

R Data Visualizations

Sam Heidler

```
# set working directory
setwd("C:/Users/samue/OneDrive/Desktop/STAT3355")
```

```
# load libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

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

```
library(RColorBrewer)
library(scales)
```

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

```
# load Dallas crime data
dallas <- read.csv("crime.csv")

str(dallas)
```

```
## 'data.frame':   61656 obs. of  18 variables:
##  $ IncidentNum      : chr  "110555-2019" "152471-2019" "235407-2021" "154267-2018" ...
##  $ Arrest.Year       : int   2019 2019 2021 2018 2018 2019 2021 2021 2020 2019 ...
##  $ Arrest.Date       : chr   "6/2/19" "7/28/19" "12/31/21" "7/15/18" ...
##  $ Arrest.Time       : chr   "4:45:00 PM" "8:40:00 PM" "9:11:00 PM" "12:10:00 AM" ...
##  $ Arrest.Address    : chr   "500 E JEFFERSON BLVD" "4800 MAPLE AVE" "13165 N CENTRAL EXPY" ...
##  $ Arrest.Zipcode    : int   75203 75235 75243 75201 75220 75226 75228 75208 75219 75243 ...
##  $ Latitude         : num   32.8 32.8 32.8 32.8 32.8 ...
##  $ Longitude        : chr   "-96.7967" "-96.7967" "-96.7967" "-96.7967" ...
##  $ Arrest.City      : chr   "Dallas" "Dallas" "Dallas" "Dallas" ...
##  $ Arrest.State     : chr   "Texas" "Texas" "Texas" "Texas" ...
##  $ Arrest.Day.of.The.Week : chr   "Sun" "Sun" "Fri" "Sun" ...
##  $ Arrest.Location   : chr   "Other" "Other" "Other" "Other" ...
##  $ Arrest.Weapon    : chr   "Unarmed" "Unarmed" "Unarmed" "Unarmed" ...
```

```
## $ Arrestee.Age.At.Arrest.Time: int 40 36 28 35 31 44 35 48 51 39 ...
## $ Arrestee.Race                : chr "Hispanic or Latino" "White" "Black" "Middle Eastern" ...
## $ Arrestee.Sex                 : chr "Male" "Female" "Male" "Male" ...
## $ Drug.Related                 : chr "No" "No" "No" "No" ...
## $ Drug.Type                    : chr "" "" "" "" ...
```

```
View(dallas)
```

```
# create value mapping to change all random weapon labels in Dallas crime
# to categorical data
```

```
value_mapping <- list(
  HANDS = c("Hands/Feet", "Strangulation"),
  KNIFE = c("Knife - Butcher", "Knife - Pocket", "Lethal Cutting Instrument", "Stabbing Instrument", "K",
  FIREARM = c("Handgun", "Shotgun", "Other Firearm", "Firearm (Type Not Stated)", "Gun"),
  UNARMED = c("Unarmed", "None"),
  OTHER = c("Club", "Drugs", "THREATS", "Rock", "Poison", "Other",
            "Vehicle", "Burn", "Explosives", "33", "Missile/Arrow"))
```

```
dallas <- dallas %>%
  mutate(Weapon_new = case_when(
    Arrest.Weapon %in% value_mapping$HANDS ~ "HANDS",
    Arrest.Weapon %in% value_mapping$KNIFE ~ "KNIFE",
    Arrest.Weapon %in% value_mapping$FIREARM ~ "FIREARM",
    Arrest.Weapon %in% value_mapping$UNARMED ~ "UNARMED",
    Arrest.Weapon %in% value_mapping$OTHER ~ "OTHER",
    TRUE ~ as.character(Arrest.Weapon)
  ))
```

```
# clean drug data
```

```
dallas <- dallas %>%
  mutate(Drug.Type = ifelse(Drug.Type %in% c("Processed Marijuana", "Cultivated Marijuana"),
                            "Marijuana", Drug.Type))
```

```
# factor Dallas crime data
```

```
dallas$Weapon_new <- factor(dallas$Weapon_new, levels = c("HANDS", "KNIFE", "FIREARM", "UNARMED", "OTHER", "THREATS", "Rock", "Poison", "Vehicle", "Burn", "Explosives", "33", "Missile/Arrow"))
dallas$Arrestee.Race <- as.factor(dallas$Arrestee.Race)
dallas$Drug.Type <- as.factor(dallas$Drug.Type)
drugs <- subset(dallas, Drug.Type != "")
```

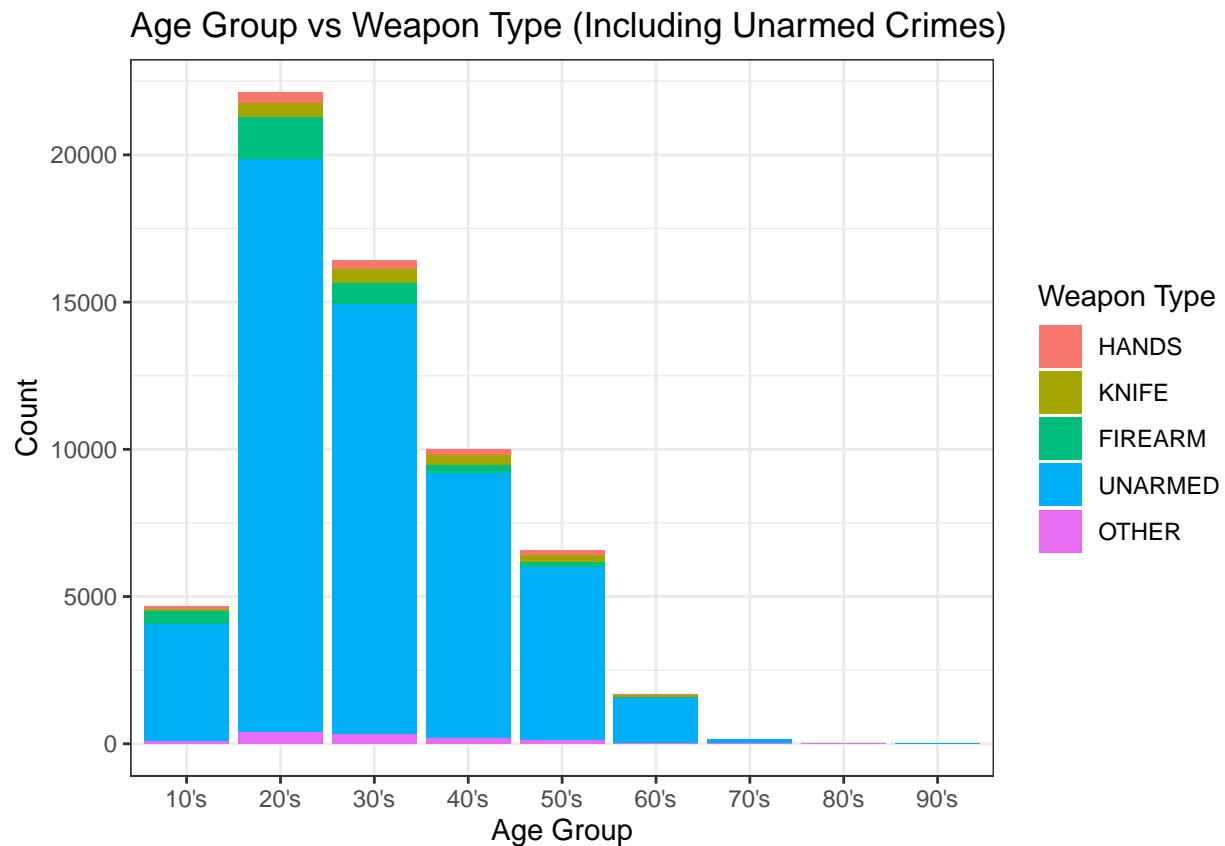
```
# use cut() to create age group data
```

