To Do 9/30/20

* Data Management of arcos\_all\_washpost.tsv
* This is a dataset of all opioid shipments to every pharmacy in the United States from 2006-2014.
* It is 74.5 GB
* I would like the dataset in the following format
  + Panel dataset by county and monthyears
  + Time in month years. The month format should be 2006m1 for January 2006
  + Transaction\_date is the original transaction date, collapse it into monthyears
* I would like the following variables
  + State (Abbreviations: MA= Massachusetts)
  + County name all caps
  + County FIPS in the following format “AK090”
    - Alaska, Fair Banks Northstar
    - http://www.naarep.com/agentsite/training/templates/FIPS.pdf
  + I want the following variables in total Purdue Dosage\_unit, Purdue\_MME, and Purdue Volume as well as Industry\_Dosage\_unit, Industry\_MME, and Industry\_Volume.
    - Industry means that it includes all companies except purdue. Drop the Purdue values then collapse it. It would be the total in the industry excluding Purdue
    - If Combined Labeler Name==Purdue Pharma LP, then it is a purdue transaction. There is another variable in the original dataset that has Purdue, I forget which variable it is. If any column has Purdue in it, please count it as Purdue totals and exclude it from Industry Totals
    - Dosage\_unit (this is the number of pills per county per monthyear)
    - MME= Volume of Morphine mg Equivalent, this is Dosage\_unit\*MME\*dos\_str in the original dataset
    - Volume of pills= dosage\_unit\*dos\_str
* Combined Labeler Name==Purdue Pharma LP
* Middle data set that goes by time by month by county by combined labeler name

Created variable is purdue dosage units