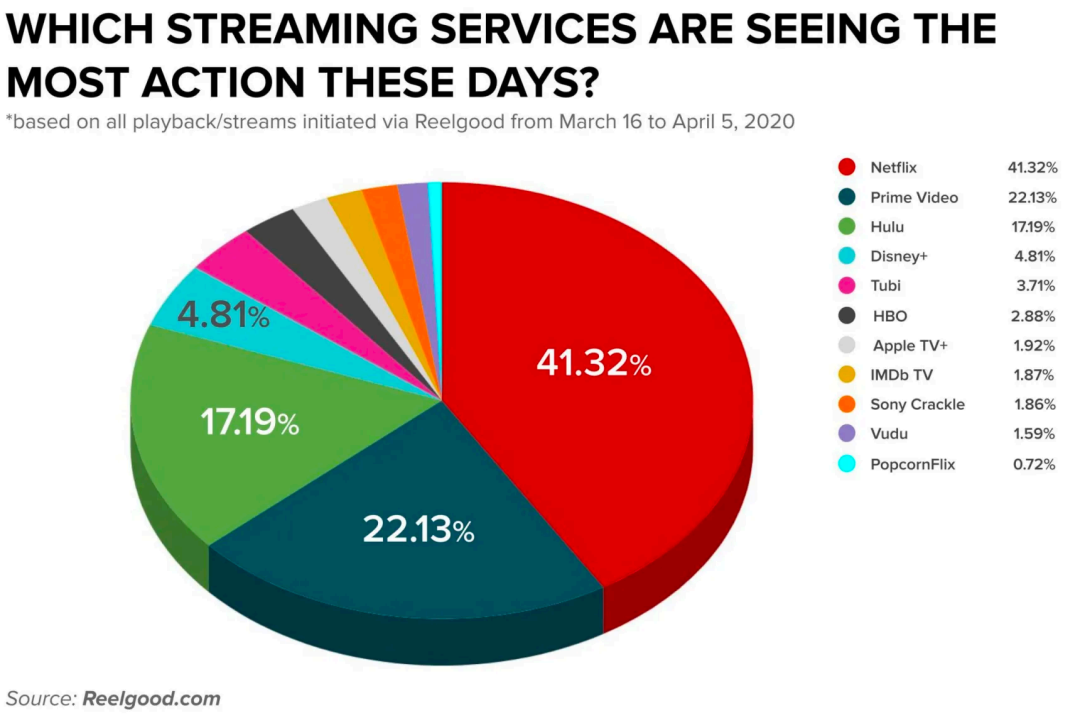
Name: Shrey Shah

UIN: 664406225

Email Id: sshah023@illinois.edu

IS 445 - Data Visualization - Assignment 1

1) Pie Chart for Streaming Platforms



The above image has been obtained from a study conducted by Reelgood organisation to understand the major player among all the streaming platforms used in America.

The type of visualization is a pie chart. The data for the same has been derived from the playback services provided by Reelgood during the Covid-19 period from March 15 to April 5, 2020. It contains of both quantitative and categorical data. The different streaming platforms are the different categories and the proportion of subscribers represent the numbers. The graph might have been generated by conducting a survey/poll among the Reelgood subscribers to choose their preferred streaming platform. The percentages might have been calculated using the number of people for 'X' platform divided by the total subscribers and multiplies by 100. These are then represented as percent of a pie.

Data Replication:

I have taken the data from Commonwealth Games 2022 based on sports categories and the medals earned in each category. The proportion represents the medals earned by all the countries in that specific category. In the process of creating the chart, the data used may not represent or correspond to the exact percentages in the visualization above.

