Project2

November 30, 2022

1 Project 2 - Life Expectancy in the US

1.0.1 Author: Tristan McCoy

This project is a continuation of Project 1. Project 1 attemped to address why, in 2015, the life expecancy in the US decreased instead of following its positive trend. I will clarify this alanysis with new datasets which will update the original datasets and allow me to revise my conlcusion. This update will add data from 2016 to 2020.

1.0.2 Importing Packages and Reading Data Files

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

[2]:
```
le1 = pd.read_csv('~/Desktop/Classes/DataScience/DAT301/Project2/
 ↪Life_Expectancy_00_15.csv', sep = ';')
DeathCause = pd.read_csv('~/Desktop/Classes/DataScience/DAT301/Project2/
 ↪cause_of_deaths.csv')
le2 = pd.read_csv("~/Desktop/Classes/DataScience/DAT301/Project2/API_SP.DYN.
 ↪LE00.IN_DS2_en_csv_v2_4700946.csv")
he1 = pd.read_csv("~/Desktop/Classes/DataScience/DAT301/Project2/API_SH.XPD.
 ↪CHEX.GD.ZS_DS2_en_csv_v2_4690193.csv")
iu1 = pd.read_csv("~/Desktop/Classes/DataScience/DAT301/Project2/API_IT.NET.
 ↪USER.ZS_DS2_en_csv_v2_4685117.csv")
```

1.1 Data Wrangling

1.1.1 Life Expectancy

[3]:
```
le3 = le2[le2['Country Name'] == 'United States']
le3 = le3[['2016', '2017', '2018', '2019', '2020']]
le3 = le3.T
le3 = le3.rename(columns={251:'Life Expectancy'})
le3['Year'] = [2016,2017,2018,2019,2020]
le3 = le3[['Year','Life Expectancy']]
LifeExp = le1[le1['Country'] == 'United States']
LifeExp = LifeExp[['Year','Life Expectancy']]
```

```
LifeExp = pd.concat([LifeExp,le3])
LifeExp = LifeExp.reset_index(drop=True)
```

1.1.2 Health Expenditure

[4]:
```
he2 = he1[he1['Country Name']=='United States']
he2 = he2[['2016','2017','2018','2019']]
he2 = he2.T
he2 = he2.rename(columns={251:'Health expenditure'})
he2['Year'] = [2016,2017,2018,2019]
he2 = he2[['Year','Health expenditure']]
HealthExp = le1[le1['Country']=='United States']
HealthExp = HealthExp[['Year','Health expenditure']]
HealthExp = pd.concat([HealthExp,he2])
HealthExp = HealthExp.reset_index(drop=True)
```

1.1.3 Internet Usage

[5]:
```
iu2 = iu1[iu1['Country Name']=='United States']
iu2 = iu2[['2016','2017','2018','2019','2020']]
iu2 = iu2.T
iu2 = iu2.rename(columns={251:'Individuals using the Internet'})
iu2['Year'] = [2016,2017,2018,2019,2020]
iu2 = iu2[['Year','Individuals using the Internet']]
IntUse = le1[le1['Country']=='United States']
IntUse = IntUse[['Year','Individuals using the Internet']]
IntUse = pd.concat([IntUse,iu2])
IntUse = IntUse.reset_index(drop=True)
```

