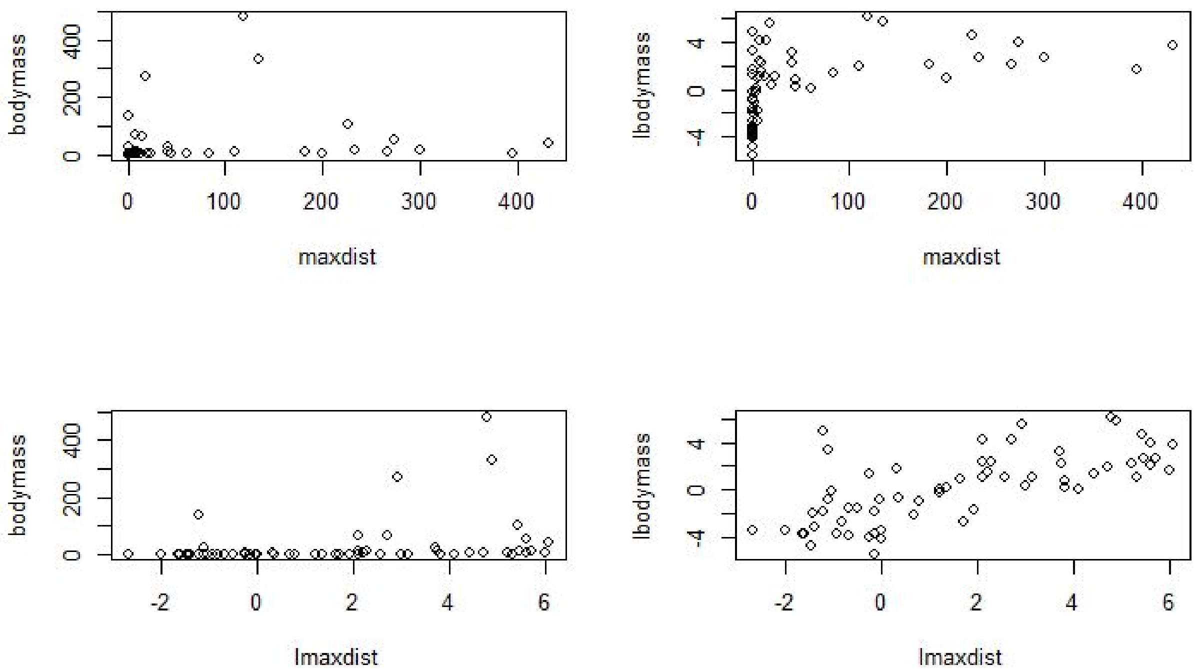
Assignment 4
Part 3

```
dir = "..."
file1 = "ex1124.csv"
dfEx1124 = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')

type = dfEx1124[,"type"]
bodymass = dfEx1124[,"bodymass"]
maxdist = dfEx1124[,"maxdist"]

# Create indicator variables for diet type.
omni = herbi = carni = rep(0,64) # make 3 variables of length 64

# all values equal to 0
omni[type=="O"] = 1 # if type ==O, then omni <- 1; else omni <- 0;
herbi[type=="H"] = 1
carni[type=="C"] = 1
lbodymass = log(bodymass)
lmaxdist = log(maxdist)

# Attach new variables to data frame
dfEx1124 = data.frame(dfEx1124,omni,herbi,carni,lbodymass,lmaxdist)
#attach(dfEx1124,pos=1)
```

3a)

```
par(mfrow=c(2,2))
plot(maxdist,bodymass)
plot(maxdist,lbodymass)
plot(lmaxdist,bodymass)
plot(lmaxdist,lbodymass)
```



The relationship between lbodymass and lmaxdist appears to be linear and certainly the most linear of the four relationships.

3b)
M1 = lm(lmaxdist~lbodymass*type, data=dfEx1124)
summary(M1)

OUTPUT

```
Call:
lm(formula = lmaxdist ~ lbodymass * type, data = dfEx1124)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9221 -0.6731  0.0952  0.7020  4.2592 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      3.01854    0.35649   8.467 1.01e-11 ***
lbodymass        0.70771    0.14168   4.995 5.72e-06 ***
typeH           -2.41569    0.42694  -5.658 4.95e-07 ***
typeO           -1.13373    0.65608  -1.728   0.0893 .  
lbodymass:typeH -0.27648    0.16140  -1.713   0.0921 .  
lbodymass:typeO -0.09824    0.21802  -0.451   0.6540    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.425 on 58 degrees of freedom
Multiple R-squared:  0.7049,	Adjusted R-squared:  0.6794 
F-statistic:  27.7 on 5 and 58 DF,  p-value: 3.232e-14
```

