UFC Data Analysis

Caleb Maddry

1 Abstract

Mixed martial arts is a relatively new sport that combines number marital arts such as boxing, wrestling, and taekwondo. The largest mixed martial arts arts (MMA) promotion company is the Ultimate Fighting Championship (UFC). Here, we examine statistical trends over the company's existence to determine how the sport has changed over the last 40 years. First, we found that the effectiveness of grappling has decreased over the lifetime of the UFC. In relation to this, we also found that kickboxing techniques have increased in frequency during the same time period.

2 Introduction

2.1 Background

The majority of professional sports leagues have been around for well over 50 years, and many of the statistics surrounding these sports were not recorded until much later. One exception to this general trend can be found in the newly popularized sport of MMA.

In MMA two fighters fight in three five-minute rounds. An exception to this is for championship fights in which there are five five-minute rounds. A fight can end in 3 possible ways. The first being technical knockout (TKO) or knockout. A TKO happens when the referee stops a fight in order to protect a fighter from further damage. Essentially, they have decided that a fighter will be knocked out if they do not stop the fight. A knockout happens when a fighter loses consciousness as a result of a punch, kick, or another impact. The second possible way is by submission. A submission happens when a fighter places there opponent in a submission hold and they 'tap-out' or tap another fighter 3 times to signal that they give up. There are many types of submission holds, but generally they can be broken into two possible categories of joint-locks and choke-holds. The two previous scenarios where a fight can be ended take place before the fight is over. However, if a fight lasts for all three rounds, then a panel of three professional judges ranks each fight on a scale 1-30 and the fighter with the most points wins the fight.

The largest promoter of this sport, the UFC, has only existed for around 30 years. Therefore, the statistics surrounding the sport are well documented for nearly the entire history of the sport. While it may seem that the young age of the sport would mean it has remained largely unchanged over the course of its existence, this is not the case. The sport has drastically changed from a group of diverse mixed martial artists coming from all sorts of backgrounds, to a highly organized and intensely regulated agency composed of professional athletes. This drastic change, in addition to the long-term statistical recordings, creates a unique situation that provides a room for an interesting analysis.

In this project, I will dive into the statistical database of the UFC and examine how the sport has changed since its inception in November of 1993.

3 General Examination

General notes on the dataset

The dataset that we will be examining in this report comes was scraped from the UFC's official statistical database (UFCstats.com) [1] by user on the statistics website 'Kaggle' [2].

The dataset itself is contains information about every bout that has taken place between two fighters in the UFC's 30 year existence. This information is sorted by the color of the gloves that a fighter wears (either Red or Blue). It contains statistics about the fight such as the date/location/referee and actual fighter statistics such as take down attempts or the number of head kicks landed. In total, there are 50 different columns of data that are available for analysis. Unfortunately, there are a few gaps in the datasets, but these will be taken into consideration as we progress through our data analysis.

One of these gaps pertains to fights being overturned. For the most part, this happens if a fighter wins a fight and then later test positive for performance enhancing drugs or other illegal substances such as marijuana.

Analysis of this dataset was done in R and the figures have been created with the package tidyverse [3]. The code can be found in the appendix.

3.1 Popularity growth

The growth and popularity of this sport is made immediately obvious when the number of fights per year is examined in figure 1:

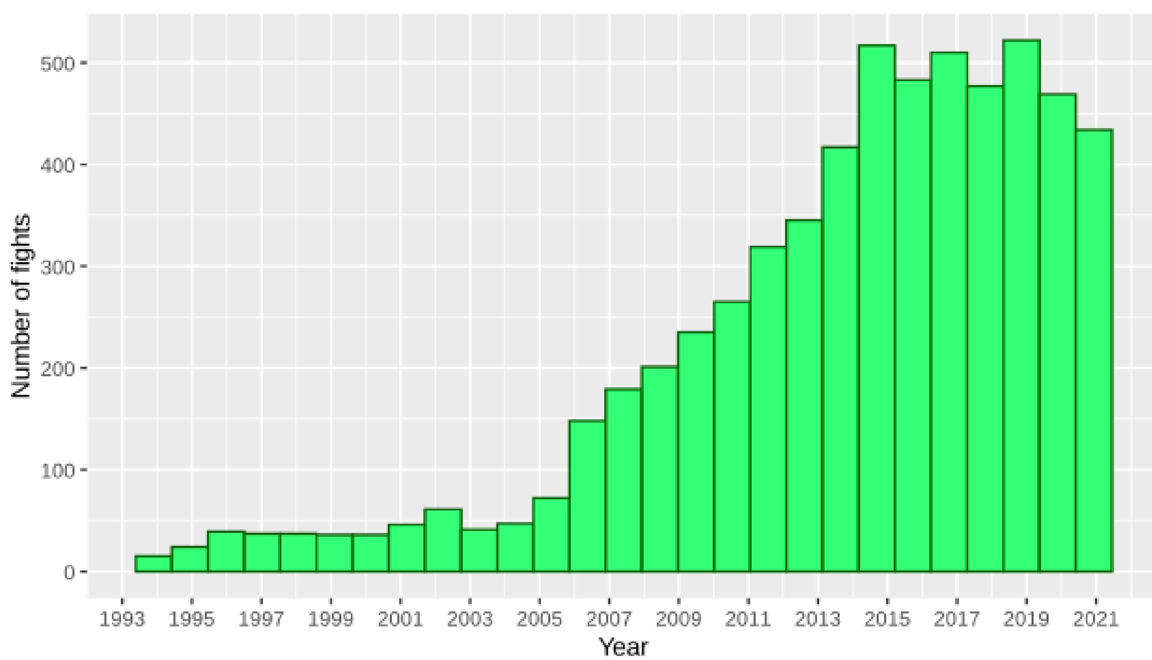


Figure 1: Fights per year

The number of events per year has increased ten-fold from the early years of the UFC. This makes the sport interesting to follow as it is suspected that a great change in the sport can be seen. Additionally, an increase in the number of fights implies that the league has become much more competitive in recent times. We look to examine general trends and assumptions made about the sport further in the following sections.

3.2 Weight classes in the UFC

In MMA, fighters are organized into weight classes, and in the UFC, the weight classes are defined as follows:

Weight Division	**Min. and Max. Weight Limits**
Straw-weight (women)	105 - 115 lbs.
Flyweight (men and women)	115 - 125 lbs.
Bantamweight (men and women)	125 - 135 lbs.
Featherweight (men and women)	135 - 145 lbs.
Lightweight (men)	145 - 155 lbs.
Welterweight (men)	155 - 170 lbs.
Middleweight (men)	170 - 185 lbs.
Light Heavyweight (men)	185 - 205 lbs.
Heavyweight (men)	205 - 265 lbs.

In this report, we will not be considering the straw-weight class as it is only a division for female fighters. Additionally, we will be including both men and women in our subsequent data analysis when speaking about specific classes.

One interesting quirk of the UFC is that fighter weigh-in $\sim$ 36 hours before they actually fight. Therefore fighter perform a 'weight cut' where they attempt to lose as much weight as possible prior to weigh-ins, and then they rapidly put on the weight again. Thus, while they have been listed as weighing 145 pounds, they are often 15 - 20 pounds heavier than this on their actual fight day. This is not important for our analysis, but it is worth mentioning to the reader.

A common belief amongst MMA fans is that fighters in the lightest and heaviest weight-classes don't knockout their opponents at the same frequency that fighters in the more common weight classes do. The thought process behind this is that lighter fighters don't generate enough power to knock their opponents unconscious, and heavier fighters have more muscle-mass in their neck that helps to stabilize their head. To test this, let's examine the average knockouts/fighter in the different weight classes:

Figure 2: K/O rate of active fighters in 2018

It is not immediately obvious if we can immediately reject or confirm that the stereotype is correct. While at first glance it appears that the flyweight and heavyweight fighters do in fact have fewer wins by KO or TKO, we look to quantify this information better.

To better quantify this data, we look to generate confidence intervals at each of the data points shown in figure 2. Typically, a confidence interval is generated using the Central Limit Theorem. Unfortunately, this method requires the underlying distribution of the data to be normally distributed and the data is certainly not as shown below:

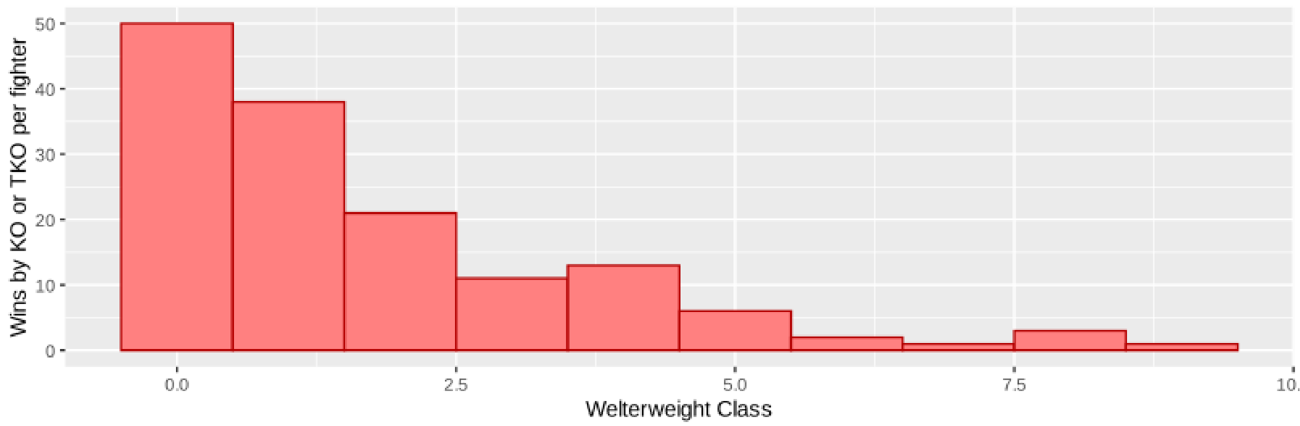


Figure 3: The underlying distribution is not normally distributed

The underlying distribution is not normally distributed, nor is it clear what the underlying distribution might be. Therefore, we must use non-parametric bootstrapping to generate confidence intervals for the means of the wins by KO or TKO at each given weight class as shown in figure 2.

Here, we create a 90% confidence interval with 10,000 bootstrapping samples. The means were taken for each bootstrapping sample and a histogram of this is shown below for the welterweight division and the blue lines represent the 90% confidence interval.

