How to Create Publication-Ready Tables in R

UseR! 2022

Raymond Balise with Anna Calderon, Francisco Cardozo, Lauren Nahodyl

Jun 20, 2022

Who are we?

- Francisco - A Ph.D. student at The University of Miami's Miller School of Medicine, quantifying policies to prevent/delay childhood alcohol use.

The rest of us work at UM:

- Ray - An Associate Professor of biostatistics (really a data scientist) studying the intersection of HIV, mental health and drug addiction.
- Lauren - A statistician/data scientist at UM's Sylvester Comprehensive Cancer Center
- Anna - A data scientist working with Ray on HIV, mental health and addiction

Disclosures

We have no financial disclosures to report.

License

General Comments

- I use and teach with `tidyverse` and `tidymodels`. Because not everybody has been able to upgrade to R 4.2, we left `%>%` pipes in this presentation but we strongly suggest that people upgrade and use the most excellent `|>` pipe.

- The slides are done with the `xaringan` package. There is bug somewhere that causes the slides to render quickly but then there is a delay (about a minute) before the details are shown on two of my Macs. Sorry! If you have the same issue add a comment:

  - https://community.rstudio.com/t/rstudio-miserably-slow-on-os-monterey-when-knitting-xaringan-slides/140183

Where did the idea for this talk come from? [1]

https://vimeo.com/585821837

Where did the idea for this talk come from? [2]

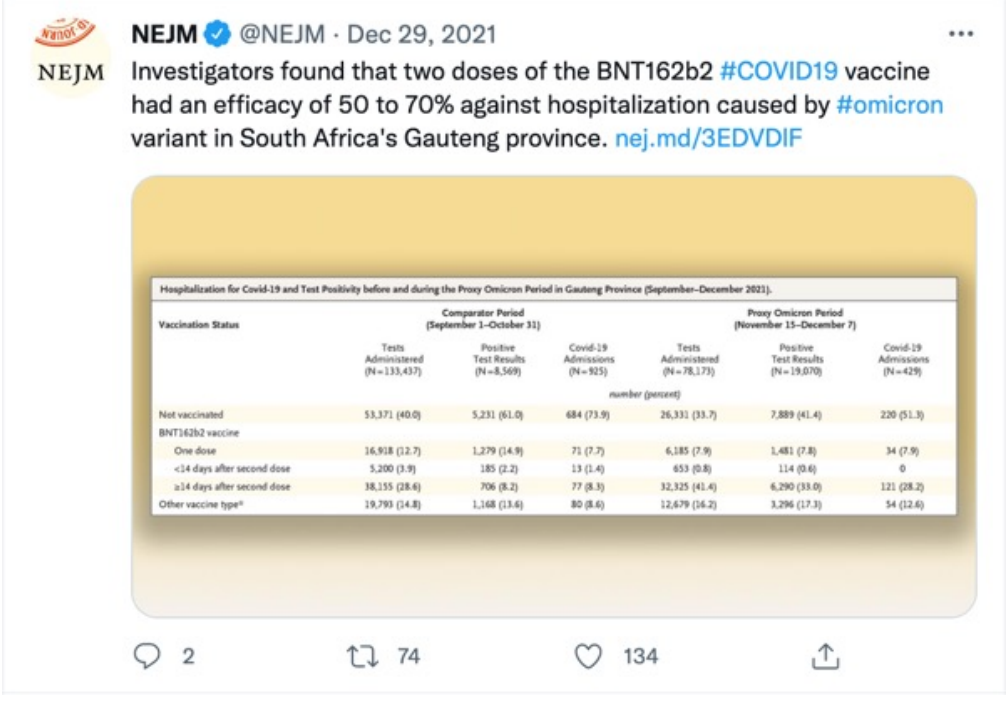
- When I used to primarily work in SAS, I would get my tables close to correct with *proc whatever* and then export to HTML, tweak things in Excel, and wait for the call saying:

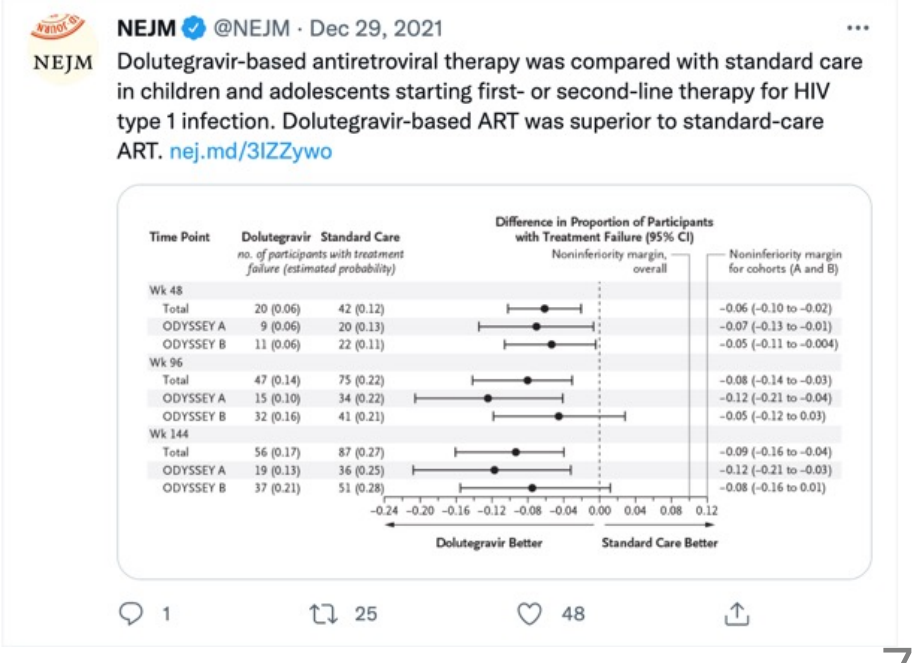
> We are really sorry but there is "this guy"...

- Of course, every number in the paper, including the tables, would change and I would mutter about needing to learn R.
- I learned R Markdown and finally stopped crying whenever my phone would ring.

When I was in school...

- Journal tables (and figures) were monochrome and simple.
- Now we have tables that are artfully formatted and have art that is ready for Twitter.
- The distinction between tables and figures has completely blurred.





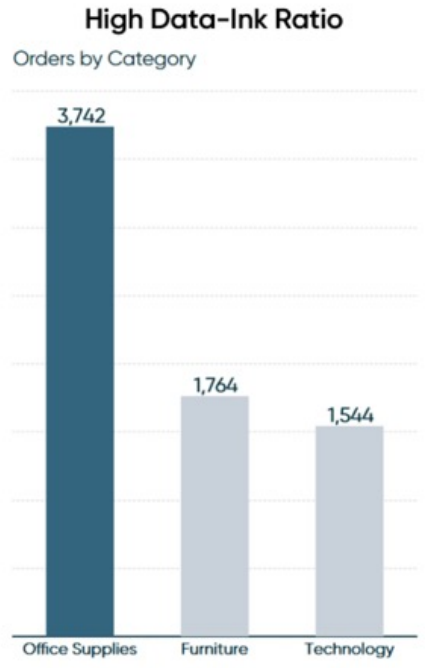
When should you use a table?

- Tables are useful to show the exact values of your data or estimates.
- They are not the best solution to show a lot of data or if you want to show the data in a compact space.
- They are not usually intended to give a quick, visual representation of data.
  - However, the core psychological principles leading to fast and accurate judgment in graphics apply to both charts and tables.
  - Read/Study/Live *Visualizing Data* and *The Elements of Graphing Data* by William S. Cleveland.

Best Practices in Plots (And Graphics)

Tufte on Visualizations

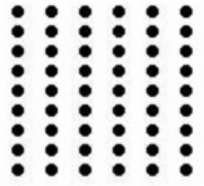
- One of the visualization thought leaders in the 1980s, Edward Tufte, stressed the importance of erasing all unnecessary ink from the page. The same holds true for tables.
- When in doubt, erase.



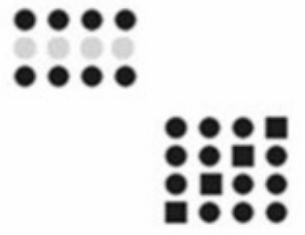
https://playfairdata.com/data-ink-ratio-animation-and-how-to-apply-it-in-tableau/

Gestalt Principles

Gestalt Law of Proximity - Things close together are automatically grouped.



Gestalt Law of Similarity - Things of the same color will be grouped. If one thing is a different color, it will pop off of the page.



Gestalt Law of Closure - You can draw part of a bounding box and the brain will fill in the rest.

https://blog.xlcubed.com/2008/05/gestalt-laws-charts-and-tables-the-way-your-brain-wants-them-to-be/

Schwabish's Rules for Better Tables

Journal of Benefit-Cost Analysis

Article contents

Abstract

References

Ten Guidelines for Better Tables

Published online by Cambridge University Press: **30 July 2020**

Jonathan A. Schwabish

Show author details

Article Metrics

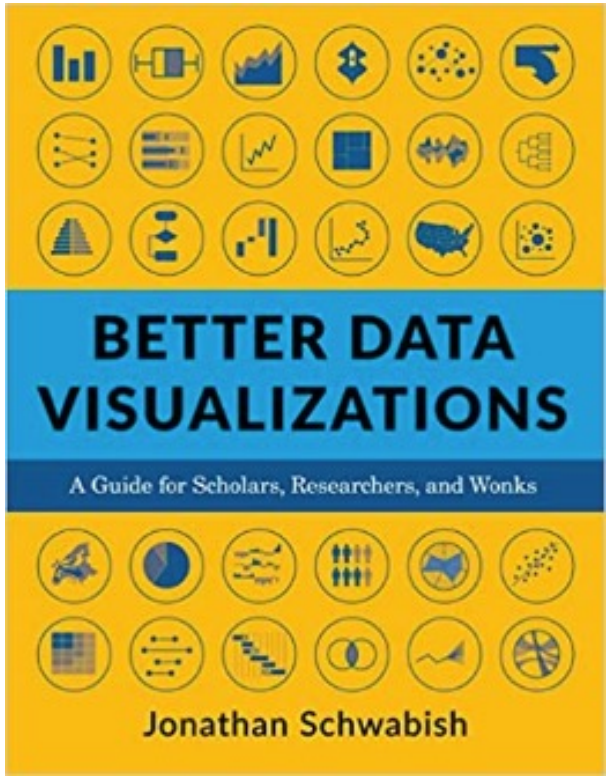
Get access | Share | Cite | Rights & Permissions

Abstract

Tables are a unique form of visualizing data because, unlike many charts, they are not usually intended to give a quick, visual representation of data. Instead, tables are useful when you want to show the exact values of your data or estimates. They are not the best solution if you want to show a lot of data or if you want to show the data in a compact space, but a well-designed table can help your reader find specific numbers and discover patterns and outliers. In this article, I present 10 guidelines for creating better, more effective tables; I then model these lessons by redesigning six tables from articles previously published in the *Journal of Benefit-Cost Analysis.*

Type	Invited Paper
Information	Journal of Benefit-Cost Analysis, Volume 11, Issue 2, Summer 2020, pp. 151 - 178 DOI: https://doi.org/10.1017/bca.2020.11
Copyright	© Society for Benefit-Cost Analysis, 2020

https://doi.org/10.1017/bca.2020.11



https://cup.columbia.edu/book/better-data-visualizations/9780231193115

Schwabish's Ten Rules for Better Tables + Balise Three

1. Offset the Heads from the Body
2. Use Subtle Dividers Rather Than Heavy Gridlines
3. Right-Align Numbers and Heads
4. Left-Align Text and Heads
5. Select the Appropriate Level of Precision
6. Guide Your Reader with Space between Rows and Columns
7. Remove Unit Repetition
8. Highlight Outliers
9. Group Similar Data and Increase White Space
10. Add Visualizations When Appropriate
11. Draw Attention to the Key Point(s)
12. Use Annotations to Explain the Statistics
13. Make Captions/Titles Self-Contained
    - Explain Sample Size, Who and When

What is the Key Take-Away Point in These Data?

