R CODA Club - Creating a descriptives table using gtsummary and other bits and bobs

Caitlin A. Howlett

2022-10-19

Basic descriptives table

An example:

Participants

Eight adults in the healthy control group had a mean age of 33.8 (11.34). Participant characteristics are displayed in Table 1.

Table 1: **Participant Characteristics** (n = 16)

Characteristic	**hc**, (n = 8)	**pp**, (n = 8)
Age in years		
Mean (SD)	33.8 (11.34)	55.9 (18.53)
Range	19 - 51	18 - 76
Gender		
Female	4 (57%)	3 (43%)
Male	2 (29%)	3 (43%)
Non-binary	1 (14%)	1 (14%)
Missing	1	1
Education		
High school	1 (12%)	1 (12%)
Higher degree	3 (38%)	5 (62%)
Tertiary	4 (50%)	2 (25%)
WCST score		
Mean (SD)	35.0 (15.46)	16.1 (11.06)
Range	13 - 56	9 - 43
Stroop score		
Mean (SD)	52.4 (31.55)	46.0 (17.27)
Range	24 - 120	22 - 67

Stacking tables together

Table 2: **Participant Characteristics** (n = 16)

Characteristic	**hc**, (n = 8)	**pp**, (n = 8)
Age in years		
Mean (SD)	33.8 (11.34)	55.9 (18.53)
Range	19 - 51	18 - 76
Gender		
Female	4 (57)	3 (43)
Male	2 (29)	3 (43)
Non-binary	1 (14)	1 (14)
Missing	1	1
Education		
High school	1 (12)	1 (12)
Higher degree	3 (38)	5 (62)
Tertiary	4 (50)	2 (25)
WCST score		
Mean (SD)	35.0 (15.46)	16.1 (11.06)
Range	13 - 56	9 - 43
Stroop score		
Mean (SD)	52.4 (31.55)	46.0 (17.27)
Range	24 - 120	22 - 67
Pain diagnosis		
CRPS	0 (0)	2 (25)
Endometriosis	0 (0)	2 (25)
Fibromyalgia	0 (0)	3 (38)
Migraine	0 (0)	1 (12)

Other useful packages:

- ggstatsplot (presents linear models as visual plots)
- jtools (similar to ggstatsplot but overlays two models onto a single visual plot)
- report (automated reporting of results from statistical tests)

