DS 294: Data Analysis and Visualization- Assignment 1

By Muttaqi Ahmad Alladin
**No:**06-18-01-10-22-20-1-18409

Aim: To install one open-source visualization program and briefly describe the program and salient features.

Introduction: I am an MTech (Res) student at CDS and am interested in carrying out research in Bioinformatics. Hence, I chose to install "circos" for this assignment as it is an open-source tool that is extensively used to visualize genomic data.

Circos is an open-source data visualization software that is written in Perl. It visualizes data in a circular layout and has a high data-to-ink ratio. This makes circos ideal for creating publication-quality visualizations, and visualizations made by circos have appeared in high-quality journals like Nature, Science and PLoS. Circos is completely command-line and has no interactive GUI (Graphical User Interphase). While command-line may be difficult initially, it is better over time as it is easy to automate the generation of visuals using command-line. Circos is highly scriptable and automatable. While circos was initially developed to visualize genomic data, it is now also used in various other scientific domains where many relationships need to be visualized efficiently.

Salient features:

1. **Circular Visualization:** Circos uses a circular layout to show connections/links between objects or positions, which are difficult to visually organize when the layout is linear. In such cases, a linear form makes it impossible to keep the relationship lines from crossing other structures, which deteriorates the graphic's effectiveness. An example of this is given below.
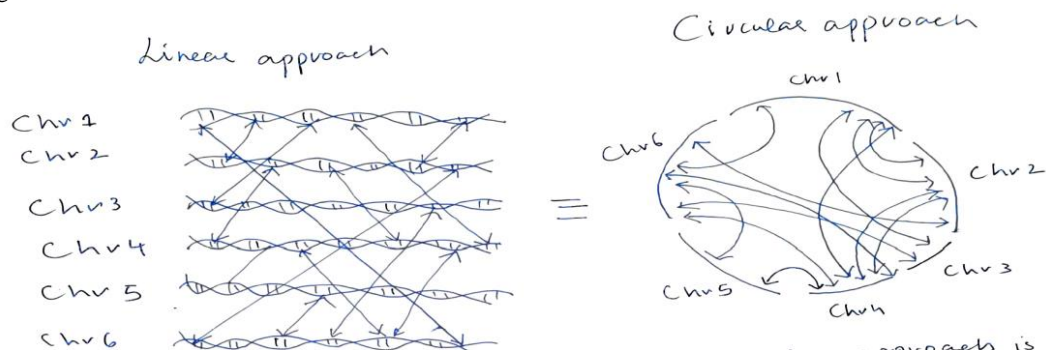
Both the figures above are trying to visualize the exact same relationship between 6 chromosomes. It is evident from the figure that it is easier to visualize relationships in a circular manner. Moreover, there are no sharp angles in circular plots, and sharp angles are difficult to track by the eye.

2. **Plot types:** Circos supports a wide variety of plot types like paired-location, scatter, line, histogram, heat map, tiles, glyph and text elements plots. Moreover, it is possible to combine different plot types in a single track. Circo also allows us to tune colours according to our needs. [source: circos.ca]
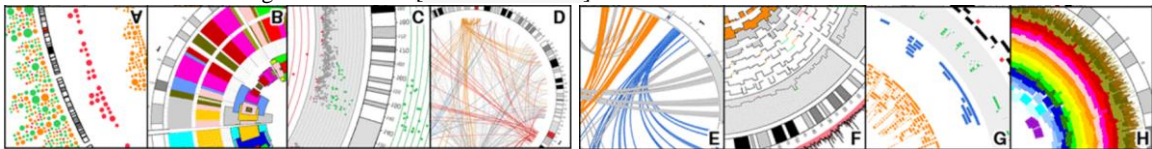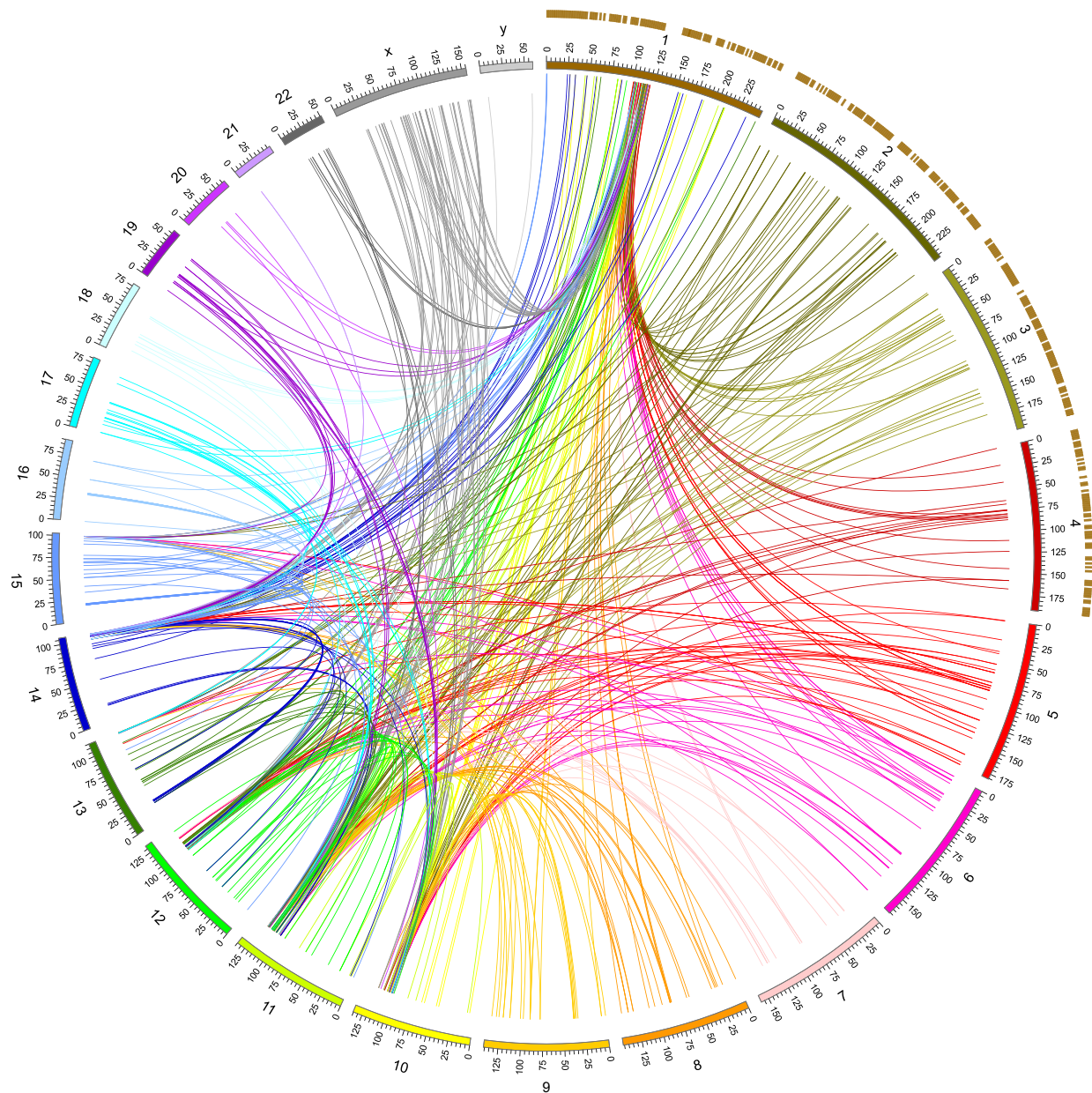

(A) glyph (B) highlight with depth control (C) scatter (D) paired-location (E) ribbon (F) histogram(G) tile (H) highlight

3. **Stack Plots:** As can also be seen from the visuals examples, it is also possible to stack different kinds of plots in the same visualization.
4. **Dynamic Rules:** Circos supports dynamic rules to adjust the plot's formatting elements based on the position, value, and formatting. Circos also has support for global and local zoom.
5. **Open-source and automatable:** Circos is an open-source command-line tool that is written in Perl and is highly automatable and scriptable.

Materials and Methods: Circos (version 0.69-9) was downloaded from http://circos.ca/software/download/ to my UNIX system. After installing all its dependencies using CPAN, circos was ready to be used. A config file conf1.conf was created to generate the visualizations. The datasets used were provided by circos.ca. The config file, along with the data sets, can be downloaded from my Github repository (https://github.com/muttaqiahmadalladin/Data-Analysis-and-Visualization/tree/main/Assignment1)

Results: The visualizations thus created are shown below.

The above visualization shows some of the most significant salient features of circos. It is a circular plot that is showing actual human karyotype data. The above plot shows 24 chromosomes (22+ X & Y). The length of the ark corresponds to the size of the chromosomes. The above visualization, despite having 726 links, is readable, and it is also possible to see patterns in the data. There is also the use of dynamic rules in the above plot. If we observe, it becomes clear that the links in chr1 follow a different set of rules than other links. I have also taken advantage of the fact that circos supports multiple plot types and can stack plots of another kind in the same visual, by having stacked a heatmap on chr1,2,3,4.

References:

- http://circos.ca/
- http://barc.wi.mit.edu/education/hot_topics/Circos/Circos.pdf
- https://www.youtube.com/channel/UC1nmdp9D424a2xNJx9FAZ-A
- https://davetang.org/muse/2012/11/26/getting-started-with-circos/