Project_01

Diana Felipe

2024-10-03

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

# new lead_iq without IQ 999:
# remove 999 value
lead_iq2 <- lead_iq[-10, ]

lead_iq <- lead_iq2
```

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$IQ) + 30,
                                  by =  ceiling(max(lead_iq$IQ)/ 10 )))
```

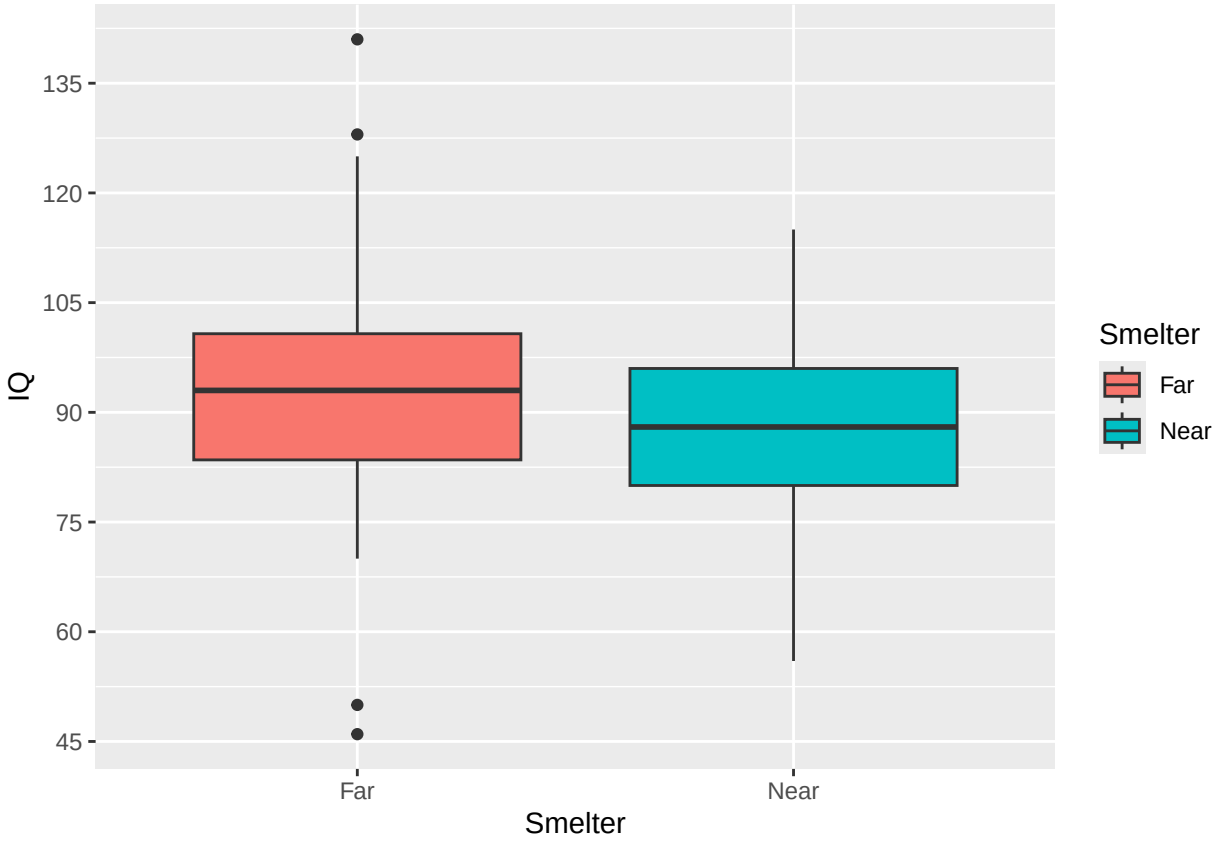
135
120
105
90
75
60
45
IQ
Far
Near
Smelter
Smelter
Far
Near

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 66  1.00  0.00      1   1   1     0  NaN      NaN 0.00
## IQ         2 66 92.59 16.08     93  46 141    95 0.07      1.3 1.98
## ------------------------------------------------------------
## 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 see that the Far group appears to generally have higher IQ's than that of the Near group. The mean of the Far group is visibly higher than that of the Near group. We can also see that both groups have a similar spread, around 10 or so for both groups. A noticeable difference between the two is that the Far group has quit a few significant outliers, both below and above. The near group does not have any outliers. Overall, we can see that there does appear to be an association between distance from the smelter and IQ.
From our table, we can see that the Far group's mean IQ is 92.59 with an SD of 16.08 and median of 93. We also have a max of 141. For the near group, we have that the mean is 89.19, sd is 12.17, and median is 88. From the table values, we arrive at the same conclusion as with the boxplots. Students that live further from the smelter seem to generally have a higher IQ compared to those that live closer to the smelter.

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$Smelter == "Far", 2]), 2) ,
                                 round(mean(x[x$Smelter == "Near", 2]), 2)),
                               collapse = " and ")
```

The mean IQ of the Far and Near groups, respectively, are 92.59 and 89.19.