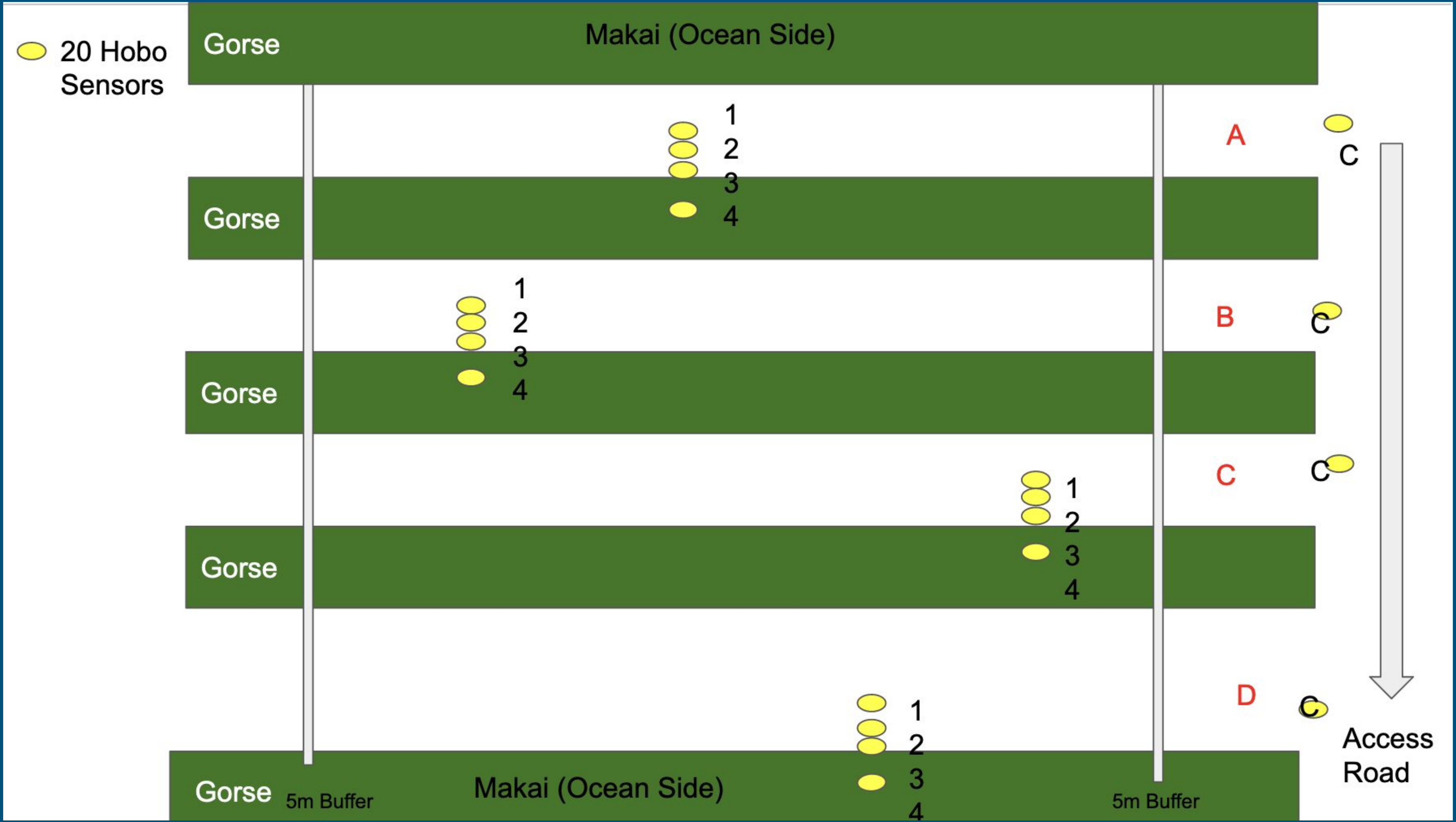
Using *Ulex europaeus* as a thermal buffer to grow *Acacia koa* on the slopes of Mauna Kea

Aaron Wehrman, Mack Jahnke, Pandu Wirabuana, Amy Knowles

Mauna Kea, Hawaiʻi

Field Site


20 Hobo Sensors
Gorse
Makai (Ocean Side)
1
2
3
4
A
C
Gorse
1
2
3
4
B
C
Gorse
C
C
1
2
3
4
Gorse
D
C
1
2
3
4
Access Road
Gorse
5m Buffer
Makai (Ocean Side)
5m Buffer

The Data Set

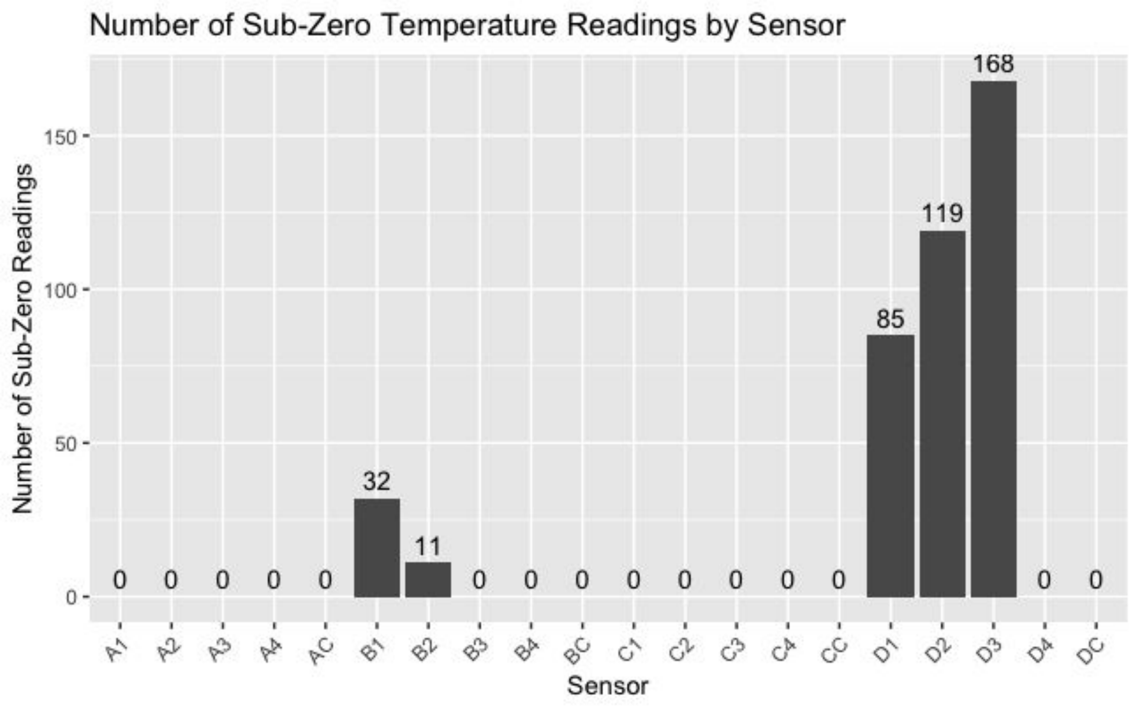
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Date-Time (HST)	AC Tempe	A1 Temperature (°C)	A2 Tempe	A3 Tempe	A4 Temperature (°C)	BC Tempe	B1 Tempe	B2 Tempe	B3 Tempe	B4 Tempe	CC Tempe	C1 Tempe	C2 Tempe	C3 Tempe	C4 Tempe	DC Tempe	D1 Tempe	D2 Tempe
2	01/11/2023 10:24:37	21.49	24.06	22.99	24.36	18.02	19.82	25.82	24.24	23.76	23.85	21.83	25.69	25.69	25.82	18.27	18.96	25.69	25.52
3	01/11/2023 10:25:37	21.79	24.06	22.95	24.28	18.19	19.99	25.82	24.24	23.85	23.98	21.49	25.57	25.65	25.52	18.32	18.74	25.69	25.44
4	01/11/2023 10:26:37	21.79	23.81	22.95	24.45	18.32	20.16	25.82	24.11	24.02	24.11	21.19	25.35	25.39	25.35	18.27	18.70	25.57	25.31
5	01/11/2023 10:27:37	21.96	23.72	23.34	24.54	18.49	20.50	25.61	24.11	24.19	24.41	21.45	25.14	25.09	25.27	18.57	18.62	25.57	25.69
6	01/11/2023 10:28:37	21.96	23.42	23.85	24.84	18.79	20.46	25.78	24.45	24.58	24.92	21.79	25.05	25.01	25.05	18.96	18.87	25.27	25.99
7	01/11/2023 10:29:37	21.66	23.29	24.02	25.01	19.17	20.59	25.82	24.84	24.75	25.09	22.01	25.01	24.49	25.48	19.00	18.87	25.69	26.34
8	01/11/2023 10:30:37	21.62	23.21	24.11	25.14	19.26	20.76	25.95	25.05	24.66	25.09	22.13	24.97	24.24	25.65	19.09	18.96	25.27	25.52
9	01/11/2023 10:31:37	21.62	23.25	24.32	25.44	19.47	20.98	25.95	24.88	24.54	25.35	22.13	24.92	24.62	26.04	19.05	19.05	25.27	25.52
10	01/11/2023 10:32:37	21.66	23.34	24.58	25.52	19.47	21.15	25.78	24.92	24.66	25.31	22.22	25.01	24.75	26.08	19.00	19.00	25.35	25.74
11	01/11/2023 10:33:37	21.62	23.42	24.84	25.61	19.6	21.10	25.78	24.88	24.32	25.35	22.09	24.88	24.66	26.04	19.09	19.09	25.35	25.22
12	01/11/2023 10:34:37	21.62	23.51	24.88	25.65	19.56	21.40	25.39	24.62	24.19	25.35	22.09	24.92	24.66	25.99	19.09	18.96	25.18	25.05
13	01/11/2023 10:35:37	21.40	23.59	25.09	25.78	19.65	21.19	25.39	24.79	24.36	25.31	22.18	25.01	24.75	25.95	19.09	18.74	25.22	25.09
14	01/11/2023 10:36:37	21.19	23.81	25.22	25.69	19.69	20.98	25.69	25.05	24.45	25.22	22.18	25.18	24.71	25.91	19.09	18.70	25.27	25.14
15	01/11/2023 10:37:37	20.85	23.89	25.18	25.65	19.73	20.93	26.04	25.22	24.54	25.14	22.05	25.09	24.75	25.74	19.09	18.96	25.48	25.31
16	01/11/2023 10:38:37	20.80	23.98	25.18	25.65	19.69	20.85	26.30	25.48	24.54	25.18	22.26	24.97	24.58	25.39	19.13	19.17	25.78	25.44
17	01/11/2023 10:39:37	20.85	24.02	25.35	25.48	19.82	20.93	26.42	25.35	24.28	25.14	22.22	24.79	24.41	25.48	19.13	19.30	25.95	25.52
18	01/11/2023 10:40:37	20.93	23.94	25.14	25.35	19.86	20.80	26.25	25.14	24.36	25.22	22.09	24.71	24.58	25.61	19.09	19.35	26.08	25.61
19	01/11/2023 10:41:37	21.06	23.89	24.92	25.22	19.99	20.89	26.34	25.31	24.54	25.35	22.13	24.84	24.75	25.65	19.05	19.30	26.12	25.52

Hypothesis #1

The sensors located within the gorse rows will be warmer than the controls as gorse will provide insulation from the cold.

