Lead-IQ Data Analysis Report

Yifan Zhao

2022-10-07

In this study, we are interested in the lead-IQ dataset, which is described in detail in the `Background/DataInfo.md` document.

Let's read in the data first, and take a look at the variables in the dataset.

```
# Read in lead-iq-01.csv
lead_IQ <- read.csv("../DataRaw/lead-iq-01.csv", stringsAsFactors = T)

# Print head and tail of the dataset
kable(head(lead_IQ))
```

Smelter	IQ
Far	70
Far	85
Far	86
Far	76
Far	96
Far	94

```
kable(tail(lead_IQ))
```

	Smelter	IQ
119	Near	95
120	Near	77
121	Near	74
122	Near	96
123	Near	91
124	Near	78

```
# Check dimensions
dim(lead_IQ)
```

```
## [1] 124   2
```

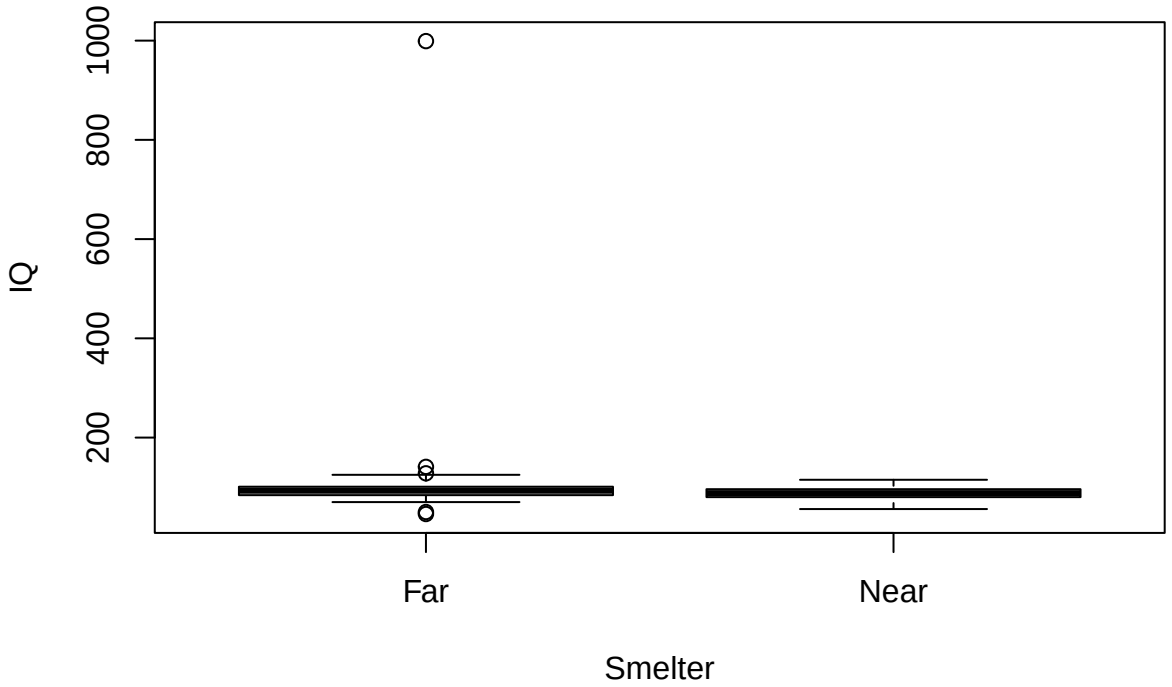
```
# Check variables
summary(lead_IQ)
```

```
##  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
```

It will be great to show the IQ levels by location status in a box plot.

```
boxplot(IQ ~ Smelter, data = lead_IQ)
```



As we can see from the figure above, there is an outlier in the Far group.