

# Cyber Research

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.1      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'purrr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

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

```
library(dplyr)
library(ggplot2)
library(fastDummies)
```

```
## Warning: package 'fastDummies' was built under R version 4.0.5
```

```
#install.packages("patchwork")
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.0.5
```

```
df.edu.safety = read.csv("https://raw.githubusercontent.com/tahlla-utd/cybersecresearch/main/CyberResearchData/edu.safety.csv",
                          header = TRUE, sep = ';')
```

```
#identifying the rows with NAs
```

```
rownames(df.edu.safety)[apply(df.edu.safety, 2, anyNA)]
```

```
## [1] "18" "19" "20" "21" "22" "46" "47" "48" "49" "50" "74" "75"  
## [13] "76" "77" "78" "102" "103" "104" "105" "106" "130" "131" "132" "133"  
## [25] "134" "158" "159" "160" "161" "162" "186" "187" "188" "189" "190" "214"  
## [37] "215" "216" "217" "218" "242" "243" "244" "245" "246" "270" "271" "272"  
## [49] "273" "274" "298" "299" "300" "301" "302" "326" "327" "328" "329" "330"  
## [61] "354" "355" "356" "357" "358" "382" "383" "384" "385" "386" "410" "411"  
## [73] "412" "413" "414" "438" "439" "440" "441" "442" "466" "467" "468" "469"  
## [85] "470" "494" "495" "496" "497" "498" "522" "523" "524" "525" "526" "550"  
## [97] "551" "552" "553" "554" "578" "579" "580" "581" "582" "606" "607"
```

```
#removing all observations with NAs
```

```
df.clean <- df.edu.safety %>% na.omit()
```

```
df.clean <- df.clean
```

```
df.clean$year = as.factor(df.clean$year)
```

```
table(df.clean$website_quality)
```

```
##
```

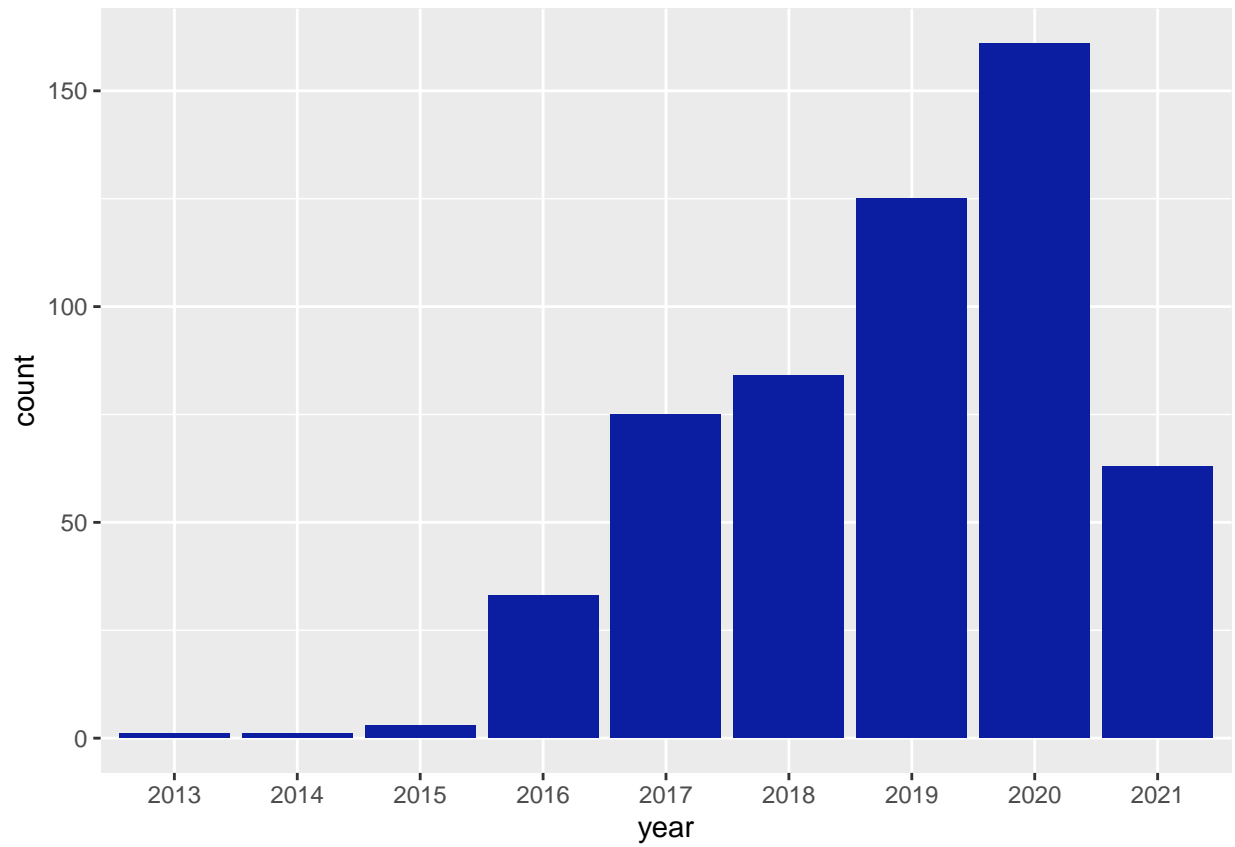
```
## high low med
```

```
## 157 86 303
```

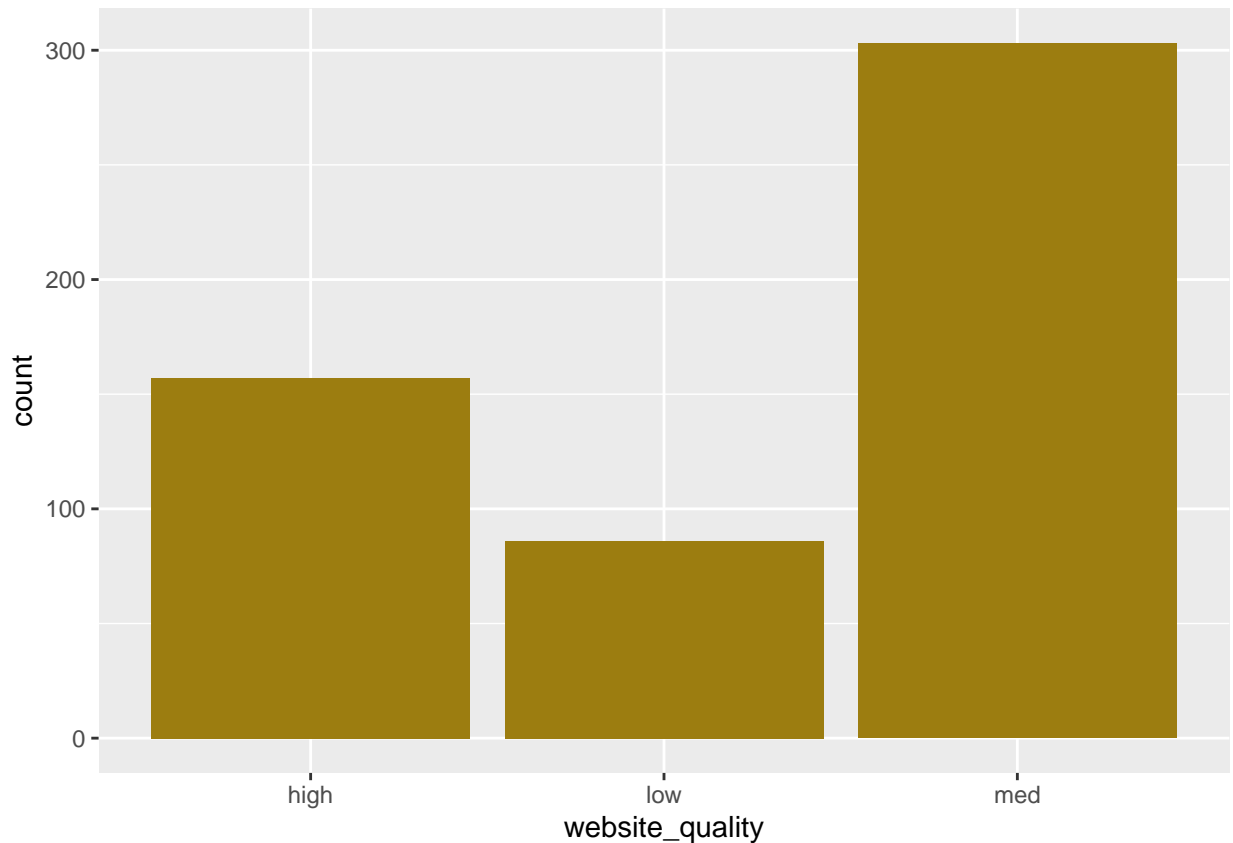
Linear Regression of Various variables

```
test.plot <- ggplot(data = df.clean, aes(x = year))+geom_bar(fill = "#0b1da1")
```

```
test.plot
```



```
test.plot <- ggplot(data = df.clean, aes(x = website_quality))+geom_bar(fill = "#9c7d10")
test.plot
```



```
#create data frame of attack type counts
```

