1) Run an MLR using the New York Air Quality dataset (*Air_qual.xls*). Your task is to predict ground level ozone pollution in the city (lnO3) in parts per billion (ppb), which is known to vary as a function of solar radiation (Solar.R), wind speed (Wind), and temperature (TempF). Determine the **effect sizes** of each of these three predictor variables and report the results in a couple of sentences. **(3 marks)**

Q1 3

Non standardized model:

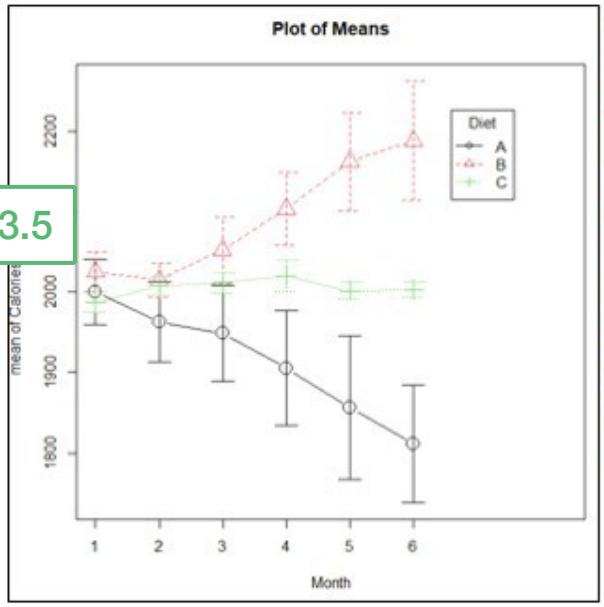
```
Residuals:
     Min       1Q   Median       3Q      Max
-1.02033 -0.31504 -0.00942  0.32296  1.12230

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2614362  0.5204961   0.502    0.617
Solar.R      0.0021904  0.0005156   4.248 4.65e-05 ***
Wind        -0.0692829  0.0145135  -4.774 5.82e-06 ***
TempF        0.0444569  0.0056785   7.829 3.95e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4666 on 106 degrees of freedom
Multiple R-squared:  0.6736,    Adjusted R-squared:  0.6643
F-statistic: 72.91 on 3 and 106 DF,  p-value: < 2.2e-16
```

Standardized model:

```
Call:
lm(formula = lnO3 ~ Z.Solar.R + Z.TempF + Z.Wind, data = air)

Residuals:
     Min       1Q   Median       3Q      Max
-1.02033 -0.31504 -0.00942  0.32296  1.12230

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.44698    0.04449  77.481  < 2e-16 ***
Z.Solar.R    0.19709    0.04640   4.248 4.65e-05 ***
Z.TempF      0.41795    0.05338   7.829 3.95e-12 ***
Z.Wind      -0.24761    0.05187  -4.774 5.82e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4666 on 106 degrees of freedom
Multiple R-squared:  0.6736,    Adjusted R-squared:  0.6643
F-statistic: 72.91 on 3 and 106 DF,  p-value: < 2.2e-16
```

For one unit increase in solar radiation, we predict that there is approximately 0.0219-unit increase in the ground level of ozone pollution in the city. Similarly, for every one unit increase in temperature, we predict that the ground level of ozone pollution of the city will increase by approximately 0.0445 units. In contrast, we predict that there is a 0.0693-unit decrease in ground level ozone pollution for every one unit increase in wind speed. To summarize, temperature has roughly twice the effect size compared to solar radiation in terms of increasing the ground level pollution while wind speed has approximately the same effect size as solar radiation except wind speed has a negative effect.

