Question 3.50

a. **[1 pt]** All measurement for the state lab are in Table 1.

Table 1. All measurements for the state lab.

```
     lab bod
1  state   6
2  state   6
3  state   8
4  state  11
5  state  18
6  state  20
7  state  28
8  state  33
9  state  34
10 state  43
11 state  71
```

b. **[1 pt]** The eighth BOD measurement for the state lab is 33 mg/L.

c. **[14 pts]** The distribution of BOD values for the private lab is approximately symmetric with no obvious outliers (Figure 1). The mean of the private data is 34.64 mg/L with a standard deviation of 10.45 mg/L (Table 2). The mean and standard deviation were used because of the symmetry of the distribution and absence of outliers. The distribution of BOD values for the state lab appears to be right-skewed with an outlier at 71 mg/L (Figure 1). The median of the state data is 20 mg/L with an IQR from 9.5 to 33.5 mg/L (Table 2). The median and IQR were used because of the presence of the outlier.

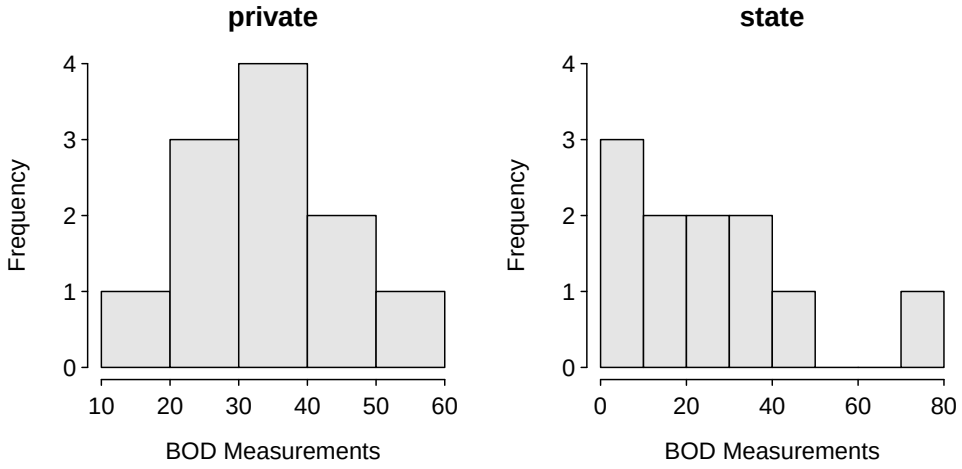

Figure 1. Histograms for BOD measurements for the private (left) and state (right) labs.

Table 2. Descriptive statistics for BOD measurements for the private and state labs.

	lab	n	mean	sd	min	Q1	median	Q3	max	percZero
1	private	11.0	34.6	10.5	15.0	28.5	35.0	40.5	54.0	0.0
2	state	11.0	25.3	19.7	6.0	9.5	20.0	33.5	71.0	0.0

d. **[2 pts]** The two most outstanding differences between the private and state labs is that the state lab BOD values are generally lower (note lower means and medians) and more dispersed (note wider IQR and larger standard deviation).

Question 3.51

a. **[2 pts]** The frequency table is shown in Table 3.

Table 3. Frequency table for perceived choice of agency that disposes of chemical waste. Note that 'o'=Occupational Safety and Health Administration, 'e'=Environmental Protection Agency, 'd'=Department of Transportation, 'n'=National Institutes of Health, and 'u'=unsure.

d	e	n	o	u
9	42	11	25	1

b. **[2 pts]** The percentage table is shown in Table 4.

Table 4. Percentages table for perceived choice of agency that disposes of chemical waste. Abbreviations are given in Table 3.

d	e	n	o	u	Sum
10.2	47.7	12.5	28.4	1.1	100.0

c. **[2 pts]** Most respondents thought that the Environmental Protection Agency was the agency that disposed of chemical wastes followed by the Occupational Safety and Health Administration. Relatively few people thought the Department of Transportation or National Institutes of Health disposed of chemical wastes.

Appendix – R Commands

```
library(NCStats)
setwd('C:/aaaWork/Books/IntroStats/HW/')

d <- read.table("3_50.txt",header=TRUE)
str(d)
( state <- Subset(d,lab=="state") )
state$bod[8]
Summarize(bod~lab,data=d,digits=1)
hist(bod~lab,data=d,xlab="BOD Measurements")

d2 <- read.table("3_51.txt",header=TRUE)
str(d2)
a.tbl <- table(d2$agency)
( p.tbl <- percTable(a.tbl) )
```

Notes From Professor

- Question 3.50c asks for a "univariate EDA" – Note that this does not mean to just make a histogram and summary statistics. Performing a univariate EDA means addressing shape, outliers, center, and dispersion from looking at a histogram (or boxplot) and summary statistics.
- In question 3.50c the shape, outliers, center, and dispersion are specifically listed, even if there was no outliers, for both the private and state labs. Also note that I clearly indicated why I chose to use the mean/sd (because of symmetry and no outliers) or median/IQR (occurrence of outliers).
- It would be reasonable to use the median and IQR for both state and private labs in question 3.50c because of the outlier in the state lab. This would allow a more appropriate comparison between labs.
- In question 3.51c I noted what I considered to be the major characteristics of the results. Your major characteristics may differ from mine. However, you should NOT just re-list all of the information in the tables.
- Note how each table and figure is labeled and referred to in the answers.
- You should use no more than one more decimal place than what was recorded for the quantitative variable and no more than one decimal place for the percentages (or three decimals for the proportions).