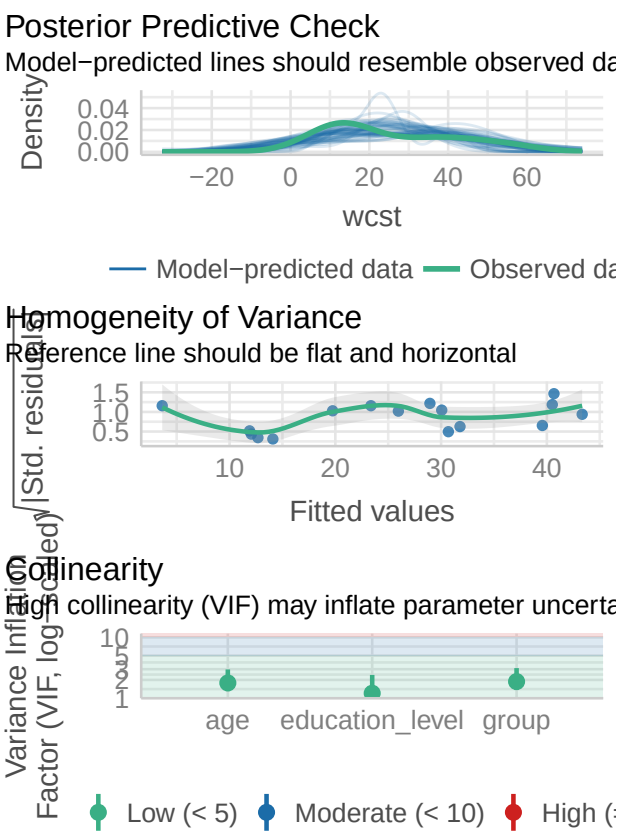
```
##
## Call:
## lm(formula = wcst ~ group + age + education_level, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.671  -5.615  -1.573   9.837  15.486
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    27.7619    13.1014   2.119   0.0577 .
## grouppp                       -12.2002     8.5361  -1.429   0.1807
## age                            -0.1570     0.2294  -0.684   0.5080
## education_levelHigher degree    7.0070    10.2511   0.684   0.5084
## education_levelTertiary        19.8156    10.2548   1.932   0.0795 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

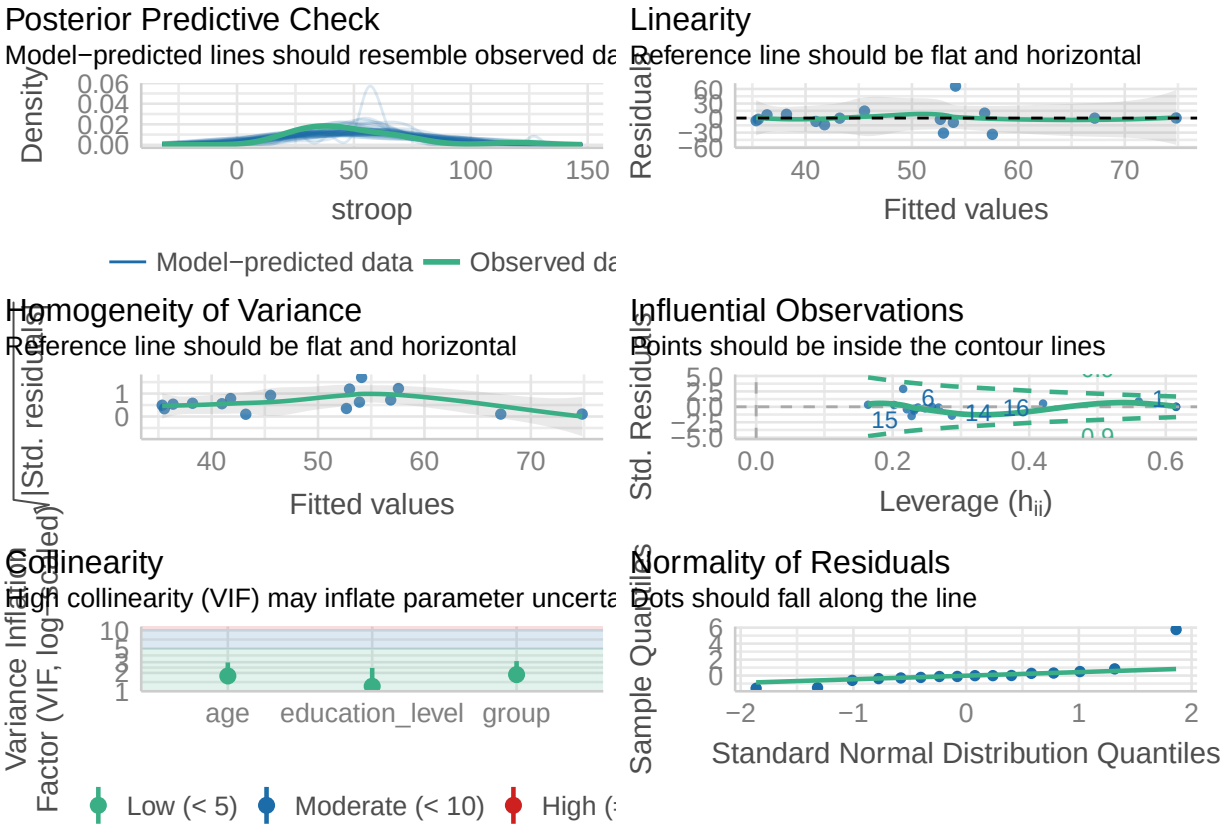
```
## Residual standard error: 12.46 on 11 degrees of freedom
## Multiple R-squared:  0.5684, Adjusted R-squared:  0.4114
## F-statistic: 3.621 on 4 and 11 DF,  p-value: 0.0406
```

