Project Proposal

2023-11-30

```
library(readr)
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(knitr)
library(ggplot2)


data <- read_csv("data.csv")
```

```
## Rows: 6012 Columns: 144
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (8): R_fighter, B_fighter, Referee, location, Winner, weight_class, B...
## dbl  (134): B_avg_KD, B_avg_opp_KD, B_avg_SIG_STR_pct, B_avg_opp_SIG_STR_pct...
## lgl    (1): title_bout
## date   (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data <- subset(data, !(weight_class %in% c("CatchWeight", "WomenBantamweight", "WomenFeatherweight", "W
```

```
data <- data %>% mutate(weight = case_when(weight_class == "Flyweight" ~ 125,
                                           weight_class == "Bantamweight" ~ 135,
                                           weight_class == "Featherweight" ~ 145,
                                           weight_class == "Lightweight" ~ 155,
                                           weight_class == "Welterweight" ~ 170,
                                           weight_class == "Middleweight" ~ 185,
                                           weight_class == "LightHeavyweight" ~ 205,
```

```
                                    weight_class == "Heavyweight" ~ 265,
                                    weight_class == "OpenWeight" ~ 300))

data <- data %>% mutate(new_class = case_when(weight_class == "Flyweight" ~ "Class 1",
                                    weight_class == "Bantamweight" ~ "Class 1",
                                    weight_class == "Featherweight" ~ "Class 1",
                                    weight_class == "Lightweight" ~ "Class 2",
                                    weight_class == "Welterweight" ~ "Class 2",
                                    weight_class == "Middleweight" ~ "Class 3",
                                    weight_class == "LightHeavyweight" ~ "Class 3",
                                    weight_class == "Heavyweight" ~ "Class 4",
                                    weight_class == "OpenWeight" ~ "Class 4"))

freq <- data %>% group_by(new_class, weight_class) %>%
  summarize(weight = median(weight),
            count = n()) %>% arrange(weight)
```

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

```
kable(freq, caption = "Frequency of Weight Class Fights")
```

Table 1: Frequency of Weight Class Fights

new_class	weight_class	weight	count
Class 1	Flyweight	125	230
Class 1	Bantamweight	135	475
Class 1	Featherweight	145	551
Class 2	Lightweight	155	1091
Class 2	Welterweight	170	1083
Class 3	Middleweight	185	813
Class 3	LightHeavyweight	205	573
Class 4	Heavyweight	265	585
Class 4	OpenWeight	300	86

```
new_class_freq <- data %>% group_by(new_class) %>% summarize(freq = n())

ggplot(data=new_class_freq, aes(x=new_class, y=freq)) +
  geom_bar(stat="identity", fill="steelblue")+
  theme_minimal() + labs(x = "New Class for Model Building", y = "Count",
                         title = "Frequency Distribution of New Weight Classes")
```


Frequency Distribution of New Weight Classes
Count
2000
1500
1000
500
0
Class 1
Class 2
Class 3
Class 4
New Class for Model Building