ECOL 610: NEON Data - Summary Data

Group - Santa Rita Experimental Range (SRER)

26 October, 2022

Contents

Group Members

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

Objective

1) Pull in inorganic N data from OneDrive (folder is called Summarized InorgN) for your site. If data is not available at your site (Toolik and Konza), use CPER data.

2) Download the total N data from NEON in the MegaPit data. Easiest way to do this is by searching MegaPit on the NEON data portal and then choosing your site in the download. Let Katie know if you need help!

3) Plots over time of nitrate, ammonium, and total inorganic N (TIN) concentrations and the ratio of nitrate to TIN.

4) Calculate mean TIN and mean TIN/(total N (from megapit data) + TIN)

Setup

Load in the needed packages

```
library(tidyverse)
library(lubridate)
library(viridis)
library(RColorBrewer)
library(scales)
library(latex2exp)
library(kableExtra)
remove(list=ls())
# what is your site name?
site <- "Santa Rita Experimental Range"
```
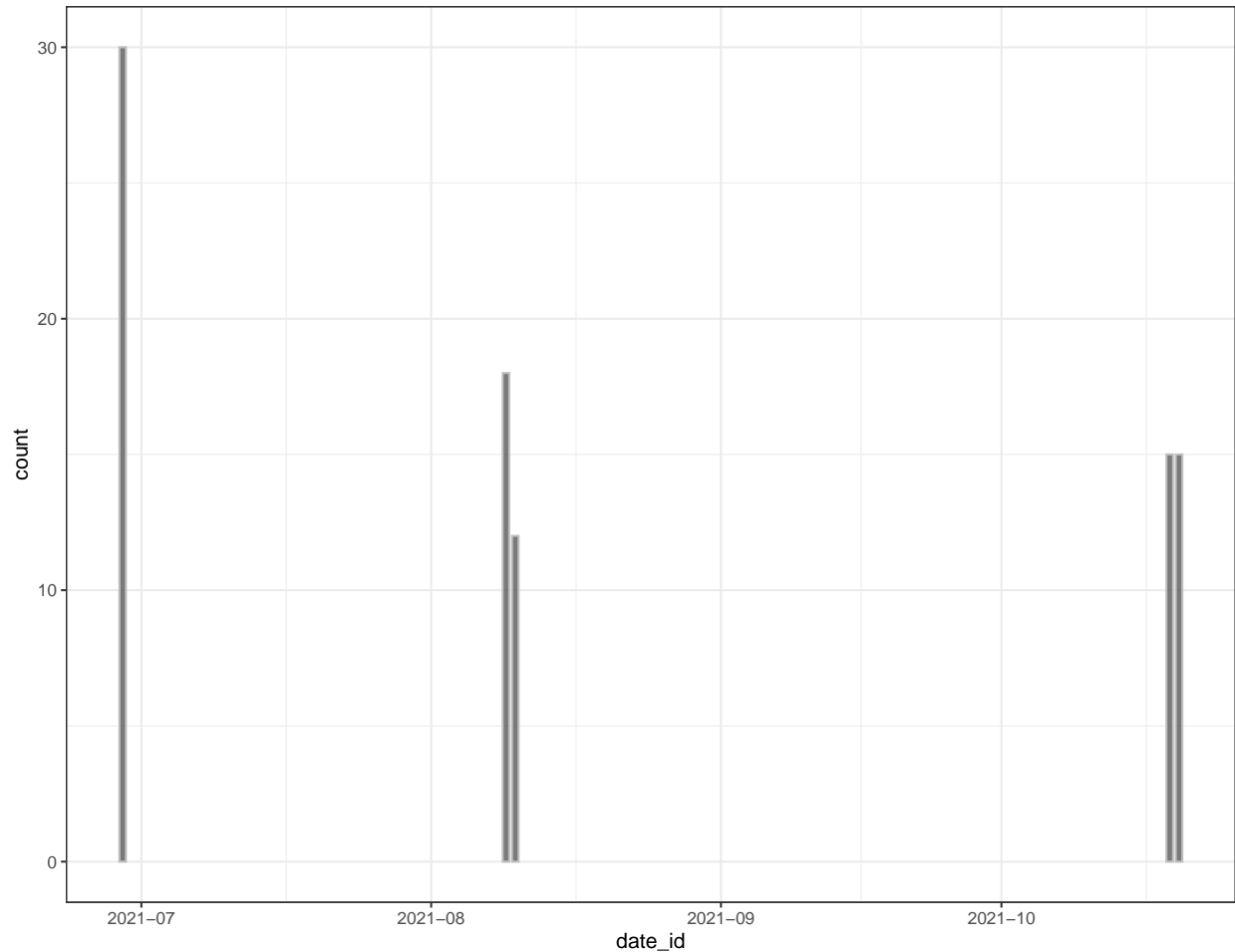
Question 1

Pull in inorganic N data from OneDrive (folder is called Summarized InorgN) for your site. If data is not available at your site (Toolik and Konza), use CPER data.

```
# inorganic n
dta_inorganic_n <- read.csv("../data/SRER_2021_N.csv") %>%
  dplyr::rename_with(~ tolower(
    gsub(" ", "_",
      stringr::str_trim(gsub("\\s+", " ", .x))
    )
  )) %>%
  dplyr::mutate(
    neon_site_name = site
    , datetime_id = lubridate::ymd_hms(collectdate)
    , date_id = lubridate::as_date(datetime_id)
    , month = lubridate::month(date_id)
    , season =
        dplyr::case_when(
          month %in% c(1:2, 11:12) ~ "Winter"
          , month %in% c(3:5) ~ "Spring"
          , month %in% c(6:8) ~ "Summer"
          , month %in% c(9:10) ~ "Fall"
          , TRUE  ~ "Other")
  ) %>%
  dplyr::relocate(neon_site_name) %>%
  dplyr::select(-c(x)) %>%
  dplyr::arrange(neon_site_name, date_id)
# quick view data
dta_inorganic_n %>%
  dplyr::glimpse()
```

```
## Rows: 90
## Columns: 13
## $ neon_site_name            <chr> "Santa Rita Experimental Range", "Santa R~
## $ sampleid                  <chr> "SRER_001-M-17.5-8.5-20210629", "SRER_001~
## $ collectdate               <chr> "2021-06-29 16:02:00", "2021-06-29 16:28:~
## $ incubationpairid          <chr> "SRER_001-M-17.5-8.5", "SRER_001-M-20.5-2~
## $ soilammoniumnugpergram    <dbl> 4.566, 2.298, 3.216, 1.357, 0.855, 0.993,~
## $ soilnitratenitritenugpergram <dbl> 3.010, 9.977, 17.872, 6.644, 2.384, 5.946~
## $ soilinorganicnugpergram   <dbl> 7.576, 12.275, 21.089, 8.001, 3.239, 6.93~
## $ netnminugpergramperday    <dbl> 0.667, 0.786, 0.539, 0.171, 0.226, 0.126,~
## $ netnitugpergramperday     <dbl> 0.849, 0.786, 0.484, 0.090, 0.245, 0.108,~
## $ datetime_id               <dttm> 2021-06-29 16:02:00, 2021-06-29 16:28:00~
## $ date_id                   <date> 2021-06-29, 2021-06-29, 2021-06-29, 2021~
## $ month                     <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,~
## $ season                    <chr> "Summer", "Summer", "Summer", "Summer", "~
```

Plot # obs per day

```
ggplot(data = dta_inorganic_n) +
  geom_bar(mapping = aes(x = date_id), width = 0.7, color = "gray", alpha = 0.8) +
  scale_x_date(date_breaks = "1 month", date_labels = "%Y-%m") +
  theme_bw()
```

There are only *5* days with data in this data set. How should the metric `netnitugpergramperday` (or any "per day" metric) be interpreted when there are multiple observations per day?

Question 2

Download the total N data from NEON in the MegaPit data. Easiest way to do this is by searching MegaPit on the NEON data portal and then choosing your site in the download. Let Katie know if you need help!

Megapit Description: Soil taxonomy, horizon names, horizon depths, soil bulk density, texture, and chemical properties in the <= 2 mm soil fraction for each soil horizon in each Megapit. Data were derived from a sampling location expected to be representative of the area where the Instrumented Soil Plots are located and were collected once during site construction, in collaboration with the USDA Natural Resources Conservation Service.

