zztable1: Advanced Publication-Ready Summary Tables

zztable1 Development Team

2026-02-18

Introduction

The `zztable1` package provides a next-generation architecture for creating publication-ready summary tables (commonly called "Table 1") used in biomedical research and clinical trials. This vignette demonstrates the key features and capabilities of the package.

Key Features

- **Lazy Evaluation Architecture**: Fast blueprint creation with computation on demand
- **Journal-Specific Theming**: NEJM, Lancet, JAMA, BMJ formatting styles
- **Advanced Footnote System**: Variable-specific, column-specific, and general footnotes with superscript markers
- **Multiple Output Formats**: Console, LaTeX, and HTML output with proper column headers
- **Flexible Statistics**: Built-in and custom summary statistics
- **Stratified Analysis**: Support for subgroup analyses
- **Full Compatibility**: Same interface as original zztable1 package
- **R Markdown Integration**: Automatic format detection for seamless PDF/HTML output

Installation and Setup

```
# Development version (when available)
# devtools::install_github("user/zztable1")

# For this vignette, source the development file
source("zztable1.R")
```

Basic Usage

Simple Summary Tables

Let's start with a basic example using the mtcars dataset:

```
# Prepare data
data(mtcars)
mtcars$transmission <- factor(
  ifelse(mtcars$am == 1, "Manual", "Automatic"),
  levels = c("Automatic", "Manual")
)
mtcars$engine_type <- factor(
  ifelse(mtcars$vs == 1, "V-shaped", "Straight"),
  levels = c("Straight", "V-shaped")
)
```

```r
# Create basic summary table
create_table(transmission ~ mpg + hp + wt, data = mtcars)
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798
wt	3.8 (0.8)	2.4 (0.6)	0

Adding Statistical Tests

Include p-values for group comparisons:

```r
create_table(transmission ~ mpg + hp + wt,
             data = mtcars,
             pvalue = TRUE)
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798
wt	3.8 (0.8)	2.4 (0.6)	0

Including Total Column

Add an overall summary column:

```r
create_table(transmission ~ mpg + hp + wt,
             data = mtcars,
             pvalue = TRUE,
             totals = TRUE)
```

variables	Automatic	Manual	Total	p.value
mpg	17.1 (3.8)	24.4 (6.2)	20.1 (6)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	146.7 (68.6)	0.1798
wt	3.8 (0.8)	2.4 (0.6)	3.2 (1)	0

Advanced Features

Custom Numeric Summaries

Built-in Options

The package provides several built-in summary statistics:

```r
# Default: Mean (SD)
cat("Mean (SD) format:\n")
```

Mean (SD) format:

```r
create_table(transmission ~ mpg + hp, data = mtcars)
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798

```r
# Median [IQR]
cat("\nMedian [IQR] format:\n")
```

Median [IQR] format:

```
create_table(transmission ~ mpg + hp, data = mtcars,
             numeric_summary = "median_iqr")
```

variables	Automatic	Manual	p.value
mpg	17.3 [14.9-19.2]	22.8 [21-30.4]	3e-04
hp	175 [116.5-192.5]	109 [66-113]	0.1798

```
# Mean +/- SE
cat("\nMean +/- SE format:\n")
```

Mean +/- SE format:

```
create_table(transmission ~ mpg + hp, data = mtcars,
             numeric_summary = "mean_se")
```

variables	Automatic	Manual	p.value
mpg	17.1 +/- 0.9	24.4 +/- 1.7	3e-04
hp	160.3 +/- 12.4	126.8 +/- 23.3	0.1798

Custom Functions

Create your own summary statistics:

```
# Custom function: Median (Min-Max)
custom_summary <- function(x) {
  med <- round(median(x, na.rm = TRUE), 1)
  min_val <- round(min(x, na.rm = TRUE), 1)
  max_val <- round(max(x, na.rm = TRUE), 1)
  paste0(med, " (", min_val, "-", max_val, ")")
}

cat("Custom Median (Min-Max) format:\n")
```

Custom Median (Min-Max) format:

```
create_table(transmission ~ mpg + hp, data = mtcars,
             numeric_summary = custom_summary)
```

variables	Automatic	Manual	p.value
mpg	17.3 (10.4-24.4)	22.8 (15-33.9)	3e-04
hp	175 (62-245)	109 (52-335)	0.1798

Stratified Analysis

Perform subgroup analyses using stratification:

```
# Create stratification variable
mtcars$cylinder_group <- factor(
  ifelse(mtcars$cyl <= 4, "4-cylinder",
  ifelse(mtcars$cyl <= 6, "6-cylinder", "8-cylinder")),
  levels = c("4-cylinder", "6-cylinder", "8-cylinder")
)

# Stratified analysis
create_table(transmission ~ mpg + hp,
             data = mtcars,
             strata = "cylinder_group",
             pvalue = TRUE)
```

variables	Automatic	Manual	p.value
Cylinder_group: 6-cylinder			
mpg	19.1 (1.6)	20.6 (0.8)	3e-04
hp	115.2 (9.2)	131.7 (37.5)	0.1798
Cylinder_group: 4-cylinder			
mpg	22.9 (1.5)	28.1 (4.5)	3e-04
hp	84.7 (19.7)	81.9 (22.7)	0.1798
Cylinder_group: 8-cylinder			
mpg	15.1 (2.8)	15.4 (0.6)	3e-04
hp	194.2 (33.4)	299.5 (50.2)	0.1798

Journal-Specific Theming

Available Themes

View all available themes:

```
themes <- list_available_themes()
print(themes)
```

[1] “console” “nejm” “lancet” “jama” “bmj” “simple”

Theme Comparison

Default Theme

```
cat("Default Theme:\n")
```

Default Theme:

```
create_table(transmission ~ mpg + hp, data = mtcars,
             theme = "default")
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798

NEJM Theme (1 decimal place)

```
cat("NEJM Theme (1 decimal place):\n")
```

NEJM Theme (1 decimal place):

```
create_table(transmission ~ mpg + hp, data = mtcars,
             theme = "nejm")
```

variables	Automatic	Manual	p.value
mpg	17.1 ± 3.8	24.4 ± 6.2	3e-04
hp	160.3 ± 53.9	126.8 ± 84.1	0.1798

JAMA Theme (2 decimal places)

```
cat("JAMA Theme (2 decimal places):\n")
```

JAMA Theme (2 decimal places):

```
create_table(transmission ~ mpg + hp, data = mtcars,
             theme = "jama")
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798

Lancet Theme

```
cat("Lancet Theme:\n")
```

Lancet Theme:

```
create_table(transmission ~ mpg + hp, data = mtcars,
             theme = "lancet")
```

variables	Automatic	Manual	p.value
mpg	17.1 (3.8)	24.4 (6.2)	3e-04
hp	160.3 (53.9)	126.8 (84.1)	0.1798

Footnote System

Variable-Specific Footnotes

Add footnotes to specific variables with superscript markers:

```
create_table(transmission ~ mpg + hp + wt,
             data = mtcars,
             theme = "nejm",
             footnotes = list(
               variables = list(
                 mpg = "EPA fuel economy rating in miles per gallon",
                 hp = "Gross horsepower measured at crankshaft",
                 wt = "Vehicle weight in thousands of pounds"
               )
             ))
```

variables	Automatic	Manual	p.value
mpg[1]	17.1 ± 3.8	24.4 ± 6.2	3e-04
hp[2]	160.3 ± 53.9	126.8 ± 84.1	0.1798
wt[3]	3.8 ± 0.8	2.4 ± 0.6	0

[1] EPA fuel economy rating in miles per gallon
[2] Gross horsepower measured at crankshaft
[3] Vehicle weight in thousands of pounds

Column-Specific Footnotes

Add footnotes to columns:

```
create_table(transmission ~ mpg + hp,
             data = mtcars,
             theme = "nejm",
             pvalue = TRUE,
             footnotes = list(
```

