

Alex Look

DSC 324

3/13/22

### Assignment 5

1a) Euclidean distance of matrix on columns 2-11

	1	2	3	4	5
2	0.7272322				
3	1.5697390	1.8492052			
4	1.8539219	2.0305045	1.2684015		
5	1.6662546	1.8829115	0.2975358	1.1916989	
6	1.0996082	1.5440580	1.3176119	1.7678896	1.3547525
7	1.5841069	1.7634645	1.1490144	1.0299818	1.0490237
8	1.5091412	1.8468267	0.3398889	1.2719303	0.4316207
9	1.8023558	2.0544616	0.6520781	0.8403389	0.6142367
10	1.4347282	1.4944984	0.7279911	1.2266398	0.6708088
11	1.6882887	1.9283144	1.3544907	1.5519973	1.3081891
12	1.7672470	2.0343733	1.4196422	1.6397532	1.4069083
13	1.4552656	1.8747545	1.1436096	1.6629513	1.1815905
14	1.4997064	1.8448304	0.5907124	0.9354137	0.5692408
15	1.4212580	1.7941160	1.2974982	1.3748080	1.2789268
16	1.2550124	1.3637588	1.1218121	1.1317151	1.0642234
17	1.7728810	1.9876043	1.6030319	1.4348174	1.5871723
18	1.0511532	1.4668203	0.9100617	1.2953081	1.0048503
19	1.4007230	1.4490070	0.9288793	1.4575319	0.8181519
20	1.9205121	2.1198904	1.2636303	0.1492539	1.1866194
21	1.5809483	1.8995743	0.4048276	1.4944746	0.5198887
22	1.6873679	1.8735937	1.2857019	0.8961782	1.3616727

	6	7	8	9	10
2					
3					
4					
5					
6					
7	1.3903578				
8	1.0596577	1.1567554			
9	1.6040403	1.2435004	0.7170009		
10	1.2503425	1.0635731	0.7637698	0.9792681	
11	1.3658479	1.1206756	1.2872281	1.5243700	1.3316102
12	1.4332747	1.2096828	1.3654066	1.6087575	1.3995626
13	0.8375832	1.2044443	0.9849233	1.4502875	1.2038583
14	1.0618428	0.8392057	0.4448509	0.6633024	0.8014891
15	0.8945117	0.8685676	1.1368609	1.4514440	1.2355117
16	1.2221023	0.4469872	1.1258005	1.3028190	0.8507743
17	1.7389280	1.0736413	1.6085895	1.6605647	1.5915576
18	1.0022757	1.2363397	0.8389299	1.0476095	0.9541119
19	1.1943371	0.8817143	0.9525259	1.2882294	0.5597161
20	1.7930692	1.0284355	1.2706732	0.8331199	1.2510384
21	1.1255875	1.3644799	0.2571313	0.8664377	0.8128402
22	1.6825281	1.3261089	1.2926246	1.1490633	1.1248803

	11	12	13	14	15
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12	0.4606777				
13	1.3053475	1.1438155			
14	1.1529193	1.2206573	0.9130613		
15	1.1818840	1.0842522	0.5668569	0.8665760	
16	1.1355754	1.2484837	1.2027871	0.8995366	0.9183382
17	0.6792585	0.8088116	1.6330287	1.3695575	1.3585262
18	1.3902247	1.3279287	0.8123873	0.7950756	0.9329151
19	1.2146529	1.3392266	1.1730249	0.9484267	1.1565091
20	1.5538330	1.6129819	1.6268183	0.9192802	1.3498559
21	1.4129162	1.4820527	1.0690075	0.6884662	1.3131275
22	1.6825437	1.6832981	1.6302255	1.0940138	1.4632114

	16	17	18	19	20
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17	1.1365256				
18	1.1051015	1.5842025			
19	0.6653321	1.4995015	1.1368606		
20	1.1690533	1.4425548	1.2965719	1.4802744	
21	1.2997845	1.7882332	0.9427259	1.0090331	1.4934093
22	1.2180429	1.5873017	1.2445166	1.5069150	0.9189203

