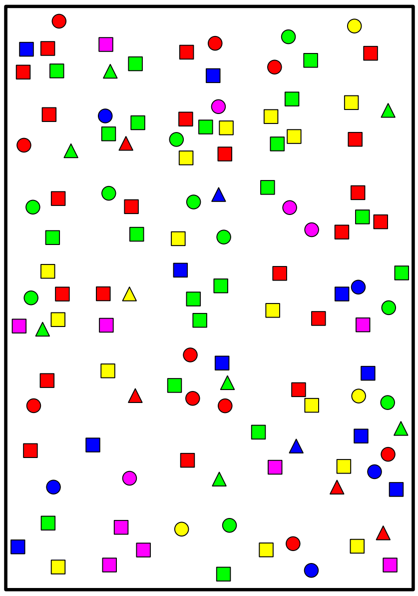
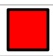
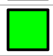
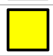
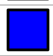
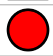
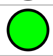
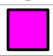
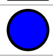
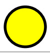
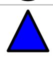
Statistics & Data Analysis in Earth Science

Lab 7 Assignment
PaleoBiodiversity Analysis

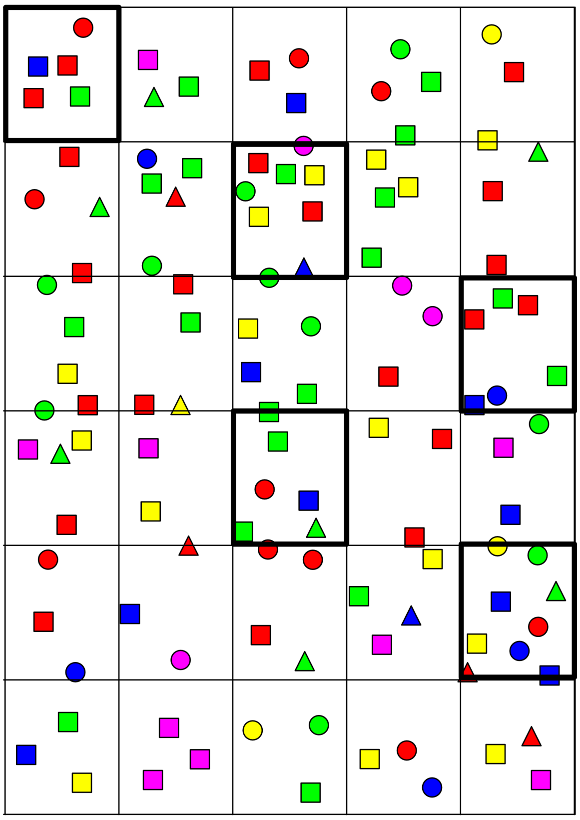


1. The Bedding Plane Map.pdf file shows the positions of fossil taxa (marked with variously colored icons) exposed on the surface of a single bedding plane of a fossil limestone bed. This exposure provides a rare opportunity to describe the preserved diversity of a fossil assemblage in terms of both species richness and relative abundance. Use this map to answer the following questions. (130 points)

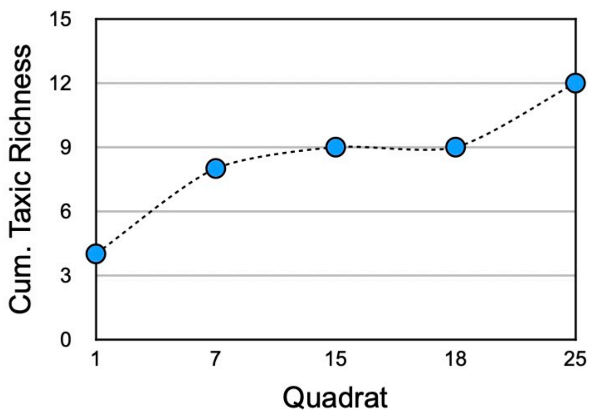
   a. How many different taxa are present on this bedding plane and what are the relative abundances of each? (20 points)

Taxon	Count	Rel. Abundance
	21	0.1736
	20	0.1653
	15	0.1240
	10	0.0826
	10	0.0826
	10	0.0826
	9	0.0744
	7	0.0579
	5	0.0413
	4	0.0331
	4	0.0331
	3	0.0248
	2	0.0165
	1	0.0083
Σ	121	1.0000

   b. Subdivide the map into 30 equal-sized quadrats. Then select five of these quadrats randomly. Show your quadrat-subdivision scheme.

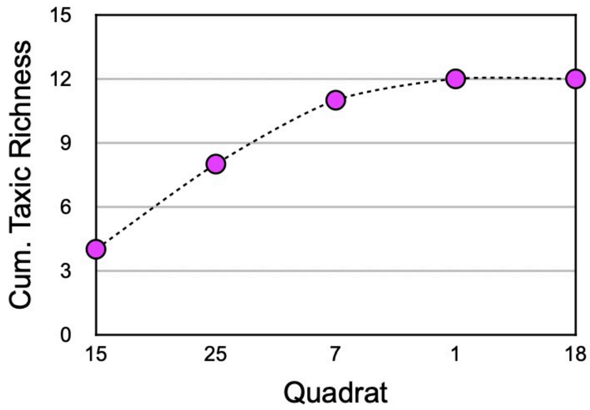


i. Plot the taxon-accumulation curve for these data. Based on this curve what is the estimated taxon richness of the assemblage? Is this estimate stable or does the curve suggest a substantial proportion of taxa have not been recorded? Explain your answers. (30 points)



The taxon accumulation curve has a distinct upward trend at the end of the distribution, suggesting a considerable proportion of taxa are unaccounted for by the quadrat-based survey.

ii. Change the order of quadrats used to construct the taxon-accumulation curve. Does changing the order make any difference to the answers you provided in 1.b.i? Provide an interpretation of your new result. (30 points)



Now taxon accumulation curve appears to be leveling off, indicating (perhaps) a useful proxy estimate for the area's taxon richness might be able to be achieved by adding a few more quadrats.

iii. Can you think of a strategy whereby the estimation of taxon-accumulation curves could be made robust to arbitrary variations induced by the order in which quadrat data was cumulated? [Note: no points will be awarded if your answer to this question is "No."]. (20 points)

Adopt a bootstrap estimation strategy involving repeated estimates based on m quadrats. Add quadrats randomly until the slope between adjacent points levels below a threshold value.

iv. How does your quadrat-based taxon-richness estimate compare with the result obtained in 1.a? Which taxa are missing from your quadrat-based results? Why? What could be done to improve your quadrat-based estimate without simply opting to count everything on the bed? (30 points)

Only two taxa were unrepresented in the cumulative taxon richness curve. This suggests it should be possible to obtain a reasonably accurate taxon-richness estimate by using a relatively small number of randomly selected quadrats, at least for this dataset.

2. The Stratigraphic Section datafile (Stratigraphic Section.csv, Stratigraphic Section.dat) contains taxon frequency-count data for 23 taxa collected from 10 horizons spanning 3 different lithologies (environments). In this exercise we want to compare biodiversity indices to get a sense of whether and/or the extent to which different indices are summarizing the same, or different, sorts of information. (130 points)

   a. Calculate the following diversity indices for each of the samples collected: Dominance, Simpson, Shannon, Evenness, Mehinick, Margalef, Equitability, Fisher-α, Chao-1, iChao-1, ACE, Squares and Hill (q = -1). (10 points)

