Data Analytics Final Project

An Analysis of Trends in Bird Obervations and Temperature in North Carolina, 2010-2020

Cate Jaffe

4/26/2021

Setup

```
##### Library #####
library(tidyverse)
library(ggplot2)
library(scales)
library(auk) # eBird Package
library(agricolae)
library(lubridate)
library(corrplot)
library(colormap)
library(ggridges)
library(RColorBrewer)
library(cowplot)


# set working directory for knitting
knitr::opts_knit$set( root.dir =
  "/Users/Kate/Documents/1.Spring 2021/JaffeWellbaumFrear_ENV872_EDA_FinalProject",
                 tidy.opts = list(width.cutoff = 60),
                 tidy = TRUE)
# check wd
#getwd()

# set ggplot theme
mytheme <- theme_light( base_size = 14) +
  theme( axis.text = element_text( color = "#222222ff"),
         legend.position = "top",
         # margins (top,right,bottom,left)
         axis.title.x = element_text( color = "black",
                                    margin = margin(20,0,0,0)),
         axis.title.y = element_text( color = "black",
                                    margin = margin(0,20,0,0)))
theme_set(mytheme)
```

Import Data

```
# import bird data
# eBird data is in text format, package "auk" used to convert to dataframe
woodduck <- read_ebd(
  "./Data/Raw/NorthCarolina/ebd_US-NC_wooduc_relFeb-2021/ebd_US-NC_wooduc_relFeb-2021.txt")
rwbbird <- read_ebd(
  "./Data/Raw/NorthCarolina/ebd_US-NC_rewbla_relMar-2021/ebd_US-NC_rewbla_relMar-2021.txt")
osprey <- read_ebd(
  "./Data/Raw/NorthCarolina/ebd_US-NC_osprey_relFeb-2021/ebd_US-NC_osprey_relFeb-2021.txt")

# import temperature data
temp <- read.csv("./Data/Raw/NorthCarolina/NCTemperature20102021.csv") %>% unique()
```

Data Cleaning

```
### create merged dataset of all bird data
allBirds <- bind_rows(woodduck, rwbbird, osprey)

### clean
allBirds <- allBirds %>%
  # filter dates to date range of interest: 2010 - 2021
  filter(year(observation_date) > 2009 & year(observation_date) < 2022) %>%
  # select only columns of interest
  select(common_name:observation_count, state, county,
                                     latitude:time_observations_started, protocol_type,
                                     duration_minutes:number_observers,
                                     all_species_reported) %>%
    # change "X" value in observation_count to 1 (X represents "present" in eBird)
    # change to numeric variable after converting to X
    mutate(observation_count = as.numeric(replace(observation_count, observation_count == 'X', '1')),
           # add column which divides # observations per minute observation
           # this controls for birding "effort" which was much higher in later years (2015-2020)
           observations_per_min = observation_count/duration_minutes,
           # add column for "Year-Month" using floor_date from lubridate
           Year_Month = floor_date(observation_date, unit = "month"),
           # add binary presence column
           Presence = 1) %>%
  # Some "observations per minute" values are NA or Inf where
  # the duration of observation was 0 minutes or was missing.
  # Exclude NA and Inf values
  filter( is.na(observations_per_min) == FALSE) %>%
  filter( observations_per_min != Inf)
  # Removing NA and Inf values removes about 10,000 observations, or ~6% of the data.

### check distribution of new "observations per minute" variable
summary(allBirds$observations_per_min)
```

```
##     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    0.000     0.015     0.036     0.659     0.111 13333.333
```

```
# plot regular distribution:
# heavy right skew in data
regDist <-
ggplot(allBirds, aes(x=observations_per_min)) +
  geom_density() +
  labs(x = "Bird Observations per Minute")
# log transformation looks more normal
logDist <-
ggplot(allBirds, aes(x=log10(observations_per_min))) +
  geom_density()  +
  labs(x = "Log10( Bird Observations per Minute )")

plot_grid(regDist,
          logDist,
          # align both horizontal and vertical axis
          nrow = 1, align = 'hv')
```

