Assignment 5

James Lopez

11/15/2016

Grade 42/50

Check out how Cole approache Question 2 using lapply and tapply.

```
haart <- read.csv('haart.csv', stringsAsFactors = FALSE)
```

1

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

```
attach(haart)
haart[,"init.date"] <- as.Date(haart[,"init.date"], format = '%m/%d/%y')
haart[,"last.visit"] <- as.Date(haart[,"last.visit"], format = '%m/%d/%y')
haart[,"date.death"] <- as.Date(haart[,"date.death"], format = '%m/%d/%y')

table(format(haart[,'init.date'], '%Y'))
```

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

Part 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?

```
haart[,'year.dead'] <- rep(0,1000)
for(i in seq(nrow(haart))) {
  if(!is.na(haart[i,'date.death'])) {
  if(difftime(haart[i,'date.death'], haart[i,'init.date'],units='days') <= 365){
    haart[i,'year.dead'] <- 1
  }
  }
}

length(which(haart[,'year.dead'] == 1))
```

```
## [1] 92
```

92 patients died within 1 year.

Part 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.

```
haart[,'followup'] <- rep(0,1000)
for(i in seq(nrow(haart))) {
  if(!is.na(haart[i,'last.visit'])) {
    haart[i,'followup'] <- difftime(haart[i,'last.visit'],haart[i,'init.date'],units='days')
  }
```

```
  else {
    haart[i,'followup'] <- difftime(haart[i,'date.death'],haart[i,'init.date'],units='days')
  }
  if(haart[i,'followup'] >= 365) {
    haart[i,'followup'] <- 365
  }
}

quantile(haart[,'followup'])
```

```
##     0%    25%    50%    75%   100%
##   0.00 320.75 365.00 365.00 365.00
```

Part 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?

```
haart[,'ltf'] <- rep(0, nrow(haart))
for(i in seq(nrow(haart))) {
  if(is.na(haart[i,'date.death'])) {
    if(difftime(haart[i,'last.visit'], haart[i,'init.date'], units='days') >= 365) {
      haart[i,'ltf'] <- 1
    }
  }
}

length(which(haart[,'ltf'] == 1))
```

```
## [1] 710
```

There are 710 records lost-to-followup.

JC Grading -2 The answer is 173 . . . see below:

```
sum(haart$followup<365 & !haart$death)
```

```
## [1] 173
```

Part 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?

```
all.reg <- strsplit(haart[,'init.reg'], ',')
all.reg <- unlist(all.reg)
all.reg <- unique(all.reg)
row.reg <- strsplit(haart[,'init.reg'], ',')
user.reg <- sapply(all.reg, function(j) sapply(row.reg, function(i) j %in% i))
colSums(user.reg)
```

```
## 3TC AZT EFV NVP D4T ABC DDI IDV LPV RTV SQV FTC TDF DDC NFV T20 ATV FPV
## 973 794 516 358 146  56  38  27  31  79  29   8  10   1   8   1   2   2
```

3TC, AZT, EFV, NVP, and D4T regimen are found over 100 times.

```
haart <- cbind(haart, user.reg)
```

Part 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.

```
haart2 <- read.csv('haart2.csv', stringsAsFactors = FALSE)
haart2[,"init.date"] <- as.Date(haart2[,"init.date"], format = '%m/%d/%y')
haart2[,"last.visit"] <- as.Date(haart2[,"last.visit"], format = '%m/%d/%y')
haart2[,"date.death"] <- as.Date(haart2[,"date.death"], format = '%m/%d/%y')

all.reg2 <- strsplit(haart2[,'init.reg'], ',')
all.reg2 <- unlist(all.reg2)
all.reg2 <- unique(all.reg2)
row.reg2 <- strsplit(haart2[,'init.reg'], ',')
user.reg2 <- sapply(all.reg, function(j) sapply(row.reg2, function(i) j %in% i))
haart2 <- cbind(haart2, user.reg2)

#haart <- rbind(haart, haart2)
head(haart)
```

```
##   male age aids cd4baseline logvl  weight hemoglobin    init.reg
## 1    1  25    0          NA    NA      NA         NA 3TC,AZT,EFV
## 2    1  49    0         143    NA 58.0608         11 3TC,AZT,EFV
## 3    1  42    1         102    NA 48.0816          1 3TC,AZT,EFV
## 4    0  33    0         107    NA 46.0000         NA 3TC,AZT,NVP
## 5    1  27    0          52     4      NA         NA 3TC,D4T,EFV
## 6    0  34    0         157    NA 54.8856         NA 3TC,AZT,NVP
##    init.date last.visit death date.death year.dead followup ltf  3TC   AZT
## 1 2003-07-01 2007-02-26     0       <NA>         0      365   1 TRUE  TRUE
## 2 2004-11-23 2008-02-22     0       <NA>         0      365   1 TRUE  TRUE
## 3 2003-04-30 2005-11-21     1 2006-01-11         0      365   0 TRUE  TRUE
## 4 2006-03-25 2006-05-05     1 2006-05-07         1       41   0 TRUE  TRUE
## 5 2004-09-01 2007-11-13     0       <NA>         0      365   1 TRUE FALSE
## 6 2003-12-02 2008-02-28     0       <NA>         0      365   1 TRUE  TRUE
##     EFV   NVP   D4T   ABC   DDI   IDV   LPV   RTV   SQV   FTC   TDF   DDC
## 1  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 2  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 3  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 4 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 5  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 6 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##     NFV   T20   ATV   FPV
## 1 FALSE FALSE FALSE FALSE
## 2 FALSE FALSE FALSE FALSE
## 3 FALSE FALSE FALSE FALSE
## 4 FALSE FALSE FALSE FALSE
## 5 FALSE FALSE FALSE FALSE
## 6 FALSE FALSE FALSE FALSE
```

```
tail(haart)
```

```
##      male age aids cd4baseline    logvl  weight hemoglobin    init.reg
## 995     1  26    1          NA       NA      NA         NA 3TC,AZT,EFV
## 996     1  42    0         164 5.281029 84.0000         12 3TC,AZT,NVP
## 997     0  39    1         125 4.625312      NA         NA 3TC,AZT,EFV
## 998     0  37    0         122       NA 86.1840         11 3TC,AZT,NVP
## 999     0  31    0         102       NA 61.6896         11 3TC,AZT,NVP
## 1000    0  40    1         131       NA 46.2672          8 3TC,D4T,NVP
```

```
##       init.date last.visit death date.death year.dead followup ltf  3TC
## 995  2003-11-01 2005-12-20     0       <NA>         0      365   1 TRUE
## 996  2005-04-30 2007-04-13     0       <NA>         0      365   1 TRUE
## 997  2007-04-24       <NA>     1 2007-08-16         1      114   0 TRUE
## 998  2005-01-12 2008-03-19     0       <NA>         0      365   1 TRUE
## 999  2003-05-22 2008-03-07     0       <NA>         0      365   1 TRUE
## 1000 2003-07-03 2008-02-29     0       <NA>         0      365   1 TRUE
##        AZT   EFV   NVP   D4T   ABC   DDI   IDV   LPV   RTV   SQV   FTC
## 995   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 996   TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 997   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 998   TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 999   TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1000 FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##        TDF   DDC   NFV   T20   ATV   FPV
## 995  FALSE FALSE FALSE FALSE FALSE FALSE
## 996  FALSE FALSE FALSE FALSE FALSE FALSE
## 997  FALSE FALSE FALSE FALSE FALSE FALSE
## 998  FALSE FALSE FALSE FALSE FALSE FALSE
## 999  FALSE FALSE FALSE FALSE FALSE FALSE
## 1000 FALSE FALSE FALSE FALSE FALSE FALSE
```

JC Grading -3

File does not compile because of error in code above. Occurs at: haart <- rbind(haart, haart2). I've commented out to run rest of file.

