Rat Sightings

Emily Martinez

2024-01-23

NYC Rat Sightings

Rats have been a persistent issue in New York City throughout the years and are commonly perceived as an integral part of the city's cultural landscape.To address this concern, the NYC Health Department oversees rodent complaints, which can be filed online or by calling 3-1-1. The NYC Health Department plays a crucial role in inspecting both private and public properties for rat infestations. This data set spans from 2010 to 2017, providing detailed information such as date, location (latitude/longitude), type of structure, and borough for reported rat sightings. Analyzing this data can yield valuable insights into the spatial and temporal distribution of rat sightings across different neighborhoods and structures in New York City. By leveraging this information, city authorities can implement targeted interventions and preventive measures to mitigate rat infestations and improve public health and sanitation standards throughout the city.

Based on our given data set, we can:

1. Investigate the number of cases between the type of structure
2. Investigate the number of cases between New York City boroughs
3. Examine the duration of case resolution for individual incidents within the top 5 versus bottom 5 areas with rat infestations
4. Explore the number of reported cases throughout the years
5. Determine if there is a significant difference in the number of days spent per case between seasons

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
```

```
rats <- read.csv("Rat_Sightings.csv")
```

```
head(rats)
```

```
##   Unique.Key           Created.Date            Closed.Date Agency
## 1   31464015 09/04/2015 12:00:00 AM 09/18/2015 12:00:00 AM  DOHMH
## 2   31464024 09/04/2015 12:00:00 AM 10/28/2015 12:00:00 AM  DOHMH
## 3   31464025 09/04/2015 12:00:00 AM                         DOHMH
```

```
## 4   31464026 09/04/2015 12:00:00 AM 09/14/2015 12:00:00 AM  DOHMH
## 5   31464027 09/04/2015 12:00:00 AM 09/22/2015 12:00:00 AM  DOHMH
## 6   31464188 09/04/2015 12:00:00 AM 09/22/2015 12:00:00 AM  DOHMH
##                               Agency.Name Complaint.Type   Descriptor
## 1 Department of Health and Mental Hygiene         Rodent Rat Sighting
## 2 Department of Health and Mental Hygiene         Rodent Rat Sighting
## 3 Department of Health and Mental Hygiene         Rodent Rat Sighting
## 4 Department of Health and Mental Hygiene         Rodent Rat Sighting
## 5 Department of Health and Mental Hygiene         Rodent Rat Sighting
## 6 Department of Health and Mental Hygiene         Rodent Rat Sighting
##                  Location.Type Incident.Zip     Incident.Address
## 1 3+ Family Mixed Use Building        10006
## 2          Commercial Building        10306 2270 HYLAN BOULEVARD
## 3          1-2 Family Dwelling        10310      758 POST AVENUE
## 4      3+ Family Apt. Building        11206   198 SCHOLES STREET
## 5 3+ Family Mixed Use Building        10462  2138 WALLACE AVENUE
## 6      3+ Family Apt. Building        11231    179 LUQUER STREET
##       Street.Name  Cross.Street.1   Cross.Street.2 Intersection.Street.1
## 1                                                          TRINITY PLACE
## 2 HYLAN BOULEVARD
## 3     POST AVENUE     CARY AVENUE GREENLEAF AVENUE
## 4  SCHOLES STREET HUMBOLDT STREET  BUSHWICK AVENUE
## 5  WALLACE AVENUE    BRADY AVENUE     LYDIG AVENUE
## 6   LUQUER STREET  CLINTON STREET     COURT STREET
##   Intersection.Street.2 Address.Type          City Landmark Facility.Type
## 1         RECTOR STREET INTERSECTION      NEW YORK                   N/A
## 2                       LATLONG STATEN ISLAND                    N/A
## 3                       ADDRESS STATEN ISLAND                    N/A
## 4                       ADDRESS      BROOKLYN                    N/A
## 5                       ADDRESS         BRONX                    N/A
## 6                       ADDRESS      BROOKLYN                    N/A
##     Status               Due.Date Resolution.Action.Updated.Date
## 1   Closed 10/04/2015 03:01:02 PM         09/18/2015 12:00:00 AM
## 2   Closed 10/04/2015 10:02:58 AM         10/28/2015 12:00:00 AM
## 3 Assigned 10/04/2015 11:58:02 AM         09/04/2015 12:03:49 PM
## 4   Closed 10/04/2015 06:56:34 AM         09/14/2015 12:00:00 AM
## 5   Closed 10/04/2015 10:08:27 AM         09/22/2015 12:00:00 AM
## 6   Closed 10/04/2015 07:59:34 PM         09/22/2015 12:00:00 AM
##              Community.Board       Borough X.Coordinate..State.Plane.
## 1               01 MANHATTAN     MANHATTAN                     980656
## 2 Unspecified STATEN ISLAND STATEN ISLAND                     955207
## 3          01 STATEN ISLAND STATEN ISLAND                     949033
## 4               01 BROOKLYN      BROOKLYN                    1000550
## 5                  11 BRONX         BRONX                    1021648
## 6               06 BROOKLYN      BROOKLYN                     984607
##   Y.Coordinate..State.Plane. Park.Facility.Name  Park.Borough School.Name
## 1                     197137        Unspecified     MANHATTAN Unspecified
## 2                     148858        Unspecified STATEN ISLAND Unspecified
## 3                     169278        Unspecified STATEN ISLAND Unspecified
## 4                     197585        Unspecified      BROOKLYN Unspecified
## 5                     250489        Unspecified         BRONX Unspecified
## 6                     186007        Unspecified      BROOKLYN Unspecified
##   School.Number School.Region School.Code School.Phone.Number School.Address
## 1   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
```

