HW11

Chriss Jordan Oboa

2022-11-08

Loading Library

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v tibble  3.1.8     v purrr   0.3.5
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggdist)
library(ggridges)
```

```
##
## Attaching package: 'ggridges'
##
## The following objects are masked from 'package:ggdist':
##
##     scale_point_color_continuous, scale_point_color_discrete,
##     scale_point_colour_continuous, scale_point_colour_discrete,
##     scale_point_fill_continuous, scale_point_fill_discrete,
##     scale_point_size_continuous
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:purrr':
```

```
##
##     some
##
## The following object is masked from 'package:dplyr':
##
##     recode
```

Loading the data needed for this lab.

```
data("faithful")
head(faithful)
```

```
##   eruptions waiting
## 1     3.600      79
## 2     1.800      54
## 3     3.333      74
## 4     2.283      62
## 5     4.533      85
## 6     2.883      55
```

1. Find the frequency distribution of the eruption waiting periods in faithful.

In order to find the frequency distribution of the eruption waiting periods in faithful, we will need first to find the range of waiting durations with the range function, then put the range into non-overlapping sub-intervals by defining a sequence of equal distance break points. After that, we will need to classify the waiting durations according to the half-unit-length sub-intervals with cut, and then compute the frequency of waiting eruptions in each sub-interval with the table function.

The result decribes the number of time waiting eruptions happen in the time range of [40,50] [50,60], [60,70], [70,80], [80,90], and [90,100].

```
#find the range of waiting durations with the range function
duration = faithful$waiting
range(duration)
```

```
## [1] 43 96
```

```
#the range into non-overlapping sub-intervals by defining a sequence of equal distance break points
breaks = seq(40, 100, by=10)    # half-integer sequence
breaks
```

```
## [1]  40  50  60  70  80  90 100
```

```
#Classify the waiting durations according to the half-unit-length sub-intervals with cut.
duration.cut = cut(duration, breaks, right=FALSE)

#Compute the frequency of waiting eruptions in each sub-interval with the table function.
duration.freq = table(duration.cut)

#The frequency distribution of the waiting eruption duration is:
duration.freq
```

```
## duration.cut
##  [40,50)  [50,60)  [60,70)  [70,80)  [80,90) [90,100)
##       21       56       26       77       80       12
```

2. Find programmatically the duration sub-interval that has the most eruptions.

In order to find programmatically the duration sub-interval that has the most eruptions, we will need first to find the range of durations with the range function, then put the range into non-overlapping sub-intervals by defining a sequence of equal distance break points. After that, we will need to classify the durations according to the half-unit-length sub-intervals with cut, and then compute the frequency of eruptions in each sub-interval with the table function.

To finish,we will use which.max to return the position of the element with the maximal value in a vector.

The result shows the time range where the number of time waiting eruptions happen the most.

```r
#find the range of eruption durations with the range function
duration2 = faithful$eruptions
range(duration2)
```

```
## [1] 1.6 5.1
```

```r
#the range into non-overlapping sub-intervals by defining a sequence of equal distance break points
breaks2 = seq(1.5, 5.5, by=0.5)    # half-integer sequence
breaks2
```

```
## [1] 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5
```

```r
#Classify the eruptions durations according to the half-unit-length sub-intervals with cut.
duration2.cut = cut(duration2, breaks2, right=FALSE)

#Compute the frequency of eruptions in each sub-interval with the table function.
duration2.freq = table(duration2.cut)

#The frequency distribution of the eruption duration is:
duration2.freq
```

```
## duration2.cut
## [1.5,2) [2,2.5) [2.5,3) [3,3.5) [3.5,4) [4,4.5) [4.5,5) [5,5.5)
##      51      41       5       7      30      73      61       4
```

