

Critique of Laufenberg Paper

Stat 512:Final Project

Steve Huysman & Parker Levinson

Friday, May 5, 2023

Contents

0.1	Introduction	1
0.2	Methods	2
0.3	Results/Summary of Statistical Findings	8
0.4	Scope of Inference	9
0.5	Critique	9
0.6	Group Work Statement	9
0.7	References	9
0.8	Appendix	10

0.1 Introduction

Whitebark pine (*Pinus albicaulis* Engelm.; abbreviation: WBP) is a conifer tree native to the mountains of the western United States and Canada. It inhabits subalpine areas where it can be found growing up to the tree line, at a higher elevation than other tree species found with it. Whitebark pine is an early successional species that is often the first to establish after disturbance such as wildfire. WBP is a keystone species of subalpine environments where it plays important ecological roles such as providing food for wildlife such as Clark's Nutcracker and the threatened Grizzly Bear.

Due to threats from climate change, mountain pine beetle, and the invasive white pine blister rust, Whitebark pine has undergone a rapid and widespread decline. It was recently estimated that over half of all standing WBP in the United States are dead. This decline has lead to its recent listing as Threatened under the Endangered Species Act. Future climate

projections indicate further deterioration of WBP's habitat. Strategies to conserve this species involve planting WBP seedlings for restoration of high-elevation forests. Successful plantings in the face of climate change require an understanding of the relationship between climate and seedling establishment and growth in this species. Competition from other tree species also plays a role in seedling establishment and was investigated here.

Laufenberg et al. (2020) investigated USFS plantings of WBP seedlings in the Greater Yellowstone Ecosystem (GYE) to answer two research questions:

- 1) What is the relationship between climate/competition and WBP seedling establishment, measured by individual growth rate?
- 2) What is the relationship between climate/competition and WBP seedling survival, measured by density change?

However, we chose to focus on a simplified question. We are interested in how annual evapotranspiration and water availability affect planted WBP seedling growth after accounting for location, competition, and climatic variables.

0.2 Methods

0.2.1 Field Methods/Study design

Over the past 40 years, the US Forest Service and National Park Service has planted more than 1500 acres of WBP in the GYE. This study investigated five planting units (Figure 1) that each contained between two and eight planting sites. This study used a hierarchical sampling design including 5 planting units, with a total of 29 planting sites across units, and thousands of white bark pine seedlings per planting site. (See Figure 2 for model design.)

In the 5 larger planting units, no data aside from name of the unit, were collected. For the 29 sites scattered across those planting units (an unbalanced design), larger climatic data were recorded such as average annual temperature and average annual precipitation. (For a full list of data collected at each location, see Figure 2. Then, seedlings within that site were sampled for annual growth rate. Seedlings were too small to measure growth rings via coring,

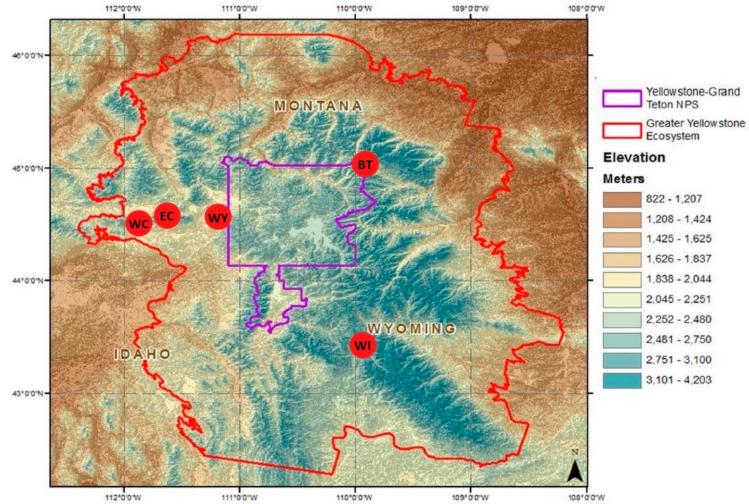


Figure 1: Map of planting units included in study in the Greater Yellowstone Ecosystem. Yellowstone National Park is outlined in purple, the greater yellowstone ecosystem is outlined in red, and the five white bark pine planting units run by the Forest Service and National Park service are circled in red. The two-letter abbreviation denotes the name: BT = Beartooth, EC = East Centennial, WC = West Centennial, WI =Wind River, WY =West Yellowstone

