

CONTENTS

サンプルの Experiment – Heat Map and Dendrogram Demo	2
クイックノート	2
フェイクのデータを作成.....	2
簡単なヒートマップ	3
デンドログラム付きのヒートマップ.....	5
行だけのデンドログラム付きヒートマップを作成する	5
デンドログラム表示の拡張.....	9
その他のいくつかの分析.....	12
2方向（行と列）デンドログラム付きヒートマップを作成する.....	13

サンプルの Experiment – Heat Map and Dendrogram Demo

クイックノート

メニュー File → Example Experiments → Analysis → Heat Map and Dendrogram Demo

この Experiment は、ヒートマップを作る方法、ヒートマップとして表示する可能性のあるデータに階層的クラスタリングを適用する方法、HeatMap_and_Dendrogram プロシージャファイルの使い方を説明します。

ヒートマップとは、2次元のマトリックスウェーブの内容を表示する画像プロットにすぎません。プロットは、偽色画像であり、つまり、画像のピクセルの色は、行列のセルの値に基づいてカラーテーブルから参照されます。

一般的な方法は、YellowHot カラーテーブルのような色テーブルを使うことです。

このテーブルでは、低い値は黒、中間の値は赤、高い値は黄色で色付けされます。

この方法は、黒体輻射の色域に似ているため、「ヒートマップ」という名前が付けられています。

フェイクのデータを作成

ステップをわかりやすくするために、新しい Experiment を作成したところからの手順で確認します。

(クイックノートで実行する場合は、コマンドの行を選択して、Ctrl+Enter で即座に実行することができます。)

1. コマンドラインで次を実行します。

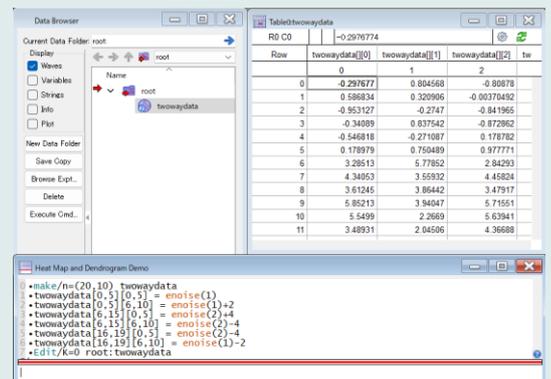
```
make/n=(20,10) twowaydata // 20 行、10 列
twowaydata[0,5][0,5] = enoise(1) // 左上
twowaydata[0,5][6,10] = enoise(1)+2 // 右上
twowaydata[6,15][0,5] = enoise(2)+4 // 左中央
twowaydata[6,15][6,10] = enoise(2)-4 // 右中央
twowaydata[16,19][0,5] = enoise(2)-4 // 左下
twowaydata[16,19][6,10] = enoise(1)-2 // 右下
```

これは、ランダムなデータが類似した値を持つ6つのエリアを持つマトリックスを作成します。

例えば、この値は多数の実験対象に多数の測定を行った結果で、各行が対象を表しているようなイメージです。

実験対象は、通常、類似したグループにあらかじめ分類されることはほとんどありません。

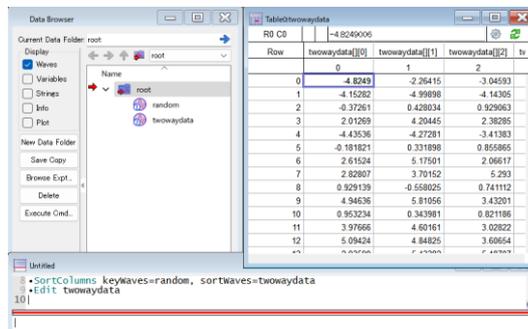
実際のデータにより近いものにするため、行を混ぜ合わせます。



(ここではテーブルを実際に表示する必要はありません)

2. コマンドラインで次を実行します。

```
Make/O/N=20 random = enoise(1)
SortColumns keyWaves=random, sortWaves=twowaydata
Edit twowaydata
```

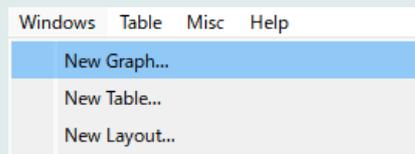


簡単なヒートマップ

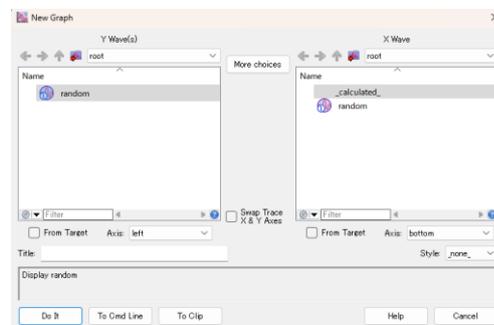
最初のウェーブの代入に記載されているコメントは、表で表示される位置を示しています。

