Mid Project Update

Kevin, Brandon, Eli, Seoyeong

Hypothesis

Precipitation, temperature, and fires will affect the abundance and diversity of various species of macrofungi (typically the Ascomycetes and Basidiomycetes; in general, fungi with visible fruiting bodies; for example in this study there will also be using slime molds, Myxomycota, because they are included in the MushroomObserver database) in their respective areas of growth in California.

Predictions

Precipitation

Fungi are made of hyphae, which are one cell thick, and which aggregate to form mycelium. Mushrooms are fungal fruiting bodies (sporocarps), made of hyphae, and have a very high water content; they require significant ambient humidity to form.[1],[2]

Therefore, we predict that increasing precipitation is positively correlated with macrofungi diversity (abundance and evenness). In other words, lower precipitation and lower water available results in fewer fruiting bodies of any species.

Temperature

Increasing temperature will decrease macrofungal abundance; fungal fruiting bodies maximize surface area via gills and pores in order to maximize spore transmission; the increased surface area and one cell thin hyphal structure put sporocarps at a high risk of water loss. Increased temperatures also means increased water evaporation and decreased water resources available for fungal fruiting bodies to form.[3],[4] Of course, near or below zero temperatures would also prevent fungal growth due to freezing; this limitation is apparent and thus we are mainly exploring the effect of increasing temperatures (When we refer to lower temperatures we mean still-ideal temperatures).

Additionally, species evenness will decrease with increasing temperature, as some fungi are more resistant to desiccation or otherwise well suited for forming fruiting bodies at higher temperatures.

[1]Sources: Lendzian KJ, Beck A. 2021. Barrier properties of fungal fruit body skins, pileipelles, contribute to protection against water loss. Scientific Reports. 11(1). doi:10.1038/s41598-021-88148-0. https://doi.org/10.1038/s41598-021-88148-0.

[2]Stojek K, Gillerot L, Jaroszewicz B. 2022. Predictors of mushroom production in the European temperate mixed deciduous forest. Forest Ecology and Management. 522:120451. doi:10.1016/j.foreco.2022.120451.

[3]Source: Straatsma G, Ayer F, Egli S. 2001. Species richness, abundance, and phenology of fungal fruit bodies over 21 years in a Swiss forest plot. Mycological Research. 105(5):515–523. https://doi.org/10.1017/s0953756201004154.

[4]A sudden decrease in temperature is also known to kickstart fungal growth (Pinna S, Gévry M -f., Côté M, Sirois L. 2010. Factors influencing fructification phenology of edible mushrooms in a boreal mixed forest of Eastern Canada. Forest Ecology and Management. 260(3):294–301. doi:10.1016/j.foreco.2010.04.024. https://doi.org/10.1016/j.foreco.2010.04.024.

Precipitation and Temperature interaction

Interactions between temperature and precipitations are sure to happen as both are important growth factors for mushrooms. The magnitude of these interactions are bound to vary, as our data relies on ambient fluctuations in the two which are determined by an area's general climate zone. For example, our primary investigation will be in California; the climate in California tends towards hot, dry summers and mild, wet winters. Since fungi require high moisture availability for growth and lower temperatures (reducing evaporative water loss), we predict that the winters (low temperature and high precipitation) will have greater macrofungi diversity – low temperature alone, for example, would have lower levels of fungal abundance since the supposed benefits of lower temperatures would not be apparent without enough precipitation to begin with.

More generally, we would predict that as temperature increases, low precipitation will lead to a greater reduction in mushroom abundance/diversity (than if precipitation was the only variable) because low precipitation is compounded by increased evaporation due to high temperature, further decreasing the moisture available to mushrooms. On the flip side, as temperature increases, high precipitation will lead to a lesser increase in mushroom abundance/diversity (than if precipitation was the only variable) because the increased evaporation caused by high temperature would reduce the effective of the precipitation.

Fire

We would predict that fire is negatively correlated with abundance because the high temperature is not tolerable by most species of fungi, which would not only burn most specimens but also evaporate most traces of water. Additionall, we predict that fire is negatively correlated with species evenness/abundance as fires significantly alter the soil composition of the affected area. These conditions are more suitable for pioneer species that are specialized to exploit the available resources, meaning that they would take advantage of this area with fewer competitors.[5]

Data description

Mushroom Dataset

There are three separate csv's (although we need to read them as tsvs) from MushroomObserver that we will be using: the primary **observations.csv** as well as **locations.csv** & **names.csv**. - From **observations.csv**, we will be using the columns *name_id* (a numerical code referencing an identification in **names.csv**), *when* (year month day), and *location_id* (a numerical code referencing a location in **location.csv**). Additionally, the *vote_cache* column gives a crowd-assessed confidence that the species identification is correct, and *is_collection_location* denotes whether or not the observation was found locally or if the specimen was brought in from a different location - **locations.csv** importantly gives the location name (column: name) and latitudinal/longitudinal (columns: North, East, South, West) bounds that each location ID refers to - **names.csv** gives the identification (column: text_name) associated with each name ID, as well as taxonomic ranking (column: rank - some observations are only prescribed a genus without species, for example).

