Ultimate Fighting Championship Statistics

Data Analysis Report

Author: Harun Tucaković   Mentor: Emin Tahirović

January 2021, Sarajevo

Table of Contents

1. Introduction

This document is final data analysis report for study done as a project for CS498 – Computing with Data course at International University of Sarajevo. All the code, graphics and conclusions within this document are original content. References to original data and the full R code that was used during the project can be found in the Section 5 ("References") of the document.

Topic of this study is possible marketing strategies of Ultimate Fighting Championship (UFC). Analysis of possible marketing strategies in our study is solely based on fights statistics and discovering appropriate audiences for future fights. Data used in this study is scraped from ufcstats.com website, official website that tracks fight statistics for UFC. Data consists of fight statistics from every fight organized under this promotion company since its creation in November 1993. By the term fight statistics in this paper, it is meant important parameters in the fight like number of significant strikes, knockdowns, takedowns, amount of control time, etc.

Goal of the study was to find out how are weight divisions related to fight outcomes and win methods in UFC, what fight parameters influence which win methods and how time format of the fight influences fight parameters and fight outcome. Close connection between these questions and UFC marketing will be discussed and explained later in the document.

This document is divided into four main sections: introduction, data, analysis, and conclusion. Section 1 (Introduction) is addressing study summary and data analysis report structure. Section 2 (Data) introduces and explains dataset used in this study. Here, we are also covering data gathering methods and process. Section 3 (Analysis) talks about actual data analysis process, explaining methods used to draw conclusions from data, explaining findings and supporting those findings with graphical proofs. Section 4 (Conclusions) presents final thoughts on findings achieved through this study. In this section, final connection is made between questions asked above and UFC marketing strategies. Also, plans for future studies on this dataset is briefly explained here.

2. Data

As mentioned in documents introduction, data used in this study is dataset about UFC fight statistics scraped from website ufcstats.com. This is official website for collecting and publishing UFC fight statistics. Scraping process will be explained in the subsection below. Here, we will briefly go over dataset format and contents.

Dataset contains every fight that happened under UFC promotion since November 1993. up to January of 2021. In that period 5917 fights happened (5917 observations in the dataset). Dataset includes 302 variables. Main variables, that were the most important for our study are fight date details, fight time format and division, fighter information, winner of the fight, number of significant strikes (attempted and landed as separate variables), number of takedowns, number of knockdowns, control time, etc. All the variables that capture fight statistics information are recorded as fight total per fighter and every variable per round per fighter. Although per round data is not used in this data analysis study, it was recorded during data scraping process due to plans for future data analysis studies that will be performed on this dataset. For detailed explanation of each variable in dataset, reader is advised to visit Appendix A.

2.1 Data scraping

Script for scraping data from the web is written in R using "rvest" package. This script also does most of data cleaning, formatting, and preparation for data analysis. For string and data manipulation and formatting following external R packages were used: stringr, sjmisc, plyr. In this subsection we will go through general idea of web-scraping process step by step.

First step in web-scraping process was to get html markdown of the ufcstats.com webpage. We achieve this with help of xml2 package and 'read_html' function.

```r
page_url <- 'http://ufcstats.com/statistics/events/completed?page=all'
ufc_stats <- xml2::read_html(page_url)
```

When html markdown is loaded, we extract URLs to every single UFC event from it. That is done by function called 'get_list_of_events', which takes as an argument html markdown. This function will return character vector containing URL links.

```r
get_list_of_events <- function(ufc_stats_html) {
  table_content <- ufc_stats_html %>%
    html_nodes('i.b-statistics__table-content')

  links <- lapply(table_content, function(element) element %>% html_node('a') %>%
 html_attr('href'))
  return(links)
}
```

With vector of URL links to every UFC event prepared, we call 'scrape_data' function and pass that vector to it. Function 'scrape_data' is main scraping function. It will call many other utility functions that are responsible for scraping, cleaning, and formatting data of each fight from each event. Firstly, dataframe that will be used for storing data is initialized to NULL. Then we iterate through vector of URLs and pass each one to the 'scrape_single_event' function, which is responsible for scraping data from all fights that happened on that event. Dataframe that we get back from 'scrape_single_event' function will be concatenated to the main dataframe.

'scrape_single_event' function will then call multiple utility functions to extract event title, event date, event location, and vector of URLs to every fight from that event. From here, we pass those URLs one by one to the 'get_fight_dataframe' function like this:

```r
for(link in fights_links_list) {
    temp_df <- get_fight_dataframe(link)
    temp_df <- cbind(event_title = event_title, event_date = event_date, event_lo
cation = event_location, temp_df)

    if (is.null(df)) {
      df <- temp_df
    } else {
      df <- rbind.fill(df, temp_df)
    }
  }
```

When all fights on that event are scraped, after 'for' loop we can see above, we return dataframe 'df' containing data about that single event. Function 'get_fight_dataframe' is responsible for scraping, gathering, cleaning, and formatting all the data about single fight. There are three types of information that need to be scraped from the webpage:

1. general fight data - information about fighters, winner, loser, judges, referee, time format, etc.,
2. fight statistics overview – number of significant strikes, takedowns, knockdowns, striking accuracy, takedown accuracy, etc.,
3. detailed fight statistics – number of significant strikes by type (head, body, leg), number of significant strikes by position (distance, clinch, ground).

This data is scraped by following functions respectively: 'get_main_fight_detailes', 'get_fight_totals_datarows', 'get_fight_sigstr_datarows'. Vectors that are returned from these functions are concatenated. Final vector is given column names and returned from 'get_fight_dataframe' function.

```r
get_fight_dataframe <- function(fight_link) {
  fight_html <- xml2::read_html(fight_link)
  main_vec <- get_main_fight_details(fight_html)
  totals_vec <- get_fight_totals_datarows(fight_html = fight_html)
  sigstr_vec <- get_fight_sigstr_datarows(fight_html = fight_html)
```

5

```r
  df_col_names <- c(names(main_vec), names(totals_vec), names(sigstr_vec))
  df_col_values <- c(main_vec, totals_vec, sigstr_vec)
  df <- data.frame(df_col_values)
  df <- as.data.frame(t(df))
  rownames(df) <- c(1)
  colnames(df) <- df_col_names
  return(df)
}
```

We will not go into more details about data scraping here. For more information on this topic and detailed analysis of the code, reader can refer to the Section 5 ("References") and check out the code on the GitHub repository.

3. Analysis

3.1 Marketability of UFC weight divisions

This section discusses marketability of different UFC weight divisions. We will try to find appropriate audience for each weight division and discover what division characteristics can be promoted to attract correct type of audience. To reach these conclusions we will take a look at these parameters: divisions activity, most frequent win methods for each division, fighters striking activity and output and control time. We want to remind the reader that R code for plotting any of the figures that are presented in this section is available in Appendix B.

3.1.1   Analysis

Before any analysis begins, some data restructuring is necessary. Firstly, variables that measure control time in a fight have format "mm:ss". This is not appropriate format for data comparison and manipulation. Hence, we will convert these variables to hold amount of control time in seconds. This is done by usage of custom 'convert_ctrl_time' function. Code of this function is available in Appendix B, if the reader is interested in more detailed analysis.

Secondly, we had to reformat 'division' column. Simple inspection of the column data with 'table' function reveled that column should be converted to 'factor' column and that some of the possible 'factor' levels have different names but very similar or the same meaning. Hence, a simple regrouping of data was necessary to form 'factor' levels that do not overlap and to drop data that is not going to be used in these studies. Additionally, new column is added which marks each row (fight) whether it is male of female division.

With data prepared, we can start analyzing data. One of the most important information about weight divisions is each divisions activity. This is achieved by comparing number of fights that happened in each division, as well as getting the proportion of female and male fights.
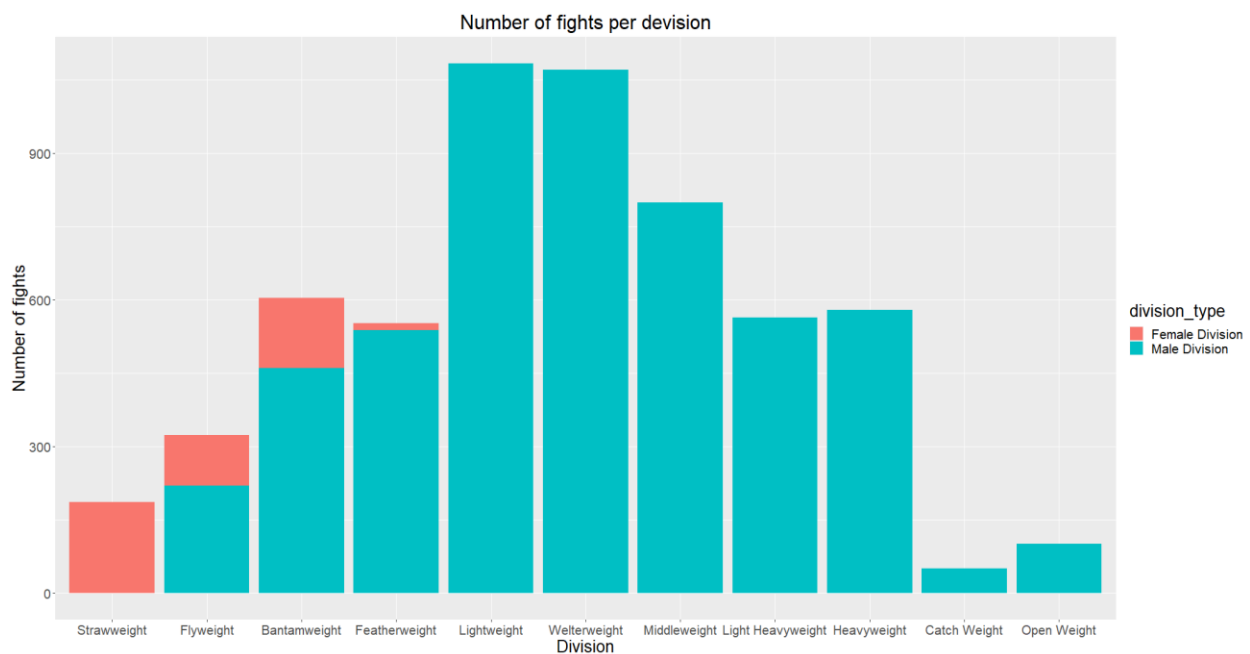

Figure 1: Number of fights per division

7

Figure 1. shows bar chart with UFC weight divisions on x-axis as bars and number of fights that happened in each division determines height of its bar. From this plot we can read that most fights in UFC happen in lightweight and welterweight male divisions, while most female fights happen in strawweight division. We can also observe proportions of fights that take place in male and female divisions.

For better insight of proportions between fights that happen in male and female divisions, we represented this data as a pie chart. From Figure 2. we can read that only 7.6% of fights in UFC are in female divisions.
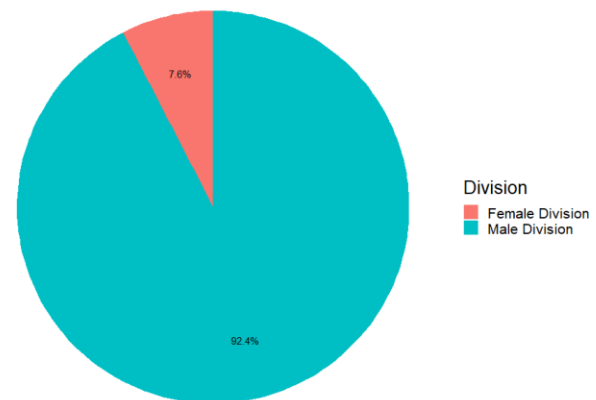

Figure 2: Proportion of fights based on division type

This study finds interesting and reveling analysis of win method frequencies over all UFC weight divisions. We want to see which win methods are most frequent in which division. Four main win methods that we are interested in are: "Decision – Split", "Decision – Unanimous", "KO/TKO", and "Submission". We are focusing on these win methods because they are the most exciting ways to end a fight and they increase crowd engagement in fight stories, which are very important marketing parameters. This is examined by plotting count of each win method in every division.
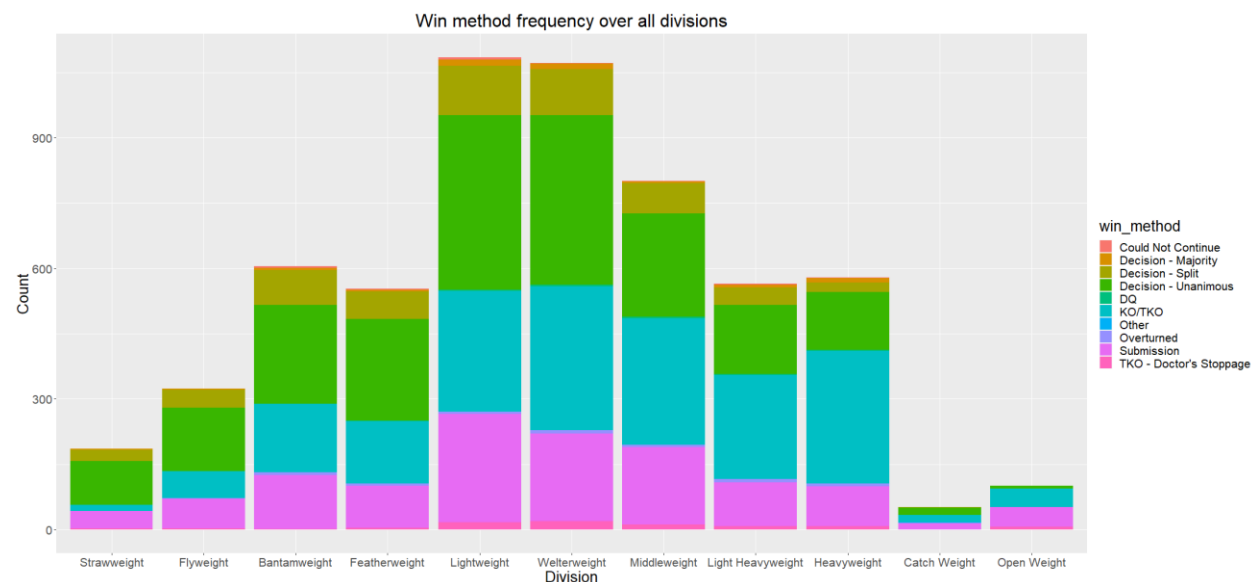

Figure 3: Win method frequency over all divisions

Figure 3. is representing number of each win method for every division in UFC. The plot above tells us that the biggest number of unanimous decision victories happen in lightweight and welterweight division, the most submission victories happen in lightweight. It is interesting to note that even with smaller number of fights in heavyweight division, there is much more KO/TKO victories then in rest of divisions. One logical reason for high number of KO/TKO victories in heavyweight division would be that bigger and heavier fighters are much stronger. And stronger the punch, the higher is probability of it knocking out the

opponent. Even though Figure 3. is useful in its own way, it is not the best for directly comparing divisions. Hence, plot in Figure 4. shows proportions of win methods in each division.
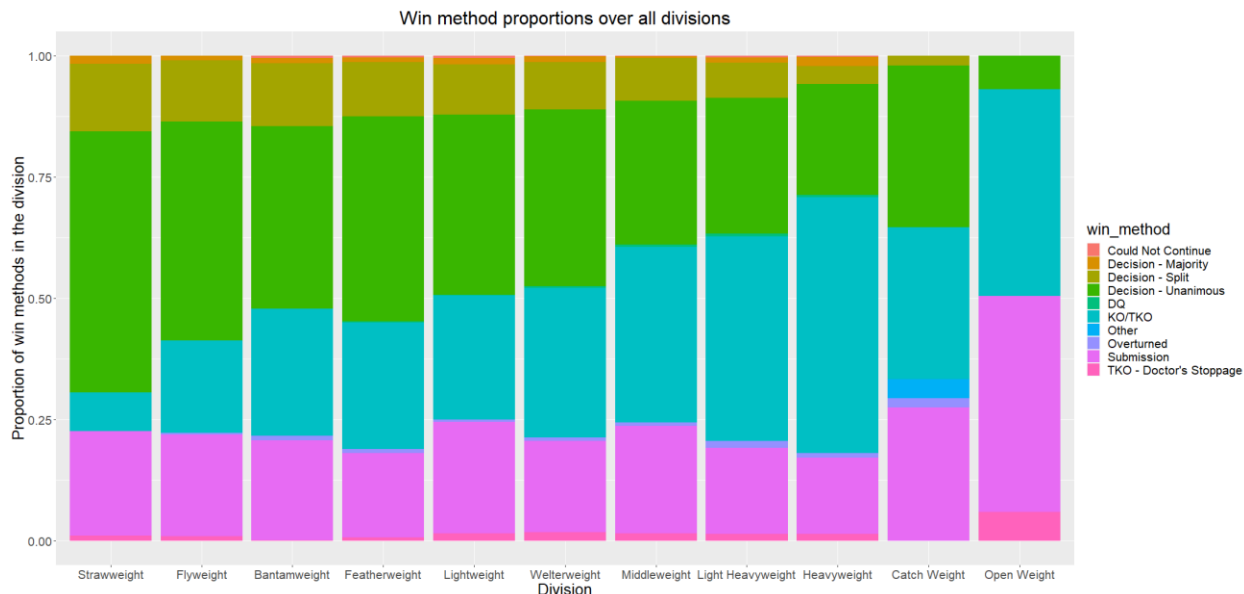

Figure 4: Win method proportions over all divisions

Plot on Figure 4. is much more reveling with respect of comparison of win method frequencies over UFC weight divisions. Here we can clearly see differences in how most of the fight are finished in each division. If we focus on four, previously mentioned, important win methods we can note that decision wins frequency heavily decreases if we read the plot from left to right. That means that the higher weight division, the smaller is the probability that the winner is going to be picked by the judges. As decision wins proportions are going down while we climb into heavier divisions, we can also notice that proportion of KO/TKO wins is steadily increasing. As mentioned before, this increase is heavily influenced by the power that heavier fighters have. Interesting thing to point out is that proportion of submission wins over divisions is very similar, with a little bigger proportion of submission wins in lighter divisions. There is one more plot on Figure 12. that readers can find in Appendix C, which shows same data plotted in another way that may help with better understanding and reading of plotted data.

One of the things that makes UFC fights exceptionally interesting is high striking work rate of fighters. Fights that have a lot of action (big number of strikes attempted and landed) attract a lot of attention of UFC fans and casual viewers. For that reason, we must treat amount of striking in the fight as crucial insight for marketing strategies. In this section we are analyzing UFC weight divisions, therefore we are analyzing striking rates over divisions. We are interested to learn which weight divisions are most attractive from striking output perspective. If we examine Figure 6. we see plot of total significant strikes thrown in a single fight (by both fighters), for every division. Boxplot of each division will tell us where most of the data lies, averages and outliers. Most important information we can read from plot on Figure 5. is that significant strikes output in a fight is drastically going down as we go from lighter to heavier divisions. Average strikes landed in a fight in lightest division is almost 100% higher in comparison to heaviest division. This phenomenon is also explainable by fighters' weight. Heavier the fighter is punches are more exhausting and take more energy, moving around takes more energy, therefore heavier fighters have much smaller striking output.
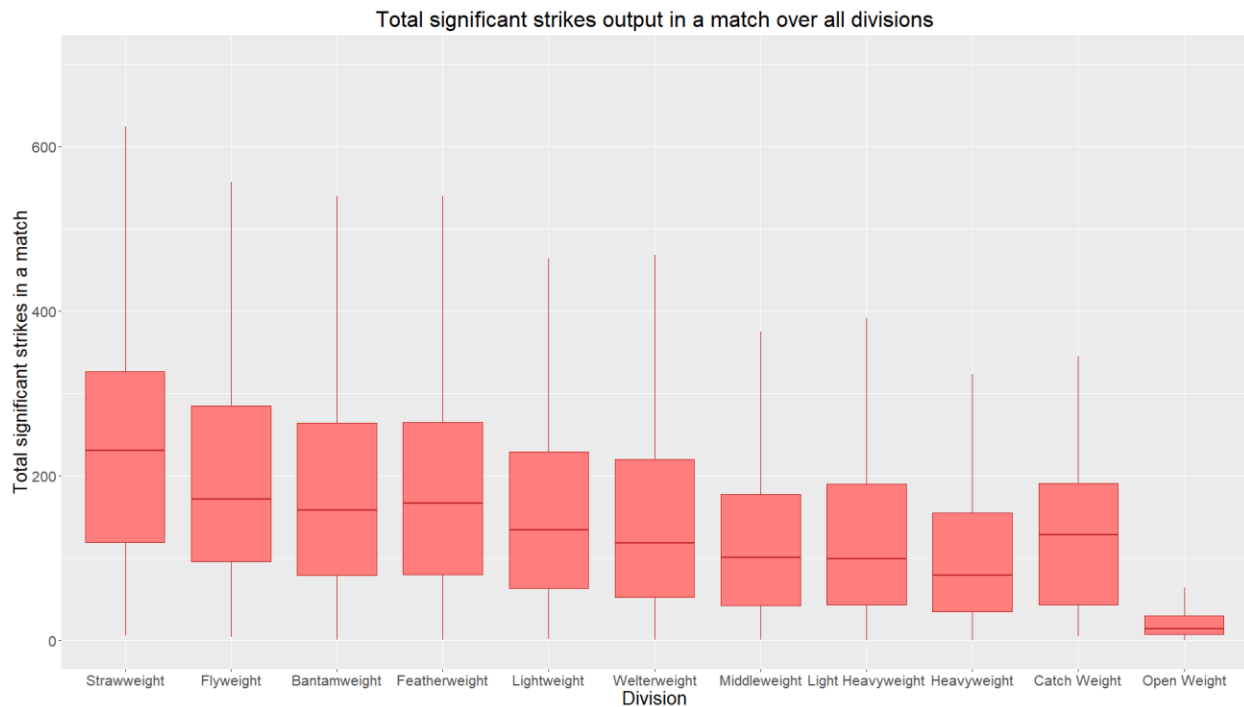
Figure 5: Total significant strikes output in a fight over all divisions

Besides the striking, another important parameter in fighting is control time. It is especially important for audience that is interested in grappling and wrestling aspects of UFC fights. If we do same analysis for control time as we did for striking we should get an idea is there difference in fighters grappling and wrestling preference based on their weight division.
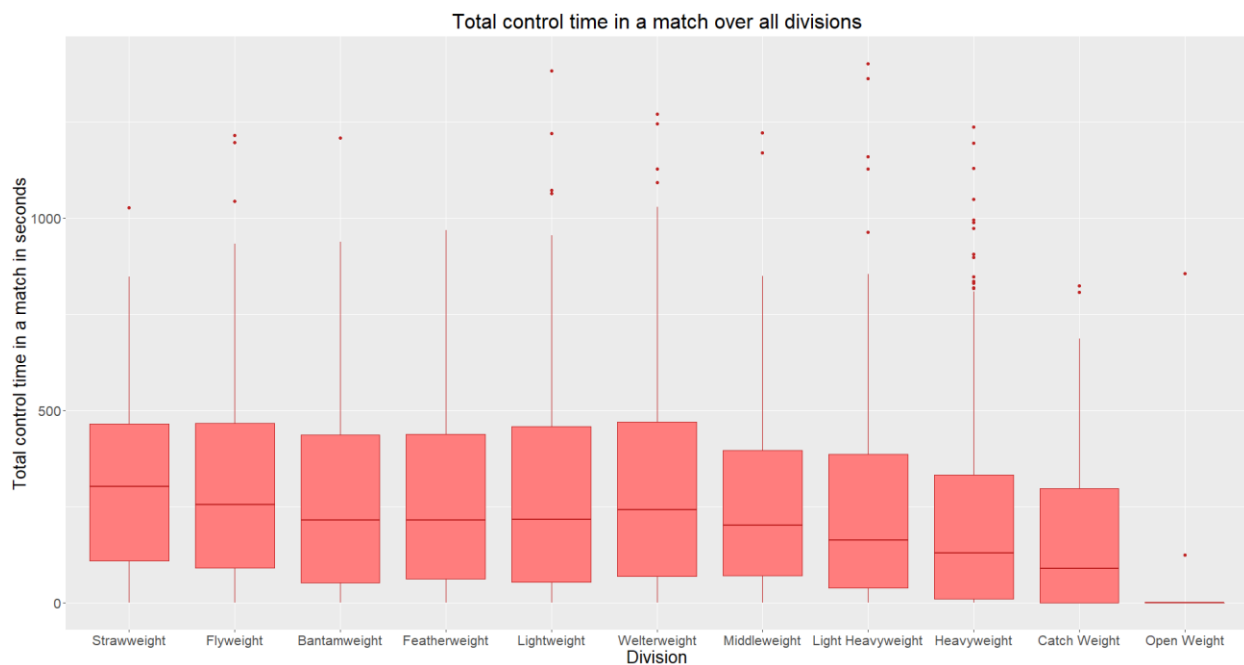

Figure 6: Total control time in a match over all divisions

Analysis of Figure 6. shows that there is not much difference in control time based on fighters' weight division. There is some difference, mainly in average control time in fights in light heavyweight and heavyweight divisions. It seems that heaviest guys do not prefer to grapple and wrestle, and if they do, those exchanges do not last for long. We also see that lightweights and welterweights have the highest range. Meaning that data spreads the most, and that in these two divisions there are fights with (almost) no control time and fights that are really grappling and wrestling heavy.

3.1.2   Conclusions

With the goal of learning about appropriate audiences for different UFC weight divisions we performed data analysis that is reported in the section above. To achieve said goal we focused on three fight parameters: fight frequency, fight striking activity and output, and fight control time. These three parameters influence how much are fans and casual viewers interested and engaged in fights.

First parameter, fight frequency, is one that influences all UFC fans. All fans want frequent fights. During the analysis of the data we found out that most active male weight divisions in UFC are lightweight and welterweight, while most active female division is strawweight. Additionally, fights in male divisions are more than eleven times more frequent. Out of all fights in UFC history only 7.6% of them were in female divisions.

Analysis of fight striking output is important from perspective of marketing fights to audience that wants to see action and brutality in the fight. From this part of the study we can conclude that lighter divisions are much more action packed and contain much more striking activity than heavier divisions. But, there is a tangent which is very important for this particular conclusion. Even though lighter weight divisions in average have much more striking activity, they have much smaller proportion of KO/TKO fight finishes. Therefore, for audience that likes to see knockouts, lighter weight divisions are not good choice, but rather the heaviest divisions (light heavyweight and heavyweight).

Fight control time is parameter that we studied for a special kind of UFC audience, and that is grappling and wrestling fanbase. Our analysis found that there is no noticeable preference for grappling in any division. Although, data shows that there is slightly less grappling and wrestling in light heavyweight and heavyweight divisions.

3.2 Win method expectation

This section will study predictability of win method based on fighting statistic and time format. Predictability of win method indirectly influences marketing strategies of UFC. Fight outcome is crucial part of planning of fighters next fights and for planning rest of divisions future. KO/TKO finish is hard to recover from, hence fighters plans are going to change after KO/TKO. Submissions can often leave hard injuries (e.g. muscle or ligament tear, joint fractures) that can take months even years to fully recover from. For this reason, it is important to have some idea of what win method is expected in which fight. It is then important for marketing reasons from planning future fights for opponents, preparing promotional video media ("Cold Open", "UFC Countdown") to planning of fighters' careers. Crucial thing to point out is that UFC fights are generally unpredictable. Upsets and big comebacks are often phenomena in UFC because of sole nature of the sport. Hence, this study does not claim that presented parameters predict fights. What is here presented is just historical relation between said parameters.

3.2.1 Analysis

Basic logic suggests that number of significant strikes thrown by winner of the fight might influence in which manner the fight is going to end. To examine this relationship let's examine plot on Figure 8. below. We can see density functions of number of significant strikes thrown by the winner of the fight plotted; functions are plotted separately for four main win methods in UFC. From the plot we read that most of the fights that finish with

Figure 7: Winners significant strikes in a fight density function by win method

KO/TKO or with submission have quite a small number of strikes thrown by fights winner. Most of fights in which winner throws more than 200 will not end with the KO/TKO or submission. On the other hand, if fight has a high number of strikes thrown it will end with winner being picked by judges, and number of strikes thrown in a fight does not seem to influence is that decision going to be split or unanimous.
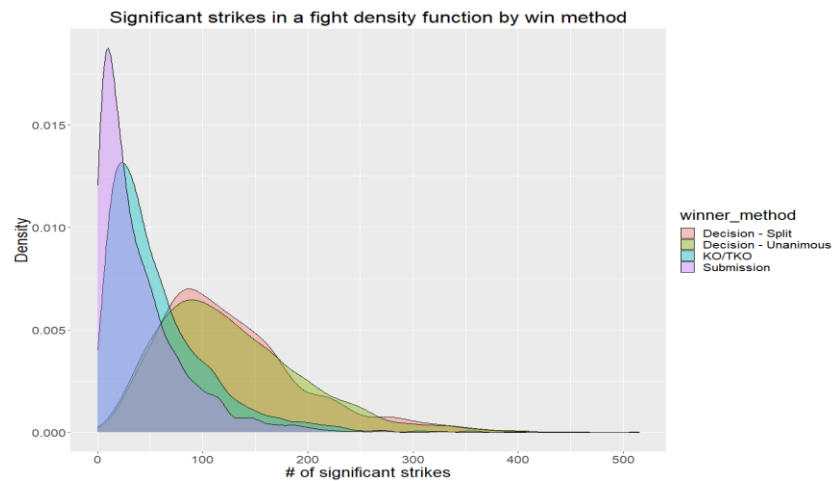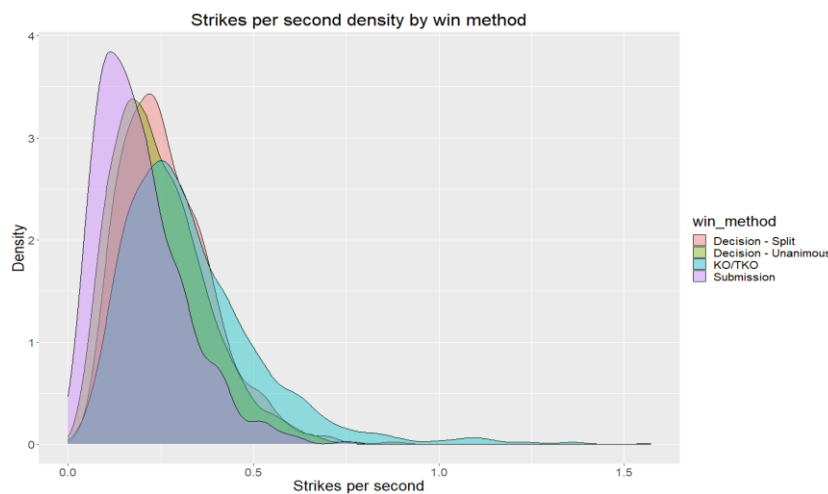

Figure 8: Strikes per second density by win method

On the other hand, if we examine number of total strikes (thrown by both fighters) per second in a fight, we will see something different (Figure 7.). Data seems well balanced for the most parts. But most right skewed density function is one for KO/TKO win method. Which means that the higher is activity of the fighters (strikes per second), the higher is probability that somebody is getting knocked out.

Accuracy of significant strikes thrown by the winner of the fight is next parameter that is examined. Does winners' accuracy in striking influence how will the fight end? Plot on the Figure 9. will help us answer this question. On this plot density functions of significant strikes accuracy of the winner in the fight are plotted; again, functions are plotted separately for each win method. From the Figure 9. we can notice that winners accuracy in striking does not make big difference in influencing fights win method. But, there are some small tangents that are worth noting. It would seem that winners accuracy in fight actually has small influence. Firstly, the more accurate winner is, the smaller is the probability that they will win by split decision. Secondly, plot shows that the higher the accuracy of the winner, it is more probable that they will win by a knockout.
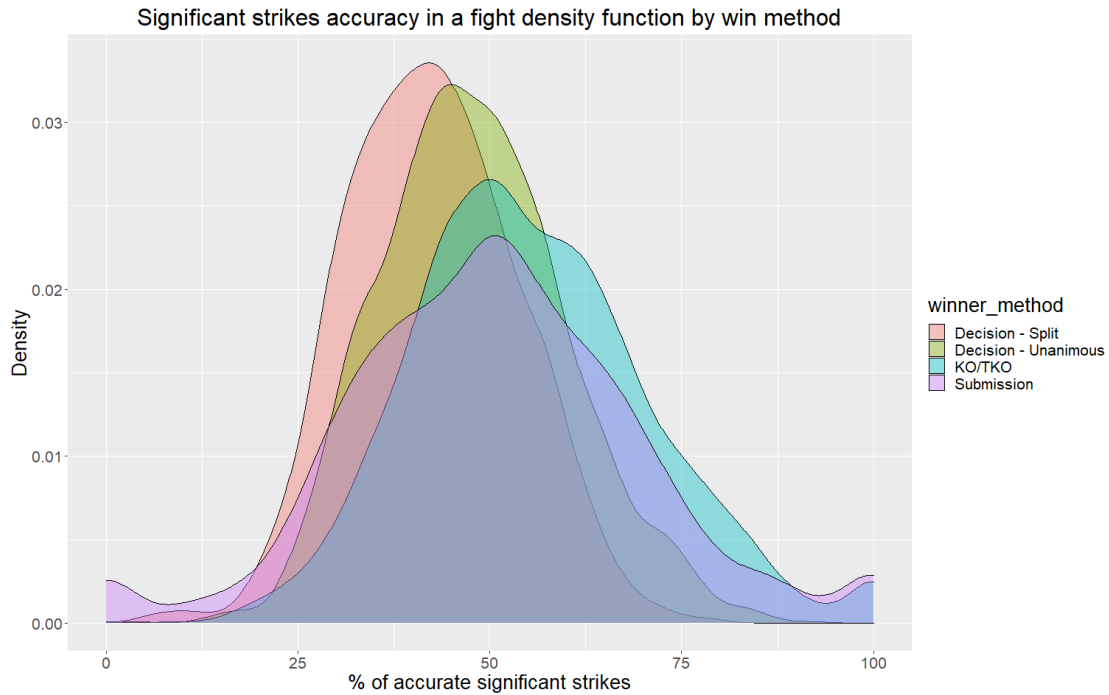
Figure 9: Significant strikes accuracy in a fight density function by win method

Third important parameters influence to fights win outcome is winners control time. To find out how winners control time in the fight influences the win method we will examine Figure 10. Two very important information will be pointed out on this plot. First, in fights that end in KO/TKO winner of the fight has a very low amount of control time. Other is that majority of fights in which winner has big amount of control time end by unanimous decision. Almost every fight in which winner has over 10 minutes of control time will end by unanimous decision.
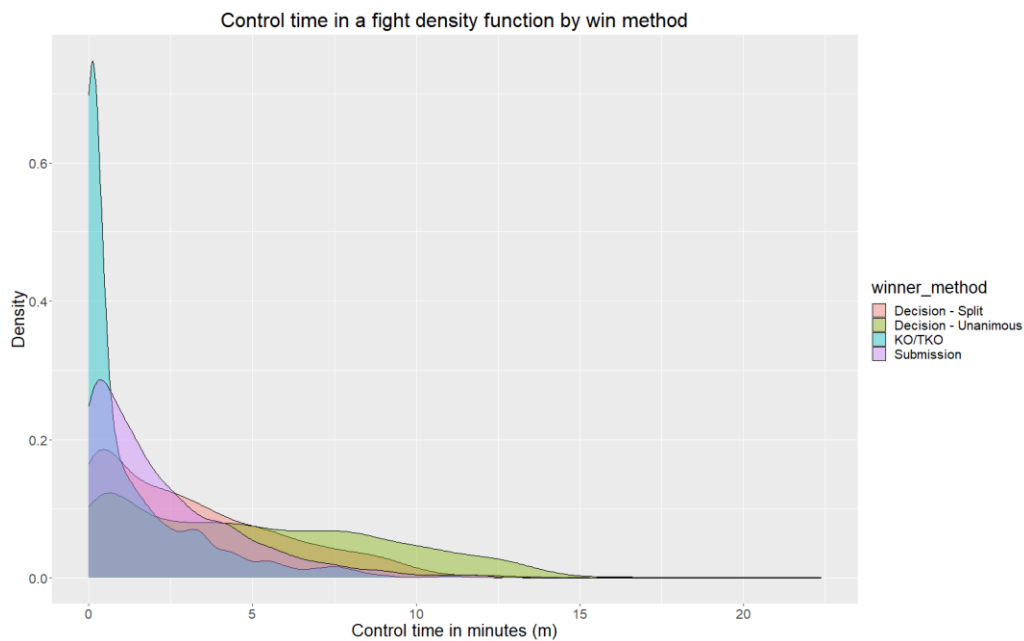

Figure 10: Control time in a fight density function by win method

Final parameter that must be mentioned is time format of the fight. Difference between three round and five round fights is significant. It changes fighters preparation, game-plan, strategy, and it also changes fighters performance due to added 10 minutes of fighting. To examine how time format of the fight influences win outcome we will take a look at plot on Figure 11. This plot shows that in five round fights there is much more KO/TKO fight finishes. Almost 50% of five rounds fights ends with KO/TKO. On the other hand, most three round fights end with unanimous decision.
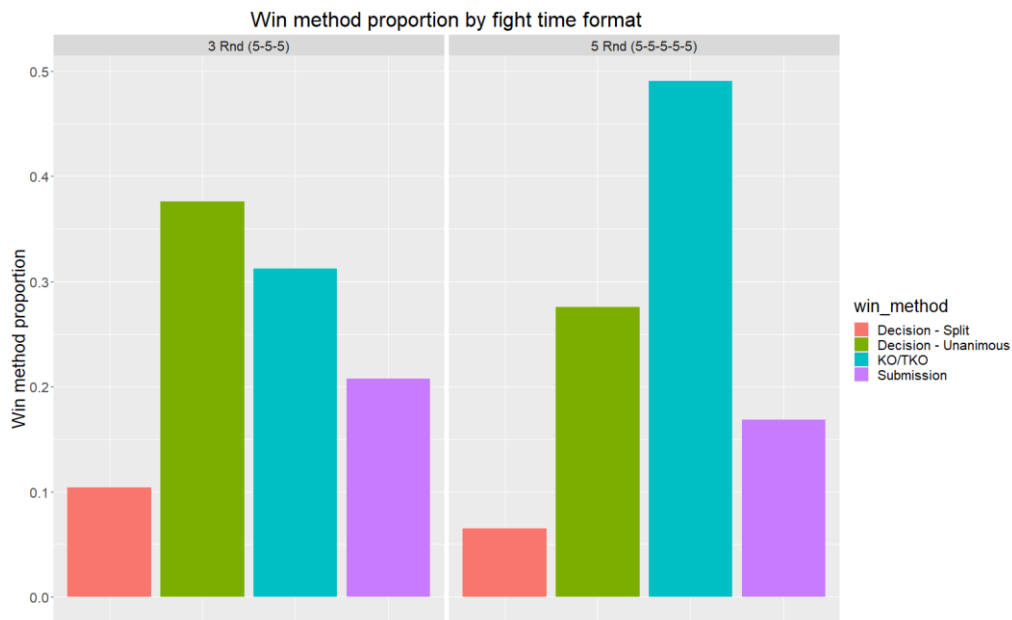

Figure 11: Win method proportion by fight time format

3.2.2   Conclusions

Every fighter has their own fight characteristics. When pairing up fighters for a fight, based on their characteristics, fight promotion (UFC) can have a basic idea what to expect from the fight, by taking into account conclusions that are going to be presented here.

In this study, we have found out that fighters with high number of significant strikes per fight historically do not finish with KO/TKO. But, fighters with huge output rate (strikes per second), do tend to finish with a knockout. With this information, we can expect fight with high cardio slow paced (but active) fighters to end with decision win. On the other side, fight with fighters with explosiveness and characterized by striking outbursts (sprints) we can expect to see KO finish. Furthermore, we found that fighters that are more accurate with their striking, tend to KO their opponents slightly more often. Control time has shown that fighters that have less control time in their fights are in the most cases likely to KO opponents. On the other side, fighters with huge control time almost exclusively finish the fight with unanimous decision.

Another important parameter that has influence over fight outcome is time format of the fight. Fights with 5 rounds almost in 50% of cases finish with knockout.

4. Discussion

Topic of this study was to explore possible marketing strategies of Ultimate Fighting Championship, which would provide directions to more effective marketing. Marketing effectiveness can be increased by discovering appropriate audiences for particular fights. Approach that this study took was to try and identify appropriate audiences by solely analyzing fighting statistics. Analysis of fighting data has provided more in-depth understanding which divisions are match for which audience. Also fighting data analysis reveled some basics of fight outcome predictions.

Goals of the study were to find out how weight divisions related to fight outcomes and win methods in Ultimate Fighting Championship, what fight parameters influence which win methods and how time format of the fight influences the fight outcome. These questions were answered and explained in Section 3. ("Analysis") of this document, and we are briefly going to sum them up here. Win methods do noticeably favor weight divisions. Heavier divisions favor KO/TKO finishes, while lighter divisions favor decision win method. From striking activity standpoint, lighter divisions are much more active, while for heavier divisions opposite is true. Weight divisions in the middle are much more balanced and diverse, and that is probably the reason why they have most frequent fights and are most popular divisions for casual viewers. In terms of influence of certain parameters on win method, there has been some interesting findings. KO/TKO finishes are mostly present in fights where winner is explosive (has high rate of striking par second), more accurate and when fight is 5 rounds long. On the other hand, decision finishes are present in fights where winner is much more durable, has high endurance, has very high number of strikes per fight (not necessarily strike per second, but over whole fight), and especially when winner has high control time.

Topic of the follow up study that is going to be done on this dataset is going to be stylistic fighting changes in UFC over time, since 1993. up to 2021. This is going to be continuation of this data analysis project, as it is going to analyze deeper relationships between fighting statistics over time, which will also provide better understanding of how fighting statistics can improve UFC marketing strategies.

5. References

http://ufcstats.com/statistics/events/completed - Data source

https://github.com/tucah1/ufcstats-data-analysis - GitHub repository of this project

Appendix A: Data set variable definitions

Suffixes

_tot – total / in the entire fight (all rounds summed up)

_rd1 – in round 1

_rd2 – in round 2

_rd3 – in round 3

_rd4 – in round 4

_rd5 – in round 5

Prefixes

red_ – name of the fighter fighting from the red corner

blue_ – name of the fighter fighting from the blue corner

Variables

event_title – name of the event

event_date – date on which event took place

event_location – location of the venue where event took place

nickname – fighters nickname

winner – winner of the fight

division – weight division of the fighters

win_method – how did winner win

win_round – in which round did winner win

win_time – what was the time on the clock when fight finished

time_format – what was time format of the fight (how many rounds, and how long are they)

referee – name of the referee

kd – number of knockdowns

sig_str – number of significant strikes landed

sig_str_attempt - number of significant strikes thrown

sig_str_prec – accuracy of thrown significant strikes

tot_str – number of total strikes landed

tot_str_attempt – number of total strikes attempted

td – number of successful takedowns

td_attempt – number of takedowns attempted

td_acc – accuracy of takedown attempts

sub_attempt – number of submission attempts

rev - /

ctrl_time – control time (how much time did fighter control body and position of another fighter)

sig_str_head – number of strikes landed to the head

sig_str_head_attempt – number of strikes to the head attempted

sig_str_body – number of strikes landed to the body

sig_str_body_attempt – number of strikes to the body attempted

sig_str_leg – number of strikes landed to the leg

sig_str_leg_attempt – number of strikes to the leg attempted

sig_str_dist – number of strikes landed from the distance

sig_str_dist_attempt – number of strikes from the distance attempted

sig_str_clinch – number of strikes landed in the clinch

sig_str_clinch_attempt – number of strikes from in the clinch

sig_str_gnd – number of strikes landed on the ground

sig_str_gnd_attempt – number of strikes on the ground attempted

Appendix B: Code

1. Control time conversion function is called on every dataframe column that holds control time data, and it looks like this:

```r
convert_ctrl_time <- function(vec) {
  vec[vec == "--"] <- "0:00"
  list_split <- strsplit(vec, ":")
  lapply(list_split, function(x) {
    return(as.numeric(x[1]) * 60 + as.numeric(x[2]))
  })
}
```

Function firstly splits by colon (":") every value in passed vector. First part is representing minutes and second is representing seconds. Therefore, for each element of passed vector (which is now a list of lists) first part is multiplied by 60 (converting minutes into seconds) and added with second part. This function returns a list of lists, and each inner list has only one element, therefore returned list is unlisted (with R function 'unlist') before use in the dataframe.