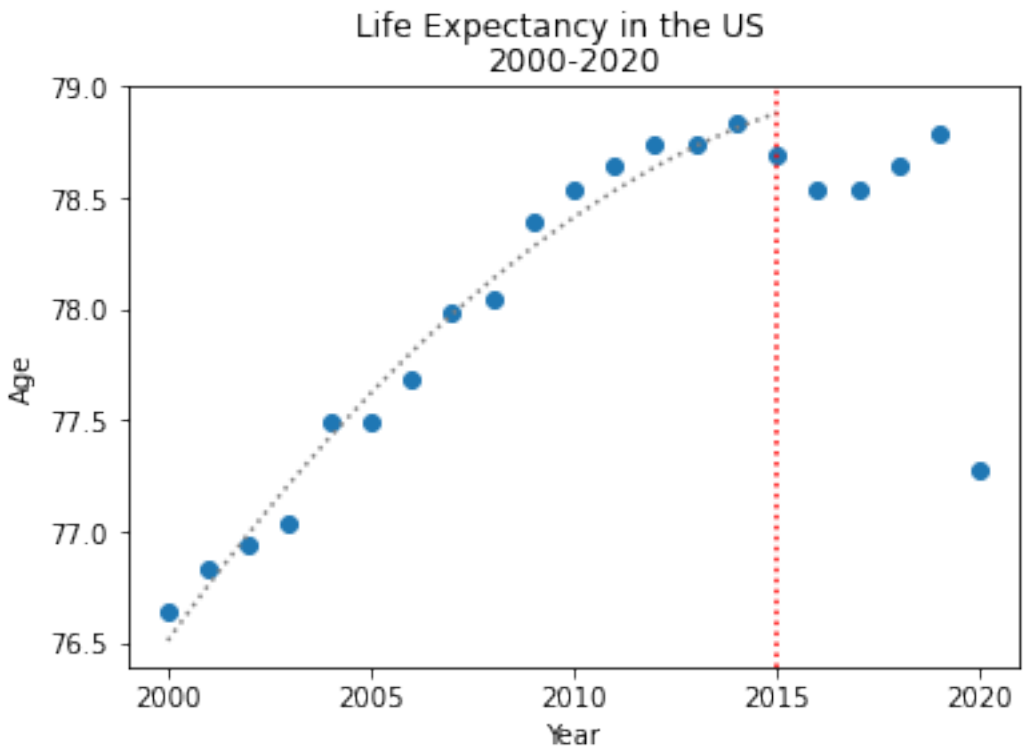
2 Introduction

2.0.1 Life Expectancy Analysis

[6]:
```
# plot
plt.scatter(LifeExp['Year'],LifeExp['Life Expectancy'])
plt.axvline(2015,c='red',ls=':')
plt.xticks(np.arange(2000,2021,step=5))
plt.xlabel("Year")
plt.ylabel("Age")
plt.suptitle("Life Expectancy in the US")
plt.title("2000-2020")

# trend line 2000-2015
x = np.polyfit(LifeExp['Year'][0:16],LifeExp["Life Expectancy"][0:16],2)
y = np.poly1d(x)
plt.plot(LifeExp['Year'][0:16],y(LifeExp['Year'][0:16]),color='grey',ls=':')
```

[6]: [<matplotlib.lines.Line2D at 0x7f923a677fa0>]



