Updated 5/30/2017 J. Connolly, 5/4/2017 J. Connolly, 4/20/2017 J. Connolly

############### User input variables
Input Sources: User driven, Bin database, output tables HUC 2 assignment tool

############### General Notes:
When collapsing bin assignments, all current assignment are considered and used to generated a single species assignment.
You will lose huc specific assignments. Typically this is only done for new huc 2 where we do not have an assignment.
Relies on the assumption that higher coded values should trump lower coded values.
    collapse_huc = True
    collapse_new_HUCS_only = True

If you opt to collapse across all HUCs for a species you will violate the assumption that higher values trump lower
values. As a result you will **need to updated the DB_code_Dict in Step_2_ReCode Bin Table** so that the values starting
with 13 are 'Yes' and not 'No'. These are huc specific assignment for land locked hucs, species can only be in these
bins in coastal HUC2s.

############### ASSUMPTIONS
1) Final columns headers are the same as those in current_bin_table
2) Columns are in the correct order ie, species info cols, bins, database info columns
3) HUC 15 should be considered land-locked because it is land-locked within US jurisdiction
4) Coded bins values are hierarchical, a higher number trumps a lower number when collapsing
    **NOTE if collapsing across all HUCs for a species the values found in the bin_code_update will need to changed to
    yes in Step_2_ReCode Bin Table

Step 1: Load data from current bin tables, tables used to update hucs and species info. Sets the columns from the
species tables that should be included in the output bin tables.

Loads input tables, updated species entity ids as needed, standardized the columns headers to current_bin_table and
master species table headers,add Spe_HUC column to all tables to be used as unique ID of species/HUC2 combinations.

Bin columns of the input table are are consider the final cols names to be use, and assumed to be in the correct order.
Bin df is split using index position of bin columns. Used index position so we could change the columns without having
to change the code. Species info columns are pulled based on user input.

Step 2: Make species data frames from woe group crosswalk from post processor, converts the woe group crosswalk from

```
38 long format to wide by adding a grouping category to the new [WoE_group] column for each entity id. Appends
   woe group
39 in wide format to species info df. Saves a wide version of the woe groups to archive folder.
40 WideWoeGroups_[date].csv - WoE group crosswalk in ide format
41
42 Finalized updated species data, appends WoE groups to end insuring all species info is updated
43
44 Step 3: Adds new HUCs and removes dropped HUCs based on species range update. Make archived of
   intermediate tables
45 Species_w_NewHUC2_[date].csv - all blank bin assignments; new species and species with new HUCs
46 RemovedHUC2_[date].csv - bins assignment for huc 2 dropped from species range
47
48 For species with a new HUC2 add updated flagged to note when there are HUC2 specific assignments, typically
   species
49 found in marine bins in both coastal and land locked HUCS.  Removed HUCs dropped from species range and
   generates an
50 archive of this data.
51
52 Step 4: Updated bins assignments in land locked hucs based on bin_code_update dictionary
53 Checks the marine bin assignments for species found in the land locked hucs, species can not be found in these
   bins
54    **NOTE when species occur in marine bins in coastal hucs and other bins in land-locked hucs the code value
   in the
55    marine bins for the land-locked huc starts with '13'
56
57 Step 5: Collapse bins assignment for a species across all hucs into a single assignment for a species, applies
   bin
58 assignment to specific hucs based on the user input variables collapse_huc, collapse_new_HUCS_only
59
60 If user wishes generate an inclusive species bin assignment that represents all HUCs this can be done just for
   the new
61 species/HUC2 combinations or for all species/HUC2 combinations.
62    ** NOTE Loads individual bin assignments for a species/HUC2 combo in list format using index position
63    ie list[list value index][int(start_index):int(end_index)] the end index must be one beyond the last value you
   want
64    to include.
65    **NOTE if doing this for all combination we will be violating the assumption that higher coded bin values
   trump
66    lower vale and the **DB_code_Dict in Step_3_ReCode Bin Table** will need to be updated
67
68 Step 6: Runs a final check on entity ids, removes old species info from bin data frame then merges the new
   species info
69 to the bin data frame.
70
71 erifies all entity id have been updated based on entityid_updated dictionary; Slices
72 working data frame to only include bin information and species identifier (entityid) removing old supporting
   species
73 information. Merges bin data frame to the updated species info data frame on the species identifier column.
   Reindex
```

```
74 col order to match user inputV
75
76 Step 7: Updated lead agency code to agency abbreviations; add in flags for new species, species without a
   range, and
77 species removed from master list
78
79 Add in flags to allow output table to be filtered to view species w/o rage, new hucs, and removed species
80
81 Step 8: Exports data frame to csv
82 UpdatedBins_[date].csv'  # working bin tables with updates
83 DroppedSpecies_[date].csv  # Species removed from master list
84
85 Reference URLs
86 https://chrisalbon.com/python/pandas_join_merge_dataframe.html
```