DSC630 Final Project - Crime Analysis - Part2

Amie Davis

28 June, 2020

Data Sources:

Uniform Crime Reporting Program Data: National Incident-Based Reporting System, [United States], 2016; United States Federal Bureau of Investigation; Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan; https://www.icpsr.umich.edu/icpsrweb/NACJD/NIBRS/
Geodetic Data for US Cities: https://simplemaps.com/data/us-cities

References:

https://www.latlong.net D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Load Libraries

```
library(readr)
library(dplyr)
library(ggmap)
```

1. Prepare Data

a) Import the Data

```
# Load Cleaned data from Part1
crime_data <- read_csv("Data/crime_top6_states.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Warning: 1334635 parsing failures.
##  row           col               expected actual
file
## 1337 ASSAULT_CIRC2 1/0/T/F/TRUE/FALSE          9
'Data/crime_top6_states.csv'
## 1338 ASSAULT_CIRC2 1/0/T/F/TRUE/FALSE          9
'Data/crime_top6_states.csv'
## 2038 OFF_CODE04    no trailing characters      A
'Data/crime_top6_states.csv'
```

```
## 2343 OFF_CODE04    no trailing characters      C
'Data/crime_top6_states.csv'
## 3141 ASSAULT_CIRC2 1/0/T/F/TRUE/FALSE          8
'Data/crime_top6_states.csv'
## .... ............. ...................... ......
.............................
## See problems(...) for more details.

#str(crime_data)
```

b) Remove Unnecessary Columns

```
# The following fields are not needed for visualizations and will be removed.

#  crime_data[ ,c(
#     "X1",
#     "VIC_INC_DATE",
#     "VICTIM_TYPE",
#     "ASSG_TYPE_OFFC",
#     "AGE_OF_VICTIM",
#     "SEX_OF_VICTIM",
#     "RACE_OF_VICTIM",
#     "ETHNIC_OF_VIC",
#
#     "INJURY_TYPE1",
#     "INJURY_TYPE2",
#     "INJURY_TYPE3",
#     "INJURY_TYPE4",
#     "INJURY_TYPE5",
#
#     "VIC_RESIDENT",
#     "NUM_RECS_PER_VICTIM",
#     "VIC_INC_YEAR",
#     "VIC_INC_MONTH",
#     "VIC_INC_DAY",
#     "VIC_INC_DOW",
#     "NUM_STATE_CODE",
#     "POP_GROUP",
#     "CTRY_DIVISION",
#     "CTRY_REGION",
#     "AGENCY_IND",
#     "CORE_CITY",
#     "FBI_OFFICE",
#     "JUDICIAL_DIST",
#     "CURRENT_POP1",
#     "UCR_COUNTY_CD1",
#     "MSA_CD1",
#     "LAST_POP1",
#     "FIPS_COUNTY1",
#     "Zip",
#
```

```
#     "city_ascii",
#     "state_name",
#     "county_fips",
#     "county_name",
#     "county_fips_all",
#     "county_name_all",
#     "population",
#     "density",
#     "source",
#     "military",
#     "incorporated",
#     "timezone",
#     "ranking",
#     "zips"
#)] <- list(NULL)

#head(crime_data)
```

c) Separate Offense Codes

```
# Create new df to include all offenses and their records
off1_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE01))
off2_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE02))
off3_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE03))
off4_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE04))
off5_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE05))
off6_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE06))
off7_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$OFF_CODE07))

names(off1_df)[names(off1_df) == "crime_data.OFF_CODE01"] <- "OFF_CODE"
names(off2_df)[names(off2_df) == "crime_data.OFF_CODE02"] <- "OFF_CODE"
names(off3_df)[names(off3_df) == "crime_data.OFF_CODE03"] <- "OFF_CODE"
names(off4_df)[names(off4_df) == "crime_data.OFF_CODE04"] <- "OFF_CODE"
names(off5_df)[names(off5_df) == "crime_data.OFF_CODE05"] <- "OFF_CODE"
names(off6_df)[names(off6_df) == "crime_data.OFF_CODE06"] <- "OFF_CODE"
names(off7_df)[names(off7_df) == "crime_data.OFF_CODE07"] <- "OFF_CODE"

off_df <- rbind(off1_df, off2_df, off3_df, off4_df, off5_df, off6_df,
off7_df)

head(off_df)
```

```
##   crime_data.ORI crime_data.INC_NUM OFF_CODE
## 1      MA0010100       83- X9Y 728N      26A
## 2      MA0010100       83-1X9Y 728N      23G
## 3      MA0010100       83-1XZ8 728N      13C
## 4      MA0010100       83-AXZ8 728N      23H
## 5      MA0010100       83-DX9Y 728N      290
## 6      MA0010100       83-DXZ8 728N      23C
```

d) Separate Aggrevated Assault Codes

```
# Convert to cahracter
crime_data$ASSAULT_CIRC1 <- as.character(crime_data$ASSAULT_CIRC1)
crime_data$ASSAULT_CIRC2 <- as.character(crime_data$ASSAULT_CIRC2)

# Create new df to include all Aggravated Assault/Homicide Circumstances and
their records
asst1_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$ASSAULT_CIRC1))
asst2_df <- na.omit(data.frame(crime_data$ORI, crime_data$INC_NUM,
crime_data$ASSAULT_CIRC2))

names(asst1_df)[names(asst1_df) == "crime_data.ASSAULT_CIRC1"] <-
"ASSAULT_CIRC"
names(asst2_df)[names(asst2_df) == "crime_data.ASSAULT_CIRC2"] <-
"ASSAULT_CIRC"

asst_df <- rbind(asst1_df, asst2_df)

head(asst_df)
```

```
##    crime_data.ORI crime_data.INC_NUM ASSAULT_CIRC
## 13      MA0010100       HE- -8ZOX5HP            1
## 14      MA0010100       HE- -8ZOX5HP            1
## 15      MA0010100       HE- -8ZOX5HP            1
## 30      MA0010100       HE- IYZOX5HP            9
## 35      MA0010100       HE- LLZOX5HP            1
## 50      MA0010100       HE- X ZOX5HP            1
```

e) Join Datasets

```
# Remove prefix from keys to match
names(off_df)[names(off_df) == "crime_data.ORI"] <- "ORI"
names(off_df)[names(off_df) == "crime_data.INC_NUM"] <- "INC_NUM"
#names(asst_df)[names(asst_df) == "crime_data.ORI"] <- "ORI"
#names(asst_df)[names(asst_df) == "crime_data.INC_NUM"] <- "INC_NUM"

# Join crime data with separate record for each offense
joint_off_df <- right_join(crime_data, off_df, c("ORI", "INC_NUM"))
#joint_all_df <- left_join(joint_off_df, asst_df, c("ORI", "INC_NUM"))

# Remove extra offense columns
```

```
  joint_off_df[ ,c(
     "OFF_CODE01",
     "OFF_CODE02",
     "OFF_CODE03",
     "OFF_CODE04",
     "OFF_CODE05",
     "OFF_CODE06",
     "OFF_CODE07"
)] <- list(NULL)

head(joint_off_df)
```

```
## # A tibble: 6 x 58
##      X1 ORI   INC_NUM VIC_INC_DATE VICTIM_TYPE ACT_TYPE_OFFC
ASSG_TYPE_OFFC
##   <dbl> <chr> <chr>   <date>       <chr>               <dbl> <chr>
## 1     1 MA00~ 83- X9~ 2016-01-01   I                      NA <NA>
## 2     2 MA00~ 83-1X9~ 2016-01-01   I                      NA <NA>
## 3     3 MA00~ 83-1XZ~ 2016-01-01   I                      NA <NA>
## 4     4 MA00~ 83-AXZ~ 2016-01-02   B                      NA <NA>
## 5     5 MA00~ 83-DX9~ 2016-01-02   I                      NA <NA>
## 6     6 MA00~ 83-DXZ~ 2016-01-02   B                      NA <NA>
## # ... with 51 more variables: AGE_OF_VICTIM <dbl>, SEX_OF_VICTIM <chr>,
## #   RACE_OF_VICTIM <chr>, ETHNIC_OF_VIC <chr>, VIC_RESIDENT <chr>,
## #   ASSAULT_CIRC1 <chr>, ASSAULT_CIRC2 <chr>, JUST_HOM_CIRC <lgl>,
## #   INJURY_TYPE1 <chr>, INJURY_TYPE2 <chr>, INJURY_TYPE3 <chr>,
## #   INJURY_TYPE4 <lgl>, INJURY_TYPE5 <lgl>, NUM_RECS_PER_VICTIM <dbl>,
## #   VIC_INC_YEAR <dbl>, VIC_INC_MONTH <dbl>, VIC_INC_DAY <dbl>,
## #   VIC_INC_DOW <chr>, NUM_STATE_CODE <dbl>, CITY <chr>, STATE <chr>,
## #   POP_GROUP <dbl>, CTRY_DIVISION <dbl>, CTRY_REGION <dbl>, AGENCY_IND
<dbl>,
## #   CORE_CITY <chr>, FBI_OFFICE <dbl>, JUDICIAL_DIST <chr>, CURRENT_POP1
<dbl>,
## #   UCR_COUNTY_CD1 <dbl>, MSA_CD1 <dbl>, LAST_POP1 <dbl>, FIPS_COUNTY1
<chr>,
## #   city_ascii <chr>, state_name <chr>, county_fips <dbl>, county_name
<chr>,
## #   county_fips_all <dbl>, county_name_all <chr>, lat <dbl>, lng <dbl>,
## #   population <dbl>, density <dbl>, source <chr>, military <lgl>,
## #   incorporated <lgl>, timezone <chr>, ranking <dbl>, zips <chr>, id
<dbl>,
## #   OFF_CODE <chr>
```

```
#summary(joint_off_df)

# Output joined file for use in modeling
write.csv(joint_off_df,"Data/crime_offenses.csv")
```

