**Scenario**: Todd has requested additional years for AIDS diagnoses data (state level) to be added to the HIV surveillance manager from the PDF surveillance reports.

*When data requests come in, clarify the years, locations, strata, source.*

**Step One:**

*Where does AIDS Diagnoses data come from and what code processes/puts this data into the surveillance manager?*

Check the documentation. Open jheem\_analyses -> documentation -> Data Manager -> outcomes -> Surveillance Manager Data Overview

This document shows us that the process code is “msa\_reports\_processing” and we currently have state level data back to 1992.

**Step two:**

Find additional reports (you can just use a Google search). CDC Stacks typically has older reports: <https://stacks.cdc.gov/view/cdc/149053>

Download and save those reports in the proper folder on the Q Drive.

**Step three:**

Put the data in a format that can be read in the code. You normally wouldn’t have to do this but it’s necessary for PDF reports.

Tables from PDF reports can be extracted using Tabula. These are even too old for that...because it’s a few data points I did this by hand and saved as CSVs.

**Step 4:**

Update the processing code to for this data. Typically this also isn’t really necessary but PDF processing is more finicky.

Typically I pull the most recent report of a data point. This is an exception to that- Todd wants as much data as possible.

**Step 5:**

Test cleaning and putting the data, check that it’s there as you expect it to be. Push any changes in the code to Github.

**Step 6:**

Run the section of the data manager where you made changes (you can check this in the documentation excel). Then merge with the other sections and save a new data manager.

Upload to cache folder one One Drive and sync metdata.

**Step 7:**

Let everyone know there is a new version (I send emails) and update any documentation to reflect changes.