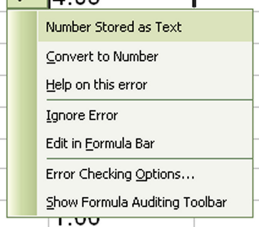
Exploring CPE Data

Don Pereira provided me with a an Excel spreadsheet that contained historic gillnet catch summary information for walleyes captured in Minnesota lakes with a primary lake classification ID of 22. These data can be used to demonstrate a variety of data analysis concepts. I will use this information to demonstrate some basic data entry principles for R and a graphical exploration of the distribution of CPE in these lakes.

1 Data Issues

The original data file that I received had one major and several minor issues. The major issue was that the the last three columns were formatted as text rather than numeric data (see below). Fortunately, this issue is easy to correct by selecting all of the cells treated as text and telling Excel to "Convert to Number"

	A	B	C	D	E	F	G	H
5	SURVEYS DATED PRIOR TO AND INCLUDING 02/28/2008							
6	CE Path: Fisl	Printed on 02	Revision: 08/	Printed at 11:06:51AM				
7	Historic Catcl	Standard gill net sets						
8								
9	Walleye (WAE)							
10	Survey Date	# of Sets	Total No.	Number Per S	Total Weight	Pounds Per S	Mean Weight (lbs)	
11	06/05/2006	24	65	2.71	172.45	7.19	2.65	
12	07/10/2006	16	41	2.56	165.18	10.32	4.03	
13	07/10/2006	15	122	8.13	135.42	9.03		
14	07/10/2006	15	89	5.93	153.94	10.26		
15	07/17/2006	12	92	7.67	87.33	7.28		
16	07/17/2006	15	128	8.53	161.90	10.79		
17	07/24/2006	10	36	3.60	70.07	7.01		
18	07/24/2006	16	22	1.38	33.73	2.11		
19	07/24/2006	15	152	10.13	251.76	16.78		
20	07/24/2006	15	105	7.00	134.93	9.00	1.29	



The minor issues include

1. Ten rows of meta-information that must be ignored by R.
2. Column names that contain spaces and other characters.
3. Column names that are too long.
4. A first column that contains data information (R will treat this numerically unless we explicitly ask it not to).

Fortunately all of these issues can be easily handled when importing the data with `read.xls()` as illustrated below,

```
> library(xlsReadWrite)
> setwd("C://aaaWork//Class Materials//MnDNR_ShortCourse//Readings//ExploreCPE//")
```

```
> wae22 <- read.xls("lakeclass22waeGNCUE_mod.xls", from = 11, colClasses = c("isodate",
+     rep("numeric", 6)), colNames = c("date", "sets", "catch", "cpe",
+     "ttl.wght", "wght.set", "mean.wght"))
> str(wae22)

'data.frame':        39 obs. of  7 variables:
 $ date     : chr  "2006-06-05" "2006-07-10" "2006-07-10" "2006-07-10" ...
 $ sets     : num  24 16 15 15 12 15 10 16 15 15 ...
 $ catch    : num  65 41 122 89 92 128 36 22 152 105 ...
 $ cpe      : num  2.71 2.56 8.13 5.93 7.67 ...
 $ ttl.wght : num  172.4 165.2 135.4 153.9  87.3 ...
 $ wght.set : num   7.19 10.32  9.03 10.26  7.28 ...
 $ mean.wght: num  2.65 4.03 1.11 1.73 0.95 1.26 1.95 1.53 1.66 1.29 ...

> rhead(wae22)

         date sets catch   cpe ttl.wght wght.set mean.wght
35 2007-08-06   15   111  7.40   174.93    11.66      1.60
24 2007-06-18   12   103  8.58   106.98     8.91      1.04
33 2007-07-30   14   132  9.43   181.86    12.99      1.38
20 2006-08-28   11   260 23.64   367.32    33.39      1.41
25 2007-07-01   20   209 10.45   222.61    11.13      1.07
14 2006-08-07   12   122 10.17   251.78    20.98      2.06
```