ヒートマップを作成するには、次の手順に従ってください。

3. メニュー Windows → New Graph を選択します。

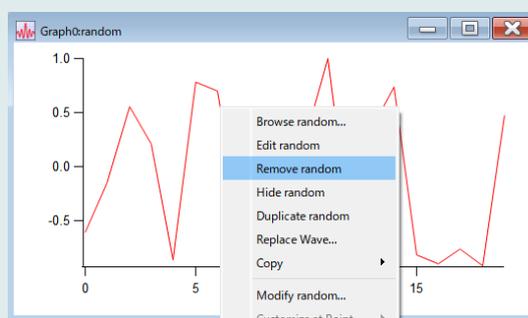


4. Do It をクリックします。

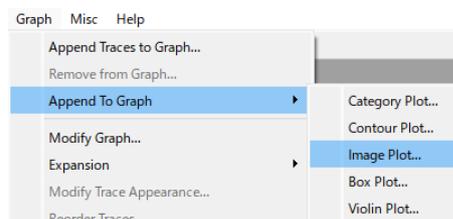


5. ブランクのグラフウィンドウにしたいため、トレース (random) を右クリックして、コンテキストメニューから Remove random を選択します。

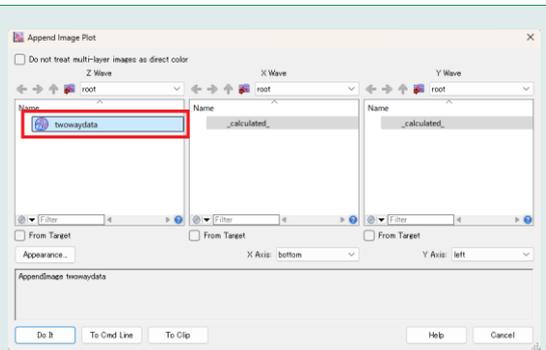
これでグラフウィンドウが、ブランクになったはずですが。



6. メニュー Graph → Append to Graph → Image Plot を選択します。

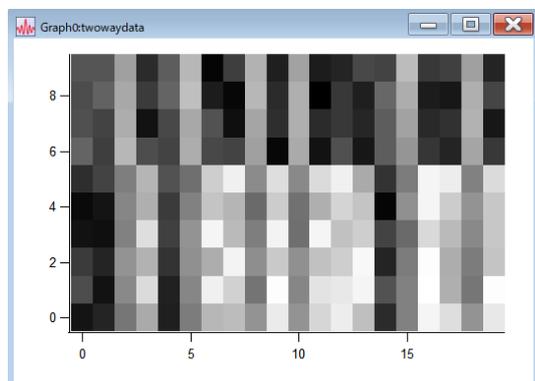


7. Z Wave リストから twowaydata を選択して、Do It をクリックします。



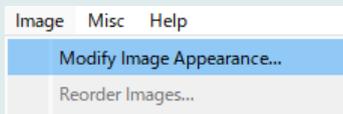
テーブルの列が水平軸に沿って配置されていることに注目してください。

画像プロットは、表の表示に対応するように変換されています。

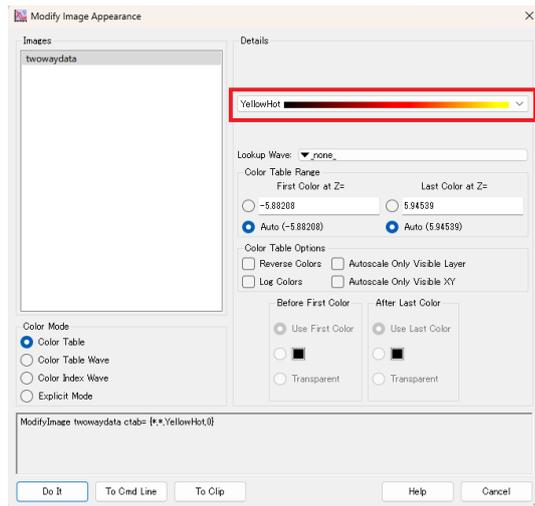


8. ここで画像の外観を変更します。

メニュー Image → Modify Image Appearance を選択します。



9. YellowHot カラーテーブルを選択します。

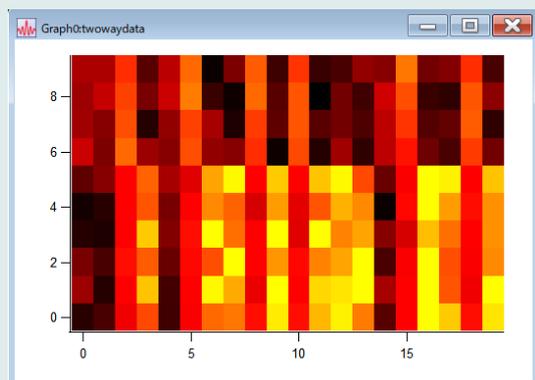


10. Do It をクリックします。

これが「ヒートマップ」と呼ばれる理由がわかると思います。このようにすると、同じような値の領域が簡単に確認できます。

このような行列は、多くの場合、データのベクトル集合を表していて、各ベクトルは1つの実験対象またはサンプルに対して行われた測定の集合を表しています。

行列の各行をベクトルとして解釈し、各列は与えられた実験対象に関する同じ種類の測定値を表すものとなります。



注記： Igor の画像プロットでは、行列の行は垂直方向に並び、画像の左から右に向かって行番号が増加します。

このデータの集合を何らかの形で整理し、グループに分けることができるのではないかと考えます。

これを行う方法の1つとして、階層型クラスタリングが考えられます。

HeatMap_and_Dendrogram プロシーダは、HCluster コマンドのための便利な GUI を提供します。

このコマンドでは、行ベクトルから距離行列を計算し、デンドログラムを描画するために必要な基本的なクラスタリング情報を計算します。

階層型クラスタリングは、クラスタを特定するものではないことに注意してください。

これは、可能性のあるクラスタリングを解釈することを助ける方法でデータベクトルをグループ化するものです。

他の方法と比較して、クラスタ数を事前に仮定する必要がないという利点があります。

デンドログラム付きのヒートマップ

HeatMap_and_Dendrogram プロシーダは、ベクトルデータから非類似度行列を計算し、階層的クラスタリングを計算し、それをヒートマップとデンドログラムの作成に使う GUI を提供します。