```r
#Highlighting the max range of value from duration.freq
fre <- which.max(duration2.freq)
duration2.freq[fre]
```

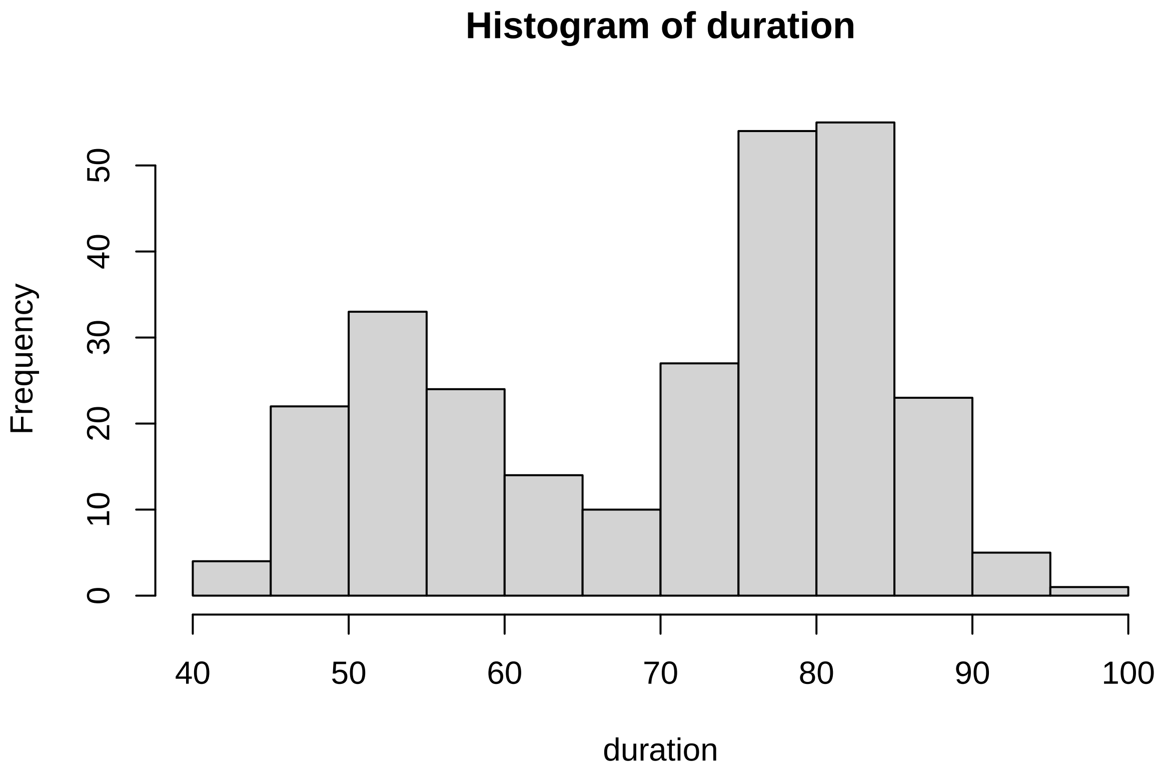
```
## [4,4.5)
##      73
```

3. Find the histogram of the eruption waiting period in faithful.

To find the histogram of the eruption waiting period in faithful, we apply the hist function to produce the histogram of the waiting eruptions variable from the dataset.

The result shows a histogram of based on duration and frequency from the variable waiting of the faithful dataset.

```r
duration = faithful$waiting
hist(duration)
```



4. Find the relative frequency distribution of the eruption waiting periods in faithful.

To find the relative frequency distribution of the eruption waiting periods in faithful, we need to to divide the duration frequency by the sample size.

The result gives us the relative frequency distribution of each eruption waiting periods range.

```
duration.relfreq = duration.freq / nrow(faithful)
duration.relfreq
```

```
## duration.cut
##    [40,50)    [50,60)    [60,70)    [70,80)    [80,90)   [90,100)
## 0.07720588 0.20588235 0.09558824 0.28308824 0.29411765 0.04411765
```

5. Find the cumulative frequency distribution of the eruption waiting periods in faithful.

To find the cumulative frequency distribution of the eruption waiting periods in faithful, we apply the cumsum function to compute the cumulative frequency distribution.

The result gives us the cumulative frequency distribution of each eruption waiting periods range or shows the frequency proportion of waiting eruptions whose durations are less than or equal to a given level.

