- a. If I were a software developer at Airbnb, I would implement a program that directly verifies if business licenses are valid. The system that is currently in place only catches invalid licenses after complaints are made by the public. This is not only an unreliable method for maintaining legal compliance, but also very inconsistent. The program I would implement would validate policy numbers by scanning them on listings. While this is a significantly better method than relying on the complaints of third parties, the organization may get some arguments about it. The first argument is that if listers were aware of the system, they could create fake policy numbers that would pass the scanning. Another argument is that listers may take offense to the fact that Airbnb does not fully trust them, so they may not want to rent out their property through such a company.
- b. After skimming through the Wikipedia page, one question I would be interested in understanding as a data scientist fighting against housing insecurity is, "does the location of housing relate to the cost of housing?" The data shows a graph of home prices in the United States in 2021 by county. I could use this data to determine if the cost of housing relates to the location of housing. Studying this relationship would help me understand the reason for lower or higher housing insecurity in certain areas or communities.
- c. One factor that I think is very important to consider when discussing the legality of web scraping is the purpose. Is web scraping being used for good intentions, like to promote equality? Or, is it being abused? Web scraping can be a great tool to promote equality. For example, it can help determine if there is gender or racial bias present in algorithms. However, web scraping can also be abused if used to scrape personal data, intellectual property, or confidential data.
- d. When deciding to use public data, I think it's important to analyze how much harm or good using the data could have on people. It could threaten privacy, security, and even increase inequality and discrimination. For this reason, it's important to protect anonymity. In cases where data is required to complete a research goal, I think it's important to collect only the data that is required rather than all of the users personal data. In summary, guidelines that should be followed when web scraping are (1) not doing it if it could harm people, (2) always protecting anonymity, and (3) only collecting data that is needed.