In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set()
import warnings
warnings.filterwarnings('ignore') # ignore warnings.
```

In [2]:
```
data=pd.read_csv('C:/Users/VIP/Downloads/U.S._Chronic_Disease_Indicators1.csv'
data
```

Out[2]:

	YearStart	YearEnd	LocationDesc	DataSource	Topic	Question	Response
0	2010	2010	Oregon	NVSS	Cardiovascular Disease	Mortality from heart failure	NaN
1	2019	2019	Arizona	YRBSS	Alcohol	Alcohol use among youth	NaN
2	2019	2019	Ohio	YRBSS	Alcohol	Alcohol use among youth	NaN
3	2019	2019	United States	YRBSS	Alcohol	Alcohol use among youth	NaN
4	2015	2015	Virgin Islands	YRBSS	Alcohol	Alcohol use among youth	NaN
...	...	...	...	...	...	...	..
1048570	2020	2020	South Carolina	BRFSS	Diabetes	Adults with diagnosed diabetes aged >= 18 year...	NaN
1048571	2020	2020	South Carolina	BRFSS	Overarching Conditions	Self-rated health status among women aged 18-4...	NaN
1048572	2015	2015	South Dakota	BRFSS	Arthritis	Physical inactivity among adults aged >= 18 ye...	NaN
1048573	2021	2021	South Dakota	BRFSS	Asthma	Pneumococcal vaccination among noninstitutiona...	NaN
1048574	2018	2018	South Carolina	BRFSS	Diabetes	Prevalence of depressive disorders among adult...	NaN

1048575 rows × 11 columns

Data cleaning

In [3]:
```
df=data.dropna(axis=1)
df=df.reset_index(drop=True)
df
```

Out[3]:

	YearStart	YearEnd	LocationDesc	DataSource	Topic	Question	Stratificat
0	2010	2010	Oregon	NVSS	Cardiovascular Disease	Mortality from heart failure	
1	2019	2019	Arizona	YRBSS	Alcohol	Alcohol use among youth	
2	2019	2019	Ohio	YRBSS	Alcohol	Alcohol use among youth	
3	2019	2019	United States	YRBSS	Alcohol	Alcohol use among youth	
4	2015	2015	Virgin Islands	YRBSS	Alcohol	Alcohol use among youth	
...	...	...	...	...	...	...	
1048570	2020	2020	South Carolina	BRFSS	Diabetes	Adults with diagnosed diabetes aged >= 18 year...	
1048571	2020	2020	South Carolina	BRFSS	Overarching Conditions	Self-rated health status among women aged 18-4...	
1048572	2015	2015	South Dakota	BRFSS	Arthritis	Physical inactivity among adults aged >= 18 ye...	
1048573	2021	2021	South Dakota	BRFSS	Asthma	Pneumococcal vaccination among noninstitutiona...	
1048574	2018	2018	South Carolina	BRFSS	Diabetes	Prevalence of depressive disorders among adult...	

