



Sample Data

陳曾基

國立陽明交通大學醫學院醫務管理研究所
國立陽明交通大學醫學院急重症醫學研究所
國立陽明交通大學醫學院醫學系家庭醫學科
臺北榮民總醫院家庭醫學部
臺北榮民總醫院醫學研究部大數據中心

示範資料說明 (1)

- 示範程式所用的資料(教學檔)，係依據舊版國家衛生研究院全民健康保險研究資料庫格式 (https://nhird.nhri.org.tw/date_02.html)，隨機仿製的虛擬資料，僅供程式練習用，內容可能有很多不合理的地方，例如：男性有婦科疾病、費用兜不攏。

示範資料說明 (2)

- 請將示範資料(教學檔)複製於D槽 SampleData子目錄(D:/SampleData/)下，將可不用更改課程提供的相關程式碼。如果個人電腦沒有D槽或習慣將檔案擺放於其他位置，則在匯入資料時，需自行於程式碼內更改檔案位置。
- 匯入資料程式與示範資料置於相同的子目錄下

注意

- 複製程式碼至MS PowerPoint或Word時，或在MS PowerPoint或Word裡更動文字時，MS Office往往會自動將程式碼內直的單引號(')與雙引號(")改為斜的單引號(')與雙引號(")。若未注意，將程式碼從MS PowerPoint或Word複製至R時，R將無法判讀，形成Error，需自行更正。

匯入示範資料

- 利用 Import_tidyverse.r，匯入資料 (可用 Notepad++ 開啟 .r 檔，複製全部內容，貼至 RGui 或 RStudio 的 Console 處，執行程式碼) (也可在 RGui 或 RStudio 開啟 .r 檔，執行程式碼)
- 如果示範資料放置於其他位置，請同時更改 Import_tidyverse.r 內資料檔案位置的子目錄
- 未來欲快速匯入資料，可直接執行：
`source("D:/SampleData/Import_tidyverse.r")`
CHANGE HERE !!! 請更動 Import_tidyverse.r 檔的位置


```

> if (! require(tidyverse)) { install.packages("tidyverse"); require(tidyverse) }
>
> directory <- "D:/SampleData/"      # CHANGE HERE !!!
>
> columnnames <- c("FEE_YM", "APPL_TYPE", "HOSP_ID", "APPL_DATE", "CASE_TYPE", "SE
>
> columnwidths <- c(6, 1, 34, 8, 2, 6, 2, 2, 2, 2, 2, 8, 8, 8, 32, 4, 1, 3, 5, 5,
>
> columntypes <- "ccccncccccccccccccccccccccccccccccccccccccccccccccccccccc"
>
> cd <- read_fwf(
+   paste(directory, "CD2009.DAT", sep = ""),
+   fwf_widths(columnwidths, columnnames),
+   col_types = columntypes,
+   progress = TRUE
+ )
|=====| 100% 171 MB
> cd
# A tibble: 598,574 x 37
  FEE_YM APPL_TYPE HOSP_ID          APPL_DATE CASE_TYPE SEQ_NO CURE_ITEM_NO1
  <chr>  <chr>      <chr>          <chr>      <chr>      <dbl> <chr>
1 200901 1          0000000000000000~ 20090204  01          3866 <NA>
2 200901 1          0000000000000000~ 20090212  01           247 <NA>
3 200901 1          0000000000000000~ 20090212  01           919 <NA>
4 200901 1          0000000000000000~ 20090212  01           752 <NA>
5 200901 1          0000000000000000~ 20090212  01           844 <NA>
6 200901 1          0000000000000000~ 20090212  01           852 <NA>
7 200901 1          0000000000000000~ 20090209  01          6542 <NA>
8 200901 1          0000000000000000~ 20090207  09           389 D0
9 200901 1          0000000000000000~ 20090207  09           369 <NA>
10 200901 1          0000000000000000~ 20090211  08          3196 33
# ... with 598,564 more rows, and 30 more variables: CURE_ITEM_NO2 <chr>,
# CURE_ITEM_NO3 <chr>, CURE_ITEM_NO4 <chr>, FUNC_TYPE <chr>,
# FUNC_DATE <chr>, TREAT_END_DATE <chr>, ID_BIRTHDAY <chr>, ID <chr>,
# CARD_SEQ_NO <chr>, GAVE_KIND <chr>, PART_NO <chr>, ACODE_ICD9_1 <chr>,
# ACODE_ICD9_2 <chr>, ACODE_ICD9_3 <chr>, ICD_OP_CODE <chr>,
# DRUG_DAY <dbl>, MED_TYPE <chr>, PRSN_ID <chr>, PHAR_ID <chr>,
# DRUG_AMT <dbl>, TREAT_AMT <dbl>, TREAT_CODE <chr>, DIAG_AMT <dbl>,
# DSVC_NO <chr>, DSVC_AMT <dbl>, BY_PASS_CODE <chr>, T_AMT <dbl>,
# PART_AMT <dbl>, T_APPL_AMT <dbl>, ID_SEX <chr>

```

匯入完畢後請測試

head(cd) # 顯示前6列

tail(cd) # 顯示最後6列

summary(cd)

summary(cd\$ID_SEX)

table(cd\$ID_SEX)

table(cd\$CURE_ITEM_NO1)

summary(cd\$T_AMT)

class(cd\$TREAT_CODE)

summary(cd\$TREAT_CODE)

務必在上課前
成功匯入資料

Thanks for
Your Attention
!