```
dallas <- dallas %>%
  mutate(Age.Group = cut(Arrestee.Age.At.Arrest.Time, breaks = seq(0, 100, 10),
                        labels = c("<10", "10's", "20's", "30's", "40's", "50's", "60's", "70's", "80's", "90's", "100's"))
```

```
# bar plot of age group vs what weapon type was most prevalent
df <- dallas %>%
  group_by(Age.Group, Weapon_new) %>%
  summarize(count = n())
```

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

```
ggplot(df, aes(x = Age.Group, y = count, fill = Weapon_new)) +
  geom_bar(stat = "identity") +
  labs(title = "Age Group vs Weapon Type (Including Unarmed Crimes)",
       x = "Age Group",
       y = "Count", fill = "Weapon Type") +
  theme_bw()
```



```
# Create a new dataframe with Age.Group, Weapon_new, and Count
dallas_agg <- dallas %>%
  filter(Weapon_new != "UNARMED") %>%
  group_by(Age.Group, Weapon_new) %>%
  summarise(Count = n())
```

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

```
# Normalize the count values by age group
```

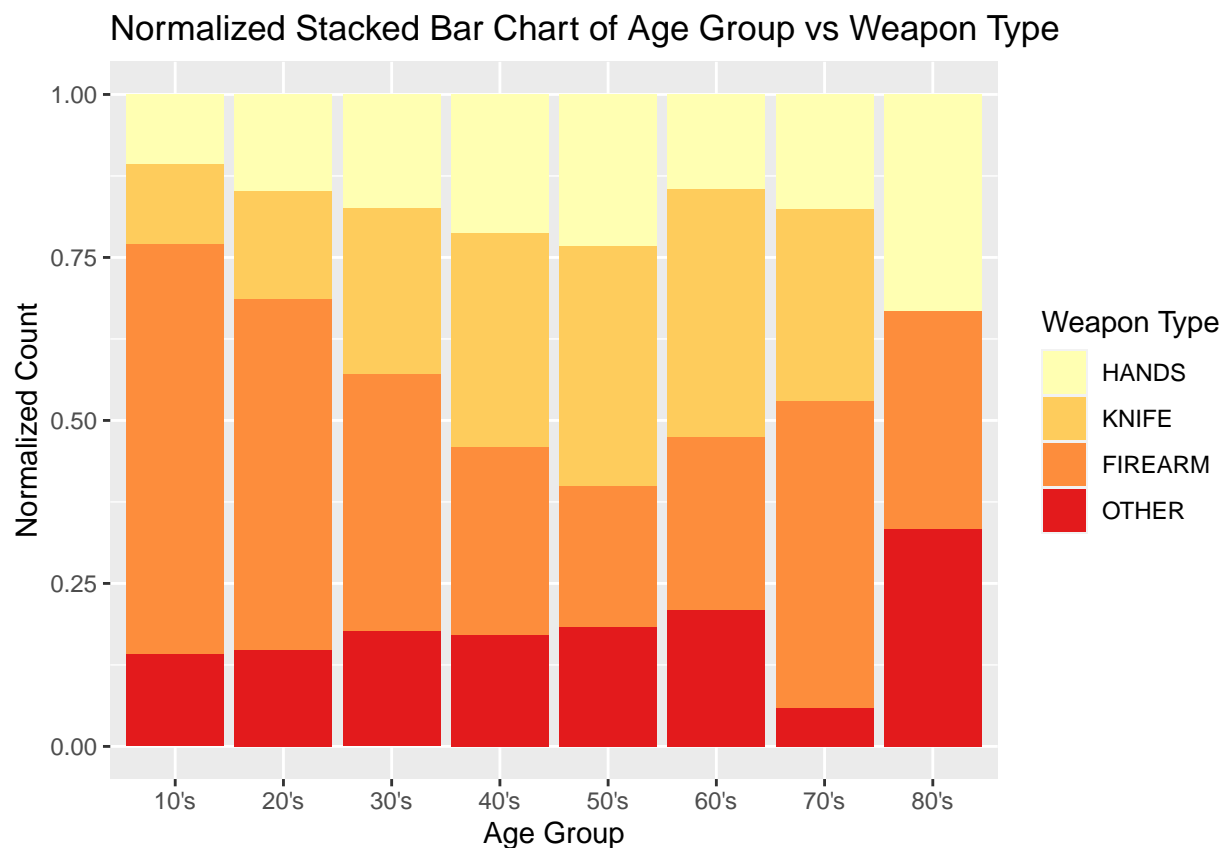
```
dallas_agg <- dallas_agg %>%
  group_by(Age.Group) %>%
  mutate(Norm_Count = Count / sum(Count))
```

```
# Subset data
```

```
dallas_agg <- dallas %>%
  filter(!is.na(Age.Group), Weapon_new != "UNARMED") %>%
  group_by(Age.Group, Weapon_new) %>%
  summarise(Count = n()) %>%
  mutate(Total_Count = sum(Count),
         Norm_Count = Count/Total_Count)
```

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

```
ggplot(dallas_agg, aes(x = Age.Group, y = Norm_Count, fill = Weapon_new)) +
  geom_bar(stat = "identity") +
  scale_fill_brewer(palette = "YlOrRd") +
  labs(title = "Normalized Stacked Bar Chart of Age Group vs Weapon Type",
       x = "Age Group",
       y = "Normalized Count", fill = "Weapon Type")
```



```

# pie charts that display breakdown of weapon type by race for Dallas arrest data
pie_charts <- list()

# Loop through each level of Arrestee.Race, excluding "NH" and "Unknown"
for (race in unique(dallas$Arrestee.Race[dallas$Arrestee.Race != "NH" & dallas$Arrestee.Race != "Unknown"])) {
  # Subset the data for the current race, excluding "UNARMED" data
  df <- dallas[dallas$Arrestee.Race == race & dallas$Weapon_new != "UNARMED", ]

  # Create a summary by Weapon_new
  df_summary <- aggregate(df$Weapon_new, by = list(df$Weapon_new), FUN = length)
  colnames(df_summary) <- c("Weapon_new", "Count")