f) Apply filters

```
# Create subset to include only records marked as justifiable homicide
just_hom_df <- filter(joint_off_df, OFF_CODE == "09C")

summary(just_hom_df)
```

```
##        X1              ORI              INC_NUM           VIC_INC_DATE
##  Min.   : 290676   Length:101         Length:101         Min.   :2016-01-
05
##  1st Qu.:1340826   Class :character   Class :character   1st Qu.:2016-03-
26
##  Median :1655552   Mode  :character   Mode  :character   Median :2016-07-
01
##  Mean   :1621686                                         Mean   :2016-06-
30
##  3rd Qu.:2172345                                         3rd Qu.:2016-09-
25
##  Max.   :2655976                                         Max.   :2016-12-
28
##
##  VICTIM_TYPE        ACT_TYPE_OFFC ASSG_TYPE_OFFC     AGE_OF_VICTIM
##  Length:101         Min.   : NA   Length:101         Min.   :15.00
##  Class :character   1st Qu.: NA   Class :character   1st Qu.:25.00
##  Mode  :character   Median : NA   Mode  :character   Median :32.00
##                     Mean   :NaN                      Mean   :36.11
##                     3rd Qu.: NA                      3rd Qu.:44.00
##                     Max.   : NA                      Max.   :73.00
##                     NA's   :101
##  SEX_OF_VICTIM      RACE_OF_VICTIM     ETHNIC_OF_VIC      VIC_RESIDENT
##  Length:101         Length:101         Length:101         Length:101
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  ASSAULT_CIRC1      ASSAULT_CIRC2      JUST_HOM_CIRC   INJURY_TYPE1
##  Length:101         Length:101         Mode :logical   Length:101
##  Class :character   Class :character   FALSE:2         Class :character
##  Mode  :character   Mode  :character   NA's :99        Mode  :character
##
##
##
##
##  INJURY_TYPE2       INJURY_TYPE3       INJURY_TYPE4   INJURY_TYPE5
##  Length:101         Length:101         Mode:logical   Mode:logical
##  Class :character   Class :character   NA's:101       NA's:101
##  Mode  :character   Mode  :character
##
```

```
##
##
##
##  NUM_RECS_PER_VICTIM  VIC_INC_YEAR  VIC_INC_MONTH     VIC_INC_DAY
##  Min.   :1.000       Min.   :2016   Min.   : 1.000   Min.   : 1.00
##  1st Qu.:1.000       1st Qu.:2016   1st Qu.: 3.000   1st Qu.: 7.00
##  Median :1.000       Median :2016   Median : 7.000   Median :13.00
##  Mean   :1.041       Mean   :2016   Mean   : 6.495   Mean   :15.21
##  3rd Qu.:1.000       3rd Qu.:2016   3rd Qu.: 9.000   3rd Qu.:25.00
##  Max.   :2.000       Max.   :2016   Max.   :12.000   Max.   :30.00
##  NA's   :4
##  VIC_INC_DOW        NUM_STATE_CODE     CITY              STATE
##  Length:101         Min.   :21.0   Length:101         Length:101
##  Class :character   1st Qu.:39.0   Class :character   Class :character
##  Mode  :character   Median :39.0   Mode  :character   Mode  :character
##                     Mean   :36.9
##                     3rd Qu.:41.0
##                     Max.   :46.0
##
##    POP_GROUP     CTRY_DIVISION    CTRY_REGION      AGENCY_IND
##  Min.   :2.000   Min.   :3.000   Min.   :2.000   Min.   :1.000
##  1st Qu.:2.750   1st Qu.:5.000   1st Qu.:3.000   1st Qu.:1.000
##  Median :3.000   Median :5.000   Median :3.000   Median :1.000
##  Mean   :3.583   Mean   :5.386   Mean   :2.901   Mean   :1.337
##  3rd Qu.:4.000   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :6.000   Max.   :9.000   Max.   :4.000   Max.   :4.000
##  NA's   :65
##   CORE_CITY           FBI_OFFICE   JUDICIAL_DIST       CURRENT_POP1
##  Length:101         Min.   :3180   Length:101         Min.   :  5253
##  Class :character   1st Qu.:3180   Class :character   1st Qu.: 65686
##  Mode  :character   Median :3220   Mode  :character   Median :177801
##                     Mean   :3346                      Mean   :304851
##                     3rd Qu.:3440                      3rd Qu.:656434
##                     Max.   :3840                      Max.   :700313
##
##  UCR_COUNTY_CD1     MSA_CD1        LAST_POP1      FIPS_COUNTY1
##  Min.   : 2.00   Min.   : 41.0   Min.   :  5250   Length:101
##  1st Qu.:21.00   1st Qu.:223.0   1st Qu.: 64683   Class :character
##  Median :42.00   Median :383.0   Median :174969   Mode  :character
##  Mean   :48.65   Mean   :424.1   Mean   :304126
##  3rd Qu.:79.00   3rd Qu.:519.0   3rd Qu.:657936
##  Max.   :94.00   Max.   :981.0   Max.   :683700
##  NA's   :2       NA's   :6
##   city_ascii         state_name         county_fips    county_name
##  Length:101         Length:101         Min.   :26049   Length:101
##  Class :character   Class :character   1st Qu.:45010   Class :character
##  Mode  :character   Mode  :character   Median :45083   Mode  :character
##                                        Mean   :43197
##                                        3rd Qu.:47157
##                                        Max.   :53077
```

```
##                                        NA's   :3
##  county_fips_all county_name_all         lat             lng
##  Min.   :26049   Length:101         Min.   :32.82   Min.   :-122.45
##  1st Qu.:45007   Class :character   1st Qu.:34.84   1st Qu.: -89.98
##  Median :46047   Mode  :character   Median :35.10   Median : -83.71
##  Mean   :42826                      Mean   :37.57   Mean   : -88.42
##  3rd Qu.:47157                      3rd Qu.:42.37   3rd Qu.: -82.36
##  Max.   :53077                      Max.   :48.01   Max.   : -78.72
##  NA's   :13                         NA's   :3       NA's   :3
##    population         density          source           military
##  Min.   :    409   Min.   : 129.0   Length:101         Mode :logical
##  1st Qu.:  44494   1st Qu.: 537.2   Class :character   FALSE:98
##  Median : 293388   Median : 793.0   Mode  :character   NA's :3
##  Mean   : 806461   Mean   : 913.2
##  3rd Qu.:1068873   3rd Qu.:1036.8
##  Max.   :3643765   Max.   :3336.0
##  NA's   :3         NA's   :3
##  incorporated      timezone           ranking          zips
##  Mode :logical   Length:101         Min.   :1.000   Length:101
##  FALSE:1         Class :character   1st Qu.:1.000   Class :character
##  TRUE :97        Mode  :character   Median :2.000   Mode  :character
##  NA's :3                            Mean   :2.031
##                                     3rd Qu.:3.000
##                                     Max.   :3.000
##                                     NA's   :3
##        id             OFF_CODE
##  Min.   :1.84e+09   Length:101
##  1st Qu.:1.84e+09   Class :character
##  Median :1.84e+09   Mode  :character
##  Mean   :1.84e+09
##  3rd Qu.:1.84e+09
##  Max.   :1.84e+09
##  NA's   :3
```

```
# There are 101 incidents of justifiable homicide reported to UCR in 2016 for
the top 6 states

# Create subset to include only records marked as aggravated assault
agg_asst_df <- filter(joint_off_df, OFF_CODE == "13A")

summary(agg_asst_df)
```

```
##        X1              ORI              INC_NUM           VIC_INC_DATE
##  Min.   :     13   Length:223507      Length:223507      Min.   :2016-01-
01
##  1st Qu.: 648953   Class :character   Class :character   1st Qu.:2016-04-
09
##  Median :1481438   Mode  :character   Mode  :character   Median :2016-07-
04
##  Mean   :1346200                                        Mean   :2016-07-
```

```
03
##  3rd Qu.:1994929                                          3rd Qu.:2016-09-
29
##  Max.   :2658812                                          Max.   :2016-12-
31
##
##  VICTIM_TYPE        ACT_TYPE_OFFC    ASSG_TYPE_OFFC     AGE_OF_VICTIM
##  Length:223507      Min.   : 1.00    Length:223507      Min.   : 0.00
##  Class :character   1st Qu.: 1.00    Class :character   1st Qu.:20.00
##  Mode  :character   Median : 4.00    Mode  :character   Median :29.00
##                     Mean   : 5.13                       Mean   :30.61
##                     3rd Qu.: 9.00                       3rd Qu.:40.00
##                     Max.   :11.00                       Max.   :99.00
##                     NA's   :217623
##  SEX_OF_VICTIM      RACE_OF_VICTIM     ETHNIC_OF_VIC      VIC_RESIDENT
##  Length:223507      Length:223507      Length:223507      Length:223507
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  ASSAULT_CIRC1      ASSAULT_CIRC2      JUST_HOM_CIRC  INJURY_TYPE1
##  Length:223507      Length:223507      Mode:logical   Length:223507
##  Class :character   Class :character   NA's:223507    Class :character
##  Mode  :character   Mode  :character                  Mode  :character
##
##
##
##
##  INJURY_TYPE2       INJURY_TYPE3       INJURY_TYPE4   INJURY_TYPE5
##  Length:223507      Length:223507      Mode:logical   Mode:logical
##  Class :character   Class :character   TRUE:7         NA's:223507
##  Mode  :character   Mode  :character   NA's:223500
##
##
##
##
##  NUM_RECS_PER_VICTIM  VIC_INC_YEAR  VIC_INC_MONTH     VIC_INC_DAY
##  Min.   : 1.00        Min.   :2016   Min.   : 1.000   Min.   : 1.00
##  1st Qu.: 1.00        1st Qu.:2016   1st Qu.: 4.000   1st Qu.: 8.00
##  Median : 1.00        Median :2016   Median : 7.000   Median :16.00
##  Mean   : 1.87        Mean   :2016   Mean   : 6.588   Mean   :15.68
##  3rd Qu.: 2.00        3rd Qu.:2016   3rd Qu.: 9.000   3rd Qu.:24.00
##  Max.   :41.00        Max.   :2016   Max.   :12.000   Max.   :31.00
##  NA's   :104030
##  VIC_INC_DOW        NUM_STATE_CODE      CITY              STATE
##  Length:223507      Min.   :20.00   Length:223507      Length:223507
##  Class :character   1st Qu.:21.00   Class :character   Class :character
##  Mode  :character   Median :39.00   Mode  :character   Mode  :character
```

```
