Bios 6301: Assignment 6

Yi Zuo

Due Thursday, 15 November, 1:00 PM

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

30 points total.

Submit a single knitr file (named `homework8.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 `homework8.rmd` or include author name may result in 5 points taken off.

Question 1

15 points

Install the `readxl` package and run the following

```
fn <- 'icd10.xlsx'
if(file.access(fn, mode = 4) == -1) {
    url <- "https://www.cdc.gov/nhsn/xls/icd10-pcs-pcm-nhsn-opc.xlsx"
    download.file(url, destfile = fn)
}
dat <- readxl::read_excel(fn, sheet = 2)
```

1. Show the class of `dat`. (1 point)

```
class(dat)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

2. Show the methods available for objects of the given class (if there are multiple classes, show methods for all classes). (3 points)

```
methods(class = class(dat)[1])
```

```
##  [1] [             [[            [[<-          [<-           $
##  [6] $<-           as.data.frame coerce        format        initialize
## [11] Ops           print         row.names<-   show          slotsFromS3
## see '?methods' for accessing help and source code
```

```
methods(class = class(dat)[2])
```

```
##  [1] [[<-        [<-         $<-         coerce      format
##  [6] initialize  Ops         print       show        slotsFromS3
## see '?methods' for accessing help and source code
```

```
methods(class = class(dat)[3])
```

```
##  [1] [             [[            [[<-          [<-           $
##  [6] $<-           aggregate     anyDuplicated as.data.frame as.list
## [11] as.matrix     by            cbind         coerce        dim
## [16] dimnames      dimnames<-    droplevels    duplicated    edit
## [21] format        formula       head          initialize    is.na
```

```
## [26] Math           merge          na.exclude     na.omit        Ops
## [31] plot           print          prompt         rbind          row.names
## [36] row.names<-    rowsum         show           slotsFromS3    split
## [41] split<-        stack          str            subset         summary
## [46] Summary        t              tail           transform      unique
## [51] unstack        within
## see '?methods' for accessing help and source code
```

3. If you call `print(dat)`, what print method is being dispatched? (1 point)

```
# print(dat)
# print.default(dat)
```

The default print method is dispatched.

4. Set the class of `dat` to be a data.frame. (1 point)

```
class(dat) <- "data.frame"
```

5. If you call `print(dat)` again, what print method is being dispatched? (1 point)

```
# print(dat)
```

The `data.frame` print method is dispatched.

Define a new generic function `nUnique` with the code below.

```
nUnique <- function(x) {
    UseMethod('nUnique')
}
```

6. Write a default method for `nUnique` to count the number of unique values in an element. (2 points)

```
nUnique.default <- function(x){
  return(length(unique(x)))
}
```

7. Check your function (2 points)

```
nUnique(letters) # should return 26
```

```
## [1] 26
```

```
nUnique(sample(10, 100, replace = TRUE)) # should return 10 (probably)
```

```
## [1] 10
```

8. Write a data.frame method for `nUnique` to operate on data.frame objects. This version should return counts for each column in a data.frame. (2 points)

```
nUnique.data.frame <- function(x){
  z <- sapply(1:ncol(x), function(i) length(unique(x[,i])))
  return(z)
}
```

9. Check your function (2 points)

```
nUnique(dat)
```

```
## [1]   39 8321 8320   15
```

Question 2

15 points

Programming with classes. The following function will generate random patient information.

```
makePatient <- function() {
  vowel <- grep("[aeiou]", letters)
  cons <- grep("[^aeiou]", letters)
  name <- paste(sample(LETTERS[cons], 1), sample(letters[vowel], 1), sample(letters[cons], 1), sep='')
  gender <- factor(sample(0:1, 1), levels=0:1, labels=c('female','male'))
  dob <- as.Date(sample(7500, 1), origin="1970-01-01")
  n <- sample(6, 1)
  doa <- as.Date(sample(1500, n), origin="2010-01-01")
  pulse <- round(rnorm(n, 80, 10))
  temp <- round(rnorm(n, 98.4, 0.3), 2)
  fluid <- round(runif(n), 2)
  list(name, gender, dob, doa, pulse, temp, fluid)
}
```

1. Create an S3 class `medicalRecord` for objects that are a list with the named elements `name`, `gender`, `date_of_birth`, `date_of_admission`, `pulse`, `temperature`, `fluid_intake`. Note that an individual patient may have multiple measurements for some measurements. Set the RNG seed to `8` and create a medical record by taking the output of `makePatient`. Print the medical record, and print the class of the medical record. (5 points)