Note: Since MushroomObserver is a citizen science platform similar to iNaturalist, data may be duplicated, inaccurate, or biased in sampling due to accessibility or environmental hazards (fires or hazardous locations, extreme weather conditions, or even the popularity of mushroom foraging may impact numbers of observations). Generally, it introduces the additional element of human behaviour which will be inherent within the data. We will need to be conscious of this and should account for it in analysis where possible.

[5] Treseder KK, Mack MC, Cross A. 2004. RELATIONSHIPS AMONG FIRES, FUNGI, AND SOIL DYNAMICS IN ALASKAN BOREAL FORESTS. Ecological Applications. 14(6):1826–1838. doi:https://doi.org/10.1890/03-5133

Data manipulation

The three csv files must be brought together into a single data frame for exploration and analysis. In short, **locations.csv** and **names.csv** were used as dictionaries to apply the location and identification data onto **observations.csv** using "left.join()". Additionally, we filtered the observations to those only found in at the specified location (*is_collection_location = 1*) and used the central latitude and longitude for each location (average of the North-South, East-West numbers).

It is unfeasible for us to use all of the observations within the data set as it has data from across the globe; our initial investigation will be in California (specifically, the San Francisco Bay area for climate) where observations are in large numbers, but we may branch out to other locations as the process unfolds. Importantly, for each new location we wish to examine, we require local climate and wildfire data to match with, which imposes a large restraint on which observations we can use. For each location, we will be accordingly subsetting the observations to match either location name or latitude/longitude area.

When we look at abundance, we group observations by the desired variable(s) and tally the number of observations seen within a certain area, time frame, etc., ignoring the identification column. For diversity, the identification column must be incorporated. As well, the *rank* and *confidence* are metrics that we will be using in analysis. We may wish to see how results vary depending on the taxonomic level we wish to examine/separate by, or if any changes in significance are observed if we choose to include/exclude unconfident identifications.

Precipitation and Temperature Dataset

For San Francisco Bay area climate, we will be using the columns: *date* (year month day), *temperature_avg*, (average temperature per time interval, both monthly and daily, in Celsius) and *precipitation* (total millimetres per time interval) retrieved from the Oakland North California climate database from the RAWS USA Climate Archive, hosted by the Western Regional Climate Center. We are using data between the years 2007 to 2022, to avoid large amounts of missing temperature data from February to July 2006 (2006 being when mushroom observations increase). There is also missing temperature data for 11/01/2008, 11/26/2008, 12/20/2008 to 02/11/2009, 02/17/2009, 02/23/2009, 03/15/2010 to 04/14/2010, 12/29/2011 to 01/04/2012, 02/20/2016 to 02/21/2016, 07/22/2021. For this, the website outputs data to be copied in a text document and then read in as a table.

Much like locations and names, this climate data is matched to the observation data frame using "left.join()", pairing the date of each observation with an (approximate) temperature and precipitation level.

Should we use data from other locations in the future, this process would be identical - taking average temperatures, total precipitations, and matching measurements to dates.

Fire Dataset:

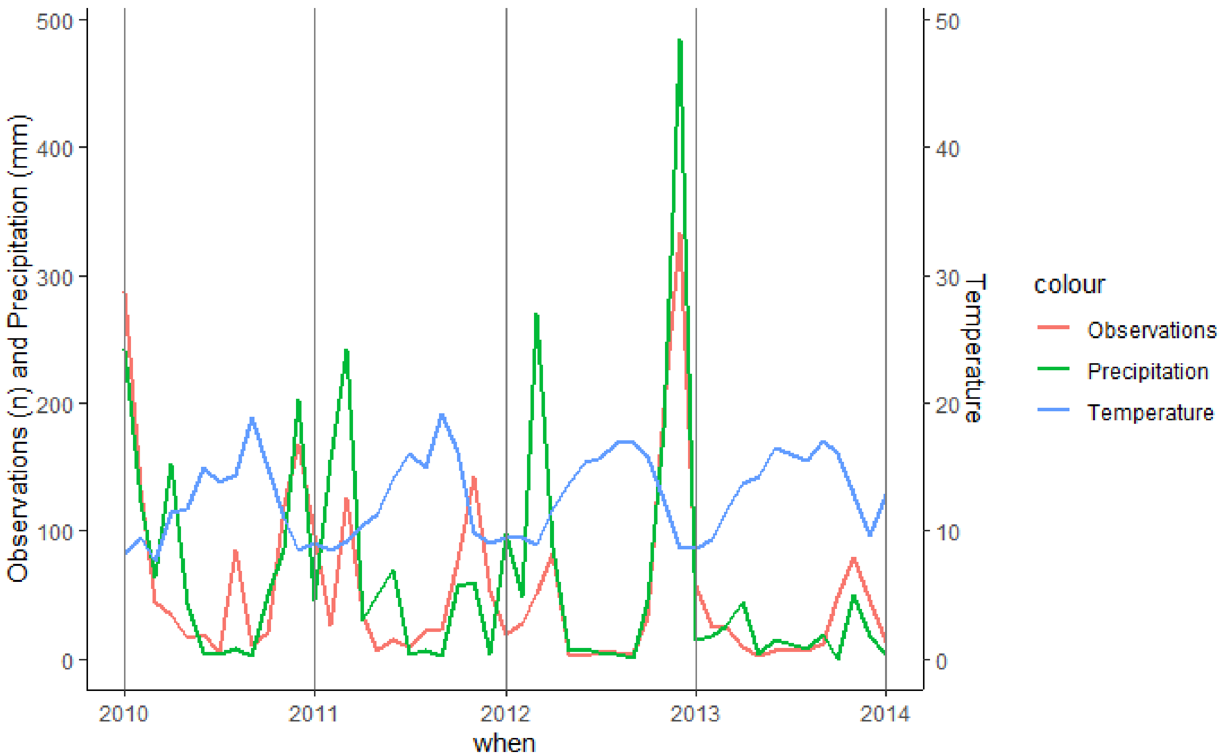
For California wildfires, We will be using the columns incident_name (fire name), incident_date_created (proxy for date of fire), incident_county (county the fire occurred in), incident_acres_burned (size, ie severity of fire), and incident_latitude & incident_longitude from the Incident Data for California from CalFire. We plan on using the mapdataall.csv. Relevant units for these columns is acres for incident_acres_burned and date for incident_date_created. Some fires do not have a date associated with them; we will be removing them.

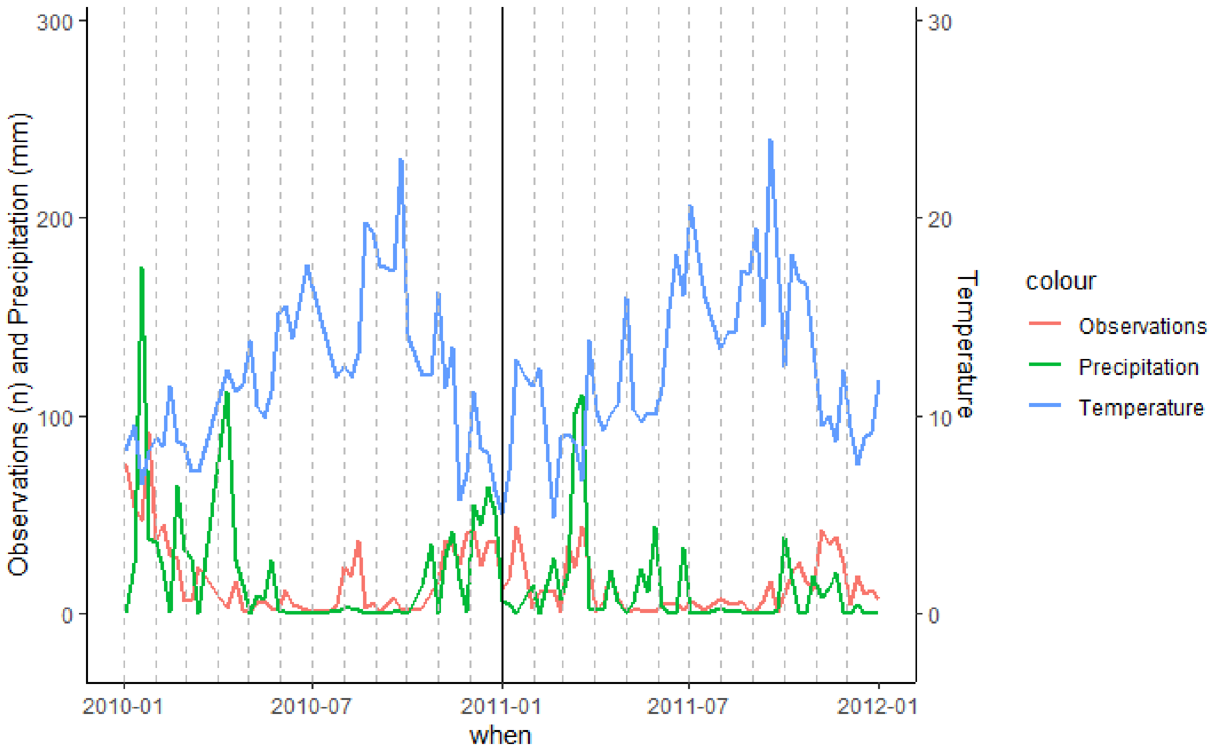
We do not plan on merging fire data with the observations file, but we will be using the fire dates, location, and size to subset the observations file to test for each respective fire's effect on observations.

