COMP 4304 / 6934 – Assignment 4 (10%)

Due: 11:59pm, Mar. 10, 2023

Learning Objectives

You will gain experience with Matplotlib and Seaborn in this assignment.

Instructions

Download the following data sets from Brightspace for this assignment:

- `population.csv`
- `olympic_athletes.csv`
- `game_consoles.csv`

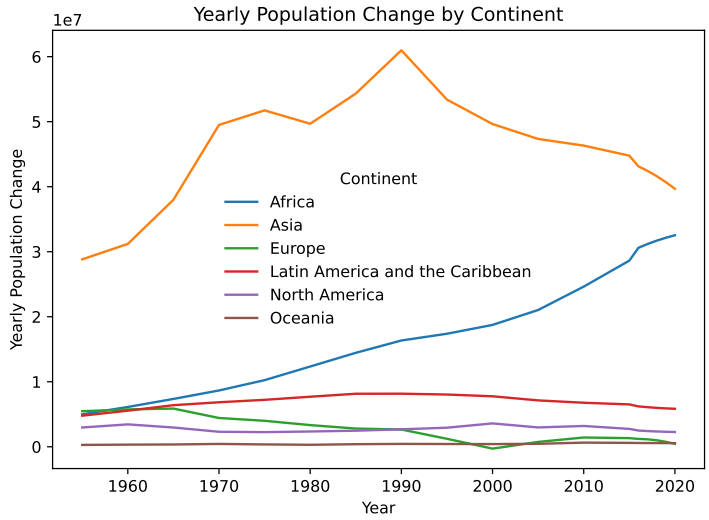
The first two data sets we have used in previous assignments. `population.csv` contains information on the population of countries around the world over time and `olympic_athletes.csv` contains information about all the athletes that have competed in the Olympics over time.

The `game_consoles.csv` data set contains information on video game consoles released by Nintendo, Microsoft and Sony over the past few decades. It has the release date for each console, the type of console (home, handheld, etc), and the number of units that were sold.

Using the provided data sets, create a Jupyter notebook that answers the following questions. You may only import the pandas, Matplotlib, Seaborn and Matplotlib-venn libraries.

Question 1: (20 pts)

Use Seaborn to recreate the plot from Question 3 of Assignment 2. This plot shows the yearly population change (raw value) per continent over the past 60 years.

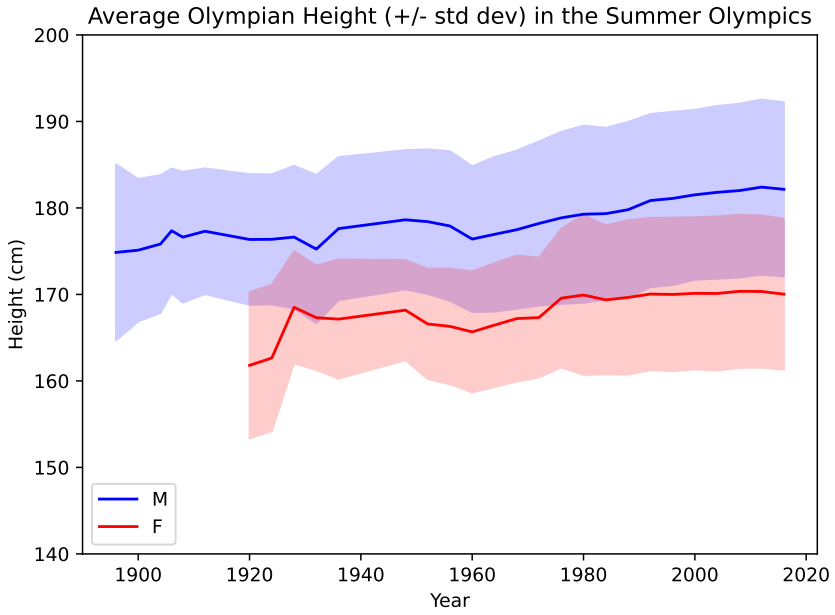


The line plot should be made using Seaborn's `lineplot()` functionality.

The legend should be in the centre of the figure and without a border. The ordering of continents should be alphabetical, and the colours should match the figure above. Make sure to include the axes labels and plot title. There should not be any shading around the lines (which Seaborn might add depending upon how you use it).

Question 2: (20 pts)

Use Seaborn to recreate the plot from Question 2 of Assignment 3. This plot shows the average height of male and female Olympians over the history of the Summer Olympics.

