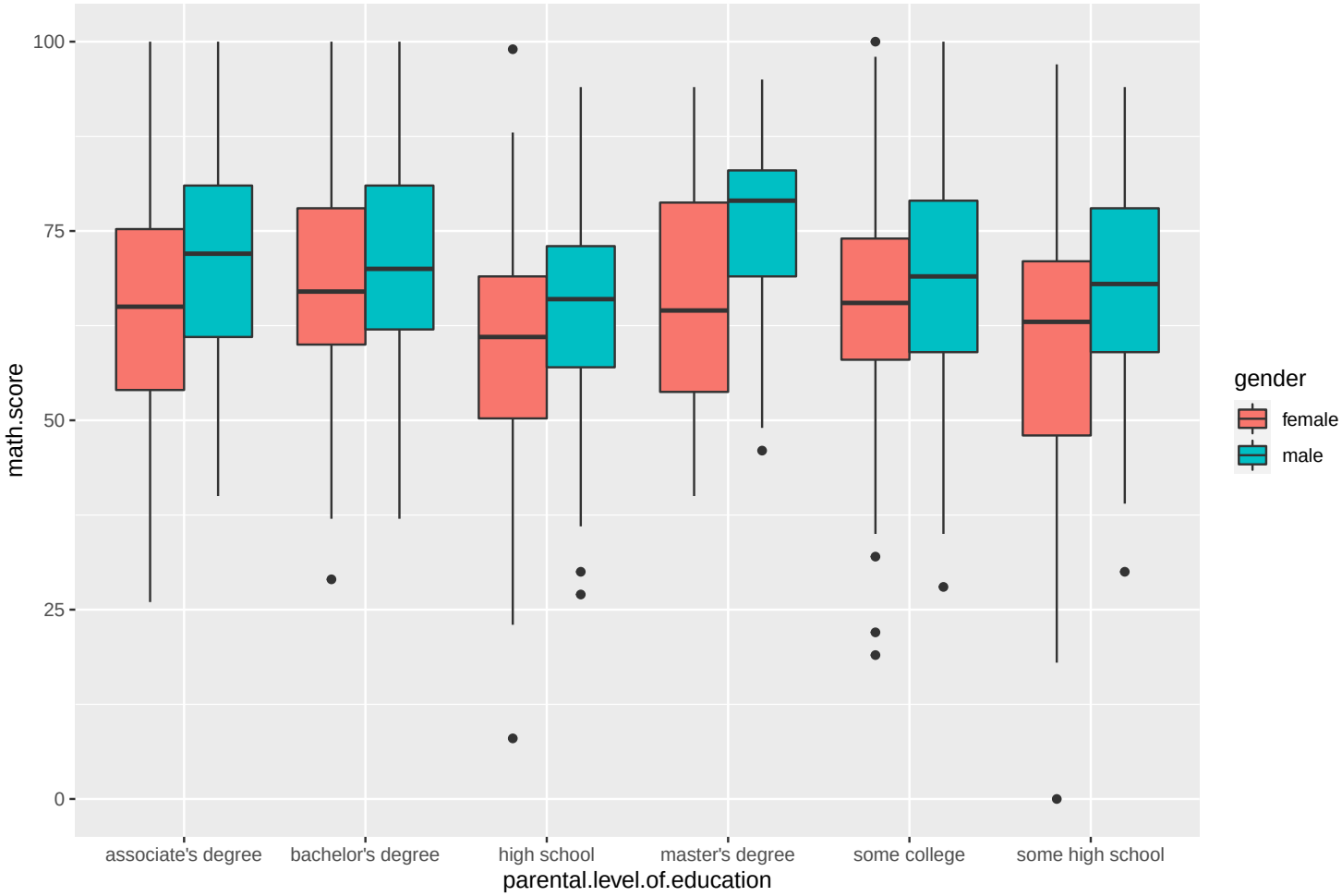
Student performance

薛貴林

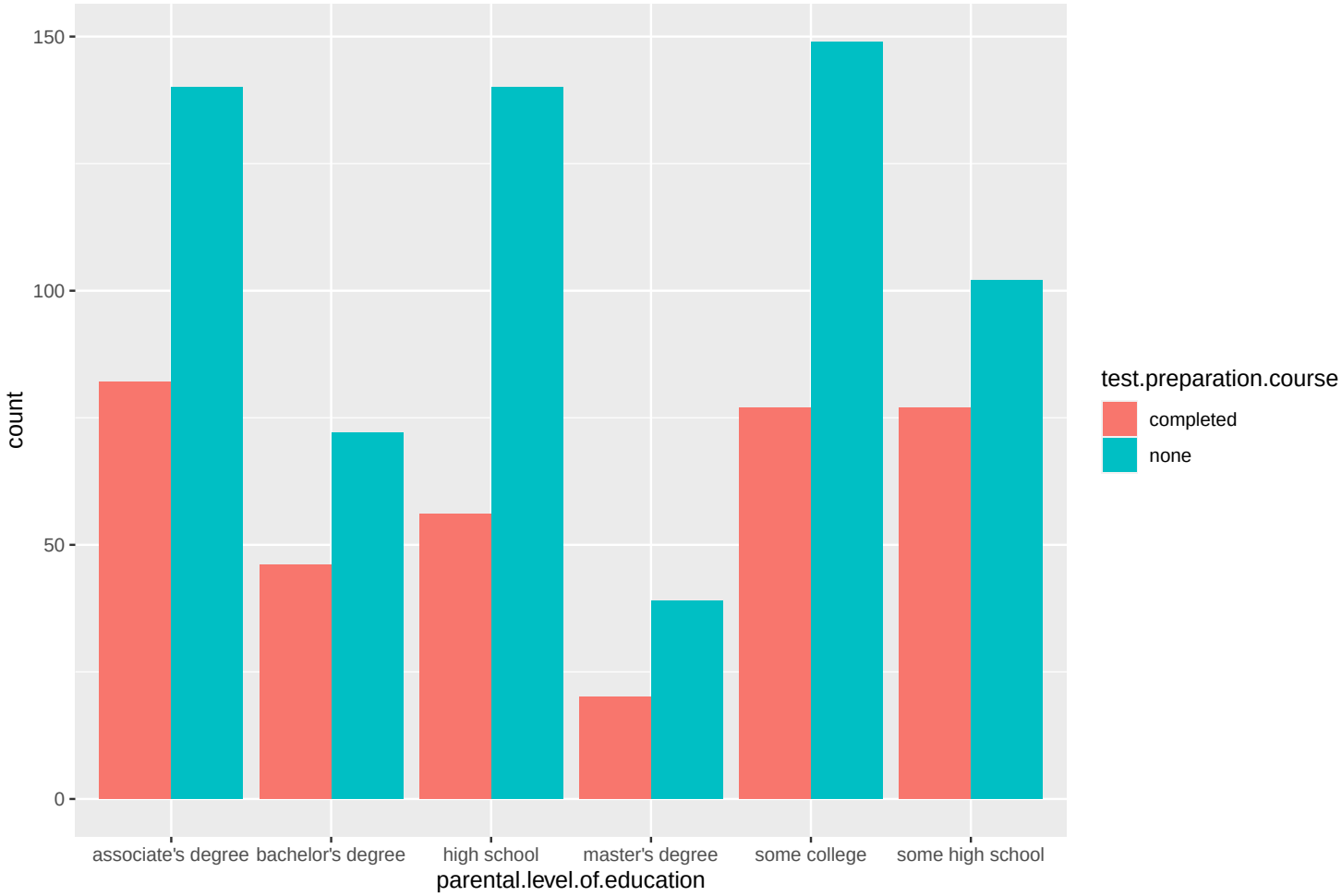
```
> setwd("D:/R/R-exercise/student_performance")
> data <- read.csv("StudentsPerformance.csv")
> library(tidyverse)
> library(kableExtra)
> knitr::kable(head(data), format = "latex", booktabs = TRUE) %>% kable_styling(latex_options = "scale_dow
```

