

NMST432
**Computational Environment for Statistical Data
Analysis**
Sample Report prepared using Sweave

Arnošt Komárek
May 5, 2021

This document was prepared using Sweave (Leisch, 2002) in R (R Core Team, 2021), version 4.0.5 (2021-03-31). Additionally, the contributed packages colorspace (Zeileis et al., 2019, 2009) and xtable (Dahl et al., 2019) were used.

1 Some Sweave examples

- Here we define our working directory.

```
> ROOT <- "/home/komarek/teach/mff_2020/nmst440_VypocProstr/Tutorial08/"  
> setwd(ROOT)
```

- Now, we load needed extension packages and provide our smaller functions.

```
> library("colorspace")  
> library("xtable")  
> source(paste(ROOT, "../Tutorial04/formatOut.R", sep = ""))  
> source(paste(ROOT, "../Tutorial04/funTabDescr.R", sep = ""))
```

- Read data (the same as those used the previous time, now directly including some derived variables):

```
> print(load(paste(ROOT, "../Tutorial04/Data/nelsNE2.RData", sep = "")))
```

```
[1] "varlabels2" "nelsNE2"
```

- Basic descriptive statistics of some variables:

```
> VARS <- c("fam.comp", "gender", "f2.sco.math", "f2.perc.math")
> summary(nelsNE2[, VARS])
```

	fam.comp	gender	f2.sco.math	f2.perc.math
Mother and father:	1601	Male :1140	Min. :30.17	Min. : 1.00
Other	: 508	Female:1172	1st Qu.:46.97	1st Qu.:40.00
NA's	: 203		Median :54.61	Median :65.00
			Mean :53.86	Mean :60.97
			3rd Qu.:61.76	3rd Qu.:85.00
			Max. :71.49	Max. :99.00
			NA's :1	NA's :1

- Here, descriptive statistics are calculated but not shown:

```
> sumnelsNE<- summary(nelsNE2[, VARS])
```

- Here, descriptive statistics are calculated, results shown but the code is not shown:

	fam.comp	gender	f2.sco.math	f2.perc.math
Mother and father:	1601	Male :1140	Min. :30.17	Min. : 1.00
Other	: 508	Female:1172	1st Qu.:46.97	1st Qu.:40.00
NA's	: 203		Median :54.61	Median :65.00
			Mean :53.86	Mean :60.97
			3rd Qu.:61.76	3rd Qu.:85.00
			Max. :71.49	Max. :99.00
			NA's :1	NA's :1

- Here, descriptive statistics are calculated but neither results nor the code are shown:

- Here, only code is shown but nothing calculated:

```
> summary(nelsNE2[, VARS])
```

- It is also possible to use a calculated number (calculated numbers) in the body of the text:

```
> meanScoMath <- mean(nelsNE2[, "f2.sco.math"], na.rm = TRUE)
> meanScoMath <- format(round(meanScoMath, 2), nsmall = 2)
> print(meanScoMath)
```

```
[1] "53.86"
```

Mean score in mathematics is 53.86 ($N = 2311$).

- If long code is shown, we may arrange that it is automatically formatted to fit on the page:

```
> meanScoMath <- format(round(mean(nelsNE2[, "f2.sco.math"], na.rm = TRUE),  
+ 2), nsmall = 2)
```

- Or we may take care ourselves for format of the code:

```
> meanScoMath <- format(round(mean(nelsNE2[, "f2.sco.math"],  
+ na.rm = TRUE), 2), nsmall = 2)
```

2 Tables

Results are seen in Table 1. Slightly extended results (by results of a t-test) are shown in Table 2.

	Mean	Std. Dev.	Std. Error	Median	Q1	Q3	N
All	54.05	9.72	0.21	54.87	47.35	61.86	2108
Mother and father	54.89	9.57	0.24	55.97	48.37	62.74	1600
Other	51.41	9.73	0.43	52.45	43.70	58.89	508

Table 1: Descriptive statistics of **score in mathematics** by **family composition**.

Table 2: Descriptive statistics of **score in mathematics** by **family composition**.

Score in mathematics by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	54.05 (0.21)	9.72	54.87	47.35 – 61.86	2108
Mother and father	54.89 (0.24)	9.57	55.97	48.37 – 62.74	1600
Other	51.41 (0.43)	9.73	52.45	43.70 – 58.89	508
Difference in means: 3.48 (2.51, 4.45)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

3 Figures

- Define what should be conducted before each plotting.

```
> figSweave <- function(){  
+   par(bty = "n", mar = c(5, 4, 4, 1) + 0.1)  
+   ## WHATEVER OTHER R COMMANDS  
+ }  
> options(SweaveHooks = list(fig = figSweave))
```

- Figure which is drawn, saved as PDF and automatically placed in a text (see Figure 1). Note that pdfL^AT_EX must then be used to process the T_EX file.
- Figure which was drawn, saved as PDF but it is nowhere placed automatically. Placing the figure into the document (see Figure 2) is the author's responsibility.

```
> COL2 <- terrain_hcl(2)  
> plot(f2.sco.math ~ fam.comp, data = nelsNE2, col = COL2,  
+       xlab = "Family composition", ylab = "Score in mathematics")
```

```
> COL <- rainbow_hcl(2, start = 90)
> plot(f2.sco.math ~ fam.comp, data = nelsNE2, col = COL,
+      xlab = "Family composition", ylab = "Score in mathematics")
```

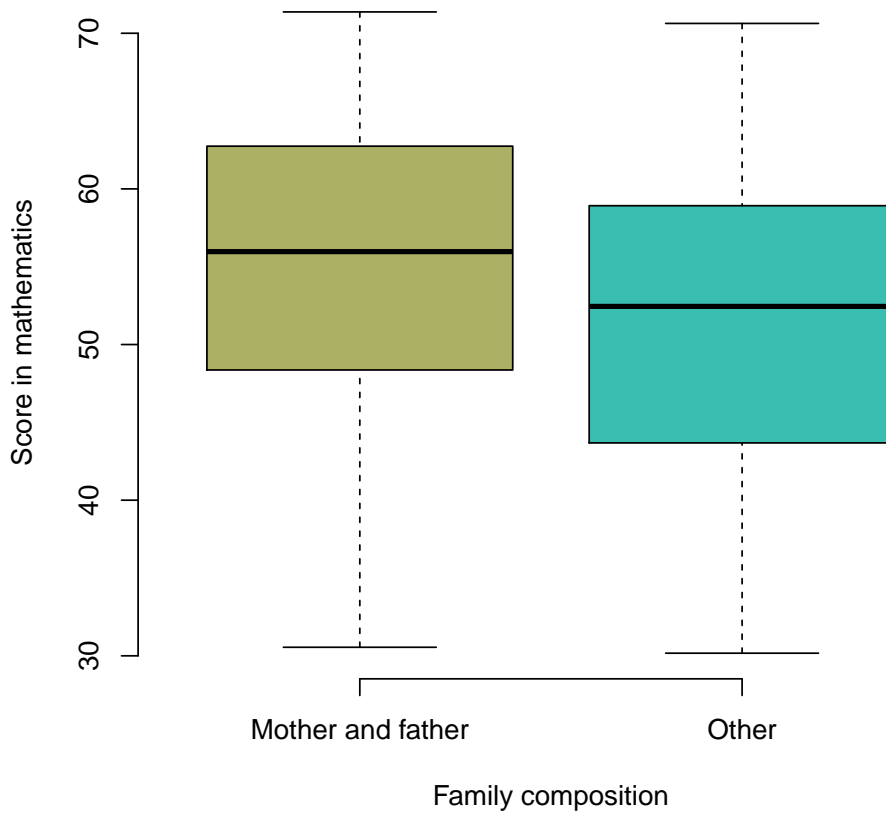


Figure 1: Score in mathematics by family composition.

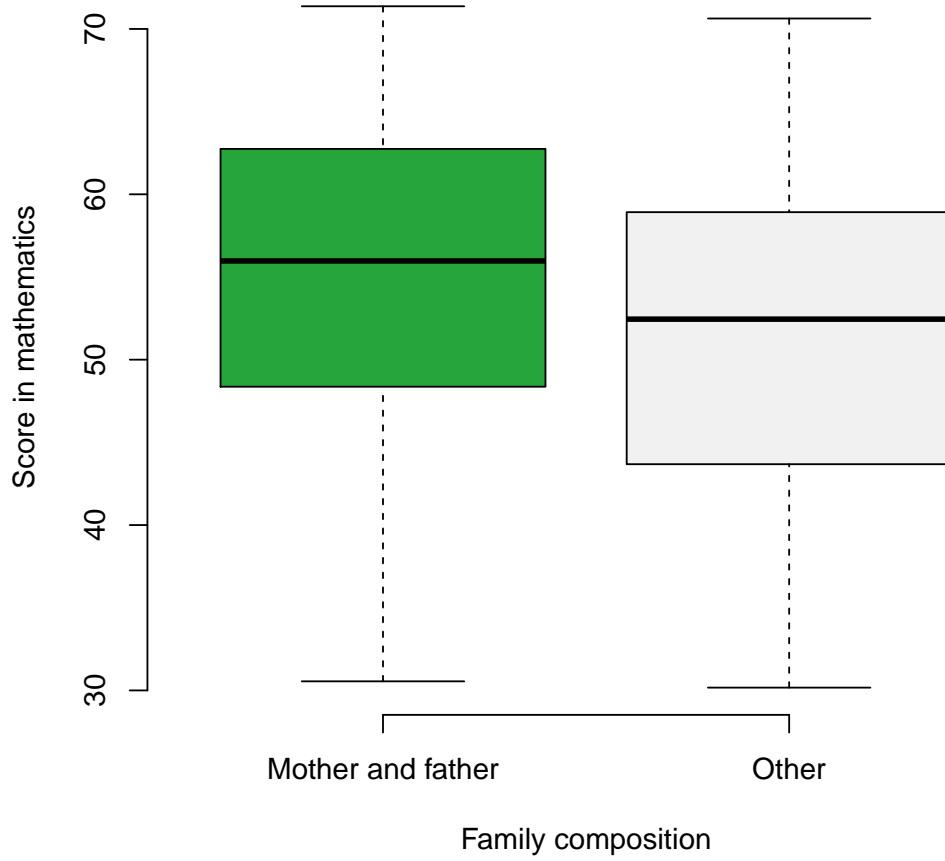


Figure 2: Score in mathematics by family composition (again).

- It is also possible to use standard functions `pdf()`, `postscript()`, `png()` etc. to save a plot in an arbitrary format on an arbitrary place with an arbitrary filename:

```
> postscript("./Figures/fig-boxplot1.eps", width = 6, height = 6,
+           horizontal = FALSE, paper = "special")
> plot(f2.sco.math ~ fam.comp, data = nelsNE2, col = COL,
+      xlab = "Family composition", ylab = "Score in mathematics")
> dev.off()
> #
> RES <- 500
> png("./Figures/fig-boxplot1.png", width = 6*RES, height = 6*RES, res = RES)
> plot(f2.sco.math ~ fam.comp, data = nelsNE2, col = COL,
+      xlab = "Family composition", ylab = "Score in mathematics")
> dev.off()
```


4 Results of a more extensive analysis

Results of analysis of dependence of score in mathematics on family composition is shown in Table 3 and on Figure 3. All results are then in Tables 4–19 and on Figures 4–19.

Table 3: Analysis of **score in mathematics** by **family composition**.

Math score by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	54.05 (0.21)	9.72	54.87	47.35 – 61.86	2108
Mother and father	54.89 (0.24)	9.57	55.97	48.37 – 62.74	1600
Other	51.41 (0.43)	9.73	52.45	43.70 – 58.89	508
Difference in means: 3.48 (2.51, 4.45)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

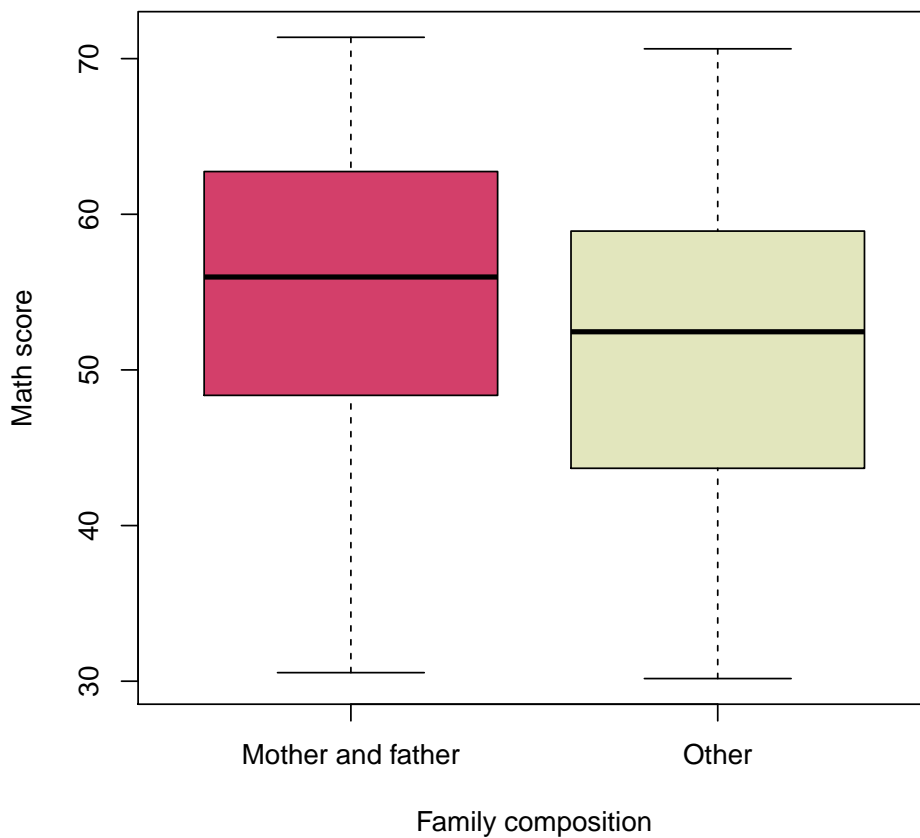


Figure 3: Score in mathematics by family composition (once again).

4.1 Math score by Family composition

Table 4: Analysis of **Math score** by **Family composition**.

Math score by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	54.05 (0.21)	9.72	54.87	47.35 – 61.86	2108
Mother and father	54.89 (0.24)	9.57	55.97	48.37 – 62.74	1600
Other	51.41 (0.43)	9.73	52.45	43.70 – 58.89	508
Difference in means: 3.48 (2.51, 4.45)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

