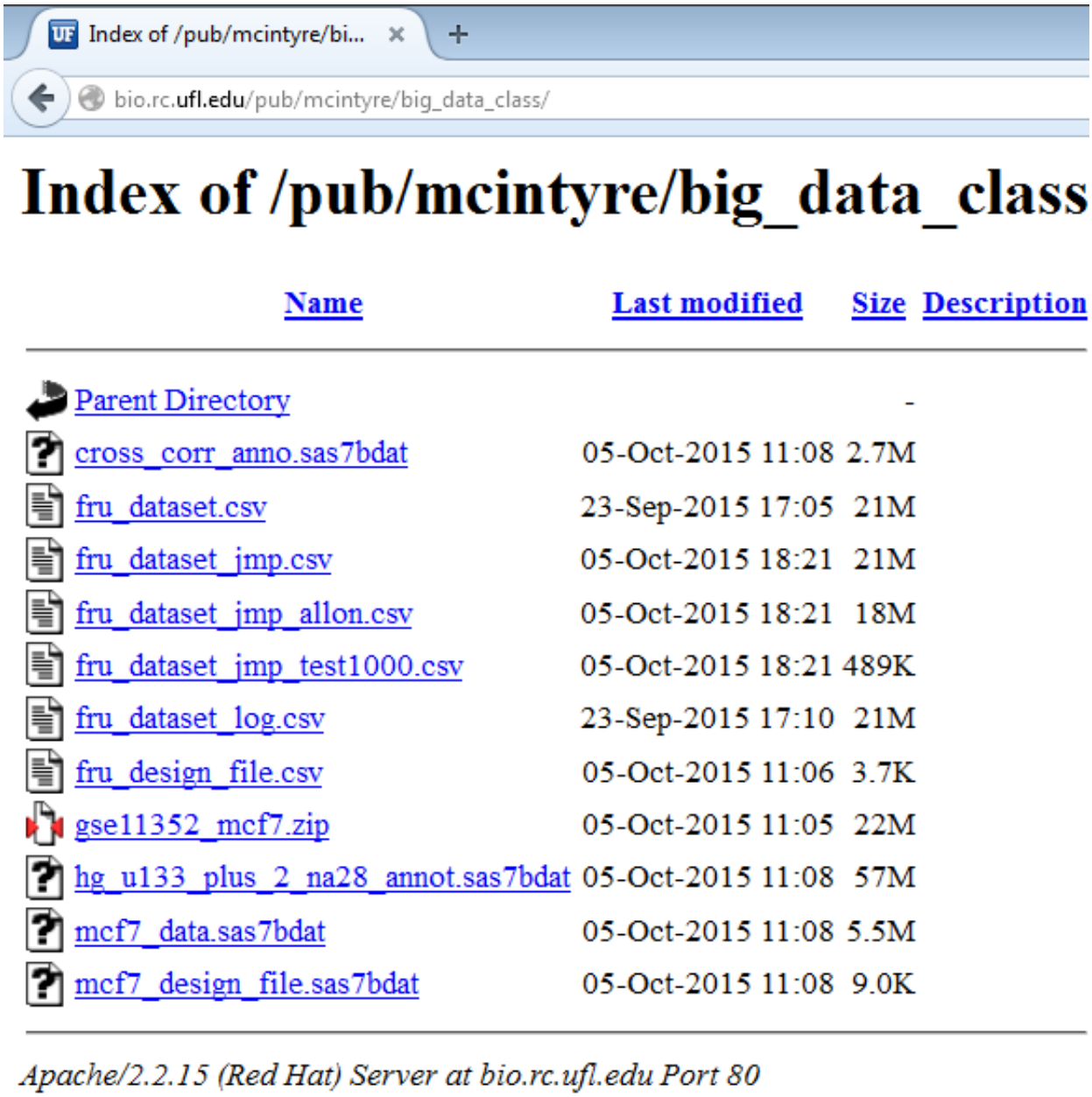
An introduction to Principal Components Analysis using JMP Genomics

This document is aimed at providing you with a step-by-step guide to performing principal components analysis (PCA) using the statistics software JMP Genomics. For the purposes of creating this tutorial, all steps were performed in JMP Genomics 7. While there will be some differences between different versions of JMP, this guide should still provide you with enough information to perform PCA using at least JMP Genomics 6 or later.

Data used in this tutorial

This tutorial will be using two datasets: GSE50515 and GSE11352. The first (GSE50515) is data used for Homework #13 and #14 and will be referred to as the "Fru" dataset. This dataset is an RNA-seq experiment comparing gene expression in *Drosophila melangaster* heads between different genotypes. The second is a microarray experiment examining the effect of 10nM estradiol exposure on gene expression in MCF-7 breast cancer cell lines. We will call it "MCF7".

These data have already been prepared and formatted for you. You can download them from http://bio.rc.ufl.edu/pub/mcintyre/big_data_class/. (http://bio.rc.ufl.edu/pub/mcintyre/big_data_class/)



The Fru dataset files you will need are:

```
fru_dataset_jmp.csv           (all data)
fru_dataset_jmp_allon.csv     (all fusions expressed in all samples)
fru_dataset_jmp_test1000.csv  (1000 fusions expressed in all samples)
fru_design_file.csv           (design file)
```

The MCF7 dataset files you will need are:

```
mcf7_data.sas7bdat
mcf7_design_file.sas7bdat
```

It is recommended that you download these to separate folders to keep the resulting outputs separated.

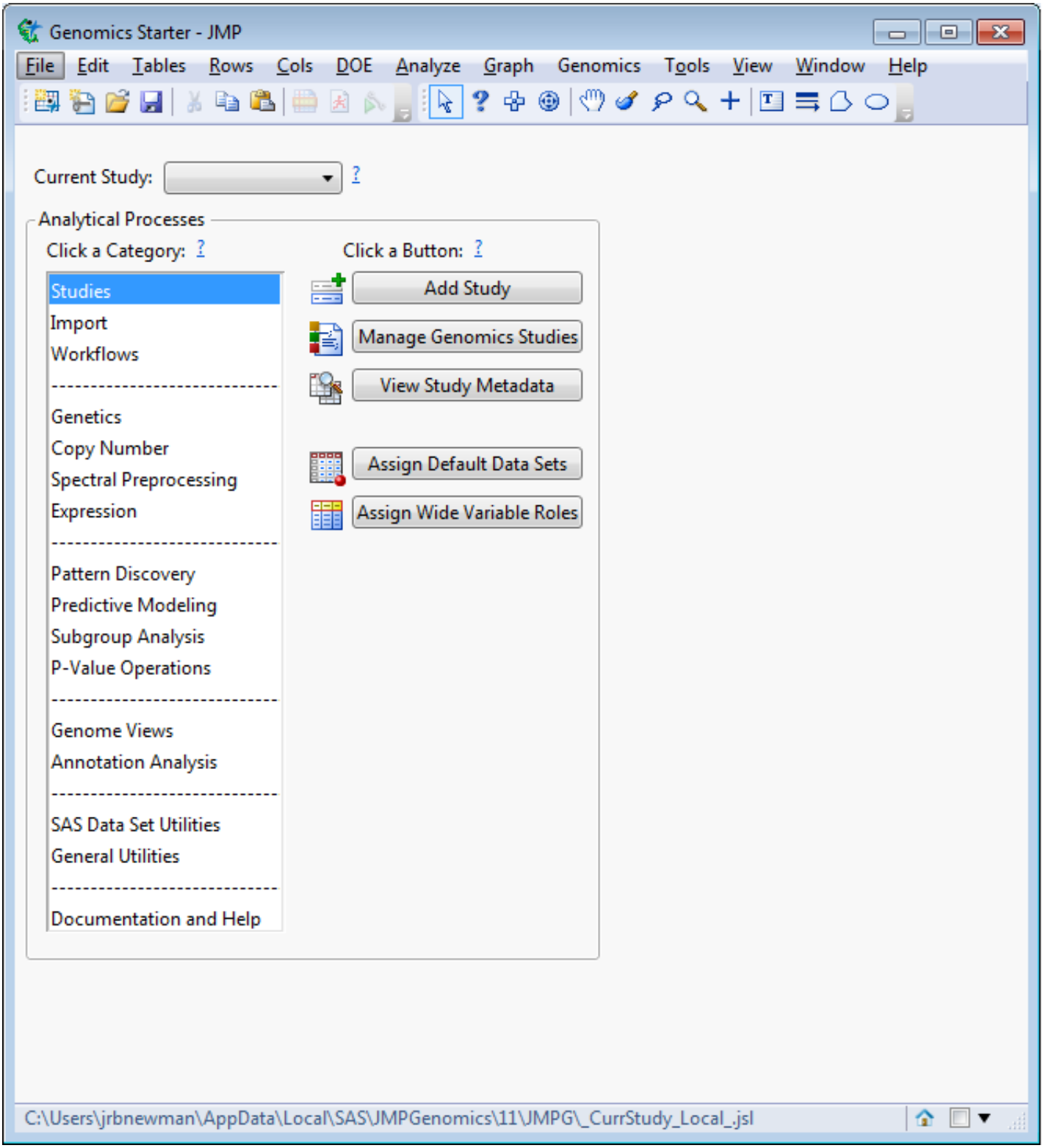
The Fru data

We will first run through how to perform PCA in JMP Genomics using the Fru dataset. In this section, we will learn how to import the data as text CSV files, transform them, then perform PCA on them using two methods available in JMP Genomics.

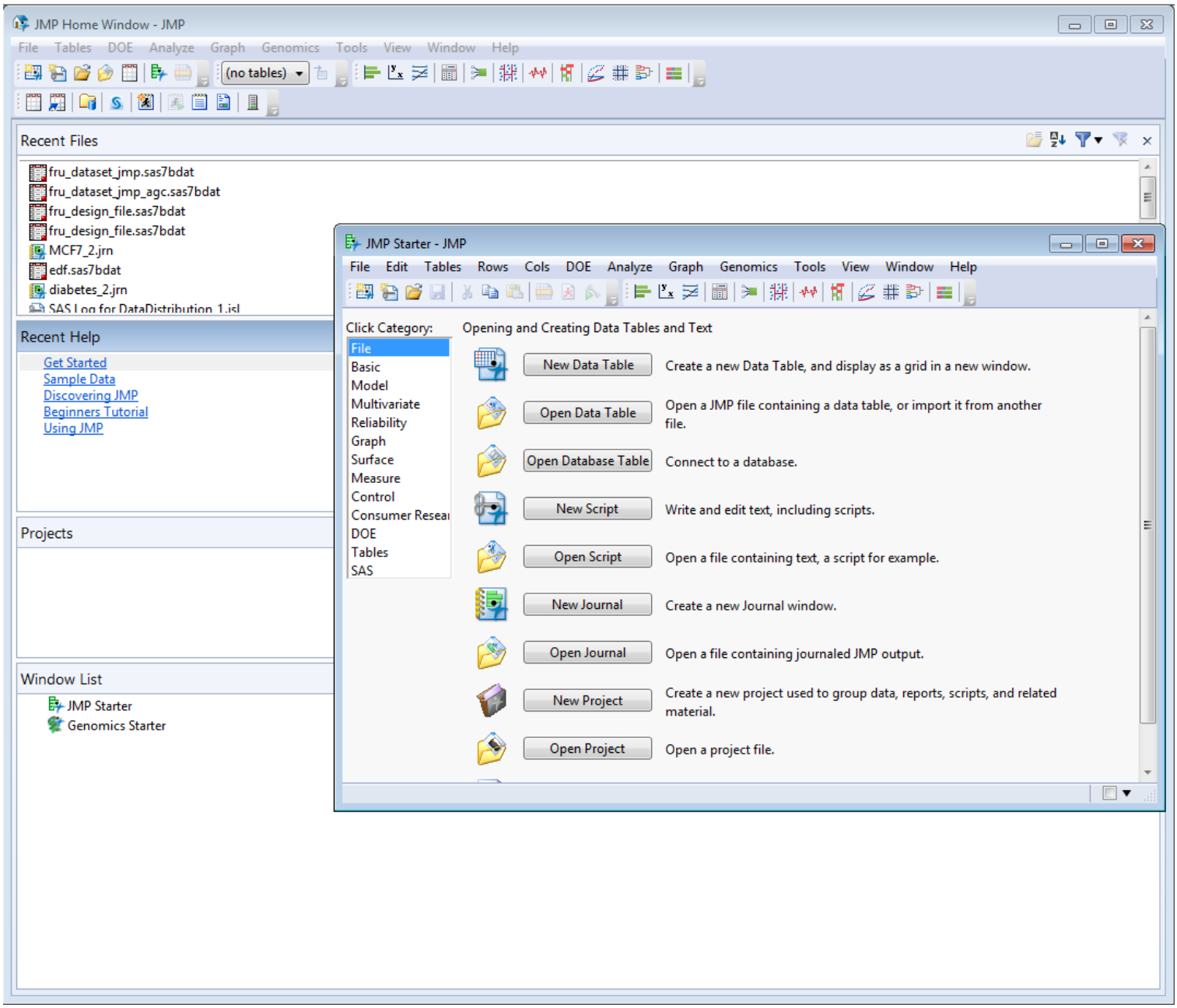
Importing the Fru data into JMP

Launch JMP

We will be begin by importing the Fru dataset into JMP. Launch JMP Genomics. You should see this window:



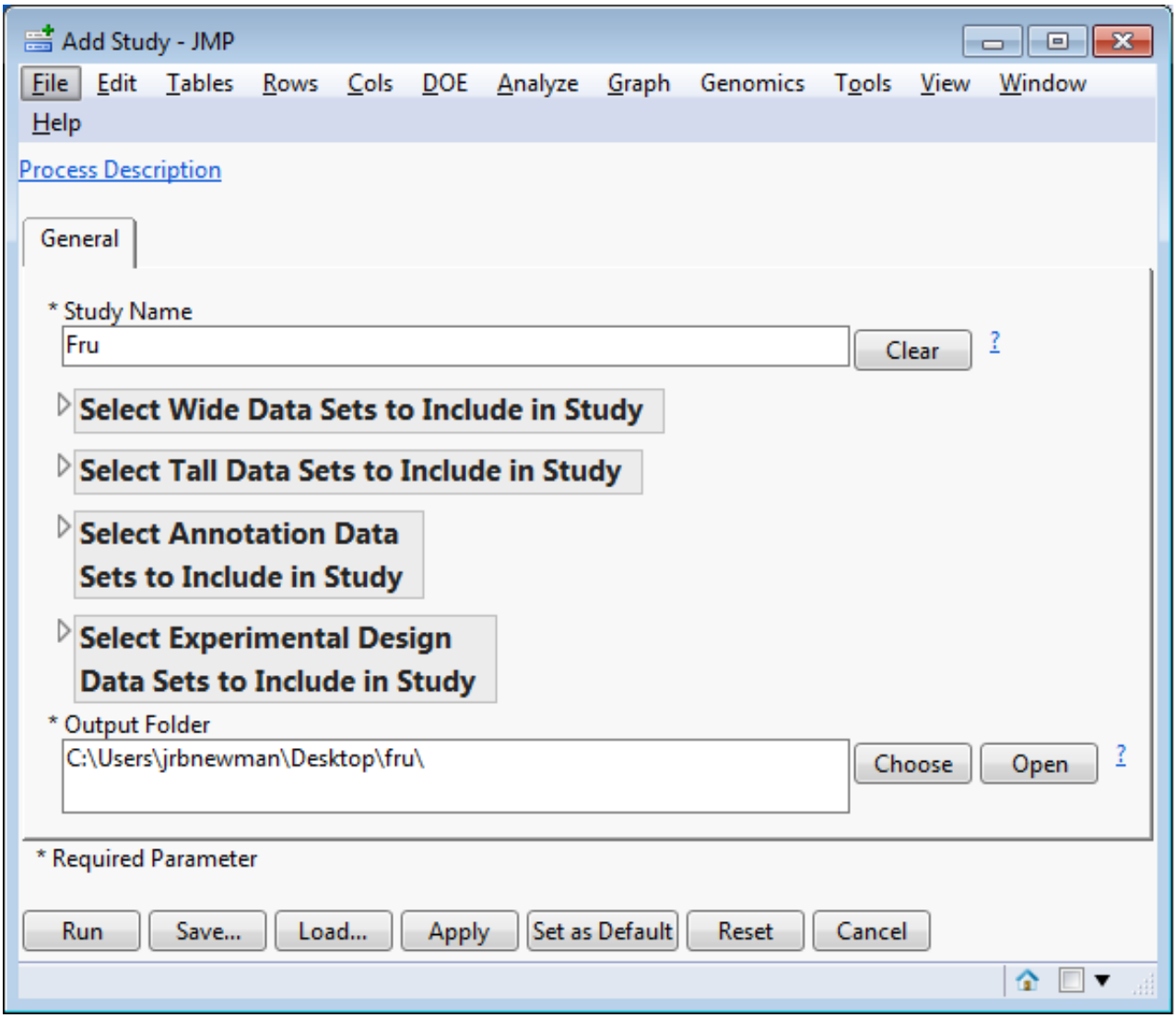
You may see these other screens:



If so, click on "View" then "Genomics Starter".

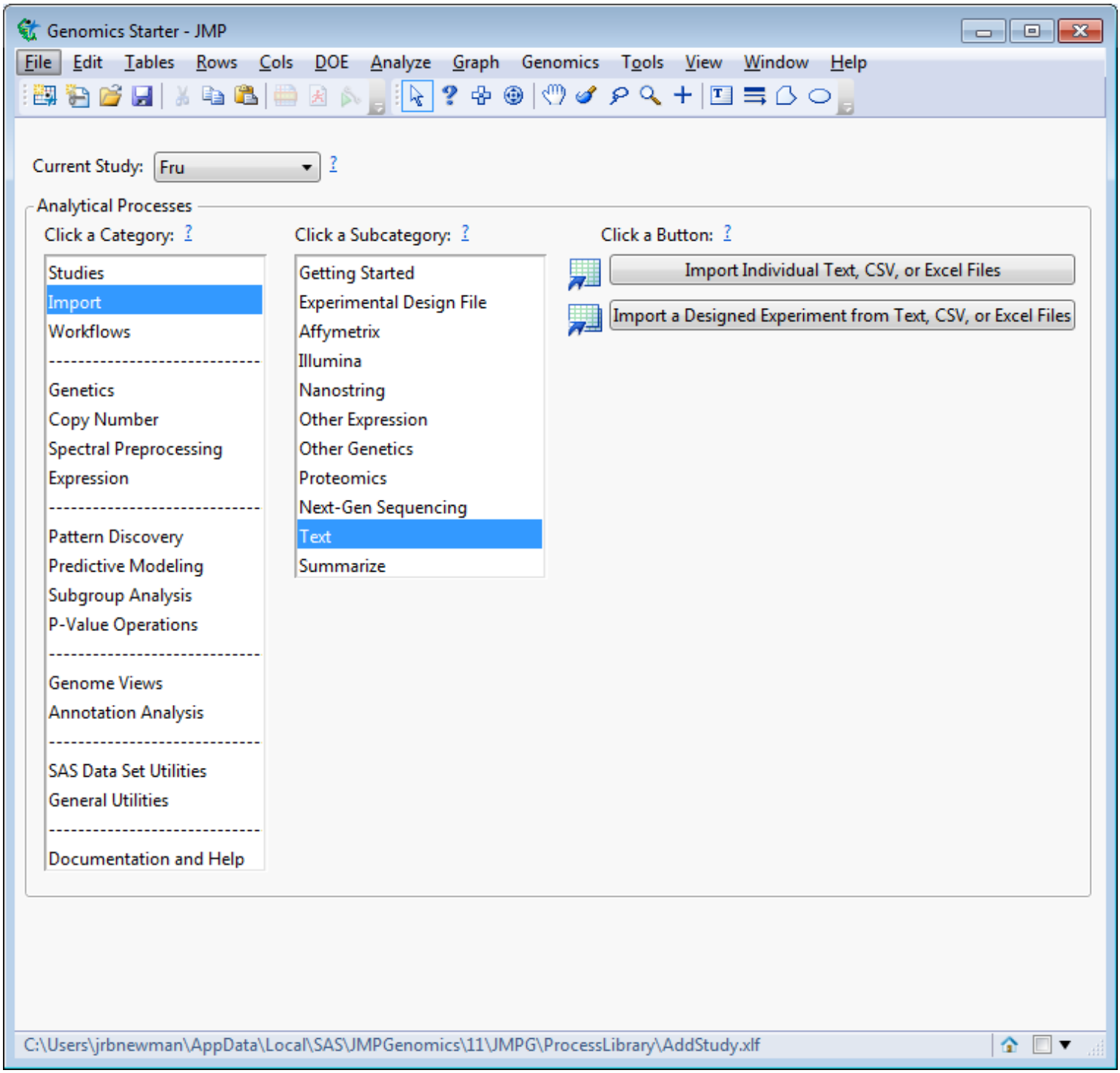
Adding Fru as a study

1.) On the Genomics Starter window, click on "Studies" in the Categories list then click "Add Study"
2.) Type "Fru" (no quotes) in the "Study Name" box.
3.) In the "Output Folder" box, select the folder where you saved the Fru datafiles.
4.) The window should look similar to this. Click "Run".



Importing the Fru data

5.) Back on the Genomics Starter window, select "Import" in the Categories list, selected the "Text" subcategory, then click on "Import individual Text, CSV, or Excel Files"



6.) In "Folder of Raw Files" add the path to the folder containing the Fru data CSVs

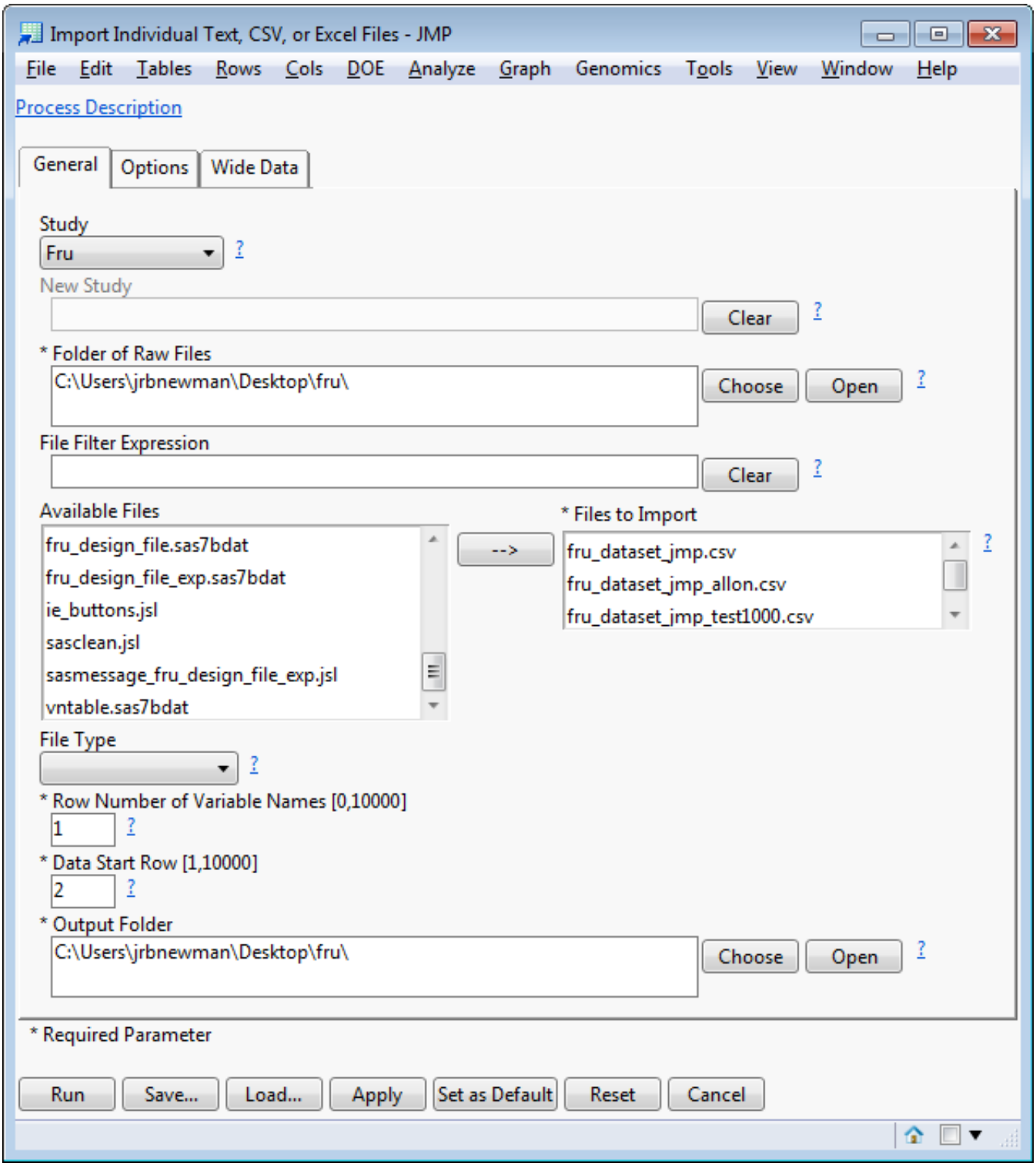
7.) In "Available Files", select **fru_dataset_jmp.csv**, **fru_dataset_jmp_allon.csv**, **fru_dataset_jmp_1000fus.csv** and "fru_design_file.csv" and click "-->".

8.) In "File Type" select "Comma Separated".

9.) Set "Row Number of Variable Names" to 1, and "Data Start Row" to 2.

10.) In "Output Folder" select the folder where you have saved your data.

11.) The window should look similar to this. Click "Run".

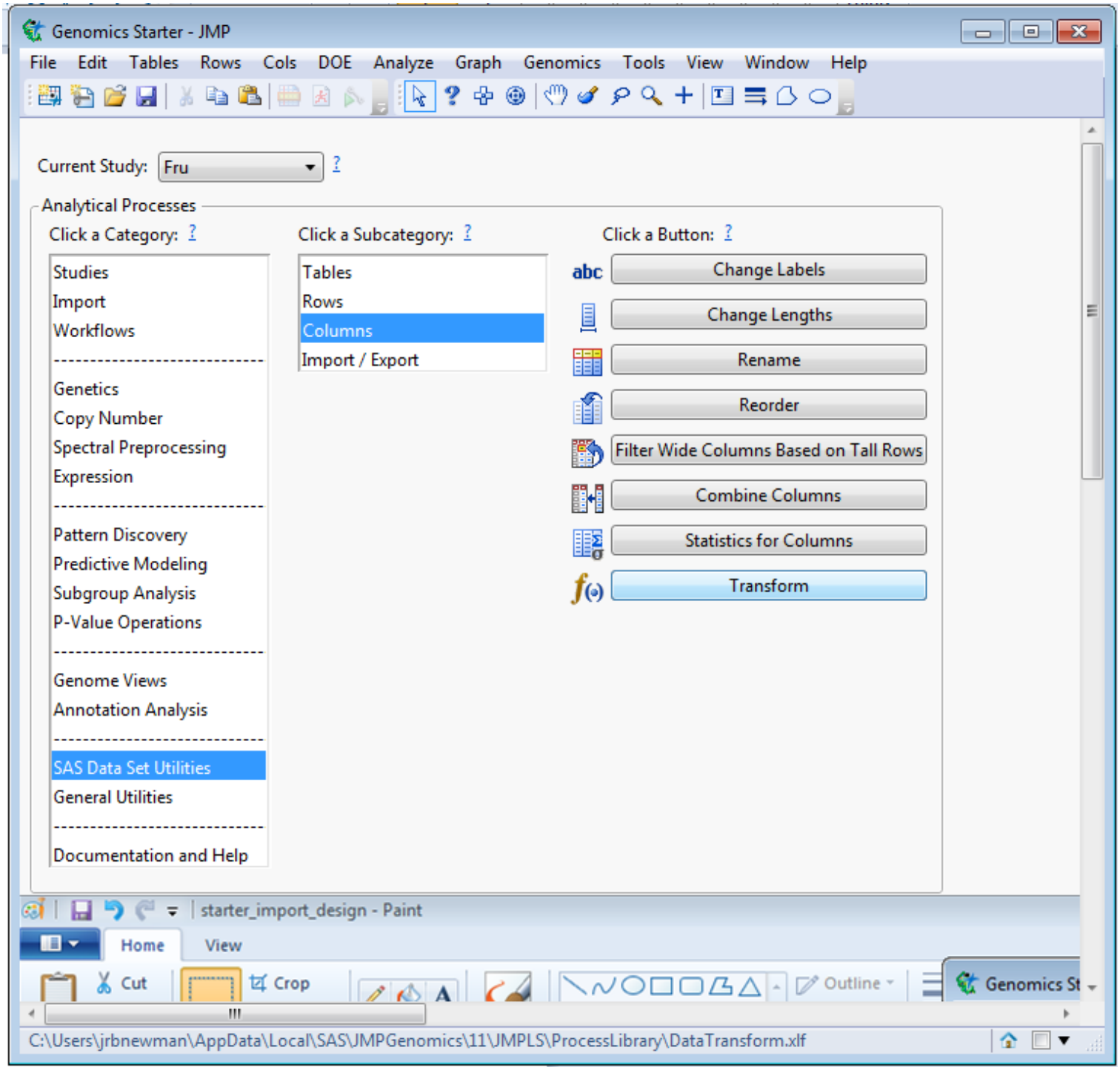


Now you should have the Fru data and associated design variables imported into JMP.

Transform data

Data transformation is a method to reduce "uninteresting" variation (e.g., technical variation, biological noise, etc.). You can remove this variation via several methods such as normalization and log-transformation. The data provided here have already been normalized and reported as "reads per kilobase per million" (RPKM). We can also transform the data by calculating the log2 of each measurement.

1.) On the Genomics Starter Window, select the "SAS Data Set Utilites" category, then the "Columns" subcategory, then click "Transform".



2.) In "Input SAS Data set" click "Choose" and select "fru_dataset_jmp.sas7bdat". Click "OK".

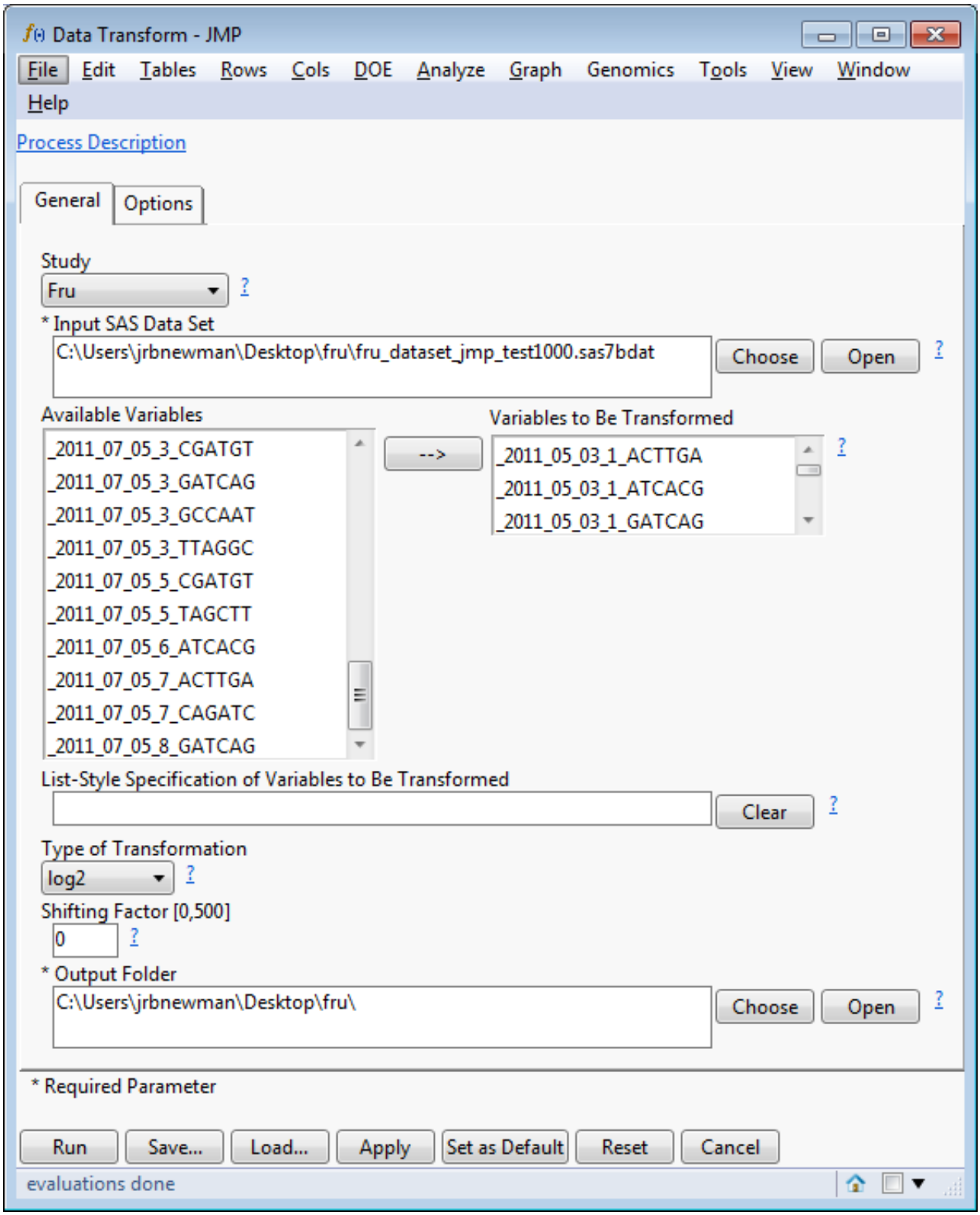
3.) In "Available Variables", select everything starting with "*2011*" and click "-->".

4.) For "Type of Transformation", select "log2".

5.) For "Shifting Factor" type in 1. This will allow "0" RPKM values to be log2 transformed by shifting the raw data by 1.

5.) Select your output folder in "Output Folder".

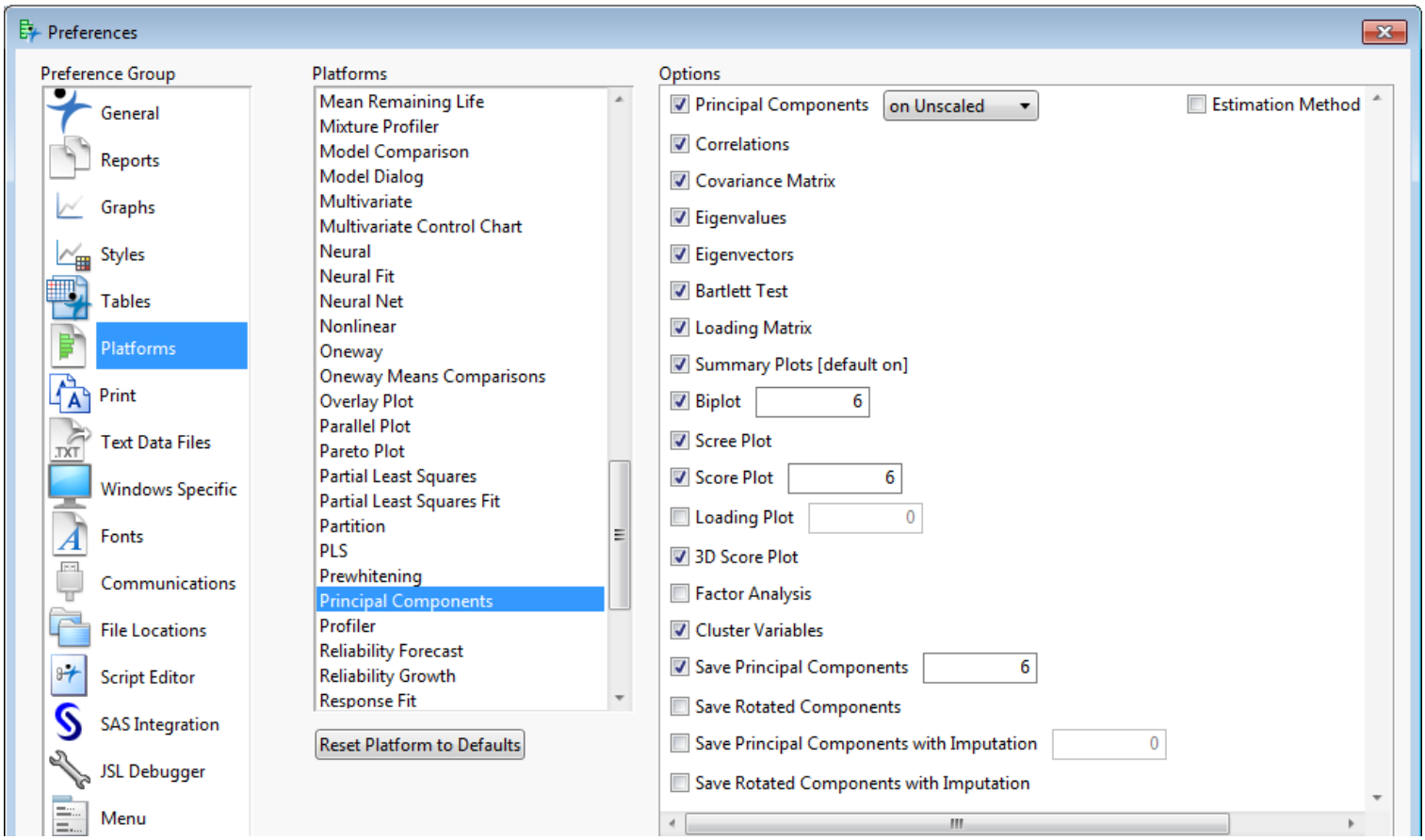
6.) The dialog box should look like the following. Click "Run".



The transformed dataset will be saved as **fru_dataset_jmp_dtf.sas7bdat**.

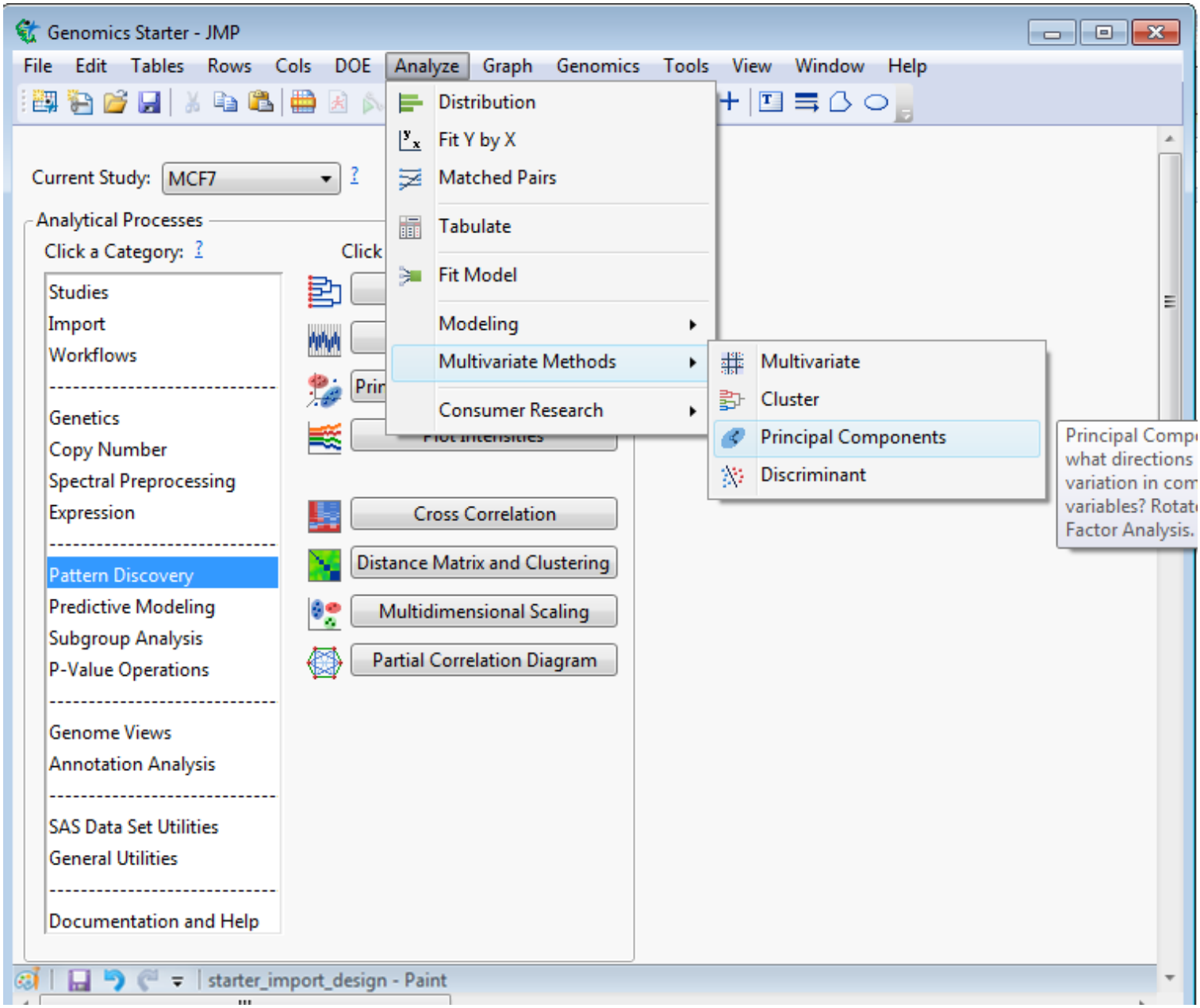
Running PCA

There are two ways to perform PCA in JMP, both providing the same outputs. One is available in all/most versions of JMP, another is native to only JMP Genomics and has some additional features. We will cover both here for completeness. First, however, we need to tell JMP what plots to output. On the menu bar, click "File" then "Preferences". On the "Preferences" dialog box, click on the "Platforms" Preference Group. Under "Platforms" scroll down and select "Principal Components". Check the boxes such that it looks like below:



Method 1 - via the Analyze menu

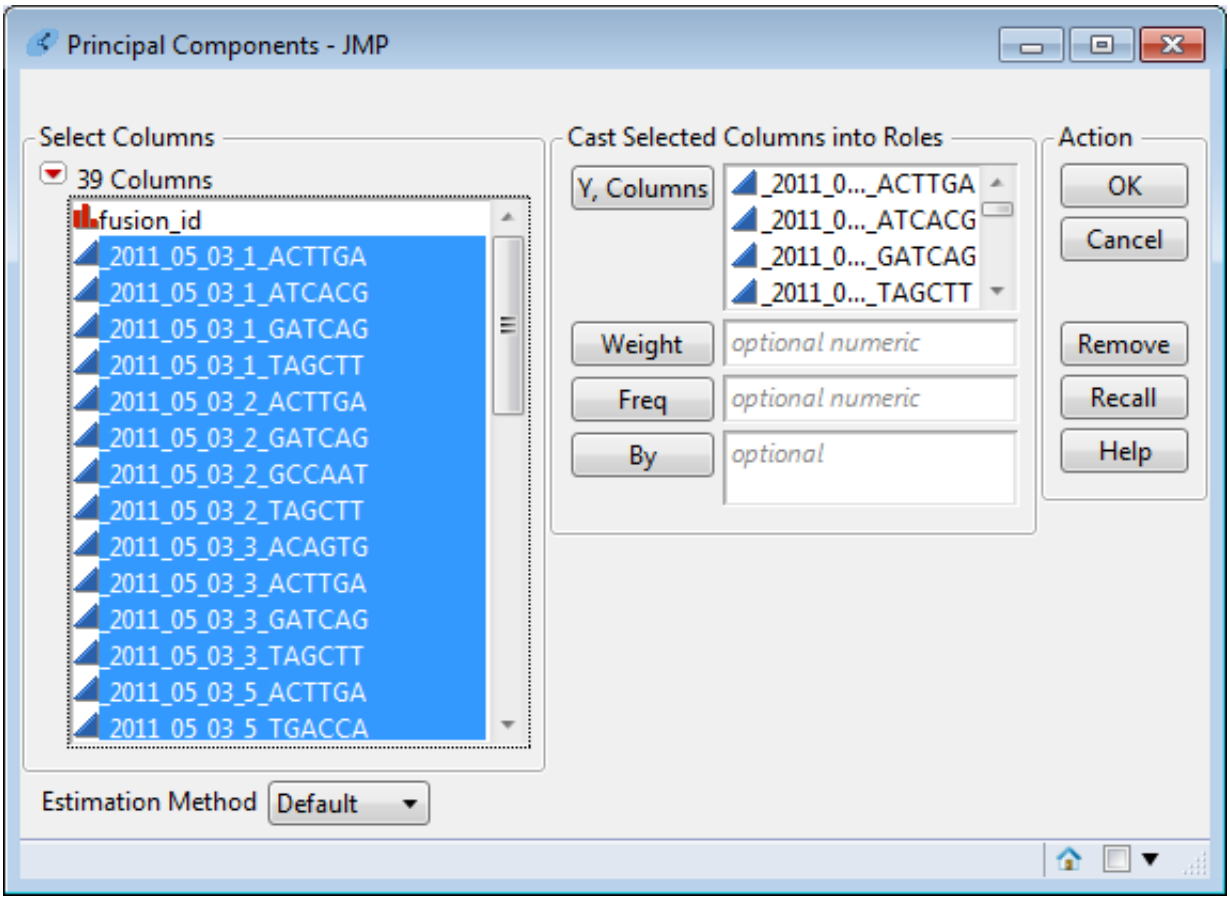
1.) Click on "Analyze" on the menu bar, select "Multivariate Methods", then "Principal Components".



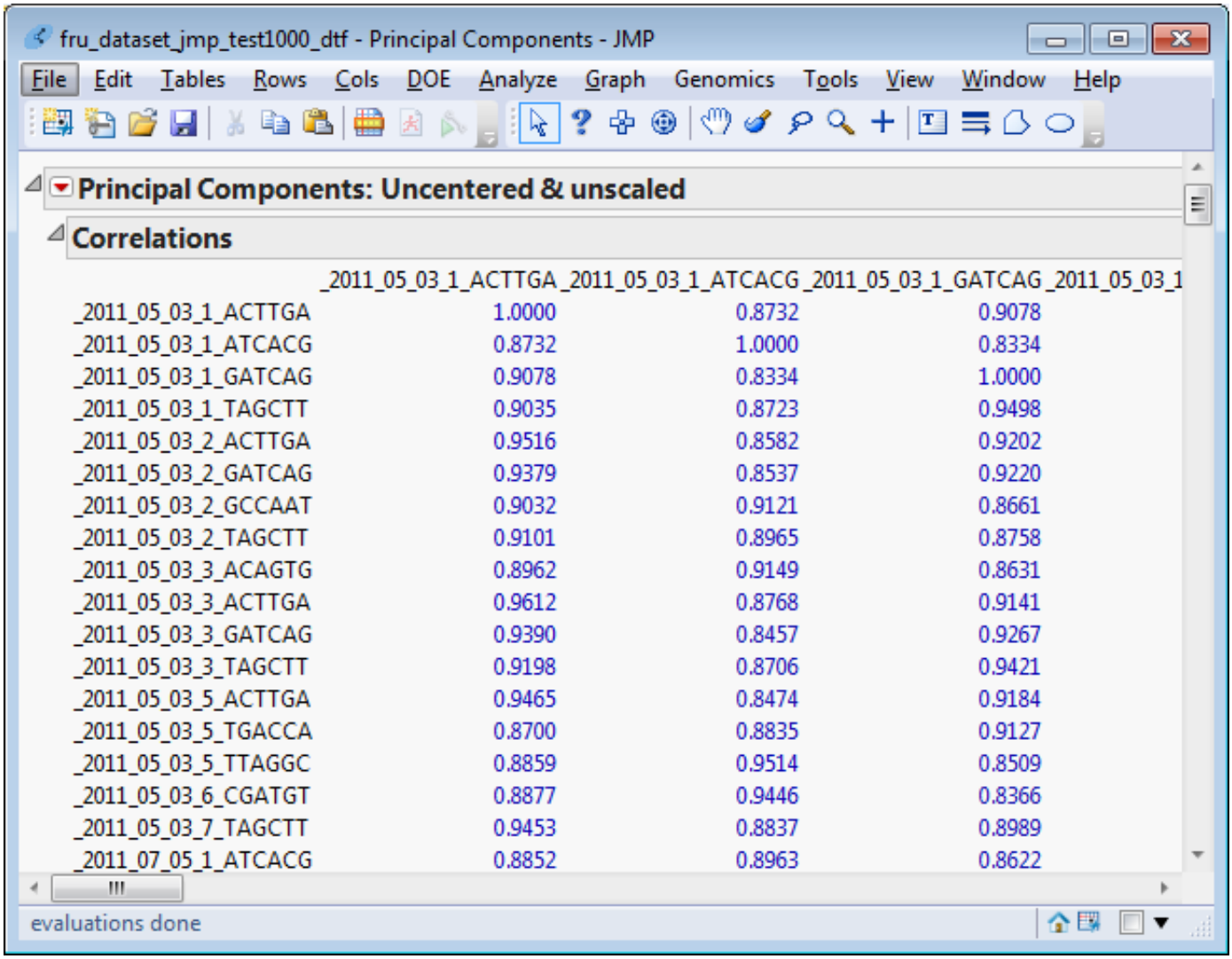
2.) A dialog box should open and ask you to select a dataset. Select "fru_dataset_jmp_dtf" (or "fru_dataset_jmp_test1000_dtf.sas7bdat") and click "Open".

3.) In "Select Columns" select everything starting with "*2011*" and click on "Y, Columns".

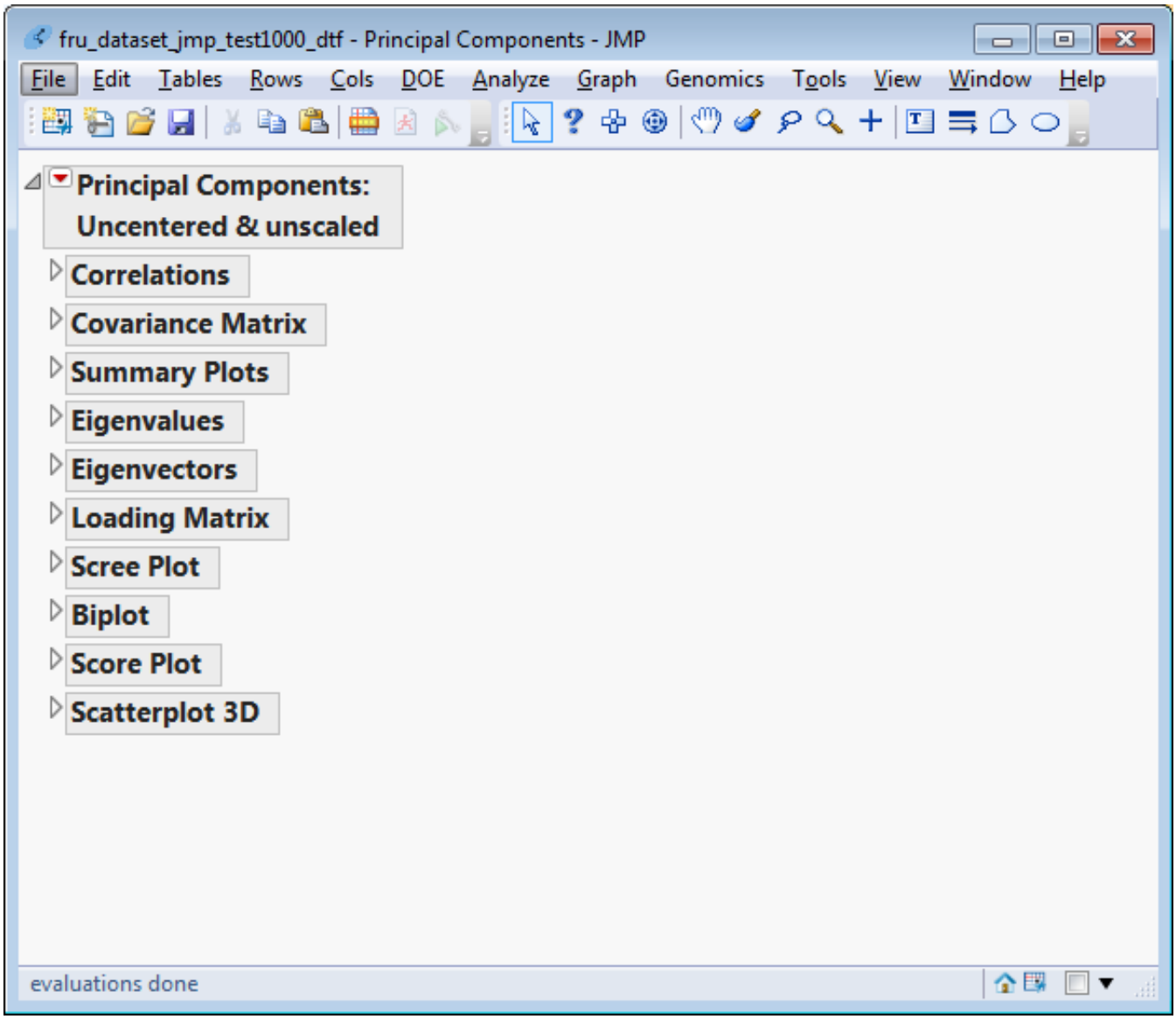
4.) You should see a dialog box similar to this. Leave everything else as its default setting and click on "Run".



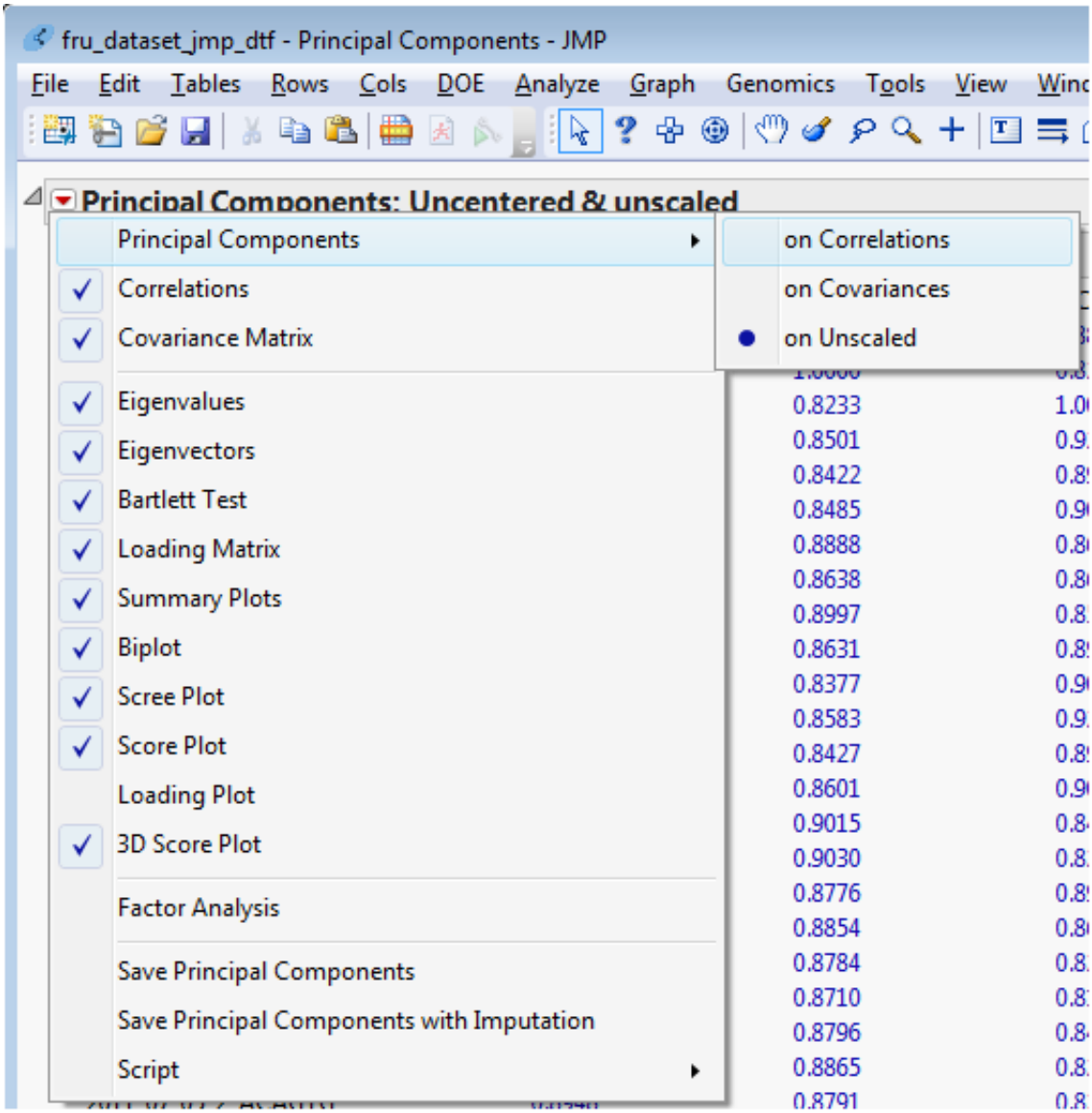
5.) The output window should appear and look like this. For now, click on the grey triangle next to everything (other than "Principal Components: Uncentered & unscaled") and we will go through each section individually.



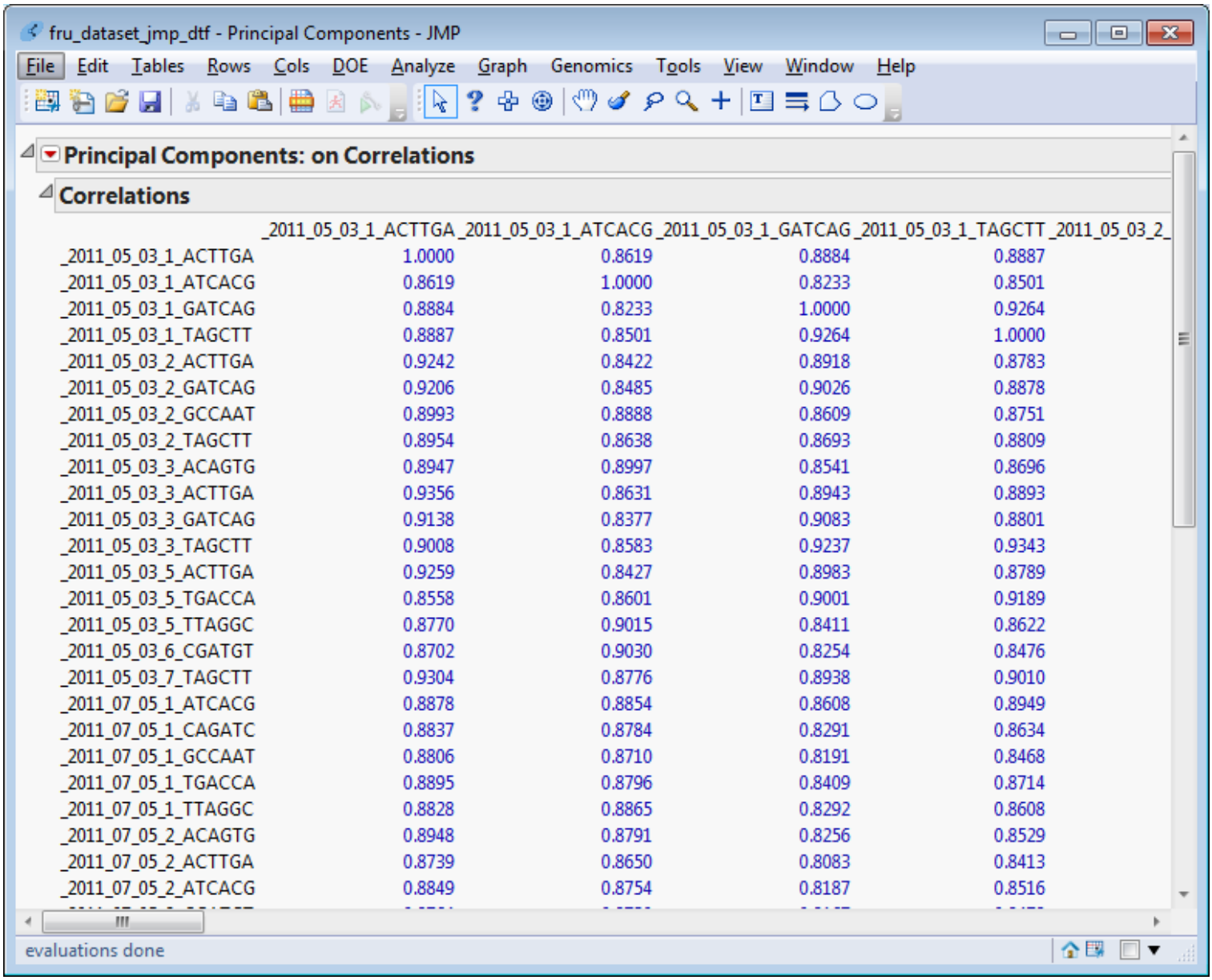
	_2011_05_03_1_ACTTGA	_2011_05_03_1_ATCACG	_2011_05_03_1_GATCAG	_2011_05_03_1
_2011_05_03_1_ACTTGA	1.0000	0.8732	0.9078	
_2011_05_03_1_ATCACG	0.8732	1.0000	0.8334	
_2011_05_03_1_GATCAG	0.9078	0.8334	1.0000	
_2011_05_03_1_TAGCTT	0.9035	0.8723	0.9498	
_2011_05_03_2_ACTTGA	0.9516	0.8582	0.9202	
_2011_05_03_2_GATCAG	0.9379	0.8537	0.9220	
_2011_05_03_2_GCCAAT	0.9032	0.9121	0.8661	
_2011_05_03_2_TAGCTT	0.9101	0.8965	0.8758	
_2011_05_03_3_ACAGTG	0.8962	0.9149	0.8631	
_2011_05_03_3_ACTTGA	0.9612	0.8768	0.9141	
_2011_05_03_3_GATCAG	0.9390	0.8457	0.9267	
_2011_05_03_3_TAGCTT	0.9198	0.8706	0.9421	
_2011_05_03_5_ACTTGA	0.9465	0.8474	0.9184	
_2011_05_03_5_TGACCA	0.8700	0.8835	0.9127	
_2011_05_03_5_TTAGGC	0.8859	0.9514	0.8509	
_2011_05_03_6_CGATGT	0.8877	0.9446	0.8366	
_2011_05_03_7_TAGCTT	0.9453	0.8837	0.8989	
_2011_07_05_1_ATCACG	0.8852	0.8963	0.8622	



You can change how the principal components are calculated, either as "uncentered/unscaled", "Correlations", "Covariances". For this guide, we will change this to "correlations". Click on the red drop down arrow next to "Principal Components", select the "Principal Components" menu item, then click "on Correlations".

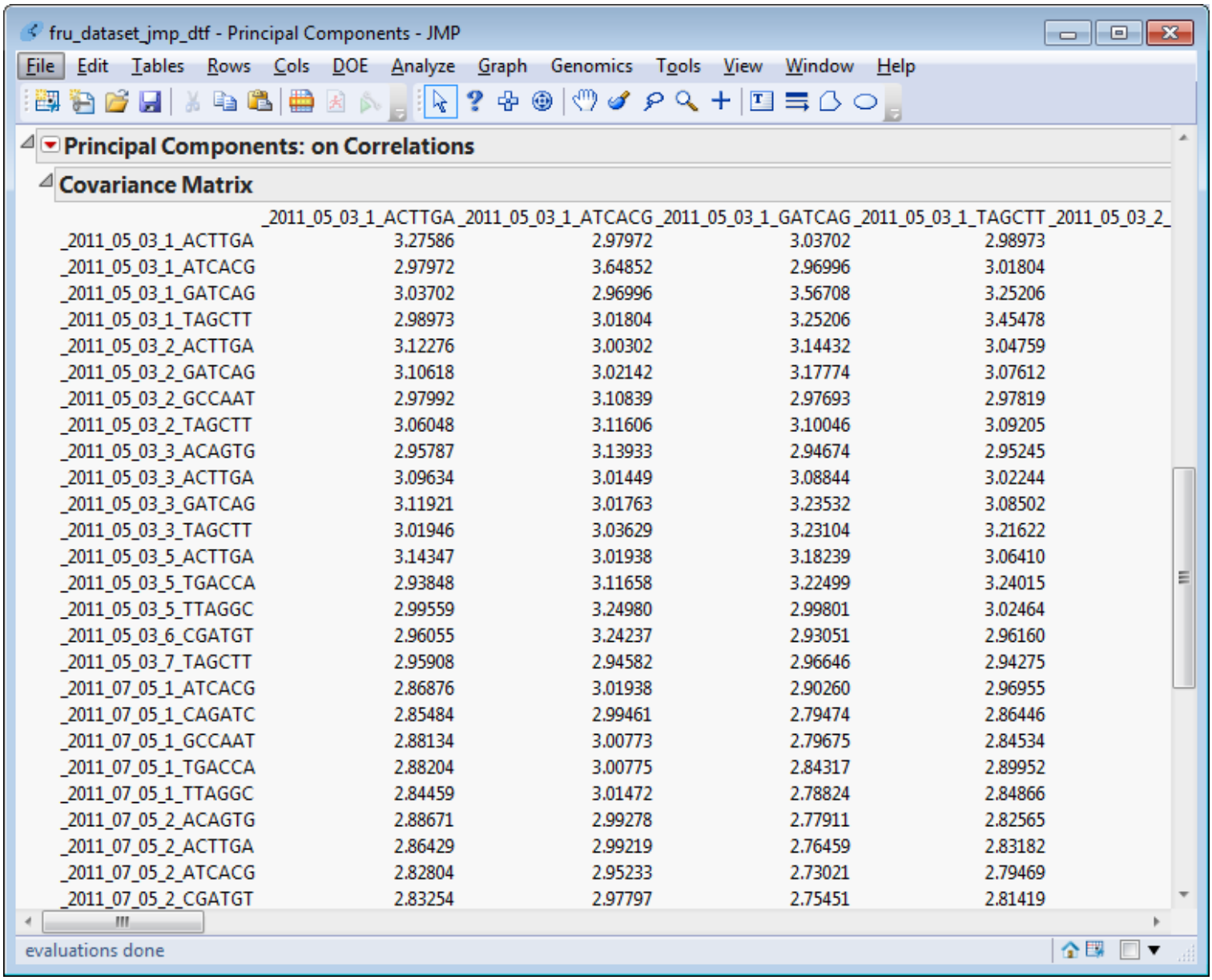


6.) First expand "Correlations". This is the correlation matrix of expression values between each sample.



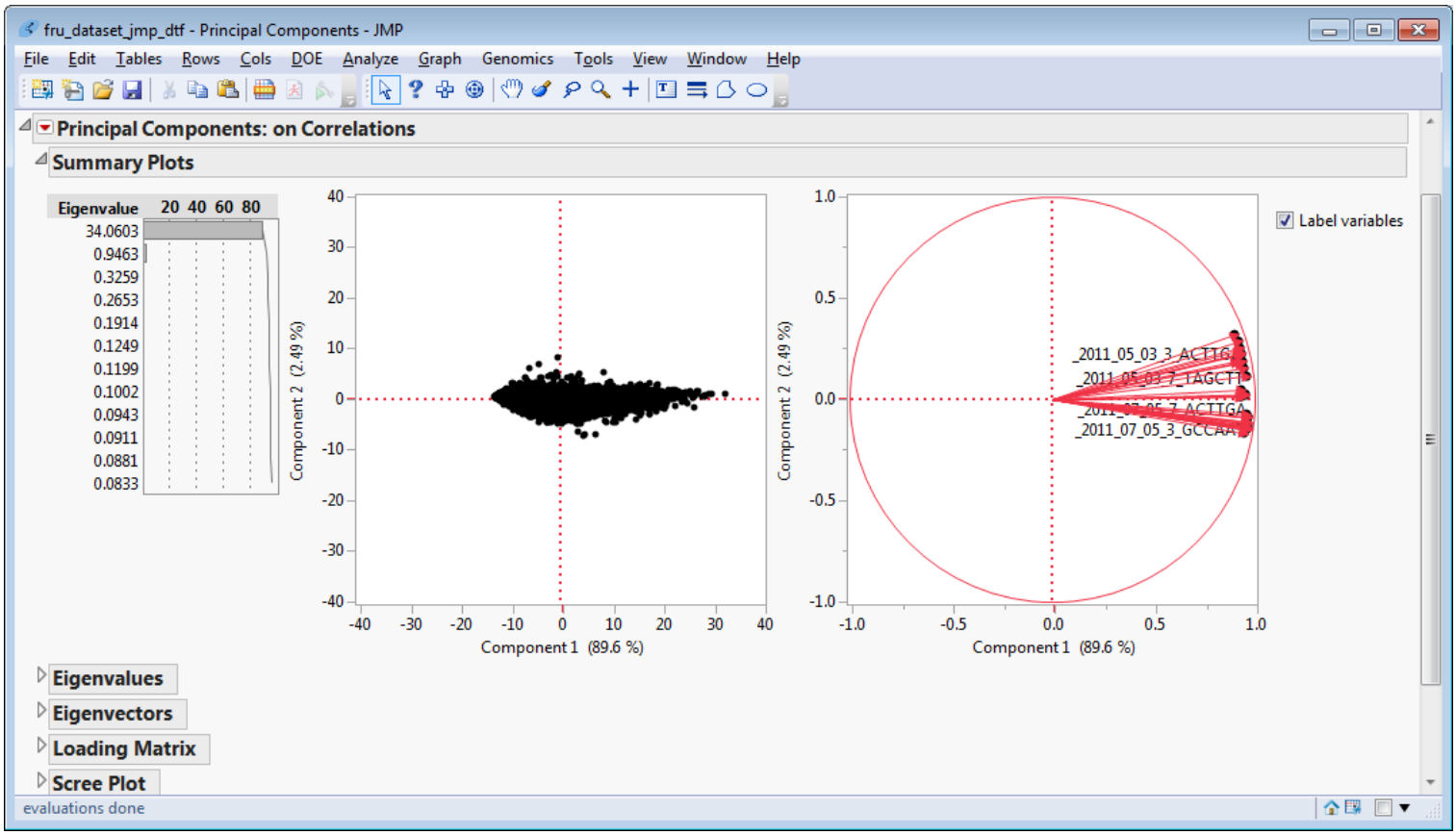
	_2011_05_03_1_ACTTGA	_2011_05_03_1_ATCACG	_2011_05_03_1_GATCAG	_2011_05_03_1_TAGCTT	_2011_05_03_2_
_2011_05_03_1_ACTTGA	1.0000	0.8619	0.8884	0.8887	
_2011_05_03_1_ATCACG	0.8619	1.0000	0.8233	0.8501	
_2011_05_03_1_GATCAG	0.8884	0.8233	1.0000	0.9264	
_2011_05_03_1_TAGCTT	0.8887	0.8501	0.9264	1.0000	
_2011_05_03_2_ACTTGA	0.9242	0.8422	0.8918	0.8783	
_2011_05_03_2_GATCAG	0.9206	0.8485	0.9026	0.8878	
_2011_05_03_2_GCCAAT	0.8993	0.8888	0.8609	0.8751	
_2011_05_03_2_TAGCTT	0.8954	0.8638	0.8693	0.8809	
_2011_05_03_3_ACAGTG	0.8947	0.8997	0.8541	0.8696	
_2011_05_03_3_ACTTGA	0.9356	0.8631	0.8943	0.8893	
_2011_05_03_3_GATCAG	0.9138	0.8377	0.9083	0.8801	
_2011_05_03_3_TAGCTT	0.9008	0.8583	0.9237	0.9343	
_2011_05_03_5_ACTTGA	0.9259	0.8427	0.8983	0.8789	
_2011_05_03_5_TGACCA	0.8558	0.8601	0.9001	0.9189	
_2011_05_03_5_TTAGGC	0.8770	0.9015	0.8411	0.8622	
_2011_05_03_6_CGATGT	0.8702	0.9030	0.8254	0.8476	
_2011_05_03_7_TAGCTT	0.9304	0.8776	0.8938	0.9010	
_2011_07_05_1_ATCACG	0.8878	0.8854	0.8608	0.8949	
_2011_07_05_1_CAGATC	0.8837	0.8784	0.8291	0.8634	
_2011_07_05_1_GCCAAT	0.8806	0.8710	0.8191	0.8468	
_2011_07_05_1_TGACCA	0.8895	0.8796	0.8409	0.8714	
_2011_07_05_1_TTAGGC	0.8828	0.8865	0.8292	0.8608	
_2011_07_05_2_ACAGTG	0.8948	0.8791	0.8256	0.8529	
_2011_07_05_2_ACTTGA	0.8739	0.8650	0.8083	0.8413	
_2011_07_05_2_ATCACG	0.8849	0.8754	0.8187	0.8516	

7.) Collapse "Correlations" and expand "Covariance Matrix". This is the variance-covariance matrix between each sample.



	_2011_05_03_1_ACTTGA	_2011_05_03_1_ATCACG	_2011_05_03_1_GATCAG	_2011_05_03_1_TAGCTT	_2011_05_03_2_
_2011_05_03_1_ACTTGA	3.27586	2.97972	3.03702	2.98973	
_2011_05_03_1_ATCACG	2.97972	3.64852	2.96996	3.01804	
_2011_05_03_1_GATCAG	3.03702	2.96996	3.56708	3.25206	
_2011_05_03_1_TAGCTT	2.98973	3.01804	3.25206	3.45478	
_2011_05_03_2_ACTTGA	3.12276	3.00302	3.14432	3.04759	
_2011_05_03_2_GATCAG	3.10618	3.02142	3.17774	3.07612	
_2011_05_03_2_GCCAAT	2.97992	3.10839	2.97693	2.97819	
_2011_05_03_2_TAGCTT	3.06048	3.11606	3.10046	3.09205	
_2011_05_03_3_ACAGTG	2.95787	3.13933	2.94674	2.95245	
_2011_05_03_3_ACTTGA	3.09634	3.01449	3.08844	3.02244	
_2011_05_03_3_GATCAG	3.11921	3.01763	3.23532	3.08502	
_2011_05_03_3_TAGCTT	3.01946	3.03629	3.23104	3.21622	
_2011_05_03_5_ACTTGA	3.14347	3.01938	3.18239	3.06410	
_2011_05_03_5_TGACCA	2.93848	3.11658	3.22499	3.24015	
_2011_05_03_5_TTAGGC	2.99559	3.24980	2.99801	3.02464	
_2011_05_03_6_CGATGT	2.96055	3.24237	2.93051	2.96160	
_2011_05_03_7_TAGCTT	2.95908	2.94582	2.96646	2.94275	
_2011_07_05_1_ATCACG	2.86876	3.01938	2.90260	2.96955	
_2011_07_05_1_CAGATC	2.85484	2.99461	2.79474	2.86446	
_2011_07_05_1_GCCAAT	2.88134	3.00773	2.79675	2.84534	
_2011_07_05_1_TGACCA	2.88204	3.00775	2.84317	2.89952	
_2011_07_05_1_TTAGGC	2.84459	3.01472	2.78824	2.84866	
_2011_07_05_2_ACAGTG	2.88671	2.99278	2.77911	2.82565	
_2011_07_05_2_ACTTGA	2.86429	2.99219	2.76459	2.83182	
_2011_07_05_2_ATCACG	2.82804	2.95233	2.73021	2.79469	
_2011_07_05_2_CGATGT	2.83254	2.97797	2.75451	2.81419	

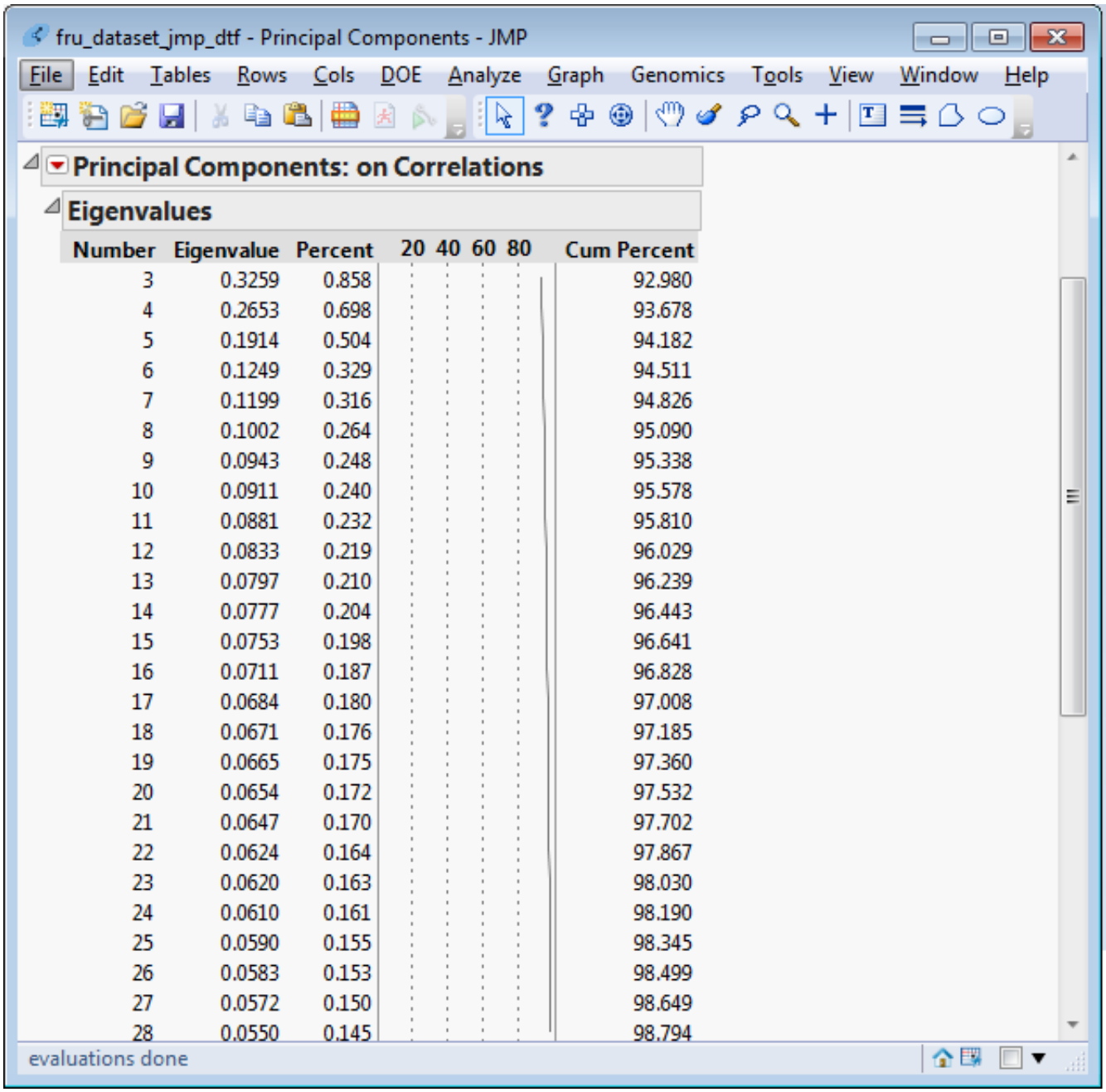
8.) Collapse "Covariance Matrix" and expand "Summary Plots". This gives you the eigenvalues for the data, and plots principal component 1 (PC1) against principal component 2 (PC2). Looking at the eigenvalues on left, we can see there is one particular large eigenvalue. What does this mean?



The plots here are the Score Plot and a Loadings Plot. The Score Plots graphs each component's calculate values in relation to the other, adjusting for mead and standard deviation. The Loadings Plots graphs the unrotated loading matrix between variables and components. The closer the value is to 1 the greater the effect of the component on the variable.

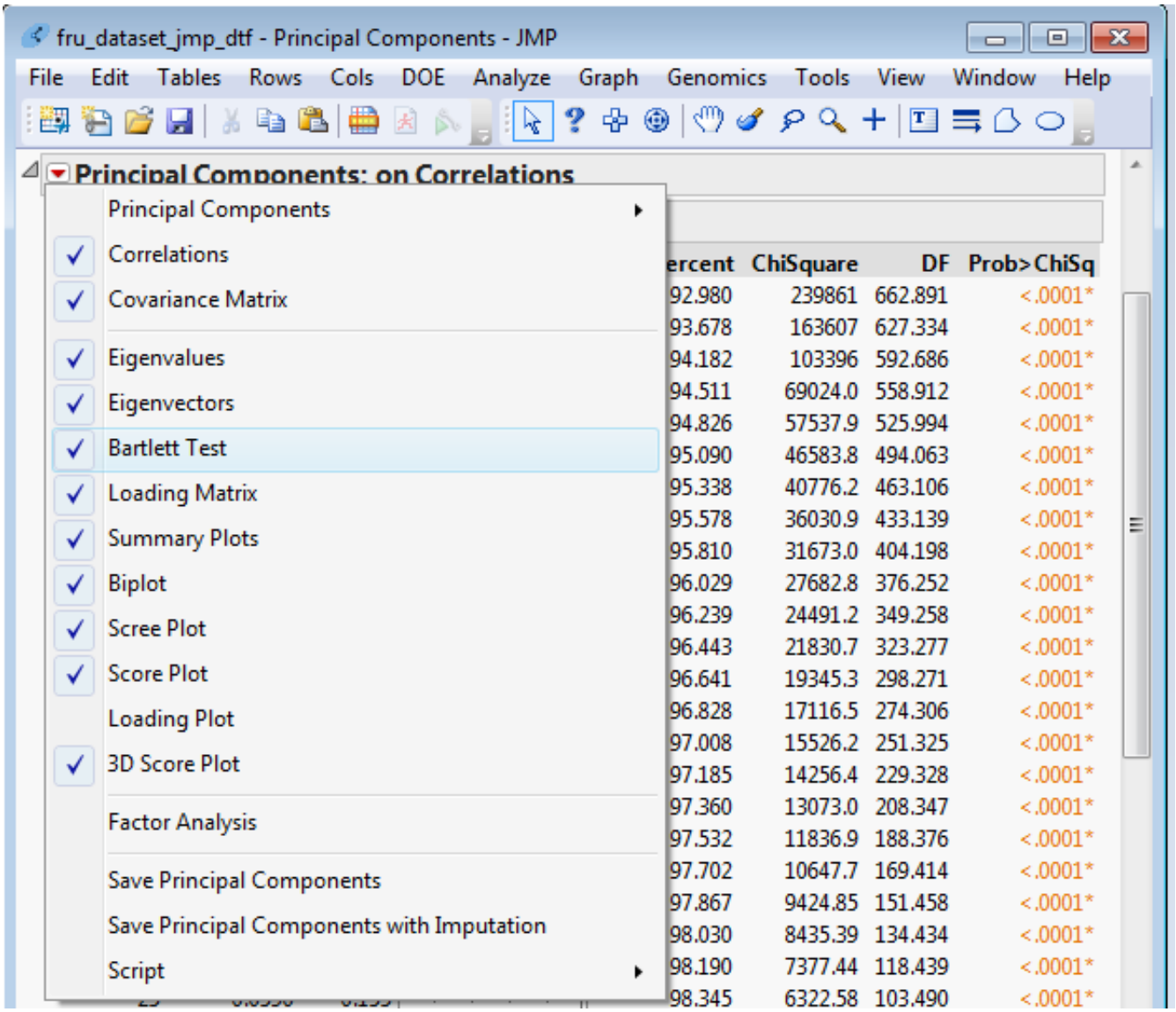
We can see here that Component 1 seems to have the greatest effect, and Component 2 has relatively small effect.

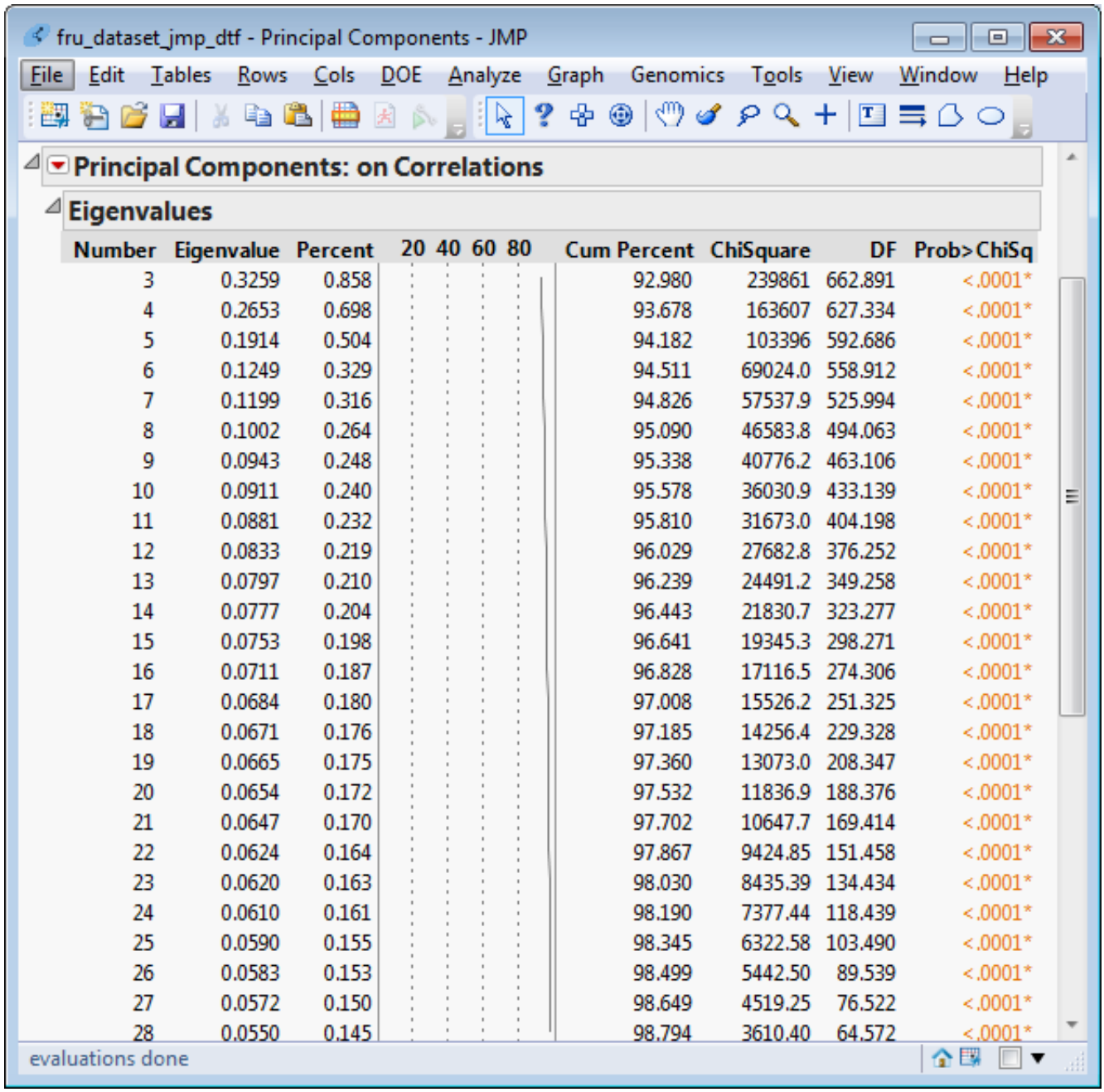
9.) Let's look further at the eigenvalues. Expand "Eigenvalues". We have 38 columns (samples) and 45339 rows (fusions). How many eigenvalues should there be? How many are there?



Let's look at the eigenvalues. How many do you think are necessary to describe the variation in the data?

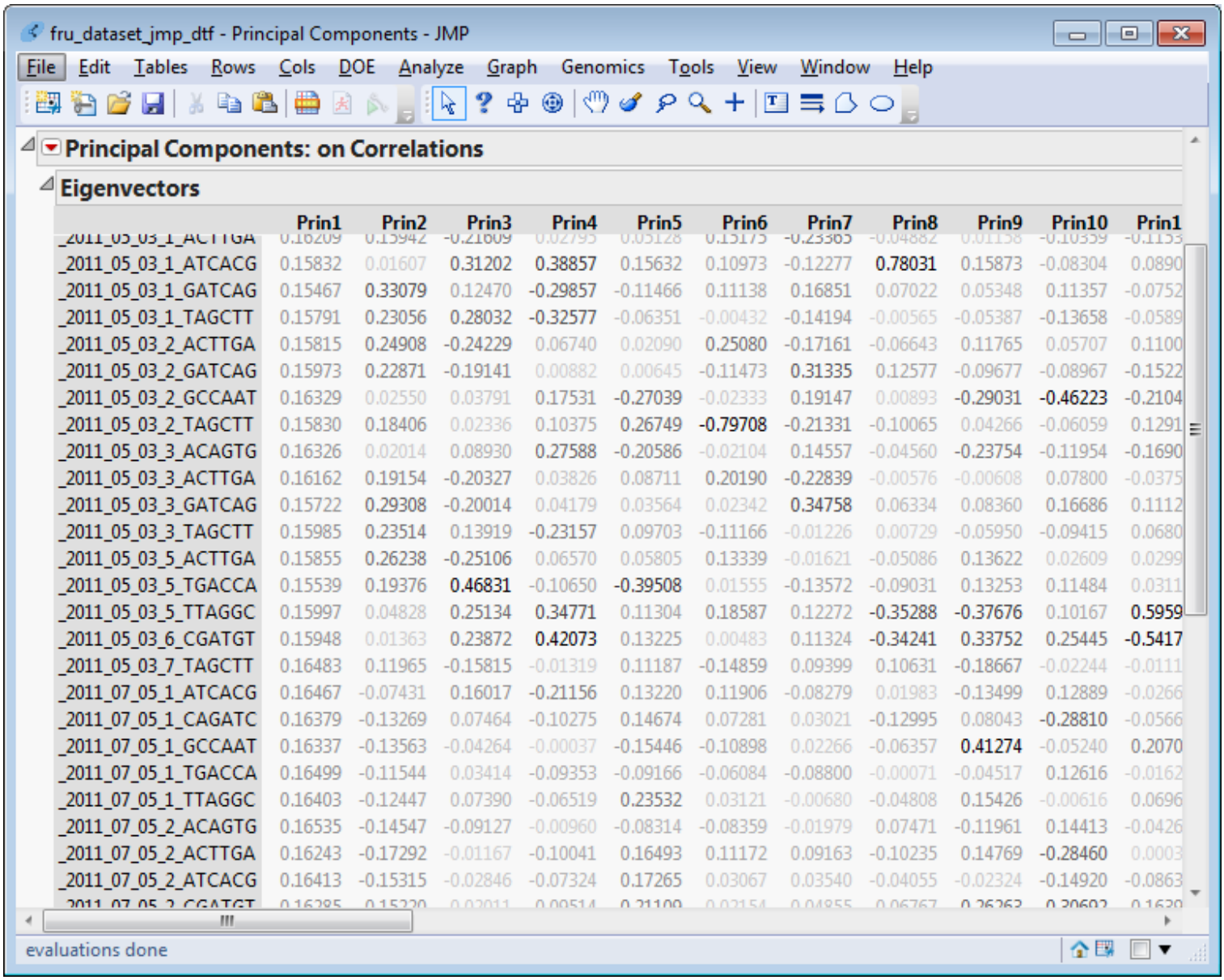
If we click on the red dropdown button next to "Principal Components" (top) and check "Bartlett Test", we can determine if the eigenvalues have the same variance by calculating the Chi-square, degrees of freedom and P value for the test. For the purposes of this tutorial, you won't need to worry about this too much.


fru_dataset_jmp_dtf - Principal Components - JMP
File Edit Tables Rows Cols DOE Analyze Graph Genomics Tools View Window Help
Principal Components: on Correlations
Principal Components
Correlations
Covariance Matrix
Eigenvalues
Eigenvectors
Bartlett Test
Loading Matrix
Summary Plots
Biplot
Scree Plot
Score Plot
Loading Plot
3D Score Plot
Factor Analysis
Save Principal Components
Save Principal Components with Imputation
Script
ercent ChiSquare DF Prob>ChiSq
92.980 239861 662.891 <.0001*
93.678 163607 627.334 <.0001*
94.182 103396 592.686 <.0001*
94.511 69024.0 558.912 <.0001*
94.826 57537.9 525.994 <.0001*
95.090 46583.8 494.063 <.0001*
95.338 40776.2 463.106 <.0001*
95.578 36030.9 433.139 <.0001*
95.810 31673.0 404.198 <.0001*
96.029 27682.8 376.252 <.0001*
96.239 24491.2 349.258 <.0001*
96.443 21830.7 323.277 <.0001*
96.641 19345.3 298.271 <.0001*
96.828 17116.5 274.306 <.0001*
97.008 15526.2 251.325 <.0001*
97.185 14256.4 229.328 <.0001*
97.360 13073.0 208.347 <.0001*
97.532 11836.9 188.376 <.0001*
97.702 10647.7 169.414 <.0001*
97.867 9424.85 151.458 <.0001*
98.030 8435.39 134.434 <.0001*
98.190 7377.44 118.439 <.0001*
98.345 6322.58 103.490 <.0001*

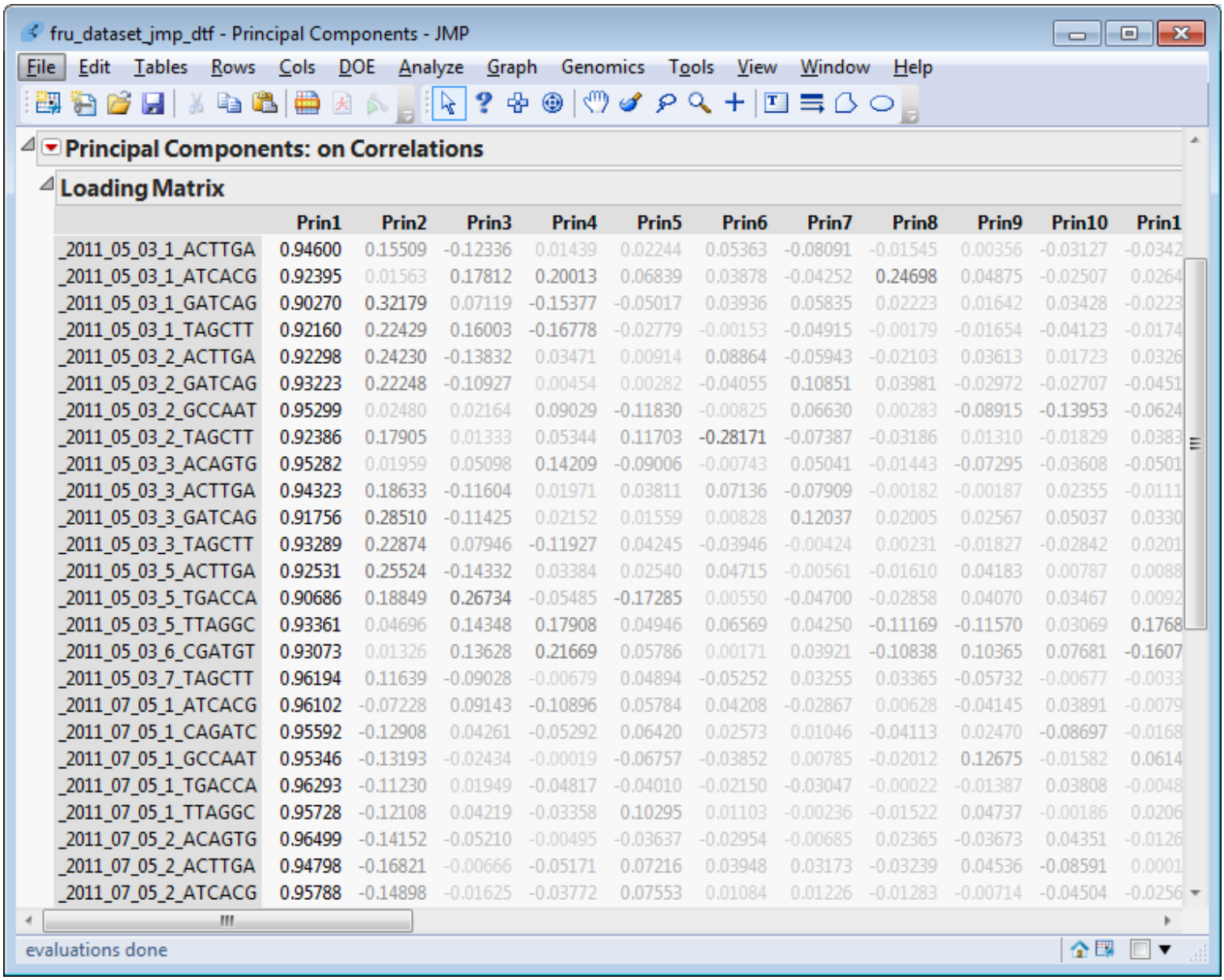


Number	Eigenvalue	Percent	Cum Percent	ChiSquare	DF	Prob>ChiSq
3	0.3259	0.858	92.980	239861	662.891	<.0001*
4	0.2653	0.698	93.678	163607	627.334	<.0001*
5	0.1914	0.504	94.182	103396	592.686	<.0001*
6	0.1249	0.329	94.511	69024.0	558.912	<.0001*
7	0.1199	0.316	94.826	57537.9	525.994	<.0001*
8	0.1002	0.264	95.090	46583.8	494.063	<.0001*
9	0.0943	0.248	95.338	40776.2	463.106	<.0001*
10	0.0911	0.240	95.578	36030.9	433.139	<.0001*
11	0.0881	0.232	95.810	31673.0	404.198	<.0001*
12	0.0833	0.219	96.029	27682.8	376.252	<.0001*
13	0.0797	0.210	96.239	24491.2	349.258	<.0001*
14	0.0777	0.204	96.443	21830.7	323.277	<.0001*
15	0.0753	0.198	96.641	19345.3	298.271	<.0001*
16	0.0711	0.187	96.828	17116.5	274.306	<.0001*
17	0.0684	0.180	97.008	15526.2	251.325	<.0001*
18	0.0671	0.176	97.185	14256.4	229.328	<.0001*
19	0.0665	0.175	97.360	13073.0	208.347	<.0001*
20	0.0654	0.172	97.532	11836.9	188.376	<.0001*
21	0.0647	0.170	97.702	10647.7	169.414	<.0001*
22	0.0624	0.164	97.867	9424.85	151.458	<.0001*
23	0.0620	0.163	98.030	8435.39	134.434	<.0001*
24	0.0610	0.161	98.190	7377.44	118.439	<.0001*
25	0.0590	0.155	98.345	6322.58	103.490	<.0001*
26	0.0583	0.153	98.499	5442.50	89.539	<.0001*
27	0.0572	0.150	98.649	4519.25	76.522	<.0001*
28	0.0550	0.145	98.794	3610.40	64.572	<.0001*

10.) Now look at the eigenvectors. Expand "Eigenvectors". If you wanted to, you could use this data to write out and calculate PC1, PC2, PC3, etc. Note that the closer the shading of the value is to black, the closer it is to 1 or -1. The closer to light grey, the closer the value is to 0.

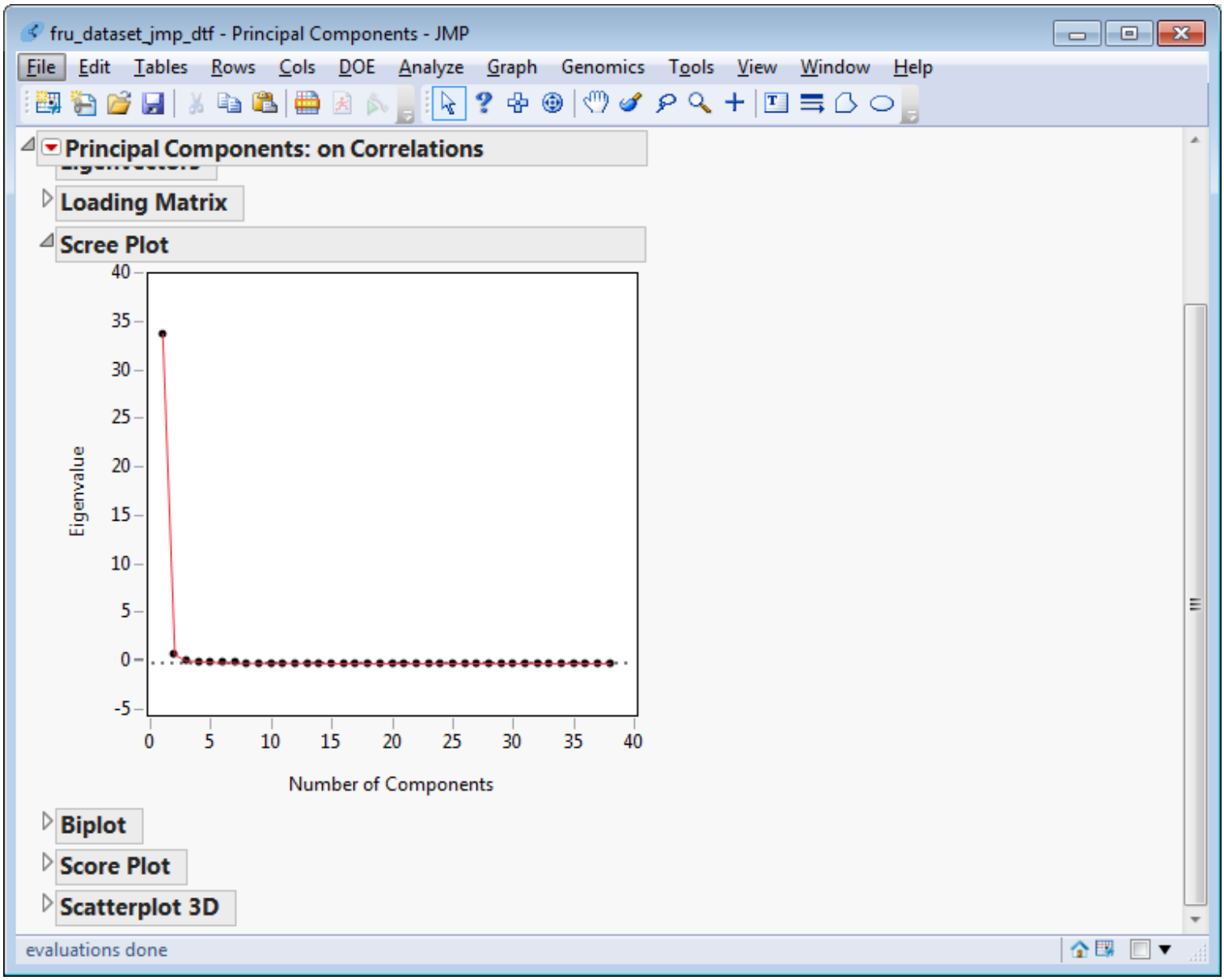


11.) Next is "Loading Matrix". This is the loading for each component for each sample and is plotted in the Loading Plot.

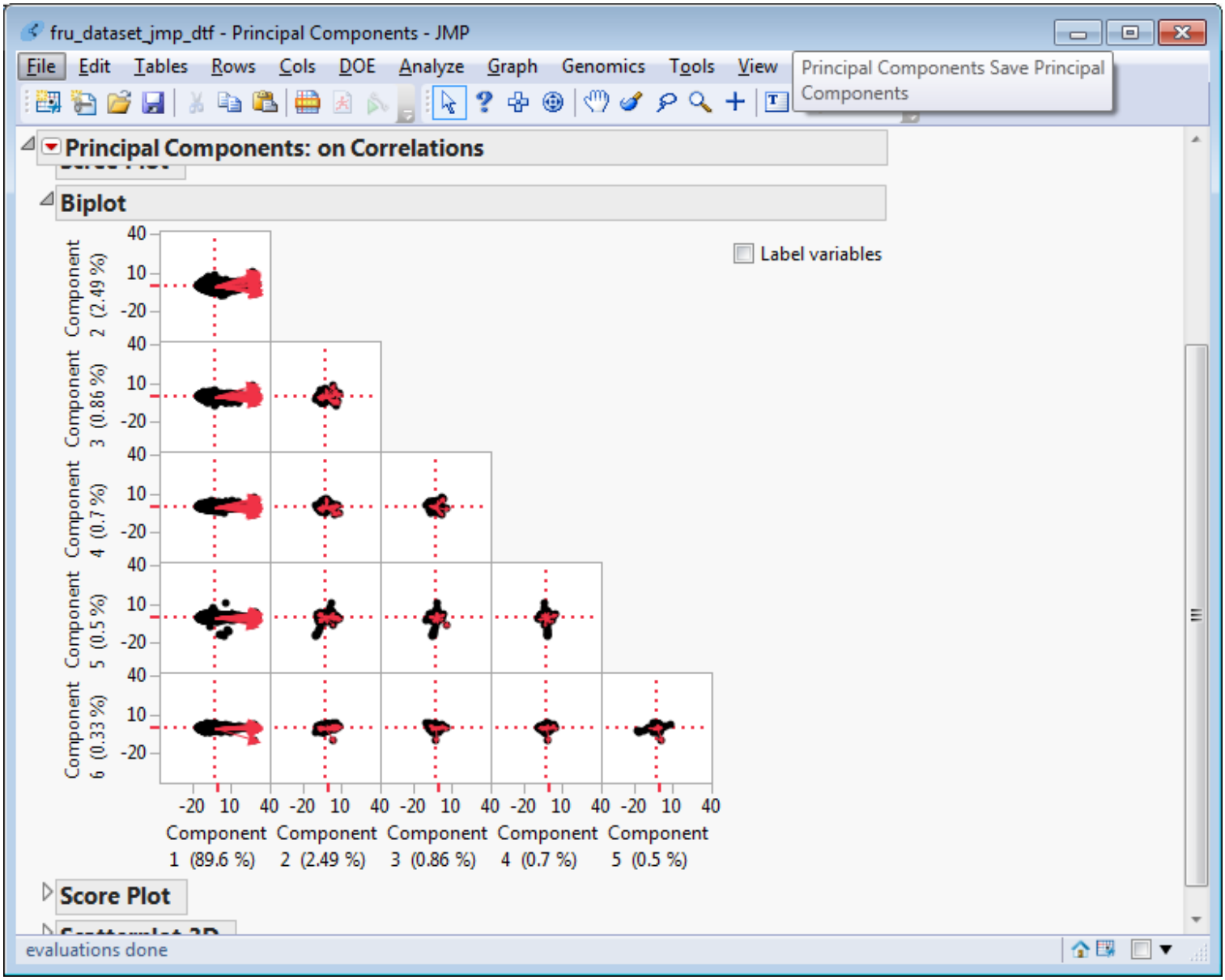


	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin1
_2011_05_03_1_ACTTGA	0.94600	0.15509	-0.12336	0.01439	0.02244	0.05363	-0.08091	-0.01545	0.00356	-0.03127	-0.0342
_2011_05_03_1_ATCACG	0.92395	0.01563	0.17812	0.20013	0.06839	0.03878	-0.04252	0.24698	0.04875	-0.02507	0.0264
_2011_05_03_1_GATCAG	0.90270	0.32179	0.07119	-0.15377	-0.05017	0.03936	0.05835	0.02223	0.01642	0.03428	-0.0223
_2011_05_03_1_TAGCTT	0.92160	0.22429	0.16003	-0.16778	-0.02779	-0.00153	-0.04915	-0.00179	-0.01654	-0.04123	-0.0174
_2011_05_03_2_ACTTGA	0.92298	0.24230	-0.13832	0.03471	0.00914	0.08864	-0.05943	-0.02103	0.03613	0.01723	0.0326
_2011_05_03_2_GATCAG	0.93223	0.22248	-0.10927	0.00454	0.00282	-0.04055	0.10851	0.03981	-0.02972	-0.02707	-0.0451
_2011_05_03_2_GCCAAT	0.95299	0.02480	0.02164	0.09029	-0.11830	-0.00825	0.06630	0.00283	-0.08915	-0.13953	-0.0624
_2011_05_03_2_TAGCTT	0.92386	0.17905	0.01333	0.05344	0.11703	-0.28171	-0.07387	-0.03186	0.01310	-0.01829	0.0383
_2011_05_03_3_ACAGTG	0.95282	0.01959	0.05098	0.14209	-0.09006	-0.00743	0.05041	-0.01443	-0.07295	-0.03608	-0.0501
_2011_05_03_3_ACTTGA	0.94323	0.18633	-0.11604	0.01971	0.03811	0.07136	-0.07909	-0.00182	-0.00187	0.02355	-0.0111
_2011_05_03_3_GATCAG	0.91756	0.28510	-0.11425	0.02152	0.01559	0.00828	0.12037	0.02005	0.02567	0.05037	0.0330
_2011_05_03_3_TAGCTT	0.93289	0.22874	0.07946	-0.11927	0.04245	-0.03946	-0.00424	0.00231	-0.01827	-0.02842	0.0201
_2011_05_03_5_ACTTGA	0.92531	0.25524	-0.14332	0.03384	0.02540	0.04715	-0.00561	-0.01610	0.04183	0.00787	0.0088
_2011_05_03_5_TGACCA	0.90686	0.18849	0.26734	-0.05485	-0.17285	0.00550	-0.04700	-0.02858	0.04070	0.03467	0.0092
_2011_05_03_5_TTAGGC	0.93361	0.04696	0.14348	0.17908	0.04946	0.06569	0.04250	-0.11169	-0.11570	0.03069	0.1768
_2011_05_03_6_CGATGT	0.93073	0.01326	0.13628	0.21669	0.05786	0.00171	0.03921	-0.10838	0.10365	0.07681	-0.1607
_2011_05_03_7_TAGCTT	0.96194	0.11639	-0.09028	-0.00679	0.04894	-0.05252	0.03255	0.03365	-0.05732	-0.00677	-0.0033
_2011_07_05_1_ATCACG	0.96102	-0.07228	0.09143	-0.10896	0.05784	0.04208	-0.02867	0.00628	-0.04145	0.03891	-0.0079
_2011_07_05_1_CAGATC	0.95592	-0.12908	0.04261	-0.05292	0.06420	0.02573	0.01046	-0.04113	0.02470	-0.08697	-0.0168
_2011_07_05_1_GCCAAT	0.95346	-0.13193	-0.02434	-0.00019	-0.06757	-0.03852	0.00785	-0.02012	0.12675	-0.01582	0.0614
_2011_07_05_1_TGACCA	0.96293	-0.11230	0.01949	-0.04817	-0.04010	-0.02150	-0.03047	-0.00022	-0.01387	0.03808	-0.0048
_2011_07_05_1_TTAGGC	0.95728	-0.12108	0.04219	-0.03358	0.10295	0.01103	-0.00236	-0.01522	0.04737	-0.00186	0.0206
_2011_07_05_2_ACAGTG	0.96499	-0.14152	-0.05210	-0.00495	-0.03637	-0.02954	-0.00685	0.02365	-0.03673	0.04351	-0.0126
_2011_07_05_2_ACTTGA	0.94798	-0.16821	-0.00666	-0.05171	0.07216	0.03948	0.03173	-0.03239	0.04536	-0.08591	0.0001
_2011_07_05_2_ATCACG	0.95788	-0.14898	-0.01625	-0.03772	0.07553	0.01084	0.01226	-0.01283	-0.00714	-0.04504	-0.0256

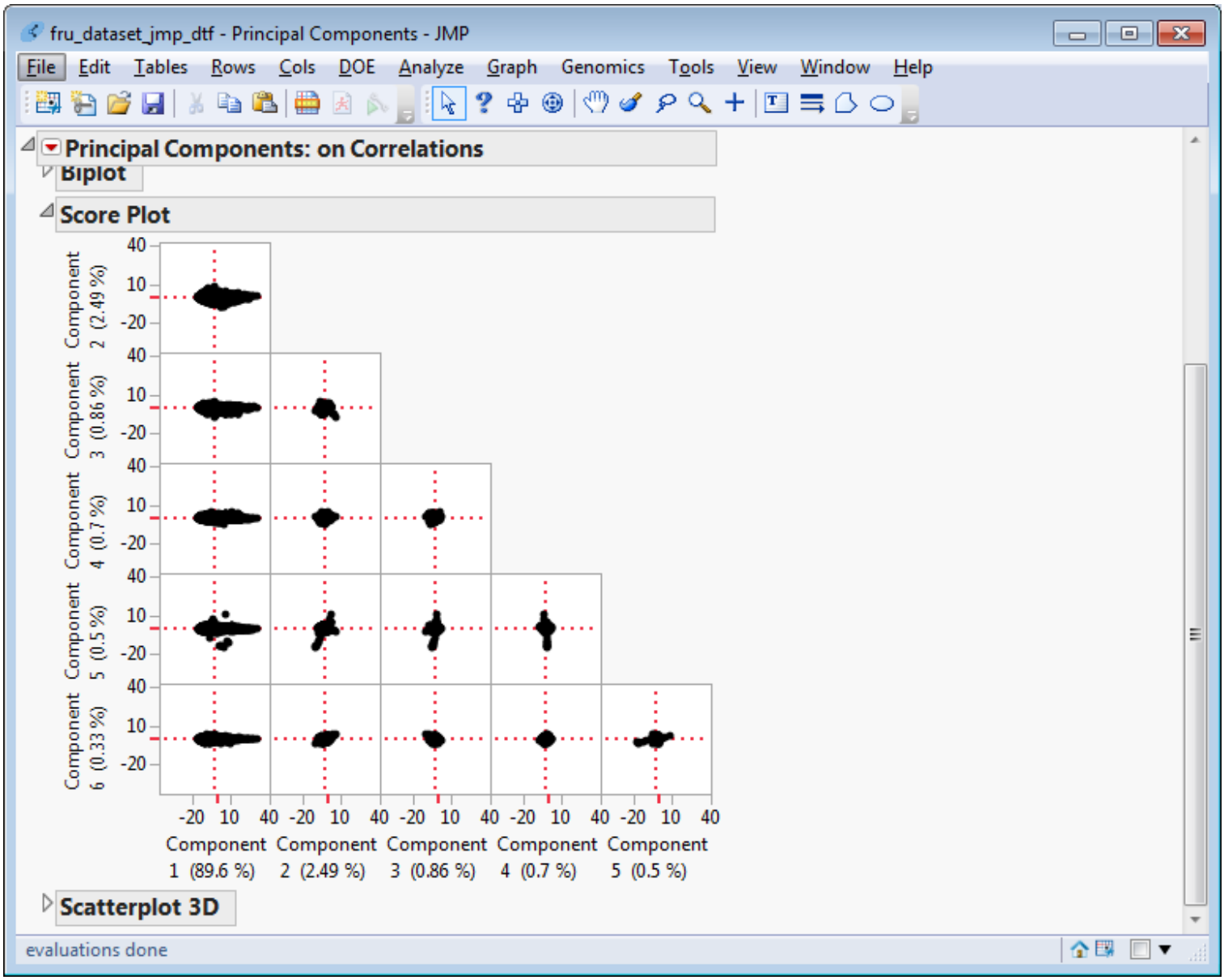
12.) Next is the "Scree Plot", which graphs the eigenvalue for each component. This can be useful in visualising the dimensionality of the data space.



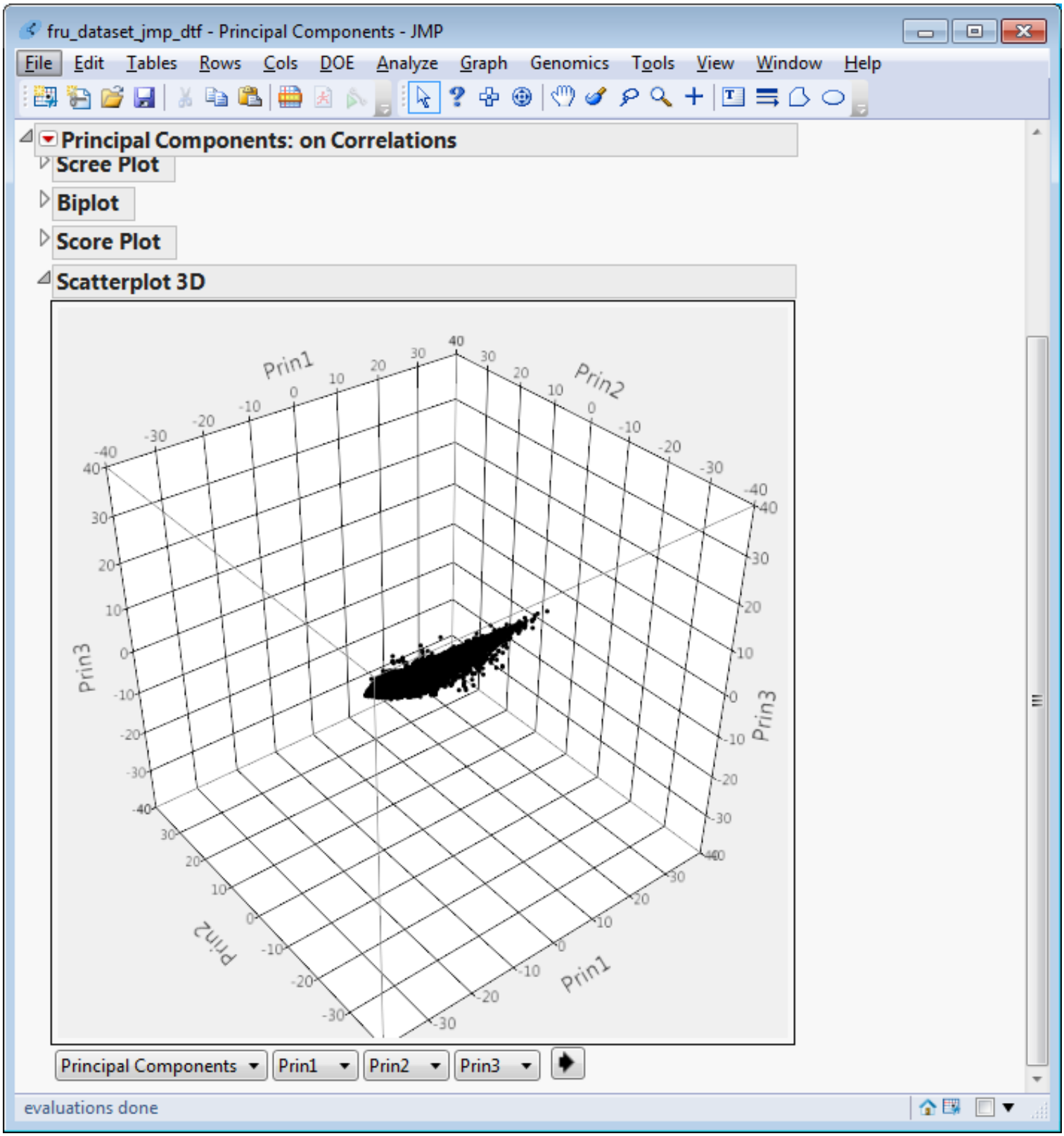
13.) Next is the "Biplot". This overlays the Score Plots and the Loading Plot for the specified number of components. Here we have set this as 6, so this is comparings PC1 to PC6. These plots tell you that PC1 is accounting for almost all the variance.



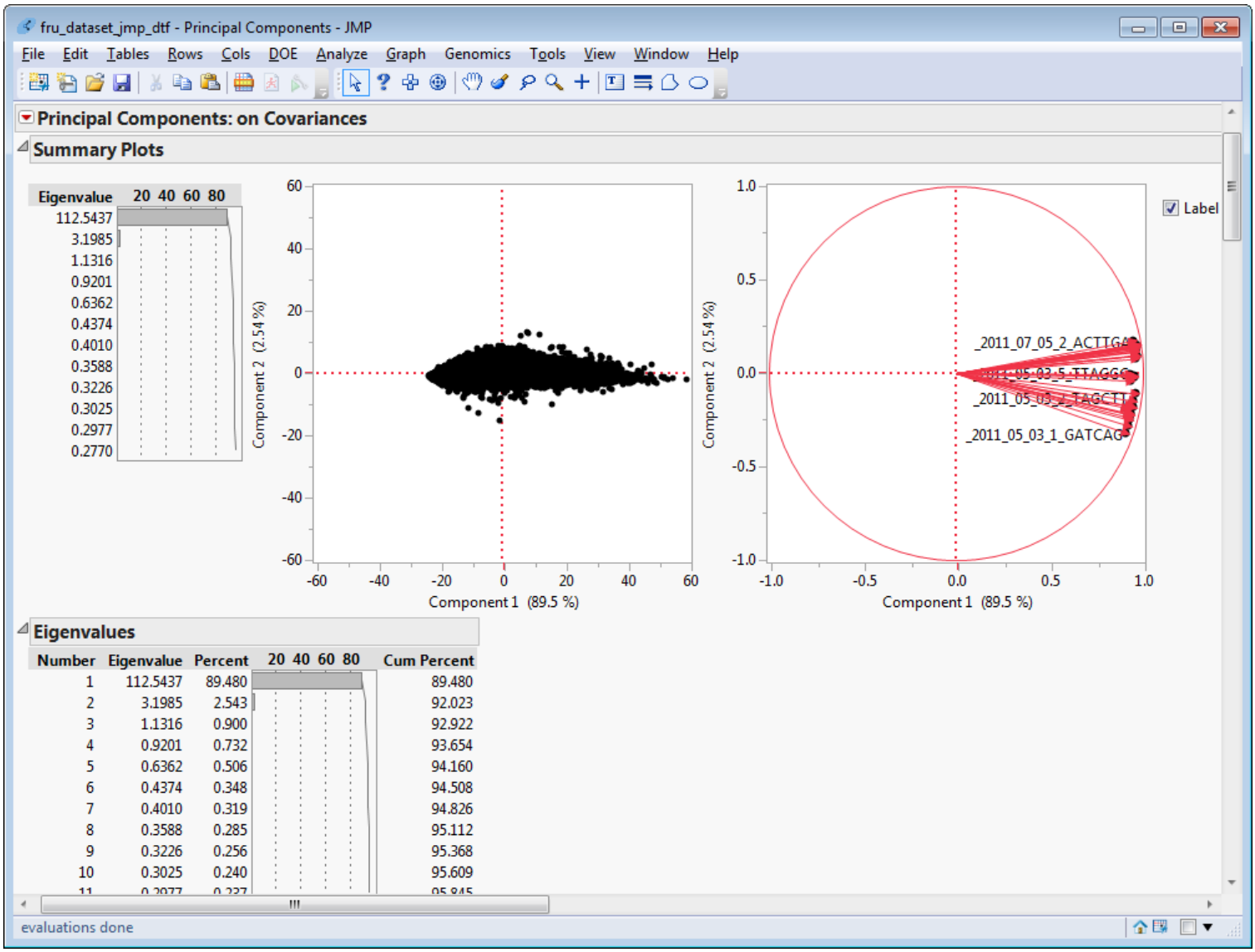
14.) Next is "Score Plot", which shows a matrix of scatterplots of the scores for each pair of principal components. Again, PC1 accounts for most of the variance. You can also see a "skew" to the data. What might this indicate?



15.) Lastly, "Scatterplot 3D". Here you can select up to three principal components and plot them. This is an interactive graph and can be rotated. You can use the arrow button at the bottom to cycle through each PC combination.

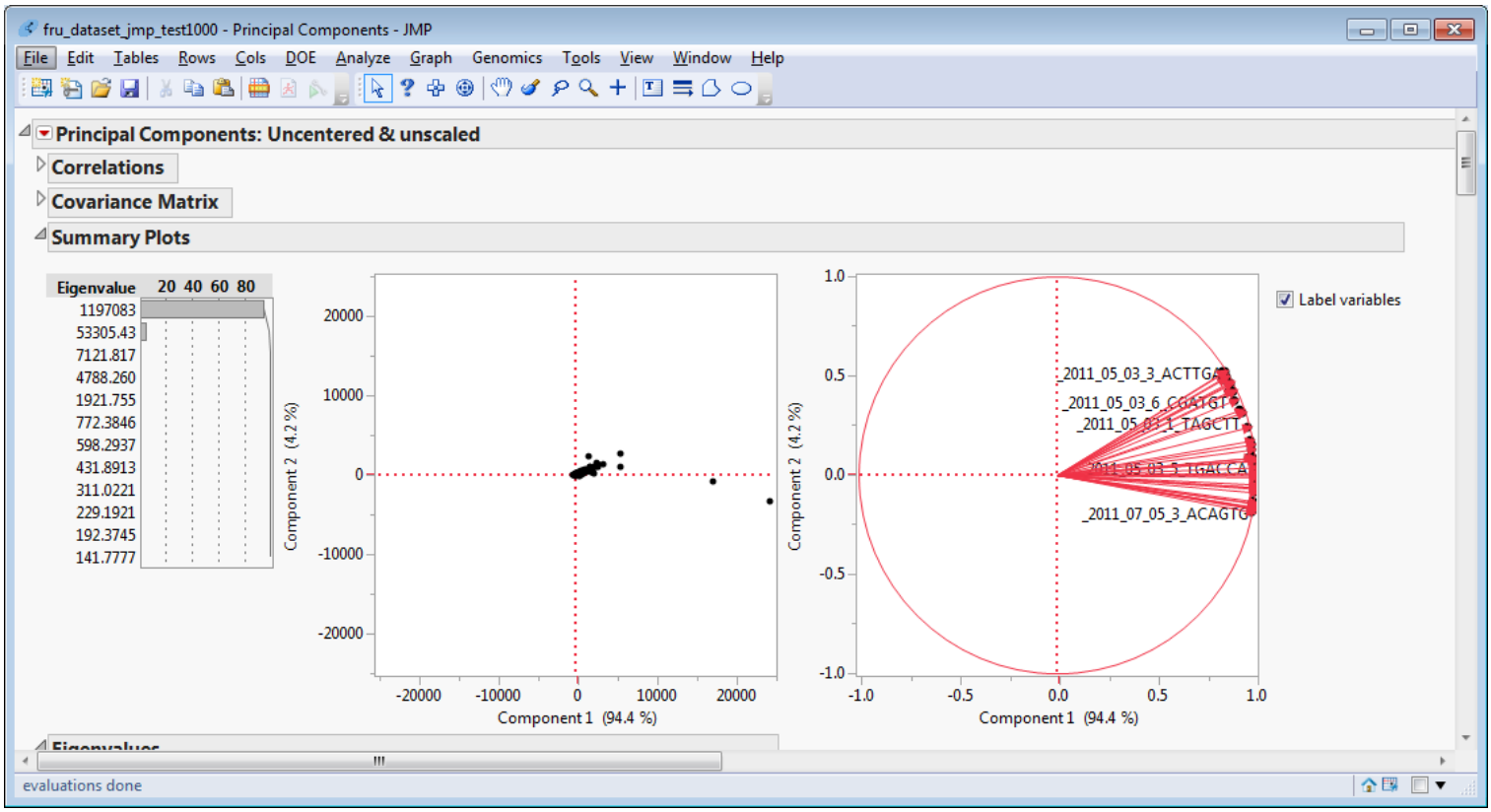


16.) Now let's change how the eigenvalues are calculate by clicking the red dropdown button next to "Principal Components" (top), select "Principal Components" then "on Covariances". How does this affect the results?



17.) Now let's repeat this with the untransformed data to compare. Go back to Analyze > Multivariate Methods > Principal Components, and select **fru_dataset_jmp** and repeat the previous steps.

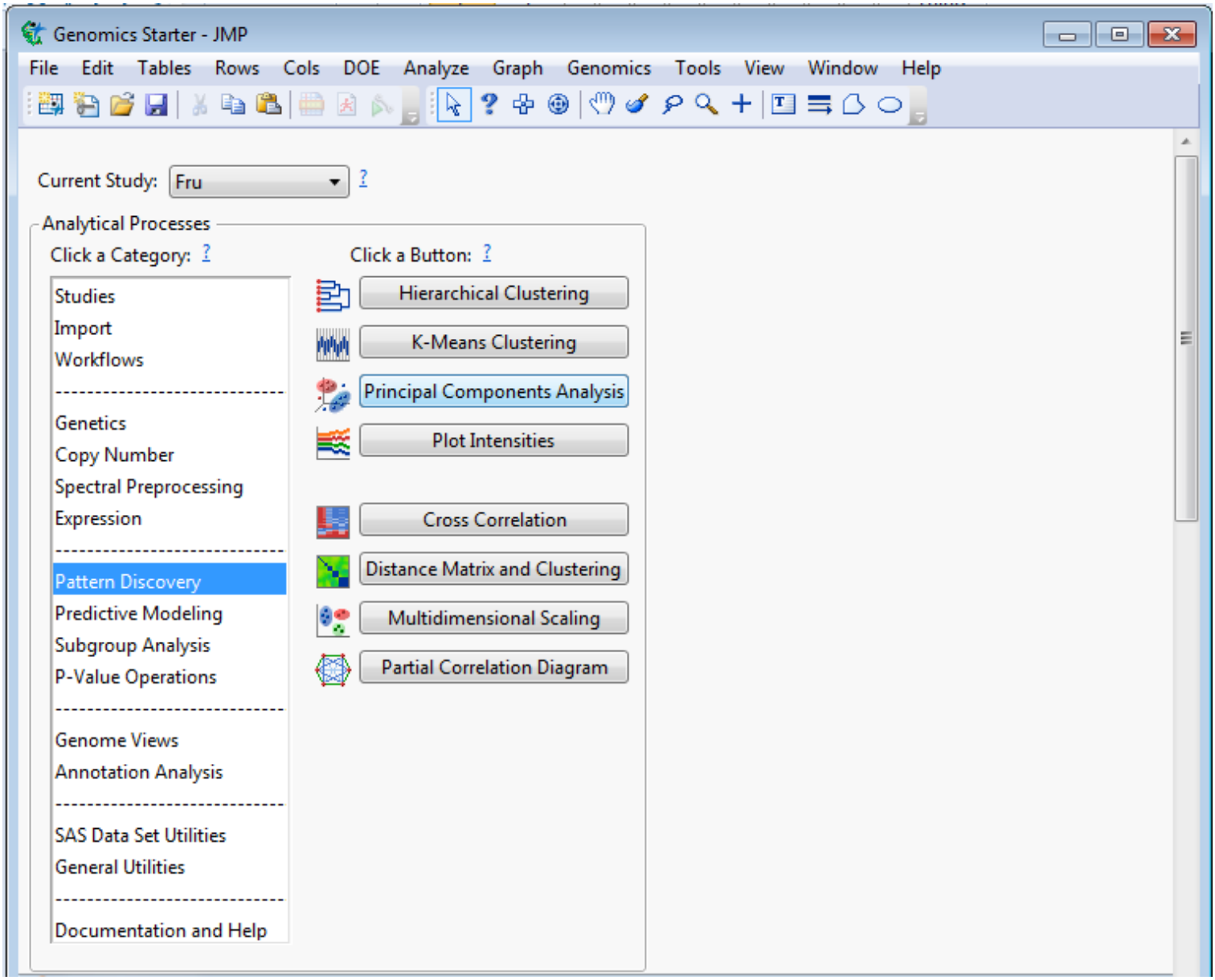
18.) Let's look at the outputs. For the purpose of this tutorial, below are the eigenvalues and summary plots to show you how transformation can change the results.



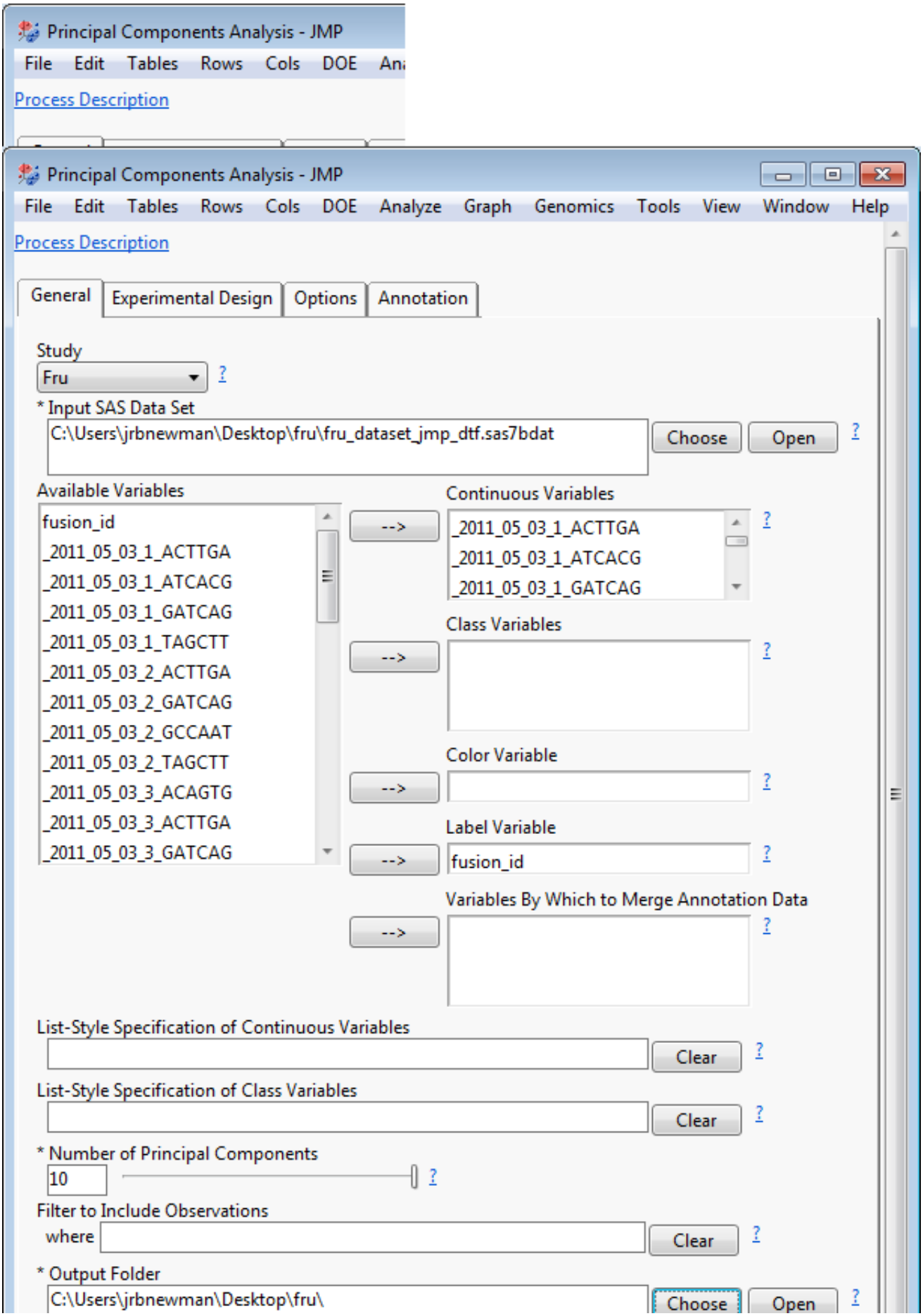
For comparison, look at the eigenvalues here and compare them to the transformed results. If you want, you can also do this with **fru_dataset_jmp_1000** and **fru_dataset_allon** to look at how the results change when the data is subsetted, either as a test set of 1000 observations, or all non-zero observations.

Method 2 - via the Genomics menu

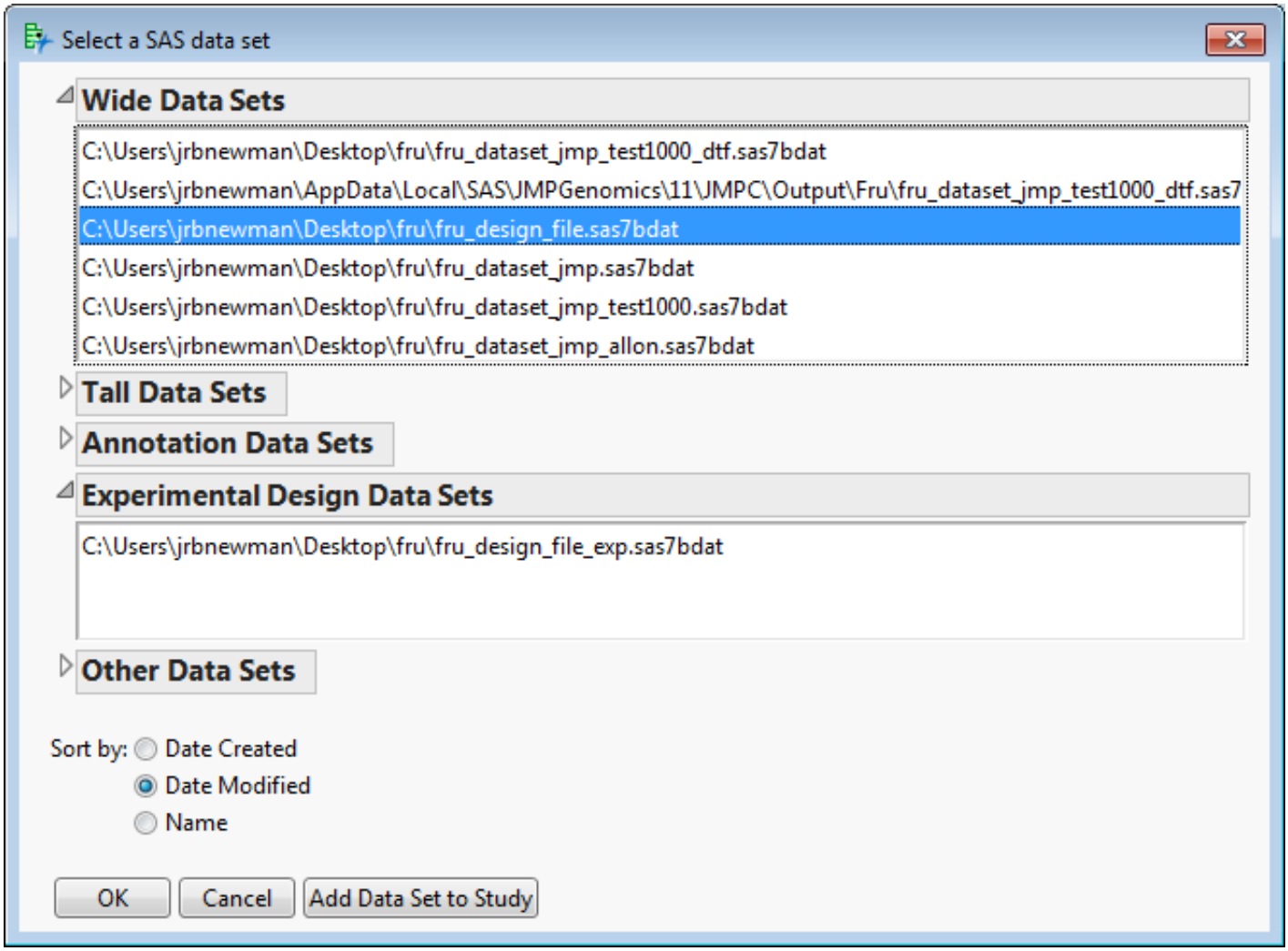
1.) On the Genomics Starter window, click on the "Pattern Discovery" category, then click "Principal Components Analysis".



2.) In the dialog box on the "General" tab, under "Input SAS Data Set" select **fru_dataset_jmp_dtf.sas7bdat**.
3.) Under "Available Variables" select everything beginning "*2011*" and click on the "-->" button next to "Continuous Variables".
4.) Under "Available Variables" select "fusion_id" and click on the "-->" button next to "Label Variable".
5.) In the "List-Style Specification of Continuous Variables" type (without quotes) "*2011*:". This allows you to specify the variables (columns) that are numbered consecutively or have a common prefix. This is an alternative way to add variables, especially if you have lots you want to include.
6.) Under "Number of Principal Components", specify the number of components you wish to calculate. For this, put in 10.
7.) Under "Output Folder" select your output folder. The dialog box should look like this:



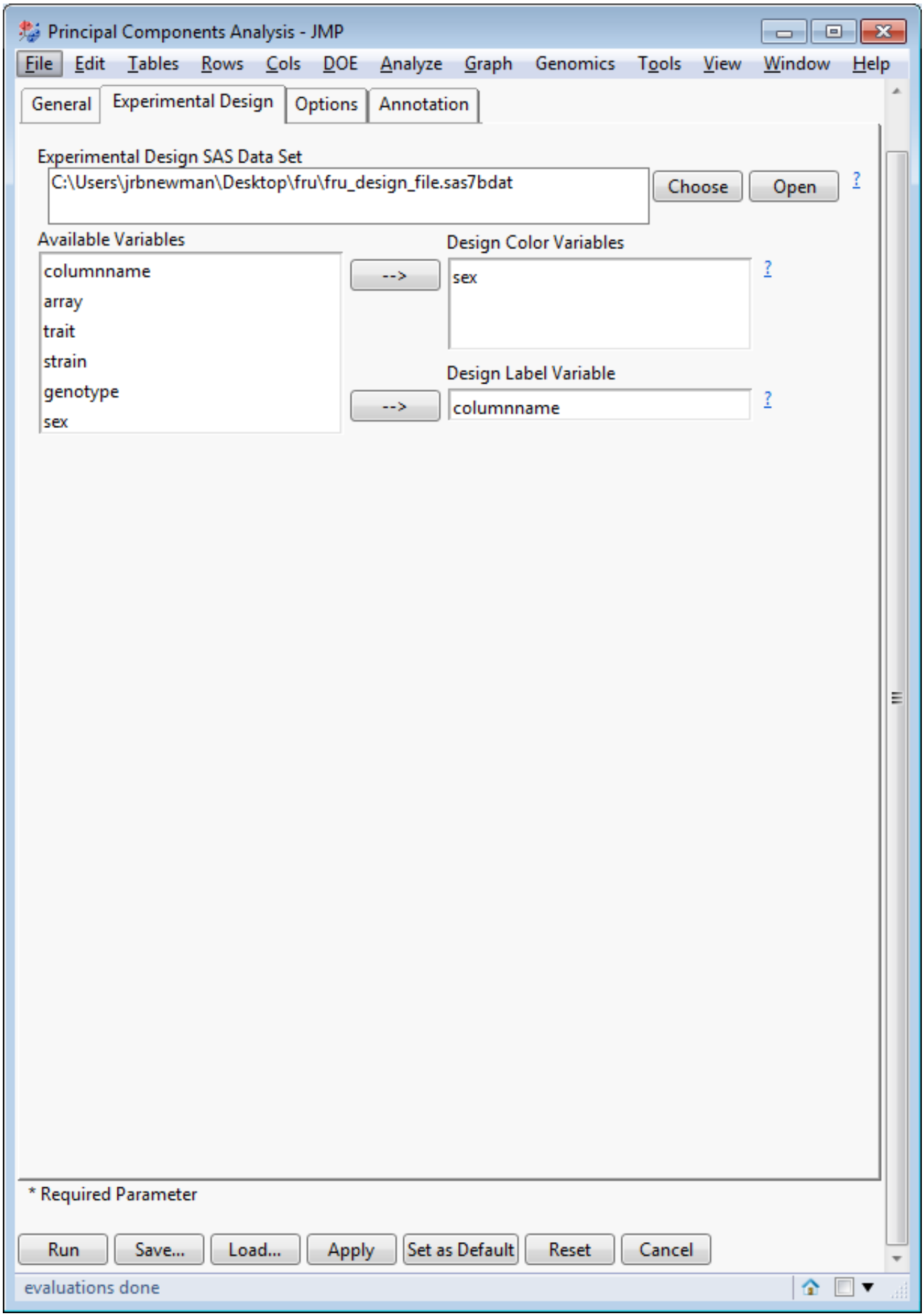
8.) Select the "Experimental Design" tab. Under "Experimental Design SAS Data Set" click "Choose". 9.) Expand "Wide Data Sets" and select **fru_design_file.sas7bdat**. Click "OK".



10.) Under "Available Variables", select "sex" and click the "-->" next to "Design Color Variables". This will color the data on the PCA plots by their trait. You can also select "strain", "genotype, or "trait".

11.) Under "Available Variables", select "columnname" and click the "-->" next to "Design Label Variable". This will match the variables here to the expression data.

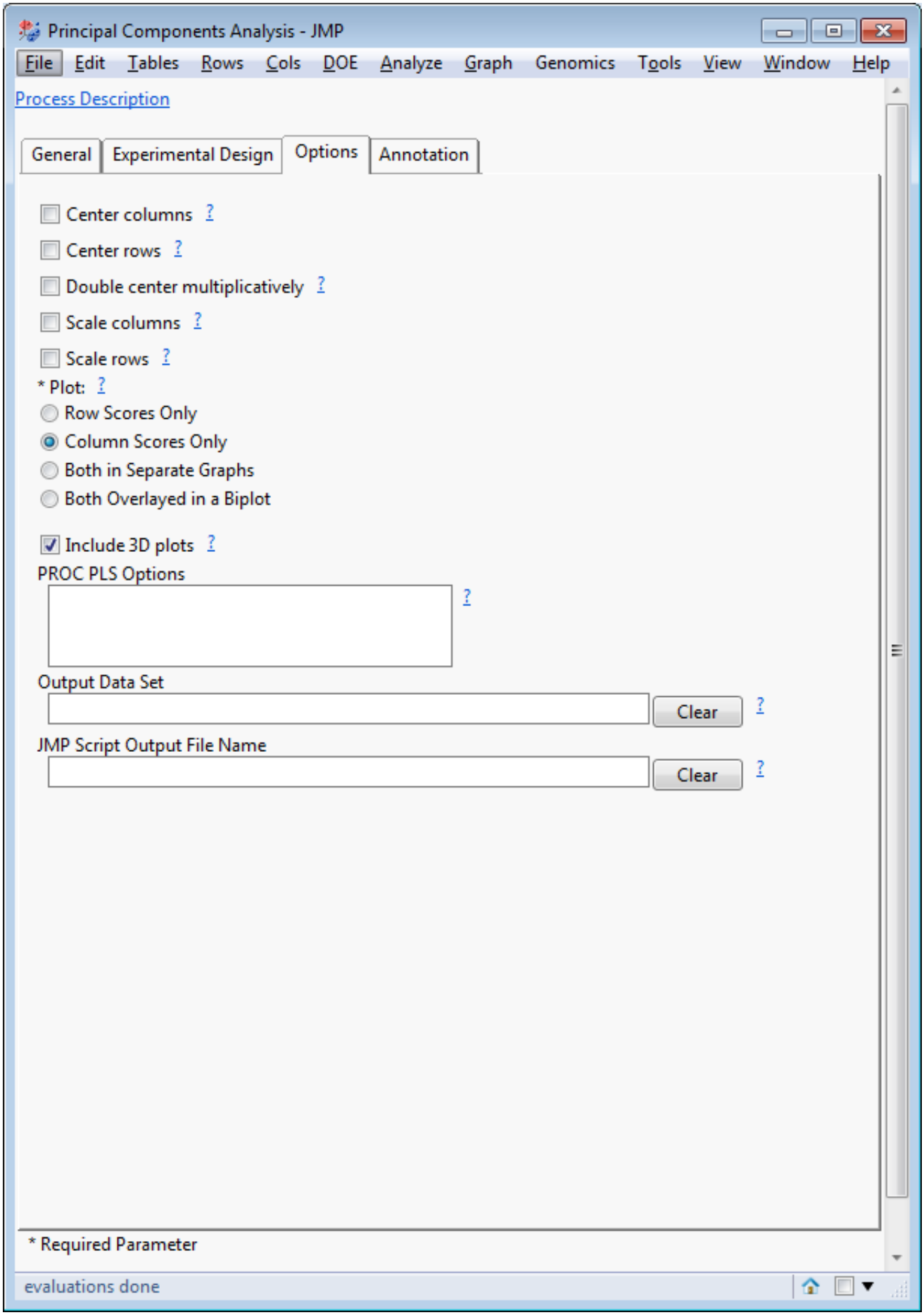
12.) The dialog box should look like this:



13.) Select the "Options" tab. For now, uncheck the five checkboxes.

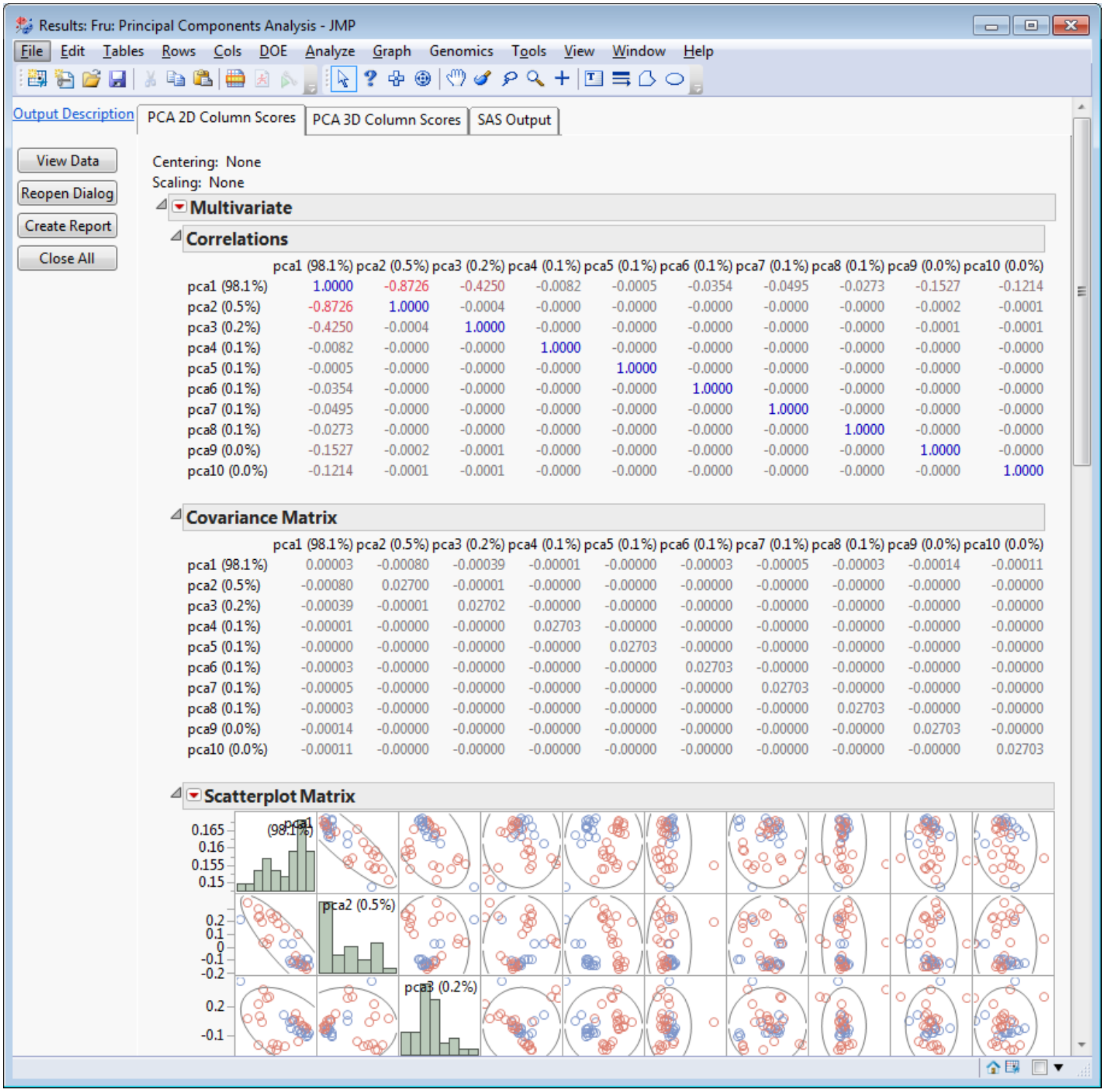
14.) The "Plot" radio selections allow you to select whether you want to perform PCA on rows (fusions), columns (samples), or both. For now, we are only interested in looking at the samples, so select "Column Scores Only".

15.) Make sure "Include 3D plots" is checked. The dialog box should look like the following. Click "Run".

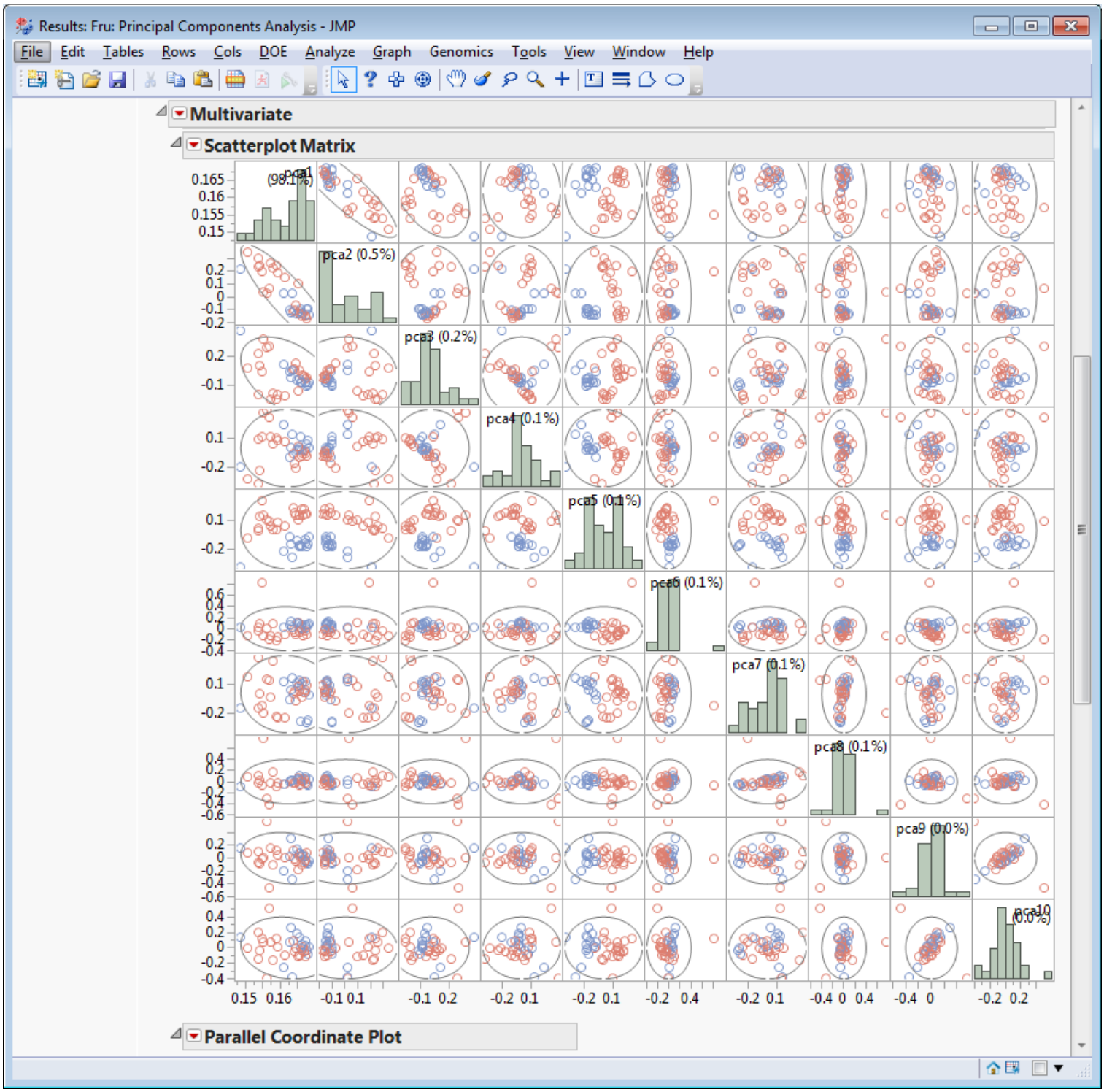


16.) The output window should appear. There are three tabs: "PCA 2D Column Scores", "PCA 3D Column Scores", and "SAS Output".

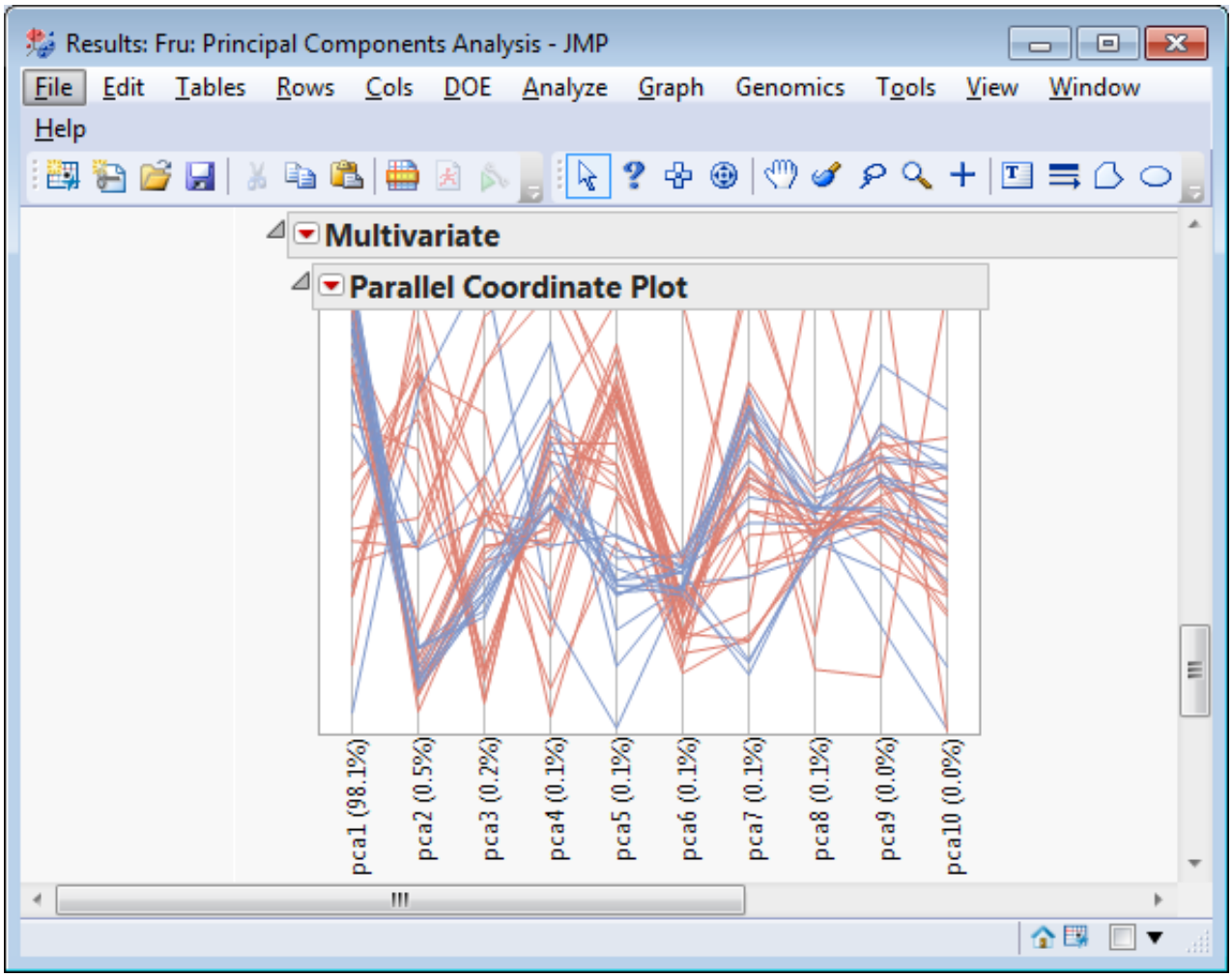
17.) Let's start with the "PCA 2D Column Scores" tab. First you should see the correlation matrix. Unlike the previous method, this is the correlation matrix between principal components. You can also view the covariance matrix between principal components by click the red dropdown button and select "Covariance Matrix"



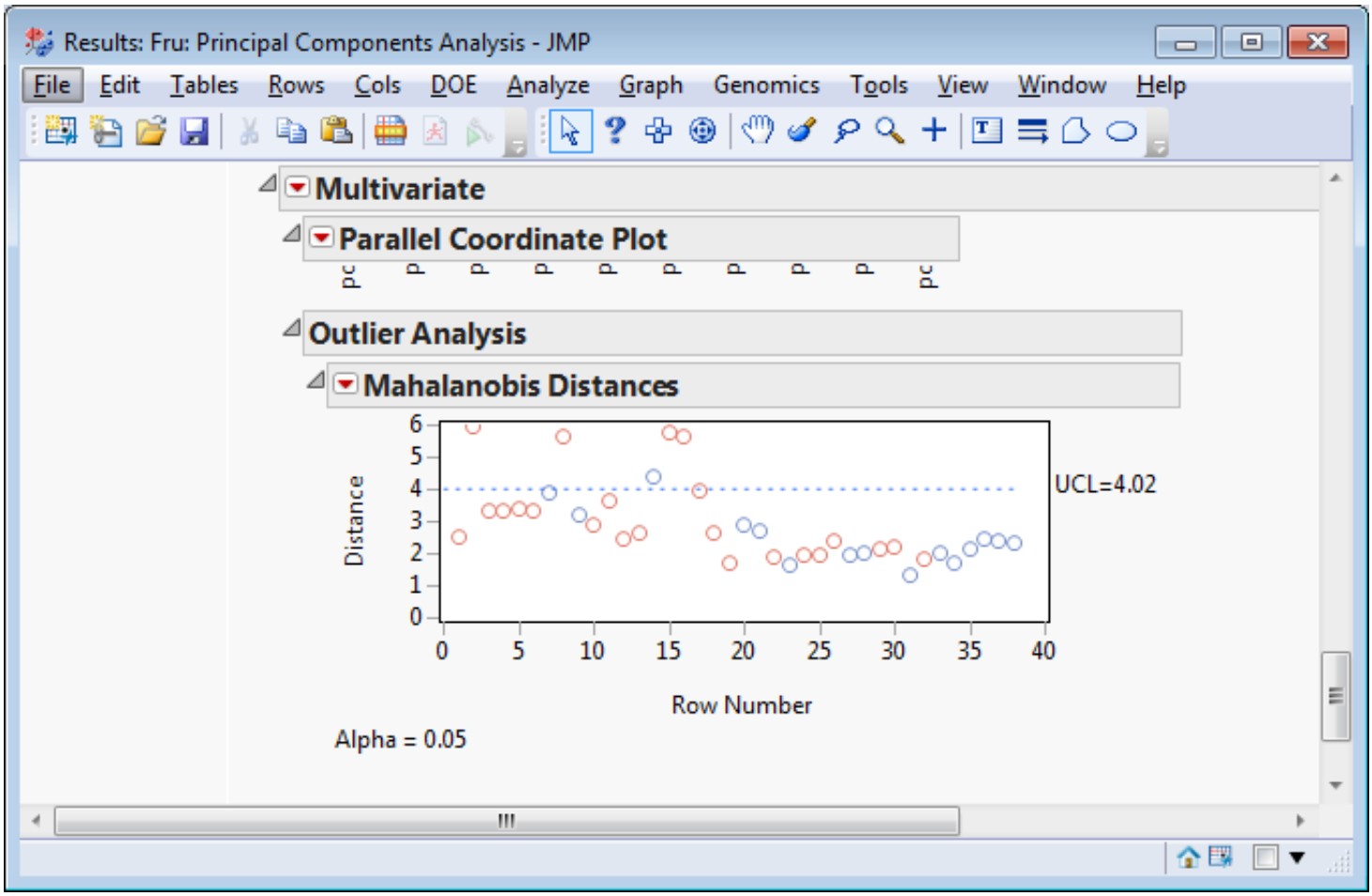
18.) Next is the "Scatterplot Matrix", which is a matrix of the 2D scatterplots between each pair of principal components.



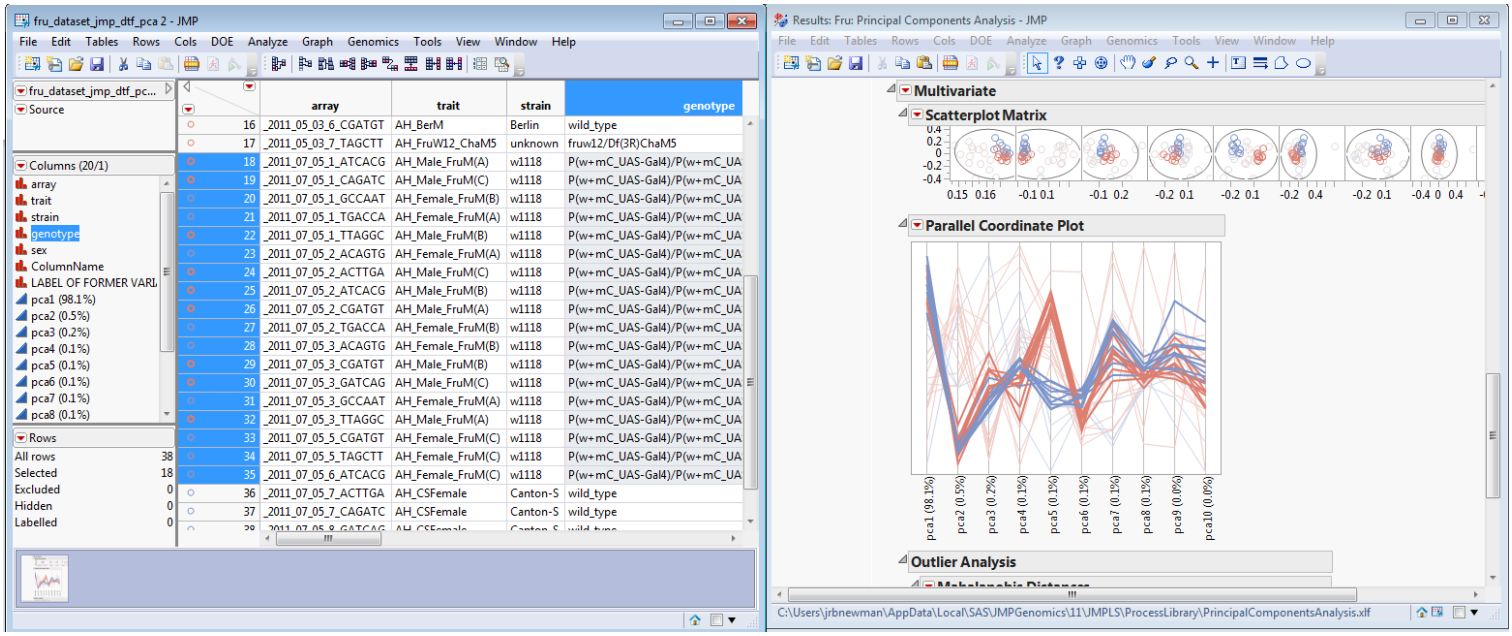
19.) Next is the "Parallel Coordinate Plot", which is a way to visualize high-dimensional data. It's a line graph connecting the calculated principal components for each sample. You can use this to identify possible relationships in the data, but for this tutorial we won't be worrying about it.



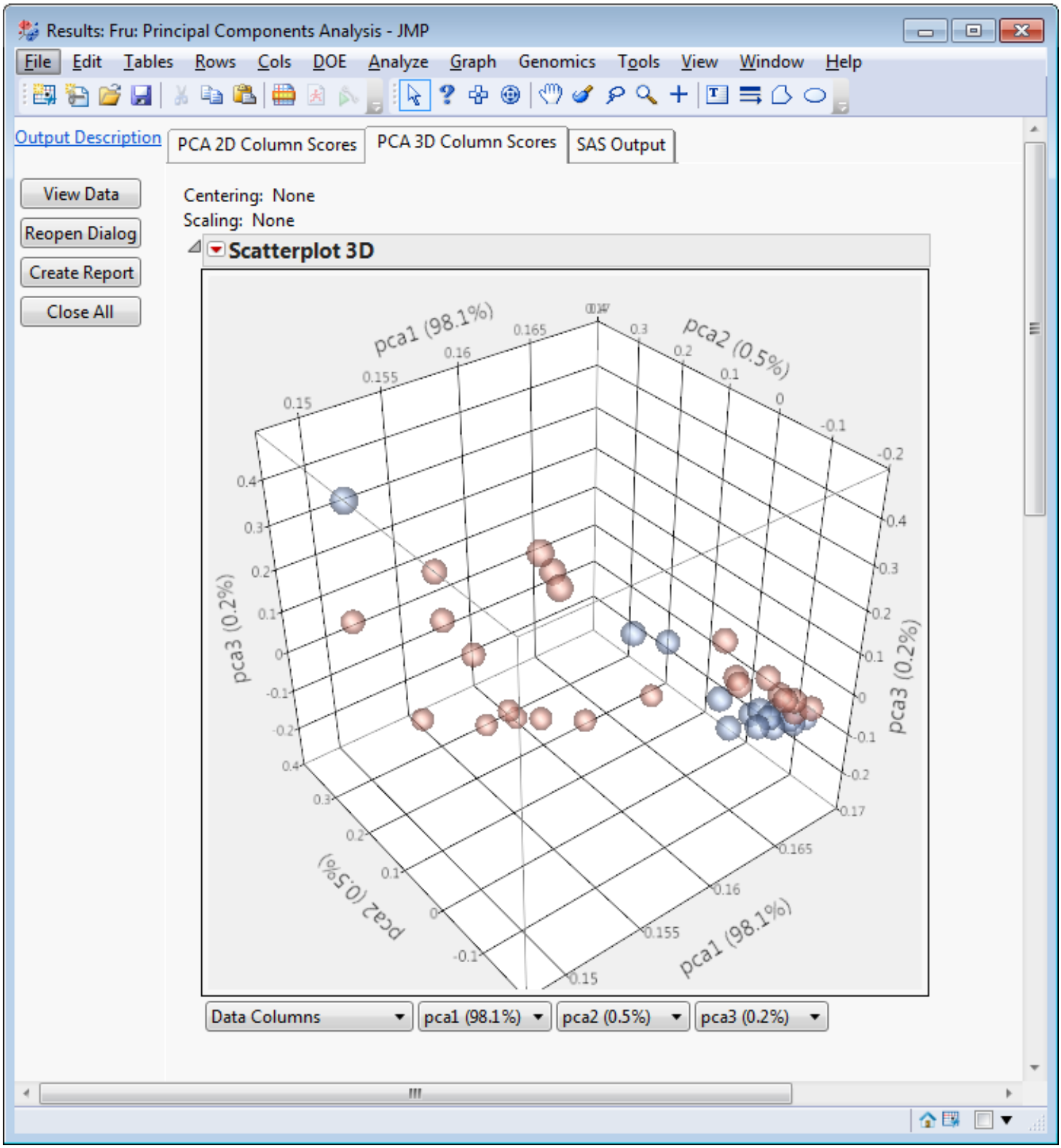
20.) Last on this tab is "Outlier Analysis". By default, this will be the calculated Mahalanobis distances for each sample. You can use this identify possible outlying samples in your data.



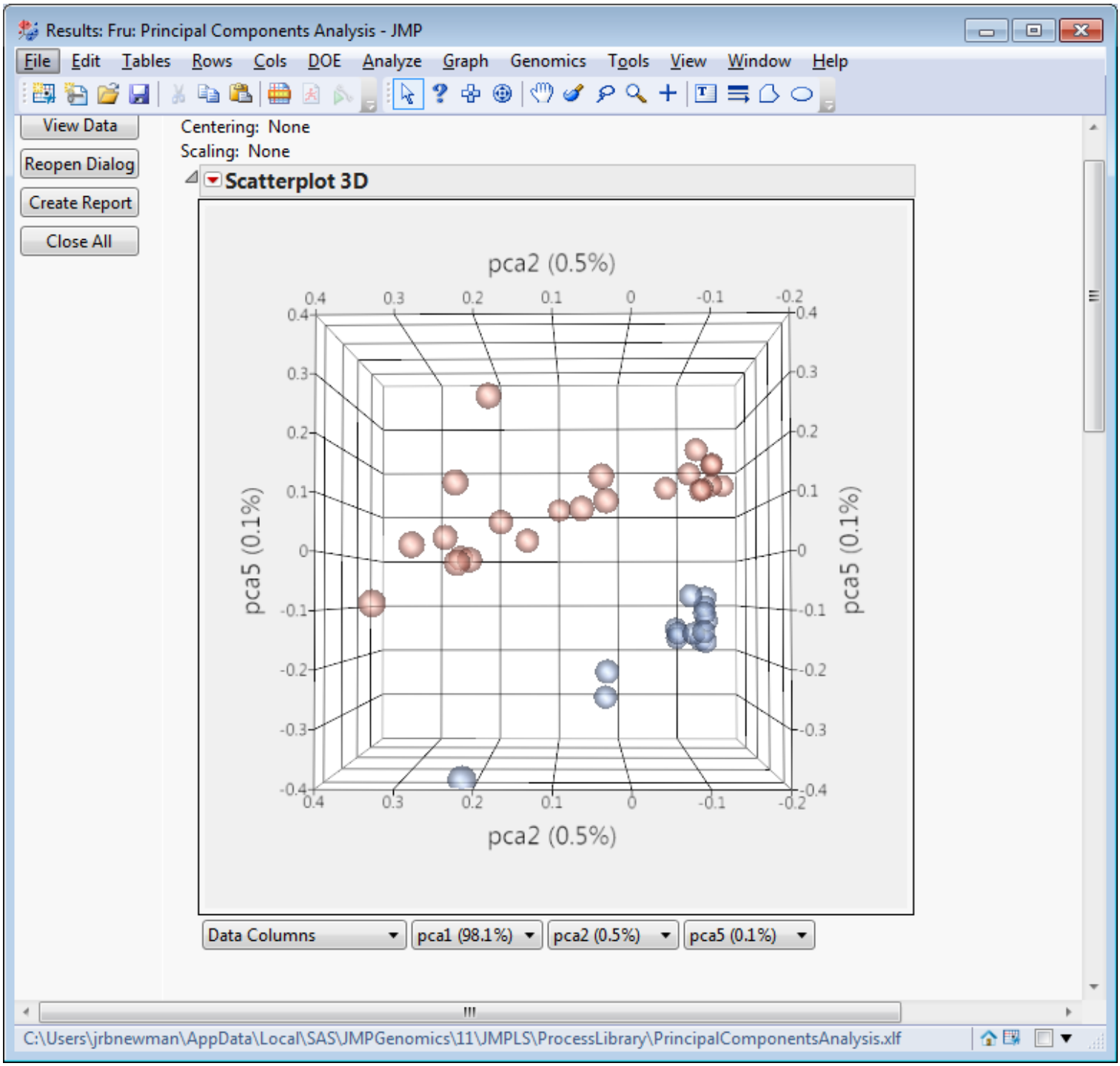
21.) You can select samples on any of these plots to highlight them. For example, if we select the samples with a Mahalanobis distance greater than 4 (above the dotted line), we can see the samples being highlighted in the other plots. You can also select samples from the data directly by clicking on the "View Data" button on the left and selecting samples.



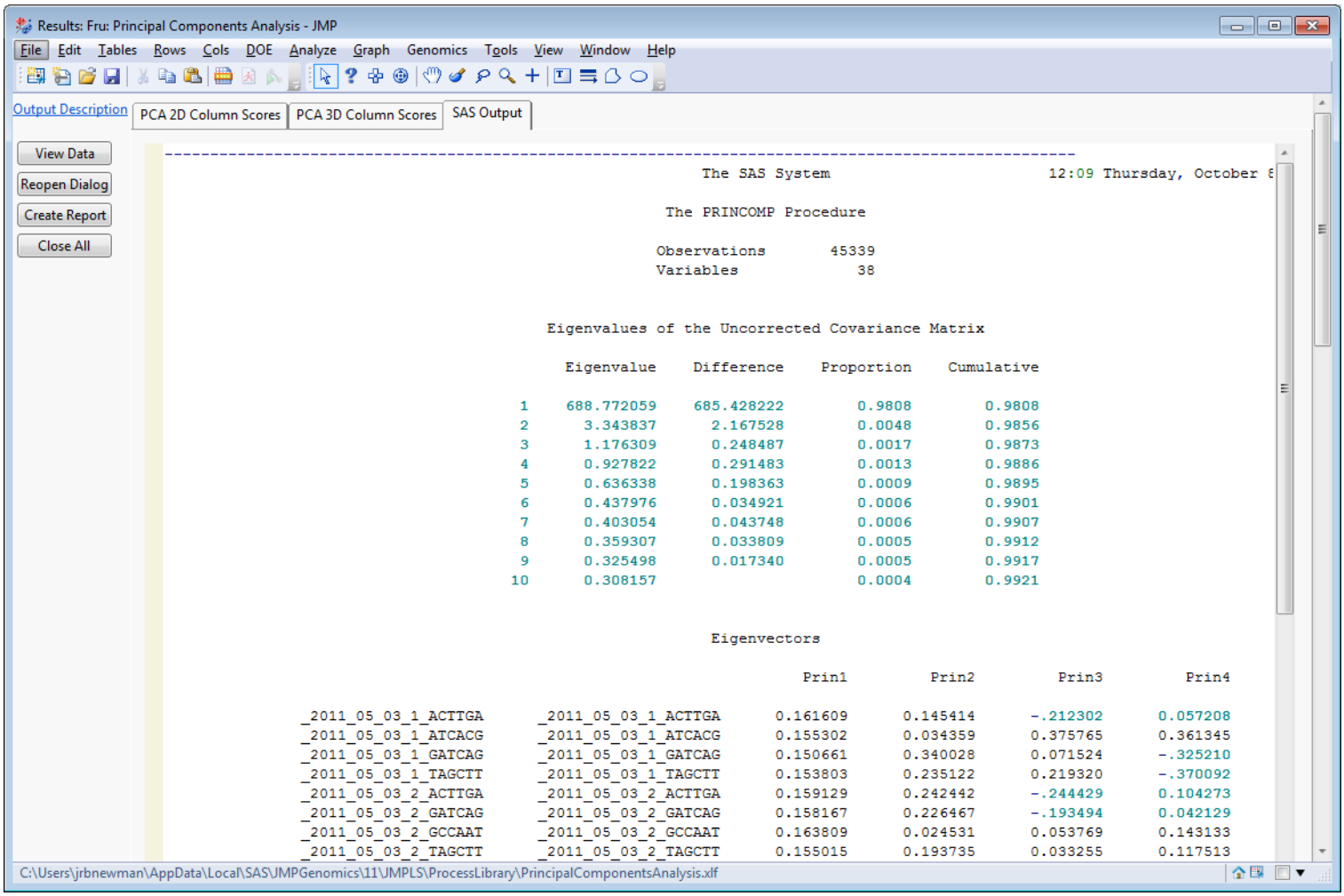
22.) Next, click on the "PCA 3D Column Scores" tab. Like the previous method, this is an interactive plot and can rotated. You can also select samples to see which samples are clustering together. You can also change what principal components to plot.



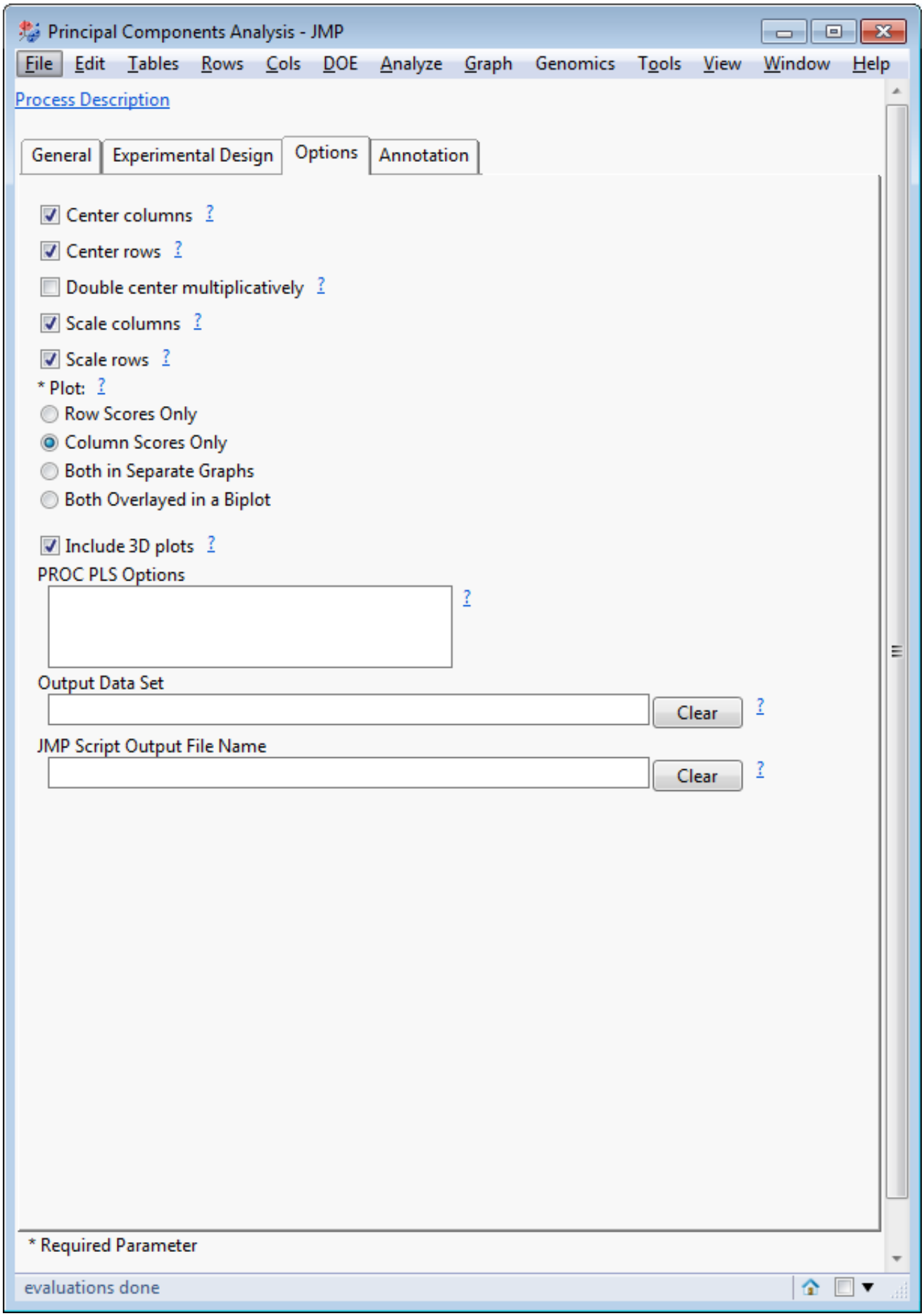
Change the third axis from "pca3" to "pca5". Click on the dropdown menu on the bottom displaying "pca3", click on "Other". Select "pca5", then click "OK". Rotate the 3D scatterplot. What do you think this tells you about the data and principal component 5?



23.) Lastly, look at the "SAS Output" tab. This will contain the eigenvalues and their eigenvectors.

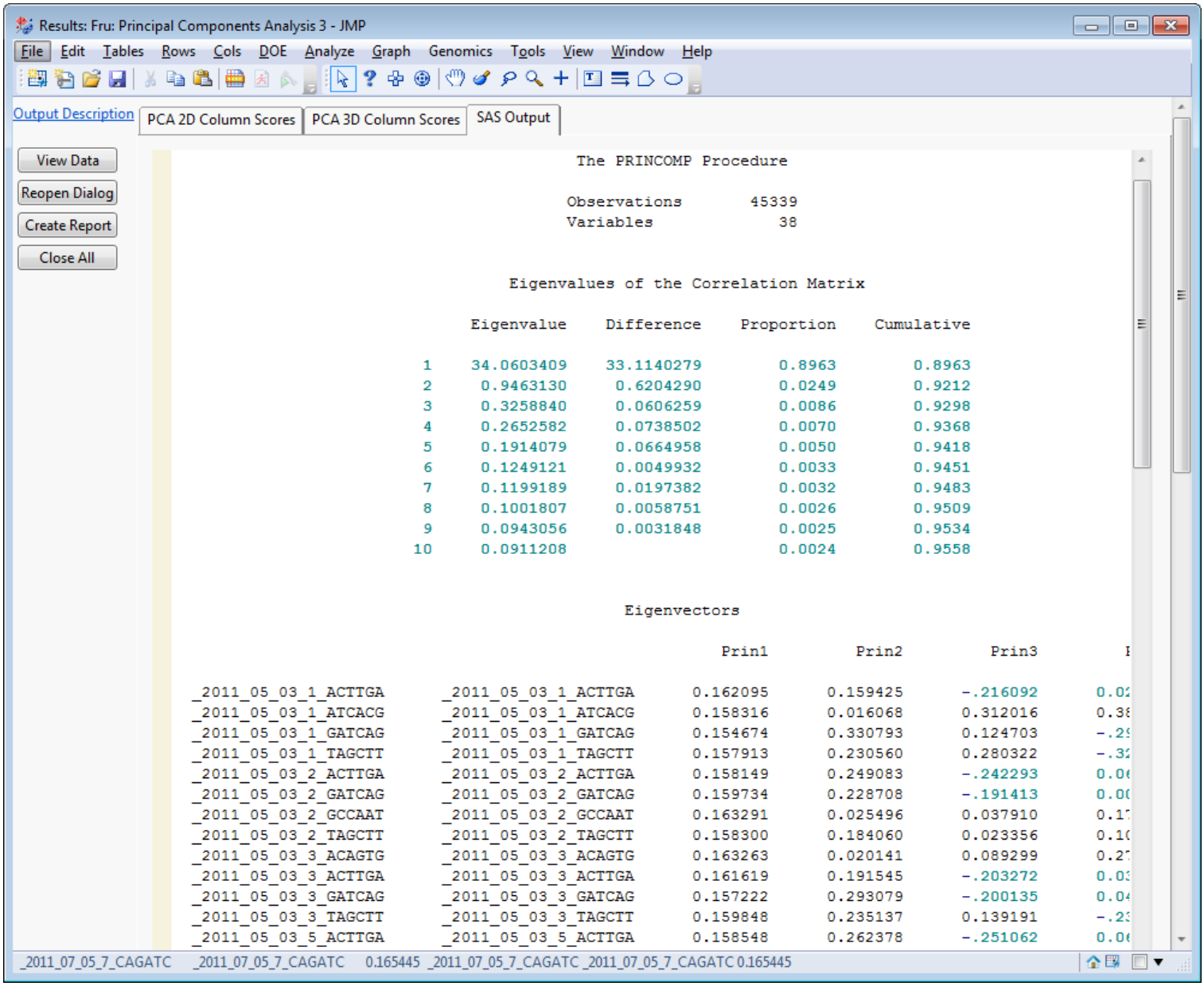


24.) Rerun the PCA but this time under the "Options" tab, and check the boxes so that the window looks like the one below, then click "Run".



25.) Let's check the outputs. Click on the "PCA 3D Column Scores" tab and make sure "pca1", "pca2", and "pca5" are the axes. How does it compare to the unscaled values?

26.) Check the "SAS Output" tab, and compare the eigenvalues with those calculated with the unscaled, uncentered data.



27.) You can repeat all this with the **fru_dataset_jmp** data (untransformed dataset), or the **fru_dataset_jmp_1000fus** and **fru_dataset_allon** and compare these to the previous data.

The MCF7 data

This is section, we will be using the MCF7 data to perform PCA. Like the Fru data, we will go through how to import the data, transform it, and run PCA. Unlike the Fru data, the MCF7 data are in SAS datasets, which JMP can handle natively (CSVs and other non-native formats are converted to SAS or JMP datasets by JMP during import). We will also only run PCA using the Genomics menu since there are additional options over Method 1. You can perform PCA using the first method if you wish however.

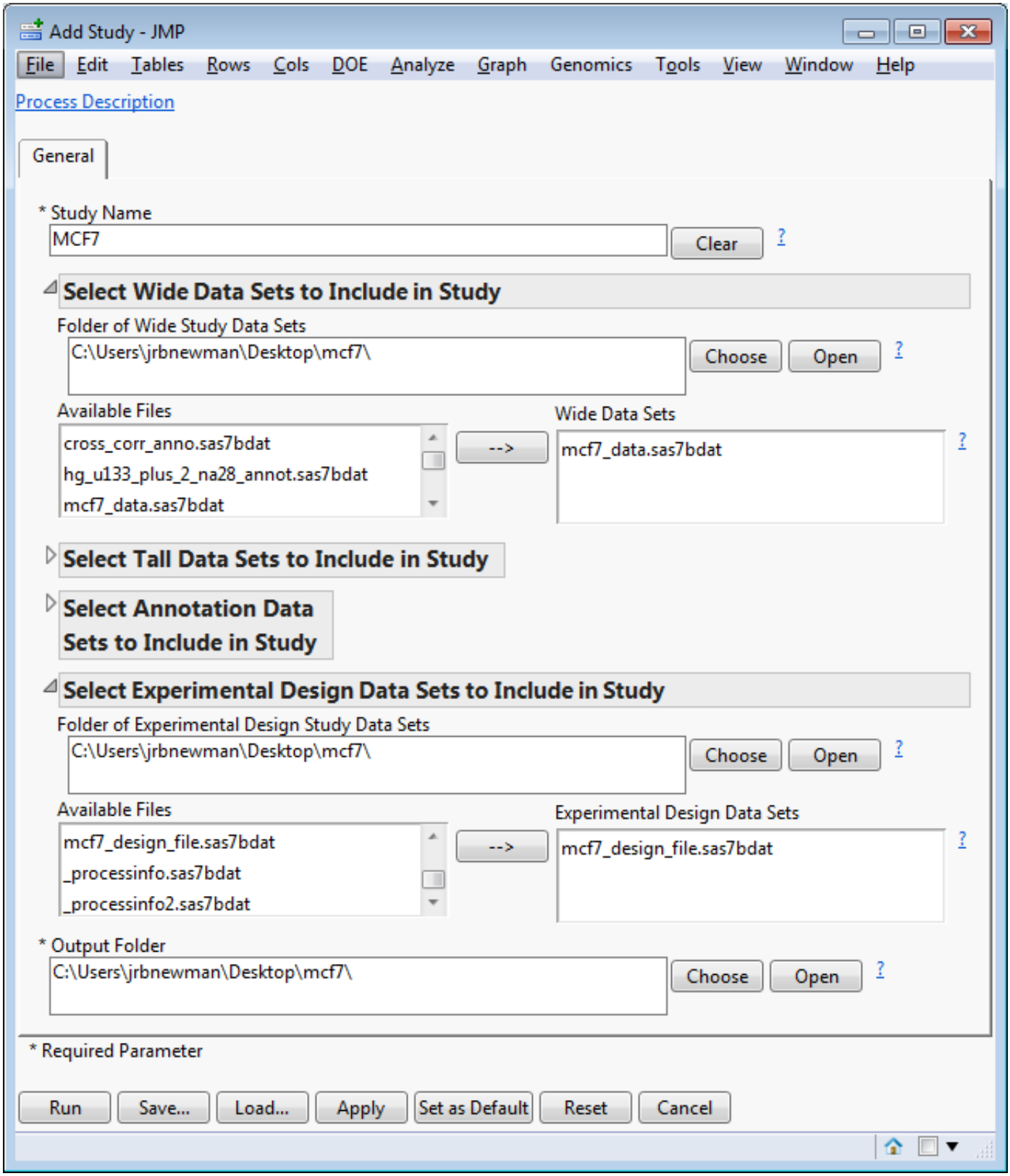
Importing the MCF7 data into JMP

1.) Return to the Genomics Starter window. Select the "Studies" category, and click "Add Study".
2.) In the dialog box, under "Study Name" type "MCF7" without quotes. Under "Output Folder" select the path to where you have saved the MCF7 datasets.

3.) Expand "Select Wide Data Sets to Include/Exclude in Study". In the "Folder of Wide Study Data Sets", select the path to the MCF7 folder. Under "Available Files", select **mcf7_data.sas7bdat**, and click the "-->" button next to "Wide Data Sets".

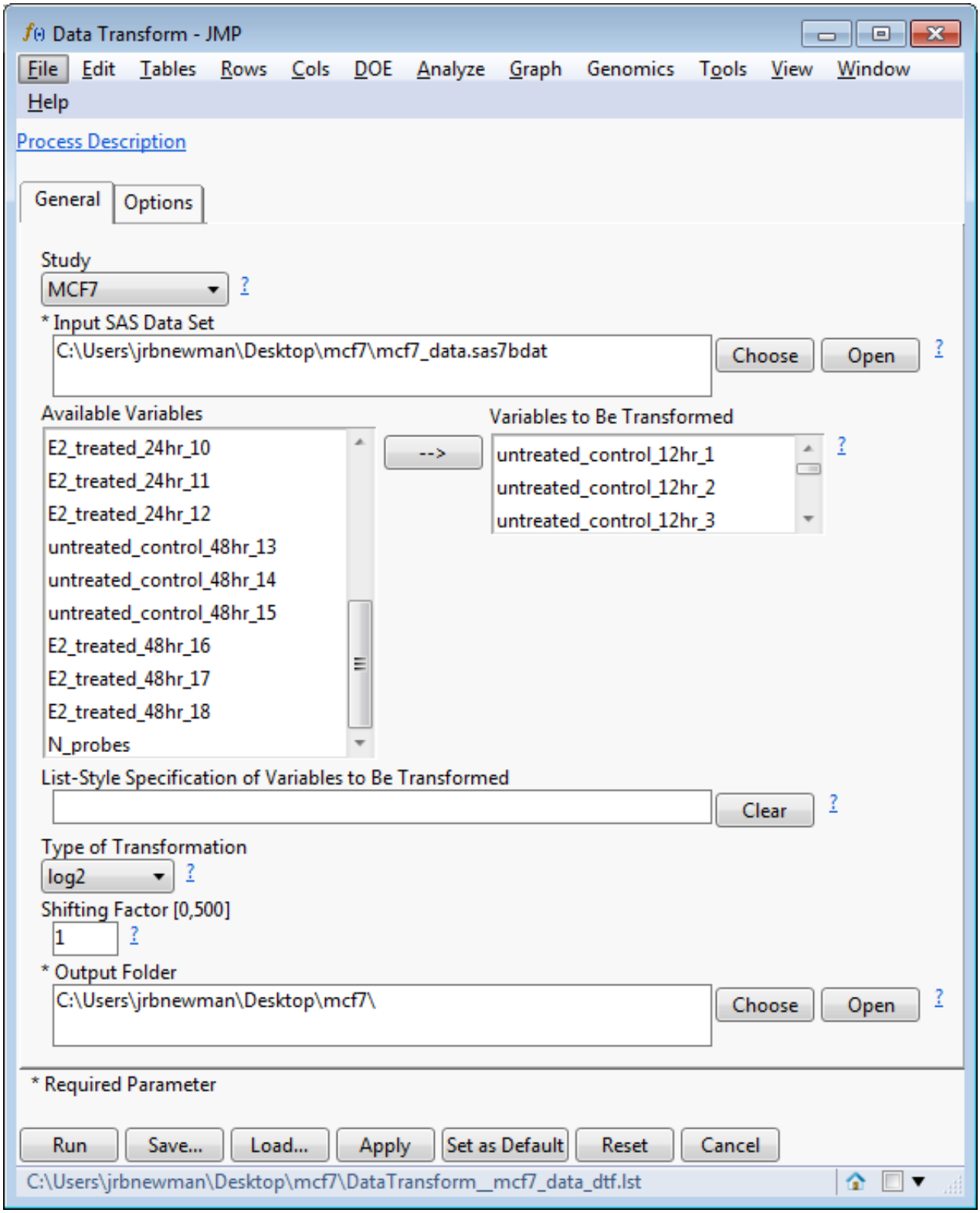
4.) Expand "Select Experimental Design Data Sets to Include/Exclude in Study". Select the path to the MCF7 folder, select **mcf7_design_file.sas7bdat**, and click the "-->" button next to "Experimental Design Data Sets".

5.) The dialog box should look like the following. Click "Run".



Transform MCF7 data

We will again log2-transform the data. Follow the same directions as with the Fru data for this. For the "Available Variables", select everything starting with "untreated_" or "E2_".



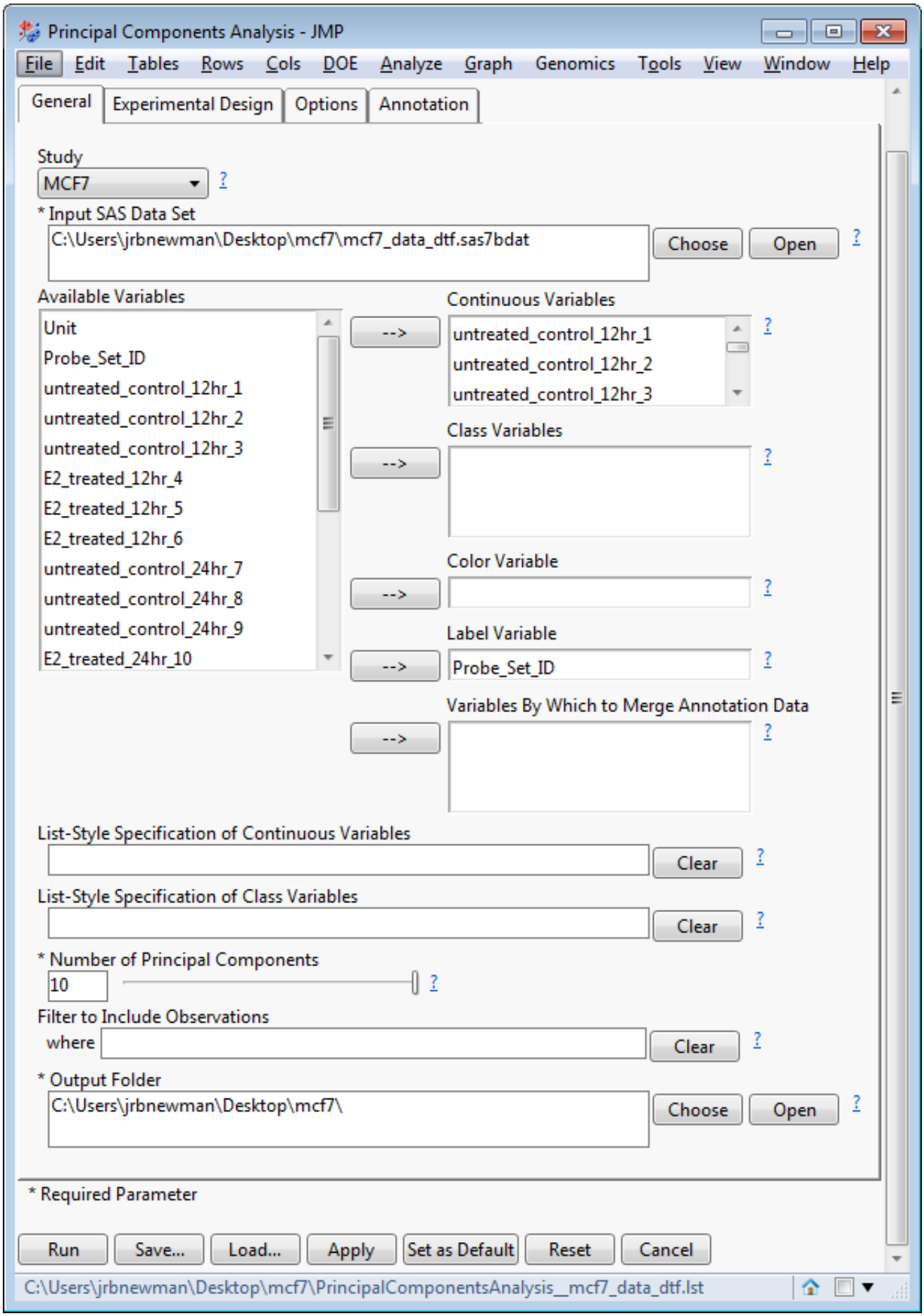
The transformed dataset will be saved as **mcf7_data_dtf.sas7bdat**.

PCA - MCF7

1.) On the Genomics Starter window, click on the "Pattern Discovery" category, then click "Principal Components Analysis".
2.) In the dialog box on the "General" tab, under "Input SAS Data Set" select **mcf7_data_dtf.sas7bdat**.

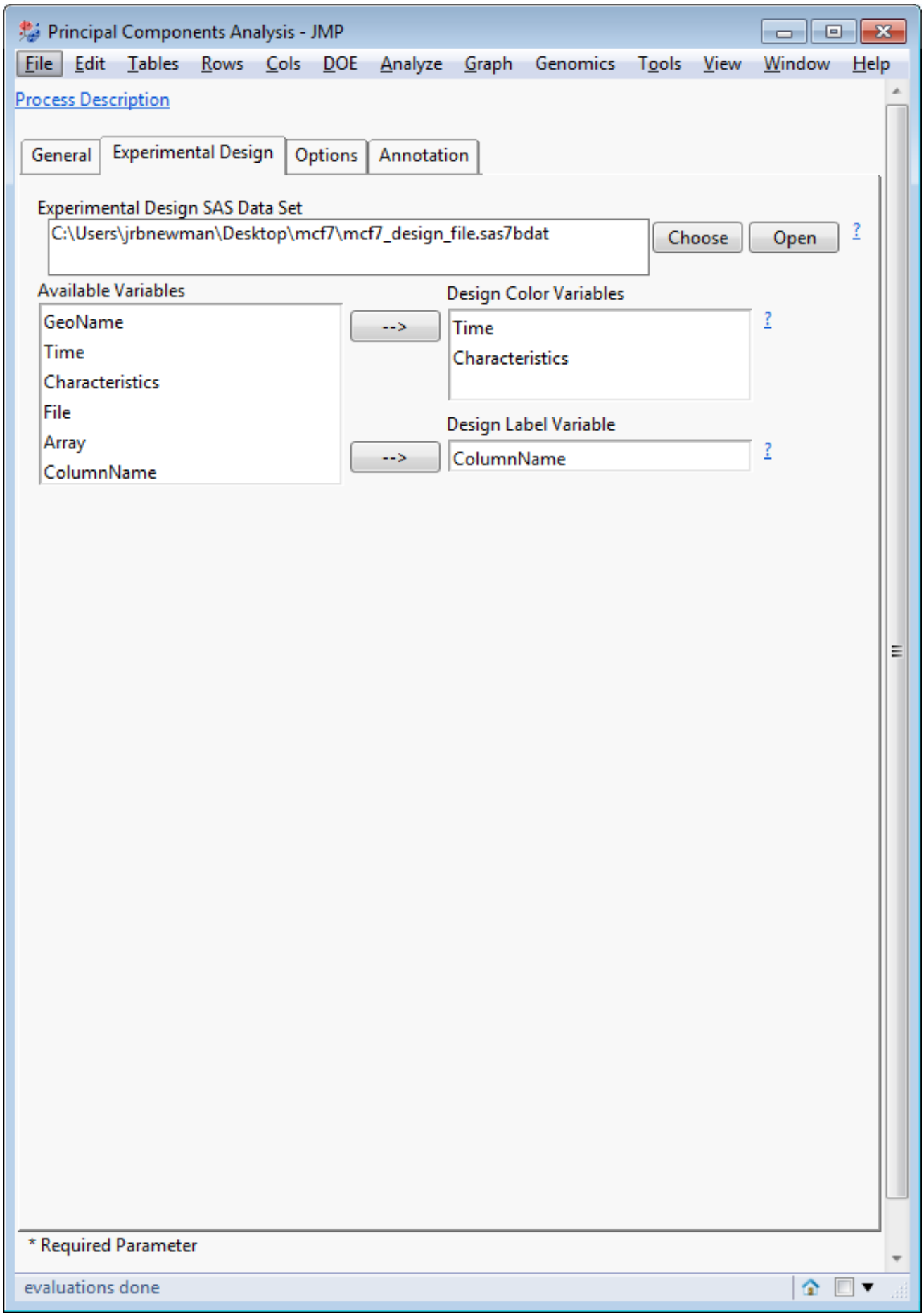
3.) Under "Available Variables" select everything beginning with "untreated_" or "E2_" and click on the "-->" button next to "Continuous Variables".
4.) Under "Available Variables" select "Probe_Set_ID" and click on the "-->" button next to "Label Variable".

5.) Under "Number of Principal Components", specify 10 as the number of components you wish to calculate.
6.) Under "Output Folder" select your output folder. The dialog box should look like this:


Principal Components Analysis - JMP
File Edit Tables Rows Cols DOE Analyze Graph Genomics Tools View Window Help
General
Experimental Design
Options
Annotation
Study
MCF7
* Input SAS Data Set
C:\Users\jrbnewman\Desktop\mcf7\mcf7_data_dtf.sas7bdat
Choose
Open
Available Variables
Unit
Probe_Set_ID
untreated_control_12hr_1
untreated_control_12hr_2
untreated_control_12hr_3
E2_treated_12hr_4
E2_treated_12hr_5
E2_treated_12hr_6
untreated_control_24hr_7
untreated_control_24hr_8
untreated_control_24hr_9
E2_treated_24hr_10
-->
Continuous Variables
untreated_control_12hr_1
untreated_control_12hr_2
untreated_control_12hr_3
Class Variables
Color Variable
Label Variable
Probe_Set_ID
Variables By Which to Merge Annotation Data
List-Style Specification of Continuous Variables
Clear
List-Style Specification of Class Variables
Clear
* Number of Principal Components
10
Filter to Include Observations
where
Clear
* Output Folder
C:\Users\jrbnewman\Desktop\mcf7\
Choose
Open
* Required Parameter
Run
Save...
Load...
Apply
Set as Default
Reset
Cancel
C:\Users\jrbnewman\Desktop\mcf7\PrincipalComponentsAnalysis__mcf7_data_dtf.lst

8.) Select the "Experimental Design" tab. Under "Experimental Design SAS Data Set" click "Choose".
9.) Under "Experimental Design Data Sets" and select **mcf7_design_file.sas7bdat**. Click "OK".
10.) Under "Available Variables", select "Time" and "Characteristics" and click the "-->" next to "Design Color Variables".
11.) Under "Available Variables", select "ColumnName" and click the "-->" next to "Design Label Variable" to match the variables here to the expression data.

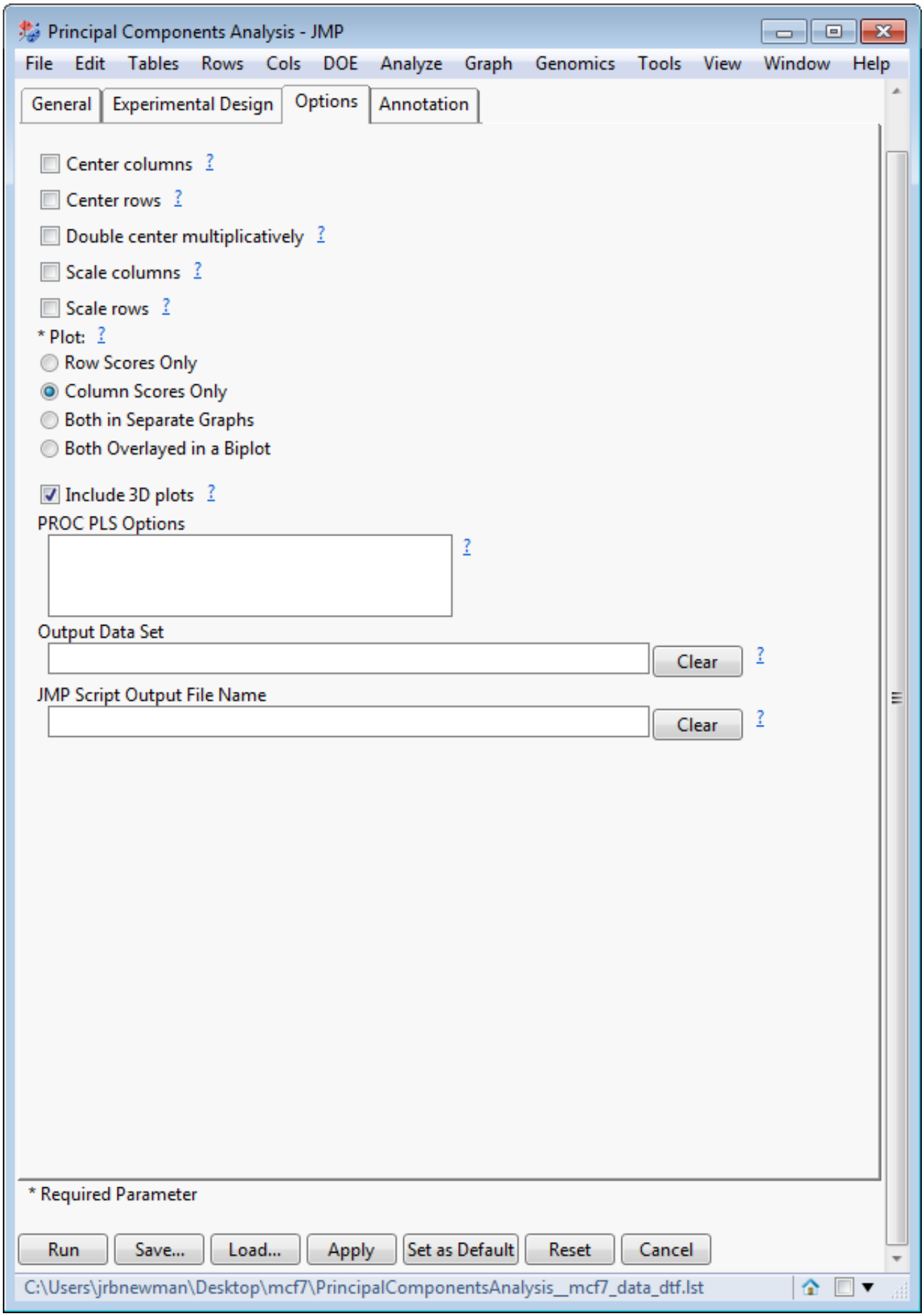
12.) The dialog box should look like this:


Principal Components Analysis - JMP
File Edit Tables Rows Cols DOE Analyze Graph Genomics Tools View Window Help
Process Description
General
Experimental Design
Options
Annotation
Experimental Design SAS Data Set
C:\Users\jrbnewman\Desktop\mcf7\mcf7_design_file.sas7bdat
Choose
Open
?
Available Variables
GeoName
Time
Characteristics
File
Array
ColumnName
-->
Design Color Variables
Time
Characteristics
?
Design Label Variable
-->
ColumnName
?
* Required Parameter
evaluations done

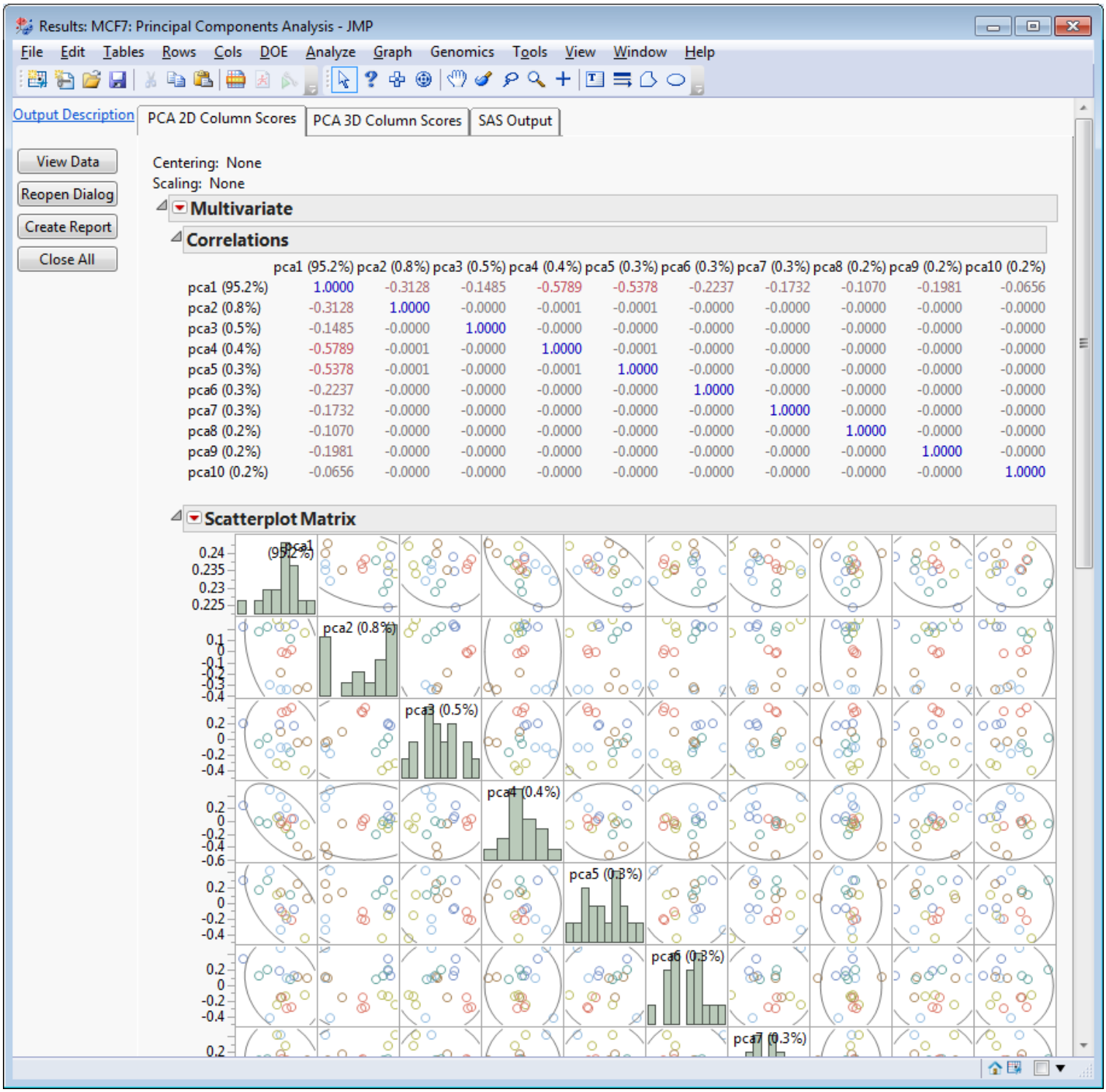
13.) Select the "Options" tab. For now, uncheck the five checkboxes.

14.) The "Plot" radio selections allow you to select whether you want to perform PCA on rows (fusions), columns (samples), or both. Since we are going to only perform PCA on the samples, select "Column Scores Only".

15.) Make sure "Include 3D plots" is checked. The dialog box should look like the following. Click "Run".

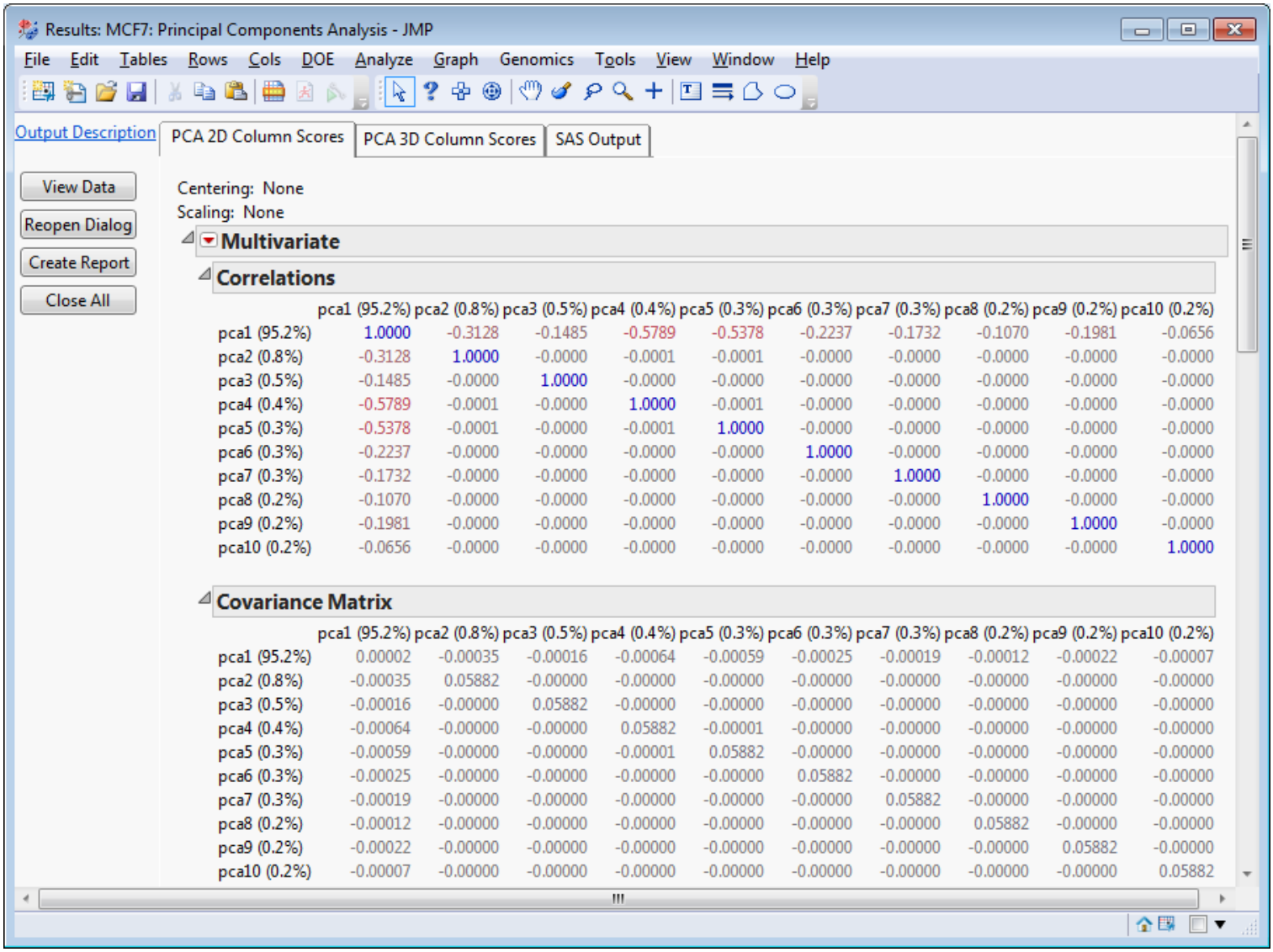


16.) The output window should appear.



	pca1 (95.2%)	pca2 (0.8%)	pca3 (0.5%)	pca4 (0.4%)	pca5 (0.3%)	pca6 (0.3%)	pca7 (0.3%)	pca8 (0.2%)	pca9 (0.2%)	pca10 (0.2%)
pca1 (95.2%)	1.0000	-0.3128	-0.1485	-0.5789	-0.5378	-0.2237	-0.1732	-0.1070	-0.1981	-0.0656
pca2 (0.8%)	-0.3128	1.0000	-0.0000	-0.0001	-0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca3 (0.5%)	-0.1485	-0.0000	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca4 (0.4%)	-0.5789	-0.0001	-0.0000	1.0000	-0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca5 (0.3%)	-0.5378	-0.0001	-0.0000	-0.0001	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca6 (0.3%)	-0.2237	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca7 (0.3%)	-0.1732	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000	-0.0000
pca8 (0.2%)	-0.1070	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000
pca9 (0.2%)	-0.1981	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000
pca10 (0.2%)	-0.0656	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000

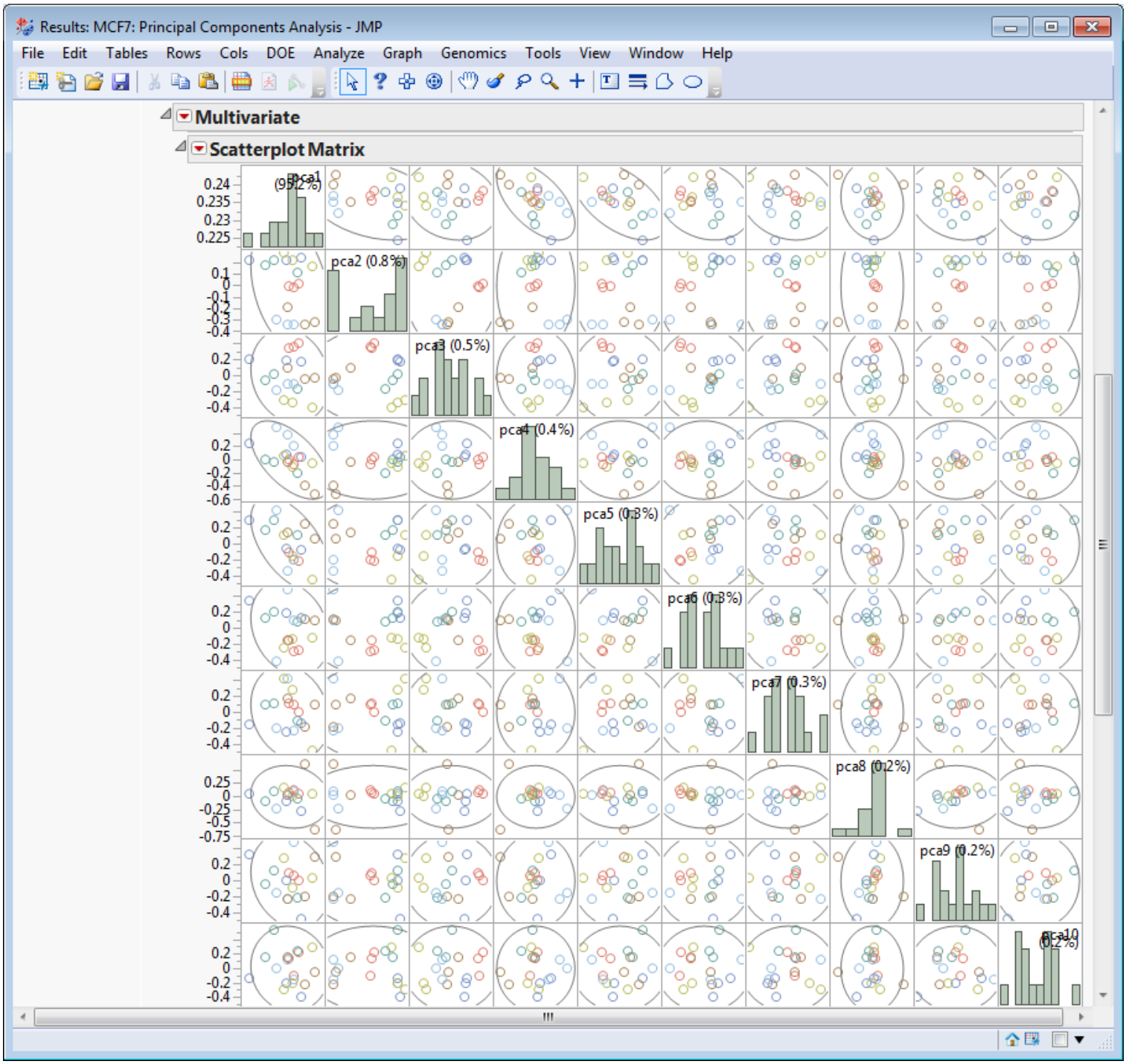
17.) Starting with the "PCA 2D Column Scores" tab, you should see the correlation matrix between principal components. You can also view the covariance matrix between principal components by click the red dropdown button and select "Covariance Matrix".



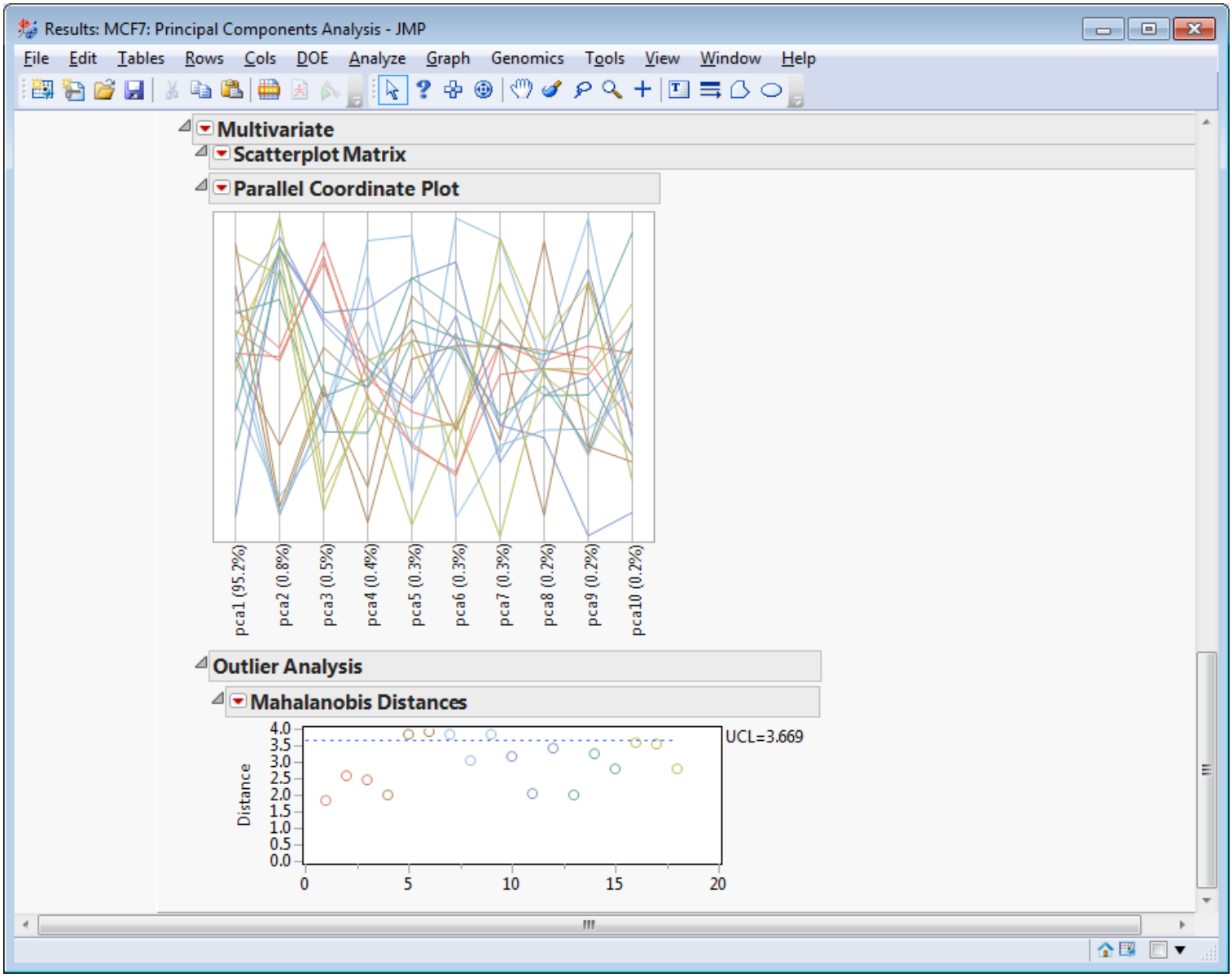
	pca1 (95.2%)	pca2 (0.8%)	pca3 (0.5%)	pca4 (0.4%)	pca5 (0.3%)	pca6 (0.3%)	pca7 (0.3%)	pca8 (0.2%)	pca9 (0.2%)	pca10 (0.2%)
pca1 (95.2%)	1.0000	-0.3128	-0.1485	-0.5789	-0.5378	-0.2237	-0.1732	-0.1070	-0.1981	-0.0656
pca2 (0.8%)	-0.3128	1.0000	-0.0000	-0.0001	-0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca3 (0.5%)	-0.1485	-0.0000	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca4 (0.4%)	-0.5789	-0.0001	-0.0000	1.0000	-0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca5 (0.3%)	-0.5378	-0.0001	-0.0000	-0.0001	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca6 (0.3%)	-0.2237	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000	-0.0000	-0.0000
pca7 (0.3%)	-0.1732	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000	-0.0000
pca8 (0.2%)	-0.1070	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000	-0.0000
pca9 (0.2%)	-0.1981	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000	-0.0000
pca10 (0.2%)	-0.0656	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	1.0000

	pca1 (95.2%)	pca2 (0.8%)	pca3 (0.5%)	pca4 (0.4%)	pca5 (0.3%)	pca6 (0.3%)	pca7 (0.3%)	pca8 (0.2%)	pca9 (0.2%)	pca10 (0.2%)
pca1 (95.2%)	0.00002	-0.00035	-0.00016	-0.00064	-0.00059	-0.00025	-0.00019	-0.00012	-0.00022	-0.00007
pca2 (0.8%)	-0.00035	0.05882	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
pca3 (0.5%)	-0.00016	-0.00000	0.05882	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
pca4 (0.4%)	-0.00064	-0.00000	-0.00000	0.05882	-0.00001	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
pca5 (0.3%)	-0.00059	-0.00000	-0.00000	-0.00001	0.05882	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
pca6 (0.3%)	-0.00025	-0.00000	-0.00000	-0.00000	-0.00000	0.05882	-0.00000	-0.00000	-0.00000	-0.00000
pca7 (0.3%)	-0.00019	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.05882	-0.00000	-0.00000	-0.00000
pca8 (0.2%)	-0.00012	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.05882	-0.00000	-0.00000
pca9 (0.2%)	-0.00022	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.05882	-0.00000
pca10 (0.2%)	-0.00007	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.05882

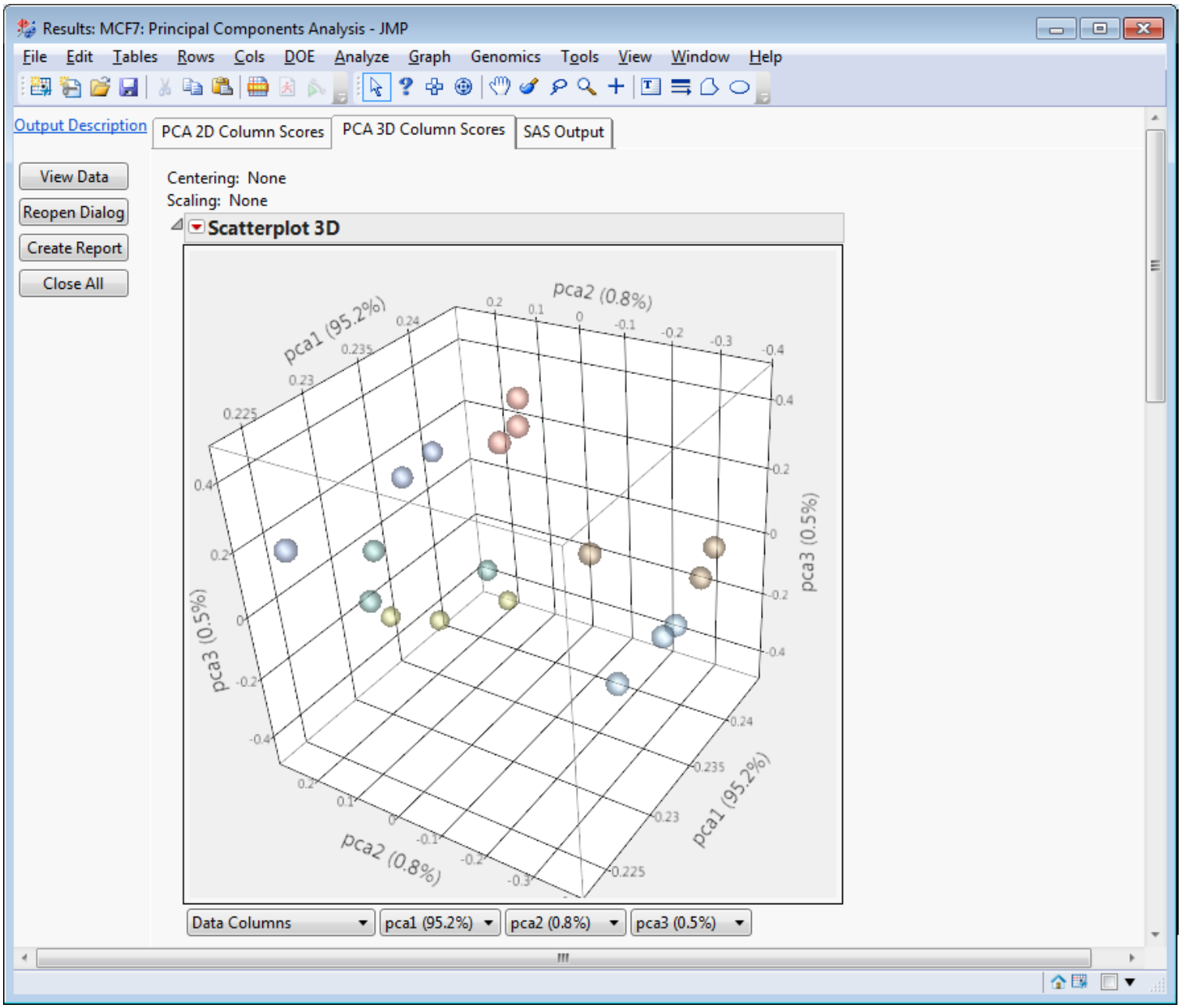
18.) Next is the "Scatterplot Matrix", which is a matrix of the 2D scatterplots between each pair of principal components.



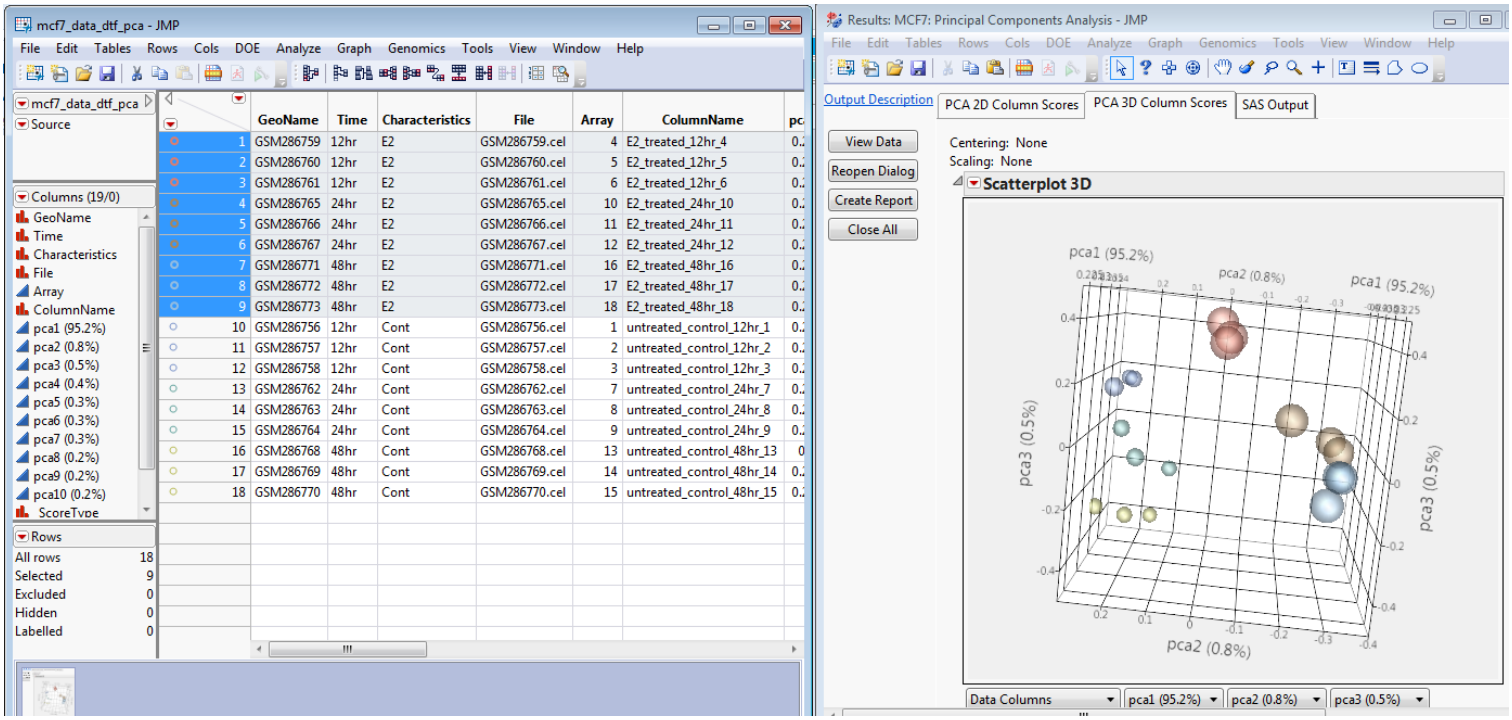
19.) Next is the "Parallel Coordinate Plot" and "Outlier Analysis". As before, you can select samples via the plots, or from the data directly (click on "View Data").



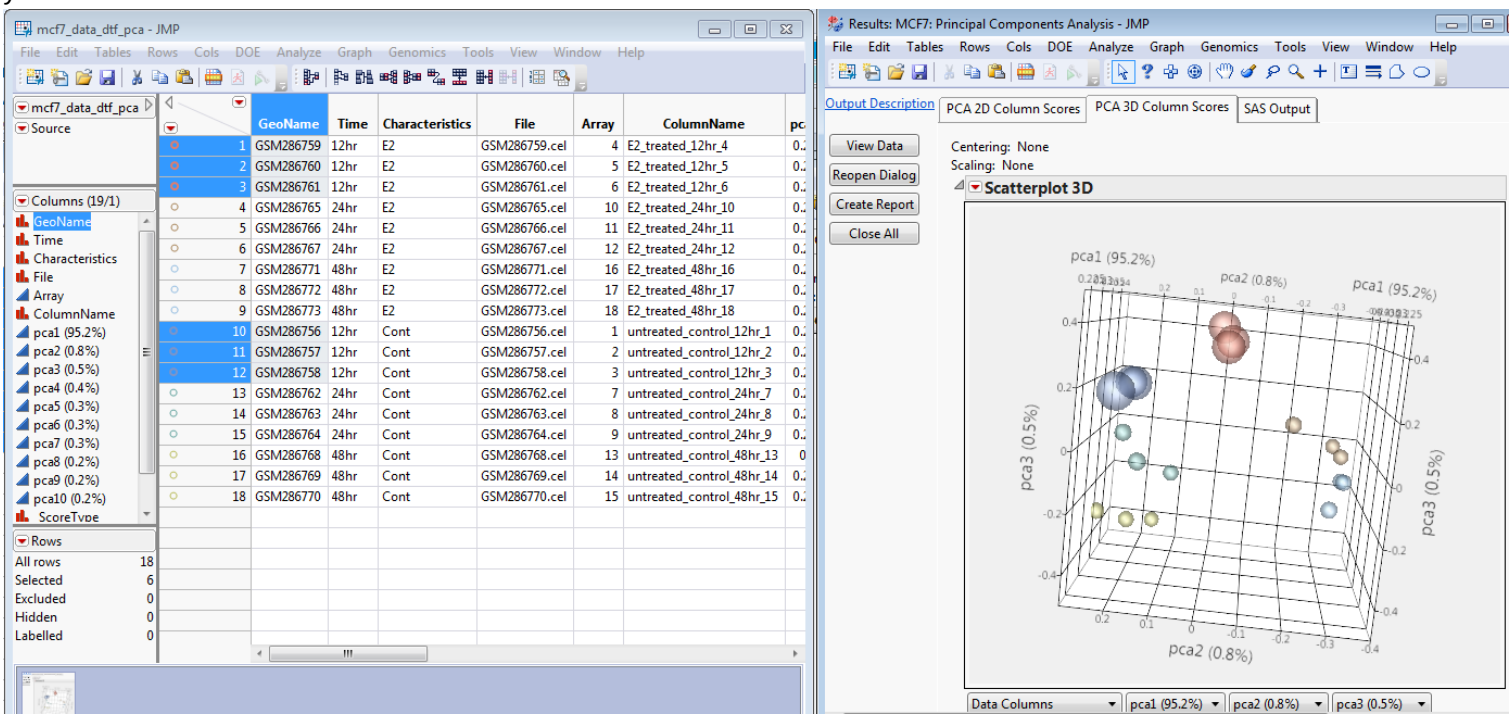
20.) Next, examine the "PCA 3D Column Scores" tab.

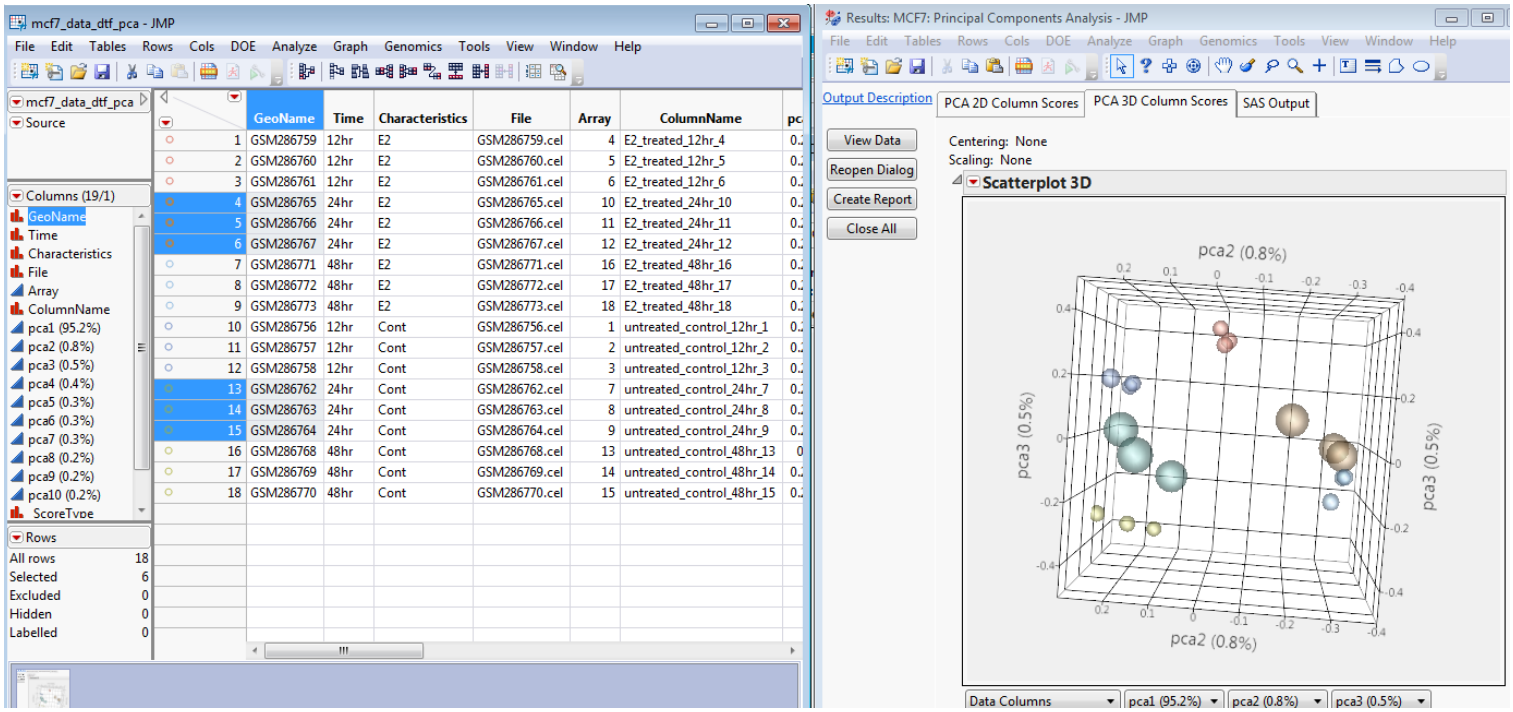


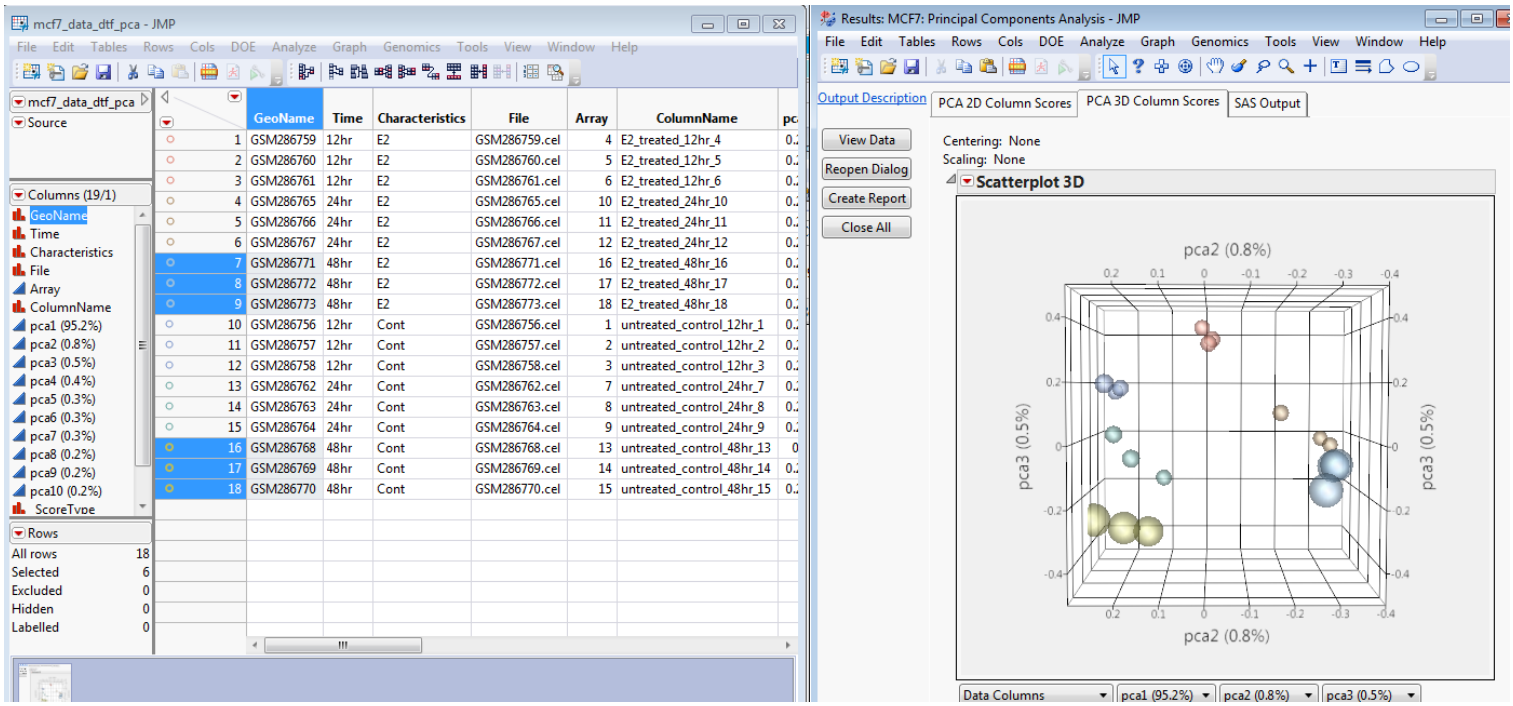
21.) Let's view the data by clicking on the "View Data" button. Select the rows where "E2" appears in the "Characteristics" column. The corresponding datapoints on the 3D plot should enlarge. Rotate the 3D PCA plot. Do the "E2" samples group separately from "Cont" (control) samples?



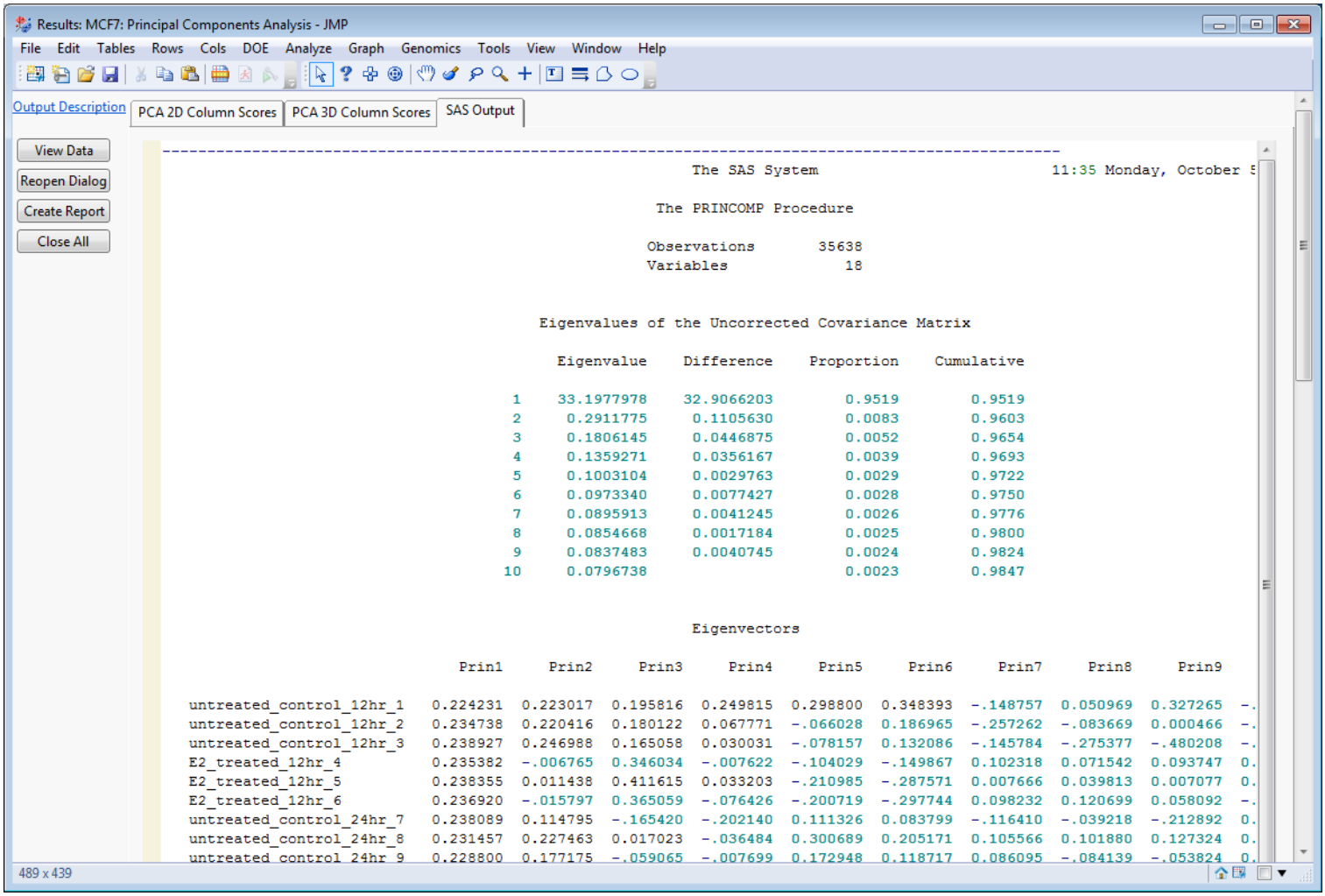
22.) Now select the rows where "12hr" appears in the "Time" column and rotate the 3D plot. How do "12hr" samples compare to other samples? Do the same for "24hr" and "48hr" samples, and see how they compare. What inferences can you make on the data based on this?







23.) Now look at the "SAS Output" tab. This will contain the eigenvalues and their eigenvectors.



24.) Rerun the PCA but this time under the "Options" tab, and check "Center columns" and "Scale columns", then click "Run".

25.) Let's check the outputs.

26.) Let's check the 3D plots, selecting samples as before ("E2" vs "Cont", "12hr"/"24hr"/"48hr"). How do these compare to the previous plots?

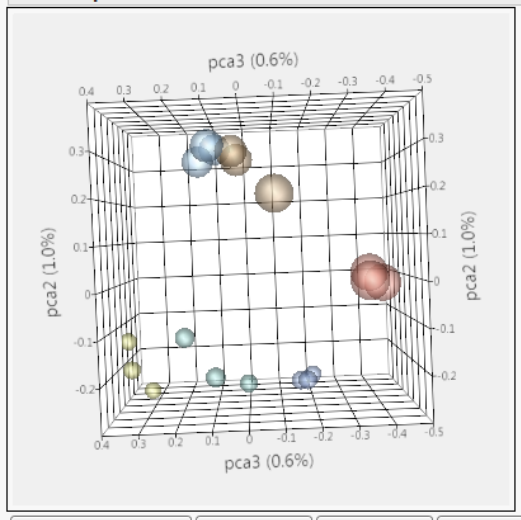
	GeoName	Time	Characteristics	File	Array	ColumnName	pca
1	GSM286759	12hr	E2	GSM286759.cel	4	E2_treated_12hr_4	0.2
2	GSM286760	12hr	E2	GSM286760.cel	5	E2_treated_12hr_5	0.2
3	GSM286761	12hr	E2	GSM286761.cel	6	E2_treated_12hr_6	0.2
4	GSM286765	24hr	E2	GSM286765.cel	10	E2_treated_24hr_10	0.2
5	GSM286766	24hr	E2	GSM286766.cel	11	E2_treated_24hr_11	0.2
6	GSM286767	24hr	E2	GSM286767.cel	12	E2_treated_24hr_12	0.2
7	GSM286771	48hr	E2	GSM286771.cel	16	E2_treated_48hr_16	0.2
8	GSM286772	48hr	E2	GSM286772.cel	17	E2_treated_48hr_17	0.2
9	GSM286773	48hr	E2	GSM286773.cel	18	E2_treated_48hr_18	0.2
10	GSM286756	12hr	Cont	GSM286756.cel	1	untreated_control_12hr_1	0.2
11	GSM286757	12hr	Cont	GSM286757.cel	2	untreated_control_12hr_2	0.2
12	GSM286758	12hr	Cont	GSM286758.cel	3	untreated_control_12hr_3	0.2
13	GSM286762	24hr	Cont	GSM286762.cel	7	untreated_control_24hr_7	0.2
14	GSM286763	24hr	Cont	GSM286763.cel	8	untreated_control_24hr_8	0.2
15	GSM286764	24hr	Cont	GSM286764.cel	9	untreated_control_24hr_9	0.2
16	GSM286768	48hr	Cont	GSM286768.cel	13	untreated_control_48hr_13	0.2
17	GSM286769	48hr	Cont	GSM286769.cel	14	untreated_control_48hr_14	0.2
18	GSM286770	48hr	Cont	GSM286770.cel	15	untreated_control_48hr_15	0.2

Output Description | PCA 2D Column Scores | PCA 3D Column Scores | SAS Output

View Data
Reopen Dialog
Create Report
Close All

Centering: Columns Only
Scaling: Columns Only

Scatterplot 3D



Data Columns | pca1 (94.3%) | pca2 (1.0%) | pca3 (0.6%)

Run Script

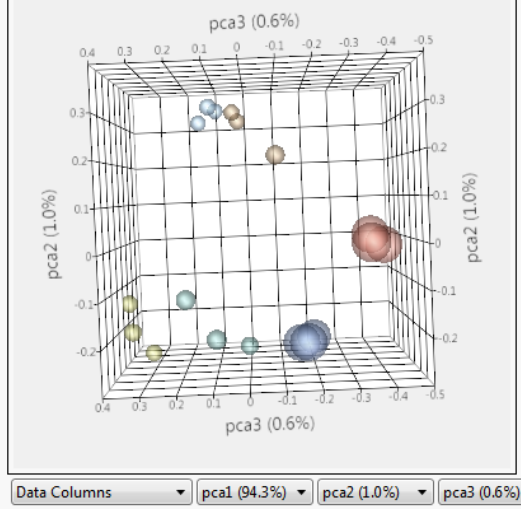
	GeoName	Time	Characteristics	File	Array	ColumnName	pca
1	GSM286759	12hr	E2	GSM286759.cel	4	E2_treated_12hr_4	0.2
2	GSM286760	12hr	E2	GSM286760.cel	5	E2_treated_12hr_5	0.2
3	GSM286761	12hr	E2	GSM286761.cel	6	E2_treated_12hr_6	0.2
4	GSM286765	24hr	E2	GSM286765.cel	10	E2_treated_24hr_10	0.2
5	GSM286766	24hr	E2	GSM286766.cel	11	E2_treated_24hr_11	0.2
6	GSM286767	24hr	E2	GSM286767.cel	12	E2_treated_24hr_12	0.2
7	GSM286771	48hr	E2	GSM286771.cel	16	E2_treated_48hr_16	0.2
8	GSM286772	48hr	E2	GSM286772.cel	17	E2_treated_48hr_17	0.2
9	GSM286773	48hr	E2	GSM286773.cel	18	E2_treated_48hr_18	0.2
10	GSM286756	12hr	Cont	GSM286756.cel	1	untreated_control_12hr_1	0.2
11	GSM286757	12hr	Cont	GSM286757.cel	2	untreated_control_12hr_2	0.2
12	GSM286758	12hr	Cont	GSM286758.cel	3	untreated_control_12hr_3	0.2
13	GSM286762	24hr	Cont	GSM286762.cel	7	untreated_control_24hr_7	0.2
14	GSM286763	24hr	Cont	GSM286763.cel	8	untreated_control_24hr_8	0.2
15	GSM286764	24hr	Cont	GSM286764.cel	9	untreated_control_24hr_9	0.2
16	GSM286768	48hr	Cont	GSM286768.cel	13	untreated_control_48hr_13	0.2
17	GSM286769	48hr	Cont	GSM286769.cel	14	untreated_control_48hr_14	0.2
18	GSM286770	48hr	Cont	GSM286770.cel	15	untreated_control_48hr_15	0.2

Output Description | PCA 2D Column Scores | PCA 3D Column Scores | SAS Output

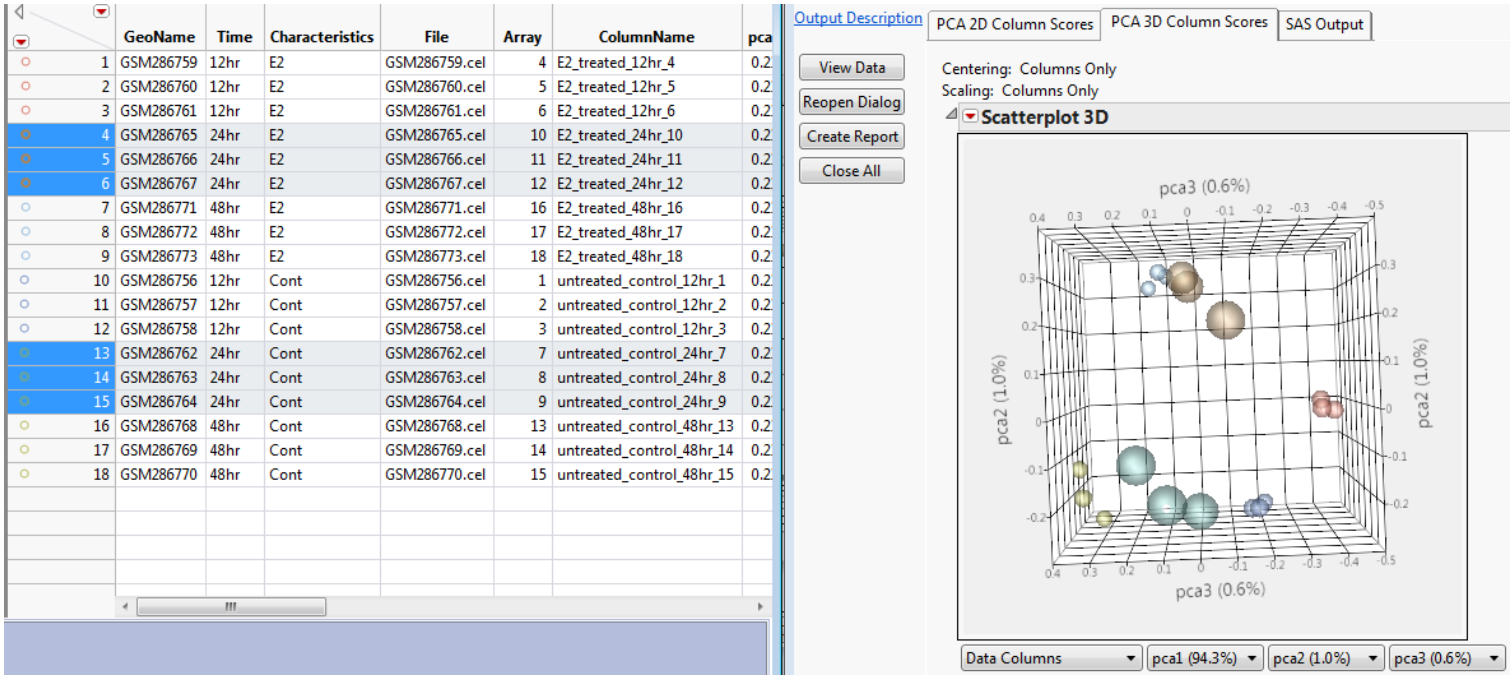
View Data
Reopen Dialog
Create Report
Close All

Centering: Columns Only
Scaling: Columns Only

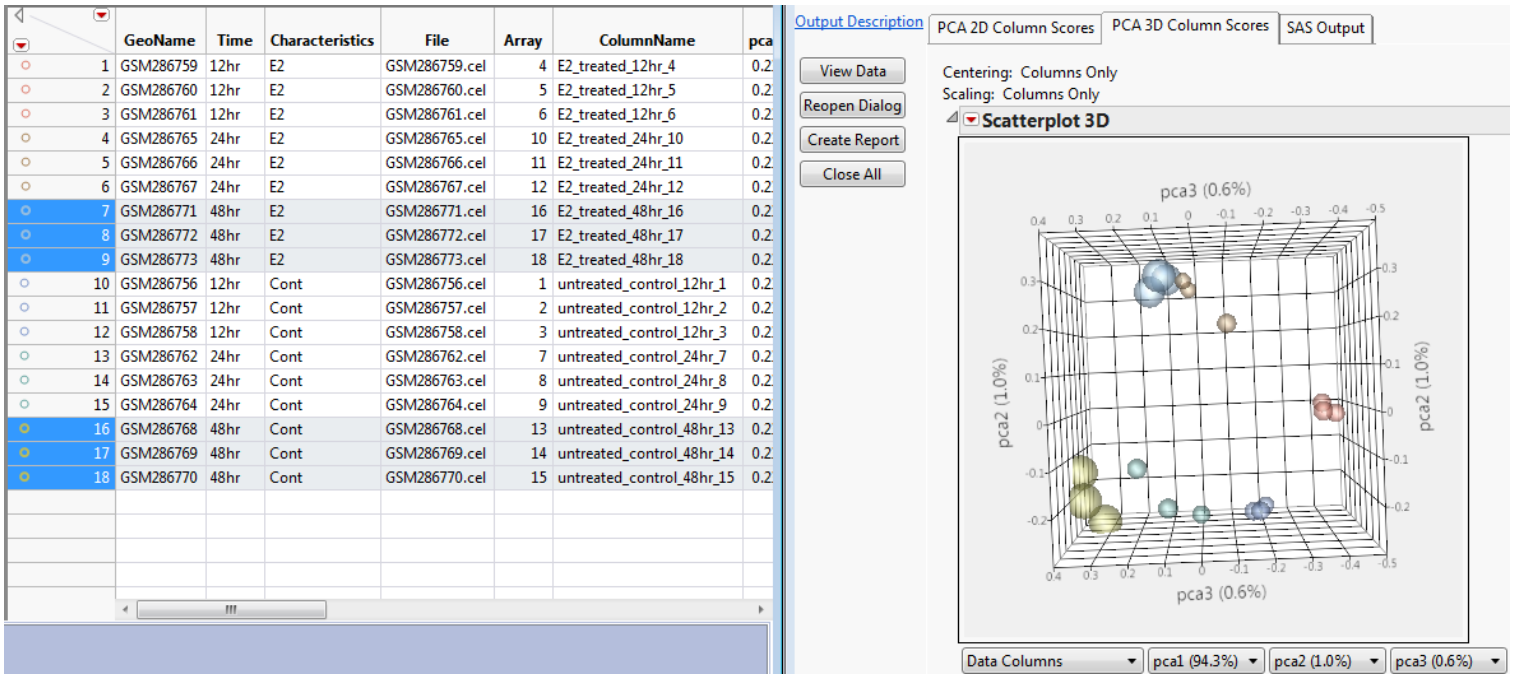
Scatterplot 3D



Data Columns | pca1 (94.3%) | pca2 (1.0%) | pca3 (0.6%)

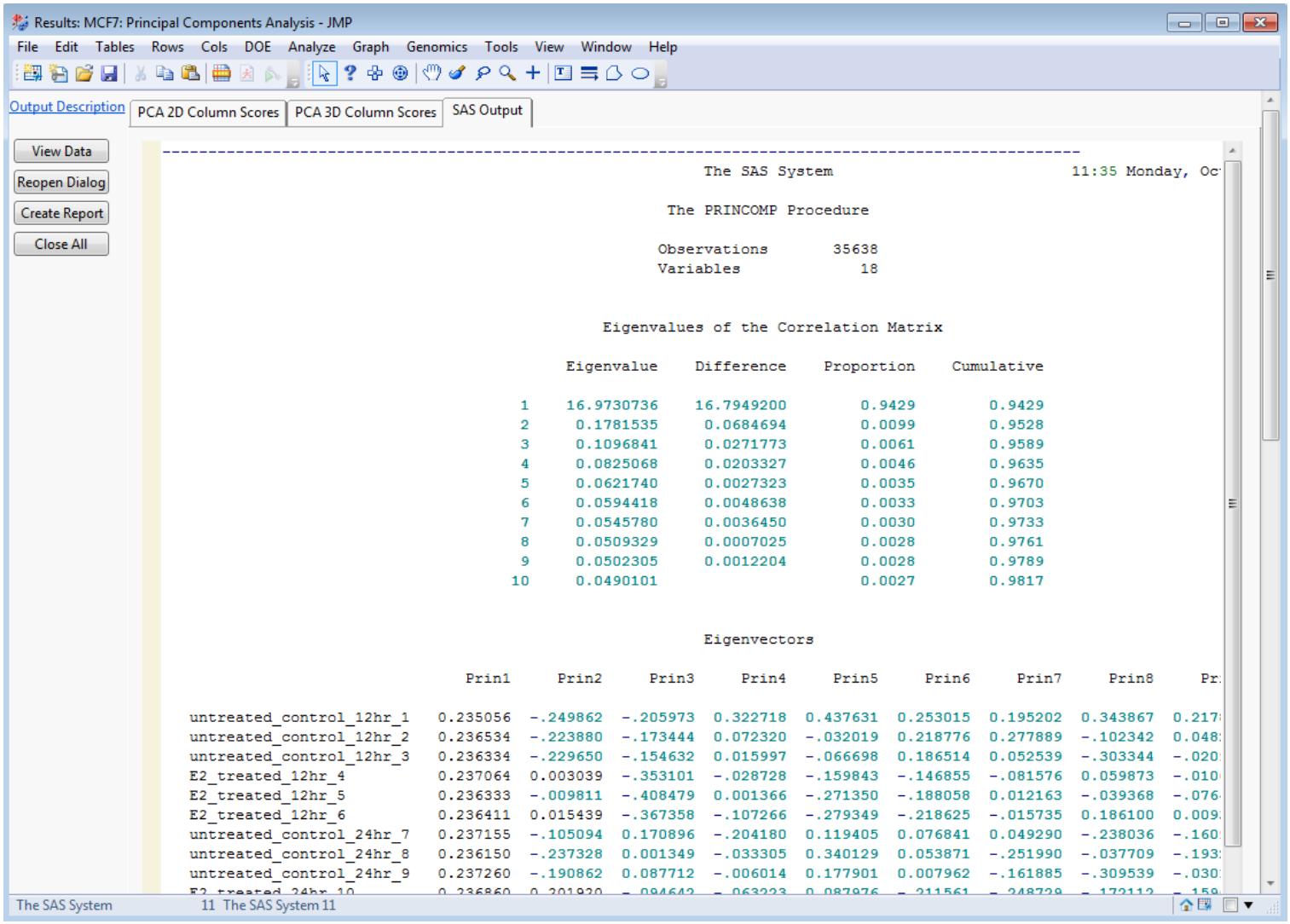


	GeoName	Time	Characteristics	File	Array	ColumnName	pca
1	GSM286759	12hr	E2	GSM286759.cel	4	E2_treated_12hr_4	0.2
2	GSM286760	12hr	E2	GSM286760.cel	5	E2_treated_12hr_5	0.2
3	GSM286761	12hr	E2	GSM286761.cel	6	E2_treated_12hr_6	0.2
4	GSM286765	24hr	E2	GSM286765.cel	10	E2_treated_24hr_10	0.2
5	GSM286766	24hr	E2	GSM286766.cel	11	E2_treated_24hr_11	0.2
6	GSM286767	24hr	E2	GSM286767.cel	12	E2_treated_24hr_12	0.2
7	GSM286771	48hr	E2	GSM286771.cel	16	E2_treated_48hr_16	0.2
8	GSM286772	48hr	E2	GSM286772.cel	17	E2_treated_48hr_17	0.2
9	GSM286773	48hr	E2	GSM286773.cel	18	E2_treated_48hr_18	0.2
10	GSM286756	12hr	Cont	GSM286756.cel	1	untreated_control_12hr_1	0.2
11	GSM286757	12hr	Cont	GSM286757.cel	2	untreated_control_12hr_2	0.2
12	GSM286758	12hr	Cont	GSM286758.cel	3	untreated_control_12hr_3	0.2
13	GSM286762	24hr	Cont	GSM286762.cel	7	untreated_control_24hr_7	0.2
14	GSM286763	24hr	Cont	GSM286763.cel	8	untreated_control_24hr_8	0.2
15	GSM286764	24hr	Cont	GSM286764.cel	9	untreated_control_24hr_9	0.2
16	GSM286768	48hr	Cont	GSM286768.cel	13	untreated_control_48hr_13	0.2
17	GSM286769	48hr	Cont	GSM286769.cel	14	untreated_control_48hr_14	0.2
18	GSM286770	48hr	Cont	GSM286770.cel	15	untreated_control_48hr_15	0.2



	GeoName	Time	Characteristics	File	Array	ColumnName	pca
1	GSM286759	12hr	E2	GSM286759.cel	4	E2_treated_12hr_4	0.2
2	GSM286760	12hr	E2	GSM286760.cel	5	E2_treated_12hr_5	0.2
3	GSM286761	12hr	E2	GSM286761.cel	6	E2_treated_12hr_6	0.2
4	GSM286765	24hr	E2	GSM286765.cel	10	E2_treated_24hr_10	0.2
5	GSM286766	24hr	E2	GSM286766.cel	11	E2_treated_24hr_11	0.2
6	GSM286767	24hr	E2	GSM286767.cel	12	E2_treated_24hr_12	0.2
7	GSM286771	48hr	E2	GSM286771.cel	16	E2_treated_48hr_16	0.2
8	GSM286772	48hr	E2	GSM286772.cel	17	E2_treated_48hr_17	0.2
9	GSM286773	48hr	E2	GSM286773.cel	18	E2_treated_48hr_18	0.2
10	GSM286756	12hr	Cont	GSM286756.cel	1	untreated_control_12hr_1	0.2
11	GSM286757	12hr	Cont	GSM286757.cel	2	untreated_control_12hr_2	0.2
12	GSM286758	12hr	Cont	GSM286758.cel	3	untreated_control_12hr_3	0.2
13	GSM286762	24hr	Cont	GSM286762.cel	7	untreated_control_24hr_7	0.2
14	GSM286763	24hr	Cont	GSM286763.cel	8	untreated_control_24hr_8	0.2
15	GSM286764	24hr	Cont	GSM286764.cel	9	untreated_control_24hr_9	0.2
16	GSM286768	48hr	Cont	GSM286768.cel	13	untreated_control_48hr_13	0.2
17	GSM286769	48hr	Cont	GSM286769.cel	14	untreated_control_48hr_14	0.2
18	GSM286770	48hr	Cont	GSM286770.cel	15	untreated_control_48hr_15	0.2

27.) Check the "SAS Output" tab, and compare the eigenvalues with those calculated with the unscaled, uncentered data.



28.) You can repeat all this with the **mcf7_data** data (untransformed dataset) if you wish. Eigenvalues and eigenvectors will be different.

In []: