HW8

Thulasiram Ruppa Krishnan

March 10, 2019

```
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(bitops)
library(gdata)
```

```
## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location
## gdata: of a valid perl intrpreter.
## gdata:
## gdata: (To avoid display of this message in the future, please
## gdata: ensure perl is installed and available on the executable
## gdata: search path.)
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.
```

```
##
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
```

```
##
```

```
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.
```

```
##
## Attaching package: 'gdata'
```

```
## The following object is masked from 'package:stats':
##
##     nobs
```

```
## The following object is masked from 'package:utils':
##
##     object.size
```

```
## The following object is masked from 'package:base':
##
##     startsWith
```

```
library(readxl)
library(ggplot2)
```

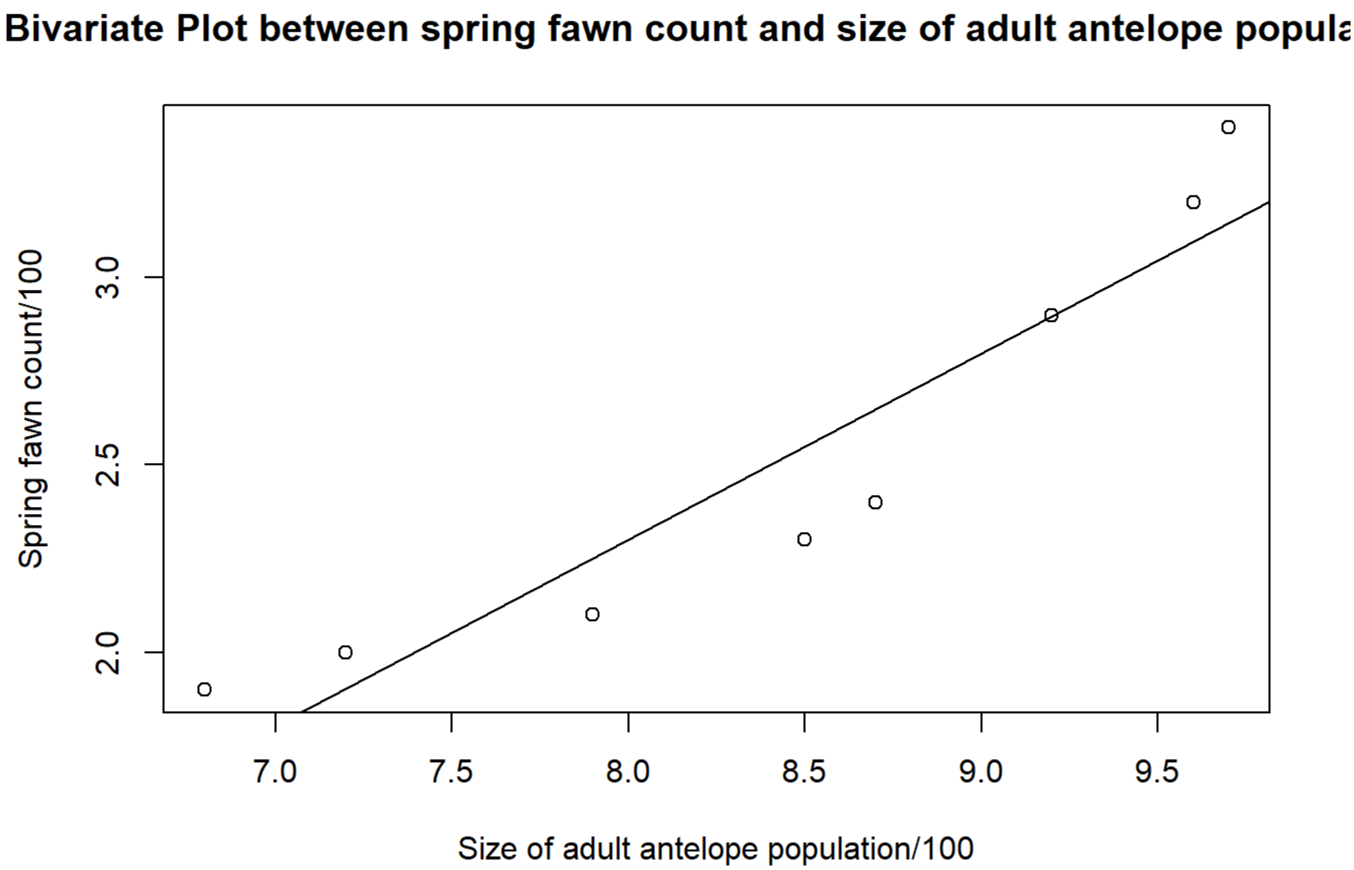
```
##Thunder Basin Antelope Study

##The data (X1, X2, X3, X4) are for each year.
##X1 = spring fawn count/100
##X2 = size of adult antelope population/100
##X3 = annual precipitation (inches)
##X4 = winter severity index (1=mild,5=severe)
mlr <- read_excel("C:/Users/rkrishnan/Documents/01 Personal/MS/IST 687/mlr01.xls")
colnames(mlr) <-c("fawn_count","antelope_size","annual_precipitation","winter_severity")
str(mlr)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    8 obs. of  4 variables:
##  $ fawn_count          : num  2.9 2.4 2 2.3 3.2 ...
##  $ antelope_size       : num  9.2 8.7 7.2 8.5 9.6 ...
##  $ annual_precipitation: num  13.2 11.5 10.8 12.3 12.6 ...
##  $ winter_severity     : num  2 3 4 2 3 5 1 3
```

