This document contains a summary of several analyses of the meadowfoam flowering data from Chapter 9, and provides examples of R code. The data is displayed on page 237 of your textbook and below. **Flowers** measures the average number of flowers per plant. **Time** is a categorical variable (or factor) **Intensity** is the light intensity (μmol/m²/sec) for whether the light treatment begins early or late.

```r
> case0901

<table>
<thead>
<tr>
<th>Flowers</th>
<th>Time</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.3</td>
<td>late</td>
<td>150</td>
</tr>
<tr>
<td>77.4</td>
<td>late</td>
<td>150</td>
</tr>
<tr>
<td>55.3</td>
<td>late</td>
<td>300</td>
</tr>
<tr>
<td>54.2</td>
<td>late</td>
<td>300</td>
</tr>
<tr>
<td>49.6</td>
<td>late</td>
<td>450</td>
</tr>
<tr>
<td>61.9</td>
<td>late</td>
<td>450</td>
</tr>
<tr>
<td>39.4</td>
<td>late</td>
<td>600</td>
</tr>
<tr>
<td>45.7</td>
<td>late</td>
<td>600</td>
</tr>
<tr>
<td>31.3</td>
<td>late</td>
<td>750</td>
</tr>
<tr>
<td>44.9</td>
<td>late</td>
<td>750</td>
</tr>
<tr>
<td>36.8</td>
<td>late</td>
<td>900</td>
</tr>
<tr>
<td>41.9</td>
<td>late</td>
<td>900</td>
</tr>
<tr>
<td>77.8</td>
<td>early</td>
<td>150</td>
</tr>
<tr>
<td>75.6</td>
<td>early</td>
<td>150</td>
</tr>
<tr>
<td>69.1</td>
<td>early</td>
<td>300</td>
</tr>
<tr>
<td>78.0</td>
<td>early</td>
<td>300</td>
</tr>
<tr>
<td>57.0</td>
<td>early</td>
<td>450</td>
</tr>
<tr>
<td>71.1</td>
<td>early</td>
<td>450</td>
</tr>
<tr>
<td>62.9</td>
<td>early</td>
<td>600</td>
</tr>
<tr>
<td>52.2</td>
<td>early</td>
<td>600</td>
</tr>
<tr>
<td>60.3</td>
<td>early</td>
<td>750</td>
</tr>
<tr>
<td>45.6</td>
<td>early</td>
<td>750</td>
</tr>
<tr>
<td>52.6</td>
<td>early</td>
<td>900</td>
</tr>
<tr>
<td>44.4</td>
<td>early</td>
<td>900</td>
</tr>
</tbody>
</table>
```

The first step is to plot the data using two different symbols for the late and early points. The method here is to make an empty plot first (`type="n"`) and then to add the points (with `points`).

```r
> early <- (Time == "early")
> late <- (Time == "late")
> plot(Intensity, Flowers, type = "n")
> points(Intensity[early], Flowers[early])
> points(Intensity[late], Flowers[late], pch = 16)
```
The first thing to try is to fit parallel lines, in other words, to assume no interaction between Intensity and Time. (We should also look at residual plots for each fit, but I am suppressing this for the moment.)

```r
> fit1 <- lm(Flowers ~ Intensity + Time)
> summary(fit1)

Call:
  lm(formula = Flowers ~ Intensity + Time)

Residuals:
    Min     1Q  Median     3Q    Max
   -9.652  -4.139  -1.558   5.632  12.165

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          83.464167  3.273772  25.495  < 2e-16 ***
Intensity            -0.040471  0.005132  -7.886  1.04e-07 ***
Timelate             -12.158333  2.629557  -4.624  0.000146 ***
---
Signif. codes:  < 0.001 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.441 on 21 degrees of freedom
Multiple R-Squared: 0.7992, Adjusted R-squared: 0.78
F-statistic: 41.78 on 2 and 21 DF, p-value: 4.786e-08

> anova(fit1)

Analysis of Variance Table

Response: Flowers

                     Df Sum Sq Mean Sq  F value    Pr(>F)
---                    ---  ---  ---  ---  ---

Bret Larget

March 12, 2003
A second fit includes an interaction between Intensity and Time. This effectively fits two lines with different intercepts and slopes. In the formula, Intensity * Time does not stand for the product of intensity and time, but rather is expanded to mean Intensity + Time + Intensity:Time, or the two main effects plus an interaction term. The categorical variable Time is treated numerically as a variable that take the value 0 for one level (early, in this case) and the value 1 for the other level late.

