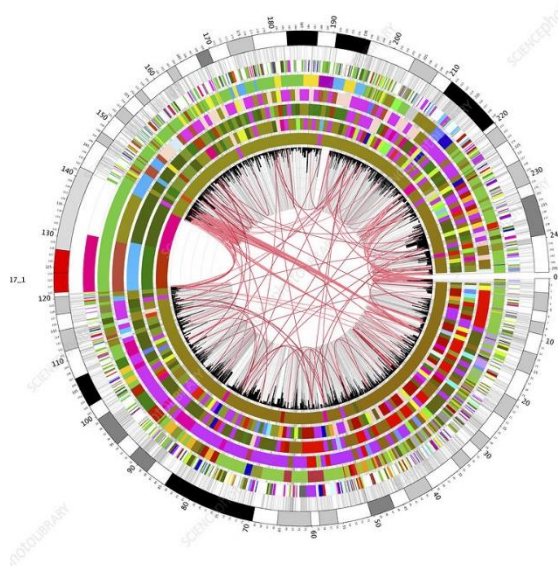


Assignment 0

Zhuoyang CHEN 50005220

1. I want to take this course because I want to learn more about data exploration. I never attempt a Kaggle competition but I do have a concept of what's it about and how does it work. I hope to see how someone as an expert deal with a data set you are completely unfamiliar with and make something reasonable out of it.
2. Exploring data distribution, data clustering, making beautiful charts etc. but these I already had some experience in it. I do hope to learning more about improve the intuition on big data, human computer interaction, dynamic visualization.
3. I proficiently work with R via *ggplot* package and familiar with python, but also want to learn how to create more advanced visualization with other tools or languages.
4. Yes. In my previous job I usually do dimension reduction like PCA, tSNE and UMAP and then used clustering algorithms to render different clusters with different colors. I also create boxplots with statistical significance like p-values between groups. Some other charts like correlation chats, scatterplot between two variables, distribution barplot etc.
5. The largest one is a singlecell sequencing gene expression matrix about 12G. It contains 120,000 columns as cells and 20,000 rows as genes. I used a package *Seurat* to deal with such a dataframe structure.
6. It's a video about the dynamic change of territories of different warlords in the late East Han dynasty and three kingdoms era across years. It presents people from all over the world that how chaotic that time period was and a direct impression of how the three kingdoms developed from a tiny place to a large territory. For people who are not familiar with Chinese history this is an amazing way to show them how the wars went on and what warlords were involved.
7. The worst visualization could be a circular genomic map from personal perspective. They usually appear on journals about a plant or an animal. They are so crowdy and most of the time just for profiling data but not yielding any useful information and I doubt people ever take a careful look at it. It's more like a paradigm for publishing paper rather really providing any insight.



<https://www.sciencephoto.com/media/211579/view/circular-genome-map>