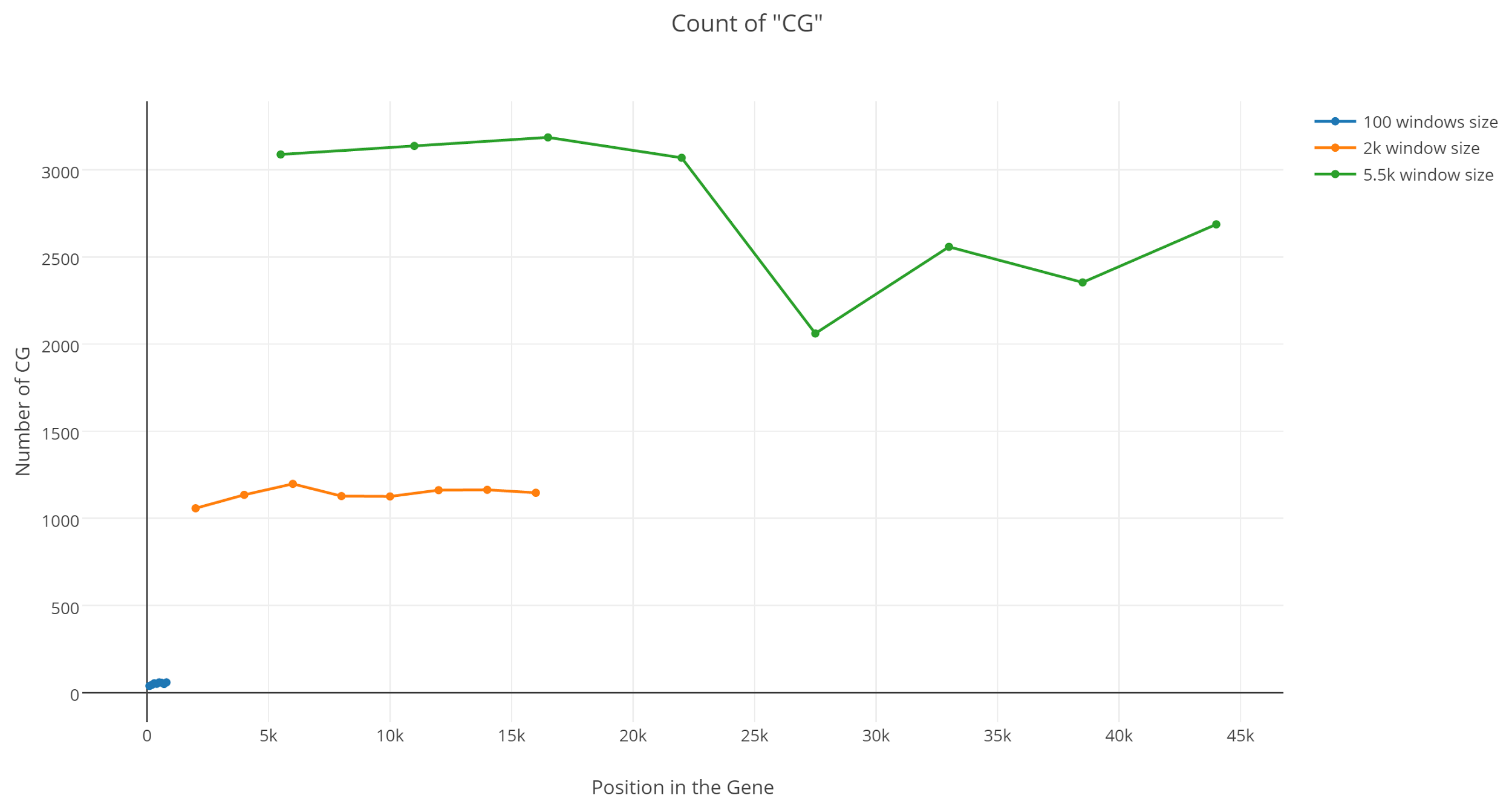
**Bio Informatics**

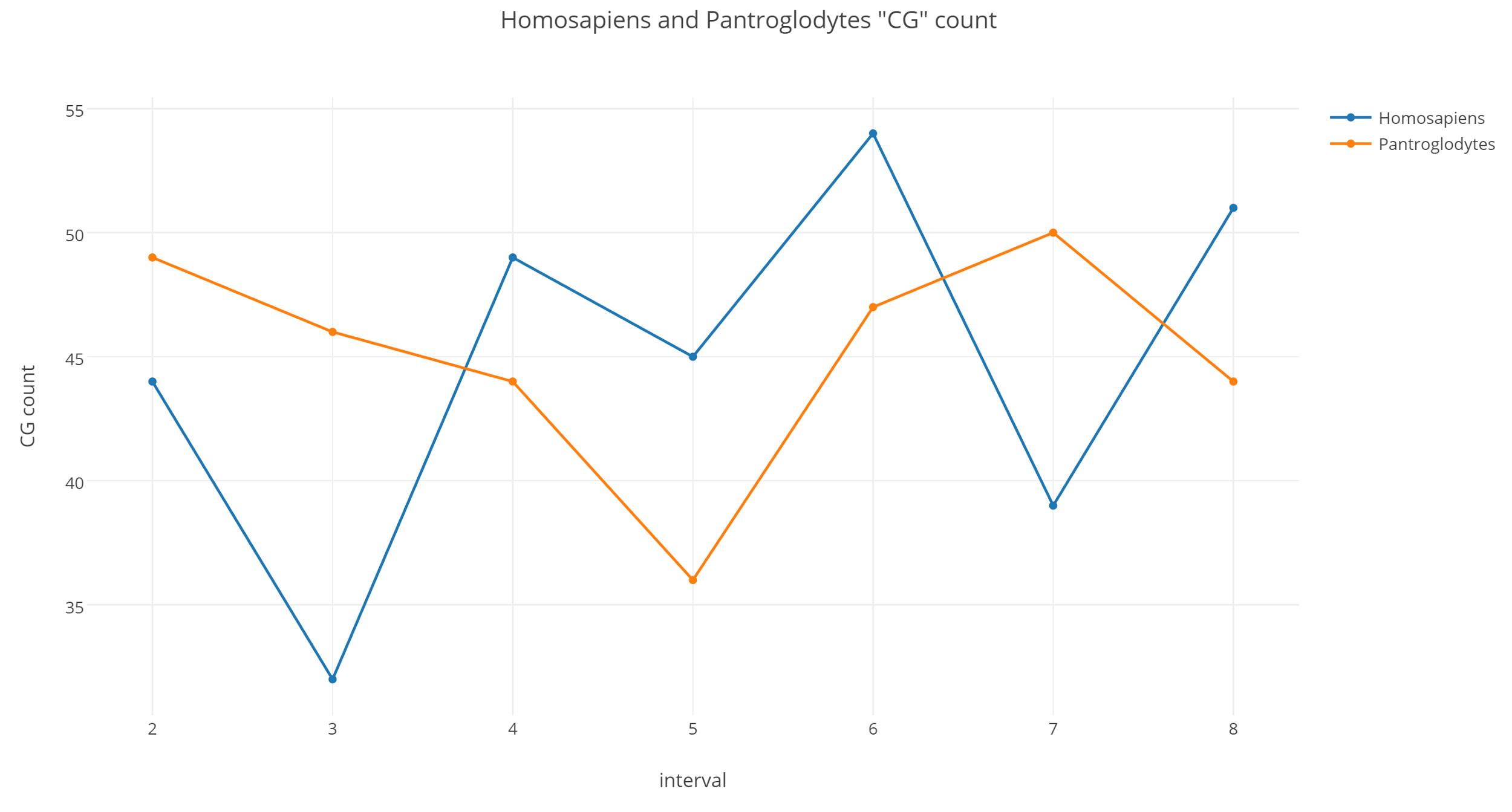
**Assignment 1**

The first part of the assignment was to download the gene sequence from the *Bacteriophage Lambda* (#  NC\_001416) from GenBank, write a program to count the “GC” count with different window sizes, and plot a chart. The code to accomplish this task was written in Python and is located on Github on the following address: [sequence.py](https://github.com/ICMC/Computation-Biology/blob/master/modulo1/NC_001416/sequence.py)

On Figure 1, the chart analyses the Bacteriophage Lambda “GC” count with window size equal to 3. You can observe that the bigger the window more “GC” count it has. I only plot 8 intervals. By this graph, you can understand that the smaller the window more information you have about the “GC” variation along the gene. 

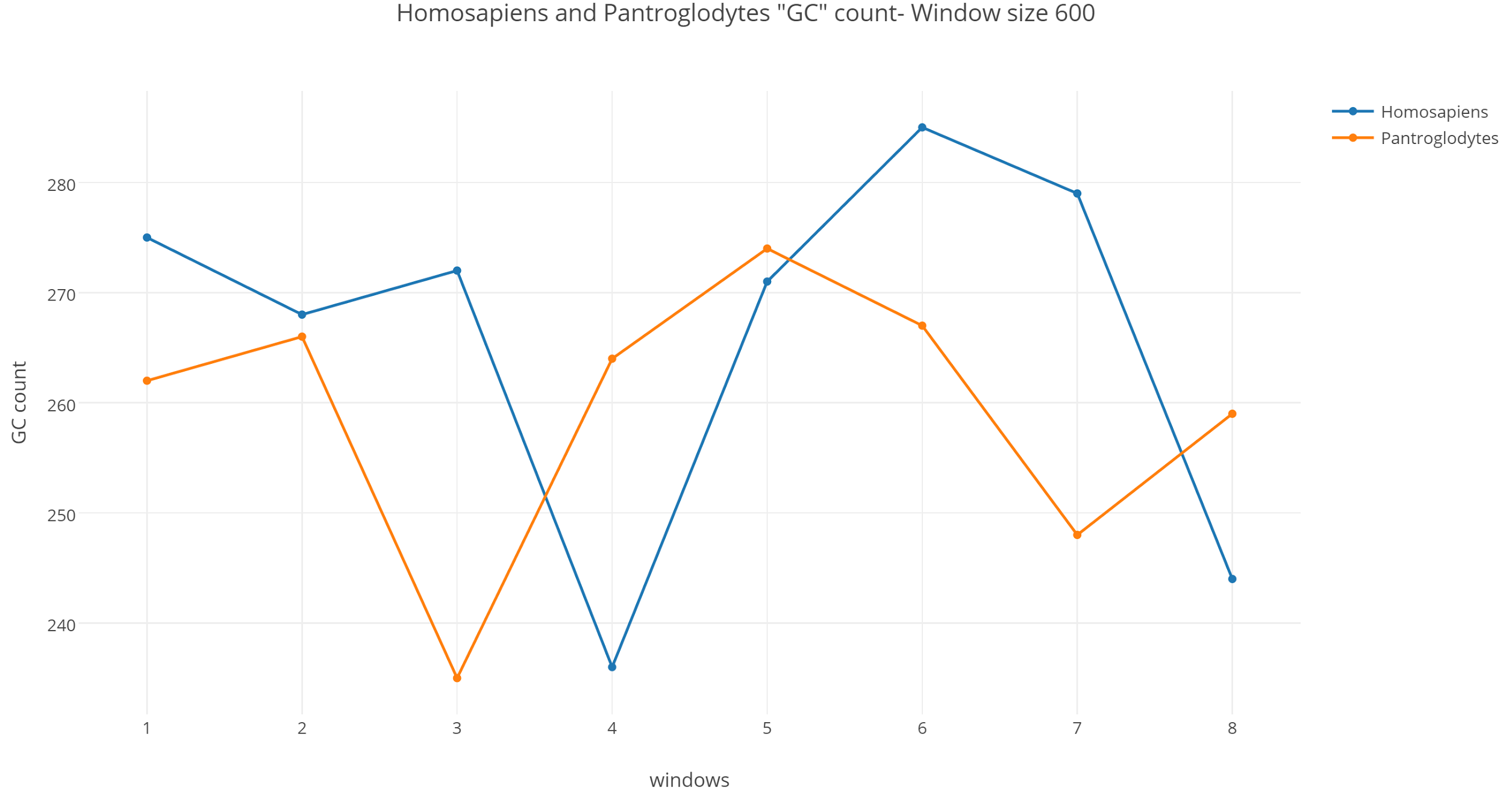
*Figure 1. Bacteriophage Lambda “GC” count chart*

The last chart, Figure 2,3,4, show the relation between the “GC” count between the Homosapiens and Pantroglodytes mitochondria with different window sizes. The Figure 2 has a window of size 100. You can observe that is very detail the differences between the two gene sequences. While Figure 3 and 4 , gthe distribution of the “GC” count look more alike between the gene that belongs to the Homosapiens and the one that belongs to the Pantroglodytes. The source code for the Figure 2,3,4 chart is located at: [mitocondria.py](https://github.com/ICMC/Computation-Biology/blob/master/modulo1/MitocondriaHomemChimpanze/mitocondria.py)