	A	B	C
1	Demographic Characteristics by PrEP Use		
2		No PrEP in the last year	On PrEP in the last year
3		(N=2700)	(N=1227)
4	Sex		
5	Male	2677 (99.1%)	1221 (99.5%)
6	Female	23 (0.9%)	6 (0.5%)
7	Sexual Orientation		
8	Gay	2275 (84.3%)	1092 (89.0%)
9	Bisexual	391 (14.5%)	121 (9.9%)
10	Straight	9 (0.3%)	3 (0.2%)
11	Other	25 (0.9%)	11 (0.9%)
12	Race		
13	White	1498 (55.5%)	636 (51.8%)
14	Black/African American	260 (9.6%)	111 (9.0%)
15	Latino	625 (23.1%)	309 (25.2%)
16	Other	317 (11.7%)	171 (13.9%)
17	Depression		
18	No Depressive Symptoms	962 (35.6%)	433 (35.3%)
19	Mild Depressive Symptoms	1176 (43.6%)	511 (41.6%)
20	Moderate Depressive Symptoms	409 (15.1%)	205 (16.7%)
21	Severe Depressive Symptoms	153 (5.7%)	78 (6.4%)
22	Racial Discrimination		
23	None	1727 (64.0%)	819 (66.7%)
24	1-2 Scenarios	633 (23.4%)	266 (21.7%)
25	3-4 Scenarios	236 (8.7%)	107 (8.7%)
26	5-6 Scenarios	104 (3.9%)	35 (2.9%)

	No PrEP	Difference (%)	On PrEP
Sex			
Male	2677 (99.1%)	0.4	1221 (99.5%)
Female	23 (0.9%)	0.4	6 (0.5%)
Sexual Orientation			
Gay	2275 (84.3%)	4.7	1092 (89.0%)
Bisexual	391 (14.5%)	4.6	121 (9.9%)
Straight	9 (0.3%)	0.1	3 (0.2%)
Other	25 (0.9%)	0	11 (0.9%)
Race			
White	1498 (55.5%)	3.7	636 (51.8%)
Black/African American	260 (9.6%)	0.6	111 (9.0%)
Latino	625 (23.1%)	2.1	309 (25.2%)
Other	317 (11.7%)	2.2	171 (13.9%)
Depressive Symptoms			
None	962 (35.6%)	0.3	433 (35.3%)
Mild	1176 (43.6%)	2	511 (41.6%)
Moderate	409 (15.1%)	1.6	205 (16.7%)
Severe	153 (5.7%)	0.7	78 (6.4%)
Racial Discrimination			
None	1727 (64.0%)	2.7	819 (66.7%)
1-2 Scenarios	633 (23.4%)	1.7	266 (21.7%)
3-4 Scenarios	236 (8.7%)	0	107 (8.7%)
5-6 Scenarios	104 (3.9%)	1	35 (2.9%)

Table 1. There are Negliable Difference in the Demographic Characteristics for 3927 Americans Taking (N=2700) or Not Taking (N=1227) PREP in the Last Year, Circa 2020.

Critique PReP Table 1 - Excel Default

	A	B	C
1	Demographic Characteristics by PrEP Use		
2		No PrEP in the last year	On PrEP in the last year
3		(N=2700)	(N=1227)
4	Sex		
5	Male	2677 (99.1%)	1221 (99.5%)
6	Female	23 (0.9%)	6 (0.5%)
7	Sexual Orientation		
8	Gay	2275 (84.3%)	1092 (89.0%)
9	Bisexual	391 (14.5%)	121 (9.9%)
10	Straight	9 (0.3%)	3 (0.2%)
11	Other	25 (0.9%)	11 (0.9%)
12	Race		
13	White	1498 (55.5%)	636 (51.8%)
14	Black/African American	260 (9.6%)	111 (9.0%)
15	Latino	625 (23.1%)	309 (25.2%)
16	Other	317 (11.7%)	171 (13.9%)
17	Depression		
18	No Depressive Symptoms	962 (35.6%)	433 (35.3%)
19	Mild Depressive Symptoms	1176 (43.6%)	511 (41.6%)
20	Moderate Depressive Symptoms	409 (15.1%)	205 (16.7%)
21	Severe Depressive Symptoms	153 (5.7%)	78 (6.4%)
22	Racial Discrimination		
23	None	1727 (64.0%)	819 (66.7%)
24	1-2 Scenarios	633 (23.4%)	266 (21.7%)
25	3-4 Scenarios	236 (8.7%)	107 (8.7%)
26	5-6 Scenarios	104 (3.9%)	35 (2.9%)

1. Heads ✅
2. Subtle Dividers ❌
3. Right-Align Numbers ❌
4. Left-Align Text ❌
5. Precision ✅
6. Space ✅
7. Remove Repetition ❌
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ❌
12. Annotations
13. Caption/Title ❌

Critique PReP Table 1 - Also Excel

	No PrEP	Difference (%)	On PrEP
Sex			
Male	2677 (99.1%)	0.4	1221 (99.5%)
Female	23 (0.9%)	0.4	6 (0.5%)
Sexual Orientation			
Gay	2275 (84.3%)	4.7	1092 (89.0%)
Bisexual	391 (14.5%)	4.6	121 (9.9%)
Straight	9 (0.3%)	0.1	3 (0.2%)
Other	25 (0.9%)	0	11 (0.9%)
Race			
White	1498 (55.5%)	3.7	636 (51.8%)
Black/African American	260 (9.6%)	0.6	111 (9.0%)
Latino	625 (23.1%)	2.1	309 (25.2%)
Other	317 (11.7%)	2.2	171 (13.9%)
Depressive Symptoms			
None	962 (35.6%)	0.3	433 (35.3%)
Mild	1176 (43.6%)	2	511 (41.6%)
Moderate	409 (15.1%)	1.6	205 (16.7%)
Severe	153 (5.7%)	0.7	78 (6.4%)
Racial Discrimination			
None	1727 (64.0%)	2.7	819 (66.7%)
1-2 Scenarios	633 (23.4%)	1.7	266 (21.7%)
3-4 Scenarios	236 (8.7%)	0	107 (8.7%)
5-6 Scenarios	104 (3.9%)	1	35 (2.9%)

Table 1. There are Negliable Difference in the Demographic Characteristics for 3927 Americans Taking (N=2700) or Not Taking (N=1227) PREP in the Last Year, Circa 2020.

1. Heads ✅
2. Subtle Dividers ✅
3. Right-Align Numbers ❌
4. Left-Align Text ✅
5. Precision ✅
6. Space ✅
7. Remove Repetition ❌
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ✅
12. Annotations
13. Caption/Title ✅

Critique PReP Table 1 - (Version 1.1)

Table 1. There are Negliable Difference in the Demographic Characteristics for 3927 Americans Taking (N=2700) or Not Taking (N=1227) PREP in the Last Year, Circa 2020.

	No PrEP	Difference (%)	On PrEP
Sex			
Male	2677 (99.1%)	0.4	1221 (99.5%)
Female	23 (0.9)	0.4	6 (0.5)
Sexual Orientation			
Gay	2275 (84.3%)	4.7	1092 (89.0%)
Bisexual	391 (14.5)	4.6	121 (9.9)
Straight	9 (0.3)	0.1	3 (0.2)
Other	25 (0.9)	0	11 (0.9)
Race			
White	1498 (55.5%)	3.7	636 (51.8%)
Black/African American	260 (9.6)	0.6	111 (9.0)
Latino	625 (23.1)	2.1	309 (25.2)
Other	317 (11.7)	2.2	171 (13.9)
Depressive Symptoms			
None	962 (35.6%)	0.3	433 (35.3%)
Mild	1176 (43.6)	2	511 (41.6)
Moderate	409 (15.1%)	1.6	205 (16.7)
Severe	153 (5.7%)	0.7	78 (6.4)
Racial Discrimination			
None	1727 (64.0%)	2.7	819 (66.7%)
1-2 Scenarios	633 (23.4)	1.7	266 (21.7)
3-4 Scenarios	236 (8.7)	0	107 (8.7)
5-6 Scenarios	104 (3.9)	1	35 (2.9)

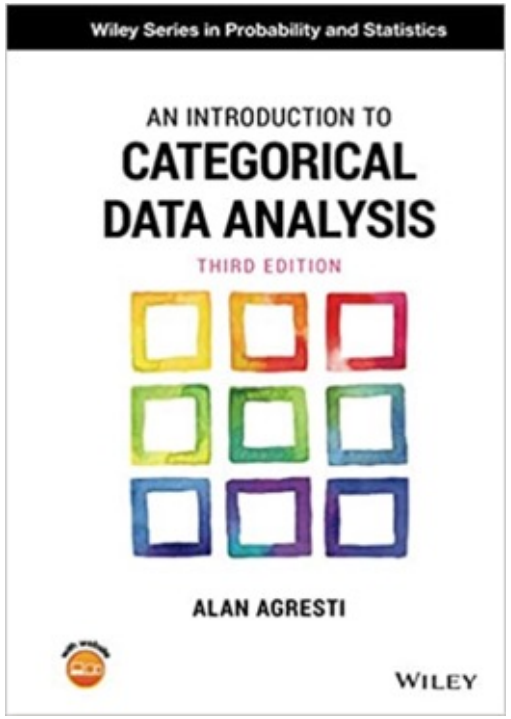
1. Heads ✅
2. Subtle Dividers ✅
3. Right-Align Numbers ❌
4. Left-Align Text ✅
5. Precision ✅
6. Space ✅
7. Remove Repetition ✅
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ✅
12. Annotations
13. Caption/Title ✅

Where We are Going

A Religious Experience
Playing with Babies
Fun with Cameras in the Surgical Theater
What is going on in Florida?

Categorical Data Analysis

- One of the best-written statistics books: *An Introduction to Categorical Data Analysis* by Alan Agresti



https://raymondbalise.github.io/Agresti_IntroToCategorical/
https://github.com/RaymondBalise/Agresti_IntroToCategorical

Good Tables vs. Really???

The physical book has lovely tables printed on horribly thin paper:

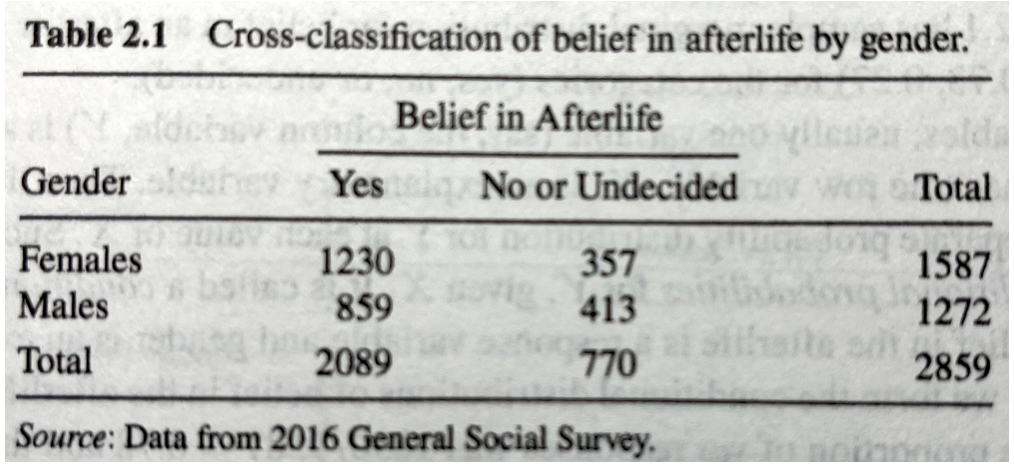


Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in Afterlife: Yes	Belief in Afterlife: No or Undecided	Total
Females	1230	357	1587
Males	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

The online version through the UM library is not beautiful:

Table 2.1 Cross-classification of belief in afterlife by gender.

	Belief in Afterlife		
Gender	**Yes**	**No or Undecided**	**Total**
Females	1230	357	1587
Males	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

Critique - Agresti Book

The physical book has lovely tables printed on horribly thin paper:

Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in Afterlife: Yes	Belief in Afterlife: No or Undecided	Total
Females	1230	357	1587
Males	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

1. Heads ✅
2. Subtle Dividers ✅
3. Right-Align Numbers ✅
4. Left-Align Text ✅
5. Precision ✅
6. Space ✅
7. Remove Repetition ✅
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ❌
12. Annotations
13. Caption/Title ✅

Critique - Agresti Online

The physical book has lovely tables printed on horribly thin paper:

Table 2.1 Cross-classification of belief in afterlife by gender.

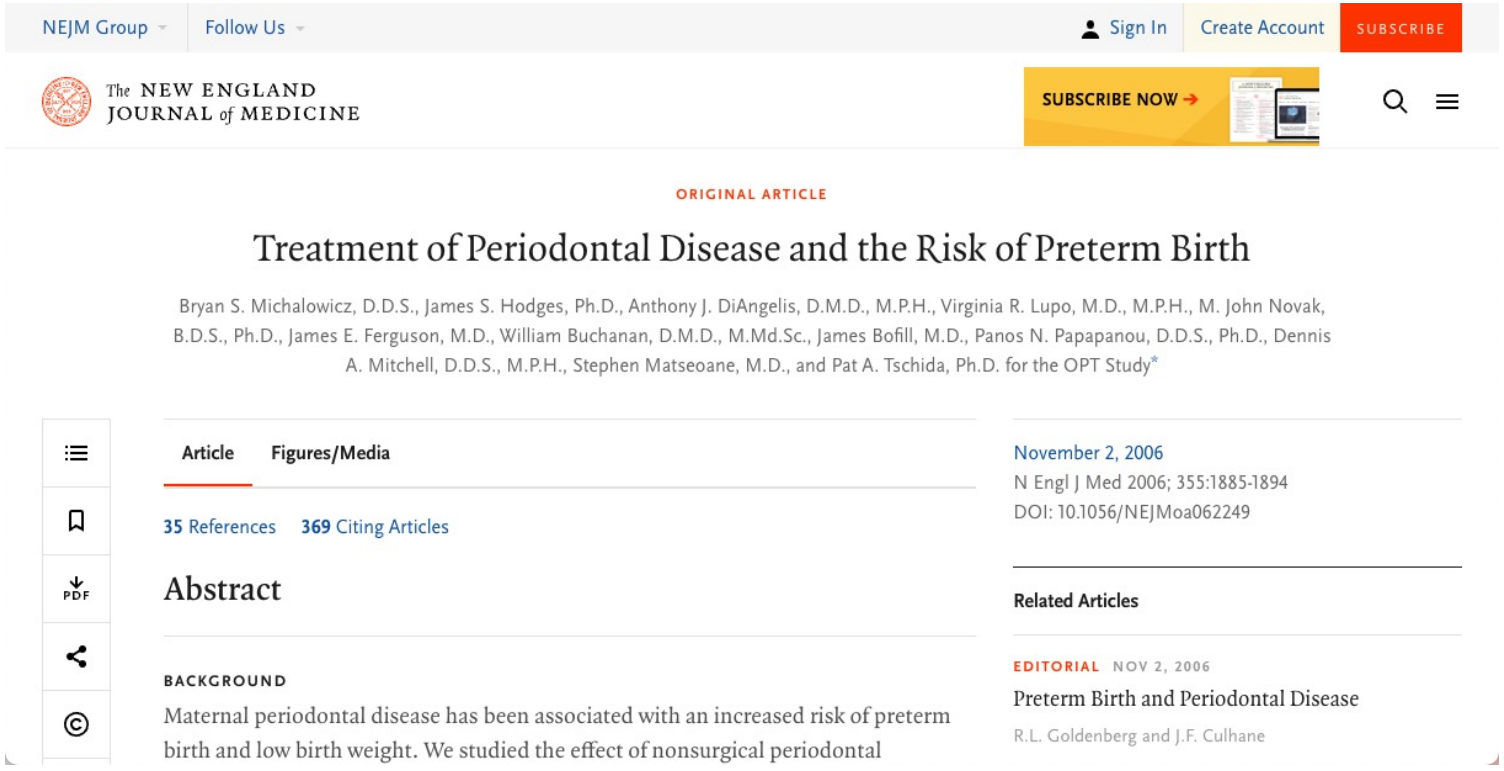
	Belief in Afterlife		
Gender	**Yes**	**No or Undecided**	**Total**
Females	1230	357	1587
Males	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

1. Heads ✅
2. Subtle Dividers ❌
3. Right-Align Numbers ✅
4. Left-Align Text ✅
5. Precision ✅
6. Space ❌
7. Remove Repetition ✅
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ❌
12. Annotations
13. Caption/Title ✅

The `medicaldata` Package

There is a dataset called `opt` that has information from a controlled trial looking to see if treatment for periodontal disease can reduce the risk of preterm and low birth weight babies.



NEJM Style Table 1

Table 1. Baseline Characteristics of the Study Patients.*

Characteristic	Control Group (N=410)	Treatment Group (N=413)	P Value
Age — yr	25.9±5.5	26.1±5.6	0.56
Race or ethnic group — no. (%)†			
White	119 (29.0)	116 (28.1)	0.77
Black	182 (44.4)	190 (46.0)	0.64
Hispanic	180 (43.9)	170 (41.2)	0.43
Education — no. (%)			0.88
≤8 yr	76 (18.5)	78 (18.9)	
9–12 yr	242 (59.0)	237 (57.4)	
>12 yr	92 (22.4)	98 (23.7)	
Mean gestational age of fetus — wk	15.0±1.3	15.0±1.3	0.85
Previous pregnancies — no. (%)‡			
Any pregnancy	305 (74.4)	306 (74.1)	0.92
Live preterm birth§	44 (16.5)	33 (12.5)	0.18
Spontaneous abortion¶	94 (30.8)	108 (35.3)	0.24
Induced abortion¶	67 (22.0)	52 (17.0)	0.12
Stillbirth¶	6 (2.0)	9 (2.9)	0.44
Coexisting medical condition — no. (%)			
Diabetes	8 (2.0)	16 (3.9)	0.10
Chronic hypertension	9 (2.2)	16 (3.9)	0.16
Self-reported drug addiction	7 (1.7)	15 (3.6)	0.09
Self-reported alcohol use	8 (2.0)	8 (1.9)	0.99
Eating disorder	0	2 (0.5)	0.16
Dental status			
Number of natural teeth	26.8±1.7	26.7±1.8	0.67
Number of qualifying teeth	14.4±6.7	15.2±6.8	0.08
Percent of tooth sites that bled on probing	69.0±17.1	69.6±17.4	0.62
Percent of tooth sites with probing depth ≥4 mm	24.8±15.9	26.5±16.6	0.13

* Plus–minus values are means ±SD.
† Race or ethnic group was reported by the patients. Some women selected more than one category and were included in all.
‡ Some patients reported more than one event.
§ Percentages are based on 266 women in the control group and 265 women in the treatment group who had had any live births.
¶ Percentages are based on 305 women in the control group and 306 women in the treatment group who had had any previous pregnancies.

1. Heads ✅
2. Subtle Dividers ✅
3. Right-Align Numbers ✅
4. Left-Align Text ✅
5. Precision ✅
6. Space ✅
7. Remove Repetition ✅
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ❌
12. Annotations ✅

Device Clinical Trial - Laryngoscope Camera

Table 1. Demographics and Airway Assessment Data

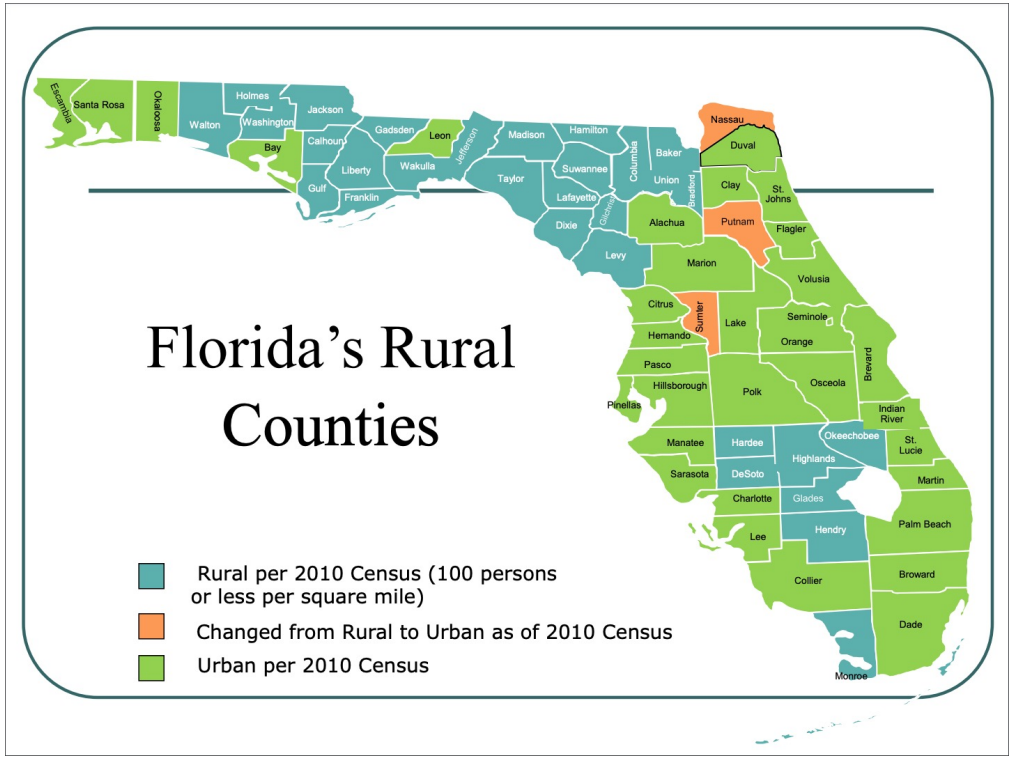
Variable	Pentax AWS (n = 50)	Macintosh (n = 49)	Standardized difference[a]
Age, y	50 ± 12	49 ± 14	0.14
Gender (male/female), *n*	11/39	10/39	0.04
Body mass index, kg/m^2	41.2 ± 4.4	42.5 ± 5.9	−0.21
ASA physical status, *n* (%)			0.41
II	15 (30)	7 (14)	
III	32 (64)	40 (82)	
IV	3 (6)	2 (4)	
Mallampati score,[b] *n* (%)			0.57
1	21 (42)	14 (29)	
2	18 (36)	21 (44)	
3	7 (14)	13 (27)	
4	4 (8)	0 (0)	

Summary statistics presented as number (%) of patients, or mean ± SD.
[a] The difference (Pentax − Macintosh) in means or proportions divided by the pooled standard deviation.
[b] One patient with missing Mallampati value in the Macintosh group.

- How hard is it to replicate these and/or make an interactive summary?

1. Heads ✅
2. Subtle Dividers ✅
3. Right-Align Numbers ❌
4. Left-Align Text ✅
5. Precision ✅
6. Space 🤷
7. Remove Repetition ✅
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s) ✅
12. Annotations ✅
13. Caption/Title ✅

What is going on in Florida?



https://www.floridahealth.gov/provider-and-partner-resources/community-health-workers/health-professional-shortage-designations/Rural%20Counties%20Map%202016.pdf https://www.flhealthcharts.gov/FLQUERY_New/Population/Count#

