Urban Phenology Project

Victoria Garfield

2025-11-11

```
library(sf)
```

```
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
library(terra)
```

```
## terra 1.8.80
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract() masks terra::extract()
## 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
```

```
library(terra)
library(ggplot2)
library(dplyr)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

. Introduce the problem and explain why: Urbanization impacts local climate patterns and species interactions with the surrounding land; this leads to ecological shifts, such as the timing of events like flowering or migration (Park et al. 2023; Leong 2014). These changes in our environments often end up leading to changes in phenology (Wu et al 2025). Changes in phenology can disrupt the surrounding ecosystems, this can impact the relationship between pollinators and plants. This in turn impacts seasonal changes, coming earlier or later, flowering times, and bee distribution (Park et al. 2023; Leong 2014). Urbanized areas tend to have warmer seasons, but the understanding of this phenomenon is not well known. Understanding why these kinds of patterns differ in urban and rural areas is important. As we see an increase in urbanization and climate change, the observation between rural and urbanized areas can help us prepare and understand what we may see more widespread in the future.

2. Past work and data available Past work on the studies of urbanized areas compared to rural found that urban environments tend to have earlier flowering times and differing arrival times of bees within these areas compared to rural areas. (Wu et al. 2025; Li, X et al. 2017). With this in mind, the observation of environmental changes like reduced vegetation, higher surface temperatures and carbon clusters, via MODIS or NDVI can help us measure these traits. The main setback to these methods of collecting data would be coarse spatial scales. This problem prevents us from looking and finer details of these urbanized areas, and potentially missing rare or special species that may benefit from these urbanized areas. With more detailed/smaller scale datasets like the Oregon Bee Atlas (OBA) and the Global Biodiversity Information Facility (GBIF) can grant us insight on localized temporal changes in species activity. In conjunction with environmental datasets like PRISM climate data and NLCD land cover data set, we can provide a more precise scale of what drives these differences between urban and rural areas.
3. Purpose of the study

To address the questions of how: Further refine your approach (e.g., what data will you combine/transorm to a new type, how will you address the question) Justify why this is needed now (e.g., visualization to test a new dimension of the question or better convey an old one)

4. Hypotheses/questions: My hypotheses and questions are as follows, Hypothesis one, how do seasonal changes of urban environments compared to rural areas impact bee phrenology. Hypothesis two would be what is driving these changes. For hypothesis three, what causes these differences in phenology.

9.5.2 Dataset identification The data I will be using primarily is Oregon Bee Atlas (OBA) dataset, in conjunction with the National Land Cover Database (NLCD). NLCD will allow me to look at urbanized areas and forest cover. This is useful because I will be able to compare these two and see how urbanization has impacted bee phenology. I like this dataset because I can also look at how forest cover has changed over times, whether that is from wildfires or human interference, and infer that way.

```r
oba_clean <- read.csv("~/Downloads/BeeUrbanizationProject/OBA_2018-2024.csv")

