

# Exploring the R gtsummary Package to Create Professional-Quality Descriptive Tables for Academic Publications

Weisi Chen

2025-02-09

## Table of contents

Install and read in R packages needed . . . . .	1
Read in the demo data . . . . .	1
Example basic table . . . . .	2
Customize the table's appearance . . . . .	2
Customize the table's appearance II . . . . .	2
Code Available . . . . .	2
Code to produce the example basic table . . . . .	2
Code to produce the customized table I . . . . .	6
Code to produce the customized table II . . . . .	9

## Install and read in R packages needed

```
library(NHANES)
library(gtsummary)
library(gt)
library(dplyr)
library(purrr)
```

## Read in the demo data

```
data <- NHANES::NHANES
```

### Example basic table

## Customize the table's appearance

- **Move the total column** to the far-right end of the table for improved readability.
- **Remove the ‘N = xxxx’** from the header to streamline the table’s appearance.
- **Add a “Total (denominator)” row** at the top of the table for better context and clarity.
- **Avoid decimal places** for both numbers and percentages for a cleaner presentation.
- **Include additional summary statistics** for continuous variables, such as mean (SD), median (IQR), and range, to provide a more comprehensive summary.
- **Customize the footnotes**

## Customize the table's appearance II

- **Separate the Number and Percentage Columns:** Split the n (count) and p (percentage) values into two separate columns in the table.
- **Right-align the Number and Percentage Columns:** Apply `cell_text(align = "right")` to these columns.
- **Label the Columns as n and col%**
- **Add Colors:** Apply `cell_fill()` for background colors and/or `cell_text()` for text colors to enhance readability.

## Code Available

### Code to produce the example basic table

```
data %>%
  # Remove missing data in the Diabetes variable for simplicity
  filter(!is.na(Diabetes)) %>%
  # Select relevant variables
  select(Gender, Age, AgeDecade, Race1, BMI_WHO, Education, MaritalStatus, H
  # Create a summary table by Diabetes group
  tbl_summary(
    by = Diabetes,
    statistic = list(
```

Table 1: Sociodemographic Characteristics of Patients With and Without Diabetes in the Demo Dataset

Characteristic	Overall N = 9,858 <sup>1</sup>	No N = 9,098 <sup>1</sup>	Yes N = 760 <sup>1</sup>	p-value <sup>2</sup>
<b>Gender</b>				0.064
female	4,949 (50%)	4,592 (50%)	357 (47%)	
male	4,909 (50%)	4,506 (50%)	403 (53%)	
<b>Age</b>	37 (22)	35 (22)	59 (15)	<0.001
<b>Age group</b>				<0.001
0-9	1,254 (13%)	1,254 (14%)	0 (0%)	
10-19	1,371 (14%)	1,354 (15%)	17 (2.5%)	
20-29	1,356 (14%)	1,344 (15%)	12 (1.7%)	
30-39	1,338 (14%)	1,295 (15%)	43 (6.2%)	
40-49	1,398 (15%)	1,302 (15%)	96 (14%)	
50-59	1,304 (14%)	1,126 (13%)	178 (26%)	
60-69	917 (9.6%)	713 (8.1%)	204 (30%)	
70+	587 (6.2%)	447 (5.1%)	140 (20%)	
Unknown	333	263	70	
<b>Ethnicity</b>				<0.001
Black	1,184 (12%)	1,053 (12%)	131 (17%)	
Hispanic	602 (6.1%)	555 (6.1%)	47 (6.2%)	
Mexican	991 (10%)	925 (10%)	66 (8.7%)	
White	6,290 (64%)	5,840 (64%)	450 (59%)	
Other	791 (8.0%)	725 (8.0%)	66 (8.7%)	
<b>BMI group</b>				<0.001
12.0_18.5	1,277 (13%)	1,274 (14%)	3 (0.4%)	
18.5_to_24.9	2,908 (30%)	2,797 (32%)	111 (15%)	
25.0_to_29.9	2,664 (28%)	2,461 (28%)	203 (27%)	
30.0_plus	2,749 (29%)	2,321 (26%)	428 (57%)	
Unknown	260	245	15	
<b>Education</b>				<0.001
8th Grade	451 (6.2%)	351 (5.4%)	100 (13%)	
9 - 11th Grade	886 (12%)	781 (12%)	105 (14%)	
High School	1,517 (21%)	1,352 (21%)	165 (22%)	
Some College	2,267 (31%)	2,039 (31%)	228 (31%)	
College Grad	2,098 (29%)	1,954 (30%)	144 (19%)	
Unknown	2,639	2,621	18	
<b>MaritalStatus</b>				<0.001
Divorced	705 (9.8%)	605 (9.3%)	100 (13%)	
LivePartner	560 (7.7%)	531 (8.2%)	29 (3.9%)	
Married	3,945 (55%)	3,519 (54%)	426 (57%)	
NeverMarried	1,380 (19%)	1,313 (20%)	67 (9.0%)	
Separated	183 (2.5%)	159 (2.5%)	24 (3.2%)	
Widowed	456 (6.3%)	361 (5.6%)	95 (13%)	
Unknown	2,629	2,610	19	
<b>Household income</b>				<0.001
0-4999	182 (2.0%)	169 (2.0%)	13 (1.9%)	
5000-9999	250 (2.8%)	223 (2.7%)	27 (3.9%)	
10000-14999	537 (5.9%)	472 (5.6%)	65 (9.3%)	
15000-19999	515 (5.7%)	461 (5.5%)	54 (7.8%)	

