

Final Write Up

Ava Wessel

Mia Hanson

2024-12-08

Note to reader: We had to combine four different R markdown scripts, so the addition of saving our models/datasets and rereading in the datasets was set up prior to combination. I guess it is still a good habit to have though. Because of this, the code from the original scripts might differ slightly to get this file to properly knit. The overall function of this markdown is the same.

Effect of Urbanization and Temperature on *Bombus* spp. Emergence

I. Introduction:

Bumble bees (*Bombus* spp.) and other bee species play an important ecological role as pollinators for various wildflowers and agricultural crops (Ahrné et al. 2009; McCabe et al. 2022). Despite their ecological importance, pollinators face numerous threats from habitat loss, pesticide exposure, climate change, and disease, with many bee populations declining worldwide (Goulson et al. 2008). Habitat loss and fragmentation caused by urbanization emerges as a growing concern due to its rapid expansion and transformation of landscapes that were once natural habitats into fragmented urban environments. Urbanization can have a large effect on *Bombus* spp. flower visitation rates and pollination success through changes in foraging behavior (Glaum et al. 2017; Harrison & Winfree 2015). Urban heat islands, or the increased temperatures often observed in cities due to concrete and reduced green spaces, further complicate the challenges faced by pollinators (Harrison & Winfree 2015). Together, these pressures can lead to declines in Oregon bee populations, leading to cascading effects impacting the numerous ecosystem services offered. Understanding how urbanization impacts emergence patterns of *Bombus* spp. is crucial, as these factors are closely tied to their ability to effectively pollinate and maintain stable populations.

Recent research has found that climate change, often resulting from urbanization, is altering phenology, with spring events generally occurring earlier (Stemkovski et al. 2020). For *Bombus* spp., increased (spring) temperatures have been shown to directly lead to advanced phenology, prompting earlier emergence and activity periods (Prestele et al. 2021). Harrison & Winfree (2015) observed that urban habitats support a greater abundance of late-season bees compared to forest habitats, likely due to the presence of late-blooming plants. They also suggest that urban warming may act as an “ecological filter,” reducing species diversity and resulting in a more homogenized bee community in cities. Similarly, Blasi et al. (2023) found that over the past 20 years, a citizen-science dataset has shown that *Bombus* spp. emergence across Sweden has advanced significantly, though responses varied based on species traits. As in Harrison & Winfree’s findings, warmer temperatures substantially advanced the flight period for early-emerging species, while late-emerging species exhibited no consistent trend. This variability in response among *Bombus* spp. may contribute to shifts in community composition, favoring species that can better synchronize their life cycles with the altered urban floral availability. Little is known about how these phenological changes impact pollination success across diverse urban ecosystems, especially in areas like Oregon where agricultural productivity and biodiversity heavily depend on stable bee populations.

Understanding these emergence patterns and phenological shifts is crucial for pollinator diversity conservation efforts. Using the Oregon Bee Atlas (OBA), census population block data, and monthly temperature reading from the PRISM group, we are able to develop a more comprehensive analysis of how urbanization and temperature variability affect *Bombus* spp. Li et al. (2020) demonstrated that human population density can serve as a reliable indicator of urbanization, allowing us to use census data as a proxy for assessing the

intensity of urban development and its impacts on bee populations. By combining monthly temperature data and population density, we are able to assess the response of *Bombus* spp. Across various levels of urbanization, allowing us to identify species-specific emergence patterns and population trends among different bee species.

Our research question is: How does peak emergence for four *Bombus* species, *vosnesenskii*, *melanopygus*, *mixtus*, and *griseocollis*, vary along climatic and urbanization gradients? Our first hypothesis is that bees in more heavily urbanized areas will have earlier emergence dates. Our second hypothesis is that areas that have higher mean temperatures will show earlier emergence times. Our third hypothesis is that earlier emerging species will be more sensitive to warming temperature than later emerging species, specifically we will see a more significant change in emergence days per °C of temperature change.

II. Data/ Data Wrangling:

The data that we be combined is the OBA dataset and merged it with US Census data to get whether those bee observations were in a more rural or urban area. We read in the US Census block data as a shapefile data type. Then we wrangled the population column into 4 even categories using the quantile function. The 4 categories were rural, subrural, urban, and suburban. We then read in the PRISM temperature data, originally as raster data, merging each of the bee observations with temperature, joining on latitude and longitude in order to see what the average temperature is in the area the bee was recorded. With the PRISM temperature data we calculated average spring temperatures, finding the mean of the average temperature from months February-June. For data exploration we made a couple tables filtering the species we were considering focusing on for our analysis, in the `Data_Exploration.Rmd` file and we took a look at any weird values like NAs in our dataframes using tables in the `Data_Cleaning.Rmd` file.

```
#Loading in OBA
bee_data <- read.csv("OBA_2018-2023_combined_dataset_working.csv")

# want to just keep our species from the main dataframe
filtered_bee_data <- bee_data[bee_data$Species %in% c("vosnesenskii", "melanopygus", "mixtus", "griseocollis"), ]

# want to keep Dec..Lat. and Dec..Long. and Collection.Date and Collection.Day.1
filtered_bee_data <- filtered_bee_data %>%
  dplyr::select(Dec..Lat., Dec..Long., Collection.Date, Species, Year.1)

head(filtered_bee_data)
```

##	Dec..Lat.	Dec..Long.	Collection.Date	Species	Year.1
## 34	44.568	-123.283	4/18/2018	mixtus	2018
## 35	44.568	-123.283	4/18/2018	mixtus	2018
## 520	42.229	-121.777	6/29/2018	vosnesenskii	2018
## 527	42.181	-121.737	7/20/2018	vosnesenskii	2018
## 562	45.643	-122.814	5/28/2018	mixtus	2018
## 603	45.642	-122.817	6/24/2017	vosnesenskii	2017

```
# removing these na rows
filtered_bee_data <- na.omit(filtered_bee_data)
```

creating a day of year column

```

# logic behind the day of year: January 1st is day 1, jan 2nd is day 2 etc

# using lubridate

calc_day_of_year <- function(date){

  # converting the string object to a Date object
  # parses mm/dd/yyyy
  date <- mdy(date)

  # get day of year
  day_of_year <- yday(date)

  return (day_of_year)

}

filtered_bee_data <- mutate(filtered_bee_data, DOY = calc_day_of_year(Collection.Date))

```

```

## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'DOY = calc_day_of_year(Collection.Date)'.
## Caused by warning:
## ! 3 failed to parse.

