Reproducible Reporting

Edward Bosko

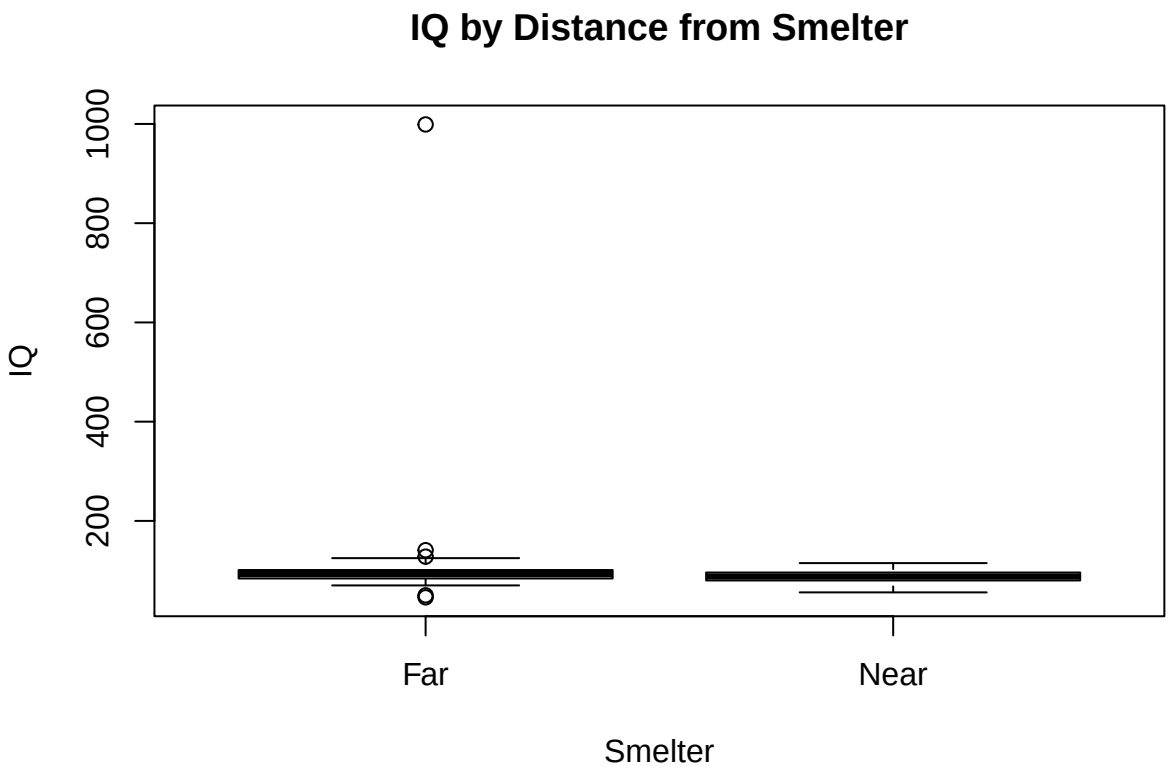
2024-10-07

Reproducible Reporting

```
dat <- read.csv('C:/Users/Eddie/OneDrive/Desktop/BIOSTAT/BIOS 6621/Week 6/lead-iq-01.csv')
```

Make graph

```
boxplot(IQ ~ Smelter, data = dat, main = 'IQ by Distance from Smelter', ylab = 'IQ')
```



Make table

```
mymod <- lm(IQ ~ as.factor(Smelter), data = dat)
summary(mymod)
```

```
##
## Call:
## lm(formula = IQ ~ as.factor(Smelter), data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -60.12 -15.42  -7.16   1.81 892.88
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               106.12      10.10  10.503   <2e-16 ***
## as.factor(Smelter)Near    -16.93      14.90  -1.136    0.258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 82.7 on 122 degrees of freedom
## Multiple R-squared:  0.01046,    Adjusted R-squared:  0.002353
## F-statistic:  1.29 on 1 and 122 DF,  p-value: 0.2582
```

```
df <- data.frame(
  estimate = c(106.12, -16.93),
  SE = c(10.10, 14.90),
  teststat <- c(10.503, -1.136),
  pval <- c('<0.0001', '0.258')
)
colnames(df) = c('Estimate', 'SE','Test Statistic', 'P-Value')

df %>% kable(caption = "Model Output of mymod")
```

Table 1: Model Output of mymod

Estimate	SE	Test Statistic	P-Value
106.12	10.1	10.503	<0.0001
-16.93	14.9	-1.136	0.258

We clearly see from our boxplot that there is a very large outlier in the "Far" smelter group. From our simple linear regression model, we see that children who are within 1 mile of the smelter are expected to have an IQ 16.93 points fewer when compared to children who live further than 1 mile from the smelter, although this difference is not statistically significant (p=0.258).

```
mean(dat$IQ)
```

```
## [1] 98.33871
```