MedTourEasy

Project description

Sepsis is a deadly illness that accounts for a large portion of in-hopital deaths. It occurs when a persons organs shut down in response to a severe infection. This public health problem is a major target for research, and hospital records can help us investigate the problem. In this R project, you will identify hospital patients with severe infection using medical record data.

To successfully complete thsi project you should have some experience with the package data.table inclding using the := operatior , grouping aggregations with by, and understandin how to use the shift function.

Process

We will work closely with you to build and fulfill the needs of this project by the end of your internship. we will do this by establishing clear goals and a comprehensive solution based on project requirements.

Our process to achieve this as follows:

Task1: Instructions

First, lets take a look at the antibiotic data.

- Load the data.table package using library().
- Read in datasets/antibioticDT.csv using the data.table function fread().
- Look at the first 30 rows.

```
library(data.table)

antibioticDT <- fread("antibioticDT.csv")
head(antibioticDT, 30, row.names = FALSE)
```

```
##     patient_id day_given antibiotic_type route
##  1:          1         2   ciprofloxacin    IV
##  2:          1         4   ciprofloxacin    IV
##  3:          1         6   ciprofloxacin    IV
##  4:          1         7     doxycycline    IV
##  5:          1         9     doxycycline    IV
##  6:          1        15      penicillin    IV
##  7:          1        16     doxycycline    IV
##  8:          1        18   ciprofloxacin    IV
##  9:          8         1     doxycycline    PO
## 10:          8         2      penicillin    IV
## 11:          8         3     doxycycline    IV
## 12:          8         6     doxycycline    PO
## 13:          8         8      penicillin    PO
## 14:          8        12      penicillin    IV
## 15:          9         8     doxycycline    IV
## 16:          9        12     doxycycline    PO
## 17:         12         4     doxycycline    PO
## 18:         12         9     doxycycline    IV
## 19:         16         1     doxycycline    IV
```

```
## 20:          16          4     amoxicillin    IV
## 21:          19          3     doxycycline    PO
## 22:          19          5     amoxicillin    IV
## 23:          19          6   ciprofloxacin    IV
## 24:          19         10     doxycycline    IV
## 25:          19         12      penicillin    IV
## 26:          23          1     doxycycline    IV
## 27:          23          1      penicillin    IV
## 28:          23          3     amoxicillin    IV
## 29:          23          3   ciprofloxacin    IV
## 30:          23          3     doxycycline    IV
##     patient_id day_given antibiotic_type route
```

Task2:

Identify rows representing "new" antibiotics.

- Use setorder() to sort the data by patients_id, antibiotic_type and day_given. Print and examine the first40 rows.
- Use shift to calculate the last day the antibiotic was given to a patient. Call the new variable, last_administration_day.
- Calculate the number of days since the antibiotic was administered to a patient. Call the new variable, days_since_last_admin.
- In a two-step process, create a new variable called antibiotic_new that is initialized to one, the reset it ot zero where it has only been one or two days since the antibiotic was given.

```
setorder(antibioticDT, patient_id, antibiotic_type, day_given)
antibioticDT[, last_administration_day := shift(day_given,type = "lag") ]
antibioticDT[, days_since_last_admin := day_given - last_administration_day, by = patient_id]
antibioticDT[, antibiotic_new := ifelse(days_since_last_admin <= 2, 0, last_administration_day) ]
head(antibioticDT)
```

```
##    patient_id day_given antibiotic_type route last_administration_day
## 1:          1         2   ciprofloxacin    IV                      NA
## 2:          1         4   ciprofloxacin    IV                       2
## 3:          1         6   ciprofloxacin    IV                       4
## 4:          1        18   ciprofloxacin    IV                       6
## 5:          1         7     doxycycline    IV                      18
## 6:          1         9     doxycycline    IV                       7
##    days_since_last_admin antibiotic_new
## 1:                    NA             NA
## 2:                     2              0
## 3:                     2              0
## 4:                    12              6
## 5:                   -11              0
## 6:                     2              0
```

Task3:

Investigate the blood culture data.

- Read in "datasets/blood_cultureDT.csv".

- Print the first 30 rows.

```
blood_cultureDT <- fread("blood_cultureDT.csv")
head(blood_cultureDT,30)
```

```
##     patient_id blood_culture_day
##  1:          1                 3
##  2:          1                13
##  3:          8                 2
##  4:          8                13
##  5:         23                 3
##  6:         39                10
##  7:         45                 4
##  8:         45                 9
##  9:         45                11
## 10:         51                 3
## 11:         51                 6
## 12:         59                 2
## 13:         64                 1
## 14:         66                 9
## 15:         66                10
## 16:         69                 2
## 17:         69                 6
## 18:         69                 7
## 19:         69                11
## 20:         69                16
## 21:         76                 1
## 22:         77                 3
## 23:         79                 5
## 24:         79                11
## 25:         79                12
## 26:         80                 3
## 27:         80                12
## 28:         81                 2
## 29:        112                 6
## 30:        115                 2
##     patient_id blood_culture_day
```

Task4:

Merge the antibiotic data with blood culture data.

