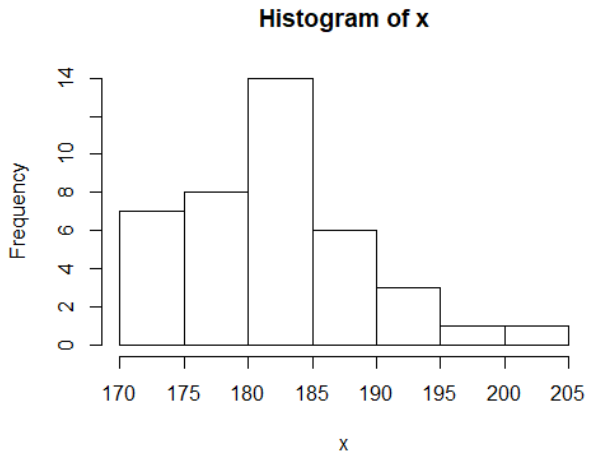
STA2023 R lab

Lab 1

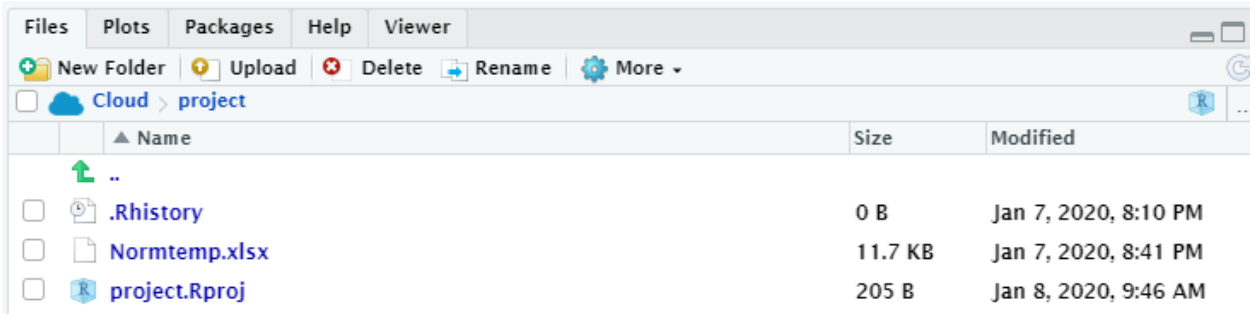
Topics: Introduction, Preview and Statistical Thinking. Types of Data: Define the terms data, statistics, parameter, sample, and population and identify them in context. Define quantitative and categorical variables and labels and identify them in context. Frequency Distributions: Construct and interpret frequency, relative frequency, and cumulative frequency distribution tables. Histograms: Construct and interpret histograms. Stem and Leaf Plots: Construct and interpret stem-and-leaf distributions.

```
> #ctrl+L, clear console. Be aware that any time you do this, it clears all
your entries.

> # For RStudio desktop (it doesn't apply to RStudio Cloud): Change the defau
lt working directory from RStudio menu under: Tools -> Global options ->
Browse to select the default working directory

x<-c(187,171,181,180,178,171,174,177,172,178,182,187,176,179,190,185,192,184,
182,178,187,173,185,184,184,183,185,197,202,181,181,191,178,187,185,186,174,1
74,182,195) # x, heights of 40 human males in centimeters. Copy and paste
into R
>#length(x): how many data values in x?
> # input the following data points as variable (object) grades: 75,71,82,89,
91,93,77,71,82,86,90
> grades<-c(75,71,82,89,91,93,77,71,82,86,90)
> hist(x)
```




```
> stem(x) # stem-and-leaf plot. It splits the "stems".

  The decimal point is 1 digit(s) to the right of the |

  17 | 1123444
  17 | 6788889
  18 | 01112223444
  18 | 555567777
  19 | 012
  19 | 57
  20 | 2
> stem(x, scale=0.5) # this is the one we will be using…
  The decimal point is 1 digit(s) to the right of the |

  17 | 11234446788889
  18 | 01112223444555567777
  19 | 01257
  20 | 2
```

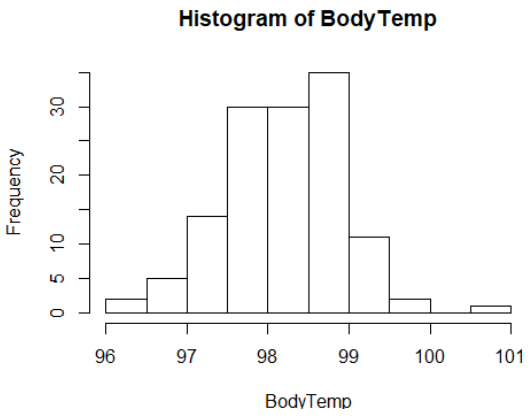
```
# generate the stem-leaf plot for grades.

# Normtemp dataset: from http://www.imathesis.com/sta2023T.html
>#  If you are on RStudio Cloud, after downloading the file, you need to
upload the file to the cloud: on right lower panel, click on Files, then
Upload, locate the file in your computer and upload it to the Cloud.
```



