MIVI

Will Bennett

11/28/2023

Japanese Stiltgrass (Microstegium vimineum)

Effects of Latitude and Elevation on Seeding Time

Data source: iNaturalist (link)
Total observations (NA): 15040
Observations needing phenology annotation: 11010

Setup

```
# set lwd
setwd("~/Documents/MIVI/")

library(dplyr)
library(ggplot2)
library(elevatr)

# number of groups for analysis
n = 60
```

Loading data

Export iNaturalist Data

1. Export all *Microstegium vimineum* observations from iNat with columns (`id`, `observed_on`, `latitude`, `longitude`, `place_state_name`, `place_country_name`): Link

2. Export all *M. vimineum* observations with phenology 'No Evidence of Flowering' with column `id`: Link

3. Export all *M. vimineum* observations with phenology 'Flowering' with column `id`: Link

4. Export all *M. vimineum* observations with phenology 'Fruiting' with column `id`: Link

```
# 1
mivi_all <- read.csv("./MIVI-ALL.csv") %>%
    mutate(date=as.Date(observed_on, format="%Y-%m-%d")) %>% select(-observed_on) %>%
    select(-place_country_name)  # not really useful
```

```r
# 2
mivi_young <- read.csv("MIVI-YOUNG.csv") %>%
    mutate(stage="Vegetation")

# 3
mivi_flowering <- read.csv("./MIVI-FLOWERING.csv") %>%
    mutate(stage="Flowering")

# 4
mivi_fruiting <- read.csv("./MIVI-FRUITING.csv") %>%
    mutate(stage="Seeding")


# join each based on id
mivi_all <- mivi_all %>% left_join(mivi_young, by="id")
mivi_all <- mivi_all %>% left_join(mivi_flowering, by="id") %>%
    mutate(stage = coalesce(stage.x, stage.y)) %>% select(-stage.x, -stage.y)
mivi_all <- mivi_all %>% left_join(mivi_fruiting, by="id") %>%
    mutate(stage = coalesce(stage.x, stage.y)) %>% select(-stage.x, -stage.y)

# memory cleanup
rm(mivi_young, mivi_flowering, mivi_fruiting)
```

Or load from processed file

```r
mivi_all <- read.csv("./MIVI-PROCESSED.csv") %>% select(-X) %>%
    mutate(date=as.Date(date, format="%Y-%m-%d"))
```

Data processing

Todo: remove known problem records
130398055 - not flowering,
32815927 - not flowering

```r
# Get Julian day
mivi_all <- mivi_all %>% mutate(julian = as.integer(strftime(date, format="%j")))

# make phenology a factor type (not necessary)
mivi_all$stage <- factor(mivi_all$stage, ordered=TRUE,
                         levels=c("Vegetation", "Flowering", "Seeding"))

# Select only observations with phenology data
mivi_annotated <- mivi_all %>% filter(!is.na(stage))

# Group into quartiles by latitude
mivi_annotated$group <- ntile(mivi_annotated$latitude, 4)
```
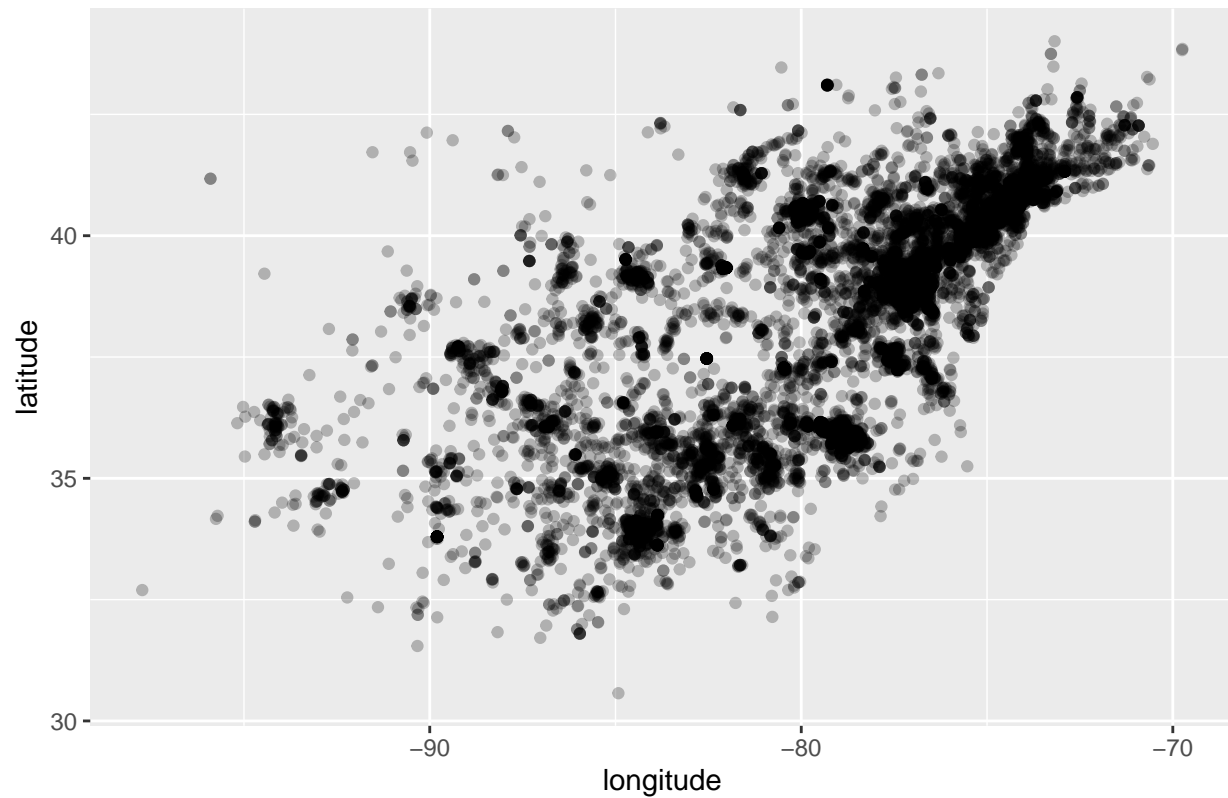
Basic Descriptive Plots

```
# Density of all observations by lat/lon
ggplot(mivi_all, aes(x=longitude,y=latitude)) + geom_point(alpha=0.25) +
    labs(title="Density of observations by location")
```

