Homework6

Yanyu Zheng yz2690

March 5, 2016

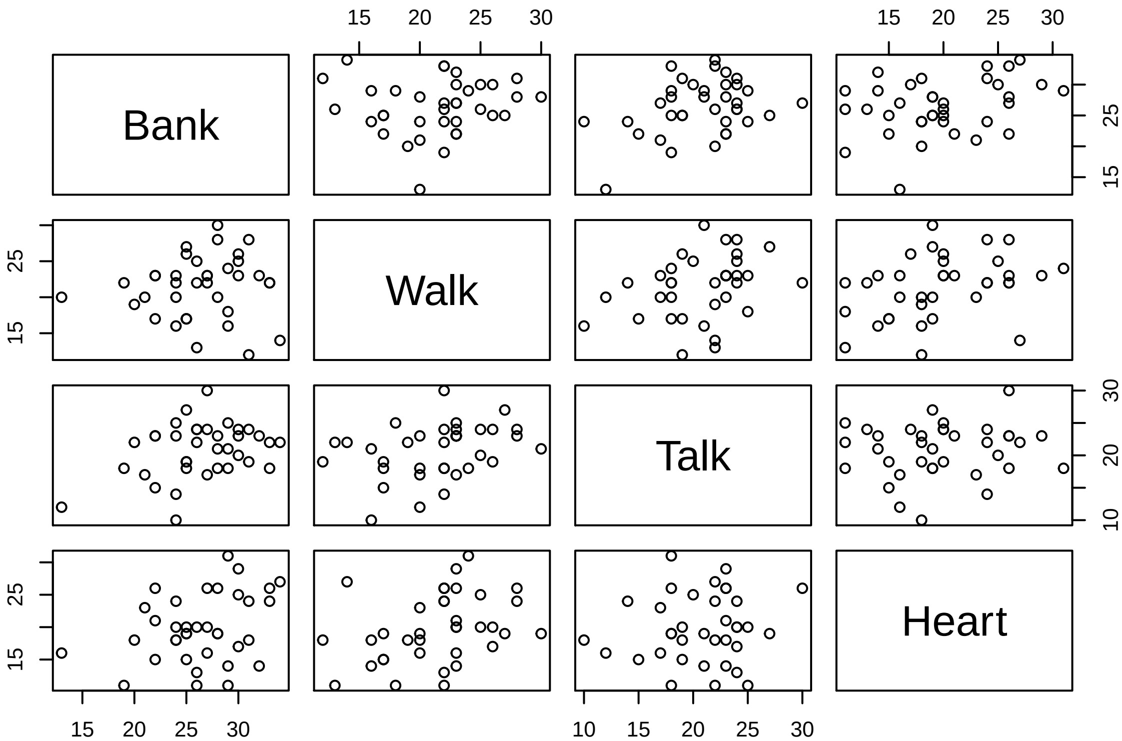
1. Chapter 9, problem 14

```
pack = c("Sleuth3", "dplyr", "ggplot2", "car")
lapply(pack, library, character.only = TRUE)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

a

```
attach(ex0914)
pairs(ex0914)
```



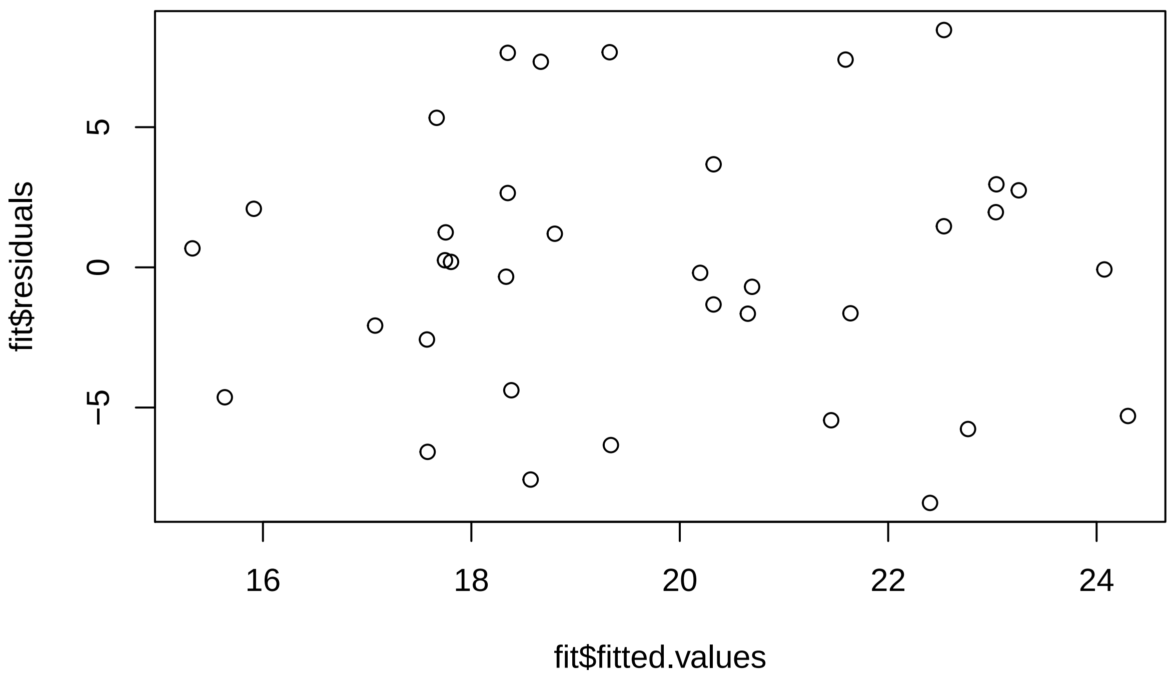
b

