

Martha Wood Capstone I Proposal

Questions

How have women's sports in the olympics changed over time? Female athletes have spent years in the fight for equality with their male counterparts. Currently the US national soccer and hockey teams are battling their respective organizations for equal coverage and equal pay. The visibility of sports is at its highest during the olympics and there were no female sports in the olympics until 1900. This dataset begins in 1924 when the only female sport in the Winter Olympics was figure skating and continues until 2014. I want to explore the way female sports were added into the olympics and investigate the trends in the countries with greater number of female medalists.

- How does the percent of medals for female athletes compare to male athletes?
 - The totals by year will compare the number of female and male sports
- Is there one country that has more female medals than male medals?
- How does the percent of female medals in the US compare to other countries?
 - China, Russia, Iran, England, Netherlands
- Are there indicators that correspond to an increase in female medals for a country?
 - Check education levels, women in government, etc

Data

Winter Olympic Medals 1924 to 2014

- Columns: Year, Sport, Event, Country, Gender, Medal Rank, Medal, Name of Athlete or Team, Age of Athlete
 - Year, Medal Rank and Age of Athlete are numerical columns, others are categorical
 - The only NaNs are in Age of Athlete. These NaNs may correspond to team medals where a single age would not be appropriate.

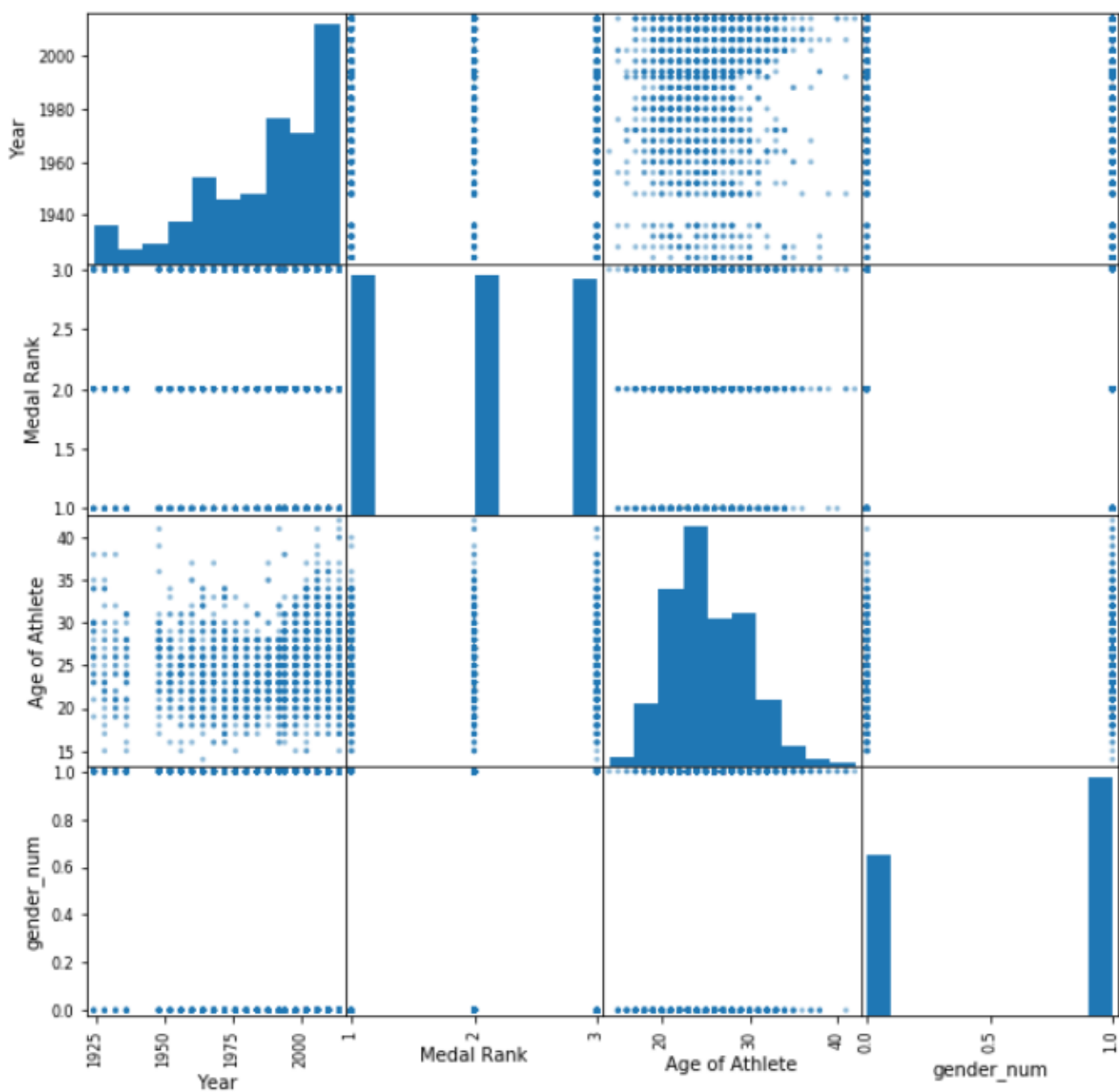
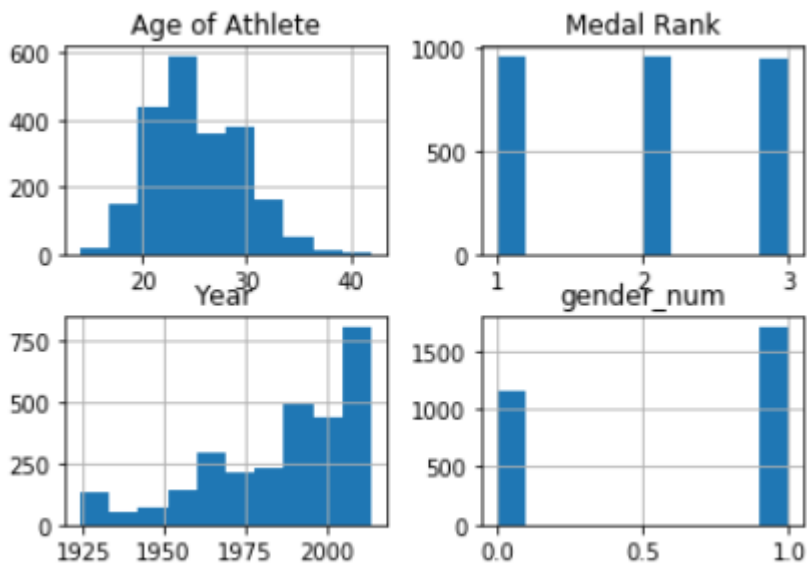
RangeIndex: 2865 entries, 0 to 2864

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	Year	2865 non-null	int64
1	Sport	2865 non-null	object
2	Event	2865 non-null	object
3	Country	2865 non-null	object
4	Gender	2865 non-null	object
5	Medal Rank	2865 non-null	int64
6	Medal	2865 non-null	object
7	Name of Athlete or Team	2865 non-null	object
8	Age of Athlete	2173 non-null	float64

dtypes: float64(1), int64(2), object(6)

- Graphs:



Other Datasets

- Summer Olympics (not on data.world, look for info in other locations)

- would building a dataset by scraping the results from Wikipedia be legal? acceptable?
- Human Development Indicators for selected countries

Minimum Viable Product

- My MVP is an evaluation of the addition of women's sports to the winter olympics over time in visual form. This will include at least three graphs: a graph of the increase in female sports over time as a percentage of total sports, a graph of the countries with the most total medals, with a breakdown of men's and women's medals, and a graph of the countries with the highest percent of female medals.
- My MVP+ would include compiling multiple HDI tables into one dataset and evaluating the provided statistics for a correlation with percentage of female medalists.
- My MVP++ is webscraping Wikipedia and developing a cleaning pipeline to create a dataset of all olympic medalists.