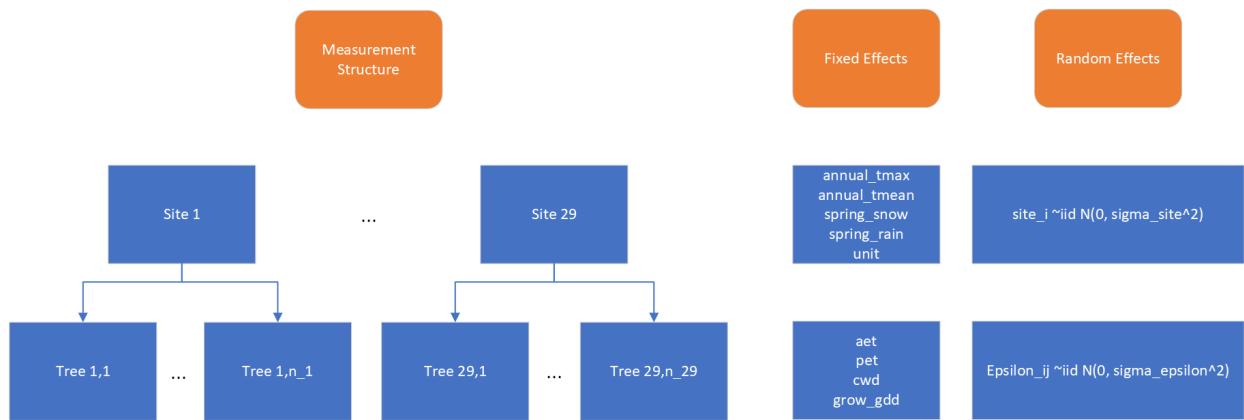


Figure 2: Hierarchical study design with fixed and random effects.

so height was used as a proxy for growth rate. Specifically, growth rate was calculated as the change in height between the study year (2018) to the relative planted height when the seedling was first planted. This was divided by the number of years since planting minus 2.5 years to account for the period of time when seedlings sequester carbon instead of focusing on their own growth.

At each site, climate data like elevation, aspect, water deficit, number of competitors, and amount of evapotranspiration were collected. Water balance variables were estimated using Daymet devices as inputs for temperature and precipitation. Then, specific data were taken for each individual tree, including height, number of competitors, and climatic variables. Sampling individual seedling growth occurred from May 2018 to October 2018. A grid cell matrix of 10 meters x 10 meters was overlaid on the study site or unit. A random starting point was decided and then every 20th grid cell from that was sampled, equating to sampling WBP in 2-15% of each site. Each seedling within that grid cell was digitally tagged, and Survey123 was used to collect field data. Researchers were most interested in the annual growth rate. Seedlings were too small to measure growth rings via coring, so height was used as a proxy for growth rate. Specifically, growth rate was calculated as the change in height between the study year (2018) to the relative planted height when the seedling was first planted. This was divided by the number of years since planting minus 2.5 years to account for the period of time when seedlings sequester carbon instead of focusing on their own growth.

A variety of data were collected at different scales. The variables that are most pertinent to the research question are:

- *growth_rate* (cm/year): WBP seedling height is used as a proxy for growth rate. This is a continuous response variable and is used as a proxy for seedling growth.

- *AET* (mm): This is a continuous predictor variable for annual evapotranspiration, the water loss through transpiration by plants on a site, given prevailing water availability. This measurement indicates the magnitude and duration of conditions favorable for plant growth on a site. This was estimated using a water balance model using historical climate data from Daymet (Thornton et al. 2016) for coordinates taken at each seedling.
- *CWD* (mm): This is a continuous predictor variable for climatic water deficit, which

measures the evaporative water demand that is not met the water supply at a site. It reflects the drought stress experienced by plants. It was estimated using the same climate modeling techniques used to estimate AET for coordinates taken at each seedling.

- *Micro (1/0)* - This was a binary variable indicating presence of favorable microsite conditions. 1 if there was a rock or other topographical feature that changed the environmental conditions where the seedling lived.

