

raport

Karol Mućk

4 04 2020

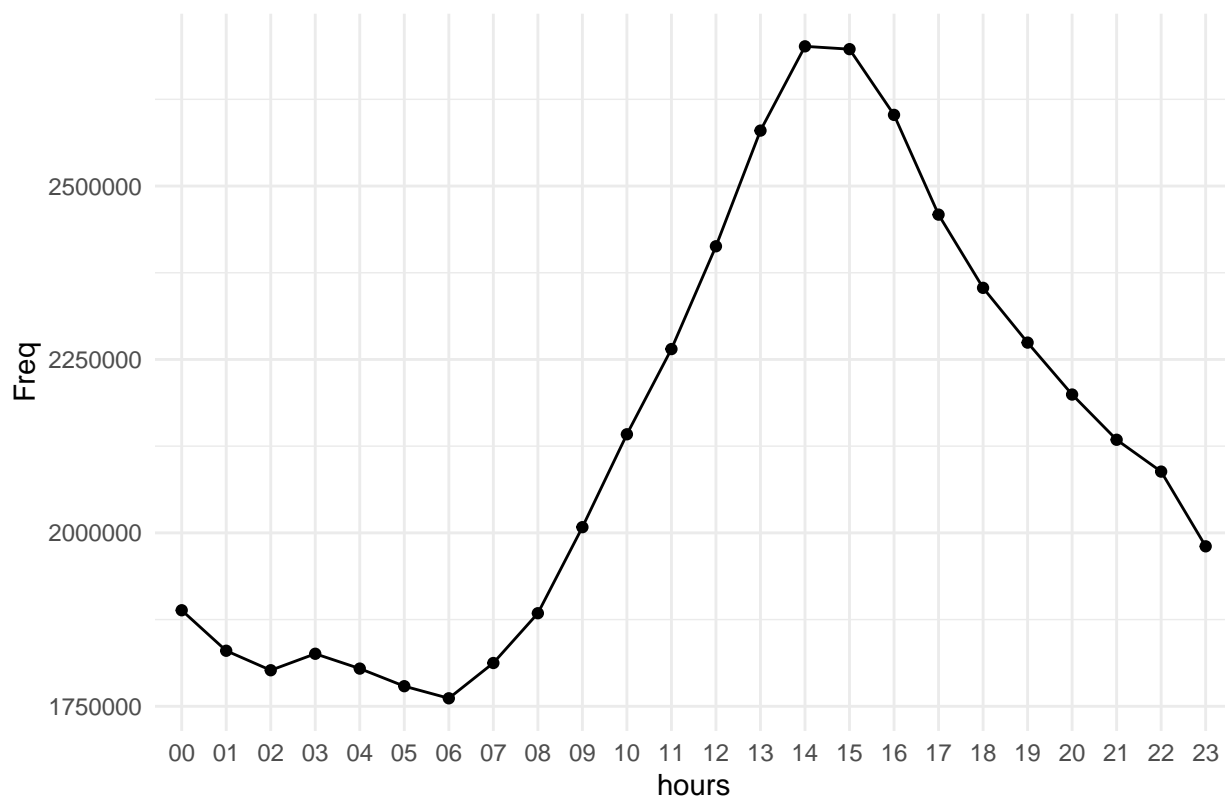
Celem projektu było przeanalizowanie danych z gry molbinej Clash Royale. Dane z których korzystałem zawierały dane odnośnie każdej rozegranej gry z stycznia 2019 roku. Jes to ponad 51 mln rozegranych gier.

Dane wyglądały następująco:

```
##          timestamp arena_id has_won      tag
## 1 2018-12-31T23:00:46Z      16      1 289VVL288
## 2 2018-12-31T23:01:34Z      14      1 2G0PRYVRU
## 3 2018-12-31T23:02:00Z      13      1  92GLVGQU
## 4 2018-12-31T23:02:00Z      14      1 2Y28RPJRC
## 5 2018-12-31T23:02:02Z      16      0  8RYLV2R
## 6 2018-12-31T23:02:21Z      12      1 8RPVVVRPV
##
##                                     player_deck
## 1          elixir-collector_golem_mega-knight_mega-minion_night-witch_rage_witch_zap
## 2              dark-prince_goblin-gang_miner_pekka_the-log_tornado_wizard_zap
## 3  baby-dragon_balloon_barbarian-barrel_bowler_freeze_inferno-dragon_lumberjack_tornado
## 4  baby-dragon_fireball_inferno-dragon_mega-minion_miner_royal-giant_the-log_tombstone
## 5              archers_barbarian-barrel_goblins_hog-rider_knight_poison_tesla_zap
## 6  goblin-barrel_hog-rider_inferno-tower_minion-horde_prince_rocket_skeleton-army_wizard
```

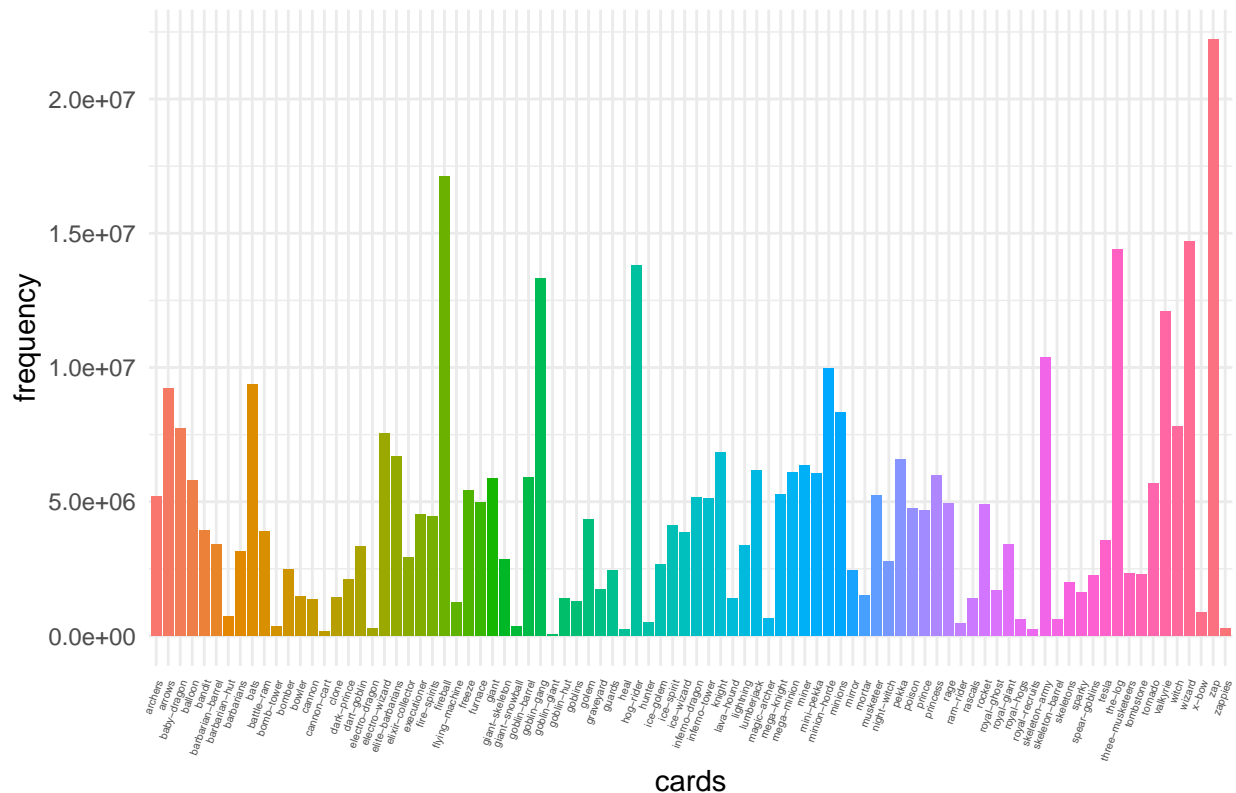
Pierwsza kolumna zawierała informację kiedy dana rozgrywka się odbyła. Druga kolumna pokazywała poziom areny na którym dana gra się toczyła, im wyższy poziom areny, tym wyższy poziom gry. Trzecia kolumna informowała czy dany gracz wygrał czy przegrał. Czwarta zawierała zakodowany nick gracza. Ostatnia zawierała informację jaką talią grał dany gracz, karty były podzielone znakiem "_". Deck składa się z 8 kart.

What time did the players play most often?



Widzimy, że godziny są rozłożone w bardzo naturalny sposób. Nikogo nie dziwi fakt że między 1 a 7 rano aktywność jest niska oraz że między 13 a 17 aktywność jest największa. Może świadczyć o tym że godziny są znormalizowane ze względu na strefy czasowe.

The frequency of occurrence of a card



Kart jest 90 więc taki wykres jest bardzo nie czytelny, wykres mówi w ilu grach dana karta wystąpiła, nie uwzględniając poszczególnych dni czy poziomu rozgrywki.

Możemy również zobaczyć jak użycie najpopularniejszych kart zmieniało się w czasie.

Figure 1 przedstawia jak popularność 15 najpopularniejszych kart zmieniała się w czasie. Widzimy, że 16 stycznia najprawdopodobniej doszło do zmiany w balansie kart.

Warto w takich analizach powiedzieć, że w użycie danej karty nie jest takie same na każdym poziomie rozgrywki. Kiedy zaczynamy grę gramy tymi kartami które są dostępne, z czasem odblokowujemy karty które są lepsze. Figure 2 przedstawia popularność 15 najbardziej popularnych kart w zależności od areny na jakiej występuje. Oczywiście liczba graczy jest różna na każdym poziomie, dlatego warto podzielić wynik przez liczbę rozegranych gier na danym poziomie.

Teraz wykorzystamy algorytm apiori do znalezienia częstych zbiorów oraz znalezienia asocjacji między poszczególnymi kartami.

#sposób wywołania funkcji apiori

```
#rules <- apriori (cards, parameter = list(supp = 0.001, conf = 0.5, maxtime=50000), control = list(mem
```

```
rules_conf <- sort (rules, by="confidence", decreasing=TRUE)
rules_lift <- sort (rules, by="lift", decreasing=TRUE)
rules_support <- sort (rules, by="support", decreasing=TRUE)
```

```
subrules_conf <- rules[quality(rules)$confidence > 0.90]
subrules_lift <- rules[quality(rules)$lift > 20]
subrules_support <- rules[quality(rules)$support > 0.01]
```

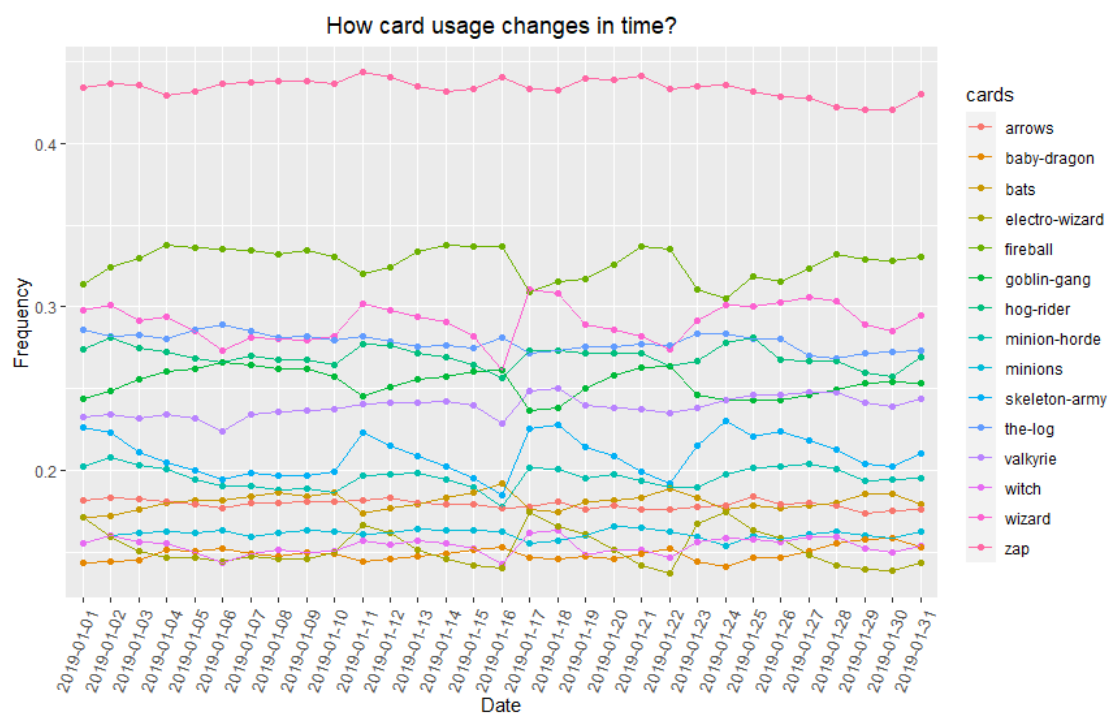
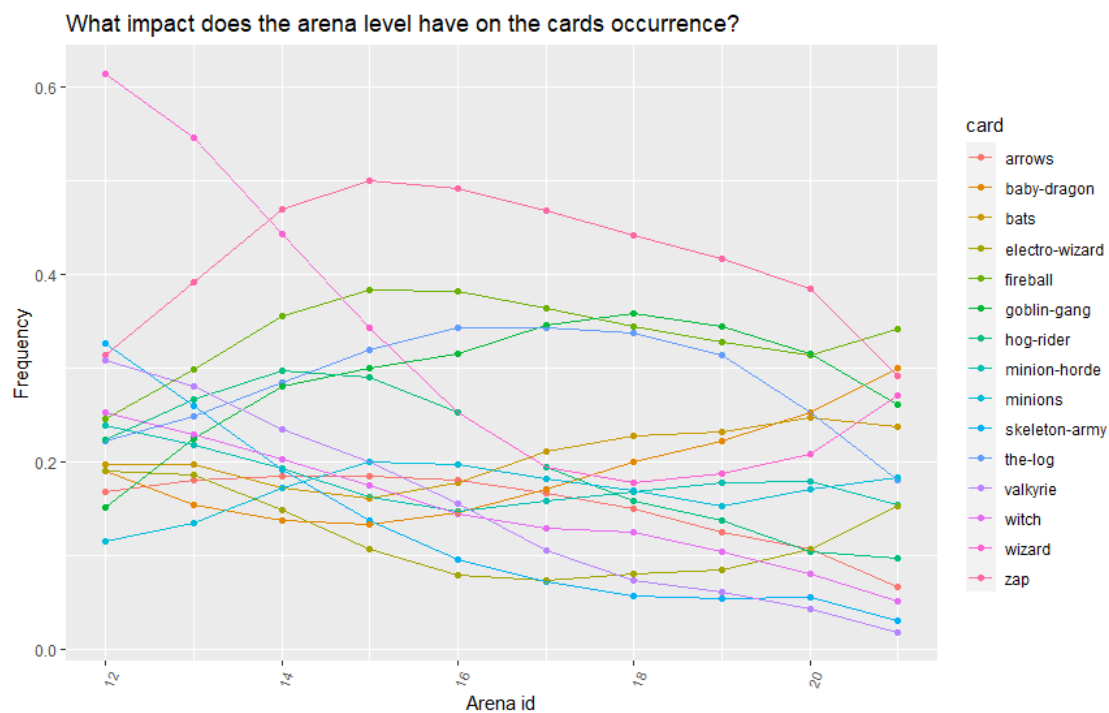


Figure 1: Top 15 most popular cards change



Zbiór danych jest ogromny, więc praca na wszystkich danych nie ma za bardzo sensu. Będziemy się zajmować danymi które mają największy support.

Wypiszmy 10 najpopularniejszych połączeń między kartami:

```
r = head(rules_support, 10)
as(r, 'data.frame')$rules
```

```
## [1] {zap} => {wizard}      {wizard} => {zap}      {wizard} => {the-log}    {witch} => {wizard}
## [5] {zap} => {the-log}      {witch} => {the-log}    {baby-dragon} => {arrows} {arrows} => {baby-
## [9] {arrows} => {bats}      {archers} => {arrows}
## 10 Levels: {archers} => {arrows} {arrows} => {baby-dragon} {arrows} => {bats} ... {zap} => {wizard}
```

Wszystkie z nich dotyczą połączeń po między dwoma kartami, sprawdźmy najpopularniejsze połączenia między większą liczbą kart.

```
more_than_one = rules[which(size(rules) > 2)]
r = head(more_than_one, 10)
as(r, 'data.frame')$rules
```

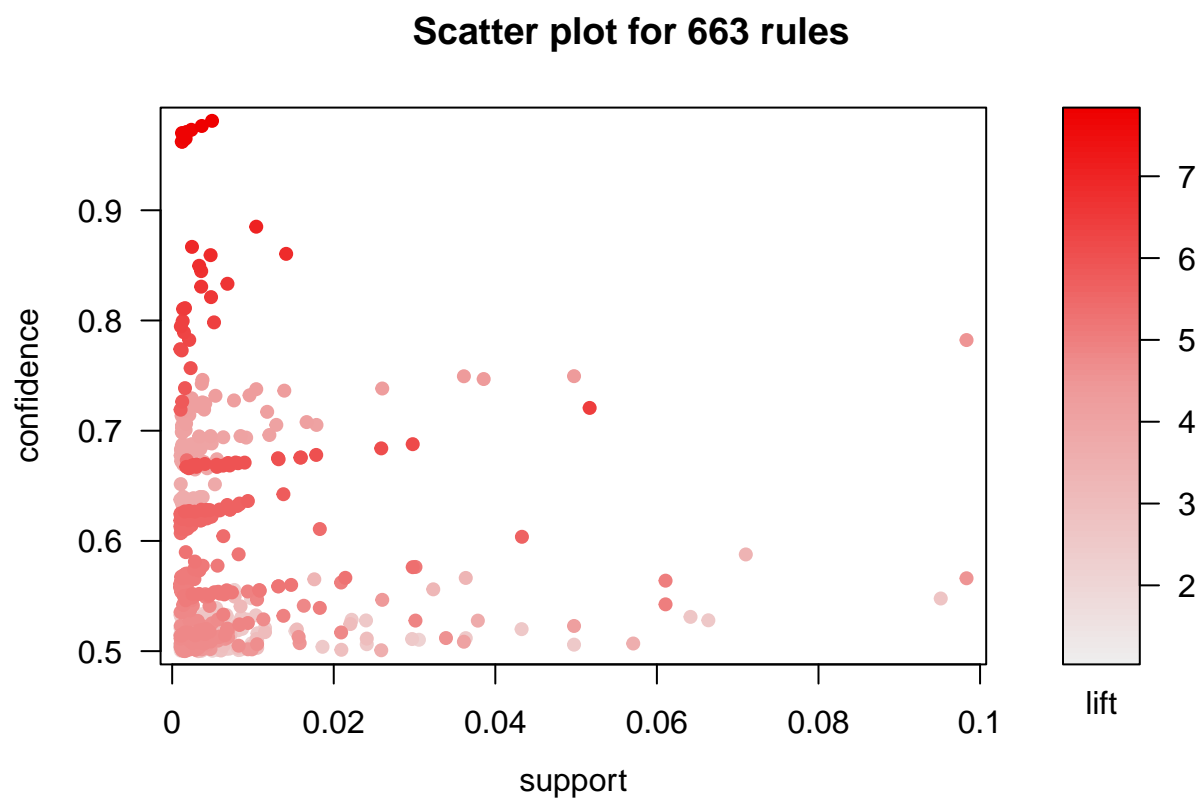
```
## [1] {tornado,zappies} => {zap}      {witch,zappies} => {zap}      {zap,zappies} => {wizard}
## [4] {wizard,zappies} => {zap}      {valkyrie,zappies} => {zap}    {the-log,zappies} => {zap}
## [7] {witch,zappies} => {wizard}    {valkyrie,zappies} => {wizard} {the-log,zappies} => {wizard}
## [10] {skeleton-barrel,zap} => {wizard}
## 10 Levels: {skeleton-barrel,zap} => {wizard} {the-log,zappies} => {wizard} ... {zap,zappies} => {wizard}
```

Również możemy zobaczyć karty które występują najczęściej razem.

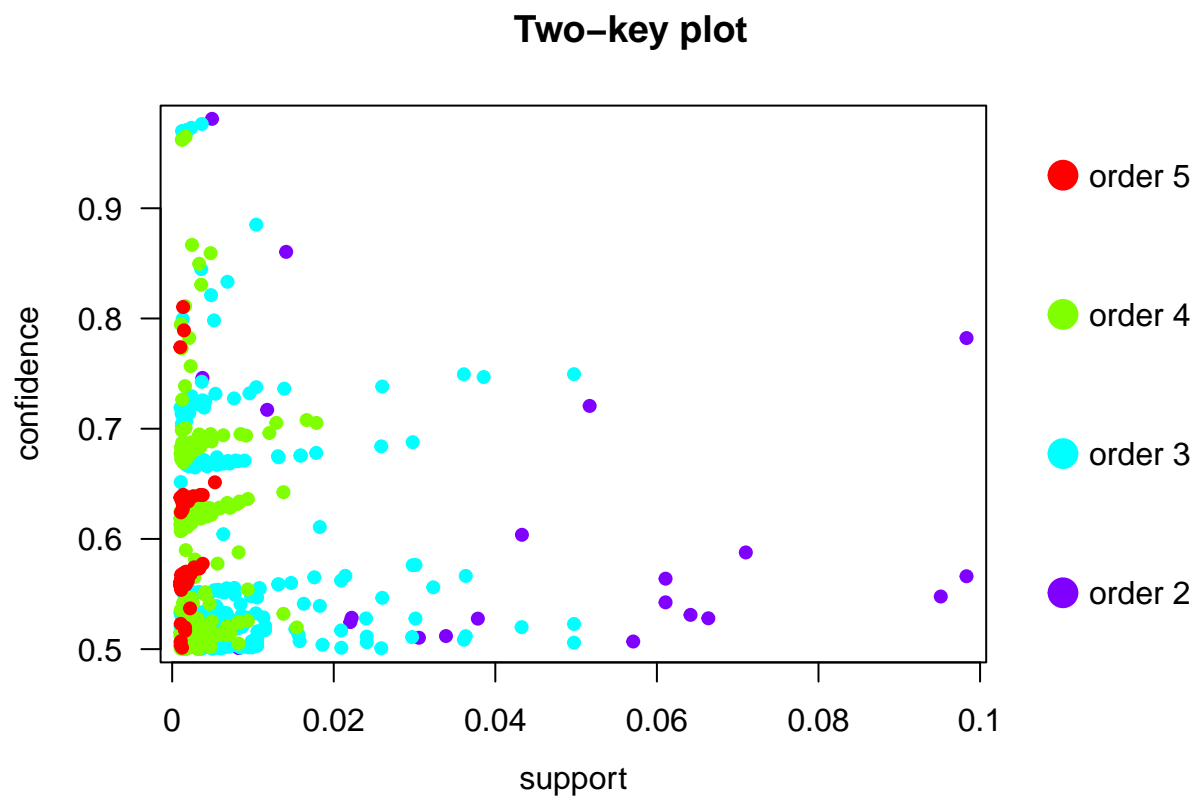
```
r = head(rules_lift, 10)
as(r, 'data.frame')$rules
```

```
## [1] {zappies} => {zap}      {wizard,zappies} => {zap}      {the-log,zappies} => {zap}
## [4] {valkyrie,zappies} => {zap} {witch,zappies} => {zap}      {tornado,zappies} => {zap}
## [7] {the-log,wizard,zappies} => {zap} {valkyrie,wizard,zappies} => {zap} {witch,wizard,zappies} => {zap}
## [10] {wizard,x-bow} => {zap}
## 10 Levels: {the-log,wizard,zappies} => {zap} {the-log,zappies} => {zap} ... {zappies} => {zap}
```

Takich ciekawych zależności jest bardzo dużo, zamiast wypisać wszystko możemy skorzystać z wykresów.



Wykres pokazuje że wszystkie zależności które mają duży lift i confidence występują rzadziej od innych.



Wykres dodatkowo za pomocą kolorów przedstawia liczbę pozycji zawartych w danej regule.

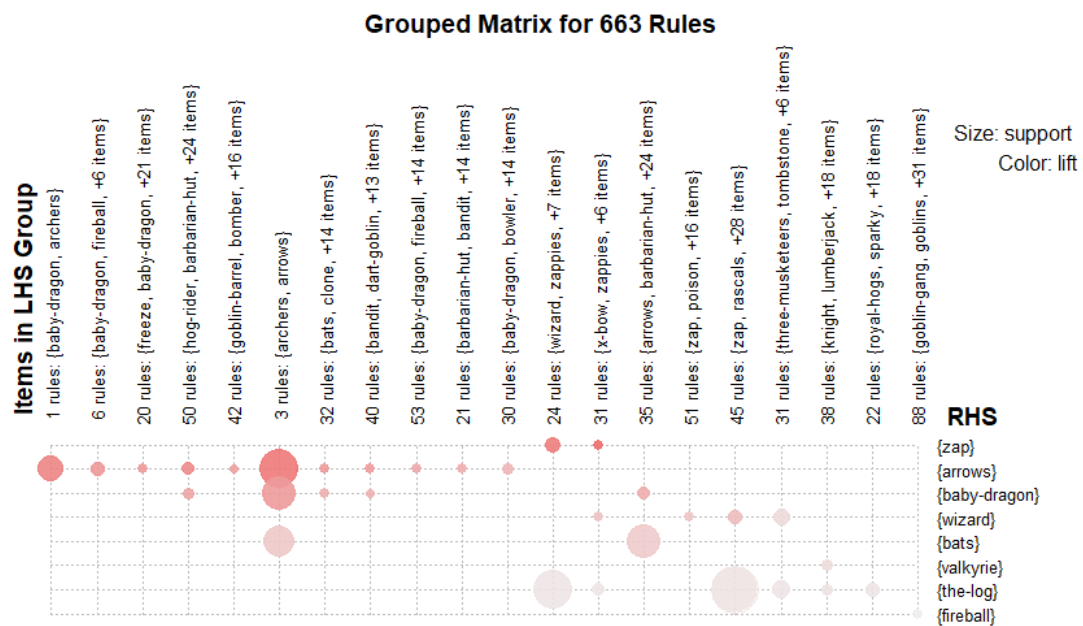


Figure 2: Grouped Matrix

Grupa najciekawszych reguł według ze względu na lift, LHS stanowią kolumny, a konsekwencje (RHS) wiersze.

Select by id ▼

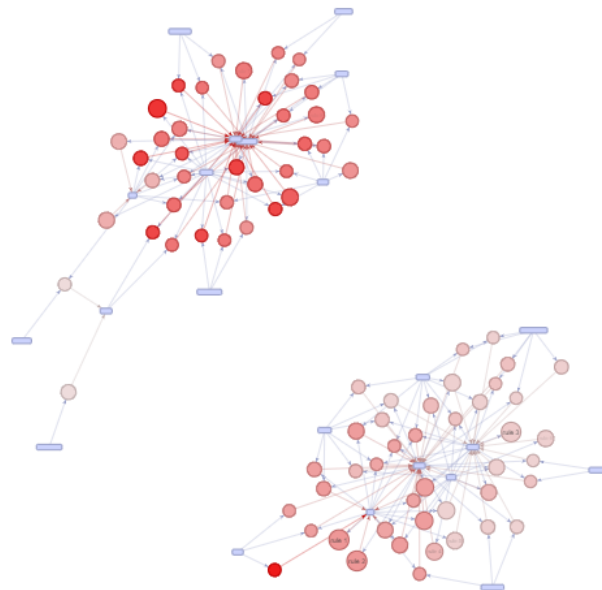


Figure 3: Graph

Figure 3 pokazuje wszystkie zależności w postaci ładnego grafu, graf jest interaktywny co ułatwia wyciągnięciu z niego odpowiednich reguł. Widzimy, że karty skupiają się przede wszystkim wokół dwóch głównych reguł.

Figure 4 przedstawia zależność po między LHS (wiersze), a RHS (kolumny), natomiast intensywność koloru stanowi lift.

Warto również zobaczyć ile unikalnych graczy zagrało w grę danego dnia.

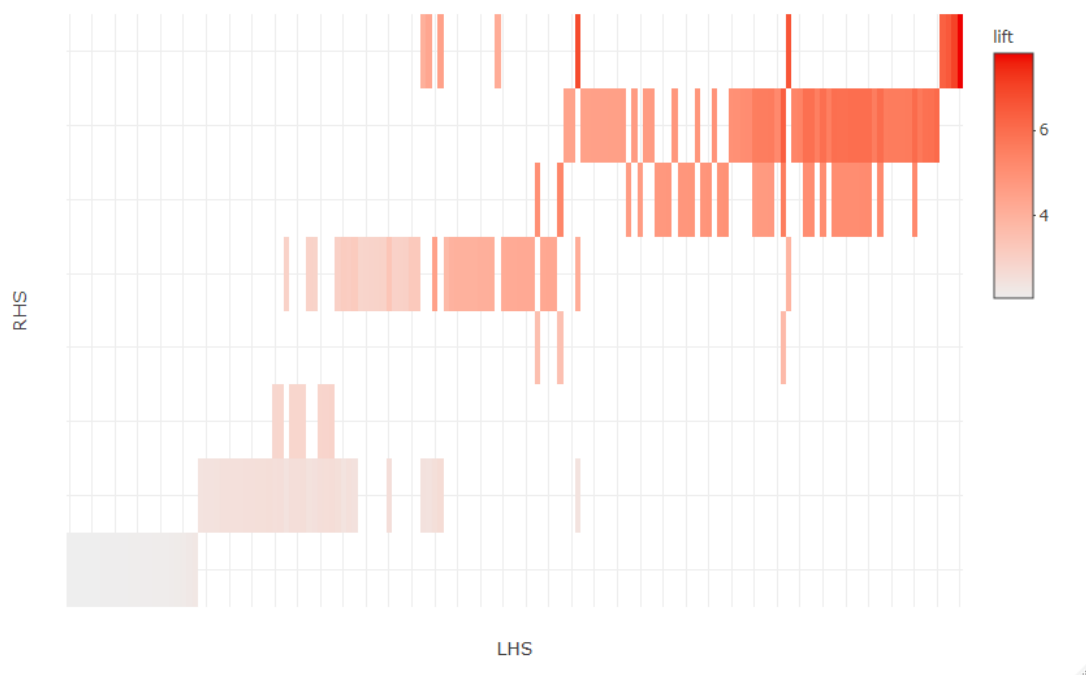
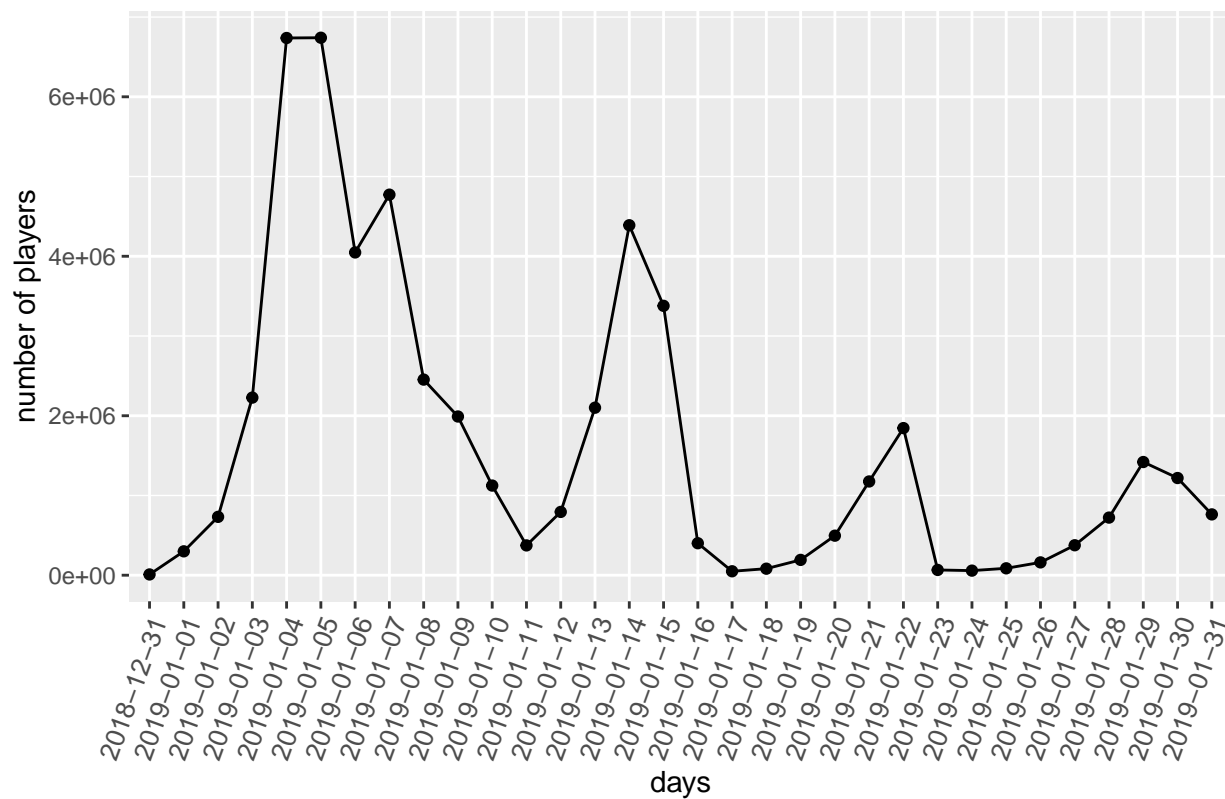


