HW8

Diego Valdes

March 9, 2019

```r
rm(list=ls()) # clear work space
#dev.off(dev.list()["RStudioGD"]) # clear plots

suppressWarnings(require(openxlsx))
```

```
## Loading required package: openxlsx
```

```r
suppressWarnings(require(ggplot2))
```

```
## Loading required package: ggplot2
```

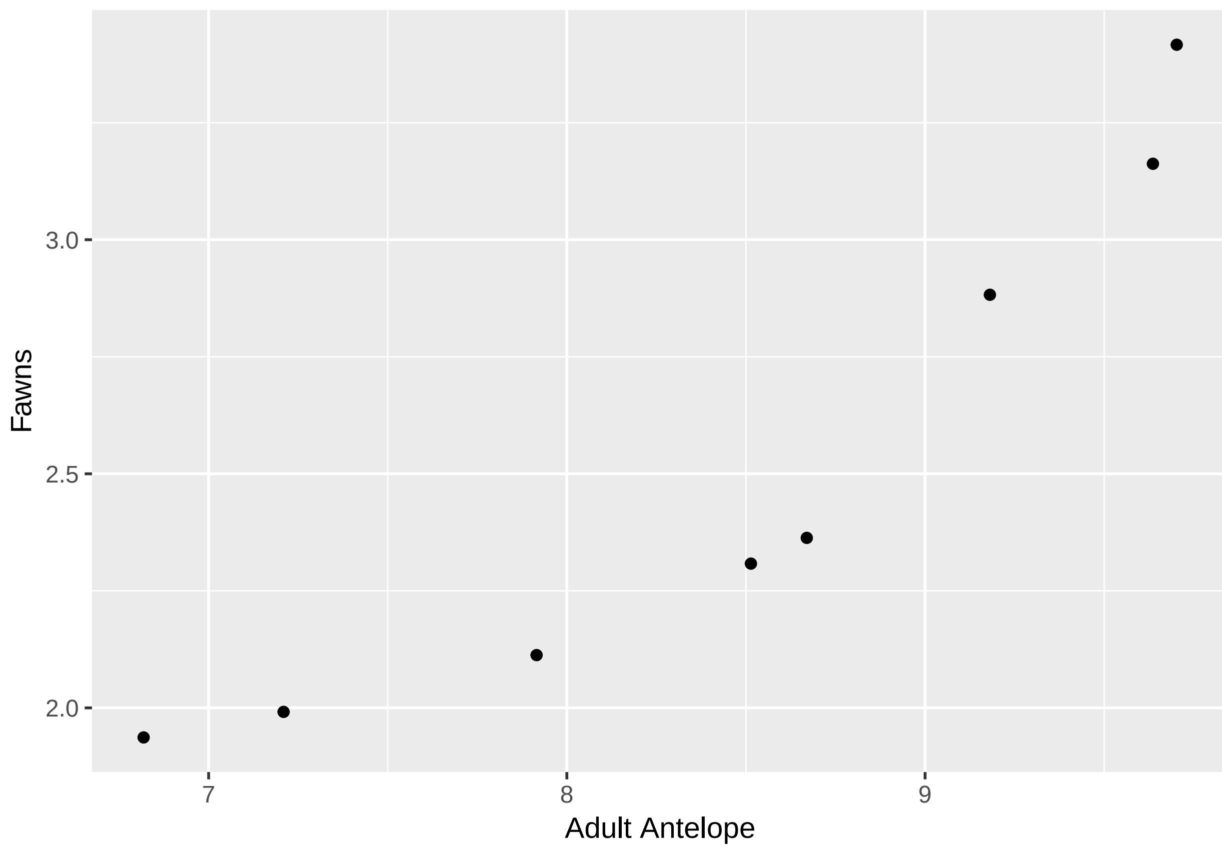
```r
# Downloaded from:
# http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/frame
# File saved as .xlsx

setwd("C:/Users/dvjr2/Google Drive/Documents/Syracuse/IST_687/HW/")
fileName = "mlr01.xlsx"

fawnData = read.xlsx(fileName)
colnames(fawnData) = c("FawnCount", "AntelopeAdultPop", "Percipitation", "WinterSev")
summary(fawnData)
```

```
##    FawnCount     AntelopeAdultPop Percipitation     WinterSev
##  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
```

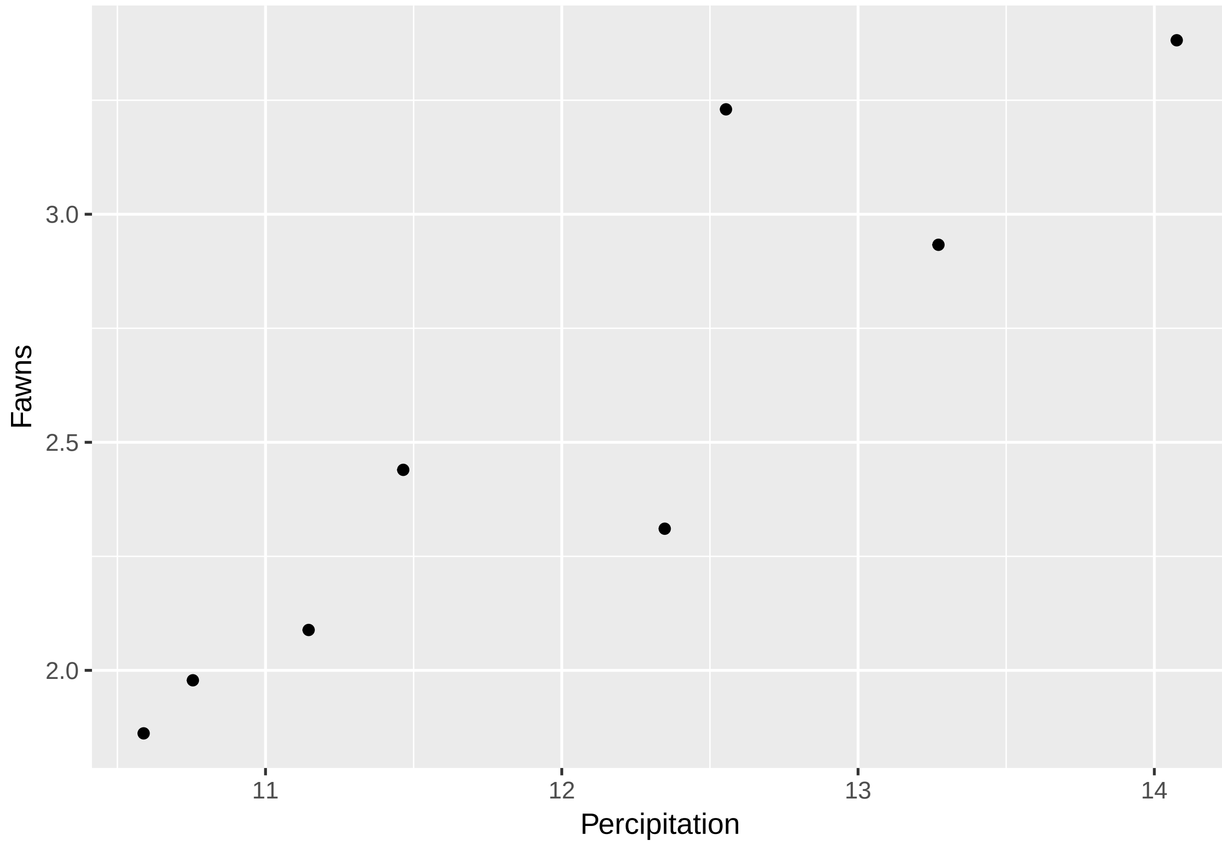
```r
str(fawnData)
```

```
## 'data.frame':    8 obs. of  4 variables:
##  $ FawnCount       : num  2.9 2.4 2 2.3 3.2 ...
##  $ AntelopeAdultPop: num  9.2 8.7 7.2 8.5 9.6 ...
##  $ Percipitation   : num  13.2 11.5 10.8 12.3 12.6 ...
##  $ WinterSev       : num  2 3 4 2 3 5 1 3
```

