Simple Effect Assignment

Pradeep Paladugula

5/4/2020

Q1) HO: their is no difference in the cholestoral from using two brands of margarine H1: Thier is a significant different in the Cholestral between using two differnet brands of cholestral Solution: 1) From the densitya and box plots it is clearly see that thier is a significant different between the Cholestoral between two different brnads of Margarine. 2) From t-test even though the p-value id less the 0.5 and thier is clear differnece in the means values, t value is greater than 3.

```
library('ggplot2')
cData <- read.csv('CholestoralData.csv')
summary(cData)
```

```
##        ID            Before           After        Margarine
##  Min.   : 1.00   Min.   : 3.910   Min.   : 3.660   A:20
##  1st Qu.:10.75   1st Qu.: 6.530   1st Qu.: 5.290   B:20
##  Median :20.50   Median : 7.860   Median : 6.415
##  Mean   :20.50   Mean   : 8.932   Mean   : 6.886
##  3rd Qu.:30.25   3rd Qu.:10.380   3rd Qu.: 7.690
##  Max.   :40.00   Max.   :17.730   Max.   :12.100
```

```
t.test(cData$Before, cData$After, paried = TRUE, alternative = 'two.sided',
var.equal = FALSE)
```

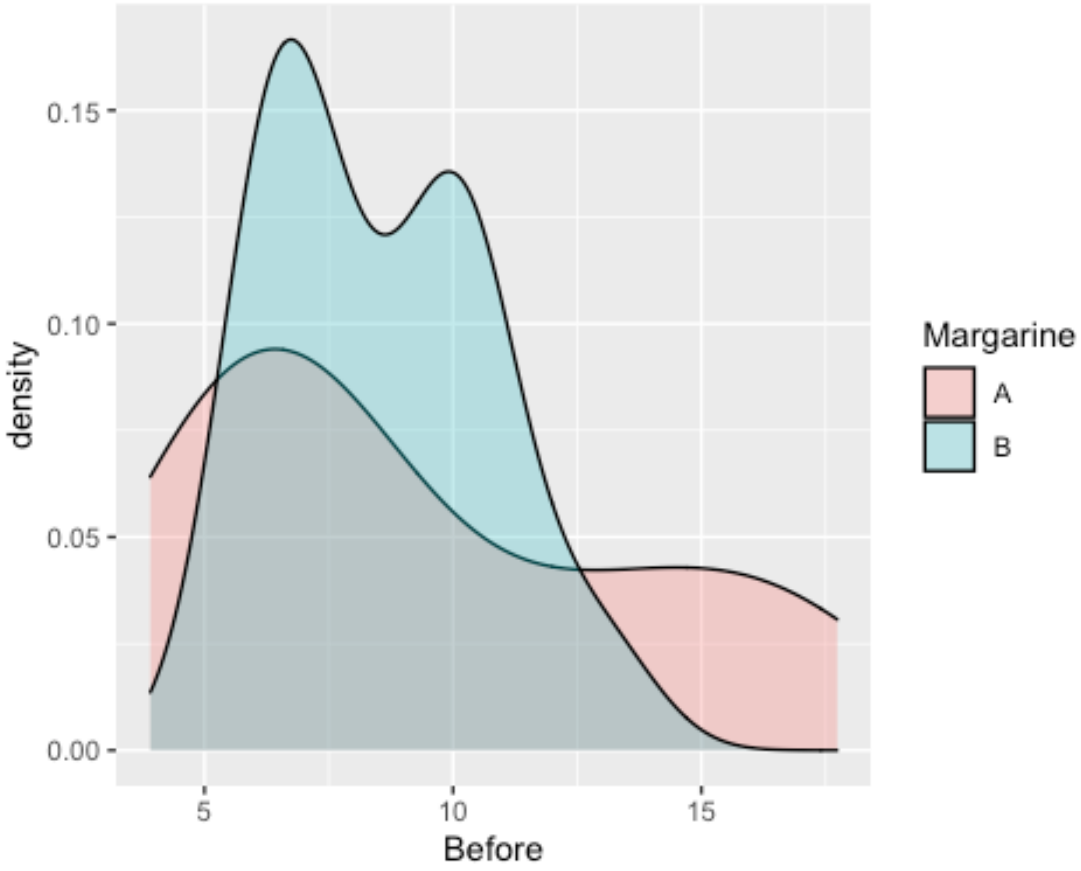
```
##
##  Welch Two Sample t-test
##
## data:  cData$Before and cData$After
## t = 3.0845, df = 67.188, p-value = 0.002958
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.7222737 3.3707263
## sample estimates:
## mean of x mean of y
##    8.9320    6.8855
```

```
t.test(cData$Before, cData$After, paried = TRUE, alternative = 'two.sided',
var.equal = TRUE)
```