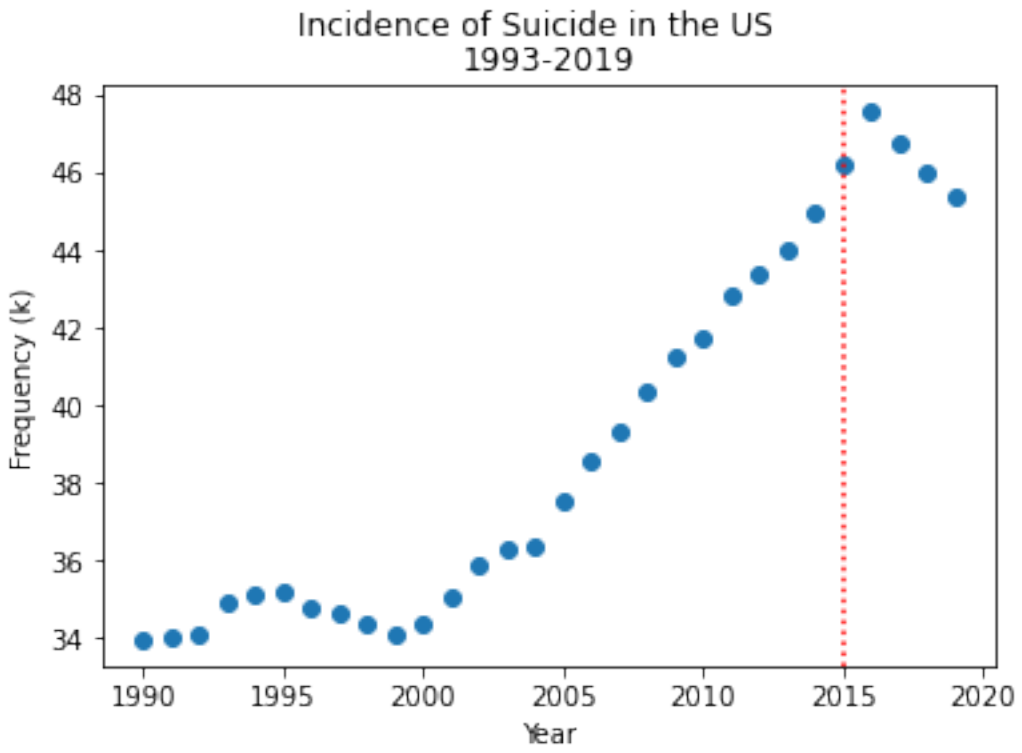
3 Analysis

3.0.1 Suicide

[7]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

# plot
plt.scatter(df['Year'],df['Self-harm']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Suicide in the US")
plt.title("1993-2019")
```

[7]: Text(0.5, 1.0, '1993-2019')

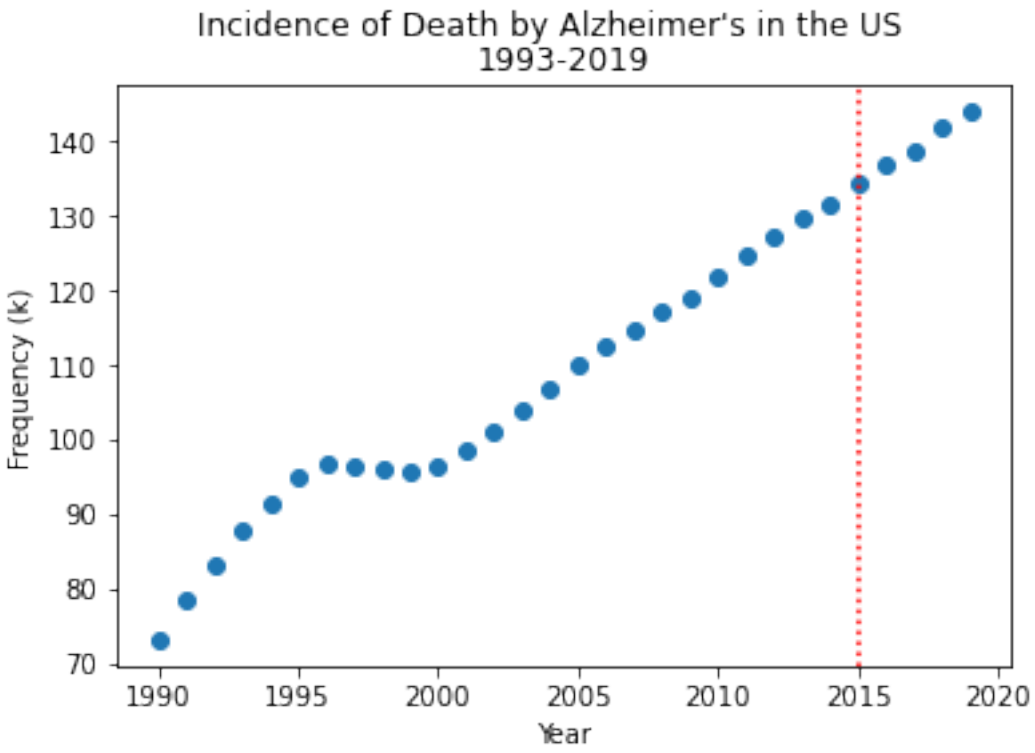


3.0.2 Alzheimer's (and other dementias)

[8]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

#plot
plt.scatter(df['Year'],df["Alzheimer's Disease and Other Dementias"]/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Alzheimer's in the US")
plt.title("1993-2019")
```

