False Reports Analysis

```r
# I. Data Cleaning
# Load Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.2.1     v dplyr   1.1.3
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(knitr)
library(here)
```

```
## Warning: package 'here' was built under R version 4.2.2
```

```
## here() starts at D:/Datos/Monica/Documents/R_Studio/migrawatch
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(stringr)
library(DT)
```

```
## Warning: package 'DT' was built under R version 4.2.3
```

```r
library(sf)
```

```
## Warning: package 'sf' was built under R version 4.2.3

## Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```r
# Load Relevant Data
data_raw <- read_csv(file = file.path(here(), '9.29.25_data.csv'),
                        # add appropriate column names
                        col_names = c('location', 'date_time_reported',
                                      'unique_id', 'date_time_raid',
                                      'day', 'time_category',
                                      'raid_calls', 'notes',
                                      'street_address', 'type_report',
                                      'detention_centers', 'tactics_reported',
                                      'rapid_response_team', 'people_detained',
                                      'verified_by_rrt', 'OCAD_operator_uploaded',
                                      'business_worksite', 'car_description', 'license_plate'))
```

```
## Warning: One or more parsing issues, see `problems()` for details

## Rows: 1290 Columns: 26
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (23): location, date_time_reported, unique_id, date_time_raid, day, time...
## dbl  (1): people_detained
## lgl  (2): raid_calls, detention_centers
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Select all rows, delete columns 20 and 21 (media links)
data_raw <-  data_raw[, c(1:19,23)]
data_raw <-  data_raw %>%
            rename("source" = X23)

# Format the date and time correctly + make certain columns strings
data_raw$date_time_reported <-  mdy_hm(data_raw$date_time_reported)
data_raw$date_time_raid <-  mdy_hm(data_raw$date_time_raid)

# Change capitalization and abbreviate to avoid problems downstream
data_raw <- data_raw %>%
  mutate(type_report = str_to_title(type_report)) %>%
  mutate(type_report = str_replace_all(type_report, "Corporate Collaboration", "Corp Collab"),
         type_report = str_replace_all(type_report, "Public Space Raid", "Pub Space Raid"),
         location = str_replace_all(location, "Chicago - |Chicago- ", ""),
         location = case_when(location == "" ~ "Chicago", TRUE ~ location))


# II. Data Transformation
```

```r
# Select all false sightings
data_labeled <- data_raw %>%
  mutate(report_status = case_when(
    # This is the step where the consultant was mixing variables.
    # There is some overlap: some cases proven false were originally tagged as Rumors.
    str_detect(type_report, "False") ~ "Confirmed False",
    # 'False' takes precedence over 'Rumors'.
    str_detect(type_report, "Rumors") ~ "Unconfirmed",
    # Default: treat everything else as Confirmed True.
    TRUE ~ "Confirmed True"))


# Clean up the labels in the refine_categories to get types of Raids
data_labeled <- data_labeled %>%
  mutate(refined_categories = str_replace_all(type_report, "False|Rumors", ""),
         refined_categories = str_replace_all(refined_categories, ",,", ","),
         refined_categories = str_replace_all(refined_categories, "^,|,$", ""),
          # Public Space Raid
         refined_categories = str_replace_all(refined_categories, "Ice Sighting,Pub Space Raid|Pub Spac
                                              "Pub Space Raid"),
          # Worksite Raid
         refined_categories = str_replace_all(refined_categories, "Ice Sighting,Worksite Raid|Worksite R
                                              "Worksite Raid"),
          # I-9 Audit
         refined_categories = str_replace_all(refined_categories, "Ice Sighting,I-9 Audit", "I-9 Audit")
          # Corporate Collaboration
         refined_categories = str_replace_all(refined_categories, "Ice Sighting,Corp Collab", "Corp Col
        # Home raid
        refined_categories = str_replace_all(refined_categories, "Ice Sighting,Home Raid|Home Raid,Ice S
        # Replace blank spaces with Replace as placeholder
        refined_categories = case_when(is.na(refined_categories)|refined_categories == "" ~ "No tags", T

# III. Manual Edits

# Clean up certain incidents manually that need additional reclassification
data_labeled <- data_labeled %>%
  mutate(refined_categories =
           case_when(unique_id == "09-11 Melrose Park-River Forest on Chicago Ave." ~ "Unclass. Raid",
                     # They specifically raided a construction site
                     unique_id == "09-18 Chicago - NWS - Hermosa-Springfield & North Ave." ~ "Worksite R
    TRUE ~ refined_categories))

# Label specific worksites
data_labeled <- data_labeled %>%
  mutate(business_worksite =
           case_when(unique_id == "09-18 Chicago - NWS - Hermosa-Springfield & North Ave." ~ "Constructi
```