2. Figure 1. plotting code:

```r
ggplot(data = ufcstats, aes(x = division)) +
  geom_bar(aes(fill = division_type)) +
  labs(title = "Number of fights per devision", x = "Division", y = "Number o
f fights") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

3. Figure 2. plotting code:

```r
division_type_df <- data.frame(
  Division = names(table(ufcstats$division_type)),
  value = as.vector(table(ufcstats$division_type))
) %>%
  mutate(
    cumulative = cumsum(rev(value)),
    midpoint = cumulative - rev(value) / 2,
    label = paste0(rev(round(value / sum(value) * 100, 1)), "%")
  )

ggplot(data = division_type_df, aes(x = "", y = value, fill = Division)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  theme_void() +
```

```
  geom_text(aes(x = 1.2, y = midpoint, label = label)) +
  ggtitle("Proportion of fights based on division type") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

4. Figure 3. plotting code:

```
ggplot(data = ufcstats, aes(x = division)) +
  geom_bar(aes(fill = win_method)) +
  labs(title = "Win method frequency over all divisions", x = "Division", y =
 "Count") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

5. Figure 4. plotting code:

```
ggplot(data = ufcstats, aes(x = division)) +
  geom_bar(aes(fill = win_method), position = "fill") +
  labs(title = "Win method proportions over all divisions", x = "Division", y
 = "Proportion of win methods in the division") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

6. Figure 5. plotting code:

```
ggplot(data = ufcstats, aes(x = division, y = total_sig_str_attempt)) +
  geom_boxplot(outlier.shape=NA, fill = "#ff7d7d", color = "#bf2222") +
  coord_cartesian(ylim = c(0, 700)) +
  labs(title = "Total significant strikes output in a match over all division
s", x = "Division", y = "Total significant strikes in a match") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

7. Figure 6. plotting code:

```
ggplot(data = ufcstats, aes(x = division, y = total_ctrl_time)) +
  geom_boxplot(fill = "#ff7d7d", color = "#bf2222") +
  labs(title = "Total control time in a match over all divisions", x = "Divis
ion", y = "Total control time in a match in seconds") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

8. Figure 7. plotting code:

```
ggplot(cropped_df, aes(x = strikes_per_second, fill = win_method)) +
  geom_density(alpha = 0.4) +
  labs(title = "Strikes per second density by win method", x = "Strikes per s
econd", y = "Density") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

9. Figure 8. plotting code:

```
ggplot(winner_df, aes(x = winner_sig_str_attempt, fill = winner_method)) +
  geom_density(alpha = 0.4) +
  labs(title = "Significant strikes in a fight density function by win method
", x = "# of significant strikes", y = "Density") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

10. Figure 9. plotting code:

```
ggplot(winner_df, aes(x = winner_sig_str_prec, fill = winner_method)) +
  geom_density(alpha = 0.4) +
  labs(title = "Significant strikes accuracy in a fight density function by w
in method", x = "% of accurate significant strikes", y = "Density") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

11. Figure 10. plotting code:

```
ggplot(winner_df, aes(x = winner_ctrl_time / 60, fill = winner_method)) +
  geom_density(alpha = 0.4) +
  labs(title = "Control time in a fight density function by win method", x =
"Control time in minutes (m)", y = "Density") +
  theme(plot.title = element_text(hjust = 0.5),
        text = element_text(size = 20))
```

12. Figure 11. plotting code:

```
ggplot(data = cropped_df, aes(x = win_method)) +
  geom_bar(aes(y = (..count..)/tapply(..count.., ..PANEL.., sum)[..PANEL..],
fill = win_method)) +
  facet_wrap(~time_format) +
  labs(title = "Win method proportion by fight time format", y = "Win method
proportion") +
  theme(plot.title = element_text(hjust = 0.5),
```

```
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        text = element_text(size = 20)
  )
```

13. Figure 12. plotting code:

```
ggplot(data = ufcstats, aes(x = win_method)) +
  geom_bar(aes(y = (..count..)/tapply(..count.., ..PANEL.., sum)[..PANEL..],
fill = win_method)) +
  scale_y_continuous(labels = percent_format()) +
  facet_wrap(~ division, nrow = 3) +
  labs(title = "Win method proportions over all divisions", y = "Percentage o
f win method frequency") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        text = element_text(size = 20)
        )
```
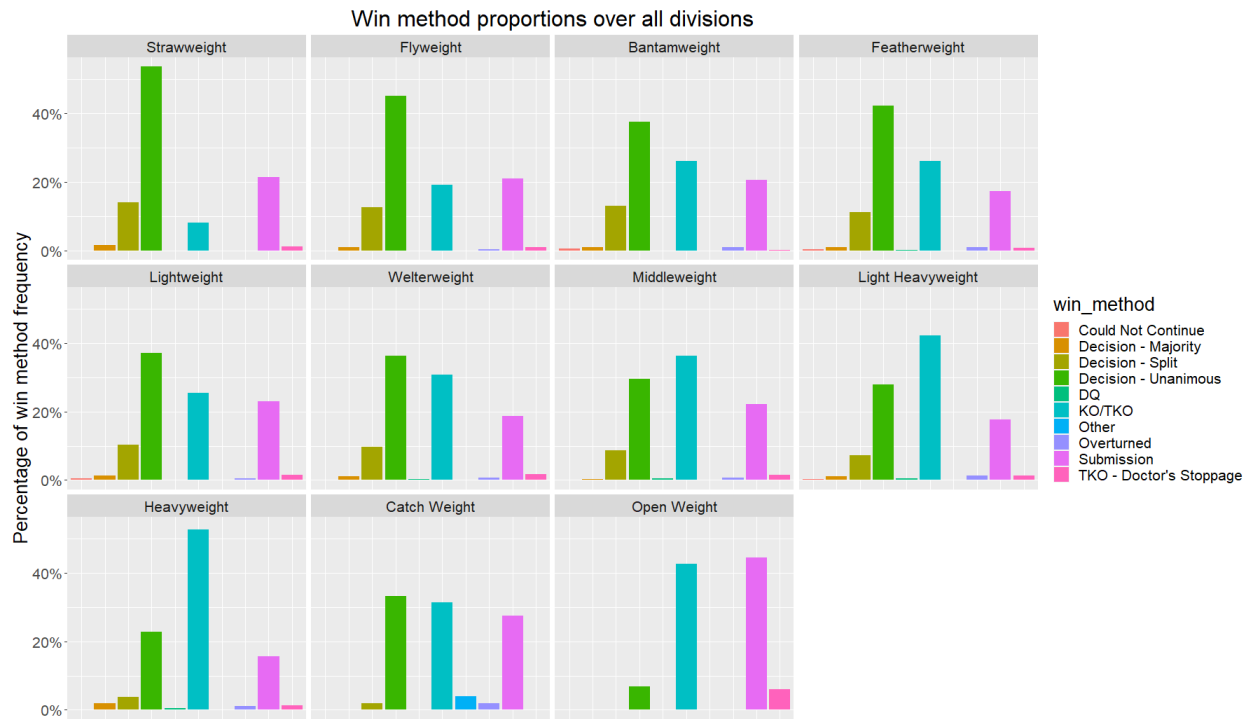
Appendix C: Additional Graphics

Figure 12: Win method proportions over all divisions