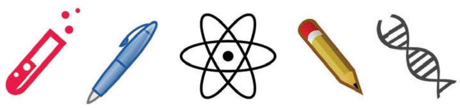
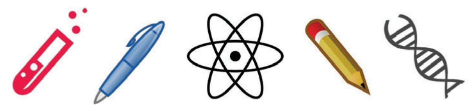
SciTools Workshop 2019



1. Scientific Writing



Writing Basics

- Everything you need to know about good science writing you can learn from Alfred Hitchcock
  1. **A good paper is about one limited concept**
  2. **With clear communication of the single concept**
     - Be able to communicate findings to a nonspecialist, don't let technical sophistication dominate the message
  3. **Every paragraph and every sentence counts**
     - The author should be able to justify each sentence and paragraph and how they flow naturally from one to the next
  4. **A good paper has suspense, not surprise**
     - Tell the reader what to expect in the introduction
     - If the reader does not understand where the argument is going they can become disinterested

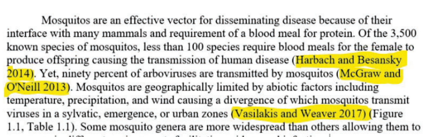
Kirshner 1996

Writing Basics: Parts of a scientific report

- Abstract: A brief summary of the work performed, its conclusions, and its broader meaning for the field
- Introduction: A treatment of the existing understanding of the problem which is being presented. Should include a hypothesis if your goal is to actually test something via experiment.
- Methods: Purely instructional, there will be no data in this section. This section is designed to give other researchers insight into how you got the results you got.
- Results: Purely data-driven, there should be *absolutely minimal* discussion of the results here, unless they DIRECTLY relate to understanding the meaning of subsequent results. Analyses and their results are presented in this section, as are most figures.
- Discussion: Lengthy examination of the data and their implications.
- Conclusion: A short statement of conclusion which summarizes the findings.

Writing Basics: Parts of a scientific report

- The meaning of a written work is never what the writer intended it to be, it is only what the reader understands it to be.
  - Understand your target audience, and use appropriate language for that audience
  - Cite you sources in text so the reader can reference previous primary literature



Mosquitos are an effective vector for disseminating disease because of their interface with many mammals and requirement of a blood meal for protein. Of the 3,500 known species of mosquitos, less than 100 species require blood meals for the female to produce offspring causing the transmission of human disease (Harbach and Besansky 2014). Yet, ninety percent of arboviruses are transmitted by mosquitos (McGraw and O'Neill 2013). Mosquitos are geographically limited by abiotic factors including temperature, precipitation, and wind causing a divergence of which mosquitos transmit viruses in a sylvatic, emergence, or urban zones (Vasilakis and Weaver 2017) (Figure 1.1, Table 1.1). Some mosquito genera are more widespread than others allowing them to persevere in different environments facilitating widespread infection.

  - The writer has a duty to proofread their work

Writing Basics: Writing sentences

- Each sentence has two components, the topic and the stress. These two work together.
  - Topic: Sets the reader's expectation that they will gain knowledge about this subject
  - Stress: Gives the reader the expected knowledge

It is linguistic commonplace

that readers of English Academic Prose naturally

emphasize the end of a sentence

Reader expects to learn what is common place

This answers the questions established in the topic position

Writing Basics: Flow and Paragraphs

- Sentences should connect to one another by their stress and topic positions. The stress position of sentence 1 should directly relate to the topic position of sentence 2. This creates an even flow and makes your writing more comprehensible.
- This practice of creating cohesive sentences makes paragraphs much neater and easier to understand

It is linguistic commonplace that readers of English Academic Prose naturally emphasize the end of a sentence.

Emphasizing the end of a sentence allows readers to more effectively understand your writing.

When TFs understand your ideas they are more likely to award you a higher grade.

- Paragraph structure is similar to sentence structure: the topic sentence introduces the idea to be discussed, there are statements of supporting evidence, and a statement of conclusion transitions to the next paragraph

Writing Basics: Hypothesis

- A hypothesis is a tentative, **testable** answer to a scientific question.
- A hypothesis leads to one or more predictions that can be tested by experimenting.