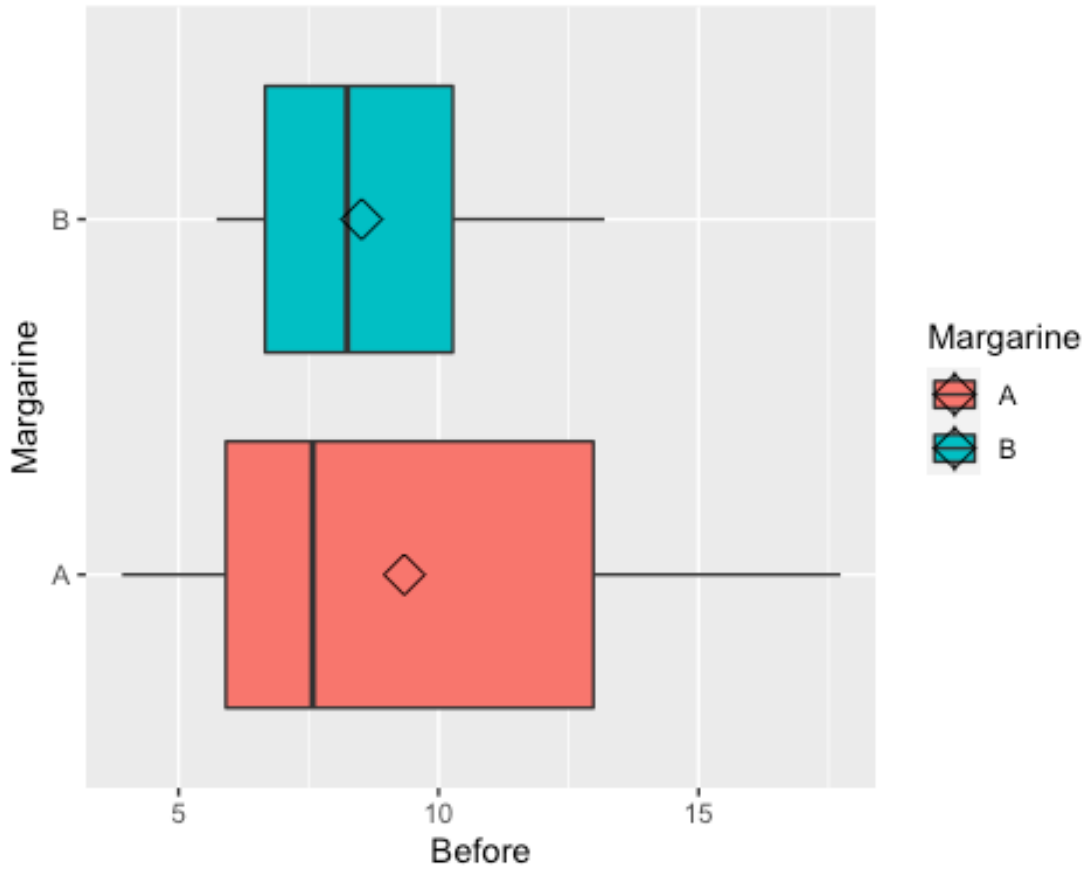
```
##
##  Two Sample t-test
##
## data:  cData$Before and cData$After
## t = 3.0845, df = 78, p-value = 0.00282
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.72563 3.36737
## sample estimates:
## mean of x mean of y
##    8.9320    6.8855
```

