# A Brief Introduction to R

Also see Appendix B of *Linear Models with R*

```
> 1+1
[1] 2
> 2^3 # Two to the power 3
[1] 8

> 1:30
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
[26] 26 27 28 29 30

> gamma(.5)^2      # Gamma(1/2) = Sqrt(Pi)
[1] 3.141593

> x = 1           # Assigns the value 1 to x
> y = 2           # Assigns the value 2 to y
> x+y
[1] 3
> z = x+y
> z
[1] 3
> x = c(1,2,3,4,5,6)    #  Collect these numbers; x is now a vector

> z # No dynamic updating; it's not a spreadsheet
[1] 3
> x+y
[1] 3 4 5 6 7 8

> y = 1 + 2*x
> cbind(x,y)
     x  y
[1,] 1  3
[2,] 2  5
[3,] 3  7
[4,] 4  9
[5,] 5 11
[6,] 6 13

> z = y[x>4]           # z gets y such that x > 4
> z
[1] 11 13
> z2 = subset(y,x>4); z2
[1] 11 13

> # If you put an array of integers inside the brackets, you get those
> # elements, in the order indicated.

> y[c(6,5,4,3,2,1)] # y in opposite order
[1] 13 11  9  7  5  3
> y[c(2,2,2,3,4)] # Repeats are okay
[1] 5 5 5 7 9
> y[7] # There is no seventh element.  NA is the missing value code
[1] NA
```

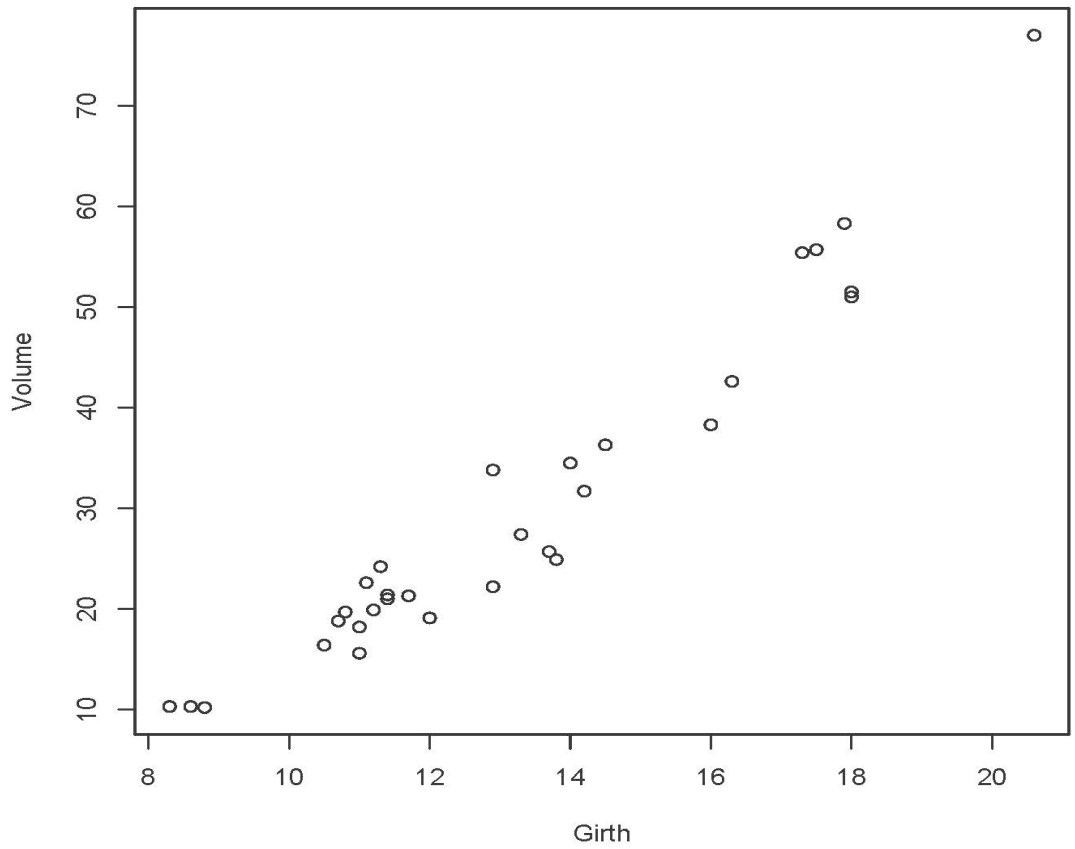
---

```
> # Computing probabilities, etc.
>
> pnorm(0) # Area less than zero for a standard normal
[1] 0.5
>
> pnorm(160,mean=100,sd=15)  # IQ of 160
[1] 0.9999683
>
> pcauchy(4)
[1] 0.9220209
>
> dnorm(0) # height of the curve
[1] 0.3989423
>
> dpois(0,lambda=3) # P(Y=0) for Y ~ Poisson(3)
[1] 0.04978707
>
> qnorm(0.975) # z value with P(Z<z) = 0.975
[1] 1.959964
>
> qf(0.975,df1=6,df2=122) # Critical value for F, not in any table
[1] 2.513606
>
> CriticalValue = qchisq(0.95,df=1:8)
> df=1:8; cbind(df,CriticalValue)
     df CriticalValue
[1,]  1      3.841459
[2,]  2      5.991465
[3,]  3      7.814728
[4,]  4      9.487729
[5,]  5     11.070498
[6,]  6     12.591587
[7,]  7     14.067140
[8,]  8     15.507313
```