I. Differentiation from AirTable Data

The dashboard in AirTable counts each of the tags and produces a bar chart. Since this chart does considers FALSE and Rumors a classification tag, instead of its a separate variable, it unintentionally overstates the I-9 audits, incidents of corporate collaboration, worksite raids, home raids, public space raids, and ICE sightings that have occurred up to 9.29.

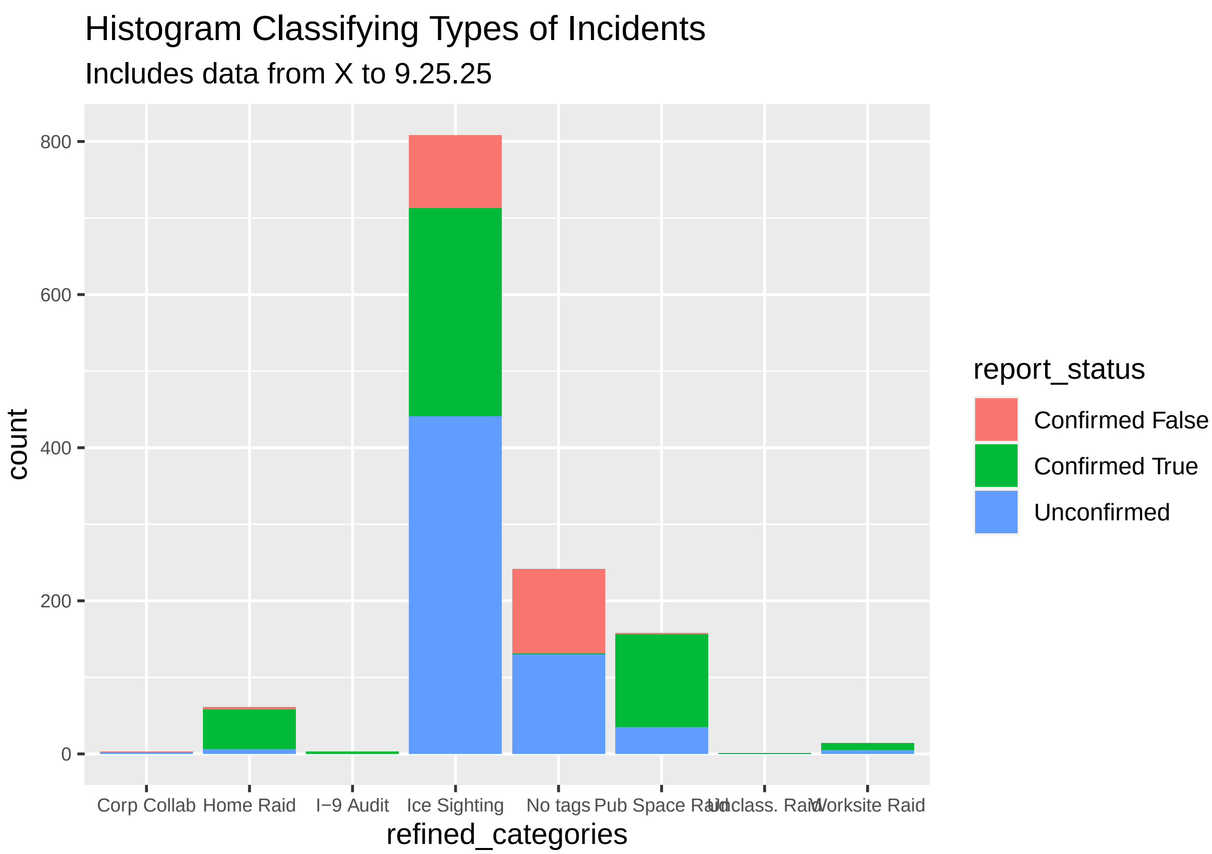
For example, the visualization includes 970 ICE sightings, but this count includes sightings that have later been proven False ("Ice Sighting, False"). If we actually break down ICE sightings by whether or not they are true, false, or unproven, we find that only 272 have been proven true, 95 are confirmed false, 441 were unable to be confirmed.

```r
# Visualization of Confirmed True, Confirmed False, and Unconfirmed
data_type <- data_labeled %>%
  group_by(refined_categories, report_status) %>%
  count() %>%
  kable(caption = "Type of Incident by Report Validity")

data_type
```

Table 1: Type of Incident by Report Validity

refined_categories	report_status	n
Corp Collab	Confirmed False	1
Corp Collab	Unconfirmed	2
Home Raid	Confirmed False	3
Home Raid	Confirmed True	52
Home Raid	Unconfirmed	6
I-9 Audit	Confirmed True	3
Ice Sighting	Confirmed False	95
Ice Sighting	Confirmed True	272
Ice Sighting	Unconfirmed	441
No tags	Confirmed False	111
No tags	Confirmed True	1
No tags	Unconfirmed	130
Pub Space Raid	Confirmed False	2
Pub Space Raid	Confirmed True	121
Pub Space Raid	Unconfirmed	35
Unclass. Raid	Confirmed True	1
Worksite Raid	Confirmed True	9
Worksite Raid	Unconfirmed	5

```r
# Bar Chart of variables
ggplot(data_labeled, aes(x = refined_categories)) +
   geom_bar(aes(fill = report_status )) +
   labs(title="Histogram Classifying Types of Incidents",
       subtitle="Includes data from X to 9.25.25") +
  theme(axis.text = element_text(size = 7))
```

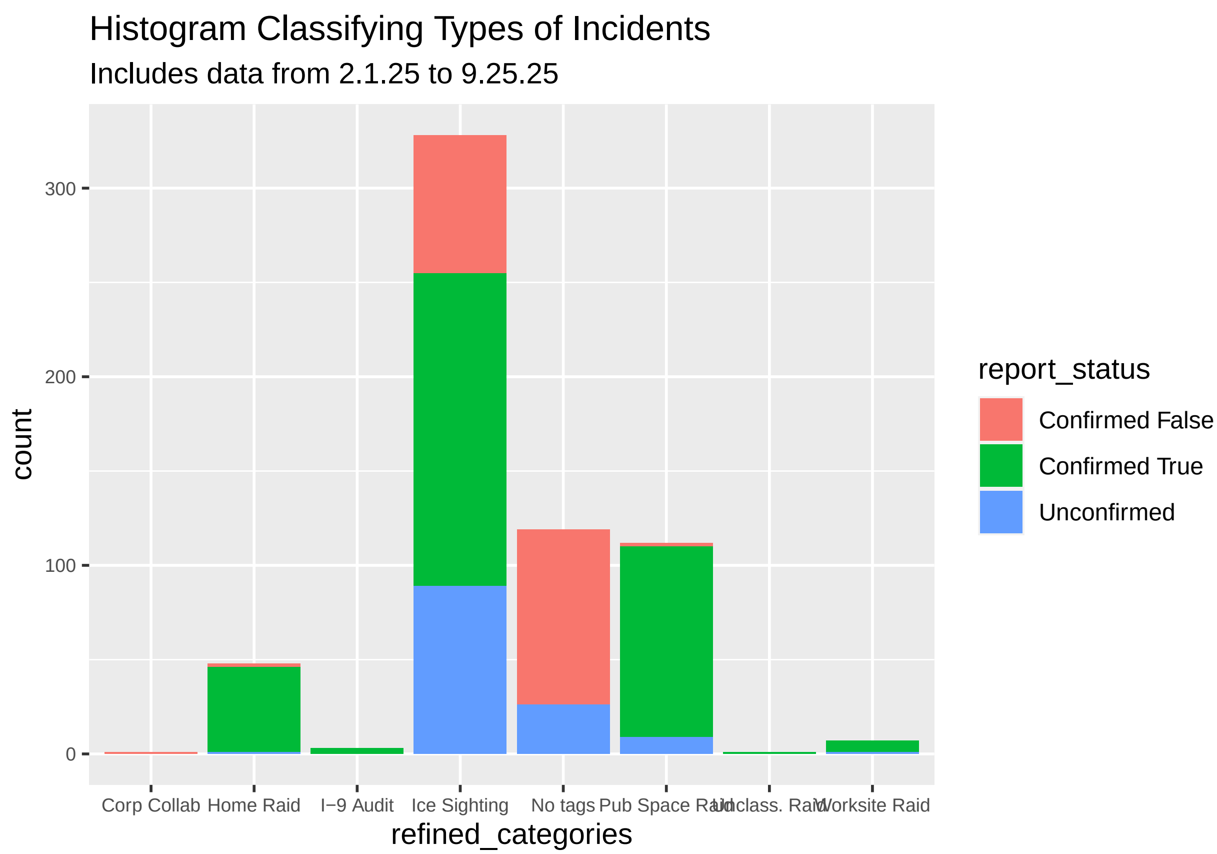


II. Analysis of Activities Verified by RRT

When we analyze incidents where a rapid responder was dispatched (619/1290 cases), we can see that the overall distribution looks similar to the broader data set.

Some notable exceptions exist. For example, false incidents comprise *12.5%* of total ICE sightings, but where a rapid responder was dispatched, incidents were proven false *22.3%* of the time. This discrepancy likely occurs because the broader data set includes referrals from the MIDAS network and relevant social media posts – driving the number of proven sightings up.

```r
# Visualization of Confirmed True, Confirmed False, and Unconfirmed
data_type_2 <- data_labeled %>%
  filter(verified_by_rrt == "checked") %>%
  group_by(refined_categories, report_status) %>%
  count()

# Visualization of activities verified by Rapid Responders
data_labeled %>%
  filter(verified_by_rrt == "checked") %>%
ggplot(aes(x = refined_categories)) +
   geom_bar(aes(fill = report_status)) +
   labs(title="Histogram Classifying Types of Incidents",
       subtitle="Includes data from 2.1.25 to 9.25.25") +
  theme(axis.text = element_text(size = 7))
```




```r
data_labeled %>%
  filter(verified_by_rrt == "checked") %>%
  group_by(refined_categories, report_status) %>%
  count()
```

```
