Plotting Data

COMP 364 - Lecture 14
March 5, 2010
Derek Ruths
Why plot data programmatically?
Different kinds of plots...

- Line plot
- Scatter plot
- Histogram
- Heatmap
Line and scatter plots
Major considerations for line/scatter plotting

- Data consists of numbers
- Each data point has an X and a Y value
  - Data is specified as two lists (X values and Y values)
- Key issue: we read our data in as strings, but need it to be two lists of numbers.
Manipulating lists

- `x.append(y)` - add the object `y` into list `x`
- `x.remove(y)` - remove the first occurrence of `y` in list `x`

Exercise: Consider a file containing x-y datapoints - each line has two numbers, separated by a space. Read these points from the file into two lists.
Line plots

• matplotlib (pylab) is a 3rd party python library that provides MANY plotting functions (http://matplotlib.sourceforge.net)

• `pylab.figure()` - creates a new blank figure

• `pylab.plot(X,Y)` - draws a line plot using data points X,Y on the current figure

• `pylab.show()` - displays the current figure on the screen

Exercise: extend our previous code to plot the data points in a line graph.
Stylizing our plot

- `pylab.plot(X,Y,fmt)` - fmt is a string that tells pylab how our points should be drawn and connected.
  - `plot(X,Y,'r')` - draw in red
  - `plot(X,Y,'b')` - draw in blue
  - `plot(X,Y,'--b')` - draw a dashed blue line
  - `plot(X,Y,'g.')` - draw a scatterplot with green points
- `pylab.hold(True)` - tells pylab to combine future plots onto the current plot (rather than replacing it)

Exercise: modify our previous script to draw a scatter plot. It also should take a threshold. All data points with a y-value > threshold should be drawn in green, otherwise blue.
Annotating a plot

• `pylab.title(s)` - set the title of the current plot to s

• `pylab.xlabel(s)` - set the label of the x axis to s

• `pylab.ylabel(s)` - set the label of the y axis to s

• `pylab.legend([c1,c2,...])` - draw a legend on the figure labeling each curve

Exercise: make the title of our plot the name of the data file, make a legend for the two colors.
Sub plots

```python
pylab.subplot(# rows, # cols, plot #)
```

`pylab.subplot(2,1,1)`
`pylab.subplot(2,1,2)`

Exercise: write a script that makes a figure with 2 subplots: one for sin, one for cos. (plot for $x = [0,6]$)
Histograms
Exercise: plot the distribution of gene lengths in a genome file

Exercise: use subplot to plot (1) the distribution of gene lengths in a genome file and (2) the length of genes along the genome (in order)