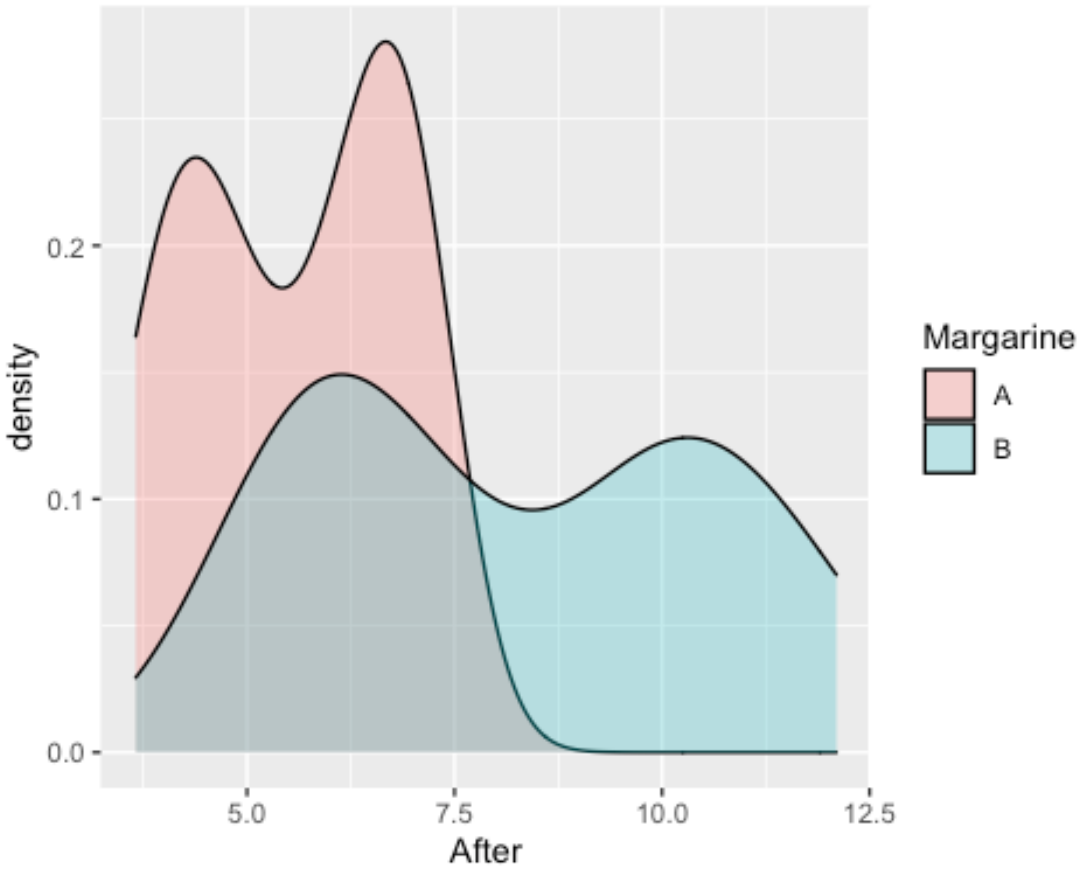
```
ggplot(cData, aes(x = Before, fill = Margarine)) +
        geom_density(alpha = .3)
```