Error in rbind(deparse.level, ...) : numbers of columns of arguments do not match

2

```
genData <- function(n) {
  if(exists(".Random.seed", envir = .GlobalEnv)) {
    save.seed <- get(".Random.seed", envir= .GlobalEnv)
    on.exit(assign(".Random.seed", save.seed, envir = .GlobalEnv))
  } else {
    on.exit(rm(".Random.seed", envir = .GlobalEnv))
  }
  set.seed(n)
  subj <- ceiling(n / 10)
  id <- sample(subj, n, replace=TRUE)
  times <- as.integer(difftime(as.POSIXct("2005-01-01"), as.POSIXct("2000-01-01"), units='secs'))
  dt <- as.POSIXct(sample(times, n), origin='2000-01-01')
  mu <- runif(subj, 4, 10)
  a1c <- unsplit(mapply(rnorm, tabulate(id), mu, SIMPLIFY=FALSE), id)
  data.frame(id, dt, a1c)
}
x <- genData(500)
```

Part 1) Order the data set by id and dt

```
y <- x[order(x['id'],x['dt']),]
```

Part 2) For each id, determine if there is more than a one year gap in between observations.

Add a new row at the one year mark, with the a1c value set to missing. A two year gap would require two new rows, and so forth.

```
z <- as.list(y)
z
```

```
## $id
##   [1]  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2
##  [24]  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3  3  3  4  4
##  [47]  4  4  4  4  4  4  4  4  5  5  5  5  5  5  5  5  5  5  5  5  6  6  6
##  [70]  6  6  6  6  6  7  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8  8
##  [93]  8  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10 10 10 10 10 10 10 10
## [116] 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 13 13 13 13 13 13 14
## [139] 14 14 14 14 14 14 14 14 14 14 15 15 15 15 15 16 16 16 16 16 16 16 17
## [162] 17 17 17 17 17 17 17 17 17 17 17 18 18 18 18 18 18 18 18 18 18 19 19
## [185] 19 19 19 19 19 19 20 20 20 20 20 20 20 21 21 21 21 21 21 21 21 21 21
## [208] 21 22 22 22 22 22 22 23 23 23 23 23 23 24 24 24 24 24 24 24 24 24 24
## [231] 24 24 24 24 24 25 25 25 25 25 25 25 25 25 25 25 25 26 26 26 26 26 26
## [254] 26 26 26 26 26 26 26 27 27 27 27 27 27 27 27 27 27 28 28 28 28 28 28
## [277] 28 28 28 28 28 28 28 29 29 29 29 29 29 29 29 29 29 30 30 30 30 30 30 31
## [300] 31 31 31 31 31 31 31 31 31 31 32 32 32 32 33 33 33 33 33 33 33 34 34
## [323] 34 34 34 34 34 34 34 35 35 35 35 35 35 35 35 35 35 36 36 36 36 36
## [346] 36 36 36 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 38 38 38 38 38
## [369] 38 38 38 38 38 38 38 38 38 39 39 39 39 39 39 40 40 40 40 40 40 41 41
## [392] 41 41 41 41 41 41 41 41 41 41 41 41 41 41 41 42 42 42 42 42 42 42 42
## [415] 42 42 42 42 42 43 43 43 43 43 43 43 43 43 43 44 44 44 44 44 44 44 44 44
## [438] 45 45 45 45 45 45 45 45 45 45 45 45 45 46 46 46 46 46 46 46 46 47 47
## [461] 47 47 47 47 47 47 47 47 47 47 48 48 48 48 48 48 48 48 48 49 49 49 49
## [484] 49 49 49 49 49 49 49 50 50 50 50 50 50 50 50 50 50
##
## $dt
##   [1] "2000-03-14 07:44:28 CST" "2000-05-04 10:29:31 CDT"
##   [3] "2000-11-13 03:32:47 CST" "2001-09-05 21:00:24 CDT"
##   [5] "2001-11-20 01:57:31 CST" "2002-08-28 04:44:21 CDT"
##   [7] "2002-10-01 19:06:42 CDT" "2003-06-01 09:43:43 CDT"
##   [9] "2003-11-27 21:46:34 CST" "2004-01-18 06:23:48 CST"
##  [11] "2004-03-24 15:24:19 CST" "2000-02-06 06:43:03 CST"
##  [13] "2000-09-27 18:35:07 CDT" "2001-02-07 23:15:11 CST"
##  [15] "2001-05-04 18:05:15 CDT" "2001-07-01 08:22:48 CDT"
##  [17] "2001-08-25 12:05:55 CDT" "2002-03-06 23:47:33 CST"
##  [19] "2002-04-07 23:22:23 CDT" "2002-04-08 07:05:28 CDT"
##  [21] "2002-06-23 02:57:34 CDT" "2002-09-23 20:36:47 CDT"
##  [23] "2002-11-30 16:56:02 CST" "2002-12-11 19:12:44 CST"
##  [25] "2003-02-19 08:55:37 CST" "2003-07-13 00:58:37 CDT"
##  [27] "2003-08-23 20:44:49 CDT" "2003-09-21 04:47:21 CDT"
##  [29] "2003-12-25 21:42:36 CST" "2004-03-07 21:49:37 CST"
##  [31] "2004-03-10 15:22:30 CST" "2001-02-14 12:42:06 CST"
##  [33] "2001-04-19 07:14:12 CDT" "2001-06-25 07:03:12 CDT"
##  [35] "2001-07-21 06:17:48 CDT" "2002-01-11 18:09:39 CST"
##  [37] "2003-05-10 11:26:11 CDT" "2003-06-25 10:43:23 CDT"
##  [39] "2003-07-02 04:03:06 CDT" "2003-11-14 02:05:22 CST"
##  [41] "2004-01-26 14:28:12 CST" "2004-02-03 13:23:23 CST"
##  [43] "2004-05-04 22:48:53 CDT" "2004-11-27 05:18:09 CST"
##  [45] "2000-05-27 12:32:42 CDT" "2000-06-03 17:25:08 CDT"
##  [47] "2001-04-19 16:42:08 CDT" "2002-04-22 01:14:00 CDT"
```

```
##  [49] "2002-11-29 09:42:26 CST" "2004-02-04 23:59:41 CST"
##  [51] "2004-06-05 15:50:07 CDT" "2004-06-25 08:44:00 CDT"
##  [53] "2004-07-24 02:46:27 CDT" "2004-08-31 09:54:18 CDT"
##  [55] "2000-01-12 12:43:36 CST" "2001-08-13 14:09:08 CDT"
##  [57] "2003-06-09 00:14:34 CDT" "2003-09-24 13:49:10 CDT"
##  [59] "2003-10-03 12:17:50 CDT" "2003-11-18 10:39:45 CST"
##  [61] "2003-11-25 06:27:58 CST" "2003-12-05 07:31:41 CST"
##  [63] "2004-07-01 15:20:47 CDT" "2004-10-26 11:11:32 CDT"
##  [65] "2004-11-21 07:43:52 CST" "2004-12-30 04:42:23 CST"
##  [67] "2000-02-16 06:11:59 CST" "2000-03-22 11:28:19 CST"
##  [69] "2000-08-12 16:59:18 CDT" "2001-09-19 18:37:44 CDT"
##  [71] "2002-03-07 23:10:31 CST" "2003-03-23 22:32:39 CST"
##  [73] "2003-07-19 13:17:54 CDT" "2004-04-25 21:32:52 CDT"
##  [75] "2000-02-01 10:21:44 CST" "2000-05-14 05:37:53 CDT"
##  [77] "2001-04-24 14:38:13 CDT" "2001-11-14 20:17:50 CST"
##  [79] "2002-02-05 14:38:36 CST" "2002-11-20 04:51:47 CST"
##  [81] "2003-03-28 16:33:34 CST" "2004-06-09 17:05:13 CDT"
##  [83] "2000-08-04 13:20:02 CDT" "2000-09-11 08:00:53 CDT"
##  [85] "2001-02-06 20:23:00 CST" "2001-02-21 20:14:08 CST"
##  [87] "2001-05-05 04:30:49 CDT" "2002-06-25 09:07:16 CDT"
##  [89] "2002-12-28 18:17:48 CST" "2003-04-18 15:55:58 CDT"
##  [91] "2004-01-06 17:55:47 CST" "2004-02-17 17:00:55 CST"
##  [93] "2004-12-25 20:55:29 CST" "2000-01-15 15:26:22 CST"
##  [95] "2000-08-03 10:25:56 CDT" "2001-02-07 05:14:52 CST"
##  [97] "2001-02-20 01:08:24 CST" "2001-12-17 03:48:47 CST"
##  [99] "2002-05-01 14:54:51 CDT" "2002-08-06 12:03:04 CDT"
## [101] "2003-03-21 20:24:14 CST" "2003-08-14 17:31:05 CDT"
## [103] "2004-03-23 06:43:11 CST" "2004-07-23 15:36:37 CDT"
## [105] "2000-04-24 14:09:06 CDT" "2000-09-02 01:50:21 CDT"
## [107] "2000-11-08 08:33:44 CST" "2001-01-21 21:30:57 CST"
## [109] "2001-07-06 16:20:16 CDT" "2001-10-10 07:44:16 CDT"
## [111] "2003-02-21 13:11:16 CST" "2003-05-07 22:28:51 CDT"
## [113] "2003-12-28 14:31:22 CST" "2004-03-10 12:12:37 CST"
## [115] "2004-12-02 19:22:10 CST" "2000-01-13 16:03:13 CST"
## [117] "2000-10-17 09:34:24 CDT" "2001-11-16 20:16:09 CST"
## [119] "2002-02-03 11:11:57 CST" "2003-04-08 11:08:06 CDT"
## [121] "2003-06-03 19:56:20 CDT" "2003-11-11 02:43:38 CST"
## [123] "2004-04-18 01:33:38 CDT" "2000-05-26 20:16:19 CDT"
## [125] "2000-07-27 04:27:43 CDT" "2000-10-02 10:30:49 CDT"
## [127] "2001-12-25 14:05:17 CST" "2002-03-16 20:03:27 CST"
## [129] "2003-04-01 21:48:21 CST" "2003-09-04 05:33:09 CDT"
## [131] "2004-01-24 18:19:20 CST" "2001-02-18 20:04:21 CST"
## [133] "2001-06-10 23:40:17 CDT" "2001-08-20 16:19:39 CDT"
## [135] "2001-10-16 17:33:00 CDT" "2003-05-30 20:05:40 CDT"
## [137] "2004-10-09 16:18:36 CDT" "2000-03-01 17:48:12 CST"
## [139] "2000-07-24 09:34:28 CDT" "2000-08-07 23:04:06 CDT"
## [141] "2001-05-10 01:23:26 CDT" "2001-08-26 22:14:13 CDT"
## [143] "2002-02-08 03:35:08 CST" "2003-04-10 16:31:27 CDT"
## [145] "2003-08-22 18:25:48 CDT" "2004-08-17 09:01:53 CDT"
## [147] "2004-09-13 20:16:31 CDT" "2004-12-10 20:49:22 CST"
## [149] "2000-04-30 00:34:50 CDT" "2001-01-17 21:11:02 CST"
## [151] "2001-04-25 06:23:05 CDT" "2003-06-06 14:06:00 CDT"
## [153] "2004-08-20 17:47:11 CDT" "2000-01-06 19:43:32 CST"
## [155] "2000-08-15 13:43:27 CDT" "2000-10-10 06:16:34 CDT"
```

```
## [157] "2001-03-13 16:56:36 CST" "2001-06-06 13:12:14 CDT"
## [159] "2002-03-04 13:59:03 CST" "2004-06-24 14:08:24 CDT"
## [161] "2000-06-19 07:41:48 CDT" "2001-03-15 04:54:28 CST"
## [163] "2001-03-20 09:37:13 CST" "2002-01-08 01:44:45 CST"
## [165] "2002-10-05 11:28:06 CDT" "2002-10-23 00:33:43 CDT"
## [167] "2003-07-17 21:34:55 CDT" "2003-09-03 05:20:15 CDT"
## [169] "2003-09-19 03:02:12 CDT" "2004-02-09 02:42:00 CST"
## [171] "2004-08-12 04:50:20 CDT" "2004-09-21 20:54:17 CDT"
## [173] "2000-01-24 06:50:13 CST" "2000-08-25 18:26:50 CDT"
## [175] "2001-05-28 23:31:00 CDT" "2001-10-20 03:55:05 CDT"
## [177] "2002-06-07 08:36:45 CDT" "2002-08-09 22:04:50 CDT"
## [179] "2003-02-16 02:56:16 CST" "2003-08-31 05:48:29 CDT"
## [181] "2004-02-12 03:52:22 CST" "2004-04-18 20:04:11 CDT"
## [183] "2001-04-27 22:31:51 CDT" "2001-06-09 20:52:51 CDT"
## [185] "2001-06-30 22:23:17 CDT" "2002-02-28 21:09:16 CST"
## [187] "2002-06-09 16:19:44 CDT" "2002-12-21 13:33:18 CST"
## [189] "2003-04-23 00:45:04 CDT" "2003-08-01 11:53:31 CDT"
## [191] "2004-09-14 07:04:17 CDT" "2000-03-12 10:37:10 CST"
## [193] "2000-09-21 09:46:42 CDT" "2001-11-18 01:08:47 CST"
## [195] "2002-04-19 23:53:01 CDT" "2002-08-26 16:18:26 CDT"
## [197] "2004-03-02 05:00:01 CST" "2004-07-02 04:13:55 CDT"
## [199] "2000-11-08 08:23:23 CST" "2001-08-18 12:28:05 CDT"
## [201] "2001-12-27 07:43:35 CST" "2002-03-28 10:35:47 CST"
## [203] "2002-08-16 04:08:42 CDT" "2002-10-28 22:49:34 CST"
## [205] "2003-07-20 02:00:00 CDT" "2003-08-23 19:26:16 CDT"
## [207] "2004-01-22 13:53:37 CST" "2004-03-28 03:37:59 CST"
## [209] "2000-01-23 07:00:42 CST" "2001-09-14 15:50:30 CDT"
## [211] "2002-11-08 03:58:11 CST" "2003-05-23 03:41:36 CDT"
## [213] "2003-10-18 22:34:28 CDT" "2004-04-29 13:06:43 CDT"
## [215] "2001-04-03 23:39:12 CDT" "2002-08-21 03:37:15 CDT"
## [217] "2003-09-27 01:30:21 CDT" "2003-11-24 11:15:58 CST"
## [219] "2004-02-01 20:33:43 CST" "2004-12-27 22:04:36 CST"
## [221] "2000-01-08 04:21:03 CST" "2000-01-20 19:57:44 CST"
## [223] "2000-03-08 09:41:19 CST" "2000-06-07 02:40:56 CDT"
## [225] "2001-02-24 01:19:03 CST" "2001-08-18 15:37:01 CDT"
## [227] "2001-09-13 04:15:49 CDT" "2001-10-06 00:28:39 CDT"
## [229] "2002-08-27 15:45:53 CDT" "2002-09-06 13:49:49 CDT"
## [231] "2003-02-05 07:15:50 CST" "2003-06-19 03:49:18 CDT"
## [233] "2003-10-18 16:37:48 CDT" "2004-02-22 13:57:40 CST"
## [235] "2004-08-23 16:33:53 CDT" "2001-01-15 12:45:56 CST"
## [237] "2001-01-28 17:55:14 CST" "2001-10-01 20:37:01 CDT"
## [239] "2001-11-26 20:10:09 CST" "2002-01-12 07:00:38 CST"
## [241] "2002-05-17 12:16:11 CDT" "2002-12-13 01:15:52 CST"
## [243] "2003-08-20 21:07:16 CDT" "2004-02-03 12:15:31 CST"
## [245] "2004-02-14 19:17:31 CST" "2004-05-14 00:55:04 CDT"
## [247] "2004-09-11 11:04:49 CDT" "2000-06-06 16:08:21 CDT"
## [249] "2000-12-12 13:35:29 CST" "2001-01-12 03:01:24 CST"
## [251] "2001-10-01 10:24:11 CDT" "2001-10-03 21:52:39 CDT"
## [253] "2001-10-18 09:24:13 CDT" "2001-12-25 13:27:11 CST"
## [255] "2002-05-04 09:28:37 CDT" "2002-06-11 02:09:20 CDT"
## [257] "2003-01-05 23:46:39 CST" "2003-03-25 05:24:26 CST"
## [259] "2004-09-08 15:28:21 CDT" "2004-09-16 18:27:03 CDT"
## [261] "2000-01-14 04:36:22 CST" "2000-05-01 00:13:49 CDT"
## [263] "2000-09-03 09:52:52 CDT" "2000-12-23 16:49:53 CST"
```

