

New York City Dog Bite Analysis

Ian Young

Contents

Load Packages	2
Connect to API & Compile Dataset	2
Initial Data Cleaning	2
Background & Initial Exploration	3
Further Cleaning	4
Analysis	4
Exploratory Data Analysis	4
Mapping Dog Bites In NYC	4
Forecasting Dog Bites In NYC	5

Load Packages

```
library(tidyverse); library(showtext); library(scales)
font_add_google("Roboto Condensed")
showtext_auto()
library(httr2)
library(gt)
```

Connect to API & Compile Dataset

The httr2 and purrr packages are used to connect to an API from New York City and convert dog bite data for the years 2015 to 2022 from an unstructured format into a data frame.

```
# Collect data on dog bites in NYC
dog_bites <-
  request("https://data.cityofnewyork.us/resource/rsgh-akpg.json") |>
  req_url_path_append("?$limit=30000") |> # change call limit to collect all records
  req_perform() |>
  resp_body_json() |>
  map_dfr(~ as_tibble(.)) # convert lists into one tibble
```

Initial Data Cleaning

An initial organizing and cleaning of the data takes place using . This involves converting the date column that is stored as character data type to a date type, and adding a individual column for year, month, and day. This is done using the lubridate package. The uniqueid column was not recorded properly in the initial data set, and a more helpful uniqueid column is created. Lastly, the species column is dropped as it provides no information (all values in it are “DOG”) and the columns are rearranged.

```
# Correct unique ID, adjust/create date columns, select relevant variables
dog_bites_clean <-
  dog_bites |>
  arrange(dateofbite) |>
  mutate(uniqueid = row_number(),
         dateofbite = as_date(dateofbite),
         biteyear = year(dateofbite),
         bitemonth = month(dateofbite),
         biteday = day(dateofbite)) |>
  select(uniqueid, dateofbite, biteyear,
         bitemonth, biteday, zipcode,
         borough, breed, gender,
         age, spayneuter)
```

Background & Initial Exploration

Below is a sample of 10 observations from the initial cleaned dog bite data. There are a total of 26,127 observations and 11 variables (3 data columns not displayed in the table below). Each observation represents a single unique dog bite incident collected by NYC between the start of 2015 and the end of 2022. The data was collected online, by mail/fax, and by phone by the health departments animal bite unit. As a result columns like zipcode, breed, and age all contain a range of values including missing data. For the spayneuter column FALSE represents dogs that have not had their reproductive organs surgically removed, as well as cases where it is unknown if the dog involved in the incident has or has not been neutered. Note: The skimr package and function skim are not rendered below, they were used in gathering a quick general understanding of the data.

```
# Sample of 10 observations
set.seed(30)
dog_bites_clean |>
  select(-biteyear, -bitemonth, -biteday) |>
  mutate(dateofbite = format(dateofbite, "%Y-%m-%d")) |>
  sample_n(10) |>
  gt() |>
  tab_header(
    title = md("**NYC Dog Bite Reports 2015-2022**"),
    subtitle = md("Sample of 10 Observations")
  ) |>
  tab_source_note(
    source_note = md("Source:
https://data.cityofnewyork.us/Health/DOHMH-Dog-Bite-Data/rsgh-akpg/about\_data")
  ) |>
  tab_footnote(
    footnote = md("**FALSE** represents dogs that have not been neutered **OR**
it was unknown."),
    locations = cells_column_labels(
      columns = spayneuter
    )
  )
)
```

NYC Dog Bite Reports 2015-2022

Sample of 10 Observations

uniqueid	dateofbite	zipcode	borough	breed	gender	age	spayneuter ¹
6474	2016-11-21	11694	Queens	Pit Bull	M	3	FALSE
23858	2022-05-15	10303	Staten Island	Chihuahua	F	1	FALSE
19725	2020-12-28	NA	Manhattan	Shih Tzu	U	NA	FALSE
9676	2017-10-20	NA	Bronx	Pit Bull	U	NA	FALSE
26122	2022-12-31	10463	Bronx	UNKNOWN	U	NA	FALSE
14651	2019-04-30	10463	Bronx	PIT BULL BLUE NOSE	M	5	FALSE
9245	2017-09-05	10306	Staten Island	American Staffordshire Terrier	F	5	TRUE
4284	2016-04-06	11362	Queens	MALTESE POODLE MIX	U	NA	FALSE
26035	2022-12-18	11229	Brooklyn	UNKNOWN	U	NA	FALSE
2884	2015-10-20	11224	Brooklyn	NA	U	NA	FALSE

¹ **FALSE** represents dogs that have not been neutered **OR** it was unknown.

Source: https://data.cityofnewyork.us/Health/DOHMH-Dog-Bite-Data/rsgh-akpg/about_data

Further Cleaning

```
# Clean up age column, join in longitude/latitude for each zip code  
# breed make everything lower case and remove extra spaces
```

Analysis

Exploratory Data Analysis

```
# Descriptive Statistics/Plots
```

Mapping Dog Bites In NYC

```
# Load mapping package  
library(sf)  
  
# Read in shape file  
shape_nyc <- st_read("Borough Boundaries.geojson", quiet = TRUE)  
  
# Begin creation of map  
ggplot() +  
  geom_sf(data = shape_nyc, aes(fill = boro_name), show.legend = FALSE) +  
  theme_void()
```



Forecasting Dog Bites In NYC

```
# Create data frame for time series analysis
dog_ts <-
  dog_bites_clean |>
  group_by(biteyear, bitemonth) |>
  count() |>
  mutate(
    season = case_when(
      bitemonth %in% c(3, 4, 5) ~ "spring",
      bitemonth %in% c(6, 7, 8) ~ "summer",
      bitemonth %in% c(9, 10, 11) ~ "fall",
      bitemonth %in% c(12, 1, 2) ~ "winter"),
    date = make_date(biteyear, bitemonth)
  ) |>
  ungroup() |>
  select(date, season, n)

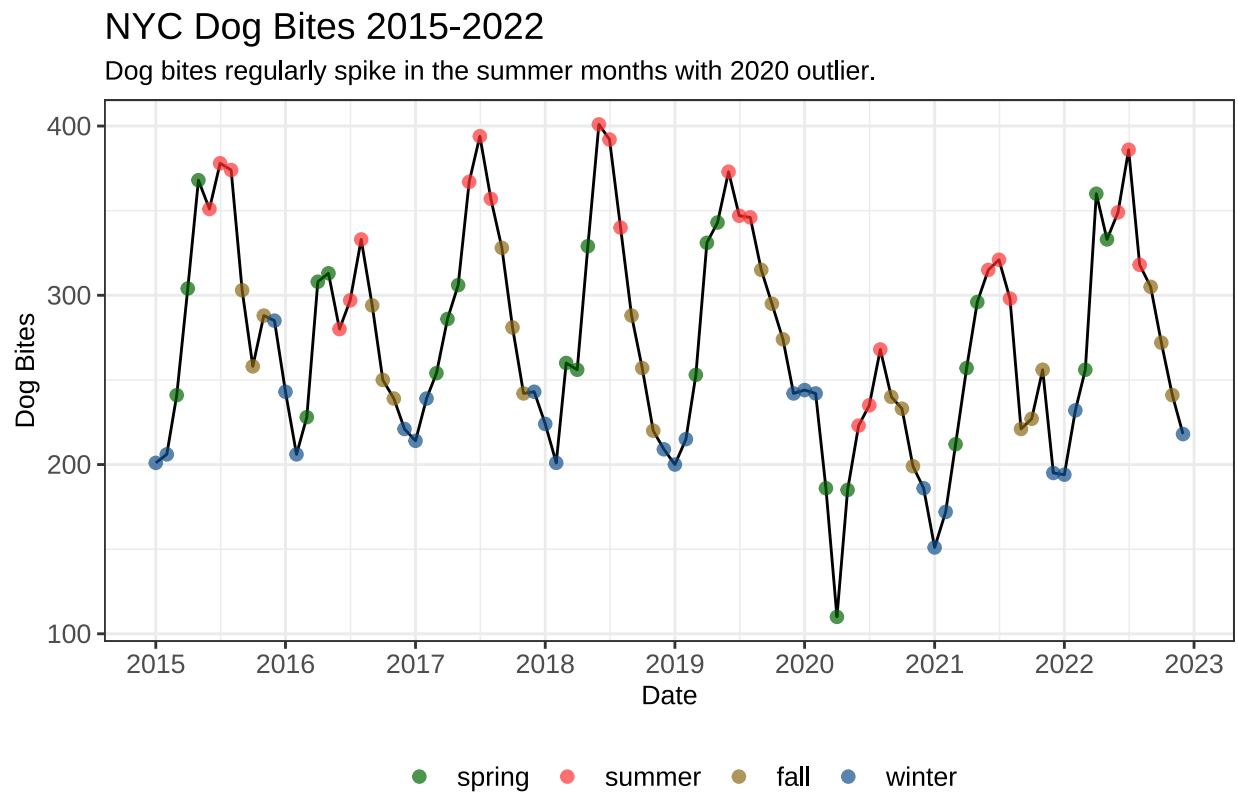
# Initial plot of time series data
dog_ts |>
  mutate(season = fct_relevel(season, "spring", "summer", "fall", "winter")) |>
  ggplot(aes(x = date, y = n)) +
    geom_line() +
    geom_point(aes(color = season), size = 3, alpha = 0.7, shape = 20) +
```

```

scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
labs(x = "Date",
     y = "Dog Bites",
     title = "NYC Dog Bites 2015-2022",
     subtitle = "Dog bites regularly spike in the summer months with 2020 outlier.",
     caption = "Source: NYCOpenData") +
scale_color_manual(values = c("spring" = "darkgreen",
                              "summer" = "firebrick1",
                              "fall" = "goldenrod4",
                              "winter" = "dodgerblue4")) +

theme_bw() +
theme(
  axis.text = element_text(size = 10),
  axis.title = element_text(size = 10),
  plot.title = element_text(size = 14),
  plot.subtitle = element_text(size = 10),
  plot.caption = element_text(size = 8),
  legend.position = "bottom",
  legend.direction = "horizontal",
  legend.title = element_blank(),
  legend.text = element_text(size = 10)
)

```



Source: NYCOpenData