

## Viz2- CyberSquirrel Attacks : Shashwat Singh

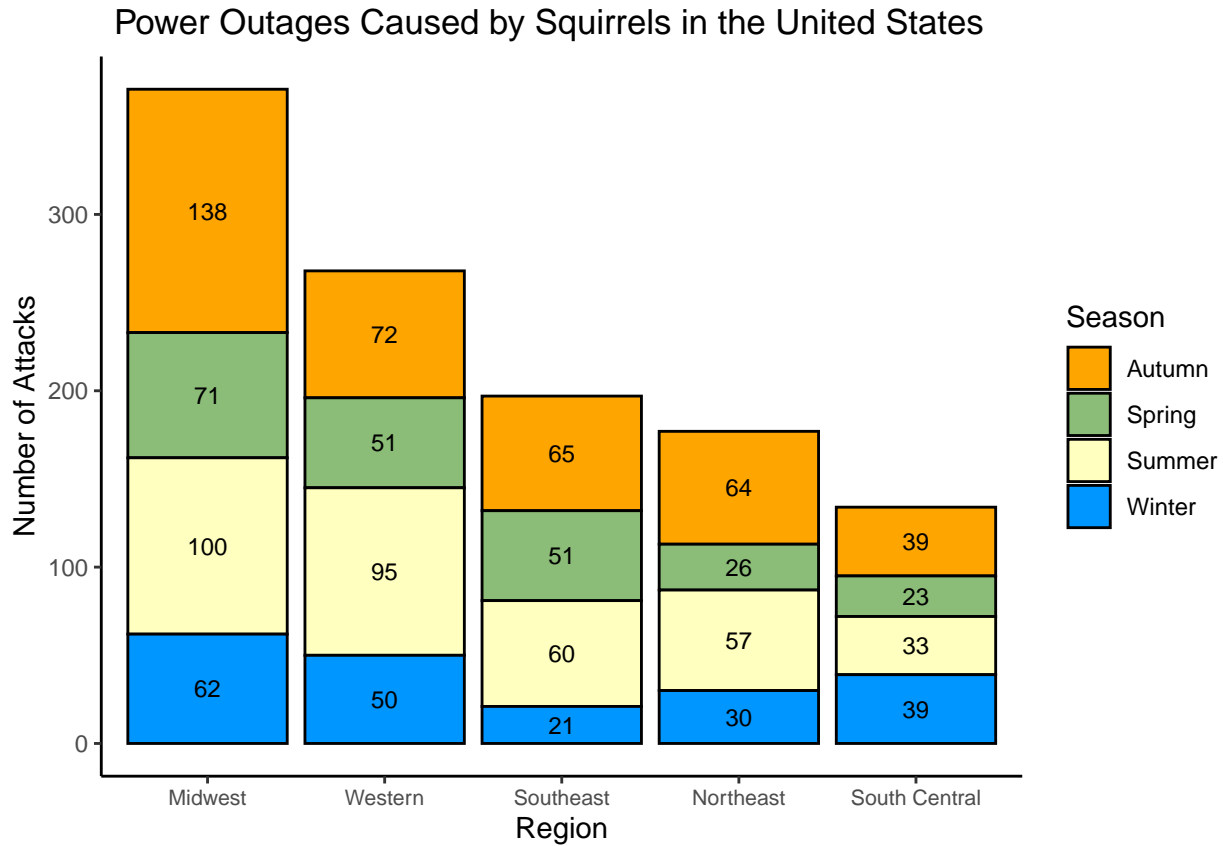


Figure 1. The above graphic shows the number of squirrel attacks in the United States which caused a power outage. Data was cleaned and grouped. . The “date” given were translated to seasons based on the month as below

1. Spring - March to May.
2. Summer - June to August.
3. Autumn - September to November.
4. Winter - December to February.

