

Verzani Problem Set

Next are considered the problems from Verzani's book on page 19

Problem 3.1

Enter in the data

```
> x <- c(60, 85, 72, 59, 37, 75, 93, 7, 98, 63, 41, 90, 5, 17, 97)
```

Make a stem and leaf plot

```
> sort(x)
[1] 5 7 17 37 41 59 60 63 72 75 85 90 93 97 98
> stem(x)
```

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

```
0 | 577
2 | 7
4 | 19
6 | 0325
8 | 50378
```

Problem 3.3

You can generate random data with the `rnorm` for example.

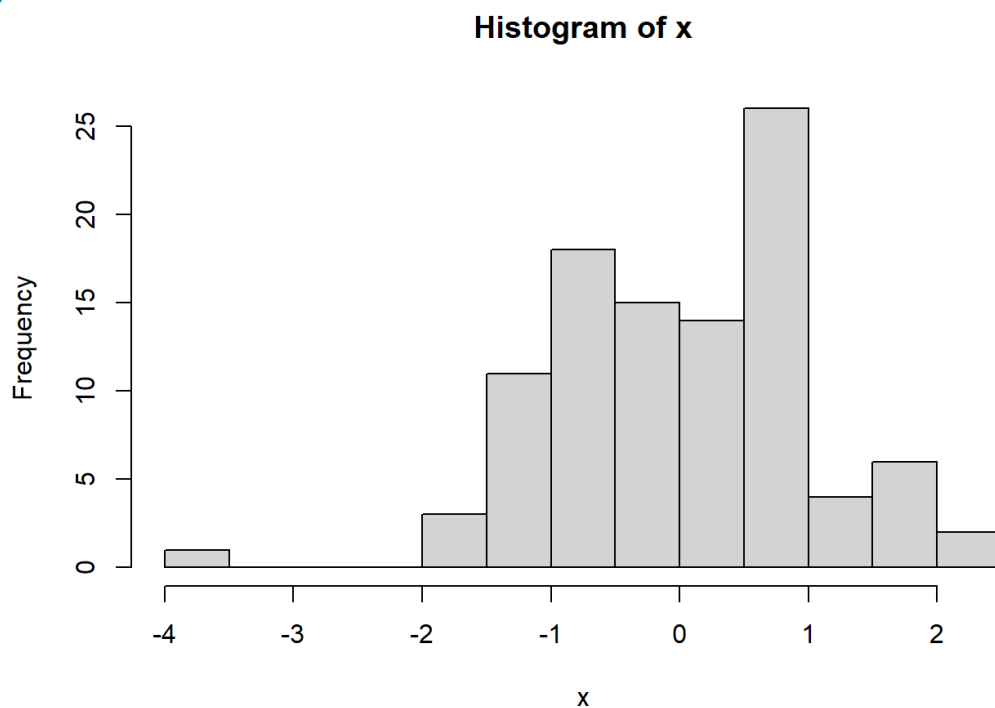
```
> x <- rnorm(100)
```

This produces 100 random numbers with normal distribution. Define it two times.

