Bivariate Data

- Many Data consists of 2 variables.
    - One of them is a Dependent variable, sometimes referred to as Response variable
    - One of them being a independent variable, sometime referred to as Predictor or Explanatory variable.
- There are situations when there is one response variable and multiple explanatory variables. We will not discuss them today.
- We will focus on Bivariate Data.

Examples

- Maternal Smoking and its effect on Birth Weight.
- Attendance in Classes and its effect on Scores in an Exam.
- Age and Heart rate
- Effect of Vitamin C on Toothgrowth

Data: Dengue Cases

```
> y = read.table("dengue.csv", header=TRUE)
```

The dataset is of 50 Dengue pediatric patients at Manipal hospital.

The patient's:

- `BICARB` reading is taken at admission.
- End resulting `DIAGNOSIS` is either severe form of Dengue Shock Syndrome (DSS) or milder form of Dengue (D).

Data: Dengue Case Data

```
> head(y)
  DIAGNOSIS BICARB
1       DSS   16.2
2       DSS   22.0
3       DSS   16.0
4       DSS   21.3
5       DSS   19.0
6       DSS   18.7
```

```
> tail(y)
   DIAGNOSIS BICARB
45         D   22.0
46         D   16.6
47         D   18.3
48         D   23.0
49         D   24.0
50         D   21.0
```

Dengue Case Data

```
> Diagnosis = y$DIAGNOSIS
> Marker = y$BICARB
```

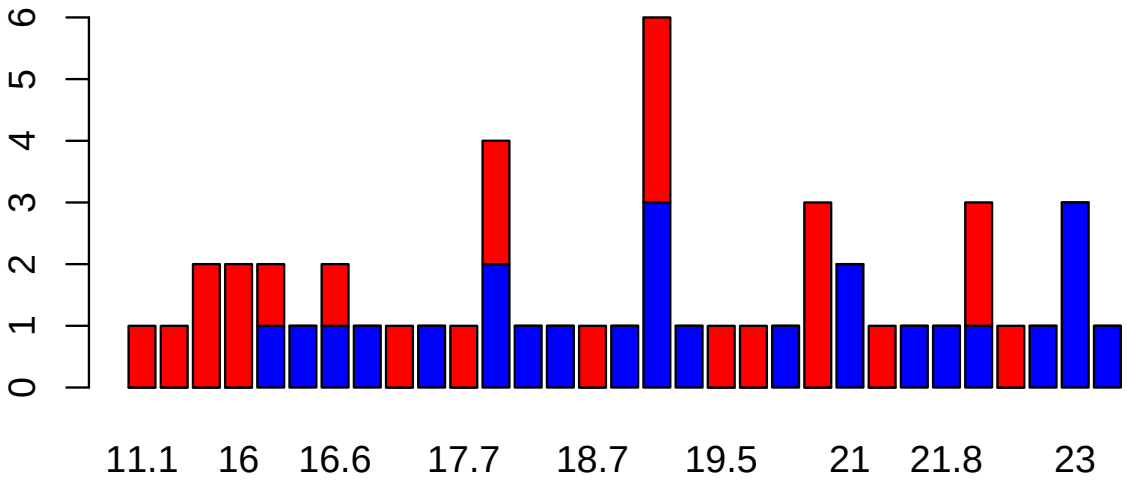
The `table` command will summarize Bivariate Data as well.

```
> table(Diagnosis, Marker)
```

```
         Marker
Diagnosis 11.1 11.7 14 16 16.2 16.3 16.6 17 17.4 17.6 17.7 18 18.2 18.3 18.7
      D      0    0  0  0    1    1    1  1    0    1    0  2    1    1    0
      DSS    1    1  2  2    1    0    1  0    1    0    1  2    0    0    1
         Marker
Diagnosis 18.9 19 19.3 19.5 19.9 20.5 20.7 21 21.3 21.6 21.8 22 22.5 22.7 23 24
      D      1  3    1    0    0    1    0  2    0    1    1  1    0    1  3  1
      DSS    0  3    0    1    1    0    3  0    1    0    0  2    1    0  0  0
```

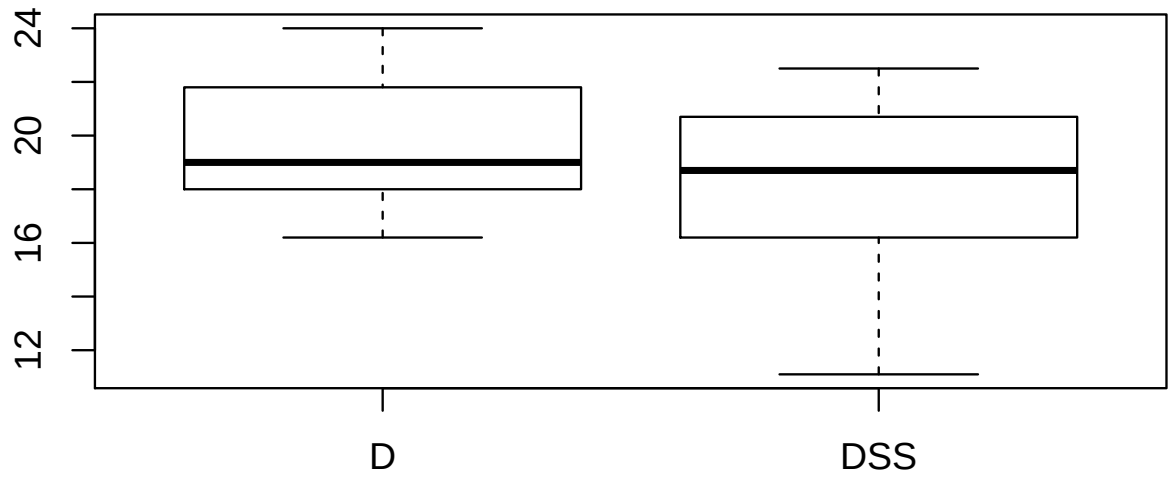
Bar Charts

```
> barplot(table(Diagnosis, Marker), col = c("blue", "Red"))
```