##                    Mean   :33.85
##                    3rd Qu.:41.00
##                    Max.   :46.00
##
##    POP_GROUP      CTRY_DIVISION    CTRY_REGION      AGENCY_IND
##  Min.   :2.00     Min.   :1.000   Min.   :1.000   Min.   :0.000
##  1st Qu.:3.00     1st Qu.:3.000   1st Qu.:2.000   1st Qu.:1.000
##  Median :4.00     Median :5.000   Median :3.000   Median :1.000
##  Mean   :3.74     Mean   :4.619   Mean   :2.559   Mean   :1.286
##  3rd Qu.:5.00     3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:1.000
##  Max.   :7.00     Max.   :9.000   Max.   :4.000   Max.   :7.000
##  NA's   :115351
##   CORE_CITY           FBI_OFFICE   JUDICIAL_DIST       CURRENT_POP1
##  Length:223507      Min.   :3090   Length:223507      Min.   :   136
##  Class :character   1st Qu.:3180   Class :character   1st Qu.: 35251
##  Mode  :character   Median :3220   Mode  :character   Median :110442
##                     Mean   :3314                      Mean   :242272
##                     3rd Qu.:3440                      3rd Qu.:398132
##                     Max.   :3840                      Max.   :844206
##
##  UCR_COUNTY_CD1     MSA_CD1        LAST_POP1      FIPS_COUNTY1
##  Min.   : 1.00   Min.   :  9.0   Min.   :   312   Length:223507
##  1st Qu.:17.00   1st Qu.:179.0   1st Qu.: 35279   Class :character
##  Median :37.00   Median :353.0   Median :109940   Mode  :character
##  Mean   :42.19   Mean   :405.2   Mean   :241131
##  3rd Qu.:77.00   3rd Qu.:570.0   3rd Qu.:393883
##  Max.   :95.00   Max.   :989.0   Max.   :829712
##  NA's   :12817   NA's   :26889
##   city_ascii         state_name         county_fips    county_name
##  Length:223507      Length:223507      Min.   :25001   Length:223507
##  Class :character   Class :character   1st Qu.:26163   Class :character
##  Mode  :character   Mode  :character   Median :45045   Mode  :character
##                                        Mean   :39954
##                                        3rd Qu.:47149
##                                        Max.   :53077
##                                        NA's   :42139
##  county_fips_all county_name_all         lat             lng
##  Min.   :25001   Length:223507      Min.   :32.21   Min.   :-124.39
##  1st Qu.:26163   Class :character   1st Qu.:35.10   1st Qu.: -87.34
##  Median :45051   Mode  :character   Median :39.14   Median : -83.69
##  Mean   :39837                      Mean   :39.03   Mean   : -86.75
##  3rd Qu.:47149                      3rd Qu.:42.38   3rd Qu.: -81.68
##  Max.   :53077                      Max.   :48.98   Max.   : -70.36
##  NA's   :55476                      NA's   :42139   NA's   :42139
##    population         density         source           military
##  Min.   :    135   Min.   :  24    Length:223507      Mode :logical
##  1st Qu.:  42174   1st Qu.: 541    Class :character   FALSE:181368
##  Median : 213418   Median : 793    Mode  :character   NA's :42139
##  Mean   : 753901   Mean   :1103
##  3rd Qu.:1068873   3rd Qu.:1551
```

```
##  Max.   :3643765   Max.   :7616
##  NA's   :42139     NA's   :42139
##  incorporated      timezone            ranking          zips
##  Mode :logical   Length:223507      Min.   :1.00    Length:223507
##  FALSE:634       Class :character   1st Qu.:1.00    Class :character
##  TRUE :180734    Mode  :character   Median :2.00    Mode  :character
##  NA's :42139                        Mean   :2.03
##                                     3rd Qu.:3.00
##                                     Max.   :3.00
##                                     NA's   :42139
##        id             OFF_CODE
##  Min.   :1.84e+09   Length:223507
##  1st Qu.:1.84e+09   Class :character
##  Median :1.84e+09   Mode  :character
##  Mean   :1.84e+09
##  3rd Qu.:1.84e+09
##  Max.   :1.84e+09
##  NA's   :42139
```

```
#head(agg_asst_df)
# There are 223,507 incidents of aggravated assaults reported to UCR in 2016
for the top 6 states

# Create subset to include only records marked as investigation of suspicious
activity
act_type_df <- filter(joint_off_df, ACT_TYPE_OFFC == "7")

summary(act_type_df)
```

```
##        X1              ORI              INC_NUM           VIC_INC_DATE
##  Min.   :   3947   Length:1522        Length:1522        Min.   :2016-01-
01
##  1st Qu.: 324337   Class :character   Class :character   1st Qu.:2016-04-
10
##  Median :1517947   Mode  :character   Mode  :character   Median :2016-06-
20
##  Mean   :1326105                                         Mean   :2016-06-
30
##  3rd Qu.:2119250                                         3rd Qu.:2016-09-
18
##  Max.   :2650603                                         Max.   :2016-12-
31
##
##  VICTIM_TYPE        ACT_TYPE_OFFC ASSG_TYPE_OFFC     AGE_OF_VICTIM
##  Length:1522        Min.   :7     Length:1522        Min.   : 0.00
##  Class :character   1st Qu.:7     Class :character   1st Qu.:27.00
##  Mode  :character   Median :7     Mode  :character   Median :33.00
##                     Mean   :7                        Mean   :33.54
##                     3rd Qu.:7                        3rd Qu.:43.00
##                     Max.   :7                        Max.   :99.00
```

```
##
##  SEX_OF_VICTIM      RACE_OF_VICTIM     ETHNIC_OF_VIC      VIC_RESIDENT
##  Length:1522        Length:1522        Length:1522        Length:1522
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  ASSAULT_CIRC1      ASSAULT_CIRC2      JUST_HOM_CIRC  INJURY_TYPE1
##  Length:1522        Length:1522        Mode:logical   Length:1522
##  Class :character   Class :character   NA's:1522      Class :character
##  Mode  :character   Mode  :character                  Mode  :character
##
##
##
##
##  INJURY_TYPE2       INJURY_TYPE3       INJURY_TYPE4   INJURY_TYPE5
##  Length:1522        Length:1522        Mode:logical   Mode:logical
##  Class :character   Class :character   NA's:1522      NA's:1522
##  Mode  :character   Mode  :character
##
##
##
##
##  NUM_RECS_PER_VICTIM  VIC_INC_YEAR  VIC_INC_MONTH     VIC_INC_DAY
##  Min.   : 1.000       Min.   :2016   Min.   : 1.000   Min.   : 1.00
##  1st Qu.: 1.000       1st Qu.:2016   1st Qu.: 4.000   1st Qu.: 8.00
##  Median : 2.000       Median :2016   Median : 6.000   Median :17.00
##  Mean   : 2.253       Mean   :2016   Mean   : 6.489   Mean   :16.04
##  3rd Qu.: 3.000       3rd Qu.:2016   3rd Qu.: 9.000   3rd Qu.:24.00
##  Max.   :10.000       Max.   :2016   Max.   :12.000   Max.   :31.00
##  NA's   :699
##  VIC_INC_DOW        NUM_STATE_CODE      CITY              STATE
##  Length:1522        Min.   :20.00   Length:1522        Length:1522
##  Class :character   1st Qu.:21.00   Class :character   Class :character
##  Mode  :character   Median :39.00   Mode  :character   Mode  :character
##                     Mean   :33.95
##                     3rd Qu.:41.00
##                     Max.   :46.00
##
##    POP_GROUP      CTRY_DIVISION    CTRY_REGION      AGENCY_IND
##  Min.   :2.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:1.000
##  Median :3.000   Median :5.000   Median :3.000   Median :1.000
##  Mean   :3.794   Mean   :4.791   Mean   :2.559   Mean   :1.209
##  3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:1.000
##  Max.   :7.000   Max.   :9.000   Max.   :4.000   Max.   :5.000
##  NA's   :420
##   CORE_CITY           FBI_OFFICE   JUDICIAL_DIST       CURRENT_POP1
```

```
##  Length:1522        Min.   :3090   Length:1522        Min.   :   532
##  Class :character   1st Qu.:3160   Class :character   1st Qu.: 29145
##  Mode  :character   Median :3220   Mode  :character   Median : 79837
##                     Mean   :3362                      Mean   :144597
##                     3rd Qu.:3440                      3rd Qu.:152526
##                     Max.   :3840                      Max.   :844206
##
##  UCR_COUNTY_CD1    MSA_CD1        LAST_POP1      FIPS_COUNTY1
##  Min.   : 1     Min.   :  9.0   Min.   :   526   Length:1522
##  1st Qu.: 9     1st Qu.:223.0   1st Qu.: 29134   Class :character
##  Median :21     Median :413.0   Median : 78214   Mode  :character
##  Mean   :31     Mean   :466.1   Mean   :142997
##  3rd Qu.:45     3rd Qu.:690.0   3rd Qu.:150319
##  Max.   :95     Max.   :981.0   Max.   :829712
##  NA's   :25     NA's   :140
##   city_ascii         state_name         county_fips    county_name
##  Length:1522        Length:1522        Min.   :25003   Length:1522
##  Class :character   Class :character   1st Qu.:26103   Class :character
##  Mode  :character   Mode  :character   Median :45051   Mode  :character
##                                        Mean   :40536
##                                        3rd Qu.:53005
##                                        Max.   :53077
##                                        NA's   :295
##  county_fips_all county_name_all         lat             lng
##  Min.   :25003   Length:1522        Min.   :32.46   Min.   :-124.15
##  1st Qu.:26099   Class :character   1st Qu.:35.96   1st Qu.:-119.17
##  Median :45051   Mode  :character   Median :42.12   Median : -83.21
##  Mean   :40451                      Mean   :40.95   Mean   : -90.92
##  3rd Qu.:53005                      3rd Qu.:45.99   3rd Qu.: -79.87
##  Max.   :53077                      Max.   :48.75   Max.   : -70.89
##  NA's   :392                        NA's   :295     NA's   :295
##    population         density          source           military
##  Min.   :    409   Min.   :  62.0   Length:1522        Mode :logical
##  1st Qu.:  30980   1st Qu.: 571.5   Class :character   FALSE:1227
##  Median :  96690   Median :1042.0   Mode  :character   NA's :295
##  Mean   : 461899   Mean   :1304.6
##  3rd Qu.: 467894   3rd Qu.:1679.0
##  Max.   :3643765   Max.   :7616.0
##  NA's   :295       NA's   :295
##  incorporated     timezone           ranking          zips
##  Mode:logical   Length:1522        Min.   :1.000   Length:1522
##  TRUE:1227      Class :character   1st Qu.:2.000   Class :character
##  NA's:295       Mode  :character   Median :2.000   Mode  :character
##                                    Mean   :2.311
##                                    3rd Qu.:3.000
##                                    Max.   :3.000
##                                    NA's   :295
##        id             OFF_CODE
##  Min.   :1.84e+09   Length:1522
##  1st Qu.:1.84e+09   Class :character
```

```
##  Median :1.84e+09   Mode  :character
##  Mean   :1.84e+09
##  3rd Qu.:1.84e+09
##  Max.   :1.84e+09
##  NA's   :295
```

```
#head(act_type_df)
# There are 1,522 incidents of investigation of suspicious activity reported
to UCR in 2016 for the top 6 states
```

