Cleaning US Census Data

You just got hired as a Data Analyst at the Census Bureau, which collects census data and finds interesting insights from it.

The person who previously had your job left you all the data they had for the most recent census. The data is spread across multiple csv files. They didn’t use R, and they would manually look through these csv files whenever they wanted to find something. Sometimes they would copy and paste certain numbers into Excel for analysis.

The thought of it makes you shiver. This is not scalable or repeatable.

Your boss wants you to dig into the data and find some insights by the end of the day. Can you get this data into R and into reasonable shape so that you can perform your analysis?

If you get stuck or want to see an experienced programmer tackle this project, click “get help” to see a walkthrough video.