

STAT4198 HW1 Min Yang

1.

```
mydata <-read.table("~/Desktop/cars01.txt",header=TRUE)
```

Q2:yes, mydata is a data frame.

```
class(mydata)
```

```
## [1] "data.frame"
```

Q3: 261 rows and 8 columns.

```
dim(mydata)
```

```
## [1] 261 8
```

```
ncol(mydata)
```

```
## [1] 8
```

```
nrow(mydata)
```

```
## [1] 261
```

Q4: list the names of the columns

mpg,cylinders,cubicinches,hp,weightlbs,time-to-60,year,brand

```
head(mydata)
```

```
##      mpg cylinders cubicinches  hp weightlbs time.to.60 year  brand
## 1 14.0          8         350 165      4209         12 1972    US.
## 2 31.9          4          89  71      1925         14 1980 Europe.
## 3 17.0          8         302 140      3449         11 1971    US.
## 4 15.0          8         400 150      3761         10 1971    US.
## 5 30.5          4          98  63      2051         17 1978    US.
## 6 23.0          8         350 125      3900         17 1980    US.
```

5.

```
names(mydata)
```

```
## [1] "mpg"          "cylinders"    "cubicinches" "hp"          "weightlbs"
## [6] "time.to.60"  "year"        "brand"
```

6.

```
mydata[[3]]
```

```
##      [1] 350  89 302 400  98 350 351 440 183  89 108  97 302 350 225 350  85
##     [18] 304 113 107  86 121 200 351  89  91  71 351 250  98  97 250 120 199
##     [35]  97 231  97  80 351 232 250 302 250 250 455 120 318 122  97 200 400
##     [52] 141 340 116 350 400 156  98  90 360 232 108 168  86 318 250  97  91
##     [69]  97 305 198 112  97 119  98 262 232 134  98 108 400 429 305 115 116
##     [86]  72 199 225  90 119  86 119 350 250  98 231 350 250 318 250 400 400
##    [103]  98 105 108 151  90  88 121 307  97 121 351  91  91 225 151 231  98
##    [120] 429 101  79 400 156  97 140 360 105 114  98  97 151 455 304 350 318
##    [137]  91 302 121 120 173  97 144  81 267 120 225 151 350 302 120 134 455
##    [154] 250 107 250  70  98 232 163 151 318 171 198  97 260 225 156 156 105
```

```
## [171] 107 131 68 112 122 96 400 304 302 98 200 97 305 318 112 200 250
## [188] 318 351 440 156 151 400 350 318 98 135 90 90 85 116 140 260 225
## [205] 140 390 85 250 307 168 146 231 79 79 225 119 85 351 173 302 258
## [222] 121 318 168 302 83 360 76 200 140 383 120 140 318 130 97 318 200
## [239] 86 122 350 225 400 110 454 318 318 90 97 135 98 121 121 350 173
## [256] 113 305 91 232 232 250
```

Q7: provide the line to retrieve 3rd column using the column name

```
mydata[["cubicinches"]]
```

```
## [1] 350 89 302 400 98 350 351 440 183 89 108 97 302 350 225 350 85
## [18] 304 113 107 86 121 200 351 89 91 71 351 250 98 97 250 120 199
## [35] 97 231 97 80 351 232 250 302 250 250 455 120 318 122 97 200 400
## [52] 141 340 116 350 400 156 98 90 360 232 108 168 86 318 250 97 91
## [69] 97 305 198 112 97 119 98 262 232 134 98 108 400 429 305 115 116
## [86] 72 199 225 90 119 86 119 350 250 98 231 350 250 318 250 400 400
## [103] 98 105 108 151 90 88 121 307 97 121 351 91 91 225 151 231 98
## [120] 429 101 79 400 156 97 140 360 105 114 98 97 151 455 304 350 318
## [137] 91 302 121 120 173 97 144 81 267 120 225 151 350 302 120 134 455
## [154] 250 107 250 70 98 232 163 151 318 171 198 97 260 225 156 156 105
## [171] 107 131 68 112 122 96 400 304 302 98 200 97 305 318 112 200 250
## [188] 318 351 440 156 151 400 350 318 98 135 90 90 85 116 140 260 225
## [205] 140 390 85 250 307 168 146 231 79 79 225 119 85 351 173 302 258
## [222] 121 318 168 302 83 360 76 200 140 383 120 140 318 130 97 318 200
## [239] 86 122 350 225 400 110 454 318 318 90 97 135 98 121 121 350 173
## [256] 113 305 91 232 232 250
```