Population by County by Year

	Population					
	2016	2017	2018	2019	2020	Total
Alachua	257,478	259,349	263,753	266,649	270,405	**1,317,634**
Baker	26,967	27,066	27,488	28,089	28,588	**138,198**
Bay	176,637	178,953	182,218	179,900	175,776	**893,484**
Bradford	27,498	27,808	28,083	28,455	28,818	**140,662**
Brevard	570,496	576,970	584,050	593,372	604,154	**2,929,042**
Broward	1,860,979	1,884,545	1,903,210	1,927,014	1,946,104	**9,521,852**
Calhoun	14,594	14,658	15,315	14,982	14,894	**74,443**
Charlotte	171,219	173,954	175,413	182,298	185,392	**888,276**
Citrus	143,458	144,922	145,164	147,735	149,781	**731,060**
Clay	206,387	210,767	213,565	217,109	219,925	**1,067,753**
Collier	351,768	358,506	367,471	377,700	386,478	**1,841,923**
Columbia	68,687	69,250	69,566	70,614	70,694	**348,811**
Miami-Dade	2,712,144	2,754,749	2,804,160	2,830,500	2,864,600	**13,966,153**
Desoto	35,215	35,454	35,940	35,718	36,388	**178,715**
Dixie	16,844	17,040	16,767	16,516	16,704	**83,871**
Duval	927,903	942,841	954,454	971,842	988,783	**4,785,823**
Escambia	310,642	312,811	317,051	322,901	324,620	**1,588,025**
Flagler	103,584	106,076	108,481	110,636	114,053	**542,830**
Franklin	11,937	12,006	12,360	12,017	12,229	**60,549**
Gadsden	48,527	48,690	48,173	47,926	46,345	**239,661**
Gilchrist	16,862	16,977	17,578	17,682	18,027	**87,126**
Glades	13,101	13,263	13,193	13,098	13,230	**65,885**
Gulf	16,718	16,957	16,235	16,507	14,716	**81,133**
Hamilton	14,666	14,749	14,706	14,787	14,618	**73,526**
Hardee	27,643	27,675	27,436	27,311	27,571	**137,636**
Hendry	38,436	38,675	39,682	40,089	40,594	**197,476**
Hernando	180,213	183,065	185,421	189,661	192,189	**930,549**
Highlands	101,727	102,590	103,317	103,391	104,384	**515,409**
Hillsborough	1,359,850	1,388,111	1,419,285	1,445,243	1,481,163	**7,093,652**
Holmes	20,037	20,132	20,404	20,218	20,184	**100,975**
Indian River	147,163	149,930	152,079	155,308	158,238	**762,718**
Jackson	50,311	50,303	50,689	50,325	47,171	**248,799**
Jefferson	14,501	14,530	14,725	14,842	14,831	**73,429**
Lafayette	8,620	8,651	8,367	8,613	8,721	**42,972**
Lake	325,887	333,598	342,356	354,537	368,828	**1,725,206**
Lee	684,465	700,837	721,053	734,630	756,912	**3,597,897**
Leon	288,495	291,879	290,223	296,717	300,519	**1,467,833**
Levy	40,599	40,832	41,550	41,354	41,634	**205,969**
Liberty	8,754	8,839	8,781	9,167	8,774	**44,315**
Madison	19,252	19,295	19,420	19,533	19,254	**96,754**
Manatee	359,486	367,130	381,071	388,729	397,727	**1,894,143**
Marion	346,956	352,067	355,325	360,053	367,247	**1,781,648**
Martin	151,081	152,333	155,705	158,006	161,017	**778,142**
Monroe	76,461	77,300	76,534	73,253	76,280	**379,828**
Nassau	78,174	79,592	83,125	85,135	87,389	**413,415**
Okaloosa	193,247	194,811	198,409	201,104	204,326	**991,897**
Okeechobee	40,983	41,469	41,492	41,347	42,187	**207,478**
Orange	1,287,703	1,317,704	1,370,447	1,389,297	1,426,631	**6,791,782**
Osceola	326,342	339,470	360,426	368,678	388,132	**1,783,048**
Palm Beach	1,395,117	1,411,054	1,442,281	1,458,576	1,469,904	**7,176,932**
Pasco	497,991	507,081	518,639	527,174	539,769	**2,590,654**
Pinellas	956,302	961,253	971,022	979,558	986,400	**4,854,535**
Polk	650,552	663,999	681,691	688,770	707,191	**3,392,203**
Putnam	73,004	73,068	73,422	73,012	73,355	**365,861**
Saint Johns	222,006	229,272	241,545	249,734	266,128	**1,208,685**
Saint Lucie	294,144	299,962	304,743	309,073	316,620	**1,524,542**
Santa Rosa	168,026	171,851	175,552	179,875	183,633	**878,937**
Sarasota	401,316	407,501	415,896	426,977	434,853	**2,086,543**
Seminole	450,706	457,028	463,627	472,775	480,417	**2,324,553**
Sumter	119,433	123,928	125,779	130,642	133,310	**633,092**
Suwannee	44,340	44,527	45,123	45,482	46,028	**225,500**
Taylor	22,400	22,220	22,258	22,652	22,654	**112,184**
Union	15,873	15,896	15,966	15,985	15,493	**79,213**
Volusia	519,037	525,121	532,926	539,563	546,612	**2,663,259**
Wakulla	31,706	32,134	32,350	32,418	33,394	**162,002**
Walton	63,562	65,724	67,926	70,352	72,528	**340,092**
Washington	24,880	24,935	25,243	25,347	25,252	**125,657**
Total	**20,231,092**	**20,555,733**	**20,957,705**	**21,268,553**	**21,640,766**	**104,653,849**

So Many Tools ... So Little Time

I wish there was a "best" tool.

- You have a lot of package options for making static (i.e., Word, PDF or HTML) and dynamic (i.e., HTML) publication-ready tables.
- All packages make web-friendly graphics.
- Others make static graphics that are beautifully formatted for the the web: `gt`, `table1`, `gtsummary`, `kableExtra`.
- Yet others make static tables that look great in Word: `flextable`.
- Some are ideal for interactive web content: `dt`, `reactable`

Output for Packages that Make Tables

Not all packages support all R Markdown output formats.

Print Engine	Function	HTML	PDF	RTF	Word
tibble	`tibble()`				
kable	`kable()`				
kableExtra	`kbl() + stuff`				
gt	`gt()`				
flextable	`flextable()`				
huxtable	`huxtable()`				
DT	`datatable()`				
reactable	`reactable()`				

Key	
	Output fully supported
	Formatted output, but missing indentation, footnotes, spanning headers
	No formatted output
	Output not supported
	Under development
	Image of first page with controls

Thanks to Daniel D. Sjoberg for the table code.

Icons from icons8

Make a Table

When You Already Have Preprocessed the Data

A religious experience...

Categorical Data Analysis

- *An Introduction to Categorical Data Analysis* by Alan Agresti has solid SAS code.
- A couple years ago, I decided to try to replicate everything in R.

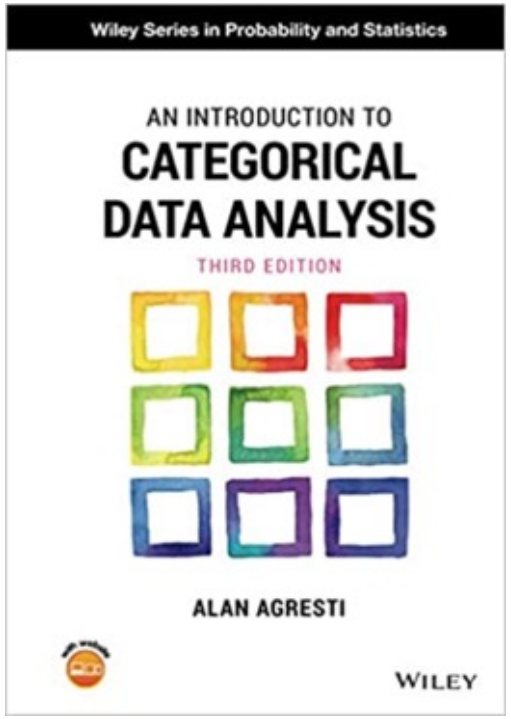


Table 2.1 Cross-classification of belief in afterlife by gender.

	Belief in the Afterlife		
Gender	Yes	No or Undecided	Total
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

Agresti in R

- This is quick code for exploratory data analysis but we can do a lot better.

```
table2_1 <- data.frame(
  Gender = c("Female", "Female", "Male", "Male"),
  Belief = c(" Yes", "No", " Yes", "No"),
  Count = c(1230, 357, 859, 413)
)

addmargins(xtabs(Count ~ Gender + Belief, table2_1))
```

```
             Belief
Gender        Yes   No  Sum
  Female     1230  357 1587
  Male        859  413 1272
  Sum        2089  770 2859
```

Notice that I used a space before "Y" in "Yes" to set the order.

`tibble` - Not Exactly Publication Ready....

```
`Table 2.1` <-
  xtabs(Count ~ Gender + Belief, table2_1) %>%
  addmargins() %>%
  as_tibble()

`Table 2.1`
```

```
  # A tibble: 9 × 3
  Gender Belief     n
  <chr>  <chr>  <dbl>
1 Female " Yes"  1230
2 Male   " Yes"   859
3 Sum    " Yes"  2089
4 Female "No"     357
5 Male   "No"     413
6 Sum    "No"     770
7 Female "Sum"   1587
8 Male   "Sum"   1272
9 Sum    "Sum"   2859
```

Pivot to Make a Table Wide - `tibble` ☹

The `Table 2.1` data set was saved with `as_tibble()`.

```
wide <-
  `Table 2.1` %>%
  pivot_wider(id_cols = Gender, names_from = Belief, values_from = n) %>%
  mutate(
    Gender = if_else(Gender == "Sum", "Total", Gender)
  ) %>%
  rename(
    Yes = ` Yes`,
    # I could also do: `No or Undecided` = No,
    Total = Sum
  )
wide
```

```
# A tibble: 3 × 4
  Gender   Yes    No Total
  <chr>  <dbl> <dbl> <dbl>
1 Female  1230   357  1587
2 Male     859   413  1272
3 Total   2089   770  2859
```

`knitr::kable()` 😐

- Drop that tibble into `kable()` and add on pretty column names with the col.names= argument.

```
wide %>%
  kable(
    col.names = c("Gender", "Yes", "No or Undecided", "Total")
  )
```

Gender	Yes	No or Undecided	Total
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Not bad...

kableExtra::kbl() + stuff

```
library(kableExtra)
wide %>%
  kbl(
    col.names = c("Gender", "Yes", "No or Undecided", "Total"),
    align = c("l", "r", "c", "r")
  ) %>%
  add_header_above(c(" ", "Belief in Afterlife" = 2, " ")) %>%
  column_spec(3, width = "10em") %>% # make the 3rd column extra wide
  footnote(
    general = "Data from 2016 General Social Survey.",
    general_title = "Source: ",
    footnote_as_chunk = T, title_format = c("italic")
  )
```

	Belief in Afterlife		
Gender	**Yes**	**No or Undecided**	**Total**
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

Closer, but not good...

Add Some CSS

- I am using the `xaringan` package to render slides, so I need a small tweak.
- Web page formatting is done with CSS. Here I am adding instructions for things tagged as being in a table. I need to set them to have a white background.

Adding CSS

````
```{css, echo=FALSE}
    .remark-slide thead, .remark-slide tr {
        background-color: white;
    }

    .remark-slide thead, .remark-slide tfoot {
        background-color: white;
    }
```
````

```
library(kableExtra)
wide %>%
  kbl(
    col.names = c("Gender", "Yes", "No or Undecided", "Total"),
    align = c("l", "r", "c", "r")
  ) %>%
  add_header_above(c(" ", "Belief in Afterlife" = 2, " ")) %>%
  column_spec(3, width = "10em") %>%
  footnote(
    general = "Data from 2016 General Social Survey.",
    general_title = "Source: ",
    footnote_as_chunk = T, title_format = c("italic")
  )
```

Close to Perfect - `kableExtra` 😁

	Belief in Afterlife		
Gender	**Yes**	**No or Undecided**	**Total**
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

How did I figure that out?

- The `kableExtra` documentation does not show the striping behavior...
- So I tried to simplify.
- I copied the code into a vanilla HTML R Markdown file.
  - It didn't include the stripes.
- I did a Google search for `xaringan kableExtra` formatting
  - It led me to this post which included the CSS: https://stackoverflow.com/questions/55319141/xaringan-kableextrakable-styling-and-wider-tables
- StackOverflow is your friend.
- To learn how to find CSS elements, watch this:

https://medium.com/@HadrienD/how-to-customize-font-size-in-r-markdown-documents-f5adff36e2cc

gt::gt()

- RStudio has been working on GT for years...

```
library(gt)
wide %>%
  gt()
```

Gender	Yes	No	Total
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

`gt::gt()` Code Duplicates the Book [1] 😊

```
library(gt)
wide %>%
  gt() %>%
  tab_header(
    title = "Table 2.1 Cross-classification of belief in afterlife by gender."
  ) %>%
  tab_spanner(
    label = "Belief in the Afterlife",
    columns = 2:3
  ) %>%
  tab_source_note(
    source_note = "<em>Source:</em> Data from 2016 General Social Survey."
  ) %>%
  tab_options(
    table_body.hlines.color = "white",
    table_body.border.bottom.color = "#D3D3D3",
    table.border.bottom.color = "white"
  )
```

`gt::gt()` Code Duplicates the Book $_2$ 😊

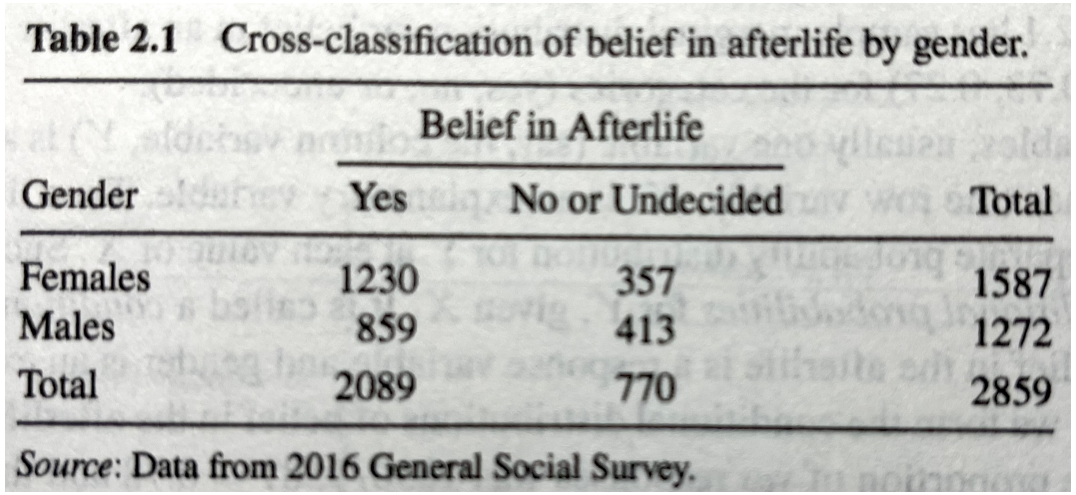


Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in Afterlife: Yes	Belief in Afterlife: No or Undecided	Total
Females	1230	357	1587
Males	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

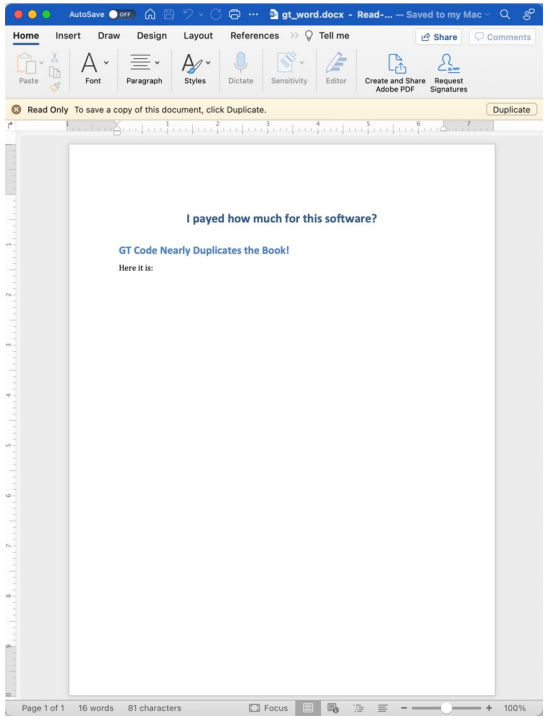
Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in the Afterlife: Yes	Belief in the Afterlife: No or Undecided	Total
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Source: Data from 2016 General Social Survey.

GT and Word

- People have been asking for a direct export into Word for *years*.
- Currently, you get nothing when you knit. 😠
- You can copy and paste GT tables from HTML output into a Word file. 🤢
- I hope something in Quarto is going to fix it. 🙏



Clean Exports into Word

I know of two options:

- The `flextable` package is what I have used for hardcore tables.
  - I have been using it for years and I still think it is work.
  - Search my Agresti notes on GitHub for examples: https://github.com/RaymondBalise/Agresti_IntroToCategorical
  - https://ardata-fr.github.io/flextable-book/index.html
- I discovered the `huxtable` package in 2022 and I am exploring it.

`flextable::flextable()` 1

- The weird thing is that you will want to specify the header first.
- Add a value for every variable/column in every row of the header.
- If you put the same words into multiple rows (like "Gender" and "Total" in the example below), the values can be merged later with `merge_*()` commands.

```
suppressPackageStartupMessages(library(flextable))

my_header <- data.frame(
  col_keys = colnames(wide),
  line1 = c("Gender", rep("Belief in the Afterlife", 2), "Total"),
  # line2 is Gender,        Yes, No,                      Total
  line2 = colnames(wide)
)
```

`flextable::flextable()` 2

- You specify the columns using the `col_keys` that you previously defined in the data frame and then tell it to use the data frame to make the columns.
- Notice the merging of the cells with identical content in the header.
- Column 1 (Gender) will be left-aligned. The rest of the columns are centered. Sadly, `flextable` can not decimal-align.

```
wide %>%
  flextable(col_keys = my_header$col_keys) %>%
  set_header_df(
    mapping = my_header,
    key = "col_keys"
  ) %>%
  theme_booktabs() %>%
  autofit(part = "all") %>%
  align_nottext_col(align = "center") %>%
  merge_h(part = "header") %>%
  merge_v(part = "header") %>%
  set_caption(caption = "Table 2.1 Cross-classification of belief in afterlife by gender.") %>%
  add_footer_lines(values = "Source: Data from 2016 General Social Survey.")
```

`flextable::flextable()` 3

Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in the Afterlife		Total
	Yes	No	
Female	1,230	357	1,587
Male	859	413	1,272
Total	2,089	770	2,859

Source: Data from 2016 General Social Survey.

- Unfortunately, the numbers are not decimal-aligned.

```
wide

my_header <- data.frame(
  col_keys = colnames(wide),
  line1 = c("Gender", rep("Belief in th
# line2 is   Gender,              Yes, N
  line2 = colnames(wide)
)

my_header

wide %>%
  flextable(col_keys = my_header$col_ke
  set_header_df(
    mapping = my_header,
    key = "col_keys"
  ) %>%
  theme_booktabs() %>%
  autofit(part = "all") %>%
  align_nottext_col(align = "center") %
  merge_h(part = "header") %>%
  merge_v(part = "header") %>%
  set_caption(caption = "Table 2.1 Cros
  add_footer_lines(values = "Source: Da
```

```
# A tibble: 3 × 4
  Gender   Yes    No Total
  <chr>  <dbl> <dbl> <dbl>
1 Female  1230   357  1587
2 Male     859   413  1272
3 Total   2089   770  2859

  col_keys                   line1  line2
1   Gender                 Gender Gender
2      Yes Belief in the Afterlife    Yes
3       No Belief in the Afterlife     No
4    Total                  Total  Total
```

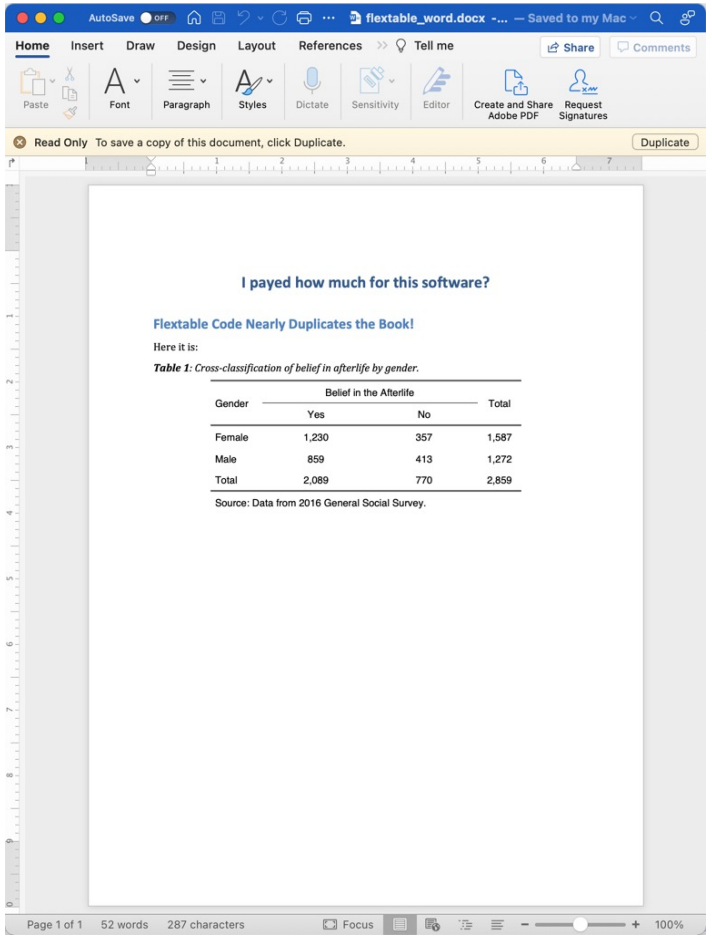
Table 2.1 Cross-classification of belief in afterlife by gender.

Gender	Belief in the Afterlife		Total
	Yes	No	
Female	1,230	357	1,587
Male	859	413	1,272
Total	2,089	770	2,859

Source: Data from 2016 General Social Survey.

`flextable::flextable()` 4

- The export to Word is cleaner than the rendering in the slides.
- Notice that it will automatically number the tables.



Calculate Statistics for a Table on the Fly

Playing with Babies

The Most Impactful Tables in Medicine

Each New England Journal of Medicine article gets cited something like 75 times and it has a weekly readership in excess of 600,000. Their tables are all over the twitterverse.

The `medicaldata` package has data from *Michalowicz et al., 'Treatment of periodontal disease and the risk of preterm birth', N Engl J Med 2006; 355:1885-1894. DOI: 10.1056/NEJMoa062249*

https://www.kolabtree.com/blog/top-20-medical-journals-for-physicians-to-publish-in/

Table 1. Baseline Characteristics of the Study Patients.*

Characteristic	Control Group (N=410)	Treatment Group (N=413)	P Value
Age — yr	25.9±5.5	26.1±5.6	0.56
Race or ethnic group — no. (%)†			
White	119 (29.0)	116 (28.1)	0.77
Black	182 (44.4)	190 (46.0)	0.64
Hispanic	180 (43.9)	170 (41.2)	0.43
Education — no. (%)			0.88
≤8 yr	76 (18.5)	78 (18.9)	
9–12 yr	242 (59.0)	237 (57.4)	
>12 yr	92 (22.4)	98 (23.7)	
Mean gestational age of fetus — wk	15.0±1.3	15.0±1.3	0.85
Previous pregnancies — no. (%)‡			
Any pregnancy	305 (74.4)	306 (74.1)	0.92
Live preterm birth§	44 (16.5)	33 (12.5)	0.18
Spontaneous abortion¶	94 (30.8)	108 (35.3)	0.24
Induced abortion¶	67 (22.0)	52 (17.0)	0.12
Stillbirth¶	6 (2.0)	9 (2.9)	0.44
Coexisting medical condition — no. (%)			
Diabetes	8 (2.0)	16 (3.9)	0.10
Chronic hypertension	9 (2.2)	16 (3.9)	0.16
Self-reported drug addiction	7 (1.7)	15 (3.6)	0.09
Self-reported alcohol use	8 (2.0)	8 (1.9)	0.99
Eating disorder	0	2 (0.5)	0.16
Dental status			
Number of natural teeth	26.8±1.7	26.7±1.8	0.67
Number of qualifying teeth	14.4±6.7	15.2±6.8	0.08
Percent of tooth sites that bled on probing	69.0±17.1	69.6±17.4	0.62
Percent of tooth sites with probing depth ≥4 mm	24.8±15.9	26.5±16.6	0.13

* Plus–minus values are means ±SD.
† Race or ethnic group was reported by the patients. Some women selected more than one category and were included in all.
‡ Some patients reported more than one event.
§ Percentages are based on 266 women in the control group and 265 women in the treatment group who had had any live births.
¶ Percentages are based on 305 women in the control group and 306 women in the treatment group who had had any previous pregnancies.

In Table 1 (Typically Demographics)

- Biostatisticians talk about:
  - Nominal - Alive/Dead
  - Ordinal - Small, Medium, Large
  - Interval/Ratio - Numbers
  - Time to Event - Survival
- Where do not mutually exclusive categories fit?
  - Notice Race and Ethnicity are mixed.
- Why are their p-values here?
  - What comparisons are done?

Table 1. Baseline Characteristics of the Study Patients.*

Characteristic	Control Group (N=410)	Treatment Group (N=413)	P Value
Age — yr	25.9±5.5	26.1±5.6	0.56
Race or ethnic group — no. (%)†			
White	119 (29.0)	116 (28.1)	0.77
Black	182 (44.4)	190 (46.0)	0.64
Hispanic	180 (43.9)	170 (41.2)	0.43
Education — no. (%)			0.88
≤8 yr	76 (18.5)	78 (18.9)	
9–12 yr	242 (59.0)	237 (57.4)	
>12 yr	92 (22.4)	98 (23.7)	

* Plus–minus values are means ±SD.
† Race or ethnic group was reported by the patients. Some women selected more than one category and were included in all.

In Table 1 (*P*-Values)

- Age - Numeric
- Race - Each category is distinct
- Education - Categorical
- It would have been nice to know what tests were done here...

