EDA

Ryan M. Allen

January 20, 2020

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.3
```

```
## -- Attaching packages ------------------------------------ tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.0       v purrr   0.2.5
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.1       v stringr 1.4.0
## v readr   1.1.1       v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.2
```

```
## Warning: package 'forcats' was built under R version 3.5.2
```

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

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(scales)
```

```
## Warning: package 'scales' was built under R version 3.5.3
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(agricolae)
library(userfriendlyscience)
library(varhandle)
```

```
## Warning: package 'varhandle' was built under R version 3.5.3
```

```
library(sf)
```

```
## Warning: package 'sf' was built under R version 3.5.3
```

```
## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
```

```
library(rnaturalearth)
```

```
## Warning: package 'rnaturalearth' was built under R version 3.5.3
```

```
library(rnaturalearthdata)
```

```
## Warning: package 'rnaturalearthdata' was built under R version 3.5.3
```

```
library(rgeos)
```

```
## Warning: package 'rgeos' was built under R version 3.5.3
```

```
## Loading required package: sp
```

```
## rgeos version: 0.5-2, (SVN revision 621)
##  GEOS runtime version: 3.6.1-CAPI-1.10.1
##  Linking to sp version: 1.3-2
##  Polygon checking: TRUE
```

```
library(rworldmap)  # getMap()
```

```
## Warning: package 'rworldmap' was built under R version 3.5.3
```

```
## ### Welcome to rworldmap ###
```

```
## For a short introduction type :   vignette('rworldmap')
```

```
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

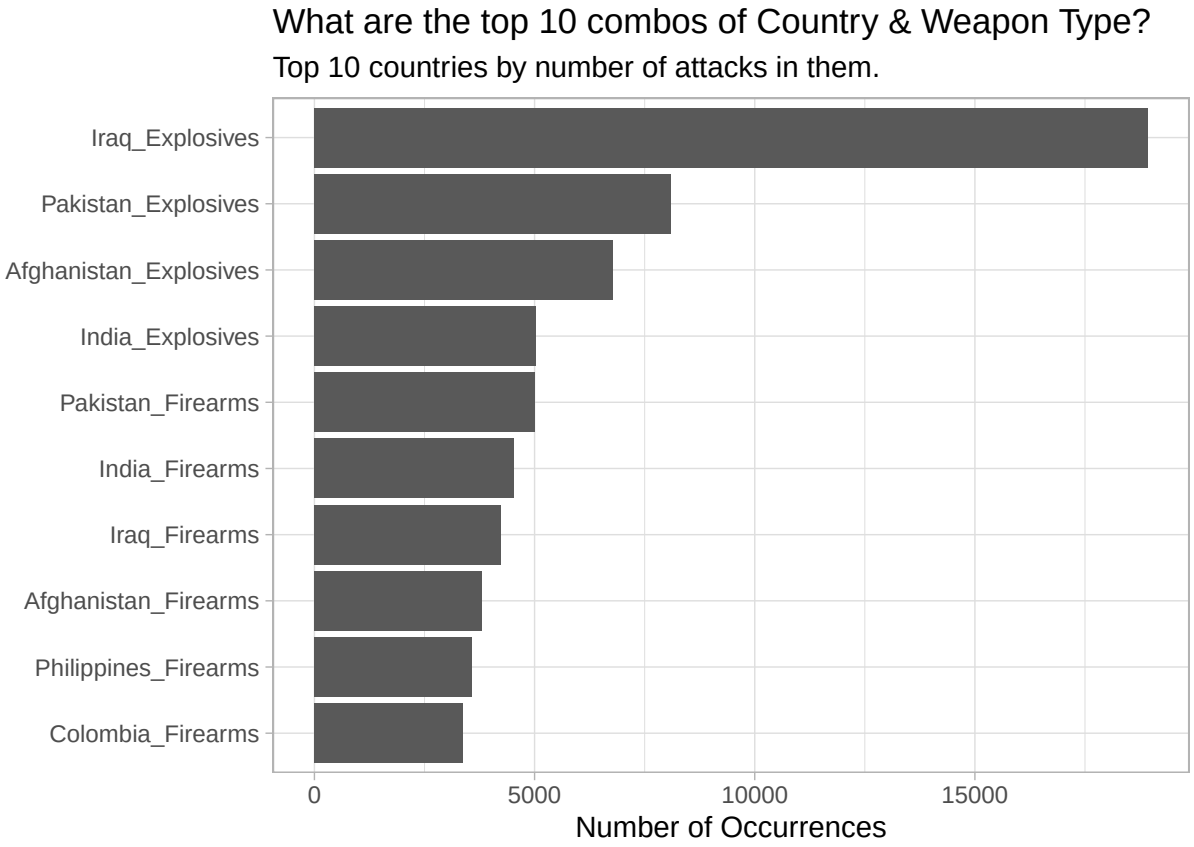
```
## The following object is masked from 'package:ggplot2':
##
##     ggsave
```

```
theme_set(theme_light())

data <- read.csv("C:/Users/Ryan Allen/Documents/Regis/Classes/Practicum_I/Data/globalterrorismdb_0718di

# dim(data)
# str(data)

# data %>%
#     select(country_txt, weaptype1_txt) %>%
#     filter(country_txt == 'United States') %>%
#     add_count(country_txt, weaptype1_txt) %>%
#     unique() %>%
#       mutate(weaptype1_txt = fct_reorder(weaptype1_txt, n, desc)) %>%
#    ggplot(aes(x = weaptype1_txt, y = n)) +
#                     geom_bar(stat = "identity")
```

```r
# top 10 countries for terrorist attacks to occur
toptenC <- data %>% select(country_txt) %>%
    add_count(country_txt) %>%
    unique() %>%
    arrange(desc(n)) %>%
    top_n(n, n=10)

# Top 10 Weapon Types
toptenW <- data %>% select(weaptype1_txt) %>%
    add_count(weaptype1_txt) %>%
    unique() %>%
    arrange(desc(n)) %>%
    top_n(n, n=10)

toptenCombo <- data %>%
    select(country_txt, weaptype1_txt) %>%
    add_count(country_txt,weaptype1_txt) %>%
    rename(Number_Attacks = n) %>%
    unique() %>%
    arrange(desc(Number_Attacks)) %>%
    top_n(Number_Attacks, n =10)
```

What Weapons used and Countries attacked combinations were most prevalent?

```r
#Top 10 Combinations of Country
toptenCombo %>%
    unite(Country_Weapon, country_txt,weaptype1_txt) %>%
    mutate(Country_Weapon = fct_reorder(Country_Weapon, Number_Attacks)) %>%
    ggplot(aes(x = Country_Weapon, y = Number_Attacks)) +
                    geom_bar(stat = "identity") +
    coord_flip() +
      labs(title = "What are the top 10 combos of Country & Weapon Type?",
       subtitle = "Top 10 countries by number of attacks in them.",
       x= "",
       y = "Number of Occurrences")
```




```
#Distributions of casualties

#Difference in Medians
howell <- data %>%
    select(country_txt,  nkill) %>%
    group_by(country_txt) %>%
    filter(country_txt %in% toptenC$country_txt) %>%
    mutate(nkill = as.numeric(nkill)) %>%
    drop_na()

