Massachusetts_Analysis

Emma Wellbaum

4/22/2021

Set up

Import and Summarize State Temperature Data

> Temperature data was obtained from the National Centers for Environmental Information Global Historical Climate Network Daily dataset. The GHCN-Daily contains daily climate summaries, including minimum and maximum temperature, from land surface stations across the world. Data was downloaded via the National Atmospheric and Oceanic Administration API using the rnoaa package and a NOAA API token. The source code used to extract and pre-process the data is located in the project repository file: './Code/TemperatureAPI.Rmd'

```
# Import state temperature data, downloaded from the National Atmospheric and
# Oceanic Administration using the rnoaa package
# The source code used to download and pre-process the temperature data is
MA_temp <- read.csv('./Data/Raw/Massachusetts/MassTemperature20102021.csv') %>% unique()

# Summarize monthly temperature data for the entire state
MA_temp <- MA_temp %>%
  mutate(Month_Year = paste0(Month,'-',Year)) %>%
  mutate(Month_Year = my(Month_Year)) %>%
  drop_na(AvgMonthlyTemp) %>%
  group_by(Month_Year) %>%
  summarize(AvgMonthlyTemp_F = mean(AvgMonthlyTemp)) %>%
  filter(year(Month_Year) %in% c(2010:2020))
```

Import Bird Observation Data

> Bird observation data was downloaded from the Cornell Lab of Ornithology eBird Database (https://ebird.org/).

```
###Import bird observation data downloaded from eBird
#Osprey
OSPR <- read_ebd(
  "./Data/Raw/Massachusetts/ebd_US-MA_osprey_relFeb-2021/ebd_US-MA_osprey_relFeb-2021.txt")

#Red-winged Blackbird
#Due to large file size, RWBL data was imported in two parts and then combined
RWBL_early <- read_ebd(
  "./Data/Raw/Massachusetts/ebd_US-MA_rewbla_201001_201412_relMar-2021/ebd_US-MA_rewbla_201001_201412_r
RWBL_late <- read_ebd(
```

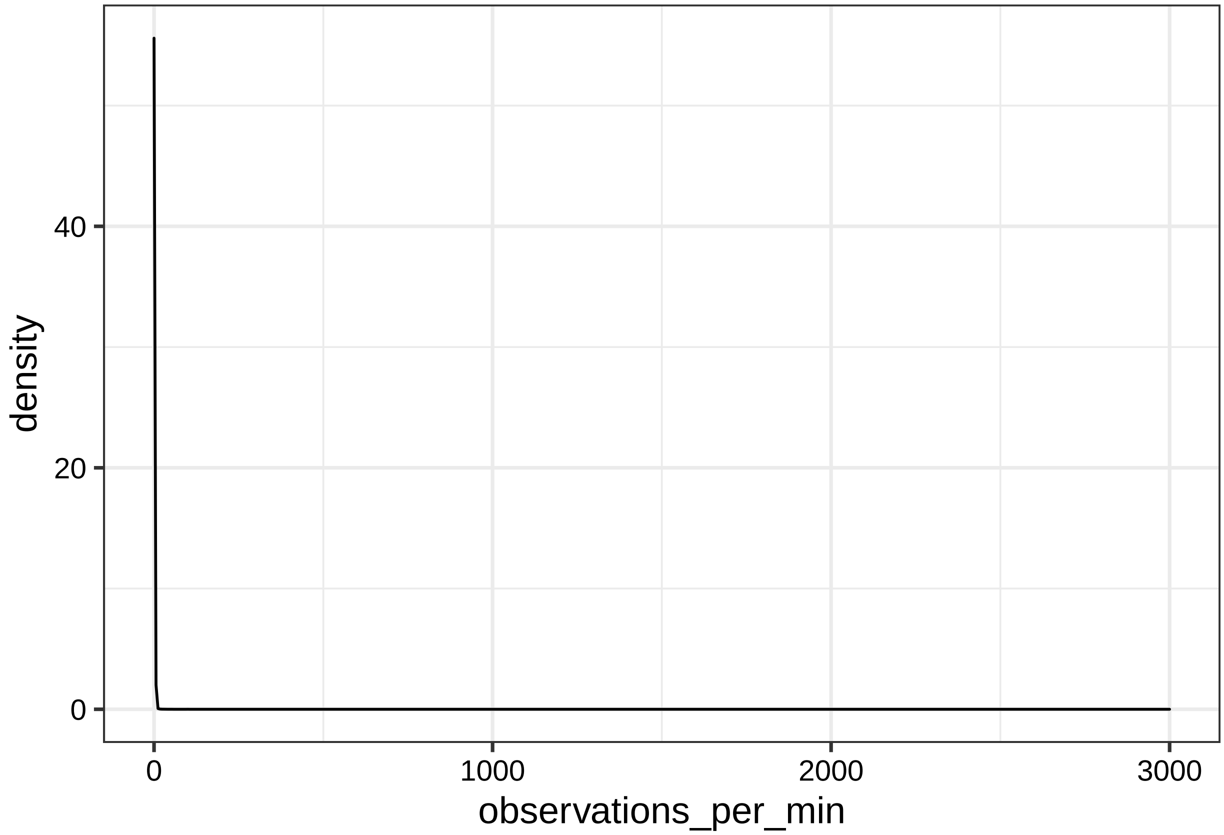
```
  "./Data/Raw/Massachusetts/ebd_US-MA_rewbla_201501_202012_relMar-2021/ebd_US-MA_rewbla_201501_202012_re

#Combine the two RWBL dataframes into a dataframe for 2010-2020
RWBL <- bind_rows(RWBL_early, RWBL_late)

#Wood Duck
WODU <- read_ebd(
  "./Data/Raw/Massachusetts/ebd_US-MA_wooduc_relFeb-2021/ebd_US-MA_wooduc_relFeb-2021.txt")

#Create a merged dataset with all three species
MAbirds <- bind_rows(OSPR, RWBL, WODU)
```

Clean and Filter Bird Data

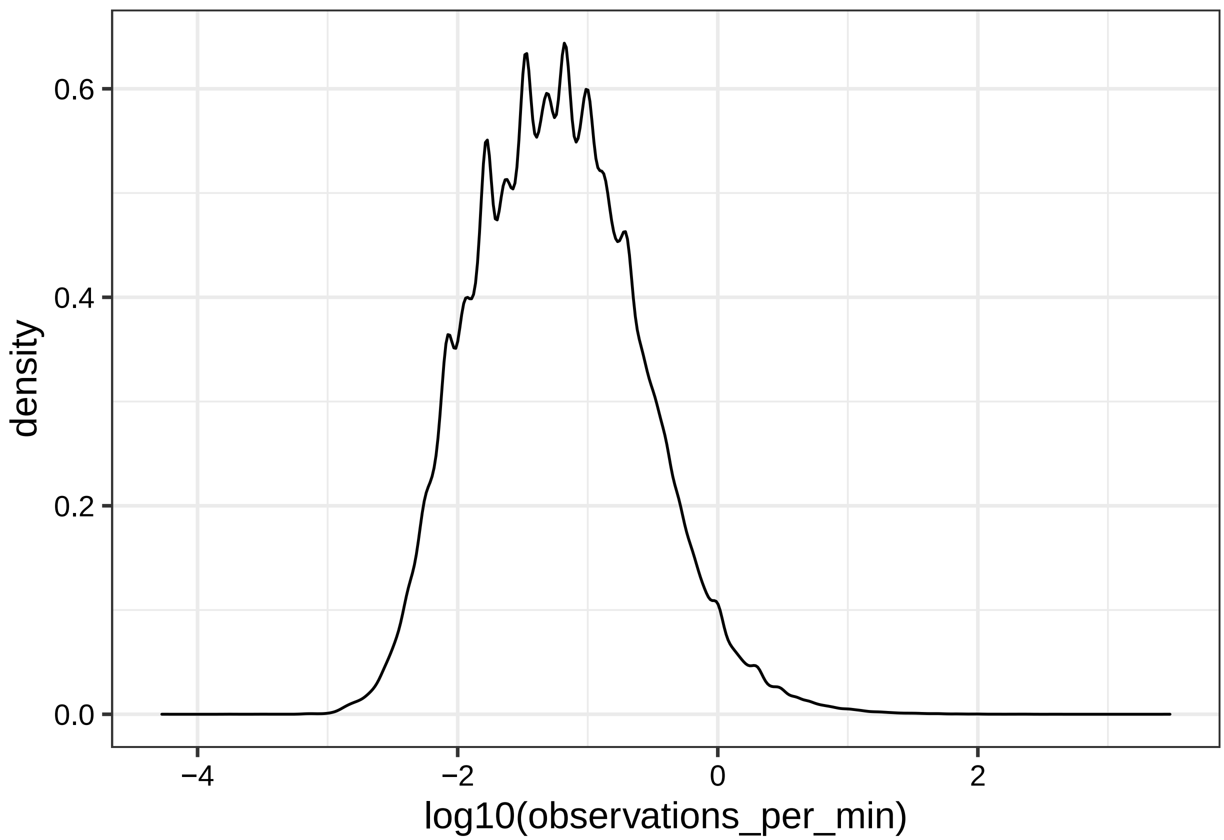
```
###Clean and filter the merged bird observation data
MAbirds <- MAbirds %>%
  #Filter species for columns of interest
  select(common_name:observation_count, state, county,
         latitude:observation_date, protocol_type, duration_minutes) %>%
  #Create/adjust columns useful for summarizing data
  mutate(#Replace "presence" notations ('X') in observation_count with values of 1
         observation_count = as.numeric(replace(
           observation_count, observation_count == 'X', '1')),
         #Add an observations_per_min column to control for increased effort
         observations_per_min = observation_count/duration_minutes,
         #Add a binary "Presence" column
         Presence = 1,
         #Add a date column set to the first day of each month
         Month_Year = floor_date(observation_date, unit = 'month')) %>%
  #Filter for years 2010-2020 (when there is data for all three species)
  filter(year(Month_Year) %in% c(2010:2020))


#Exclude NA and Inf values in the observations_per_min column
MAbirds <- MAbirds %>%
  filter( is.na(observations_per_min) == FALSE) %>%
  filter( observations_per_min != Inf)
#This filtering removed 5.1% of the original three-species dataset (21,781 rows).

#Examine the results
summary(MAbirds$observations_per_min)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    0.0001    0.0200    0.0556    0.2295    0.1556 3000.0000
```

```
#Visualize distribution
ggplot(MAbirds, aes(x=observations_per_min)) +
  geom_density() #Normal distribution has a heavy right skew
```




```
ggplot(MAbirds, aes(x=log10(observations_per_min))) +
  geom_density() #The log10 transformed data looks decent
```




```
#Save output to csv
write.csv(MAbirds, row.names = FALSE,
          './Data/Processed/Massachusetts/MAbirds.csv')
```

Summarize MA Bird Data

```
## Create summary dataset of bird data by month
#Statistics per species: (1) sum of observation_count per month,
#(2) sum of presence per month, and (3) average observations per minute per month
MAbirds_month <- MAbirds %>%
  group_by(common_name, Month_Year) %>%
  summarize(state = first(state),
            observations_per_min_avg = mean(observations_per_min),
            presence_count = sum(Presence),
            observation_count = sum(observation_count)) %>%
  select(common_name, state, observation_count,
         observations_per_min_avg, presence_count, Month_Year)

##Join the bird observation to the temperature data
MAbirds_temp <- full_join(MAbirds_month, MA_temp, by='Month_Year')


#### Prepare the data for statistical analysis ####
```

```r
#Create numeric Month and Year columns for correlation plots
MAbirds_temp <- MAbirds_temp %>%
  mutate(Month = month(Month_Year),
         Year = year(Month_Year))

#Create a  numeric "Season" column based on Month and the meteorological seasons:
#Winter=1 (Dec-Feb), Spring=2 (Mar=May), Summer=3 (Jun-Aug), Fall=4 (Sept-Nov)
#The value mapping is easiest to remember by singing "You've Got a Friend in Me"
MAbirds_temp$Season[MAbirds_temp$Month %in% c(1,2,12)] <- 1 #Winter
MAbirds_temp$Season[MAbirds_temp$Month %in% c(3:5)] <- 2 #Spring
MAbirds_temp$Season[MAbirds_temp$Month %in% c(6:8)] <- 3 #Summer
MAbirds_temp$Season[MAbirds_temp$Month %in% c(9:11)] <- 4 #Fall

#Save output to csv
write.csv(MAbirds_temp, row.names = FALSE,
          './Data/Processed/Massachusetts/MAbirds_temp.csv')
```

MA Exploratory Plots

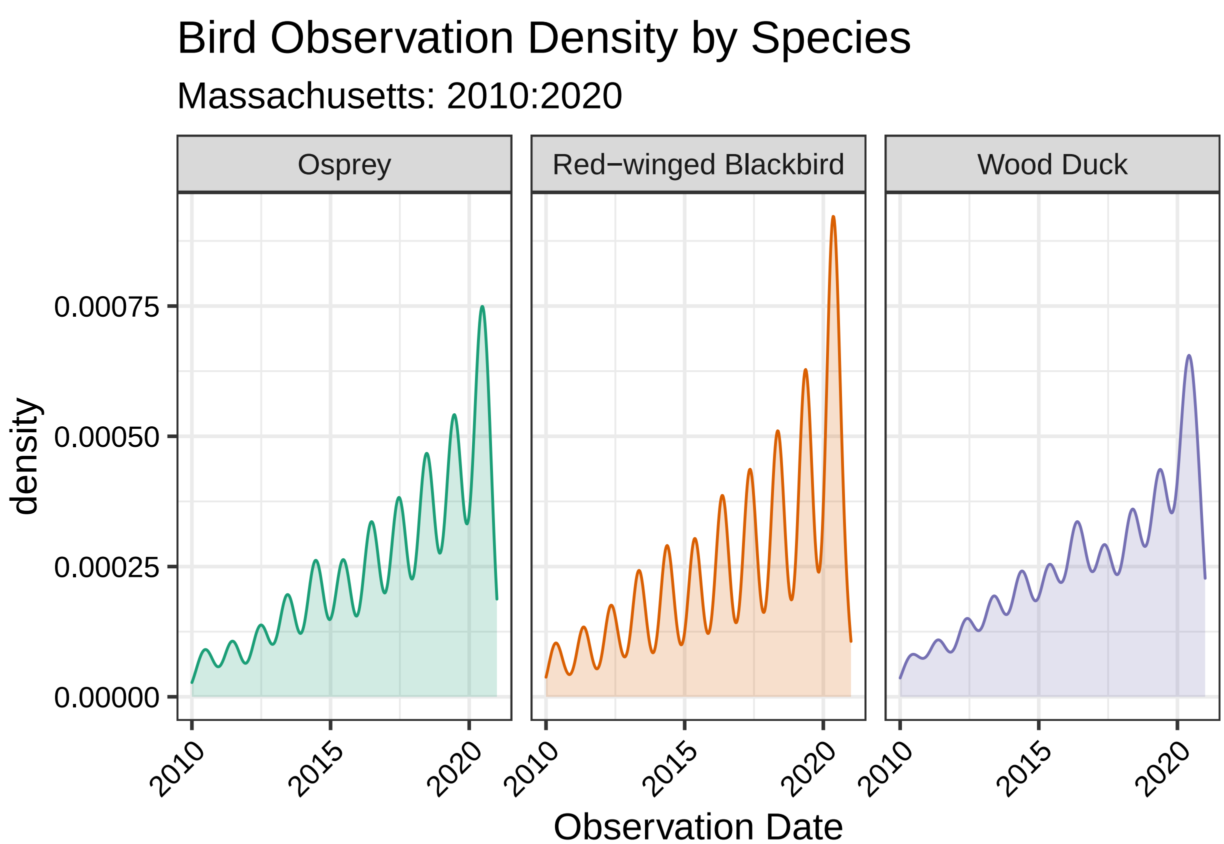


Figure 1: Basic density plot for all MA species across the entire date range. This plot represents bird observations uncorrected for effort.

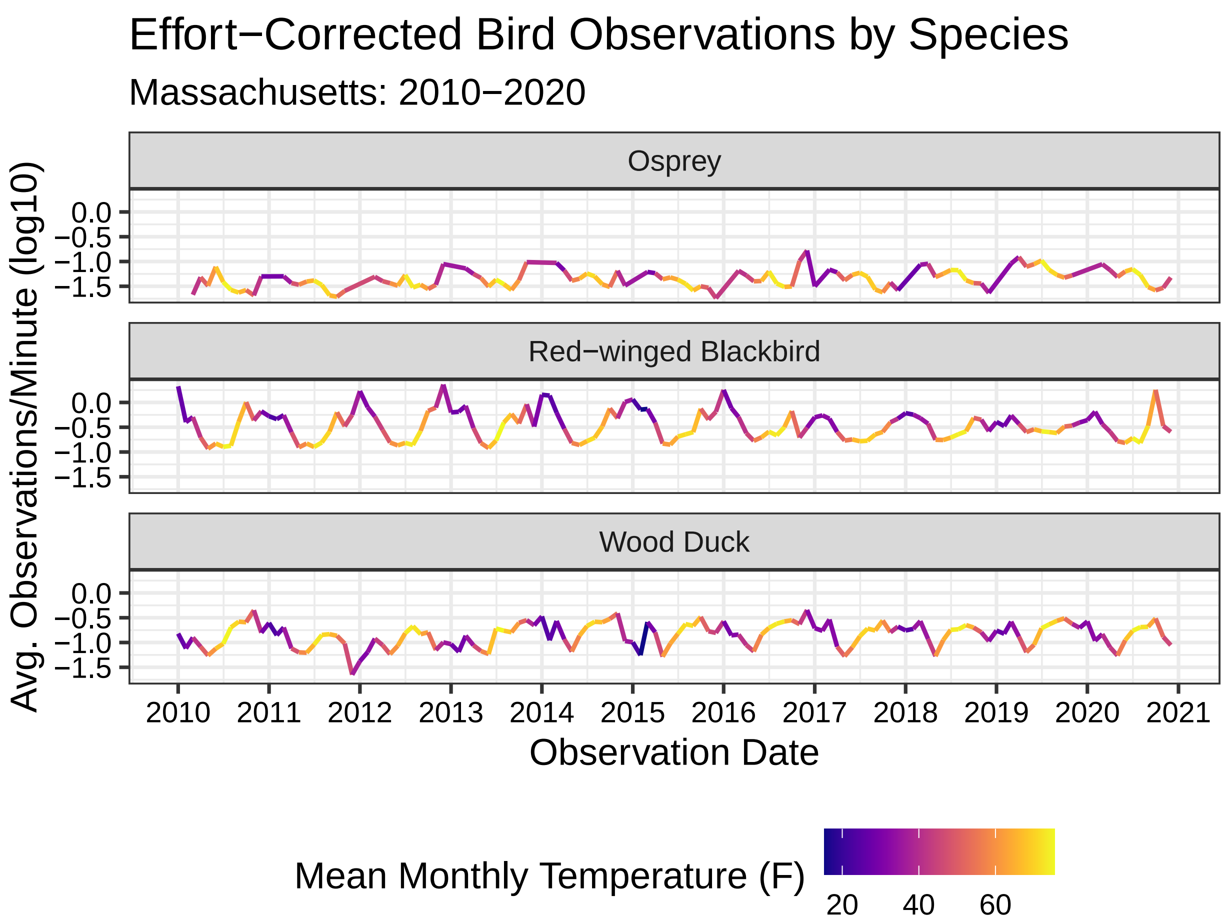


Figure 2: Average Observatons per Minute during 2020-2020 by Species in Massachusetts

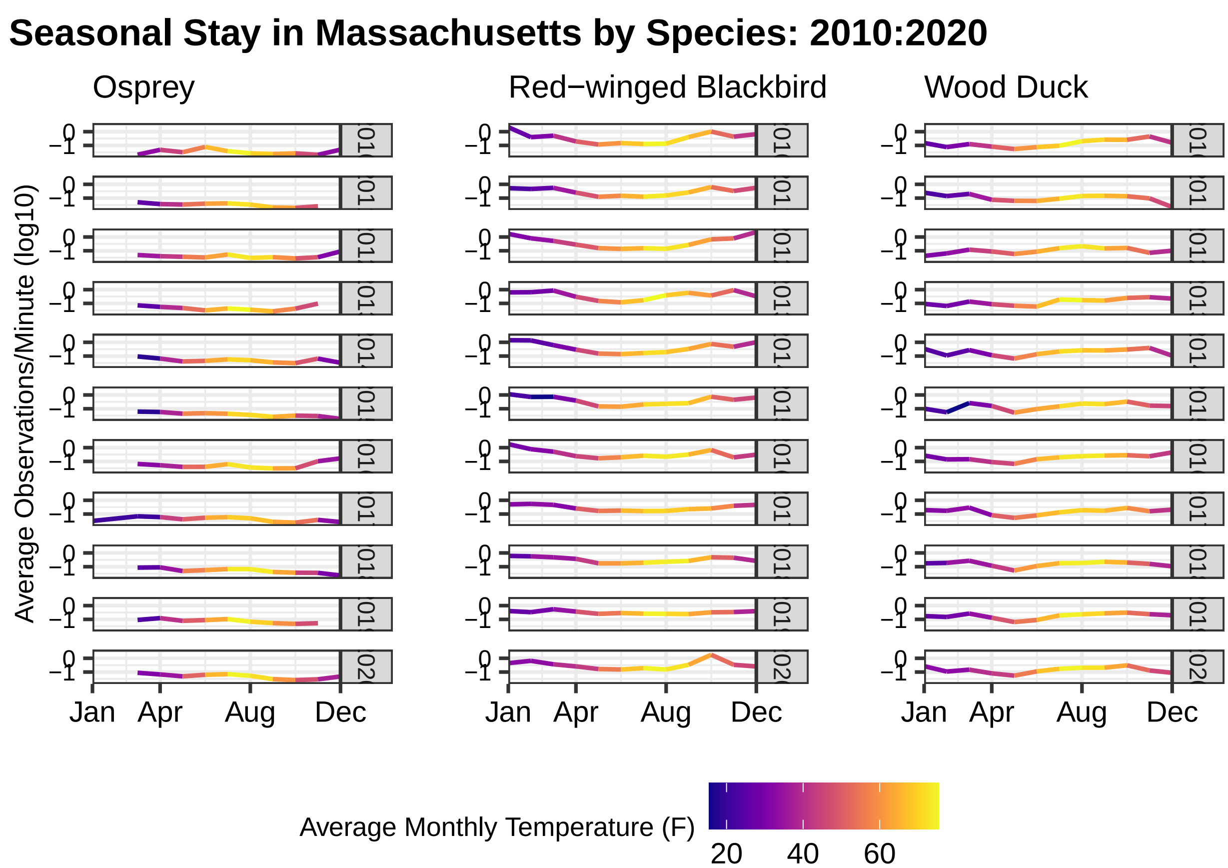


Figure 3: Cowplot of Average Observations per Minute by Year and Species in Massachusetts

MA Statistical Analysis

> Question: For each bird species, what is the relationship between the average monthly observations per minute and average monthly temperature?
>
> Is month (i.e., time) or average monthly temperature a better predictor of species-specific trends in average monthly observations per minute?
>
> Hypothesis: This is likely species depended. I expect that temperature will have the largest affect on the warmer temperatures will result in less birds, effect of temperature will depend on month.

Correlation Tests

> Question: How are changes in temperature associated with time in Massachusetts?

```
cor.test(MAbirds_temp$Month, MAbirds_temp$AvgMonthlyTemp_F)
```

```
##
##  Pearson's product-moment correlation
##
## data:  MAbirds_temp$Month and MAbirds_temp$AvgMonthlyTemp_F
## t = 5.1206, df = 370, p-value = 4.91e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1597466 0.3497773
## sample estimates:
##       cor
## 0.2572472
```

```
cor.test(MAbirds_temp$Season, MAbirds_temp$AvgMonthlyTemp_F)
```

```
##
##  Pearson's product-moment correlation
##
## data:  MAbirds_temp$Season and MAbirds_temp$AvgMonthlyTemp_F
## t = 14.147, df = 370, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5222638 0.6547172
## sample estimates:
##       cor
## 0.5924802
```

```
cor.test(MAbirds_temp$Year, MAbirds_temp$AvgMonthlyTemp_F)
```

```
##
##  Pearson's product-moment correlation
##
## data:  MAbirds_temp$Year and MAbirds_temp$AvgMonthlyTemp_F
## t = -0.29765, df = 370, p-value = 0.7661
```

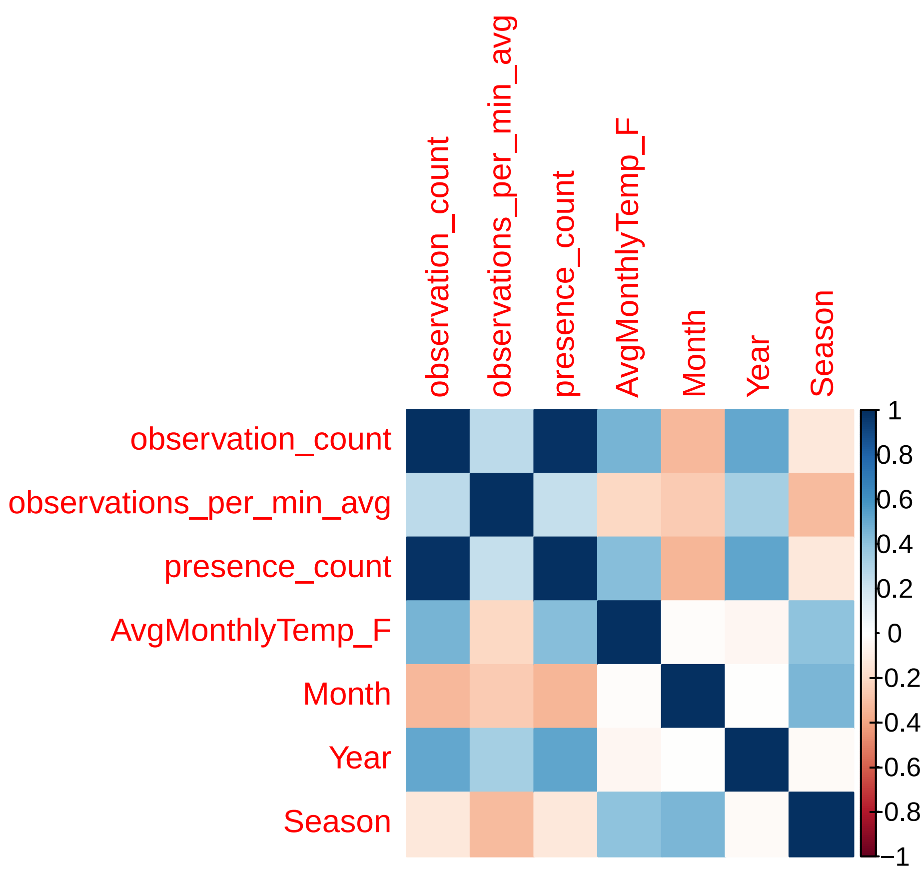
```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.11696747  0.08634245
## sample estimates:
##         cor
## -0.01547243
```

Answer: Temperature is not associated with year (cor= -0.01547243). This indicates that we would need to download data for a significantly larger date range to observe potential species shifts related to climate change. There are stronger associations between temperature and both month and season (cor=0.2572472 and 0.5924802, respectively). This is expected given the known seasonal variation in Massachusetts.

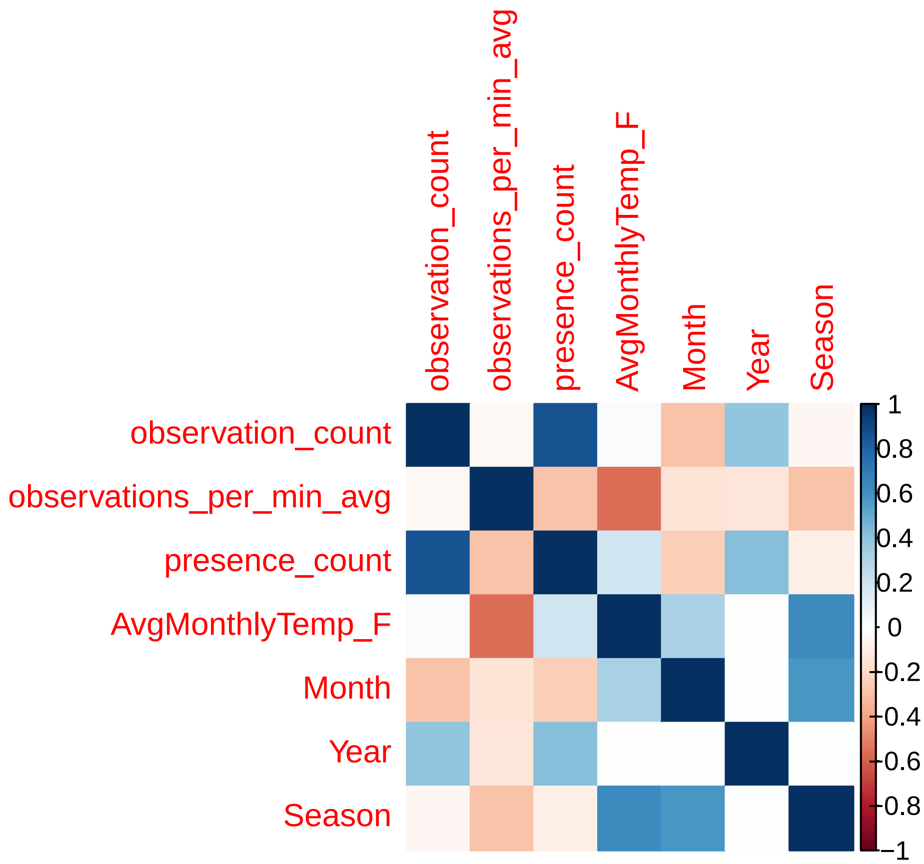
Correlation Plots

Create correlation plots for the numeric variables in our processed dataset: observation count, presence, average observations per minute, month, year, temp, and season Question: What does the relationship between bird observations, time of year, and temperature look like?

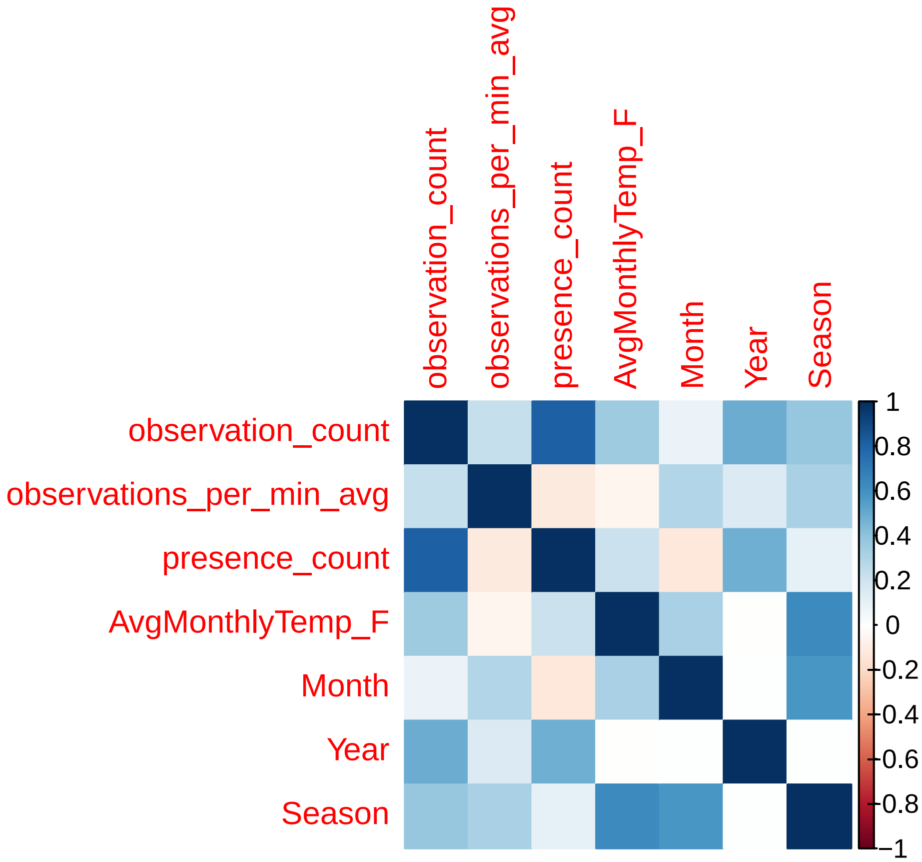
```
## Osprey
corMatrix_OSPR <- cor(OSPR_MA_temp[,c(3:5,7:10)])
corrplot(corMatrix_OSPR, method = "color")
```



```r
## Red winged Blackbird
corMatrix_RWBL <- cor(RWBL_MA_temp[,c(3:5,7:10)])
corrplot(corMatrix_RWBL, method = "color")
```




```r
## Wood Duck
corMatrix_WODU <- cor(WODU_MA_temp[,c(3:5,7:10)])
corrplot(corMatrix_WODU, method = "color")
```



> Answer: Based on the correlation plots, the strongest, non-colinear relationships appear to be between observations per minute and average monthly temperature, observations per minute and month, and observations per minute and season. The osprey correlation plot had the strongest association between month and the observational variables of the three species while the red-winged blackbird ahd the strongest association between temperature and the observational variables (specifically observatons per minute). Across the three species, there also appear to be associations between temperature/time and observation count/presence. However, the statistical analysis for MA will focus on effort-adjusted observation data only.

Linear Regressions

Question: For each species, is month, season, or average monthly temperature a better predictor of effort-adjusted observations (average observations per minute)?

NOTE: For each species, Month and Season are evaluated separately (with Average Monthly Temperature) to avoid regressions colinearity. Season is a numeric variable that was created by grouping the Month variable into four groups, which will cause singularities when running a regression.

Hypothesis: Time of year (i.e., month or season) will have a stronger relationship to observations per minute than average monthly temperature. In other words, there will be a stronger association with the time of year that birds tend to be observed than trends in temperature associated with those times of year. Between month and season, I expect that month will be a better predictor of effort-adjusted observations for the osprey, who are not present in Massachusetts year round. I expect the grouping of months into seasons will make the the broader season variable be a better predictor of bird-observations for the red-winged blackbird and wood duck, both of which have resident populations.

```r
#### Osprey ####
#Observations per minute vs.Temperature and Month
lm_OSPR.MA.month <- lm(data = OSPR_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month)-1)
#Summarize output
summary(lm_OSPR.MA.month)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1, data = OSPR_MA_temp)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.041553 -0.011391 -0.004116  0.005296  0.107502
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## AvgMonthlyTemp_F   -0.0018163  0.0008105  -2.241 0.027345 *
## as.factor(Month)1   0.0899431  0.0340552   2.641 0.009648 **
## as.factor(Month)3   0.1339084  0.0301607   4.440 2.41e-05 ***
## as.factor(Month)4   0.1488153  0.0386350   3.852 0.000212 ***
## as.factor(Month)5   0.1502754  0.0476114   3.156 0.002135 **
## as.factor(Month)6   0.1724525  0.0539924   3.194 0.001898 **
## as.factor(Month)7   0.1915439  0.0598136   3.202 0.001849 **
## as.factor(Month)8   0.1730429  0.0579618   2.985 0.003592 **
## as.factor(Month)9   0.1485052  0.0525129   2.828 0.005702 **
## as.factor(Month)10  0.1275760  0.0435130   2.932 0.004212 **
## as.factor(Month)11  0.1240831  0.0344950   3.597 0.000511 ***
## as.factor(Month)12  0.1181575  0.0284783   4.149 7.22e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0219 on 96 degrees of freedom
## Multiple R-squared:  0.8598, Adjusted R-squared:  0.8423
## F-statistic: 49.08 on 12 and 96 DF,  p-value: < 2.2e-16
```

```r
#Select model based on AIC
step(lm_OSPR.MA.month) #no recommended change to model
```

```
## Start:  AIC=-814.09
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1
##
##                    Df Sum of Sq      RSS     AIC
## <none>                          0.046054 -814.09
## - AvgMonthlyTemp_F  1  0.002409 0.048463 -810.58
## - as.factor(Month) 11  0.045784 0.091838 -761.55
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
```

```
##     1, data = OSPR_MA_temp)
##
## Coefficients:
##   AvgMonthlyTemp_F   as.factor(Month)1   as.factor(Month)3   as.factor(Month)4
##          -0.001816            0.089943            0.133908            0.148815
##  as.factor(Month)5   as.factor(Month)6   as.factor(Month)7   as.factor(Month)8
##           0.150275            0.172452            0.191544            0.173043
##  as.factor(Month)9  as.factor(Month)10  as.factor(Month)11  as.factor(Month)12
##           0.148505            0.127576            0.124083            0.118158
```

```
#Observations per minute vs. Temperature and Season
lm_OSPR.MA.season <- lm(data = OSPR_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season)-1)
#Summarize output
summary(lm_OSPR.MA.season)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = OSPR_MA_temp)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042486 -0.013448 -0.003708  0.006479  0.111866
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## AvgMonthlyTemp_F   -0.0009340  0.0002743  -3.405 0.000945 ***
## as.factor(Season)1  0.0853535  0.0117907   7.239 8.39e-11 ***
## as.factor(Season)2  0.1027368  0.0134904   7.616 1.32e-11 ***
## as.factor(Season)3  0.1171087  0.0196268   5.967 3.45e-08 ***
## as.factor(Season)4  0.0865979  0.0150458   5.756 8.96e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02203 on 103 degrees of freedom
## Multiple R-squared:  0.8479, Adjusted R-squared:  0.8405
## F-statistic: 114.8 on 5 and 103 DF,  p-value: < 2.2e-16
```

```
#Select model based on AIC
step(lm_OSPR.MA.season) #no recommended change to model
```

```
## Start:  AIC=-819.26
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1
##
##                     Df Sum of Sq      RSS     AIC
## <none>                           0.049978 -819.26
## - AvgMonthlyTemp_F   1  0.005625 0.055603 -809.74
## - as.factor(Season)  4  0.041860 0.091838 -761.55
```

```
##
```

```
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = OSPR_MA_temp)
##
## Coefficients:
##   AvgMonthlyTemp_F  as.factor(Season)1  as.factor(Season)2  as.factor(Season)3
##          -0.000934            0.085353            0.102737            0.117109
## as.factor(Season)4
##           0.086598
```

```
#### Red winged Blackbird ####

#Observations per minute vs. Temperature and Month
lm_RWBL.MA.month <- lm(data = RWBL_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month)-1)
#Summarize output
summary(lm_RWBL.MA.month)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1, data = RWBL_MA_temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62749 -0.06796 -0.02212  0.02810  1.65874
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## AvgMonthlyTemp_F   -0.006075   0.008811  -0.689  0.49186
## as.factor(Month)1   1.195381   0.259142   4.613 1.01e-05 ***
## as.factor(Month)2   0.839803   0.273838   3.067  0.00268 **
## as.factor(Month)3   0.781747   0.332721   2.350  0.02044 *
## as.factor(Month)4   0.577319   0.423793   1.362  0.17569
## as.factor(Month)5   0.515344   0.520664   0.990  0.32429
## as.factor(Month)6   0.567522   0.589670   0.962  0.33778
## as.factor(Month)7   0.629107   0.652688   0.964  0.33706
## as.factor(Month)8   0.637402   0.632635   1.008  0.31572
## as.factor(Month)9   0.707748   0.573662   1.234  0.21973
## as.factor(Month)10  1.026863   0.476400   2.155  0.03314 *
## as.factor(Month)11  0.719123   0.379237   1.896  0.06035 .
## as.factor(Month)12  0.861388   0.309921   2.779  0.00633 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3032 on 119 degrees of freedom
## Multiple R-squared:  0.7671, Adjusted R-squared:  0.7416
## F-statistic: 30.15 on 13 and 119 DF,  p-value: < 2.2e-16
```

```
#Select model based on AIC
step(lm_RWBL.MA.month) #stepwise function recommends dropping temperature
```

```
## Start:  AIC=-302.77
```

```
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1
##
##                    Df Sum of Sq    RSS     AIC
## - AvgMonthlyTemp_F  1    0.0437 10.980 -304.24
## <none>                          10.937 -302.77
## - as.factor(Month) 12   18.6883 29.625 -195.23
##
## Step:  AIC=-304.24
## observations_per_min_avg ~ as.factor(Month) - 1
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          10.980 -304.24
## - as.factor(Month) 12    35.976 46.956 -136.43
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ as.factor(Month) - 1,
##     data = RWBL_MA_temp)
##
## Coefficients:
##  as.factor(Month)1   as.factor(Month)2   as.factor(Month)3   as.factor(Month)4
##             1.0282              0.6618              0.5612              0.2920
##  as.factor(Month)5   as.factor(Month)6   as.factor(Month)7   as.factor(Month)8
##             0.1619              0.1659              0.1835              0.2058
##  as.factor(Month)9  as.factor(Month)10  as.factor(Month)11  as.factor(Month)12
##             0.3173              0.7045              0.4654              0.6572
```

```
##Observations per minute vs. Month
lm_RWBL.MA.month.only <- lm(data = RWBL_MA_temp,
                observations_per_min_avg ~ as.factor(Month)-1)
#Summarize output
summary(lm_RWBL.MA.month.only)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ as.factor(Month) - 1,
##     data = RWBL_MA_temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62266 -0.07043 -0.01674  0.02181  1.64674
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## as.factor(Month)1   1.02819    0.09121  11.273  < 2e-16 ***
## as.factor(Month)2   0.66182    0.09121   7.256 4.28e-11 ***
## as.factor(Month)3   0.56117    0.09121   6.153 1.04e-08 ***
## as.factor(Month)4   0.29200    0.09121   3.202 0.001750 **
## as.factor(Month)5   0.16193    0.09121   1.775 0.078367 .
## as.factor(Month)6   0.16587    0.09121   1.819 0.071469 .
## as.factor(Month)7   0.18352    0.09121   2.012 0.046442 *
## as.factor(Month)8   0.20578    0.09121   2.256 0.025865 *
```

```
## as.factor(Month)9   0.31727    0.09121   3.479 0.000703 ***
## as.factor(Month)10  0.70449    0.09121   7.724 3.78e-12 ***
## as.factor(Month)11  0.46535    0.09121   5.102 1.27e-06 ***
## as.factor(Month)12  0.65721    0.09121   7.206 5.55e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3025 on 120 degrees of freedom
## Multiple R-squared:  0.7662, Adjusted R-squared:  0.7428
## F-statistic: 32.76 on 12 and 120 DF,  p-value: < 2.2e-16
```

```
step(lm_RWBL.MA.month.only)
```

```
## Start:  AIC=-304.24
## observations_per_min_avg ~ as.factor(Month) - 1
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          10.980 -304.24
## - as.factor(Month) 12    35.976 46.956 -136.43
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ as.factor(Month) - 1,
##     data = RWBL_MA_temp)
##
## Coefficients:
##  as.factor(Month)1   as.factor(Month)2   as.factor(Month)3   as.factor(Month)4
##             1.0282              0.6618              0.5612              0.2920
##  as.factor(Month)5   as.factor(Month)6   as.factor(Month)7   as.factor(Month)8
##             0.1619              0.1659              0.1835              0.2058
##  as.factor(Month)9  as.factor(Month)10  as.factor(Month)11  as.factor(Month)12
##             0.3173              0.7045              0.4654              0.6572
```

```
#Observations per minute vs. Temperature and Season
lm_RWBL.MA.season <- lm(data = RWBL_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season)-1)
#Summarize output
summary(lm_RWBL.MA.season)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = RWBL_MA_temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48755 -0.12733 -0.04050  0.06133  1.58505
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## AvgMonthlyTemp_F   -0.011671   0.003759  -3.105 0.002350 **
## as.factor(Season)1  1.134187   0.126038   8.999 2.79e-15 ***
```

```
## as.factor(Season)2  0.888633   0.185638   4.787 4.63e-06 ***
## as.factor(Season)3  1.003981   0.269488   3.726 0.000292 ***
## as.factor(Season)4  1.114685   0.206873   5.388 3.33e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3171 on 127 degrees of freedom
## Multiple R-squared:  0.7281, Adjusted R-squared:  0.7174
## F-statistic: 68.01 on 5 and 127 DF,  p-value: < 2.2e-16
```

```
#Select model based on AIC
step(lm_RWBL.MA.season) #no recommended change to model
```

```
## Start:  AIC=-298.32
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1
##
##                     Df Sum of Sq    RSS     AIC
## <none>                           12.769 -298.32
## - AvgMonthlyTemp_F   1    0.9691 13.738 -290.67
## - as.factor(Season)  4   16.8561 29.625 -195.23
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = RWBL_MA_temp)
##
## Coefficients:
##   AvgMonthlyTemp_F  as.factor(Season)1  as.factor(Season)2  as.factor(Season)3
##           -0.01167             1.13419             0.88863             1.00398
## as.factor(Season)4
##            1.11468
```

```
#### Wood Duck ####

#Observations per minute vs. Temperature and Month
lm_WODU.MA.month <- lm(data = WODU_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month)-1)
#Summarize output
summary(lm_WODU.MA.month)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1, data = WODU_MA_temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15380 -0.03682 -0.00058  0.03555  0.28167
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
```

```
## AvgMonthlyTemp_F   -0.004237   0.001887  -2.245 0.026586 *
## as.factor(Month)1   0.303124   0.055489   5.463 2.61e-07 ***
## as.factor(Month)2   0.239650   0.058636   4.087 7.97e-05 ***
## as.factor(Month)3   0.356859   0.071245   5.009 1.92e-06 ***
## as.factor(Month)4   0.300167   0.090746   3.308 0.001244 **
## as.factor(Month)5   0.305684   0.111488   2.742 0.007053 **
## as.factor(Month)6   0.375718   0.126264   2.976 0.003542 **
## as.factor(Month)7   0.471930   0.139758   3.377 0.000991 ***
## as.factor(Month)8   0.508552   0.135465   3.754 0.000271 ***
## as.factor(Month)9   0.485032   0.122837   3.949 0.000134 ***
## as.factor(Month)10  0.479962   0.102010   4.705 6.91e-06 ***
## as.factor(Month)11  0.393145   0.081205   4.841 3.91e-06 ***
## as.factor(Month)12  0.308072   0.066363   4.642 8.94e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06491 on 119 degrees of freedom
## Multiple R-squared:  0.8897, Adjusted R-squared:  0.8777
## F-statistic: 73.86 on 13 and 119 DF,  p-value: < 2.2e-16
```

```
#Select model based on AIC
step(lm_WODU.MA.month) #no recommended change to model
```

```
## Start:  AIC=-709.64
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1
##
##                    Df Sum of Sq     RSS     AIC
## <none>                          0.50146 -709.64
## - AvgMonthlyTemp_F  1   0.02125 0.52270 -706.16
## - as.factor(Month) 12   0.84687 1.34832 -603.08
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Month) -
##     1, data = WODU_MA_temp)
##
## Coefficients:
##   AvgMonthlyTemp_F   as.factor(Month)1   as.factor(Month)2   as.factor(Month)3
##          -0.004237            0.303124            0.239650            0.356859
##  as.factor(Month)4   as.factor(Month)5   as.factor(Month)6   as.factor(Month)7
##           0.300167            0.305684            0.375718            0.471930
##  as.factor(Month)8   as.factor(Month)9  as.factor(Month)10  as.factor(Month)11
##           0.508552            0.485032            0.479962            0.393145
## as.factor(Month)12
##           0.308072
```

```
#Observations per minute vs. Temperature and Season
lm_WODU.MA.season <- lm(data = WODU_MA_temp,
                observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season)-1)
#Summarize output
summary(lm_WODU.MA.season)
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = WODU_MA_temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19354 -0.04441 -0.01163  0.04986  0.30297
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## AvgMonthlyTemp_F   -0.0027304  0.0008809  -3.100  0.00239 **
## as.factor(Season)1  0.2382152  0.0295330   8.066 4.69e-13 ***
## as.factor(Season)2  0.2498859  0.0434984   5.745 6.45e-08 ***
## as.factor(Season)3  0.3463776  0.0631461   5.485 2.14e-07 ***
## as.factor(Season)4  0.3728286  0.0484742   7.691 3.51e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0743 on 127 degrees of freedom
## Multiple R-squared:  0.8458, Adjusted R-squared:  0.8398
## F-statistic: 139.3 on 5 and 127 DF,  p-value: < 2.2e-16
```

```
#Select model based on AIC
step(lm_WODU.MA.season) #no recommended change to model
```

```
## Start:  AIC=-681.41
## observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1
##
##                     Df Sum of Sq     RSS     AIC
## <none>                           0.70109 -681.41
## - AvgMonthlyTemp_F   1   0.05304 0.75412 -673.78
## - as.factor(Season)  4   0.64724 1.34832 -603.08
```

```
##
## Call:
## lm(formula = observations_per_min_avg ~ AvgMonthlyTemp_F + as.factor(Season) -
##     1, data = WODU_MA_temp)
##
## Coefficients:
##   AvgMonthlyTemp_F  as.factor(Season)1  as.factor(Season)2  as.factor(Season)3
##           -0.00273             0.23822             0.24989             0.34638
## as.factor(Season)4
##            0.37283
```

Analysis

Osprey: The osprey month-temperature model explained slightly more variance in effort-adjusted observations than the season-temperature model(**84.23%** and **84.05%**, respectively). Looking at the month and season coefficients, this is likely due to the lack/absence of osprey observational data during January and February (assumed to be from migration). The months surrounding this period (i.e., March, April, November, December) are the most statistically significant in the month-temperature model (p-value <0.001). Given that three seasons are represented in those four months, it makes sense that modeling the data by month was more accurate than modeling the data by season for this species. Average monthly temperature was statistically significant in both models: month-temperature p-value: **0.027345** and season-temperature p-value: **0.000945**. This result was expected, as the correlation between season and temperature was greater than the correlation between month and temperature based on earlier correlation tests. That being said, the coefficient of average monthly temperature was not more significant than any of the seasonal coefficients for the osprey, which supports the hypothesis that time of year is a better predictor of changes in bird observations than temperature.

Red-winged Blackbird: Similar to the osprey, the red-winged blackbird month-temperature model explained the most variance in effort-corrected observations – **74.28%**. However, unlike the osprey, the stepwise function for the red-winged blackbird month-temperature model recommended removing average monthly temperature from the model (**Starting AIC: -302.77; Final AIC: -304.24**). Prior to removing temperature, the red-winged blackbird month-temperature model explained 74.16% of the variance in observation data, which was still a better fit than the season-temperature model (explaining 71.74% of variance). This result is consistent with the hypothesis that time of year is a better predictor of bird observation effort than temperature. While only ten months had statistically significant p-values > 0.5, month is still appears to be a better predictor of changes in effort-corrected observations for the red-winged blackbird than season. Seasonal trends in observations may be difficult to analyze for the red-winged blackbird in Massachusetts given the prevalence of the species and the fact that there are both resident and migratory populations. Looking at the month and season factors together, the change that occurs in species observations during the winter months is definitely worth exploring (p-value for Season 1 (Winter) is the most significant: 2.79e-15). However, I would recommend attempting to narrow down observations to a subpopulation of birds in Massachusetts.

Wood Duck: The wood duck month-temperature model explained more variance in effort-adjusted observations than the season-temperature model by the largest margin of the three bird species – **87.77%** to **83.98%**. Similar to the osprey, average monthly temperature was statistically significant for both models, but had a larger impact on bird observations in the season-temperature model (season-temperature p-value: 0.00239; month-temperature p-value: 0.026586). The stepwise functions for the wood duck models did not recommend dropping any variables or factors. Like the red-winged blackbird, wood ducks have both a resident and migratory population in Massachusetts. The resident population of wood ducks can make it difficult to evaluate their migration patterns using observation data because the species is observed year round. Viewing the month and season regressions together, there appear to be differences in wood duck observations in the Spring & Summer vs. Fall & Winter, which could be interesting to explore in more detail.

Post-hoc Analysis

For each species, what months have similar effort-adjusted bird observations? While not explored in great detail here, these groupings may indicate a more focused period of time to study migratory patterns in the future.

```
#### Osprey ####

#Fit an observations per minute vs. month model as an anova
aov_OSPR.MA.month <-
  aov(data = OSPR_MA_temp, observations_per_min_avg ~ as.factor(Month))
```

```r
#Summarize the aov
summary(aov_OSPR.MA.month)
```

```
##                   Df  Sum Sq   Mean Sq F value  Pr(>F)
## as.factor(Month) 10 0.01536 0.0015358   3.074 0.00197 **
## Residuals        97 0.04846 0.0004996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Reveal groupings with similar bird observations: Tukey HSD Test
OSPR.MA.month.grps <-
  HSD.test(aov_OSPR.MA.month, "as.factor(Month)", group = TRUE)
#Print the groupings
OSPR.MA.month.grps
```

```
## $statistics
##       MSerror Df       Mean       CV
##   0.000499615 97 0.04951444 45.14253
##
## $parameters
##    test           name.t ntr StudentizedRange alpha
##   Tukey as.factor(Month)  11         4.662438  0.05
##
## $means
##    observations_per_min_avg         std  r        Min        Max        Q25
## 1                0.03150912          NA  1 0.03150912 0.03150912 0.03150912
## 10               0.03120192 0.008061459 11 0.01941345 0.04719129 0.02641372
## 11               0.04821692 0.028482008 11 0.02069841 0.10261373 0.02950917
## 12               0.05674811 0.049944994  8 0.01815336 0.16711047 0.02553826
## 3                0.06796462 0.022006178 11 0.02131816 0.09363848 0.05589840
## 4                0.06351637 0.024502202 11 0.03666224 0.12323924 0.05036492
## 5                0.04461960 0.012462815 11 0.03225148 0.07840467 0.03817219
## 6                0.05237522 0.018251962 11 0.03111153 0.08844467 0.03990803
## 7                0.05833365 0.018954592 11 0.03846931 0.10501586 0.04308420
## 8                0.04400849 0.014331523 11 0.02668239 0.06759434 0.03425112
## 9                0.03176914 0.009281524 11 0.02075382 0.05359931 0.02628351
##           Q50        Q75
## 1  0.03150912 0.03150912
## 10 0.03064158 0.03413999
## 11 0.03593820 0.05884582
## 12 0.04015537 0.05972222
## 3  0.06869301 0.08717331
## 4  0.05812776 0.06664814
## 5  0.04163302 0.04749687
## 6  0.04775824 0.05979236
## 7  0.05736762 0.06525593
## 8  0.03608055 0.05184904
## 9  0.03049639 0.03433993
##
## $comparison
## NULL
##
## $groups
```

```
##    observations_per_min_avg groups
## 3                0.06796462      a
## 4                0.06351637      a
## 7                0.05833365     ab
## 12               0.05674811     ab
## 6                0.05237522     ab
## 11               0.04821692     ab
## 5                0.04461960     ab
## 8                0.04400849     ab
## 9                0.03176914      b
## 1                0.03150912      b
## 10               0.03120192      b
##
## attr(,"class")
## [1] "group"
```

```
#### Red-winged Blackbird ####

#Fit an observations per minute vs. month model as an anova
aov_RWBL.MA.month <-
  aov(data = RWBL_MA_temp, observations_per_min_avg ~ as.factor(Month))
#Summarize the aov
summary(aov_OSPR.MA.month)
```

```
##                  Df  Sum Sq   Mean Sq F value  Pr(>F)
## as.factor(Month) 10 0.01536 0.0015358   3.074 0.00197 **
## Residuals        97 0.04846 0.0004996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Reveal groupings with similar bird observations: Tukey HSD Test
RWBL.MA.month.grps <-
  HSD.test(aov_RWBL.MA.month, "as.factor(Month)", group = TRUE)
#Print the groupings
RWBL.MA.month.grps
```

```
## $statistics
##      MSerror  Df      Mean       CV       MSD
##   0.09150359 120 0.4503825 67.16416 0.4299805
##
## $parameters
##    test           name.t ntr StudentizedRange alpha
##   Tukey as.factor(Month)  12         4.714396  0.05
##
## $means
##    observations_per_min_avg        std  r       Min       Max       Q25
## 1                 1.0281892 0.63260078 11 0.4055289 2.1194947 0.5147117
## 10                0.7044928 0.42538785 11 0.2542612 1.8005514 0.4323437
## 11                0.4653518 0.21001654 11 0.1953581 0.9207245 0.3383975
## 12                0.6572056 0.59199109 11 0.2547917 2.3039455 0.3182619
## 2                 0.6618238 0.28007535 11 0.3331879 1.3767917 0.5047950
## 3                 0.5611657 0.13491236 11 0.3665382 0.8553286 0.4948463
## 4                 0.2919956 0.06254871 11 0.1964582 0.3921841 0.2524270
```

```
## 5                 0.1619271 0.03517748 11 0.1167685 0.2522183 0.1500835
## 6                 0.1658657 0.04615557 11 0.1198366 0.2877202 0.1397636
## 7                 0.1835212 0.04518296 11 0.1264296 0.2613480 0.1588802
## 8                 0.2057835 0.07423937 11 0.1340778 0.3937784 0.1541437
## 9                 0.3172680 0.10424697 11 0.2264929 0.5872847 0.2556231
##          Q50       Q75
## 1  0.6289867 1.5648851
## 10 0.6723131 0.7532237
## 11 0.4337148 0.4637343
## 12 0.4719875 0.6577372
## 2  0.6503527 0.7436118
## 3  0.5191808 0.5888495
## 4  0.2797512 0.3372797
## 5  0.1554748 0.1689643
## 6  0.1483978 0.1772454
## 7  0.1709739 0.1996734
## 8  0.1944319 0.2278492
## 9  0.2673782 0.3349191
##
## $comparison
## NULL
##
## $groups
##    observations_per_min_avg groups
## 1                 1.0281892      a
## 10                0.7044928     ab
## 2                 0.6618238     ab
## 12                0.6572056     ab
## 3                 0.5611657     bc
## 11                0.4653518     bc
## 9                 0.3172680     bc
## 4                 0.2919956     bc
## 8                 0.2057835      c
## 7                 0.1835212      c
## 6                 0.1658657      c
## 5                 0.1619271      c
##
## attr(,"class")
## [1] "group"
```

```
#### Wood Duck ####

#Fit an observations per minute vs. month model as an anova
aov_WODU.MA.month <-
  aov(data = WODU_MA_temp, observations_per_min_avg ~ as.factor(Month))
#Summarize the aov
summary(aov_WODU.MA.month)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(Month)  11 0.4321 0.03928   9.018 1.45e-11 ***
## Residuals        120 0.5227 0.00436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Reveal groupings with similar bird observations: Tukey HSD Test
WODU.MA.month.grps <-
  HSD.test(aov_WODU.MA.month, "as.factor(Month)", group = TRUE)
#Print the groupings
WODU.MA.month.grps
```

```
## $statistics
##       MSerror  Df      Mean       CV        MSD
##   0.004355854 120 0.1649739 40.00567 0.09381373
##
## $parameters
##    test           name.t ntr StudentizedRange alpha
##   Tukey as.factor(Month)  12         4.714396  0.05
##
## $means
##    observations_per_min_avg        std  r        Min        Max        Q25
## 1                0.18653383 0.08828403 11 0.04197384 0.33674581 0.12630096
## 10               0.25515923 0.06322874 11 0.13669684 0.32861908 0.22499162
## 11               0.21617893 0.11867951 11 0.07044320 0.44813680 0.14514660
## 12               0.16568648 0.11255537 11 0.02239183 0.45317437 0.10356129
## 2                0.11553720 0.04628985 11 0.05545918 0.18857344 0.07039925
## 3                0.20303749 0.06906292 11 0.11966843 0.29218367 0.14085621
## 4                0.10119781 0.02689736 11 0.07589466 0.16060193 0.08232092
## 5                0.05922994 0.00608517 11 0.05183867 0.06741072 0.05385341
## 6                0.09562493 0.02805372 11 0.05855119 0.14405027 0.07764859
## 7                0.16120271 0.04118061 11 0.09094674 0.21476571 0.14170847
## 8                0.20756508 0.03422376 11 0.14240651 0.26226330 0.18832114
## 9                0.21273278 0.04896648 11 0.14732904 0.27423900 0.16784343
##           Q50        Q75
## 1  0.17560490 0.25747734
## 10 0.27763752 0.30336951
## 11 0.16863047 0.26031567
## 12 0.15640067 0.20378681
## 2  0.11091035 0.14526519
## 3  0.20120242 0.26695744
## 4  0.08865882 0.11717578
## 5  0.05810698 0.06473925
## 6  0.09015785 0.11167753
## 7  0.17220874 0.19278901
## 8  0.20447278 0.23442266
## 9  0.21679726 0.26014967
##
## $comparison
## NULL
##
## $groups
##    observations_per_min_avg groups
## 10               0.25515923      a
## 11               0.21617893     ab
## 9                0.21273278     ab
## 8                0.20756508    abc
## 3                0.20303749    abc
## 1                0.18653383   abcd
```

```
## 12                0.16568648   abcd
## 7                 0.16120271    bcd
## 2                 0.11553720    cde
## 4                 0.10119781     de
## 6                 0.09562493     de
## 5                 0.05922994      e
##
## attr(,"class")
## [1] "group"
```