```
mrf <- function(){
  set.seed(8)
  mr <- makePatient()
  names(mr)<-c("name","gender","date_of_birth", "date_of_admission", "pulse", "temperature", "fluid_inta
  class(mr) <- "medicalRecord"
  return(mr)
}

mr <- mrf()
mr
```

```
## $name
## [1] "Mev"
##
## $gender
## [1] male
## Levels: female male
##
## $date_of_birth
## [1] "1976-08-09"
##
## $date_of_admission
## [1] "2011-03-14" "2013-10-30" "2013-02-27" "2012-08-23" "2011-11-16"
##
## $pulse
## [1] 67 81 95 74 81
##
## $temperature
## [1] 98.33 98.16 99.00 98.49 98.67
##
```

```
## $fluid_intake
## [1] 0.62 0.93 0.18 0.39 0.34
##
## attr(,"class")
## [1] "medicalRecord"
```

2. Write a `medicalRecord` method for the generic function `mean`, which returns averages for pulse, temperature and fluids. Also write a `medicalRecord` method for `print`, which employs some nice formatting, perhaps arranging measurements by date, and `plot`, that generates a composite plot of measurements over time. Call each function for the medical record created in part 1. (5 points)

```
# mean function
mean.medicalRecord <- function(x){
  cat(sprintf("Average for pulse: %f\nAverage for temperature: %f\nAverage for fluids: %f", mean(x$puls
}

mean(mr)
```

```
## Average for pulse: 79.600000
## Average for temperature: 98.530000
## Average for fluids: 0.492000
```

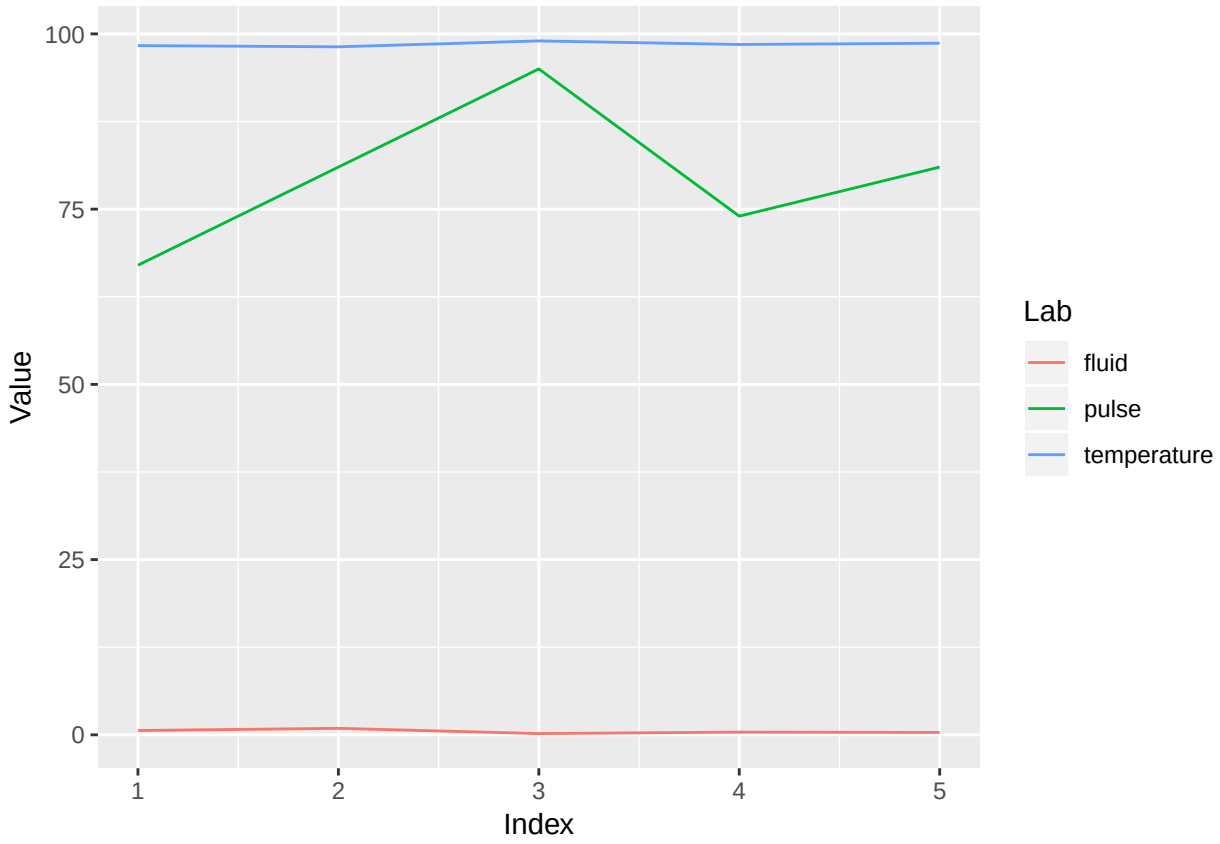
```
# print function
print.medicalRecord <- function(x) {

