

# Fun Text Mining Project

Xi Fang

6/23/2020

A survey was conducted among sports fan about their impressions on convention sports and e-sport. This project uses modified data from Hui Du at UGA.

Let's dig into their response to see what are some common impressions

```
df1 <- read.csv("words.csv")
# combine columns 2 to 21
col <- colnames(df1)[2:21]
df1$text <- apply(df1[,col], 1, paste, collapse = " ")
df2 <- df1[,c(1,22)]
head(df2)
```

```
##      ID
## 1     1
## 2     2
## 3     3
## 4     4
## 5     5
## 6     6
##
## 1
## 2
## 3 It is old school It brings many families together It is a good conversation point Almost everyone l
## 4
## 5
## 6
```

```
# formatting
```

```
library(tm)
```

```
## Loading required package: NLP
```

```
library(quanteda)
```

```
## Package version: 2.0.1
```

```
## Parallel computing: 2 of 8 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
##
```

```
## Attaching package: 'quanteda'
```

```
## The following objects are masked from 'package:tm':
```

```
##
```

```
##      as.DocumentTermMatrix, stopwords
```

```
## The following objects are masked from 'package:NLP':
```

```
##
```

```
##      meta, meta<-
```

```
## The following object is masked from 'package:utils':
```

```
##
```

```
##      View
```

```
corpus <- corpus(df2, docid_field = "ID",
```

```
                text_field = "text",
```

```
                unique_docnames = TRUE)
```

```
corpus <- Corpus(VectorSource(corpus))
```

```
# clean text
```

```
corpus <- tm_map(corpus, tolower)
```

```
corpus <- tm_map(corpus, removePunctuation)
```

```
corpus <- tm_map(corpus, removeNumbers)
```

```
corpus <- tm_map(corpus, stemDocument)
```

```
clean <- tm_map(corpus, removeWords, stopwords('english'))
```

```
# stopwords('english')
```

```
clean <- tm_map(clean, removeWords, c('veri'))
```

```
inspect(corpus[1:3])
```

```
## <<SimpleCorpus>>
```

```
## Metadata: corpus specific: 1, document level (indexed): 0
```

```
## Content: documents: 3
```

```
##
```

```
##
```

```
##
```

```
##
```

```
##
```

```
##
```

```
## it is old school it bring mani famili togeth it is a good convers point almost everyon know someth a
```

```
# Inspect word frequency
```

```
freq <- TermDocumentMatrix(clean)
```

```
freqset <- as.data.frame(as.matrix(freq))
```

```
write.table(freqset, file='wordfreq_all.csv', sep = ',')
```

```
freq
```

```
## <<TermDocumentMatrix (terms: 1989, documents: 450)>>
```

```
## Non-/sparse entries: 9832/885218
```

```
## Sparsity           : 99%
```

```
## Maximal term length: 15
```

```
## Weighting          : term frequency (tf)
```

```
findFreqTerms(freq, lowfreq = 5)
```

```
## [1] "basebal"      "big"          "broadcast"    "busi"         "can"
## [6] "competit"    "comput"      "entertain"    "eye"          "fair"
## [11] "fast"        "great"       "hand"         "lucrat"       "nba"
## [16] "new"         "nfl"         "nhl"         "particip"    "physic"
## [21] "realli"     "salari"      "skill"        "small"       "smart"
## [26] "attend"     "creativ"     "enjoy"        "friend"      "game"
## [31] "good"       "happi"       "innov"        "interest"    "join"
## [36] "joy"        "leagu"       "lot"          "practic"     "team"
## [41] "train"      "athlet"      "audienc"      "bring"       "differ"
## [46] "dont"       "event"       "everyon"      "famili"      "get"
## [51] "grow"       "know"        "mani"         "media"       "money"
## [56] "old"        "onli"        "peopl"        "popular"     "rise"
## [61] "school"     "someth"      "spend"        "sport"       "think"
## [66] "togeth"     "tradi"       "young"        "contract"    "injury"
## [71] "internet"   "sponsorship" "tournament"   "uniform"     "box"
## [76] "fantasi"    "focus"      "larg"         "onlin"       "pace"
## [81] "stream"     "teamwork"    "televis"      "twitch"      "championship"
## [86] "cheer"      "commit"      "disput"       "electron"    "organ"
## [91] "play"       "prepar"      "profess"      "talent"     "varieti"
## [96] "video"      "addict"      "best"         "involv"     "jump"
## [101] "natur"      "run"         "view"         "crowd"       "difficult"
## [106] "fun"        "intens"      "rich"         "stadium"     "sweat"
## [111] "access"     "conveni"     "countri"      "high"        "learn"
## [116] "like"       "make"        "much"         "profession"  "take"
## [121] "thing"      "watch"       "wish"         "exercis"     "fame"
## [126] "overpaid"   "thrill"      "win"          "bar"         "crazi"
## [131] "gambl"      "group"       "hockey"       "import"      "korea"
## [136] "opportun"   "polit"       "reflex"       "requir"      "sad"
## [141] "season"     "starcraft"   "tribal"       "nice"        "odd"
## [146] "quick"      "wast"        "better"       "career"      "control"
## [151] "energi"     "enthusiast"  "fit"          "free"        "gain"
## [156] "healthi"    "mind"        "pleasant"     "social"      "stress"
## [161] "challeng"   "hardwork"    "incom"        "real"        "sourc"
## [166] "beer"       "cool"        "drink"        "excit"       "parti"
## [171] "communic"   "dure"        "esport"       "just"        "one"
## [176] "player"     "tactic"      "cant"         "funni"       "help"
## [181] "honest"     "improv"      "knowledg"     "posit"       "time"
## [186] "basketbal"  "headset"     "paid"         "ticket"      "brain"
## [191] "coach"      "compet"      "famous"       "footbal"     "legend"
## [196] "platform"   "scholarship" "uniti"        "world"       "ball"
## [201] "hope"       "jolli"       "expert"       "hobbi"       "live"
## [206] "relax"      "interact"    "lazi"         "strong"      "counter"
## [211] "cup"        "fifa"        "lol"          "soccer"      "strike"
## [216] "videogam"   "youtub"      "nerdi"        "bet"         "brand"
## [221] "cheerlead"  "food"        "loud"         "action"      "bros"
## [226] "call"       "duti"        "expens"       "manag"       "super"
## [231] "equip"      "graphic"     "citi"         "playstat"    "pubg"
## [236] "bodi"       "intellig"    "life"         "short"       "sponsor"
## [241] "advertis"   "announc"     "fan"          "field"       "prize"
## [246] "strateg"    "strength"    "success"      "uniqu"       "dumb"
## [251] "full"       "nerd"        "spectat"     "stupid"      "dollar"
```

## [256]	"gamer"	"gear"	"huge"	"million"	"six"
## [261]	"spirit"	"trophi"	"adrenalin"	"champion"	"glori"
## [266]	"rivalri"	"stake"	"winner"	"hard"	"home"
## [271]	"hour"	"mental"	"pressur"	"strategi"	"work"
## [276]	"aggress"	"awesom"	"bore"	"long"	"realist"
## [281]	"technic"	"cheat"	"playoff"	"shooter"	"speed"
## [286]	"youth"	"award"	"connect"	"nation"	"avail"
## [291]	"kid"	"love"	"pay"	"see"	"top"
## [296]	"want"	"way"	"year"	"coordin"	"digit"
## [301]	"less"	"sunday"	"activ"	"health"	"name"
## [306]	"payment"	"perform"	"score"	"stand"	"star"
## [311]	"come"	"easi"	"generat"	"look"	"understand"
## [316]	"communiti"	"histori"	"miss"	"well"	"colleg"
## [321]	"level"	"role"	"technolog"	"virtual"	"athletic"
## [326]	"instinct"	"trade"	"amaz"	"confid"	"eager"
## [331]	"engag"	"fierc"	"madden"	"choos"	"dota"
## [336]	"fortnit"	"non"	"male"	"streamer"	"support"
## [341]	"sit"	"tough"	"individu"	"industri"	"match"
## [346]	"start"	"tech"	"worldwid"	"attent"	"problem"
## [351]	"relief"	"solv"	"status"	"competitor"	"concentr"
## [356]	"dedic"	"distanc"	"experi"	"victori"	"pro"
## [361]	"race"	"reward"	"age"	"corpor"	"sometim"
## [366]	"follow"	"need"	"multiplay"	"muscl"	"set"
## [371]	"earn"	"job"	"passion"	"disciplin"	"logic"
## [376]	"rule"	"alway"	"feel"	"foot"	"believ"
## [381]	"favorit"	"now"	"capac"	"critic"	"excercis"
## [386]	"goal"	"increas"	"memori"	"outdoor"	"power"
## [391]	"respons"	"screen"	"self"	"talk"	"use"
## [396]	"visual"	"arena"	"invest"	"cultur"	"danger"
## [401]	"enough"	"tension"	"proud"	"dream"	"potenti"
## [406]	"seat"	"endur"	"boost"	"develop"	"modern"
## [411]	"espn"	"final"	"hall"	"abil"	"loyalti"
## [416]	"major"	"special"	"hero"	"lame"	"celebr"
## [421]	"everi"	"sound"	"encourag"	"enthusiasm"	"qualiti"
## [426]	"elit"	"intern"	"becom"	"calm"	"consol"
## [431]	"person"	"movement"	"overwatch"	"techniqu"	"part"
## [436]	"fight"	"fighter"	"seri"	"inspir"	"teenag"
## [441]	"futur"	"owner"	"comfort"	"creat"	"give"
## [446]	"stay"	"bad"	"motiv"	"draft"	"serious"
## [451]	"back"	"deal"	"low"	"still"	"advanc"
## [456]	"effort"	"curios"	"pass"	"fake"	"competet"
## [461]	"greedi"	"cricket"	"attitud"	"confus"	"mlb"
## [466]	"statist"	"xbox"	"lose"	"loser"	"profit"
## [471]	"market"	"base"	"asian"	"wonder"	"general"
## [476]	"usual"	"simul"	"friendship"	"achiev"	"energet"
## [481]	"pride"	"soon"	"summer"	"smash"	"men"
## [486]	"theyr"	"olymp"	"relat"	"trendi"	"benefit"
## [491]	"reduc"	"right"	"man"	"domin"	"escap"
## [496]	"noth"	"american"	"form"	"concuss"	"regular"

*# delete words appears less than 0.05 % frequency*

```
word <- removeSparseTerms(freq, 0.995)
word1 <- as.data.frame(as.matrix(word))
```

```
rownames(word1)<-make.names(rownames(word))
word <- as.matrix(word1)
head(word1)
```

```
##      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
## basebal 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
## big     2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## broadcast 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi     2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can      1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## competit 1 0 0 0 1 3 0 2 1 1 0 0 0 0 0 0 0 0 0 0 2 1 1 1 0 0
##      27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49
## basebal  0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## competit 1 0 3 0 4 0 0 2 1 0 0 0 0 0 0 1 1 0 2 2 2 0 0 2
##      50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
## basebal  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 0 0 0 0 0 2 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2
## competit 2 0 0 0 0 0 1 0 0 0 3 3 2 1 1 0 4 2 0 0 0 0 2
##      73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
## basebal  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
## competit 0 0 0 0 2 0 0 2 0 0 1 5 0 1 1 0 1 1 0 0 2 0 4
##      96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113
## basebal  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## competit 1 1 0 0 0 2 0 2 2 2 2 0 2 0 0 0 2 0
##      114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130
## basebal  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0
## competit 2 0 0 2 0 0 1 0 2 1 0 0 0 2 0 0 0
##      131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147
## basebal  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## big       0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## competit 0 0 2 0 4 0 0 0 0 2 2 1 1 1 1 1 0
```

##	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164
## basebal	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	0	4	1	0	2	0	0	0	2	1	4	2	0	0	3	0	3
##	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	3
## competit	0	1	1	1	4	0	1	0	0	0	2	0	0	3	0	0	2
##	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
## competit	1	0	0	1	0	0	0	0	2	0	2	0	0	0	2	1	1
##	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
## busi	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
## competit	0	1	0	2	0	2	0	0	3	1	3	1	1	0	2	3	0
##	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	1	1	2	3	1	0	2	0	0	0	1	0	1	2	0	0	4
##	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
## competit	1	0	0	0	0	0	0	3	0	1	1	3	0	0	1	2	4
##	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	3	1	1	1	1	2	1	3	2	1	0	1	1	0	0	0	0
##	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

## can	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
## competit	0	0	4	1	0	1	2	0	1	1	1	2	2	0	1	0	1
##	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
## busi	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	0	2	0	0	1	1	1	0	1	3	4	0	0	4	1	3	0
##	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317
## basebal	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
## competit	0	1	0	0	0	2	1	0	1	0	0	2	1	1	2	4	1
##	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334
## basebal	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	2	0	0	1	0	1	0	0	2	0	0	0
## broadcast	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
## can	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0
## competit	0	3	2	1	0	0	0	2	1	1	1	1	2	2	0	0	0
##	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351
## basebal	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	4
## competit	2	1	4	1	0	0	1	0	1	1	2	0	2	2	0	0	0
##	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368
## basebal	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	0	1	2	0	2	1	0	1	0	0	1	2	2	2	0	3	0
##	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1
## broadcast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	0	1	0	1	0	0	0	0	0	1	0	0	3	0	0	0	0
##	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402
## basebal	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
## broadcast	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
## busi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## can	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## competit	1	0	0	2	0	1	0	3	1	2	0	0	0	1	2	1	0
##	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419
## basebal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## big	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 1 0 0 6 0 0 0 0 0 0 0 0
## competit 0 1 1 2 0 2 0 0 0 3 4 3 0 1 0 0 0
##           420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436
## basebal   0 0 0 0 0 0 0 0 1 1 0 1 1 1 2 1 1
## big       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0
## competit 0 2 0 0 0 3 0 0 0 2 2 1 1 0 0 1 1
##           437 438 439 440 441 442 443 444 445 446 447 448 449 450
## basebal   0 0 1 0 1 1 0 0 0 0 0 0 0 1
## big       0 0 0 0 0 0 0 0 0 0 0 1 0 0
## broadcast 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## busi      0 0 0 0 0 0 0 0 0 0 0 0 0 0
## can       0 0 0 0 0 0 0 0 0 0 0 0 0 3
## competit 0 0 0 0 1 2 0 0 4 0 0 0 1 0
```

```
dim(word1)
```

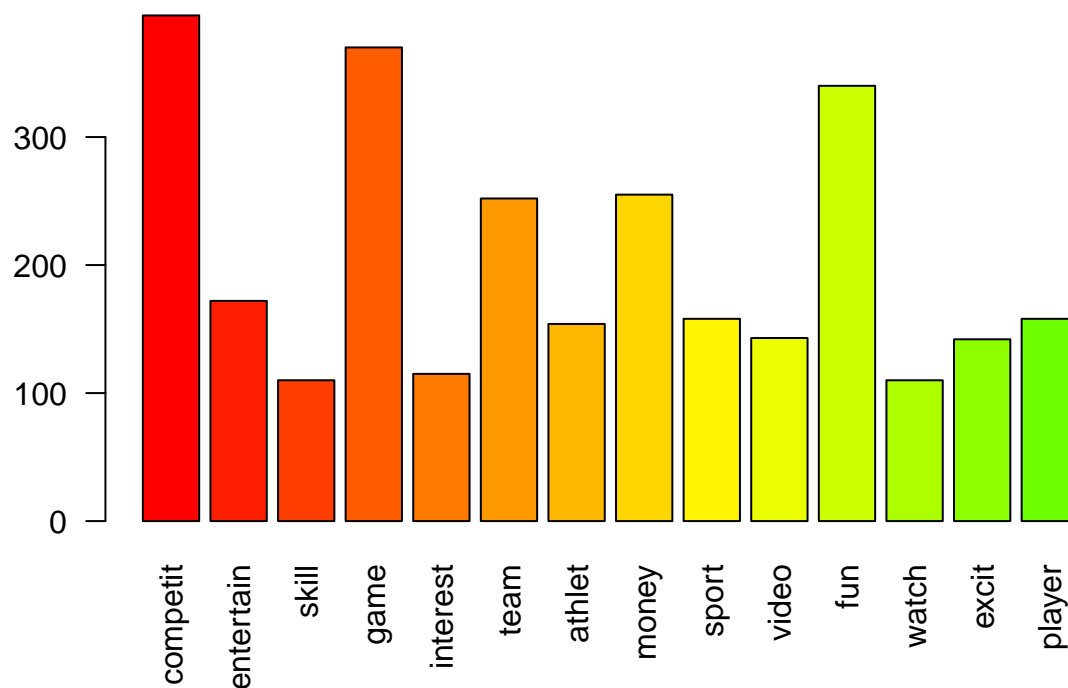
```
## [1] 692 450
```

```
write.table(word1, file='wordfreq_trimmed.csv', sep=',')
```

## Visualization

```
# frequency visualization
c <- rowSums(word1)
c <- subset(c, c >= 100)
barplot(c,
        las = 2,
        col = rainbow(50))
```





```
# Word cloud
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
v <- sort(rowSums(word1), decreasing = TRUE)
wordcloud(words = names(v),
  freq = v,
  max.words = 30,
  random.order = F,
  min.freq = 30,
  colors = brewer.pal (8, 'Dark2'), scale = c(4, 0.4))
```