```

```

#Creating winter month, day, and year for PRISM extraction
month <- seq(as.Date("2020-01-01"),
             as.Date("2020-12-01"),
             by = "1 month")
month_label <- lubridate::month(month, label = TRUE)

winter.months <- seq(as.Date("2020-09-01"),
                    as.Date("2021-08-01"),
                    by = "1 month")
winter.month_label <- lubridate::month(winter.months, label = TRUE)

filtered_bee_data <- filtered_bee_data %>%
  mutate(Collection.Date=mdy(Collection.Date)
         ,year=as.numeric(year(Collection.Date))
         ,month=month(Collection.Date)
         ,doy = yday(Collection.Date)
  ) %>%
  arrange(year,doy)

```

```

## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'Collection.Date = mdy(Collection.Date)'.
## Caused by warning:
## ! 3 failed to parse.

```

```

filtered_bee_data <- filtered_bee_data %>% rename(latitude = Dec..Lat., longitude = Dec..Long.)

filtered_bee_data <- mutate(filtered_bee_data,
                           winter.year = ifelse(month>8, year+1, year)

```

```

        ,winter.month = ifelse(month>8, month-8, month+4)
        ,winter.day = ifelse(month>8, doy-243, doy+120)
    )

filtered_bee_data <- na.omit(filtered_bee_data)

dim(filtered_bee_data)

```

```
## [1] 5116 12
```

Adding in PRISM climate data:

Note: I could not for the life of me figure out how to bypass this knitting error without completely changing the code. I check to see if they were all numeric and they were, not really sure the solution. To see working code on the original markdown file, see `Data_Cleaning.rmd`. I included the code up until it stopped working and the execution halt error occurs. I included the error below: Error in FUN(): ! only defined on a data frame with all numeric-alike variables Backtrace: 1. ... %>% print(bee_long) 4. pillar:::print.tbl(., bee_long) 5. pillar:::print_tbl(...) 8. pillar:::format.tbl(...) 9. pillar:::format_tbl(...) ... 17. pillar:::ctl_new_pillar.tbl(controller, vector(), width, title = title) 20. pillar:::max0(width) 21. base::Summary.data.frame(<tibble[,21]>, na.rm = FALSE) 22. base::lapply(...) 23. base (local) FUN(X[[i]], ...)

Quitting from lines 130-202 [unnamed-chunk-3] (Final_Writeup.Rmd) Execution halted

```

#Adding PRISM Data and calculating values for 'winter year' so that we take into account the months pri

prism_set_dl_dir("prism_data")

# Had to comment out as it was printing all output in the pdf
#get_prism_monthlys(
# type = "tmean",
# year=2016:2021,
# mon=1:12,
# keepZip = FALSE
# )

sub.tmean <- prism_archive_subset("tmean", "monthly", mon = 1:12)

RS <- pd_stack(sub.tmean) ##raster file

bee.test <- filtered_bee_data
bee.spdf <- SpatialPointsDataFrame(coords=bee.test[,c('longitude','latitude')],
                                   data=bee.test, proj4string = CRS("+proj=longlat +ellps=WGS84 +no_de

bee.clim <- extract(RS, bee.spdf, fun=mean, na.rm=TRUE, sp=TRUE)
bee.clim <- as.data.frame(bee.clim)

#...(rest of code in Data_Cleaning.rmd)

# Read-in from original script due to knitting error
bee_data_wTemp <- read_rds("Data/bee_data_wTemp")

```

Census Block Data

```
# Adding in Census Block Data
```

```
census_shp <- st_read("Data/CensusBlockGroups/CensusBlockGroups.shp")
```

```
## Reading layer 'CensusBlockGroups' from data source
##   '/Users/avawessel/Bee-Urbanization-BI410/Data/CensusBlockGroups/CensusBlockGroups.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 2970 features and 77 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -13881940 ymin: 5159794 xmax: -12964630 ymax: 5828411
## Projected CRS: WGS 84 / Pseudo-Mercator
```

```
census_sf <- census_shp %>%
  st_transform(crs = "EPSG:4326") %>%
  rename(latitude = INTPTLAT, longitude = INTPTLON) %>%
  dplyr::select(latitude, longitude, geometry, POP20, POPDENS20, COUNTY)
```

```
bee_sf <- st_as_sf(bee_data_wTemp,
  coords = c("longitude", "latitude"),
  crs = crs("EPSG:4326"))
census_sf <- st_cast(census_sf, "POLYGON")
```

```
## Warning in st_cast.sf(census_sf, "POLYGON"): repeating attributes for all
## sub-geometries for which they may not be constant
```

```
bee_with_census <- st_join(bee_sf, census_sf)
dim(bee_with_census)
```

```
## [1] 5116    31
```

```
saveRDS(bee_with_census, "Data/final_bee_dataset")
```

III. Analysis / Hypothesis Testing:

To test our hypothesis we used a linear mixed model, and we completed our analysis in the file `Bee_Analysis.Rmd`. We predicted the day of year of peak emergence for our bee species. We ran the model on all of the species together and on each species separately to see the specific effects of temperature and urbanization on each species. Our formula was: $DOY \sim 1 + \text{urbanizationLevel} + \text{springTemp} + \text{pop} + (1 | \text{species})$. So our fixed effects to predict the day of year were urbanization level, average spring temperature, and population. And our random effect was species to account for variation between species. We ran this model using the “lmer” function. We also used backward selection and the Variance Inflation Factor to make sure that the formula we were using was the most effective at predicting day of year and that we didn’t have any concerning covariance in our model. The output of our model found significance ($p < 0.05$) in each of our predicting effects, showing that when compared to rural, there was a significant difference in emergence time. Specifically the effect of urbanization resulted in an earlier emergence of about 15.8 days per °C compared to rural areas. Warmer spring temperatures resulted in an earlier emergence day by about 8.5 days per °C. We also found that *Bombus* spp. that on average have an earlier peak emergence were

more sensitive to changes in temperature, going along a gradient where the later peak emergence species were found to be gradually less sensitive to temperature changes.

Adding urbanization levels

```
complete_bee_data <- read_rds("Data/final_bee_dataset")

# function to add urbanization levels, rural, subrural, suburban, and urban
# split into 4 even categories using the quantile function
assign_urbanization_level <- function(data) {
  quantiles <- quantile(data$POP20, probs = seq(0, 1, 0.25), na.rm = TRUE)

  data <- data %>%
    mutate(urbanization_level = cut(POP20,
                                   breaks = quantiles,
                                   labels = c("rural", "subrural", "suburban", "urban"),
                                   include.lowest = TRUE))

  return(data)
}

complete_bee_data <- assign_urbanization_level(complete_bee_data)
saveRDS(complete_bee_data, "~/Bee-Urbanization-BI410/Data/complete_bee_data")
```

