What is R-squared – Most Memorable Visualizations?

February 4, 2022, Jill Wang

What is R^2?

In a regression model, the Total Sum of Square (**TSS**) is a measurement of the sum of squared difference between the observed dependent variable and its mean, $\sum_{i=1}^{n}(y_i - \bar{y})^2$. It can be split into two components:

1. The amount of variability in the response that is explained by a regression model. It is called Explained Sum of Squares or Regression Sum of Squares, abbreviated as **ESS**. It is the sum of squared difference between the predicted value and the mean of the dependent variable, $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$.

2. The amount of variability that is unexplained by a regression model. It is known as Residual Sum of Squares or Error Sum of Squares, abbreviated as **RSS**. It is the sum of squared difference between the observed value and the predicted value, $\sum_{i=1}^{n} e_i^2$.

R-squared is the proportion of variability in Y that can be explained by a linear model using X. To calculate R^2, we use the following formula

$$R^2 = \frac{Explained\ Variation}{Total\ Variation} = 1 - \frac{RSS}{TSS}$$

The larger the R-squared, the great amount of explainable variation, and the better the regression model that fits your observations.

```
install.packages ("correlation")

install.packages ("ggplot2")

install.packages ("patchwork")
```
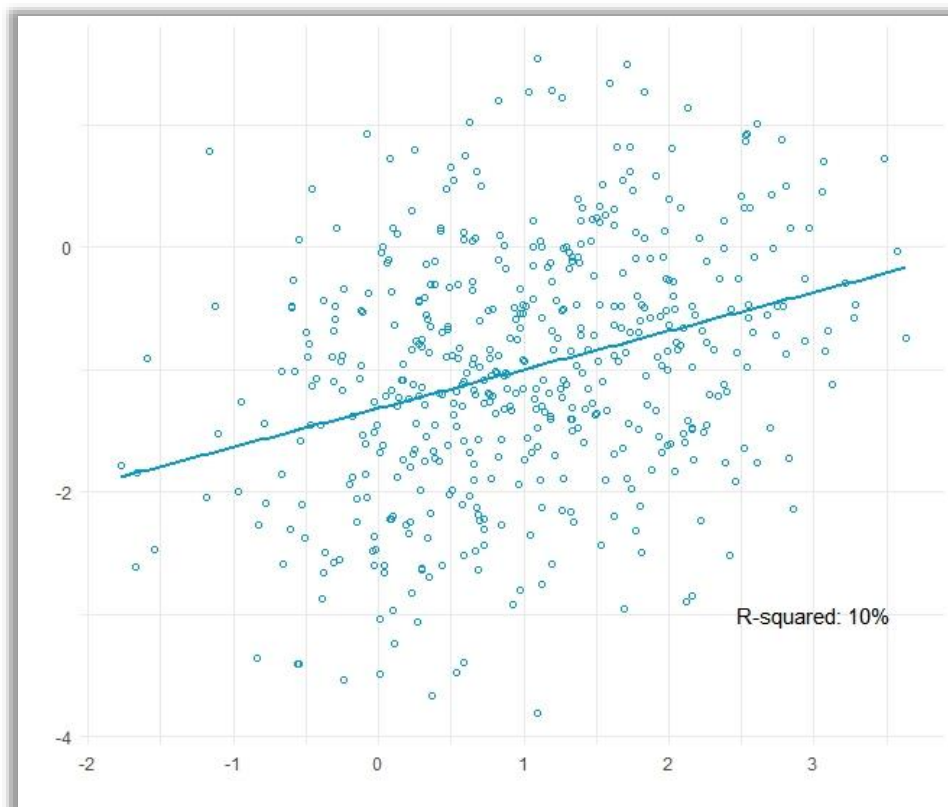
```
library (correlation)

library (ggplot2)

library (patchwork)
```