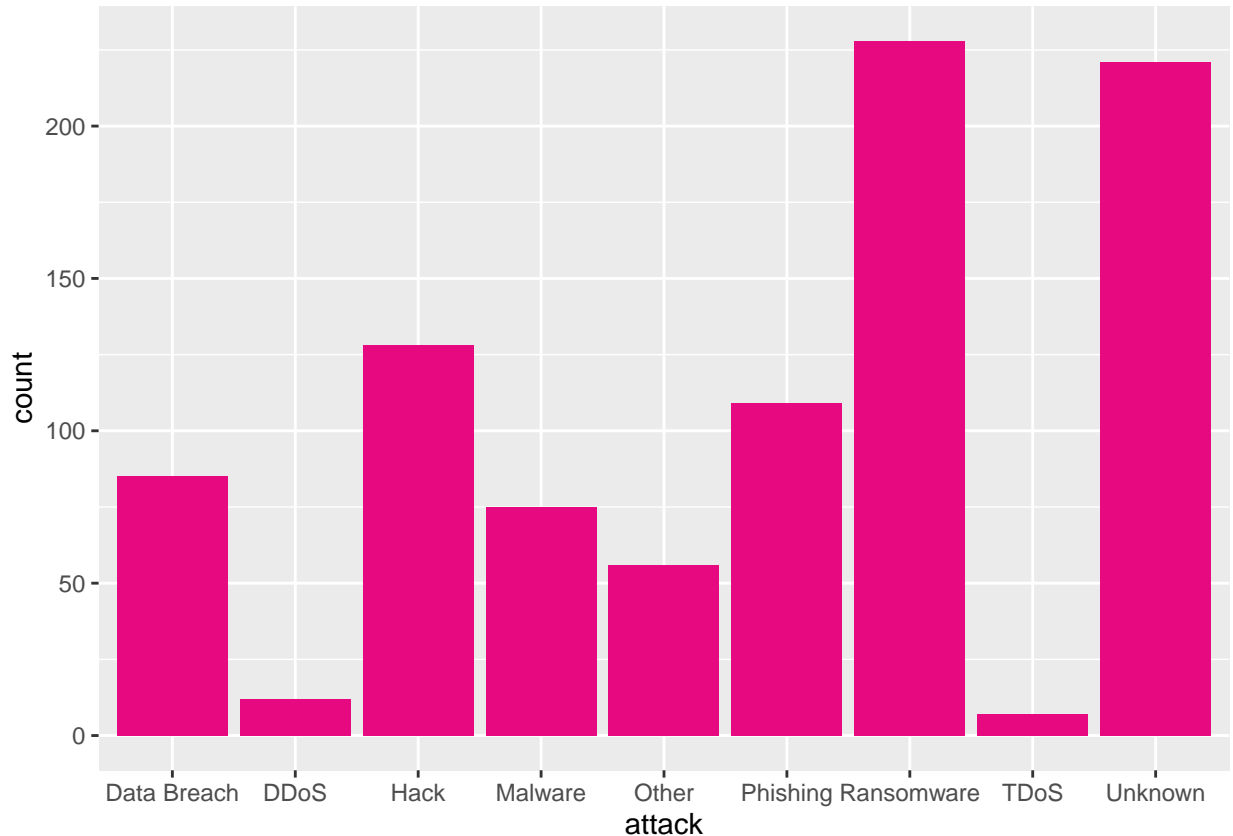
```
df.attack.count <- data.frame(matrix(ncol = 9, nrow = 1))
x <- c("Unknown","Hack","Data Breach","Ransomware","Phishing","TDoS","Malware","DDoS","Other")
colnames(df.attack.count)<- x
```

```
df.attack.count[1, "Unknown"] <- sum(df.clean$Unknown)
df.attack.count[1, "Hack"] <- sum(df.clean$Hack)
df.attack.count[1, "Data Breach"] <- sum(df.clean$Data_Breach)
df.attack.count[1, "Ransomware"] <- sum(df.clean$Ransomware)
df.attack.count[1, "Phishing"] <- sum(df.clean$Phishing)
df.attack.count[1, "TDoS"] <- sum(df.clean$TDoS)
df.attack.count[1, "Malware"] <- sum(df.clean$Malware)
df.attack.count[1, "DDoS"] <- sum(df.clean$DDoS)
df.attack.count[1, "Other"] <- sum(df.clean$Other)
```

```
head(df.attack.count)
```

```
##   Unknown Hack Data Breach Ransomware Phishing TDoS Malware DDoS Other
## 1    221  128         85         228    109    7      75   12   56
```

```
df.long <- df.attack.count %>%
  pivot_longer(Unknown:Other, names_to = "attack", values_to = "count")
ggplot(df.long, aes(x = attack, y = count)) +
  geom_col(fill = "#e6097f")
```



```
graph.attack.by.filter <- function(fill.choice){
  df.unk <- df.clean %>% as_tibble() %>%
    mutate(Unknown = as.factor(if_else(Unknown == 1, "True", "False")))

  unk.graph <- ggplot(data = df.unk, aes(Unknown == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
  #####
  df.hack <- df.clean %>% as_tibble() %>%
    mutate(Hack = as.factor(if_else(Hack == 1, "True", "False")))

  hack.graph <- ggplot(data = df.hack, aes(Hack == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
  #####
  df.data <- df.clean %>% as_tibble() %>%
    mutate(Data_Breach = as.factor(if_else(Data_Breach == 1, "True", "False")))

  data.graph <- ggplot(data = df.data, aes(Data_Breach == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
  #####
  df.ran <- df.clean %>% as_tibble() %>%
    mutate(Ransomware = as.factor(if_else(Ransomware == 1, "True", "False")))

  ran.graph <- ggplot(data = df.ran, aes(Ransomware == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
  #####
  df.phish <- df.clean %>% as_tibble() %>%
    mutate(Phishing = as.factor(if_else(Phishing == 1, "True", "False")))
```

```

phish.graph <- ggplot(data = df.phish, aes(Phishing == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####
df.tdos <- df.clean %>% as_tibble() %>%
  mutate(TDoS = as.factor(if_else(TDoS == 1, "True", "False")))

tdos.graph <- ggplot(data = df.tdos, aes(TDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####
df.mal <- df.clean %>% as_tibble() %>%
  mutate(Malware = as.factor(if_else(Malware == 1, "True", "False")))

mal.graph <- ggplot(data = df.mal, aes(Malware == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####
df.ddos <- df.clean %>% as_tibble() %>%
  mutate(DDoS = as.factor(if_else(DDoS == 1, "True", "False")))

ddos.graph <- ggplot(data = df.ddos, aes(DDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####
df.other <- df.clean %>% as_tibble() %>%
  mutate(Other = as.factor(if_else(Other == 1, "True", "False")))

other.graph <- ggplot(data = df.other, aes(Other == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####
df.serv <- df.clean %>% as_tibble() %>%
  mutate(Server_Shutdown = as.factor(if_else(Server_Shutdown == 1, "True", "False")))

serv.graph <- ggplot(data = df.serv, aes(Server_Shutdown == "True", fill = fill.choice))+geom_bar() + ylim(0, 250)
#####

(data.graph + scale_x_discrete(limit = c(TRUE)) | ddos.graph + scale_x_discrete(limit = c(TRUE)) |
  hack.graph + scale_x_discrete(limit = c(TRUE)))/ (mal.graph + scale_x_discrete(limit = c(TRUE)) |
  other.graph + scale_x_discrete(limit = c(TRUE)) | phish.graph + scale_x_discrete(limit = c(TRUE)))/
  (ran.graph + scale_x_discrete(limit = c(TRUE)) | tdos.graph + scale_x_discrete(limit = c(TRUE)) | unk
  scale_x_discrete(limit = c(TRUE)))
}

```

```
graph.attack.by.filter(df.clean$year)
```

```

## Warning: Removed 461 rows containing non-finite values (stat_count).

## Warning: Removed 534 rows containing non-finite values (stat_count).

## Warning: Removed 418 rows containing non-finite values (stat_count).

## Warning: Removed 471 rows containing non-finite values (stat_count).

## Warning: Removed 490 rows containing non-finite values (stat_count).

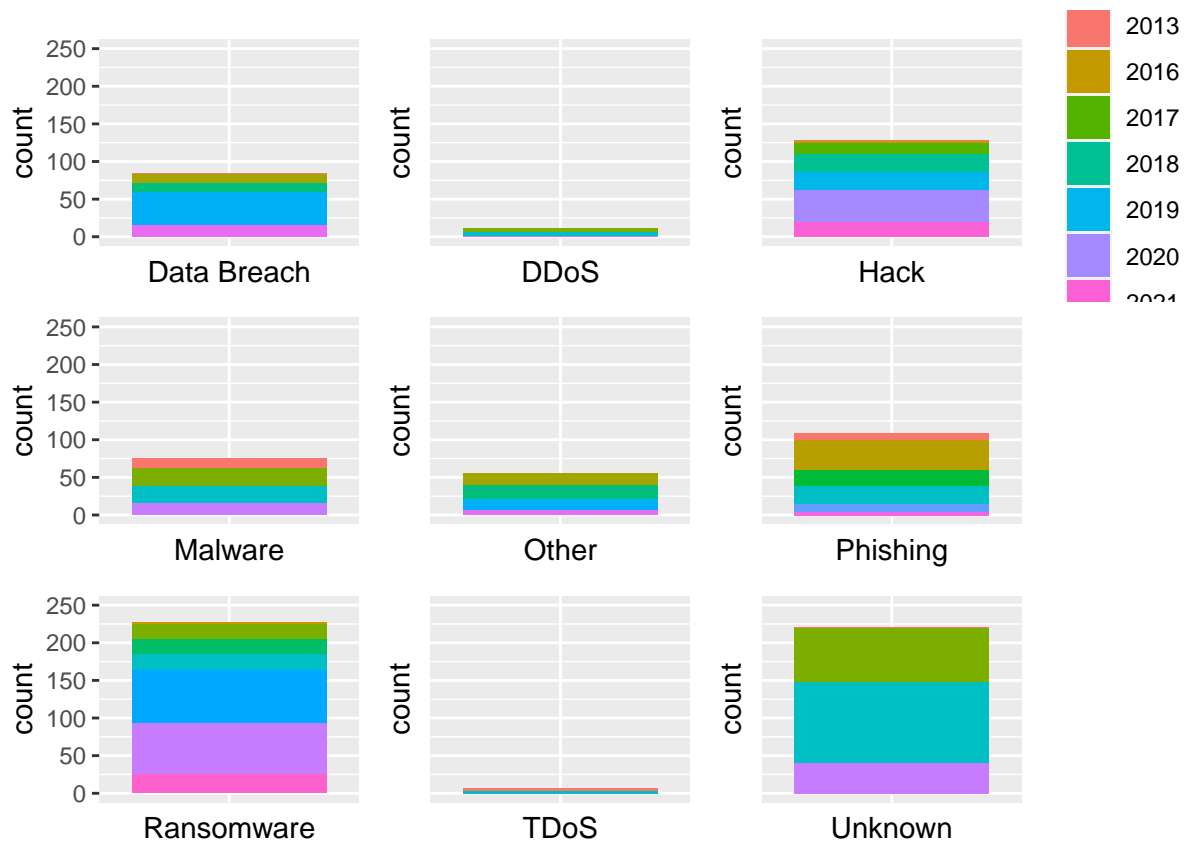
## Warning: Removed 437 rows containing non-finite values (stat_count).

## Warning: Removed 318 rows containing non-finite values (stat_count).