Therefore, there will be more temperature readings below zero in the controls than the other sensors.

Results



Functions Used

- [data < 0]
- Colsums
- Data frame
  - X = Sensor
  - Y = Count
- Ggplot2
  - Geom_bar

Hypothesis #2

The nighttime (12am-5am) average temperatures of Sensors 1-4, located within or between gorse rows, will be higher than those of the control positions.

Hypothesis #3

The nighttime (12am-5am) average temperature of Position 4, located within the gorse shrub, will be warmer than the average temperatures of any other positions- 1, 2, 3, and controls, due to the insulation from the gorse shrubs.

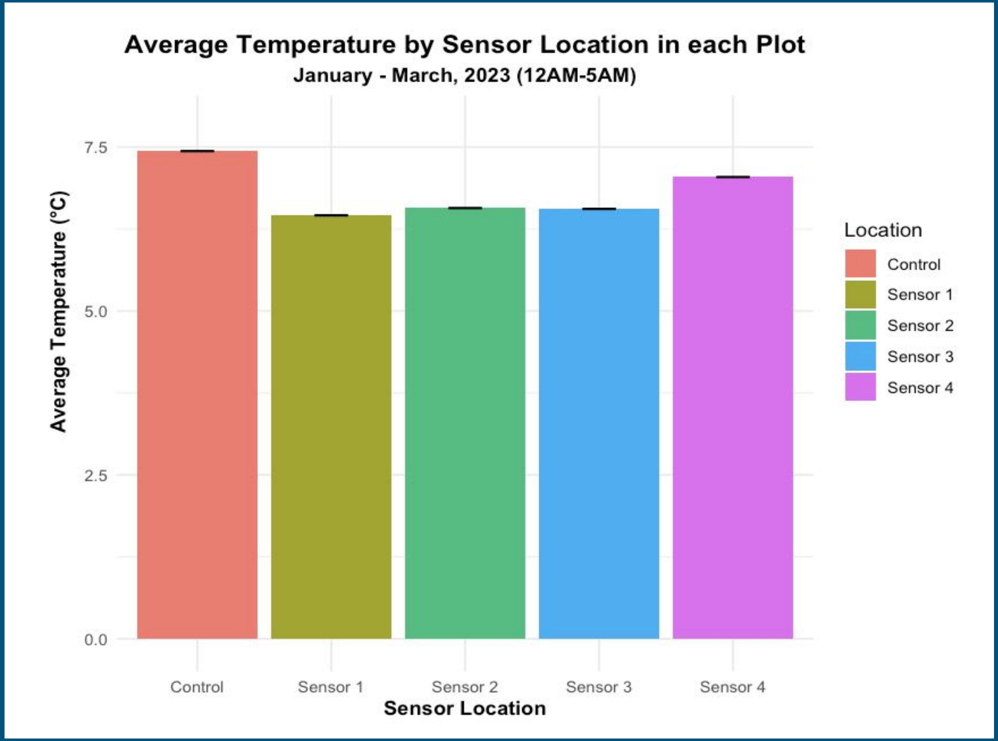
Data Wrangling and General Coding methods for H1 & H2

- Abbreviated Column Names
- Separated Date and Time columns
- Combined columns based on positional replication
  - Ie. Position 1= A1, B1, C1, D1
- Filtered data to only include 12am-5am

- Wrote code to calc means of positions
- Ran ANOVA and Tukey tests to find if temp averages at sensor positions were significantly different
- Made tables of means and statistical results

✘ Hypothesis 2

- Positions 1-4 **colder** than controls



✘ Hypothesis 3

- **Control** positions had warmest temperatures
- Position 4 had second warmest night temperatures

Position	Avg_Daily_Temp_Celsius	Avg_Nighttime_Temp_Celsius
P1	11.79454	6.458640
P2	11.94855	6.569659
P3	11.90748	6.557116
P4	11.03361	7.042890
PC	11.79687	7.438489

Anova Test

term	df	sumsq	meansq	statistic	p.value
ind	4	16599261	4.149815e+06	629715.2	0
Residuals	119531515	787711190	6.589988e+00	NA	NA

There is a significant difference in temperature between the test positions

Tukey Test

term	contrast	null.value	estimate	conf.low	conf.high	adj.p.value
ind	P2-P1	0	0.11101913	0.10899373	0.11304452	0
ind	P3-P1	0	0.09847629	0.09645090	0.10050169	0
ind	P4-P1	0	0.58424944	0.58222405	0.58627484	0
ind	PC-P1	0	0.97984848	0.97782308	0.98187387	0
ind	P3-P2	0	-0.01254283	-0.01456823	-0.01051744	0
ind	P4-P2	0	0.47323032	0.47120492	0.47525571	0
ind	PC-P2	0	0.86882935	0.86680396	0.87085475	0
ind	P4-P3	0	0.48577315	0.48374776	0.48779855	0
ind	PC-P3	0	0.88137218	0.87934679	0.88339758	0
ind	PC-P4	0	0.39559903	0.39357364	0.39762443	0

All positions experienced temperatures significantly different from one another

Hypothesis #4

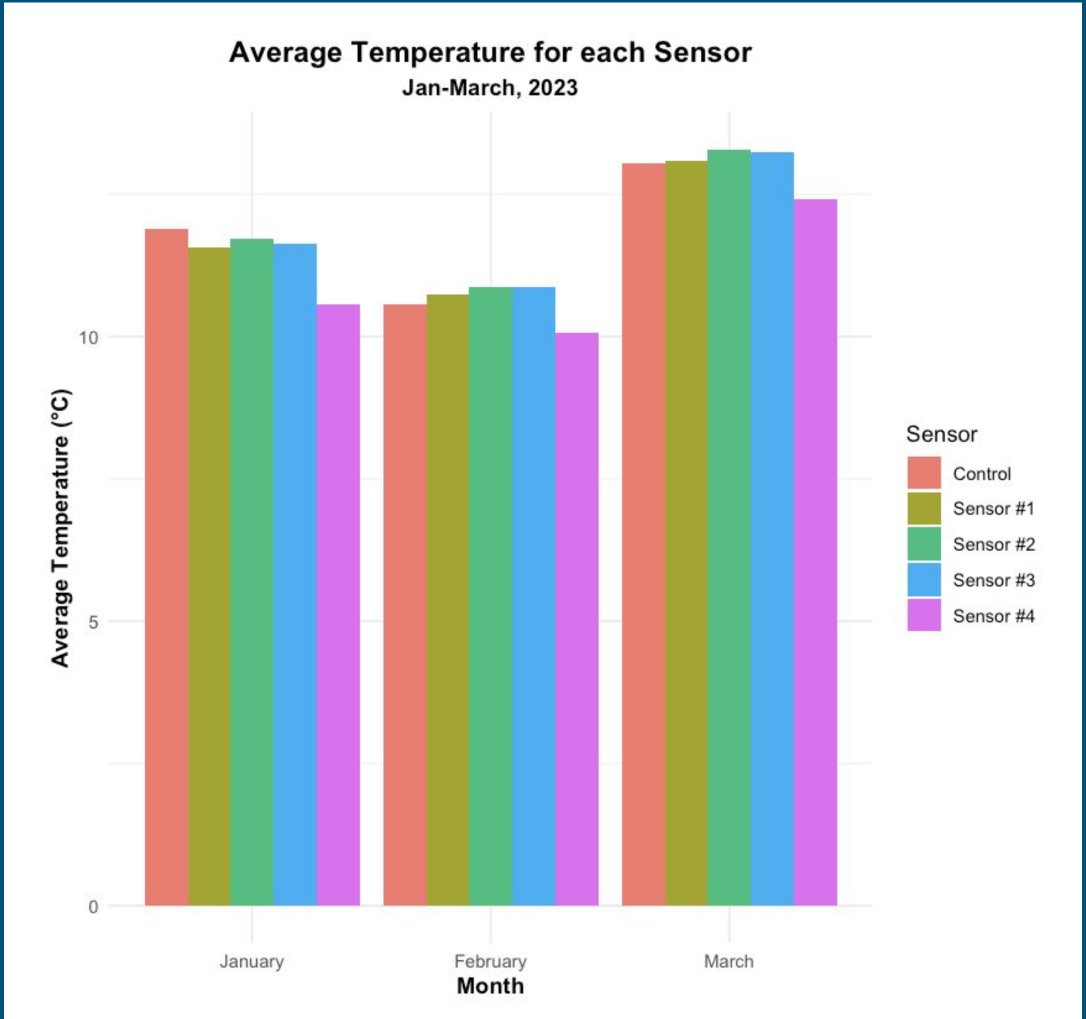
On the coldest recorded day, the average temperature of the control sensors will continue to be most similar to the average temperature of the #4 sensors, mirroring the results of the 12AM-5AM comparisons.