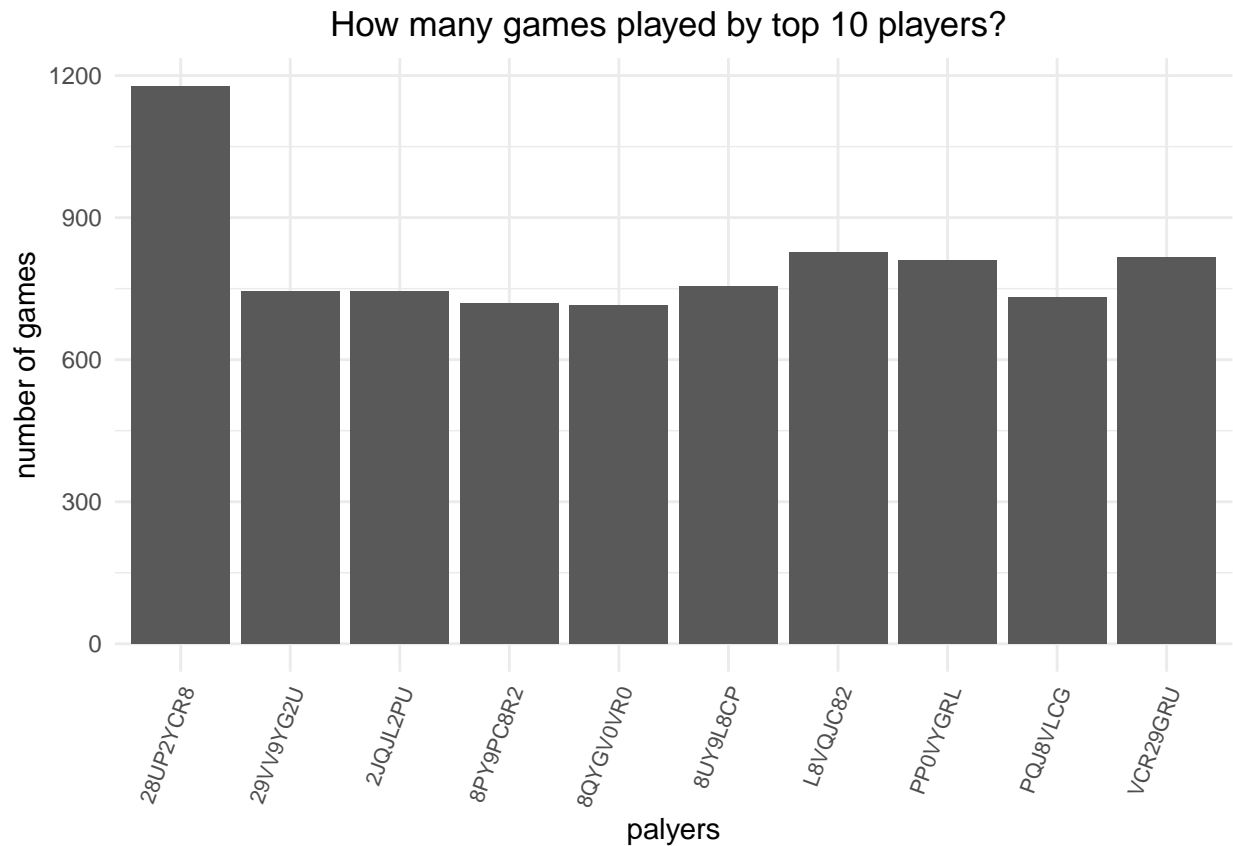
Figure 4: plot

How number of players changes in time?



Widzimy na początku roku dużo większe zainteresowanie grą. Popularność gry znacząco się różniła na

przestrzeni miesiąca. Piki w popularności gry wypadają na weekendy co jest bardzo naturalne.



Wykres przedstawia 10 graczy którzy zagraли najwięcej gier. Możemy sprawdzić ile gier zagrał rekordzista.

```
print(daily_players[1,2])
```

```
## [1] 1178
```

Zatem rekordzista zagrał 1178 gier w miesiącu! Co daje 38 gier dziennie!

Również możemy zobaczyć iloma deckami grał średnio jeden gracz.

```
## [1] 1.417698
```

Czyli średnio gracz grał około 1,5 deckami przez miesiąc. Oczywiście to tylko średnia, możemy wynioskować z tego że gracze nie zmieniali często decków.

Możemy również zobaczyć ile unikalnych graczy zagrało w Clash Royale w styczniu.

```
print(length(unic_players))
```

```
## [1] 5982309
```

Dzięki algorytmowi apiori możemy również znaleźć deck z największym win ratio. Do decków które wygrały dodałem kolejną “karte” o nazwie “win”. Jeśli po prawej stronie (RHS) znajduje się win to confidence będzie win ratio danego decku.

Warto też wypisać karty które najczęściej pojawiały się w deckach o największym win ratio.

```
r <- head(best_decks, 10)
as(r, 'data.frame')$rules
```

```
## [1] {cannon,fireball,hog-rider,ice-golem,ice-spirit,musketeer,skeletons,the-log} => {win}
## [2] {balloon,fireball,guards,lava-hound,mega-minion,minions,tombstone,zap} => {win}
## [3] {bandit,battle-ram,electro-wizard,minions,pekka,potion,royal-ghost,zap} => {win}
## [4] {battle-ram,elixir-collector,goblin-gang,ice-golem,miner,minion-horde,three-musketeers,zap} => {win}
## [5] {bats,battle-ram,elixir-collector,giant,goblin-gang,minion-horde,three-musketeers,zap} => {win}
## [6] {baby-dragon,barbarian-barrel,golem,lumberjack,mega-minion,night-witch,potion,tornado} => {win}
## [7] {goblin-barrel,goblin-gang,ice-spirit,knight,princess,rocket,tesla,the-log} => {win}
## [8] {goblin-barrel,goblin-gang,ice-spirit,inferno-tower,knight,princess,rocket,the-log} => {win}
## 8 Levels: {baby-dragon,barbarian-barrel,golem,lumberjack,mega-minion,night-witch,potion,tornado} => {win}
```

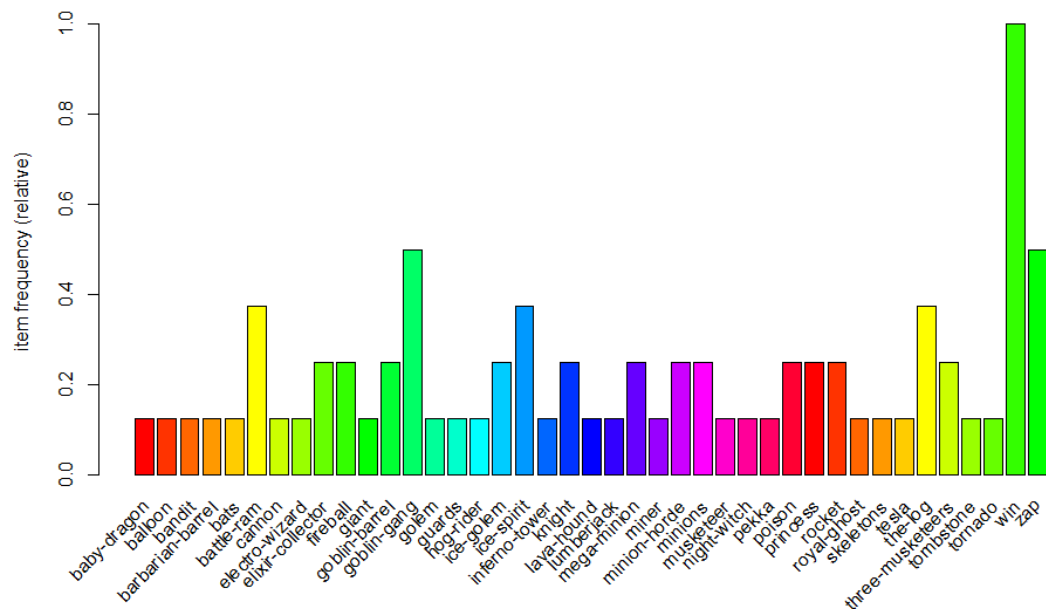


Figure 1: Cards frequency