The states were mapped to corresponding regions in the United States as below

1. Midwest: NE, KS, IN, OH, IA, IL, MO, WI, MI, MN, SD, ND
2. Southeast: VA, TN, LA, AL, MS, FL, KY, NC, GA
3. Western: CA, CO, TX, WA, MT, ID, OR, WY, UT, AZ, AK, NV
4. Northeast: NY, NJ, VT, PA, MD, NH, CT, RI, DE, DC
5. South Central: OK, AR, NM, SC, ME, MA

The number within each block represents the count of attacks specific to that Region and Season

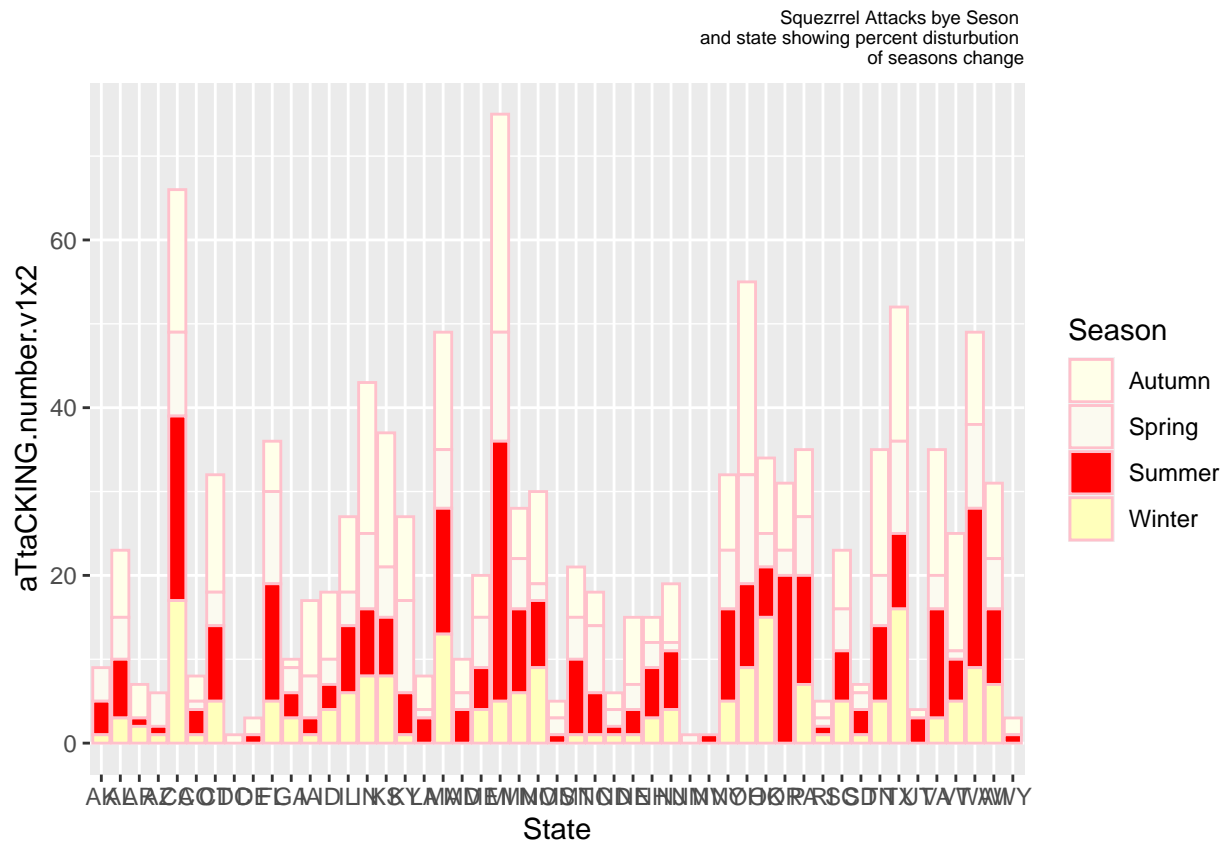


Figure 2. Squezrrel Attacks bye Month and Season showing percent disturbtion of squirrels attack

