hw8

```
library(gdata)
```

```
## Warning: package 'gdata' was built under R version 4.1.1
```

```
## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2
```

```
## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2
```

```
## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
```

```
##
```

```
## 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
```

```
require(readxl)
```

```
## Loading required package: readxl
```

```
## Warning: package 'readxl' was built under R version 4.1.1
```

```
library(ggplot2)
library(RCurl)
```

First Exercise

Load Data

```
data <- read_excel("C:/Users/klein/Desktop/mlr01.xls")
```

3. Inspect Data

```
str(data)
```

```
## tibble [8 x 4] (S3: tbl_df/tbl/data.frame)
##  $ X1: num [1:8] 2.9 2.4 2 2.3 3.2 ...
##  $ X2: num [1:8] 9.2 8.7 7.2 8.5 9.6 ...
##  $ X3: num [1:8] 13.2 11.5 10.8 12.3 12.6 ...
##  $ X4: num [1:8] 2 3 4 2 3 5 1 3
```

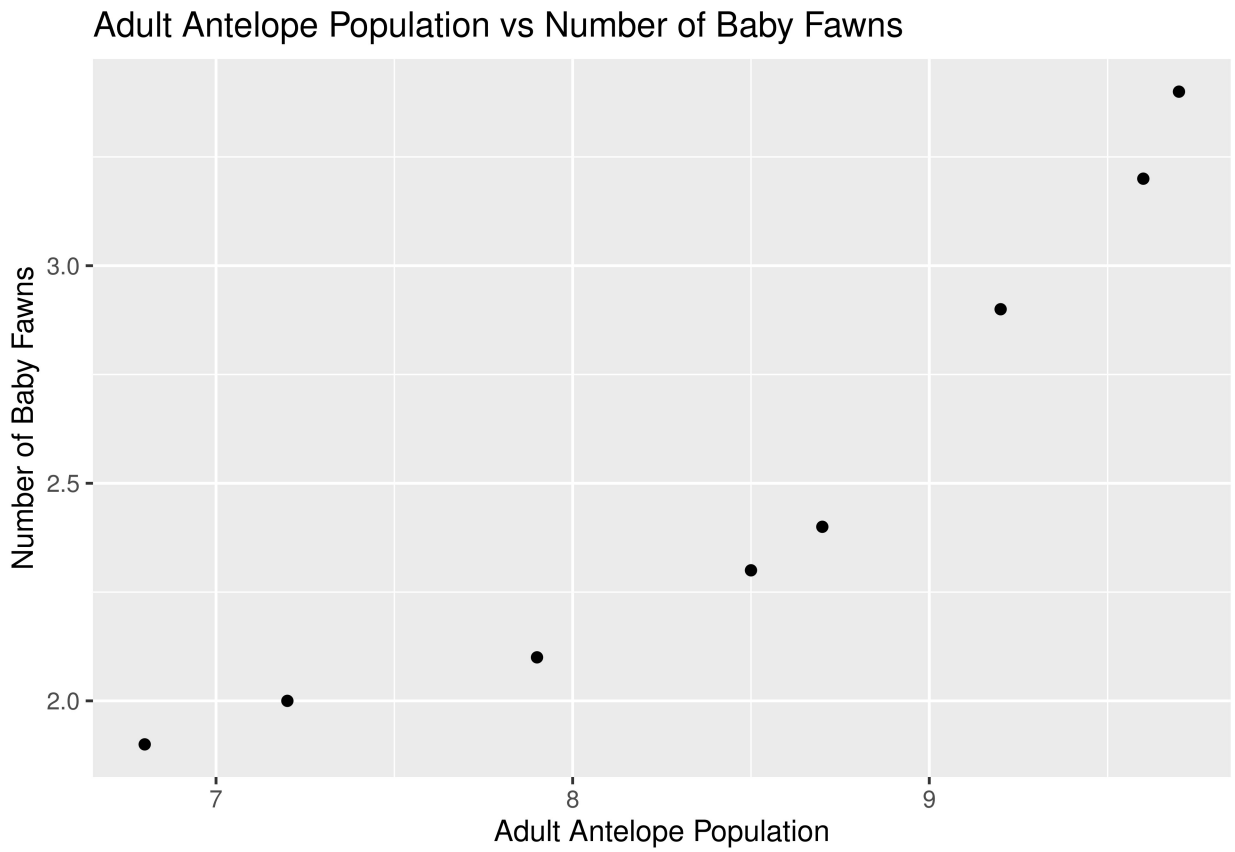
4. Bivariate

Bivariate Plot of fawns vs. adult antelope population

```
colnames(data) <- c("numfawnspring", "popadultante", "annpercip", "winter")
data$year <- seq(1,8, by =1)
```