2. Visualizations for Features of Concern

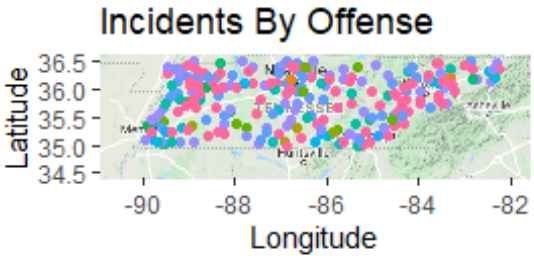
Using ggmap to use Google maps to display geodetic information. The API key is passed, but is hidden from Markdown.

a) Tennessee

```
# Create TN subset of all offense
tn_df <- filter(joint_off_df, STATE == "TN")

# Use Tennesee coordinates as center
tn_map <- ggmap(get_googlemap(
                  center = c(lon = -86.660156, lat = 35.860119),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
tn_map +

  # Re-scale to focus on TN rectangle
  scale_y_continuous(limits=c(34.5, 36.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = tn_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())
```

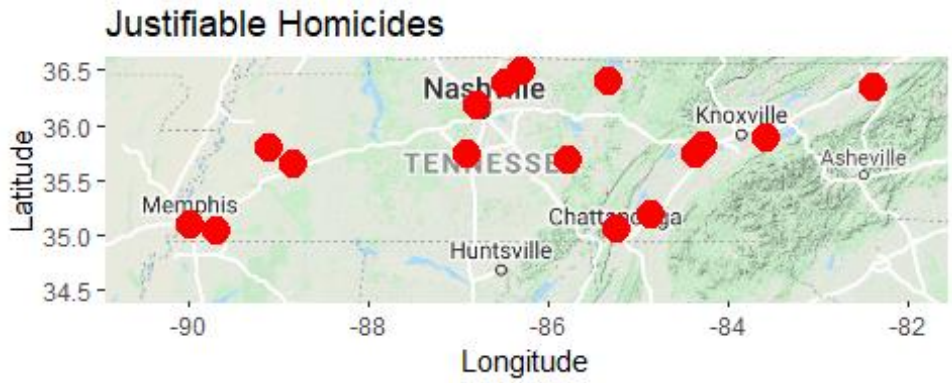
```
## Warning: Removed 1 rows containing missing values (geom_rect).
```

```
## Warning: Removed 191448 rows containing missing values (geom_point).
```