Density of observations by location


```
# Histogram of all observations by lat
ggplot(mivi_all, aes(x=latitude)) + geom_histogram(binwidth=0.2) +
    labs(title="Histogram of observations by latitude")
```
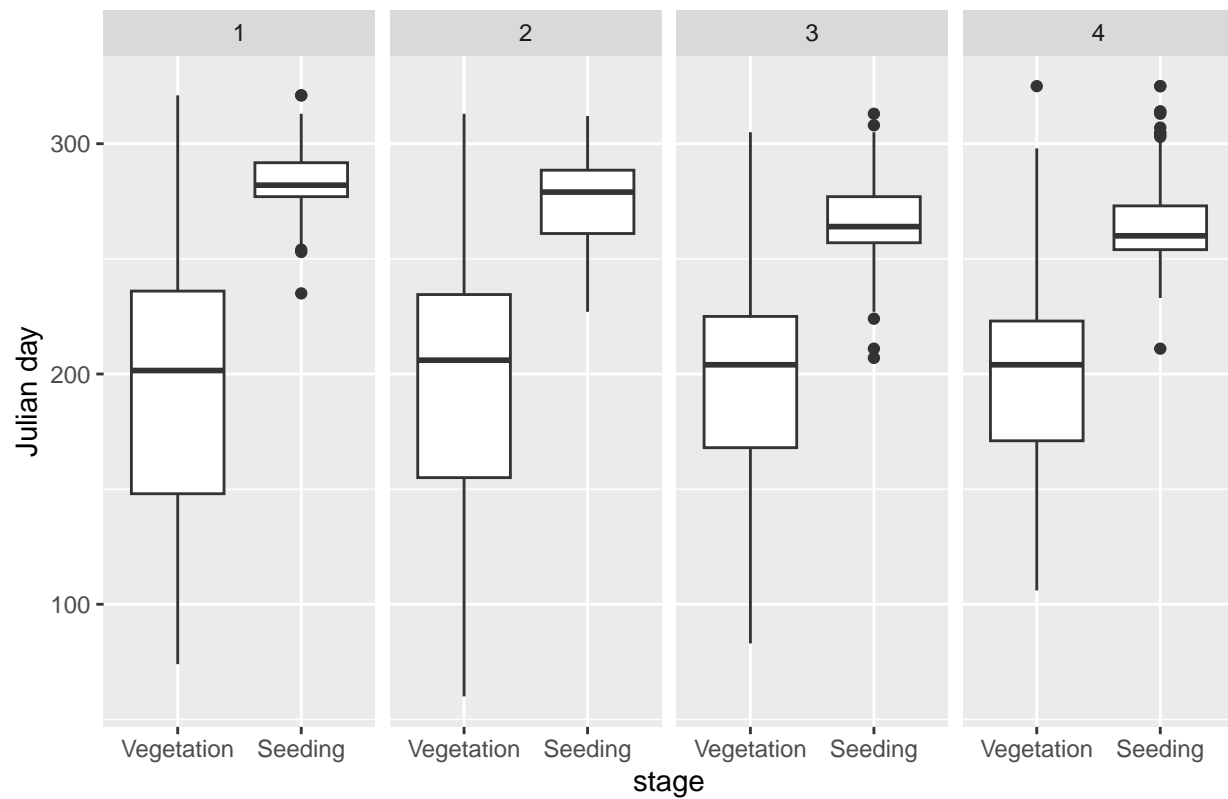
Histogram of observations by latitude
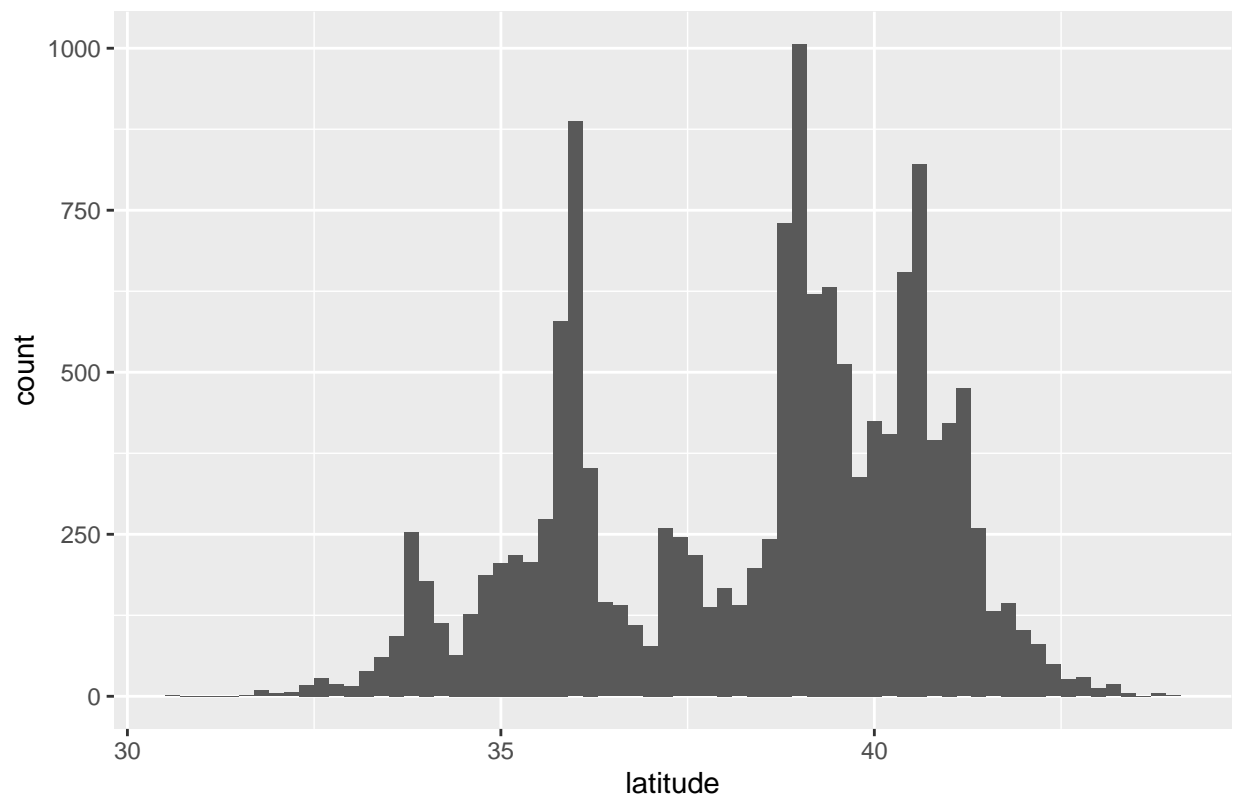

```r
# Histogram with city labels
ggplot(mivi_all, aes(x=latitude)) + geom_histogram(binwidth=0.2) +
    geom_vline(xintercept=33.75) + annotate("text", x=33.5, y=850, label="Atlanta", angle=90) +
    geom_vline(xintercept=35.84) + annotate("text", x=35.59, y=850, label="Raleigh", angle=90) +
    geom_vline(xintercept=38.9) + annotate("text", x=38.65, y=850, label="Washington", angle=90) +
    geom_vline(xintercept=40.75) + annotate("text", x=41, y=850, label="New York", angle=90) +
    ylab("count") + labs(title="Histogram of observations by latitude")
```

