**ReadMe for HHS Protect data**

**Updated October 21, 2020**

**Commuting data**

* Use “L:\PRIV\Weaver\COVID-19\Commute data\commute\_results\_v8.csv” created by Jeremy Baynes.
  + Code “Import commute data.sas” reads into SAS and creates SAS file “L:\PRIV\Weaver\COVID-19\ Commute data\commute.sas7bdat”

**Hospital capacity data**

* Used for metrics of ICU capacity
* Log into <https://protect.hhs.gov/workspace/carbon/>
  + Click on “Data Catalog”
  + Click on “Hospital Data”
  + Click on dataset “[Unified] Hospitals Priorities – Latest(hospitals\_unified\_prioritized\_latest)”
    - The dataset should appear. On the top bar, click Analyze. Save as intermediate “contour object” in My Projects/Hospitals.
    - Click on Filter, then “Remove columns”.
    - Remove columns other than hospital information (name, ccn, state, zip, fips code, date, source, is\_va is\_ihs) and ICU information (icu\_beds\_available icu\_beds\_used total\_icu\_beds).
    - Save in My Projects/Hospitals/Hospital\_data\_MM\_DD\_YY. Click on dataset to open it.
    - click Actions (drop down menu) and Download as CSV.
    - A pop-up box will inform you of a restricted action. In the box, enter something similar to “These data will be aggregated and used to inform re-opening decisions at EPA facilities.”
    - Download the data, append date to end, and save to L:\PRIV\Weaver\COVID-19\Hospitals
* This dataset can be read into SAS, summarized by commuting area, and exported into semicolon-delimited txt file using “L:\PRIV\Weaver\COVID-19\Hospitals \Hospital cleaning code.sas”
  + Will also need “L:\PRIV\Weaver\COVID-19\ Commute data\commute.sas7bdat”
* The SAS file “L:\PRIV\Weaver\COVID-19\Hospitals\Archived and Error check codes\Checks for Hospital code.sas” contains checks for common data issues in the hospital data.
  + NOTE: dates in code will need to be updated

**Symptom data**

* Used to determine daily counts of emergency department (ED) visits for influenza-like illness (ILI) and COVID-19-like illness (CLI)
* Log into <https://protect.hhs.gov/workspace/carbon/>
  + Go to [ed\_data\_county\_aggregates\_timeseries](https://gcc01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fprotect.hhs.gov%2Fworkspace%2Fdata-integration%2Fdataset%2Fpreview%2Fri.foundry.main.dataset.e0cfccc8-5c72-4c2b-8b6a-4600cbe1251c%2Fmaster&data=02%7C01%7Cweaver.anne%40epa.gov%7C20480cf02c72435cfbbd08d86483e13c%7C88b378b367484867acf976aacbeca6a7%7C0%7C0%7C637369864860353968&sdata=0zUqhAVlnUMTsr0CcPORJ3uliAIYUYvZCVF0ITinefA%3D&reserved=0)
    - The dataset should appear. On the top bar, click Analyze. Save as intermediate “contour object” in My Projects/Symptoms.
      * Click “Filter”
      * Filter to keep all rows where “date” is between (inclusive) eight days and one day prior (for example, on 7/15/2020, select dates between 7/7/2020 and 7/14/2020)
      * Under “Result” click “calculate results”
      * Under “Result” click “Save as dataset” and save dataset to “my files” in a folder “Symptoms” (create folder if it doesn’t exist) as ed \_MM\_DD\_YY (where MM is month, DD is day, and YY is year)
      * Click on dataset once it’s created. The dataset should appear.
      * On the top bar, click Actions (drop down menu) and Download as CSV.
      * A pop-up box will inform you of a restricted action. In the box, enter something similar to “These data will be aggregated and used to inform re-opening decisions at EPA facilities.”
        + Download the data and save to L:\PRIV\Weaver\COVID-19\Symptoms\Data from HHS Protect\ed\_MM\_DD\_YY.csv
* This dataset can be read into SAS, summarized by commuting area, and exported into semicolon-delimited txt file using “L:\PRIV\Weaver\COVID-19\Symptoms\Symptom data V3.sas”
  + R code may also be used “L:\PRIV\Weaver\COVID-19\Symptoms\readsympsv3.R”
  + Will also need “L:\PRIV\Weaver\COVID-19\ Commute data\commute.sas7bdat” (for SAS code) or “L:\PRIV\Weaver\COVID-19\Commute data\commute\_results\_v7.csv” (for R code)
  + Will also need prior week’s ED data “L:\PRIV\Weaver\COVID-19\Symptoms\Data for Dashboard\ILI\_CLI\_by\_facility\_MM\_DD\_YY.txt”. New data will be appended to prior three weeks of data.
  + NOTE: This code may need to be updated depending on formatting changes in the data
  + NOTE: dates in code will need to be updated

**Testing data**

* Used for metrics of test positivity
* Log into <https://protect.hhs.gov/workspace/carbon/>
  + Click on “Data Catalog”
  + Search on left bar for “all\_pcr\_tests”
  + Click on first dataset “all\_pcr\_tests” with sub-head “/COVID/Analysis: Lab Testing Results/data/transform/all\_pcr\_tests”
    - The dataset should appear. On the top bar, click “Analyze”
      * Click “Filter”
      * Filter to keep all rows where “order\_date” is between (inclusive) 28 days prior to the previous Friday (for example, on 7/15/2020, select dates between 6/12/2020 and 7/10/2020)
        + Note: if there are bandwidth issues, just download 14 days and use SAS code to append prior two weeks’ data to dataset
      * Save dataset to “my files” as tests\_MM\_DD\_YY (where MM is month and DD is day)
      * Open the dataset. On the top bar, click Actions (drop down menu) and Download as CSV.
      * A pop-up box will inform you of a restricted action. In the box, enter something similar to “These data will be aggregated and used to inform re-opening decisions at EPA facilities.”
      * Download the data and save L:\PRIV\Weaver\COVID-19\Testing\Data from HHS Protect\Tests\_MM\_DD\_YY.csv
* This dataset can be read into SAS, summarized by commuting area, and exported into semicolon-delimited txt file using “L:\PRIV\Weaver\COVID-19\Tests\Clean test data.sas”
  + Will also need “L:\PRIV\Weaver\COVID-19\ Commute data\commute.sas7bdat”
  + Will also need prior week’s testing data “L:\PRIV\Weaver\COVID-19\Testing\Data for dashboard\ PCR\_tests\_MM\_DD\_YY.txt”. New data will be appended to prior three weeks of data.
  + NOTE: dates will need to be updated in code.
* Autoregression models for testing data can be run using “L:\PRIV\Weaver\COVID-19\Testing\PercentPositive7-22-20withinstructions.sas”
  + Need test data from above step: “L:\PRIV\Weaver\COVID-19\Testing\Data for dashboard\ PCR\_tests\_MM\_DD\_YY.txt”.
  + NOTE: dates will need to be updated in code.
  + NOTE: check log for convergence errors in autoregression models.