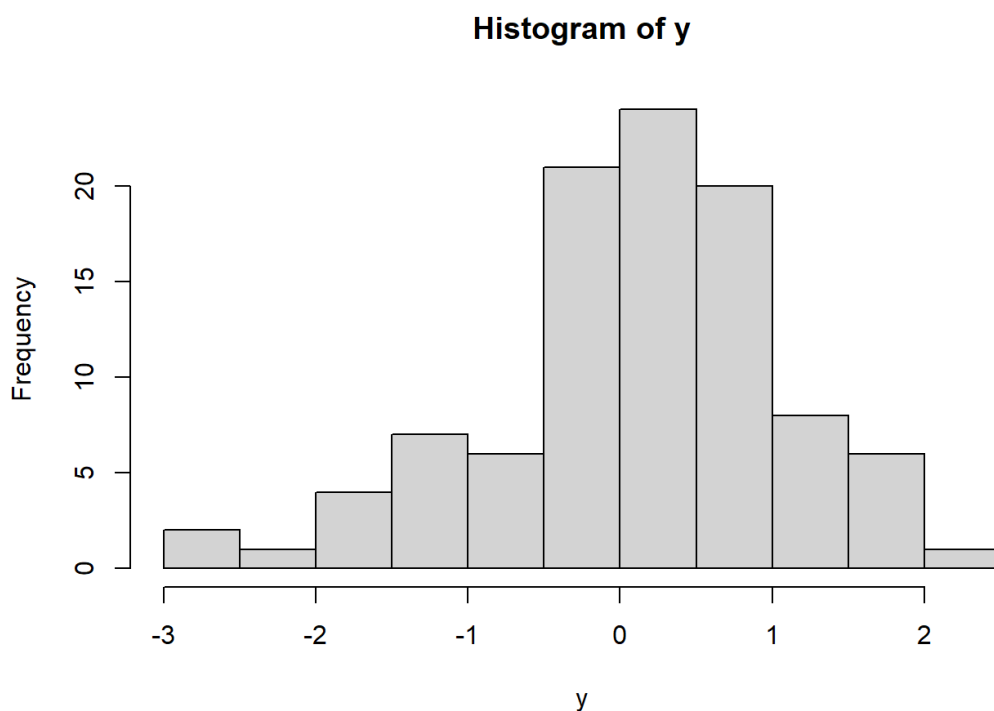
```
> y <- rnorm(100)
```

Create the two histograms. Do you get the same histogram?

```
> hist(x)
```



```
> hist()
```



Problem 3.4

Let's review `south`, `crime` and `aid` data frames from `UsingR` package. What data are they containing?

```
> library(UsingR)
```

Warning: package 'UsingR' was built under R version 4.0.3

Loading required package: MASS

Loading required package: HistData

Loading required package: Hmisc

Loading required package: lattice

Loading required package: survival

Loading required package: Formula

Loading required package: ggplot2

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

`format.pval`, `units`

Attaching package: 'UsingR'

The following object is masked from 'package:survival':

`cancer`

```
> ?south
```

```
> ?crime
```

```
> ?aid
```

`south` data frame contains murder rates for 30 Southern US cities. `crime` data frame contains violent `crime` rates in 50 states of US in 1983 and 1993. `aid` data frame contains monthly payment for federal program.

Now lets review what are this data sets containing

```
> head(south)
```

```
[1] 12 10 10 13 12 12
```

```
> head(crime)
```

```
      y1983 y1993
```

```
Alabama  416.0 871.7
```

```
Alaska   613.8 660.5
```

```
Arizona  494.2 670.8
```

```
Arkansas 297.7 576.5
```

```
California 772.6 1119.7
```

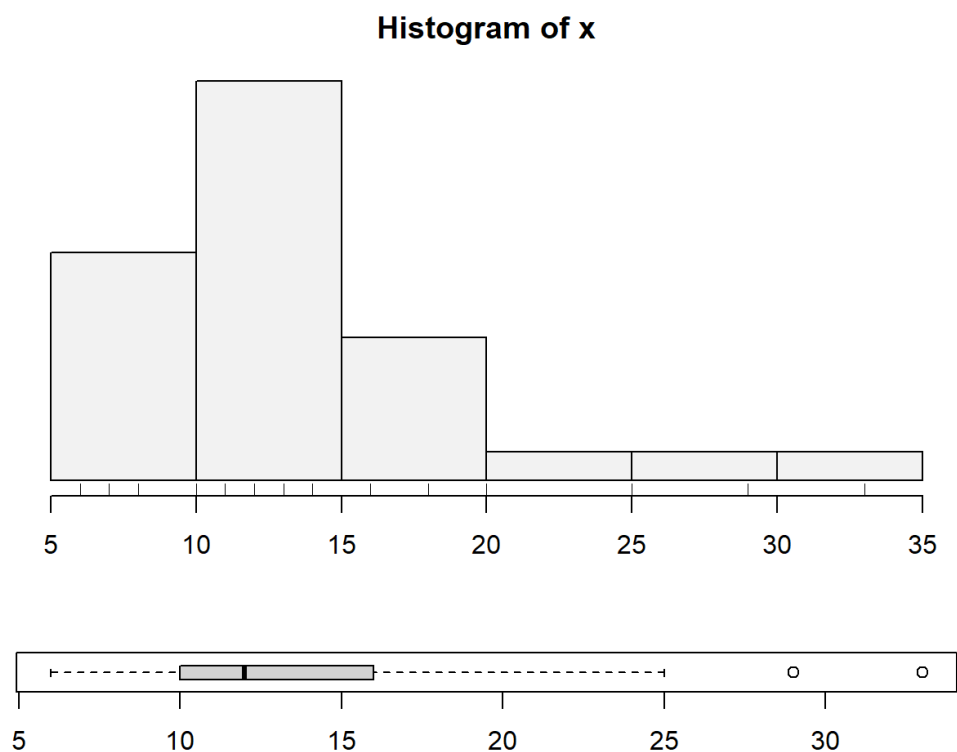
```
Colorado 476.4 578.8
```

```
> head(aid)
```

