

Exploratory data visualizations

We requested this data through Cornell, who approved and granted access to the full eBird data. We chose to restrict our modeling to the Northern Cardinal in Illinois to have a manageable size to work with.

```
library(tidyverse)

##  Attaching packages tidyverse 1.2.1
##    ggplot2 3.3.2      purrr   0.3.3
##    tibble  2.1.3      dplyr   0.8.3
##    tidyr   1.0.0      stringr 1.4.0
##    readr   1.1.1      forcats 0.2.0
##    Conflicts          tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(maps)

##
## Attaching package: 'maps'
## The following object is masked from 'package:purrr':
## 
##     map
library(ggmap)
library(ggthemes)
library(mapdata)

ebd <- read_tsv("../data/raw/ebd/ebd_US-IL_norcar_relSep-2020/ebd_US-IL_norcar_relSep-2020.txt", col_na

## Warning: Missing column names filled in: 'X47' [47]

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `LAST EDITED DATE` = col_datetime(format = ""),
##   `TAXONOMIC ORDER` = col_integer(),
##   `BCR CODE` = col_integer(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double(),
##   `OBSERVATION DATE` = col_date(format = ""),
##   `TIME OBSERVATIONS STARTED` = col_time(format = ""),
##   `DURATION MINUTES` = col_integer(),
##   `EFFORT DISTANCE KM` = col_double(),
##   `EFFORT AREA HA` = col_double(),
##   `NUMBER OBSERVERS` = col_integer(),
##   `ALL SPECIES REPORTED` = col_integer(),
##   `HAS MEDIA` = col_integer(),
##   APPROVED = col_integer(),
##   REVIEWED = col_integer()
## )

## See spec(...) for full column specifications.

## Warning in rbind(names(probs), probs_f): number of columns of result is not
```

```

## a multiple of vector length (arg 1)
## Warning: 1852 parsing failures.
## row # A tibble: 5 x 5 col      row col      expected      actual file
## ... ..... . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
## See problems(...) for more details.

# ebd
names(ebd) <- gsub(" ", "_", names(ebd))

states <- map_data("state")
IL <- subset(states, region %in% c("illinois"))
counties <- map_data("county")
IL_county <- subset(counties, region == "illinois")

il_base <- ggplot(data = IL, mapping = aes(x = long, y = lat, group = group)) +
  coord_fixed(1.3) +
  geom_polygon(color = "black", fill = NA) +
  theme_tufte()

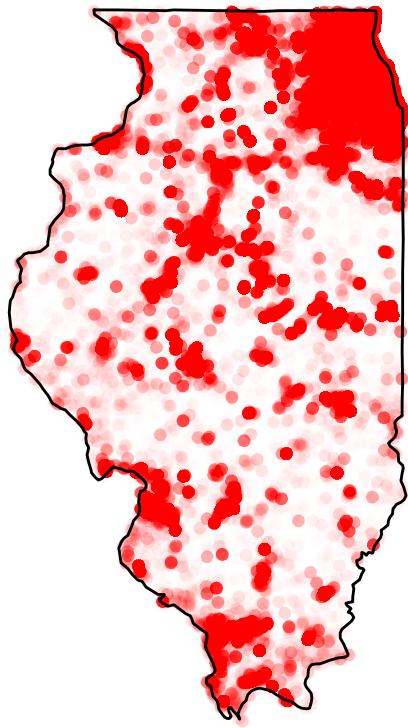
il_observations <- ggplot(data = IL, mapping = aes(x = long, y = lat, group = group)) +
  geom_point(data = ebd, inherit.aes = FALSE, aes(x = LONGITUDE, y = LATITUDE), alpha = .01, color = 'red') +
  labs(title = "eBird Observations of Northern Cardinals") +
  geom_polygon(color = "black", fill = NA) +
  theme_tufte() + theme(
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.y = element_blank(),
    axis.text.y = element_blank(),
    axis.title.y = element_blank()
  )

il_observations

## Warning: Removed 2 rows containing missing values (geom_point).

```

eBird Observations of Northern Cardinals

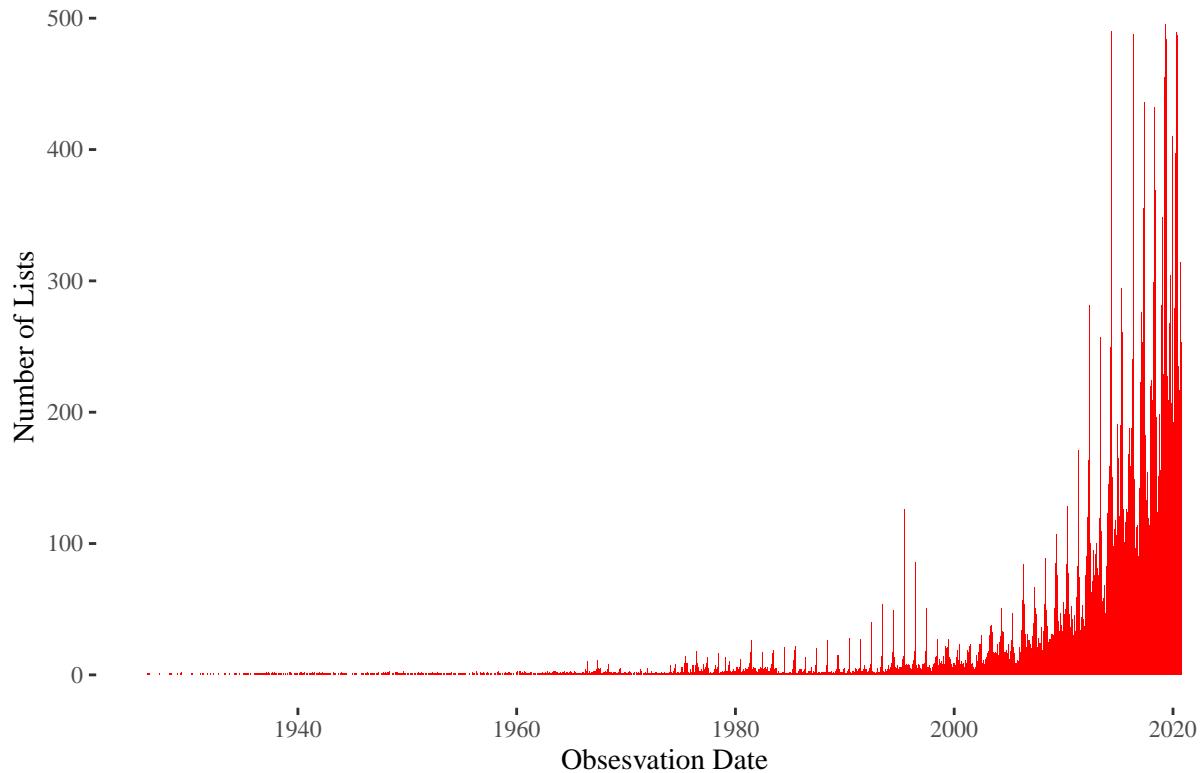


```
date_observations <- ggplot(data = ebd, aes(x = OBSERVATION_DATE)) +
  geom_histogram(stat = 'count', fill = 'red') +
  ylim(0,500) +
  labs(title = "eBird Observations of Northern Cardinals", y = "Number of Lists", x = "Observation Date")
  theme_tufte()

## Warning: Ignoring unknown parameters: binwidth, bins, pad
date_observations

## Warning: Removed 2 rows containing non-finite values (stat_count).
## Warning: Removed 15 rows containing missing values (geom_bar).
```

eBird Observations of Northern Cardinals



```
filter(ebd, OBSERVATION_DATE >= '2010-09-01')

## # A tibble: 353,013 x 47
##   GLOBAL_UNIQUE_I... LAST_EDITED_DATE    TAXONOMIC_ORDER CATEGORY
##   <chr>              <dttm>                <int> <chr>
## 1 URN:CornellLab0... 2013-10-31 01:56:17     33194 species
## 2 URN:CornellLab0... 2014-03-25 14:23:01     33194 species
## 3 URN:CornellLab0... 2011-03-05 19:22:47     33194 species
## 4 URN:CornellLab0... 2019-02-17 20:48:12     33194 species
## 5 URN:CornellLab0... 2018-08-03 16:48:42     33194 species
## 6 URN:CornellLab0... 2017-08-16 06:56:17     33194 species
## 7 URN:CornellLab0... 2016-06-26 18:24:35     33194 species
## 8 URN:CornellLab0... 2020-05-11 09:53:40     33194 species
## 9 URN:CornellLab0... 2018-08-03 16:48:57     33194 species
## 10 URN:CornellLab0... 2014-10-15 20:53:32    33194 species
## # ... with 353,003 more rows, and 43 more variables: COMMON_NAME <chr>,
## #   SCIENTIFIC_NAME <chr>, SUBSPECIES_COMMON_NAME <chr>,
## #   SUBSPECIES_SCIENTIFIC_NAME <chr>, OBSERVATION_COUNT <chr>,
## #   BREEDING_BIRD_ATLAS_CODE <chr>, BREEDING_BIRD_ATLAS_CATEGORY <chr>,
## #   `AGE/SEX` <chr>, COUNTRY <chr>, COUNTRY_CODE <chr>, STATE <chr>,
## #   STATE_CODE <chr>, COUNTY <chr>, COUNTY_CODE <chr>, IBA_CODE <chr>,
## #   BCR_CODE <int>, USFWS_CODE <chr>, ATLAS_BLOCK <chr>, LOCALITY <chr>,
## #   LOCALITY_ID <chr>, LOCALITY_TYPE <chr>, LATITUDE <dbl>,
## #   LONGITUDE <dbl>, OBSERVATION_DATE <date>,
## #   TIME_OBSERVATIONS_STARTED <drtn>, OBSERVER_ID <chr>,
## #   SAMPLING_EVENT_IDENTIFIER <chr>, PROTOCOL_TYPE <chr>,
## #   PROTOCOL_CODE <chr>, PROJECT_CODE <chr>, DURATION_MINUTES <int>,
```

```

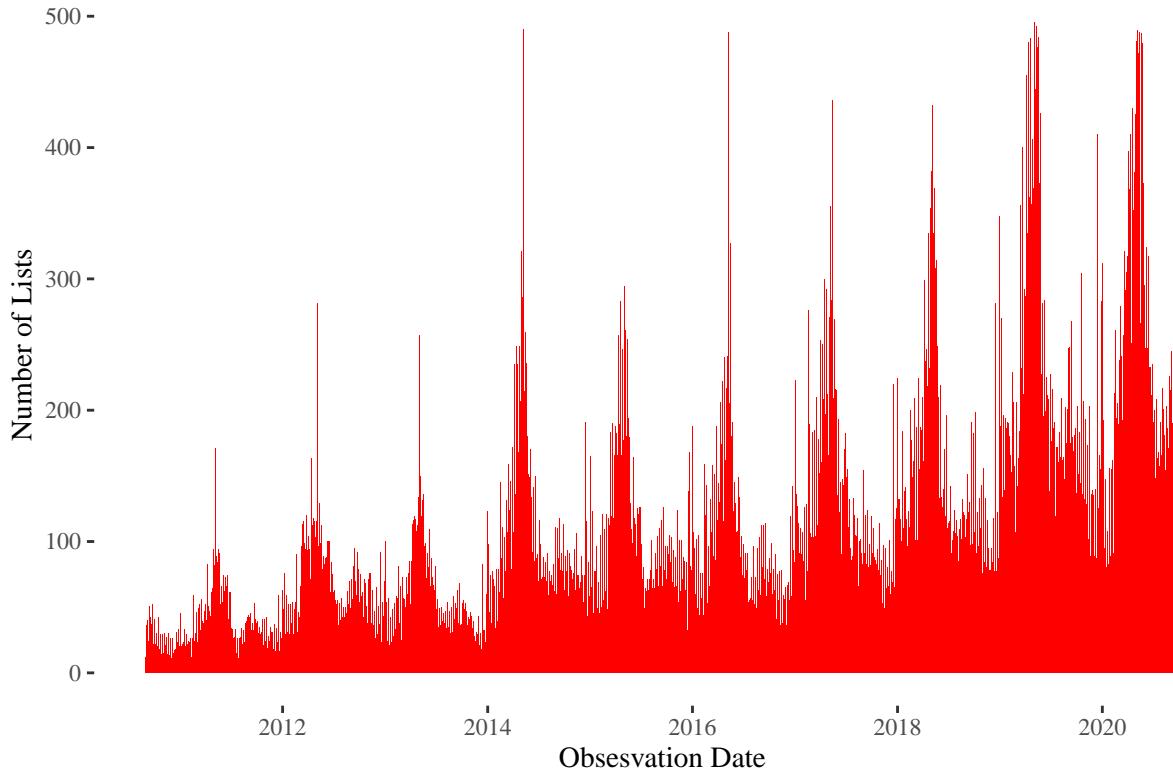
## #  EFFORT_DISTANCE_KM <dbl>, EFFORT_AREA_HA <dbl>,
## #  NUMBER_OBSERVERS <int>, ALL_SPECIES_REPORTED <int>,
## #  GROUP_IDENTIFIER <chr>, HAS_MEDIA <int>, APPROVED <int>,
## #  REVIEWED <int>, REASON <chr>, TRIP_COMMENTS <chr>,
## #  SPECIES_COMMENTS <chr>, X47 <chr>
decade_observations <- ggplot(data = filter(ebd, OBSERVATION_DATE >= '2010-09-01'), aes(x = OBSERVATION_DATE))
  geom_histogram(stat = 'count', fill = 'red') +
  ylim(0,500) +
  labs(title = "eBird Observations of Northern Cardinals", y = "Number of Lists", x = "Observation Date")
  theme_tufte()

## Warning: Ignoring unknown parameters: binwidth, bins, pad
decade_observations

## Warning: Removed 15 rows containing missing values (geom_bar).

```

eBird Observations of Northern Cardinals



References

- Drawing beautiful maps programmatically with R, sf and ggplot2 — Part 1: Basics
- (<https://stackoverflow.com/questions/53601185/plotting-illinois-with-ggmap-and-ggplot-in-r>)