

# Lab Assignment 2:

**Objective: To apply linear regression on a dataset.**

**Name: Aakash Verma**

**Registration No. : 24-08-26** 

**Course : M.Tech.(Cyber Security)**

```
In [1]: # installing necessary packages
install.packages("tidyverse")

install.packages("datarium")
install.packages("gridExtra")
```

...

```
In [3]: library(tidyverse)
library(gridExtra)
```

```
In [4]: data("marketing", package = "datarium")
```

```
In [5]: head(marketing)
```

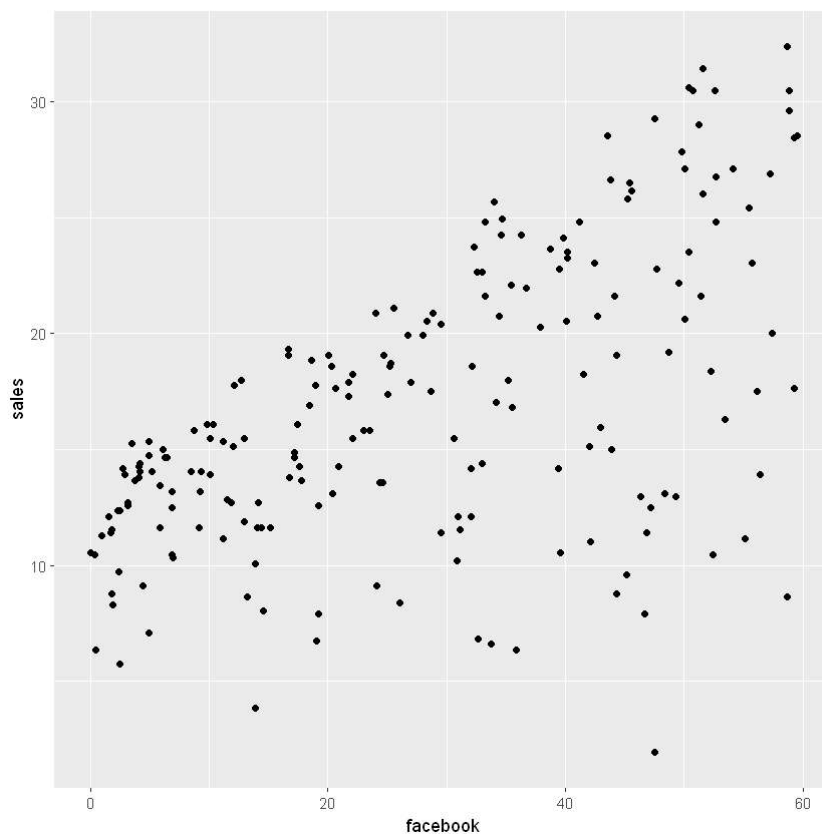
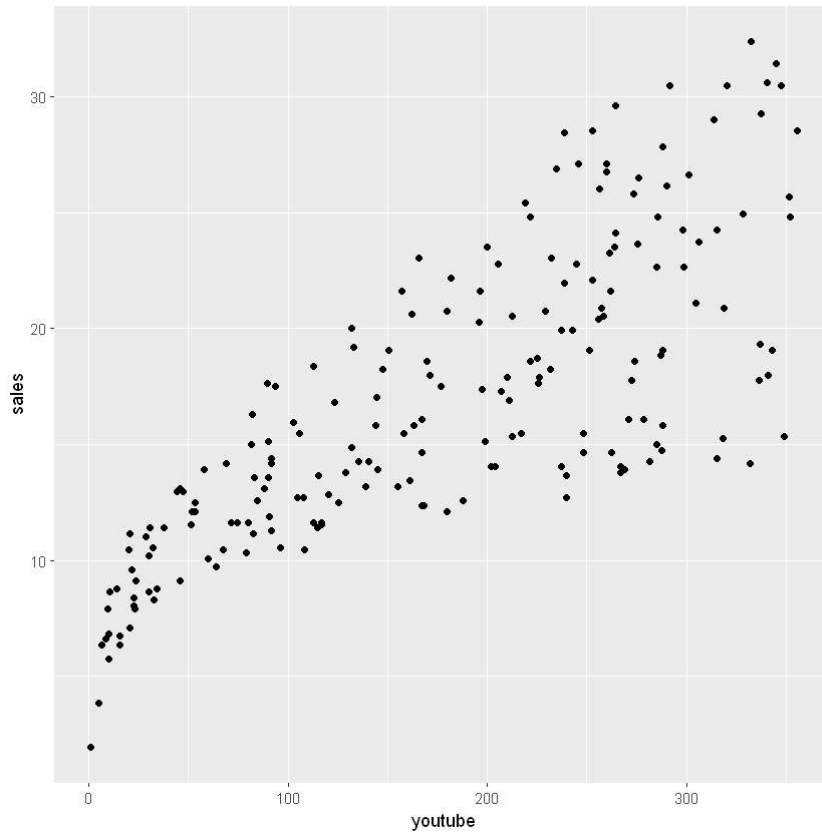
A data.frame: 6 × 4

