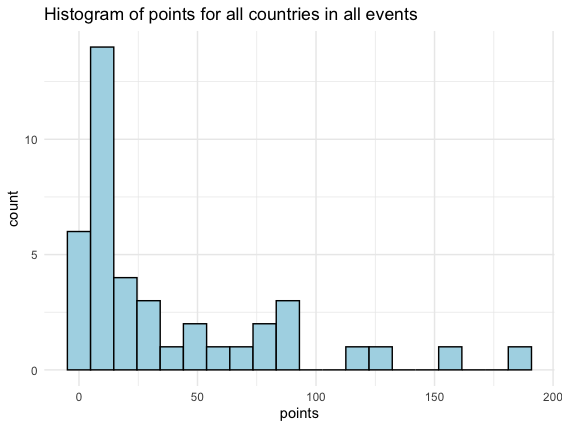
1. Create a histogram of points for all countries. Describe the distribution of points.

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. Using <rowing_medals.csv>, 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.

A table of numbers with text

Description automatically generatedUK 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

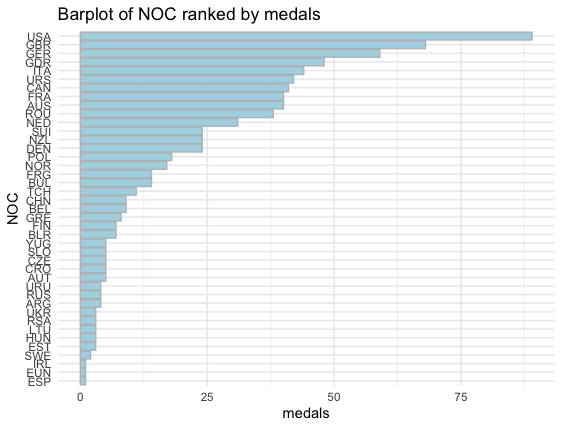
121<133

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

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, 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. 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. 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 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.