Finalizing the columns and data that we want to use for our model and running the model

```
# creating the column spring temp, averaging temperatures from Feb to June
complete_bee_data <- complete_bee_data %>%
  rowwise() %>%
  mutate(springTemp = mean(c_across(c("tmean_6", "tmean_7",
                                     "tmean_8", "tmean_9", "tmean_10")), na.rm = TRUE)) %>%
  ungroup()

# creating a dataframe with all of our variables that we want to use in the model
dat <- data.frame(
  DOY = complete_bee_data$DOY,
  springTemp = scale(complete_bee_data$springTemp),
  urbanizationLevel = complete_bee_data$urbanization_level,
  species = complete_bee_data$Species,
  pop = scale(complete_bee_data$POP20)
)

sum(is.na(dat))

## [1] 14

dat$springTemp <- as.numeric(dat$springTemp)

# Model
formula <- DOY ~ 1 + urbanizationLevel + springTemp + pop + (1 | species)
```

```
model1 <- lmer(formula, data = dat)
```

```
summary(model1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula
##   Data: dat
##
## REML criterion at convergence: 50825.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.5922 -0.6562 -0.0230  0.6675  5.1729
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   species  (Intercept)         428.2    20.69
##   Residual                  1224.2    34.99
## Number of obs: 5109, groups:  species, 4
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    181.0662    10.4405    3.0755  17.343 0.000361 ***
## urbanizationLevelsbrural   -5.5531     1.4619  5100.2632  -3.799 0.000147 ***
## urbanizationLevelsbrurban  -4.3300     1.6705  5100.7001  -2.592 0.009569 **
## urbanizationLevelurban    -15.8093     2.6026  5100.5548  -6.074 1.33e-09 ***
## springTemp             -8.5344     0.5243  5100.2921 -16.278 < 2e-16 ***
## pop                   3.8036     0.9437  5100.5545   4.030 5.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) urbnztnLvlsbrb urbnztnLvlsbrb urbnztnLvlsbrb urbnztnLvlsbrb
## urbnztnLvlsbrb -0.089
## urbnztnLvlsbrb -0.102  0.570
## urbnztnLvlsbrb -0.107  0.525      0.690
## springTemp      0.016 -0.108      -0.086      -0.164
## pop             0.082 -0.306      -0.545      -0.823      -0.035
```

```
step(model1)
```

```
## Backward reduced random-effect table:
##
##              Eliminated npar logLik   AIC   LRT Df Pr(>Chisq)
## <none>              8 -25413 50841
## (1 | species)      0    7 -25799 51612 772.24  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Backward reduced fixed-effect table:
## Degrees of freedom method: Satterthwaite
```

```
##
##               Eliminated Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## urbanizationLevel      0  54126   18042     3  5100.3   14.738 1.488e-09 ***
## springTemp             0 324371  324371     1  5100.3  264.974 < 2.2e-16 ***
## pop                    0  19884   19884     1  5100.6   16.243 5.652e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Model found:
## DOY ~ 1 + urbanizationLevel + springTemp + pop + (1 | species)
```

```
vif(model1)
```

```
##               GVIF Df GVIF^(1/(2*Df))
## urbanizationLevel 3.788848  3      1.248584
## springTemp        1.138555  1      1.067031
## pop               3.679843  1      1.918292
```

```
saveRDS(model1, "~/Bee-Urbanization-BI410/model1")
```

```
# calculating R^2 values
r2 <- r.squaredGLMM(model1)
r2
```

```
##               R2m      R2c
## [1,] 0.0562828 0.3008351
```

Applying the model to each of our 4 species

```
# Model applied to each species

select_species <- c("vosnesenskii", "melanopygus", "mixtus", "griseocollis")

model_results <- list()
r2_val <- list()

for (i in seq_along(select_species)) {

  formula_for_species <- DOY ~ 1 + urbanization_level + scale(springTemp) + scale(POP20) + (1 | COUNTY)

  single_species <- na.omit(complete_bee_data[complete_bee_data$Species == select_species[i], ])

  species_model <- lmer(formula_for_species, data = single_species)
  model_results[[select_species[i]]] <- species_model
  r2 <- r.squaredGLMM(species_model)
  r2_val[i] <- r2[2]
}

model_summaries <- lapply(model_results, summary)

model_summaries[1]
```



```
## $vosnesenskii
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula_for_species
## Data: single_species
##
## REML criterion at convergence: 27192
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.1521 -0.6593  0.0327  0.6746  4.7610
##
## Random effects:
## Groups Name Variance Std.Dev.
## COUNTY (Intercept) 186.3 13.65
## Residual 1406.8 37.51
## Number of obs: 2692, groups: COUNTY, 32
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 198.190 3.511 47.866 56.442 < 2e-16 ***
## urbanization_levelsubrural -13.316 2.603 2386.687 -5.115 3.38e-07 ***
## urbanization_levelsuburban -14.138 2.753 2657.817 -5.135 3.03e-07 ***
## urbanization_levelurban -18.097 4.150 2684.825 -4.361 1.34e-05 ***
## scale(springTemp) -10.491 1.088 1420.238 -9.641 < 2e-16 ***
## scale(POP20) 2.556 1.510 2685.990 1.692 0.0907 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) urbnztn_lvlsbrr urbnztn_lvlsbrb urbnztn_lvlr scl(T)
## urbnztn_lvlsbrr -0.473
## urbnztn_lvlsbrb -0.498 0.556
## urbnztn_lvlr -0.541 0.518 0.713
## scl(sprngT) 0.120 -0.042 -0.029 -0.163
## scal(POP20) 0.428 -0.306 -0.552 -0.822 0.039
```

```
model_summaries[2]
```

```
## $melanopygus
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula_for_species
## Data: single_species
##
## REML criterion at convergence: 5383.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.0795 -0.5595 -0.0019  0.3904  6.4945
##
## Random effects:
## Groups Name Variance Std.Dev.
## COUNTY (Intercept) 63.75 7.984
```

```
## Residual          800.84   28.299
## Number of obs: 566, groups: COUNTY, 24
##
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      151.344    4.511 133.750  33.552  <2e-16 ***
## urbanization_levelsubrural    1.336    4.819 362.665   0.277   0.782
## urbanization_levelsuburban   -4.219    4.710 482.816  -0.896   0.371
## urbanization_levelurban     -4.562    6.679 521.705  -0.683   0.495
## scale(springTemp)     -17.042    1.689 221.947 -10.090  <2e-16 ***
## scale(POP20)           3.400    2.164 558.585   1.571   0.117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##               (Intr) urbnztn_lvlsbrr urbnztn_lvlsbrb urbnztn_lvlr scl(T)
## urbnztn_lvlsbrr -0.697
## urbnztn_lvlsbrb -0.780  0.633
## urbnztn_lvlr    -0.811  0.616      0.760
## scl(sprngT)      0.306 -0.291     -0.211     -0.307
## scal(POP20)      0.543 -0.276     -0.495     -0.706     -0.001
```

```
model_summaries[3]
```

```
## $mixtus
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula_for_species
## Data: single_species
##
## REML criterion at convergence: 14575.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9260 -0.5747  0.0109  0.6169  3.7272
##
## Random effects:
## Groups Name Variance Std.Dev.
## COUNTY (Intercept) 170.8 13.07
## Residual 713.3 26.71
## Number of obs: 1545, groups: COUNTY, 33
##
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      155.8984    3.3549  60.5571  46.469  < 2e-16 ***
## urbanization_levelsubrural    1.6646    2.2786 1506.2666   0.731   0.4652
## urbanization_levelsuburban   16.4316    2.8959 1534.3003   5.674 1.66e-08 ***
## urbanization_levelurban      6.5760    4.8208 1536.3872   1.364   0.1727
## scale(springTemp)     -14.0885    0.9468  925.6481 -14.881  < 2e-16 ***
## scale(POP20)          -3.0419    1.6548 1536.4067  -1.838   0.0662 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##              (Intr) urbznztn_lvlsbrr urbznztn_lvlsbrb urbznztn_lvlr scl(T)
## urbznztn_lvlsbrr -0.491
## urbznztn_lvlsbrb -0.532  0.625
## urbznztn_lvlr    -0.575  0.578          0.752
## scl(sprngT)      0.229 -0.087          -0.061          -0.155
## scal(POP20)      0.485 -0.389          -0.635          -0.867          0.033
```

```
model_summaries[4]
```

```
## $griseocollis
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula_for_species
## Data: single_species
##
## REML criterion at convergence: 2980.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.9289 -0.6003 -0.1740  0.5238  3.3036
##
## Random effects:
## Groups Name Variance Std.Dev.
## COUNTY (Intercept) 682.5 26.13
## Residual 960.9 31.00
## Number of obs: 306, groups: COUNTY, 28
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)      204.799      7.596  53.775  26.960 < 2e-16 ***
## urbanization_levelsubrural    -2.690      6.932  285.353  -0.388  0.69822
## urbanization_levelsuburban    -5.743      7.371  299.612  -0.779  0.43647
## urbanization_levelurban     -29.469      9.584  299.102  -3.075  0.00230 **
## scale(springTemp)     -9.603      3.024  214.661  -3.176  0.00171 **
## scale(POP20)          4.838      3.461  297.500   1.398  0.16314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) urbznztn_lvlsbrr urbznztn_lvlsbrb urbznztn_lvlr scl(T)
## urbznztn_lvlsbrr -0.519
## urbznztn_lvlsbrb -0.499  0.542
## urbznztn_lvlr    -0.587  0.533          0.658
## scl(sprngT)      0.135 -0.118          -0.191          -0.245
## scal(POP20)      0.401 -0.209          -0.287          -0.720          0.094
```

```
r2_val
```

```
## [[1]]
## [1] 0.1973274
##
## [[2]]
## [1] 0.2950551
```

```
##
## [[3]]
## [1] 0.364078
##
## [[4]]
## [1] 0.4899885
```

```
saveRDS(model_summaries, "~/Bee-Urbanization-BI410/species_model_summaries")
```

IV. Visualizations:

These first two figures were made to summarize the raw data in two different ways.

Bombus Observations across Oregon:

```
# making our data into a shapefile for plotting
bombus_shp <- st_as_sf(complete_bee_data,
                      coords = c("Dec..Long.", "Dec..Lat."),
                      crs = st_crs("EPSG:4326"), na.fail = FALSE
)

# Calculating the same average spring temperatures that we used for the model
complete_bee_data <- bombus_shp %>%
  rowwise() %>%
  mutate(springTemp = mean(c_across(c("tmean_6", "tmean_7",
                                     "tmean_8", "tmean_9", "tmean_10")), na.rm = TRUE)) %>%
  ungroup()

# downloading in the Oregon base map
or <- map_data("county", "oregon")%>%
  dplyr::select(lon = long, lat, group, id = subregion)

# first need to create the extent from the or dataframe
min_long <- min(or$lon)
max_long <- max(or$lon)
min_lat <- min(or$lat)
max_lat <- max(or$lat)

or_extent <- c(xmin = min_long, xmax = max_long, ymin = min_lat, ymax = max_lat)

# create bounding box
bbox <- st_bbox(or_extent, crs = st_crs(bombus_shp))
# concert bbox to sf polygon
bbox_poly <- st_as_sfc(bbox)
# crop shapefile using bounding box
cropped_bombus_shp <- st_intersection(bombus_shp, bbox_poly)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```
# plotting!
ggplot() +
  geom_polygon(data=or, aes(x=lon, y=lat, group=group)) +
```

```
geom_sf(data = cropped_bombus_shp, aes(color = Species)) +
scale_color_viridis_d(name = "Species") +
ggtitle("Bombus Species Observations Across Oregon") +
labs(x = "longitude", y = "latitude") +
scale_color_manual(
  values = c("red", "white", "forestgreen", "violet"), # Custom colors for each species
  name = "Species"
)
```

```
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

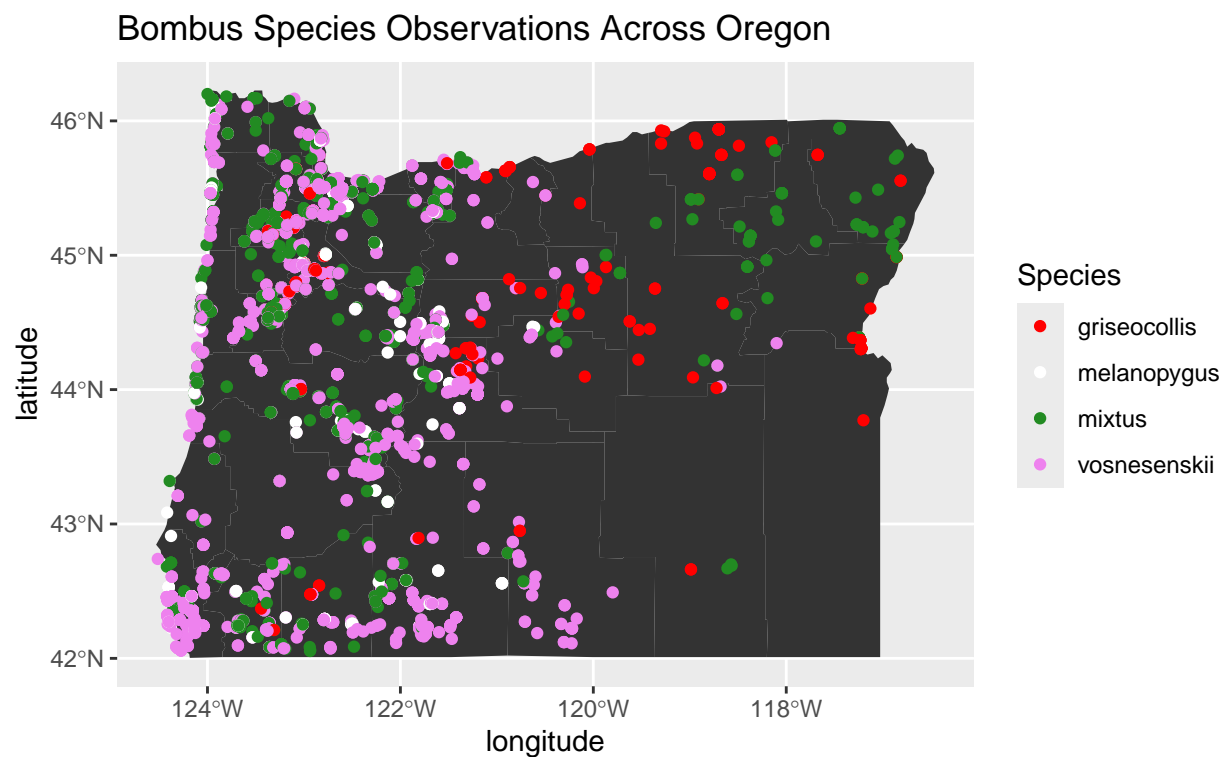


Figure 1: Map showing the average spring temperatures across Oregon, each point is a bee observation.

Census Blocks Plot

