DaigleInClassLabWk10D2.R

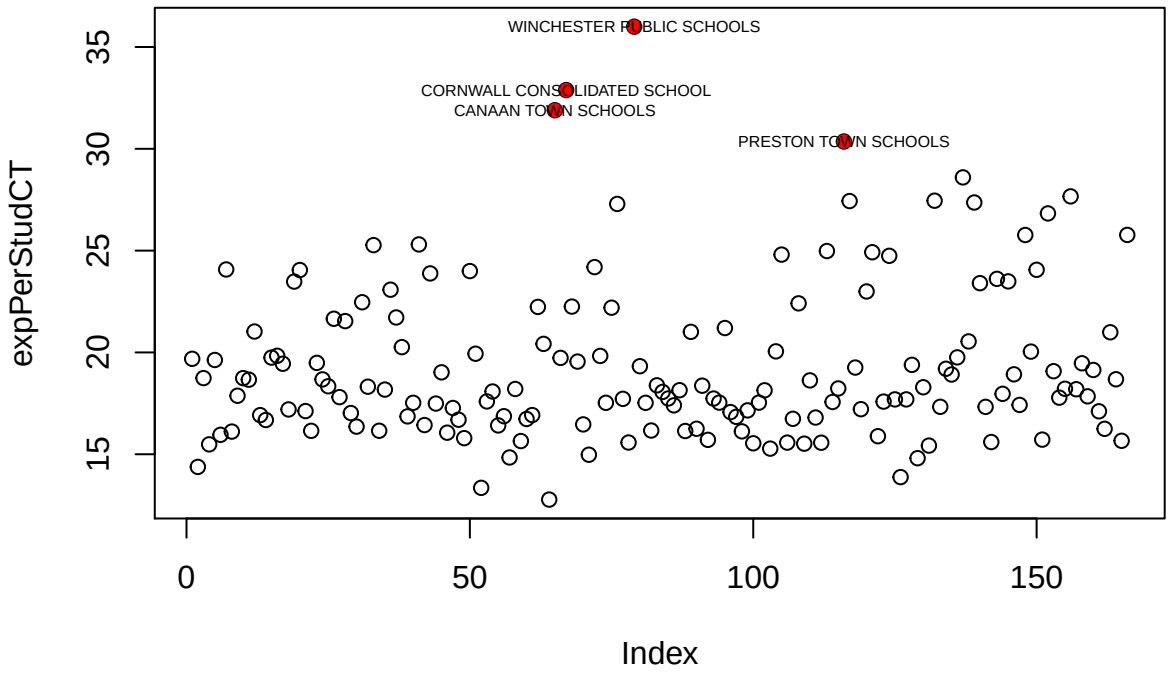
daiglechris

Wed Mar 28 23:23:37 2018

```
# Christopher Daigle
# Week10D2 In Class Lab - Histogram

# Exercise

setwd("/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/Da
dataSet <- read.csv("histogram.csv", stringsAsFactors = FALSE)
dataSet <- dataSet[dataSet$YRDATA == 2013, ]

# Using plot() and points() functions, highlight the 4 best schools in CT in terms of expenditure per s
head(dataSet)
```

```
##     STATE ENROLL                          NAME YRDATA TOTALREV TFEDREV
## 1 Alabama   9717 AUTAUGA COUNTY SCHOOL DISTRICT   2013    79651    7259
## 2 Alabama  29419 BALDWIN COUNTY SCHOOL DISTRICT   2013   294249   22266
## 3 Alabama   1061 BARBOUR COUNTY SCHOOL DISTRICT   2013    10444    2650
## 4 Alabama   2805   EUFAULA CITY SCHOOL DISTRICT   2013    24320    3588
## 5 Alabama   3475    BIBB COUNTY SCHOOL DISTRICT   2013    31799    3909
## 6 Alabama   8341  BLOUNT COUNTY SCHOOL DISTRICT   2013    67328    6812
##   TSTREV TLOCREV TOTALEXP TCURINST TCURSSVC TCURONON TCAPOUT
## 1  51182   21210    77028    42480    21483     6996    4563
## 2 122616  149367   264079   139820    85770    17963   11220
## 3   5850    1944    10628     5318     3858     1223       0
## 4  14592    6140    24630    13974     7149     2276    1192
## 5  21660    6230    30276    16113     9405     3184     725
## 6  47058   13458    65016    36785    21422     5487     744
```

```
onlyPosEnrCT <- dataSet[dataSet$STATE == "Connecticut" & dataSet$ENROLL > 0, ]
expPerStudCT <- onlyPosEnrCT$TOTALEXP / onlyPosEnrCT$ENROLL
greaterThan30 <- which(expPerStudCT > 30)

plot(expPerStudCT)
points(greaterThan30, expPerStudCT[greaterThan30], pch=16, col="red")
text(greaterThan30, expPerStudCT[greaterThan30], labels = onlyPosEnrCT$NAME[greaterThan30], cex = 0.5)
```




```
# Make two histograms together for expenditure per student (TOTALEXP / ENROLL) of Connecticut and Alaba

onlyPosEnrAB <- dataSet[dataSet$STATE == "Alabama" & dataSet$ENROLL > 0, ]
expPerStudAB <- onlyPosEnrAB$TOTALEXP / onlyPosEnrAB$ENROLL

summary(expPerStudCT)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.77   16.84   18.20   19.38   21.01   36.00
```

```
sd(expPerStudCT)
```

```
## [1] 3.911911
```