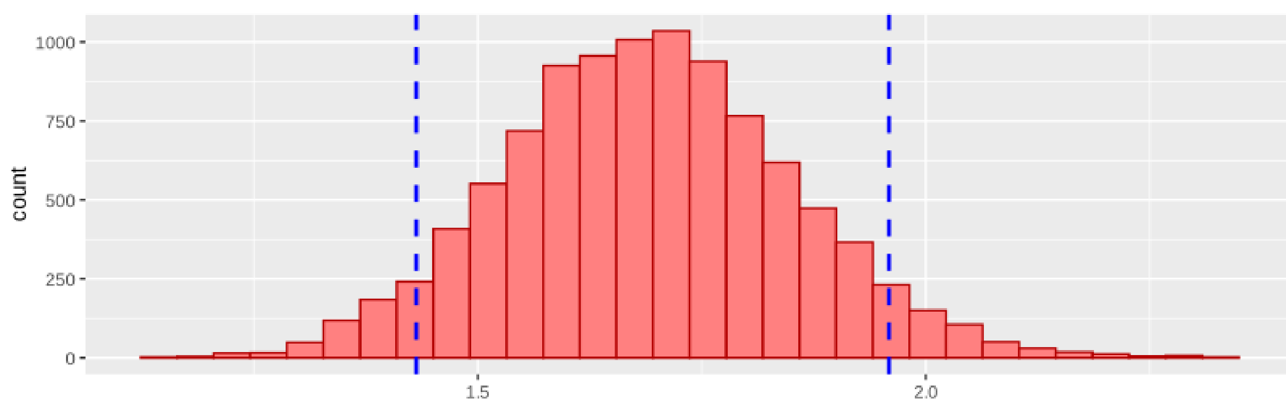


Figure 4: Estimator for the mean of 10,000 bootstrapping samples

This process was carried out for each of the 8 weight classes and these confidence intervals were then overlaid onto the previous plot shown in figure 2, resulting in the following plot:

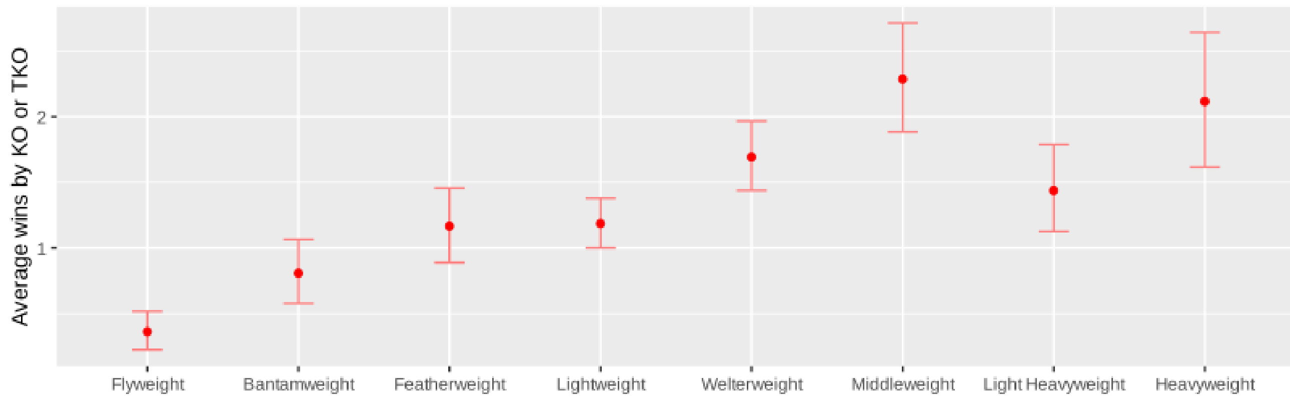


Figure 5: K/O rate of active fighters in 2018 with bootstrapping confidence intervals

Now that we have created confidence intervals for our data, it is much easier to interpret the data. Now, we can confidently say that the first part of the initial hypothesis is true, that is, fighters in lighter weight classes do not finish fights with knockouts at the same rate as fighters in the middle weight classes. Additionally, we see that the second part of this initial assumption is most likely false. While the confidence interval for the Light Heavyweight and Heavyweight classes are quite large, we can say that we are 90% confident that the Heavyweight division has a higher knockout rate than most of the other weight classes.

4 Technique Trend Analysis

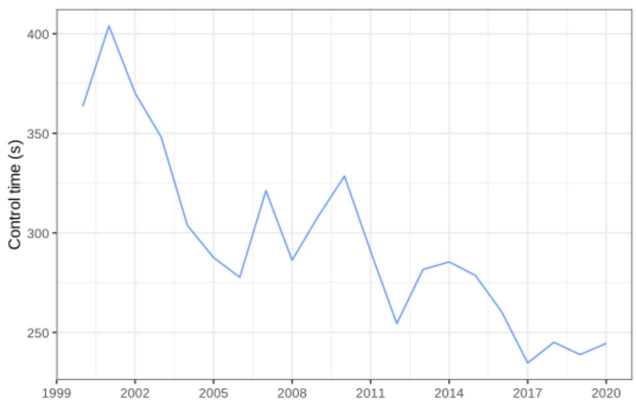
Now that we have begun to explore this dataset in some detail, the main question that I want to answer is: how have techniques in the UFC changed since its inception?

Before this question can be answered, it is important to define the various statistics that we will be using and analyzing.

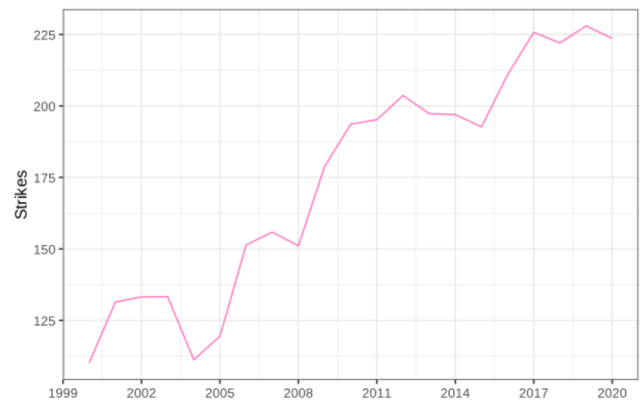
- A **take down** happens when a fighter brings another fighter to the canvas. Additionally, the fighter must maintain control and end up in a dominant position for it to be recorded as a **successful** take down. If the fighter doesn't maintain control, then the the take down is only recorded as a **attempt**.
- **Control time** is the amount of time that have fighter maintains control of another fighter while on the ground. This time is for the duration of the fight and not calculated on a round-by-round basis.
- A punch or elbow are types of **strikes**. Similar to take downs, these can be recorded as an attempt or success. Additionally, a subcategory of strikes include the location of the strike (head or body).
- **Kicks** are recorded separately from strikes and can also be categorized based on location (head, body, or leg). These are also recorded as an attempt or success.

4.1 Grappling vs. striking

Originally, MMA was seen as much closer to boxing, but with the addition of wrestling and kicking. However, Brazilian Jiu-Jistu (BJJ) has become a dominant force in the sport. This has also moved over into the fan base, as there are now thousands of BJJ gyms in the United States. While this may lead to the assumption that grappling has increased in the UFC. Below, we plot the average control time and the average strike attempts per fight over the existence of these statistics (after 1999). We see the following trend:



(a) Average control time per fight



(b) Average strike attempts per fight

Figure 6: Control time and strike attempts comparison

There is a very clear trend seen in this analysis. With the average control time and average strike attempts per fight being anti-correlated. To quantify this correlation, we calculate the correlation coefficient (c) for these two quantities using the following equation:

$$c = \text{Corr}[X, Y] = \frac{\text{Cov[X,Y]}}{sd[X] \cdot sd[Y]}$$

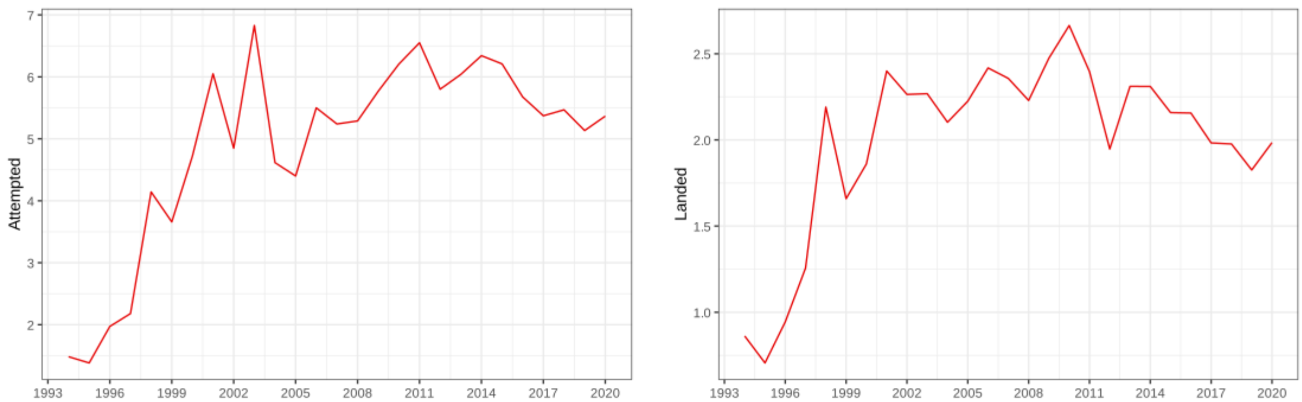
And we found $c = -0.753$. Since, $|c| \leq 1$, there is a strong negative correlation between the two quantities.

I found such a strong correlation to be quite counter-intuitive. One theory that could be help explain this correlation is that the level of grappling has evened out across the sport. In the early days of the UFC, fighters who knew jiu-jistu were incredibly successful, however, the rise in the popularity of BJJ allowed many fighters to become competent in this area of MMA. In fact, nearly every MMA fighter is at least a brown belt in BJJ. An accomplishment that takes an average of 7 years of consistent training to obtain.

Unfortunately, control time does not show the entire picture, because it is only one aspect of grappling. Take-downs play a large part as well. Next, we will look into take down attempts and successful take downs to find out how these have changed over the years. This may provide greater insight into how the UFC strategies have changed.

Figure 4 shows both the take down attempts, and successful take downs for each year of the UFC. The take downs were taken from each fight and averaged across any given year. Early on, the number of take downs were low, but quickly flattened out. Both the attempts and successes also lowered in recent years. This is not very illuminating and doesn't provide any additional information that might help us determine the general grappling trend in the UFC.

However, the story changes when we examine the take down percentage that is shown in figure 5. There is a clear downward trend that has once again leveled off in recent years. This backs up the theory that a high level of grappling has spread throughout the UFC, and the great advantage early on, is not nearly as valuable as it once was.



(a) Take down attempts per fight

(b) Successful take downs per fight

Figure 7: Total take down attempts and successes

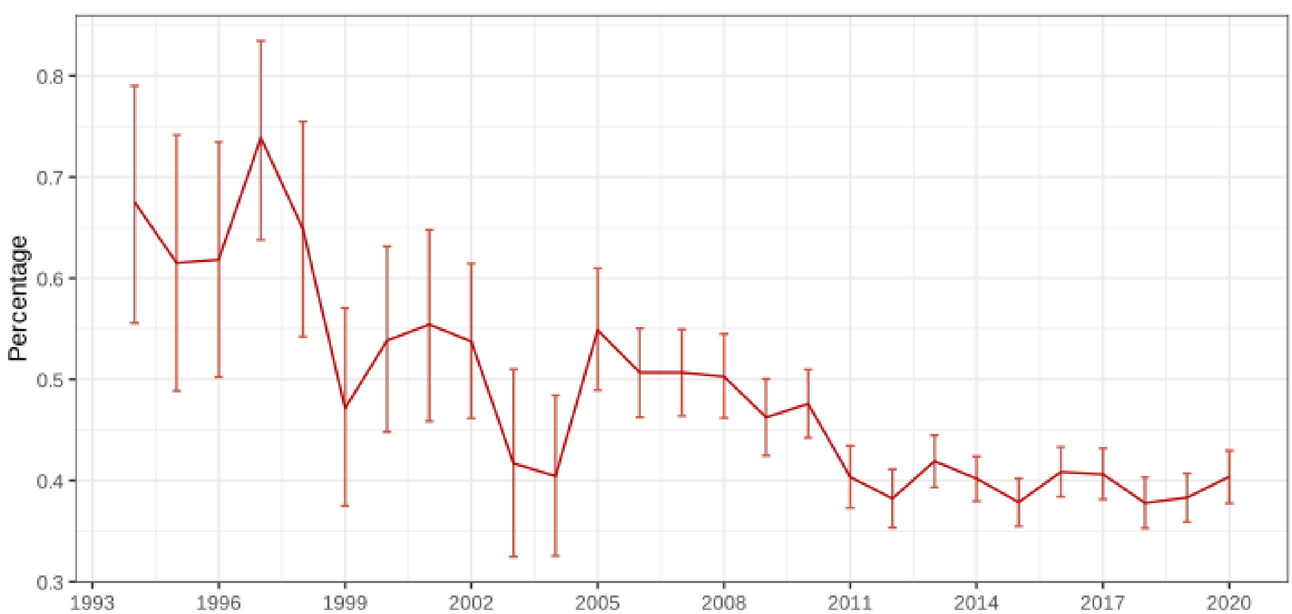


Figure 8: Take down percentage per fighter with 10,000 bootstrap samples to generate a 90% confidence interval

Overall, this was quite a surprising trend to me. I expected to see that grappling was and is the dominant skill in the UFC. And while this may have been true early on, it is no longer the case. However, it begs the question: what other skills have increased in the UFC?

4.2 Leg kicks in the UFC

One observation that I have made as a fan of the sport, is that fighters who are well versed in leg kicks tend to do dominate their opponents using this method. Fighters such as Alex Pereira, Sean Strickland, and Israel Adesanya are some of the most successful kick boxers in the UFC. It takes a MMA fighter years to become proficient in leg checks, a technique used to combat leg kicks. I have always noticed that many fighters can get completely dominated if they are not able to defend these using leg checks.

First, in continuation with the previous trend examination, we will first look at how leg kicks have changed over the history of the UFC. This plot looks incredibly similar to figure 3b. and further shows that the UFC has moved more towards kickboxing as opposed to Jiu-Jitsu which is what I had previously believed.

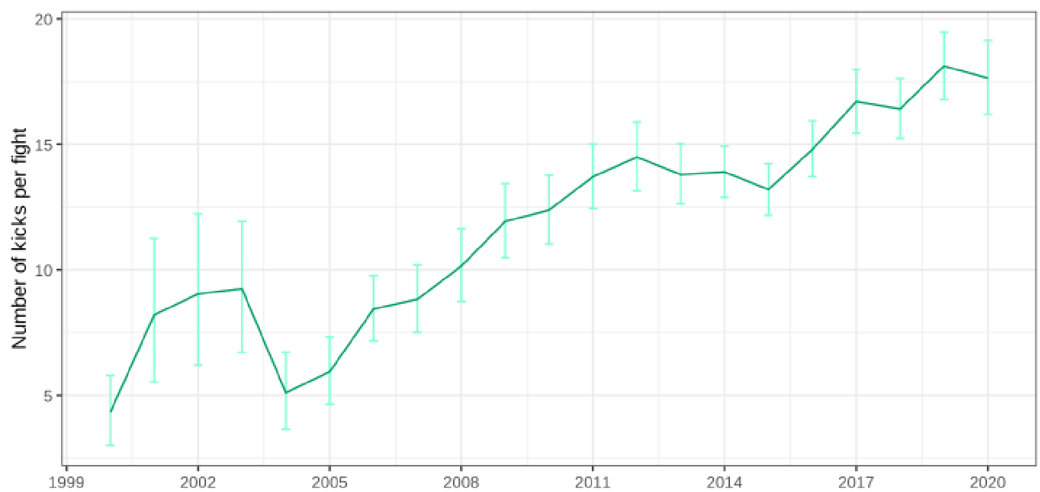
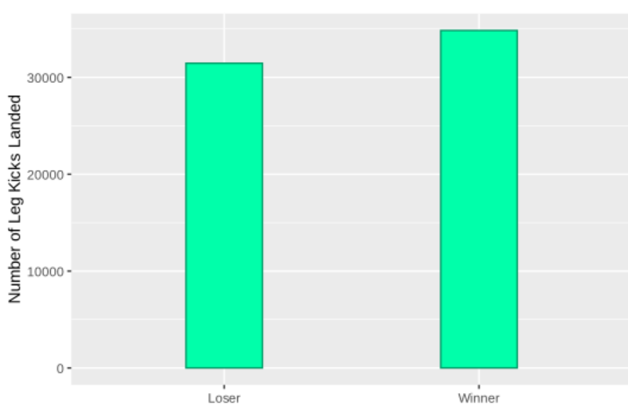
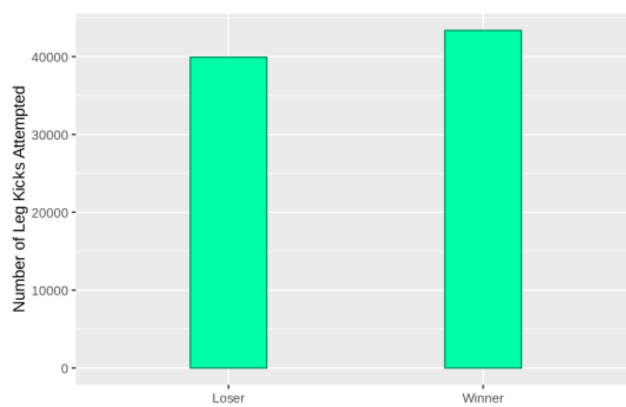


Figure 9: Number of kicks per fight over the years with 10,000 bootstrap samples for a 90% confidence interval

Next, we look to if this strategy is as dominant as I believe it to be. First, the total number of kicks for was sorted for the winner and loser of all fights across the UFC. This led to the following plot shown in figure 7.



(a) Total landed kicks



(b) Total attempted kicks

Figure 10: Landed and attempted kicks by who won and lost the the fight.

This isn't the result that I was expecting and nothing can be said about the effectiveness of leg kicks, since the winner of a fight will most likely have thrown more punches, kicks, and scored more take downs. After all, they most likely won the fight for a reason. This information doesn't isn't incredibly illuminating, so to obtain a better idea of how leg kicks might effect a fight, we look at the success percentage of leg kicks for both the winning and losing fighters. That is, how often is a leg kick successfully landed. The mean of the leg kicks percentage for losing fighters is 78.8% and for winning fighters it is 81.2%.

This still doesn't provide evidence for my original assumption and thus we reject this assumption. Leg kicks are not a dominant force in the UFC and are not a significant determining factor into whether a fighter wins or loses a fight.

5 Discussion and Conclusion

Mixed martial arts, while a relatively new sport, has seen a significant amount of change since the sports largest promoter, the UFC, has been in existence. Not only has the popularity of the sport drastically increased in the last thirty years, but the use of various techniques has also changed. Ground-based techniques such as jiu-jistu and wrestling have decreased in effectiveness, while kick-boxing techniques have increased in popularity.

Take downs, a major ground-based technique has dramatically increased in popularity, but has decreased in effectiveness. That is, more take downs are attempted, but they are landed less and less successfully. There are a few reasons behind this. For one, the number of attempts is increasing, however, there is nothing guaranteeing these to be high quality attempts. Fighters could be taking poor shots which are naturally less likely to work.

Similar to the number of strikes each fighter throws, the number of leg kicks is also increasing. I had originally believed that this strategy was highly effective for fighters who were good at them, making them a dominant technique. However, we found that leg kicks are not a technique that is decisive in determining the winner of a fight, despite their growing popularity.

While this statistical analysis shows a few surprising trends, there are a few limitations that could be improved on to provide a more comprehensive analysis. For example, a lot of our analysis groups fights by year, but this doesn't provide as clear of a trend analysis as desired. Therefore, we would group statistics into 3 month or 6 month bins. Additionally, doing a bootstrapping confidence interval determination for each statistic would be beneficial, but unfortunately this is too computationally heavy for a time-constrained project.

Overall, we have gained more insight into the UFC and we will continue to perform more analysis in the future.

6 Appendix

References

[1]

[2] Rajeev Warrier. Ufc-fight historical data from 1993 to 2021, Mar 2021.

[3] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephan Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *J. Open Source Softw.*, 4(43):1686, November 2019.

6.1 UFC Analysis Jupyter Notebook

STAT4000 FinalProject

May 1, 2024

1 STAT 4000 Final Project Code

1.0.1 Caleb Maddry

[1]:
```
## Installing of packages
install.packages('anytime')
# install.packages('chron')

## Libraries
library(ggplot2)
# library(chron)
library(hms)
library(anytime)
library(tidyr)
library(tidyverse)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Warning message in system("timedatectl", intern = TRUE):
"running command 'timedatectl' had status 1"
  Attaching packages                            tidyverse
1.3.2
  tibble  3.1.8        dplyr   1.0.10
  readr   2.1.3        stringr 1.5.0
  purrr   1.0.0        forcats 0.5.2
  Conflicts
tidyverse_conflicts()
  dplyr::filter() masks stats::filter()
  dplyr::lag()    masks stats::lag()
```

[2]:
```
## Storing the data
data = read.csv("data.csv")
preprocessed_data = read.csv("preprocessed_data.csv")
raw_fighter_details = read.csv("raw_fighter_details.csv")
raw_total_fight_data = read.csv("raw_total_fight_data.csv", sep = ';')
```

```
head(raw_total_fight_data)
head(raw_fighter_details)
head(data)
head(preprocessed_data)
```

A data.frame: 6 × 41

	R_fighter <chr>	B_fighter <chr>	R_KD <int>	B_KD <int>	R_SIG_STR. <chr>	B_SIG_ST <chr>
1	Adrian Yanez	Gustavo Lopez	2	0	41 of 103	23 of 51
2	Trevin Giles	Roman Dolidze	0	0	27 of 57	32 of 67
3	Tai Tuivasa	Harry Hunsucker	1	0	14 of 18	2 of 6
4	Cheyanne Buys	Montserrat Conejo	0	0	31 of 65	15 of 41
5	Marion Reneau	Macy Chiasson	0	0	30 of 63	51 of 138
6	Leonardo Santos	Grant Dawson	0	0	30 of 67	46 of 84

A data.frame: 6 × 14

	fighter_name <chr>	Height <chr>	Weight <chr>	Reach <chr>	Stance <chr>	DOB <chr>	SLpM <dbl>
1	Tom Aaron		155 lbs.			Jul 13, 1978	0.00
2	Papy Abedi	5' 11"	185 lbs.		Southpaw	Jun 30, 1978	2.80
3	Shamil Abdurakhimov	6' 3"	235 lbs.	76"	Orthodox	Sep 02, 1981	2.45
4	Danny Abbadi	5' 11"	155 lbs.		Orthodox	Jul 03, 1983	3.29
5	Hiroyuki Abe	5' 6"	145 lbs.		Orthodox		1.71
6	Ricardo Abreu	5' 11"	185 lbs.		Orthodox	Apr 27, 1984	3.79

A data.frame: 6 × 144

	R_fighter <chr>	B_fighter <chr>	Referee <chr>	date <chr>	location <chr>
1	Adrian Yanez	Gustavo Lopez	Chris Tognoni	2021-03-20	Las Vegas, Neva
2	Trevin Giles	Roman Dolidze	Herb Dean	2021-03-20	Las Vegas, Neva
3	Tai Tuivasa	Harry Hunsucker	Herb Dean	2021-03-20	Las Vegas, Neva
4	Cheyanne Buys	Montserrat Conejo	Mark Smith	2021-03-20	Las Vegas, Neva
5	Marion Reneau	Macy Chiasson	Mark Smith	2021-03-20	Las Vegas, Neva
6	Leonardo Santos	Grant Dawson	Chris Tognoni	2021-03-20	Las Vegas, Neva

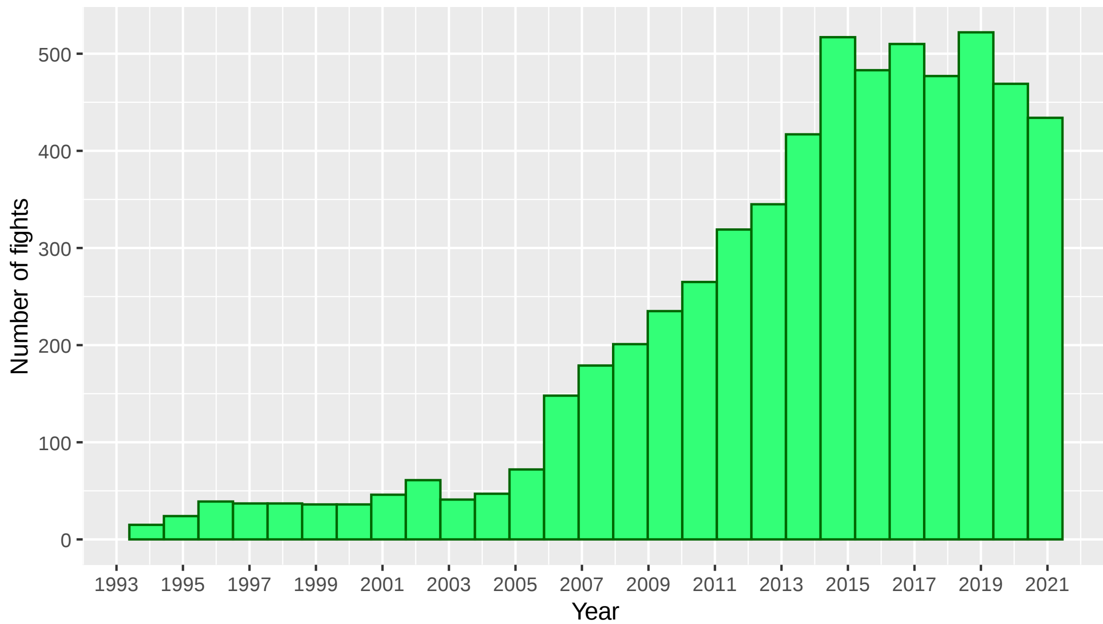
A data.frame: 6 × 160

	Winner <chr>	title_bout <chr>	B_avg_KD <dbl>	B_avg_opp_KD <dbl>	B_avg_SIG_STR_pct <dbl>	B_ <(
1	Red	False	0.000000	0	0.420000	0.4
2	Red	False	0.500000	0	0.660000	0.3
3	Red	False	0.015625	0	0.450000	0.4
4	Blue	False	0.015625	0	0.450000	0.4
5	Blue	False	0.125000	0	0.535625	0.5
6	Blue	False	0.000000	0	0.515000	0.4

1.0.2 Number of fights per year

[3]:
```
data$date = as.Date(data$date)
data$year = as.numeric(format(data$date, '%Y'))

options(repr.plot.width = 7, repr.plot.height = 4)
ggplot(data, aes(date)) +
```

```
  geom_histogram(bins = max(data$year)-min(data$year), color = '#006600',␣
↪fill = '#33ff77') +
  scale_x_date(date_labels = "%Y", date_breaks = "2 years") +
  labs(y = "Number of fights", x = "Year")
```

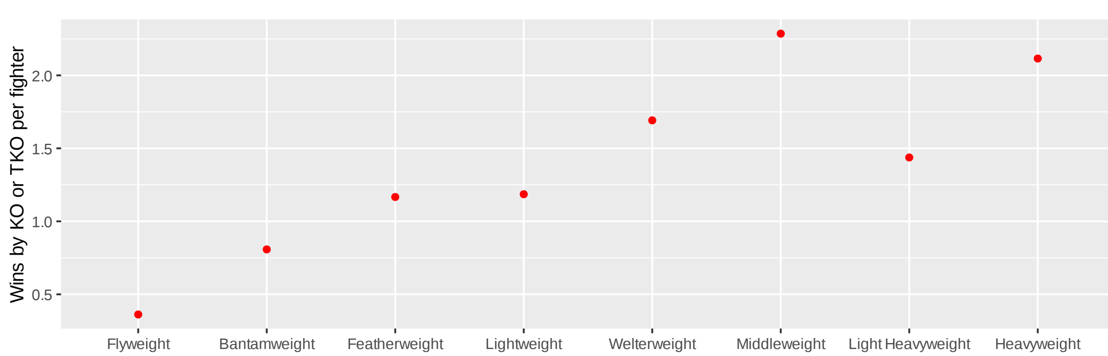


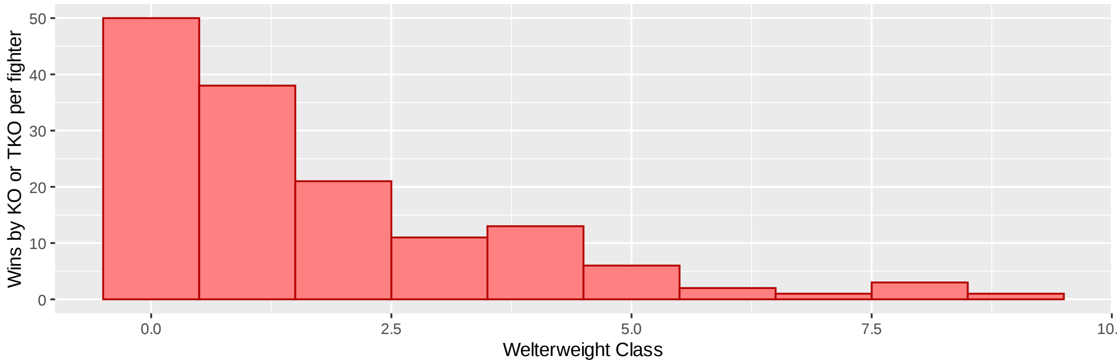
1.0.3 Knockout weight dependence

[4]:

```
## 2018 KO/TKO per weight class
dataYear2018 = subset(data, year==paste0('20',17))
dataYear2018Flyweight = subset(dataYear2018, weight_class == 'Flyweight')
dataYear2018Bantamweight = subset(dataYear2018, weight_class == 'Bantamweight')
dataYear2018Featherweight = subset(dataYear2018, weight_class ==␣
↪'Featherweight')
dataYear2018Lightweight = subset(dataYear2018, weight_class == 'Lightweight')
dataYear2018Welterweight = subset(dataYear2018, weight_class == 'Welterweight')
dataYear2018Middleweight = subset(dataYear2018, weight_class == 'Middleweight')
dataYear2018LightHeavyweight = subset(dataYear2018, weight_class ==␣
↪'LightHeavyweight')
dataYear2018Heavyweight = subset(dataYear2018, weight_class == 'Heavyweight')

a = mean(c(dataYear2018Flyweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Flyweight[,"B_win_by_KO.TKO"]))
b = mean(c(dataYear2018Bantamweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Bantamweight[,"B_win_by_KO.TKO"]))
```

```r
c = mean(c(dataYear2018Featherweight[,"R_win_by_KO.TKO"], dataYear2018Featherweight[,"B_win_by_KO.TKO"]))
d = mean(c(dataYear2018Lightweight[,"R_win_by_KO.TKO"], dataYear2018Lightweight[,"B_win_by_KO.TKO"]))
e = mean(c(dataYear2018Welterweight[,"R_win_by_KO.TKO"], dataYear2018Welterweight[,"B_win_by_KO.TKO"]))
f = mean(c(dataYear2018Middleweight[,"R_win_by_KO.TKO"], dataYear2018Middleweight[,"B_win_by_KO.TKO"]))
g = mean(c(dataYear2018LightHeavyweight[,"R_win_by_KO.TKO"], dataYear2018LightHeavyweight[,"B_win_by_KO.TKO"]))
h = mean(c(dataYear2018Heavyweight[,"R_win_by_KO.TKO"], dataYear2018Heavyweight[,"B_win_by_KO.TKO"]))

## Plotting of the dataframe
df = data.frame(x = c('Flyweight','Bantamweight','Featherweight','Lightweight',
                      'Welterweight','Middleweight','Light Heavyweight','Heavyweight'), y = c(a,b,c,d,e,f,g,h))
options(repr.plot.width = 9, repr.plot.height =3)

df$x = factor(df$x, levels = c('Flyweight','Bantamweight','Featherweight','Lightweight',
                               'Welterweight','Middleweight','Light Heavyweight','Heavyweight'))
ggplot(df) +
    geom_point(aes(x,y), color = 'red') +
    labs(y = "Wins by KO or TKO per fighter", x = "")

df = data.frame(c(dataYear2018Welterweight[,"R_win_by_KO.TKO"],dataYear2018Welterweight[,"B_win_by_KO.TKO"]))
colnames(df) = 'num'

ggplot(df) +
    geom_histogram(aes(num),color = '#b30000', fill = '#ff8080', bins = 10) +
    labs(y = "Wins by KO or TKO per fighter", x = "Welterweight Class")
```





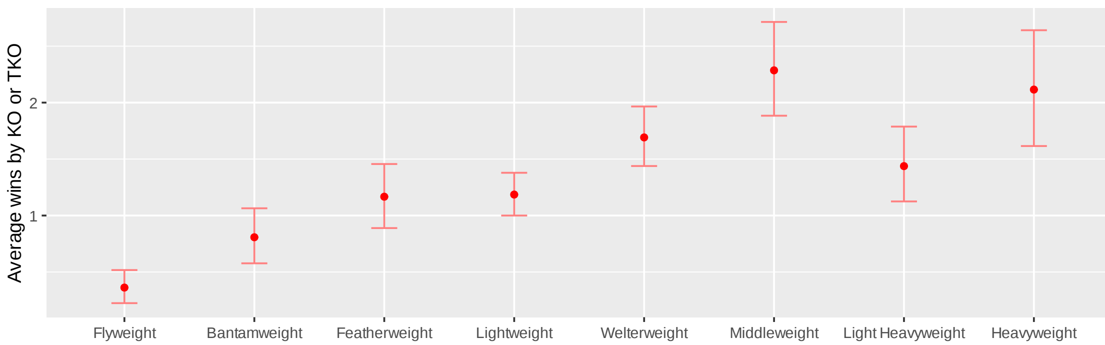
[5]:
```
set.seed(344)
ciMatrix = matrix(NA, 8, 2)
alpha = 0.10
B = 10000

sampleData = c(dataYear2018Flyweight[,"R_win_by_KO.TKO"],␣
 ↪dataYear2018Flyweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
 ↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[1, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[1,2] = quantile(thetaStar, 1 - alpha/2)

sampleData =  c(dataYear2018Bantamweight[,"R_win_by_KO.TKO"],␣
 ↪dataYear2018Bantamweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
 ↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[2, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[2,2] = quantile(thetaStar, 1 - alpha/2)

sampleData = c(dataYear2018Featherweight[,"R_win_by_KO.TKO"],␣
 ↪dataYear2018Featherweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
 ↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[3, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[3,2] = quantile(thetaStar, 1 - alpha/2)

sampleData = c(dataYear2018Lightweight[,"R_win_by_KO.TKO"],␣
 ↪dataYear2018Lightweight[,"B_win_by_KO.TKO"])
```

```
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[4, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[4,2] = quantile(thetaStar, 1 - alpha/2)

sampleData = c(dataYear2018Welterweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Welterweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[5, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[5,2] = quantile(thetaStar, 1 - alpha/2)

sampleData = c(dataYear2018Middleweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Middleweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[6, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[6,2] = quantile(thetaStar, 1 - alpha/2)


sampleData = c(dataYear2018LightHeavyweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018LightHeavyweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[7, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[7,2] = quantile(thetaStar, 1 - alpha/2)

sampleData = c(dataYear2018Heavyweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Heavyweight[,"B_win_by_KO.TKO"])
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)
    ciMatrix[8, 1] = quantile(thetaStar, alpha/2)
    ciMatrix[8,2] = quantile(thetaStar, 1 - alpha/2)

plus = ciMatrix[,2]
minus = ciMatrix[,1]

df = data.frame(x = c('Flyweight','Bantamweight','Featherweight','Lightweight',
                      'Welterweight','Middleweight','Light␣
↪Heavyweight','Heavyweight'),
                y = c(a,b,c,d,e,f,g,h), plus, minus)
```

```
df$x = factor(df$x, levels =␣
↪c('Flyweight','Bantamweight','Featherweight','Lightweight',
                              'Welterweight','Middleweight','Light␣
↪Heavyweight','Heavyweight'))
df

ggplot(df) +
    geom_errorbar(aes(x, y, ymax = plus,
                  ymin = minus), width = 0.2, color = '#ff8080') +
    geom_point(aes(x,y), color = 'red') +
    labs(y = "Average wins by KO or TKO", x = "")
```

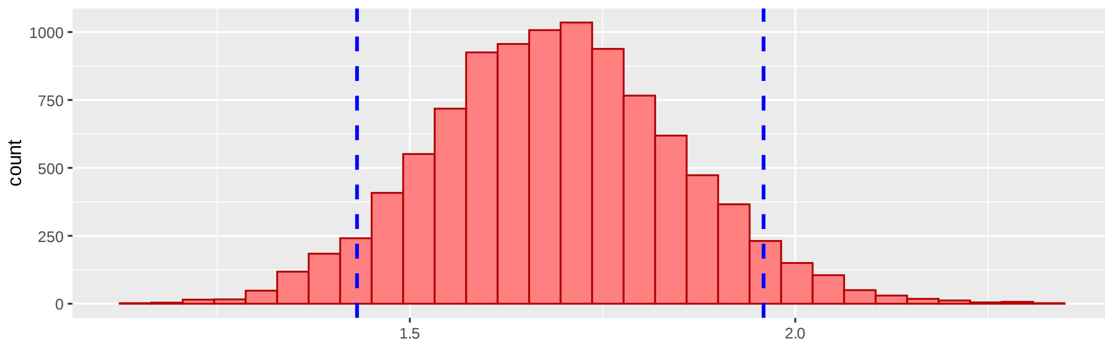
A data.frame: 8 × 4

x <fct>	y <dbl>	plus <dbl>	minus <dbl>
Flyweight	0.3620690	0.5172414	0.2241379
Bantamweight	0.8076923	1.0641026	0.5769231
Featherweight	1.1666667	1.4561111	0.8888889
Lightweight	1.1857143	1.3785714	1.0000000
Welterweight	1.6917808	1.9657534	1.4383562
Middleweight	2.2857143	2.7142857	1.8839286
Light Heavyweight	1.4375000	1.7875000	1.1250000
Heavyweight	2.1153846	2.6410256	1.6153846



[6]:
```
B = 10000
sampleData = c(dataYear2018Welterweight[,"R_win_by_KO.TKO"],␣
↪dataYear2018Welterweight[,"B_win_by_KO.TKO"])
bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData), TRUE)))

MLEofTheta = mean(sampleData)
thetaStar = apply(bootstrapSamples, 1, mean)

alpha = 0.10
lowerCI = quantile(thetaStar, alpha/2)
```

```
upperCI = quantile(thetaStar, 1 - alpha/2)

thetaStarDF = data.frame(thetaStar)

ggplot(thetaStarDF, aes(thetaStar)) +
    geom_histogram(color = '#b30000', fill = '#ff8080', bins = 30) +
    geom_vline(aes(xintercept = lowerCI), color="blue", linetype = "dashed",␣
 ↪linewidth = 1) +
    geom_vline(aes(xintercept = upperCI), color="blue", linetype = "dashed",␣
 ↪linewidth = 1) +
    labs(x = '')
```



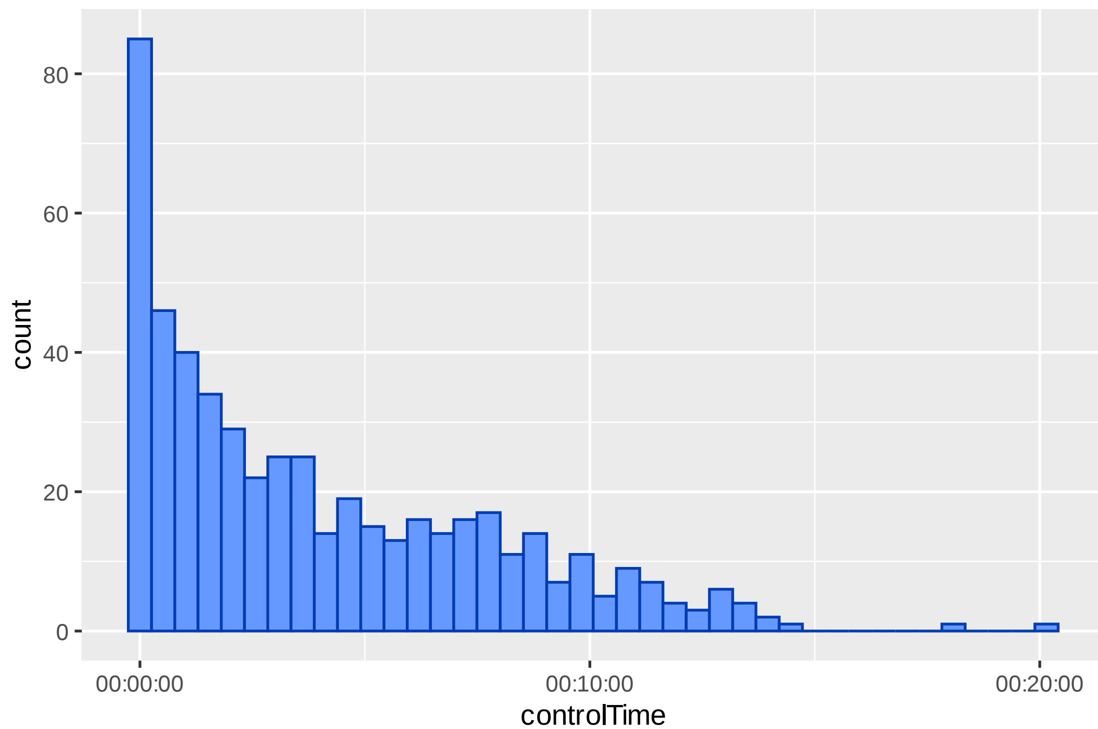
1.1 Control time vs Strike attempts

[7]:
```
# options(repr.plot.width = 15, repr.plot.height = 10)
# ggplot(data, aes( B_avg_TD_att)) +
#  geom_histogram(bins = 20) +
#  facet_wrap(data$year)
options(repr.plot.width = 6, repr.plot.height = 4)
```

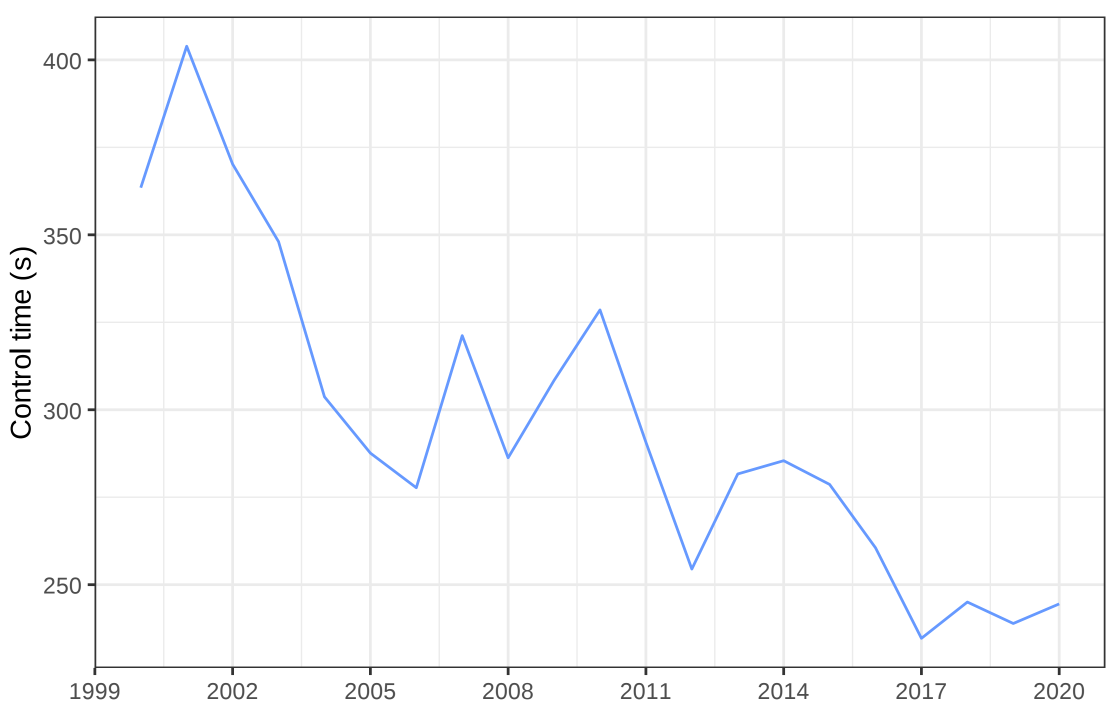
1.1.1 Control time analysis

[8]:
```
invisible(na.omit(raw_total_fight_data))

## Coverting the data so that there can be a year column
raw_total_fight_data$date = anytime(raw_total_fight_data$date)
raw_total_fight_data$date = as.Date(raw_total_fight_data$date, format =␣
 ↪'%m%d%Y')
raw_total_fight_data$year = as.numeric(format(raw_total_fight_data$date, '%Y'))

## Converting the raw control time so that it can be used in hms form
tempVec = 0
for(i in 1:5823){
```

```
   tempVec[i] =  as_hms(paste0('00:', raw_total_fight_data$B_CTRL[i])) +␣
 ↪as_hms(paste0('00:', raw_total_fight_data$R_CTRL[i]))
}

## Temp dataframe for plotting reasons
df = data.frame(hms(tempVec), raw_total_fight_data$year[1:5823])
colnames(df) = c('controlTime', 'year')

    ## Plotting to see if the data comes from a normal distribution or not
    ggplot(df[df$year == 2019,], aes(controlTime)) +
        geom_histogram(fill = '#6699ff', color = '#003cb3', bins = 40)

    ## Calculating the mean and storing it in a vector so that it can be plotted
    meanVector = 0
    yearVector = 0
    for(i in 2000:2020){
        meanVector[i - 1999] = mean(df[df$year == paste0(i),]$controlTime)
        yearVector[i - 1999] = i
    }
    controlTimeVyearDF = data.frame(yearVector, meanVector)

    ## Control time vs. year plot
    ggplot(controlTimeVyearDF, aes(yearVector, meanVector)) +
         geom_line(color = '#6699ff') +
         scale_x_continuous(breaks=seq(1999,2020,by=3)) +
         labs(x = '', y = 'Control time (s)') +
         theme_bw()
```



80
60
40
20
0
count
00:00:00
00:10:00
00:20:00
controlTime



400
350
300
250
Control time (s)
1999
2002
2005
2008
2011
2014
2017
2020

Data conversion

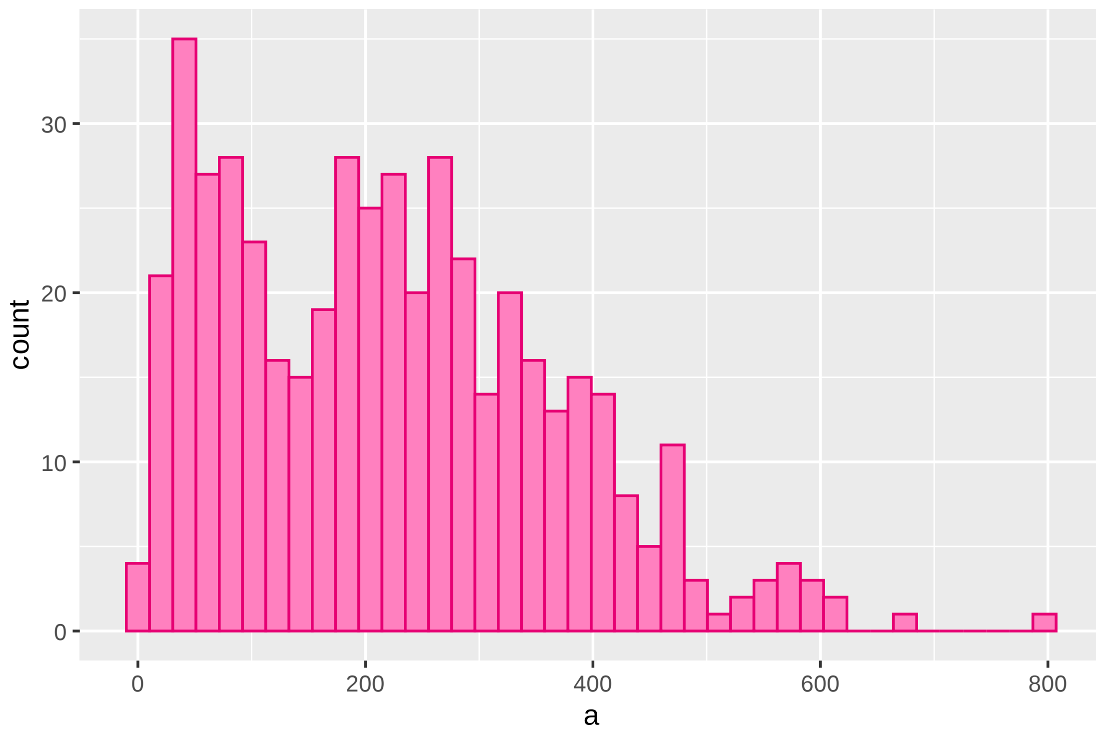
[9]:
```
## Data conversion of raw data to usable form for analysis
# Conversion of strike attempts to a usable form
    raw_total_fight_data = separate(raw_total_fight_data, R_TOTAL_STR.,␣
↪into=c('RTotalStrikesLanded', 'RTotalStrikesAttempted'), sep='of')
    raw_total_fight_data = separate(raw_total_fight_data, B_TOTAL_STR.,␣
↪into=c('BTotalStrikesLanded', 'BTotalStrikesAttempted'), sep='of')

    raw_total_fight_data$RTotalStrikesAttempted  = as.
↪numeric(raw_total_fight_data$RTotalStrikesAttempted)
    raw_total_fight_data$BTotalStrikesAttempted  = as.
↪numeric(raw_total_fight_data$BTotalStrikesAttempted)

    raw_total_fight_data$TotalStrikesAttempted =␣
↪raw_total_fight_data$RTotalStrikesAttempted +␣
↪raw_total_fight_data$BTotalStrikesAttempted

# Conversion of leg kick attempts to a usable form
    raw_total_fight_data = separate(raw_total_fight_data, R_LEG,␣
↪into=c('RLegKicksLanded', 'RLegKicksAttempted'), sep='of')
    raw_total_fight_data = separate(raw_total_fight_data, B_LEG,␣
↪into=c('BLegKicksLanded', 'BLegKicksAttempted'), sep='of')

    raw_total_fight_data$RLegKicksAttempted  = as.
↪numeric(raw_total_fight_data$RLegKicksAttempted)
    raw_total_fight_data$BLegKicksAttempted  = as.
↪numeric(raw_total_fight_data$BLegKicksAttempted)

    raw_total_fight_data$RLegKicksLanded  = as.
↪numeric(raw_total_fight_data$RLegKicksLanded)
    raw_total_fight_data$BLegKicksLanded  = as.
↪numeric(raw_total_fight_data$BLegKicksLanded)

    raw_total_fight_data$TotalKicksAttempted =␣
↪raw_total_fight_data$RLegKicksAttempted +␣
↪raw_total_fight_data$BLegKicksAttempted
    raw_total_fight_data$TotalKicksLanded =␣
↪raw_total_fight_data$RLegKicksLanded + raw_total_fight_data$BLegKicksLanded

# Time conversion and the addition of a year vector
    raw_total_fight_data$date = anytime(raw_total_fight_data$date)
    raw_total_fight_data$date = as.Date(raw_total_fight_data$date, format =␣
↪'%m%d%Y')
    raw_total_fight_data$year = as.numeric(format(raw_total_fight_data$date,␣
↪'%Y'))
```

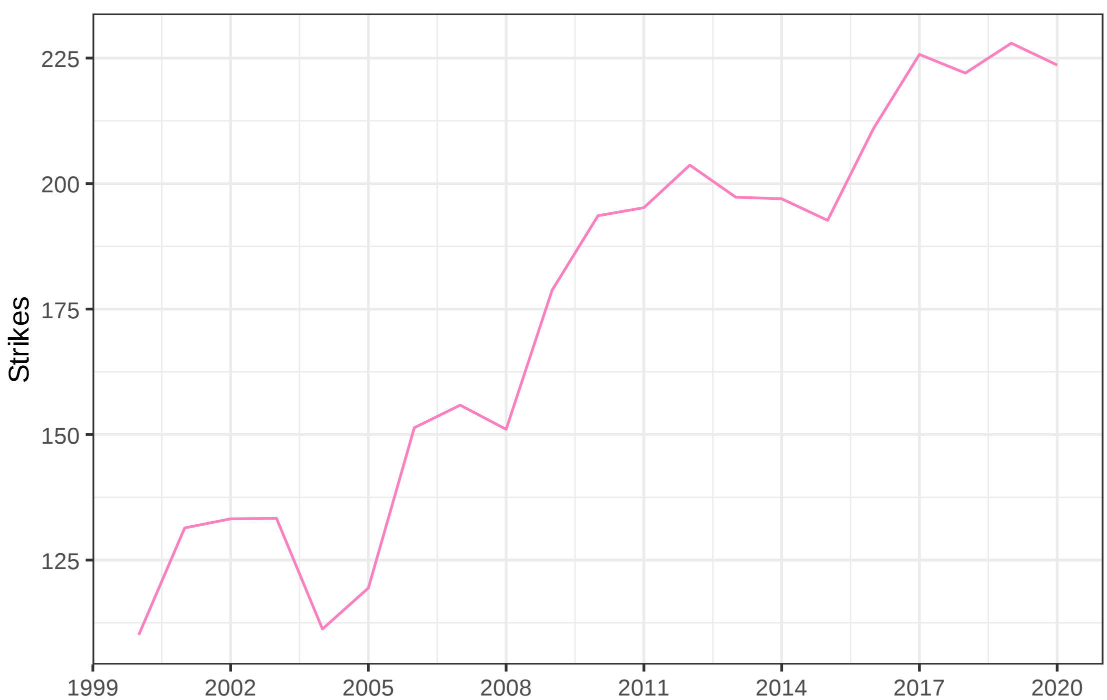
```
# Figuring out which fighter won
    for(i in 1: length(raw_total_fight_data)){
        if(raw_total_fight_data$R_fighter[i] == raw_total_fight_data$Winner[i]){
            raw_total_fight_data$BlueWon[i] = 0
            raw_total_fight_data$RedWon[i] = 1
        }
        else if(raw_total_fight_data$B_fighter[i] == raw_total_fight_data$Winner[i]){
            raw_total_fight_data$BlueWon[i] = 1
            raw_total_fight_data$RedWon[i] = 0
        }
        else{
            raw_total_fight_data$BlueWon[i] = 10
            raw_total_fight_data$RedWon[i] = 10
        }
    }
# Conversion of takedown attempts to a usable form
    raw_total_fight_data = separate(raw_total_fight_data, R_TD, into=c('RTotalTDLanded', 'RTotalTDAttempted'), sep='of')
    raw_total_fight_data = separate(raw_total_fight_data, B_TD, into=c('BTotalTDLanded', 'BTotalTDAttempted'), sep='of')

    raw_total_fight_data$RTotalTDAttempted  = as.numeric(raw_total_fight_data$RTotalTDAttempted)
    raw_total_fight_data$BTotalTDAttempted  = as.numeric(raw_total_fight_data$BTotalTDAttempted)

    raw_total_fight_data$RTotalTDLanded  = as.numeric(raw_total_fight_data$RTotalTDLanded)
    raw_total_fight_data$BTotalTDLanded  = as.numeric(raw_total_fight_data$BTotalTDLanded)

    raw_total_fight_data$TotalTDAttempted = raw_total_fight_data$RTotalTDAttempted + raw_total_fight_data$BTotalTDAttempted
    raw_total_fight_data$TotalTDLanded = raw_total_fight_data$RTotalTDLanded + raw_total_fight_data$BTotalTDLanded

    raw_total_fight_data$TDPercentage  = raw_total_fight_data$TotalTDLanded / raw_total_fight_data$TotalTDAttempted
```

1.1.2 Strike attempts over time

[10]:

```
## Total strikes attempted analysis
## Plotting of the histogram of strikes attempted to see if it follows a normal␣
↪distribution
a = raw_total_fight_data[raw_total_fight_data$year ==␣
↪paste0(2018),]$TotalStrikesAttempted

x = data.frame(a)
ggplot(x, aes(a)) +
    geom_histogram(fill = '#ff80bf', color = '#e60073', bins = 40)

## Calculation of the average strikes attempted each year
meanVector = 0
yearVector = 0
for(i in 2000:2020){
    meanVector[i - 1999] = mean(raw_total_fight_data[raw_total_fight_data$year␣
↪== paste0(i),]$TotalStrikesAttempted)
    yearVector[i - 1999] = i
}
strikesVyearDF = data.frame(yearVector, meanVector)

    ## Plotting of the strike attempts per year
    ggplot(strikesVyearDF, aes(yearVector, meanVector)) +
        geom_line(color = '#ff80bf') +
        scale_x_continuous(breaks=seq(1999,2020,by=3)) +
        labs(x = '', y = 'Strikes') +
        theme_bw()
```


count
30
20
10
0
0
200
400
600
800
a


Strikes
225
200
175
150
125
1999
2002
2005
2008
2011
2014
2017
2020

1.1.3 Correlation calculations for control time vs strikes

[11]:
```
controlTimevStrikes = data.frame(strikesVyearDF$meanVector,␣
↪controlTimeVyearDF$meanVector, controlTimeVyearDF$year)
colnames(controlTimevStrikes) = c('Strikes', 'controlTime', 'year')

paste0('The correlation coeffecient between control time and strikes is: ',␣
↪round(cor(controlTimevStrikes$controlTime,controlTimevStrikes$Strikes), 3))
```

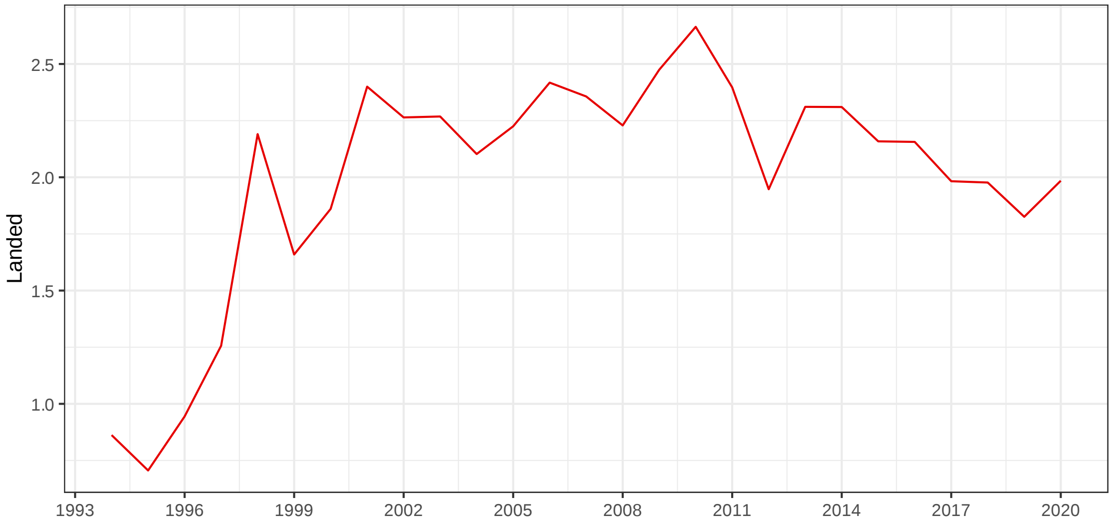
'The correlation coeffecient between control time and strikes is: -0.753'

