

# Project\_01

Diana Felipe

2024-10-03

```
# read in data
library(readr)
lead_iq <- read_csv("../DataRaw/lead-iq-01.csv")
```

## Part a

Produce a graph showing IQ levels by location status (with outlier)

```
# graph showing IQ levels by location status
library(ggplot2)
ggplot(lead_iq, aes(x = Smelter, y = IQ, fill = Smelter)) +
  geom_boxplot() +
  scale_y_continuous(breaks = seq(0, max(lead_iq$IQ) + 30, by = 100))
```

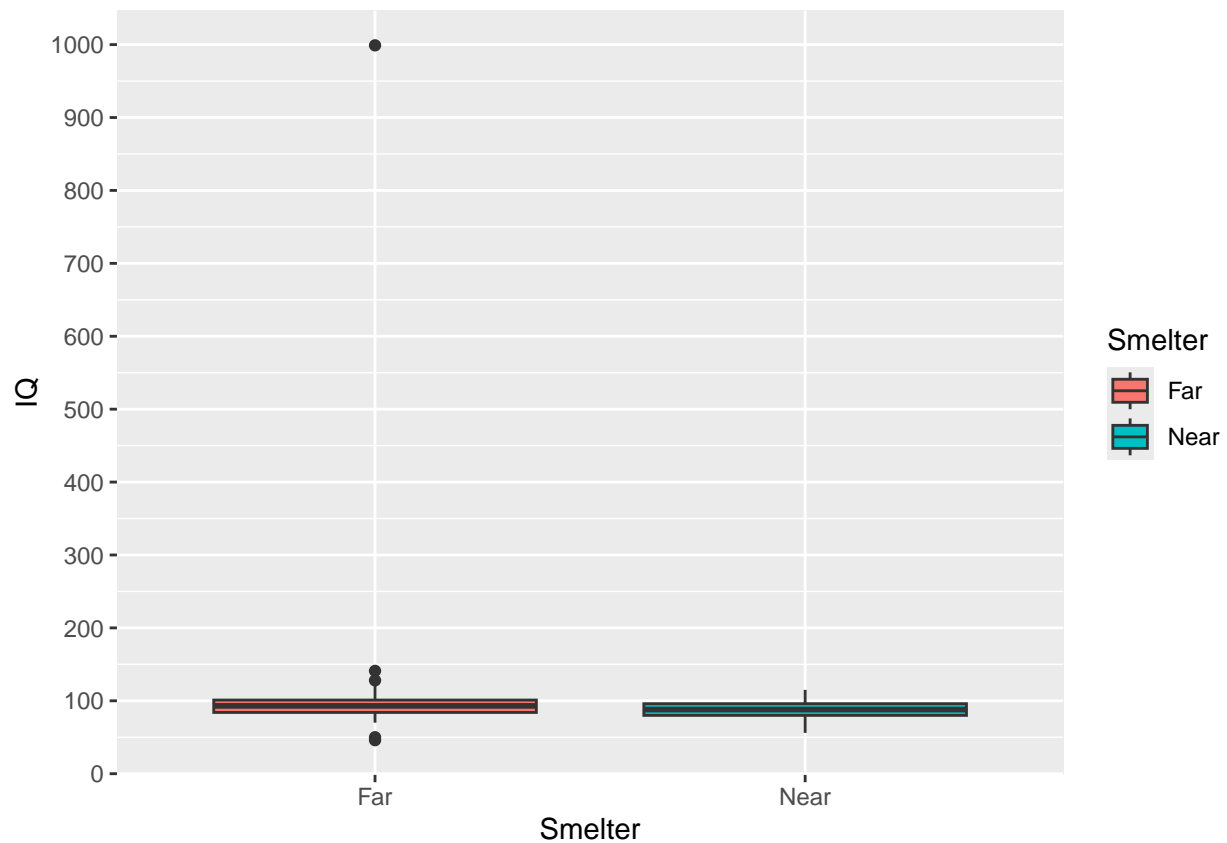


Table 1: Tables of Smelter (Near and Far) and IQ

	Smelter	IQ	Smelter	IQ
68	Near	84	Far	70
69	Near	56	Far	85
70	Near	77	Far	86
71	Near	80	Far	76
72	Near	86	Far	96
73	Near	88	Far	94

## Part b

Produce a nicely formatted table

```
# table with kable function
d1 <- head(lead_iq[lead_iq$Smelter == "Near", ])
d2 <- head(lead_iq[lead_iq$Smelter == "Far", ])
knitr::kable(
  list(d1, d2),
  caption = 'Tables of Smelter (Near and Far) and IQ',
  valign = 't')
```

```
# second table for summary stats
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##   %+%, alpha
```

```
describeBy(lead_iq, group=lead_iq$Smelter, fast = TRUE)
```

```
##
## Descriptive statistics by group
## group: Far
##      vars  n  mean    sd median min max range skew kurtosis  se
## Smelter   1 67   1.00  0.00     1   1   1     0  NaN      NaN  0.00
## IQ        2 67 106.12 111.88    93  46 999   953 7.58   57.55 13.67
## -----
## group: Near
##      vars  n  mean    sd median min max range skew kurtosis  se
## Smelter   1 57   2.00  0.00     2   2   2     0  NaN      NaN  0.00
## IQ        2 57  89.19 12.17    88  56 115    59 0.14   -0.19 1.61
```

## Part c

A couple sentences of text (not on the graph, but in the body of the Rmarkdown document) describing the graph and the table

From the boxplot, we can noticeably see an outlier at around 1000. With this plot, it is difficult to compare those who lived far or near to the Smelter. However, we can see that they appear to have a similar range of IQ's. Once adjusting the y limits in the plot to exclude the outlier, we are more easily able to see that those who live further away from the smelter seem to generally have a higher IQ. The mean of the Far group is higher and the range is slightly higher than than of the near group. We also see some noticeable outliers in the Far group. Overall, with the boxplots there appears to be a slight difference in the range of IQ of those who live near to the smelter compared to those who live further away.

From our table, we can see that the Far group's mean IQ is 106.12 with an SD of 111.88 and median of 93. We also have a max of 999 which obviously effects these summary stats. For the near group, we have that the mean is 89.19, sd is 12.17, and median is 88. With the outlier, it is a bit challenging to tell how different the groups vary in IQ through the histogram and table.

## Part d

In-line calculations giving the values of the means, that will be updated if the data are changed

```
mean_func <- function(x) paste(c(round(mean(x[x$Smelter == "Far", 2]), 4) ,  
                                round(mean(x[x$Smelter == "Near", 2]), 4)),  
                                collapse = " and ")
```

The mean IQ of the Far and Near groups, respectively are 106.1194 and 89.193.