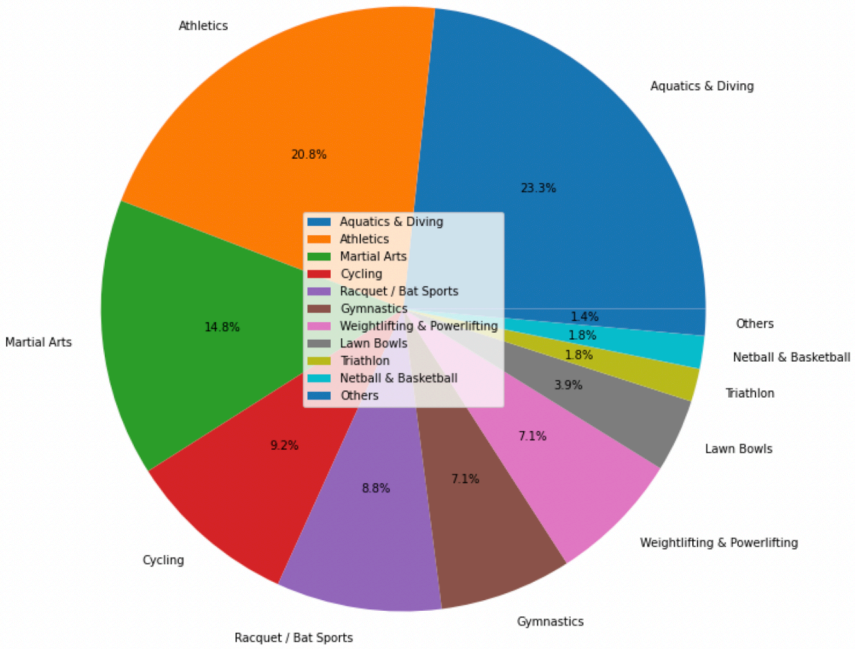
Visualization 1

```
In [1]: import matplotlib.pyplot as plt

        %matplotlib inline
```

```
In [2]: sport_category = ['Aquatics & Diving', 'Athletics', 'Martial Arts', 'Cycling', 'Racquet / Bat Sports', 'Gymnastics',
                          'Weightlifting & Powerlifting', 'Lawn Bowls', 'Triathlon', 'Netball & Basketball', 'Others']
```

```
In [52]: medals = [66,59,42,26,25,20,20,11,5,5,4]
```

```
In [69]: fig, ax = plt.subplots(1,1,figsize=(15,12))

         ax.pie(medals, labels = sport_category, autopct='%1.1f%%')
         plt.legend(sport_category, loc = 'center')
         plt.show()
```



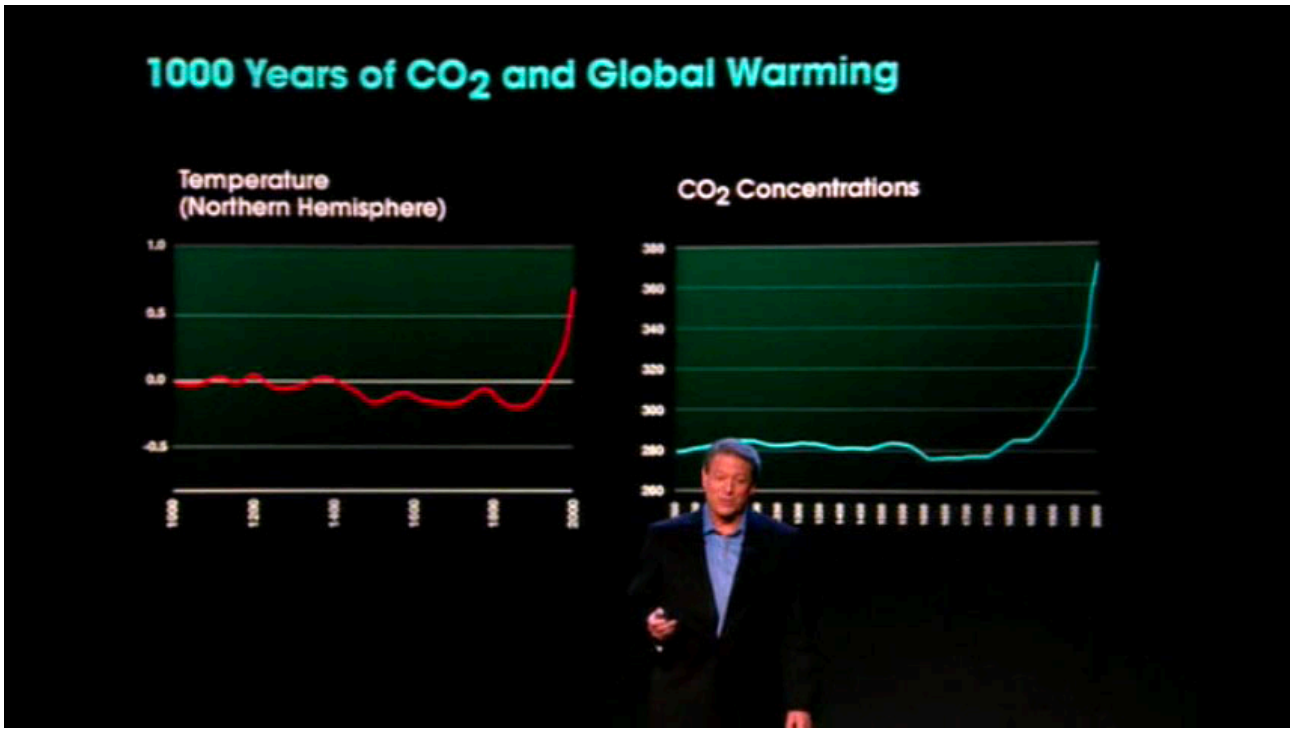
References:

Effects of covid-19 and social distancing on streaming video and ott. UpNext. (2020, August 2). Retrieved September 3, 2022, from https://blog.reelgood.com/effects-of-covid-19-and-social-distancing-on-streaming-video-and-ott

Google. (n.d.). *Battle of the streaming services (SVOD + AVOD).JPG*. Google Drive. Retrieved September 3, 2022, from https://drive.google.com/file/d/1sCp5-fr8etd89lmEb4zzOtDTbn2HC8t1/view

Birmingham Commonwealth Games Facts & Statistics 2022. img:First Mats. (n.d.). Retrieved September 3, 2022, from https://www.firstmats.co.uk/blogs/buying-guides/2022-birmingham-commonwealth-games-facts-figures

2) Line Charts for CO_2 levels and Global Warming



The above image has been obtained from the movie/documentary 'An Inconvenient Truth' from the year 2006.

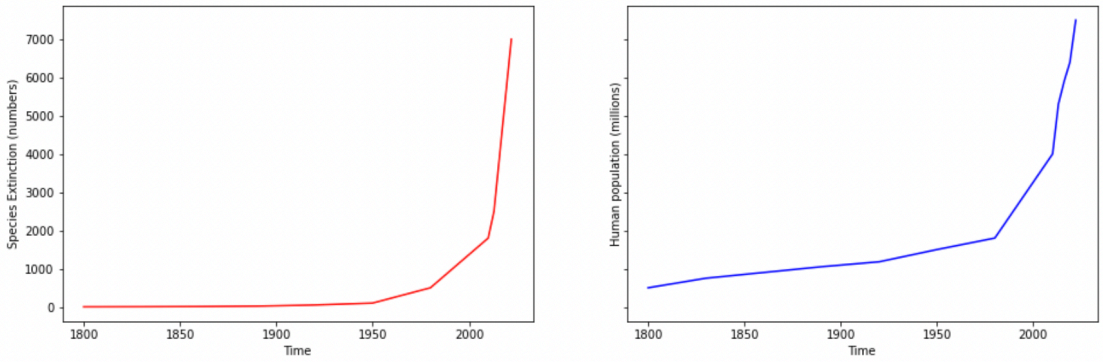
It appears in the documentary in the first 30 minutes where Al Gore is trying to explain the correlation between the increase in temperature in the northern hemisphere as a result of the increase in the CO_2 concentrations from the year 1900 to 2000.The temperature, CO_2 concentrations and time are quantitative and continuous variables in nature. Certain data for the visualisations had been obtained from multiple space missions like 'Earthrise' and 'The Blue Marble'. The temperature data might have been collected from the weather stations and the CO_2 levels might have been collected based on the AQI (Air Quality Index) or ppm levels (parts per million) of CO_2 the northern hemisphere regions. The data, probably, has been collected for every single day for the year and then to depict a year the mean was calculated as a part of processing. The average values for both temperature and CO_2 levels were then plotted against time in years in the form of smoothened line charts.