  date_admission <- x$date_of_admission[order(x$date_of_admission)]
  pulse <- x$pulse[order(x$date_of_admission)]
  temperature <- x$temperature[order(x$date_of_admission)]
  fluid <- x$fluid_intake[order(x$date_of_admission)]
  cat(sprintf("name: %s\ngender: %d\ndate of birth: %s\ndate of admission: %s\npulse: %d\ntemperature: %
}

mr
```

```
## name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2011-03-14
## pulse: 67
## temperature: 98.330000
## fluid intake: 0.620000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2011-11-16
## pulse: 81
## temperature: 98.670000
## fluid intake: 0.340000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2012-08-23
## pulse: 74
## temperature: 98.490000
## fluid intake: 0.390000
##  name: Mev
```

```
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2013-02-27
## pulse: 95
## temperature: 99.000000
## fluid intake: 0.180000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2013-10-30
## pulse: 81
## temperature: 98.160000
## fluid intake: 0.930000
##
```

```
# plot function

plot.medicalRecord <- function(x){

  pulse <- x$pulse
  temperature <- x$temperature
  fluid <- x$fluid_intake
  Value <- c(pulse, temperature, fluid)
  Index <- 1:length(pulse)
  Lab <- c(rep("pulse",length(pulse)),rep("temperature",length(temperature)),rep("fluid",length(fluid)))

  df <- data.frame(Index,Value,Lab)

  library(ggplot2)

  ggplot(data=df,aes(x=Index,y=Value,group=Lab)) + geom_line(aes(color=Lab))

}

plot(mr)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```



3. Create a further class for a cohort (group) of patients, and write methods for `mean` and `print` which, when applied to a cohort, apply mean or print to each patient contained in the cohort. Hint: think of this as a "container" for patients. Reset the RNG seed to 8 and create a cohort of ten patients, then show the output for `mean` and `print`. (5 points)

```
set.seed(8)
patients <- replicate(10,list(makePatient()))

class(patients) <- "Patients"

# mean
mean.Patients <- function(x){
  pulse <- sapply(1:10, function(i) mean(x[[i]][[5]]))
  temperature <- sapply(1:10, function(i) mean(x[[i]][[6]]))
  fluid <- sapply(1:10, function(i) mean(x[[i]][[7]]))

  message_pulse <- sprintf("Mean pulse for the subject %d: %f",1:10,pulse)
  message_temp <- sprintf("Mean temperature for the subject %d: %f",1:10,temperature)
  message_fluid <- sprintf("Mean fluid for the subject %d: %f",1:10,fluid)
  return(c(message_pulse,message_temp, message_fluid))
}

mean(patients)
```

```
##  [1] "Mean pulse for the subject 1: 79.600000"
##  [2] "Mean pulse for the subject 2: 78.000000"
##  [3] "Mean pulse for the subject 3: 81.500000"
##  [4] "Mean pulse for the subject 4: 78.000000"
```

```
##  [5] "Mean pulse for the subject 5: 88.333333"
##  [6] "Mean pulse for the subject 6: 83.500000"
##  [7] "Mean pulse for the subject 7: 83.000000"
##  [8] "Mean pulse for the subject 8: 77.500000"
##  [9] "Mean pulse for the subject 9: 77.000000"
## [10] "Mean pulse for the subject 10: 79.333333"
## [11] "Mean temperature for the subject 1: 98.530000"
## [12] "Mean temperature for the subject 2: 98.495000"
## [13] "Mean temperature for the subject 3: 98.440000"
## [14] "Mean temperature for the subject 4: 98.600000"
## [15] "Mean temperature for the subject 5: 98.050000"
## [16] "Mean temperature for the subject 6: 98.450000"
## [17] "Mean temperature for the subject 7: 98.010000"
## [18] "Mean temperature for the subject 8: 98.148333"
## [19] "Mean temperature for the subject 9: 98.830000"
## [20] "Mean temperature for the subject 10: 98.300000"
## [21] "Mean fluid for the subject 1: 0.492000"
## [22] "Mean fluid for the subject 2: 0.245000"
## [23] "Mean fluid for the subject 3: 0.403333"
## [24] "Mean fluid for the subject 4: 0.650000"
## [25] "Mean fluid for the subject 5: 0.586667"
## [26] "Mean fluid for the subject 6: 0.452500"
## [27] "Mean fluid for the subject 7: 0.970000"
## [28] "Mean fluid for the subject 8: 0.336667"
## [29] "Mean fluid for the subject 9: 0.445000"
## [30] "Mean fluid for the subject 10: 0.658333"
```

```
# print
print.Patients <- function(x){

  # get the number of visits from each person
  n_visit <- sapply(1:10, function(i) length(x[[i]][[4]]))

  # store the admission date info
  date_admission <- sapply(1:10, function(i) x[[i]][[4]][order(x[[i]][[4]])])

  # pulse
  pulse <- sapply(1:10, function(i) x[[i]][[5]][order(x[[i]][[4]])])
  temperature <- sapply(1:10, function(i) x[[i]][[6]][order(x[[i]][[4]])])
  fluid <- sapply(1:10, function(i) x[[i]][[7]][order(x[[i]][[4]])])

  # print
  sapply(1:10, function(i) cat(sprintf("name: %s\ngender: %d\ndate of birth: %s\ndate of admission: %s\n

}

patients
```

```
## name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2011-03-14
## pulse: 67
## temperature: 98.330000
```

```
## fluid intake: 0.620000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2011-11-16
## pulse: 81
## temperature: 98.670000
## fluid intake: 0.340000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2012-08-23
## pulse: 74
## temperature: 98.490000
## fluid intake: 0.390000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2013-02-27
## pulse: 95
## temperature: 99.000000
## fluid intake: 0.180000
##  name: Mev
## gender: 2
## date of birth: 1976-08-09
## date of admission: 2013-10-30
## pulse: 81
## temperature: 98.160000
## fluid intake: 0.930000
##
## name: Yul
## gender: 2
## date of birth: 1988-06-28
## date of admission: 2012-01-16
## pulse: 76
## temperature: 98.920000
## fluid intake: 0.140000
##  name: Yul
## gender: 2
## date of birth: 1988-06-28
## date of admission: 2013-08-07
## pulse: 80
## temperature: 98.070000
## fluid intake: 0.350000
##
## name: Zet
## gender: 1
## date of birth: 1970-06-13
## date of admission: 2010-03-21
## pulse: 79
## temperature: 98.580000
## fluid intake: 0.220000
##  name: Zet
## gender: 1
```