3c)
M2 = lm(lmaxdist~lbodymass+type, data=dfEx1124)
summary(M2)
anova(M2)

OUTPUT

```
Call:
lm(formula = lmaxdist ~ lbodymass + type, data = dfEx1124)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3571 -0.6168 -0.0674  0.9038  4.1355 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.1839     0.3439   9.258 3.64e-13 ***
lbodymass     0.5112     0.0635   8.049 4.04e-11 ***
typeH        -2.5381     0.4216  -6.020 1.14e-07 ***
typeO        -1.1672     0.6133  -1.903   0.0618 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.441 on 60 degrees of freedom
Multiple R-squared:  0.6878,	Adjusted R-squared:  0.6722 
F-statistic: 44.07 on 3 and 60 DF,  p-value: 3.546e-15

> anova(M2)
Analysis of Variance Table

Response: lmaxdist
          Df  Sum Sq Mean Sq F value    Pr(>F)    
lbodymass  1 198.024 198.024  95.323 5.250e-14 ***
type       2  76.622  38.311  18.442 5.715e-07 ***
Residuals 60 124.643   2.077                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

3d)
anova(M2,M1)
Fcritical = qf(0.95, df1=2, df2=58) = 3.1559

Table

```
> anova(M2,M1)
Analysis of Variance Table

Model 1: lmaxdist ~ lbodymass + type
Model 2: lmaxdist ~ lbodymass * type
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     60 124.64
2     58 117.85  2    6.7972 1.6727 0.1967
```

5 Step Hypothesis test:

1. H0: The Parallel Lines model is sufficient to describe the data
   H1: The Three-Lines model is necessary to describe the data

2. Significance level, a = 0.05

3. From Table 1 we can see that our Fobs value for the Three-Lines model is 1.6727

4. We will reject the null hypothesis if Fobs > Fcritical. We can see that 1.6727 < 3.1559 so we do not have evidence to reject the null.

5. There is not enough evidence to reject the null hypothesis and so we can see that the parallel lines model is sufficient to describe the data.

3e)
M3=lm(lmaxdist~lbodymass, data=dfEx1124)
summary(M3)
anova(M3)

```
Call:
lm(formula = lmaxdist ~ lbodymass, data = dfEx1124)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.6635 -1.0135 -0.1455  1.3091  3.4661 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.52351    0.22531   6.762 5.55e-09 ***
lbodymass    0.59854    0.07663   7.810 8.44e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.802 on 62 degrees of freedom
Multiple R-squared:  0.4959,	Adjusted R-squared:  0.4878 
F-statistic:    61 on 1 and 62 DF,  p-value: 8.444e-11

> anova(M3)
Analysis of Variance Table

Response: lmaxdist
          Df Sum Sq Mean Sq F value    Pr(>F)    
lbodymass  1 198.02 198.024  61.002 8.444e-11 ***
Residuals 62 201.26   3.246                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3f)
anova(M3,M2)
Fcritical = qf(0.95, df1=2, df2=60) = 3.1504

Table 2:

```
> anova(M3,M2)
Analysis of Variance Table

Model 1: lmaxdist ~ lbodymass
Model 2: lmaxdist ~ lbodymass + type
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     62 201.26
2     60 124.64  2    76.622 18.442 5.715e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5 Step Hypothesis test:

1. H0: The SLR is sufficient to describe the data
   H1: The Parallel Lines model is necessary

2. Significance level, a = 0.05

3. From Table 2 we can see that our F value for Parallel Lines model is 18.442

4. We will reject the null hypothesis if Fobs > Fcritical. We can see that 18.442 > 3.1504 so we have evidence to reject the null.

5. There is evidence to reject the null hypothesis and so we can see that the SLR Model is not sufficient to describe the data.

3g)
B1 = 0.5112, the coefficient on lbodymass, is the expected change of lmaxdistance if we increase lbodymass by 1 unit while holding the type variable constant.
B2 = -2.5381, the coefficient on typeH, is the expected difference of lmaxdistance between typeH and typeO if we hold lbodymass constant.
B3 = -1.1672, the coefficient on typeO is the expected difference in lmaxdistance between typeO and typeH if we hold lbodymass constant.