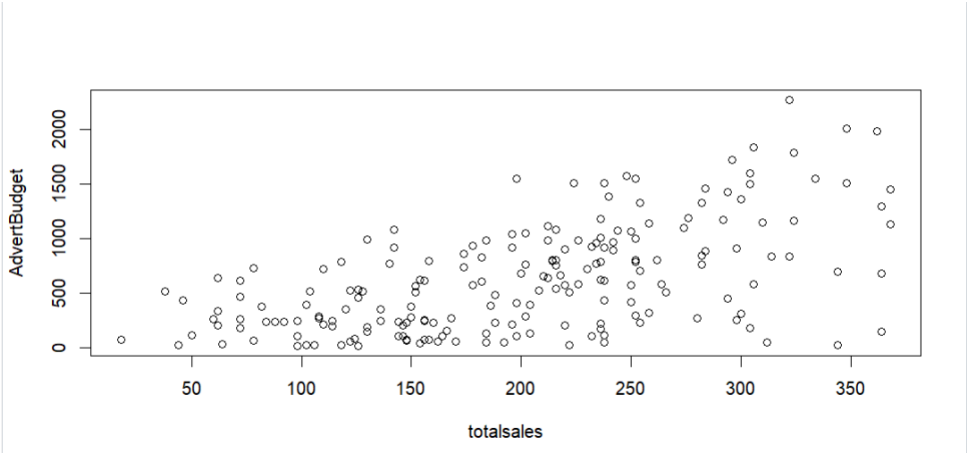
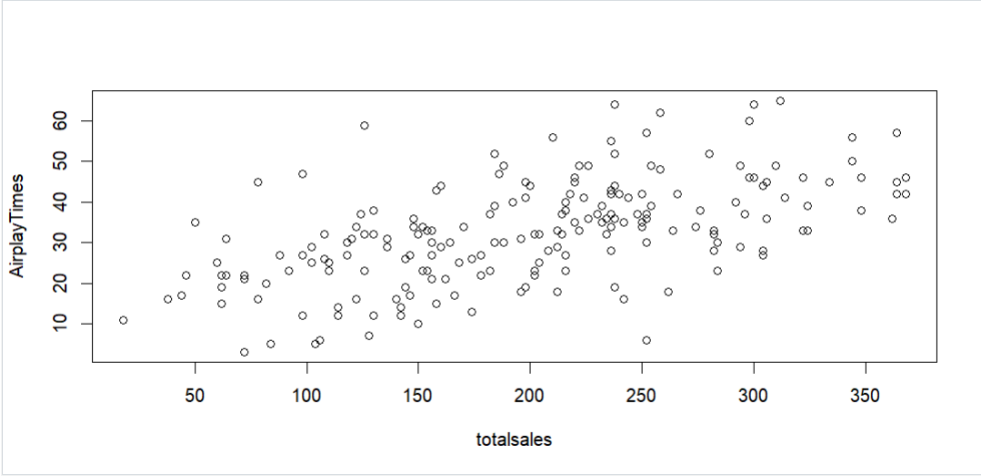
CSCI-4047-901
Tyler Burleson
Exercise 5

Visualization -

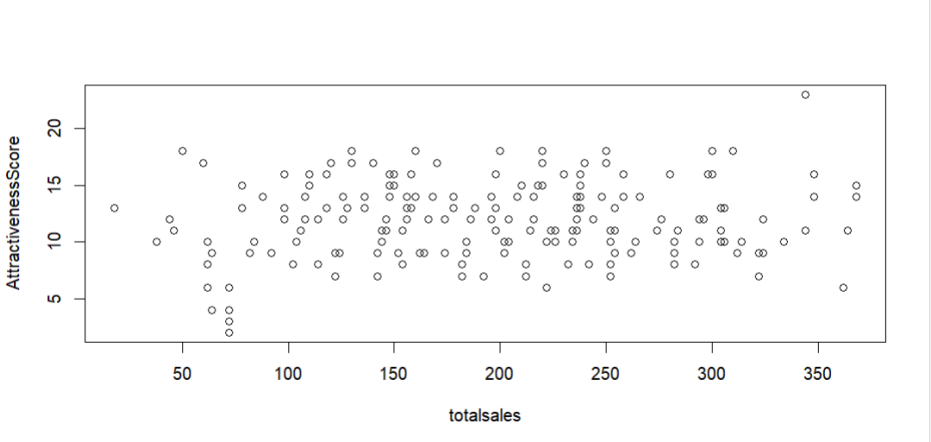
1. Sales and advertising -



2. Sales and airplay -



3. Sales and attractiveness -



CSCI-4047-901
Tyler Burleson
Exercise 5

Linear Regression -

4. Conduct a linear regression to construct a linear model between Sales and adverts and write down the F-statistic and P-value.

```
Call:
lm(formula = data$totalsales ~ data$AdvertBudget)

Residuals:
     Min       1Q   Median       3Q      Max
-150.880  -43.217    0.541   35.795  210.079

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.396e+02  7.589e+00  18.395   <2e-16 ***
data$AdvertBudget 9.552e-02  9.647e-03   9.902   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 66.02 on 198 degrees of freedom
Multiple R-squared:  0.3312,    Adjusted R-squared:  0.3278
F-statistic: 98.04 on 1 and 198 DF,  p-value: < 2.2e-16
```

F-statistic: 98.04 P-value: <2.2e-16

5. Discuss what these values (F-statistic and P-value) describe about our linear regression model? Is it good? Bad? Can't say?
This is a strong model due to the ratio we have with our F-statistic and our P-value. Our F-statistic has a high value above 1 (98.04) while keeping a low P-value below 0.05 (2.2e-16)

Model Coefficients

6. What is the intercept value and coefficient (adverts) value of your linear regression model?
Knowing that the regression line is described using this equation
Y = $bb0$ + $bb1$ $XX1$
Where,

- Y denotes the sales
- $bb0$ denotes the intercept value
- $bb1$ denotes the advertising budget coefficient
- $XX1$ denotes the advertising budget

The equation becomes:
Album sales = intercept value + (coefficient * advertising budget)
Intercept value: 139.6 AdvertBudget-Coefficient: 0.09552
7. Using the intercept value and coefficient of your linear model, please calculate how many records will be sold if we spent $135 000 on advertising the latest album "Dear Agony" by Breaking Benjamin
$13,034.8 = 139.6 + (0.09552 * 135000)

CSCI-4047-901
Tyler Burleson
Exercise 5

Multiple Regression

8. Conduct a multiple regression to construct a model between Sales and the predictors (adverts, airplay, attract) and report the F-statistic and P-value.

```
Call:
lm(formula = data$totalsales ~ data$AdvertBudget + data$AirplayTimes +
    data$AttractivenessScore)

Residuals:
     Min       1Q   Median       3Q      Max
-118.972  -31.086    5.338   30.384  164.176

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              35.206800  15.188398   2.318   0.0215 *
data$AdvertBudget         0.089167   0.007395  12.057   <2e-16 ***
data$AirplayTimes         3.448470   0.295765  11.660   <2e-16 ***
data$AttractivenessScore -0.358069   1.105602  -0.324   0.7464
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 50.22 on 196 degrees of freedom
Multiple R-squared:  0.617,     Adjusted R-squared:  0.6111
F-statistic: 105.2 on 3 and 196 DF,  p-value: < 2.2e-16
```

F-statistic: 105.2 P-value: < 2.2e-16

9. We know that the R-squared value can be used to evaluate the overall fit of a linear model. Also that higher R-squared values are better if their p-values is < 0.05.

Based on this, discuss which one of the two models that you constructed is Better?

- Model 1: the linear model between Sales and advertising you constructed in (4).
- Model 2: the multiple regression model between outcome: Sales and the predictors (advertising, airplay, attractiveness) that you constructed in (8).

Model 2 is superior to model one: **When we used the multiple regression on our sales we were able to generate a higher F-statistic with a similar P-value, additionally the R-squared in model 2 is higher (0.617) than the one in model 1 (0.3312)**