```
##
## Call:
## lm(formula = stroop ~ group + age + education_level, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.556  -7.450  -1.385   6.918  65.906
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    80.5277    26.9205   2.991   0.0123 *
## grouppp                         2.1247    17.5398   0.121   0.9058
## age                            -0.2037     0.4714  -0.432   0.6741
## education_levelHigher degree -33.4480    21.0637  -1.588   0.1406
## education_levelTertiary      -17.4727    21.0713  -0.829   0.4246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.59 on 11 degrees of freedom
## Multiple R-squared:  0.2185, Adjusted R-squared:  -0.06563
## F-statistic: 0.769 on 4 and 11 DF,  p-value: 0.5673
```

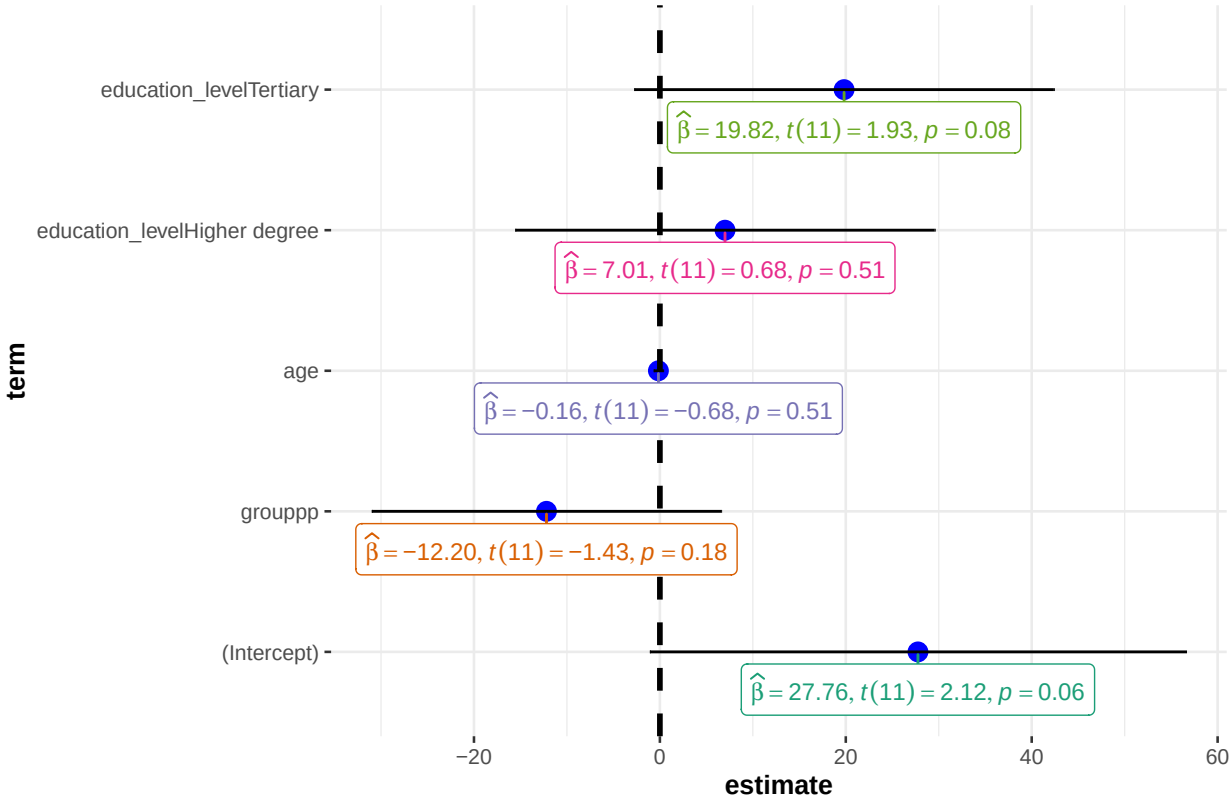
```
## Anova Table (Type II tests)
##
## Response: wcst
##                  Sum Sq Df F value Pr(>F)
## group            316.92  1  2.0428 0.1807
## age               72.62  1  0.4681 0.5080
## education_level  785.86  2  2.5327 0.1245
## Residuals       1706.58 11
```

```
## Anova Table (Type II tests)
##
## Response: stroop
##                 Sum Sq Df F value Pr(>F)
## group              9.6  1  0.0147 0.9058
## age              122.3  1  0.1867 0.6741
## education_level 1845.8  2  1.4089 0.2853
## Residuals       7205.4 11
```

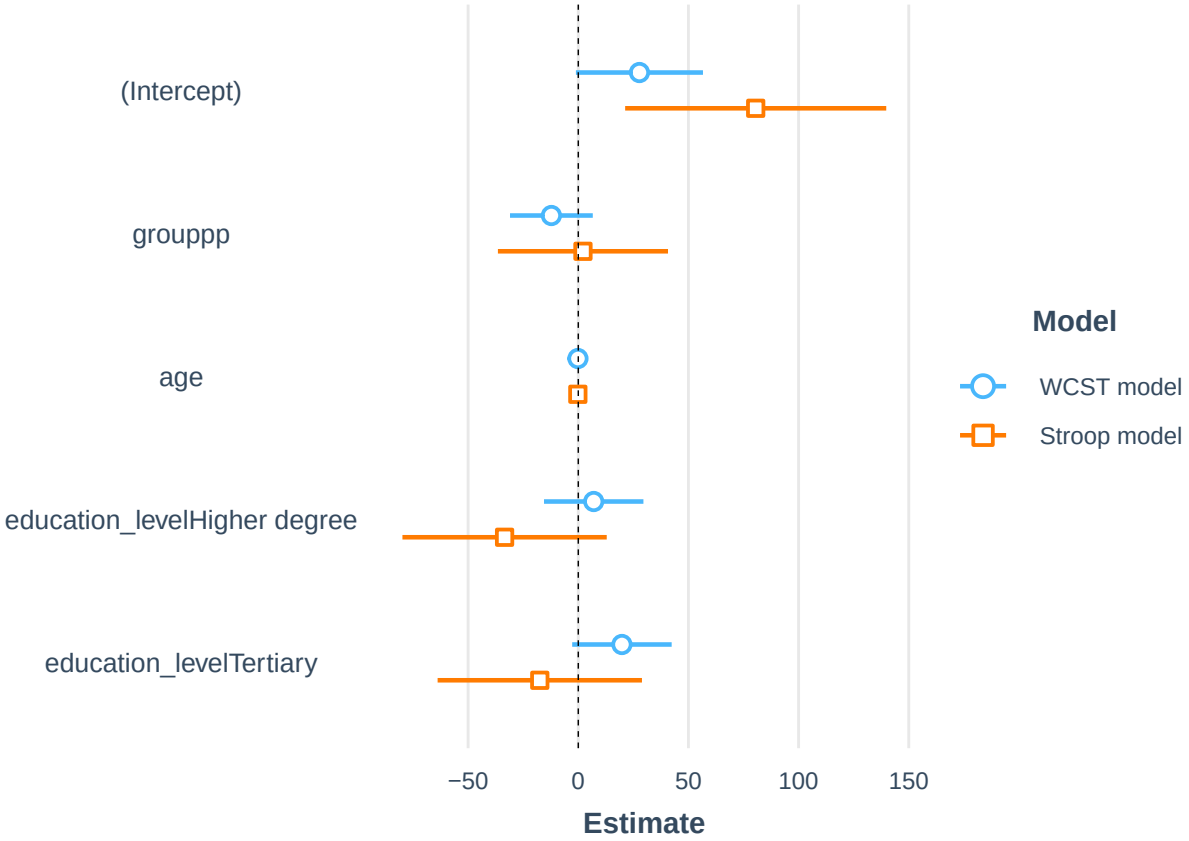

Posterior Predictive Check
Model−predicted lines should resemble observed da
Density
0.04
0.02
0.00
−20
0
20
40
60
wcst
Model−predicted data
Observed da
Linearity
Reference line should be flat and horizontal
Residuals
20
0
−20
10
20
30
40
Fitted values
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
1.5
1.0
0.5
10
20
30
40
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
5.0
2.5
0.0
−2.5
−5.0
0.0
0.2
0.4
0.6
Leverage (hii)
16
10
6
1
5
Collinearity
High collinearity (VIF) may inflate parameter uncerta
Variance Inflation Factor (VIF, log-scaled)
10
5
3
2
1
age
education_level
group
Low (< 5)
Moderate (< 10)
High (
Normality of Residuals
Dots should fall along the line
Sample Quantiles
2
0
−2
−2
−1
0
1
2
Standard Normal Distribution Quantiles



Characteristic	Beta	95% CI	p-value
group			0.2
hc			
pp	-14	-36, 7.4	
age	0.00	-0.62, 0.63	>0.9
education_level			0.2
High school			
Higher degree	12	-13, 37	
Tertiary	22	-3.1, 47	
gender			0.3
Male			
Female	-3.7	-21, 14	
Non-binary	-19	-47, 8.3	



AIC = 132, BIC = 137



$$\widehat{\text{wcst}} = 27.76 - 12.2(\text{group}_{\text{pp}}) - 0.16(\text{age}) + 7.01(\text{education_level}_{\text{Higher degree}}) + 19.82(\text{education_level}_{\text{Tertiary}}) \quad (1)$$

$$\widehat{\text{stroop}} = 80.53 + 2.12(\text{group}_{\text{pp}}) - 0.2(\text{age}) - 33.45(\text{education_level}_{\text{Higher degree}}) - 17.47(\text{education_level}_{\text{Tertiary}}) \quad (2)$$

