# Creating / updating the mini artifact

# **WHY**

The inputs to PRL are much larger than some of our other models. Running with the full artifact locally will cause you machine to OOM.

# WHAT

We’ve created a mini artifact for local testing. The code to create mini artifact on this branch:  
test/mini-artifact

The mini artifact building code has the following differences:

* It reads in mini input data versions of:
  + The ACS person data
  + The ACS household data
  + The address data
* It subsets, within the loader code, using
  + the first name data to a random subset of the inputs
* It subsets (within the loader code) the last name data to a random subset of the inputs

The reason we’ve created mini inputs for the mini artifact (instead of doing all subsetting just within the loader code), is that the inputs are too large to debug locally.

# MINI ARTIFACT INPUTS

The mini inputs were created in a python notebook called “create\_mini\_artifact\_data\_inputs.ipynb” that beatrixh shared to the #vivarium-census-prl channel on 12/8/2022 along with the message “@James Albright notebook that created inputs for the mini artifact”

* This could be stuck in the github repo? But kinda sucks to have more notebooks in there
* I used the notebook format because these inputs needed to be created on the cluster (will OOM locally) and I wanted to be able to look at what was happening to the data I was manipulating. Ideally this code wouldn’t be in a notebook, perhaps.

The ACS person and household data need to contain the same households, so at the top of the notebook there is some code that selects a subset of ACS households that have either positive household weights, or are group quarters

(note: group quarters all have weight zero. I don’t know that the other non-GQ households are with weight zero; I asked RT at one point and I think they didn’t have the bandwidth to look into it)

I then double-checked that the selected households contain some GQ households.

Then both the mini household and person files are created by subsetting to the selected households.

# RECOMMENDED WORKFLOW

If you need to update the artifact with new data, I would recommend you update the artifact, then merge main + those updates into the test/mini-artifact branch. (As of 12/8/2022, doing this

If you don’t need to touch the mini inputs, great.

If you do, go look at the notebook that created them.

Update the loader code as needed and run.