```
## [265] "2001-08-18 08:50:34 CDT" "2002-04-25 16:27:51 CDT"
## [267] "2002-06-21 00:37:36 CDT" "2004-01-17 12:53:28 CST"
## [269] "2004-07-13 01:57:50 CDT" "2004-07-20 04:24:26 CDT"
## [271] "2000-04-29 22:38:23 CDT" "2000-11-20 00:14:40 CST"
## [273] "2000-12-21 04:51:28 CST" "2001-01-13 02:55:28 CST"
## [275] "2001-04-18 11:08:19 CDT" "2001-04-24 01:52:55 CDT"
## [277] "2001-07-13 13:53:47 CDT" "2001-09-12 17:44:54 CDT"
## [279] "2002-02-20 06:17:03 CST" "2003-04-09 00:42:42 CDT"
## [281] "2003-09-10 23:00:00 CDT" "2003-10-12 16:06:55 CDT"
## [283] "2004-09-10 16:41:01 CDT" "2000-06-19 16:06:44 CDT"
## [285] "2000-12-28 23:09:04 CST" "2001-05-27 18:12:12 CDT"
## [287] "2002-06-11 19:26:36 CDT" "2002-08-02 09:28:33 CDT"
## [289] "2003-04-20 22:12:56 CDT" "2003-07-06 13:52:12 CDT"
## [291] "2003-07-22 18:07:53 CDT" "2003-08-05 03:47:31 CDT"
## [293] "2000-09-21 17:48:13 CDT" "2001-01-26 13:49:01 CST"
## [295] "2002-12-24 11:29:36 CST" "2003-04-28 13:33:01 CDT"
## [297] "2003-12-30 04:30:14 CST" "2004-04-17 21:50:21 CDT"
## [299] "2001-08-27 18:17:25 CDT" "2002-03-03 22:03:35 CST"
## [301] "2002-04-07 23:29:50 CDT" "2002-04-13 18:29:18 CDT"
## [303] "2002-09-01 16:24:15 CDT" "2002-09-12 20:02:02 CDT"
## [305] "2002-10-25 03:27:37 CDT" "2003-05-03 17:36:01 CDT"
## [307] "2003-05-12 18:12:12 CDT" "2003-09-29 12:05:29 CDT"
## [309] "2004-01-21 05:14:22 CST" "2001-05-10 09:00:03 CDT"
## [311] "2002-06-10 04:06:16 CDT" "2003-04-10 21:44:17 CDT"
## [313] "2004-02-18 14:14:22 CST" "2001-03-01 00:38:47 CST"
## [315] "2002-02-26 11:18:01 CST" "2002-10-31 20:47:50 CST"
## [317] "2003-02-28 20:22:59 CST" "2003-06-04 08:34:14 CDT"
## [319] "2003-11-01 15:51:13 CST" "2004-11-08 18:27:18 CST"
## [321] "2000-03-10 06:22:47 CST" "2000-06-10 21:05:34 CDT"
## [323] "2001-01-02 23:11:28 CST" "2001-06-20 13:56:46 CDT"
## [325] "2001-08-16 00:33:00 CDT" "2002-03-14 03:08:23 CST"
## [327] "2003-05-22 09:29:38 CDT" "2004-07-10 17:28:02 CDT"
## [329] "2004-08-03 12:29:29 CDT" "2004-11-28 02:47:14 CST"
## [331] "2000-08-25 03:07:22 CDT" "2000-12-15 14:15:33 CST"
## [333] "2001-08-14 05:04:57 CDT" "2002-04-04 18:16:28 CST"
## [335] "2002-04-08 01:03:54 CDT" "2002-04-14 22:17:24 CDT"
## [337] "2003-06-12 11:35:17 CDT" "2004-04-01 15:06:38 CST"
## [339] "2004-10-10 23:40:20 CDT" "2004-11-24 23:37:58 CST"
## [341] "2002-02-10 18:55:33 CST" "2003-03-04 18:56:56 CST"
## [343] "2003-04-22 08:50:58 CDT" "2003-04-24 10:21:04 CDT"
## [345] "2003-09-22 15:34:24 CDT" "2003-11-26 08:13:01 CST"
## [347] "2003-12-10 13:11:04 CST" "2004-06-20 13:22:32 CDT"
## [349] "2000-01-04 10:31:04 CST" "2000-02-09 13:14:40 CST"
## [351] "2000-03-02 10:10:39 CST" "2000-03-15 00:29:47 CST"
## [353] "2000-06-16 22:41:07 CDT" "2000-08-17 03:58:50 CDT"
## [355] "2000-11-20 15:56:15 CST" "2001-01-29 17:35:03 CST"
## [357] "2002-06-29 21:02:17 CDT" "2002-08-23 08:49:51 CDT"
## [359] "2003-02-08 13:21:06 CST" "2003-02-15 09:48:43 CST"
## [361] "2004-07-20 02:29:53 CDT" "2004-10-22 22:31:57 CDT"
## [363] "2004-12-12 09:52:54 CST" "2000-03-01 21:52:46 CST"
## [365] "2000-08-17 09:04:23 CDT" "2001-02-01 09:15:56 CST"
## [367] "2001-03-16 03:11:29 CST" "2001-03-25 15:48:47 CST"
## [369] "2001-07-02 09:49:23 CDT" "2001-10-26 14:06:20 CDT"
## [371] "2003-02-03 01:25:43 CST" "2003-03-08 08:13:04 CST"
```

```
## [373] "2003-05-06 07:54:29 CDT" "2003-09-08 18:37:25 CDT"
## [375] "2003-10-11 11:39:25 CDT" "2003-12-25 20:38:23 CST"
## [377] "2004-11-15 12:40:59 CST" "2000-12-14 15:17:08 CST"
## [379] "2001-02-13 10:26:45 CST" "2001-06-11 20:44:02 CDT"
## [381] "2002-08-25 09:05:13 CDT" "2003-10-16 20:41:19 CDT"
## [383] "2004-06-20 18:58:23 CDT" "2000-09-09 05:29:37 CDT"
## [385] "2001-01-06 18:12:24 CST" "2001-05-29 00:16:18 CDT"
## [387] "2001-10-18 16:38:27 CDT" "2002-10-01 04:04:32 CDT"
## [389] "2003-11-19 05:52:33 CST" "2000-03-11 11:02:13 CST"
## [391] "2000-07-31 18:23:24 CDT" "2000-08-27 13:05:42 CDT"
## [393] "2001-01-06 08:23:48 CST" "2001-01-26 11:12:23 CST"
## [395] "2001-07-27 11:36:03 CDT" "2001-10-13 15:40:21 CDT"
## [397] "2001-12-21 15:05:20 CST" "2002-01-01 17:23:23 CST"
## [399] "2002-01-07 14:40:00 CST" "2002-08-03 11:36:19 CDT"
## [401] "2002-12-17 21:54:24 CST" "2003-01-20 15:07:51 CST"
## [403] "2003-07-14 14:50:41 CDT" "2004-01-22 11:52:51 CST"
## [405] "2004-03-21 04:34:34 CST" "2004-04-25 10:17:23 CDT"
## [407] "2000-01-24 20:17:31 CST" "2000-06-03 06:28:31 CDT"
## [409] "2001-03-12 16:31:46 CST" "2001-03-29 19:15:23 CST"
## [411] "2001-04-23 11:29:26 CDT" "2001-04-30 21:21:00 CDT"
## [413] "2001-08-21 15:52:03 CDT" "2002-02-27 16:43:27 CST"
## [415] "2002-07-12 09:55:18 CDT" "2002-10-04 03:25:33 CDT"
## [417] "2003-11-02 23:55:16 CST" "2003-12-22 20:16:02 CST"
## [419] "2004-02-27 18:44:23 CST" "2000-02-07 07:04:09 CST"
## [421] "2001-06-30 18:39:38 CDT" "2001-08-12 05:28:00 CDT"
## [423] "2001-08-13 08:38:36 CDT" "2001-10-29 13:43:11 CST"
## [425] "2003-06-18 04:12:25 CDT" "2004-02-15 03:55:53 CST"
## [427] "2004-04-19 15:21:31 CDT" "2004-07-24 14:17:10 CDT"
## [429] "2000-03-02 22:23:01 CST" "2000-05-13 02:50:47 CDT"
## [431] "2001-03-10 20:44:09 CST" "2001-04-15 09:31:09 CDT"
## [433] "2001-12-30 13:45:27 CST" "2003-05-01 06:47:31 CDT"
## [435] "2003-08-12 04:08:05 CDT" "2003-09-09 11:39:05 CDT"
## [437] "2004-10-14 05:28:02 CDT" "2000-01-24 02:24:39 CST"
## [439] "2000-07-03 17:31:48 CDT" "2001-02-20 13:31:18 CST"
## [441] "2001-03-31 12:17:12 CST" "2001-09-20 22:52:26 CDT"
## [443] "2001-09-24 08:02:44 CDT" "2002-05-10 09:26:15 CDT"
## [445] "2003-05-15 19:37:51 CDT" "2003-09-24 02:54:24 CDT"
## [447] "2003-12-25 09:59:41 CST" "2004-06-17 12:12:21 CDT"
## [449] "2004-07-21 18:09:35 CDT" "2004-12-21 08:14:11 CST"
## [451] "2000-05-21 06:23:36 CDT" "2000-10-27 22:02:36 CDT"
## [453] "2001-06-22 04:49:32 CDT" "2001-07-17 12:29:17 CDT"
## [455] "2002-08-29 06:50:10 CDT" "2002-09-20 15:41:04 CDT"
## [457] "2003-01-12 03:47:29 CST" "2003-03-03 16:14:08 CST"
## [459] "2000-03-20 13:18:02 CST" "2000-04-20 17:21:57 CDT"
## [461] "2001-03-11 15:40:18 CST" "2001-06-26 11:26:07 CDT"
## [463] "2001-10-20 06:55:15 CDT" "2002-02-09 13:46:35 CST"
## [465] "2003-02-06 08:18:32 CST" "2003-06-08 15:15:39 CDT"
## [467] "2003-09-05 23:10:03 CDT" "2004-07-24 09:41:27 CDT"
## [469] "2004-10-19 12:22:55 CDT" "2004-11-22 15:27:41 CST"
## [471] "2000-08-04 01:39:44 CDT" "2001-08-17 23:45:51 CDT"
## [473] "2002-02-20 14:30:26 CST" "2002-03-20 23:41:33 CST"
## [475] "2003-07-18 22:13:39 CDT" "2003-12-10 02:15:58 CST"
## [477] "2003-12-31 21:15:31 CST" "2004-01-21 15:54:42 CST"
## [479] "2004-08-28 19:41:51 CDT" "2000-03-19 18:28:14 CST"
```

```
## [481] "2000-06-21 02:44:02 CDT" "2002-01-26 23:46:45 CST"
## [483] "2002-11-20 08:26:22 CST" "2003-02-26 08:44:05 CST"
## [485] "2003-06-16 21:34:48 CDT" "2003-08-03 01:57:16 CDT"
## [487] "2003-11-10 09:20:14 CST" "2003-12-27 21:48:49 CST"
## [489] "2004-10-19 09:31:47 CDT" "2004-12-16 14:02:13 CST"
## [491] "2001-02-10 12:14:19 CST" "2001-07-02 03:46:14 CDT"
## [493] "2001-07-31 22:54:57 CDT" "2001-12-03 12:00:58 CST"
## [495] "2002-07-29 17:13:57 CDT" "2002-10-14 15:02:07 CDT"
## [497] "2003-01-27 21:06:01 CST" "2003-03-11 20:06:00 CST"
## [499] "2003-11-22 16:07:03 CST" "2004-05-28 06:02:54 CDT"
##
## $a1c
##   [1]  5.452057  4.609312  4.219065  2.234179  2.748069  4.904277  4.343939
##   [8]  4.981675  3.589598  5.159701  2.455222  7.737220  7.453772  8.645883
##  [15]  7.579143  7.444163  6.089039  8.986401  7.138702  8.048449  6.789743
##  [22]  7.338620  7.233755  7.223148  7.692685  7.428381  7.545006  7.130214
##  [29]  8.859482  6.933565  7.595479  5.458973  7.781191  6.966075  6.408238
##  [36]  7.393190  6.726370  5.537395  8.179432  8.274279  5.168455  7.276415
##  [43]  5.894145  6.785159  3.566135  2.955887  5.759313  5.989879  3.148383
##  [50]  4.085015  4.855899  2.362547  3.485858  2.712358  9.110934  9.649331
##  [57] 10.586352  9.821794 10.116674  9.841045  8.681195  8.700747  8.692254
##  [64]  9.829726  8.967686 10.149990  6.624953  7.387311  7.171313  8.478958
##  [71]  6.857083  7.756622  7.536041  8.635439  9.880476  8.787287 10.400148
##  [78]  9.388096  8.568382  9.915630  7.816330  8.537136  7.518224  7.151946
##  [85]  6.224875  8.328739  6.555329  7.459567  6.190650  6.830623  7.382025
##  [92]  8.054384  7.383348  8.046467  9.189446  9.263242  8.527757 10.468543
##  [99] 10.473206  9.112783  9.652115 11.095655  7.957322  8.336069  8.005756
## [106]  6.519573  8.843217  7.598683  9.137850  7.890126  6.784908  9.293014
## [113]  6.126086  7.796419  9.731754  8.770823  6.807899  5.936558  5.745135
## [120]  6.795538  6.461254  6.777114  8.046176  7.770349  6.857469  8.414811
## [127]  6.905767  7.517219  6.613855  4.556653  7.636045  9.685656  9.467547
## [134]  8.810552  8.229774  9.241735  9.436427  5.497147  5.919141  6.751341
## [141]  7.530131  6.625343  7.163004  6.440273  7.435674  6.608967  6.980222
## [148]  5.910078  7.527105  5.898371  8.566593  9.133769  8.936190  3.702627
## [155]  5.857281  3.447221  5.948422  4.632084  2.774031  3.193438  4.695063
## [162]  4.839526  3.830901  5.285508  4.054559  4.483293  2.671586  4.811637
## [169]  3.496204  2.095877  4.847168  2.841083  8.583458  9.972658  8.069042
## [176] 11.311360 10.669313  7.771634  8.046465  9.401234  8.657878  9.165691
## [183]  6.048896  5.595630  6.263994  4.731299  4.401713  4.695538  6.093509
## [190]  6.129478  5.604833  2.511383  3.307914  4.224824  4.929644  3.514134
## [197]  3.815193  3.783632  8.415140  7.000966  7.514169  7.952595  8.946419
## [204]  7.932334  7.318638  8.992822  8.337196  8.999109  5.609340  5.519465
## [211]  6.322255  5.498381  7.048272  3.827296  6.971693  6.330971  8.809112
## [218]  8.018845  6.990753  7.079960  7.648908  8.006692  7.775008  5.544651
## [225]  9.474541  7.477594 10.116559  5.965407  6.659937  7.482505  7.594292
## [232]  6.117872  7.367876  8.068536  6.289358  6.810806  7.405325  6.646405
## [239]  8.097374  6.434536  5.421498  5.726077  7.143020  6.357675  8.222245
## [246]  6.279511  7.981147  6.674702  5.082672  4.583922  5.144310  7.282685
## [253]  6.206772  8.283471  8.100256  4.987150  7.708757  7.255256  8.063651
## [260]  5.864257  5.188675  5.441924  4.639696  4.973571  4.053654  4.706853
## [267]  5.124321  4.316378  4.860044  5.959458  7.574393  8.577672  8.067329
## [274]  6.712182  5.725540  7.714023  7.774396  6.702442  9.126001  8.015406
## [281]  6.216750  6.834094  7.600696  4.273429  3.662980  4.477118  5.535907
## [288]  4.782558  5.190293  3.128745  5.869510  3.652233  6.354781  5.845323
```

```
## [295]  5.324100  5.581342  7.336617  5.831301  7.833619  5.560845  8.359121
## [302]  8.226234  7.843916  6.673637  5.651674  5.879353  7.188015  7.577177
## [309]  7.488850  7.422034  5.338268  6.042607  7.472256  6.258884  7.714164
## [316]  4.629184  6.273297  6.103795  7.582219  6.896945  6.672449  7.135051
## [323]  6.255340  6.123402  6.999220  7.817720  6.818496  6.718016  5.742919
## [330]  7.403538  8.053235  8.563920  8.101484  9.280775  6.825296  8.040744
## [337]  9.589349  9.178030  7.986970  9.147193 10.088599  9.409205  8.990238
## [344] 10.283910  9.446568  9.155634  9.537207  9.923918  7.674761 11.049303
## [351] 10.940435  8.207832  8.773848  9.625890 10.380595  8.622952  9.650118
## [358]  8.471523 10.528959  9.644036 10.145173 10.055101 10.323264  6.024135
## [365]  6.512096  5.101166  4.706317  4.739199  6.137582  5.343153  3.814121
## [372]  5.840423  6.734259  5.312007  5.529102  5.309972  3.880181  5.993376
## [379]  6.482377  7.407109  7.511896  7.003631  7.103687  9.374750  9.536771
## [386]  8.961818 10.246297  9.131978  9.929203  8.857702 11.297718  8.930901
## [393]  8.948475 11.579820  9.819376  9.147303 10.291154 11.320047 10.577416
## [400]  9.429560 10.987847  9.400568  9.736937  9.383457  8.792973  8.139962
## [407]  4.374044  3.892797  3.224010  4.497949  2.519393  3.611041  3.432115
## [414]  5.652292  3.182028  4.290907  4.204516  3.052397  4.659514  7.194290
## [421]  6.207837  3.688395  6.503561  6.229327  4.926351  5.836864  6.951236
## [428]  7.324779 10.662976  9.276947  9.730451  7.689378  8.211858 10.762428
## [435]  9.359516  7.141636  8.989838  6.623787  5.482196  6.095120  7.575338
## [442]  5.938463  8.035306  6.083796  6.095493  6.536981  7.018600  7.548331
## [449]  7.307657  7.242591  8.638051  9.998967 10.731243  9.319496  7.896552
## [456] 10.189552 10.503969  9.696275  9.509746  9.984591 10.927428  8.052130
## [463] 10.420549 10.972189  8.216927  8.376289  8.548913  9.776965  7.039840
## [470]  8.952297  6.036881  5.733356  5.489845  7.478090  5.054100  7.833596
## [477]  7.653118  6.869489  5.492926  6.090393  6.704109  6.930944  5.607635
## [484]  5.962446  5.036806  5.574253  6.747143  7.020386  5.158865  6.223859
## [491]  8.636432  8.688873  9.066142  8.586003  9.245883  9.626216  9.234505
## [498] 10.263365  7.925987  8.349782
```

```
order(z$id)
```

```
##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
##  [18]  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34
##  [35]  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51
##  [52]  52  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68
##  [69]  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85
##  [86]  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 101 102
## [103] 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119
## [120] 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
## [137] 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153
## [154] 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
## [171] 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187
## [188] 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204
## [205] 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221
## [222] 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238
## [239] 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255
## [256] 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272
## [273] 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289
## [290] 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306
## [307] 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323
## [324] 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
## [341] 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357
## [358] 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374
```

```
## [375] 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
## [392] 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408
## [409] 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425
## [426] 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442
## [443] 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459
## [460] 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476
## [477] 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493
## [494] 494 495 496 497 498 499 500
```

```
a <- table(y$id)
a[[1]]
```

```
## [1] 11
```

```
for(i in seq(length(z$dt))) {
  if(i == 500) break
  difference <- difftime(z$dt[i+1],z$dt[i],units='days')
  if(difference >= 0) {
    if(difference >= 365) {
      newy <- rep(z$dt[i]+365*trunc(difference/365), trunc(difference/365))
      z$dt <- append(z$dt, newy, after=i)
      newyid <- rep(z$id[i],trunc(difference/365))
      z$id <- append(z$id, newyid, after=i)
      newya1c <- rep(NA, trunc(difference/365))
      z$a1c <- append(z$a1c,newya1c, after=i)
    }
  }
}

a <- do.call(cbind.data.frame,z)
```

