Daniella Assing

Geosci 541

Feb 29 2016

Lab #6

#### **Problem Set 1**

1) As part of our matching procedure between the Macrostrat database and the Paleobiology database, we lost some data - i.e., there was some Paleobiology Database occurrences that could not be matched to the Macrostrat database. How many occurrences were lost? Show your code.

> dim(DataPBDB)

[1] 34166 26

> dim(MacroPBDB)

[1] 9013 13

> dim(DataPBDB)-dim(MacroPBDB)

[1] 25153 13

25153 occurrences were lost.

2) Using what you know about the Macrostrat database can you speculate as to why so much data was lost?

Macrostrat currently only contains data for North America while the Paleobiology Database is more complete and contains data for every continent. Therefore some of the data would not match up.

#### **Problem Set 2**

1) A good candidate for classification as a lagerstätten should have high diversity of different taxnomic orders. Calculate the order-level richness of each stratigraphic formation. Find the top 10 candidate units by this criteria. State them and give the code you used. Also, create a **character** vector( ) listing the names of these stratigraphic units. Name this vector CandidateUnits.

##### **Hints**

1. Remember how to calculate the richness of a sample, and *apply* this to each sample of the OrderMatrix.
2. If you do not remember how to calculate richness, consult the [previous lab](https://github.com/aazaff/teachPaleobiology/blob/master/LabExercise5.md#problem-set-2).
3. Use the sort( ) function to put your richnesses in order.

> specnumber(GenusMatrix, , MARGIN = 1)

> GenusMatrixRichness<-specnumber(GenusMatrix, , MARGIN = 1)

> GenusMatrixSort<-sort(GenusMatrixRichness)

> CandidateUnits<-names(tail(GenusMatrixSort,n=10))

> CandidateUnits

[1] "Tavsens Iskappe Gp"

[2] "Bison Creek Fm"

[3] "Marjum Limestone"

[4] "Pioche Shale"

[5] "Sekwi Fm"

[6] "Stephen Fm"

[7] "Rabbitkettle"

[8] "Cow Head"

[9] "Chancellor"

[10] "Deadwood Fm"

2) A good candidate for classification as a lagerstätten should have a large number of genera that are relatively rare. Using the GenusMatrix column to find out how many samples each genus occurs in. Show the code you used. Name your outputGenusFrequencies.

> GenusFrecuencies<-apply(GenusMatrix,2,sum)

3) What is the necessary code to make a [frequency barplot](https://github.com/aazaff/startLearn.R/blob/master/expertConcepts.md#describing-distributions-with-statistics) of the GenusFrequencies?

> barplot(table(GenusFrecuencies))

4) After looking at this barplot, you should see a familiar paleobiological pattern. What do we call this type of curved distribution in paleobiology and ecology. [Hint: Think back to the lectures]

This type of pattern is the hollow curve distribution.

5) Subset your new vector GenusFrequencies so that only those genera from the lower 50% of the distribution are present. In other words genera with occurrences <= to the median( ). Show your code. Name this new subset vector as RareGenera.

> median(GenusFrecuencies)

[1] 2

> RareGenera<-which(GenusFrecuencies <= 2)

#### **Problem Set 3**

We will now take the list of rare genera you just made, RareGenera and see ***what percentage of genera in our top 10 candidate stratigraphic units CandiateUnits come from this list***. We will then decide which of our 10 candidate units is the most likely to be a Lagerstätte.

1) Subset GenusMatrix so that it only shows the 10 stratigraphic units we determined were potential candidates. Show your code and name your new matrix, CandidateMatrix.

> CandidateMatrix<-GenusMatrix[CandidateUnits, ]

Load in the following code that I wrote for you. It will find the percentage of genera in each sample (stratigraphic unit) that occur in our vector of RareGenera.

# A function that I wrote just for you, don't you feel special?  
percentRare<-function(CandidateUnit,RareGenera) {  
 CandidateUnit<-CandidateUnit[CandidateUnit>0] # Limit the data only to taxa prensent (non-zero) in the unit  
 PercentIn<-length(which(names(CandidateUnit) %in% names(RareGenera))) # Find the number of genera in the CandidateUnit that are in RareGenera  
 TotalGenera<-length(CandidateUnit) # Find the total number of genera in the unit  
 PercentShared<-PercentIn/TotalGenera  
 return(PercentShared)  
 }  
  
# Apply the percentRare( ) function to each row of CandidateMatrix  
PercentShared<-apply(CandidateMatrix,1,percentRare,RareGenera)

2) Based on the output of PercentShared and your answer to [Problem Set 2 - Question 1](https://github.com/aazaff/teachPaleobiology/blob/master/LabExercise6.md#problem-set-2) - i.e., the ranking of candidate units based on the number of orders represented, which four units do you think are most likely to qualify as Lagerstätten? Explain your reasoning.

> PercentSort<-sort(PercentShared)

> CandidateUnits

[1] "Tavsens Iskappe Gp"

[2] "Bison Creek Fm"

[3] "Marjum Limestone"

[4] "Pioche Shale"

[5] "Sekwi Fm"

[6] "Stephen Fm"

[7] "Rabbitkettle"

[8] "Cow Head"

[9] "Chancellor"

[10] "Deadwood Fm"

These are the four units which I think most likely qualify as Lagerstatten:

Stephen Fm

Deadwood Fm

Sekwi Fm

Chancellor

These have high percents and they also occur lower down in the CandidateUnit list (arranged in descending order).

3) Look closer into the into the four units you chose - using Google and information in the Paleobiology Database. One of these should be a very famous Lagerstätten. Which one? What is famous about it? What is its significance to Paleobiology?

The Stephen Formation is a unit in the Burgess Shale. It is significant in paleontology because it preserves the soft parts of many organisms (which are usually not preserved).