```
## 2   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
## 3   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
## 4   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
## 5   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
## 6   Unspecified   Unspecified Unspecified         Unspecified    Unspecified
##   School.City School.State  School.Zip School.Not.Found
## 1 Unspecified  Unspecified Unspecified                N
## 2 Unspecified  Unspecified Unspecified                N
## 3 Unspecified  Unspecified Unspecified                N
## 4 Unspecified  Unspecified Unspecified                N
## 5 Unspecified  Unspecified Unspecified                N
## 6 Unspecified  Unspecified Unspecified                N
##   School.or.Citywide.Complaint Vehicle.Type Taxi.Company.Borough
## 1                           NA           NA                   NA
## 2                           NA           NA                   NA
## 3                           NA           NA                   NA
## 4                           NA           NA                   NA
## 5                           NA           NA                   NA
## 6                           NA           NA                   NA
##   Taxi.Pick.Up.Location Bridge.Highway.Name Bridge.Highway.Direction Road.Ramp
## 1                    NA                  NA                        NA        NA
## 2                    NA                  NA                        NA        NA
## 3                    NA                  NA                        NA        NA
## 4                    NA                  NA                        NA        NA
## 5                    NA                  NA                        NA        NA
## 6                    NA                  NA                        NA        NA
##   Bridge.Highway.Segment Garage.Lot.Name Ferry.Direction Ferry.Terminal.Name
## 1                     NA              NA              NA                  NA