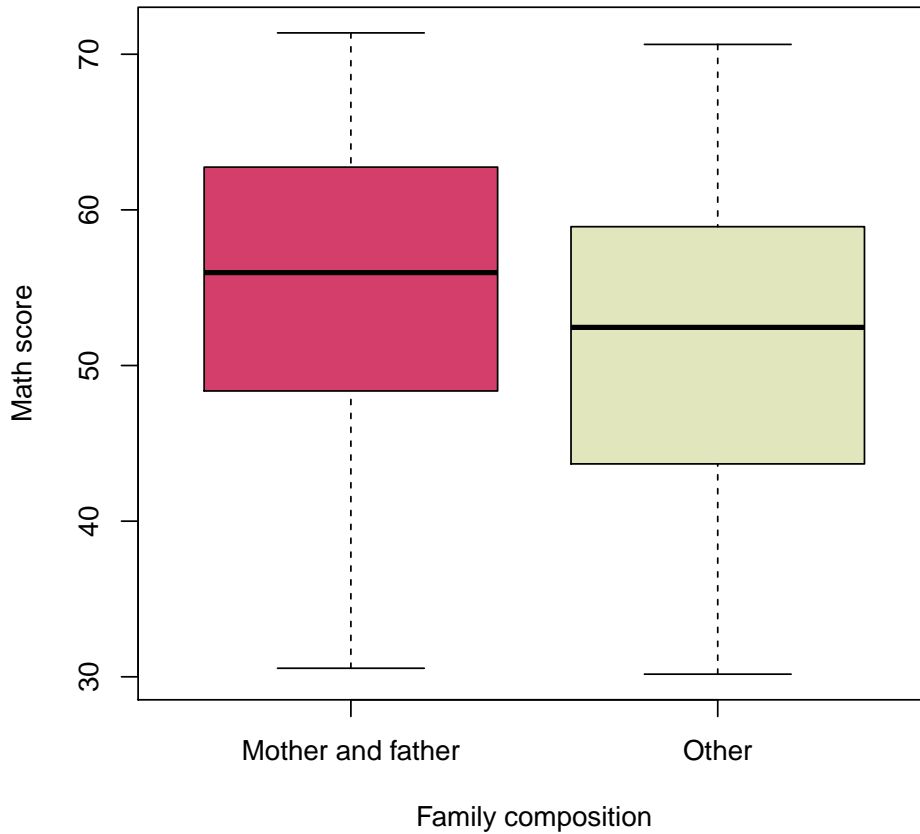


Figure 4: Boxplots of **Math score** by **Family composition**.

4.2 Math score by Gender

Table 5: Analysis of **Math score** by **Gender**.

Math score by Gender					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.86 (0.20)	9.79	54.61	46.97 – 61.76	2311
Male	54.25 (0.30)	9.97	55.37	47.03 – 62.31	1139
Female	53.47 (0.28)	9.60	54.11	46.91 – 61.19	1172
Difference in means: 0.78		(-0.01, 1.58) [†] ,		P: 0.054 [‡]	

[†] 95% confidence interval

[‡] Welch two-sample t-test

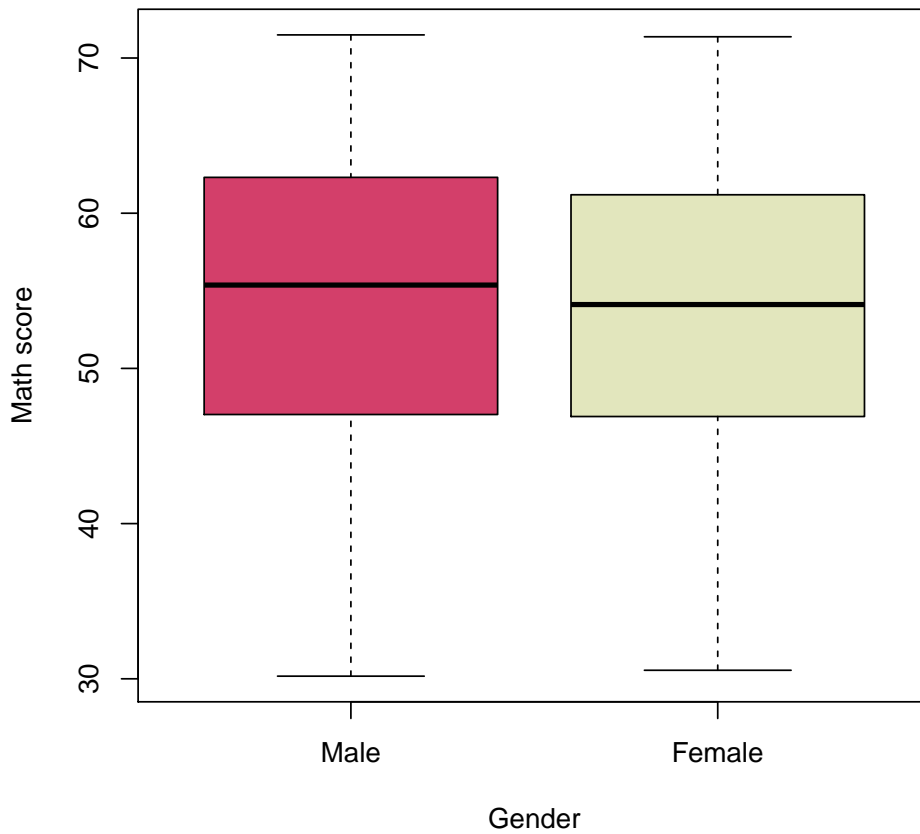


Figure 5: Boxplots of **Math score** by **Gender**.

4.3 Math score by Math enrollment past 2 years

Table 6: Analysis of **Math score** by **Math enrollment past 2 years**.

Math score by Math enrollment past 2 years					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	54.19 (0.20)	9.66	55.15	47.59 – 61.90	2249
Yes	54.72 (0.20)	9.42	55.80	48.36 – 62.20	2122
No	45.43 (0.83)	9.33	42.98	37.88 – 52.95	127
Difference in means: 9.29			(7.60, 10.97)[†] ,		P: <0.001[‡]

[†] 95% confidence interval

[‡] Welch two-sample t-test

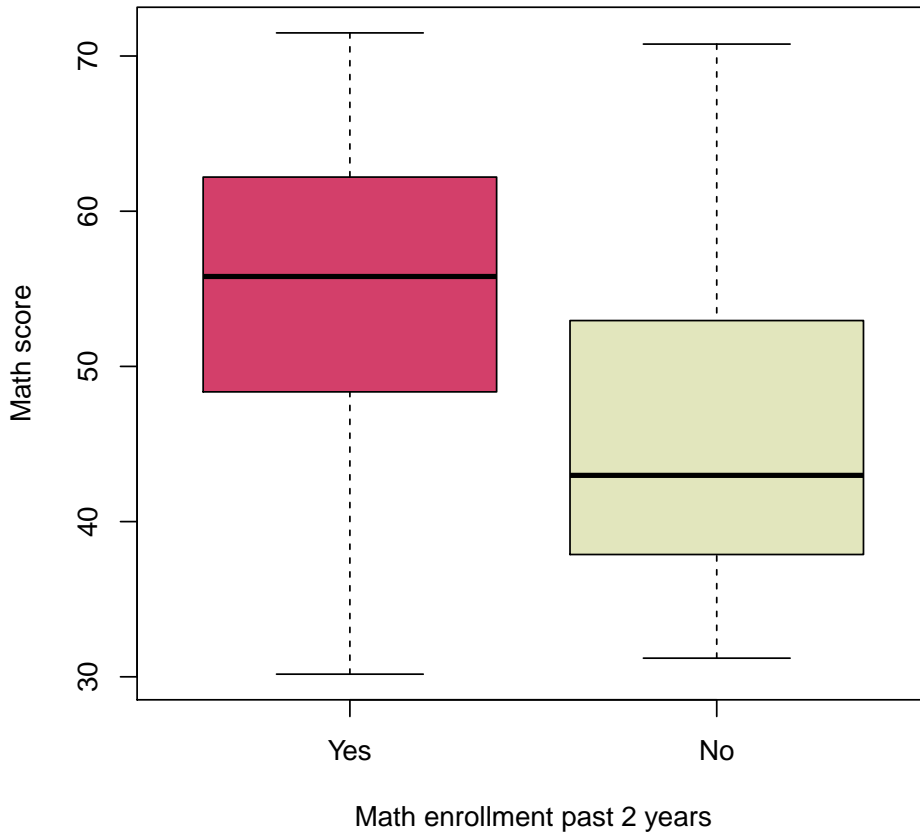


Figure 6: Boxplots of **Math score** by **Math enrollment past 2 years**.

4.4 Math score by Arrested

Table 7: Analysis of **Math score** by **Arrested**.

Math score by Arrested					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	54.19 (0.20)	9.67	55.16	47.56 – 61.92	2246
Never	54.35 (0.21)	9.62	55.31	47.80 – 62.08	2187
Ever	48.27 (1.29)	9.89	48.87	39.11 – 56.74	59
Difference in means: 6.09 (3.48, 8.69) [†] , P: <0.001 [‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

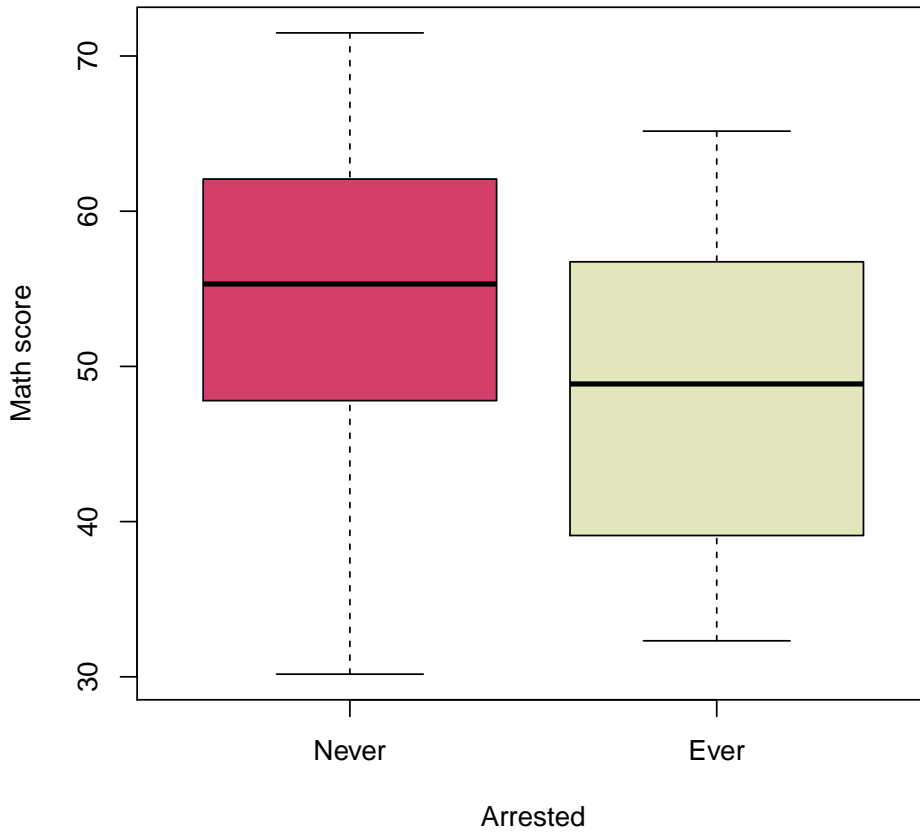


Figure 7: Boxplots of **Math score** by **Arrested**.

4.5 Science score by Family composition

Table 8: Analysis of **Science score** by **Family composition**.

Science score by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.88 (0.21)	9.63	54.89	46.59 – 61.90	2095
Mother and father	54.55 (0.24)	9.46	55.75	47.62 – 62.37	1593
Other	51.74 (0.44)	9.88	51.92	44.29 – 60.15	502
Difference in means: 2.81 (1.83, 3.79)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

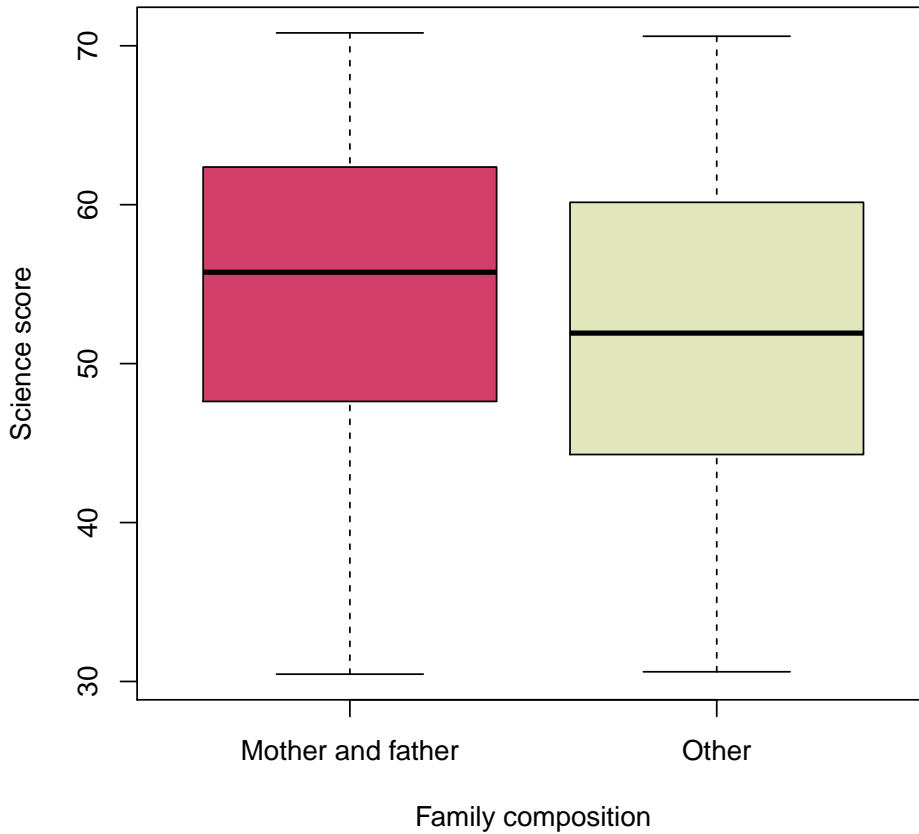


Figure 8: Boxplots of **Science score** by **Family composition**.

4.6 Science score by Gender

Table 9: Analysis of **Science score** by **Gender**.

Science score by Gender					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.51 (0.21)	9.82	54.53	46.00 – 61.74	2294
Male	54.82 (0.30)	9.93	56.22	47.72 – 63.36	1133
Female	52.23 (0.28)	9.54	53.15	44.86 – 59.95	1161
Difference in means: 2.59 (1.79, 3.39)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

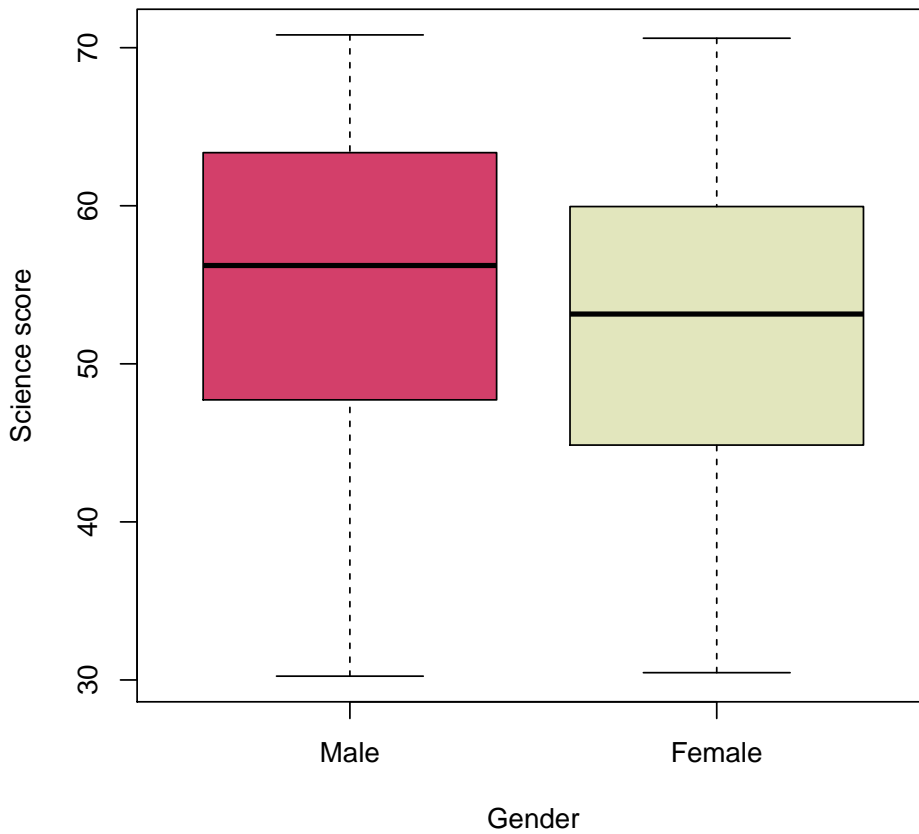


Figure 9: Boxplots of **Science score** by **Gender**.