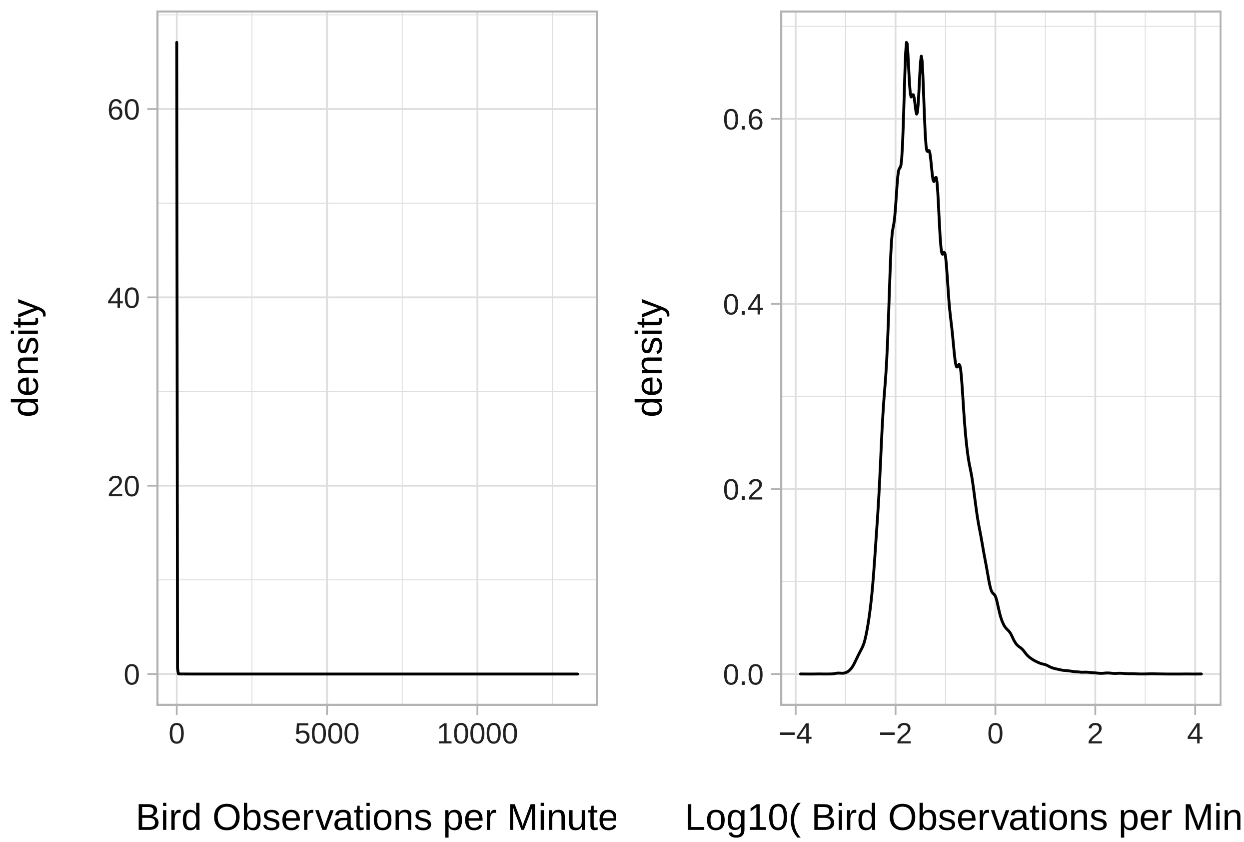


Figure 1: A comparison of Bird observations, raw data (left) and log transformed (right). Log transformed data have a more normal distribution

```
### Create grouped dataset, with observations per minute averaged by year-month
allBirds_YMgrouped <-
  allBirds %>%
  group_by(common_name, Year_Month) %>%
  summarize(state = first(state),
            observation_count_sum = sum(observation_count),
```

```
            # sum presence column: how many times did bird appear on a checklist?
            presence_count = sum(Presence),
            # average obsv/min: how many times was the bird observed, corrected for effort
            observation_per_min_avg = mean(observations_per_min)
            )
```

```
#  create temperature dataset grouped by month & year
temp_YM <- temp %>%
  # omit NAs
  na.omit() %>%
  # make Year-Month column
  mutate( Year_Month = ydm((paste0(Year,"-01-",Month)))) %>%
  group_by(Year_Month) %>%
  # take statewide average
  summarise(AvgMonthlyTemp_Statewide = mean(AvgMonthlyTemp)) %>%
  # re-add month and year columns
  mutate(Month = month(Year_Month),
         Year = year(Year_Month))
```

Data Exploration

```
### Birds, ungrouped, uncorrected for effort
# basic density plot across all years, faceted
birdsRaw <-
ggplot(allBirds, aes(x = observation_date, fill = common_name, color = common_name)) +
  geom_density(alpha = .2) +
  facet_wrap(vars(common_name), nrow = 1, ncol = 3) +
  theme( legend.position = "none",
         axis.text.x = element_text(angle = 45,
                                    hjust = 1)) +
  # disable scientific notation in y axis
  # set color and fill manually
  scale_color_brewer( palette = "Dark2") +
  scale_fill_brewer( palette = "Dark2") +
  # make labels legible
  labs(x = " ",
       y = "Density of Bird Observations",
       title = "Comparision of Bird Observation Data:",
       subtitle = "corrected and uncorrected for birding effort")

### Birds, grouped by year-month, corrected for effort
birdsEffort <-
ggplot(allBirds_YMgrouped %>% filter(year(Year_Month) != 2021),
       aes(x = Year_Month, y = log10(observation_per_min_avg),
           group = common_name, color = common_name)) +
  geom_line(alpha = .8) +
  facet_wrap(vars(common_name), nrow = 1, ncol = 3) +
  theme( legend.position = "none",
         axis.text.x = element_text(angle = 45,
                                    hjust = 1),
         axis.title.x = element_text(vjust = -2)) +
```

```
  # disable scientific notation in y axis
  # set color and fill manually
  scale_color_brewer( palette = "Dark2") +
  scale_fill_brewer( palette = "Dark2") +
  # make labels legible
  labs(x = "Observation Date",
       y = "Average Birds per \nMinute of Observation (log10)")

plot_grid(birdsRaw,
          birdsEffort,
          nrow = 2, align = 'v')
```

```
# Temperature trends by year-month groups
ggplot(temp_YM %>% filter(Year != 2021),
       aes(x = Year_Month, y = AvgMonthlyTemp_Statewide,
           color = AvgMonthlyTemp_Statewide)) +
  geom_point(alpha = .4) +
  geom_line(alpha = .8) +
  scale_x_date(breaks = "6 months",
               date_labels = "%b %Y") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1),
        legend.position = "none") +
  scale_color_colormap(colormap = "plasma") +
  labs(y = "Average Monthly Temperature (F)",
       x = "Date")
```

Join Bird and Temperature Data

```
birdsTemp_YM <- full_join(allBirds_YMgrouped, temp_YM,
                          by = "Year_Month")
```

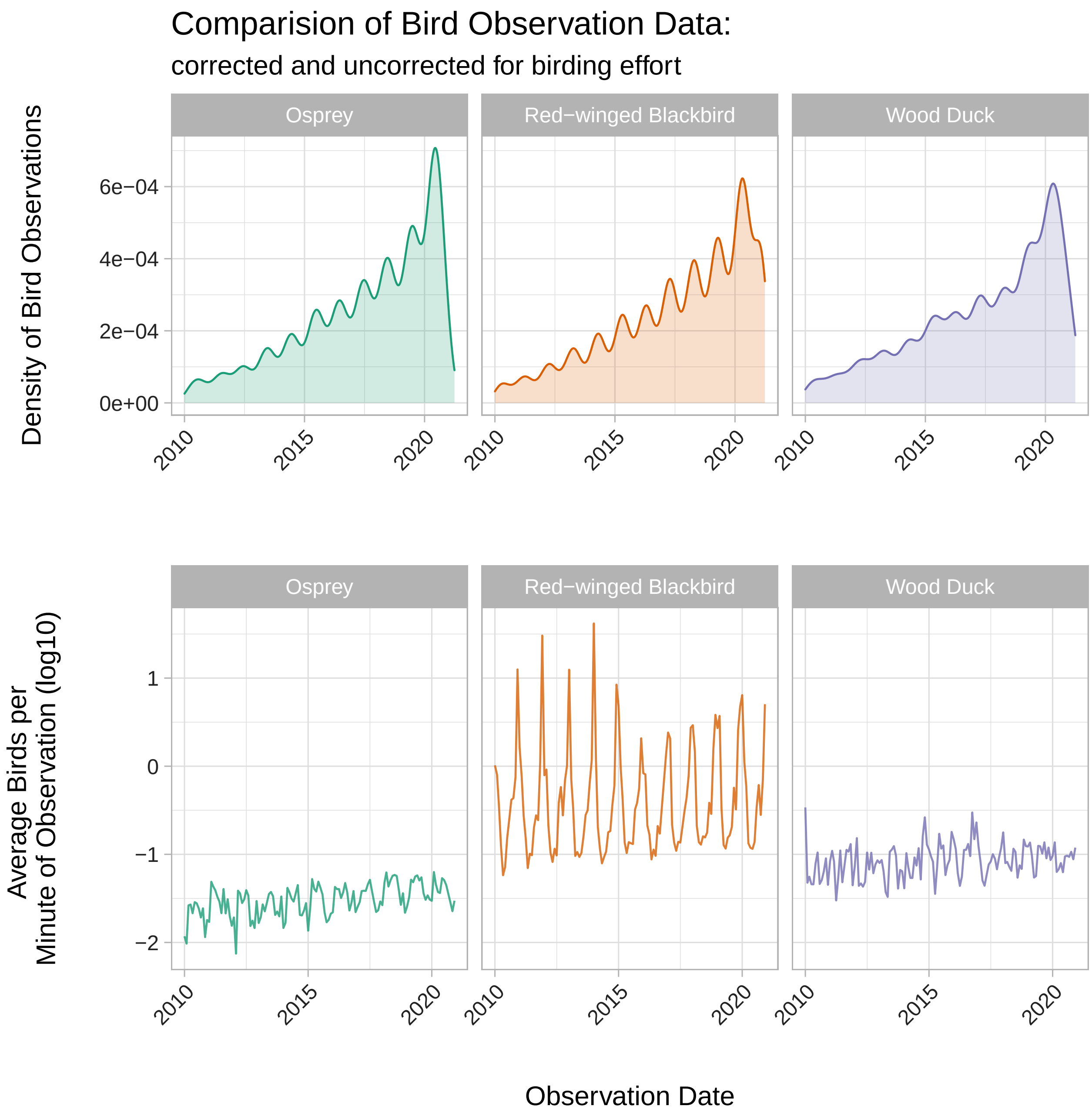


Figure 2: The top plot is raw bird observation data, the bottom plot is observations per minute of observation.