Histogram of observations by latitude
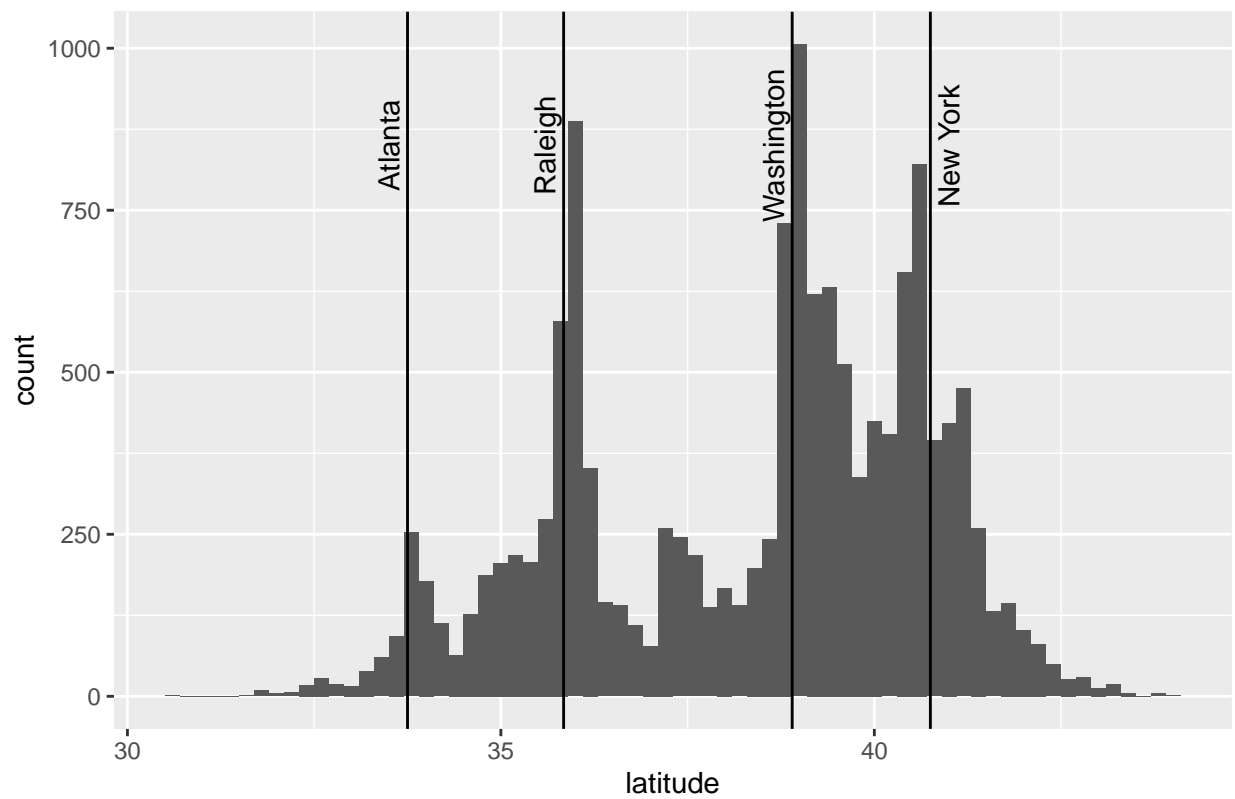

```
# Boxplots of phenology by quartile
# Slight earlier trend in Seeding visible
ggplot(mivi_annotated, aes(julian, stage)) + geom_boxplot() + facet_grid(~group) +
    coord_flip() + xlab("Julian day") + labs(title="Phenology stage, quartiles by latitude")
```
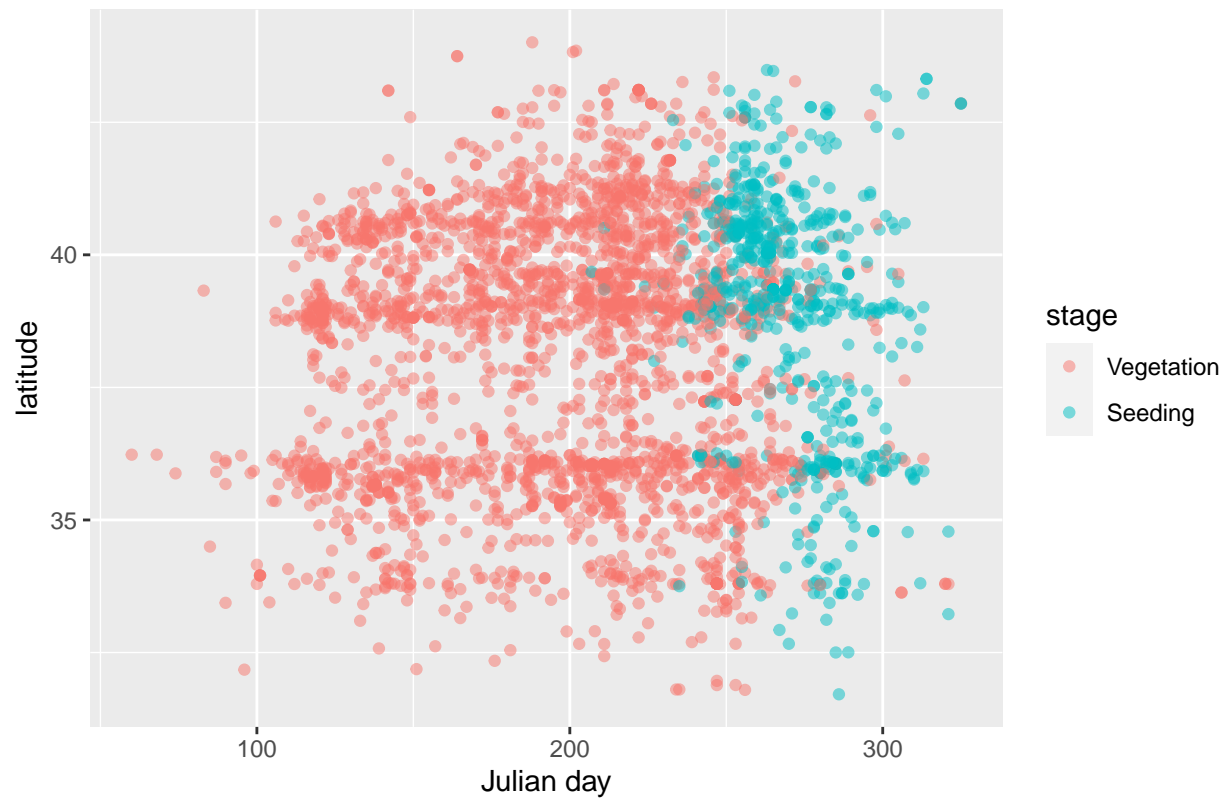
Phenology stage, quartiles by latitude


