
DOTA2
SQUARE 2011 SQUAR3C.DEVIANTART.COM

# Top-level Data

- match.csv:

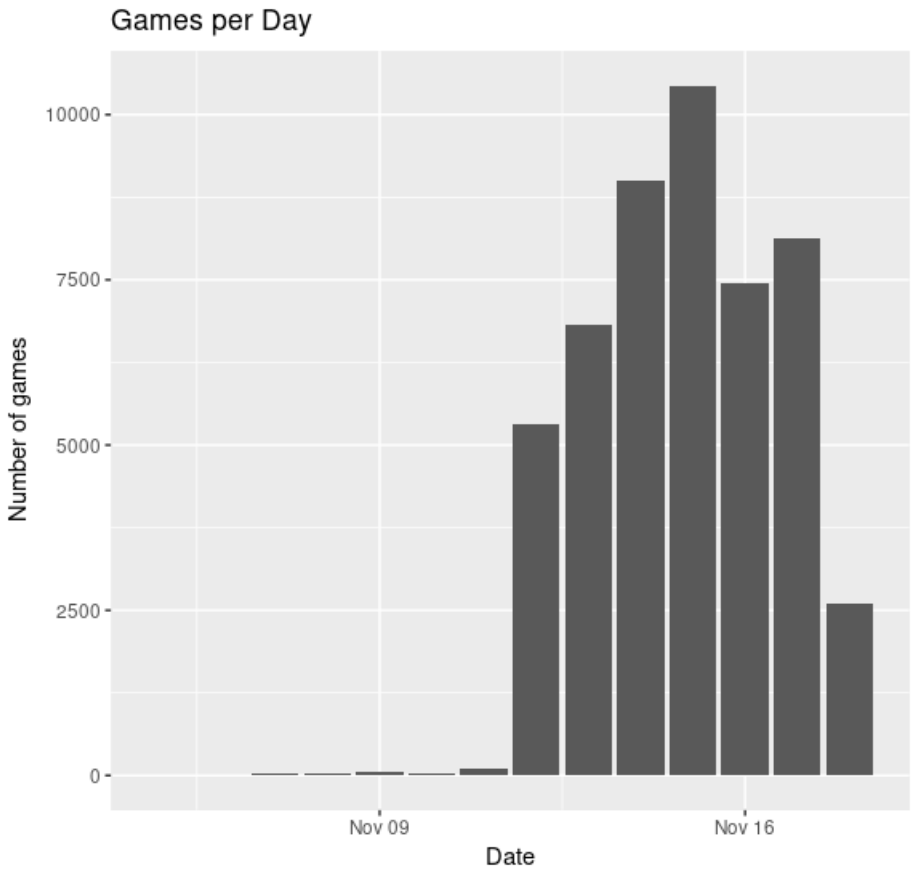
```
> match %>% str
Classes ‘data.table’ and 'data.frame':  50000 obs. of  13 variables:
 $ match_id               : int  0 1 2 3 4 5 6 7 8 9 ...
 $ start_time             : int  1446750112 1446753078 1446764586 1446765723 1446796385 1446798766 1446800938 14468
04030 1446819063 1446837251 ...
 $ duration               : int  2375 2582 2716 3085 1887 1574 2124 2328 2002 2961 ...
 $ tower_status_radiant   : int  1982 0 256 4 2047 2047 1972 2046 0 0 ...
 $ tower_status_dire      : int  4 1846 1972 1924 0 4 0 0 1982 1972 ...
 $ barracks_status_dire   : int  3 63 63 51 0 3 3 0 63 63 ...
 $ barracks_status_radiant: int  63 0 48 3 63 63 63 63 0 0 ...
 $ first_blood_time       : int  1 221 190 40 58 113 4 255 4 85 ...
 $ game_mode              : int  22 22 22 22 22 22 22 22 22 22 ...
 $ radiant_win            : logi  TRUE FALSE FALSE FALSE TRUE TRUE ...
 $ negative_votes         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ positive_votes         : int  1 2 0 0 0 0 0 0 0 0 ...
 $ cluster                : int  155 154 132 191 156 155 151 138 182 133 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

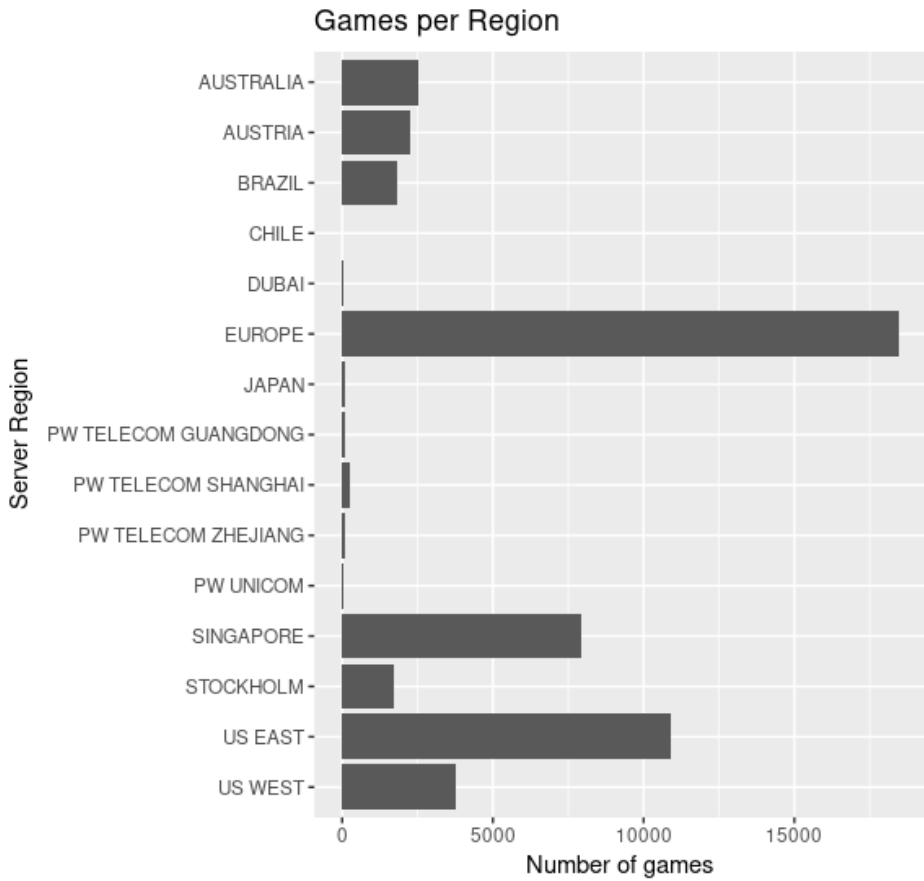
# Top-level Data

- players.csv:
  - Lots of detailed data, but only a summary

```
> players %>% str
Classes ‘data.table’ and 'data.frame':  500000 obs. of  73 variables:
 $ match_id                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ account_id                        : int  0 1 0 2 3 4 0 5 0 6 ...
 $ hero_id                           : int  86 51 83 11 67 106 102 46 7 73 ...
 $ player_slot                       : int  0 1 2 3 4 128 129 130 131 132 ...
 $ gold                              : int  3261 2954 110 1179 3307 476 317 2390 475 60 ...
 $ gold_spent                        : int  10960 17760 12195 22505 23825 12285 10355 13395 5035 17550 ...
 $ gold_per_min                      : int  347 494 350 599 613 397 303 452 189 496 ...
 $ xp_per_min                        : int  362 659 385 605 762 524 369 517 223 456 ...
 $ kills                             : int  9 13 0 8 20 5 4 4 1 1 ...
 $ deaths                            : int  3 3 4 4 3 6 13 8 14 11 ...
 $ assists                           : int  18 18 15 19 17 8 5 6 8 6 ...
 $ denies                            : int  1 9 1 6 13 5 2 31 0 0 ...
 $ last_hits                         : int  30 109 58 271 245 162 107 208 27 147 ...
 $ stuns                             : chr  "76.7356" "87.4164" "None" "None" ...
 $ hero_damage                       : int  8690 23747 4217 14832 33740 10725 15028 10230 4774 6398 ...
 $ hero_healing                      : int  218 0 1595 2714 243 0 764 0 0 292 ...
 $ tower_damage                      : int  143 423 399 6055 1833 112 0 2438 0 0 ...