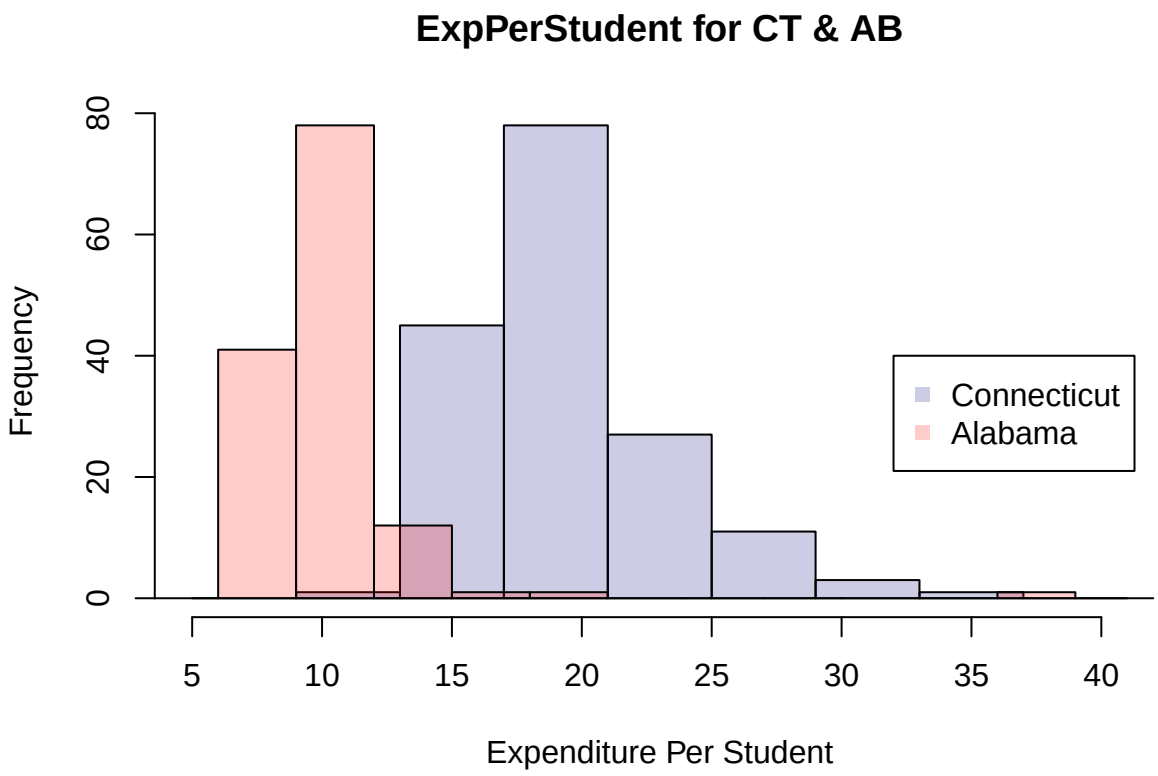
```
hist(expPerStudCT, col = rgb(0,0,0.5, 0.2), freq = TRUE, breaks = seq(5, 44, by = 4), xlab = "Expenditu

summary(expPerStudAB)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.362   8.845   9.620  10.229  10.654  36.898
```

```
sd(expPerStudAB)
```

```
## [1] 2.911242
```

```
hist(expPerStudAB, col = rgb(1,0,0,.2), freq = TRUE, breaks = seq(0, 44, by = 3), add = TRUE)
legend(x = 32, y = 40, c("Connecticut", "Alabama"), col = c(rgb(0,0,0.5, 0.2), rgb(1,0,0,.2)), pch = 15
```




```
# In comparison, AB spends less per student than CT does, more often (as measured by the frequency of
# occurence), but there is at least one outlier where AB spends more.

# graphics.off()
#
# hist(expPerStudCT, col = rgb(0,0,0.5, 0.2), freq = FALSE, breaks = seq(5, 44, by = 4), xlab = "Expend
# hist(expPerStudAB, col = rgb(1,0,0,.2), freq = FALSE, breaks = seq(0, 44, by = 3), add = TRUE)
# legend(x = 32, y = 40, c("Connecticut", "Alabama"), col = c(rgb(0,0,0.5, 0.2), rgb(1,0,0,.2)), pch =
```