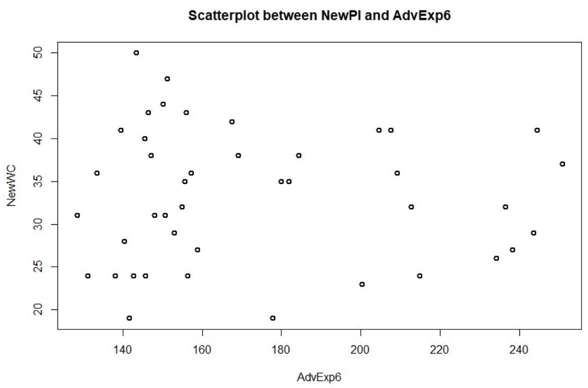
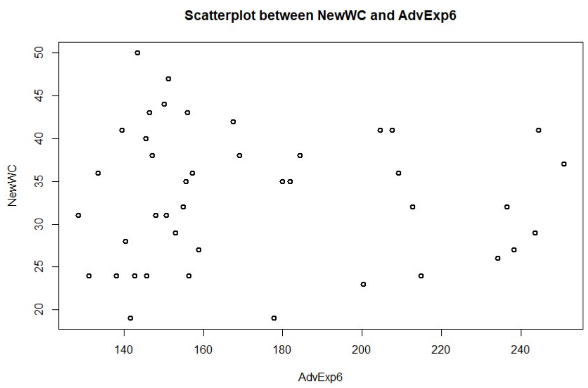
Question 1:

```
#Line of code below, is to include only non-NA values
LegalAdv = LegalAdv[!is.na(LegalAdv$AdvExp6), ]
```

```
plot(NewPI~AdvExp6, data=LegalAdv,
main="Scatterplot between NewPI
and AdvExp6", lwd=2)
```




```
plot(NewWC~AdvExp6, data=LegalAdv,
main="Scatterplot between NewWC
and AdvExp6", lwd=2)
```



Question 2:

The first plot shows AdvExp6 on x-axis and NewPI on y-axis ranging from 0 to 260, and 0 to 50 respectively. From the plot we can see that NewPI vs AdvExp6 the data is scattered little bit more in between \$0 to \$160. From \$160 onwards there are comparatively less data points. This could indicate that on further increase of expenditure on advertising crossing the \$160 threshold, there may not be a significant rise in PI cases.

The second plot shows AdvExp6 on x-axis and NewWC on y-axis ranging from 0 to 260, and 0 to 50 respectively. From the plot we can see that NewWC vs AdvExp6 the data is scattered little bit more in between \$0 to \$170. From \$170 onwards there are comparatively less data points. This could indicate that on further increase of expenditure on advertising crossing the \$170 threshold, there may not be a 40-unit rise in WC cases.

Question 3:

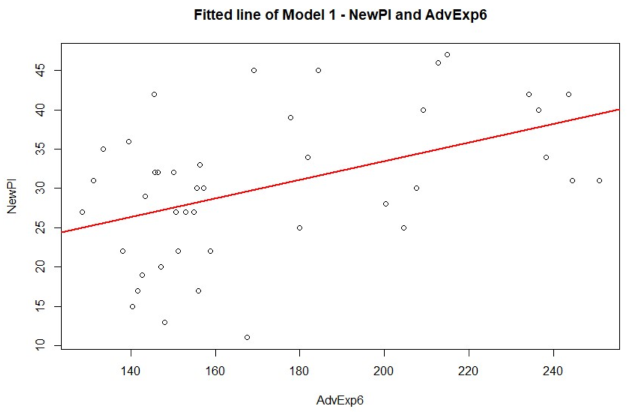
```
model1=lm(NewPI~AdvExp6,
data=LegalAdv)

> model1

Call:
lm(formula = NewPI ~ AdvExp6, dat
a = LegalAdv)

Coefficients:
(Intercept)      AdvExp6
     9.8214       0.1182

plot(NewPI~AdvExp6, main="Fitted
line of Model 1 - NewPI and AdvEx
p6", data=LegalAdv)
abline(model1, col="red", lwd=2)
```



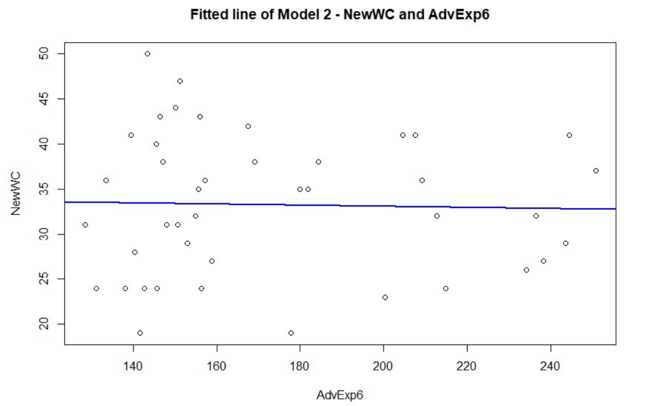

```
model2=lm(NewWC~AdvExp6,
data=LegalAdv

> model2

Call:
lm(formula = NewWC ~ AdvExp6, dat
a = LegalAdv)

Coefficients:
(Intercept)      AdvExp6
  34.349974    -0.006284

plot(NewWC~AdvExp6, main="Fitted
line of Model 2 - NewWC and AdvEx
p6", data=LegalAdv)
abline(model2, col="blue", lwd=2)
```



Consider the model1 as NewPI = $\beta_0 + \beta_1$AdvExp6 + ϵ where ϵ is a normal random variable.
After fitting, the estimates are $\widehat{\beta 0}$ = 9.8214, $\widehat{\beta 1}$ = 0.1182 and the fitted line is:
$\widehat{NewPI}$ = 9.8214 + 0.1182AdvExp6

Consider the model2 as NewWC = $\beta'_0 + \beta'_1$AdvExp6 + ϵ' where ϵ is a normal random var
After fitting, the estimates are $\widehat{\beta' 0}$ = 34.349974, $\widehat{\beta' 1}$ = -0.006284 and the fitted line is:
$\widehat{NewWC}$ = 34.349974 - 0.006284AdvExp6

Question 4:

From the above fitted lines it can be inferred as follows:

AdvExp6 has a positive impact on NewPI. This is due to the postive value of $\widehat{\beta 1}$ = 0.1182. This indicates a positive relationship between the two. Therefore on increase of $1000 AdvExp, the NewWC increases by $8 approximately.

AdvExp6 has a negative impact on NewWC. This is due to the negative value of $\widehat{\beta' 1}$ = -0.006284. This indicates a negative relationship between the two. Therefore, an increase of $1000 AdvExp, the NewWC decreases by $6 approximately.

Question 5:

We do not want to print the entire summary of the model. So, we use the below code to print only the coefficients. We are interested in the t-test and getting the p-values for determining statistical significance between AdvExp6 and the output variables – NewPI and NewWC

```
> summary(model1)$coefficient
             Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 9.8213652 6.30502661 1.557704 0.127181173
AdvExp6     0.1181839 0.03564751 3.315349 0.001953163
> summary(model2)$coefficient
               Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 34.349974128 5.96916985  5.7545647 1.045975e-06
AdvExp6     -0.006283724 0.03374864 -0.1861919 8.532353e-01
```

Question 6:

From the test outcomes of question 5
H_0: $\beta_1 = 0$ vs H_1: $\beta_1 \neq 0$,
In this test we are looking at testing whether AdvExp6(β_1) is contributing towards a change in NewPI.
We can look the *p*-value from the output.
The *p*-value is **0.001953163** < 0.05 = α, so the model1 **is** statistically significant at the level 5%.
Looking at the t-value we can see that 3.315349 > 0.05 so the model1 **is** statistically significant at the level 5%.

Also, H_0: $\beta'_1 = 0$ vs H_1: $\beta'_1 \neq 0$,
In this test we are looking at testing whether AdvExp6(β'_1)is contributing towards a change in NewWC.
We can look the *p*-value from the output.
The *p*-value is **8.532353e-01**> 0.05 = α, so the model2 is **NOT** statistically significant at the level 5%
Looking at the t-value we can see that -0.1861919 < 0.05 so the model1 **is NOT** statistically significant at the level 5%.

Therefore, it can be concluded that only NewPI is statistically linearly related to cumulative 6-month advertising expenditures,

Question 7:

```
> confint(model1, level=0.95)
                  2.5 %     97.5 %
(Intercept) -2.92156892 22.5642993
AdvExp6      0.04613762  0.1902302
```

For Model1, the 95% CI of β_1 is given by [0.04613762, 0.1902302]. 0 is **excluded.**

```
> confint(model2, level=0.95)
                  2.5 %      97.5 %
(Intercept) 22.28583185 46.41411641
AdvExp6     -0.07449226  0.06192481
```

For Model2, the 95% CI of β'_1 is given by [-0.07449226 0.06192481]. 0 is **included.**

Question 8:

```
> predict(model1, newdata=data.frame(AdvExp6=170),interval="co
nfidence" )
       fit     lwr      upr
1 29.91263 27.3048 32.52047
```

Upon entering 170 into the 95% CI for the AdvExp6 variable the fitted value is 29.91263. Therefore, it can be inferred that upon spending 170$ into advertising, the predicted value of **NewPI** is $29 approximately with the 95% CI being [27.3048 , 32.52047].

```
> predict(model2, newdata=data.frame(AdvExp6=170),interval="co
nfidence" )
       fit      lwr      upr
1 33.28174 30.81282 35.75067
```

Upon entering 170 into the 95% CI for the AdvExp6 variable the fitted value is 33.28174. Therefore, it can be inferred that upon spending 170$ into advertising, the predicted value of **NewWC** is $36 approximately with the 95% CI being [30.81282 , 35.75067].

Question 9:

We can find R^2(Residual sum of squares) from the summary of the model1. We know that R^2 is equal to r^2(correlation coefficient).
Therefore, r = sqrt(R)

```
> summary(model1)$r.squared
[1] 0.2155561
> sqrt(summary(model1)$r.squared)
[1] 0.4642802
```

```
> summary(model2)$r.squared
[1] 0.0008659354
> sqrt(summary(model2)$r.squared)
[1] 0.02942678
```

Question 10:

From question 9 we get the R^2 for model1 is 0.215, and so just under 21.5% of the variability in expenditure is explained by a linear regression on (NewPI) new personal injury cases, R^2 for model2 is 0.008, and so the variability in expenditure explained by a linear regression on (NewWC) workers' compensation cases is very **low**. As the number of model2 is near 0, it has **no effect** on the variability in responses or we may need more data for model2.

Question 11:

Based on the above analysis, the court can give the following verdict:

"Partner A has wrongly sued Partner B in the hope of equal contribution towards advertising expenditure. The graphs and R^2 values indicate that there is negligible (less than 0.8%) influence of advertising on workers' compensation cases when compared to the 21.5% influence over personal injury cases. The verdict should be given in favour of Partner B. Partner A, alone should pay for the advertising expenditure."