4.7 Science score by Math enrollment past 2 years

Table 10: Analysis of **Science score** by **Math enrollment past 2 years**.

Science score by Math enrollment past 2 years					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.77 (0.21)	9.75	54.85	46.39 – 61.90	2233
Yes	54.14 (0.21)	9.66	55.39	47.02 – 62.17	2108
No	47.53 (0.81)	9.01	47.09	40.59 – 53.66	125
Difference in means: 6.61 (4.96, 8.26) [†] , P: <0.001 [‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

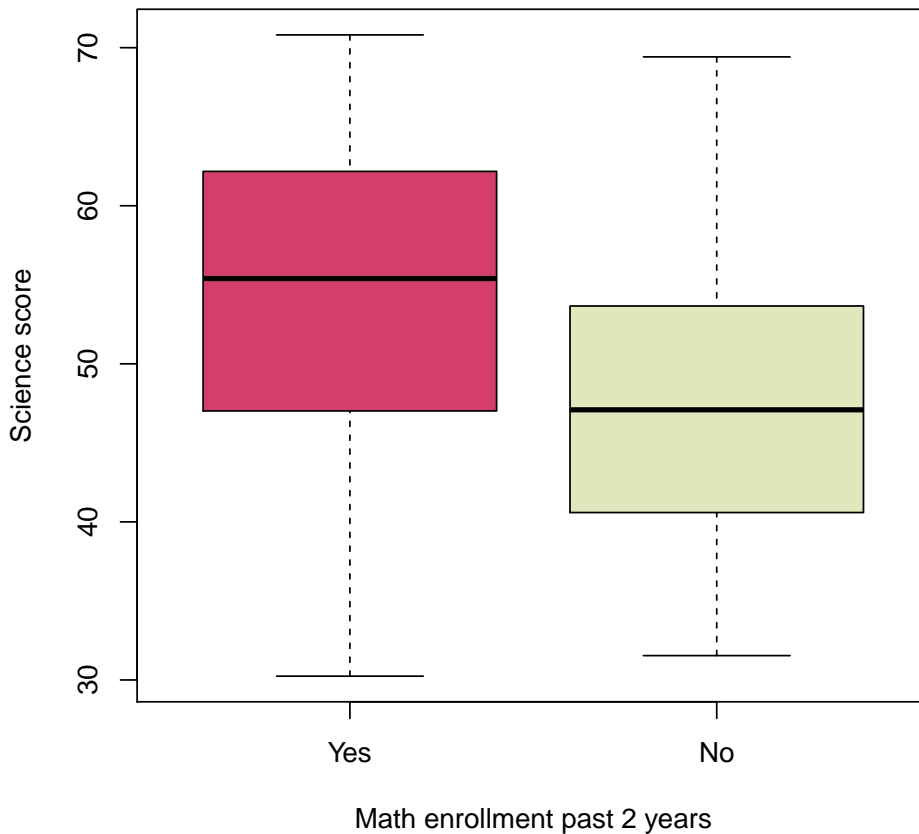


Figure 10: Boxplots of **Science score** by **Math enrollment past 2 years**.

4.8 Science score by Arrested

Table 11: Analysis of **Science score** by **Arrested**.

Science score by Arrested					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.78 (0.21)	9.75	54.87	46.40 – 61.90	2230
Never	53.93 (0.21)	9.64	54.95	46.72 – 61.91	2172
Ever	48.01 (1.57)	11.95	45.76	37.97 – 58.72	58
Difference in means: 5.92 (2.75, 9.09)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

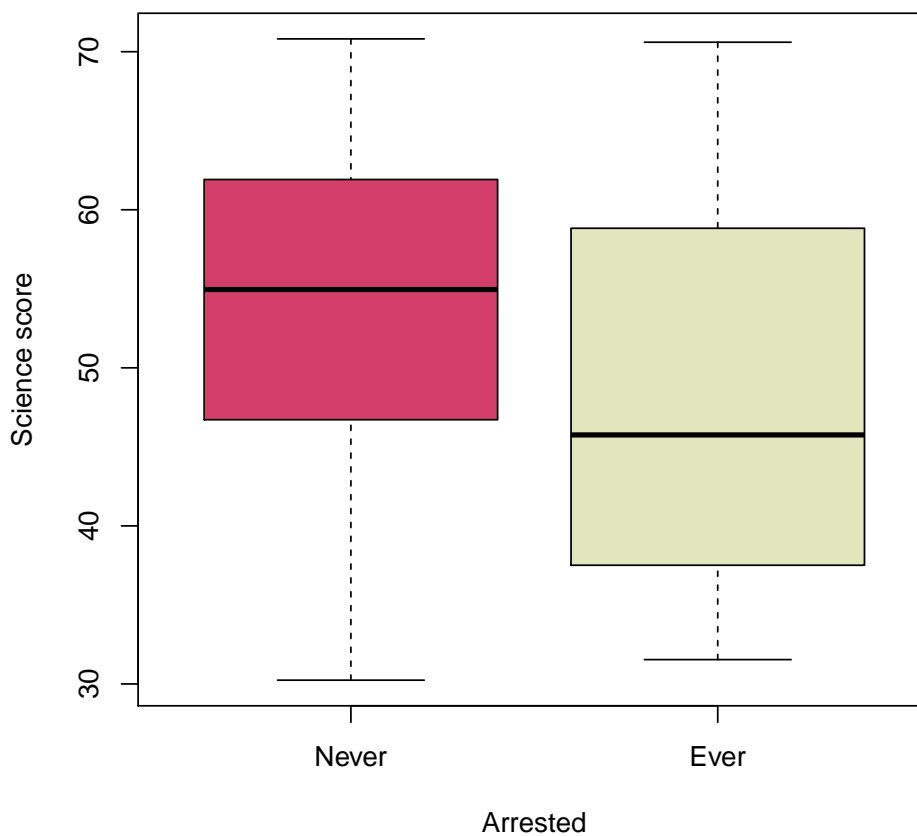


Figure 11: Boxplots of **Science score** by **Arrested**.

4.9 Social science score by Family composition

Table 12: Analysis of **Social science score** by **Family composition**.

Social science score by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.51 (0.21)	9.47	54.40	46.03 – 61.48	2081
Mother and father	54.25 (0.24)	9.36	55.28	46.91 – 61.95	1584
Other	51.15 (0.42)	9.42	51.33	44.31 – 58.73	497
Difference in means: 3.10 (2.15, 4.05)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

