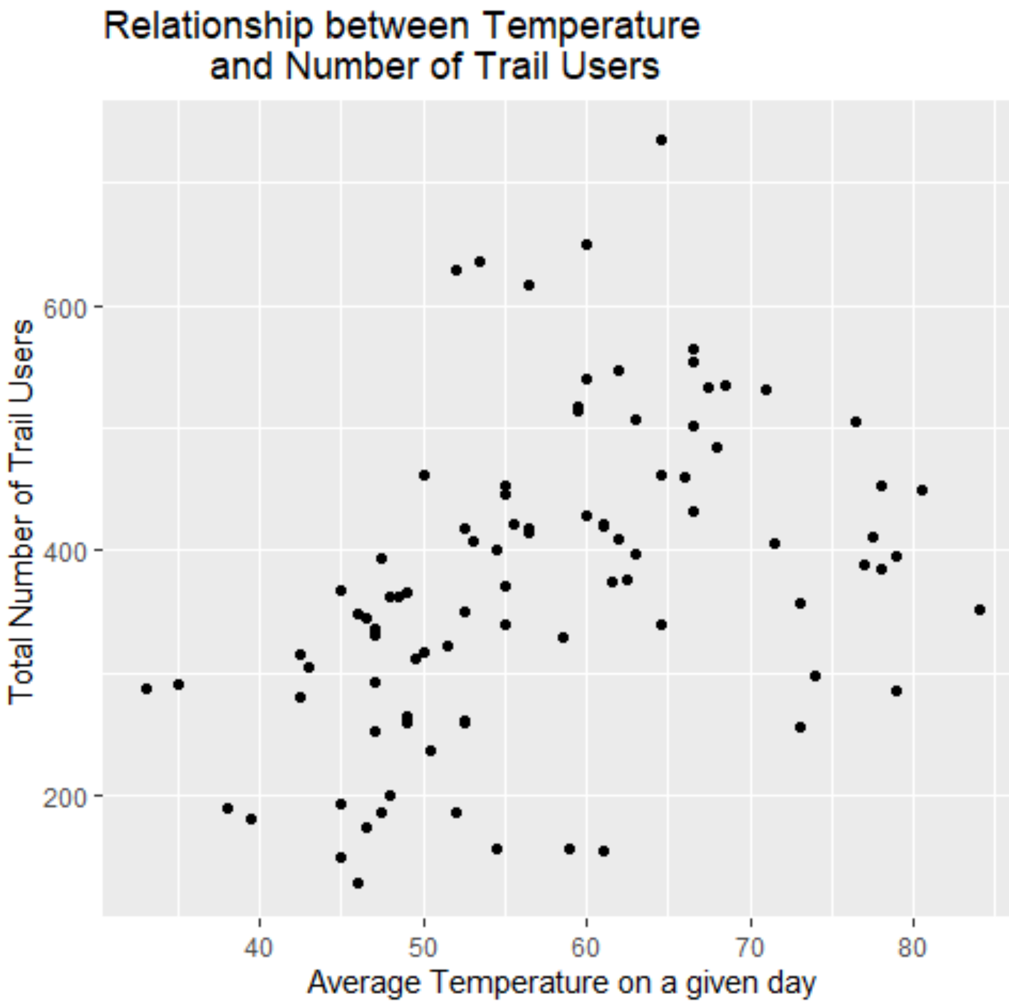
Simple Linear Regression

Scatter Plot

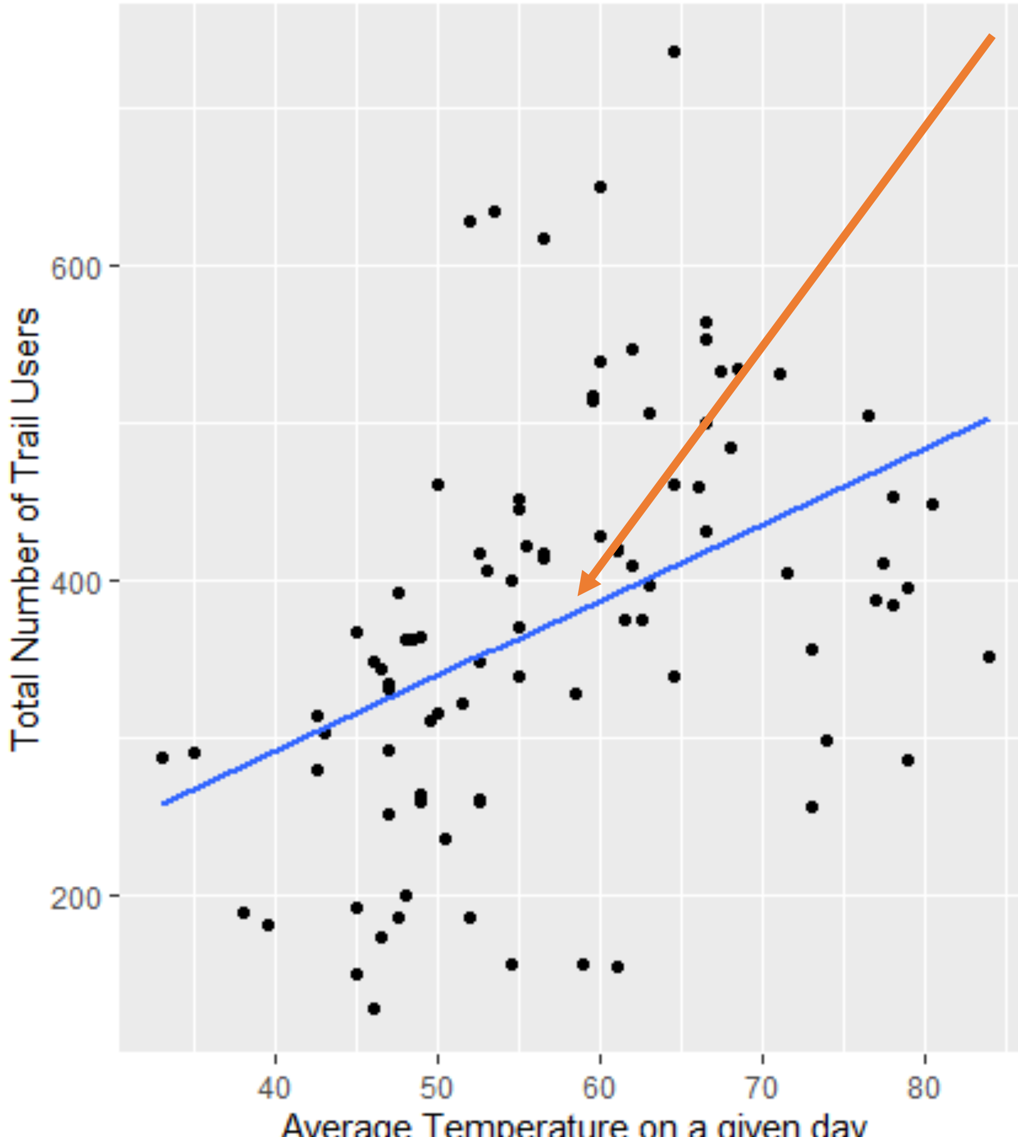


Pearson Correlation Test

```
        Pearson's product-moment correlation

data:  RailTrail$volume and RailTrail$avgtemp
t = 4.4279, df = 88, p-value = 2.723e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2410727 0.5824569
sample estimates:
      cor
0.4268535
```

