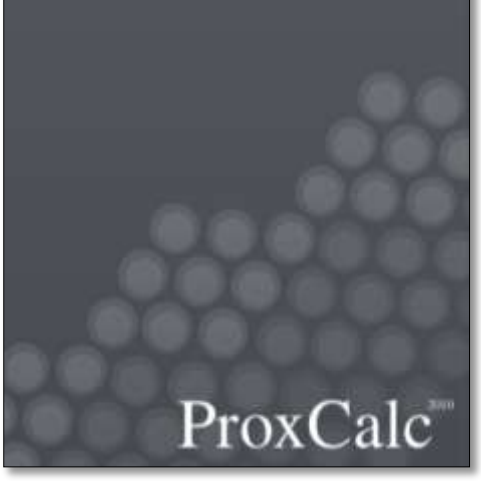


ProxCalc 2010: Proximity coefficient calculator software 1.1

2007, updated in 2010

Lancaster University

8/23/2010

Index Page

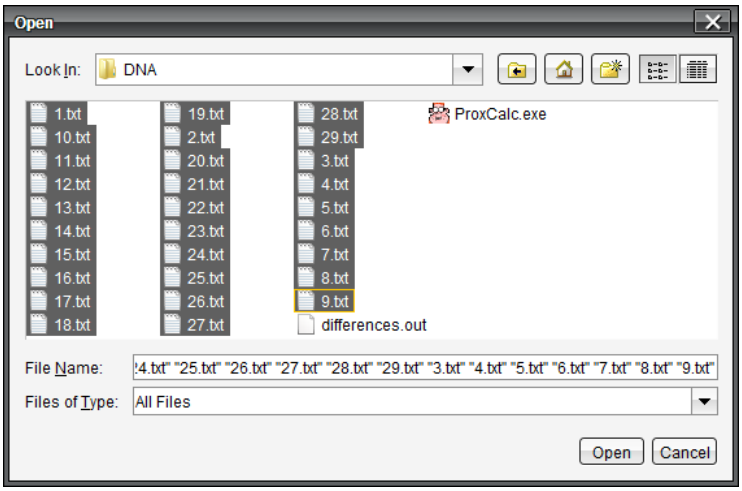
Table of Contents

Managing Files

ProxCalc begins with a preliminary analysis of files to allow the user to inspect the distribution of the occurrence of codes across the cases. This allows for basic but critical file management (e.g., ensuring the files are as expected). To load a set of files for analysis, do the following:

1) Use the 'Add' button to open the file selection dialogue.

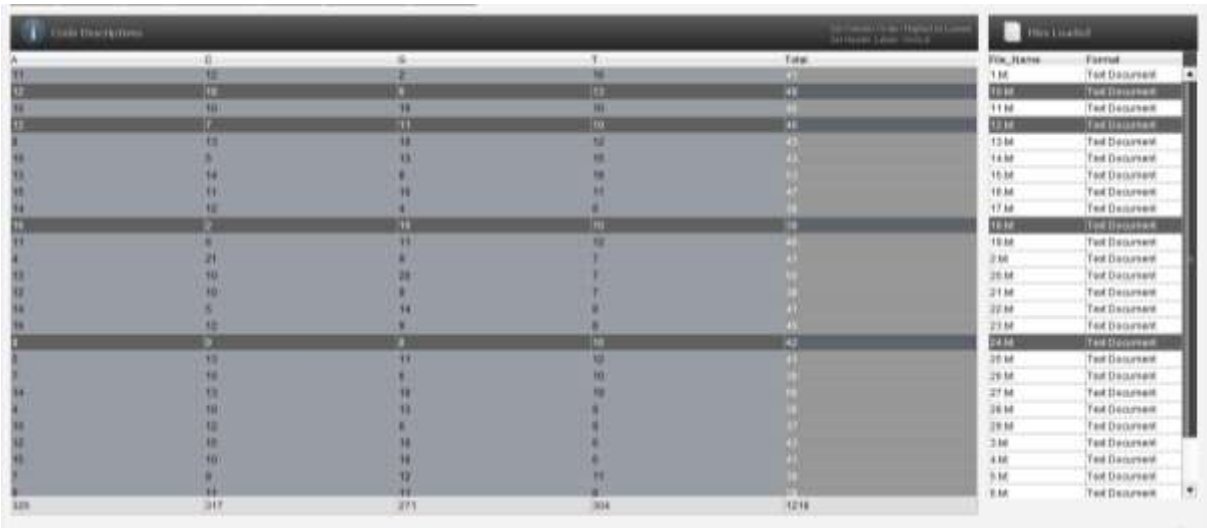
2) Select a file or multiple files from the screen. Click 'Open' when done.



3) The files you selected will now be processed by ProxCalc. ProxCalc will collect and count the codes found in the sequences contained in the files. These results will be displayed in the 'Code Descriptives' table. In this table, the right hand column gives a list of the files loaded along with their file type. Adjacent to this in the main, left-side of the table are counts recorded for each code as it appeared in the adjacent file. The totals for each code in the set and codes in each file will be displayed.

Multiple File Selection

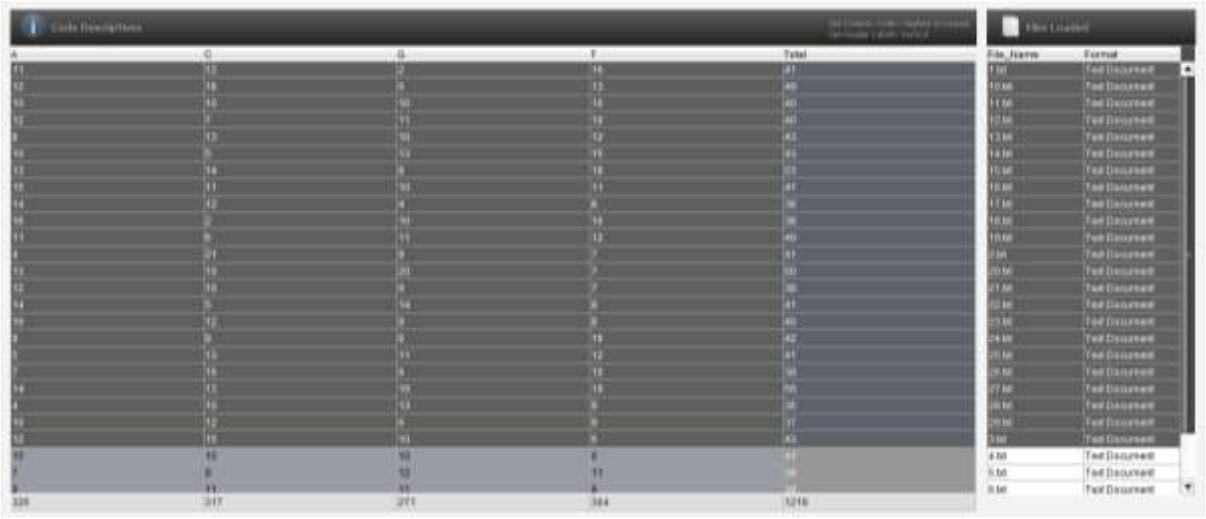
By selecting a row that represents a file in the ‘Files Descriptives table, further options (which are disabled when no rows are selected) become available. You can select multiple files by holding down the ‘SHIFT’ and ‘CTRL’ keys or your machines equivalent.



