MIVI

Will Bennett

12/02/2023

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

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

library(dplyr)
library(ggplot2)
library(elevatr)
library(corrtable)
library(table1)
library(knitr)

# number of groups for analysis
n = 200
```

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

# 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

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

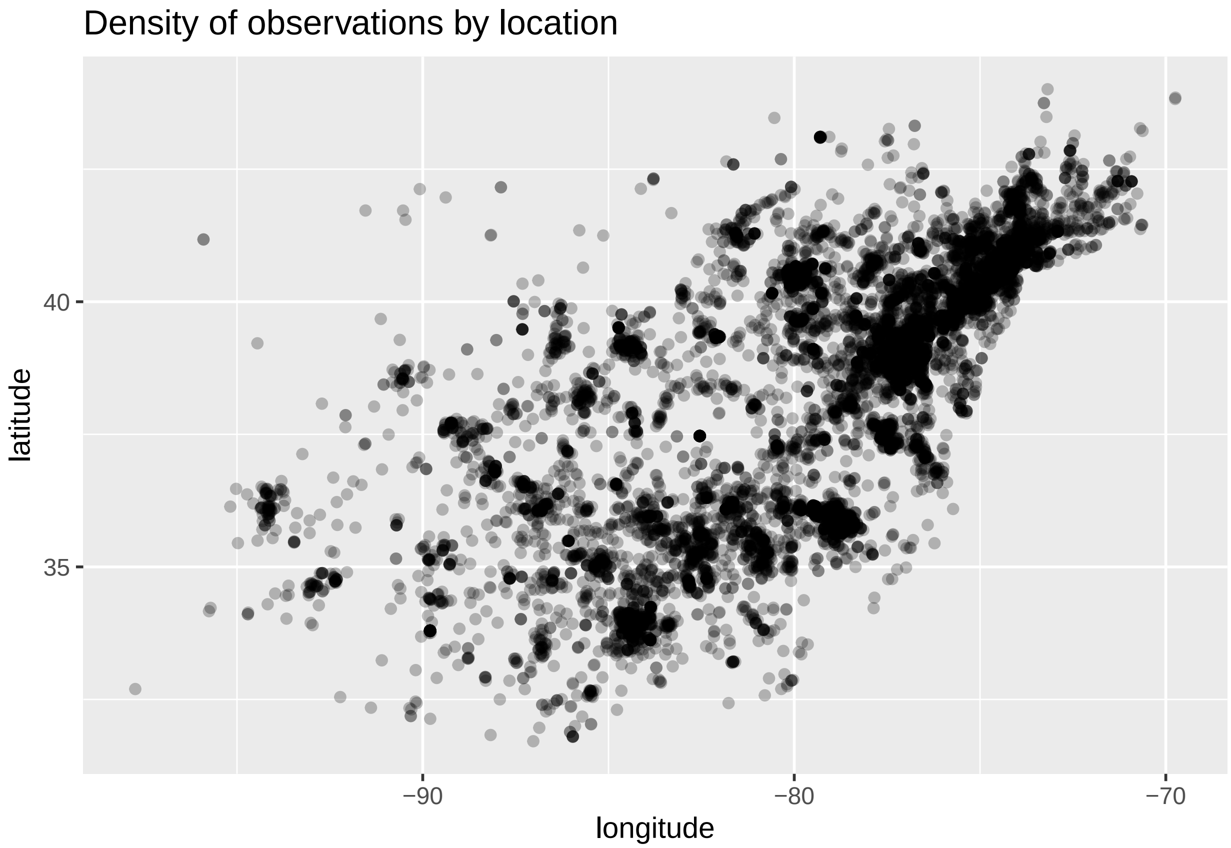
Data processing

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

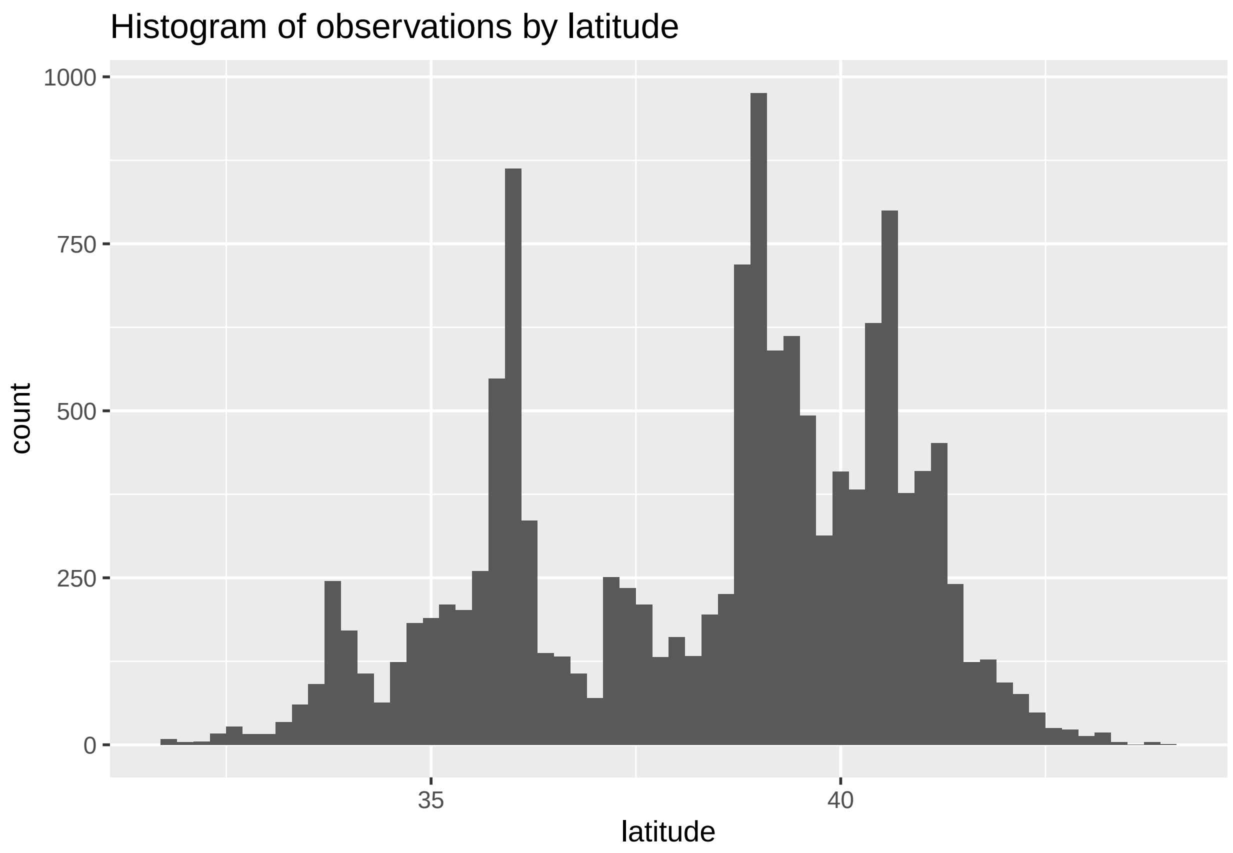
```
# Get Julian day
mivi_all <- mivi_all %>% mutate(julian = as.integer(strftime(date, format="%j")))

# 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

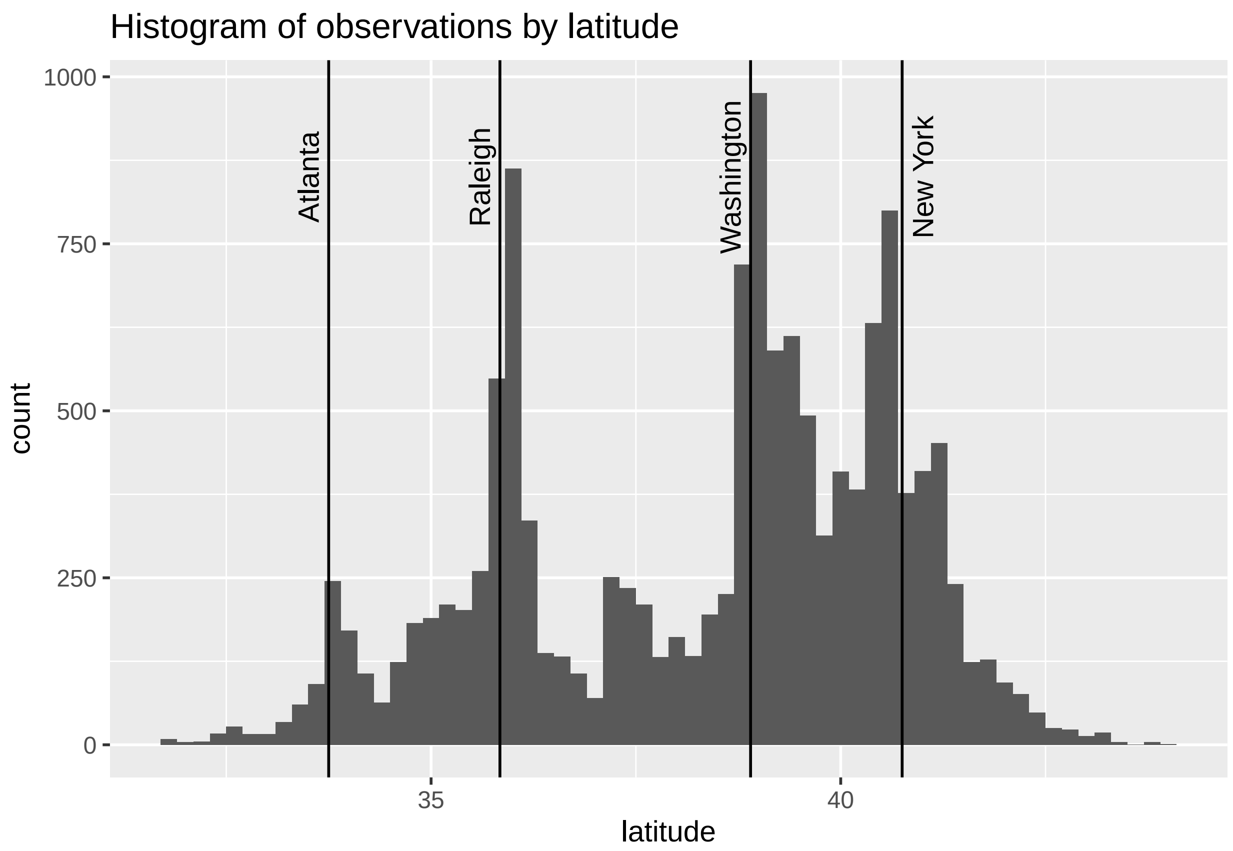
```
# Density of all observations by lat/lon
ggplot(mivi_all, aes(x=longitude,y=latitude)) + geom_point(alpha=0.25) +
    labs(title="Density of observations by location")
```



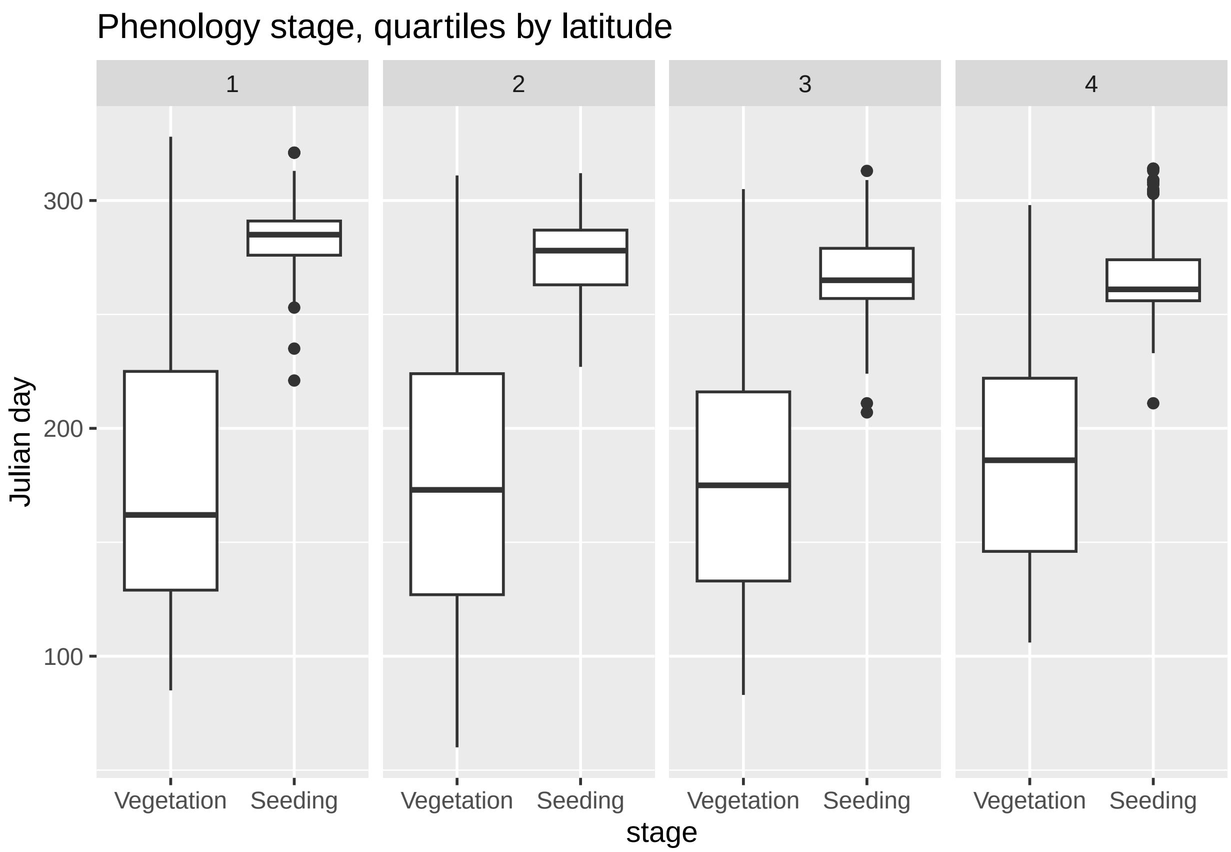
```
# Histogram of all observations by lat
ggplot(mivi_all, aes(x=latitude)) + geom_histogram(binwidth=0.2) +
    labs(title="Histogram of observations by latitude")
```



```
# Histogram 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")
```