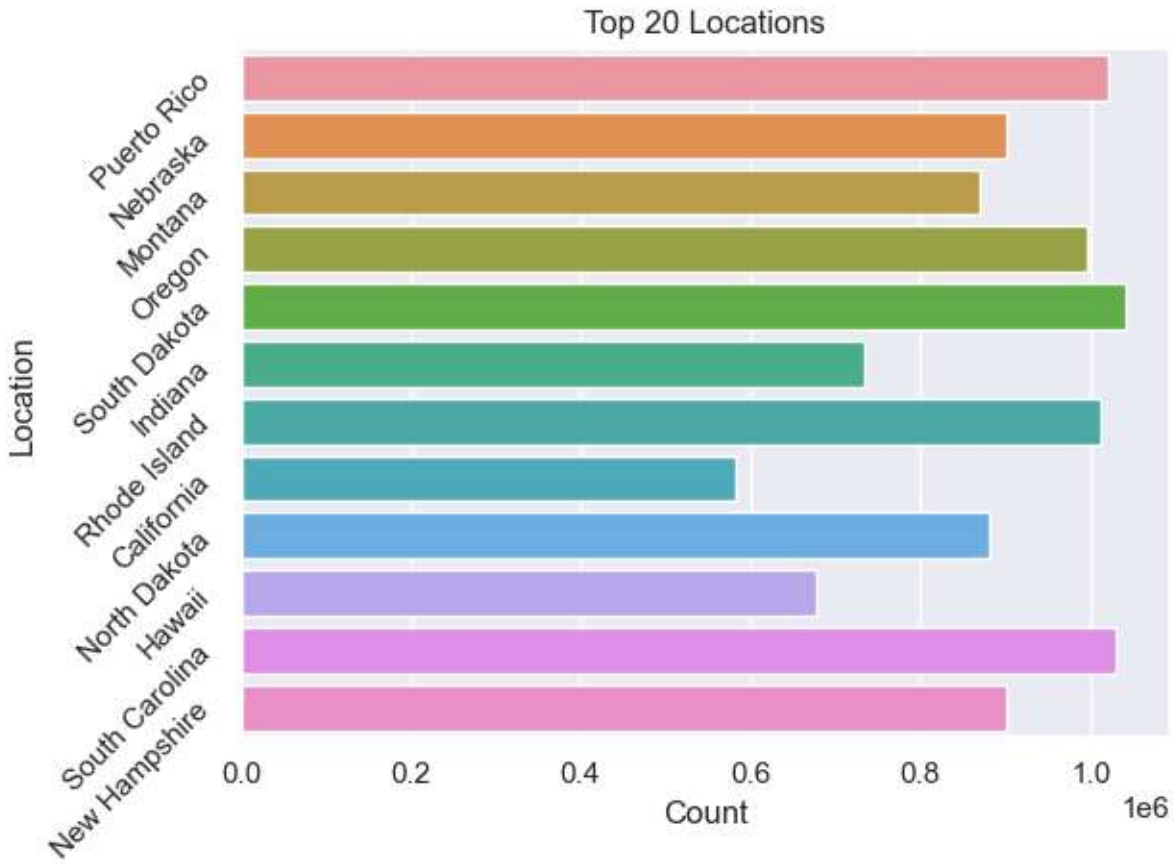
1048575 rows × 7 columns

In [4]:
```
top20LOC= df.sort_values(by='YearStart', ascending=False)['LocationDesc'].head(
print(top20LOC)
```

```
1022286       Puerto Rico
901909           Nebraska
869018            Montana
995255             Oregon
1041490      South Dakota
734651            Indiana
1012926      Rhode Island
734645            Indiana
582135         California
881772       North Dakota
734635            Indiana
582148         California
677616             Hawaii
582979         California
677619             Hawaii
1029636    South Carolina
901907      New Hampshire
995263             Oregon
734621            Indiana
995264             Oregon
Name: LocationDesc, dtype: object
```

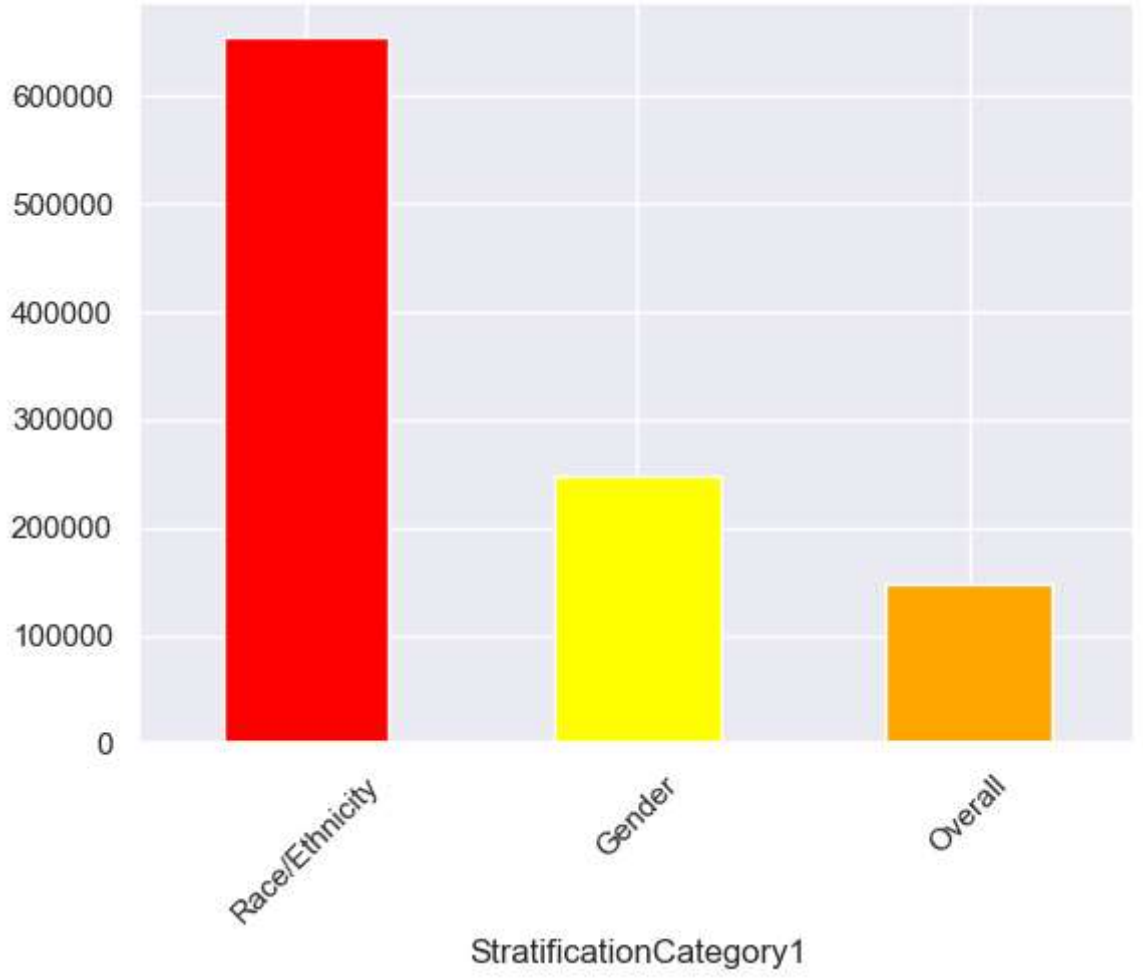
```
# to illustarte the top 20 locationDesc grouped by yearstart for showing relationship between yearstart and location in chronic disease
```

In [5]:
```
sns.barplot(x=df.sort_values(by='YearStart', ascending=False)['LocationDesc'].
plt.ylabel('Location')
plt.xlabel('Count')
plt.title('Top 20 Locations')