```

```
## Warning: Removed 539 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 325 rows containing non-finite values (stat_count).
```



```
graph.attack.by.filter(df.clean$type)
```

```
## Warning: Removed 461 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 534 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 418 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 471 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 490 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 437 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 318 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 539 rows containing non-finite values (stat_count).
```

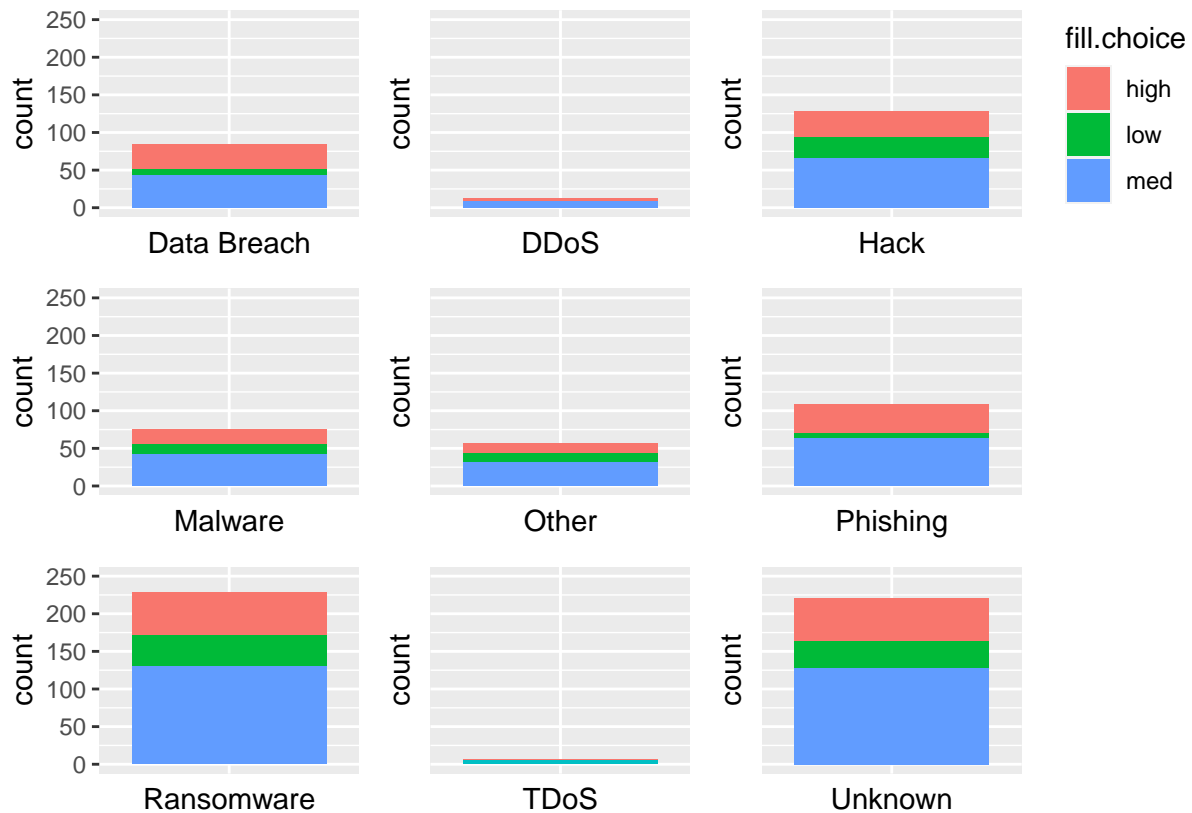
```
## Warning: Removed 325 rows containing non-finite values (stat_count).
```



```
graph.attack.by.filter((df.clean$website_quality))
```

```
## Warning: Removed 461 rows containing non-finite values (stat_count).
## Warning: Removed 534 rows containing non-finite values (stat_count).
## Warning: Removed 418 rows containing non-finite values (stat_count).
## Warning: Removed 471 rows containing non-finite values (stat_count).
## Warning: Removed 490 rows containing non-finite values (stat_count).
## Warning: Removed 437 rows containing non-finite values (stat_count).
## Warning: Removed 318 rows containing non-finite values (stat_count).
## Warning: Removed 539 rows containing non-finite values (stat_count).
## Warning: Removed 325 rows containing non-finite values (stat_count).
```





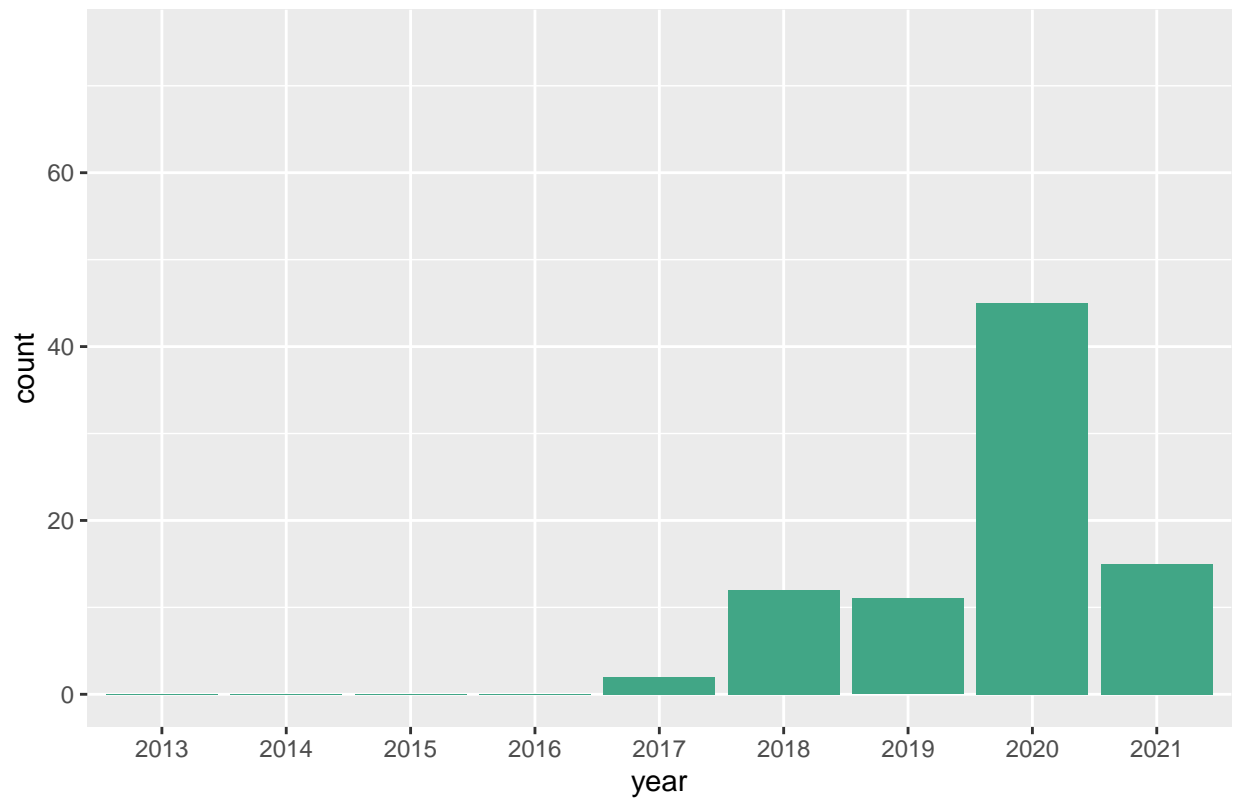
```
year.trend <-function(attack.type, color.choice){
  year.list <- c("2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021")
  df.tmp <- data.frame(matrix(ncol = 9, nrow = 1))
  colnames(df.tmp) <- year.list

  df.tmp[1, "2013"] <- sum((df.clean[which(df.clean$year == "2013"), attack.type]))
  df.tmp[1, "2014"] <- sum((df.clean[which(df.clean$year == "2014"), attack.type]))
  df.tmp[1, "2015"] <- sum((df.clean[which(df.clean$year == "2015"), attack.type]))
  df.tmp[1, "2016"] <- sum((df.clean[which(df.clean$year == "2016"), attack.type]))
  df.tmp[1, "2017"] <- sum((df.clean[which(df.clean$year == "2017"), attack.type]))
  df.tmp[1, "2018"] <- sum((df.clean[which(df.clean$year == "2018"), attack.type]))
  df.tmp[1, "2019"] <- sum((df.clean[which(df.clean$year == "2019"), attack.type]))
  df.tmp[1, "2020"] <- sum((df.clean[which(df.clean$year == "2020"), attack.type]))
  df.tmp[1, "2021"] <- sum((df.clean[which(df.clean$year == "2021"), attack.type]))

  df.long <- df.tmp %>%
    pivot_longer(1:9, names_to = "year", values_to = "count")
  ggplot(df.long, aes(x = year, y = count)) +
    geom_col(fill = color.choice) + labs(title = attack.type) + ylim(0, 75)
}
```

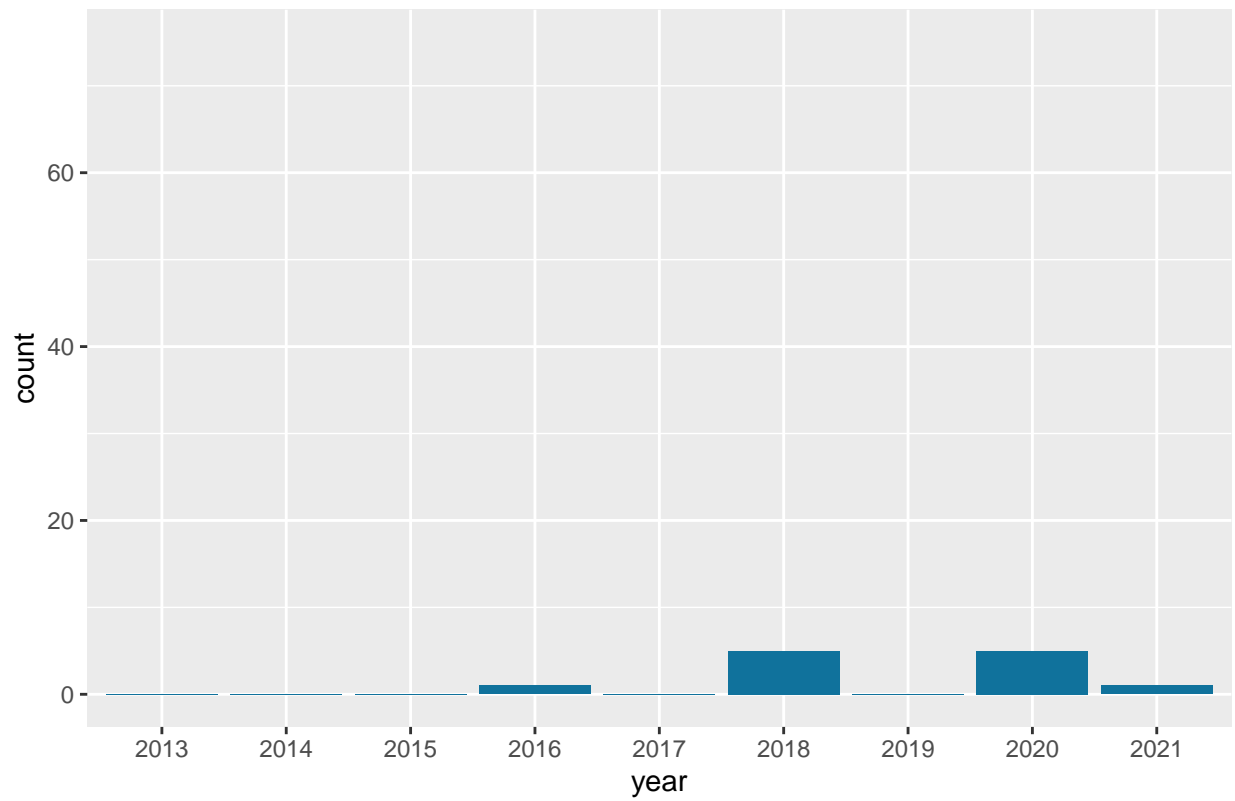
```
year.trend("Data_Breach", "#41a686")
```

Data\_Breach



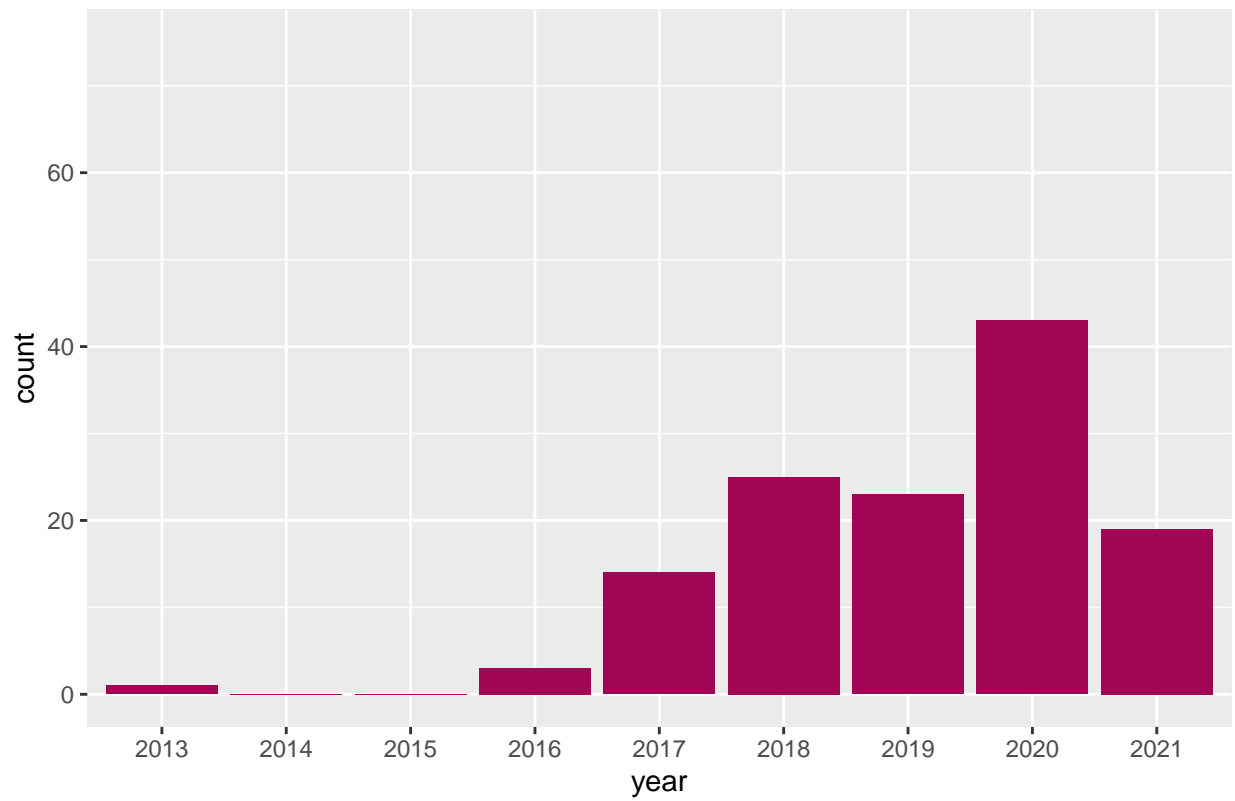
```
year.trend("DDoS", "#10729c")
```

## DDoS



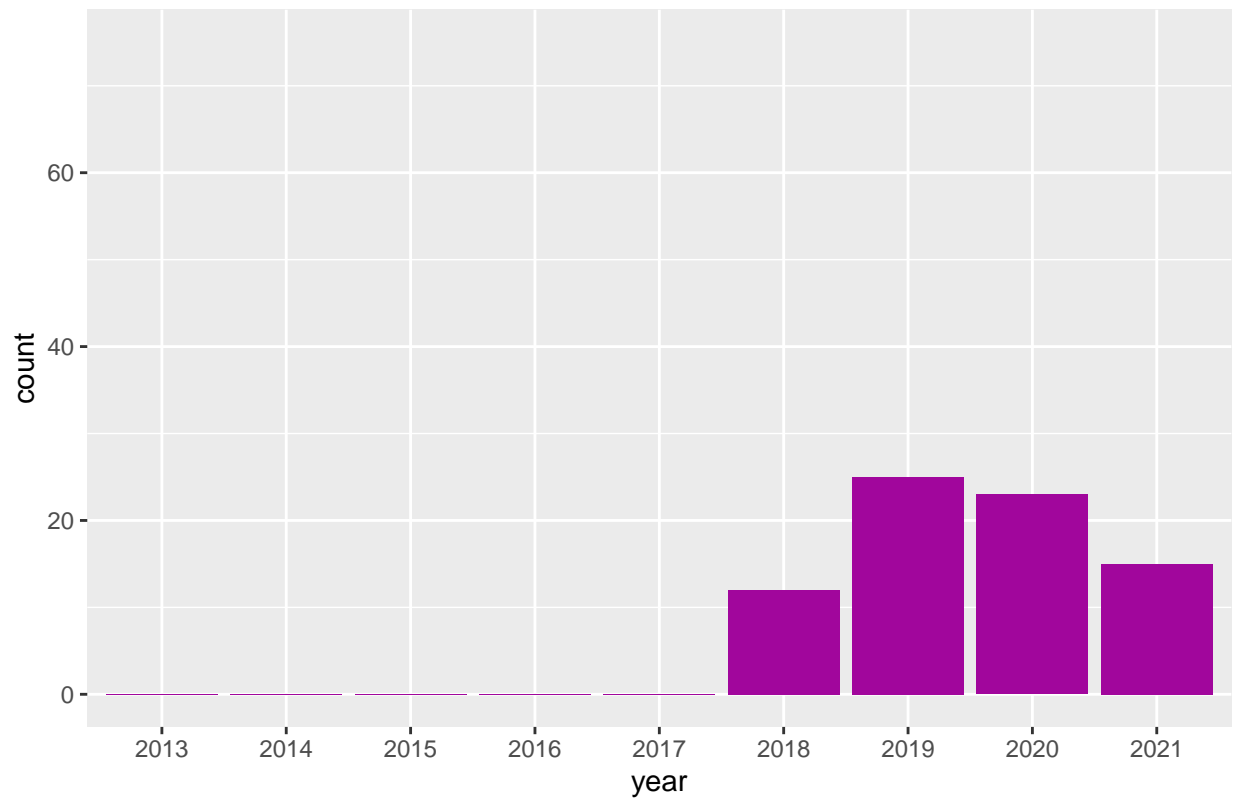
```
year.trend("Hack", "#a10654")
```

## Hack

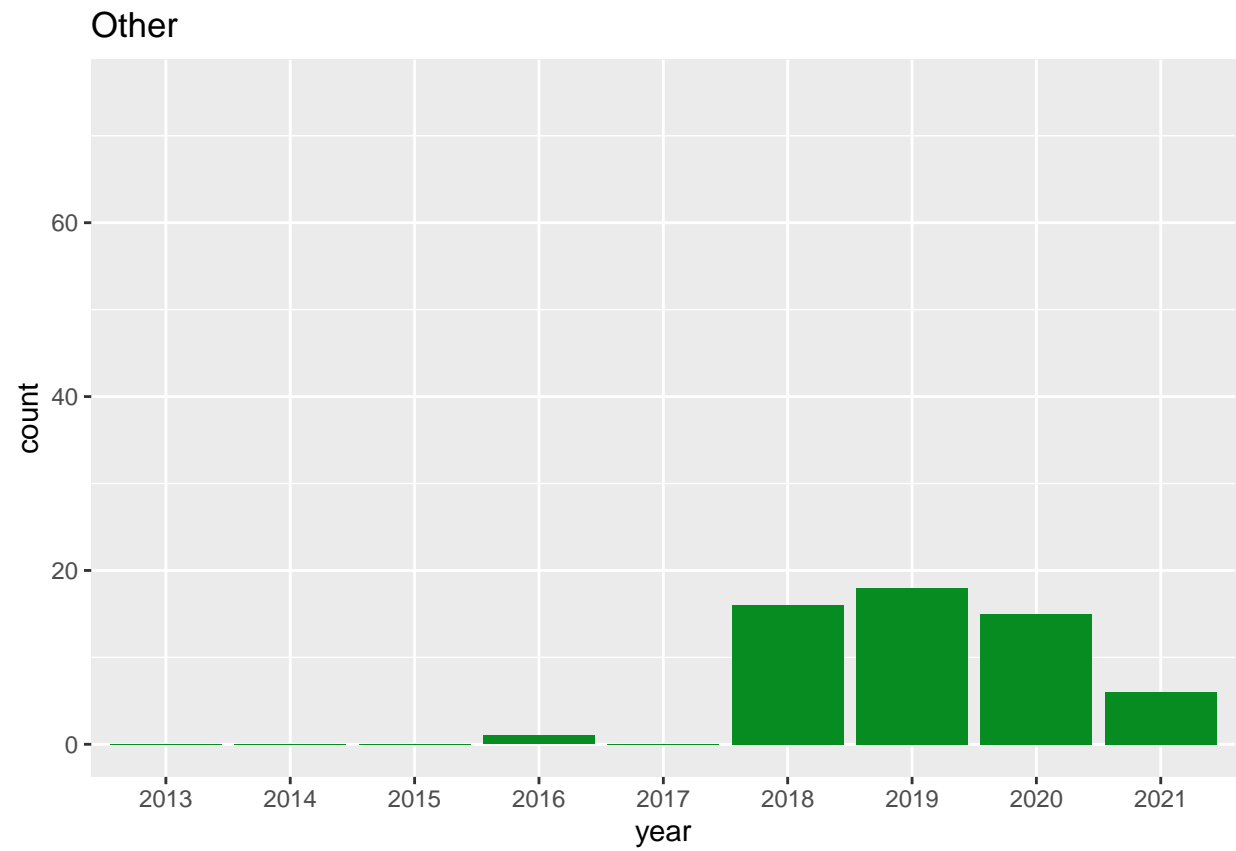


```
year.trend("Malware", "#a1069c")
```

