

Lead IQ Report

Mahfuza Haque Mahi

2025-11-12

Data

We analyze the CDC “lead-iq-01.csv” dataset, which contains information on 124 children aged 3–15 years who lived at varying distances from a lead smelter in El Paso, Texas.

The dataset includes two variables:

Smelter - indicates whether the child lived Near (within 1 mile) or Far (more than 1 mile) from the lead smelter.

IQ – the child’s score on the Wechsler Intelligence Scale for Children (WISC), a standardized measure of cognitive ability.

The goal of this analysis is to compare IQ levels between children who lived near the smelter and those who lived farther away, to explore whether proximity to the smelter (and potential lead exposure) is associated with lower IQ scores.

```
## Smelter      IQ
## Far :67   Min.   : 46.00
## Near:57   1st Qu.: 81.50
##                   Median : 91.00
##                   Mean   : 91.08
##                   3rd Qu.: 99.00
##                   Max.   :141.00
```

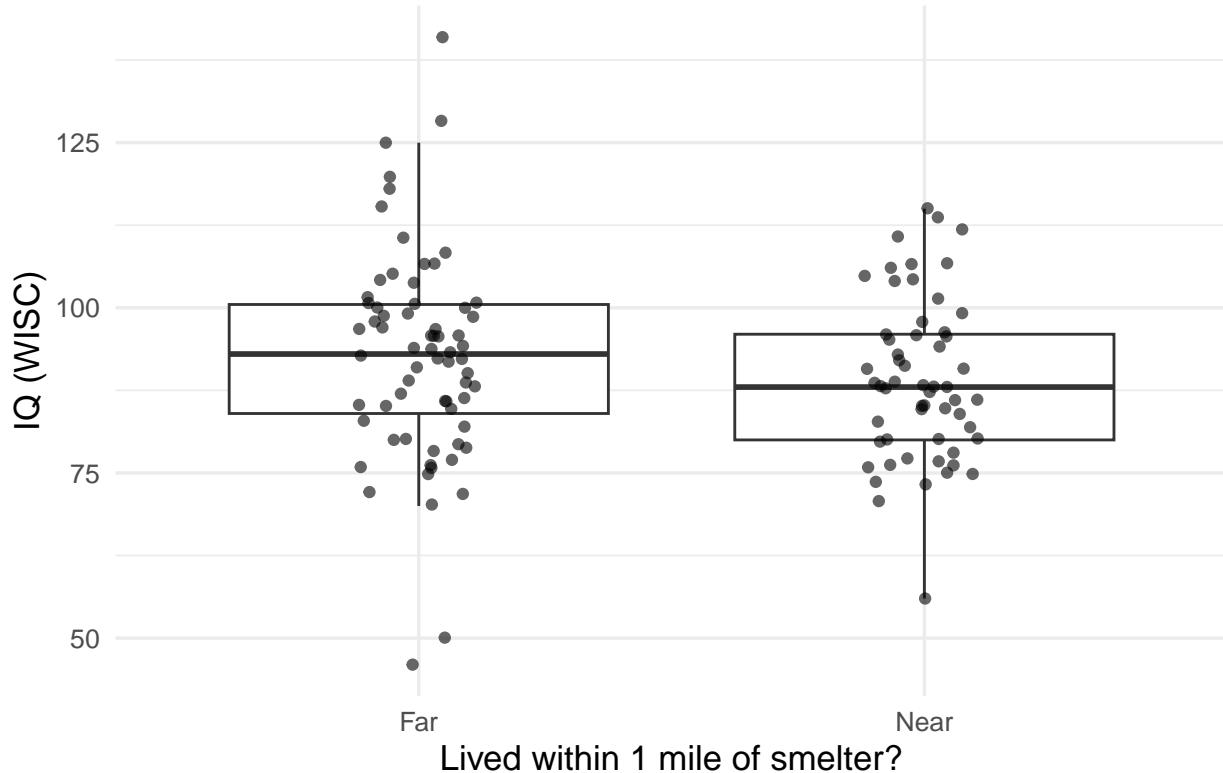
Descriptives

Overall mean IQ is 91.1.

Children **Near** the smelter have mean IQ 89.2, while those **Far** have mean IQ 92.7.

Graph: IQ by smelter status

IQ by smelter proximity



Interpretation:

Boxplots and jittered points show the IQ distribution by proximity. We see lower average IQ among children living Near compared with those Far.

Table: summary by group

Table 1: IQ summary statistics by smelter proximity

Smelter	n	mean_IQ	sd_IQ	median_IQ	q1	q3
Far	67	92.7	16.0	93	84	100.5
Near	57	89.2	12.2	88	80	96.0

Interpretation: The table summarizes center and spread of IQ within each group. Children living near the smelter ($n = 57$) had a slightly lower mean IQ (89.2 ± 12.2) compared with those living far from the smelter ($n = 67$, mean = 92.7 ± 16.0). Median IQ values and interquartile ranges also suggest slightly lower central tendency and somewhat narrower variability among the Near group.

Code Appendix

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)
required_pkgs <- c("tidyverse", "here", "knitr", "gtsummary")
to_install <- setdiff(required_pkgs, rownames(installed.packages()))
if (length(to_install)) install.packages(to_install, quiet = TRUE)
lapply(required_pkgs, library, character.only = TRUE)
# Read raw data from DataRaw/
lead_raw <- readr::read_csv(here::here("DataRaw", "lead-iq-01.csv"), show_col_types = FALSE)

lead <- lead_raw |>
  mutate(
    Smelter = factor(Smelter, levels = c("Far", "Near")),
    IQ = suppressWarnings(as.numeric(IQ)),
    IQ = ifelse(IQ == 999, 99, IQ) # correction per PI conversation
  )

stopifnot(all(c("Smelter", "IQ") %in% names(lead)))
summary(lead)
ggplot(lead, aes(x = Smelter, y = IQ)) +
  geom_boxplot(outlier.alpha = 0) +
  geom_jitter(width = 0.12, alpha = 0.6) +
  labs(x = "Lived within 1 mile of smelter?", y = "IQ (WISC)",
       title = "IQ by smelter proximity") +
  theme_minimal(base_size = 13)
tbl <- lead |>
  group_by(Smelter) |>
  summarise(
    n = dplyr::n(),
    mean_IQ = mean(IQ, na.rm = TRUE),
    sd_IQ = sd(IQ, na.rm = TRUE),
    median_IQ = median(IQ, na.rm = TRUE),
    q1 = quantile(IQ, 0.25, na.rm = TRUE),
    q3 = quantile(IQ, 0.75, na.rm = TRUE)
  )

knitr::kable(
  tbl,
  digits = c(0,0,1,1,1,1,1),
  caption = "IQ summary statistics by smelter proximity"
)
```