  # Calculate the percentage of each category
  df_summary$Percent <- round(df_summary$Count/sum(df_summary$Count)*100, 1)

  # Create the pie chart plot for the current race
  pie_chart <- ggplot(df_summary, aes(x = "", y = Count, fill = Weapon_new)) +
    geom_bar(stat = "identity", width = 1) +
    coord_polar("y", start = 0) +
    labs(title = paste("Weapon Type by Race: ", race),
         x = NULL, y = NULL, fill = "Weapon Type") +
    scale_fill_brewer(palette = "Inferno") +
    theme_void() +
    # Add percentage labels to the pie chart
    geom_text(aes(label = paste0(Percent, "%")), position = position_stack(vjust = 0.5))

  # Add the pie chart plot to the list
  pie_charts[[race]] <- pie_chart
}

```

```

## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno
## Warning in pal_name(palette, type): Unknown palette Inferno

```

```

# Print the list of pie chart plots
pie_charts

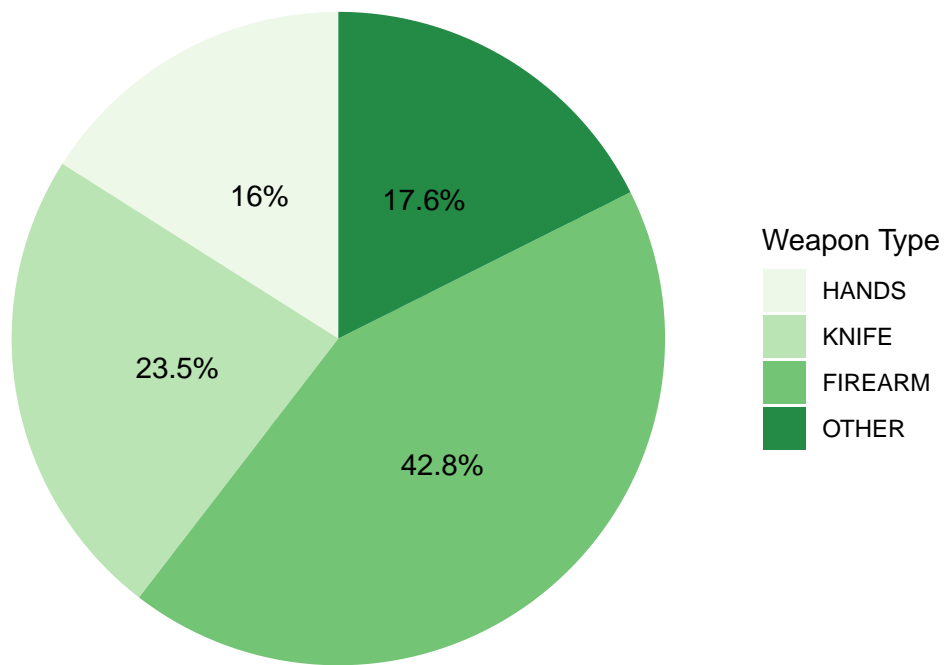
```

```

## $'Hispanic or Latino'

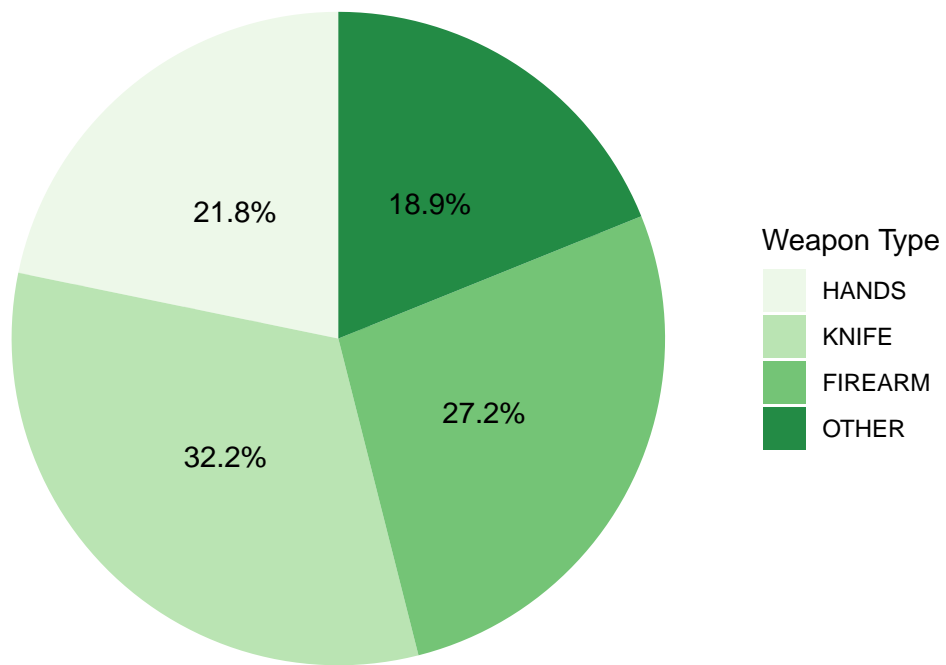
```

Weapon Type by Race: Hispanic or Latino



\$White

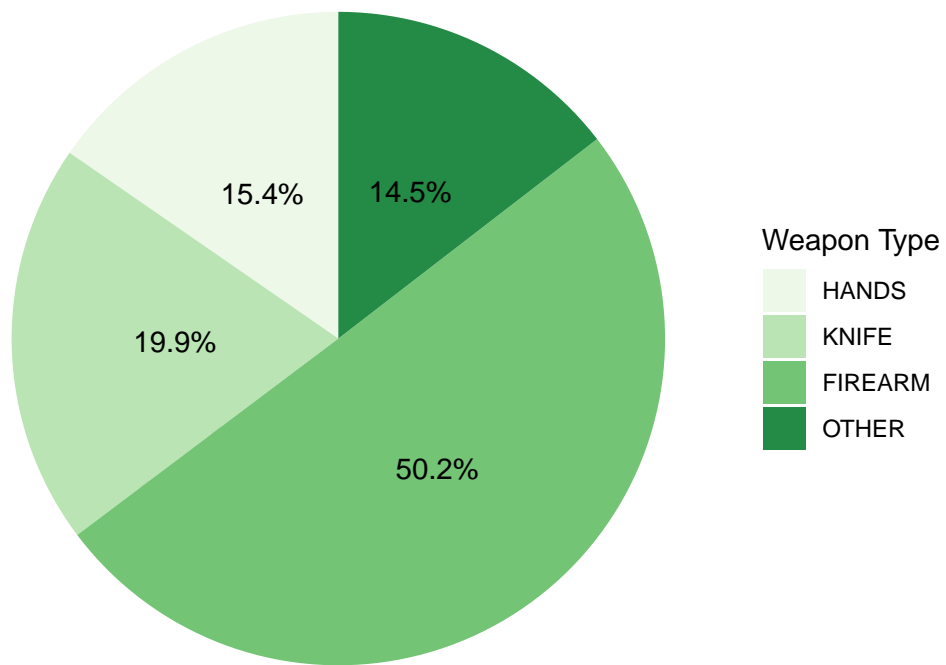
Weapon Type by Race: White



##

\$Black

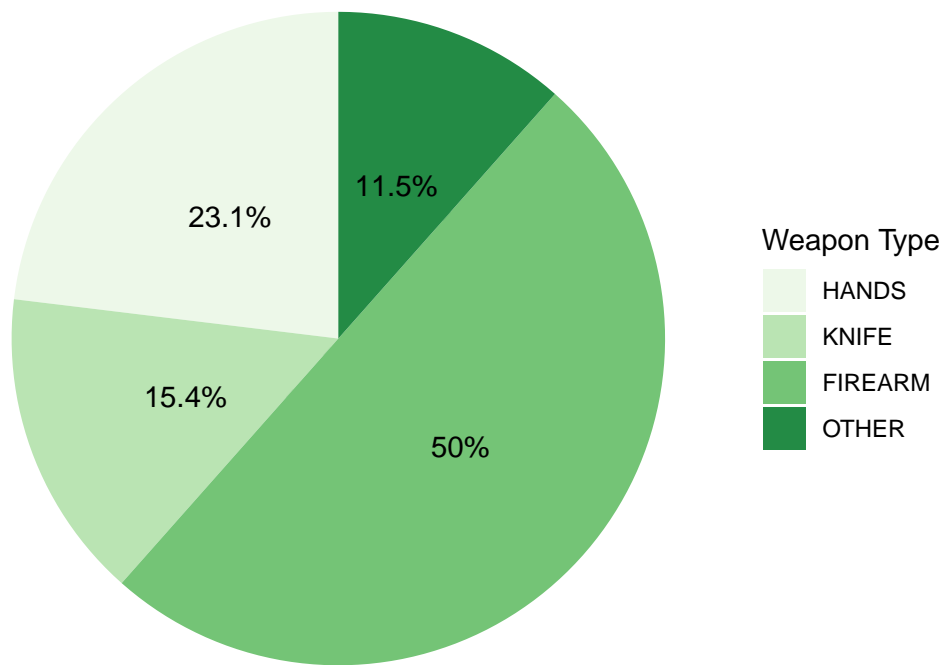
Weapon Type by Race: Black



##

\$'Middle Eastern'

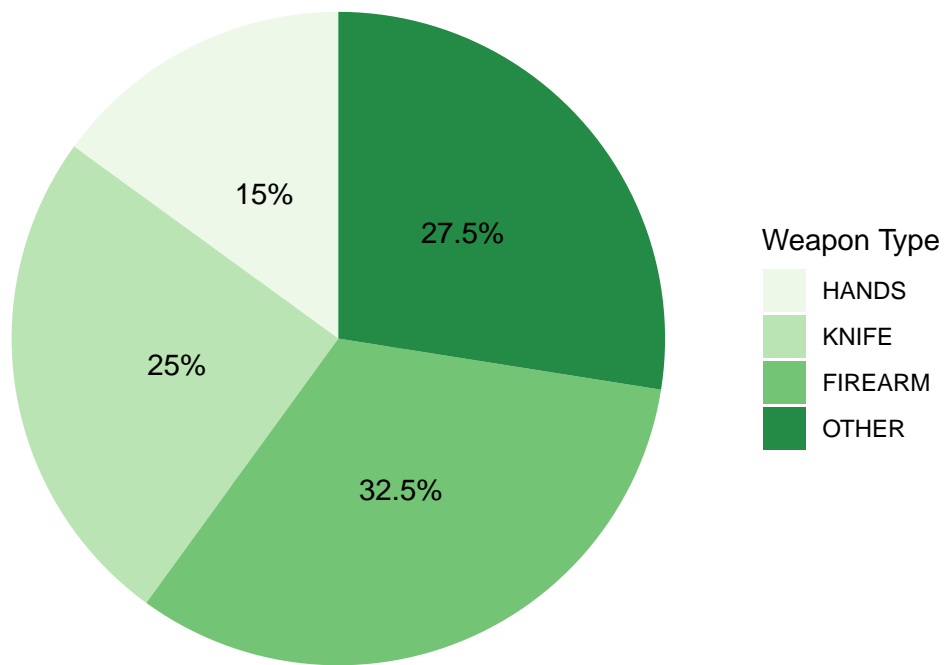
Weapon Type by Race: Middle Eastern



##

\$Asian

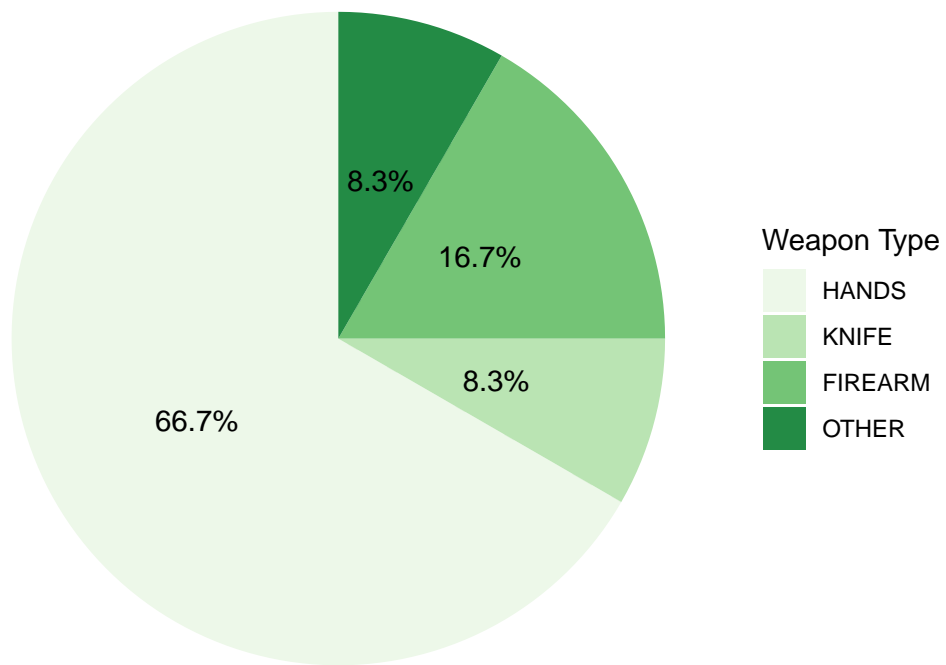
Weapon Type by Race: Asian



##

\$'American Indian or Alaska Native'

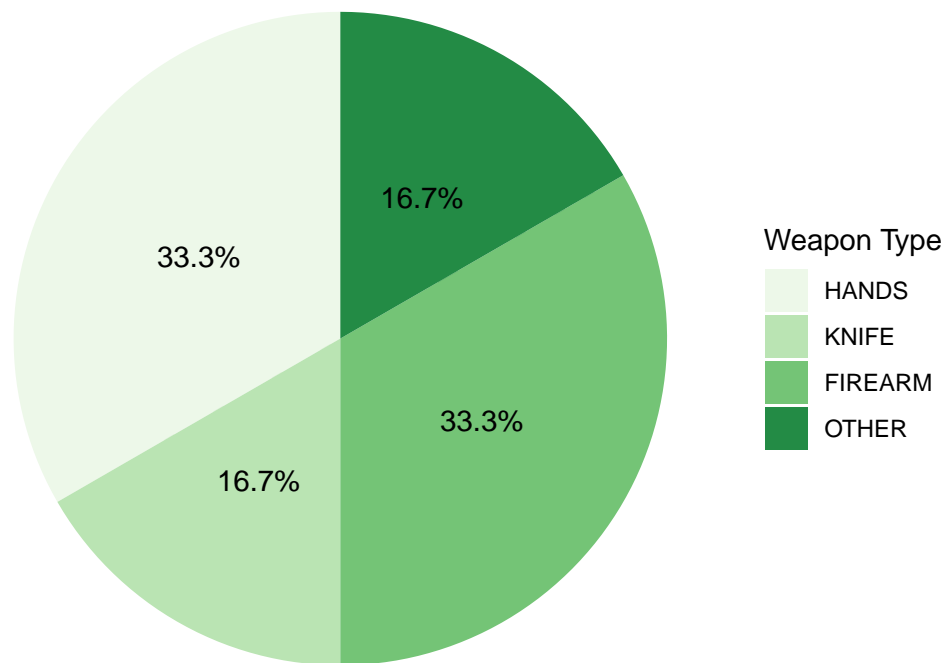
Weapon Type by Race: American Indian or Alaska Native



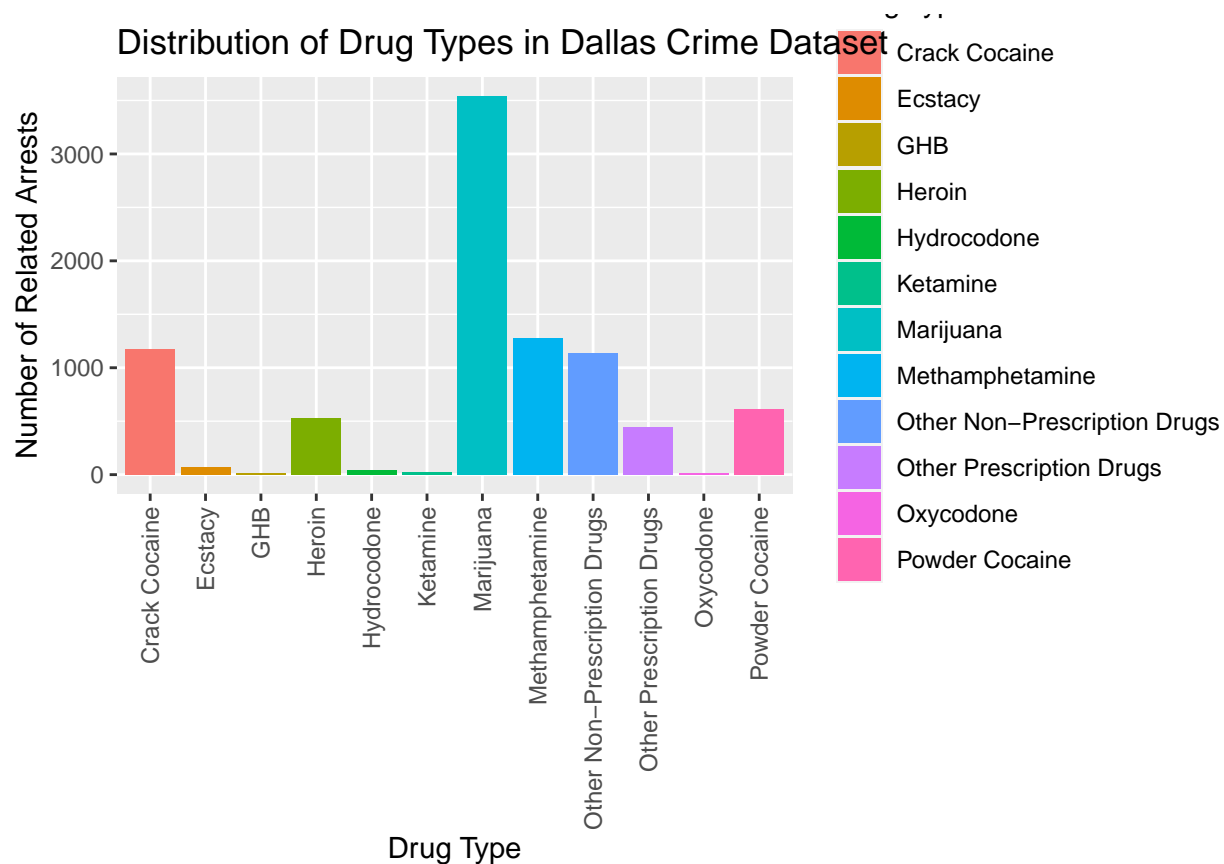
##

\$'Native Hawaiian/Pacific Islander'

Weapon Type by Race: Native Hawaiian/Pacific Islander