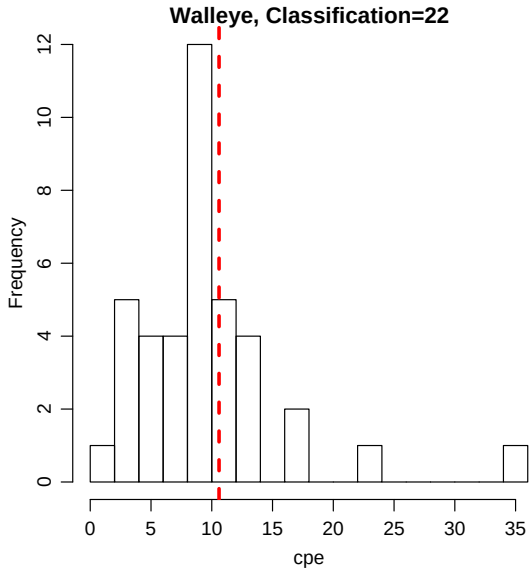
2 CPE Comparison Graph

Suppose we want a graphic to show how "our lake" with a CPE=10.6 compares to the historical catches.

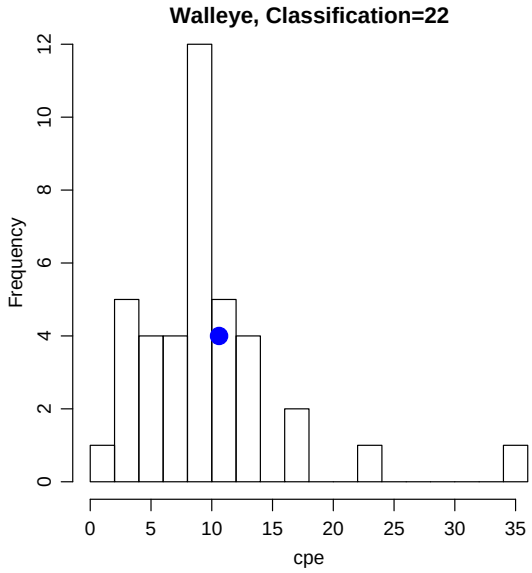
```
> attach(wae22)
> my.cpe <- 10.6
```

2.1 Histograms

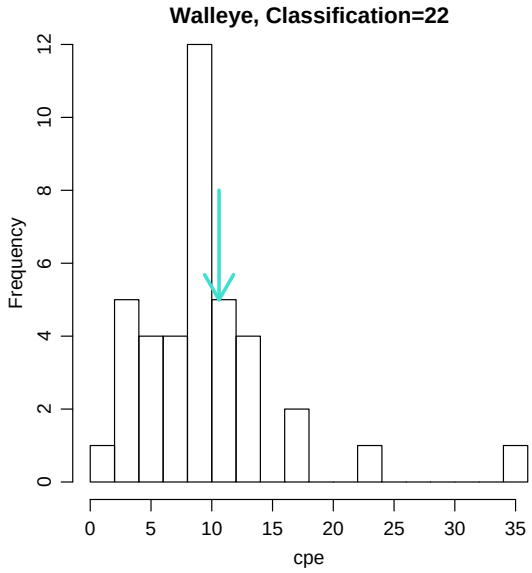
```
> hist(cpe, main = "Walleye, Classification=22", breaks = 20)
> abline(v = my.cpe, col = "red", lty = 2, lwd = 3)
```




```
> hist(cpe, main = "Walleye, Classification=22", breaks = 20)
> points(my.cpe, 4, col = "blue", pch = 19, cex = 2)
```

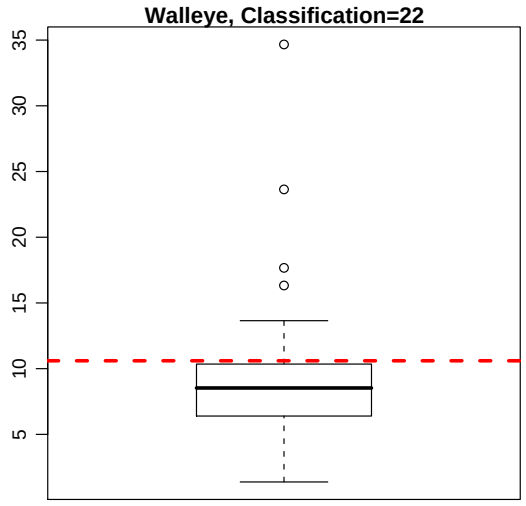



```
> hist(cpe, main = "Walleye, Classification=22", breaks = 20)
> arrows(my.cpe, 8, my.cpe, 5, col = "turquoise", lwd = 3)
```

