Kari Char Big Reed

Sander Elliott - Kari Minisale

2026-01-29

Load Packages

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.5.1
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.5.1
```

```
## Warning: package 'tidyr' was built under R version 4.5.1
```

```
## Warning: package 'readr' was built under R version 4.5.1
```

```
## Warning: package 'purrr' was built under R version 4.5.1
```

```
## Warning: package 'stringr' was built under R version 4.5.1
```

```
## Warning: package 'forcats' was built under R version 4.5.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr   1.1.4     v readr   2.1.5
## v forcats 1.0.1     v stringr 1.5.1
## v ggplot2 3.5.2     v tibble  3.2.1
## v purrr   1.1.0     v tidyr   1.3.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
```

Kari's Script

```
# note - the top temp logger only took one point of data on the first day
#        so I didn't include it

hobo_2 <- read.csv("10040750(2nd_down) - 10040750(2nd_down).csv")
hobo_3 <- read.csv("10040752_0(3rd_down) - 10040752_0(3rd_down).csv")
hobo_4 <- read.csv("10040751(4th_down) - 10040751(4th_down).csv")
hobo_5 <- read.csv("10040749(5th_down) - 10040749(5th_down).csv")
hobo_6 <- read.csv("10040748(Bottom) - 10040748(Bottom).csv")

# Library Load ----
#library(tidyverse)

# Data tidying ----
# . Adding column for hobo logger number and depth ----
hobo_2$depth <- 3
hobo_3$depth <- 5
hobo_4$depth <- 7
hobo_5$depth <- 9
hobo_6$depth <- 11

# . Adding column for hobo logger number ----
hobo_2$hobo_num <- 2
hobo_3$hobo_num <- 3
hobo_4$hobo_num <- 4
hobo_5$hobo_num <- 5
hobo_6$hobo_num <- 6

# . Smooshing everything together ----
hobo_list <- list(hobo_2, hobo_3, hobo_4, hobo_5, hobo_6)
br <- bind_rows(hobo_list)

# . Just checking ----
head(br)
```

```
##   Number       Date_Time Temp_C Batt_V Coupler.Attached Host.Connected Stopped
## 1      1 11/22/2011 0:00  4.831   3.01
## 2      2 11/22/2011 0:30  4.831   3.01
## 3      3 11/22/2011 1:00  4.831   3.01
## 4      4 11/22/2011 1:30  4.831   3.01
## 5      5 11/22/2011 2:00  4.831   3.01
## 6      6 11/22/2011 2:30  4.831   3.01
##   End.Of.File depth hobo_num
## 1                 3        2
## 2                 3        2
## 3                 3        2
## 4                 3        2
```

```
## 5                 3        2
## 6                 3        2
```

```
unique(br$depth)
```

```
## [1]  3  5  7  9 11
```

```
unique(br$hobo_num)
```

```
## [1] 2 3 4 5 6
```

Look at some stuff

Make a plot that shows temperature by month, year, and HOBO

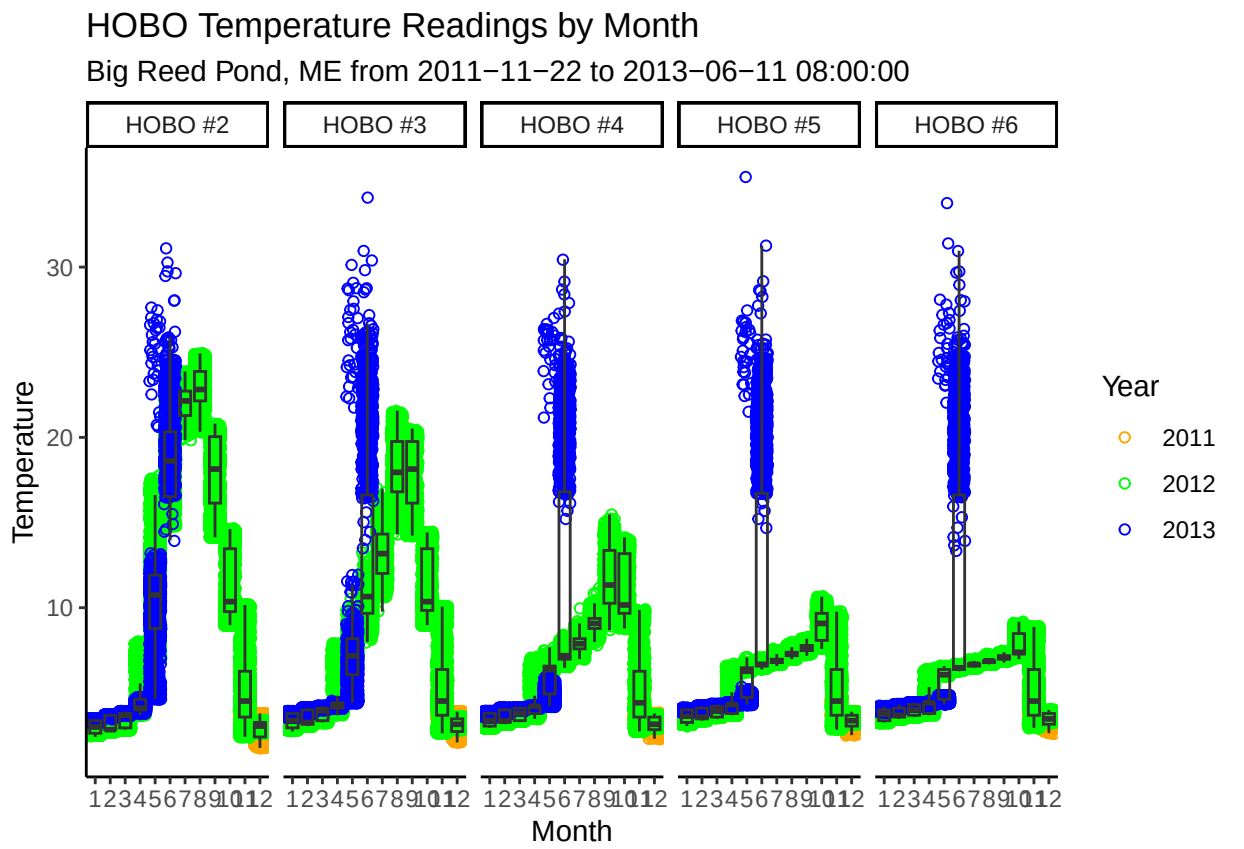
```
class(br$Date_Time)
```

```
## [1] "character"
```

```
datebr <- br %>%
  mutate(Date_Time = mdy_hm(Date_Time),
         month = month(Date_Time), ## Get Month
         year = year(Date_Time)) %>%  ## Get Year
  drop_na(Temp_C) ## Drop rows with no temp readings

start <- min(datebr$Date_Time) ## Get a start Date
end <- max(datebr$Date_Time) ## End Date

ggplot(datebr, aes(x = factor(month), y = Temp_C)) +
  ## Make month a factor so it bins
  geom_jitter(aes(color = factor(year)), shape = 21, fill = NA, na.rm = TRUE) +
  ## Display points with small amount of random variation along x axis
  geom_boxplot(fill = NA, outlier.shape = NA,
               na.rm = TRUE) +
  scale_color_manual(
    name = "Year",
    values = c("2011" = "orange","2012" = "green", "2013" = "blue")) +
   labs(
    title = "HOBO Temperature Readings by Month",
    subtitle = paste("Big Reed Pond, ME from", paste(start, end, sep = " to ")),
    x = "Month",
    y = "Temperature",
    color = "Year"
  ) +
  facet_wrap(~hobo_num, ## separate plots by HOBO number
             nrow = 1,
             labeller = labeller(hobo_num = function(x) paste0("HOBO #", x))) +
  theme_classic()
```



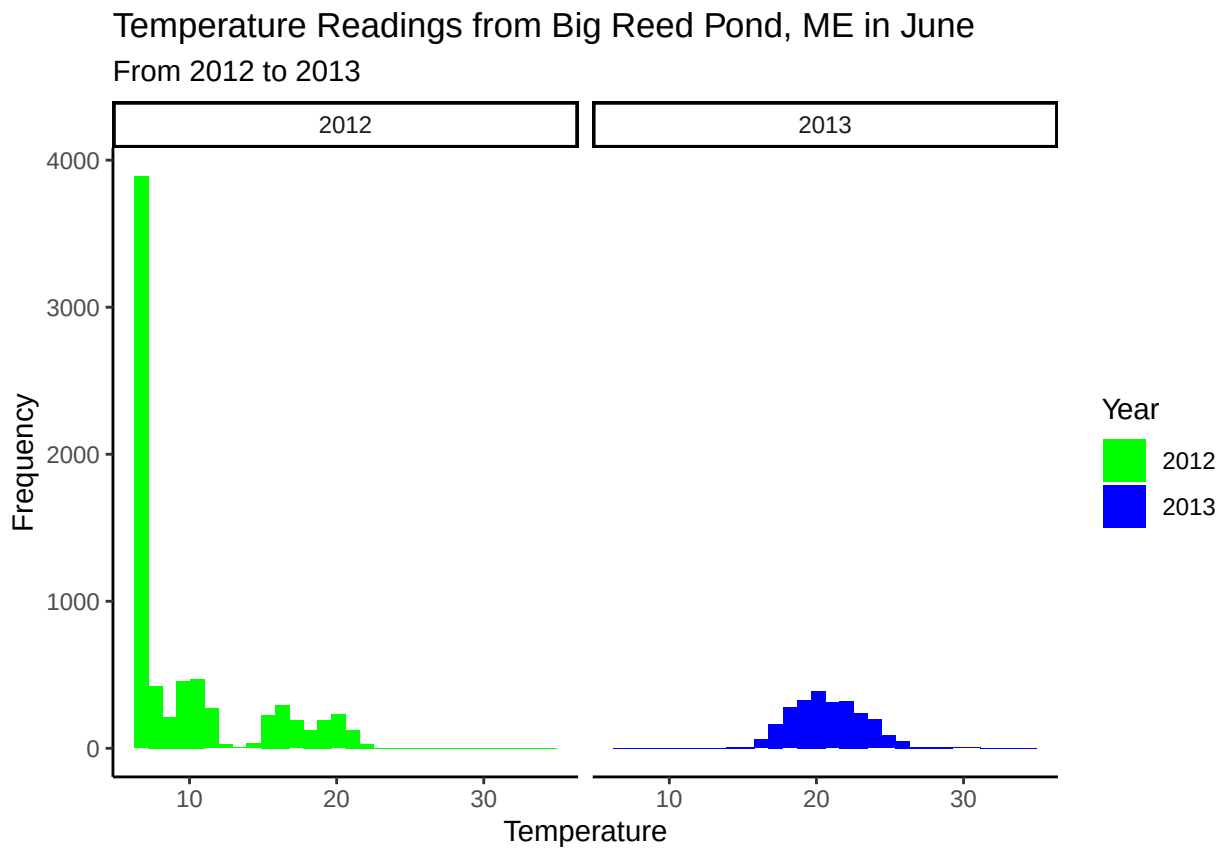
HOBOs 3-6 have interesting temperature distributions for June. There seems to be way higher temperatures in 2013. Almost looks like bimodal distribution but will have to check. Otherwise, the full year (2012), looks like a normal seasonal temperature fluctuation.

Check out distributions for June and compare to overall distribution.

```
junebr <- datebr %>%
  filter(month == 6) ## Just looking at June

ggplot(junebr, aes(x = Temp_C, fill = factor(year))) +
  geom_histogram() +
  scale_fill_manual(name = "Year",
                    values = c("2012" = "green", "2013" = "blue")) +
  theme_classic() +
  labs(
    title = "Temperature Readings from Big Reed Pond, ME in June",
    subtitle = "From 2012 to 2013",
    x = "Temperature",
    y = "Frequency"
  ) +
  facet_wrap(~year) ## separate plots by year
```

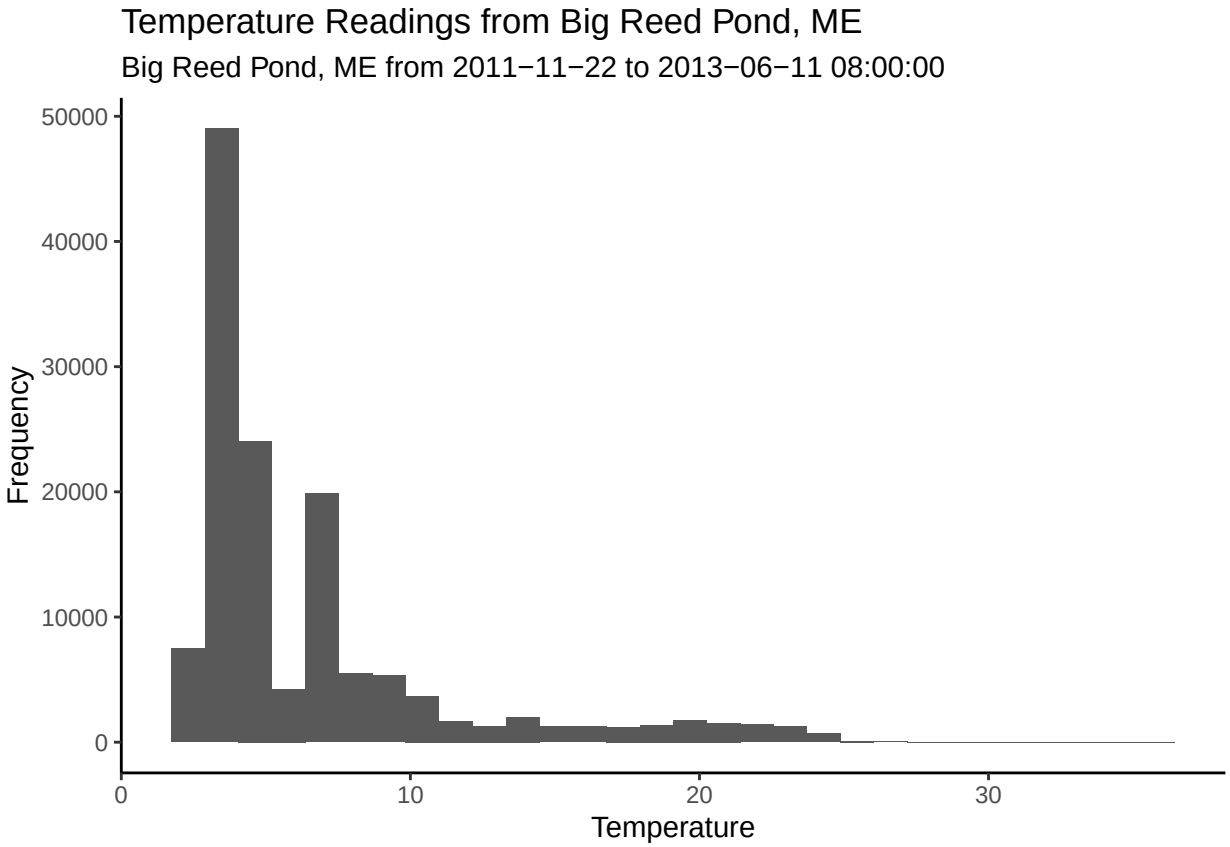
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## Make a separate plot without breaking apart year or filtering month

ggplot(datebr, aes(x = Temp_C)) +
  geom_histogram() +
  theme_classic() +
  labs(
    title = "Temperature Readings from Big Reed Pond, ME",
    subtitle =  paste("Big Reed Pond, ME from",
                      paste(start, end, sep = " to ")),
    x = "Temperature",
    y = "Frequency"
  )
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



In June, 2013 has a more normal distribution, but 2012 more closely resembles the overall distribution of the data. It is worth looking closer at the differences between 2012 and 2013. 2012 is the only full year in the data set. Year will likely have a large effect on how the analysis is conducted.

Make Table to Summarise Depth, HOBO #, and Temperature.

```
brsumstat  <- datebr %>%
  group_by(depth, hobo_num) %>%
  summarise(Low_Temp = min(Temp_C),
            High_Temp =  max(Temp_C),
            AVG_Temp = mean(Temp_C),
            SE_Temp = sd(Temp_C)/sqrt(n())
            )
```

```
## `summarise()` has grouped output by 'depth'. You can override using the
## `.groups` argument.
```

```
brsumstat
```

```
## # A tibble: 5 x 6
## # Groups:   depth [5]
##   depth hobo_num Low_Temp High_Temp AVG_Temp SE_Temp
##   <dbl>    <dbl>    <dbl>     <dbl>    <dbl>   <dbl>
```

```
## 1     3        2     1.76      31.1     8.69  0.0434
## 2     5        3     2.09      34.1     7.32  0.0322
## 3     7        4     2.30      30.5     5.88  0.0212
## 4     9        5     2.52      35.3     5.43  0.0172
## 5    11        6     2.62      33.7     5.31  0.0164
```

It looks like the average temperature decreases with lower depths along with the variation of temperature. Interestingly, the high temperature seems to vary by depth, alternating between high and low. The low temperature increases with depth. It is worthwhile to see how this table changes when year is counted for.

Just one full year (2012)

```
br2012sumstat  <- datebr %>%
  filter(year == 2012) %>% ## just 2012
  group_by(depth, hobo_num) %>%
  summarise(Low_Temp = min(Temp_C),
            High_Temp =  max(Temp_C),
            AVG_Temp = mean(Temp_C),
            SE_Temp = sd(Temp_C)/sqrt(n())
            )
```

```
## `summarise()` has grouped output by 'depth'. You can override using the
## `.groups` argument.
```

```
br2012sumstat
```

```
## # A tibble: 5 x 6
## # Groups:   depth [5]
##   depth hobo_num Low_Temp High_Temp AVG_Temp SE_Temp
##   <dbl>    <dbl>    <dbl>     <dbl>    <dbl>   <dbl>
## 1     3        2     2.41      24.9     10.5   0.0579
## 2     5        3     2.62      21.6      8.54  0.0415
## 3     7        4     2.73      15.5      6.42  0.0233
## 4     9        5     2.84      10.7      5.68  0.0150
## 5    11        6     2.94       9.18     5.46  0.0125
```

Looking at just 2012, the fluctuation in high temperature goes away and instead decreases with lower depths. The values for low temperature, average temperature, and variation in temperature change, but the same trends remain.

AI Statement: Microsoft CoPilot was used to find geom_jitter which helped me to overlay points on a box plot without them all being on one line. I found this useful and will use this function more in the future. CoPilot also helped me troubleshoot some issues I was having with facet_wrap(). I had the right function but did not know to use a ~ before the facet or how to use labeller() inside facet_wrap().