Q8: list elements of the second row.

```
mydata[2,]
```

```
## mpg cylinders cubicinches hp weightlbs time.to.60 year brand
## 2 31.9 4 89 71 1925 14 1980 Europe.
```

9.

```
mydata$newcolumn<-NA
```

10.

```
mydata$newcolumn<-mydata$brand=="US."
```

Q11:

```
table(mydata$newcolumn)
```

```
##
## FALSE TRUE
## 99 162
```

12.

```
newmydata<-rbind(mydata,list(30.5,4 , 98, 63, 2051, 17, 1978, "US." ," TRUE"))
```

Q13: mean mpg is 23.1729

```
mean(newmydata$mpg)
```

```
## [1] 23.1729
```

Q14: median mpg is 22

```
median(newmydata$mpg)
```

```
## [1] 22
```

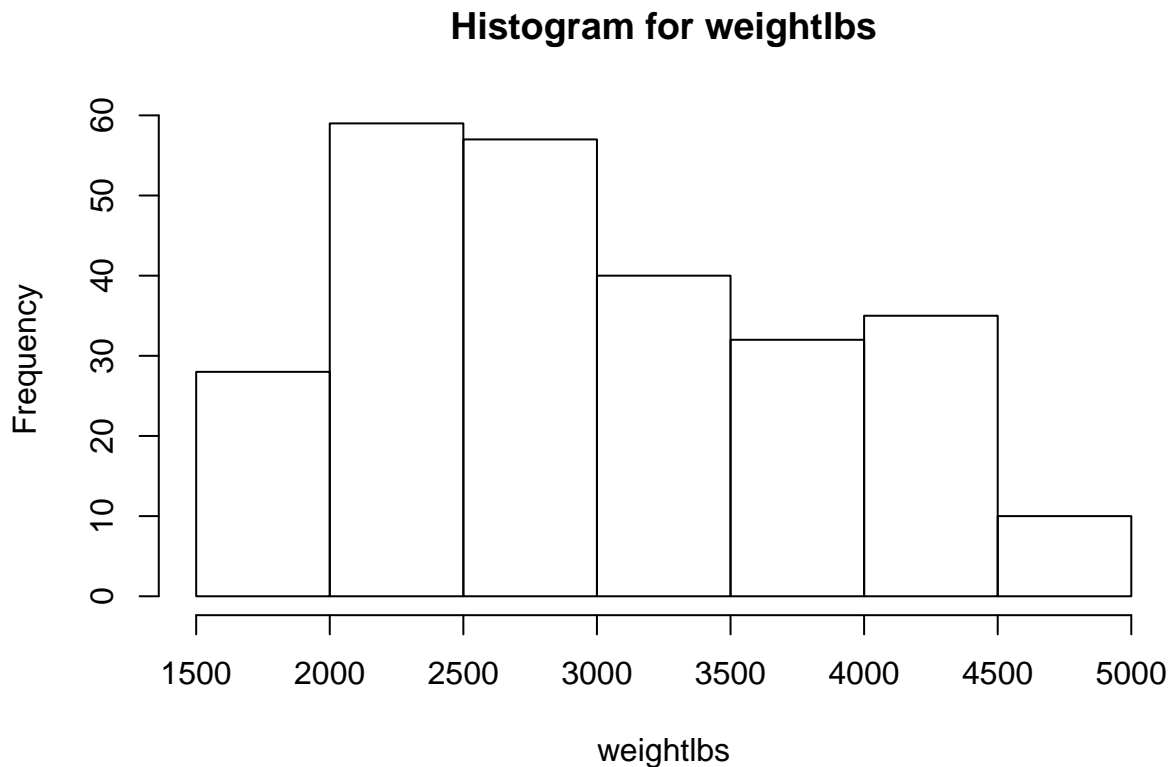
Q15: Variance of mpg is 61.18022

```
var(newmydata$mpg)
```

```
## [1] 61.18022
```

Q16:

```
hist(mydata[["weightlbs"]],breaks=5,  
     main="Histogram for weightlbs",  
     xlab="weightlbs")
```



Q17: The first histogram has a frequency scale from 0 to 120, while the second histogram has a frequency scale from 0 to 30. The second plot on the right side has 16 breaks which is more specific than the first one. With 16 breaks, the second histogram can provide more details on the distribution of the frequency of “weightlbs”.

```
par(mfrow = c(1, 2))
```

```
hist(mydata[["weightlbs"]],breaks=3,  
     main="Histogram for weightlbs",  
     xlab="weightlbs")
```

```
hist(mydata[["weightlbs"]],breaks=16,  
     main="Histogram for weightlbs",  
     xlab="weightlbs"  
)
```