```
fit = lm(Heart ~ Bank + Walk + Talk, ex0914)
fit
```

```
##
## Call:
## lm(formula = Heart ~ Bank + Walk + Talk, data = ex0914)
##
## Coefficients:
## (Intercept)         Bank         Walk         Talk
##      3.1787       0.4052       0.4516      -0.1796
```

c

```r
plot(fit$fitted.values,fit$residuals)
```



The variance of residuals seem to be constant throughout all levels of the fitted value. And no evidence for outliers was found.

d

```r
summary(fit)
```

```
##
## Call:
## lm(formula = Heart ~ Bank + Walk + Talk, data = ex0914)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4014 -3.0263  0.0602  2.6748  8.4646
##
```

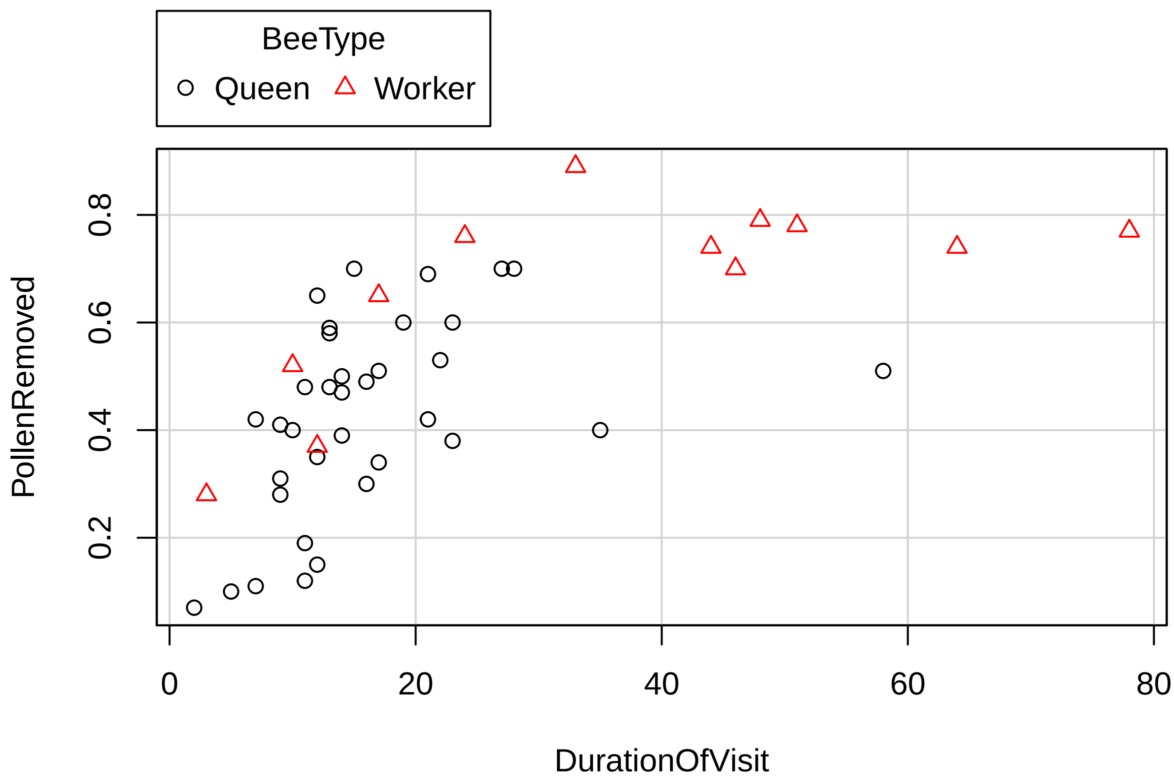
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.1787     6.3369   0.502   0.6194
## Bank          0.4052     0.1971   2.056   0.0480 *
## Walk          0.4516     0.2009   2.248   0.0316 *
## Talk         -0.1796     0.2222  -0.808   0.4249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.805 on 32 degrees of freedom
## Multiple R-squared:  0.2236, Adjusted R-squared:  0.1509
## F-statistic: 3.073 on 3 and 32 DF,  p-value: 0.04162
```

```
detach(ex0914)
```