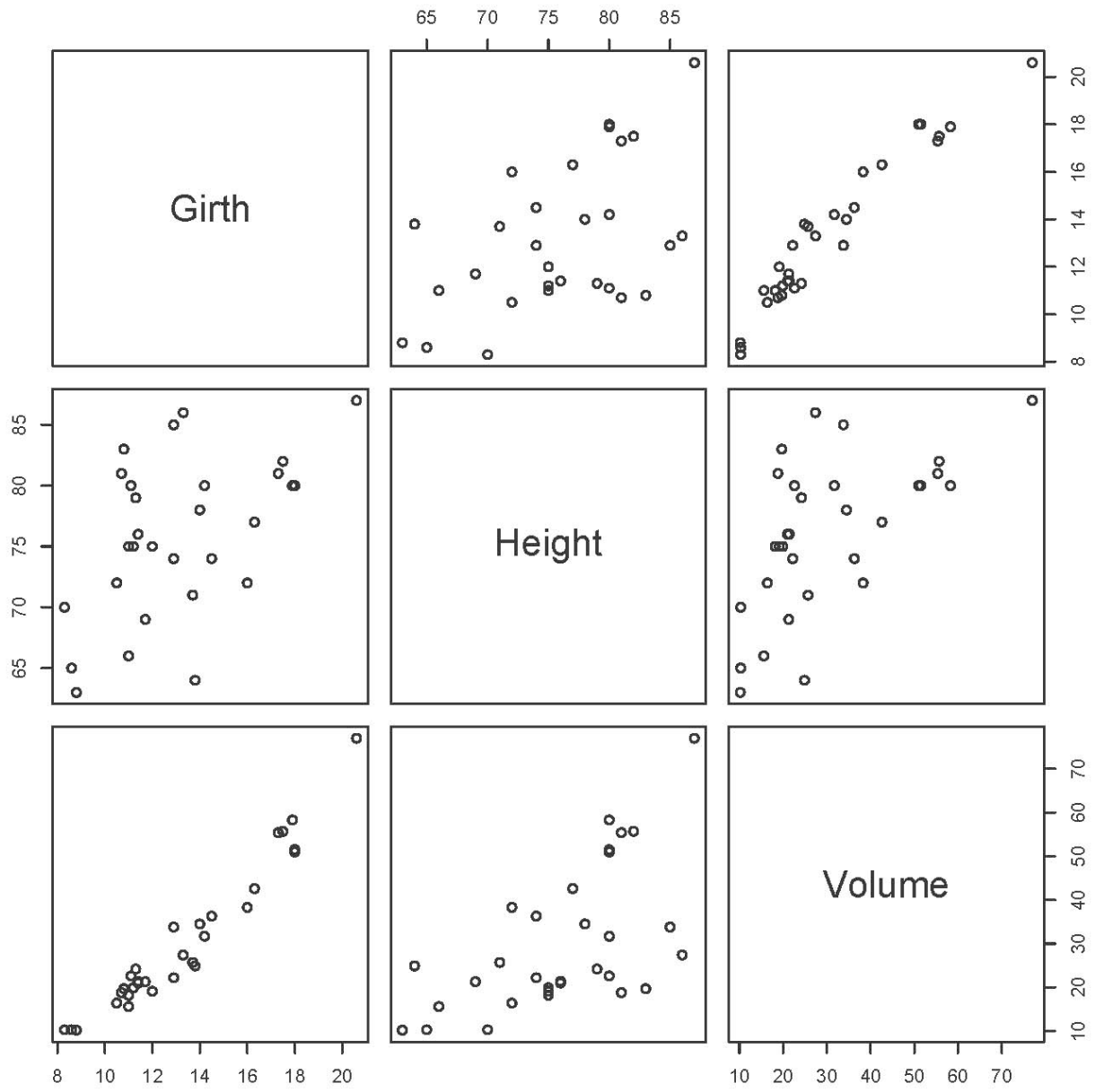
```
> # data() lists available data sets
> trees
   Girth Height Volume
1    8.3     70   10.3
2    8.6     65   10.3
3    8.8     63   10.2
4   10.5     72   16.4
5   10.7     81   18.8
6   10.8     83   19.7
7   11.0     66   15.6
8   11.0     75   18.2
9   11.1     80   22.6
10  11.2     75   19.9
11  11.3     79   24.2
12  11.4     76   21.0
13  11.4     76   21.4
14  11.7     69   21.3
15  12.0     75   19.1
16  12.9     74   22.2
17  12.9     85   33.8
18  13.3     86   27.4
19  13.7     71   25.7
20  13.8     64   24.9
21  14.0     78   34.5
22  14.2     80   31.7
23  14.5     74   36.3
24  16.0     72   38.3
25  16.3     77   42.6
26  17.3     81   55.4
27  17.5     82   55.7
28  17.9     80   58.3
29  18.0     80   51.5
30  18.0     80   51.0
31  20.6     87   77.0
> mean(trees$Height); var(trees$Height)
[1] 76
[1] 40.6
> mean(Height)
Error in mean(Height) : object 'Height' not found
> attach(trees)
> mean(Height)
[1] 76
```

```
> summary(trees)
     Girth           Height       Volume     
 Min.   : 8.30   Min.   :63   Min.   :10.20  
 1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40  
 Median :12.90   Median :76   Median :24.20  
 Mean   :13.25   Mean   :76   Mean   :30.17  
 3rd Qu.:15.25   3rd Qu.:80   3rd Qu.:37.30  
 Max.   :20.60   Max.   :87   Max.   :77.00  

> cor(trees)
           Girth    Height    Volume
Girth  1.0000000 0.5192801 0.9671194
Height 0.5192801 1.0000000 0.5982497
Volume 0.9671194 0.5982497 1.0000000
> plot(Girth,Volume)
```



```
> pairs(trees) # Can also give it a matrix, in which the columns are
variables
```