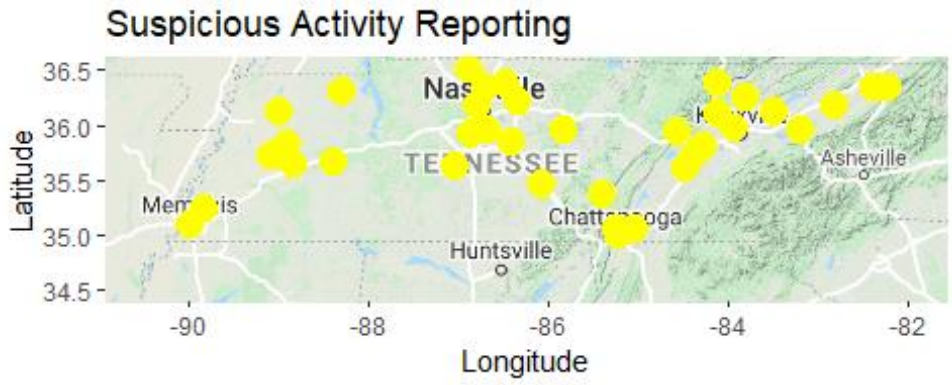


09A	23C	290
09B	23D	35A
09C	23E	35B
100	23F	36A
11A	23G	36B
11B	23H	370
11C	240	39A
11D	250	39B
120	26A	39C
13A	26B	39D
13B	26C	40A
13C	26D	40B
200	26E	40C
210	26F	510
220	26G	520
23A	270	64A
23B	280	720

```
# Create TN subset of Justifiable Homicides
tn_spec_df <- filter(joint_off_df, STATE == "TN" & (OFF_CODE == "09C"))

# Build map
tn_map +

  # Re-scale to focus on TN rectangle
  scale_y_continuous(limits=c(34.5, 36.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = tn_spec_df,
             aes(x=lng, y=lat, size=3),
             color="red",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Justifiable Homicides")

## Warning: Removed 1 rows containing missing values (geom_rect).

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



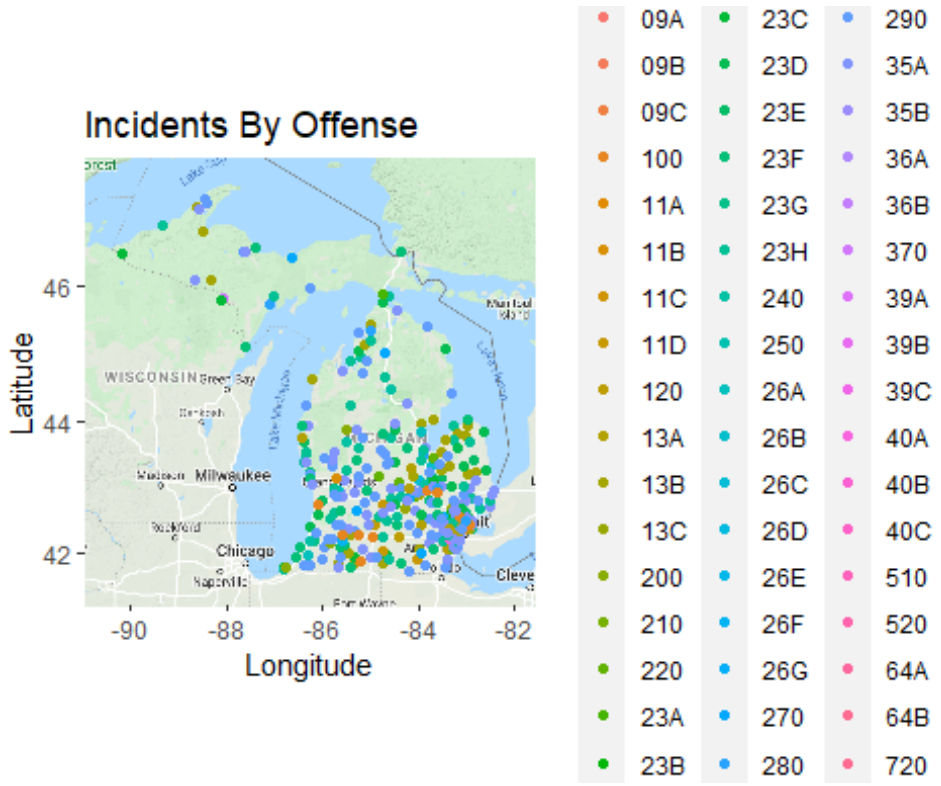

```
# Create TN subset of Suspicious Activity Reporting
tn_spec_df <- filter(joint_off_df, STATE == "TN" & (ACT_TYPE_OFFC == "7"))

# Build map
tn_map +

  # Re-scale to focus on TN rectangle
  scale_y_continuous(limits=c(34.5, 36.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = tn_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")

## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 143 rows containing missing values (geom_point).
```



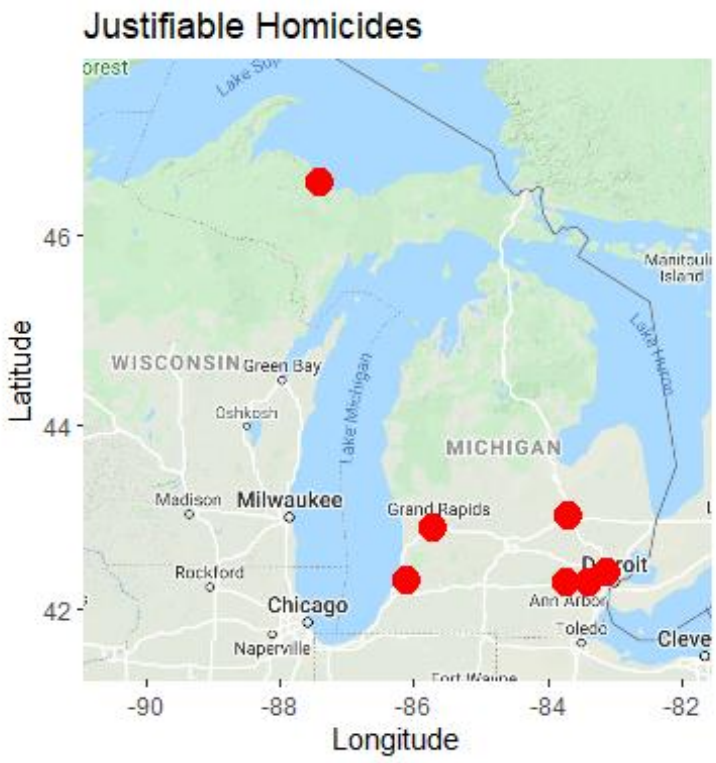
b) Michigan

```
# Create MI subset of all offense
mi_df <- filter(joint_off_df, STATE == "MI")

# Use Michigan coordinates as center
mi_map <- ggmap(get_googlemap(
                  center = c(lon = -84.506836, lat = 44.182205),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
mi_map +

  # Re-scale to focus on MI rectangle
  scale_y_continuous(limits=c(41.5, 47.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = mi_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())

## Warning: Removed 1 rows containing missing values (geom_rect).
```

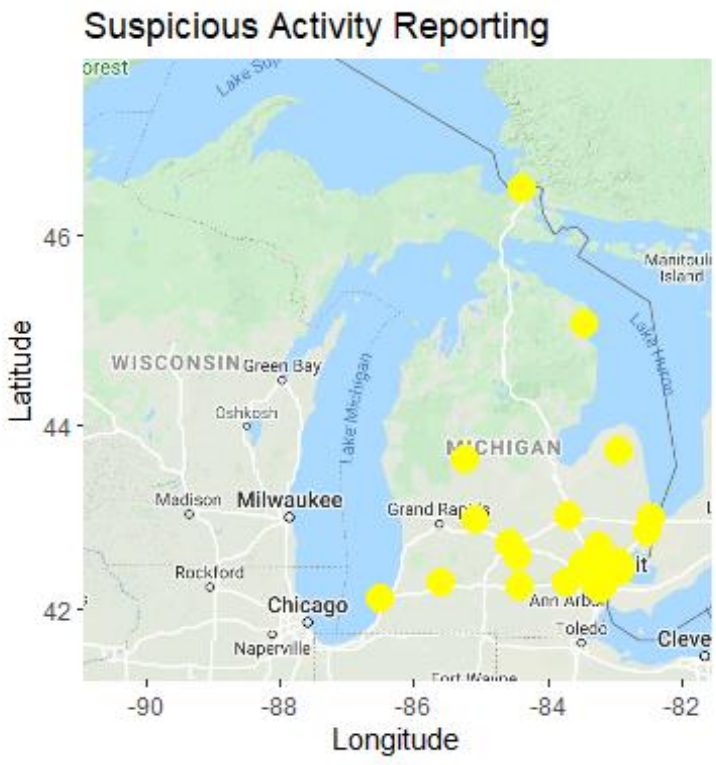
```
## Warning: Removed 222211 rows containing missing values (geom_point).
```




```
# Create MI subset of Justifiable Homicides
mi_spec_df <- filter(joint_off_df, STATE == "MI" & (OFF_CODE == "09C"))

# Build map
mi_map +

  # Re-scale to focus on MI rectangle
  scale_y_continuous(limits=c(41.5, 47.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = mi_spec_df,
             aes(x=lng, y=lat, size=3),
             color="red",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Justifiable Homicides")