- make a combined dataset by merging antibioticDT with blood_cultureDT.
- Sort by patient_id, blood_culture_day, day_given, and antibiotic_type.
- Print and examine the first 3o rows.

```
combinedDT <- merge(antibioticDT,blood_cultureDT, by = "patient_id")
setorder(combinedDT, patient_id, blood_culture_day, day_given, antibiotic_type)
head(combinedDT)
```

```
##    patient_id day_given antibiotic_type route last_administration_day
## 1:          1         2   ciprofloxacin    IV                     NA
```

```
## 2:          1         4   ciprofloxacin    IV                     2
## 3:          1         6   ciprofloxacin    IV                     4
## 4:          1         7     doxycycline    IV                    18
## 5:          1         9     doxycycline    IV                     7
## 6:          1        15      penicillin    IV                    16
##    days_since_last_admin antibiotic_new blood_culture_day
## 1:                    NA             NA                 3
## 2:                     2              0                 3
## 3:                     2              0                 3
## 4:                   -11              0                 3
## 5:                     2              0                 3
## 6:                    -1              0                 3
```

Task5:

Make a new variable indicating whether or not the antibiotic administration and blood culture are withn two days of each other.

- make anew vriable called drug_in_bcx_window which is 1 if the drug was given in the 2-day window and 0 otherwise.

For indicator functions, it can be handy to use as.numeric to convert logical values (TRUE or False) to 0 or 1.

```
combinedDT[, drug_in_bcx_window :=
             ifelse(abs(blood_culture_day - day_given) <=2, 1, 0)]
head(combinedDT)
```

```
##    patient_id day_given antibiotic_type route last_administration_day
## 1:          1         2   ciprofloxacin    IV                     NA
## 2:          1         4   ciprofloxacin    IV                      2
## 3:          1         6   ciprofloxacin    IV                      4
## 4:          1         7     doxycycline    IV                     18
## 5:          1         9     doxycycline    IV                      7
## 6:          1        15      penicillin    IV                     16
##    days_since_last_admin antibiotic_new blood_culture_day drug_in_bcx_window
## 1:                    NA             NA                 3                  1
## 2:                     2              0                 3                  1
## 3:                     2              0                 3                  0
## 4:                   -11              0                 3                  0
## 5:                     2              0                 3                  0
## 6:                    -1              0                 3                  0
```

Task6:

For each patient/blood culture day combination, determine if at least one I.V. antibiotic was given in the +/- 2 day window.

- Create a new variable, any_iv_in_bcx_window, whether or not an I.V. drug was given a +/- 2 day window of a blood culture day.
- Exclude rows in which the blood_culture_day does not have any I.V. drug in the window.

Use any() to check if there are any row that are both: (1) in +/-2 day window, and (2) have an I.V. drug administeres. Use by = to make sure this is calculated within each blood culture day for each patient.

```
combinedDT[, any_iv_in_bcx_window := any(route == "IV" |drug_in_bcx_window== 1),
           by = .(patient_id,blood_culture_day)]
combinedDT[route == "IV"]
```

```
##       patient_id day_given antibiotic_type route last_administration_day
##    1:          1         2   ciprofloxacin    IV                      NA
##    2:          1         4   ciprofloxacin    IV                       2
##    3:          1         6   ciprofloxacin    IV                       4
##    4:          1         7     doxycycline    IV                      18
##    5:          1         9     doxycycline    IV                       7
##   ---
## 5025:       2998         6      penicillin    IV                       5
## 5026:       2998         9   ciprofloxacin    IV                       3
## 5027:       2998        14   ciprofloxacin    IV                       9
## 5028:       2998        15     doxycycline    IV                       8
## 5029:       2998        17   ciprofloxacin    IV                      14
##       days_since_last_admin antibiotic_new blood_culture_day drug_in_bcx_window
##    1:                    NA             NA                 3                  1
##    2:                     2              0                 3                  1
##    3:                     2              0                 3                  0
##    4:                   -11              0                 3                  0
##    5:                     2              0                 3                  0
##   ---
## 5025:                     1              0                 4                  1
## 5026:                     6              3                 4                  0
## 5027:                     5              9                 4                  0
## 5028:                     7              8                 4                  0
## 5029:                     3             14                 4                  0
##       any_iv_in_bcx_window
##    1:                 TRUE
##    2:                 TRUE
##    3:                 TRUE
##    4:                 TRUE
##    5:                 TRUE
##   ---
## 5025:                 TRUE
## 5026:                 TRUE
## 5027:                 TRUE
## 5028:                 TRUE
## 5029:                 TRUE
```

```
head(combinedDT)
```

```
##    patient_id day_given antibiotic_type route last_administration_day
## 1:          1         2   ciprofloxacin    IV                      NA
## 2:          1         4   ciprofloxacin    IV                       2
## 3:          1         6   ciprofloxacin    IV                       4
## 4:          1         7     doxycycline    IV                      18
## 5:          1         9     doxycycline    IV                       7
## 6:          1        15      penicillin    IV                      16
```

```
##    days_since_last_admin antibiotic_new blood_culture_day drug_in_bcx_window
## 1:                    NA             NA                 3                  1
## 2:                     2              0                 3                  1
## 3:                     2              0                 3                  0
## 4:                   -11              0                 3                  0
## 5:                     2              0                 3                  0
## 6:                    -1              0                 3                  0
##    any_iv_in_bcx_window
## 1:                 TRUE
## 2:                 TRUE
## 3:                 TRUE
## 4:                 TRUE
## 5:                 TRUE
## 6:                 TRUE
```

