# Example: Ride share data analysis

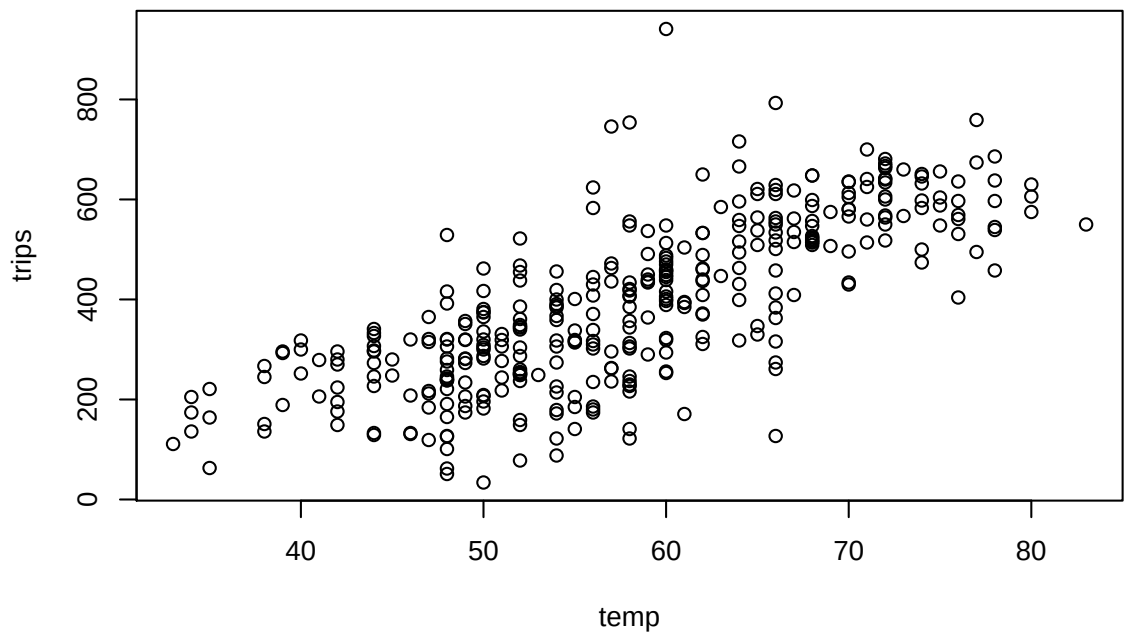
Peter Hoff

STAT 423

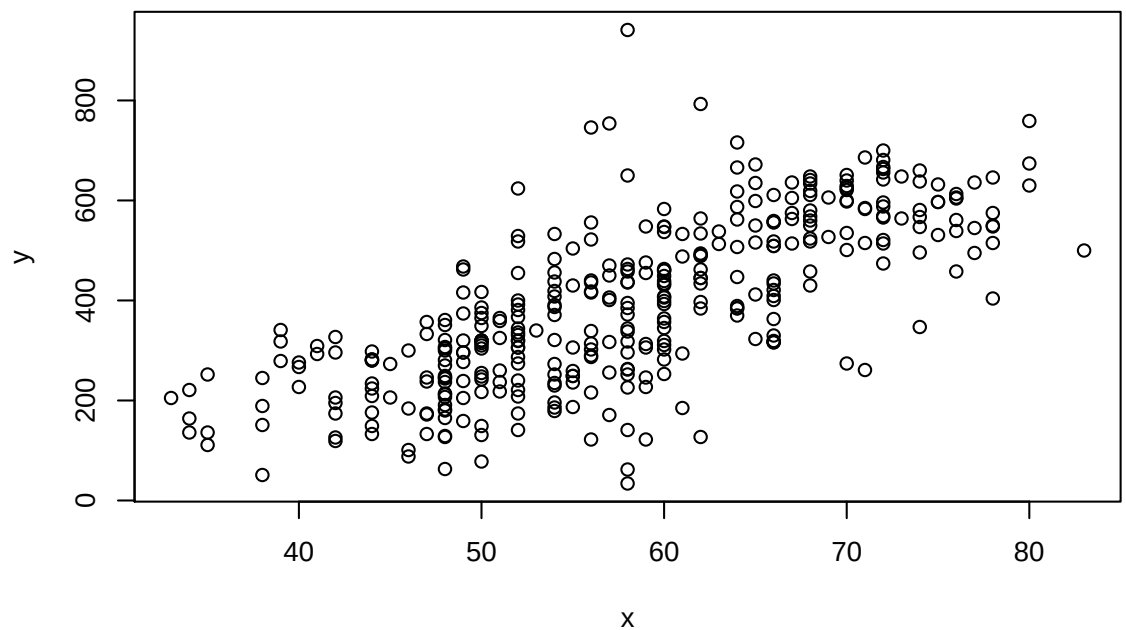
Applied Regression and Analysis of Variance

