ECOL 610: NEON Assignment 1

Group - Santa Rita Experimental Range (SRER)

14 September, 2022

- Emily Swartz
- Shahriar Shah Heydari
- Stephanie Cardinalli
- George Woolsey

In Class

Setup

This is the R script for ECOL610 NEON assignment 1 This script will provide code to carry out assignment 1 for the Central Plains Experimental Range - you will need to modify this code to carry it out for your site.

First, load in the needed packages. Install the packages if needed.

```
library(tidyverse)
library(lubridate)
library(viridis)
library(scales)
library(latex2exp)
```

Load in the data. You will need to change the file path for your directory.

```
#load in the data
CPER_30 <- read.csv("../data/Central Plains Experimental Range - 30 min.csv")
CPER_daily <- read.csv("../data/Central Plains Experimental Range - daily.csv")
```

Question 1

Calculate GPP and compare to NEON GPP for summer and winter week

NEE is defined, by convention, as CO_2 flux from the ecosystem to the atmosphere. It corresponds to a negative carbon input to ecosystems. NEE is defined in this way because atmospheric scientists, who originated the term, seek to document net sources of CO_2 to the atmosphere (i.e., NEE) that account for rising atmospheric CO_2 concentration. Therefore, CO_2 input to the ecosystem is a negative NEE.

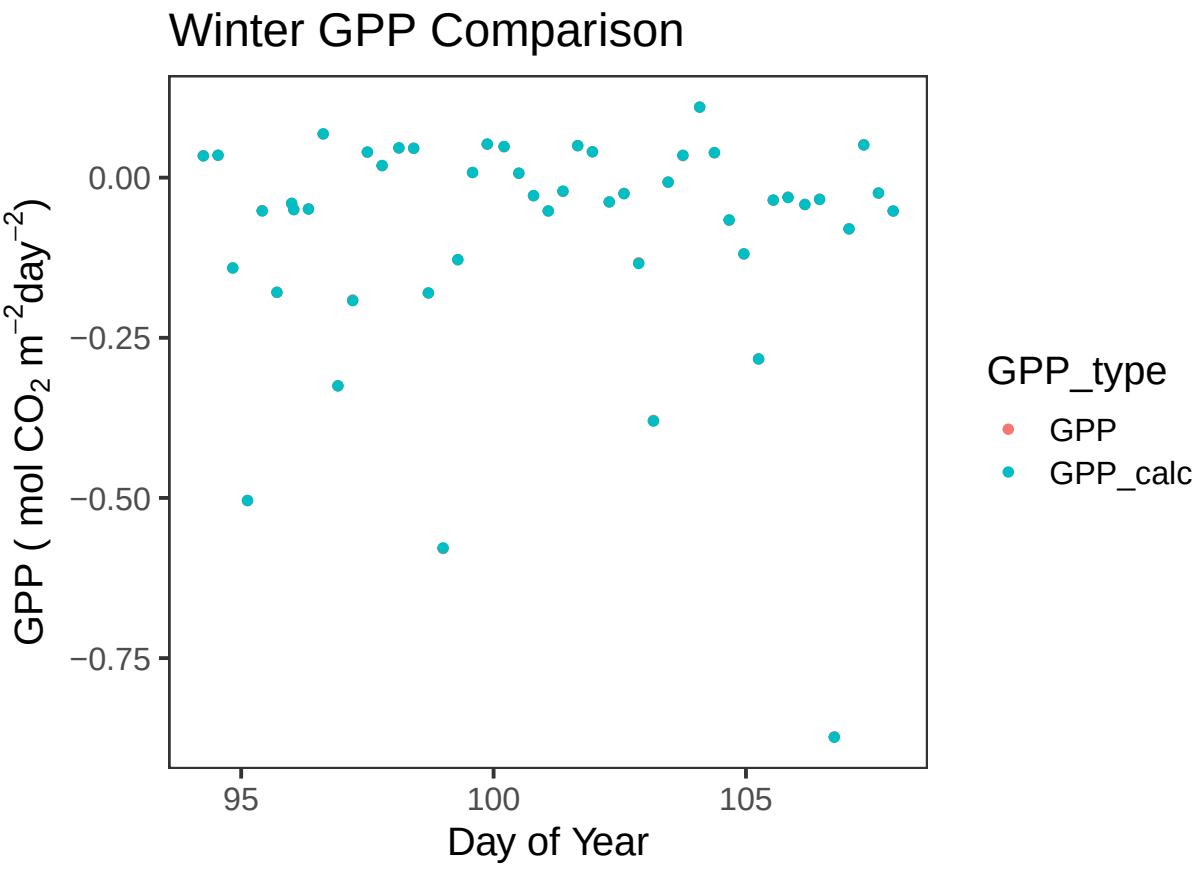
Chapin, F. S., Matson, P. A., Mooney, H. A., & Vitousek, P. M. (2002). Principles of terrestrial ecosystem ecology. p.208

$$NEE = R_E - GPP$$
$$GPP = R_E - NEE$$

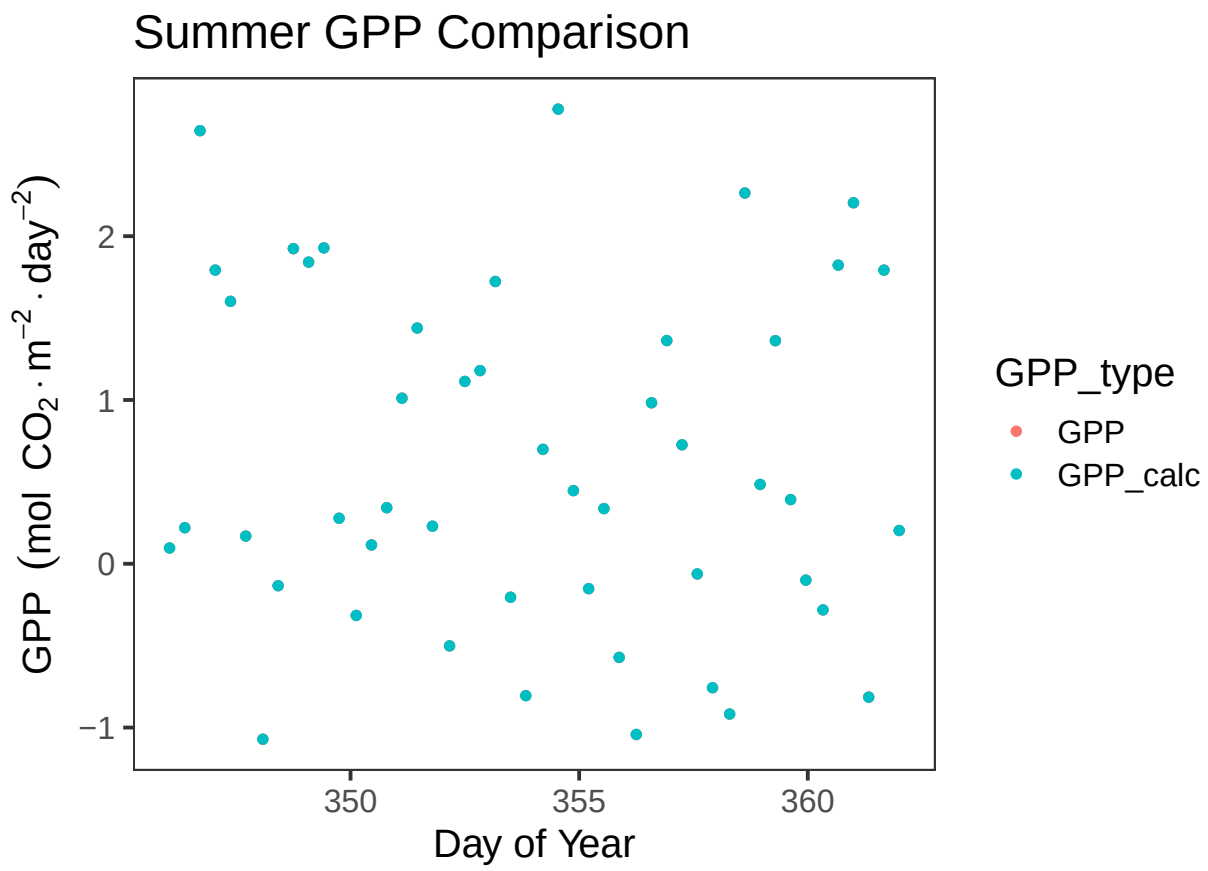
```r
#calculate GPP from NEE and Re using equation NEE = Re - GPP (negative values land C storage)
CPER_30 <- CPER_30 %>%
  dplyr::mutate(
    GPP_calc = as.numeric(RE - NEE)
    , GPP_is_equal = round(GPP_calc, 3) == round(GPP, 3)
  )

# are there rows where the data is not equal?
  nrow(
    CPER_30 %>%
      dplyr::filter(
        GPP_is_equal == FALSE
        & !is.na(GPP)
      )
  )
```

```
## [1] 8499
```

```r
#clipping the data to a winter and summer week
 #choose your own weeks (without data gaps) for your site!
#February 16-22, 2020
CPER_winter_week <- filter(CPER_30, Year == "2020" & DOY == 47:53)
#June 21-27, 2020
CPER_summer_week <- filter(CPER_30, Year == "2020" & DOY == 173:180)

#creating long format for plotting
CWW_stacked <-gather(CPER_winter_week, "GPP_type", "GPP_value", c(9,15))
CSW_stacked <-gather(CPER_summer_week, "GPP_type", "GPP_value", c(9,15))

#winter
#plot of NEON vs calculated GPP
ggplot(CWW_stacked,aes(x=DOY.total,y=GPP_value, color = GPP_type)) + geom_point() +
    xlab("Day of Year") +
    ylab(expression(paste("GPP ( mol CO"[2]*" m"^{-2}, "day"^{-2}*")"))) +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
```