**Table 1: Sociodemographic Characteristics of Patients With and Without Diabetes in the Demo Dataset**

	With Diabetes	Without Diabetes	Total	p-value <sup>1</sup>
<b>Total (column denominator)</b>	760 (100%)	9,098 (100%)	9,858 (100%)	
<b>Gender<sup>2</sup></b>				0.064
female	357 (47%)	4,592 (50%)	4,949 (50%)	
male	403 (53%)	4,506 (50%)	4,909 (50%)	
<b>Age<sup>3</sup></b>				<0.001
Mean, (SD)	59, (15)	35, (22)	37, (22)	
Median, (IQR)	61, (51, 70)	34, (17, 52)	37, (18, 54)	
Range	11, 80	1, 80	1, 80	
<b>Age group</b>				<0.001
0-9	0 (0%)	1,254 (14%)	1,254 (13%)	
10-19	17 (2%)	1,354 (15%)	1,371 (14%)	
20-29	12 (2%)	1,344 (15%)	1,356 (14%)	
30-39	43 (6%)	1,295 (15%)	1,338 (14%)	
40-49	96 (14%)	1,302 (15%)	1,398 (15%)	
50-59	178 (26%)	1,126 (13%)	1,304 (14%)	
60-69	204 (30%)	713 (8%)	917 (10%)	
70+	140 (20%)	447 (5%)	587 (6%)	
<b>Ethnicity</b>				<0.001
Black	131 (17%)	1,053 (12%)	1,184 (12%)	
Hispanic	47 (6%)	555 (6%)	602 (6%)	
Mexican	66 (9%)	925 (10%)	991 (10%)	
White	450 (59%)	5,840 (64%)	6,290 (64%)	
Other	66 (9%)	725 (8%)	791 (8%)	
<b>BMI group</b>				<0.001
12.0_18.5	3 (0%)	1,274 (14%)	1,277 (13%)	
18.5_to_24.9	111 (15%)	2,797 (32%)	2,908 (30%)	
25.0_to_29.9	203 (27%)	2,461 (28%)	2,664 (28%)	
30.0_plus	428 (57%)	2,321 (26%)	2,749 (29%)	
<b>Education</b>				<0.001
8th Grade	100 (13%)	351 (5%)	451 (6%)	
9 - 11th Grade	105 (14%)	781 (12%)	886 (12%)	
High School	165 (22%)	1,352 (21%)	1,517 (21%)	
Some College	228 (31%)	2,039 (31%)	2,267 (31%)	
College Grad	144 (19%)	1,954 (30%)	2,098 (29%)	
<b>MaritalStatus</b>				<0.001
Divorced	100 (13%)	605 (9%)	705 (10%)	
LivePartner	29 (4%)	531 (8%)	560 (8%)	
Married	426 (57%)	3,519 (54%)	3,945 (55%)	
NeverMarried	67 (9%)	1,313 (20%)	1,380 (19%)	
Separated	24 (3%)	159 (2%)	183 (3%)	
Widowed	95 (13%)	361 (6%)	456 (6%)	
<b>Household income</b>				<0.001
0-4999	13 (2%)	169 (2%)	182 (2%)	
5000-9999	27 (4%)	223 (3%)	250 (3%)	
10000-14999	65 (9%)	472 (6%)	537 (6%)	
15000-19999	54 (8%)	461 (6%)	515 (6%)	

