WLE 650: Lab 1

Lizbeth G Amador

2024-01-19

## Your Assignment

In this lab, you will use the R CheatSheet and any online resources to:

### 1.

Create an RMarkdown file that will be knitted as a word document you turn in. Please be sure to show all code and results in your knitted word document. Also, create sub-headings for tasks 2-5 below.

### 2.

Read in your multivariate data set.

require(tidyverse, quietly = TRUE)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

#Setting directory  
root\_dir = "G:/Shared drives/MultivariateStatistics"  
multivar\_stud\_data = "Data/StudentDataSets"  
  
#read in phenology data  
neonnpn = read.csv(file.path(root\_dir, multivar\_stud\_data, "AmadorL\_NeonNpn\_OpenFLowers\_conus.CSV"))  
  
#remove extra columns  
phe = neonnpn %>%  
 select(-c("elevation\_in\_metersStat", "phenophase\_status", "intensity\_category\_id", "intensity\_value", "abundance\_value", "elevation\_in\_meters", "first\_yes\_julian\_date", "numdays\_since\_prior\_no", "last\_yes\_julian\_date", "numdays\_until\_next\_no"))

### 3.

Use the **head** function to show the columns and first several rows of your dataset.

head(phe, 6)

## observation\_id update\_datetime site\_id state species\_id genus species  
## 1 18208742 2015-03-25 00:00:00 35855 NH 3 Acer rubrum  
## 2 18208843 2015-03-25 00:00:00 35855 NH 3 Acer rubrum  
## 3 18208781 2015-03-25 00:00:00 35855 NH 3 Acer rubrum  
## 4 18340387 2015-04-02 00:00:00 35875 MA 3 Acer rubrum  
## 5 18209053 2015-03-25 00:00:00 35855 NH 3 Acer rubrum  
## 6 18209177 2015-03-27 00:00:00 35855 NH 3 Acer rubrum  
## common\_name kingdom individual\_id phenophase\_id phenophase\_description  
## 1 red maple Plantae 214366 501 Open flowers  
## 2 red maple Plantae 214368 501 Open flowers  
## 3 red maple Plantae 214370 501 Open flowers  
## 4 red maple Plantae 215797 501 Open flowers  
## 5 red maple Plantae 214371 501 Open flowers  
## 6 red maple Plantae 214367 501 Open flowers  
## observation\_date day\_of\_year first\_yes\_year first\_yes\_month first\_yes\_day  
## 1 2014-05-07 127 2014 4 29  
## 2 2014-05-07 127 2014 4 29  
## 3 2014-05-07 127 2014 4 29  
## 4 2014-05-07 127 2014 4 21  
## 5 2014-05-09 129 2014 4 29  
## 6 2014-05-12 132 2014 4 29  
## first\_yes\_doy last\_yes\_year last\_yes\_month last\_yes\_day last\_yes\_doy  
## 1 119 2014 5 9 129  
## 2 119 2014 5 9 129  
## 3 119 2014 5 9 129  
## 4 111 2014 5 9 129  
## 5 119 2014 5 5 125  
## 6 119 2014 5 12 132  
## first\_yes\_sample\_size mean\_first\_yes\_doy data\_name DomainID longitude  
## 1 9 144 NEON D01 -71.28844  
## 2 9 144 NEON D01 -71.28844  
## 3 9 144 NEON D01 -71.28844  
## 4 23 110 NEON D01 -72.17213  
## 5 9 144 NEON D01 -71.28844  
## 6 9 144 NEON D01 -71.28844  
## latitude  
## 1 44.06695  
## 2 44.06695  
## 3 44.06695  
## 4 42.54278  
## 5 44.06695  
## 6 44.06695

### 4.

Create a bulleted list describing the columns in your dataset and the type of variable that each column represents (e.g., ordinal, categorical, continuous, etc. per the lab lecture slides)

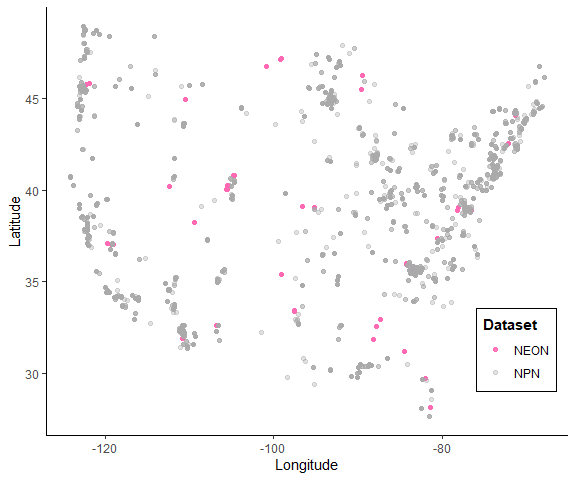
* observation\_id : Unique observation identifier, integer
* update\_datatime : Data (yyyy-MM-DD) and time (hh:mm:ss), character
* site\_id : Unique identifier of site, integer
* genus : Organism’s genus level, character
* species : Organism’s species level, character
* common\_name : Organism’s common name, character
* kingdom : Organism’s kingdom level, character
* individual\_id : Unique identifier for each individual plant surveyed, integer
* phenophase\_id : Unique identifier of phenophase groups, integer
* phenophase\_description : Description of the phenophase, character
* observation\_date : Date (yyyy-mm-dd) of when an observation was originally made, character
* day\_of\_year : converted observation data into a stadardised data measurement, integer
* first\_yes\_year : Year of when phenophase was first oberserved, integer
* first\_yes\_month: Month of when phenophase was first oberserved, integer
* first\_yes\_day : Day of when phenophase was first oberserved, integer
* first\_yes\_doy : Day of year of when phenophase was first oberserved for a given site, integer
* first\_yes\_sample\_size : Sample size of individuals that first flowered at a given site, integer
* mean\_first\_yes\_doy : Average day of year for first flowering individuals at a given site, integer
* data\_name : Dataset origin (NEON or NPN), character
* DomainID : Unique identifier for NEON domains
* longitude : Decimal degree longitude for observation point, numerical
* latitude : Decimal degree latitude for observation point, numerical

### 5.

Create a plot of some of your data.

phe %>%  
 ggplot() +   
 geom\_point(mapping = aes(x=longitude, y = latitude, color = data\_name)) +   
 xlab("Longitude") +   
 ylab("Latitude") +  
 scale\_color\_manual(name="Dataset", values = alpha(c("hotpink", "darkgrey"),   
 c(0.9, 0.35))) +   
 theme(panel.background = element\_rect(fill='transparent'),  
 legend.title=element\_text(face="bold"),  
 legend.key = element\_rect(fill = "white"),  
 legend.position = c(0.9, 0.2),   
 legend.background = element\_rect(size=0.5, linetype="solid", colour ="black"),  
 axis.line = element\_line(color = "black",linewidth = 0.5, linetype = 1))

## Warning: The `size` argument of `element\_rect()` is deprecated as of ggplot2 3.4.0.  
## ℹ Please use the `linewidth` argument instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



par(mfrow=c(1,2))  
#All FFD  
hist(phe$first\_yes\_doy, main = "First Flower Date (FFD)",  
 xlab="Day of Year (DOY)")  
#All avg FFD  
hist(phe$mean\_first\_yes\_doy, main = "Average First Flower Date (FFD)",  
 xlab="Day of Year (DOY)")

