Syed Noor Razi Ali
2070326

DSC 324/424

Assignment 5

Due: June 8th at 11:59pm CST

1) MDS and Hierarchical Clustering

Download the "kellog.csv" file which contains 10 variables on 22 cereals from various US brands. We'll look at MDS with two different distance metrics and evaluate the results of a hierarchical clustering.

a. Compute the Euclidean distance matrix on columns 2-11. In R, you can use the "dist" function which uses the Euclidean distance by default.

```
> distance = dist(cer)
> print(distance)
            1          2          3          4          5          6          7          8          9         10
2  196.606553
3  136.332663 313.207176
4   60.845846 250.949947 126.992546
5  294.600424 231.286021 330.144228 347.213095
6  223.249581 179.498135 267.390429 277.840976  74.062599
7  256.034222 154.672737 323.699839 312.838984  81.471028  68.264130
8  208.861226 200.233623 234.909473 260.238826 103.567133  48.190197 112.392640
9  172.745301 190.150532 205.497722 222.276451 128.499567  65.215692 124.078156  49.940791
10 123.368547 207.404829 141.037117 166.923545 193.314025 127.746428 183.667409 100.059977  65.989287
11 270.845562 230.117736 295.302727 323.140542  45.673414  55.380366  96.902100  69.620011 102.126448 162.336216
12 241.617351 278.717913 221.302623 280.726555 140.298109 117.968405 182.719737  84.123564  94.620514 118.700602
13 258.209478 216.770780 286.950326 311.134239  50.339156  41.609109  88.137954  58.800210  91.067568 151.992114
14 182.318907 131.629867 250.355729 239.352024 121.968222  54.444184  77.217011  73.228902  67.015432 112.733270
           11         12         13         14         15         16         17         18         19         20
2
3
4
5
6
7
8
9
10
11
12 105.579302
13  14.634581 105.681701
14 109.247032 150.575534  95.065558
           21         22         23         24         25         26         27         28         29         30
2
3
4
5
6
7
8
9
10
11
12
13
14
           31         32         33         34         35         36         37         38         39         40
```

b. Compute the distance matrix again, but this time use the Gower distance (in the StatMatch package). library(StatMatch)
gower.dist(DATAFRAME)

Syed Noor Razi Ali
2070326

```
> gdx = gower.dist(cereal[4:16])
> gdx
            [,1]      [,2]       [,3]      [,4]       [,5]       [,6]       [,7]       [,8]       [,9]     [,10]
 [1,] 0.00000000 0.3318711 0.06638538 0.1400190 0.30459941 0.36082141 0.35607798 0.31214533 0.28007211 0.1790866
 [2,] 0.33187106 0.0000000 0.35619274 0.4317362 0.20435699 0.25976330 0.25261602 0.20331279 0.31958789 0.2636819
 [3,] 0.06638538 0.3561927 0.00000000 0.1605691 0.29526724 0.36110463 0.38039966 0.29800547 0.27073995 0.1553314
 [4,] 0.14001901 0.4317362 0.16056912 0.0000000 0.39965687 0.45587888 0.42517391 0.40720280 0.37512958 0.2337595
 [5,] 0.30459941 0.2043570 0.29526724 0.3996569 0.00000000 0.12676803 0.14954097 0.10164405 0.18478371 0.1707051
 [6,] 0.36082141 0.2597633 0.36110463 0.4558789 0.12676803 0.00000000 0.13443859 0.18991734 0.11921084 0.2525681
 [7,] 0.35607798 0.2526160 0.38039966 0.4251739 0.14954097 0.13443859 0.00000000 0.23677647 0.21222382 0.2378888
 [8,] 0.31214533 0.2033128 0.29800547 0.4072028 0.10164405 0.18991734 0.23677647 0.00000000 0.21668860 0.1734433
 [9,] 0.28007211 0.3195879 0.27073995 0.3751296 0.18478371 0.11921084 0.21222382 0.21668860 0.00000000 0.1461778
[10,] 0.17908662 0.2636819 0.15533138 0.2337595 0.17070510 0.25256813 0.23788880 0.17344333 0.14617780 0.0000000
[11,] 0.39680938 0.2817652 0.37786183 0.4918668 0.12176299 0.11291104 0.12622843 0.18788083 0.21289111 0.2645176
[12,] 0.41118315 0.3634778 0.36338945 0.4581637 0.28091625 0.21534339 0.30835637 0.26442371 0.19599382 0.2936350
[13,] 0.39443784 0.2486245 0.38029798 0.4894953 0.11762912 0.12015488 0.15735097 0.17909903 0.20410932 0.2557358
           [,11]     [,12]      [,13]      [,14]       [,15]       [,16]      [,17]      [,18]       [,19]      [,20]
 [1,] 0.39680938 0.4111832 0.39443784 0.21515018 0.375588936 0.438860951 0.42356395 0.37634272 0.373610575 0.19709924
 [2,] 0.28176524 0.3634778 0.24862447 0.16752361 0.274530818 0.372204990 0.37987788 0.25971202 0.272552457 0.15240002
 [3,] 0.37786183 0.3633895 0.38029798 0.23466416 0.375872155 0.391067247 0.37577024 0.40066440 0.373893794 0.21661323
 [4,] 0.49186684 0.4581637 0.48949530 0.31020764 0.470646404 0.474303035 0.44939065 0.44543864 0.468668043 0.29215671
 [5,] 0.12176299 0.2809163 0.11762912 0.08944923 0.150223011 0.210701825 0.22302264 0.15517977 0.152892569 0.13826939
 [6,] 0.11291104 0.2153434 0.12015488 0.16169688 0.122459830 0.171539106 0.17921200 0.16000768 0.120481469 0.19449140
 [7,] 0.12622843 0.3083564 0.15735097 0.18932524 0.073919217 0.178927437 0.18892429 0.04479985 0.076588775 0.22532489
 [8,] 0.18788083 0.2644237 0.17909903 0.09699515 0.225838711 0.264416673 0.27208957 0.24422069 0.223860349 0.14581532
 [9,] 0.21289111 0.1959938 0.20410932 0.18228319 0.201286056 0.158788840 0.14349184 0.21966804 0.199307695 0.19973075
[10,] 0.26451763 0.2936350 0.25573583 0.13102510 0.257720266 0.259774332 0.24447733 0.24533302 0.255741905 0.13220493
[11,] 0.00000000 0.2994083 0.03955103 0.18806946 0.061605055 0.222852755 0.24151466 0.11461803 0.063583417 0.23047936
[12,] 0.29940827 0.0000000 0.29543417 0.25749266 0.297418602 0.174793181 0.15905911 0.31580059 0.295440241 0.30621408
[13,] 0.03955103 0.2954342 0.00000000 0.17928766 0.088079669 0.223526571 0.24218848 0.13933031 0.090058031 0.19733859
          [,21]      [,22]      [,23]      [,24]      [,25]      [,26]      [,27]      [,28]      [,29]      [,30]
 [1,] 0.3615062 0.33211172 0.23236308 0.27427952 0.33686829 0.40828590 0.27455499 0.22877454 0.27240867 0.34937886
 [2,] 0.2932835 0.27653869 0.18051832 0.25824667 0.23340633 0.33799701 0.22694917 0.20085293 0.27080854 0.27908997
 [3,] 0.3865994 0.30354878 0.25187706 0.26013966 0.36118997 0.39895373 0.29887667 0.23867314 0.24384574 0.37129669
 [4,] 0.3729099 0.36755380 0.32742054 0.32895237 0.43673345 0.47257413 0.34365092 0.32383201 0.33669691 0.44684017
 [5,] 0.2466188 0.11735948 0.09146710 0.09634553 0.13033129 0.14654031 0.20639057 0.14750319 0.18540752 0.12529353
 [6,] 0.2868484 0.22971896 0.13166347 0.18530563 0.11326366 0.08215079 0.20812540 0.19410981 0.25445004 0.11788437
 [7,] 0.1908713 0.16018421 0.12852260 0.16513179 0.02117493 0.11422715 0.16084991 0.23135356 0.22053994 0.07647395
 [8,] 0.2864567 0.16875037 0.13153573 0.12930451 0.21756679 0.23609977 0.21393649 0.10260156 0.12339050 0.21485299
 [9,] 0.2541212 0.20588576 0.16062430 0.15286126 0.19301413 0.13462404 0.16099656 0.21762776 0.22456269 0.19030033
[10,] 0.2436952 0.15302510 0.14013543 0.10480829 0.24944834 0.24362235 0.15377566 0.12516917 0.10293744 0.24673455
[11,] 0.2914587 0.21509854 0.16444630 0.20137601 0.10505350 0.11368422 0.23449797 0.20649637 0.22837506 0.05848870
[12,] 0.2692288 0.25073625 0.30848397 0.27131315 0.28914668 0.26152582 0.27260597 0.32411109 0.34258449 0.31720211
[13,] 0.2873249 0.21577236 0.17169014 0.19259421 0.13617604 0.12878111 0.22571618 0.21374022 0.25324711 0.08320097
           [,31]      [,32]      [,33]      [,34]      [,35]      [,36]      [,37]      [,38]      [,39]      [,40]
```


```
[13,] 0.2873249 0.21577236 0.17169014 0.19259421 0.13617604 0.12878111 0.22571618 0.21374022 0.25324711 0.08320097
           [,31]      [,32]      [,33]      [,34]      [,35]      [,36]      [,37]      [,38]      [,39]      [,40]
```


```
 [1,] 0.40525792 0.37586597 0.20984954 0.22406361 0.23285704 0.39567559 0.34384759 0.44507951 0.33872953 0.3624794
 [2,] 0.29466948 0.30557708 0.20455598 0.27510770 0.16656379 0.24986222 0.27355871 0.34402139 0.26206686 0.2832631
 [3,] 0.42957960 0.33768765 0.21974815 0.21953914 0.24756333 0.37672804 0.31047697 0.44536273 0.34382044 0.3675703
 [4,] 0.47435385 0.47092344 0.29529163 0.28042877 0.33441450 0.49073305 0.43890506 0.50936775 0.43378700 0.4575369
 [5,] 0.21911896 0.12341622 0.12677550 0.16922810 0.16217069 0.12527713 0.16460287 0.19759723 0.13125449 0.1936935
 [6,] 0.12490744 0.13235225 0.20543341 0.24788601 0.23441834 0.11177725 0.05703792 0.11502733 0.22005482 0.2627360
 [7,] 0.08680908 0.13877910 0.18629254 0.22839899 0.25483516 0.10858872 0.15432472 0.12257549 0.19042537 0.2740182
 [8,] 0.29994242 0.19246190 0.13488305 0.16219118 0.15573060 0.18674704 0.18287942 0.28250877 0.18096649 0.1049494
 [9,] 0.18448068 0.17271693 0.14925081 0.19351931 0.22572082 0.21175732 0.11698061 0.18103304 0.22692665 0.2814457
[10,] 0.28706873 0.23716396 0.09518600 0.10617021 0.16484734 0.26338384 0.21476098 0.29003135 0.21925830 0.2430081
[11,] 0.18495531 0.06902033 0.23180599 0.27425859 0.24680490 0.03190302 0.14430794 0.14362730 0.23628497 0.2751225
[12,] 0.30936847 0.26115717 0.24139868 0.26547604 0.29268853 0.26750525 0.18907470 0.24639635 0.31023869 0.3647578
[13,] 0.21375388 0.06504623 0.22302419 0.26547679 0.20725388 0.05356994 0.15635947 0.15872420 0.22750318 0.2663407
           [,41]      [,42]      [,43]      [,44]      [,45]      [,46]      [,47]      [,48]      [,49]      [,50]
 [1,] 0.40652590 0.23340062 0.35133193 0.29331705 0.33508748 0.34046790 0.33585640 0.30524576 0.24802175 0.28098064
 [2,] 0.33538867 0.22924160 0.25027381 0.24338653 0.25249832 0.28507521 0.26142766 0.23429308 0.25732905 0.19670574
 [3,] 0.35873220 0.24810692 0.35161514 0.30802335 0.34979378 0.32152035 0.33133193 0.28629821 0.24393918 0.28607155
 [4,] 0.47273722 0.32845809 0.44638939 0.36433606 0.43014495 0.43552537 0.43091387 0.40030323 0.27384845 0.37603811
 [5,] 0.18085739 0.14812186 0.12596600 0.23413034 0.19195146 0.16477253 0.09573655 0.14656810 0.18221868 0.11728224
 [6,] 0.21164586 0.13703618 0.09820282 0.24135967 0.24309990 0.09919967 0.10303215 0.23484139 0.26087659 0.16550511
 [7,] 0.17752827 0.16947223 0.05927759 0.19474353 0.25250664 0.16144342 0.10629173 0.28170052 0.21062033 0.21309383
 [8,] 0.22760035 0.15799175 0.20158170 0.24216872 0.10692125 0.19651766 0.15792786 0.04492405 0.16877150 0.06928499
 [9,] 0.20337687 0.13505928 0.17702905 0.19701267 0.27993694 0.06039579 0.13751506 0.21211299 0.18838402 0.23048545
[10,] 0.25820852 0.16072426 0.23346326 0.21735588 0.22523163 0.20176589 0.20676978 0.16654376 0.10103491 0.18243248
[11,] 0.18836040 0.16340876 0.07624668 0.25124873 0.23625569 0.18326456 0.08018374 0.22799718 0.28724917 0.17789168
[12,] 0.19822736 0.19348553 0.27316159 0.20699204 0.36261075 0.16089896 0.27426299 0.28517142 0.23808200 0.31266680
[13,] 0.18671025 0.16103722 0.10272129 0.24246694 0.25632005 0.17929045 0.07372591 0.22402308 0.27846737 0.17872526
           [,51]      [,52]      [,53]      [,54]      [,55]     [,56]      [,57]      [,58]      [,59]      [,60]
 [1,] 0.22887426 0.39025577 0.38448903 0.34357364 0.16415795 0.3504445 0.28336290 0.19477935 0.23065370 0.45891788
 [2,] 0.26434295 0.33770481 0.31929301 0.28295167 0.22019829 0.2906425 0.29158508 0.15792504 0.22645895 0.39338821
 [3,] 0.21954209 0.34246206 0.40190090 0.36555956 0.18607578 0.3779713 0.26922304 0.21429334 0.25497538 0.42073956
 [4,] 0.32393173 0.42569785 0.36792063 0.32700524 0.26161926 0.4118481 0.37842037 0.28983682 0.29974963 0.48474458
 [5,] 0.19817270 0.21626672 0.28695603 0.28611317 0.14044146 0.3455011 0.20618406 0.10982006 0.17362519 0.23420900
 [6,] 0.24798445 0.32165431 0.39026174 0.35712553 0.21909936 0.2478769 0.19830350 0.16604206 0.25228309 0.19272232
 [7,] 0.22561281 0.25909145 0.29428469 0.27364944 0.21355465 0.3484220 0.19516263 0.19687555 0.21116146 0.20011065
 [8,] 0.14096338 0.22991649 0.37576751 0.35865694 0.14798738 0.3530470 0.14416704 0.11736598 0.20553009 0.28559989
 [9,] 0.24378350 0.29479904 0.35636904 0.33925847 0.15885008 0.2331405 0.18183345 0.18490989 0.17163754 0.17884577
[10,] 0.15132491 0.21116915 0.28696865 0.26985808 0.11320594 0.3226399 0.16790050 0.12667991 0.12467305 0.27983126
[11,] 0.23793512 0.31168182 0.35311385 0.37832314 0.24547194 0.3607880 0.17703622 0.20203003 0.27865567 0.21326674
[12,] 0.38103606 0.28921247 0.42905237 0.41194180 0.24957323 0.2193925 0.30443216 0.28889572 0.33533714 0.18846549
[13,] 0.26440973 0.30770771 0.34897997 0.37595160 0.23669015 0.3584164 0.19710058 0.19965849 0.26987388 0.21394056
            [,61]      [,62]      [,63]      [,64]      [,65]     [,66]      [,67]      [,68]     [,69]      [,70]
 [1,] 0.439840445 0.38058563 0.33847373 0.35012227 0.33657982 0.3832635 0.29614178 0.38213440 0.3717167 0.30186851
```

Syed Noor Razi Ali
2070326

```
[12,] 0.19822736 0.19348553 0.27316159 0.20699204 0.36261075 0.16089896 0.27426299 0.28517142 0.23808200 0.31266680
[13,] 0.18671025 0.16103722 0.10272129 0.24246694 0.25632005 0.17929045 0.07372591 0.22402308 0.27846737 0.17872526
           [,51]      [,52]      [,53]      [,54]      [,55]     [,56]      [,57]      [,58]      [,59]      [,60]
 [1,] 0.22887426 0.39025577 0.38448903 0.34357364 0.16415795 0.3504445 0.28336290 0.19477935 0.23065370 0.45891788
 [2,] 0.26434295 0.33770481 0.31929301 0.28295167 0.22019829 0.2906425 0.29158508 0.15792504 0.22645895 0.39338821
 [3,] 0.21954209 0.34246206 0.40190090 0.36555956 0.18607578 0.3779713 0.26922304 0.21429334 0.25497538 0.42073956
 [4,] 0.32393173 0.42569785 0.36792063 0.32700524 0.26161926 0.4118481 0.37842037 0.28983682 0.29974963 0.48474458
 [5,] 0.19817270 0.21626672 0.28695603 0.28611317 0.14044146 0.3455011 0.20618406 0.10982006 0.17362519 0.23420900
 [6,] 0.24798445 0.32165431 0.39026174 0.35712553 0.21909936 0.2478769 0.19830350 0.16604206 0.25228309 0.19272232
 [7,] 0.22561281 0.25909145 0.29428469 0.27364944 0.21355465 0.3484220 0.19516263 0.19687555 0.21116146 0.20011065
 [8,] 0.14096338 0.22991649 0.37576751 0.35865694 0.14798738 0.3530470 0.14416704 0.11736598 0.20553009 0.28559989
 [9,] 0.24378350 0.29479904 0.35636904 0.33925847 0.15885008 0.2331405 0.18183345 0.18490989 0.17163754 0.17884577
[10,] 0.15132491 0.21116915 0.28696865 0.26985808 0.11320594 0.3226399 0.16790050 0.12667991 0.12467305 0.27983126
[11,] 0.23793512 0.31168182 0.35311385 0.37832314 0.24547194 0.3607880 0.17703622 0.20203003 0.27865567 0.21326674
[12,] 0.38103606 0.28921247 0.42905237 0.41194180 0.24957323 0.2193925 0.30443216 0.28889572 0.33533714 0.18846549
[13,] 0.26440973 0.30770771 0.34897997 0.37595160 0.23669015 0.3584164 0.19710058 0.19965849 0.26987388 0.21394056
            [,61]      [,62]      [,63]      [,64]      [,65]     [,66]      [,67]      [,68]     [,69]      [,70]
 [1,] 0.439840445 0.38058563 0.33847373 0.35012227 0.33657982 0.3832635 0.29614178 0.38213440 0.3717167 0.30186851
 [2,] 0.371385269 0.32760147 0.32577166 0.32410723 0.23744478 0.3711389 0.24959363 0.31018105 0.3228058 0.27078901
 [3,] 0.392046741 0.41320636 0.37143421 0.38308275 0.36090150 0.3498929 0.31084808 0.36318685 0.3671922 0.28292096
 [4,] 0.475282528 0.39198925 0.33753507 0.35258164 0.43644498 0.4187056 0.35562232 0.44834572 0.4667742 0.36807982
 [5,] 0.214530024 0.33702660 0.35378847 0.34515216 0.14751125 0.2543487 0.21823327 0.15332581 0.2737974 0.19055141
 [6,] 0.170719385 0.26183835 0.27860022 0.26996391 0.12060294 0.2012423 0.21996810 0.26336132 0.3281191 0.27882470
 [7,] 0.180431675 0.26687056 0.31440166 0.30576535 0.05950873 0.2202504 0.14807723 0.22691977 0.2824295 0.26414537
 [8,] 0.263596952 0.36484520 0.33569337 0.32705706 0.23009883 0.2761711 0.25478609 0.20720043 0.2015877 0.18527682
 [9,] 0.159768334 0.18702182 0.16900476 0.17368142 0.20554617 0.1728857 0.14747552 0.25590845 0.3298708 0.20641178
[10,] 0.260753826 0.28428810 0.22852993 0.24017847 0.26198038 0.2349461 0.16781873 0.23862471 0.2612218 0.15835881
[11,] 0.222033034 0.36513401 0.38189588 0.37325956 0.11792186 0.2728408 0.24634067 0.24736728 0.3146032 0.29558189
[12,] 0.170964983 0.26428391 0.26998760 0.26135129 0.33244795 0.0929175 0.26960941 0.31037433 0.4265080 0.26381767
[13,] 0.222706850 0.35635221 0.37311408 0.36447777 0.14672044 0.2688667 0.23755888 0.24323340 0.3346675 0.28680010
           [,71]      [,72]      [,73]      [,74]      [,75]
 [1,] 0.31499594 0.36624240 0.30686553 0.31928884 0.35653921
 [2,] 0.27435356 0.26518428 0.32276291 0.29824048 0.29073154
 [3,] 0.27200993 0.38575639 0.27349490 0.30034128 0.34720704
 [4,] 0.38120726 0.46129987 0.37307684 0.38550015 0.45159668
 [5,] 0.09834889 0.13975323 0.21408003 0.21800288 0.11472687
 [6,] 0.19676460 0.13234406 0.14850717 0.15243001 0.05929643
 [7,] 0.20271208 0.06256826 0.24152015 0.21467376 0.14216698
 [8,] 0.13579601 0.22931269 0.19758749 0.20631803 0.17153257
 [9,] 0.18876998 0.20476003 0.05891927 0.07501223 0.07646710
[10,] 0.16667855 0.26119424 0.15854814 0.17577914 0.22264490
[11,] 0.18277497 0.07239438 0.23257205 0.24611028 0.14283427
[12,] 0.24996663 0.30089258 0.13707455 0.12266355 0.19695862
[13,] 0.17880087 0.09710666 0.22859795 0.23732849 0.13405248
 [ reached getOption("max.print") -- omitted 62 rows ]
```

c. Run multidimensional scaling on each of your distance matrices with the "isoMDS" function in the MASS library. Report the stress values for each model and interpret them to evaluate which distance metric is performing better. Remember the stress value from R is a percentage.

fit = isoMDS(DISTANCE MATRIX)
fit$stress

```
> model  = isoMDS(distance, k=2)
initial  value 6.878292
final  value 6.877016
converged
> model$stress
[1] 6.877016
> # Compute MDS for this set by gower dist
> model2 = isoMDS(gdx, k=2)
initial  value 23.446502
iter   5 value 19.293034
final  value 19.076612
converged
> model2$stress
[1] 19.07661
>
```

Using Euclidean Distance, stress value is 6.87% means 0.06 which indicates the good fit in MDS (Multi-Dimensional Scaling).

Using Gower, Stress value (19.07%) of MDS of non-matric method seems high.

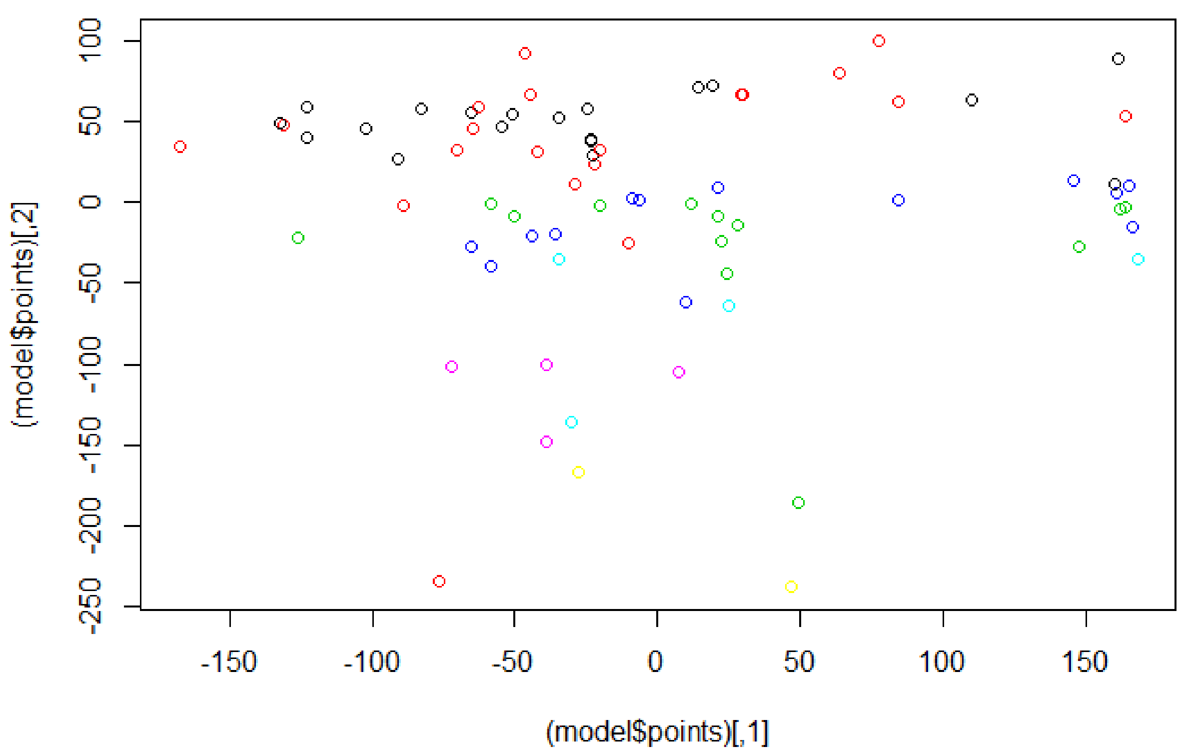
Syed Noor Razi Ali
2070326

Euclidian distance matric perform well because it gives us less stress value compared to Gower distance method.

d. For the model you chose to be the better performer in c), plot the MDS and color by "param8". Look at HClust.R line 58 for an example of how to do this:

plot(fit$points, col=(kellog$param8)+1)

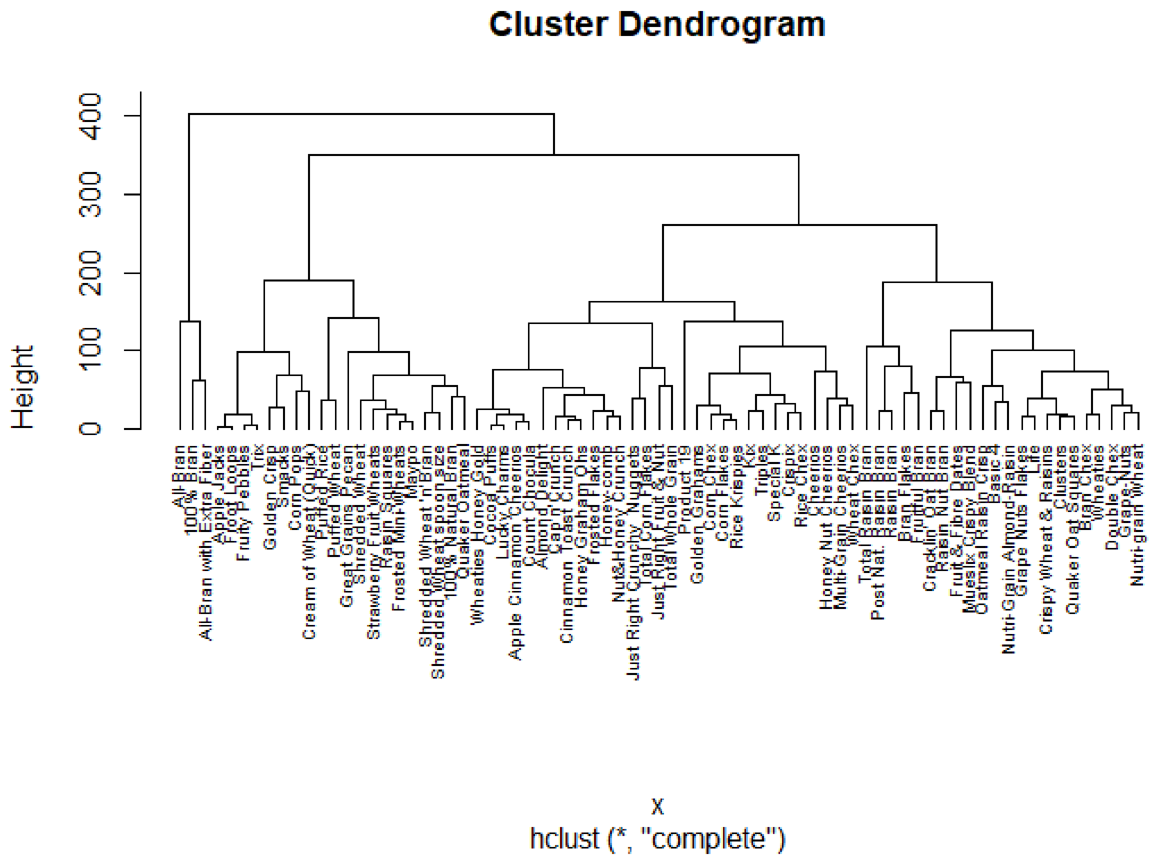
Include a screenshot of the graph and determine how many, if any, clusters you see emerging. How well separable are they?



e. Run agglomerative hierarchical clustering on the dataset using the hclust() function, and plot the results as a dendogram. Include a screenshot.

Use this before plotting the dendogram to get proper labels. Replace "kellog" with your dataframe: row.names(kellog)<-kellog[,1]

Syed Noor Razi Ali
2070326



f. By cutting off the tree, we can specify which cereal belongs to which cluster. Use this command:

clustCut = cutree(clust, k = 3)

- **Create a new variable in your original dataframe with the new cluster labels:**

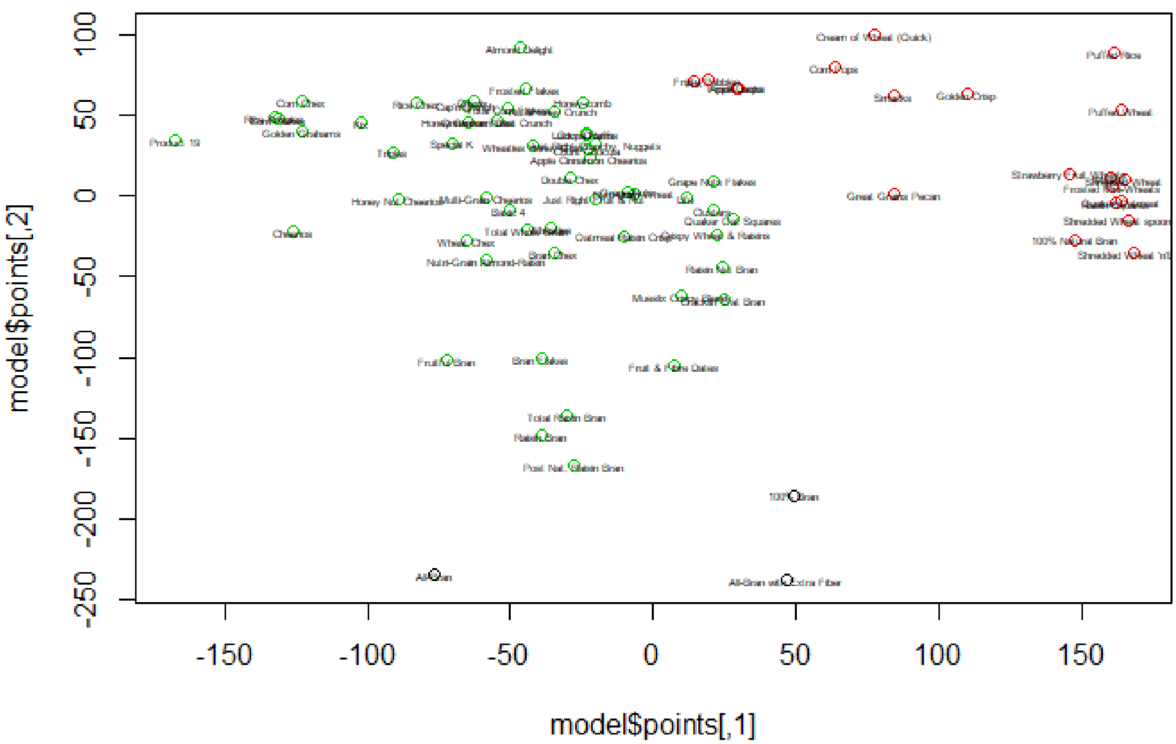
  library(dplyr)
  kellog = mutate(as.data.frame(kellog), cluster = clustCut)

- **Repeat part d), but use this new "cluster" variable for the color. You can also use this to get labels: text(fit$points, labels = kellog[,1], cex=.35)**

Based on the cereals included in each cluster, come up with a label/interpretation for what kinds of cereals that cluster represents. Look at what the cereals have in common and try to think of some characteristics unique to those cereals that might be defining the cluster.

Syed Noor Razi Ali
2070326

```
> cereal2 = mutate(as.data.frame(cereal), cluster = clustCut)
> head(cereal2)
                                               name brand type calories protein fat sodium fiber carbo sugars potass
100% Bran                                 100% Bran     N    C       70       4   1    130  10.0   5.0      6    280
100% Natural Bran                 100% Natural Bran     Q    C      120       3   5     15   2.0   8.0      8    135
All-Bran                                   All-Bran     K    C       70       4   1    260   9.0   7.0      5    320
All-Bran with Extra Fiber All-Bran with Extra Fiber     K    C       50       4   0    140  14.0   8.0      0    330
Almond Delight                       Almond Delight     R    C      110       2   2    200   1.0  14.0      8     -1
Apple Cinnamon Cheerios     Apple Cinnamon Cheerios     G    C      110       2   2    180   1.5  10.5     10     70
                          vitamins shelf weight cups   rating cluster
100% Bran                       25     3      1 0.33 68.40297       1
100% Natural Bran                0     3      1 1.00 33.98368       2
All-Bran                        25     3      1 0.33 59.42551       1
All-Bran with Extra Fiber       25     3      1 0.50 93.70491       1
Almond Delight                  25     3      1 0.75 34.38484       3
Apple Cinnamon Cheerios         25     1      1 0.75 29.50954       3
```



Syed Noor Razi Ali
2070326

```
> table(cereal2$cluster)

 1  2  3 
 3 20 52 
> table(cereal2$cluster,cereal$vitamins)
   
     0 25 100
  1  0  3   0
  2  8 12   0
  3  0 46   6
> table(cereal2$cluster,cereal$protein)
   
     1  2  3  4  5  6
  1  0  0  0  3  0  0
  2  4  8  6  1  1  0
  3  9 17 22  2  0  2
> table(cereal2$cluster,cereal$calories)
   
    50 70 80 90 100 110 120 130 140 160
  1  1  2  0  0   0   0   0   0   0   0
  2  2  0  1  4   5   6   2   0   0   0
  3  0  0  0  3  12  23   8   2   3   1
> 
```

2) Canonical Correlation Analysis

Water, soil, and mosquito fish samples were collected at 165 stations in southern Florida. The data is recorded in data_marsh_cleaned.csv.

The following variables apply to the water samples:

MEHGSWB Methyl Mercury in surface water, ng/L

TURB in situ surface water turbidity

DOCSWD Dissolved organic carbon in surface water, mg/L

SRPRSWFB Soluble reactive phosphorus in surface water, mg/L or ug/L

THGFSFC Total Mercury in mosquitofish, average of 7 individuals, ug/kg

The following variables apply to the soil samples:

THGSDFC Total mercury in soil, ng/g

TCSDFB Total carbon in soil, %

TPRSDFB Total phosphorus in soil, ug/g

a. Give the formulae for the first canonical variate for the soil and water variables.

Syed Noor Razi Ali
2070326

```
Canonical Correlation Analysis

Canonical Correlations:
     CV 1      CV 2      CV 3
0.3855843 0.3449978 0.2675698

X Coefficients:
                CV 1         CV 2         CV 3
THGSDFC -0.011415578 -0.010169482  0.014106076
TCSDFB   0.077556675 -0.037720634 -0.072787341
TPRSDFB  0.002969355  0.002268621  0.004222605

Y Coefficients:
                 CV 1         CV 2         CV 3
MEHGSWB  -0.720571333 -0.613310304 -0.442819677
TURB     -0.014902006  0.003947628 -0.046585662
DOCSWD    0.122898091 -0.045649299  0.038307498
SRPRSWFB 15.972715690 77.864165952 98.959103678
THGFSFC  -0.004124619 -0.009849176  0.009493841
```

<u>For soil:</u>
CV1: (-0.011416678) THGSDFC + (0.077556675) TCSDFB + (0.002969355) TPRSDFB

<u>For water:</u>
CV1: (-0.720571333) MEHGSWB + (0.014902006) TURB + (0.122898091) DOCSWD + (15.972715690) SRPRSWFB + (- 0.004124619) THGFSFC

Syed Noor Razi Ali
2070326

b. Give the correlations between the significant canonical variates for soils and the soil variables.

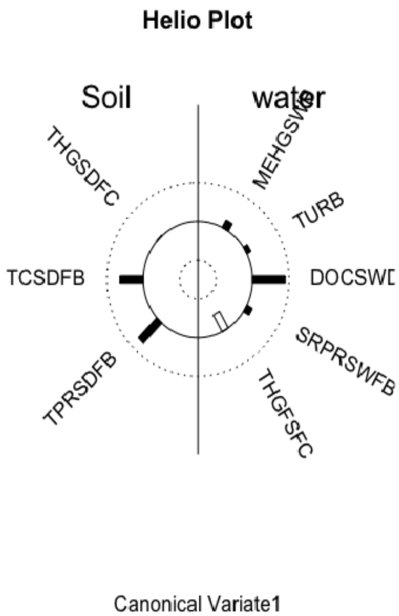
```
> round(-loadingsSoil$corr.Y.yscores, 2)
        [,1]  [,2]  [,3]
THGSDFC 0.01  0.88  0.47
TCSDFB  0.64  0.77 -0.04
TPRSDFB 0.71 -0.15  0.68
```

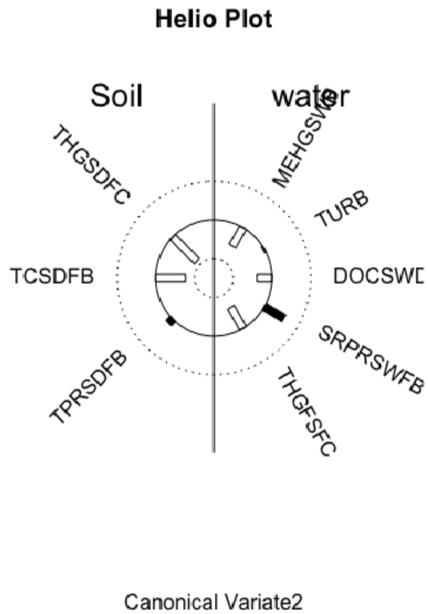
c. Give the correlations between the significant canonical variates for water and the water variables.

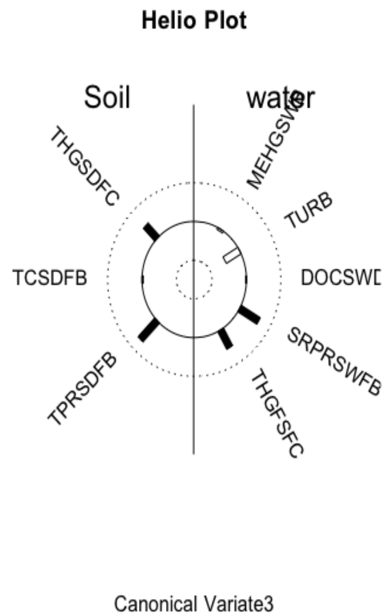
```
> round(-loadingsSoil$corr.X.xscores, 2)
          [,1]  [,2]  [,3]
MEHGSWB   0.21  0.54 -0.06
TURB      0.12  0.03 -0.50
DOCSWD    0.89  0.39 -0.02
SRPRSWFB  0.17 -0.58  0.64
THGFSFC  -0.49  0.62  0.53
```

d. Use parts b and c to interpret the variates. Do this as best you can. Even with a lack of domain knowledge, you should be able to draw some general conclusions based on the variables involved and the correlations.

Syed Noor Razi Ali
2070326







Increased DOCSWD breakdown organic carbon, which results to more phosphorus and carbon in soil, as seen in Canonical variation 1.

In Canonical variant 2 we can see that more Methyl Mercury and phosphorus in water leads to more mercury and carbon in soil.

We can see that more soluble reactive phosphorus in the water surface and Mercury in mosquitofish leads to higher mercury and phosphorus in the soil in Canonical variation 3.