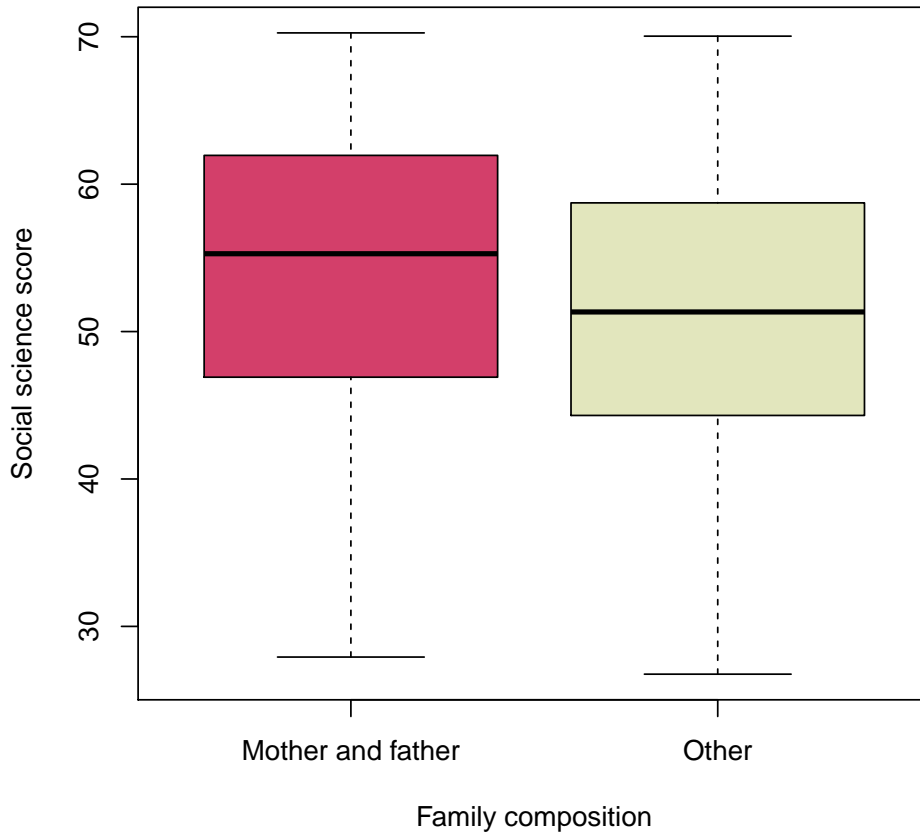


Figure 12: Boxplots of **Social science score** by **Family composition**.

4.10 Social science score by Gender

Table 13: Analysis of **Social science score** by **Gender**.

Social science score by Gender					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.24 (0.20)	9.52	53.85	45.74 – 61.30	2275
Male	53.99 (0.29)	9.78	55.15	46.19 – 62.11	1125
Female	52.50 (0.27)	9.19	52.52	45.30 – 60.36	1150
Difference in means: 1.49 (0.71, 2.28)[†] , P: <0.001[‡]					

[†]95% confidence interval

[‡]Welch two-sample t-test

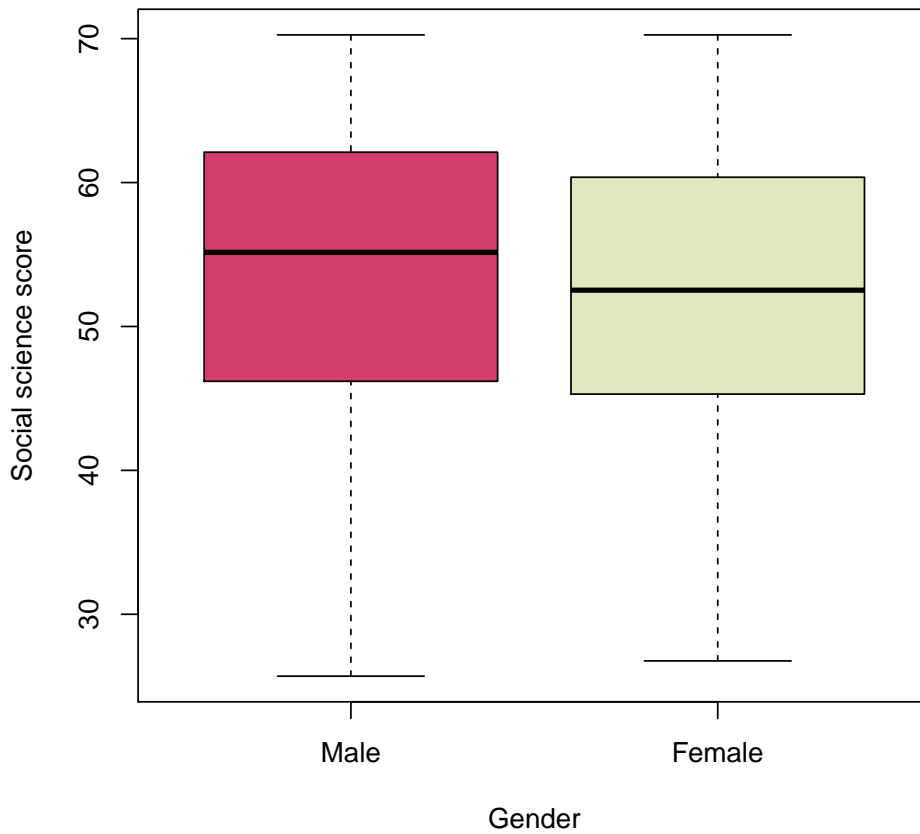


Figure 13: Boxplots of **Social science score** by **Gender**.

4.11 Social science score by Math enrollment past 2 years

Table 14: Analysis of **Social science score** by **Math enrollment past 2 years**.

Social science score by Math enrollment past 2 years					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.52 (0.20)	9.43	54.29	46.06 – 61.44	2216
Yes	53.84 (0.20)	9.34	54.83	46.44 – 61.60	2094
No	48.10 (0.85)	9.37	46.54	41.46 – 54.92	122
Difference in means: 5.74 (4.02, 7.47) [†] , P: <0.001 [‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

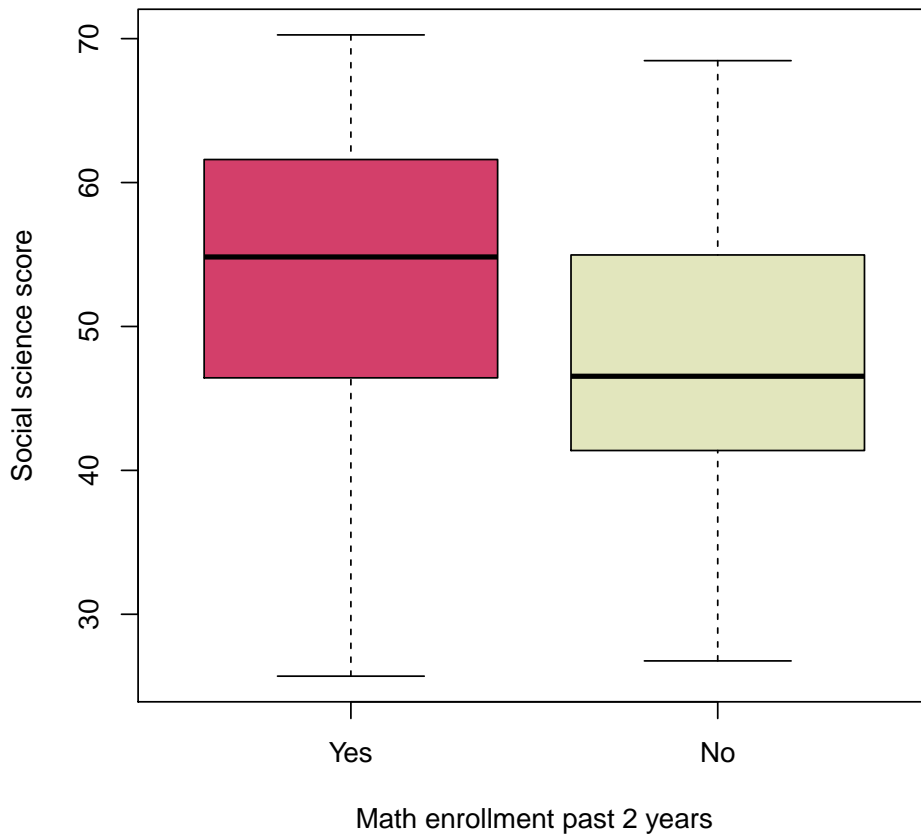


Figure 14: Boxplots of **Social science score** by **Math enrollment past 2 years**.

4.12 Social science score by Arrested

Table 15: Analysis of **Social science score** by **Arrested**.

