Lead IQ Report

Sarah Bird

2022-10-09

First we must read in the data.

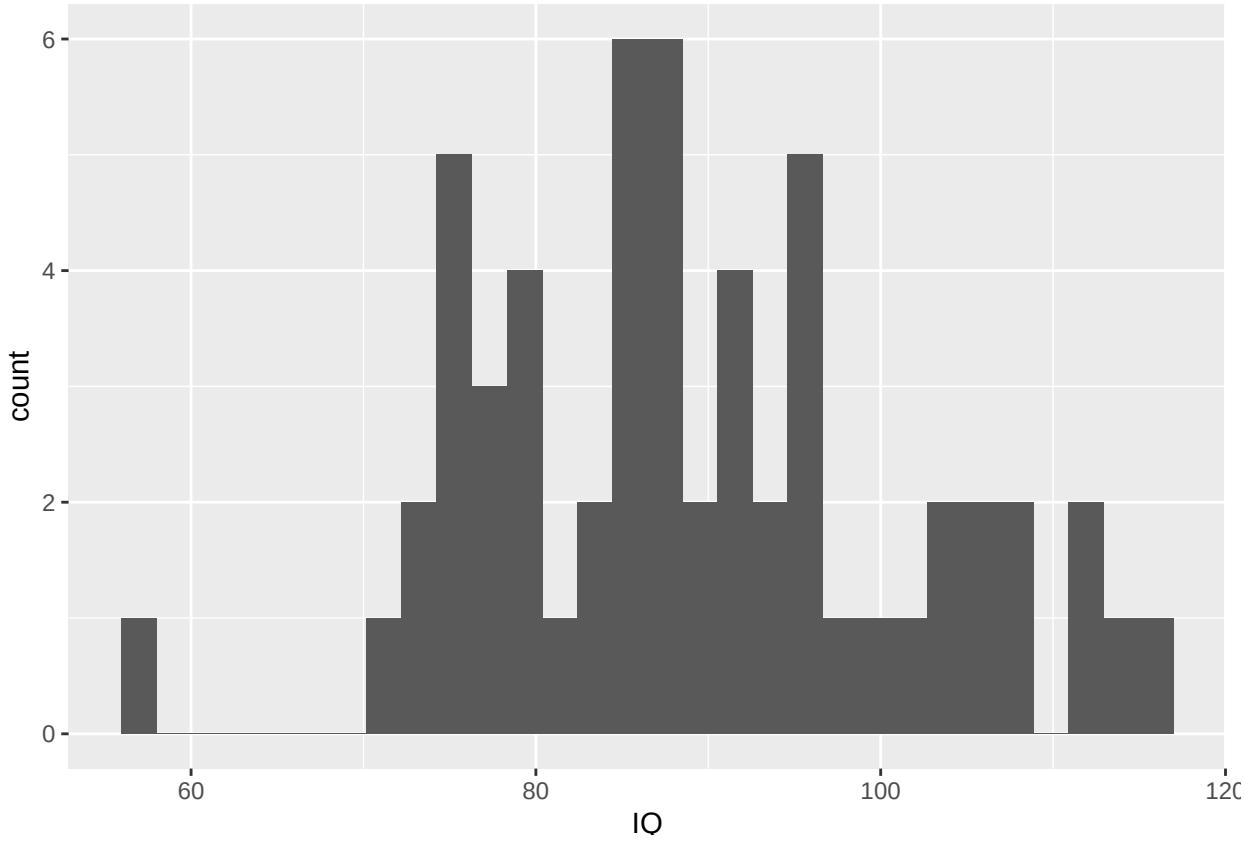
```
data <- read.csv("lead-iq-01.csv")
data[data ==  999] <- 99


dataNear <- data %>% filter(Smelter == "Near")
dataFar <- data %>% filter(Smelter == "Far")
```

Now we can use this data to create a graph showing IQ levels by location status

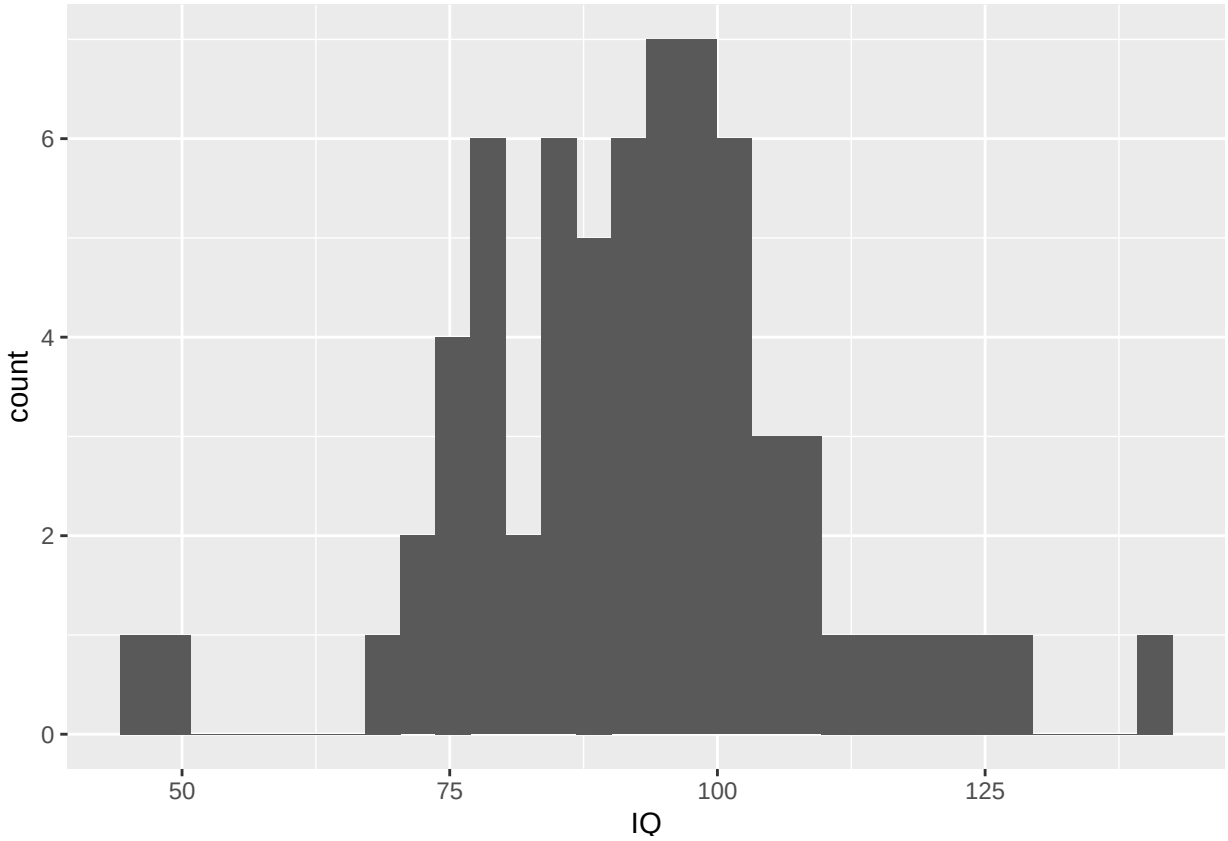
```
ggplot(data=dataNear, aes(x=IQ)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(data=dataFar, aes(x=IQ)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Let's view all of the observations of IQ in the near

```
tableInfo <- matrix( data = c(mean(dataNear$IQ), mean(dataFar$IQ), sd(dataNear$IQ), sd(dataFar$IQ)), nr
                     ncol = 2, byrow = TRUE)
colnames(tableInfo) <- c("Near", "Far")
rownames(tableInfo) <- c("Mean", "SD")

kable(tableInfo)
```

	Near	Far
Mean	89.19298	92.68657
SD	12.17497	15.97461

Thus we can see that the mean of the near variable is 89.193 while the mean of the far variable is 92.687.

We can see that the variation in the far variable is much larger than that of the variation in the near variable. Additionally, we can see that the mean is much higher in the far variable as well.

[UPDATE: we have removed the outlier observation, and now the far observation has more variability but it is a better value.]