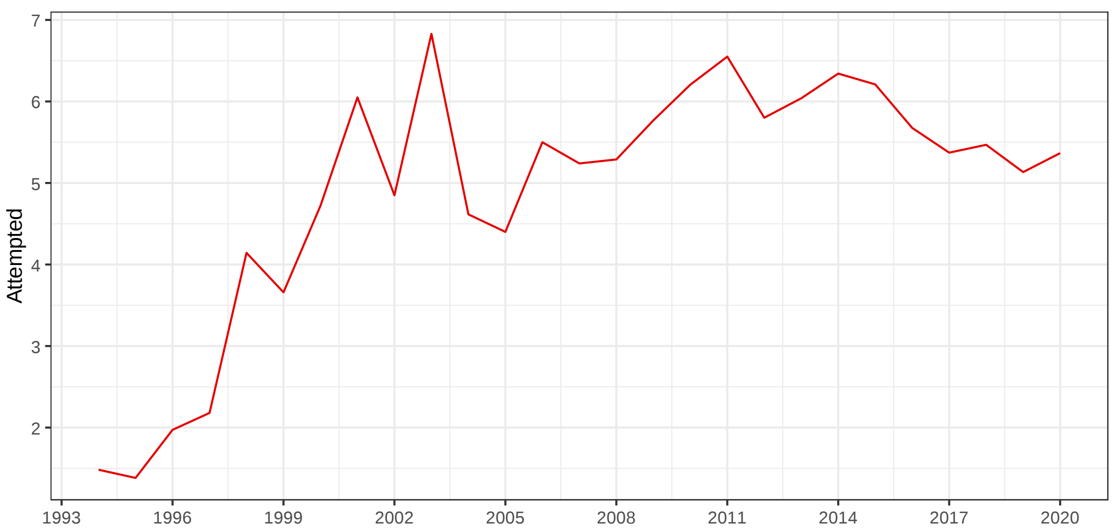
1.1.4 Takedown attempts

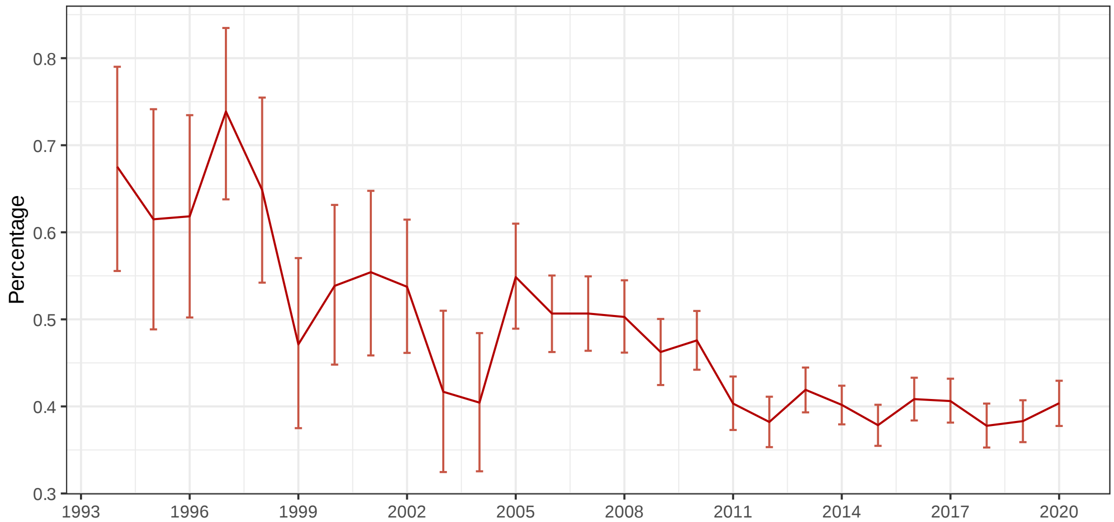
[37]:
```
# Calculation of the average strikes attempted each year
meanVector = 0
yearVector = 0
for(i in 1994:2020){
    meanVector[i - 1993] = mean(raw_total_fight_data[raw_total_fight_data$year␣
↪== paste0(i),]$TotalTDLanded)
    yearVector[i - 1993] = i
}
TDLandedDF = data.frame(yearVector, meanVector)

    ## Plotting of the strike attempts per year
    ggplot(TDLandedDF, aes(yearVector, meanVector)) +
        geom_line(color = '#e60000') +
        scale_x_continuous(breaks=seq(1993,2020,by=3)) +
        labs(x = '', y = 'Landed') +
        theme_bw()

## Calculation of the average strikes attempted each year
meanVector = 0
yearVector = 0
for(i in 1994:2020){
    meanVector[i - 1993] = mean(raw_total_fight_data[raw_total_fight_data$year␣
↪== paste0(i),]$TotalTDAttempted)
    yearVector[i - 1993] = i
}
TDAttemptedDF = data.frame(yearVector, meanVector)

    ## Plotting of the strike attempts per year
    ggplot(TDAttemptedDF, aes(yearVector, meanVector)) +
        geom_line(color = '#e60000') +
        scale_x_continuous(breaks=seq(1993,2020,by=3)) +
        labs(x = '', y = 'Attempted') +
        theme_bw()

## Calculation of the average strikes attempted each year
```

```
meanVector = 0
yearVector = 0
for(i in 1994:2020){
    meanVector[i - 1993] = mean(na.
 ↪omit(raw_total_fight_data[raw_total_fight_data$year ==␣
 ↪paste0(i),]$TDPercentage))
    yearVector[i - 1993] = i
}
# TDPercentageDF = data.frame(yearVector, meanVector)

options(repr.plot.width = 8, repr.plot.height = 4)
#     ## Plotting of the strike attempts per year
#     ggplot(TDPercentageDF, aes(yearVector, meanVector)) +
#         geom_line(color = '#b30000') +
#         scale_x_continuous(breaks=seq(1993,2020,by=3)) +
#         labs(x = '', y = 'Percentage') +
#         theme_bw()

# options(repr.plot.width = 6, repr.plot.height = 4)

confidenceInterval = matrix(NA, 27, 2)


## Calculation of the average strikes attempted each year
B = 10000
alpha = 0.10
for(i in 1994:2020){
    sampleData = na.omit(raw_total_fight_data[raw_total_fight_data$year ==␣
 ↪paste0(i),]$TDPercentage)
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
 ↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)

    confidenceInterval[i - 1993, 1] = quantile(thetaStar, alpha/2)
    confidenceInterval[i - 1993, 2] = quantile(thetaStar, 1 - alpha/2)

}

plus = confidenceInterval[,2]
minus = confidenceInterval[,1]

TDPercentageDF = data.frame(yearVector, meanVector, plus, minus)

    ggplot(TDPercentageDF, aes(yearVector, meanVector)) +
        geom_errorbar(aes(x = yearVector, y = meanVector, ymax = plus,
                          ymin = minus), width = 0.2, color = '#C75645') +
        geom_line(color = '#b30000') +
```

```
    scale_x_continuous(breaks=seq(1993,2020,by=3)) +
    labs(x = '', y = 'Percentage') +
    theme_bw()
```



