

in class activity 8

Jiaqi Wang

2025-09-25

METHOD 1 : kable

```
# ---- Load package ----  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(tidyr)  
library(palmerpenguins)
```

```
##  
## Attaching package: 'palmerpenguins'
```

```
## The following objects are masked from 'package:datasets':  
##  
## penguins, penguins_raw
```

```
library(knitr)  
  
# ---- Load and prepare data ----  
data(penguins)  
  
# Remove rows with missing data for cleaner comparison  
penguins_clean <- penguins %>%  
  filter(!is.na(sex),  
         !is.na(bill_length_mm),  
         !is.na(bill_depth_mm),  
         !is.na(body_mass_g))  
  
# Preview the data  
glimpse(penguins_clean)
```

```
## Rows: 333
## Columns: 8
## $ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel...
## $ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse...
## $ bill_length_mm <dbl> 39.1, 39.5, 40.3, 36.7, 39.3, 38.9, 39.2, 41.1, 38.6...
## $ bill_depth_mm <dbl> 18.7, 17.4, 18.0, 19.3, 20.6, 17.8, 19.6, 17.6, 21.2...
## $ flipper_length_mm <int> 181, 186, 195, 193, 190, 181, 195, 182, 191, 198, 18...
## $ body_mass_g   <int> 3750, 3800, 3250, 3450, 3650, 3625, 4675, 3200, 3800...
## $ sex           <fct> male, female, female, female, male, female, male, fe...
## $ year          <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007...
```

```

# ---- Create overall summary ----
overall_summary <- penguins_clean %>%
  summarise(
    species = "Overall",
    N = n(),
    `Bill Length (mm)` = paste0(round(mean(bill_length_mm), 1), " (",
                                round(sd(bill_length_mm), 1), ")"),
    `Bill Depth (mm)` = paste0(round(mean(bill_depth_mm), 1), " (",
                                round(sd(bill_depth_mm), 1), ")"),
    `Body Mass (g)` = paste0(round(mean(body_mass_g), 0), " (",
                              round(sd(body_mass_g), 0), ")"),
    `Female (%)` = paste0(sum(sex == "female"), " (",
                          round(100*sum(sex=="female")/n(), 1), "%"),
    `Male (%)` = paste0(sum(sex == "male"), " (",
                        round(100*sum(sex=="male")/n(), 1), "%")
  )

# ---- Create summary by species ----
summary_by_species <- penguins_clean %>%
  group_by(species) %>%
  summarise(
    N = n(),
    `Bill Length (mm)` = paste0(round(mean(bill_length_mm), 1), " (",
                                round(sd(bill_length_mm), 1), ")"),
    `Bill Depth (mm)` = paste0(round(mean(bill_depth_mm), 1), " (",
                                round(sd(bill_depth_mm), 1), ")"),
    `Body Mass (g)` = paste0(round(mean(body_mass_g), 0), " (",
                              round(sd(body_mass_g), 0), ")"),
    `Female (%)` = paste0(sum(sex == "female"), " (",
                          round(100*sum(sex=="female")/n(), 1), "%"),
    `Male (%)` = paste0(sum(sex == "male"), " (",
                        round(100*sum(sex=="male")/n(), 1), "%"),
    .groups = "drop"
  )

# ---- Combine and transpose ----
summary_stats <- bind_rows(overall_summary, summary_by_species)

# Create column names (species names)
column_names <- c("Variable", summary_stats$species)

# Transpose: make species columns, variables rows
summary_transposed <- summary_stats %>%
  select(-species) %>%
  t() %>%
  as.data.frame() %>%
  tibble::rownames_to_column(var = "Variable")

colnames(summary_transposed) <- column_names

# ---- Create kable table ----
kable_table <- summary_transposed %>%
  kable(caption = "Table 1. Baseline Characteristics by Penguin Species",
        row.names = FALSE)

```

`kable_table`

Table 1. Baseline Characteristics by Penguin Species

Variable	Overall	Adelie	Chinstrap	Gentoo
N	333	146	68	119
Bill Length (mm)	44 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)
Bill Depth (mm)	17.2 (2)	18.3 (1.2)	18.4 (1.1)	15 (1)
Body Mass (g)	4207 (805)	3706 (459)	3733 (384)	5092 (501)
Female (%)	165 (49.5)	73 (50)	34 (50)	58 (48.7)
Male (%)	168 (50.5)	73 (50)	34 (50)	61 (51.3)

METHOD 2: gt

```
library(gt)
summary_transposed %>%
  gt() %>%
  tab_header(title = "Table 1. Baseline Characteristics by Penguin Species")
```

Table 1. Baseline Characteristics by Penguin Species				
Variable	Overall	Adelie	Chinstrap	Gentoo
N	333	146	68	119
Bill Length (mm)	44 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)
Bill Depth (mm)	17.2 (2)	18.3 (1.2)	18.4 (1.1)	15 (1)
Body Mass (g)	4207 (805)	3706 (459)	3733 (384)	5092 (501)
Female (%)	165 (49.5)	73 (50)	34 (50)	58 (48.7)
Male (%)	168 (50.5)	73 (50)	34 (50)	61 (51.3)