- *Don't retroactively think about your hypothesis. You should come up with one before your experiment, not during or after.*

Writing Basics: Figures and Tables

- Caption below for a figure and above for a table
- Provide a brief **title in bold** followed by a caption
- Every figure and table included must be referenced in text (Figure 1)
- A good caption should:
  - Stand alone
  - Define all the symbols, lines, colors, error bars, and other components that are not intuitive to the reader
  - Location if applicable
  - Any associated statistics should be cited

Writing Basics: Figures and Tables

t (time) = 15', T (temperature) = 32°C; t (time) = 12', T (temperature) = 32°C; t (time) = 6', T (temperature) = 29°C; t (time) = 9', T (temperature) = 31°C; t (time) = 3', T (temperature) = 27°C; t (time) = 0', T (temperature) = 25°C

vs.

Time (minutes)	Temperature (°C)
0	25
3	27
6	29
9	31
12	32
15	32

Writing Basics: Outlines

- Outlines are an extremely useful tool for drastically improving the flow of your writing
- They help to organize ideas into a coherent structure before those ideas are fully developed into sentences
- Outlines take time, but you save that time by reducing the amount of time you spend on editing ordering your ideas when they're written in paragraph style
  - Will be especially helpful for longer, full lab reports

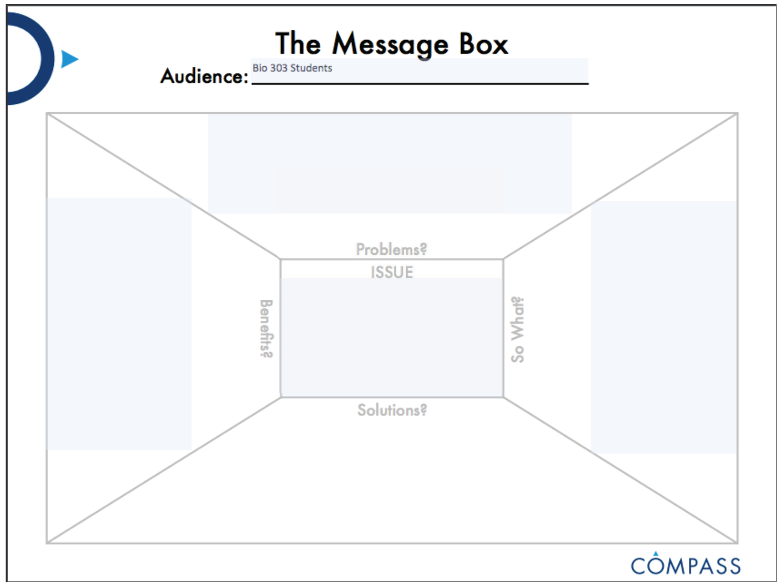
Writing Basics: Outline Structure

1. Abstract
   - Summarize motivation for, brief method, and main findings of a study
   - Should be written last
2. Introduction
   - Background info
   - Present the problem
3. Methods
   - Set up
   - Data collection
   - Analysis techniques
4. Results
5. Discussion
   - Interpretation of results
   - Implication of results
   - Suggested ideas for further research
6. Conclusion – not always necessary
   - Try not to blur the line with your Discussion. Think of this like your message box. Synthesize the problem, the "so what", and results and/or solution <u>if applicable</u>. Keep this short

Writing activity: Message Box

COMPASS: *"Our vision is to see more scientists engage, and engage effectively, in the public discourse about the environment."*

Your issue can be anything science related you care about. It can be the forest succession lab, an issue you have studied or seen in the media – anything!



