Cause Name",y="Age-adjusted Death Rate",data=data2,color="red",ci= None
plt.show()
```



.

3) Crime Dataset Visualisation

In [27]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
data3 = pd.read_csv('C:\\Users\\Lenovo\\Desktop\\Crime1.csv')
data3
```

Out[27]:

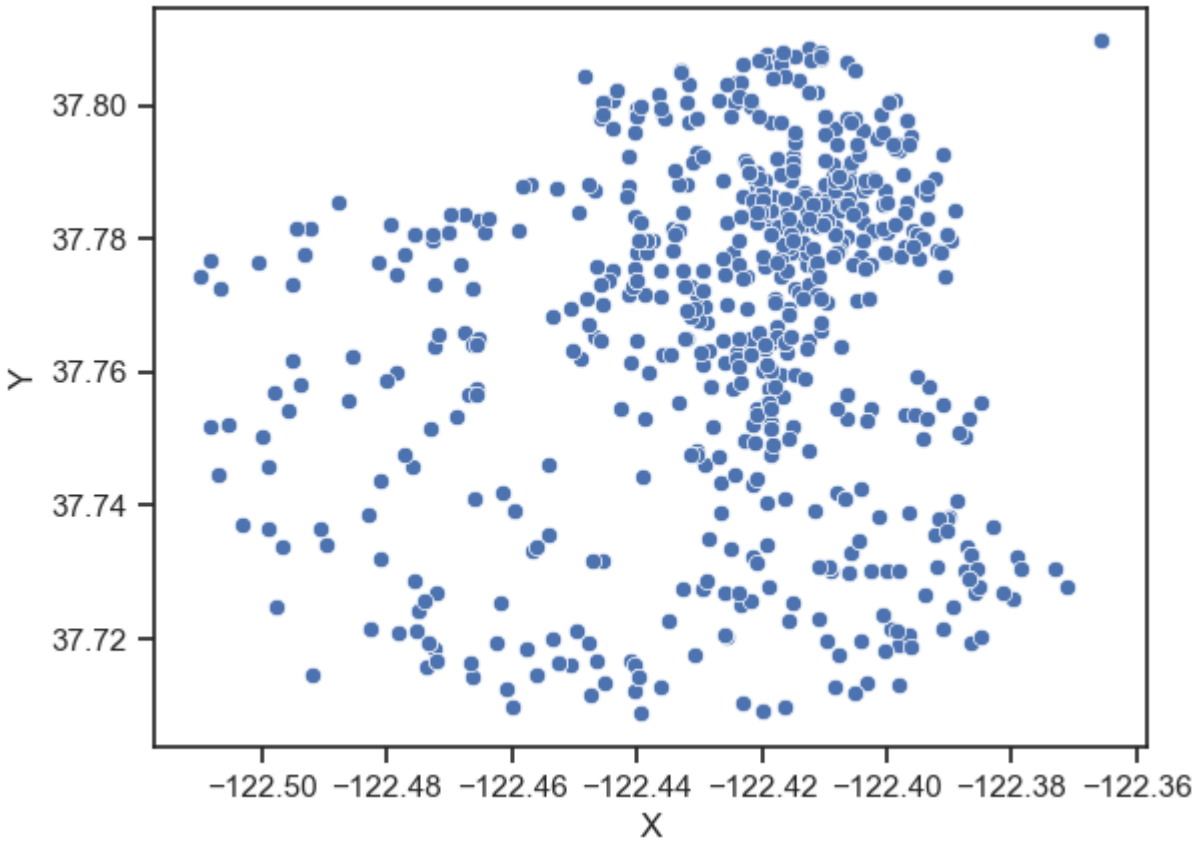
	Dates	Category	Descript	DayOfWeek	PdDistrict	Resolution	Address	
0	5/13/2015 23:53	WARRANTS	WARRANT ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.4
1	5/13/2015 23:53	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.4
2	5/13/2015 23:33	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	VANNESS AV / GREENWICH ST	-122.4
3	5/13/2015 23:30	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Wednesday	NORTHERN	NONE	1500 Block of LOMBARD ST	-122.4
4	5/13/2015 23:30	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Wednesday	PARK	NONE	100 Block of BRODERICK ST	-122.4
...	...	...	...	...	...	...	...	
830	4/30/2015 11:04	NON-CRIMINAL	AIDED CASE, MENTAL DISTURBED	Thursday	MISSION	NONE	3700 Block of 17TH ST	-122.4
831	4/30/2015 11:00	WEAPON LAWS	POSSESSION OF AIR GUN	Thursday	INGLESIDE	NONE	500 Block of YALE ST	-122.4
832	4/30/2015 11:00	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Thursday	SOUTHERN	NONE	400 Block of 10TH ST	-122.4
833	4/30/2015 11:00	OTHER OFFENSES	TAMPERING WITH A VEHICLE	Thursday	SOUTHERN	NONE	0 Block of NORFOLK ST	-122.4
834	4/30/2015 11:00	FRAUD	FALSE PRETENSES, GRAND THEFT	Thursday	MISSION	NONE	0 Block of VIRGIL ST	-122.4

835 rows × 9 columns

Scatter Plot

In [31]:
```
sns.scatterplot(x="X",y="Y",palette="ch:r=-.2,d=.3_r",data=data3)
```

Out[31]:
```
<AxesSubplot:xlabel='X', ylabel='Y'>
```



In [36]:
```
sns.factorplot(x='X', y = 'Category', data=data3,kind="bar", size=4.25, aspect=1.9, pa
plt.title('Factorplot of the category of crime and number of occurences ')
```

Out[36]:
```
Text(0.5, 1.0, 'Factorplot of the category of crime and number of occurences ')
```