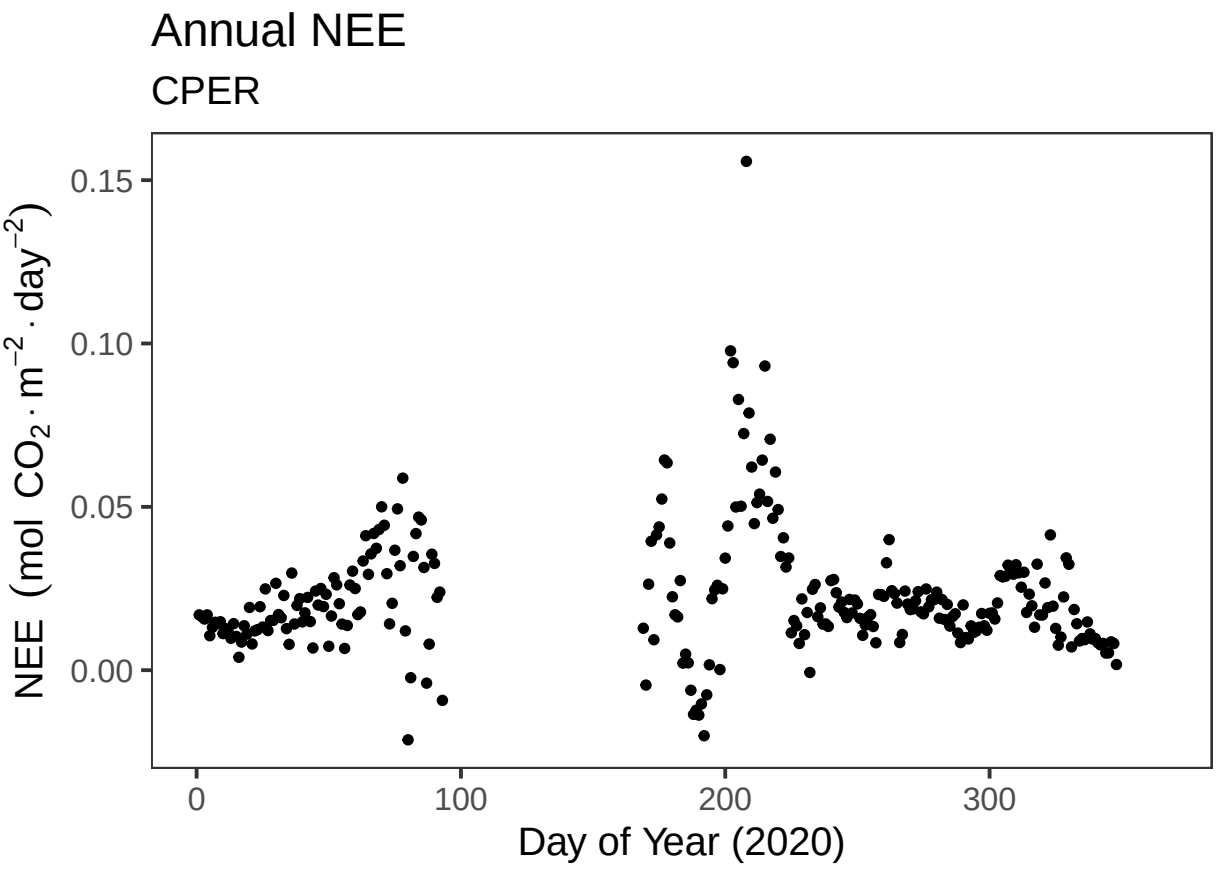

```
#summer
#plot of NEON vs calculated GPP
ggplot(CSW_stacked,aes(x=DOY.total,y=GPP_value, color = GPP_type)) + geom_point() +
    xlab("Day of Year") +
    ylab(latex2exp::TeX("$GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
```



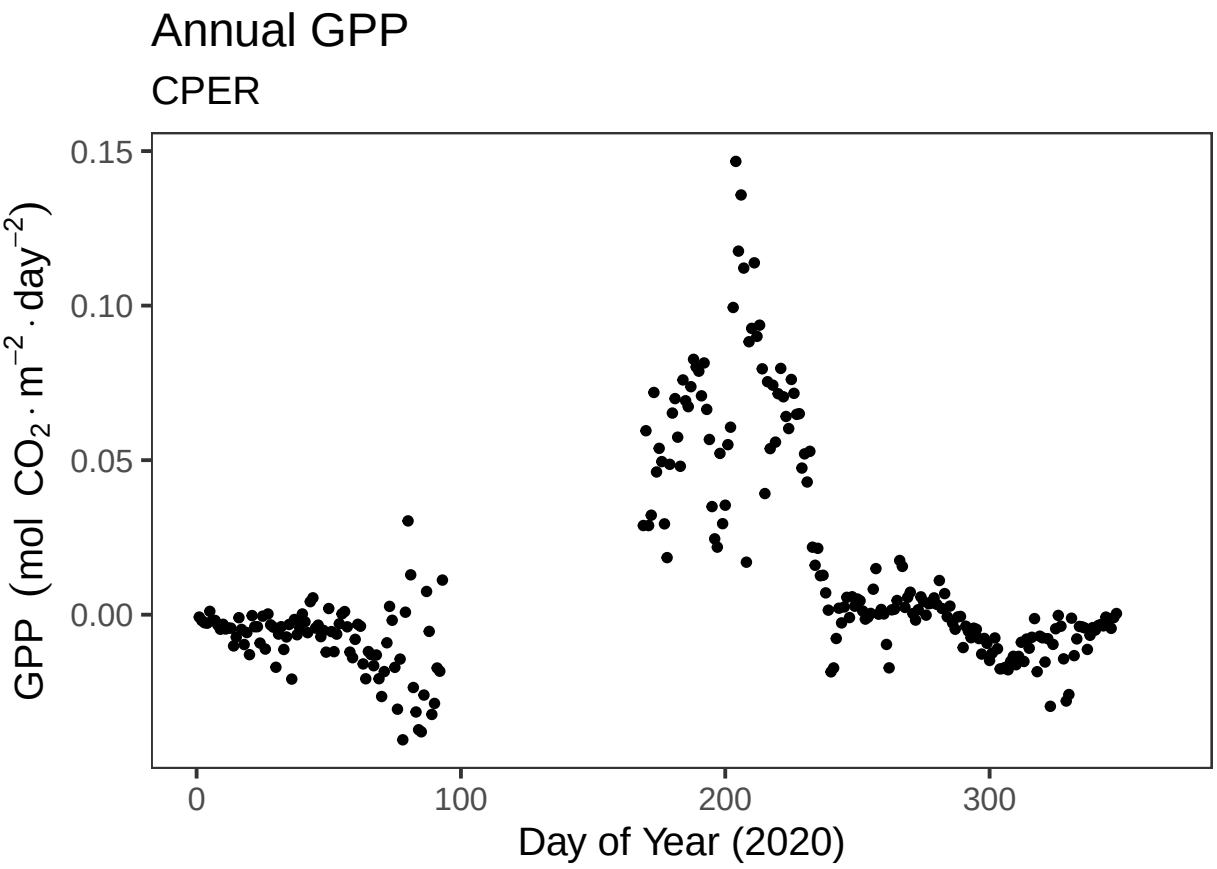
Question 2

annual patterns in C exchange and environmental properties

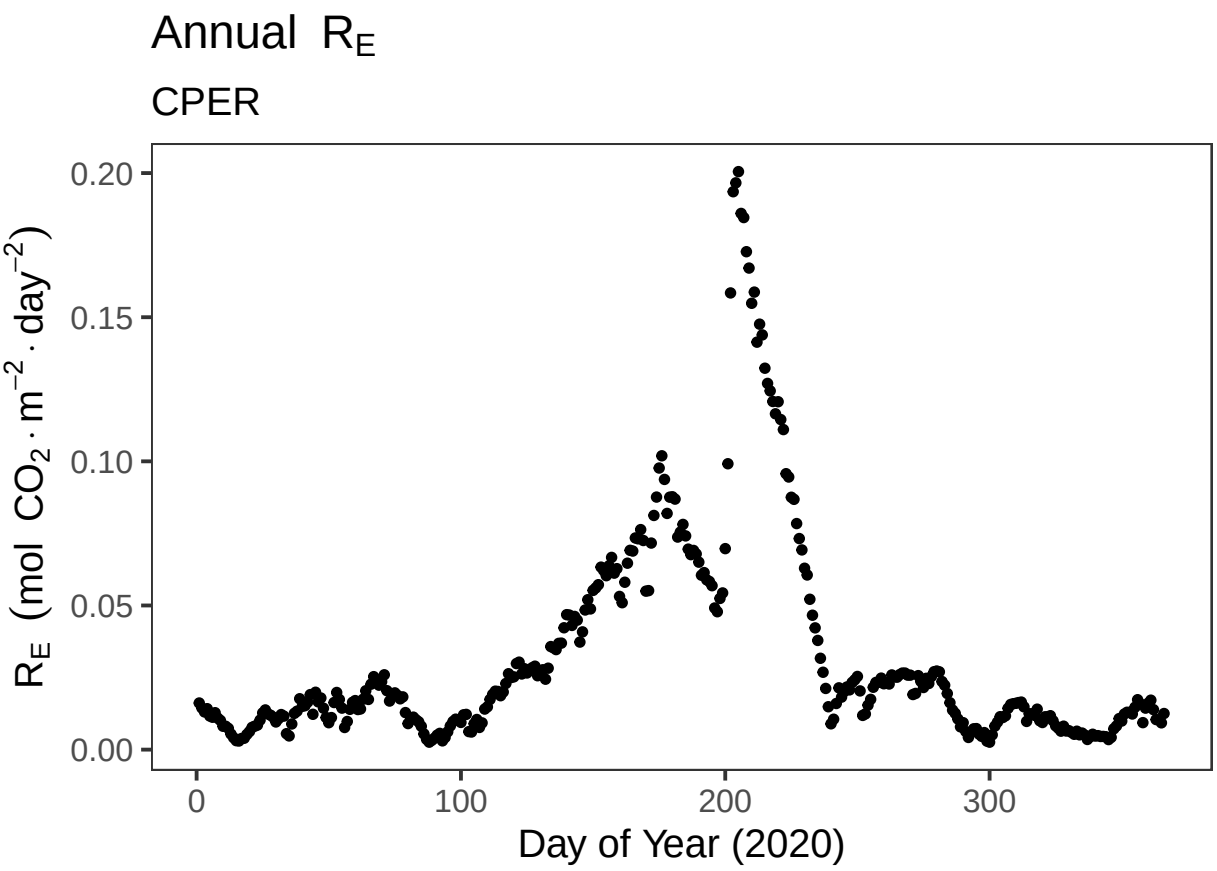
```
#reduce data to just 2020
CPER_2020 <- filter(CPER_daily, Year == "2020")
#plot of NEE
ggplot(CPER_2020,aes(x=DOY,y=NEE)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab(latex2exp::TeX("$NEE \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = "Annual NEE",
         subtitle = "CPER")
```



