COMP 4304 / 6934 – Assignment 3 (9%)

Due: 11:59pm, Feb. 24, 2023

Learning Objectives

You will gain experience using pandas and Matplotlib in this assignment.

Instructions

Download the following data set from Brightspace for this assignment:

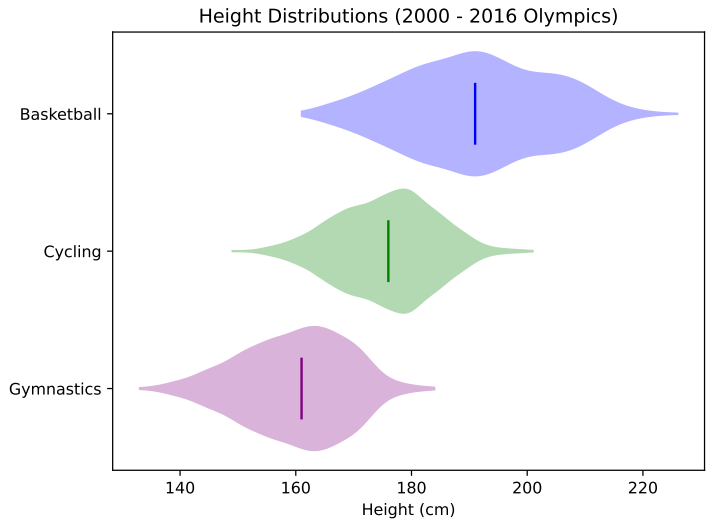
- `olympic_athletes.csv`

This data set contains information about all the athletes that have participated in the Olympics up until the 2016 games.

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

Question 1: (35 pts)

The following violin plot shows the height distributions of athletes in the Gymnastics, Cycling and Basketball sports at the Olympic games between 2000 to 2016 (inclusive). The line inside each violin shows the location of the median value.

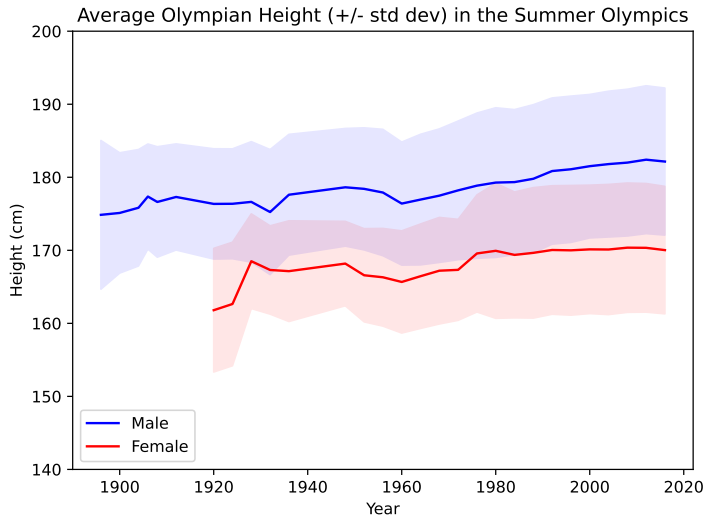


Recreate the above violin plot using Matplotlib. The three violins should be purple, green and blue in colour, including their filled colour and the colour of the median lines (Hint: consider the return values of the violin plot method). Include the x-axis and y-axis labels, as well as the plot title.

Since some athletes may compete in multiple events or multiple Olympics, make sure to include only one instance of an athlete (no duplicates).

Question 2: (35 pts)

This line plot shows the average height of male (blue) and female (red) Olympians over the history of the Summer Olympics. The shaded region represents the standard deviation about the average.

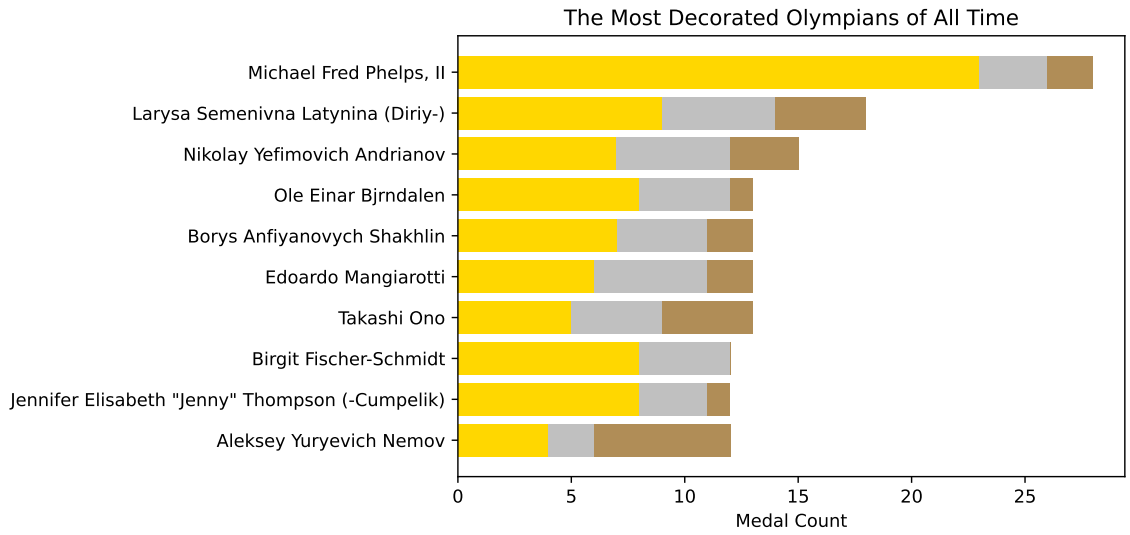


Recreate the above line plot using Matplotlib. Include the axes labels, plot title, and legend in the lower left corner. The average male heights should be blue and average female heights red. Each athlete should only be represented once in a given year.

Each line should include shading that has width equal to twice the standard deviation, that is, the top of the shaded region is the average value plus the standard deviation, and the bottom of the shaded region is the average value minus the standard deviation (hint: try `fill_between()`). The shading uses a transparency value of 0.1.

6934 Students Only – Question 3: (35 pts)

The bar plot below shows the top 10 most decorated Olympians. That is, the Olympic athletes that have won the highest total number of medals. Each bar is colour coded according to the number of gold, silver and bronze medals won by that athlete. In this way, the total number of medals won is represented, along with the breakdown per medal type.



Recreate the above bar plot using Matplotlib. The colours for the medal types are #FFD700, #C0C0C0, and #B08D57. Make sure to include the axes labels and plot title. The bars should be ordered as given in the figure.

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.