University of Washington

```r
temp<-weather$Mean_Temperature_F
trips<-weather$ttrips

cor(temp,trips)


## [1] 0.7542428
```



```r
x<-temp[-length(temp)]
y<-trips[-1]

cor(x,y)


## [1] 0.7239651
```



```r
fit<-lm(y~x)

summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -358.59  -74.64    8.42   71.70  548.41
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -268.4622    33.5743  -7.996 1.75e-14 ***
## x             11.3975     0.5708  19.968  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 113.6 on 362 degrees of freedom
## Multiple R-squared:  0.5241, Adjusted R-squared:  0.5228
## F-statistic: 398.7 on 1 and 362 DF,  p-value: < 2.2e-16
```

```r
sqrt( sum(fit$res^2)/(n-2) )
```

```
## [1] 113.5972
```

```r
sqrt( sum(fit$res^2)/(n-2) )
```

```
## [1] 113.5972
```

```r
s2<-sum(fit$res^2)/(n-2)

sqrt(s2/sum( (x-mean(x))^2 ) )
```

```
## [1] 0.5708012
```

```r
sqrt( sum(fit$res^2)/(n-2) )
```

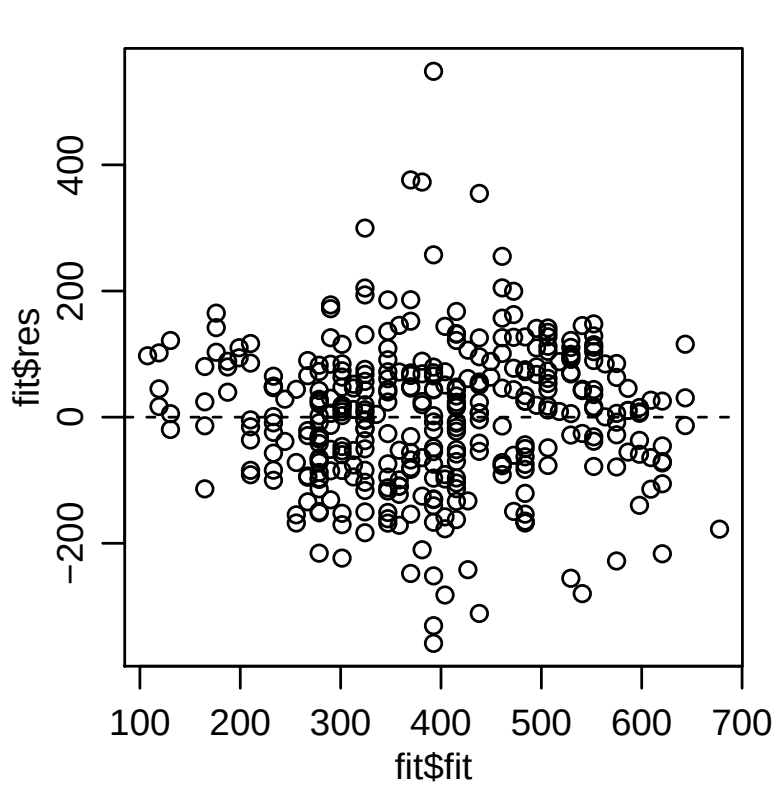
```
## [1] 113.5972
```

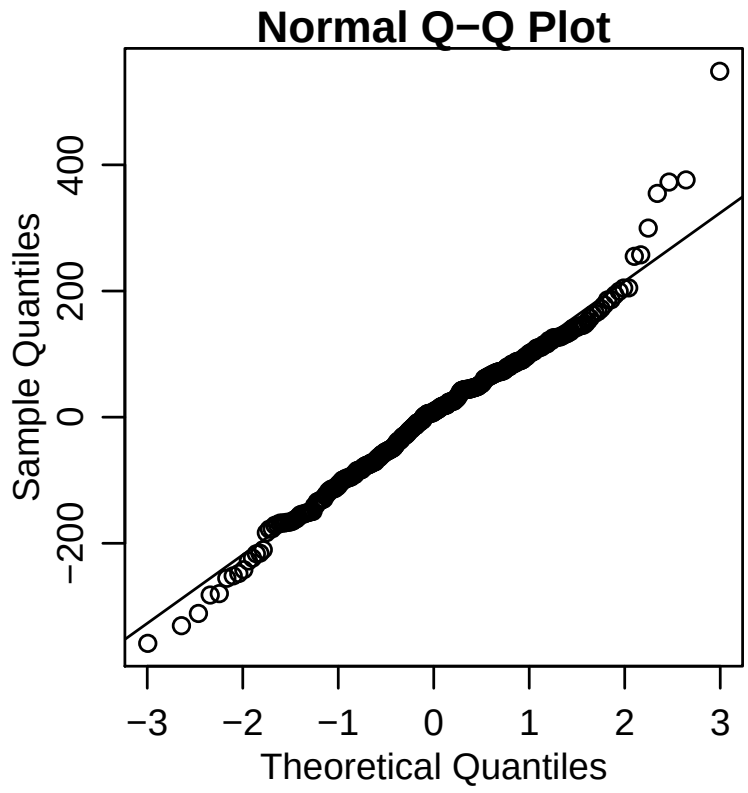
```r
s2<-sum(fit$res^2)/(n-2)

sqrt(s2/sum( (x-mean(x))^2 ) )
```

```
## [1] 0.5708012
```

```r
se_b1<-sqrt(s2/sum( (x-mean(x))^2 ) )
```

```r
plot(fit$res~fit$fit) ; abline(h=0,lty=2)

qqnorm(fit$res) ; qqline(fit$res)
```



```
fit$coef
```

```
## (Intercept)           x
##  -268.46222    11.39753
```

```
b1_ols<-fit$coef[2]
```

```r
fit$coef
```

```
## (Intercept)           x
##  -268.46222    11.39753
```

```r
b1_ols<-fit$coef[2]
```

```r
qt_alpha<-qt(.975,n-2)

qt_alpha
```

```
## [1] 1.966539
```

```r
fit$coef
```

```
## (Intercept)           x
##  -268.46222    11.39753
```

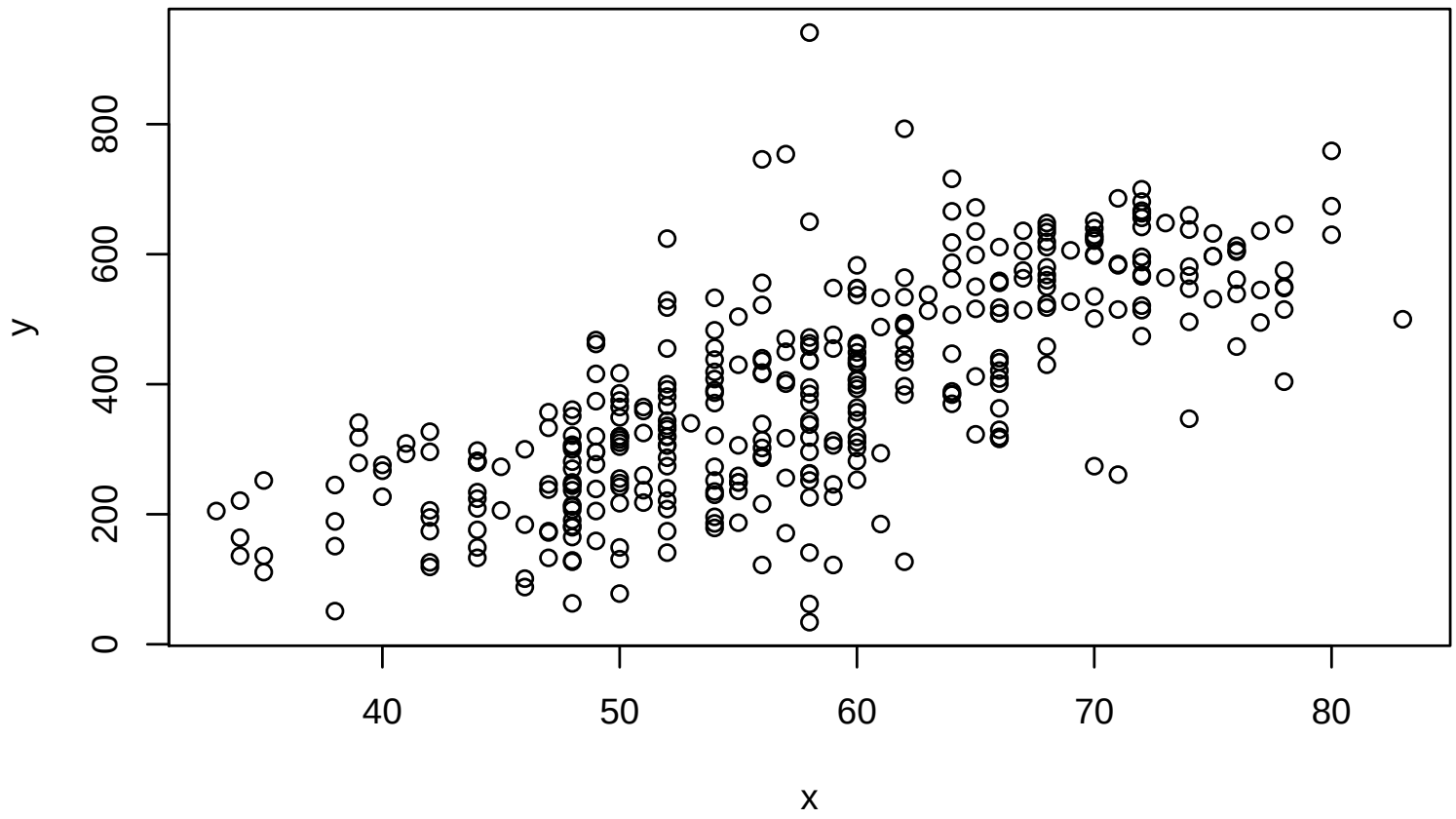
```r
b1_ols<-fit$coef[2]
```

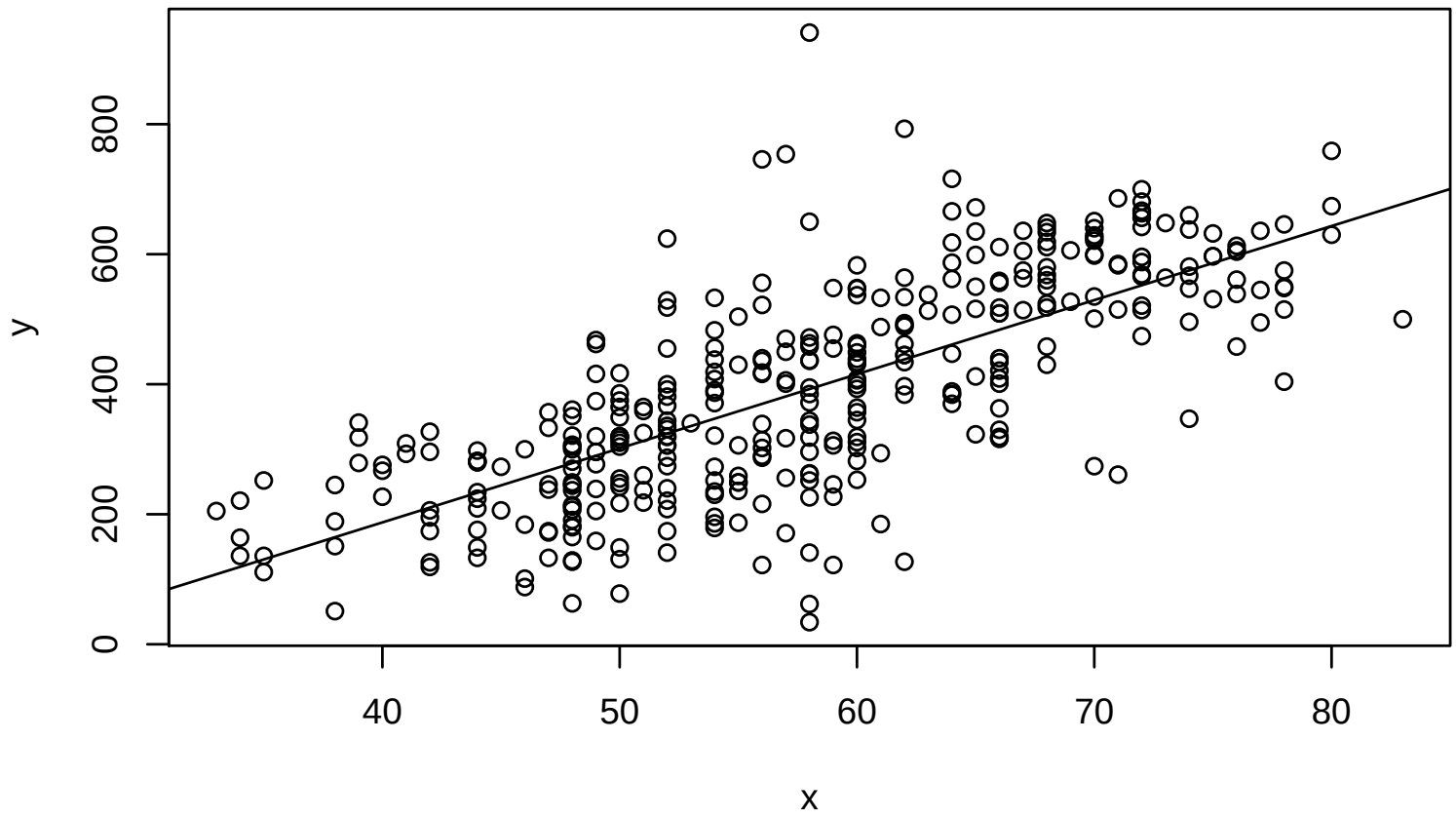
```r
qt_alpha<-qt(.975,n-2)

qt_alpha
```

```
## [1] 1.966539
```

```r
ci_b1<- b1_ols + c(-1,1)*qt_alpha*se_b1

ci_b1
```