Parts of the Message Box

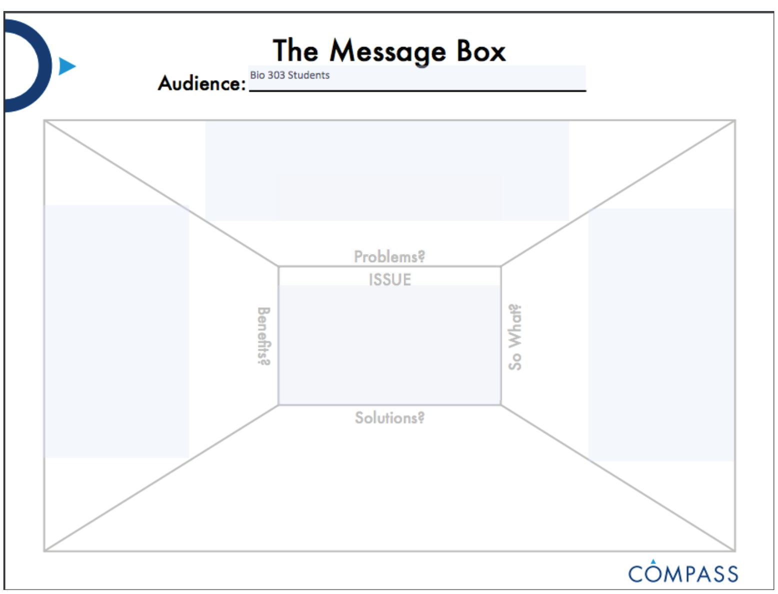
1. Who is your **intended audience**?
   - Remember that your audience matters!
   - Simplify to a level your audience can understand (and understand quickly!)
2. What is the **main issue**?
   - This should be short, one sentence or less. What is the main issue you'd like to discuss?
3. What are the problems/conflicts/issues involved **(Problems?)**
   - What problems do we face when trying to overcome the main issue?
   - What are some of the competing interests when addressing this problem?
   - Are there technical hurdles?

Parts of the Message Box

4. Why does this information matter to my listener **(So What?)**
   - Tell me why I should care – your *hook*

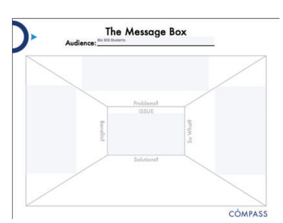
5. What are some of the possible solutions to the problem **(Solutions?)**

6. What are the potential benefits of resolving this problem **(Benefits?)**
   - Describe the positive outcomes of fixing the problem
   - List as many as possible, ideally more than the number of problems presented in the "problems" box

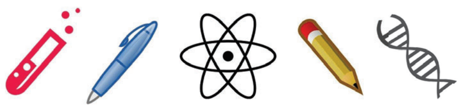


Activity: Message Box Swap

- Swap your message box with a partner at your table
- Read and discuss message boxes with each other
- I will ask volunteers from each table to outline the subject of their partners message box for the class
  - –These will also be collected

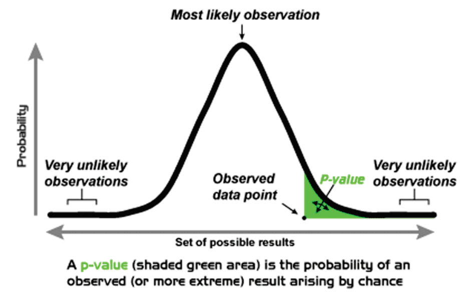


2. Statistics



Statistics Basics

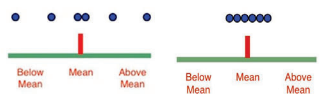
- Determining *significance* of differences between treatments in research studies
- How do we define significance?



Statistics Basics: Standard deviation vs. standard error

Standard deviation

- Measures variability in data
- How close the data is to the mean



Standard error

- A type of standard deviation that depends on the standard deviation and sample size
- SE = SD/√(sample size)
- The variability between your sample's mean and the entirety of the sample type
- How precise/certain is your sample mean?

Statistics Basics: Standard deviation vs. standard error

- We will use excel for calculations
- Example:

Geese	
10m Proportion of vigilant behavior	**3m** Proportion of vigilant behavior
0.55	0.3
0.6	0.4
0.6	0.4
0.7	0.5
0.9	0.55
0.9	0.7

- SE in excel for 10m: =(STDEV(array1))/(SQRT(COUNT(array1))
- SE in excel for 3m: =(STDEV(array2))/(SQRT(COUNT(array2))

Statistics Basics: T-tests and ANOVA

Unpaired T-test

- Compares the means of two sets of data when observations are made using distinct groups of test subjects

Paired T-test

- Compares the means of two sets of data when observations both from samples consist of the same test subjects.