Q2 3.5



Students receiving diet A gain the highest number of calories compared to students who received diet B and diet C. It is apparent that students of diet A steadily increase their caloric intake over the 6-month period. Meanwhile, students of diet C steadily decrease their monthly calorie intake over the span of 6 months. Students of diet C are also have the least number of daily caloric intake after the 6-month period. On the other hand, students of diet B neither gain nor reduce their calorie means per day. It can be seen that students of diet B's daily caloric intake levels remain relatively equal over the 6-month period.

```
Type II Repeated Measures MANOVA Tests: Pillai test statistic
             Df test stat approx F num Df den Df    Pr(>F)
(Intercept)   1   0.99866   6694.4      1      9 3.082e-14 ***
Diet          2   0.49069      4.3      2      9   0.04802 *
Student       1   0.20715      0.3      5      5   0.91647
Diet:Student  2   1.27762      2.1     10     12   0.10876
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The diet has a p-value of 0.04802, which is marginally smaller than the p-value threshold of 0.05 – this means that the diet has a significant effect on the daily caloric intake levels of the students. Although the diet is statistically significant, we are not particularly interested in the effect of the diet because we know, through literature, that diet would always have an effect on the caloric levels of the subject. What we are more interested in is the interaction term between the student and the diet.

After running the RmANOVA, we find that the interaction term between the type of diet and the student is not significant, with a p-value of 0.10876, which is greater than the p-value threshold of 0.05. This means that differing type of diet does not have any significant impact on the daily caloric intake levels of the students.

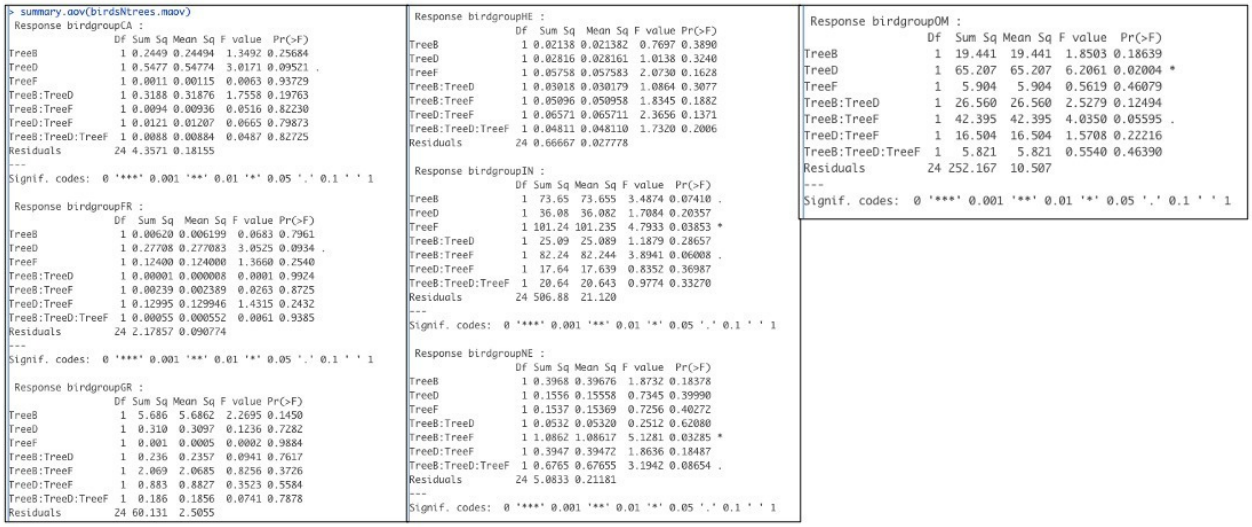
What is a possible reason the interaction term was not significant despite the graph looking significant?

```
>     manova(cbind(birdgroupCA, birdgroupFR, birdgroupGR, birdgroupHE, birdgroupIN, birdgroupNE, birdgroupOM)~ TreeB + TreeD + TreeF + TreeB*TreeD*TreeF,
 data=birdsNtrees)
Call:
   manova(cbind(birdgroupCA, birdgroupFR, birdgroupGR, birdgroupHE,
    birdgroupIN, birdgroupNE, birdgroupOM) ~ TreeB + TreeD +
    TreeF + TreeB * TreeD * TreeF, data = birdsNtrees)

