EDA Scratch

This file is purely used to draft ideas and such. Ommitted graphs and code will be here. So if you are interested in some of the different ideas we went through but didn't include, here you go!

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

```
aoe2 <- read.csv("../Data/aoe2_leaderboard_sample.csv")
head(aoe2)
```

```
##   profile_id           name rank rating country games wins losses drops
## 1     199170          dogao    4   2232      BR   459  349    110    14
## 2     196240       TheViper    7   2207      DE   420  294    126     1
## 3     199419      gkt_cloud   30   2074      TW   361  221    140     3
## 4     312938           BacT   33   2060      VN   853  532    321     8
## 5    2431776    ItsOver9000   54   2009      AR    62   39     23     1
## 6     268565 RoR | Luca8761   67   1965      CA   619  320    299     7
##        game_type
## 1 1v1 Random Map
## 2 1v1 Random Map
## 3 1v1 Random Map
## 4 1v1 Random Map
## 5 1v1 Random Map
## 6 1v1 Random Map
```

```
str(aoe2)
```

```
## 'data.frame':    6539 obs. of  10 variables:
##  $ profile_id: int  199170 196240 199419 312938 2431776 268565 560474 225365 547152 1788064 ...
##  $ name      : chr  "dogao" "TheViper" "gkt_cloud" "BacT" ...
##  $ rank      : int  4 7 30 33 54 67 68 77 84 86 ...
##  $ rating    : int  2232 2207 2074 2060 2009 1965 1963 1953 1933 1932 ...
```

```
##  $ country   : chr  "BR" "DE" "TW" "VN" ...
##  $ games     : int  459 420 361 853 62 619 81 484 251 456 ...
##  $ wins      : int  349 294 221 532 39 320 58 275 150 254 ...
##  $ losses    : int  110 126 140 321 23 299 23 209 101 202 ...
##  $ drops     : int  14 1 3 8 1 7 2 6 1 1 ...
##  $ game_type : chr  "1v1 Random Map" "1v1 Random Map" "1v1 Random Map" "1v1 Random Map" ...
```

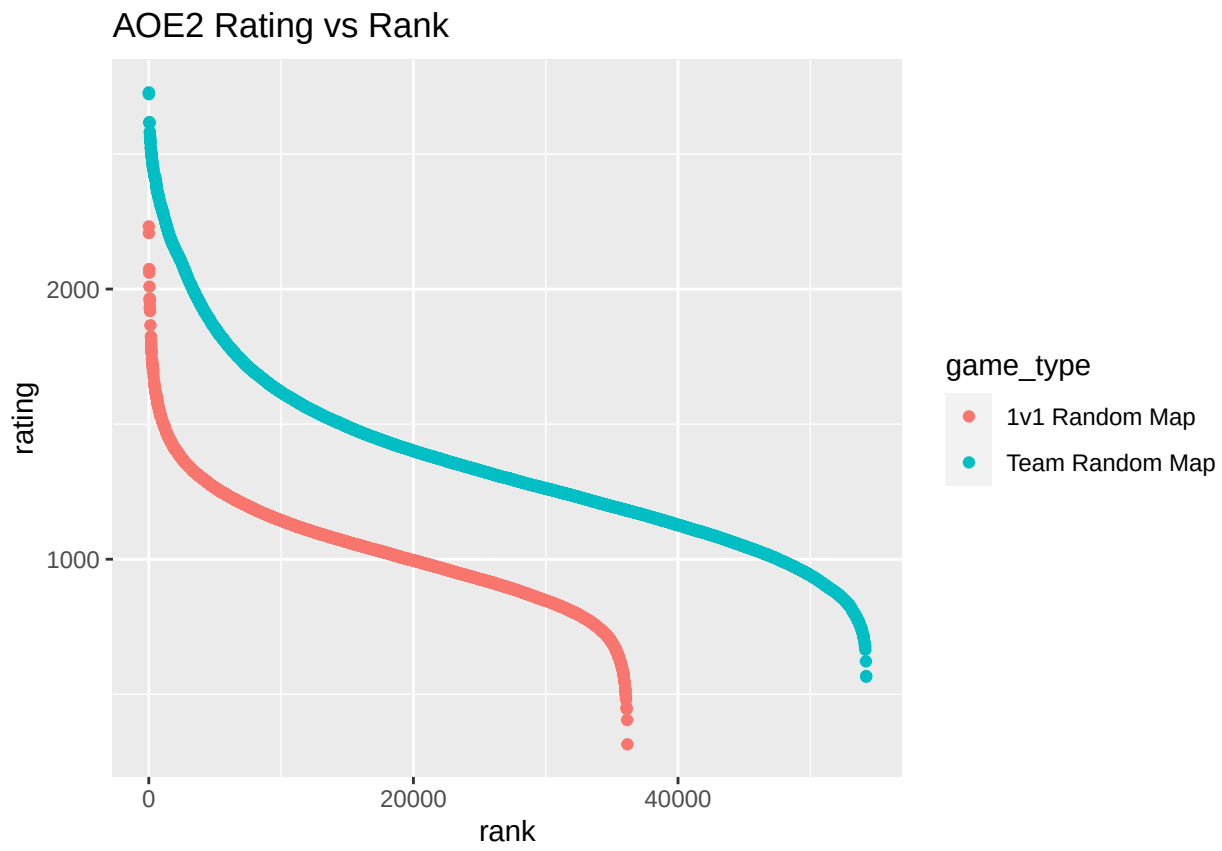
```
summary(aoe2)
```

```
##    profile_id          name               rank           rating     
##  Min.   :   1082   Length:6539        Min.   :    4   Min.   : 315  
##  1st Qu.: 549275   Class :character   1st Qu.:10040   1st Qu.:1002  
##  Median :1437310   Mode  :character   Median :21079   Median :1183  
##  Mean   :1375678                      Mean   :22203   Mean   :1243  
##  3rd Qu.:2103808                      3rd Qu.:32526   3rd Qu.:1409  
##  Max.   :2766539                      Max.   :54224   Max.   :2728  
##    country              games             wins            losses       
##  Length:6539        Min.   :  10.00   Min.   :  0.00   Min.   :  0.00  
##  Class :character   1st Qu.:  23.00   1st Qu.: 11.00   1st Qu.: 12.00  
##  Mode  :character   Median :  52.00   Median : 27.00   Median : 25.00  
##                     Mean   :  94.26   Mean   : 48.96   Mean   : 45.31  
##                     3rd Qu.: 117.50   3rd Qu.: 61.00   3rd Qu.: 57.00  
##                     Max.   :1757.00   Max.   :967.00   Max.   :851.00  
##      drops          game_type        
##  Min.   :  0.000   Length:6539        
##  1st Qu.:  0.000   Class :character   
##  Median :  1.000   Mode  :character   
##  Mean   :  2.881                      
##  3rd Qu.:  3.000                      
##  Max.   :127.000
```

```
head(aoe2[aoe2$game_type == 'Team Random Map',])
```

```
##      profile_id           name rank rating country games wins losses drops
## 2817    1195260         Kellar    8   2728      NO   364  319     45     2
## 2818     312774    Sun | Keno_    9   2725      MX   300  238     62     4
## 2819     560474        Sunzets   10   2721      MX   253  191     62     4
## 2820     199419      gkt_cloud   49   2618      TW   749  536    213     8
## 2821     332603 teutonic_tanks   54   2616      AT   280  237     43     9
## 2822    1892228            KaN   57   2614      AR   211  153     58     2
##            game_type
## 2817 Team Random Map
## 2818 Team Random Map
## 2819 Team Random Map
## 2820 Team Random Map
## 2821 Team Random Map
## 2822 Team Random Map
```

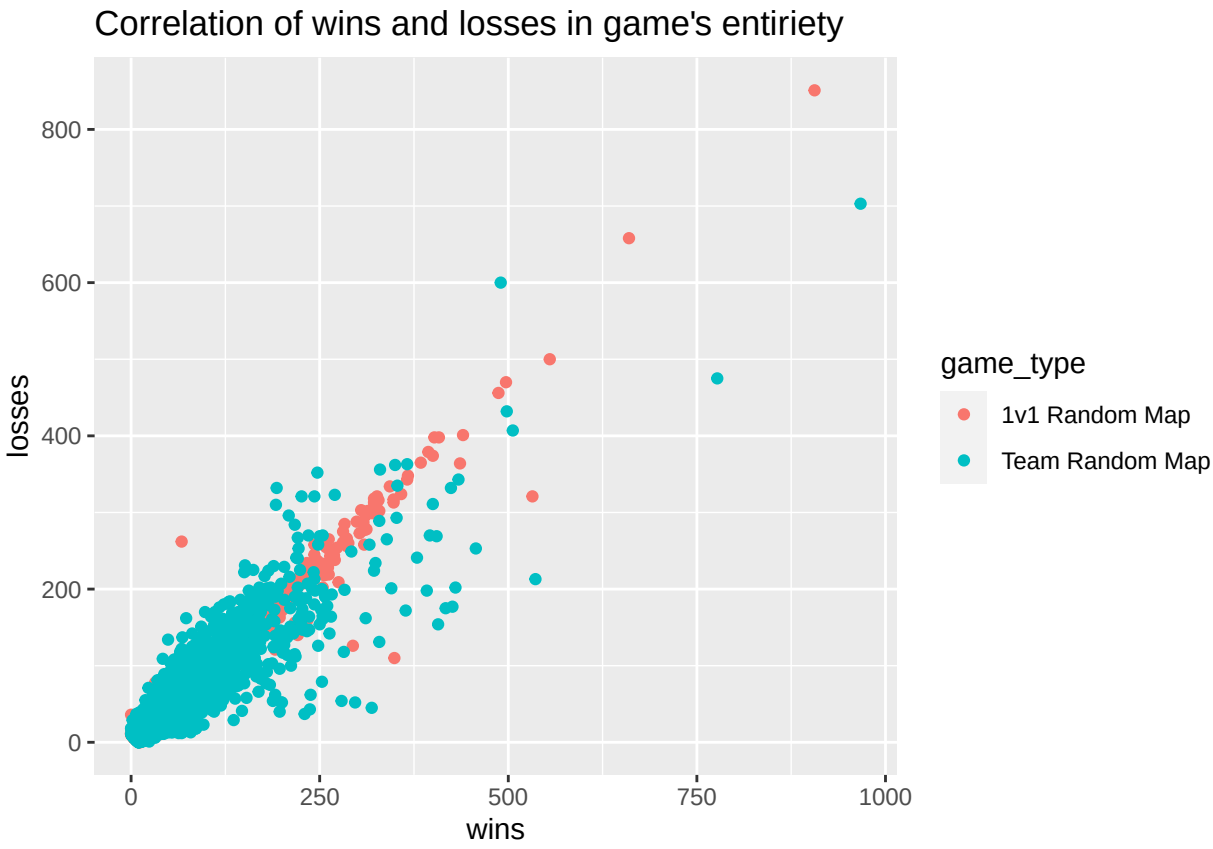
```
ggplot(aoe2, aes(x = rank, y = rating, color = game_type)) +
  geom_point() +
  ggtitle("AOE2 Rating vs Rank")
```



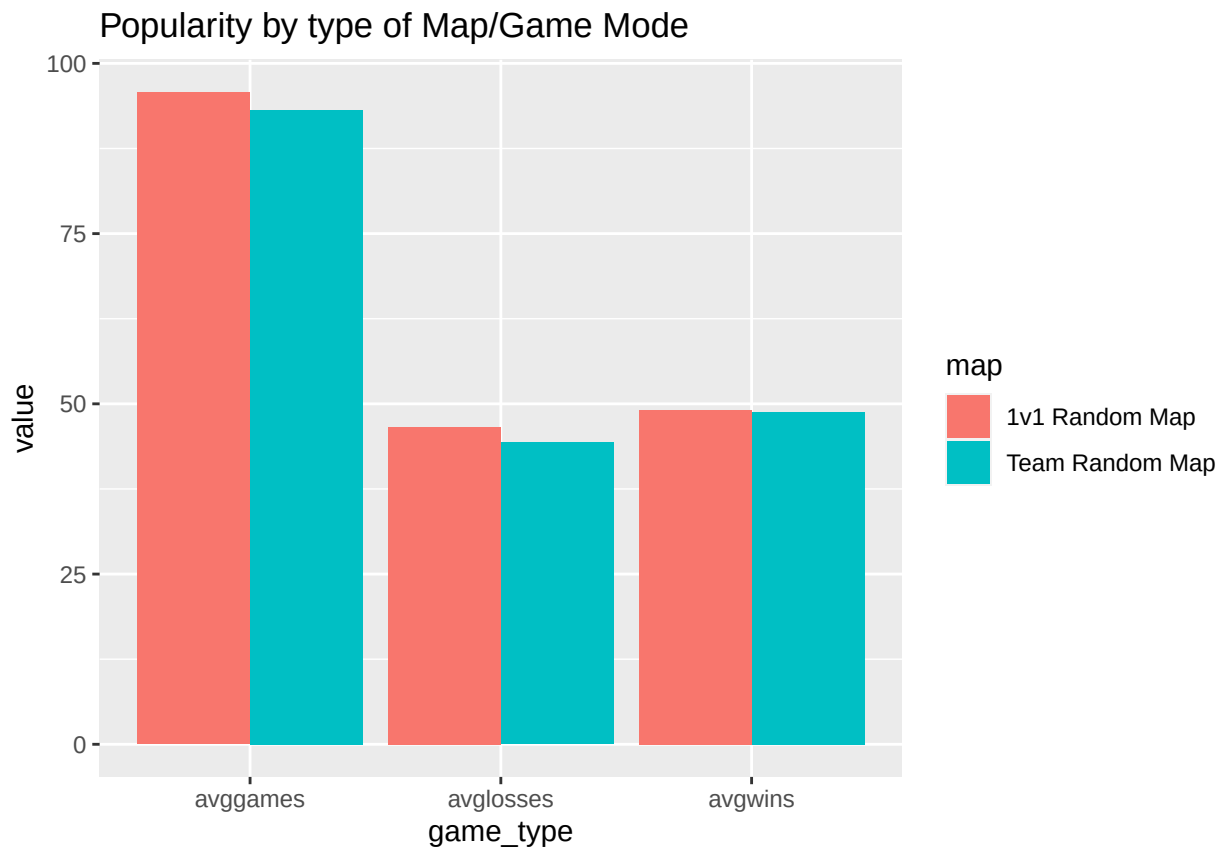

```
aoe2 %>%
  group_by(country) %>%
  summarise(avggames = mean(games), avgwins = mean(wins), avglosses = mean(losses), avgdrops = mean(dro
```

```
## # A tibble: 93 x 5
##    country avggames avgwins avglosses avgdrops
##    <chr>      <dbl>   <dbl>     <dbl>    <dbl>
##  1 AE          74.3    40.5      33.8     2
##  2 AL          34      22        12       0
##  3 AM          21       9        12       2
##  4 AR          69.9    36.5      33.4     2.29
##  5 AT          91.2    48.7      42.5     2.22
##  6 AU          77.2    38.6      38.7     1.97
##  7 BD          45      18        27       0
##  8 BE         112.     58.0      53.9     2.54
##  9 BG         141.     71.8      69       4.67
## 10 BH          99      43        56      10.3
## # ... with 83 more rows
```

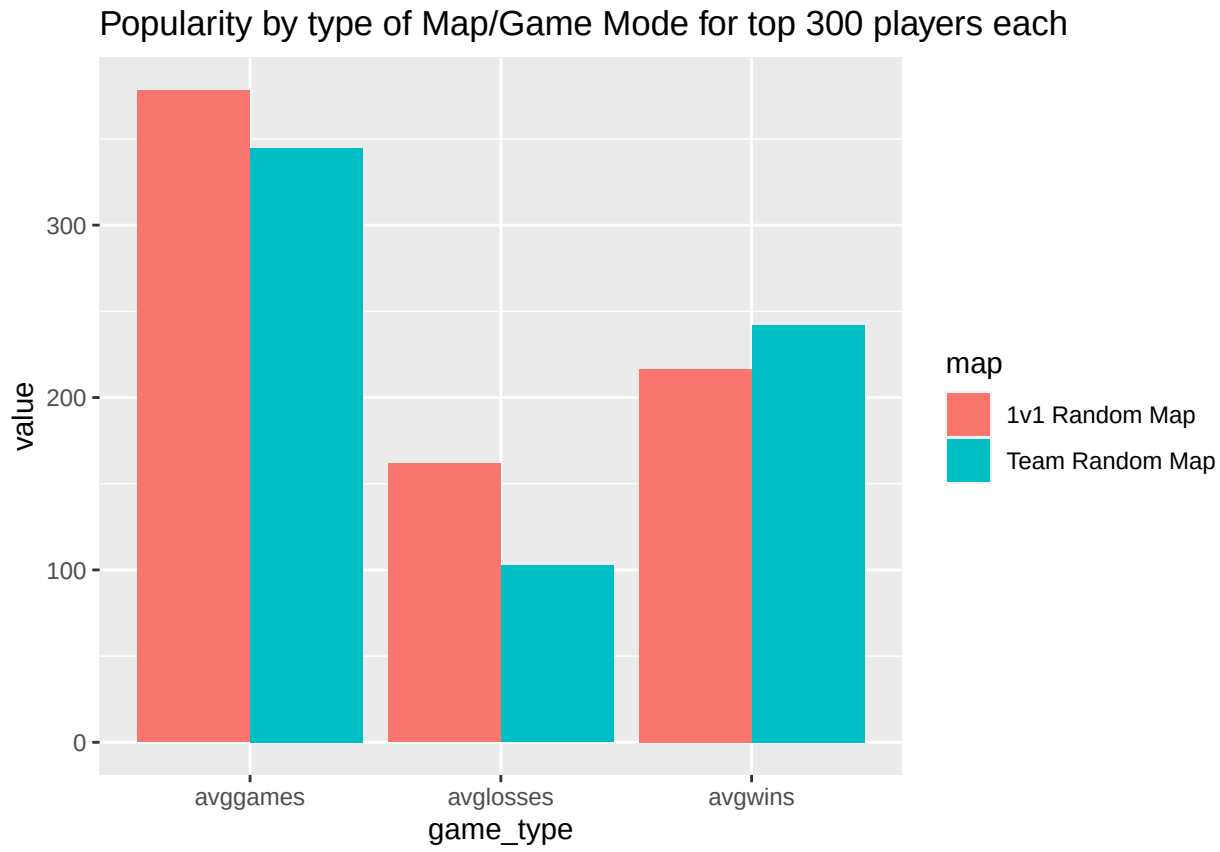
```
ggplot(aoe2) +
  geom_point(aes(wins, losses, color = game_type)) +
  ggtitle("Correlation of wins and losses in game's entiriety")
```




```
gametype <- aoe2 %>%
  group_by(game_type) %>%
  summarise(avggames = mean(games), avgwins = mean(wins), avglosses = mean(losses))

gametype_long <- gather(gametype, game_type)
gametype_long$map <- rep(c("1v1 Random Map", "Team Random Map"), 3)

ggplot(gametype_long, aes(fill = map, y = value, x = game_type)) +
  geom_bar(position="dodge", stat="identity") +
  ggtitle("Popularity by type of Map/Game Mode")
```



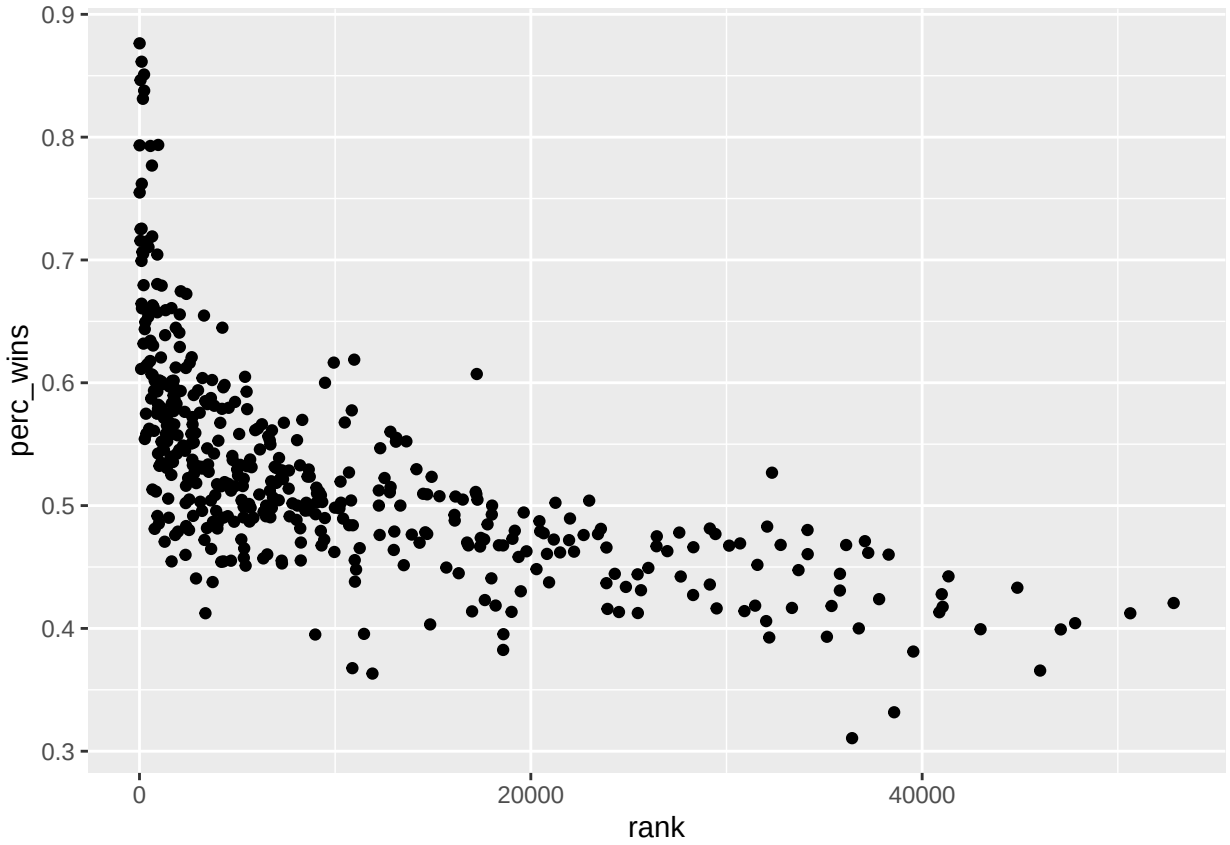

```
gametype <- aoe2[aoe2$rank <= 300, ] %>%
  group_by(game_type) %>%
  summarise(avggames = mean(games), avgwins = mean(wins), avglosses = mean(losses))

gametype_long <- gather(gametype, game_type)
gametype_long$map <- rep(c("1v1 Random Map", "Team Random Map"), 3)

ggplot(gametype_long, aes(fill = map, y = value, x = game_type)) +
  geom_bar(position="dodge", stat="identity") +
  ggtitle("Popularity by type of Map/Game Mode for top 300 players each")
```




```
aoe2$perc_wins  = aoe2$wins / aoe2$games


aoe2common <- subset(aoe2, games > 200)
head(arrange(aoe2common, desc(perc_wins)))
```

```
##   profile_id           name rank rating country games wins losses drops
## 1    1195260         Kellar    8   2728      NO   364  319     45     2
## 2     199170          dogao  110   2548      BR   267  230     37     8
## 3     268565 RoR | Luca8761  234   2487      CA   349  297     52     5
## 4     332603 teutonic_tanks   54   2616      AT   280  237     43     9
## 5     216736     JaySquared  240   2485      DE   333  279     54     1
## 6     216073        MaxiArg  175   2510      AR   237  197     40     5
##         game_type perc_wins
## 1 Team Random Map 0.8763736
## 2 Team Random Map 0.8614232
## 3 Team Random Map 0.8510029
## 4 Team Random Map 0.8464286
## 5 Team Random Map 0.8378378
## 6 Team Random Map 0.8312236
```

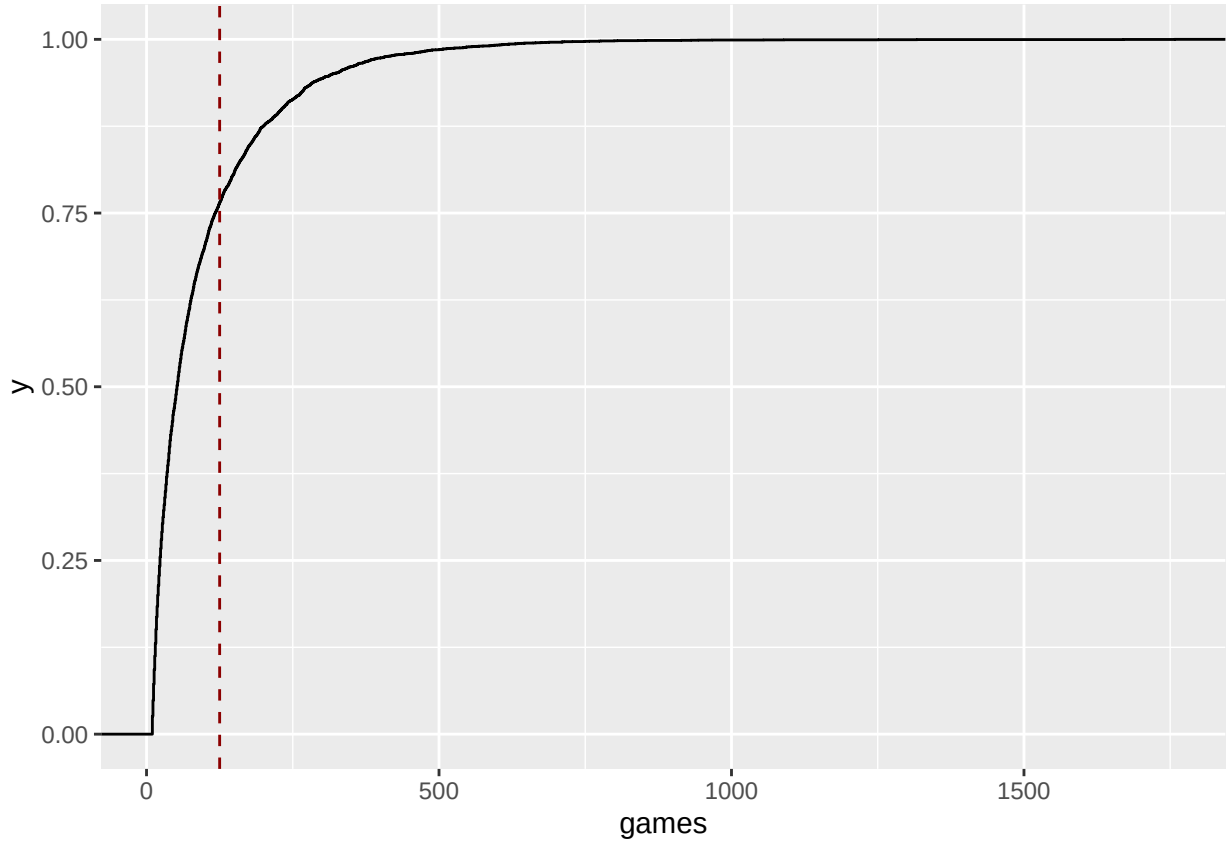
```
ggplot(aoe2common[aoe2common$game_type == "Team Random Map", ], aes(rank, perc_wins)) +
  geom_point()
```




```
head(aoe2common)
```

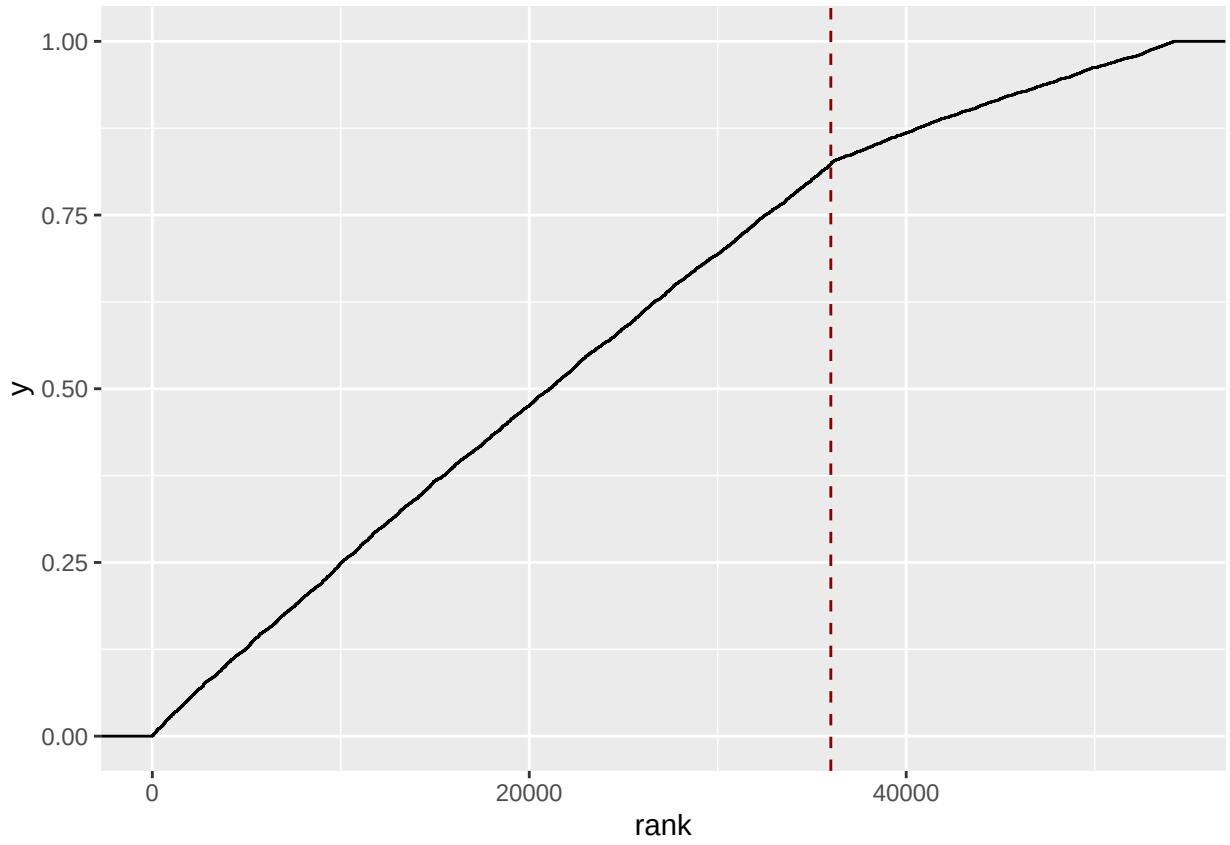
```
##   profile_id           name rank rating country games wins losses drops
## 1     199170          dogao    4   2232      BR   459  349    110    14
## 2     196240       TheViper    7   2207      DE   420  294    126     1
## 3     199419      gkt_cloud   30   2074      TW   361  221    140     3
## 4     312938           BacT   33   2060      VN   853  532    321     8
## 6     268565 RoR | Luca8761   67   1965      CA   619  320    299     7
## 8     225365        Obadiah   77   1953      CA   484  275    209     6
##        game_type perc_wins
## 1 1v1 Random Map 0.7603486
## 2 1v1 Random Map 0.7000000
## 3 1v1 Random Map 0.6121884
## 4 1v1 Random Map 0.6236811
## 6 1v1 Random Map 0.5169628
## 8 1v1 Random Map 0.5681818
```

```
aoe2 %>%
  ggplot(aes(x = games)) +
  stat_ecdf() +
  geom_vline(xintercept = 125,
             linetype = "dashed",
             color = "darkred")
```

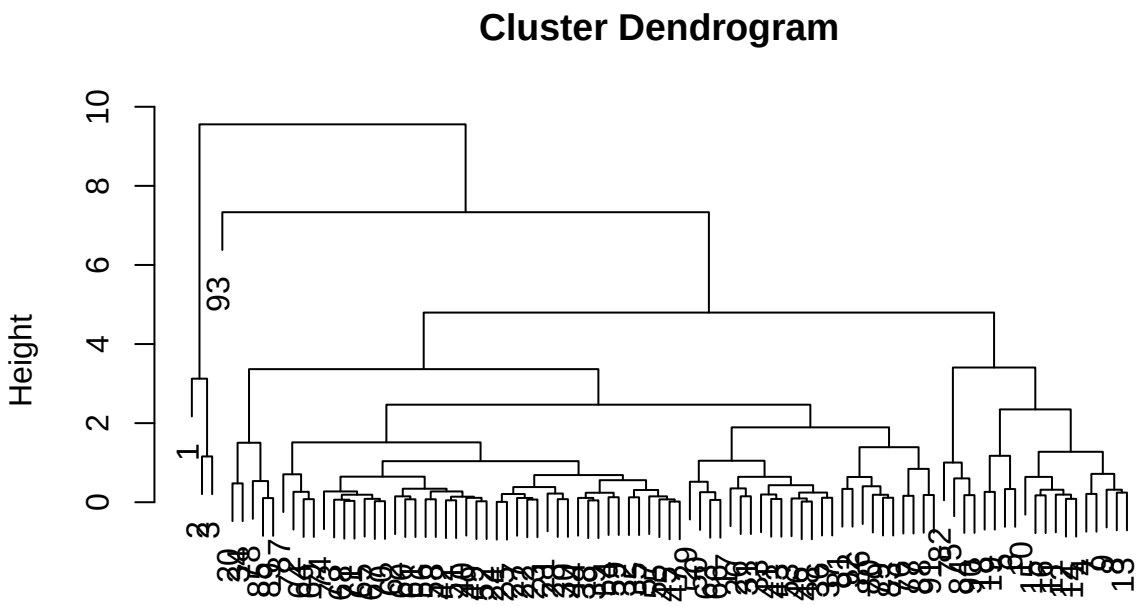



```
aoe2 %>%
  ggplot(aes(x = rank)) +
  stat_ecdf() +
  geom_vline(xintercept = 36000,
             linetype = "dashed",
             color = "darkred",
             stat.identity = TRUE)
```

```
## Warning: Ignoring unknown parameters: stat.identity
```

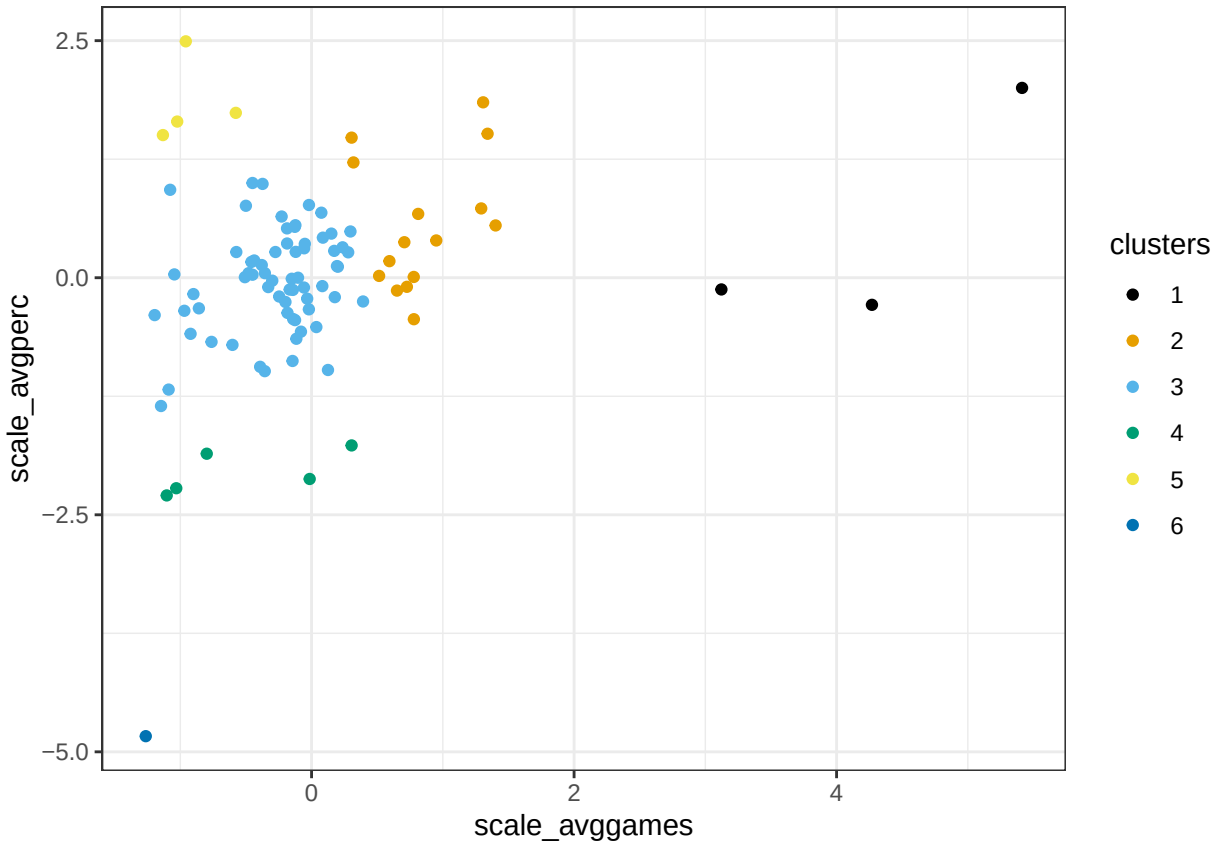



```
aoe2_avgcountry <- aoe2 %>%
  group_by(country) %>%
  summarise(avggames = mean(games), avgwins = mean(wins), avglosses = mean(losses), avgdrops = mean(dro
  arrange(desc(avggames))

aoe2_avgcountry$scale_avggames <- scale(aoe2_avgcountry$avggames)
aoe2_avgcountry$scale_avgperc <- scale(aoe2_avgcountry$avgperc)



aoe2_hclust <-
  hclust(dist(
    dplyr::select(aoe2_avgcountry,
                  scale_avggames, scale_avgperc)),
         method = "complete")
plot(aoe2_hclust)
```