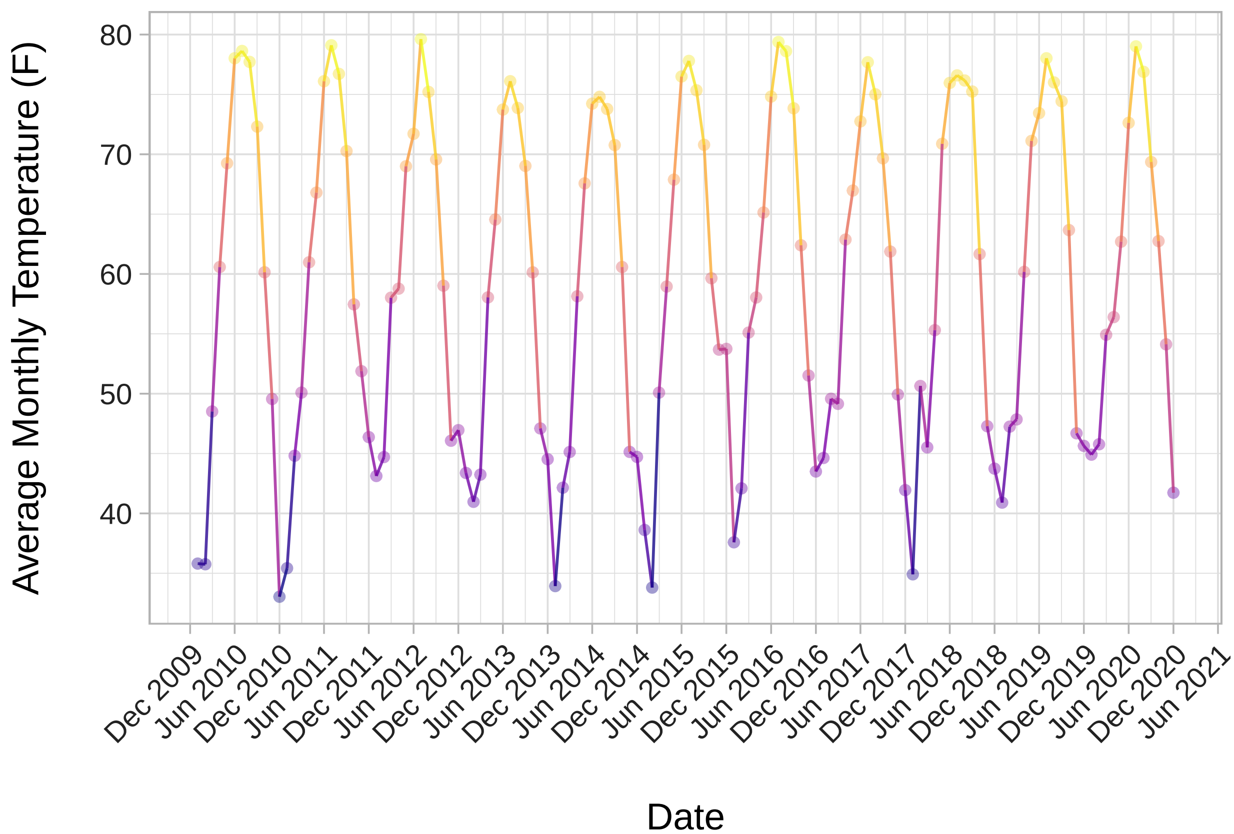


Figure 3: Average Monthly Temperature for the Study Period: 2010-2020

Analysis: Bird Observations & Temperature

```
#  observations per minute to overtime, by temperature
ggplot(birdsTemp_YM %>% filter(Year != 2021),
       aes(x = Year_Month, y = log10(observation_per_min_avg),
           color = AvgMonthlyTemp_Statewide)) +
  geom_line(lwd = .8) +
  facet_wrap(vars(common_name), nrow = 3, scale = "free") +
  theme(legend.position = "right") +
  scale_color_colormap(colormap = "plasma") +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  labs(y = "Average Birds per Minute of Observation (log10)",
       x = "Year",
       color = "Average Monthly \nTemperature (F)",
       title = "Bird Observations in North Carolina",
       subtitle = "2010 - 2020")
```