```
census_shp <- st_read("Data/CensusBlockGroups/CensusBlockGroups.shp")
```

```
## Reading layer 'CensusBlockGroups' from data source
##   '/Users/avawessel/Bee-Urbanization-BI410/Data/CensusBlockGroups/CensusBlockGroups.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 2970 features and 77 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -13881940 ymin: 5159794 xmax: -12964630 ymax: 5828411
## Projected CRS: WGS 84 / Pseudo-Mercator
```

```

census_csv <- read.csv("CensusBlockGroups.csv")

#ggplot() +
# geom_polygon(data=or, aes(x=lon, y=lat, group=group), fill = "tan")

census_shp <- census_shp %>%
  mutate(pop_bin = cut(POP20, breaks = c(0, 3000, 4000, 5000, 6000, Inf),
    labels = c("0-3k", "3k-4k", "4k-5k", "5k-6k", "6k+"),
    include.lowest = TRUE))

ggplot() +
  geom_sf(data = census_shp, aes(fill = POP20)) +
  scale_fill_gradient(low="beige", high = "brown", name = "Population by Census Block") +

  geom_sf(data = complete_bee_data, aes(color = springTemp)) +
  scale_color_gradient(low="blue", high="red", name = "Avg Spring Temp °C (Feb-June)") + ggtitle("All B")

  xlab("lon") + ylab("lat")

```

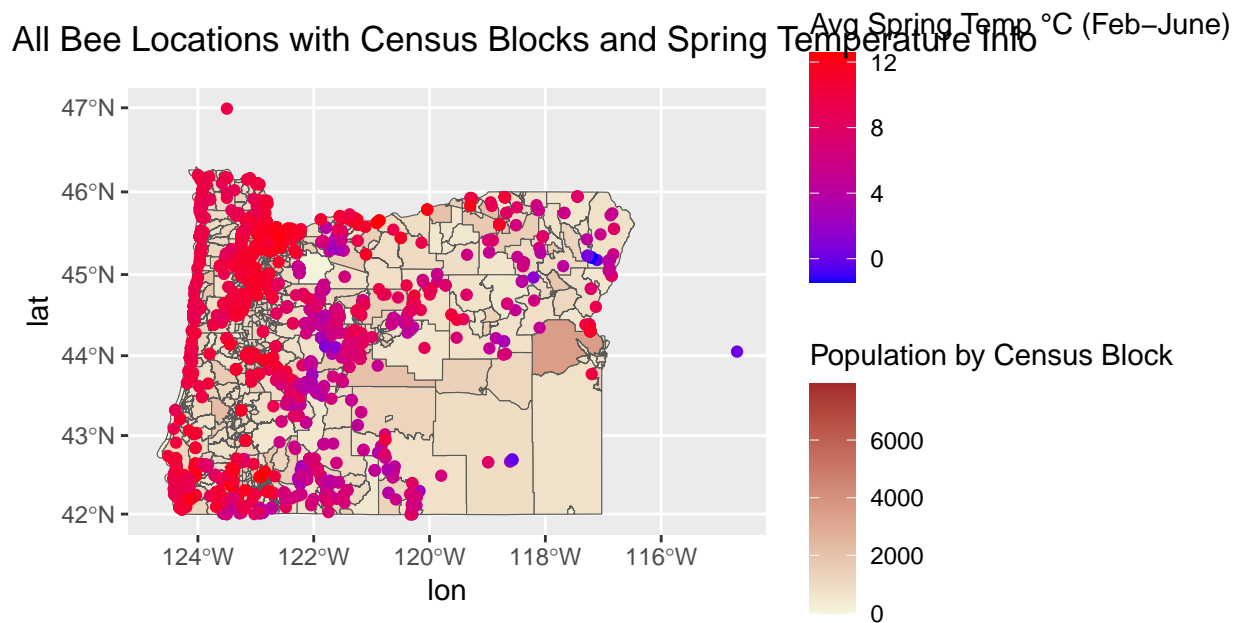


Figure 2: Map with average spring temperatures and population by Census Block across Oregon, each point is a bee observation.

The three figures below showcases the predicted linear trend for spring temperature, urbanization, and species specific response.

Creating Temperature variables and checking linear model. (model1):

```
complete_bee_data <- complete_bee_data %>%
  rowwise() %>%
  mutate(springTemp = mean(c_across(c("tmean_6", "tmean_7",
                                      "tmean_8", "tmean_9", "tmean_10")), na.rm = TRUE)) %>%
  ungroup()

anova(model1)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## urbanizationLevel  54126   18042      3 5100.3  14.738 1.488e-09 ***
## springTemp        324371  324371      1 5100.3 264.974 < 2.2e-16 ***
## pop               19884   19884      1 5100.6  16.243 5.652e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: formula
##   Data: dat
##
## REML criterion at convergence: 50825.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.5922 -0.6562 -0.0230  0.6675  5.1729
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   species  (Intercept)    428.2      20.69
##   Residual                    1224.2     34.99
## Number of obs: 5109, groups: species, 4
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    181.0662    10.4405    3.0755  17.343 0.000361 ***
## urbanizationLevelsubrural  -5.5531     1.4619 5100.2632  -3.799 0.000147 ***
## urbanizationLevelsuburban  -4.3300     1.6705 5100.7001  -2.592 0.009569 **
## urbanizationLevelurban    -15.8093     2.6026 5100.5548  -6.074 1.33e-09 ***
## springTemp        -8.5344     0.5243 5100.2921 -16.278 < 2e-16 ***
## pop                3.8036     0.9437 5100.5545   4.030 5.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) urbnztnLvlsbrr urbnztnLvlsbrb urbnztnLvlr sprngT
## urbnztnLvlsbrr -0.089
## urbnztnLvlsbrb -0.102  0.570
## urbnztnLvlr    -0.107  0.525      0.690
## springTemp      0.016 -0.108     -0.086     -0.164
## pop             0.082 -0.306     -0.545     -0.823     -0.035
```

Spring Temperature vs DOY Visualization:

```
tempDOY <- ggpredict(model1, terms = c("springTemp"))

tempDOY_plot <- ggplot(tempDOY, aes(x = x, y = predicted)) +
  geom_point(data = complete_bee_data,
            aes(x = scale(springTemp), y = DOY, color = "Observed Data"),
            alpha = 0.5) +
  geom_line(aes(color = "Model Prediction"), linewidth = 1) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.2) +
  scale_color_manual(values = c("Observed Data" = "green3", "Model Prediction" = "darkgreen")) +
  labs(x = "Scaled Spring Temperature", y = "DOY", color = "Legend", fill = "Legend",
       title = "Effect of Temperature on Day of Emergence") +
  theme_minimal()

#calculating slope
slope_tempDOY <- diff(tempDOY$predicted) / diff(tempDOY$x)
print(slope_tempDOY)
```

```
## [1] -8.53439 -8.53439 -8.53439 -8.53439 -8.53439 -8.53439 -8.53439
```

```
tempDOY_plot
```

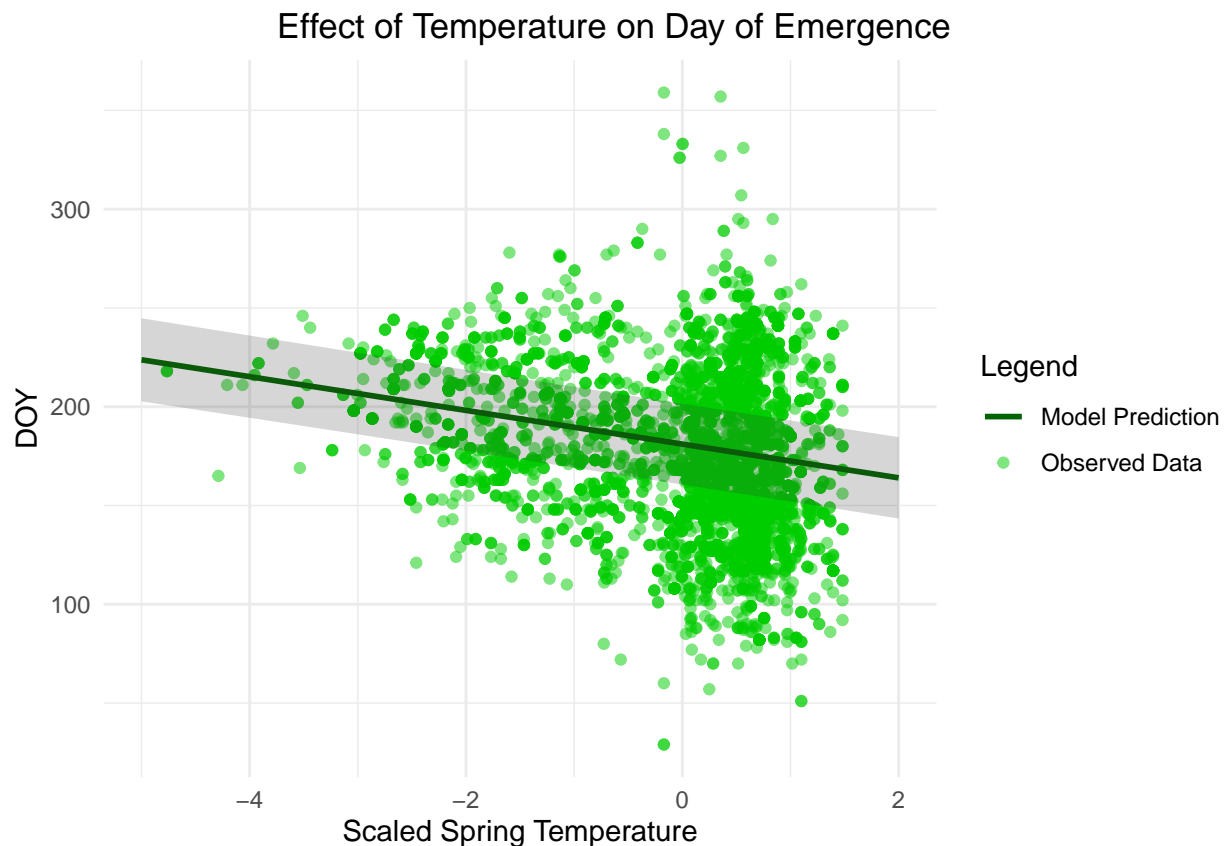


Figure 3: Average peak *Bombus* spp. emerge earlier with increasing spring temperatures. Green line represents model prediction with 95% confidence interval. Green dots represent observed data.

Urbanization Level vs DOY Visualization:

```
urbanizationDOY <- ggpredict(model1, terms = c("urbanizationLevel"))

urbanizationDOY_plot <- ggplot(urbanizationDOY, aes(x = x, y = predicted)) +
  geom_line(aes(group = 1), size = 1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high),
               width = 0.2, size = 0.8, color = "darkgrey") +
  geom_point(size = 3, color = "forestgreen") +
  labs(x = "Urbanization Level",
       y = "Predicted DOY",
       title = "Effect of Urbanization on Day of Emergence") +
  theme_minimal()
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