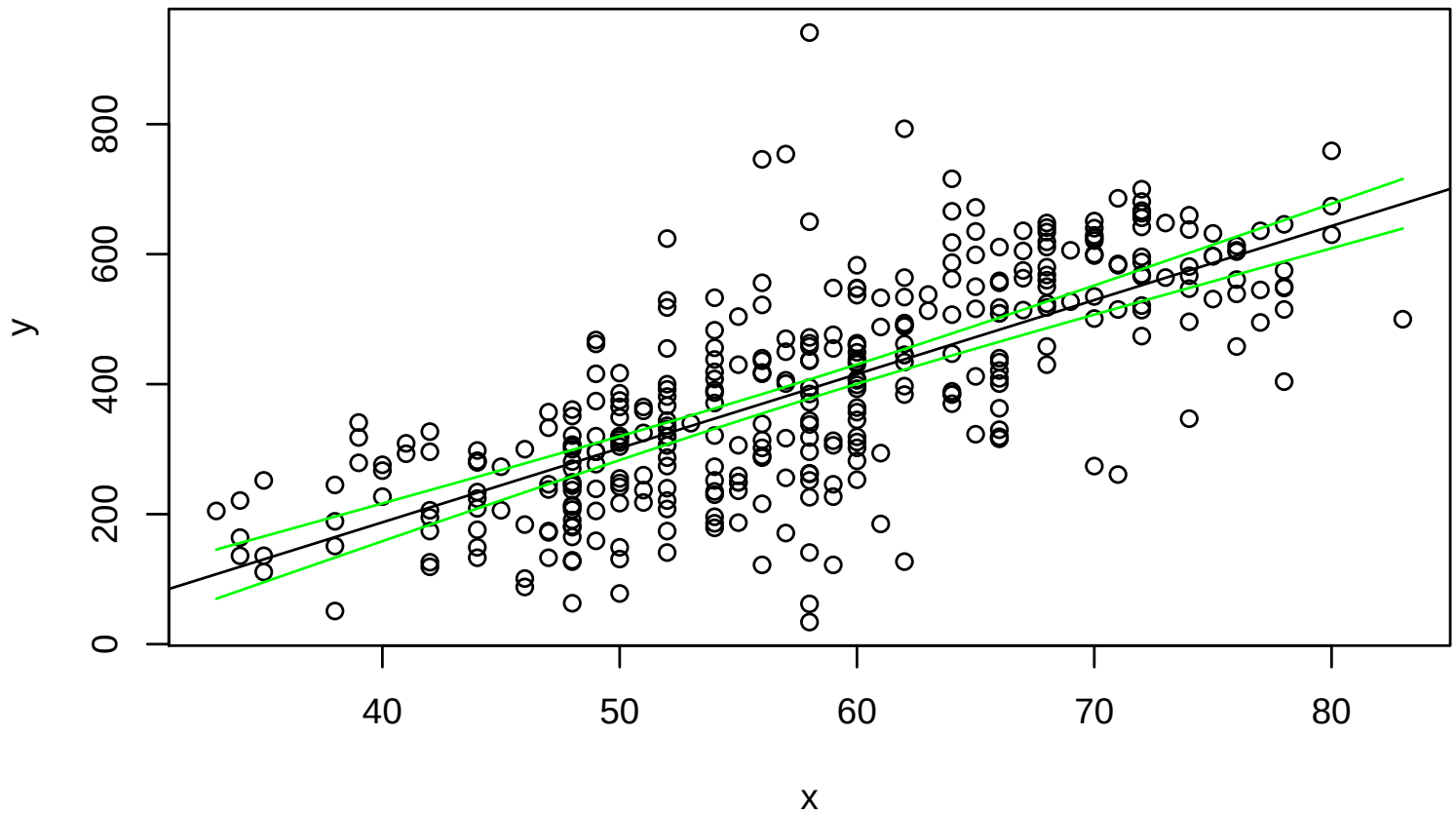
```
## [1] 10.27502 12.52003
```

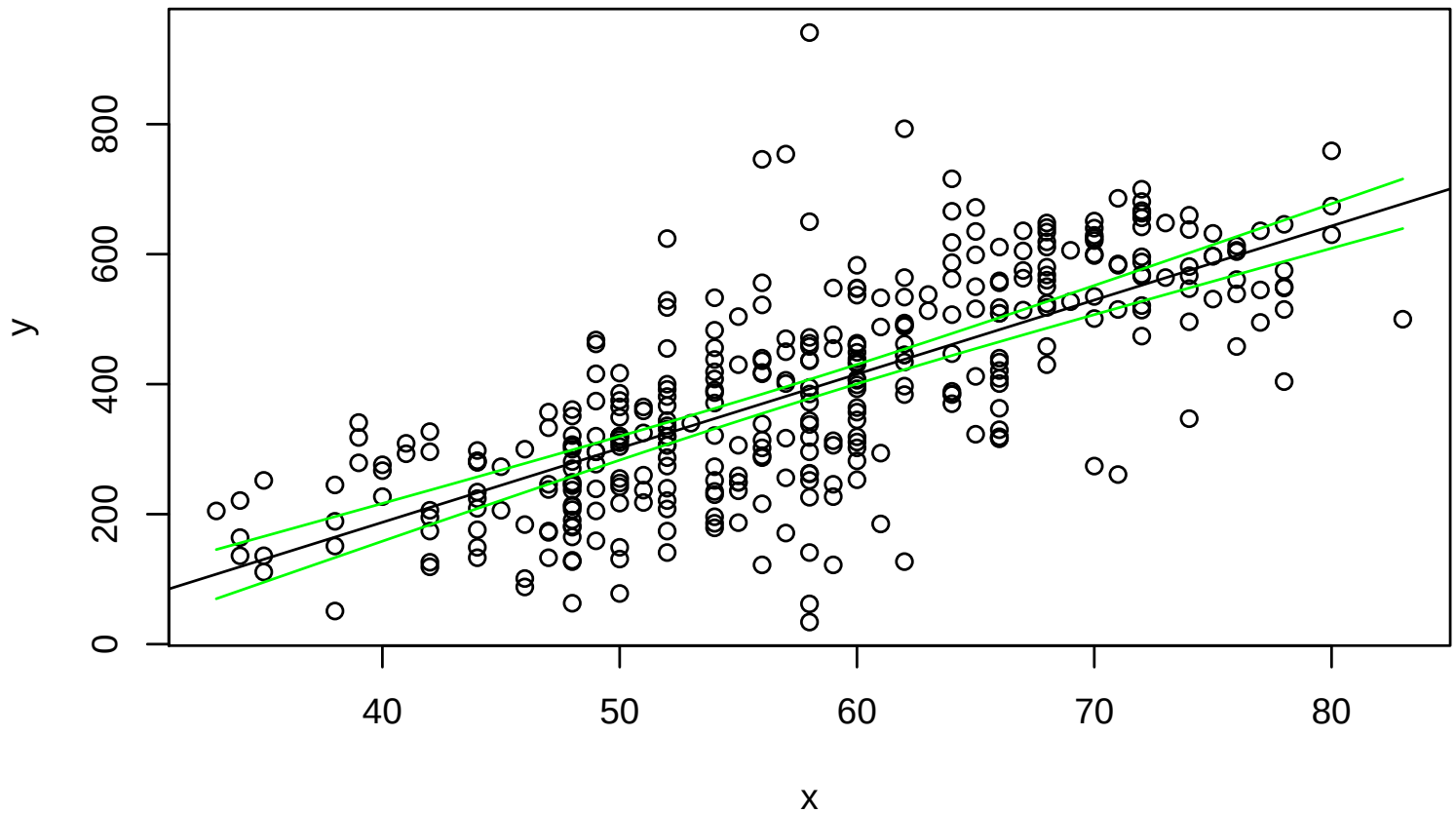

y
0
200
400
600
800
40
50
60
70
80
x


y
0
200
400
600
800
40
50
60
70
80
x

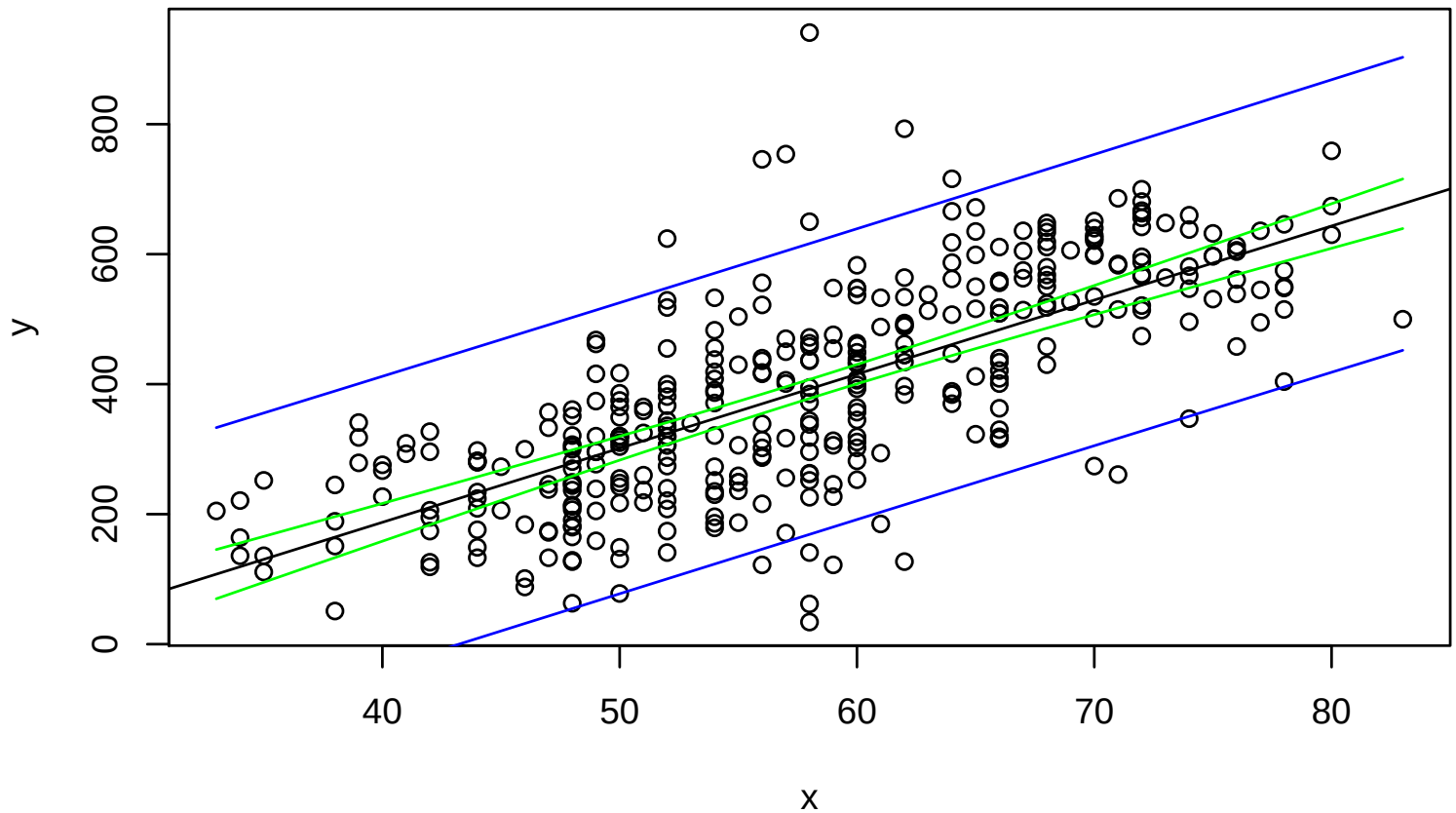
```
sigma<-sqrt( sum(fit$res^2)/(n-2) )
xbar<-mean(x)
SXX<-sum( (x-xbar)^2 )
```

```
xseq<-seq(min(x),max(x),by=1)

fit_x<-fit$coef[1] + fit$coef[2]*xseq

se_fit<- sigma*sqrt( 1/n + (xseq-xbar)^2/SXX)

fit_bound<-cbind( fit_x-se_fit*sqrt( (2*qf(.95,2,n-2) ) ),
                  fit_x+se_fit*sqrt( (2*qf(.95,2,n-2) ) ) )
```