## Warning: Removed 1 rows containing missing values (geom_rect).

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



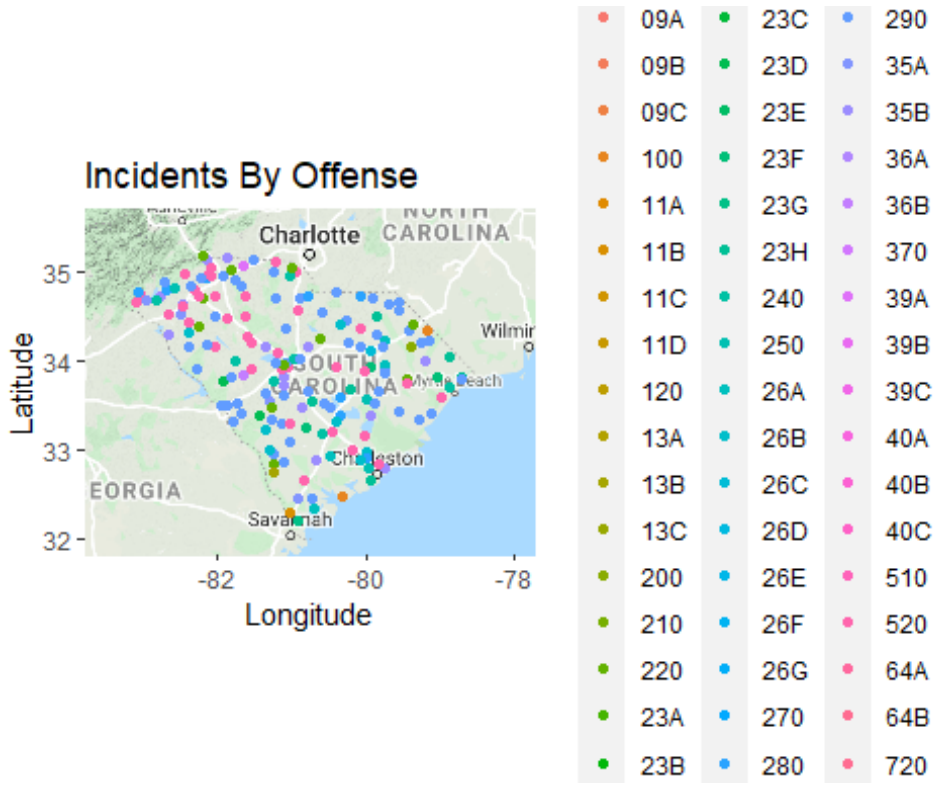

```
# Create MI subset of Suspicious Activity Reporting
mi_spec_df <- filter(joint_off_df, STATE == "MI" & (ACT_TYPE_OFFC == "7"))

# Build map
mi_map +

  # Re-scale to focus on MI rectangle
  scale_y_continuous(limits=c(41.5, 47.5)) +
  scale_x_continuous(limits=c(-90.5, -82)) +

  # Add data points to the map
  geom_point(data = mi_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")

## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 28 rows containing missing values (geom_point).
```



c) South Carolina

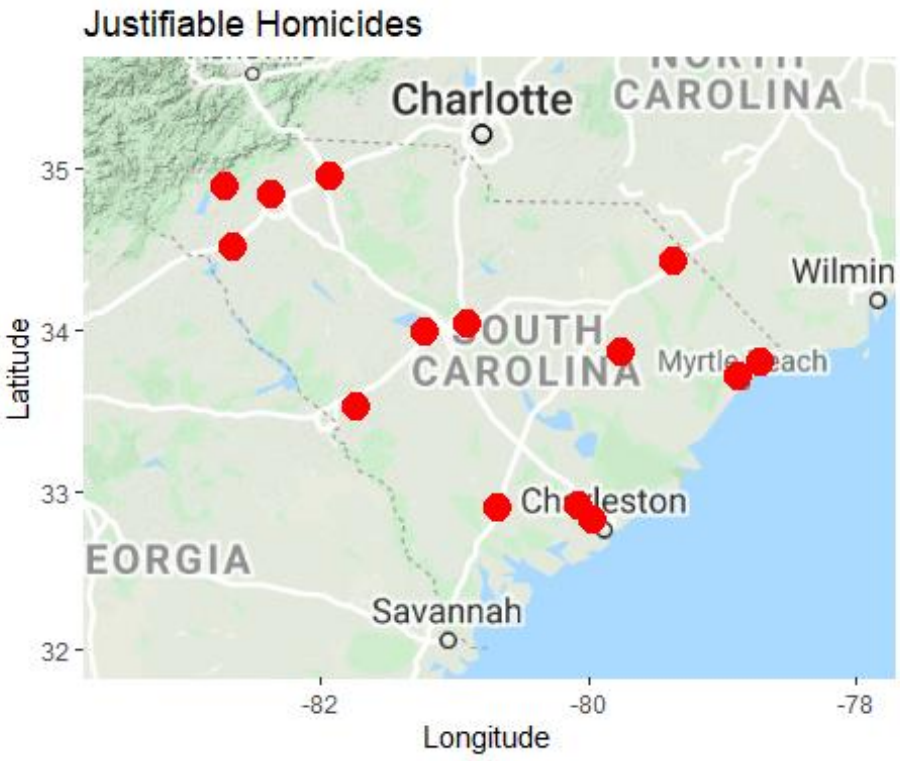
```
# Create SC subset of all offense
sc_df <- filter(joint_off_df, STATE == "SC")

# Use South Carolina coordinates as center
sc_map <- ggmap(get_googlemap(
                  center = c(lon = -81.163727, lat = 33.836082),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
sc_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(32, 35.5)) +
  scale_x_continuous(limits=c(-83.5, -78)) +

  # Add data points to the map
  geom_point(data = sc_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())

## Warning: Removed 1 rows containing missing values (geom_rect).
```

```
## Warning: Removed 74225 rows containing missing values (geom_point).
```

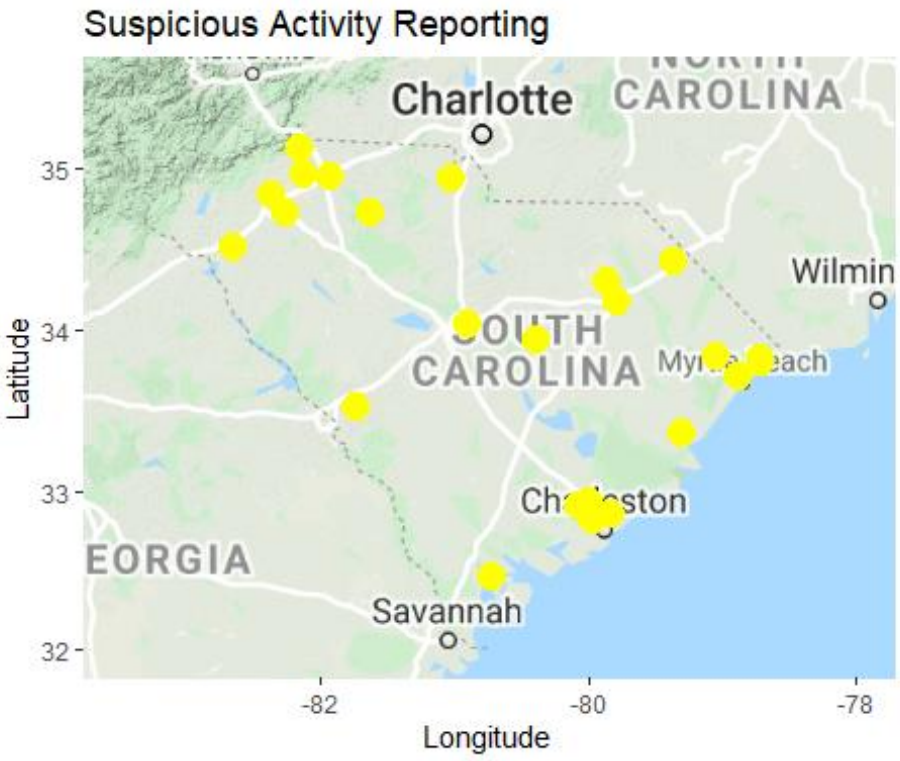



```
# Create SC subset of Justifiable Homicides
sc_spec_df <- filter(joint_off_df, STATE == "SC" & (OFF_CODE == "09C"))

# Build map
sc_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(32, 35.5)) +
  scale_x_continuous(limits=c(-83.5, -78)) +

  # Add data points to the map
  geom_point(data = sc_spec_df,
             aes(x=lng, y=lat, size=3),
             color="red",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Justifiable Homicides")
```

```
## Warning: Removed 1 rows containing missing values (geom_rect).
```



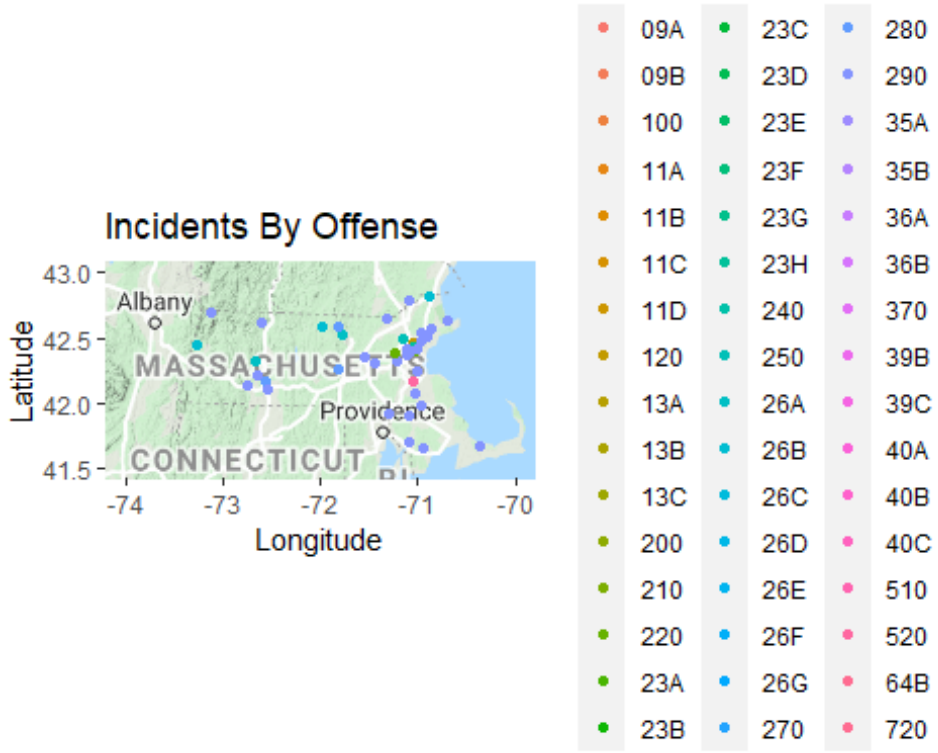

```
# Create SC subset of Suspicious Activity Reporting
sc_spec_df <- filter(joint_off_df, STATE == "SC" & (ACT_TYPE_OFFC == "7"))

# Build map
sc_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(32, 35.5)) +
  scale_x_continuous(limits=c(-83.5, -78)) +

  # Add data points to the map
  geom_point(data = sc_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")

## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 5 rows containing missing values (geom_point).
```



d) Massachusetts

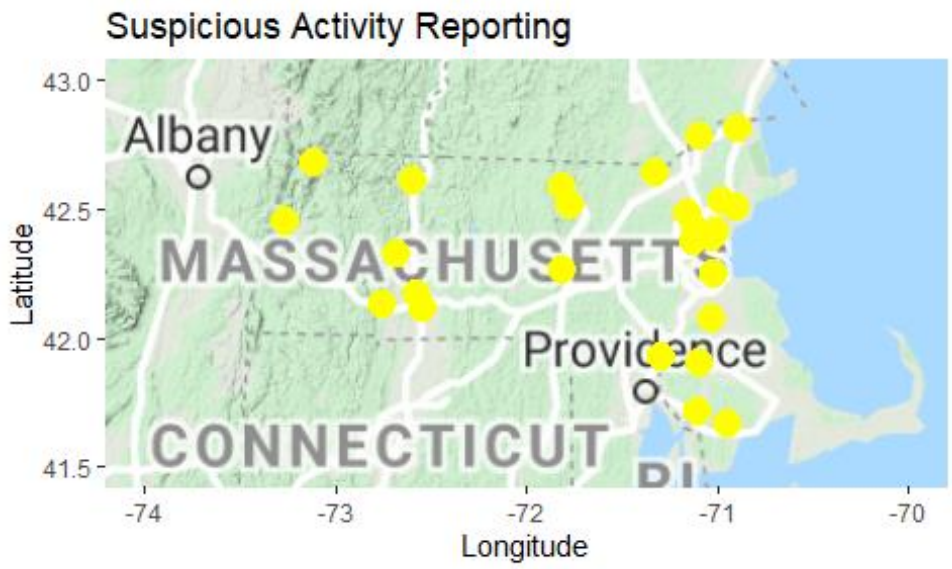
```
# Create MA subset of all offense
ma_df <- filter(joint_off_df, STATE == "MA")

# Use Massachusetts coordinates as center
ma_map <- ggmap(get_googlemap(
                  center = c(lon = -71.382439, lat = 42.407211),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
ma_map +

  # Re-scale to focus on MA rectangle
  scale_y_continuous(limits=c(41.5, 43)) +
  scale_x_continuous(limits=c(-74, -70)) +

  # Add data points to the map
  geom_point(data = ma_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())

## Warning: Removed 1 rows containing missing values (geom_rect).
```

```
## Warning: Removed 166538 rows containing missing values (geom_point).
```




```
# Create MA subset of Justifiable Homicides
ma_spec_df <- filter(joint_off_df, STATE == "MA" & (OFF_CODE == "09C"))
#summary(ma_spec_df)
# There are were no Justifiable Homicides reported in MA for 2016.


# Create MA subset of Suspicious Activity Reporting
ma_spec_df <- filter(joint_off_df, STATE == "MA" & (ACT_TYPE_OFFC == "7"))

# Build map
ma_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(41.5, 43)) +
  scale_x_continuous(limits=c(-74, -70)) +

  # Add data points to the map
  geom_point(data = ma_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")

## Warning: Removed 1 rows containing missing values (geom_rect).
```

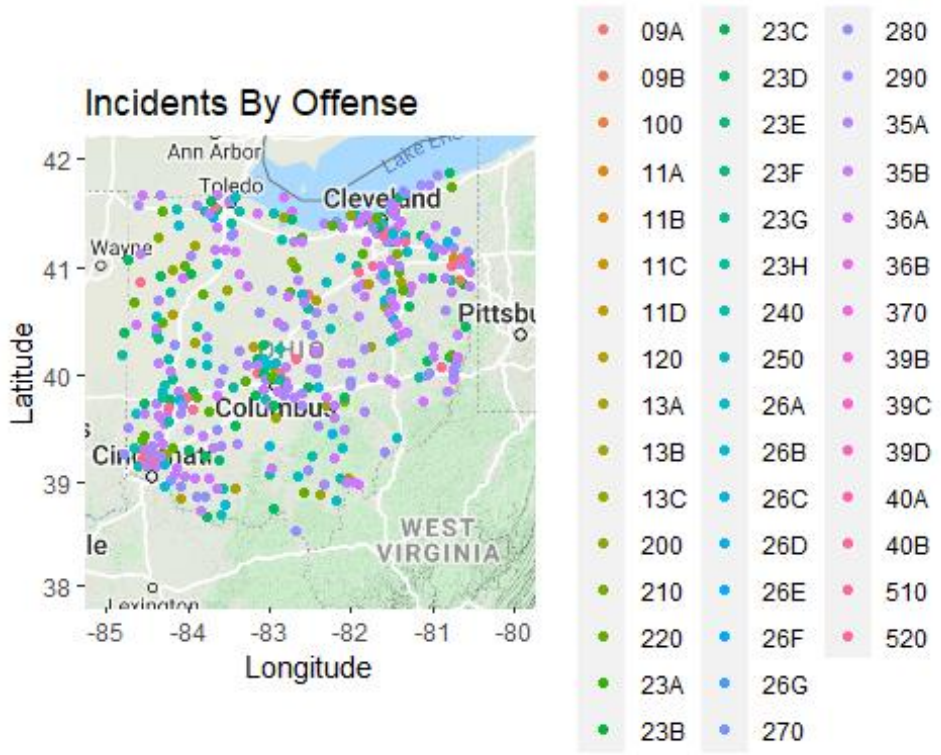
```
## Warning: Removed 113 rows containing missing values (geom_point).
```



e) Ohio

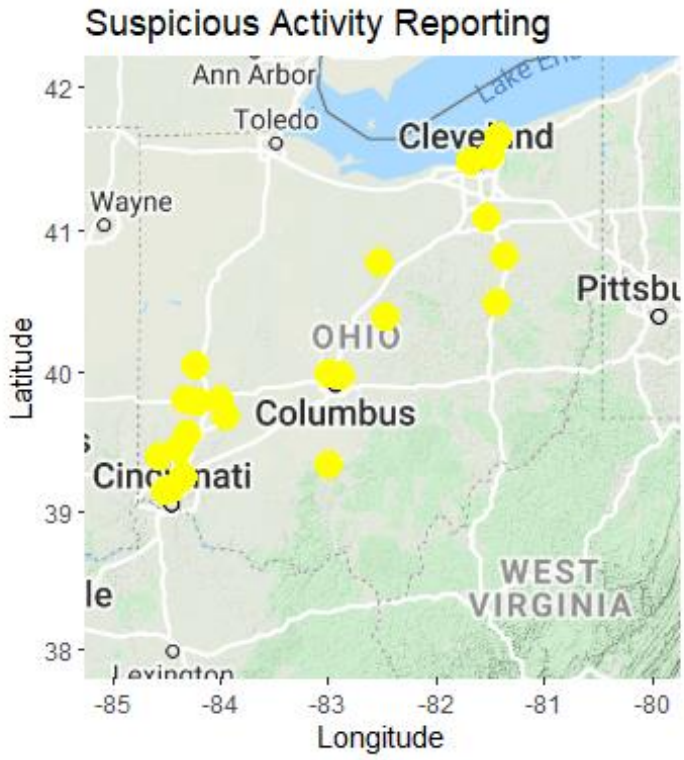
```
# Create OH subset of all offense
oh_df <- filter(joint_off_df, STATE == "OH")

# Use Ohio Carolina coordinates as center
oh_map <- ggmap(get_googlemap(
                  center = c(lon = -82.996216, lat = 40.367474),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
oh_map +

  # Re-scale to focus on OH rectangle
  scale_y_continuous(limits=c(38, 42)) +
  scale_x_continuous(limits=c(-85, -80)) +

  # Add data points to the map
  geom_point(data = oh_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())
```

```
## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 123581 rows containing missing values (geom_point).
```




```
# Create OH subset of Justifiable Homicides
oh_spec_df <- filter(joint_off_df, STATE == "OH" & (OFF_CODE == "09C"))
#summary(oh_spec_df)
# There are were no Justifiable Homicides reported in OH for 2016.


# Create OH subset of Suspicious Activity Reporting
oh_spec_df <- filter(joint_off_df, STATE == "OH" & (ACT_TYPE_OFFC == "7"))

# Build map
oh_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(38, 42)) +
  scale_x_continuous(limits=c(-85, -80)) +

  # Add data points to the map
  geom_point(data = oh_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")
```

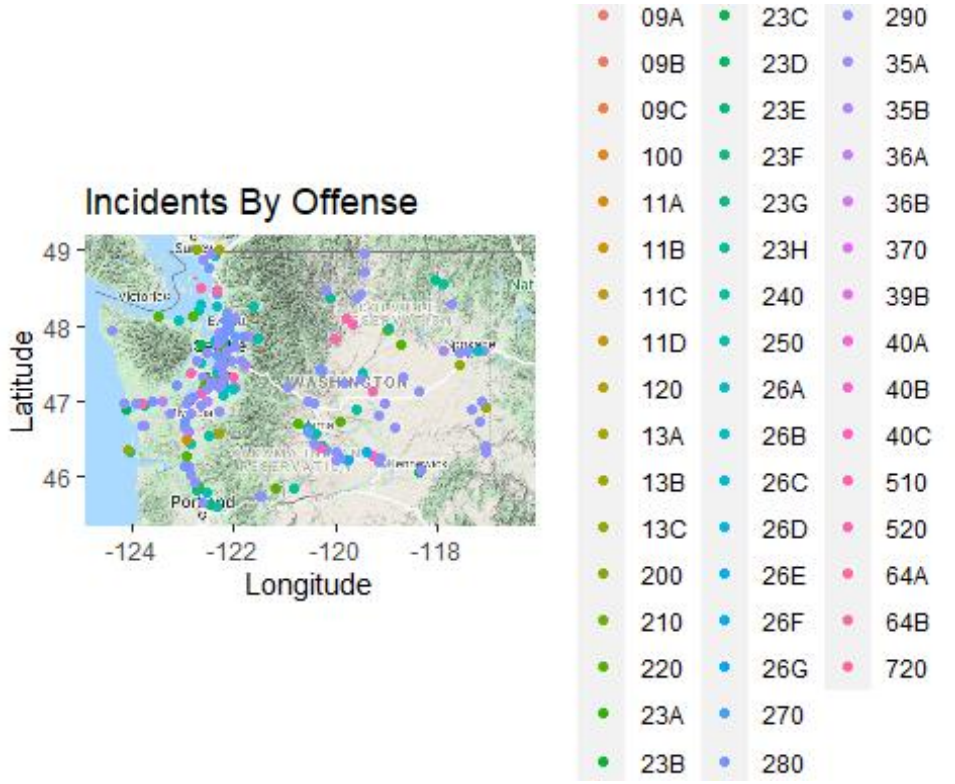
```
## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 14 rows containing missing values (geom_point).
```



f) Washington State

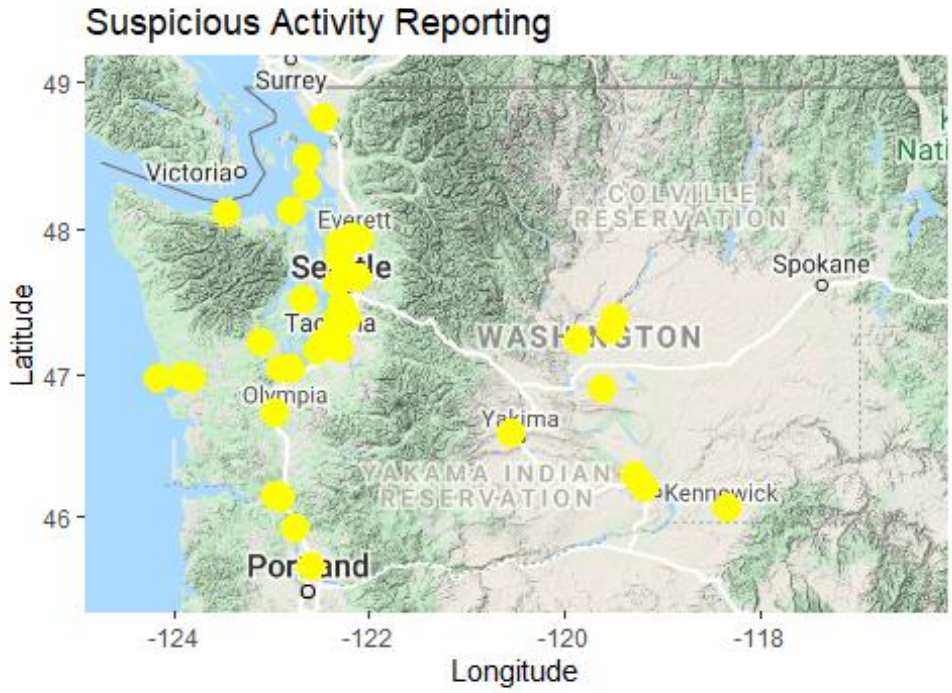
```
# Create WA subset of all offense
wa_df <- filter(joint_off_df, STATE == "WA")

# Use Washington coordinates as center
wa_map <- ggmap(get_googlemap(
                  center = c(lon = -120.740135, lat = 47.751076),
                  zoom = 6, scale = 2,
                  maptype = "terrain",
                  color="color"))

# Build map
wa_map +

  # Re-scale to focus on SC rectangle
  scale_y_continuous(limits=c(45.5, 49)) +
  scale_x_continuous(limits=c(-124.5, -116.5)) +

  # Add data points to the map
  geom_point(data = wa_df,
             aes(x=lng, y=lat, color = OFF_CODE)) +
```

```
  labs(x="Longitude", y="Latitude", title="Incidents By Offense") +
  theme(legend.title = element_blank())

## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 104325 rows containing missing values (geom_point).
```




```
# Create WA subset of Justifiable Homicides
wa_spec_df <- filter(joint_off_df, STATE == "WA" & (OFF_CODE == "09C"))

# Build map
wa_map +

  # Re-scale to focus on WA rectangle
  scale_y_continuous(limits=c(45.5, 49)) +
  scale_x_continuous(limits=c(-124.5, -116.5)) +

  # Add data points to the map
  geom_point(data = wa_spec_df,
             aes(x=lng, y=lat, size=3),
             color="red",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Justifiable Homicides")

## Warning: Removed 1 rows containing missing values (geom_rect).
```


```
# Create WA subset of Suspicious Activity Reporting
wa_spec_df <- filter(joint_off_df, STATE == "WA" & (ACT_TYPE_OFFC == "7"))

# Build map
wa_map +

  # Re-scale to focus on WA rectangle
  scale_y_continuous(limits=c(45.5, 49)) +
  scale_x_continuous(limits=c(-124.5, -116.5)) +

  # Add data points to the map
  geom_point(data = wa_spec_df,
             aes(x=lng, y=lat, size=3),
             color="yellow",
             show.legend = FALSE) +
  labs(x="Longitude", y="Latitude", title="Suspicious Activity Reporting")

## Warning: Removed 1 rows containing missing values (geom_rect).

## Warning: Removed 28 rows containing missing values (geom_point).
```


Suspicious Activity Reporting
Surrey
Victoria
Everett
Spokane
COLVILLE RESERVATION
WASHINGTON
Olympia
Yakima
YAKAMA INDIAN RESERVATION
Kennewick
Portland
49
48
47
46
-124
-122
-120
-118
Latitude
Longitude