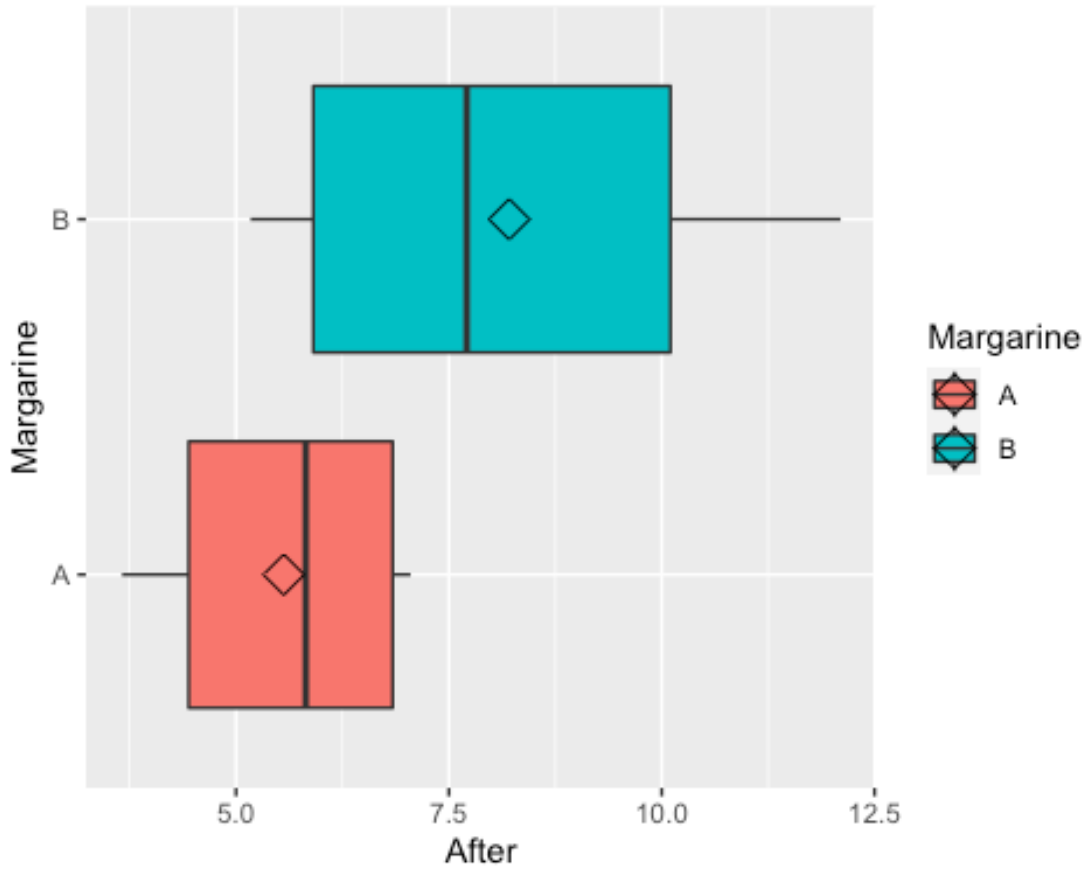

```
ggplot(cData, aes(x = Margarine, y = Before, fill = Margarine)) +
        geom_boxplot() + coord_flip() + stat_summary(fun=mean, geom="point",
shape=5, size=4)
```



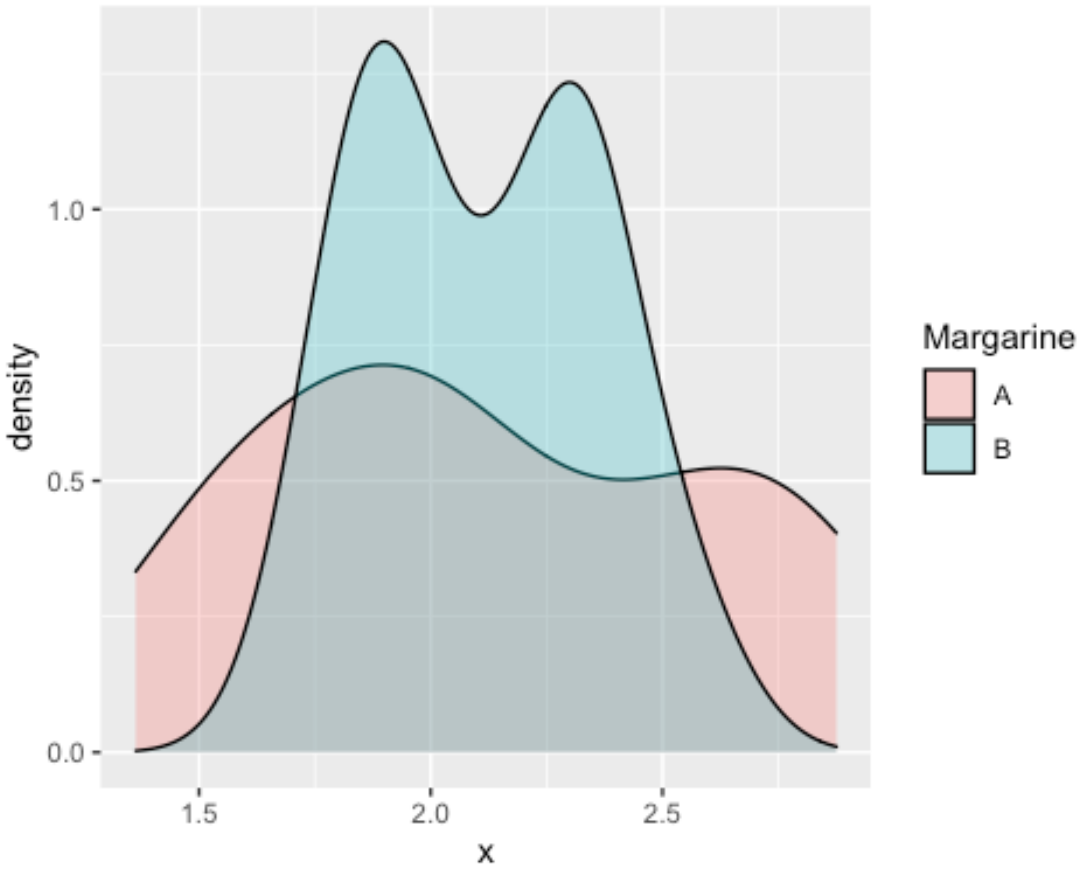

```
ggplot(cData, aes(x = After, fill = Margarine)) +
         geom_density(alpha = .3)
```