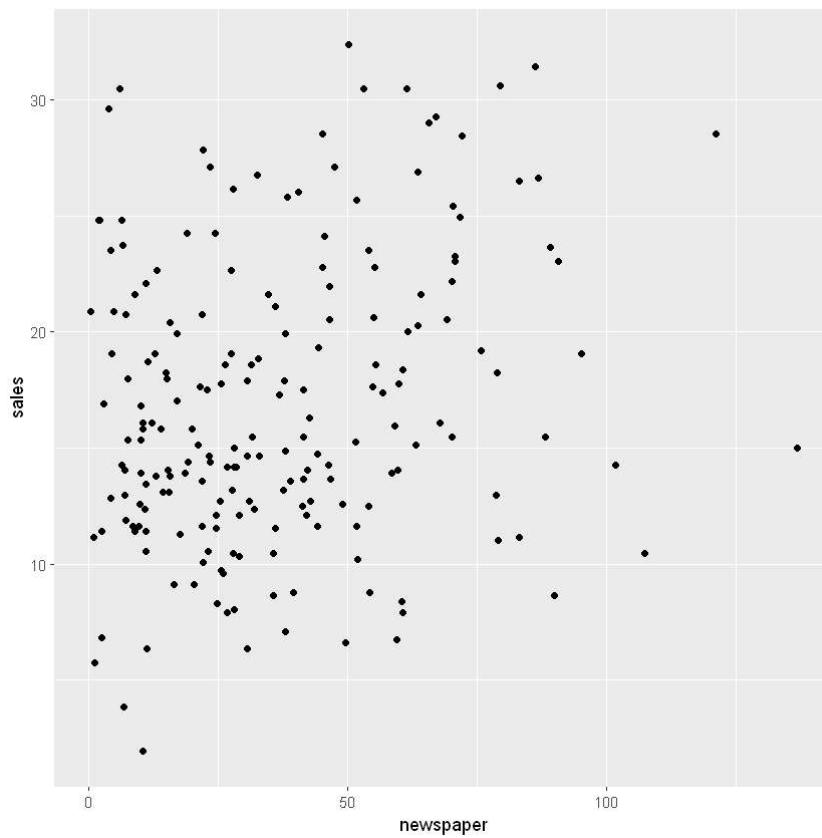
	youtube	facebook	newspaper	sales
	<dbl>	<dbl>	<dbl>	<dbl>
1	276.12	45.36	83.04	26.52
2	53.40	47.16	54.12	12.48
3	20.64	55.08	83.16	11.16
4	181.80	49.56	70.20	22.20
5	216.96	12.96	70.08	15.48
6	10.44	58.68	90.00	8.64

```
In [6]: summary(marketing)
```