Time Series Plots

```
# Latitude against Julian day
ggplot(mivi_annotated, aes(julian, latitude)) + geom_point(aes(color=stage), alpha=0.5) +
    scale_color_hue() + xlab("Julian day") +
    labs(title="Annotated Observations by Latitude and Julian Day")
```

Annotated Observations by Latitude and Julian Day


```r
# Time series by latitude, color by stage
ggplot(mivi_all, aes(date, latitude)) + geom_point(aes(color=stage), alpha=0.5) +
    scale_x_date(date_breaks = "1 year", date_labels = "%Y") + scale_color_hue() +
    labs(title="Observations by latitude over time")
```

Observations by latitude over time
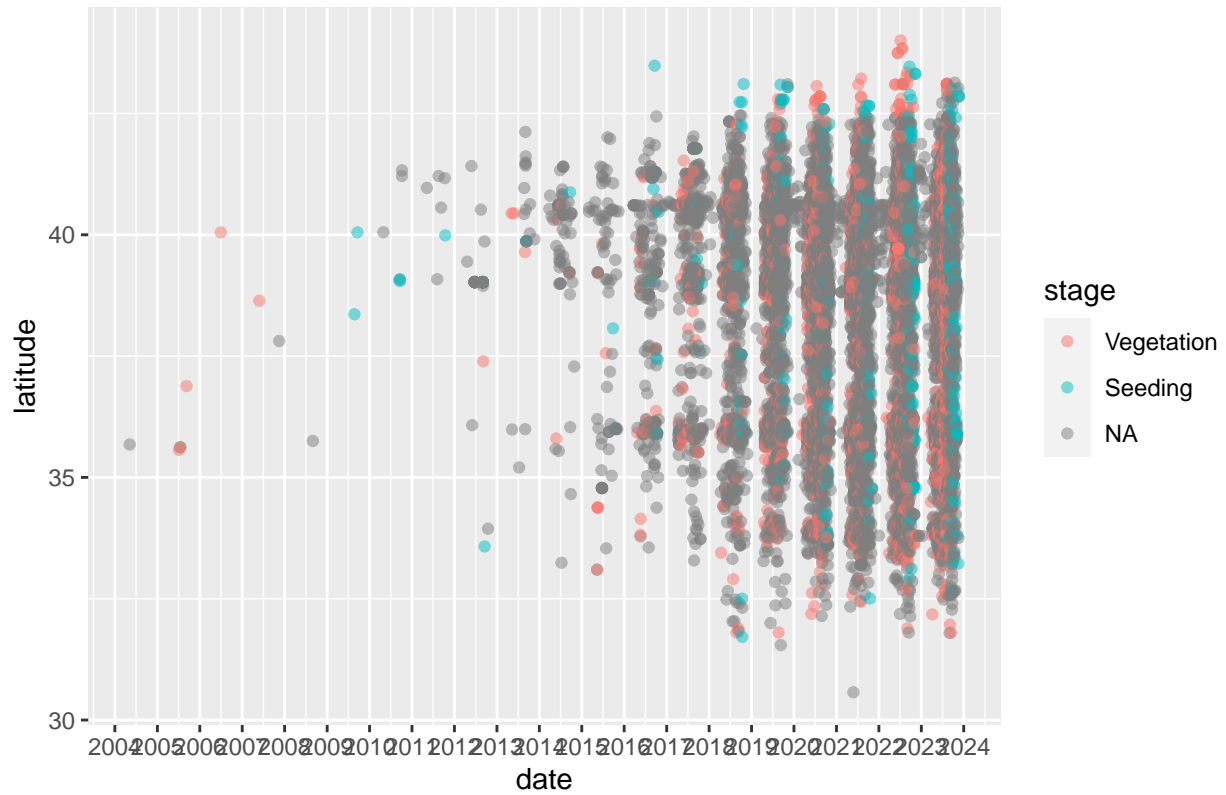

```r
# Zoom in on recent data
timeclip <- c(as.Date("2019-02-01"), as.Date("2023-11-30"))
ggplot(mivi_all, aes(date, latitude)) + geom_point(aes(color=stage), alpha=0.5) +
    scale_x_date(limits=timeclip, date_breaks = "1 year", date_labels = "%Y") + scale_color_hue() +
    labs(title="Observations by latitude over time (2019-2023)")
```

```
## Warning: Removed 1735 rows containing missing values ('geom_point()').
```

Observations by latitude over time (2019–2023)
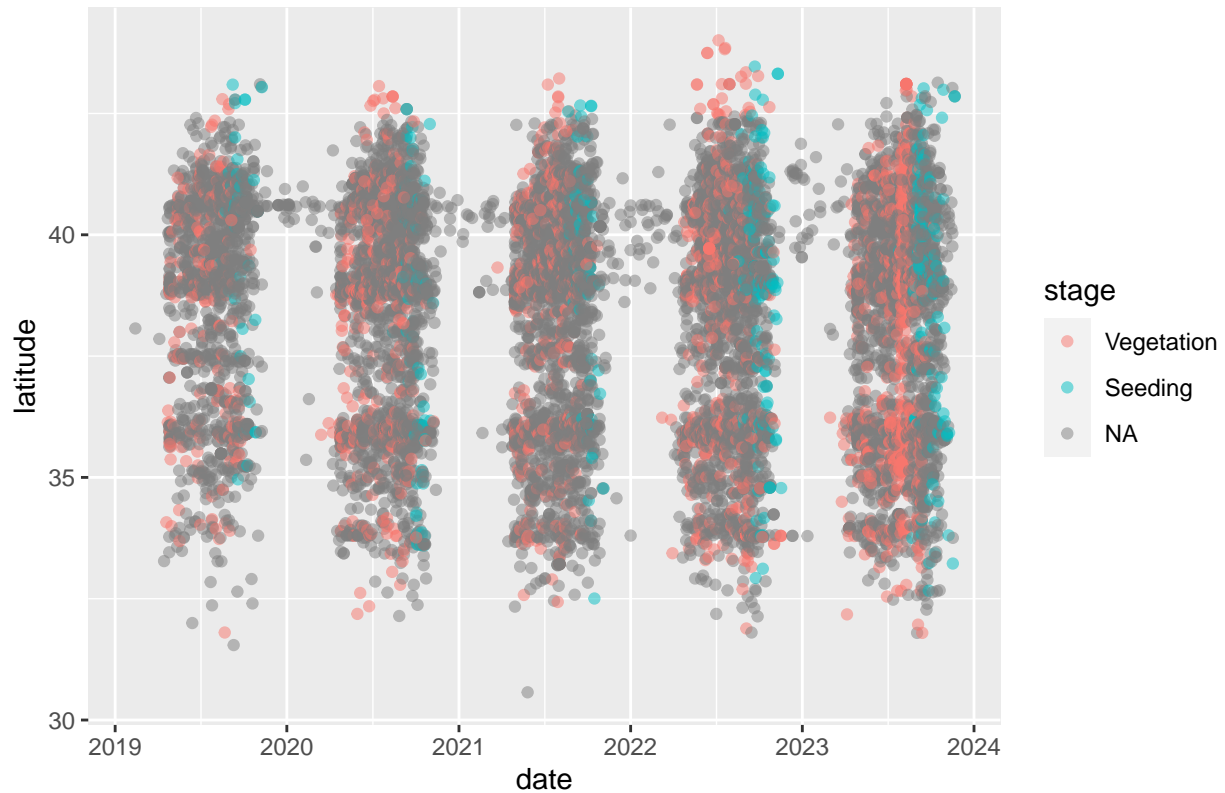
Retrieve Elevation Information

```
coords <- data.frame(x=mivi_annotate$longitude,
                     y=mivi_annotate$latitude, ele_id=mivi_annotate$id)
# use slice(1:100) to get a subset for reducing retrieval time

# retrieve elevation from USGS (takes a while)
elevations <- get_elev_point(coords, prj=4326, src="epqs")

mivi_all <- mivi_all %>% left_join(elevations, by=join_by("id" == "ele_id")) %>%
    select(-elev_units, -geometry)
rm(elevations)


write.csv(mivi_all, file="./MIVI-PROCESSED.csv", na='')
```

Analysis!

```
# Note: returns Inf if there are none in the selection
first_seed <- function(df) {
```

```r
    df <- df %>% filter(stage == "Seeding")
    return (min(df$julian, na.rm=TRUE))
}
```

Latitude

```r
# Create groups
mivi_annotated$group <- ntile(mivi_annotated$latitude, n)

# mean lat for each group
a <- mivi_annotated %>% group_by(group) %>% summarise(mlat=mean(latitude, na.rm=TRUE))

# first seeding date in each group
b <- mivi_annotated %>% group_by(group) %>% group_map(~first_seed(.x))

a <- bind_cols(a, do.call(rbind.data.frame, b)[,1]) %>% mutate(fseed = ...3) %>% select(-...3)
```
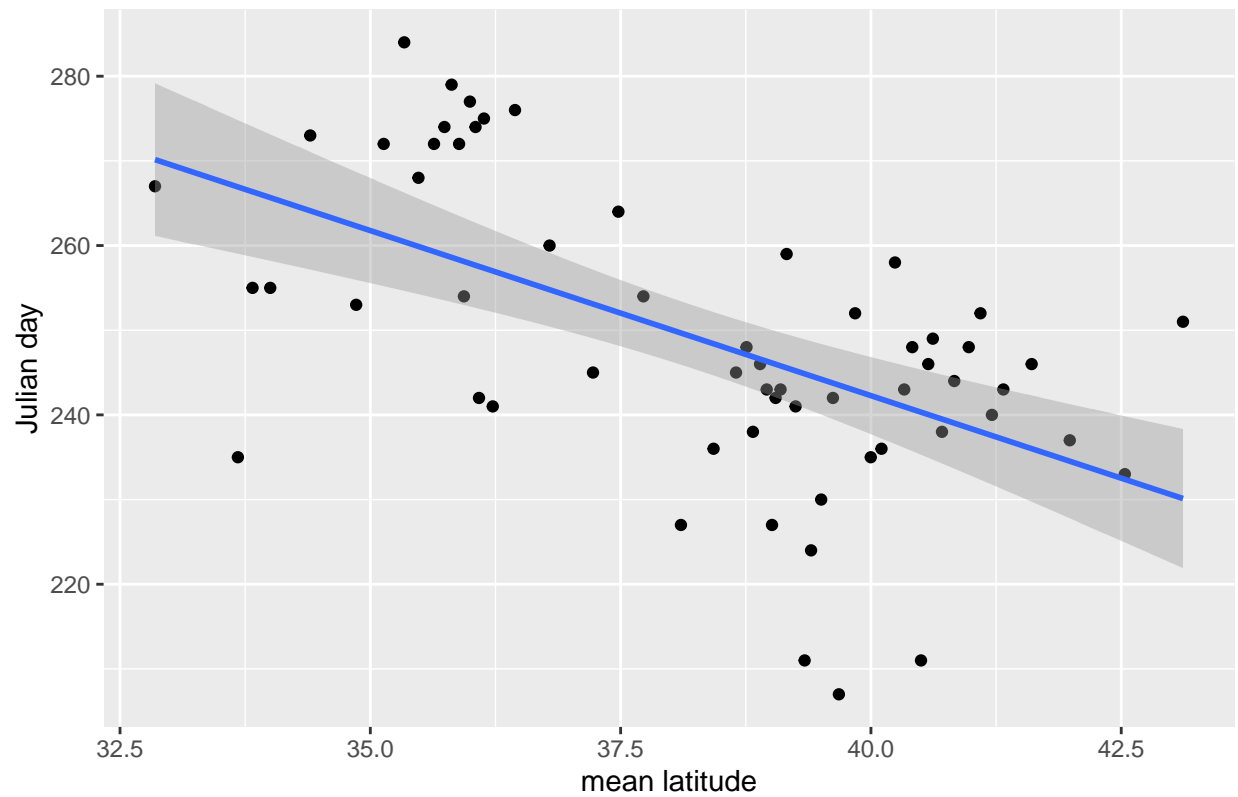
```
## New names:
## * '' -> '...3'
```

```r
a <- remove_missing(a, finite=TRUE)  # remove any Inf's

# Plot with linear model
ggplot(a, aes(mlat, fseed)) + geom_point() + geom_smooth(method='lm') +
    ylab("Julian day") + xlab("mean latitude") + labs(title="First Seeding Day by Latitude")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```
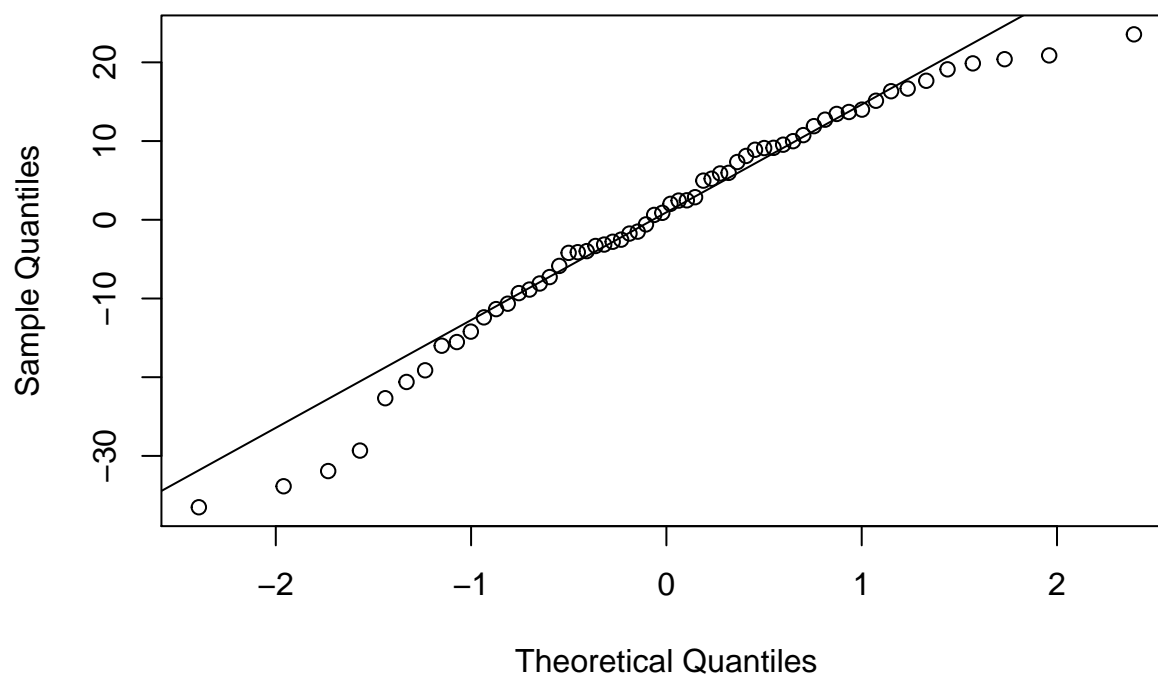
First Seeding Day by Latitude


```
# Q-Q residual plot
model_lat <- lm(fseed~mlat, data=a)
res <- resid(model_lat)
qqnorm(res)
qqline(res)
```

Normal Q–Q Plot


```r
# Pearson's correlation test
cor.test(a$fseed, a$mlat, alternative="less")
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  a$fseed and a$mlat
## t = -5.1603, df = 58, p-value = 1.568e-06
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.0000000 -0.3938359
## sample estimates:
##        cor 
## -0.5609363
```

```r
rm(a, b, res)
```

Elevation

```r
# Create groups
mivi_annotated$group <- ntile(mivi_annotated$elevation, n)

# mean lat for each group
```

```
a <- mivi_annotated %>% group_by(group) %>% summarise(mele=mean(elevation, na.rm=TRUE))

# first seeding date in each group
b <- mivi_annotated %>% group_by(group) %>% group_map(~first_seed(.x))

a <- bind_cols(a, do.call(rbind.data.frame, b)[,1]) %>% mutate(fseed = ...3) %>% select(-...3)
```
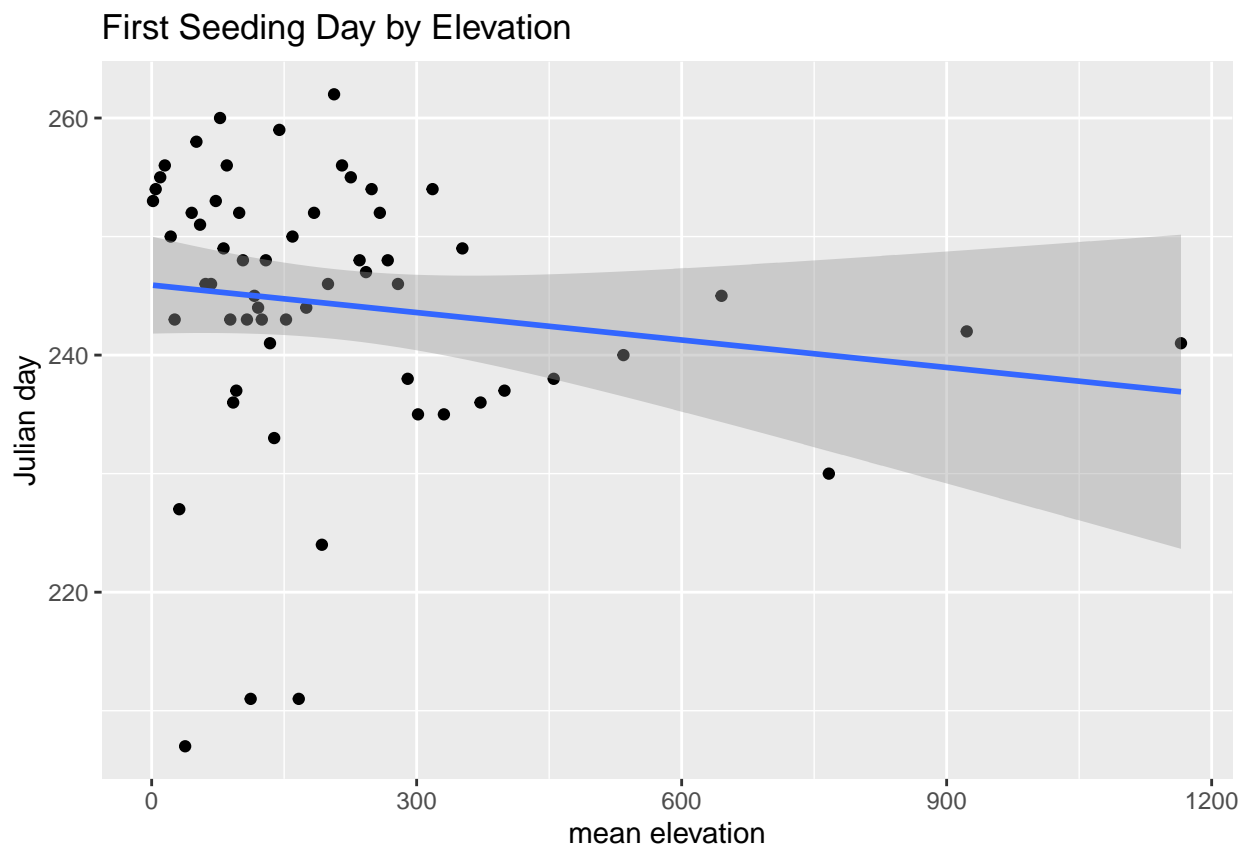
```
