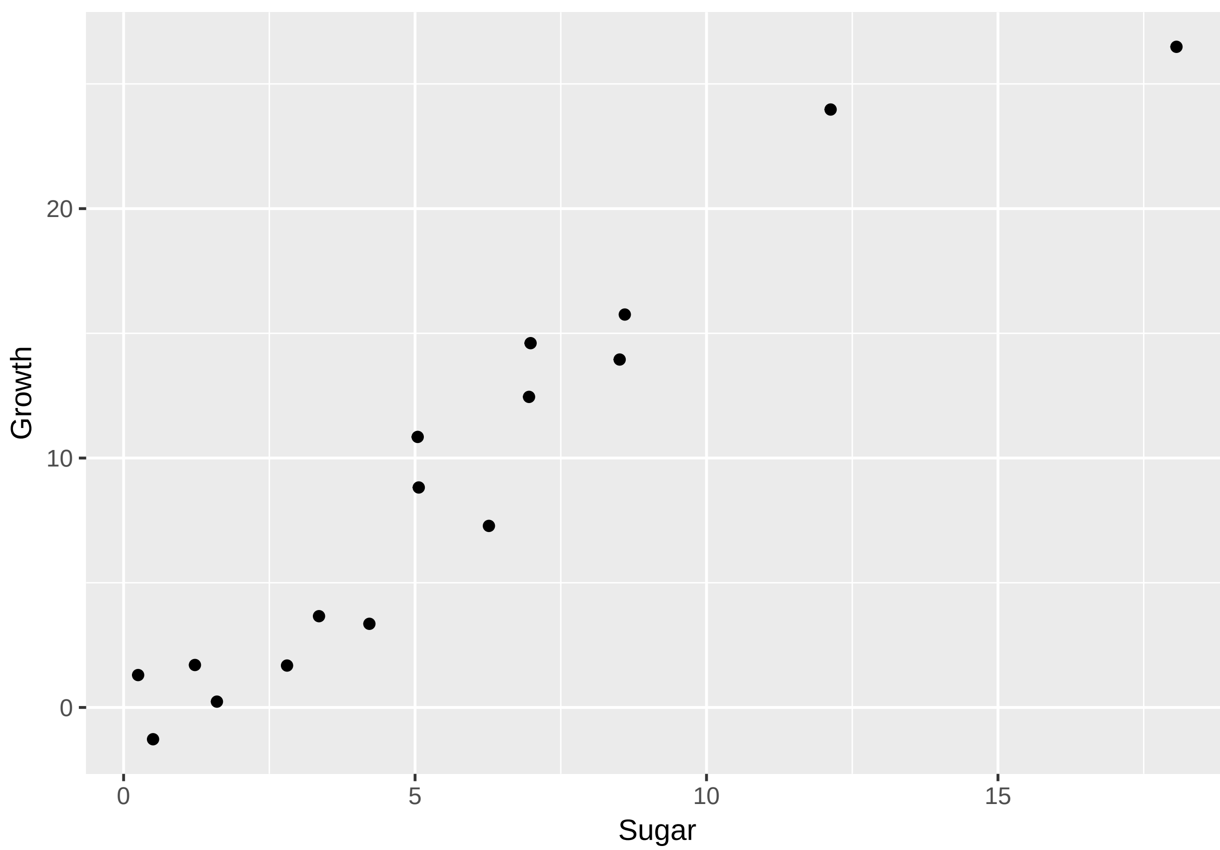
Sugar Regression

We have data from an experiment evaluating the effect of sugar concentration on growth of *E. Coli* in lab cultures. An overview of the data is presented below, both in tabular and graphical form:

```
##      sugar            growth      
##  Min.   : 0.249   Min.   :-1.274  
##  1st Qu.: 2.503   1st Qu.: 1.698  
##  Median : 5.054   Median : 8.050  
##  Mean   : 5.723   Mean   : 9.052  
##  3rd Qu.: 7.364   3rd Qu.:14.114  
##  Max.   :18.062   Max.   :26.488  
```



Superficially, at least, this appears to indicate some sort of linear relationship. We will test this by comparing the negative log likelihood functions of a constant model and a linear model using a likelihood ratio test. First, however, we need hypotheses:

- **Null Hypothesis:** The slope of a linear model fitted to the data is 0.
- **Alternate Hypothesis:** The slope of a linear model fitted to the data is not 0.

Let G be a variable denoting the values for growth, S be a variable denoting the values for sugar, and ϵ be an error term which is normally distributed with standard deviation σ. Then we fit models of the form $G = \beta_0 + \epsilon$ and $G = \beta_0 + \beta_1 S + \epsilon$ to the data. An optimization run on the first model returned values of $\beta_0 \approx 9.0559$ and $\sigma \approx 8.134$; an optimization run on thes second model returend values of $\beta_0 \approx -0.85101$, $\beta_1 \approx 1.7304$, and $\sigma \approx 2.3363$. The negative log-likelihood for the constat model was approximately 56.2412; for the linear model it was approximately 36.27864.

To implement the likelihood ratio test, we found the absolute values of the difference between the negative log-likelihood values, multiplied this by two. As the size of the sample grows, this value approaches a chi-square distribution with 1 degree of freedom, so we used a chi-square distribution with 1 degree of freedom to find the p-value for the experiment.

The p-value was $p \approx 2.6389 \times 10^{-10}$. This value is quite small, so we reject the null hypothesis, and conclude that a linear fit correctly explains the patterns we see in the data. We summarize this with another graph: