Assignment #3: Big Data in R

Data: The data for the book *Data Science in R* is located at the book website.

http://rdatasciencecases.org/

We will use the flight data from 1987, 1988, and 1989 for this assignment.

Assignment Instructions:

This assignment will follow the case study in Chapter 5. However, students should note that the case study walks you through the steps of the analysis, however, your statistical report is NOT a report or diary of your actions or the case study.

For this case study we will read but not complete Sections 5.3.2 and 5.3.3 and focus Section 5.3.4 and Section 5.4. For the analysis in Section 5.3.4 and Section 5.4 we will use the flight data from 1987, 1988, and 1989 only. That is enough data to balance the analysis against computation time. Note that this assignment is about HOW you should approach larger data sets using R, not the actual analysis. This analysis is actually quite boring, but it forces us to use some different R tools that we should all know.

What we will turn in is a short report (maybe 2-3 pages) of variations of the graphic on p. 235.

Plot #1: run the analysis as in the book by hour of the day but with myProbs <- c(0.50,0.75,0.90).

Plot #2: run the analysis but change it to day of the week and use myProbs <- c(0.50,0.75,0.90).

Plot #3: run the analysis but change it to month of the year and use myProbs <- c(0.50,0.75,0.90).

Note that your report is a somewhat degenerate report compared to our usual style. Your report should follow the following format. Your report is not just a single page with three plots on it.

(1) Introduction:
(2) Plot #1: Plot and discussion of the plot.
(3) Plot #2: Plot and discussion of the plot.
(4) Plot #3: Plot and discussion of the plot.

Assignment Document:

The report document should be submitted in pdf format. File should be named Assignment3_LastName_FirstName.pdf.