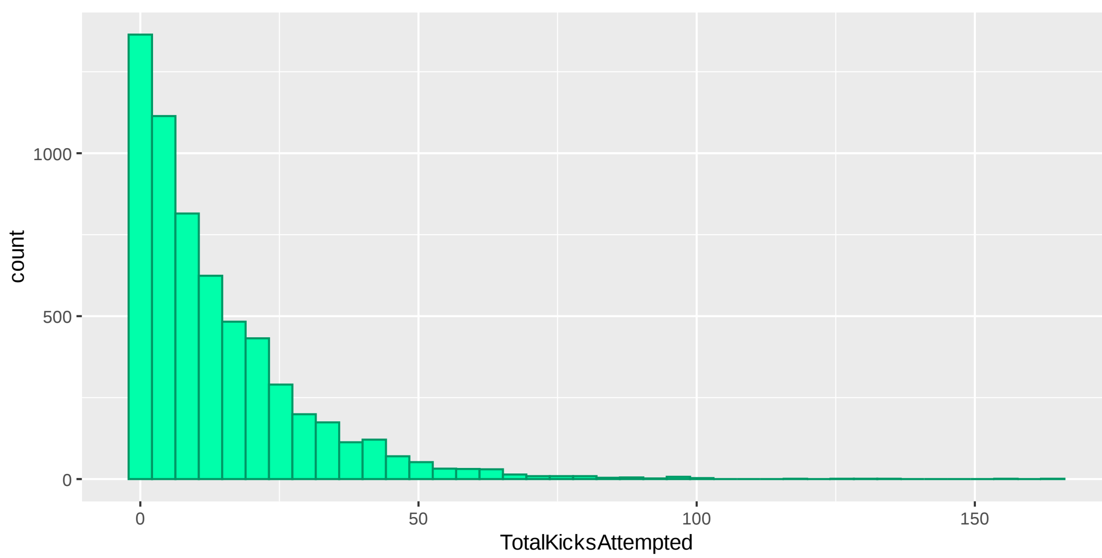


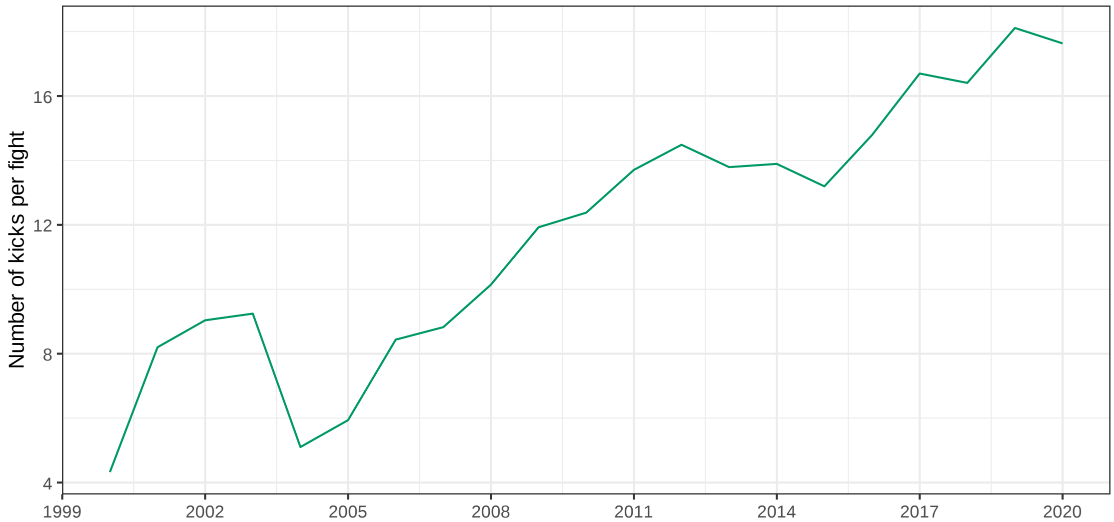


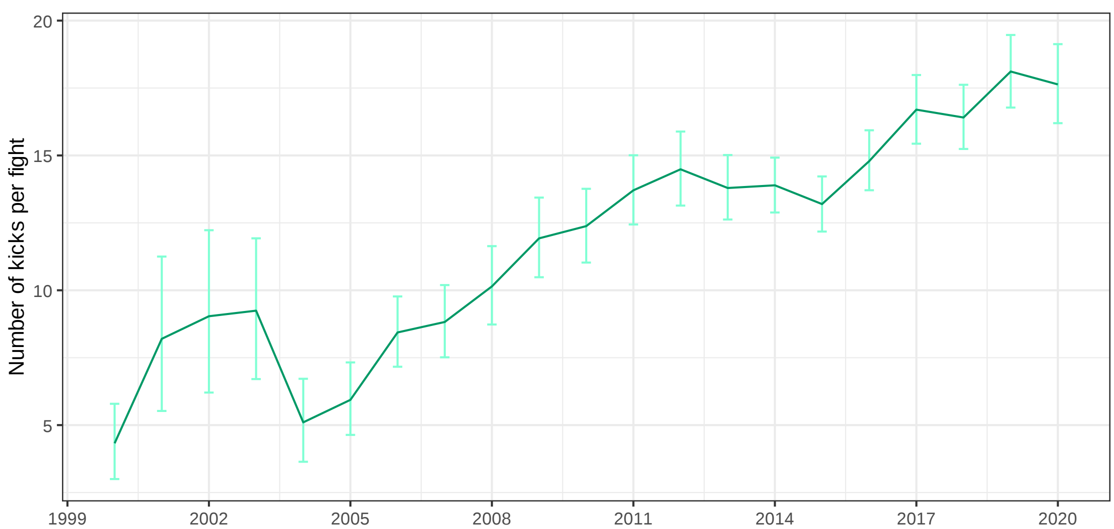
1.2 Leg kick analysis

```
[43]: ggplot(raw_total_fight_data, aes(TotalKicksAttempted)) +
          geom_histogram(bins = 40, fill = '#00ffaa', color = '#009966')

      ## Calculation of the average strikes attempted each year
      meanVector = 0
      yearVector = 0
      for(i in 2000:2020){
          meanVector[i - 1999] = mean(raw_total_fight_data[raw_total_fight_data$year
       ↪== paste0(i),]$TotalKicksAttempted)
          yearVector[i - 1999] = i
      }
      kicksVyearDF = data.frame(yearVector, meanVector)

          ## Plotting of the strike attempts per year
          ggplot(kicksVyearDF, aes(yearVector, meanVector)) +
              geom_line(color = '#009966') +
              scale_x_continuous(breaks=seq(1999,2020,by=3)) +
              labs(x = '', y = 'Number of kicks per fight') +
              theme_bw()

      StrikesvKicks = data.frame(kicksVyearDF$meanVector, strikesVyearDF$meanVector,
       ↪strikesVyearDF$year)
      colnames(StrikesvKicks) = c('Kicks', 'Strikes', 'year')

      confidenceInterval = matrix(NA, 21, 2)
```

```
B = 10000
alpha = 0.10
for(i in 2000:2020){
    sampleData = raw_total_fight_data[raw_total_fight_data$year ==␣
 ↪paste0(i),]$TotalKicksAttempted
    bootstrapSamples = t(replicate(B, sample(sampleData, length(sampleData),␣
 ↪TRUE)))
    thetaStar = apply(bootstrapSamples, 1, mean)

    confidenceInterval[i - 1999, 1] = quantile(thetaStar, alpha/2)
    confidenceInterval[i - 1999, 2] = quantile(thetaStar, 1 - alpha/2)

}

plus = confidenceInterval[,2]
minus = confidenceInterval[,1]

kicksVyearDF = data.frame(yearVector, meanVector, plus, minus)

    ggplot(kicksVyearDF, aes(yearVector, meanVector)) +
        geom_errorbar(aes(x = yearVector, y = meanVector, ymax = plus,
                            ymin = minus), width = 0.2, color = '#80ffd4') +
        geom_line(color = '#009966') +
        scale_x_continuous(breaks=seq(1993,2020,by=3)) +
        labs(x = '', y = 'Number of kicks per fight') +
        theme_bw()
```





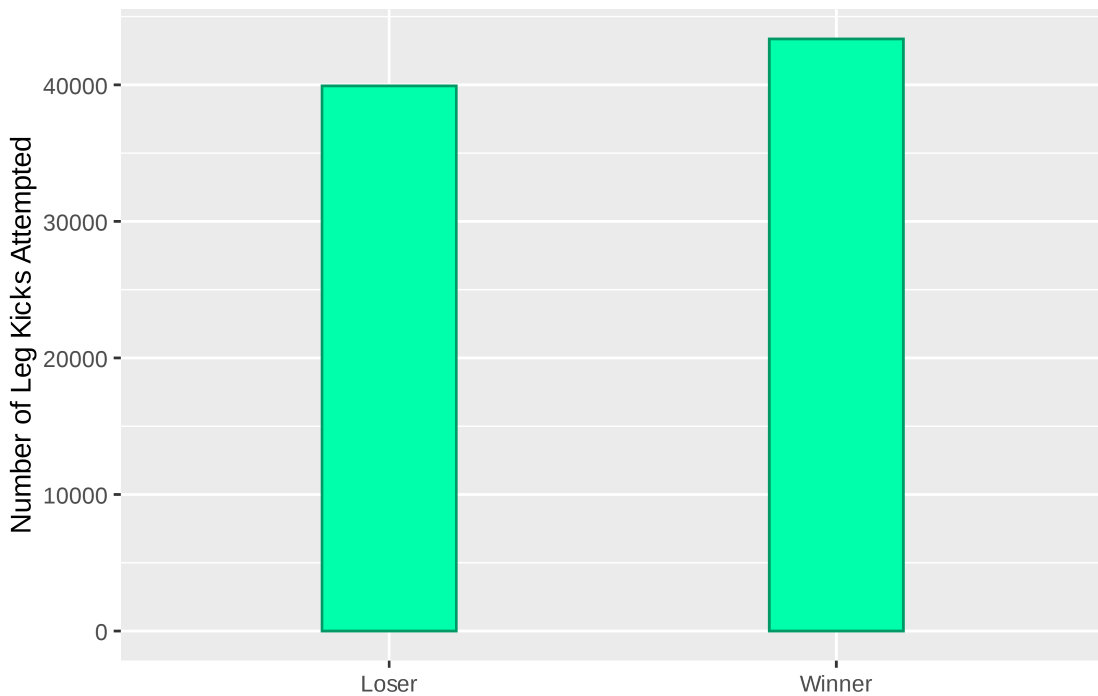


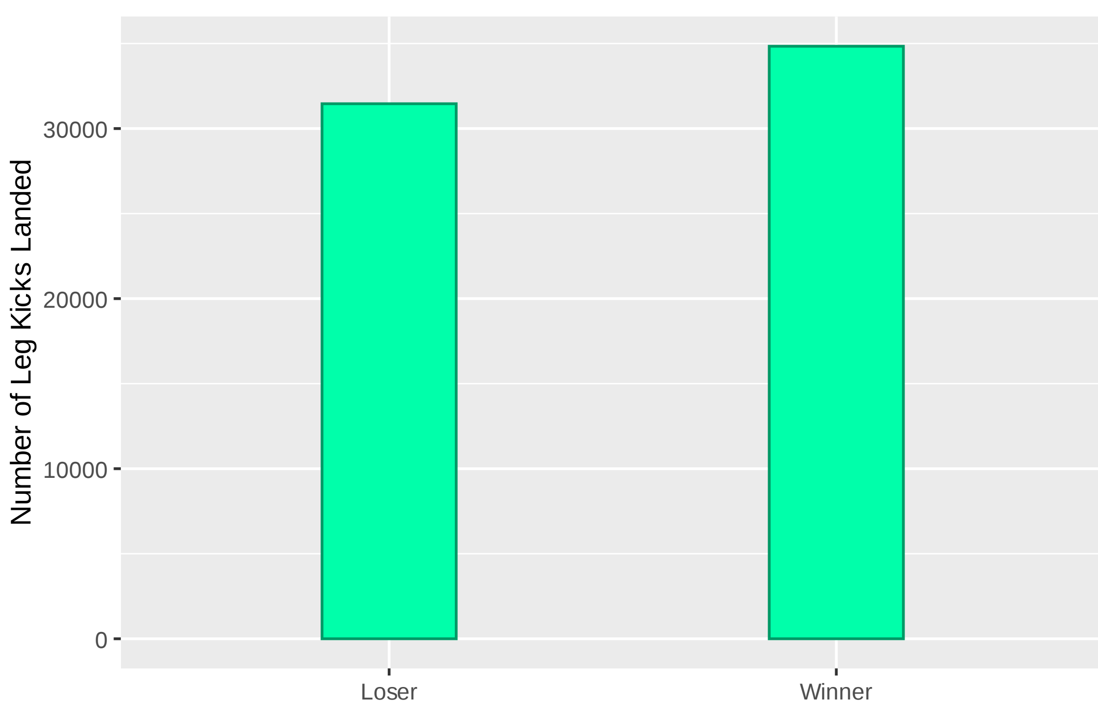
[15]:
```
paste0('The correlation coeffecient between Strikes and Kicks is: ',␣
↪round(cor(StrikesvKicks$Strikes, StrikesvKicks$Kicks), 3))
```

'The correlation coeffecient between Strikes and Kicks is: 0.98'

[16]:
```
## Kicks attempted
```

```
  WinnerLegAttempts =␣
↪c(raw_total_fight_data[which(raw_total_fight_data$RedWon == 1),␣
↪"RLegKicksAttempted"],␣
↪raw_total_fight_data[which(raw_total_fight_data$BlueWon == 1),␣
↪"BLegKicksAttempted"])
  LoserLegAttempts = c(raw_total_fight_data[which(raw_total_fight_data$RedWon␣
↪== 1), "BLegKicksAttempted"],␣
↪raw_total_fight_data[which(raw_total_fight_data$BlueWon == 1),␣
↪"RLegKicksAttempted"])

  legKickDF = data.frame(WinnerLegAttempts, LoserLegAttempts)

  kickAttemptedDF = data.
↪frame(c('Winner','Loser'),c(sum(legKickDF$WinnerLegAttempts),␣
↪sum(legKickDF$LoserLegAttempts)))
  colnames(kickAttemptedDF) = c('who','kicks')

  ggplot(kickAttemptedDF, aes(x = who, y = kicks)) +
      geom_bar(stat = "identity", fill = '#00ffaa', color = '#009966', width␣
↪= 0.3) +
      labs(x = '', y = 'Number of Leg Kicks Attempted')



## Kicks landed
  WinnerLegLanded = c(raw_total_fight_data[which(raw_total_fight_data$RedWon␣
↪== 1), "RLegKicksLanded"],␣
↪raw_total_fight_data[which(raw_total_fight_data$BlueWon == 1),␣
↪"BLegKicksLanded"])
  LoserLegLanded = c(raw_total_fight_data[which(raw_total_fight_data$RedWon␣
↪== 1), "BLegKicksLanded"],␣
↪raw_total_fight_data[which(raw_total_fight_data$BlueWon == 1),␣
↪"RLegKicksLanded"])

  legKickDF = data.frame(WinnerLegLanded, LoserLegLanded)

  kickLandedDF = data.frame(c('Winner', 'Loser'),␣
↪c(sum(legKickDF$WinnerLegLanded), sum(legKickDF$LoserLegLanded)))
  colnames(kickLandedDF) = c('who', 'kicks')

      ggplot(kickLandedDF, aes(x = who, y = kicks)) +
          geom_bar(stat = "identity", fill = '#00ffaa', color = '#009966',␣
↪width = 0.3) +
          labs(x = '', y = 'Number of Leg Kicks Landed')
```


Number of Leg Kicks Attempted
40000
30000
20000
10000
0
Loser
Winner


Number of Leg Kicks Landed
30000
20000
10000
0
Loser
Winner

[17]:

```
## Percentages
    paste0('Winner land percentage: ', round(mean(WinnerLegLanded/
 ↪WinnerLegAttempts, na.rm = TRUE), 3))
    paste0('Loser land percentage: ', round(mean(LoserLegLanded/
 ↪LoserLegAttempts, na.rm = TRUE), 3))


sd(WinnerLegLanded/WinnerLegAttempts, na.rm = TRUE)
sd(LoserLegLanded/LoserLegAttempts, na.rm = TRUE)
```

'Winner land percentage: 0.812'

'Loser land percentage: 0.788'

0.236230790151112

0.255862570675009