COMP364: Working with Matplotlib (2)

Jérôme Waldispühl, McGill University

Analysis of human mRNA

Using CURL download human mRNA file on your account from:

http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/refMrna.fa.gz

Unzip the file unzip gunzip : gunzip refMrna.fa.gz

Write a parser that reads the file and store the mRNA into a dictionary such that:
- entry : sequence id
- content : sequence in IUPAC code

Using matplotlib, create histograms that plots:
- An histogram of the distribution of mRNA lengths.
- A figure that plots the GC-content (x-axis) and AG-content (y-axis) of each mRNA.
- Same as above but length vs. GC-content

Multiple figures

Continue your work with the mRNA file. Build a graph that plots:

• The length vs. GC-content (in red) and the length vs. AG-content (in blue)

• A figure that plots the GC-content (x-axis) and AG-content (y-axis) of each mRNA.

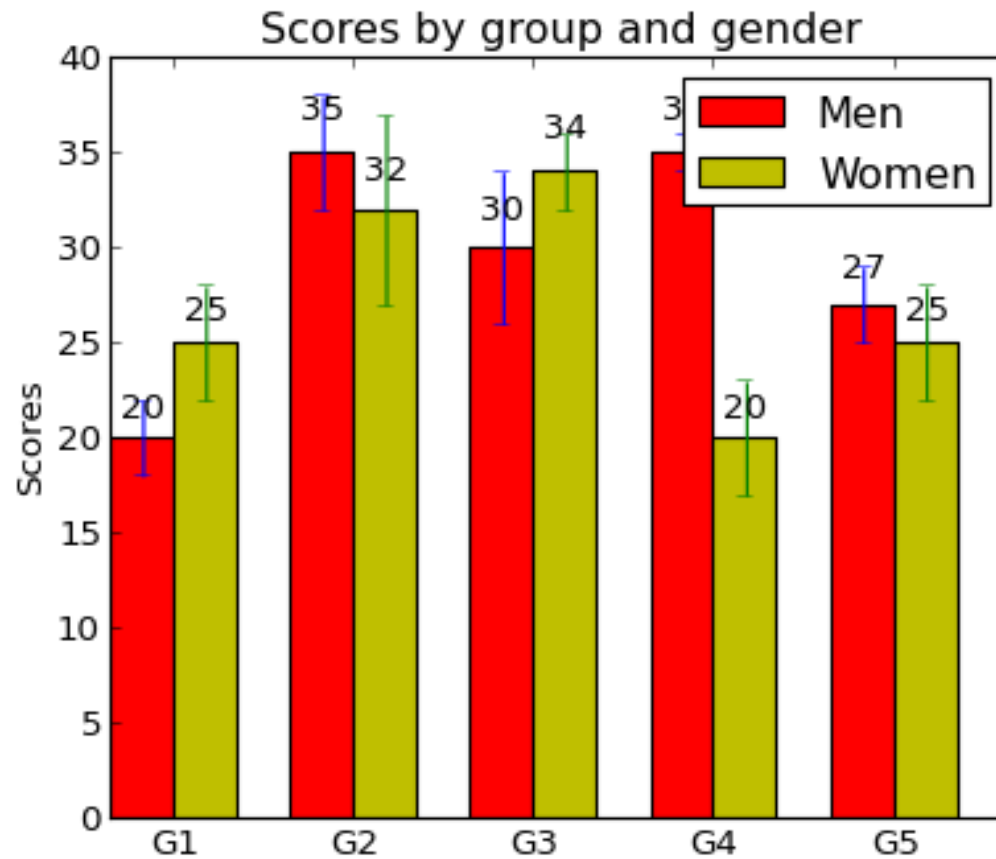In each figure label the axis, include a legend and a title.

Using the command:

    fig.savefig('myfigure.pdf',transparent=True)

Save your figures in a pdf file.

Assemble your two graphs within the same figure.

Bar charts

http://matplotlib.org/examples/api/barchart_demo.html

Amino acids distribution

Using the code you created for the last question of the first assignment, modify the code available at

http://matplotlib.org/examples/api/barchart_demo.html

to build an histogram that shows the distribution of amino acids.

The names of the amino acid must be displayed at the bottom of each bar.

N.B.: You will be mainly interested by the functions:

.bar(), .set_xticks() and .set_xticklabels()