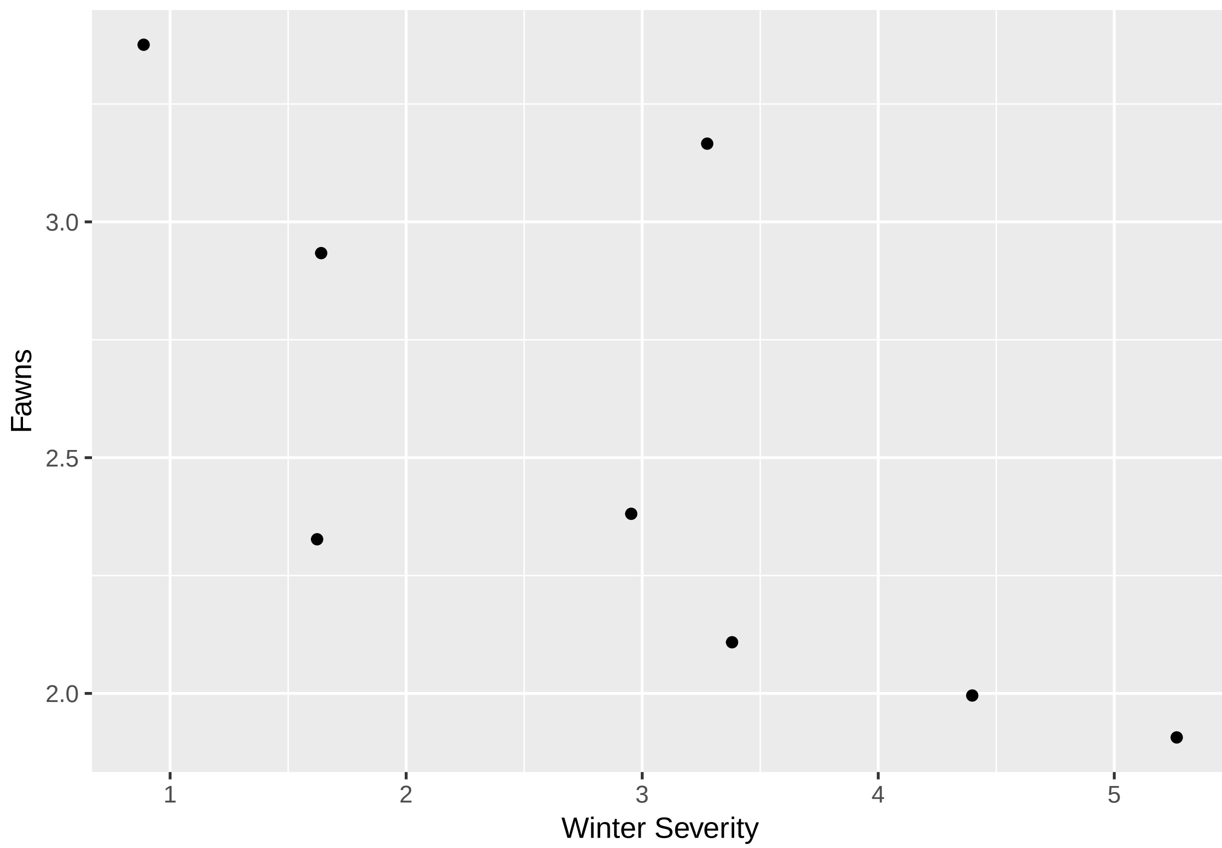
```r
# fawn vs adult antelopes
ggplot(fawnData, aes(AntelopeAdultPop, FawnCount)) + geom_jitter() +
  xlab("Adult Antelope") + ylab("Fawns")
```




```
# fawn vs precipitation
ggplot(fawnData, aes(Percipitation, FawnCount)) + geom_jitter() +
  xlab("Percipitation") + ylab("Fawns")
```




```r
# fawn vs winter
ggplot(fawnData, aes(WinterSev, FawnCount)) + geom_jitter() +
  xlab("Winter Severity") + ylab("Fawns")
```




```
# linear models

# fawns ~ winter
model_001 = lm(FawnCount ~ WinterSev, data = fawnData)
summary(model_001)
```

```
##
## Call:
## lm(formula = FawnCount ~ WinterSev, data = fawnData)
##
## 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 ***
## WinterSev    -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
```

```
#plot(fawnData$WinterSev, fawnData$FawnCount)
#abline(model_001)
```

```r
# fawns ~ winter + percipitation
model_002 = lm(FawnCount ~ WinterSev + Percipitation, data = fawnData)
summary(model_002)
```

```
##
## Call:
## lm(formula = FawnCount ~ WinterSev + Percipitation, data = fawnData)
##
## Residuals:
##         1         2         3         4         5         6         7
## -0.165458  0.188313  0.006417 -0.193358  0.289080 -0.193312 -0.010695
##         8
##  0.079013
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.7791     2.2139  -2.610  0.04765 *
## WinterSev       0.2269     0.1490   1.522  0.18842
## Percipitation   0.6357     0.1511   4.207  0.00843 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2133 on 5 degrees of freedom
## Multiple R-squared:    0.9,  Adjusted R-squared:   0.86
## F-statistic: 22.49 on 2 and 5 DF,  p-value: 0.003164
```

```r
# fawns ~ winter + percipitation + adults
model_003 = lm(FawnCount ~ WinterSev + Percipitation + AntelopeAdultPop, data = fawnData)
summary(model_003)
```

```
##
## Call:
## lm(formula = FawnCount ~ WinterSev + Percipitation + AntelopeAdultPop,
##     data = fawnData)
##
## 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 **
## WinterSev         0.26295    0.08514   3.089   0.0366 *
## Percipitation     0.40150    0.10990   3.653   0.0217 *
## AntelopeAdultPop  0.33822    0.09947   3.400   0.0273 *
## ---
## 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
```

```r
# Which model   works   best?   Model_003 works best Both R^2 values are the highest at .97 and .955
# Which of  the predictors  are statistically   significant in  each    model?
```

```
# Percipitation is the most significant in models 2 and 3
# If    you wanted  to  create  the most    parsimonious    model,  what    would   it  contain?
# I would create a model with Percipitation and Antelope Population and see how well that
#performs on it's own
# b/c in model 3, those were the most significant variables.  Based on that result, I'd make
# a choice between model 3 and the new model.
```