2. Chapter 9, problem 16

a

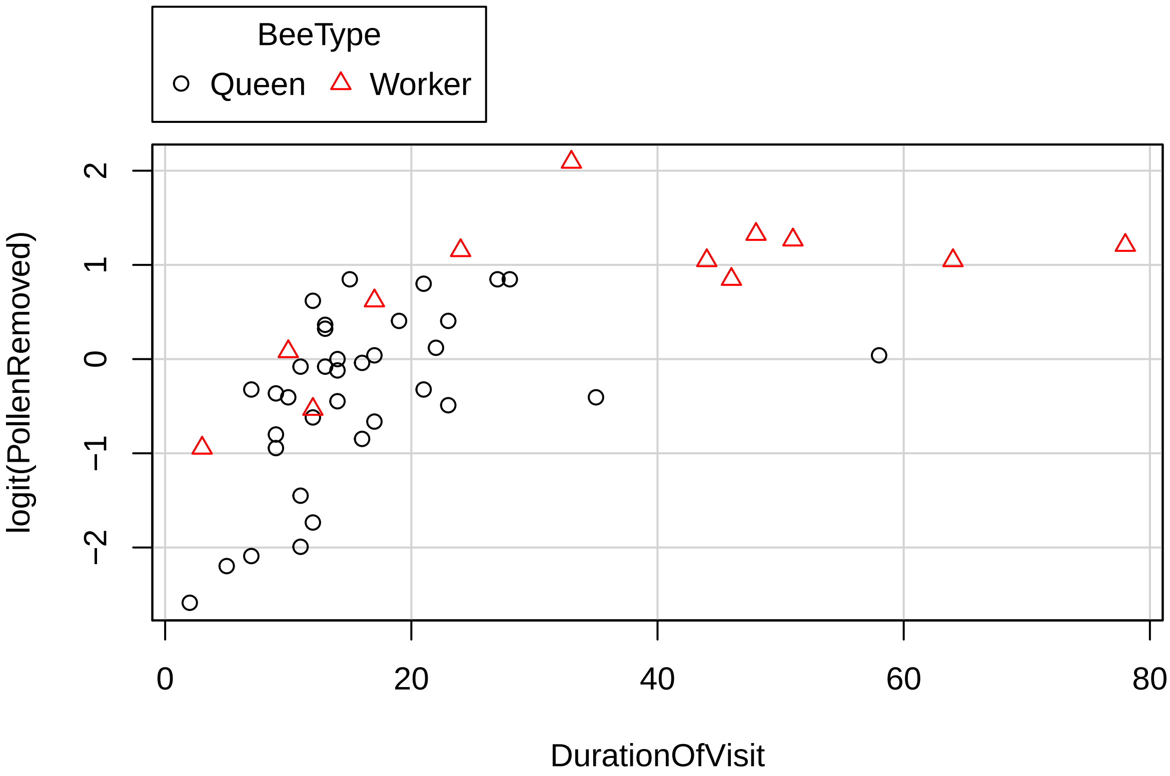
```
attach(ex0327)
scatterplot(PollenRemoved ~ DurationOfVisit | BeeType, ex0327, smoother = FALSE, reg.line = FALSE)
```



No. It does not appear to be a straight line.

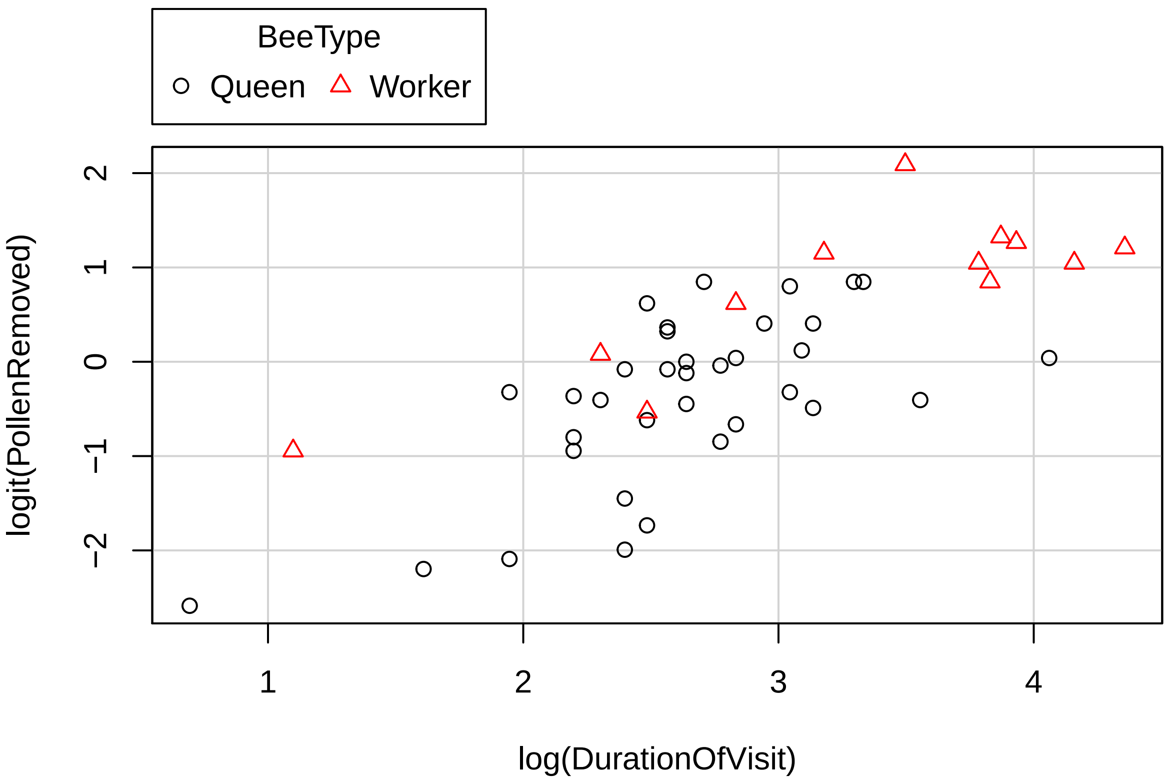
b

```
scatterplot(logit(PollenRemoved) ~ DurationOfVisit | BeeType, ex0327, smoother = FALSE, reg.line = FALSE
```



c

```
scatterplot(logit(PollenRemoved) ~ log(DurationOfVisit) | BeeType, ex0327, smoother = FALSE, reg.line =
```



