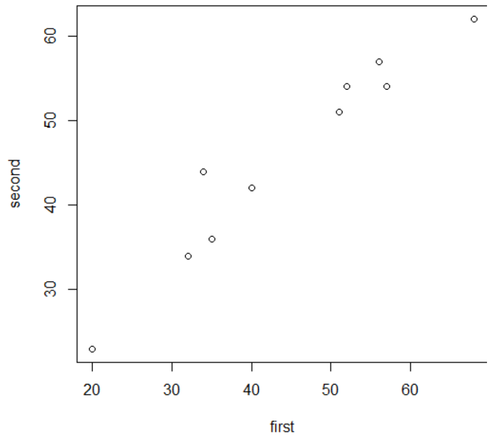
```
> first=c(20,32,35,34,40,51,52,56,57,68)
> second=c(23,34,36,44,42,51,54,57,54,62)
> #a. Display the relationship between first and second visit dollar amounts?
> plot(first,second)
>
> #b. Describe the pattern in part (a) briefly. Is there a relationship? Is it positive or negative? Is it linear or non-linear? Is it weak or strong?
> #It seems like a positive linear relationship.
> #c. Calculate the correlation coefficient between the amount of money spent on the first visit and the second visit.
> cor(first,second)
[1] 0.9690331
> cor.test(first,second)

        Pearson's product-moment correlation

data:  first and second
t = 11.0996, df = 8, p-value = 3.876e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8705612 0.9928768
sample estimates:
      cor
0.9690331

> r=0.9690331
> #d. What does the standard error in part (c) refer to?
> se=sqrt((1-r^2)/(length(first)-2))
> se
[1] 0.08730324
> #0.08730325
> #e. Calculate an approximate 95% confidence interval for ρ.
> #cor.test(first,second)
> #95 percent confidence interval:
> #0.8705612 0.9928768
> #sample estimates:
>         #cor
> #0.9690331
> #2. Answer the following question using the data from question (2).
> #a. Adding $30 to each of the observations for the second visit. How is the correlation coefficient between first and second visits affected? What can you
```

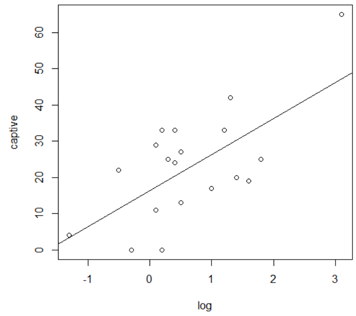
conclude about the effects on the correlation coefficient of adding a constant to one or both of the variables?

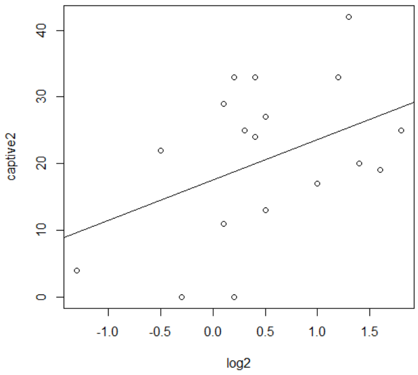
```
> second2=second+30
> cor(first,second2)
[1] 0.9690331
> #0.9690331
> #it's correlation coefficient does not change. Because correlation coefficient is a standardized value.
> #b. Convert the first visit to cents (i.e., multiply by 100). How does this affect the correlation between the first and second visits? What can you conclude about the effects on the correlation coefficient of multiplying one or both of the variables by a constant?
> first2=100*first
> cor(first2,second)
[1] 0.9690331
> #it's correlation coefficient does not change. Because correlation coefficient is a standardized value.
>
> #3. Some species seem to thrive in captivity, whereas others are prone to health and behavior difficulties when caged. Maternal care problems in some captive species, for example, lead to high infant mortality. Can these differences be predicted? The following data are measurements of the infant mortality (percentage of births) of 20 carnivore species in captivity along with the log (based-10) of the minimal home range sizes (in km2) of the same species in the wild (Clubb and Mason 2003). For example, -1.3 is the home range and 4 is the captive infant mortality percentage.
> #Log home range size: -1.3 (4),-0.5 (22),-0.3 (0),0.2 (0),0.1 (11),0.5 (13),1.0 (17),0.3 (25),0.4 (24),0.5 (27),0.1 (29),0.2 (33),0.4 (33),1.3 (42),1.2 (33),1.4 (20),1.6 (19),1.6 (19),1.8 (25),3.1 (65)
> #a. Draw a scatter plot of these data, with log of home range size as the explanatory variable. Describe the association between the two variables in words.
> log=c(-1.3,-0.5,-0.3,0.2,0.1,0.5,1.0,0.3,0.4,0.5,0.1,0.2,0.4,1.3,1.2,1.4,1.6,1.6,1.8,3.1)
> captive =c(4,22,0,0,11,13,17,25,24,27,29,33,33,42,33,20,19,19,25,65)
> plot(log,captive)
>
> #It seems like a positive linear relationship.
> #b. Estimate the slope and intercept of the least squares regression line, with the log of home range size as the explanatory variable. Add this line to your plot.
> esti=lm(captive~log)
> abline(esti)
> summary(esti)
```

```
Call:
lm(formula = captive ~ log)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.271  -9.481   2.199  10.954  17.858 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   16.280      3.219   5.057  8.2e-05 ***
log            9.955      2.766   3.600  0.00205 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.69 on 18 degrees of freedom
Multiple R-squared:  0.4186,	Adjusted R-squared:  0.3863 
F-statistic: 12.96 on 1 and 18 DF,  p-value: 0.002049

> #slope=9.955 intercept=16.280
> #c. Does home range size in the wild predict the mortality of captive carnivores? 
Carry out a formal test. Assume that the species data are independent.
> bptest(esti)

	studentized Breusch-Pagan test

data:  esti
BP = 2.7642, df = 1, p-value = 0.09639

> #homoscedasticity is ok
> shapiro.test(esti$residuals)

	Shapiro-Wilk normality test

data:  esti$residuals
W = 0.9201, p-value = 0.09969

> #normality is ok
> dwtest(esti)

	Durbin-Watson test
```

```
data:  esti
DW = 0.8856, p-value = 0.001185
alternative hypothesis: true autocorrelation is greater than 0

> #independence is not ok.
> #But independent is already assumed, so this model can be used.
> #and p-value is 0.002049
> #d. Outliers should be investigated because they might have a substantial effect on the estimate so of the slope
> #20th obs is outlier. So, i remove it.
> log2=c(-1.3,-0.5,-0.3,0.2,0.1,0.5,1.0,0.3,0.4,0.5,0.1,0.2,0.4,1.3,1.2,1.4,1.6,1.6,1.8)
> captive2 =c(4,22,0,0,11,13,17,25,24,27,29,33,33,42,33,20,19,19,25)
> esti2=lm(captive2~log2)
> summary(esti2)

Call:
lm(formula = captive2 ~ log2)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.704  -7.311  -3.405   7.886  16.627 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   17.492      3.002   5.827 2.02e-05 ***
log2           6.063      3.129   1.938   0.0694 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.7 on 17 degrees of freedom
Multiple R-squared:  0.1809,    Adjusted R-squared:  0.1327 
F-statistic: 3.755 on 1 and 17 DF,  p-value: 0.06944

> plot(log2,captive2)
> abline(esti2)
> #slope=6.063 #intercept=17.492
> #slope is more slowly increasing than include outlier.
```


second
first
20
30
40
50
60


captive
log


captive2
log2