Data Exploration

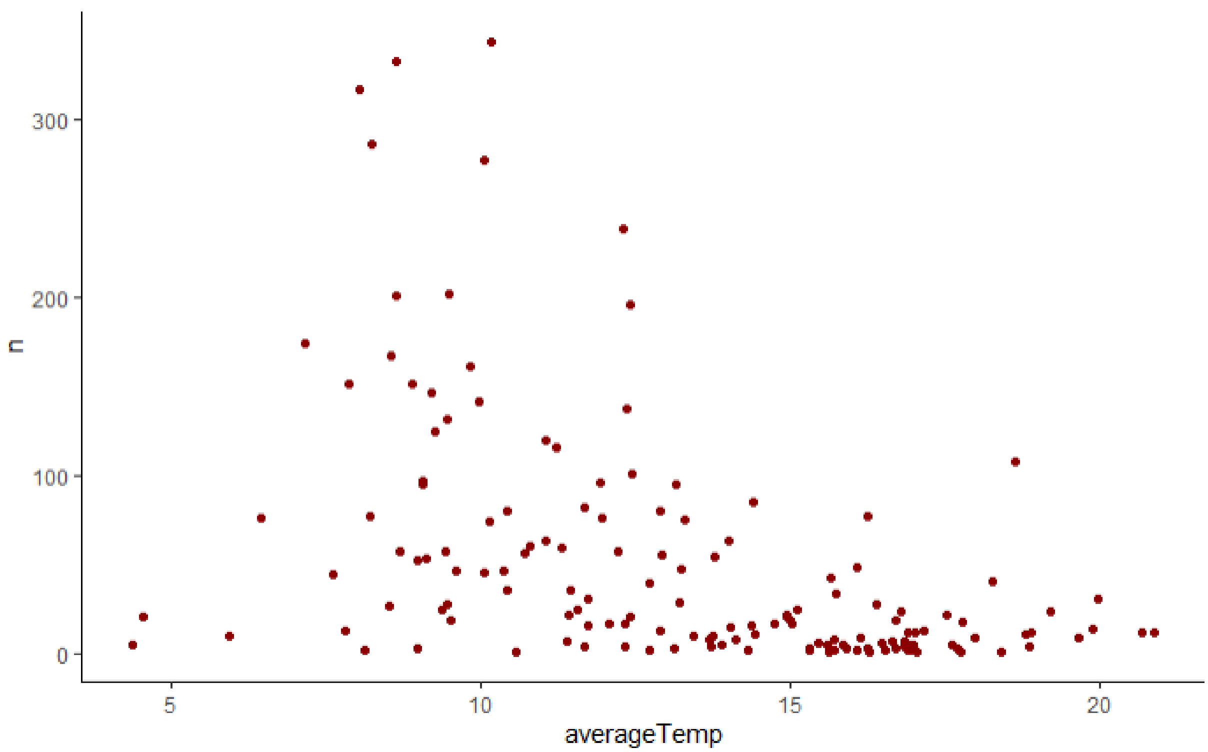
Climate

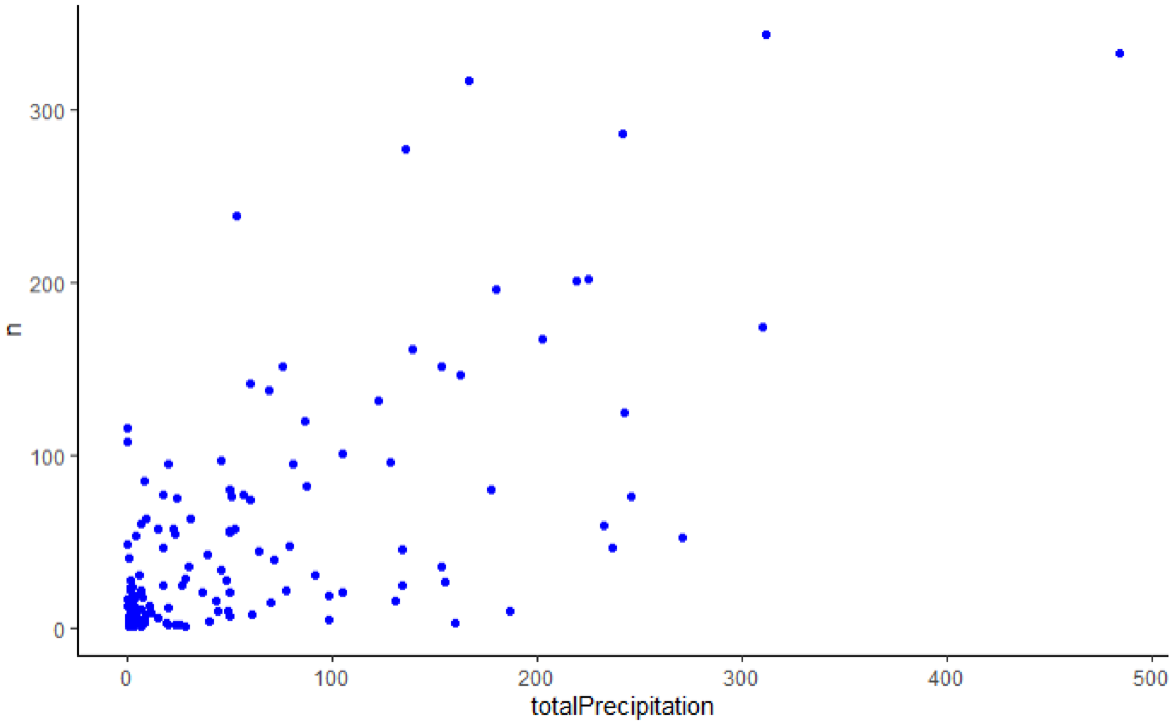
We started our look in the San Francisco Bay Area, and observed the fluctuations in mushroom observations alongside average