Task7:

For each blood culture, find the first day of potential 4-day antibiotic sequences. This day will be the first day that is both in the window, and a new antibiotic was given.

- Create a new variable called day_of_first_new_abx_in_window.
- Remove rows where the day is before this first qualifying day.

Since we are looking for the day, start with day_given and index from there. Then select only the first, using Indexing[1] only works if the data are sorted by day, which we did in a previous step. Remember, this will be the first day that is both in the window and a new antibiotic was given.

```
combinedDT[, day_of_first_new_abx_in_window := day_given[1],
             by = blood_culture_day]

combinedDT[!day_given < day_given[1]]
```

```
##       patient_id day_given antibiotic_type route last_administration_day
##    1:          1         2   ciprofloxacin    IV                      NA
##    2:          1         4   ciprofloxacin    IV                       2
##    3:          1         6   ciprofloxacin    IV                       4
##    4:          1         7     doxycycline    IV                      18
##    5:          1         9     doxycycline    IV                       7
##   ---
## 5820:       2998         8     doxycycline    PO                       4
## 5821:       2998         9   ciprofloxacin    IV                       3
## 5822:       2998        14   ciprofloxacin    IV                       9
## 5823:       2998        15     doxycycline    IV                       8
## 5824:       2998        17   ciprofloxacin    IV                      14
##       days_since_last_admin antibiotic_new blood_culture_day drug_in_bcx_window
##    1:                    NA             NA                 3                  1
##    2:                     2              0                 3                  1
##    3:                     2              0                 3                  0
##    4:                   -11              0                 3                  0
##    5:                     2              0                 3                  0
##   ---
## 5820:                     4              4                 4                  0
```

```
## 5821:                    6             3                4                 0
## 5822:                    5             9                4                 0
## 5823:                    7             8                4                 0
## 5824:                    3            14                4                 0
##       any_iv_in_bcx_window day_of_first_new_abx_in_window
##    1:                 TRUE                             2
##    2:                 TRUE                             2
##    3:                 TRUE                             2
##    4:                 TRUE                             2
##    5:                 TRUE                             2
##   ---
## 5820:                 TRUE                             1
## 5821:                 TRUE                             1
## 5822:                 TRUE                             1
## 5823:                 TRUE                             1
## 5824:                 TRUE                             1
```

```
head(combinedDT)
```

```
##    patient_id day_given antibiotic_type route last_administration_day
## 1:          1         2   ciprofloxacin    IV                     NA
## 2:          1         4   ciprofloxacin    IV                      2
## 3:          1         6   ciprofloxacin    IV                      4
## 4:          1         7     doxycycline    IV                     18
## 5:          1         9     doxycycline    IV                      7
## 6:          1        15      penicillin    IV                     16
##    days_since_last_admin antibiotic_new blood_culture_day drug_in_bcx_window
## 1:                    NA             NA                3                  1
## 2:                     2              0                3                  1
## 3:                     2              0                3                  0
## 4:                   -11              0                3                  0
## 5:                     2              0                3                  0
## 6:                    -1              0                3                  0
##    any_iv_in_bcx_window day_of_first_new_abx_in_window
## 1:                 TRUE                            2
## 2:                 TRUE                            2
## 3:                 TRUE                            2
## 4:                 TRUE                            2
## 5:                 TRUE                            2
## 6:                 TRUE                            2
```

Task8:

Make a new dataset that only contains what we need to check the remainig criteria.

