DESCRIPTION OF DOCUMENTATION FOR:

ZUPAN (2017)

"WALKABILITY AND OBESITY TRENDS ACROSS GEOGRAPHICAL REGIONS IN THE UNITED STATES"

The data management and analysis for the paper were conducted using R version 3.3.1 "Bug in Your Hair" using RStudio version 1.0.136 It has been written to run on a Windows or Mac operating system. In addition to the base R package, several packages are used. These can be seen at the top of the Rmd document loaded in the first code chunk.

The paper was knitted to pdf using the **elsevier article** format contained in the **rticles** package. Knitting to pdf requires a LATEX installation on your computer, though you will not need to write any LATEX code directly.

The documentation is organized into three main directories (listed in order of importance for reproduction of results):

- **Source code:** Here you will find the Rmd file that reproduces the entire paper, along with all auxiliary files created in process of knitting the pdf.
- **Processing and analysis:** contains all data called by the Rmd file in analysis, as well as supplementary code that is called by the **source** command in the Rmd file. The **Supplementary code** folder contains script that is called by the Rmd code and do not need to be run directly by the user. This code does the data cleaning necessary to put the data into analysis form. It is excluded from the Rmd file only to keep the document cleaner. The **importable data** folder contains the original data files modified to the minimal extent necessary to import them into R. All further processing is done in the code.
- **Original Data and Metadata:** contains the data we collected in the most original form possible. HTML files of the webpages are available in each of the folders. These are the websites from which I gathered the data, and then I manually transferred all of the data into an Excel spreadsheet. Websites include: Governing.com (Obesity 2010), Gallup Analytics (Obesity 2013), American Fact Finder (Census/Demographic Information), Walkscore (Walkscore), and National Geographic Regions (division of US states into 5 regions).

To reproduce all results reported in paper:

Copy the **Documentation** folder to your computer. Make sure to leave the folder structurally intact. The code has relative paths to the data and supplementary code that assume the data is in this folder structure. The Documentation folder can be moved or renamed as necessary, but everything else must remain unchanged.

Open the Rmd file in the **Source Code** folder using R Studio. Make sure all required packages are installed.

Run all code in the file. If the output format is changed to an R notebooks format, the chunks can be run one at a time to see individual results.

Knit the file to **elsevier_article**. (Knitting to other pdf formats should change only the formatting. You must have a LATEX installation on your computer to do this.