[8]:
```
Text(0.5, 1.0, '1993-2019')
```

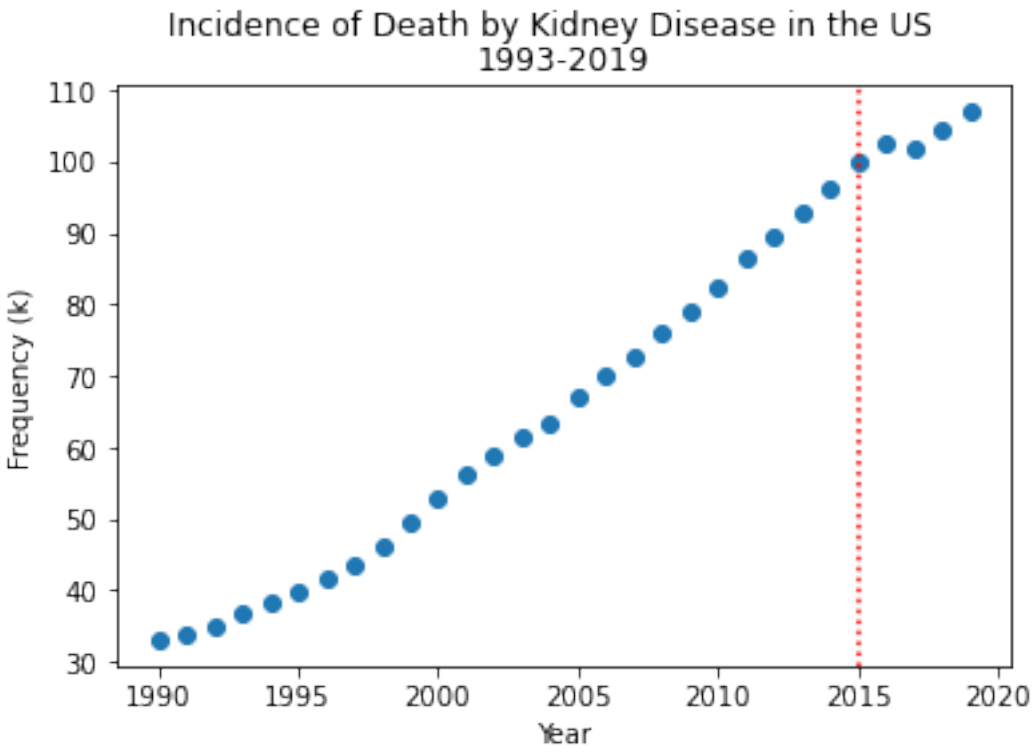


3.0.3 Chronic Kidney Disease

[9]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

#plot
plt.scatter(df['Year'],df['Chronic Kidney Disease']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Kidney Disease in the US")
plt.title("1993-2019")
```

[9]:
```
Text(0.5, 1.0, '1993-2019')
```

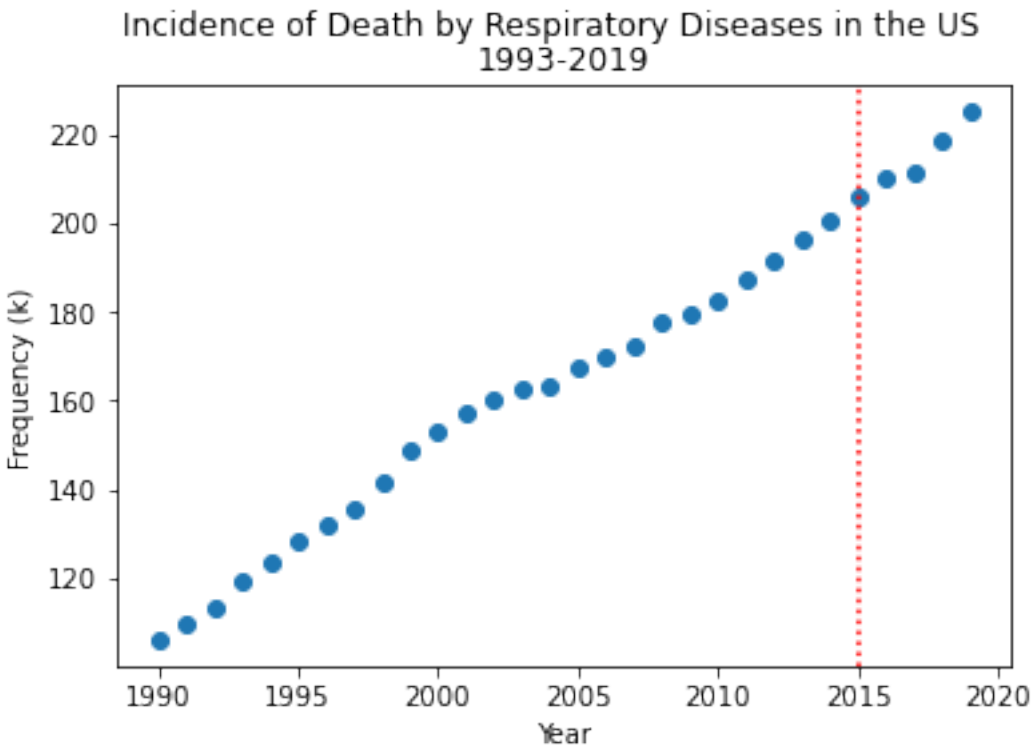


3.0.4 Chronic Respiratory Diseases

[10]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

# plot
plt.scatter(df['Year'],df['Chronic Respiratory Diseases']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Respiratory Diseases in the US")
plt.title("1993-2019")
```