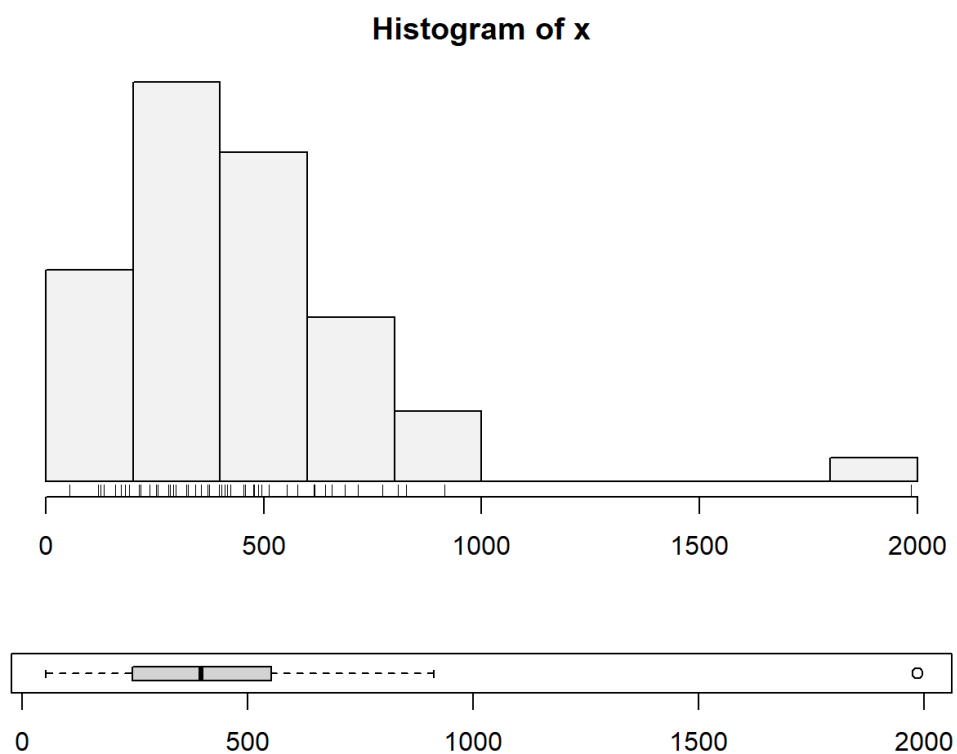
```
Alabama  Alaska  Arizona  Arkansas  California  Colorado  
57.16    253.54   114.23    68.22    199.57    110.86
```

Make a histogram and boxplot for every one of them. Which of them are symmetric? Which of them are skewed? Which of them has outliers?

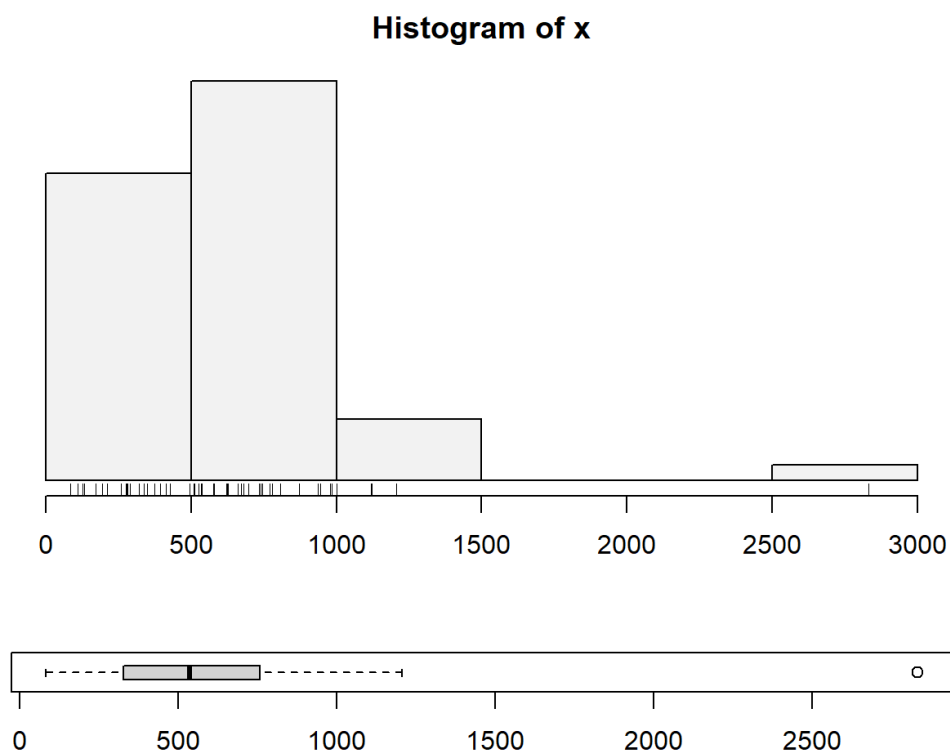
```
> simple.hist.and.boxplot(south)
```



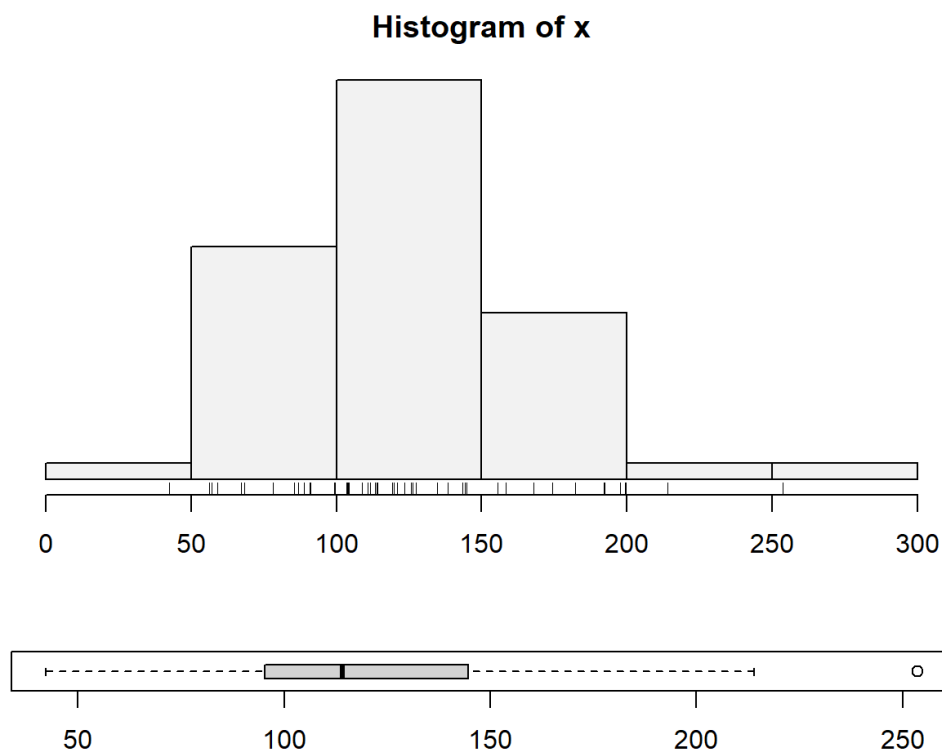
```
> simple.hist.and.boxplot(crime$y1983)
```



```
> simple.hist.and.boxplot(crime$y1993)
```



```
> simple.hist.and.boxplot(aid)
```



Problem 3.5

Let's review `bumpers`, `firstchi` and `math` data frames from `UsingR` package. What data are they containing?

```
> ?bumpers
> ?firstchi
> ?math
```

`bumpers` data frame contains bumper repair costs for various automobiles. `firstchi` data frame contains age of mother at birth of first child. `math` data frame contains standardized math scores.