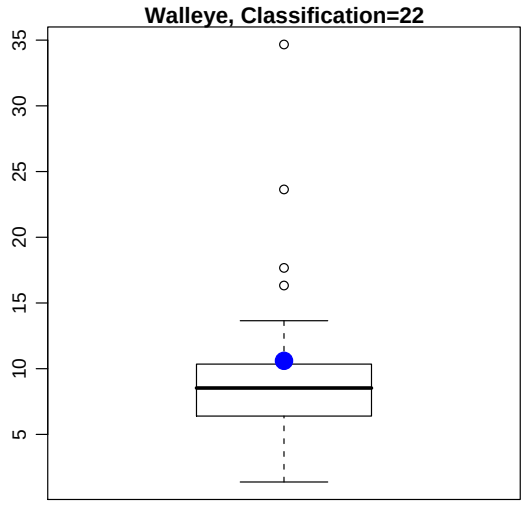


2.2 Boxplots

```
> boxplot(cpe, main = "Walleye, Classification=22")
> abline(h = my.cpe, col = "red", lty = 2, lwd = 3)
```

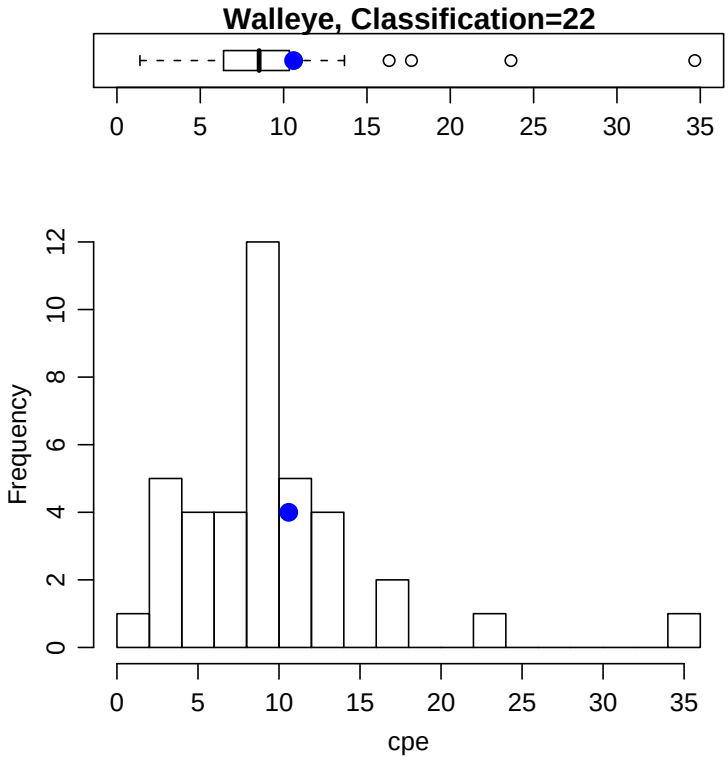



```
> boxplot(cpe, main = "Walleye, Classification=22")
> points(1, my.cpe, col = "blue", pch = 19, cex = 2)
```



