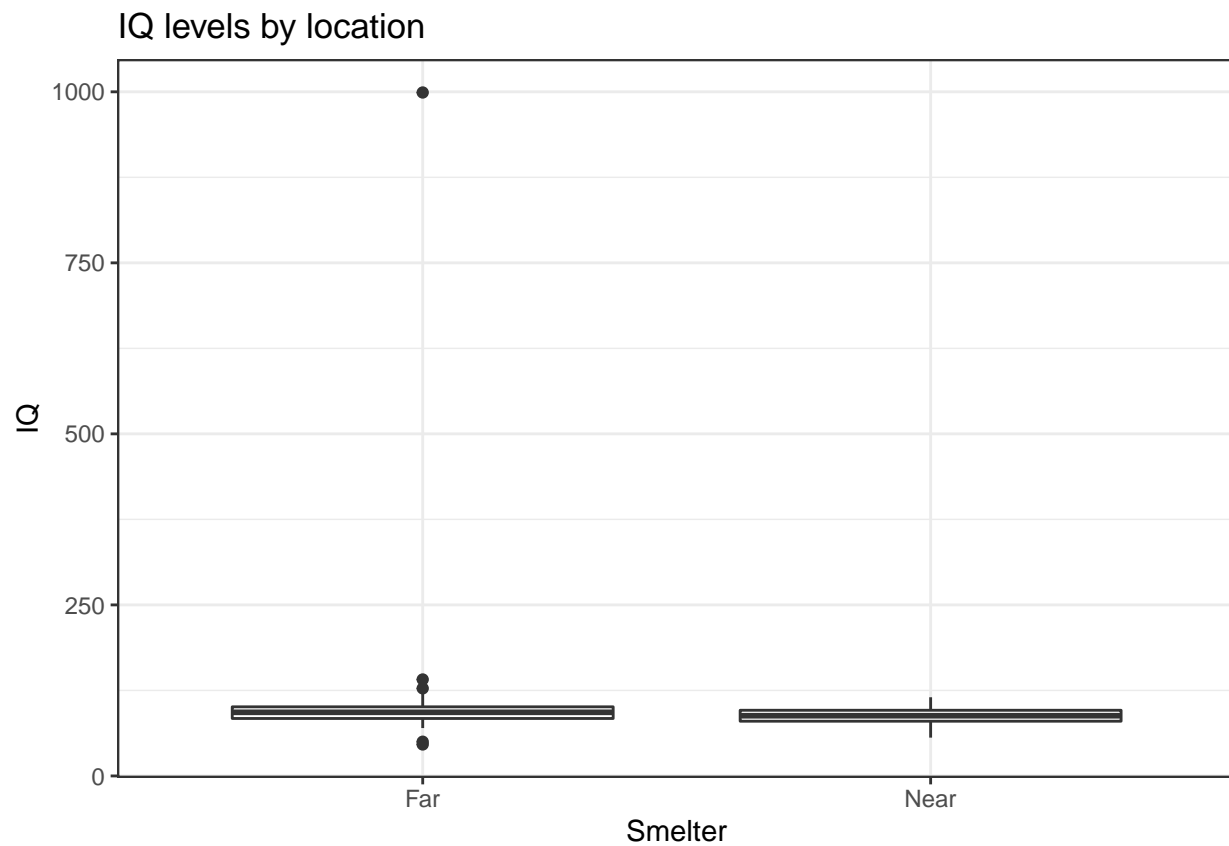# Reproducible Reporting

Arunangshu Sarkar

2022-10-09

Exercise 2

Exercise 2A

Below is a graph showing IQ levels by location status:

```
datatab <- read.csv("../DataRaw/lead-iq-01.csv")
ggplot(datatab, aes(Smelter, IQ)) +
  geom_boxplot() +
  theme_bw() +
  ggtitle("IQ levels by location")
```

Exercise 2B

Below is a formatted table using the `kable()` function

```
knitr::kable(datatab[c(1:8, 68:76), ],
             caption = "A few rows for each location status",
             booktabs = TRUE)
```

Table 1: A few rows for each location status

	Smelter	IQ
1	Far	70
2	Far	85
3	Far	86
4	Far	76
5	Far	96
6	Far	94
7	Far	115
8	Far	97
68	Near	84
69	Near	56
70	Near	77
71	Near	80
72	Near	86
73	Near	88
74	Near	96
75	Near	96
76	Near	107

Exercise 2C

Boxplots: From the box plots, there doesn't seem to be much difference in IQ levels for each location status, though it's hard to interpret the plots due to one extreme outlier present in the "Far" group. Now, the IQ of a person cannot be near 1000, so there seems to be an error when the data was entered.

Table: This nicely formatted table shows a few rows in the data set for each location status.

Exercise 2D

Calculating the means by location status:

```
mean_df <- datatab %>%
  group_by(Smelter) %>%
  summarise(mean = mean(IQ))
```

The mean IQ for the subjects who were near the smelter was 89.19, and for those far from the smelter was 106.12