Introduction to the R Statistical Computing Environment
Getting Started With R: Exercises

John Fox
(McMaster University)
ICPSR

2018

1. If you haven't already done so, install R and RStudio on your computer (instructions are on the lecture-series website) and verify that it works. Install the R packages used in the lecture series (again, instructions are on the lecture-series website), including the **car** package.

2. One of the data sets in the **car** package, called `States`, contains education and other data for the 50 U.S. states and Washington DC. Using the R Commander, find out what's in the data set by looking at its help page and at the data, and then perform a linear least-squares regression of the average SAT math score of graduating high-school students on the average teachers' salary in the states. Perform a second regression of SAT math score on both teachers' salary and percentage of students taking the SAT exam. Compare the coefficients for teachers' salary in the two regressions. How do you account for the difference? Make some graphs of the data.

3. * Write a function `MAD()` to compute the median absolute deviation from the median,

$$\text{MAD} = \text{median}\left[|x_i - \text{median}(x_i)|\right]$$

Confirm that your function works by comparing its results with those from the standard R `mad()` function, using, e.g., `MAD(1:100)` and `mad(100, constant=1)`. Then look at how `mad()` is programmed and compare it to your solution. In making the comparison to your function, be sure to set the argument `constant=1` in the call to `mad()` (see `?mad` for the explanation).