Oscar Beresford 201722873

Task One

The researcher first started by loading the data into R using the built in *'Import Dataset'* button – Assigning the dataset the `buffalo` variable. The data was formatted with 2 columns labelled *'year'* and *'snowfall'*. The first recorded year was `1910` and the last recorded year was `2018`. The data for the snowfall was recorded in inches with a minimum value of `25.0` and a maximum value of `199.4` giving a range of `177.4`.

The minimum and maximum values for the snowfall in inches were assigned the variables `min_value` and `max_value` respectively.

```
min_value <- min(buffalo$snowfall)

max_value <- max(buffalo$snowfall)

min_value

max_value
```

output

```
25

199.4
```

Task Two

This task was split into 6 sub-tasks labelled PartA - PartF.

Part A

The researcher initially identified the index corresponding to the minimum amount of snowfall using the `which()` function. Storing said calculated index in `index_year_min` then using this index to relate to the corresponding value for the *'year'*.

```
index_year_min <- which.min(buffalo$snowfall)

min_year <- buffalo$year[index_year_min]

min_year
```

output

```
1919
```

Part B

The researcher took a similar approach, note the change of variable and the use of the `max()` function

```
index_year_max <- which.max(buffalo$snowfall)

max_year <- buffalo$year[index_year_max]

max_year
```

output

```
1977
```

Part C

The average and the statistical mean are synonymous with one another, thus the built-in `mean()` function is used to calculate the average snowfall.

```
average_snow <- mean(buffalo$snowfall)

average_snow
```

output

```
86.69174
```

Part D

The `sd()` function is built into R and outputs the standard deviation. Taking in the *'snowfall'* column as an input.

```
standard_dev_snow <- sd(buffalo$snowfall)

standard_dev_snow
```

output

```
28.23302
```

Part E

Here the researcher used the built in `cm()` function which converts the previously calculated standard deviation from inches to cm

```
cm(standard_dev_snow)
```

output

```
71.71188
```

Part F

The researcher multiplied 3 by 39.3701 to get 118.11 (M to Inch conversion factor) storing the equivalent value in `amount_of_snow`. Then they used the > operator to make a direct comparison between each row in the 'snowfall' column and `amount_of_snow`.

The > operator in R outputs a Boolean value of 1 or 0 for if the statement is true or false respectively. Hence `buffalo$snowfall > amount_of_snow` will output a 1 when this statement is true and 0 when it's false. The `sum()` function then sums up the 1's and 0's to give the answer.

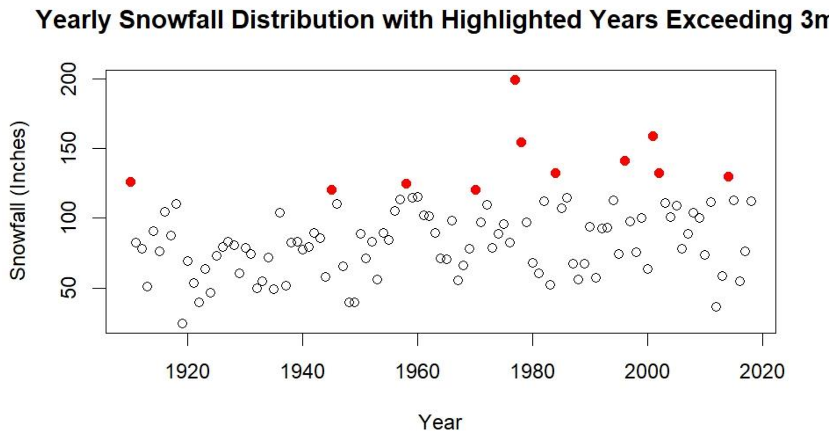
```
amount_of_snow <- 118.11

num_of_years <- sum(buffalo$snowfall > amount_of_snow)

num_of_years
```

output

```
11
```

Task Three

The graph below depicts years in which snowfall exceeded 3 meters, highlighted in red. A total of 11 occurrences are represented by the 11 red data points on the histogram. Notably, the plot reveals a significant peak in snowfall during 1977, as evidenced by the red data point reaching 199.4.




```
plot(buffalo$year, buffalo$snowfall, xlab="Year", ylab ="Snowfall (Inches)", main=”Yearly Snowfall Distribution with Highlighted Years Exceeding 3m”)

years_greater_than_3m <- buffalo$snowfall > 118.11

points(buffalo$year[years_greater_than_3m], buffalo$snowfall[years_greater_than_3m], col = "red", pch=19)
```