youtube	facebook	newspaper	sales
Min. : 0.84	Min. : 0.00	Min. : 0.36	Min. : 1.92
1st Qu.: 89.25	1st Qu.:11.97	1st Qu.: 15.30	1st Qu.:12.45
Median :179.70	Median :27.48	Median : 30.90	Median :15.48
Mean :176.45	Mean :27.92	Mean : 36.66	Mean :16.83
3rd Qu.:262.59	3rd Qu.:43.83	3rd Qu.: 54.12	3rd Qu.:20.88
Max. :355.68	Max. :59.52	Max. :136.80	Max. :32.40

```
In [7]: ggplot(marketing, aes(x = youtube, y = sales)) +  
  geom_point()  
  
ggplot(marketing, aes(x = facebook, y = sales)) +  
  geom_point()  
  
ggplot(marketing, aes(x = newspaper, y = sales)) +  
  geom_point()
```





```
In [8]: # training linear regression model for different column with sales
model1 = lm(sales ~ youtube, marketing)
model2 = lm(sales ~ facebook, marketing)
model3 = lm(sales ~ newspaper, marketing)
```

```
In [9]: # details of model params
model1
print('-----')
model2
print('-----')
model3
```

```
Call:
lm(formula = sales ~ youtube, data = marketing)
```

```
Coefficients:
(Intercept)      youtube
   8.43911      0.04754
```

```
[1] "-----"
```

```
Call:
lm(formula = sales ~ facebook, data = marketing)
```

```
Coefficients:
(Intercept)      facebook
  11.1740      0.2025
```

```
[1] "-----"
```

```
Call:
lm(formula = sales ~ newspaper, data = marketing)
```