Data were provided by Laufenberg, but the process of cleaning and structuring the data were not explicit in the paper. As such, we had to experiment to figure out how data were cleaned. We ultimately decided to create an annual evapotranspiration rate, given that it is extremely pertinent to our research question. To get this annual evapotranspiration (AET), we multiplied the average monthly evapotranspiration by 7 for the 7 months of the growing season. The same was done for mean water deficit (*grow_dmean*) to get climatic water deficit (CWD).

Climate and water balance predictor variables were tested for collinearity with a cutoff of $r \geq +/- 0.6$ (Figure 3). A parsimonious list of water balance variables was selected by choosing the more biologically relevant variable from pairs that exceeded this threshold. This list of variables was then combined with the variables pertinent to the research question, which are *AET* and *WD*. Following this process, our list of environmental components we would like to account for in the model are *comp_number* and microsite status.

After data were selected based on their correlation, we looked at linearity and potential other polynomial forms (Figure 4). We then looked at variance between sites (Figure @ref(fig:site_var)), which confirmed the need to include a random effect of site.

0.2.2 Statistical Procedures Used

Mixed effects models were used to study what covariates were correlated with WBP performance. Random effects were always used for the site and fixed units were used for the study unit, due to only having 5 units. A variety of linear and polynomial functions, and fixed effects were explored. Cubic forms were explored due to the biological idea that WBP seedling establishment may vary drastically depending on the conditions. In below average conditions, we expect low growth rate. We expect there to be an average growth rate in average conditions, and then for high quality conditions, we expect a much higher growth rate - therefore following a cubic form. A corrected AIC was used to compare models. We aimed to find the most parsimonious set of water balance variables while still incorporating the biologically relevant variables and incorporating environmental variables related to the research questions (AET and CWD).

The theoretical model for our full model analyzed is as follows:

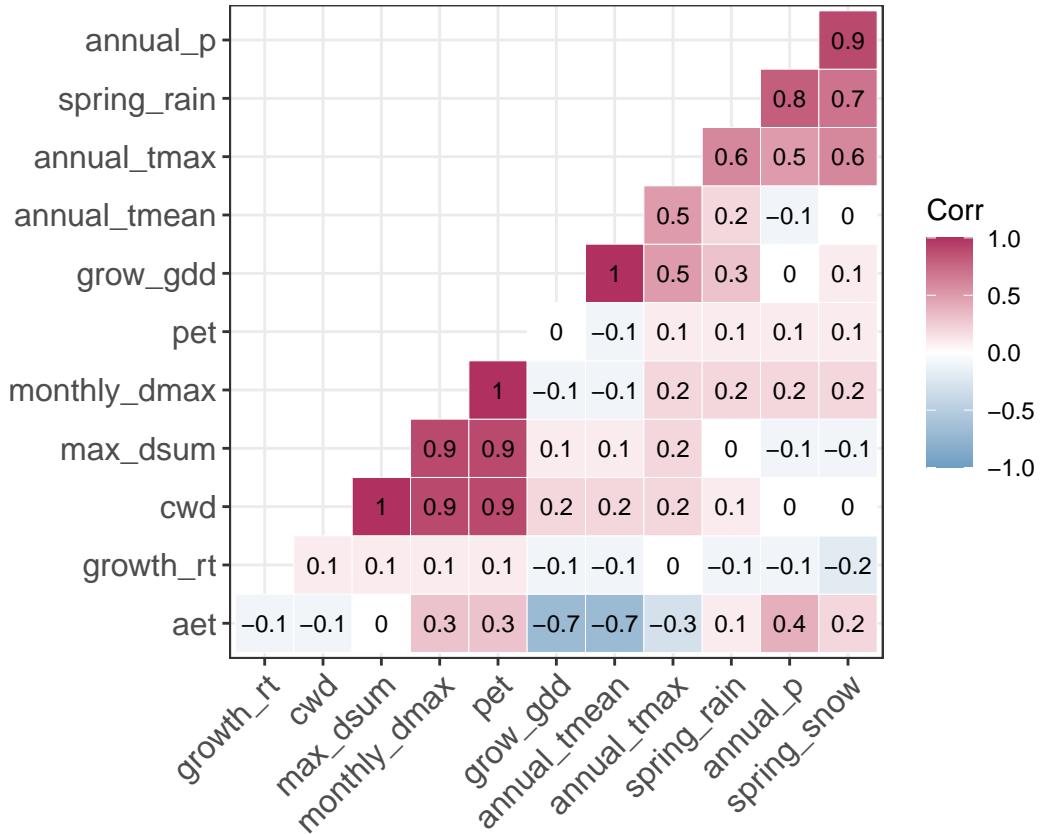


Figure 3: Correlation matrix of predictor variables. Darker squares indicate a higher degree of correlation. For variables that were at least 0.6 correlated, the most ecologically relevant one was selected