**Table 1: Sociodemographic Characteristics of Patients With and Without Diabetes in the Demo Dataset**

	With Diabetes		Without Diabetes		Total	
	n	col%	n	col%	n	col%
<b>Total (column denominator)</b>	760	(100%)	9,098	(100%)	9,858	(100%)
<b>Gender<sup>1</sup></b>						
female	357	(47%)	4,592	(50%)	4,949	(50%)
male	403	(53%)	4,506	(50%)	4,909	(50%)
<b>Age<sup>2</sup></b>						
Mean (SD)	59 (15)		35 (22)		37 (22)	
Median (Q1, Q3)	61 (51, 70)		34 (17, 52)		37 (18, 54)	
Range	11, 80		1, 80		1, 80	
<b>Age group</b>						
0-9	0	(0%)	1,254	(14%)	1,254	(13%)
10-19	17	(2%)	1,354	(15%)	1,371	(14%)
20-29	12	(2%)	1,344	(15%)	1,356	(14%)
30-39	43	(6%)	1,295	(15%)	1,338	(14%)
40-49	96	(14%)	1,302	(15%)	1,398	(15%)
50-59	178	(26%)	1,126	(13%)	1,304	(14%)
60-69	204	(30%)	713	(8%)	917	(10%)
70+	140	(20%)	447	(5%)	587	(6%)
<b>Ethnicity</b>						
Black	131	(17%)	1,053	(12%)	1,184	(12%)
Hispanic	47	(6%)	555	(6%)	602	(6%)
Mexican	66	(9%)	925	(10%)	991	(10%)
White	450	(59%)	5,840	(64%)	6,290	(64%)
Other	66	(9%)	725	(8%)	791	(8%)
<b>BMI group</b>						
12.0_18.5	3	(0%)	1,274	(14%)	1,277	(13%)
18.5_to_24.9	111	(15%)	2,797	(32%)	2,908	(30%)
25.0_to_29.9	203	(27%)	2,461	(28%)	2,664	(28%)
30.0_plus	428	(57%)	2,321	(26%)	2,749	(29%)
<b>Education</b>						
8th Grade	100	(13%)	351	(5%)	451	(6%)
9 - 11th Grade	105	(14%)	781	(12%)	886	(12%)
High School	165	(22%)	1,352	(21%)	1,517	(21%)
Some College	228	(31%)	2,039	(31%)	2,267	(31%)
College Grad	144	(19%)	1,954	(30%)	2,098	(29%)
<b>MaritalStatus</b>						
Divorced	100	(13%)	605	(9%)	705	(10%)
LivePartner	29	(4%)	531	(8%)	560	(8%)
Married	426	(57%)	3,519	(54%)	3,945	(55%)
NeverMarried	67	(9%)	1,313	(20%)	1,380	(19%)
Separated	24	(3%)	159	(2%)	183	(3%)
Widowed	95	(13%)	361	(6%)	456	(6%)
<b>Household income</b>						
0-4999	13	(2%)	169	(2%)	182	(2%)
5000-9999	27	(4%)	223	(3%)	250	(3%)
10000-14999	65	(9%)	472	(6%)	537	(6%)

```

    all_continuous() ~ "{mean} ({sd})",
    all_categorical() ~ "{n} ({p}%)"
  ),
  label = list(
    AgeDecade = "Age group",
    Race1 = "Ethnicity",
    BMI_WHO = "BMI group",
    HHIncome = "Household income",
    Work = "Employment status"
  )
) %>%
add_overall() %>%
add_p() %>% # Test for differences between groups
bold_labels() %>%
modify_header(label = "**Characteristic**") %>% # Update column header
as_gt() %>%
gt::tab_header(
  "Table 1: Sociodemographic Characteristics of Patients With and Without Diabetes in 1