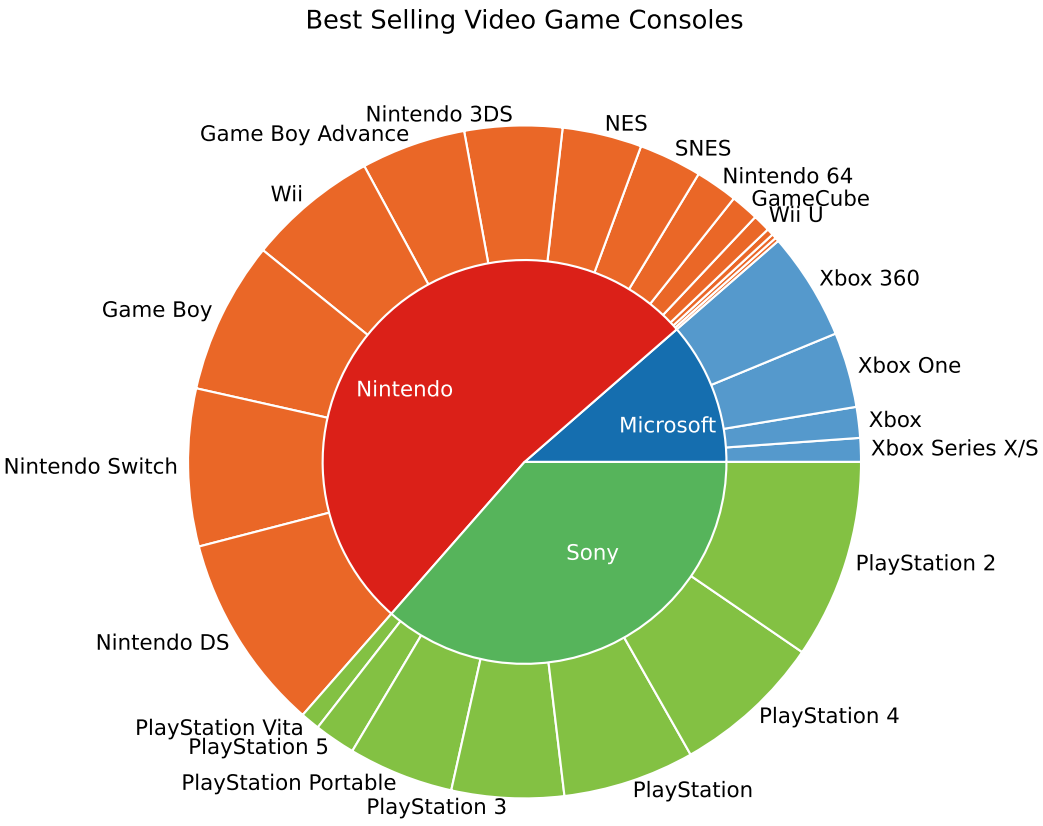


The line plot should be made using Seaborn's `lineplot()` functionality. Do not use Matplotlib's `fill_between()`.

Include the axes labels, plot title, and legend in the lower left corner. The average male heights should be blue and average female heights red. Each athlete should only be represented once in a given year. The width of the shading should be equal to 1 standard deviation on either side of the line.

Question 3: (40 pts)

The figure below is a sunburst plot showing the total number of video game consoles sold by Nintendo, Microsoft and Sony. The inner ring shows the total number of consoles sold per company, and the outer ring refines each company into their individual consoles.



Recreate the above sunburst plot using Matplotlib (hint: use nested pie charts). The arc length of each wedge should be proportional to the number of units sold. Importantly, the outer ring should be aligned with the inner ring. For example, all of the Nintendo consoles on the outer ring should be within the arc spanned by Nintendo.

