

>data()

Gives all built in data sets

> state.x77

| | Population | Income | Illiteracy | Life Exp | Murder | HS Grad | Frost | Area |
|---------------|------------|--------|------------|----------|--------|---------|-------|--------|
| Alabama | 3615 | 3624 | 2.1 | 69.05 | 15.1 | 41.3 | 20 | 50708 |
| Alaska | 365 | 6315 | 1.5 | 69.31 | 11.3 | 66.7 | 152 | 566432 |
| Arizona | 2212 | 4530 | 1.8 | 70.55 | 7.8 | 58.1 | 15 | 113417 |
| Arkansas | 2110 | 3378 | 1.9 | 70.66 | 10.1 | 39.9 | 65 | 51945 |
| California | 21198 | 5114 | 1.1 | 71.71 | 10.3 | 62.6 | 20 | 156361 |
| Colorado | 2541 | 4884 | 0.7 | 72.06 | 6.8 | 63.9 | 166 | 103766 |
| Connecticut | 3100 | 5348 | 1.1 | 72.48 | 3.1 | 56.0 | 139 | 4862 |
| Delaware | 579 | 4809 | 0.9 | 70.06 | 6.2 | 54.6 | 103 | 1982 |
| Florida | 8277 | 4815 | 1.3 | 70.66 | 10.7 | 52.6 | 11 | 54090 |
| Georgia | 4931 | 4091 | 2.0 | 68.54 | 13.9 | 40.6 | 60 | 58073 |
| Hawaii | 868 | 4963 | 1.9 | 73.60 | 6.2 | 61.9 | 0 | 6425 |
| Idaho | 813 | 4119 | 0.6 | 71.87 | 5.3 | 59.5 | 126 | 82677 |
| Illinois | 11197 | 5107 | 0.9 | 70.14 | 10.3 | 52.6 | 127 | 55748 |
| Indiana | 5313 | 4458 | 0.7 | 70.88 | 7.1 | 52.9 | 122 | 36097 |
| Iowa | 2861 | 4628 | 0.5 | 72.56 | 2.3 | 59.0 | 140 | 55941 |
| Kansas | 2280 | 4669 | 0.6 | 72.58 | 4.5 | 59.9 | 114 | 81787 |
| Kentucky | 3387 | 3712 | 1.6 | 70.10 | 10.6 | 38.5 | 95 | 39650 |
| Louisiana | 3806 | 3545 | 2.8 | 68.76 | 13.2 | 42.2 | 12 | 44930 |
| Maine | 1058 | 3694 | 0.7 | 70.39 | 2.7 | 54.7 | 161 | 30920 |
| Maryland | 4122 | 5299 | 0.9 | 70.22 | 8.5 | 52.3 | 101 | 9891 |
| Massachusetts | 5814 | 4755 | 1.1 | 71.83 | 3.3 | 58.5 | 103 | 7826 |
| Michigan | 9111 | 4751 | 0.9 | 70.63 | 11.1 | 52.8 | 125 | 56817 |
| Minnesota | 3921 | 4675 | 0.6 | 72.96 | 2.3 | 57.6 | 160 | 79289 |
| Mississippi | 2341 | 3098 | 2.4 | 68.09 | 12.5 | 41.0 | 50 | 47296 |
| Missouri | 4767 | 4254 | 0.8 | 70.69 | 9.3 | 48.8 | 108 | 68995 |
| Montana | 746 | 4347 | 0.6 | 70.56 | 5.0 | 59.2 | 155 | 145587 |
| Nebraska | 1544 | 4508 | 0.6 | 72.60 | 2.9 | 59.3 | 139 | 76483 |
| Nevada | 590 | 5149 | 0.5 | 69.03 | 11.5 | 65.2 | 188 | 109889 |
| New Hampshire | 812 | 4281 | 0.7 | 71.23 | 3.3 | 57.6 | 174 | 9027 |
| New Jersey | 7333 | 5237 | 1.1 | 70.93 | 5.2 | 52.5 | 115 | 7521 |