[10]:
```
Text(0.5, 1.0, '1993-2019')
```

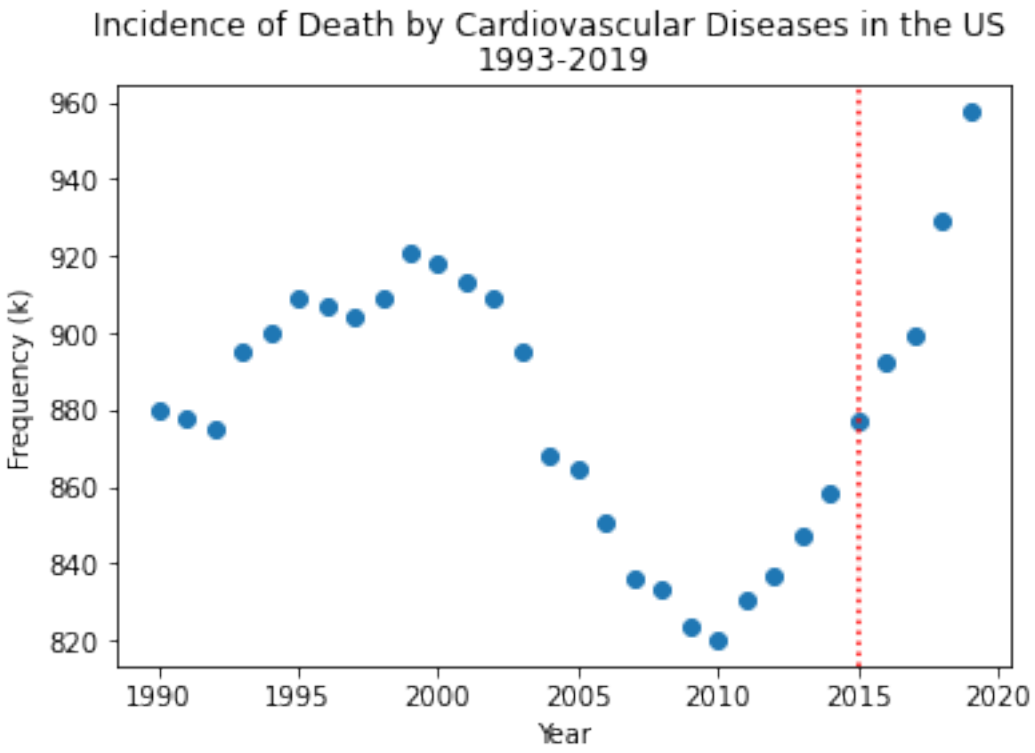


3.0.5 Cardiovascular Diseases (including stroke)

[11]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

plt.scatter(df['Year'],df['Cardiovascular Diseases']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Cardiovascular Diseases in the US")
plt.title("1993-2019")
```

[11]:
```
Text(0.5, 1.0, '1993-2019')
```

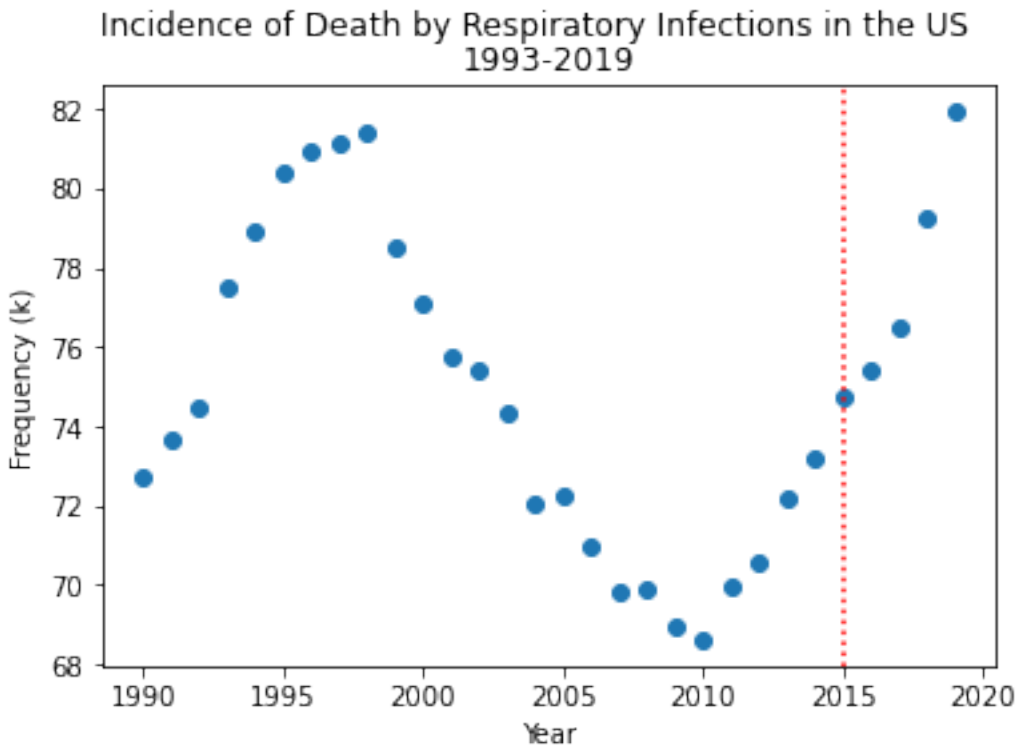


3.0.6 Chronic Respiratory Infections (including pneumonia)

[12]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

plt.scatter(df['Year'],df['Lower Respiratory Infections']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Respiratory Infections in the US")
plt.title("1993-2019")
```