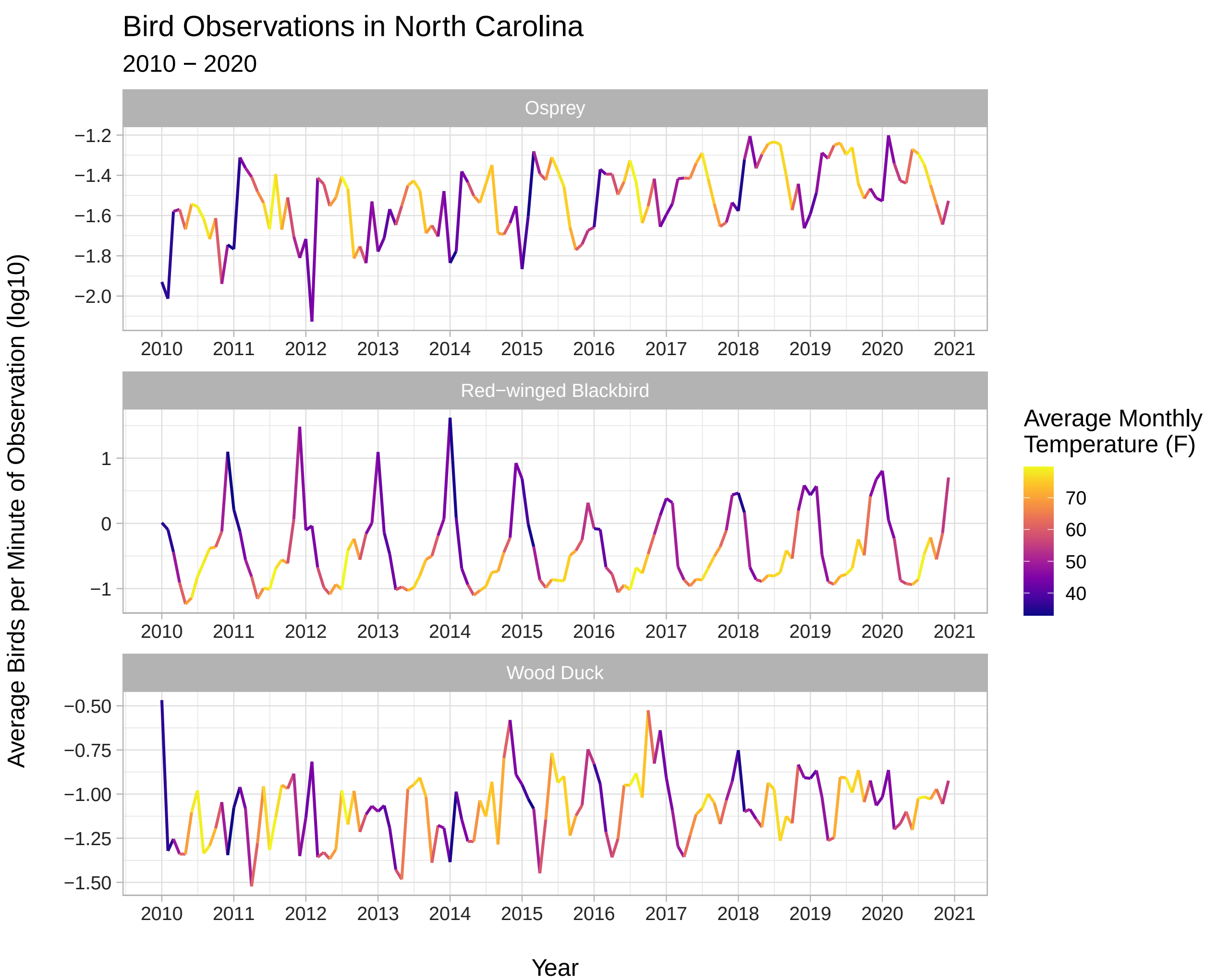


Figure 4: Bird Observations by Temperature in North Carolina

Linear Regression

```
### Osprey ###########
# Observations per minute vs. Temperature, Year, and Month
# as an lm()
lm_osprey <- lm(data = birdsTemp_YM %>% filter(common_name == "Osprey"),
                observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
                  # have to convert month and year to factors
                  as.factor(Year) + as.factor(Month))
# summarize output
summary(lm_osprey)
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
##     as.factor(Year) + as.factor(Month), data = birdsTemp_YM %>%
##     filter(common_name == "Osprey"))
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0182578 -0.0044408  0.0000751  0.0034037  0.0240445
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.0014202  0.0083056   0.171 0.864545
## AvgMonthlyTemp_Statewide  0.0001840  0.0002039   0.902 0.368933
## as.factor(Year)2011       0.0088992  0.0028450   3.128 0.002253 **
## as.factor(Year)2012       0.0047090  0.0028565   1.649 0.102094
## as.factor(Year)2013       0.0055760  0.0028320   1.969 0.051474 .
## as.factor(Year)2014       0.0078695  0.0028345   2.776 0.006462 **
## as.factor(Year)2015       0.0100104  0.0028465   3.517 0.000636 ***
## as.factor(Year)2016       0.0129689  0.0028570   4.539 1.45e-05 ***
## as.factor(Year)2017       0.0128259  0.0028571   4.489 1.77e-05 ***
## as.factor(Year)2018       0.0228491  0.0028417   8.041 1.11e-12 ***
## as.factor(Year)2019       0.0211754  0.0028648   7.392 3.02e-11 ***
## as.factor(Year)2020       0.0186843  0.0028551   6.544 1.96e-09 ***
## as.factor(Year)2021       0.0181480  0.0054667   3.320 0.001223 **
## as.factor(Month)2         0.0103753  0.0029355   3.534 0.000599 ***
## as.factor(Month)3         0.0205126  0.0036007   5.697 1.03e-07 ***
## as.factor(Month)4         0.0135575  0.0049270   2.752 0.006936 **
## as.factor(Month)5         0.0105254  0.0064065   1.643 0.103252
## as.factor(Month)6         0.0146578  0.0077221   1.898 0.060295 .
## as.factor(Month)7         0.0148735  0.0083585   1.779 0.077928 .
## as.factor(Month)8         0.0135240  0.0079874   1.693 0.093255 .
## as.factor(Month)9        -0.0003171  0.0071321  -0.044 0.964621
## as.factor(Month)10        0.0003866  0.0052466   0.074 0.941396
## as.factor(Month)11        0.0018421  0.0035510   0.519 0.604966
## as.factor(Month)12        0.0043840  0.0030775   1.425 0.157123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006934 on 110 degrees of freedom
## Multiple R-squared:  0.7363, Adjusted R-squared:  0.6812
```

```
## F-statistic: 13.35 on 23 and 110 DF,  p-value: < 2.2e-16
```

```r
# stepwise selection of most parsimonious model
step(lm_osprey)
```

```
## Start:  AIC=-1310.75
## observation_per_min_avg ~ AvgMonthlyTemp_Statewide + as.factor(Year) +
##     as.factor(Month)
##
##                            Df Sum of Sq       RSS     AIC
## - AvgMonthlyTemp_Statewide  1 0.0000391 0.0053282 -1311.8
## <none>                                  0.0052891 -1310.8
## - as.factor(Month)         11 0.0056056 0.0108947 -1235.9
## - as.factor(Year)          11 0.0060664 0.0113555 -1230.4
##
## Step:  AIC=-1311.77
## observation_per_min_avg ~ as.factor(Year) + as.factor(Month)
##
##                    Df Sum of Sq       RSS     AIC
## <none>                          0.0053282 -1311.8
## - as.factor(Year)  11 0.0064252 0.0117535 -1227.8
## - as.factor(Month) 11 0.0083842 0.0137124 -1207.1
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ as.factor(Year) + as.factor(Month),
##     data = birdsTemp_YM %>% filter(common_name == "Osprey"))
##
## Coefficients:
##         (Intercept)  as.factor(Year)2011  as.factor(Year)2012
##            0.008475             0.009155             0.005053
## as.factor(Year)2013  as.factor(Year)2014  as.factor(Year)2015
##            0.005505             0.007741             0.010279
## as.factor(Year)2016  as.factor(Year)2017  as.factor(Year)2018
##            0.013317             0.013174             0.023073
## as.factor(Year)2019  as.factor(Year)2020  as.factor(Year)2021
##            0.021572             0.019019             0.018259
##   as.factor(Month)2    as.factor(Month)3    as.factor(Month)4
##            0.011076             0.022406             0.017134
##   as.factor(Month)5    as.factor(Month)6    as.factor(Month)7
##            0.015667             0.021105             0.021937
##   as.factor(Month)8    as.factor(Month)9   as.factor(Month)10
##            0.020229             0.005551             0.004316
##  as.factor(Month)11   as.factor(Month)12
##            0.003658             0.005245
```

```r
# post analysis Tukey Test, only run with categorical explanatory variables.
# this post analysis test will reveal which groups of years and/or
# months had similar observations of birds

# make aov for tukey test
aov_osprey <- aov(data = birdsTemp_YM %>% filter(common_name == "Osprey"),
                observation_per_min_avg ~
```

```r
                  as.factor(Year) + as.factor(Month))

# create and print group labels - for Month
osprey.groups.yr <-
  HSD.test(aov_osprey, "as.factor(Year)", group = TRUE)
osprey.groups.yr$groups
```

```
##      observation_per_min_avg groups
## 2018              0.04390803      a
## 2019              0.04240748     ab
## 2020              0.03985486    abc
## 2016              0.03415242    bcd
## 2017              0.03400990    bcd
## 2021              0.03227183   bcde
## 2015              0.03111446    cde
## 2011              0.02999051     de
## 2014              0.02857630     de
## 2013              0.02634011     de
## 2012              0.02588886     de
## 2010              0.02083548      e
```

```r
# create and print group labels - for Year
osprey.groups.month <-
  HSD.test(aov_osprey, "as.factor(Month)", group = TRUE)
osprey.groups.month$groups
```

```
##    observation_per_min_avg groups
## 3               0.04250742      a
## 7               0.04203857      a
## 6               0.04120636     ab
## 8               0.04033054     ab
## 4               0.03723535     ab
## 5               0.03576862     ab
## 2               0.03172998     bc
## 9               0.02565231     cd
## 12              0.02534643     cd
## 10              0.02441694     cd
## 11              0.02375897     cd
## 1               0.02065403      d
```

```r
### Red winged Blackbird ###########
lm_rwbb <- lm(data = birdsTemp_YM %>% filter(common_name == "Red-winged Blackbird"),
               observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
               as.factor(Year) + as.factor(Month))

summary(lm_rwbb)
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
##     as.factor(Year) + as.factor(Month), data = birdsTemp_YM %>%
##     filter(common_name == "Red-winged Blackbird"))
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.988 -1.384  0.231  0.694 31.649
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              10.15604    5.09522   1.993  0.04869 *
