In Lab 06

CS 251: Lab 07: Gnuplot (September 15)

- Please write (only if true) the honor code. If you used any source (person or thing) explicitly state it. You can find the honor code on the web page.

Overview

As they say, **Data is king**. Or queen if you like.

Data has exploded and we don't know what to do with it. Arguably, visualization of data has become very important. Further, no matter what position you take up in your later life – researcher, entrepreneur, administrator – charts and pictorial representation are a must. In this lab, you will hone your `gnuplot` skills.

Pre-tasks

1. Link your submission submitted in Lab 05 and Lab 06 if you have not already done.

2. Read up on gnuplot

Inlab tasks

There is only task here and it is arguably and deliberately ill-specified. You will find in the data folder three files with names starting with combined. You will also find another folder called individual, which has been grouped for you. Each of these files contain data.

We don't really don't know (or care) what these numbers are. They might indicate prices of various items in Flipkart, Snapdeal, and Amazon. Or they might indicate marks of students. Or they might indicate prices of shares. For sake of uniformity, let's assume "more" means "better."

The combined files may indicate, oh, let's say, average share prices in three groups. Stocks, Bonds, and International. These numbers probably came from the individual files, so let's go that folder. In each of the these folders you will see names coded with a prefix. These groups have been explained as being in three categories: blueChip (companies that historically well), ancient (companies that have been around for ever), and new (newer, dotCom companies). If I missed anything put them in the dotCom category.

Your goal is to help the CEO of the company to visualize the data. Here are some things you can consider.

1. Plot all the charts nicely. You can do multiplot, a mosaic, or whatever. You may have to decide how to bin them the companies. For example, the histogram bin size should be it 5 units or 10 units.

2. Group data and replot. You want to help the CEO make conclusions such as "while some dotComs do well, overall dotComs are a poor substitute for traditional".

3. Make sure you plot things like average and standard deviations.

How we will grade you

This assignment is intended to be like a friendly competition. So we will bring them up in the Tuesday and you can vote on which ones are nice.

Files To Submit

Submit a `tar.gz` of the folder containing all your script files and the usual `readme.txt`. This time the readme.txt file is going to be very important – it will summarize your findings. We have increased file size upload, so do submit important images (upto 10). As an alternate to readme.txt, you are encouraged to create a presentation using the Google Docs Presentation tool.

The folder and its compressed version should both be named `lab06_groupXY_inlab`. For example, for group 07, folder should be named `lab06_group07_inlab` and the related `tar.gz` should be named `lab06_group07_inlab.tar.gz`.

Please don't submit any unnecessary file which we can produce. Marks might be deducted for that.