Table 1. Baseline Characteristics of the Study Patients.*

Characteristic	Control Group (N=410)	Treatment Group (N=413)	P Value
Age — yr	25.9±5.5	26.1±5.6	0.56
Race or ethnic group — no. (%)†			
White	119 (29.0)	116 (28.1)	0.77
Black	182 (44.4)	190 (46.0)	0.64
Hispanic	180 (43.9)	170 (41.2)	0.43
Education — no. (%)			0.88
≤8 yr	76 (18.5)	78 (18.9)	
9–12 yr	242 (59.0)	237 (57.4)	
>12 yr	92 (22.4)	98 (23.7)	

* Plus–minus values are means ±SD.
† Race or ethnic group was reported by the patients. Some women selected more than one category and were included in all.

When I Do Exploratory Data Analysis (EDA)

- I start with `skimr::skim()`.
- I quickly move to the `table1` package.
  - Excellent out of the box behavior!
  - It does not do well on the *remove unit repetition* criteria.

```
opt_df <- medicaldata::opt %>%
  select(Age, Black, White, Hisp, Education, Group)

library(table1)
table1(~ . | Group, data = opt_df, overall = FALSE)
```

	C (N=410)	T (N=413)
Age		
Mean (SD)	25.9 (5.51)	26.1 (5.62)
Median [Min, Max]	25.0 [16.0, 44.0]	25.0 [16.0, 44.0]
Black		
No	228 (55.6%)	223 (54.0%)
Yes	182 (44.4%)	190 (46.0%)
White		
No	291 (71.0%)	297 (71.9%)
Yes	119 (29.0%)	116 (28.1%)
Hisp		
	70 (17.1%)	75 (18.2%)
No	160 (39.0%)	168 (40.7%)
Yes	180 (43.9%)	170 (41.2%)
Education		
8-12 yrs	242 (59.0%)	237 (57.4%)
LT 8 yrs	76 (18.5%)	78 (18.9%)
MT 12 yrs	92 (22.4%)	98 (23.7%)

Process a Couple of Variables

```
make_logical <- function(x) {
  as.logical(x == "Yes") # Combine no and missing ... ummmm.
}

opt_data <-
  as_tibble(medicaldata::opt) %>%
  mutate(
    Group = case_when(
      Group == "C" ~ "Control Group",
      Group == "T" ~ "Treatment Group"
    )
  ) %>%
  mutate(
    `Education (yrs)` = case_when(
      Education == "LT 8 yrs " ~ "≤8",
      Education == "8-12 yrs " ~ "9-12",
      Education == "MT 12 yrs" ~ ">12"
    ),
    `Education (yrs)` = factor(`Education (yrs)`, levels = c("≤8", "9-12", ">12"))
  ) %>%
  mutate(across(c(Black, White, Hisp), make_logical)) %>%
  select(
    Age, Black, White, Hisp, `Education (yrs)`, Group
  )
```

{table1} For Quick, Presentable Tables

```
library(table1)
table1(~ . | Group, data = opt_data, overall = FALSE)
```

	Control Group (N=410)	Treatment Group (N=413)
Age		
Mean (SD)	25.9 (5.51)	26.1 (5.62)
Median [Min, Max]	25.0 [16.0, 44.0]	25.0 [16.0, 44.0]
Black		
Yes	182 (44.4%)	190 (46.0%)
No	228 (55.6%)	223 (54.0%)
White		
Yes	119 (29.0%)	116 (28.1%)
No	291 (71.0%)	297 (71.9%)
Hisp		
Yes	180 (43.9%)	170 (41.2%)
No	230 (56.1%)	243 (58.8%)
Education (yrs)		
≤8	76 (18.5%)	78 (18.9%)
9-12	242 (59.0%)	237 (57.4%)
>12	92 (22.4%)	98 (23.7%)

1. Heads ✅
2. Dividers ✅
3. Right-Align Numbers ❌
4. Left-Align Text ✅
5. Precision ✅
6. Space ✅
7. Remove Repetition ❌
8. Outliers
9. Group Similar ✅
10. Visualizations
11. Key Point(s)
12. Annotations
13. Caption/Title ✅

Drop a Binary Level

- If you only want to show a single level of binary factor, you can tweak the render function. Be careful because this masks missing data (like in the NEJM article).

```
# https://github.com/benjaminrich/table1/issues/48
rndr <- function(x, ...) {
  y <- render.default(x, ...)
  if (is.logical(x)) y[2] else y
}

table1(~ . | Group, data = opt_data, overall = FALSE, render = rndr)
```

	Control Group (N=410)	Treatment Group (N=413)
Age		
Mean (SD)	25.9 (5.51)	26.1 (5.62)
Median [Min, Max]	25.0 [16.0, 44.0]	25.0 [16.0, 44.0]
Black	182 (44.4%)	190 (46.0%)
White	119 (29.0%)	116 (28.1%)
Hisp	180 (43.9%)	170 (41.2%)
Education (yrs)		
≤8	76 (18.5%)	78 (18.9%)
9-12	242 (59.0%)	237 (57.4%)
>12	92 (22.4%)	98 (23.7%)

`table1::table1()`

- If you don't need to add subheaders (like over Race/Ethnicity here) and you don't need p-values, the `table1` package is ideal.
- The author is wonderfully quick to answer questions.

The `gtsummary` Package is Ridiculously Great

- The `table1` package gives a quick, excellent summary but use `{gtsummary}` when you want to include inferential statistics (p-values and confidence limits).

```
suppressMessages(library(gtsummary))

reset_gtsummary_theme()
suppressMessages(theme_gtsummary_compact())

opt_data %>%
  select(
    Age, Black, White, Hisp,
    `Education (yrs)`, Group
  ) %>%
  tbl_summary(by = Group) %>%
  add_p()
```

Characteristic	**Control Group**, N = 410[1]	**Treatment Group**, N = 413[1]	**p-value**[2]
Age	25 (22, 30)	25 (22, 30)	0.6
Black	182 (44%)	190 (46%)	0.6
White	119 (29%)	116 (28%)	0.8
Hisp	180 (44%)	170 (41%)	0.4
Education (yrs)			0.9
≤8	76 (19%)	78 (19%)	
9-12	242 (59%)	237 (57%)	
>12	92 (22%)	98 (24%)	

[1] Median (IQR); n (%)

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

Tweaking the Display Details is Easy

```
suppressMessages(library(gtsummary))

reset_gtsummary_theme()
suppressMessages(theme_gtsummary_compact())

opt_data %>%
  select(
    Age, Black, White, Hisp,
    `Education (yrs)`, Group
  ) %>%
  # show mean to 1 decimal place and SD to 2 dec
  tbl_summary(
    by = Group,
    digits = list(
      all_continuous() ~ c(1, 2, 2),
      all_categorical() ~ c(0, 1)
    )
  ) %>%
  add_p()
```

Characteristic	Control Group, N = 410[1]	Treatment Group, N = 413[1]	p-value[2]
Age	25.0 (22.00, 29.75)	25.0 (22.00, 30.00)	0.6
Black	182 (44.4%)	190 (46.0%)	0.6
White	119 (29.0%)	116 (28.1%)	0.8
Hisp	180 (43.9%)	170 (41.2%)	0.4
Education (yrs)			0.9
≤8	76 (18.5%)	78 (18.9%)	
9-12	242 (59.0%)	237 (57.4%)	
>12	92 (22.4%)	98 (23.7%)	

[1] Median (IQR); n (%)

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

Adding Subgroups is Not Too Bad

```
suppressMessages(library(gtsummary))

reset_gtsummary_theme()
suppressMessages(theme_gtsummary_compact())

t1 <- opt_data %>%
  select(Age, `Education (yrs)`, Group) %>%
  tbl_summary(by = Group) %>%
  bold_labels() %>%
  add_p()

t2 <- opt_data %>%
  select(Black, White, Hisp, Group) %>%
  tbl_summary(by = Group) %>%
  bold_labels() %>%
  add_p()

tbl_stack(
  list(t1, t2),
  group_header = c(" ", "Race/Ethnicity")
)
```

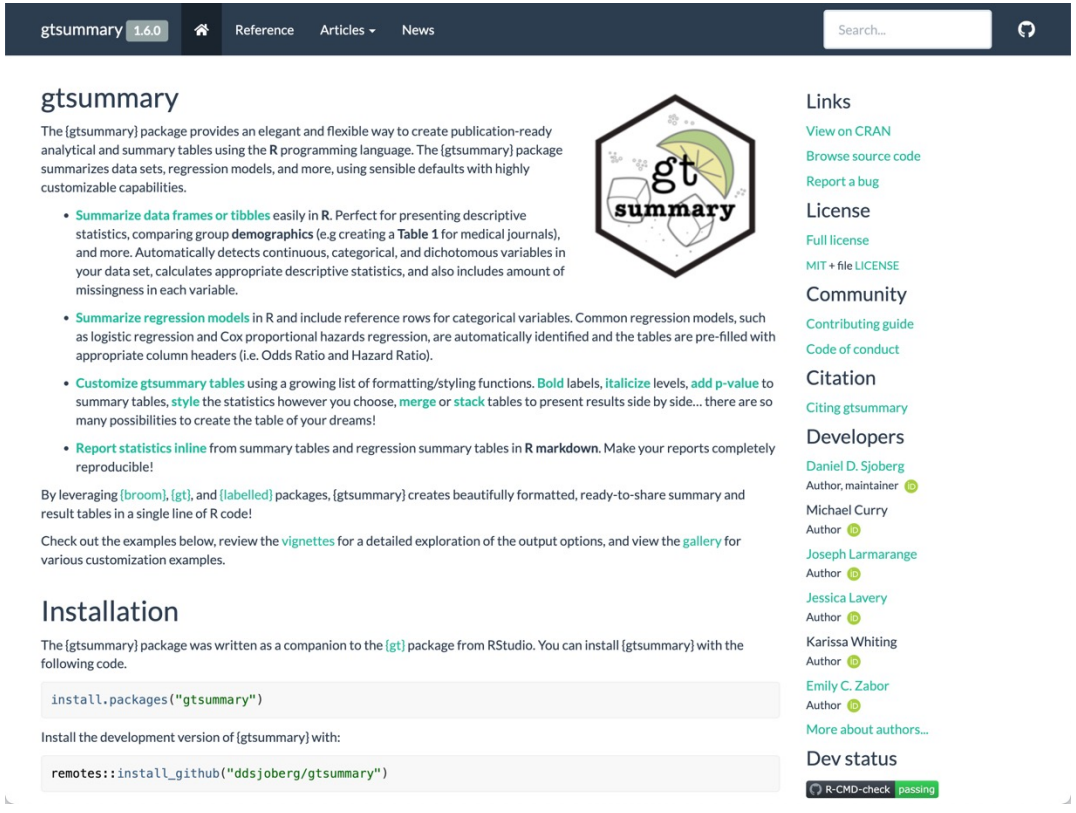
Characteristic	**Control Group**, N = 410[1]	**Treatment Group**, N = 413[1]	**p-value**[2]
Age	25 (22, 30)	25 (22, 30)	0.6
Education (yrs)			0.9
≤8	76 (19%)	78 (19%)	
9-12	242 (59%)	237 (57%)	
>12	92 (22%)	98 (24%)	
Race/Ethnicity			
Black	182 (44%)	190 (46%)	0.6
White	119 (29%)	116 (28%)	0.8
Hisp	180 (44%)	170 (41%)	0.4

[1] Median (IQR); n (%)

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

The vignettes for the `gtsummary` package are extraordinary.

https://www.danieldsjoberg.com/gtsummary/index.html



https://www.danieldsjoberg.com/gtsummary/articles/tbl_summary.html
https://www.danieldsjoberg.com/gtsummary/articles/gallery.html

Surgical Outcomes and Bartending with R...

gt + gtsummary + flextable + rUM

Remember that `laryngoscope` data?

- The online version of the publication has this (blurry) static image made from the table.

