Assignment Week 2
Getting started with MySQL and Mathematica

A. Refresh yourself on the following

1. Geological Time Scale
2. Basic Statistics (Mean, Standard Deviation, Variance, Range, T-Test, ANOVA, Regression)
3. Basic Operation of MS Excel
4. GoogleEarth, GoogleMaps, GoogleScholar

B. Familiarize yourself with the following online data sources

1. List of Paleo Research Databases (http://www.ucmp.berkeley.edu/pdn/pdnhomelinks.htm)
2. Paleobiology Database (http://www.pbdb.org/)
3. NOW Database (http://www.helsinki.fi/science/now/)
5. Ocean Drilling Database (http://www.oceandrilling.org/Data_Samples/default.html)
6. NOAA Paleoclimate Data (http://www.ncdc.noaa.gov/paleo/data.html)
7. GBIF Biodiversity Data (http://data.gbif.org/welcome.htm)
8. ORNL Spatial Data (http://webmap.ornl.gov/wcsdown/)
9. USGS Geospatial Data (http://datagateway.nrcs.usda.gov/)

C. Useful resources

1. SLASHTMP server for transferring large files (https://www.slashtmp.iu.edu/)
2. GoogleScholar IULink (Add “Indiana University” under Advanced Preferences)
3. IUWare Online for university licensed software (http://iuware.iu.edu/)
4. Wikipedia for good background information on most statistical, mathematical, or analytical topics.
In Class Assignment

1. Load the Carnivore data set into MySQL.

Using phpMyAdmin

- Practice selecting the entire table.
- Select only rows for Felidae.
- Select only columns with measurements.
- Select rows where body mass is greater than average.
- Count number of data entries for each family (Group By).

Using Mathematica

- Practice selecting the entire table.
- Select only rows for Felidae.
- Select only columns with measurements.
- Select rows where body mass is greater than average.
- Count number of data entries for each family (Group By).

Take Home Assignment

1. Find new data of your own, at least 30 rows and 10 columns. These might be data from a paper, from an online supplementary data file, or data that you have collected. The data should have both text information (e.g., species names) and quantitative information (e.g., measurements or geological ages). At least one column should be a grouping category that is shared by more than one entry in the data (e.g., Family, sex, locomotor type, habitat, geological age).
2. Load these into a new table in your database.

In Mathematica:

3. Use SQLExecute and the Group By function to create a summary table for your data, summarizing the number of individuals in the group, the maximum and minimum of all the numeric columns, the mean of all the numeric columns.
4. Export this table to Excel and format it nicely with column headers, lines, etc. as though you intended to publish it.
5. Using the Graphics[] function, make a bivariate plot of two of your measurements. Give the points for each group a different color, add labels to each point, and label the axes.

Turn in by e-mail:

- Original data table in CSV format
- Mathematica notebook
- Excel version of the final table
Next Session

No meeting next week (11 September)

Reading for 18 September:
