Table 1: List of programs to process ghgrp files, from initial download to readying file for analysis.

| Step | Function | Function File | Remarks (including run file) |
| --- | --- | --- | --- |
| Download file from api and put into panda | grab\_ghg\_url\_2\_pandas | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\ghgrp\_vars\_topandas\_v1.py | This will result in a number of small pandas—usually can download only 1500 observations at a time. Best to set this up to run at night.  Will need to modify for annual updates—write only the latest year |
| Join several pandas into 1 file | join\_pandas | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\ghgrp\_vars\_topandas\_v1.py | This is the step after downloading pandas |
| Join pub\_dim\_facility files | intake\_ghgrp\_PUB\_ff | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | Run in one program with data\_3NAICS. See C:\Users\KPE\.spyder-py3\api\_scraping\_project\pub\_dim\_panda\_unify4.py |
| Join pub\_dim\_facility files | data\_3NAICS | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | Run in one program with intake\_ghgrp\_PUB\_ff. See C:\Users\KPE\.spyder-py3\api\_scraping\_project\pub\_dim\_panda\_unify4.py |
| Join pub\_facts files (two kinds –sector or subpart} | misc\_ghg\_df (for subpart)  misc\_ghg\_sector (for sector) | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | At this point, the pub\_factsX and pub\_facility\_X datasets are separate |
| Join pub\_facility and pub\_facts dataset with detailed emissions on each gas | merge\_ghgrp\_SUBP\_files (also merge\_ghgrp\_PUB\_files) | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | Can use program either for merging either the subpart or sector files; this is a terminal file for all subparts and gas types if you want detail on gas types but not subpart types. Filename is PUB\_FACTS\_SUBP\_GHG\_EMISSION\_all\_desc.gzip |
| Consolidating gas emissions for each facility to prepare to merge with a subpart file or sector file | groupies | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | Sums up the gasses so there’s one observation for facility\_id, reporting\_year, and subpart. This is a transitional file and doesn't have much use on its own. Use this file to merge when starting to merge with a subpart. Filename: pub\_dim\_subp\_2mrg\_pubdimfac.gzip |
| Create file with total GHG emissions from all gasses and one subpart. | merge\_ghgrp\_SUBP\_files | L:\mid\kpe\data\api\_scraping\ghgrp\_write\_files\data\_clean\_fns.py | Result is terminal file. Level of observation of facility\_id, reporting\_year, and subpart. Each facility may have more than one subpart. Emissions will be for that particular subpart only. The subpartW terminal file is "pub\_dim\_facility\_subpW.gzip" |