Summarize Data

library(apa7)  
library(ftExtra)  
library(flextable)  
library(gt)  
library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(stringr)  
library(purrr)

The flextable and gt packages have many, many options and helper functions. The tabulator, summarizor, and proc\_freq functions from the flextable package are particularly powerful ways of summarizing and describing data. Functions that automate your descriptive statistics tables for you (e.g., the amazing [gtsummary package](https://www.danieldsjoberg.com/gtsummary/index.html)) inevitably entail some level of compromise. It is unreasonable for a single function to anticipate the diversity of needs out there. Sometimes you need a table to be a particular way, and you need to do the heaving lifting yourself. After all that, the apa\_style function will get a flextable or gt table close to APA style. You may need to do some additional styling with flextable or gt as well.

Here I create a table of means, standard deviations, and sample sizes for several variables across groups. The particulars are not so important here, just that flextable and gt can do most of the rest after that.

# Data Setup

d <- diamonds %>%   
 select(cut, price, carat, depth, table) %>%   
 arrange(cut) %>%   
 rename\_with(str\_to\_title) %>%   
 pivot\_longer(where(is.numeric), names\_to = "Variable") %>%   
 summarise(  
 M = mean(value, na.rm = TRUE),  
 SD = sd(value, na.rm = TRUE),  
 n = n(),  
 .by = c(Variable, Cut)) %>%   
 pivot\_longer(c(M, SD)) %>%   
 unite(Variable, Variable, name) %>%  
 pivot\_wider(names\_from = Variable)

# Flextable

The flextable package has a theme\_apa function that can get things very close to full APA style. With a little extra care, the formatting in [Table 1](#tbl-flextable) is pretty much perfect.

flextable::flextable(d) |>  
 theme\_apa() |>  
 flextable::separate\_header() |>  
 colformat\_double(j = 3:4,  
 prefix = "$",  
 digits = 2) |>  
 italic(j = -1, i = 2, part = "header") |>  
 italic(j = 2, i = 1, part = "header") |>  
 align(align = "center", part = "header") |>  
 align(j = 1) |>  
 surround(  
 i = 2,  
 part = "header",  
 border.top = flextable::fp\_border\_default(  
 color = "black",  
 width = 1)  
 )

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Price, Carat, Depth, Table Using flextable   | Cut | *n* | Price | | Carat | | Depth | | Table | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | | Fair | 1,610 | $4,358.76 | $3,560.39 | 1.05 | 0.52 | 64.04 | 3.64 | 59.05 | 3.95 | | Good | 4,906 | $3,928.86 | $3,681.59 | 0.85 | 0.45 | 62.37 | 2.17 | 58.69 | 2.85 | | Very Good | 12,082 | $3,981.76 | $3,935.86 | 0.81 | 0.46 | 61.82 | 1.38 | 57.96 | 2.12 | | Premium | 13,791 | $4,584.26 | $4,349.20 | 0.89 | 0.52 | 61.26 | 1.16 | 58.75 | 1.48 | | Ideal | 21,551 | $3,457.54 | $3,808.40 | 0.70 | 0.43 | 61.71 | 0.72 | 55.95 | 1.25 | |

## The apa\_flextable function

I prefer related columns to separated by a small break in the headers, so I use the add\_break\_columns function to insert the breaks after each column ending with \_SD (omitting the last instance). The apa\_flextable function handles the rest automatically except for the dollar signs, which flextable can add later. [Table 2](#tbl-apaflextable) also decimal aligns all the numbers, though this is apparent in column n only.

d |>   
 add\_break\_columns(ends\_with("\_SD"), omit\_last = TRUE) |>  
 apa\_flextable() |>   
 colformat\_char(j = 3:4, prefix = "$")

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Price, Carat, Depth, Table Using apa\_flextable   | Cut | *n* | Price | |  | Carat | |  | Depth | |  | Table | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | | Fair | 1,610 | $4,358.76 | $3,560.39 |  | 1.05 | 0.52 |  | 64.04 | 3.64 |  | 59.05 | 3.95 | | Good | 4,906 | $3,928.86 | $3,681.59 |  | 0.85 | 0.45 |  | 62.37 | 2.17 |  | 58.69 | 2.85 | | Very Good | 12,082 | $3,981.76 | $3,935.86 |  | 0.81 | 0.46 |  | 61.82 | 1.38 |  | 57.96 | 2.12 | | Premium | 13,791 | $4,584.26 | $4,349.20 |  | 0.89 | 0.52 |  | 61.26 | 1.16 |  | 58.75 | 1.48 | | Ideal | 21,551 | $3,457.54 | $3,808.40 |  | 0.70 | 0.43 |  | 61.71 | 0.72 |  | 55.95 | 1.25 | |

# gt

The gt package is not really supported by the apa7 package, but I include this here because I did not want to delete my initial efforts. Getting gt to have APA style borders took me a long time to figure out because of the complex rules by which borders overwrite each other. The apa\_style function saves me from having to figure it all out repeatedly when working with gt tables.

gt(d) %>%  
 fmt\_number(decimals = 2, columns = is.double) %>%  
 fmt\_number(columns = is.integer, decimals = 0) %>%  
 fmt\_currency(columns = contains("Price")) %>%   
 tab\_spanner\_delim(delim = "\_") %>%  
 cols\_align(align = "center", columns = -1) %>%  
 cols\_align(align = "left", columns = "Cut") %>%  
 tab\_style(cell\_text(style = "italic"),   
 locations = cells\_column\_labels(-Cut)) %>%  
 tab\_style(cell\_text(align = "right"),   
 locations = cells\_body(-Cut)) %>%  
 apa\_style(horizontal\_padding = 10) %>%   
 tab\_options(table.align = "left")

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Price, Carat, Depth, Table Using gt   |  |  | Price | | Carat | | Depth | | Table | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Cut | *n* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | | Fair | 1,610 | $4,358.76 | $3,560.39 | 1.05 | 0.52 | 64.04 | 3.64 | 59.05 | 3.95 | | Good | 4,906 | $3,928.86 | $3,681.59 | 0.85 | 0.45 | 62.37 | 2.17 | 58.69 | 2.85 | | Very Good | 12,082 | $3,981.76 | $3,935.86 | 0.81 | 0.46 | 61.82 | 1.38 | 57.96 | 2.12 | | Premium | 13,791 | $4,584.26 | $4,349.20 | 0.89 | 0.52 | 61.26 | 1.16 | 58.75 | 1.48 | | Ideal | 21,551 | $3,457.54 | $3,808.40 | 0.70 | 0.43 | 61.71 | 0.72 | 55.95 | 1.25 | |