Understanding R-squared is super useful, especially while building a more complex model. Let us plot multiple R-squared values in R – the graphs that you will never forget.
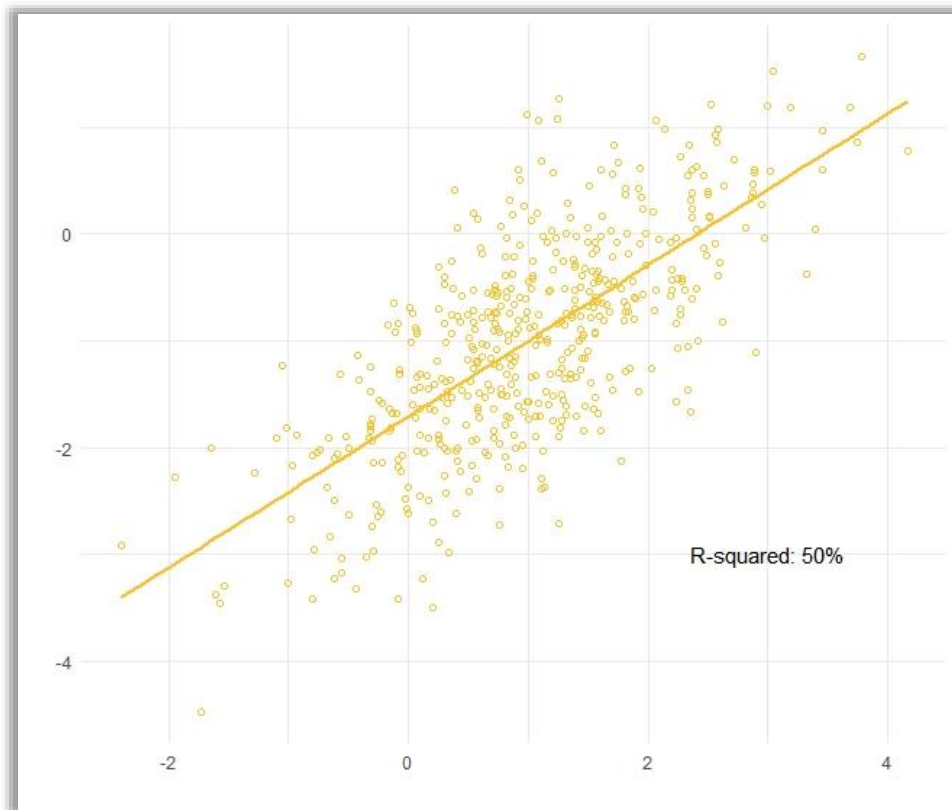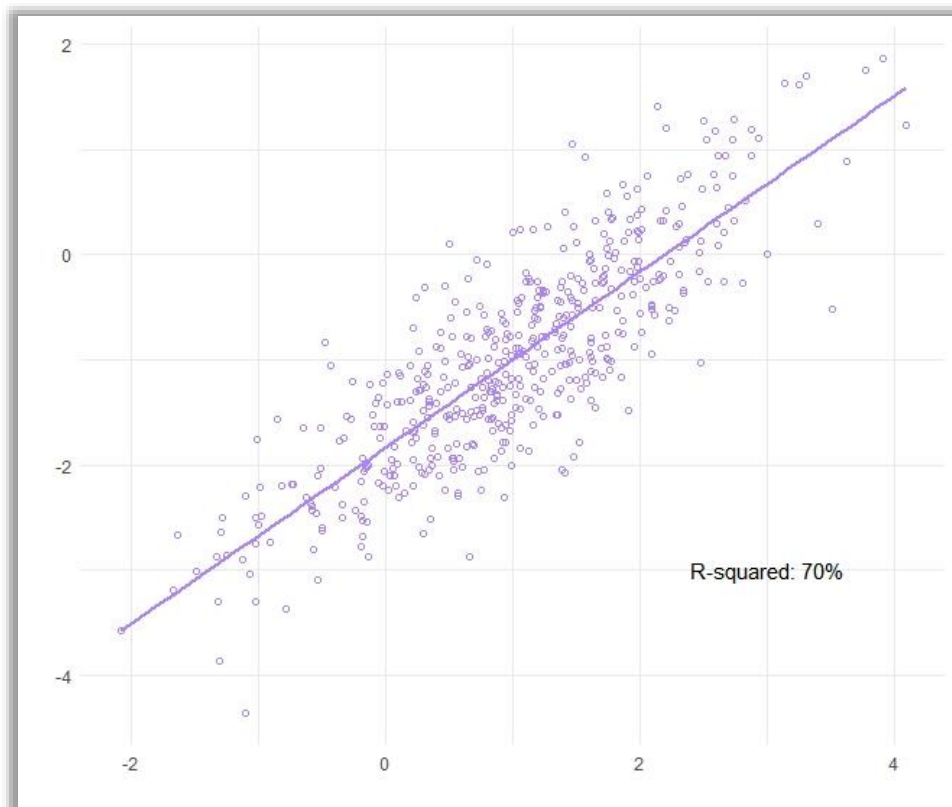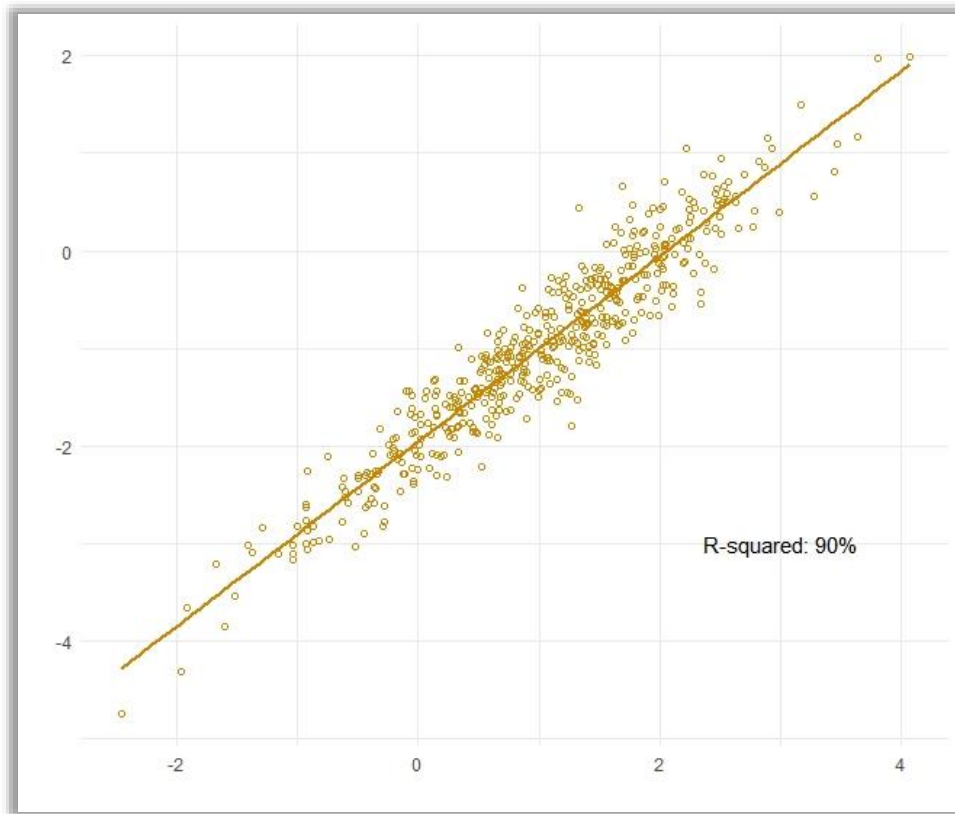
Graph 1. $R^2 = 0\%$