Social science score by Arrested					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.52 (0.20)	9.44	54.27	46.06 – 61.45	2213
Never	53.68 (0.20)	9.39	54.53	46.20 – 61.55	2155
Ever	47.61 (1.24)	9.41	47.42	41.21 – 55.07	58
Difference in means: 6.07 (3.57, 8.58) [†] , P: <0.001 [‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

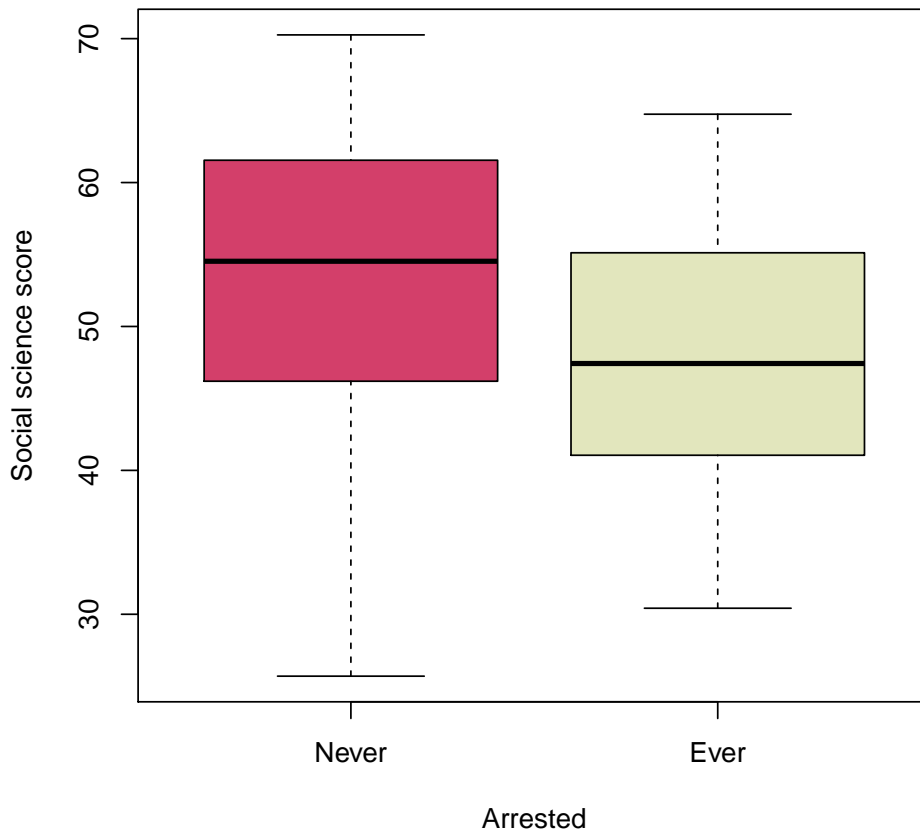


Figure 15: Boxplots of **Social science score** by **Arrested**.

4.13 Reading score by Family composition

Table 16: Analysis of **Reading score** by **Family composition**.

Reading score by Family composition					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.38 (0.21)	9.53	54.86	46.39 – 61.08	2107
Mother and father	53.85 (0.24)	9.56	55.54	47.03 – 61.82	1600
Other	51.92 (0.41)	9.30	53.40	45.16 – 59.16	507
Difference in means: 1.93 (0.99, 2.87)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

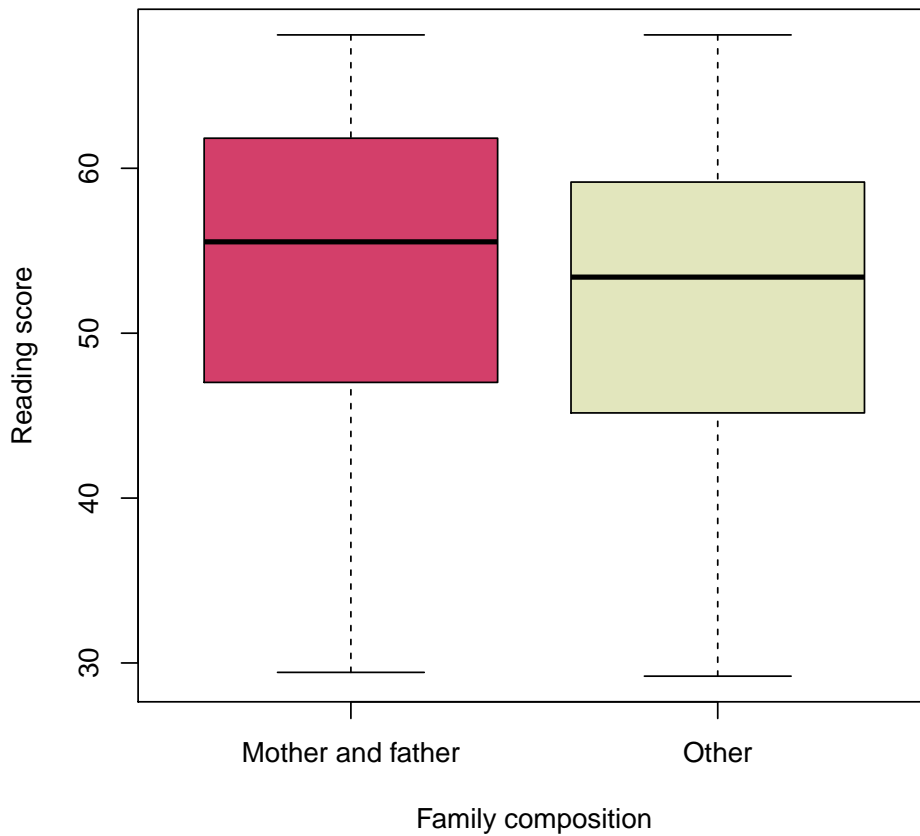


Figure 16: Boxplots of **Reading score** by **Family composition**.

4.14 Reading score by Gender

Table 17: Analysis of **Reading score** by **Gender**.

Reading score by Gender					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.06 (0.20)	9.69	54.47	45.80 – 60.98	2306
Male	51.77 (0.30)	10.12	53.11	43.76 – 60.23	1137
Female	54.31 (0.27)	9.08	55.81	48.16 – 61.76	1169
Difference in means:		-2.54	(-3.33, -1.76)[†] ,	P: <0.001[‡]	

[†] 95% confidence interval

[‡] Welch two-sample t-test

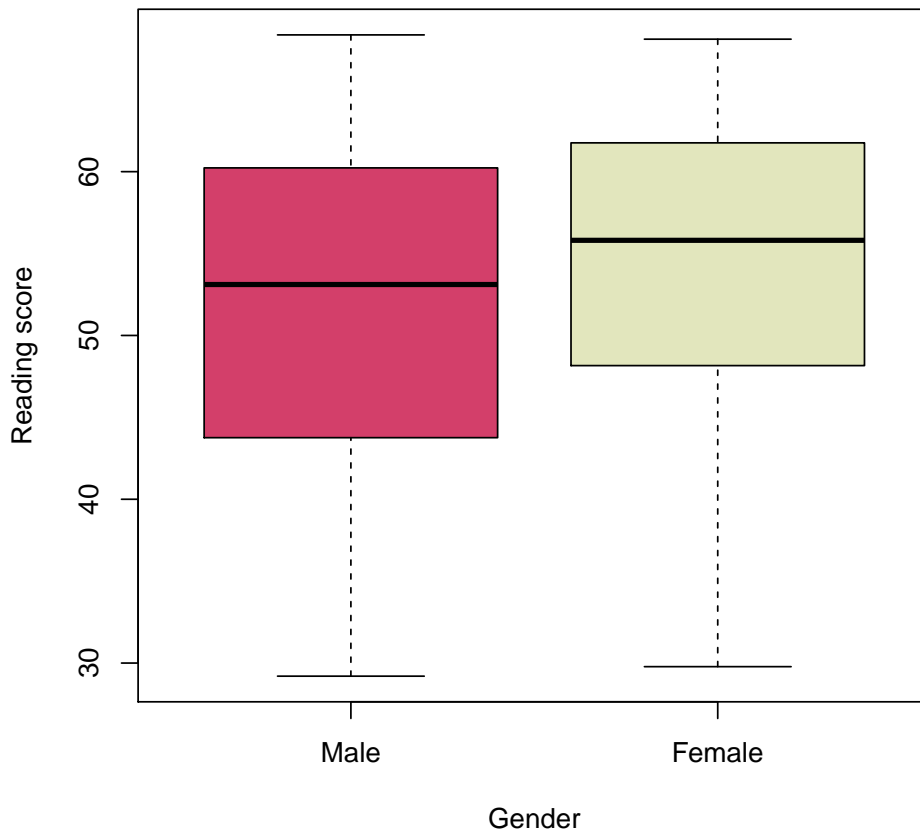


Figure 17: Boxplots of **Reading score** by **Gender**.