oneway(as.numeric(howell$nkill), factor(howell$country_txt), posthoc = "games-howell")
```

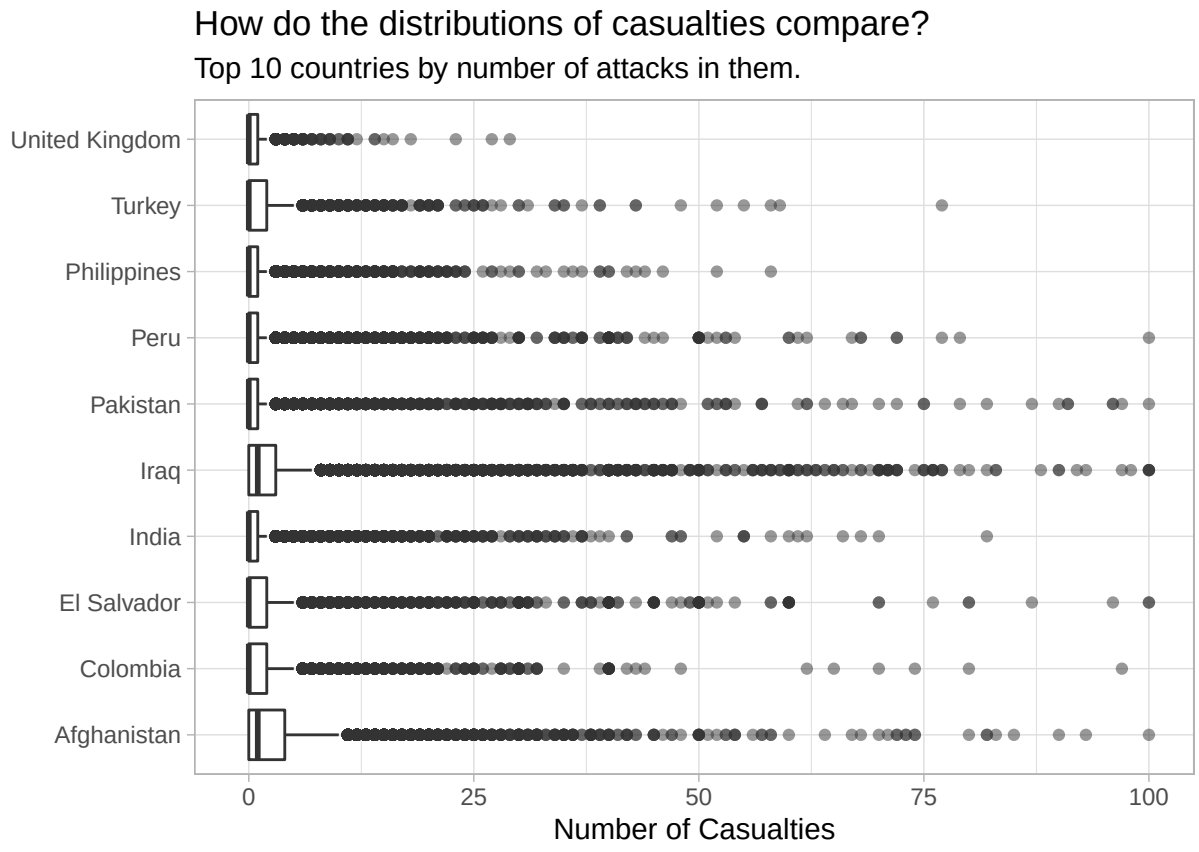
```
## ### Oneway Anova for y=nkill and x=country_txt (groups: Afghanistan, Colombia, El Salvador, India, I
##
## Omega squared: 95% CI = [.01; .01], point estimate = .01
## Eta Squared: 95% CI = [.01; .01], point estimate = .01
##
##                                         SS    Df      MS     F     p
## Between groups (error + effect)   67200.13     9 7466.68 82.92 <.001
## Within groups (error only)      8580457.68 95285   90.05
##
##
## ### Post hoc test: games-howell
##
##                             diff ci.lo ci.hi     t       df     p
## Colombia-Afghanistan       -1.31 -1.60 -1.03 14.58 19204.39 <.001
```

```
## El Salvador-Afghanistan    -0.13 -0.69  0.44  0.71  5138.23  .999
## India-Afghanistan          -1.54 -1.78 -1.29 19.98 21545.26 <.001
## Iraq-Afghanistan            0.10 -0.28  0.48  0.83 35595.42  .998
## Pakistan-Afghanistan       -1.50 -1.75 -1.25 19.05 23144.44 <.001
## Peru-Afghanistan           -0.85 -1.19 -0.50  7.78 11290.80 <.001
## Philippines-Afghanistan    -1.76 -2.01 -1.51 22.07 19047.69 <.001
## Turkey-Afghanistan         -1.53 -1.83 -1.23 16.36 11569.30 <.001
## United Kingdom-Afghanistan -2.51 -2.78 -2.24 29.65 15832.54 <.001
## El Salvador-Colombia        1.19  0.62  1.75  6.65  5117.01 <.001
## India-Colombia             -0.23 -0.47  0.02  2.93 14657.78  .098
## Iraq-Colombia               1.41  1.03  1.80 11.65 31621.99 <.001
## Pakistan-Colombia          -0.19 -0.43  0.06  2.36 15818.78  .350
## Peru-Colombia               0.47  0.12  0.81  4.30 10627.90  .001
## Philippines-Colombia       -0.44 -0.70 -0.19  5.59 13976.18 <.001
## Turkey-Colombia            -0.22 -0.51  0.08  2.32 10361.91  .375
## United Kingdom-Colombia    -1.20 -1.47 -0.93 14.16 12813.17 <.001
## India-El Salvador          -1.41 -1.96 -0.87  8.20  4478.28 <.001
## Iraq-El Salvador            0.23 -0.39  0.85  1.16  7367.56  .979
## Pakistan-El Salvador       -1.37 -1.92 -0.83  7.94  4556.74 <.001
## Peru-El Salvador           -0.72 -1.32 -0.12  3.82  6102.94  .005
## Philippines-El Salvador    -1.63 -2.18 -1.08  9.41  4593.69 <.001
## Turkey-El Salvador         -1.40 -1.97 -0.83  7.79  5236.93 <.001
## United Kingdom-El Salvador -2.39 -2.94 -1.83 13.58  4825.19 <.001
## Iraq-India                  1.64  1.28  1.99 14.64 31067.61 <.001
## Pakistan-India              0.04 -0.16  0.24  0.63 25838.27 1.000
## Peru-India                  0.69  0.38  1.00  7.06  8175.26 <.001
## Philippines-India          -0.22 -0.42 -0.02  3.41 16009.73  .023
## Turkey-India                0.01 -0.25  0.26  0.10  7701.53 1.000
## United Kingdom-India       -0.97 -1.20 -0.75 13.79 11204.62 <.001
## Pakistan-Iraq              -1.60 -1.96 -1.24 14.14 32220.45 <.001
## Peru-Iraq                  -0.95 -1.38 -0.52  6.97 21500.35 <.001
## Philippines-Iraq           -1.86 -2.22 -1.50 16.33 30282.48 <.001
## Turkey-Iraq                -1.63 -2.02 -1.24 13.16 23495.16 <.001
## United Kingdom-Iraq        -2.61 -2.99 -2.24 22.25 28452.55 <.001
## Peru-Pakistan               0.65  0.34  0.97  6.57  8605.35 <.001
## Philippines-Pakistan       -0.26 -0.47 -0.05  3.90 17692.99  .004
## Turkey-Pakistan            -0.03 -0.29  0.23  0.38  8290.81 1.000
## United Kingdom-Pakistan    -1.01 -1.24 -0.78 14.00 12274.38 <.001
## Philippines-Peru           -0.91 -1.23 -0.60  9.10  8534.95 <.001
## Turkey-Peru                -0.68 -1.04 -0.33  6.14  9492.43 <.001
## United Kingdom-Peru        -1.67 -2.00 -1.34 15.98  9133.62 <.001
## Turkey-Philippines          0.23 -0.04  0.49  2.73  8018.20  .162
## United Kingdom-Philippines -0.75 -0.99 -0.52 10.27 10807.43 <.001
## United Kingdom-Turkey      -0.98 -1.26 -0.70 11.12  8462.33 <.001
```

```
data %>%
    select(country_txt, weaptype1_txt, nkill) %>%
    group_by(country_txt, weaptype1_txt) %>%
    filter(country_txt %in% toptenC$country_txt) %>%
    drop_na() %>%
    arrange(desc(nkill)) %>%
    mutate(nkill - as.numeric(nkill)) %>%
    ggplot(aes(x= country_txt, y = nkill)) +
    geom_boxplot(outlier.alpha = .5)+
```

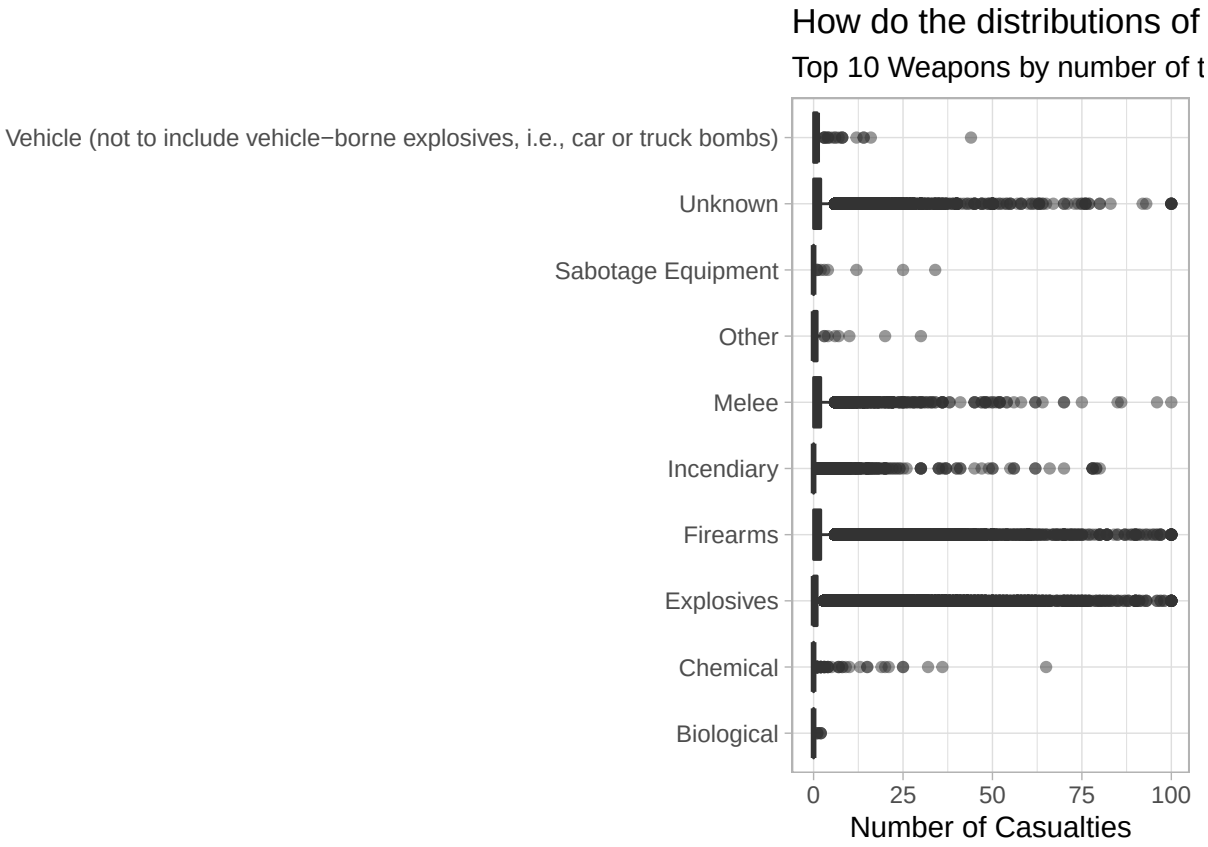
```
    ylim(-1, 100) +
    coord_flip() +
     labs(title = "How do the distributions of casualties compare?",
       subtitle = "Top 10 countries by number of attacks in them.",
       x= "",
       y = "Number of Casualties")
```