```
## date of birth: 1970-06-13
## date of admission: 2010-04-01
## pulse: 73
## temperature: 98.320000
## fluid intake: 0.610000
##  name: Zet
## gender: 1
## date of birth: 1970-06-13
## date of admission: 2012-08-29
## pulse: 88
## temperature: 98.470000
## fluid intake: 0.590000
##  name: Zet
## gender: 1
## date of birth: 1970-06-13
## date of admission: 2013-06-01
## pulse: 84
## temperature: 98.220000
## fluid intake: 0.250000
##  name: Zet
## gender: 1
## date of birth: 1970-06-13
## date of admission: 2013-11-03
## pulse: 72
## temperature: 98.540000
## fluid intake: 0.030000
##  name: Zet
## gender: 1
## date of birth: 1970-06-13
## date of admission: 2014-02-05
## pulse: 93
## temperature: 98.510000
## fluid intake: 0.720000
##
## name: Qih
## gender: 1
## date of birth: 1987-08-30
## date of admission: 2011-06-22
## pulse: 78
## temperature: 98.600000
## fluid intake: 0.650000
##
## name: Wut
## gender: 2
## date of birth: 1974-06-28
## date of admission: 2010-04-12
## pulse: 76
## temperature: 98.050000
## fluid intake: 0.650000
##  name: Wut
## gender: 2
## date of birth: 1974-06-28
## date of admission: 2011-02-16
## pulse: 93
```

```
## temperature: 98.260000
## fluid intake: 0.970000
##  name: Wut
## gender: 2
## date of birth: 1974-06-28
## date of admission: 2012-04-12
## pulse: 96
## temperature: 97.840000
## fluid intake: 0.140000
##
## name: Juy
## gender: 2
## date of birth: 1983-06-09
## date of admission: 2010-03-10
## pulse: 81
## temperature: 99.110000
## fluid intake: 0.660000
##  name: Juy
## gender: 2
## date of birth: 1983-06-09
## date of admission: 2010-03-25
## pulse: 90
## temperature: 98.580000
## fluid intake: 0.260000
##  name: Juy
## gender: 2
## date of birth: 1983-06-09
## date of admission: 2010-04-18
## pulse: 75
## temperature: 98.580000
## fluid intake: 0.600000
##  name: Juy
## gender: 2
## date of birth: 1983-06-09
## date of admission: 2010-06-10
## pulse: 88
## temperature: 97.530000
## fluid intake: 0.290000
##
## name: God
## gender: 1
## date of birth: 1990-02-12
## date of admission: 2010-03-12
## pulse: 83
## temperature: 98.010000
## fluid intake: 0.970000
##
## name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2011-04-07
## pulse: 80
## temperature: 97.870000
## fluid intake: 0.360000
```

```
##  name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2011-04-14
## pulse: 83
## temperature: 97.910000
## fluid intake: 0.000000
##  name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2011-08-16
## pulse: 66
## temperature: 98.490000
## fluid intake: 0.130000
##  name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2013-03-15
## pulse: 74
## temperature: 98.380000
## fluid intake: 0.310000
##  name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2013-06-20
## pulse: 74
## temperature: 98.410000
## fluid intake: 0.490000
##  name: Fut
## gender: 2
## date of birth: 1970-01-11
## date of admission: 2013-11-12
## pulse: 88
## temperature: 97.830000
## fluid intake: 0.730000
##
## name: Pet
## gender: 2
## date of birth: 1979-01-01
## date of admission: 2010-10-30
## pulse: 85
## temperature: 98.840000
## fluid intake: 0.600000
##  name: Pet
## gender: 2
## date of birth: 1979-01-01
## date of admission: 2012-05-10
## pulse: 69
## temperature: 98.820000
## fluid intake: 0.290000
##
## name: Yed
## gender: 2
## date of birth: 1977-11-11
```

```
## date of admission: 2010-01-28
## pulse: 63
## temperature: 97.950000
## fluid intake: 0.940000
##  name: Yed
## gender: 2
## date of birth: 1977-11-11
## date of admission: 2010-03-06
## pulse: 81
## temperature: 98.450000
## fluid intake: 0.670000
##  name: Yed
## gender: 2
## date of birth: 1977-11-11
## date of admission: 2010-07-10
## pulse: 98
## temperature: 98.650000
## fluid intake: 0.790000
##  name: Yed
## gender: 2
## date of birth: 1977-11-11
## date of admission: 2010-08-27
## pulse: 66
## temperature: 97.680000
## fluid intake: 0.360000
##  name: Yed
## gender: 2
## date of birth: 1977-11-11
## date of admission: 2011-06-18
## pulse: 83
## temperature: 98.000000
## fluid intake: 0.690000
##  name: Yed
## gender: 2
## date of birth: 1977-11-11
## date of admission: 2013-01-06
## pulse: 85
## temperature: 99.070000
## fluid intake: 0.500000
##
```