The Summer Olympic Games are an international athletics event held every four years and hosted in different countries around the world. Rowing was added to the Olympics in 1896 and has been in every Summer Olympics since. Rowing races in the Olympic context are typically regatta style, meaning that there are multiple boats racing head-to-head against each other in multiple lanes. Since 1912, the standard distance for Olympic regattas has been 2000m, but until then there had been a range in distances. The boat that is first to cross the finish line is awarded a gold medal, the second a silver medal, and the third a bronze. Over the course of its time as an Olympic sport there have been 25 different event entries. These events are separated by gender and range with the number of rowers in the boat (1, 2, 4, 6, 8, 17), the rigging (inrigged, outrigged), sculling, sweeping, and whether or not they are coxed. An inrigged shell means the riggers (where the oar is attached to the boat) are on the inside of the boat, outrigged shells mean the riggers are on the outside. Sculling is where the rowers have an oar on each side and sweeping is when each rower only has one oar on one side. The coxswain steers the boat and guides the rowers, some events have coxed boats whereas some others do not. Since the data spans so far back in time, the original dataset had some interesting issues in which some of the numbers were off for the number of medals awarded to athletes or nations. For example in the 1964 Rowing Men's Coxless 4 event, the US had 5 medals awarded instead of 4 because one of the rowers passed out after the first heat and the sub had to row in the final heat so both were given medals. Situations like this were not super common but they were frequent enough to through out the data. Using dyplyr the data has been adjusted to avoid these errors. The total\_medals and total\_points are counted as one medal towards each boat as opposed to each athlete in the boat. In looking at the total medals and total points for each nation, it is interesting to see which nations dominate in Olympic rowing. Additionally, looking at the overall distribution of the medals for all countries provides insight on just how lob-sided medaling can be in rowing at the Olympic level. This effect could likely be attributed towards how much funding nations are placing towards their rowing teams.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| NOC | National Olympic Committee or the nation competing. |
| total\_medals | The total number of medals for that country in that event. |
| total\_points | The total number of points for that country in that event. The points are scaled with a gold medal counting for 3 points, a silver for 2, and a bronze for 1. |
| total\_gold | The total number of gold medals for that country. |
| total\_silver | The total number of silver medals for that country. |
| total\_bronze | The total number of bronze medals for that country. |

1. Create a histogram of total\_points for all countries. Describe the distribution of total\_points. Is there a skew, if so, what is it and what does it mean?
2. Using <rowing_medals.csv>, obtain the summary statistics for total\_points and fill them in below.

**Minimum**: **Lower Quartile:**  **Median**: **Mean**:

**Upper Quartile**: **Maximum:**

1. Determine whether or not the total\_points for the UK is an outlier. Use calculations to justify your answer.
2. The UK has 154 total\_points in the dataset, determine whether or not that is an outlier. Use calculations to justify your answer.
3. Check to see if there are any other outliers.
4. Using your answers to questions 2-4 to create a boxplot of total\_points.
5. Create a barplot of NOC ranked by total\_medals to showcase the distribution of medals per nation. Based on this plot would it be fair to assume that economic welfare is a confounding variable in Olympic rowing medals? Explain why or why not.
6. What could be a reason for the distribution of medals and points being so heavily skewed towards certain nations winning more than others?
7. There is a lot of debate about how to best weigh the points for the different types of medals. Read [this link](https://www.topendsports.com/events/summer/medal-tally/rankings-weighted.htm) about different medal point weighing and decide on a method you think would be best. How would the new method alter the data?
8. Using the total\_gold, total\_silver, and total\_bronze variables, create a new total\_points variable based on your idea for medal scaling. Create a histogram of the total\_points with this new variable and comment on the skew.