Hypothesis #5

The average temperature of the control sensors and the #4 sensors on the coldest day will remain higher than the average temperatures of the #1, #2, and #3 sensors on the coldest day recorded in the dataset.

Steps to Find the Lowest Temp Month:

- Filtered the data out for each month
- Filtered out the columns for each sensor position from each plot
- Got the total average temp for each of the columns
- Extracted the average for each column and convert the column averages into a matrix to allow for efficient mathematical operations
- Assigned the matrix to a new variable
- Got overall average of the new variable
- Repeated for each sensor in each month
- For the bar chart, a data frame was made by creating 2 vectors, one for the sensors and one for the months
- The data frame was modified to have a month column as a factor and the months as levels in order to allow for the correct order of variables on the x-axis
- We then created the bar chart using ggplot

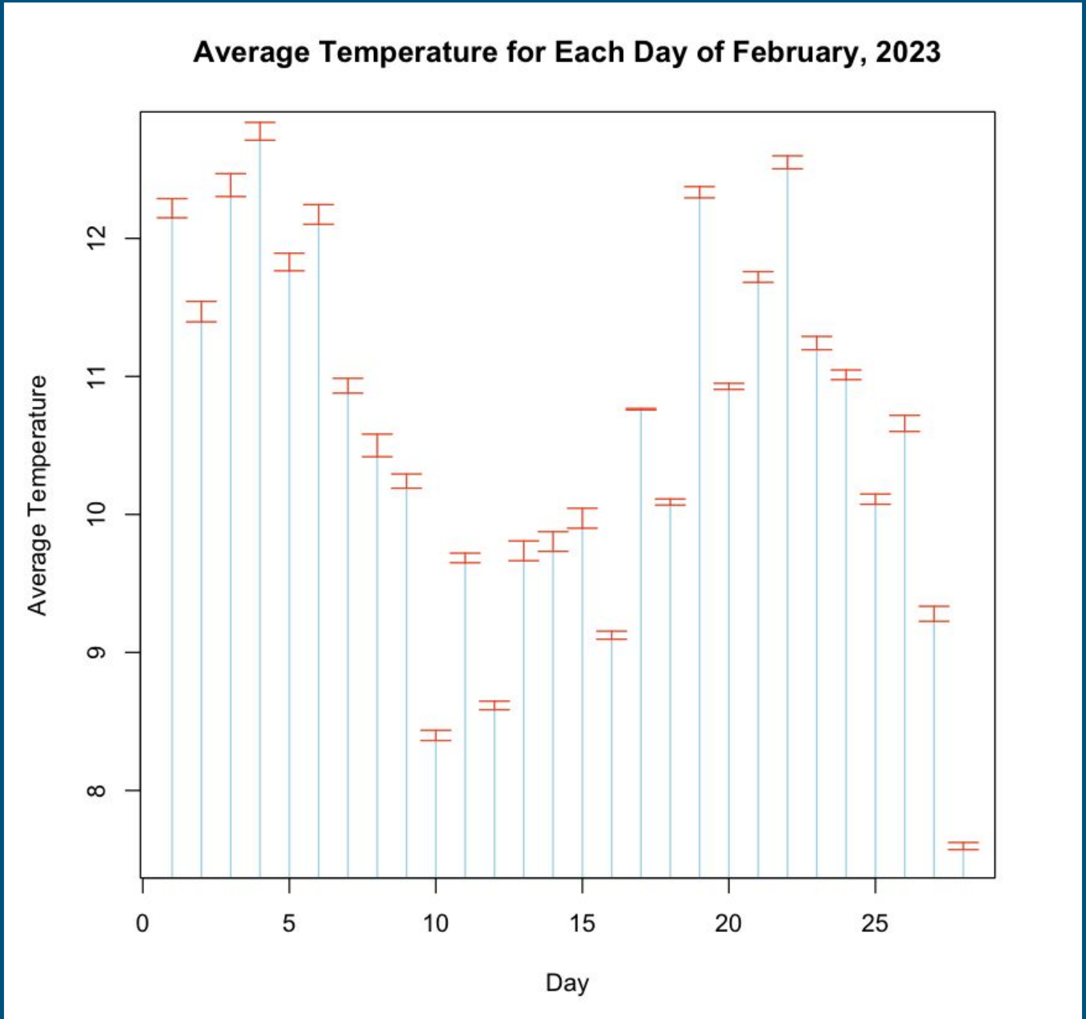


When comparing the average temperature for each sensor, per month, February was revealed to be the coldest month in our dataset.

However, the control sensor and the #4 sensor positions recorded the lowest average temperatures for the months of February and January.

Steps to Find the Lowest Temp Day of the Month:

- Changed the format of the Date_Time_HST column in the dataframe into a date-time format, POSIXct
- Filtered out all of the data from February
- Partitioned the February data by day
- Created 28 empty vectors, 1 vector to store the data from each day in February
- Created a for-loop to store the data from each day into one of the 28 empty vectors
- Used “which.min” function to find the day in February that had the lowest average temperature
- Created a bar chart using lubridate and dplyr

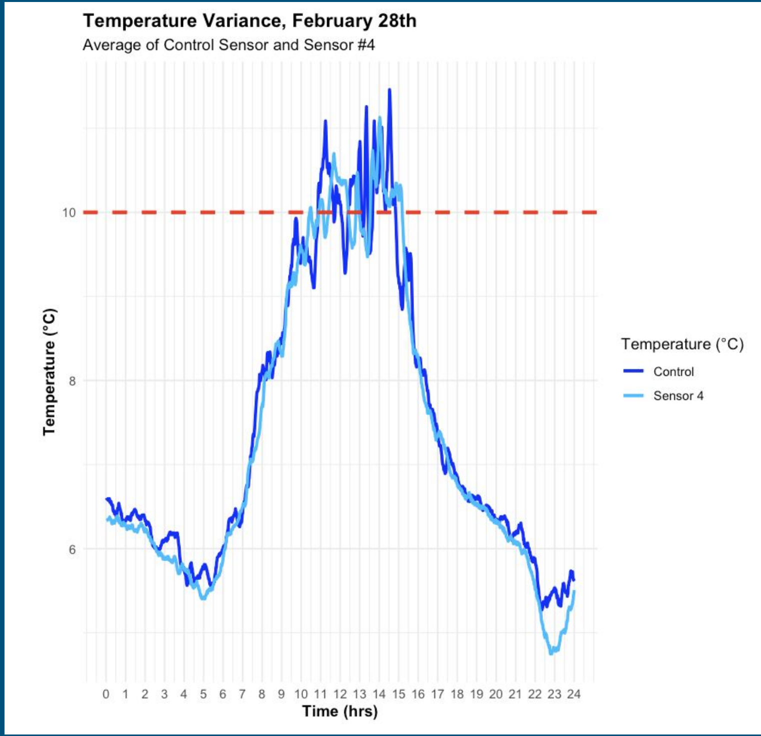


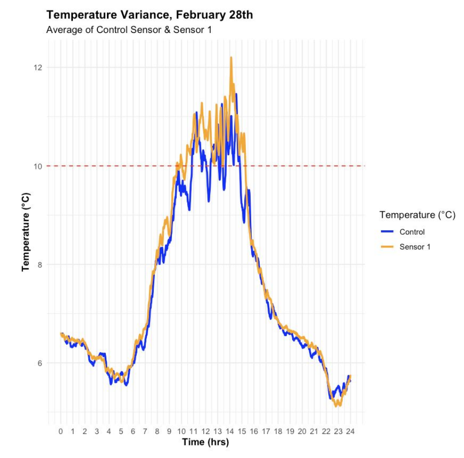
When each day of the month of February was compared, we found that Tuesday, February 28th had the lowest average temperature.

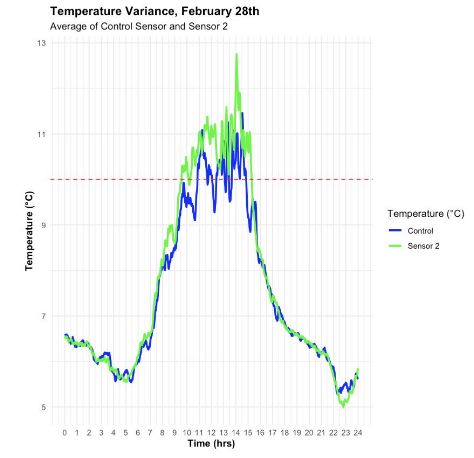
Steps to Compare the Control Sensors with the #4 Sensors and to Compare all Sensors to the Control :

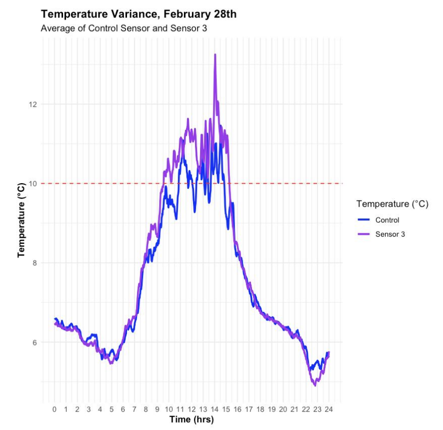
- Created a vector to specify the columns from the 28th day of February that we wanted the averages of
- Used “rowMeans” function to calculate the row-wise averages in the specified columns. This was done for each sensor on day #28 of february
- Using ggplot we showed the average temp for the control sensors throughout the 24 hours of the 28th day of February and made multiple graphs to compare the average temperature line of the control to the average temperature lines of sensors #1, #2, #3, & #4.

Comparing the control sensors to sensors #1, #2, #3, & #4







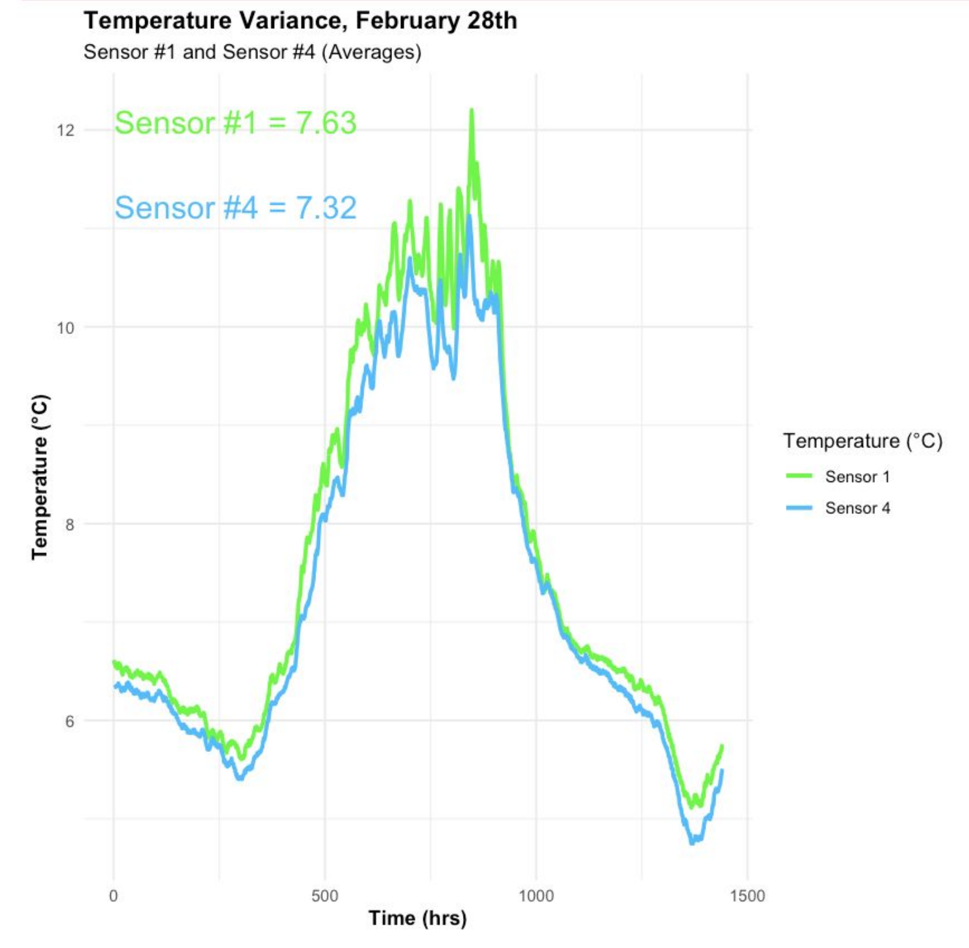


✔/✘ The control and #4 sensors have the most similar temperature lines. However, the average temperature of these sensors is less than that of sensors #1, #2, #3

Hypothesis #6

The average temperature of the #1 sensors will be lower than the average of the #4 sensors, as the #1 sensors are located outside of the canopy, while the #4 sensors are located beneath the canopy.

Steps for Coding:
Used same steps as graphs in hypothesis #5, but compared the average of sensor #1 to the average sensor #4



✘ The hypothesis is rejected as on the coldest day, the average temperature for the #1 sensors is higher than the average temperature of the #4 sensors. Therefore, the canopy cover is ineffective at minimizing the risk of frost for young plants.

```
T-Test Result:

        Welch Two Sample t-test

data:  row_averages_S1 and row_averages_S4
t = 4.5399, df = 2860.4, p-value = 5.861e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1777029 0.4479047
sample estimates:
mean of x mean of y
 7.634691  7.321887
```

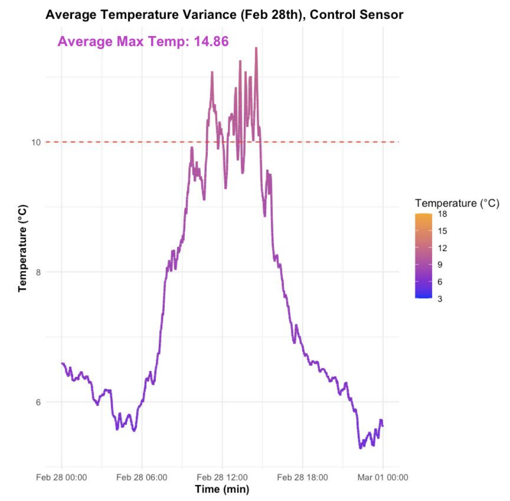
The mean of sensor #1 is significantly higher than the mean of sensor #4

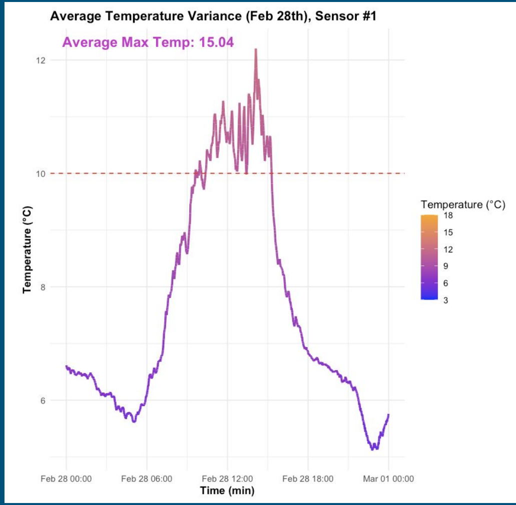
Hypothesis #7

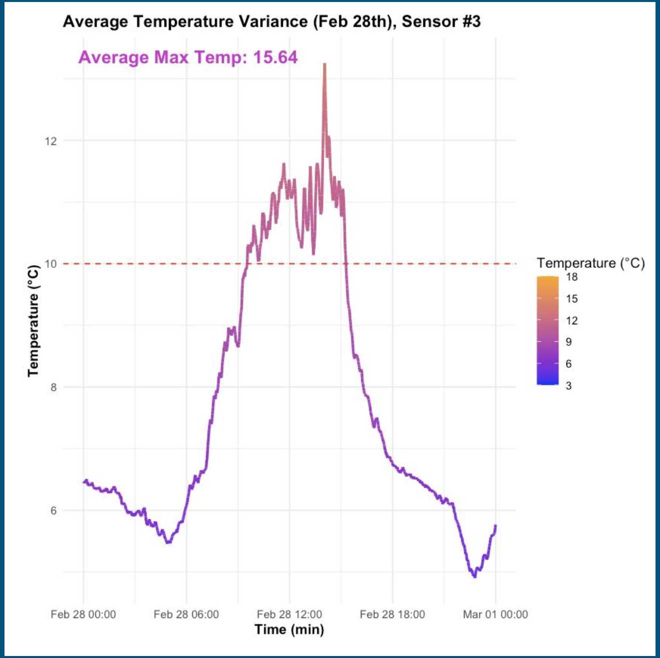
The #1 and #2 sensors will have the highest average maximum temperature as these sensors are in the middle of the gorse rows where they are insulated and they are not covered by canopy.

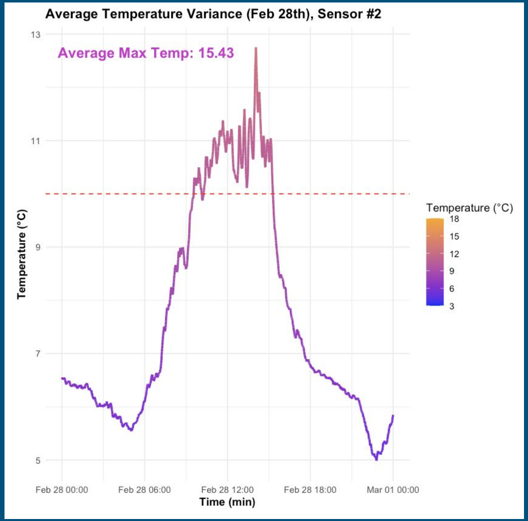
Steps to find the average maximum temperature for each sensor:

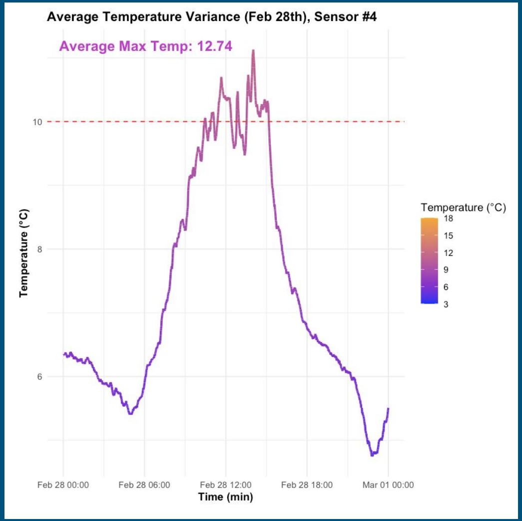
- Made a vector that contained all of the columns for each individual sensor in each plot
- Used the function “sapply” to convert the values in the previously made vector into numeric values and to find the max temp in each column
- Used a for loop to scan each temperature column for the highest temperature
- Used “mean” function to calculate the average of these highest temperatures
- Converted the values in the day_28_data dataframe temperature columns into a numeric format
- Changed the format of the Date_Time_HST column in the dataframe into a date-time format, POSIXct
- Calculated the average of each row for the specified columns for each individual sensor using “rowMeans” function
- Used ggplot to show average temp for the sensor for each hour on day #28, and used the average max temp calculation to display the value at the top of the graph











✘ Our hypothesis is rejected, the #3 sensors have the highest maximum average temp. Followed by the #2 sensors and the #1 sensors

Anova Test

```
                Df Sum Sq Mean Sq F value Pr(>F)  
Sensor_Location  4  20.24   5.061   3.884 0.0233 *
Residuals       15  19.55   1.303                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The null hypothesis is rejected and it can be concluded that there is a significant difference in mean temperatures among the different sensor locations

Tukey Test

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Temperature ~ Sensor_Location, data = tukey_data)

$Sensor_Location
                    diff       lwr        upr     p adj
Sensor 1-Control   0.1800 -2.312524  2.6725245 0.9993834
Sensor 2-Control   0.2550 -2.237524  2.7475245 0.9975879
Sensor 3-Control   0.7825 -1.710024  3.2750245 0.8646191
Sensor 4-Control  -2.1250 -4.617524  0.3675245 0.1138057
Sensor 2-Sensor 1  0.0750 -2.417524  2.5675245 0.9999810
Sensor 3-Sensor 1  0.6025 -1.890024  3.0950245 0.9418294
Sensor 4-Sensor 1 -2.3050 -4.797524  0.1875245 0.0766106
Sensor 3-Sensor 2  0.5275 -1.965024  3.0200245 0.9633038
Sensor 4-Sensor 2 -2.3800 -4.872524  0.1125245 0.0646896
Sensor 4-Sensor 3 -2.9075 -5.400024 -0.4149755 0.0188174
```