urbanizationDOY_plot

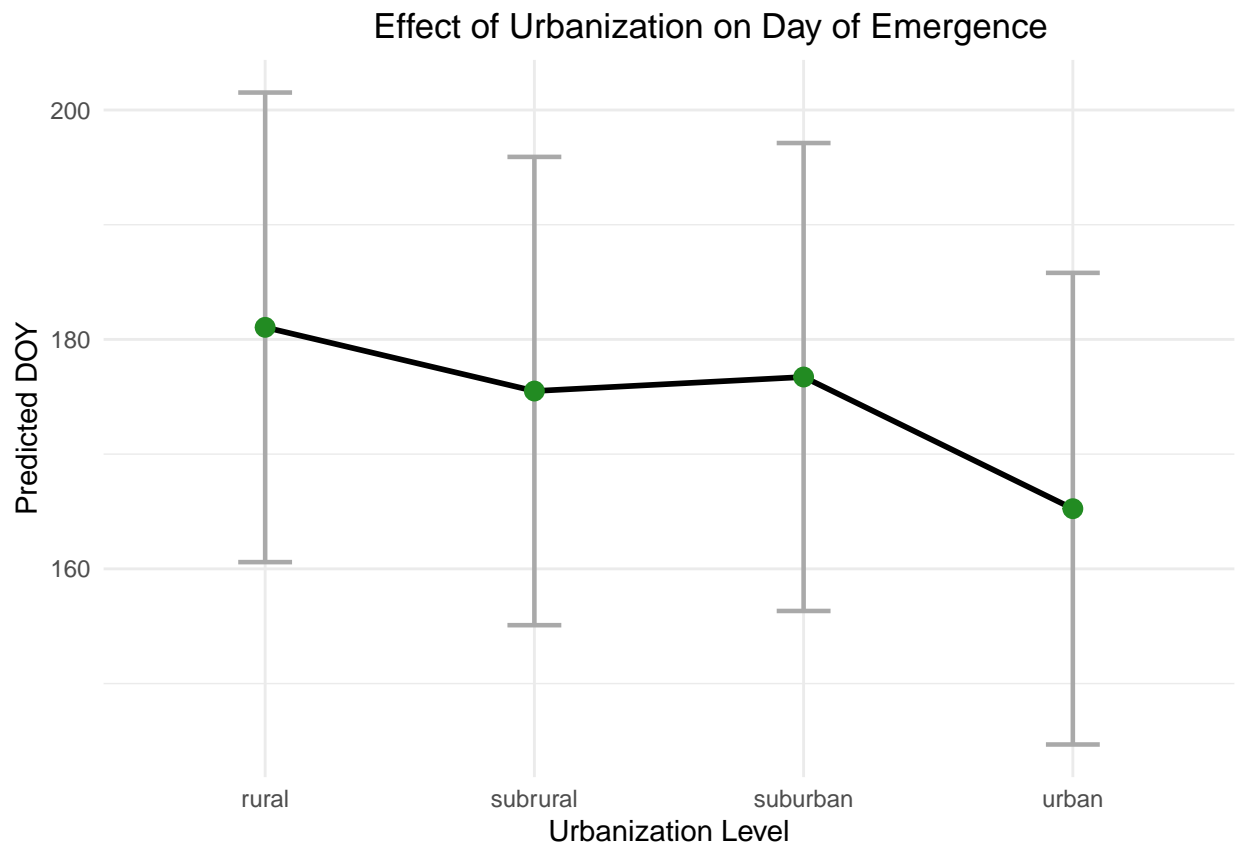


Figure 4: Dense urban areas lead to an earlier average peak emergence when compared to rural areas. Green dots represent slopes for each urbanization level. Gray bars represent 95% confidence interval.

Species Specific Figure:

```

results <- data.frame(
  species = character(),
  slope = numeric(),
  conf.low = numeric(),
  conf.high = numeric(),
  stringsAsFactors = FALSE
)

for (species_name in names(model_summaries)) {
  coefficients <- model_summaries[[species_name]][["coefficients"]]

  if ("scale(springTemp)" %in% rownames(coefficients)) {
    slope <- coefficients["scale(springTemp)", "Estimate"]
    std_error <- coefficients["scale(springTemp)", "Std. Error"]
    df <- coefficients["scale(springTemp)", "df"]

    t_critical <- qt(0.975, df = df)
    conf.low <- slope - t_critical * std_error
    conf.high <- slope + t_critical * std_error

    results <- rbind(
      results,
      data.frame(
        species = species_name,
        slope = slope,
        conf.low = conf.low,
        conf.high = conf.high,
        stringsAsFactors = FALSE
      )
    )
  }
}

ggplot(results, aes(x = reorder(species, slope), y = slope)) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2, color = "gray50") +
  geom_point(size = 3, color = "forestgreen") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red", size = 0.8) +
  labs(
    y = "Emergence Sensitivity (days/°C)",
    x = "Species",
    title = "Species-Specific Emergence Sensitivity to Temperature"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

```

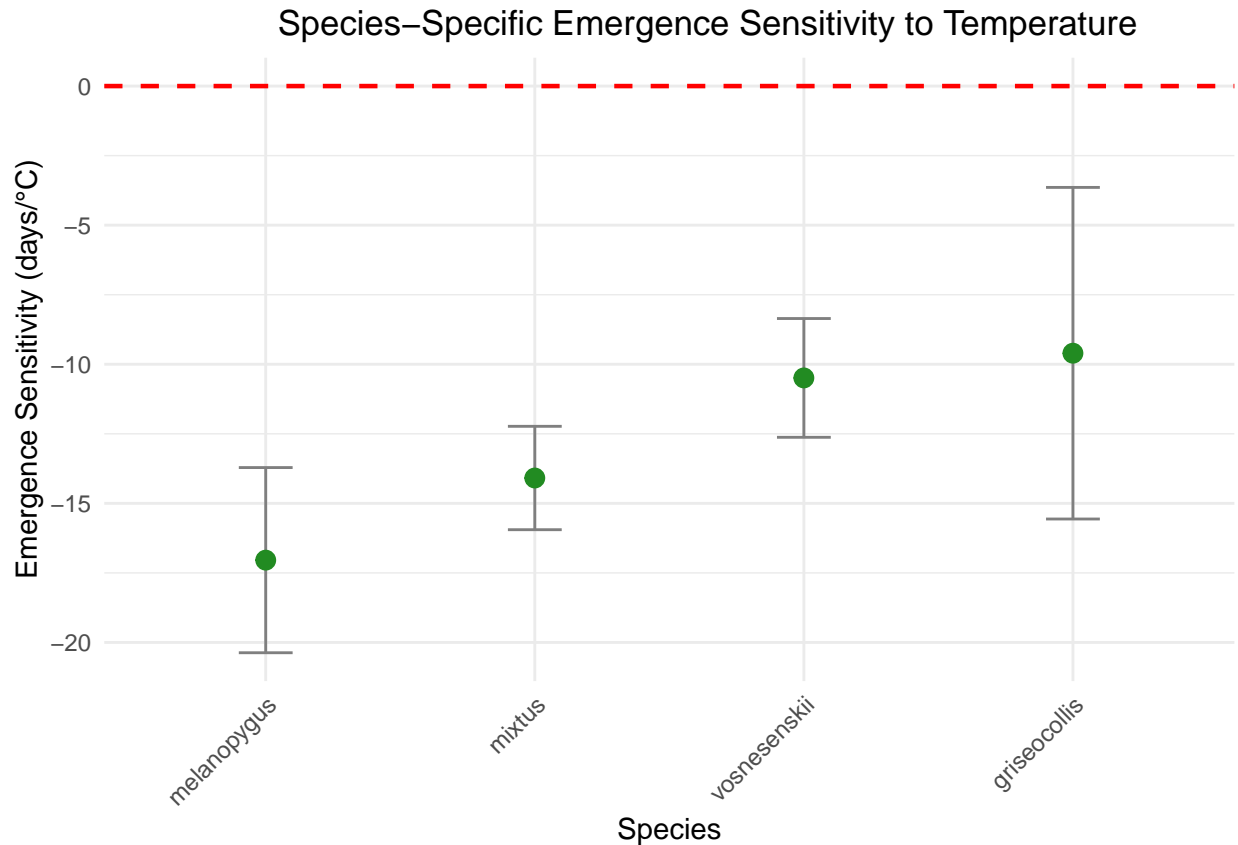


Figure 5: *Melanopygus* saw the greatest emergence sensitivity, with *mixtus*, *vosnesenskii*, and *griseocollis* following. Green dots represent slopes for each species. Black bars represent 95% confidence intervals for each estimate. Red dashed line at 0 represent no emergence sensitivity to temperature.