gender	race.ethnicity	parental.level.of.education	lunch	test.preparation.course	math.score	reading.score	writing.score
female	group B	bachelor's degree	standard	none	72	72	74
female	group C	some college	standard	completed	69	90	88
female	group B	master's degree	standard	none	90	95	93
male	group A	associate's degree	free/reduced	none	47	57	44
male	group C	some college	standard	none	76	78	75
female	group B	associate's degree	standard	none	71	83	78

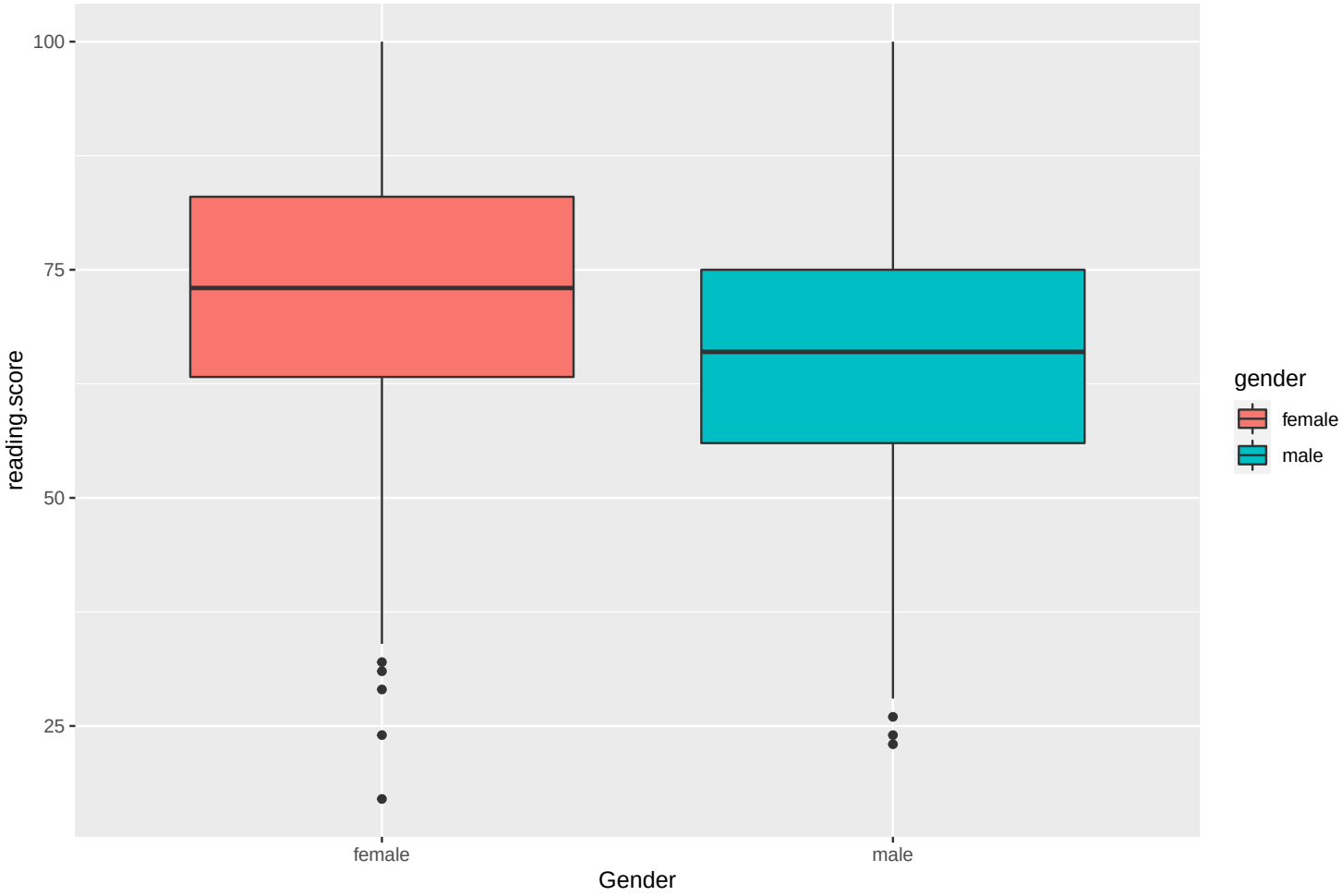
```
>
>
> # 分布情况 父母教育水平性别和数学分数
> ggplot(data, aes(x = parental.level.of.education, y = math.score, fill = gender)) +
+     geom_boxplot(position = "dodge")  # position =dodge是用于分组绘图的
```



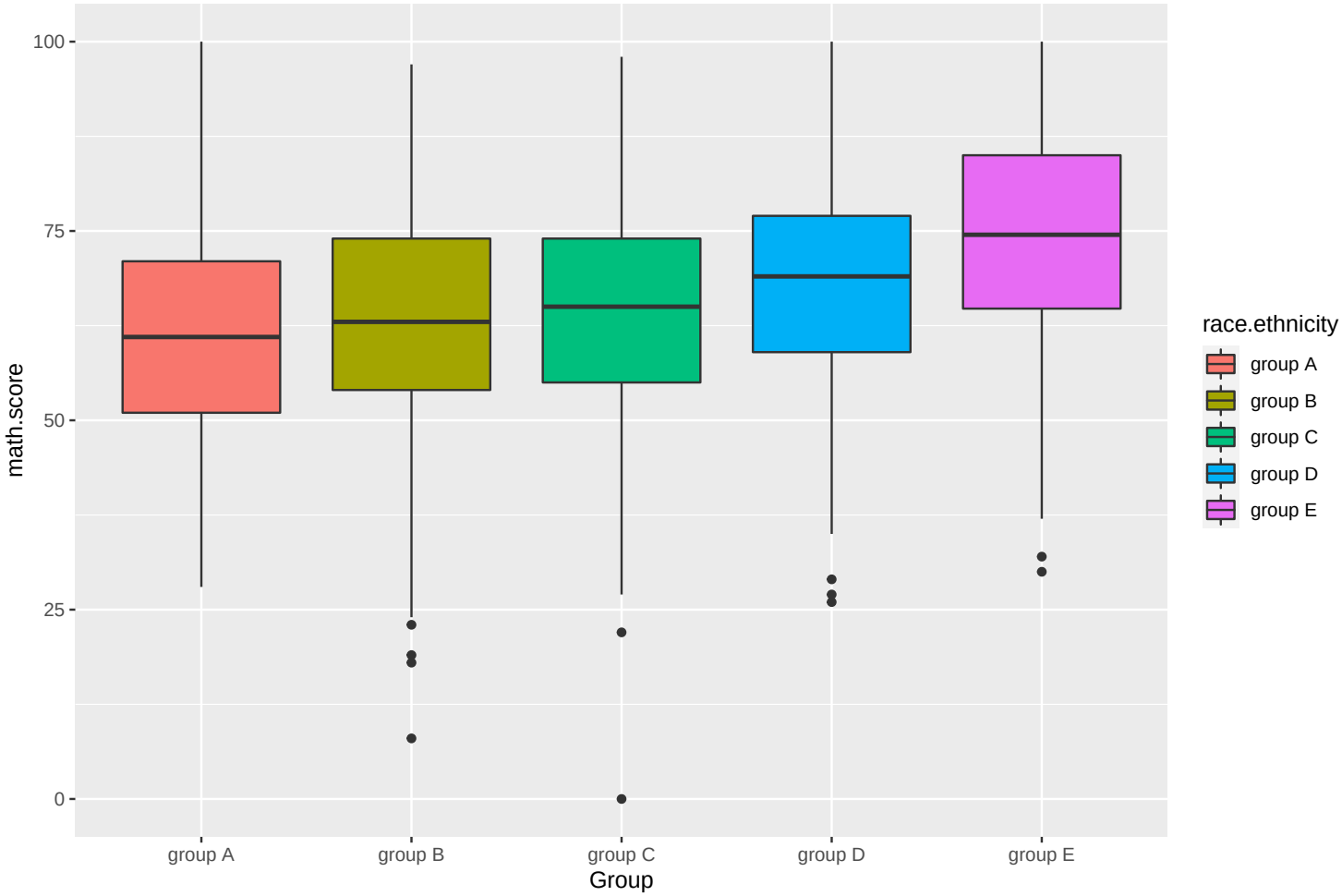

```
> # 父母教育水平和成绩通过
> ggplot(data, aes(x = parental.level.of.education, fill = test.preparation.course)) +
+     geom_bar(position = "dodge")
```



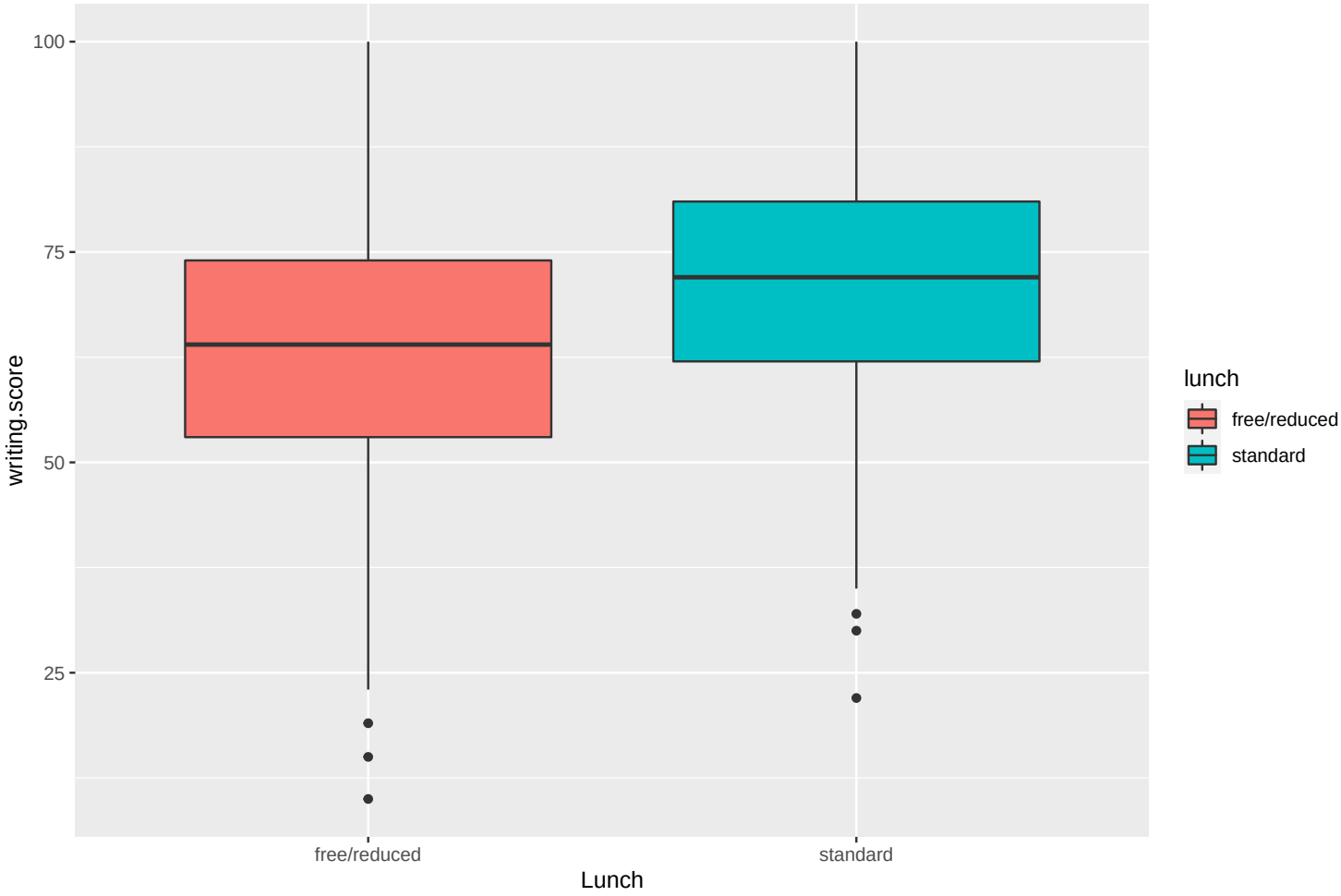

```
>
>
>
> # 性别和阅读成绩
> ggplot(data, aes(x = factor(gender), y = reading.score, fill = gender)) + geom_boxplot() +
+     xlab("Gender")
```



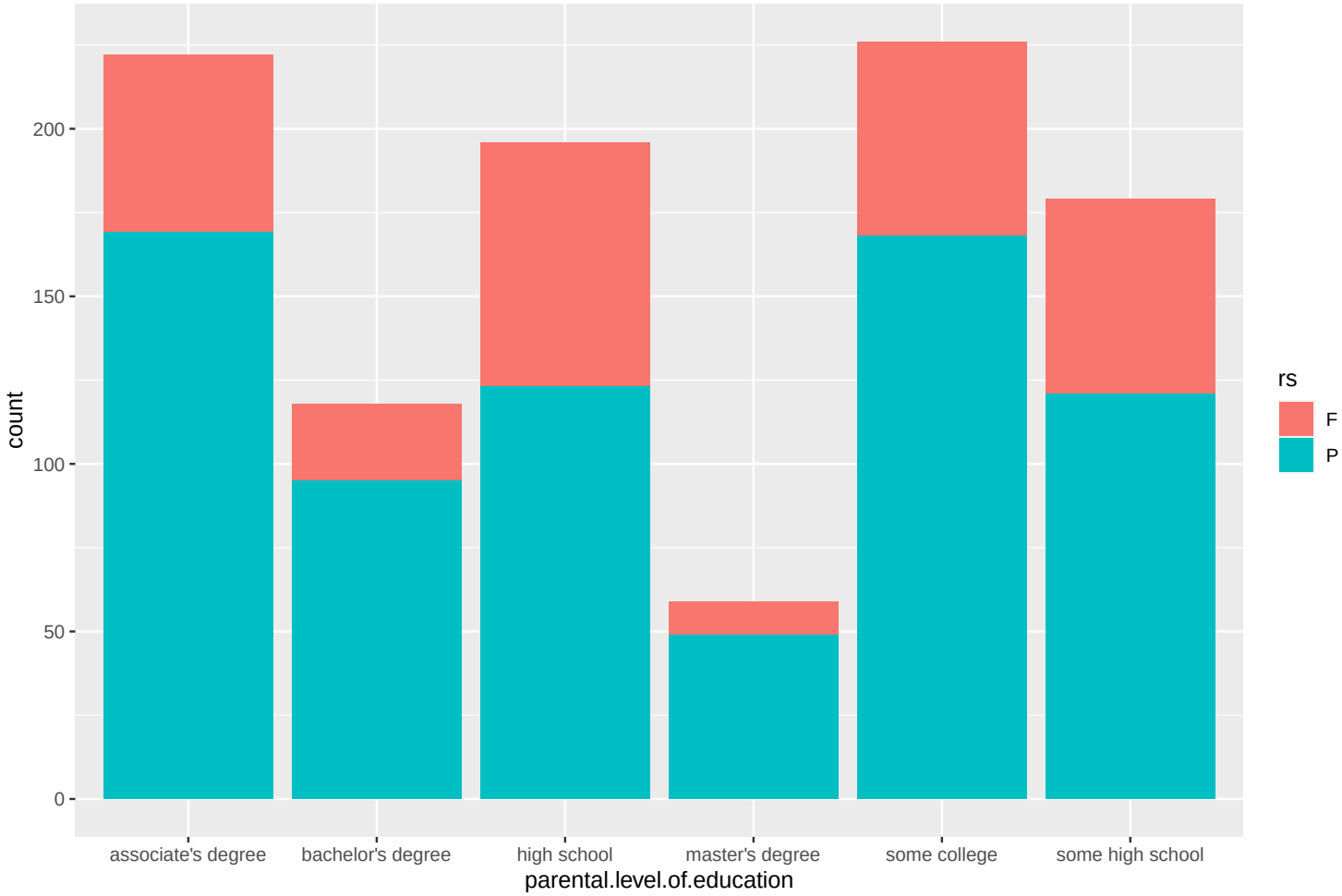

```
> # 种族和数学成绩
> ggplot(data, aes(x = factor(race.ethnicity), y = math.score, fill = race.ethnicity)) +
