

DS5110 HW 5 - Due March 22

Kylie Ariel Bemis

3/12/2019

Instructions

Create a directory with the following structure:

- `hw5-your-name/hw5-your-name.Rmd`
- `hw5-your-name/hw5-your-name.pdf`

where `hw5-your-name.Rmd` is an R Markdown file that compiles to create `hw5-your-name.pdf`.

Do not include data in the directory. Compress the directory as `.zip`.

Your solution should include all of the code necessary to answer the problems. All of your code should run (assuming the data is available). All plots should be generated using `ggplot2`. Missing values and overplotting should be handled appropriately. Axes should be labeled clearly and accurately.

To submit your solution, create a new private post of type “Note” on Piazza, select “Individual Student(s) / Instructor(s)” and type “Instructors”, select the folder “hw5”, go to Insert->Insert file in the Rich Text Editor, upload your `.zip` homework solution. Title your note “[hw5 solutions] - your name” and post the private note to Piazza. **Be sure to post it only to instructors**

Part A

Problem 1

Choose one of the “miniposters” created by your fellow classmates and posted on Piazza for Homework 3. Cite both the name of the student whose miniposter you chose and the original source of the dataset used in that miniposter.

Download and import that dataset into R, put it into a tidy format (if necessary), and print the first ten observations of the dataset.

Problem 2

To the best of your ability, reproduce the figures from the miniposter you chose. You may contact the author of the original miniposter; if you do, cite and describe any information you receive from them.

(If you are contacted for information on reproducing figures from your own miniposter, you may provide it, but you are not obligated respond.)

Part B

Problems 3–5 use the `BostonHousing` data from the `mlbench` package. Install the `mlbench` package from CRAN and use `data(BostonHousing)` to load the dataset.

Problem 3

Fit a model that predicts per capita crime rate by town (`crim`) using only *one* predictor variable. Use plots to justify your choice of predictor variable and the appropriateness of any transformations you use. Print the values of the fitted model parameters.

Problem 4

Plot the residuals of the fitted model from Problem 3 against the predictor variable already in the model *and* against other potential predictor variables in the dataset. Comment on what you observe in each residual plot.

(You do not need to print plots for predictor variables not in the model if you observe no systematic patterns in them; you can use `eval=FALSE` in your code block to suppress that code block in your knitted output.)

Problem 5

Fit a new model for predicting per capita crime rate by town, adding or removing variables based on the residual plots from Problem 4.

Interpret the model.