

I found the data file for this assignment on data.gov with the following description:


Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease and Cancer

 Metadata Updated: Sep 04, 2015

Community Health Status Indicators (CHSI) to combat obesity, heart disease, and cancer are major components of the Community Health Data Initiative. This dataset provides key health indicators for local communities and encourages dialogue about actions that can be taken to improve community health (e.g., obesity, heart disease, cancer). The CHSI report and dataset was designed not only for public health professionals but also for members of the community who are interested in the health of their community. The CHSI report contains over 200 measures for each of the 3,141 United States counties. Although CHSI presents indicators like deaths due to heart disease and cancer, it is imperative to understand that behavioral factors such as obesity, tobacco use, diet, physical activity, alcohol and drug use, sexual behavior and others substantially contribute to these deaths.

Access & Use Information

 **Public:** This dataset is intended for public access and use.

 **License:** No license information was provided. If this work was prepared by an officer or employee of the United States government as part of that person's official duties it is considered a [U.S. Government Work](#).

To set the stage for your work, imagine that you have been called by one of the 3,141 communities included in the report – Worcester, MA. The community leaders of Worcester want to know how they compare with other communities with regard to the health measures included in the report.

The community leaders don't have much money, so you need to be able to help them in a reasonable amount of time – by the end of next week sounds about right. So . . .

1. Explore the data. Organize it so that you can assess and describe comparability between communities.
2. Make sure you examine and unpack the metadata file:
d5190a7b-361e-4f50-9ab4-bf526cc67124 – it is a json file.
3. I have suggested to you that clustering might be a way to put the 3,141 communities into groups which share similar descriptions in this dataset. Use the K-means clustering method I demonstrated in class and posted on Blackboard. This means that you can restrict yourself to variables with “decimal” data values.

Submit a written report which is not more than five pages long. Evaluation of the quality of your work will not necessarily improve with the length of your report. Use R and Rstudio. Describe what you did and provide files so that I can reproduce your work with a minimum of effort. Put your work in a folder in your github and submit the URL on Blackboard.

You may get help from class members, but only class members. Please give credit to those who have helped you. This means that you must be able to explain the work you submit or give me access to the person who helped you.