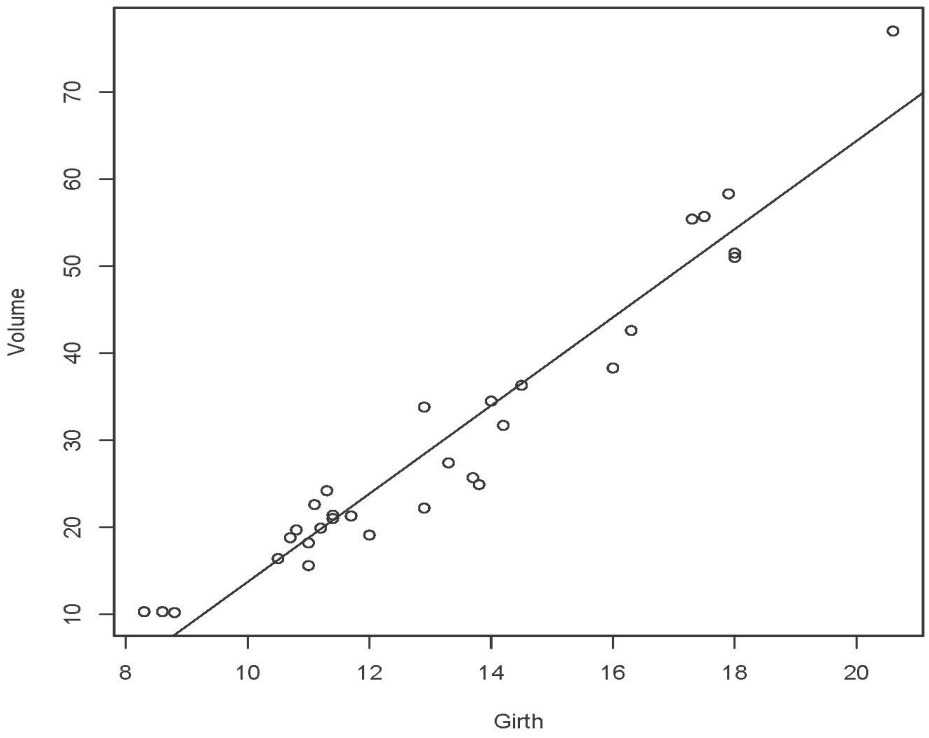


```
> lm(Volume~Girth) # Means V_i = beta0 + beta1 G_i + epsilon_i

Call:
lm(formula = Volume ~ Girth)

Coefficients:
(Intercept)        Girth
    -36.943        5.066

> # The intercept of the least squares line is -36.943, and the slope is 5.066.
> x = c(5,25); y = -36.943 + 5.066*x
> cbind(x,y)
      x       y
[1,]  5 -11.613
[2,] 25  89.707
>
> # Add the least squares line to the plot.
> lines(x,y,lty=1) # Line Type 1 is a solid line.
```




---