I. Reflection on previous weeks
   1. Some folks had trouble constructing the linear regression equations.
      a. Feel free to interrupt during lecture.
      b. Or we can work out the details during the lab on Fri. We’ve done a good job of this the past couple weeks.
   2. Grading Lab Report 3 right now – they’re good overall.
   3. A couple of R hints (leaving the comfort zone of R Commander):
      a. Call variables using `dataset$variable`
      b. Name models using `<-(e.g. model <- lm(dataset$y ~ dataset$x))`
      c. Use the “up arrow” key to call up previous lines of code.

II. Non-parametric (“distribution free”) statistics
   1. We’ve talked a lot in this course about the importance of the normal distribution but what happens if our data are not normally distributed?
   2. Most non-parametric tests are based on ranking values.
   3. Non-parametric methods are preferred when:
      a. Distribution assumptions strongly violated (e.g. outliers, non-normality)
      b. Small samples sizes, can’t tell if the data are normally distributed.
      c. Detection limit problems can be accommodated.
   4. Why not just use non-parametric tests all the time?
      a. Parametric methods are more powerful – when data are indeed normally distributed (more likely to detect differences that are actually there).
      b. If data are not normally distributed non-parametric methods are more powerful.
      c. Difficult to extrapolate beyond the data.
      d. But, if you’re unsure, then the safe bet is to use non-parametric approaches.
   5. For descriptive statistics, median and IQR are non-parametric (i.e. robust)

III. Non-parametric statistical tests
   1. We’ve already discusses one non-parametric approach – Spearman’s rank correlation! (review the Regression lecture for details)
      a. Tests the monotonic association between two variables.
      b. **Parametric alternative = Pearson’s product moment correlation.**
      c. See! Non-parametric methods are easy!
   2. Mann-Whitney U-test/ Wilcoxon rank sum test (compare medians from two different groups/ samples)
      a. Rank values in both groups jointly (ties assigned by average)
      b. Sum ranks in one group
      c. Compute $U=R1-n1(n1+1)/2$
      d. Compute $U'=n1n2 – U$
      e. Take smaller of the Us and check “U-table” for significance (note that in this case a smaller U means greater significance)
      f. Or just use R!
g.  **Parametric alternative = t-test**

3. Paired-samples Wilcoxon test  
   a. Examine the differences between paired samples and rank the magnitude differences (split ties, ignore zeroes).  
   b. Then sum the ranks for each sign and sum for the test statistic and evaluate relative to a critical value table.  
   c. Or just use R!  
   d. **Parametric alternative = paired t-test**

4. Kruskal-Wallis test (compare medians from multiple groups/samples)  
   a. As with other non-parametric tests we rank all the values.  
   b. Sum the ranks, square the sums and divide by \( n \) (sound familiar?!)
   c. Use \( H = \frac{12}{N(N+1)} \sum \frac{R^2}{n} - 3(N+1) \)  
   d. We can check a Chi-square table for the \( p \)-value based on the test statistic \( H \) (of again, we can just use R).  
   e. Note that a Kruskal-Wallis test with two groups is functionally (but not quantitatively) the same as a Mann-Whitney U test.  
   f. **Parametric alternative = ANOVA**

III. Multiple comparisons (Tukey, Bonferroni etc.)  
1. Not accounting for multiple comparisons is one of the biggest, most rookie and most egregious mistakes in stats (in my humble opinion).  
2. We therefore must correct for the multiple comparisons and commonly (and easily) can do so using the “Tukey test” which is a modified t-test (modified to account for the number of comparisons).
   a. Rank means in ascending order  
   b. For each comparison compute Tukey’s post-hoc t-values  
   c. Find the associated values (using tables or R)  
   d. So now we can find which groups are different from one another.  
   e. But why not just do a regular old t-test?  
      i. \( p \)-values will be different and in fact, entire interpretation might be altered.  
      ii. Multiple t-tests across groups is a no-no!  
3. We can also use Bonferroni  
   a. Our \( p(\text{crit}) = \frac{\alpha}{k} \) where \( k \) = # of comparisons and \( \alpha \) is our significance level.  
   b. So if we have 10 groups and an alpha of 0.05 then we use 0.005 as our significance level.  
   c. Compare again to the t-test. How is it different?  
4. In addition to Bonferroni and Tukey there are Holm, Benjamini-Hochberg, Dunnett…  
   a. Bonferroni and Tukey probably the most common/standard.  
   b. Bonferroni is generally more conservative than Tukey.
5. What about our first ANOVA example, Chlosyne morphology?  
6. Also, non-parametric multiple comparisons are quite doable in R (but not R Commander) but are not standard.