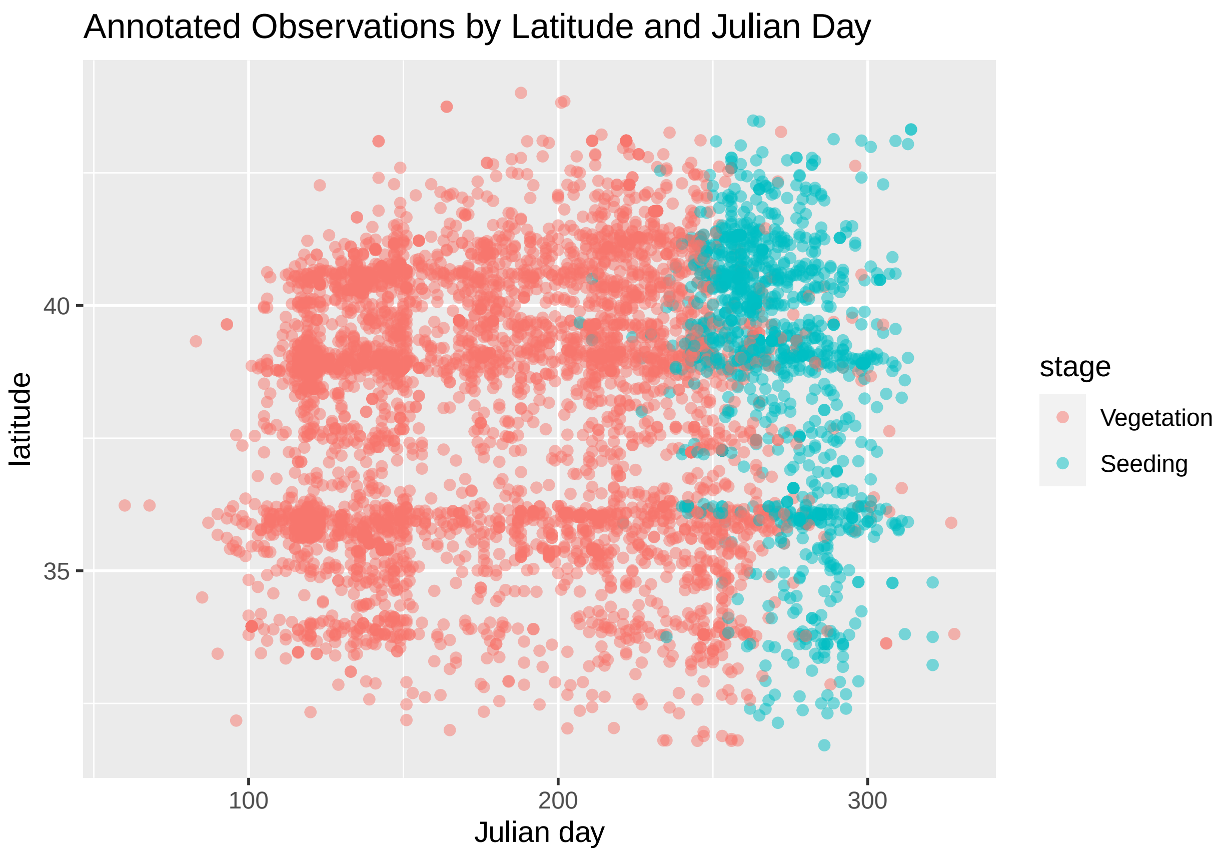


```
# Boxplots of phenology by quartile
# Slight earlier trend in Seeding visible
ggplot(mivi_annotated, aes(julian, stage)) + geom_boxplot() + facet_grid(~group) +
    coord_flip() + xlab("Julian day") + labs(title="Phenology stage, quartiles by latitude")
```

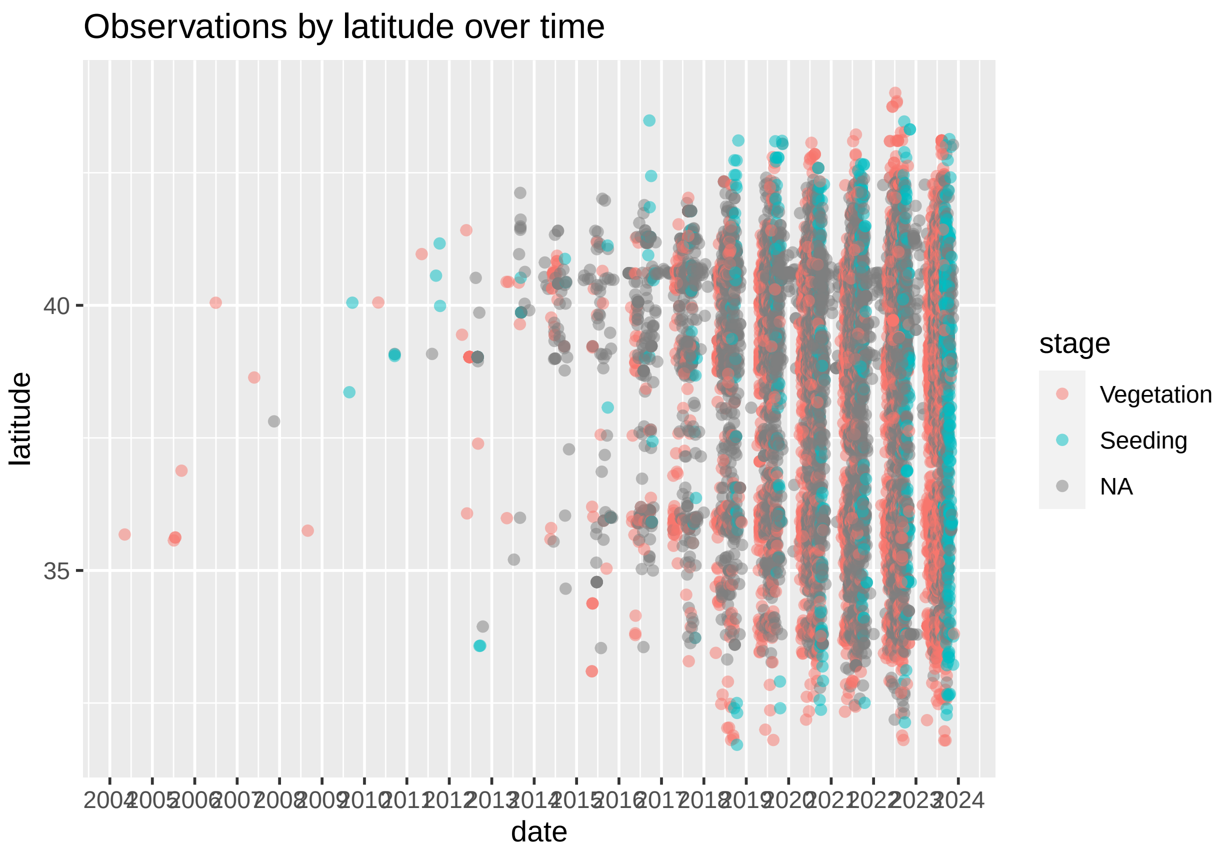


Time Series Plots

```
# 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")
```

