

Predicting UFC Fight Results

2023-12-11

```
##
## 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

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':
##
##   combine

## The following object is masked from 'package:ggplot2':
##
##   margin

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##   lift
```

```
## Rows: 6012 Columns: 144
## -- Column specification -----
## Delimiter: ","
## chr   (9): R_fighter, B_fighter, Referee, date, location, Winner, weight_cla...
## dbl (134): B_avg_KD, B_avg_opp_KD, B_avg_SIG_STR_pct, B_avg_opp_SIG_STR_pct,...
## lgl   (1): title_bout
##
## 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.
```

Introduction

The UFC began as a professional mixed martial arts organization in 1993 serving as an alternative hand-to-hand combat sport that combined traditional boxing with wrestling, karate, kickboxing, and jiu-jitsu fighting techniques. The entity was acquired by a group led by Dana White in 2001, who has served as the President for over two decades. Since establishing control, Dana White has exponentially grown the reach of UFC's product while also creating more structure and sanction to the sport of MMA. UFC currently has over 60 global broadcasting partners and is able to be accessed in over 165 different countries. With a traveling, tour-like model, the UFC has been able to sell-out many arenas across the world as equally become a highly-touted event to attend similar to boxing matches with well-known participants involved.

In many other North-american based sports, organizations have invested and founded their own analytics departments. These departments are responsible for using data to acquire and develop the right talent that will lead to on-field success and improve the team's product. Since the UFC's participants are individual fighters that often follow their own training regiment, there is a smaller focus on analytics within the sport.

The purpose of our project is two-fold. We want to evaluate fighters' historical data to determine fight styles that may possess a stronger correlation to success within the octagon. Identifying important factors will allow UFC fighters and their hired trainers to optimize their training regime, and will also benefit commentators in pointing out facets of the match the audience should keep in mind while spectating.

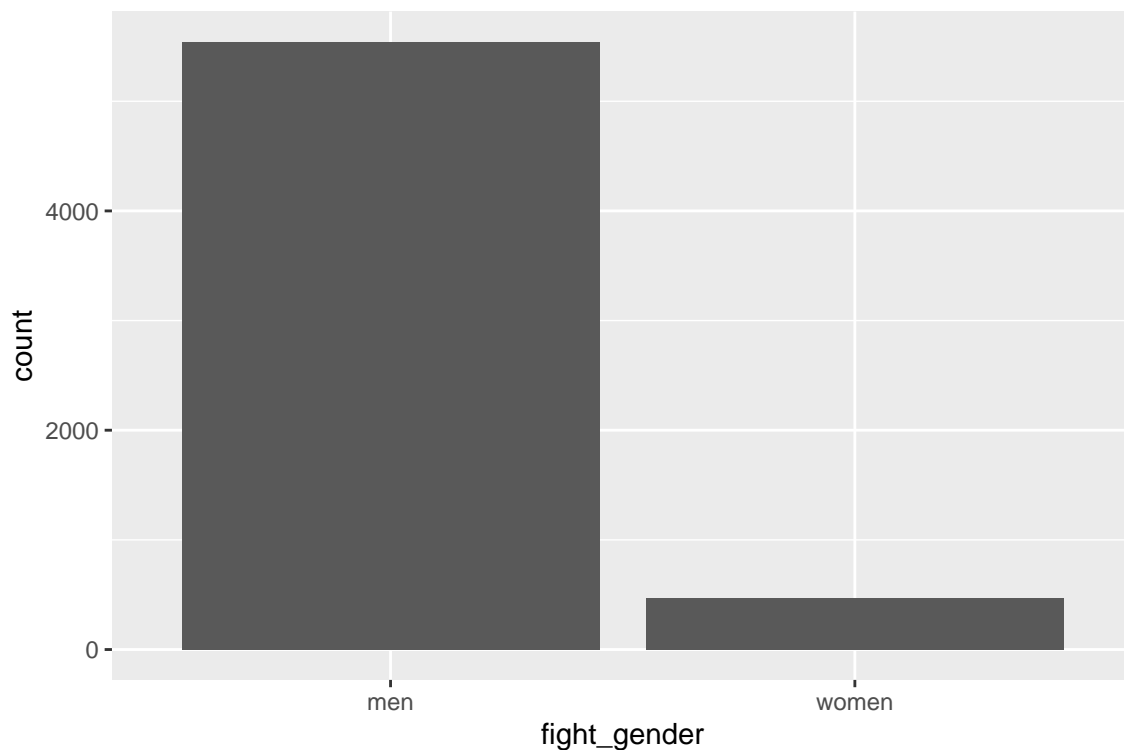
Additionally, we want to create a model that maximizes predictive accuracy for the purposes of assisting sports bettors in finding potential opportunities for value not seen by the public. Dana White and the UFC have fully embraced the recent popularity of sports betting, forming sponsorships with companies such as bet365 in the UK, DraftKings in the US, as well as many others located worldwide. There is an established market for sports betting in the UFC, and we hope to create a model that provides an estimation of a winner between two fighters along with some form of uncertainty that allows a sports-better to determine if the predicted odds over or under-estimate a fighter's chance of winning compared to the sportsbook odds given to the public.