plt.yticks(rotation=45)
plt.tight_layout()
plt.show()
```



In [6]:
```
c = ['red', 'yellow', 'orange']
df['StratificationCategory1'].value_counts().plot(kind='bar',color=c)
plt.xticks(rotation=45)
```

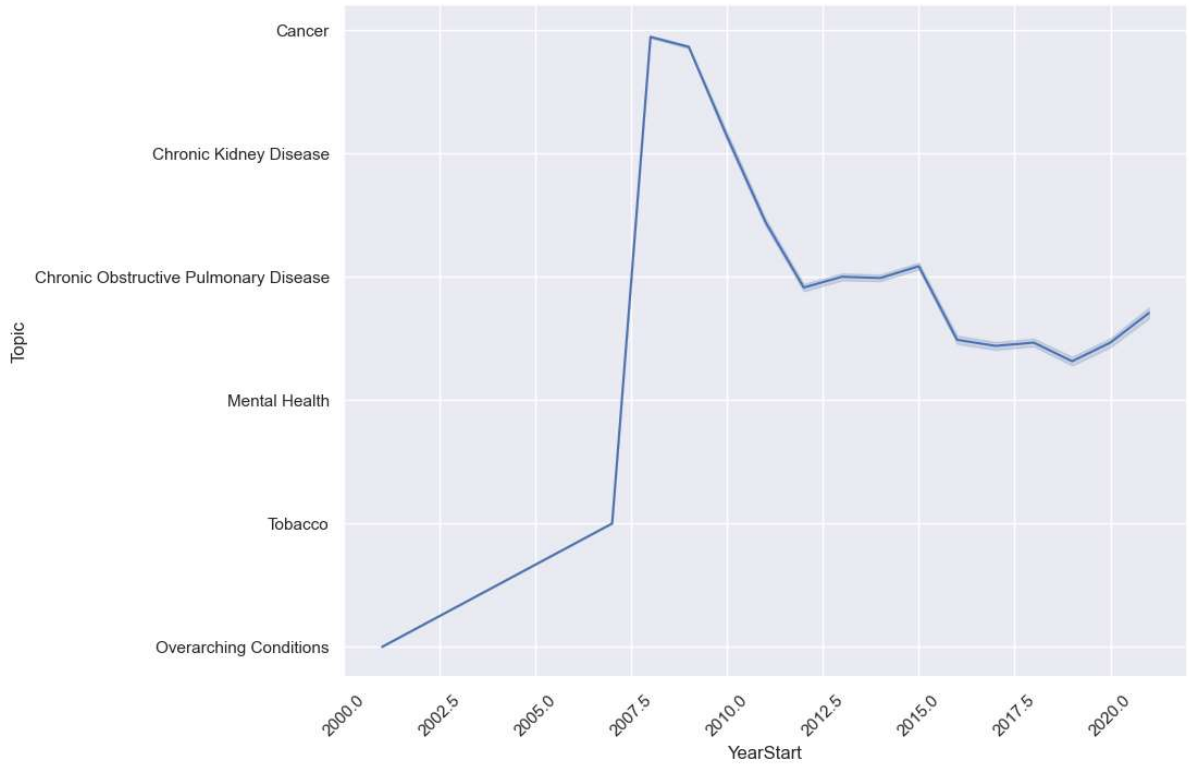
Out[6]:
```
(array([0, 1, 2]),
 [Text(0, 0, 'Race/Ethnicity'), Text(1, 0, 'Gender'), Text(2, 0, 'Overall')])
```



to show topics(diseases) through yearstart

In [7]:
```
plt.figure(figsize=(10,8))
sns.lineplot(x='YearStart', y='Topic',data=df)
plt.xticks(rotation=45)
```

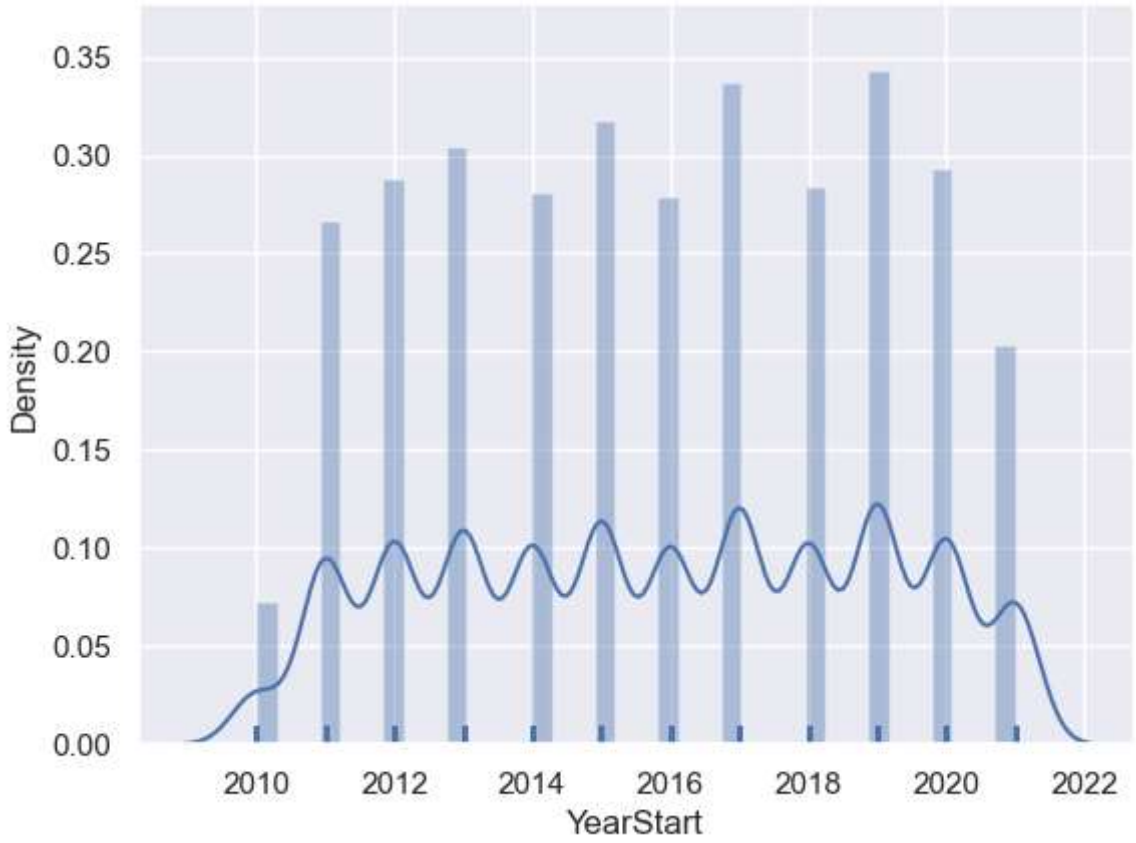
Out[7]:
```
(array([2000. , 2002.5, 2005. , 2007.5, 2010. , 2012.5, 2015. , 2017.5,
        2020. , 2022.5]),
 [Text(2000.0, 0, '2000.0'),
  Text(2002.5, 0, '2002.5'),
  Text(2005.0, 0, '2005.0'),
  Text(2007.5, 0, '2007.5'),
  Text(2010.0, 0, '2010.0'),
  Text(2012.5, 0, '2012.5'),
  Text(2015.0, 0, '2015.0'),
  Text(2017.5, 0, '2017.5'),
  Text(2020.0, 0, '2020.0'),
  Text(2022.5, 0, '2022.5')])
```



to show relationship between alcohol and yearstart that are showwn the most influence in 2019 for people drink alcohol