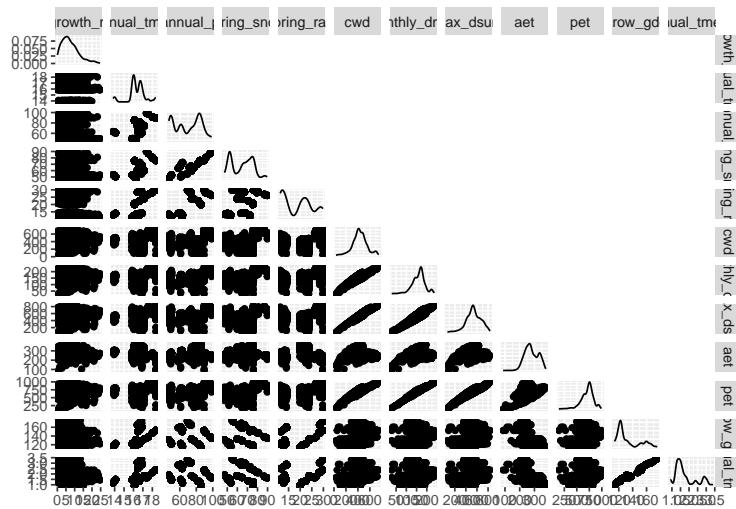


Figure 4: Raw data visualization of variables of interest

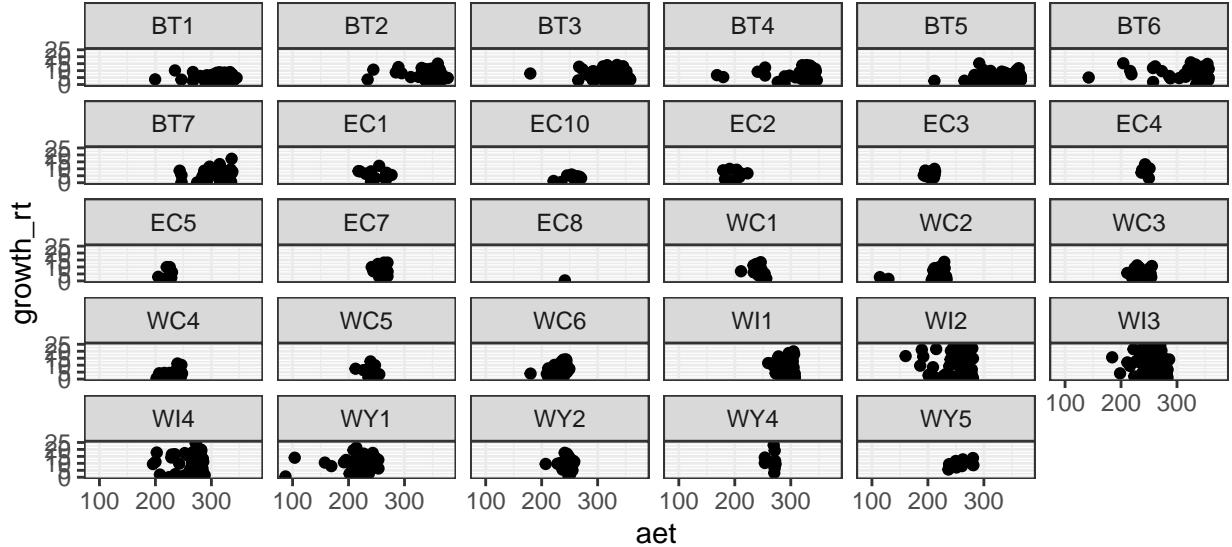


Figure 5: (#fig:site_var)Growth rate compared with AET by each site

$$\log(growth_rate)_{ij} = \mu_{ij} + Site_i + \epsilon_{ij} \quad (1)$$

$$Site_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{Site}^2)$$

$$\epsilon_{ij} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{error}^2)$$

$$\begin{aligned} \mu_{ij} = & \beta_0 + \beta_1 * AET + \beta_2 * CWD + \beta_3 * I_{Unit=EC} + \beta_4 * I_{Unit=WC} \\ & + \beta_5 * I_{Unit=WI} + \beta_6 * I_{Unit=WY} + \dots \beta_i * climate\ variables \end{aligned}$$

$$I_{unit=WC} = \begin{cases} 1, & \text{if } unit = West\ Centennial \\ 0, & \text{otherwise} \end{cases} \quad I_{unit=EC} = \begin{cases} 1, & \text{if } unit = East\ Centennial \\ 0, & \text{otherwise} \end{cases}$$

$$I_{unit=WI} = \begin{cases} 1, & \text{if } unit = Wind\ River \\ 0, & \text{otherwise} \end{cases} \quad I_{unit=WY} = \begin{cases} 1, & \text{if } unit = West\ Yellowstone \\ 0, & \text{otherwise} \end{cases}$$