```
# Create a side-by-side bar plot of Drug.Type
ggplot(drugs, aes(x=Drug.Type, fill=Drug.Type)) +
  geom_bar() +
  labs(title = "Distribution of Drug Types in Dallas Crime Dataset",
        x = "Drug Type",
        y = "Number of Related Arrests", fill = "Drug Type") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



```
# load baltimore crime data
baltimore <- read.csv("baltimore_crime.csv")

# change blank values to unarmed in baltimore crime weapon data
baltimore$Weapon <- ifelse(baltimore$Weapon == "", "UNARMED", baltimore$Weapon)

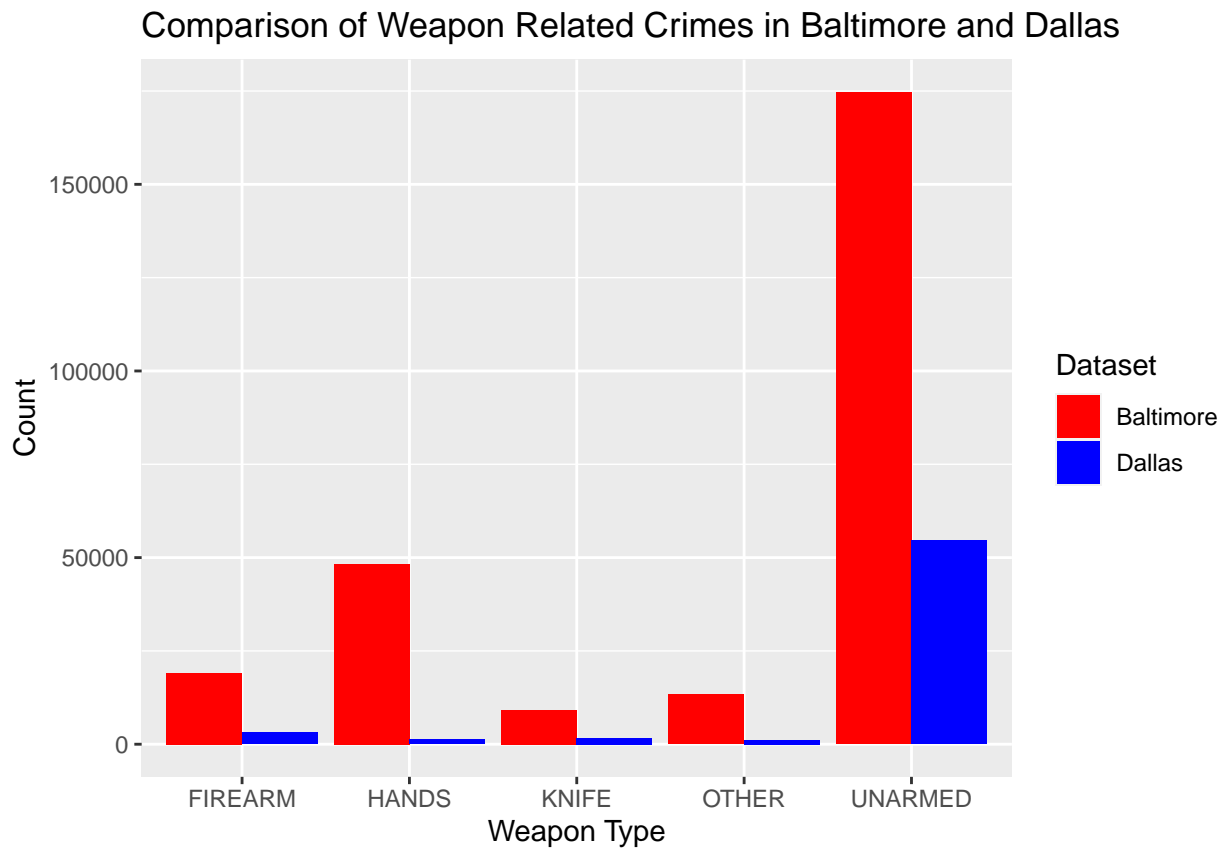
# Create a data frame with counts of each weapon type in Baltimore
baltimore_counts <- data.frame(table(baltimore$Weapon))
colnames(baltimore_counts) <- c("Weapon", "Count")

# Create a data frame with counts of each weapon type in Dallas
dallas_counts <- data.frame(table(dallas$Weapon_new))
colnames(dallas_counts) <- c("Weapon", "Count")

# Combine the two data frames
combined_counts <- rbind(baltimore_counts, dallas_counts)
combined_counts$Dataset <- ifelse(combined_counts$Count %in% baltimore_counts$Count, "Baltimore", "Dallas")

# Create a side-by-side barplot
ggplot(combined_counts, aes(x = Weapon, y = Count, fill = Dataset)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Comparison of Weapon Related Crimes in Baltimore and Dallas") +
  xlab("Weapon Type") +
  ylab("Count") +
```

```
scale_fill_manual(values = c("red", "blue"), name = "Dataset")
```



```
# Subset the data frame to exclude rows where Weapon_new == "UNARMED"
dallas_counts <- subset(dallas, Weapon_new != "UNARMED")

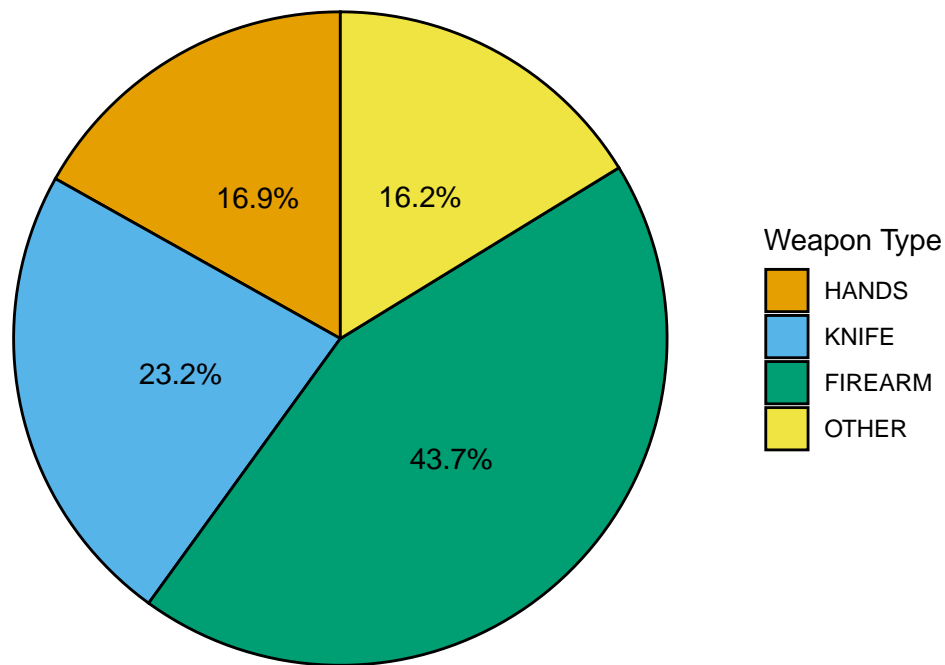
# Aggregate counts by weapon type
dallas_counts_agg <- aggregate(dallas_counts$Weapon_new, by = list(dallas_counts$Weapon_new), FUN = length)
colnames(dallas_counts_agg) <- c("Weapon_new", "Count")

# Calculate percentage of each category
dallas_counts_agg$Percent <- round(dallas_counts_agg$Count/sum(dallas_counts_agg$Count)*100, 1)

# Create the pie chart plot
pie_chart <- ggplot(dallas_counts_agg, aes(x = "", y = Count, fill = Weapon_new)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar("y", start = 0) +
  labs(title = "Weapon Types in Dallas (for Arrests Involving Weapons)",
       x = NULL, y = NULL, fill = "Weapon Type") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9", "#009E73", "#F0E442"), name = "Weapon Type") +
  theme_void() +
  # Add percentage labels to the pie chart
  geom_text(aes(label = paste0(Percent, "%")), position = position_stack(vjust = 0.5))

# Display the pie chart
pie_chart
```

