Histograms.stem

Astudent

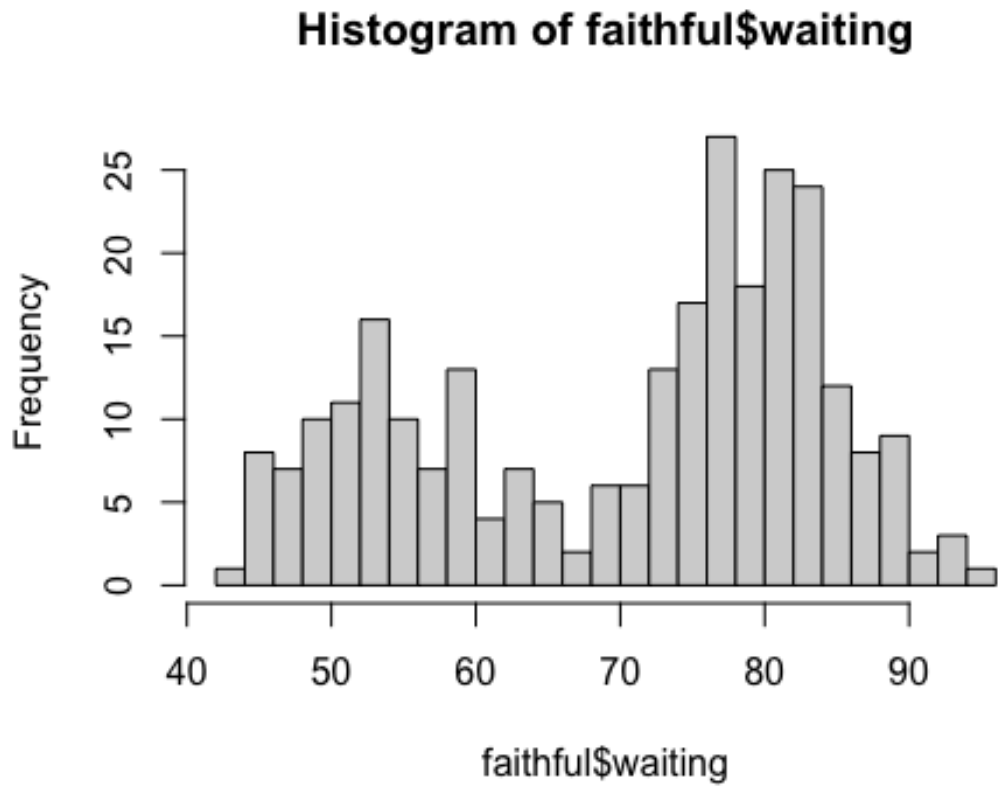
8/28/2023

D.Numerical data Some graphs to help us understand the data visually. For example, what is shape of the distribution of the data? Is it symmetric, is it skewed right or left, does it have one peak or two or barely any peak? Where is the center? What is the spread? Answer these questions for the graph below.

Histograms explained. The graph shows how many data values fall in each bin. For the graph below, the bins are [40,45], (45,50], (50,55], etc. Usually bins are right closed.

Adding more divisions will change what the graph looks like. Add breaks = 20 to the graph below

```
hist(faithful$waiting,breaks = 20)
```




```
#creates a histogram of the variable waiting of the built-in dataset faithful
```

