eBirds Data Analysis

Vivian Lee, Shun Yamamoto, and Vivian Chen

2019-10-21

Analysis of Bird Sightings Recorded on eBirds

Source: data gathered from ebird.org, managed by the Cornell Lab of Orthinology

```r
# Functions to extract and organize bird sighting data

get_bird_sightings <- function(url) {
  # Scrape the results from the given url from eBirds for a specific region
  # Args:
  #   url - the url for the eBirds region page
  # Return:
  #   A data frame containing the bird information organized by species code, name, and
  #   total sightings for specified date range, default = all sightings to date
  region_page <- read_html(url)

  bird_sightings <- html_node(region_page, 'body') %>%
    html_node(css='#sightingsTable') %>%
    html_node('tbody') %>%
    html_nodes(xpath='//tr[contains(@id, "has-det")]')

  species <- bird_sightings %>%
    html_nodes(css='.species-name') %>%
    html_node('a')

  species_codes <- species %>%
    html_attr('data-species-code')

  species_names <- species %>%
    html_text()

  # X connotes 1 or more birds sighted. Replace X with 1 since at least 1 bird was sighted.
  counts <- bird_sightings %>%
    html_node(xpath='td[@headers="c"]') %>%
    html_text() %>%
    map_chr(function(x) ifelse(x=="X", 1, x)) %>%
    as.integer()

  # Suppress warnings when NA is assigned to missing dates
  dates <- suppressWarnings(bird_sightings %>%
    html_node(xpath = 'td[@headers="d"]') %>%
    html_text() %>%
    dmy())

  # Construct dataframe of bird sighting data, omit records without a name column value
  # as these are sightings without where the bird species was not properly identified
  # and they do not follow a proper naming convention
  tibble(
    code = species_codes,
    name = species_names,
    count = counts,
    date = dates
```

```
  ) %>%
    filter(!is.na(date))
}
```

Create data tables for bird sightings in Washington County, RI and Rhode Island

We chose to analyze Washington County, which contains many towns including Weekapaug, as opposed to analyzing Weekapaug alone, due to the larger amount of sightings logged for the county. We also analyzed the bird sightings recorded for the entire state of Rhode Island. Our functions and methodology can be applied to any data table on eBirds for additional analyses of different regions.

Washington County, RI Bird Sightings

```
(washington_county_df <-
    get_bird_sightings('https://ebird.org/subnational2/US-RI-009?yr=all&m=&rank=mrec'))
```

```
## # A tibble: 480 x 4
##    code    name               count date
##    <chr>   <chr>              <int> <date>
##  1 mallar3 Mallard                7 2019-10-20
##  2 gnwtea  Green-winged Teal      1 2019-10-20
##  3 rinduc  Ring-necked Duck       1 2019-10-20
##  4 <NA>    <NA>                  27 2019-10-20
##  5 pibgre  Pied-billed Grebe      2 2019-10-20
##  6 grbher3 Great Blue Heron       2 2019-10-20
##  7 blujay  Blue Jay               2 2019-10-20
##  8 amerob  American Robin        20 2019-10-20
##  9 sonspa  Song Sparrow           1 2019-10-20
## 10 reshaw  Red-shouldered Hawk    1 2019-10-20
## # ... with 470 more rows
```

Rhode Island Bird Sightings

```
(rhode_island_df <- get_bird_sightings('https://ebird.org/region/US-RI?yr=all&m=&rank=mrec'))
```

```
## # A tibble: 522 x 4
##    code   name                count date
##    <chr>  <chr>               <int> <date>
##  1 grhowl Great Horned Owl        2 2019-10-21
##  2 perfal Peregrine Falcon        1 2019-10-20
##  3 rocpig Rock Pigeon             2 2019-10-20
##  4 coohaw Cooper's Hawk           1 2019-10-20
##  5 eursta European Starling       8 2019-10-20
##  6 carwre Carolina Wren           2 2019-10-20
##  7 cangoo Canada Goose           75 2019-10-20
##  8 mutswa Mute Swan              75 2019-10-20
##  9 gadwal Gadwall                 1 2019-10-20
## 10 amewig American Wigeon        25 2019-10-20
## # ... with 512 more rows
```

Organize Sightings by Date and Season of the Year

eBirds divides the bird migration year into the following four seasons:

Spring Migration (Mar-May)
Breeding Season (Jun-Jul)
Fall Migration (Aug-Nov)
Winter (Dec-Feb)
All (Mar-Feb)

```r
get_bird_sightings_for_date_range <- function(df, start_date, end_date) {
  # Get bird sightings that occurred within the specified date range for a given dataframe
  # Args:
  #   df - the dataframe to filter
  #   start_date - the start date in YYYY-MM-DD format, inclusive
  #   end_date - the end date in YYYY-MM-DD format, non-inclusive
  # Return:
  #   A data frame containing the bird sightings from the given df, filtered for the
  #   specified date range
  if (missing(start_date) & missing(end_date)) {
    df
  }
  else if (missing(start_date)) {
    df %>% filter(date < as.Date(end_date))
  }
  else if (missing(end_date)) {
    df %>% filter(date >= as.Date(start_date))
  }
  else {
    df %>% filter(date >= as.Date(start_date) & date < as.Date(end_date))
  }
}
```

```r
get_bird_sightings_for_season_year <- function(df, season, year) {
  # Get bird sightings that occurred for the specified season and year for a given dataframe
  # Args:
  #   df - the dataframe to filter
  #   season - one of ['spring', 'breeding', 'fall', 'winter', 'all']
  #   year - the year to filter for
  # Return:
  #   A data frame containing the bird sightings from the given df, filtered for the
  #   specified season and year

  if (!season %in% c("spring", "breeding", "fall", "winter", "all"))
    stop("season arg must be one of ['spring', 'breeding', 'fall', 'winter', 'all']")

  if (year < 1970)
    stop("year must be between 1970-present")

  start_month <- switch(season, "spring" = 3,"breeding" = 6,"fall" = 8,"winter" = 12, "all" = 3)
  end_month <- switch(season, "spring" = 6, "breeding" = 8, "fall" = 12, "winter" = 3, "all" = 3)
  start_year <- year
  end_year <- ifelse(season %in% c("winter", "all"), year+1, year)
  start_date <- str_c(start_year, start_month, 1, sep="-")
  end_date <- str_c(end_year, end_month, 1, sep="-")

  get_bird_sightings_for_date_range(df, start_date, end_date)
}

get_count_per_season_data <- function(df, year) {
  seasons <- c("spring", "breeding", "fall", "winter", "all")
  c(year, unlist(map(seasons,
```

```
                    function(x) sum(get_bird_sightings_for_season_year(df, x, year)$count))))
}

# start_year (inclusive) to end_year (inclusive)
get_count_per_season_for_year_range <- function(df, start_year, end_year) {
  years <- start_year:end_year
  year_row_data <- map(years, function(x) get_count_per_season_data(df, x))
  result_df <- data.frame(do.call(rbind, year_row_data))
  colnames(result_df) <- c("year","spring", "breeding", "fall", "winter", "all")
  result_df
}

# input: must be dataframe returned by get_count_per_season_for_year_range
plot_sightings_by_season <- function(df) {
  df %>% gather(`spring`, `breeding`, `fall`, `winter`, `all`, key="season", value="count") %>%
    ggplot(mapping=aes(x=year, y=count, color=season)) +
    geom_smooth(method="loess")
}
```

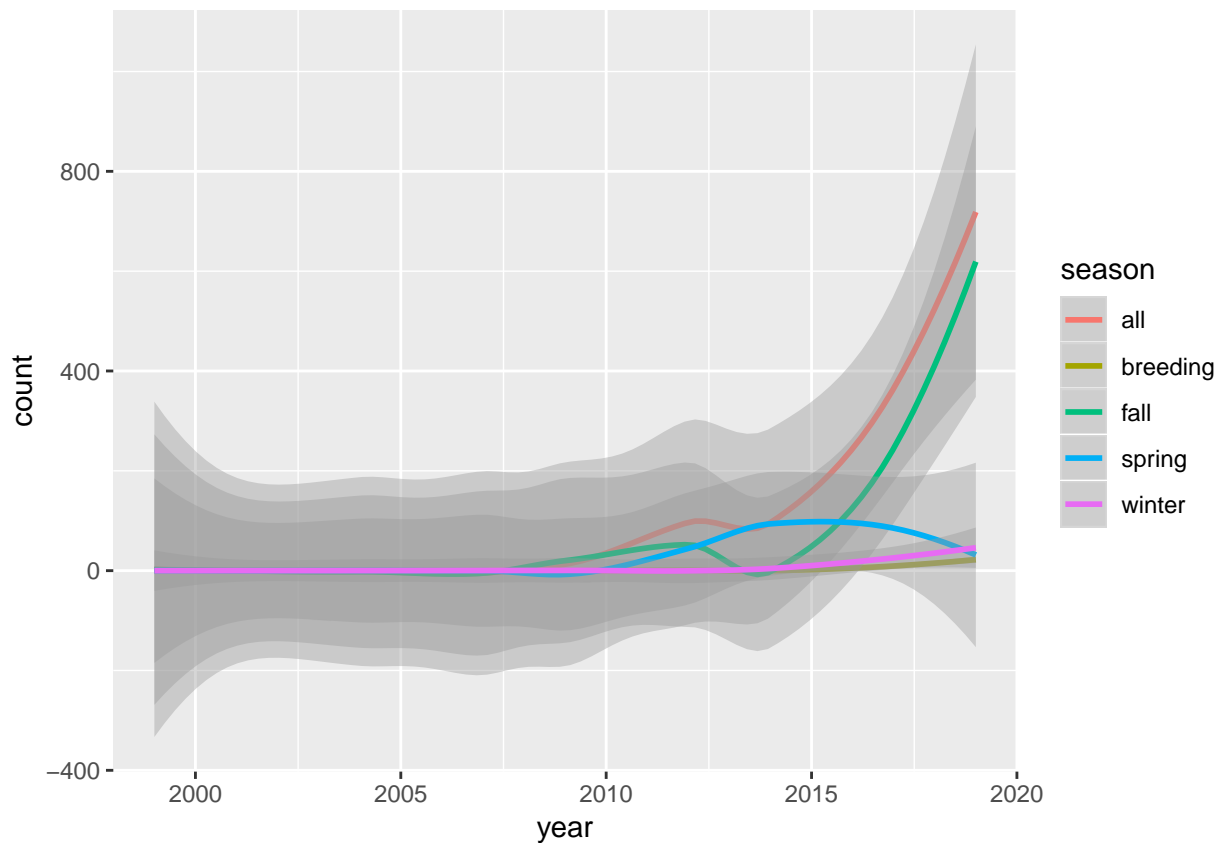**Birds Sighted in Washington County, RI by Season Between 1999-2019**

```
washington_sightings_by_season_1999_2019 <-
    get_count_per_season_for_year_range(washington_county_df, 1999, 2019)

head(washington_sightings_by_season_1999_2019, 10)
```

```
##     year spring breeding fall winter all
## 1   1999      0        1    1      0   2
## 2   2000      0        0    0      0   0
## 3   2001      0        0    0      0   0
## 4   2002      0        0    0      0   0
## 5   2003      0        1    1      0   2
## 6   2004      1        0    0      0   1
## 7   2005      0        0    0      0   0
## 8   2006      0        0    2      0   2
## 9   2007      0        0    1      0   1
## 10  2008      1        0    1      1   3
```

```
plot_sightings_by_season(washington_sightings_by_season_1999_2019)
```

Birds Sighted in Rhode Island by Season Between 1999-2019

```
ri_sightings_by_season_1999_2019 <-
    get_count_per_season_for_year_range(rhode_island_df, 1999, 2019)

head(ri_sightings_by_season_1999_2019, 10)
```

```
##     year spring breeding fall winter all
## 1  1999      0        1    1      0   2
## 2  2000      0        0    0      0   0
## 3  2001      0        0    0      0   0
## 4  2002      1        0    0      0   1
## 5  2003      0        1    1      0   2
## 6  2004      1        0    0      0   1
## 7  2005      0        0    0      1   1
## 8  2006      0        0    2      0   2
## 9  2007      0        0    2      0   2
## 10 2008      1        1    0      3   5
```

```
plot_sightings_by_season(ri_sightings_by_season_1999_2019)
```
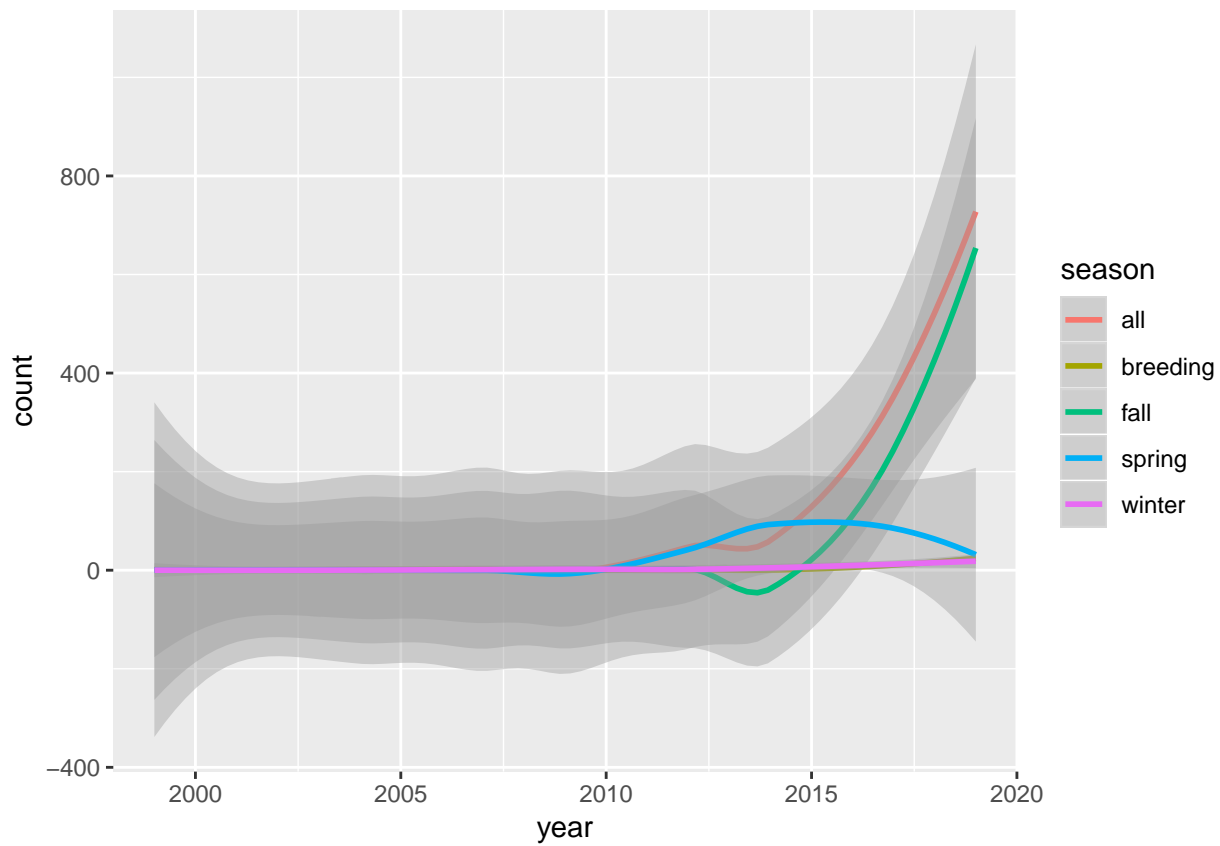
Note that the Fall line dips below 0 between 2010 and 2015. The actual count of birds sighted can never be below 0; this dip is due to the smoothing function used to smooth the line plot and show the trend over time. The loess function, typically used for plotting small data sets with < 1000 observations, was used to plot the line graphs.