HW6

Harrison Fringer

2025-03-19

Problem 1

a.) First, we wil use the provided SDev output to find the PVE.

```
## [1] 0.9655342206 0.0278173366 0.0057995349 0.0008489079
```

b.) We will now find the proportions of variance using Equation 12.10.

```
##                        PC1         PC2          PC3        PC4
## Alabama          64.802164 -11.4480074  -2.49493284 -2.4079009
## Alaska           92.827450 -17.9829427  20.12657487  4.0940470
## Arizona         124.068216   8.8304030  -1.68744836  4.3536852
## Arkansas         18.340035 -16.7039114   0.21018936  0.5209936
## California      107.422953  22.5200698   6.74587299  2.8118259
## Colorado         34.975986  13.7195840  12.27936280  1.7214637
## Connecticut     -60.887282  12.9325302  -8.42065719  0.6999023
## Delaware         66.731025   1.3537978 -11.28095735  3.7279812
## Florida         165.244370   6.2746901  -2.99793315 -1.2476807
## Georgia          40.535177  -7.2902396   3.60952946 -7.3436728
## Hawaii         -123.536106  24.2912079   3.72444284 -3.4728494
## Idaho           -51.797002  -9.4691910  -1.52006356  3.3478283
## Illinois         78.992097  12.8970605  -5.88326477 -0.3676407
## Indiana         -57.550961   2.8462647   3.73816049 -1.6494302
## Iowa           -115.586790  -3.3421305  -0.65402935  0.8694960
## Kansas          -55.789694   3.1572339   0.38436416 -0.6527917
## Kentucky        -62.383181 -10.6732715   2.23708903 -3.8762164
## Louisiana        78.277631  -4.2949175  -3.82786965 -4.4835590
## Maine           -89.261044 -11.4878272  -4.69240562  2.1161995
## Maryland        129.330136  -5.0070315  -2.34717282  1.9283242
## Massachusetts   -21.266283  19.4501790  -7.50714835  1.0348189
## Michigan         85.451527   5.9045567   6.46434210 -0.4990479
## Minnesota       -98.954816   5.2096006   0.00657376  0.7318957
## Mississippi      86.856358 -27.4284196  -5.00343624 -3.8797577
## Missouri          7.986289   5.2756414   5.50057972 -0.6794055
## Montana         -62.483635  -9.5105021   1.83835536 -0.2459426
## Nebraska        -69.096544  -0.2111959   0.46802086  0.6565664
## Nevada           83.613578  15.1021839  15.88869482 -0.3341962
## New Hampshire  -114.777355  -4.7345584  -2.28238693  0.9359106
## New Jersey      -10.815725  23.1373389  -6.31015739 -1.6124273
## New Mexico      114.868163  -0.3364531   2.26126996  1.3812478
## New York         84.294231  15.9239655  -4.72125960 -0.8920194
```

```
## North Carolina  164.325514 -31.0966153 -11.69616350  2.1111927
## North Dakota   -127.495597 -16.1350394  -1.31182982  2.3009639
## Ohio            -50.086822  12.2793244   1.65733077 -2.0291157
## Oklahoma        -19.693723   3.3701310  -0.45314329  0.1803457
## Oregon          -11.150240   3.8660682   8.12998050  2.9140109
## Pennsylvania    -64.689142   8.9115466  -3.20646858 -1.8749353
## Rhode Island      3.063973  18.3739704 -17.47001970  2.3082597
## South Carolina  107.281069 -23.5361159  -2.03279501 -1.2517463
## South Dakota    -86.106720 -16.5978586   1.31437998  1.2522874
## Tennessee        17.506264  -6.5065756   6.10012753 -3.9228558
## Texas            31.291122  12.9849566  -0.39340922 -4.2420040
## Utah            -49.913397  17.6484577   1.78816852  1.8677052
## Vermont        -124.714469 -27.3135591   4.80277765  2.0049857
## Virginia        -14.817448  -1.7526150   1.04538813 -1.1738408
## Washington      -25.075839   9.9679669   4.78112764  2.6910819
## West Virginia   -91.544647 -22.9528778  -0.40198344 -0.7368781
## Wisconsin      -118.176328   5.5075792  -2.71132077 -0.2049724
## Wyoming         -10.434539  -5.9244529  -3.79444682  0.5178674
```

```
##          PC1          PC2          PC3          PC4
## 0.9655342206 0.0278173366 0.0057995349 0.0008489079
```

Problem 2

a.) We will now perform hierarchical clustering of the USArrests dataset

b.) We will now cut the tree and see what we get:

```
##        Alabama         Alaska        Arizona       Arkansas     California
##              1              1              1              2              1
##       Colorado    Connecticut       Delaware        Florida        Georgia
##              2              3              1              1              2
##         Hawaii          Idaho       Illinois        Indiana           Iowa
##              3              3              1              3              3
##         Kansas       Kentucky      Louisiana          Maine       Maryland
##              3              3              1              3              1
##  Massachusetts       Michigan      Minnesota    Mississippi       Missouri
##              2              1              3              1              2
##        Montana       Nebraska         Nevada  New Hampshire     New Jersey
##              3              3              1              3              2
##     New Mexico       New York North Carolina   North Dakota           Ohio
##              1              1              1              3              3
##       Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
##              2              2              3              2              1
##   South Dakota      Tennessee          Texas           Utah        Vermont
##              3              2              2              3              3
##       Virginia     Washington  West Virginia      Wisconsin        Wyoming
##              2              2              3              3              2
```

c/d.) We will now scale the variables, then perform the same operations.

```
##        Alabama         Alaska        Arizona       Arkansas     California
##              1              1              2              3              2
```

```
##       Colorado    Connecticut       Delaware        Florida        Georgia
##              2              3              3              2              1
##         Hawaii          Idaho       Illinois        Indiana           Iowa
##              3              3              2              3              3
##         Kansas       Kentucky      Louisiana          Maine       Maryland
##              3              3              1              3              2
##  Massachusetts       Michigan      Minnesota    Mississippi       Missouri
##              3              2              3              1              3
##        Montana       Nebraska         Nevada New Hampshire     New Jersey
##              3              3              2              3              3
##     New Mexico       New York North Carolina   North Dakota           Ohio
##              2              2              1              3              3
##       Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
##              3              3              3              3              1
##   South Dakota      Tennessee          Texas           Utah        Vermont
##              3              1              2              3              3
##       Virginia     Washington  West Virginia      Wisconsin        Wyoming
##              3              3              3              3              3
```

```
##            cutscale
## cut_arrests  1  2  3
##           1  6  9  1
##           2  2  2 10
##           3  0  0 20
```