## 2                     NA              NA              NA                  NA
## 3                     NA              NA              NA                  NA
## 4                     NA              NA              NA                  NA
## 5                     NA              NA              NA                  NA
## 6                     NA              NA              NA                  NA
##   Latitude Longitude                                Location
## 1 40.70777 -74.01296 (40.70777155363643, -74.01296309970473)
## 2 40.57521 -74.10455 (40.575209242947444, -74.1045465185469)
## 3 40.63124 -74.12688 (40.63123555151668, -74.12687759748677)
## 4 40.70899 -73.94121 (40.70898692345805, -73.94120690238431)
## 5 40.85413 -73.86481 (40.85413014360452, -73.86481331044513)
## 6 40.67722 -73.99871 (40.67722299833378, -73.99871293635606)
```

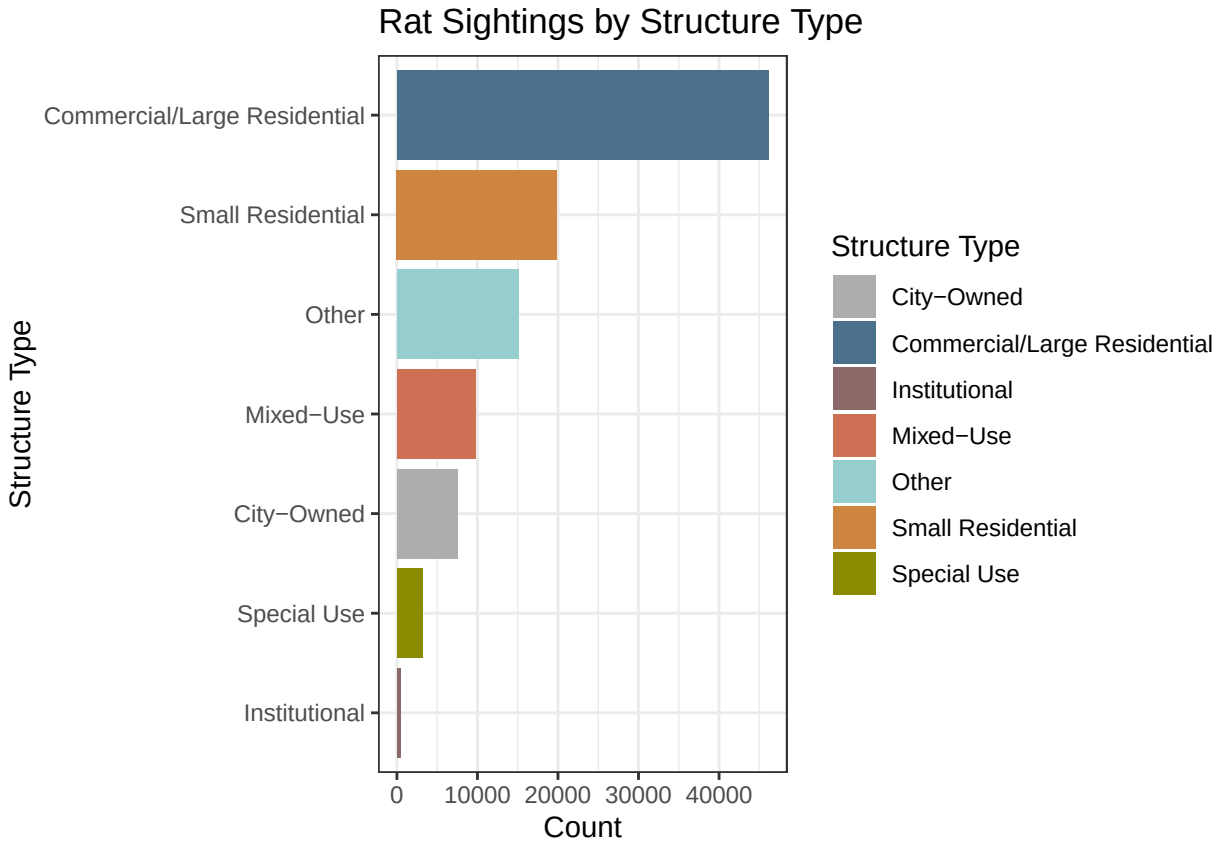
```r
rats_data <- rats %>%
  select(Unique.Key, Created.Date, Location.Type, Incident.Zip, Borough, Resolution.Action.Updated.Date
  filter(Borough != "Unspecified") %>%
  mutate(structure_type = case_when (Location.Type %in% c( "1-2 Family Dwelling", "Single Room Occupancy
                                     Location.Type %in% c("Commercial Building", "Office Building"
                                     Location.Type %in% c("3+ Family Mixed Use Building", "1-2 Fami
                                     Location.Type %in% c("Hospital", "School/Pre-School", "Day Ca
                                     Location.Type %in% c("Government Building", "Public Stairs",
                                     Location.Type %in% c("Parking Lot/Garage", "Construction Site
                                     Location.Type %in% c( "Other (Explain Below)") ~ "Other")) %>%
  distinct() %>%
  mutate(Created.Date = as.Date(Created.Date, format = "%m/%d/%Y"),
         Resolution.Action.Updated.Date = as.Date(Resolution.Action.Updated.Date, format = "%m/%d/%Y"),
```