Removing Files

You may want to remove some files after they have been loaded by mistake or used in some previous analysis. To remove a file from the analysis, do the following:

1) Select the files you want to remove.



2) The ‘Remove’ button will now be available for selection

3) After pressing the 'Remove' button the selected files will be deleted from the list

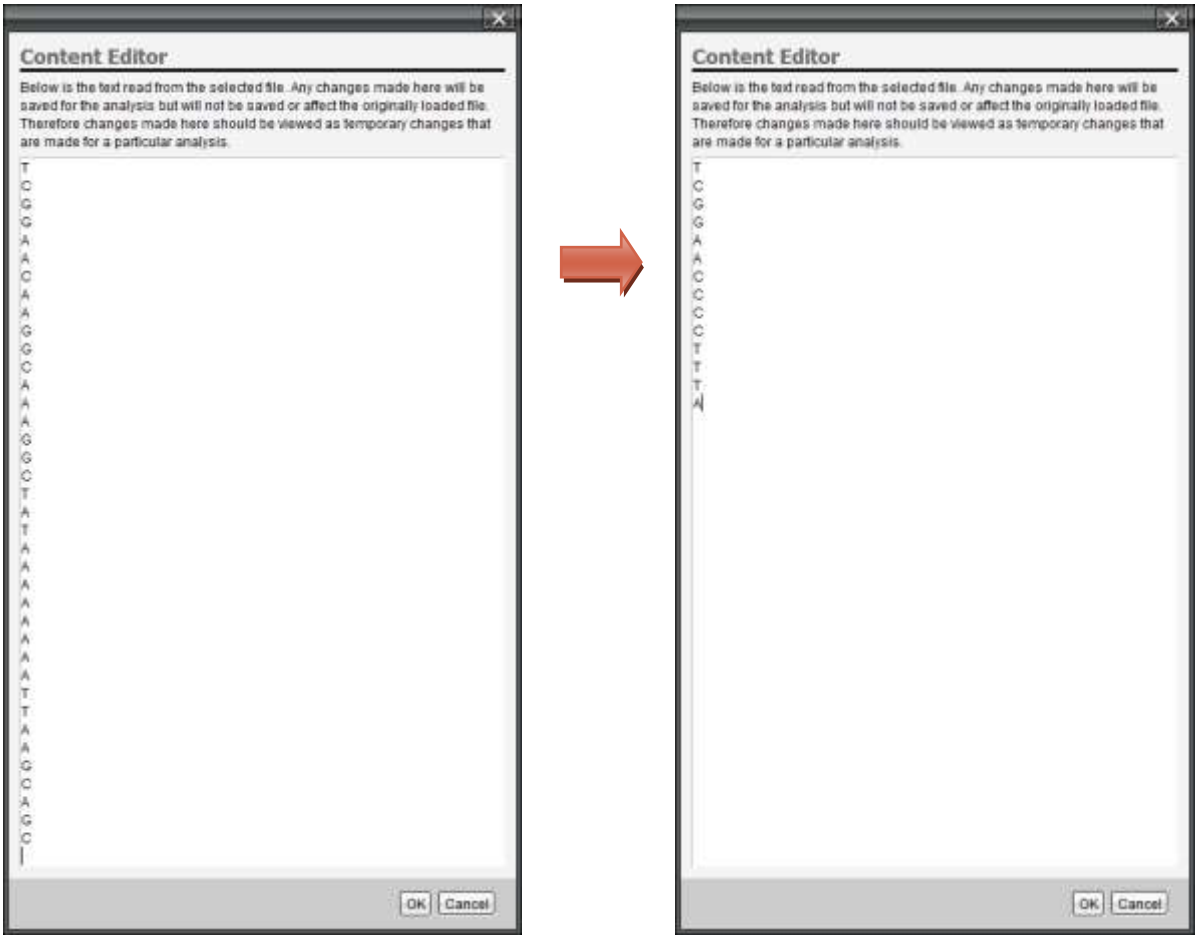
Editing a File Sequence

It is recommended that you edit sequences outside of the ProxCalc software before you load these files for analysis. However, there may be a case where you wish to temporarily remove a particular code or even correct or mark a particular code. To edit a file sequence, do the following:

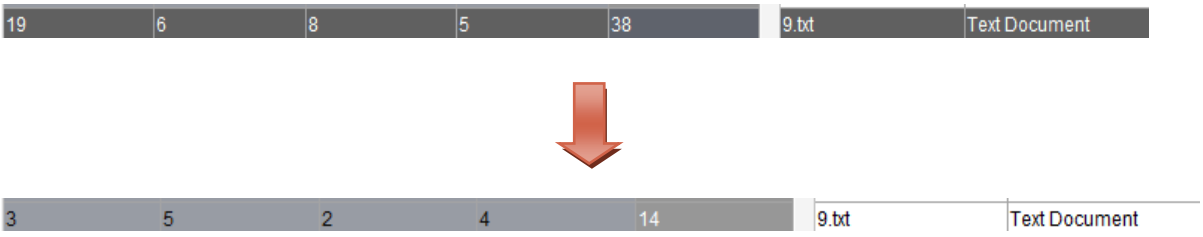
1) Select the file you want to edit.

2) The 'Edit' button will now be available for selection.

3) The 'Content Editor' window will allow you to edit the sequences. Be sure to maintain correct separation of each sequence to allow ProxCalc to segment the text into codes. (Each code must be on a separate line)



3) Click 'OK' to view updated results back in the 'Code Descriptive' Table.



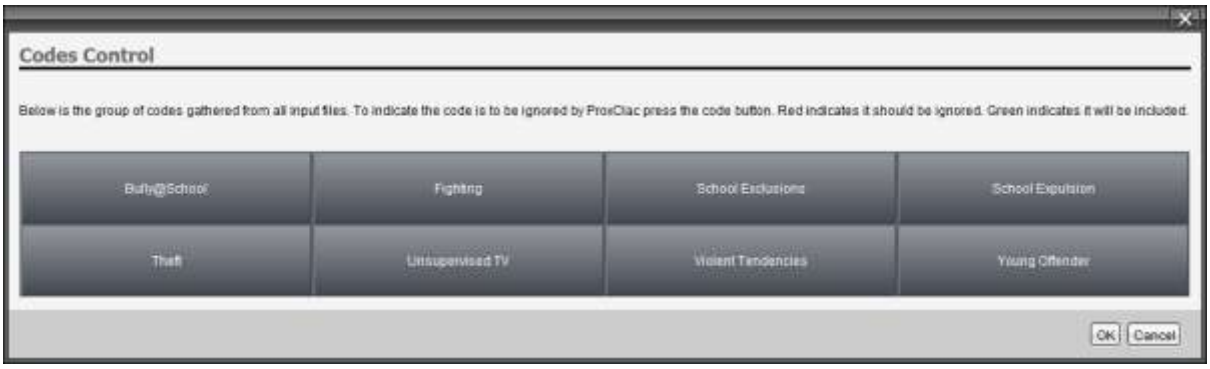
Managing Codes

For different types of analysis, you may find it useful to remove or rather discount certain codes temporarily in order to focus on particular codes. One reason you may wish to do this is because of the lack of occurrence of a code within the cases. To help you identify non-occurring codes, ProxCalc highlights cells within the Code Descriptives table in red when the code has not occurred. Columns of mainly red indicate a code that hardly ever occurred within the cases.

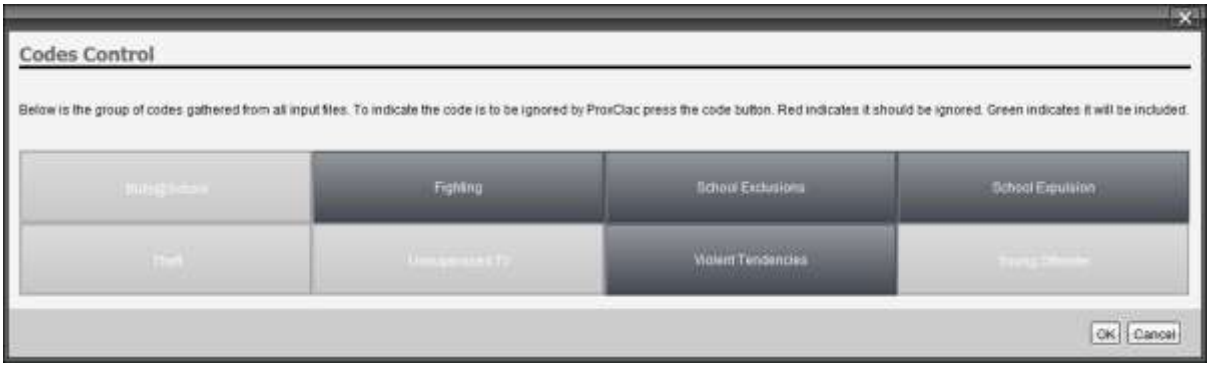
The following figure shows a data set has been loaded that contains many codes, some of which are not of interest to the current research hypothesis. The researcher therefore wishes to remove these codes from the analysis before continuing (Note this is not always the appropriate thing to do. For example, if a researcher were interested in certain aspects of a larger sequence of interaction, one would still retain the larger sequence of interaction but focus on the results for the codes of interest). To manage the codes in the analysis, do the following:

1) Select the 'Codes Control' button.

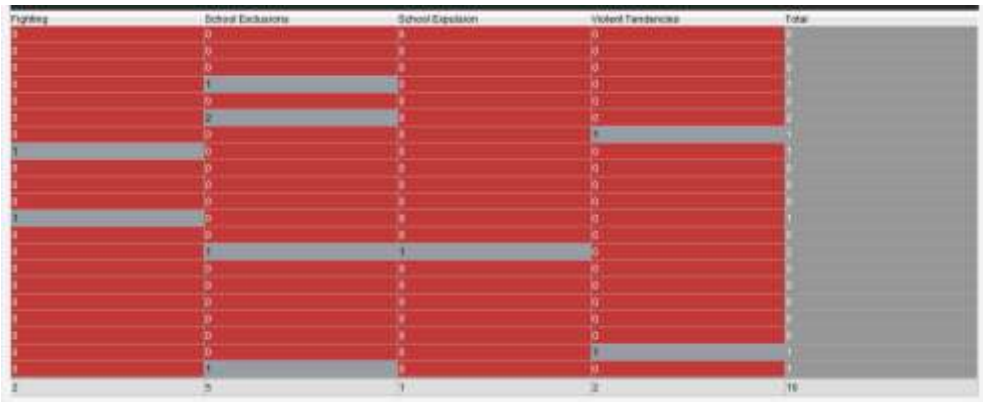
2) The codes available from the current dataset are displayed as a push buttons.



3) Deselect the codes you do not wish to include in the analysis.



4) The 'Code Descriptives' table will be updated according to the codes you deselected removed from the table and all subsequent analysis.

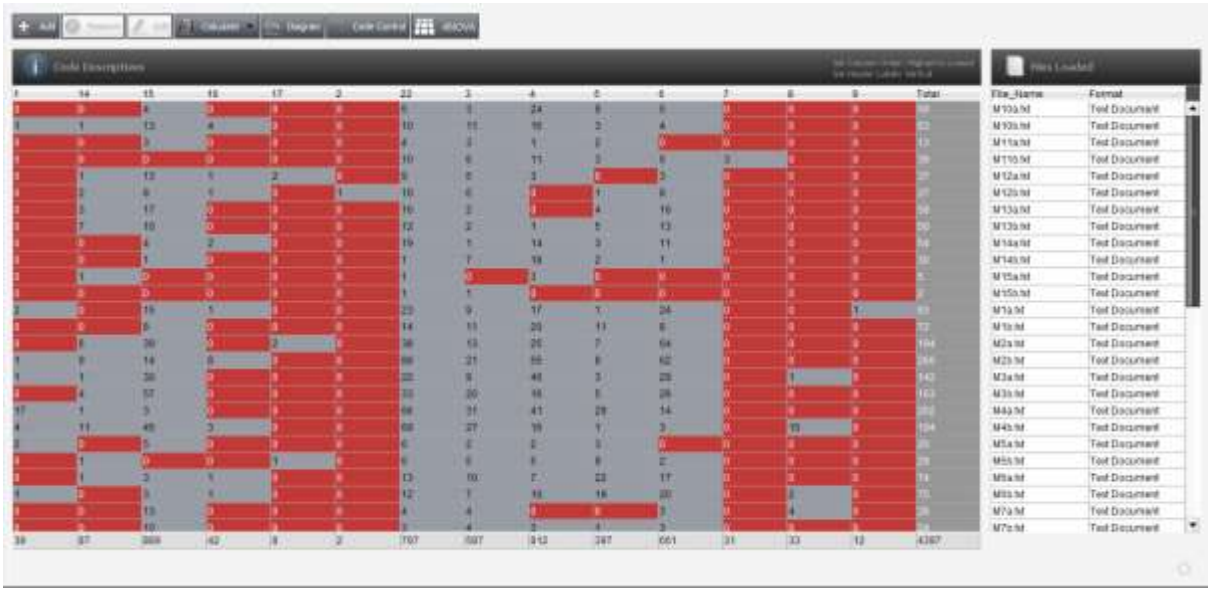


Proximity Coefficient

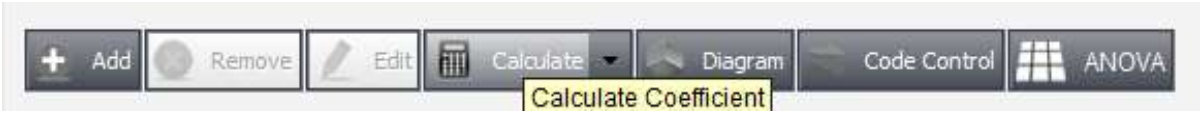
Default

This section outlines basic calculation of the proximity coefficient without making use of the software's advanced analytic features. To calculate the proximity coefficients from a file or range of files, do the following:

1) Load Files (See above). After you have loaded the files, you will see a screen similar to this:



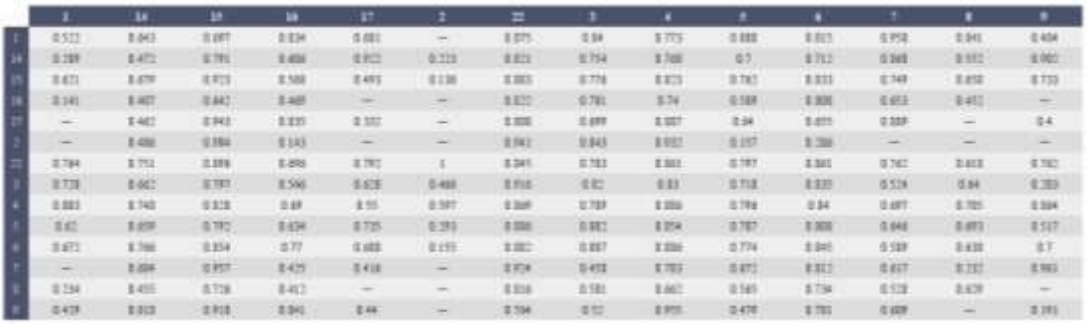
2) The 'Calculate' button will now be available. Press Calculate.



3) Wait for ProxCalc software to finish calculating results. The activity indicator bar will be active when ProxCalc is busy.

4) When the results have been calculated a new results tab will be added to the ProxCalc windowpane.

5) The results are displayed starting with a table that gives the overall averaged proximities found from all contingencies across all files.



NOTE: You should read Proximity scores horizontally then up, with the 'cue' shown in the rows and the 'response' shown in the columns. Therefore, in this example '14->1' has a proximity score of 0.289 and '1->14' has a proximity score of 0.643.

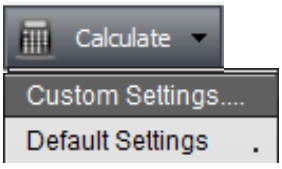
Following the overall table is each files individual proximity scores tables label for each file they represent.

	15	22	3	4	5	6
15	0.972	0.979	—	—	—	0.969
22	0.596	0.797	—	0.951	0.896	0.892
3	0.194	0.965	0.99	0.861	0.979	0.715
4	0.636	0.688	—	0.993	0.846	0.795
5	0.623	0.85	1	0.965	0.927	0.877
6	0.865	0.812	—	1	0.75	0.849

Advanced

The advanced option allows you to calculate the proximity coefficient using a custom weight. A custom weight is appropriate when there are theoretical reasons to restrict the scope of proximity. Examples of where this might be appropriate are described in Taylor (2006). To change the weighting of a coefficient:

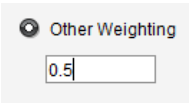
1) Use the dropdown menu and select 'Custom Settings'.



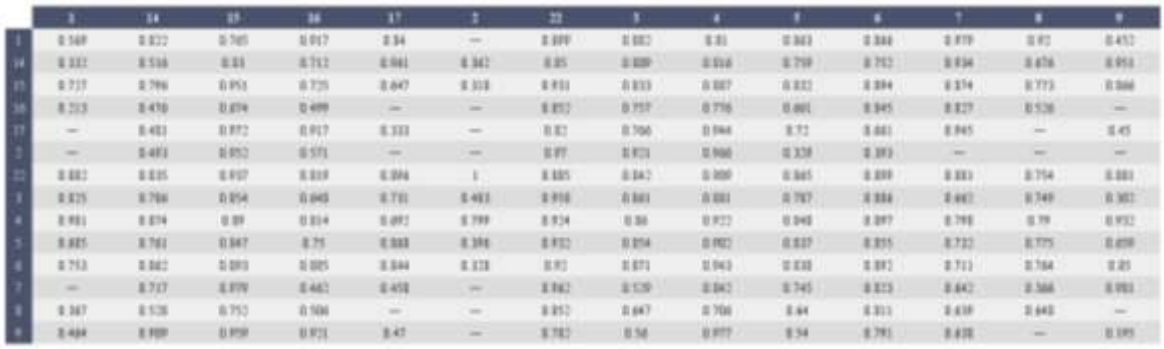
2) Using the radio buttons select the appropriate option. No weighting (technically w = 1.00) produces a coefficient matrix in which a P(A,B) = 1.00 indicates that A always immediately occurred before B, while P(A,B) = 0.00 indicates that A occurs only once at the beginning of the sequence and B occurs only once at the end of the sequence.

Absolute proximity produces a coefficient matrix in which the proximity between two codes is represented according to the absolute number of intervening codes, rather than as a proportion of the sequence length (i.e., maximum proximity).

Selecting 'Other Weighting' allows you to type in your own weight. This must be specified in numeric terms, using only numbers ~~and mathematical functions~~.



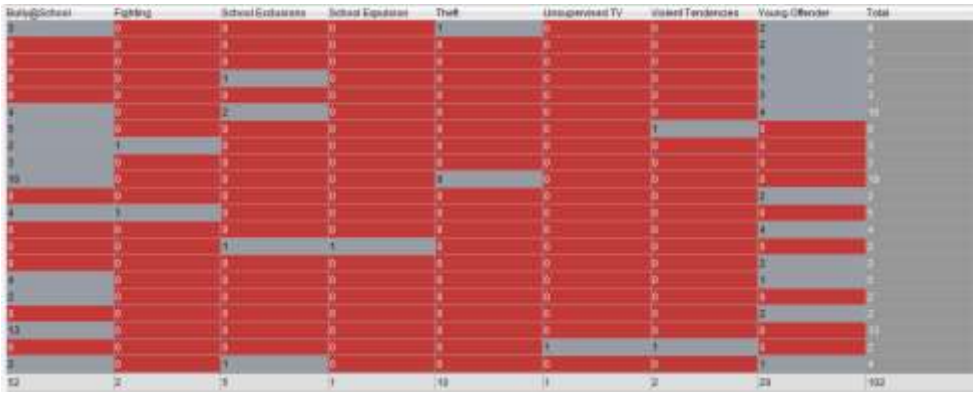
3) After you have selected the option required for the analysis press the 'OK' button to start the Proximity Coefficient calculation. Results are displayed in the same way as previously described in the 'Default' section of the help guide.



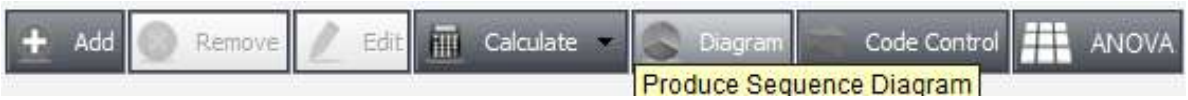
Proximity Diagram

This option allows you to present contingencies and proximities among codes graphically using a state transition diagram. In these diagrams nodes represent the events (variables) within a set of cases. An arrow is drawn between two nodes when the events associated with the nodes are found to occur next to one another in one or more of the cases. The direction of the arrow corresponds with the temporal order in which the events occurred. To derive a state transition diagram:

1) Load up files (See...).



2) Select 'Diagram' button from the menu.



4) When the results have been calculated a new results tab will be added to the ProxCalc windowpane.

3) The output will be added to a new results tab on the ProxCalc windowpane.

4) The sequence diagram output ready for customisation. This includes manual manipulation of the positioning of the nodes.

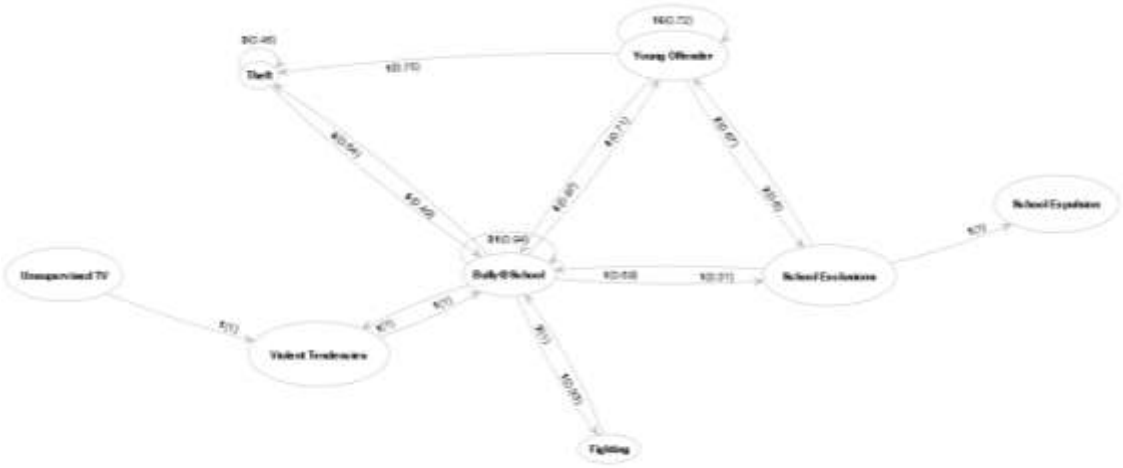
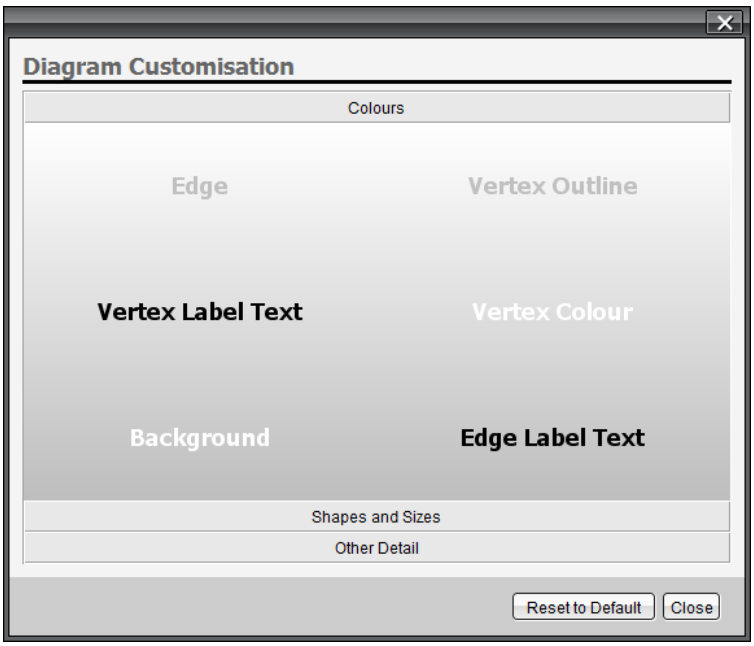


Diagram Customisation

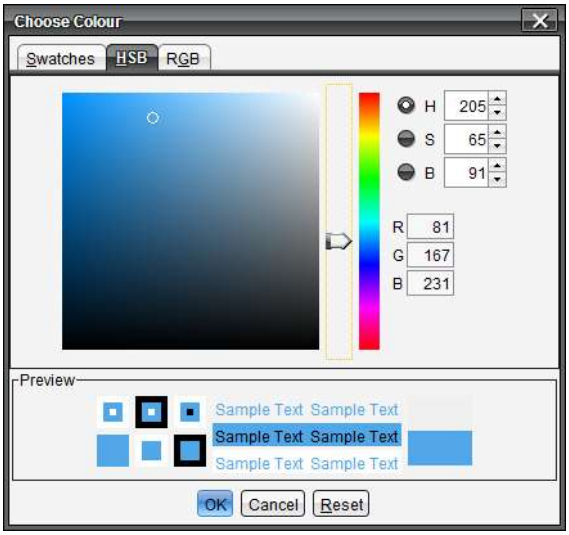
Although the diagram layout is controlled by a algorithm that attempts to work out the clearest layout for the number of vertices and edges, it may not be the best layout. You can drag vertices around on screen and zoom in and out either using the controls on the side of the screen or the mouse roller.

1) Click the settings button.

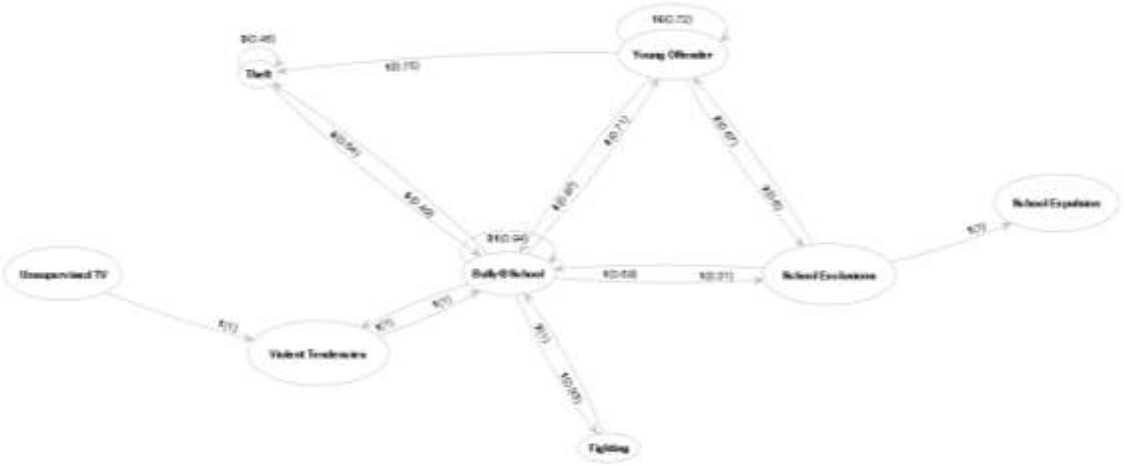
2) With the 'Diagram Customisation' window, you can change colours, sizes of the diagram components and arrow shapes etc.

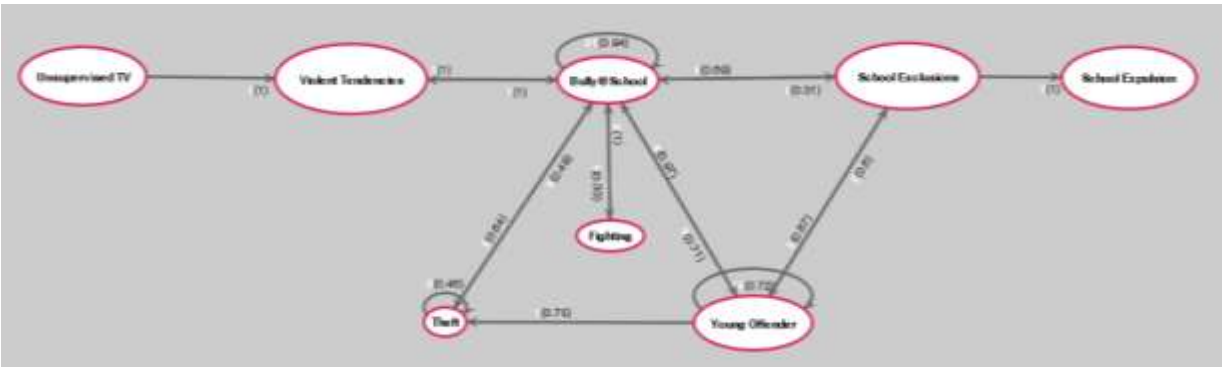


3) Click on a link label. A colour selection window enabled you to pick or make your own colours.



4) Using the settings it is possible to change the look and layout of the diagram.



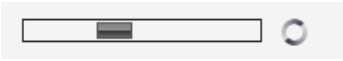


Group Difference Testing (Analysis of Variance)

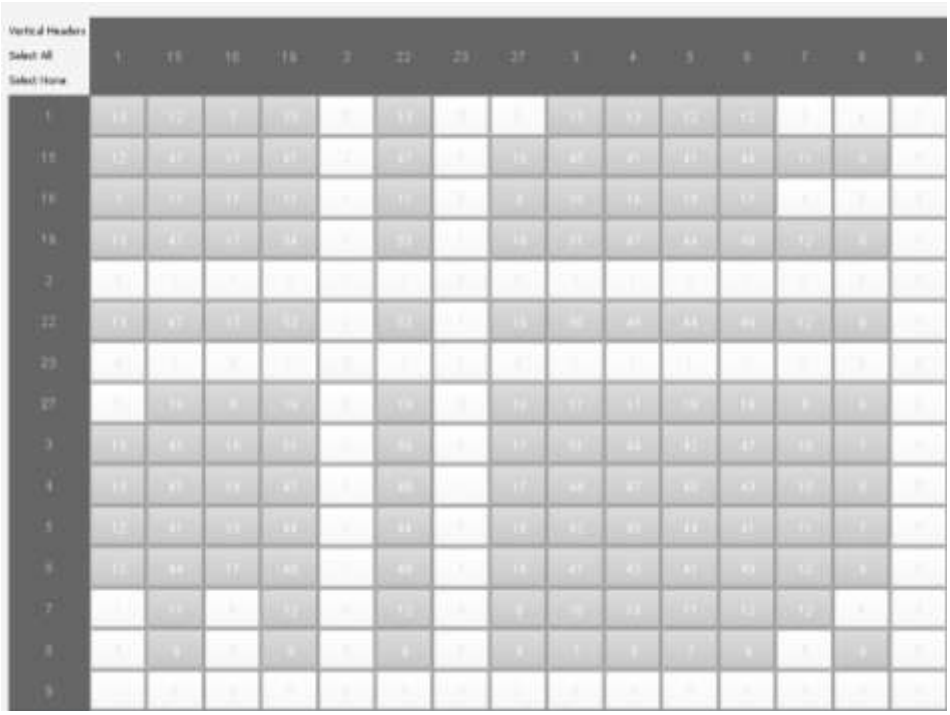
Often researchers want to go beyond providing descriptive accounts of their sequence structures to make inferences about coefficients found in groups of cases. One approach to making such comparisons is to use a randomisation test in which we assess the likelihood that the difference observed between groups is unlikely to have occurred by chance. ProxCalc implements a randomisation test using the Analysis of Variance statistic. This test indicates the probability with which an *F* value higher than that observed in the data occurred across a set number of randomisations. For further information on this approach, please see Taylor (2006, or Giebels & Taylor, 2009) and the references therein.

In order to run an ANOVA test, complete the following steps:

1) After you have pressed the ANOVA button the proximities are calculated for each contingency in each file so the activity indicator bar will be active for a few moments.

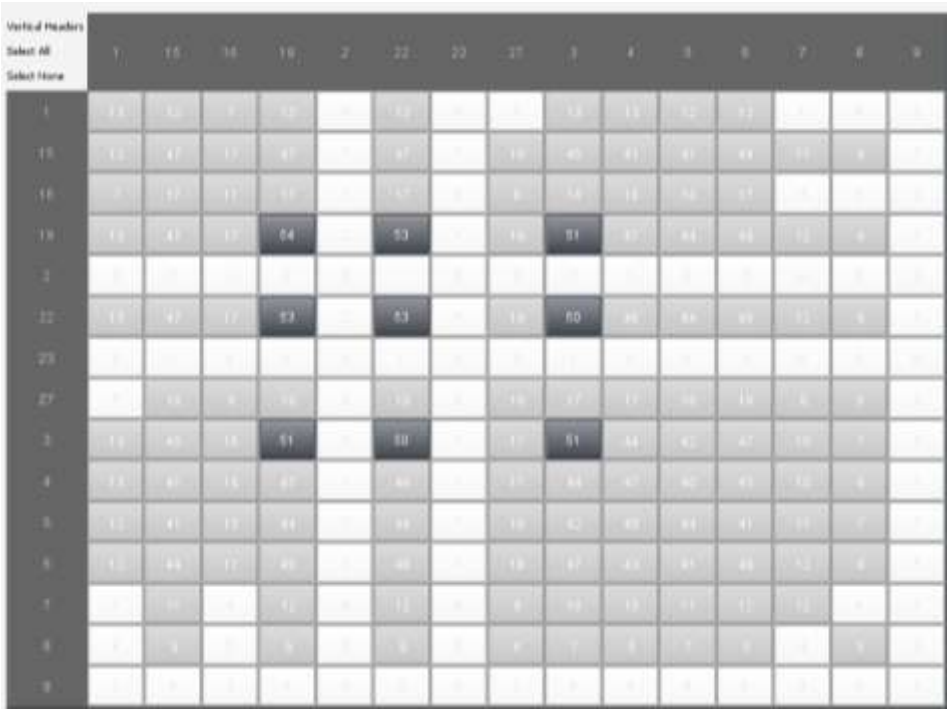


2) After the proximities are calculated, you will see the contingency selection screen similar to this:



Here you can select which contingencies you want to run the ANOVA tests with. Some options are disabled because the number of files that contain that contingency are too few. It is recommended that you focus your analysis on only those contingencies that test your hypotheses (i.e., don't go fishing). Hovering over a contingency push button will highlight what the contingency is and its related codes.

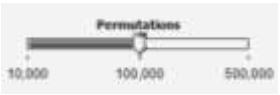
3) Select the contingencies you wish to use and press the ‘Next ‘ button.



2) On the ‘Analysis Design’ screen. First, decide and select the number of Independent Variables you wish to examine. ProxCalc is currently designed to run up to a three-way ANOVA design. It assumes a between-subjects design (note p is calculated through randomisation so this should allow meaningful tests of reapeated measures) and utilises the formulas presented in Howell (1996).

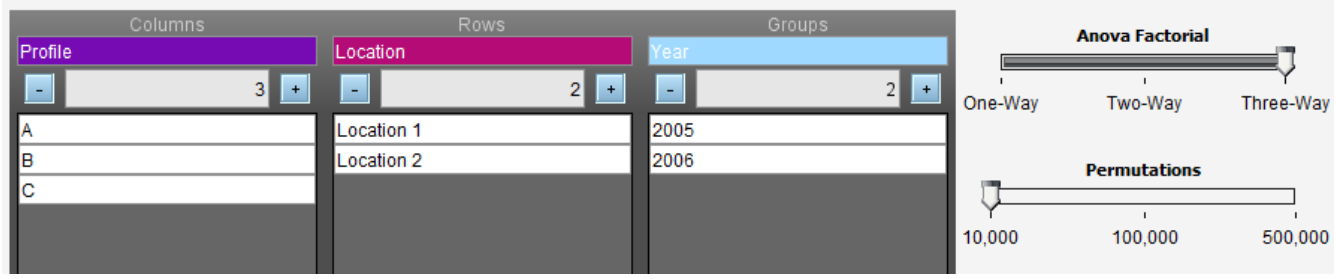
The dialogue boxes not only allow you to select the number of Independent Variables but it also allows you to lable those variables and the levels within those variables.

3) Select the number of 'Permutations'. This is the number of times the ANOVA is repeated with the values from each cell shuffled and redistributed. A greater number of permutations will provide more accuracy in the estimate of *p*. (We are currently testing the accuracy of different numbers of permutations and hope to present the results of this in Summer 2011).



Example Design

To illustrate the ANOVA test in actions we will use the following example.



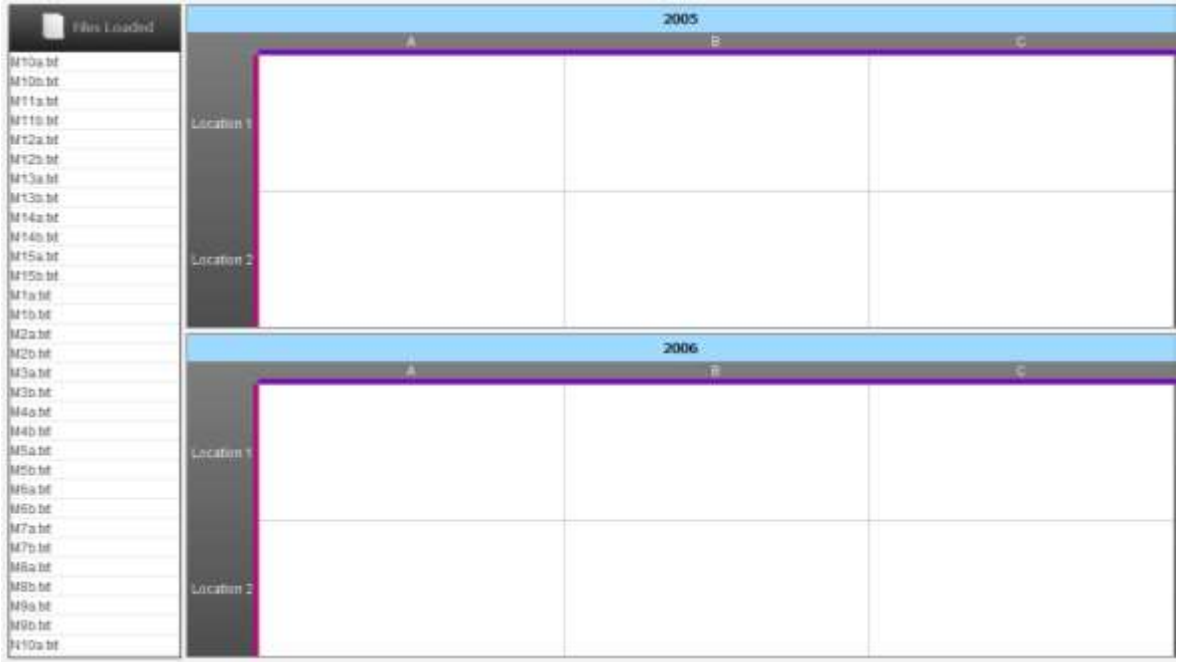
In this design we have selected Three-Way ANOVA design. Renamed the independent variables 'Profile', 'Location' and 'Year'. We have three profiles 'A', 'B' and 'C'. We have two locations 'Location 1' and 'Location 2'. And we have two different years '2005' and '2006'. We are keeping the number of Permutations at 10,000.

File Drop Interface

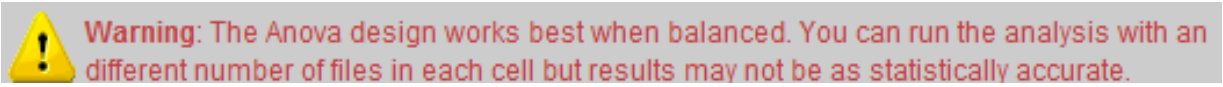
After designing our ANOVA test we must now populate the cells of the design. Based on the example design the 'File Drop Interface'.

The Key explains the variables and how they are presented in the table.

As you can see the design is grouped into two years with a grid of locations and profiles. Each white space represents a cell in which files can be dragged from the left hand side and dropped into the cell.



As we begin to drop files into the cells a 'Warning' message will be displayed at the bottom of the screen. This is a just a reminder that the ANOVA design should be, where possible, 'Balanced'. However, ProxCalc does use an unweighted-means solution to calculate the ANOVA with unbalanced design.



After we have finished dragging the files to the correct cells, we will end up with a number of files in each cell.

Clicking the 'Start' will begin the ANOVA test. (See What happens in ANOVA for details)

This is an example of the results we can expect when there was enough data for the ANOVA to run. The output mirrors a standard ANOVA output from popular statistics packages such as SPSS. The one shown below is for the contingency Variable 19 → Variable 19. The single difference between this and the output of SPSS is that the *P* value reflects the results of the randomisation test, rather than the original value derived from the *F* distribution. (You can of course compare the P given by ProxCalc to that obtained by the traditional approach by looking up the relevant *F* and df in a significance table).

19 --> 19 Analysis of Variance

Source	df	SS	MS	F	P
Location	1	0.00431	0.00431	1.3318	0.2908
Profile	2	0.00278	0.00139	0.42944	0.7859
Year	1	0.00174	0.00174	0.53736	0.5687
Location * Profile	2	0.01114	0.00557	1.7192	0.1654
Location * Year	1	0.00055	0.00055	0.16969	0.7882
Profile * Year	2	0.00332	0.00166	0.51259	0.6995
Location * Profile * Year	2	0.00286	0.00143	0.44184	0.7532
Error	36	0.11663	0.00324		
Total	47	0.14334			

The 'Cell subject Information' gives details about how each cell was populated when the ANOVA test ran. This is important in deciding on the robustness of the results.

Cell Subject Information

Anova cell: [Location 1 A 2005] had a subject count of 4
Anova cell: [Location 1 B 2005] had a subject count of 4
Anova cell: [Location 1 C 2005] had a subject count of 4
Anova cell: [Location 2 A 2005] had a subject count of 4
Anova cell: [Location 2 B 2005] had a subject count of 4
Anova cell: [Location 2 C 2005] had a subject count of 4
Anova cell: [Location 1 A 2006] had a subject count of 4
Anova cell: [Location 1 B 2006] had a subject count of 4
Anova cell: [Location 1 C 2006] had a subject count of 4
Anova cell: [Location 2 A 2006] had a subject count of 4
Anova cell: [Location 2 B 2006] had a subject count of 4
Anova cell: [Location 2 C 2006] had a subject count of 4

In this example you can see the cells had 4 subjects in each. This is a balanced design. However in practise you can expect to see counts that vary. When a cell has a subject count of 0 then the ANOVA is unable to run, since there is not enough information to conduct the test.

Controls in Result Tab Pages

In the results tab page for proximity, proximity diagrams, and Group difference testing, you will see an array of button at the top left-hand side. These button control options for that particular results page.

Saves the output as a file. If this a diagram then it is saved as png image. If this text then saves it was a Rtf file.

Copies into clipboard. If this is a text output then this will be Html formatted ready to paste into another application. If this an image output then it takes a screenshot of the image view and puts that into your clipboard.

Zooms into the diagram.

Zooms out of the diagram.

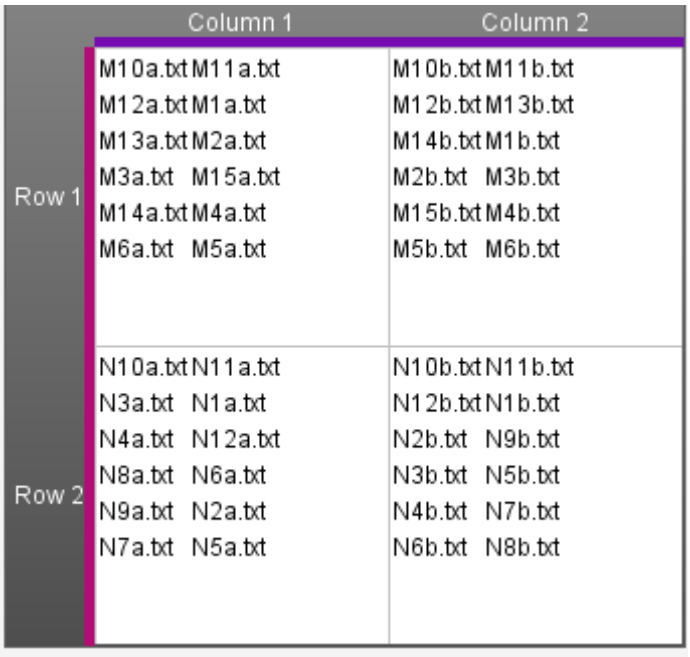
Shows settings options for diagrams.

Deletes results page.

What Happens in ANOVA

Using a simple 2x2 example this is a brief breakdown of the process that happens when the ANOVA is executed.

The files have been distributed to their corresponding cells.



	Column 1	Column 2
Row 1	M10a.txt M11a.txt M12a.txt M1a.txt M13a.txt M2a.txt M3a.txt M15a.txt M14a.txt M4a.txt M6a.txt M5a.txt	M10b.txt M11b.txt M12b.txt M13b.txt M14b.txt M1b.txt M2b.txt M3b.txt M15b.txt M4b.txt M5b.txt M6b.txt
Row 2	N10a.txt N11a.txt N3a.txt N1a.txt N4a.txt N12a.txt N8a.txt N6a.txt N9a.txt N2a.txt N7a.txt N5a.txt	N10b.txt N11b.txt N12b.txt N1b.txt N2b.txt N9b.txt N3b.txt N5b.txt N4b.txt N7b.txt N6b.txt N8b.txt

The ANOVA will cycle through each proximity coefficient for the contingencies selected. So based on the above design when the algorithm cycles through, it will collect all the proximities. So for the contingency '19-->15' (19 and 15 represent codes) it will collect the proximity coefficient from the files corresponding to each cell. In this case:

	Column 1	Column 2
Row 1	0.6761086603204087 0.8476658476658476 0.9617834394904459 0.9139837398373983 0.8833927510398099 0.9317508673944317 0.9322250639386189 0.7775910364145658 0.7793853591160221 0.9479305740987984 0.8847826086956522	0.9446280991735537 0.9044303797468355 0.7852180339985217 0.5352056962025317 0.9796747967479675 0.9145650414098013 0.9852449340940389 0.9586921296296296 0.9488372093023256
Row 2	0.9089053463481014 0.9897959183673469 0.9147321428571429 0.9547619047619048 0.6388594476829772 0.8463087248322148	0.9122983870967742 0.9669016842929886 0.9260398860398861 0.985239852398524 0.8302350591488707 0.9311160857051487

	0.8312606292517006 0.9656035846072747 0.897087868218683 0.9531233595800525	0.9719946013689386 0.7784842756487369 0.8400203873598369 0.7490720118782479 0.8567689399447433

As you can see this design is not too unbalanced this is because the contingency '19-->15' has a score in majority of the files put into the cells as subjects.

The following ANOVA results table is given for this contingency. Notice the SS Total is not there this is due to the type of method used to calculate the ANOVA when the design is unbalanced.

19 --> 15 Analysis of Variance

Source	df	SS	MS	F	P
Variable 2	1	0.00162	0.00162	0.14876	0.7078
Variable 1	1	0.00045	0.00045	0.04102	0.8463
Variable 2 * Variable 1	1	0.00112	0.00112	0.10251	0.7541
Error	37	0.4027	0.01088		
Total	40				

Next we will look at a contingency that was not as well represented in the files of the data set. The contingency '220 --> 14' did not have as many scores present in the particular files from that data set. This can be seen by the unbalanced representation in the ANOVA design.

	Column 1	Column 2
Row 1	0.7870791628753412 0.9240924092409241 0.9002557544757033 0.875 0.3598130841121495	0.45999999999999996 0.6786216596343179 0.8407553107789142 0.8970611346848971 0.8836990173643828 0.9049342105263158
Row 2	0.8274932614555256 0.7225950782997763	0.8122184684684685 0.93275932759327

This is just something to keep in mind when looking at results. Here are the ANOVA results for this contingency:

220 --> 14 Analysis of Variance

Source	df	SS	MS	F	P
Variable 2	1	0.00743	0.00743	0.20988	0.7323
Variable 1	1	0.00818	0.00818	0.23094	0.6392
Variable 2 * Variable 1	1	0.00582	0.00582	0.16437	0.6879
Error	11	0.38944	0.0354		
Total	14				