

Homework6

Yanyu Zheng yz2690

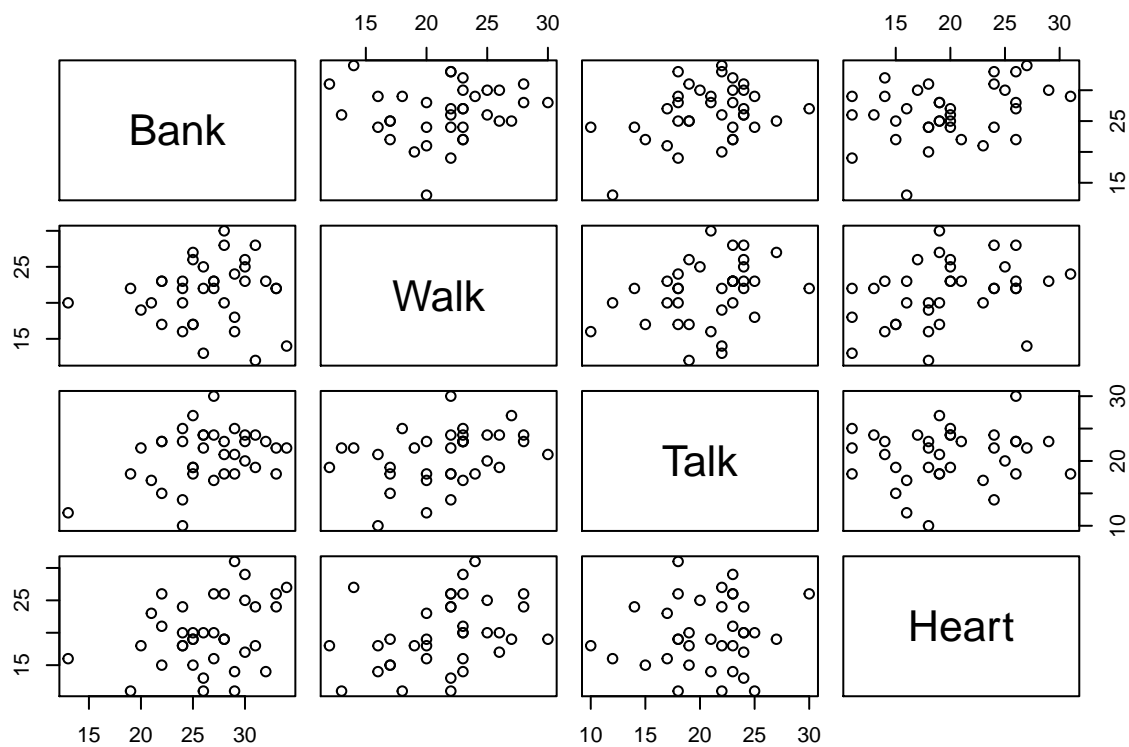
March 5, 2016

1. Chapter 9, problem 14

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

a

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



b

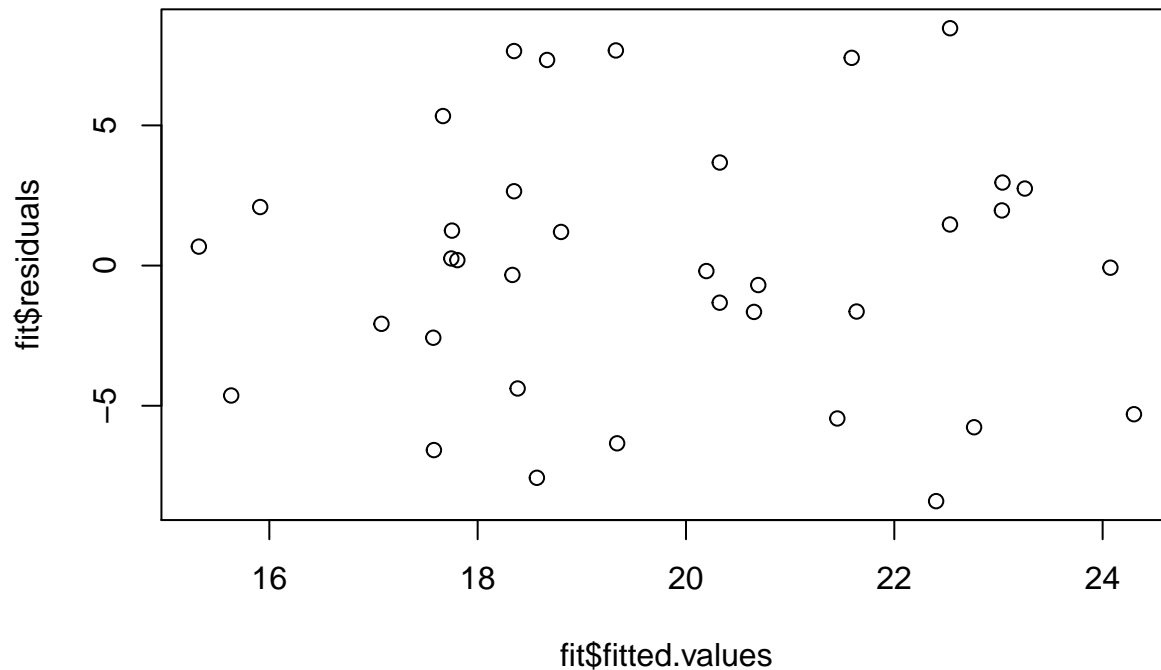
```
fit = lm(Heart ~ Bank + Walk + Talk, ex0914)  
fit
```

```
##
```

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

c

```
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

