eda

Kate Colvin

2024-12-11

Data Cleaning

```
score_df <- read_csv("Project_1_data.csv") %>%
  janitor::clean_names()
```

```
## Rows: 948 Columns: 14
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr (10): Gender, EthnicGroup, ParentEduc, LunchType, TestPrep, ParentMarita...
## dbl  (4): NrSiblings, MathScore, ReadingScore, WritingScore
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
score_df %>%
  tbl_summary(by = gender) %>%
  add_overall()
```

Characteristic	Overall N = 948[1]	female N = 488[1]	male N = 460[1]
ethnic_group			
group A	80 (9.0%)	32 (7.0%)	48 (11%)
group B	171 (19%)	91 (20%)	80 (19%)
group C	277 (31%)	156 (34%)	121 (28%)
group D	237 (27%)	117 (25%)	120 (28%)
group E	124 (14%)	63 (14%)	61 (14%)
Unknown	59	29	30
parent_educ			
associate's degree	198 (22%)	101 (22%)	97 (23%)
bachelor's degree	104 (12%)	57 (12%)	47 (11%)
high school	176 (20%)	83 (18%)	93 (22%)
master's degree	55 (6.1%)	33 (7.1%)	22 (5.1%)
some college	199 (22%)	107 (23%)	92 (21%)
some high school	163 (18%)	86 (18%)	77 (18%)
Unknown	53	21	32
lunch_type			
free/reduced	331 (35%)	179 (37%)	152 (33%)
standard	617 (65%)	309 (63%)	308 (67%)
test_prep			
completed	322 (36%)	162 (35%)	160 (37%)
none	571 (64%)	297 (65%)	274 (63%)
Unknown	55	29	26
parent_marital_status			
divorced	146 (16%)	75 (16%)	71 (16%)
married	516 (57%)	255 (55%)	261 (60%)
single	213 (24%)	120 (26%)	93 (21%)
widowed	24 (2.7%)	12 (2.6%)	12 (2.7%)
Unknown	49	26	23
practice_sport			
never	112 (12%)	57 (12%)	55 (12%)
regularly	343 (37%)	177 (37%)	166 (37%)
sometimes	477 (51%)	249 (52%)	228 (51%)
Unknown	16	5	11
is_first_child	604 (66%)	321 (68%)	283 (64%)
Unknown	30	13	17
nr_siblings			
0	101 (11%)	40 (8.7%)	61 (14%)
1	245 (27%)	129 (28%)	116 (26%)
2	213 (24%)	109 (24%)	104 (24%)
3	198 (22%)	106 (23%)	92 (21%)
4	76 (8.4%)	43 (9.3%)	33 (7.5%)
5	50 (5.5%)	26 (5.7%)	24 (5.4%)
6	8 (0.9%)	6 (1.3%)	2 (0.5%)
7	11 (1.2%)	1 (0.2%)	10 (2.3%)
Unknown	46	28	18
transport_means[2]			
private	337 (40%)	174 (39%)	163 (40%)
school_bus	509 (60%)	268 (61%)	241 (60%)
Unknown	102	46	56