

Statistical Methods for Data Science Mini Project 2

Exercise 1 (10 points):

- (a) (10 points) Use R to make three maps of the states in the USA. The first map should plot state level income share of the top 1% of income earners in 2012. The second map should plot the same variable but for 1999. The third map should plot the 2012-1999 difference of the variable.

Visit the website:

http://www.shsu.edu/eco_mwf/inequality.html

to download the data and learn about the variable of interest.

Read the handout about map making in R for a prototype example.

Simply replace “India” with “USA” in the getData function in the handout to get a shape file of the states in the US.

2 Bonus points: The handout does not show how to label the states with their abbreviations. Label the states with their abbreviations as has been done on the maps shown on the website: http://www.shsu.edu/eco_mwf/inequality.html. You should be able to figure this out by googling.

- (b) What do the maps show? Justify your conclusions.

Exercise 2 (10 points):

- (a) Read about the Happy Planet Index on Wikipedia:

https://en.wikipedia.org/wiki/Happy_Planet_Index

Be sure to read the methodology behind the index and its criticism. Download the data for 2012 from <http://www.happyplanetindex.org/data/>.

- (b) Examine the distribution of the HPI variable graphically. What would be appropriate measures of center and spread of this distribution --- (mean, SD) or (median, IQR). Justify your answers.
- (c) Make scatterplots of HPI against each of the three variables on which the index is best. Comment on what you see. Will it be appropriate to use correlation to summarize the relationship of HPI with the other three variables? If yes, provide

the correlations. Explain your answers. Sample correlation can be obtained using the 'cor' function in R.

Instructions:

- Due date: Tuesday, October 6.
- Total points = 20
- Submit a typed report and include any relevant plots.
- You can work on the project either individually or in a group of no more than two students. In case of the latter, submit only one report for the group, and include a description of the contribution of each member.
- Do a good job.
- You must use the following template for your report:

Mini Project #

Name

Names of group members (if applicable)

Contributions of group members

Your answers and justifications for each exercise.

Provide the R codes in an appendix. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.