## Malware

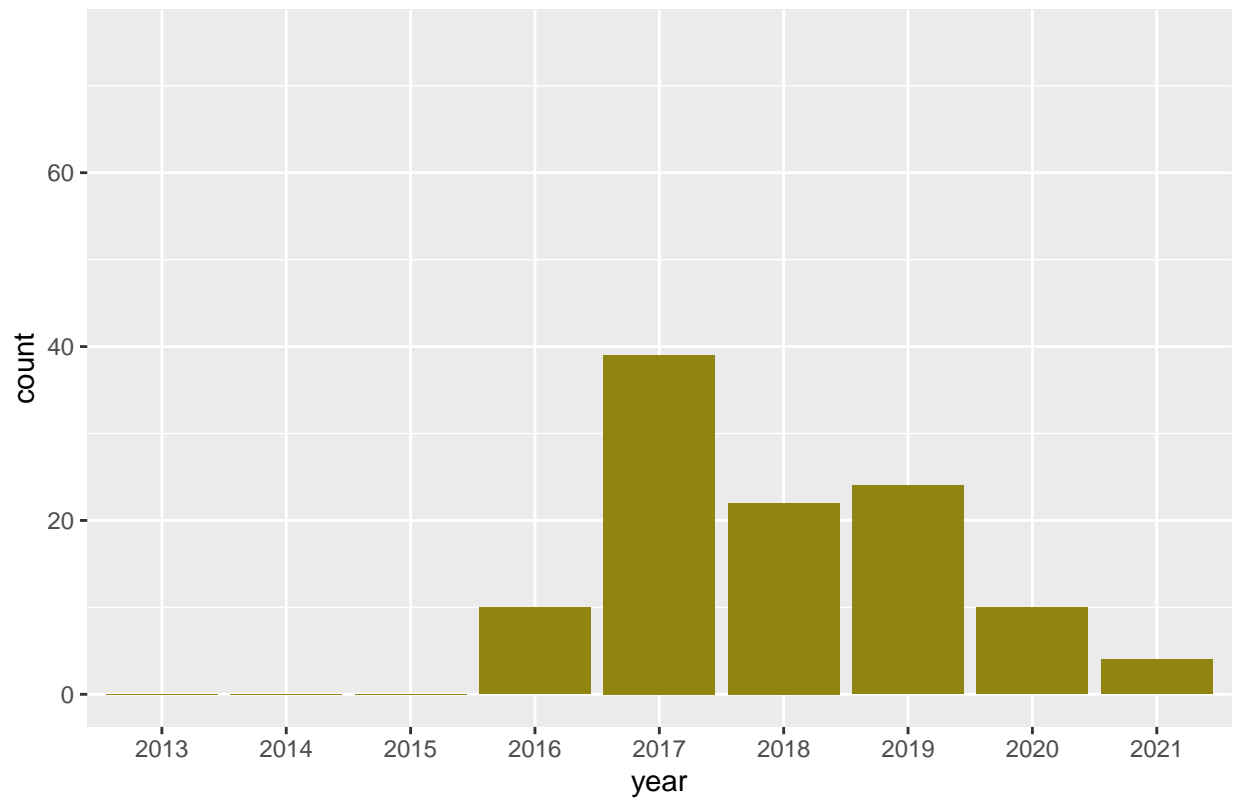


```
year.trend("Other", "#078c22")
```



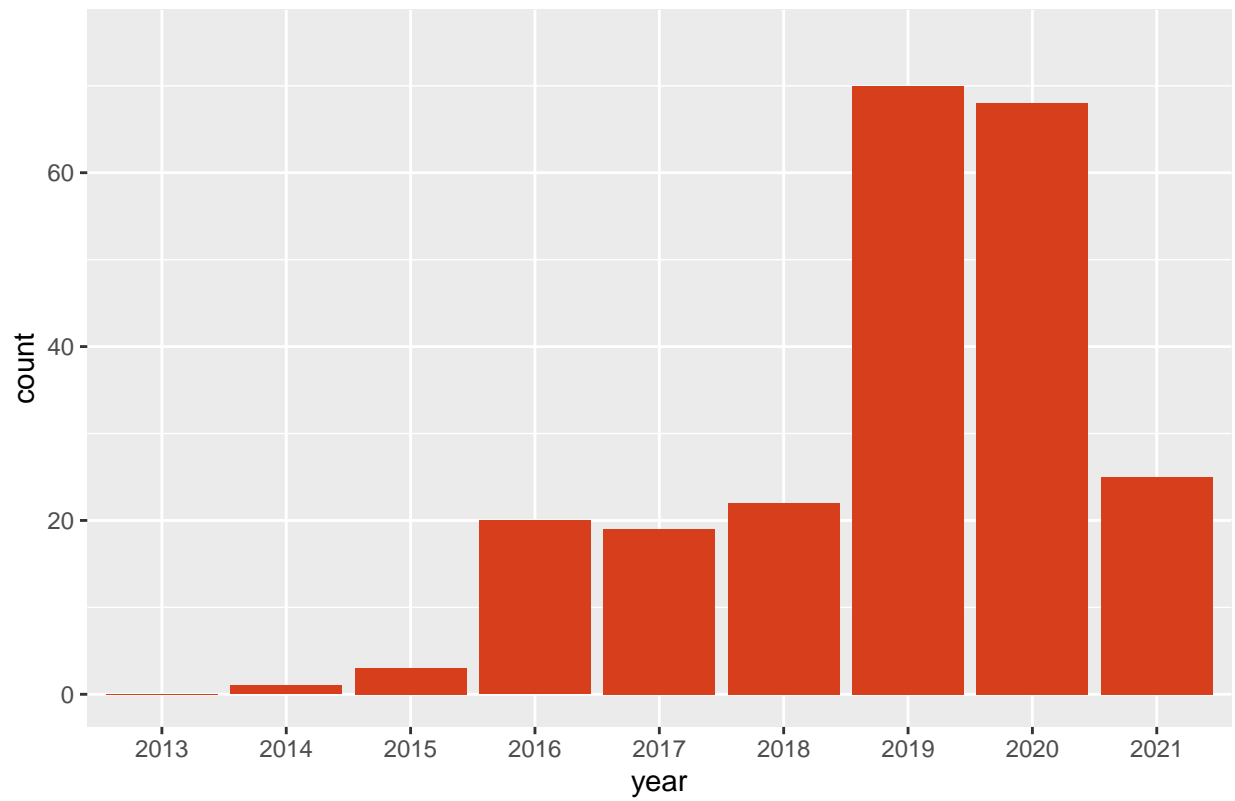
```
year.trend("Phishing", "#918511")
```

## Phishing



```
year.trend("Ransomware", "#d63e1c")
```

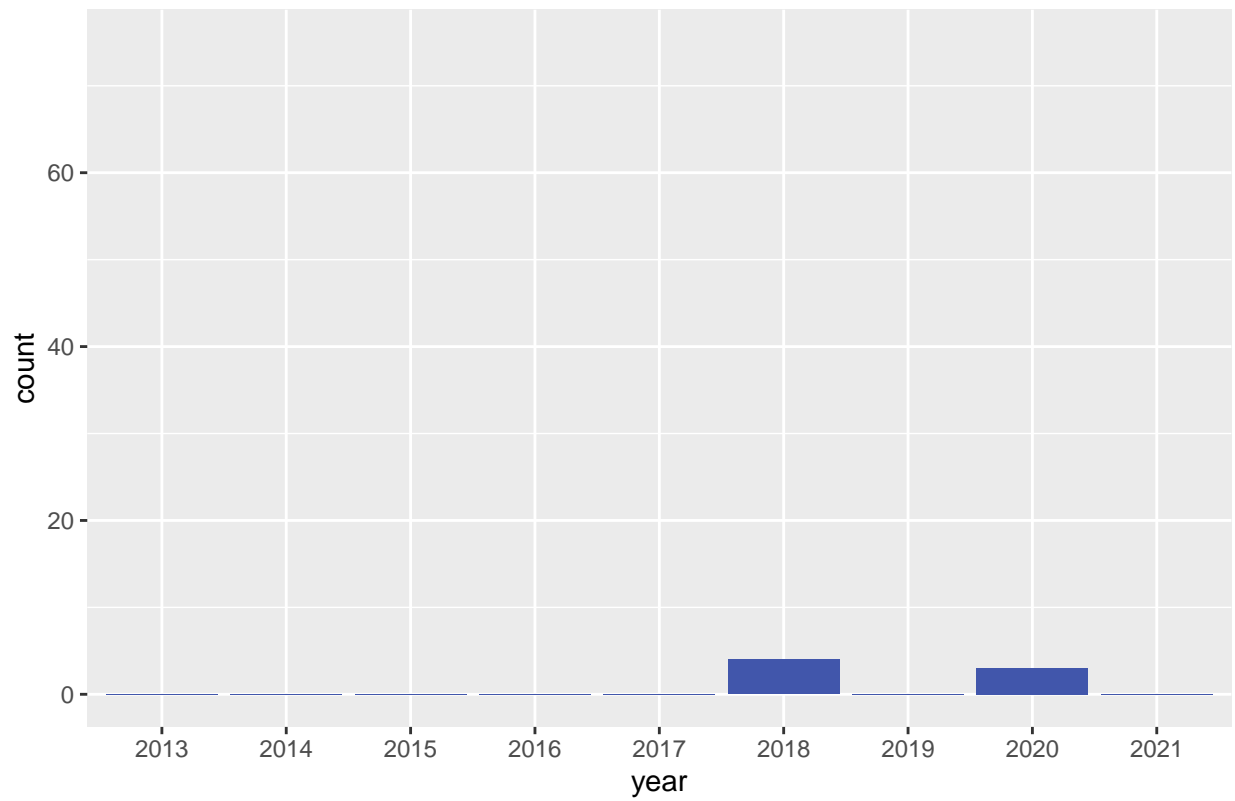
## Ransomware



```
year.trend("TDoS", "#4156ab")
```