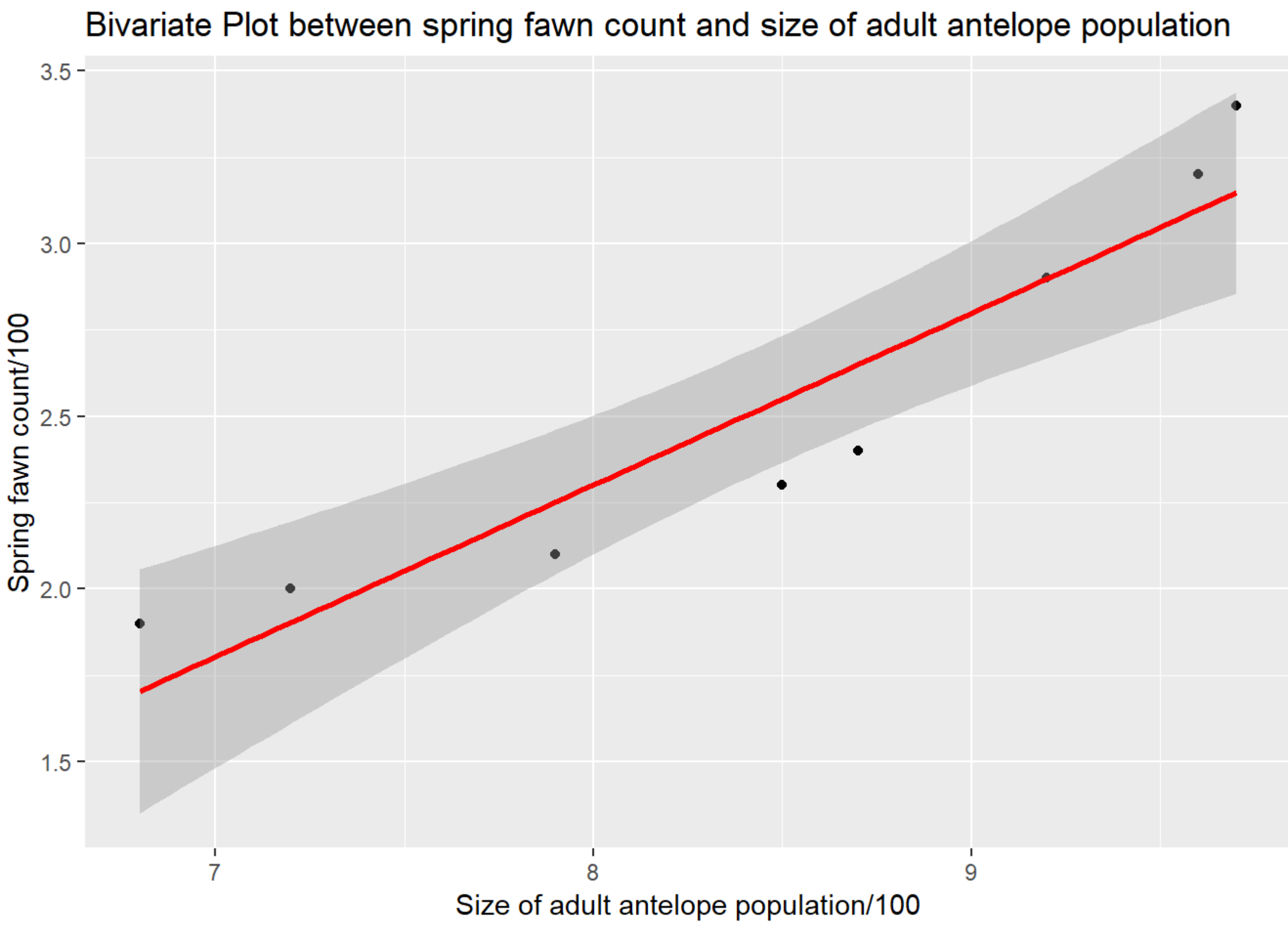
```
summary(mlr)
```

```
##    fawn_count    antelope_size   annual_precipitation winter_severity
##  Min.   :1.900   Min.   :6.800   Min.   :10.60        Min.   :1.000
##  1st Qu.:2.075   1st Qu.:7.725   1st Qu.:11.10        1st Qu.:2.000
##  Median :2.350   Median :8.600   Median :11.90        Median :3.000
##  Mean   :2.525   Mean   :8.450   Mean   :12.04        Mean   :2.875
##  3rd Qu.:2.975   3rd Qu.:9.300   3rd Qu.:12.75        3rd Qu.:3.250
##  Max.   :3.400   Max.   :9.700   Max.   :14.10        Max.   :5.000
```

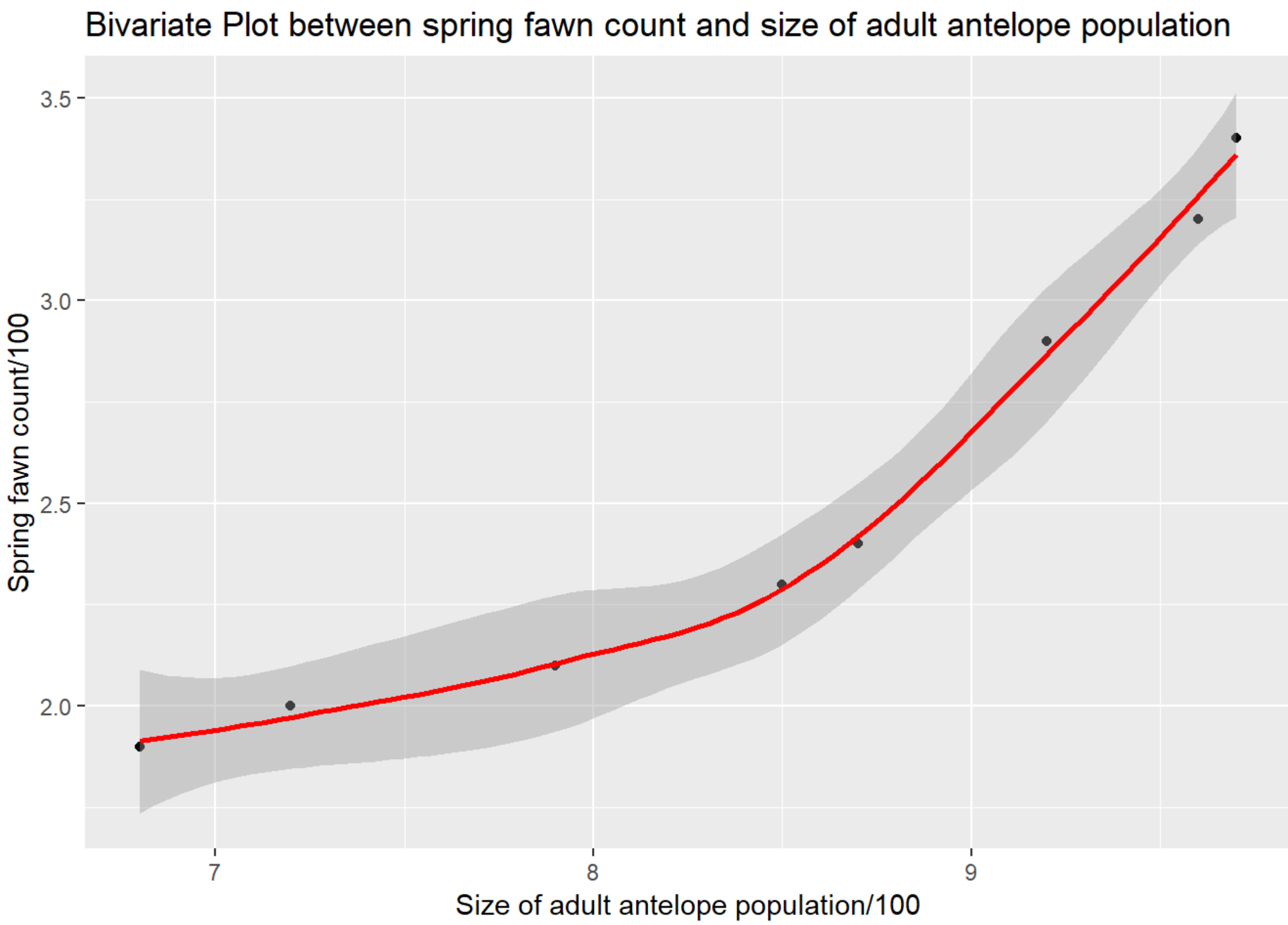
```
# Bivariate plot between fawn_count and other variables (antelope_size,precipitation,winter_seve
rity)

# Let's plot the relationship between  antelope_size and fawn_count
lm_apVsfc <- lm(formula = mlr$fawn_count ~ mlr$antelope_size,data=mlr)
plot(mlr$antelope_size,mlr$fawn_count,main ="Bivariate Plot between spring fawn count and size o
f adult antelope population",xlab = "Size of adult antelope population/100",ylab = "Spring fawn
 count/100",type = "p"  )
abline(lm_apVsfc)
```



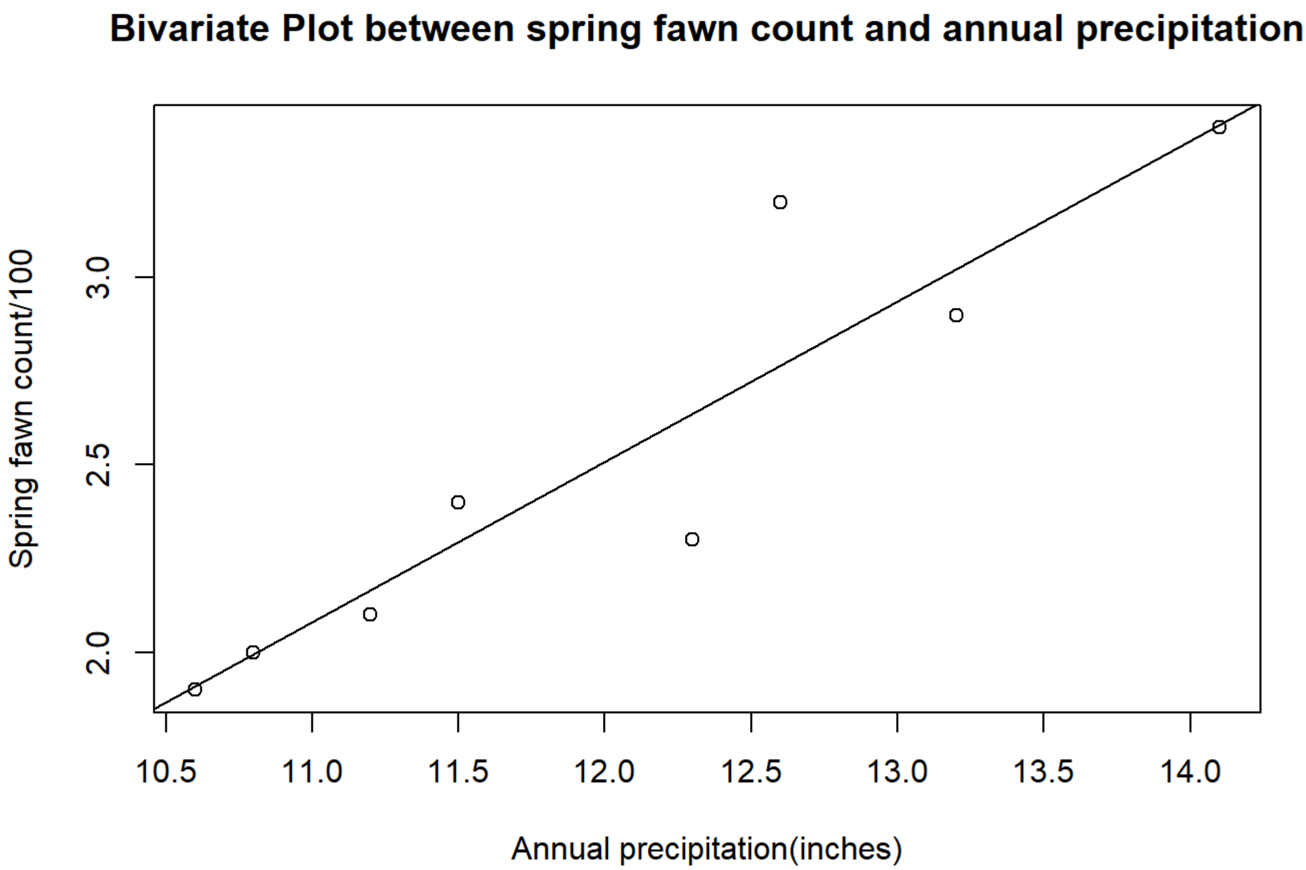

```
# how they look in ggplot  using lm
ggplot(data = mlr,aes(x=mlr$antelope_size,y=mlr$fawn_count))+labs(x="Size of adult antelope popu
lation/100",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and size
of adult antelope population") + geom_point() + geom_smooth(method="lm", color="red",size=1)
```