In [8]:
```
sns.distplot(df[df['Topic']=='Alcohol']['YearStart'], rug=True)
```

Out[8]:
```
<Axes: xlabel='YearStart', ylabel='Density'>
```



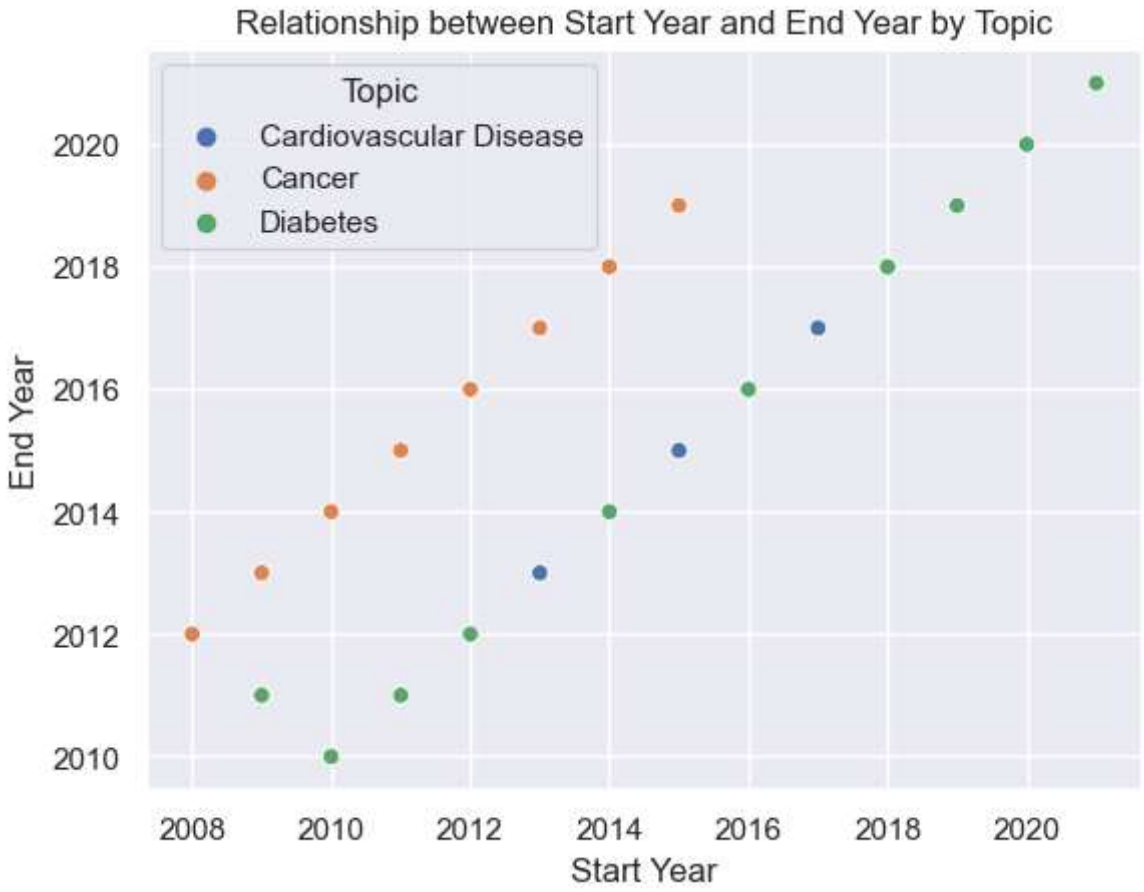
 # i did a selected topic from topic column to show relation between (diabetes,cancer and cardiovascular) on yearstart and endstart

In [31]:
```
selected_topics = ['Diabetes', 'Cancer', 'Cardiovascular Disease']
filtered_data = df[df['Topic'].isin(selected_topics)]
print(filtered_data)
```

```
         YearStart  YearEnd    LocationDesc  \
0             2010     2010          Oregon
104           2013     2013         Indiana
264           2013     2017          Alaska
265           2014     2018          Alaska
266           2009     2013     Connecticut
...            ...      ...             ...
1048549       2015     2015  South Carolina
1048562       2013     2013  South Carolina
1048569       2016     2016  South Carolina
1048570       2020     2020  South Carolina
1048574       2018     2018  South Carolina

                                  DataSource                    Topic  \