- Create a new data.table containing only patient_id, blood_culture_day, and day_given.
- Remove duplicate rows.

```
simplified_data <- combinedDT[, .(patient_id, blood_culture_day, day_given)]

setkey(simplified_data, NULL)
simplified_data <-  unique(simplified_data)
head(simplified_data)
```

```
##    patient_id blood_culture_day day_given
## 1:          1                 3         2
## 2:          1                 3         4
## 3:          1                 3         6
## 4:          1                 3         7
## 5:          1                 3         9
## 6:          1                 3        15
```

Task9:

Extract the first antibiotic days.

- Make a new variable, num_antibiotic_days, showing the number of antibiotic days each patient/blood culture day combination had.
- Remove blood culture days with less than four antibiotic days(rows).
- Select the first four days(rows) for each blood culture.

The special symbol .N counts the number of observation. When used with by =, it counts the number of rows in each by= group. You can use this to get the number of antibiotic days in each patient-blood culture day.

Selecting the first four row for each patient ID/ blood culture day combination is a little tricky. Use the data.table special symbol .SD.

print(".SD[1:4]")

```
simplified_data[, num_antibiotic_days := .N, by = .(patient_id,blood_culture_day)]
simplified_data[!num_antibiotic_days <= 4 ]
```

```
##       patient_id blood_culture_day day_given num_antibiotic_days
##    1:          1                 3         2                   8
##    2:          1                 3         4                   8
##    3:          1                 3         6                   8
##    4:          1                 3         7                   8
##    5:          1                 3         9                   8
##   ---
## 3716:       2998                 4         8                  11
## 3717:       2998                 4         9                  11
## 3718:       2998                 4        14                  11
## 3719:       2998                 4        15                  11
## 3720:       2998                 4        17                  11
```

```
simplified_data <- simplified_data[, .SD[1:4], by = blood_culture_day]
head(simplified_data)
```

```
##    blood_culture_day patient_id day_given num_antibiotic_days
## 1:                 3          1         2                   8
## 2:                 3          1         4                   8
## 3:                 3          1         6                   8
## 4:                 3          1         7                   8
## 5:                13          1         2                   8
## 6:                13          1         4                   8
```

Task10:

Find which four-day sequences qualify.

- Make a new 0/1 variable, four_in_seq, indicating whether or not the antibiotic sequences has no skips of more than one day.

diff() takes a vector of numbers and calculates the differences between each pair of adjacent numbers. If there is a gap of one day, the difference will be two .max() of the diff() would be useful here too.

Do not forget as.numeric() when making four_in_seq a 0/1 indicator.

```
first_four_days <-
  simplified_data[, four_in_seq :=
                    ifelse((day_given - shift(simplified_data$day_given,type = "lag")) ==

head(first_four_days)
```

```
##    blood_culture_day patient_id day_given num_antibiotic_days four_in_seq
## 1:                 3          1         2                    8          NA
## 2:                 3          1         4                    8           1
## 3:                 3          1         6                    8           1
## 4:                 3          1         7                    8           0
## 5:                13          1         2                    8           0
## 6:                13          1         4                    8           1
```

Task11:

Create a new data frame with one row for each patient_id with suspected infection.

- Select the rows which have four_in_seq equal to 1.
- Retain only the patient_id column.
- Get rid of duplicates.
- Make a new indicator, infection, setting it to 1 for everyone.

To select one column of a data table as a new data table, use .() with the column inside the parantheses.

```
suspected_infection <- first_four_days[four_in_seq == 1]
suspected_infection <- suspected_infection[,.(patient_id)]
suspected_infection <- unique(suspected_infection)
suspected_infection <- suspected_infection[, infection_indicator := 1 ]
head(suspected_infection,7)
```

```
##    patient_id infection_indicator
## 1:          1                   1
## 2:         45                   1
## 3:         64                   1
## 4:         80                   1
## 5:        379                   1
## 6:        157                   1
## 7:        846                   1
```

Task12:

Find the percentage of presumed serious infections in the data.

- Use fread() to read in "datasets/all_patients.csv", which contains a record of all patients who were in the hospital during the same two-week timeframe.
- Merge this dataset with the infection flag data. Make sure to retain all patients.
- The patients who were not in the antibiotic and blood culture data will have missing values for the infection flag. Set these to 0.
- Calculate the percentage of patients who met the criteria for presumed infection.

```
all_patientsDT <- fread("all_patients.csv")
all_patientsDT <- merge(all_patientsDT,suspected_infection, by = "patient_id",all.x = T)


all_patientsDT <- all_patientsDT[, infection_indicator := ifelse(is.na(infection_indicator), 0 , 1)]
prop.table(table(all_patientsDT[,infection_indicator]))
```

```
##
##           0           1
## 0.992134831 0.007865169
```