Jan 2023 to do:

Generate table of singletons per site, plus at level of contract area/APEI so can get estimate of proportion of species which only occur once, or only occur in one APEI, contract area etc

(currently doing a matrix table- spp by site with abundance and sum across to see which are singletons

Get estimate of S (ie E(S)2000, or S as derived from estimateS Cowell,

Think is default output of one of vegan functions- to check

WoRMS/WorDSS data- all accepted metazoan spp by phyla- M to get data from the db editors and do breakdown of no of spp by phyla

iNEXT- M and J to redo and see if plots are the same.

Also do plot of sampling completeness (ggiNEXT(Hills\_q, type=2))

Look at Reviewer 1 comment- suggest you do Hills nos (don’t think we need to include Simpson and Shannon ie q=1, q=2. Maybe they mean we use the estimated values only?)

New S Fig 1 – no of samples by depth, all phyla combined- final version

GEBCO grid readd to old repo

Generate clean final script- check fully reproducible

reviewer 3- could more be done with mapping the data- have a look at- mapping species, map diversity by region- east central west or something? Total spp by contract area?

To do (2022):

**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?

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

**Do community matrix plot by contract area-apei to figure out if any singletons over area (**ie species known from one specimen- generally per deployment- but here by region ie Spionis sp. A is only in APEI-10)

**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

🡪 done but minor data updates- need to update file and rerun (ie 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

🡪 done and added into folder – 5th nov plots

**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)

🡪 Added (CCZ\_ALL\_TAXA\_v1\_2022-11-06 – nb current version missing species names

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

**do depth plot all taxa records** (of above)- violin plot- overay somehow the total record distribution – e.g. transparent boxplot of total records of depth by contract area – and overlay of violin dist of records of depth by phyla?

**do metadata files**