Weapon Types in Dallas (for Arrests Involving Weapons)



```
# Subset the data frame to exclude rows where Weapon_new == "UNARMED"
baltimore_counts <- subset(baltimore, Weapon != "UNARMED")

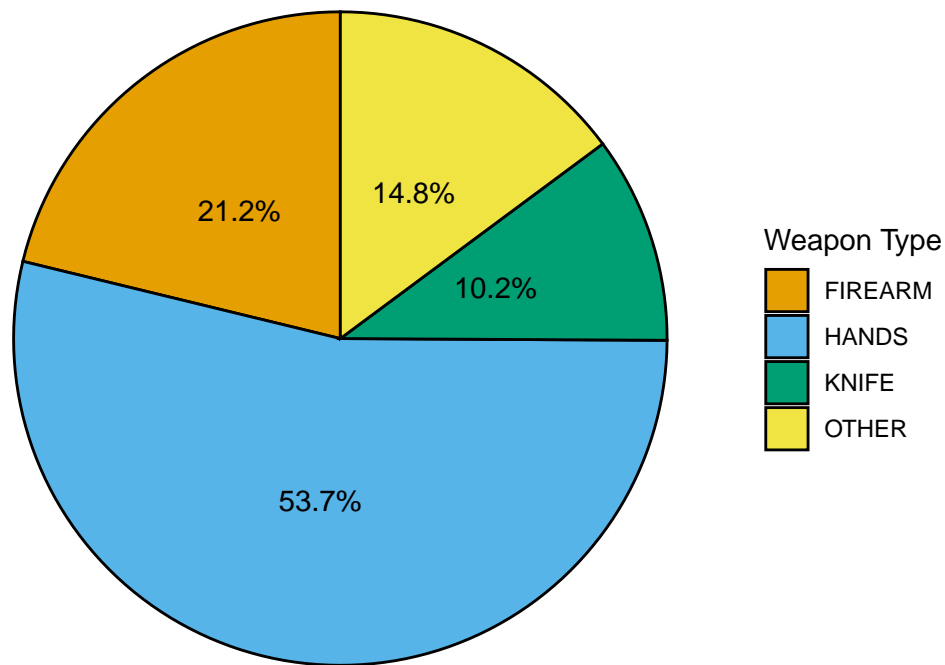
# Aggregate counts by weapon type
baltimore_counts_agg <- aggregate(baltimore_counts$Weapon, by = list(baltimore_counts$Weapon), FUN = length)
colnames(baltimore_counts_agg) <- c("Weapon", "Count")

# Calculate percentage of each category
baltimore_counts_agg$Percent <- round(baltimore_counts_agg$Count/sum(baltimore_counts_agg$Count)*100, 1)

# Create the pie chart plot
pie_chart <- ggplot(baltimore_counts_agg, aes(x = "", y = Count, fill = Weapon)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar("y", start = 0) +
  labs(title = "Weapon Types in Baltimore (for Arrests Involving Weapons)",
       x = NULL, y = NULL, fill = "Weapon Type") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9", "#009E73", "#F0E442"), name = "Weapon Type") +
  theme_void() +
  # Add percentage labels to the pie chart
  geom_text(aes(label = paste0(Percent, "%")), position = position_stack(vjust = 0.5))

# Display the pie chart
pie_chart
```

Weapon Types in Baltimore (for Arrests Involving Weapons)



```
# Specify the order of the categories
order <- c("FIREARM", "HANDS", "KNIFE", "OTHER")

# Subset the data frame to exclude rows where Weapon_new == "UNARMED"
dallas_counts <- subset(dallas, Weapon_new != "UNARMED")

# Aggregate counts by weapon type
dallas_counts_agg <- aggregate(dallas_counts$Weapon_new, by = list(dallas_counts$Weapon_new), FUN = length)
colnames(dallas_counts_agg) <- c("Weapon_new", "Count")

# Convert Weapon_new to a factor with the specified order
dallas_counts_agg$Weapon_new <- factor(dallas_counts_agg$Weapon_new, levels = order)

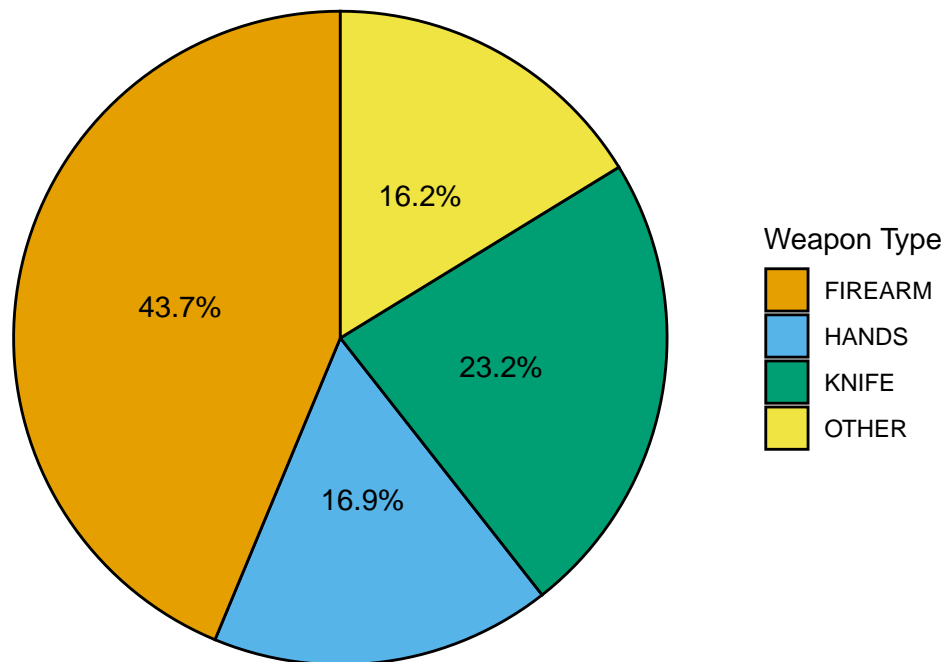
# Calculate percentage of each category
dallas_counts_agg$Percent <- round(dallas_counts_agg$Count/sum(dallas_counts_agg$Count)*100, 1)

# Create the pie chart plot
pie_chart <- ggplot(dallas_counts_agg, aes(x = "", y = Count, fill = Weapon_new)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar("y", start = 0) +
  labs(title = "Weapon Types in Dallas (for Arrests Involving Weapons)",
       x = NULL, y = NULL, fill = "Weapon Type") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9", "#009E73", "#F0E442"), name = "Weapon Type") +
  theme_void() +
  # Add percentage labels to the pie chart
  geom_text(aes(label = paste0(Percent, "%")), position = position_stack(vjust = 0.5))
```



```
# Display the pie chart
pie_chart
```