```
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:
```

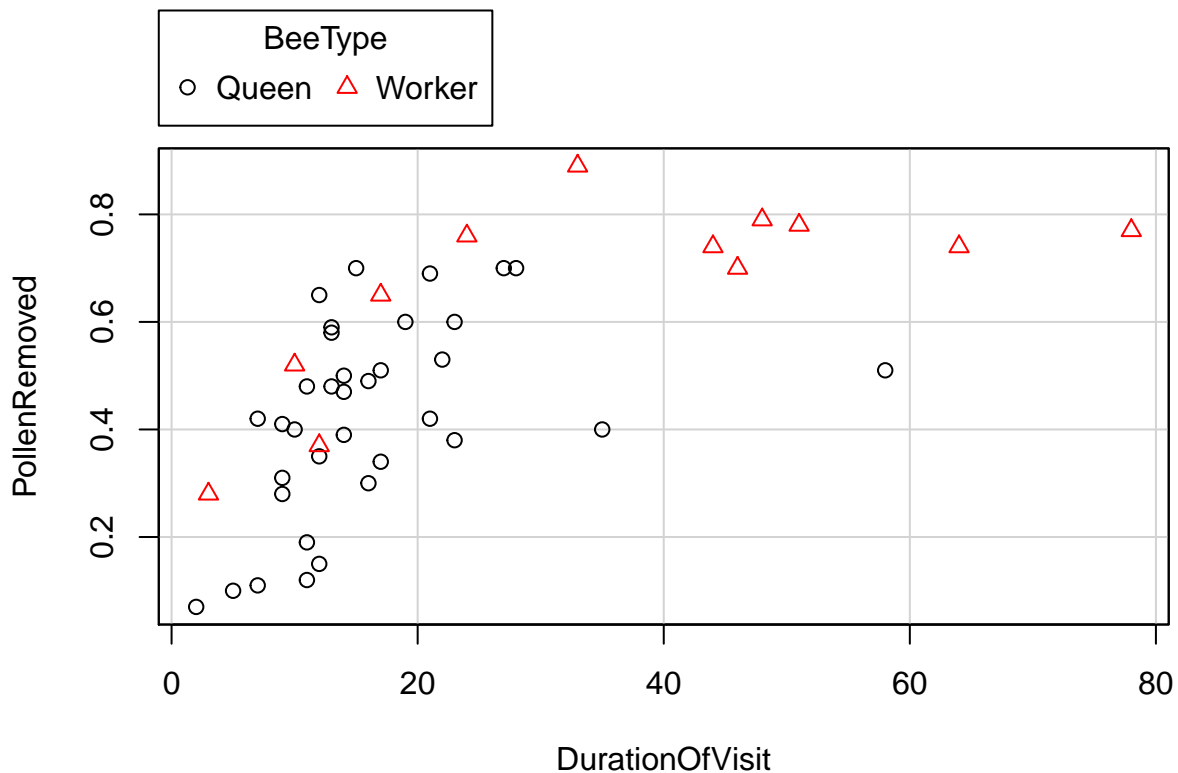
```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.1787     6.3369   0.502  0.6194
## Bank          0.4052     0.1971   2.056  0.0480 *
## Walk          0.4516     0.2009   2.248  0.0316 *
## Talk         -0.1796     0.2222  -0.808  0.4249
## ---
## Signif. codes:  0 '***' 0.001 '**' 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

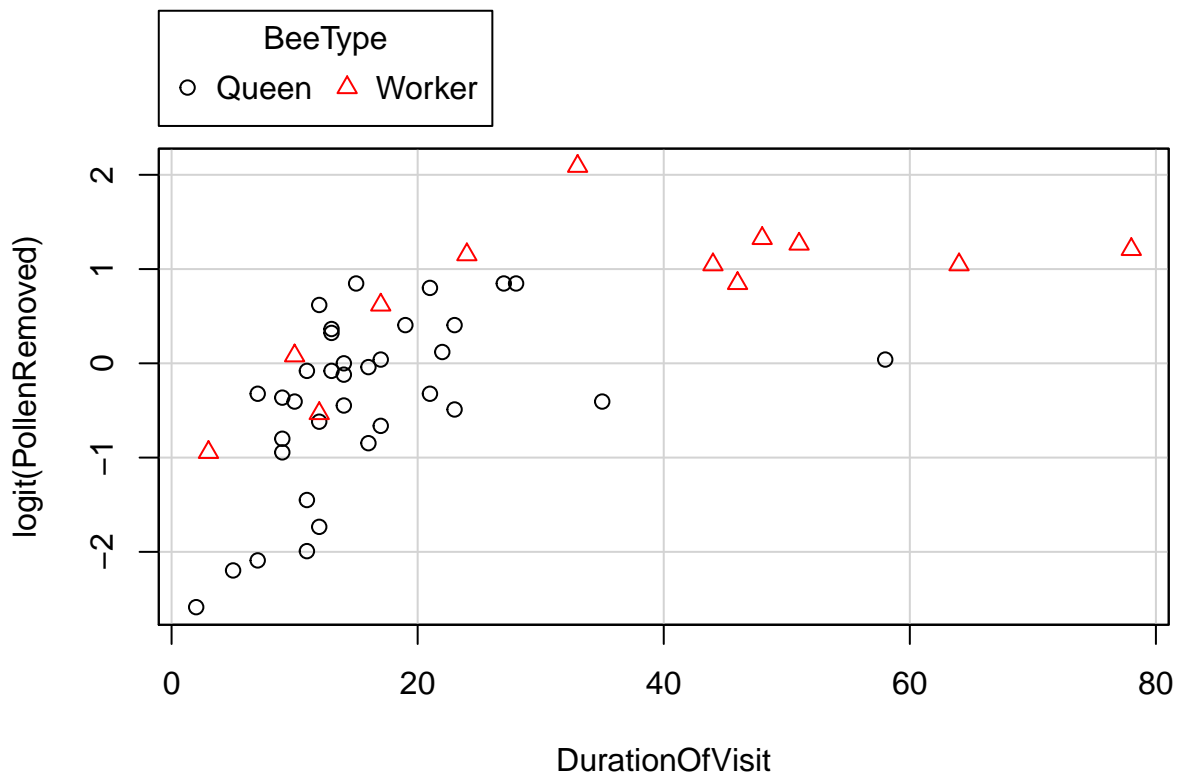
```
attach(ex0327)
scatterplot(PollenRemoved ~ DurationOfVisit | BeeType, ex0327, smoother = FALSE, reg.line = FALSE)
```