 $ item_0                            : int  180 46 48 63 114 145 50 41 36 63 ...
 $ item_1                            : int  37 63 60 147 92 73 11 63 0 9 ...
 $ item_2                            : int  73 119 59 154 147 149 102 36 0 116 ...
 $ item_3                            : int  56 102 108 164 0 48 36 147 46 65 ...
 $ item_4                            : int  108 24 65 79 137 212 185 168 0 229 ...
 $ item_5                            : int  0 108 0 160 63 0 81 21 180 79 ...
 $ level                             : int  16 22 17 21 24 19 16 19 12 18 ...
 $ leaver_status                     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ xp_hero                           : num  8840 14331 6692 8583 15814 ...
 $ xp_creep                          : num  5440 8440 8112 14230 14325 ...
 $ xp_roshan                         : num  NA 2683 NA 894 NA ...
 $ xp_other                          : num  83 671 453 293 62 1 1 244 27 933 ...
 $ gold_other                        : num  50 395 259 100 NA ...
 $ gold_death                        : num  -957 -1137 -1436 -2156 -1437 ...
 $ gold_buyback                      : num  NA NA -1015 NA -1056 ...
 $ gold_abandon                      : num  NA NA NA NA NA NA NA NA NA NA ...
 $ gold_sell                         : num  212 1650 NA 938 4194 ...
 $ gold_destroying_structure         : num  3120 3299 3142 4714 3217 ...
 $ gold_killing_heros                : num  5145 6676 2418 4104 7467 ...
 $ gold_killing_creeps               : num  1087 4317 3697 10432 9220 ...
 $ gold_killing_roshan               : num  400 937 400 400 400 NA NA NA NA NA ...
 $ gold_killing_couriers             : num  NA NA NA NA NA NA NA NA NA NA ...
 $ unit_order_none                   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ unit_order_move_to_position       : num  4070 5894 7053 4712 3853 ...
 $ unit_order_move_to_target         : num  1 214 3 133 7 166 63 11 55 2 ...
 $ unit_order_attack_move            : num  25 165 132 163 7 76 100 214 5 105 ...
 $ unit_order_attack_target          : num  416 1031 645 690 1173 ...
 $ unit_order_cast_position          : num  51 98 36 9 31 196 13 122 68 64 ...
 $ unit_order_cast_target            : num  144 39 160 15 84 3 173 NA 18 102 ...
 ...
```

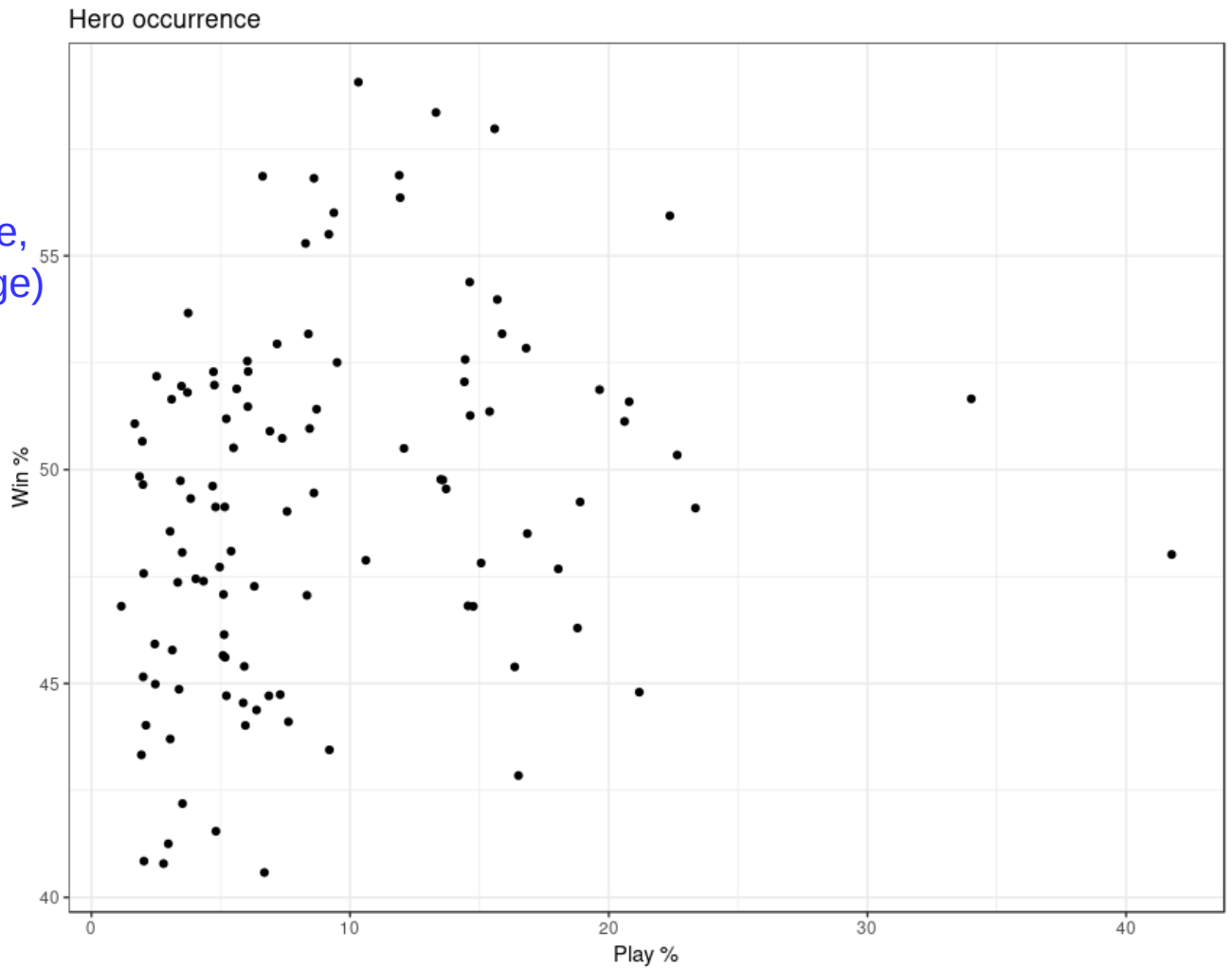
# Summaries


Games per Day
10000
7500
5000
2500
0
Number of games
Nov 09
Nov 16
Date


Games per Region
AUSTRALIA
AUSTRIA
BRAZIL
CHILE
DUBAI
EUROPE
JAPAN
PW TELECOM GUANGDONG
PW TELECOM SHANGHAI
PW TELECOM ZHEJIANG
PW UNICOM
SINGAPORE
STOCKHOLM
US EAST
US WEST
Server Region
0
5000
10000
15000
Number of games

# Summaries

```
> cor(
  heroes$win_percentage,
  heroes$play_percentage)
0.25
```



# In-game Data

- During the match
  - player_time.csv:
    Gold & Experience for each player at each minute
  - objectives.csv:
    Objective type, exact game time, involved players, ...
  - teamfights.csv & teamfights_players.csv:
    Starttime and endtime of teamfights, summary, involved players, result, ...
  - purchase_log.csv:
    Item buys for each player with exact time
  - ability_upgrades.csv
  - chat.csv

=> Very different data types with different formatting
=> Not equispaced in time

# Looking at a single game

```
> id <- sample(match$match_id, 1)
> match[match_id==id] %>% dim
[1]  1 13
> players[match_id==id] %>% dim
[1] 10 73
> player_time[match_id==id] %>% dim
[1] 56 32
> objectives[match_id==id] %>% dim
[1] 28  9
> ability_upgrades[match_id==id] %>% dim
[1] 207   5
> purchase_log[match_id==id] %>% dim
[1] 434   4
> teamfights[match_id==id] %>% dim
[1] 13  5
> teamfights_players[match_id==id] %>% dim
[1] 130   8
> chat[match_id==id] %>% dim
[1] 35  5
```

- Total of 914 rows
- ~150 different variables

# Ideas & Challenges

- General win-prediction:
  - Use player history and team composition to predict wins
- On-line win-prediction:
  - Use all available in-game data to predict
  - Update prediction each time new information is available
  - Challenge:
    - Very different types of information
    - Many variables
    - Information is not equispaced
- On-line prediction for events in general:
  - Similar approach as above, while we do not limit us to the binary outcome „win“