ANOVA

- Compares the means between two or more sets of data with two variables

t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

$$SD = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

$\bar{x}_1$ = mean of first sample
$\bar{x}_2$ = mean of second sample
s_1 = standard deviation of first sample
s_2 = standard deviation of second sample
n_1 = number of measurements in first sample
n_2 = number of measurements in second sample

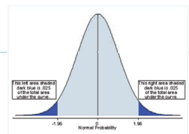
Statistics Basics: T-tests and ANOVA

- T-test function in excel:

TTEST(array1, array 2, 2, 2)

Alfalfa	
Control Proportion of seeds germinated	**Salt** Proportion of seeds germinated
0.9	0.3
0.7	0.5
0.6	0.4
0.7	0.5
0.55	0.2
0.8	0.7

<u>2 Tailed T-test</u>: Testing for a relationship in any direction, positive or negative



<u>Type 2 T-test</u>: 2 groups of data with the null hypothesis that both group means are equal

---

Alfalfa	
Control Proportion of seeds germinated	**Salt** Proportion of seeds germinated
0.9	0.3
0.7	0.5
0.6	0.4
0.7	0.5
0.55	0.2
0.8	0.7

=TTEST(I14:I19,J14:J19,2,2)

TTEST(**array1**, array2, tails, type)

Alfalfa	
Control Proportion of seeds germinated	**Salt** Proportion of seeds germinated
0.9	0.3
0.7	0.5
0.6	0.4
0.7	0.5
0.55	0.2
0.8	0.7

0.011167

Excel will return a **p value**

Are these groups significantly different?

Statistics Basics: Regression and correlation

Linear Regression

- Allows us to summarize and study relationships between two continuous (quantitative) variables
- How much do two variables correlate or "vary together"
- R-squared is a statistical measure of how close the data are to the fitted regression line
  - The R2 value is the percentage of the response variable variation that is explained by a linear model
  - Values closer to 1 = higher correlation

Statistics Basics: Chi-squared

- Used to test if two or more groups have a certain proportions
  - I.E. 1:1 sex ratio
  - Compares observed ratios to expected ratios
- Example: Do we see the same proportion of each color of M&Ms in a bowl?

$X^2 = (\Sigma(O-E)^2)/E$

O = observed

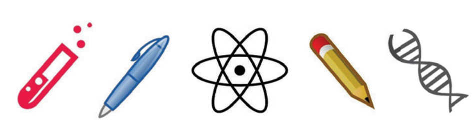
E = expected

Statistics activity: Pick the appropriate test

- I will show you a graph with a word bank of potential statistical tests that could be carried out to determine relationships
- On the back of your message box, write the graph number and the appropriate stats test
- When you are done, hand in your message box and then we will discuss the answers
  - For a grade!

---

3. Video projects – Questions?



Introduction
- Effectively presents relevant background information
- States objectives, hypotheses and justifies experiment
- All statements of fact are supported by references

Methods
- Methods for all experiments are described and rationalized
- Relevant information is present
- Statistics used are explained in your own words

Results
- Text summarizes major findings in the context of the objectives
- All relevant results and appropriate statistics are included
- Statistical results are correctly described
- Writing conveys good understanding of the topic
- Correct data is graphed
- Figures are appropriately captioned
- Error bars are present and correct (if required)
- Axes are appropriately labeled
- Figures are referenced in the text

Discussion
- Objectives and hypotheses are stated
- Results are briefly restated in context of hypotheses
- Results are analyzed correctly using background information
- Problems with experimental procedure are addressed
- Suggestions for applications/future research are made
- Appropriate closure (statement of greater importance of findings)
- Synthesizes results with previous empirical and theoretical findings

Evaluation of overall content
- Maximum length (1 page) not exceeded
- No/minimal typos or spelling mistakes
- Correct interpretation of results
- Overall knowledge of the topic well conveyed

Below is an example of a very good lab report from one of my past students. Notice that this student has fulfilled virtually every guideline in the above rubric/criterion list. The student includes only necessary information, effectively communicates the idea and the experiment, and contextualizes the data. This example report is not meant to be a format guide, but only as an example of what a good student report might look like.

Food Caching Behavior in Squirrels *(Sciurus carolinensis)*

Introduction: Foraging behavior can have both costs and benefits. To get through the seasons, animals can either choose to consume food which gives immediate benefits, or choose to cache food for later consumption which incurs costs (such as increased predation risk) and benefits (such as energy reserved for a later time) (Wasserman, 2015). The Eastern Grey squirrel *(Sciurus carolinensis)* exhibits just such foraging behavior. The objectives of this lab are to determine if the perishability of a food item affects a squirrel's choice to eat, cache, or reject the food. The Optimal Foraging Theory (OFT) predicts how animals behave in order to maximize benefits and minimize costs (Pyke, 1984). Hadj-Chikh et al previously demonstrated that perishability plays a role on the squirrel's decision to cache food items (1996). As such, we can hypothesize that squirrels will follow the OFT by preferring to eat items with higher perishability and to cache items with lower perishability, and that this behavior pattern will be more prominent during the summer than the autumn due to a greater food abundance. The null hypothesis is that squirrels will equally eat, cache or reject the food items randomly.