Summary stats for species specific response comparison:

Mean DOY and total number of observations for all species.

```
doy_avg <- complete_bee_data %>%
  group_by(Species) %>%
  summarise(doy_avg = mean(DOY),
            observations = sum(n())) %>%
  arrange(doy_avg, desc = TRUE)
```

```
doy_avg
```

```
## Simple feature collection with 4 features and 3 fields
## Geometry type: MULTIPOINT
## Dimension: XY
## Bounding box: xmin: -124.515 ymin: 41.994 xmax: -114.69 ymax: 46.99
## Geodetic CRS: WGS 84
## # A tibble: 4 x 4
##   Species    doy_avg observations      geometry
##   <chr>      <dbl>      <int>      <MULTIPOINT [°]>
## 1 melanopygus    150.        566 ((-123.91 46.1), (-123.951 46.16), (-124 46~
## 2 mixtus         165.       1549 ((-123.5 46.99), (-123.91 46.1), (-123.855 ~
## 3 vosnesenskii   189.       2695 ((-123.855 46.087), (-123.855 46.086), (-12~
## 4 griseocollis   193.        306 ((-122.841 45.745), (-123.082 45.644), (-12~
```

V. Conclusion:

We found that more urbanized areas resulted in an earlier peak emergence, by around 15.8 days per °C. This is likely caused by the Urban Heat Island Effect which causes warmer microclimates due to less green spaces and a higher concentration of concrete in more urbanized areas. The effect of temperature also causes earlier emergence dates, across all species, by around 8.5 days per °C (more specifics on model predictions can be found in the `Bee_Analysis.rmd` script). This lines up with our understanding of temperature as an environmental cue that regulates phenology, directly or through changes in plant phenological sensitivity. The effect of urbanization and temperature on bees might indirectly affecting them, as studies have shown that plant phenology is highly synchronized with temperature. Further research would need to be conducted assess the relationship between the plant the bee was found on, the climate, and the bees species itself. As for the species-specific phenological sensitivity, *Bombus* species that have an earlier average peak emergence were more sensitive to the warming changes in temperature, the reliance on temperature could make early-season species more vulnerable to climate change with premature emergence that could cause a mismatch with floral resource availability. This result aligns with our initial hypothesis as *melanopygus*, the earliest emerging species also is the most sensitive to temperature. As you move to later emerging *Bombus* species, the less sensitive to temperature they appear to be.

Things that could be causing bias include the fact that urban areas are more populated and easier to access, so there might be an overestimation of the effects of urbanization due to the increased number of observations in urban areas. We also have to consider, due to our usage of quantiles, just because something was labeled urban or rural, there is a chance it might not fit the true description of what a rural or urban environment looks like. The choice of creating categorical values for urbanization allowed us to easily test the two extremes, but possibly ignored what the true defined urbanization levels are for a given observation. Additionally, there are more common species with more observations: *Bombus vosnesenskii* had 2,695 observations, *Bombus mixtus* had 1549, *Bombus melanopygus* had 566, and *Bombus griseocollis* had 306. Because of the substantially lower observations for *melanopygus* and *griseocollis*, the possibility of skewing their data within the overall model was feasible. Though we created a species specific model to account for variations in observation, this wasn't taken into account within the overall model. More specifically, we could be underestimating the effects on species like *griseocollis* and overestimating the effects of species like *vosnesenskii*.

VI. References:

Ahrné, K., Bengtsson, J., & Elmqvist, T. (2009). Bumble Bees (*Bombus* spp) along a Gradient of Increasing Urbanization. *PLoS ONE*, 4(5), e5574. <https://doi.org/10.1371/journal.pone.0005574> McCabe, L. M., Aslan, C. E., & Cobb, N. S. (2022). Decreased bee emergence along an elevation gradient: Implications for climate change revealed by a transplant experiment. *Ecology*, 103(2), e03598. <https://doi.org/10.1002/ecy.3598> Harrison, T., & Winfree, R. (2015). Urban drivers of plant-pollinator interactions. *Functional Ecology*, 29(7), 879–888. <https://doi.org/10.1111/1365-2435.12486> Glaum, P., Simao, M.-C., Vaidya, C., Fitch, G., & Iulinao, B. (2017). Big city *Bombus*: using natural history and land-use history to find significant environmental drivers in bumble-bee declines in urban development. *Royal Society Open Science*, 4(5), 170156. <https://doi.org/10.1098/rsos.170156> Goulson, D., Lye, G. C., & Darvill, B. (2008). Decline and Conservation of Bumble Bees. *Annual Review of Entomology*, 53(1), 191–208. <https://doi.org/10.1146/annurev.ento.53.103106.093454> Stenkovski, M., Pearse, W. D., Griffin, S. R., Pardee, G. L., Gibbs, J., Griswold, T., Neff, J. L., Oram, R., Rightmyer, M. G., Sheffield, C. S., Wright, K., Inouye, B. D., Inouye, D. W., & Irwin, R. E. (2020). Bee phenology is predicted by climatic variation and functional traits. *Ecology Letters*, 23(11), 1589–1598. <https://doi.org/10.1111/ele.13583> Blasi, M., Carrié, R., Fägerström, C., Svensson, E., & Persson, A. S. (2023). Historical and citizen-reported data show shifts in bumblebee phenology over the last century in Sweden. *Biodiversity and Conservation*, 32(5), 1523–1547. <https://doi.org/10.1007/s10531-023-02563-5> Prestele, R., Brown, C., Polce, C., Maes, J., & Whitehorn, P. (2021). Large variability in response to projected climate and land-use changes among European bumblebee species. *Global Change Biology*, 27(19), 4530–4545. <https://doi.org/10.1111/gcb.15780> Li, D., Barve, N., Brenskelle, L., Earl, K., Barve, V., Belitz, M. W., Doby, J., Hantak, M. M., Oswald, J. A., Stucky, B. J., Walters, M., & Guralnick, R. P. (2021). Climate, urbanization, and species traits interactively drive flowering duration. *Global Change Biology*, 27(4),

892–903. <https://doi.org/10.1111/gcb.15461> iNaturalist community. Observations of *Bombus* spp. from Oregon, recorded between 2017 and 2024. Data exported from <https://www.inaturalist.org> on December 4, 2024.