行のみ、または行と列の両方について、デンドログラムによる表示を作成できます。

まずは行だけから始めます。

注記： データベクトルとして、行ではなく列を使いたい場合もあります。

HCluster コマンドとこのプロシーダファイルは、行のベクトルに対してのみ動作します。

データセットの列に対して処理を行うには、MatrixTranspose コマンドを使います。

まず、プロシーダパッケージをロードします。

メニュー Analysis → Packages → Hierarchical Clustering and Dendrogram を選択します。

これにより、プロシーダパッケージがロードされ、元のデータの選択や非類似度行列とデンドログラムのクラスタリング情報の計算に使う各種オプションを選択するためのコントロールパネルが表示されます。

このコントロールパネルでは、必要に応じて独自の非類似度行列を使って作業を行うことができます。

または、プロシーダウィンドウのコードの上部付近に次の行を入力することもできます：

```
#include <HeatMap_and_Dendrogram>
```

この場合、コントロールパネルは表示されません。

プロシーダパッケージがロードされると、メニュー Analysis → Packages → Hierarchical Clustering →

Display Heat Map with Dendrogram を選択すれば、いつでもコントロールパネルを表示できるようになります。

行だけのデンドログラム付きヒートマップを作成する

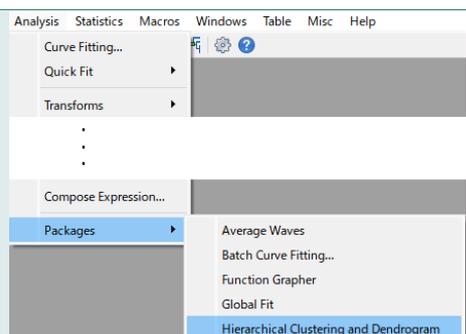
このセクションでは、行がデータのベクトルを表す行列の非類似度データを計算し、その非類似度行列に HCluster コマンドを適用する手順を説明します。

次に、HCluster の出力からデンドログラムを描画できるように、元のデータを再編成したグラフを作成します。

最終的に、HCluster 出力からデンドログラムが描画されます。

1. メニュー Analysis → Packages → Hierarchical Clustering → Display Heat Map with Dendrogram を選択します。

パッケージが読み込まれ、Analysis メニューに項目が追加されます。

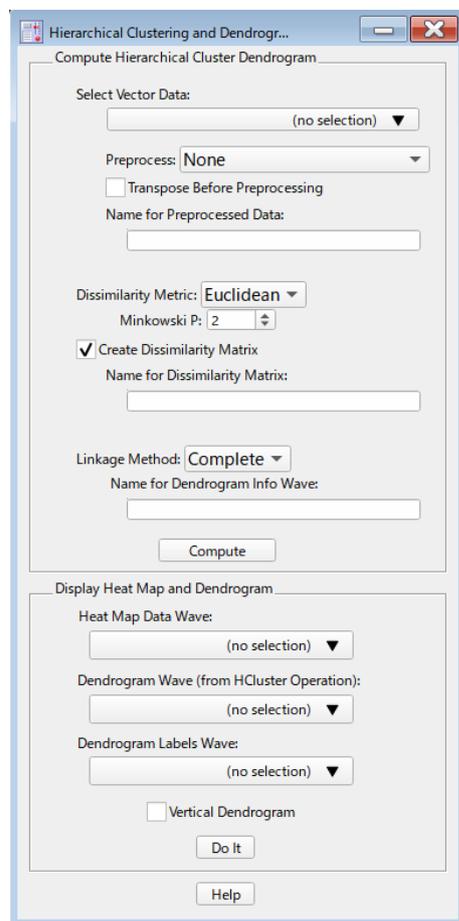


2. メニュー Analysis → Hierarchical Clustering → Display Heat Map with Dendrogram を選択します。

コントロールパネルに多くのオプションが表示されます。

すでに表示されている場合は、そのメニュー項目を選択するとコントロールパネルウィンドウが前面に表示されます。

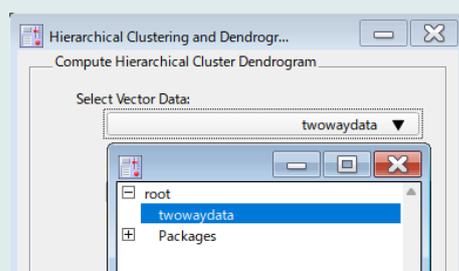
画面に収まりきらない場合は、パネルを右クリックして、コンテキストメニューから Expansion で縮小してください。



3. 画面上部の Select Vector Data のポップアップメニューから twowaydata を選択します。

このウェーブは 20 行 10 列です。

これは双方向のデンドログラムを示すために作られたものですが、単純な一方向のデンドログラムの表示という今回のニーズにも十分対応できます。

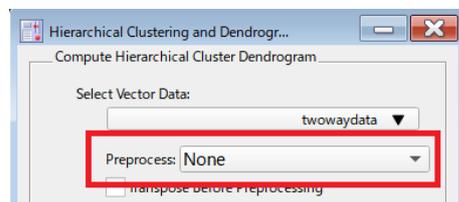


4. Preprocess で None が選択されていることを確認してください。

twowaydata ウェーブは、ある列のデータが他の列のデータと類似するように構築されているため、クラスタリング計算の前にデータを整える必要はありません。

データ列にさまざまな種類の測定値が含まれ、異なる範囲と大きさを持つ場合は、列に何らかの正規化を適用する必要があるかもしれません。

None を選択しているため、前処理データの名前を入力する必要はありません。



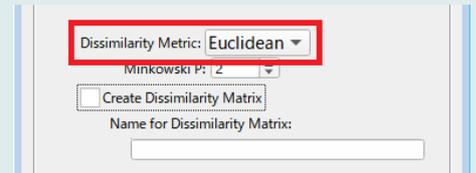
5. Dissimilarity Metric で Euclidean が選択されていることを確認してください。

これは、ユークリッド距離を非類似性の尺度として選択します。つまり、各行の値の二乗和の平方根を単純に計算します。

これはもっとも一般的な方法で、他の方法が必要になることはあまりないと思います。

さまざまなオプションに関する説明は、Hierarchical Clustering (Analysis.ihf ファイル内) を参照してください。

Euclidean を選択するということは Minkowski P (ミンコフスキーP) 設定は必要ないということです。



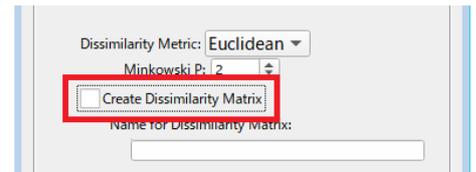
6. Create Dissimilarity Matrix のチェックが外れていることを確認してください。

これも、名前を付ける必要がないことを意味します。

明示的な非類似性ウェーブを作らないことで、HCluster コマンドでは、より効率的なアルゴリズムを使うことができる可能性があります。

また、ヒートマップは距離行列ではなく、元のデータ行列に基づいて作成されることも意味します。

距離行列または非類似度行列が重要である場合、名前を入力すると、HCluster が非類似度行列で正方行列を埋めます。



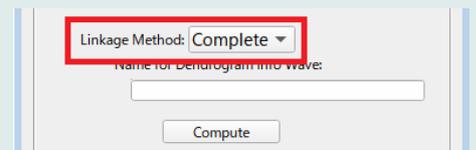
7. Linkage Method として Complete を選択します。

階層型クラスタリングが進むにつれて、データ行列の複数の行を表すノード間の非類似度を計算する方法が必要になります。

この選択は非常に重要です。

プロセスにおける特定のステップで、どのベクトルがグループ化されるかに大きな影響を与えます。

今回の場合、Average と Complete は同様の結果を示します。

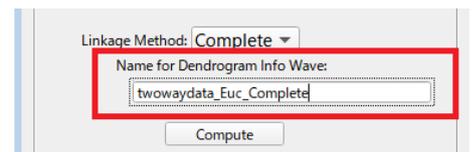


8. デンドログラムウェーブの名前を設定します。

一般に計算によるクラスタリングは試行錯誤のプロセスであるため、適切に名前を付けることは不可欠です。

このケースでは「twowaydata_Euc_Complete」を使います。

これは「元のウェーブ名_非類似度法_連結法」を名前にコード化したものです。



9. Compute ボタンをクリックします。

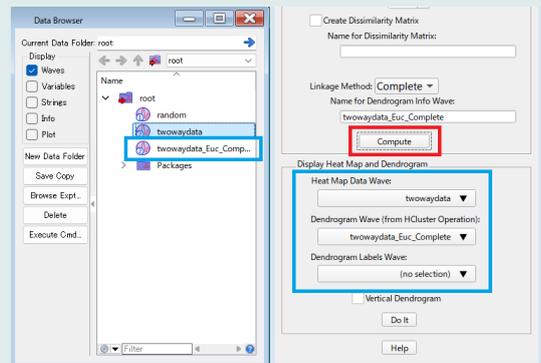
あまり目立った変化は起こっていませんが、デンドログラムを記述したウェーブが作成されています。

デンドログラム付きヒートマップを表示する手順を早めるために、下部のセクションでいくつかの選択が行われています。

Data Browser で twowaydata_Euc_Complete というウェーブを見つけることができます。

このウェーブの内容については、Hierarchical Clustering (Analysis.ihf ファイル内) を参照してください。

このウェーブのデータは解析が非常に難しく、デンドログラムのノードへの参照の連鎖を含んでいます。



Heat Map Data Wave で twowaydata が選択されていることを確認してください。

Dendrogram Wave で twowaydata_Euc_Complete が選択されていることを確認してください。

10. データマトリックスの各行にラベルが付いたテキストウェーブが必要です。

コマンドウィンドウで次を実行してください。

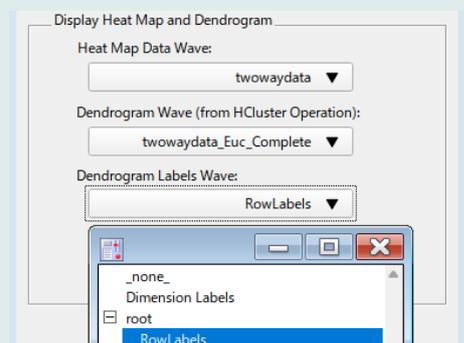
```
Make/N=20/T RowLabels = "Row "+num2str(p)
```

11. Dendrogram Labels Wave で RowLabels を選択します。

必ずしもラベルウェーブを選択する必要はありませんが、ラベルのないデンドログラムは通常、あまり有益ではありません。

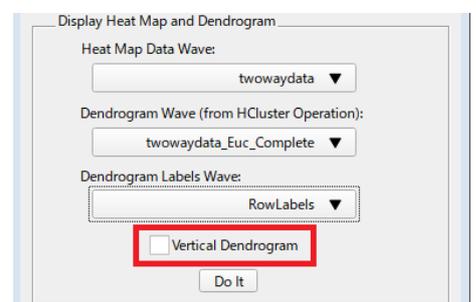
表示が作成されると、データ行列の行は、クラスタリングを反映するように並び替えられます。

ラベルは、並び替えをした後に何が何であるかがわかりやすくなります。



12. Vertical Dendrogram チェックボックスがオフになっていることを確認してください。

水平方向のデンドログラムでは、データ行列の列が縦方向に表示され、表で表示される形式と一致します。



13. Do It をクリックして、グラフウィンドウを大きくして見やすくします。

類似したベクトルをまとめるために、行がどのように並べ替えられたかを確認できます。

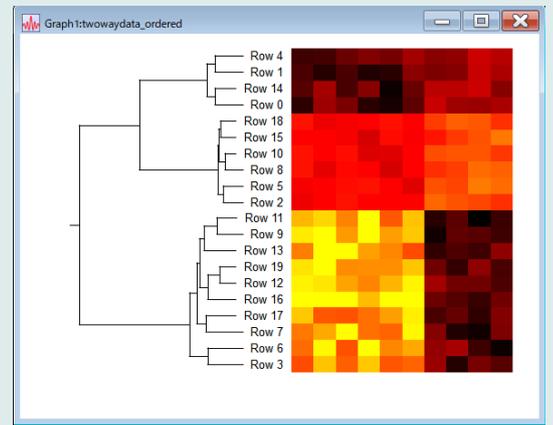
ベクトルのペアを繋ぐ縦線は、接続されたノードペア間の非類似性に比例する水平位置に描画されます。

3つの明らかなクラスタがあります。

データをそのように作成したので、これは良い結果です。

上部の4つのデータのクラスタは、下部の10行のもののもれよりも、中段の6行のデータに似ています。

これらの2つのグループは、一番下のグループの縦線よりも右側に縦線で繋がれています。



ennoise で元のデータ行列を作成したため、正確な内容はデータ行列を作る前の乱数発生器の状態に依存します。そのため、デンドログラムの全体的な外観は非常に似ていますが、任意の位置の行番号は上図とは異なる場合があります。

行は、縦のノード接続線が交差しないように、上から下に向かって順番に並べられています。

いくつかのメモ：

デンドログラムは、Igor の 描画ツールを使って線とラベルを描画します。

必要に応じて、メニュー Graph → Show Tools を選択して、さまざまな方法でグラフを修正することができます。

しかし、あまり良い考えではないかもしれません。ウィンドウのサイズを変更したときなどは、必要があればデンドログラムは自動的に消去されて、再描画されてしまうためです。もし手動で変更していた場合、その変更は失われてしまいます。

デンドログラムは User Front レイヤーに描画され、デンドログラムの線とラベルは注釈以外のグラフのすべての部分の上に配置されます。

すべてのデンドログラムの描画は、名前の付いたグループを使って行われ、デンドログラムが再描画される時には、それらのグループのみが消去されます。

これにより、独自に追加した描画オブジェクトはそのまま残ります。

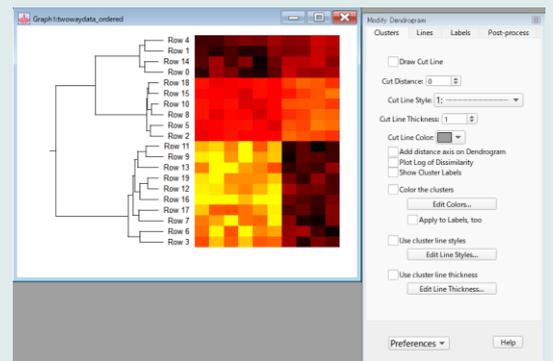
デンドログラム表示の拡張

ここからは、色などを変えていきます。

1. ヒートマップとデンドログラムを前面に表示し、メニュー Analysis → Hierarchical Clustering → Modify Dendrogram を選択します。

コントロールパネルには多くのオプションが表示されます。

グラフウィンドウにドッキングされているため、どのグラフ用のものか、すぐにわかります。



2. コントロールパネルでは、「Clusters」タブが選択されているはずですが。

Draw Cut Line チェックボックスをオンにします。

デンドログラムの右端に点線がひかれます。

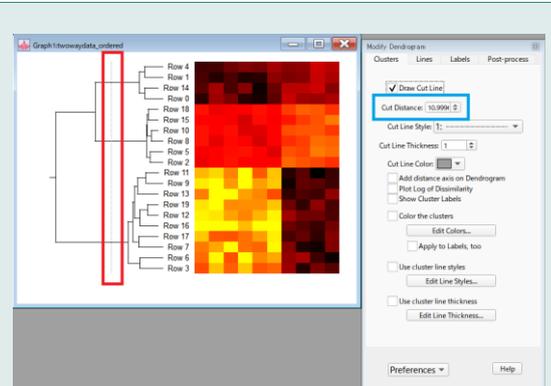
カットラインはカット値を示していて、これは異なるクラスタを定義する非類似値です。

現状は明らかに誤った位置にあります。



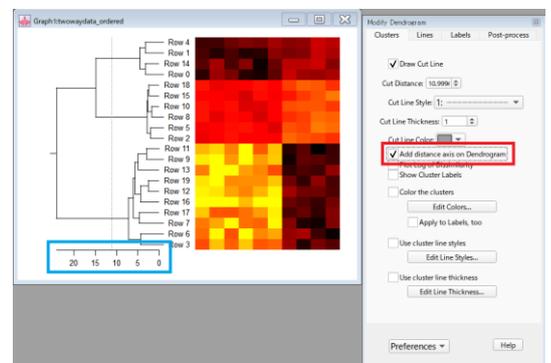
3. カットラインを左にドラッグして、3つの明白なクラスタに接続された3本の水平線と交差するようにします。

また、コントロールパネルの「Cluster」タブの Cut Distance でカット値を設定することもできます。



4. 「Cluster」タブで、Add distance axis on Dendrogram チェックボックスをオンにします。

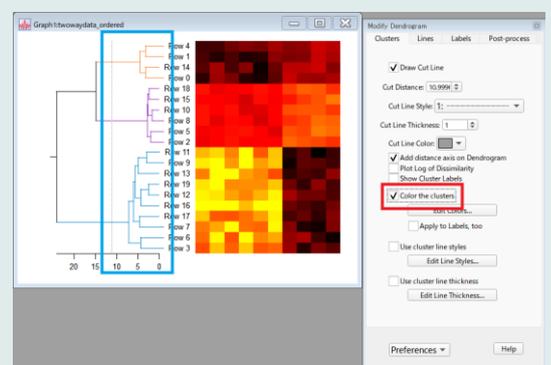
これを使うと、カットラインをドラッグすることが少しは楽になります。



5. Color the clusters チェックボックスをオンにします。

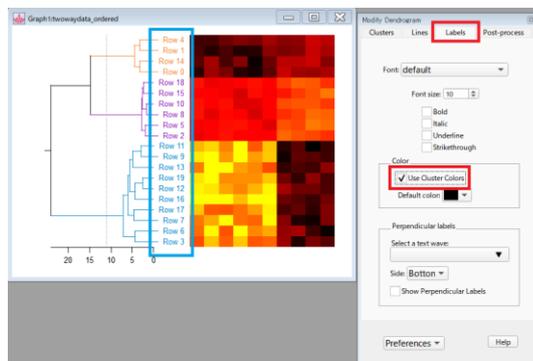
デンドログラムに色が付けられます。

線が細すぎて、色が少し見づらいかもしれません。



6. 「Labels」タブを選択します。

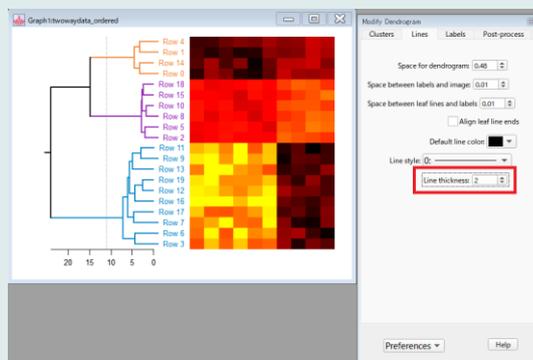
Color ボックスで、Use Cluster Colors チェックボックスをオンにします。



7. 「Lines」タブをクリックし、Line Thickness を 2 にします。

「Lines」タブには、デンドログラムのライン全体の外観の設定があります。

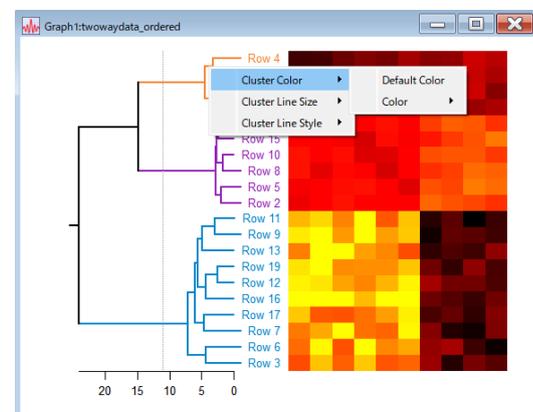
これらの設定は、「Cluster」タブの設定を使って、個々のクラスターごとに上書きすることができます。



8. 色のついたクラスタのうちの1つを右クリックします。

各クラスタの色、線の太さ、線のスタイルを個別に変更することができます。

「Default」を選択すると、「Lines」タブで設定したラインスタイル、色、太さに一致する関連した属性が設定されます。



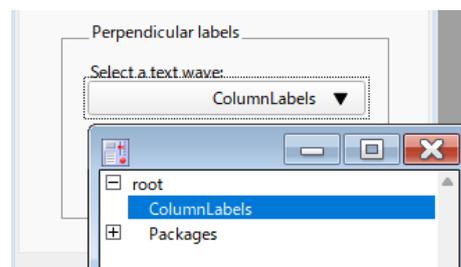
9. 列にラベルを付けるためのテキストウェーブを作成します。

次のコマンドを実行します。

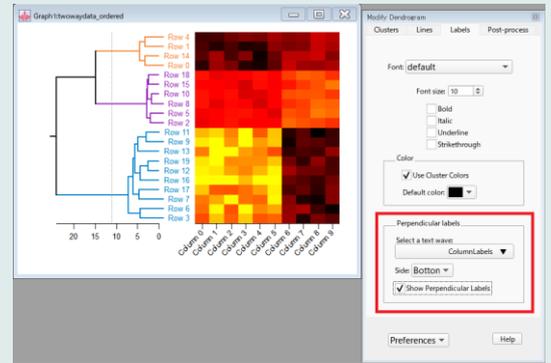
```
Make/N=10/T ColumnLabels = "Column "+num2str(p)
```

列（データベクトルの要素）が異なる種類の測定値を表している場合は、列にラベルを付けるとよいでしょう。この疑似データの場合、データベクトルの要素は互いに似ているデータです。

10. 「Labels」タブの「Perpendicular labels」ボックスで、Select a text wave で「ColumnLabels」を選択します。



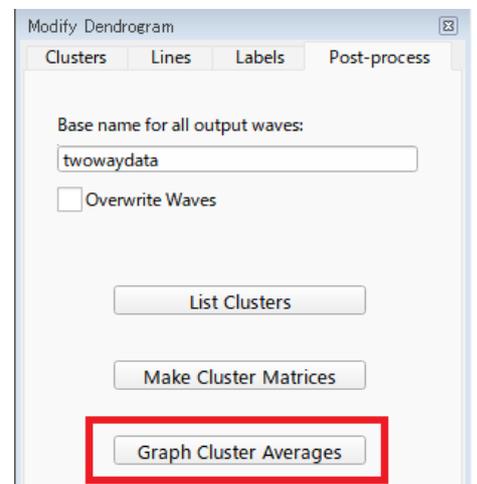
11. ポップアップメニューでラベルを表示する場所を選択し、Show Perpendicular Labels チェックボックスをオンにします。



ここでは説明しなかったヒートマップとデンドログラムの外観を変更する方法がいくつかあります。コントロールパネルを見てみると、ほとんどの設定は、説明を読まなくても理解できるはずです。このプロシージャファイルは、一般によく使われる YellowHot カラーテーブルを使ってヒートマップを作成します。これは、高輝度であるほど高温であるように見えます（したがって「ヒートマップ」と呼びます）。しかし、これは単なる画像プロットで、自由にカラーテーブルを使うことができます。

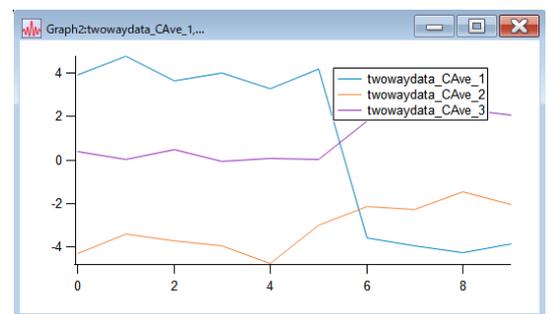
その他のいくつかの分析

コントロールパネルの「Post-process」タブに「Graph Cluster Averages」というボタンがあります。



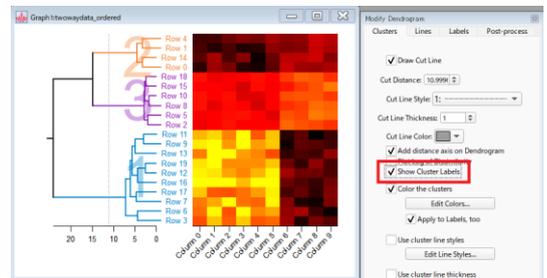
そのボタンをクリックすると、各クラスターのデータベクトルを抽出し、ベクトルを平均化してグラフ化します。3つのクラスターを持つ今回のケースでは、グラフは次のようになります。

トレースにはクラスターと同じ色が割り当てられ、トレースを生成するウェーブには、デンドログラムのウェーブ名に始まり、指定したクラスター番号の平均であることを示す表記で終わる名前が付けられます。



内部的には、クラスターには番号が割り当てられています。しかし、デンドログラムの構成とクラスター番号の決定方法により、それらはデンドログラム上にランダムに配置されます。この番号は便利で、色、線の太さ、線のスタイルの編集機能では、すべてのこれらの番号を使います。

「Clusters」タブの Show Cluster Labels チェックボックスをオンにすると、クラスタ番号が表示されます。



2方向（行と列）デンドログラム付きヒートマップを作成する

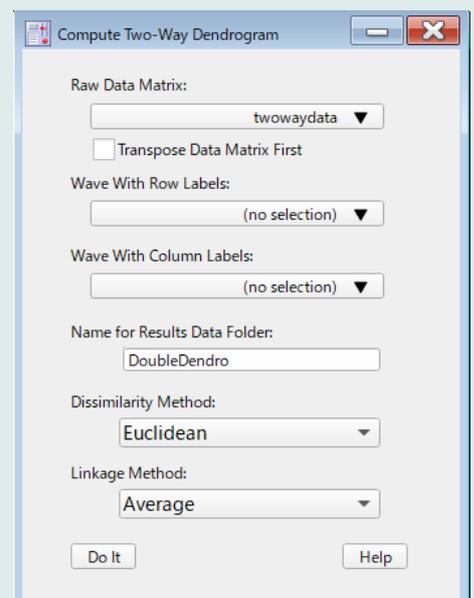
実際のデータは、サンプル抽出された母集団内だけでなく、測定自体にクラスタリング分析が必要な母集団の測定値であるかもしれません。

その場合、2方向デンドログラムが必要になります。

この例での疑似データは、2方向クラスタリングができるようになっています。

1. メニュー Analysis → Hierarchical Clustering → Heatmap and Two-way Dendrogram を選択します。

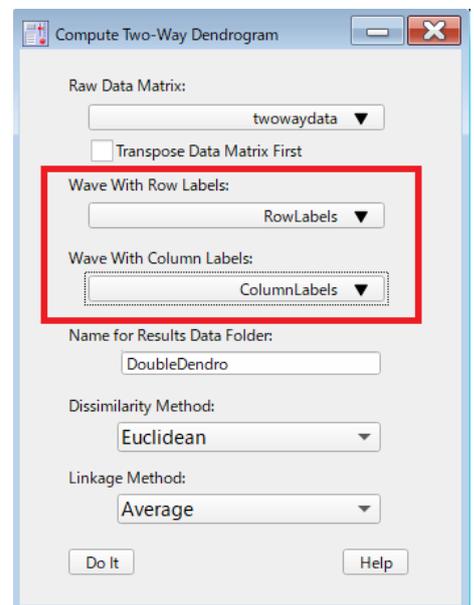
データセットに HCluster のコマンドを2回適用し、2つの垂直デンドログラムを描画するための新しいコントロールパネルが表示されます。



2. Row Data Matrix に twowaydata が表示されていることを確認してください。

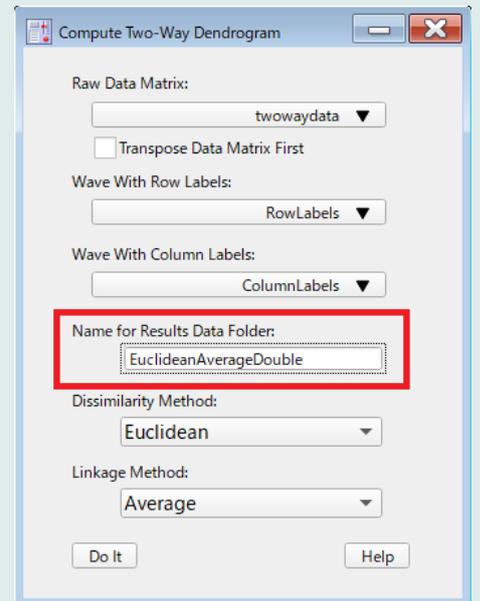
Row Labels ポップアップメニューで **RowLabels** を選択します。

Column Labels ポップアップメニューで **ColumnLabels** を選択します。



3. HCluster コマンドの結果を保存するデータフォルダーの名前として「EuclideanAverageDouble」と入力します。

コントロールパネルは、生のデータ行列に HCluster コマンドを適用し、その後、行列を転置して再度 HCluster を適用します。さらに、ヒートマップを作成し、2つの垂直デンドログラムを描画するために、並べ替えられたデータ行列が作成されます。このデータフォルダーには、すべてを整理しておくために必要な複数のウェーブが含まれます。

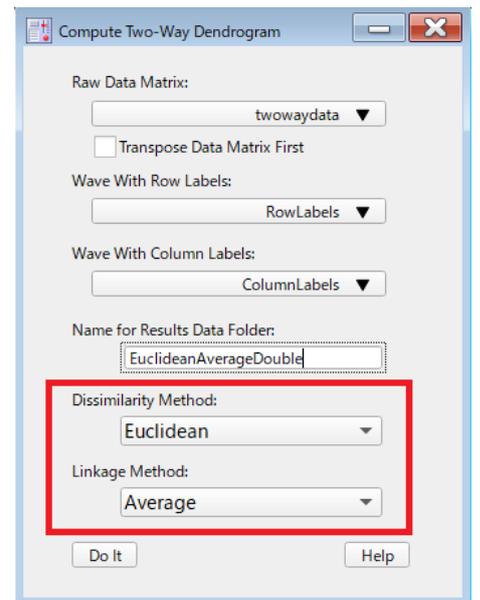


4. Dissimilarity Method で Euclidean が選択されていることを確認してください。

Linkage Method で Average が選択されていることを確認してください。

このコントロールパネルでは、2つの方向に対して異なる非類似度行列やリンク方法の適用はサポートされていません。

このコントロールパネルは、処理中の距離行列は保持していません。

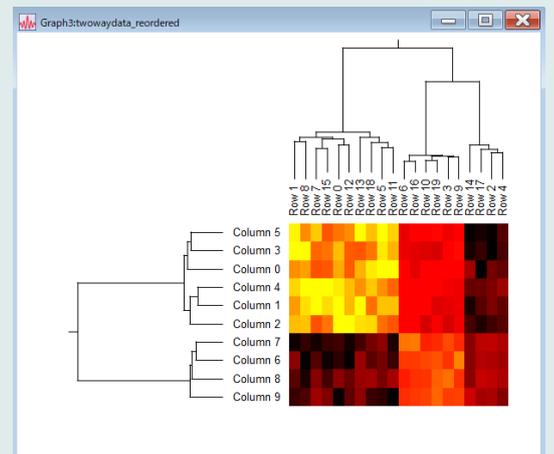


5. Do It をクリックします。

Igor では、行と列は通常通り、X 軸に沿って行、Y 軸に沿って列が表示されます。

転置が必要な場合は、Transpose Data Matrix First チェックボックスをオンにします。

こうする場合は、最初からの手順を繰り返す必要があります。



6. メニュー Analysis → Hierarchical Clustering →

Modify Horizontal Dendrogram を選択します。

これは、単一のデンドログラムの外観を変えるために使ったものと同じコントロールパネルを表示します。

色付けなどを変更してみてください。

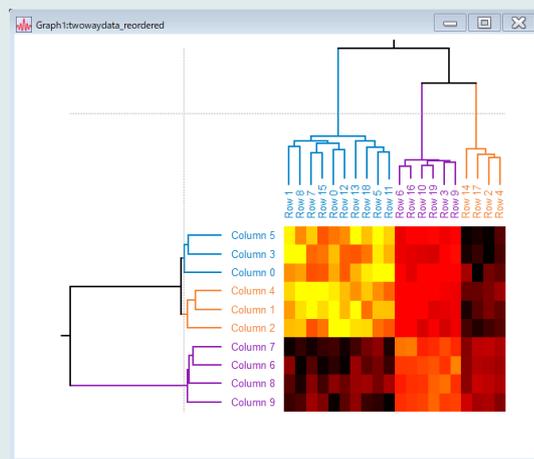
メニュー Analysis → Hierarchical Clustering → Modify

Vertical Dendrogram を選択します。

今度は垂直方向のデンドログラムを変更できるコントロールパネルが表示されます。

こちらでも色付けなどを変更してみてください。

少し変更してみた結果が右図です。



これで終了です。