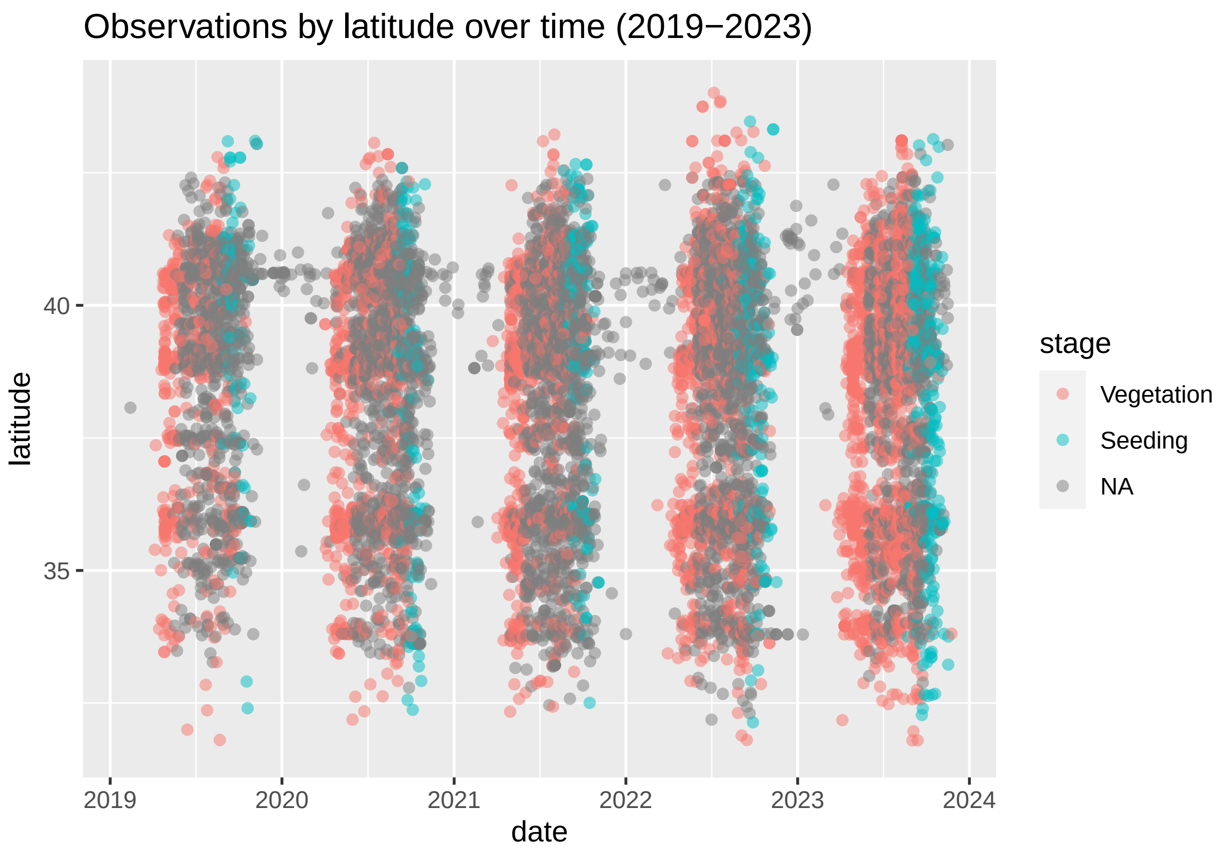


```
# Latitude against Julian day
# ggplot(mivi_all, aes(julian, latitude)) + geom_point(aes(color=stage), alpha=0.5) +
#     scale_color_hue() + xlab("Julian day") +
#     labs(title="Observations by Latitude and Julian Day")

# 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")
```



```
# 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 1699 rows containing missing values (‘geom_point()‘).
```



Retrieve Elevation Information

```
# TODO: replace elevation value with value from 'field:elevation (feet)'

coords <- data.frame(x=mivi_annotated$longitude,
                     y=mivi_annotated$latitude, ele_id=mivi_annotated$id)
# pipe to '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(coords, elevations)

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

Analysis!

