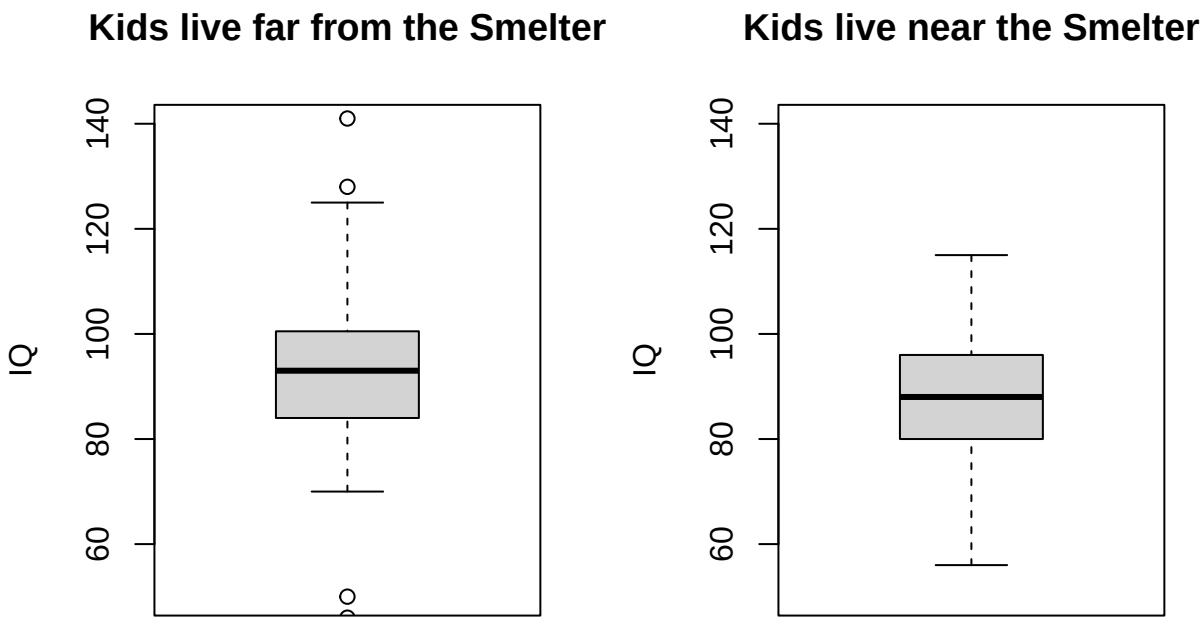
Consulting Report

Guangyu Qiu

2024-10-07

```
lead_data <- read.csv("C:/Users/16462/Documents/Qiu_Project_01/DataRaw/lead-iq-01.csv")
```

```
# Change the 999 outlier to 99
lead_data[lead_data$IQ == 999, "IQ"] <- 99
lead_data_far <- lead_data$IQ[lead_data$Smelter == "Far"]
lead_data_near <- lead_data$IQ[lead_data$Smelter == "Near"]
par(mfrow = c(1, 2))
boxplot(lead_data_far, main = "Kids live far from the Smelter",
     ylab = "IQ",ylim = c(50, 140))
boxplot(lead_data_near, main = "Kids live near the Smelter",
     ylab = "IQ",ylim = c(50, 140))
```



```r
par(mfrow = c(1, 1))
```

From the boxplot, we can see that the box representing the IQ of kids living far from the smelter is higher than the box representing the IQ of kids living near the smelter, indicating a higher overall IQ. This can also be seen from the heights of the black lines representing the mean and the 25th and 75th percentiles, which are all higher for children living far away.

To be more quantitive, The mean IQ of the kids living far from the smelter is 92.6865672. The mean IQ of the kids living near the smelter is 89.1929825.

```r
library(knitr)
iq_breaks <- c(0, seq(60, 130, by = 10), Inf)
iq_labels <- c("Below 60", "60-769", "70-79", "80-89", "90-99", "100-109",
               "110-119", "120-129", "Above 130")
lead_data$IQ_range <- cut(lead_data$IQ, breaks = iq_breaks,
                          labels = iq_labels, right = FALSE)
lead_summary <- table(lead_data$Smelter, lead_data$IQ_range)
kable(lead_summary, caption = "Count of IQ by Location")
```

Table 1: Count of IQ by Location

	Below 60	60-769	70-79	80-89	90-99	100-109	110-119	120-129	Above 130
Far	2	0	11	14	21	12	3	3	1
Near	1	0	11	21	13	7	4	0	0

From the table, we can observe that the majority of children living far away have IQs in the 90-99 range, while most children living near the smelter have IQs in the 80-89 range. Additionally, there are 4 children living far away with IQs above 120, whereas there are no children living nearby with IQs above 120.