**Troubleshooting for Big Data Biology – Pseudosuchia data set**

**Useful links**

* Big Data Biology home page: <https://www-users.york.ac.uk/~dj757/BIO00047I/BIO00047I_index.html>
* Big Data Biology data set descriptions: <https://www-users.york.ac.uk/~dj757/BIO00047I/misc/BIO00047I-data-description-2021.pdf>
* Big Data Biology report guide, marking scheme & data set analysis guidance: <https://www-users.york.ac.uk/~dj757/BIO00047I/misc/BIO00047I-report-guide-and-marking-scheme-2021.pdf>
* Google document from the Q&A sessions: <https://docs.google.com/document/d/16Ux020nGcrhgd3KVPehdpbBRsX5PGrqIDtg_pcJLBtA/edit>

**Common problems with *R***

* Not setting your working directory before you start – if you’re getting an error message as soon as you get started, double check that you’ve set your working directory correctly - setwd()
* Not installing/loading packages – it’s easy to forget to install/load packages so if you have problems make sure you double check you have all the necessary packages loaded.
* Typos! - this is probably the most common problem and it happens to all of us! If you’re having trouble getting something to work double-check your spellings (variable names, functions etc) and double-check the case. Sometimes this means going through your code line by line, even character by character to spot the issue.
* Anything to do with ggplot2 take a look at this cheat sheet: <https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

**Common problems that are data set specific**

* Error when trying to plot the extinction or net diversification rate through time for the marine partition.
  + This comes up a lot and is caused by a bug in the *R* package causing a small number of NAs in the extinction time series for the marine partition. This is how to fix it.
  + First check for yourself that this is in fact the problem:

which(is.na(rtt\_M$mu))

This will show you which points in your time series contain NAs instead of an extinction rate (remember to check variable names!).

* + Now you can set the Nas to zero using this line of code:

rtt\_M$mu[is.na(rtt\_M$mu)] <- 0

* + You will then need to rerun the plot command for extinction using this code (your settings may vary, the important part is to replace subtree with your variable name, which in this example is rtt\_M:

plotRateThroughTime(rtt\_M, ratetype='extinction', avgCol="blue", ylim=c(0,0.5), cex.axis=2, intervalCol='blue', intervals=c(0.05, 0.95), opacity=0.1)

* How to change axes on rate through time plots.
  + This is a tricky one as there are no settings in BAMMtools (at least that are listed in the documentation) that allow you to do this. My suggestions are any of the following
    - Check the BAMM Google group to see if anyone has asked this previously: <https://groups.google.com/g/bamm-project>
    - Post a message asking the question on the Google Group.
    - Do it in Inkscape or similar ([https://inkscape.org/release/](https://inkscape.org/release/inkscape-1.0.2/)).
  + Sorry I can't give you an easier solution!
* What are the histogram axes?
  + These are frequency histograms therefore the y-axis should be labelled as either “Count” or “Frequency”. The x-axis is the correlation-coefficient between the environmental variable and the rates through time so I recommend renaming the current label to something more sensible.
* Data citations:
  + Temperature: Zachos, J. Trends, rhythms, and aberrations in global climate 65 ma to present. *Science* 292, 686–693 (2001). <https://science.sciencemag.org/content/292/5517/686>
  + Temperature: Veizer, J. et al. 87Sr/86Sr, δ13C and δ18O evolution of Phanerozoic seawater. *Chem. Geol.* 161, 59–88 (1999).  
    <https://www.sciencedirect.com/science/article/pii/S0009254199000819>
  + Sea level: Haq, B.U., Hardenbol, J., and Vail, P.R., 1987, The chronology of fluctuating sea level since the Triassic: *Science*, v. 235, p. 1156–1167, <https://doi.org/10.1126/science.235.4793.1156>.
  + Phylogeny: Payne ARD, Lloyd GT, Mannion PD & Davis KE. Decoupling speciation and extinction reveals both abiotic and biotic drivers shaped 250 million years of diversity on crocodile-line archosaurs. *In prep*.
  + Diversification rate data: Payne ARD, Lloyd GT, Mannion PD & Davis KE. Decoupling speciation and extinction reveals both abiotic and biotic drivers shaped 250 million years of diversity on crocodile-line archosaurs. *In prep*.
* Some useful papers:
  + <https://www.nature.com/articles/ncomms9438>
  + <https://www.nature.com/articles/ncomms13003>
  + <https://science.sciencemag.org/content/323/5915/728.full>