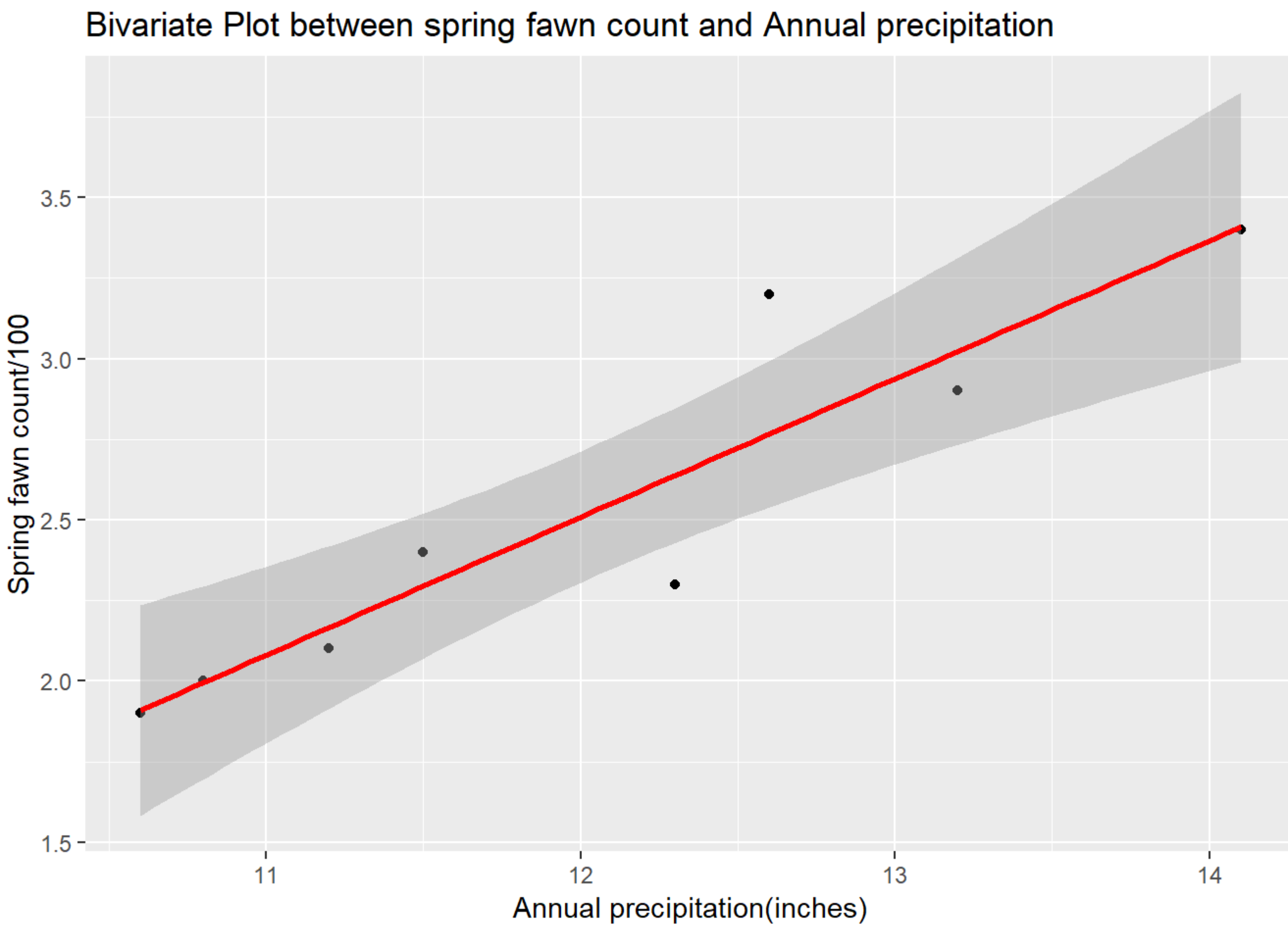

```
# how they look in ggplot using loess
ggplot(data = mlr,aes(x=mlr$antelope_size,y=mlr$fawn_count))+labs(x="Size of adult antelope popu
lation/100",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and size
of adult antelope population") + geom_point() + geom_smooth(method="loess", color="red",size=1)
```