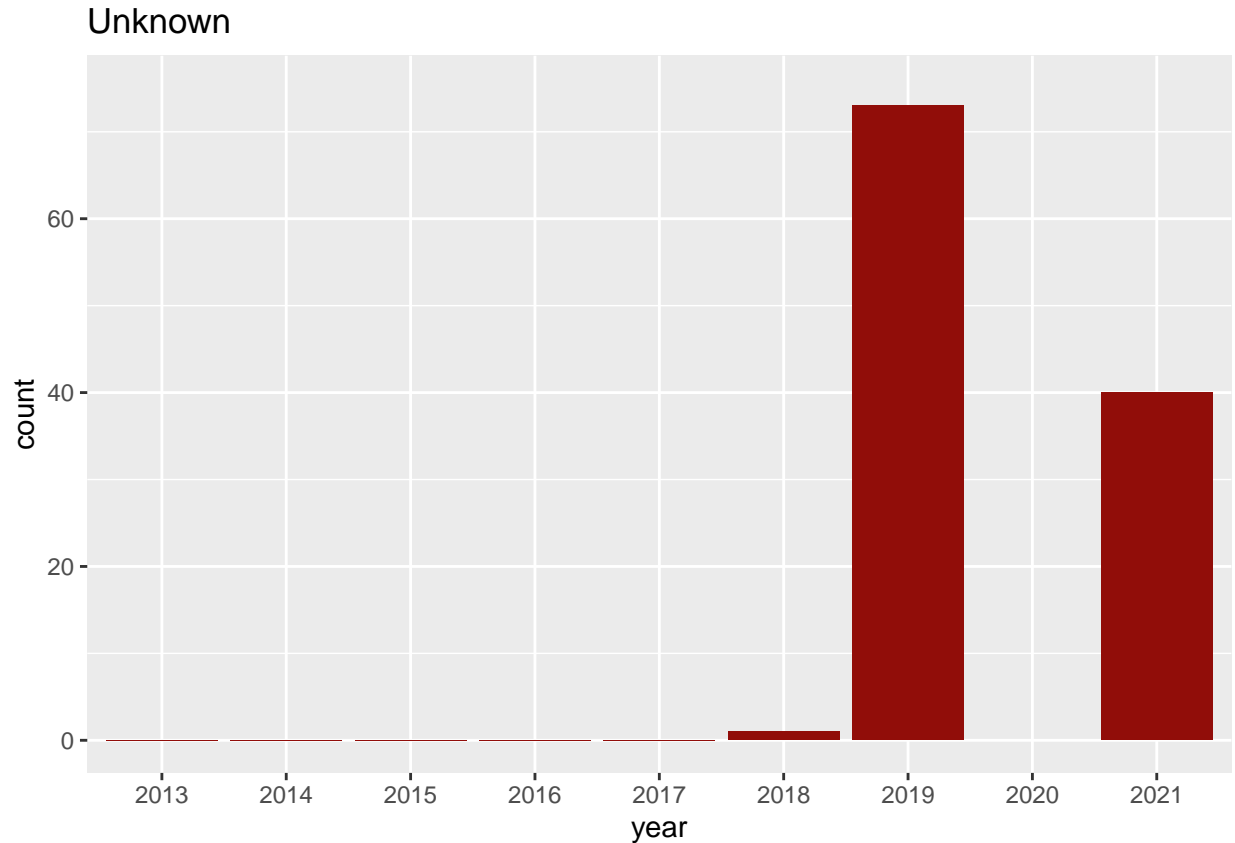


## TDoS



```
year.trend("Unknown", "#910d09")
```

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



*#subset for education and one for safety*

```
df.education <- subset(df.clean, type == "Education")
df.safety <- subset(df.clean, type == "Safety")
```

*#a work around to not have to change a bunch of small code;*

```
df.temporary <- df.clean
df.clean <- df.education
```

```
graph.attack.by.filter <- function(fill.choice){
  df.unk <- df.clean %>% as_tibble() %>%
  mutate(Unknown = as.factor(if_else(Unknown == 1, "True", "False")))
```

```
unk.graph <- ggplot(data = df.unk, aes(Unknown == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
```

#####

```
df.hack <- df.clean %>% as_tibble() %>%
  mutate(Hack = as.factor(if_else(Hack == 1, "True", "False")))
```

```
hack.graph <- ggplot(data = df.hack, aes(Hack == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
```

#####

```
df.data <- df.clean %>% as_tibble() %>%
  mutate(Data_Breach = as.factor(if_else(Data_Breach == 1, "True", "False")))
```

```
data.graph <- ggplot(data = df.data, aes(Data_Breach == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
```

#####

```

df.ran <- df.clean %>% as_tibble() %>%
  mutate(Ransomware = as.factor(if_else(Ransomware == 1, "True", "False")))

ran.graph <- ggplot(data = df.ran, aes(Ransomware == "True", fill = fill.choice))+geom_bar() + ylim(0, 1)
#####

df.phish <- df.clean %>% as_tibble() %>%
  mutate(Phishing = as.factor(if_else(Phishing == 1, "True", "False")))

phish.graph <- ggplot(data = df.phish, aes(Phishing == "True", fill = fill.choice))+geom_bar() + ylim(0, 1)
#####

df.tdos <- df.clean %>% as_tibble() %>%
  mutate(TDoS = as.factor(if_else(TDoS == 1, "True", "False")))

tdos.graph <- ggplot(data = df.tdos, aes(TDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####

df.mal <- df.clean %>% as_tibble() %>%
  mutate(Malware = as.factor(if_else(Malware == 1, "True", "False")))

mal.graph <- ggplot(data = df.mal, aes(Malware == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####

df.ddos <- df.clean %>% as_tibble() %>%
  mutate(DDoS = as.factor(if_else(DDoS == 1, "True", "False")))

ddos.graph <- ggplot(data = df.ddos, aes(DDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####

df.other <- df.clean %>% as_tibble() %>%
  mutate(Other = as.factor(if_else(Other == 1, "True", "False")))

other.graph <- ggplot(data = df.other, aes(Other == "True", fill = fill.choice))+geom_bar() + ylim(0, 1)
#####

df.serv <- df.clean %>% as_tibble() %>%
  mutate(Server_Shutdown = as.factor(if_else(Server_Shutdown == 1, "True", "False")))

serv.graph <- ggplot(data = df.serv, aes(Server_Shutdown == "True", fill = fill.choice))+geom_bar() + ylim(0, 1)
#####

(data.graph + scale_x_discrete(limit = c(TRUE)) | ddos.graph + scale_x_discrete(limit = c(TRUE)) |
  hack.graph + scale_x_discrete(limit = c(TRUE)))/ (mal.graph + scale_x_discrete(limit = c(TRUE)) |
  other.graph + scale_x_discrete(limit = c(TRUE)) | phish.graph + scale_x_discrete(limit = c(TRUE)))/
  (ran.graph + scale_x_discrete(limit = c(TRUE)) | tdos.graph + scale_x_discrete(limit = c(TRUE)) | unk
  scale_x_discrete(limit = c(TRUE)))
}

graph.attack.by.filter(df.clean$year)

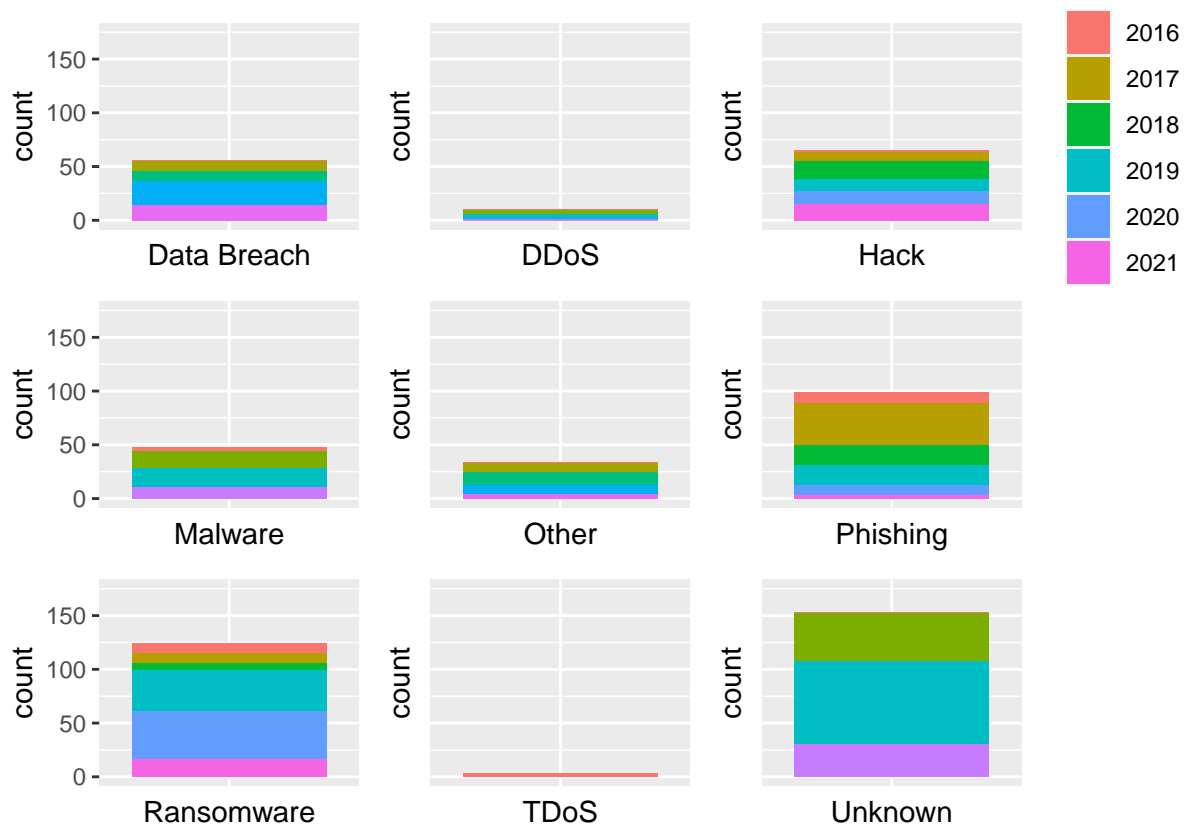
```

```
## Warning: Removed 291 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 337 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 282 rows containing non-finite values (stat_count).
```

