Homework 6 Report

Isabelle Meredith

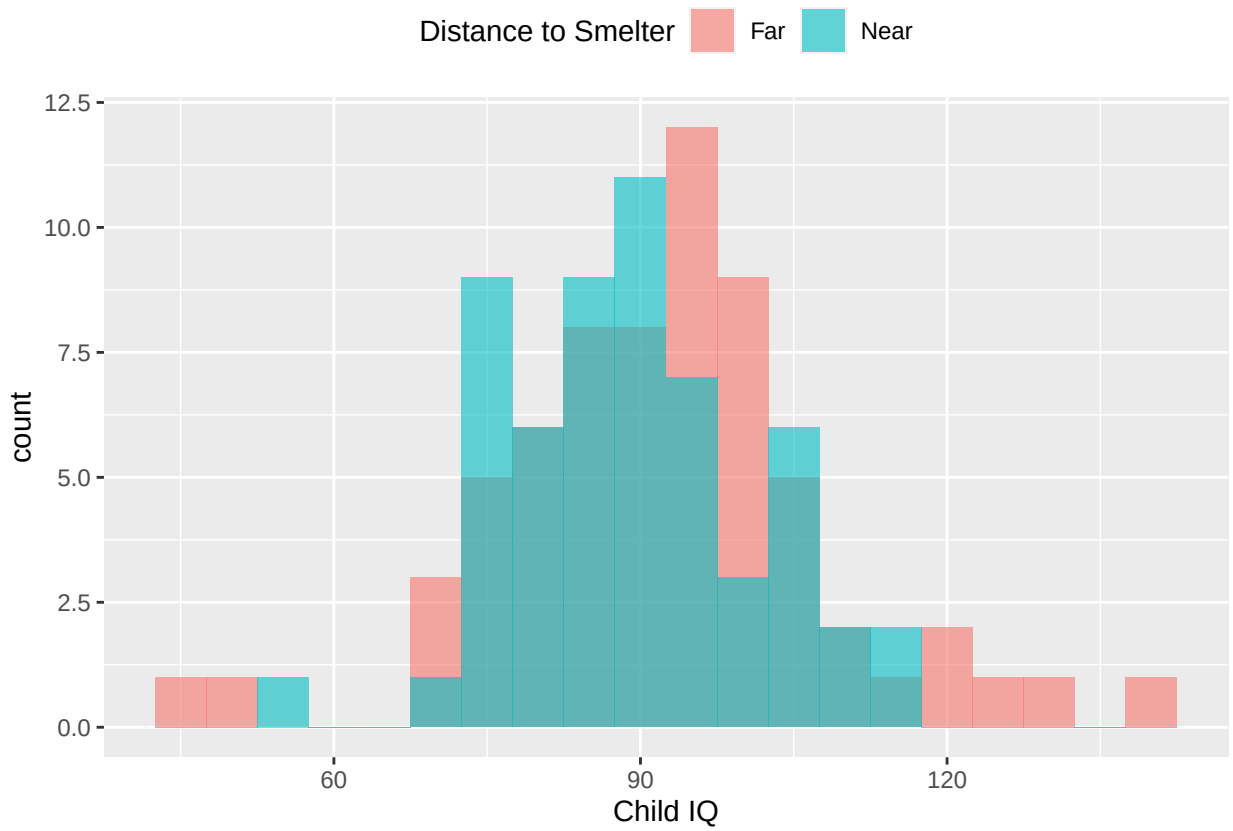
2024-10-08

```
dfLeadRaw <- read.csv("../DataRaw/lead-iq-01.csv")

head(dfLeadRaw)
```

```
##   Smelter IQ
## 1     Far 70
## 2     Far 85
## 3     Far 86
## 4     Far 76
## 5     Far 96
## 6     Far 94
```