temperature and precipitation across months and weeks:





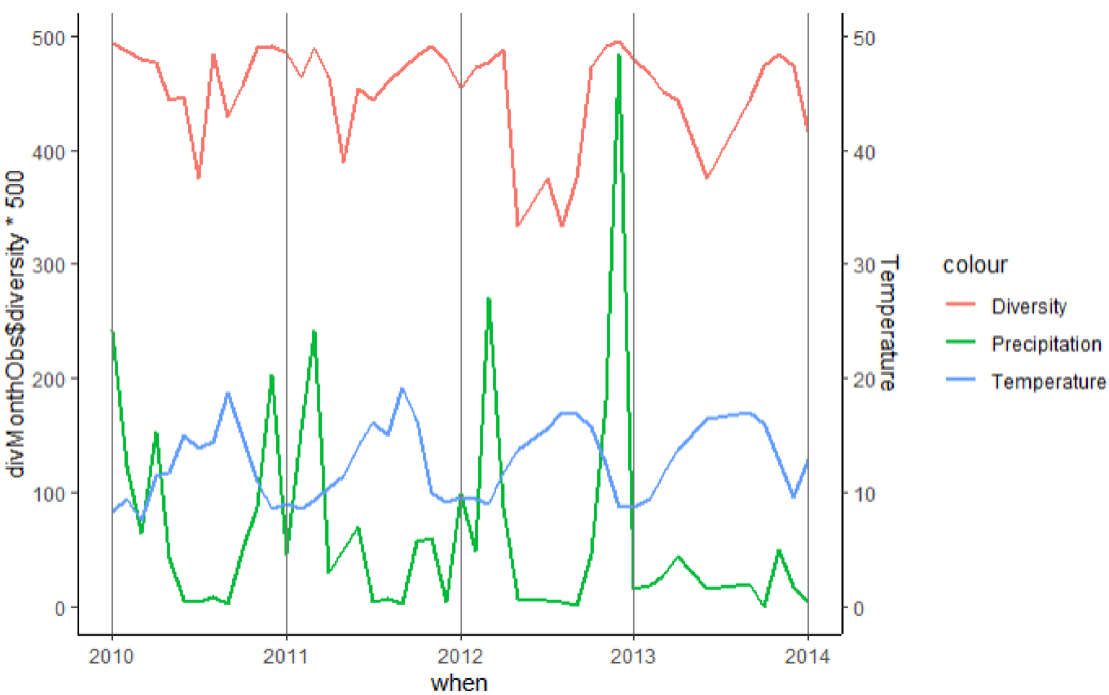
This gives a pretty good indication that observations co-vary with precipitation and temperature, so we looked at plotting this relation

