

MA 386 - Statistical Programming

Portfolio: Data Input/Output

Background:

Terre Haute, Indiana is a small college town, home to both Rose-Hulman Institute of Technology as well as Indiana State University (Saint Mary's of the Woods is located in nearby West Terre Haute). The city is home to approximately 61,040 individuals with a median household income of \$33,317 (based on 5-year estimates from the American Community Survey). As with many college towns, the city draws many popular retail chains.

Prior to 2015, K-Mart and Walmart each had a presence in the city. Following the opening of a Meier in the summer of 2015, K-Mart closed its doors. Rumors immediately swirled that Target was considering moving into the vacated building, which most residents would welcome. However, there has been a long-held belief (and widely circulated rumor) that Target would not open a store in Terre Haute because the median income of the city is not high enough. If such a standard exists, and Terre Haute does indeed fall below this standard, we are not likely to see Target invest in a local branch without extreme incentives. We are interested in determining if such a rumor is plausible.

The American Community Survey is conducted annually (as opposed to the US census which occurs every few years). The US Census Bureau uses the survey to provide estimated characteristics for various regions. In addition to population estimates, median household income is available. Further, this data is available at the zip-code level across the United States. Using the FactFinder application on the US Census Bureau website, we collected median household income data from the American Community Survey. The data, as well as additional supporting documents, are provided on the course page ([aff download zip.zip](#)). Note: data is available on nearly every zip code in the United States; there are a few zip codes for which there were not enough residents (business districts) and these were automatically excluded from the available data.

The locations of all Target stores in the US is not as easy to obtain. However, Target does make the locations of stores available by state on their website:

http://gam.target.com/store-locator/state-listing?lnk=findstore_viewall_storesbystate

We are interested in combining these two data sources in order to assess whether the data supports the rumors that Target only places stores in locations above some

threshold income level. Further, does it appear Terre Haute may be below such a threshold?

Assignment:

Using the data available from both the American Community Survey as well as the Target website, comment on the rumor that Terre Haute does not meet the minimum median household income for Target to consider opening a location in the city. You should construct some type of graphic to support your conclusions.

Tips:

The following may be useful in completing the assignment.

- On the Target website provided, note that clicking on each state provides a unique URL for the resulting information.
- Examine the source of one of the state-specific URL's above; what can you say about how the data is stored/presented? (Hint: it is not standard HTML).
- In order to extract the data, consider using a css selector.