Weapon Types in Dallas (for Arrests Involving Weapons)



```
# Subset the data frame to exclude rows where Weapon_new == "UNARMED"
baltimore_counts <- subset(baltimore, Weapon != "UNARMED")

# Aggregate counts by weapon type
baltimore_counts_agg <- aggregate(baltimore_counts$Weapon, by = list(baltimore_counts$Weapon), FUN = length)
colnames(baltimore_counts_agg) <- c("Weapon", "Count")

# Change order of factor levels for Weapon variable
baltimore_counts_agg$Weapon <- factor(baltimore_counts_agg$Weapon, levels = c("FIREARM", "KNIFE", "HANDS", "OTHER"))

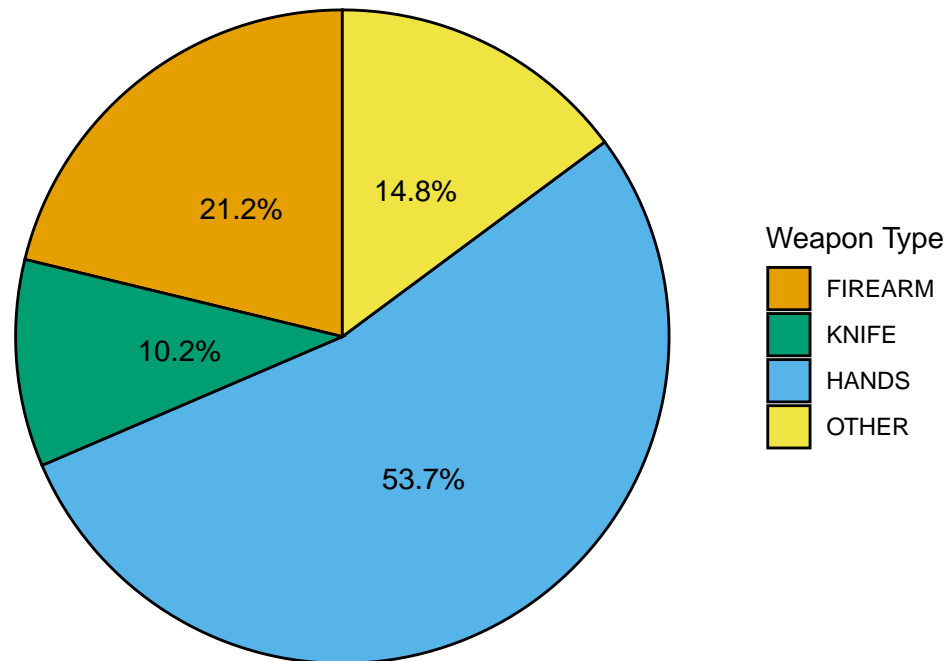
# Calculate percentage of each category
baltimore_counts_agg$Percent <- round(baltimore_counts_agg$Count/sum(baltimore_counts_agg$Count)*100, 1)

# Create the pie chart plot
pie_chart <- ggplot(baltimore_counts_agg, aes(x = "", y = Count, fill = Weapon)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar("y", start = 0) +
  labs(title = "Weapon Types in Baltimore (for Arrests Involving Weapons)",
       x = NULL, y = NULL, fill = "Weapon Type") +
  scale_fill_manual(values = c("#E69F00", "#009E73", "#56B4E9", "#F0E442"), name = "Weapon Type") +
  theme_void() +
```

```
# Add percentage labels to the pie chart
geom_text(aes(label = paste0(Percent, "%")), position = position_stack(vjust = 0.5))

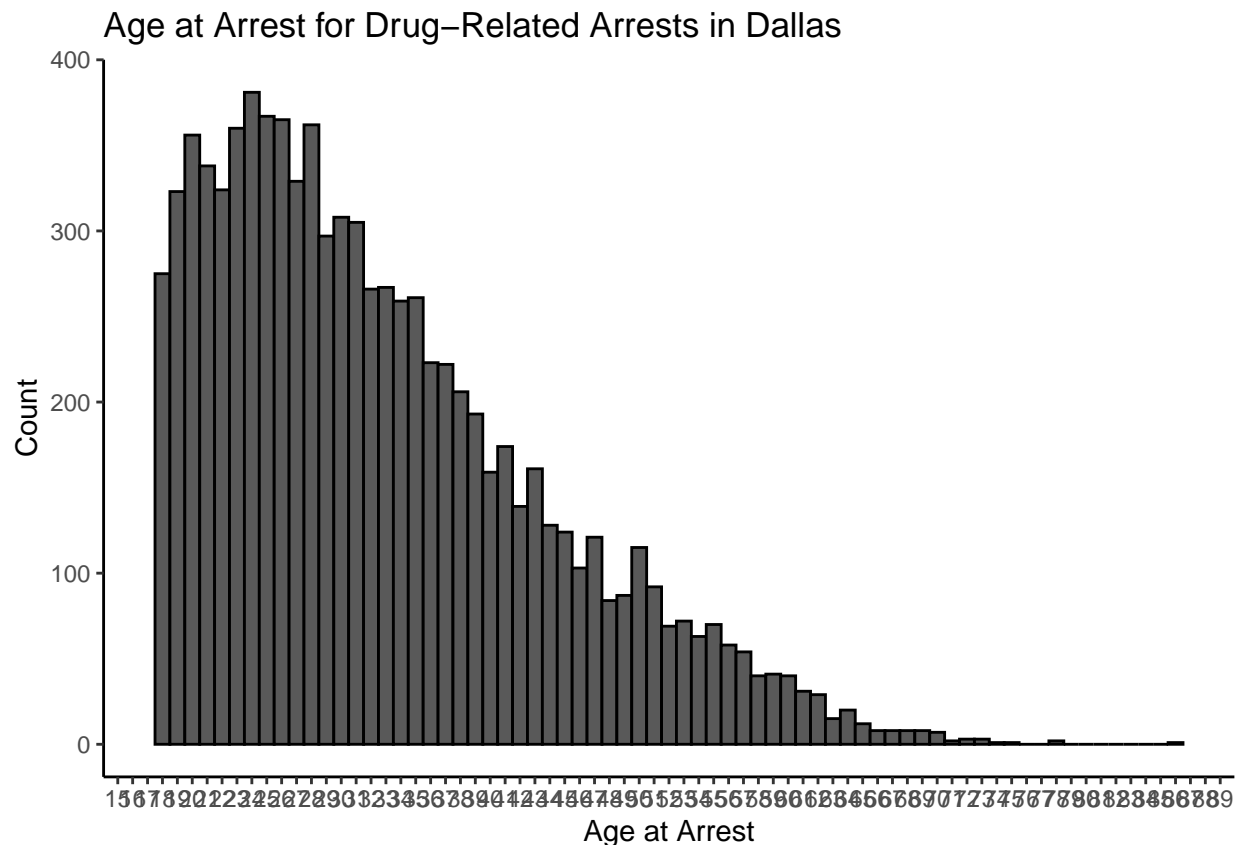
# Display the pie chart
pie_chart
```

Weapon Types in Baltimore (for Arrests Involving Weapons)



```
# Subset the data frame to include only rows where Drug.Related == "Yes"
dallas_drug_related <- subset(dallas, Drug.Related == "Yes")

# Create the histogram plot
ggplot(dallas_drug_related, aes(x = Arrestee.Age.At.Arrest.Time)) +
  geom_histogram(binwidth = 1, color = "black") +
  labs(title = "Age at Arrest for Drug-Related Arrests in Dallas",
       x = "Age at Arrest",
       y = "Count") +
  theme_classic() +
  scale_x_continuous(breaks = seq(0, 90, by = 1))
```



```
# Create Age.Group variable for dallas_drug_related
dallas_drug_related <- dallas %>%
  filter(Drug.Related == "Yes") %>%
  mutate(Age.Group = cut(Arrestee.Age.At.Arrest.Time, breaks = seq(0, 100, 10),
    labels = c("<10", "10's", "20's", "30's", "40's", "50's", "60's", "70's", "80's", "80+"))

# Create histogram with colored bars by Age.Group
ggplot(dallas_drug_related, aes(x = Arrestee.Age.At.Arrest.Time, fill = Age.Group)) +
  geom_histogram(binwidth = 1, color = "black", alpha = 0.5) +
  scale_fill_manual(values = c("#D7191C", "#F0E442", "#0072B2", "#D55E00", "#E69F00", "#009E73", "#56B4E9")) +
  labs(title = "Distribution of Ages for Drug-Related Arrests in Dallas",
    x = "Age at Arrest",
    y = "Count",
    fill = "Age Group") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 90, by = 1))
```