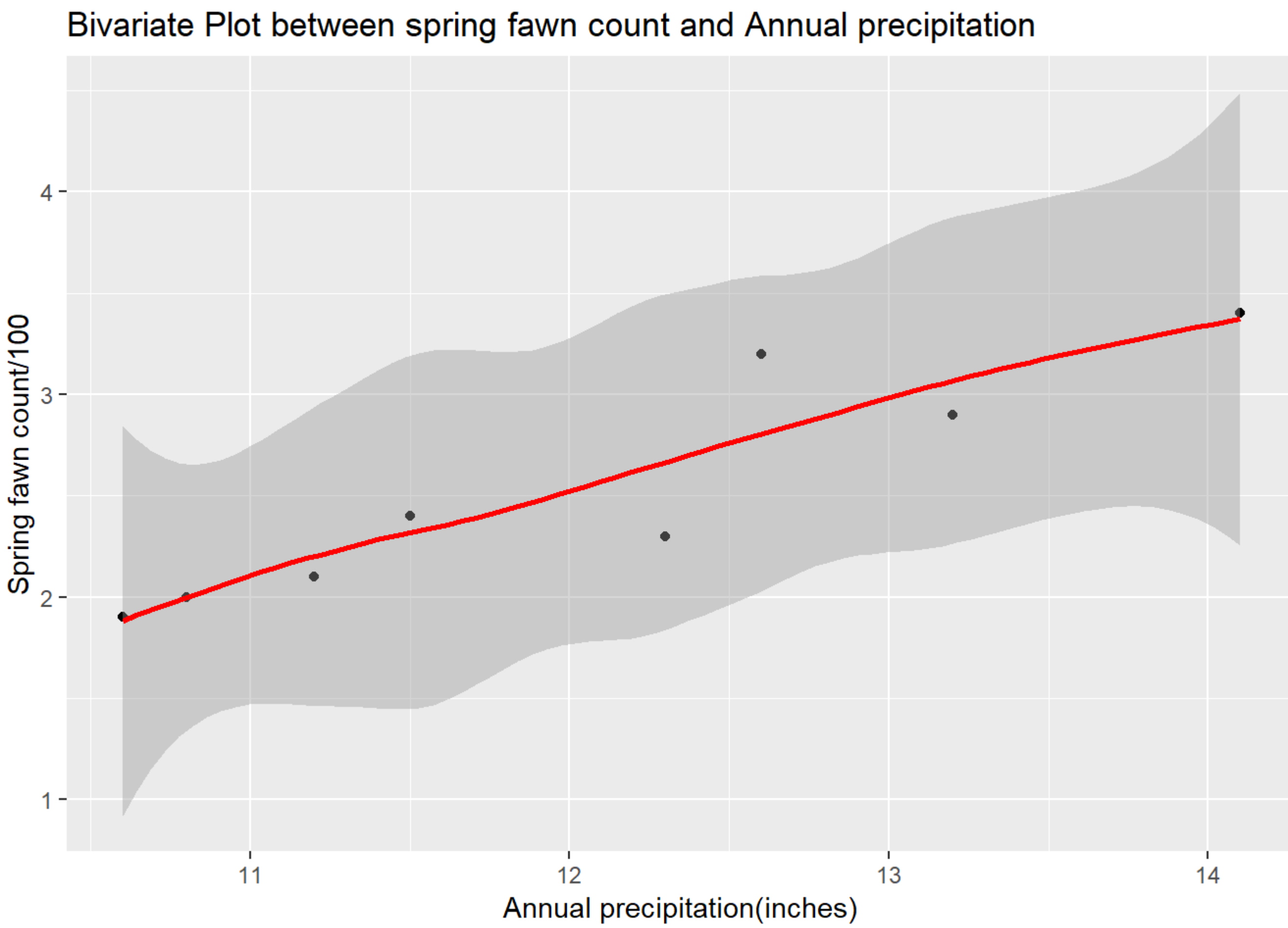

```
# Let's plot the relationship between Annual precipitation and fawn_count
lm_pVsfc <- lm(formula = mlr$fawn_count ~ mlr$annual_precipitation,data=mlr)

plot(mlr$annual_precipitation,mlr$fawn_count,main ="Bivariate Plot between spring fawn count and
annual precipitation",xlab = "Annual precipitation(inches)",ylab = "Spring fawn count/100",type
 = "p"  )
abline(lm_pVsfc)
```



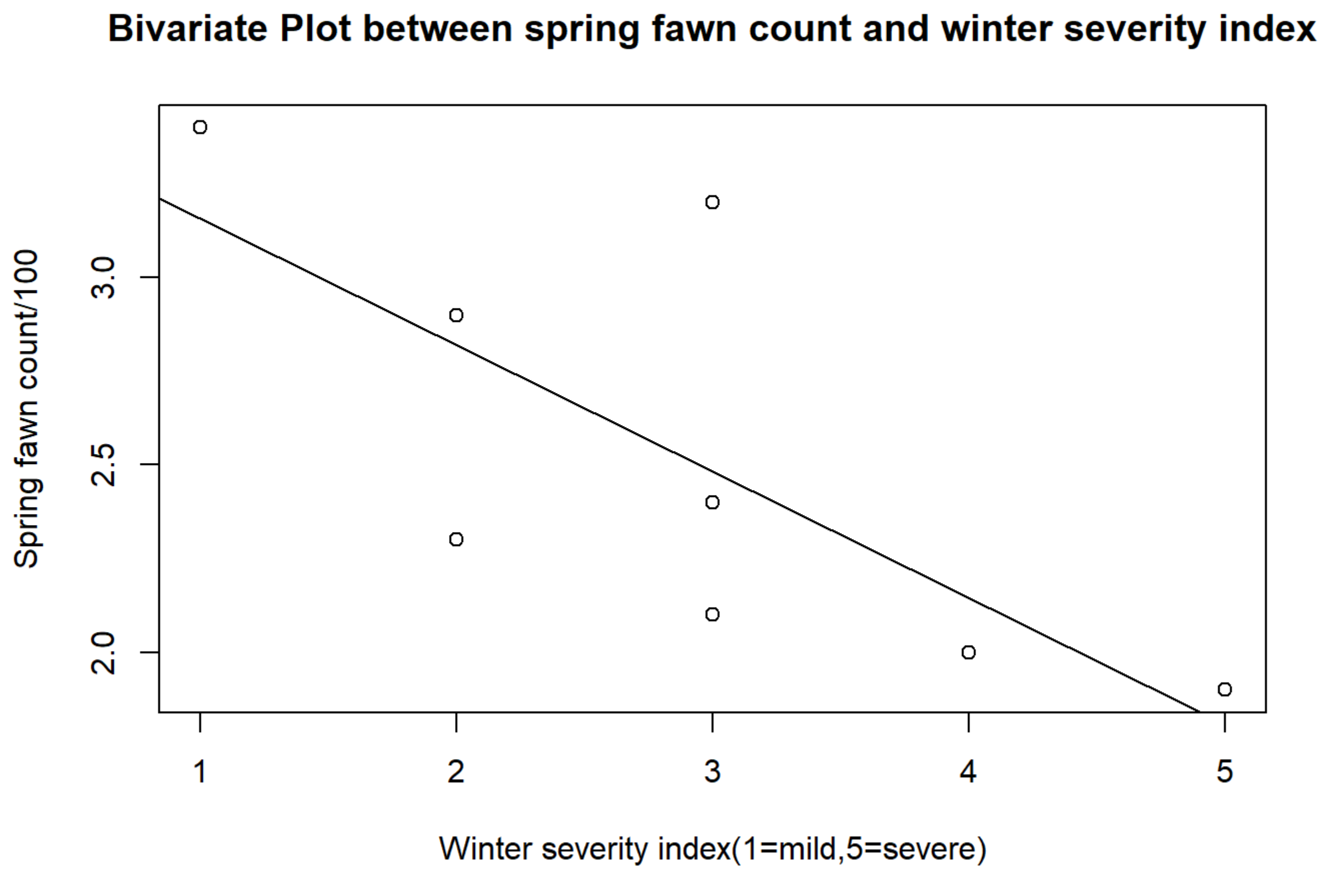

```
# how they look in ggplot  using lm
ggplot(data = mlr,aes(x=mlr$annual_precipitation,y=mlr$fawn_count))+labs(x="Annual precipitation
(inches)",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and Annual
precipitation") + geom_point() + geom_smooth(method="lm", color="red",size=1)
```



