

Moreno_Project_01

Mauricio Moreno

2025-11-16

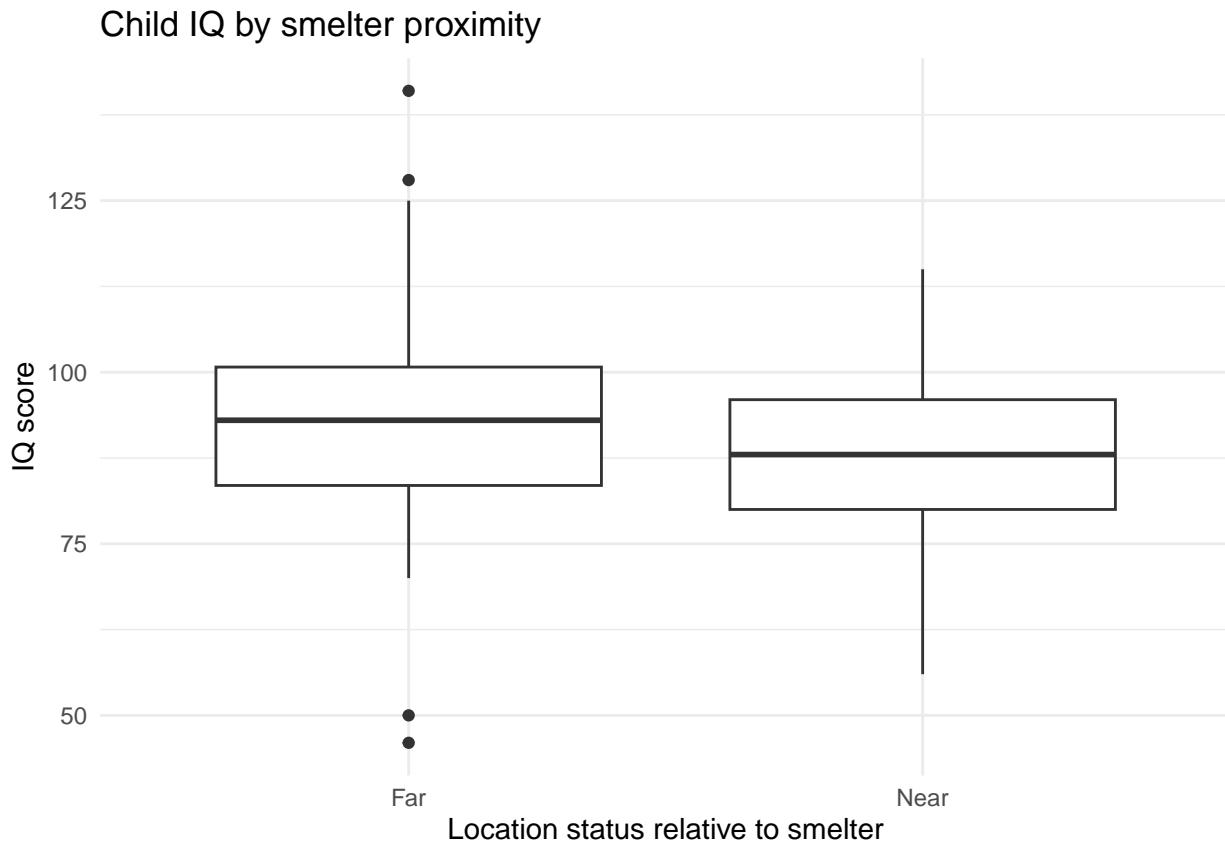
```
lead <- readr::read_csv("DataRaw/lead-iq-01.csv", show_col_types = FALSE)

# Basic cleaning, treat 999 IQ as missing
lead <- lead %>%
  mutate(IQ = if_else(IQ == 999, NA_real_, as.numeric(IQ)),
        Smelter = factor(Smelter)
  )
```

Overall description

The dataset contains 124 children. The mean IQ for children living near the smelter is 89.2, and for those living far from the smelter is 92.6.

```
# IQ levels by location status
ggplot(lead, aes(x = Smelter, y = IQ)) +
  geom_boxplot() +
  labs(x = "Location status relative to smelter",
       y = "IQ score",
       title = "Child IQ by smelter proximity") +
  theme_minimal()
```



The boxplot shows how IQ scores vary between children living near versus far from the smelter. We can visually compare the medians, spread, and any potential outliers across location status.

```

iq_summary <- lead %>%
  group_by(Smelter) %>%
  summarize(
    n = sum(!is.na(IQ)),
    mean_IQ = mean(IQ, na.rm = TRUE),
    sd_IQ   = sd(IQ, na.rm = TRUE),
    .groups = "drop"
  )

kable(
  iq_summary,
  caption = "Summary of IQ by smelter proximity",
  digits = 1
) %>%
  kable_styling(
    full_width = FALSE,
    position   = "center",
    font_size  = 14
)

```

Table 1: Summary of IQ by smelter proximity

Smelter	n	mean_IQ	sd_IQ
Far	66	92.6	16.1
Near	57	89.2	12.2

The table above summarizes IQ levels by smelter location status. For each group (near vs. far), it reports the number of non-missing IQ observations, the mean IQ, and the standard deviation. Together with the boxplot, this lets us see whether IQ tends to be lower or higher among children living near the smelter and how variable the scores are in each group.