R-squared: 0%

Graph 2. $R^2 = 10\%$

R-squared: 10%

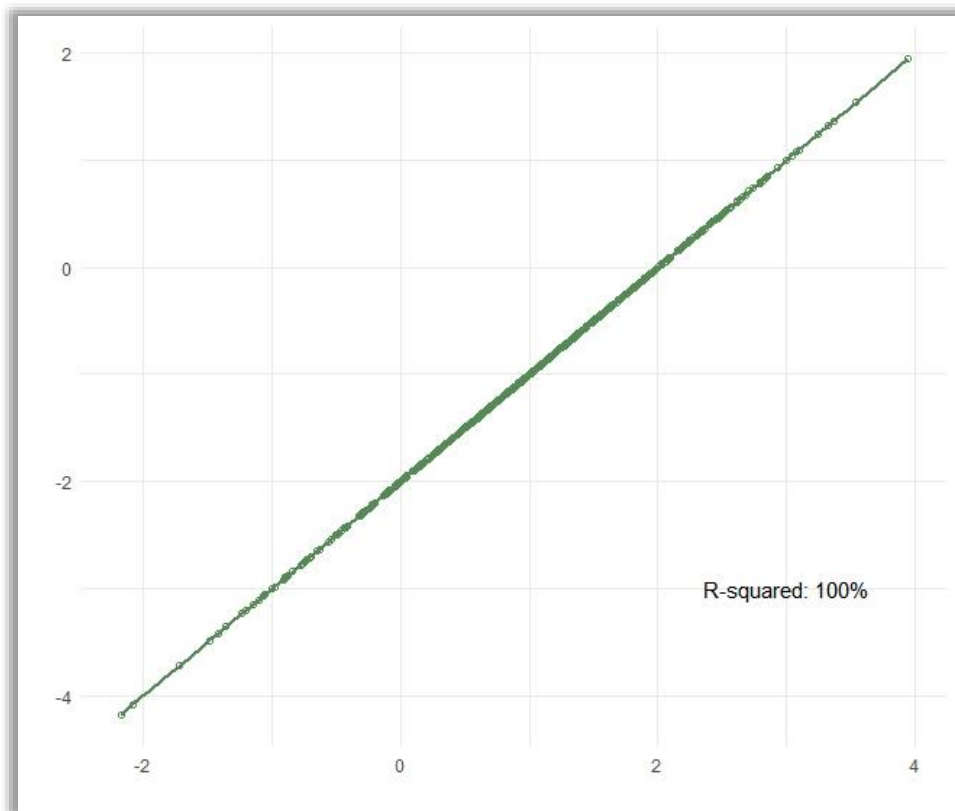**Graph 3. $R^2 = 50\%$**

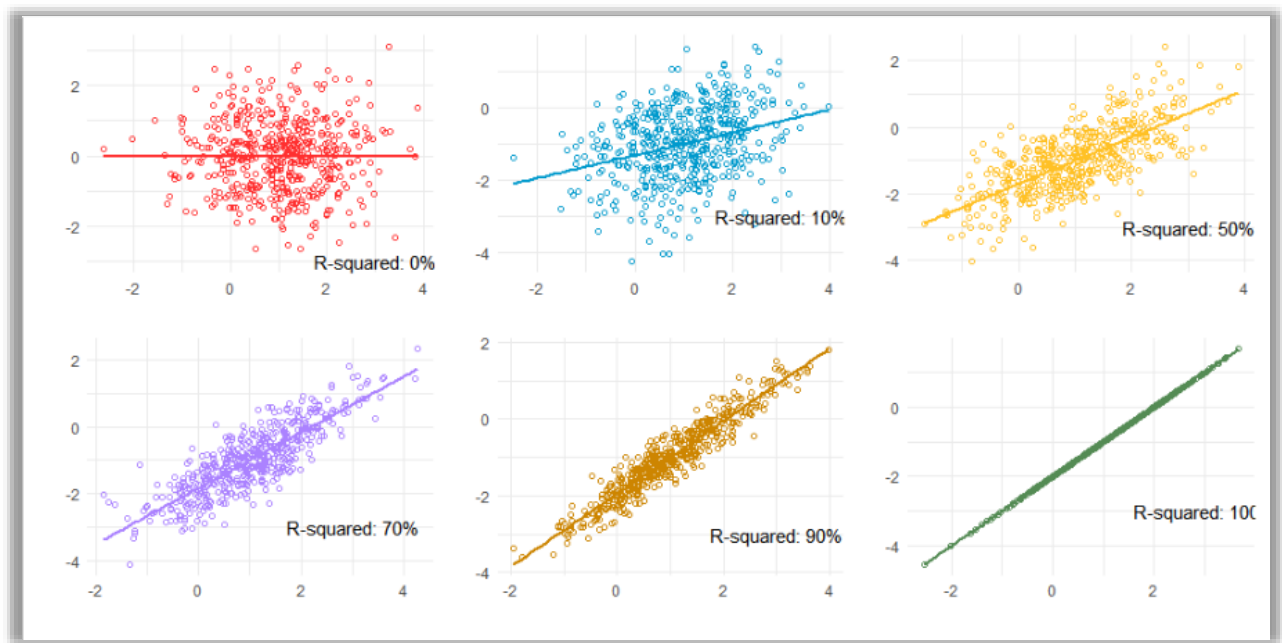R-squared: 50%