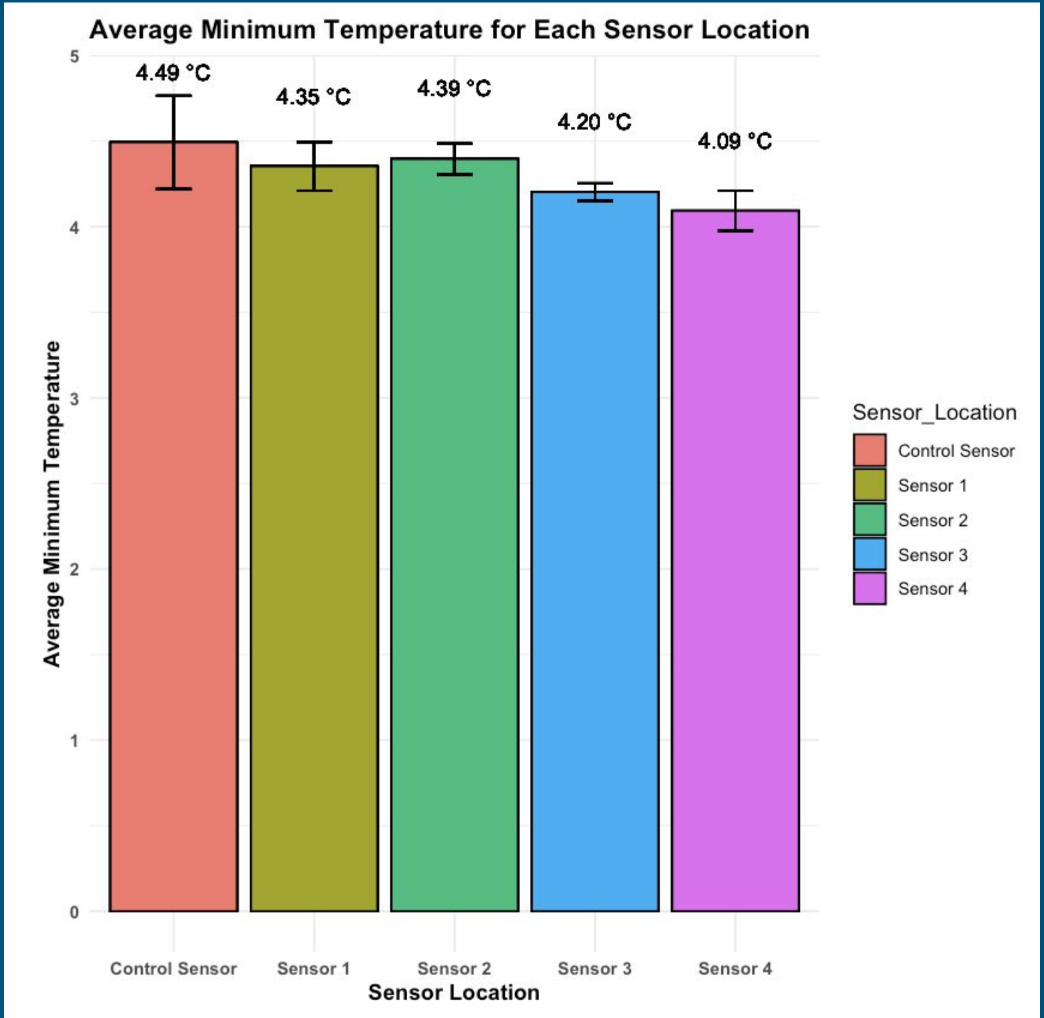
- There is a significant difference between Sensor 4 and Sensor #3. No other sensors were found to be significantly different from each other

Hypothesis #8

The #4 sensors will have the lowest average minimum temperatures as they are covered by the canopy and therefore do not have as much access to sunlight.

Steps to find the average minimum temperature for each sensor:

- Made a vector that contained all of the columns for each individual sensor in each plot
- Used the function “sapply” to convert the values in the previously made vector into numeric values and to find the min temp in each column
- Used a for loop to scan each temperature column for the lowest temperature
- Used “mean” function to calculate the average of these minimum temperatures
- We then put the average minimum temperature found for each sensor into a bar chart using the values found above to define the data frame used by the ggplot2 “geom_bar” function to make the graph.



✔ Our hypothesis is supported as sensor #4 has the lowest average minimum temperature.

Anova Test

```
                Df Sum Sq Mean Sq F value  Pr(>F)   
Sensor_Location  4  1.401  0.3502   4.369 0.00314 **
Residuals       75  6.012  0.0802                   
```

We can reject the null hypothesis. We can conclude that there is a significant difference in mean temperatures between the sensors.

Tukey Test

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Min_Temperature ~ Sensor_Location, data = anova_data)

$Sensor_Location
                    diff       lwr         upr     p adj
Sensor_1-Control -0.0250 -0.304809  0.25480905 0.9991223
Sensor_2-Control  0.0950 -0.184809  0.37480905 0.8766184
Sensor_3-Control -0.2925 -0.572309 -0.01269095 0.0360084
Sensor_4-Control -0.1325 -0.412309  0.14730905 0.6774681
Sensor_2-Sensor_1  0.1200 -0.159809  0.39980905 0.7520518
Sensor_3-Sensor_1 -0.2675 -0.547309  0.01230905 0.0678183
Sensor_4-Sensor_1 -0.1075 -0.387309  0.17230905 0.8192961
Sensor_3-Sensor_2 -0.3875 -0.667309 -0.10769095 0.0020841
Sensor_4-Sensor_2 -0.2275 -0.507309  0.05230905 0.1653135
Sensor_4-Sensor_3  0.1600 -0.119809  0.43980905 0.5031281
```

- Sensor position #3 shows a significant difference compared to the Control group
- Sensor position #3 shows a significant difference compared to Sensor #2

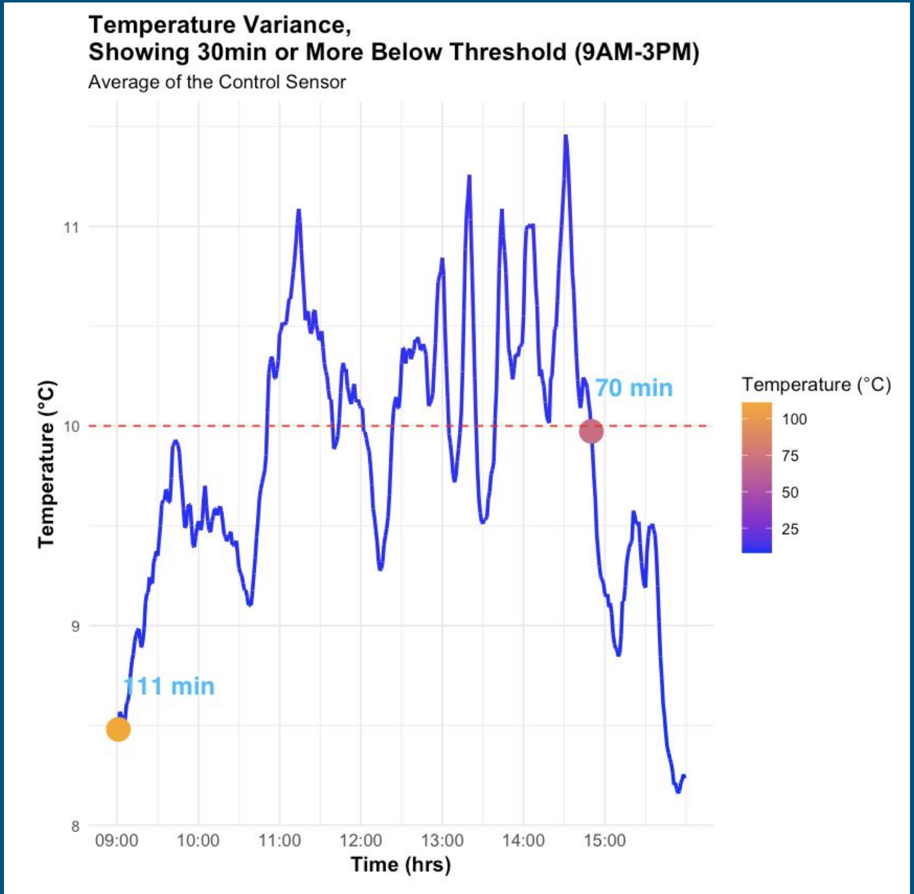
Hypothesis #9

The positions of sensors #3 and sensors # 4 are the least advantageous in terms of photosynthesis as these position are under the canopy.

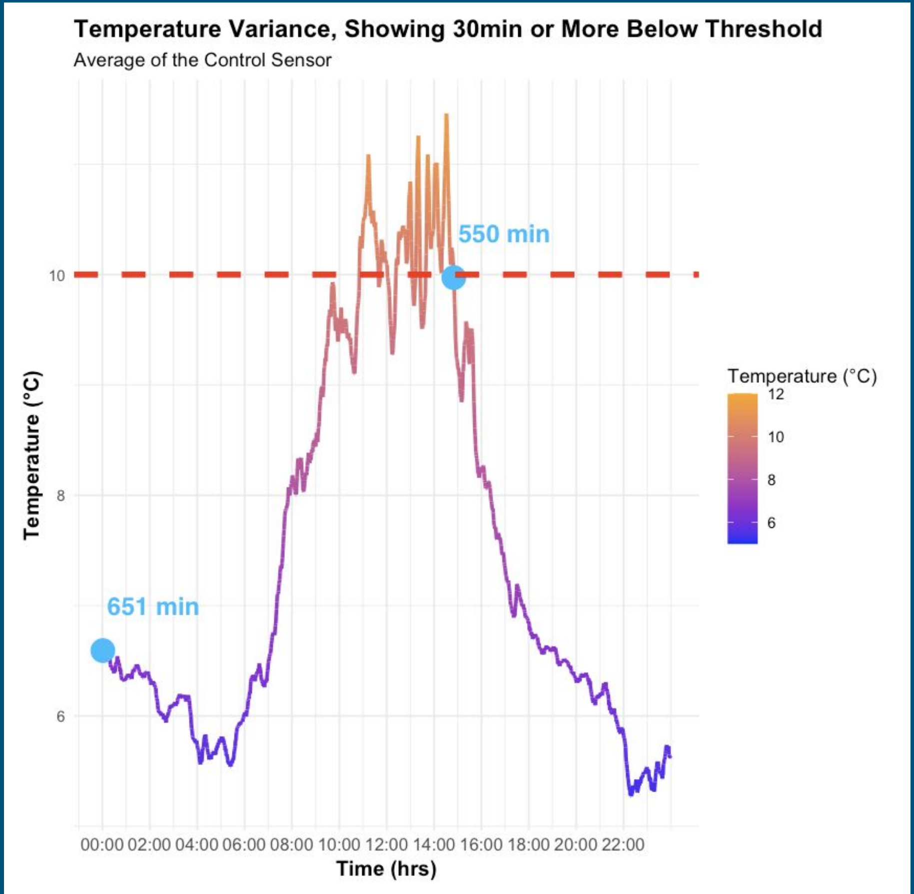
Photosynthesis slows if temperature stays below 10 degrees celsius for 30 minutes or longer. The sensor position with the most minutes under 10 degrees celsius for 30 min or more will be considered the least advantageous for photosynthesis.

Steps to find the amount of time each sensor position recorded temp 10 degrees or below for 30 minutes or more between 9AM- 3PM:

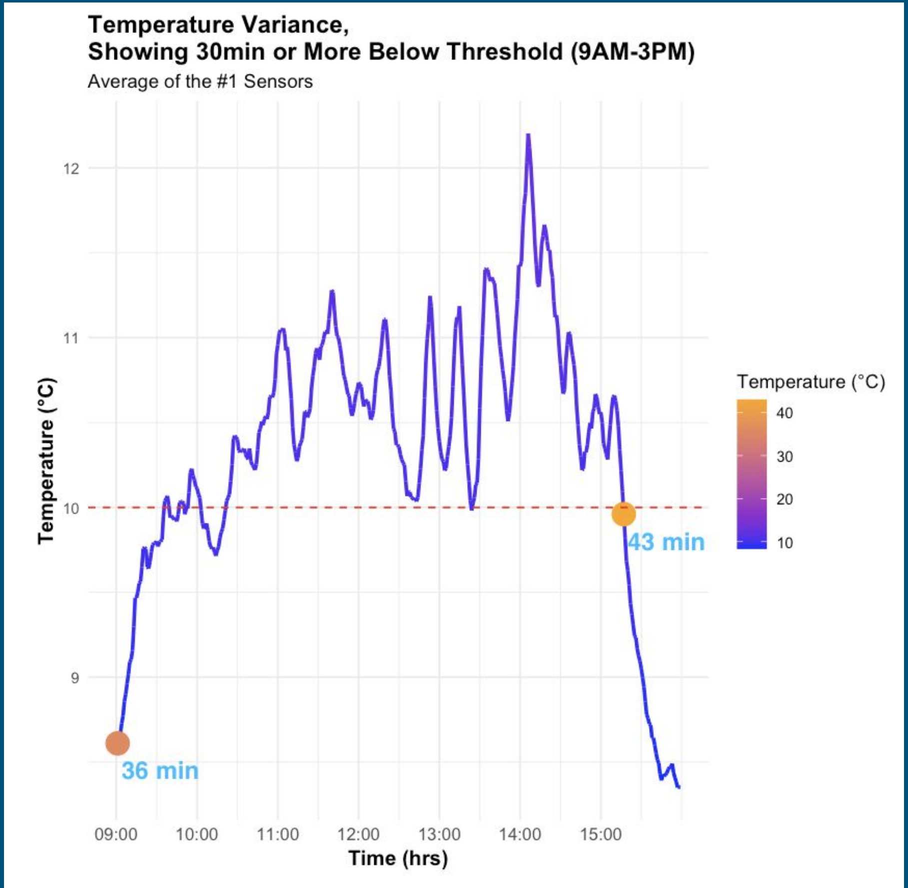
- Used the pipe operator to chain the multiple instructions together
- Used "mutate" function to change the format of the Date_Time_HST column in the dataframe into a date-time format, POSIXct
- Used filter function to extract the data recorded between 9AM-3PM
- Created a vector to store the data in the specified columns for each sensor position
- Used "rowMeans" function to find the mean of the specified columns
- Used the comparison operator "< 10" to check to see of the averages calculated above were less than 10 degrees, giving result of TRUE or FALSE