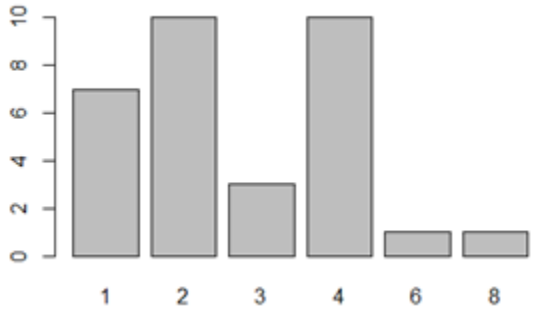

```
>#  how many data values in Normtemp? Find dim of Normtemp. Classify variables
continuous, discrete, quantitative, categorical.

> View(Normtemp)#open the excel file Normtemp in R Studio.
> attach(Normtemp)
> range(HeartRate)
[1] 57 89
> cc2<-cut(HeartRate, breaks = c(50,60,70,80,90), right=F)
> table10<-table(cc2)
> table10
cc2
[50,60) [60,70) [70,80) [80,90)
      4       32       66       28

> #Use Excel to construct a frequency table.
```

Rates	Count	cum	rel frq	cum frq
50-60	4	4	0.03	0.03
60-70	32	36	0.25	0.28
70-80	66	102	0.51	0.78
80-90	28	130	0.22	1.00

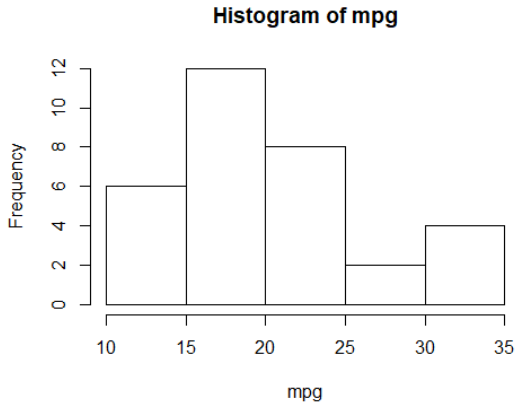
```
> with(Normtemp, hist(BodyTemp)) # create a hist of Bodytemp.
```



```
> dettach(Normtemp)
> data(mtcars) # loads mtcars dataset from R
> head(mtcars) # first six row of the dataframe
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

> # for a description of dataset mtcars: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

> # classify variables mpg, cyl, wt, vs, am as categorical (qualitative) or quantitative variables. Explain.

> attach(mtcars)
> dim(mtcars)# dimensions: rows x columns
[1] 32 11

> B<-with(mtcars, table (carb))
> B
carb
 1  2  3  4  6  8
 7 10  3 10  1  1

> # Barplots, histograms, stem-leaf plots
> barplot(B) # B is the table we just created
```

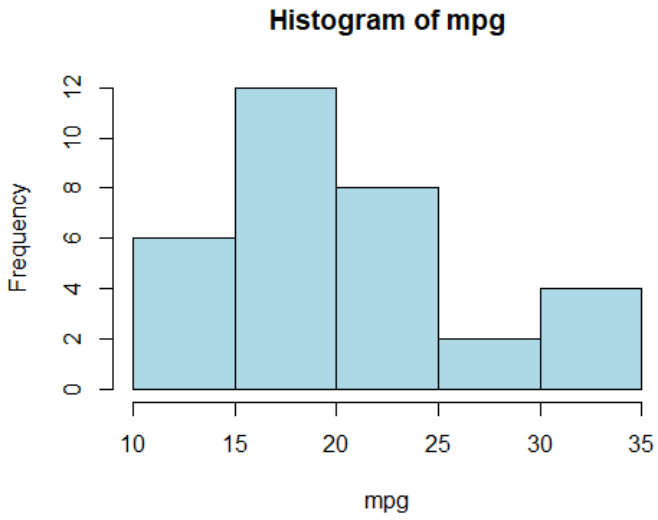



```
> with(mtcars, hist(mpg))# " plain" histogram of a continuous variable, mpg
```

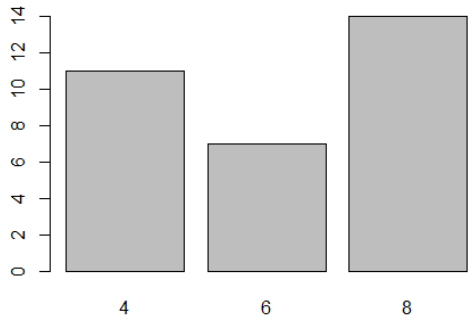


Or:

```
>with(mtcars, hist(mpg, col="lightblue")) # with color
```




```
> #what these charts (barplots or histograms) tell us about the sample data
distribution?
> table(cyl)# create a table of Cylinders.

 4  6  8
11  7 14

Or, assign a name to te table, say C:

> C<-table(cyl)# create a table of Cylinders.
> C
cyl
 4  6  8
11  7 14

> with(mtcars, barplot(table(cyl)))# obtain the bar plot of cylinders.
Why the bar plot is the appropriate graph here and not a histogram?

Or,
> with(mtcars, barplot(C))# since C corresponds to table(cyl)
```




```
> dettach(mtcars)
```