Group 3: Project One Proposal

1) What is Your Dataset and Why?
120 Years Olympic History: https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results
We chose this dataset because we wanted a dataset that would be fun to breakdown but also one that had enough substance to do an interesting analysis and pull meaningful results.

2) Three Research Guiding Questions

- Which city has hosted the most Olympics? Does hosting the Olympics impact total medals won for home country?

- What is the age distribution for summer and winter Olympics? Which sport has the highest average age?

- Does height or weight impact which sport someone is most likely to compete in? Is the distribution of height and weight different for summer and Winter Olympics?

3) Inspiration - other code/articles similar to your dataset
https://www.kaggle.com/code/flyingmonkey/10-questions-about-olympics
https://www.kaggle.com/code/joshuaswords/does-hosting-the-olympics-improve-performance
https://www.kaggle.com/code/fangya/summer-olympic-medal
https://www.kaggle.com/code/heesoo37/olympic-history-data-a-thorough-analysis

4) Example Visualization?

- Bar chart showing how many times each city has hosted the Olympics
- Bar chart showing medals won by country the year they hosted, as percentage of total medals available in the given year
- Violin plot of age ratio for summer Olympics
- Violin plot of age for winter Olympics
- Scatter plot of height/weight ratio for summer
- Scatter plot of height/weight ratio for winter

5) what are you regressing?
Is there a correlation to age of participant and medals won?

6) color palette
Olympic Colors - blue, yellow, black, green and red

7) roles & responsibilities
Jessie - Question Two & PowerPoint Creation
Jorge - Question Three
Sabina - Question One & Analysis Write Up
Leon - Regression
Gavin - Regression