King Sejong's National Poll on Tax Reform

coop711

2015년 3월 29일

Data Management

- Original data came from intenet version of Sejong silok, summarized by Oh, Ki-Soo.

```
sejong.poll<-read.table("sejong_poll.txt",header=TRUE,sep="")
str(sejong.poll)
```

```
## 'data.frame':    44 obs. of  4 variables:
##  $ counts: int  21 194 259 393 443 117 1123 71 29 5 ...
##  $ vote  : chr  "yes" "no" "yes" "no" ...
##  $ class : chr  "high" "high" "third.current" "third.current" ...
##  $ region: chr  "Seoul" "Seoul" "Seoul" "Seoul" ...
```

```
sejong.poll
```

```
##    counts vote         class      region
## 1      21  yes          high       Seoul
## 2     194   no          high       Seoul
## 3     259  yes third.current       Seoul
## 4     393   no third.current       Seoul
## 5     443  yes      third.ex       Seoul
## 6     117   no      third.ex       Seoul
## 7    1123  yes      ordinary        yuhu
## 8      71   no      ordinary        yuhu
## 9      29  yes         chief     gyunggi
## 10      5   no         chief     gyunggi
## 11  17076  yes      ordinary     gyunggi
## 12    236   no      ordinary     gyunggi
## 13      1   no          high     pyungan
## 14      6  yes         chief     pyungan
## 15     35   no         chief     pyungan
## 16   1326  yes      ordinary     pyungan
## 17  28474   no      ordinary     pyungan
## 18     17  yes         chief    hwanghae
## 19     17   no         chief    hwanghae
## 20   4454  yes      ordinary    hwanghae
## 21  15601   no      ordinary    hwanghae
## 22      2   no          high chungcheong
## 23     35  yes         chief chungcheong
## 24     26   no         chief chungcheong
## 25   6982  yes      ordinary chungcheong
## 26  14013   no      ordinary chungcheong
## 27      5  yes         chief     kangwon
## 28     10   no         chief     kangwon
## 29    939  yes      ordinary     kangwon
## 30   6888   no      ordinary     kangwon
## 31      1   no          high      hamgil
## 32      3  yes         chief      hamgil
## 33     14   no         chief      hamgil
## 34     75  yes      ordinary      hamgil
## 35   7387   no      ordinary      hamgil
## 36     55  yes         chief   gyungsang
## 37     16   no         chief   gyungsang
## 38  36262  yes      ordinary   gyungsang
## 39    377   no      ordinary   gyungsang
## 40      2   no          high      jeolla
## 41     42  yes         chief      jeolla
## 42     12   no         chief      jeolla
## 43  29505  yes      ordinary      jeolla
## 44    257   no      ordinary      jeolla
```

- We need vote, class, region as factors. If you leave them as chr, it will be coerced to factor when you tabulate it according to alphabetical order, which is not what you want. So, use factor() to convert them. First, make a working copy vesion of sejong.poll

```
sejong.poll.2<-sejong.poll
sejong.poll.2$vote<-factor(sejong.poll.2$vote, levels=c("yes","no"), label
s=c("yes","no"))
str(sejong.poll.2)
```

```
## 'data.frame':    44 obs. of  4 variables:
##  $ counts: int  21 194 259 393 443 117 1123 71 29 5 ...
##  $ vote  : Factor w/ 2 levels "yes","no": 1 2 1 2 1 2 1 2 1 2 ...
##  $ class : chr  "high" "high" "third.current" "third.current" ...
##  $ region: chr  "Seoul" "Seoul" "Seoul" "Seoul" ...
```

- You can check that “labels=” is not necessary if same as levels. Continue with class and region.

```
sejong.poll.2$class<-factor(sejong.poll.2$class, levels=c("high","third.curren
t", "third.ex", "chief", "ordinary"), labels=c("High","3rd.current", "3rd.forme
r", "Chief", "Commons"))
sejong.poll.2$region<-factor(sejong.poll.2$region, levels=c("Seoul","yuhu", "gy
unggi", "pyungan", "hwanghae", "chungcheong", "kangwon", "hamgil", "gyungsang",
"jeolla"), labels=c("Seoul","Yuhu", "Gyunggi", "Pyungan", "Hwanghae", "Chungche
ong", "Kangwon", "Hamgil", "Gyungsang", "Jeolla"))
str(sejong.poll.2)
```

```
## 'data.frame':    44 obs. of  4 variables:
##  $ counts: int  21 194 259 393 443 117 1123 71 29 5 ...
##  $ vote  : Factor w/ 2 levels "yes","no": 1 2 1 2 1 2 1 2 1 2 ...
##  $ class : Factor w/ 5 levels "High","3rd.current",..: 1 1 2 2 3 3 5 5 4 4
...
##  $ region: Factor w/ 10 levels "Seoul","Yuhu",..: 1 1 1 1 1 1 2 2 3 3 ...
```