```
## Warning: Removed 60 rows containing non-finite values (stat_boxplot).
```



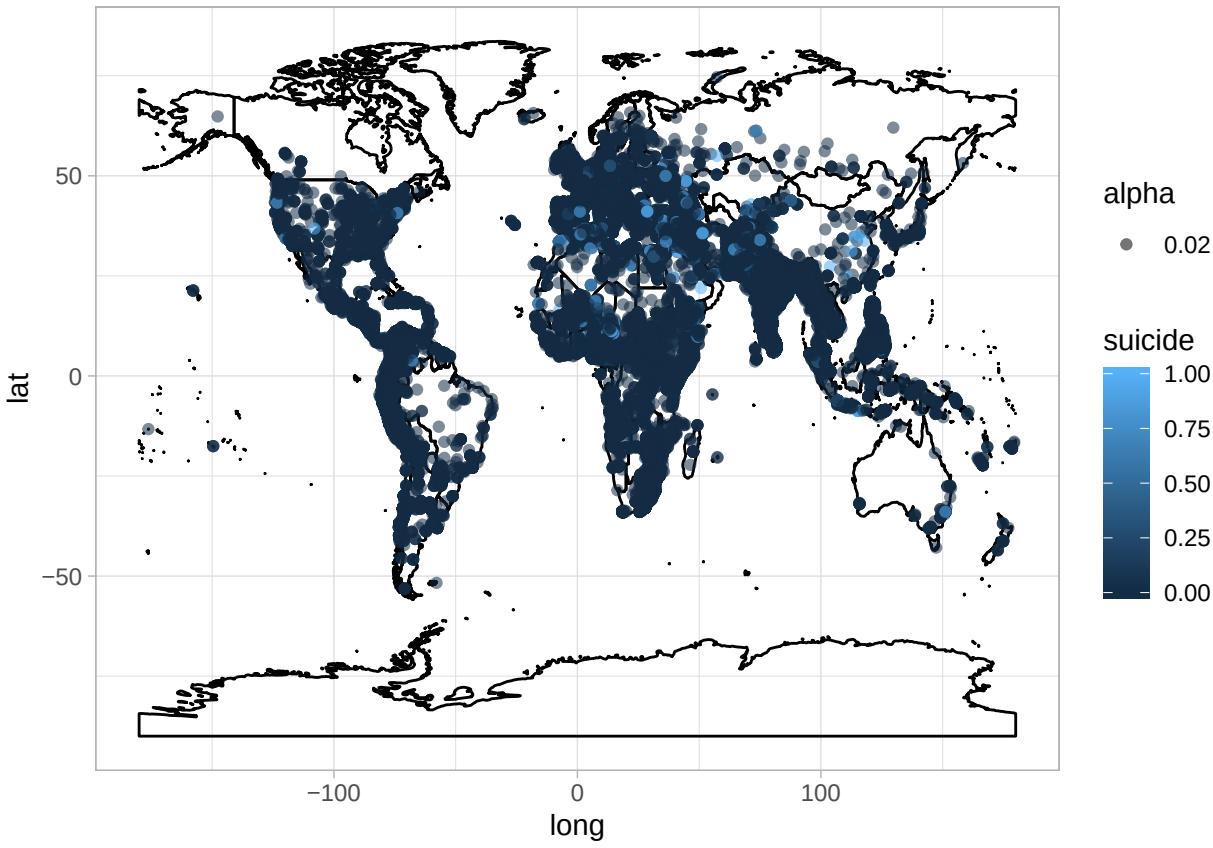

```
# Distributions by Weapon Type
data %>%
    select(country_txt, weaptype1_txt, nkill) %>%
    filter(weaptype1_txt %in% toptenW$weaptype1_txt) %>%
    drop_na() %>%
    arrange(desc(nkill)) %>%
    mutate(nkill - as.numeric(nkill)) %>%
    ggplot(aes(x= weaptype1_txt, y = nkill)) +
    geom_boxplot(outlier.alpha = .5)+
    ylim(-1, 100) +
    coord_flip() +
     labs(title = "How do the distributions of casualties compare?",
       subtitle = "Top 10 Weapons by number of times used.",
       x= "",
       y = "Number of Casualties")
```

```
## Warning: Removed 193 rows containing non-finite values (stat_boxplot).
```




```
mapping <- data %>% select(latitude,longitude, suicide) %>%
    drop_na() %>%
    filter(longitude> -180)

world <- getMap(resolution = "low")

ggplot() +
    geom_polygon(data = world, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = mapping, aes(x = longitude, y = latitude, col = suicide, alpha = .02))
```

```
## Regions defined for each Polygons
```




```
# ggplot() +
#     geom_polygon(data = world, aes(x = long, y = lat, group = group),
#                  fill = NA, colour = "black") +
#     geom_density2d(data = mapping, aes(x = longitude, y = latitude, col = suicide))

par(mfrow=c(2,4))

#North American Regional Map
regionsNA <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>% filter (region_
NARegion <- unique(regionsNA$country_txt)
world_NA <- world[world@data$ADMIN %in% NARegion, ]

NorthAmerica <- ggplot() +
    geom_polygon(data = world_NA, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsNA,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
labs(title = "North American Region Terrorist Attack Locations",
       subtitle = "<1% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none')
```

```
## Regions defined for each Polygons
# Central America Map
regionsCAC <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt == 'Central America & Caribbean') %>% filter(longitude>-125)
world_CAC <- world[world@data$ADMIN %in% unique(regionsCAC$country_txt), ]