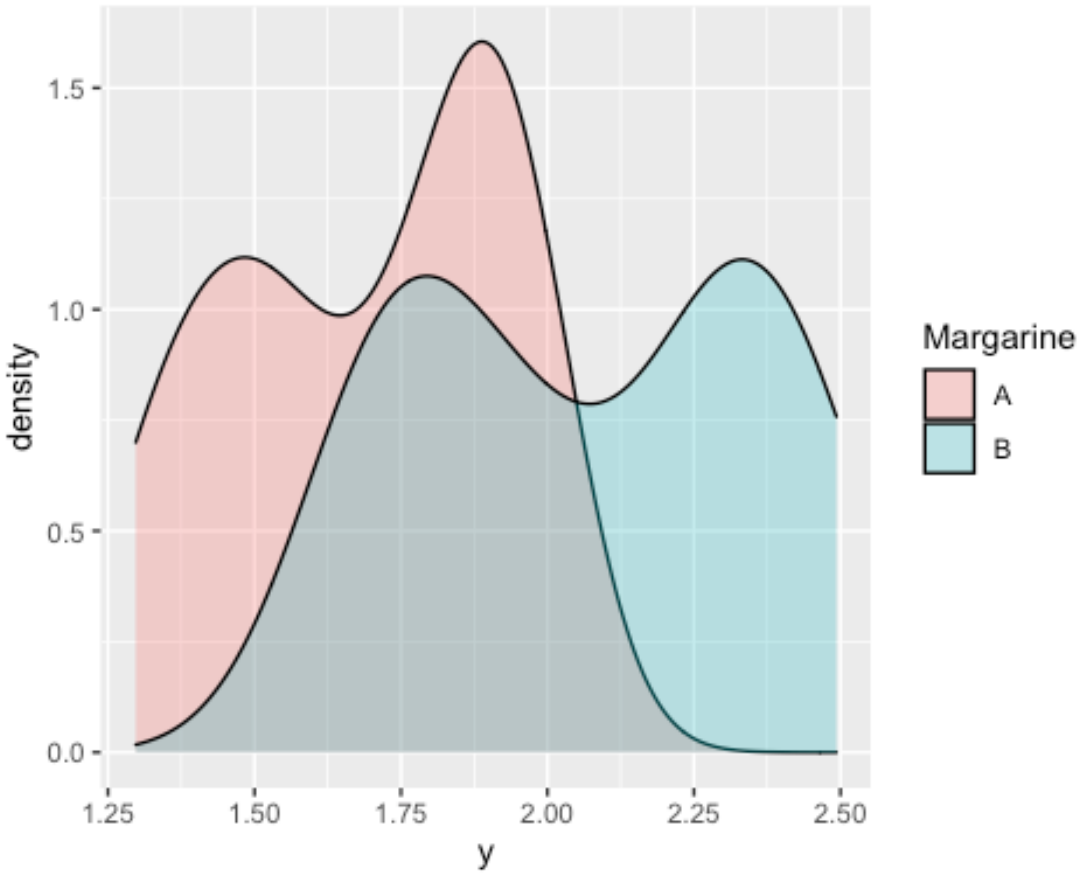

```
ggplot(cData, aes(x = Margarine, y = After, fill = Margarine)) +
         geom_boxplot() + coord_flip() + stat_summary(fun=mean, geom="point",
shape=5, size=4)
```




```
x <- log(cData$Before)
ggplot(cData, aes(x = x, fill = Margarine)) +
         geom_density(alpha = .3)
```




```
y <- log(cData$After)
ggplot(cData, aes(x = y, fill = Margarine)) +
         geom_density(alpha = .3)
```