The Megapit data package content includes: * mgp_perhorizon: Per soil horizon metadata * mgp_permegapit: Data collected per megapit * mgp_perarchivesample: Archived sample information * mgp_perbiogeosample: Data collected on biogeochemistry sample * mgp_perbulksample: Data collected on bulk density sample * variables: Description and units for each column of data in data tables * readme: Data product description, issue log, and other metadata about the data product * validation: Description of data validation applied at the points of collection and ingest

```
# files available in the download include:
dir_megapit <- list.dirs("../data/NEON_soil-megapit", recursive = FALSE)[1]
fls_megapit <- list.files(path = dir_megapit)
fls_megapit %>%
  kableExtra::kable(
    caption = "Megapit data files"
    , col.names = c(
      "file name"
      )
  ) %>%
  kableExtra::kable_styling(font_size = 11) %>%
  # kableExtra::column_spec(1, bold = TRUE, width = "18em") %>%
  kableExtra::kable_styling(latex_options = "HOLD_position")
```

Table 1: Megapit data files

file name
NEON.D14.SRER.DP0.00096.001.validation.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.EML.20131104-20131104.20220120T173946Z.xml
NEON.D14.SRER.DP1.00096.001.mgp_perarchivesample.2013-11.basic.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.mgp_perbiogeosample.2013-11.basic.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.mgp_perbulksample.2013-11.basic.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.mgp_perhorizon.2013-11.basic.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.mgp_permegapit.2013-11.basic.20211220T234810Z.csv
NEON.D14.SRER.DP1.00096.001.readme.20220120T173946Z.txt
NEON.D14.SRER.DP1.00096.001.variables.20211220T234810Z.csv

Question 3

Plots over time of nitrate, ammonium, and total inorganic N (TIN) concentrations and the ratio of nitrate to TIN.

Summarize data to daily level

```
# vars to aggregate
summary_vars <- c(
  "soilammoniumnugpergram"
  , "soilnitratenitritenugpergram"
  , "soilinorganicnugpergram"
)
# aggregate data to daily level
# named list of functions
summary_fns <- list(
  mean = ~mean(.x, na.rm = TRUE)
  , median = ~median(.x, na.rm = TRUE)
  , min = ~min(.x, na.rm = TRUE)
  , max = ~max(.x, na.rm = TRUE)
  , N = ~sum(ifelse(is.na(.x), 0, 1))
```

```r
)
dta_inorganic_n_summ <- dta_inorganic_n %>%
  dplyr::group_by(neon_site_name, date_id) %>%
  dplyr::summarise(
    dplyr::across(
      tidyselect::all_of(summary_vars)
      , summary_fns
      , .names = "{.col}_{.fn}"
    )
  ) %>%
  dplyr::ungroup() %>%
  dplyr::group_by(neon_site_name, date_id)
# data for plots
dta_temp <- dta_inorganic_n_summ %>%
  dplyr::select(neon_site_name, date_id, tidyselect::ends_with("_median")) %>%
  dplyr::rename_with(function(x){gsub("_median","",x)}, tidyselect::ends_with("_median")) %>%
  # total inorganic N (TIN) concentrations and the ratio of nitrate to TIN
  dplyr::mutate(
    tin = soilammoniumnugpergram + soilnitratenitritenugpergram
    , ammonium_pct = soilammoniumnugpergram / tin * 100
    , nitrate_pct = soilnitratenitritenugpergram / tin * 100
  )
  # tidyr::pivot_longer(
  #   cols = -tidyselect::all_of(c("neon_site_name", "date_id"))
  #   , names_to = "var_name"
  #   , values_to = "var_value"
  #   , values_drop_na = FALSE
  # ) %>%
  # dplyr::mutate(
  #   var_name = gsub("_median", "", var_name)
  # ) %>%
```