β_1 and β_2 will allow us to answer the research question of how AET and CWD affect WBP seedling growth. $\beta_3 - \beta_6$ help get at the question of what variability exists in WBP seedling growth between platning units.

We ran the saturated cubic model without interactions (Equation (1)), and then refined it using backwards selection from the *step()* function and a AIC cutoff of AIC<2.

After running the model initially, we found that the residuals had a lot of funneling (Figure 6), and we found that log transforming the response variable (*growth_rate*) reduced this skew. However, log transforming does result in a non-random distribution of residuals, so we tried a variety of different functional transformations to both the response and predictor

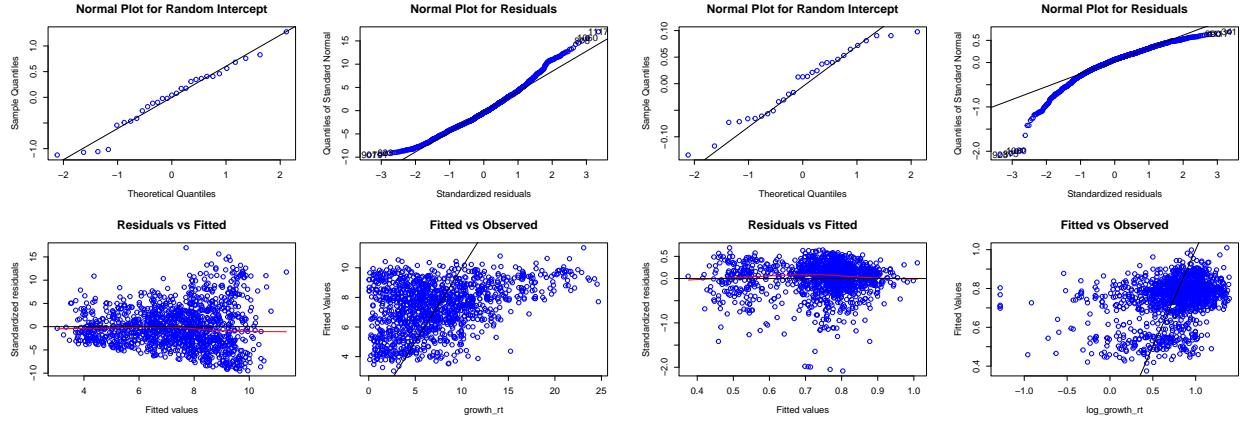


Figure 6: Diagnostic plots of saturated model before and after log transformation of growth rate

variables but were unable to reduce this violation. All analysis was done use R statistical software (v4.2.2; R Core Team 2021).