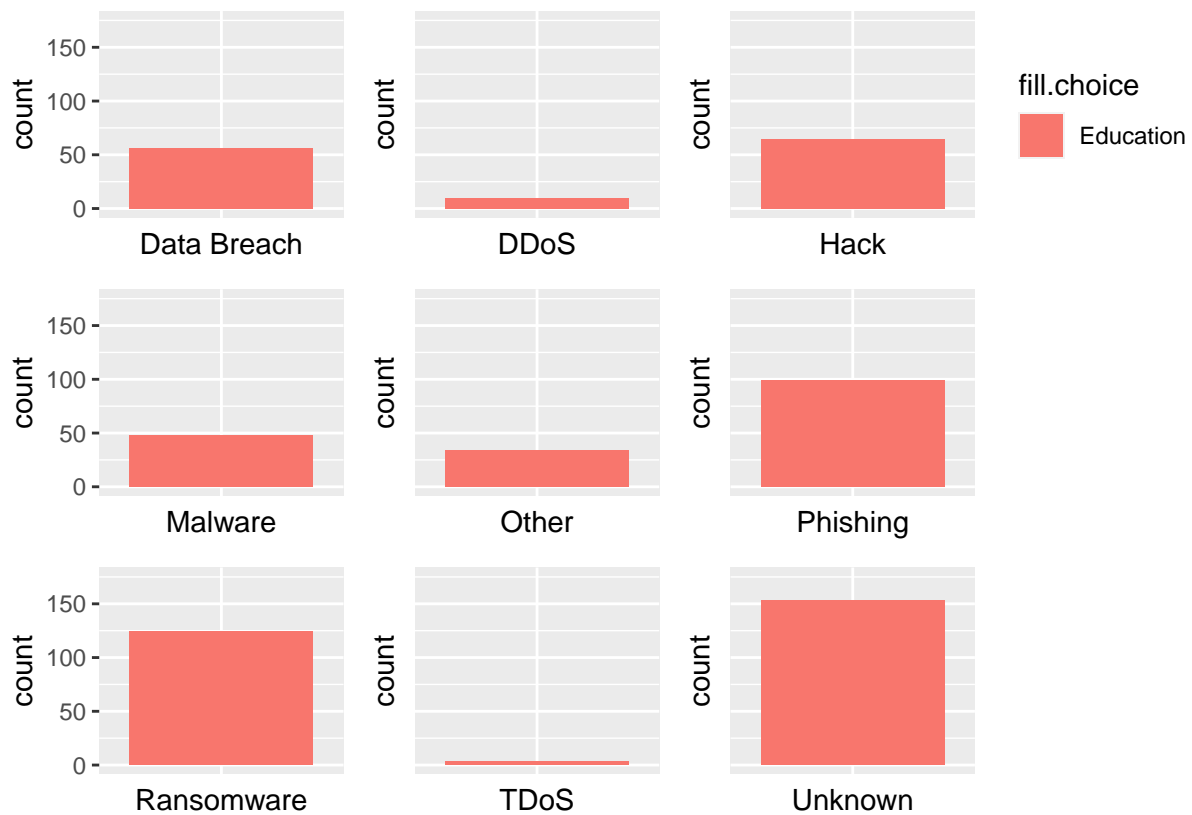
```
## Warning: Removed 299 rows containing non-finite values (stat_count).
## Warning: Removed 313 rows containing non-finite values (stat_count).
## Warning: Removed 248 rows containing non-finite values (stat_count).
## Warning: Removed 223 rows containing non-finite values (stat_count).
## Warning: Removed 344 rows containing non-finite values (stat_count).
## Warning: Removed 194 rows containing non-finite values (stat_count).
```



```
graph.attack.by.filter(df.clean$type)
```

```
## Warning: Removed 291 rows containing non-finite values (stat_count).
## Warning: Removed 337 rows containing non-finite values (stat_count).
## Warning: Removed 282 rows containing non-finite values (stat_count).
## Warning: Removed 299 rows containing non-finite values (stat_count).
## Warning: Removed 313 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 248 rows containing non-finite values (stat_count).
## Warning: Removed 223 rows containing non-finite values (stat_count).
## Warning: Removed 344 rows containing non-finite values (stat_count).
## Warning: Removed 194 rows containing non-finite values (stat_count).
```