```r
> fit2 <- lm(Flowers ~ Intensity * Time)
> summary(fit2)

Call:
  lm(formula = Flowers ~ Intensity * Time)

Residuals:
       Min        1Q   Median        3Q       Max
-9.51600 -4.27500 -1.42167  5.47325 11.93850

Coefficients:              Estimate Std. Error   t value  Pr(>|t|)
(Intercept)       83.146667    4.343305  19.14404   2.49e-14 ***
Intensity         -0.039867     0.007435  -5.36162   3.01e-05 ***
Timelate          -11.523333    6.142360  -1.87578   0.07534 .
Intensity:Timelate -0.001210     0.010515  -0.11513   0.90962

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.598 on 20 degrees of freedom
Multiple R-Squared: 0.7993,    Adjusted R-squared: 0.7692
F-statistic: 26.55 on 3 and 20 DF,  p-value: 3.549e-07
```

Up to this point, we have treated Intensity as a quantitative variable. We could also treat it as a factor with six levels. This would essentially be a two-way analysis of variance. Here is how to do this in R, first without an interaction term and then with one.

```r
> anova(fit2)

Analysis of Variance Table

Response: Flowers

            Df Sum Sq Mean Sq  F value Pr(>F)
Intensity    1 2579.75  2579.75  59.2597  2.1e-07 ***
Time         1  886.95   886.95  20.3742  6.8e-04 ***
Intensity:Time 1   0.58    0.58  0.0132  0.90968
Residuals 20  870.66   43.53

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Bret Larget

March 12, 2003
> fit3 <- lm(Flowers ~ as.factor(Intensity) + Time)
> summary(fit3)

Call:
lm(formula = Flowers ~ as.factor(Intensity) + Time)

Residuals:
     Min      1Q  Median      3Q     Max
-8.9790 -4.3080 -1.3420  5.2040 10.2040

Coefficients:
                        Estimate Std. Error t value  Pr(>|t|)
(Intercept)              79.354      3.629  21.869 6.92e-14 ***
as.factor(Intensity)300  -9.125      4.751  -1.921  0.071715 .
as.factor(Intensity)450 -13.375      4.751  -2.815  0.011919 *
as.factor(Intensity)600 -23.225      4.751  -4.888 1.38e-05 ***
as.factor(Intensity)750 -27.750      4.751  -5.841 1.38e-05 ***
as.factor(Intensity)900 -29.350      4.751  -6.178 1.97e-05 ***
    Time late        -12.158      2.743  -4.432 0.000365 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.719 on 17 degrees of freedom
Multiple R-Squared: 0.8231,   Adjusted R-squared: 0.7606
F-statistic: 13.18 on 6 and 17 DF,  p-value: 1.427e-05

> anova(fit3)

Analysis of Variance Table

Response: Flowers

            Df Sum Sq Mean Sq  F value    Pr(>F)
as.factor(Intensity)  5 2683.5  536.70  11.888 4.63e-05 ***
Time                1  886.95  886.95  19.646 0.0003649 ***
Residuals         17  767.47   45.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> fit4 <- lm(Flowers ~ as.factor(Intensity) * Time)
> summary(fit4)

Call:
lm(formula = Flowers ~ as.factor(Intensity) * Time)

Residuals:
     Min      1Q  Median      3Q     Max
-7.550e+00 -4.675e+00  9.992e-16  4.675e+00 7.550e+00

Coefficients:

Bret Larget                  March 12, 2003
Estimate | Std. Error | t value | Pr(>|t|)

(Intercept)  | 76.700      | 5.228   | 14.671 | 5.01e-09  ***
as.factor(Intensity)300 | -3.150    | 7.393  | -0.426 | 0.67760 |
as.factor(Intensity)450 | -12.650    | 7.393  | -1.711 | 0.11278  |
as.factor(Intensity)600 | -19.150    | 7.393  | -2.590 | 0.02365  *
as.factor(Intensity)750 | -23.750    | 7.393  | -3.212 | 0.00746  **
as.factor(Intensity)900 | -28.200    | 7.393  | -3.814 | 0.00246  **
Timelate | -6.850      | 7.393   | -0.927 | 0.37244  |
as.factor(Intensity)300:Timelate | -11.950 | 10.456 | -1.143 | 0.27536  |
as.factor(Intensity)450:Timelate | -1.450    | 10.456 | -0.139 | 0.89200  |
as.factor(Intensity)600:Timelate | -8.150    | 10.456 | -0.779 | 0.45080  |
as.factor(Intensity)750:Timelate | -8.000    | 10.456 | -0.765 | 0.45897  |
as.factor(Intensity)900:Timelate | -2.300    | 10.456 | -0.220 | 0.82959  |
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.393 on 12 degrees of freedom
Multiple R-Squared: 0.8488, Adjusted R-squared: 0.7102
F-statistic: 6.124 on 11 and 12 DF,  p-value: 0.002028

> anova(fit4)
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