CentralAmerica <- ggplot() +
    geom_polygon(data = world_CAC, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsCAC,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "Central America Terrorist Attack Locations",
       subtitle = "<1% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none')
```

```
## Regions defined for each Polygons
# Asia
regionsAsia <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt %in% c('Southeast Asia', 'East Asia', 'Central Asia', 'South Asia')) %>% filter(l
world_Asia <- world[world@data$ADMIN %in% unique(regionsAsia$country_txt), ]

Asia <- ggplot() +
    geom_polygon(data = world_Asia, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsAsia,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "Asia Terrorist Attack Locations",
       subtitle = "3.4% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none')
```

```
## Regions defined for each Polygons
# Middle East
regionsME <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt == 'Middle East & North Africa') %>% filter(longitude < 70)
world_ME <- world[world@data$ADMIN %in% unique(regionsME$country_txt), ]

MiddleEast <- ggplot() +
    geom_polygon(data = world_ME, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsME,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
```

```
    labs(title = "Middle East Terrorist Attack Locations",
       subtitle = "7.6% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none')
```

```
## Regions defined for each Polygons
```

```
# Europe
regionsEU <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt %in% c('Eastern Europe', 'Western Europe'))
world_EU <- world[world@data$ADMIN %in% unique(regionsEU$country_txt), ]

Europe <- ggplot() +
    geom_polygon(data = world_EU, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsEU,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "Europe Terrorist Attack Locations",
       subtitle = "<1% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none') +
    xlim(-50, 160) +
    ylim(20, 80)
```

```
## Regions defined for each Polygons
```

```
# Sub-Saharan Africa
regionsSA <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt == 'Sub-Saharan Africa')
world_SA <- world[world@data$ADMIN %in% unique(regionsSA$country_txt), ]

SubSaharahAfrica <- ggplot() +
    geom_polygon(data = world_SA, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsSA,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "Sub-Saharan Africa Terrorist Attack Locations",
       subtitle = "4.2% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none') +
    ylim(-38, 25)
```

```
## Regions defined for each Polygons
```

```
# South America
regionsSAm <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
```

```
    filter(region_txt == 'South America')
world_SAm <- world[world@data$ADMIN %in% unique(regionsSAm$country_txt), ]

SouthAmerica <- ggplot() +
    geom_polygon(data = world_SAm, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsSAm,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "South America Terrorist Attack Locations",
       subtitle = "<1% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none')
```