```
## We fitted a linear model (estimated using OLS) to predict wcst with group
## (formula: wcst ~ group + age + education_level). The model explains a
## statistically significant and substantial proportion of variance (R2 = 0.57,
## F(4, 11) = 3.62, p = 0.041, adj. R2 = 0.41). The model's intercept,
## corresponding to group = hc, is at 27.76 (95% CI [-1.07, 56.60], t(11) = 2.12,
## p = 0.058). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and negative (beta
## = -12.20, 95% CI [-30.99, 6.59], t(11) = -1.43, p = 0.181; Std. beta = -0.75,
## 95% CI [-1.91, 0.41])
##   - The effect of age is statistically non-significant and negative (beta =
## -0.16, 95% CI [-0.66, 0.35], t(11) = -0.68, p = 0.508; Std. beta = -0.18, 95%
## CI [-0.76, 0.40])
##   - The effect of education level [Higher degree] is statistically
## non-significant and positive (beta = 7.01, 95% CI [-15.56, 29.57], t(11) =
## 0.68, p = 0.508; Std. beta = 0.43, 95% CI [-0.96, 1.82])
```

```
##   - The effect of education level [Tertiary] is statistically non-significant and
## positive (beta = 19.82, 95% CI [-2.75, 42.39], t(11) = 1.93, p = 0.079; Std.
## beta = 1.22, 95% CI [-0.17, 2.61])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation., We fitted a linear model
## (estimated using OLS) to predict wcst with age (formula: wcst ~ group + age +
## education_level). The model explains a statistically significant and
## substantial proportion of variance (R2 = 0.57, F(4, 11) = 3.62, p = 0.041, adj.
## R2 = 0.41). The model's intercept, corresponding to age = 0, is at 27.76 (95%
## CI [-1.07, 56.60], t(11) = 2.12, p = 0.058). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and negative (beta
## = -12.20, 95% CI [-30.99, 6.59], t(11) = -1.43, p = 0.181; Std. beta = -0.75,
## 95% CI [-1.91, 0.41])
##   - The effect of age is statistically non-significant and negative (beta =
## -0.16, 95% CI [-0.66, 0.35], t(11) = -0.68, p = 0.508; Std. beta = -0.18, 95%
## CI [-0.76, 0.40])
##   - The effect of education level [Higher degree] is statistically
## non-significant and positive (beta = 7.01, 95% CI [-15.56, 29.57], t(11) =
## 0.68, p = 0.508; Std. beta = 0.43, 95% CI [-0.96, 1.82])
##   - The effect of education level [Tertiary] is statistically non-significant and
## positive (beta = 19.82, 95% CI [-2.75, 42.39], t(11) = 1.93, p = 0.079; Std.
## beta = 1.22, 95% CI [-0.17, 2.61])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation. and We fitted a linear
## model (estimated using OLS) to predict wcst with education_level (formula: wcst
## ~ group + age + education_level). The model explains a statistically
## significant and substantial proportion of variance (R2 = 0.57, F(4, 11) = 3.62,
## p = 0.041, adj. R2 = 0.41). The model's intercept, corresponding to
## education_level = High school, is at 27.76 (95% CI [-1.07, 56.60], t(11) =
## 2.12, p = 0.058). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and negative (beta
## = -12.20, 95% CI [-30.99, 6.59], t(11) = -1.43, p = 0.181; Std. beta = -0.75,
## 95% CI [-1.91, 0.41])
##   - The effect of age is statistically non-significant and negative (beta =
## -0.16, 95% CI [-0.66, 0.35], t(11) = -0.68, p = 0.508; Std. beta = -0.18, 95%
## CI [-0.76, 0.40])
##   - The effect of education level [Higher degree] is statistically
## non-significant and positive (beta = 7.01, 95% CI [-15.56, 29.57], t(11) =
## 0.68, p = 0.508; Std. beta = 0.43, 95% CI [-0.96, 1.82])
##   - The effect of education level [Tertiary] is statistically non-significant and
## positive (beta = 19.82, 95% CI [-2.75, 42.39], t(11) = 1.93, p = 0.079; Std.
## beta = 1.22, 95% CI [-0.17, 2.61])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.

## We fitted a linear model (estimated using OLS) to predict stroop with group
```

```
## (formula: stroop ~ group + age + education_level). The model explains a
## statistically not significant and moderate proportion of variance (R2 = 0.22,
## F(4, 11) = 0.77, p = 0.567, adj. R2 = -0.07). The model's intercept,
## corresponding to group = hc, is at 80.53 (95% CI [21.28, 139.78], t(11) = 2.99,
## p = 0.012). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and positive (beta
## = 2.12, 95% CI [-36.48, 40.73], t(11) = 0.12, p = 0.906; Std. beta = 0.09, 95%
## CI [-1.47, 1.64])
##   - The effect of age is statistically non-significant and negative (beta =
## -0.20, 95% CI [-1.24, 0.83], t(11) = -0.43, p = 0.674; Std. beta = -0.15, 95%
## CI [-0.94, 0.63])
##   - The effect of education level [Higher degree] is statistically
## non-significant and negative (beta = -33.45, 95% CI [-79.81, 12.91], t(11) =
## -1.59, p = 0.141; Std. beta = -1.35, 95% CI [-3.22, 0.52])
##   - The effect of education level [Tertiary] is statistically non-significant and
## negative (beta = -17.47, 95% CI [-63.85, 28.90], t(11) = -0.83, p = 0.425; Std.
## beta = -0.70, 95% CI [-2.58, 1.17])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation., We fitted a linear model
## (estimated using OLS) to predict stroop with age (formula: stroop ~ group + age
## + education_level). The model explains a statistically not significant and
## moderate proportion of variance (R2 = 0.22, F(4, 11) = 0.77, p = 0.567, adj. R2
## = -0.07). The model's intercept, corresponding to age = 0, is at 80.53 (95% CI
## [21.28, 139.78], t(11) = 2.99, p = 0.012). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and positive (beta
## = 2.12, 95% CI [-36.48, 40.73], t(11) = 0.12, p = 0.906; Std. beta = 0.09, 95%
## CI [-1.47, 1.64])
##   - The effect of age is statistically non-significant and negative (beta =
## -0.20, 95% CI [-1.24, 0.83], t(11) = -0.43, p = 0.674; Std. beta = -0.15, 95%
## CI [-0.94, 0.63])
##   - The effect of education level [Higher degree] is statistically
## non-significant and negative (beta = -33.45, 95% CI [-79.81, 12.91], t(11) =
## -1.59, p = 0.141; Std. beta = -1.35, 95% CI [-3.22, 0.52])
##   - The effect of education level [Tertiary] is statistically non-significant and
## negative (beta = -17.47, 95% CI [-63.85, 28.90], t(11) = -0.83, p = 0.425; Std.
## beta = -0.70, 95% CI [-2.58, 1.17])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation. and We fitted a linear
## model (estimated using OLS) to predict stroop with education_level (formula:
## stroop ~ group + age + education_level). The model explains a statistically not
## significant and moderate proportion of variance (R2 = 0.22, F(4, 11) = 0.77, p
## = 0.567, adj. R2 = -0.07). The model's intercept, corresponding to
## education_level = High school, is at 80.53 (95% CI [21.28, 139.78], t(11) =
## 2.99, p = 0.012). Within this model:
##
##   - The effect of group [pp] is statistically non-significant and positive (beta
## = 2.12, 95% CI [-36.48, 40.73], t(11) = 0.12, p = 0.906; Std. beta = 0.09, 95%
## CI [-1.47, 1.64])
```

```
##   - The effect of age is statistically non-significant and negative (beta =
## -0.20, 95% CI [-1.24, 0.83], t(11) = -0.43, p = 0.674; Std. beta = -0.15, 95%
## CI [-0.94, 0.63])
##   - The effect of education level [Higher degree] is statistically
## non-significant and negative (beta = -33.45, 95% CI [-79.81, 12.91], t(11) =
## -1.59, p = 0.141; Std. beta = -1.35, 95% CI [-3.22, 0.52])
##   - The effect of education level [Tertiary] is statistically non-significant and
## negative (beta = -17.47, 95% CI [-63.85, 28.90], t(11) = -0.83, p = 0.425; Std.
## beta = -0.70, 95% CI [-2.58, 1.17])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.
```

Useful resources for gtsummary:

- gtsummary paper: https://journal.r-project.org/archive/2021/RJ-2021-053/RJ-2021-053.pdf
- gtsummary cheatsheet: https://www.rstudio.com/resources/cheatsheets/
- R Ladies STL recently held a workshop with the package developer of gtsummary (Daniel Sjoberg) - the workshop will be posted on the R Ladies STL Youtube channel in a few days time: https://www.youtube.com/channel/UCQ7DcXWJm3es06U7lGmBDkw

Citing R packages:

citation("tidyverse")

citation("gtsummary")

citation("report")

citation("equatiomatic")

citation("jtools")

citation("ggstatsplot")