Data Replication:

I have created a dataset which shows the rate of human population explosion and species extinction from 1800 to 2022 in gaps of 30 years. The graph shows how with increase in human population the species have undergone extinction and the rate at which both have increased is exponential. In my case, I have not smoothened the data and they are connected via straight lines.

Visualization 2

In [163]:
```
time = [1800, 1830, 1860, 1890, 1920, 1950, 1980, 2010, 2013, 2016, 2019, 2022]
human_population = [500, 750, 900, 1050, 1180, 1500, 1800, 4000, 5300, 5900, 6400, 7500]
species_extinction = [2, 5, 10, 20, 50, 100, 500, 1800, 2500, 4000, 5500, 7000]

fig, ax = plt.subplots(1,2,figsize=(16,5), sharex=True, sharey=True)

ax[0].plot(time, species_extinction, color='r')
plt.setp(ax[0:2], xlabel='Time')
plt.setp(ax[0:1], ylabel='Species Extinction (numbers)')

ax[1].plot(time, human_population, color='b')
plt.setp(ax[1:2], ylabel='Human population (millions)')

plt.show()
```



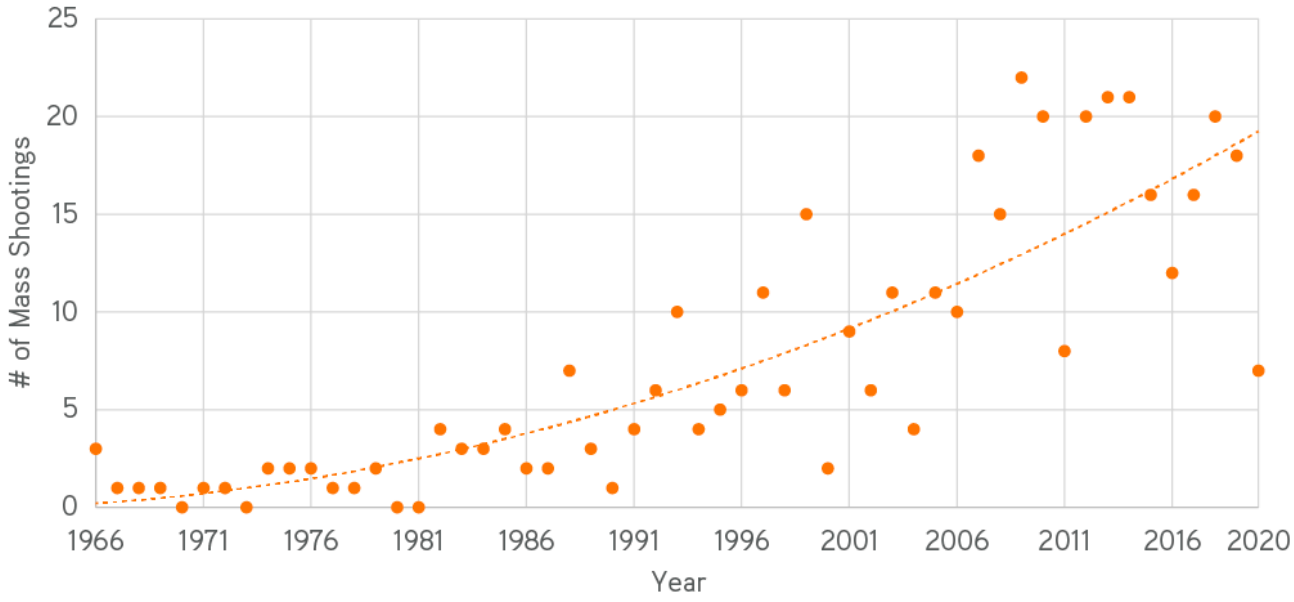
References:

Charts and figures from "An inconvenient truth". (n.d.). Retrieved September 3, 2022, from https://smallpond.ca/jim/ref/inconvenientTruth/

Wikimedia Foundation. (2022, August 29). *An inconvenient truth*. Wikipedia. Retrieved September 3, 2022, from https://en.wikipedia.org/wiki/An_Inconvenient_Truth

Linking behaviour and conservation. Behavioural Ecology And Conservation. (n.d.). Retrieved September 3, 2022, from https://bio-ns6201.weebly.com/linking-behaviour-and-conservation.html

3) Scatter Plot for Mass Shootings in US over the years



The above graph indicates the trend of total number of mass shootings in the US from the year 1966 to 2020.

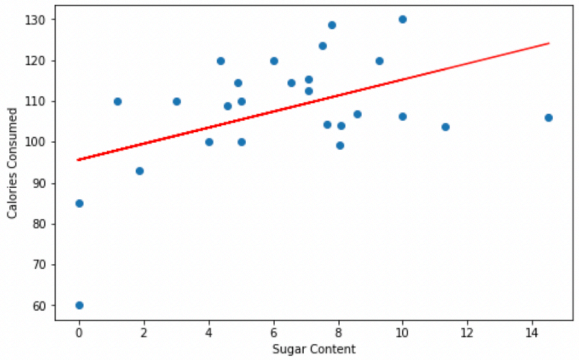
The data has been taken from the Mass Shooting Factsheet by the Rockefeller Institute of Government. The data of 'Year' and the '# of Mass Shootings' both come under the quantitative data. The 'Year' is a continuous variable whereas the '# of Mass Shootings' is a discrete variable. The data might have been collected based on the number of casualties, injured and the number of cases that would have been filed after the incident. The total number of such incidents would have been recorded and the sum of them would be the total number of mass shooting cases for that year. The graph plotted is a scatter plot which takes 2 quantitative variables - Year on the X-axis and the sum on the Y-axis and the intersection point of the 2 values indicates the # of Mass Shootings for that year. If we ignore the smooth line in the graph and draw a linear line of best fit, we can conclude that the overall # of Mass Shootings in the US has increased.

Data Replication:

I have replicated the above graph by considering the 2 variables of sugar content in different cereals and the calories it provides. It consists data points for 26 cereals and the corresponding sugar and calories contents. I have added a trend line using Python code and it shows that the general trend is positive. It means that more sugary the cereal becomes, the calories consumed also increases.

Visualization 3

In [189]:
```
import numpy as np

sugar = [0, 0, 1.2, 1.87, 3, 4, 4.36, 4.6, 4.91, 5, 5, 6, 6.55, 7.09, 7.1, 7.5, 7.65, 7.8, 8.04, 8.1, 8.6, 9.27, 10,
         10, 11.3, 14.5]
calories = [60, 85, 110, 93, 110, 100, 120, 108.9, 114.5, 100, 110, 120, 114.5, 115.3, 112.5, 123.5, 104.3, 128.6,
            99.3, 103.9, 106.9, 120, 130, 106.4, 103.7, 106]

fig, ax = plt.subplots(1,1,figsize=(8,5))

ax.scatter(sugar, calories)
plt.xlabel("Sugar Content")
plt.ylabel("Calories Consumed")

b, a = np.polyfit(sugar, calories, deg=1)
xseq = np.linspace(0, sugar)
plt.plot(xseq, a + b * xseq, color="r");

plt.show()
```



References:

Mass shooting factsheet. Rockefeller Institute of Government. (2022, June 8). Retrieved September 3, 2022, from https://rockinst.org/gun-violence/mass-shooting-factsheet/

Holtz, Y. (n.d.). *Scatterplot with regression line in Matplotlib*. The Python Graph Gallery. Retrieved September 3, 2022, from https://www.python-graph-gallery.com/scatterplot-with-regression-fit-in-matplotlib