To do:

**Make grid** (1 degree and 5 degree) for CCZ polygon, aggregate observations by grids

🡪 Finished making 5 degree grid, need site coordinates to aggregate. Will check data (M- added site table as excel file – tabs split by data source- may need some explan)

**try log regression plot** – fig 3c currently – total phyla-classes/orders/families/genera- log total and get a predicted species value from slope intercept (try predict function?)

**get estimates of S** ie Cowell et al., 2012– from specaccum? From specpool? From estimateS?

**Rerun species-level richness estimates and curves for** – species accumulation, rarefaction in vegan, chao1 and chao2 in iNEXT using final data-

CCZ\_ALL\_SPP\_DATA\_2022-11-05

NB 1st 3 rows relevant ones- site/abundance/species + extra columns-

‘rec’ for row records from each data source- (literature/DeepData/OBIS/GBIF- recorded in row ‘SOURCE’)

‘site\_cat’ for size fraction- can use to subset data and run curves separately for mega-macro and meiofauna

‘REGION’- can use to subset data and run curves separately for regions- west-central-east-

**Do UpSet plot** using above data (CCZ\_ALL\_SPP\_DATA\_2022-11-05) by region

**Do MDS plot** by region or depth bins?

**Compile final data- all taxa** final data for analysis but at all taxa levels not just spp-level (i.e. CCZ\_ALL\_SPP\_DATA\_2022-11-05 but records id’d to all taxon levels not filtered for spp only)

**rerun family accumulation curve** (fig3a) and supp fig violin plot phyla by depth using above data

M to do metadata files