- We add color for the vote.

```
sejong.poll.2$color[sejong.poll.2$vote=="yes"]<-"cyan"
sejong.poll.2$color[sejong.poll.2$vote=="no"]<-"red"
```

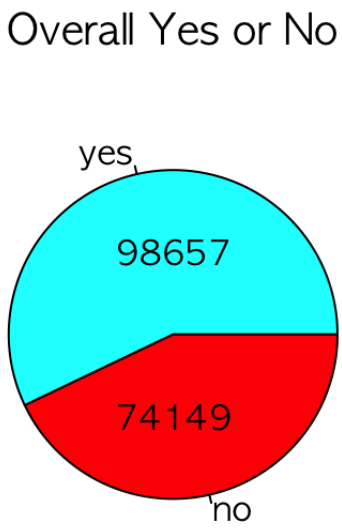
- Check the total vote with xtabs()

```
xtabs(counts~vote, data=sejong.poll.2)
```

```
## vote
##   yes    no
## 98657 74149
```

- We can check the color. Coordinates of text() are found by locator(2). Try!

```
pie(xtabs(counts~vote, data=sejong.poll.2), col=sejong.poll.2$color)
title(main="Overall Yes or No")
text(x=0, y=c(0.4,-0.4), labels=c("98657", "74149"))
```



- Vote by class

```
xtabs(counts~vote+class, data=sejong.poll.2)
```

```
##      class
## vote   High 3rd.current 3rd.former Chief Commons
##   yes    21         259         443   192   97742
##   no    200         393         117   135   73304
```

- We need to analyse Commons separately.

```
sejong.poll.2$class.2<-ifelse(sejong.poll.2$class=="Commons", "Commons", "Burea
us")
```

- Compare the votes by class.2, (Bureaucrats vs Commons)

```
xtabs(counts~vote+class.2, data=sejong.poll.2)
```

```
##      class.2
## vote  Bureaus Commons
##   yes     915   97742
##   no      845   73304
```

- Add subtotals to the margins,

```
addmargins(xtabs(counts~vote+class.2, data=sejong.poll.2))
```

```
##      class.2
## vote  Bureaus Commons    Sum
##   yes     915   97742  98657
##   no      845   73304  74149
##   Sum    1760  171046 172806
```

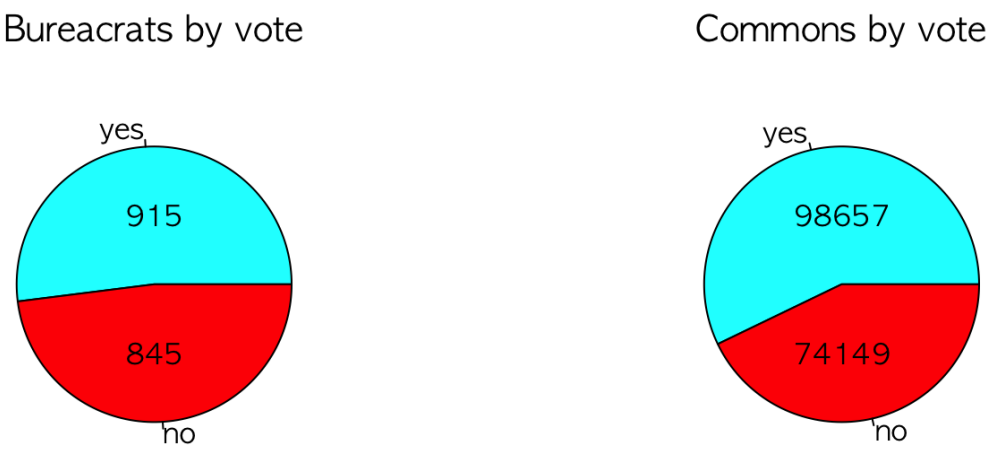
- Compute the marginal proportions. Note the use of digits=3.

```
options(digits=3)
prop.table(xtabs(counts~vote+class.2, data=sejong.poll.2), margin=2)
```

```
##      class.2
## vote  Bureaus Commons
##   yes   0.520   0.571
##   no    0.480   0.429
```