## # A tibble: 16 x 3
## # Groups:   refined_categories, report_status [16]
##    refined_categories report_status       n
##    <chr>              <chr>           <int>
##  1 Corp Collab        Confirmed False     1
##  2 Home Raid          Confirmed False     2
##  3 Home Raid          Confirmed True     45
##  4 Home Raid          Unconfirmed         1
##  5 I-9 Audit          Confirmed True      3
##  6 Ice Sighting       Confirmed False    73
##  7 Ice Sighting       Confirmed True    166
##  8 Ice Sighting       Unconfirmed        89
##  9 No tags            Confirmed False    93
## 10 No tags            Unconfirmed        26
## 11 Pub Space Raid     Confirmed False     2
## 12 Pub Space Raid     Confirmed True    101
## 13 Pub Space Raid     Unconfirmed         9
## 14 Unclass. Raid      Confirmed True      1
## 15 Worksite Raid      Confirmed True      6
## 16 Worksite Raid      Unconfirmed         1
```

III. Geographic Distribution of Reports

Reports – regardless of their validity – are concentrated in the following neighborhoods:

```
# Preparing the GeoJSON
# I. Read GeoJSON file containing Chicago neighborhoods ..........................................
geojson_file <- "chicago_community_boundaries.json"
chicago_communities <- st_read(geojson_file) %>%
mutate(shape_area = as.numeric(shape_area),
       shape_len = as.numeric(shape_len))
```

```
## Reading layer `chicago_community_boundaries' from data source
##   `D:\Datos\Monica\Documents\R_Studio\migrawatch\canvassing_analysis\chicago_community_boundaries.js
##   using driver `GeoJSON'
## Simple feature collection with 77 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.94011 ymin: 41.64454 xmax: -87.52414 ymax: 42.02304
## Geodetic CRS:  WGS 84
```

```
# II. Read Evanston GeoJSON file.............................................................
evanston_boundaries <- st_read("https://maps.cityofevanston.org/arcgis/rest/services/OpenData/ArcGISOpe
   mutate(community = "EVANSTON")
```

```
## Reading layer `OGRGeoJSON' from data source
##   `https://maps.cityofevanston.org/arcgis/rest/services/OpenData/ArcGISOpenData/MapServer/0/query?whe
##   using driver `GeoJSON'
## Simple feature collection with 1 feature and 4 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.73251 ymin: 42.01914 xmax: -87.66501 ymax: 42.07176
## Geodetic CRS:  WGS 84
```

```
colnames(evanston_boundaries)[2] <- "shape_area"
colnames(evanston_boundaries)[3] <- "shape_len"

# III. Read Cicero and Bewryn GeoJSON file .....................................................
cicero_berwyn_boundaries <- st_read("https://geoservices.epa.illinois.gov/arcgis/rest/services/Politica
  rename("community" = "DIST_NAME")
```