## AvgMonthlyTemp_Statewide -0.09020    0.12536  -0.719  0.47336
## as.factor(Year)2011       1.65246    1.75286   0.943  0.34787
## as.factor(Year)2012      -0.80818    1.75988  -0.459  0.64697
## as.factor(Year)2013      -0.07711    1.74484  -0.044  0.96483
## as.factor(Year)2014       2.96019    1.74637   1.695  0.09287 .
## as.factor(Year)2015      -0.43233    1.75374  -0.247  0.80573
## as.factor(Year)2016      -0.82571    1.76022  -0.469  0.63992
## as.factor(Year)2017      -0.43772    1.76026  -0.249  0.80408
## as.factor(Year)2018      -0.35463    1.75080  -0.203  0.83986
## as.factor(Year)2019       0.09076    1.76497   0.051  0.95908
## as.factor(Year)2020       0.05109    1.75905   0.029  0.97688
## as.factor(Year)2021      -1.52446    2.83887  -0.537  0.59234
## as.factor(Month)2        -4.94836    1.80832  -2.736  0.00723 **
## as.factor(Month)3        -5.36088    2.18781  -2.450  0.01583 *
## as.factor(Month)4        -4.87605    3.03564  -1.606  0.11106
## as.factor(Month)5        -4.14219    3.94630  -1.050  0.29616
## as.factor(Month)6        -3.48072    4.75558  -0.732  0.46576
## as.factor(Month)7        -3.17058    5.14699  -0.616  0.53915
## as.factor(Month)8        -3.24959    4.91878  -0.661  0.51021
## as.factor(Month)9        -3.50882    4.39268  -0.799  0.42612
## as.factor(Month)10       -4.49868    3.23243  -1.392  0.16679
## as.factor(Month)11       -4.89397    2.18681  -2.238  0.02722 *
## as.factor(Month)12        0.32407    1.89283   0.171  0.86437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.272 on 111 degrees of freedom
## Multiple R-squared:  0.3189, Adjusted R-squared:  0.1777
## F-statistic: 2.259 on 23 and 111 DF,  p-value: 0.002624
```

```
step(lm_rwbb)
```

```
## Start:  AIC=413.66
## observation_per_min_avg ~ AvgMonthlyTemp_Statewide + as.factor(Year) +
##     as.factor(Month)
##
##                            Df Sum of Sq    RSS    AIC
## - as.factor(Year)          11    159.45 2185.5 401.88
## - AvgMonthlyTemp_Statewide  1      9.45 2035.5 412.28
## <none>                                  2026.0 413.66
## - as.factor(Month)         11    409.42 2435.5 416.50
##
## Step:  AIC=401.88
## observation_per_min_avg ~ AvgMonthlyTemp_Statewide + as.factor(Month)
##
##                            Df Sum of Sq    RSS    AIC
```

```
## - AvgMonthlyTemp_Statewide  1     29.89 2215.4 401.72
## <none>                                 2185.5 401.88
## - as.factor(Month)         11    414.79 2600.3 403.34
##
## Step:  AIC=401.72
## observation_per_min_avg ~ as.factor(Month)
##
##                    Df Sum of Sq    RSS    AIC
## <none>                          2215.4 401.72
## - as.factor(Month) 11    759.09 2974.5 419.49
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ as.factor(Month), data = birdsTemp_YM %>%
##     filter(common_name == "Red-winged Blackbird"))
##
## Coefficients:
##       (Intercept)   as.factor(Month)2   as.factor(Month)3   as.factor(Month)4
##           6.62255            -5.29185            -6.31114            -6.49269
##  as.factor(Month)5   as.factor(Month)6   as.factor(Month)7   as.factor(Month)8
##          -6.52626            -6.50477            -6.49687            -6.40010
##  as.factor(Month)9  as.factor(Month)10  as.factor(Month)11  as.factor(Month)12
##          -6.24891            -6.28816            -5.64729             0.03865
```

```r
# stepwise selection suggests a model with only month is the most parsimonious
lm_rwbb_monthOnly <- lm(data = birdsTemp_YM %>% filter(common_name == "Red-winged Blackbird"),
                observation_per_min_avg ~ as.factor(Month))

summary(lm_rwbb_monthOnly)
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ as.factor(Month), data = birdsTemp_YM %>%
##     filter(common_name == "Red-winged Blackbird"))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.833 -0.192 -0.018  0.026 35.083
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.62255    1.22513   5.406 3.22e-07 ***
## as.factor(Month)2  -5.29185    1.73259  -3.054 0.002766 **
## as.factor(Month)3  -6.31114    1.73259  -3.643 0.000396 ***
## as.factor(Month)4  -6.49269    1.77153  -3.665 0.000366 ***
## as.factor(Month)5  -6.52626    1.77153  -3.684 0.000343 ***
## as.factor(Month)6  -6.50477    1.77153  -3.672 0.000358 ***
## as.factor(Month)7  -6.49687    1.77153  -3.667 0.000363 ***
## as.factor(Month)8  -6.40010    1.77153  -3.613 0.000440 ***
## as.factor(Month)9  -6.24891    1.77153  -3.527 0.000591 ***
## as.factor(Month)10 -6.28816    1.77153  -3.550 0.000547 ***
## as.factor(Month)11 -5.64729    1.77153  -3.188 0.001818 **
## as.factor(Month)12  0.03865    1.77153   0.022 0.982631
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.244 on 123 degrees of freedom
## Multiple R-squared:  0.2552, Adjusted R-squared:  0.1886
## F-statistic: 3.831 on 11 and 123 DF,  p-value: 9.415e-05
```

```
# post analysis Tukey Test, only run with categorical explanatory variables.
# this post analysis test will reveal which groups of years and/or
# months had similar observations of birds

aov_rwbb <- aov(data = birdsTemp_YM %>% filter(common_name == "Red-winged Blackbird"),
                observation_per_min_avg ~ as.factor(Month))

# create and print group labels - for Month
rwbb.groups.month <-
  HSD.test(aov_osprey, "as.factor(Month)", group = TRUE)
rwbb.groups.month$groups
```

```
##    observation_per_min_avg groups
## 3               0.04250742      a
## 7               0.04203857      a
## 6               0.04120636     ab
## 8               0.04033054     ab
## 4               0.03723535     ab
## 5               0.03576862     ab
## 2               0.03172998     bc
## 9               0.02565231     cd
## 12              0.02534643     cd
## 10              0.02441694     cd
## 11              0.02375897     cd
## 1               0.02065403      d
```

```
### Wood Duck ###########
lm_duck <- lm(data = birdsTemp_YM %>% filter(common_name == "Wood Duck"),
                observation_per_min_avg ~ AvgMonthlyTemp_Statewide + Year + Month)

summary(lm_duck)
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
##     Year + Month, data = birdsTemp_YM %>% filter(common_name ==
##     "Wood Duck"))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.070410 -0.030131 -0.008806  0.024946  0.255294
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -4.1484198  2.4889519  -1.667   0.0980 .
## AvgMonthlyTemp_Statewide -0.0007204  0.0002990  -2.409   0.0174 *
```

```
## Year                        0.0021180  0.0012351   1.715   0.0887 .
## Month                       0.0025514  0.0011854   2.152   0.0332 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04602 on 130 degrees of freedom
## Multiple R-squared:  0.07764,    Adjusted R-squared:  0.05636
## F-statistic: 3.648 on 3 and 130 DF,  p-value: 0.01445
```

```
step(lm_duck)
```

```
## Start:  AIC=-821.12
## observation_per_min_avg ~ AvgMonthlyTemp_Statewide + Year + Month
##
##                             Df Sum of Sq     RSS     AIC
## <none>                                   0.27537 -821.12
## - Year                       1 0.0062293 0.28159 -820.13
## - Month                      1 0.0098127 0.28518 -818.43
## - AvgMonthlyTemp_Statewide  1 0.0122926 0.28766 -817.27
```

```
##
## Call:
## lm(formula = observation_per_min_avg ~ AvgMonthlyTemp_Statewide +
##     Year + Month, data = birdsTemp_YM %>% filter(common_name ==
##     "Wood Duck"))
##
## Coefficients:
##              (Intercept)  AvgMonthlyTemp_Statewide                      Year
##               -4.1484198                -0.0007204                 0.0021180
##                    Month
##                0.0025514
```

```
# stepwise selection indicates no variables should be removed from model
# No Tukey HSD is run for this model because it includes a continuous numerical variable (Temperature)
```