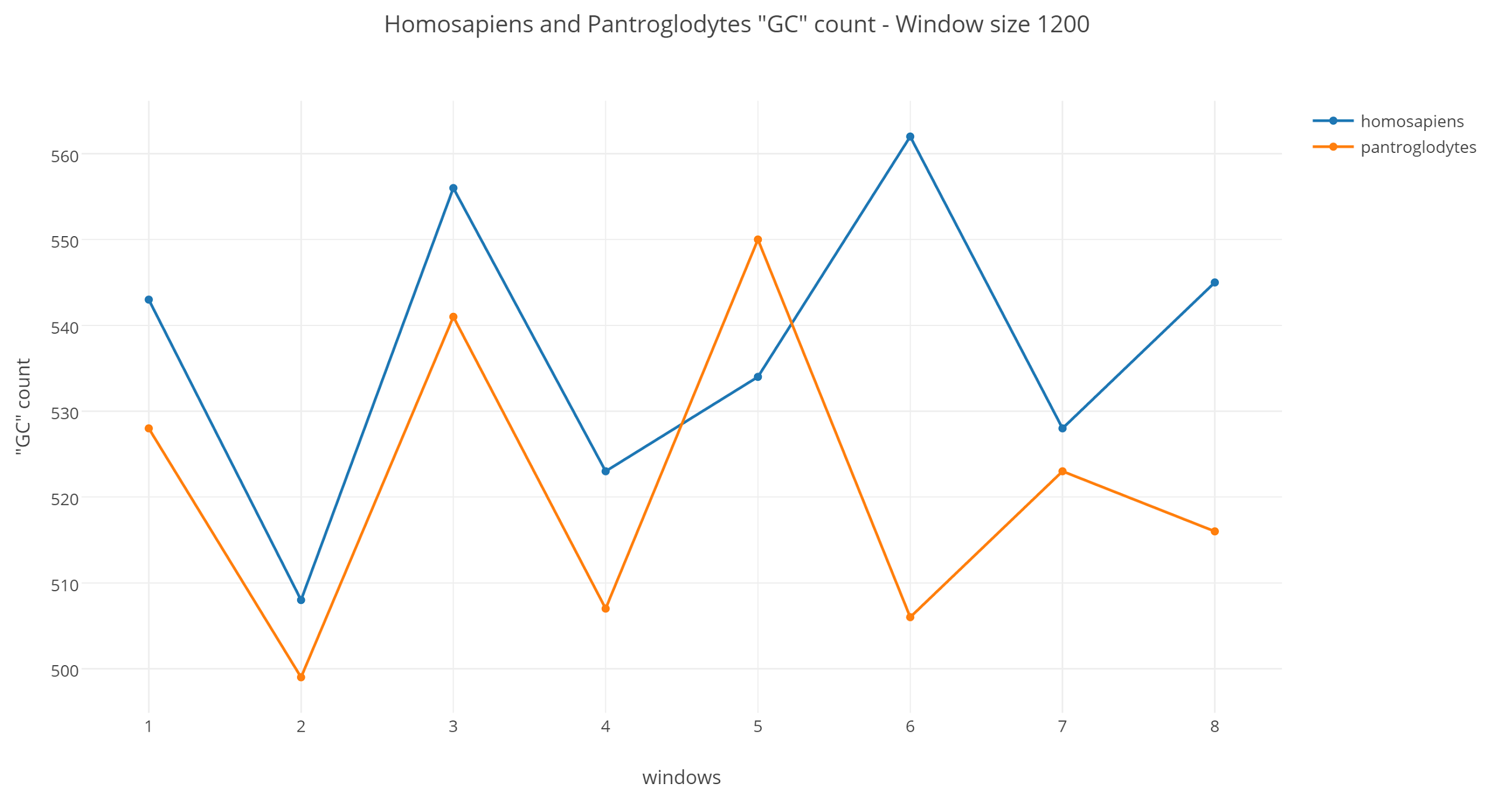


*Figure 2. Mitochondria “GC” count Window Size 100*

There are more similarities between the “GC” count of Homosapiens and Pantroglodytes when the window size is bigger. It is important to notice that since the window size is bigger is covering a larger piece of the gene sequence. The first graph, Figure 2, is covering from position 1 until 800, which means it is covering only the first 5% of the gene. The following graph, Figure 3, is representing the gene sequence from position 1 until position 4800. Since the gene sequence has size approximately 16000, this graph is only covering the first 30% of the gene. The last graph, Figure 4, is covering the first 60% of the gene sequence.



*Figure 3. Mitochondria “GC” count Window Size 600*

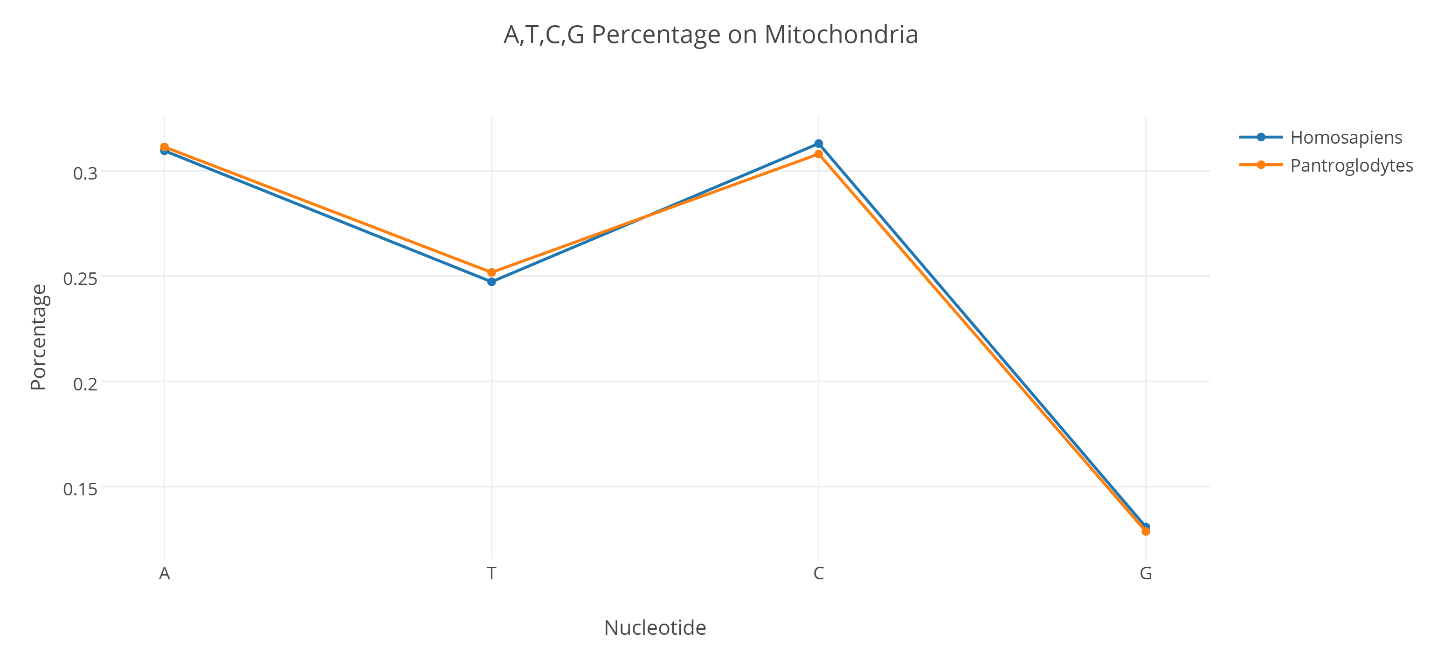


*Figure 3. Mitochondria “GC” count Window Size 1200*

The task for the following charts was to analyze the gene sequence of the Mitochondria of two different species. The first specie is Homosapiens Sapiens, the second species is Pantroglodytes.

Figure 3 shows the percentage of the nucleotides A,T,C, and G of both species. It is possible to observe that both genes have almost an identical percentage of those nucleotides on the mitochondria gene sequence. Meaning that the mitochondria that belongs to the Homosapiens and Pantroglodytes function alike.

The source code for Figure 3 is located at: [ACGTprobability.py](https://github.com/ICMC/Computation-Biology/blob/master/modulo1/MitocondriaHomemChimpanze/ACGTprobability.py)



*Figure 3. A,T,C,G Nucleotides Percentage*