1. Create a histogram of *points* for all countries. Describe the distribution of points.

The data is right-skewed meaning that the values are clustered around the smaller values for points, so it is more common for nations to get fewer medals in Olympic rowing.

R KEY:

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

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

1. Obtain the summary statistics for *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$points)

1. Check to see if there are any outliers for *points*.

GBR is 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

USA is an outlier.

187>133.5

131<133

121<133

87<133

```{r}

medals\_df |> arrange(desc(points)) |> head()

```

1. Create a barplot of *NOC* ranked by *medals* to showcase the distribution of *medals* per nation. Describe any noticeable patterns in the plot.

The country with leading medals is the USA followed by the UK and Germany, all of which are fairly large nations. The countries with the least medals seem to be smaller.

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

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

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

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

```

1. Why does it seem that some nations seem to win more medals than others?

Economics could certainly be a factor in medaling as rowing programs are expensive to fund due to the price of boats. Additionally, countries’ access to rowable water bodies could come into play. Also the population size of countries could certainly matter as rowing programs need a lot of people.

1. 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) (<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 (or make up your own weighting system). How would the new method alter the data?

Open ended, no right or wrong answer, credit for answering.

1. Using the gold, silver, and bronze variables, create a new *points* variable with a different name based on your idea for medal scaling. Create a histogram of the points with this new variable and comment on the skew.

Open ended, no right or wrong answer, credit for answering, so long as the data makes sense and there are no errors.