- Pie charts for Bureacrats by vote and Commons by vote.

```
attach(sejong.poll.2)
par(mfrow=c(1,2))
pie(xtabs(counts~vote+class.2, data=sejong.poll.2[class.2=="Bureaus",], dro
p=T), labels=c("yes", "no"), col=color)
title(main="Bureacrats by vote")
text(x=0, y=c(0.4,-0.4), labels=c("915", "845"))
pie(xtabs(counts~vote+class.2, data=sejong.poll.2[class.2=="Commons",], dro
p=T), labels=c("yes", "no"), col=color)
title(main="Commons by vote")
text(x=0, y=c(0.4,-0.4), labels=c("98657", "74149"))
```




```
par(mfrow=c(1,1))
```

- Count the vote by region class.2 wise.

```
xtabs(counts~vote+region, data=sejong.poll.2[class.2=="Bureaus",], drop=T)
```

```
##      region
## vote  Seoul Gyunggi Pyungan Hwanghae Chungcheong Kangwon Hamgil Gyungsang
##   yes   723      29       6       17           35       5      3         55
##   no    704       5      36       17           28      10     15         16
##      region
## vote  Jeolla
##   yes     42
##   no      14
```

```
xtabs(counts~vote+region, data=sejong.poll.2[class.2=="Commons",], drop=T)
```

```
##      region
## vote   Yuhu Gyunggi Pyungan Hwanghae Chungcheong Kangwon Hamgil Gyungsang
##   yes  1123   17076    1326     4454        6982     939     75     36262
##   no     71     236   28474    15601       14013    6888   7387       377
##      region
## vote  Jeolla
##   yes  29505
##   no     257
```

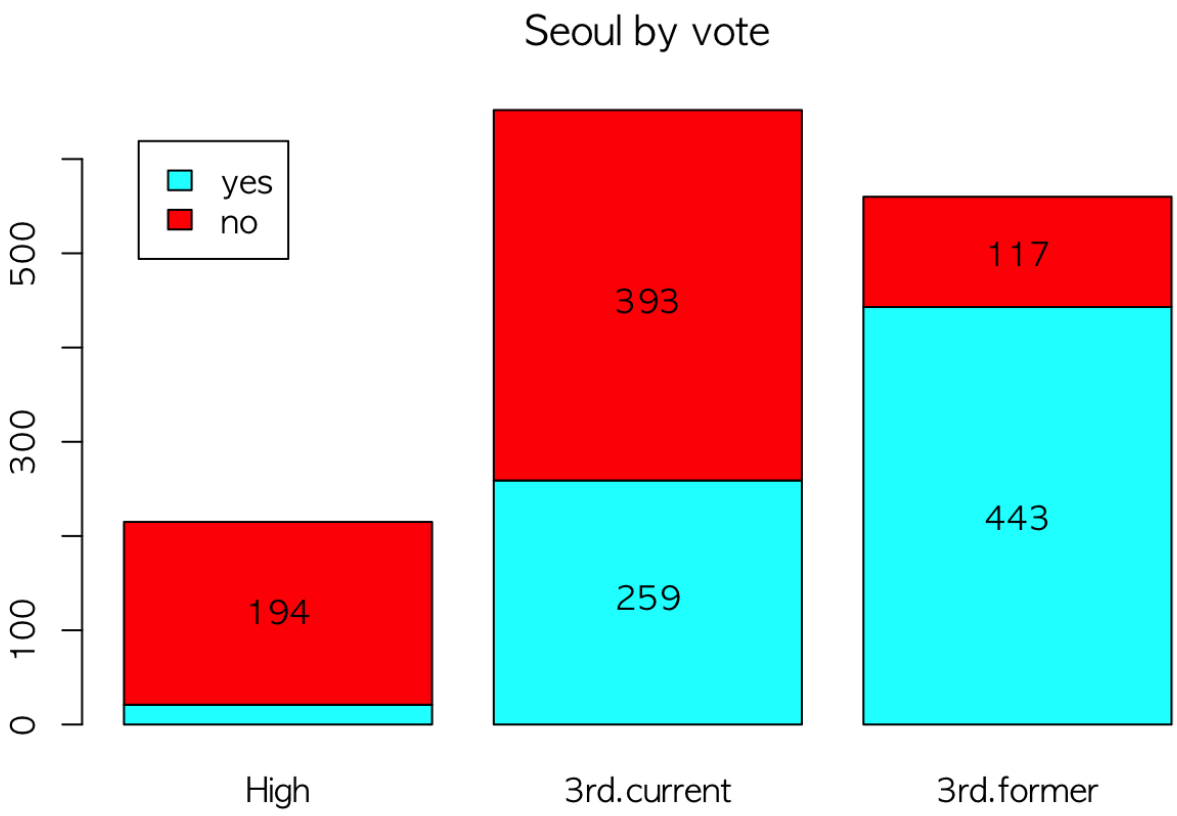
- Seoul has three times more Bureaucrats than other regions, so analyse further.

```
xtabs(counts~vote+class, data=sejong.poll.2[region=="Seoul",], drop=T)
```

```
##      class
## vote  High 3rd.current 3rd.former
##   yes   21         259        443
##   no   194         393        117
```

