Helen Cain
PPUA 5262 - Midterm

Boston 911 - Read-Me Document

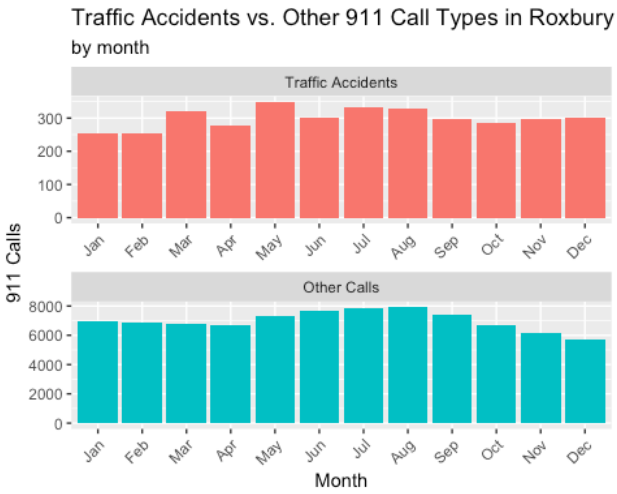
Overview

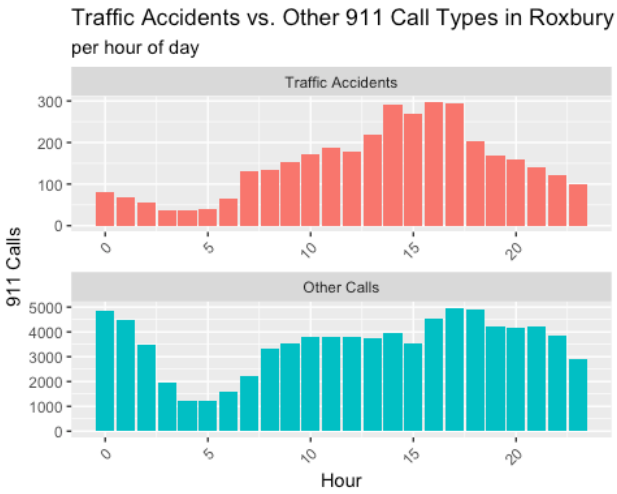
The city of Boston implements a CAD system, or computer-assisted dispatch system, to maintain records of all calls made to 911. The Boston Area Research Initiative, or BARI, has obtained the raw data from Boston's 911 system from the year 2024. There are 625,928 records in this data set, each cataloged with 24 variables that describe the nature of each call, and where and when those calls occurred. Data has been anonymized to protect the identity of callers. This Read-Me document provides an overview of the dataset, specifically in regards to calls related to motor vehicle accidents, hit-and-runs, and accidents involving pedestrians. The data also focuses on Roxbury, and the main thoroughfares through the neighborhood such as Blue Hill Ave, Warren St, and the residential areas surrounding them.

Fun Facts

1. 87,790 of all calls in 2024, or about 14% of the total, came from the Census Block Groups surrounding Blue Hill Avenue.
2. 14.5% of calls in the area around Blue Hill Avenue were related to traffic accidents, only very slightly higher than the rest of Boston, where 14% of all calls were related to traffic accidents.
3. The month of May had the highest volume of 911 calls related to traffic accidents in the area around Blue Hill Ave, at 349 calls. February had the fewest, at 253.
4. The highest volume of calls regarding traffic accidents in Roxbury are made between 2:00 and 5:00 pm. Other calls in Roxbury were most frequent at midnight, and the hours between 4 and 6pm.
5. Census block group 250250819002 had the fewest motor vehicle accidents in 2024, at 8 total. This block group is a very small residential area between Trotter and Ellis elementary schools, with few streets.
6. Census block group 250250801001 had the highest number of motor vehicle accidents in 2024, at 487 total. This block group encompasses a large commercial and industrial area in Lower Roxbury, bordered by the I-93, Melnea Cass Boulevard, and the Commuter Rail's Franklin Line.
7. There was an average of 64.30 accidents per block group in 2024. The median number of accidents per block group was 45.
8. Out of all traffic accidents in the block groups surrounding Blue Hill avenue, about 22% were classified as hit and runs. This is similar to the proportion for the rest of Boston: 23%.
9. 8.6% of all traffic accidents in the Blue Hill Ave area involved a pedestrian, a slightly higher percentage than elsewhere in Boston, with a rate of 6.8%.

