Problem Set 3: Databases

Background

Instructions

1. Find 5 different sources of publically available biomedical or public health data. Briefly describe (in about one or two sentences what each source is). Hint: We may have used data from some of these sources in previous assignments.
2. Pick one of these sources and download a dataset from that source.
3. Load the dataset into R and calculate some simple summary statistics (e.g. mean, median, min, max, variance, if the variable is continuous) for 5 variables.

The report

Develop a report (I recommend a Word (or other text editor) document) for your problem set that includes answers to all of the questions posed above, showing plots where appropriate.

Save your report as a PDF file and submit your report through the course 2GW site. Clean up your code and submit it as a supplementary file, along with your main report.

Due date

Day 7, Week 6