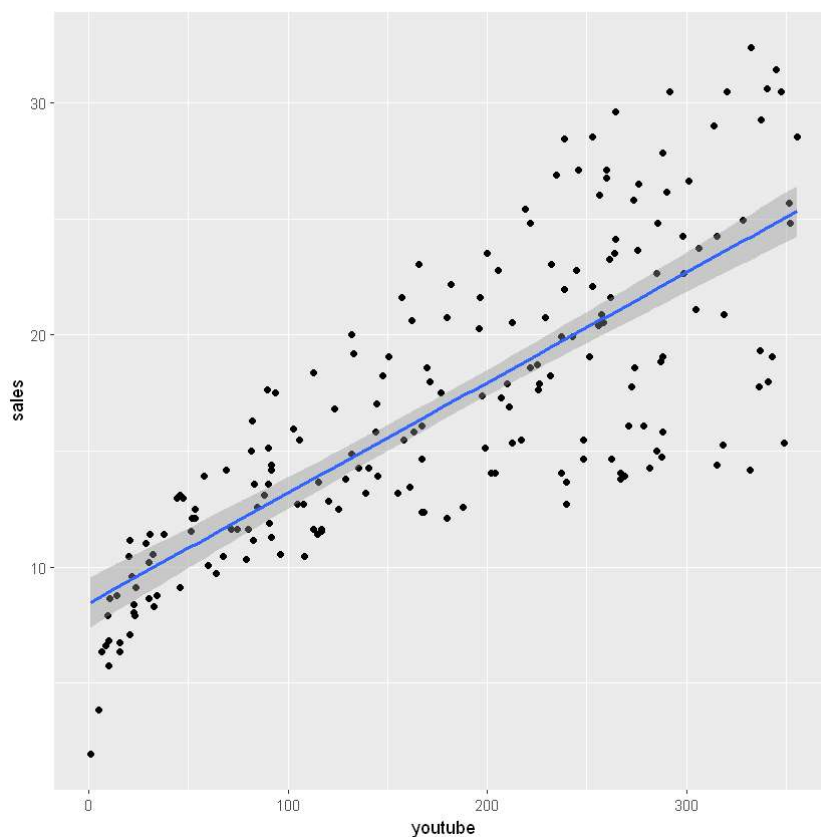
```
Coefficients:
(Intercept)      newspaper
  14.82169      0.05469
```

```
In [18]: # plotting of data points w.r.t trained model
ggplot(marketing, aes(youtube, sales)) +
  geom_point() +
  stat_smooth(method = lm)

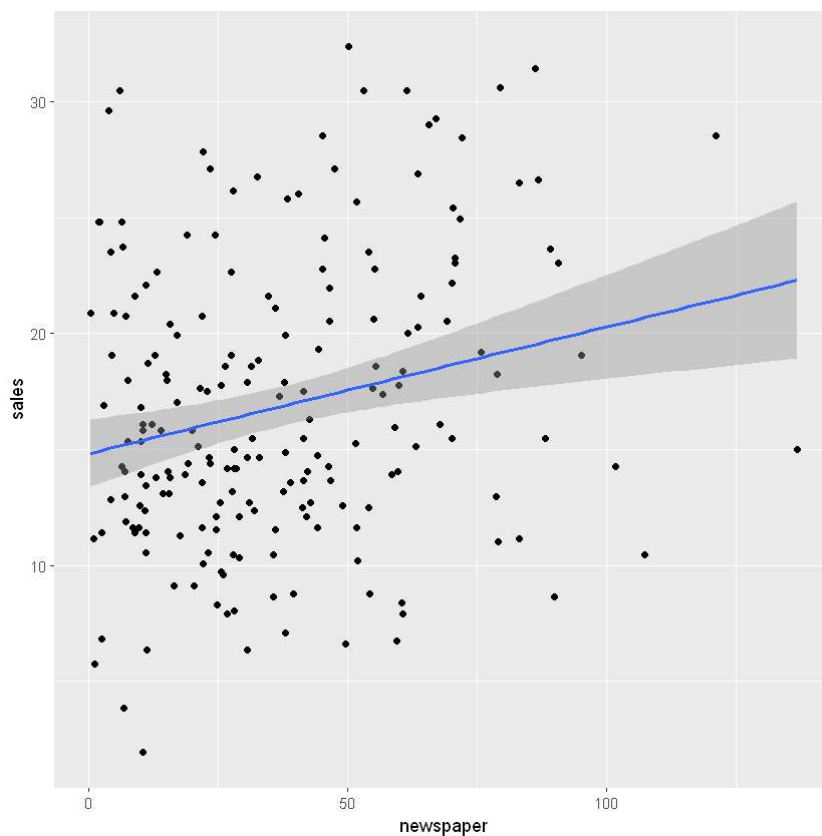
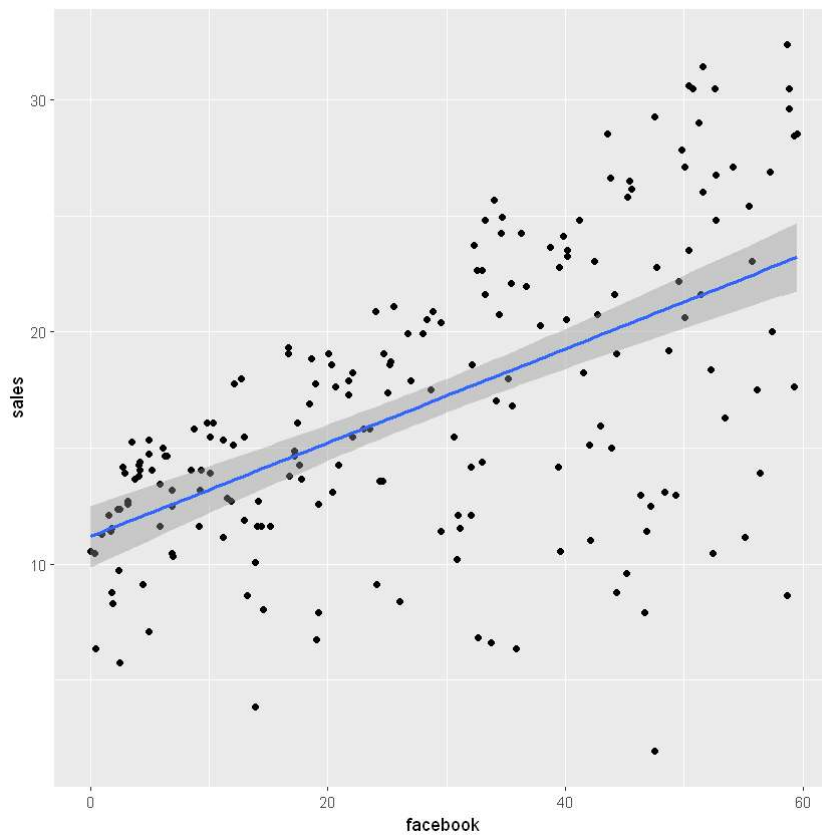
ggplot(marketing, aes(facebook, sales)) +
  geom_point() +
  stat_smooth(method = lm)

ggplot(marketing, aes(newspaper, sales)) +
  geom_point() +
  stat_smooth(method = lm)
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```