0.3 Results/Summary of Statistical Findings

Using backwards selection, the most parsimonious model included a random effect of site and fixed effects for unit, aet, and number of competitors. Although water deficit was not selected, we included it as it directly helps us answer our research question. The select model had an AICc of 1148.38. The full estimation is below:

$$\begin{aligned}
 \log(growth_rate)_{ij} &= \hat{\mu}_{ij} + Site_i + \epsilon_{ij} \\
 Site_i &\stackrel{iid}{\sim} \mathcal{N}(0, \hat{\sigma}_{Site}^2) \\
 \epsilon_{ij} &\stackrel{iid}{\sim} \mathcal{N}(0, \hat{\sigma}_{error}^2) \\
 \hat{\mu}_{ij} &= 0.22 + 0.0014 * AET - 0.000094 * CWD + 0.15 * I_{Unit=EC} + 0.14 * I_{Unit=WC} \\
 &\quad + 0.18 * I_{Unit=WI} + 0.36 * I_{Unit=WY} + 0.0084 * comp_number
 \end{aligned}$$

The final model suggests that WBP seedling growth is most correlated with the annual evapotranspiration and the number of competitors around the seedling. For a one **UNIT** change in AET, there is a multiplicative change in the median seedling growth rate by 1.001447 **UNIT** (95% CI: -2.64e-05 to 2.94e-03, profile-likelihood CI) and for a one **UNIT** change in CWD, there is a multiplicative change in the median seedling growth rate by 0.9999055 **UNIT** (95% CI: -5.25e-04 to 3.373e-04, profile-likelihood CI), after accounting for number of competitors, unit, and a random effect of site. After accounting for AET, water deficit, and number of competitors, the estimated correlation between any two seedlings in the same site $\hat{ICC} = 0.217$.

WHAT DOES THIS MEAN FOR WATER WBP RESTORATION

0.4 Scope of Inference

Because there was random sampling of the WPB within each site, the inferences can be applied to the larger WBP population at that site, but because the sites and units were not randomly selected, no further inferences can be drawn. Furthermore, there was no treatment applied since this was an observational study, so inferences are only correlative, not causative.

0.5 Critique

Trying to follow and reproduce what Laufenberg et al. 2020 did in the paper proved to be extremely complicated. There appeared to have been extensive data cleaning and manipulation, which was not communicated in the paper. We know that they chose the “more ecologically relevant” variable when looking at correlation, but there was no clear instructions on what that was. Furthermore, the paper had no clear model selection processes, making it difficult to reproduce. Many of their choices, such as using a cubic form for AET or log transforming growth rate, were not explained and the reasoning behind was unclear. The paper itself is fishing for an explanation. Furthermore, the models selected in the paper severely violate many of the assumptions of linear models, specifically normality. Unfortunately, our simplified model also violated the normality assumption, suggesting that perhaps the dataset is missing an important climatic variable that may help explain WBP seedling growth rate.

After trying and failing to reproduce the results from the paper, we streamline our research question and attempted to just use the data collected to answer what we were interested in. However, even that proved to be challenging.

0.6 Group Work Statement

Steve understands the ecological implications of this work as it closely mirrors his thesis research, and thus he has a very good understanding of the types of data and how to analyze them. He spearheaded the data wrangling and analysis. While Parker does not understand the study system as well, she is quite comfortable in R and in communicate results, so Parker took lead on the write-up and formatting. Both people contributed to the model selection, interpretation, discussion, and critique. We mostly worked remotely, sharing our progress via GitHub so we could track the changes that were made. We met once or twice a week to discuss variable and model selection.

We allocate 10 points to each person, because this was a true collaboration. Steve was integral in handling this messy dataset, figuring out which variables were ecologically important, and in interpreting results. Parker was crucial in writing and formatting the report, in thinking of alternative analyses we could do, and in diagnosing issues. We worked closely to figure out how we wanted to do our model selection.

0.7 References

- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:

10.18637/jss.v067.i01.

- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Laufenberg, David, et al. “Biophysical gradients and performance of whitebark pine plantings in the Greater Yellowstone Ecosystem.” *Forests* 11.1 (2020): 119.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Thornton, P.E.; Thornton, M.M.; Mayer, B.W.; Wei, Y.; Devarakonda, R.; Vose, R.S.; Cook, R.B. Daymet: Daily Surface Weather Data on a 1-km Grid for North America; Version 3; ORNL DAAC: Oak Ridge, TN, USA, 2016.

0.8 Appendix

- Must include a compiled RMarkdown with all of our results
- All code is currently in an R script, will attach for final paper