

Sandbox

Eren Akgunduz

20 April 2023

The data

```
library(tidyverse)
```

Clean it

```
df <- read.csv("data/all.csv", stringsAsFactors = FALSE)
df$Damage <- gsub(".*\\((\\d+)\\).*", "\\1", df$Damage)
df$Damage <- gsub("(\\d+)\\s*\\((\\d+)\\)", "pmax(\\1, \\2)", df$Damage)
df$Damage <- gsub("(\\d+)\\*(\\d+)", "\\1+\\2", df$Damage)
df$Damage <- gsub("(\\d+)x(\\d+)", "\\1*\\2", df$Damage)
df$Damage <- sapply(df$Damage, function(x) eval(parse(text = x)))
df$Damage <- as.numeric(df$Damage)
df$Stun <- gsub(".*\\((\\d+)\\).*", "\\1", df$Stun)
df$Stun <- gsub("(\\d+)\\s*\\((\\d+)\\)", "pmax(\\1, \\2)", df$Stun)
df$Stun <- gsub("(\\d+)\\*(\\d+)", "\\1+\\2", df$Stun)
df$Stun <- gsub("(\\d+)x(\\d+)", "\\1*\\2", df$Stun)
df$Stun <- sapply(df$Stun, function(x) eval(parse(text = x)))
df$Stun <- as.numeric(df$Stun)
df$health <- as.numeric(df$health)
df$stun <- as.numeric(df$stun)
df$vgauge1 <- as.numeric(df$vgauge1)
df$vgauge2 <- as.numeric(df$vgauge2)
df$fDash <- as.numeric(df$fDash)
df$bDash <- as.numeric(df$bDash)
df$fWalk <- as.numeric(df$fWalk)
df$bWalk <- as.numeric(df$bWalk)
df$throwHurt <- as.numeric(df$throwHurt)
df$throwRange <- as.numeric(df$throwRange)
na_rows <- df[!complete.cases(df), ]
na_rows
write.csv(df, "data/all.csv", row.names = FALSE)

df <- read.csv("data/all.csv", stringsAsFactors = FALSE)

# Replace values in parentheses with the minimum value
regx <- ".*?(\\-?\\d+)\\((\\-?\\d+)\\).*"
df$onBlock <- gsub(regx, "\\1", df$onBlock)
df$onBlock <- gsub("(\\-?\\d+)[\\*/](\\-?\\d+)", "\\1", df$onBlock)

# Use ifelse to replace NA values w/ NA and non-NA values with the minimum value
df$onBlock <- ifelse(
```

```

is.na(df$onBlock), # condition: check if value is NA
NA, # value to replace with if condition is TRUE
# value to replace with if condition is FALSE
pmin(as.numeric(df$onBlock), as.numeric(gsub(regex, "\\2", df$onBlock)))
)

# Convert column to numeric
df$onBlock <- as.numeric(df$onBlock)

write.csv(df, "data/all.csv", row.names = FALSE)

df <- read.csv("data/all.csv", stringsAsFactors = FALSE)
# Check for missing values
na_rows <- df[!complete.cases(df), ]
na_rows

## [1] Character Move      onBlock    plnCmd     airmove    followUp
## [7] projectile moveType  health     stun       vgauge1    vgauge2
## [13] fDash      bDash      fWalk      bWalk      throwHurt  throwRange
## [19] Damage     Stun
## <0 rows> (or 0-length row.names)

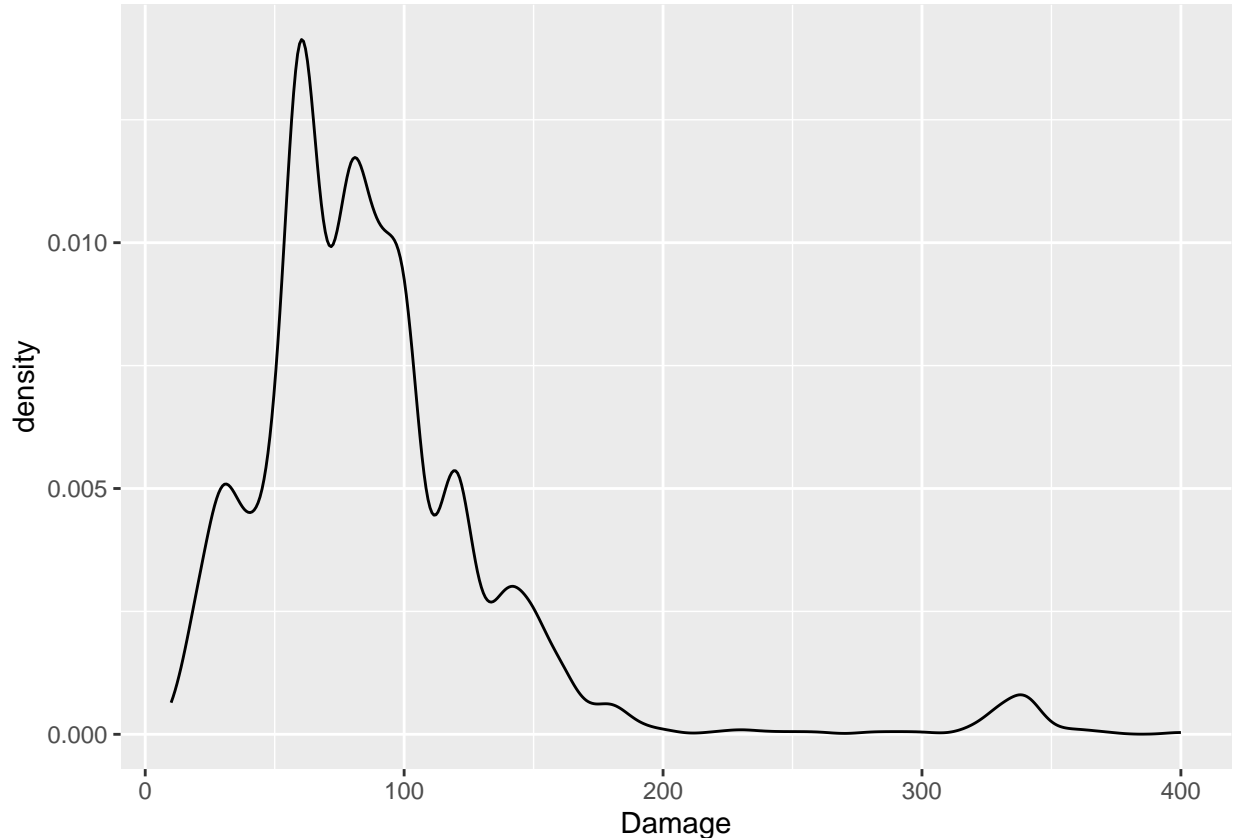
```

KDE

```

ggplot(df, aes(Damage)) +
  geom_density()

```



```
ggplot(df, aes(Stun)) +  
  geom_density()
```