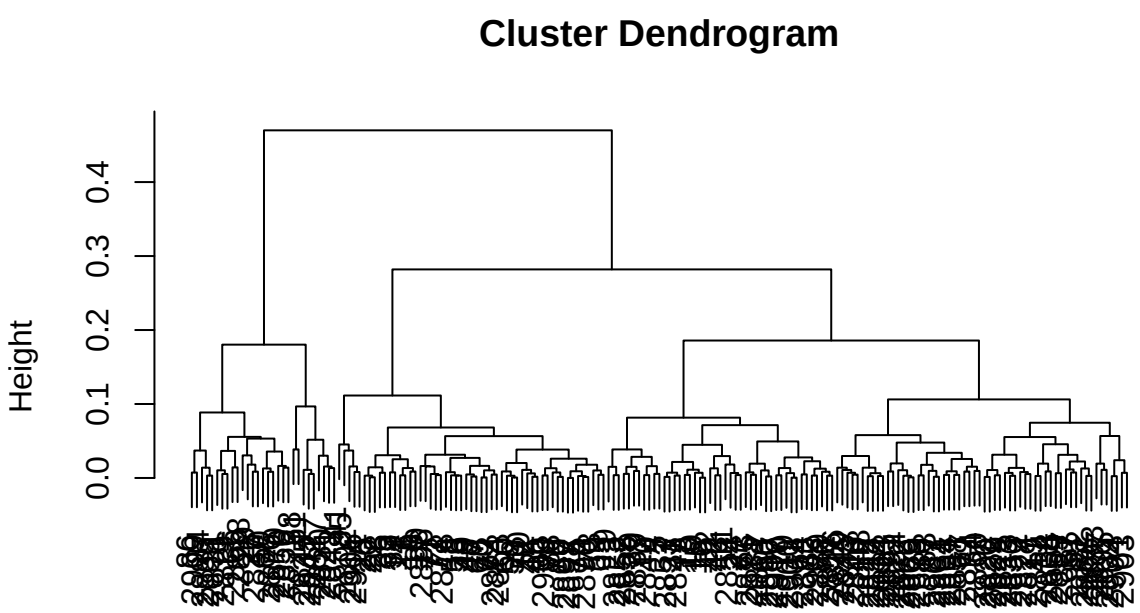


dist(dplyr::select(aoe2_avgcountry, scale_avggames, scale_avgperc))
hclust (*, "complete")

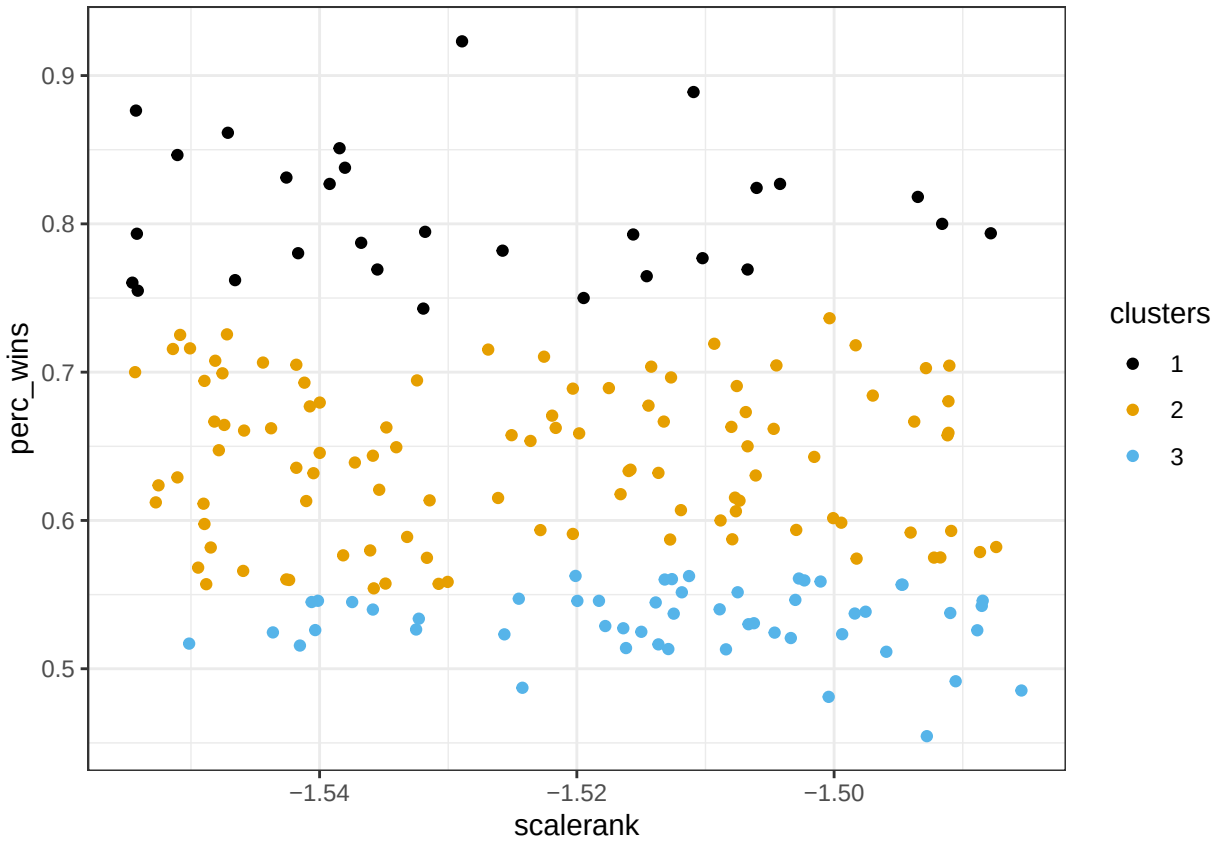
```
aoe2_player_clusters <-
  cutree(aoe2_hclust,
         k = 6)


aoe2_avgcountry <- aoe2_avgcountry %>%
  mutate(clusters =
           as.factor(aoe2_player_clusters))


aoe2_avgcountry %>%
  ggplot(aes(x = scale_avggames, y = scale_avgperc,
             color = clusters)) +
  geom_point() +
  ggthemes::scale_color_colorblind() +
  theme_bw()
```




```
aoe2$scalerank <- scale(aoe2$rank)

filtered_aoe2 <- aoe2[aoe2$rank < 1000, ]

aoe2_hclust <-
  hclust(dist(
    dplyr::select(filtered_aoe2,
                  scalerank, perc_wins)),
         method = "complete")
plot(aoe2_hclust)
```




```
aoe2_player_clusters <-
  cutree(aoe2_hclust,
         k = 3)


filtered_aoe2 <- filtered_aoe2 %>%
  mutate(clusters =
           as.factor(aoe2_player_clusters))


filtered_aoe2 %>%
  ggplot(aes(x = scalerank, y = perc_wins,
             color = clusters)) +
  geom_point() +
  ggthemes::scale_color_colorblind() +
  theme_bw()
```


0.9
0.8
0.7
0.6
0.5
perc_wins
−1.54
−1.52
−1.50
scalerank
clusters
1
2
3