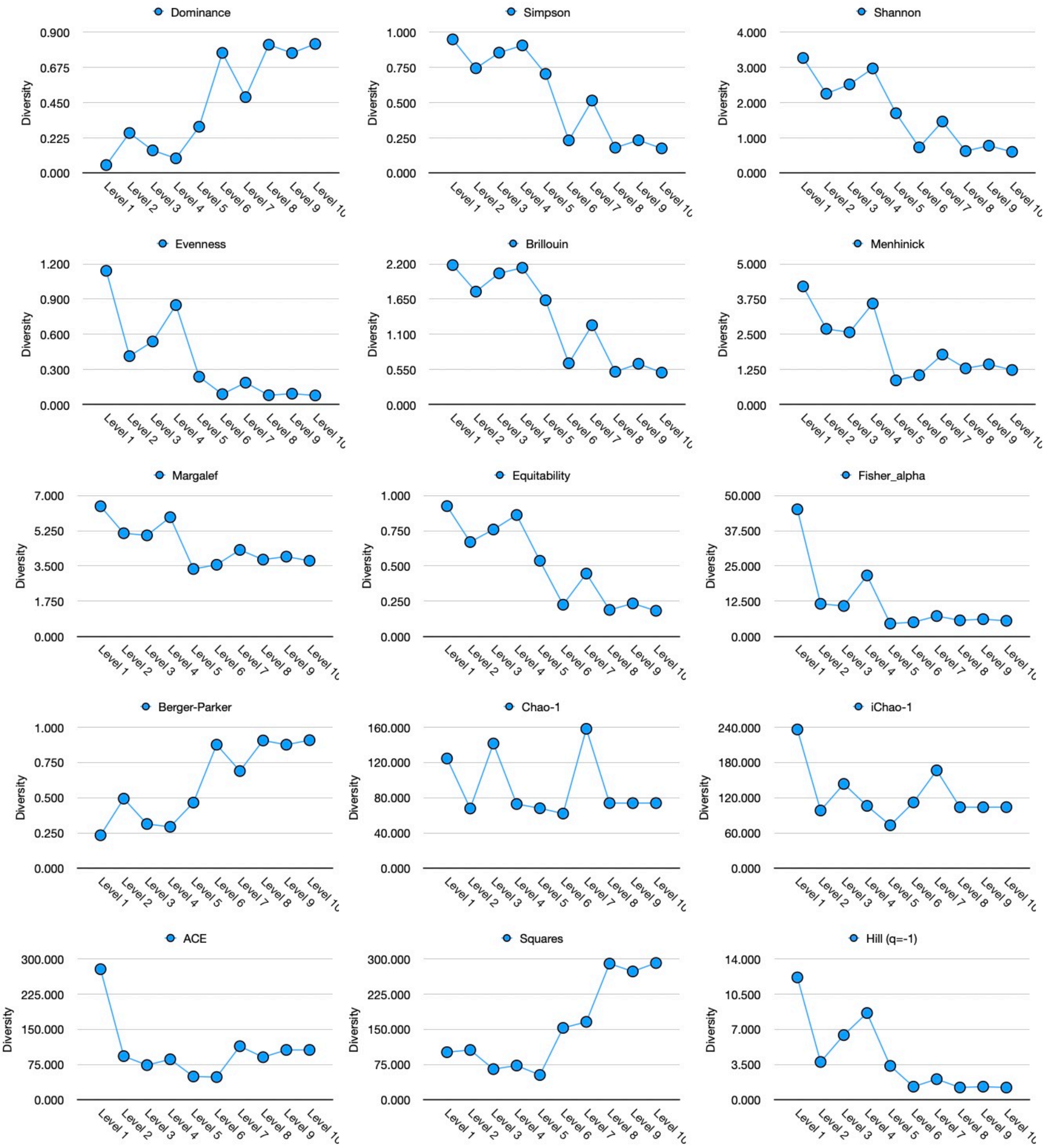
Diversity Indices

Level	Lithology	Dominance	Simpson	Shannon	Evenness	Brillouin	Menhinick
Level 1	Silt	0.051	0.949	3.268	1.141	2.181	4.199
Level 2	Silt	0.256	0.744	2.253	0.414	1.767	2.692
Level 3	Silt	0.145	0.855	2.518	0.539	2.053	2.571
Level 4	Silt	0.094	0.906	2.971	0.849	2.139	3.592
Level 5	Silt	0.296	0.704	1.699	0.238	1.632	0.863
Level 6	Lmst	0.768	0.232	0.729	0.090	0.648	1.048
Level 7	B & C	0.485	0.515	1.463	0.188	1.242	1.780
Level 8	B & C	0.820	0.180	0.624	0.081	0.515	1.290
Level 9	B & C	0.767	0.233	0.776	0.094	0.640	1.435
Level 10	B & C	0.825	0.175	0.603	0.079	0.502	1.233

Diversity Indices (continued)

Level	Margalef	Equitability	Fisher_alpha	Berger-Parker	Chao-1	iChao-1	ACE	Squares	Hill (q=-1)
Level 1	6.468	0.925	45.120	0.233	124.500	236.200	278.100	101.300	12.162
Level 2	5.128	0.670	11.560	0.493	67.710	98.100	92.890	106.000	3.755
Level 3	5.021	0.759	10.800	0.313	141.500	143.400	73.750	65.270	6.426
Level 4	5.924	0.862	21.640	0.293	72.760	106.000	86.120	72.870	8.621
Level 5	3.351	0.537	4.548	0.465	67.940	72.920	49.540	52.710	3.367
Level 6	3.561	0.225	5.030	0.876	61.920	111.900	48.290	153.000	1.302
Level 7	4.299	0.446	7.227	0.689	158.200	166.500	114.000	165.900	2.046
Level 8	3.818	0.188	5.692	0.906	73.840	103.700	90.660	290.000	1.218
Level 9	3.965	0.234	6.114	0.876	73.800	103.700	106.100	273.300	1.303
Level 10	3.759	0.182	5.532	0.908	73.850	103.800	106.100	291.300	1.212

i. Create profile plots for the values of each index across all 10 stratigraphic samples. Which indices (if any) reflect clear distinctions between the different depositional environments? (30 points)



Across almost all indices the limestone (level 6) and lowermost breccia-conglomerate (level 7) samples stand out as representing somewhat unique and contrasting biotas. Silt samples 1–5, and breccia-conglomerate samples 8–10 are often contrasted clearly by various indices (Dominance, Simpson, Shannon, Brillouin, Mehinick, Margalef, Equitability, Berger-Parker, Squares), but less so in others (Evenness, Fisher-α, Chao-1, iChao-1, Ace, Hill$^{q=-1}$). However, the limestone biota's diversity is often placed with the breccia-conglomerate biota's index range whereas the lowermost breccia-conglomerate biota often occupies a position intermediate between typical silt and breccia-conglomerate biotas. If these distinctions are acknowledged as reflecting genuine assessments of environmental character a wide range of diversity indices could be used to reflect these similarities/differences.

ii. Attempt a qualitative categorization of these different diversity indices based on the diversity index-profile results alone. Are any groups present? If so, which diversity indices belong to which group. (30 points)

Diversity Indices (All) copy

Level	Group 1	Group 2	Group 3
	Increasing to right	Steady State	Increasing to left
Dominance	✔		
Simpson			✔
Shannon			✔
Evenness			✔
Brillouin			✔
Menhinick			✔
Margalef			✔
Equitability			✔
Fisher_alpha		✔	
Berger-Parker	✔		
Chao-1		✔	
iChao-1		✔	
ACE	✔		
Squares			✔
Hill (q=-1)			✔

iii. In order to summarize the behavior of these indices quantitatively, select an appropriate ordination/discrimination procedure and analyze the diversity-index results for structured covariation. Which ordination procedure did you select? Why? Provide full details of your data-analysis method (with justification) and list all results obtained. (20 points)

