Assignment:  
  
 Use the following data sources and steps to answer this question:

How many single-family homes with central AC and heat with natural gas in Minnesota grouped by wall type?

[2020 RECS Data Survey](https://www.eia.gov/consumption/residential/data/2020/index.php?view=microdata)

[Variable And Response Codebook](https://www.eia.gov/consumption/residential/data/2020/index.php?view=microdata#:~:text=CSV-,XLS,-EIA%2D457%20A)

1. Design, create, and populate a database from the above source data that enforces proper referential integrity. You can choose from the following database management systems: MS SQL Server, Postgres, MySQL (or an online dbms as long we can access and inspect it).
   1. You do not need to create type tables for each variable in the codebook. Just enough to answer the question above. It should be at least 4 tables.
2. Write a python script that populates the database from the step above and performs the following data validation checks and cleanup:
   1. Remove any records where heating degree days are less than 7000.
   2. Implement 3 other validation/qa checks that you feel would be important (the data is actually pretty clean, just have the checks in the script)

Store the database create scripts, python script, and query text in a public source control repository such as github and send a link to Tom ([tspielman@mncee.org](mailto:tspielman@mncee.org))

Due Date: End of Day on July 12th.  
  
Follow up question:   
  
An important aspect of the role is to be able to meet with non-technical team members and gather requirements for the database. Using the requirements, you would be asked to research different database options and propose a solution that meets the team’s needs. Please share a list of no more than 10 questions you would ask the team to develop a list of requirements you could use to evaluate different database options.