EnvironmentalStatisticsIRP

Getting Started

Always have to start with loading in a few libraries I'm sure I'll need

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   4.0.0     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(ggplot2)
```

Getting the Data

Need to read in 2 critical things: my actual data, as well as my shapefiles for PNF. This data is in a very raw state, so it will take quite a bit of modification to get it into my actual database.

First, I'll read in my raw data, clean it, and make it into a single database. This is a lot of data and will require a lot of cleaning to get into a single database.

```
# Reading in the CSVs. There were multiple files, due to the size of the database being difficult to do

RawBradshaw1 <- read_csv("RawData/RawPNFBradshawData.csv")
```

```
## Rows: 32907 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
RawBradshaw2 <- read_csv("RawData/RawPNFBradshawData2.csv")
```

```
## Rows: 23673 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
RawChinoValley1 <- read_csv("RawData/RawPNFChinoValleyData.csv")
```

```
## Rows: 43557 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
RawChinoValley2 <- read_csv("RawData/RawPNFChinoValleyData3.csv")
```

```
## Rows: 50598 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
RawVerde1 <- read_csv("RawData/RawPNFVerdeData.csv")
```

```
## Rows: 33078 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
RawVerde2 <- read_csv("RawData/RawPNFVerdeData2.csv")
```

```
## Rows: 29910 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): ProtocolType, ProtocolName, EventType, EventName, FormName, Date, ...
## dbl  (9): Slope, Aspect, Elevation, Transect, SampleNumber, Element, nValue,...
## lgl (12): SubElement, cParameter, cParameter2, cParameter3, cValue, cValueCo...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
#From here, I'm going to treat to A) have only the columns I need (site name, location, species, and ra

CleanPNF <- function(df) {
  CleanData <- df %>%
    select(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName, CommonName, SampleNumber, nValue) %>%
    rename(Rank = nValue)
  return(CleanData)
}

Bradshaw1Refined <- CleanPNF(RawBradshaw1)
Bradshaw2Refined <- CleanPNF(RawBradshaw2)
ChinoValley1Refined <- CleanPNF(RawChinoValley1)
ChinoValley2Refined <- CleanPNF(RawChinoValley2)
Verde1Refined <- CleanPNF(RawVerde1)
Verde2Refined <- CleanPNF(RawVerde2)


#Combine the main group into one dataframe

PNFFirst <- Bradshaw1Refined %>%
  full_join(Bradshaw2Refined) %>%
  full_join(ChinoValley1Refined) %>%
  full_join(ChinoValley2Refined) %>%
  full_join(Verde1Refined) %>%
  full_join(Verde2Refined)
```

```
## Joining with ‘by = join_by(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName,
## CommonName, SampleNumber, Rank)‘
## Joining with ‘by = join_by(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName,
## CommonName, SampleNumber, Rank)‘
## Joining with ‘by = join_by(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName,
## CommonName, SampleNumber, Rank)‘
## Joining with ‘by = join_by(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName,
## CommonName, SampleNumber, Rank)‘
## Joining with ‘by = join_by(Date, Ancestry, SiteID, SpeciesSymbol, SpeciesName,
## CommonName, SampleNumber, Rank)‘
```

That’s my first dataframe, so now I’m going to narrow this data down to the point where I can get the composition for each species on each site. To do this, I will count each rank each species got per site (How many times was BOCU ranked 1 on X site in Y event?). This will involve creating a new dataframe, as the per sample data will become irrelevant.

```
PNFSecond <- PNFFirst %>%
  group_by(Ancestry, SiteID, Date, SpeciesSymbol, Rank) %>%
  summarize(RankCount = n())
```

```
## ‘summarise()‘ has grouped output by ’Ancestry’, ’SiteID’, ’Date’,
## ’SpeciesSymbol’. You can override using the ‘.groups‘ argument.
```

```
PNFSecond <- PNFSecond %>%
  mutate(Rank = case_when(
    Rank == 1 ~ 7,
    Rank == 2 ~ 2,
    Rank == 3 ~ 1,
    TRUE ~ Rank
  ))

PNFSecond
```

```
## # A tibble: 20,184 x 6
## # Groups:   Ancestry, SiteID, Date, SpeciesSymbol [8,730]
##    Ancestry                           SiteID Date  SpeciesSymbol  Rank RankCount
##    <chr>                              <chr>  <chr> <chr>         <dbl>     <int>
##  1 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ 2FA               7         1
##  2 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ 2FA               2         1
##  3 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ 2FA               1        21
##  4 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ AGAVE             7         1
##  5 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ AGAVE             2         1
##  6 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ AGAVE             1         2
##  7 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ ARPU5             2         1
##  8 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ ARTE3             7        13
##  9 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ ARTE3             2        13
## 10 USFS > Region 03 > Prescott Natio~ 03-09~ 11/1~ ARTE3             1        10
## # i 20,174 more rows
```

```
PNFPreComposition <- PNFSecond %>%
  group_by(Ancestry, SiteID, Date, SpeciesSymbol) %>%
  summarize(PreComp = Rank*RankCount)
```

```
## Warning: Returning more (or less) than 1 row per ‘summarise()‘ group was deprecated in
## dplyr 1.1.0.
## i Please use ‘reframe()‘ instead.
## i When switching from ‘summarise()‘ to ‘reframe()‘, remember that ‘reframe()‘
##   always returns an ungrouped data frame and adjust accordingly.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## ‘summarise()‘ has grouped output by ’Ancestry’, ’SiteID’, ’Date’,
## ’SpeciesSymbol’. You can override using the ‘.groups‘ argument.
```

```
PNFComposition <- PNFPreComposition %>%
  group_by(Ancestry, SiteID, Date, SpeciesSymbol) %>%
  summarize("Composition(%)" = sum(PreComp)/10)
```

```
## ‘summarise()‘ has grouped output by ’Ancestry’, ’SiteID’, ’Date’. You can
## override using the ‘.groups‘ argument.
```

```
PNFComposition
```

```
## # A tibble: 8,730 x 5
## # Groups:   Ancestry, SiteID, Date [479]
##    Ancestry                          SiteID Date  SpeciesSymbol ‘Composition(%)‘
##    <chr>                             <chr>  <chr> <chr>                    <dbl>
##  1 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ 2FA                        3
##  2 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ AGAVE                      1.1
##  3 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ ARPU5                      0.2
##  4 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ ARTE3                     12.7
##  5 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ BOCU                      25.8
##  6 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ BOHI2                      4.8
##  7 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ CAREX                      0.1
##  8 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ CEMO2                      8.3
##  9 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ ELEL5                      0.3
## 10 USFS > Region 03 > Prescott Nati~ 03-09~ 11/1~ ERIN                       0.3
## # i 8,720 more rows
```