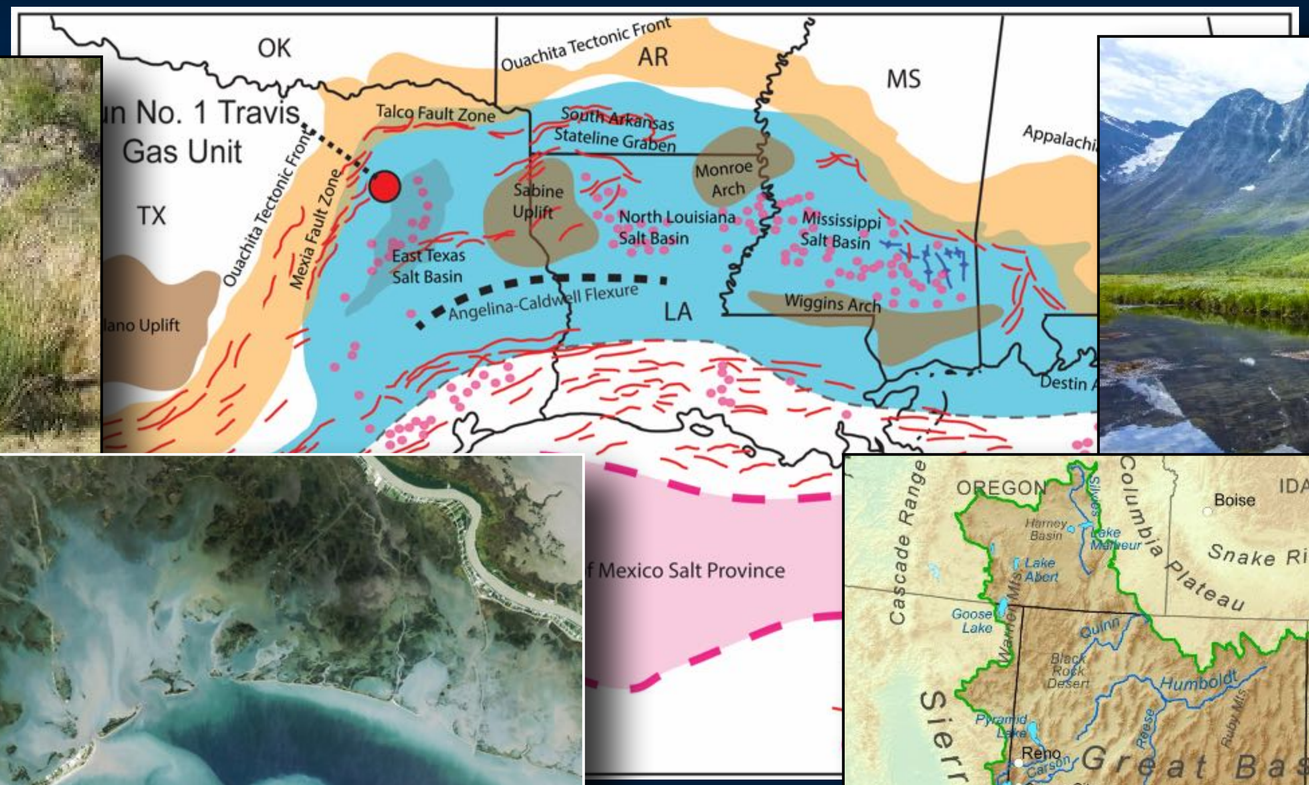
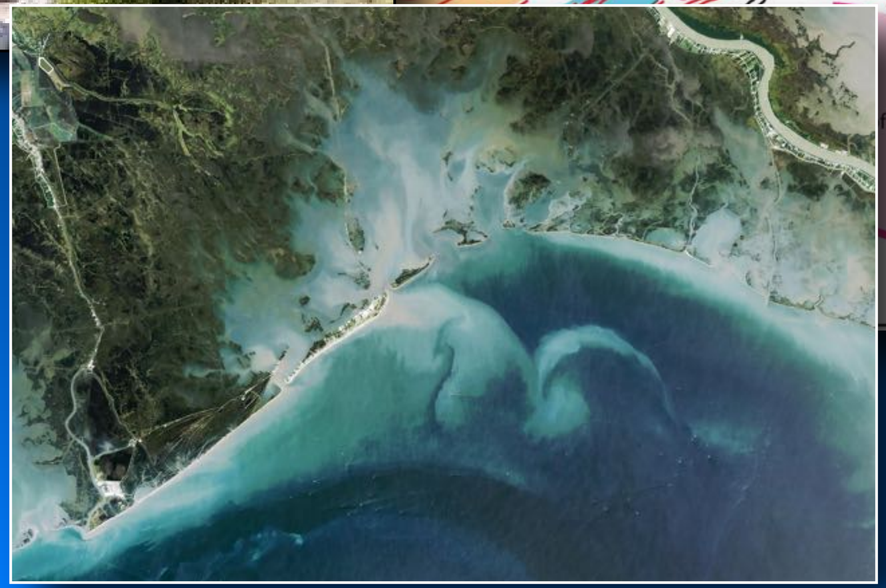
Lab Practical Assignment III

Multivariate Data Analysis & Dimensionality Reduction





Lab Practical Assignment III

1. The US government has exploded a large number of nuclear explosive devices underground at its Oasis Valley Test Site (Nevada), many at or below the water table. Farnham et al. (2000) provide a table of analyses for 19 trace elements collected from 22 wells at this location. These data are provided in the Oasis Valley datafiles, along the locations of the sampled wells and springs. Determine whether similarities exist among the trace element compositions of these water sources using principal components analysis (PCA).

   i. Decide whether the data need to be transformed prior to analysis. Explain the logic behind your decision.

   ii. Decide what the appropriate basis matrix for your analysis is, given the decision you made above. Explain your logic.

   iii. Decide how many principal components (eigenvectors) are necessary to represent the major features of these data. Explain your logic.

   iv. Do natural trace-element groupings of water bodies exist in these data? If so, define these groups and characterize both their similarities and differences in terms of the original variables. Show your plots and the equations of the eigenvector axes.

   v. Which of these groupings (if any) reflect the geographic location of the sampled locations? Summarize the data you used to answer this question on a geographic location map. Note: some localities may be represented by multiple samples.

Lab Practical Assignment III

Oasis Valley, Nevada

Site	Sample	X-Coord	Y-Coord	Li	Ti	V	Cr	Mn	Ni	Ge	As	Se	Rb	Sr
1	Oasis Valley	5.86	19.82	93	1.15	10.3	1.42	0.15	0.12	0.32	13.9	1.78	22.5	148
1	Oasis Valley	5.86	19.82	103	1.87	6.96	0.67	0.09	0.18	0.26	15.1	0.92	23.1	158
1	Oasis Valley	5.86	19.82	104	1.84	7.07	0.67	0.09	0.18	0.27	15.2	0.89	23.6	160
2	Oasis Valley	14.56	43.3	166	0.47	2.67	0.8	0.34	0.1	1.55	6.4	0.57	18.1	163
2	Oasis Valley	14.56	43.3	179	0.34	2.16	0.37	0.49	0.09	2.36	6.4	0.64	18.4	167
3	Oasis Valley	61.63	85.69	52	2.78	2.2	0.93	26.6	4.76	0.51	5.8	0.45	11.4	23.7
3	Oasis Valley	61.63	85.69	51	3.04	2	0.87	26.1	4.33	0.5	5.5	0.51	11.1	23
4	Oasis Valley	76.75	44.22	63	0.16	4.93	2.43	17	0.67	0.19	7.8	0.57	3.7	6.7
5	Oasis Valley	19.28	51.38	175	0.42	1.78	1.47	0.54	1.07	1.12	6.3	0.82	25.3	4.7
5	Oasis Valley	19.28	51.38	178	0.45	1.77	1.48	0.51	1.04	1.13	6.5	0.82	25.4	4.6

Lab Practical Assignment III

2. Different sedimentary depositional environments may be able to Be “typed” by their characteristic grain size distribution. Krumbein and Aberdeen (1937) tabulated grain-size data from a set of 98 samples collected from different near-shore depositional environments in Barataria Bay, Louisiana (see the Barataria Bay dataset). The problem here is to explore which representation of these data is most informative when it comes to typing these environments using grain-size data, which consist of the percentage sand in each of seven grain-size categories. Compare and contrast the following representations of these data:

   - mean sand grain size,
   - pooled sample PC-1 score,
   - percent sand in each sand grain-size category.

Lab Practical Assignment III

2. Essentially, this is a comparison of alternative dimensionality-reduction strategies. The following questions pertain (mostly) to the PCA part of this analysis.

   i. Decide whether the data need to be transformed prior to analysis. Explain the logic behind your decision.

   ii. Decide what the appropriate basis matrix for your PCA analysis is, given the decision you made above. Explain your logic.

   iii. Decide how many principal components (eigenvectors) are necessary to represent the major features of these data. Explain your logic.

   iv. Do natural grain-size groupings of water bodies exist in these data? Show your plots and the equations of the retained eigenvector axes.

   v. The ratio of the sum of squares among groups (SSA) to the total sum of squares (SST) can be used as a measure of how tightly a set of data is clustered and separated from other groups. Determine which representation is most effective for characterizing the different depositional environments. Show your calculations for each analysis along with your interpretation.