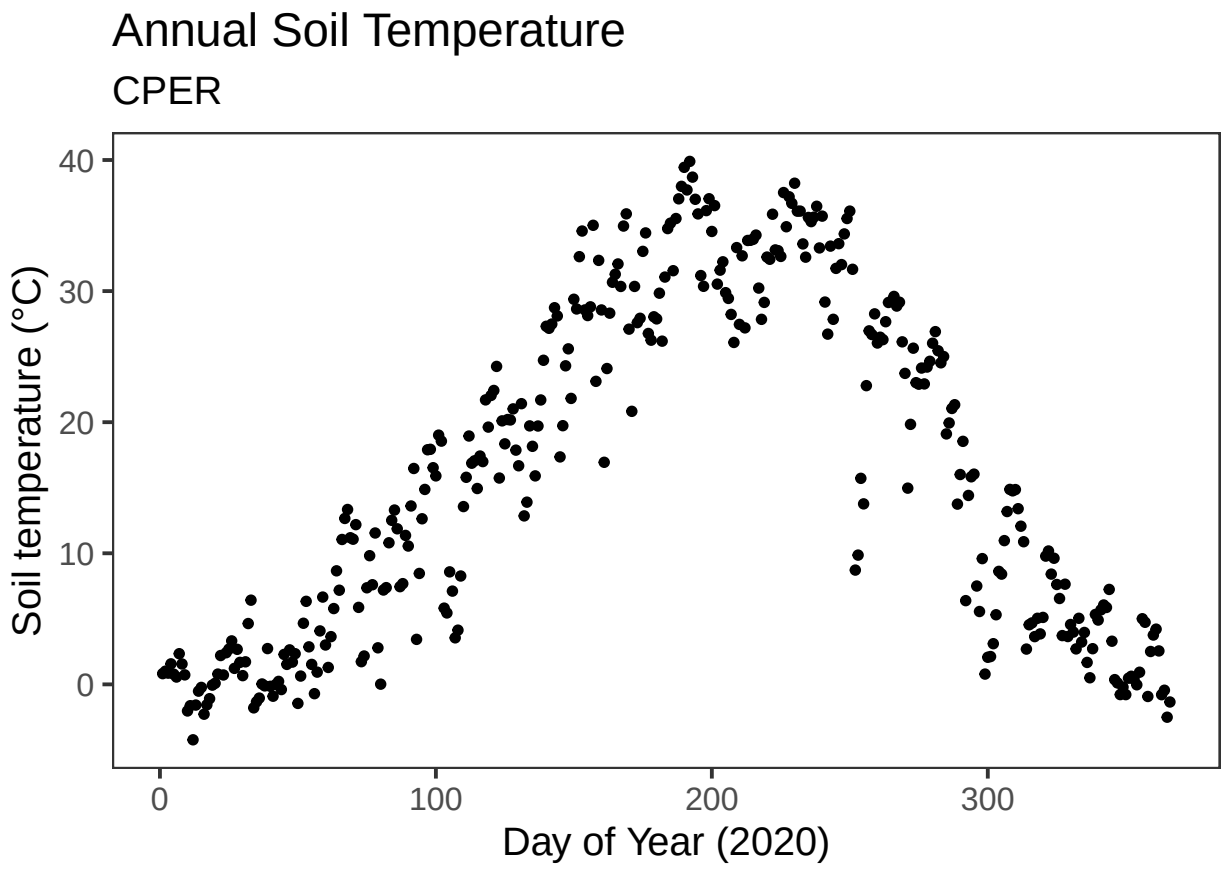

```
#plot of GPP
ggplot(CPER_2020,aes(x=DOY,y=GPP)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab(latex2exp::TeX("$GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = "Annual GPP",
         subtitle = "CPER")
```



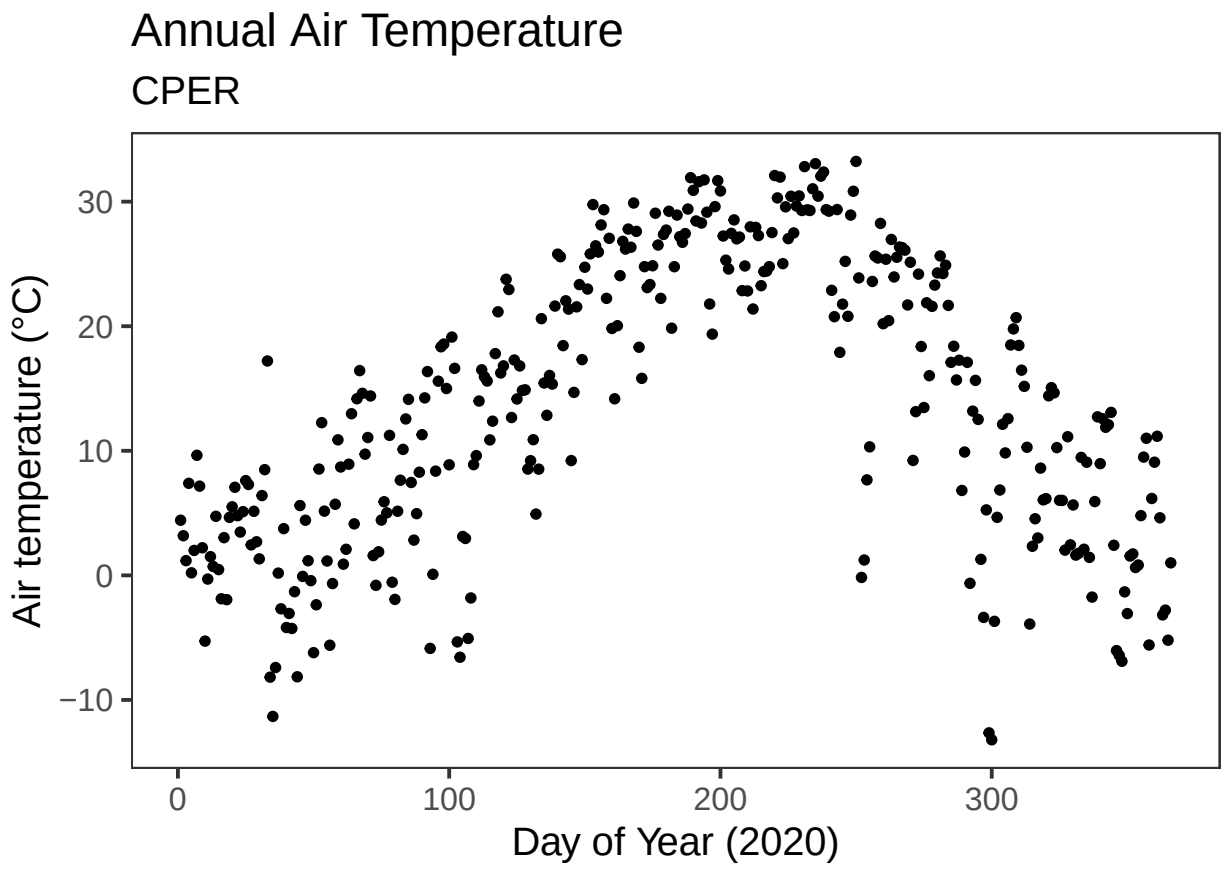

```
#plot of Re
ggplot(CPER_2020,aes(x=DOY,y=RE)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab(latex2exp::TeX("$R_{E} \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = latex2exp::TeX("$Annual \\; R_{E}$"),
         subtitle = "CPER")
```



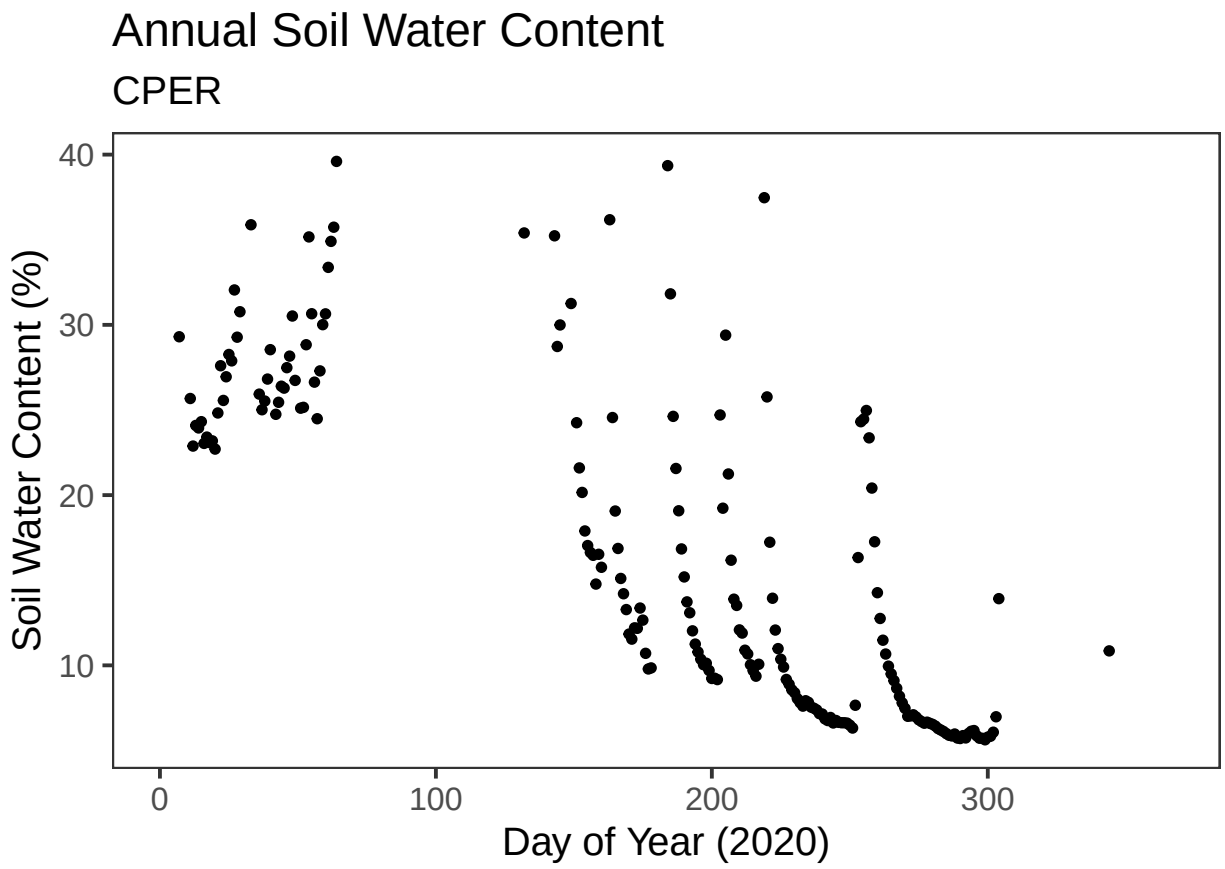

```
#plot of soil temp
ggplot(CPER_2020,aes(x=DOY,y=TS)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab("Soil temperature (\u00B0C)") +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = "Annual Soil Temperature",
         subtitle = "CPER")
```




```
#plot of air temp
ggplot(CPER_2020,aes(x=DOY,y=TA)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab("Air temperature (\u00B0C)") +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = "Annual Air Temperature",
         subtitle = "CPER")
```




```
#plot of soil moisture
ggplot(CPER_2020,aes(x=DOY,y=SWC)) + geom_point() +
    xlab("Day of Year (2020)") +
    ylab("Soil Water Content (%)") +
    theme_bw(base_size = 16) + theme(panel.grid.major = element_blank(), panel.grid.minor = element_bla
    labs(title = "Annual Soil Water Content",
         subtitle = "CPER")
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Assignment

Load data

```
remove(list=ls())
# what is your site name?
site <- "Santa Rita Experimental Range"
# load 30 min data
f_list <- list.files(path = "../data/", pattern="*30 min.csv")
for (i in 1:length(f_list)){
  nm <- stringr::word(f_list[i], start = 1, sep = "30 min") %>%
    stringr::str_replace_all(pattern = "[[:punct:]]", replacement = "") %>%
    stringr::str_trim() %>%
    stringr::str_squish()
  temp <- read.csv(paste0("../data/", f_list[i])) %>%
    dplyr::mutate(neon_site_name = nm) %>%
    dplyr::relocate(neon_site_name) %>%
    dplyr::rename_with(~ tolower(
      gsub(" ", "_",
         str_trim(gsub("\\s+", " ", .x))
```

```
      )
    ))
  if(i==1){
    dta_30min <- temp
  }else{
    dta_30min <- dplyr::union_all(dta_30min, temp)
  }
  remove(temp)
}
# load daily data
f_list <- list.files(path = "../data/", pattern="*daily.csv")
for (i in 1:length(f_list)){
  nm <- stringr::word(f_list[i], start = 1, sep = "daily") %>%
    stringr::str_replace_all(pattern = "[[:punct:]]", replacement = "") %>%
    stringr::str_trim() %>%
    stringr::str_squish()
  temp <- read.csv(paste0("../data/", f_list[i])) %>%
    dplyr::mutate(neon_site_name = nm) %>%
    dplyr::relocate(neon_site_name) %>%
    dplyr::rename_with(~ tolower(
      gsub(" ", "_",
         str_trim(gsub("\\s+", " ", .x))
      )
    ))
  if(i==1){
    dta_1day <- temp
  }else{
    dta_1day <- dplyr::union_all(dta_1day, temp)
  }
  remove(temp)
}
# create dates and record counts
# day
dta_1day <- dta_1day %>%
  dplyr::mutate(
    date_id = lubridate::make_date(year = year, month = month, day = day)
    , week = lubridate::week(date_id)
    , has_gpp = ifelse(!is.na(gpp), 1, 0)
    , season =
        dplyr::case_when(
          month %in% c(1:3, 12) ~ "Winter"
          , month %in% c(6:8) ~ "Summer"
          , TRUE  ~ "other")
  ) %>%
  dplyr::group_by(neon_site_name, week, year) %>%
  dplyr::mutate(is_full_week = sum(has_gpp)==7) %>%
  dplyr::ungroup()
# 30
dta_30min <- dta_30min %>%
  dplyr::mutate(
    date_id = lubridate::make_date(year = year, month = month, day = day)
    , time_id = lubridate::make_datetime(year = year, month = month, day = day
                          , hour = floor(hour)
```

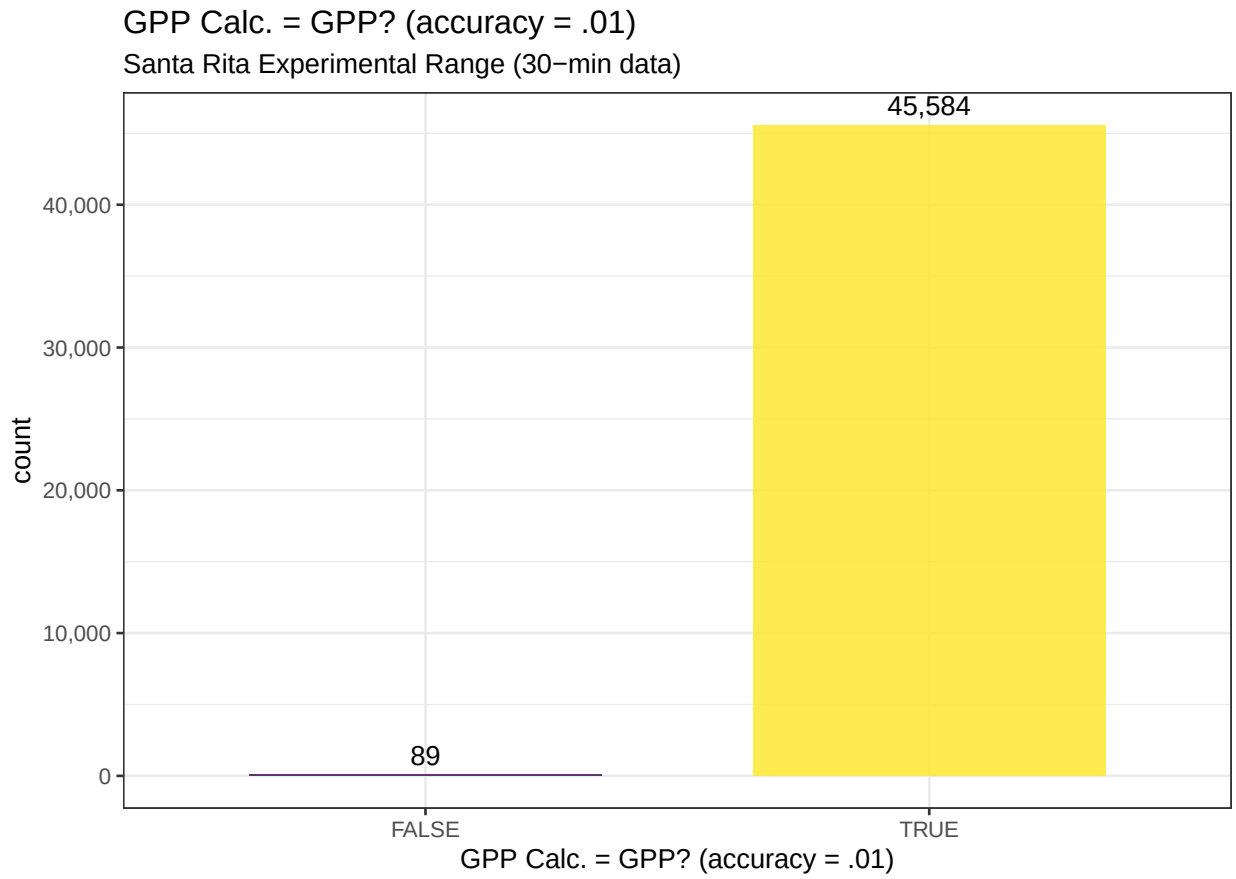
```
                        , min = (hour-floor(hour))*60
                )
    , week = lubridate::week(date_id)
    , has_gpp = ifelse(!is.na(gpp), 1, 0)
    , season =
        dplyr::case_when(
          month %in% c(1:3, 12) ~ "Winter"
          , month %in% c(6:8) ~ "Summer"
          , TRUE  ~ "other")
  ) %>%
  dplyr::group_by(neon_site_name, week, year) %>%
  dplyr::mutate(is_full_week = sum(has_gpp)==24*2*7) %>%
  dplyr::ungroup()
```

The draft questions for that assignment are:

Question 1

Using the 30 minute data. Calculate GPP for your site.

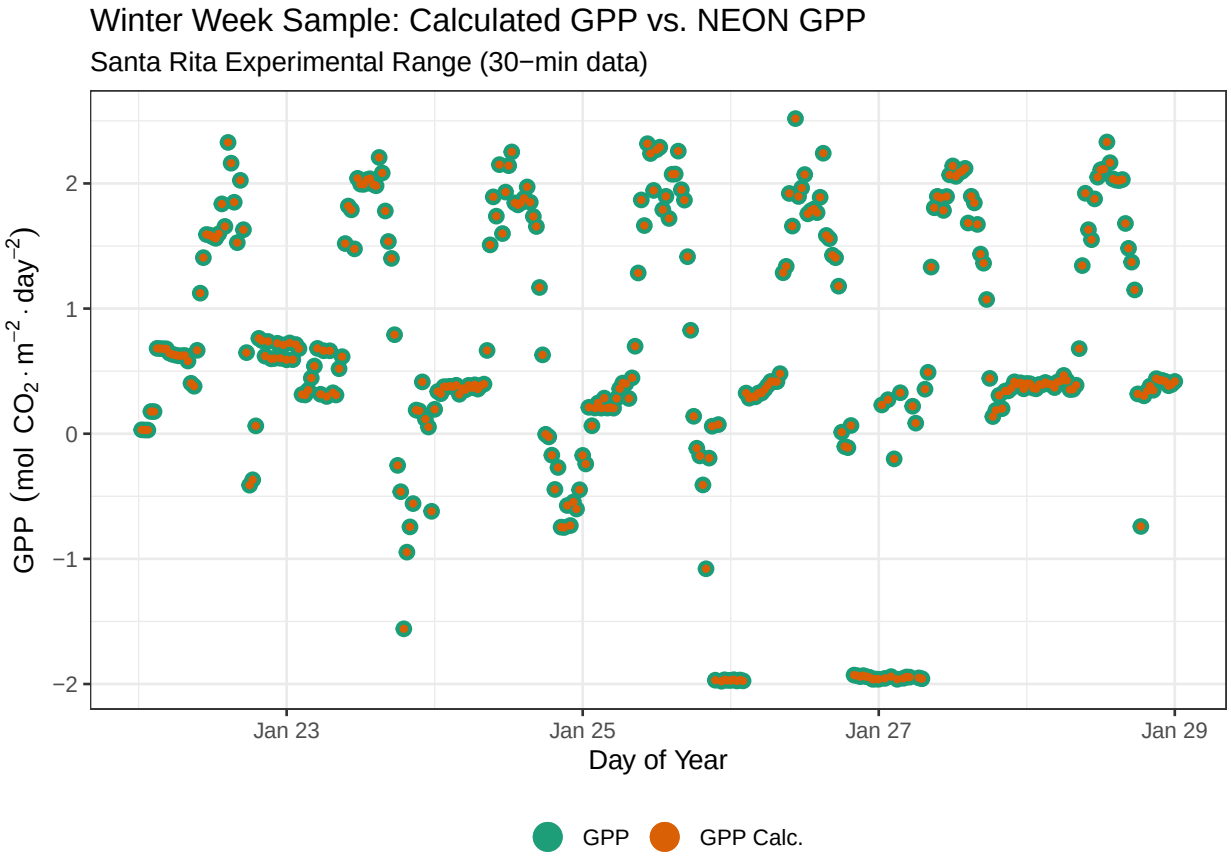
```
dta_30min <- dta_30min %>%
  dplyr::mutate(
    gpp_calc = as.numeric(re - nee)
    , gpp_is_equal = round(gpp_calc, 2) == round(gpp, 2)
  )
# quick plot of gpp equality
dta_30min %>%
  dplyr::filter(
    neon_site_name == site
    & !is.na(gpp)
  ) %>%
ggplot(., aes(x = gpp_is_equal)) +
  geom_bar(
    aes(fill = gpp_is_equal)
    , width = 0.7
  ) +
  geom_text(
    aes(label = scales::comma(..count..))
    , stat = "count"
    , vjust = -0.5
    , color = "black"
  ) +
  scale_fill_viridis_d(alpha = 0.8) +
  scale_y_continuous(labels = scales::comma) +
  xlab("GPP Calc. = GPP? (accuracy = .01)") +
  labs(
    title = "GPP Calc. = GPP? (accuracy = .01)"
    , subtitle = paste0(site, " (30-min data)")
  ) +
  theme_bw() +
  theme(
    legend.position = "none"
  )
```



Choose a winter and summer week in your dataset. Create a plot with both your calculated GPP and the NEON GPP.

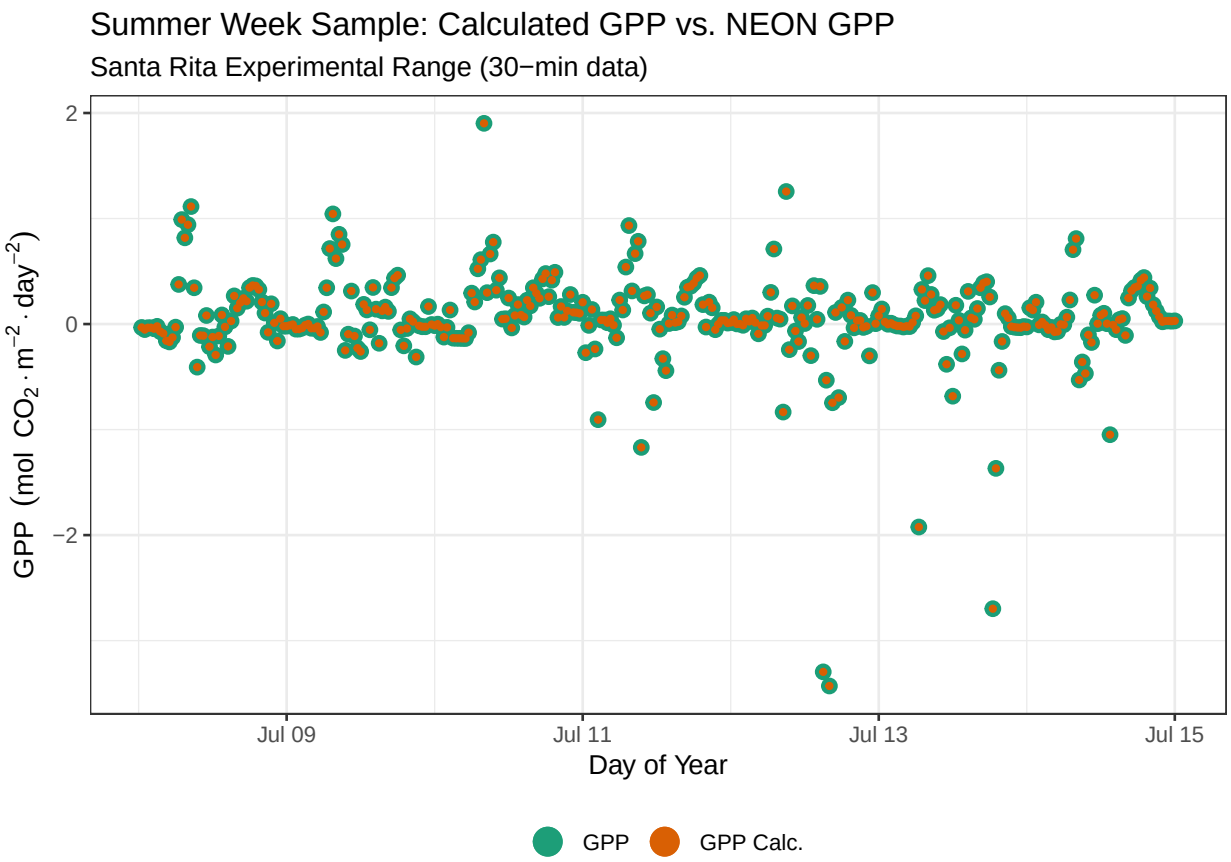
```
# select week with full data for winter and summer
sum_wk <- dta_30min %>%
  dplyr::filter(
    neon_site_name == site
    & season == "Summer"
    & year == 2020
    & is_full_week == TRUE
  ) %>%
  dplyr::slice_sample(n = 1)
wint_wk <- dta_30min %>%
  dplyr::filter(
    neon_site_name == site
    & season == "Winter"
    & year == 2020
    & is_full_week == TRUE
  ) %>%
  dplyr::slice_sample(n = 1)

# plot winter week
dta_30min %>%
  dplyr::filter(
    neon_site_name == site
```

```
    & year == 2020
    & week == wint_wk$week[1]
  ) %>%
ggplot(., aes(x = time_id)) +
  geom_point(
    aes(y = gpp, color = "GPP")
    , size = 2.5
  ) +
  geom_point(
    aes(y = gpp_calc, color = "GPP Calc.")
    , size = 0.9
  ) +
  scale_color_brewer(type = "qual", palette = "Dark2") +
  xlab("Day of Year") +
  ylab(latex2exp::TeX("$GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
  labs(
    title = "Winter Week Sample: Calculated GPP vs. NEON GPP"
    , subtitle = paste0(site, " (30-min data)")
  ) +
  theme_bw() +
  theme(
    legend.position = "bottom"
    , legend.direction = "horizontal"
    , legend.title = element_blank()
  ) +
  guides(color = guide_legend(override.aes = list(size=5, alpha = 0.9)))
```




```
# plot summer week
dta_30min %>%
  dplyr::filter(
    neon_site_name == site
    & year == 2020
    & week == sum_wk$week[1]
  ) %>%
ggplot(., aes(x = time_id)) +
  geom_point(
    aes(y = gpp, color = "GPP")
    , size = 2.5
  ) +
  geom_point(
    aes(y = gpp_calc, color = "GPP Calc.")
    , size = 0.9
  ) +
  scale_color_brewer(type = "qual", palette = "Dark2") +
  xlab("Day of Year") +
  ylab(latex2exp::TeX("$GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
  labs(
    title = "Summer Week Sample: Calculated GPP vs. NEON GPP"
    , subtitle = paste0(site, " (30-min data)")
  ) +
  theme_bw() +
  theme(
```

```
    legend.position = "bottom"
    , legend.direction = "horizontal"
    , legend.title = element_blank()
  ) +
  guides(color = guide_legend(override.aes = list(size=5, alpha = 0.9)))
```



a.

How do the calculated and NEON GPP's compare for your site? Why are they the same or different? (Note: NEON uses eddy covariance to calculate these metrics - more about that here).

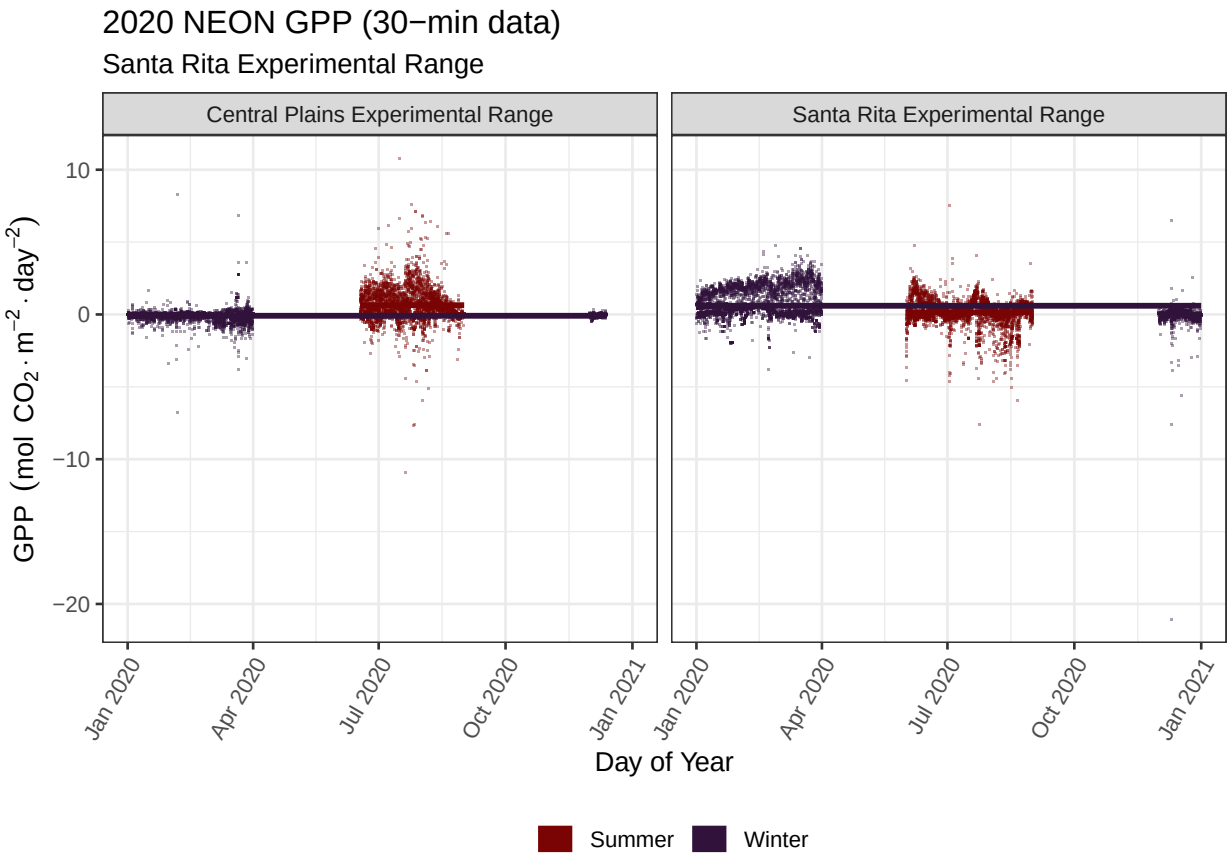
The majority of data points have a calculated GPP ($GPP = R_E - NEE$) equal to the NEON GPP. There are 45,584 records for which the GPP is equal (accuracy = 0.01) and 89 records for which the GPP is not equal. In cases where the GPP is not equal, the difference is negligible and likely due to rounding due to number accuracy.

b.

How do your sites GPP values vary between summer and winter? What do you think is driving these differences?

```
# plot points
dta_30min %>%
```

```
  dplyr::filter(
    # neon_site_name == site
    year == 2020
    & season %in% c("Summer", "Winter")
  ) %>%
ggplot(., aes(x = time_id, y = gpp, color = season)) +
  geom_point(
    # size = 0.2
    alpha = 0.4
    , shape = "."
  ) +
  geom_smooth(
    method = "lm"
    , formula=y~1
    , se = FALSE
    , lwd = 1.1
    , linetype = "solid"
    , alpha = 1
  ) +
  facet_wrap(~neon_site_name) +
  scale_color_viridis_d(option = "turbo", direction = -1) +
  xlab("Day of Year") +
  ylab(latex2exp::TeX("$GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
  labs(
    title = "2020 NEON GPP (30-min data)"
    , subtitle = paste0(site)
  ) +
  theme_bw() +
  theme(
    legend.position = "bottom"
    , legend.direction = "horizontal"
    , legend.title = element_blank()
    , axis.text.x = element_text(angle = 60, hjust=1)
  ) +
  guides(color = guide_legend(override.aes = list(size=5, alpha = 0.9)))
```



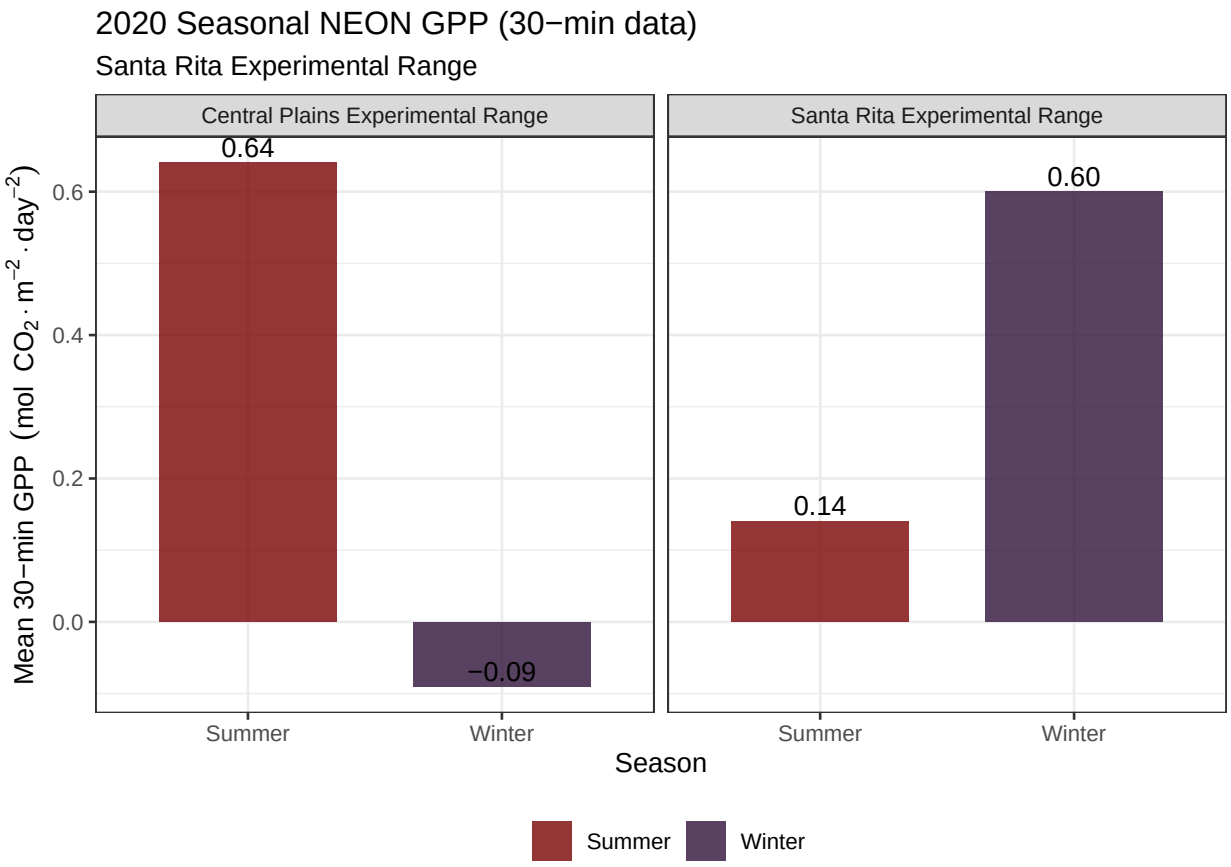
Based on 2020 data, The average 30-minute GPP for the NEON site Santa Rita Experimental Range is higher in the winter months than in the summer months. This is opposite of the pattern for GPP seen at the CPER site. At the Santa Rita Experimental Range, this pattern is likely driven by low summer precipitation and high summer temperatures to the winter.

c.

CPER is a semi-arid grassland. Compare GPP values between your site and CPER - why might they be different or similar?

```
# plot points
dta_30min %>%
  dplyr::filter(
    # neon_site_name == site
    year == 2020
    & season %in% c("Summer", "Winter")
  ) %>%
  dplyr::group_by(neon_site_name, season, year) %>%
  dplyr::summarise(mean_gpp = mean(gpp, na.rm = TRUE)) %>%
ggplot(., aes(x = season, y = mean_gpp, fill = season)) +
  geom_col(
    width = 0.7
    , alpha = 0.8
  ) +
  geom_text(
```

```
  aes(label = scales::comma(mean_gpp, accuracy = 0.01))
  , vjust = -0.3
  , color = "black"
) +
facet_wrap(~neon_site_name) +
scale_fill_viridis_d(option = "turbo", direction = -1) +
xlab("Season") +
ylab(latex2exp::TeX("Mean 30-min $GPP \\; ( mol \\; CO_{2} \\cdot m^{-2} \\cdot day^{-2})$")) +
labs(
  title = "2020 Seasonal NEON GPP (30-min data)"
  , subtitle = paste0(site)
) +
theme_bw() +
theme(
  legend.position = "bottom"
  , legend.direction = "horizontal"
  , legend.title = element_blank()
) +
guides(color = guide_legend(override.aes = list(size=5, alpha = 0.9)))
```

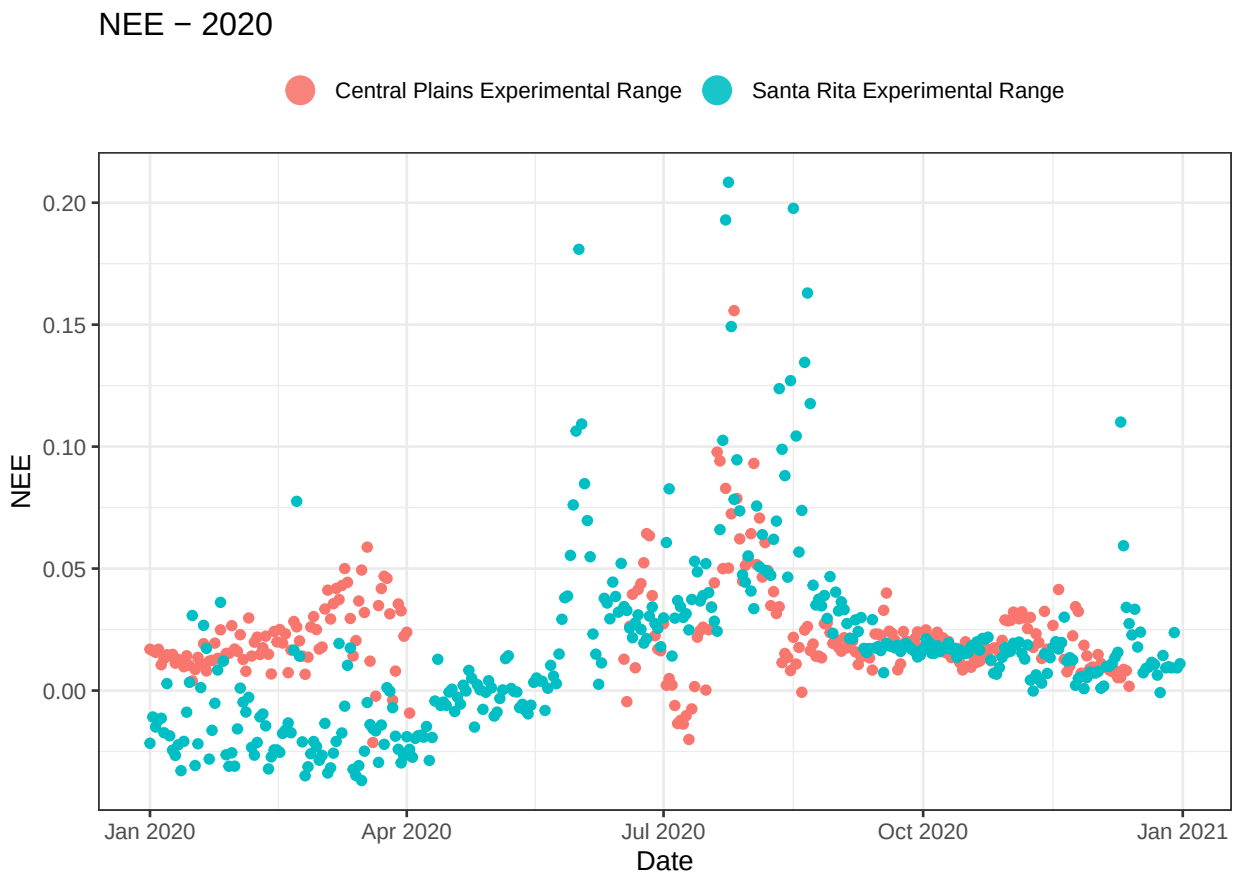


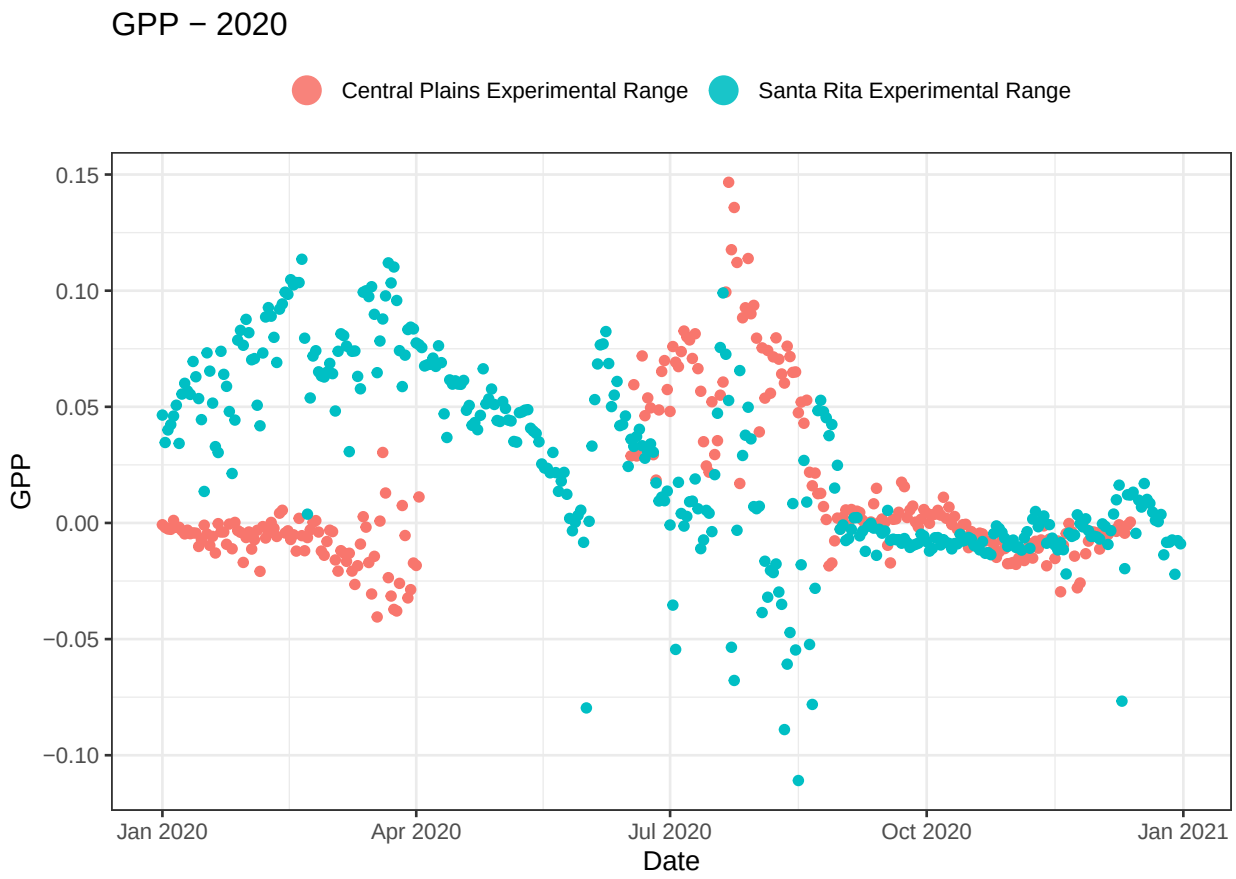
Based on 2020 data, The average 30-minute GPP for the NEON site Santa Rita Experimental Range is higher in the winter months (0.60) than in the summer months (0.14). This is opposite of the pattern for GPP seen at the CPER site. At the Santa Rita Experimental Range, this pattern is likely driven by low summer precipitation and high summer temperatures to the winter.

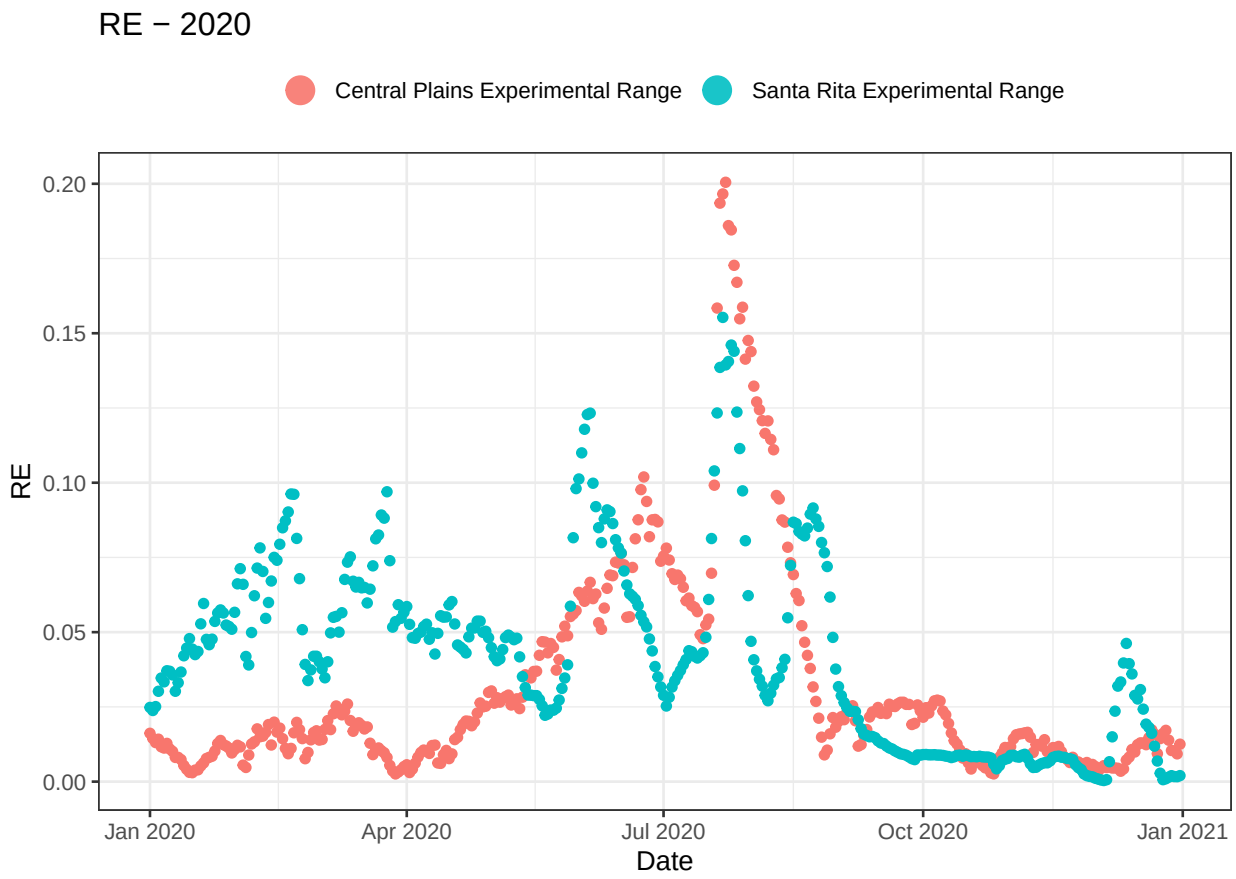
Question 2

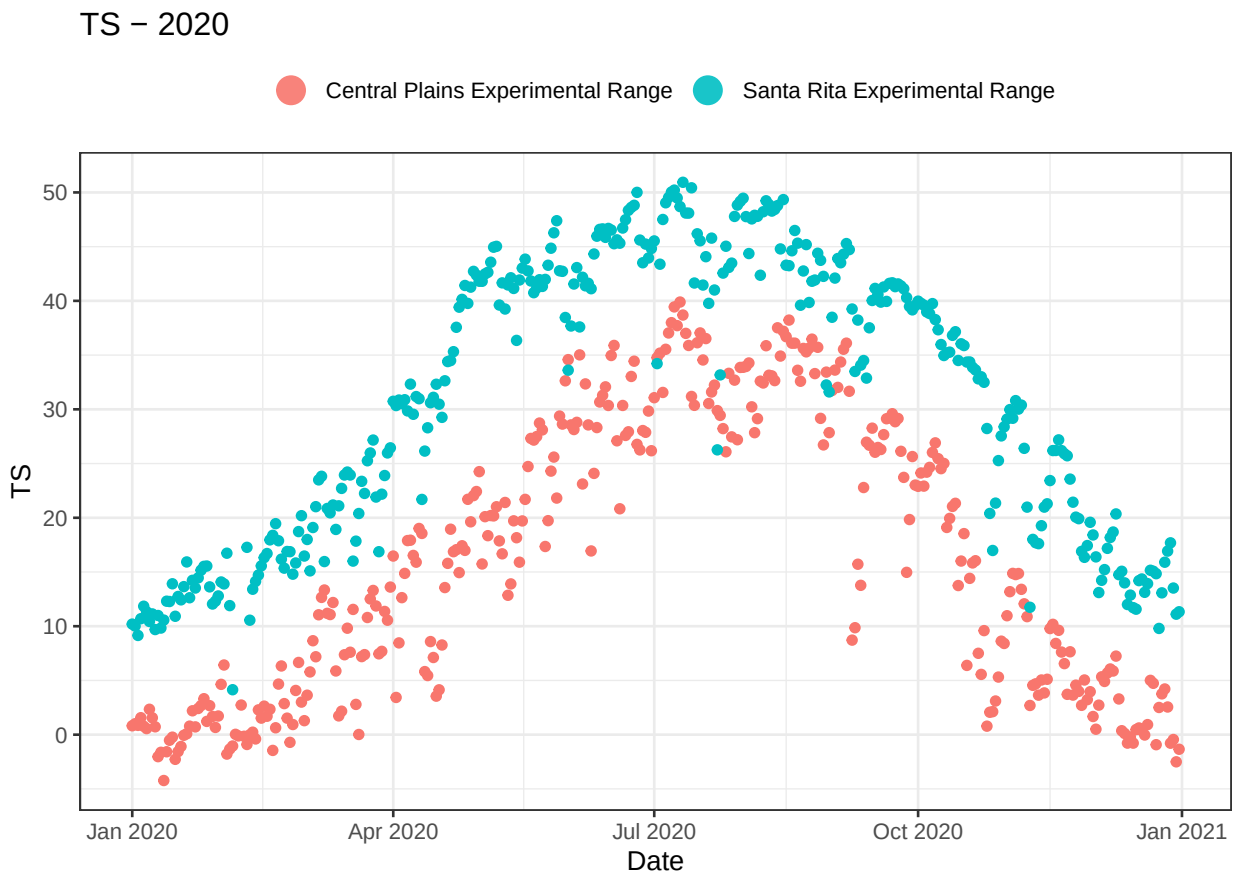
Using the daily data. Select a single year of data for your site. Plot NEE, GPP, Re, soil temperature, air temperature, and soil water content against Day of Year.

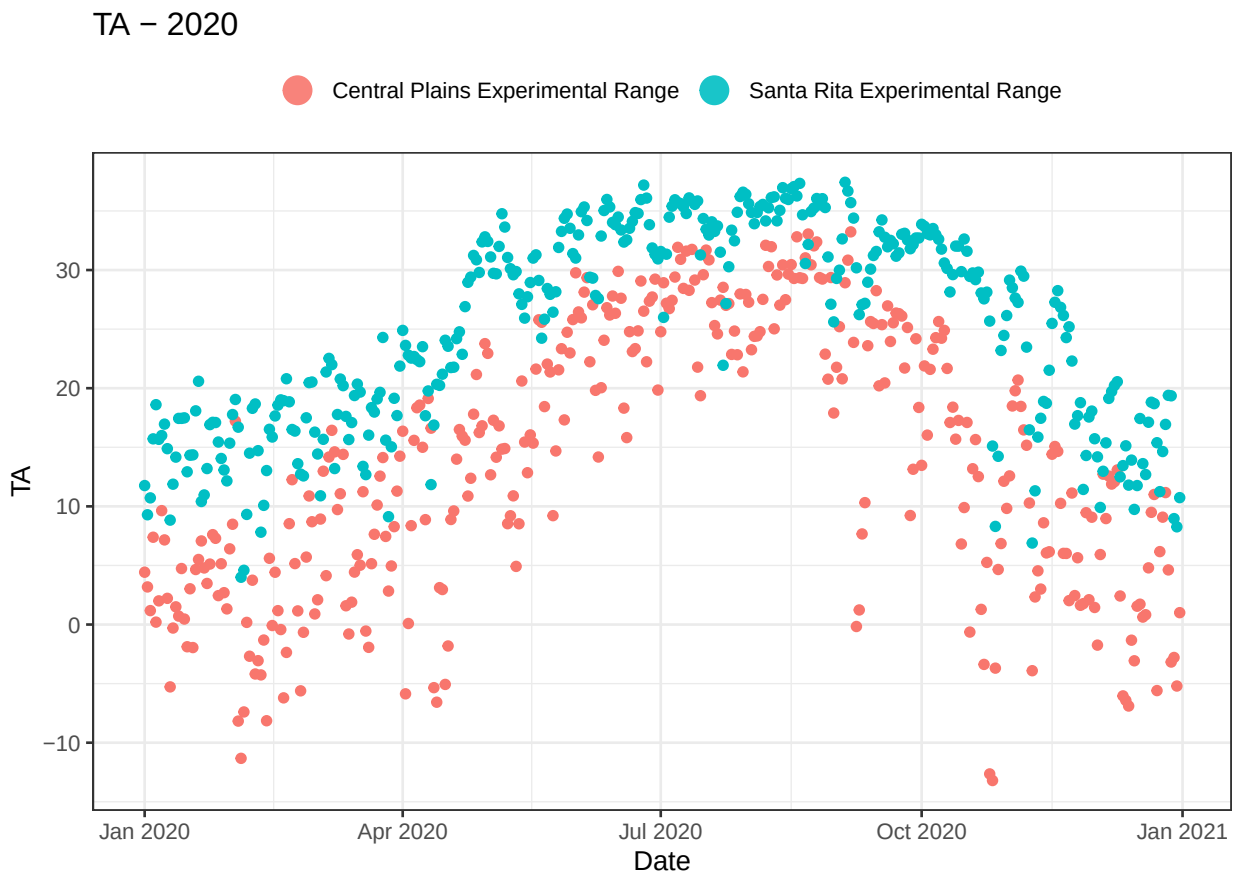
```
p_fn <- function(my_var, my_year) {
  my_title <- paste0(toupper(my_var), " - ", as.character(my_year))
  #plot
  (
    dta_1day %>%
    dplyr::filter(
      year == my_year
    ) %>%
    ggplot(., aes_string(x = "date_id", y = my_var, color = "neon_site_name")) +
      geom_point() +
      xlab("Date") +
      ylab(toupper(my_var)) +
      labs(
        title = my_title
        # , subtitle = site
      ) +
      scale_fill_brewer(type = "qual", palette = "Dark2", direction = -1) +
      theme_bw() +
      theme(
        legend.position = "top"
        , legend.direction = "horizontal"
        , legend.title = element_blank()
      ) +
      guides(color = guide_legend(override.aes = list(size=5, alpha = 0.9)))
  )

}
# vars
v <- c("nee", "gpp", "re", "ts", "ta", "swc")
for (i in 1:length(v)) {
  print( p_fn(my_var = v[i], my_year = 2020) )
}
```

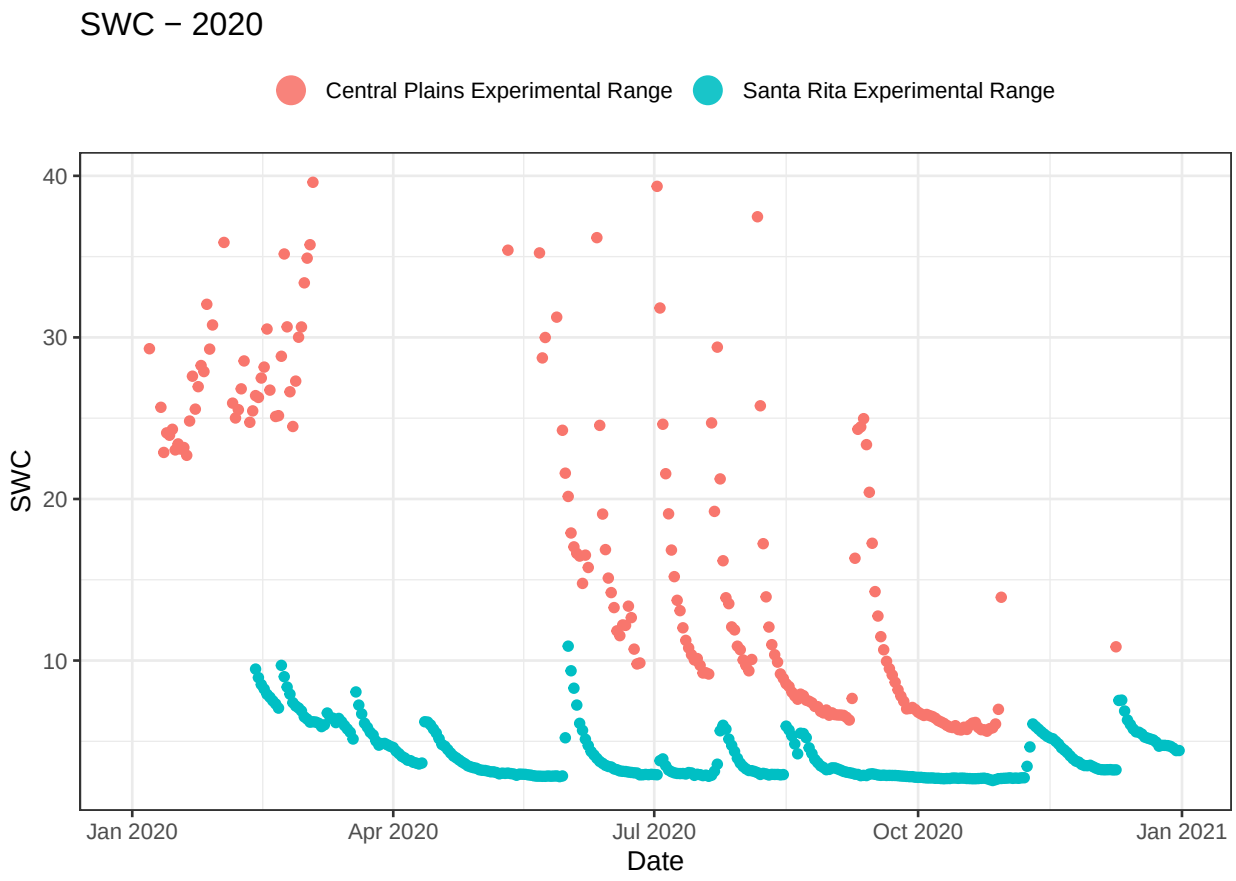

NEE − 2020
Central Plains Experimental Range
Santa Rita Experimental Range
0.20
0.15
0.10
0.05
0.00
NEE
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date


GPP − 2020
Central Plains Experimental Range
Santa Rita Experimental Range
0.15
0.10
0.05
0.00
−0.05
−0.10
GPP
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date


RE − 2020
Central Plains Experimental Range
Santa Rita Experimental Range
RE
0.20
0.15
0.10
0.05
0.00
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date


TS − 2020
Central Plains Experimental Range
Santa Rita Experimental Range
TS
50
40
30
20
10
0
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date


TA − 2020
Central Plains Experimental Range
Santa Rita Experimental Range
TA
30
20
10
0
−10
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date



a.

Describe the annual patterns in each plot and what you think drives them.

b.

Compare annual values of NEE at your site vs. CPER. Which exchanges more carbon? Why do you think that is?