140.886: Intro to R

Homework 1

1/7/2013

Name:

Please submit:

1. Annotated R analysis script (emailed to me)
2. Answers written/typed on this sheet (turned in tomorrow morning)

For (a), the script should be source-able (ie have no errors) and divided into sections (corresponding to questions) using comments. The answer to the question should be the last line(s) of code evaluated in that section. For example:

#####

# q1

[code goes here]

[have answer print here]

Instructions:

1. Download the ‘Charm City Circulator Ridership’ dataset from OpenBaltimore: <https://data.baltimorecity.gov/Transportation/Charm-City-Circulator-Ridership/wwvu-583r>
2. Consult the Charm City Circulator website (<http://www.charmcitycirculator.com/>) to better understand the data that you are working with
3. All questions require a code and answer, unless the question is designated with [NC] – no code – just require a response and do not require code or [CO] – code only, which do not require a response, and only require code

Questions:

1. How many lines are currently served by the Charm City Circulator? What is/are the name(s) of the line(s)? [NC]
2. How would you get to Camden Yards (the baseball stadium where the Orioles play) after class? [NC]
3. How much does it cost for a one-way trip on the Charm City Circulator? [NC]
4. Create a second dataset that only contains the columns: day, date, average ridership per line, and daily. Display the first 4 rows of this new dataset. [CO]
5. How many lines were in service on:
   1. 01/16/2010:
   2. 06/17/2010:
   3. 03/7/2012:
   4. 07/4/2012:
   5. 10/29/2012:
6. How many days in the dataset did the circulator not operate (i.e. how many days had 0 riders)?
7. What day did service start for each of the lines?
8. Which day of the week had the highest average ridership, and what was the overall average ridership?