## New names:
## * '' -> '...3'
```

```
a <- remove_missing(a, finite=TRUE)   # remove any Inf's
```

```
## Warning: Removed 1 rows containing non-finite values.
```

```
# Plot with linear model
ggplot(a, aes(mele, fseed)) + geom_point() + geom_smooth(method='lm') +
    ylab("Julian day") + xlab("mean elevation") + labs(title="First Seeding Day by Elevation")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```
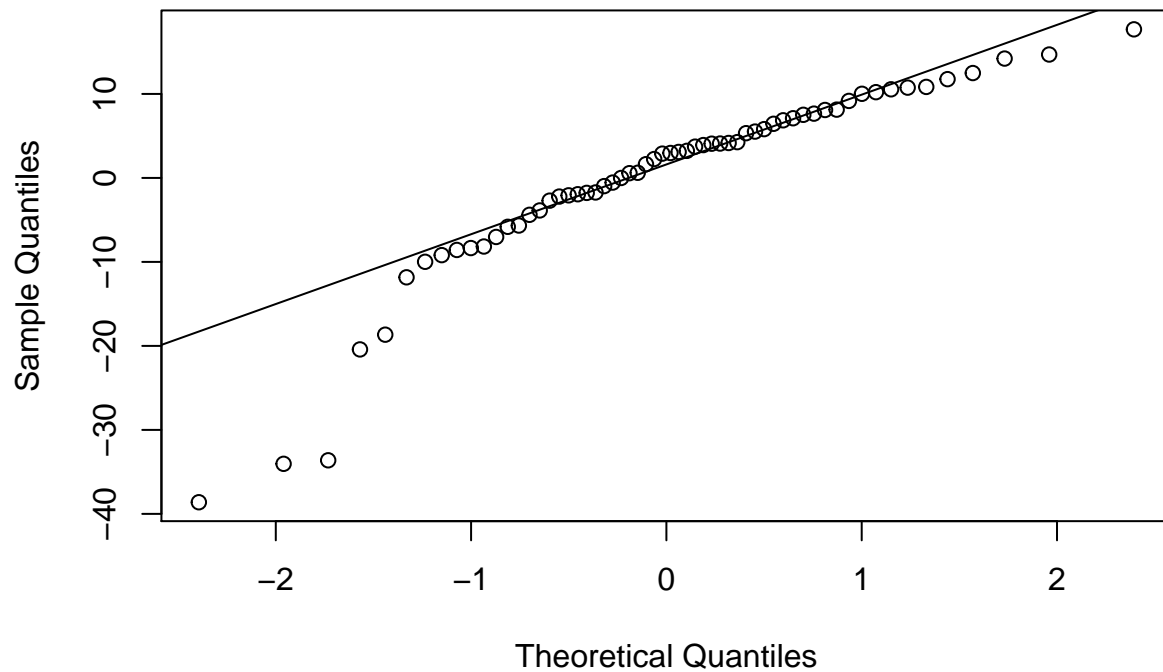


```
# Q-Q residual plot
model_ele <- lm(fseed~mele, data=a)
res <- resid(model_ele)
qqnorm(res)
qqline(res)
```

Normal Q–Q Plot


```
# Pearson's correlation test
cor.test(a$fseed, a$mele, alternative="less")
```

```
##
##   Pearson's product-moment correlation
##
## data:  a$fseed and a$mele
## t = -1.1429, df = 58, p-value = 0.1289
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.00000000  0.06824594
## sample estimates:
##        cor
## -0.1484098
```

```
rm(a, b, res)
```