```
## Regions defined for each Polygons
```

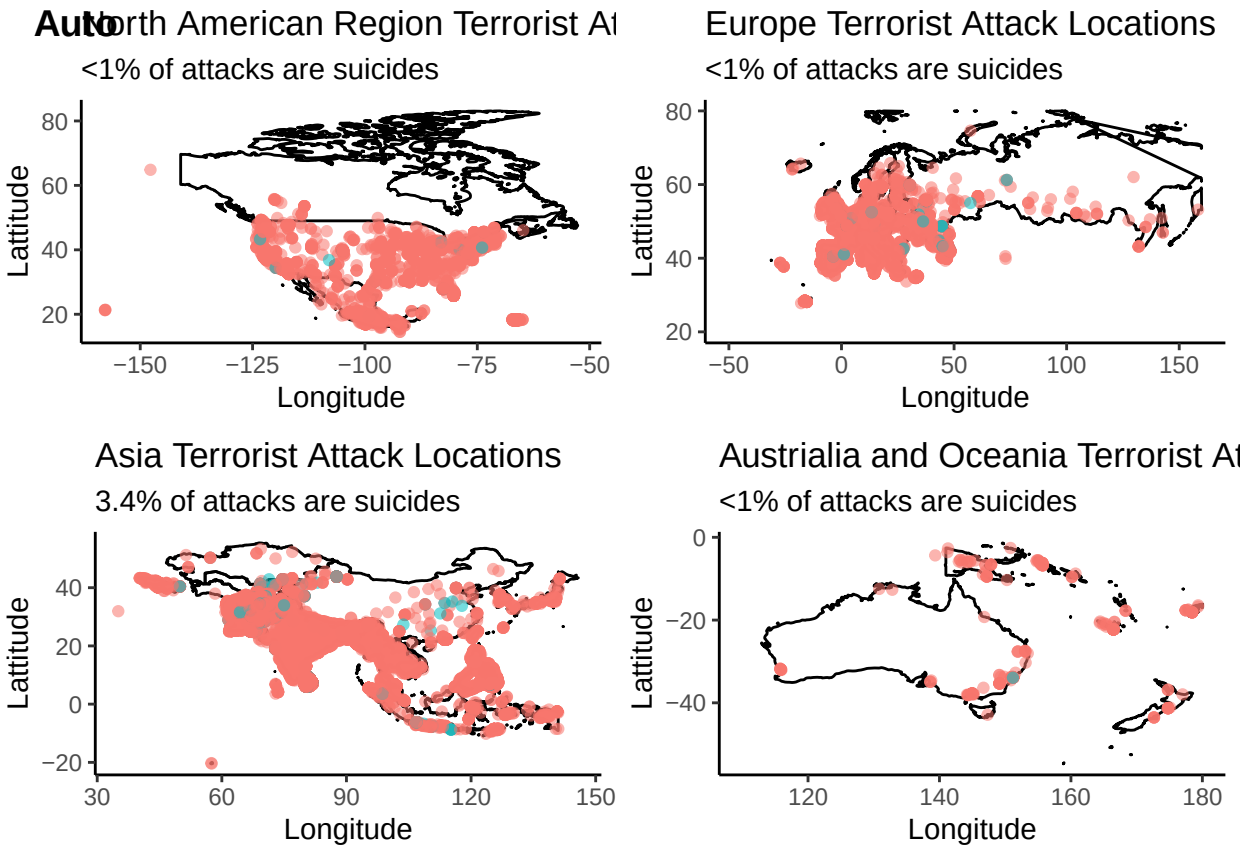
```
# Australia and Oceania
regionsAus <- data %>% select(country_txt, region_txt, latitude, longitude, suicide) %>%
    filter(region_txt == 'Australasia & Oceania')
world_Aus <- world[world@data$ADMIN %in% unique(regionsAus$country_txt), ]

Australia <- ggplot() +
    geom_polygon(data = world_Aus, aes(x = long, y = lat, group = group),
                 fill = NA, colour = "black") +
    geom_point(data = regionsAus,
               aes(x = longitude, y = latitude, col = factor(suicide),
                   alpha = .02), show.legend = F) +
    theme_classic() +  # Remove ugly grey background
    labs(title = "Austrialia and Oceania Terrorist Attack Locations",
       subtitle = "<1% of attacks are suicides",
       x= "Longitude",
       y = "Lattitude") +
    guides(colour=guide_legend(title="Suicide Attack")) +
    scale_alpha(guide = 'none') +
    xlim(110, 180)
```

```
## Regions defined for each Polygons
```

```
plot_grid(
NorthAmerica,
Europe,
Asia,
Australia,
labels = "Auto", nrow =2, ncol = 2)
```

```
## Warning: Removed 184 rows containing missing values (geom_point).
```

```
## Warning: Removed 10 rows containing missing values (geom_point).
```




```
# Deadliest Weapons
data %>%
    select(weaptype1_txt, nkill) %>%
    drop_na() %>%
    group_by(weaptype1_txt) %>%
    summarise(Total= sum(nkill)) %>%
    arrange(desc(Total))
```

```
## # A tibble: 12 x 2
##    weaptype1_txt                                                  Total
##    <fct>                                                          <int>
##  1 Firearms                                                      174894
##  2 Explosives                                                    174277
##  3 Unknown                                                        42620
##  4 Melee                                                          10575
##  5 Incendiary                                                      5476
##  6 Vehicle (not to include vehicle-borne explosives, i.e., car or t~  3184
##  7 Chemical                                                         624
##  8 Other                                                            123
##  9 Sabotage Equipment                                                83
## 10 Biological                                                         9
## 11 Radiological                                                       2
## 12 Fake Weapons                                                       1
```

```
data %>%
    select(country_txt, weaptype1_txt, nkill) %>% filter(nkill ==1384)
```

```
##     country_txt
## 1 United States
##                                                                 weaptype1_txt
## 1 Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)
##   nkill
## 1  1384
```