## Code Used To Make These Plots

Below are the code chunks used to make these plots. Use `install.packages()` to install any missing packages before you run.

```
library(ggplot2)
library(dplyr)
library(lubridate)
#Reading File
squirrels<-read.csv("CyberSquirrel-map_data.csv")

#Filtering for squirrels in the United States
squirrels<- squirrels %>%
  filter(Operative=="Squirrel", Country=="United States")

#Mapped the States to the corresponding regions as
#Midwest: NE, KS, IN, OH, IA, IL, MO, WI, MI, MN, SD, ND
#Southeast: SC, ME, MA, VA, TN, LA, AL, MS, FL, KY, NC, GA
#Western: CA, CO, TX, WA, MT, ID, OR, WY, UT, AZ, AK, NV
#Northeast: NY, NJ, VT, PA, MD, NH, CT, RI, DE, DC
#South Central: OK, AR, NM

state_region_mapping <- data.frame(
  State.Provence = c("NE", "KS", "IN", "OH",
                     "IA", "IL", "MO", "WI",
                     "MI", "MN", "SD", "ND",
                     "SC", "ME", "MA", "VA",
                     "TN", "LA", "AL", "MS",
                     "FL", "KY", "NC", "GA",
                     "CA", "CO", "TX", "WA",
                     "MT", "ID", "OR", "WY",
                     "UT", "AZ", "AK", "NV",
                     "NY", "NJ", "VT", "PA",
                     "MD", "NH", "CT", "RI",
                     "DE", "DC", "OK", "AR", "NM"),

  Region = c("Midwest", "Midwest", "Midwest", "Midwest",
             "Midwest", "Midwest", "Midwest", "Midwest",
             "Midwest", "Midwest", "Midwest", "Midwest",
             "South Central", "South Central", "South Central", "Southeast",
             "Southeast", "Southeast", "Southeast", "Southeast", "Southeast",
             "Southeast", "Southeast", "Southeast",
             "Western", "Western", "Western", "Western",
             "Western", "Western", "Western", "Western",
             "Western", "Western", "Western", "Western",
             "Northeast", "Northeast", "Northeast", "Northeast",
             "Northeast", "Northeast", "Northeast", "Northeast", "Northeast", "Northeast",
             "South Central", "South Central",
             "South Central")
)
```

```

#Joining the Region column on squirrels dataset
squirrels <- squirrels %>%
  left_join(state_region_mapping, by = "State.Provence")

#Removed any unmapped regions due to ambiguity in the statecode
squirrels<- squirrels %>%
  filter(!is.na(Region))

# Extracting month from date
squirrels <- squirrels %>%
  mutate(Date = as.Date(Date, format = "%m/%d/%Y"),
         Month = month(Date))

# Create the "Season" column based on the month
#Spring - March to May.
#Summer - June to August.
#Autumn - September to November.
#Winter - December to February.

squirrels <- squirrels %>%
  mutate(Season = ifelse(Month %in% 3:5, "Spring",
                        ifelse(Month %in% 6:8, "Summer",
                              ifelse(Month %in% 9:11, "Autumn", "Winter"))))

#Get final table with relevant data
squirrels_final <- squirrels %>%
  group_by(Region, Season) %>%
  summarise(count = n())

# Create order for a pretty descending graph
region_order <- c("Midwest", "Western", "Southeast", "Northeast", "South Central")

#The Pretty Plot

ggplot(squirrels_final, aes(x = factor(Region, levels = region_order),
                           y = count, fill = Season, label = count)) +
  geom_bar(stat = "identity", col = "black") +
  labs(
    title = "Power Outages Caused by Squirrels in the United States",
    x = "Region",
    y = "Number of Attacks"
  ) +
  scale_fill_manual(values = c("Spring" = "#8bbd78",
                              "Summer" = "#FFFFBF",
                              "Autumn" = "orange",
                              "Winter" = "#0096FF")) +
  theme_classic() +
  theme(
    axis.text.x = element_text(size = 8),
    plot.title = element_text(hjust = 0.25)
  ) +
  geom_text(position = position_stack(vjust = 0.5), size = 3)

```

*#The Not so Pretty Plot*

```
ggplot(squirrels, aes(x = State.Provence, fill = Season)) +  
  geom_bar(col="pink"  
  ) +  
  labs(  
    title = "Squezzrrel Attacks bye Seson  
    and state showing percent disturbutio  
    of seasons change",  
    x = "State",  
    y = "aTtaCKING.number.v1x2"  
  ) +  
  scale_fill_manual(values = c("Spring" = "#FBFAF0", "Summer" = "red",  
                                "Autumn" = "#FFFFE9", "Winter" = "#FFFFBB")) +  
  theme(plot.title = element_text(hjust = 1, size=7))
```

All of the Errors in Figure 2 :

- The caption says it is showing percentage distribution of squirrel attack, which it is not.
- The X-axis label is practically unreadable due to overlap.
- The colors are terrible.
- The y axis does not have a easy to read name
- The data is distributed so much that we can not identify a trend.
- The title is also right aligned with uneven line breaks.