Discussion of Linear Regressions:

Osprey: The most parisimonious model for the Osprey observations (corrected for effort) included year and month as explanatory variables, but not temperature. This model explained **68%** of the variation in Osprey observations. Like the Blackbird, the spring and summar months had similar observations (Group ab: April, May, June, August) which were statistically different than the mean of observations in fall and winter months (Group cd: September, October, November, December).

Red-Winged Blackbird: The most parisimonious model for the Blackbird observations (corrected for effort) included only month as an explanatory variable. This model explained only 18.9% of the variation in Blackbird observations. Like the Osprey, the spring and summar months had similar observations of Blackbird (Group ab: April, May, June, August) which were statistically different than the mean of observations in fall and winter months (Group cd: September, October, November, December).

Wood Duck: The most parisimonious model for the Wood Duck observations (corrected for effort) included temperature, year and month as explanatory variables. Together, these variables explain only 5.6% of the variation in wood duck observations. For every 1 degree *increase* in temperature (with month and year held constant) we would expect the observations of wood ducks (per minute of observation) to *decrease* by .00072 duck per minute observation. There is likely some other variable, not measured here, explaining the variation in wood duck observations in North Carolina between 2010 and 2020.

Overall, the Wood Duck appears to be the only bird of the three examined in this study for which average monthly temperature has a statistically significant relationship with bird abundance (corrected for observation effort). The Month of the Year was included in the final model for all three birds, and observations tended to be most similar in non-migratory periods (namely late spring to summer and late fall to winter).

Since the linear regression revealed that across species, month tended to have a strong relationship with bird observation - and that the mean observations per month tended to be similar between seasonal groups of months (for instance, spring months grouped together in group ab of the Tukey test), we visualized how bird observations might vary by "season" and temperature.

```
# add "seasonal" dummy variable to dataset and summarize temperature.
birdsTemp_season <- birdsTemp_YM %>%
  mutate( season =
            if_else( Month %in% c(3, 4, 5), "spring",
                     if_else(Month %in% c(6,7,8), "summer",
                             if_else( Month %in% c(9,10,11), "fall",
                                      if_else( Month %in% c(12, 1, 2), "winter", "NA")))))


# PLot
ggplot(birdsTemp_season,
       aes(x = AvgMonthlyTemp_Statewide, y = log10(observation_per_min_avg), color = season)) +
  geom_point(alpha = .8) +
  stat_ellipse(alpha = .4) +
  facet_wrap(vars(common_name), nrow = 1, scales = "free") +
  scale_color_manual(values=c('#e75f2dff', '#008066ff', '#ffd42bff', '#0b6ca8ff'))  +
  theme(legend.title = element_blank(), legend.position = "bottom",
        axis.title.x = element_text(vjust = -1),
        axis.title.y = element_text(vjust = 3)) +
  labs(y = "Average Birds per \nMinute of Observation (log10)",
       x = "Average Monthly Temperature (F)",
       title = "Bird Observations vs. Temperature, \nby Season in North Carolina",
       subtitle = "2010 - 2020")
```

Figure 5: The relationship between bird observations and temperature, broken down by season and bird.

Analysis: Date of First and Last Annual Observations

```r
allBirdsTemp <- full_join(allBirds, temp_YM,
                          by = "Year_Month")

#Summarize the number of observations by week
allBirds_week <- allBirdsTemp %>%
  # make column of "week"
  mutate(Week = floor_date(observation_date, unit = "week")) %>%
  group_by(Week, Year_Month, common_name, state) %>%
  summarize(observations_per_min_avg = mean(observations_per_min),
            Presence = sum(Presence),
            observation_count_sum = sum(observation_count),
            AvgWeeklyTemp_F = mean(AvgMonthlyTemp_Statewide)) %>%
  mutate(Month = month(Year_Month),
         Year = year(Year_Month)) %>%
  filter(Year %in% c(2010:2020))

# Plot
ggplot(allBirds_week, aes(x=month(Week), y=log10(observation_count_sum), color=common_name)) +
  geom_smooth(method='loess', se=F, lwd=0.8) +
  scale_color_manual(values=c('#1E88E5', '#D81B60', '#004D40')) +
#  scale_y_continuous(n.breaks = 3) +
  scale_x_continuous(expand = c(0,0),
                     breaks = c(1, 4, 8, 12),
                     labels = c('Jan','Apr', 'Aug', 'Dec')) +
  labs(x='', y='', color='',
       title='Seasonal Stay by Species Over Time',
       subtitle = 'North Carolina') +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        panel.spacing.x = unit(2, 'lines')) +
  facet_grid(vars(year(Year_Month)), vars(common_name))
```

Discussion of Seasonal Stays by each species

All birds appear to reside in North Carolina year round. Osprey appear most populus in the summer months and almost absent in the winter, with populations potentially increasing overtime between 2010 and 2020. The opposite pattern appears to be true for Blackbird, which is more abundant in the summers and less in the winters, and is generally the most populus of the three birds examined here. Wood Duck appears from this plot to be least seaonsal, and to be present in constant numbers (at State scale) throughout the year.