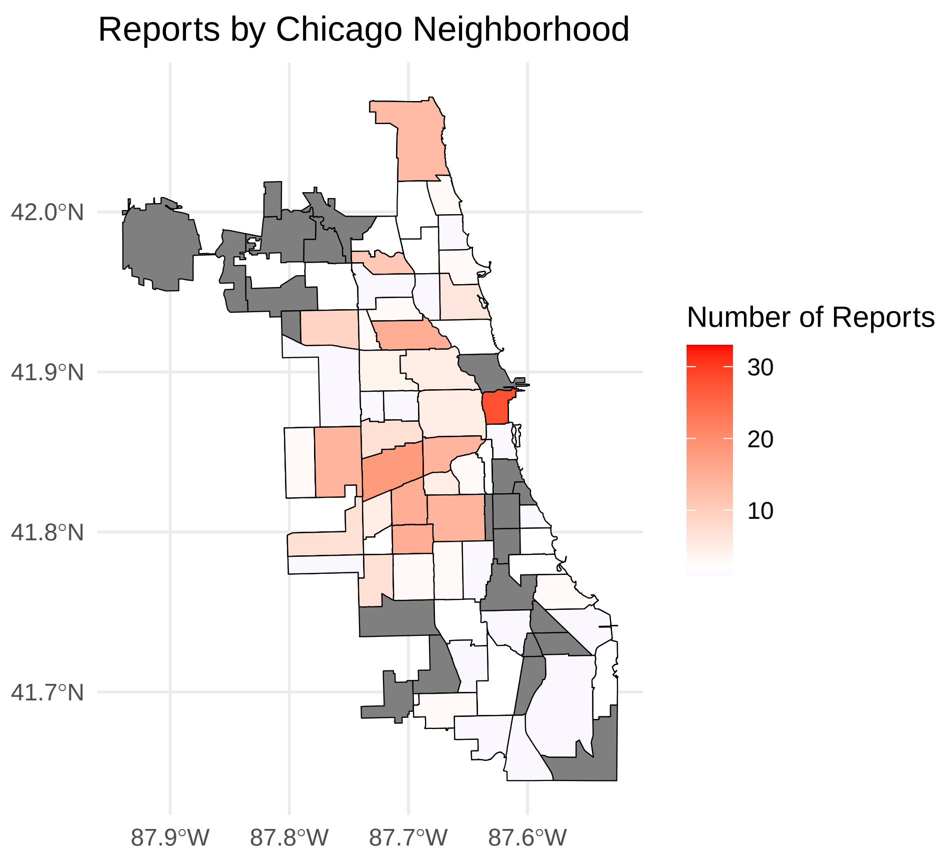
```
## Reading layer `OGRGeoJSON' from data source
##   `https://geoservices.epa.illinois.gov/arcgis/rest/services/Political/IllinoisPoliticalBoundaries/Ma
##   using driver `GeoJSON'
## Simple feature collection with 2 features and 12 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.80421 ymin: 41.82118 xmax: -87.73849 ymax: 41.86597
## Geodetic CRS:  WGS 84
```

```
colnames(cicero_berwyn_boundaries)[11] <- "shape_area"
colnames(cicero_berwyn_boundaries)[12] <- "shape_len"
```

```r
# IV. Combine GeoSON file .........................................................

# Ensure that our coordinate systems match
evanston_boundaries <- st_transform(evanston_boundaries, st_crs(chicago_communities))
cicero_berwyn_boundaries <- st_transform(cicero_berwyn_boundaries, st_crs(chicago_communities))

# Bind our data together
combined_sf_data <- bind_rows(chicago_communities, evanston_boundaries, cicero_berwyn_boundaries)

# Label neighborhoods to correspond with Chicago community areas

location_data <- data_labeled %>%
  mutate(community = str_to_upper(location),
         community = str_replace_all(community, "SWS -|SWS / ", ""),
         community = case_when(
                    str_detect(community, "PILSEN") ~ "LOWER WEST SIDE",
                    str_detect(community, "LITTLE VILLAGE") ~ "SOUTH LAWNDALE",
                    str_detect(community, "BACK OF THE YARDS|BOTY") ~ "NEW CITY",
                    str_detect(community, "DOWNTOWN|THE LOOP") ~ "LOOP",
                    str_detect(community, "SOUTH LOOP") ~ "NEAR SOUTH SIDE",
                    str_detect(community, "CHICAGO-PORTAGE PARK") ~ "PORTAGE PARK",
                    str_detect(community, "CHICAGO UPTOWN") ~ "UPTOWN",
                    str_detect(community, "CHICAGO ALBANY PARK") ~ "ALBANY PARK",
                    str_detect(community, "HUMBOLT PARK") ~ "HUMBOLDT PARK",
                    str_detect(community, "WEST LOOP|WICKER PARK|FULTON MARKET DISTRICT") ~ "NEAR WEST S
                    str_detect(community, "WEST 54TH AND PULASKI") ~ "WEST ELSDON",
                    str_detect(community, "BUCKTOWN") ~ "WEST TOWN",
                    str_detect(community, "MIDWAY") ~ "GARFIELD RIDGE",
                    str_detect(community, "CHINATOWN") ~ "ARMOUR SQUARE",
                    str_detect(community, "BELMONT/CAGRIN") ~ "BELMONT CRAGIN",
                    str_detect(community, "CALUMET PARK") ~ "WEST PULLMAN",
                    str_detect(community, "BLUE ISLAND") ~ "MORGAN PARK",
                    str_detect(community, "NORTH PULASKI|IRVING PARK") ~ "IRVING PARK",
                    str_detect(community, "ROSCOE VILLAGE") ~ "NORTH CENTER",
                    str_detect(community, "LAKEVIEW") ~ "LAKE VIEW",
                    str_detect(community, "BERWYN CICERO") ~ "BERWYN",
                    # relabel southwest data
                    unique_id == "09-25 Hermosa/Humboldt Park-4770 W Grand Ave, Chicago, IL 60639" ~ "HU
                    unique_id == "02-24 Chicago - SWS-South Shields and West 43rd PL." ~ "FULLER PARK",
                    unique_id == "02-26 Chicago - SWS-51st and St. Louis" ~ "GAGE PARK",
                    unique_id == "03-07 Chicago - SWS-51st and Spaulding" ~ "GAGE PARK",
                    unique_id == "03-14 Chicago - SWS-1352 32nd St." ~ "GAGE PARK",
                    unique_id == "04-16 Chicago - SWS-47th and Sacramento" ~ "BRIGHTON PARK",
                    unique_id == "04-27 Chicago - SWS -65th and Long" ~ "CLEARING",
                    unique_id == "09-25 Chicago - SWS-W 63rd St & S California Ave" ~ "NEW CITY",
                    unique_id == "05-14 Chicago - SWS-65th and Harlem" ~ "GARFIELD RIDGE",
                    unique_id == "06-16 Chicago - SWS-53rd and Pulaski" ~ "ARCHER HEIGHTS",
                    unique_id == "09-18 Chicago - SWS-42nd & Ashland" ~ "NEW CITY",
                    TRUE ~ community))
```

```r
all_reports_count <- location_data %>%
group_by(community, report_status) %>%
count()
```

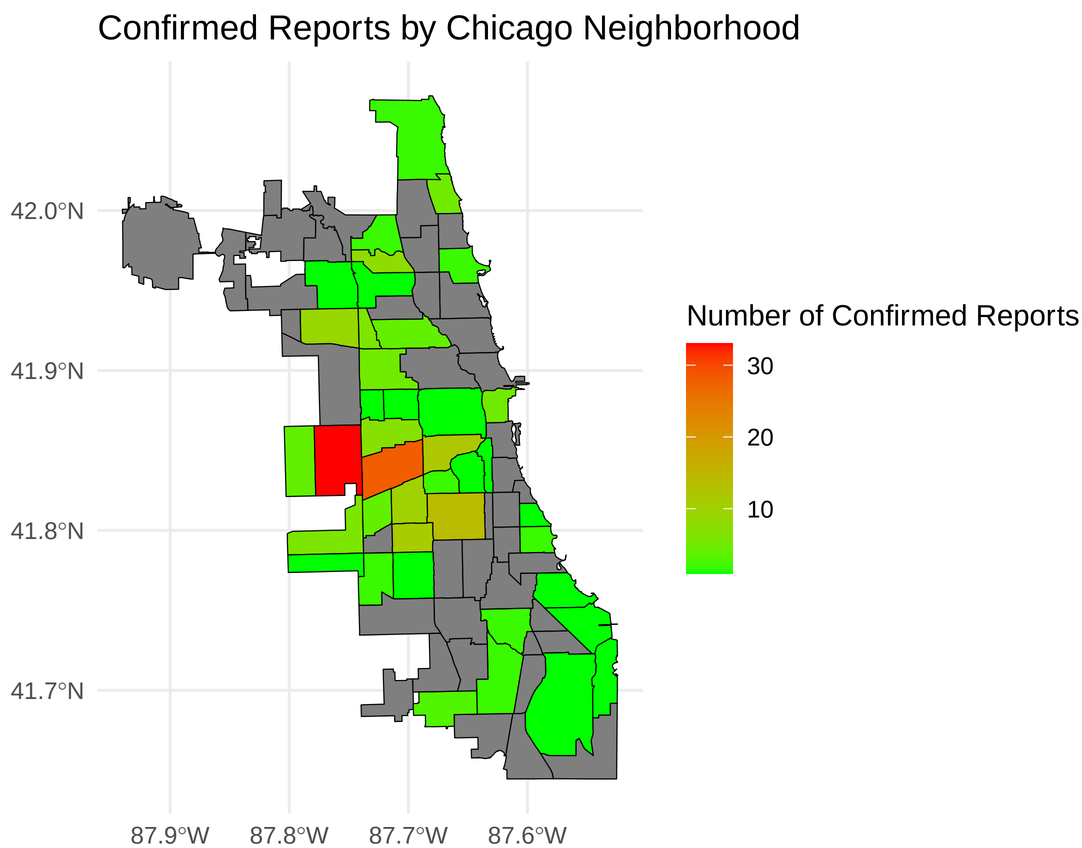
```
# join the counts with evanston data
map_df <- left_join(combined_sf_data, all_reports_count, by = "community")

# VI. Map my data ..............................................................
ggplot(map_df) +
  geom_sf(aes(fill = n), color = "black", size = 0.2) +
scale_fill_gradient2(
  low = "blue", high = "red", midpoint = 2, name = "Number of Reports") +
  labs(title = "Reports by Chicago Neighborhood") +
  theme_minimal() +
  theme(legend.position = "right")
```



Confirmed reports are clustered in the following neighbors:

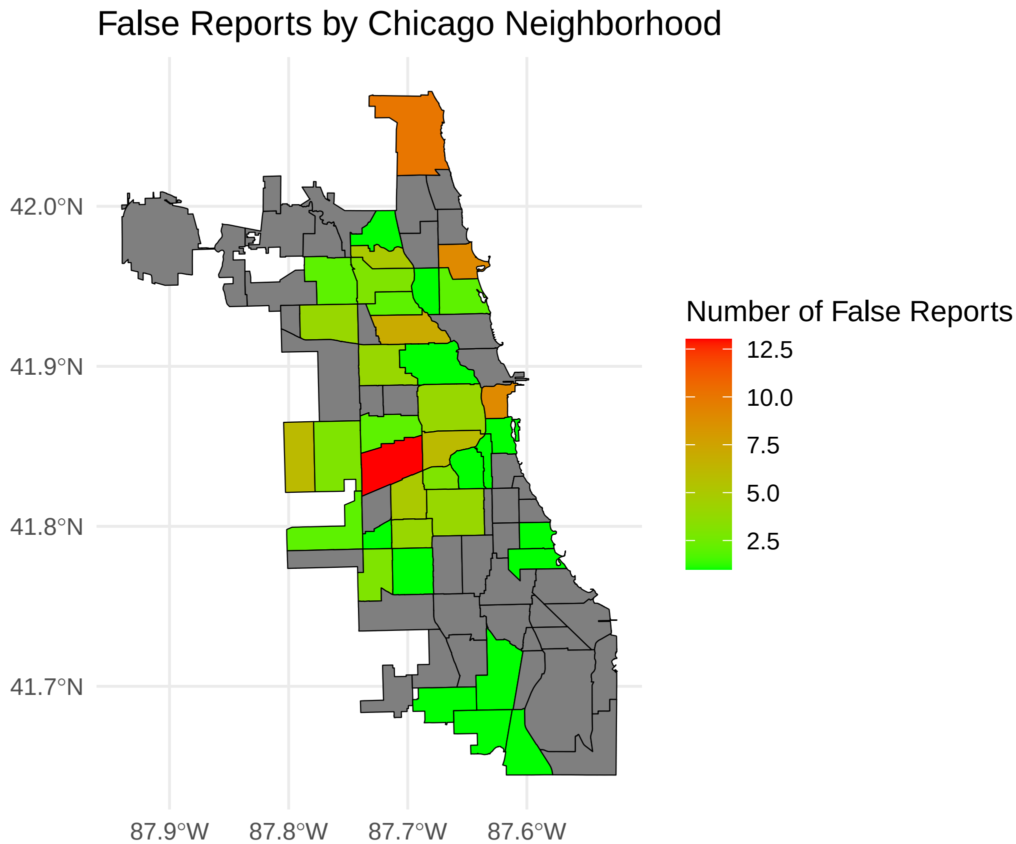
```
confirmed_reports_count <- location_data %>%
group_by(community, report_status) %>%
filter(report_status == "Confirmed True") %>%
count()

# join the counts with evanston data
map_df <- left_join(combined_sf_data, confirmed_reports_count, by = "community")

# VI. Map my data ..............................................................
ggplot(map_df) +
  geom_sf(aes(fill = n), color = "black", size = 0.2) +
scale_fill_gradient(
```

```
  low = "green", high = "red", name = "Number of Confirmed Reports") +
  labs(title = "Confirmed Reports by Chicago Neighborhood") +
  theme_minimal() +
  theme(legend.position = "right")
```



False reports are clustered in the following neighborhoods:

```
false_reports_count <- location_data %>%
group_by(community, report_status) %>%
filter(report_status == "Confirmed False") %>%
count()

# join the counts with evanston data
map_df <- left_join(combined_sf_data, false_reports_count, by = "community")

# VI. Map my data ...............................................................
ggplot(map_df) +
  geom_sf(aes(fill = n), color = "black", size = 0.2) +
scale_fill_gradient(
  low = "green", high = "red", name = "Number of False Reports") +
  labs(title = "False Reports by Chicago Neighborhood") +
  theme_minimal() +
  theme(legend.position = "right")
```



IV. Interpretation and Recommendations

1. ICE Sightings are by far the common type of the incident. Thus far, the network has recorded 272 confirmed ICE Sightings, 121 Public Space Raids, 52 Home Raids, 9 work site raids, and 6 I-9 raids.

2. Nearly 1 in 4 Ice Sightings were proven false by a rapid responder (22.3%). In contrast, only 2.0% percent of public space raids were proven false and 4.4% of home raids were proven false by a responder.

Considering this large discrepancy in valid ICE Sightings – a gap that raises to 49.4% when we add in sightings that were unable to be validated – I would recommend the FSN implement a triage system to prioritize responding to raids over ICE sightings.

Would also ask leadership to consider the following question: Is every sighting worth validating, especially in times when we are hurting for volunteer capacity? Is there a threshold that needs to be met for passing along a sighting? By developing a better protocol for identifying ICE sightings, OCAD and ICIRR can better direct teams.

3. In terms of the geographic distribution of reports, have a couple of takeaways I'd like to stress: A. High numbers of false reports are, in some ways, a good thing. That means people have utilized the ICIRR hotline and understand it's a resource (likely what we are seeing in Evanston, downtown, Little Village, and Uptown). In this case, it is safe to say that they are aware of the hotline's existence – they just likely need more KYR training.

Would recommend the Little Village team, in particular, do more KYR trainings since this neighborhood has been hit particularly heavily and also has a high quantity of false reports.

b. Lack of data can also tell us about a community. If we look at the South Side, we can see that there are few reports to begin with – and almost no false sightings. Would recommend some community canvasses that prioritizing raising awareness of the FSN and then move into KYR.