```r
# Note: returns Inf if there are none in the selection
first_seed <- function(df) {
    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), minlat=min(latitu

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

```
## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf

## Warning in min(df$julian, na.rm = TRUE): no non-missing arguments to min;
## returning Inf
```

```r
a <- bind_cols(a, do.call(rbind.data.frame, b)[,1]) %>% mutate(fseed = ...5) %>% select(-...5)
```

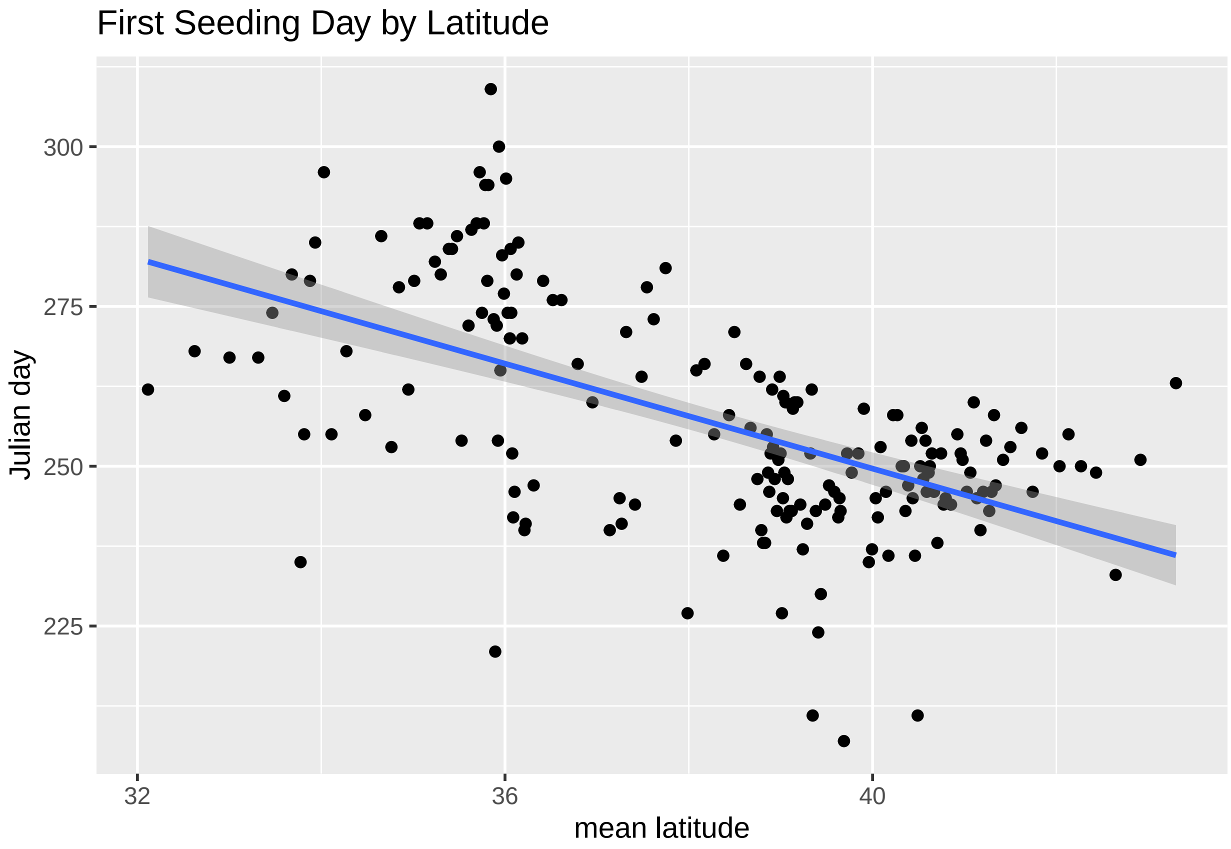
```
## New names:
## * ‘‘ -> ‘...5‘
```

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

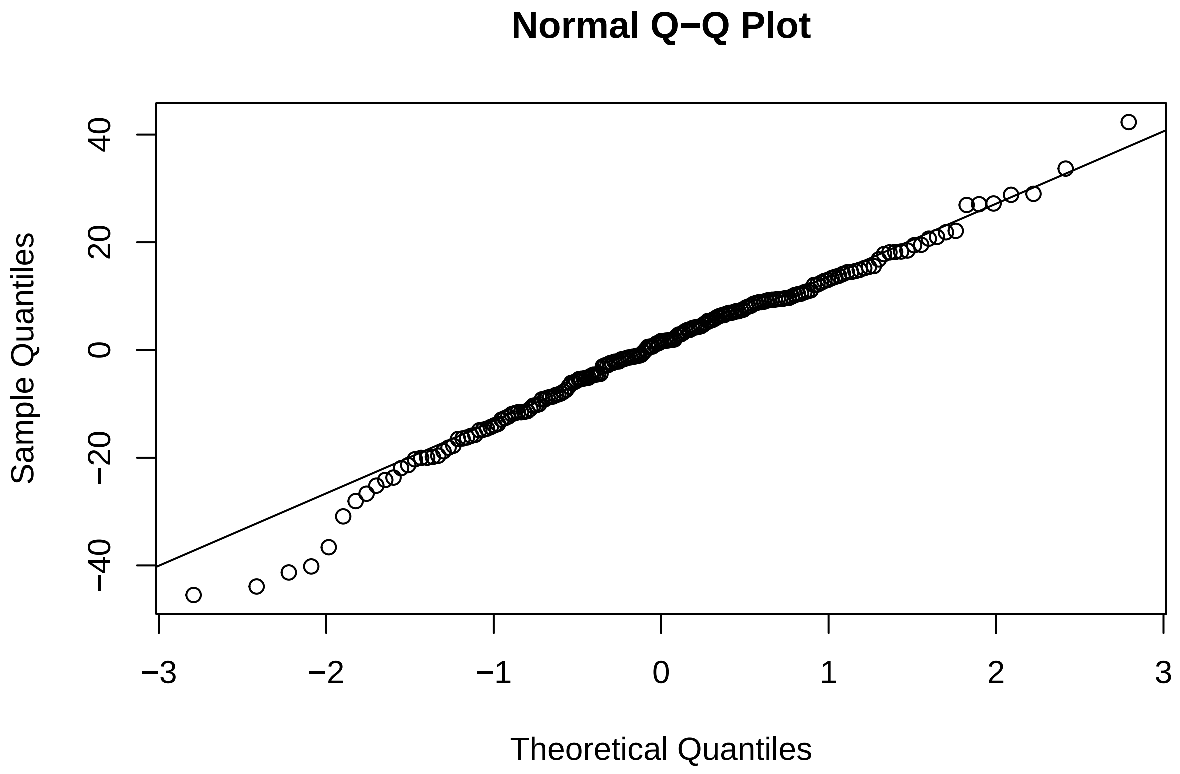
```
## Warning: Removed 9 rows containing non-finite values.
```

```
# 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'
```