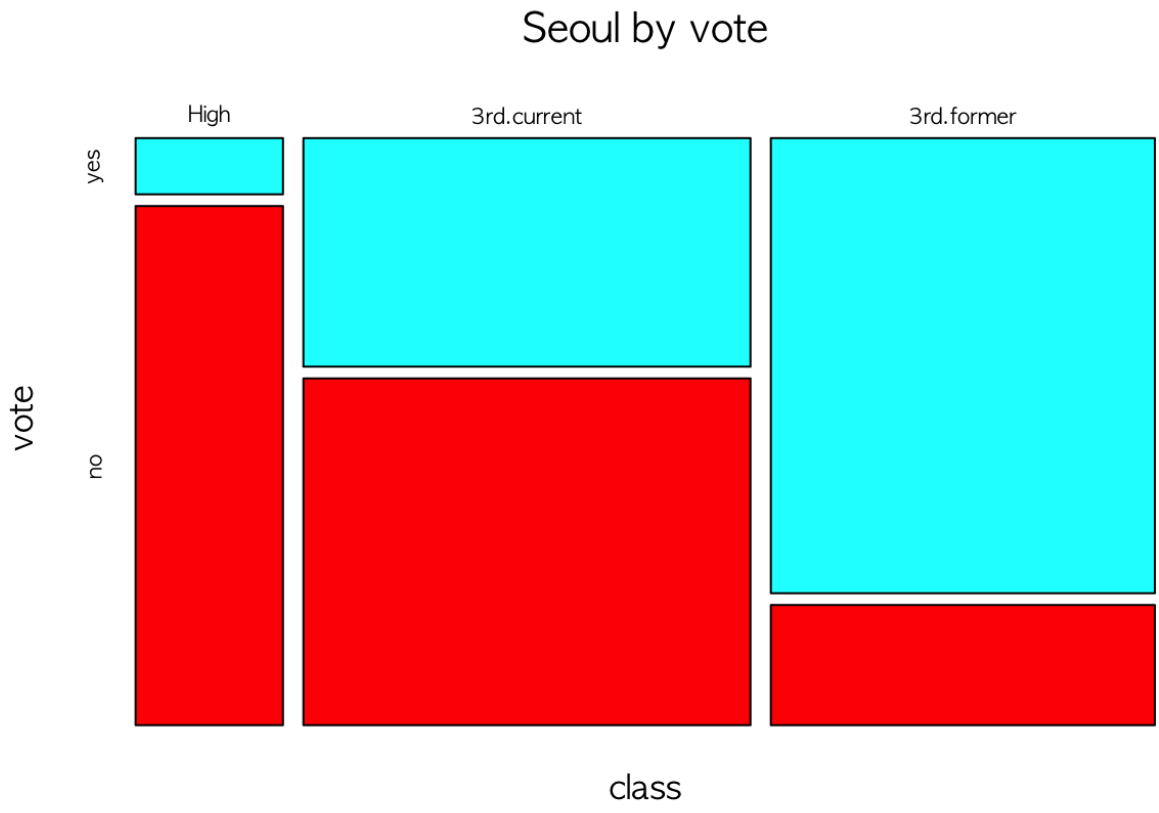
- Draw barplot for the vote by class in Seoul. Text positions were obtained by locator().

```
barplot(xtabs(counts~vote+class, data=sejong.poll.2[region=="Seoul",], drop=T),
col=color)
title(main="Seoul by vote")
text(x=c(0.7, 1.9, 1.9, 3.1, 3.1), y=c(120, 450, 135, 500, 220), labels=c("19
4","393", "259", "117", "443"))
legend("topleft", inset=0.05, fill=c("cyan", "red"), legend=c("yes", "no"))
```



- Using mosaicplot()

```
mosaicplot(xtabs(counts~class+vote, data=sejong.poll.2[region=="Seoul",], drop=T), col=color, main="Seoul by vote")
```

