PSTAT131 Final Project

Yifei Zhang

2022-05-11

Contents

```
library(ggplot2)
library(tidyverse)
library(tidymodels)
library(corrplot)
library(ggthemes)
library(discrim)
library(poissonreg)
library(corrr)
library(klaR)
library(ISLR)
library(ISLR2)
library(purrr)
library(janitor)
tidymodels_prefer()
loldata <- read_csv("high_diamond_ranked_10min.csv")
```

Begin tidying

```
lol <- clean_names(loldata)
lol
```

```
## # A tibble: 9,879 x 40
##      game_id blue_wins blue_wards_placed blue_wards_destroyed blue_first_blood
##        <dbl>     <dbl>             <dbl>                <dbl>            <dbl>
##  1 4519157822         0                28                    2                1
##  2 4523371949         0                12                    1                0
##  3 4521474530         0                15                    0                0
##  4 4524384067         0                43                    1                0
##  5 4436033771         0                75                    4                0
##  6 4475365709         1                18                    0                0
##  7 4493010632         1                18                    3                1
##  8 4496759358         0                16                    2                0
##  9 4443048030         0                16                    3                0
```

```
## 10 4509433346             1                  13                   1                   1
## # ... with 9,869 more rows, and 35 more variables: blue_kills <dbl>,
## #   blue_deaths <dbl>, blue_assists <dbl>, blue_elite_monsters <dbl>,
## #   blue_dragons <dbl>, blue_heralds <dbl>, blue_towers_destroyed <dbl>,
## #   blue_total_gold <dbl>, blue_avg_level <dbl>, blue_total_experience <dbl>,
## #   blue_total_minions_killed <dbl>, blue_total_jungle_minions_killed <dbl>,
## #   blue_gold_diff <dbl>, blue_experience_diff <dbl>, blue_cs_per_min <dbl>,
## #   blue_gold_per_min <dbl>, red_wards_placed <dbl>, ...
```

```
lol <-  lol[ , 0:21]
lol
```

```
## # A tibble: 9,879 x 21
##        game_id blue_wins blue_wards_placed blue_wards_destroyed blue_first_blood
##          <dbl>     <dbl>             <dbl>                <dbl>            <dbl>
##  1 4519157822         0                28                    2                1
##  2 4523371949         0                12                    1                0
##  3 4521474530         0                15                    0                0
##  4 4524384067         0                43                    1                0
##  5 4436033771         0                75                    4                0
##  6 4475365709         1                18                    0                0
##  7 4493010632         1                18                    3                1
##  8 4496759358         0                16                    2                0
##  9 4443048030         0                16                    3                0
## 10 4509433346         1                13                    1                1
## # ... with 9,869 more rows, and 16 more variables: blue_kills <dbl>,
## #   blue_deaths <dbl>, blue_assists <dbl>, blue_elite_monsters <dbl>,
## #   blue_dragons <dbl>, blue_heralds <dbl>, blue_towers_destroyed <dbl>,
## #   blue_total_gold <dbl>, blue_avg_level <dbl>, blue_total_experience <dbl>,
## #   blue_total_minions_killed <dbl>, blue_total_jungle_minions_killed <dbl>,
## #   blue_gold_diff <dbl>, blue_experience_diff <dbl>, blue_cs_per_min <dbl>,
## #   blue_gold_per_min <dbl>
```

```
lol %>%
  ggplot(aes(x = blue_wins)) +
  geom_histogram(bins = 3)
```
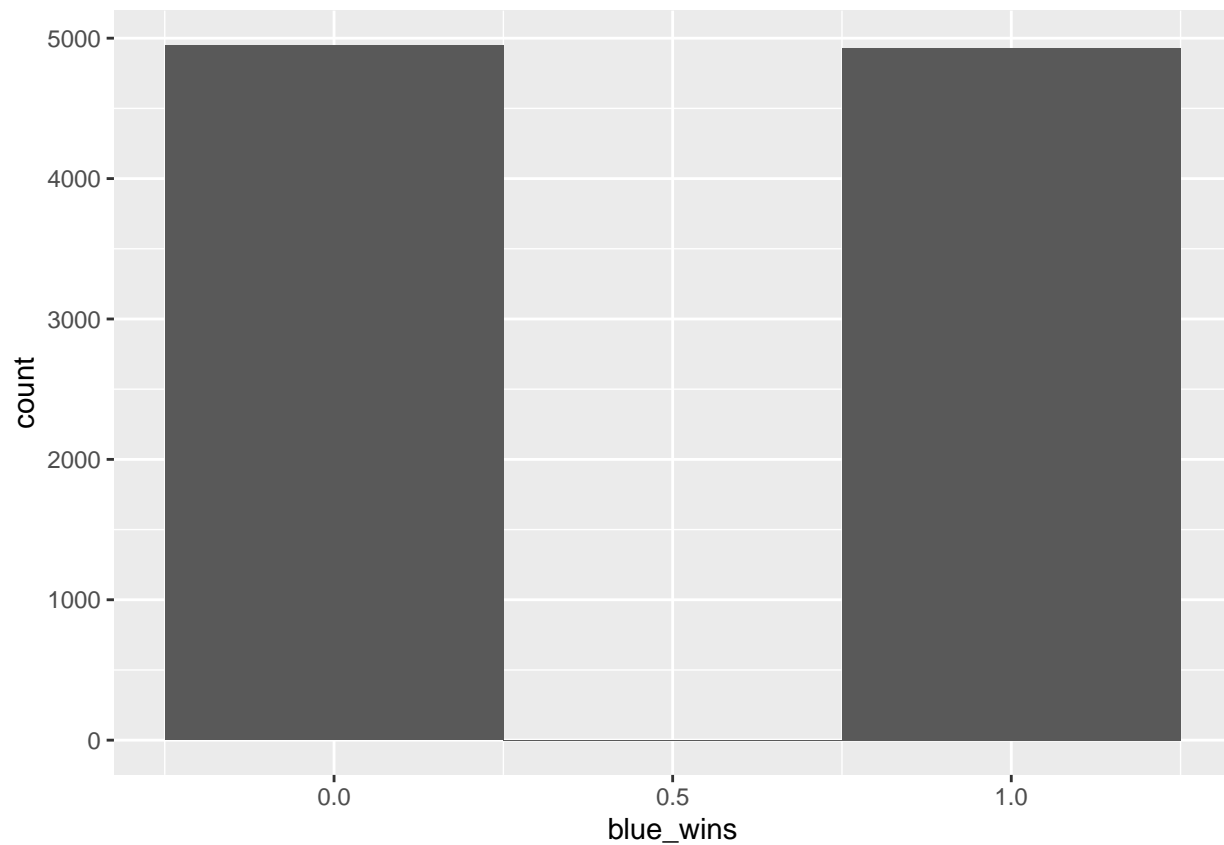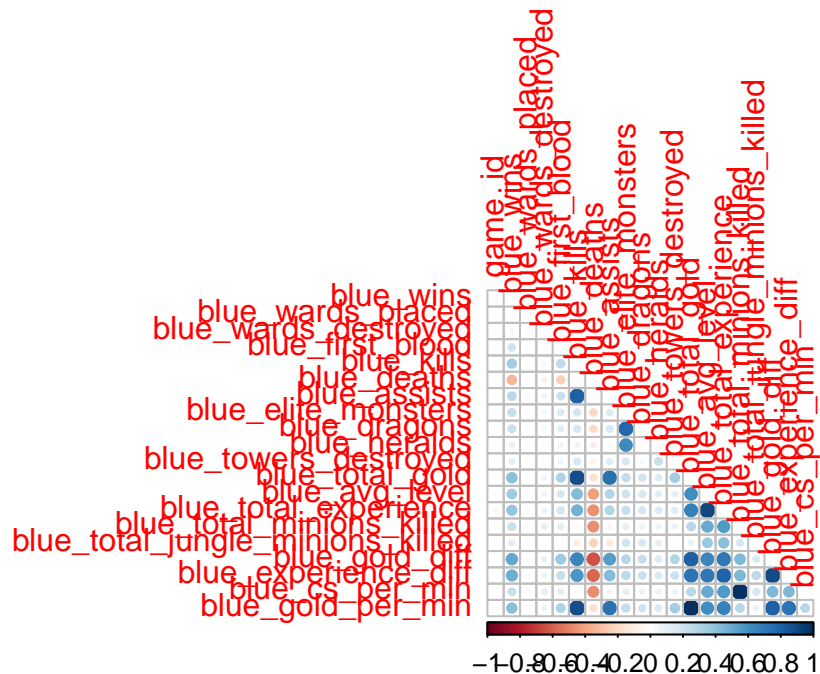
```
# there is a very slight difference between the number of blue win and lose
```

Setting

```
set.seed(1010)
lol_split <- lol %>%
  initial_split(strata = blue_wins, prop = 0.75)
lol_train <- training(lol_split)
lol_test <- testing(lol_split)
dim(lol_train)
```

```
## [1] 7408    21
```

```
lol_train %>%
  select(is.numeric) %>%
  cor(use = "complete.obs") %>%
  corrplot(type = "lower", diag = FALSE)
```

```
pokemon_folds <- vfold_cv(lol_train, v = 5, strata = 'blue_wins')
pokemon_folds
```

```
## #  5-fold cross-validation using stratification
## # A tibble: 5 x 2
##   splits              id
##   <list>              <chr>
## 1 <split [5925/1483]> Fold1
## 2 <split [5926/1482]> Fold2
## 3 <split [5927/1481]> Fold3
## 4 <split [5927/1481]> Fold4
## 5 <split [5927/1481]> Fold5
```

cols = ["gameId", "redFirstBlood", 'redKills', 'redEliteMonsters', 'redDragons', 'redTotalMinionsKilled', 'redTotalJungleMinionsKilled', 'redGoldDiff', 'redExperienceDiff', 'redCSPerMin', 'redGoldPerMin', 'redHeralds', 'blueGoldDiff', 'blueExperienceDiff', 'blueCSPerMin', 'blueGoldPerMin', 'blueTotalMinionsKilled']