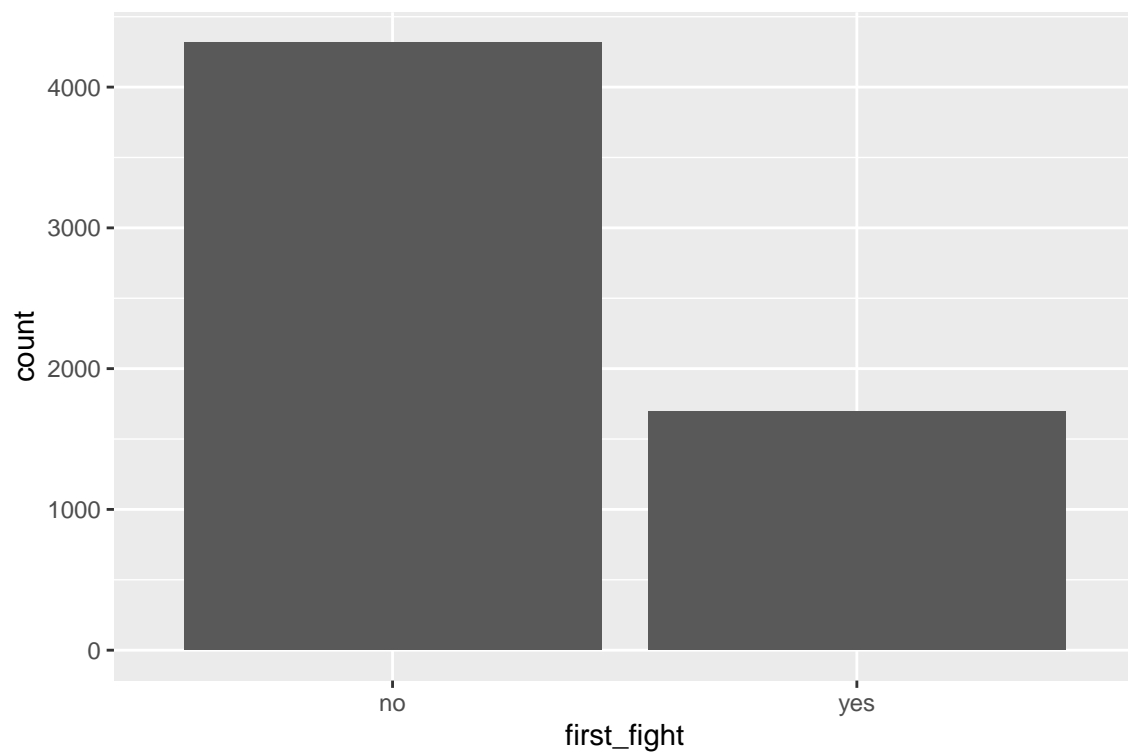
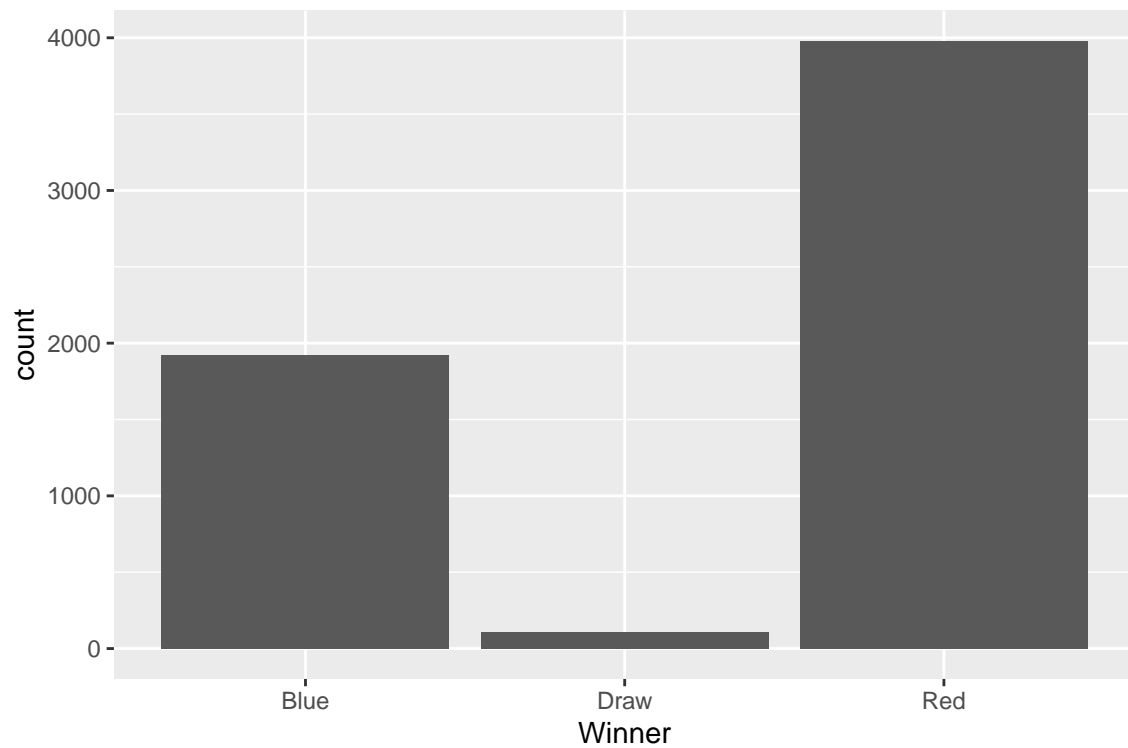
Our data is each UFC fight from 1993-2021, containing each fighter's names, physical information (age, height, weight), the amount of wins and losses in their UFC history, and various fighting data. The fighting data includes the average amount of attempted and landed attacks over their UFC career, as well as the amount of different types of attacks they have faced from their previous opponents. We plan to initially fit an elastic net model that incorporates the standardization of a ridge regression model, and the variable selection of a lasso regression model, to determine which predictors in our dataset are most influential. Our response variable will be the winner in each fight, with that value randomized dependent on the color of the corner assigned during the fight (red or blue). The elastic net model will be best served for UFC commentators and trainers interested in how prior fighting strategies can lead to success in the future. From there, we plan to evaluate the inclusion of nonlinear effects within the most influential predictors in hopes to maximize predictive accuracy.

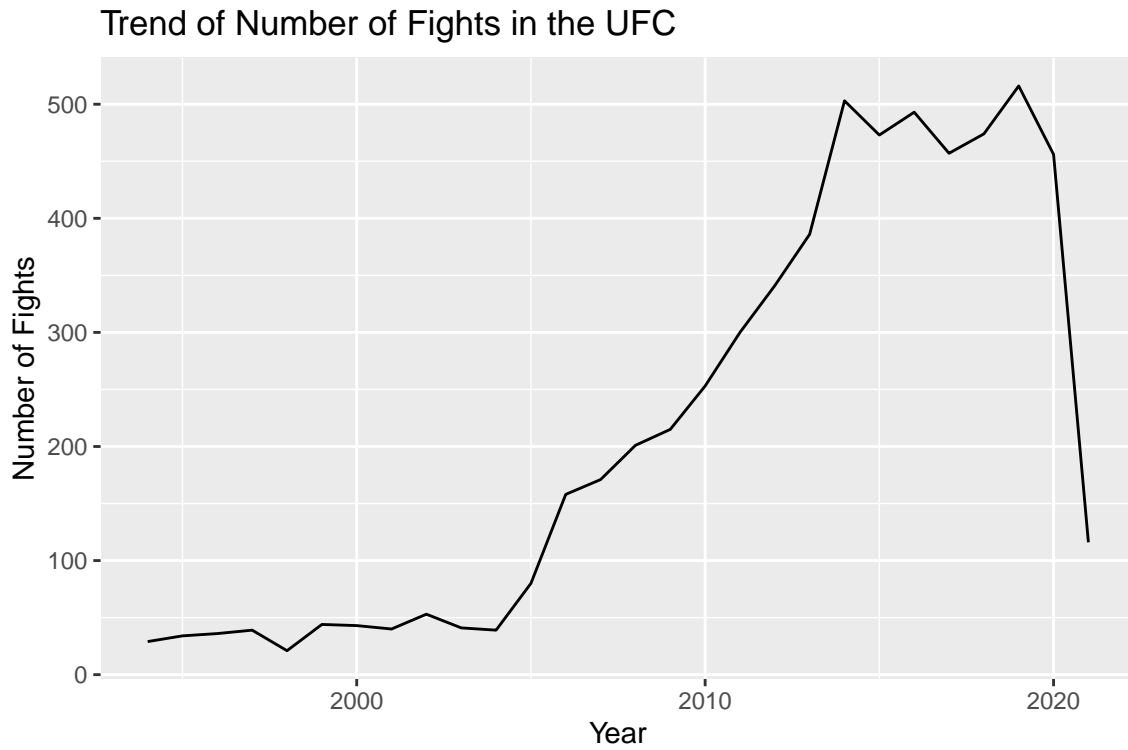
Data Description and EDA

The dataset used for modeling was initially sourced from ufcstats.com, where the data was processed and published on [kaggle.com](https://www.kaggle.com). Using

```
## # A tibble: 14 x 2
##   weight_class      count
##   <chr>          <int>
## 1 Bantamweight      475
## 2 CatchWeight       54
## 3 Featherweight    551
## 4 Flyweight        230
## 5 Heavyweight       585
## 6 LightHeavyweight  573
## 7 Lightweight     1091
## 8 Middleweight     813
## 9 OpenWeight        86
## 10 Welterweight    1083
## 11 WomenBantamweight 151
## 12 WomenFeatherweight 16
## 13 WomenFlyweight   112
## 14 WomenStrawweight 192
```

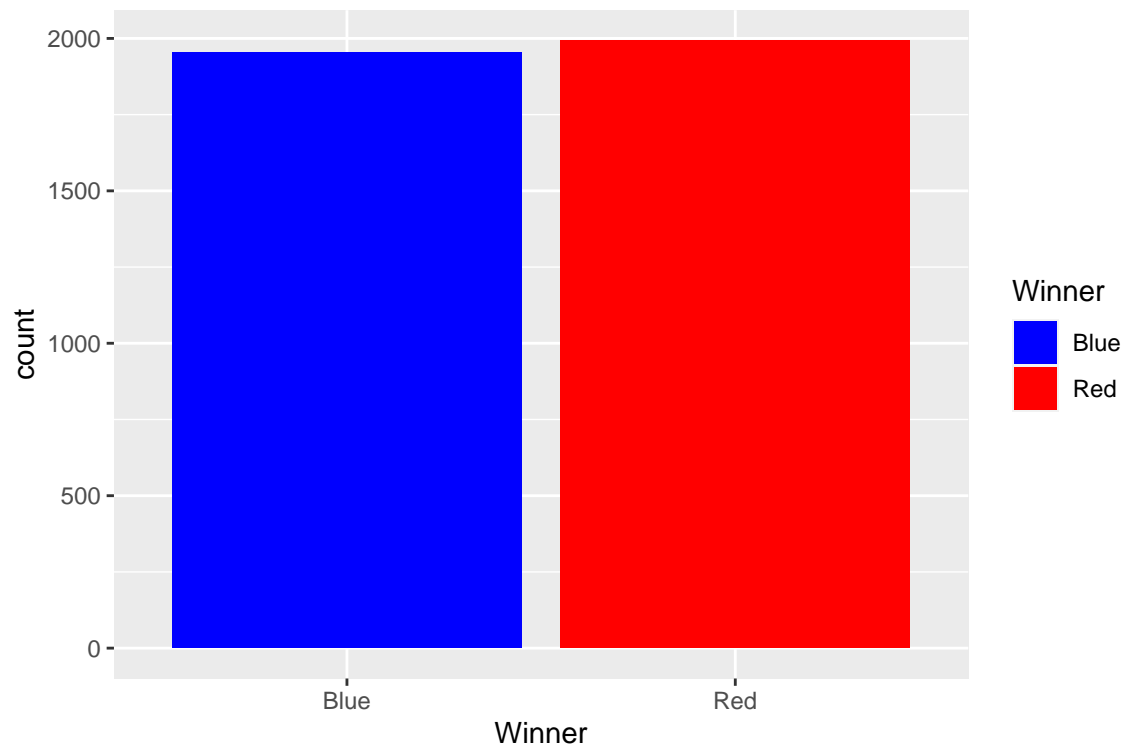






```
## [1] 3948
```

```
#randomizing blue and red
```



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

Table 1: Frequency of Weight Class Fights

new_class	weight_class	weight	count
Class 1	Flyweight	125	175
Class 1	Bantamweight	135	339
Class 1	Featherweight	145	408
Class 2	Lightweight	155	797
Class 2	Welterweight	170	797
Class 3	Middleweight	185	590
Class 3	LightHeavyweight	205	424
Class 4	Heavyweight	265	389
Class 4	OpenWeight	300	29

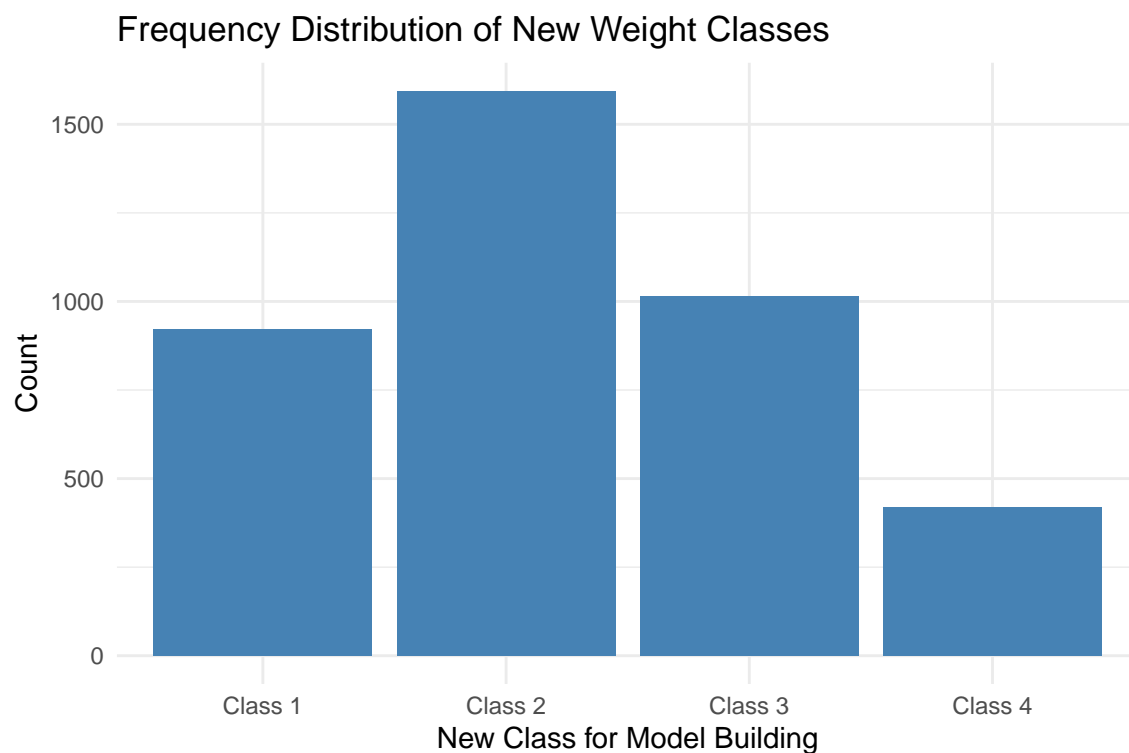


Table 2: Avg Attempt of Different Style Attacks

new_class	Rmean_head	Rmean_body	Rmean_leg	Rmean_clinch	Rmean_ground	Rmean_ctrltime
Class 1	73.2	12.0	8.1	7.4	8.6	141.3
Class 2	62.0	10.2	7.0	7.9	8.3	153.8
Class 3	52.3	8.1	6.1	8.1	8.9	136.3
Class 4	43.0	6.7	4.9	7.2	8.5	106.8

```
## [1] 371
```

Random Forest Model Creation

```
## [1] Red Blue Blue Blue Red Blue Red Red Red Red Blue Blue Red Red Blue
## [16] Blue Red Blue Red Red Blue Blue Blue Blue Red Red Red Blue Blue Red
## [31] Blue Blue Red Blue Blue Blue Red Blue Blue Blue Blue Red Red Red Red
## [46] Blue Blue Red Red Blue Red Red Blue Red Red Red Red Red Blue Blue Blue
## [61] Red Blue Red Blue Red Blue Red Blue Blue Red Red Red Red Red Red Blue
## [76] Red Red Blue Red Blue Red Red Blue Blue Blue Blue Blue Blue Blue Blue
## [91] Blue Blue Blue Blue Red Blue Red Red Blue Red Red Blue Blue Blue Blue
## [106] Red Blue Red Blue Blue Red Red Red Red Red Red Red Red Red Red Blue
## [121] Blue Red Red Red Blue Blue Red Red Blue Red Red Red Red Red Red Red
## [136] Blue Red Red Red Blue Blue Blue Red Red Red Red Red Red Red Red Red
## [151] Blue Blue Blue Red Blue Red Blue Blue Red Red Red Red Red Red Red Red
## [166] Red Red Red Red Blue Blue Blue Blue Red Blue Red Red Red Red Red Red
## [181] Blue Blue Red Red Blue Blue Blue Red Blue Red Blue Red Red Red Red Red
## [196] Red Red Blue Red Red Blue Blue Red Blue Blue Red Red Red Red Red Red
## [211] Red Blue Red Blue Blue Red Red Blue Blue Blue Red Red Red Red Red Red
## [226] Red Red Red Red Blue Blue Blue Red Red Blue Red Red Red Red Red Red
## [241] Red Red Blue Red Blue Red Red Blue Blue Red Red Red Red Red Red Red
## [256] Red Blue Blue Blue Blue Blue Blue Blue Red Red Red Red Red Red Red
## [271] Red Blue Blue Red Red Blue Red Red Blue Blue Blue Red Red Red Red
## [286] Red Blue Red Red Red Red Red Red Red Red Red Red Red Red Red Red
## [301] Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red Red Red
## [316] Blue Blue Red Red Red Blue Blue Red Blue Red Red Red Red Red Red Red
## [331] Red Blue Red Blue Red Red Blue Blue Red Blue Blue Red Red Red Red Red
## [346] Red Blue Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red
## [361] Blue Blue Red Red Red Blue Blue Red Red Red Red Red Red Red Red Red
## [376] Blue Blue Red Blue Red Blue Blue Red Blue Red Red Red Red Red Red Red
## [391] Red Red Red Blue Blue Blue Red Red Red Red Red Red Red Red Red Red
## [406] Blue Red Blue Red Red Red Red Red Red Red Red Red Red Red Red Red
## [421] Blue Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red Red
## [436] Red Blue Blue Red Blue Red Red Red Red Red Red Red Red Red Red Red
## [451] Red Blue Blue Blue Red Blue Red Red Red Red Red Red Red Red Red Red
## [466] Blue Red Blue Blue Red Blue Blue Blue Red Red Red Red Red Red Red Red
## [481] Blue Blue Red Red Red Red Red Red Red Red Red Red Red Red Red Red
## [496] Blue Red Red Red Red Red Red Red Red Red Red Red Red Red Red Red
## [511] Red Red Red Blue Blue Red Red Red Red Red Red Red Red Red Red Red
## [526] Red Blue Blue Red Red Red Red Red Red Red Red Red Red Red Red Red
## [541] Red Blue Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red
## [556] Red Red Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red
## [571] Blue Blue Red Blue Red Blue Red Red Red Red Red Red Red Red Red Red
## [586] Blue Blue Blue Blue Blue Red Red Red Red Red Red Red Red Red Red Red
## [601] Red Red Blue Red Blue Red Red Red Red Red Red Red Red Red Red Red
## Levels: Blue Red
```