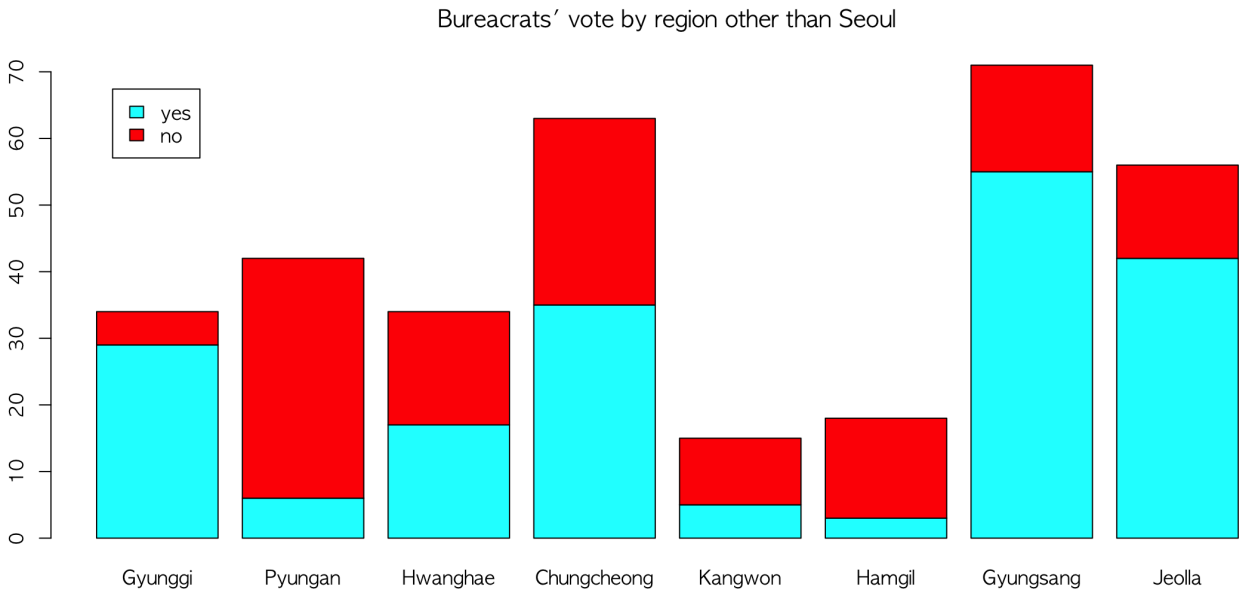


- Draw barplot() for the Bureaus by region.

```
xtabs(counts~vote+region, data=sejong.poll.2[class.2=="Bureaus" & !region=="Seo
ul",], drop=T)
```

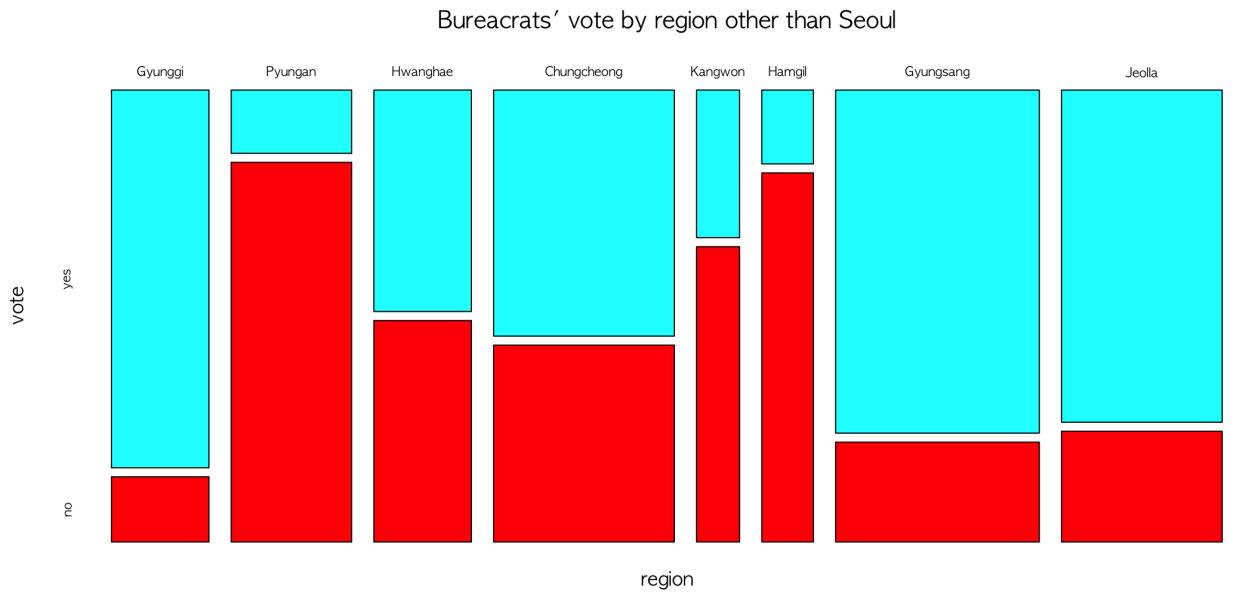
```
##      region
## vote  Gyunggi Pyungan Hwanghae Chungcheong Kangwon Hamgil Gyungsang Jeolla
##   yes      29       6       17          35       5      3         55     42
##   no        5      36       17          28      10     15         16     14
```

```
barplot(xtabs(counts~vote+region, data=sejong.poll.2[class.2=="Bureaus" & !regi
on=="Seoul",], drop=T), col=color)
title(main="Bureacrats' vote by region other than Seoul")
legend("topleft", inset=0.05, fill=c("cyan", "red"), legend=c("yes", "no"))
```



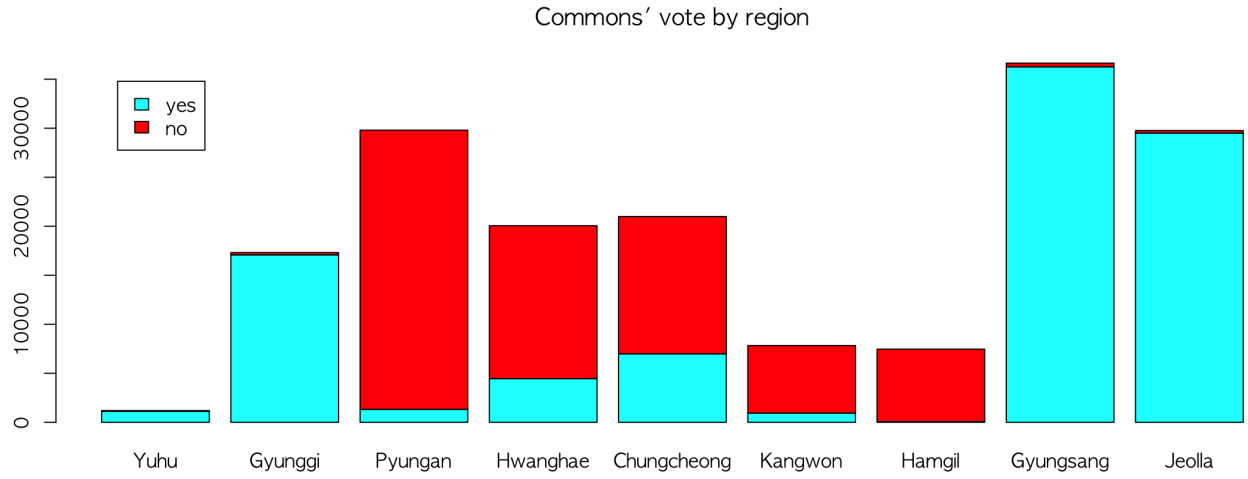
- Using mosaicplot()

```
mosaicplot(xtabs(counts~region+vote, data=sejong.poll.2[class.2=="Bureaus" & !r
egion=="Seoul",], drop=T), col=color, main="")
title(main="Bureacrats' vote by region other than Seoul")
```



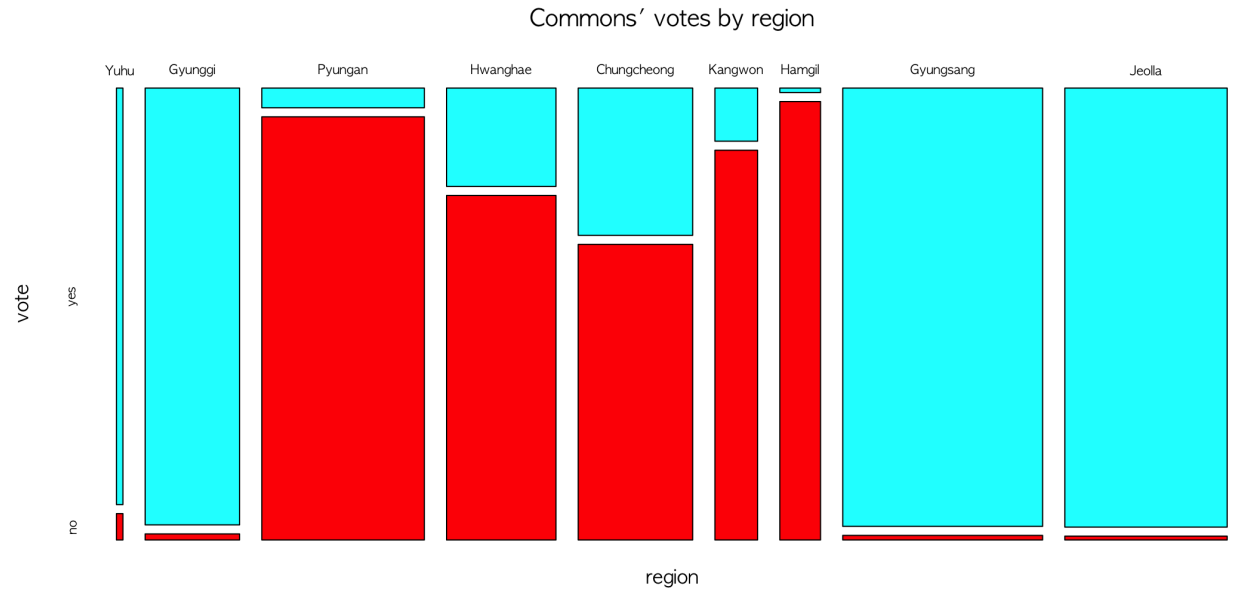
- Draw barplot() for the Commons by region.

```
barplot(xtabs(counts~vote+region, data=sejong.poll.2[class.2=="Commons",], dro
p=T), col=color)
title(main="Commons' vote by region")
legend("topleft", inset=0.05, fill=c("cyan", "red"), legend=c("yes", "no"))
```



- Draw by mosaicplot() in base graphics.

```
mosaicplot(xtabs(counts~region+vote, data=sejong.poll.2[class.2=="Commons",], d
rop=T), col=color, main="Commons' votes by region")
```



- Chungcheong's case.

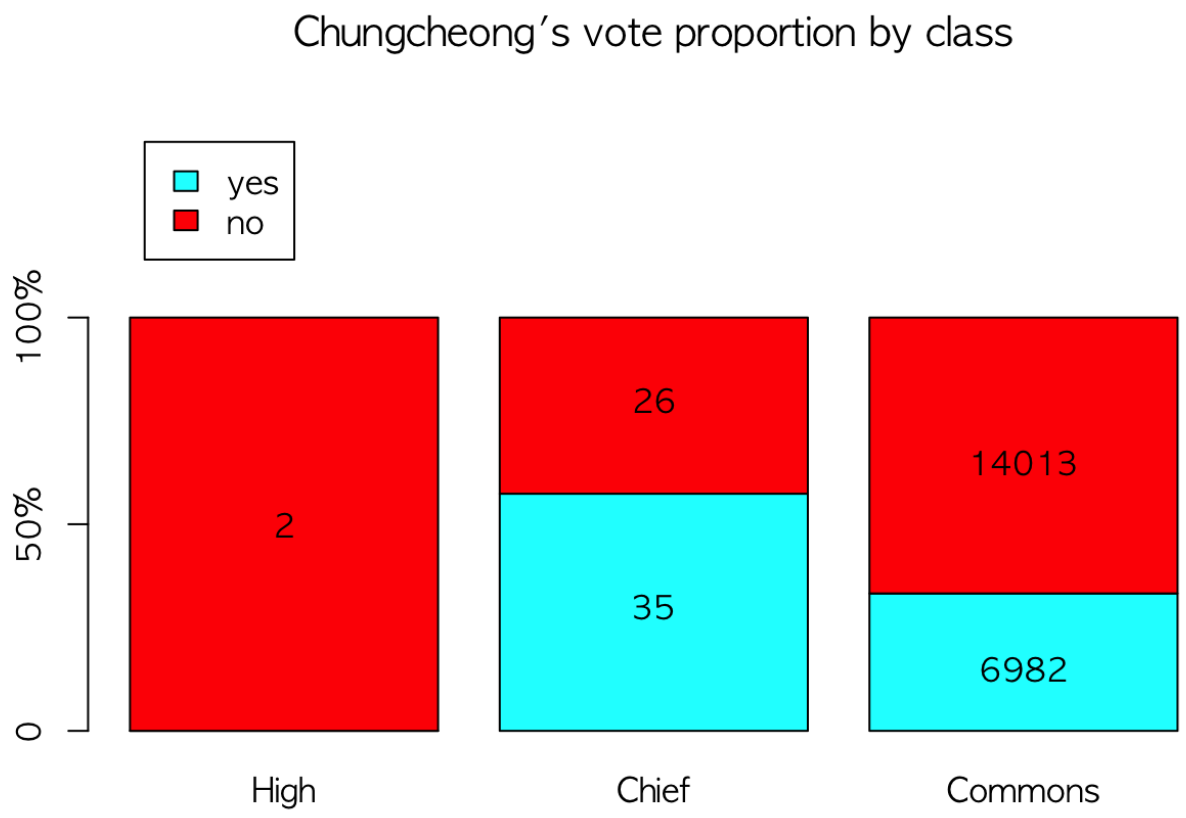
```
xtabs(counts~vote+class, data=sejong.poll.2[region=="Chungcheong",], drop=T)
```

```
##      class
## vote   High Chief Commons
##   yes     0    35    6982
##   no      2    26   14013
```

```
prop.table(xtabs(counts~vote+class, data=sejong.poll.2[region=="Chungcheong",],
drop=T), margin=2)
```

