Create a Demographics Table with gt Package in R

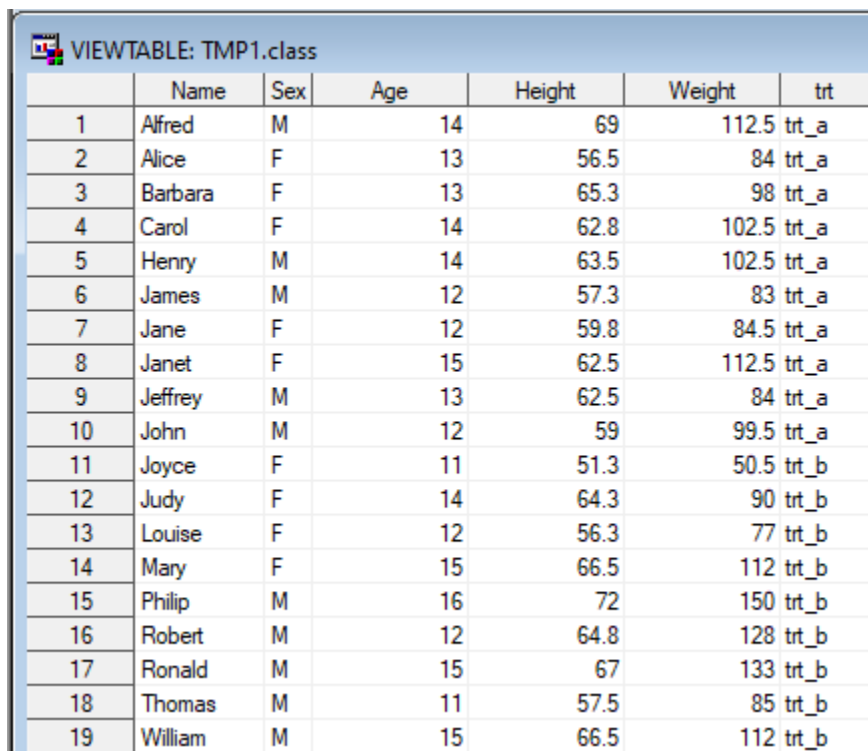Hengwei Liu, Hengrui USA, Princeton, New Jersey

ABSTRACT

SAS® programmers create demographics and baseline characteristics table in each study. In the table descriptive statistics are given for continuous variables; count and percentage are given for categorical variables. In this paper we will attempt to create the demographics table with the gt package in R.

INTRODUCTION

We will use the dataset SASHELP.CLASS with some modification. Below is the SAS program used to create the dataset class.

```
data class; set sashelp.class;
if _n_<=10 then trt='trt_a';
else trt='trt_b';
run;
```

The display 1 shows the dataset class.

	Name	Sex	Age	Height	Weight	trt
1	Alfred	M	14	69	112.5	trt_a
2	Alice	F	13	56.5	84	trt_a
3	Barbara	F	13	65.3	98	trt_a
4	Carol	F	14	62.8	102.5	trt_a
5	Henry	M	14	63.5	102.5	trt_a
6	James	M	12	57.3	83	trt_a
7	Jane	F	12	59.8	84.5	trt_a
8	Janet	F	15	62.5	112.5	trt_a
9	Jeffrey	M	13	62.5	84	trt_a
10	John	M	12	59	99.5	trt_a
11	Joyce	F	11	51.3	50.5	trt_b
12	Judy	F	14	64.3	90	trt_b
13	Louise	F	12	56.3	77	trt_b
14	Mary	F	15	66.5	112	trt_b
15	Philip	M	16	72	150	trt_b
16	Robert	M	12	64.8	128	trt_b
17	Ronald	M	15	67	133	trt_b
18	Thomas	M	11	57.5	85	trt_b
19	William	M	15	66.5	112	trt_b

Display 1. The Dataset Class.

The display 2 shows a demographics table in RTF format created with SAS.

Table 14.1 Demographics and Baseline Characteristics
Safety Population

Statistics	Treatment A (N=10)	Treatment B (N=9)
Height (in)		
n	10	9
Mean	61.82	62.91
SD	3.788	6.504
Median	62.50	64.80
Min, Max	56.5, 69.0	51.3, 72.0
Weight (kg)		
n	10	9
Mean	96.30	104.17
SD	11.708	31.243
Median	98.75	112.00
Min, Max	83.0, 112.5	50.5, 150.0
Sex		
Male	5 (50.0)	5 (55.6)
Female	5 (50.0)	4 (44.4)

Note: the source data is class.
Program Source: table1.sas

Executed: (Draft) 2022-02-09

Display 2. Demographics Table Created with SAS.

We will try to use the gt package in R to re-produce this table. See the reference [1] and [2] for more details about the gt package.

THE DETAILS

We will need three packages in the R program. We use the haven package to read the SAS datasets, use the psych package to get the descriptive statistics, and use the gt package to create the table in HTML format.

This is the R program. It is also available in reference [3].

```
library(haven)
library(psych)
library(gt)


the_date <- as.character(Sys.Date())



# read in the data
class <- read_sas("class.sas7bdat")


# get the descriptive statistics
ht <- describeBy(class$Height, group=class$trt, mat=TRUE)
wt <- describeBy(class$Weight, group=class$trt, mat=TRUE)


# get the count and percentage
bign <- table(group=class$trt)
bign1 <- as.numeric(bign[1])
bign2 <- as.numeric(bign[2])



freq <- table(class$Sex, group=class$trt)
prop <- 100*prop.table(table(class$Sex, class$trt), 2)


# handle the decimals
ht$n <- format(ht$n, nsmall=0)
ht$mean <- format(round(ht$mean,2), nsmall=2)
ht$sd <- format(round(ht$sd,3), nsmall=3)
ht$median <- format(round(ht$median,2), nsmall=2)
```

```r
ht$min <- format(round(ht$min,1), nsmall=1)
ht$max <- format(round(ht$max,1), nsmall=1)


wt$n <- format(wt$n, nsmall=0)
wt$mean <- format(round(wt$mean,2), nsmall=2)
wt$sd <- format(round(wt$sd,3), nsmall=3)
wt$median <- format(round(wt$median,2), nsmall=2)
wt$min <- format(round(wt$min,1), nsmall=1)
wt$max <- format(round(wt$max,1), nsmall=1)


# create a variable minmax, and do transpose
ht$minmax <- paste(ht$min, ',', ht$max)
ht <- ht[c("n","mean","sd","median","minmax")]
ht2 <- t(ht)


wt$minmax <- paste(wt$min, ',', wt$max)
wt <- wt[c("n","mean","sd","median","minmax")]
wt2 <- t(wt)


# combine count and percentage
X11 <- paste(freq[,1], '(', format(prop[,1], digit=3), ')')
X12 <- paste(freq[,2], '(', format(prop[,2], digit=3), ')')
sex <- cbind(X11, X12)


# create a new column called statistics and get the final data for
reporting
rownames(sex) <- c('Female','Male')
rownames(ht2)  <- c("n","mean","sd","median","min, max")


sex3 <- data.frame(statistics=rownames(sex), sex)
ht3 <- data.frame(statistics=rownames(ht2), ht2)
wt3 <- data.frame(statistics=rownames(ht2), wt2)


final <- rbind(ht3, wt3, sex3)



# use gt to do the reporting
tab_rtf <- final %>%
  gt() %>%
```

```
  tab_row_group(

    label = "Sex",

    rows = 11:12

  ) %>%


  tab_row_group(

    label = "Weight (kg)",

    rows = 6:10

  ) %>%


tab_row_group(

  label = "Height (in)",

  rows = 1:5

) %>%


tab_header(

  title = "Table 14.1 Demographics and Baseline Characteristics",

  subtitle = "Safety Population"

) %>%

tab_source_note(

  source_note = "Note: the source data is class."

) %>%


tab_source_note(

    source_note = paste('Program Source: demog.R          Executed:
(Draft)',  the_date)

  ) %>%


cols_label(

  X11 = html(paste("Treatment A <br> (N=", bign1, ")")),

  X12 = html(paste("Treatment B <br> (N=", bign2, ")"))

) %>%


  tab_options(

    table.border.top.color = "white",

    heading.border.bottom.color = "black",

    table.border.bottom.color = "white",

    table_body.border.bottom.color = "black",

    table_body.hlines.color = "white",
```

```
        row_group.border.bottom.color = "white",

        row_group.border.top.color = "white",

        column_labels.border.top.color = "black",

        column_labels.border.bottom.color = "black",

    )


# as_rtf(tab_rtf, page_numbering=c("header"))


# output the HTML table


tab_rtf %>%
gtsave("demog.html", path = "C:\\demog_R" )
```

The table created with this R program is shown in display 3. It is not exactly a replicate of the table in display 2.

The author has yet to figure out the solutions to the issues listed below. People with more knowledge in R programming may have some ready answers.

- How to add the two titles in the top showing company name, page number and compound name?
- How to add indentation to the statistics under each row group?
- How to add a straight line under the column labels, without adding any other lines to the table?
- How to create the table with orientation = landscape?

Table 14.1 Demographics and Baseline Characteristics
Safety Population

statistics	Treatment A (N= 10)	Treatment B (N= 9)
Height (in)		
n	10	9
mean	61.82	62.91
sd	3.788	6.504
median	62.50	64.80
min, max	56.5 , 69.0	51.3 , 72.0
Weight (kg)		
n	10	9
mean	96.30	104.17
sd	11.708	31.243
median	98.75	112.00
min, max	83.0 , 112.5	50.5 , 150.0
Sex		
Female	5 (50)	4 (44.4)
Male	5 (50)	5 (55.6)

Note: the source data is class.

Program Source: demog.R Executed: (Draft) 2022-03-14

CONCLUSION

The package gt is a powerful tool to create good-looking tables, although the author is not able to create a table exactly like the one created in SAS.

REFERENCES

[1] Some detailed discussion about gt package, available at:

https://aosmith16.github.io/spring-r-topics/slides/week04_gt_tables.html#1

[2] Presentation by Rich Iannone, available at:

https://www.youtube.com/watch?v=h1KAjSfSbmk&t=872s

[3] the R program in this paper is available at:

https://github.com/hengweiliu2020/demographics-table-with-gt-package

CONTACT INFORMATION

Hengwei Liu
Hengrui USA
400 Alexander Park
Princeton, NJ 08540
Hengwei_liu@yahoo.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.