Terms:
                      TreeB    TreeD    TreeF TreeB:TreeD TreeB:TreeF TreeD:TreeF
birdgroupCA          0.2449   0.5477   0.0011     0.3188     0.0094     0.0121
birdgroupFR          0.0062   0.2771   0.1240     0.0000     0.0024     0.1299
birdgroupGR          5.6862   0.3097   0.0005     0.2357     2.0685     0.8827
birdgroupHE          0.0214   0.0282   0.0576     0.0302     0.0510     0.0657
birdgroupIN         73.6550  36.0821 101.2354    25.0893    82.2440    17.6391
birdgroupNE          0.3968   0.1556   0.1537     0.0532     1.0862     0.3947
birdgroupOM         19.4413  65.2071   5.9037    26.5605    42.3950    16.5044
Deg. of Freedom           1        1        1          1          1          1
                    TreeB:TreeD:TreeF Residuals
birdgroupCA                    0.0088    4.3571
birdgroupFR                    0.0006    2.1786
birdgroupGR                    0.1856   60.1310
birdgroupHE                    0.0481    0.6667
birdgroupIN                   20.6431  506.8810
birdgroupNE                    0.6765    5.0833
birdgroupOM                    5.8213  252.1667
Deg. of Freedom                     1        24

Residual standard errors: 0.4260841 0.3012869 1.582863 0.1666667 4.595654 0.4602234 3.241442
Estimated effects may be unbalanced
```

```
> birdsNtrees.maov <- manova(cbind(birdgroupCA, birdgroupFR, birdgroupGR, birdgroupHE, birdgroupIN, birdgroupNE, birdgroupOM)~ TreeB + TreeD + TreeF + Tre
eB*TreeD*TreeF, data=birdsNtrees)
> summary.manova(birdsNtrees.maov)
                Df  Pillai approx F num Df den Df  Pr(>F)
TreeB            1 0.52104  2.79733      7     18 0.03716 *
TreeD            1 0.46963  2.27694      7     18 0.07561 .
TreeF            1 0.40696  1.76456      7     18 0.15673
TreeB:TreeD      1 0.46755  2.25798      7     18 0.07765 .
TreeB:TreeF      1 0.37685  1.55505      7     18 0.21203
TreeD:TreeF      1 0.28144  1.00717      7     18 0.45835
TreeB:TreeD:TreeF  1 0.14253  0.42743      7     18 0.87241
Residuals       24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q3 3