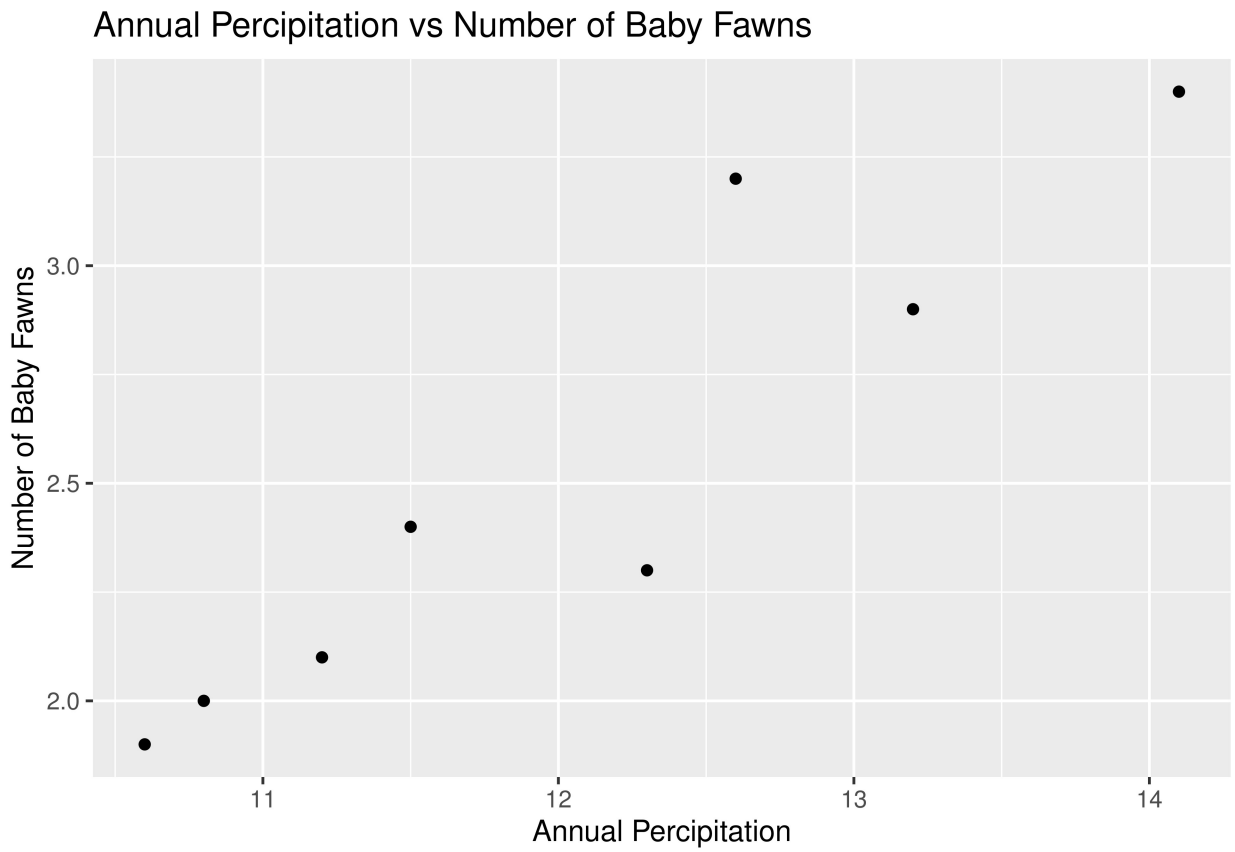
Plot 1: Fawns vs. Adult Antelope

```
ggplot(data, aes(popadultante, numfawnspring)) + geom_point() +
  labs(x= 'Adult Antelope Population', y= 'Number of Baby Fawns') +
  ggtitle('Adult Antelope Population vs Number of Baby Fawns')
```