Using the table below, it is clearly shown that the scaling affects the clustering. That being said, we should take note of the different variables used. Most are in per 100,000 units, but UrbanPop is expressed as percentages. To make sure that all variables can have a more balanced impact on the clustering, I would recommend scaling the variables before making clusters.

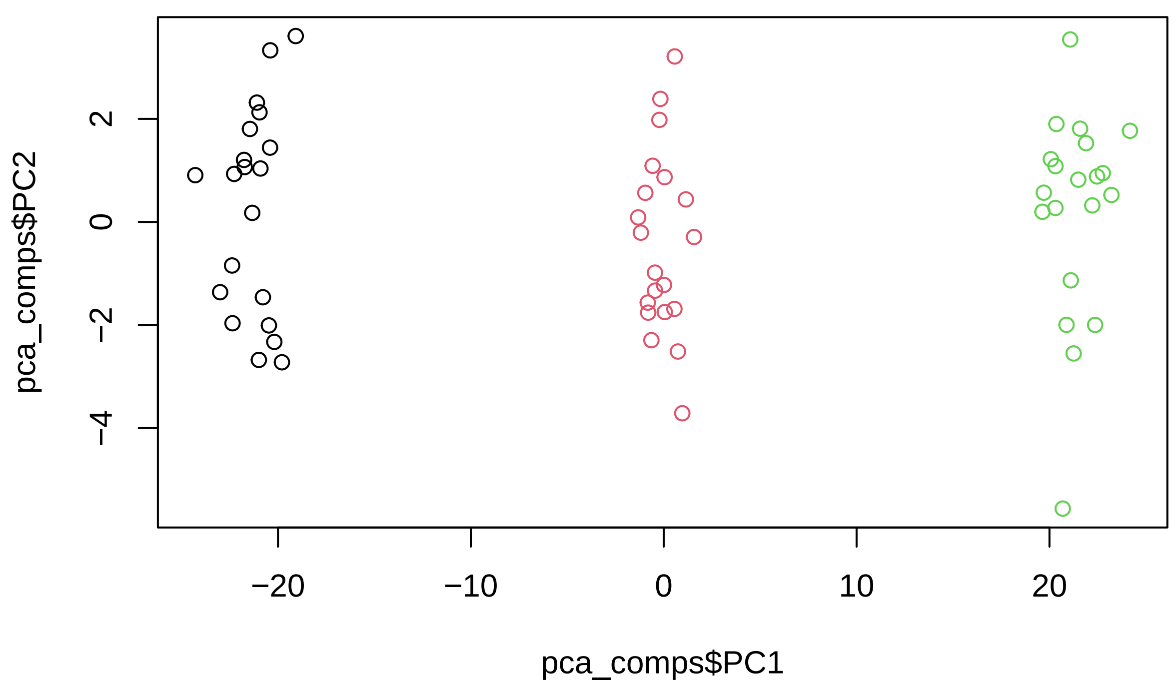
Problem 3

a. We will first make the requested data:

b.) We will now perform PCA on the simulated data:

```
## Importance of components:
##                            PC1     PC2     PC3    PC4     PC5     PC6     PC7
## Standard deviation     17.5963 1.92330 1.82018 1.8042 1.71470 1.67690 1.59535
## Proportion of Variance  0.8563 0.01023 0.00916 0.0090 0.00813 0.00778 0.00704
## Cumulative Proportion   0.8563 0.86652 0.87568 0.8847 0.89282 0.90059 0.90763
##                            PC8     PC9    PC10   PC11    PC12    PC13    PC14
## Standard deviation     1.53114 1.45437 1.43933 1.4224 1.37844 1.30575 1.27929
## Proportion of Variance 0.00648 0.00585 0.00573 0.0056 0.00525 0.00472 0.00453
## Cumulative Proportion  0.91412 0.91997 0.92570 0.9313 0.93655 0.94126 0.94579
##                           PC15    PC16    PC17    PC18    PC19   PC20   PC21
## Standard deviation     1.20986 1.17089 1.13522 1.12060 1.07328 1.0415 1.0062
## Proportion of Variance 0.00405 0.00379 0.00356 0.00347 0.00319 0.0030 0.0028
## Cumulative Proportion  0.94983 0.95363 0.95719 0.96066 0.96385 0.9668 0.9697
##                           PC22    PC23    PC24    PC25    PC26    PC27    PC28
## Standard deviation     0.99242 0.94456 0.92245 0.89426 0.84476 0.84138 0.83722
## Proportion of Variance 0.00272 0.00247 0.00235 0.00221 0.00197 0.00196 0.00194
```

```
## Cumulative Proportion  0.97237 0.97484 0.97719 0.97940 0.98138 0.98334 0.98527
##                           PC29    PC30    PC31    PC32    PC33    PC34    PC35
## Standard deviation     0.80378 0.74452 0.72890 0.68052 0.64468 0.62737 0.62198
## Proportion of Variance 0.00179 0.00153 0.00147 0.00128 0.00115 0.00109 0.00107
## Cumulative Proportion  0.98706 0.98859 0.99006 0.99134 0.99249 0.99358 0.99465
##                           PC36    PC37    PC38    PC39    PC40   PC41    PC42
## Standard deviation     0.54655 0.52101 0.50575 0.47212 0.43657 0.4245 0.36914
## Proportion of Variance 0.00083 0.00075 0.00071 0.00062 0.00053 0.0005 0.00038
## Cumulative Proportion  0.99548 0.99623 0.99694 0.99755 0.99808 0.9986 0.99895
##                           PC43    PC44   PC45    PC46    PC47    PC48    PC49
## Standard deviation     0.32098 0.29072 0.2677 0.19726 0.17448 0.15869 0.11447
## Proportion of Variance 0.00028 0.00023 0.0002 0.00011 0.00008 0.00007 0.00004
## Cumulative Proportion  0.99924 0.99947 0.9997 0.99978 0.99986 0.99993 0.99997
##                           PC50
## Standard deviation     0.10483
## Proportion of Variance 0.00003
## Cumulative Proportion  1.00000
```



c.) We will now perform K-Means with 3 clusters:

```
##
##       1  2  3
##    1  0 20  0
##    2  0  0 20
##    3 20  0  0
```