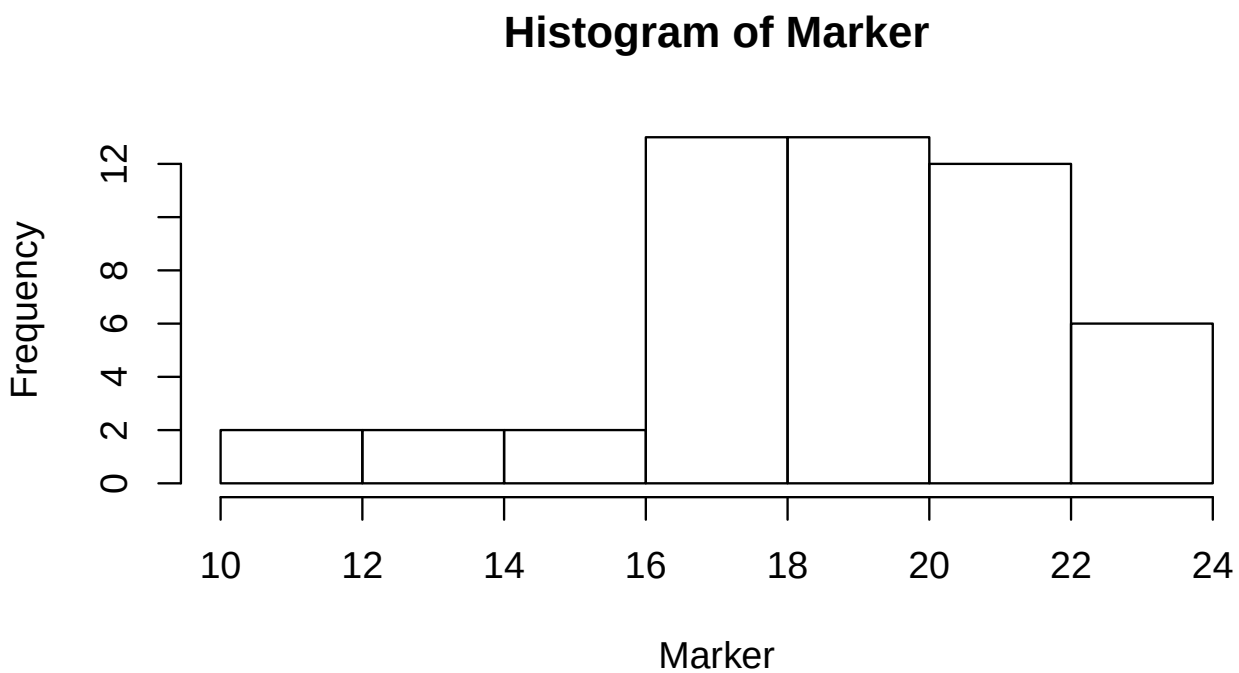
Box Plot

```
> plot(Diagnosis, Marker)
```



Dengue Case Data

```
> hist(Marker)
```



Dengue Case Data

We can categorize the data in to say 0,1,2,3

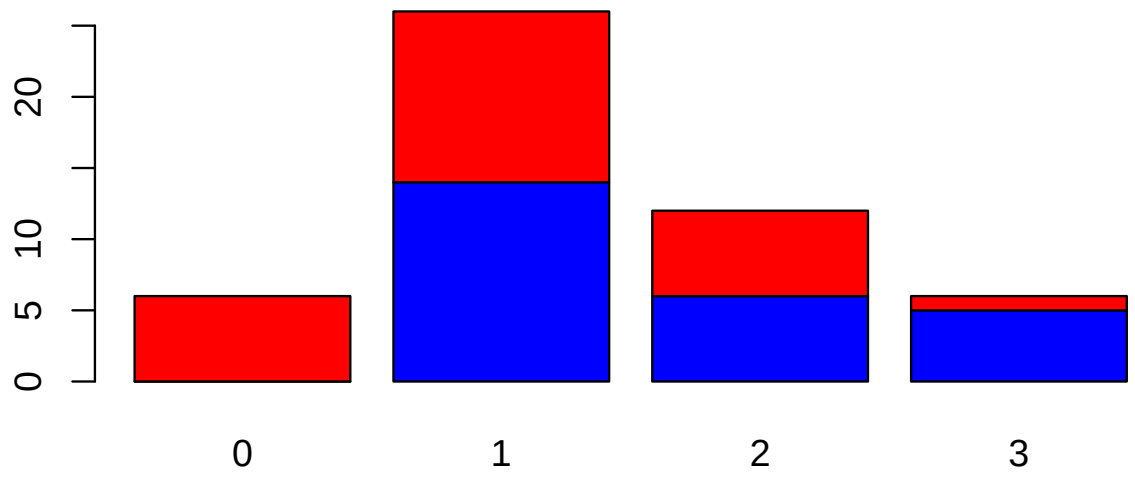
```
> z = replace(Marker, Marker <= 16,0)
> z1 = replace(z, 16< z & z<=20,1)
> z2 = replace(z1, 20< z1 & z1<=22,2)
> z3 = replace(z2, 22< z2,3)
```

Dengue Case Data

```
> table(Diagnosis, z3)

         z3
Diagnosis  0  1  2  3
      D    0 14  6  5
      DSS  6 12  6  1
```

Bar Charts

```
> barplot(table(Diagnosis, z3), col = c("blue", "Red"))
```

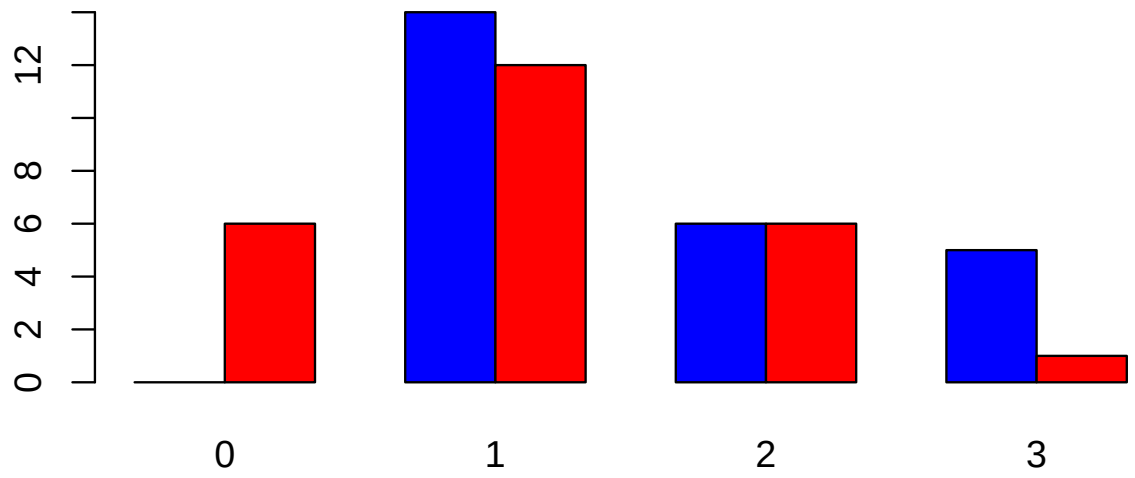


Bar Charts

- `Barplot` plots each row of data.
- Stacked manner is the default,
- It can be beside each other (by setting option `beside=TRUE`).

Bar Charts

```
> barplot(table(Diagnosis, z3), col = c("blue", "Red"),
+ beside=TRUE)
```

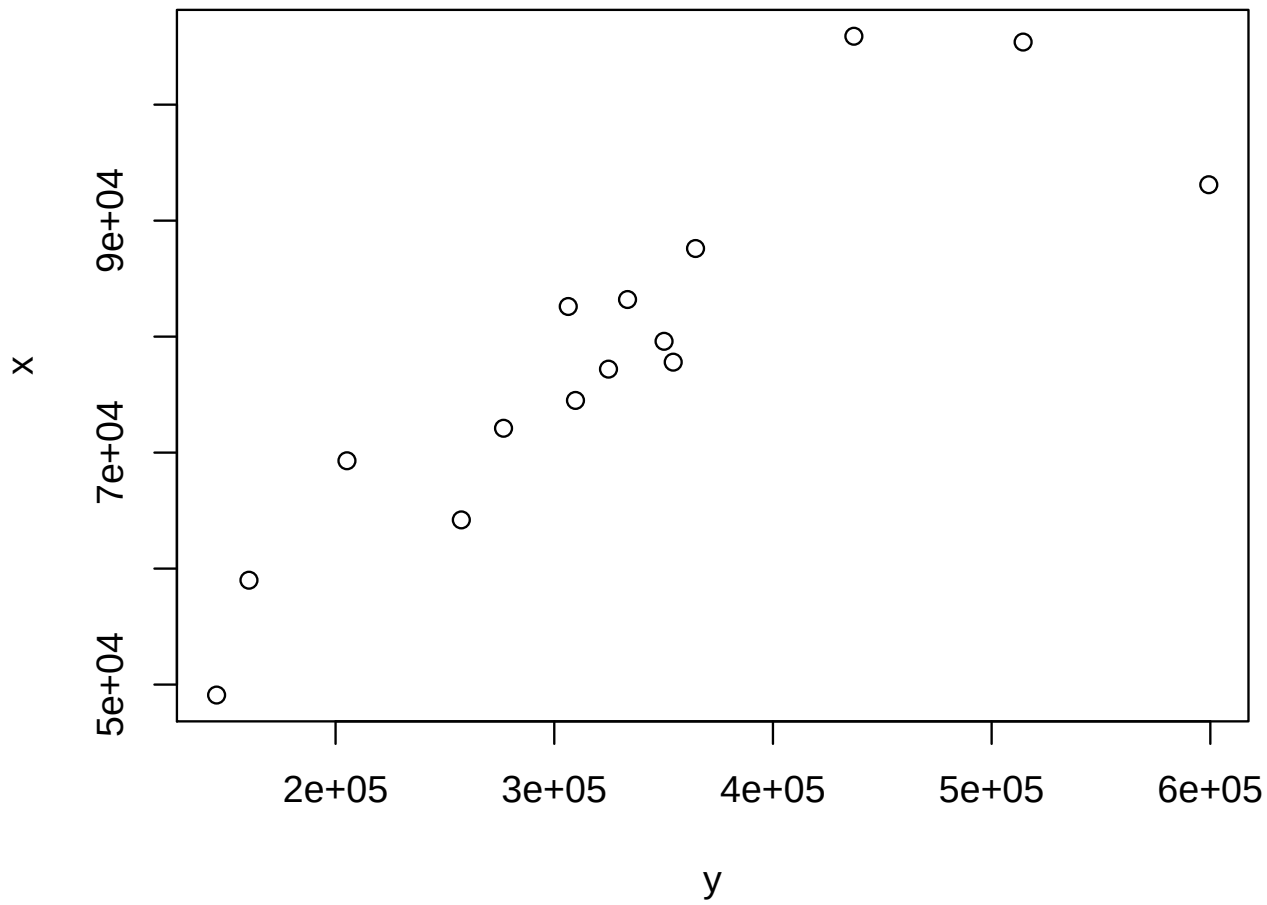


Simple Linear Regression

```
> require(UsingR)

> data(home)
> head(home)

    old    new
1 64200 257500
2 72100 276800
3 87600 364600
4 59000 160400
5 83200 333500
6 49100 145600

> x = home$old
> y = home$new
```

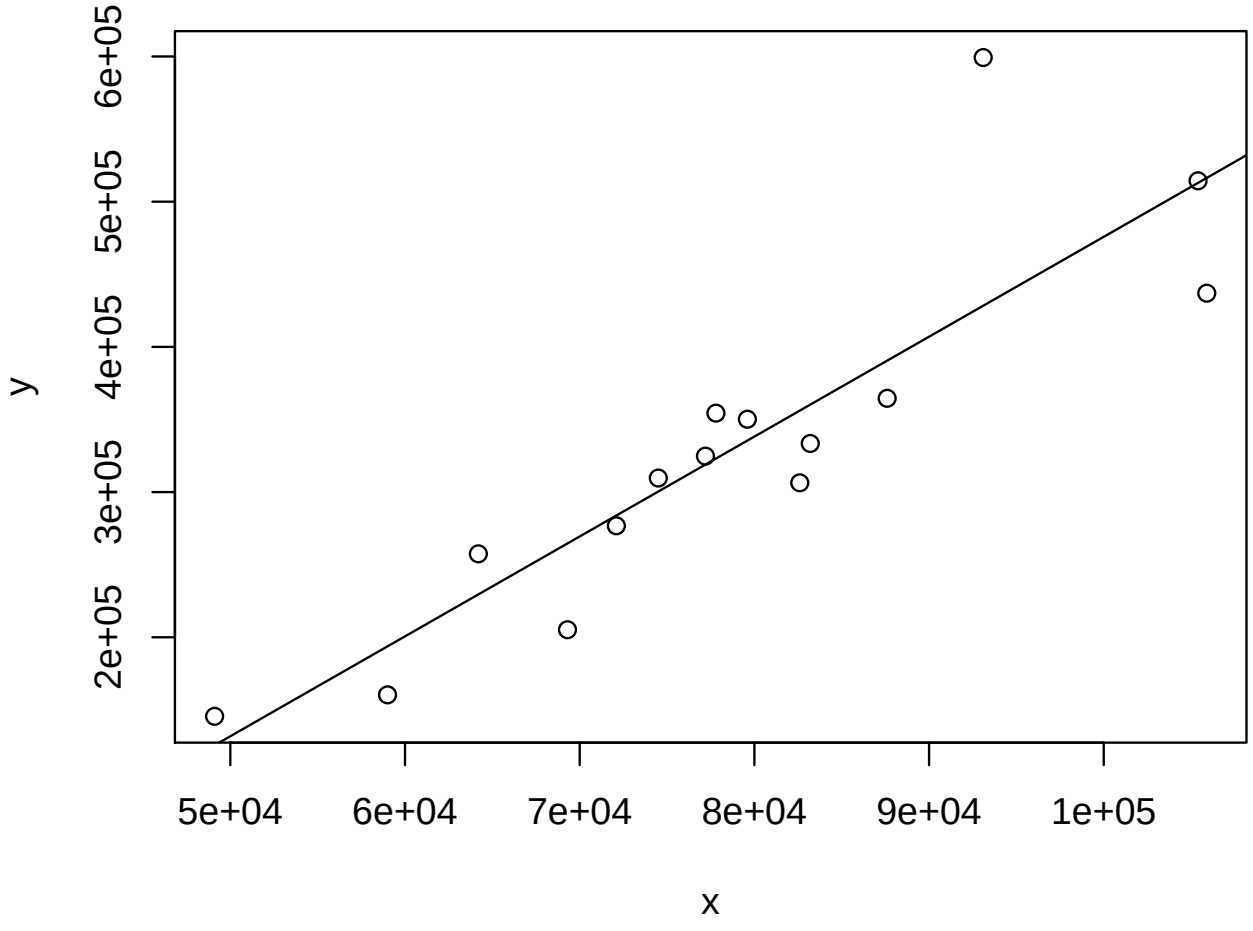
Simple Linear Regression

```
> plot(y,x)
```



Simple Linear Regression

There could be a Linear Relationship between y and x.



Simple Linear Regression

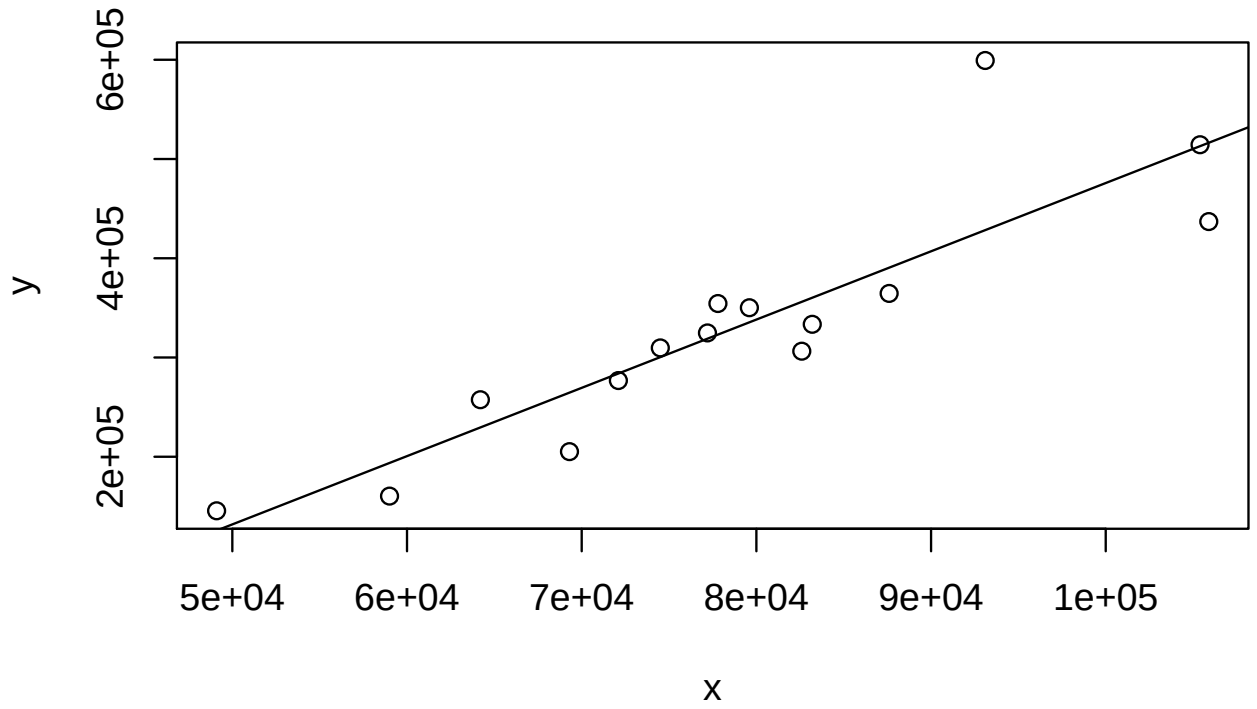
We can use the inbuilt function `lm`

```
> residual = lm(y~x)
> residual

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
 -2.121e+05    6.879e+00
```

Simple Linear Regression

We can use the inbuilt function `lm`

```
> plot(y~x)
> abline(lm(y~x))
```



Simple Linear Regression

Consider the following dataset in `HistData` in `UsingR` by F. Galton.

```
> head(GaltonFamilies)
```

	family	father	mother	midparentHeight	children	childNum	gender	childHeight
1	001	78.5	67.0	75.43	4	1	male	73.2
2	001	78.5	67.0	75.43	4	2	female	69.2
3	001	78.5	67.0	75.43	4	3	female	69.0
4	001	78.5	67.0	75.43	4	4	female	69.0
5	002	75.5	66.5	73.66	4	1	male	73.5
6	002	75.5	66.5	73.66	4	2	male	72.5

Simple Linear Regression

Consider the following dataset in `HistData` in `UsingR` by F. Galton.

```
> require(UsingR)
> mh = GaltonFamilies$mother
> head(mh)
```

```
[1] 67.0 67.0 67.0 67.0 66.5 66.5
```

```
> fh = GaltonFamilies$father
> head(fh)
```

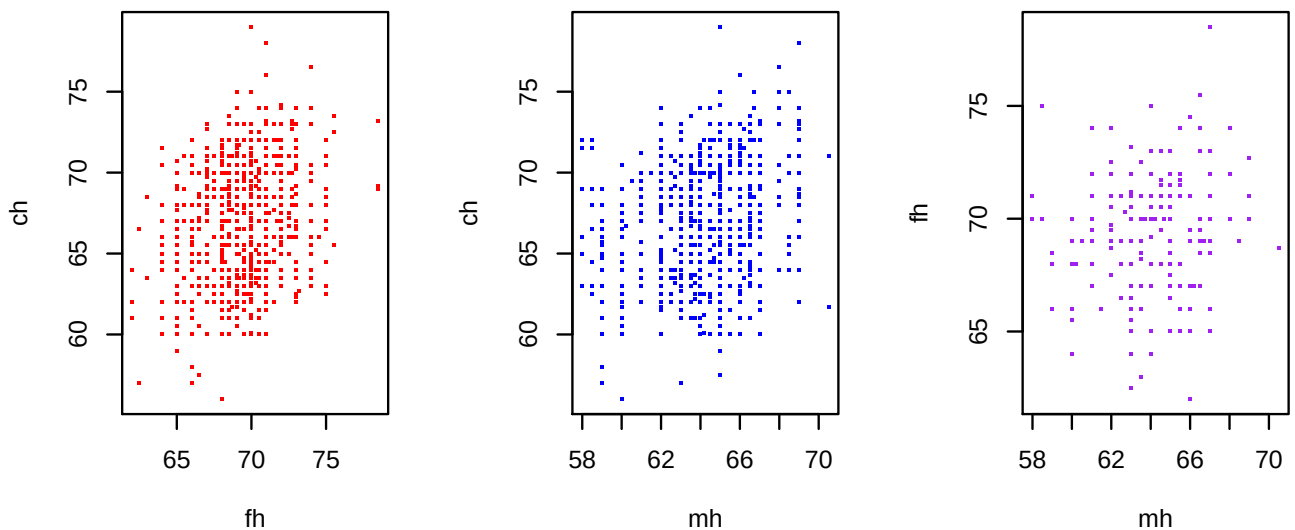
```
[1] 78.5 78.5 78.5 78.5 75.5 75.5
```

```
> ch = GaltonFamilies$childHeight
> head(ch)
```

```
[1] 73.2 69.2 69.0 69.0 73.5 72.5
```

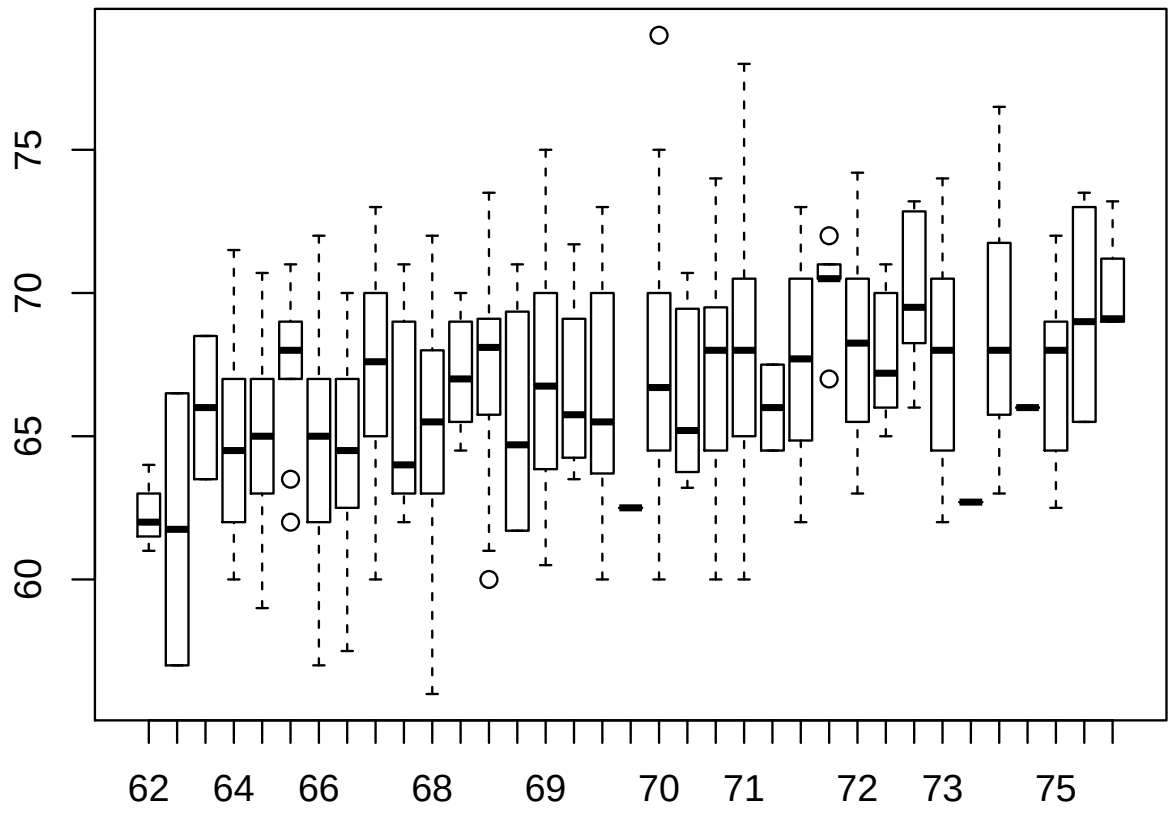
Simple Linear Regression

```
> par(mfrow=c(1,3))
> plot(ch~fh, pch=".", col="Red")
> plot(ch~mh, pch=".", col="Blue")
> plot(fh~mh, pch=".", col="green")
```



Simple Linear Regression

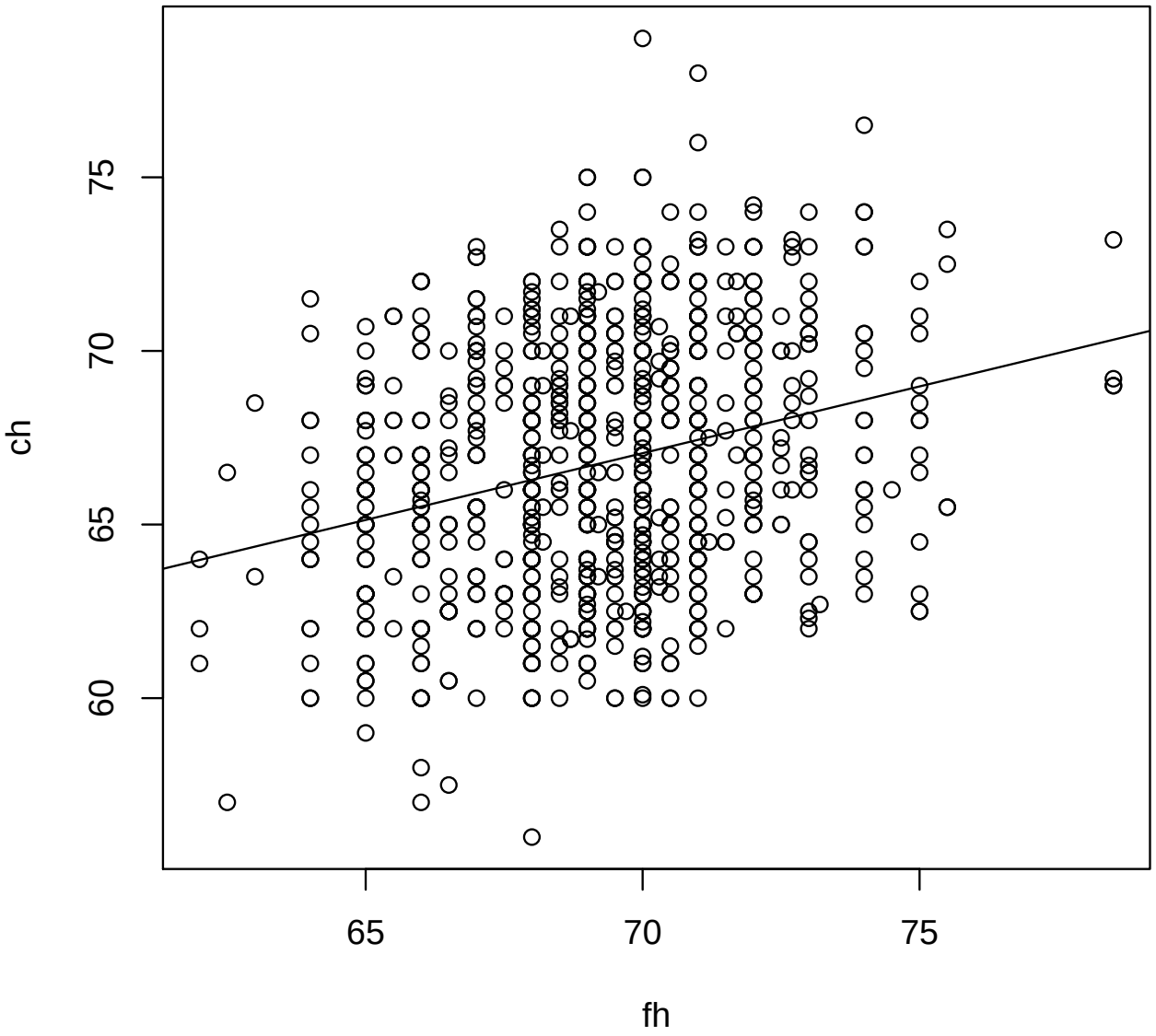
```
> boxplot(ch~fh)
```



Simple Linear Regression

```
> res = lm(ch~fh)
> res

Call:
lm(formula = ch ~ fh)

Coefficients:
(Intercept)           fh
    40.1393       0.3845
```

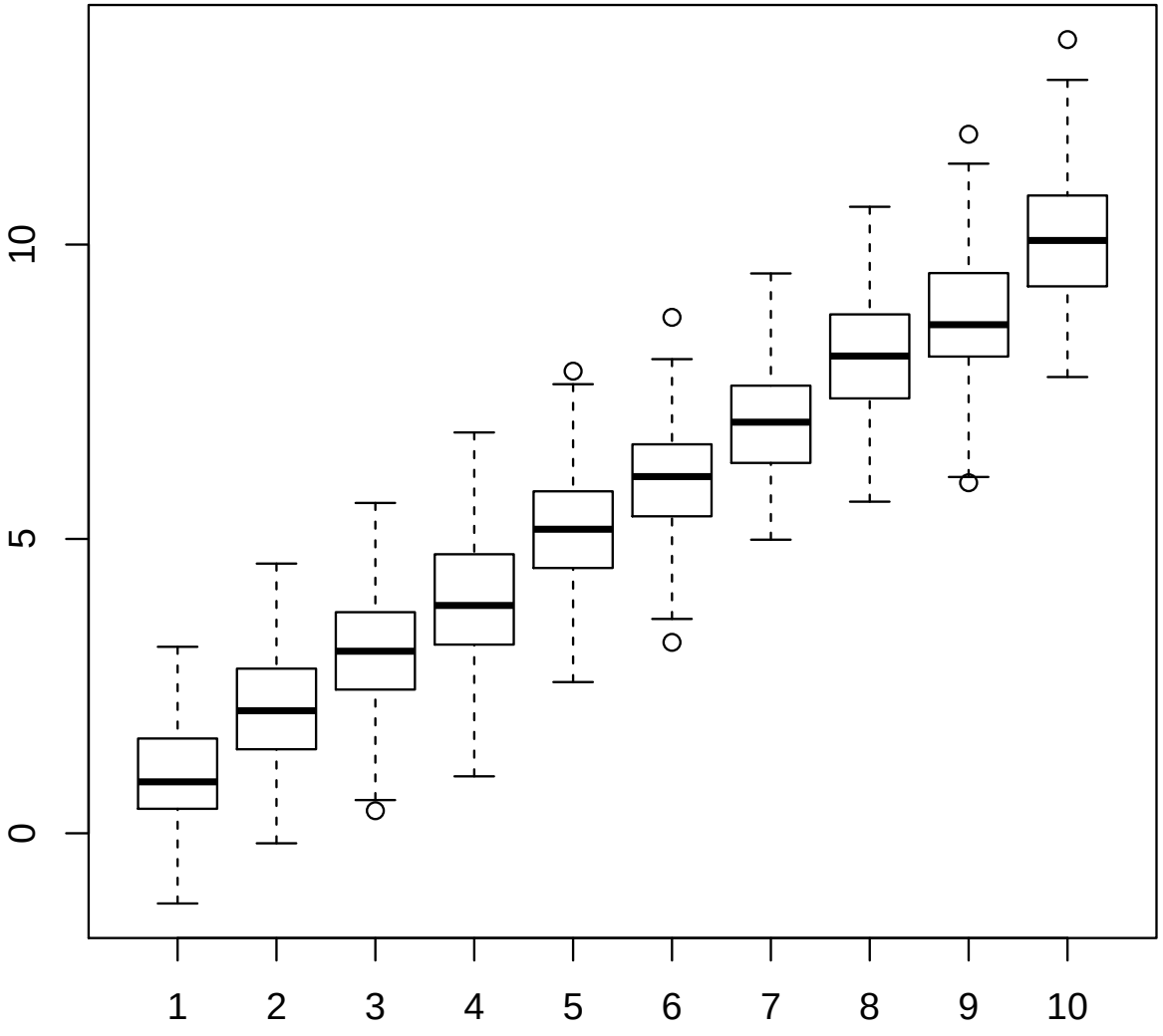
Simple Linear Regression



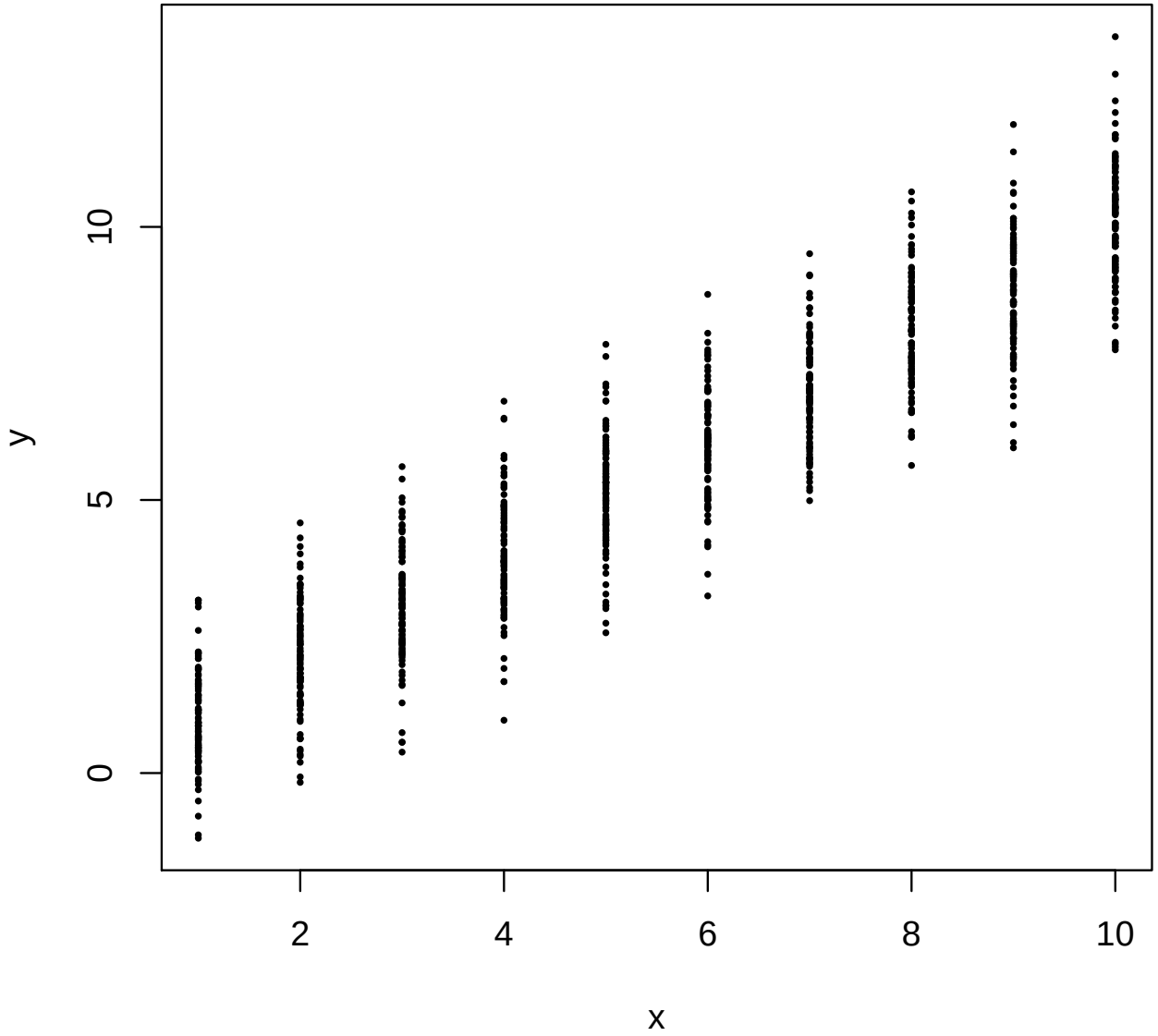
Simple Linear Regression

```
> x = rep(1:10, each =100)
> y = rnorm(1000, mean = x, sd=1)
```

Simple Linear Regression



Simple Linear Regression



Simple Linear Regression

```
> res = lm(y~x)
> res

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    0.08185      0.98977
```

Simple Linear Regression