+     geom_boxplot() + xlab("Group")
```



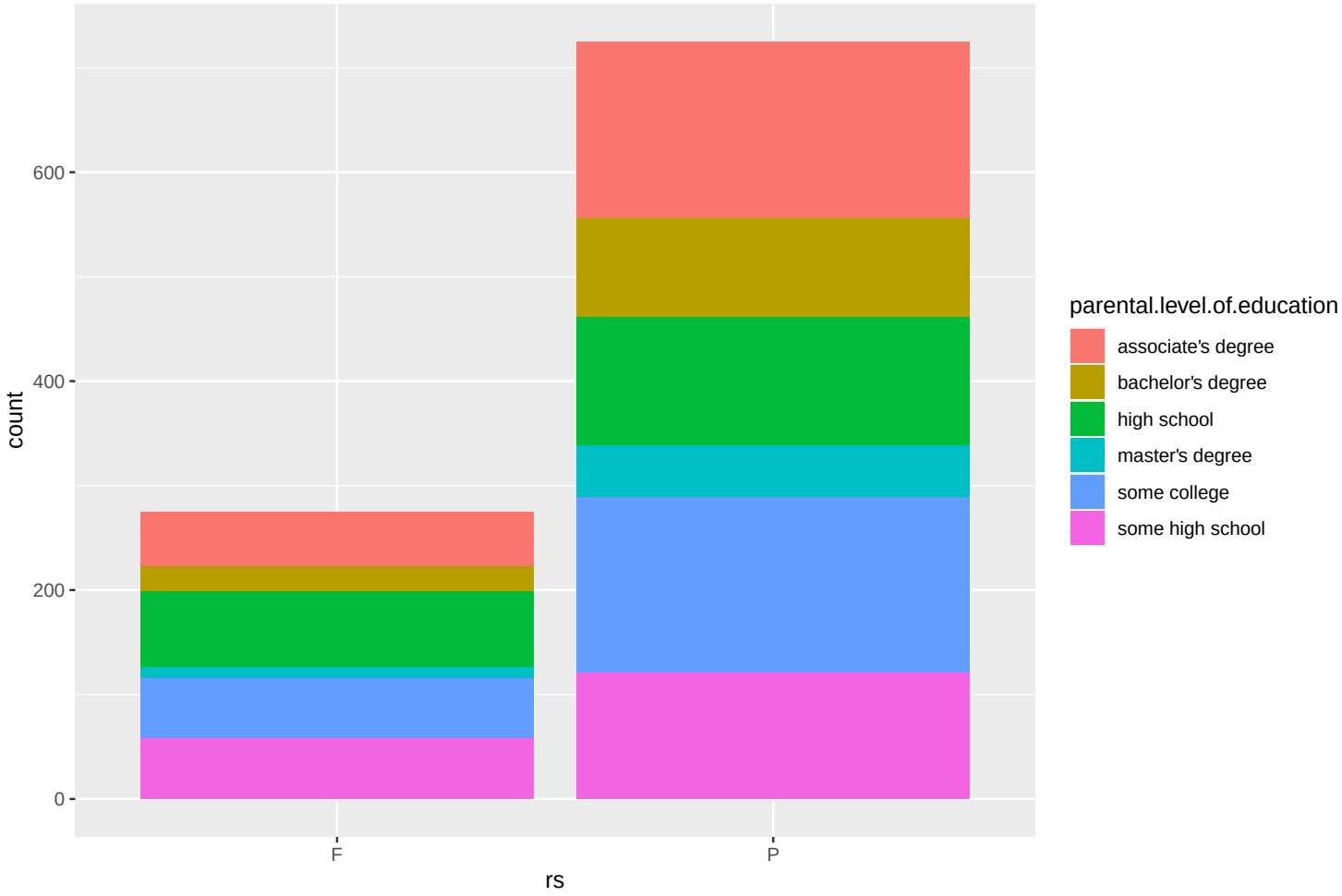

```
> # 午饭和写作成绩
> ggplot(data, aes(x = factor(lunch), y = writing.score, fill = lunch)) + geom_boxplot() +
+     xlab("Lunch")
```



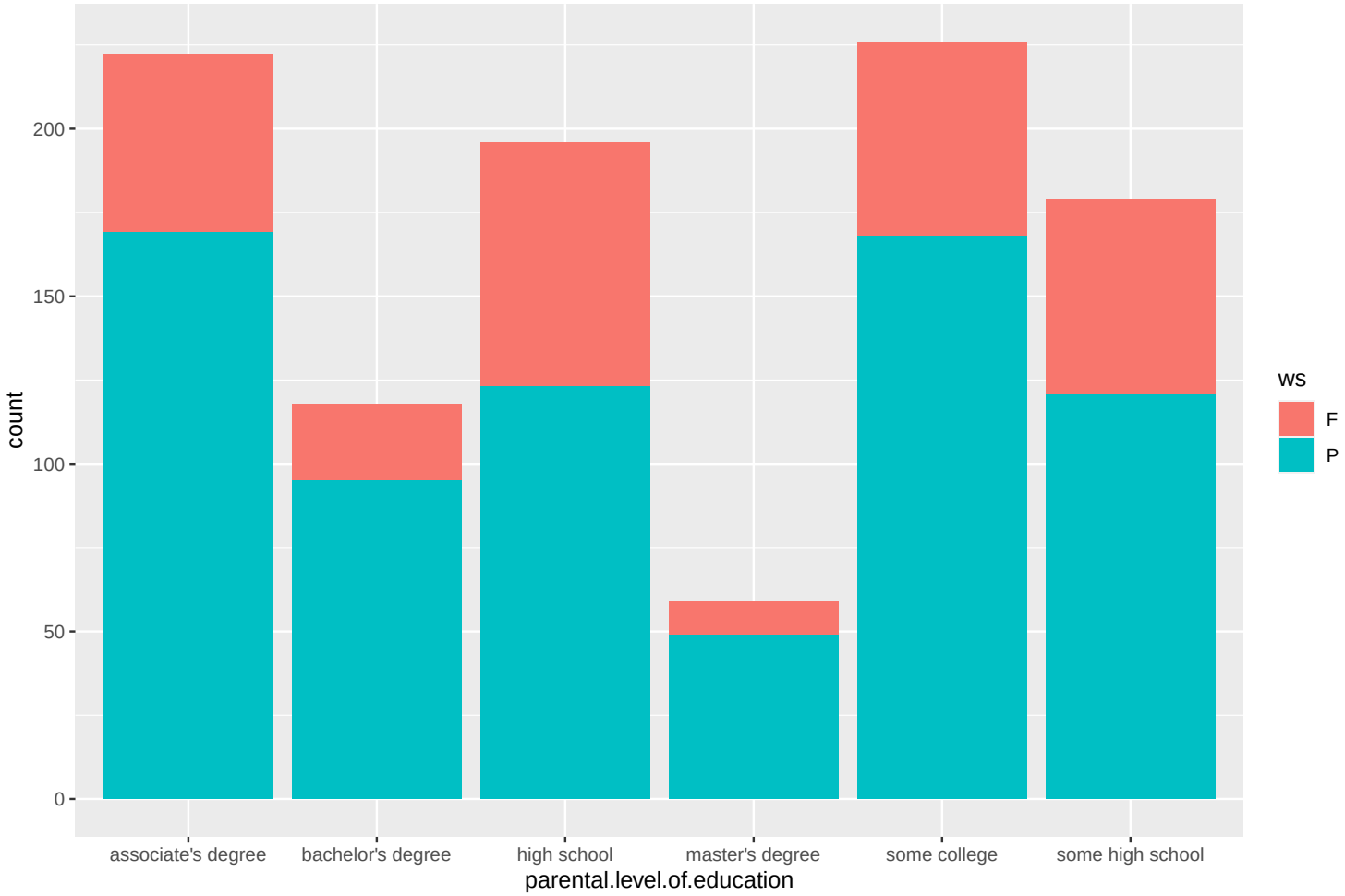

```
>
> # 父母教育水平和阅读成绩是否合格
> rs <- ifelse(data$reading.score > 60, "P", "F")
>
> data$rs <- rs
>
> ggplot(data, aes(x = parental.level.of.education, fill = rs)) + geom_bar()
```



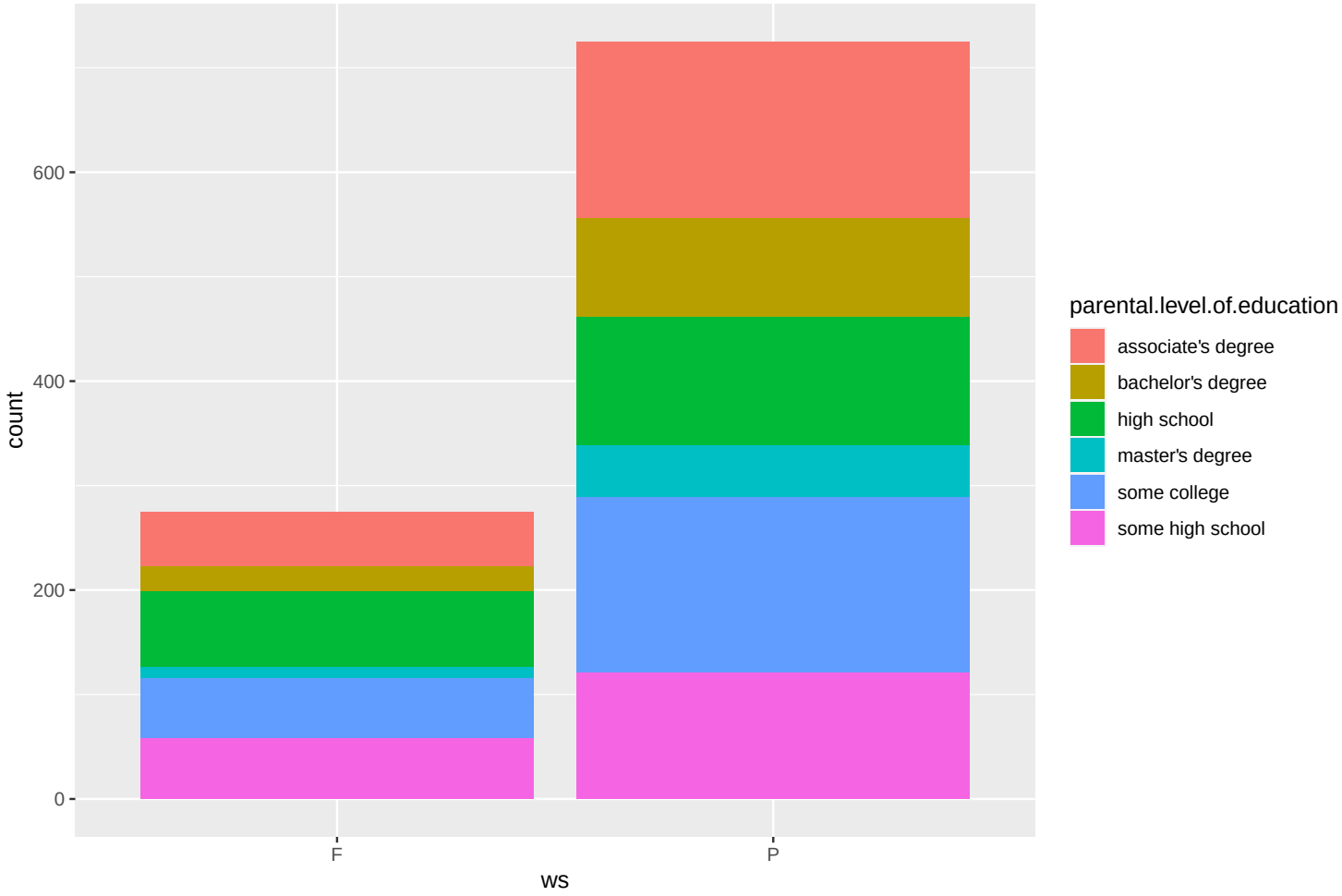

```
>
> ggplot(data, aes(x = rs, fill = parental.level.of.education)) + geom_bar()
```



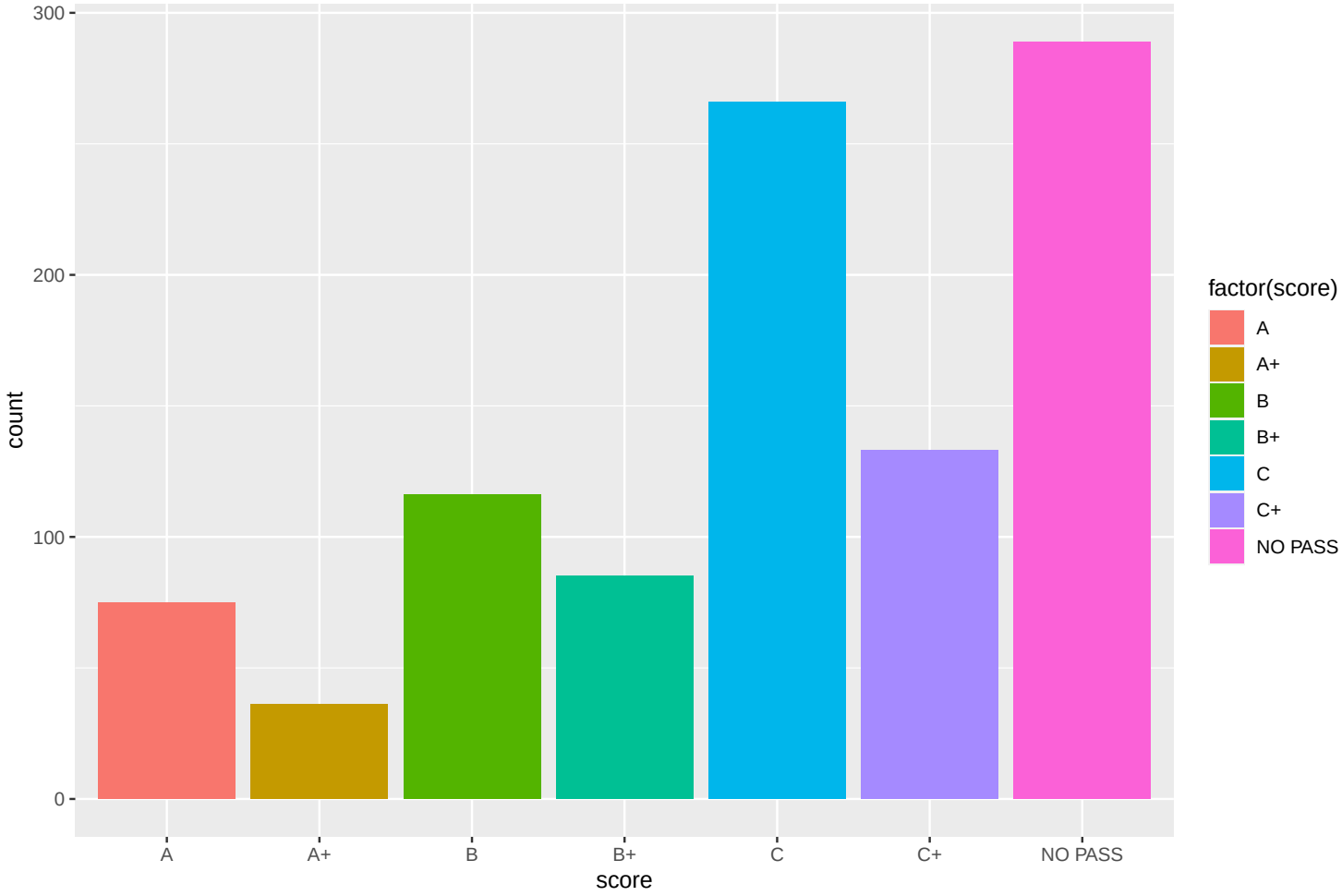

```
>
> # 父母教育水平和写作成绩是否合格
> ws <- ifelse(data$reading.score > 60, "P", "F")
> data$ws <- ws
> ggplot(data, aes(x = parental.level.of.education, fill = ws)) + geom_bar()
```



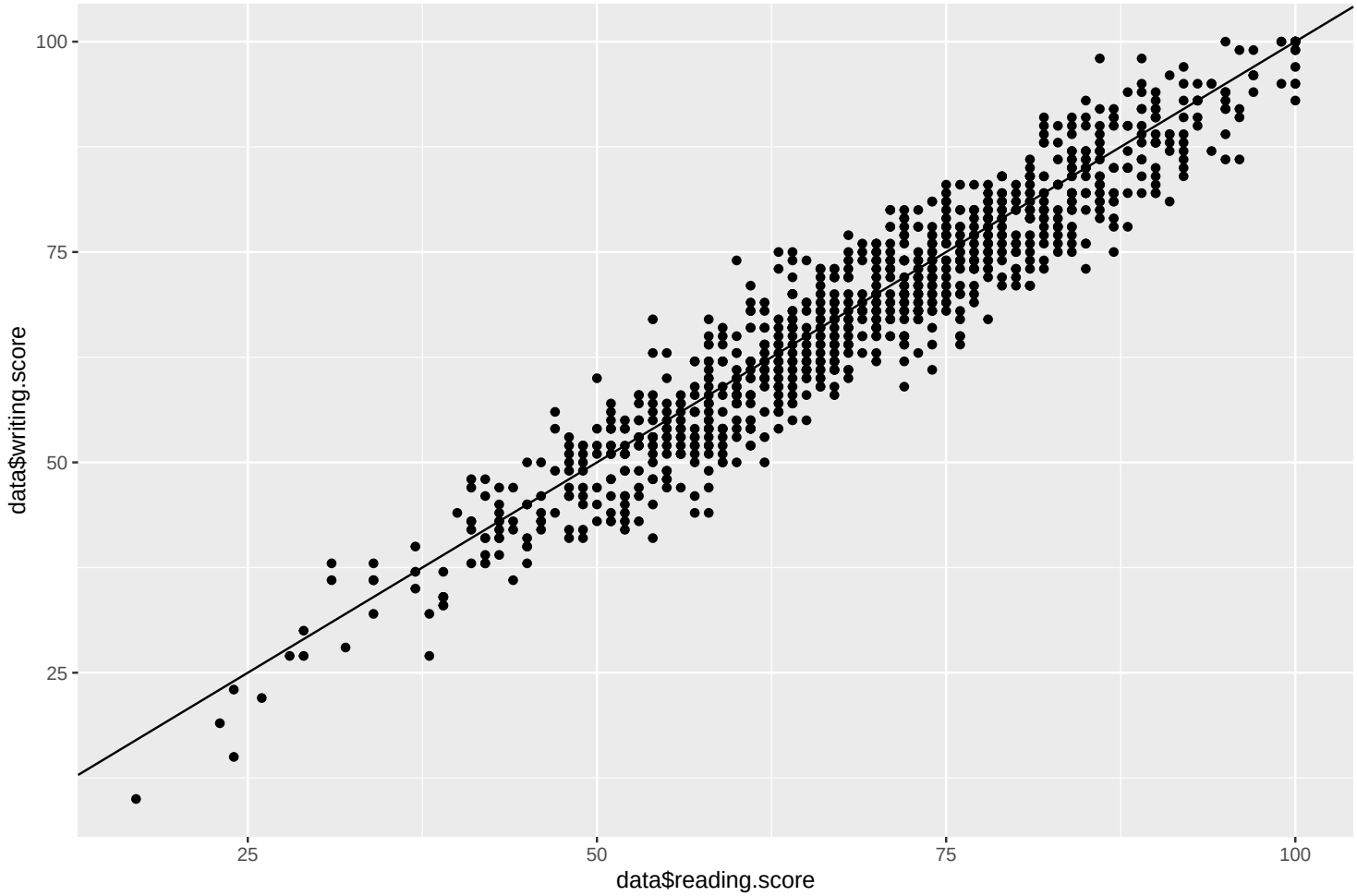

```
>
> ggplot(data, aes(x = ws, fill = parental.level.of.education)) + geom_bar()
```




```
>
>
> # 总评成绩分布
> a <- data$math.score * 0.4 + data$reading.score * 0.3 + data$writing.score * 0.3
> score <- function(a) {
+     ifelse(a > 92, "A+", ifelse(a > 85, "A", ifelse(a > 80, "B+", ifelse(a > 75,
+         "B", ifelse(a > 70, "C+", ifelse(a > 65, "C", ifelse(a > 60, "C", "NO PASS")))))))
+ }
> data$score <- score(a)
>
> ggplot(data, aes(x = score, fill = factor(score))) + geom_bar()
```




```
>
> # 阅读成绩和写作成绩的关系
> ggplot(data, aes(x = data$reading.score, y = data$writing.score)) + geom_point() +
+     geom_abline()
```


100
75
50
25
data$writing.score
25
50
75
100
data$reading.score