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



```
print(n)
```

```
## [1] 200
```

```
print(mad(res))
```

```
## [1] 12.67488
```

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

```
##
##  Pearson’s product-moment correlation
##
## data:  a$fseed and a$mlat
## t = -9.6063, df = 189, p-value < 2.2e-16
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.0000000 -0.4866653
## sample estimates:
##        cor
## -0.5727761
```

```
correlation_matrix(a, use="lower")
```

```
##        group       mlat        minlat      maxlat      fseed
## group  " 1.000   " ""          ""          ""          ""
## mlat   " 0.984***" " 1.000   " ""          ""          ""
## minlat " 0.983***" " 1.000***" " 1.000   " ""          ""
## maxlat " 0.984***" " 1.000***" " 0.999***" " 1.000   " ""
## fseed  "-0.575***" "-0.573***" "-0.572***" "-0.571***" " 1.000   "
```

```
# correlation_matrix(a, use="lower")$estimate

# kable(correlation_matrix(a, use="lower"), booktabs=TRUE, format="latex")
kable(correlation_matrix(a, use="lower"), booktabs=TRUE)
```

	group	mlat	minlat	maxlat	fseed
group	1.000				
mlat	0.984***	1.000			
minlat	0.983***	1.000***	1.000		
maxlat	0.984***	1.000***	0.999***	1.000	
fseed	-0.575***	-0.573***	-0.572***	-0.571***	1.000

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

Elevation

```
# 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)
```

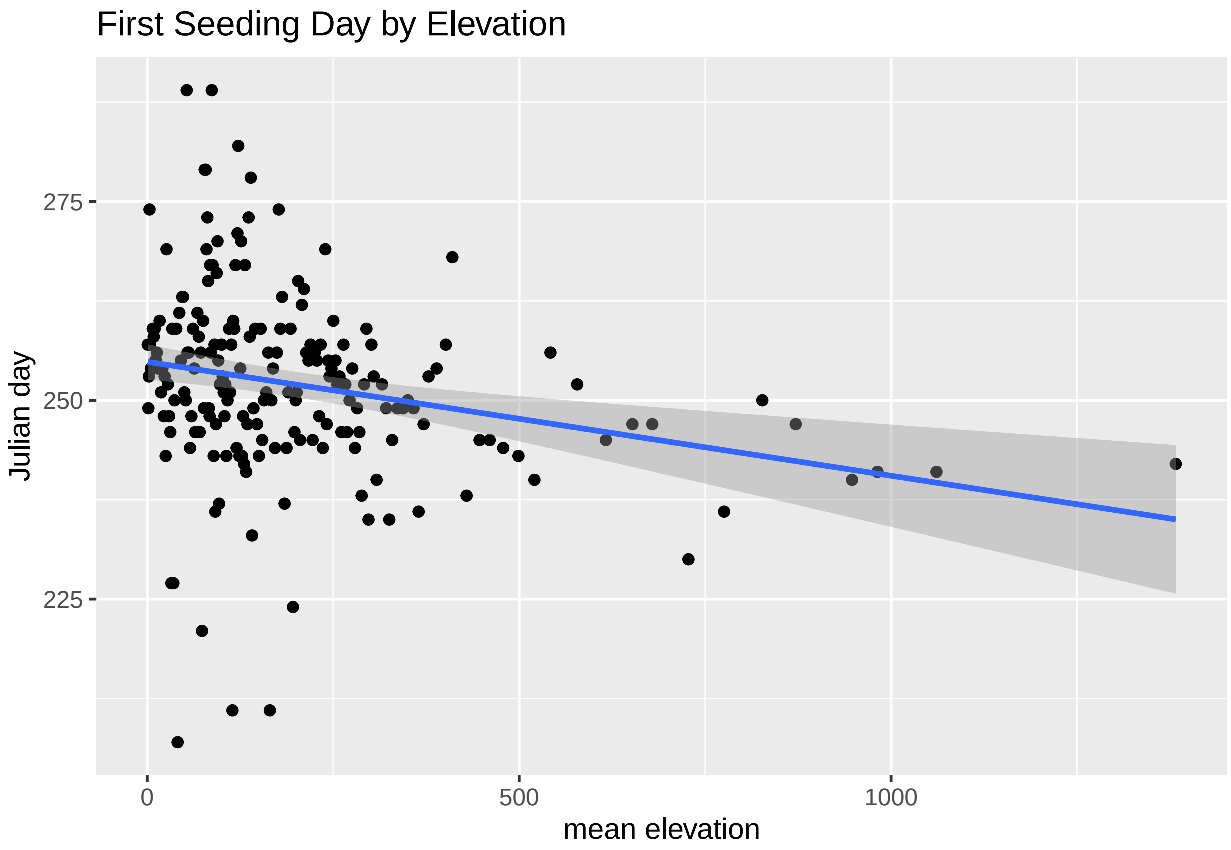
```
## New names:
## * ‘‘ -> ‘...3‘
```

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

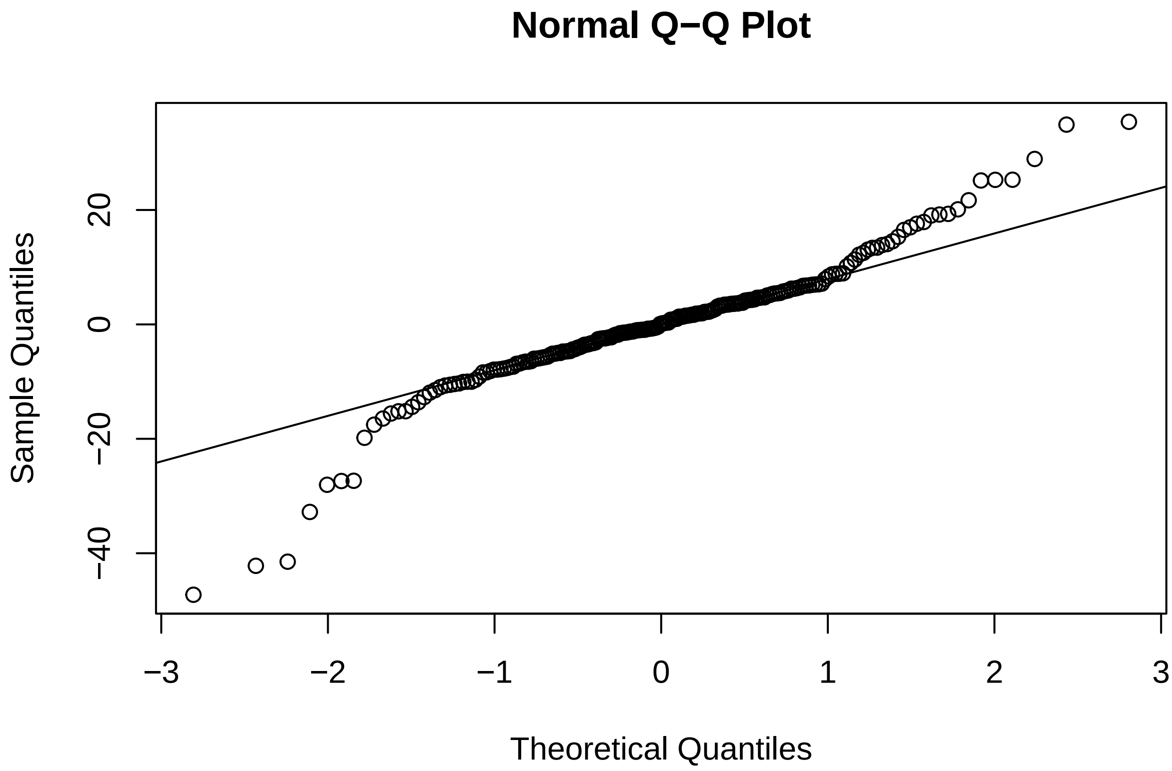
```
## Warning: Removed 1 rows containing non-finite values.
```

```
# 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)
```



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

```
##
##  Pearson's product-moment correlation
##
## data:  a$fseed and a$mele
## t = -3.6327, df = 198, p-value = 0.0001786
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.0000000 -0.1373168
## sample estimates:
##        cor
## -0.2499699
```

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

Tables

```
table1(~ latitude + longitude + elevation + julian | stage, data=mivi_all %>% mutate(stage=addNA(stage)
```

	Vegetation	Seeding	NA	Overall
	(N=4364)	(N=1180)	(N=8186)	(N=13730)
latitude				
Mean (SD)	38.0 (2.44)	38.9 (2.39)	38.5 (2.21)	38.4 (2.32)
Median [Min, Max]	38.8 [31.8, 44.0]	39.3 [31.7, 43.5]	39.1 [32.2, 43.0]	39.0 [31.7, 44.0]
longitude				
Mean (SD)	-79.3 (4.37)	-78.7 (4.60)	-79.0 (4.23)	-79.1 (4.31)
Median [Min, Max]	-78.4 [-97.7, -69.7]	-77.7 [-95.8, -70.6]	-77.7 [-95.2, -70.8]	-77.9 [-97.7, -69.7]
elevation				
Mean (SD)	194 (201)	208 (220)	NA (NA)	197 (205)
Median [Min, Max]	132 [-0.850, 1720]	139 [0.200, 1600]	NA [NA, NA]	133 [-0.850, 1720]
Missing	18 (0.4%)	4 (0.3%)	8186 (100%)	8208 (59.8%)
julian				
Mean (SD)	179 (49.2)	271 (17.4)	223 (47.2)	213 (53.3)
Median [Min, Max]	175 [60.0, 328]	269 [207, 321]	227 [1.00, 365]	219 [1.00, 365]