[12]:
```
Text(0.5, 1.0, '1993-2019')
```

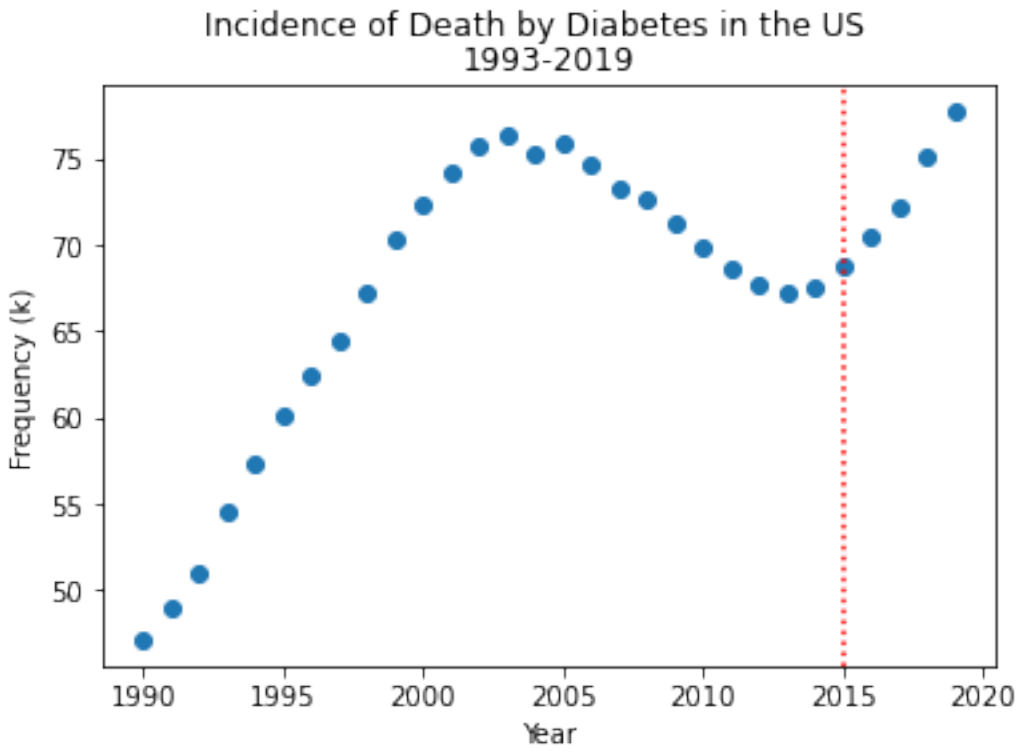


3.0.7 Diabetes

[13]:
```
df = DeathCause[DeathCause['Country/Territory'] == 'United States']

# plot
plt.scatter(df['Year'],df['Diabetes Mellitus']/1000)
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("Frequency (k)")
plt.suptitle("Incidence of Death by Diabetes in the US")
plt.title("1993-2019")
```