Methods: We performed this experiment at Boston Common during autumn and observed Eastern Grey squirrels because of their proclivity to cache food items. We approached one squirrel from behind and tossed either a peanut or a grape, which differ in perishability, from a height that simulated an acorn falling from a tree. We observed whether the squirrel ate, cached (dug a hole and buried the item), or rejected (ignored or partially ate) the food item. We repeated these steps for a total of 15 peanuts and 15 grapes tossed to a new squirrel each trial. We made sure not serially present food to the same squirrel, and attempted to observe as many different squirrels as possible so as to make each trial independent. We used the Chi-squared test for goodness of fit to see if our observed data fit the expected 1:1:1 ratio for behavior choice. We also used the two-tailed Binomial test to determine if squirrels preferred to cache one food item over the other (Wasserman, 2015).

Results: When we performed Chi-squared tests, results demonstrated that squirrel behavior significantly deviated from the null hypothesis of its behavior to eat, cache or reject food items equally for both grapes (X^2=87.31, $p<0.001$, n=165) and peanuts (X^2=84.25, $p<0.001$, n=165). See Figures 1 and 2. When we performed the two-tailed Binomial test, squirrels significantly preferred to cache peanuts over grapes ($p<0.001$, n=62). See Table 1.

Discussion: The first objective was to determine if a squirrel's decision to eat, cache or reject food was dependent on a food's perishability, and we hypothesized that it was dependent. In support of our first hypothesis, Chi-squared values were significant for both peanuts and grapes (Figures 1 and 2), suggesting that perishability may be an important factor (Hadj-Chikh et al, 1996). Interestingly, squirrels often ate the peanuts despite that nuts are of lower perishability. Perhaps squirrels had cached enough food in preparation for the winter at this time point. However, squirrels were still more likely to cache peanuts than grapes ($p<0.001$, n=62), and ended up eating slightly more grapes than peanuts (Table 1). Our results generally appear to be consistent with the predictions of the OFT. Additionally, we posit that squirrels probably exhibit this overall behavior pattern constantly and may cache more non-perishable foods more often during summer, as there may be a greater food supply due to warmer temperatures (Kochert et al, 1996). Thus, squirrels may have a greater reserve of energy to consume during the winter when resources became scarce (Delgado et al, 2014). However, we would need to conduct this experiment during the summer to test the veracity of that statement. Sources of error include human interference and inconsistent throwing of food items. We can repeat this experiment in a more natural environment without human presence, and can perform future studies to test other factors that affect caching behavior, such as competition. In conclusion, food perishability does play a significant role in foraging behavior in Eastern Grey squirrels, and may also apply to other animals with similar food caching behaviors.

