



Pearson Correlation

QAC 201

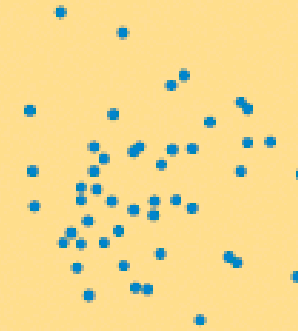
“ r ” ranges from -1 to $+1$

“ r ” quantifies the strength and direction of a linear relationship between two quantitative variables.

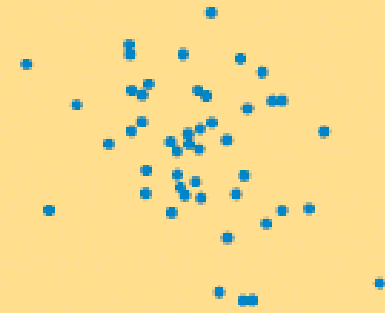
Strength: How closely the points follow a straight line.

Direction is positive when individuals with higher x values tend to have higher values of y .

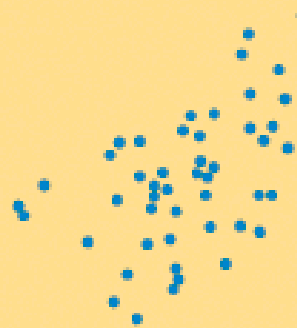
Significance: Is the correlation coefficient significantly different from 0? This requires a statistical test to determine.



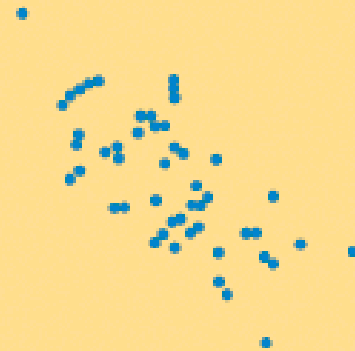
Correlation $r = 0$



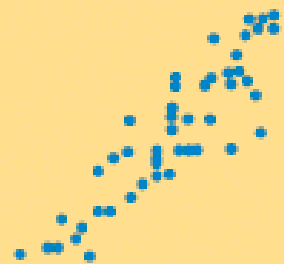
Correlation $r = -0.3$



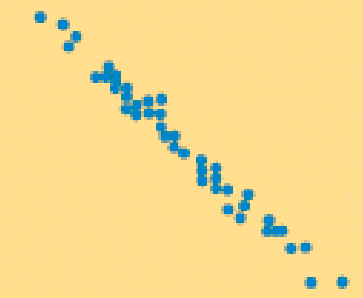
Correlation $r = 0.5$



Correlation $r = -0.7$

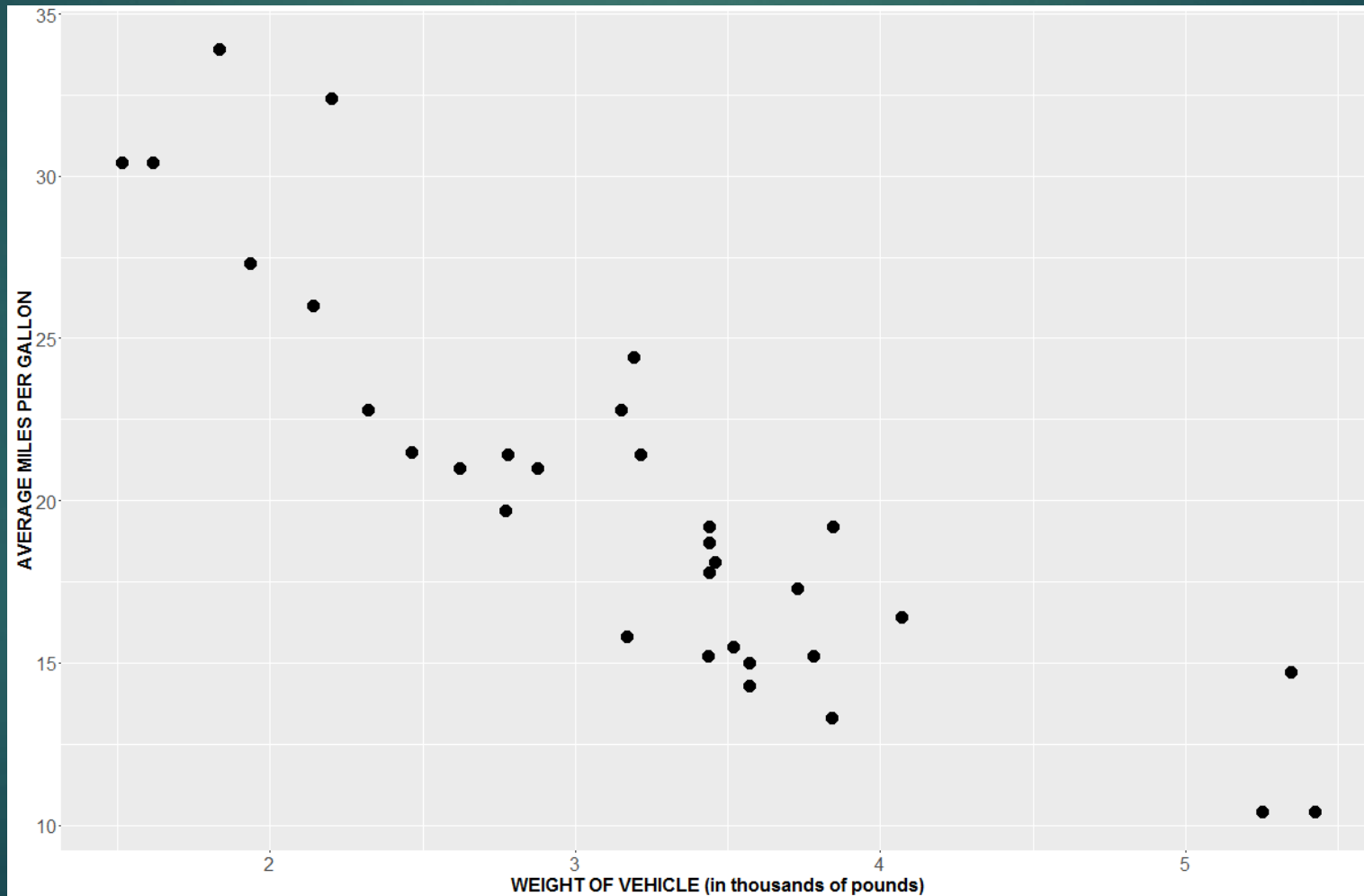


Correlation $r = 0.9$



Correlation $r = -0.99$

Suppose we want to look at the relationship between mpg of a vehicle and the corresponding weight of a vehicle. Ultimately we would like to see whether weight can help us predict mpg of a car. Weight is in thousands of pounds.



Is there a significant linear relationship between weight and mpg?

One way to test this is with the Pearson Correlation test.

R OUTPUT:

```
Pearson's product-moment correlation
data:  mtcars$wt and mtcars$mpg
t = -9.559, df = 30, p-value = 1.294e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9338264 -0.7440872
sample estimates:
      cor
-0.8676594
```

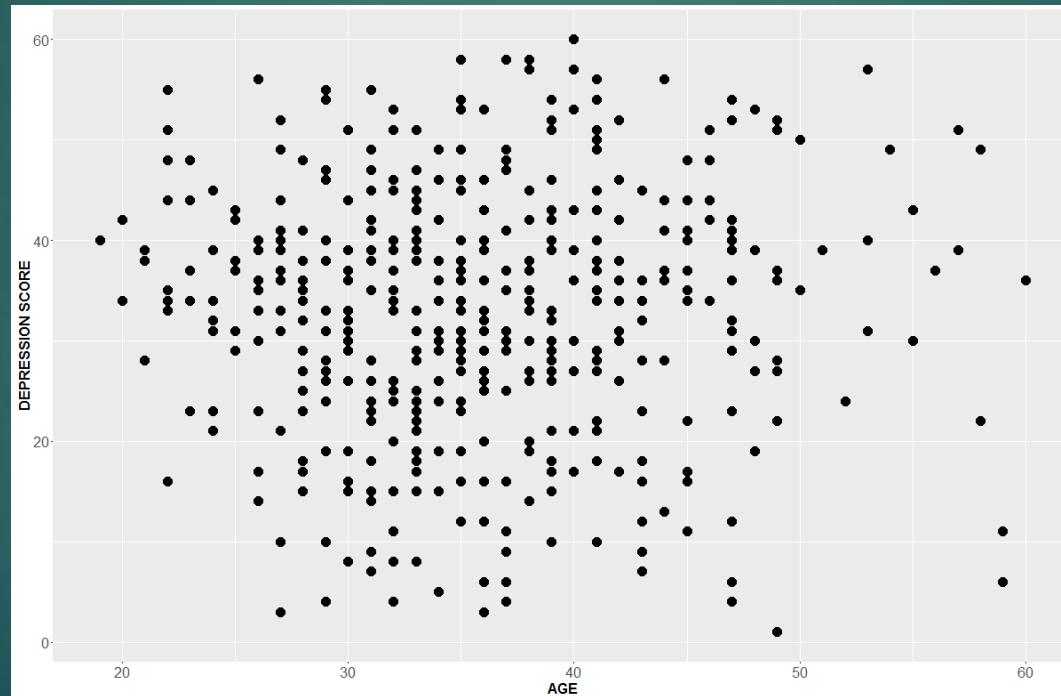
H0: There is no linear association between weight of vehicle and MPG of vehicle

HA: There is a linear association between weight of vehicle and MPG of vehicle

Stata OUTPUT:

	mpg	wt
mpg	1.0000	
wt	-0.8677 0.0000	1.0000

Suppose we were interested in the linear relationship between age and depression scores.



Pearson's product-moment correlation

```
data: HELPrct$age and HELPrct$cesd
t = 0.17778, df = 451, p-value = 0.859
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.08382526  0.10042509
sample estimates:
      cor
0.008370962
```

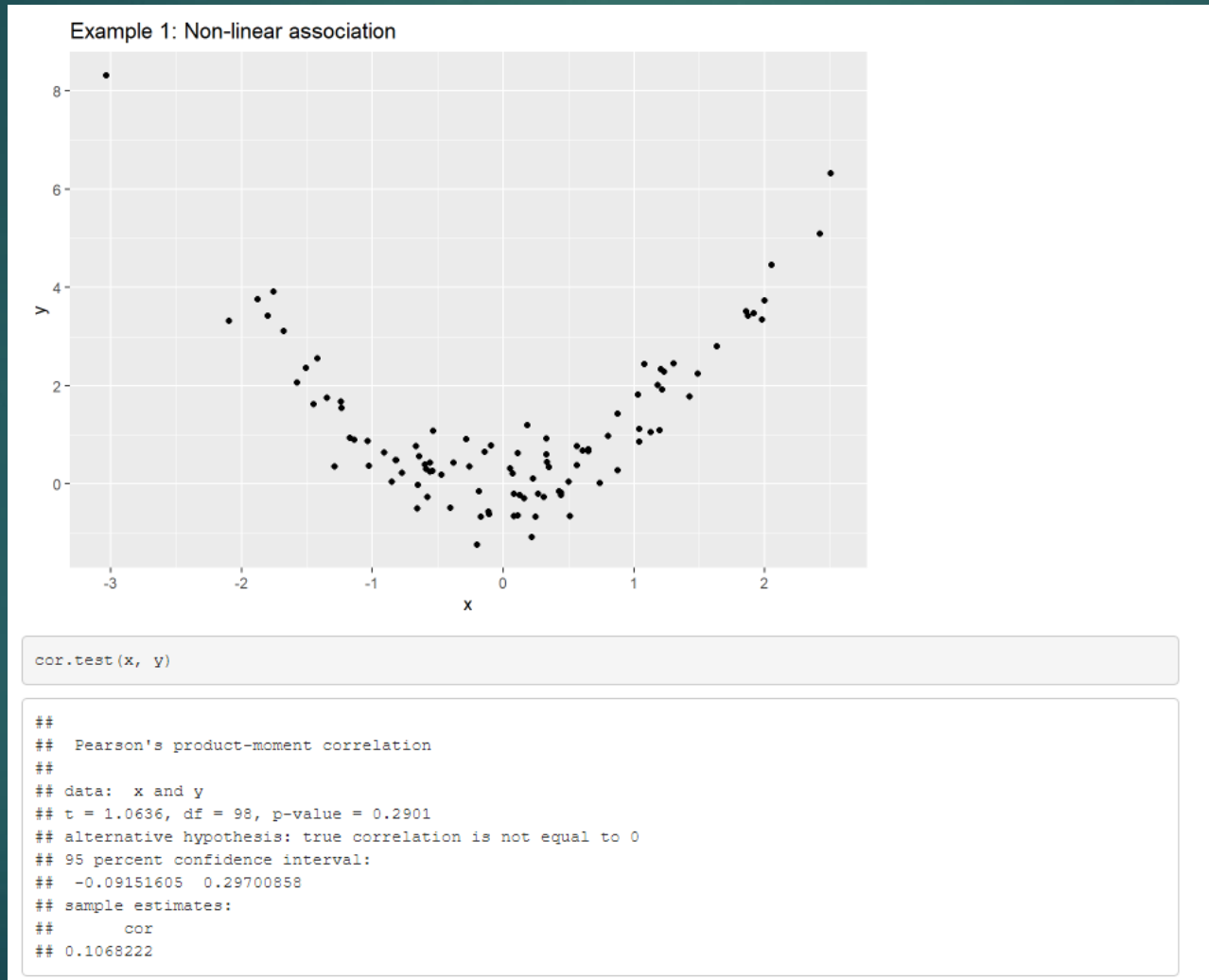
	cesd	age
cesd	1.0000	
age	0.0084 0.8590	1.0000

H0: There is no linear association between age and depression

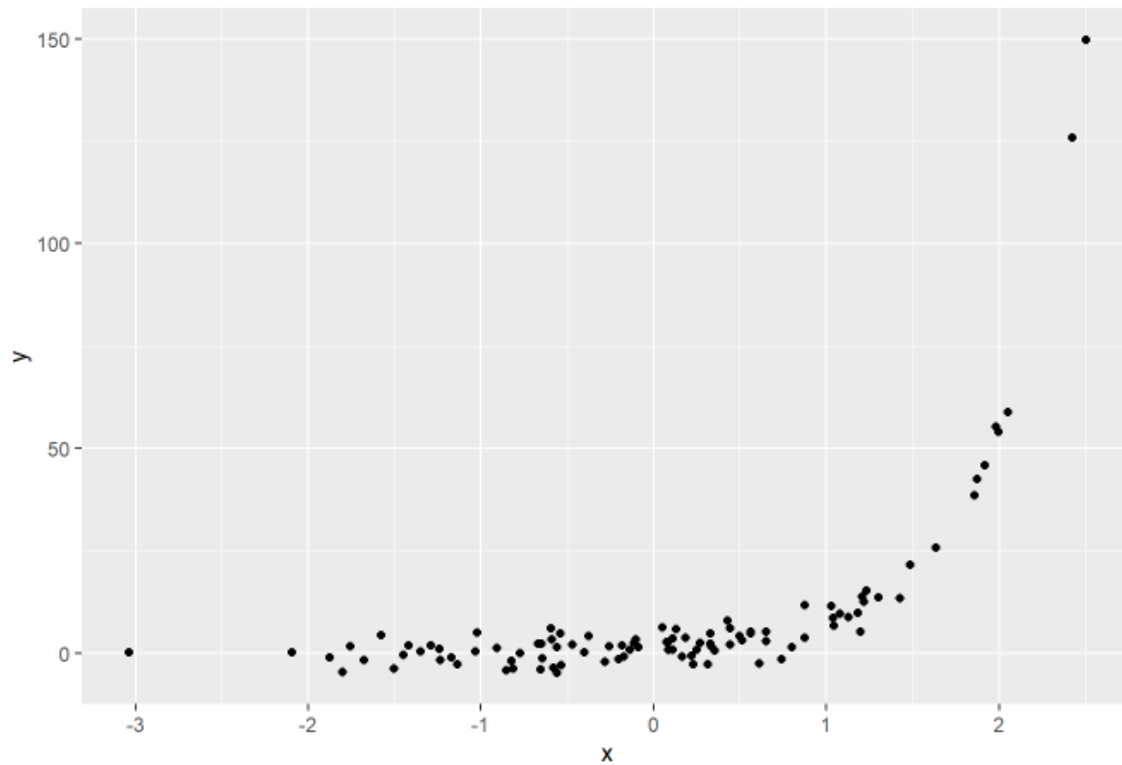
HA: There is a linear association between age and depression

Pearson correlation only tests the degree of the linear association between two variables.

If two variables have a non-linear association, Pearson correlation will not adequately describe the strength of the relationship.



Example 2: Non-linear association



```
cor.test(x, y)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: x and y  
## t = 7.7489, df = 98, p-value = 8.676e-12  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4778092 0.7250210  
## sample estimates:  
## cor  
## 0.6163798
```