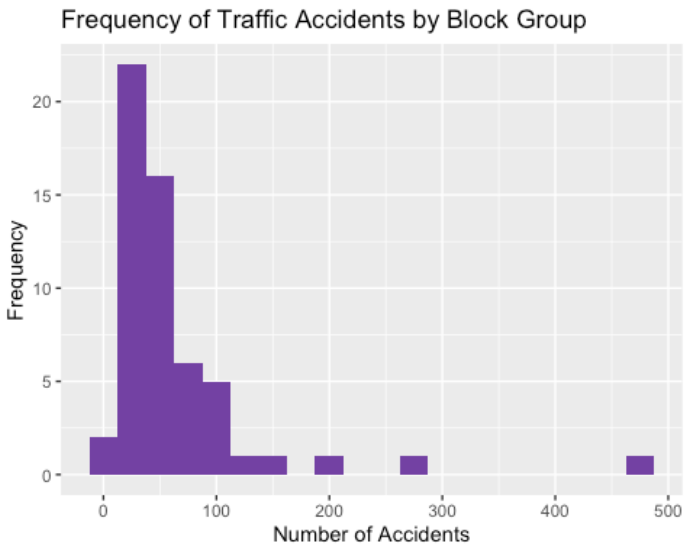
Helen Cain
PPUA 5262 - Midterm





The two graphs above illustrate changes in the volume of calls from month to month and throughout the day, as well as differences in the patterns between calls about traffic accidents and other call types. During a workshop with the Dudley Street Neighborhood initiative, community members noted an uptick in recent years of local traffic through residential areas during rush hour. As can be seen in the chart on the right, traffic accidents are reported most frequently between 3 and 5 pm, or during rush hour.

The graph below shows the distribution of traffic accidents throughout the various census block groups in Roxbury. The bar farthest to the right represents a block group that includes a major highway (see fun fact 6.) Most block groups reported within the range of 25-75 accidents in the year 2024.



The city of Boston has made street infrastructure changes in recent years in an effort to improve pedestrian safety. The Slow Streets and Safety Surge Initiatives, in particular, saw the addition of speed humps and intersection improvements in many neighborhoods, including Roxbury. Community members expressed a degree of frustration at how some of those changes affect drivers, however. Namely, bike lanes make turns at certain intersections more difficult, and speed humps on residential streets can be annoying to drive over, illustrating that benefits for pedestrians must be balanced with the needs of drivers as well.

Helen Cain
PPUA 5262 - Midterm

Appendix A: Data Dictionary

- eid: the unique identifier for each 911 call
- date: the date and time the call was recieved
- month: newly created variable, the month in which the call was recieved
- hour: newly created variable, the hour of day in which the call was recieved (24-hour clock)
- typ_eng: brief description of the type of call
- sub_eng: brief description of the call's subtype
- BG_ID_20: the 2020 census block group from which the call originated
- location: newly created variable, designates which of the calls originated from the block groups within Roxbury, or elsewhere in Boston
- traffic_accident: newly created variable, designates whether or not the 911 call was related to a traffic accident
- hit_and_run: newly created variable, designates whether or not the 911 call was related to a hit and run report
- pedestrian: newly created variable, designates whether or not the call was related to a traffic accident involving a pedestrian

Helen Cain
PPUA 5262 - Midterm

Appendix B: Annotated R Syntax

Load dataset:

```
bos_911 <- read.csv('911_Calls_From_Jan-Dec_2024.csv',
na.strings='NULL')
```

