1. A graph of a bar graph

   Description automatically generated 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?

The data is right-skewed meaning that the values are clustered around the smaller values for total\_points, so it is more common for nations to do poorly in Olympic rowing than well.

R KEY:

medals\_hist <- medals\_df |> filter(total\_points >= 5) |> arrange(desc(total\_points))

ggplot(medals\_df, aes(x = total\_points)) + geom\_histogram(fill = "lightblue", colour = "black", bins = 20) + theme\_minimal() + labs(title = "Histogram of total\_points for all countries in all events")

1. Using <rowing_medals.csv>, obtain the summary statistics for total\_points and fill them in below.

**Minimum**: 1.00 **Lower Quartile:** 6.00 **Median**: 15.00 **Mean**: 38.15

**Upper Quartile**: 57.00 **Maximum:** 187.00

R KEY:

summary(medals\_df$total\_points)

1. Determine whether or not the total\_points for the UK is an outlier. Use calculations to justify your answer.

An outlier

IQR = Q3-Q1 = 57-6 = 51

1.5\*IQR = 1.5\*51 = 76.5

Q3 +1.5\* IQR = 57 + 76.5 = 133.5

154 > 133.5

R KEY:

medals\_df |> filter(NOC == "GBR")

1. 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.

A graph of a bar chart

Description automatically generated

Based on this graph it would not be fair to assume that economic welfare is a confounding variable in predicting nations’ performance in Olympic rowing. If there was a GDP variable it may be easier to assess this.

R KEY:

```{r, fig.height= 7, fig.width=2}

sum\_medals <- medals\_df |> group\_by(NOC) |> summarise(total\_medals = sum(total\_medals))

medals\_reordered <- sum\_medals |> mutate(NOC = fct\_reorder(NOC, total\_medals))

ggplot(data = medals\_reordered, aes(x = NOC, y = total\_medals)) + geom\_col(fill = "lightblue", colour = "grey") + coord\_flip() + theme\_minimal() + labs(title = "Barplot of NOC ranked by total\_medals")

```

1. What could be a reason for the distribution of medals and points being so heavily skewed towards certain nations winning more than others?

Rowing shells (boats) are expensive, meaning rowing is a sport that requires a great deal of funding. Some nations make funding rowing a priority while others do not. This can also ultimately come down to which countries are wealthier in general than others.