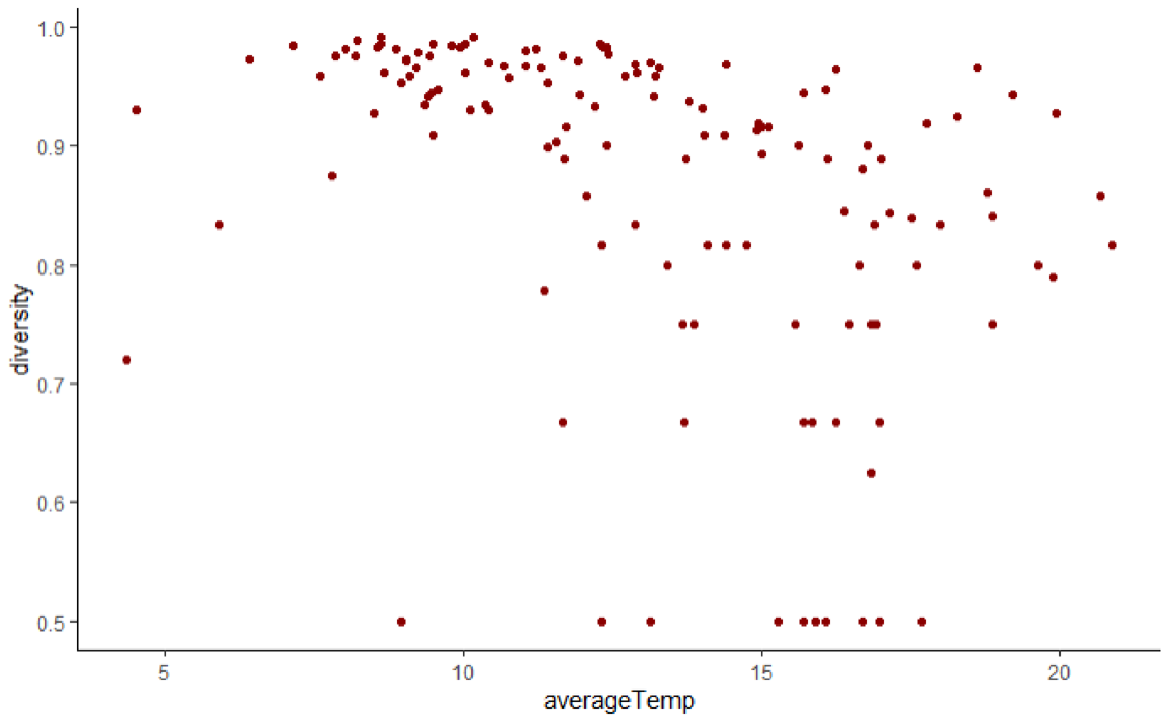


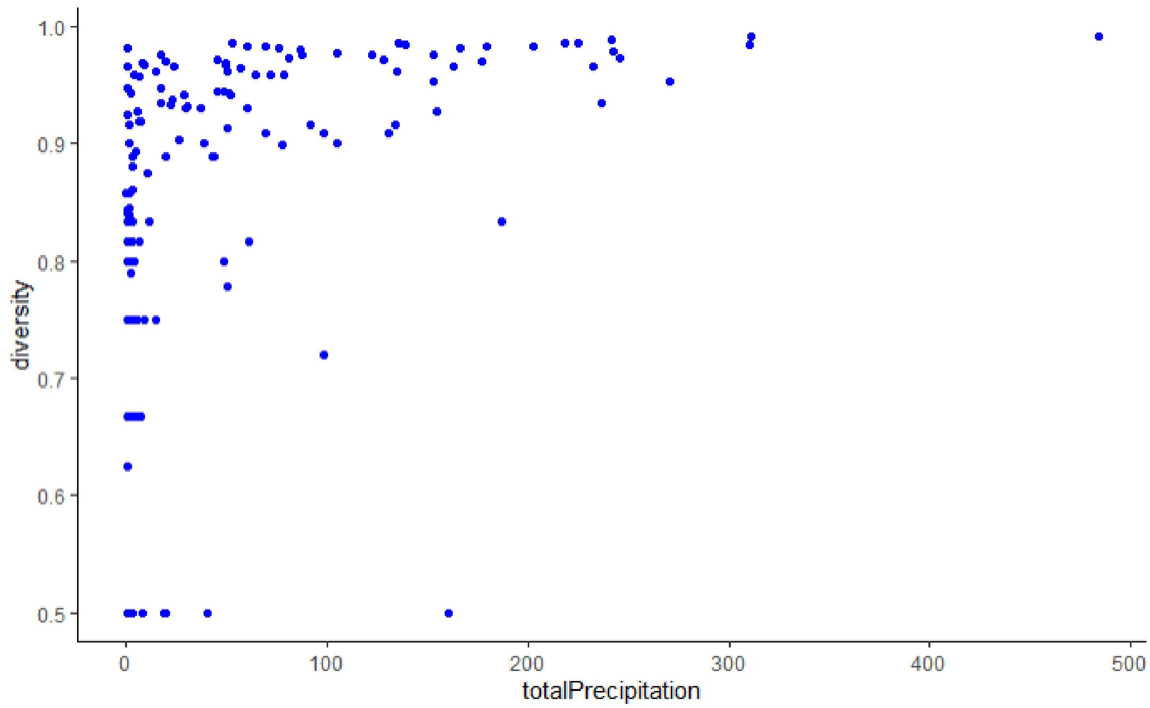


These give an even bigger impression that the larger numbers of observations are observed only for increasing precipitation levels and for a lower temperature range

For diversity, we used the simpson index for an initial exploration, giving not as clear results



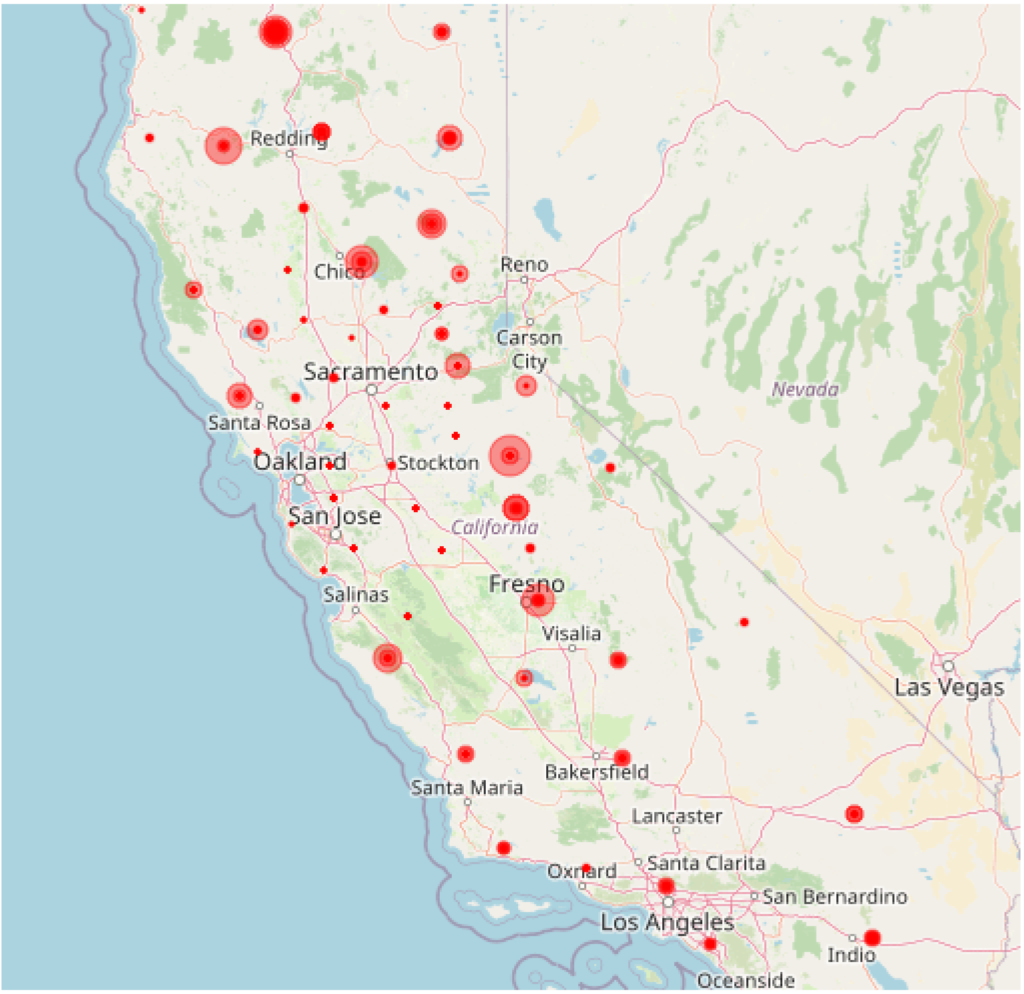




There may be less of an effect, or perhaps we will need to check with other diversity indices - this requires more exploration

Fire

To get an idea of fire distribution and size, we matched the incidents to their respective counties and mapped them out. To do this, we used an additional data set (we used the xlsx) describing the county borders (With columns *NAMELSAD*, *INTPTLAT*, *INTPTLON* - from **ca-county-coords.xlsx**) and matched the fires by county, also with left.join():



This motivated a look at the larger fires in the data set, under the thought that we'd find the most impact on observations with larger fires. Through some manipulation in dplyr, we found the largest fire in the record was in Tuolumne. With a location, we sifted through the mushroom observer data to graph data from the region 2 years before and after the time of the fire - a good number of observations are seen in Yosemite park:

We can see no major reports of mushrooms in a period after the fire. For further analysis, we need to match all possible fires to their respective mushroom observations in location in time and see if a similar result is seen where mushroom abundance appears to decline after a major fire. An automation of this process is due.