```
        days_spent = as.numeric(abs(difftime(Resolution.Action.Updated.Date, Created.Date, units = "da
        month = lubridate::month(Created.Date),
        season = case_when(month %in% c(12,1,2) ~ "Winter",
                           month %in% c(3,4,5) ~ "Spring",
                           month %in% c(6,7,8) ~ "Summer",
                           month %in% c(9,10,11) ~ "Fall"),
        year = lubridate::year(Created.Date),
        month = ordered(month),
        year = ordered(year))
```

```
head(rats_data)
```

```
##   Unique.Key Created.Date               Location.Type Incident.Zip
## 1   31464015   2015-09-04 3+ Family Mixed Use Building        10006
## 2   31464024   2015-09-04          Commercial Building        10306
## 3   31464025   2015-09-04          1-2 Family Dwelling        10310
## 4   31464026   2015-09-04      3+ Family Apt. Building        11206
## 5   31464027   2015-09-04 3+ Family Mixed Use Building        10462
## 6   31464188   2015-09-04      3+ Family Apt. Building        11231
##         Borough Resolution.Action.Updated.Date Latitude Longitude
## 1     MANHATTAN                     2015-09-18 40.70777 -74.01296
## 2 STATEN ISLAND                     2015-10-28 40.57521 -74.10455
## 3 STATEN ISLAND                     2015-09-04 40.63124 -74.12688
## 4      BROOKLYN                     2015-09-14 40.70899 -73.94121
## 5         BRONX                     2015-09-22 40.85413 -73.86481
## 6      BROOKLYN                     2015-09-22 40.67722 -73.99871
##                                 Location               structure_type
## 1 (40.70777155363643, -74.01296309970473)                    Mixed-Use
## 2 (40.575209242947444, -74.1045465185469) Commercial/Large Residential
## 3 (40.63123555151668, -74.12687759748677)            Small Residential
## 4 (40.70898692345805, -73.94120690238431) Commercial/Large Residential
## 5 (40.85413014360452, -73.86481331044513)                    Mixed-Use
## 6 (40.67722299833378, -73.99871293635606) Commercial/Large Residential
##   days_spent month season year
## 1         14     9   Fall 2015
## 2         54     9   Fall 2015
## 3          0     9   Fall 2015
## 4         10     9   Fall 2015
## 5         18     9   Fall 2015
## 6         18     9   Fall 2015
```

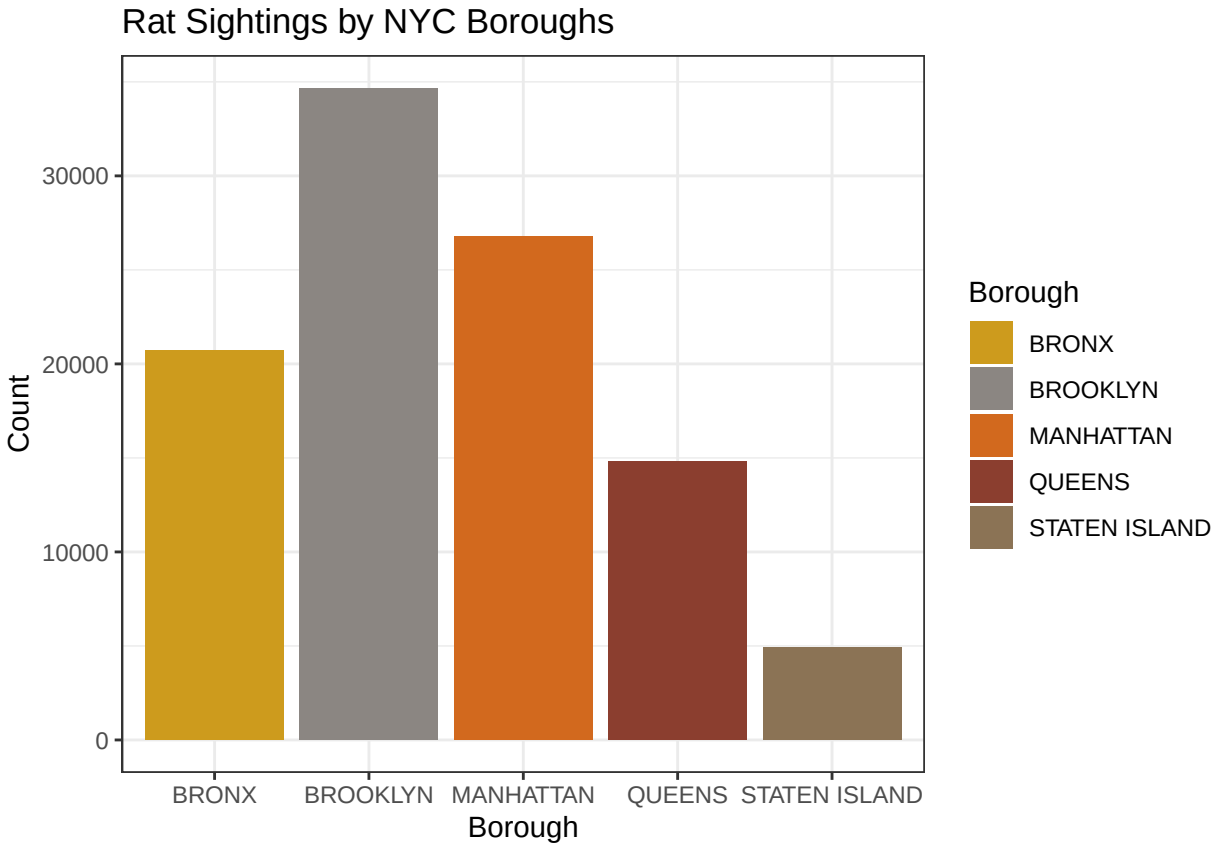
```
# Structure type visualization
rats_data %>%
  drop_na(structure_type) %>%
  count(structure_type) %>%
  ggplot(aes(x = reorder(structure_type, n), y = n, fill = structure_type)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("gray68", "skyblue4", "rosybrown4", "salmon3", "paleturquoise3", "tan3",
  coord_flip() +
  labs(title = "Rat Sightings by Structure Type",
       x = "Structure Type",
       y = "Count",
       fill = "Structure Type") +
  theme_bw()
```



Between the years 2010 and 2017, Commercial/Large Residential buildings have reported the most rat sightings in New York City with over 45,000 complaints. Commercial/Large Residential buildings include the following: Commercial buildings, office buildings, and 3+ family apartment buildings. Institutional buildings such as hospitals, schools, and day care/nurseries have the least reported rat sightings in the city.