```

## Code to produce the customized table I

```

data %>%
  # Remove missing data in the Diabetes variable for simplicity
  filter(!is.na(Diabetes)) %>%

  # Format the Diabetes variable
  mutate(
    Diabetes = case_when(
      Diabetes == "Yes" ~ "With Diabetes",
      Diabetes == "No" ~ "Without Diabetes"
    ),
    Diabetes = factor(Diabetes, levels = c("With Diabetes", "Without Diabetes"))
  ) %>%

  # Add total number
  mutate(total = TRUE) %>%

  # Select relevant variables
  select(
    total, Gender, Age, AgeDecade, Race1, BMI_WHO, Education,

```

```

    MaritalStatus, HHIncome, Work, Diabetes
  ) %>%

# Create a summary table by Diabetes group
tbl_summary(
  by = Diabetes,
  type = all_continuous() ~ "continuous2",
  statistic = list(
    # Include additional summary statistics for continuous variables
    all_continuous() ~ c("{mean}, ({sd})",
                        "{median}, ({p25}, {p75})",
                        "{min}, {max}"),
    all_categorical() ~ "{n} ({p}%)"
  ),
  label = list(
    total = "Total (column denominator)",
    AgeDecade = "Age group",
    Race1 = "Ethnicity",
    BMI_WHO = "BMI group",
    HHIncome = "Household income",
    Work = "Employment status"
  ),
  missing = "no",

  # Remove decimal places for all numbers and percentages
  digits = list(
    all_continuous() ~ c(0, 0),
    all_categorical() ~ c(0, 0)
  )
) %>%

# Add total column
add_overall() %>%

# Move the total column to the far end of the table
modify_table_body(
  ~ .x %>%
  dplyr::relocate(stat_0, .after = stat_2) %>%

  # Change label name
  dplyr::mutate(
    label = ifelse(label == "Median, (Q1, Q3)", "Median, (IQR)", label)
  )
) %>%

```

```

    ) %>%
    dplyr::mutate(
      label = ifelse(label == "Min, Max", "Range", label)
    )
  ) %>%

# Modify the header
modify_header(
  update = list(
    all_stat_cols(TRUE) ~ "**{level}**",
    label = "",
    stat_0 = "**Total**",
    stat_1 = "**{level}**",
    stat_2 = "**{level}**"
  )
) %>%

# Test for differences between groups
add_p() %>%

# Bold labels for readability
bold_labels() %>%

# Modify footnotes
modify_footnote(
  c(all_stat_cols()) ~ NA
) %>%

# Add more footnotes to specific rows
modify_table_styling(
  columns = label,
  row = label == list("Gender"),
  footnote = "This is a sample footnote 1."
) %>%
modify_table_styling(
  columns = label,
  row = label == list("Age"),
  footnote = "This is a sample footnote 2."
) %>%

# Convert to gt table
as_gt() %>%

```



```

# Add table header with title
gt::tab_header(
  title = md("**Table 1: Sociodemographic Characteristics of Patients With and Without Dia
) %>%

# Prevent footnotes from being split across multiple lines
tab_options(footnotes.multiline = FALSE)

```

## Code to produce the customized table II

```

tab <- c("{n}", "{p}%") %>%
  map(
    ~data %>%
      # Remove missing data in the Diabetes variable for simplicity
      filter(!is.na(Diabetes)) %>%

      # Format the Diabetes variable
      mutate(
        Diabetes = case_when(
          Diabetes == "Yes" ~ "With Diabetes",
          Diabetes == "No" ~ "Without Diabetes"
        ),
        Diabetes = factor(Diabetes, levels = c("With Diabetes", "Without Diabetes"))
      ) %>%

      # Add total number
      mutate(total = TRUE) %>%

      # Select relevant variables
      select(
        total, Gender, Age, AgeDecade, Race1, BMI_WHO, Education,
        MaritalStatus, HHIncome, Work, Diabetes
      ) %>%

      # Create a summary table by Diabetes group
      tbl_summary(
        by = Diabetes,
        type = all_continuous() ~ "continuous2",
        statistic = list(
          # Include additional summary statistics for continuous variables