4.15 Reading score by Math enrollment past 2 years

Table 18: Analysis of **Reading score** by **Math enrollment past 2 years**.

Reading score by Math enrollment past 2 years					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.30 (0.20)	9.62	54.78	46.31 – 61.11	2244
Yes	53.68 (0.21)	9.49	55.22	46.79 – 61.41	2117
No	46.90 (0.84)	9.49	47.06	39.45 – 53.14	127
Difference in means: 6.79 (5.07, 8.50)[†] , P: <0.001[‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

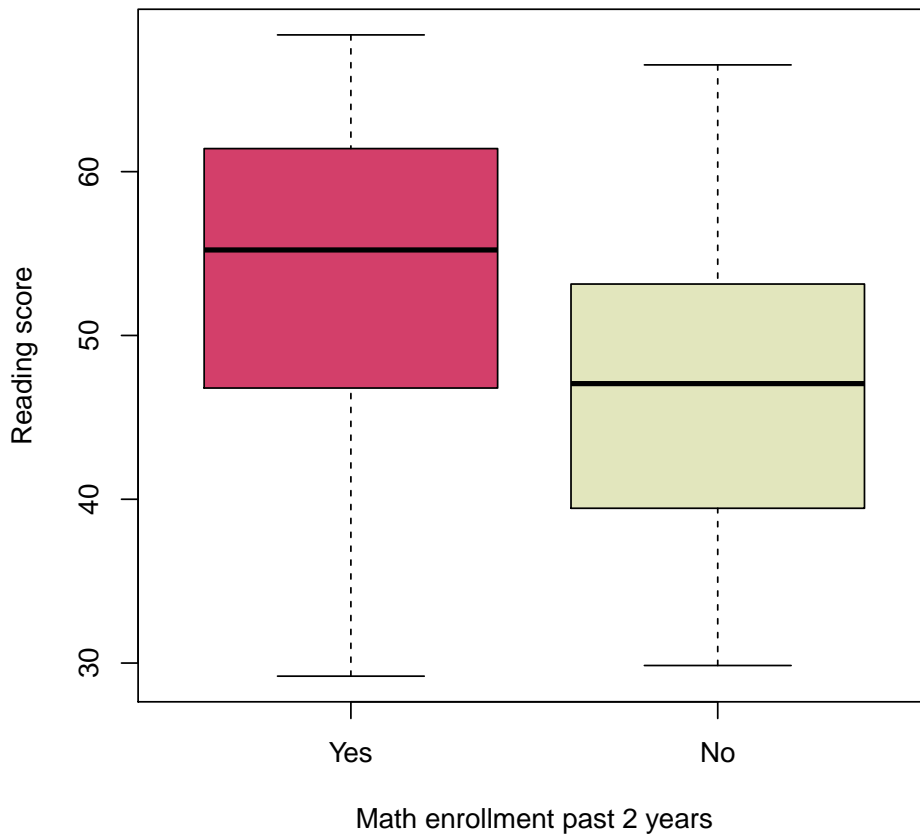


Figure 18: Boxplots of **Reading score** by **Math enrollment past 2 years**.

4.16 Reading score by Arrested

Table 19: Analysis of **Reading score** by **Arrested**.

Reading score by Arrested					
Group	Mean (S.E.)	Std. Dev.	Median	Q ₁ – Q ₃	N
All	53.31 (0.20)	9.61	54.81	46.32 – 61.14	2241
Never	53.50 (0.20)	9.53	54.94	46.46 – 61.22	2182
Ever	46.49 (1.31)	10.09	47.06	36.50 – 52.50	59
Difference in means: 7.01 (4.35, 9.66) [†] , P: <0.001 [‡]					

[†] 95% confidence interval

[‡] Welch two-sample t-test

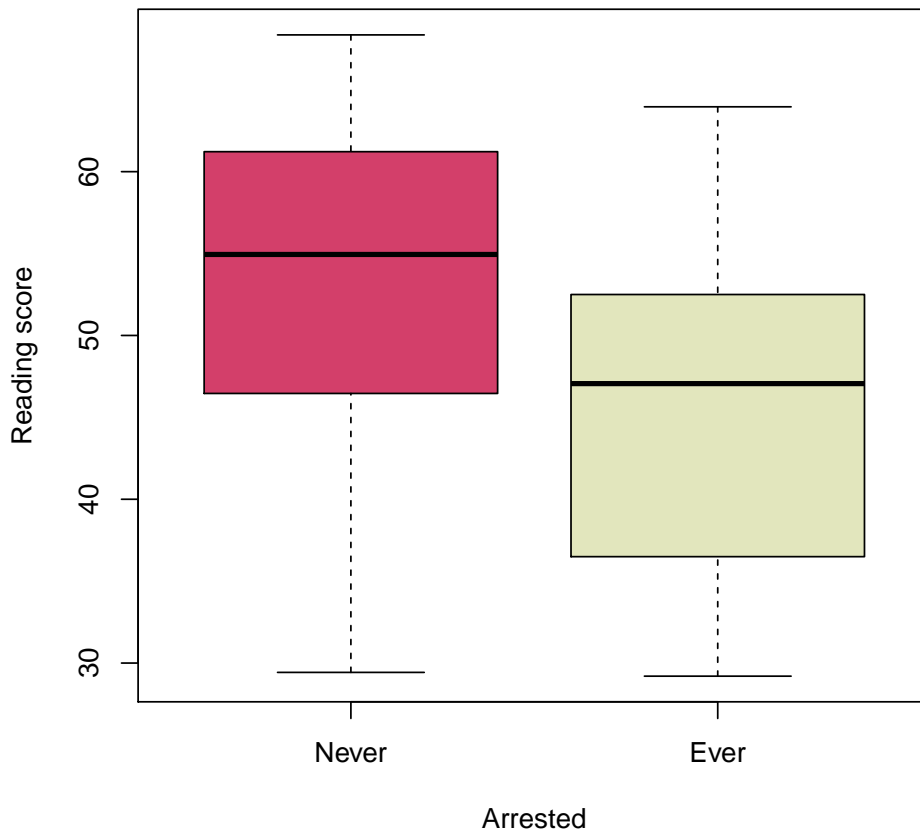


Figure 19: Boxplots of **Reading score** by **Arrested**.

References

- DAHL, D. B., SCOTT, D., ROOSEN, C., MAGNUSSON, A., and SWINTON, J. (2019). *xtable: Export tables to L^AT_EX or HTML*. URL <http://CRAN.R-project.org/package=xtable>. R package version 1.8-4.
- LEISCH, F. (2002). Dynamic generation of statistical reports using literate data analysis. In HÄRDLE, W. and RÖNZ, B., editors, *COMPSTAT 2002 – Proceedings in Computational Statistics*, pages 575–580, Heidelberg, 2002. Physica-Verlag.
- R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- ZEILEIS, A., HORNIK, K., and MURRELL, P. (2009). Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics and Data Analysis*, **53**(9), 3259–3270. doi:10.1016/j.csda.2008.11.033.
- ZEILEIS, A., FISHER, J. C., HORNIK, K., IHAKA, R., MCWHITE, C. D., MURRELL, P., STAUFFER, R., and WILKE, C. O. (2019). *colorspace: A toolbox for manipulating and assessing colors and palettes*. arXiv 1903.06490, arXiv.org E-Print Archive. URL <http://arxiv.org/abs/1903.06490>.