Task Four

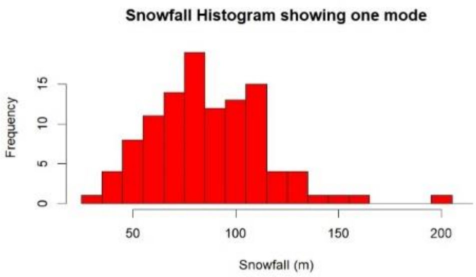
For this task, an alternative method was used to determine the breakpoints in the histogram. Instead of specifying an integer value for the number of breaks, a vector was utilized.

A sequence of numbers was generated using the `seq()` function, starting from 25 and finishing at 215, with incremental adjustments controlled by the `by()` function. These numbers corresponded to the positions of the breaks in the histogram, influencing the number of bins. Subsequently, three histograms were constructed, each exhibiting a different number of modes.

Histogram with one mode

The `col=”red”` describes the colour of the histogram and the labels relate to the x and y variables. The researcher used 20 breaks in this histogram. The histogram shows a single modal peak.

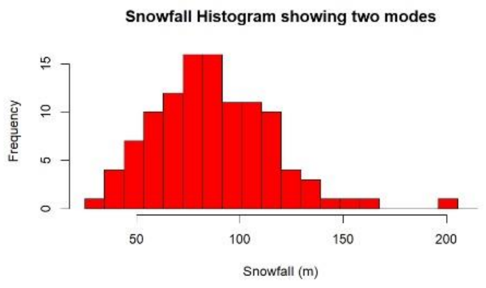
```
break_vector = seq(25,215,by=10)

hist(buffalo$snowfall, breaks = break_vector, col = "red", xlab = "Snowfall (m)", ylab = "Frequency", main = "Snowfall Histogram showing one mode")
```



Histogram with two modes

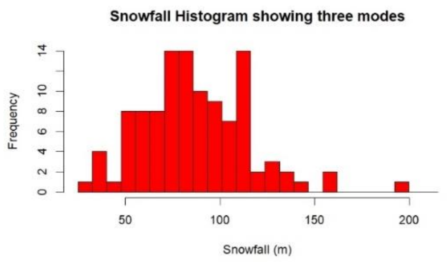
By changing the value of `by=` to 9.5 the researcher was able to get 21 breaks which led to two centralised modal peaks.

```
break_vector = seq(25,215,by=9.5)
```

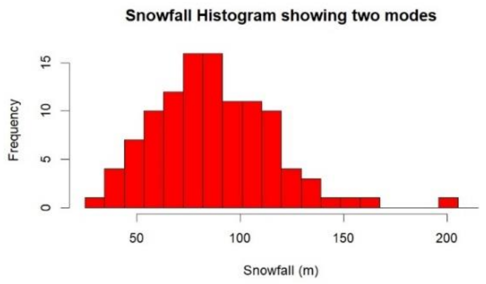


Histogram with three modes

Again by changing `by=9.5` to `by=7.60` to get 26 breaks which led to three modal peaks



Conclusion



After reviewing and analysing the three previous histograms the researcher believes there to be **two modes** in the `buffalo` data-set. Looking at the two mode histogram the two peaks are larger in comparison to the surrounding peaks while dominating the centre of the distribution. The peaks of the histogram are also in a pyramid-like structure which is a further indicator of a strong mode.

Academic integrity statement

You must sign this Oscar Beresford and include it with each piece of work you submit.

I am aware that the University defines plagiarism as presenting someone else's work, in whole or in part, as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work, in whole or in part, as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have not resubmitted my own work or part thereof without specific written permission to do so from the University staff concerned when any of this work has been or is being submitted for marks or credits even if in a different module or for a different qualification or completed prior to entry to the University. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the University's policy on mitigation and the School's procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Name	Oscar William Beresford
Student ID	201722873