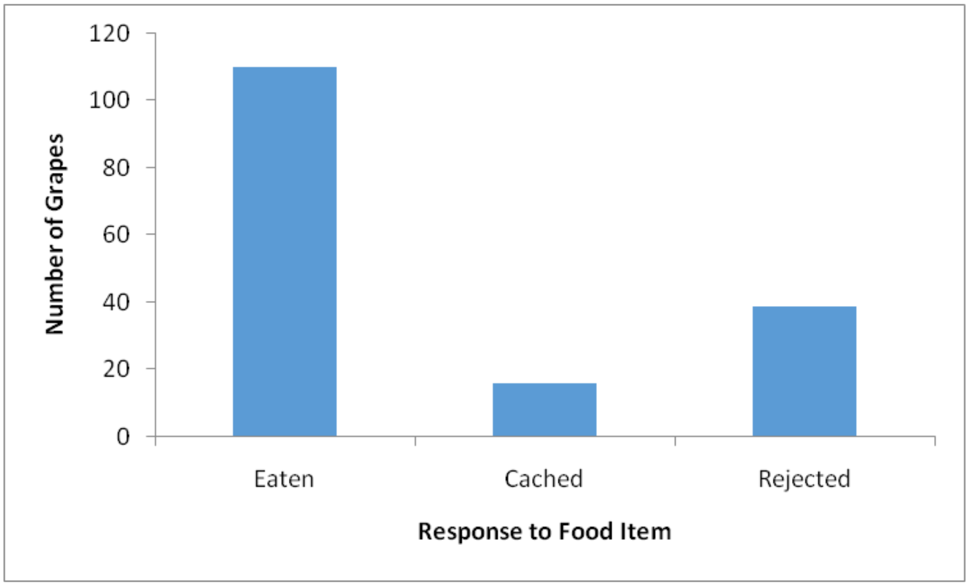


Figure 1. Total number of grapes eaten, cached or rejected by Eastern Grey squirrels. The greatest number of grapes presented were eaten by squirrels (110). The Chi-squared test demonstrated the results were significant (X^2=87.31, p<0.001, n=165).

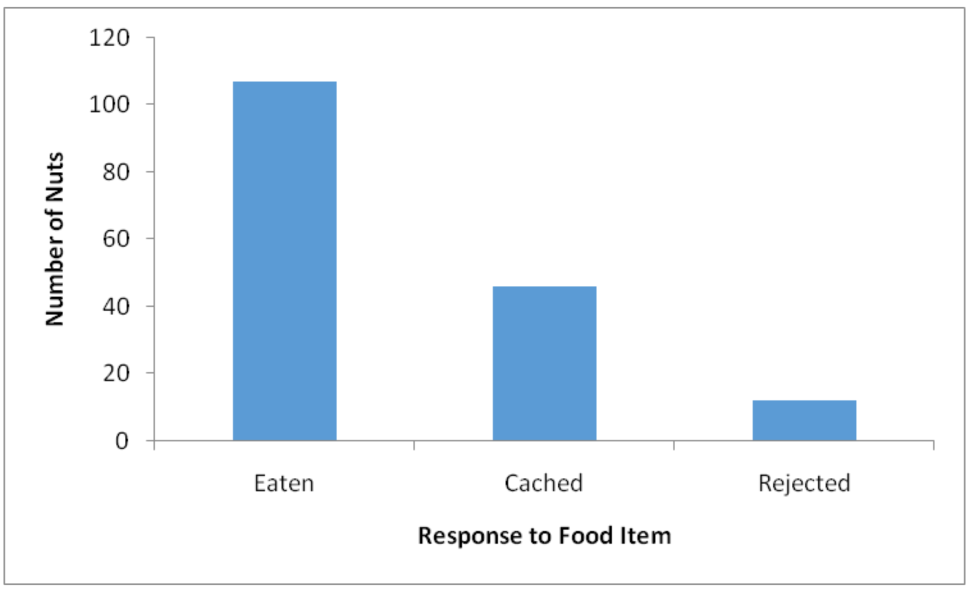


Figure 2. Total number of peanuts eaten, cached or rejected by Eastern Grey squirrels. The greatest number of peanuts presented were eaten by squirrels (107). The Chi-squared test demonstrated the results were significant (X^2=84.25, p<0.001, n=165).

Table 1. Total number of peanuts and grapes eaten, cached or rejected by Eastern Grey squirrels. The two-tailed Binomial test demonstrated that the results were significant for caching (p<0.001, n=62).

	Eaten	Cached	Rejected
Nuts	107	46	12
Grapes	110	16	39

Works Cited

Delgado MM, Nicholas M, Petrie DJ, Jacobs LF. 2014. Fox Squirrels Match Food Assessment and Cache Effort to Value and Scarcity. PloS one, 9(3), e92892.

Hadj-Chikh LZ, Steele MA, Smallwood PD. 1996. Caching decisions by grey squirrels: A test of the handling time and perishability hypotheses. Animal Behaviour. 52:941-948.

Kochert G, Stalker TH, Gimenes M, Galgaro L, Lopes CR, Moore K. 1996. "RFLP and Cytogenetic Evidence on the Origin and Evolution of Allotetraploid Domesticated Peanut, *Arachis hypogaea (Leguminosae).*" American Journal of Botany 83 10:1282–1291.

Lima SL and Valone TJ. 1986. Influence of predation risk on diet selection: a simple example in the grey squirrel. Animal Behaviour. 24:536-544.

Pyke GH. 1984. Optimal foraging theory: a critical review. Annual Review of Ecology and Systematics. 15:523-575.

Wasserman F. 2015. “Food Caching Behavior in Squirrels (*Sciurus carolinensis)*.” Lab procedure. Boston University. Boston.