BIOS 6301: Assignment 6

Charlie Rhea

Due Tuesday, 24 October, 1:00 PM

$5^{n=day}$ points taken off for each day late.

40 points total.

Submit a single knitr file (named `homework6.rmd`), along with a valid PDF output file. Inside the file, clearly indicate which parts of your responses go with which problems (you may use the original homework document as a template). Add your name as `author` to the file's metadata section. Raw R code/output or word processor files are not acceptable.

Failure to name file `homework6.rmd` or include author name may result in 5 points taken off.

Question 1

16 points

Obtain a copy of the football-values lecture. Save the five 2023 CSV files in your working directory.

Modify the code to create a function. This function will create dollar values given information (as arguments) about a league setup. It will return a data.frame and write this data.frame to a CSV file. The final data.frame should contain the columns 'PlayerName', 'pos', 'points', 'value' and be ordered by value descendingly. Do not round dollar values.

Note that the returned data.frame should have `sum(posReq)*nTeams` rows.

Define the function as such (10 points):

```
# path: directory path to input files
# file: name of the output file; it should be written to path
# nTeams: number of teams in league
# cap: money available to each team
# posReq: number of starters for each position
# points: point allocation for each category

path = '/Users/charlesrhea/Desktop/BIOS 6301 - Introduction to Statistical Computing/#Homework Assignmen
ffvalues <- function(path, file='outfile.csv', nTeams=12, cap=200, posReq=c(qb=1, rb=2, wr=3, te=1, k=1)
                     points=c(fg=4, xpt=1, pass_yds=1/25, pass_tds=4, pass_ints=-2,
                              rush_yds=1/10, rush_tds=6, fumbles=-2, rec_yds=1/20, rec_tds=6)) {

#1: Read in the 5 CSV files
positions = c('k','qb','rb','te','wr')
csvfile = paste('proj_', positions,'23.csv', sep='')
datafiles = file.path(path, csvfile)
names(datafiles) = positions

k = read.csv(datafiles['k'])
```

```
qb = read.csv(datafiles['qb'])
rb = read.csv(datafiles['rb'])
te = read.csv(datafiles['te'])
wr = read.csv(datafiles['wr'])

#2: Calculate Dollar Values
#Merge into 1 dataset
cols = unique(c(names(k), names(qb), names(rb), names(te), names(wr), "pos"))

k[, setdiff(cols, names(k))] = 0
k$pos = "k"
qb[, setdiff(cols, names(qb))] = 0
qb$pos = "qb"
rb[, setdiff(cols, names(rb))] = 0
rb$pos = "rb"
te[, setdiff(cols, names(te))] = 0
te$pos = "te"
wr[, setdiff(cols, names(wr))] = 0
wr$pos = "wr"

x = rbind(k[,cols], qb[,cols], rb[,cols], te[,cols], wr[,cols])

#3 Calculate Points per Player
for (i in 1:length(points)){
  x[, paste("p_", names(points[i]), sep ='' )] = x[, names(points[i])] * points[i]
  }
x$points = rowSums(x[ , grep("^p_", names(x))])

#4 Calculating Player Value
#Rank players by points
x = x[order(x['points'], decreasing=TRUE),]

#Calculate marginal points by position
x$marg = NA

for (i in 1:length(names(posReq))){
  if (posReq[i] != 0){
x$marg[which(x$pos == names(posReq)[i])] = x$points[which(x$pos == names(posReq)[i])] - x$points[which(
}}

#Drop negative/missing marginal points
x = na.omit(x)
x2 = x[x$marg >= 0,]

#Order by marginal points
x2 = x2[order(x2['marg'], decreasing=TRUE),]

#Calculate dollar value (from Lecture #1)
rownames(x2) <- NULL
x2$value = ((cap*nTeams) - (nTeams*sum(posReq))) * (x2$marg/sum(x2$marg)) + 1

#Final data.frame with needed information
x3 = x2[,c('PlayerName','pos','points', 'value')]
```

```
#Save as CSV file
write.csv(x3, file = file)

#Return data.frame with dollar values
return(x3)
}
```