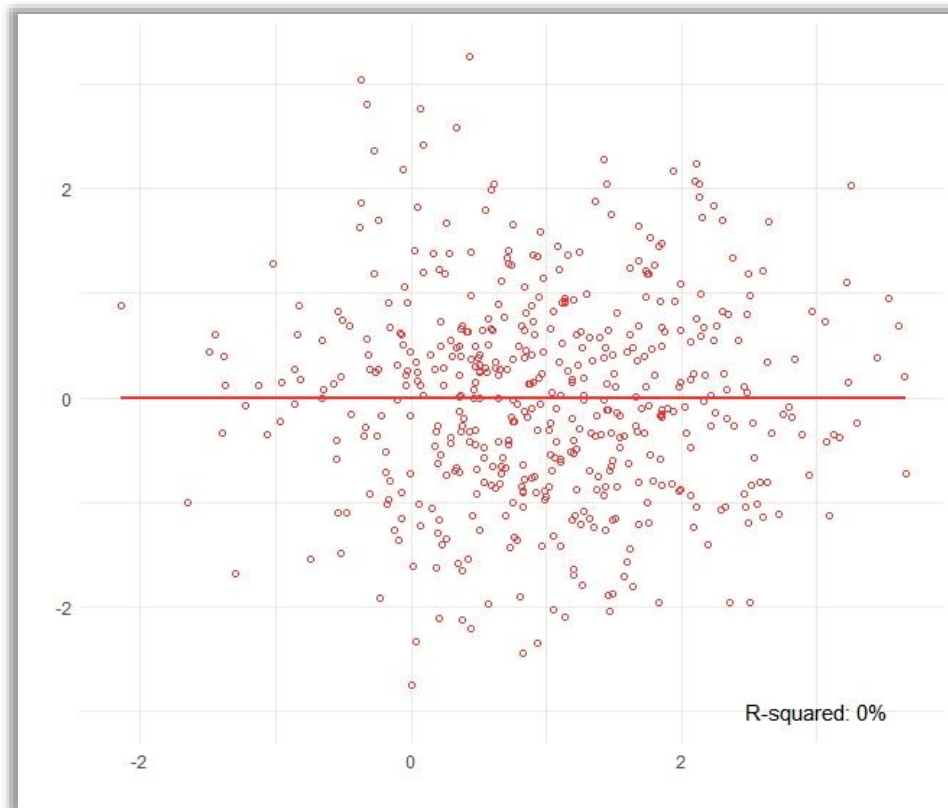
Graph 4. $R^2 = 70\%$

R-squared: 70%

Graph 5. R^2 = 90%

Graph 6. R^2 = 100%

Graph 7. A merge graph from multiple graphs

Reference

[1] The Elements of Statistical Learning