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

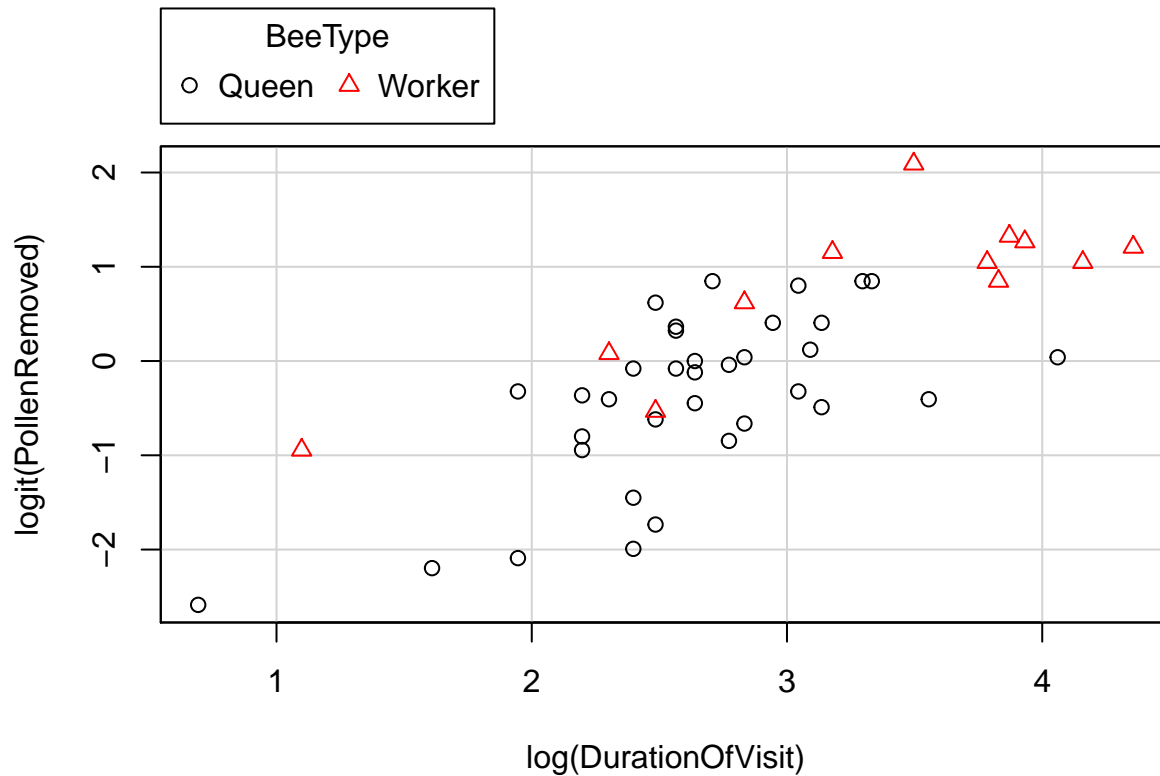
b

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



c

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



Logit VS Log seems most resonable to persuit.

d

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

```
##
## Call:
## lm(formula = logit(PollenRemoved) ~ log(DurationOfVisit) + BeeType +
##     DurationOfVisit * BeeType)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45766 -0.31708  0.04816  0.40354  1.05768
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.823174   0.577762  -6.617 5.14e-08 ***
## log(DurationOfVisit)  1.548448   0.302326   5.122 7.16e-06 ***
## BeeTypeWorker    0.723702   0.398926   1.814  0.0768 .
## DurationOfVisit  -0.038601   0.018918  -2.040  0.0476 *
## BeeTypeWorker:DurationOfVisit  0.004891   0.014069   0.348  0.7299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.6229 on 42 degrees of freedom
## Multiple R-squared: 0.6574, Adjusted R-squared: 0.6247
## F-statistic: 20.14 on 4 and 42 DF, p-value: 2.524e-09
```