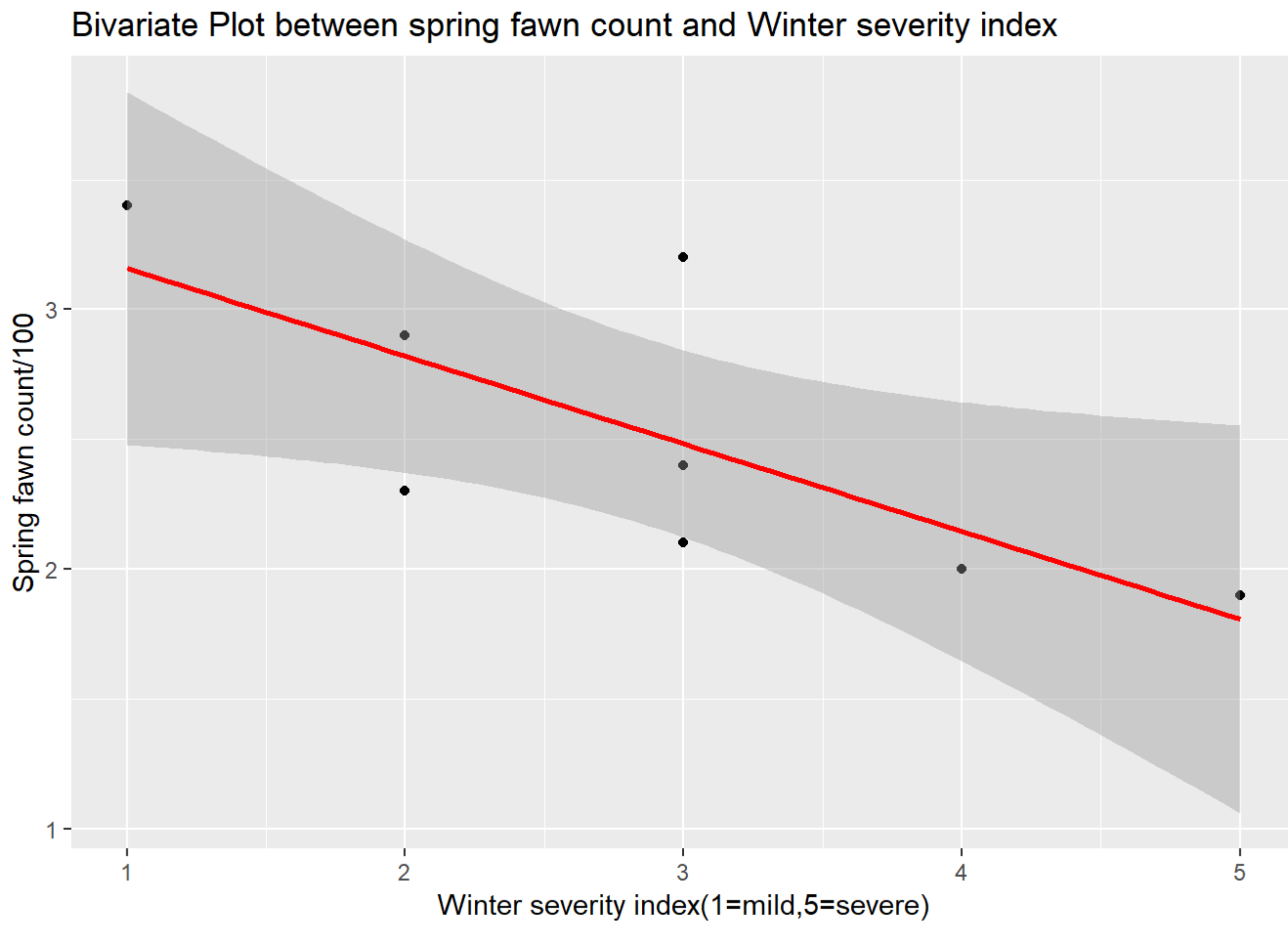

```
# how they look in ggplot using loess
ggplot(data = mlr,aes(x=mlr$annual_precipitation,y=mlr$fawn_count))+labs(x="Annual precipitation
(inches)",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and Annual
precipitation") + geom_point() + geom_smooth(method="loess", color="red",size=1)
```




```
# Let's plot the relationship between winter severity index(1=mild,5=severe) and fawn_count
lm_wiVsfc <- lm(formula = mlr$fawn_count ~ mlr$winter_severity,data=mlr)


plot(mlr$winter_severity,mlr$fawn_count,main ="Bivariate Plot between spring fawn count and wint
er severity index",xlab = "Winter severity index(1=mild,5=severe)",ylab = "Spring fawn count/10
0",type = "p"  )
abline(lm_wiVsfc)
```




```
# how they look in ggplot  using lm
ggplot(data = mlr,aes(x=mlr$winter_severity,y=mlr$fawn_count))+labs(x="Winter severity index(1=mild,5=severe)",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and Winter severity index") + geom_point() + geom_smooth(method="lm", color="red",size=1)
```




```
# how they look in ggplot using loess
ggplot(data = mlr,aes(x=mlr$winter_severity,y=mlr$fawn_count))+labs(x="Winter severity index(1=m
ild,5=severe)",y="Spring fawn count/100",title = "Bivariate Plot between spring fawn count and W
inter severity index") + geom_point() + geom_smooth(method="loess", color="red",size=1)
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.98
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 2.02
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 2.3712e-017
```

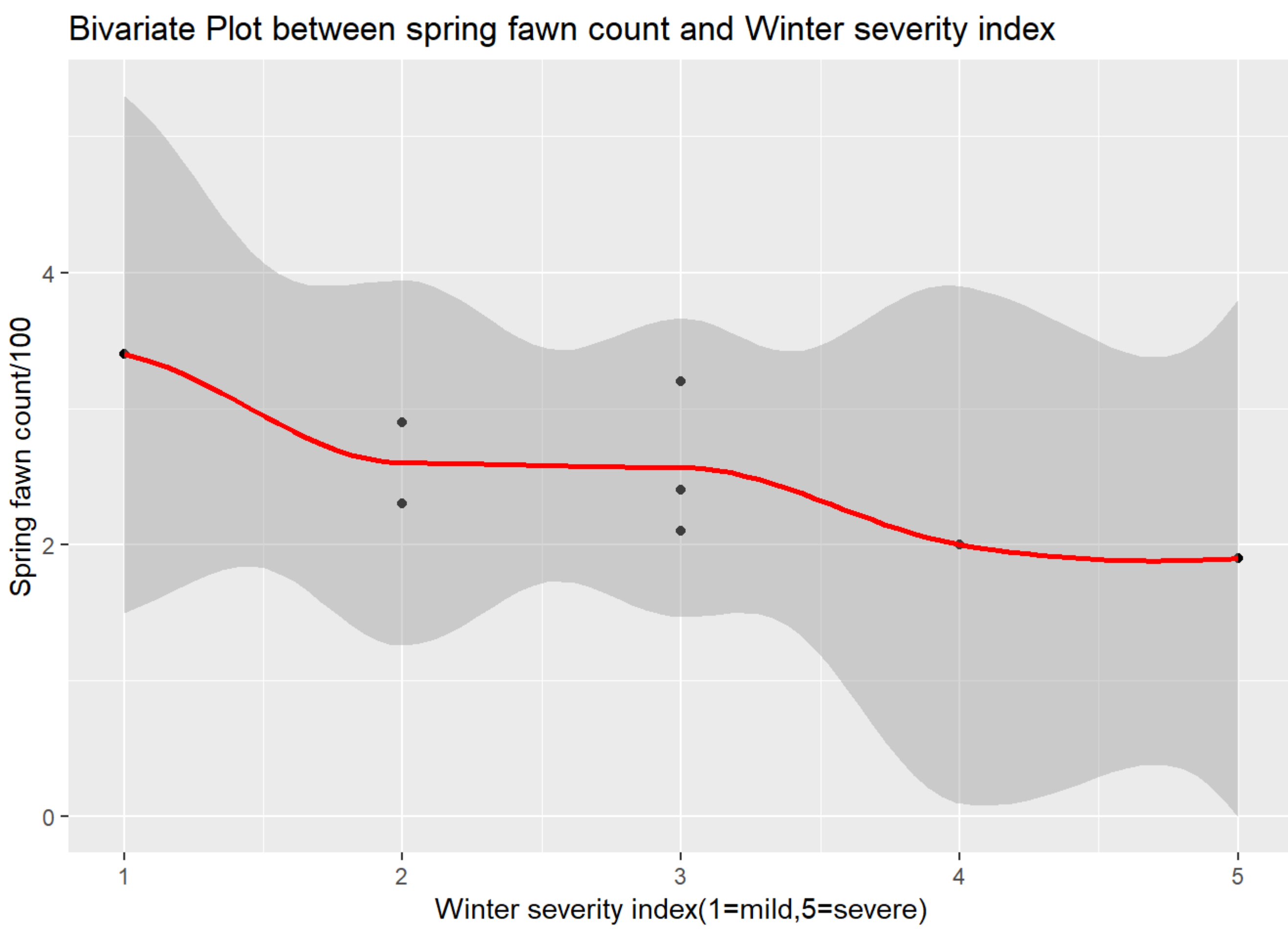
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 0.98
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 2.02
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 2.3712e-017
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other
## near singularities as well. 1
```




```
# Regression Model,     predict the number  of  fawns   from    Antelope Size
sum.lm_apVsfc <- summary(lm_apVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_apVsfc$coef[,4]
```

```
##       (Intercept) mlr$antelope_size
##      0.0381522675      0.0005471422
```

```
paste("Adjusted R squared: " ,sum.lm_apVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.861563784952523"
```

```
# Regression Model,     predict the number  of  fawns   from    Annual Precipitation
sum.lm_pVsfc <- summary(lm_pVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_pVsfc$coef[,4]
```

```
##              (Intercept) mlr$annual_precipitation
##              0.023840288              0.001039425
```

```
paste("Adjusted R squared: " ,sum.lm_pVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.829212576293043"
```

```
# Regression Model,     predict the number  of  fawns   from Winter Severity Index
sum.lm_wiVsfc <- summary(lm_wiVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_wiVsfc$coef[,4]
```

```
##         (Intercept) mlr$winter_severity
##         0.000108158         0.036263036
```

```
paste("Adjusted R squared: " ,sum.lm_wiVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.47020333641798"
```

```
# Let's predict fawn count from two variables antelope_size and annual_precipitation
lm_ap.pVsfc <- lm(formula = mlr$fawn_count ~ mlr$antelope_size + mlr$annual_precipitation,data=m
lr)
sum.lm_ap.pVsfc <- summary(lm_ap.pVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_ap.pVsfc$coef[,4]
```

```
##              (Intercept)       mlr$antelope_size mlr$annual_precipitation
##               0.02846521              0.12406648               0.23548196
```

```
paste("Adjusted R squared: " ,sum.lm_ap.pVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.878160928538559"
```

```
# Addition of one more variable to annual_precipitation has increased the model accuracy from 8
6% to 87% from the original model with just antelope size

# Let's predict fawn count from two variables antelope_size and winter_index

lm_wi.pVsfc <- lm(formula = mlr$fawn_count ~ mlr$antelope_size + mlr$winter_severity,data=mlr)
sum.lm_wi.pVsfc <- summary(lm_wi.pVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_wi.pVsfc$coef[,4]
```

```
##         (Intercept)   mlr$antelope_size mlr$winter_severity
##          0.16977988          0.01118699          0.59557987
```

```
paste("Adjusted R squared: " ,sum.lm_wi.pVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.84389367311556"
```

```
# Addition of winter index to annual_precipitation has dropped the model accuracy from 86% to 8
4% from the original model with just antelope size

# Let's predict fawn count from two variables annual_precipitation and winter_index

lm_wi.apVsfc <- lm(formula = mlr$fawn_count ~ mlr$annual_precipitation + mlr$winter_severity,dat
a=mlr)
sum.lm_wi.apVsfc <- summary(lm_wi.apVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_wi.apVsfc$coef[,4]
```

```
##              (Intercept) mlr$annual_precipitation      mlr$winter_severity
##              0.047647946              0.008431877              0.188417206
```

```
paste("Adjusted R squared: " ,sum.lm_wi.apVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.859962754071404"
```

```
# Addition of winter index to annual_precipitation has dropped the model accuracy from 86% to 8
4% from the original model with just antelope size

# Let's predict fawn count from three variables antelope_size , annual_precipitation and winter
 index
lm_ap.p.wiVsfc <- lm(formula = mlr$fawn_count ~ mlr$antelope_size + mlr$annual_precipitation + m
lr$winter_severity,data=mlr)
summary(lm_ap.p.wiVsfc)
```

```
##
## Call:
## lm(formula = mlr$fawn_count ~ mlr$antelope_size + mlr$annual_precipitation +
##     mlr$winter_severity, data = mlr)
##
## Residuals:
##        1        2        3        4        5        6        7        8
## -0.11533 -0.02661  0.09882 -0.11723  0.02734 -0.04854  0.11715  0.06441
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -5.92201    1.25562  -4.716   0.0092 **
## mlr$antelope_size         0.33822    0.09947   3.400   0.0273 *
## mlr$annual_precipitation  0.40150    0.10990   3.653   0.0217 *
## mlr$winter_severity       0.26295    0.08514   3.089   0.0366 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1209 on 4 degrees of freedom
## Multiple R-squared:  0.9743, Adjusted R-squared:  0.955
## F-statistic: 50.52 on 3 and 4 DF,  p-value: 0.001229
```

```
sum.lm_ap.p.wiVsfc <- summary(lm_ap.p.wiVsfc)
paste("p values:")
```

```
## [1] "p values:"
```

```
sum.lm_ap.p.wiVsfc$coef[,4]
```

```
##             (Intercept)        mlr$antelope_size mlr$annual_precipitation
##             0.009196072              0.027272444              0.021707219
##      mlr$winter_severity
##             0.036626174
```

```
paste("Adjusted R squared: " ,sum.lm_ap.p.wiVsfc$adj.r.squared)
```

```
## [1] "Adjusted R squared:  0.955004704934087"
```

```
# After we added the winter index to the model, accuracy has increased from 87% to 95% from the
 original model with just antelope size and annual_precipitation

#Compare all the models

cat(paste("One Variable Prediction ", "Adjusted R squared for Antelope Size : ",sum.lm_apVsfc$ad
j.r.squared, "Adjusted R squared for Annual Precipitation: ",sum.lm_pVsfc$adj.r.squared , "Adjus
ted R squared for Winter Index: ",sum.lm_wiVsfc$adj.r.squared, "Two Variable Prediction ", "Adju
sted R squared for Antelope Size and Annual Precipitation : ",sum.lm_ap.pVsfc$adj.r.squared, "Ad
justed R squared for Antelope Size and Winter Index : ",sum.lm_wi.pVsfc$adj.r.squared, "Adjusted
R squared for Annual Precipitaion and Winter Index : ",sum.lm_wi.apVsfc$adj.r.squared, "Three Va
riable Prediction ", "Adjusted R squared for Antelope Size, Annual Precipitation and Winter Inde
x : ",sum.lm_ap.p.wiVsfc$adj.r.squared,sep='\n'  ))
```

```
## One Variable Prediction
## Adjusted R squared for Antelope Size :
## 0.861563784952523
## Adjusted R squared for Annual Precipitation:
## 0.829212576293043
## Adjusted R squared for Winter Index:
## 0.47020333641798
## Two Variable Prediction
## Adjusted R squared for Antelope Size and Annual Precipitation :
## 0.878160928538559
## Adjusted R squared for Antelope Size and Winter Index :
## 0.84389367311556
## Adjusted R squared for Annual Precipitaion and Winter Index :
## 0.859962754071404
## Three Variable Prediction
## Adjusted R squared for Antelope Size, Annual Precipitation and Winter Index :
## 0.955004704934087
```

```
# The best predction with fewer variable is given by the model using Antelope Size and Annual Pr
ecipitation
```