Coding game plan

* Cleaning data
  + Take one more crack at problematic documents
  + ~~Make new spread sheet with updated information (make note of changes being made for each species~~ 
    - [Still a couple of problems but have made new data excel doc with an extra notes section (currently column C) to note changes being made]
  + Check that years information has been updated
    - Pretty much done – see comments in notes sections for ones that were ambiguous and didn’t/haven’t gotten deleted so far
  + Question of how to sort folder – need to organize folder more – how do so without messing up r working directory?? --> set up github
    - Need to call IT desk to find out where to get Github help
    - Made new github account because Cornell interface doesn’t allow outside collaborators
  + Run through (on R) data cleaning including
    - ~~Deleting notes columns~~ 
      * ~~Manually removed notes columns~~
    - Correcting spelling errors
    - Fill in 1s and 0s for information columns
      * The internet isn’t very helpful
      * Need to write loop
      * Want to go through a column and change blanks to 0s and text to 1s while skipping over boxes that already have 0s and 1s
      * Tried two loops and weird things started to happen
    - Want to make sure data is being read in as the correct type --> text, numeric etc
      * Have checked structure – all information will be read in as factor unless stated otherwise. Unsure what this means
    - Logical “check” on numbers in years sheet vs salafsky
      * Where there are differences, need to write a code to fix this
    - Some cases in Salafsky where have multiple partners as “all partners”, need to fix this and if weird acronym or group, then identify (if possible and if not make a note) who is in the group
      * Are we interested in looking at actions that are done by all partners? Point of analysis to see if multi-partner actions are similar across conservation plan?
* Pre analysis set up
  + Partners per doc per species
    - Separate partners per document in salasfsky
    - Remaining total number of partners should belong to other doc (if more than one applicable to a species, will have to figure out how to code it in)
  + Also have to separate the partners that are collaboratively working together on actions
    - Unsure how to approach for network analysis
    - For now should split the given action in 2 and then place action back into partner specific line?
      * If BLM and USFWS are doing action 6 together, then want to add action 6 to both partner specific rows, making sure not to add if it partner is already doing action themselves?
    - Or would changing to a .5 work better for analysis?
* Analysis