y
0
200
400
600
800
40
50
60
70
80
x


y
0
200
400
600
800
40
50
60
70
80
x

```
se_prd<- sigma*sqrt(1+ 1/n + (xseq-xbar)^2/SXX)

prd_bound<-cbind( fit_x-se_prd*qt(.975,n-2),
                  fit_x+se_prd*qt(.975,n-2) )
```


y
0
200
400
600
800
40
50
60
70
80
x

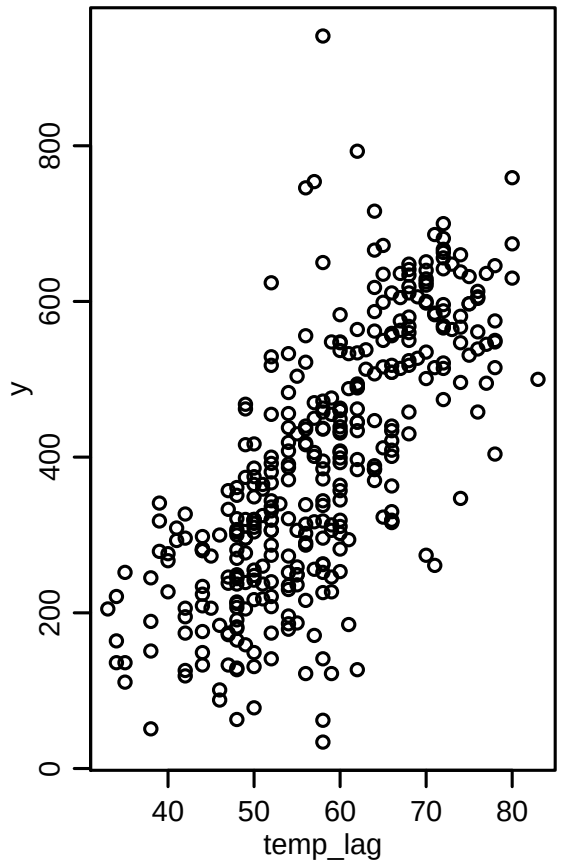
```r
in_ci<-0

for(xs in xseq)
{
  in_ci<-in_ci + sum( prd_bound[xseq==xs,1]<=y[x==xs] &
                      y[x==xs]<=prd_bound[xseq==xs,2] )
}
```

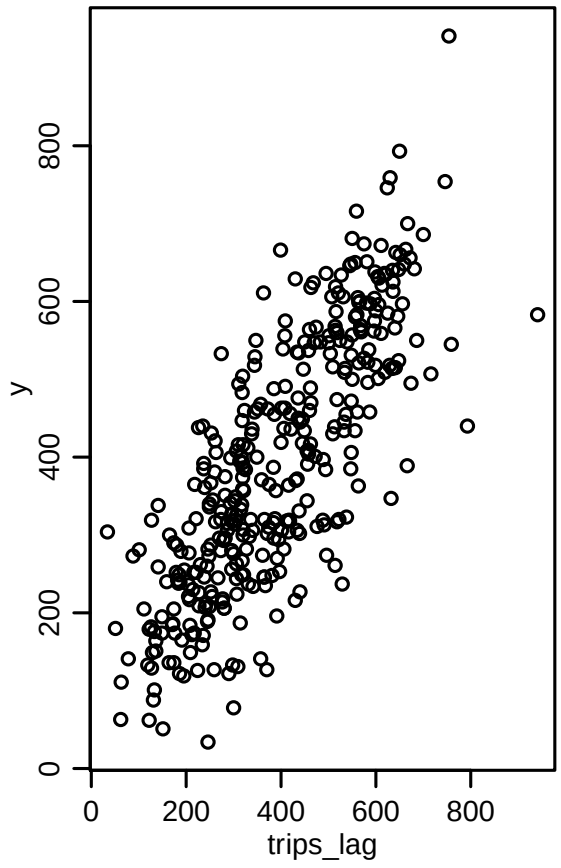
```r
in_ci<-0

for(xs in xseq)
{
  in_ci<-in_ci + sum( prd_bound[xseq==xs,1]<=y[x==xs] &
                      y[x==xs]<=prd_bound[xseq==xs,2] )
}
```

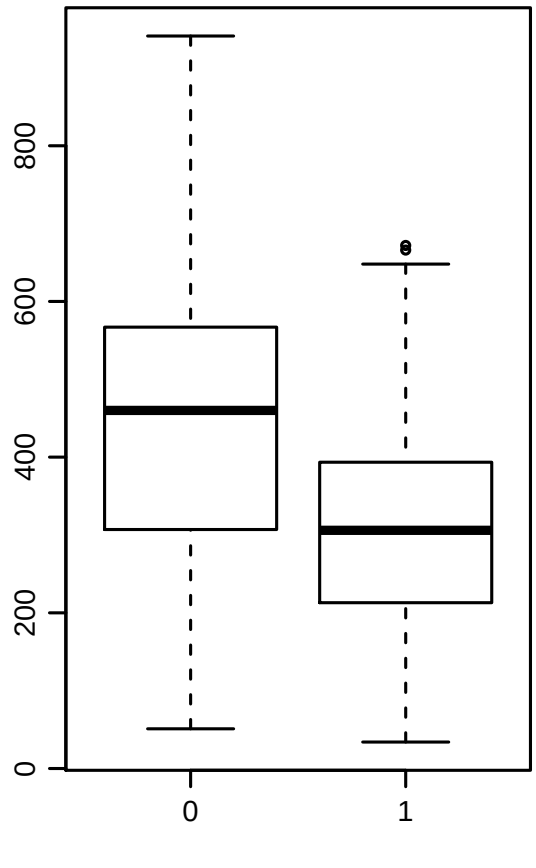
```r
in_ci/n


## [1] 0.9532967
```

```
temp_lag<-x

trips_lag<-trips[ -nrow(weather) ]

rain_lag<-1 * ( weather$Precipitation_In[ -nrow(weather) ]  >0 )
```

```r
temp_lag<-x

trips_lag<-trips[ -nrow(weather) ]

rain_lag<-1 * ( weather$Precipitation_In[ -nrow(weather) ]  >0 )
```



```r
fit1<-lm(y~temp_lag)

sqrt( sum(fit1$res^2)/(n-2) )

## [1] 113.5972
```

```r
fit1<-lm(y~temp_lag)

sqrt( sum(fit1$res^2)/(n-2) )

## [1] 113.5972
```

```r
fit2<-lm(y~temp_lag + trips_lag + rain_lag)

sqrt( sum(fit2$res^2)/(n-4) )

## [1] 94.72035
```

```r
fit1<-lm(y~temp_lag)

sqrt( sum(fit1$res^2)/(n-2) )

## [1] 113.5972
```

```r
fit2<-lm(y~temp_lag + trips_lag + rain_lag)

sqrt( sum(fit2$res^2)/(n-4) )

## [1] 94.72035
```

```r
2*( sqrt( sum(fit1$res^2)/(n-2) )  - sqrt( sum(fit2$res^2)/(n-4) ) )

## [1] 37.7536
```

```
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ temp_lag + trips_lag + rain_lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -268.73  -60.90    6.14   58.48  350.85
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -91.35368   31.31021  -2.918  0.00375 **
## temp_lag      4.97449    0.75868   6.557 1.91e-10 ***
## trips_lag     0.52120    0.05446   9.570  < 2e-16 ***
## rain_lag    -23.59353   12.30028  -1.918  0.05589 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 94.72 on 360 degrees of freedom
## Multiple R-squared:  0.671, Adjusted R-squared:  0.6682
## F-statistic: 244.7 on 3 and 360 DF,  p-value: < 2.2e-16
```