Article\_Code File Map

* Data\_article
  + Filters
    - Results of gravimetric filter measurements
  + Homes
    - Home assessment data compiled
    - Spreadsheet of UPAS IDs, filter IDs, date deployed, and other metadata for UPASs deployed
* Omni
  + Omni\_raw\_data\_import
    - Imports raw (uncalibrated and uncleaned) data.
    - Pulls from omni\_all\_locations\_raw.rds, which is raw data from all omnis. No significant cleaning done from Nick’s code beforehand
    - Omits data collected after May 31st 2021
    - Cleans duplicate data from overlapping sensor deployments and daylight savings time
    - Copies outdoor sensor data for home 13/14 and 15/16 duplexes
    - Creates cleaned but uncalibrated data in minutely, hourly, and daily resolution and deposits files in ./csv\_created\_article/pre-calibration folder
  + Omni\_calibration
    - Calibrates PM2.5 data using gravimetric sample results in output\_article>pm25
    - Creates correction factor boxplot and model plot
    - Adds energy cluster column to all data using ../sense/csv\_created\_sense/daily\_energy\_data\_clustered.csv
    - Creates calibrated rds files in ./csv\_created\_article folder
  + Omni\_autocorrelation
    - Optional function to count amount of missing days for each home in a specified room sensor
    - Takes hourly omni data and makes daily data, filling missing days with NA
    - Adds in a home type column from ../sense/csv\_created\_sense\_article/home\_type\_df.csv
    - Adds in an energy cluster column from ../sense/csv\_created\_sense\_article/energy\_cluster\_df.csv
    - Creates csv of clustered data with gap NA days in ./csv\_created\_article/acf/omni\_daily\_data\_complete.csv
    - *Deleted section that tested for impacts of stationarity*
    - Tests autocorrelation for each sensor-season condition and makes csv of acf values for correlograms at ./csv\_created\_article/acf/acf\_data.csv
      * Includes n of lag pairs for each lag
      * Includes fraction of missing data for sensor sample
      * **Run time is very long**
    - Creates correlogram for each sensor-season condition
    - Creates example correlogram used in thesis
    - Creates csv of lags before insignificant acf is reached for each sensor-season condition identify data that will be omitted due to omission criteria, but do not omit before creating csv at ./csv\_created\_article/acf/acf\_lags.csv
    - Omission criteria
      * If less than 25 days within sampling period
      * If more than 11.11% of missing data
      * If acf not insignificant by the 30th day
    - Boxplots of lags before insignificant ACF, colored by room, and season
    - Create csv of lags before insignificance by season and total for each IEQ indicator at csv\_created\_article/acf/lag\_summary.csv
  + Omni\_averaging
    - Averages data in ./csv\_created\_article/omni\_**daily\_**calibrated.rds for each ieq data with averaging periods determined in ./csv\_created\_article/acf/lag\_summary.csv
    - Plots boxplots of time-averaged values individually by home Groups specified in ../sense/csv\_created\_sense\_article/home\_type\_df.csv
    - Uses **hourly** data and Makes csvs of summary stats (by-home data used in thesis), csvs in ./csv\_created\_article/averages/
  + Omni\_cov
    - Uses **hourly** data to make cov csvs (by-home csv used in thesis); csvs created in ./csv\_created\_article/cov/
  + Omni\_ts\_summary
    - Uses hourly data to make time series plots
    - Hour-of-day plots
      * Pooled (all homes)
      * By home (all homes)
      * By season (omits homes without sense data or with uncertain season classification)
    - Monthly plots
    - Weekday plots
  + Omni\_time\_representativeness
    - Creates csv of representativeness of samples
    - Only considers homes in group 2 and group 3
    - No overlap between sampling periods
    - Calculates how representative a sample from a given season is of the **entire monitoring period**
    - **Takes a long time…**
    - Csvs in ./csv\_created\_article/representativeness\_data/
    - Makes csvs for IEQ indicators individually, and have to manually bind them together
    - Manually create ./csv\_created\_article/representativeness\_data/rep\_data\_time.rds
  + Omni\_indoor\_correlation
    - Uses hourly omni data
    - Day data include 6:00am to 8:59pm
    - Spearman correlation and n only includes data where both living and bed (or living and kitchen) were available
    - Bedroom and kitchen concentrations vs livingroom concentrations for all homes pooled
      * If I run correlations, do I ignore homes? **Yes**
      * Do I run pearson or spearman correlations? **spearman**
      * Spearman, p-value is not exact, as there are ties. Can likely just say it is an approximation, [but also can look at if distribution agrees with student t-test?](https://rpubs.com/aaronsc32/spearman-rank-correlation)
      * Do I omit outliers to allow for clearer plot? **No**
      * Removed day/night stratification, 4/22/22
  + Omni\_temporal\_rep\_plots
    - Import data from ./csv\_created\_article/representativeness\_data/rep\_data\_time.rds
    - Clean rep data
      * Can omit a sample length for a givens based on fraction of samples available, **but currently not omitting based on this criteria**
      * omit extreme outliers for K:D
        + group\_by(method, metric, sample\_length)
        + define extreme outliers within each group as median KLD +3\*IQR **need to redefine with Q#, not median**
      * counts the percentage of samples omitted by group
    - group cleaned data by IEQ indicator and scale the KLD values to calculate representativeness
      * kld\_max is the least representative sample of the given IEQ indicator
      * Rep = 1-kld/kld\_max
    - Example plot of sample hour of day averages compared to long-term hour of day averages with Rep values included
      * 1,3,7 day sample lengths
      * 1,3,7,14 day sample lengths
    - Density plots of representativeness NOTE, ONLY GROUP 3 HOMES INCLUDED
      * Thesis version
        + Thresholds at 0.8 and 0.9
        + 1,3,7,14 day sample lengths
      * Article version
        + Threshold at 0.8
        + 1,3,7 day sample lengths
  + Omni\_thesis\_compare
    - Compare thesis precalibrated data (used for Gravimetric Correction factor) with article precalibrated data
    - **Thesis data had used dates after may 31**
    - **Thesis and article datasets are equal to one another when dates after may 31 are omitted from thesis data**
* Sense
  + Sense\_season\_clustering
    - Clusters days for each home into seasons using ../output\_article/sense/sense\_hourly.rds
    - Creates csv of clustered days in ./csv\_created\_sense\_article/daily\_energy\_data\_clustered.csv
    - Identifies “start and end dates” for seasons and makes csv of these dates for each home in ./csv\_created\_sense\_article/energy\_cluster\_df.csv
    - Creates season classification plot
    - Creates csv of percentages of days classified in each season in ./csv\_created\_sense\_article/season\_pcts.csv
    - Creates csv of home grouping in ./csv\_created\_sense\_article/home\_type\_df.csv
  + Sense\_thesis\_compare
    - check that the season categorization of each day was the same for thesis and article datasets (./csv\_created\_sense\_article/daily\_energy\_data\_clustered.csv')
    - **they are equal**
* Upas
  + Upas\_analysis
    - Uses gracvimetric and upas data and meta data from data\_article and output\_article
    - Checks for transcription errors of gravimetric data entry
    - Calculates PM2.5 concentrations from gravimetric data
    - Converts samples with less than 24 hours of collection to NA values
    - Calculates LOD from blanks
    - Creates gravimetric concentrations csv at ../output\_article/pm25/pm25.rds
* Home\_map
  + Home\_map
    - Changed slightly from version sent to Kelsey
    - Only need google key if want to
      * geocode in new homes
      * or make new background map of FoCo
    - Can edit csv to edit info
    - Imports
      * Black and white map of FoCo at home\_map.rda
      * Home lat/lons and occupant info at home\_coord\_article.csv
    - Makes jittered map and saves as image file at home\_map\_image.jpeg (**dated)**