```
    columns = list(
      "p.value" = "Two-tailed t-test, alpha = 0.05"
    )
  ))
```

variables	Automatic	Manual	p.value[1]
mpg	17.1 ± 3.8	24.4 ± 6.2	3e-04
hp	160.3 ± 53.9	126.8 ± 84.1	0.1798

[1] Two-tailed t-test, alpha = 0.05

Comprehensive Footnotes

Combine multiple footnote types:

```
create_table(transmission ~ mpg + hp,
             data = mtcars,
             theme = "nejm",
             pvalue = TRUE,
             footnotes = list(
               variables = list(
                 mpg = "EPA fuel economy standard",
                 hp = "Gross horsepower"
               ),
               columns = list(
                 "p.value" = "Statistical significance testing"
               ),
               general = list(
                 "Data source: Henderson and Velleman (1981)",
                 "Missing values excluded from analysis"
               )
             ))
```

variables	Automatic	Manual	p.value[3]
mpg[1]	17.1 ± 3.8	24.4 ± 6.2	3e-04
hp[2]	160.3 ± 53.9	126.8 ± 84.1	0.1798

[1] EPA fuel economy standard
[2] Gross horsepower
[3] Statistical significance testing

Clinical Trial Example

Simulated Clinical Trial Data

Let's create a more realistic clinical trial example:

```
set.seed(123)
n <- 200

# Generate clinical trial data
trial_data <- data.frame(
  patient_id = 1:n,
  treatment = factor(
    sample(c("Placebo", "Drug A", "Drug B"), n, replace = TRUE),
    levels = c("Placebo", "Drug A", "Drug B")
```

```
  ),
  age = round(rnorm(n, 65, 12)),
  sex = factor(sample(c("Male", "Female"), n, replace = TRUE)),
  race = factor(
    sample(c("White", "Black", "Hispanic", "Asian", "Other"),
           n, replace = TRUE, prob = c(0.6, 0.2, 0.1, 0.08, 0.02)),
    levels = c("White", "Black", "Hispanic", "Asian", "Other")
  ),
  baseline_bmi = round(rnorm(n, 28, 5), 1),
  diabetes = factor(sample(c("No", "Yes"), n, replace = TRUE, prob = c(0.7, 0.3))),
  hypertension = factor(sample(c("No", "Yes"), n, replace = TRUE, prob = c(0.6, 0.4))),
  center = factor(sample(paste("Center", 1:4), n, replace = TRUE))
)

# Preview the data
head(trial_data, 10)
```

patient_id treatment age sex race baseline_bmi diabetes hypertension 1 1 Drug B 70 Female Black 25.5 Yes No 2 2 Drug B 65 Male White 20.9 No Yes 3 3 Drug B 60 Male Black 28.6 No Yes 4 4 Drug A 40 Male White 37.7 Yes No 5 5 Drug B 79 Male White 32.0 Yes Yes 6 6 Drug A 47 Male White 33.8 No No 7 7 Drug A 74 Female White 29.8 No No 8 8 Drug A 88 Female White 25.0 No No 9 9 Drug B 48 Female White 27.0 Yes Yes 10 10 Placebo 73 Male White 26.6 No No center 1 Center 3 2 Center 1 3 Center 3 4 Center 4 5 Center 1 6 Center 1 7 Center 4 8 Center 2 9 Center 3 10 Center 3

Basic Clinical Table 1