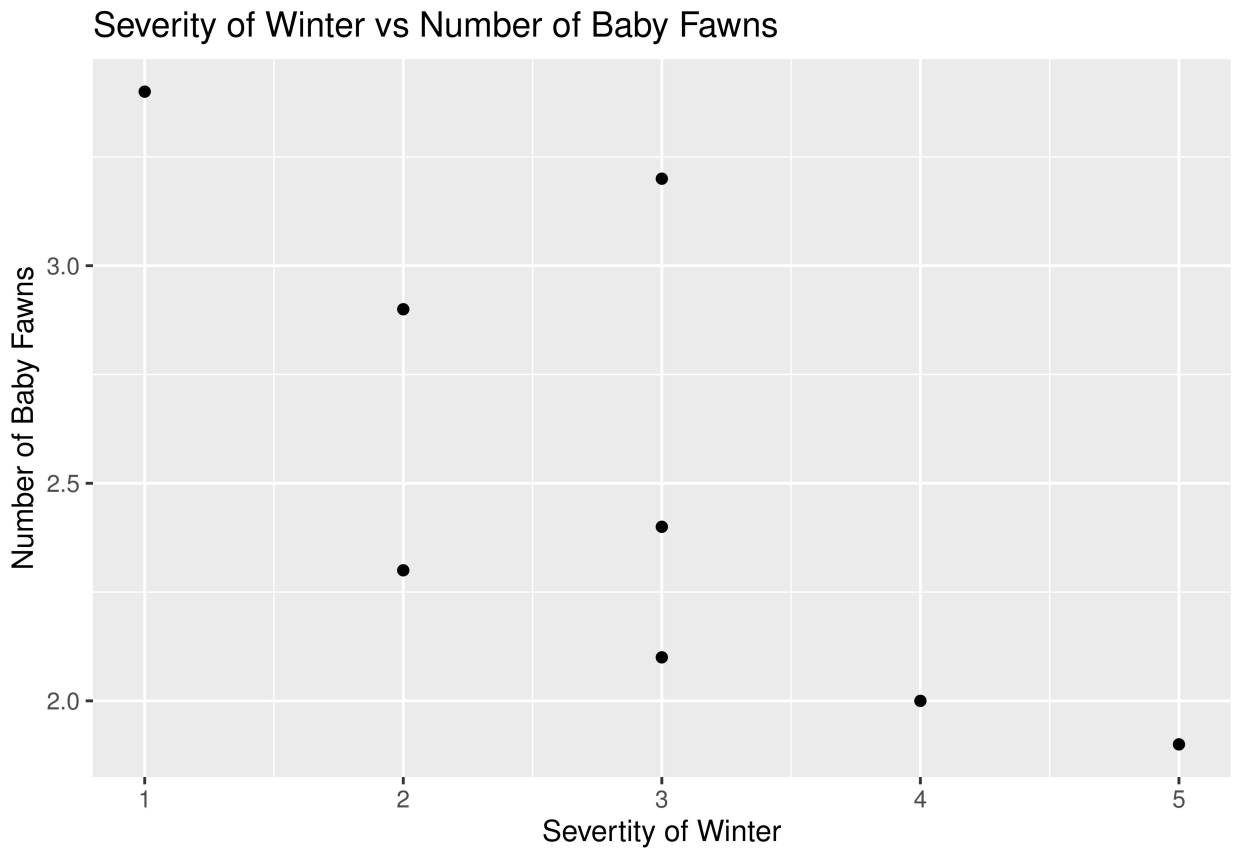
Plot 2: Fawns vs. Percipitation

```
ggplot(data, aes(annpercip, numfawnspring)) + geom_point() +
  labs(x= 'Annual Percipitation', y= 'Number of Baby Fawns') +
  ggtitle('Annual Percipitation vs Number of Baby Fawns')
```



Plot 3: Fawns vs. Severity of Winter

```
ggplot(data, aes(winter, numfawnspring)) + geom_point() +
  labs(x= 'Severtity of Winter', y= 'Number of Baby Fawns') +
  ggtitle('Severity of Winter vs Number of Baby Fawns')
```



5. Regression Models

Regression Model: Number of Fawns from teh Severity of the Winter

```
fawnwinter <- lm(numfawnspring ~ winter, data= data)
summary(fawnwinter)
```

```
##
## Call:
## lm(formula = numfawnspring ~ winter, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52069 -0.20431 -0.00172  0.13017  0.71724
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.4966     0.3904   8.957 0.000108 ***
## winter       -0.3379     0.1258  -2.686 0.036263 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.415 on 6 degrees of freedom
## Multiple R-squared:  0.5459, Adjusted R-squared:  0.4702
## F-statistic: 7.213 on 1 and 6 DF,  p-value: 0.03626
```

Regression Model: Number of Fawns from Severtity of Winter and

```
fawnwinteradult <- lm(numfawnspring ~ winter + popadultante, data= data)
summary(fawnwinteradult)
```

```
##
## Call:
## lm(formula = numfawnspring ~ winter + popadultante, data = data)
##
## Residuals:
##        1        2        3        4        5        6        7        8
##  0.01231 -0.27531  0.10301 -0.19154  0.01535  0.15880  0.29992 -0.12256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.46009    1.53443  -1.603   0.1698
## winter        0.07058    0.12461   0.566   0.5956
## popadultante  0.56594    0.14439   3.920   0.0112 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2252 on 5 degrees of freedom
## Multiple R-squared:  0.8885, Adjusted R-squared:  0.8439
## F-statistic: 19.92 on 2 and 5 DF,  p-value: 0.004152
```