Lab Practical Assignment III

Barataria Bay (Louisiana)

Site	Dep. Environment	φ1-2	φ2-3	φ3-4	φ4-5	φ5-6	φ6-7	φ7-8	φ8-9	φ9-10	φ10-11
1	Beach & Foreshore Sands	0.6	70.2	29.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Beach & Foreshore Sands	1.0	69.9	29.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Beach & Foreshore Sands	0.8	73.7	25.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Beach & Foreshore Sands	0.9	75.3	23.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	Beach & Foreshore Sands	0.6	62.5	36.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	Beach & Foreshore Sands	1.1	68.8	30.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	Silty Channel Margin Sands	0.6	5.9	33.0	24.9	9.4	7.8	5.5	5.4	4.4	3.1
8	Organic Muds	1.0	2.3	6.6	16.2	12.0	11.4	13.3	11.0	7.5	18.7
9	Organic Bottom Sands	1.2	1.6	15.3	38.4	13.0	9.5	5.6	5.3	4.2	5.9
10	Silty Channel Sands	9.5	15.8	59.0	8.4	0.9	0.9	1.4	2.3	1.8	0.0

Lab Practical Assignment III

Diameter (mm)	Phi units	Particle Size		Sedimentary Rocks				Settling velocity (cm/s)	Entrainment velocity (cm/s)
					Grain/Clast Composition				
					Quartz >90%	Feldspar > 10% Feldspar > Lithics	Lithics[1] > 10% Lithics > Feldspar		
256	-8	Boulder	Gravel (rounded) or Rubble (angular)	Conglomerate (rounded) or Breccia (angular)	X	X	_____[2] Boulder Conglomerate (Breccia)		500
64	-6	Cobble			Quartz Cobble Conglomerate (Breccia)	X	_____[2] Cobble Conglomerate (Breccia)	163	175
4	-2	Pebble			Quartz Pebble Conglomerate (Breccia)	X	_____[2] Pebble Conglomerate (Breccia)	43	60
2	-1	Granule			Quartz Granule Conglomerate (Breccia)	Feldspathic Granule Conglomerate (Breccia)	_____[2] Granule Conglomerate (Breccia)	26	50
1	0	Very coarse	Sand	Sandstone	Quartz Arenite (<10% matrix) or Quartz Wacke[4] (>10% matrix)	Feldspathic Arenite[3] (<10% matrix) or Feldspathic Wacke[3,4] (>10% matrix)	Lithic Arenite (<10% matrix) or Lithic Wacke[4] (>10% matrix)	15	38
1/2	1	Coarse						7	28
1/4	2	Medium						3	23
1/8	3	Fine						1	20
1/16	4	Very fine						0.3	Cohesion may become important and entrainment velocities may increase
1/256	8	Silt	Mud	Mudrock (massive) or Shale (fissile)	Siltstone (silt > clay) Mudstone (silt ≈ mud) Claystone (silt < clay)			0.08	
		Clay							

X = clasts of this size and composition are unlikely.
[1]Lithic fragments could be composed of chert, limestone, igneous, or metamorphic rock fragments (and many others).
[2]Lithologic descriptor of most abundant clast type should be used in the blank space (ex: basalt cobble conglomerate); use the term polymictic if no one lithology is dominant (ex: polymictic pebble conglomerate).
[3]Feldspar-rich sandstones are informally termed arkose.
[4]Sandstones with abundant muddy matric are informally called graywackes.

Table from Rygel & Quinton (2026)

Lab Practical Assignment III

3. A collection of ten trilobite taxa has been made from localities in the Great Basin for the purpose of determining whether these species can be identified using simple length measurements of the cephalon and pygidium. There data are provided in the Great Basin Trilobites Dataset.

Variable	Description
x1	Border length
x2	Brim length
x3	Palpebral lobe length
x4	Glabella width
x5	Fixed cheek length
x6	Genal spine length/free cheek length
x7	Pygidial axis width
x8	Pleural axis width
x9	Pygidial axis length
x10	Pygidium length

In order to standardize for size differences among individuals all cephalon lengths were divided by the length of the glabella and all pygidum lengths were divided by the width of the pygidium.

Lab Practical Assignment III

Great Basin Trilobites (Utah)

N	Species	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1	*Aphelaspis brachyphasis*	0.208	0.250	0.542	0.237	0.875	0.292	0.284	0.925	0.343	0.373
2	*Aphelaspis haguei*	0.318	0.318	0.545	0.428	1.000	0.318	0.296	0.796	0.444	0.537
3	*Aphelaspis subditus*	0.174	0.304	0.391	0.375	0.913	0.304	0.297	0.946	0.405	0.486
4	*Dicanthopyge convergens*	0.259	0.370	0.370	0.859	0.852	0.333	0.500	0.591	0.591	0.818
5	*Dicanthopyge quadrata*	0.250	0.350	0.500	0.615	0.900	0.351	0.434	0.783	0.478	0.652
6	*Dicanthopyge reductus*	0.316	0.421	0.474	0.736	1.158	0.421	0.500	0.675	0.500	0.775
7	*Prehousia alata*	0.136	0.409	0.273	0.469	1.000	0.136	0.269	0.769	0.327	0.423
8	*Prehousia indenta*	0.192	0.308	0.269	0.628	0.923	0.154	0.308	0.795	0.308	0.436
9	*Prehousia prima*	0.261	0.261	0.261	0.545	0.956	0.261	0.296	0.833	0.333	0.407
10	*Prehousia longispina*	0.259	0.370	0.556	0.444	0.852	0.296	0.372	0.824	0.431	0.706

Lab Practical Assignment III

3. Use multivariate dimensionality reduction methods to summarize these data and determine whether representation of the associations between variables or distances between individuals performs best in recovering the standard taxonomy.

   i. Decide whether the data need to be transformed prior to analysis. Explain the logic behind your decision.

   ii. Decide what the appropriate basis matrices are for your analysis. Explain your logic.

   iii. Decide how many eigenvectors are necessary to represent the major features of these data. Explain your logic.

   iv. Identify the variables and species most closely aligned with each of your eigenvector axes.

   v. Show the comparative ordination plots that support your (written) interpretation.

Lab Practical Assignment III

4. The Swedish government was interested in evaluating the potential for exploiting heavy mineral deposits in the heavily forested northern part of that country. As airborne magnetometer surveys proved to be of limited value a program of ground sampling from stream sediments was undertaken. The Sweden Sediments HM dataset contains measurements (in ppm) of 13 heavy-metal concentrations for 46 locality samples divided into three groups: 1 - samples from known productive regions, 2 - samples from known non-productive regions and 3 - samples from regions whose productivity was unknown. Perform a multivariate discriminant analysis on the data from groups 1 and 2.

   i. Decide whether the data need to be transformed prior to analysis. Explain the logic behind your decision.

   ii. Identify the variables most closely aligned with your discriminant function.

   iii. Show the comparative ordination plots that support your (written) interpretation.

   iv. Test your results for statistical significance.

   v. Use your result to predict the potential for mining operations to be successful in the regions of unknown productivity.

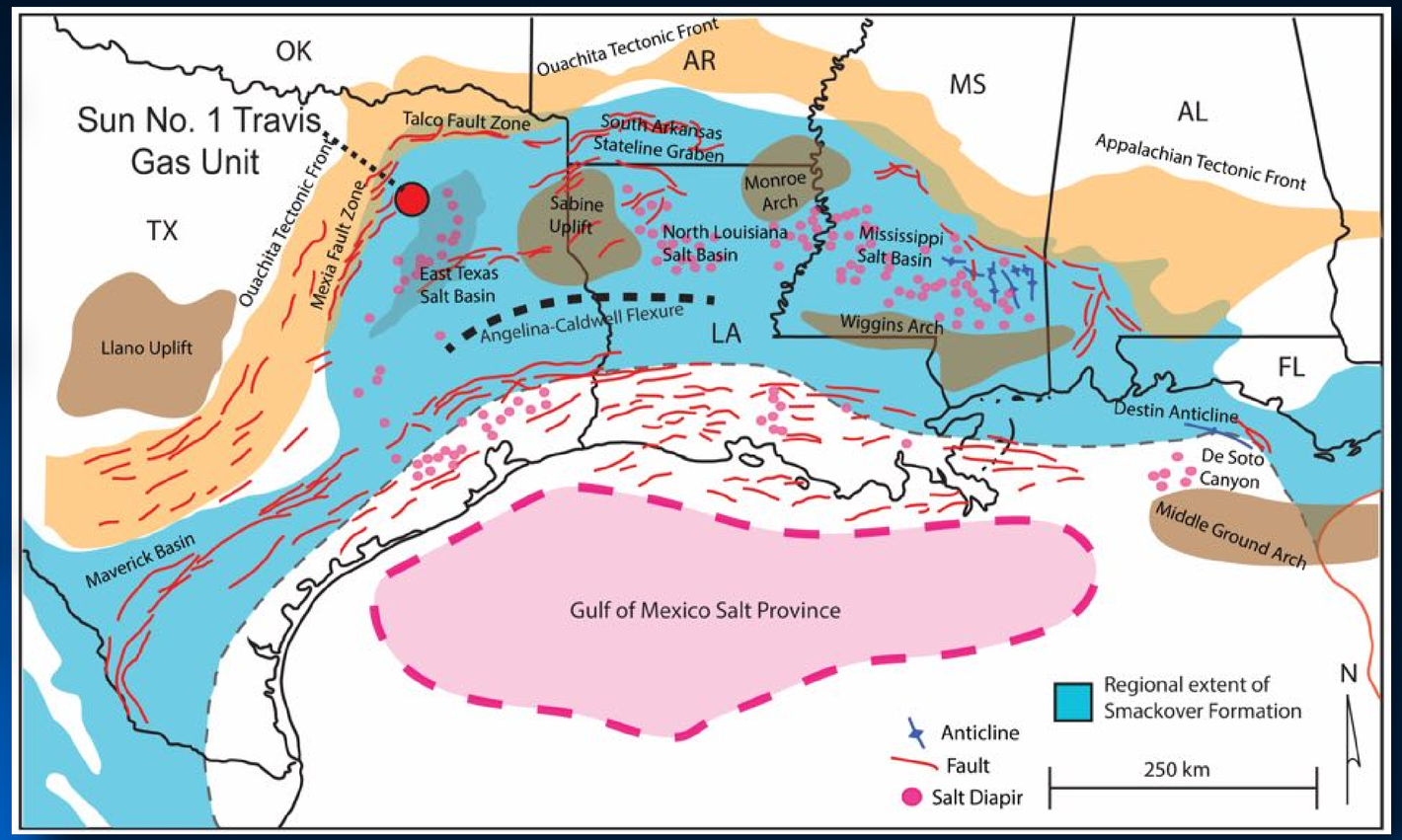
Lab Practical Assignment III

Sweden Sediments HM

Sample	Group	Ti	Mn	Ag	Ba	Co	Cr	Cu	Ni	Pb	Sr	V	Zn	Au
1	Non-Productive	7280	1300	30	720	30	150	73	50	70	60	70	190	0.02
2	Non-Productive	10300	1200	0.7	1280	20	160	25	50	70	90	50	50	0.02
3	Non-Productive	7000	1500	0.7	760	30	160	70	40	110	240	40	250	0.01
4	Non-Productive	5100	1000	0.5	740	20	140	39	50	80	50	60	130	0.02
5	Non-Productive	10600	2100	0.3	980	30	50	25	30	70	150	160	110	0.01
6	Productive	10600	2900	0.4	1640	25	20	21	30	30	320	90	200	0.01
7	Productive	11500	3200	0.7	710	30	30	15	20	20	260	270	180	0.01
8	Productive	7100	1800	0.9	490	75	50	8	10	30	80	180	100	0.02
9	Unknown	3100	500	0.2	100	15	30	65	10	20	710	100	40	0.01
10	Unknown	6500	700	1	1070	20	200	48	70	100	210	50	170	0.01

Lab Practical Assignment III

5. The carbonate sediments of the Upper Jurassic Smackover Fm. on the eastern Gulf Coast of the USA are extensively dolomitized with the source of dolomitization varying regionally. Among the known mechanisms are seawater seepage (SS), reflux of evaporite brines (EB), shallow burial alteration by freshwater seepage (SF), shallow burial alteration by mixed connate and freshwater (SM) and deep burial (secondary) alteration (DB). The Smackover Dolomites dataset contains a variety of isotopic, elemental and physical variables collected from 97 samples for which the dominant mechanism of dolomitization has been established based on independent criteria.



However, since these variables can be assessed from sediment cores we would like to know the extent to which these variables reflect different dolomitization environments. Show the comparative ordination plots that support your (written) interpretation.

Lab Practical Assignment III

Smackover Dolomites

Sample	Deposiitonal Envisonment	δO^{18}	δC^{13}	Fe (ppm)	Mg (%)	Na (ppm)	Mn (ppm)	Sr (ppm)	Ca (%)	Porosity (%)	Permeability (md)
1	EB	0.27	5.59	4818	10.27	673	761	98	18.2	3.1	0
2	DB	-7.53	3.95	8461	10.44	388	3214	68	19.09	2.8	0
3	DB	-7.7	3.98	6907	10.78	590	2837	55	19.22	3.7	0
4	DB	-6.17	4.92	3300	11.98	1600	1500	50	21.18	22.2	417
5	DB	-6.37	4.8	4500	10.8	1750	90	20	20.22	19.3	107
6	SM	-5.24	4.75	1611	11.71	785	1752	81	21.39	8.7	0.2
7	SM	-4.01	4.74	787	11.86	6658	930	69	21.17	12.5	2.38
8	SM	-3.2	4.95	3904	11.28	613	635	172	20.88	7.9	0
9	SM	-3.94	5.11	2666	11.39	288	968	62	20.31	2.8	0
10	SM	-5.66	5.04	3877	10.35	322	880	239	20.79	3.6	0

Lab Practical Assignment III

4. Use multivariate dimensionality reduction analysis to assess the potential use of these measurements'

   i. Decide whether these data need to be transformed prior to analysis. Explain the logic behind your decision.

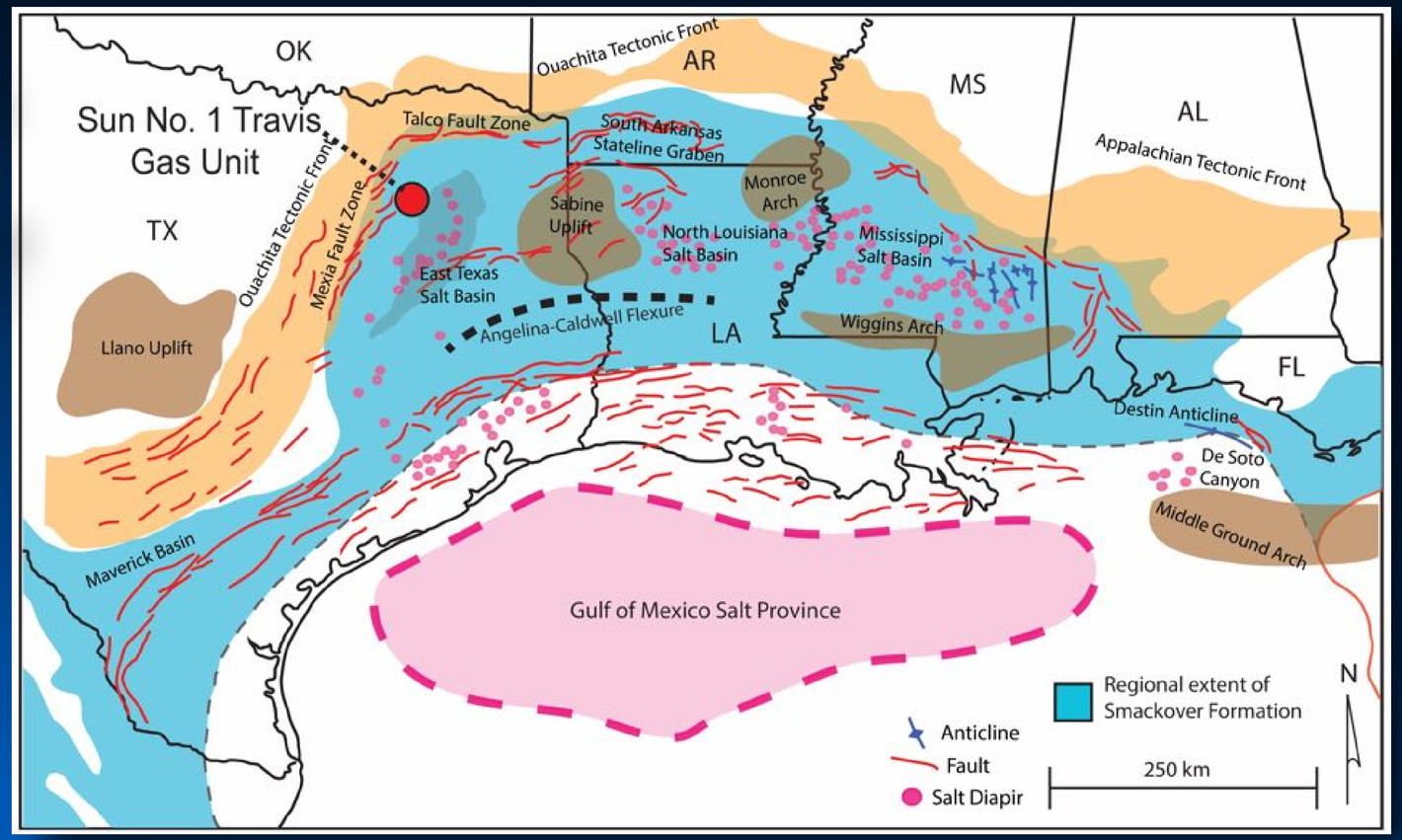
   ii. Select an appropriate data-analysis method for use in assessing these data. Justify your decision.

   iii. Decide how many dimensions are necessary to represent the major features of these data. Explain your logic.

   iv. Identify the variables most indicative of each dolomitization mechanism.

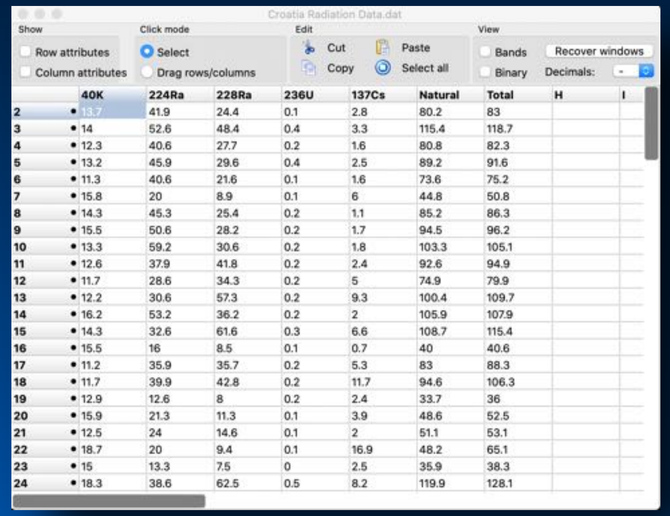
   v. Show the ordination plot(s) that support your (written) interpretation.

   vi. If appropriate, test the reliability and/or statistical significance of your results.

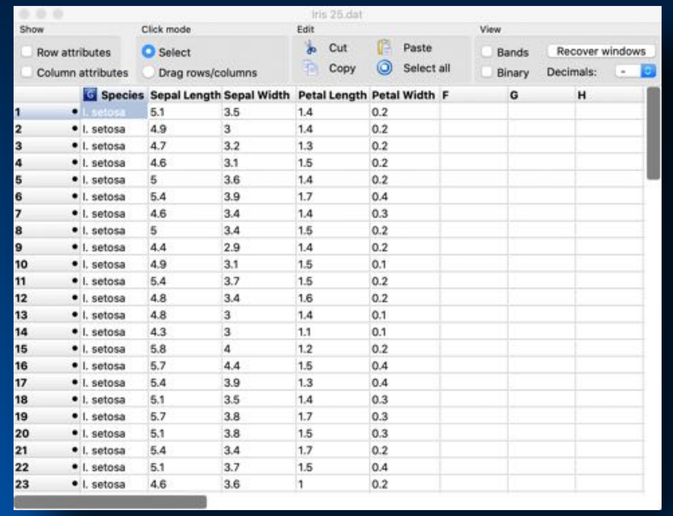


Lab Practical Assignment III

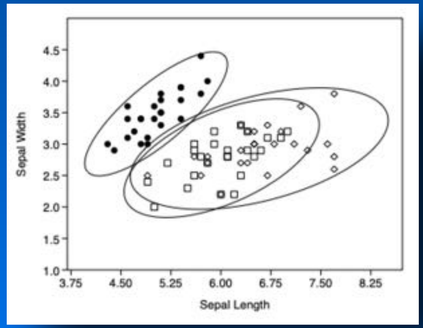
Multivariate Data Analysis & Dimensionality Reduction In PAST

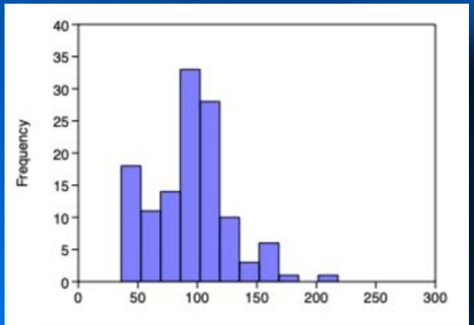


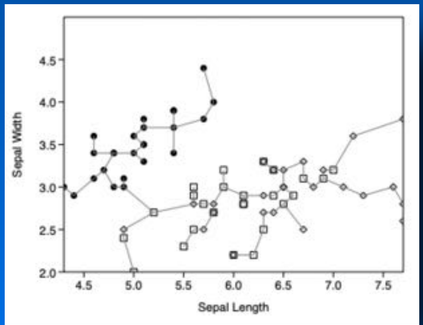
	40K	224Ra	228Ra	236U	137Cs	Natural	Total	H	I
2	13.7	41.9	24.4	0.1	2.8	80.2	83		
3	14	52.6	48.4	0.4	3.3	115.4	118.7		
4	12.3	40.6	27.7	0.2	1.6	80.8	82.3		
5	13.2	45.9	29.6	0.4	2.5	89.2	91.6		
6	11.3	40.6	21.6	0.1	1.6	73.6	75.2		
7	15.8	20	8.9	0.1	6	44.8	50.8		
8	14.3	45.3	25.4	0.2	1.1	85.2	86.3		
9	15.5	50.6	28.2	0.2	1.7	94.5	96.2		
10	13.3	59.2	30.6	0.2	1.8	103.3	105.1		
11	12.6	37.9	41.8	0.2	2.4	92.6	94.9		
12	11.7	28.6	34.3	0.2	5	74.9	79.9		
13	12.2	30.6	57.3	0.2	9.3	100.4	109.7		
14	16.2	53.2	36.2	0.2	2	105.9	107.9		
15	14.3	32.6	61.6	0.3	6.6	108.7	115.4		
16	15.5	16	8.5	0.1	0.7	40	40.6		
17	11.2	35.9	35.7	0.2	5.3	83	88.3		
18	11.7	39.9	42.8	0.2	11.7	94.6	106.3		
19	12.9	12.6	8	0.2	2.4	33.7	36		
20	15.9	21.3	11.3	0.1	3.9	48.6	52.5		
21	12.5	24	14.6	0.1	2	51.1	53.1		
22	18.7	20	9.4	0.1	16.9	48.2	65.1		
23	15	13.3	7.5	0	2.5	35.9	38.3		
24	18.3	38.6	62.5	0.5	8.2	119.9	128.1		



	Species	Sepal Length	Sepal Width	Petal Length	Petal Width	F	G	H
1	I. setosa	5.1	3.5	1.4	0.2			
2	I. setosa	4.9	3	1.4	0.2			
3	I. setosa	4.7	3.2	1.3	0.2			
4	I. setosa	4.6	3.1	1.5	0.2			
5	I. setosa	5	3.6	1.4	0.2			
6	I. setosa	5.4	3.9	1.7	0.4			
7	I. setosa	4.6	3.4	1.4	0.3			
8	I. setosa	5	3.4	1.5	0.2			
9	I. setosa	4.4	2.9	1.4	0.2			
10	I. setosa	4.9	3.1	1.5	0.1			
11	I. setosa	5.4	3.7	1.5	0.2			
12	I. setosa	4.8	3.4	1.6	0.2			
13	I. setosa	4.8	3	1.4	0.1			
14	I. setosa	4.3	3	1.1	0.1			
15	I. setosa	5.8	4	1.2	0.2			
16	I. setosa	5.7	4.4	1.5	0.4			
17	I. setosa	5.4	3.9	1.3	0.4			
18	I. setosa	5.1	3.5	1.4	0.3			
19	I. setosa	5.7	3.8	1.7	0.3			
20	I. setosa	5.1	3.8	1.5	0.3			
21	I. setosa	5.4	3.4	1.7	0.2			
22	I. setosa	5.1	3.7	1.5	0.4			
23	I. setosa	4.6	3.6	1	0.2			

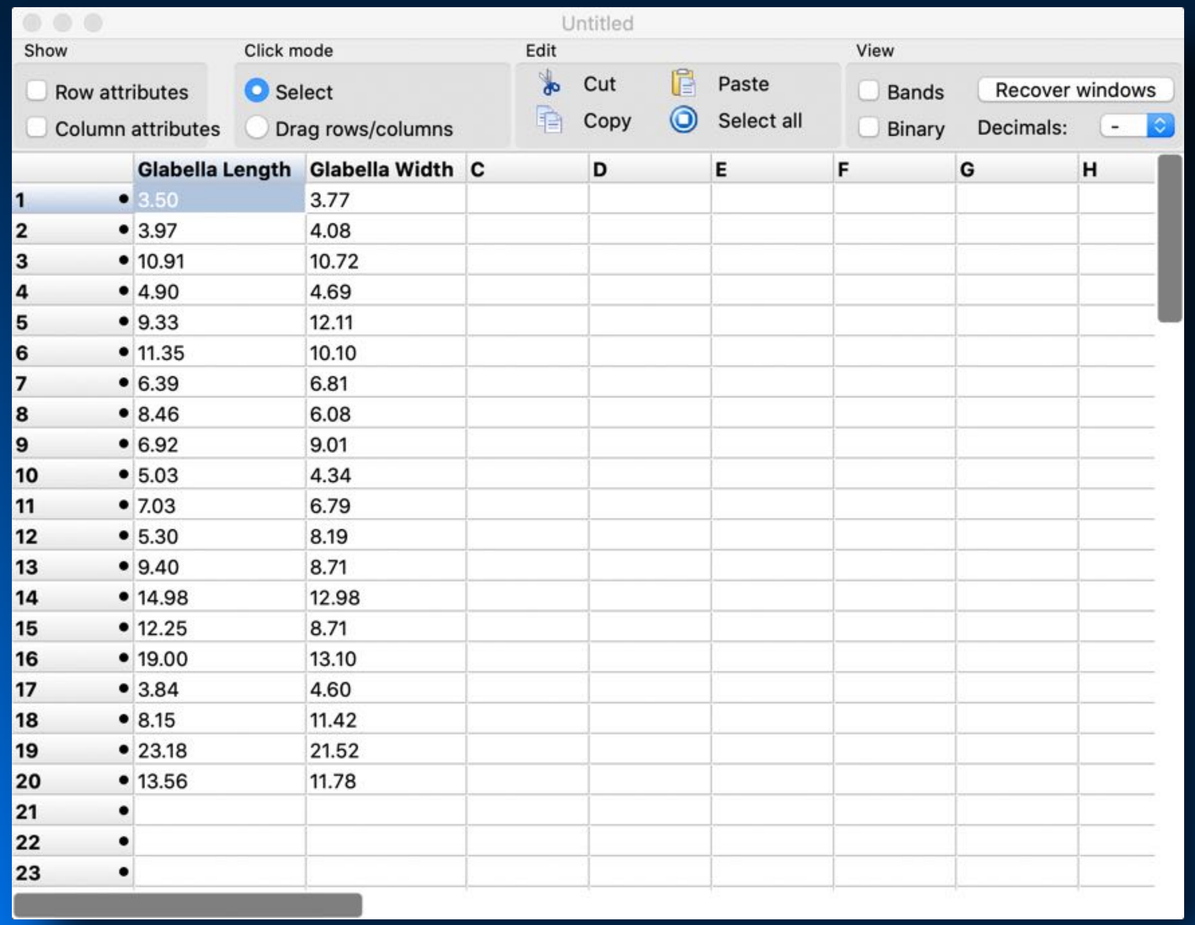






Lab Practical Assignment III

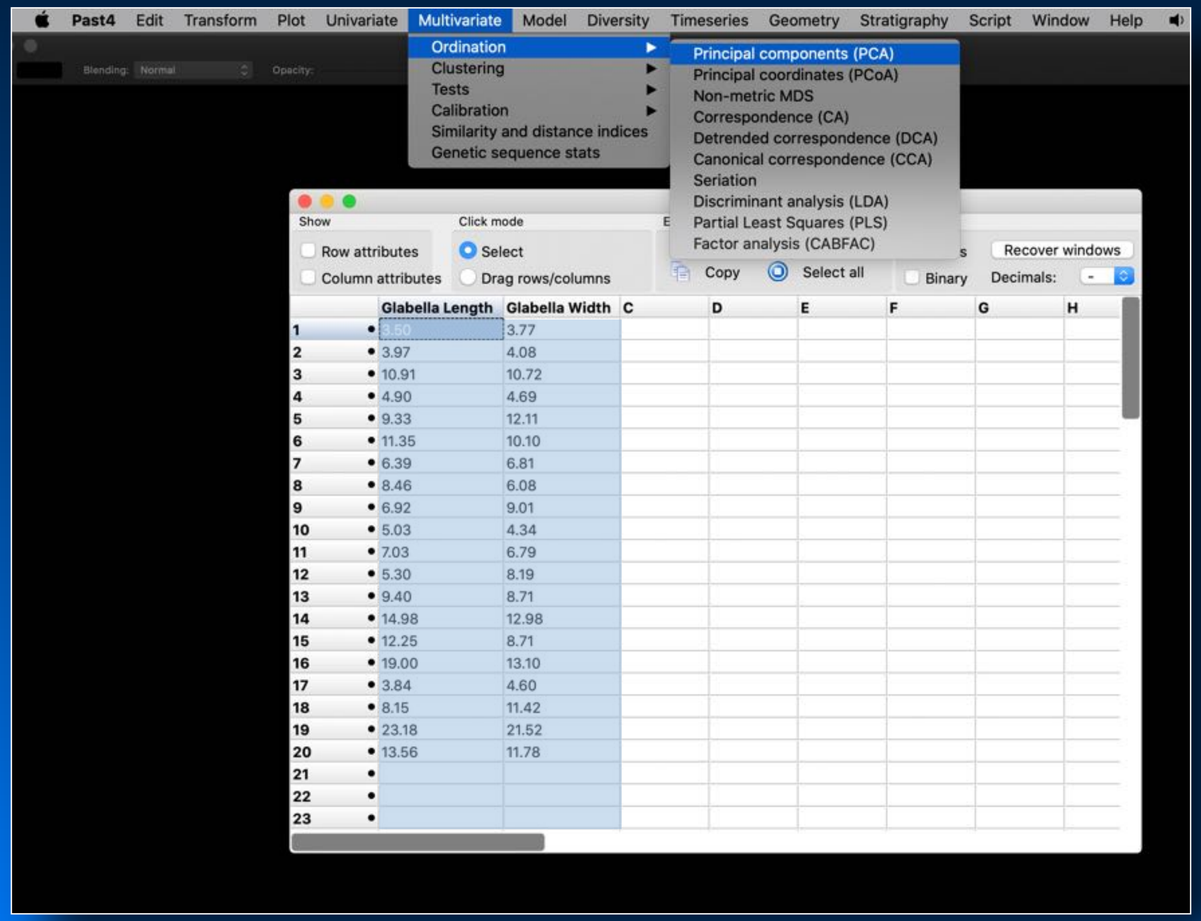
Principal Component Analysis (PCA) in PAST



	Glabella Length	Glabella Width	C	D	E	F	G	H
1	3.50	3.77						
2	3.97	4.08						
3	10.91	10.72						
4	4.90	4.69						
5	9.33	12.11						
6	11.35	10.10						
7	6.39	6.81						
8	8.46	6.08						
9	6.92	9.01						
10	5.03	4.34						
11	7.03	6.79						
12	5.30	8.19						
13	9.40	8.71						
14	14.98	12.98						
15	12.25	8.71						
16	19.00	13.10						
17	3.84	4.60						
18	8.15	11.42						
19	23.18	21.52						
20	13.56	11.78						
21								
22								
23								

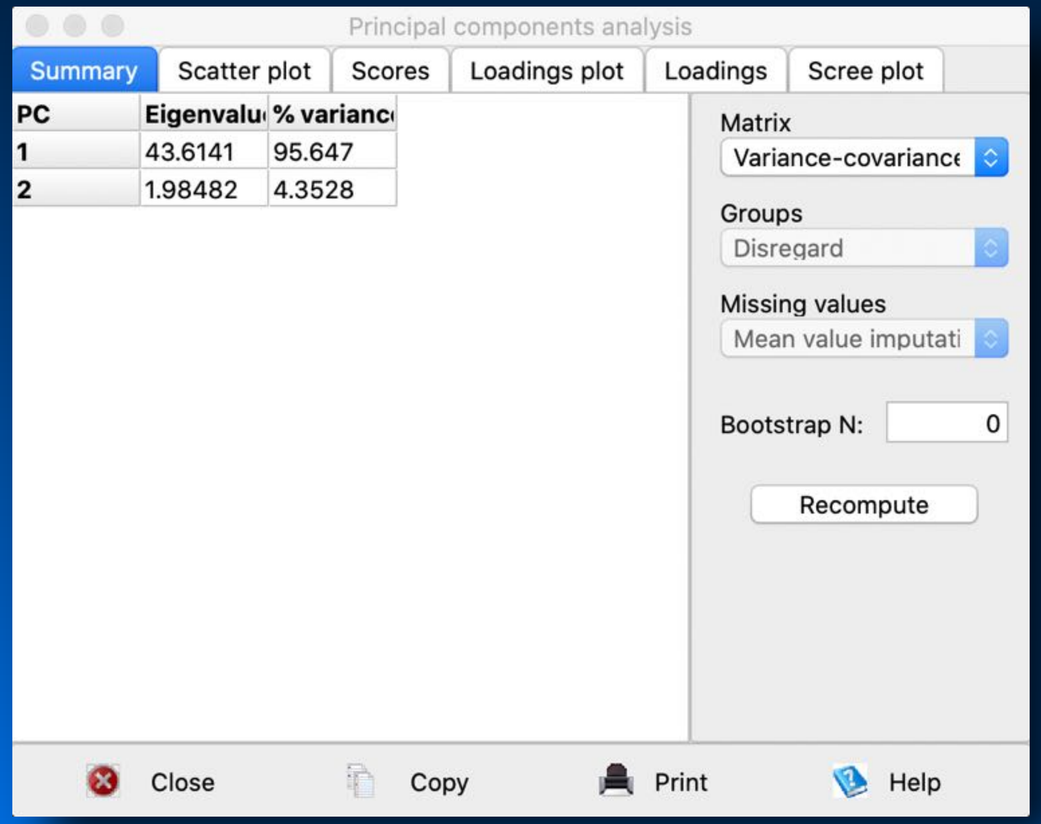
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



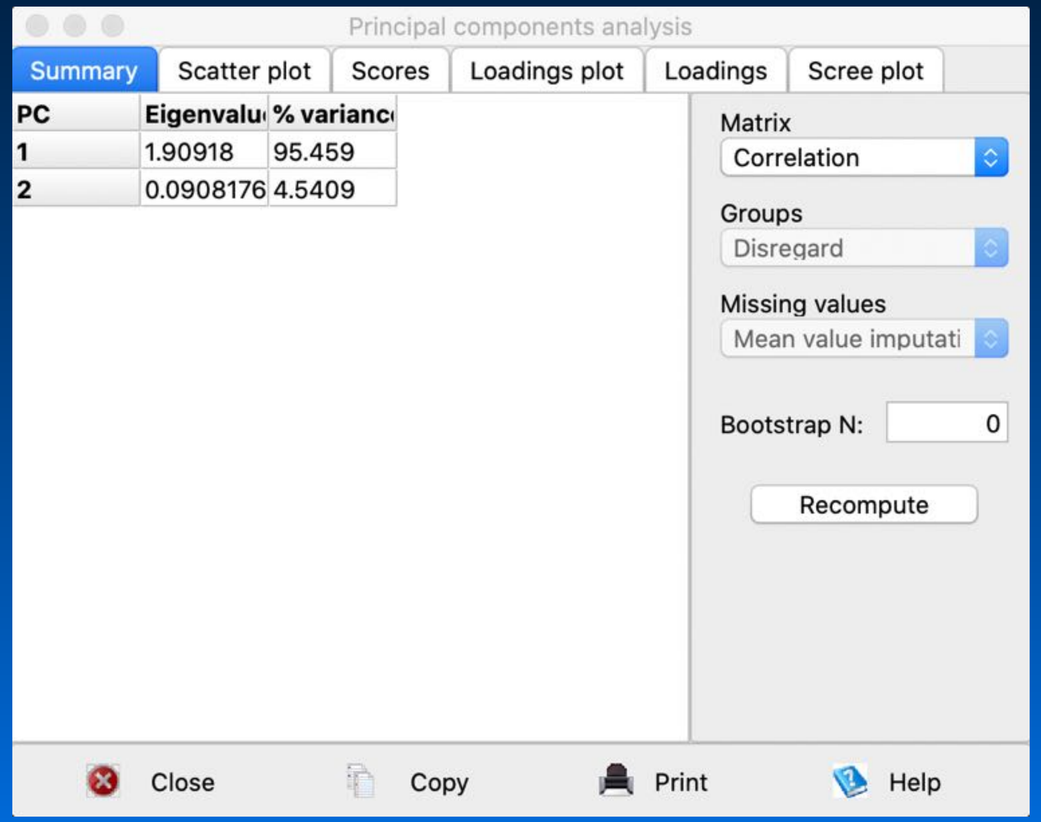
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



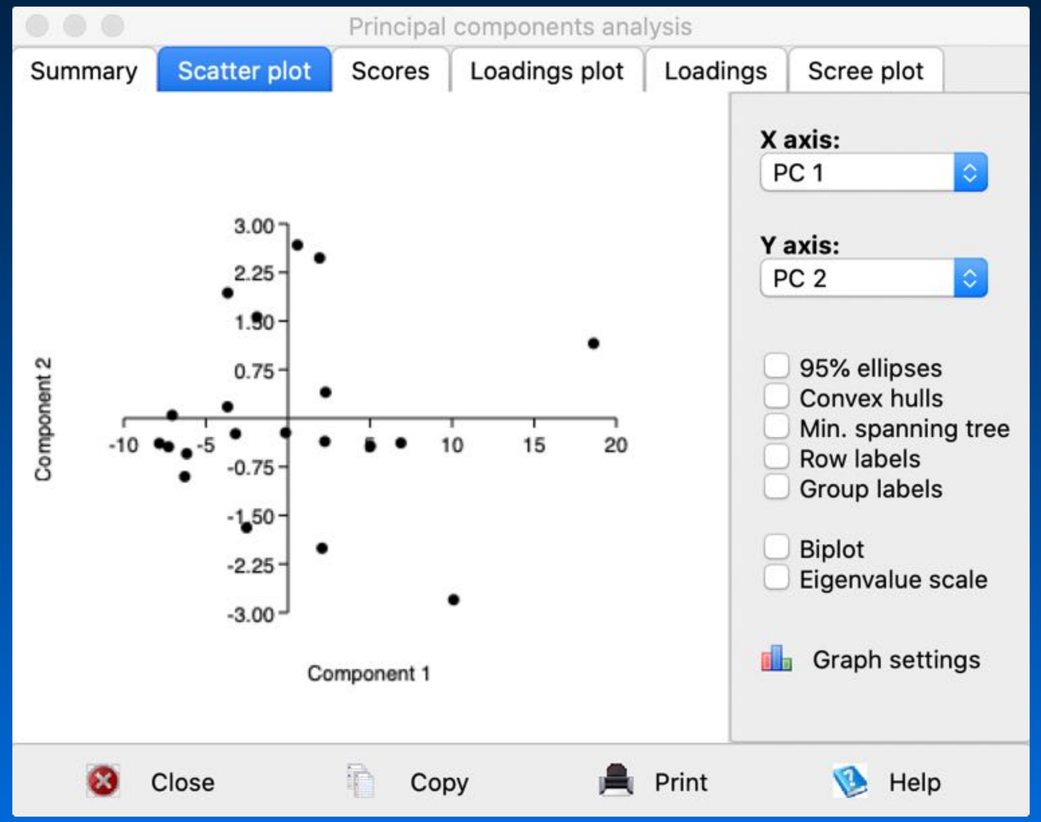
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



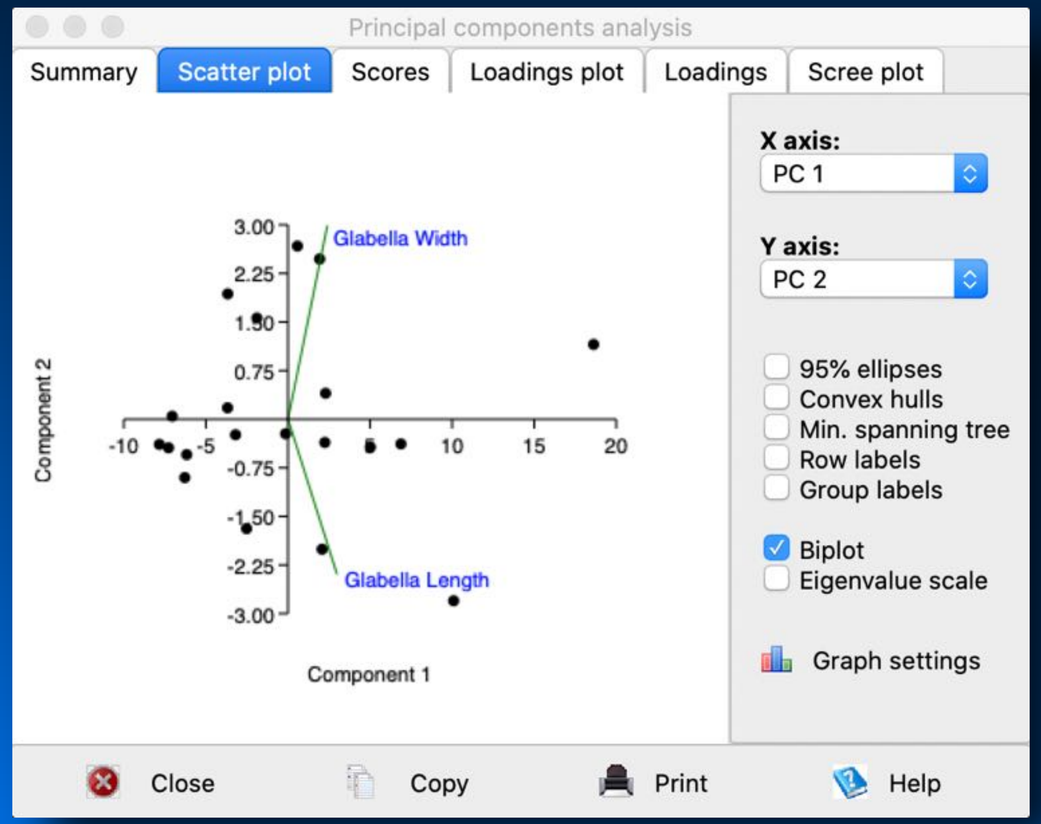
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



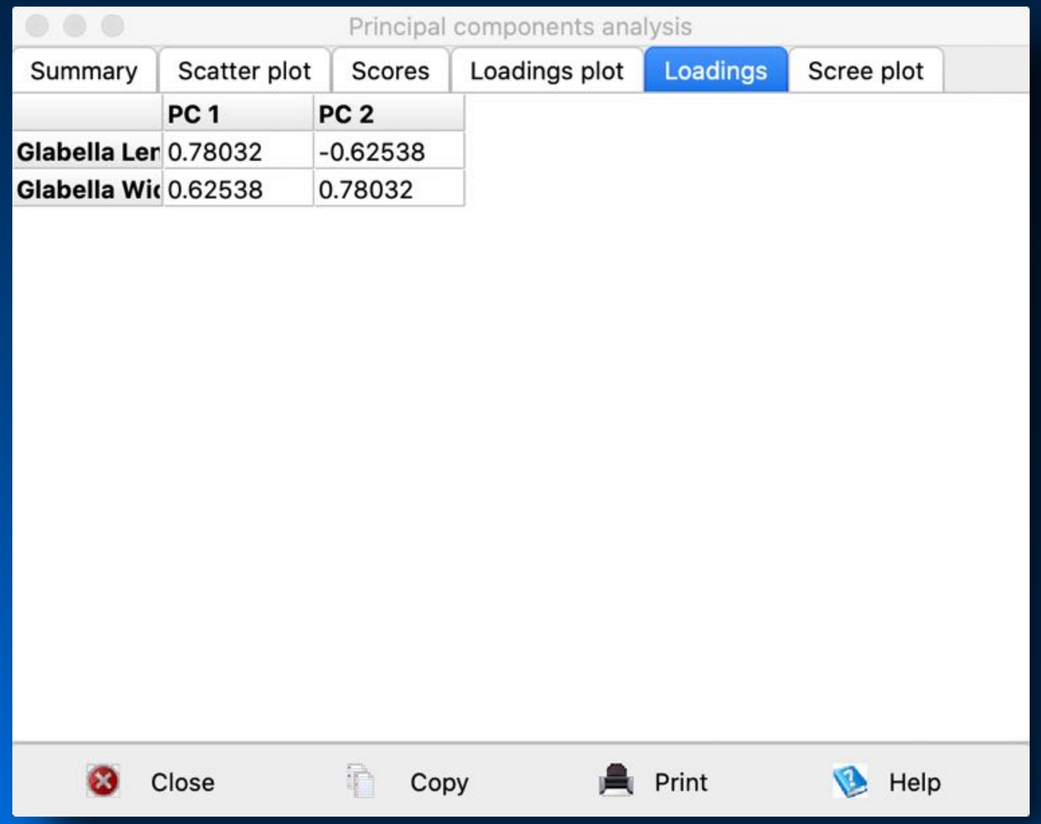
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



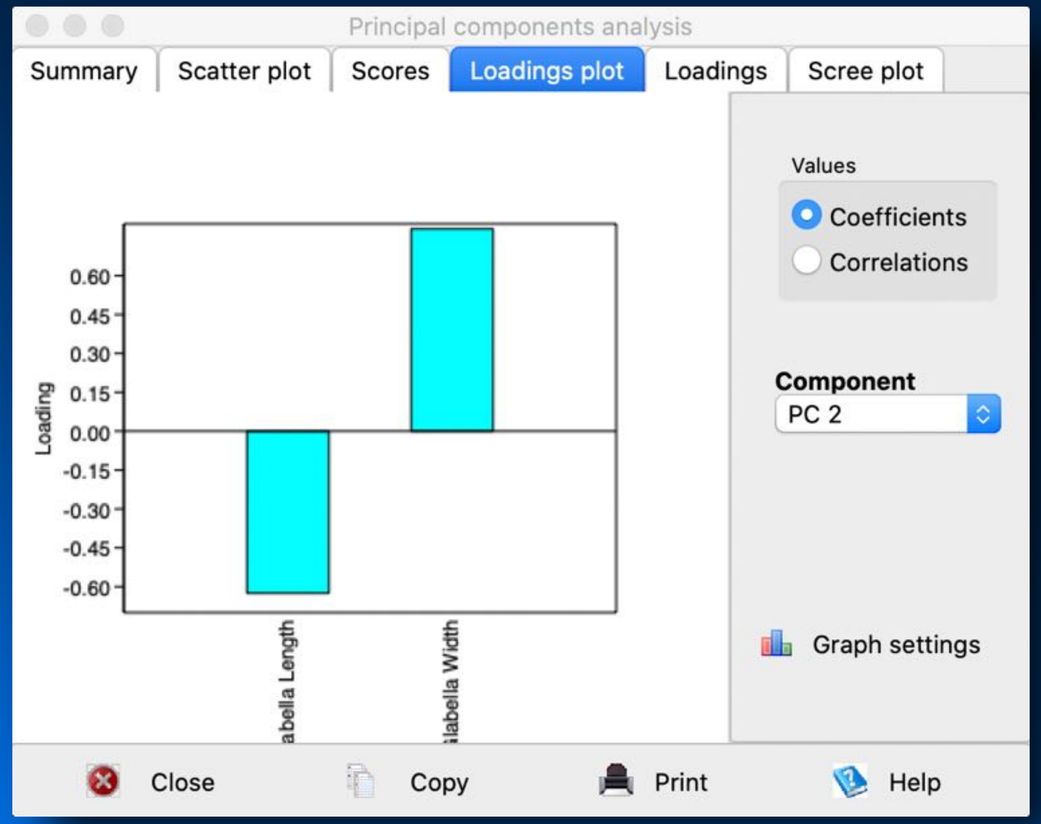
	PC 1	PC 2
Glabella Ler	0.78032	-0.62538
Glabella Wic	0.62538	0.78032

Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST

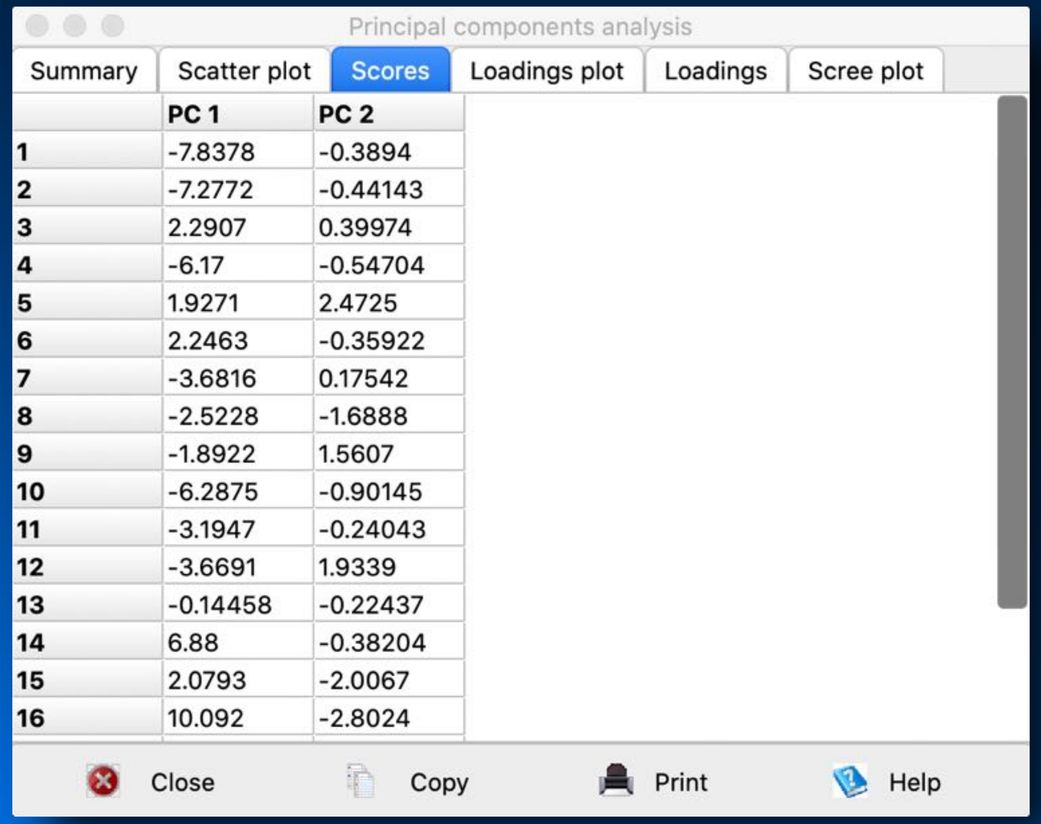
Lab Practical Assignment III

Principal Component Analysis (PCA) in PAST



Lab Practical Assignment III

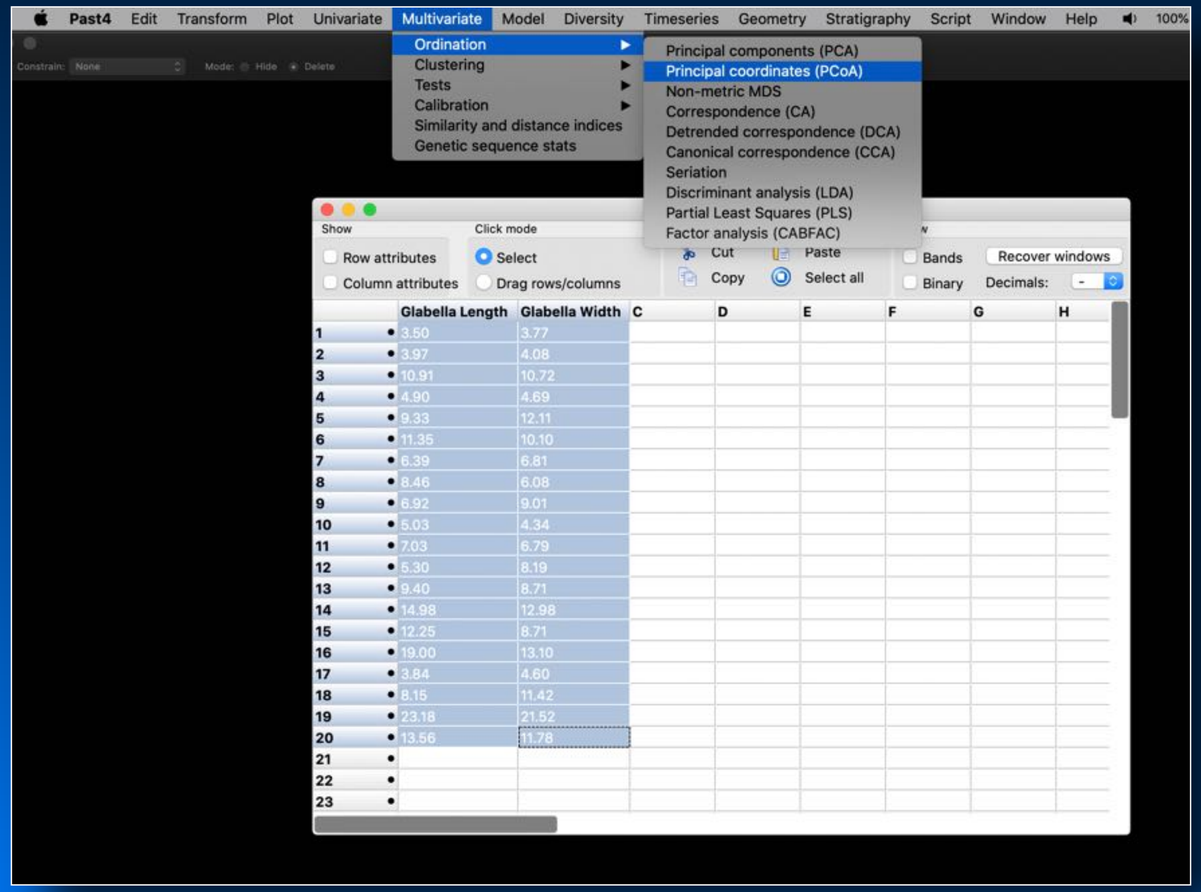
Principal Component Analysis (PCA) in PAST



	PC 1	PC 2
1	-7.8378	-0.3894
2	-7.2772	-0.44143
3	2.2907	0.39974
4	-6.17	-0.54704
5	1.9271	2.4725
6	2.2463	-0.35922
7	-3.6816	0.17542
8	-2.5228	-1.6888
9	-1.8922	1.5607
10	-6.2875	-0.90145
11	-3.1947	-0.24043
12	-3.6691	1.9339
13	-0.14458	-0.22437
14	6.88	-0.38204
15	2.0793	-2.0067
16	10.092	-2.8024

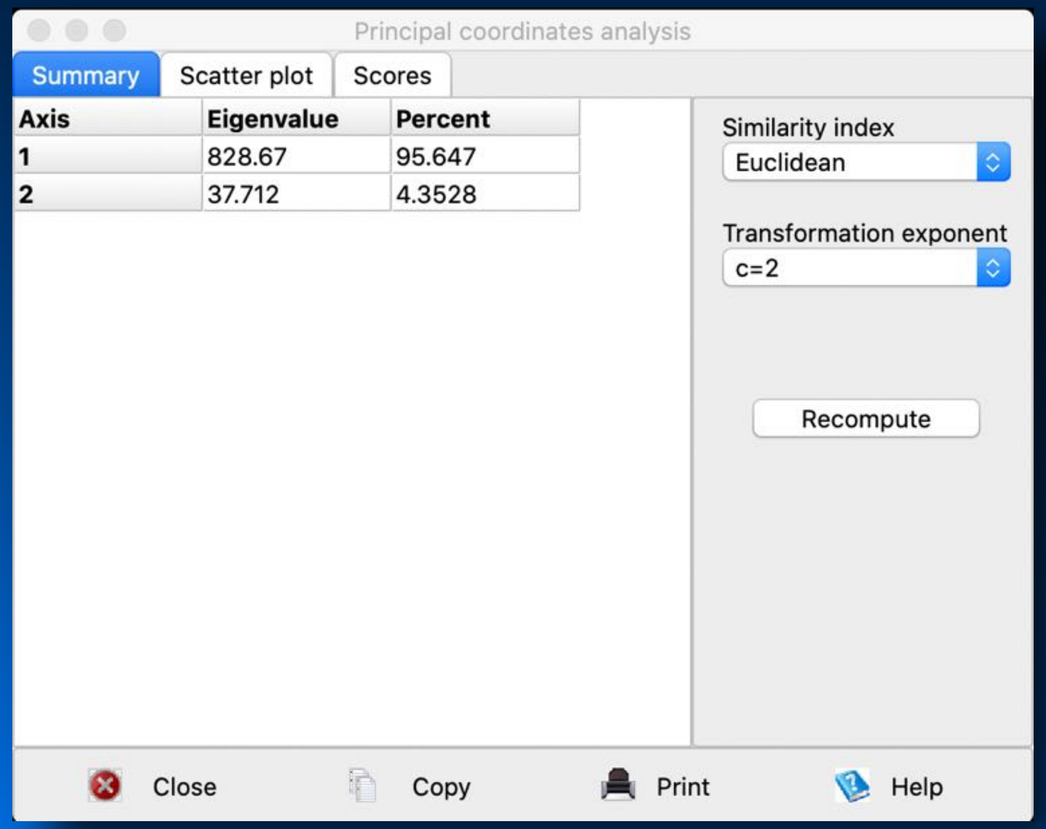
Lab Practical Assignment III

Principal Coordinates Analysis (PCA) in PAST



Lab Practical Assignment III

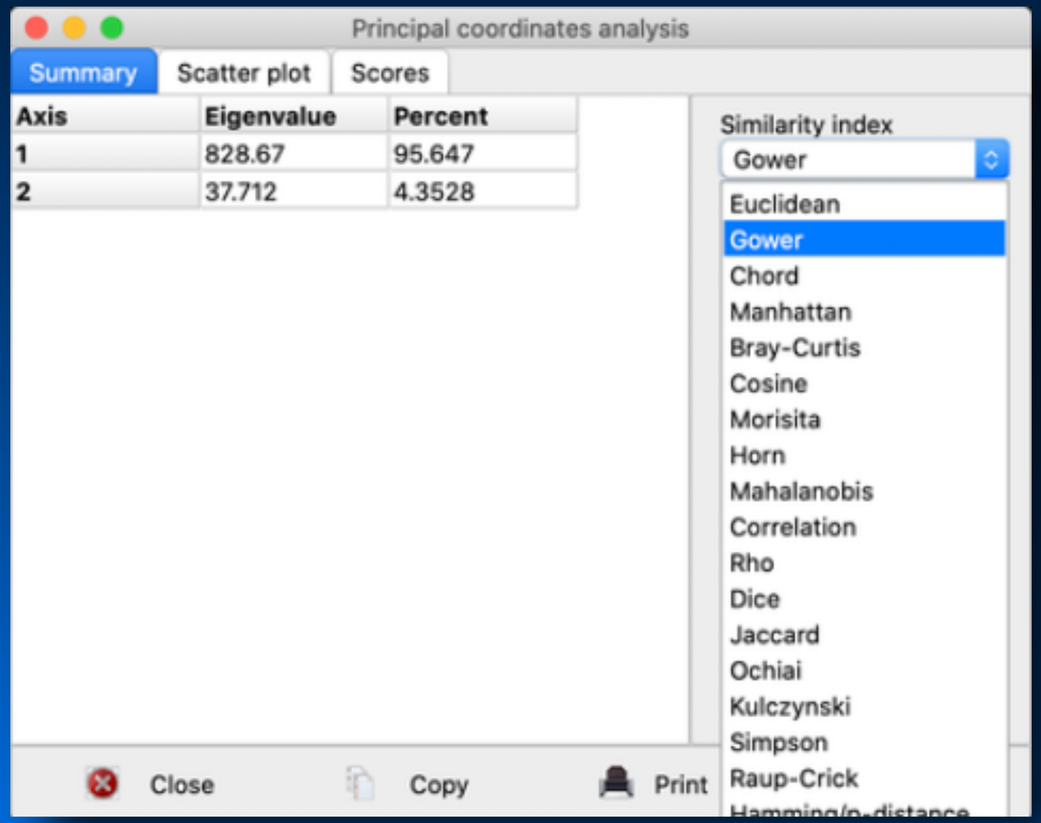
Principal Coordinates Analysis (PCA) in PAST



Axis	Eigenvalue	Percent
1	828.67	95.647
2	37.712	4.3528

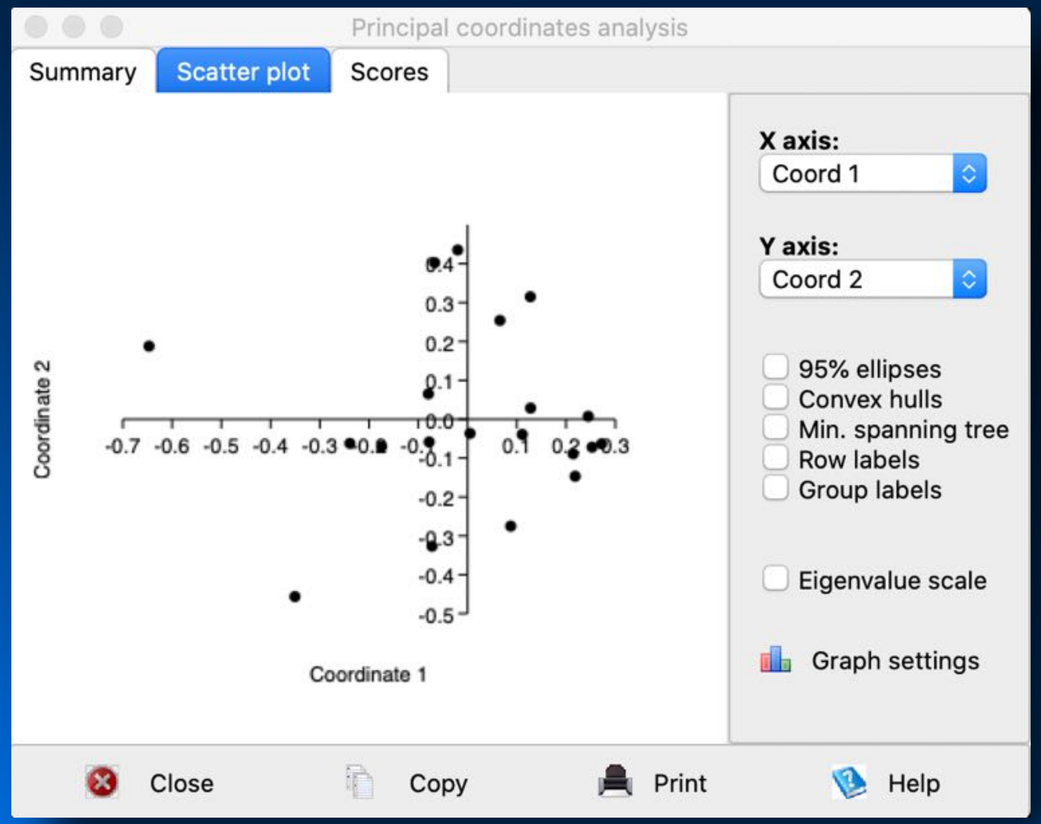
Lab Practical Assignment III

Principal Coordinates Analysis (PCA) in PAST



Lab Practical Assignment III

Principal Coordinates Analysis (PCA) in PAST



Lab Practical Assignment III

Principal Coordinates Analysis (PCA) in PAST

Principal coordinates analysis

Summary | Scatter plot | Scores

	Coord 1	Coord 2
1	0.27227	-0.063411
2	0.2528	-0.071883
3	-0.079576	0.065095
4	0.21434	-0.089081
5	-0.066944	0.40262
6	-0.078034	-0.058496
7	0.12789	0.028565
8	0.087639	-0.275
9	0.06573	0.25414
10	0.21842	-0.14679
11	0.11098	-0.039152
12	0.12746	0.31492
13	0.0050225	-0.036537
14	-0.239	-0.062211
15	-0.072233	-0.32677
16	-0.35058	-0.45635

Close Copy Print Help

Lab Practical Assignment III

Multivariate Data Analysis & Dimensionality Reduction