DSC 424: Advanced Data Analysis and Regression

Assignment 05

Name: Sidhant Thakur

Student ID: 2020181

Problem 1

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.

```
            1          2          3          4          5          6          7          8          9         10         11         12
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 105.579302
13 258.209478 216.770780 286.950326 311.134239  50.339156  41.609109  88.137954  58.800210  91.067568 151.992114  14.634581 105.681701
14 182.318907 131.629867 250.355729 239.352024 121.968222  54.444184  77.217011  73.228902  67.015432 112.733270 109.247032 150.575534
           13         14         15         16         17         18         19         20         21         22         23         24
2
3
4
5
6
7
8
9
10
11
12
13
14  95.065558
           25         26         27         28         29         30         31         32         33         34         35         36
2
3
4
5
6
7
8
9
10
11
12
13
14
           37         38         39         40         41         42         43         44         45         46         47         48
2
```

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

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	0.0000000	0.3542882	0.1242007	0.1880165	0.33065282	0.37937856	0.37526758	0.33719262	0.30939583	0.22187507	0.41056813	0.4230254
[2,]	0.3542882	0.0000000	0.3753670	0.4408381	0.24377606	0.29179486	0.28560055	0.24287109	0.34364284	0.29519096	0.24419654	0.3816808
[3,]	0.1242007	0.3753670	0.0000000	0.1391599	0.32256494	0.37962401	0.32967971	0.32493808	0.30130795	0.20128719	0.39414692	0.3816042
[4,]	0.1880165	0.4408381	0.1391599	0.0000000	0.41303596	0.46176170	0.36848406	0.41957576	0.39177897	0.26925821	0.49295126	0.4637419
[5,]	0.3306528	0.2437761	0.3225649	0.4130360	0.00000000	0.17653229	0.19626884	0.15475818	0.16014588	0.21461108	0.17219460	0.3101274
[6,]	0.3793786	0.2917949	0.3796240	0.4617617	0.17653229	0.00000000	0.18318011	0.16459503	0.16998273	0.28555904	0.16452290	0.1866309
[7,]	0.3752676	0.2856005	0.3296797	0.3684841	0.19626884	0.18318011	0.00000000	0.27187294	0.25059398	0.27283696	0.17606464	0.3339088
[8,]	0.3371926	0.2428711	0.3249381	0.4195758	0.15475818	0.16459503	0.27187294	0.00000000	0.25446346	0.21698422	0.22949672	0.2291672
[9,]	0.3093958	0.3436428	0.3013080	0.3917790	0.16014588	0.16998273	0.25059398	0.25446346	0.00000000	0.19335409	0.25117230	0.2365280
[10,]	0.2218751	0.2951910	0.2012872	0.2692582	0.21461108	0.28555904	0.27283696	0.21698422	0.19335409	0.00000000	0.29591528	0.3211503
[11,]	0.4105681	0.2441965	0.3941469	0.4929513	0.17219460	0.16452290	0.17606464	0.22949672	0.25117230	0.29591528	0.00000000	0.3261538
[12,]	0.4230254	0.3816808	0.3816042	0.4637419	0.31012742	0.18663094	0.33390885	0.22916722	0.23652798	0.32115033	0.32615384	0.0000000
[13,]	0.4085128	0.2821412	0.3962582	0.4908959	0.16861190	0.10413423	0.20303750	0.15521916	0.24356141	0.28830439	0.10094422	0.2560429

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
[1,]	0.40851279	0.25313015	0.392177078	0.44701282	0.43375542	0.39283035	0.390462498	0.2374860	0.3799721	0.35449682	0.26804800	0.30437558
[2,]	0.28214121	0.21185380	0.304593376	0.38924432	0.39589417	0.29175042	0.302878796	0.1987467	0.3875124	0.30633353	0.22311588	0.29048045
[3,]	0.39625825	0.27004228	0.392422534	0.40559161	0.32566754	0.34724248	0.390707954	0.1877315	0.4683862	0.26307561	0.28496012	0.29212104
[4,]	0.49089593	0.33551329	0.474560217	0.47772930	0.38947189	0.38604683	0.472845637	0.2532025	0.4565219	0.31854663	0.35043114	0.35175872
[5,]	0.16861190	0.14418933	0.196859943	0.18260825	0.25995295	0.20115580	0.199173560	0.1865001	0.3470696	0.16837822	0.14593815	0.08349946
[6,]	0.10413423	0.14013729	0.106131853	0.21533389	0.22198373	0.20533999	0.104417273	0.2352259	0.3819353	0.26575643	0.11410834	0.22726488
[7,]	0.20303750	0.23074854	0.130729988	0.22173711	0.16373439	0.03882654	0.133043605	0.1952816	0.2987551	0.13882632	0.17805292	0.20978089
[8,]	0.15521916	0.08406247	0.195726883	0.29582778	0.30247763	0.27832460	0.194012303	0.1930399	0.3815958	0.21291699	0.11399763	0.17873057
[9,]	0.24356141	0.22464543	0.241114582	0.13761699	0.19102626	0.25704563	0.239400002	0.2397666	0.3535717	0.24510099	0.20587439	0.13247975
[10,]	0.28830439	0.18022175	0.290024231	0.29180442	0.27854702	0.27928862	0.288309651	0.1812443	0.3445358	0.19928842	0.18811737	0.15750051
[11,]	0.10094422	0.22966020	0.120057714	0.25980572	0.27597937	0.16600229	0.121772295	0.2664154	0.3859309	0.25308540	0.20918679	0.24119254
[12,]	0.25604295	0.22316031	0.257762789	0.21815409	0.20451790	0.34036051	0.256048209	0.3320522	0.3666649	0.28397142	0.26735277	0.30180473
[13,]	0.00000000	0.15538264	0.076335713	0.26038970	0.27656335	0.18741960	0.078050293	0.2376934	0.3823482	0.25366938	0.14879813	0.23358165

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]	[,36]
[1,]	0.35861919	0.42051444	0.30461433	0.26493793	0.30275418	0.36946168	0.4178902	0.39241717	0.24853627	0.26085513	0.26847610	0.40958551
[2,]	0.26895215	0.35959741	0.26335595	0.24073920	0.30136740	0.30854464	0.3220469	0.33150014	0.24394852	0.30509334	0.21102195	0.21654726
[3,]	0.31303131	0.34575990	0.25902645	0.27351672	0.21133297	0.38845713	0.4389690	0.35932930	0.25711506	0.25693392	0.28122155	0.39316430
[4,]	0.37850233	0.40956425	0.29783080	0.34732107	0.29180399	0.45392815	0.4777733	0.47480031	0.32258608	0.30970493	0.35649257	0.49196865
[5,]	0.17962045	0.19366827	0.24553849	0.19450276	0.22735318	0.17525439	0.2565698	0.17362739	0.17653877	0.21333102	0.20721460	0.17524018
[6,]	0.16482851	0.13786402	0.24704201	0.23489517	0.28719003	0.16883312	0.1749198	0.11470528	0.24470895	0.28150121	0.26982923	0.16354028
[7,]	0.01835160	0.09899686	0.13940326	0.26717308	0.19113461	0.13294409	0.1419012	0.18694188	0.22812020	0.26461246	0.28752381	0.16077689
[8,]	0.25522455	0.27128647	0.25207829	0.15558802	0.17360510	0.25287259	0.3266168	0.16680031	0.18356531	0.20723236	0.20163319	0.22851410
[9,]	0.23394558	0.18334084	0.20619702	0.25527739	0.26128767	0.23159362	0.2265499	0.21635468	0.19601737	0.23438340	0.26229138	0.25018968
[10,]	0.28285523	0.27780604	0.19993891	0.10847995	0.15587911	0.21383661	0.2487929	0.27220877	0.08249453	0.09201418	0.14286770	0.29493266
[11,]	0.15771304	0.16519299	0.26989824	0.24563019	0.26459172	0.11735687	0.2269613	0.12648429	0.26756519	0.30435744	0.28056425	0.02764928
[12,]	0.31726045	0.29332238	0.30292517	0.34756295	0.36357322	0.34157516	0.3347860	0.22633622	0.27587886	0.29674590	0.32033006	0.29850455
[13,]	0.18468590	0.17827696	0.26228735	0.25190819	0.28614750	0.13877418	0.2519200	0.05637340	0.25995430	0.29674655	0.24628669	0.11309395

	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]	[,47]	[,48]
[1,]	0.36466791	0.45240224	0.36023226	0.38081546	0.41898912	0.26894720	0.37115434	0.3875414	0.35707582	0.36173884	0.35774222	0.33121299
[2,]	0.30375088	0.36481854	0.29379128	0.31216133	0.35733685	0.19867605	0.28357063	0.3442683	0.28549854	0.31373185	0.29323730	0.26972067
[3,]	0.33574670	0.45264770	0.29797772	0.31856092	0.37756791	0.28169266	0.37139979	0.4002869	0.30315461	0.34531763	0.28715434	0.24812512
[4,]	0.44705105	0.50811871	0.37594873	0.39653194	0.47637226	0.35133034	0.45353747	0.4490913	0.37279229	0.44412198	0.37345869	0.34692946
[5,]	0.20932249	0.23791760	0.18042056	0.23453432	0.22340973	0.19503895	0.17583720	0.3362463	0.23302460	0.20946953	0.14963834	0.19369235
[6,]	0.04943287	0.16635702	0.25738084	0.29437117	0.18342641	0.18543135	0.08510911	0.3425117	0.27735324	0.08597305	0.15596119	0.27019587
[7,]	0.20041475	0.17289876	0.16503532	0.23748242	0.22052450	0.21354260	0.11804058	0.3021111	0.21883909	0.20658429	0.09211950	0.24414045
[8,]	0.15849550	0.31150760	0.22350429	0.15762284	0.19725364	0.20359285	0.17470414	0.3432129	0.15933175	0.17031531	0.20353747	0.10560084
[9,]	0.16804986	0.22356197	0.26333643	0.31058630	0.24292662	0.18371804	0.22009184	0.3040776	0.30927868	0.11900968	0.18584639	0.25049793
[10,]	0.25279285	0.25136050	0.25669053	0.27727373	0.29044738	0.20596102	0.26900149	0.3217084	0.26186742	0.24153044	0.24586715	0.21100459
[11,]	0.19173354	0.19114366	0.27144698	0.30510620	0.22991234	0.14162092	0.13274712	0.3510822	0.27142160	0.22549595	0.13615924	0.26426422
[12,]	0.16386474	0.28021017	0.33554019	0.38279006	0.17179705	0.23435413	0.23674005	0.3127264	0.38092931	0.13944577	0.30436126	0.31381523
[13,]	0.13551154	0.20422764	0.26383609	0.29749531	0.16181555	0.20623225	0.08902512	0.3434713	0.28881071	0.15538506	0.13056245	0.26082000

	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]	[,58]	[,59]	[,60]
[1,]	0.2816188	0.31018322	0.26502436	0.40488833	0.3998905	0.3644305	0.20893689	0.4370519	0.31224785	0.23547544	0.26656654	0.46439549
[2,]	0.2896852	0.23714497	0.29576389	0.35934417	0.2767206	0.2452248	0.19083852	0.3185569	0.31937373	0.20353503	0.26293109	0.40760311
[3,]	0.2114140	0.31459534	0.25693648	0.29680045	0.4149808	0.3834849	0.22793234	0.4609085	0.23332664	0.25238756	0.22097867	0.43130762
[4,]	0.2373353	0.39256636	0.34740750	0.36893814	0.3855312	0.3500712	0.29340336	0.4902684	0.32796432	0.31785858	0.25978302	0.48677863
[5,]	0.2245895	0.16831127	0.23841634	0.25409782	0.3153619	0.3146314	0.18838260	0.4327676	0.24535952	0.16184405	0.21714183	0.20298114
[6,]	0.2927597	0.14343776	0.28158652	0.34543374	0.4048935	0.3761755	0.25655278	0.3481600	0.23852970	0.14390312	0.28531202	0.23369268
[7,]	0.1825376	0.25134798	0.26219777	0.22454592	0.3217134	0.3038295	0.25174736	0.4352990	0.16914095	0.23729215	0.18300660	0.24009590
[8,]	0.2129353	0.06004699	0.18883493	0.26592763	0.3923318	0.3775027	0.19492240	0.4393074	0.19161144	0.10171718	0.24479274	0.31418657
[9,]	0.2299328	0.26642072	0.27794570	0.32215917	0.3755198	0.3606907	0.20433674	0.3353884	0.22425565	0.22692190	0.21541920	0.15499966
[10,]	0.1542303	0.22477482	0.13114826	0.24967993	0.3153728	0.3005437	0.16477848	0.4129546	0.21218043	0.17645592	0.17471665	0.30918709
[11,]	0.3156159	0.22083945	0.27287710	0.33679091	0.3060320	0.3278801	0.21274235	0.3793496	0.22009806	0.24175936	0.30816825	0.25149784
[12,]	0.2730044	0.27097789	0.39689792	0.31731748	0.4385121	0.4236829	0.28296346	0.3234735	0.33050787	0.25037629	0.35729219	0.23000343
[13,]	0.3080051	0.15489523	0.29582177	0.33334668	0.3691160	0.3924914	0.27179813	0.4439609	0.23748717	0.17303735	0.30055736	0.25208182

```
         [,61]     [,62]      [,63]      [,64]      [,65]     [,66]     [,67]      [,68]     [,69]      [,70]      [,71]      [,72]
 [1,] 0.44786172 0.3298409 0.29334390 0.30343930 0.35836918 0.3988284 0.2566562 0.39784982 0.3888211 0.32828604 0.33966315 0.38407675
 [2,] 0.38853390 0.3505879 0.34900210 0.34755960 0.27245214 0.3883204 0.2829811 0.33549025 0.3464317 0.30135048 0.30443975 0.29649304
 [3,] 0.33977384 0.4247788 0.38857631 0.39867171 0.31278130 0.3032405 0.3360683 0.38142861 0.3848999 0.31186483 0.30240861 0.40098887
 [4,] 0.41191152 0.4063907 0.35919706 0.37223742 0.37825232 0.3628782 0.3748727 0.45523295 0.4712043 0.38566918 0.39704629 0.46645989
 [5,] 0.25259269 0.3587564 0.37328334 0.36579854 0.19450975 0.2871022 0.2558022 0.19954903 0.3039578 0.23181122 0.15190237 0.18778613
 [6,] 0.21462347 0.2935932 0.30812019 0.30063539 0.17118921 0.2410766 0.2573057 0.22824648 0.2843699 0.24164807 0.17052932 0.11469819
 [7,] 0.15637412 0.2979545 0.33914811 0.33166330 0.05157424 0.1908837 0.1950003 0.26333047 0.3114389 0.29559265 0.24235047 0.12089250
 [8,] 0.29511736 0.3828658 0.35760092 0.35011612 0.26608565 0.3060150 0.2874813 0.17957371 0.1747094 0.16057325 0.11768988 0.19873766
 [9,] 0.20513256 0.2287522 0.21313746 0.21719057 0.24480668 0.2165009 0.1944788 0.28845399 0.3525547 0.24555688 0.23026732 0.24412536
[10,] 0.29265332 0.3130497 0.26472594 0.27482134 0.29371633 0.2702866 0.2121096 0.27347475 0.2930589 0.20391097 0.21112141 0.29303501
[11,] 0.25909530 0.3831161 0.39764309 0.39015829 0.16886562 0.3031287 0.2801619 0.28105164 0.3393228 0.32283764 0.22507164 0.12940847
[12,] 0.21483632 0.2957127 0.30065592 0.29317112 0.35478822 0.1471952 0.3003282 0.26899108 0.3696403 0.22864198 0.21663775 0.26077357
[13,] 0.25967927 0.3755052 0.39003220 0.38254740 0.19382438 0.2996844 0.2725510 0.21080228 0.2900452 0.24856008 0.15496075 0.08415911
         [,73]      [,74]      [,75]
 [1,] 0.33261679 0.34338366 0.37566731
 [2,] 0.34639452 0.32514175 0.31863400
 [3,] 0.30369558 0.32696245 0.36757944
 [4,] 0.38999993 0.40076680 0.45805045
 [5,] 0.18553603 0.25560249 0.16609662
 [6,] 0.19537288 0.13210601 0.05139024
 [7,] 0.27598413 0.25271726 0.18987805
 [8,] 0.23790916 0.17880896 0.14866156
 [9,] 0.05106337 0.13167727 0.13293815
[10,] 0.20407505 0.21900859 0.25962558
[11,] 0.26822911 0.27996225 0.19045637
[12,] 0.18546461 0.10630841 0.17069747
[13,] 0.26478489 0.20568469 0.11617881
 [ 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.

Multidimensional scaling using Euclidean distance

```
> model = isoMDS(x, k=2)
initial  value 6.878292
final  value 6.877016
converged
> model
$points
                                [,1]         [,2]
100% Bran                  49.399052 -185.4277764
100% Natural Bran         147.713630  -27.2441718
All-Bran                  -76.471415 -234.9514171
All-Bran with Extra Fiber  46.807191 -238.4597038
Almond Delight            -46.429732   92.0466060
Apple Cinnamon Cheerios   -21.798526   23.4987744
Apple Jacks                29.685665   67.4174455
Basic 4                   -50.373453   -8.5315188
Bran Chex                 -34.356571  -34.7988024
Bran Flakes               -39.046760 -100.5187778
Cap'n'Crunch              -65.542563   56.2387414
Cheerios                 -126.066646  -21.6921818
Cinnamon Toast Crunch     -54.815149   47.0091867
Clusters                   20.997484   -8.7850661
Cocoa Puffs               -23.471930   38.9777853
Corn Chex                -122.938055   59.4129008
Corn Flakes              -130.860176   47.9795329
Corn Pops                  63.607686   80.0365074
Count Chocula             -22.760594   29.0777182
Cracklin' Oat Bran         25.123970  -63.4968825
Cream of Wheat (Quick)     77.401349   99.8323441
Crispix                   -63.009794   58.8598792
Crispy Wheat & Raisins     22.634516  -23.6976803
Double Chex               -29.075964   11.4943207
Froot Loops                29.619451   67.4874990
Frosted Flakes            -44.756923   66.4855881
Frosted Mini-Wheats       160.758048    5.8900760
Fruit & Fibre Dates         7.513536 -104.5290901
Fruitful Bran             -72.102354 -101.0692568
Fruity Pebbles             19.044642   72.0277126
Golden Crisp              110.234667   63.5587380
Golden Grahams           -122.698883   40.7515550
Grape Nuts Flakes          21.215044    9.8837049
Grape-Nuts                 -8.735703    2.7411509
Great Grains Pecan         84.281759    1.3326936
Honey Graham Ohs          -64.504752   45.9961976
Honey Nut Cheerios        -89.201525   -2.1417983
Honey-comb                -24.518972   58.4199710
Just Right Crunchy  Nuggets -19.901390  32.4439477
Just Right Fruit & Nut    -19.900184   -1.3352293
Kix                      -102.269976   46.2908140
```

```
Kix                            -102.269976   46.2908140
Life                             11.637833   -0.3815562
Lucky Charms                    -23.162507   38.6694704
Maypo                           160.103059   11.2342953
Mueslix Crispy Blend             10.164063  -61.8139413
Multi-Grain Cheerios            -58.166327   -0.6347161
Nut&Honey Crunch                -34.750957   52.8754246
Nutri-Grain Almond-Raisin       -58.619124  -39.0510168
Nutri-grain Wheat                -6.610026    1.6620974
Oatmeal Raisin Crisp             -9.904918  -24.7117318
Post Nat. Raisin Bran           -27.714802 -167.1470693
Product 19                     -167.668488   34.5008295
Puffed Rice                     161.446299   89.0493654
Puffed Wheat                    164.261773   54.0464344
Quaker Oat Squares               27.801574  -14.3250962
Quaker Oatmeal                  163.830288   -3.3105511
Raisin Bran                     -38.986637 -148.1232774
Raisin Nut Bran                  24.402658  -43.8470907
Raisin Squares                  162.083870   -4.0699298
Rice Chex                       -83.125960   57.6865565
Rice Krispies                  -132.121415   48.7423740
Shredded Wheat                  165.362782   10.0004909
Shredded Wheat 'n'Bran          168.266278  -34.7949663
Shredded Wheat spoon size       166.642210  -14.7698469
Smacks                           84.506852   62.0667082
Special K                       -70.449680   32.6433206
Strawberry Fruit Wheats         146.141389   14.2322693
Total Corn Flakes               -51.203181   54.7580244
Total Raisin Bran               -30.179314 -136.5211730
Total Whole Grain               -44.141170  -20.7967200
Triples                         -90.950268   27.2598074
Trix                             14.097488   71.6420030
Wheat Chex                      -65.458984  -26.9852048
Wheaties                        -36.129010  -19.7202497
Wheaties Honey Gold             -41.835350   31.4226288

$stress
[1] 6.877016

> model$stress
[1] 6.877016
> 
```

Stress measures the of goodness of fit in multidimensional scaling. Here, it seems to be good i.e., 6.87%.

Multidimensional scaling using gower distance

```
initial  value 23.446502
iter   5 value 19.293034
final  value 19.076612
converged
> model
$points
             [,1]         [,2]
 [1,] -0.34344239 -0.194398156
 [2,] -0.30763462 -0.003864870
 [3,] -0.34687207 -0.188106735
 [4,] -0.42988493 -0.363617113
 [5,] -0.04083116  0.089767229
 [6,]  0.10643536  0.093375418
 [7,]  0.06304987  0.109783898
 [8,] -0.13488290  0.093445598
 [9,]  0.08774089 -0.044945389
[10,] -0.12701090 -0.081534655
[11,]  0.07439790  0.202075290
[12,]  0.31308907 -0.108229445
[13,]  0.06970183  0.200513497
[14,] -0.10976469  0.013242088
[15,]  0.07901856  0.160167628
[16,]  0.22338537 -0.003583315
[17,]  0.21789051 -0.037486945
[18,]  0.10975821  0.122322340
[19,]  0.07846649  0.160252067
[20,] -0.18220689  0.005663748
[21,]  0.10818279 -0.202624720
[22,]  0.01387958 -0.039231935
[23,] -0.07181205  0.048132523
[24,] -0.02767318 -0.023487761
[25,]  0.05313144  0.119232595
[26,]  0.16108866  0.106115990
[27,]  0.02141935 -0.114472730
[28,] -0.19820508  0.097199871
[29,] -0.20871538  0.093157964
[30,]  0.05890161  0.143294468
[31,]  0.21875648  0.098672039
[32,]  0.09070442  0.138693396
[33,] -0.06489640 -0.048063279
[34,] -0.08712042 -0.089740785
[35,] -0.17367937 -0.053390059
[36,]  0.07789746  0.175921862
[37,]  0.09388122  0.057317004
[38,]  0.21563733  0.138080346
[39,] -0.06471083  0.131889776
[40,] -0.19954555  0.201296891
```

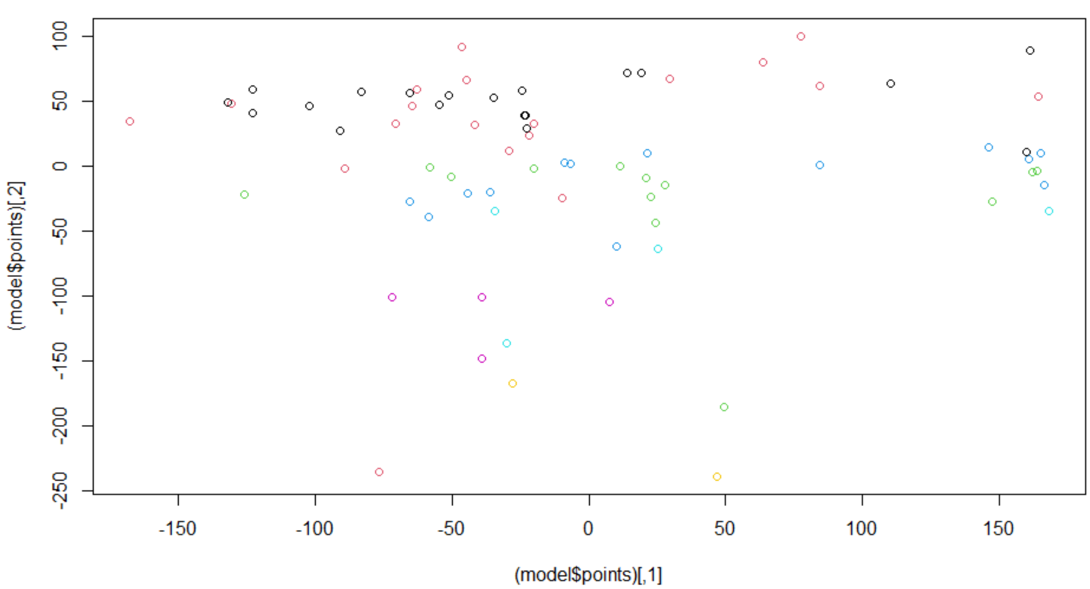
```
[41,]  0.17839117  0.045568586
[42,] -0.02105630 -0.010889897
[43,]  0.05321240  0.120396259
[44,]  0.06182996 -0.139235056
[45,] -0.25119182  0.174643985
[46,]  0.10780371 -0.004146728
[47,]  0.02852576  0.090255128
[48,] -0.18674353  0.062741909
[49,] -0.03681414 -0.114847287
[50,] -0.15923409  0.116532978
[51,] -0.25042246  0.140615622
[52,] -0.12766638 -0.181186643
[53,] -0.03842100 -0.383264932
[54,] -0.01912571 -0.383436073
[55,] -0.10224204 -0.049616544
[56,]  0.22901583 -0.338488130
[57,] -0.13376988  0.168245860
[58,] -0.13607662  0.010572216
[59,] -0.06232602 -0.140120689
[60,]  0.26493739  0.017279982
[61,]  0.22479876 -0.002160911
[62,]  0.23981247 -0.263159137
[63,]  0.20677191 -0.265336691
[64,]  0.18890663 -0.268674648
[65,]  0.03278051  0.154897345
[66,]  0.22011157 -0.087358718
[67,]  0.06024927 -0.125106354
[68,]  0.02338390  0.221774168
[69,] -0.26621532  0.279699783
[70,] -0.10947465 -0.103269740
[71,] -0.01787267  0.015575169
[72,]  0.08065950  0.142154100
[73,]  0.09082748 -0.073520238
[74,]  0.10566901 -0.072972492
[75,]  0.10343980  0.043002186

$stress
[1] 19.07661

> model$stress
[1] 19.07661
> 
```

Stress measures the of goodness of fit in multidimensional scaling. Here, it seems to be bad i.e., 19.07%

As a result, multidimensional scaling using the Euclidean distance metric outperforms multidimensional scaling using the Gower distance metric because the stress value in Euclidean distance is lower as compared to Gower distance

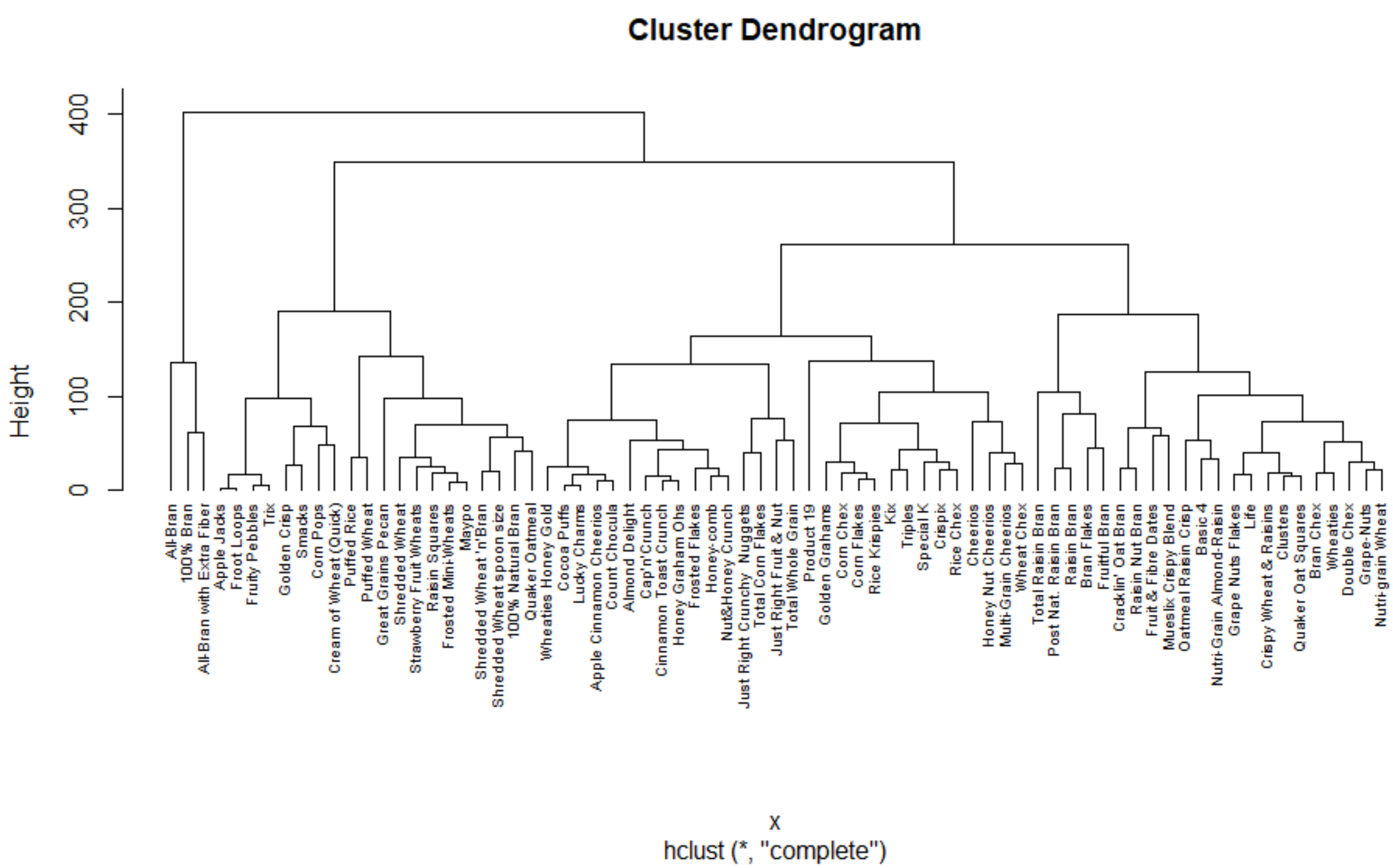
d. For the model you chose to be the better performer in c), plot the MDS and color by "param8". Look at HClust.R line 58 for an example of how to do this:
plot (fit$points, col=(kellog$param8)+1)
Include a screenshot of the graph and determine how many, if any, clusters you see emerging. How well separable are they?

I went with a multidimensional scaling model based on Euclidean distance.



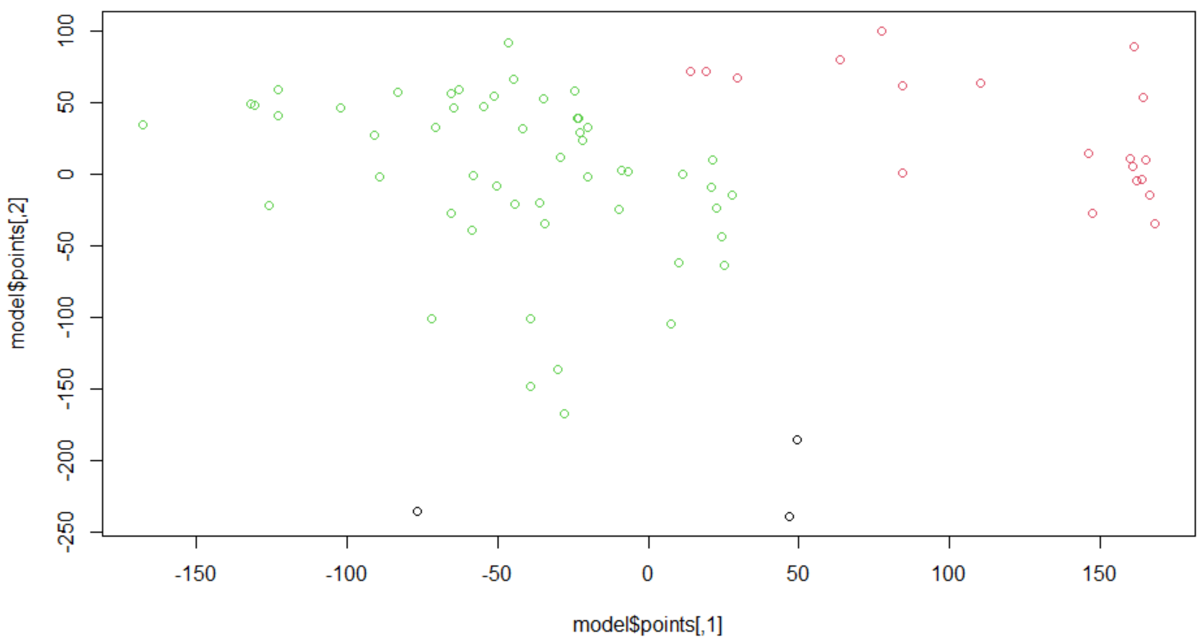
The above figure clearly shows that we have three clusters in black, red, and green colors, but these three clusters are difficult to separate because they are mixed together.

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

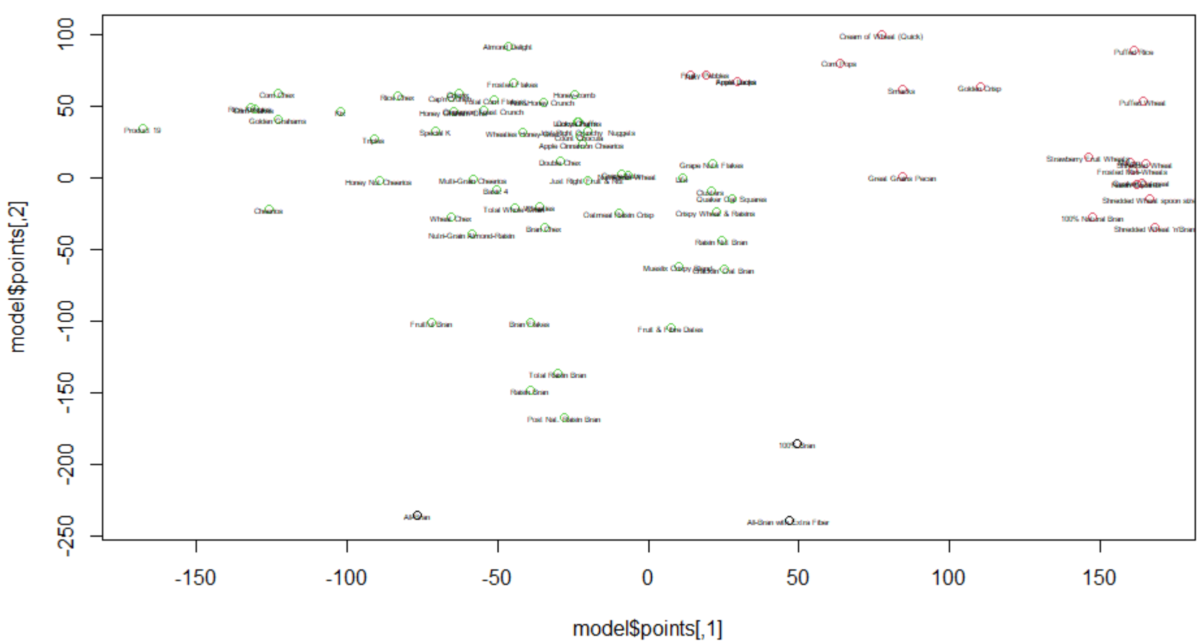


f. By cutting off the tree, we can specify which cereal belongs to which cluster. Use this command:
clustCut = cutree(clust, k = 3)

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)


model$points[,2]
model$points[,1]
100
50
0
-50
-100
-150
-200
-250
-150
-100
-50
0
50
100
150

With Text:


model$points[,2]
model$points[,1]
100
50
0
-50
-100
-150
-200
-250
-150
-100
-50
0
50
100
150

```
> table(cereal$name,cereal$cluster)

                               1 2 3
  100% Bran                    1 0 0
  100% Natural Bran            0 1 0
  All-Bran                     1 0 0
  All-Bran with Extra Fiber    1 0 0
  Almond Delight               0 0 1
  Apple Cinnamon Cheerios      0 0 1
  Apple Jacks                  0 1 0
  Basic 4                      0 0 1
  Bran Chex                    0 0 1
  Bran Flakes                  0 0 1
  Cap'n'Crunch                 0 0 1
  Cheerios                     0 0 1
  Cinnamon Toast Crunch        0 0 1
  Clusters                     0 0 1
  Cocoa Puffs                  0 0 1
  Corn Chex                    0 0 1
  Corn Flakes                  0 0 1
  Corn Pops                    0 1 0
  Count Chocula                0 0 1
  Cracklin' Oat Bran           0 0 1
  Cream of Wheat (Quick)       0 1 0
  Crispix                      0 0 1
  Crispy Wheat & Raisins       0 0 1
  Double Chex                  0 0 1
  Froot Loops                  0 1 0
  Frosted Flakes               0 0 1
  Frosted Mini-Wheats          0 1 0
  Fruit & Fibre Dates          0 0 1
  Fruitful Bran                0 0 1
  Fruity Pebbles               0 1 0
  Golden Crisp                 0 1 0
  Golden Grahams               0 0 1
  Grape-Nuts                   0 0 1
  Grape Nuts Flakes            0 0 1
  Great Grains Pecan           0 1 0
  Honey-comb                   0 0 1
  Honey Graham Ohs             0 0 1
  Honey Nut Cheerios           0 0 1
  Just Right Crunchy  Nuggets  0 0 1
  Just Right Fruit & Nut       0 0 1
  Kix                          0 0 1
  Life                         0 0 1
  Lucky Charms                 0 0 1
  Maypo                        0 1 0
```

```
Maypo                         0 1 0
Mueslix Crispy Blend          0 0 1
Multi-Grain Cheerios          0 0 1
Nut&Honey Crunch              0 0 1
Nutri-Grain Almond-Raisin     0 0 1
Nutri-grain Wheat             0 0 1
Oatmeal Raisin Crisp          0 0 1
Post Nat. Raisin Bran         0 0 1
Product 19                    0 0 1
Puffed Rice                   0 1 0
Puffed Wheat                  0 1 0
Quaker Oat Squares            0 0 1
Quaker Oatmeal                0 1 0
Raisin Bran                   0 0 1
Raisin Nut Bran               0 0 1
Raisin Squares                0 1 0
Rice Chex                     0 0 1
Rice Krispies                 0 0 1
Shredded Wheat                0 1 0
Shredded Wheat 'n'Bran        0 1 0
Shredded Wheat spoon size     0 1 0
Smacks                        0 1 0
Special K                     0 0 1
Strawberry Fruit Wheats       0 1 0
Total Corn Flakes             0 0 1
Total Raisin Bran             0 0 1
Total Whole Grain             0 0 1
Triples                       0 0 1
Trix                          0 1 0
Wheat Chex                    0 0 1
Wheaties                      0 0 1
Wheaties Honey Gold           0 0 1
> 
```

The above table and graphs show three clusters: green, red, and black. Two large clusters of green and red, and one small black cluster

Black cluster consist of 3 cereals: 100% Bran, All-Bran, All-Bran with Extra Fiber

Red cluster consist of following cereals: 100% Natural Bran, Apple Jacks, Corn Pops, Cream of Wheat (Quick), Froot Loops, Frosted Mini-Wheats, Fruity Pebbles, Golden Crisp etc.

Green cluster consist of following cereals: Almond Delight, Apple Cinnamon Cheerios, Basic 4, Bran Chex, Bran Flakes, Cap'n'Crunch, Cheerios etc.

Problem 2

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

```
> c = cancor(soil, water)
> c
$cor
[1] 0.3855843 0.3449978 0.2675698

$xcoef
                [,1]          [,2]          [,3]
THGSDFC -0.0008914069 -0.0007941031  0.0011014995
TCSDFB   0.0060561589 -0.0029454867 -0.0056837364
TPRSDFB  0.0002318677  0.0001771496  0.0003297301

$ycoef
                 [,1]          [,2]          [,3]          [,4]          [,5]
MEHGSWB  -0.0562671679 -0.0478914887 -0.0345784074  1.343294e-01  0.1393652286
TURB     -0.0011636512  0.0003082580 -0.0036377291  4.729215e-03 -0.0044666489
DOCSWD    0.0095967286 -0.0035646114  0.0029913130 -3.350867e-03 -0.0033336791
SRPRSWFB  1.2472595484  6.0801698565  7.7274077473  6.386519e+00 -0.3976446079
THGFSFC  -0.0003220786 -0.0007690914  0.0007413444  2.748318e-05 -0.0005735501
```

Formulae for the canonical variate for the soil is
-0.0008914069 THGSDFC + 0.0060561589 TCSDFB + 0.0002318677 TPRSDFB

Formulae for the canonical variate for the water is
-0.0562671679 MEHGSWB -0.0011636512 TURB+ 0.0095967286 DOCSWD+ 1.2472595484 SRPR-0.000322078 THGFSFC

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

```
> c = matcor(soil, water)
> c
$Xcor
          THGSDFC    TCSDFB   TPRSDFB
THGSDFC 1.0000000 0.6677804 0.1966074
TCSDFB  0.6677804 1.0000000 0.3178176
TPRSDFB 0.1966074 0.3178176 1.0000000

$Ycor
             MEHGSWB        TURB      DOCSWD    SRPRSWFB     THGFSFC
MEHGSWB   1.00000000  0.04286195  0.53653344 -0.05729504  0.04523356
TURB      0.04286195  1.00000000  0.26262016 -0.03127880 -0.08426556
DOCSWD    0.53653344  0.26262016  1.00000000  0.01784706 -0.20284406
SRPRSWFB -0.05729504 -0.03127880  0.01784706  1.00000000 -0.08581679
THGFSFC   0.04523356 -0.08426556 -0.20284406 -0.08581679  1.00000000

$XYcor
             THGSDFC      TCSDFB     TPRSDFB     MEHGSWB        TURB
THGSDFC   1.00000000  0.66778043  0.19660738  0.15971021 -0.05151880
TCSDFB    0.66778043  1.00000000  0.31781764  0.19749008  0.04374098
TPRSDFB   0.19660738  0.31781764  1.00000000  0.02092839 -0.05980083
MEHGSWB   0.15971021  0.19749008  0.02092839  1.00000000  0.04286195
TURB     -0.05151880  0.04374098 -0.05980083  0.04286195  1.00000000
DOCSWD    0.11909492  0.32344092  0.22121653  0.53653344  0.26262016
SRPRSWFB -0.09647552 -0.11800127  0.19411633 -0.05729504 -0.03127880
THGFSFC   0.25310209  0.03809560 -0.07060351  0.04523356 -0.08426556
              DOCSWD    SRPRSWFB     THGFSFC
THGSDFC   0.11909492 -0.09647552  0.25310209
TCSDFB    0.32344092 -0.11800127  0.03809560
TPRSDFB   0.22121653  0.19411633 -0.07060351
MEHGSWB   0.53653344 -0.05729504  0.04523356
TURB      0.26262016 -0.03127880 -0.08426556
DOCSWD    1.00000000  0.01784706 -0.20284406
SRPRSWFB  0.01784706  1.00000000 -0.08581679
THGFSFC  -0.20284406 -0.08581679  1.00000000
```

```
      WilksL    F df1     df2    p
[1,]   0.70 4.05  15 433.81 0.00
[2,]   0.82 4.18   8 316.00 0.00
[3,]   0.93 4.09   3 159.00 0.01
```

Since all the three correlations have small p value, it means we need to reject null hypothesis, which means the correlations are not equal to zero.

```
> round(-loadingsSales$corr.X.xscores, 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.

```
          [,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.

To Enhance the amount of carbon and phosphorus in soil can be done by adding more dissolved organic carbon to surface water and by reducing mercury to mosquitofish, in Canonical variant 1.

In Canonical variant 2, we need to add more methyl mercury in surface water, more mercury in mosquitofish, and less soluble reactive phosphorus in surface water led to more mercury in soil and more carbon in soil.

In canonical variant 3, adding soluble reactive phosphorus in surface water and mercury in mosquitofish, as well as reducing situ surface water turbidity, can be used to increase the quantity of mercury and phosphorus in soil.