Plot median values per day

```r
vars_temp <- c(
  "soilammoniumnugpergram"
  , "soilnitratenitritenugpergram"
  , "tin"
  , "ammonium_pct"
  , "nitrate_pct"
)
vars_lab_temp <- c(
  latex2exp::TeX("Soil Ammonium ($\\nu g \\cdot gram^{-1}$)")
  , latex2exp::TeX("Soil Nitrate ($\\nu g \\cdot gram^{-1}$)")
  , latex2exp::TeX("Tot. Inorganic Nitrogen ($\\nu g \\cdot gram^{-1}$)")
  , "Ammonium %"
  , "Nitrate %"
)
# define plot function
  p_fn <- function(my_var) {
  #plot
```

```
  (
    dta_temp %>%
    dplyr::filter(
      neon_site_name == site
    ) %>%
    ggplot(., aes_string(x = "date_id", y = my_var)) +
      geom_col(
        width = 0.7
        , alpha = 0.8
        , fill = RColorBrewer::brewer.pal(name = "Dark2", n = length(vars_temp))[which(vars_temp==my_var
      ) +
      xlab("Date") +
      ylab(vars_lab_temp[which(vars_temp == my_var)]) +
      labs(
        title = site
        , subtitle = vars_lab_temp[which(vars_temp == my_var)]
      ) +
      scale_y_continuous(breaks = scales::extended_breaks(n = 5)) +
      scale_x_date(date_breaks = "1 month", date_labels = "%Y-%m") +
      theme_bw() +
      theme(
        legend.position = "none"
        , axis.text.x = element_text(angle = 60, hjust=1)
      )
  )

}
# call function
vars_temp %>%
  purrr::map(p_fn)
```
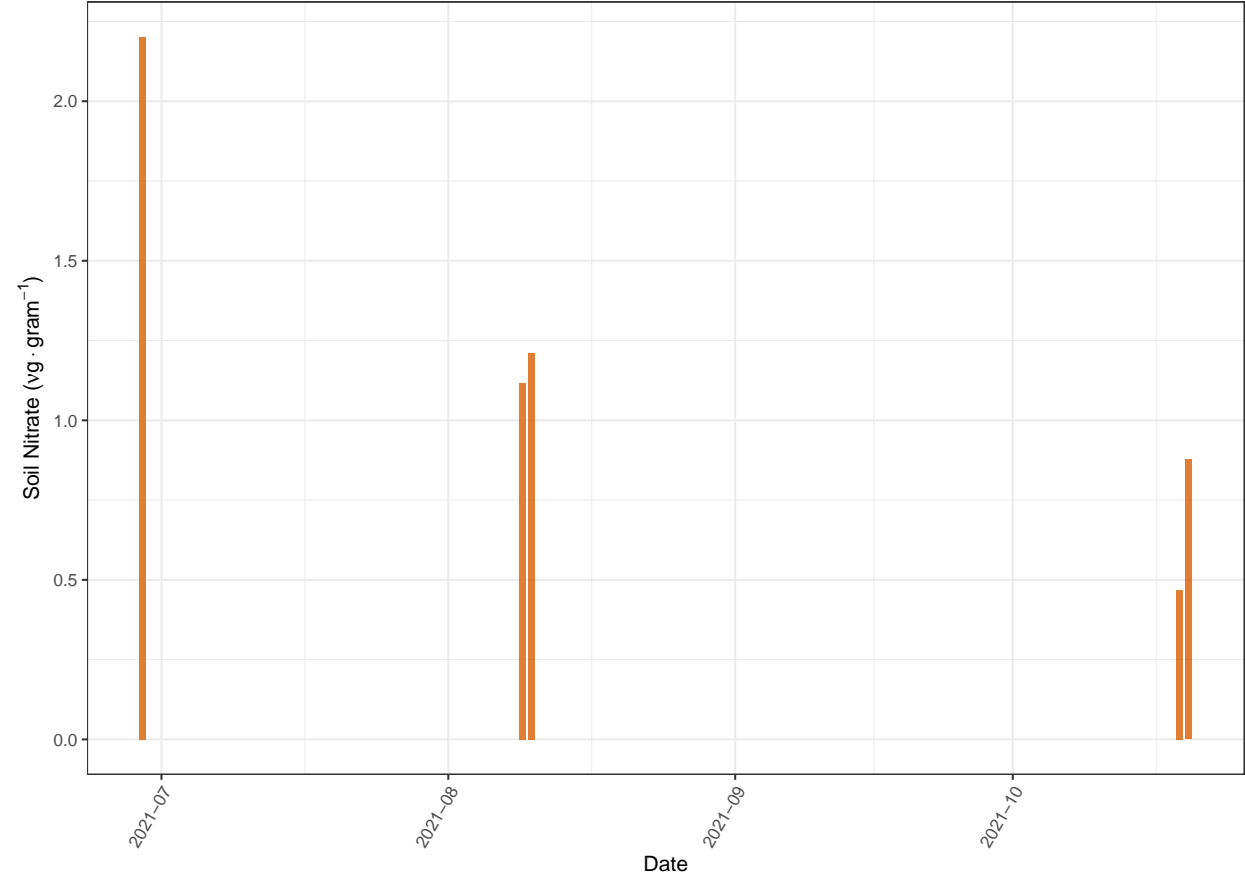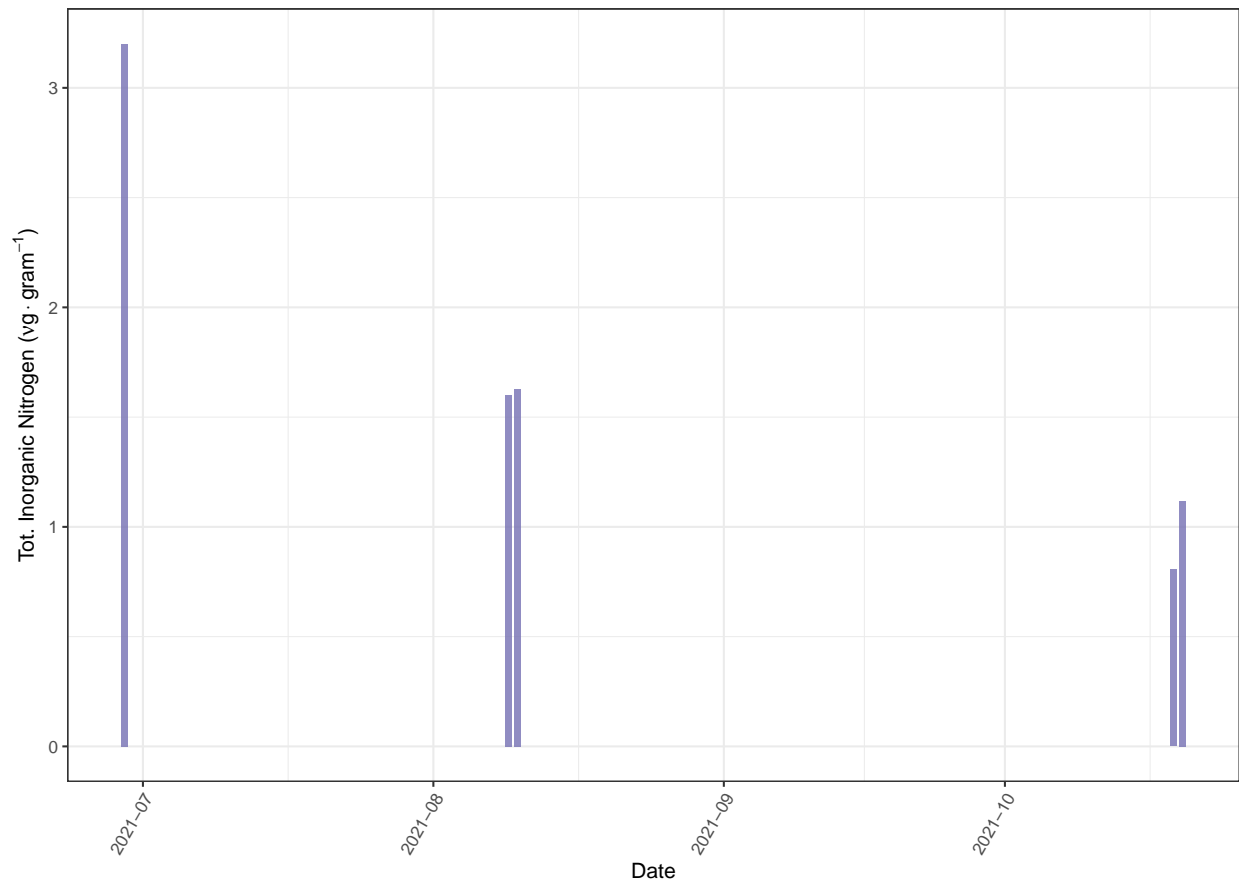
Santa Rita Experimental Range