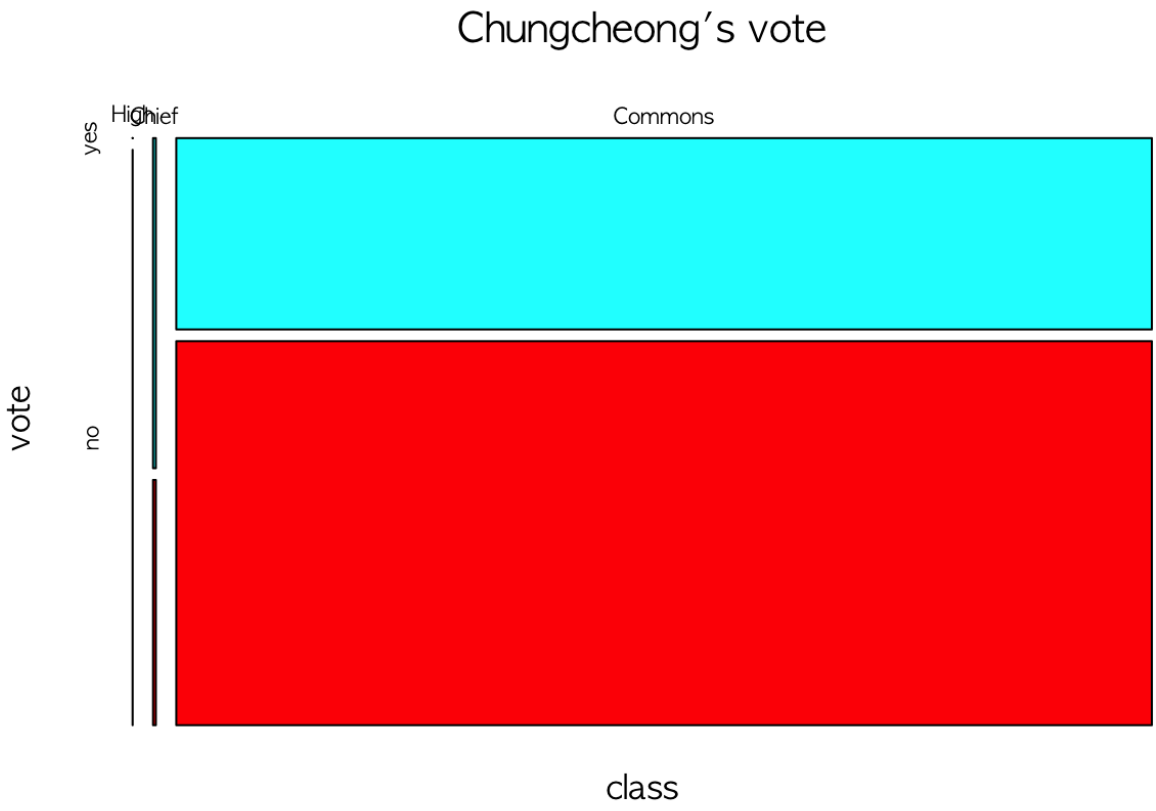
```
##      class
## vote   High Chief Commons
##   yes 0.000 0.574   0.333
##   no  1.000 0.426   0.667
```

```
barplot(prop.table(xtabs(counts~vote+class, data=sejong.poll.2[region=="Chungch
eong",], drop=T), margin=2), col=color, ylim=c(0, 1.5), axes=F)
axis(side=2, at=c(0, 0.5, 1.0), labels=c("0", "50%", "100%"))
title(main="Chungcheong's vote proportion by class")
legend("topleft", inset=0.05, fill=c("cyan", "red"), legend=c("yes", "no"))
text(x=c(0.7, 1.9, 1.9, 3.1, 3.1), y=c(0.5, 0.3, 0.8, 0.15, 0.65), labels=c(2,
35, 26, 6982, 14013))
```



- With mosaicplot, it's hard to compare.

```
mosaicplot(xtabs(counts~class+vote, data=sejong.poll.2[region=="Chungcheong",],
drop=T), col=color, main="")
title(main="Chungcheong's vote")
```



- Save the working directory image, save history and quit.

```
save.image(file="sejong_poll0328.rda")
savehistory(file="sejong_poll0328.Rhistory")
q("no")
```