Creation of new variables:

```
# traffic_accident
bos_911$traffic_accident <- ifelse(bos_911$typ_eng == "MOTOR VEHICLE
ACCIDENT", 1, 0)

# pedestrian
bos_911$pedestrian <- ifelse(str_detect(bos_911$sub_eng, 'PEDESTRIAN
STRUCK'), 1, 0)

# hit_and_run
bos_911$hit_and_run <- ifelse(str_detect(bos_911$sub_eng, 'HIT AND
RUN'), 1, 0)

# location
source('blue_hill_region.R')

bos_911$location <- ifelse(bos_911$BG_ID_20 %in% blue_hill_cbgs,
"Roxbury", "Boston")

# month and hour
bos_911$date <- ymd_hms(bos_911$date, tz = "US/Eastern")
bos_911$month <- month(bos_911$date, label=TRUE, abbr=TRUE)
bos_911$hour <- hour(bos_911$date)
```

Visualizations

Bar chart by month:

```
bos_911 |> filter(location == "Roxbury" & !is.na(month)) |>
  mutate(traffic_accident = factor(traffic_accident, levels = c(1,
0))) |>
  ggplot(aes(x = month, fill = traffic_accident)) +
  geom_bar() +
  facet_wrap(~traffic_accident, ncol=1, scales="free",
             labeller = labeller(traffic_accident = c("0" = "Other
Calls",
                                                      "1" = "Traffic
Accidents"))) +
  labs(title = "Traffic Accidents vs. Other 911 Call Types in
Roxbury", subtitle = "by month",
       x="Month", y = "911 Calls") +
```

```
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1),
        legend.position = "none")
```

Bar chart by hour:

```
bos_911 |> filter(location == "Roxbury" & !is.na(hour)) |>
  mutate(traffic_accident = factor(traffic_accident, levels = c(1,
0))) |>
  ggplot(aes(x = hour, fill = traffic_accident)) +
  geom_bar() +
  facet_wrap(~traffic_accident, ncol=1, scales="free",
             labeller = labeller(traffic_accident = c("0" = "Other
Calls",
                                                      "1" = "Traffic
Accidents"))) +
  labs(title = "Traffic Accidents vs. Other 911 Call Types in
Roxbury", subtitle = "per hour of day",
       x="Hour", y = "911 Calls") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1),
        legend.position = "none")
```

Histogram:

```
bg_crash <- bos_911 |> filter(location == "Roxbury") |>
group_by(BG_ID_20) |> summarize(count = sum(traffic_accident == 1)) |>
arrange(by = count)
bg_crash |> ggplot(aes(x = count)) +
  geom_histogram(binwidth = 25, fill="#7340A3") +
  labs(title = "Frequency of Traffic Accidents by Block Group",
       x = "Number of Accidents",
       y = "Frequency")
```

Fun facts:

```
#1
bos_911 |> filter(location == "Roxbury") |> dim()

#2
bos_911 |> select(c(location, traffic_accident)) |>
  table() |> prop.table(margin = 2)

#3
bos_911 |> filter(location=="Roxbury" & traffic_accident == 1) |>
  select(month) |> table()

#4
bos_911 |> filter(location == "Roxbury") |>
  select(traffic_accident, hour) |> table()
```

5. and 6.

```
bg_crash <- bos_911 |> filter(location == "Roxbury") |>
  group_by(BG_ID_20) |> summarize(count = sum(traffic_accident == 1))
|>
  arrange(by = count)

#7
summary(bg_crash$count)

#8
bos_911 |> filter(traffic_accident == 1) |>
  select(c(hit_and_run, location)) |>
  mutate(hit_and_run = recode(hit_and_run, '0' = "Other", '1' = "Hit
and Run")) |>
  table() |> prop.table(margin = 2) |>
  kable()

#9
bos_911 |> filter(traffic_accident == 1) |>
  select(c(pedestrian, location)) |>
  mutate(pedestrian = recode(pedestrian, '0' = "Other", '1' =
"Pedestrian Struck")) |>
  table() |> prop.table(margin = 2) |>
  kable()
```