NAME : SRAVAN KUMAR VASAM

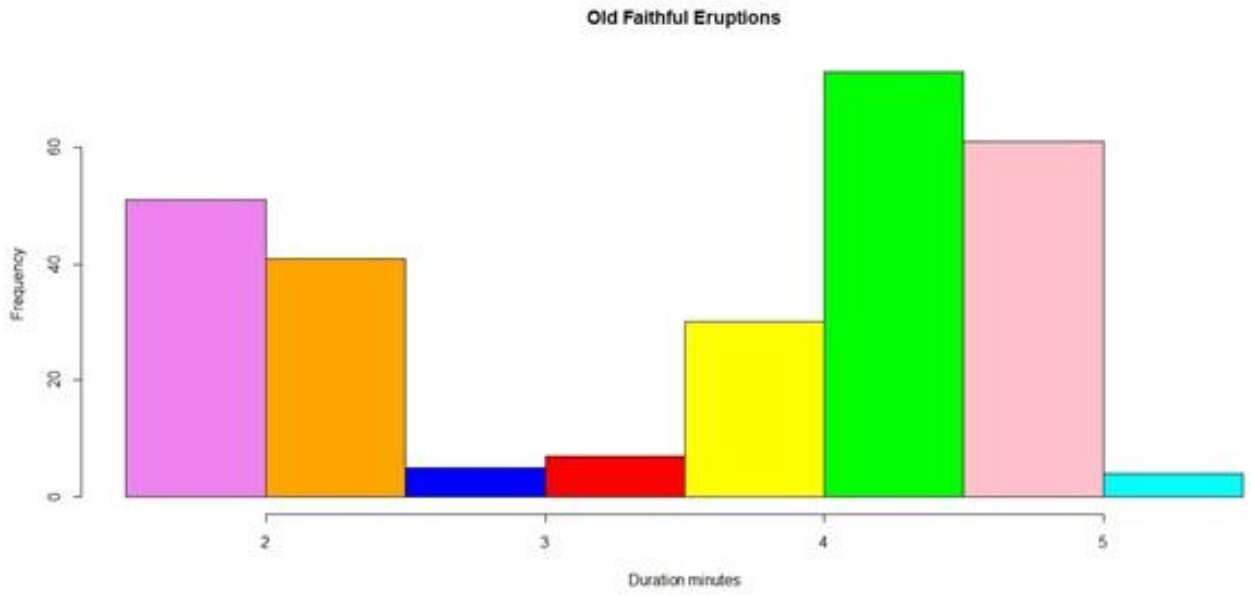
Statistical Analysis :

TITLE: QUANTITATIVE DATA ANALYSIS TO DISCOVER FREQUENCY, RELATIVE FREQUENCY & CUMULATIVE DISTRIBUTIONS

Data Analysis of Frequency & Relative Frequency Distribution, Cumulative Frequency & Cumulative Relative frequency distribution, Scatter plots of Quantitative Data set with visualization reports. (Histogram, graph, scatter plot)

Frequency Distribution of Quantitative Data

```
> duration = faithful$eruptions
> colors = c("red", "yellow", "green", "violet", "orange",
+   "blue", "pink", "cyan")
> hist(duration,     # apply the hist function
+   right=FALSE,     # intervals closed on the left
+   col=colors,      # set the color palette
+   main="Old Faithful Eruptions", # the main title
+   xlab="Duration minutes")       # x-axis label
```

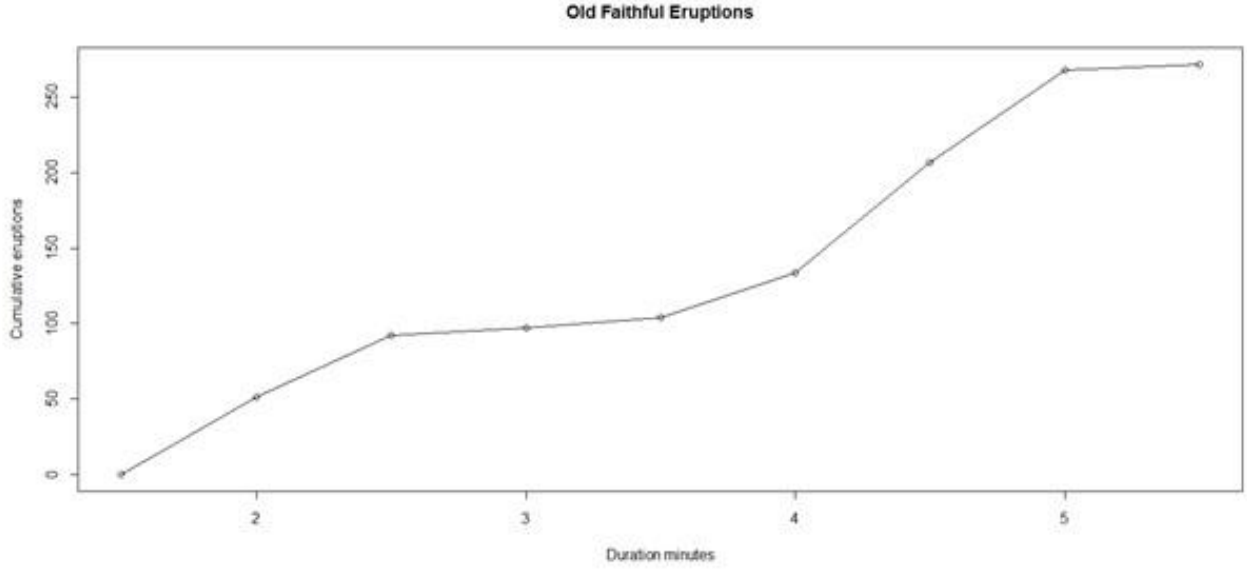


Relative Frequency Distribution of Quantitative Data

```
duration = faithful$eruptions

breaks = seq(1.5, 5.5, by=0.5)

duration.cut = cut(duration, breaks, right=FALSE)

duration.freq = table(duration.cut)

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

colors=c("red","blue","black","white","pink","purple","violet","orange")

pie(duration.relfreq,col=colors)
```

CUMULATIVE FREQUENCY GRAPH:

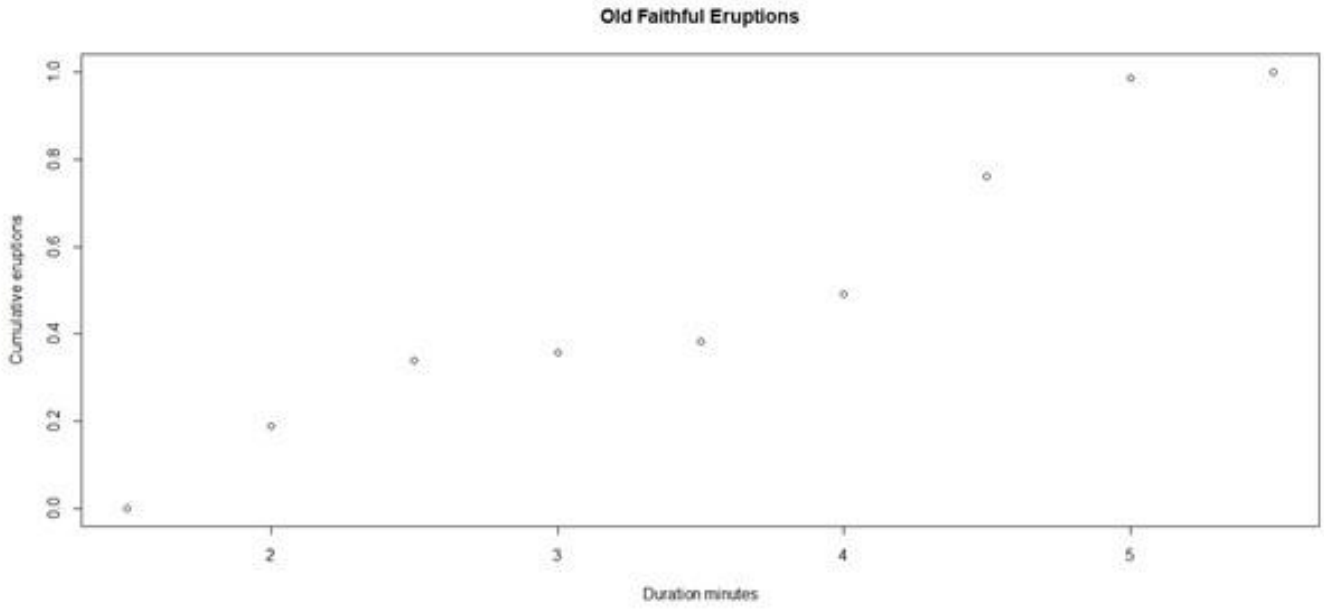
```
> duration = faithful$eruptions
> breaks = seq(1.5, 5.5, by=0.5)
> duration.cut = cut(duration, breaks, right=FALSE)
> duration.freq = table(duration.cut)

> cumfreq0 = c(0, cumsum(duration.freq))
> plot(breaks, cumfreq0,             # plot the data
+   main="Old Faithful Eruptions",   # main title
+   xlab="Duration minutes",         # x-axis label
+   ylab="Cumulative eruptions")    # y-axis label
> lines(breaks, cumfreq0)            # join the points
```



CUMULATIVE RELATIVE FREQUENCY GRAPH:

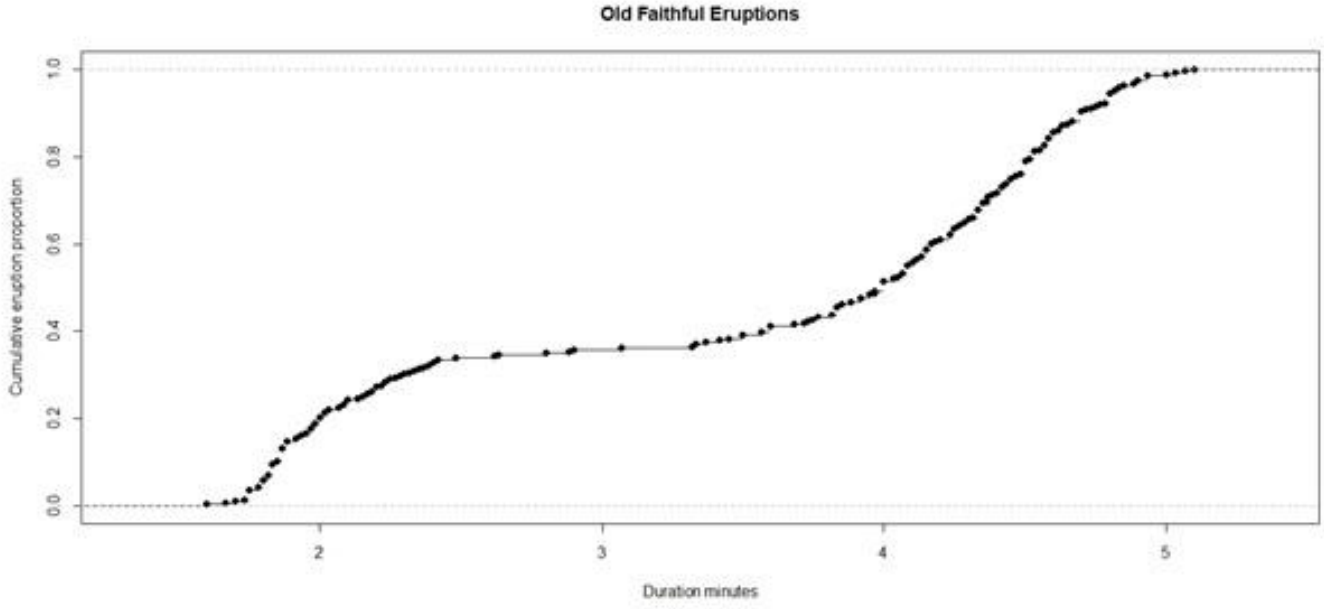
```
> duration = faithful$eruptions
> breaks = seq(1.5, 5.5, by=0.5)
> duration.cut = cut(duration, breaks, right=FALSE)
> duration.freq = table(duration.cut)
```

```
> duration.cumfreq = cumsum(duration.freq)
> duration.cumrelfreq = duration.cumfreq / nrow(faithful)

> cumrelfreq0 = c(0, duration.cumrelfreq)
> plot(breaks, cumrelfreq0,
+   main="Old Faithful Eruptions",  # main title
+   xlab="Duration minutes",
+   ylab="Cumulative eruption proportion")
> lines(breaks, cumrelfreq0)        # join the points
```



Interpolate function Fn with the built-in function ecdf

```
Fn = ecdf(duration)

plot(Fn,main="Old Faithful Eruptions",xlab="Duration minutes",ylab="Cumulative eruption proportion")
```



Scatter Plot

A **scatter plot** pairs up values of two quantitative variables in a data set and display them as geometric **points** inside a *Cartesian diagram*.

```
> duration = faithful$eruptions       # the eruption durations
> waiting = faithful$waiting          # the waiting interval
> plot(duration, waiting,             # plot the variables
+   xlab="Eruption duration",         # x-axis label
+   ylab="Time waited")               # y-axis label

abline(lm(waiting ~ duration))
```