We can see it performed perfectly, which is to be expected

d.) Now we perform it with 2 clusters:

```
##
##       1  2  3
##    1  0  0 20
##    2 20 20  0
```

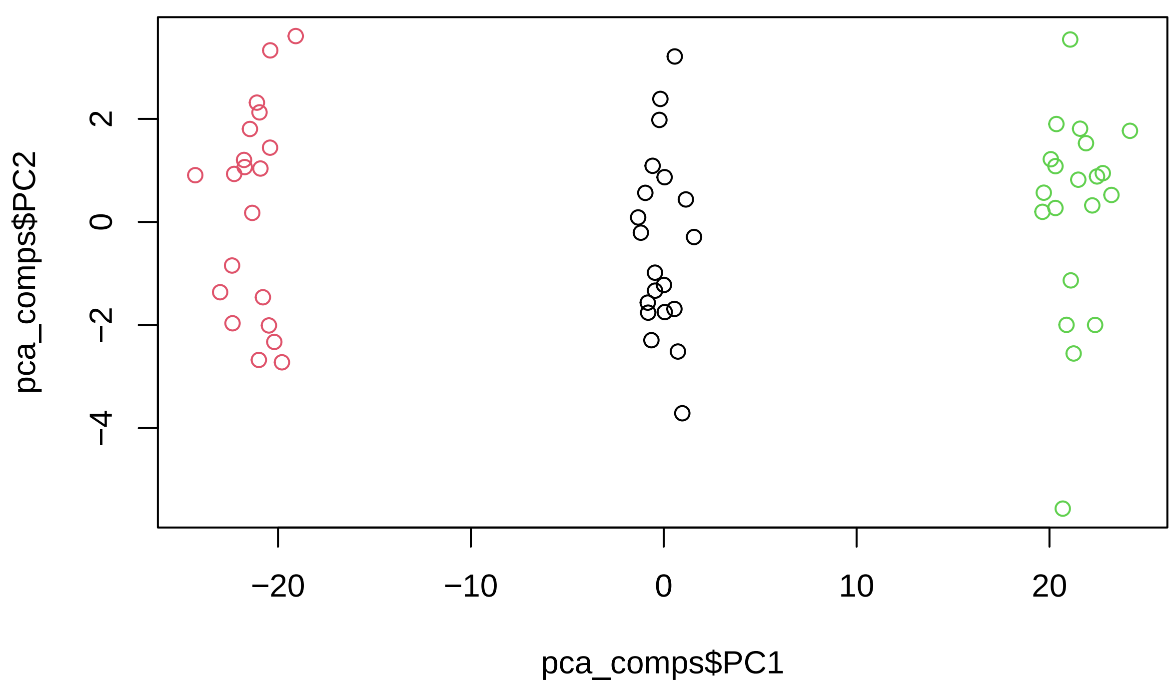
Here we can see that two of the defined clusters are being combined, this makes sense.

e.) Now we perform this with 4 clusters:

```
##
##       1  2  3
##    1  9  0  0
##    2 11  0  0
##    3  0  0 20
##    4  0 20  0
```

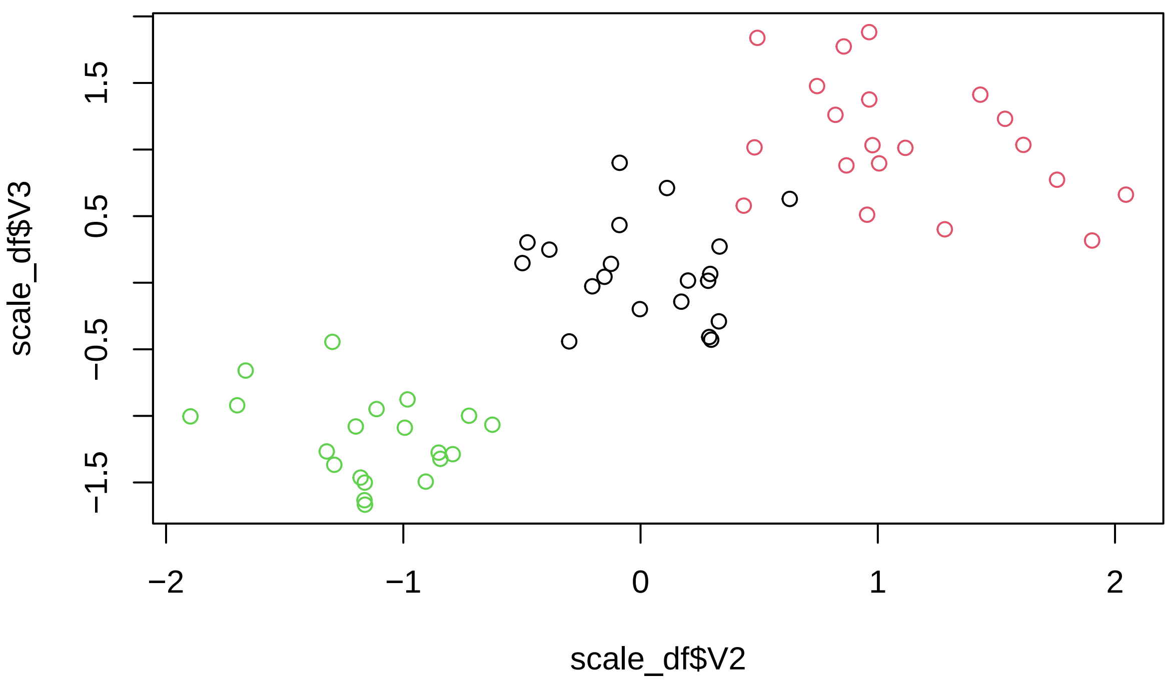
Here, one of our clusters is getting split into two separate clusters. This makes sense, as whichever cluster was most variable became two separate clusters.

f.) Now we perform K-means of 3, on the first two principle components.



It is clear that the group was properly separated. This makes sense as most of our variance was explained by the first two principle componenets.

g.) We will now scale the data and do this again:



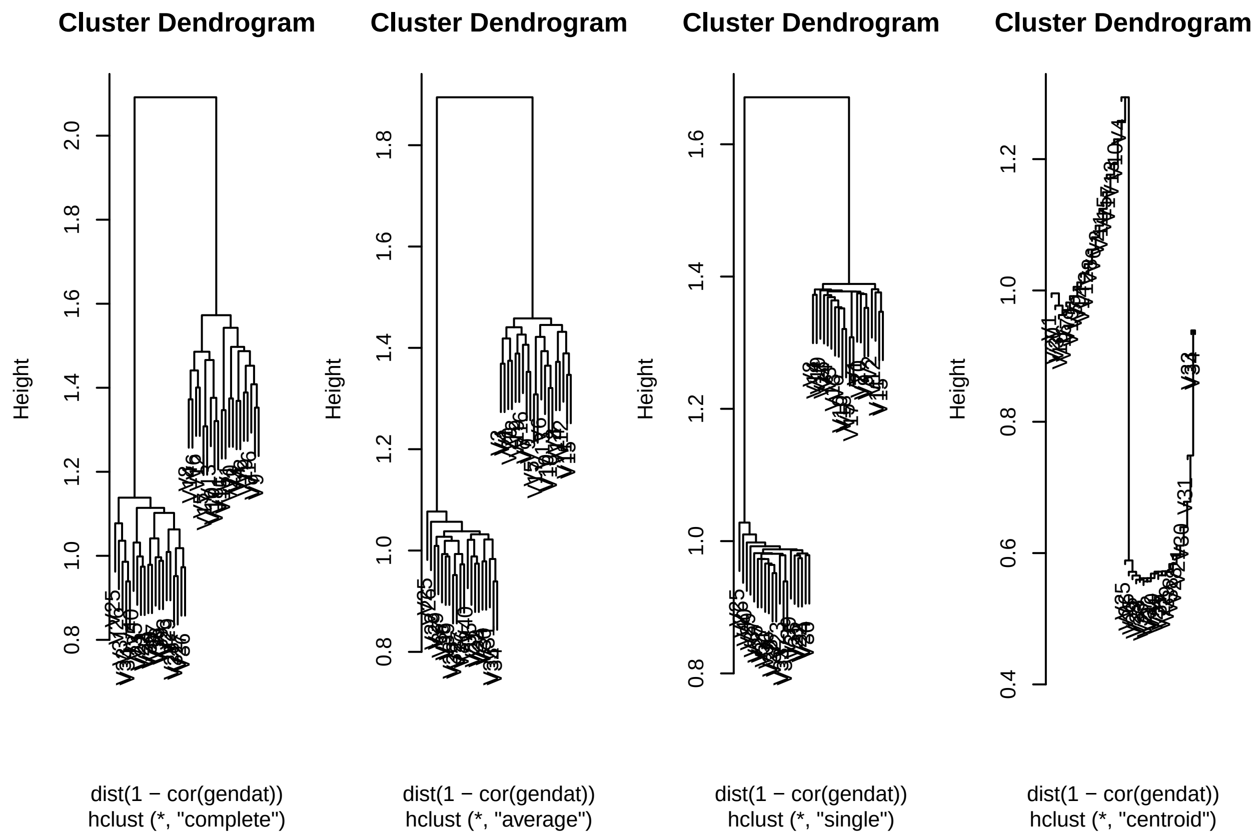

```
##
##       1  2  3
##    1  0 20  0
##    2  0  0 20
##    3 20  0  0
```

After performing the scaling and placing the appropriate plots, kmeans still found the correct subtrends. This makes sense because rescaling shouldn't significantly change the separation of each group because the standard deviation of each variable has (roughly) the same variance before we scale them.

Problem 4

a.) Let's read in the data from the csv:

b.) Time to make a dendrogram, using correlation-based distance:



It looks that regardless of the linkage type utilized, we see results that split the samples into two distinct groups, irregardless of linkage type. However, the specific orders slightly vary between the groups depending on the type of linkage used. Most notably, when applying centroid-based linkage, we get a considerably different looking dendrogram.

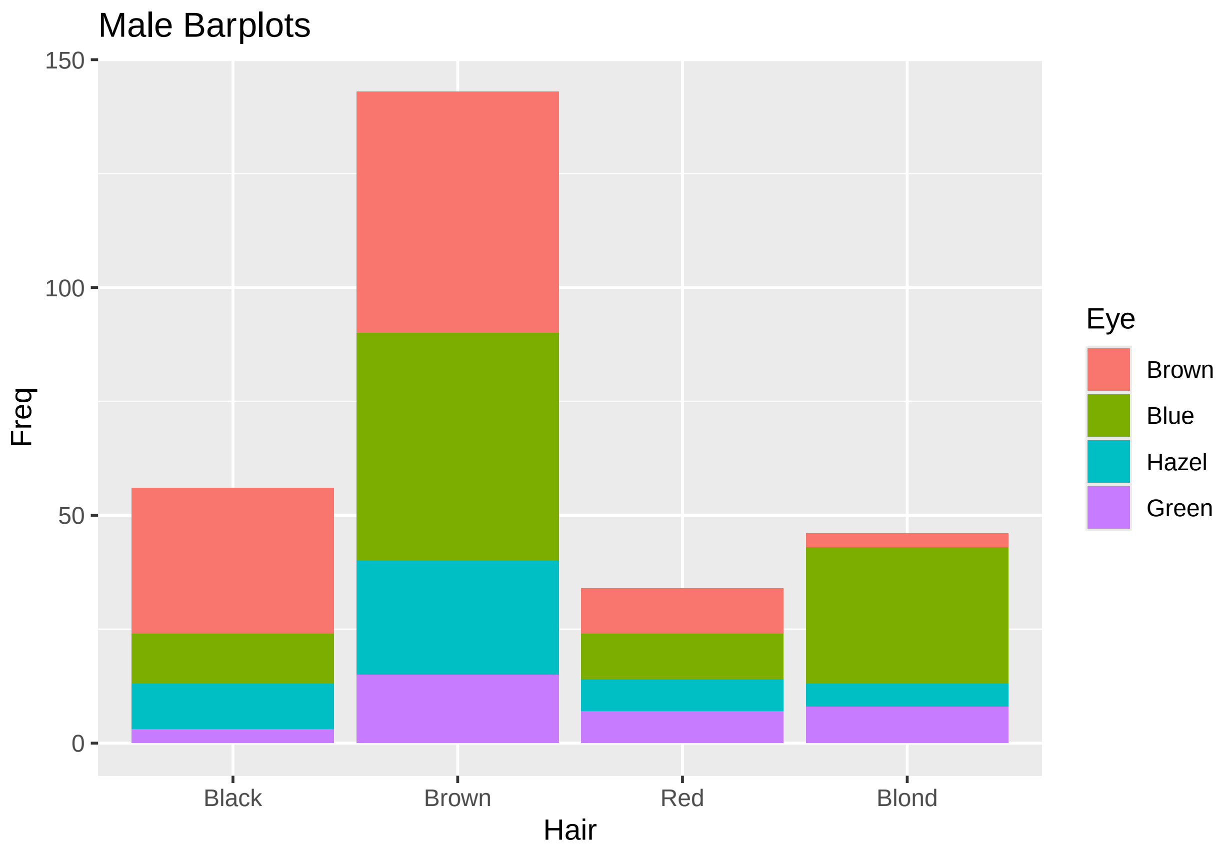
c.) Albeit not a statistically sound method, we can take the two distinct groups, then sum the rows and do a simple value comparison to see which genes show the greatest difference. This is more of a point estimate way, and could likely be answered further through use of methods such as t-tests. We will explore this point estimate approach using the complete linkage dendrogram. That said, we should note the method can easily be extended to the other linkage types. We will NOT take the absolute value until the end as that may interfere with the measure of variability. Again, I'd like to argue this isn't an optimal approach, but does provide us with an interpretable result.

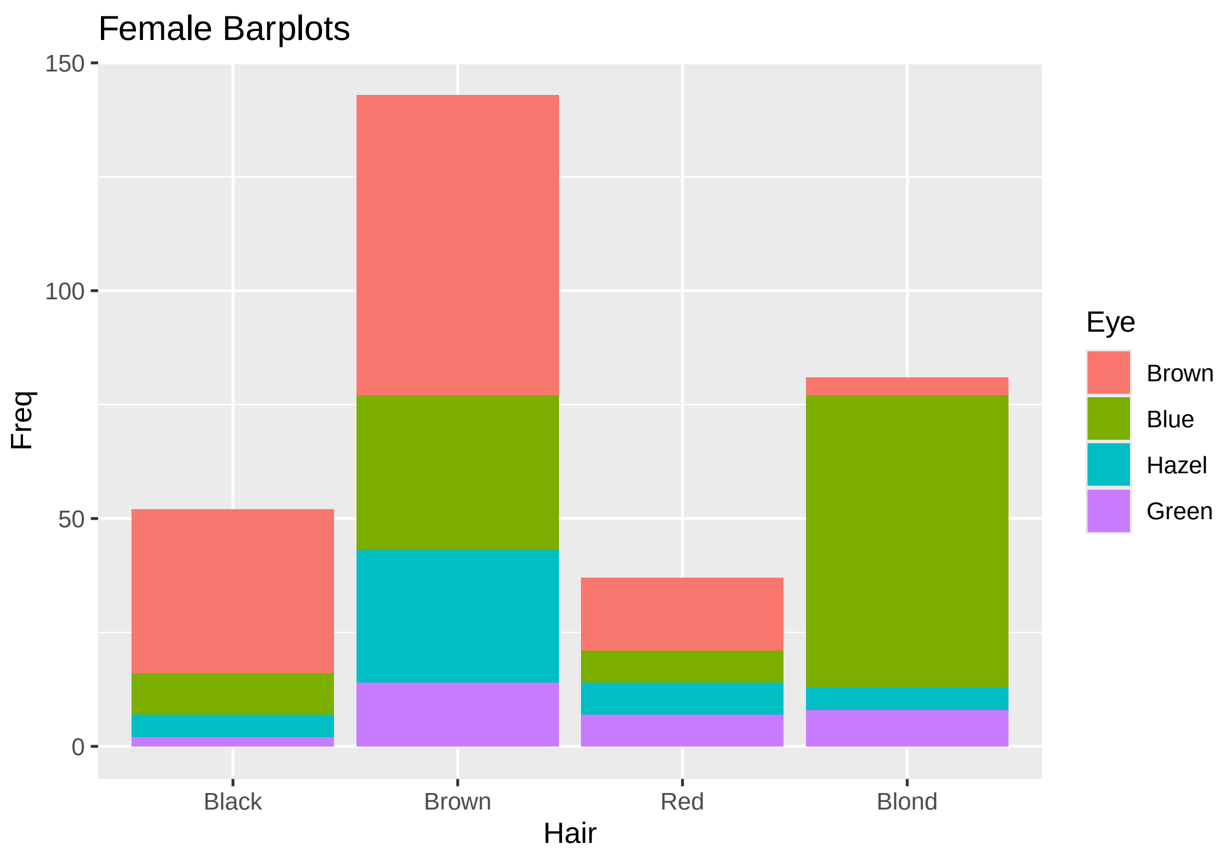
```
##     mean_diffs gene_num
## 600   2.747577      600
## 584   2.601985      584
## 549   2.550757      549
## 540   2.545174      540
## 502   2.544461      502
## 568   2.519418      568
```

Taking the head of the dataframe, we can see that the genes with the largest differences between means of the two groups are genes 600, 584, 549, 540, 502, and 568.

