Forest Grassland Analysis

Sergio

2025-03-29

Introduction

Bird populations serve as key indicators of ecosystem health. This study explores bird species distributions across **forest** and **grassland** habitats in the National Capital Region (NCR) using data from the **NCRN LAND Bird Monitoring Data (2007 - 2017)** dataset. By analyzing species richness, abundance, and habitat differences, this study provides insights to inform conservation strategies.

Scenario

The dataset represents **10 years** of bird monitoring across **11 National Parks**. Observers recorded bird detections at over **384 forest locations** and later expanded to **grassland sites** in 2014. The objective is to compare the diversity and abundance of bird species between these two habitats.

Ask

Business Problem:
How do bird species distributions vary between **forest** and **grassland** habitats in the National Capital Region?

Key Questions:
- Which species are most commonly found in forests vs. grasslands?
- How does species richness (number of unique species) differ by habitat?
- Are certain species more abundant in one habitat type compared to the other?
- What trends emerge when visualizing species distributions?

Prepare

Dataset: NCRN LAND Bird Monitoring Data (2007 - 2017)
Source: Catalog.Data.Gov

Data Summary:
- Observations collected across **forest** and **grassland** habitats
- Variables include **species, count, location type, weather conditions, and observer details**
- Data is structured with **point-count survey records**

Process

Data Cleaning Steps:
1. Remove missing or inconsistent values
2. Standardize column names for easy analysis

3. Convert categorical variables (e.g., species names) into factors
4. Aggregate data by **location type** and **species**

```
# Load necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readxl)
library(knitr)

# Load dataset
bird_data <- read_excel("NCRN LAND Bird Monitoring Data 2007 - 2017_Public.xlsx")
str(bird_data)
```

```
## tibble [14,230 x 29] (S3: tbl_df/tbl/data.frame)
##  $ Admin_Unit_Code            : chr [1:14230] "ANTI" "ANTI" "ANTI" "ANTI" ...
##  $ Sub_Unit_Code              : chr [1:14230] "NA" "NA" "NA" "NA" ...
##  $ Site_Name                  : chr [1:14230] "ANTI 2" "ANTI 2" "ANTI 2" "ANTI 2" ...
##  $ Plot_Name                  : chr [1:14230] "ANTI-0207" "ANTI-0207" "ANTI-0207" "ANTI-0207" ...
##  $ Location_Type              : chr [1:14230] "Forest" "Forest" "Forest" "Forest" ...
##  $ Year                       : num [1:14230] 2008 2008 2008 2008 2008 ...
##  $ Date                       : POSIXct[1:14230], format: "2008-05-28" "2008-05-28" ...
##  $ Start_Time                 : POSIXct[1:14230], format: "1899-12-31 06:42:00" "1899-12-31 06:42:00
##  $ End_Time                   : POSIXct[1:14230], format: "1899-12-31 06:52:00" "1899-12-31 06:52:00
##  $ Observer                   : chr [1:14230] "Desiree Narango" "Desiree Narango" "Desiree Narango"
##  $ Visit                      : num [1:14230] 1 1 1 1 1 1 1 1 1 ...
##  $ Interval_Length            : chr [1:14230] "0-2.5 min" "0-2.5 min" "0-2.5 min" "0-2.5 min" ...
##  $ ID_Method                  : chr [1:14230] "Calling" "Calling" "Calling" "Singing" ...
##  $ Distance                   : chr [1:14230] "<= 50 Meters" "50 - 100 Meters" "50 - 100 Meters" "<=
##  $ Flyover_Observed           : logi [1:14230] FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ Sex                        : chr [1:14230] "Undetermined" "Undetermined" "Undetermined" "Male" ..
##  $ Common_Name                : chr [1:14230] "Acadian Flycatcher" "Blue-gray Gnatcatcher" "Acadian
##  $ Scientific_Name            : chr [1:14230] "Empidonax virescens" "Polioptila caerulea" "Empidonax
##  $ AcceptedTSN                : num [1:14230] 178339 179853 178339 179236 179731 ...
##  $ NPSTaxonCode               : num [1:14230] 85824 88038 85824 95300 87106 ...
##  $ AOU_Code                   : chr [1:14230] "ACFL" "BGGN" "ACFL" "AMGO" ...
##  $ PIF_Watchlist_Status       : logi [1:14230] FALSE FALSE FALSE FALSE FALSE TRUE ...
##  $ Regional_Stewardship_Status: logi [1:14230] TRUE FALSE TRUE FALSE FALSE TRUE ...
##  $ Temperature                : num [1:14230] 15.6 15.6 15.6 15.6 15.6 ...
##  $ Humidity                   : num [1:14230] 57 57 57 57 57 57 57 57 57 57 ...
##  $ Sky                        : chr [1:14230] "Cloudy/Overcast" "Cloudy/Overcast" "Cloudy/Overcast"
##  $ Wind                       : chr [1:14230] "Calm (< 1 mph) smoke rises vertically" "Calm (< 1 mph
##  $ Disturbance                : chr [1:14230] "No effect on count" "No effect on count" "No effect o
##  $ Initial_Three_Min_Cnt      : logi [1:14230] TRUE TRUE TRUE TRUE TRUE TRUE ...
```

```r
# Clean the dataset
bird_data_clean <- bird_data %>%
  filter(!is.na(Common_Name), !is.na(Location_Type)) %>%
  mutate(location_type = as.factor(Location_Type))

# Verify clean dataset
summary(bird_data_clean)
```

```
##  Admin_Unit_Code    Sub_Unit_Code       Site_Name          Plot_Name
##  Length:14230       Length:14230       Length:14230       Length:14230
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Location_Type           Year           Date
##  Length:14230       Min.   :2007   Min.   :2007-05-11 00:00:00.00
##  Class :character   1st Qu.:2015   1st Qu.:2015-06-22 00:00:00.00
##  Mode  :character   Median :2016   Median :2016-05-24 00:00:00.00
##                     Mean   :2015   Mean   :2015-11-12 10:17:29.39
##                     3rd Qu.:2017   3rd Qu.:2017-05-28 00:00:00.00
##                     Max.   :2017   Max.   :2017-07-17 00:00:00.00
##
##    Start_Time                      End_Time
##  Min.   :1899-12-31 05:12:00.00   Min.   :1899-12-31 05:22:00.0
##  1st Qu.:1899-12-31 06:26:00.00   1st Qu.:1899-12-31 06:37:00.0
##  Median :1899-12-31 07:34:00.00   Median :1899-12-31 07:44:00.0
##  Mean   :1899-12-31 07:38:00.86   Mean   :1899-12-31 07:48:09.3
##  3rd Qu.:1899-12-31 08:42:00.00   3rd Qu.:1899-12-31 08:52:00.0
##  Max.   :1899-12-31 10:53:00.00   Max.   :1899-12-31 11:03:00.0
##
##    Observer            Visit       Interval_Length     ID_Method
##  Length:14230       Min.   :1.000   Length:14230       Length:14230
##  Class :character   1st Qu.:1.000   Class :character   Class :character
##  Mode  :character   Median :2.000   Mode  :character   Mode  :character
##                     Mean   :1.869
##                     3rd Qu.:3.000
##                     Max.   :3.000
##
##    Distance         Flyover_Observed     Sex             Common_Name
##  Length:14230       Mode :logical    Length:14230       Length:14230
##  Class :character   FALSE:12339      Class :character   Class :character
##  Mode  :character   TRUE :1891       Mode  :character   Mode  :character
##
##
##
##
##  Scientific_Name     AcceptedTSN      NPSTaxonCode      AOU_Code
##  Length:14230       Min.   :174773   Min.   :     0   Length:14230
##  Class :character   1st Qu.:178581   1st Qu.: 84865   Class :character
##  Mode  :character   Median :179150   Median : 87409   Mode  :character
##                     Mean   :201009   Mean   :108632
##                     3rd Qu.:179443   3rd Qu.: 94215
```

```
##                    Max.   :950061   Max.   :926917
##                    NA's   :125
##  PIF_Watchlist_Status Regional_Stewardship_Status  Temperature
##  Mode :logical        Mode :logical               Min.   : 0.00
##  FALSE:14046          FALSE:11085                 1st Qu.:18.80
##  TRUE :184            TRUE :3145                  Median :22.40
##                                                   Mean   :22.16
##                                                   3rd Qu.:25.20
##                                                   Max.   :41.20
##
##     Humidity         Sky                Wind           Disturbance
##  Min.   : 0.00   Length:14230       Length:14230       Length:14230
##  1st Qu.:72.80   Class :character   Class :character   Class :character
##  Median :81.70   Mode  :character   Mode  :character   Mode  :character
##  Mean   :78.46
##  3rd Qu.:87.70
##  Max.   :98.00
##
##  Initial_Three_Min_Cnt   location_type
##  Mode :logical         Forest   : 2801
##  FALSE:6207            Grassland:11429
##  TRUE :8023
##
##
##
##
```

#Analyze ## 1. Bird Species Count by Habitat

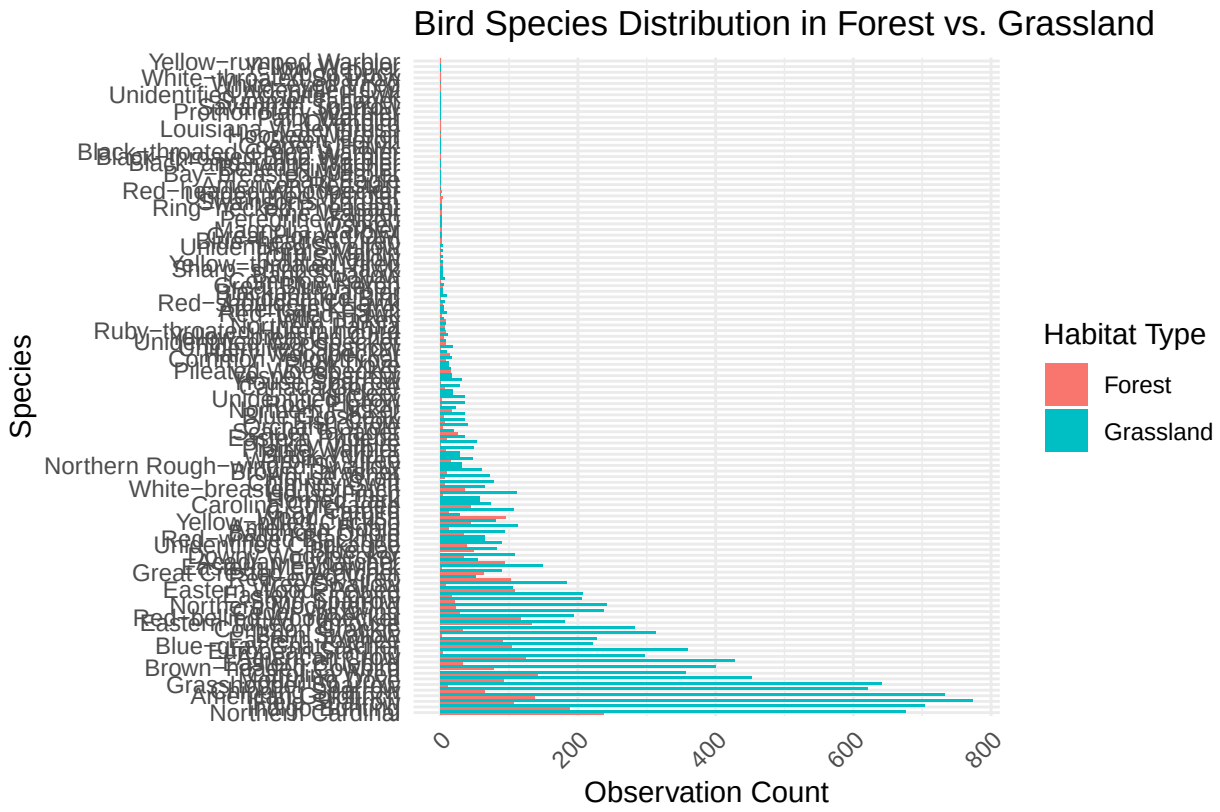
```r
species_distribution <- bird_data_clean %>%
  group_by(location_type, Common_Name) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(location_type, desc(count))

# Display top species per habitat
kable(head(species_distribution, 10), caption = "Top 10 Bird Species by Habitat Type")
```

Table 1: Top 10 Bird Species by Habitat Type

location_type	Common_Name	count
Forest	Northern Cardinal	238
Forest	Indigo Bunting	188
Forest	Carolina Wren	141
Forest	American Goldfinch	137
Forest	Eastern Tufted Titmouse	133
Forest	American Crow	124
Forest	Red-bellied Woodpecker	117
Forest	Eastern Wood-Pewee	108
Forest	Field Sparrow	107
Forest	Blue-gray Gnatcatcher	104

2. Visualization: Species Distribution in Forest vs. Grassland

```
ggplot(species_distribution, aes(x = reorder(Common_Name, -count), y = count, fill = location_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip() +
  labs(
    title = "Bird Species Distribution in Forest vs. Grassland",
    x = "Species",
    y = "Observation Count",
    fill = "Habitat Type",
    caption = "Source: NCRN LAND Bird Monitoring Data (2007 - 2017), catalog.data.gov"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Source: NCRN LAND Bird Monitoring Data (2007 − 2017), catalog.data.gov

3. Species Richness by Habitat

```
species_richness <- bird_data_clean %>%
  group_by(location_type) %>%
  summarise(total_species = n_distinct(Common_Name))

# Display species richness
kable(species_richness, caption = "Species Richness in Forest and Grassland Habitats")
```

Table 2: Species Richness in Forest and Grassland Habitats

location_type	total_species
Forest	90
Grassland	107

Share

Key Findings

1. Species richness and abundance differ significantly between forests and grasslands.
2. Forests support a higher diversity of species, while grasslands are home to species adapted to open environments.
3. Certain species are exclusively found in forested areas, while others thrive in grasslands.

These insights are valuable for conservation planning and ecological management in the National Capital Region.

Act

Key Takeaways

- Forests support a wider variety of bird species than grasslands, but grassland species are uniquely adapted to open environments.
- Conservation efforts should prioritize maintaining habitat diversity to protect both forest-dwelling and grassland-specialist birds.
- Ongoing monitoring and data collection are essential for tracking long-term changes in avian populations.

Recommendations

1. **Enhance Habitat Protection:** Ensure conservation policies support both forest and grassland ecosystems.
2. **Increase Monitoring Efforts:** Expand data collection to assess long-term populations trends.
3. **Community Engagement:** Educate local communities on the importance of habitat conservation.

This analysis provides data-driven insights that can guide ecological decision-making and habitat making strategies.