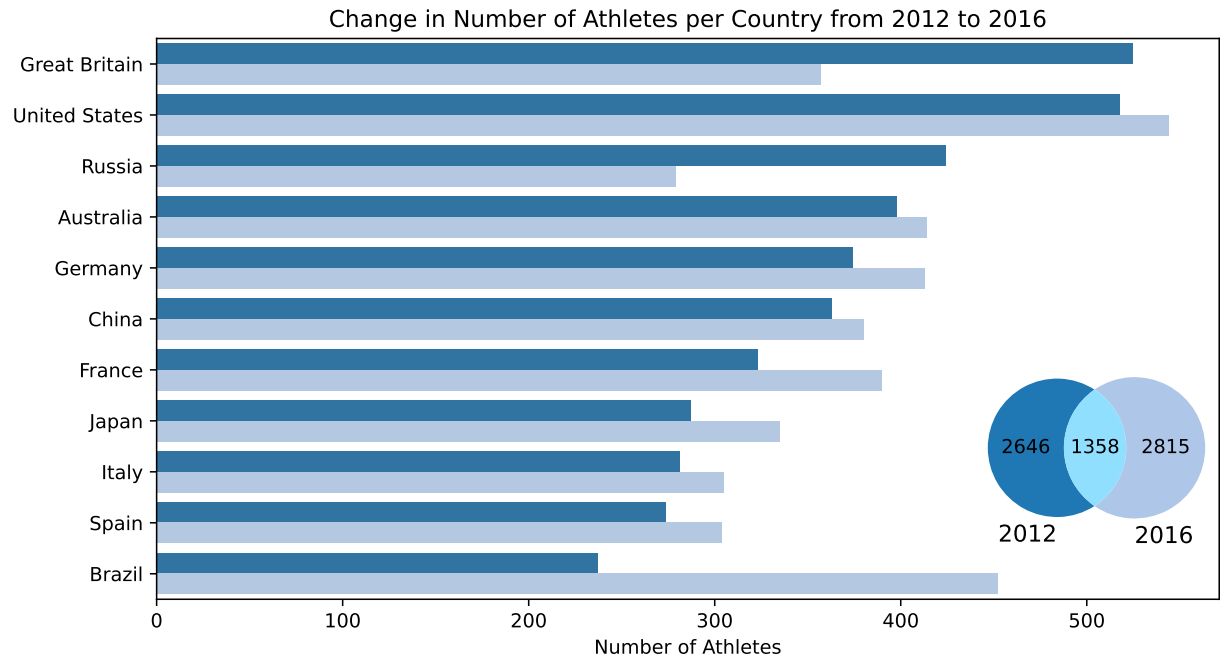
The consoles on the outer ring should be ordered from largest number of units sold to smallest number of units sold in a clockwise fashion. The three consoles with the smallest number of units sold for Nintendo are not displayed (since that would create visual clutter).

The colours of the inner ring are [#156EAF, #DB2018, #56B45B] for Microsoft, Nintendo and Sony, and [#5599CC, #EA6727, #83C143] for the outer ring.

The edge colours of the wedges should be white. The labels of the inner ring should be white in colour and on the interior, and the labels of the outer ring on the exterior.

6934 Students Only – Question 4: (40 pts)

The bar plot below shows the number of athletes that competed in the 2012 (dark blue) and 2016 (light blue) Olympics from a set of different countries. The Venn diagram acts as the legend – it informs the meaning of each colour. Additionally, it shows the overlap on the number of athletes (from these countries) that competed in both Olympics, and how many competed in just one of these Olympics.



Recreate the above bar plot using Seaborn, Matplotlib and the Matplotlib-venn libraries. The colours are given by the 'tab20' colour palette. Make sure to include the axes labels, plot title, legend annotations, and that the countries are listed in descending order based on number of athletes in the 2012 Olympics.

Submission

Submit your Jupyter notebook (`.ipynb`) through Brightspace. Late submissions will be subject to a 10% penalty for each hour past the deadline.

Attribution

Submissions should include an attribution section indicating any sources of material, ideas or contribution of others to the submission.

Submissions must represent your independent work.

You are encouraged to use any resources to help with your solution, but your solution must represent independent work. If your submitted work includes unacknowledged collaboration, code materials, ideas or other elements that are not your original work, it may be considered plagiarism or some other form of cheating under MUN general regulations 6.12.4.2 (4.12.4.2 for graduate students) and academic penalties will be applied accordingly.

Avoid academic penalties by properly attributing any contribution to your submission by others, including internet sources and classmates. This will also help distinguish what elements of the submission are original. You may not receive full credit if your original elements are insufficient, but you can avoid penalties for plagiarism or copying if you acknowledge your sources.

Github

I encourage you to store and version your work on GitHub. It is good practice to do so as everyone uses git in the real world.

However, **it is a requirement that git repositories containing assignment material be private.** University regulations (undergraduate 6.12.4.2 and graduate 4.12.4.2) consider it cheating if you allow your work to be copied. There will be zero tolerance for this.