6/17/2019

* There are some species in the list that are “least concern” and “data deficient” which we might choose to omit from the study? If they are data deficient, there may not be enough data for the parameters we want to test. If least concern, they may not compare well with the other species despite still being in decline?
* Might want to wait until looking through the Sea Around Us database for information before making assumptions about removing species from the data set.
* I will look at the FishBase data to gather depth information first. If I can’t find depth data from FishBase then I probably won’t be able to find other data for the fish. Might be a good way to start filtering.
* \*\*\*How are we going to analyze or standardize depth as a representation of cost? Do we need to? What do we do about a depth range, find the middle and call it an average?
* As of today I’ve gotten FishBase available depth data, exploring the Sea Around Us data
* Having trouble finding useful data from Sea Around Us, so I’m kind of just navigating the website and finding all kinds of other resources.

6/18/2019

* Interesting/useful sources for data:
  + Catches by EEZ in the waters of FAO area Atlantic, Northeast (27): <http://www.seaaroundus.org/data/#/fao/27?chart=catch-chart&dimension=eez&measure=tonnage&limit=20>
    - Has a download button for data, downloaded in my DropBox as “SAU FAO 27 v47-1”
    - Advanced search option?
  + Marine Tropic Index info, data on their GitHub in R: <http://www.seaaroundus.org/regional-mti-tools/>
  + EEZ data: <http://www.seaaroundus.org/data/#/eez>
    - Can download data by country, might be useful?
    - Advanced search option
  + <https://github.com/seaaroundus?utf8=%E2%9C%93&q=&type=&language=>
  + FAO **Fishery Statistical Collections** Global Capture Production: <http://www.fao.org/fishery/statistics/global-capture-production/en>
    - Some datasets?
  + FAO FishStatJ data/program: <http://www.fao.org/fishery/statistics/software/fishstatj/en>
    - Might have useful data, but it looks like I have to download the program to access it?
  + FAO FishFinder Fact Sheets, like FishBase: <http://www.fao.org/fishery/species/search/en>
  + Would “tonnes of catch” catch allocation data from SAU be useful for max catch/MSY, or is that not direct enough?
* May want to refer to the FishBase distribution maps to see if a fish’s distribution is really were I want—some of the species didn’t seem to have very Atlantic regions—some were more Mediterranean Europe.
* Costello notes:
  + Does the Ê mean estimate? Saying that the estimated MSY=…?
  + Are the 1.78 and other numerical constants/values something that will remain no matter what? Or is that something I’ll have to figure out? (I assume it’s standard, as it’s what was published as THE equation, but I just want to make sure)
  + How much should I understand about how they got to this conclusion/why it works? I’m still unsure of this.
* What units should MaxCatch be in? Or what standard unit should I convert to?
* Could the RAM Legacy database be of use? <https://www.ramlegacy.org/explore-the-database/regions/>
  + Costello et. al cited a paper about the RAM database and it seems good? Still need to look at the data though
  + <https://zenodo.org/record/2542919#.XEk3hM9KjBK> to access the data, saved as “RLSADB v4.44” in my DropBox

6/24/19

* Separated the Ram species that had a Bmsy value listed from those that do not
* Color coded according to general location
  + Substantial number in Pacific and Atlantic, could narrow down to NZ? Or N/S Atl?Pacif? or type of fish? Maybe rockfish in Pacific?
* Looks like I’ll **need to use R to calculate the means for the timeseries data**, it’s so long. I think I’ll at least go through and enter “NA” into the blank cells to clean up the data a bit more, but that’s it. I think I’ll make it its own csv to avoid weird formatting errors
  + Use a for loop to avoid redundancy, see list of chosen species from other tab (or see the methods that Michelle went over with me and Cassiel on 7/1)

6/25/2019

* SAU data is for each country,
  + Assess at the FAO region primarily
    - Use cost and price data that Ed recommends or take price from the global dataset and take only from countries in my FAO region (get average price/cost by country per species)
    - Compare to the catch values from the FAO
  + **Total landed for each species each year**
  + To get max catch:
    - Go to tools &data 🡪 basic search 🡪 FAO Area 🡪 area 27 🡪 download data
    - Learn how to read in csv file in R, manipulate data tables in R (tidyverse, dplyr),
    - **Sum the tonnes for each species each year—across dif countries and categories (total catch of that species in that FAO area in that year), look across all years for each species and get a max (write in dplyr and as for a max across years) (take the timeseries and get the max over that time)**
    - Price (landed value): tools & data 🡪 fisheries economics 🡪 advanced search 🡪 search by FAO area 🡪 go to area 27 🡪 download data (wait! But this gives me the same as the other data download???)
      * Ed should have some of this data for his paper, ask him where he got it
  + FAO area 27 countries: Belgium, Denmark, Faeroe Isl. (Denmark), Finland, France, Germany, Iceland, Ireland, Malta, Netherlands, Norway, Portugal, Russian Federation, Spain, Sweden, United Kingdom, Channel Isl. (UK), Azores Isl. (Portugal), Estonia, Latvia, Lithuania, Poland, USA, Greenland
* Tidyverse? Makes data easier for you to manipulate in R?
* Reading through Ed’s paper to find the resources he used for cost/price:
  + Smsy = greatest harvestable surplus production
  + Fmsy = associated harvest rate
  + “The distributions of cost/benefit ratios γ, stock
  + MSY, and mean harvest rates for fisheries in the RAM Legacy database … γ was computed as the simple average of the (variable cost + government subsidy)/total landing value ratios from countries or regions involved in fishery management, averaged over the years 1990–2000 (22, 27). This γ estimator was derived by ref. 27 from various sources (European Commission, Food and Agriculture Organization, websites, gray literature, and survey)”
* Ed Meeting:
  + Talking with Malin, it sounds like I could still just use the Sea Around Us data. I can use the landed value data from the global dataset that match the catch data from the regional FAO dataset. Here I’m getting price and harvest data
  + Now we need to find cost data, Malin said you have access or used cost data in your paper. Is there anything that can be used with R? If not, where should I look/is depth a good enough proxy still?
  + I’ll try to see how many of my IUCN species overlap with the SAU database and use those species if I can, otherwise maybe do a random selection?
  + Evolution conference in Boston
  + Global market should still apply for regional
  + His cost data was country-based
  + Malin’s idea for max catch sounds good, to estimate MSY/ carrying capacity
  + Stick with the RAM might be easier, need to talk to Malin to coordinate our next moves
    - RAM is good at harvest rate, SAU biomass?
    - Can look at relative biomass and infer extinction risk, compare to model that Ed and Juan are working on
    - This approach easiest to get to on paper, Ed and Juan will be using similar methods to Ed’s paper for their analysis
    - Focus on cleaning up data and analysis, this is a side project?
      * SAU would be a side project, higher quality/more data than RAM and could focus on a smaller geographic region, goal would be to explain species extinction probability
      * Depends on what we want to focus on/how much work I want to do (RAM easier)
  + Meeting with Malin tomorrow, will try to come together
  + Ask Juan to see his simulation results
    - Used Ed’s model and tried assimilating stock/inst groups to see how they do—how many species remained
  + UC Santa Barbara RAM help
  + RAM excel
  + Have data organized to copy it into R, start playing around
    - Can call Ed and screenshare
* Try to replace text with graphs if possible for presentation

6/27/19

* Today:
  + Try to find overlap between IUCN and SAU species from region 27
  + Try to load in and start finding the means for timeseries data for RAM stocks of interest
    - Review R workshop notes/docs, read the R book in the lab
  + Make a pro/con list for each database and why should/not use each one

RAM

|  |  |
| --- | --- |
| PRO | CON |
| Easier analysis, quicker | Few species |
| Readily accessible MSY proxy/measure | Stock, not species based |
| Well-known and recognized | No cost or price data associated |
| Mentors and other faculty are familiar with it | Very little overlap with IUCN species selection |

\*stock-based fishery groups, not species-based

\*loosely tied to a given region

\*MSY data focused on relative biomass and not just directly MSY

SAU

|  |  |
| --- | --- |
| PRO | CON |
| Close MSY proxy | May require more work/time to get the information we want how we want it? |
| Potentially accessible/useful price data | Catch data on regional scale, price on global? |
| Some mentors familiar with it | Doesn’t have cost data |
| Many species overlap with IUCN |  |

\*species-by-country-based (kind of like stocks, but can look at the data in a way that looks at species across the various countries)

\*Regionally-based

\*can use landed catch to calculate maximum catch to calculate MSY

* + Meet with everyone and go over next steps
    - Which database are we moving forward with? **SAU seems more favorable if we still want to work with IUCN species. Also more favorable if we can use the global landed value with the regional/country landed catch data.**
    - What analysis or data reformatting might that require? Should I use R for all of it or can I do some in excel? What packages do I need in R to do those analyses? **Likely to use tidyverse, and I downloaded it today (should be able to run it now). What are some of the things I should be reading up on how to do? Wat are the first few things I should be doing and what resources do I have for that?**
    - What should we move forward with for cost data? Should we continue to use depth, and if so how? (Depth is given as a range on FishBase, not a discrete ideal depth value. Should we find the middle depth in the range and go with that? **Should we maybe base them off of the lowest depth? Should we find other sources aside from FishBase to reference for this? Is there a way that we could rank a depth range to give it a singular numerical value we can look at?**

6/2719 Group meeting:

* For the theory (working with Juan):
  + Cost, benefit/price/total revenue, harvest/MSY/biomass
  + Y axis is B/Bmsy and X is same as Ed’s paper
  + Testing to see if different institutions will perform (lower range rarely reached, not equal distribution of bimodality
  + Predictions about how often extinctions, frequency
* Malin’s concerns with RAM: it would be essentially the same as Ed’s project, not novel
  + Not harvest rate, but biomass—extinction risk determine by harvest rate and if a stock is close to the bifurcation
* **B/Bmsy biomass RAM data would be more accurate than SAU**
* **Cost/price data that Ed used for his paper, would be easiest route—depth?**
  + Depth could be an additional variable to add into cost knowledge
  + Cost is the hardest thing to get, in general
  + Distribution of B/Bmsy over x axis to show the bifurcation and data dist across this, find the mean?
  + Plot as average biomass including ones that go extinct is the prediction we want to compare
  + Juan could calculate the potential and visualized form my data? Few points but can give a sense of the relative importance of one outcome over the other
  + **Differentiated from Ed’s paper since his was human-centric and this is species-centric**
* Ramp up a paper by end of summer, lead a paper on the data analysis
  + Entails: doing the data analysis, writing it up and producting figures and discussion
  + Could have a draft that could be published over next few months
  + Or a paper that I am listed as a contributor, could come out earlier
  + Could split it into 2 papers, one where I do data and then we push out a big one with the predictions with Juan and Ed?
  + Have a report-scale paper ready, but would need to be in touch after to make something real
* Base on RAM for best stock, but check SAU for price data and use depth (mean depth?)
  + **If using depth as cost, we need some good cost/depth data (do a literature review looking at cost/depth relationship for fishing, 2011 LAM paper referenced in Ed’s paper price by gear type, see if RAM has gear type)**
  + Deliverable: cost/benefit for RAM stocks and on the side they work on their theory+data paper
  + Maybe narrow down to the IUCN list-RAM overlap species and really dive dep into their biology and life history info, still compare to theoretical model and see if there’s anything
    - More qualitative, have model be a separate paper
    - **Assess things like cc on habitat? Could incorporate into the model? Fishing history could be important, reflected in the data?**
    - Still will set up a general method of how to relate the cost/benefit/harvest data
    - **Research how ecological characteristics relate to the cost, country GDP, find all kinds of other data**
    - That paper on the distance from shore could be a good proxy for cost, can look into this paper and similar things for the historical data and stuff
* Next steps: lit search, IUCN-RAM overlap good next steps
* **Get B/Bmsy values**
* R/analysis:
  + **assume all species in IUCN have low biomass**
  + stats tests: not sure, only 9 points isn’t much, try to get a better sense of what factors influence extinction the most (**regression analysis/trend line fit bc probably not a linear relationship**, quadratic function? \*\*\*familiarize how to transform and fit a best fit/regression into a linear format)
  + mean MSY, mean cost/benefit ratio
* **Joyce has good code for pulling data from RAM!**
* Michelle will show me a better way to move and filter the data, less error prone and 100% going to get all the data points I need, yay!

6/28/2019

* Today:
  + **Assess how many IUCN species are in the filtered RAM and overall RAM**
    - **There were very few species that overlapped AND had Bmsy data**
    - **Will need to discuss and perhaps look into other regions (Pacific or NZ)**
  + ~~Start researching those species, just download a ton of the literature~~
  + ~~Begin to make a list of set parameters that all have in common~~
* Way to find species info in RAM:
  + Latin name and stockID both in “stock” tab
  + Can use stockID to search in the “bioparams\_values\_views” tab for the “Tbmsybest” value
  + Use stockID to see if the species is in my reduced list
  + Use latin name to search for information
  + **Latin name 🡪 stockID 🡪 Tbmsybest data**
  + **Tbmesybest 🡪 stockID 🡪 latin name 🡪 IUCN check**
* RAM overlap stock ID (searched latin name in the “stock” tab and will use stock ID to search in filtered data):
  + **BLSHARNPAC, Pacific species**
  + **ANCHMEDGSA16 (Sicily),**
  + **ANCHOSA, Argentina**
  + **ALBANATL, Atlantic**
  + **ALBASATL, Atlantic**
  + **BLUEFISHATLC**

Not overlapping: GOLDREDNEAR, GOLDREDV-VI-XII-XIV, SBARSHARATL, SCHHEADATL, TUR2232, PORSHARATL, PORSHARNEATL, STRAYII-IIIa-IV, GTRIGGM --Gulf of Mexico, ATBTUNAEATL, ATBTUNAWATL, ASHARNEATL, TPSHARNEATL, SDOG4VWX5 –CA, SDOGATLC –US, SDOGBLKGSA29 --black sea, BLRAYIVa-VI, BLRAYIVc-VIId, BLRAYIXa, BLRAYVIIafg, BLRAYVIIe, ANCHIXa (these are all from dif regions, most in Mediterranean Sea though), ANCHMEDGSA1, ANCHMEDGSA17, ANCHMEDGSA17-18, ANCHMEDGSA19, ANCHMEDGSA20, ANCHMEDGSA22, ANCHMEDGSA29, ANCHMEDGSA6, ANCHMEDGSA7, ANCHMEDGSA9, ANCHOBAYB --Europe/EU, ANCHOWA, RBRMIX, RBRMMEDGSA1-3 (Med sea), RBRMVI-VII-VIII, RBRMX,

* Next steps:
  + Find species that overlap with RAM New Zealand/Pacific and IUCN, should be at least 10 species
  + Start finding biological factors/data for them
  + Keep looking for good price/cost proxies or datsets

6/30/19

* New Zealand species: **Not sufficient numbers, will assess different region next (Pacific, may need to go global in scope or assess the need not to select for IUCN declining species?)**
  + Metanephrops challenger (Metanephrops challenger? LC and unknown): Scampi Bay of Plenty (SCMPBP), Scampi Wairapa/Hawke Bay (SCMPWHB)
  + Pseudocyttus maculatus (DD, decreasing): Smooth Oreo Bounty Plateau (SMOOTHOREOBP), Smooth oreo Chatham Rise (SMOOTHOREOCR), Smooth Oreo East Pukaki Rise (SMOOTHOREOEPR), Smooth Oreo Southland (SMOOTHOREOSLD), Smooth oreo West end of Chatham Rise (SMOOTHOREOWECR)
  + Hoplostethus atlanticus (assessed as VU and unknown trends in Euro?): Orange roughy New Zealand Challenger Plateau (OROUGHYNZ7A), Orange roughy Northwest Chatham Rise (OROUGHYNWCR), Orange roughy New Zealand Mid East Coast (OROUGHYNZMEC)
  + Pseudocaranx dentex (LC unknown): Trevally New Zealand Areas TRE 7 (TREVALLYTRE7)
  + Arripis trutta: Australian salmon New Zealand (AUSSALMONNZ)
  + Allocyttus niger: Black Oreo Pukaki Rise (BLACKOREOPR), Black oreo West end of Chatham Rise (BLACKOREOWECR)
  + Hyperoglyphe Antarctica: Bluenose New Zealand (BNSNZ)
  + Rexea solandri: common gemfish New Zealand (GEMFISHNZ)
  + Kathetostoma giganteum: Giant stargazer NZ Area STA7 (GSTRGZRSTA7)
  + Macruronus novaezelandiae: Hoki Eastern New Zealand (HOKIENZ), Hoki Western New Zealand (HOKIWNZ)
  + Genypterus blacodes: New Zealand ling New Zealand Areas LIN 3 and 4 \* (NZLINGLIN3-4), New Zealand ling New Zealand Areas LIN 5 and 6 (NZLINGLIN5-6), New Zealand ling New Zealand Area LIN 6b (NZLINGLIN6b), New Zealand ling New Zealand Area LIN 72 (NZLINGLIN72) , New Zealand ling New Zealand Area LIN 7WC – WCSI (NZLINGLIN7WC)
  + Chrysophrys auratus: New Zealand snapper New Zealand Area 8 (NZSNAPNZ8)
  + Haliotis iris: Black Foot Paua NZ North PAUA5A (PAUANPAU5A), Black Foot Paua NZ South PAUA5A (PAUASPAU5A), New Zealand abalone species New Zealand Area PAU 5B (PAUAPAU5B), New Zealand abalone species New Zealand Area PAU 5D (PAUAPAU5D), New Zealand abalone species New Zealand Area PAU 7 (PAUAPAU7),
  + Merluccius australis: Southern hake Chatham Rise (SOUTHHAKECR)
  + Nemadactylus macropterus: Tarakihi New Zealand (TARAKNZ)

7/1/19

* R workshop goals/ideas:
  + How to copy/paste large sections of excel without losing data?
    - My copy/pated data seems to be ok, same total rows between original and mine
  + How can I do my searches for info more efficiently?
    - See the row/column filtering functions/operations discussed, see our first tutorial stuff
  + Regression and trendline fit for something nonlinear
    - Wait and see if we even need this, might not…
* Pacific species:
  + ALBASPAC, Thunnus alalunga (Albacore Tuna, NT, decreasing)
  + BIGEYECWPAC, Thunnus obesus (Bigeye Tuna, VU, decreasing)
  + BLSHARNPAC, Prionace glauca (Blue Shark, NT, unknown if decreasing)
  + CMACKPCOAST, Scomber japonicas (Pacific Chub Mackerel, LC stable)
  + PSOLEPCOAST, Eopsetta jordani (Petrale Sole, LC stable)
  + LNOSESKAPCOAST, Raja rhina (Longnose Skate, LC stable)
  + PACBTUNA, Thunnus orientalis (Pacific Bluefin Tuna, VU decline)
  + SKJCWPAC, Katsuwonus pelamis (Skipjack Tuna, LC stable)
  + YFINEPAC, Thunnus albacares (Yellowfin Tuna, NT decreasing)
  + SPSDOGPCOAST, Squalus suckleyi (North Pacific Spiny Dogfish, LC stable)
  + SSTHORNHPCOAST, Sebastolobus alascanus (EN unspecified)
  + STFLOUNNPCOAST, Platichthys stellatus (Starry Flounder, LC stable)
  + STMARLINNEPAC, Kajikia audax (Striped Marlin, NT decreasing)
  + SWORDEPAC, Xiphias gladius (Swordfish, LC decreasing)
  + PHAKEPCOAST, Merluccius productus (LC, unknown)
  + ARFLOUNDPCOAST, Atheresthes stomias
  + BGROCKPCOAST, Sebastes melanostomus
  + BRNROCKPCOAST, Sebastes auriculatus
  + CHILISPCOAST, Sebastes goodei
  + CPRROCKPCOAST, Sebastes caurinus
  + LSTHORNHPCOAST, Sebastolobus altivelis
  + OFLOUNPAC, Paralichthys olivaceus
  + PCODNPAC, Gadus macrocephalus
  + POPERCHPCOAST, Sebastes alutus
  + PTOOTHFISHMI, Dissostichus eleginoides
  + RSOLEHSTR, Lepidopsetta bilineata
  + SABLEFEBSAIGA, Anoplopoma fimbria
  + SAURNWPAC, Cololabis saira
  + SNOWCRABNPAC, Chionoecetes opilio
  + SNROCKPCOAST, Sebastes diploproa
  + WLFLOUNNPAC, Tanakius kitaharae
  + WROCKPCOAST, Sebastes entomelas
  + CROCKPCOAST, Sebastes pinniger
  + DSOLEPCOAST, Microstomus pacificus
  + ESOLEPCOAST, Parophrys vetulus
  + GOPHERSPCOAST, Sebastes carnatus
  + GRNSTROCKPCOAST, Sebastes elongatus
  + GRSPROCKNCAL, Sebastes chlorostictus
  + KCROCKNPAC, Sebastolobus macrochir
  + LINGCODNPCOAST, Ophiodon elongatus
  + YTROCKNPCOAST, Sebastes flavidus
* Talked to Malin, might think of doing global scale review and try to get less profile info and more trends/analysis? Ask Ed.

7/2/19

* Today:
  + Get list of global species (used list of species I already looked at
  + Look at Ed’s data from paper (see email and paper citations)
  + Call Ed at 1pm
  + Practice R, try calculating timeseries means etc?
  + More literature search?
* Global list:
  + I took the stockID from my extracted list and searched for it in the stock tab in the full database. I copied over the latin names and used those as a search for the IUCN and deleted the species/stocks that were not in the IUCN list.
  + Should I also look and see how many are also decreasing or remove the DD/unknown ones? (questions for Ed)
  + Total species found: 53 (should run r and get a nice orderly list with exact number)
    - Run something to look for number of unique latinname and stockid
  + List (with the individual stocks too):
  + Should we take the mean Bmsy value for each species across all stocks?
* Meeting with Ed:
  + Global species list is 53 (just my counting, could use R to get exact in case I miscounted)
  + Multiple stocks for some species, should we take the average values? **Keep stocks separate!**
  + \_\_\_\_\_\_\_\_\_\_\_\_
  + COST/BENEFIT RATIO
  + PNAS paper didn’t look at biomass, but I wouldn’t do much to add
  + See how biomass relates to bio/econ factors
  + Focus on IUCN/RAM overlap
  + Look specifically at those species
  + Paper start with Juan’s simulation and analysis of endangered status matches simulation, refer to my work assessing this (co-author, then lead later?)
    - Creating useful doc/database can lead to a lot later
  + Pursue both a project and database
  + Record IUCN status for each of the RAM species, codify status as number—also trend
  + Count the number that have term endangered in their status (match the occurrence of a stock being endangered according to the status and matching with the model)
    - % species with similar characteristics that are endangered
    - Some measure of the number of species on record that have similar characteristics
      * If **on IUCN list**, species with **MSY** x tons and **cost/benefit** ratio has 3/6 endangered or critically endangered
      * Then compare to prediction, maybe 30% and compare the match between them
  + SAU prices still useful? Use cost/benefit data for later? Find this by region and match to RAM/IUCN,
  + Look for depth data for each species
  + Next steps: match RAM/IUCN, codify IUCN status, get depth data/cost/benefit/msy data over years, get another independent estimate of cost/benefit
    - Cost/benefit following Ed’s methods:
      * Use the excel from Ed’s data, can use it by
      * Check most likely gear type used for each species, get estimate for cost
      * Cost data tab has data by region, should use that
      * Pacific coast: North America
      * Tuna?: go to that stock and see which countries involved in management and then group them by general region (if multiple regions, average the regions)
    - Try to add notes to the excel, maybe make a new tab like in RAM
  + For each species: include cost data by country/region
    - Variable cost, subsidy, landed cost, landed values, cost by gear (can average if harvested in different ways)
    - Cost data tab B-E is raw data, use this and cost by gear B-D
  + Landing value/landing gets you average price
    - Could refine by gear specific price to see if fishing method has higher price/value
  + Start with the easiest data/cost columns and get data for all species and then move onto other columns
    - Record important harvesters for each species and then average cost/price etc
    - Can automate in R, recording **country and gear** involved and write a script that records this for stock and then add in stuff about cost data
    - Have separate column per gear type and then fill in the % that that species is caught with that method and then write a program to calculate the cost (write script that does the conversion)
    - Send Ed a blank excel sheet with all variable names as different columns before I start filling it in and the list of matches
* Moving forward summary:
  + ~~match RAM/IUCN~~
  + ~~codify IUCN status (by endangered status)~~
  + ~~make list of how many species have “endangered” in their status~~
  + ~~put together a blank excel sheet with the variables we will look at as column headers, send to Ed~~
  + ~~send Ed my Ram/IUCN overlap data~~
  + Find a way to quantify the relationship between endangered species/status and the model Ed and Juan are developing
    - % species with similar characteristics that are endangered (endangered status)
    - Some measure of the number of species on record that have similar characteristics
      * If **on IUCN list**, species with **MSY** x tons and **cost/benefit** ratio indicate 3/6 endangered or critically endangered
      * Then **compare to prediction,** maybe 2/6 (50% of IUCN and 30% or model are predicted as at risk) and **compare the match between them**
  + Start gathering data for major countries as managers for each stock
  + Start assessing gear type used for each stock
  + Can being to use the cost/gear data from Ed’s excel to put values into the gear/cost columns
  + get depth data/cost/benefit/msy data over years, get another independent estimate of cost/benefit
* the regional/management data is in RAM stock tab under areaid and region

7/3/19

* Today:
  + ~~List of countries/regions as managers of stocks~~
    - ~~Use areaid and region from stock tab, get exact country/region using the areaid and then the area/country data~~
  + List of gear types for stocks
    - So far not seeing much for this in RAM, may need to see if there’s any data in SAU for overlapping species and move forward with those, otherwise look at other sources that we have yet to discuss or at worst scarp the cost data.
  + Start to bring over the cost data by country (what to do about multinational designation? Scrap those stocks? Look into other sources?)
* Ed wants these parameters in the preliminary datasheet added:
  + Row specifying if columns are stock/species/country level
  + Top row indicating if factor is ecological or economic
  + Add harvest rate (U or F, F and Fmsy are available in RAM)
  + Add MSY
  + Add depth
  + Add distance from shore
  + Have separate B, Bmsy, and B/Bmsy columns
  + Separate B and F values into initial and average values for the time series (initial is the first value, average is of all years but the first)
  + Row for units?
  + Add initial and final year in RAM timeseries columns
  + Add price ($/volume) and corresponding harvest volume from the SAU if possible
  + Cost data: note if total cost ($) or marginal cost ($/volume)
  + \*\*\*standardize cost, check and see but it should be $US converted to year 2000
* **Find countries within these multinational regions:**
  + Mediterranean sea: Spain, France, Monaco, Italy, Slovenia, Croatia, Bosnia and Herzegovina, Montenegro, Albania, Greece, Turkey, Syria, Lebanon, Israel, Egypt, Libya, Tunisia, Algeria, and Morocco; Malta and Cyprus
  + South of Sicily: Tunisia, Libya, Malta, Egypt
  + Atlantic Ocean:
  + Northern Atlantic: Iceland, Greenland, Canada, USA, Mexico, Cuba, Haiti, Dominican Republic, Puerto Rico, Guatemala, Honduras, Nicaragua, Costa Rica, Panama, Cayman Islands, Jamaica, Bahamas, Turks and Caicos, St. Kitts& Nevis, Antigua and Barbuda, Montserrat, Guadeloupe, Dominica, Martinique, St. Lucia, St Vincent and Grenadines, Barbados, Grenada, Colombia, Venezuela, Trinidad and Tobago, Guyana, Suriname, French Guiana, Norway, Denmark, UK, Netherlands, Germany, Belgium, France, Ireland, Spain, Portugal, Morocco, Western Sahara, Mauritania, Senegal, The Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Cote d’Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Equatorial Guinea, Cape Verde
  + South Atlantic: Gabon, Republic of the Congo, Democratic Republic of the Congo, Angola, Namibia, South Africa, Argentina, Uruguay, Brazil,
  + Eastern Atlantic: Gabon, Republic of the Congo, Democratic Republic of the Congo, Angola, Namibia, South Africa, Norway, Denmark, UK, Netherlands, Germany, Belgium, France, Ireland, Spain, Portugal, Morocco, Western Sahara, Mauritania, Senegal, The Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Cote d’Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Equatorial Guinea,
  + Western Atlantic: Argentina, Uruguay, Brazil, : Iceland, Greenland, Canada, USA, Mexico, Cuba, Haiti, Dominican Republic, Puerto Rico, Guatemala, Honduras, Nicaragua, Costa Rica, Panama, Cayman Islands, Jamaica, Bahamas, Turks and Caicos, St. Kitts& Nevis, Antigua and Barbuda, Montserrat, Guadeloupe, Dominica, Martinique, St. Lucia, St Vincent and Grenadines, Barbados, Grenada, Colombia, Venezuela, Trinidad and Tobago, Guyana, Suriname, French Guiana
  + West Africa: Morocco, Western Sahara, Mauritania, Senegal, The Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Cote d’Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Equatorial Guinea, Gabon, Republic of the Congo, Democratic Republic of the Congo, Angola, Namibia, South Africa, cape Verde
  + West Africa Zone C:
  + Pacific:
  + North Pacific: USA, Canada, Mexico, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panama, Colombia, Russia, japan, Korea, China, Philippines, Vietnam, Indonesia, Malaysia, Cambodia, Thailand, Singapore
  + South Pacific: Ecuador, Peru, Chile, New Zealand, Australia, Papua New Guinea, French Polynesia, Cook Islands, American Samoa, Niue, Tokelau, Tonga, Tuvalu, Fiji, Vanuatu, Solomon Islands
  + Eastern Pacific: Ecuador, Peru, Chile, USA, Canada, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panama, Colombia
  + Western Pacific: Australia, Brunei, Cambodia, China, Cook Islands, Fiji, Japan, Kiribati, Malaysia, Marshall Islands, Micronesia, Mongolia, Nauru, New Zealand, Niue, Palau, Papua New Guinea, Philippines, Republic of Korea, Samoa, Singapore, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuata, Vietnam, Russia, Indonesia,
  + Northeast Pacific: USA, Canada, Mexico, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panama, Colombia
  + Central and Western Pacific: Australia, Brunei, Cambodia, China, Cook Islands, Fiji, Japan, Kiribati, Malaysia, Marshall Islands, Micronesia, Mongolia, Nauru, New Zealand, Niue, Palau, Papua New Guinea, Philippines, Republic of Korea, Samoa, Singapore, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuata, Vietnam, Russia, Indonesia,
  + Indian Ocean: South Africa, Mozambique, Madagascar, Tanzania, Kenya, Somalia, Yemen, Oman, Iran, Pakistan, India, Sri Lanka, Bangladesh, Myanmar, Thailand, Malaysia, Indonesia, Timor, Australia, Comoros
  + Southern Oceans: Chile, Argentina, New Zealand
* Moving forward:
  + Thursday:
    - Filter out the species that don’t have an MSY or MSYbest
    - Practice some of the R4 data science book exercises
    - Review the R workshops with Michelle and Joe
    - Pick out the MSYbest values for my species
    - Start finding the initial and averages for each stock (TB and F)
    - Work on country list when tired/burned out
  + Friday:
    - Finish country list
    - Start doing country data for all stocks/cost info
    - See if the cost data I’m using is standardized
    - If not standardized then do so

7/5/19

* Didn’t end up working yesterday, took the day off after all. Will try to get some work done Sunday to make up for it.
* Goals for today:
  + ~~Finish country list~~
  + Get MSYbest values for my species, ID the ones that don’t have that data
  + Do same for Fmsy and Flim values
* Mediterranean sea: Libya, Morocco, Tunisia, Egypt, Israel, Lebanon, Syria, Turkey, Albania, Croatia, Cyprus, Greece, Italy, Malta,
* South of Sicily: Libya, Tunisia, Egypt, Malta,
* Atlantic Ocean: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Morocco, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, South Africa, Togo, Suriname, Belgium, Denmark, France, Germany, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, UK, Canada, US, Antigua Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Jamaica, Haiti, Honduras, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Uruguay, Venezuela
* Northern Atlantic: Cameroon, Benin, Eq Guinea, Gambia, Ghana, Guinea, Cote d’Ivoire, Sao Tome Principe, Senegal, Sierra Leone, Togo, Suriname, Belgium, Denmark, France, Germany, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, UK, Canada, US, Bahamas, Barbados, Belize, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Jamaica, Haiti, Honduras, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Venezuela
* South Atlantic: Angola, Congo Rep, DCR, Gabon, Liberia, Mauritania, Namibia, Nigeria, Brazil, GuineaBissau, South Africa, Argentina, Chile, Uruguay,
* Eastern Atlantic: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, Sierra Leone, South Africa, Togo, Belgium, Denmark, France, Germany, Ireland, Netherlands, Norway, Portugal, Spain, UK,
* Western Atlantic: Suriname, Canada, US, Antigua Barbuda, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Jamaica, Haiti, Honduras, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Uruguay, Venezuela
* West Africa: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, Sierra Leone, Togo,
* West Africa Zone C:
* Pacific:
* North Pacific: China, Japan, Korea, Philippines, Taiwan, Vietnam, Russia, Canada, US, Micronesia, Palau, Papua N Guinea, Solomon Is., Colombia, Costa Rica, El Salvador, Guatemala, Haiti, Honduras,
* South Pacific: Australia, Fiji, Kiribati, New Zealand, Samoa, Tonga, Vanuatu, Ecuador, Peru,
* Eastern Pacific: Canada, US, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Peru, Panama, Nicaragua, Chile
* Western Pacific: China, Japan, Korea, Philippines, Taiwan, Vietnam, Russia, Australia, Fiji, Micronesia, Nauru, New Zealand, Palau, Papua N Guinea, Solomon Is., Vanuatu,
* Northeast Pacific: Canada, US,
* Central and Western Pacific: Australia, Fiji, Kiribati, Marshall Is, Nauru, Samoa, Tonga,
* Indian Ocean: Comoros, Kenya, Madagascar, Maldives, Mauritius, Mozambique, Seychelles, Somalia, Sudan, South Africa, Tanzania, Bangladesh, India, Indonesia, Malaysia, Myanmar, Pakistan, Sri Lanka, Yemen, Australia,

7/8/19:

* Today (DIDN’T GET MUCH DONE ON Friday and headache yesterday so lots to do today)
  + ~~Get MSYbest values for my species, ID the ones that don’t have that data~~
  + ~~Do same for Fmsy~~
  + Try to load timeseries into R and get initial and means for my species that remain after the above data collection
* Ed meeting:
  + Questions:
    - Should I include data for Flim if available? It represents the mortality threshold before stock decline.
    - Should I record the TBdivTBmsy and FdivFmsy data in my sheet? Should I check with the TBmsy and timeseries TB data that the listed TBdivTBmsy come out with the same results?
    - F= harvest rate
    - B (or TB) = biomass
    - MSY =max sustainable harvest
    - B/Bsmy = relative biomass
    - So we want countries listed in the database? I’m just thinking about the best way to include that when there are a lot of countries in each region listed for the stocks. I’m inclined to make a separate sheet with region and the countries we’ll include and then in the main data sheet just reference the region? Maybe set it up in a way that makes calculating the regional cost data easier?
    - Also, there are a lot of countries in each region. Should I try and use all of the countries or find a way to narrow it down for simpler calculations? What is more reproducible/reasonable?
      * Pick the ones with the most of the catch share, maybe use R to do this later, collect everything else later.
    - What do you want to go over tomorrow with Juan?
    - Do you have any advice for preparing for the SEES meeting?
  + Juan’s simulation results will show more about what we want to do for tests wth our data now
    - Hoping to explore human noise/variability and how that may influence extinction prediction
    - 100 species wit same characteristics with various x axis values
      * Increase in cost/MSY lower persistence of species, overharvest possibility after the critical point (in model, 1)
      * Also higher risk as you get lower (low cost, high price, low productivity)
      * Closer to the bifurcation the better the persistence
      * High temporal human decision at Fmsy, can either overharvest or under harvest
        + Higher than f/fmsy can lead to extinction
  + IUCN species should be on the higher or lower end of the x axis if threatened
    - Can’t just do a linear regression, maybe quickpoint analysis?
  + Test for bimodality, not useful don’t need
    - Extinction is bistable/multiple stability, but need to account for changing dynamics (human decision and equilibrium)
  + Recognize that all data is biased toward non-extinct species
  + Jude is helping to make their model more realistic, PNAS model assumes only one manager
* Adding Fmsy data REALLY cut down the stock list, likely cutting out several species too.
  + Stocks: 50
  + Species:

7/9/19

* Today and moving forward:
  + ~~Determine the countries in each region that make up 90% of the catch share, can use R?~~
    - Average the regions in contact with what’s specified (refer to the regions at bottom of the country cost table)
    - Go through all stocks, record region, record countries involved, figure out how to **use R to get the averages for the region**
      * **For loop for each stock region, find countries that correspond (copy paste the regions lists as a separate sheet and store, can try to just turn it directly into a vector?)**
      * **Get countries from region and store it in a vector(?) another for loop for this vector and assess cost value for each, store in another vector**
      * **Take the average of that vector to get average marginal cost**
    - Stats analysis: make model to predict marginal cost based on things like country/location, gear type, distance, depth, etc. look for relevant data to refine the model for cost (see Lam and Sumaila papers, they address similar things)
  + **Hypothesis: More of the endangered species should be on either end of the x axis, at least more on the edges than in the middle, predicting bimodality in terms of economic and ecological characteristics**

7/10/19

* Today:
  + ~~Email Kevin about fact sheet~~, Eli about NOAA price database
    - Sounds like we’ll probably still go with SAU though. The data quality might not be as good, but the resolution (global country) fits better with our cost/general data
    - This is the first test run of these methods, we’re just developing things. SAU will work well enough for our purposes. Using other data like NOAA/NZ/Australia management resources may require us to rethink/reduce our species list. I sent the email out anyway though, we’ll see what he has to say if he replies.
  + Collect distance, gear, ~~depth data~~ (SAU, FishBase?)
    - Don’t think I can get distance data from FB, other ways? This might be stock/fishery/effort specific, as different depth is probably the biggest factor and is related to underwater topography? In theory, you have to go farther out to go deeper (unless the continental shelf is close to the shore, but not always?)
    - Not sure what the best way to get the gear data would be either? Take a similar approach to selecting the main countries? Oof, but it’s over timeseries though, and with a number of other factors too…
  + ~~Tidy the database, make an organized notes section with where everything comes from and what it’s for~~
  + ~~Try to find good NOAA price database~~ (only if SAU doesn’t work or if the NOAA/Eli data fits better. Wait and see if he sends us anything good)
  + ~~Read through R materials and practice relevant stuff (see intro chapter of R for data science, review workshop files/notes, TAKE NOTES ON SYNTAX AND TOOLS/FUNCTIONS)~~
    - More or less did this, just started.
  + Try to load timeseries into R and get initial and means for my species that remain after the above data collection
* SAU global data:
  + No mention of species, it’s all by country/gear/use/ etc in a timeseries format. In the “landed\_value” it probably assumes all fish caught of the same end use, gear type, reporting, catch type, and fishing sector of the same country within a year. This is mixing fish species and I’m not sure if that’s the direction we want to go in
    - How will the cost/price data reflect anything specific about each stock? Or is that not the main goal? Is the fact that they are fished by those countries/regions enough of an indicator for stocks/species?
    - I suppose it’s our best bet though, as species/stock specific data is hard to find, especially globally for what we’re looking at. Would have to be looking at a more regional scale like PNW coast fisheries and using NOAA data or something similarly specific to get the good data that’s species/stock related

7/11/19

* Today:
  + ~~Take notes on R resources to get an idea of how to load stuff and familiarize with syntax and functions~~
  + Try to work out how I’ll setup the region/country data (make into objects?) and set up the for loop
  + ~~Lab meeting~~
  + ~~Meeting with Malin about price data~~
  + ~~Go through RAM and my other data files to collect the more specific country/region data (look for what says UAS, Canada, Australia, multinational… I know I have it somewhere, also need to track down which sheet from RAM I got it from)~~
* Malin Meeting notes:
  + Eli said that SAU might not be very reputable. It’s a fair concern but is it significant enough to have to look for alternative? Eli sent over an excel sheet that could be useful, and I’ve looked at other NOAA data. I don’t really think it’s what we’re looking for though.
  + The SAU global price data is organized according to a number of different factors and then by country in a given year. What is the best approach to this, given how I don’t think I can get much data for gear type for the stocks?
  + Should I add up all of the landed values for a given year for each country?
  + But Ed’s cost data from Sumaila also has a column for landed value? Why do we even need the SAU then?
* R stuff: use R to get the averages for the region
  + For loop for each stock region, find countries that correspond (copy paste the regions lists as a separate sheet and store, can try to just turn it directly into a vector?)
  + Get countries from region and store it in a vector(?) another for loop for this vector and assess cost value for each, store in another vector
  + Take the average of that vector to get average marginal cost
* Getting the more specific country/region lists:
  + List the stockid (take from my working list of species)
  + Find the assessorid that goes with that (go to bioparams and search the stockid to find associated assessorid)
  + Find the region associated with that assessorid (search assessorids in assessor and scroll over to find country/multinational)
  + Regions listed for multinational stocks can be found in stock tab

7/12/19

* Today:
  + ~~Follow directions above about getting the full region data, finish this~~
  + Write out notes/try to set up the country cost data collection using for loop
  + Write out notes/try to set up timeseries data using for loop
  + Keep reading through R for data science, taking notes on workshop worksheets
  + **Meet with Ed about how to move forward?**
  + **Look through Ed’s paper and data to see how he got his cost/price**
* Meeting with Ed:
  + In the RAM timeseries sheet, they have listed TBdivTBmsy and FdivFmsy columns. In theory they should be the same as if I calculate the stock TBmsy and Fmsy values divided by the average F and TB values. Should I use/reference these or ignore them?
  + I’m trying to figure out if there’s a way to put everything in one sheet, or the best way to use multiple sheets/files for the same for loop
    - Will determine how I make subsets
  + Multiple sheets? Manually delete?
  + Ed’s code:
    - Had whole RAM saved and then code read and extract it automatically
    - Select the data that has enough data points, read text in the file
    - Match IUCN manually for the RAM timeseries?
  + Most relevant part of RAM is stock and country/region
  + For each stock, just need cost/benefit ratio and MSY
    - Don’t have to look at biomass and harvest rate right now
      * Wanted them more for future work looking at actual human decision going into extinction (harvest rate, affecting biomass, affecting extinction risk)
      * Intermediate variables
  + Focus on the cost/price extraction right now
    - Extract country list, put into a list/subset
    - Get average cost/benefit value for those countries and take the average
    - (Cost+subsidy)/value, match to the stock
      * Would be easier to copy manually, but future expansion of database and reproducibility use code
      * Can still do analysis using R, histogram and plot the endangered species along the histogram (x axis =average cost/benefit \*ln(msy)) with any species remotely threatened
      * Scatterplot for IUCN status as y and x same as histogram?
        + Try to fit a regression for the scatterplot? Non-linear
  + Next steps: compile the data and get back to Ed
    - See I we can use Juan’s graph for my poster
  + Data analysis after next day or 2, then maybe get more data if time/interest/need
    - Can make histograms in Excel with 2019 version, could do on new workstation
* Country regions I need:
  + Mediterranean, N Atlantic, S Atlantic, S Pacific, S of Sicily, Atlantic, CW Pacific, E Pacific, Indian Ocean, E Pacific, Pacific, W Africa, E Atlantic, W Atlantic, SW Pacific (=CW Pacific)
* Steps to find initial and average values for timeseries stuff:
  + Set working directory: Session --> Set Working Directory --> Choose Directory
  + Import data:
    - view the file names in the directory with dir()
    - whatyouwanttocallthedatafile <- read.csv(file=’type the exact file name here’, header=T)
  + check that the data is what you want using str(whatyouwantocallthedatafile)
  + make list of all the stockid names I need, make a subset
    - usedstocks <- subset(x = RAM, subset = stockid == c('ALBAMED',… type out the whole damn list) , select = c(TB,F))
  + For loop to collect: tbinitial (not sure how to get that? Can just look in excel and ctrl+F, same for year stuff), tbyearinitial, tbyear last, tb average (do a mean for the subset where i=stocks from the subset)
    - Make list for stocks: stocks = unique(RAM$stockid)
    - For (i in 1:length (stocks)) {

#subset the data

stocks.sub = subset(RAM, stockid==stock[i])

#calculate

stock.mean.tb = mean(stock.sub$tb)

stock.mean.f = mean(stock.sub$f)

#record values

results$mean.length[i] = stock.mean.tb

results$mean.length[i] = stock.mean.f

}

**\*\*\*\*\*\*\*\*** to import all excel sheets:

library(readxl)

read\_excel\_allsheets <- function(filename, tibble = FALSE) {

# I prefer straight data.frames

# but if you like tidyverse tibbles (the default with read\_excel)

# then just pass tibble = TRUE

sheets <- readxl::excel\_sheets(filename)

x <- lapply(sheets, function(X) readxl::read\_excel(filename, sheet = X))

if(!tibble) x <- lapply(x, as.data.frame)

names(x) <- sheets

x

}

Or:

library(readxl)

path <- readxl\_example("datasets.xls")

sheetnames <- excel\_sheets(path)

mylist <- lapply(excel\_sheets(path), read\_excel, path = path)

# name the dataframes

names(mylist) <- sheetnames

plus If you want to bring the dataframes out of the list use the next bit of code.