glimpse(oba_clean)
```

```
## Rows: 240,752
## Columns: 58
## $ fieldNumber                <chr> "1800001", "1800002", "1800003", "180000~
## $ catalogNumber              <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ occurrenceID               <chr> "https://osac.oregonstate.edu/OBS/OBA_18~
## $ userId                     <chr> "429964", "429964", "429964", "429964", ~
## $ userLogin                  <chr> "amelathopoulos", "amelathopoulos", "ame~
## $ firstName                  <chr> "Andony", "Andony", "Andony", "Andony", ~
## $ firstNameInitial           <chr> "A.", "A.", "A.", "A.", "A.", "A.", "A."~
## $ lastName                   <chr> "Melathopoulos", "Melathopoulos", "Melat~
## $ recordedBy                 <chr> "Andony Melathopoulos", "Andony Melathop~
## $ sampleId                   <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ specimenId                 <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ day                        <int> 4, 4, 4, 4, 16, 22, 22, 22, 22, 22, 22, ~
## $ month                      <chr> "5", "5", "5", "5", "5", "6", "6", "6", ~
## $ year                       <int> 2018, 2018, 2018, 2018, 2018, 2018, 2018~
## $ verbatimEventDate          <chr> "5/4/18", "5/4/18", "5/4/18", "5/4/18", ~
## $ day2                       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ month2                     <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ year2                      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ startDayofYear             <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
## $ endDayofYear                  <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ country                       <chr> "USA", "USA", "USA", "USA", "USA", "USA"~
## $ stateProvince                 <chr> "OR", "OR", "OR", "OR", "OR", "OR", "OR"~
## $ county                        <chr> "Benton", "Benton", "Benton", "Benton", ~
## $ locality                      <chr> "Corvallis", "Corvallis", "Corvallis", "~
## $ verbatimElevation             <chr> "71", "71", "71", "71", "62", "1271", "1~
## $ decimalLatitude               <dbl> 44.5599, 44.5599, 44.5599, 44.5599, 45.6~
## $ decimalLongitude              <dbl> -123.2883, -123.2883, -123.2883, -123.28~
## $ coordinateUncertaintyInMeters <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ samplingProtocol              <chr> "aerial net", "aerial net", "aerial net"~
## $ relationshipOfResource        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ resourceID                    <chr> "https://osac.oregonstate.edu/OBS/OBA_18~
## $ relatedResourceID             <chr> "1444041d-1464-49d0-9c5c-f297aa90e2ae", ~
## $ relationshipRemarks           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ phylumPlant                   <chr> "Tracheophyta", "Tracheophyta", "Tracheo~
## $ orderPlant                    <chr> "Ranunculales", "Ranunculales", "Ranuncu~
## $ familyPlant                   <chr> "Berberidaceae", "Berberidaceae", "Berbe~
## $ genusPlant                    <chr> "", "", "", "", "Vicia", "Eschscholzia",~
## $ speciesPlant                  <chr> "", "", "", "", "Vicia villosa", "Eschsc~
## $ taxonRankPlant                <chr> "family", "family", "family", "family", ~
## $ url                           <chr> "https://www.inaturalist.org/observation~
## $ phylum                        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ class                         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ order                         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ family                        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ genus                         <chr> "Andrena", "Andrena", "Andrena", "Andren~
## $ subgenus                      <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ specificEpithet               <chr> "prunorum", "", "hippotes", "angustitars~
## $ taxonomicNotes                <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ scientificName                <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ sex                           <chr> "male", "male", "male", "female", "femal~
## $ caste                         <chr> "", "", "", "", "worker", "worker", "", ~
## $ taxonRank                     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ identifiedBy                  <chr> "L.R.Best", "L.R.Best", "A.S.Jackson", "~
## $ familyVolDet                  <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ genusVolDet                   <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ specificEpithetVolDet         <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ sexVolDet                     <chr> "", "", "", "", "", "", "", "", "", "", ~
## $ casteVolDet                   <chr> "", "", "", "", "", "", "", "", "", "", ~
```

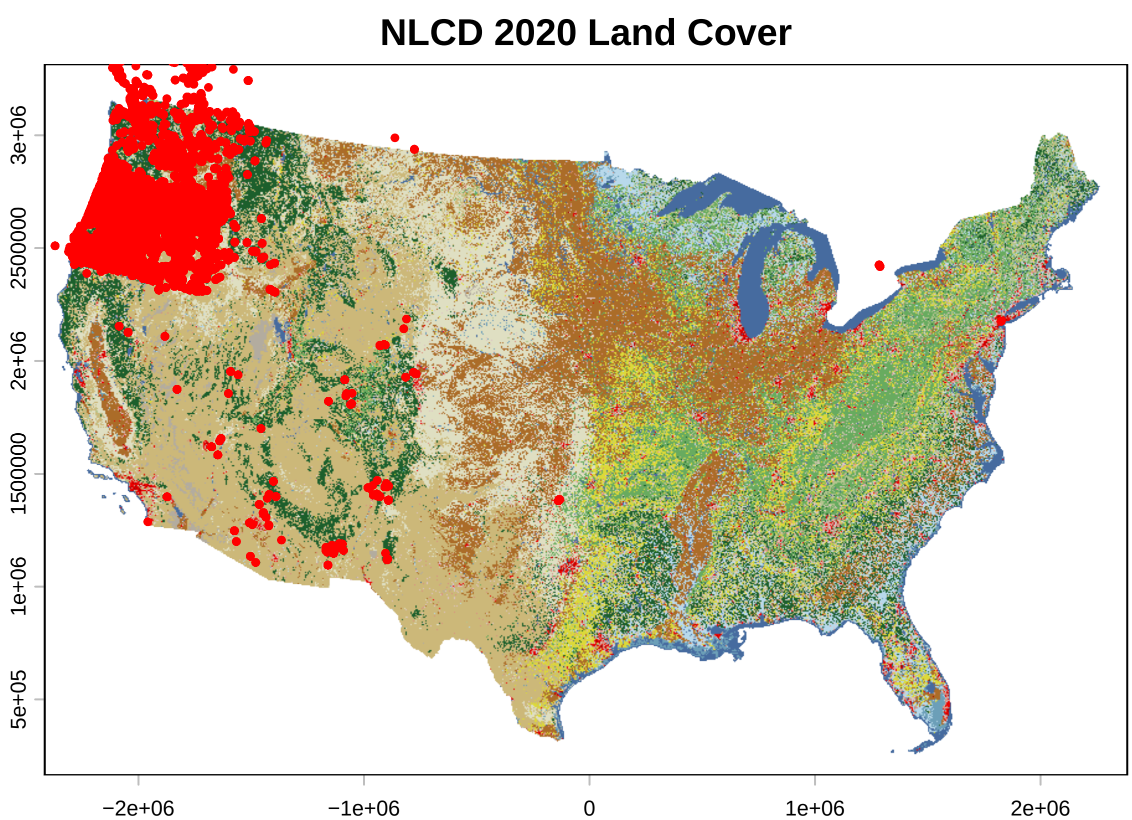
```r
## cleaning Data, Getting rid of NA's
oba_clean <- oba_clean %>% filter(!is.na(decimalLatitude), !is.na(decimalLongitude))


## Convert to sf points (lat/long fields)
oba_sf <- st_as_sf(oba_clean, coords = c("decimalLongitude", "decimalLatitude"), crs = 4326)

## NCLD 2020 RIF RASTER
nlcd_2020_landcover <- rast("~/Downloads/BeeUrbanizationProject/Annual_NLCD_LndCov_2020_CU_C1V1.tif")

nlcd_2020_impervious <- rast("~/Downloads/BeeUrbanizationProject/Annual_NLCD_FctImp_2020_CU_C1V1.tif")
```

```r
# reprojecting OBA points, to align with NLCD
oba_sf <- st_transform(oba_sf, crs(nlcd_2020_landcover))
```

```r
## Extract NLCD values at each bee point
oba_clean$nlcd_2020_landcover <- terra::extract(
  nlcd_2020_landcover,
  vect(oba_sf)
)[,2]
```

```r
## Extracting Impervious surface value
oba_clean$nlcd_2020_impervious <- terra::extract(
  nlcd_2020_impervious,
  vect(oba_sf)
)[,2]
```

```r
##quick confirmation of nlcd land cover, making sure its what I am looking for
plot(nlcd_2020_landcover, main = "NLCD 2020 Land Cover")
points(oba_sf, pch = 20, col = "red")
```




```r
## Define and add rural/urban classification, impervous
oba_clean$imperv_class <- dplyr::case_when(
  oba_clean$nlcd_2020_impervious >= 20 ~ "Urban",
  oba_clean$nlcd_2020_impervious > 0 & oba_clean$nlcd_2020_impervious < 20 ~ "Suburban",
  oba_clean$nlcd_2020_impervious == 0 ~ "Rural",
```

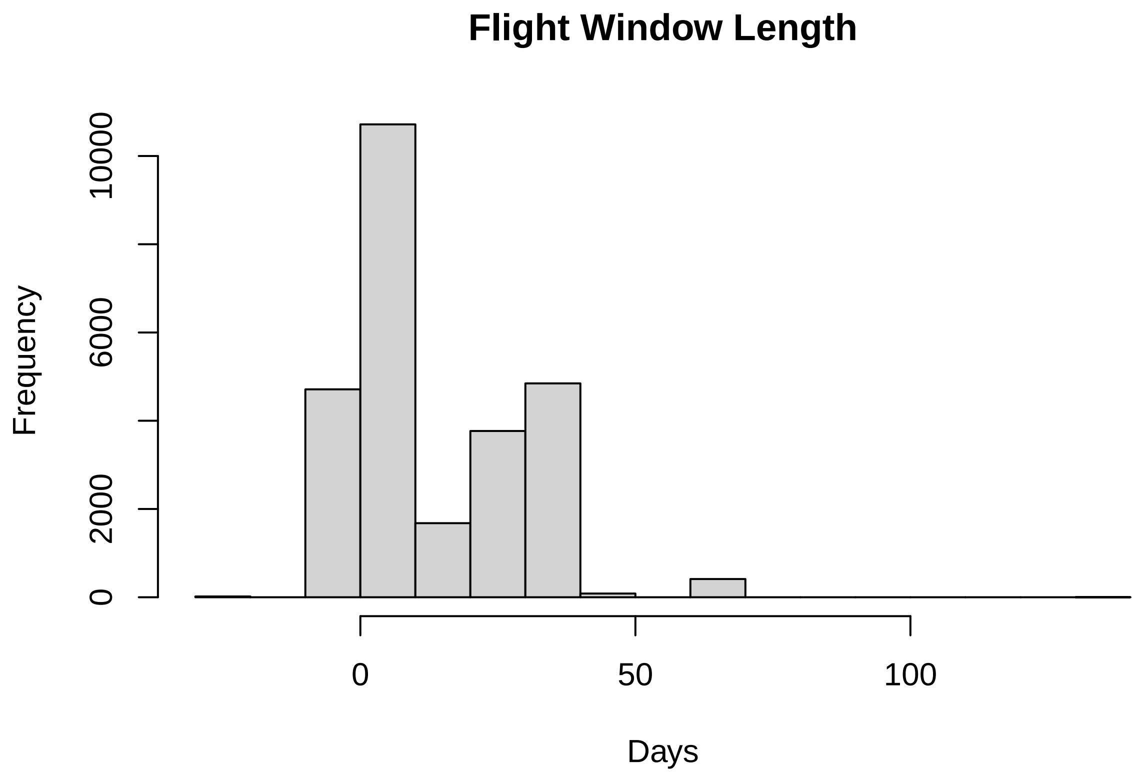
```r
  TRUE ~ NA_character_
)

##Error in `mutate()`: In argument: `doy_mid = (startDayofYear + endDayofYear)/2`.Caused by error in `s

oba_clean <- oba_clean %>%
  mutate(
    endDayofYear = ifelse(endDayofYear == "", NA, endDayofYear),
    endDayofYear = as.numeric(endDayofYear)
  )
```

```
## Warning: There was 1 warning in ‘mutate()‘.
## i In argument: ‘endDayofYear = as.numeric(endDayofYear)‘.
## Caused by warning:
## ! NAs introduced by coercion
```

```r
oba_clean <- oba_clean %>%
  mutate(
    endDayofYear = ifelse(endDayofYear == "", NA, endDayofYear),
    endDayofYear = as.numeric(endDayofYear)
  )


##Phenology metrics
oba_clean <- oba_clean %>%
  mutate(
    doy_start = startDayofYear,
    doy_end = endDayofYear,
    doy_mid = (startDayofYear + endDayofYear) / 2,
    flight_window = endDayofYear - startDayofYear
  )
```