- Used "consecutive count" variable to set up a counter consecutive occurrences in the next steps
- Created an empty data frame to later populate it with data points that meet the conditions of the proceeding for loop.
- Created a for loop to keep track of every time the temp went below 10 degrees for 30 min or more, and records the starting time and the corresponding average temp
- Used ggplot to create a graph to show the information
- We omitted the filter function to view the amount of time below 10 degrees for 30 min or more over the entire 24 hours



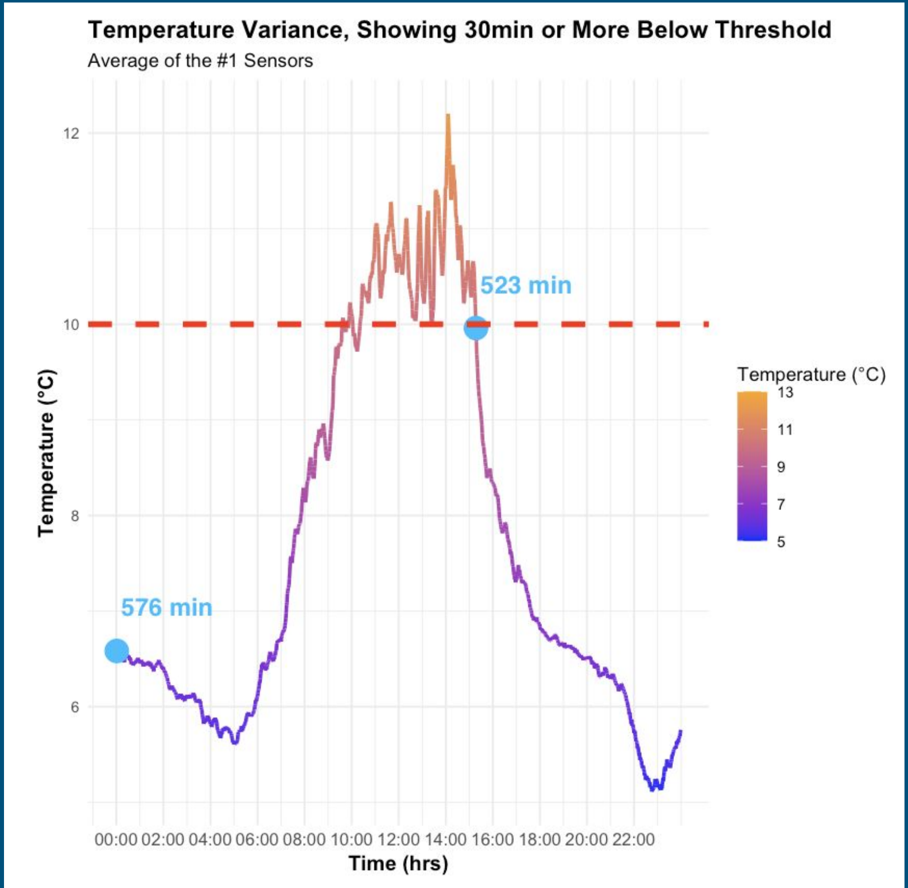
3 hours



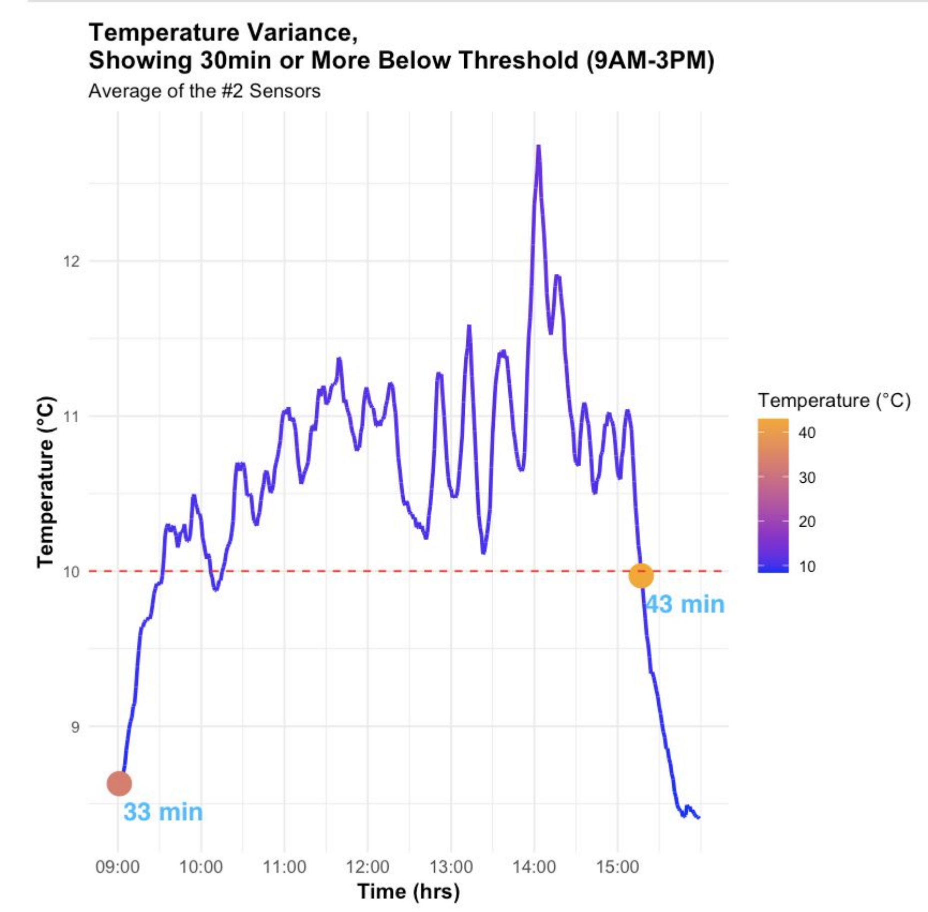
20.02 hours



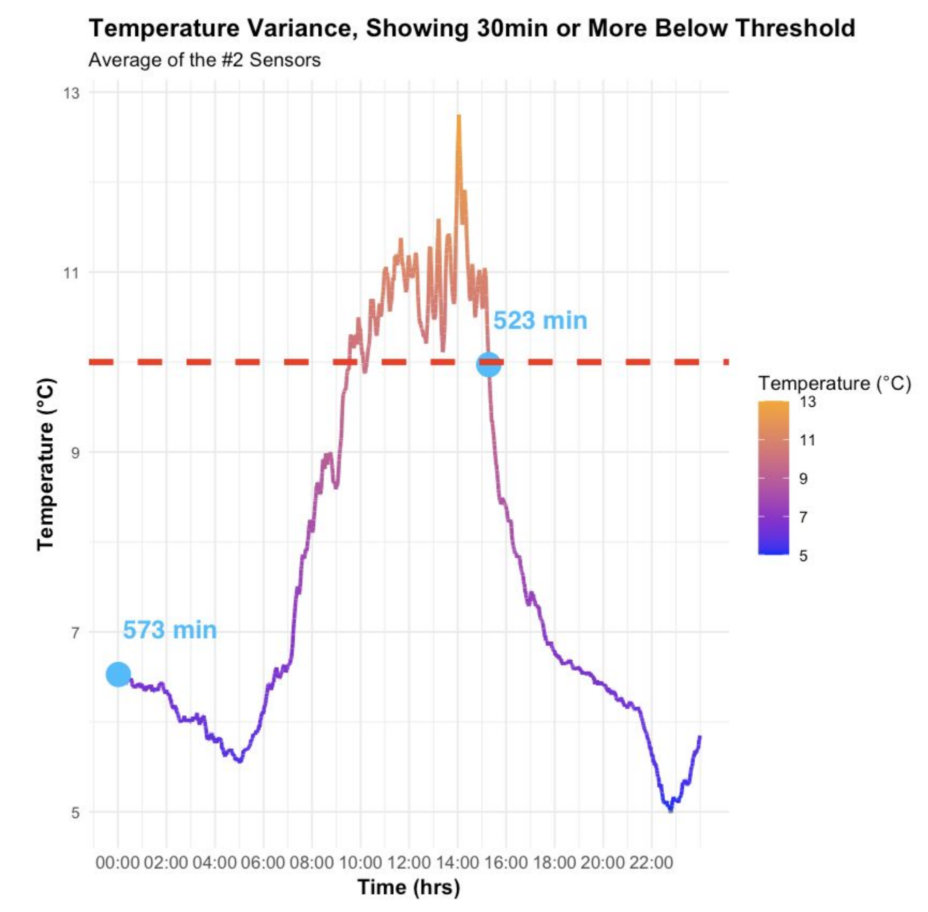
1.32 hours



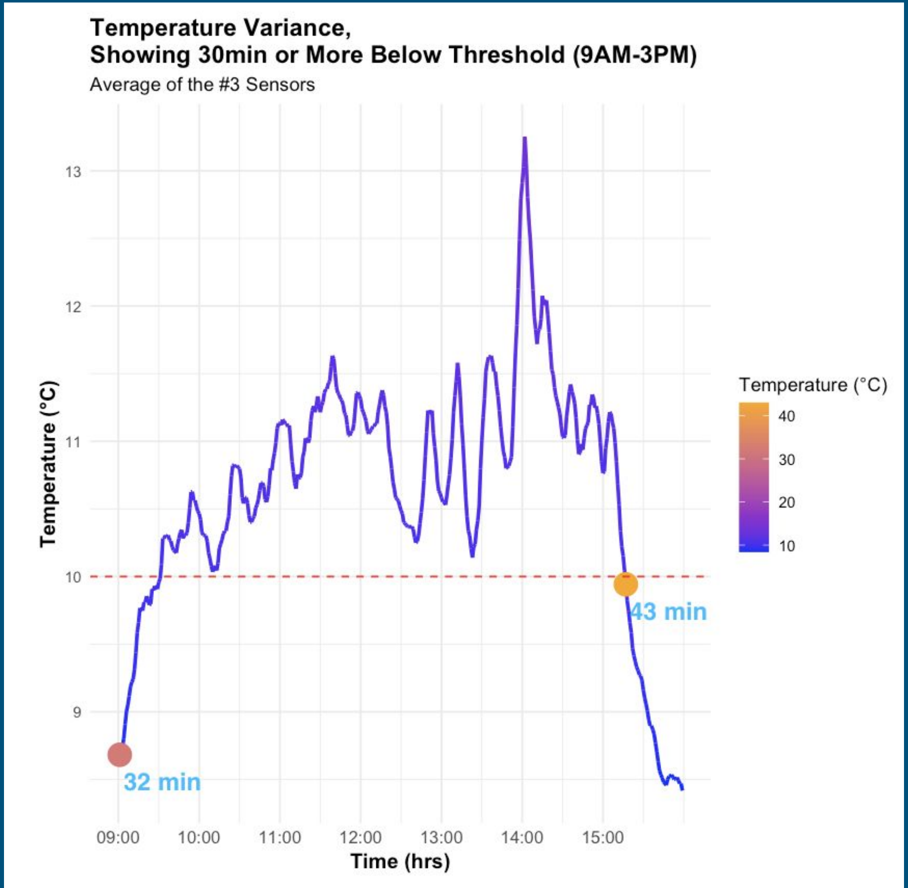
18.32 hours



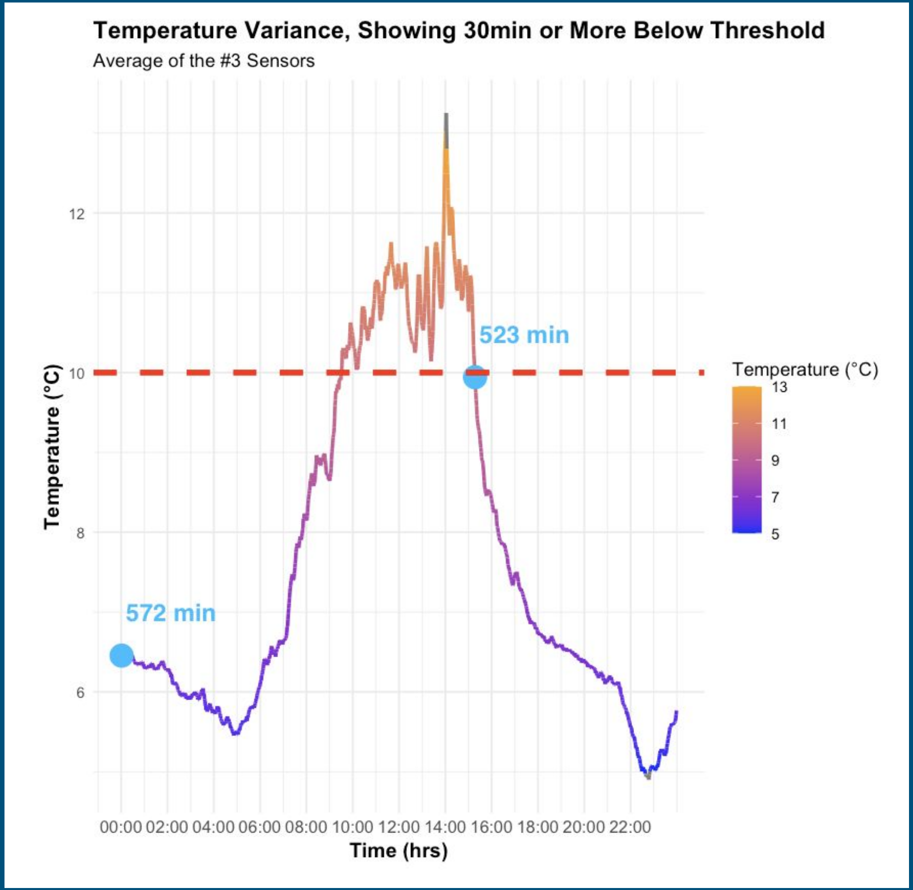
1.27 hours



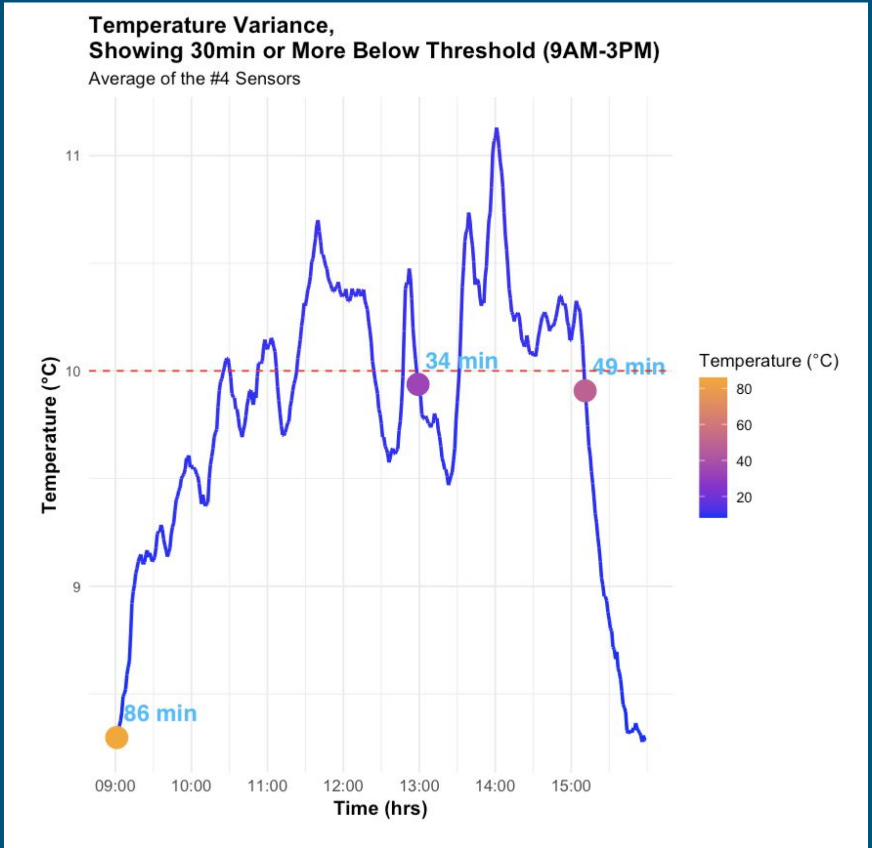
18.23 hours



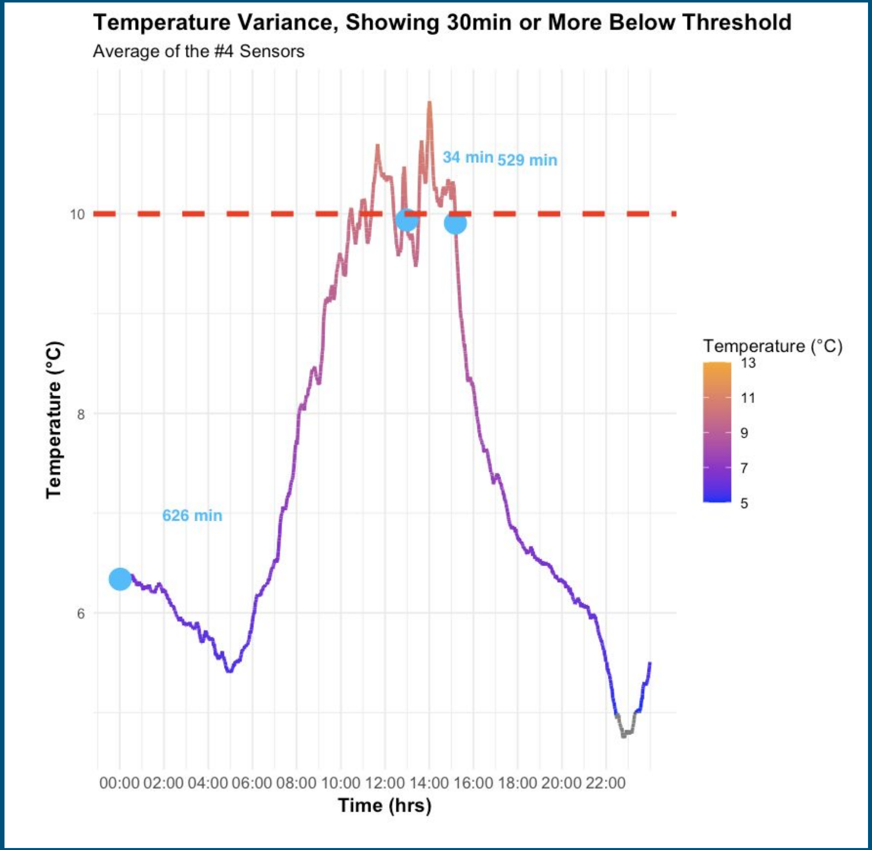
1.25 hours



18.25 hours



2.81 hours



19.82 hours

Sensor #	Hours within 9AM- 3PM	Hours within the entire day
Control	3	20.02
One	1.32	18.32
Two	1.27	18.23
Three	1.25	18.25
Four	2.81	19.82

✔/✘ The control and sensor #4 positions are the least advantageous for photosynthesis as they have the most 30+ minute intervals under ten degrees celsius.

The sensor #3 positions have the least 30+ minute intervals under ten degrees celsius, meaning this is the most advantageous position in terms of photosynthesis.

Hypothesis 10: Logistic regression can be used to predict frost incidence using logistic regression?

Method	: Binomial Logistic Regression
Y-variable	: Frost (0 = not happen, 1 = happen)
X1-variable	: Month (1 = January, 2 = February, 3 = March)
X2-variable	: Hours (continuous data 0-23)
X3-variable	: Minute (continuous data 0-59)
X4-variable	: Canopy (0 = openness, 1 = closure)

Parameter Evaluation : 1. Significant fit-parameter
2. Akaike Information Criterion (AIC)
3. Features Important
4. Variance Inflation Factor
5. Accuracy
6. Sensitivity
7. Specificity

Working Step

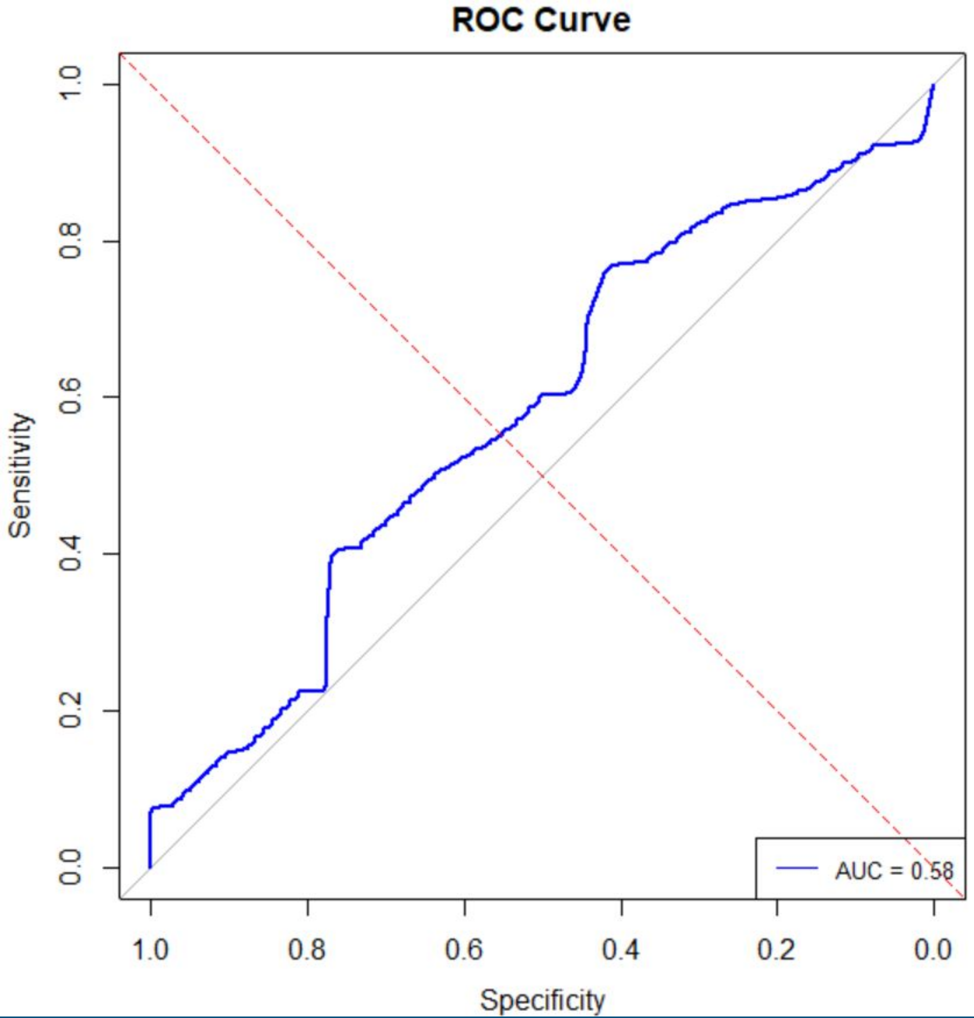