```
duration.cumfreq = cumsum(duration.freq)
duration.cumfreq
```

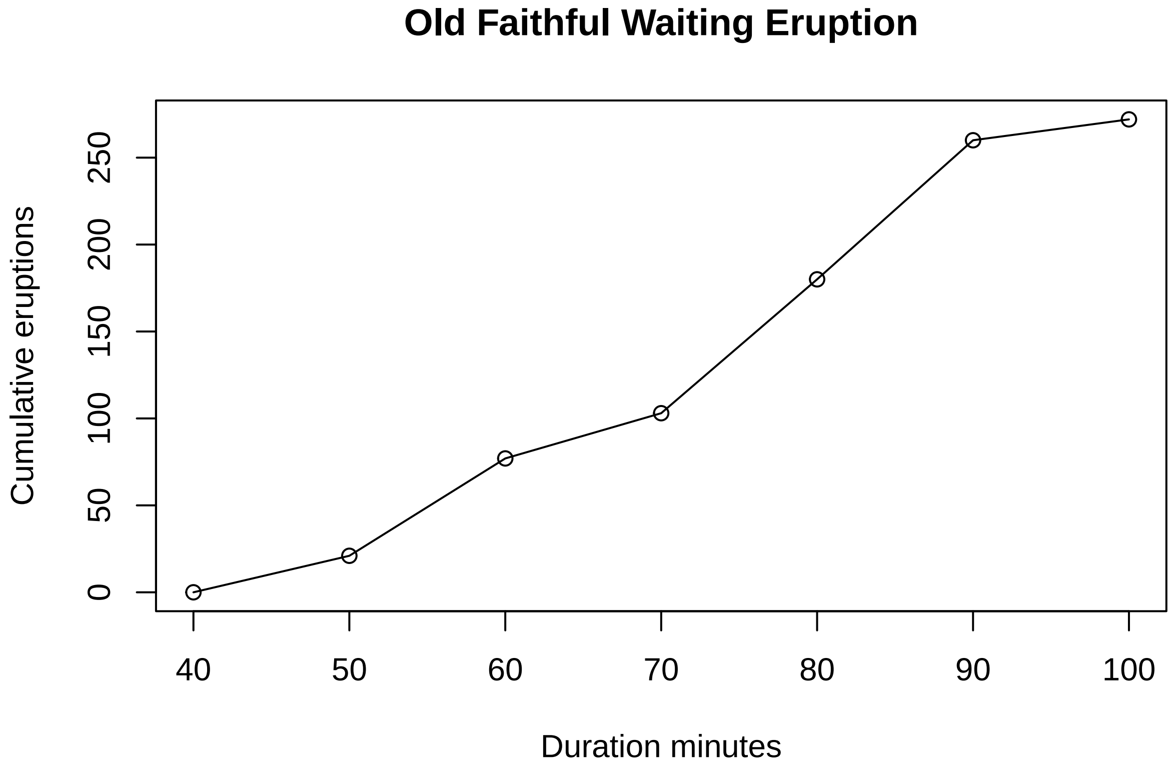
```
##  [40,50)  [50,60)  [60,70)  [70,80)  [80,90) [90,100)
##       21       77      103      180      260      272
```

6. Find the cumulative frequency graph of the eruption waiting periods in faithful.

To find the cumulative frequency graph of the eruption waiting periods in faithful, we compute its cumulative frequency with cumsum, add a starting zero element, and plot the graph. We use breaks as x variable which represent the duration in minutes, and cumfreq0 as y cumulative eruptions.

The result portrays the cumulative frequency graph of the eruption durations or shows the frequency proportion of waiting eruptions whose durations are less than or equal to a given level

```
cumfreq0 = c(0, cumsum(duration.freq))
plot(breaks, cumfreq0, main="Old Faithful Waiting Eruption", xlab="Duration minutes", ylab="Cumulative
lines(breaks, cumfreq0)
```



7. Find the stem-and-leaf plot of the eruption waiting periods in faithful.

To find the stem-and-leaf plot of the eruption waiting periods in faithful, we apply the stem function to compute the stem-and-leaf plot of eruptions.

The result identifies durations with the same two most significant digits, and queue them up in rows.