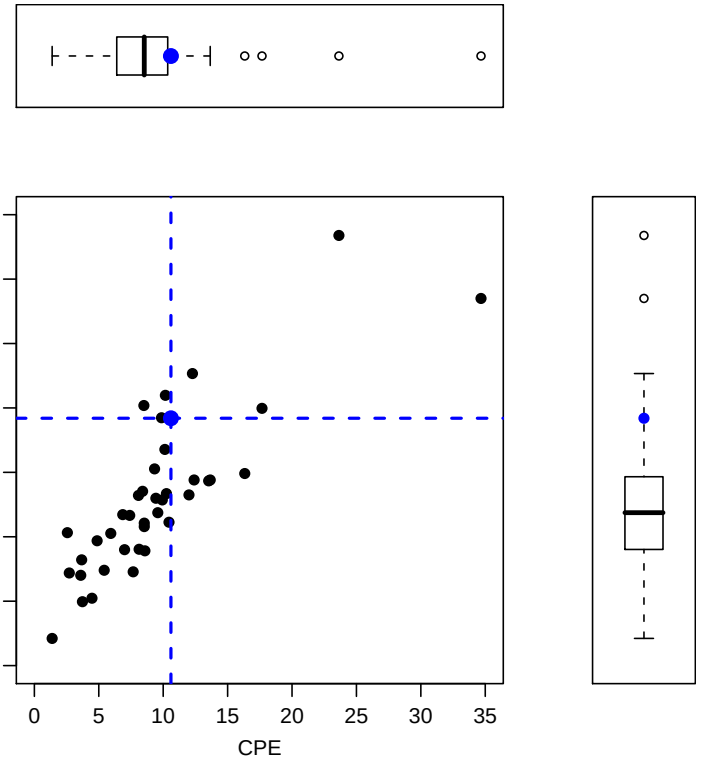
2.3 Combined

```
> layout(matrix(c(2, 1), 2, 1), heights = c(25, 75))
> hist(cpe, main = "", breaks = 20)
> points(my.cpe, 4, col = "blue", pch = 19, cex = 1.5)
> boxplot(cpe, main = "Walleye, Classification=22", horizontal = TRUE,
+     ylim = c(0, 35))
> points(my.cpe, 1, col = "blue", pch = 19, cex = 1.5)
```



Suppose that my lake also had an average weight per set of 19.2.

```
> my.wght.set <- 19.2
> layout(matrix(c(2, 0, 1, 3), 2, 2, byrow = TRUE), heights = c(25, 75),
+     widths = c(75, 25))
> par(mar = c(3.5, 3.5, 0, 0))
> plot(wght.set ~ cpe, pch = 19, xlab = "CPE", ylab = "Weight per Set",
+     xlim = c(0, 35), ylim = c(0, 35))
> abline(v = my.cpe, col = "blue", lwd = 2, lty = 2)
> abline(h = my.wght.set, col = "blue", lwd = 2, lty = 2)
> points(my.cpe, my.wght.set, col = "blue", pch = 19, cex = 1.5)
> boxplot(cpe, horizontal = TRUE, ylim = c(0, 35), xaxt = "n")
> points(my.cpe, 1, col = "blue", pch = 19, cex = 1.5)
> boxplot(wght.set, ylim = c(0, 35), yaxt = "n")
> points(1, my.wght.set, col = "blue", pch = 19)
```



2.4 As a Function

```
> comp.wae22 <- function(my.cpe, my.wght.set, col = "blue") {
+     layout(matrix(c(2, 0, 1, 3), 2, 2, byrow = TRUE), heights = c(25,
+         75), widths = c(75, 25))
+     par(mar = c(3.5, 3.5, 0, 0))
+     plot(wght.set ~ cpe, pch = 19, xlab = "CPE", ylab = "Weight per Set",
+         xlim = c(0, 35), ylim = c(0, 35))
+     abline(v = my.cpe, col = col, lwd = 2, lty = 2)
+     abline(h = my.wght.set, col = col, lwd = 2, lty = 2)
+     points(my.cpe, my.wght.set, col = col, pch = 19, cex = 1.5)
+     boxplot(cpe, horizontal = TRUE, ylim = c(0, 35), xaxt = "n")
+     points(my.cpe, 1, col = col, pch = 19, cex = 1.5)
+     boxplot(wght.set, ylim = c(0, 35), yaxt = "n")
+     points(1, my.wght.set, col = col, pch = 19)
+ }
```

Now try an example lake with a cpe of 3 and weight set of 20.

```
> comp.wae22(3, 20, "red")
```