Regression



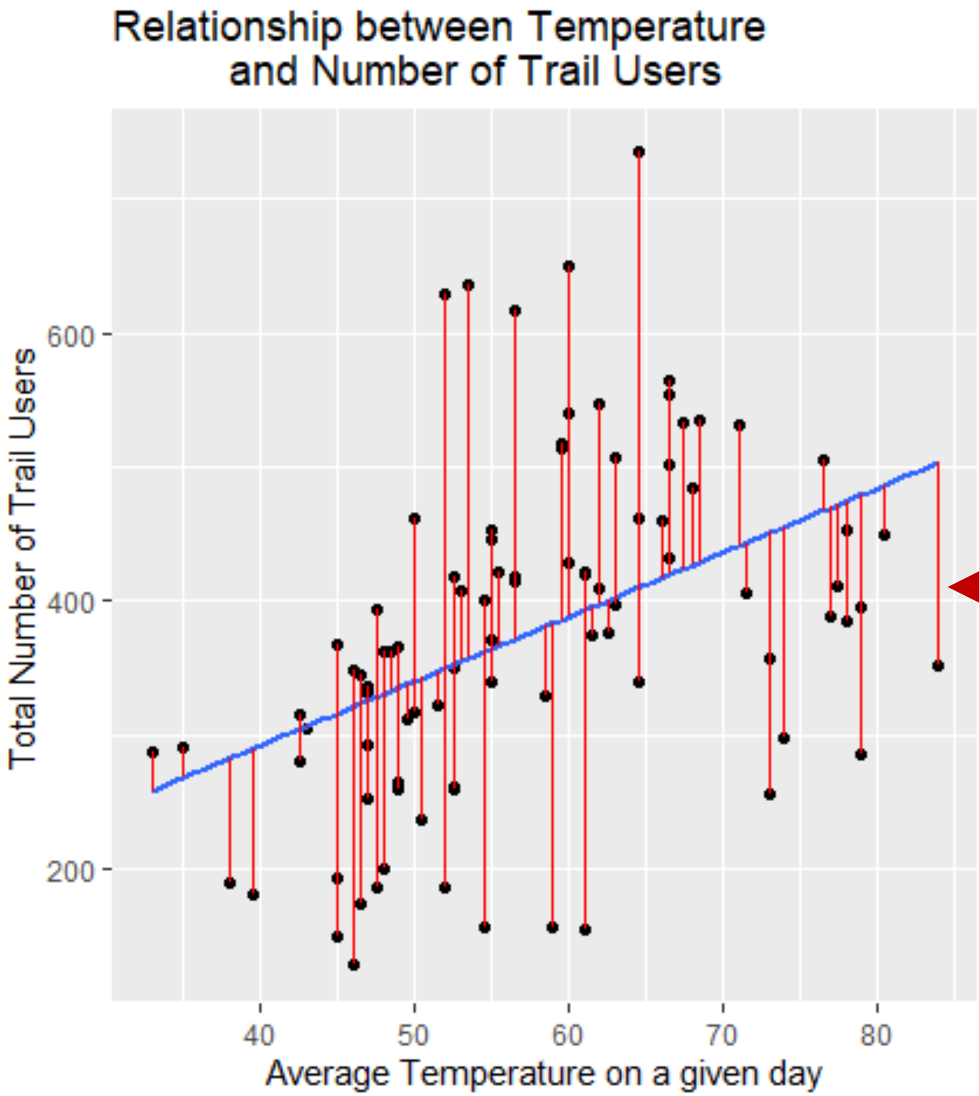
line of
best fit:

Least squares
regression
line

Scatter Plot

line that minimizes errors

$$MSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{Y} - Y\right)^2}{n}}$$



error $(\hat{Y} - Y)$

Regression Output

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   99.602     63.474   1.569     0.12    
avgtemp        4.802      1.084   4.428 2.72e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 115.9 on 88 degrees of freedom
Multiple R-squared:  0.1822,	Adjusted R-squared:  0.1729 
F-statistic: 19.61 on 1 and 88 DF,  p-value: 2.723e-05
```

Least Squares Regression Equation:
Y=99.6 + 4.802X

That is,
Volume=99.6+4.902(AvgTemp)

As average temperature increases by 1 degree F, we expect the number of trail users to increase by 4.9 people.

If the average temperature is 0 degrees F, we expect there to be 99.6 people using the trail on average.

If it is 80 degrees F outside, how many people do we expect to be using this trail?