1. Call `x1 <- ffvalues('.')`

    1. How many players are worth more than $20? (1 point) *44 players are worth more than $20*

```
#x1 <- ffvalues('.')
#sum(x1$value>20)
```

    2. Who is 15th most valuable running back (rb)? (1 point) *Kenneth Walker III*

```
#x1$PlayerName[which(x1$pos == 'rb')[15]]
```

2. Call `x2 <- ffvalues(getwd(), '16team.csv', nTeams=16, cap=150)`

    1. How many players are worth more than $20? (1 point) *46 players are worth more than $20*

```
#x2 <- ffvalues(getwd(), '16team.csv', nTeams=16, cap=150)
#sum(x2$value>20)
```

    2. How many wide receivers (wr) are in the top 40? (1 point) *11 wide receivers are in the top 40*

```
#sum(which(x2$pos == 'wr')<41)
```

3. Call:

```
#x3 <- ffvalues('.', 'qbheavy.csv', posReq=c(qb=2, rb=2, wr=3, te=1, k=0),
                      points=c(fg=0, xpt=0, pass_yds=1/25, pass_tds=6, pass_ints=-2,
                      rush_yds=1/10, rush_tds=6, fumbles=-2, rec_yds=1/20, rec_tds=6))
```

    1. How many players are worth more than $20? (1 point) *39 players are worth more than $20*

```
#sum(x3$value>20)
```

    2. How many quarterbacks (qb) are in the top 30? (1 point) *17 quarterbacks are in the top 40*

```
#sum(which(x3$pos == 'qb')<30)
```

Question 2

24 points

Import the HAART dataset (`haart.csv`) from the GitHub repository into R, and perform the following manipulations: (4 points each)

```r
library(readr)
haart.ds <- read_csv("~/Desktop/BIOS 6301 - Introduction to Statistical Computing/datasets/haart.csv")
```

```
## Rows: 1000 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): init.reg, init.date, last.visit, date.death
## dbl (8): male, age, aids, cd4baseline, logvl, weight, hemoglobin, death
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
View(haart.ds)
```

1. Convert date columns into a usable (for analysis) format. Use the `table` command to display the counts of the year from `init.date`.

```r
haart.ds$init.date <- as.Date(haart.ds$init.date,"%m/%d/%y")
haart.ds$last.visit <- as.Date(haart.ds$last.visit,"%m/%d/%y")
haart.ds$date.death <- as.Date(haart.ds$date.death,"%m/%d/%y")

haart.ds$years <- format(haart.ds$init.date, format = "%Y")
table(haart.ds$years)
```

```
##
## 1998 2000 2001 2002 2003 2004 2005 2006 2007
##    1    5   17   60  270  292  207  104   44
```

2. Create an indicator variable (one which takes the values 0 or 1 only) to represent death within 1 year of the initial visit. How many observations died in year 1? *92 observations died in year 1*

```r
haart.ds$oneyeardeath_years <- difftime(haart.ds$date.death, haart.ds$init.date, unit = "weeks")/52.25
haart.ds$oneyeardeath_years <- as.numeric(haart.ds$oneyeardeath_years)
haart.ds$oneyeardeath_indi <- as.numeric(haart.ds$oneyeardeath_years < 1)
table(haart.ds$oneyeardeath_indi)
```

```
##
##  0  1
## 25 92
```

3. Use the `init.date`, `last.visit` and `death.date` columns to calculate a followup time (in days), which is the difference between the first and either the last visit or a death event (whichever comes first). If these times are longer than 1 year, censor them (this means if the value is above 365, set followup to 365). Print the quantile for this new variable. *See output below*

```r
#Collapse the last.visit and death.date into a single column and take the value of the date that occurr
haart.ds$followup_refdate <- apply(haart.ds[,c(10,12)],1,min,na.rm = TRUE)

#Determine follow-up time from init.date to followup_refdate by number of days
haart.ds$followup_days <- round(difftime(haart.ds$followup_refdate, haart.ds$init.date, unit = "days"),
#Censor, or replace those with a follow-up time > 365 with the value 365; print quantiles of this varia
haart.ds$followup_daysCENSOR <- replace(haart.ds$followup_days, haart.ds$followup_days > 365, 365)
print(quantile(haart.ds$followup_daysCENSOR, na.rm = TRUE))
```

```
## Time differences in days
##     0%    25%    50%    75%   100%
##   0.00 320.75 365.00 365.00 365.00
```

4. Create another indicator variable representing loss to followup; this means the observation is not known to be dead but does not have any followup visits after the first year. How many records are lost-to-followup? *173 records were lost to follow-up*

```
#Create variable to indicate observations who are not known to be dead (death = 0) and whose follow-up
haart.ds$losstofollowup <- NA
haart.ds[haart.ds$death == 0 & haart.ds$followup_days < 365, "losstofollowup"] <- 1
table(haart.ds$losstofollowup)
```

```
##
##   1
## 173
```

5. Recall our work in class, which separated the `init.reg` field into a set of indicator variables, one for each unique drug. Create these fields and append them to the database as new columns. Which drug regimen are found over 100 times? *Two regimens are found over 100 times - 3TC, AZT, and EVF (421 times), and 3TC, AZT, and NVP (284 times)*

```
#Create a variable that contains a list to determine the number of times a regimen has be perscribed
init.reg <- as.character(haart.ds$init.reg)
(haart.ds[['init.reg_list']] <- strsplit(init.reg, ","))[1:3]
```

```
## [[1]]
## [1] "3TC" "AZT" "EFV"
##
## [[2]]
## [1] "3TC" "AZT" "EFV"
##
## [[3]]
## [1] "3TC" "AZT" "EFV"
```

```
table(haart.ds$init.reg)
```

```
##
##         3TC,ABC,AZT 3TC,ABC,AZT,LPV,RTV 3TC,ABC,AZT,RTV,SQV         3TC,ABC,EFV
##                  29                   1                   1                  11
##     3TC,ABC,IDV,RTV         3TC,ABC,NVP         3TC,ABC,RTV     3TC,ABC,RTV,SQV
##                   1                   2                   1                   4
##         3TC,AZT,DDI         3TC,AZT,EFV     3TC,AZT,EFV,NFV     3TC,AZT,FPV,RTV
##                   1                 421                   1                   1
##         3TC,AZT,IDV     3TC,AZT,IDV,RTV     3TC,AZT,LPV,RTV         3TC,AZT,NFV
##                  12                   8                  16                   4
##         3TC,AZT,NVP     3TC,AZT,RTV,SQV         3TC,D4T,EFV     3TC,D4T,IDV,RTV
##                 284                  13                  54                   6
##     3TC,D4T,LPV,RTV         3TC,D4T,NFV         3TC,D4T,NVP     3TC,D4T,RTV,SQV
##                   2                   3                  61                   8
##         3TC,DDI,EFV     3TC,DDI,LPV,RTV         3TC,DDI,NVP         3TC,EFV,TDF
```

```
##                  15                   4                   6                   1
##     3TC,FPV,RTV,TDF     3TC,LPV,RTV,TDF     ABC,ATV,DDI,RTV         ABC,D4T,EFV
##                   1                   1                   1                   2
##     ABC,D4T,LPV,RTV     ABC,D4T,RTV,SQV     ABC,DDI,LPV,RTV     ATV,FTC,RTV,TDF
##                   1                   1                   1                   1
##         AZT,DDI,EFV         D4T,DDC,EFV         D4T,DDI,EFV         D4T,DDI,NVP
##                   2                   1                   4                   2
##         D4T,RTV,SQV         DDI,EFV,FTC DDI,LPV,RTV,SQV,T20         EFV,FTC,TDF
##                   1                   1                   1                   3
##     FTC,LPV,RTV,TDF         FTC,NVP,TDF         LPV,NVP,RTV
##                   2                   1                   2
```

```
#Unlist to identify the name and total number of each unique drug
unlist(haart.ds$init.reg_list)[seq(50)]
```

```
##  [1] "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "AZT" "NVP"
## [13] "3TC" "D4T" "EFV" "3TC" "AZT" "NVP" "3TC" "AZT" "NVP" "3TC" "AZT" "EFV"
## [25] "3TC" "ABC" "AZT" "3TC" "DDI" "NVP" "3TC" "AZT" "NVP" "3TC" "AZT" "IDV"
## [37] "3TC" "AZT" "NVP" "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "D4T" "NVP"
## [49] "3TC" "AZT"
```

```
(all_drugs <- unique(unlist(haart.ds$init.reg_list)))
```

```
##  [1] "3TC" "AZT" "EFV" "NVP" "D4T" "ABC" "DDI" "IDV" "LPV" "RTV" "SQV" "FTC"
## [13] "TDF" "DDC" "NFV" "T20" "ATV" "FPV"
```

```
#Create a matrix (1000 rows for 1000 observations, and 18 variables for 18 unique drugs) to and search
reg_drugs <- matrix(0, nrow=nrow(haart.ds), ncol=length(all_drugs))
for(i in seq_along(all_drugs)) {
  reg_drugs[,i] <- sapply(haart.ds$init.reg_list, function(x) all_drugs[i] %in% x)
}

#Martix results in a data frame to ease evaluation and allow for appending to original dataset
reg_drugs <- data.frame(reg_drugs)
names(reg_drugs) <- all_drugs
head(reg_drugs)
```

```
##   3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV FPV
## 1   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 5   1   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 6   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
#Appending drug data frame to original dataset
haart.ds_2 <- cbind(haart.ds, reg_drugs)
head(haart.ds_2)
```

```
##   male age aids cd4baseline logvl  weight hemoglobin    init.reg  init.date
## 1    1  25    0          NA    NA      NA         NA 3TC,AZT,EFV 2003-07-01
```

```
## 2    1  49    0         143    NA 58.0608         11 3TC,AZT,EFV 2004-11-23
## 3    1  42    1         102    NA 48.0816          1 3TC,AZT,EFV 2003-04-30
## 4    0  33    0         107    NA 46.0000         NA 3TC,AZT,NVP 2006-03-25
## 5    1  27    0          52     4      NA         NA 3TC,D4T,EFV 2004-09-01
## 6    0  34    0         157    NA 54.8856         NA 3TC,AZT,NVP 2003-12-02
##   last.visit death date.death years oneyeardeath_years oneyeardeath_indi
## 1 2007-02-26     0       <NA>  2003                 NA                NA
## 2 2008-02-22     0       <NA>  2004                 NA                NA
## 3 2005-11-21     1 2006-01-11  2003          2.6985646                 0
## 4 2006-05-05     1 2006-05-07  2006          0.1175666                 1
## 5 2007-11-13     0       <NA>  2004                 NA                NA
## 6 2008-02-28     0       <NA>  2003                 NA                NA
##   followup_refdate followup_days followup_daysCENSOR losstofollowup
## 1       2007-02-26     1336 days            365 days             NA
## 2       2008-02-22     1186 days            365 days             NA
## 3       2005-11-21      936 days            365 days             NA
## 4       2006-05-05       41 days             41 days             NA
## 5       2007-11-13     1168 days            365 days             NA
## 6       2008-02-28     1549 days            365 days             NA
##   init.reg_list 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20
## 1 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4 3TC, AZT, NVP   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0
## 5 3TC, D4T, EFV   1   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0
## 6 3TC, AZT, NVP   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0
##   ATV FPV
## 1   0   0
## 2   0   0
## 3   0   0
## 4   0   0
## 5   0   0
## 6   0   0
```

6. The dataset `haart2.csv` contains a few additional observations for the same study. Import these and append them to your master dataset (if you were smart about how you coded the previous steps, cleaning the additional observations should be easy!). Show the first five records and the last five records of the complete (and clean) data set. *See head output for first 5 observations, and tail output for last 5 observations below*

```
haart2.ds <- read_csv("~/Desktop/BIOS 6301 - Introduction to Statistical Computing/datasets/haart2.csv")
```

```
## Rows: 4 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): init.reg, init.date, last.visit
## dbl (8): male, age, aids, cd4baseline, logvl, weight, hemoglobin, death
## lgl (1): date.death
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
haart_orig.ds <- read_csv("~/Desktop/BIOS 6301 - Introduction to Statistical Computing/datasets/haart.cs
```

```
## Rows: 1000 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): init.reg, init.date, last.visit, date.death
## dbl (8): male, age, aids, cd4baseline, logvl, weight, hemoglobin, death
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
final_haart.ds <- rbind(haart_orig.ds, haart2.ds)
View(final_haart.ds)

final_haart.ds$init.date <- as.Date(final_haart.ds$init.date,"%m/%d/%y")
final_haart.ds$last.visit <- as.Date(final_haart.ds$last.visit,"%m/%d/%y")
final_haart.ds$date.death <- as.Date(final_haart.ds$date.death,"%m/%d/%y")

final_haart.ds$oneyeardeath_years <- difftime(final_haart.ds$date.death, final_haart.ds$init.date, unit
final_haart.ds$oneyeardeath_years <- as.numeric(final_haart.ds$oneyeardeath_years)
final_haart.ds$oneyeardeath_indi <- as.numeric(final_haart.ds$oneyeardeath_years < 1)

final_haart.ds$followup_refdate <- apply(final_haart.ds[,c(10,12)],1,min,na.rm = TRUE)
final_haart.ds$followup_days <- round(difftime(final_haart.ds$followup_refdate, final_haart.ds$init.dat
final_haart.ds$followup_daysCENSOR <- replace(final_haart.ds$followup_days, final_haart.ds$followup_days

final_haart.ds$losstofollowup <- NA
final_haart.ds[final_haart.ds$death == 0 & final_haart.ds$followup_days < 365, "losstofollowup"] <- 1

#Drug regimen, and indicator variables for each unique drug
init.reg <- as.character(final_haart.ds$init.reg)
(final_haart.ds[['init.reg_list']] <- strsplit(init.reg, ","))[1:3]
```

```
## [[1]]
## [1] "3TC" "AZT" "EFV"
##
## [[2]]
## [1] "3TC" "AZT" "EFV"
##
## [[3]]
## [1] "3TC" "AZT" "EFV"
```

```
unlist(final_haart.ds$init.reg_list)[seq(50)]
```

```
##  [1] "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "AZT" "NVP"
## [13] "3TC" "D4T" "EFV" "3TC" "AZT" "NVP" "3TC" "AZT" "NVP" "3TC" "AZT" "EFV"
## [25] "3TC" "ABC" "AZT" "3TC" "DDI" "NVP" "3TC" "AZT" "NVP" "3TC" "AZT" "IDV"
## [37] "3TC" "AZT" "NVP" "3TC" "AZT" "EFV" "3TC" "AZT" "EFV" "3TC" "D4T" "NVP"
## [49] "3TC" "AZT"
```

```
(all_drugs <- unique(unlist(final_haart.ds$init.reg_list)))
```

```
##  [1] "3TC" "AZT" "EFV" "NVP" "D4T" "ABC" "DDI" "IDV" "LPV" "RTV" "SQV" "FTC"
## [13] "TDF" "DDC" "NFV" "T20" "ATV" "FPV"
```

```
reg_drugs <- matrix(0, nrow=nrow(final_haart.ds), ncol=length(all_drugs))
for(i in seq_along(all_drugs)) {
  reg_drugs[,i] <- sapply(final_haart.ds$init.reg_list, function(x) all_drugs[i] %in% x)
}

reg_drugs <- data.frame(reg_drugs)
names(reg_drugs) <- all_drugs
head(reg_drugs)
```

```
##   3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV FPV
## 1   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 5   1   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 6   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

```
final2_haart.ds <- cbind(final_haart.ds, reg_drugs)
head(final2_haart.ds, 5)
```

```
##   male age aids cd4baseline logvl  weight hemoglobin    init.reg  init.date
## 1    1  25    0          NA    NA      NA         NA 3TC,AZT,EFV 2003-07-01
## 2    1  49    0         143    NA 58.0608         11 3TC,AZT,EFV 2004-11-23
## 3    1  42    1         102    NA 48.0816          1 3TC,AZT,EFV 2003-04-30
## 4    0  33    0         107    NA 46.0000         NA 3TC,AZT,NVP 2006-03-25
## 5    1  27    0          52     4      NA         NA 3TC,D4T,EFV 2004-09-01
##   last.visit death date.death oneyeardeath_years oneyeardeath_indi
## 1 2007-02-26     0       <NA>                NA                NA
## 2 2008-02-22     0       <NA>                NA                NA
## 3 2005-11-21     1 2006-01-11         2.6985646                 0
## 4 2006-05-05     1 2006-05-07         0.1175666                 1
## 5 2007-11-13     0       <NA>                NA                NA
##   followup_refdate followup_days followup_daysCENSOR losstofollowup
## 1       2007-02-26     1336 days            365 days             NA
## 2       2008-02-22     1186 days            365 days             NA
## 3       2005-11-21      936 days            365 days             NA
## 4       2006-05-05       41 days             41 days             NA
## 5       2007-11-13     1168 days            365 days             NA
##   init.reg_list 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20
## 1 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3 3TC, AZT, EFV   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4 3TC, AZT, NVP   1   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0
## 5 3TC, D4T, EFV   1   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0
##   ATV FPV
## 1   0   0
## 2   0   0
```

```
## 3   0   0
## 4   0   0
## 5   0   0
```

```
tail(final2_haart.ds, 5)
```

```
##      male      age aids cd4baseline    logvl  weight hemoglobin    init.reg
## 1000    0 40.00000    1         131       NA 46.2672          8 3TC,D4T,NVP
## 1001    0 27.00000    0         232       NA      NA         NA 3TC,AZT,NVP
## 1002    1 38.72142    0         170       NA 84.0000         NA 3TC,AZT,NVP
## 1003    1 23.00000   NA         154 3.995635 65.5000         14 3TC,DDI,EFV
## 1004    0 31.00000    0         236       NA 45.8136         NA 3TC,D4T,NVP
##       init.date last.visit death date.death oneyeardeath_years
## 1000 2003-07-03 2008-02-29     0       <NA>                 NA
## 1001 2003-12-01 2004-01-05     0       <NA>                 NA
## 1002 2002-09-26 2004-03-29     0       <NA>                 NA
## 1003 2007-01-31 2007-04-16     0       <NA>                 NA
## 1004 2003-12-03 2007-10-11     0       <NA>                 NA
##      oneyeardeath_indi followup_refdate followup_days followup_daysCENSOR
## 1000                NA       2008-02-29     1702 days            365 days
## 1001                NA       2004-01-05       35 days             35 days
## 1002                NA       2004-03-29      550 days            365 days
## 1003                NA       2007-04-16       75 days             75 days
## 1004                NA       2007-10-11     1408 days            365 days
##      losstofollowup init.reg_list 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV
## 1000             NA 3TC, D4T, NVP   1   0   0   1   1   0   0   0   0   0   0
## 1001              1 3TC, AZT, NVP   1   1   0   1   0   0   0   0   0   0   0
## 1002             NA 3TC, AZT, NVP   1   1   0   1   0   0   0   0   0   0   0
## 1003              1 3TC, DDI, EFV   1   0   1   0   0   0   1   0   0   0   0
## 1004             NA 3TC, D4T, NVP   1   0   0   1   1   0   0   0   0   0   0
##      FTC TDF DDC NFV T20 ATV FPV
## 1000   0   0   0   0   0   0   0
## 1001   0   0   0   0   0   0   0
## 1002   0   0   0   0   0   0   0
## 1003   0   0   0   0   0   0   0
## 1004   0   0   0   0   0   0   0
```