```
create_table(treatment ~ age + sex + race + baseline_bmi +
             diabetes + hypertension,
             data = trial_data,
             theme = "nejm",
             pvalue = TRUE)
```

variables	Placebo	Drug A	Drug B	p.value
age	64.2 ± 9.3	66.6 ± 13.6	66.3 ± 12.3	0.2458
sex				
Female	33 (52%)	29 (41%)	34 (51%)	0.3849
Male	30 (48%)	41 (59%)	33 (49%)	
race				
White	42 (67%)	46 (66%)	40 (60%)	0.2309
Black	10 (16%)	11 (16%)	17 (25%)	
Hispanic	5 (8%)	6 (9%)	9 (13%)	
Asian	5 (8%)	7 (10%)	1 (1%)	
Other	1 (2%)	0 (0%)	0 (0%)	
baseline_bmi	27 ± 4.8	27.7 ± 5.2	28.1 ± 5	0.4149
diabetes				
No	38 (60%)	51 (73%)	45 (67%)	0.3331
Yes	25 (40%)	19 (27%)	22 (33%)	
hypertension				
No	36 (57%)	42 (60%)	30 (45%)	0.1646
Yes	27 (43%)	28 (40%)	37 (55%)	

With Footnotes and Stratification

```
create_table(treatment ~ age + sex + race + baseline_bmi +
               diabetes + hypertension,
               data = trial_data,
               strata = "center",
               theme = "nejm",
               pvalue = TRUE,
               footnotes = list(
                 variables = list(
                   age = "Age at enrollment (years)",
                   baseline_bmi = "Body mass index at baseline (kg/m²)",
                   diabetes = "Type 2 diabetes mellitus diagnosis",
                   hypertension = "Hypertension diagnosis"
                 ),
                 columns = list(
                   "p.value" = "ANOVA for continuous, chi-squared for categorical"
                 ),
                 general = list(
                   "Data are mean (SD) or n (%)",
                   "ITT population (N=200)"
                 )
               ))
```

variables	Placebo	Drug A	Drug B	p.value[5]
Center: Center 3				
age	63.2 ± 7.7	65.5 ± 12.4	64.2 ± 11.5	0.2458
sex				
Female	3 (25%)	9 (45%)	10 (62.5%)	0.3849
Male	9 (75%)	11 (55%)	6 (37.5%)	
race				
White	9 (75%)	14 (70%)	11 (68.8%)	0.2309
Black	1 (8.3%)	6 (30%)	4 (25%)	
Hispanic	0 (0%)	0 (0%)	1 (6.2%)	
Asian	2 (16.7%)	0 (0%)	0 (0%)	
Other	0 (0%)	0 (0%)	0 (0%)	
baseline_bmi	26 ± 4.2	28 ± 4.1	28.4 ± 5.1	0.4149
diabetes				
No	8 (66.7%)	15 (75%)	7 (43.8%)	0.3331
Yes	4 (33.3%)	5 (25%)	9 (56.2%)	
hypertension				
No	5 (41.7%)	13 (65%)	7 (43.8%)	0.1646
Yes	7 (58.3%)	7 (35%)	9 (56.2%)	
Center: Center 1				
age	67.2 ± 10.2	63.3 ± 10.9	68.2 ± 13	0.2458
sex				
Female	7 (50%)	10 (47.6%)	15 (62.5%)	0.3849
Male	7 (50%)	11 (52.4%)	9 (37.5%)	
race				
White	7 (50%)	12 (57.1%)	18 (75%)	0.2309
Black	5 (35.7%)	3 (14.3%)	3 (12.5%)	
Hispanic	1 (7.1%)	3 (14.3%)	3 (12.5%)	
Asian	0 (0%)	3 (14.3%)	0 (0%)	
Other	1 (7.1%)	0 (0%)	0 (0%)	
baseline_bmi	29.1 ± 3.5	27.8 ± 4.5	27.3 ± 5.1	0.4149
diabetes				
No	7 (50%)	14 (66.7%)	17 (70.8%)	0.3331
Yes	7 (50%)	7 (33.3%)	7 (29.2%)	
hypertension				
No	9 (64.3%)	15 (71.4%)	10 (41.7%)	0.1646
Yes	5 (35.7%)	6 (28.6%)	14 (58.3%)	
Center: Center 4				
age	62.2 ± 9.6	68.8 ± 13.9	66.6 ± 14.7	0.2458
sex				
Female	10 (71.4%)	6 (42.9%)	5 (55.6%)	0.3849
Male	4 (28.6%)	8 (57.1%)	4 (44.4%)	
race				
White	9 (64.3%)	10 (71.4%)	3 (33.3%)	0.2309
Black	2 (14.3%)	0 (0%)	4 (44.4%)	
Hispanic	1 (7.1%)	1 (7.1%)	1 (11.1%)	
Asian	2 (14.3%)	3 (21.4%)	1 (11.1%)	
Other	0 (0%)	0 (0%)	0 (0%)	
baseline_bmi	25.9 ± 5.9	27.2 ± 7.3	27.1 ± 6.9	0.4149
diabetes				
No	9 (64.3%)	10 (71.4%)	6 (66.7%)	0.3331
Yes	5 (35.7%)	4 (28.6%)	3 (33.3%)	
hypertension				
No	9 (64.3%)	7 (50%)	4 (44.4%)	0.1646
Yes	5 (35.7%)	7 (50%)	5 (55.6%)	
Center: Center 2				
age	64 ± 9.5	70.5 ± 17.8	65.3 ± 11.3	0.2458
sex				
Female	13 (56.5%)	4 (26.7%)	4 (22.2%)	0.3849
Male	10 (43.5%)	11 (73.3%)	14 (77.8%)	

Different Output Formats

Console Output (Default)