```
data <- as.data.frame(table(rats_data$Borough, rats_data$Incident.Zip)) %>% filter(Freq!=0) %>%
  rename(Zipcode = Var2, Borough = Var1) %>%
  distinct()
```

```
# Borough Visualization
rats_data %>%
  ggplot(aes(x = Borough, fill = Borough)) +
  geom_bar() +
  scale_fill_manual(values = c("goldenrod3", "seashell4", "chocolate", "coral4", "burlywood4")) +
  labs(title = "Rat Sightings by NYC Boroughs",
       y = "Count") +
  theme_bw()
```



With nearly 35,000 rat sightings, rodents are most likely to be found in Brooklyn. With Manhattan following with over 25,000 rodent complaints. The analysis showed that Staten Island had the least number of rat sightings out of the five boroughs. This is likely due to its population density compared to Brooklyn and Manhattan.

```
# Top 5 and Bottom 5 Borough  & Zip Code Pairs
rats_descriptive <- rats_data %>%
  group_by(Borough, Incident.Zip) %>%
  summarise(Freq = n(),
            mean_days = round(mean(days_spent, na.rm = T), 0),
            sd_days = round(sd(days_spent, na.rm = T), 0),
            median_days = round(median(days_spent, na.rm = T), 0)) %>%
  ungroup()
```

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

```
inaccurate_pairs <- c("BROOKLYN",   "11416",
           "MANHATTAN", "10000",
           "QUEENS",    "11208",
           "QUEENS",    "11237")

# Filtering out inaccurate pairs from rats_descriptive
cleaned_rats_descriptive <- rats_descriptive %>%
  filter(!(Borough %in% inaccurate_pairs & Incident.Zip %in% inaccurate_pairs))