Further, a look at diversity will also be needed - looking at how it changes before/after fire and maybe discovering what species might favour fire for growth

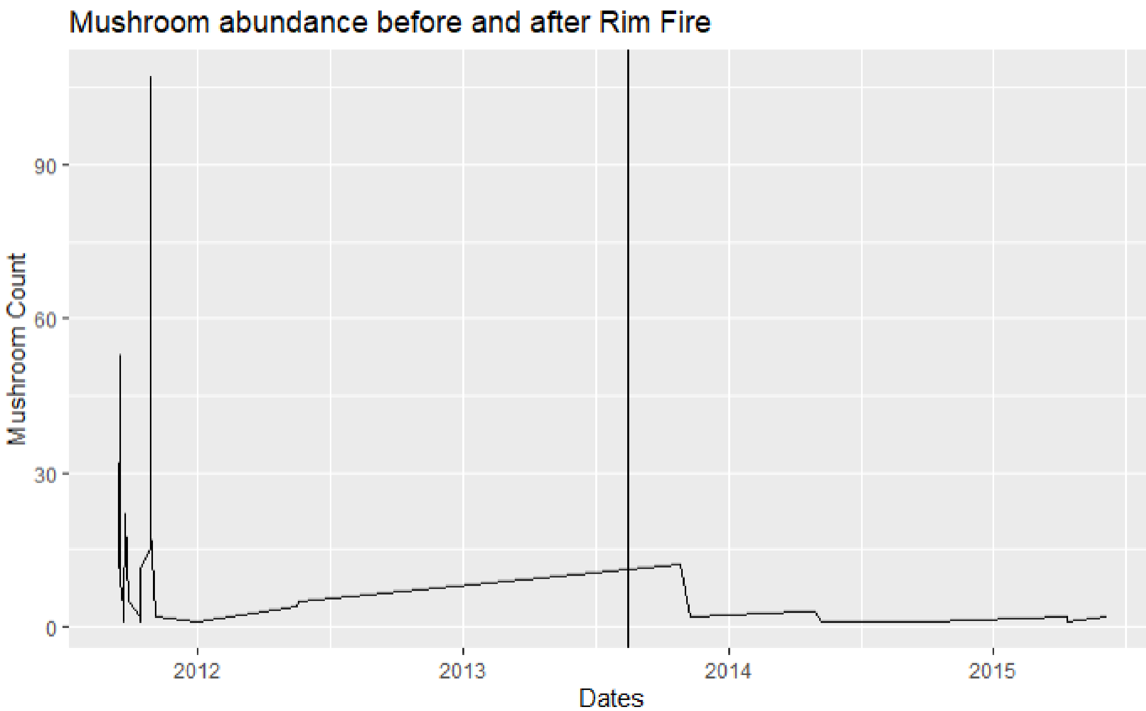


Figure 1: FirePlot

Statistical Analysis

Climate

For climate we plan on forming a linear model that incorporates the effects of temperature and precipitation on mushroom observation abundance and diversity. This model will include the interaction between the two as well as the separate effects to test to see which are significant. The preliminary look at abundance seems to have promise, but the look at diversity will need more consideration, possibly with a different diversity index or with different levels of taxonomic rank. Looking at other areas may help. It is also possible that our observations are too few or too biased to get a reasonable estimation of diversity - we may see if a power analysis would be possible to see if more observations would be necessary for data like this. - We have yet to consider how to work through assumptions and confounders (Like how our data is crowd-sourced), if possible

Wildfires

We plan on conducting a test of a difference of means (likely t-test) on mushroom diversity/abundance for each fire location before and after the fire. The exact time frame(s) in which this will happen will need to be further looked into. This will test for the alternative hypothesis that fires do have an impact on mushroom abundance and diversity. The values tested will be a mean of before fire observations vs. after fire observations from all of the fires in the data set(s) that we use to see if there is a significance. - If using a t-test we will have to see if the assumptions are met (normality, even variance), which we will have to further consider

Contributions

Brandon - Fire data analysis and writing
Seoyeong - Fire data analysis, writing, and research
Kevin - Climate data analysis and mushroom observer data manipulation
Eli - Climate effect research and writing

*These roles were for this portion of the project - work subject to change