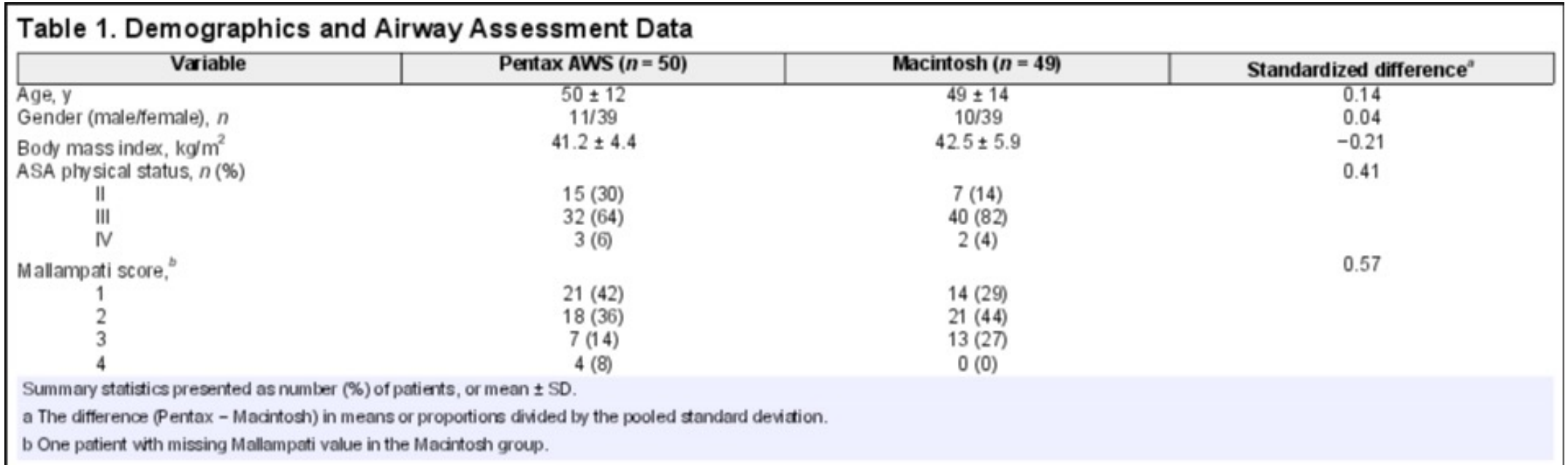


Table 1. Demographics and Airway Assessment Data

Variable	Pentax AWS (*n* = 50)	Macintosh (*n* = 49)	Standardized difference[a]
Age, y	50 ± 12	49 ± 14	0.14
Gender (male/female), *n*	11/39	10/39	0.04
Body mass index, kg/m^2	41.2 ± 4.4	42.5 ± 5.9	−0.21
ASA physical status, *n* (%)			0.41
II	15 (30)	7 (14)	
III	32 (64)	40 (82)	
IV	3 (6)	2 (4)	
Mallampati score,[b]			0.57
1	21 (42)	14 (29)	
2	18 (36)	21 (44)	
3	7 (14)	13 (27)	
4	4 (8)	0 (0)	

Summary statistics presented as number (%) of patients, or mean ± SD.

a The difference (Pentax – Macintosh) in means or proportions divided by the pooled standard deviation.

b One patient with missing Mallampati value in the Macintosh group.

- What is it like to recreate it?

Load a couple of packages.

```
library(conflicted)
conflict_prefer("select", "dplyr", quiet = TRUE)
conflict_prefer("filter", "dplyr", quiet = TRUE)

conflict_prefer("continuous_summary", "gtsummary", quiet = TRUE)
conflict_prefer("as_flextable", "gtsummary", quiet = TRUE)
conflict_prefer("void", "reactablefmtr", quiet = TRUE)

library(medicaldata)
library(gtable)
suppressPackageStartupMessages(library(gt))
suppressPackageStartupMessages(library(gtsummary))
library(flextable)
library(dplyr)
library(forcats)
library(labelled)
suppressPackageStartupMessages(require(gdtools))
library(officer)
library(smd)
```

Preprocess

```
library(forcats) # for fct_rev()
analysis <- medicaldata::laryngoscope %>%
  mutate(
    male = gender == 1,
    female = gender == 0,
    asa_rm = factor(as.character(as.roman(asa))), # asa as roman numerals
    Mallampati_f = factor(Mallampati, ordered = T)
  ) %>%
  mutate(
    laryngoscope = if_else(
      Randomization == 0,
      "MacIntosh",
      "Pentax AWS"
    ),
    laryngoscope = fct_rev(laryngoscope) # reverse order
  ) %>%
  select(laryngoscope, age, male, female, BMI, asa_rm, Mallampati_f)
```

Add labels from the `{labelled}` package.

```
analysis <-
  analysis %>%
  set_variable_labels(
    age = "Age, y",
    male = "Males",
    female = "Females",
    BMI = "Body mass index, kg/m²",
    asa_rm = "ASA physical status, n (%)",
    Mallampati_f = "Mallampati score,ᵇ"
  )
```

Add a function to calculate the standardized differences.

```
smd_calc <- function(data, variable, by, ...) {
  smd::smd(x = data[[variable]], g = data[[by]], na.rm = TRUE)["estimate"]
}
```

Make the {gtsummary} table.

- This is the trickery you saw before, plus we are adding in custom formatting for the continuous variables a new statistic.

```
lar_data_gt <- analysis %>%
  tbl_summary(
    by = "laryngoscope",
    statistic = list(all_continuous() ~ "{mean} ± {sd}")
  ) %>%
  add_stat(
    fns = list(
      all_continuous() ~ smd_calc,
      all_categorical() ~ smd_calc
    )
  ) %>%
  modify_header(estimate ~ "Standardized difference ª") %>%
  modify_fmt_fun(update = estimate ~
    function(x) {
      scales::number(x, accuracy = 0.01)
    })

lar_data_gt
```

Characteristic	**Pentax AWS**, N = 50[1]	**MacIntosh**, N = 49[1]	Standardized difference [a]
Age, y	50 ± 12	49 ± 14	0.14
Males	11 (22%)	10 (20%)	0.04
Females	39 (78%)	39 (80%)	-0.04
Body mass index, kg/m²	41.4 ± 4.4	42.5 ± 5.9	-0.21
Unknown	2	0	
ASA physical status, n (%)			0.41
II	15 (30%)	7 (14%)	
III	32 (64%)	40 (82%)	
IV	3 (6.0%)	2 (4.1%)	
Mallampati score,[b]			0.57
1	21 (42%)	14 (29%)	
2	18 (36%)	21 (44%)	
3	7 (14%)	13 (27%)	
4	4 (8.0%)	0 (0%)	
Unknown	0	1	

[1] Mean ± SD; n (%)

{gtsummary} Plays Well with Others...

This is optimized for Word....

```
# Convert to flextable + additional customization:
big_border <- fp_border(color = "black", width = 2) # create bold border type
fontname <- "Avenir Next Condensed" # create font closely matching manuscript

flex <-
  as_flex_table(
    lar_data_gt,
  ) %>%
  delete_part(part = "footer") %>%
  add_header_lines(values = "Table 1. Demographics and Airway Assessment Data") %>%
  add_footer_lines(values = "Summary statisticis presented as number (%) of patients , or mean ± SD.
  a The difference (Pentax - Macintosh) in means or proportions divided by the pooled standard deviati
  b One Patient with missing Mallampati value in the Macintosh group.") %>%
  bg(bg = "Light Grey", i = 2, part = "header") %>%
  fontsize(i = 1, size = 14, part = "header") %>%
  bold(part = "header") %>%
  bg(bg = "lavender", part = "footer") %>%
  border_remove() %>%
  surround(i = 2, j = 1:4, border = big_border, part = "header") %>%
  font(fontname = fontname, part = "all") %>%
  line_spacing(space = 1.5, part = "footer") %>%
  line_spacing(space = .3, part = "body")
```

Strong `flex`... Suboptimal in `{xaringan}`

```
flex
```

Table 1. Demographics and Airway Assessment Data

Characteristic	Pentax AWS, N = 50[1]	MacIntosh, N = 49[1]	Standardized difference [a]
Age, y	50 ± 12	49 ± 14	0.14
Males	11 (22%)	10 (20%)	0.04
Females	39 (78%)	39 (80%)	-0.04
Body mass index, kg/m²	41.4 ± 4.4	42.5 ± 5.9	-0.21
Unknown	2	0	
ASA physical status, n (%)			0.41
II	15 (30%)	7 (14%)	
III	32 (64%)	40 (82%)	
IV	3 (6.0%)	2 (4.1%)	
Mallampati score,[b]			0.57
1	21 (42%)	14 (29%)	
2	18 (36%)	21 (44%)	
3	7 (14%)	13 (27%)	
4	4 (8.0%)	0 (0%)	
Unknown	0	1	

Summary statisticis presented as number (%) of patients , or mean ± SD.

a The difference (Pentax - Macintosh) in means or proportions divided by the pooled standard deviation.

b One Patient with missing Mallampati value in the Macintosh group.

Strong `flex` Rendered in HTML

Elit libero montes tristique scelerisque donec arcu. Dis mi ad suspendisse libero senectus. Purus tempus orci neque volutpat. Pellentesque porta torquent posuere accumsan id litora fringilla rutrum metus consequat convallis sed donec mauris inceptos cras nostra diam praesent sollicitudin volutpat euismod.

Sit tempus erat dapibus nisl tortor! Praesent egestas dictumst ridiculus sem enim gravida penatibus. Posuere arcu torquent ullamcorper pretium lobortis pretium malesuada tortor malesuada! Cum quisque dictum primis dis dis neque sagittis aptent egestas ac nunc arcu.

Table 1. Demographics and Airway Assessment Data

Characteristic	Pentax AWS, N = 50[1]	MacIntosh, N = 49[1]	Standardized difference [a]
Age, y	50 ± 12	49 ± 14	0.14
Males	11 (22%)	10 (20%)	0.04
Females	39 (78%)	39 (80%)	-0.04
Body mass index, kg/m²	41.4 ± 4.4	42.5 ± 5.9	-0.21
Unknown	2	0	
ASA physical status, n (%)			0.41
II	15 (30%)	7 (14%)	
III	32 (64%)	40 (82%)	
IV	3 (6.0%)	2 (4.1%)	
Mallampati score,[b]			0.57
1	21 (42%)	14 (29%)	
2	18 (36%)	21 (44%)	
3	7 (14%)	13 (27%)	
4	4 (8.0%)	0 (0%)	
Unknown	0	1	

Summary statisticis presented as number (%) of patients , or mean ± SD.

a The difference (Pentax - Macintosh) in means or proportions divided by the pooled standard deviation.

b One Patient with missing Mallampati value in the Macintosh group.

5 Conclusion

Lorem ridiculus nunc consequat praesent facilisi rhoncus dis. Imperdiet quisque felis risus urna porttitor. Tortor vel scelerisque nibh pellentesque vehicula feugiat integer ante tristique morbi ligula tellus mattis a leo vivamus facilisi sodales curabitur inceptos rutrum.

Adipiscing eleifend nisl suscipit massa ante parturient ligula enim phasellus fusce proin. At eget himenaeos porta tempor mollis in litora

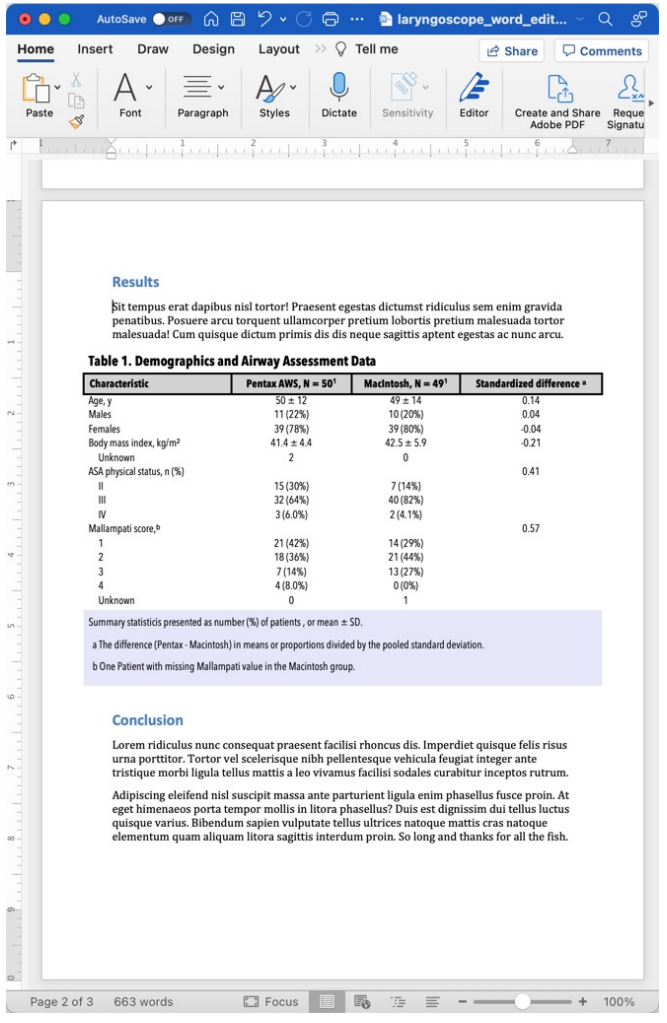
Add Some `{rUM}`

- Your R-loving friends at the University of Miami made a package called rUM https://raymondbalise.github.io/rUM/
- It has a function `rUM::make_project()` that will make an RStudio project that contains:
  - An R Markdown file that has an academic paper shell
  - A bibliography file for used R packages
  - A `.gitignore` file designed to help protect data leakage

rUM + Word

With a tiny tweak to the output, you get a word document.

```
---
title: "Analysis of medicaldata::laryngoscope"
author: "Raymond R. Balise and Lauren Nahodyl"
date: "2022-06-20"
output:
  rmarkdown::word_document
bibliography: [packages.bib]
csl: the-new-england-journal-of-medicine.csl
---
```



References - HTML

If I am making HTML output, I typically make a custom title, in a different font and color. I add a "tag" to that title, then link to it.

In your Rmd files, you can reference tables with Markdown like:

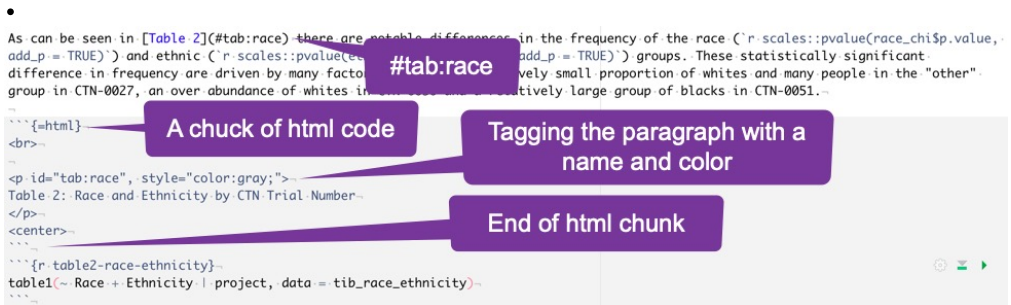
```
[Table 6.1](#tab:tab6-1)
```

Race and Ethnicity

As can be seen in Table 2 there are notable differences in the frequency of the race (p<0.001) and ethnic (p<0.001) groups. These statistically significant difference in frequency are driven by many factors including a relatively small proportion of whites and many people in the “other” group in CTN-0027, an over abundance of whites in CTN-0030 and a relatively large group of blacks in CTN-0051.

Table 2: Race and Ethnicity by CTN Trial Number

	CTN-0027 (N=1269)	CTN-0030 (N=653)	CTN-0051 (N=570)	Overall (N=2492)
Race				
Refused/Missing	6 (0.5%)	1 (0.2%)	6 (1.1%)	13 (0.5%)
Black	128 (10.1%)	21 (3.2%)	73 (12.8%)	222 (8.9%)
Other	229 (18.0%)	35 (5.4%)	70 (12.3%)	334 (13.4%)
White	906 (71.4%)	596 (91.3%)	421 (73.9%)	1923 (77.2%)
Ethnicity				
Hispanic	206 (16.2%)	31 (4.7%)	99 (17.4%)	336 (13.5%)
Not Hispanic	1063 (83.8%)	622 (95.3%)	471 (82.6%)	2156 (86.5%)



References - Work with the `bookdown` Package

- If you are writing to Word or PDFs, you can use the built-in R Markdown tools for creating links.
- You will want to change the output type to use the `bookdown` package.

```
output:
- bookdown::html_document2
- bookdown::pdf_document2
- bookdown::word_document2
```

- You can label tables with captions and then use `\@ref(label)` to refer to them.

Using `kable` or `flextable`

After changing the output to be bookdown, name a code chunk with a name like *simple-example*

```
wide %>%
  knitr::kable(
    caption = "The raw religon data printed with kable."
  )
```

The raw religon data printed with kable.

Gender	Yes	No	Total
Female	1230	357	1587
Male	859	413	1272
Total	2089	770	2859

Then reference the name of the table with code like this:

```
The simple kable example: \@ref(tab:simple-example)
```

Word with Table Reference

````
---
title: "I payed how much for this software?"
output: bookdown::word_document2
---

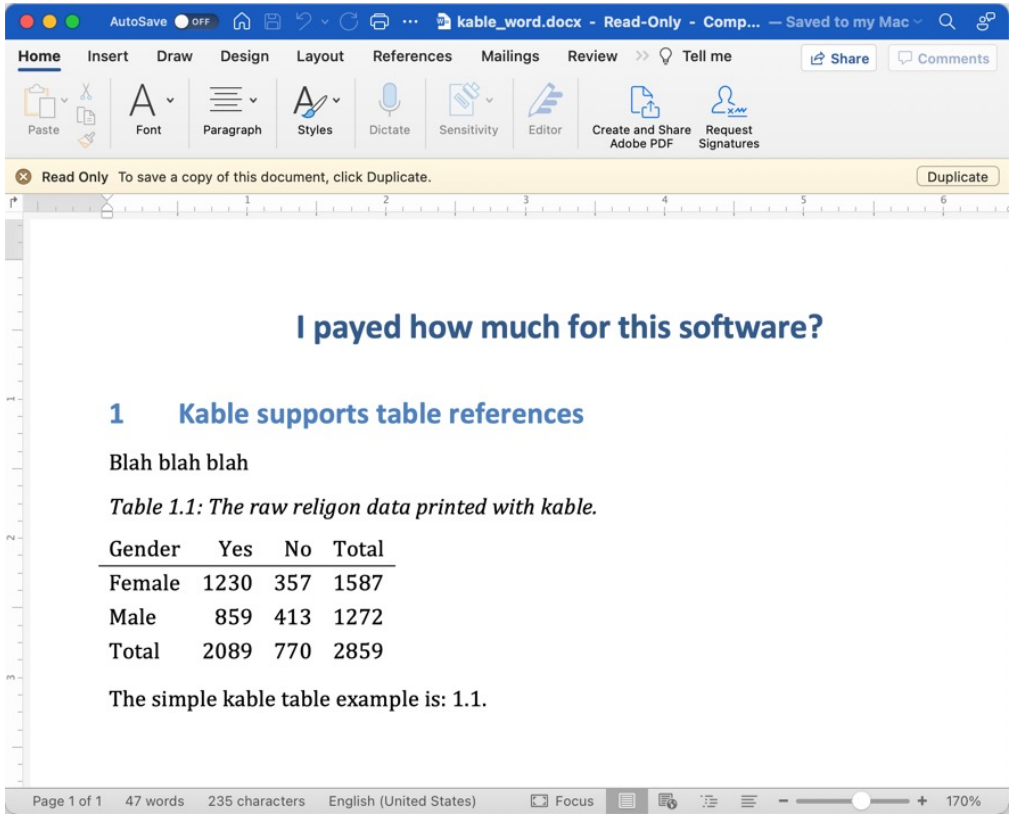
# Kable supports table references

Blah blah blah

```{r echo=FALSE}```


```{r simple-example, echo=FALSE}
knitr::kable(
  wide,
  caption = "The raw religon data printed with kable."
)
```

The simple kable table example is: \@ref(tab:simple-example).
````


````

Be Dynamic

Static tables are so last century!

Dynamic Tables

- There are two excellent packages for making dynamic tables, `DT` and `reactable`
- `DT` is widely used.
- `reactable` is new(ish) but it is built on top of React which is a **huge** deal (it is made by Facebook/Meta).

DT

- With `DT` you get a fully interactive set of controls.

```
set.seed(123)
the_start <-
  medicaldata::laryngoscope %>%
  slice_head(n = 100)

the_start %>%
  DT::datatable()
```

Show 10 entries Search:

	age	gender	asa	BMI	Mallampati	Randomization	attempt1_time	attempt1_S
1	51	0	3	56.2	1	0	29	
2	52	0	3	44.6	2	0	29	

reactable

- With `reactable` you also get dynamic displays with controls.

```
the_start %>%
  reactable::reactable()
```

age	gender	asa	BMI	Mallampati	Randomization	attempt1_time	attempt1_S_F	attempt2_time	attempt2_assigned_method	attempt2_S_F
51	0	3	56.2	1	0	29	1			
52	0	3	44.6	2	0	29	1			
37	0	3	41.6	1	0	31	0	60	1	1
20	0	3	46.32	2	0	31	0	46	1	1

Favorite `reactable` Featuers

- The ability to collapse/expand groups
- Embed graphics

Grab a Few Variables

```
lar_data <- as_tibble(medicaldata::laryngoscope) %>% #
  mutate(
    Sex =
      case_when(
        gender == 1 ~ "Male",
        gender == 0 ~ "Female"
      )
  ) %>%
  mutate(
    Laryngoscope =
      if_else(Randomization == 0, "MacIntosh", "Pentax AWS")
  ) %>%
  mutate(
    # asa as roman numerals to match the table. This messes with sorting.
    asa_rm = factor(as.character(as.roman(asa)), levels = c("II", "III", "IV"), ordered = TRUE)
  ) %>%
  mutate(
    Mallampati_f = factor(Mallampati, ordered = T) # as factor and ordered
  ) %>%
  mutate(Time = round(total_intubation_time)) %>%
  rename(
    `American Society of Anesthesiologists physical status` = asa_rm
  ) %>%
  select(Laryngoscope, Time, `American Society of Anesthesiologists physical status`, Sex) %>%
  arrange(
      "Laryngoscope", "American Society of Anesthesiologists physical status"
    )
```

Subgroups 1

```
library(reactable)
lar_data %>%
  reactable(
    groupBy = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    ),
    defaultSorted = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    )
  )
```

Subgroups 2

Laryngoscope ↑	American Society of Anesthesiologists physical status ↑	Time	Sex
▸ MacIntosh (3)			
▸ Pentax AWS (3)			

Agregate data 1

```
lar_data %>%
  reactable(
    groupBy = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    ),
    defaultSorted = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    ),
    columns = list(
      `American Society of Anesthesiologists physical status` =
        colDef(
          aggregate = "frequency"
        ),
      Sex = colDef(
        aggregate = "frequency"
      ),
      Time = colDef(
        aggregate = "mean", format = colFormat(digits = 0)
      )
    )
```

Agregate data 2

Laryngoscope ↑	American Society of Anesthesiologists physical status ↑	Time	Sex
▸ MacIntosh (3)	III (40), II (7), IV (2)	30	Female (39), Male (10)
▸ Pentax AWS (3)	III (32), II (15), IV (3)	45	Female (39), Male (11)

Add Art With or Without {reactablefmtr} 1

```
library(htmltools)
library(reactablefmtr)

lar_data %>%
  reactable(
    groupBy = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    ),
    defaultSorted = c(
      "Laryngoscope",
      "American Society of Anesthesiologists physical status"
    ),
    columns = list(
      `American Society of Anesthesiologists physical status` =
        colDef(
          aggregate = "frequency"
        ),
      Sex = colDef(
        aggregate = "frequency"
      ),
      Time = colDef(
        aggregate = "mean", format = colFormat(digits = 0)
```

Add Art With or Without `{reactablefmtr}` 2

```
Attaching package: 'reactablefmtr'

The following object is masked from 'package:flextable':

    void
```

Laryngoscope ↑	American Society of Anesthesiologists physical status ↑	Time	Sex
▸ MacIntosh (3)	III (40), II (7), IV (2)	30	Female (39), Male (10)
▸ Pentax AWS (3)	III (32), II (15), IV (3)	45	Female (39), Male (11)

I expect incredible things from `reactable`.

- Expect crazy fast performance and beautiful magic.
- Explore: https://glin.github.io/reactable/articles/examples.html
  - The Grouping and Aggregation example is like whoa. 😍
  - The embedding of html widgets is like my brain goes ... 💥.
- Watch for add-on packages like https://kcuilla.github.io/reactablefmtr/index.html

Where to Learn More

Gina Reynolds, author of the `flextable` package, has made great tutorials:

- kableExtra: https://evamaerey.github.io/tables/kableextra.html
- gt: https://evamaerey.github.io/tables/gt.html
- huxtable: https://evamaerey.github.io/tables/huxtable.html