top5_descriptive <- cleaned_rats_descriptive %>%
  drop_na() %>%
```

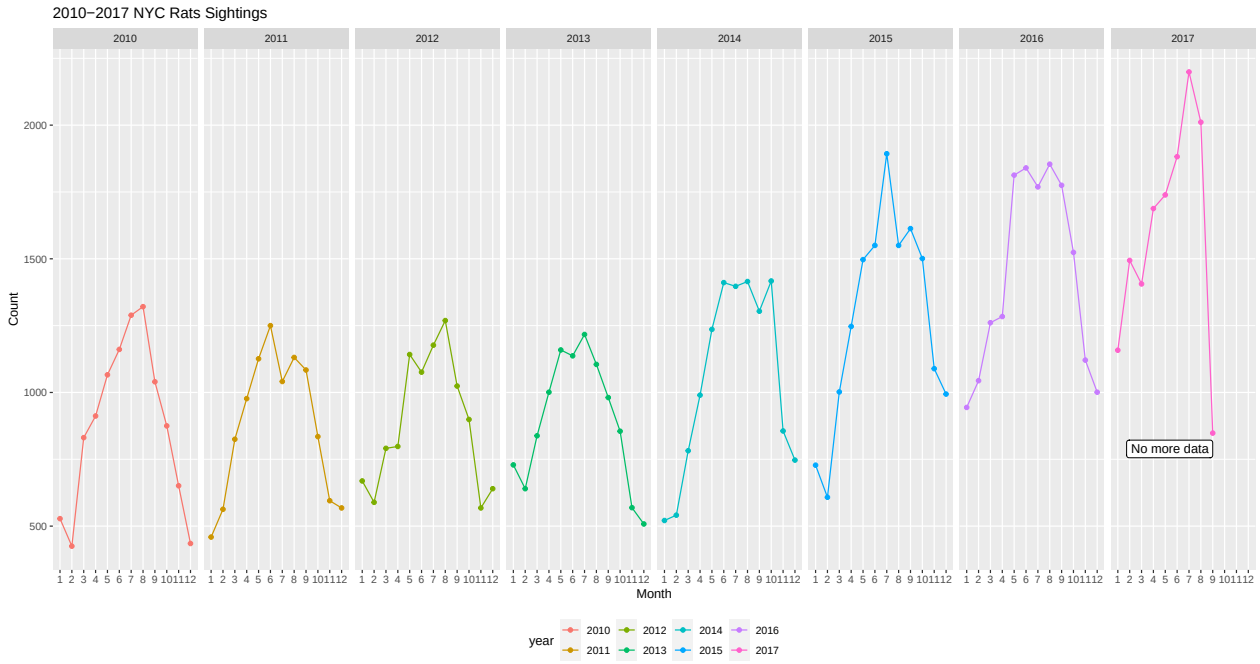
```
  arrange(desc(Freq)) %>%
  head(n = 5) %>%
  mutate(position = "Top 5")

bottom5_descriptive <-cleaned_rats_descriptive %>%
  drop_na() %>%
  arrange(Freq) %>%
  head(n = 5) %>%
  mutate(position = "Bottom 5")

top_bottom <- rbind(top5_descriptive, bottom5_descriptive) %>%
  rename(`N cases` = Freq)
top_bottom
```

