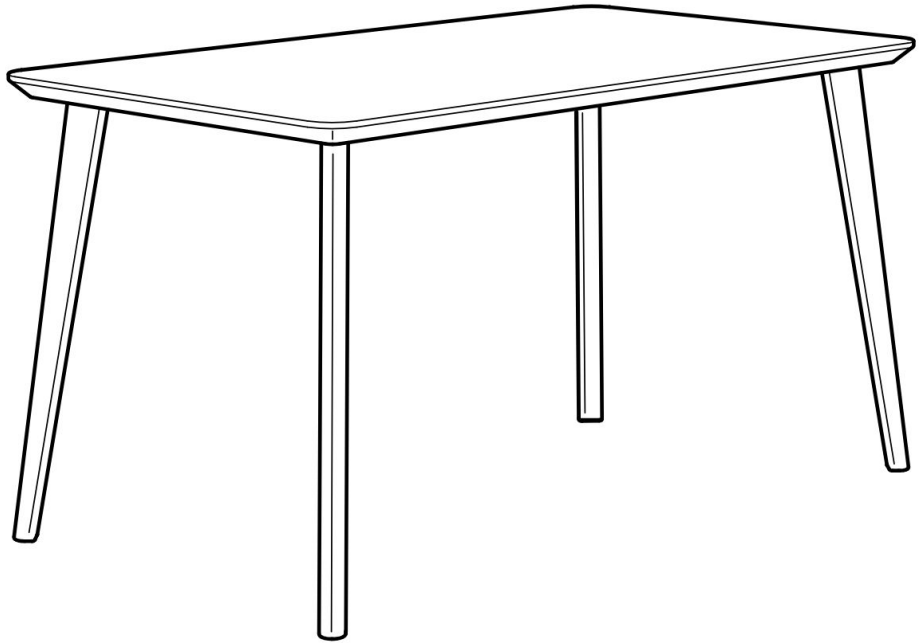
TABLE TALK



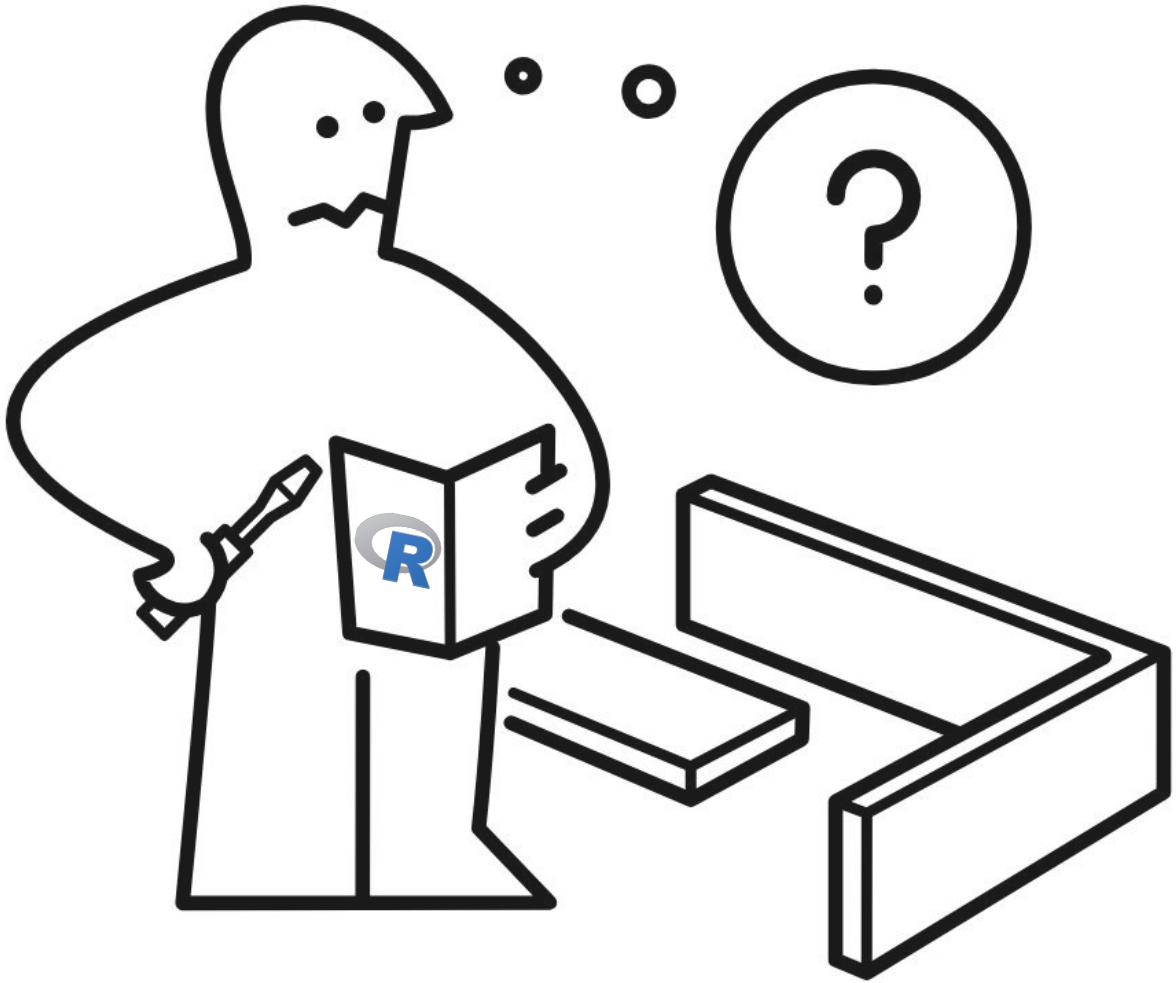
Reiko Okamoto®

Bruyère Research Institute
9 July, 2024
useR! Salzburg

TABLE 1. Baseline Characteristics of Hospitalized Home Care Recipients in Ontario From 2010 to 2015, by Linguistic Group

Baseline Characteristic	Anglophone (n = 156,186)	Francophone (n = 5110)	Allophone (n = 29,428)	*P**
Age at admission, mean ± SD, y	76.4 ± 13.4	78.1 ± 11.5	80.0 ± 10.4	<0.001
Sex, no. (%)				0.002
Female, no. (%)	88,786 (56.8)	2962 (58.0)	16,451 (55.9)	
Male, no. (%)	67,400 (43.2)	2148 (42.0)	12,977 (44.1)	
Marital status, no. (%)				<0.001
Not married	87,794 (56.2)	3040 (59.5)	15,238 (51.8)	
Married or common-law	65,558 (42.0)	1992 (39.0)	14,018 (47.6)	
Other	2834 (1.8)	78 (1.5)	172 (0.6)	
Education, no. (%)				<0.001
Less than high school	38,851 (24.9)	2499 (48.9)	11,921 (40.5)	
High school	25,982 (16.6)	527 (10.3)	2230 (7.6)	
Some postsecondary	21,163 (13.5)	455 (8.9)	2109 (7.2)	
University graduate	14,801 (9.5)	294 (5.8)	1510 (5.1)	
Missing	55,389 (35.5)	1335 (26.1)	11,658 (39.6)	
Income quintile, no. (%)				<0.001
1 (lowest)	37,254 (23.9)	1453 (28.4)	7491 (25.5)	
2	32,332 (20.7)	1162 (22.7)	6928 (23.5)	
3	30,363 (19.4)	1008 (19.7)	5715 (19.4)	
4	28,721 (18.4)	826 (16.2)	5330 (18.1)	
5 (highest)	26,765 (17.1)	645 (12.6)	3826 (13.0)	
Missing	751 (0.5)	17 (0.3)	138 (0.5)	
Urban/rural residence, no. (%)				<0.001
Urban	129,918 (83.1)	3554 (69.5)	28,673 (97.4)	
Rural	26,214 (16.8)	1554 (30.4)	740 (2.5)	
Missing	54 (0.0)	2 (0.0)	15 (0.1)	
Recent immigrant (i.e., after 1985), no. (%)				<0.001
Yes	3618 (2.3)	47 (0.9)	7348 (25.0)	
No	152,568 (97.7)	5063 (99.1)	22,080 (75.0)	
English proficiency, no. (%)				<0.001
Low	—	396 (7.7)	15,570 (52.9)	
Moderate-to-high	—	4714 (92.3)	13,858 (47.1)	

*Baseline characteristics were compared using analysis of variance for continuous variables and χ^2 tests for categorical variables.


?
R

All-purpose table-making packages

Stars on GitHub	678	541	587
Questions on StackOverflow	739	622	2016
CRAN Downloads	83K/month	50K/month	260K/month

* data as of 02-07-2024

table1

```r
df <- survival::pbc |>
  dplyr::select(
    age, bili, albumin,
    stage, hepato, sex
  )

table1::table1(
  x = ~ age + bili + albumin +
    factor(stage) + factor(hepato) | sex,
  data = df, overall = F,
  render.continuous = c(. = "Mean (SD)")
)
```

	m (N=44)	f (N=374)
age		
Mean (SD)	55.7 (11.0)	50.2 (10.2)
bili		
Mean (SD)	2.87 (2.32)	3.26 (4.59)
albumin		
Mean (SD)	3.54 (0.457)	3.49 (0.422)
factor(stage)		
1	3 (6.8%)	18 (4.8%)
2	8 (18.2%)	84 (22.5%)
3	16 (36.4%)	139 (37.2%)
4	17 (38.6%)	127 (34.0%)
Missing	0 (0%)	6 (1.6%)
factor(hepato)		
0	15 (34.1%)	137 (36.6%)
1	21 (47.7%)	139 (37.2%)
Missing	8 (18.2%)	98 (26.2%)

table1

```
df <- survival::pbc |>
  dplyr::select(
    age, bili, albumin,
    stage, hepato, sex
  )

table1::table1(
  x = ~ age + bili + albumin +
    factor(stage) + factor(hepato) | sex,
  data = df, overall = F,
  render.continuous = c(. = "Mean (SD)")
)
```

	m (N=44)	f (N=374)
age		
Mean (SD)	55.7 (11.0)	50.2 (10.2)
bili		
Mean (SD)	2.87 (2.32)	3.26 (4.59)
albumin		
Mean (SD)	3.54 (0.457)	3.49 (0.422)
factor(stage)		
1	3 (6.8%)	18 (4.8%)
2	8 (18.2%)	84 (22.5%)
3	16 (36.4%)	139 (37.2%)
4	17 (38.6%)	127 (34.0%)
Missing	0 (0%)	6 (1.6%)
factor(hepato)		
0	15 (34.1%)	137 (36.6%)
1	21 (47.7%)	139 (37.2%)
Missing	8 (18.2%)	98 (26.2%)

gtsummary

```
library(gtsummary)
survival::pbc |>
  dplyr::select(
    age, bili, albumin,
    stage, hepato, sex
  ) |>
  tbl_summary(
    by = sex,
    statistic = all_continuous()
    ~ "{mean} ({sd})"
  ) |>
  add_p()
```

Characteristic	**m**, N = 44[1]	**f**, N = 374[1]	**p-value**[2]
age	56 (11)	50 (10)	0.003
bili	2.9 (2.3)	3.3 (4.6)	0.029
albumin	3.54 (0.46)	3.49 (0.42)	0.3
stage			0.8
1	3 (6.8%)	18 (4.9%)	
2	8 (18%)	84 (23%)	
3	16 (36%)	139 (38%)	
4	17 (39%)	127 (35%)	
Unknown	0	6	
hepato	21 (58%)	139 (50%)	0.4
Unknown	8	98	