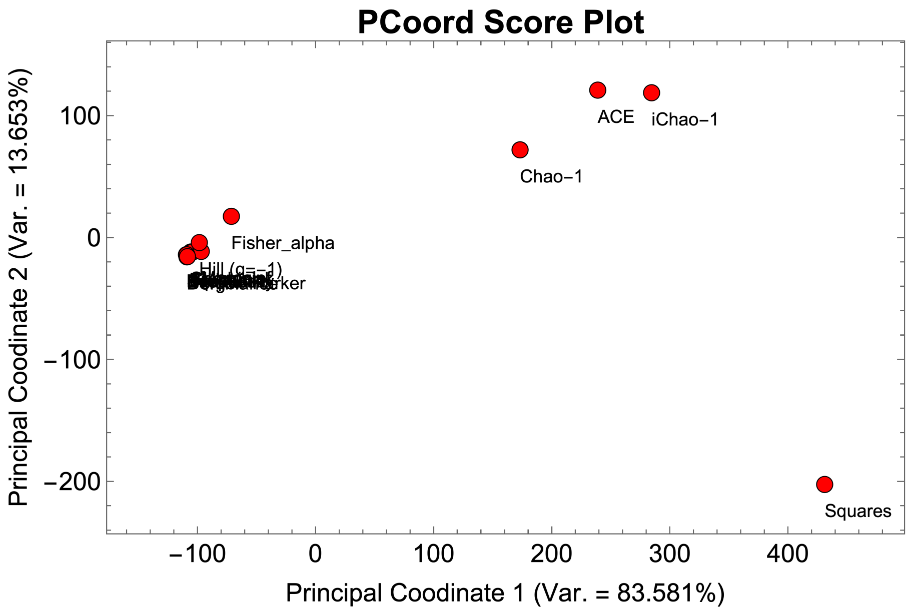
I chose a squared Euclidean distance PCoord as a general-purpose, linear, dimensionality-reduction method that would respond to both associations between the patterns resident in the various diversity indices but also to the magnitudes of those values. This option was preferred over PCA because decomposing the matrix in the PCA mode would have run into effective dimensionality problems ($n < m$).

Eigenvalues

Component	Eigenvalue	Variance (%)	Cum. Variance (%)
1	470796.794	83.581	83.581
2	76902.724	13.653	97.234
3	14426.934	2.561	99.795
4	1056.082	0.187	99.982
5	91.487	0.016	99.999
6	5.837	0.001	100.000
7	2.277	0.000	100.000
8	0.222	0.000	100.000

9	0.011	0.000	100.000
10	0.000	0.000	100.000

Only two eigenvectors are needed to capture the overwhelming majority of the variation present in this dataset.



iv. Examine the results you generated in 2.a.iii for evidence of structure in the summaries provided by these diversity indices. Are any clusters of diversity indices evident in your results? If a random-distribution model can be rejected attempt to identify the within-group similarities and between-group differences both in terms of value delivered for this dataset and with respect to the different diversity-index concepts. [Note: a brief description of all the diversity indices except the Hill can be found in the PAST ver. 5.0 User's Manual.] The Hill diversity index will be discussed in the course lecture. Other, more technical, descriptions are, of course, available on the internet. (20 points)

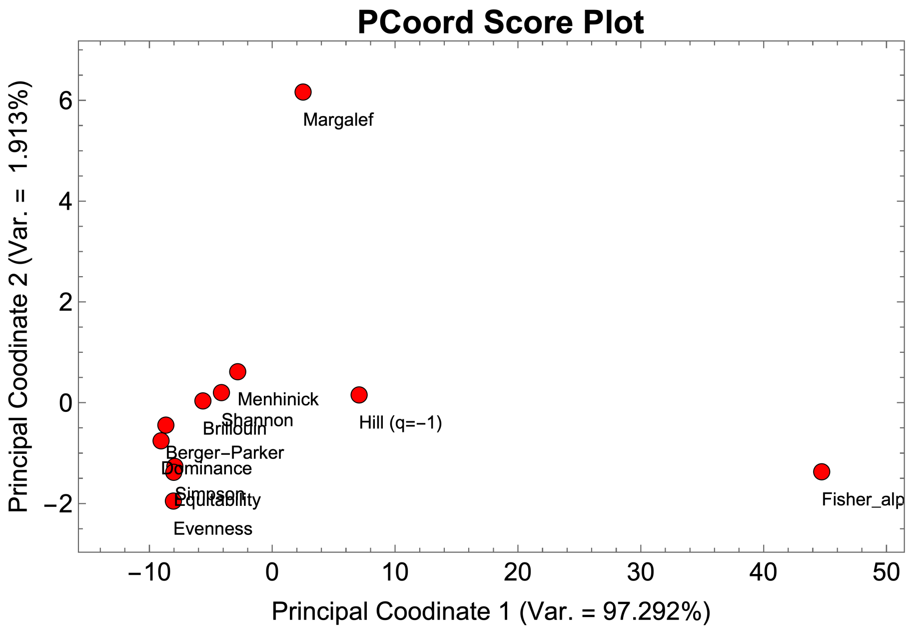
There are three clusters of diversity indices:
Group 1: Squares
Group 2: Chao-1, iChao-1, ACE
Group 3: The remainder

Group 1 represents the single index (Squares) that assumes abundance distributions are even. Since none of the other indices make this assumption the distinction of this index is not unexpected.

Group 2 unites the indices that focus on dominant aspects of the data but treat low-abundance taxa (e.g., singleton's doubletons) differently. Again, since none of the other indices make this distinction the internal similarity of these indices, as well as their differences from the main dominance/evenness group indices (Group 3) is not unexpected.

Group 3 collects the large group of simple dominance/evenness indices together.

v. If any obvious outliers exist in the results you generated in 2.a.iv, remove these and re-run your ordination to test for subsidiary structure that was being masked by the presence of the outliers. Are any subsidiary clusters of diversity indices evident in your results? If a random-distribution model can be rejected attempt to identify the within-group similarities and between-group differences both in terms of value delivered for this dataset and with respect to these different diversity index concepts. (20 points)



Removal of the outliers clarifies relations among the dominance/evenness diversity indices. The distinction of the Fisher α index can be explained by noting this index does not incorporate any relative abundance information. Similarly, the Hill (q = -1) index is the alternative of the family of Hill indices that places emphasis on the presence of common taxa. In this sense its location close to the main dominance/evenness index closer (that also take greater note of common taxa) is not unexpected. Finally the distinctiveness of the Margalef index (especially from the Menhinick index) reflects the fact that these two indices make very different assumptions regarding the character of differences in the relative abundance values (Margalef assumes they are distributed via square roots, Menhinick via logarithms). Placement of the Menhinick index within the main group of dominance/evenness indices implies these are making the sample (implicit?) assumption.

3. The Provinces datafile (Provinces.csv, Provinces.dat) contains frequency-count data for taxonomic assemblages collected from three different Ordovician provinces. We would like to know whether the species richnesses are the same both within and across provinces. (110 points)

   a. Convert the frequency-count data into taxon relative abundances. (10 points)

Appalachian Basin

		CAT1	CAT10	CAT11	CAT2	CAT3	CAT5	CAT6	CAT7	CAT8	CAT9	CF1	CF2	CF3	CF4	JG10	JG11	JG15	JG3	JG4	JG6	JG8	JG9		Total	Rel Abund.
1	Dalmanella	4	18	3	24	14	65	21	9	0	0	32	30	28	16	14	1	0	0	6	9	10	5		309	0.342
2	Rafinesquina	5	6	4	9	10	37	6	1	2	1	5	15	7	4	23	2	2	0	16	5	1	6		167	0.185
3	Sowerbyella	4	3	0	5	4	7	1	2	0	4	90	0	0	0	0	0	0	0	0	0	0	0		120	0.133
4	Zygospira	4	0	4	0	11	0	0	11	1	4	6	6	3	5	3	1	1	0	1	4	0	11		76	0.084
5	Tancrediopsis	0	6	17	0	6	0	16	4	2	4	0	0	0	0	0	0	1	0	1	1	0	3		61	0.067
6	Branching Bryozoan	0	6	0	0	0	0	6	3	2	3	0	0	0	0	0	0	6	14	0	0	0	0		40	0.044
7	Ischyrodonta	0	5	3	2	4	0	9	2	3	2	0	0	0	0	0	4	2	0	0	0	0	0		36	0.040
8	Modiolopsis	0	2	11	2	3	0	6	2	3	2	0	0	0	0	0	0	2	0	1	1	0	0		35	0.039
9	Sheet Bryozoan	0	3	0	9	0	6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		20	0.022
10	Ambonychia	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	2		10	0.011
11	Lyrodesma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0		6	0.007
12	Caritodens	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1		6	0.007
13	Ctenodonta	0	2	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5	0.006
14	Cornulites	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0		3	0.003
15	Encrusting Bryozoan	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		3	0.003
16	Dalmanellid (small)	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		3	0.003
17	Hebertella	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0		2	0.002
18	Lingulid	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		1	0.001
19	Praenucula	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		1	0.001
																									904	1.000

Mississippi Valley Basin

		RH8	RH12	RH17	RH13	RH6	RH10	RH11	RH16	RH15		Total	Rel Abund.
1	Thaerodonta	1	11	30	20	3	0	0	25	0		90	0.234
2	Paucicrura	12	2	9	0	2	13	6	3	24		71	0.184
3	Austinella/Plaesiomys	5	2	1	6	5	21	11	0	12		63	0.164
4	Strophomena	0	7	6	5	1	0	0	6	6		31	0.081
5	Plectodontid (small)	0	0	0	0	0	0	0	0	30		30	0.078
6	Onniella	6	3	0	0	9	0	8	3	0		29	0.075
7	Zygospira	0	0	2	1	7	10	0	0	6		26	0.068
8	Megamyonia	0	0	0	11	0	1	0	0	8		20	0.052
9	Dalmanellid (small)	0	0	0	0	0	0	8	0	0		8	0.021
10	Rafinesquina	0	2	1	0	0	0	0	0	1		4	0.010
11	Leptaena	0	0	0	3	0	0	0	0	0		3	0.008
12	Dalmanellid	0	0	0	3	0	0	0	0	0		3	0.008
13	Strophomenid (small)	0	0	0	0	0	0	0	0	2		2	0.005
14	Circular Bivalve	0	0	0	0	0	0	0	0	1		1	0.003
15	Lepidocyclus	0	0	0	0	0	0	0	1	0		1	0.003
16	Orthodesma	0	0	0	1	0	0	0	0	0		1	0.003
17	Plectodontid	0	0	0	0	0	1	0	0	0		1	0.003
18	Bivalve (small)	0	0	0	0	0	0	0	0	1		1	0.003
												385	1.000

Manitoba Basin

		SM4	SM13	SM12	SM5	SM14	SM16	SM6	SM18	SM8	SM20	SM17	SM2		Total	Rel Abund.
1	Diceromyonia	4	10	11	6	5	7	8	8	14	6	10	10		99	0.236
2	Dinorthis	6	12	8	9	5	4	13	7	10	5	10	5		94	0.224
3	Oepikina	6	2	2	3	8	0	5	0	3	2	0	5		36	0.086
4	Branching Bryozoan	4	1	0	3	9	0	1	0	6	0	2	0		26	0.062
5	Batostoma	2	4	0	2	3	3	3	0	0	1	3	2		23	0.055
6	Pachydictya	3	7	2	1	1	2	1	3	0	0	3	0		23	0.055
7	Sticopora	1	1	2	0	3	0	0	2	0	5	3	2		19	0.045
8	Strophomena	3	2	0	0	10	0	2	0	0	0	0	2		19	0.045
9	Megamyonia	0	1	0	0	9	0	0	0	0	2	1	0		13	0.031
10	Paleofavosites	2	2	0	3	1	0	0	0	0	0	3	0		11	0.026
11	Hypsiptycha	0	2	0	0	3	2	0	0	0	0	3	0		10	0.024
12	Sceptopora	0	0	0	0	0	0	0	5	0	0	2	0		7	0.017
13	Bighornua	2	1	0	0	0	0	1	0	0	0	1	1		6	0.014
14	Deiracorallium	3	0	0	2	0	0	0	0	0	0	0	0		5	0.012
15	Inflated Gastropod	2	0	0	0	3	0	0	0	0	0	0	0		5	0.012
16	Encrusting Bryozoan	0	2	0	0	2	0	0	0	0	0	0	0		4	0.010
17	Hiscobeccus	0	0	0	0	0	2	1	0	1	0	0	0		4	0.010
18	Isotelus	0	1	1	0	0	0	1	0	0	0	0	1		4	0.010
19	Nasutmena	1	1	1	0	0	0	0	0	0	0	0	1		4	0.010
20	Salvadorea	0	0	0	0	0	1	0	0	0	2	1	0		4	0.010
21	Planispiral Gastropod	0	0	0	0	1	0	0	0	2	0	0	0		3	0.007
22	Lobocorallium	0	0	0	0	0	0	0	0	0	1	0	0		1	0.002
															420	1.000

i. Compare the relative abundance rank orders across provinces. What is the ratio of cosmopolitan (= occurring in two or more provinces) to indigenous (= restricted to one province) taxa? Which province has the greatest number of cosmopolitan taxa? (20 points)

		Appalachian Basin	Mississippi Valley Basin	Manitoba Basin	Basin Count
11	Dalmanellid (small)	16	9		2
13	Dalmanella	1	12		2
41	Rafinesquina	2	10		2
50	Zygospira	4	7		2
47	Strophomena		4	8	2
29	Megamyonia		8	9	2
6	Bryozoan (branching)	6		4	2
17	Bryozoan (encrusting)	15		15	2

45	Sowerbyella	3			
48	Tancrediopsis	5			
22	Ischyrodonta	7			
30	Modiolopsis	8			
44	Bryozoan (sheet)	9			
1	Ambonychia	10			
28	Lyrodesma	11			
7	Caritodens	12			
10	Ctenodonta	13			
9	Cornulites	14			
18	Hebertella	17			
26	Lingulid	18			
40	Praenucula	19			
49	Thaerodonta		1		
37	Paucicrura		2		
2	Austinella/Plaesiomys		3		
	Plectodontid (small)		5		
33	Onniella		6		
25	Leptaena		11		
	Strophomena (small)		13		
8	Bivalve (circular)		14		
24	Lepidocyclus		15		
34	Orthodesma		16		
39	Plectodontid		17		
5	Bivalve (small)		18		
3	Batostoma			5	
4	Bighornua			13	
14	Deiracorallium			14	
15	Diceromyonia			1	
16	Dinorthis			2	
21	Gastropod (inflated)			15	
38	Gastropod (planispiral)			21	
19	Hiscobeccus			17	
20	Hypsiptycha			11	
23	Isotelus			18	
27	Lobocorallium			22	
31	Nasutmena			19	
32	Oepikina			3	
35	Pachydictya			6	
36	Paleofavosites			10	
42	Salvadorea			20	
43	Sceptopora			12	

46	Sticopora			7	
	Cosmopolitan	6	6	4	8
	Indiggenous	15	12	18	50
	C/I Ratio	0.400	0.500	0.222	0.160

Overall C/I index: 0.16
Appalachian Basin: 0.400
Mississippi Valley Basin: 0.500
Manitoba Basin: 0.222

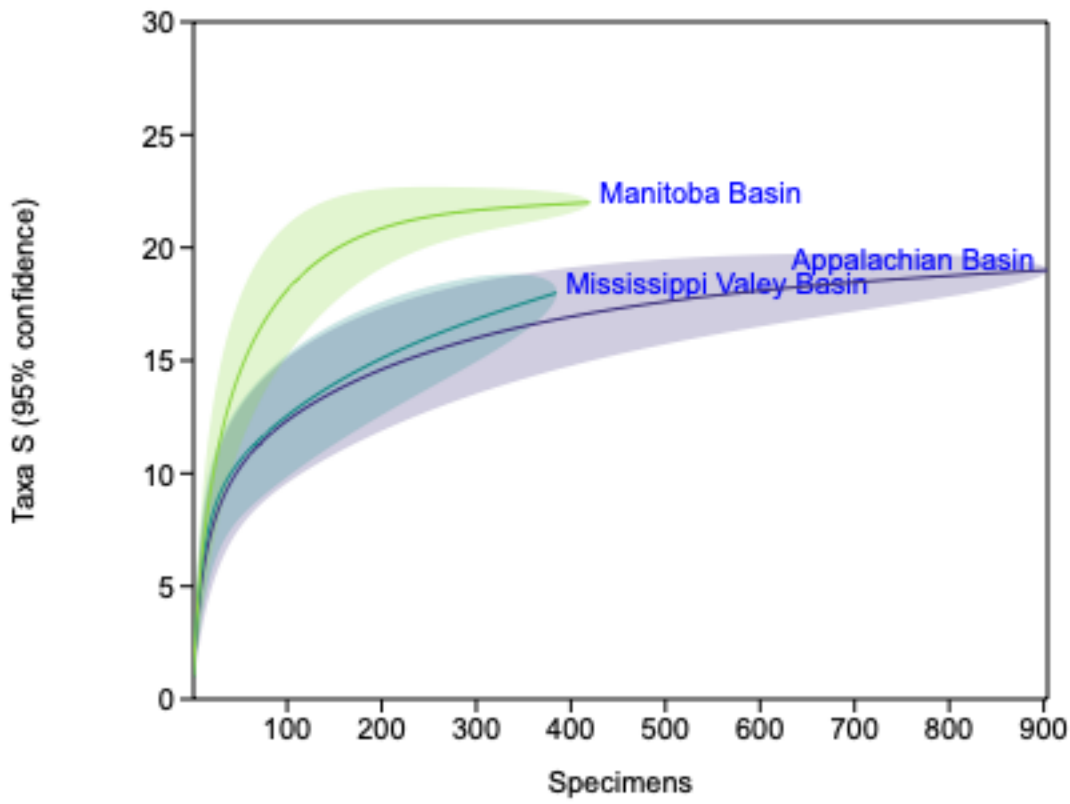
ii. Among the cosmopolitan taxa, what is the average difference in their rank orders across provinces? Does this value strike you are unusually high or unusually low, based on your qualitative perception? Describe a strategy whereby the significance of this value might be tested. [Note: up to 10 extra points will be awarded to those who implement their strategy and provide a quantitative answer to the previous question.] (30 points)

		Appalachian Basin	Mississippi Valley Basin	Manitoba Basin	Average
1	Bryozoan (branching)	6	-	4	5.0
2	Bryozoan (encrusting)	15	-	16	15.5
3	Dalmanella	1	-		1.0
4	Dalmanellid (small)	-	9		9.0
5	Megamyonia	-	8	9	8.5
6	Rafinesquina	2	10		6.0
7	Strophomena	-	4	8	6.0
8	Zygospira	4	7		5.5

I would not regard this result as unusual. All cosmopolitan taxa except for encrusting bryozoans exhibit average rank orders in the top 20 percent of the total faunal list. Indigenous taxa would, arguably, be expected to be more completely, and more complexly, adapted to local basin conditions than cosmopolitan taxa with cosmopolitan taxa exhibiting broader tolerance for environmental variation.

With regard to testing the previous interpretation I would set up a simulation experiment by creating a faunal list of names that reflected the total relative abundance of each taxon across all three basins. I would then create a series of (say) 1000 pseudo-samples for each of the basins of exactly the same size as the three basins in question by drawing names from this faunal list randomly, with replacement. from these sets of simulations I would calculate the mean rank order of the cosmopolitan taxa along with their rank-order variances. These statistical summaries could then be used to evaluate (i.) whether these were they cosmopolitan taxa most likely to be observed and (ii.) whether their mean rank orders were unusually high or unusually low.

iii. Use rarefaction analysis to obtain size-standardized taxon richness estimates for each province at m = 25, 50 and 100 specimens. Evaluate the null hypothesis that there are no species richness differences between these provinces. Which province has the most unusual estimated species-richness profile compared to the other two? Why might this be the case? (50 points)



Samp size	Appalachian Basin	Std.err 1s	Mississippi Valey Basin	Std.err 1s	Manitoba Basin	Std.err 1s
25	8.12	1.32	8.67	1.23	10.65	1.62
50	10.25	1.36	10.56	1.29	14.48	1.68
100	12.34	1.41	12.55	1.38	18.12	1.48

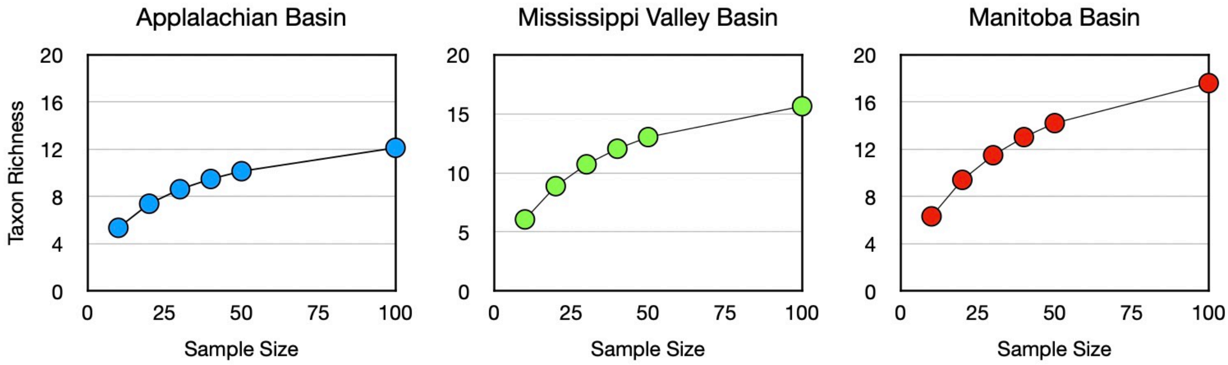
No taxic-richness differences between the Appalachian and Mississippi Valley Basins, quite a large taxic-richness difference between these two and the Manitoba Basin. This difference manifests itself primarily at sample sizes greater than 100. This structure seems to reflect the physical proximity of these basins.

4. Take the provinces data for Part 3 and conduct a parallel coverage-based analysis of species richness variations across the three provinces. For this exercise you may use either the equations listed by Chao and Jost (2012) or the ones used by Alroy in his Shareholder Quorum Subsampling (SQS) procedure (available in PAST). (160 points)

    a. Convert the frequency-count data from into taxon relative abundances.

        i. Based on these relative abundance data, estimate mean taxon diversity ($E(S_m)$) for subsamples of m = 10, 20, 30, 40 and 50 for all three provinces. Plot the estimated taxon diversity vs subsample size profiles for each province. According to these results what sample size might be necessary in order to obtain an accurate and stable estimate of taxon richness values for each province? (40 points)

E(Sm)	Appalachian Basin	Mississippi Valley Basin	Manitoba Basin
10	5.352	6.062	6.321
20	7.390	8.877	9.412
30	8.621	10.727	11.496
40	9.486	12.046	13.030
50	10.143	13.031	14.213
100	12.125	15.648	17.602



Greater than 100 specimens.

ii. Based on these relative abundance data, estimate the coverage for subsamples at m = 10, 25 and 50 individuals across all three provinces. Which province has the most efficient coverage profile? Which the least? What factors are responsible for this difference. (40 points)

E(Sm)	Appalachian Basin	Mississippi Valley Basin	Manitoba Basin
10	0.731	0.617	0.619
25	0.881	0.794	0.800
50	0.942	0.889	0.899

Most Efficient: Appalachian Basin
Least Efficient: Mississippi Valley Basin

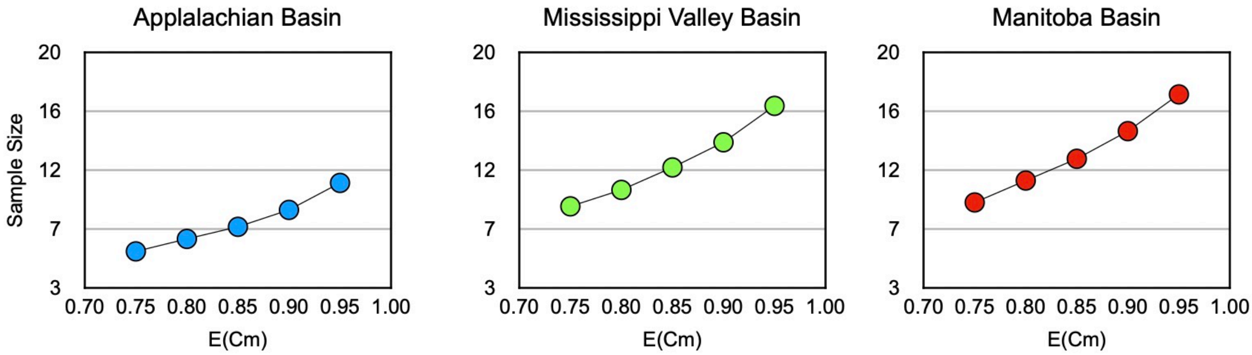
Differences in the evenness of the taxon relative abundances across the different basins.

iii. Repeat the exercise in 4.a.i for all provinces, but this time correct the sample sizes to ensure equal coverage at $E(C_m)$ values of 0.75, 0.85 and 0.95. Document how sample sizes need to vary across provinces in order to achieve equal coverage. (20 points)

E(Cm)	Appalachian Basin	Mississippi Valley Basin	Manitoba Basin
0.75	11	20	19
0.85	20	37	35
0.95	58	118	87

iv. For each province estimated taxon accumulation curves based on five nodal values than employ the constraint of achieving equal coverage values for each node. (20 points)

E(Cm)	Appalachian Basin	Mississippi Valley Basin	Manitoba Basin
0.75	6	9	9
0.80	7	10	11
0.85	7	12	12
0.90	9	14	14
0.95	11	16	17



v. Compare and contrast the results you obtained above with those obtained in Part 4.a.i. What are the similarities? What are the differences? Which set of results do you believe to be more accurate? Why? (40 points)

Estimates of the sample sizes needed to ensure comparable coverage are very different between the different basins. Similarly, standardization by sample size more-or-less ensures dramatic differences in coverage, especially at small sample sizes. Since the purpose of standardization is to minimize differences between samples, it would appear standardization by coverage, rather than simple sample size, is a far better basis on which to compare samples.