```
`geom_smooth()` using formula = 'y ~ x'
```



```
In [13]: calculating residuals and adding it to marketing dataframe columns
marketing$residuals_youtube <- residuals(model1)
marketing$residuals_facebook <- residuals(model2)
marketing$residuals_newspaper <- residuals(model3)
```

In [14]: head(marketing)

A data.frame: 6 × 7

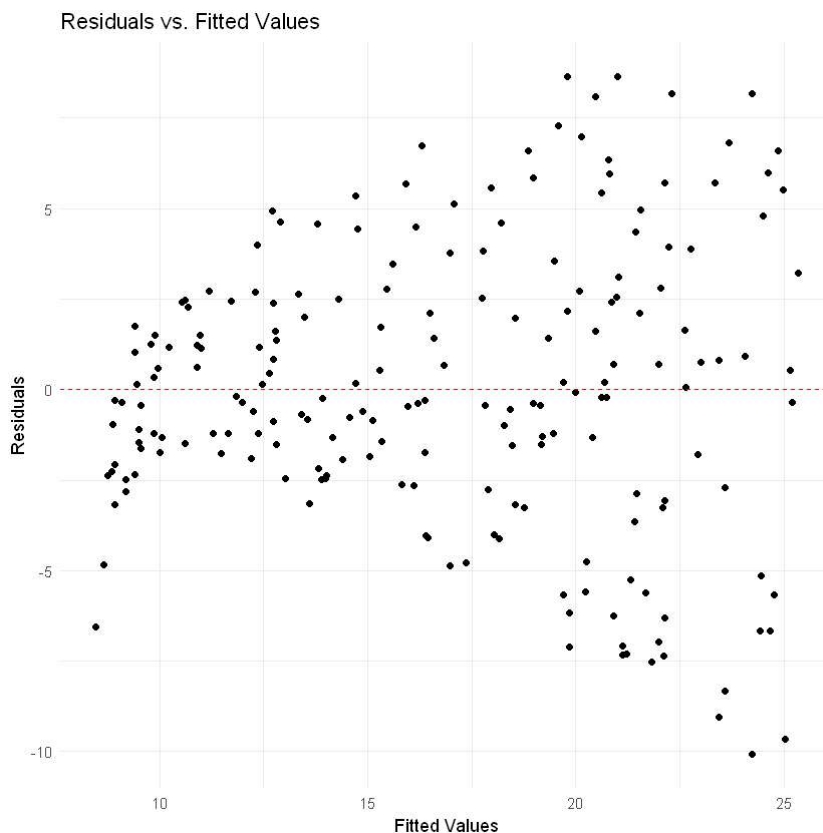
	youtube	facebook	newspaper	sales	residuals_youtube	residuals_facebook	residuals_ne
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	276.12	45.36	83.04	26.52	4.9550706	6.1608256	
2	53.40	47.16	54.12	12.48	1.5024311	-8.2436669	-
3	20.64	55.08	83.16	11.16	1.7397315	-11.1674335	-
4	181.80	49.56	70.20	22.20	5.1187265	0.9903433	
5	216.96	12.96	70.08	15.48	-3.2726618	1.6816889	-
6	10.44	58.68	90.00	8.64	-0.2953948	-14.4164183	-1



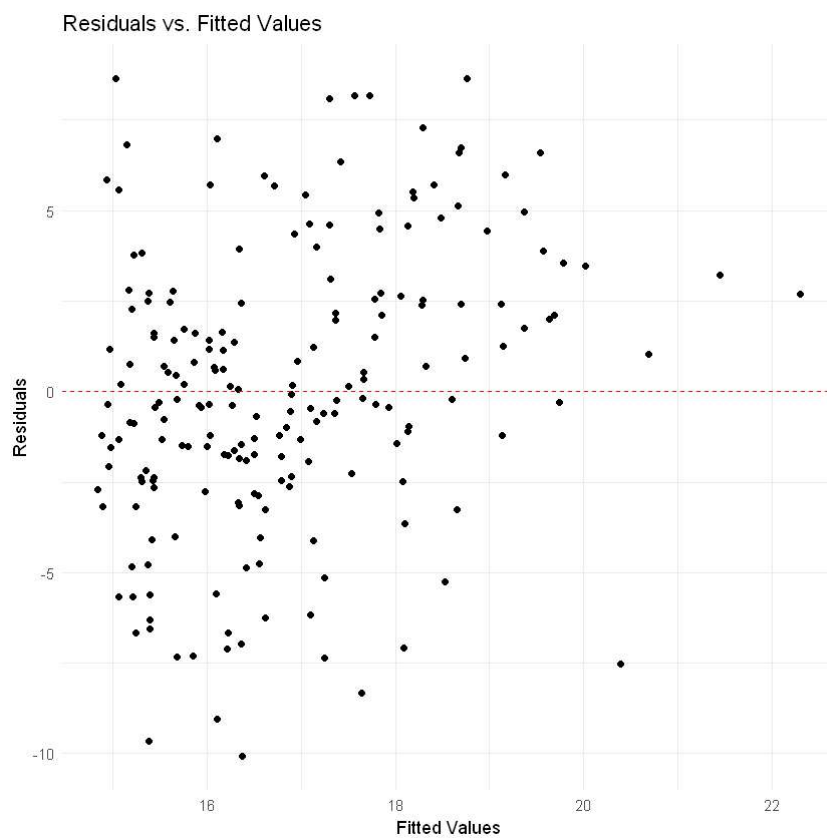
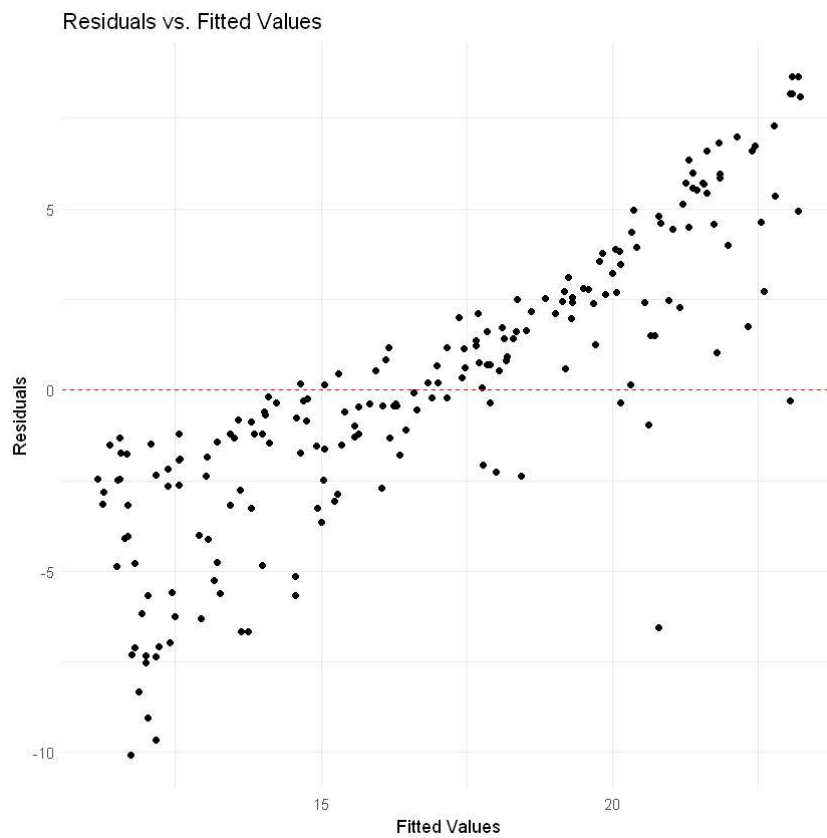
```
In [19]: library(ggplot2)
# plotting of residuals w.r.t. the fitted data.
ggplot(marketing, aes(x = fitted(model1), y = residuals_youtube)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values",
        x = "Fitted Values",
        y = "Residuals") +
  theme_minimal()

ggplot(marketing, aes(x = fitted(model2), y = residuals_youtube)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values",
        x = "Fitted Values",
        y = "Residuals") +
  theme_minimal()

ggplot(marketing, aes(x = fitted(model3), y = residuals_youtube)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values",
        x = "Fitted Values",
        y = "Residuals") +
  theme_minimal()
```







In [ ]:

In [ ]:

## Conclusion

- 1. Scatter plot gives better understanding whether the linear model will work for the dataset or not.**
- 2. Facebook Advertising: Most impactful, increasing sales by 0.203 units per unit spent**
- 3. Newspaper Advertising: Moderate impact, increasing sales by 0.055 units per unit spent.**
- 4. YouTube Advertising: Least impact, increasing sales by 0.048 units per unit spent.**
- 5. Data Overview: Facebook has the highest mean spend (27.92 units), while YouTube has the highest range (0.84 to 355.68 units).**