After performing MANOVA, it was found that none of the residuals are significant in the model, so a summary of MANOVA was done. It was found that only the presence of Tree B, with p-value of 0.03716 (lower than alpha=0.05) had a significant impact on the overall response in the different bird guild communities. However, this information is still limited because we are still not necessarily sure whether each bird group significantly respond to Tree B positively. Because of this, we have to perform a univariate results through MANOVA.



```
> summary.aov(birdsNtrees.maov)
 Response birdgroupCA :
              Df Sum Sq Mean Sq F value  Pr(>F)
TreeB          1 0.2449 0.24494  1.3492 0.25684
TreeD          1 0.5477 0.54774  3.0171 0.09521 .
TreeF          1 0.0011 0.00115  0.0063 0.93729
TreeB:TreeD    1 0.3188 0.31876  1.7558 0.19763
TreeB:TreeF    1 0.0094 0.00936  0.0516 0.82230
TreeD:TreeF    1 0.0121 0.01207  0.0665 0.79873
TreeB:TreeD:TreeF  1 0.0088 0.00884  0.0487 0.82725
Residuals     24 4.3571 0.18155
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response birdgroupFR :
              Df  Sum Sq  Mean Sq F value Pr(>F)
TreeB          1 0.00620 0.006199  0.0683 0.7961
TreeD          1 0.27708 0.277083  3.0525 0.0934 .
TreeF          1 0.12400 0.124000  1.3660 0.2540
TreeB:TreeD    1 0.00001 0.000008  0.0001 0.9924
TreeB:TreeF    1 0.00239 0.002389  0.0263 0.8725
TreeD:TreeF    1 0.12995 0.129946  1.4315 0.2432
TreeB:TreeD:TreeF  1 0.00055 0.000552  0.0061 0.9385
Residuals     24 2.17857 0.090774
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response birdgroupGR :
              Df Sum Sq Mean Sq F value Pr(>F)
TreeB          1  5.686  5.6862  2.2695 0.1450
TreeD          1  0.310  0.3097  0.1236 0.7282
TreeF          1  0.001  0.0005  0.0002 0.9884
TreeB:TreeD    1  0.236  0.2357  0.0941 0.7617
TreeB:TreeF    1  2.069  2.0685  0.8256 0.3726
TreeD:TreeF    1  0.883  0.8827  0.3523 0.5584
TreeB:TreeD:TreeF  1  0.186  0.1856  0.0741 0.7878
Residuals     24 60.131  2.5055

 Response birdgroupHE :
              Df  Sum Sq  Mean Sq F value Pr(>F)
TreeB          1 0.02138 0.021382  0.7697 0.3890
TreeD          1 0.02816 0.028161  1.0138 0.3240
TreeF          1 0.05758 0.057583  2.0730 0.1628
TreeB:TreeD    1 0.03018 0.030179  1.0864 0.3077
TreeB:TreeF    1 0.05096 0.050958  1.8345 0.1882
TreeD:TreeF    1 0.06571 0.065711  2.3656 0.1371
TreeB:TreeD:TreeF  1 0.04811 0.048110  1.7320 0.2006
Residuals     24 0.66667 0.027778

 Response birdgroupIN :
              Df Sum Sq Mean Sq F value  Pr(>F)
TreeB          1  73.65  73.655  3.4874 0.07410 .
TreeD          1  36.08  36.082  1.7084 0.20357
TreeF          1 101.24 101.235  4.7933 0.03853 *
TreeB:TreeD    1  25.09  25.089  1.1879 0.28657
TreeB:TreeF    1  82.24  82.244  3.8941 0.06008 .
TreeD:TreeF    1  17.64  17.639  0.8352 0.36987
TreeB:TreeD:TreeF  1  20.64  20.643  0.9774 0.33270
Residuals     24 506.88  21.120
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response birdgroupNE :
              Df Sum Sq Mean Sq F value  Pr(>F)
TreeB          1 0.3968 0.39676  1.8732 0.18378
TreeD          1 0.1556 0.15558  0.7345 0.39990
TreeF          1 0.1537 0.15369  0.7256 0.40272
TreeB:TreeD    1 0.0532 0.05320  0.2512 0.62080
TreeB:TreeF    1 1.0862 1.08617  5.1281 0.03285 *
TreeD:TreeF    1 0.3947 0.39472  1.8636 0.18487
TreeB:TreeD:TreeF  1 0.6765 0.67655  3.1942 0.08654 .
Residuals     24 5.0833 0.21181
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response birdgroupOM :
              Df  Sum Sq Mean Sq F value  Pr(>F)
TreeB          1  19.441  19.441  1.8503 0.18639
TreeD          1  65.207  65.207  6.2061 0.02004 *
TreeF          1   5.904   5.904  0.5619 0.46079
TreeB:TreeD    1  26.560  26.560  2.5279 0.12494
TreeB:TreeF    1  42.395  42.395  4.0350 0.05595 .
TreeD:TreeF    1  16.504  16.504  1.5708 0.22216
TreeB:TreeD:TreeF  1   5.821   5.821  0.5540 0.46390
Residuals     24 252.167  10.507
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After running the univariate MANOVA, we can see that not all species respond to tree species the same way. Bird groups CA, FR, GR and HE do not seem to have a significant response to the tree species, but bird groups IN, NE and OM do. Bird group IN particularly have a significant response (p-value=0.0385 < alpha=0.05) to Tree F, bird group NE have a significant response to the presence of both Tree B and Tree F (interaction term p-value = 0.03285 < alpha=0.05), and bird group OM have a significant response to the presence of Tree D (p-value=0.02004 < alpha=0.05).

Proportion that died $= \frac{74}{145} \times 100\%$

$= 51.0345\%$

Q4 3

51.0345% of song sparrows on Mandarte Island died. ✓

Odds of death $= \frac{p_i}{1-p_i}$

$= \frac{0.510345}{1-0.510345}$

$= 1.042254$ ✓

~ 1

For every one song sparrow that survives, one song sparrow dies.

Logit of odds $= \ln\left(\frac{p_i}{1-p_i}\right)$

$= \ln\left(\frac{0.510345}{1-0.510345}\right)$

$= 0.041395$ ✓

∴ The logit of odds is 0.41395.