To further investigate seasonal and long-term (2010-2020) trends, we conducted a time-series analysis:

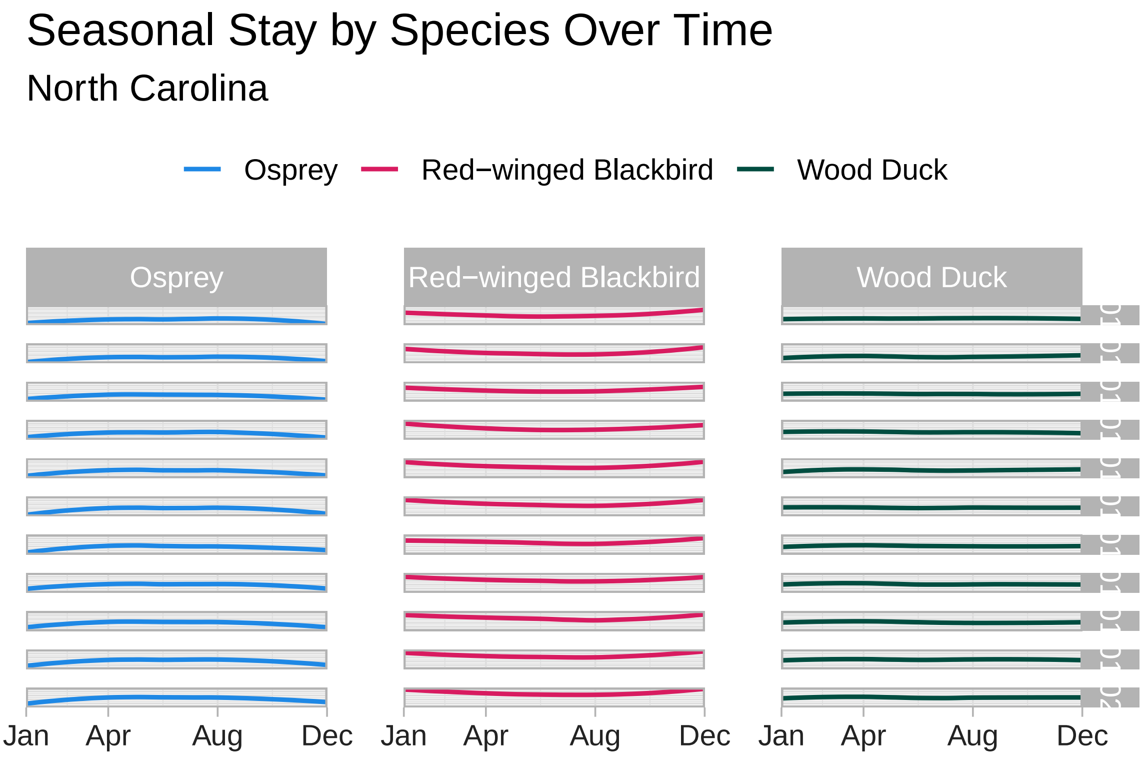


Figure 6: Seasonal Stays of Birds in North Carolina

Analysis: Time Series

```
# Subset data for running time series
birdsTemp_YM_ospr <- birdsTemp_YM %>% filter(common_name == "Osprey" & Year %in% c(2010:2020))

birdsTemp_YM_rwbb <- birdsTemp_YM %>% filter(common_name == "Red-winged Blackbird" & Year %in% c(2010:2

birdsTemp_YM_wodu <- birdsTemp_YM %>% filter(common_name == "Wood Duck" & Year %in% c(2010:2020))


## Osprey
osprey.ts <- ts(birdsTemp_YM_ospr$observation_per_min_avg,
                             start = c(2010,1), frequency = 12)
osprey.ts.decomposed <- decompose(osprey.ts, type = "multiplicative")

# Seasonal Mann Kendall
monthly_ospr_trend <- Kendall::SeasonalMannKendall(osprey.ts)
summary(monthly_ospr_trend)
```

```
## Score =  356 , Var(Score) = 1980
## denominator =  660
## tau = 0.539, 2-sided pvalue =1.3323e-15
```

```
## Red-winged Blackbird
rwbb.ts <- ts(birdsTemp_YM_rwbb$observation_per_min_avg,
```

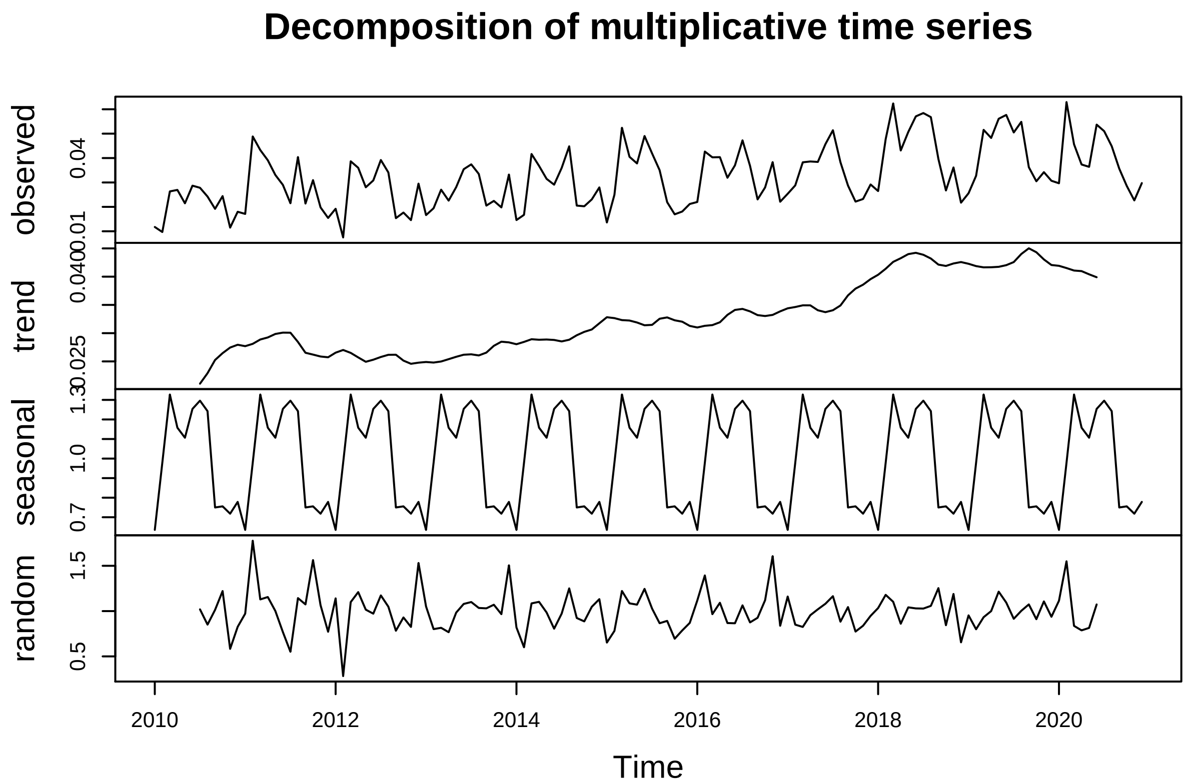
```
                              start = c(2010,1), frequency = 12)

rwbb.ts.decomposed <- decompose(rwbb.ts, type = "multiplicative")


monthly_rwbb_trend <- Kendall::SeasonalMannKendall(rwbb.ts)
summary(monthly_rwbb_trend)
```

```
## Score =  178 , Var(Score) = 1980
## denominator =  660
## tau = 0.27, 2-sided pvalue =6.3275e-05
```