| | | | | | | | | |
|----------------|-------|------|-----|-------|------|------|-----|--------|
| New Mexico | 1144 | 3601 | 2.2 | 70.32 | 9.7 | 55.2 | 120 | 121412 |
| New York | 18076 | 4903 | 1.4 | 70.55 | 10.9 | 52.7 | 82 | 47831 |
| North Carolina | 5441 | 3875 | 1.8 | 69.21 | 11.1 | 38.5 | 80 | 48798 |
| North Dakota | 637 | 5087 | 0.8 | 72.78 | 1.4 | 50.3 | 186 | 69273 |
| Ohio | 10735 | 4561 | 0.8 | 70.82 | 7.4 | 53.2 | 124 | 40975 |
| Oklahoma | 2715 | 3983 | 1.1 | 71.42 | 6.4 | 51.6 | 82 | 68782 |
| Oregon | 2284 | 4660 | 0.6 | 72.13 | 4.2 | 60.0 | 44 | 96184 |
| Pennsylvania | 11860 | 4449 | 1.0 | 70.43 | 6.1 | 50.2 | 126 | 44966 |
| Rhode Island | 931 | 4558 | 1.3 | 71.90 | 2.4 | 46.4 | 127 | 1049 |
| South Carolina | 2816 | 3635 | 2.3 | 67.96 | 11.6 | 37.8 | 65 | 30225 |
| South Dakota | 681 | 4167 | 0.5 | 72.08 | 1.7 | 53.3 | 172 | 75955 |
| Tennessee | 4173 | 3821 | 1.7 | 70.11 | 11.0 | 41.8 | 70 | 41328 |
| Texas | 12237 | 4188 | 2.2 | 70.90 | 12.2 | 47.4 | 35 | 262134 |
| Utah | 1203 | 4022 | 0.6 | 72.90 | 4.5 | 67.3 | 137 | 82096 |
| Vermont | 472 | 3907 | 0.6 | 71.64 | 5.5 | 57.1 | 168 | 9267 |
| Virginia | 4981 | 4701 | 1.4 | 70.08 | 9.5 | 47.8 | 85 | 39780 |
| Washington | 3559 | 4864 | 0.6 | 71.72 | 4.3 | 63.5 | 32 | 66570 |
| West Virginia | 1799 | 3617 | 1.4 | 69.48 | 6.7 | 41.6 | 100 | 24070 |
| Wisconsin | 4589 | 4468 | 0.7 | 72.48 | 3.0 | 54.5 | 149 | 54464 |

```
> coef<-cor(state.x77,method="pearson")
```

```
> coef
```

```
> coef<-cor(state.x77,method="pearson")
> coef
```

| | Population | Income | Illiteracy | Life Exp | Murder | HS Grad | Frost | Area |
|------------|-------------|------------|-------------|-------------|------------|-------------|------------|-------------|
| Population | 1.00000000 | 0.2082276 | 0.10762237 | -0.06805195 | 0.3436428 | -0.09848975 | -0.3321525 | 0.02254384 |
| Income | 0.20822756 | 1.0000000 | -0.43707519 | 0.34025534 | -0.2300776 | 0.61993232 | 0.2262822 | 0.36331544 |
| Illiteracy | 0.10762237 | -0.4370752 | 1.0000000 | -0.58847793 | 0.7029752 | -0.65718861 | -0.6719470 | 0.07726113 |
| Life Exp | -0.06805195 | 0.3402553 | -0.58847793 | 1.0000000 | -0.7808458 | 0.58221620 | 0.2620680 | -0.10733194 |
| Murder | 0.34364275 | -0.2300776 | 0.70297520 | -0.78084575 | 1.0000000 | -0.48797102 | -0.5388834 | 0.22839021 |
| HS Grad | -0.09848975 | 0.6199323 | -0.65718861 | 0.58221620 | -0.4879710 | 1.0000000 | 0.3667797 | 0.33354187 |
| Frost | -0.33215245 | 0.2262822 | -0.67194697 | 0.26206801 | -0.5388834 | 0.36677970 | 1.0000000 | 0.05922910 |
| Area | 0.02254384 | 0.3633154 | 0.07726113 | -0.10733194 | 0.2283902 | 0.33354187 | 0.0592291 | 1.0000000 |

```
> |
```

```
> states<-state.x77
```

```
> x<-states[,c("Illiteracy")]
```

```
> y<-states[,c("Murder")]
```

```
> coef<-cor(x,y,method="pearson")
```

```
> cat("Pearson correlation between Illiteracy and Murder:",coef)
```

Pearson correlation between Illiteracy and Murder: 0.7029752

Highest correlation in the table is between Murder and Illiteracy: 0.7029752

There is highly positive correlation between murder and illiteracy so we can fit a regression model on that.

```
> coefs<-cor(state.x77,method="spearman")
```

```
> coefs
```

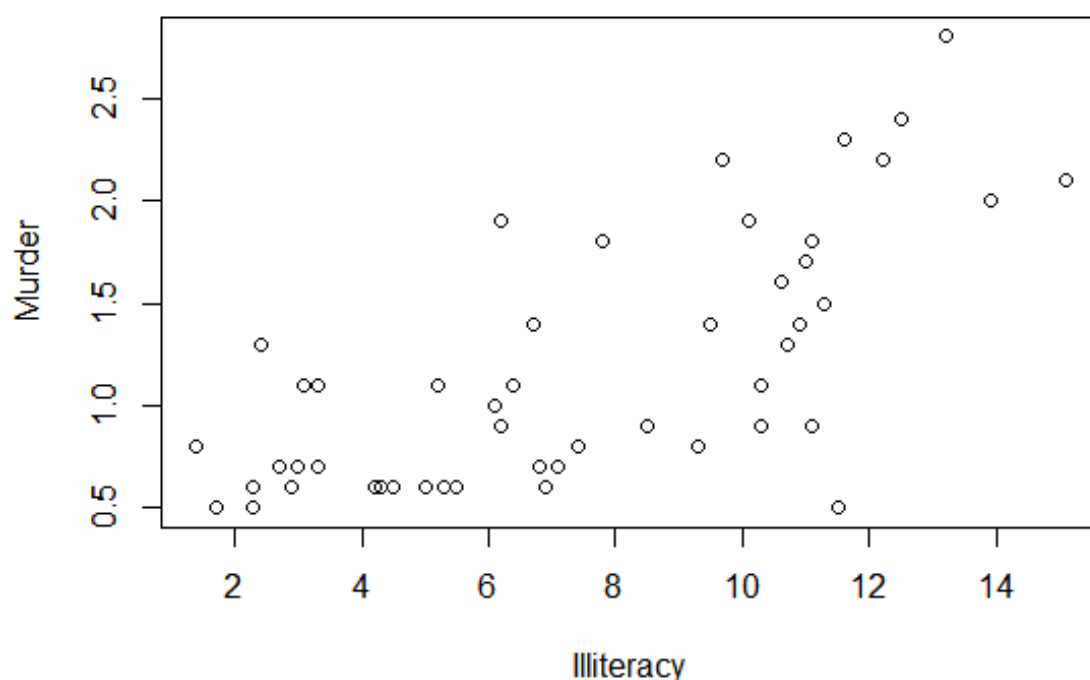
| | Population | Income | Illiteracy | Life Exp | Murder | HS Grad | Frost | Area |
|------------|------------|-------------|------------|------------|------------|------------|------------|-------------|
| Population | 1.0000000 | 0.12460984 | 0.3130496 | -0.1040171 | 0.3457401 | -0.3833649 | -0.4588526 | -0.12067227 |
| Income | 0.1246098 | 1.00000000 | -0.3145948 | 0.3241050 | -0.2174623 | 0.5104809 | 0.1968638 | 0.05709484 |
| Illiteracy | 0.3130496 | -0.31459482 | 1.00000000 | -0.5553735 | 0.6723592 | -0.6545396 | -0.6831936 | -0.25037208 |
| Life Exp | -0.1040171 | 0.32410498 | -0.5553735 | 1.00000000 | -0.7802406 | 0.5239410 | 0.2983910 | 0.12750018 |
| Murder | 0.3457401 | -0.21746230 | 0.6723592 | -0.7802406 | 1.00000000 | -0.4367330 | -0.5438432 | 0.10642590 |
| HS Grad | -0.3833649 | 0.51048095 | -0.6545396 | 0.5239410 | -0.4367330 | 1.00000000 | 0.3985351 | 0.43897520 |
| Frost | -0.4588526 | 0.19686382 | -0.6831936 | 0.2983910 | -0.5438432 | 0.3985351 | 1.00000000 | 0.11228778 |
| Area | -0.1206723 | 0.05709484 | -0.2503721 | 0.1275002 | 0.1064259 | 0.4389752 | 0.1122878 | 1.00000000 |

Highest correlation in the table is between Murder and Illiteracy: 0.6723592

```
> coefsp<-cor(x,y,method="spearman")
```

```
> cat("Spearman correlation between Illiteracy and Murder:",coefsp)
```

Spearman correlation between Illiteracy and Murder: 0.6723592



```
> plot(y,x,xlab="Illiteracy",ylab="Murder")
```

```
> model=lm(Murder~Illiteracy,data=data.frame(state.x77))
```

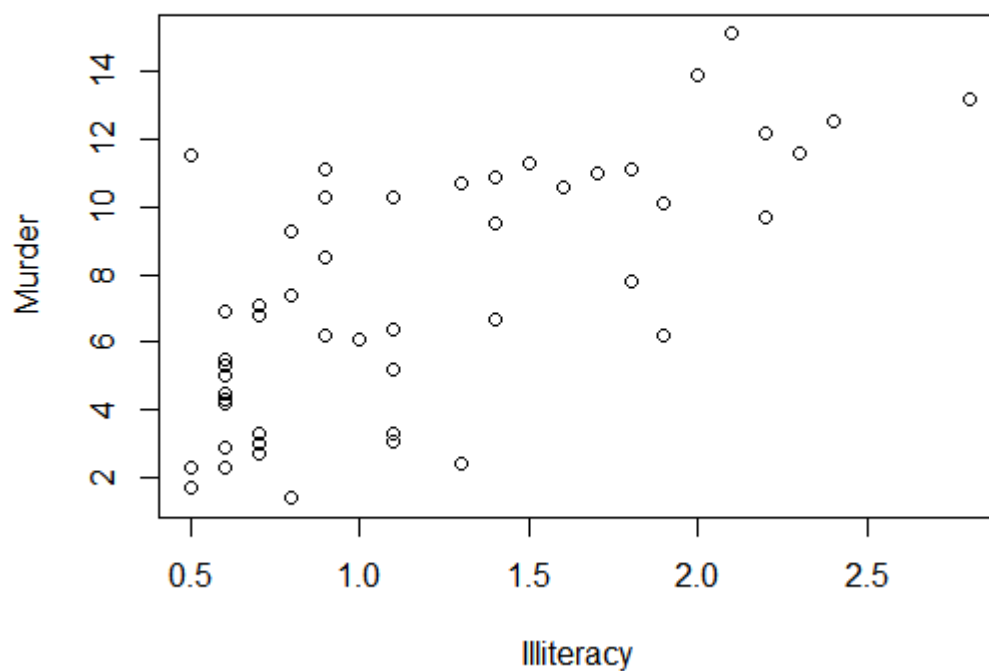
Here Murder is dependent variable and Illiteracy is independent variable

```
> coeffs=coefficients(model)
```

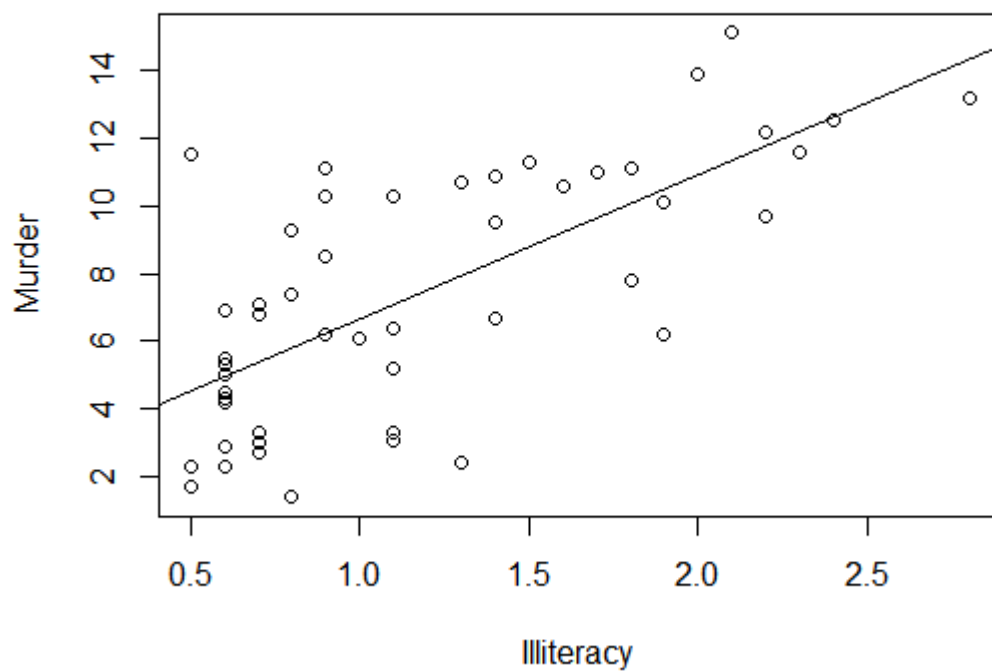
```
> coeffs
```

(Intercept) Illiteracy

2.396776 4.257457



```
> abline(lm(Murder~Illiteracy,data=data.frame(state.x77)))
```



```
> test_data=data.frame(Illiteracy=c(2.5,8.3,12.5))
```

```
> df<-predict(model,test_data)
```

```
> names(df)<-c(2.5,8.3,12.5)
```

```
> df
```

```
2.5 8.3 12.5
```

```
13.04042 37.73367 55.61498
```

(for 12.5 illiteracy murder rate is 55.61498)

```
> model1<-lm(Murder~Illiteracy+Area,data=data.frame(state.x77))
```

```
> coeffs1=coefficients(model1)
```

```
> coeffs1
```

```
(Intercept) Illiteracy Area
```

```
1.956727e+00 4.175513e+00 7.576393e-06
```

```
> summary(model1)
```

Call:

```
lm(formula = Murder ~ Illiteracy + Area, data = data.frame(state.x77))
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|--------|--------|--------|-------|-------|
| -4.993 | -2.018 | -0.469 | 1.692 | 6.623 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | 1.957e+00 | 8.409e-01 | 2.327 | 0.0243 * |
| Illiteracy | 4.176e+00 | 6.109e-01 | 6.835 | 1.45e-08 *** |
| Area | 7.576e-06 | 4.364e-06 | 1.736 | 0.0891 . |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.599 on 47 degrees of freedom

Multiple R-squared: 0.5247, Adjusted R-squared: 0.5044

F-statistic: 25.94 on 2 and 47 DF, p-value: 2.568e-08

```
> model2=lm(Murder~Illiteracy+Life.Exp,data=data.frame(state.x77))
```

```
> coef2=coefficients(model2)
```

```
> coef2
```

| (Intercept) | Illiteracy | Life.Exp |
|-------------|------------|-----------|
| 114.216241 | 2.255651 | -1.544576 |

```
> model3=lm(Murder~Illiteracy+Population+Area,data=data.frame(state.x77))
```

```
> coef3=coefficients(model3)
```

```
> coef3
```

```
(Intercept) Illiteracy Population      Area
```

```
1.227941e+00 4.002200e+00 2.221474e-04 7.410016e-06
```