```
create_table(transmission ~ mpg + hp, data = mtcars, theme = "nejm")
```

variables	Automatic	Manual	p.value
mpg	17.1 ± 3.8	24.4 ± 6.2	3e-04
hp	160.3 ± 53.9	126.8 ± 84.1	0.1798

LaTeX Output

```
bp_latex <- table1(transmission ~ mpg + hp, data = mtcars,
                   layout = "latex", theme = "nejm")

# Note: LaTeX output would contain LaTeX markup
cat("LaTeX theme config:\n")
```

LaTeX theme config:

```
cat("Font size:", bp_latex$metadata$theme$latex$font_size, "\n")
```

Font size:

```
cat("Packages:", paste(bp_latex$metadata$theme$latex$packages, collapse = ", "), "\n")
```

Packages:

HTML Output

```
bp_html <- table1(transmission ~ mpg + hp, data = mtcars,
                  layout = "html", theme = "nejm")

# Note: HTML output would contain HTML markup
cat("HTML theme ready for web display\n")
```

HTML theme ready for web display

Performance and Architecture

Blueprint Architecture

The lazy evaluation approach provides several benefits:

```
# Large dataset simulation
large_data <- data.frame(
  group = factor(sample(c("A", "B", "C"), 10000, replace = TRUE)),
  var1 = rnorm(10000),
  var2 = rnorm(10000),
  var3 = rnorm(10000),
  var4 = rnorm(10000),
  var5 = rnorm(10000)
)

# Fast blueprint creation (no computations yet)
```

```r
system.time({
  bp_large <- table1(group ~ var1 + var2 + var3 + var4 + var5,
                     data = large_data)
})
```

user system elapsed 0.002 0.000 0.002

```r
# Computations happen only during display
cat("Blueprint created instantly. Computations happen during display.\n")
```

Blueprint created instantly. Computations happen during display.

```r
cat("Blueprint dimensions:", dim(bp_large), "\n")
```

Blueprint dimensions: 5 5

Memory Efficiency

```r
# Blueprint object structure
bp_small <- table1(transmission ~ mpg, data = mtcars)

cat("Blueprint components:\n")
```

Blueprint components:

```r
cat("- Cells: ", length(bp_small$cells), "\n")
```

- Cells: 4

```r
cat("- Dimensions: ", dim(bp_small), "\n")
```

- Dimensions: 1 4

```r
cat("- Metadata keys: ", names(bp_small$metadata), "\n")
```

- Metadata keys: formula options data_info data dimensions footnote_markers footnote_list created optimized version cell_count theme stat_cache

Best Practices

Recommendations

1. **Choose Appropriate Themes**: Use journal-specific themes for manuscript preparation
2. **Add Informative Footnotes**: Explain variables and statistical methods
3. **Use Stratification Wisely**: For meaningful subgroup analyses
4. **Custom Functions**: Create domain-specific summary statistics
5. **Validate Results**: Check statistical assumptions and interpret p-values carefully

Common Patterns

```r
# Standard clinical trial baseline table
create_baseline_table <- function(data, treatment_var, theme = "nejm") {
  formula_str <- paste(treatment_var, "~ .")
  bp <- table1(as.formula(formula_str),
               data = data,
               theme = theme,
```

```
                pvalue = TRUE,
                footnotes = list(
                  general = list(
                    "Data are mean (SD) or n (%)",
                    "P-values from ANOVA or chi-squared test"
                  )
                ))
  return(bp)
}

# Example usage
# bp_standard <- create_baseline_table(trial_data, "treatment")
cat("Utility function created for standardized baseline tables\n")
```

Utility function created for standardized baseline tables

Troubleshooting

Common Issues

1. **Missing Variables**: Ensure all formula variables exist in the data
2. **Factor Levels**: Check factor level ordering for expected display
3. **Missing Values**: Use `missing = TRUE` to show missing counts
4. **Theme Application**: Themes affect decimal places and formatting
5. **Large Tables**: Use stratification to break down complex tables

Error Handling

```
# Example of error handling
tryCatch({
  # This will cause an error - variable doesn't exist
  bp_error <- table1(nonexistent_var ~ mpg, data = mtcars)
}, error = function(e) {
  cat("Error caught:", e$message, "\n")
  cat("Solution: Check that all variables in formula exist in data\n")
})
```

Error caught: Variables not found in data: nonexistent_var

Available variables: mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb, transmission, engine_type, cylinder_group Solution: Check that all variables in formula exist in data

Conclusion

The `zztable1` package provides a powerful, flexible system for creating publication-ready summary tables. Key advantages include:

- **Performance**: Lazy evaluation for fast blueprint creation
- **Flexibility**: Multiple themes, custom statistics, advanced footnotes
- **Compatibility**: Same interface as original zztable1
- **Publication-Ready**: Journal-specific formatting out of the box

For more information, see the package documentation and function help files.

Session Information

```
sessionInfo()
```

R version 4.5.2 (2025-10-31) Platform: aarch64-apple-darwin20 Running under: macOS Tahoe 26.2

Matrix products: default BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.frame
LAPACK: /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib; LA-
PACK version 3.12.1

locale: [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: America/Los_Angeles tzcode source: internal

attached base packages: [1] stats graphics grDevices utils datasets methods base

other attached packages: [1] kableExtra_1.4.0 htmltools_0.5.9

loaded via a namespace (and not attached): [1] dplyr_1.2.0 compiler_4.5.2 tinytex_0.58
[4] tidyselect_1.2.1 xml2_1.5.2 stringr_1.6.0
[7] parallel_4.5.2 snakecase_0.11.1 tidyr_1.3.1
[10] textshaping_1.0.4 systemfonts_1.3.1 scales_1.4.0
[13] yaml_2.3.12 fastmap_1.2.0 R6_2.6.1
[16] generics_0.1.4 knitr_1.51 berryFunctions_1.22.13 [19] backports_1.5.0 zztable1_0.1.0 tibble_3.3.1
[22] janitor_2.2.1 svglite_2.2.2 lubridate_1.9.4
[25] RColorBrewer_1.1-3 pillar_1.11.1 rlang_1.1.7
[28] broom_1.0.11 stringi_1.8.7 xfun_0.56
[31] otel_0.2.0 viridisLite_0.4.3 timechange_0.3.0
[34] cli_3.6.5 magrittr_2.0.4 digest_0.6.39
[37] rstudioapi_0.18.0 lifecycle_1.0.5 vctrs_0.7.1
[40] evaluate_1.0.5 glue_1.8.0 farver_2.1.2
[43] abind_1.4-8 rmarkdown_2.30 purrr_1.2.1
[46] tools_4.5.2 pkgconfig_2.0.3