[1] Mean (SD); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

gtsummary

```
library(gtsummary)
survival::pbc |>
  dplyr::select(
    age, bili, albumin,
    stage, hepato, sex
  ) |>
  tbl_summary(
    by = sex,
    statistic = all_continuous()
    ~ "{mean} ({sd})"
  ) |>
  add_p()
```

Characteristic	m, N = 44[1]	f, N = 374[1]	p-value[2]
age	56 (11)	50 (10)	0.003
bili	2.9 (2.3)	3.3 (4.6)	0.029
albumin	3.54 (0.46)	3.49 (0.42)	0.3
stage			0.8
1	3 (6.8%)	18 (4.9%)	
2	8 (18%)	84 (23%)	
3	16 (36%)	139 (38%)	
4	17 (39%)	127 (35%)	
Unknown	0	6	
hepato	21 (58%)	139 (50%)	0.4
Unknown	8	98	

[1] Mean (SD); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

Elevating gtsummary

```
data_dict <- readr::read_csv("data-dictionary.csv")
head(data_dict)
```

```
#> # A tibble: 6 × 2
#>   variable variable_label
#>   <chr>    <chr>
#> 1 age      Age (years)
#> 2 bili     Serum bilirubin (mg/dl)
#> 3 albumin  Serum albumin (g/dl)
#> 4 stage    Histologic stage of disease
#> 5 hepato   Presence of hepatomegaly or enlarged liver
#> 6 sex      Sex
```

Elevating gtsummary

```r
# create a named vector
labels_vec <- tibble::deframe(data_dict)

# assign each variable with its respective label attribute
df <- labelled::set_variable_labels(df, !!!labels_vec)

# render summary table
tbl_summary(
  df,
  by = sex,
  statistic = all_continuous() ~ "{mean} ({sd})"
)
```

* `df` defined in slide 5

Characteristic	m, N = 44[1]	f, N = 374[1]
age	56 (11)	50 (10)
bili	2.9 (2.3)	3.3 (4.6)
albumin	3.54 (0.46)	3.49 (0.42)
stage		
1	3 (6.8%)	18 (4.9%)
2	8 (18%)	84 (23%)
3	16 (36%)	139 (38%)
4	17 (39%)	127 (35%)
Unknown	0	6
hepato	21 (58%)	139 (50%)
Unknown	8	98

[1] Mean (SD); n (%)

Characteristic	m, N = 44[1]	f, N = 374[1]
Age (years)	56 (11)	50 (10)
Serum bilirunbin (mg/dl)	2.9 (2.3)	3.3 (4.6)
Serum albumin (g/dl)	3.54 (0.46)	3.49 (0.42)
Histologic stage of disease		
1	3 (6.8%)	18 (4.9%)
2	8 (18%)	84 (23%)
3	16 (36%)	139 (38%)
4	17 (39%)	127 (35%)
Unknown	0	6
Presence of hepatomegaly or enlarged liver	21 (58%)	139 (50%)
Unknown	8	98

[1] Mean (SD); n (%)

By combining the strengths of different packages, you can easily create visually polished tables in a way that is conducive to reproducible analysis.

References

- In-Hospital Patient Harm Across Linguistic Groups: A Retrospective Cohort Study of Home Care Recipients
- kableExtra: Construct Complex Table with 'kable' and Pipe Syntax
- flextable: Functions for Tabular Reporting
- DT: A Wrapper of the JavaScript Library 'DataTables'
- table1: Tables of Descriptive Statistics in HTML
- gtsummary: Presentation-Ready Data Summary and Analytic Result Tables
- PIPING HOT DATA: The case for variable labels in R
- Lefkios Paikousis: A workflow with labelled data

Images from:

- IKEA
- The R Foundation