METHOD 3: gtsummary

```
library(gtsummary)

gtsummary_table <- penguins_clean %>%
  select(species, bill_length_mm, bill_depth_mm, body_mass_g, sex) %>%
  tbl_summary(
    by = species,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    )
  ) %>%
  add_overall() %>%
  modify_header(label ~ "**Variable**") %>%
  modify_caption("**Table 1. Baseline Characteristics by Penguin Species**")

gtsummary_table
```

Table 1. Baseline Characteristics by Penguin Species

Variable	Overall N = 333 ¹	Adelie N = 146 ¹	Chinstrap N = 68 ¹	Gentoo N = 119 ¹
bill_length_mm	44.0 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)
bill_depth_mm	17.16 (1.97)	18.35 (1.22)	18.42 (1.14)	15.00 (0.99)
body_mass_g	4,207 (805)	3,706 (459)	3,733 (384)	5,092 (501)
sex				
female	165 (50%)	73 (50%)	34 (50%)	58 (49%)
male	168 (50%)	73 (50%)	34 (50%)	61 (51%)

¹ Mean (SD); n (%)

1. Which package was easiest to use? The gtsummary package was the easiest to use. With a single call to `tbl_summary()`, it automatically produces a professional Table 1 that includes means, standard deviations, counts, and percentages. It also allows the addition of an overall column and p-values with minimal extra code.
2. Which package produced the most professional output? The gt package produced the most professional-looking tables. Its design is publication-ready and highly customizable in terms of fonts, alignment, headers, footnotes, and export options.
3. Which package gave the most control over formatting? The combination of manual summarization + kable provided the most control over formatting because every number and label was calculated and arranged by hand. However, this also means it is the most time-consuming approach. Among the table-focused packages, gt provides the broadest set of formatting tools.
4. Which package would you use for quick exploratory analysis? For quick exploratory analysis, gtsummary or kable are best. gtsummary is ideal when you need descriptive statistics quickly, while kable is very lightweight for simply displaying data frames during initial data exploration.

key for METHOD 1 : kable

```
# Summary by species
summary_by_species <- penguins_clean %>%
  group_by(species) %>%
  summarise(
    N = n(),
    `Bill Length (mm)` = paste0(round(mean(bill_length_mm), 1), " (",
                                round(sd(bill_length_mm), 1), ")"),
    `Bill Depth (mm)` = paste0(round(mean(bill_depth_mm), 1), " (",
                                round(sd(bill_depth_mm), 1), ")"),
    `Body Mass (g)` = paste0(round(mean(body_mass_g), 0), " (",
                              round(sd(body_mass_g), 0), ")"),
    `Female (%)` = paste0(sum(sex == "female"), " (",
                          round(100 * sum(sex == "female") / n(), 1), "%"),
    `Male (%)` = paste0(sum(sex == "male"), " (",
                        round(100 * sum(sex == "male") / n(), 1), "%")
  )

# Overall summary
overall_summary <- penguins_clean %>%
  summarise(
    species = "Overall",
    N = n(),
    `Bill Length (mm)` = paste0(round(mean(bill_length_mm), 1), " (",
                                round(sd(bill_length_mm), 1), ")"),
    `Bill Depth (mm)` = paste0(round(mean(bill_depth_mm), 1), " (",
                                round(sd(bill_depth_mm), 1), ")"),
    `Body Mass (g)` = paste0(round(mean(body_mass_g), 0), " (",
                              round(sd(body_mass_g), 0), ")"),
    `Female (%)` = paste0(sum(sex == "female"), " (",
                          round(100 * sum(sex == "female") / n(), 1), "%"),
    `Male (%)` = paste0(sum(sex == "male"), " (",
                        round(100 * sum(sex == "male") / n(), 1), "%")
  )

# Combine & transpose
summary_stats <- bind_rows(overall_summary, summary_by_species)
column_names <- c("Variable", summary_stats$species)

summary_transposed <- summary_stats %>%
  select(-species) %>%
  t() %>%
  as.data.frame() %>%
  tibble::rownames_to_column(var = "Variable")

colnames(summary_transposed) <- column_names

# Table with kable
summary_transposed %>%
  knitr::kable(caption = "Table 1. Baseline Characteristics by Penguin Species",
               row.names = FALSE)
```

