

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

EP219 Data analysis and interpretation

Assignment 1

Dated : 5 - 8 - 2018

Initial steps

1. Take a look at the text file `pre_primary_school_data.csv` that shows the statewise data of the number of pre-primary schools and number of male and female teachers.
2. If you want to find other interesting data sets like this released by the government of India you can take a look at `data.gov.in`

You assignment

1. Extract the full data set to a numpy array. For this assignment we will only be interested in Pre - Primary Schools Having LKG / UKG / Nursery Classes (relevant data is in columns M, N, O)
2. To this array add a column which shows the total number of teachers per pre-primary school.
3. Make a histogram of number states with a certain number of teachers per pre-primary school. To be more explicit the x-axis of your histogram should be the number of teachers per school, the y-axis is the number of states. You will have to choose the bin-size of your histogram x-axis wisely so that your bins are neither too sparsely populated nor too densely populated. Make 3 separate histograms - one each for urban, rural and combined data.

Deadline

1. Add all plots to your report along with a discussion. Upload your code and report to your website by Monday, August 13th at 10 am.

Notes:

- Make sure to label all your plots, axes, title etc. Install latex so that you can use latex symbols in the plot legends.
- Try to experiment with histogram bins, axes range, colors, linestyle, plot markers, displaying multiple plots on the same image, saving plots to pdfs etc.
- Comment your code with detailed comments! Uncommented code will receive no credit.
- Try to follow best programming practices in python. <https://gist.github.com/slوريا/7001839>