```

```

    all_continuous() ~ c("{mean} ({sd})",
                        "{median} ({p25}, {p75})",
                        "{min}, {max}"),
    all_categorical() ~ .x
  ),
  label = list(
    total = "Total (column denominator)",
    AgeDecade = "Age group",
    Race1 = "Ethnicity",
    BMI_WHO = "BMI group",
    HHIncome = "Household income",
    Work = "Employment status"
  ),
  missing = "no",

  # Remove decimal places for all numbers and percentages
  digits = list(
    all_continuous() ~ c(0, 0),
    all_categorical() ~ c(0, 0)
  )
) %>%

# Add total column
add_overall() %>%

# Bold labels for readability
bold_labels() %>%
tbl_merge() %>%
modify_spanning_header(everything()~NA) %>%

# Re-arrange the number and percentage columns
modify_table_body(
  ~ .x %>%
  dplyr::relocate(stat_1_2, .after=stat_1_1) %>%
  dplyr::relocate(stat_2_2, .after=stat_2_1) %>%
  dplyr::relocate(stat_0_1, .after=stat_2_2) %>%
  dplyr::relocate(stat_0_2, .after=stat_0_1)
) %>%

# Change label name
dplyr::mutate(
  label = ifelse(label == "Median, (Q1, Q3)", "Median, (IQR)", label)
) %>%

```

```

dplyr::mutate(
  label = ifelse(label == "Min, Max", "Range", label)
) %>%

# Remove the summary statistics for the continuous variable in the % column
dplyr::mutate(
  stat_0_2 = ifelse(label == "Mean (SD)", "", stat_0_2 ),
  stat_0_2 = ifelse(label == "Median (Q1, Q3)", "", stat_0_2 ),
  stat_0_2 = ifelse(label == "Range", "", stat_0_2 ),
  stat_1_2 = ifelse(label == "Mean (SD)", "", stat_1_2 ),
  stat_1_2 = ifelse(label == "Median (Q1, Q3)", "", stat_1_2 ),
  stat_1_2 = ifelse(label == "Range", "", stat_1_2 ),
  stat_2_2 = ifelse(label == "Mean (SD)", "", stat_2_2 ),
  stat_2_2 = ifelse(label == "Median (Q1, Q3)", "", stat_2_2 ),
  stat_2_2 = ifelse(label == "Range", "", stat_2_2 ),
)
) %>%

# Modify the header
modify_header(
  update = list(
    all_stat_cols(TRUE) ~ "**{level}**",
    label = "",
    stat_0_1 = "**n**",
    stat_0_2 = "**col%**",
    stat_1_1 = "**n**",
    stat_1_2 = "**col%**",
    stat_2_1 = "**n**",
    stat_2_2 = "**col%**"
  )
) %>%

# Modify spanning header
modify_spanning_header(
  stat_1_1:stat_1_2 ~ "**With Diabetes**",
  stat_2_1:stat_2_2 ~ "**Without Diabetes**",
  stat_0_1:stat_0_2 ~ "**Total**",
) %>%

# Modify footnotes
modify_footnote(
  c(all_stat_cols()) ~ NA

```

```

) %>%

# Add more footnotes to specific rows
modify_table_styling(
  columns = label,
  row = label == list("Gender"),
  footnote = "This is a sample footnote 1."
) %>%
modify_table_styling(
  columns = label,
  row = label == list("Age"),
  footnote = "This is a sample footnote 2."
) %>%

# Convert to gt table
as_gt() %>%

# Add table header with title
gt::tab_header(
  title = md("**Table 1: Sociodemographic Characteristics of Patients With and Without Dialysis")
) %>%

# Prevent footnotes from being split across multiple lines
tab_options(footnotes.multiline = FALSE) %>%

# Right-align all columns except the label column
tab_style(
  style = cell_text(align = "center"),
  locations = cells_column_labels(
    columns = everything()
  )
) %>%
tab_style(
  style = cell_text(align = "right"),
  locations = cells_body(
    columns = !label
  )
)

# Adding some colors to the tables
tab %>%

```

```

tab_style(
  style = cell_fill(color = "#E8E4E6"), # Apply the background color
  locations = cells_body(
    rows = seq(2, nrow(tab$`_data`), by = 2) # Select every second row (alternating)
  )
) %>%
tab_style(
  style = cell_fill(color = "#DAE9F7"),
  locations = cells_column_labels()
) %>%
tab_style(
  style = cell_fill(color = "#DAE9F7"),
  locations = cells_column_spanners()
)

```