```
## # A tibble: 10 x 7
##    Borough   Incident.Zip `N cases` mean_days sd_days median_days position
##    <chr>     <chr>            <int>     <dbl>   <dbl>       <dbl> <chr>
##  1 BROOKLYN  11221             3124        18      27          10 Top 5
##  2 BROOKLYN  11216             2494        21     161           9 Top 5
##  3 MANHATTAN 10025             2285        16      29           6 Top 5
##  4 BROOKLYN  11238             2158        18      22          11 Top 5
##  5 BROOKLYN  11233             1925        14      19           8 Top 5
##  6 MANHATTAN 10020                2        46      43          46 Bottom 5
##  7 BROOKLYN  11239                7        85     128          45 Bottom 5
##  8 QUEENS    11040                8        63     146           6 Bottom 5
##  9 MANHATTAN 10044               10         4       5           2 Bottom 5
## 10 MANHATTAN 10162               10        14       7          16 Bottom 5
```

Four neighborhoods in Brooklyn have experienced the highest number of rat infestations in New York City with nearly 9,700 cases. Additionally, one neighborhood in Manhattan has also reported a significant number of cases. It typically takes around twenty days to resolve a single case within these neighborhoods.

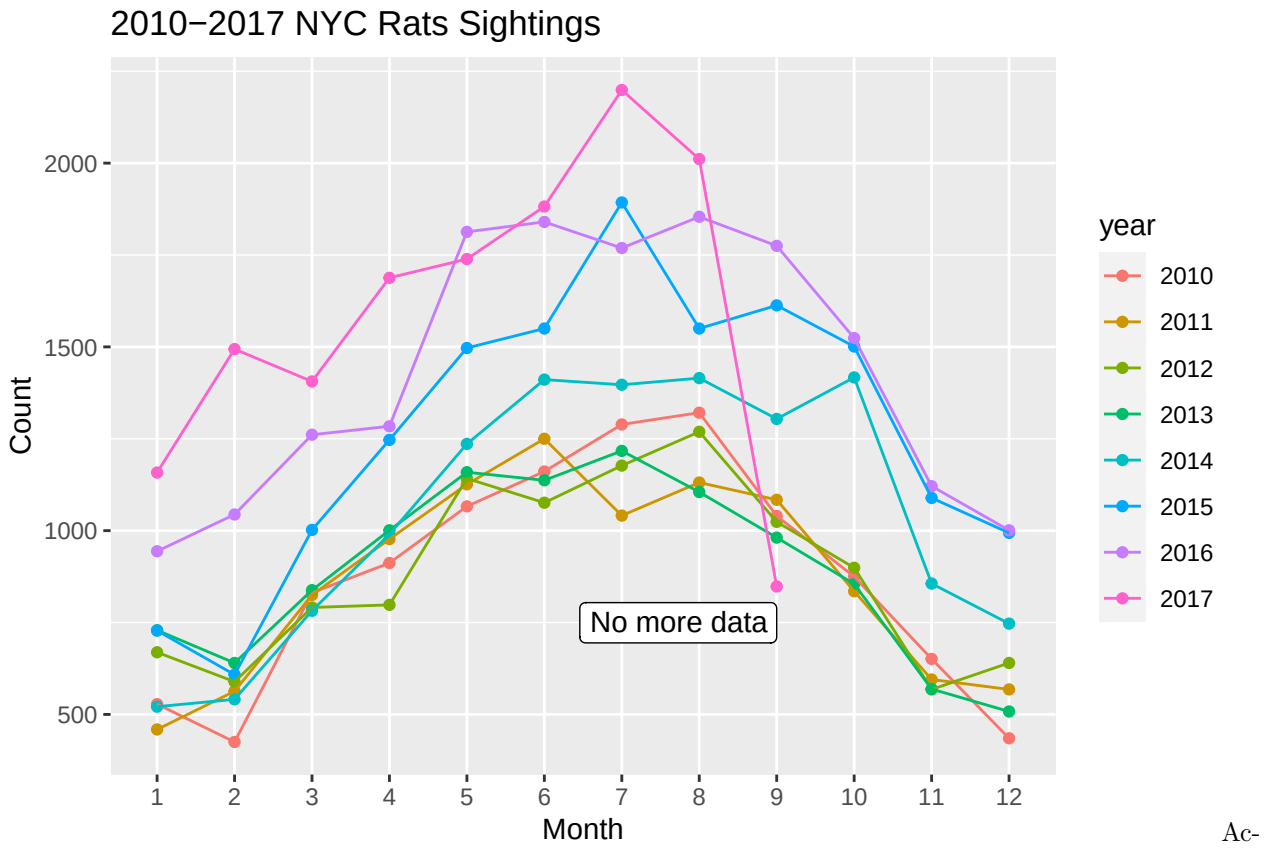
On the other hand, neighborhoods with significantly fewer rodent cases experience a broader range in resolution times, spanning from four to eighty-five days per case.

```
# Rat Sightings Over Time
rats_line <- as.data.frame(table(rats_data$year, rats_data$month)) %>%
  rename(year = Var1, month = Var2, count = Freq) %>%
  filter(count != 0) %>%
  arrange(year, month)

label_data <- rats_line %>%
  filter(year == "2017" & month == 9)

rats_line %>%
  ggplot(aes(x = month, y = count, group = year)) +
  geom_point(aes(color = year)) +
  geom_line(aes(color = year)) +
    geom_label(data = label_data, aes(x = month, y = count, label = "No more data"), hjust = 1.0, vjust
  facet_grid(cols = vars(year)) +
  theme(legend.position = "bottom") +
  labs(title = "2010-2017 NYC Rats Sightings",
       x = "Month",
       y = "Count")
```



From 2010 to 2017, there has been a noticeable increase in rat sightings in New York City. According to the line charts, the number of cases peak during the middle of each year – around summer time. It is important to note that during the year 2017, data was not reported after September 16th by the New York City Health Department.