Logit VS Log seems most resonable to persuit.

d

```
fit2 = lm(logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType + log(DurationOfVisit)*BeeType)
fit2
```

```
##
## Call:
## lm(formula = logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType +
##     log(DurationOfVisit) * BeeType)
##
## Coefficients:
##                  (Intercept)                log(DurationOfVisit)
##                      -3.0390                              1.0121
##                BeeTypeWorker  log(DurationOfVisit):BeeTypeWorker
##                       1.3770                             -0.2709
```

```
summary(fit2)
```

```
##
## Call:
## lm(formula = logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType +
##     log(DurationOfVisit) * BeeType)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3803 -0.3699  0.0307  0.4552  1.1611
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -3.0390     0.5115  -5.941 4.45e-07
## log(DurationOfVisit)             1.0121     0.1902   5.321 3.52e-06
## BeeTypeWorker                    1.3770     0.8722   1.579    0.122
## log(DurationOfVisit):BeeTypeWorker  -0.2709     0.2817  -0.962    0.342
##
## (Intercept)                    ***
## log(DurationOfVisit)           ***
## BeeTypeWorker
## log(DurationOfVisit):BeeTypeWorker
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6525 on 43 degrees of freedom
## Multiple R-squared:  0.6151, Adjusted R-squared:  0.5882
## F-statistic:  22.9 on 3 and 43 DF,  p-value: 5.151e-09
```

The pValue is 0.73, wchich means there's no evidence indicating the proportion of pollen depends on duration of visit differently for queens than for workers.

e

```r
fit2e = lm(logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType)
fit2e
```

```
##
## Call:
## lm(formula = logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType)
##
## Coefficients:
##          (Intercept)  log(DurationOfVisit)         BeeTypeWorker
##              -2.7146                0.8886                0.5697
```

```r
summary(fit2e)
```

```
##
## Call:
## lm(formula = logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40852 -0.49627  0.08815  0.43598  1.15562
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -2.7146     0.3842  -7.065 9.18e-09 ***
```

```
## log(DurationOfVisit)    0.8886     0.1402   6.339 1.07e-07 ***
## BeeTypeWorker           0.5697     0.2364   2.409   0.0202 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.652 on 44 degrees of freedom
## Multiple R-squared:  0.6068, Adjusted R-squared:  0.5889
## F-statistic: 33.95 on 2 and 44 DF,  p-value: 1.206e-09
```

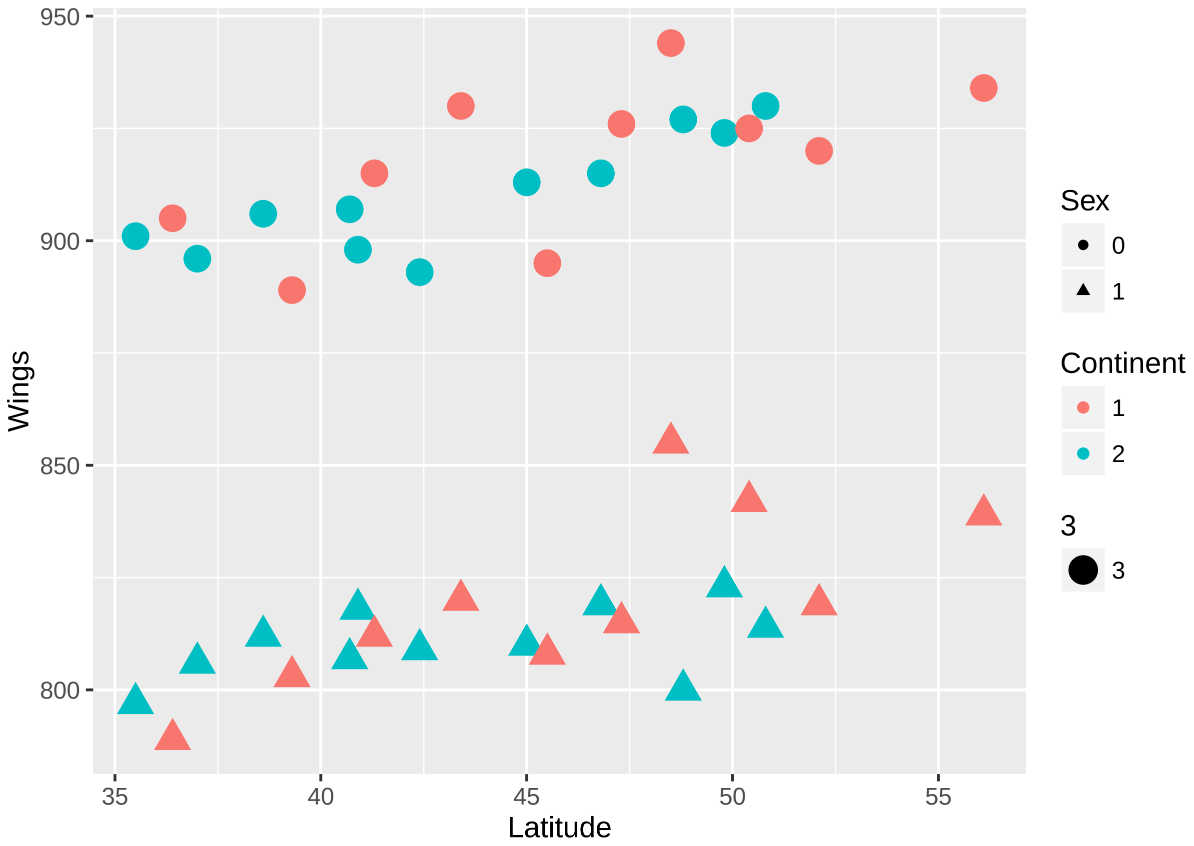
```r
detach(ex0327)
```

Yes, queens tend to remove a smaller proportion. The cross term has high correlation with the BeeType variable, making the model highly unstable. So the diffrence in pVlue is not surprising when removing the cross term.

3. Chapter 9, problem 18

a

```r
attach(ex0918)
library("ggplot2")
Wings = c(Females, Males)
data3 = data.frame(Continent = as.factor(c(Continent, Continent)), Latitude = c(Latitude, Latitude), Wi
ggplot(data3, aes(x = Latitude, y = Wings, colour = Continent, shape = Sex)) + geom_point(aes(size = 3)
```



##

Yes. There's no significant difference between the data from NA and EU

b

```
fit3 = lm(Wings ~ Latitude + Continent + Sex + Continent*Sex, data3)
fit3
```

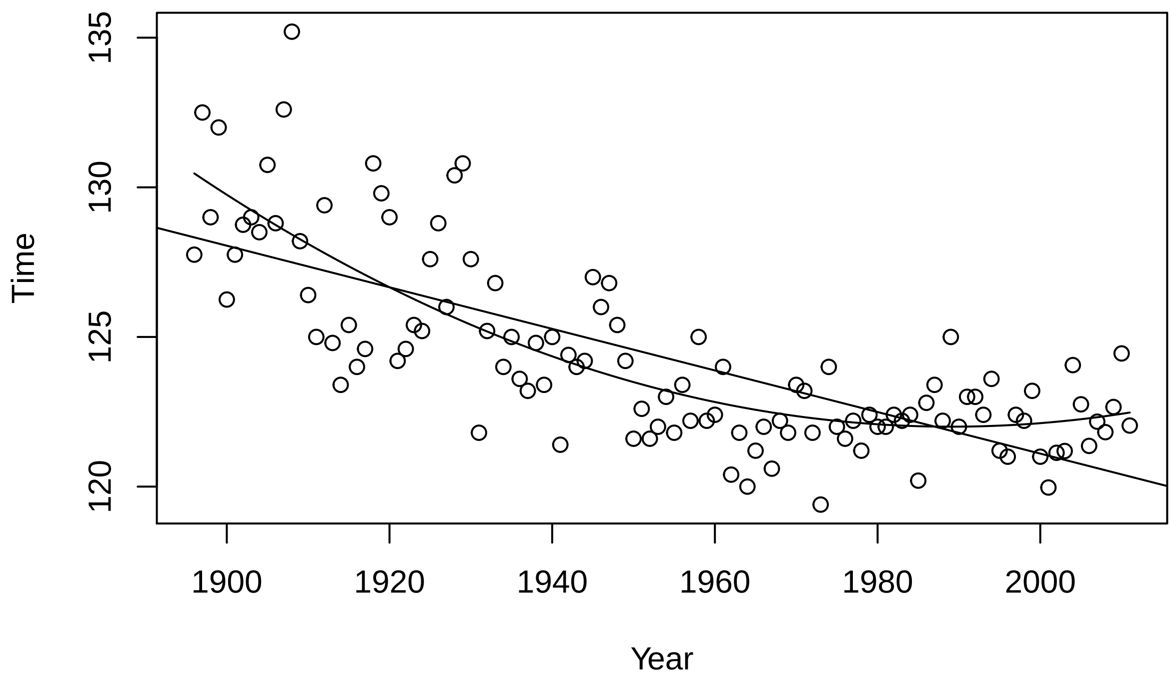
```
##
## Call:
## lm(formula = Wings ~ Latitude + Continent + Sex + Continent *
##     Sex, data = data3)
##
## Coefficients:
##     (Intercept)         Latitude       Continent2             Sex1
##         835.786            1.793           -3.406          -98.100
## Continent2:Sex1
##          -1.445
```

```
detach(ex0918)
```

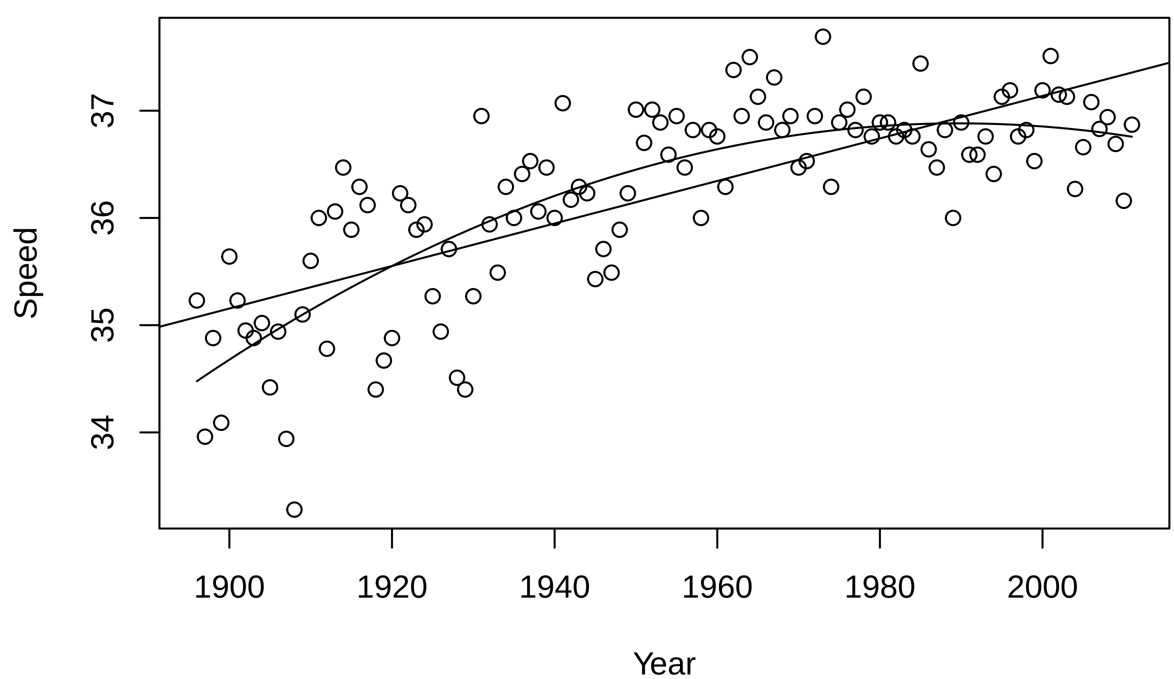
4. Chapter 9, problem 20

a

```
attach(ex0920)
plot(Year, Time)
fit41 = lm(Time~Year)
fit42 = lm(Time~Year + I(Year^2))
abline(fit41)
lines(Year,predict(fit42))
```




```
plot(Year, Speed)
fit43 = lm(Speed~Year)
fit44 = lm(Speed~Year + I(Year^2))
abline(fit43)
lines(Year,predict(fit44))
```



Chooes the fit43, Speed described by quadratic curve of Year.

b

```r
fit4b = lm(Speed~Year + I(Year^2) + Conditions)
fit4b
```

```
##
## Call:
## lm(formula = Speed ~ Year + I(Year^2) + Conditions)
##
## Coefficients:
##    (Intercept)            Year       I(Year^2)  ConditionsSlow
##     -9.791e+02       1.023e+00      -2.575e-04      -9.861e-01
```

Fast tracks exceeds the mean on slow tracks for a speed of 0.9861 miles per hour.

c

```r
fit4c1 = lm(Speed~Year + I(Year^2) + Conditions + Starters)
fit4c2 = lm(Speed~Year + I(Year^2) + Conditions + Starters + Starters * Conditions)
summary(fit4c1)
```

```
##
## Call:
## lm(formula = Speed ~ Year + I(Year^2) + Conditions + Starters)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14854 -0.25191 -0.00883  0.23850  0.80152
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.042e+03  1.408e+02  -7.396 2.84e-11 ***
## Year            1.085e+00  1.440e-01   7.536 1.40e-11 ***
## I(Year^2)      -2.730e-04  3.681e-05  -7.416 2.58e-11 ***
## ConditionsSlow -9.672e-01  9.774e-02  -9.896  < 2e-16 ***
## Starters       -2.060e-02  9.667e-03  -2.131   0.0353 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3894 on 111 degrees of freedom
## Multiple R-squared:  0.819,  Adjusted R-squared:  0.8124
## F-statistic: 125.5 on 4 and 111 DF,  p-value: < 2.2e-16
```

```r
summary(fit4c2)
```

```
##
## Call:
```

```
## lm(formula = Speed ~ Year + I(Year^2) + Conditions + Starters +
##     Starters * Conditions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08958 -0.24451 -0.02678  0.24784  0.77824
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.028e+03  1.418e+02  -7.247 6.23e-11 ***
## Year                     1.071e+00  1.451e-01   7.383 3.15e-11 ***
## I(Year^2)               -2.693e-04  3.709e-05  -7.261 5.82e-11 ***
## ConditionsSlow          -1.175e+00  2.542e-01  -4.624 1.03e-05 ***
## Starters                -2.496e-02  1.085e-02  -2.300   0.0233 *
## ConditionsSlow:Starters  1.622e-02  1.827e-02   0.888   0.3767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3898 on 110 degrees of freedom
## Multiple R-squared:  0.8203, Adjusted R-squared:  0.8121
## F-statistic: 100.4 on 5 and 110 DF,  p-value: < 2.2e-16
```

```r
detach(ex0920)
```

There's evidence for the effect of Starters, while no evidence for the effect of the cross term of Starters and Conditions.