```
Call:
glm(formula = survival ~ bdepth + blength + bwidth + mass + tarsus +
    wing, family = binomial(logit), data = songsparrow)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9165  -1.0206  -0.4834   1.0070   1.7138

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0432    12.9519   0.081 0.935804
bdepth       -1.2845     1.0661  -1.205 0.228262
blength       1.3559     0.6724   2.017 0.043738 *
bwidth        0.9117     1.0732   0.849 0.395617
mass          0.1417     0.1525   0.930 0.352608
tarsus       -1.3745     0.3841  -3.578 0.000346 ***
wing          0.1917     0.1686   1.138 0.255304
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 200.95  on 144  degrees of freedom
Residual deviance: 174.99  on 138  degrees of freedom
  (3 observations deleted due to missingness)
AIC: 188.99

Number of Fisher Scoring iterations: 3


Rcmdr>  exp(coef(GLM.1))  # Exponentiated coefficients ("odds ratios")
(Intercept)      bdepth     blength      bwidth        mass      tarsus
  2.8382863   0.2767986   3.8802291   2.4885440   1.1522759   0.2529741
       wing
  1.2113563
```

Q5 3.5

The tarsus length has the least p-value of 0.000346, which is less than the p-threshold of 0.05. This means that the tarsus length has the most significant impact on the survival of song sparrows on Mandarte island. The tarsus length has an odd ratio of approximately 0.25297, which is less than 1. This means that the there is a negative relationship between the length of tarsus and chances of dying in song sparrows.

% Odds of death = (Odd ratio – 1)*100 = (0.25297 – 1)*100 = -74.70%

In conclusion, that for every 1 mm increase of tarsus length, the odds of death is decreased by 74.70%

Odds of survival decreases if tarsus length increases

Remember to specify that the other traits are being controlled for when discussing the odds of survival -0.5

```
Call:
glm(formula = survival ~ blength + tarsus, family = binomial(logit),
    data = songsparrow)

Devi  Residuals:
Q6 3       1Q   Median       3Q      Max
-1.7740  -1.0435  -0.5051   1.0211   1.9192

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  11.0541     8.6275   1.281 0.200105
blength       1.4057     0.6023   2.334 0.019613 *
tarsus       -1.1829     0.3442  -3.437 0.000589 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 200.95  on 144  degrees of freedom
Residual deviance: 179.29  on 142  degrees of freedom
  (3 observations deleted due to missingness)
AIC: 185.29

Number of Fisher Scoring iterations: 4


Rcmdr>  exp(coef(GLM.2))  # Exponentiated coefficients ("odds ratios")
 (Intercept)      blength       tarsus
63200.192779     4.078304     0.306384
```

$$z = a + b_1 X_1 + b_2 X_2$$

$$z = 11.0541 + 1.4057\,(\text{blength}) - 1.1829\,(\text{tarsus})$$

blength = 8 , tarsus = 20

$$z = 11.0541 + 1.4057(8) - 1.1829(20)$$
$$= 11.0541 + 11.2456 - 23.658$$
$$= -1.3583$$

$$p = \frac{e^z}{1+e^z} = \frac{e^{-1.3583}}{1+e^{-1.3583}} = 0.2045167$$

∴ The probability of survival for a bird w/ 8mm of beak length & 20mm of tarsus is 0.20. This is a relatively high chance of survival. ✓