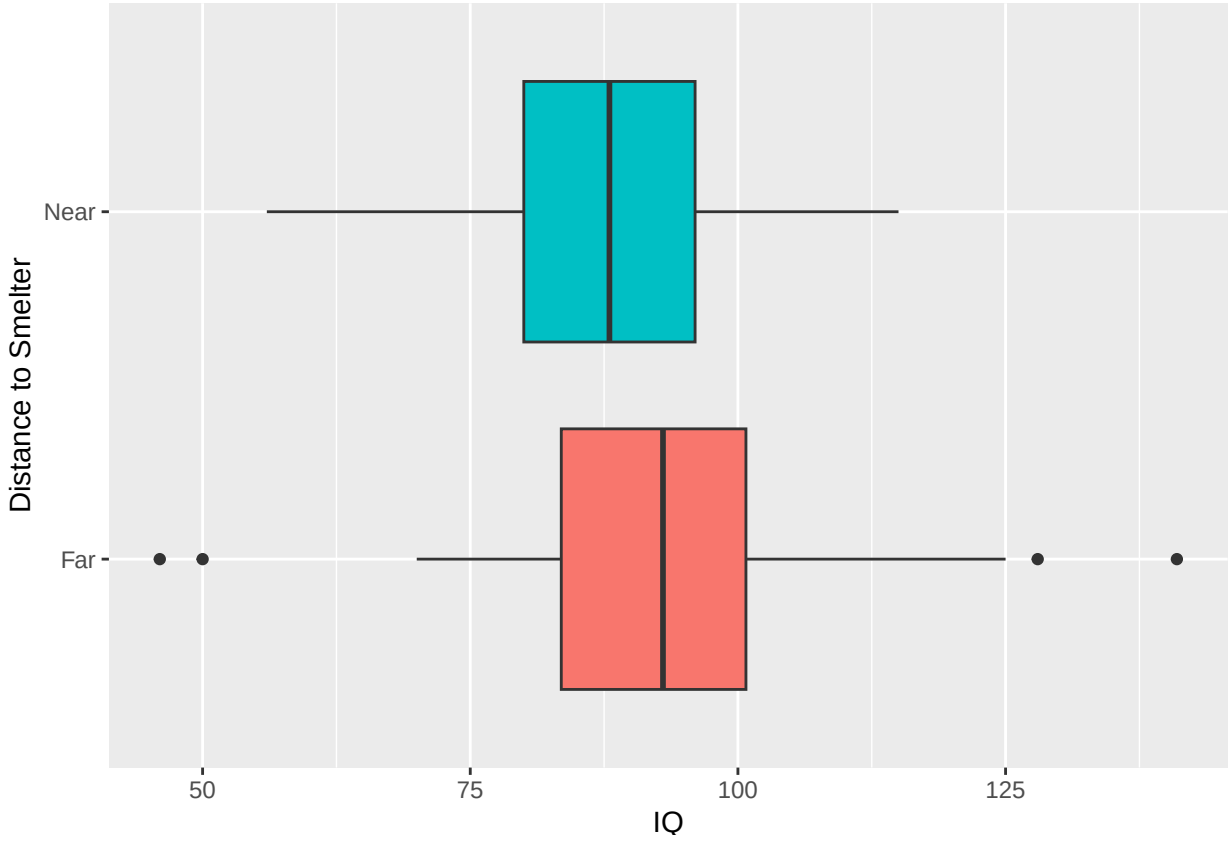
```
# Data processing takes place here.
# Removing an outlier at 999
dfLeadRaw <- dfLeadRaw[(dfLeadRaw$IQ != 999), ]
write.csv(dfLeadRaw, "../DataProcessed/lead-iq-cleaned.csv")

# Read in the processed form of the data instead
dfLead <- read.csv("../DataProcessed/lead-iq-cleaned.csv")
```

```
ggplot(dfLead, aes(IQ, fill = Smelter)) +
   geom_histogram(alpha = 0.6, position = 'identity', binwidth = 5) +
   xlab("Child IQ") +
   guides(fill = guide_legend(title = "Distance to Smelter")) +
   theme(legend.position = "top")
```



Histogram of IQs comparing children who were near to the smelter with children who were far from the smelter.

```
ggplot(dfLead, aes(x=IQ, y = Smelter, fill = Smelter)) +
   geom_boxplot(position = 'identity') +
   ylab("Distance to Smelter") +
   guides(fill="none")
```



Boxplot of IQs comparing children who were near to the smelter with children who were far from the smelter.

```
dfLeadSummary <- dfLead %>%
  group_by(Smelter) %>%
  rename("Smelter Distance" = Smelter) %>%
  summarise("Mean IQ" = mean(IQ), "IQ SD" = sd(IQ), "Group Count" = n())
kable(dfLeadSummary)
```

Smelter Distance	Mean IQ	IQ SD	Group Count
Far	92.59091	16.07767	66
Near	89.19298	12.17497	57

The table contains the values for the group means, standard deviations, and counts for the group near and far from the smelter.

The mean IQ for close to the smelter is 89.1929825 and the mean IQ far from the smelter is 92.5909091.