5. Chapter 10, problem 19

a

```r
data(case0901)
fit5a = lm(Flowers ~ Intensity + Time, case0901)
anova(fit5a)
```

```
## Analysis of Variance Table
##
## Response: Flowers
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Intensity  1 2579.75 2579.75  62.181 1.037e-07 ***
## Time       1  886.95  886.95  21.379 0.0001464 ***
## Residuals 21  871.24   41.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b

```r
fit5b = lm(Flowers ~ as.factor(Intensity) + Time + as.factor(Intensity)*Time, case0901)
anova(fit5b)
```

```
## Analysis of Variance Table
##
## Response: Flowers
##                           Df  Sum Sq Mean Sq F value    Pr(>F)
## as.factor(Intensity)       5 2683.51  536.70  9.8189 0.0006388 ***
## Time                       1  886.95  886.95 16.2266 0.0016745 **
## as.factor(Intensity):Time  5  111.55   22.31  0.4081 0.8341569
## Residuals                 12  655.92   54.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c

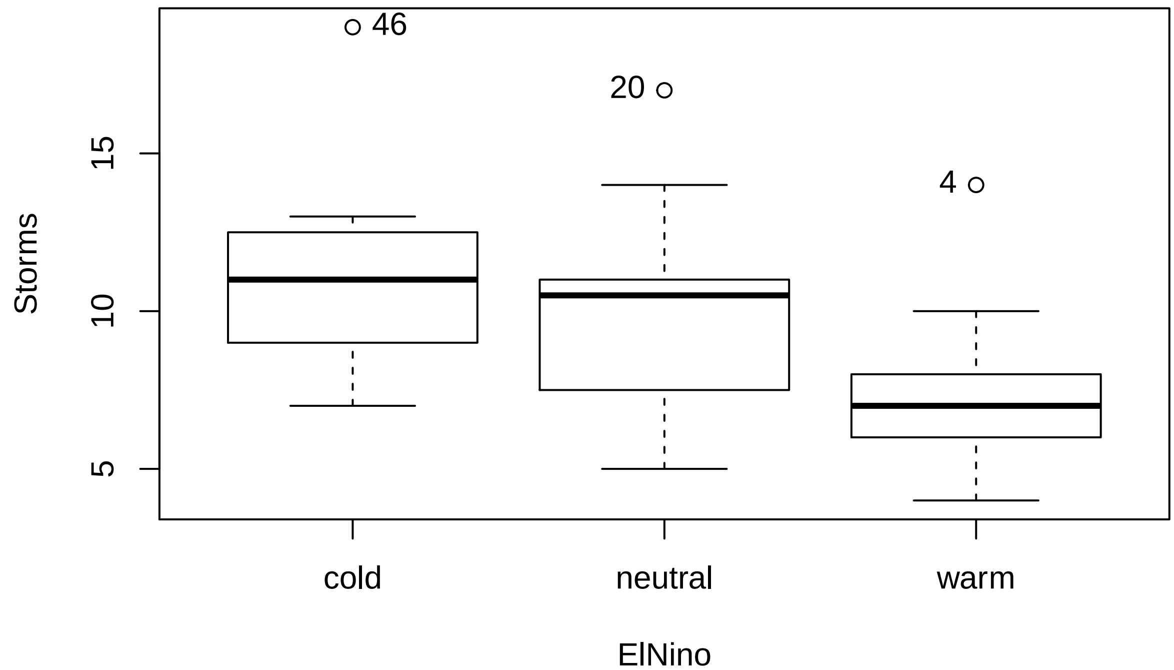
```r
anova(fit5b,fit5a)
```

```
## Analysis of Variance Table
##
## Model 1: Flowers ~ as.factor(Intensity) + Time + as.factor(Intensity) *
##     Time
## Model 2: Flowers ~ Intensity + Time
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     12 655.92
## 2     21 871.24 -9   -215.31 0.4377 0.8894
```

6. Chapter 10, problem 28

a

```r
attach(ex1028)
scatterplot(Storms~ElNino, ex1028, smoother = FALSE)
```

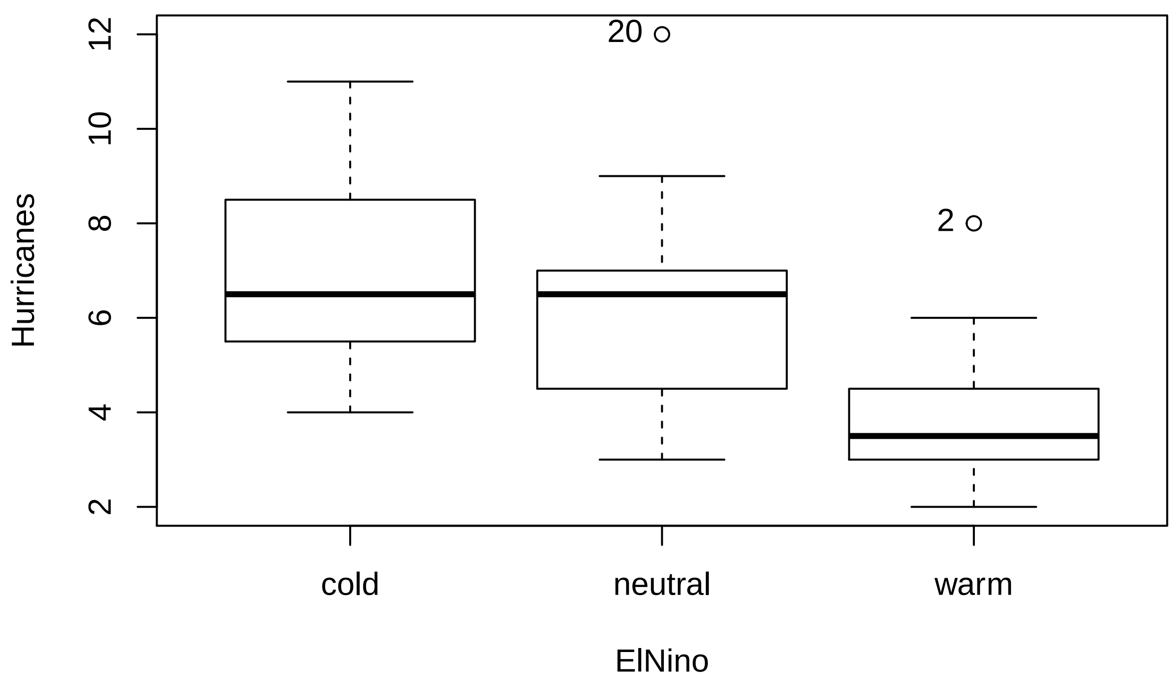



```
## [1] "46" "20" "4"
```

We can clearly see the effect of ElNino on the Storms from the plot. The warmer it is, the less possible there's storms.

b

```r
scatterplot(Hurricanes~ElNino, ex1028, smoother = FALSE)
```



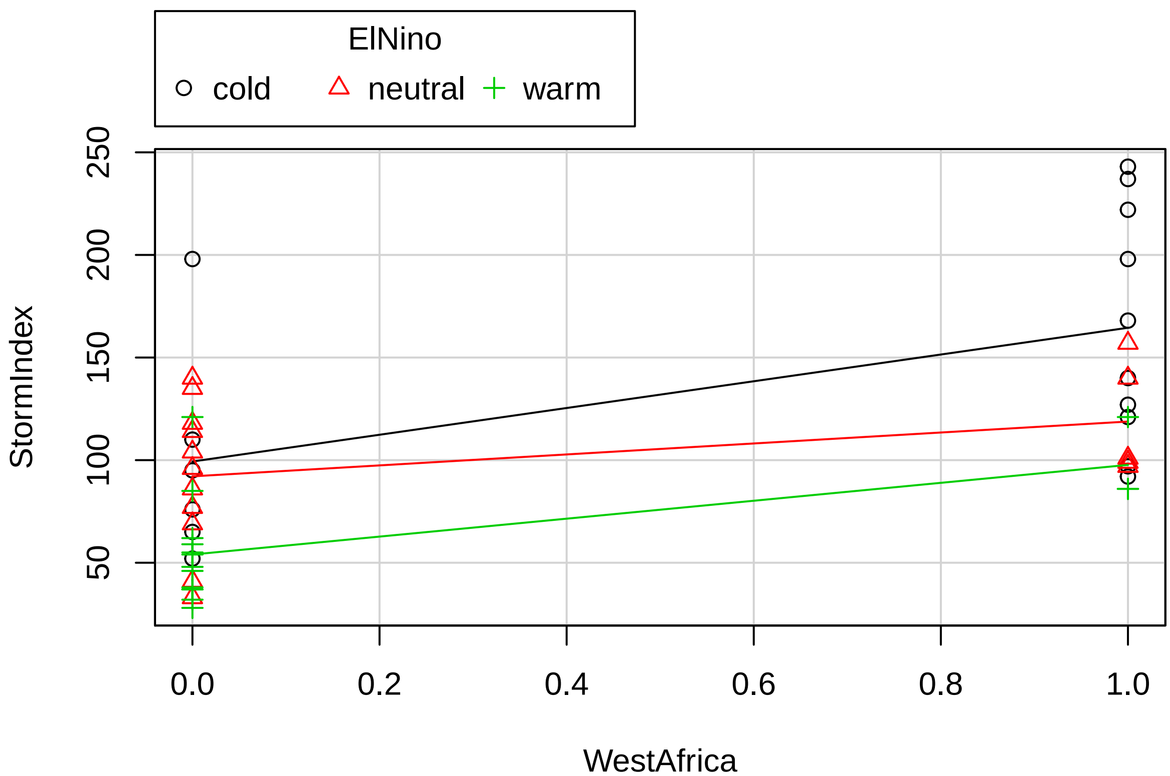
```
## [1] "20" "2"
```

We can clearly see the effect of ElNino on the Hurricanes from the plot. The warmer it is, the less possible there's Huricanes.

```r
anova(lm(StormIndex ~ WestAfrica + ElNino + Year))
```

```
## Analysis of Variance Table
##
## Response: StormIndex
##            Df Sum Sq Mean Sq F value    Pr(>F)
## WestAfrica  1  45554   45554 28.0290 3.842e-06 ***
## ElNino      2  23323   11661  7.1752  0.002047 **
## Year        1     76      76  0.0465  0.830242
## Residuals  43  69885    1625
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## Year is not significant, so I exclude if from the plot
scatterplot(StormIndex ~ WestAfrica + ElNino, ex1028, smoother = FALSE)
```



From the plot and F test, we can see ElNino still have inpact on StormIndex after consideration of the impact of WestAfrica and Time.