Show **all** of your work on this assignment and answer each question fully in the given context. This homework consists of two parts: the first part will be problems from the book the second part will be an exploratory data analysis in R. There is no magical connection between the two except that they are both worth the same amount (i.e., 50% of the grade on this HW will come from the first part and 50% will come from the second part).

1. **Problems from Rice Section 3.8**

   (a) Problem 1
   (b) Problem 7
   (c) Problem 8
   (d) Problem 11
   (e) Problem 57
   (f) Problem 58

2. **Exploratory Analysis of the Behavioral Risk Factor Surveillance System**

   According to the Centers for Disease Control and Prevention:

   > The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

   The results and related documentation of the 2016 survey are public and can be found here:

   `https://www.cdc.gov/brfss/annual_data/annual_2016.html`

   For this problem, perform an exploratory data analysis of the 2016 BRFSS survey. The survey itself is pretty interesting, since it deals with behaviors that people engage in that could have consequences in personal or public health. Your analysis should include an (brief) overview of the dataset, summary plots and tables, and possible simple statistical models. Your goal is to find one *interesting* thing in the data and report on it. You have two pages (including graphs) to do this in so be very careful which plots and tables you decide to include. Write the report up as if you are sharing information with your fellow students (i.e., you can assume they know how many rows of data there are).

   Of particular interest to you will be the dataset we are analyzing, which can be directly downloaded as a zip here:

   `https://www.cdc.gov/brfss/annual_data/2016/files/LLCP2016ASC.zip`

   and the code book which explains the variable names, which can be viewed here:

   `https://www.cdc.gov/brfss/annual_data/2016/pdf/codebook16_llcp.pdf`

   Notice that the main file is a "fixed-width" file. This means that the data in the file is meant to be read by knowing which column each field begins and ends on. I have written a simple R function that will do this for you. This is provided in the code book but I have written a

---

function that can read the data in for you. Notice that the dataset is very large - you may want to consider working with only part of it initially (so that each operation you try cuts down on some time).

```
source("https://imouzon.github.io/stat430/hw/hw4/read_brfss.R")
read_brfss("<path_to_your_brfss_data>")
```