[13]:
```
Text(0.5, 1.0, '1993-2019')
```



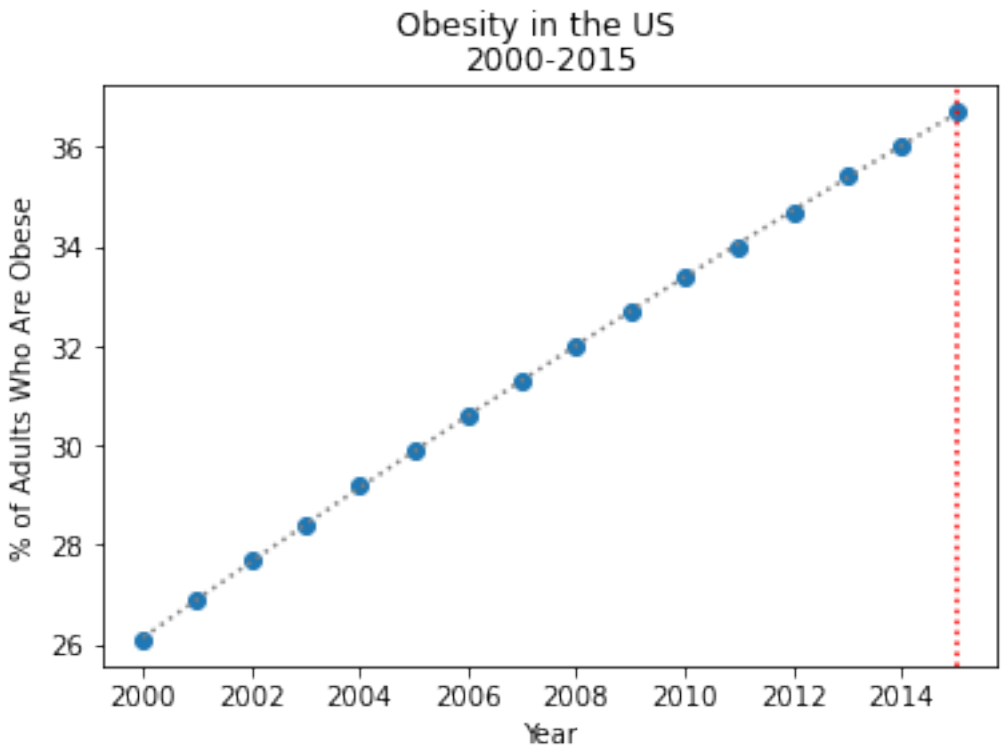
3.0.8 Incidence of Obesity

[14]:
```
df = le1[le1['Country']=='United States']

# plot
plt.scatter(df['Year'],df['Obesity among adults'])
plt.axvline(2015,c='red',ls=':')
plt.xlabel("Year")
plt.ylabel("% of Adults Who Are Obese")
plt.suptitle("Obesity in the US")
plt.title("2000-2015")

# trend line
x = np.polyfit(df['Year'],df['Obesity among adults'],2)
y = np.poly1d(x)
plt.plot(df['Year'],y(df['Year']),color='grey',ls=':')
```

[14]:
```
[<matplotlib.lines.Line2D at 0x7f923bf428e0>]
```

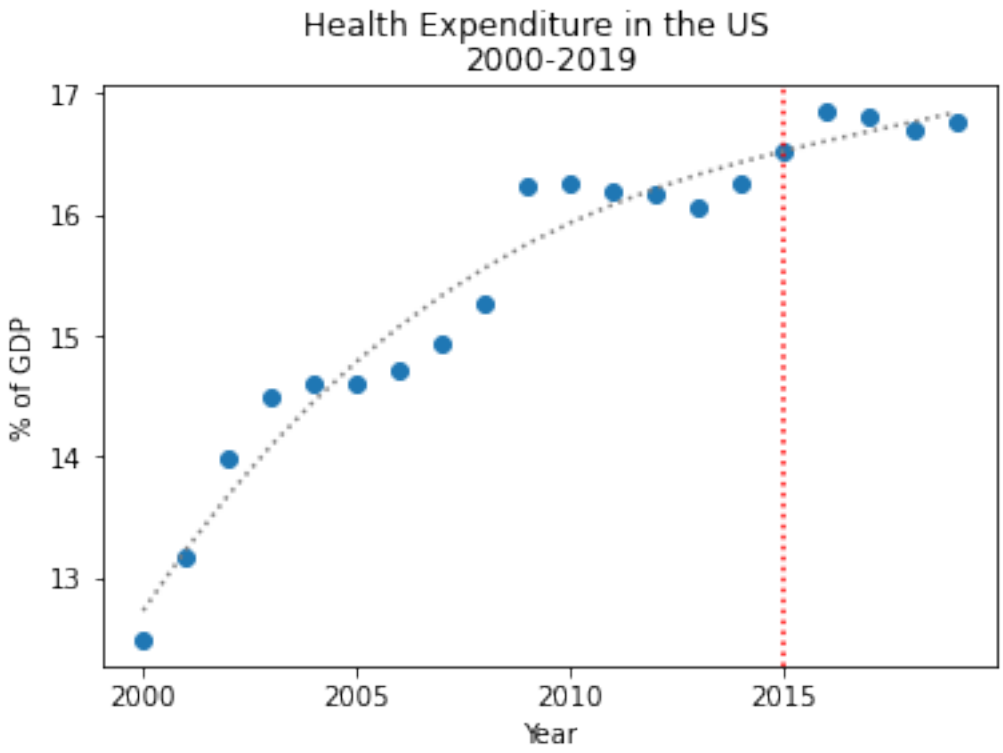


3.0.9 Health Expenditure

[15]:
```
#plot
plt.scatter(HealthExp['Year'],HealthExp['Health expenditure'])
plt.axvline(2015,c='red',ls=':')
plt.xticks(np.arange(2000,2021,step=5))
plt.xlabel("Year")
plt.ylabel("% of GDP")
plt.suptitle("Health Expenditure in the US")
plt.title("2000-2019")

# trend line
x = np.polyfit(HealthExp['Year'],HealthExp['Health expenditure'],3)
y = np.poly1d(x)
plt.plot(HealthExp['Year'],y(HealthExp['Year']),color='grey',ls=':')
```