```
# Rat Sightings Over Time
label_data <- rats_line %>%
  filter(year == "2017" & month == 9)

rats_line %>%
  ggplot(aes(x = month, y = count, group = year)) +
  geom_point(aes(color = year)) +
  geom_line(aes(color = year)) +
  geom_label(data = label_data, aes(x = month, y = count, label = "No more data"), hjust = 1.0, vjust =
  labs(title = "2010-2017 NYC Rats Sightings",
       x = "Month",
       y = "Count")
```



According to the line chart, the number of reported cases start to increase during the month of February and decrease after August of each year. Additionally, rodents are most likely to occupy areas of New York City during warmer seasons.

```
# One-Way ANOVA
rats_season <- aov(days_spent ~ season, data = rats_data)
summary(rats_season)
```

```
##                 Df    Sum Sq Mean Sq F value  Pr(>F)
## season           3     92057   30686    22.4 1.7e-14 ***
## Residuals   101906 139573613    1370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3 observations deleted due to missingness
```

The results of the one-way ANOVA indicates a significant difference in the number of days spent per case for different seasons.

```
group_by(rats_data, season) %>%
  summarise(
    count = n(),
    mean = mean(days_spent, na.rm = TRUE),
    sd = sd(days_spent, na.rm = TRUE)
  )
```

```
## # A tibble: 4 x 4
##   season count  mean    sd
##   <chr>  <int> <dbl> <dbl>
## 1 Fall   23024  17.3  34.9
## 2 Spring 27411  15.8  53.2
```

```
## 3 Summer 34945  15.5  26.8
## 4 Winter 16533  14.3  23.0
```

```
# Post-HOC Test
TukeyHSD(rats_season)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = days_spent ~ season, data = rats_data)
##
## $season
##                       diff       lwr        upr     p adj
## Spring-Fall   -1.4547501 -2.304703 -0.6047976 0.0000648
## Summer-Fall   -1.7852004 -2.592250 -0.9781510 0.0000001
## Winter-Fall   -3.0034027 -3.972628 -2.0341778 0.0000000
## Summer-Spring -0.3304502 -1.097561  0.4366607 0.6854218
## Winter-Spring -1.5486526 -2.484883 -0.6124221 0.0001259
## Winter-Summer -1.2182023 -2.115664 -0.3207410 0.0027482
```

The Tukey test results shows significant differences in the number of days spent per case for all seasons except the Summer-Spring pair.

Therefore, we can conclude that there are significant differences in the number of days spent per case for different seasons, $F(3, 101906) = 22.4$, $p < 0.05$. Specifically, the number of days per case spent is different between Spring ($M = 15.82$, $SD = 53.25$) and Fall ($M = 17.28$, $SD = 34.88$), Summer ($M = 15.49$, $SD = 26.82$) and Fall, Winter ($M = 14.27$, $SD = 22.97$) and Fall, Winter and Spring, and Winter and Summer. No other significant differences were found.

Conclusion

In conclusion, Brooklyn has the highest concentration of rat sightings within New York City between 2010 and 2017, with four of its neighborhoods bearing a significant burden of infestations. With more than 45,000 complaints filed, these rodents are prone to frequenting both commercial establishments and large residential buildings. With each passing year, the number of reported sightings increases, particularly during warmer weather.

Recommendations

Based on our findings, city authorities can take preventive measures to mitigate rat infestations within New York City. 1. City authorities can allocate resources and efforts towards implementing targeted rat control measures in the identified neighborhoods within Brooklyn where the burden of infestations is highest. This could involve intensified pest control efforts, increased sanitation measures, and community education programs to raise awareness about rat prevention and management. 2. Additionally, city authorities can prioritize pest control efforts during the warmer months to effectively manage and reduce rat populations. Implementing proactive measures such as regular inspections, baiting programs, and sealing potential entry points to prevent rat infestations from escalating during peak seasons may contribute to long-term sustainability and resilience within affected communities. 3. Launch public awareness campaigns to educate residents and businesses about the importance of proper waste management, sanitation practices, and early detection of rat infestations may also significantly reduce reported rat sightings in New York City.