KRC_Mosq_TidyData

```r
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tibble' was built under R version 4.3.3
```

```
## Warning: package 'purrr' was built under R version 4.3.3
```

```
## Warning: package 'lubridate' was built under R version 4.3.3
```

```
## -- 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   3.5.2     v tibble    3.3.0
## 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 errors
```

```r
library(leaflet)
library(lubridate)
```

```r
mosquito_data <- read.csv("mosquito_data.csv", row.names = 1)
```

```r
# About the project:
#   Output of Arizona Map, focusing on Metro PHX and Maricopa
# Add layers of NDVI, Precip
# coordinates --> gis spatial
# create variable "group by" cooridnate set "group_by" (2013-2023) ()
# I want to see the abundance of mosquitos in (, ) species
# Females vs Males
# Filter out rows with NAs/Nones for Date or species of location
# # Mosquito pool testing
# Trapped mosquitoes are separated by species and sex - female *Culex* and *Aedes aegypti* mosquitoes a
#
# ## Arizona Mosquito Dataset
#
# Arizona conducts extensive surveillance of mosquitoes year round, with majority of surveillance occur
```

```r
#change date to year and week (1-52), use lubridate for dates, mutate to create an additional column
mosq_clean <- mosquito_data %>%
  mutate(
    Date = ymd(Date),
    year = year(Date),
    week = isoweek(Date)
  )

#view current species in dataset
mosquito_data %>%
  count(Species)
```

```
##                                         Species      n
## 1                                         Aedes      1
## 2                                 Aedes aegypti 107528
## 3                              Aedes albopictus      1
## 4                                 Aedes epactius     76
## 5                             Aedes nigromaculis      6
## 6                              Aedes purpureipes     30
## 7                              Aedes sollicitans     28
## 8                       Aedes spencerii spencerii      1
## 9                              Aedes trivittatus     69
## 10                                 Aedes vexans  11379
## 11                        Aedes vexans nocturnus     94
## 12                                     Amblyomma      1
## 13                                     Anopheles      3
## 14                         Anopheles franciscanus    383
## 15                               Anopheles hermsi     21
## 16                                          Bird      5
## 17                              Culesita incidens      3
## 18                                         Culex     58
## 19                                Culex erraticus    199
## 20                             Culex erythrothorax    531
## 21                                  Culex pipiens      3
## 22                Culex pipiens/quinquefasciatus     83
## 23 Culex pipiens/quinquefasciatus/restuans      4
## 24        Culex pipiens/restuans/salinarius      1
## 25                     Culex quinquefasciatus 117691
## 26                                 Culex tarsalis  44917
## 27                               Culex territans      1
## 28                               Culex thriambus      2
## 29                                      Culiseta      1
## 30                            Culiseta incidens     44
## 31                            Culiseta inornata    319
## 32                      No Mosquitoes Collected     70
## 33                                         None 262157
## 34                           Ochlerotatus dorsalis      2
## 35                         Ochlerotatus hexodontus     40
## 36                        Ochlerotatus nigromaculis     11
## 37                           Ochlerotatus pullatus      1
## 38                         Ochlerotatus sierrensis      1
## 39                         Ochlerotatus trivittatus     30
## 40                                     Psorophora      3
```

```
## 41                    Psorophora ciliata      1
## 42                  Psorophora columbiae  27025
## 43                  Psorophora confinnis     50
## 44                   Psorophora howardii     25
## 45                Psorophora signipennis     87
## 46                        Psorophora sp  29206
## 47                               Unknown     24
## 48                                  <NA>  17428
```

#questions: 1) should aedes remain anonymous (count=1); 2) for 'Culex' should it be left as standalone

#NA / None / cleaning #keep in mind: Females = transmission relevance & Males = species presence / population signal

```
mosq_clean <- mosq_clean %>%
  filter(
    # Remove inconsistent None rows
    !(Species == "None" & (Males > 0 | Females > 0)),

    # Remove NA species if mosquitoes were caught
    !(is.na(Species) & (Males > 0 | Females > 0))
  )
```

#No mosquitos caught vs. mosquitos caught (Presence/absence models)

#Decision point: some of the species list have multiple species string, I am choosing to create a new column and name it 'Culex Hybrid'. #Culex quinquefasciatus, Culex tarsalis, Aedes aegypti, Culex hybrid (any multi-species Culex string containing /)

##Decision point: We will keep 'zero traps' because it shows the trap was deployed and the species was NOT detected.If we remove zero traps, our model only learns from "where species occurred," which biases suitability upward.

```
mosq_clean <- mosq_clean %>%
  mutate(
    total_mosq = Males + Females,
    trap_zero = if_else(total_mosq == 0, TRUE, FALSE)
  ) %>%
  mutate(
    species_clean = case_when( #abundance + species classification is chained
      str_detect(Species, "Culex") & str_detect(Species, "/") ~ "Culex hybrid",
      Species == "Culex quinquefasciatus" ~ "Culex quinquefasciatus",
      Species == "Culex tarsalis" ~ "Culex tarsalis",
      Species == "Aedes aegypti" ~ "Aedes aegypti",
      TRUE ~ NA_character_
    )
  )
```

```
#view work and filter the 4 groups
mosq_clean <- mosq_clean %>%
  filter(!is.na(species_clean))
```

```r
mosq_clean %>% count(species_clean)
```

```
##             species_clean      n
## 1           Aedes aegypti 107528
## 2            Culex hybrid     88
## 3 Culex quinquefasciatus 117691
## 4          Culex tarsalis  44917
```

Now lets create a weekly summary to sum across the info

```r
weekly_summary <- mosq_clean %>%
  filter(!is.na(species_clean)) %>%
  group_by(year, week, species_clean) %>%
  summarise(
    total_traps = n(), #Number of trap × species records that week.
    traps_detected = sum(total_mosq > 0),#How many traps detected this species that week
    total_males = sum(Males, na.rm = TRUE),
    total_females = sum(Females, na.rm = TRUE),
    total_mosq = sum(total_mosq, na.rm = TRUE), #Total abundance of that species that week.

    total_tests = sum(!is.na(Result)), #Number of pools tested.
    positives = sum(Result == 1, na.rm = TRUE), #Number of positive pools.
    positivity_rate = if_else(total_tests > 0, #Positive pools / tested pools.
                              positives / total_tests,
                              NA_real_),

    .groups = "drop"
  )
#NOTE: This summary is species-specific weekly aggregation.
#It does NOT yet include zero traps where species absent.
```

```r
glimpse(weekly_summary)
```

```
## Rows: 1,662
## Columns: 11
## $ year            <dbl> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, ~
## $ week            <dbl> 1, 1, 1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8, ~
## $ species_clean   <chr> "Aedes aegypti", "Culex quinquefasciatus", "Culex tars~
## $ total_traps     <int> 1, 2, 2, 22, 9, 6, 6, 1, 40, 7, 1, 13, 3, 58, 12, 15, ~
## $ traps_detected  <int> 1, 2, 2, 22, 9, 6, 6, 1, 40, 7, 1, 13, 3, 58, 12, 15, ~
## $ total_males     <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, ~
## $ total_females   <int> 1, 2, 4, 36, 9, 6, 12, 4, 111, 14, 1, 29, 5, 297, 30, ~
## $ total_mosq      <int> 1, 2, 4, 36, 9, 6, 12, 4, 111, 14, 1, 29, 5, 297, 30, ~
## $ total_tests     <int> 0, 2, 2, 22, 9, 6, 6, 0, 40, 7, 0, 13, 3, 58, 12, 15, ~
## $ positives       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ positivity_rate <dbl> NA, 0, 0, 0, 0, 0, 0, NA, 0, 0, NA, 0, 0, 0, 0, 0, 0, ~
```

```r
arrange(weekly_summary, year, week, species_clean)
```

```
## # A tibble: 1,662 x 11
##     year  week species_clean         total_traps traps_detected total_males
```

```
##      <dbl> <dbl> <chr>                          <int>          <int>          <int>
##  1  2013     1 Aedes aegypti                      1              1              0
##  2  2013     1 Culex quinquefasciatus             2              2              0
##  3  2013     1 Culex tarsalis                     2              2              0
##  4  2013     2 Culex quinquefasciatus            22             22              0
##  5  2013     2 Culex tarsalis                     9              9              0
##  6  2013     3 Culex quinquefasciatus             6              6              0
##  7  2013     3 Culex tarsalis                     6              6              0
##  8  2013     4 Aedes aegypti                      1              1              0
##  9  2013     4 Culex quinquefasciatus            40             40              0
## 10  2013     4 Culex tarsalis                     7              7              0
## # i 1,652 more rows
## # i 5 more variables: total_females <int>, total_mosq <int>, total_tests <int>,
## #   positives <int>, positivity_rate <dbl>
```

Next steps:

```
# Species-specific abundance modeling
#
# Species-specific positivity rates
#
# Spatial risk mapping by species
#
# Presence/absence modeling
```