[15]:
```
[<matplotlib.lines.Line2D at 0x7f923c067520>]
```

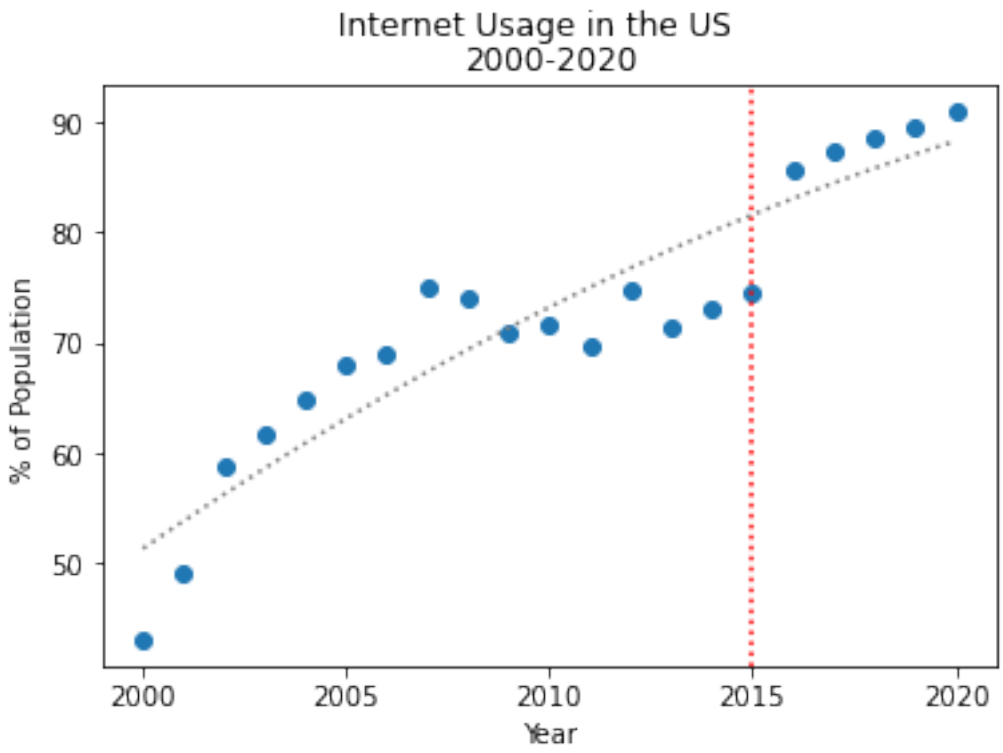


3.0.10 Internet Usage

[16]:
```
# plot
plt.scatter(IntUse['Year'],IntUse['Individuals using the Internet'])
plt.axvline(2015,c='red',ls=':')
plt.xticks(np.arange(2000,2021,step=5))
plt.xlabel("Year")
plt.ylabel("% of Population")
plt.suptitle("Internet Usage in the US")
plt.title("2000-2020")

# trend line
x = np.polyfit(IntUse['Year'],IntUse['Individuals using the Internet'],2)
y = np.poly1d(x)
plt.plot(IntUse['Year'],y(IntUse['Year']),color='grey',ls=':')
```

[16]:
```
[<matplotlib.lines.Line2D at 0x7f923c16f760>]
```



4 Conclusion

The sharp increase in internet usage is a key factor for why life expectancy in the US began to decline in 2015. Inceased internet usage suggests that more Americans are adpopting a sedentary lifestyle which contributes to lower levels of physical wellbeing if adequate exercise is avoided. The trend in sedentism can explain why the incidence of deaths by diabetes and cardiovascular diseases spiked around 2015 and why the incidence of obesity in the US has continued on its strongly positive linear trend. Health expenditure is less of a contributing factor (if at all) considering that it has maintained its irregular positive trend. Of course, correlation does not equal causation so we would need more data exploration to make a certain conclusion as to why life expectancy began to decrease in 2015.

5 Sources

https://www.cdc.gov/nchs/products/databriefs/db267.htm

5.0.1 Life Expectancy

https://www.kaggle.com/datasets/vrec99/life-expectancy-2000-2015

5.0.2 Causes of Death

https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world

5.0.3 Life Expectancy (updated)

https://data.worldbank.org/indicator/SP.DYN.LE00.IN?view=chart

5.0.4 Health Expenditure (updated)

https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

5.0.5 Internet Usage (updated)

https://data.worldbank.org/indicator/IT.NET.USER.ZS