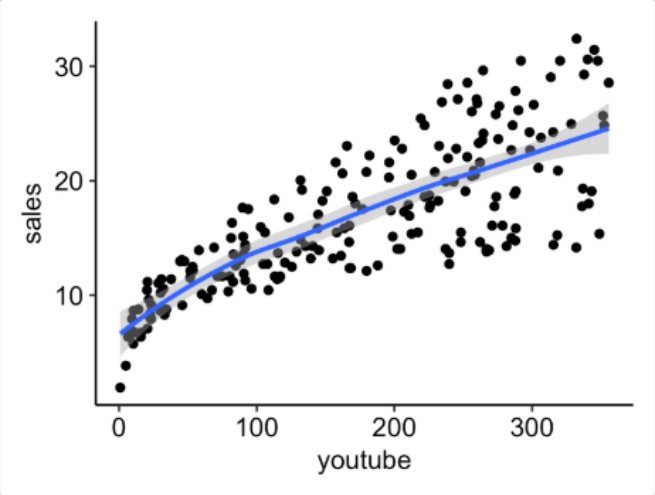
DATA 106 Simple Linear Regression Worksheet

Consider the following dataset and results:
Data – first 4 observations

```
##   youtube facebook newspaper sales
## 1   276.1     45.4      83.0  26.5
## 2    53.4     47.2      54.1  12.5
## 3    20.6     55.1      83.2  11.2
## 4   181.8     49.6      70.2  22.2
```

Visualization

```
ggplot(marketing, aes(x = youtube, y = sales)) +
  geom_point() +
  stat_smooth()
```



Model

```
model <- lm(sales ~ youtube, data = marketing)
```

```
summary(model)
```

```
##
## Call:
## lm(formula = sales ~ youtube, data = marketing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.06  -2.35  -0.23   2.48   8.65
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.43911    0.54941    15.4   <2e-16 ***
## youtube      0.04754    0.00269    17.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.91 on 198 degrees of freedom
## Multiple R-squared:  0.612,  Adjusted R-squared:  0.61
## F-statistic:  312 on 1 and 198 DF,  p-value: <2e-16
```

DATA 106 Simple Linear Regression Worksheet

1. Is there a relationship between the YouTube and sales? Why?

2. Is the relationship linear? Why?

3. How strong is the relationship between the YouTube and sales? Why?

4. Does YouTube contribute to sales? Why?

5. Use the effect of YouTube to estimate sales. Describe.

6. How accurately can we predict future sales using YouTube? Why?

Here are some more results:

```
> model2=lm(sales ~ facebook, data = train.data)
> summary(model2)

Call:
lm(formula = sales ~ facebook, data = train.data)

Residuals:
    Min      1Q  Median      3Q     Max
-18.812  -2.577   1.044   3.350   9.901

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.3301     0.7503  15.100  < 2e-16 ***
facebook      0.1978     0.0228   8.676 4.41e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.239 on 160 degrees of freedom
Multiple R-squared:  0.3199,    Adjusted R-squared:  0.3157
F-statistic: 75.27 on 1 and 160 DF,  p-value: 4.413e-15
```

```
> model3=lm(sales ~ newspaper, data = train.data)
> summary(model3)

Call:
lm(formula = sales ~ newspaper, data = train.data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.3208  -4.3284  -0.6912   4.0602  15.4768

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.68194    0.87008  16.874  < 2e-16 ***
newspaper    0.05353    0.01847   2.899  0.00427 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.192 on 160 degrees of freedom
Multiple R-squared:  0.04991,    Adjusted R-squared:  0.04397
F-statistic: 8.405 on 1 and 160 DF,  p-value: 0.004267
```

7. **Which of facebook, newspaper or YouTube will more accurately predict sales? Why?**

8. **Which of these variables have a stronger relationship/association with sales? Why?**

9. **Are all these variables good predictors of sales? Why?**

10. **Which of these variables are have a biggest effect on sales? Why?**