Q2) H0: Their is no difference in children help or doesn't help in cooking the food for the calories they consumed H1: Their is a significant difference in children help or doesn't help in cooking the food for the calories they consumed Solution: 1) From Levens `Test: the p-value is 0.8716 which is great than 0.05 and significatly supports the null hypothesis.  2) From the independent t-tests`: I have conducted t-test twice for both clasical and Welch t-tests. I noticed that their is significant diference between the mean values of those whose help cooking and doesn't help in cooking.

```
library(lawstat)
kcData <- read.csv('kidscalories.csv')
levene.test(
  kcData$calorieintake,
  kcData$helpedinprep,
  location = c("median", "mean", "trim.mean"),
  trim.alpha = 0.25,
  bootstrap = FALSE,
  num.bootstrap = 1000,
  kruskal.test = FALSE,
  correction.method = c("none", "correction.factor", "zero.removal",
"zero.correction")
)
```

```
##
##  Modified robust Brown-Forsythe Levene-type test based on the absolute
##  deviations from the median
##
## data:  kcData$calorieintake
## Test Statistic = 0.026441, p-value = 0.8716
```

```
t.test(calorieintake ~ helpedinprep, data = kcData, alternative =
c("two.sided"), var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  calorieintake by helpedinprep
## t = 2.8248, df = 44.779, p-value = 0.007039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   24.27227 144.92875
## sample estimates:
## mean in group 1 mean in group 2
##        431.3996        346.7991
```

```
t.test(calorieintake ~ helpedinprep, data = kcData, alternative =
c("two.sided"), var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  calorieintake by helpedinprep
## t = 2.8137, df = 45, p-value = 0.007236
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   24.04243 145.15859
## sample estimates:
## mean in group 1 mean in group 2
##        431.3996        346.7991
```

```
tapply(kcData$calorieintake, kcData$helpedinprep, sd)
```

```
##         1         2
## 105.70124  99.50114
```

Q3) H0: Kids Priorities in Goals have not differed from the regions they lived in H1: Kids Priorities in Goals have significance difference from the regions they lived in Solution: I have conducted the ChiSquare statistics to investiagte the whether the ditribution of categorical variable differ from one another. 1) For rural reagion the p-value is 0.4677 whihc is greateer than the conventionally accepted significance level of 0.05 we fail to reject the null hypothesis. 2) For Suburban and Urban reagion the p-value is less than 0.001 whihc is far less than the conventionally accepted significance level of 0.05 we we reject the null hypothesis in favor of the alternative hypothesis. when we clarified the pearsons ChiSquare method for Urban and Suburban inclusively, when considered together both the

regions are showing same interests in thier goals. 3) When compared between rural and (Urban and Subarban), there is much differece seen between the goals of the kids.

```
library(reshape)
pData <- read.csv('PrioritiesData.csv')

## Warning in read.table(file = file, header = header, sep = sep, quote =
quote, :
## incomplete final line found by readTableHeader on 'PrioritiesData.csv'

modifiedPData <- melt(pData, id=c("Goal"))

#levene.test(value ~ variable, modifiedPData, center = mean)

chisq.test(pData$Rural)

##
##  Chi-squared test for given probabilities
##
## data:  pData$Rural
## X-squared = 1.52, df = 2, p-value = 0.4677

chisq.test(pData$Suburban)

##
##  Chi-squared test for given probabilities
##
## data:  pData$Suburban
## X-squared = 30.32, df = 2, p-value = 2.607e-07

chisq.test(pData$Urban)

##
##  Chi-squared test for given probabilities
##
## data:  pData$Urban
## X-squared = 57.38, df = 2, p-value = 3.468e-13

chisq.test(pData[,3:4])

##
##  Pearson's Chi-squared test
##
## data:  pData[, 3:4]
## X-squared = 3.6416, df = 2, p-value = 0.1619
```

Q4)

```
vData <- read.csv('VotingData.csv')
#colnames <- c("VotedFor", "DRL", "RLD", "LDR")
#rownames <- c("DEM", "REP", "LIB")
```

```
#P <- matrix(vData[c(1:2)], nrow = 4, byrow = TRUE, dimnames = list(rownames, colnames))
```