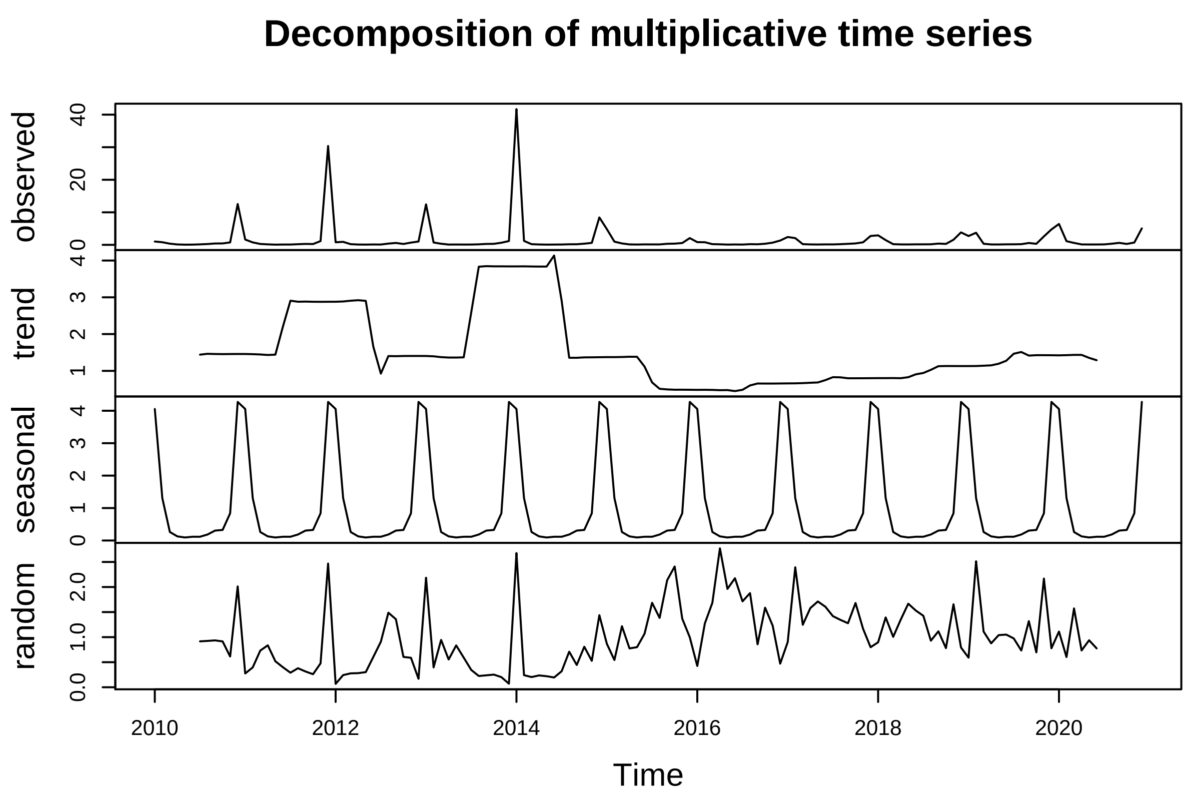
```
## Wood Duck

wodu.ts <- ts(birdsTemp_YM_wodu$observation_per_min_avg,
                              start = c(2010,1), frequency = 12)
wodu.ts.decomposed <-  decompose(wodu.ts, type = "multiplicative")


monthly_wodu_trend <- Kendall::SeasonalMannKendall(wodu.ts)
summary(monthly_wodu_trend)
```

```
## Score =  154 , Var(Score) = 1980
## denominator =  660
## tau = 0.233, 2-sided pvalue =0.00053839
```

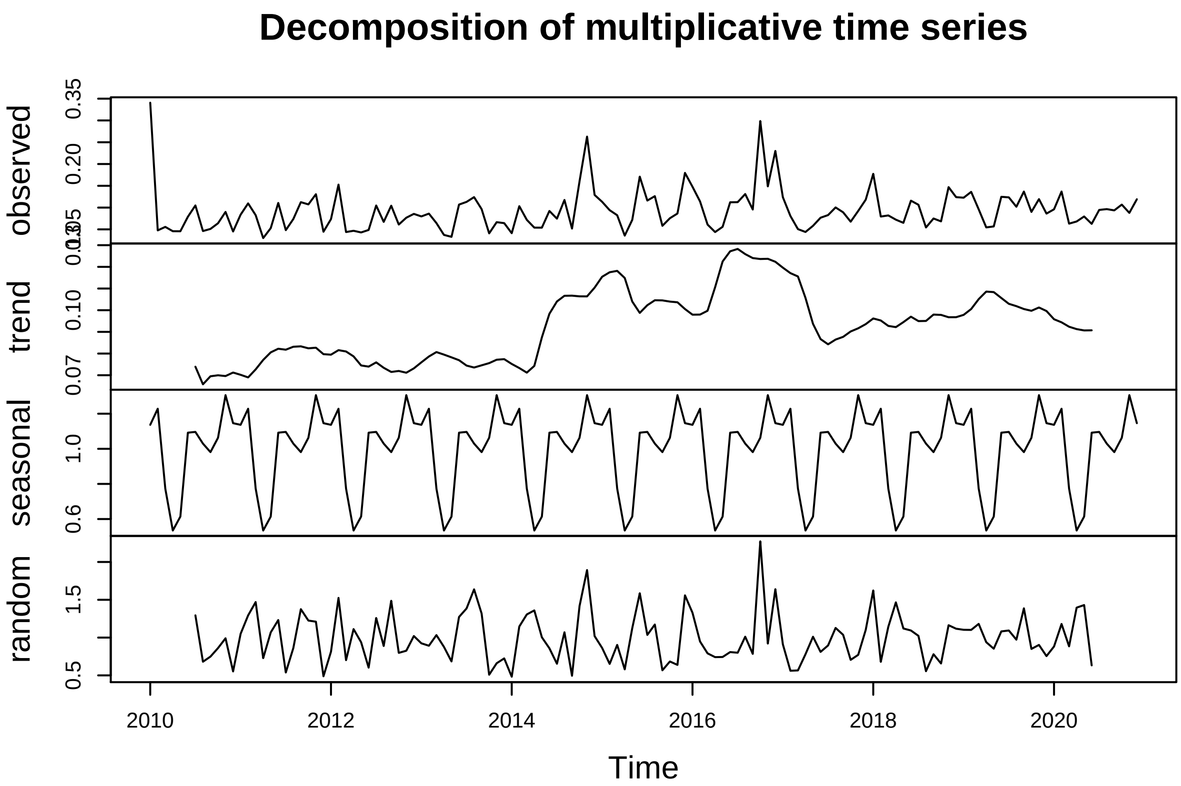
```
# plot time series

#### Osprey
plot(osprey.ts.decomposed)
```




```r
### Red-winged Blackbird
plot(rwbb.ts.decomposed)
```



```
### Wood Duck
plot(wodu.ts.decomposed)
```



Discussion of Time Series Results: