Activity: Bird Species

Mohil Patel

03/02/2022

Set Up

```
# Load `tidyverse` and `dcData` packages
library(tidyverse)
library(dcData)



# Load the `OrdwayBirds` data set into your RStudio environment from the `dcData` package
data("OrdwayBirds")
library(mosaic)



# Show a few commands used to inspect the data set a few different ways

glimpse(OrdwayBirds)
```

```
Rows: 15,829
Columns: 26
$ bogus              <chr> ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", ".", "."~
$ Timestamp          <chr> "4/14/2010 13:20:56", "", "5/13/2010 16:00:30", "5/13/2010 16:02:15", "5/13/2010 16~
$ Year               <chr> "1972", "", "1972", "1972", "1972", "1972", "1972", "1972", "1972", "1972", "1972",~
$ Day                <chr> "16", "", "16", "16", "16", "16", "16", "16", "16", "16", "17", "18", "18", "18", "~
$ Month              <chr> "7", "", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7", "7",~
$ CaptureTime        <chr> "7:00:00", "", "7:00:00", "7:00:00", "7:00:00", "7:00:00", "13:00:00", "13:00:00", ~
$ SpeciesName        <chr> "Song Sparrow", "", "Song Sparrow", "Field Sparrow", "Field Sparrow", "Song Sparrow~
$ Sex                <chr> "U", "", "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "U", "M", "U", "U",~
$ Age                <chr> "AHY (After Hatch Year)", "", "AHY (After Hatch Year)", "AHY (After Hatch Year)", "~
$ BandNumber         <chr> "107-151187", "", "107-151187", "1260-74572", "1260-74541", "107-151188", "1260-745~
$ TrapID             <chr> "warbler", "", "Warbler", "Warbler", "Elevator", "Elevator", "Net-Blue", "Warbler",~
$ Weather            <chr> "none", "", "none", "none", "none", "none", "none", "none", "none", "none", "Cloudy~
$ BandingReport      <chr> "check mark in record", "", "check mark in record", "no check mark in record", "no ~
$ RecaptureYN        <chr> "no check", "", "no check", "check or yes", "check or yes", "no check", "no check",~
$ RecaptureMonth     <chr> "none", "", "none", "7", "4", "none", "none", "none", "none", "none", "none", "none~
$ RecaptureDay       <chr> "none", "", "none", "9", "25", "none", "none", "none", "none", "none", "none", "non~
$ Condition          <chr> "none", "", "none", "none", "none", "none", "none", "none", "none", "none", "none",~
$ Release            <chr> "none", "", "none", "none", "none", "none", "none", "none", "none", "none", "none",~
$ Comments           <chr> "none", "", "none", "none", "none", "none", "none", "none", "none", "none", "none",~
$ DataEntryPerson    <chr> "Jerald Dosch", "Caitlin Baker", "Caitlin Baker", "Caitlin Baker", "Caitlin Baker",~
$ Weight             <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
$ WingChord          <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
$ Temperature        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
$ RecaptureOriginal <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
$ RecapturePrevious <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
$ TailLength         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "",~
```

Activity

When you're finished with this activity, you will have a graph that shows what time of year various species appear at the Katherine Ordway Natural History Study Area in Inver Grove Heights, MN.

Step 0

Before we begin, the book gives us some instructions to select a few key variables and clean up the date formatting. Use the commands given in the book, but be sure to study the functions used so you know what is happening. In this case, we select `SpeciesName`, `Month`, and `Day`. Then a `mutate()` command converts each variable to character, and then converts the resulting "character string" to a number.

```
# Get the data table & clean up dates (see Data Computing eBook)
OrdwayBirds <-
  OrdwayBirds %>%
  select( SpeciesName, Month, Day ) %>%
  mutate( Month = as.numeric(as.character(Month)),
          Day = as.numeric(as.character(Day)))



# Inspect resulting data
```

Step 1

```
# unique species in original data (including mis-spellings)
OrdwayBirds %>%
  summarise(count = n_distinct(SpeciesName))
```

count <int>
275

1 row

```



# unique species in the clean list
OrdwaySpecNameCount <-
  OrdwaySpeciesNames %>%
  summarise(count = n_distinct(SpeciesNameCleaned))
OrdwaySpecNameCount
```

count <int>
109

1 row

```
NA
NA
```

Step 2

```
# inspect OrdwaySpeciesNames
Corrected <-
  OrdwayBirds %>%
  inner_join( OrdwaySpeciesNames) %>%
  select(Species = SpeciesNameCleaned, Month, Day) %>%
  na.omit()
```

```
Joining, by = "SpeciesName"
```

```
# inner join




# inspect the data after the join
```

Step 3

We want to identify the top 5-6 bird species sighted at the Katherine Ordway Natural History Study Area.

```
# count sightings by species in descending order
CountCorrect <-
  Corrected %>%
  group_by(Species) %>%
  summarise(count=n()) %>%
  arrange(desc(count))




# a hint on p.164 recommends displaying the top 10 species to choose our threshold
topSixSpec <-
  CountCorrect %>%
  head(n = 6) %>%
  .$Species
topSixSpec
```

```
[1] "Slate-colored Junco"    "Tree Swallow"           "Black-capped Chickadee" "American Goldfinch"
[5] "Field Sparrow"          "Lincoln's Sparrow"
```

```
# define a sightings threshold; produce a table with only birds belonging to major species

Majors <-
  Corrected %>%
  filter(Species %in% topSixSpec)
```

Step 4

We now want to produce a month-by-month count for each of the major species, and then construct a chart to tell the story of what time of year the various species appear.

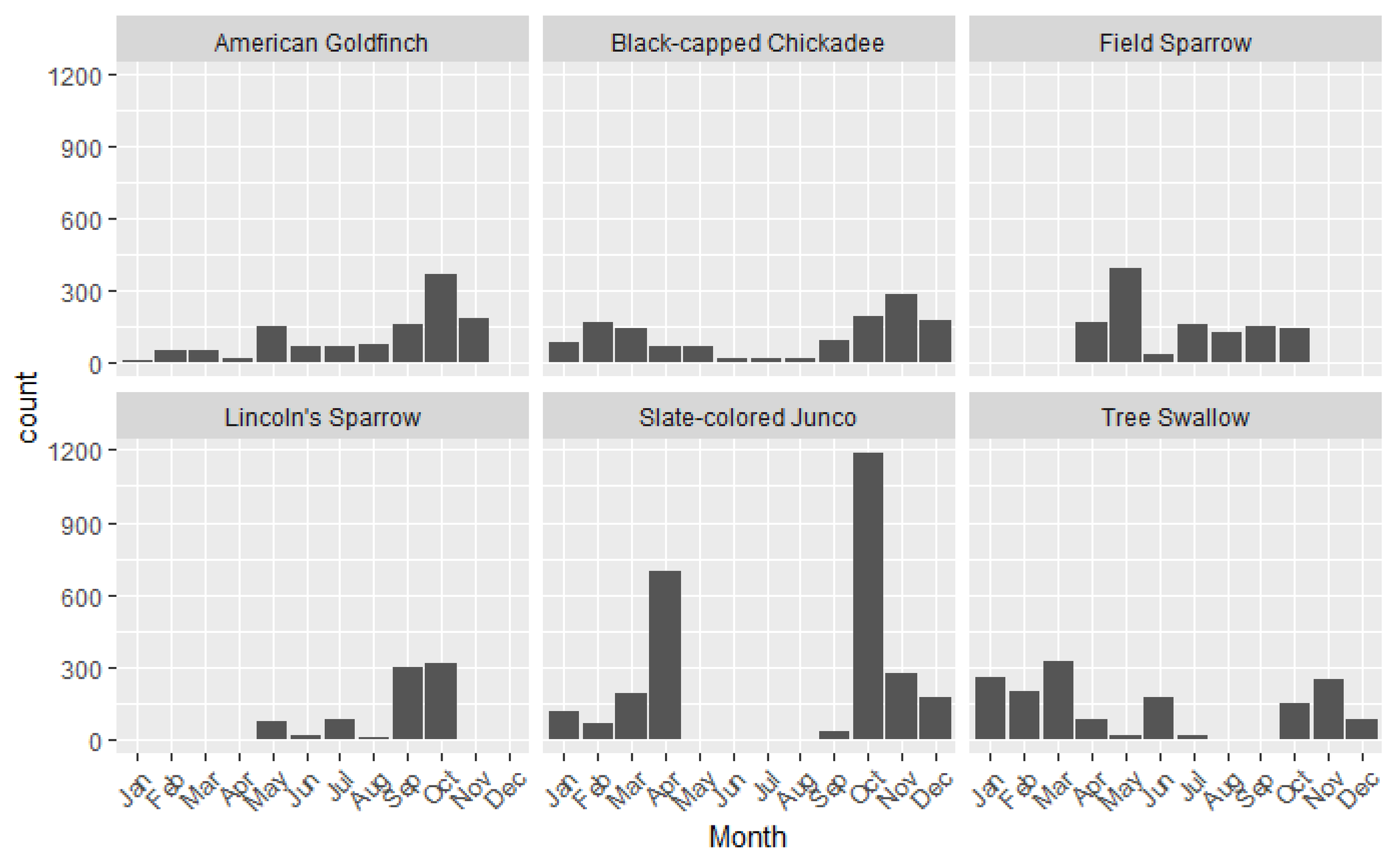
```
# prepare glyph-ready data table called `ByMonth`
ByMonth <- Majors %>%
  group_by(Species, Month) %>%
  summarise(count = n()) %>%
  arrange(Month)
```

```
`summarise()` has grouped output by 'Species'. You can override using the `.groups` argument.
```

```
CrazyMonth <- monthAbbr <- with(Majors,
                  plyr::mapvalues(Month, from = 1:12,
                                  to = month.abb))
CrazyMonth <- factor(CrazyMonth, levels = month.abb)
Majors$Month <- CrazyMonth
```

Now, using `mplot()` or `esquisser()` in the console, we can configure a reasonable graph and choose "show expression" to get `ggplot2` that can be embedded in our report.

```
# ggplot command
Majors %>%
  ggplot(aes(x = Month)) +
  geom_bar() +
  facet_wrap(~ Species) +
  theme(axis.text.x=element_text(angle=45,hjust=1))
```




```
NA
NA
```

According to the graph, we can answer the questions posed:

1. Which species are present year-round?

American Goldfinch Black-capped chickadee

2. Which species are migratory, that is, primarily present in one or two seasons?

Lincolns Sparrow Slate-colored Junco

3. What is the peak month for each major species? (bullet list)

American Goldfinch: October Black-capped Chickadee: November Field Sparrow: May Lincolns Sparrow: October Slate-colored Junco: October Tree swallow: March

4. Which major species are seen in good numbers for at least 6 months of the year?

American Goldfinch Black-capped Chickadee Tree swallow