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

e

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

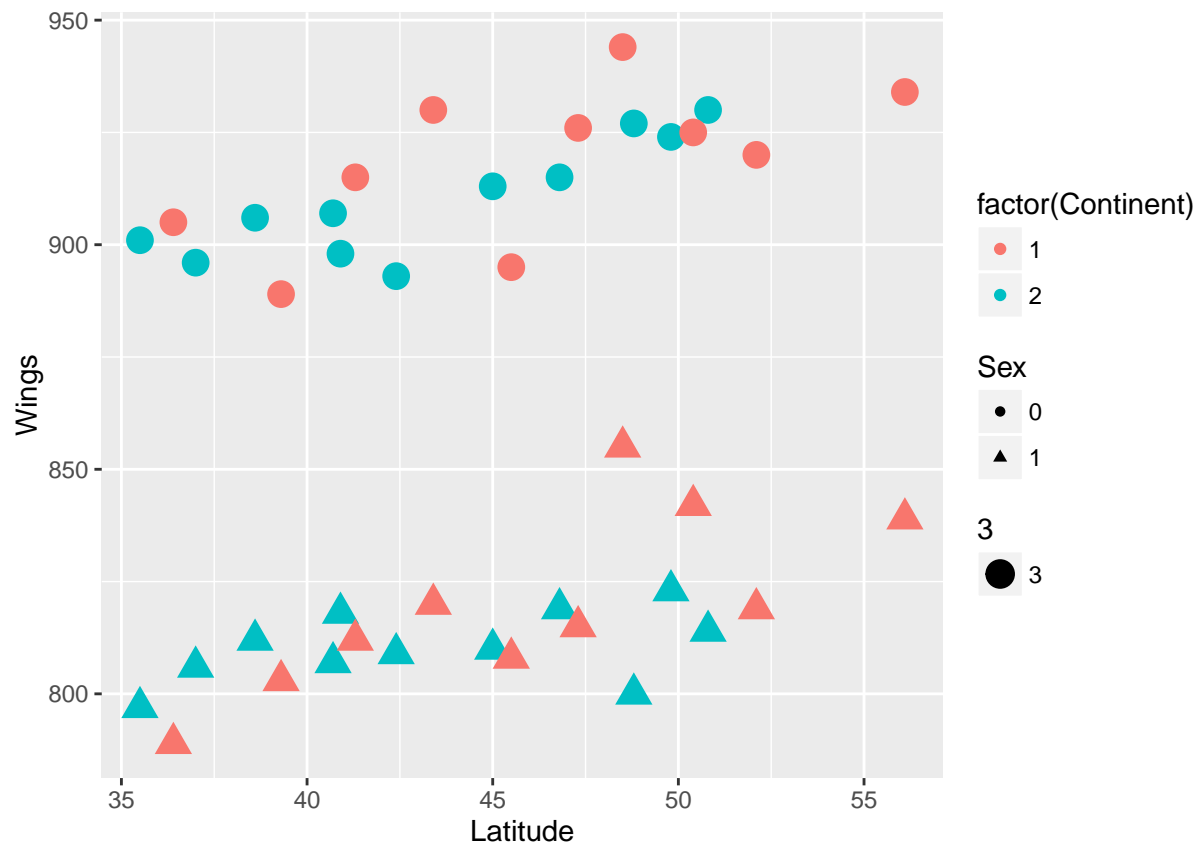
```
detach(ex0327)
```

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 difference in pValue is not surprising when removing the cross term.

3. Chapter 9, problem 18

a

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



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

4. Chapter 9, problem 20

5. Chapter 10, problem 19

6. Chapter 10, problem 28