Now let's review what are these data sets containing

```
> head(bumpers)
```

Honda Accord	Chevrolet Cavalier	Toyota Camry	Saturn SL2
618	795	1304	1308
Mitsubishi Galant	Dodge Monaco		
1340	1456		

```
> head(firstchi)
```

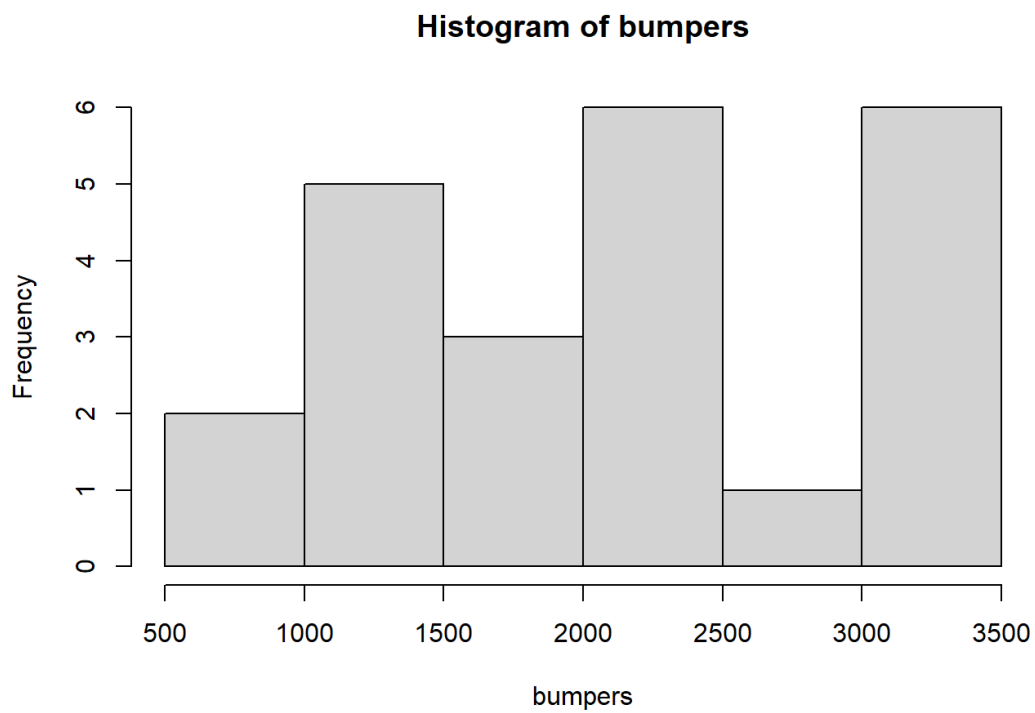
```
[1] 30 18 35 22 23 22
```

```
> head(math)
```

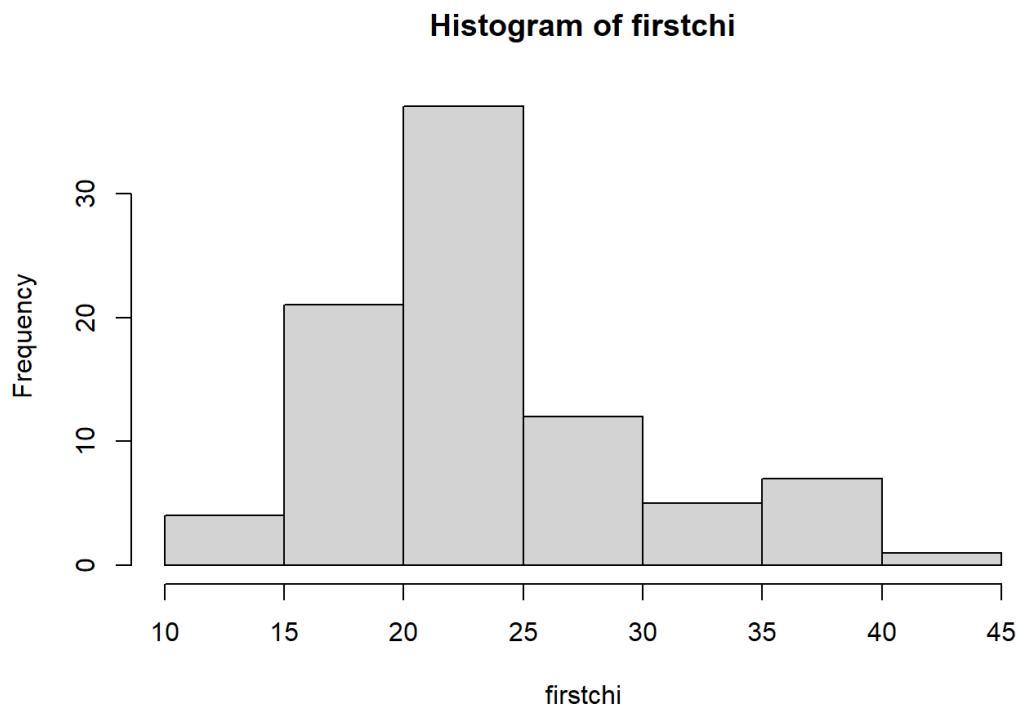
```
[1] 44 49 62 45 51 59
```

Make a histogram for every one of them.

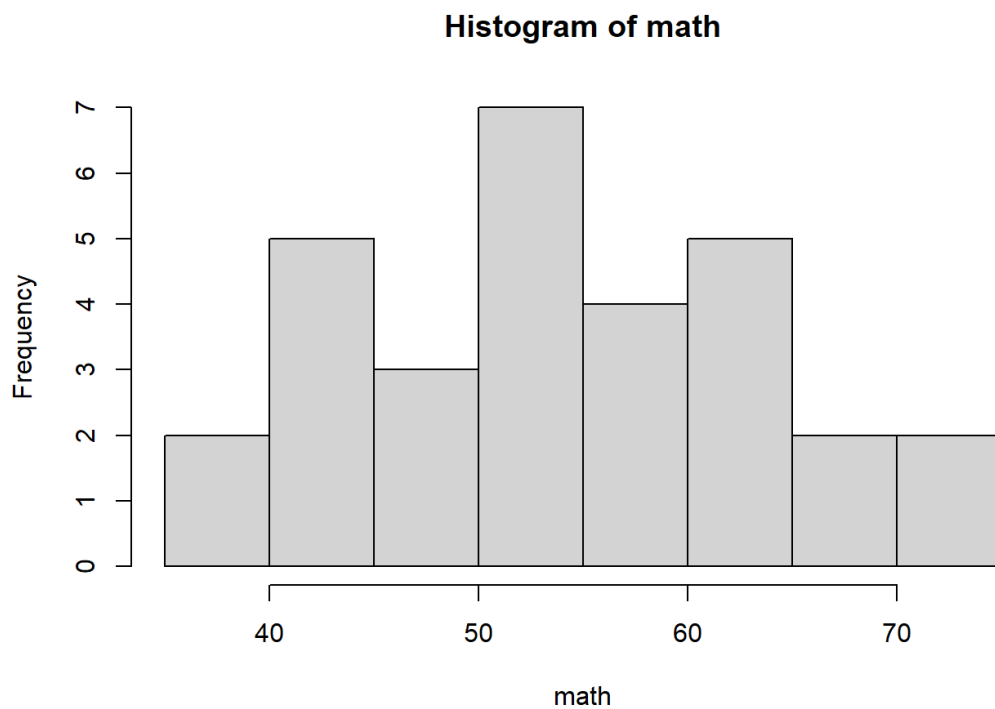
```
> hist(bumpers)
```



```
> hist(firstchi)
```



```
> hist(math)
```



Try to predict the mean, median and standard deviation of every one of them. Check your guesses with the appropriate R commands.

```
> summary(bumpers)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
618	1478	2129	2122	2774	3298

```
> summary(firstchi)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.00	20.00	23.00	23.98	26.00	42.00

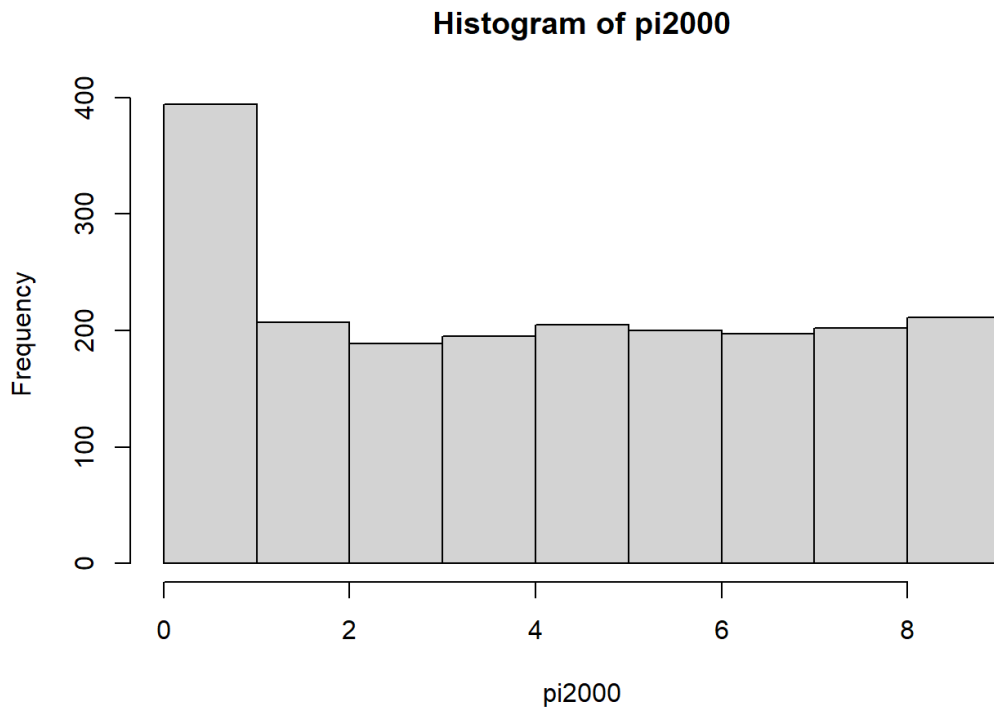
```
> summary(math)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
38.00	49.00	54.00	54.90	61.75	75.00

Problem 3.7

The pi2000 data frames from UsingR package contains the first 2000 digits of π . Make a histogram.

```
> hist(pi2000)
```



Is it surprising? Find the proportion of 1's, 2's and so on.

```
> table(pi2000)
```

```
pi2000
 0  1  2  3  4  5  6  7  8  9
181 213 207 189 195 205 200 197 202 211
```

```
> table(pi2000) / length(pi2000)
```

```
pi2000
 0  1  2  3  4  5  6  7  8  9
0.0905 0.1065 0.1035 0.0945 0.0975 0.1025 0.1000 0.0985 0.1010 0.1055
```

```
> prop.table(table(pi2000))
```

```
pi2000
 0  1  2  3  4  5  6  7  8  9
0.0905 0.1065 0.1035 0.0945 0.0975 0.1025 0.1000 0.0985 0.1010 0.1055
```

```
> proportions(table(pi2000))
```

```
pi2000
 0  1  2  3  4  5  6  7  8  9
0.0905 0.1065 0.1035 0.0945 0.0975 0.1025 0.1000 0.0985 0.1010 0.1055
```


Problem 3.8

Fit a density estimate to the pi2000 data frame.

```
> density(pi2000)
```

Call:

```
density.default(x = pi2000)
```

Data: pi2000 (2000 obs.); Bandwidth 'bw' = 0.5657

```
      x      y  
Min. :-1.697 Min. :0.0007207  
1st Qu.: 1.401 1st Qu.:0.0755293  
Median : 4.500 Median :0.0990100  
Mean   : 4.500 Mean   :0.0805830  
3rd Qu.: 7.599 3rd Qu.:0.1007573  
Max.   :10.697 Max.   :0.1046462  
> hist(pi2000, probability = TRUE)  
> lines(density(pi2000), lwd = 2)
```

