**Statistical Methods for Data Science**

**Mini Project 2**

**Exercise 1** (3 points):

Use R to make a map of the states/provinces/regions showing GDP or income inequality in a country of your choice (except USA). Please mention the source of the data you used to create the map.

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

Simply replace “USA” with “the country of your choice” in the Data function in the handout to get a shape file of the states/provinces/regions in the country of your choice.

The maps should label the states/regions/provinces. 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.

**Exercise 2** (7 points):

Mixture models form one of the most fundamental classes of generative models for clustered data. This will have a multimodal distribution.

The following are run times (in seconds) on unix servers in 100 universities.

30.16 30.36 97.83 101.59 106.42 30.75 100.10 103.30 101.73 25.48 98.90 31.41 26.33 32.35 96.52 31.93 108.32 99.72 101.11 103.92 97.87 97.83 99.22 97.51 103.24 29.31 29.82 98.42 34.28 27.12 99.28 103.77 102.61 27.22 97.71 105.96 102.41 30.38 101.73 98.59 100.14 99.09 27.44 100.37 99.84 97.34 101.17 99.14 97.41 99.92 101.31 104.61 100.71 30.62 103.57 28.35 108.12 100.05 31.84 28.80

98.47 27.99 105.05 33.33 100.09 23.57 101.68 95.62 102.10 98.77 100.93 98.68 27.00 102.04 100.88 98.79 102.58 27.40 29.01 29.57 97.16 96.60 105.35 97.74 100.97 101.88 96.75 29.01 98.08 99.63 99.41 101.96 26.70 31.66 98.29 103.51 99.28 99.10 33.36 100.36

Using mixture normal distribution (a bimodal distribution)

0.7×N(µ1,σ12)+0.3×N(µ2,σ22)

(a) (4 points) find the maximum likelihood estimates of the unknown parameters and their standard errors. Use any appropriate R package or otherwise.

(b) (3 points) Draw a histogram of the data and superimpose the density of the above mixture normal distribution using maximum likelihood estimates of the unknown parameters.