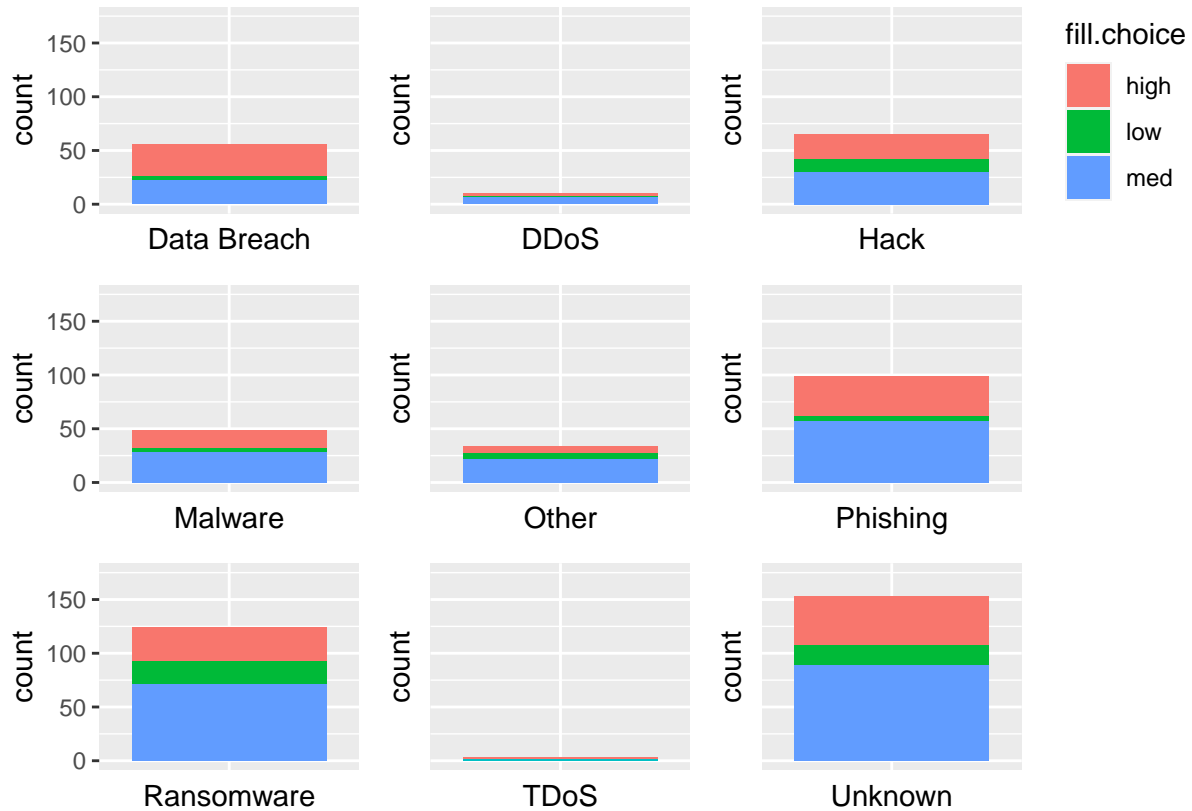


```
graph.attack.by.filter((df.clean$website_quality))
```

```
## Warning: Removed 291 rows containing non-finite values (stat_count).
## Warning: Removed 337 rows containing non-finite values (stat_count).
## Warning: Removed 282 rows containing non-finite values (stat_count).
## Warning: Removed 299 rows containing non-finite values (stat_count).
## Warning: Removed 313 rows containing non-finite values (stat_count).
## Warning: Removed 248 rows containing non-finite values (stat_count).
## Warning: Removed 223 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 344 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 194 rows containing non-finite values (stat_count).
```



```
df.clean <- df.temporary
```

```
#a work around to not have to change a bunch of small code;
```

```
df.temporary <- df.clean
df.clean <- df.safety
```

```
graph.attack.by.filter <- function(fill.choice){
  df.unk <- df.clean %>% as_tibble() %>%
  mutate(Unknown = as.factor(if_else(Unknown == 1, "True", "False")))
```

```
unk.graph <- ggplot(data = df.unk, aes(Unknown == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
```

```
#####
```

```
df.hack <- df.clean %>% as_tibble() %>%
  mutate(Hack = as.factor(if_else(Hack == 1, "True", "False")))
```

```
hack.graph <- ggplot(data = df.hack, aes(Hack == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
```

```
#####
```

```
df.data <- df.clean %>% as_tibble() %>%
  mutate(Data_Breach = as.factor(if_else(Data_Breach == 1, "True", "False")))
```

```

data.graph <- ggplot(data = df.data, aes(Data_Breach == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.ran <- df.clean %>% as_tibble() %>%
  mutate(Ransomware = as.factor(if_else(Ransomware == 1, "True", "False")))

ran.graph <- ggplot(data = df.ran, aes(Ransomware == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.phish <- df.clean %>% as_tibble() %>%
  mutate(Phishing = as.factor(if_else(Phishing == 1, "True", "False")))

phish.graph <- ggplot(data = df.phish, aes(Phishing == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.tdos <- df.clean %>% as_tibble() %>%
  mutate(TDoS = as.factor(if_else(TDoS == 1, "True", "False")))

tdos.graph <- ggplot(data = df.tdos, aes(TDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.mal <- df.clean %>% as_tibble() %>%
  mutate(Malware = as.factor(if_else(Malware == 1, "True", "False")))

mal.graph <- ggplot(data = df.mal, aes(Malware == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.ddos <- df.clean %>% as_tibble() %>%
  mutate(DDoS = as.factor(if_else(DDoS == 1, "True", "False")))

ddos.graph <- ggplot(data = df.ddos, aes(DDoS == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.other <- df.clean %>% as_tibble() %>%
  mutate(Other = as.factor(if_else(Other == 1, "True", "False")))

other.graph <- ggplot(data = df.other, aes(Other == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####
df.serv <- df.clean %>% as_tibble() %>%
  mutate(Server_Shutdown = as.factor(if_else(Server_Shutdown == 1, "True", "False")))

serv.graph <- ggplot(data = df.serv, aes(Server_Shutdown == "True", fill = fill.choice))+geom_bar() + ylim(0, 175)
#####

(data.graph + scale_x_discrete(limit = c(TRUE)) | ddos.graph + scale_x_discrete(limit = c(TRUE)) |
  hack.graph + scale_x_discrete(limit = c(TRUE)))/ (mal.graph + scale_x_discrete(limit = c(TRUE)) |
  other.graph + scale_x_discrete(limit = c(TRUE)) | phish.graph + scale_x_discrete(limit = c(TRUE)))/
  (ran.graph + scale_x_discrete(limit = c(TRUE)) | tdos.graph + scale_x_discrete(limit = c(TRUE)) | unk
  scale_x_discrete(limit = c(TRUE)))

}

graph.attack.by.filter(df.clean$year)

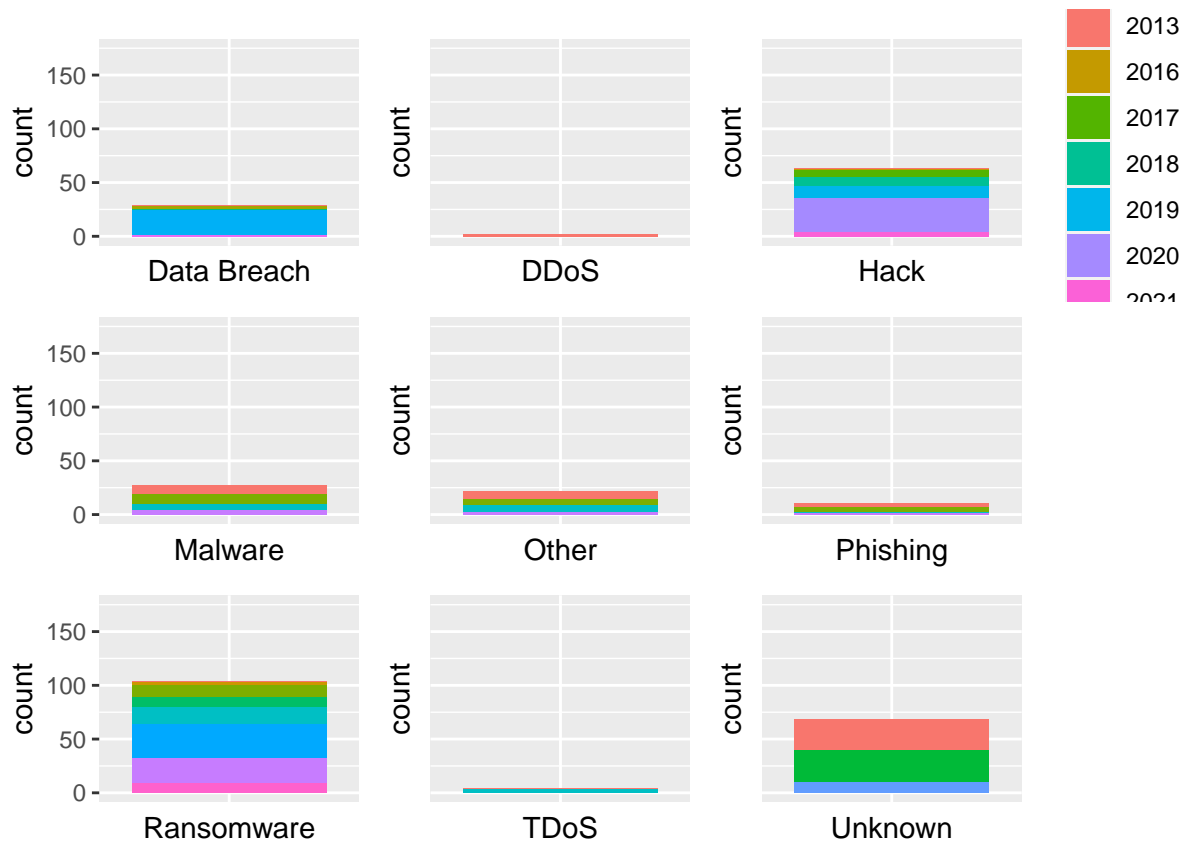
```

```
## Warning: Removed 170 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 197 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 136 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 172 rows containing non-finite values (stat_count).
## Warning: Removed 177 rows containing non-finite values (stat_count).
## Warning: Removed 189 rows containing non-finite values (stat_count).
## Warning: Removed 95 rows containing non-finite values (stat_count).
## Warning: Removed 195 rows containing non-finite values (stat_count).
## Warning: Removed 131 rows containing non-finite values (stat_count).
```



```
graph.attack.by.filter(df.clean$type)
```

```
## Warning: Removed 170 rows containing non-finite values (stat_count).
## Warning: Removed 197 rows containing non-finite values (stat_count).
## Warning: Removed 136 rows containing non-finite values (stat_count).
## Warning: Removed 172 rows containing non-finite values (stat_count).
## Warning: Removed 177 rows containing non-finite values (stat_count).
```



```
## Warning: Removed 189 rows containing non-finite values (stat_count).
## Warning: Removed 95 rows containing non-finite values (stat_count).
## Warning: Removed 195 rows containing non-finite values (stat_count).
## Warning: Removed 131 rows containing non-finite values (stat_count).
```

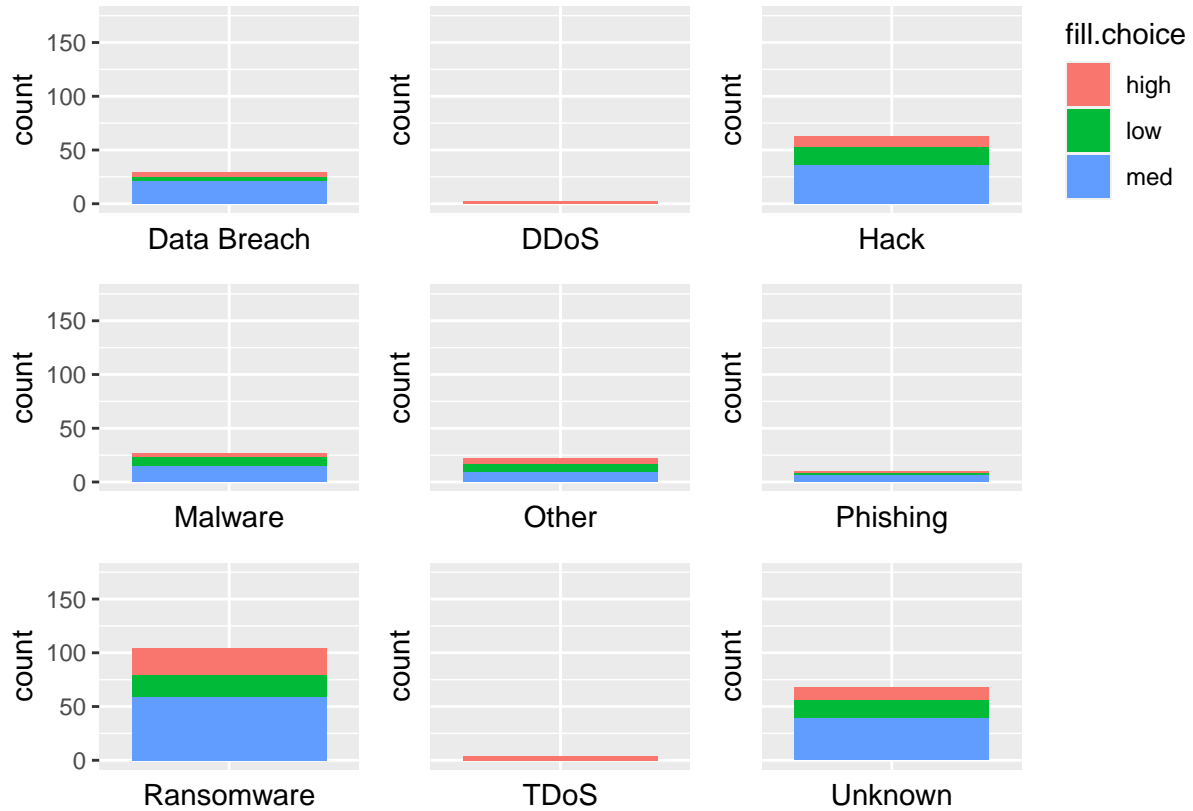


```
graph.attack.by.filter((df.clean$website_quality))
```

```
## Warning: Removed 170 rows containing non-finite values (stat_count).
## Warning: Removed 197 rows containing non-finite values (stat_count).
## Warning: Removed 136 rows containing non-finite values (stat_count).
## Warning: Removed 172 rows containing non-finite values (stat_count).
## Warning: Removed 177 rows containing non-finite values (stat_count).
## Warning: Removed 189 rows containing non-finite values (stat_count).
## Warning: Removed 95 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 195 rows containing non-finite values (stat_count).
```

```
## Warning: Removed 131 rows containing non-finite values (stat_count).
```



```
df.clean <- df.temporary
```

```
““
```