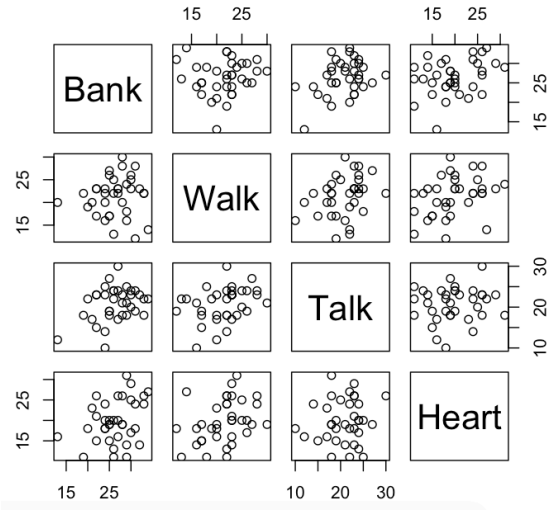
Name: Parth Pareek
UNI: PP2547
Date: 3/9/2016
Assignment: HW6

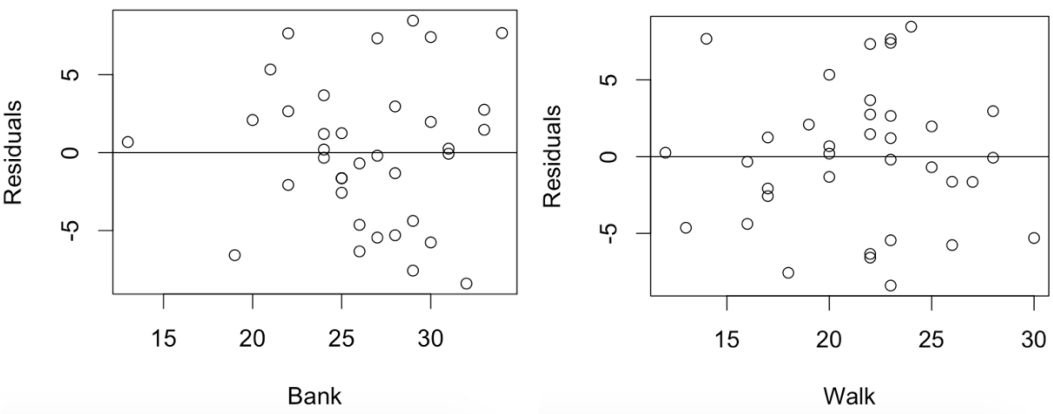
1. a.

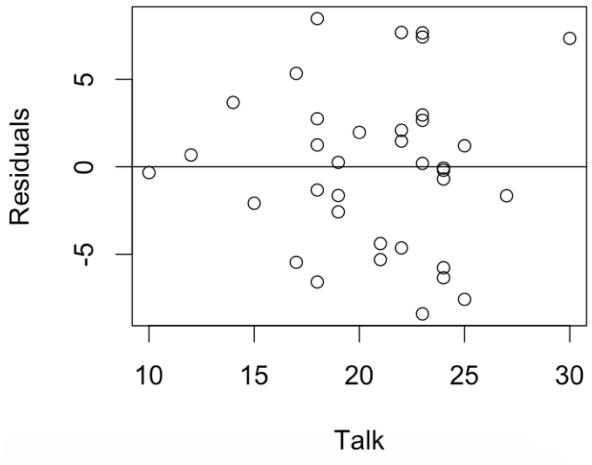


b.

```
Call:
lm(formula = Heart ~ ., data = dat)

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
```

c.



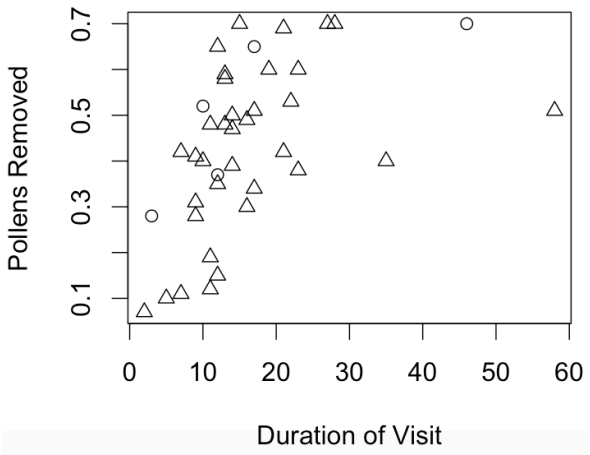


No evidence that variance of residuals increases with increasing fitted values.

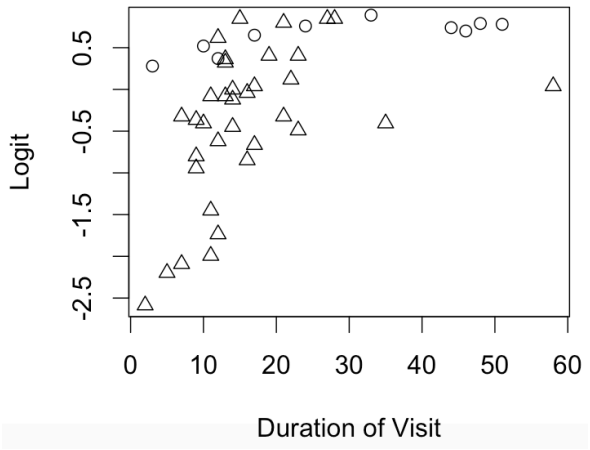
d. Noted in (b)

2.

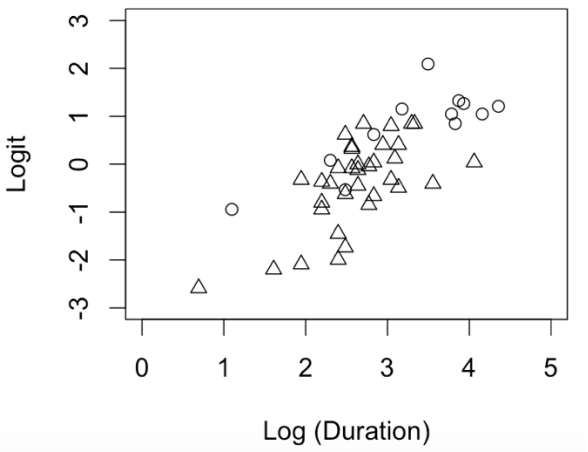
a.



b.



c.



d.

```
Call:
lm(formula = logit ~ log(DurationOfVisit) * BeeType, data = dat.trans)

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 ***
BeeType                        1.3770     0.8722   1.579    0.122
log(DurationOfVisit):BeeType  -0.2709     0.2817  -0.962    0.342
---
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
```

Two sided p-value = 0.342. Therefore, no evidence that time of duration depends on Bee Type

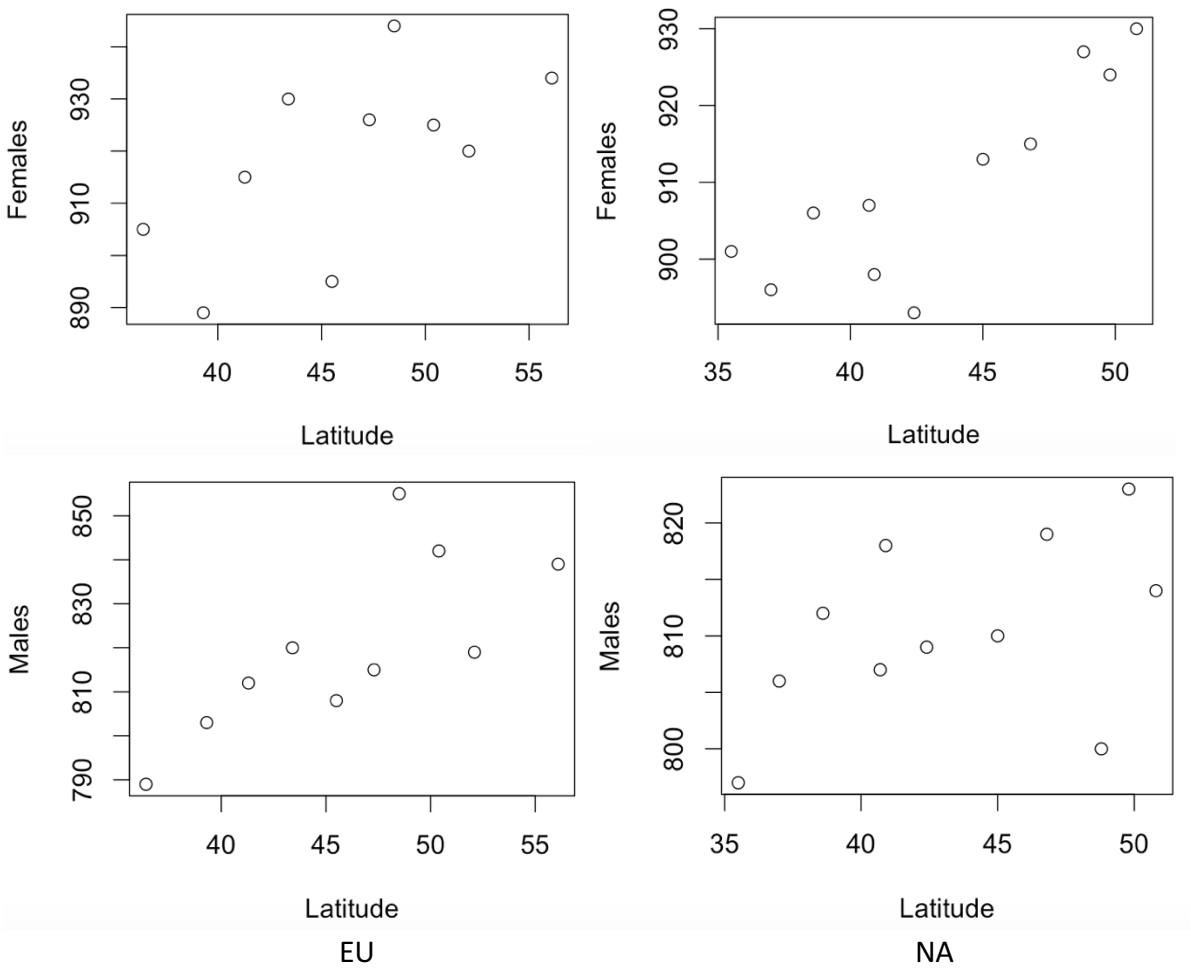
e.

```
Call:
lm(formula = logit ~ log(DurationOfVisit) + BeeType, data = dat.trans)

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 ***
BeeType               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
```

No evidence that queens remove more. Two sided P-value for workers removing more is 0.02 (1 for workers 0 for queens)

Possible collinearity that leads to more p-value. Variable is being accounted for in interaction term also. But cannot say for sure since models are different

3. a.



EU and NA have similar slopes for both males and females

b. 1 for females and 0 for males

```
Call:
lm(formula = WingSize ~ Sex * Latitude, data = newdat)

Residuals:
    Min      1Q  Median      3Q     Max
-22.545  -5.575  -0.844   5.055  32.987

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  735.9839    19.7815  37.206  < 2e-16 ***
Sex1          89.0957    27.9753   3.185 0.002891 **
Latitude       1.7738     0.4401   4.030 0.000258 ***
Sex1:Latitude  0.2189     0.6224   0.352 0.727050
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.23 on 38 degrees of freedom
Multiple R-squared:  0.9572,    Adjusted R-squared:  0.9538
F-statistic: 283.3 on 3 and 38 DF,  p-value: < 2.2e-16
```

4. a.

```
Call:
lm(formula = Speed ~ Year + I(Year^2), data = ex0920)

Residuals:
     Min       1Q   Median       3Q      Max
-1.77397 -0.29516  0.02933  0.31150  1.15734

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.048e+03  1.912e+02  -5.481 2.61e-07 ***
Year         1.090e+00  1.957e-01   5.569 1.75e-07 ***
I(Year^2)   -2.740e-04  5.010e-05  -5.468 2.76e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5411 on 113 degrees of freedom
Multiple R-squared:  0.6442,    Adjusted R-squared:  0.6379
F-statistic: 102.3 on 2 and 113 DF,  p-value: < 2.2e-16
```

b.

```
Call:
lm(formula = Speed ~ Year + I(Year^2) + Conditions, data = ex0920)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11599 -0.22085  0.03084  0.24188  0.85527

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -9.791e+02  1.399e+02  -6.999 2.01e-10 ***
Year            1.023e+00  1.432e-01   7.142 9.83e-11 ***
I(Year^2)      -2.575e-04  3.666e-05  -7.025 1.76e-10 ***
ConditionsSlow -9.861e-01  9.886e-02  -9.975  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3955 on 112 degrees of freedom
Multiple R-squared:  0.8116,    Adjusted R-squared:  0.8065
F-statistic: 160.8 on 3 and 112 DF,  p-value: < 2.2e-16
```

c.

```
Call:
lm(formula = Speed ~ Year + I(Year^2) + Conditions + I(Starters *
    Conditions), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05633 -0.24425 -0.02043  0.26174  0.77515

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -1.013e+03  1.382e+02  -7.332 3.94e-11 ***
Year                      1.055e+00  1.414e-01   7.461 2.06e-11 ***
I(Year^2)                -2.652e-04  3.617e-05  -7.333 3.92e-11 ***
Conditions                1.279e+00  1.623e-01   7.882 2.40e-12 ***
I(Starters * Conditions) -2.406e-02  1.068e-02  -2.252   0.0263 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3885 on 111 degrees of freedom
Multiple R-squared:  0.8198,    Adjusted R-squared:  0.8133
F-statistic: 126.2 on 4 and 111 DF,  p-value: < 2.2e-16
```

d.

```
Call:
lm(formula = Speed ~ Year + I(Year^2) + Starters * Conditions,
    data = dat)

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)         -1.029e+03  1.418e+02  -7.257 5.93e-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 ***
Starters            -8.741e-03  1.650e-02  -0.530    0.597
Conditions           1.175e+00  2.542e-01   4.624 1.03e-05 ***
Starters:Conditions -1.622e-02  1.827e-02  -0.888    0.377
---
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
```

Effect on horses on speed is irrelevant (Two sided p-value 0.597)

5.

a.

```
Analysis of Variance Table

Response: dat$Flowers
              Df  Sum Sq Mean Sq F value    Pr(>F)
dat$Intensity  1 2579.75 2579.75  62.181 1.037e-07 ***
dat$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.

```
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
```

6.

a.

```
Call:
lm(formula = Storms ~ ElNino)

Residuals:
   Min     1Q Median     3Q    Max
-4.938 -1.938 -0.125  1.062  7.875

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.1250     0.6996  15.901  < 2e-16 ***
ElNinoneutral  -1.1875     0.9894  -1.200 0.236353
ElNinowarm     -4.0000     0.9894  -4.043 0.000204 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.799 on 45 degrees of freedom
Multiple R-squared:  0.277,     Adjusted R-squared:  0.2449
F-statistic: 8.621 on 2 and 45 DF,  p-value: 0.0006766
```

b.

```
Call:
lm(formula = Hurricanes ~ ElNino)

Residuals:
   Min     1Q Median     3Q    Max
-3.250 -1.062 -0.125  1.000  5.750

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.0000     0.5069  13.810  < 2e-16 ***
ElNinoneutral  -0.7500     0.7169  -1.046 0.301043
ElNinowarm     -3.0000     0.7169  -4.185 0.000131 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.028 on 45 degrees of freedom
Multiple R-squared:  0.2966,    Adjusted R-squared:  0.2653
F-statistic: 9.486 on 2 and 45 DF,  p-value: 0.0003651
```

c.

```
Call:
lm(formula = StormIndex ~ WestAfrica * ElNino)

Residuals:
    Min      1Q  Median      3Q     Max
-72.500 -22.273  -6.045  22.265  98.667

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 99.333     16.341   6.079 3.06e-07 ***
WestAfrica                  65.167     20.670   3.153  0.00298 **
ElNinoneutral               -7.242     20.314  -0.357  0.72324
ElNinowarm                 -45.333     19.755  -2.295  0.02681 *
WestAfrica:ElNinoneutral   -38.458     29.888  -1.287  0.20524
WestAfrica:ElNinowarm      -21.500     32.932  -0.653  0.51741
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.03 on 42 degrees of freedom
Multiple R-squared:  0.5153,    Adjusted R-squared:  0.4576
F-statistic: 8.931 on 5 and 42 DF,  p-value: 7.715e-06
```

```
#Question 6
library(Sleuth3)
attach(ex1028)
dat <- ex1028
model1 <- lm(Storms ~ ElNino)
summary(model1)

model2 <- lm(Hurricanes ~ ElNino)
summary(model2)

model3 <- lm(StormIndex ~ WestAfrica*ElNino)
summary(model3)
```

CODES

```
#Question 1
library(lattice)
library(ggplot2)
library(Sleuth3)
attach(ex0914)
dat <- ex0914

#Part a
pairs(dat)

#Part b
model <- lm(Heart~.,data = dat)
summary(model)

#Part c
plot(dat$Bank,model$residuals,xlab = "Bank", ylab = "Residuals")
abline(0,0)

plot(dat$Walk,model$residuals,xlab = "Walk", ylab = "Residuals")
abline(0,0)

plot(dat$Talk,model$residuals,xlab = "Talk", ylab = "Residuals")
abline(0,0)


#Question 2
library(lattice)
library(ggplot2)
library(car)
library(Sleuth3)
attach(ex0327)

#Part a
dat <- ex0327
datQ <- dat[which(dat$BeeType == "Queen"),]
datW <- dat[which(dat$BeeType == "Worker"),]
plot(datQ$DurationOfVisit, datQ$PollenRemoved, pch = 2,
     xlab = "Duration of Visit", ylab = "Pollens Removed",
     xlim = range(0:80), ylim = range(0:1))
points(datW$DurationOfVisit,datW$PollenRemoved)

#Part b
dat.trans <- ex0327
dat.trans$logit <- log(dat$PollenRemoved/(1-dat$PollenRemoved))
#dat.trans$logit <- logit(dat$PollenRemoved)
datQ.trans <- dat.trans[which(dat$BeeType == "Queen"),]
datW.trans <- dat.trans[which(dat$BeeType == "Worker"),]
plot(datQ.trans$DurationOfVisit, datQ.trans$logit, pch = 2,
     xlab = "Duration of Visit", ylab = "Logit",
     xlim = range(0:80), ylim = range(-3:3))
points(datW.trans$DurationOfVisit,datW.trans$logit)
```

```
#Part c
plot(log(datQ.trans$DurationOfVisit), datQ.trans$logit, pch = 2,
     xlab = "Log (Duration)", ylab = "Logit",
     xlim = range(0:5), ylim = range(-3:3))
points(log(datW.trans$DurationOfVisit),datW.trans$logit)

#Part d
dat.trans$BeeType <- ifelse(BeeType == "Worker", 1, 0)
model1 <- lm(logit ~ log(DurationOfVisit)*BeeType, data = dat.trans)
summary(model1)

#Part e
model2 <- lm(logit ~ log(DurationOfVisit) + BeeType, data = dat.trans)
summary(model2)


#Question 3
library(lattice)
library(ggplot2)
library(Sleuth3)
attach(ex0918)
dat <- ex0918
datNA <- ex0918[which(ex0918$Continent == "NA"),]
datEU <- ex0918[which(ex0918$Continent == "EU"),]
plot(datNA$Latitude,datNA$Females, xlab = "Latitude", ylab = "Females")
plot(datEU$Latitude,datEU$Females, xlab = "Latitude", ylab = "Females")
plot(datNA$Latitude,datNA$Males, xlab = "Latitude", ylab = "Males")
plot(datEU$Latitude,datEU$Males, xlab = "Latitude", ylab = "Males")

newdat <- matrix(data = NA, nrow = 42, ncol = 4)
newdat <- rbind(dat[,1:2],dat[,1:2])
newdat$Sex[1:21] <- 1
newdat$Sex[22:42] <- 0
newdat$Sex <- as.factor(newdat$Sex)
newdat$WingSize[1:21] <- dat$Females
newdat$WingSize[22:42] <- dat$Males
model2 <- lm(WingSize ~ Sex*Latitude, data = newdat)
summary(model2)


#Question 4
library(lattice)
library(ggplot2)
library(Sleuth3)
attach(ex0920)
dat <- ex0920
#Part a
model1 <- lm (Speed ~ Year + I(Year^2), data = ex0920)
summary(model1)

model2 <- lm (Speed ~ Year + I(Year^2) + Conditions, data = ex0920)
summary(model2)

dat$Conditions <- ifelse(dat$Conditions == "Fast", 1, 0)
```

```
model3 <- lm (Speed ~ Year + I(Year^2) + Conditions +
I(Starters*Conditions),
              data = dat)
summary(model3)

model4 <- lm (Speed ~ Year + I(Year^2) + Starters*Conditions,
              data = dat)
summary(model4)


#Question 5
library(Sleuth3)
attach(case0901)
dat <- case0901
dat_test1 <- aov(dat$Flowers~dat$Intensity+dat$Time)
dat_anova1 <- anova(dat_test)

model2 <- lm(Flowers ~ as.factor(Intensity)*Time, data = dat)
dat_test2 <- aov(model2)
dat_anova2 <- anova(dat_test2)
```