Problem 5

a.) Let's begin getting the data organized and plotted as requested:





b.) We will now split the table into respective two way tables, and then find the appropriate chi-square independence tests.

```
##
##  Pearson's Chi-squared test
##
## data:  twes
## X-squared = 1.5298, df = 3, p-value = 0.6754

##
##  Pearson's Chi-squared test
##
## data:  twhs
## X-squared = 7.9942, df = 3, p-value = 0.04613

##
##  Pearson's Chi-squared test
##
## data:  twhe
## X-squared = 138.29, df = 9, p-value < 2.2e-16
```

From these 3 separate analyses, we can see that both Hair/Sex and Hair/Eye suggest that there is an association between these two pairs of features.

Problem 6

a.) Let's first put the data into a dataframe, then build the appropriate marginal tables to find the analyses of interest.

```
## [1] "Chi-Sq Test for Dept A"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 16.372, df = 1, p-value = 5.205e-05
##
## [1] "Chi-Sq Test for Dept B"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 0.085098, df = 1, p-value = 0.7705
##
## [1] "Chi-Sq Test for Dept C"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 0.63322, df = 1, p-value = 0.4262
##
## [1] "Chi-Sq Test for Dept D"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 0.22159, df = 1, p-value = 0.6378
##
## [1] "Chi-Sq Test for Dept E"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 0.80805, df = 1, p-value = 0.3687
##
## [1] "Chi-Sq Test for Dept F"
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  twsub
## X-squared = 0.21824, df = 1, p-value = 0.6404
```

With these results, it appears that only Department A shows signs of an association between gender and admissions.

b.)

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
```

```
##
## data:  table.ucb
## X-squared = 91.61, df = 1, p-value < 2.2e-16
```

This result shows that ignoring department, there is an association between gender and admission rates overall.

c.) We will now carry out a CMH test and report the results:

```
##
##  Mantel-Haenszel chi-squared test with continuity correction
##
## data:  UCBAdmissions
## Mantel-Haenszel X-squared = 1.4269, df = 1, p-value = 0.2323
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.7719074 1.0603298
## sample estimates:
## common odds ratio
##         0.9046968
```

This CMH test concludes that there is not an association between gender and admissions.

d.) There is a slight confliction between parts b.) and c.), as one suggests no association while the other does suggest an association. That being said, investigation of each department shows that only a singular department shows a potential association between gender and admissions. With this in mind, I believe that it is safe to conclude that there is not an association between gender and admissions.

e.) We will now calculate the success rates across each department, as well as the overall success rate. We will then draw some conclusions:

```
##       M     F
## A 0.621 0.824
## B 0.630 0.680
## C 0.369 0.341
## D 0.331 0.349
## E 0.277 0.239
## F 0.059 0.070
```

```
## [1] "Overall:"
```

```
## [1] "M    F"
```

```
## [1] 0.445 0.304
```

Disregarding the previous tests, acting solely on this information, I would conclude that certain departments (A) have gender-bias in admissions. The overall value also seems to suggest that there is some form of gender bias in admissions.

Problem 7

a.) Let's load and analyze some of the base features of this dataset:

```
## transactions as itemMatrix in sparse format with
##  9835 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.02609146
##
## most frequent items:
##       whole milk other vegetables       rolls/buns             soda
##             2513             1903             1809             1715
##           yogurt          (Other)
##             1372            34055
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46
##   17   18   19   20   21   22   23   24   26   27   28   29   32
##   29   14   14    9   11    4    6    1    1    1    1    3    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   4.409   6.000  32.000
##
## includes extended item information - examples:
##        labels  level2           level1
## 1 frankfurter sausage meat and sausage
## 2     sausage sausage meat and sausage
## 3  liver loaf sausage meat and sausage
```

```
##
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46
##   17   18   19   20   21   22   23   24   26   27   28   29   32
##   29   14   14    9   11    4    6    1    1    1    1    3    1
```

```
## [1] "Percentage of >20 Item Transactions"
```

```
## [1] 0.386
```

```
## [1] "Average Number of items per transaction:"
```

```
## [1] 4.408541
```

Looking at the provided, summary, we can see that there are i.) 9,835 rows (transactions), ii.) with the item most frequently being bought being whole milk. iii.) The number of transactions involving 20 or more items is 0.386% iv.) with the average number of items per transaction being 4.408 items, meaning we should either round to 4 or 5 items depending on our rounding criterion.

b.) We will now find all rules with support > 1% and confidence > 50%. This gives:

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.01      1
##  maxlen target  ext
```

```
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [15 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
##      lhs                                    rhs                support
## [1]  {curd, yogurt}                      => {whole milk}       0.01006609
## [2]  {other vegetables, butter}          => {whole milk}       0.01148958
## [3]  {other vegetables, domestic eggs}   => {whole milk}       0.01230300
## [4]  {yogurt, whipped/sour cream}        => {whole milk}       0.01087951
## [5]  {other vegetables, whipped/sour cream} => {whole milk}    0.01464159
## [6]  {pip fruit, other vegetables}       => {whole milk}       0.01352313
## [7]  {citrus fruit, root vegetables}     => {other vegetables} 0.01037112
## [8]  {tropical fruit, root vegetables}   => {other vegetables} 0.01230300
## [9]  {tropical fruit, root vegetables}   => {whole milk}       0.01199797
## [10] {tropical fruit, yogurt}            => {whole milk}       0.01514997
## [11] {root vegetables, yogurt}           => {other vegetables} 0.01291307
## [12] {root vegetables, yogurt}           => {whole milk}       0.01453991
## [13] {root vegetables, rolls/buns}       => {other vegetables} 0.01220132
## [14] {root vegetables, rolls/buns}       => {whole milk}       0.01270971
## [15] {other vegetables, yogurt}          => {whole milk}       0.02226741
##      confidence coverage   lift     count
## [1]  0.5823529  0.01728521 2.279125  99
## [2]  0.5736041  0.02003050 2.244885 113
## [3]  0.5525114  0.02226741 2.162336 121
## [4]  0.5245098  0.02074225 2.052747 107
## [5]  0.5070423  0.02887646 1.984385 144
## [6]  0.5175097  0.02613116 2.025351 133
## [7]  0.5862069  0.01769192 3.029608 102
## [8]  0.5845411  0.02104728 3.020999 121
## [9]  0.5700483  0.02104728 2.230969 118
## [10] 0.5173611  0.02928317 2.024770 149
## [11] 0.5000000  0.02582613 2.584078 127
## [12] 0.5629921  0.02582613 2.203354 143
## [13] 0.5020921  0.02430097 2.594890 120
## [14] 0.5230126  0.02430097 2.046888 125
## [15] 0.5128806  0.04341637 2.007235 219
```

From this analysis, we see that there are 15 rules with confidence higher than .01 and confidence higher than 0.5. We can additionally observe that the highest support comes from the rule {citrus fruit, root vegetables} -> {other vegetables}, with a confidence of 0.5862. We see that the rule with the highest support is {other vegetables, yogurt} -> {whole milk}, with a support of 0.0223. For the first rule, we can interpret that: Support (.0103): 1.03% of all transactions contained the item pair citrus fruits, root vegetables, other vegetables. Confidence (.5862): 58.62% of the time that citrus fruits and root vegetabes were purchased, so were other vegetables. Lift (3.029): Purchasing root vegetables and citrus fruits saw a 3.029 times increase in the purchasing of other vegetables.

For the second rule, we can interpret that: Support (.0223): 2.23% of all transactions contained the item pair other vegetables, yogurt, and whole milk. Confidence (.5128): 51.28% of the time that other vegetables and yogurt were purchased, so was whole milk. Lift (3.029): Purchasing other vegetables and yogurt saw a 3.029 times increase in the purchasing of whole milk.

c.) We will now perform the analysis such that we view only rules with > 1% support, > 20% confidence, and a lhs containing 'whole milk':

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.2    0.1    1 none FALSE            TRUE       5    0.01      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [3 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
##     lhs             rhs                support    confidence coverage lift
## [1] {whole milk} => {yogurt}           0.05602440 0.2192598  0.255516 1.571735
## [2] {whole milk} => {rolls/buns}       0.05663447 0.2216474  0.255516 1.205032
## [3] {whole milk} => {other vegetables} 0.07483477 0.2928770  0.255516 1.513634
##     count
## [1] 551
## [2] 557
## [3] 736
```

d.) Lastly, we will use the same parameter levels, but look for rhs being 'whole milk':

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.2    0.1    1 none FALSE            TRUE       5    0.01      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
```

```
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 98
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [71 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

##      lhs                              rhs           support confidence   coverage      lift count
## [1]  {}                            => {whole milk} 0.25551601  0.2555160 1.00000000 1.0000000  2513
## [2]  {hard cheese}                 => {whole milk} 0.01006609  0.4107884 0.02450432 1.6076815    99
## [3]  {butter milk}                 => {whole milk} 0.01159126  0.4145455 0.02796136 1.6223854   114
## [4]  {ham}                         => {whole milk} 0.01148958  0.4414062 0.02602949 1.7275091   113
## [5]  {sliced cheese}               => {whole milk} 0.01077783  0.4398340 0.02450432 1.7213560   106
## [6]  {oil}                         => {whole milk} 0.01128622  0.4021739 0.02806304 1.5739675   111
## [7]  {onions}                      => {whole milk} 0.01209964  0.3901639 0.03101169 1.5269647   119
## [8]  {berries}                     => {whole milk} 0.01179461  0.3547401 0.03324860 1.3883281   116
## [9]  {hamburger meat}              => {whole milk} 0.01474326  0.4434251 0.03324860 1.7354101   145
## [10] {hygiene articles}            => {whole milk} 0.01281139  0.3888889 0.03294357 1.5219746   126
## [11] {salty snack}                 => {whole milk} 0.01118454  0.2956989 0.03782410 1.1572618   110
## [12] {sugar}                       => {whole milk} 0.01504830  0.4444444 0.03385867 1.7393996   148
## [13] {waffles}                     => {whole milk} 0.01270971  0.3306878 0.03843416 1.2941961   125
## [14] {long life bakery product} => {whole milk} 0.01352313  0.3614130 0.03741739 1.4144438   133
## [15] {dessert}                     => {whole milk} 0.01372649  0.3698630 0.03711235 1.4475140   135
## [16] {cream cheese }               => {whole milk} 0.01647178  0.4153846 0.03965430 1.6256696   162
## [17] {chicken}                     => {whole milk} 0.01759024  0.4099526 0.04290798 1.6044106   173
## [18] {white bread}                 => {whole milk} 0.01708185  0.4057971 0.04209456 1.5881474   168
## [19] {chocolate}                   => {whole milk} 0.01667514  0.3360656 0.04961871 1.3152427   164
## [20] {coffee}                      => {whole milk} 0.01870869  0.3222417 0.05805796 1.2611408   184
## [21] {frozen vegetables}           => {whole milk} 0.02043721  0.4249471 0.04809354 1.6630940   201
## [22] {beef}                        => {whole milk} 0.02125064  0.4050388 0.05246568 1.5851795   209
## [23] {curd}                        => {whole milk} 0.02613116  0.4904580 0.05327911 1.9194805   257
## [24] {napkins}                     => {whole milk} 0.01972547  0.3766990 0.05236401 1.4742678   194
## [25] {pork}                        => {whole milk} 0.02216573  0.3844797 0.05765125 1.5047187   218
## [26] {frankfurter}                 => {whole milk} 0.02053889  0.3482759 0.05897306 1.3630295   202
## [27] {bottled beer}                => {whole milk} 0.02043721  0.2537879 0.08052872 0.9932367   201
## [28] {brown bread}                 => {whole milk} 0.02521607  0.3887147 0.06487036 1.5212930   248
## [29] {margarine}                   => {whole milk} 0.02419929  0.4131944 0.05856634 1.6170980   238
## [30] {butter}                      => {whole milk} 0.02755465  0.4972477 0.05541434 1.9460530   271
## [31] {newspapers}                  => {whole milk} 0.02735130  0.3426752 0.07981698 1.3411103   269
## [32] {domestic eggs}               => {whole milk} 0.02999492  0.4727564 0.06344687 1.8502027   295
## [33] {fruit/vegetable juice}       => {whole milk} 0.02663955  0.3684951 0.07229283 1.4421604   262
## [34] {whipped/sour cream}          => {whole milk} 0.03223183  0.4496454 0.07168277 1.7597542   317
## [35] {pip fruit}                   => {whole milk} 0.03009659  0.3978495 0.07564820 1.5570432   296
## [36] {pastry}                      => {whole milk} 0.03324860  0.3737143 0.08896797 1.4625865   327
## [37] {citrus fruit}                => {whole milk} 0.03050330  0.3685504 0.08276563 1.4423768   300
## [38] {shopping bags}               => {whole milk} 0.02450432  0.2487100 0.09852567 0.9733637   241
## [39] {sausage}                     => {whole milk} 0.02989324  0.3181818 0.09395018 1.2452520   294
## [40] {bottled water}               => {whole milk} 0.03436706  0.3109476 0.11052364 1.2169396   338
```

```
## [41] {tropical fruit}           => {whole milk} 0.04229792  0.4031008 0.10493137 1.5775950   416
## [42] {root vegetables}          => {whole milk} 0.04890696  0.4486940 0.10899847 1.7560310   481
## [43] {soda}                     => {whole milk} 0.04006101  0.2297376 0.17437722 0.8991124   394
## [44] {yogurt}                   => {whole milk} 0.05602440  0.4016035 0.13950178 1.5717351   551
## [45] {rolls/buns}               => {whole milk} 0.05663447  0.3079049 0.18393493 1.2050318   557
## [46] {other vegetables}         => {whole milk} 0.07483477  0.3867578 0.19349263 1.5136341   736
## [47] {curd,
##       yogurt}                   => {whole milk} 0.01006609  0.5823529 0.01728521 2.2791250    99
## [48] {pork,
##       other vegetables}         => {whole milk} 0.01016777  0.4694836 0.02165735 1.8373939   100
## [49] {other vegetables,
##       butter}                   => {whole milk} 0.01148958  0.5736041 0.02003050 2.2448850   113
## [50] {other vegetables,
##       domestic eggs}            => {whole milk} 0.01230300  0.5525114 0.02226741 2.1623358   121
## [51] {other vegetables,
##       fruit/vegetable juice}    => {whole milk} 0.01047280  0.4975845 0.02104728 1.9473713   103
## [52] {yogurt,
##       whipped/sour cream}       => {whole milk} 0.01087951  0.5245098 0.02074225 2.0527473   107
## [53] {other vegetables,
##       whipped/sour cream}       => {whole milk} 0.01464159  0.5070423 0.02887646 1.9843854   144
## [54] {pip fruit,
##       other vegetables}         => {whole milk} 0.01352313  0.5175097 0.02613116 2.0253514   133
## [55] {other vegetables,
##       pastry}                   => {whole milk} 0.01057448  0.4684685 0.02257245 1.8334212   104
## [56] {citrus fruit,
##       yogurt}                   => {whole milk} 0.01026945  0.4741784 0.02165735 1.8557678   101
## [57] {citrus fruit,
##       other vegetables}         => {whole milk} 0.01301474  0.4507042 0.02887646 1.7638982   128
## [58] {sausage,
##       other vegetables}         => {whole milk} 0.01016777  0.3773585 0.02694459 1.4768487   100
## [59] {other vegetables,
##       bottled water}            => {whole milk} 0.01077783  0.4344262 0.02480935 1.7001918   106
## [60] {tropical fruit,
##       root vegetables}          => {whole milk} 0.01199797  0.5700483 0.02104728 2.2309690   118
## [61] {tropical fruit,
##       yogurt}                   => {whole milk} 0.01514997  0.5173611 0.02928317 2.0247698   149
## [62] {tropical fruit,
##       rolls/buns}               => {whole milk} 0.01098119  0.4462810 0.02460600 1.7465872   108
## [63] {tropical fruit,
##       other vegetables}         => {whole milk} 0.01708185  0.4759207 0.03589222 1.8625865   168
## [64] {root vegetables,
##       yogurt}                   => {whole milk} 0.01453991  0.5629921 0.02582613 2.2033536   143
## [65] {root vegetables,
##       rolls/buns}               => {whole milk} 0.01270971  0.5230126 0.02430097 2.0468876   125
## [66] {root vegetables,
##       other vegetables}         => {whole milk} 0.02318251  0.4892704 0.04738180 1.9148326   228
## [67] {yogurt,
##       soda}                     => {whole milk} 0.01047280  0.3828996 0.02735130 1.4985348   103
## [68] {other vegetables,
##       soda}                     => {whole milk} 0.01392984  0.4254658 0.03274021 1.6651240   137
## [69] {yogurt,
##       rolls/buns}               => {whole milk} 0.01555669  0.4526627 0.03436706 1.7715630   153
## [70] {other vegetables,
##       yogurt}                   => {whole milk} 0.02226741  0.5128806 0.04341637 2.0072345   219
```

```
## [71] {other vegetables,
##       rolls/buns}              => {whole milk} 0.01789527  0.4200477 0.04260295 1.6439194   176
```