Regression Model: Number of Fawns from All Varaibles

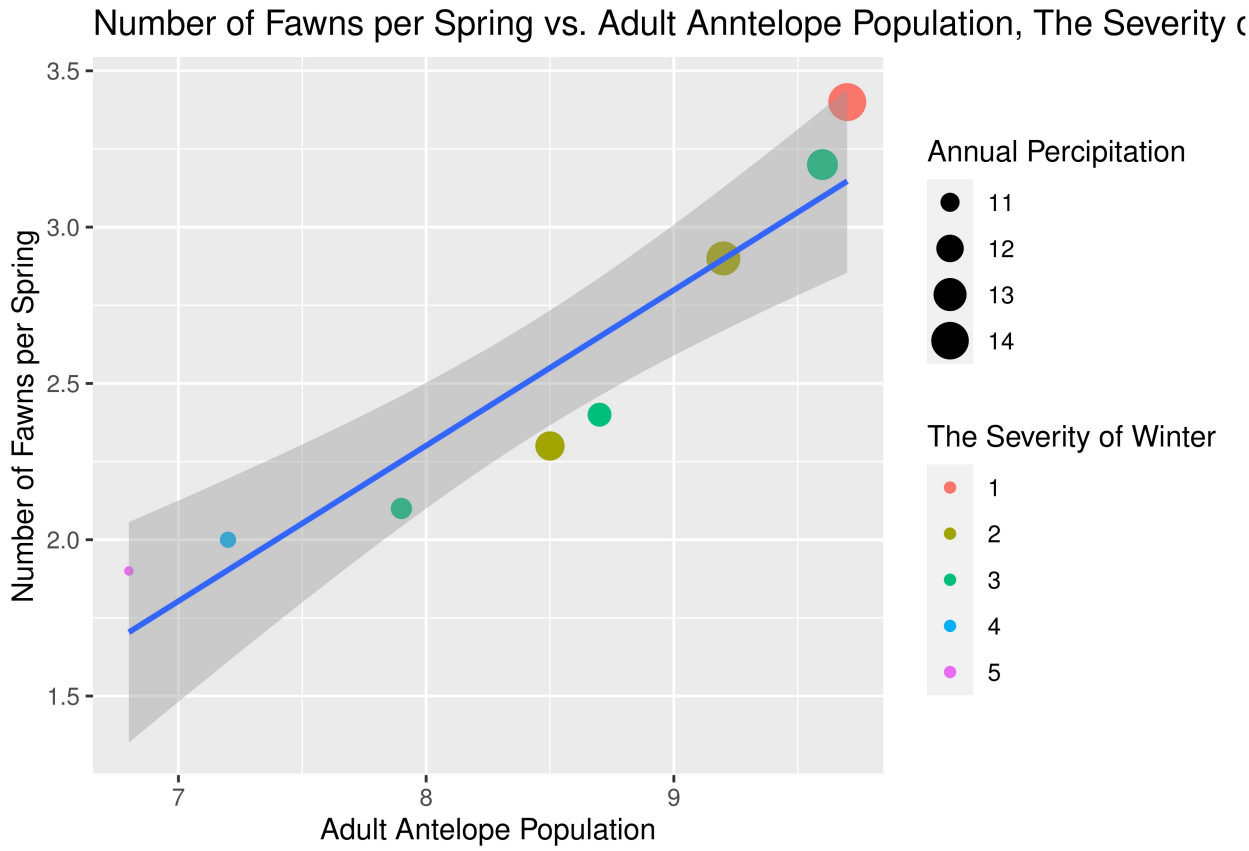
```
allandfawn <- lm(numfawnspring ~ winter + popadultante + annpercip, data= data)
summary(allandfawn)
```

```
##
## Call:
## lm(formula = numfawnspring ~ winter + popadultante + annpercip,
##     data = data)
##
## 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 **
## winter        0.26295    0.08514   3.089   0.0366 *
## popadultante  0.33822    0.09947   3.400   0.0273 *
## annpercip     0.40150    0.10990   3.653   0.0217 *
## ---
## 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
```

Relate Variables

```
ggplot(data, aes(x=popadultante, y=numfawnspring)) +
  geom_point(aes(size= annpercip, color = as.factor(winter))) +
  geom_smooth(method= 'lm', aes(x = popadultante, y= numfawnspring)) +
  labs(x = 'Adult Antelope Population', y='Number of Fawns per Spring',
  color='The Severity of Winter', size='Annual Percipitation') +
  ggtitle('Number of Fawns per Spring vs. Adult Anntelope Population, The Severity of Winter, and Annual
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Which model works the best? Which of the predictors are stastically significant in each model? If you wanted to create the most parsimonious model? What would it contain?

The third model works best.

The third model because it has the R^2 closest to 1, though all are stastically significant.

To create the most parsimonious model, it would need to neither one that overfits or underfits. Using AIC, the lower AIC value is removed and after that step the AIC increases. So, we'd need all the attributes for the parismonious model.