getwd()

setwd('C:/Users/sy09p033/Documents/SY09/TP1')

books <- read.csv("anonymous-betting-data.csv")

source("pretraitements.R")

ls()

books

summary(books)

Questions

On pourra adopter la stratégie d’étude suivante.

1. Faire une analyse descriptive générale des données, en répondant par exemple aux questions

suivantes :

**combien de matches concernent-elles : 26532 ; length(levels(books$match\_uid)) (class(book$match\_uid)) -> factor**

**combien de joueurs : 1527 length(levels(books$winner)) // length(levels(books$loser))**

**length(levels(as.factor(c(books$winner, books$loser)))) : 1527**

**sur quelle période de temps ? :**

**class(books$year)**

**min(books$year) : 2009**

**max(books$year) : 2015**

**période 2009 à 2015**

**Nombre de paris :**

**length(levels(books$match\_book\_uid)) -> 129271 // nrow(books)**

**Nombre de bookmakers :**

**length(levels(books$book)) : 7**

**Nombre de paris d’annulés or walkover :**

**Summary(books) : (logicial) 2151**

**Restants : 127120**

(Ces suggestions ne constituent évidemment pas une liste exhaustive.)

2 : On s’intéressera ensuite aux joueurs. En particulier, on commencera par calculer le nombre de matches gagnés, perdus, et donc joués par chaque joueur.

/\*\*\*

**FAUX : levels(as.factor(c(books$match\_uid, books$winner)))**

aggregate**(books$match\_uid**, by=list(**books$winner, books$loser**), max)

aggregate(books$winner, list(books$match\_uid), unique)

levels(books$winner) \*\*\*/

**levels(as.factor(c(levels(books$winner), levels(books$loser)))) : nb de joueurs totals**

aggregate(**levels(as.factor(c(levels(books$winner), levels(books$loser))))**, list(books$match\_uid), unique)

On pourra catégoriser les joueurs en fonction de leur propension à gagner les matches, et représenter l’information du nombre de matches joués (gagnés, perdus) en fonction du niveau du joueur.