Lead IQ Report

Mahfuza Haque Mahi

2025-11-12

```
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)
```

```
## [[1]]
##  [1] "lubridate" "forcats"   "stringr"   "dplyr"     "purrr"     "readr"
##  [7] "tidyr"     "tibble"    "ggplot2"   "tidyverse" "stats"     "graphics"
## [13] "grDevices" "utils"     "datasets"  "methods"   "base"
##
## [[2]]
##  [1] "here"      "lubridate" "forcats"   "stringr"   "dplyr"     "purrr"
##  [7] "readr"     "tidyr"     "tibble"    "ggplot2"   "tidyverse" "stats"
## [13] "graphics"  "grDevices" "utils"     "datasets"  "methods"   "base"
##
## [[3]]
##  [1] "knitr"     "here"      "lubridate" "forcats"   "stringr"   "dplyr"
##  [7] "purrr"     "readr"     "tidyr"     "tibble"    "ggplot2"   "tidyverse"
## [13] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [19] "base"
##
## [[4]]
##  [1] "gtsummary" "knitr"     "here"      "lubridate" "forcats"   "stringr"
##  [7] "dplyr"     "purrr"     "readr"     "tidyr"     "tibble"    "ggplot2"
## [13] "tidyverse" "stats"     "graphics"  "grDevices" "utils"     "datasets"
## [19] "methods"   "base"
```

Data and cleaning

We analyze the CDC "lead-iq-01.csv" dataset (124 children, variables: Smelter: Near/Far; IQ: WISC IQ). During QC we learned one IQ was mistakenly entered as 999 and should be 99 (see code below).

```
# 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))
  )
```

```r
stopifnot(all(c("Smelter","IQ") %in% names(lead)))
summary(lead)
```

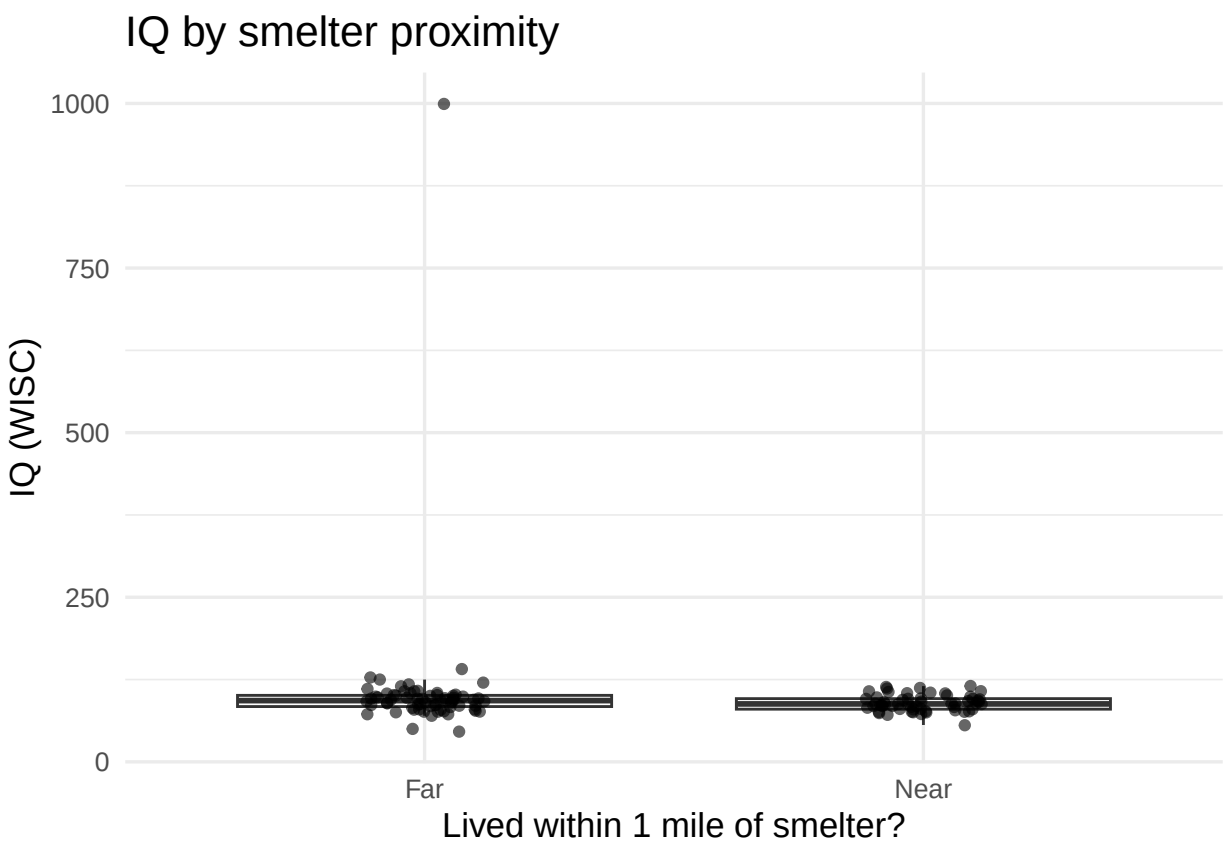
```
##  Smelter         IQ
##  Far :67   Min.   : 46.00
##  Near:57   1st Qu.: 81.50
##            Median : 91.00
##            Mean   : 98.34
##            3rd Qu.: 99.25
##            Max.   :999.00
```

Descriptive text with inline values

Overall mean IQ is 98.3.
Children **Near** the smelter have mean IQ 89.2,
while those **Far** have mean IQ 106.1.

Graph: IQ by smelter status

```r
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)
```



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

```
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"
)
```

Table 1: IQ summary statistics by smelter proximity

Smelter	n	mean_IQ	sd_IQ	median_IQ	q1	q3
Far	67	106.1	111.9	93	84	101
Near	57	89.2	12.2	88	80	96

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.