```r
## create midpoint of my DOY
oba_clean <- oba_clean %>%
  mutate(
    doy_mid = (startDayofYear + endDayofYear) / 2
  )
```

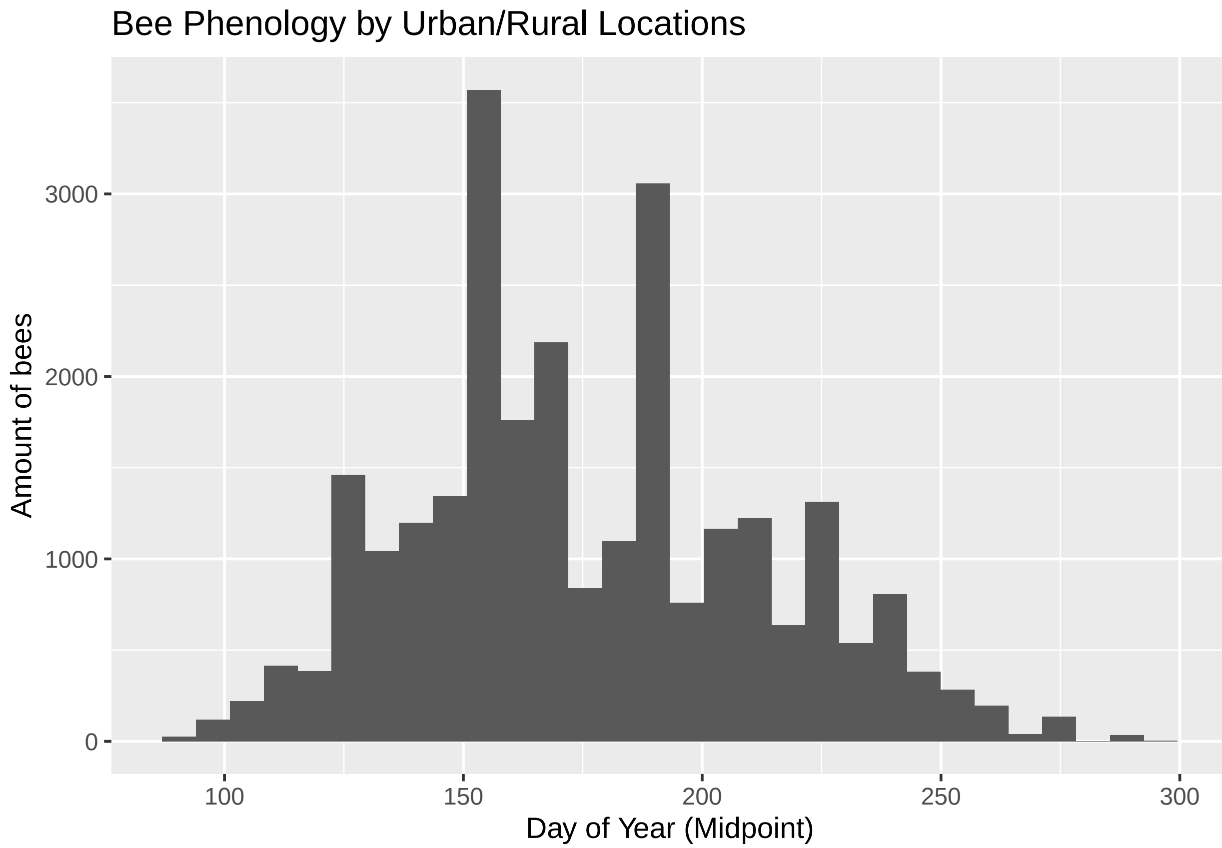
```r
hist(oba_clean$flight_window,
     main="Flight Window Length",
     xlab="Days")
```




```
## Urban vs rural phenology comparison
library(ggplot2)

ggplot(oba_clean, aes(x = doy_mid)) +
  geom_histogram() +
  labs(
    title = "Bee Phenology by Urban/Rural Locations",
    x = "Day of Year (Midpoint)",
    y = "Amount of bees",
  )
```

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```

```
## Warning: Removed 214477 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



Now that we have collected DOY bee phenology and window of flight time, I want to compare these and see how be phenology compares to the window of flight time in urban and rural areas.

```
# filter and clean oba_clean, so it can be plotted as a ggplot
oba_clean_plot <- oba_clean %>%
  filter(!is.na(doy_mid), !is.na(flight_window), !is.na(imperv_class)) %>%
  mutate(imperv_class = factor(imperv_class, levels = c("Rural", "Urban")))

## ggplot
```

Workflow Description: Data Cleaning 1. removing missing spatial data - found that latitude and longitude had some missing coordniates 2. filtered for non-bee species. - removed any non scientific/species names. Using scientific name. 3. missing phenology data. - conveerted empty strings in "startofDay" and "endDayofYear" to "NA", converted columns to numeric, this was causing me issues when I first started trying to analyze my data. 4. removing multiple of the same records, to prevent bias.

Data Transformation: 1. looked at how to calculate phenology metrics in R studio(will attach reference into this workflow detail) to analyze middle of the year time period, this helped me look at flight window times as well. I did this to analyze what the flight period of bees, to see in the future when they are arriving in certain areas (urban and rural). 2. Environmental factors. wanted to use terra::extract() to NCLD to look at land cover values and surface estimates on bee observation

Aggregation: 1. Remove incomplete spatial data - organize data by country, year(2020) and urbanization to observe temporal triends

Vizualization: 1. Histograms - Want to show distribution of middle of the year and flight windows across Urban and rural areas. 2. Spatial plots: - map bee distribution using NLCD raster to show urban/rural gradients

Functions: 1. function to compute doy_mid and flight_window for data. - using histograms to observe time periods of bee flights/travel to an area - use ggplot to identify trends of how bees are distributed/travel according to urban and rural areas