	21
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	1.4530108

1b) Gower distance matrix

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.00000	0.16795	0.43925	0.47239	0.47341	0.28121
[2,]	0.16795	0.00000	0.45220	0.52034	0.48636	0.36082
[3,]	0.43925	0.45220	0.00000	0.25894	0.06082	0.29138
[4,]	0.47239	0.52034	0.25894	0.00000	0.26644	0.44440
[5,]	0.47341	0.48636	0.06082	0.26644	0.00000	0.32554
[6,]	0.28121	0.36082	0.29138	0.44440	0.32554	0.00000
[7,]	0.37955	0.41584	0.21968	0.13926	0.22718	0.30836
[8,]	0.39926	0.47887	0.03999	0.28561	0.08749	0.25139
[9,]	0.48743	0.53788	0.13506	0.16712	0.09932	0.37706
[10,]	0.38568	0.39863	0.17169	0.27589	0.15733	0.25601
[11,]	0.41885	0.51722	0.30040	0.34132	0.30790	0.30308
[12,]	0.44768	0.54605	0.36611	0.40703	0.37361	0.31857
[13,]	0.38401	0.46988	0.29282	0.43958	0.31366	0.17572
[14,]	0.39194	0.50281	0.13309	0.22819	0.14059	0.27533
[15,]	0.34030	0.46991	0.35349	0.32319	0.36099	0.17575
[16,]	0.31001	0.30172	0.26256	0.21862	0.27006	0.24396
[17,]	0.44521	0.50650	0.37820	0.24592	0.38570	0.41722
[18,]	0.27880	0.40217	0.21863	0.35455	0.22613	0.24059
[19,]	0.35050	0.36345	0.22019	0.29773	0.20583	0.25235
[20,]	0.48863	0.54992	0.25698	0.02958	0.26448	0.44244
[21,]	0.41321	0.49282	0.07374	0.32946	0.11270	0.27868
[22,]	0.43503	0.47758	0.27754	0.15430	0.28504	0.41010

	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	0.37955	0.39926	0.48743	0.38568	0.41885	0.44768
[2,]	0.41584	0.47887	0.53788	0.39863	0.51722	0.54605
[3,]	0.21968	0.03999	0.13506	0.17169	0.30040	0.36611
[4,]	0.13926	0.28561	0.16712	0.27589	0.34132	0.40703
[5,]	0.22718	0.08749	0.09932	0.15733	0.30790	0.37361
[6,]	0.30836	0.25139	0.37706	0.25601	0.30308	0.31857
[7,]	0.00000	0.24635	0.22786	0.25805	0.20528	0.27099
[8,]	0.24635	0.00000	0.16173	0.19836	0.26041	0.32612
[9,]	0.22786	0.16173	0.00000	0.22675	0.32732	0.39303
[10,]	0.25805	0.19836	0.22675	0.00000	0.31943	0.33492
[11,]	0.20528	0.26041	0.32732	0.31943	0.00000	0.10571
[12,]	0.27099	0.32612	0.39303	0.33492	0.10571	0.00000
[13,]	0.30354	0.25283	0.35892	0.29745	0.29826	0.23255
[14,]	0.19215	0.09310	0.14751	0.19168	0.21227	0.25980
[15,]	0.18715	0.31350	0.36167	0.27812	0.26643	0.18738
[16,]	0.11412	0.28923	0.28948	0.17545	0.22882	0.24755
[17,]	0.17184	0.40487	0.38638	0.34871	0.15414	0.17985
[18,]	0.31851	0.17864	0.24015	0.20992	0.33823	0.30888
[19,]	0.17989	0.24686	0.27525	0.10278	0.24127	0.31010
[20,]	0.13730	0.28365	0.16516	0.29213	0.33936	0.40507
[21,]	0.29342	0.04707	0.20880	0.20463	0.30104	0.36675
[22,]	0.22452	0.30421	0.18572	0.25979	0.36036	0.42607

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
[1,]	0.38401	0.39194	0.34030	0.31001	0.44521	0.27880
[2,]	0.46988	0.50281	0.46991	0.30172	0.50650	0.40217
[3,]	0.29282	0.13309	0.35349	0.26256	0.37820	0.21863
[4,]	0.43958	0.22819	0.32319	0.21862	0.24592	0.35455
[5,]	0.31366	0.14059	0.36099	0.27006	0.38570	0.22613
[6,]	0.17572	0.27533	0.17575	0.24396	0.41722	0.24059
[7,]	0.30354	0.19215	0.18715	0.11412	0.17184	0.31851
[8,]	0.25283	0.09310	0.31350	0.28923	0.40487	0.17864
[9,]	0.35892	0.14751	0.36167	0.28948	0.38638	0.24015
[10,]	0.29745	0.19168	0.27812	0.17545	0.34871	0.20992
[11,]	0.29826	0.21227	0.26643	0.22882	0.15414	0.33823
[12,]	0.23255	0.25980	0.18738	0.24755	0.17985	0.30888
[13,]	0.00000	0.25235	0.11639	0.23914	0.41240	0.20521
[14,]	0.25235	0.00000	0.23468	0.24109	0.34423	0.15314
[15,]	0.11639	0.23468	0.00000	0.19399	0.30933	0.23440
[16,]	0.23914	0.24109	0.19399	0.00000	0.20478	0.27911
[17,]	0.41240	0.34423	0.30933	0.20478	0.00000	0.42737
[18,]	0.20521	0.15314	0.23440	0.27911	0.42737	0.00000
[19,]	0.28047	0.22686	0.26114	0.13339	0.31055	0.30626
[20,]	0.43762	0.21195	0.32123	0.24820	0.26216	0.35259
[21,]	0.29346	0.13373	0.35413	0.32986	0.44550	0.21927
[22,]	0.41862	0.24357	0.35555	0.24252	0.32712	0.32025
	[,19]	[,20]	[,21]	[,22]		
[1,]	0.35050	0.48863	0.41321	0.43503		
[2,]	0.36345	0.54992	0.49282	0.47758		
[3,]	0.22019	0.25698	0.07374	0.27754		
[4,]	0.29773	0.02958	0.32946	0.15430		
[5,]	0.20583	0.26448	0.11270	0.28504		
[6,]	0.25235	0.44244	0.27868	0.41010		
[7,]	0.17989	0.13730	0.29342	0.22452		
[8,]	0.24686	0.28365	0.04707	0.30421		
[9,]	0.27525	0.16516	0.20880	0.18572		
[10,]	0.10278	0.29213	0.20463	0.25979		
[11,]	0.24127	0.33936	0.30104	0.36036		
[12,]	0.31010	0.40507	0.36675	0.42607		
[13,]	0.28047	0.43762	0.29346	0.41862		
[14,]	0.22686	0.21195	0.13373	0.24357		
[15,]	0.26114	0.32123	0.35413	0.35555		
[16,]	0.13339	0.24820	0.32986	0.24252		
[17,]	0.31055	0.26216	0.44550	0.32712		
[18,]	0.30626	0.35259	0.21927	0.32025		
[19,]	0.00000	0.31397	0.25313	0.32163		
[20,]	0.31397	0.00000	0.32750	0.15234		
[21,]	0.25313	0.32750	0.00000	0.34484		
[22,]	0.32163	0.15234	0.34484	0.00000		

1c) Stress values

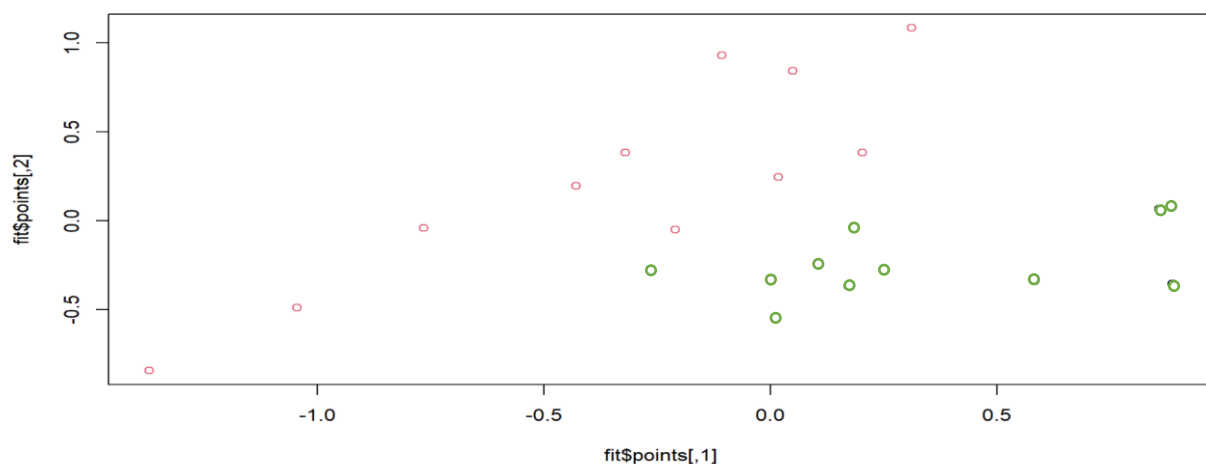
```

initial value 19.915627
iter 5 value 14.639105
iter 10 value 14.220989
final value 14.179482
converged
initial value 19.501867
iter 5 value 15.562812
iter 10 value 15.308617
iter 15 value 14.798679
final value 14.719101
converged
[1] 14.17948
[1] 14.7191

```

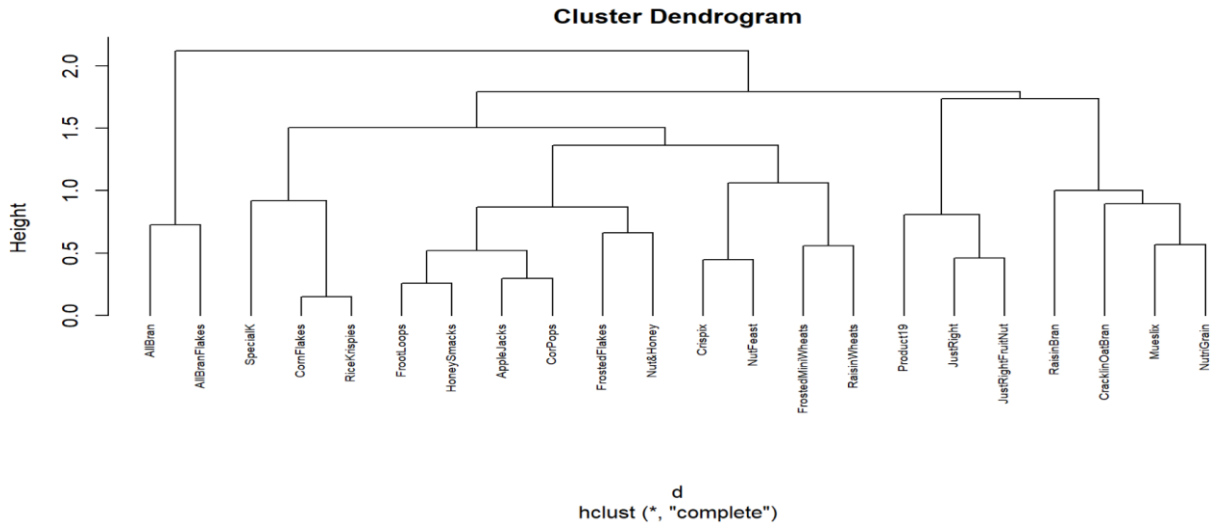
The stress test results of the Euclidean distance are 14.17948. The Gower distance was higher at 14.7191. There is not much of a different between the two but technically speaking Euclidean is better handling stress than Gower.

1d)

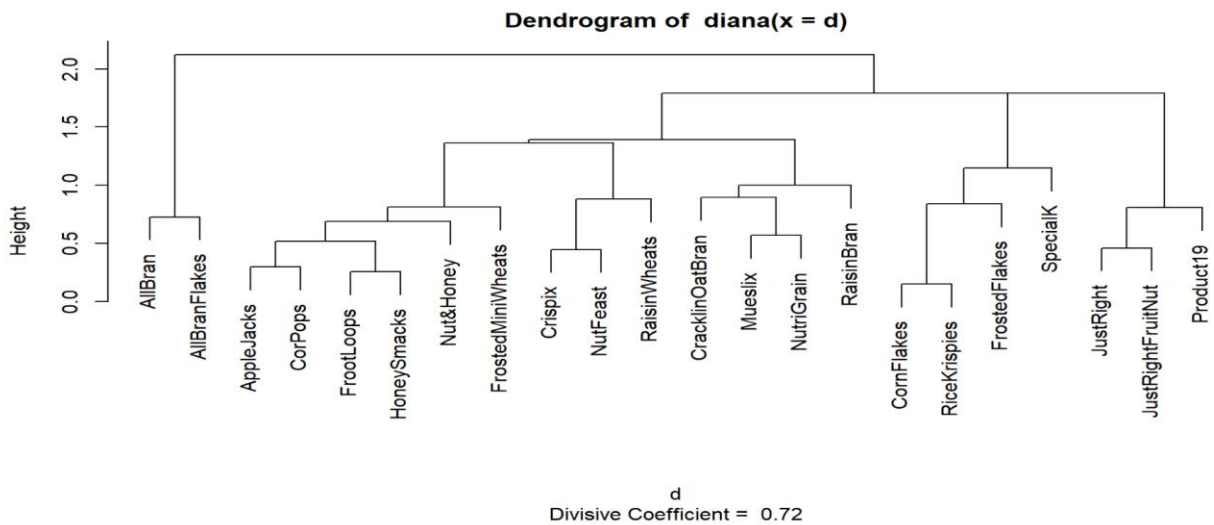


These colors should be reversed not sure why R messed it up for me. The graph in 1f) is the right one where the colors should be represented correctly. When looking at potential clusters there could be a of what is supposed to be of red cluster creating around  $x=0.0-0.3$  and  $y=-0.5-0.0$ . I would say this is a cluster because of the condense space that multiple points are in. Another possible cluster is from  $x=-0.6-0.25$  and  $y=0.0-1.0$  that is green. This is not as likely of one because of the space in between many of the points themselves from an eye view.

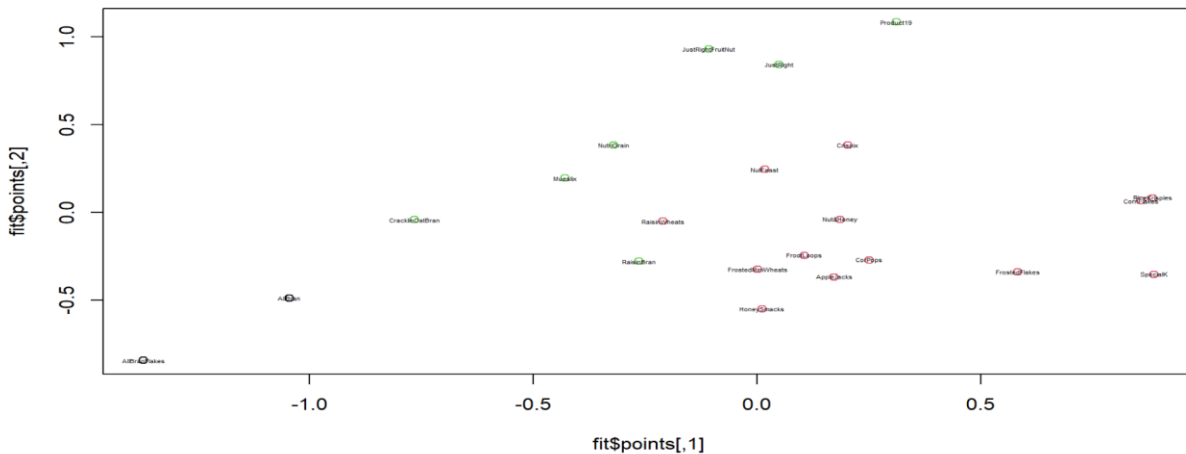
1e)



1f)



From this dendrogram when looking at the lowest level clusters there is grouping types. There is those sugar covered cereals AppleJack, CorPops, FrootLoops, Frosted Mini Wheats, Nuts&Honey. From there is the fruit fiber ceareals NutriGran, RaisinBran, CracklinOatBran. There is the AllBran group with AllBranFlakes. There is the corn-based cereals CornFlakes, RiceKrispies, FrostedFlakes, SpecialK. Then there is this JustRight and JustRightFruitNut grouping. From there the next level starts to combine these individual clusters. It would be sugar meets sugar high fiber grouping. There are the named brand groupings with RiceKripsies SpecialK and JustRight products. From there the overarching group is cereal.



From looking at this cluster graph it seems that cluster in red between -.3 - .3 turned out to be a lot of high sugar with fruits cereals which makes sense. Any natural or fake fruit product will have higher sugar content which would mean that is the close cluster. When trying to think of the cluster above in green the red cluster was the JustRight brand and nut cereals. It makes sense seeing nut cereal seems like a close cluster but not always connected. Sometimes it could be straight nuts but other times it could be nuts and fruit because some points went further down.

2a)

```
$cor
[1] 0.3855843 0.3449978 0.2675698
```

The equation could look like  $0.3855843 + 0.3449978 * \text{variable1} + 0.265698 \text{ variable2}$

2b)

```
$Ycor
as.double.df_marsh.THGSDFC. as.double.df_marsh.THGSDFC.
as.double.df_marsh.THGSDFC. 1.0000000
as.double.df_marsh.TCSDFB. 0.6677804
as.double.df_marsh.TPRSDFB. 0.1966074
as.double.df_marsh.TCSDFB. as.double.df_marsh.TCSDFB.
as.double.df_marsh.THGSDFC. 0.6677804
as.double.df_marsh.TCSDFB. 1.0000000
as.double.df_marsh.TPRSDFB. 0.3178176
as.double.df_marsh.TPRSDFB. as.double.df_marsh.TPRSDFB.
as.double.df_marsh.THGSDFC. 0.1966074
as.double.df_marsh.TCSDFB. 0.3178176
as.double.df_marsh.TPRSDFB. 1.0000000
```

2c)



```

$Xcor
as.double.df_marsh.MEHGSWB. 1.00000000
as.double.df_marsh.TURB. 0.04286195
as.double.df_marsh.DOCSWD. 0.53653344
as.double.df_marsh.SRPRSWFB. -0.05729504
as.double.df_marsh.THGFSFC. 0.04523356
as.double.df_marsh.TURB. 0.04286195
as.double.df_marsh.MEHGSWB. 1.00000000
as.double.df_marsh.TURB. 0.26262016
as.double.df_marsh.DOCSWD. -0.03127880
as.double.df_marsh.SRPRSWFB. -0.08426556
as.double.df_marsh.THGFSFC. 0.53653344
as.double.df_marsh.MEHGSWB. 0.26262016
as.double.df_marsh.TURB. 1.00000000
as.double.df_marsh.DOCSWD. 0.01784706
as.double.df_marsh.SRPRSWFB. -0.20284406
as.double.df_marsh.THGFSFC. -0.05729504
as.double.df_marsh.MEHGSWB. -0.03127880
as.double.df_marsh.TURB. 0.01784706
as.double.df_marsh.DOCSWD. 1.00000000
as.double.df_marsh.SRPRSWFB. -0.08581679
as.double.df_marsh.THGFSFC. 0.04523356
as.double.df_marsh.MEHGSWB. -0.08426556
as.double.df_marsh.TURB. -0.20284406
as.double.df_marsh.DOCSWD. -0.08581679
as.double.df_marsh.SRPRSWFB. 1.00000000

```

2d) From the soil sample it shows that in total mercury it is related to carbon at 0.6677804 which is pretty high meaning both are in the soil. I personally think because of plants taking in CO<sub>2</sub> that would mean the soil would contain higher levels of this. Thus, meaning for carbon it is the exact same value as mercury. For Phosphorus there is not as high of correlation to the other chemicals, but it is strongest is with carbon. This would mean carbon is the main chemical then everything else is associated with it. Why correlation values are higher is because this is straight soil and not considered run off into the water. Maybe the water dilutes the concentration of chemicals since the water has O<sub>2</sub> in it. Since it is a bigger body of O<sub>2</sub> could mean chemicals are taken away. From looking at these correlation scores mercury his closest to dissolved organic carbon surface water at 0.53653344. The next situ surface water turbidity is with dissolved organic carbon in surface water at 0.2626016. That would make sense why visibility in turbid water the higher these values the less visible the surface water is. The dissolved organic carbon is again associated back with methyl mercury at the same correlation value. The next is soluble reactive phosphorus in surface water at 0.01784706 this means this goes to the dissolved organic carbon again like the rest. Finally, the total mercury is with methyl mercury at 0.04523356. This means that mercury is associated with mercury why it is not with dissolved organic carbon. What I can tell is that when chemicals go into the water the lower the correlation. This could be again because when run off soil goes into the water it holds less chemicals because the soil disperses, and levels would go down.