# Bring the dataframes to the global environment

list2env(mylist ,.GlobalEnv)

* Steps to make a for loop to calculate all the values I need from the RAM\_Analysis file from Ed
  + Set working directory: Session --> Set Working Directory --> Choose Directory
  + Import data:
    - view the file names in the directory with dir()
    - whatyouwanttocallthedatafile <- read.csv(file=’type the exact file name here’, header=T)
    - check that the data is what you want using str(whatyouwantocallthedatafile)
  + create subset for each region:
    - pacific <- subset(x = costprice, subset = country == c('Angola',… type out the whole damn list) , select = c(cost, etc))
      * #I don’t think there’s a better way to automate this step so that I don’t have to make each list on its own.
      * #this is to get the subset of those countries and their values for each of the columns I select for in the vector c()
* Country/region lists:
* ~~Mediterranean sea: Libya, Morocco, Tunisia, Egypt, Israel, Lebanon, Syria, Turkey, Albania, Croatia, Cyprus, Greece, Italy, Malta,~~
* ~~South of Sicily: Libya, Tunisia, Egypt, Malta,~~
* ~~Atlantic Ocean: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Morocco, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, South Africa, Togo, Suriname, Belgium, Denmark, France, Germany, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, UK, Canada, US, Antigua Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Nicaragua, Panama, Jamaica, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Uruguay, Venezuela~~
* ~~Northern Atlantic: Cameroon, Benin, Eq Guinea, Gambia, Ghana, Guinea, Cote d’Ivoire, Sao Tome Principe, Senegal, Sierra Leone, Togo, Suriname, Belgium, Denmark, France, Germany, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, UK, Canada, Mexico, US, Bahamas, Barbados, Belize, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Jamaica, Nicaragua, Panama, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Venezuela~~
* ~~South Atlantic: Angola, Congo Rep, DCR, Gabon, Liberia, Mauritania, Namibia, Nigeria, Brazil, GuineaBissau, South Africa, Argentina, Chile, Uruguay,~~
* ~~Eastern Atlantic: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, Sierra Leone, South Africa, Togo, Belgium, Denmark, France, Germany, Ireland, Netherlands, Norway, Portugal, Spain, UK,~~
* ~~Western Atlantic: Suriname, Canada, Mexico, US, Antigua Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Grenada, Guyana, Haiti, Honduras, Jamaica, Haiti, Honduras, St Kitts, St Lucia, St Vincent, Trinidad and Tobago, Uruguay, Venezuela~~
* ~~North Pacific: China, Japan, Korea, Philippines, Taiwan, Vietnam, Russia, Canada, Mexico, US, Micronesia, Palau, Papua N Guinea, Solomon Is., Colombia, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama,~~
* ~~Indian Ocean: Comoros, Kenya, Madagascar, Maldives, Mauritius, Mozambique, Seychelles, Somalia, Sudan, South Africa, Tanzania, Bangladesh, India, Indonesia, Malaysia, Myanmar, Pakistan, Sri Lanka, Yemen, Australia,~~
* ~~South Pacific: Australia, Fiji, Kiribati, New Zealand, Samoa, Tonga, Vanuatu,~~
* ~~Eastern Pacific: Canada, US, Mexico, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Peru, Panama, Nicaragua, Chile~~
* ~~Western Pacific: China, Japan, Korea, Philippines, Taiwan, Vietnam, Russia, Australia, Fiji, Micronesia, Nauru, New Zealand, Palau, Papua N Guinea, Solomon Is., Vanuatu,~~
* ~~Central and Western Pacific: Australia, Fiji, Kiribati, Marshall Is, Nauru, Samoa, Tonga,~~
* ~~West Africa: Angola, Cameroon, Congo Rep, DCR, Benin, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Cote d’Ivoire, Liberia, Mauritania, Namibia, Nigeria, GuineaBissau, Sao Tome Principe, Senegal, Sierra Leone, Togo,~~

7/14/19

* ~~Today: finish calculating regional averages~~
* Tomorrow:
  + Add those averages into the datatable
  + Try to figure out how to do a histogram
  + If bored, do more of the TB and F stuff by hand

7/15/19

* Today:
  + ~~Add other countries to the lists~~ ~~(Djibouti, Eritrea, Saudi Arabia, UAE, Iran to Indian Ocean) (Guatemala to N and W Atlantic) (Sweden and Poland to N and E Atlantic)~~
    - This classification is kind of hard and not super standardized? I kind of just looked at the map and if a country was behind a small passage to get to that location or kind of far up that passage then it didn’t count
    - It would also be great if I could find a way to say countries within x miles from the shore/coast of that ocean section are included, but IDK and don’t have time to figure that out.
    - Or a way to know which countries fish there or is they just stay in the sea/gulf closest (does Mexico go out to the Atlantic, is it right to kind of say that the Gulf of Mexico and Caribbean are basically Atlantic?)
  + ~~Redo the country cost values in data table~~
  + ~~Make data .CSV and load into R~~
    - I made a new file with just the data we want, the first section and saved it as a CSV in dropbox “data” as RAM\_IUCN\_DATA
  + ~~Take notes on plotting from r4data science and look up histogram specific support elsewhere~~
  + Make plans with Michelle to go over using Git and what I need to put on there
* Questions for R:
  + How do I tell R what row the column name starts at? Otherwise I think it starts at line 1 which is wrong (I need line 4)
    - I just cut out the other stuff, it was easier and makes the most sense
  + ~~Can I run natural log or should I make a data column in the excel/csv for the cost/price ln(msy) thing?~~
    - Maybe go in and make a new column for this? Might be simpler?
    - I think I have to do this? And it’d still be simpler either way.
  + Is there a specific plotting function to make a histogram, or are we kind of just manipulating a bar chart to suit our needs?
* Unit for x value: average cost/benefit \*ln(msy))
* Cost unit: (cost+subsidy)/value
* Tomorrow:
  + Rename the columns above in the datafile so it’s easier to work with in R, make good notes for this in the big excel data file
  + Play around and keep trying different types of plots
  + Look more into histograms, try to plot the x variable we want and test out different Y variables?
  + Try some scatter plots?
    - IUCN code x MSY, IUCN code x costs, IUCN costsln(MSY), region x and cost/MSY values, region x IUCN code,

7/16/19

* Today:
  + ~~Rename the columns above in the datafile so it’s easier to work with in R, make good notes for this in the big excel data file~~
  + ~~Look more into histograms, try to plot the x variable we want~~
    - Test out different Y variables?
  + Put code on Git! Ask Michelle if running into issues
  + ~~Try some scatter plots?~~
    - IUCN code x MSY, IUCN code x costs, IUCN costsln(MSY), region x and cost/MSY values, region x IUCN code,
  + ~~Play around and keep trying different types of plots~~
  + ~~Look into bimodality tests?~~
    - Nothing super obvious, or ones that are apparently compatible with R yet
  + ~~Write out a few practice talk runs (put best ones on notecard to practice with tomorrow)~~
  + ~~Work on internship report when tired/out of focus~~
* Renamed “(cost+sub)/value” to “cost\_value”
* Renamed “((cost+subsidies)/value)\*ln(MSY)” to “final\_value”
* Made a new project and connected it to Git
  + It was empty and I opened the analysis file I’d been using, not sure where to go from here? Not much is showing up in the Git repository
    - Do I need to commit? I think I did and nothing showed up despite along processing thing popped up
* What would be the units for ((cost+subsidies)/value)\*ln(MSY) when cost is in US$ 2005/ton and MSY is metric tons?

7/17/19

* Today:
  + Put code on Git! Ask Michelle if running into issues
    - I tried doing it myself, but I’m really not sure what I’m doing here… I’ll ask Michelle when she gets in
  + ~~In the IUCN\_status column, remove or create new column for the decreasing/increasing/etc info, it’s adding too many columns to my graphs~~
    - New column? It’s one other variable I could look at
  + ~~Keep improving histograms/other plots~~
    - ~~Add nicer axes and titles, colors~~
  + ~~Look into bimodality tests~~
  + Work on Git README file, update it for what I want
  + Work on internship report when tired/out of focus
  + Practice talk for tomorrow (out of office in Cape May)
  + Make some of the edits from the meeting with Ed!
* Ed meeting:
  + What would be the units for ((cost+subsidies)/value)\*ln(MSY) when cost is in US$ 2005/ton and MSY is metric tons?
    - No units, $/kg, ln is always unitless (multiply tons by 1000, if do this reference the PNAS paper)
    - Convert MSY, if I do anything, look into this now
    - Have a ln(MSY) in kg
* I’m not sure what we could do to add a factor to the histogram Y axis or how that would help? Default options are frequency or proportion density (it only changes the axes)
  + - He meant that dividing threatened counts by total counts (have a species proportion ranging from 0-1, manifesting ratio as Y axis (see threat ratio notes)
  + It’s hard to tell if the shape of the histogram matches yours very well?
  + Opinion/impression of graphs?
  + When would be good to try and have all analysis done?
    - Try to have done in a week and a half
    - Theoretical expectations, expect bimodality? How present in poster?
      * Reference Ed’s work, compare to mine? Explain in words
      * Focus more on the data and verbally represent hypothesis
      * Most of project data collection, present data vis
  + Bother with bi/multi-modal test yet?
    - Come back later once I have more plots
    - Bimodality for non-standard stats, not used often
  + Do histogram with just threatened species
  + Do just cost/benefit (economic) and lnMSY(ecological) scatterplots
  + Make status is order of how threatened, not alphabetical
    - On plot axis, reorder them to show increasing threat
    - Have legends to explain the color differences too
  + Make histograms and remove lower threat species and keep removing until get to the last CR (see which gives best impression)
    - Just make a ton of histograms
    - Don’t need the curve, but it’s nice
  + See if occurrence of threatened species is hard, sampling bias b/c more records of species in certain parameter space and not indicative of risk
    - Can correct for it by dividing occurrence of threatened species by the occurrence of actual reports?
    - Have histogram with same number of breaks, threatened species only, then subset by full set to get ratios closer to extinction risk (ratio of vulnerable to all reports gives extinction risk? At x=5 there might be 7 total species but if all 7 were threatened then 100% of species in that range are threatened so high extinction risk—but if only 1 was threatened, 1/7 ratio lower risk)
    - New vector with number of counts per bin, just fewer points (divide that new vector by original vector and get extinction risk ratio)
  + Do this later, but keep in mind and try to get this done: Plot all RAM species, then get cost/benefit and MSY to show the whole RAM histogram, get ratio for extinction risk with my subset and compare to threatened subset
    - Assess how many species are persisting despite fishing!
    - Divide threatened stock by this distribution, might see bimodal pattern
  + Think about the database and data bias (why are they included, what extinction threats face each species/stock)
  + Try with different break numbers while making new histograms (fewer breaks means less resolution… higher is better but there’s an upper limit)
  + For MSY x status it should be ln(MSY) and x country, make new column
    - Biases: bigger MSY stocks could be more threatened, but should be multiplied by cost/benefit so this isn’t the best story
  + Scatterplot with cost with final value and IUCN status on y
    - A good look at raw data
    - Look out for if most of threatened species cluster around one “price point”
* Ed meeting note Summary:
  + Go in and edit database/working data:
    - Make a new column next to MSY that is ln(MSY\*1000) which will be unitless, but will essentially have MSY converted from metric tons to Kg which is what Ed used (trying to be consistent in matching up with his PNAS paper)
      * He’s not 100% sure on needing to convert, but it shouldn’t make a super huge difference anyway since the \*1000 will be log-ed anyway
  + Replot all the MSY plots with the new converted ln(MSY) values instead
  + In all histograms, try with as many bins/breaks as you can, just see what looks best
    - After some testing around, figure out what works best and make all histograms in that bin size (try to be consistent to make all histograms comparable)
    - More is better(higher resolution) but not always the best fit for what we’re looking for/the data
  + Make scatterplots of just the ln(MSY), final\_value, and cost\_value
  + Make histograms for the threatened status species:
    - Make a histogram with all threat levels (already have this)
    - Make a histogram with only those considered at risk (VU, EN, CR)
    - Make a histogram with just VU/EN species
    - Make a histogram with just EN and CR
    - Make a histogram with just VU and EN species
  + From the VU, EN, and CR histogram, try to find out ratio with the overall dataset with same bin size, use for y axis (is this possible, how?) to show threat ratio
    - This will take a lot of working things out and maybe help with Ed and Michelle
  + Get the cost/value and MSY data for all the RAM stocks/species and make a histogram for that
    - Compare to the threatened species histogram as well as the one for my data set

7/19/19

* Today:
  + ~~Keep editing the main data, see meeting summary notes~~
  + Remake the MSYx plots with ln(MSY) instead
  + Rename all plots to follow set naming (plottype\_x-variable\_y-variable)
  + Start making new histograms
    - Make subsets, vectors, other lists etc to help facilitate this as needed
  + Meet with Michelle a bit?
  + Work on Git README file, update it for what files I put on there/explain the work and make it a nice landing page for my research (something I want to link on my resume)
  + Work on internship report when tired/out of focus
    - Self-reflection and methods sections
* List of plots I should make:
  + plot(datafile$IUCN\_codified)
  + ~~plot(datafile$final\_value)~~
  + ~~plot(datafile$lnMSY)~~
  + ~~plot(datafile$cost\_value)~~
  + plot(x=datafile$countries, y=datafile$IUCN\_status, xlab='Management Country', ylab='IUCN Status', main="Proportion of IUCN Status by Managing Country")
  + plot(x=datafile$countries, y=datafile$lnMSY, xlab="Management Country", ylab="ln(MSY)", main="ln(MSY) by Country Manager")
  + plot(x=datafile$countries, y=datafile$cost\_value)
  + plot(datafile$IUCN\_status, y=datafile$lnMSY, xlab="IUCN Status", ylab="ln(MSY)", main="ln(MSY) and IUCN Status")
  + plot(x=datafile$final\_value, y=datafile$IUCN\_codified)
  + plot(x=datafile$IUCN\_codified, y=datafile$final\_value)
  + plot(x=datafile$IUCN\_status, y=datafile$final\_value, xlab="IUCN Status", ylab="((cost+subsidy)/value) \* ln(MSY)", main="IUCN Status and Socioeconomic Value", col=4)
  + hist(datafile$final\_value,breaks=5, freq=F, col=4, xlab= "((cost+subsidy)/value)\*ln(MSY)", main="Breaks=5")
  + hist(datafile$final\_value,breaks=10, freq=F, col=4, xlab= "((cost+subsidy)/value)\*ln(MSY)", main="Breaks=10")
  + hist(datafile$final\_value,breaks=20, freq=F, col=4, xlab= "((cost+subsidy)/value)\*ln(MSY)", main="Breaks=20")
  + hist(datafile$final\_value,breaks=20, col=4, prob=T, xlab= "((cost+subsidy)/value)\*ln(MSY)", main="Density, Breaks=20", xlim=c(0,30))
  + lines(density(datafile$final\_value, na.rm=T),lwd=2, col="black")
  + histogram of VU
  + Histogram of EN and CR
  + Histogram of VU, EN, CR
  + Histogram of CR,EN, VU, NT (no DD or LC)
* List of good plots, save code for when adding them at the top of the file:
* This was a very low-productivity day. Will try to get the remaining to-do items done over the weekend

7/22

* Today:
  + ~~Keep making/improving graphs~~
  + ~~Remake the MSYx plots with ln(MSY) instead~~
  + ~~Rename all plots to follow set naming (plottype\_x-variable\_y-variable)~~
  + ~~Start making new histograms, fix boxplots~~
    - Make subsets, vectors, other lists etc to help facilitate this as needed
    - Made improved box plot of the final value and status!
    - Made histogram of just threatened species (CR,EN,VU)
  + ~~Work on Git README file, update it for what files I put on there/explain the work and make it a nice landing page for my research (something I want to link on my resume)~~
  + Work on internship report when tired/out of focus
* Didn’t get any work done over weekend, need to make progress today and tomorrow!
* **\*\*\*Figured out boxplot issues, make all of those today!**
* **Try to figure out histograms**

7/23

* Today:
  + ~~Keep improving graphs~~
    - Try to get more histograms
    - ~~Ask for name of the weird plot, so I can look up how to edit it/options~~
    - ~~Come up with better color palettes~~
    - Try to get all graphs done that Ed asked for, minus the ratio stuff, no idea
    - Ask about how to do the ratio thing that Ed wanted to see
  + Work on Git README file, update it for what I want
    - Find link to Ed’s paper and insert
    - Figure out how to put other files (excel files) on Git (maybe wait until project is done and all files are essentially complete)
    - Figure out more formatting stuff to make it pretty
  + Work on internship report when tired/out of focus
  + **Get the cost/value and MSY data for all the RAM stocks/species and make a histogram for that**
    - Compare to the threatened species histogram as well as the one for my data set
* Graph colors:
  + Blue: darkcyan
  + Blue scale: c("turquoise4", "turquoise3", "turquoise2", "turquoise1", "turquoise", “paleturquoise”))
    - That doesn’t show up nicely on a boxplot, find other color group:

c(“midnightblue”, “navy”,”mediumblue”, “blue3”, “blue”, “steeleblue1”)

c(“paleturquoise4”, “turquoise4”, “lightseagreen”, “cyan3”, “cyan”, “aquamarine”)

* Working with the “whole RAM” database is actually a lot trickier than I thought it’d be, the stockid and other parameters are not all arranged by stock alphabetically, bust sometimes by other parameters alphabetically or there are more stocks??? I’ve been taking my stockid list from bioparams since that’s where I take the MSY from but there are only 629 stockids but in stock there are 1373?
  + Bioparams\_values\_views sheet: 629 stockid and MSY (less MSY points, not all stocks have it)
  + Stock sheet: 1373 stockid, areaid, region
  + What I need to do:
    - Sheet: A=stockid, B=areaid, C=country, D=areaname, E=region estimate, F= cost/price value, G=MSYbest H=ln(MSY\*1000),
    - Create a list with stockid and
  + Getting the more specific country/region lists:
    - List the stockid from bioparams (the MSY is really what we need/will be plotted)
    - Find the assessorid that goes with that (go to bioparams and search the stockid to find associated assessorid)
    - Find the region associated with that assessorid (search assessorids in assessor and scroll over to find country/multinational)
    - Regions listed for multinational stocks can be found in stock tab

7/24

* Today:
  + ~~Histogram for all RAM~~
    - ~~Get country, region, cost/price, msy, lnMSY, final value~~
    - ~~Follow directions up top for how to get this all~~
      * ~~OR JUST AUTOMATE IT (see the FULL\_RAM\_MSY.R file)~~
  + Try to have all graphs ready to go
    - Have a ton of them, go through them tomorrow and Friday
    - Fix, add \*1000 to any mention of MSY
    - Add “n=” above the bars in spineplot (say specifically what coordinates you want the text at)
    - Improve the colors and point type/notch/border/etc
    - Ask about how to do the ratio thing that Ed wanted to see
  + Work on internship report when tired/out of focus
* Canada E Coast: ~~Canada, USA, Iceland~~
* EU: ~~Sweden, Finland, Estonia, latvia, lithuania, poland, Germany, denmark, netherlands, UK, Ireland, france, spain, portugal, italy~~,croatia, greece, malta, cyprus
* Euro non-EU: ~~Russia~~, ~~Albania~~, ~~Norway~~
* NE Pacific: ~~Russia, Japan, China, Korea, Philippines, vietnam, Indonesia~~
* SW Pacific: ~~Ecuador, peru, chile~~
* S America: ~~Colombia, Eciador, Peru, chile, argentina, uruguay, brazil, suriname, guyana, venezuela~~
* Antarctic: ~~chile, argentina, South Africa, Australia, NZ~~

7/25

* Today:
  + Try to finish up analysis
    - Have nice plots saved for viewing, PDF is higher resolution but PNG is better formatting? We’ll have to see if that holds up by saving things in both formats (maybe delete the PNGs and remake them as necessary for the right dimensions later if I have to?
  + Go to workshop to learn poster makings
  + Start a general template for poster
  + Lab meeting
  + Ed meeting
  + Make some adjustments, but go home
* Ed meeting:
  + Go over notes/expectations from the workshop earlier
    - 3x4 ft poster
    - Logos should be sent over soon
  + Do you have a template/advice for how to make the poster? Powerpoint or inkscape? Keynote?
    - Ed suggested keynote as a good option for good resolution
  + What are our expectations for the poster?
    - Explain what the problem is, general for fisheries/conservation and technical problems of getting relevant data for these extinction hypotheses (intro)
  + What should I focus on in terms of theory or anything else?
    - Talk about the different hypotheses out there verbally, body size correlation, and more, talk about ecological factors and the database background
  + What should be some of the central figures? Are any of the ones I have so far good, should I keep refining them?
    - Histograms as focus,
    - Others are good too
  + \*\*\*issue I’m running into, the histogram for IUCN stock/final\_value dataset is coming out different (14 breaks instead of 20 like I ask it to).
  + Look through the plots, pick out the really good ones and go over edits that others need, which can kind of be discarded
  + What are the best next steps, any more analysis?
  + What should I refer to ((cost+subsidies)/value)\*ln(MSY) as? “Bioeconomic factor” or just state it as is?
  + I took notes on paper, easier than when also trying to look at graphs typing here.
* Ed meeting notes summary:
  + Biggest takeaways:
    - FOCUS ON HISTOGRAM RATION THING
    - Then make scatterplots
      * Cost/benefit on x, status on y (scatter not box plots)
      * Cost/benefit and lnMSY
      * Cost/benefit on x axis and threat on y (to prevent frombeing a boxplot, it’ll be a scatterplot)
    - Convert spineplot to stacked barplot and make more barplots that are similar
    - Make text larger for all plots, also reduce the actual plot size (without making text smaller) to make them more manageable
  + Histogram ratio:
    - Compare across FULL, IUCN, and the subset I made that’s called “threatened” in the IUCN R file
    - Basically see if I can use R or excel to collect the n value (frequency?) for each bin (need to standardize the bins though somehow) and then one by the other to get stuff
    - Need to set all bin s to the same size/sections for the finalvalue values so they’re standard and easier to compare across
    - Ed said something about using a for loop for this, but it didn’t make sense and the video kept cutting out
    - Base everything on 10 breaks?
    - Use frequency for everything
    - For the threatened subset, there aren’t any finalvalue values that go up to 45, may need to arbitrarily put something at 45 and then say not to include it? I’m confused but ok.
    - I might be dividing with vectors, might need to use notation “./” inseatd of just “/”
  + Remove title from all plots I’ll use, don’t worry about making them for new plots. I’ll just explain everything in my captions anyway, don’t worry.
  + Refer to ((cost+subsidies)/value)\*ln(MSY) as relative cost \* potential productivity
  + The index on the scatter plots could probably be removed, but also just don’t use those. Just plot according to the scatterplots we talked about. The index was just the number for the stock (assigned row number, basically) and has no real meaning.
  + Make the text larger for all plots, also see if I can reduce the actual plot space so that they’re more condensed and fit better. At the very least, MAKE TEXT LARGER

7/26

* Today:
  + Start working on poster! Try to have a first draft ready, drawn on paper at least (bring notecards!)
    - ~~Get basic template/format going, title, sectioning, etc~~
    - Determine which plots to put in the poster (scatterplots, etc)
  + Make scatterplots
    - See the one Ed made and put in dropbox and make more like that, see the meeting summary above
    - ~~Cost/price vs IUCN code~~
    - ~~Finalvalue vs IUCN code~~
    - ~~Cost/price vs lnMSY~~
    - ~~Save those plots~~
  + Work on getting the histogram stuff figured out, ask for help and check in with Ed before 1pm
    - ~~Fix font size, make larger~~
    - ~~Set the bins so they’re at the same intervals, easy comparisons for frequencies~~
    - ~~Collect frequencies for each bin for each plot~~
    - ~~Put in an excel sheet~~
    - ~~Divide them by each other and organize those values in a way that can be used by R~~
    - Plot these and see how they look, bar plot (find way to make NA and 0 look different!)
  + Figure out how to weight the risk level? See emails with Ed from today
* Yes, I went in and used log() instead of ln() and it produced finalvalues of under 10. I don’t recall using log(), but at any rate I know that the values are right, as they are now using ln()
* Use to increase font size: cex.axis=1.5, cex.lab=2
* I’ll probably need to work over the weekend, but that will probably just be poster content, practicing my talk, and trying to finish everything else in R
  + Background section
  + Talk briefly about/make flowchart for methods
  + Try and add to conclusions?
  + Acknowledgements
  + References
  + Make the barplots for the ratio data I calculated
  + Try to make the spineplot a stacked barplot with help from Jennifer’s git repo

7/29

* Today:
  + ~~Plot the histogram ratio data and see how they look, bar plot (find way to make NA and 0 look different!)~~
  + ~~Figure out how to weight the risk level? See emails with Ed from Friday and Sunday~~
    - For bin for cost/benefit\*ln(MSY)=10 [7.5 to 12.5?]: risk=sum(#extinct x 1+ #extinct in wild x 4/5 + …)/sum(#records in RAM or in IUCN list)
      * EX=1, EW=4/5, CR=3/5, EN=2/5, VU=1/5, NT/LC/DD=0
      * In R : sum for IUCN (1679.752) and for FULL (4954.48)
      * In a bin means +/- 2.5 from the bin assigned number?
      * Should I do:
        + (#EN\*(2/5)+#VU\*(1/5))/sum( #stocks of all FULL RAM or IUCN list, NOT the sum of final\_value)
        + (#EN\*(2/5)+#VU\*(1/5))/sum(final\_value of all FULL RAM or IUCN list)
        + ((2/5)\*(final\_value sum for all EN)+(1/5)\*(final\_value sum for all VU))/ sum(final\_value of all FULL RAM or IUCN list)
      * Problem? The only species in the FULL RAM that have IUCN status listed are ones that are in the IUCN subset… so doing the weighting will actually only change by the sum of all records
    - Plot as barplot similar to the bin ratios
  + **Have a poster draft ready for tomorrow, edit what I made on Friday/Saturday/Sunday**
  + ~~KEEP WORKING on plots/weighting~~
  + ~~Make edits to plots as needed for poster (size, color, png, etc)~~
  + ~~Check in with Ed about poster/presentation tomorrow if he wants to skype in~~
  + ~~Make plots a new color that matches poster better~~ 
    - **“royalblue4”**
* Ed meeting sumary:
  + Change the bin ranges on all histograms/dependent things to interval of 5 (7, 11, 16, etc) because we want to illustrate the model’s prediction that 11 is a tipping point of sorts
    - But wait to do this until after tomorrow, use current breaks but talk about how that will change a little bit
    - This means updating the histogram ratio (find the frequency and update the histinfo objects to get this) and those graphs
    - Update the weighting intervals/data
  + For the weighting, only base it on the number of stocks (I have this already set up in general, see the need to update things in the above point)
    - (#EN\*(2/5)+#VU\*(1/5))/sum( #stocks of all FULL RAM or IUCN list, NOT the sum of final\_value)
  + Try to make a stacked barplot of the IUCNscatterplot\_finalvalue\_IUCNcodified
    - X=final\_value, y=status and shown with legend
  + Title:
    - Quantifying Extinction Risk in Commercial Marine Fisheries
    - This is from the interest in emphasizing the role of data/collection
  + Intro:
    - Use terms economic value, potential productivity (bold these) for cost and MSY (but the actual things in parentheses)
    - Remove the Used RAM part, cut it out
    - Say that the economic value and potential productivity are correlated with extinction risk
    - Make text larger
  + First figure/hypotheses:
    - H1: (value/productivity) ∝ (is proportional to) extinction risk
      * High value and low productivity
    - H2: (productivity/value) ∝ (is proportional to) extinction risk
      * Low value and high productivity
      * Cite Ed’s paper, see same citation scripting as for Lam below in methods
  + Methods: (call Data)
    - Bullet points only
    - RAM Legacy, variables collected
    - IUCN (but spell it out) Red List, status (see their language on how to describe status... extinction? Threat?)
    - Lam et. al, economic value (cost+subsidy/price) (cite this! Use a superscript or subscript number to illustrate this)
  + Results figures:
    - Label all plots with the same x axis
      * Productivity/value (give the full final\_value formula)
    - THREATdivIUCN
      * Refer to as “fraction of commercial species threatened”
      * Rename file to show that it’s not histogram
      * Explain:
    - THREATdivFULL
      * Refer to as “fraction of IUCN listed commercial species threatened”
      * Rename file to show that it’s not histogram
      * Explain:
  + Conclusions:
    - ??? Ed had to go before we could talk about this, but I imagine it’s related to the hypotheses
  + References:
    - Use end notes format basically to indicate citation (Lam and Ed)

7/30

* Today:
  + ~~Prep everything in first hour to get ready for poster presentation at lunch~~
    - ~~Last changes to poster layout/content~~
    - ~~Practice talk (3 mins ish) in lab room~~
  + ~~Present!~~
  + Make changes that people suggest, edit poster
    - ~~Increase axis size, axis up to 2x? See if there’s a way to avoid weird cutoff~~
    - Good layout and text amount/size, in general, no need to redo this
    - State the main question big and large somewhere
      * Essentially we’re trying to link threat level to extinction risk? Linking the prod/value variable to threat? To extinction?
    - Reword the hypotheses:
      * H1: productivity/value increases with extinction risk
      * H2: productivity/value decreases with extinction risk
      * (the idea behind this is to use a consistent version or prod/val or val/prod, it’s confusing to use two different things.
      * Make a little diagram showing where dif types of fish would be (high prod/low val anchovy on left, low prod/high val tuna on right)
    - Is high economic value high on the x axis (just making sure, there was confusion about that and I had trouble/second guessed my understanding)
      * High value (high cost+subsidy/price) is to the right side of the x axis?
    - Find a better term for “economic value” maybe, read Ed’s paper to see if he refers to it at all as something else?
      * Use the hypothesis graphic to help explain this
    - Be clearer about the differences between the two middle plots. Why have the bottom one? Is it to show that we can use this variable for any species, even if we don’t know its status?
      * Have this in the conclusion section
        + Have a little version of the question next to each plot/as the title:

Top: out of the known status IUCN stocks, is the variable a good predictor of extinction risk?

Bottom: out of all stocks, is the variable a good…?

* + - Move the full equation out of the plot axes, put it nearer to the background or the hypotheses
    - For the threatened histogram, also have a little question thing telling what it’s about
    - Stacked barplots for the middle plots? Do I have time?
    - Be more optimistic/open in the conclusion, the first part sounds really dry and dismissive
      * Focus on the second part talking about “as extinction risk is similar at both low and high”
      * Focus on what we do observe, nit what we don’t
      * Talk about how this research can be applied (very productive or very valuable species are most at risk?, can use this to better monitor them or assess their extinction risk?)
    - Do talk a bit about how the EN are on the left and the CR is on the right, VU in the middle. Not the main takeaway, but an interesting conclusion.
    - Assess how many of the IUCN are DD, it might skew the data in the middle plots one way or another?
    - Weighting to say that no status=0 suggests that no status means they’re not threatened, quite an assumption? Not really a better way to emphasize the risk though. It works I guess
    - State the main question big and large somewhere
  + Make histogram edits
    - Change bin ranges so that 11 is a break (6-11, 11-16, 16-21, 21-26, 26-31, 31-36, 36-41, 41-46) for ALL histograms
    - Save the new histograms in the poster folder (png) and plots folder (pdf)
    - Find the new frequencies for the bins (update/redo the histinfo stuff)
    - Use these frequencies to make new ratio plots (see the ones that Ed said to put on the poster)
    - Save the new barplots for the ratios same as above for histograms
    - Send Ed the new plots/let him know about progress
  + Update the poster with new plots/data/conclusions
* Ed poster edits:
  + Make graphs pdf, google it or just try copy/paste
  + ~~Also, if easy to change, the y-axes should probably read:~~
    - ~~Fraction of commercial species threaten~~
    - ~~Fraction of IUCN-listed commercial species threaten~~
  + ~~For the author list, I think we can include Juan and Malin? Also take out me, Juan and Malin from the acknowledgement list, and make the Reference heading the same size as the other sections.~~
  + ~~Also, I think commercial marine species is more accurate, as commercial marine fisheries may imply it’s the fisheries going extinct. “Commercial marine” should be sufficient to highlight that the species are harvested by fisheries.~~
  + Fix center box alignments
  + ~~Make center box white with colored outline~~
  + ~~Spell out MSY on first go~~
  + ~~Change conclusions:~~
    - ~~neither H1 nor H2 is obviously supported, with extinction risk being comparable at low and high value/productivity levels.~~
    - ~~future work includes a more detailed data analysis and quantitative theoretical predictions~~

7/31

* Today:
  + ~~Find way to increase axis font size but not get cut off~~
  + Stacked barplots for weighted and ratio
    - See Jennifer’s code (start 291) and other help sites/helpfile
    - Text for the stacked barplot table:

stackedIUCN<-read.table(text= “6 11 16 21 26 31 36 41

CR 0 0 0 0 0 1 0 0

EN 0 3 0 0 0 0 0 0

VU 2 9 3 2 0 1 0 0

NT 0 6 4 4 4 0 0 0

LC 1 18 15 3 2 3 4

DD 0 5 1 0 0 1 0 0” header=T)

barplot(as.matrix(stackedIUCN))

* + Main question things, add them
  + Talk with Ed about the questions I/everyone else had
  + ~~Finish up with the new break analysis~~
    - ~~Weighted risk:~~



* + - EX=5/5
    - EW=4/5
    - CR=3/5
    - EN=2/5
    - VU=1/5
    - NT=0
    - LC=0

8/1

* Today:
  + Try to finish edits to poster
    - Change to benefit/(cost+subsidy)
    - Why? Just do it
  + ~~Stacked barplots if not done yesterday~~
  + ~~Look into how to add NT to the “threatened” related plots~~
    - ~~Not consistent with IUCN ranking/classification, but will give more data points~~
    - ~~Not doing this, no worries~~
  + ~~Talk with Ed if didn’t do so yesterday~~
    - We leave the hypotheses as prod/value and value/prod. Should we add the 3rd?
    - Should we go back to the 5/10/15 bins? I could explain that we ran models and other tests and found that 11 was a critical point, I need a refresher on the best way to explain it, but I feel like that generally would be an ok reasoning for that increment.
    - So kind of what we’re seeing is that species are not more likely to go extinct with H2 is not the case because we’re seeing low extinction risk at the first bin, representing lower value/higher productivity? But are we also seeing a greater/proportional extinction risk at the upper end where the high value/low productivity species are?
    - The 2 ratio plots on top are good, I kept the FULL RAM weighted plot, it’s the same trend as the IUCN but represents the whole data set.
    - I also added the stacked barplot, but I feel like that could just be explained? It might be good since I sort of have the space?
    - “The original (cost+sub)/benefit x ln(msy) is correct for the analyses. I just meant that your definition of value on the poster intro was incorrect. economic value=benefit/(cost+sub). What you are plotting is 1/value or ln(msy)/value. Value is technically the benefit-cost ratio.”
      * Why would I not use the benefit/cost+sub for calculations then if that’s the better representation? Is it essentially that we’re manipulating the value term so that it plays nice within the whole finalvalue equation with ln(MSY)?
  + ~~Ethics/reflection workshop~~
  + Start compiling a section on my methods/materials
    - For final turn in and add to internship report
  + Work on Git README file, update it for what I want
    - Add data files
    - Add Ed’s paper as a reference
* Ed meeting notes/to-dos:
  + Change bins back to the 5,10, etc
    - Change everything else back as well
    - Change axis labels to be at an angle and write 0-5, 5-10, 10-15, etc so it’s clearer
  + Make a stacked barplot for the FULL RAM
    - Add status column
    - Assign the 92 species’ statuses and leave all others “none”
    - May need to alter y axis, but we’ll see
    - Set increasing risk as moving up column (CR at top) to match with the legend
    - Make colors more easily distinguished, rainbow and grey could convey this nicely, despite breaking with color palette
  + For each bar (in each bar graph) place text on top of bar with #threatened/#total in that bin
  + Get rid of the lower right histogram, the trend will mirror in the weighted plot
  + Add caption for stacked barplot
  + Simplify the hypotheses:
    - Don’t refer to as “hypotheses,” call “guiding questions” or just “questions”
    - Rephrase them to fit as questions
    - Have one big fish and a few smaller fish of the same kind (big on left, small on right)
    - Add a line graph showing the trend (see notecards)
    - Cite first one
  + Explain why the 3 middle plots are important and how they’re different
    - All have ambiguous trends
    - Have all 3 as an attempt to reduce/identify bias and look at different angles
    - Have them with small cations of “risk measure 1” etc and a **short title** (bolded)(don’t repeat axes, but say a little about them)
    - Have a text box talking more in-depth about their relationships
  + Conclusions:
    - 1st point: we compiled a new extinction risk database coupling economic and ecological data.
    - 2nd point: species with very different economic and ecological characteristics can have similar extinction risks
    - Low extinction risk in the first bin
      * H2 is worth looking into more
    - We’ll conduct/this data will allow for further quantitative analyses and test bioeconomic extinction models
  + We’ll do more analyses in Toronto and look into doing a standalone paper!
    - Logistic and ordinal regressions
  + Next week skype with Ariel and plan out logistics for how I’ll get there from the airport
    - Saturday will be chill, I can go and explore the area/be a tourist that day
    - Sunday dinner with Ed’s family and Ariel, maybe museum too?
    - Monday: 2pm meeting with Shelby, Ed’s in his office 10-5:30 so we can work on stuff then
    - Tuesday: 2pm meeting with Chealsea and drinks after work, office with Ed
    - Wednesday: 10PM to most of the day maybe conference practice talks with Marie Josee’s lab (free conference, yass)
    - M and W are pretty free if I have anything I want to do/meet other people
    - Customs via Canada takes a long time, I should arrive 2+ hours ahead of time when leaving (and go to the right airport, there’s 2 of them!)
    - The Royal Ontario Museum is really close to where I’ll be!

8/2

* Today:
  + ~~I’m getting really tired of all the names I have for my files and objects in R, sometimes the names don’t fit with the content anymore (“datafile” for the RAM\_IUCN\_DATA.CSV is so irrelevant, renaming so it’s easier to understand later and by someone who hasn’t been working with this like me)~~
    - ~~Rename “datafile” object to “RAMIUCN”~~
    - ~~Rename file “RAM\_INTEGRATED\_DATA.CSV” to “RAM\_FULL\_DATA.CSV” in FULL\_RAM\_MSY.R~~
    - ~~Rename “”integrated” object to “RAMFULL” in FULL\_RAM\_MSY.R~~
      * ~~Left all the old intermediates, renamed them and saved them on their own~~
  + ~~Simplify the hypotheses:~~
    - ~~Don’t refer to as “hypotheses,” call “guiding questions” or just “questions”~~
    - ~~Rephrase them to fit as questions~~
    - ~~Have one big fish and a few smaller fish of the same kind (big on left, small on right)~~
    - ~~Add a line graph showing the trend (see notecards)~~
    - ~~Cite first one~~
  + ~~Conclusions:~~
    - ~~1~~~~st~~ ~~point: we compiled a new extinction risk database coupling economic and ecological data.~~
    - ~~2~~~~nd~~ ~~point: species with very different economic and ecological characteristics can have similar extinction risks~~
    - ~~Low extinction risk in the first bin~~
      * ~~H2 is worth looking into more~~
    - ~~We’ll conduct/this data will allow for further quantitative analyses and test bioeconomic extinction models~~
  + ~~Change bins back to the 5,10, etc~~
    - ~~Change everything else back as well~~
  + ~~Get rid of the lower right histogram, the trend will mirror in the weighted plot~~

8/3-8/4 weekend

* Today:
  + Make a stacked barplot for the FULL RAM
    - Add status column
    - Assign the 92 species’ statuses and leave all others “none”
    - May need to alter y axis, but we’ll see
    - Set increasing risk as moving up column (CR at top) to match with the legend
    - Make colors more easily distinguished, rainbow and grey could convey this nicely, despite breaking with color palette
  + For each bar (in each bar graph) place text on top of bar with #threatened/#total in that bin
  + Change axis labels to be at an angle and write 0-5, 5-10, 10-15, etc so it’s clearer

#from Jennifer’s code, after line 291

text(seq(0.7, 5.5, by=1.2), -0.11, srt = 45, labels = c("NJ", "DE", "VA", "NC", "SC"), xpd = TRUE) #(saying that the x axis text will be in these locations at this angle, and say these things. I could try to just to the srt part and if that doesn’t work, add this line)

text(seq(0.7, 5.5, by=1.2), 1.03, labels = c("n = 52", "n = 50", "n = 44", "n = 146", "n = 1"), cex = 0.8) #(this is for the text at the top of each bar and what they say. Do this for the ratios)

* + Add caption for stacked barplot
  + Explain why the 3 middle plots are important and how they’re different
    - All have ambiguous trends
    - Have all 3 as an attempt to reduce/identify bias and look at different angles
    - Have them with small cations of “risk measure 1” etc and a **short title** (bolded)(don’t repeat axes, but say a little about them)
    - Have a text box talking more in-depth about their relationships
  + Finalize poster content and try to have 100% done if possible (due Tuesday)
  + Materials and methods
  + Work on Git README file, update it for what I want
    - Commit my R files
    - Find link to/upload Ed’s paper as a primary reference for this project
    - Figure out how to put other files (excel files) on Git (maybe wait until project is done and all files are essentially complete)
    - Figure out more formatting stuff to make it pretty
  + Show Michael my poster
* On 8/4 I found out that I actually had missed a few stocks’ classifications! Somewhere in the process of doing all of this by had I missed some stocks. So I spent most of my time working fixing this error instead of making progress with the stacked barplot for the full list (I initially didn’t have the IUCN status listed in a column so I had to add it and subsequently found out that there’s a whole set of stocks of the same species where only a few are given the status. I think that may have happened by some error when I was tired or I may have done something else? I’m really at a loss for words here. )

8/5

* Today:
  + ~~Fix IUCN status mistake~~
    - ~~Reload the IUCN and Threatened file/subset~~
    - ~~Rerun those tests~~
    - ~~Fix all ratio and weighting plots/data~~
  + ~~Make a stacked barplot for the FULL RAM~~
    - ~~Add status column~~
    - ~~Assign the 92 species’ statuses and leave all others “none”~~
    - May need to alter y axis, but we’ll see?
    - Set increasing risk as moving up column (CR at top) to match with the legend
    - Make colors more easily distinguished, rainbow and grey could convey this nicely, despite breaking with color palette
  + For each bar (in each bar graph) place text on top of bar with #threatened/#total in that bin
  + ~~Change axis labels to be at an angle and write 0-5, 5-10, 10-15, etc so it’s clearer~~

#from Jennifer’s code, after line 291

text(seq(0.7, 5.5, by=1.2), -0.11, srt = 45, labels = c("NJ", "DE", "VA", "NC", "SC"), xpd = TRUE) #(saying that the x axis text will be in these locations at this angle, and say these things. I could try to just to the srt part and if that doesn’t work, add this line)

text(seq(0.7, 5.5, by=1.2), 1.03, labels = c("n = 52", "n = 50", "n = 44", "n = 146", "n = 1"), cex = 0.8) #(this is for the text at the top of each bar and what they say. Do this for the ratios)

* + ~~Add caption for stacked barplot~~
  + ~~Explain why the 3 middle plots are important and how they’re different~~
    - ~~All have ambiguous trends~~
    - ~~Have all 3 as an attempt to reduce/identify bias and look at different angles~~
    - ~~Have them with small cations of “risk measure 1” etc and a~~ **~~short title~~** ~~(bolded)(don’t repeat axes, but say a little about them)~~
    - ~~Have a text box talking more in-depth about their relationships~~
  + Work on Git README file, update it for what I want
    - Commit my R files
    - Find link to/upload Ed’s paper as a primary reference for this project
    - Figure out how to put other files (excel files) on Git (maybe wait until project is done and all files are essentially complete)
    - Figure out more formatting stuff to make it pretty
  + Work on internship report when tired/out of focus

8/6

* Today:
  + ~~Fix the stacked barplot ordering to get what we want~~
  + ~~Review the text I have on the poster~~
  + ~~Meet with Ed at 11~~
  + Make all plots have the same dimensions (1500x1500), but stacked plot can have its own
  + ~~Make text in plots x3~~
    - ~~Widen the margins (y and x)~~
    - ~~May need to use a text line for the x lab~~
  + ~~Edit captions (see below, also any other blue highlighted changes)~~
  + ~~Have a last quick review by other people to see if I need to change anything on the poster~~
  + ~~FINISH POSTER~~
  + ~~SUBMIT POSTER FOR PRINTING by 5pm~~
  + Start materials and methods!!!
  + Work on Git README file, update it for what I want
  + Work on internship report when tired/out of focus
* Ed Meeting:
  + General layout is good
  + 3rd graph looks smaller, make it larger to fit size of others
    - ~~Risk Measure3: Give weighting scheme: keep the last line, but get rid of the rest~~
      * ~~Weight, last line with the numbers~~
  + ~~Remove the takeaways section~~
    - Reader should be able to see and understand a lot
    - Add the stacked
  + Make plot text larger and higher resolution
    - **Save as tiff**
    - Make 3 middle barplots shorter
    - Make stacked barplot narrower and taller, same dimensions as the others
    - TRY WITH THE legend!!!!
    - See if pdf images look better once the powerpoint is saved as a PDF
  + Stacked barplot caption
    - ~~Add a caption with bold font “title”~~
    - Don’t really need an explanation
    - ~~“data distribution”~~
  + ~~Work on the second point in conclusions~~
  + Reduce text from measures 1 and 2:
    - ~~Removed “as taken from histogram counts~~
    - ~~Remove “regardless of IUCN listing”~~
    - Add numbers #threatened in bar/#in bar total
      * For the first 2 graphs
      * Don’t do it for the 3rd plot
    - ~~Take out the n=30 and leave the~~ n=628
    - ~~“Take out the sections without a bar” sentence~~
  + Increase font size for middle column text, as possible
  + Make the plots horizontal instead, see how it looks
  + ~~References: add the superscript at the first, put at the end of the sentence~~
    - ~~Lam et al3~~
    - ~~Fish image4~~
    - ~~Etc~~
  + Illustration:
    - ~~Vertical line not fully straight, make a box instead~~
    - Can’t use images without permission
    - Use an icon instead of fish image, plain black fish sillouette
      * Use built in clipart in keynote (apple powerpoint)
      * ~~Creative commons is ok, don’t need to cite~~
      * ~~Find an ambiguous fish, can find 2 and use one for each (hi or low)~~
      * ~~Change x axis to reflect hypotheses (replace high with productivity/value)~~
      * ~~Y axis= extinction risk~~
      * ~~Label x as productivity/value~~
      * Our model suggests blue question?
  + Hypotheses as one sentence?
    - ~~Does extinction risk increase or decrease with productivity/value~~
    - ~~Reduce it to one sentence using productivity/value for all~~
    - Instead of: “Is (value/productivity) proportional to extinction risk? (High value, low productivity)

Is (productivity/value) proportional to extinction risk? (Low value, high productivity)

* + Background:
    - Reword the “human activity”, focus on “human harvesting”
    - Reword “species are going extinct”
      * ~~“species are at risk of extinction due to human harvest”~~
    - This study aims to…
      * ~~Don’t cap the Max Sust Yield, add “, MSY)”~~
      * ~~Benefit, not price (replace this, as benefit is how much in dollars, price is per pound)~~
      * ~~“ this factor is related to extinction risk” (we’re not predicting, we’re just collecting the data rn), combine with 4~~~~th~~ ~~point (this combined factor, which has been suggested to be a predictor [use predictor, not indicator] of harvest behavior)~~
    - ~~Last point: “here we matched data from RAM and IUCN~~

8/7

* Today:
  + TURN IN MATERIALS AND METHODS
  + Practice presentation

8/8

* Today:
  + Practice presentation
  + Lunch thing?