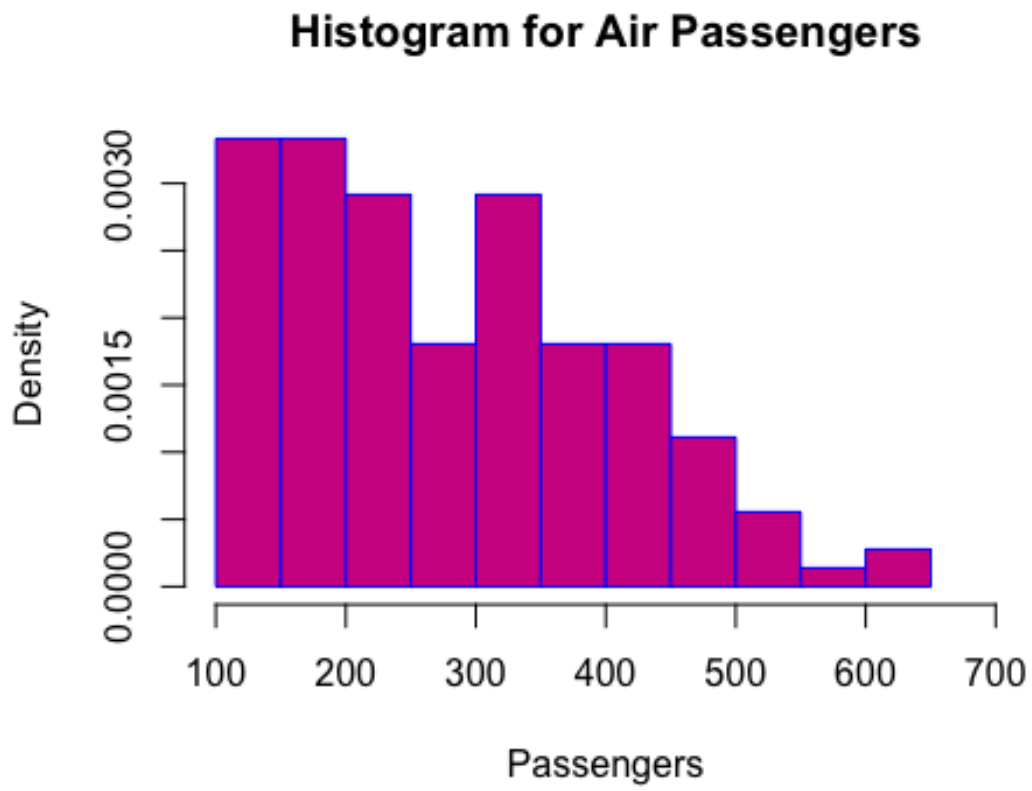
Make a histogram by hand for the data 2,3,4,1,6,7,10,20,1,5 with intervals [0,5], (5,10],(10,15],(15,20]

stem plot: Gives the numerical values in the columns. Run the code and then try dividing the variable by 10.

```
stem(faithful$waiting) #What is the difference in the information obtained
from a stem plot and that from a histogram? #What are a couple pieces of
information you can get from this stem plot? Why would you use one instead of
the other?
```

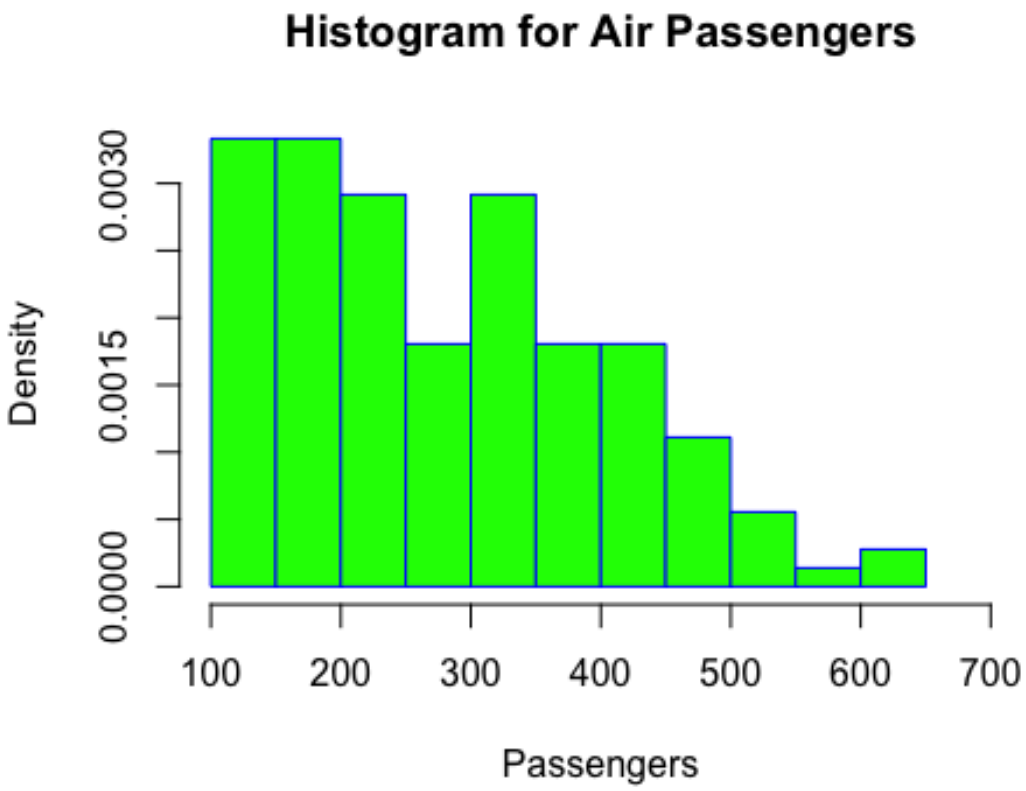
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 3
##   4 | 55566666777788899999
##   5 | 00000111111222223333333444444444
##   5 | 555555666677788889999999
##   6 | 00000022223334444
##   6 | 555667899
##   7 | 00001111123333333444444
##   7 | 55555555666666666777777777777888888888888889999999999
##   8 | 00000000111111111111222222222223333333333333344444444444
##   8 | 55555566666677888888999
##   9 | 00000012334
##   9 | 6
```

```
hist(AirPassengers, main="Histogram for Air Passengers", xlab="Passengers",
border="blue", col='violetred' ,  xlim=c(100,700),  breaks=10,  prob = TRUE)
```




```r
#monthly international airline passengers 1949-1960-built in dataset
#describe the shape
#note the labeling : the title and the x variable
#limiting the x axis
#coloring
#prob=T : The area inside the bar ( the height of the bar times its width) is
the proportion of observations in that interval.

u=hist(AirPassengers, main="Histogram for Air Passengers", xlab="Passengers",
border="blue", col='green' ,  xlim=c(100,700),  breaks=10,  prob = TRUE)
```




```
#explain what each of the parameters does-play around with changing them

str(u)

## List of 6
##  $ breaks  : int [1:12] 100 150 200 250 300 350 400 450 500 550 ...
##  $ counts  : int [1:11] 24 24 21 13 21 13 13 8 4 1 ...
##  $ density : num [1:11] 0.00333 0.00333 0.00292 0.00181 0.00292 ...
##  $ mids    : num [1:11] 125 175 225 275 325 375 425 475 525 575 ...
##  $ xname   : chr "AirPassengers"
##  $ equidist: logi TRUE
##  - attr(*, "class")= chr "histogram"

sum(u$density*50) # the sum of the areas of the bars=sum of the proportions
in each intervals =?

## [1] 1
```

Look at the dataset rivers.

```
str(rivers)

##  num [1:141] 735 320 325 392 524 ...
```

```r
help(rivers)

#stem(rivers)  #What is the length of the shortest river? How many rivers of
length 310 are there according to the plot?
#rivers<=319&rivers>=310
#sum(rivers<=319&rivers>=310)  #says how many rivers meet the condition in
parenthesis
#rivers[1]
rivers[rivers<=319&rivers>=310]
```

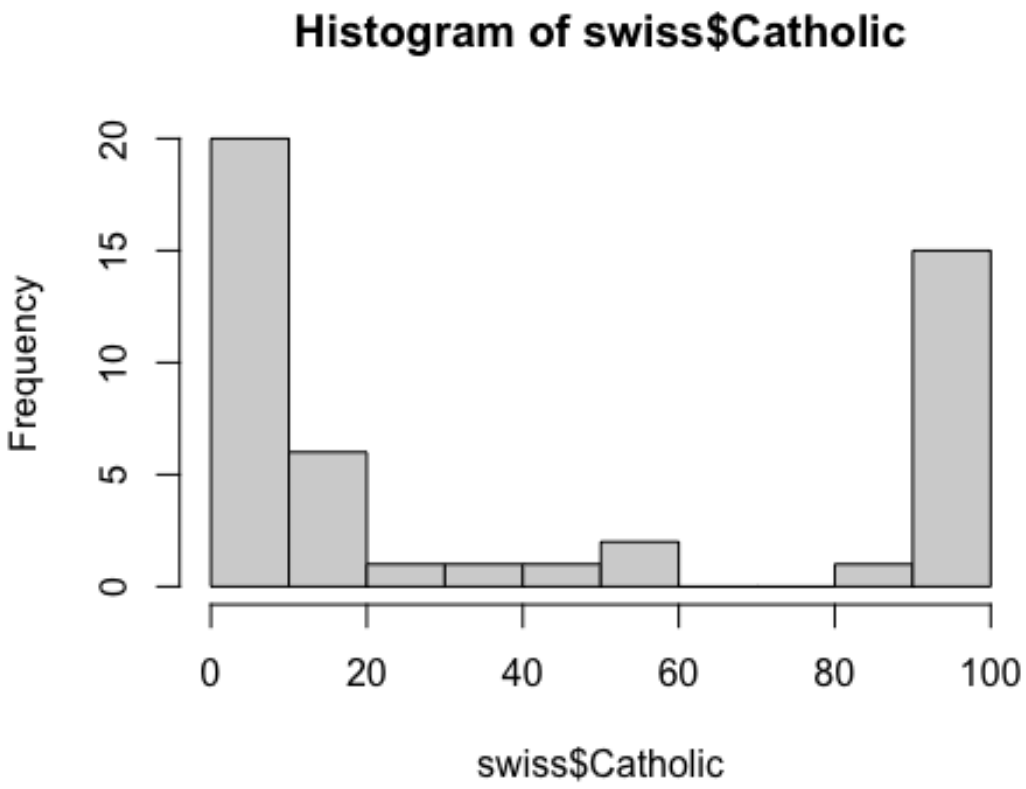
```
## [1] 315 310 314 310
```

```r
# see what variables are in the built-in dataset swiss
str(swiss)
```

```
## 'data.frame':    47 obs. of  6 variables:
##  $ Fertility       : num  80.2 83.1 92.5 85.8 76.9 76.1 83.8 92.4 82.4
82.9 ...
##  $ Agriculture     : num  17 45.1 39.7 36.5 43.5 35.3 70.2 67.8 53.3 45.2
...
##  $ Examination     : int  15 6 5 12 17 9 16 14 12 16 ...
##  $ Education       : int  12 9 5 7 15 7 7 8 7 13 ...
##  $ Catholic        : num  9.96 84.84 93.4 33.77 5.16 ...
##  $ Infant.Mortality: num  22.2 22.2 20.2 20.3 20.6 26.6 23.6 24.9 21 24.4
...
```

```r
#create a histogram for one of those variables
#describe its shape

#create a histogram for another one of those variables
#describe its shape
hist(swiss$Catholic)
```

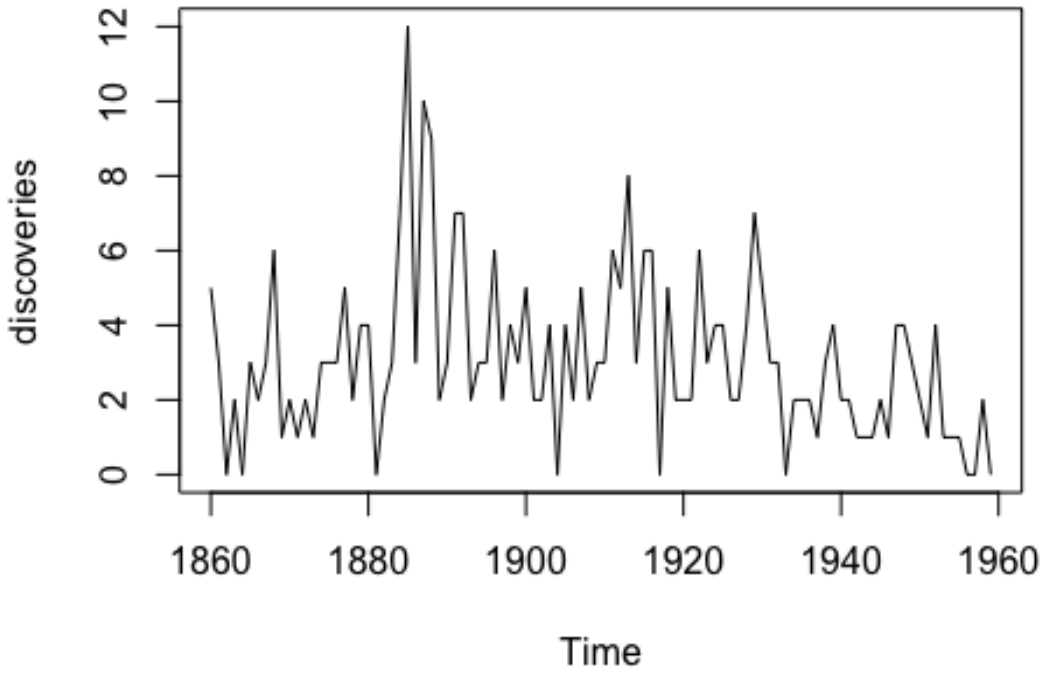


line/time plots -look for peaks, where increasing, decreasing

```
str(discoveries)
```

```
##  Time-Series [1:100] from 1860 to 1959: 5 3 0 2 0 3 2 3 6 1 ...
```

```
plot(discoveries) #What might explain some of the peaks and low points?
```




```
max(discoveries)
```

```
## [1] 12
```

```
which.max(discoveries) #What year?
```

```
## [1] 26
```

knit this file to a word document by going to the knit icon and selecting word. This would be the report you would turn in.Make sure have the rmarkdown library installed

summary: R code: hist, stem, plot, str hist: number of breaks, frequency or probability, labels