Soil Ammonium (vg · gram^{-1})
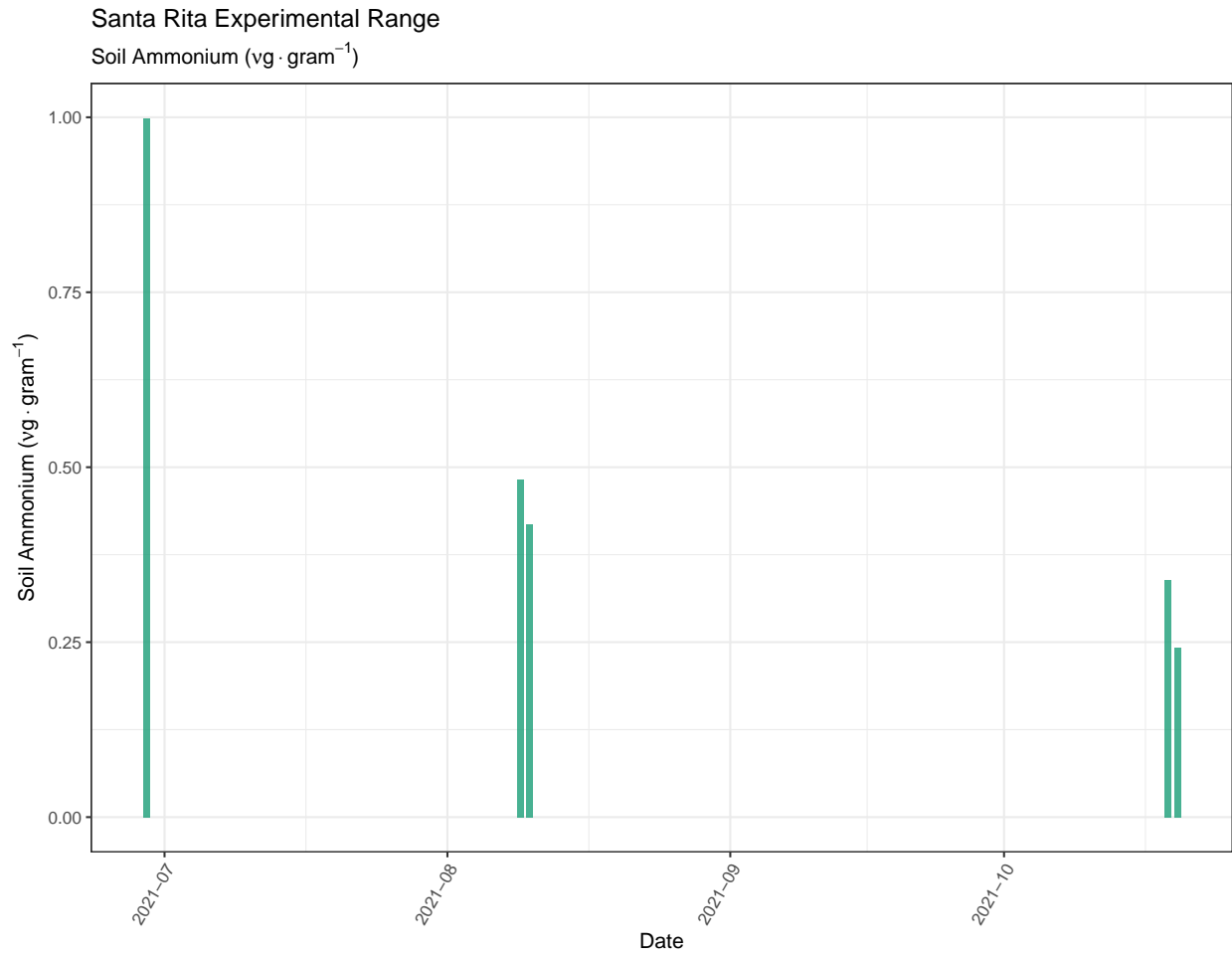
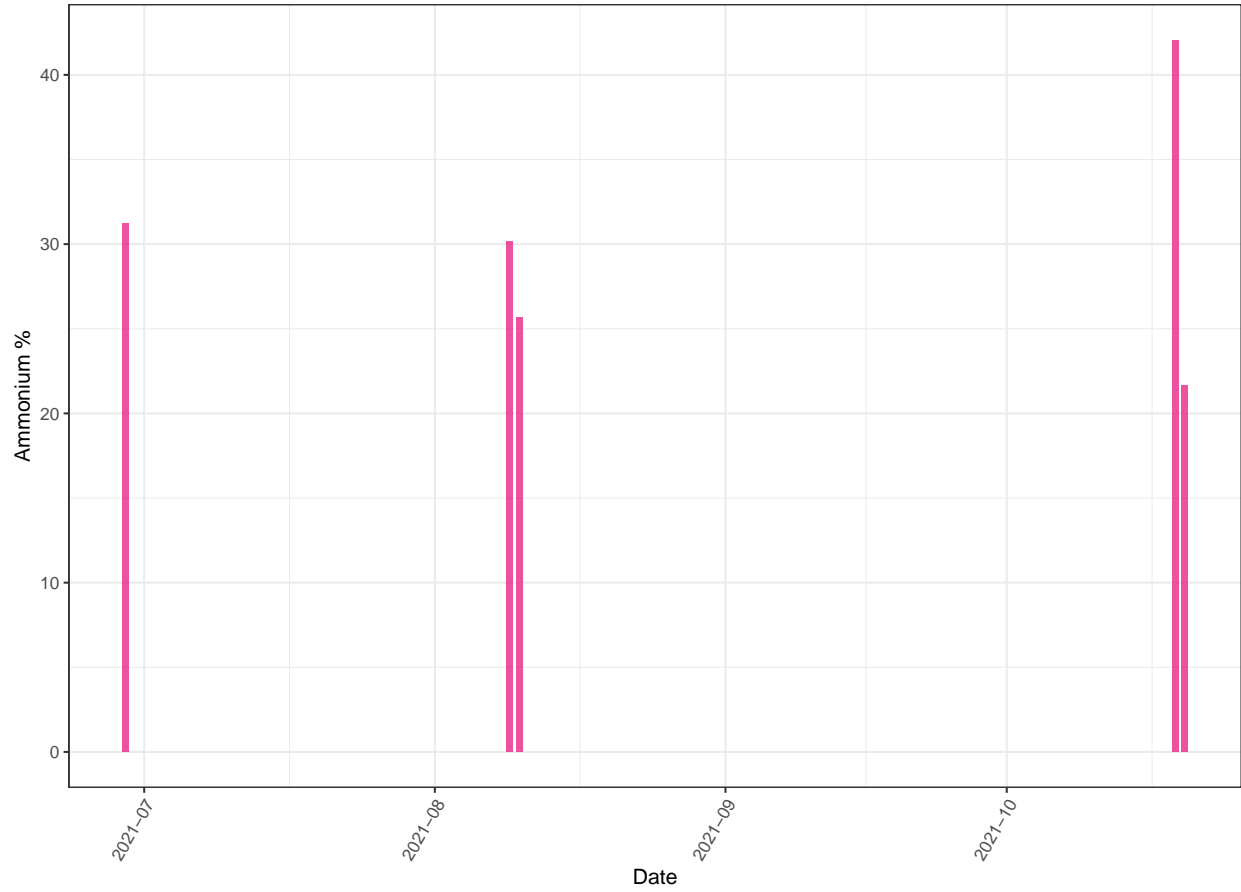## Santa Rita Experimental Range

Soil Nitrate (vg · gram^{-1})

Santa Rita Experimental Range

Tot. Inorganic Nitrogen (vg · gram^{-1})

Santa Rita Experimental Range

Santa Rita Experimental Range

Nitrate %
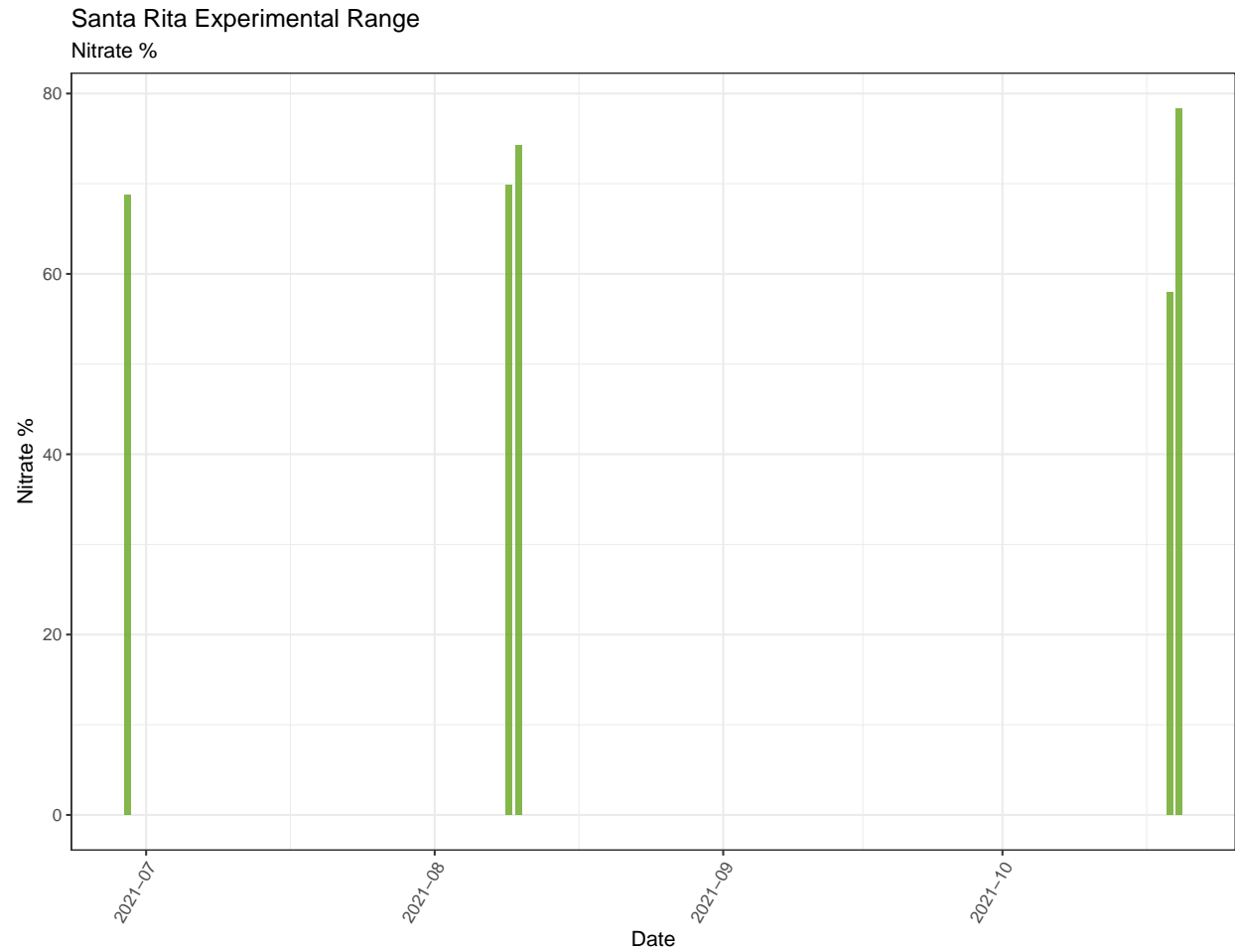
Plots above are of the median daily value of each metric based on multiple measurements per day.

Question 4

Calculate mean TIN and mean TIN/(total N (from megapit data) + TIN)

```
# in progress
```