Process with Parser.

Games:

GA522 Game 1

GA523 Game 2

GA525 Game 3a

GA526 Game 4

(GA524 "Game 3" does not appear in Client logs)

gamep <- read.table("game.csv", header = TRUE, sep=",")

#day hour event clientId playerId gameId year

> summary(gamep)

day hour event clientId

Min. :14810 Min. : 6.24 gamep:6224 Min. :3.550e+14

1st Qu.:14811 1st Qu.:12.21 1st Qu.:3.550e+14

Median :14813 Median :13.95 Median :3.550e+14

Mean :14813 Mean :13.81 Mean :3.550e+14

3rd Qu.:14814 3rd Qu.:15.16 3rd Qu.:3.550e+14

Max. :14814 Max. :21.13 Max. :3.550e+14

playerId gameId year

P541 : 237 GA516 : 953 Min. :1900

P544 : 171 GA523 : 794 1st Qu.:1924

P543 : 164 GA525 : 688 Median :1949

P538 : 158 GA514 : 626 Mean :1952

P537 : 136 GA526 : 391 3rd Qu.:1978

P529 : 132 GA510 : 368 Max. :2020

(Other):5226 (Other):2404 NA's : 79

> inc <- (gamep$gameId=='GA522' | gamep$gameId=='GA523' | gamep$gameId=='GA525' | gamep$gameId=='GA526')

> gameId <- factor(gamep$gameId[inc])

> playerId <- factor(gamep$playerId[inc])

> hour <- gamep$hour[inc]

> year <- gamep$year[inc]

> plot(hour,year)



For each player, min year, max year, num diff year.

> minyear <- tapply(year, playerId, min)

> minyear

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

NA NA NA NA 1961 NA NA NA NA NA NA 1900 NA NA NA NA

P584 P585 P586 P587 P588 P589 P590 P591 P592

NA NA NA 1901 NA NA NA NA NA

> maxyear <- tapply(year[!is.na(year)], playerId[!is.na(year)], max)

> maxyear

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

1944 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011

P584 P585 P586 P587 P588 P589 P590 P591 P592

2011 2011 2011 2011 2011 2011 1972 1928 2011

> minyear2 <- tapply(year[!is.na(year)], playerId[!is.na(year)], min)

> minyear2

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

1900 1900 1900 1901 1961 1900 1900 1900 1900 1900 1900 1900 1900 1900 1900 1900

P584 P585 P586 P587 P588 P589 P590 P591 P592

1900 1900 1900 1901 1900 1900 1900 1900 1900

> years <- tapply(seq(1,length(year))[!is.na(year)],playerId[!is.na(year)],length)

> years

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

11 113 32 59 31 112 98 87 97 112 99 92 90 112 89 100

P584 P585 P586 P587 P588 P589 P590 P591 P592

97 99 89 95 89 105 67 27 98

> summary(years)

Min. 1st Qu. Median Mean 3rd Qu. Max.

11 87 95 84 99 113

P569 (uniquely) receives 2011 twice.

> summary(years / 1.12)

Min. 1st Qu. Median Mean 3rd Qu. Max.

9.821 77.680 84.820 75.000 88.390 100.900

hist(years,breaks=seq(0,115,by=5))



> hist(years/1.12,breaks=seq(0,105,by=5))



> actions <- read.table("action.csv", header = TRUE, sep=",")

> summary(actions)

day hourlon lat age

Min. :14810 Min. : 6.255 Min. :-1.18481 Min. : 0.00

1st Qu.:14813 1st Qu.:12.316 1st Qu.: 0.06505 1st Qu.:51.49

Median :14814 Median :12.982 Median : 0.06798 Median :51.49

Mean :14813 Mean :13.370 Mean :-0.05914 Mean :48.54

3rd Qu.:14814 3rd Qu.:14.819 3rd Qu.: 0.06945 3rd Qu.:51.49

Max. :14814 Max. :20.624 Max. : 0.07691 Max. :52.98

r X action playerId

Min. : 0.000 Min. : 4 CommunitiesButton :1034 P573 : 244

1st Qu.: 0.000 1st Qu.: 97 CreateMemberButton : 696 P588 : 154

Median : 0.000 Median : 240 createMember.done : 601 P578 : 143

Mean : 7.716 Mean : 18910 placeMember.done : 314 P592 : 130

3rd Qu.: 0.000 3rd Qu.: 494 createMember.failed: 133 P575 : 111

Max. :3044.000 Max. :185303 CreateStoryButton : 69 P589 : 97

NA's : 182 (Other) : 191 (Other):2159

gameId

GA523 :849

GA526 :441

GA525 :389

GA522 :296

GA514 :158

GA516 :155

(Other):750

> ainc <- (actions$gameId=='GA522' | actions$gameId=='GA523' | actions$gameId=='GA525' | actions$gameId=='GA526')

> actionf <- factor(actions$action[ainc])

> summary(actionf)

addStory.done addStory.failed carryMember.done

12 6 23

carryMember.failed CommunitiesButton createMember.done

8 624 449

createMember.failed createMember.started CreateMemberButton

85 14 507

CreateStoryButton placeMember.done placeMember.failed

37 169 39

placeMember.started

2

> actions$playerId <- factor(actions$playerId)

createMember.done or .failed !

createMemberF <- (actions$action=="createMember.done" | actions$action=="createMember.failed")

createMember <- tapply(seq(1,length(year))[createMemberF],playerId[createMemberF],length)

> createMember

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

3 26 4 10 5 34 19 26 20 21 22 17 18 22 34 23

P584 P585 P586 P587 P588 P589 P590 P591 P592

34 34 19 19 17 42 11 9 47

> summary(createMember)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.00 17.00 20.00 21.44 26.00 47.00

> hist(createMember,breaks=seq(0,50,by=1))



createMemberF2 <- (actions$action=="createMember.done")

createMember2 <- mapply(function(x) if (is.na(x)) 0 else x, tapply(seq(1,length(year))[createMemberF2],playerId[createMemberF2],length))

> createMember2

P568 P569 P570 P571 P572 P573 P574 P575 P576 P577 P578 P579 P580 P581 P582 P583

0 14 3 6 5 34 14 22 15 21 18 4 15 20 34 21

P584 P585 P586 P587 P588 P589 P590 P591 P592

12 24 15 18 12 31 11 0 47

> summary(createMember2)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 11.00 15.00 16.64 21.00 47.00

> createMember2/createMember\*100

P568 P569 P570 P571 P572 P573 P574 P575

0.00000 53.84615 75.00000 60.00000 100.00000 100.00000 73.68421 84.61538

P576 P577 P578 P579 P580 P581 P582 P583

75.00000 100.00000 81.81818 23.52941 83.33333 90.90909 100.00000 91.30435

P584 P585 P586 P587 P588 P589 P590 P591

35.29412 70.58824 78.94737 94.73684 70.58824 73.80952 100.00000 0.00000

P592

100.00000

hist(createMember2,breaks=seq(0,50,by=1))



> summary(createMember2/createMember\*100)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 70.59 78.95 72.68 94.74 100.00

hist(createMember2/createMember\*100,breaks=seq(0,100,by=5))