Part 3) Create a new column visit. For each id, add the visit number. This should be 1 to n where n is the number of observations for an individual. This should include the observations created with missing a1c values.

```
b <- a[,'id']
visit <- sapply(1:length(b), function(x) sum(b[1:x] == b[x]))
c <- cbind(a, visit)
```

Part 4) For each id, replace missing values with the mean a1c value for that individual.

```
for(i in 1:ncol(c)){
  c[is.na(c[,i]), i] <- mean(c[,i], na.rm = TRUE)
}
```

JC Grading -3 The means calculated above are column means, rather than patient-specific means.

Part 5) Print mean a1c for each id.

```
means <- tapply(c[,'a1c'],c['id'],mean)
means
```

```
## id
##        1        2        3        4        5        6        7        8
## 4.063372 7.544643 6.784320 4.431966 9.172146 7.471004 8.936071 7.184239
##        9       10       11       12       13       14       15       16
## 9.283873 7.904879 6.960281 7.053448 8.641751 6.666040 7.681938 4.868602
##       17       18       19       20       21       22       23       24
## 3.996034 9.164873 5.669605 4.483227 8.140939 6.010916 7.307956 7.439316
```

```
##       25       26       27       28       29       30       31       32
## 6.877135 6.597787 5.126885 7.412292 4.770393 6.200660 7.116586 6.681265
##       33       34       35       36       37       38       39       40
## 6.573706 6.829039 8.354378 9.329604 9.315065 5.474325 6.970549 9.187425
##       41       42       43       44       45       46       47       48
## 9.802424 4.123154 6.284087 8.735213 6.765344 9.621763 9.231489 6.404600
##       49       50
## 6.096076 8.962319
```

Part 6) Print total number of visits for each id.

```
totalvisits <- tapply(c[,'visit'],c[,'id'],length)
totalvisits
```

```
##  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
## 11 20 14 12 14 10  9 12 11 12 10 10  8 12  8  9 12 10 10  9 10  8  8 15 12
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
## 14 11 14 10  7 11  5  8 12 11  9 17 15  8  7 17 14 11 11 14  8 12  9 11 10
```

Part 7) Print the observations for id = 15.

```
c[which(c[,'id'] == 15),]
```

```
##     id                  dt      a1c visit
## 166 15 2000-04-30 00:34:50 7.527105     1
## 167 15 2001-01-17 21:11:02 5.898371     2
## 168 15 2001-04-25 06:23:05 8.566593     3
## 169 15 2003-04-25 06:23:05 7.131159     4
## 170 15 2003-04-25 06:23:05 7.131159     5
## 171 15 2003-06-06 14:06:00 9.133769     6
## 172 15 2004-06-05 14:06:00 7.131159     7
## 173 15 2004-08-20 17:47:11 8.936190     8
```

3

```
addr = readLines("addr.txt")
addr = as.data.frame(do.call(rbind, strsplit(addr, split=" {2,10}")))
stre <- sub(" ", "~",addr[,3])
stre
```

```
##  [1] "725~Commonwealth Ave."    "373~W. Geneva St."
##  [3] "373~W. Geneva St."        "725~Commonwealth Ave."
##  [5] "933~E. 56th St."          "111~Nowelo St."
##  [7] "2145~Sheridan Rd"         "933~E. 56th St."
##  [9] "174~W. 18th Ave."         "5000~Forbes Ave."
## [11] "933~E. 56th St."          "373~W. Geneva St."
## [13] "725~W. Commonwealth Ave." "725~W. Commonwealth Ave."
## [15] "725~W. Commonwealth Ave." "373~W. Geneva St."
## [17] "5640~S. Ellis Ave."       "5640~S. Ellis Ave."
## [19] "1002~W. Green St."        "373~W. Geneva St."
## [21] "4201~Wilson Blvd"         "174~W. 18th Ave."
## [23] "933~E. 56th St."          "373~W. Geneva St."
## [25] "5000~Forbes Ave."         "2145~Sheridan Rd"
## [27] "373~W. Geneva St."        "373~W. Geneva St."
## [29] "373~W. Geneva St."        "5000~Forbes Ave."
```

```
## [31] "933~E. 56th St."          "5640~S. Ellis Ave."
## [33] "2145~Sheridan Rd"         "8730~W. Mountain View Ln"
## [35] "933~E. 56th St."          "373~W. Geneva St."
## [37] "373~W. Geneva St."        "933~N. Cherry St."
## [39] "5000~Forbes Ave."         "373~W. Geneva St."
## [41] "791~Holmdel-Keyport Rd."  "5640~S. Ellis Ave."
```

```
street2 <- as.data.frame(do.call(rbind, strsplit(stre, split='~')))
street2
```

```
##      V1                  V2
## 1   725    Commonwealth Ave.
## 2   373        W. Geneva St.
## 3   373        W. Geneva St.
## 4   725    Commonwealth Ave.
## 5   933          E. 56th St.
## 6   111           Nowelo St.
## 7  2145          Sheridan Rd
## 8   933          E. 56th St.
## 9   174         W. 18th Ave.
## 10 5000          Forbes Ave.
## 11  933          E. 56th St.
## 12  373        W. Geneva St.
## 13  725 W. Commonwealth Ave.
## 14  725 W. Commonwealth Ave.
## 15  725 W. Commonwealth Ave.
## 16  373        W. Geneva St.
## 17 5640        S. Ellis Ave.
## 18 5640        S. Ellis Ave.
## 19 1002         W. Green St.
## 20  373        W. Geneva St.
## 21 4201          Wilson Blvd
## 22  174         W. 18th Ave.
## 23  933          E. 56th St.
## 24  373        W. Geneva St.
## 25 5000          Forbes Ave.
## 26 2145          Sheridan Rd
## 27  373        W. Geneva St.
## 28  373        W. Geneva St.
## 29  373        W. Geneva St.
## 30 5000          Forbes Ave.
## 31  933          E. 56th St.
## 32 5640        S. Ellis Ave.
## 33 2145          Sheridan Rd
## 34 8730  W. Mountain View Ln
## 35  933          E. 56th St.
## 36  373        W. Geneva St.
## 37  373        W. Geneva St.
## 38  933        N. Cherry St.
## 39 5000          Forbes Ave.
## 40  373        W. Geneva St.
## 41  791  Holmdel-Keyport Rd.
## 42 5640        S. Ellis Ave.
```

```
addr2 <- cbind(addr[,1:2], street2, addr[,4:6])
addr2
```

```
##             V1         V2   V1                   V2         V4 V5
## 1        Bania  Thomas M.  725    Commonwealth Ave.     Boston MA
## 2      Barnaby      David  373        W. Geneva St.   Wms. Bay WI
## 3       Bausch       Judy  373        W. Geneva St.   Wms. Bay WI
## 4      Bolatto    Alberto  725    Commonwealth Ave.     Boston MA
## 5    Carlstrom       John  933          E. 56th St.    Chicago IL
## 6   Chamberlin Richard A.  111           Nowelo St.       Hilo HI
## 7        Chuss       Dave 2145          Sheridan Rd   Evanston IL
## 8        Davis      E. J.  933          E. 56th St.    Chicago IL
## 9        Depoy     Darren  174         W. 18th Ave.   Columbus OH
## 10     Griffin       Greg 5000          Forbes Ave. Pittsburgh PA
## 11   Halvorsen       Nils  933          E. 56th St.    Chicago IL
## 12     Harper         Al  373        W. Geneva St.   Wms. Bay WI
## 13       Huang     Maohai  725 W. Commonwealth Ave.     Boston MA
## 14     Ingalls   James G.  725 W. Commonwealth Ave.     Boston MA
## 15     Jackson   James M.  725 W. Commonwealth Ave.     Boston MA
## 16     Knudsen      Scott  373        W. Geneva St.   Wms. Bay WI
## 17       Kovac       John 5640        S. Ellis Ave.    Chicago IL
## 18   Landsberg      Randy 5640        S. Ellis Ave.    Chicago IL
## 19          Lo  Kwok-Yung 1002         W. Green St.     Urbana IL
## 20 Loewenstein  Robert F.  373        W. Geneva St.   Wms. Bay WI
## 21       Lynch       John 4201          Wilson Blvd  Arlington VA
## 22     Martini       Paul  174         W. 18th Ave.   Columbus OH
## 23       Meyer    Stephan  933          E. 56th St.    Chicago IL
## 24      Mrozek       Fred  373        W. Geneva St.   Wms. Bay WI
## 25     Newcomb       Matt 5000          Forbes Ave. Pittsburgh PA
## 26       Novak      Giles 2145          Sheridan Rd   Evanston IL
## 27      Odalen      Nancy  373        W. Geneva St.   Wms. Bay WI
## 28      Pernic       Dave  373        W. Geneva St.   Wms. Bay WI
## 29      Pernic        Bob  373        W. Geneva St.   Wms. Bay WI
## 30    Peterson    Jeffrey 5000          Forbes Ave. Pittsburgh PA
## 31       Pryke       Clem  933          E. 56th St.    Chicago IL
## 32      Rebull      Luisa 5640        S. Ellis Ave.    Chicago IL
## 33   Renbarger     Thomas 2145          Sheridan Rd   Evanston IL
## 34     Rottman        Joe 8730  W. Mountain View Ln  Littleton CO
## 35   Schartman      Ethan  933          E. 56th St.    Chicago IL
## 36       Spotz        Bob  373        W. Geneva St.   Wms. Bay WI
## 37       Thoma       Mark  373        W. Geneva St.   Wms. Bay WI
## 38      Walker      Chris  933        N. Cherry St.     Tucson AZ
## 39      Wehrer     Cheryl 5000          Forbes Ave. Pittsburgh PA
## 40       Wirth      Jesse  373        W. Geneva St.   Wms. Bay WI
## 41      Wright       Greg  791  Holmdel-Keyport Rd.    Holmdel NY
## 42     Zingale    Michael 5640        S. Ellis Ave.    Chicago IL
##             V6
## 1        02215
## 2        53191
## 3        53191
## 4        02215
## 5        60637
## 6        96720
## 7  60208-3112
```

```
## 8        60637
## 9        43210
## 10       15213
## 11       60637
## 12       53191
## 13      O2215
## 14      O2215
## 15      O2215
## 16       53191
## 17       60637
## 18       60637
## 19       61801
## 20       53191
## 21       22230
## 22       43210
## 23       60637
## 24       53191
## 25       15213
## 26 60208-3112
## 27       53191
## 28       53191
## 29       53191
## 30       15213
## 31       60637
## 32       60637
## 33 60208-3112
## 34       80125
## 35       60637
## 36       53191
## 37       53191
## 38       85721
## 39       15213
## 40       53191
## 41 O7733-1988
## 42       60637
```

```
colnames(addr2) <- c('lastname','firstname','streetno','streetname','city','state','zip')
```