Table 1. Baseline Characteristics by Penguin Species

Variable	Overall	Adelie	Chinstrap	Gentoo
N	333	146	68	119
Bill Length (mm)	44 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)
Bill Depth (mm)	17.2 (2)	18.3 (1.2)	18.4 (1.1)	15 (1)
Body Mass (g)	4207 (805)	3706 (459)	3733 (384)	5092 (501)
Female (%)	165 (49.5)	73 (50)	34 (50)	58 (48.7)
Male (%)	168 (50.5)	73 (50)	34 (50)	61 (51.3)

key for METHOD 2: gt

```
library(gt)

gt_table <- summary_transposed %>%
  gt() %>%
  tab_header(
    title = "Baseline Characteristics by Penguin Species",
    subtitle = "Palmer Station Antarctica LTER Dataset"
  ) %>%
  cols_label(
    Variable = "Characteristic",
    Overall = "Overall",
    Adelie = "Adelie",
    Chinstrap = "Chinstrap",
    Gentoo = "Gentoo"
  ) %>%
  tab_spanner(
    label = "Mean (SD)",
    columns = c(Overall, Adelie, Chinstrap, Gentoo),
    gather = FALSE
  ) %>%
  tab_style(
    style = list(cell_fill(color = "#f8f9fa"),
                  cell_text(weight = "bold")),
    locations = cells_body(columns = Variable)
  ) %>%
  tab_style(
    style = list(cell_fill(color = "#e3f2fd"),
                  cell_text(weight = "bold")),
    locations = cells_body(columns = Overall)
  ) %>%
  tab_footnote(
    footnote = "Data represents complete cases only",
    locations = cells_title(groups = "subtitle")
  ) %>%
  tab_options(
    heading.background.color = "#f8f9fa",
    column_labels.font.weight = "bold",
    table.font.size = 12
  )

gt_table
```

Baseline Characteristics by Penguin Species				
Palmer Station Antarctica LTER Dataset ¹				
Characteristic	Mean (SD)			
	Overall	Adelie	Chinstrap	Gentoo
N	333	146	68	119
Bill Length (mm)	44 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)

¹ Data represents complete cases only

Baseline Characteristics by Penguin Species

Palmer Station Antarctica LTER Dataset[†]

Characteristic	Mean (SD)			
	Overall	Adelie	Chinstrap	Gentoo
Bill Depth (mm)	17.2 (2)	18.3 (1.2)	18.4 (1.1)	15 (1)
Body Mass (g)	4207 (805)	3706 (459)	3733 (384)	5092 (501)
Female (%)	165 (49.5)	73 (50)	34 (50)	58 (48.7)
Male (%)	168 (50.5)	73 (50)	34 (50)	61 (51.3)

[†] Data represents complete cases only

key for METHOD 3: gtsummary

```
library(gtsummary)

gtsummary_table <- penguins_clean %>%
  select(species, bill_length_mm, bill_depth_mm, body_mass_g, sex) %>%
  tbl_summary(
    by = species,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    ),
    digits = list(
      bill_length_mm ~ c(1, 1),
      bill_depth_mm ~ c(1, 1),
      body_mass_g ~ c(0, 0),
      sex ~ c(0, 1)
    ),
    label = list(
      bill_length_mm ~ "Bill Length (mm)",
      bill_depth_mm ~ "Bill Depth (mm)",
      body_mass_g ~ "Body Mass (g)",
      sex ~ "Sex"
    ),
    missing = "no"
  ) %>%
  add_overall(col_label = "**Overall**") %>%
  modify_header(all_stat_cols() ~ "**{level}**") %>%
  modify_caption("**Table 1. Baseline Characteristics by Penguin Species**") %>%
  modify_footnote(
    all_stat_cols() ~ "Mean (SD) for continuous variables; n (%) for categorical variables"
  ) %>%
  modify_footnote(
    update = all_stat_cols() ~ "Data from Palmer Station Antarctica LTER Dataset",
    abbreviation = TRUE
  )
```

```
## Warning: The `update` argument of `modify_footnote()` is deprecated as of gtsummary
## 2.0.0.
## i Use `modify_footnote(...)` input instead. Dynamic dots allow for syntax like
## `modify_footnote(!!!list(...))`.
## i The deprecated feature was likely used in the gtsummary package.
## Please report the issue at <https://github.com/ddsjoberg/gtsummary/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

gtsummary_table

Table 1. Baseline Characteristics by Penguin Species

Characteristic	Overall ¹	Adelie ¹	Chinstrap ¹	Gentoo ¹
Bill Length (mm)	44.0 (5.5)	38.8 (2.7)	48.8 (3.3)	47.6 (3.1)
Bill Depth (mm)	17.2 (2.0)	18.3 (1.2)	18.4 (1.1)	15.0 (1.0)
Body Mass (g)	4,207 (805)	3,706 (459)	3,733 (384)	5,092 (501)
Sex				
female	165 (49.5%)	73 (50.0%)	34 (50.0%)	58 (48.7%)
male	168 (50.5%)	73 (50.0%)	34 (50.0%)	61 (51.3%)

¹ Mean (SD) for continuous variables; n (%) for categorical variables

Abbreviation: Data from Palmer Station Antarctica LTER Dataset