

Homework 2

Dr. Isuru Dassanayake

2023-04-01

Question 1:

For this question use the `Japan_earthquakes.csv`. This data set contains information regarding earthquakes from 2001-2018 in/around Japan.

Please create an interactive map using `leaflet`. Answer the following questions.

(Please export your maps as images for each part for the pdf file as html maps will not be supported when converted in to pdf files.

You can use `always_allow_html: true` on the header of your rmd file, so that you can execute the codes. Please see the rmd file attached.

You must include the all relevant codes as well.)

1. Use geolocation of Japan (lat and long values) to set up a base map. (You can use any `providerTiles` that is appropriate).
2. Create subset, of all earthquake data which has a magnitude (`mag`) of 6 and higher.
3. Using the subset from part 2) plot all the earthquake locations on the map of Japan.
4. Use `Circles` to represent the location of earthquakes and make sure that the `area` of the circle corresponds to the magnitude of each earthquake. *(Hint: Use the following formula for the radius, $\sqrt{(10^x) * c}$: Here x is the magnitude and the c is a arbitrary constant that you can choose your own).*
5. Adjust the circles that you got from part 4), so that the map is aesthetically pleasing. *(Hint: colors, opacity etc).*
6. Add `labels` to each circle indicating the `place` that each earthquake occurred.
7. Add a `popups` to each circle to see the following information:
 - Time of the Earthquake
 - Magnitude of the Earthquake
 - Depth in (km) of the Earthquake
 - Place of the Earthquake.

Question 2:

Is there a relationship between student math test scores and socioeconomic variables? The data set `CASchools.csv` contains data on test performance, school demographics, and student demographic background for school districts in California. Remove the variable `county` before the analysis.

Please use the Data set description document to learn more about the data set.

1. Fit a linear regression model to predict `math` test scores using all variables in the data set. Discuss your results, making sure to cover the following points:
 - a. What do you observe about the relationship between these predictors and math test scores?

- b. Are there any insignificant variables?
 - c. Which predictor is explaining math scores the best?
 - d. Explain the interpretation of the coefficient on lunch in terms of the response variable.
- 2. Do you think you can fit a better model than the model in part (1)? Fit a new model that you think is a better fit for the data. What do you observe about this new model?
- 3. Write the equation for the least square fit from part 2).
- 4. Compare the R^2 and RSE for models from part (1) and (2).
- 5. Using a residuals plot for part 1 and part 2 models, state the appropriateness of using a linear model for this problem.
- 6. Incorporate an interaction term into your multiple linear regression and respond to the following:
 - a. Provide analytical justification for your interaction term.
 - b. Compare your results with the models from parts 1) and 2).