Almost done. Notice in 'zip' col. that O's should be zeroes. Let's fix that really quickly.

```
addr2[,'zip'] <- as.character(addr2[,'zip'])
addr2[,'zip'] <- sub("O","0",addr2[,'zip'])
addr2
```

```
##       lastname  firstname streetno          streetname       city state
## 1        Bania  Thomas M.      725    Commonwealth Ave.     Boston    MA
## 2      Barnaby      David      373        W. Geneva St.   Wms. Bay    WI
## 3       Bausch       Judy      373        W. Geneva St.   Wms. Bay    WI
## 4      Bolatto    Alberto      725    Commonwealth Ave.     Boston    MA
## 5    Carlstrom       John      933          E. 56th St.    Chicago    IL
## 6   Chamberlin Richard A.      111           Nowelo St.       Hilo    HI
## 7        Chuss       Dave     2145          Sheridan Rd   Evanston    IL
## 8        Davis      E. J.      933          E. 56th St.    Chicago    IL
## 9        Depoy     Darren      174         W. 18th Ave.   Columbus    OH
## 10     Griffin       Greg     5000          Forbes Ave. Pittsburgh    PA
## 11   Halvorsen       Nils      933          E. 56th St.    Chicago    IL
```

```
## 12      Harper         Al      373        W. Geneva St.   Wms. Bay    WI
## 13       Huang     Maohai      725 W. Commonwealth Ave.     Boston    MA
## 14     Ingalls   James G.      725 W. Commonwealth Ave.     Boston    MA
## 15     Jackson   James M.      725 W. Commonwealth Ave.     Boston    MA
## 16     Knudsen      Scott      373        W. Geneva St.   Wms. Bay    WI
## 17       Kovac       John     5640        S. Ellis Ave.    Chicago    IL
## 18   Landsberg      Randy     5640        S. Ellis Ave.    Chicago    IL
## 19          Lo  Kwok-Yung     1002         W. Green St.     Urbana    IL
## 20 Loewenstein  Robert F.      373        W. Geneva St.   Wms. Bay    WI
## 21       Lynch       John     4201          Wilson Blvd  Arlington    VA
## 22     Martini       Paul      174         W. 18th Ave.   Columbus    OH
## 23       Meyer    Stephan      933          E. 56th St.    Chicago    IL
## 24      Mrozek       Fred      373        W. Geneva St.   Wms. Bay    WI
## 25     Newcomb       Matt     5000          Forbes Ave. Pittsburgh    PA
## 26       Novak      Giles     2145          Sheridan Rd   Evanston    IL
## 27      Odalen      Nancy      373        W. Geneva St.   Wms. Bay    WI
## 28      Pernic       Dave      373        W. Geneva St.   Wms. Bay    WI
## 29      Pernic        Bob      373        W. Geneva St.   Wms. Bay    WI
## 30    Peterson    Jeffrey     5000          Forbes Ave. Pittsburgh    PA
## 31       Pryke       Clem      933          E. 56th St.    Chicago    IL
## 32      Rebull      Luisa     5640        S. Ellis Ave.    Chicago    IL
## 33   Renbarger     Thomas     2145          Sheridan Rd   Evanston    IL
## 34     Rottman        Joe     8730  W. Mountain View Ln  Littleton    CO
## 35   Schartman      Ethan      933          E. 56th St.    Chicago    IL
## 36       Spotz        Bob      373        W. Geneva St.   Wms. Bay    WI
## 37       Thoma       Mark      373        W. Geneva St.   Wms. Bay    WI
## 38      Walker      Chris      933        N. Cherry St.     Tucson    AZ
## 39      Wehrer     Cheryl     5000          Forbes Ave. Pittsburgh    PA
## 40       Wirth      Jesse      373        W. Geneva St.   Wms. Bay    WI
## 41      Wright       Greg      791  Holmdel-Keyport Rd.    Holmdel    NY
## 42     Zingale    Michael     5640        S. Ellis Ave.    Chicago    IL
##            zip
## 1        02215
## 2        53191
## 3        53191
## 4        02215
## 5        60637
## 6        96720
## 7   60208-3112
## 8        60637
## 9        43210
## 10       15213
## 11       60637
## 12       53191
## 13       02215
## 14       02215
## 15       02215
## 16       53191
## 17       60637
## 18       60637
## 19       61801
## 20       53191
## 21       22230
## 22       43210
```

```
## 23       60637
## 24       53191
## 25       15213
## 26 60208-3112
## 27       53191
## 28       53191
## 29       53191
## 30       15213
## 31       60637
## 32       60637
## 33 60208-3112
## 34       80125
## 35       60637
## 36       53191
## 37       53191
## 38       85721
## 39       15213
## 40       53191
## 41 07733-1988
## 42       60637
```

4

```
url <- "https://github.com/fonnesbeck/Bios6301/raw/master/datasets/haart.csv"
haart_df <- read.csv(url)[,c('death','weight','hemoglobin','cd4baseline')]
coef(summary(glm(death ~ ., data=haart_df, family=binomial(logit))))
```

```
##                 Estimate  Std. Error   z value     Pr(>|z|)
## (Intercept)  3.576411744 1.226870535  2.915069 0.0035561039
## weight      -0.046210552 0.022556001 -2.048703 0.0404911395
## hemoglobin  -0.350642786 0.105064078 -3.337418 0.0008456055
## cd4baseline  0.002092582 0.001811959  1.154872 0.2481427160
```

```
myfun <- function(dat, response) {
  form <- as.formula(response ~ .)
  coef(summary(glm(form, data=dat, family=binomial(logit))))
}

tryCatch(myfun(haart_df, death), error = function(e) e)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
##                   Estimate  Std. Error       z value  Pr(>|z|)
## (Intercept) -2.656607e+01 115935.1524 -2.291459e-04 0.9998172
## death        5.313213e+01  69028.2910  7.697153e-04 0.9993859
## weight      -1.610484e-15   1939.0567 -8.305501e-19 1.0000000
## hemoglobin   1.697890e-14   9774.8170  1.737004e-18 1.0000000
## cd4baseline  4.076548e-17    184.0846  2.214497e-19 1.0000000
```

```
debug(glm.fit)
```

According to the warning message, the function doesn't converge.