```
duration = faithful$waiting
stem(duration)
```

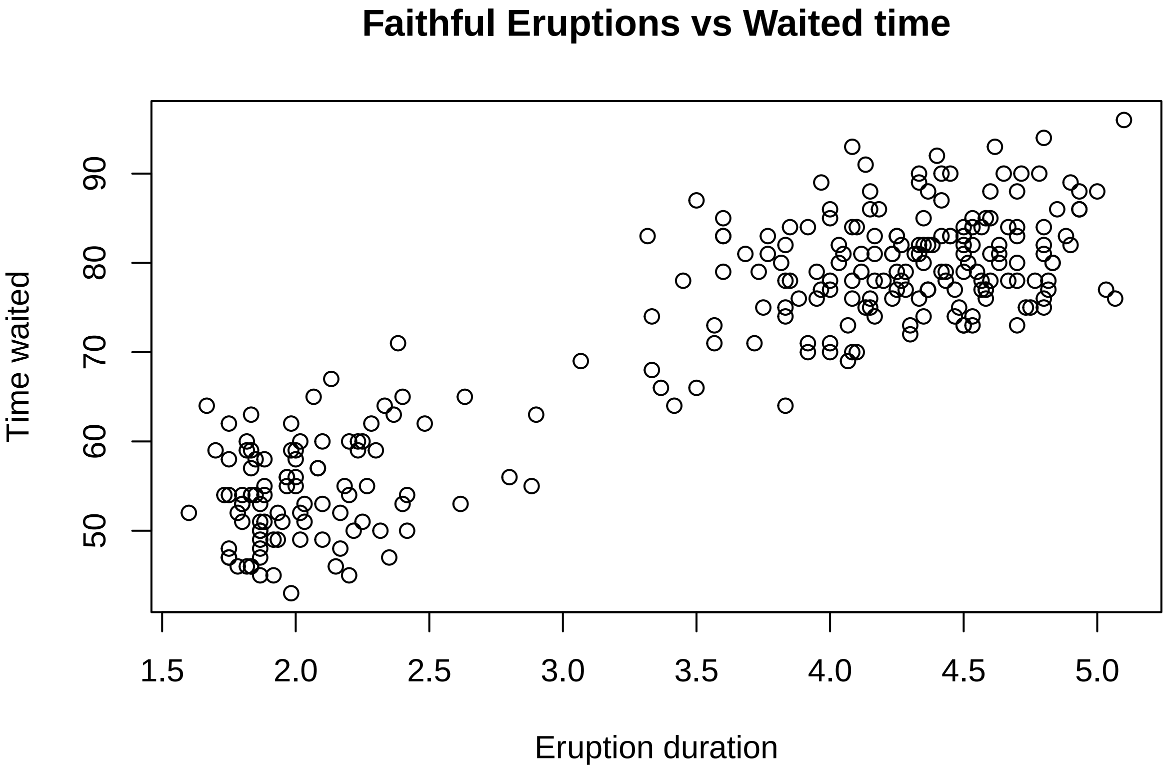
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 3
##   4 | 55566666777788899999
##   5 | 00000111111222223333333444444444
##   5 | 555555666677788889999999
##   6 | 00000022223334444
##   6 | 555667899
##   7 | 00001111123333333444444
##   7 | 555555556666666667777777777778888888888888889999999999
##   8 | 000000001111111111111222222222222333333333333334444444444
##   8 | 55555566666677888888999
##   9 | 00000012334
##   9 | 6
```

8. Find the scatter plot of the eruption durations and waiting intervals in faithful. Does it reveal any relationship between the variables?

To find the scatter plot of the eruption durations and waiting intervals in faithful, we apply the plot function to compute the scatter plot of eruptions and waiting.

This graph reveal a correlation between the eruption durations and waiting intervals. This correlation is a positive linear relationship between the two variables.

```
duration = faithful$eruptions
waiting = faithful$waiting
plot(duration, waiting,
main="Faithful Eruptions vs Waited time",
xlab="Eruption duration",
ylab="Time waited")
```



end