1. Assign independent variable as binary factor using conversion script
2. Developing a binary equation using GLM method
3. Summary GLM model for assessing fit-parameter and AIC
4. Calculating Feature Important and VIF from the model
5. Computing Accuracy, Sensitivity, and Specificity

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  7.620e-01  1.652e-02  46.117  < 2e-16 ***
Month       -3.406e-01  5.417e-03 -62.886  < 2e-16 ***
Hours       -6.552e-03  6.161e-04 -10.635  < 2e-16 ***
Minute      -7.061e-05  2.464e-04  -0.287    0.774    
Canopy       6.363e-02  8.533e-03   7.457 8.87e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 310366  on 223881  degrees of freedom
Residual deviance: 306201  on 223877  degrees of freedom
AIC: 306211
```

```
> caret::varImp(model)
         Overall
Month  62.885960
Hours  10.634747
Minute  0.286595
Canopy  7.456662
> car::vif(model)
   Month    Hours   Minute   Canopy 
1.000311 1.000293 1.000000 1.000018 
```



Indicator	Value
Accuracy	0.549
Sensitivity	0.546
Specificity	0.552

Conclusions

✘/✔	Hypothesis #	Results
✘	Hypothesis 1	
✘	Hypothesis 2	Positions 1-4 warmer than controls
✔	Hypothesis 3	Position 4 had the warmest average night temperature- gorse insulated row
✔	Hypothesis 4	On the coldest day, the average temp of the control sensors was most similar to the average temperature of the #4 sensors
✘	Hypothesis 5	The average temp of the control sensors and the #4 sensors was not higher than the average temps of the #1, #2, and #3 sensors on the coldest day
✘	Hypothesis 6	The average temp of sensor #1 was not lower than that of sensor #4
✘	Hypothesis 7	The #1 and #2 sensors had the highest average maximum temperature
✔	Hypothesis 8	The #4 sensors had the lowest average minimum temperatures
✘/✔	Hypothesis 9	The positions of sensors #4 is the least advantageous in terms of photosynthesis along with the control sensor. The sensor #3 position actually turned out to be the most advantageous.
✘	Hypothesis 10	Logistic regression provided a low accuracy