BACS-hw09-107070004

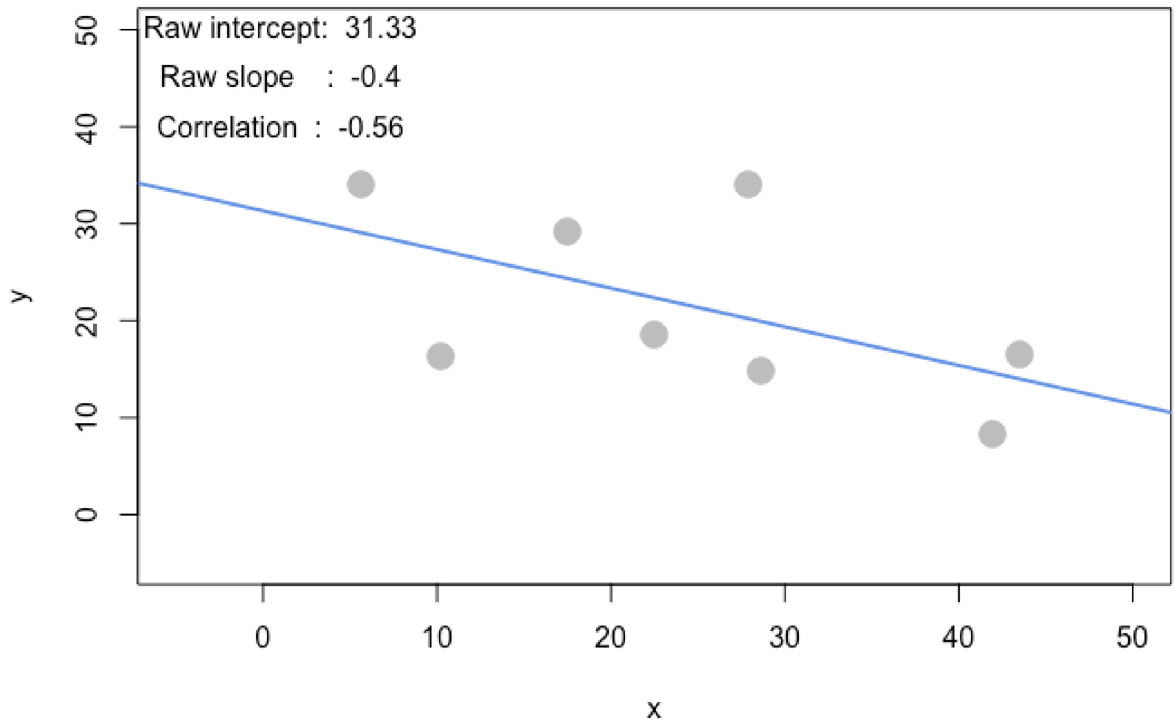
(e) Apart from any of the above scenarios, find another pattern of data points with no correlation (r ~= 0). (can create a pattern that visually suggests a strong relationship but produces r ~= 0?)

(f) Apart from any of the above scenarios, find another pattern of data points with perfect correlation (r ~= 1). (can you find a scenario where the pattern visually suggests a different relationship?)

(g) Let's see how correlation relates to simple regression, by simulating any linear relationship you wish:

(i) Run the simulation and record the points you create: pts <- interactive_regression() (simulate either a positive or negative relationship)

```
source("demo_simple_regression.R")
pts <- interactive_regression()
pts
```



	x	y
1	5.625473	34.049378
2	27.889029	34.049378
3	22.479193	18.563136
4	28.617276	14.831511
5	41.933796	8.301168
6	43.494325	16.510743
7	17.485498	29.198266
8	10.203026	16.324161

(ii) Use the lm() function to estimate the regression intercept and slope of pts to ensure they are the same as the values reported in the simulation plot: summary(lm(pts$y\ pts$x))

```
summary( lm( pts$y ~ pts$x ))
```

```
Call:
lm(formula = pts$y ~ pts$x)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.9373  -5.3984  -0.6459   4.8700  13.8353 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  31.3271     6.6344   4.722  0.00325 **
pts$x        -0.3985     0.2384  -1.671  0.14572   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.619 on 6 degrees of freedom
Multiple R-squared:  0.3176,	Adjusted R-squared:  0.2039 
F-statistic: 2.793 on 1 and 6 DF,  p-value: 0.1457
```

(iii) Estimate the correlation of x and y to see it is the same as reported in the plot: cor(pts)

```
cor(pts)
```

```
           x          y
x  1.0000000 -0.5635847
y -0.5635847  1.0000000
```

(iv) Now, standardize the values of both x and y from pts and re-estimate the regression slope

```
x <- scale(pts$x,center=TRUE,scale=TRUE)
y <- scale(pts$y,center=TRUE,scale=TRUE)
summary( lm( y ~ x ))
```

```
Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.13222 -0.55884 -0.06686  0.50414  1.43221 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.128e-16  3.155e-01   0.000    1.000
x           -5.636e-01  3.372e-01  -1.671    0.146

Residual standard error: 0.8922 on 6 degrees of freedom
Multiple R-squared:  0.3176,	Adjusted R-squared:  0.2039 
F-statistic: 2.793 on 1 and 6 DF,  p-value: 0.1457
```

(v) What is the relationship between correlation and the standardized simple-regression estimates?

Between the stadardized and non-stadardized data, they have same R-squared and f-statistic. However, the stadard error is different.