0                                       NVSS  Cardiovascular Disease
104                                     NVSS  Cardiovascular Disease
264      Statewide central cancer registries                  Cancer
265      Statewide central cancer registries                  Cancer
266      Statewide central cancer registries                  Cancer
...                                      ...                     ...
1048549                                BRFSS  Cardiovascular Disease
1048562                                BRFSS  Cardiovascular Disease
1048569                                BRFSS                Diabetes
1048570                                BRFSS                Diabetes
1048574                                BRFSS                Diabetes

                                                 Question  \
0                            Mortality from heart failure
104                          Mortality from heart failure
264      Invasive cancer of the oral cavity or pharynx,...
265      Invasive cancer of the oral cavity or pharynx,...
266      Invasive cancer of the oral cavity or pharynx,...
...                                                    ...
1048549  Pneumococcal vaccination among noninstitutiona...
1048562  Influenza vaccination among noninstitutionaliz...
1048569  Dilated eye examination among adults aged >= 1...
1048570  Adults with diagnosed diabetes aged >= 18 year...
1048574  Prevalence of depressive disorders among adult...

        StratificationCategory1
0                Race/Ethnicity
104                      Gender
264                      Gender
265                     Overall
266                      Gender
...                         ...
1048549          Race/Ethnicity
1048562          Race/Ethnicity
1048569                  Gender
1048570          Race/Ethnicity
1048574          Race/Ethnicity

[450788 rows x 7 columns]
```

In [34]:
```
sns.scatterplot(data=filtered_data,x='YearStart',y='YearEnd',hue='Topic')
plt.xlabel('Start Year')
plt.ylabel('End Year')
plt.title('Relationship between Start Year and End Year by Topic')
```

Out[34]: Text(0.5, 1.0, 'Relationship between Start Year and End Year by Topic')



In [ ]:

In [ ]: