Take Home Assignment

General Guidelines and Evaluation Criteria

1. The solution for this assignment is expected to be a web app using R-Shiny/python-dash/streamlit or any other framework of your choice.

2. Your webapp should ideally not use absolute paths when loading data files. The UI is expected to be self-explanatory.

3. It is encouraged that you make good use of comments as explanations of your code.

4. You are encouraged to keep your code as modular as possible - feel free to make use of custom functions.

About the data

There are 4 datasets (chronos.csv, cn.csv, expression.csv and metadata.csv) attached to this assignment document.

-   *chronos.csv* - contains the chronos data
-   *cn.csv* - contains the copy number data
-   *expression.csv* - contains the expression data
-   *metadata.csv* - contains description about the data available in the rest 3 files

Apart from metadata.csv, all 3 datasets contain 100 gene names as identifiers in the columns and 908 sample IDs as identifiers in the rows.

**Please feel free to be as creative as possible.**

However, here are a few things which SHOULD be present in your work -

1. **Data Overview [MANDATORY]**
    a. Allow user to select the dataset.

b. Upon selection, display the dataset as an interactive table (preferably with options to search/sort/download data).

c. Display all summary statistics for the selected dataset - like number of rows, number of columns and others as you feel appropriate.

2. **Data Visualization [MANDATORY]**

   a. Scatter plot

      i. The webapp should allow the user to select genes from datasets and plot 2D **scatter plots** between 2 variables(expression/copy_number/chronos) for any pair of genes.

      ii. The user should be able to filter and color data points using metadata information available in the file "metadata.csv".

      iii. The visualization could be interactive - It would be great if the user can hover over the data-points on the plot and get the relevant information (hint - visit https://plotly.com/r/, https://plotly.com/python)

      iv. Here is a quick reference for you. The scatter plot is between chronos score for TTBK2 gene and expression for MORC2 gene with coloring defined by Gender/Sex column from the metadata file**.**
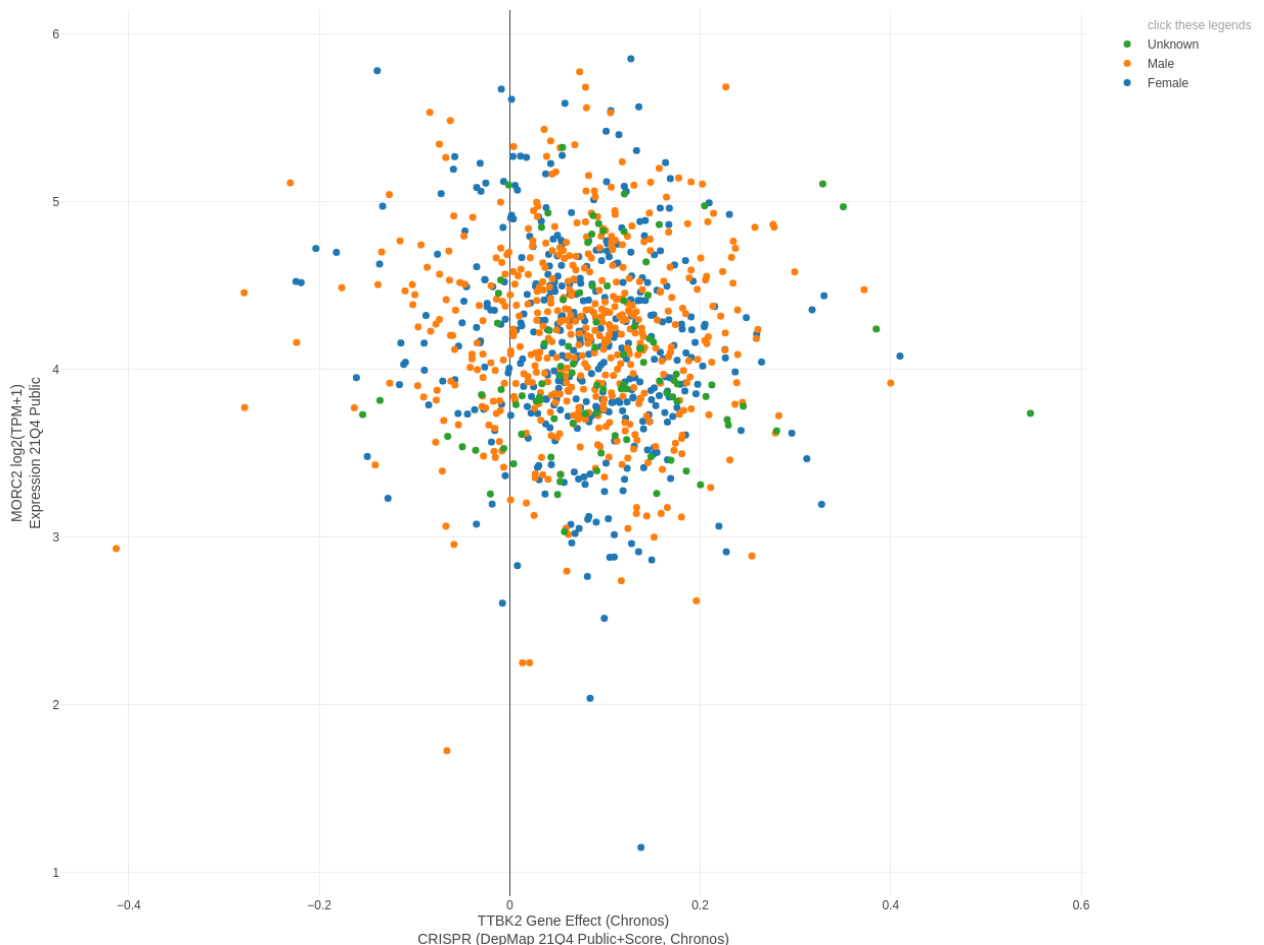
b. Boxplot/violin plot

   i. User should be able to select a gene and a variable (expression/chronos/copy_number) and generate a boxplot to display its distribution across multiple categories as defined by user selected variable (a column from the metadata file)

   ii. Here is an example for your reference where violin plot for CHRONOS score for gene CCL22 is plotted and grouped by 'Lineage'
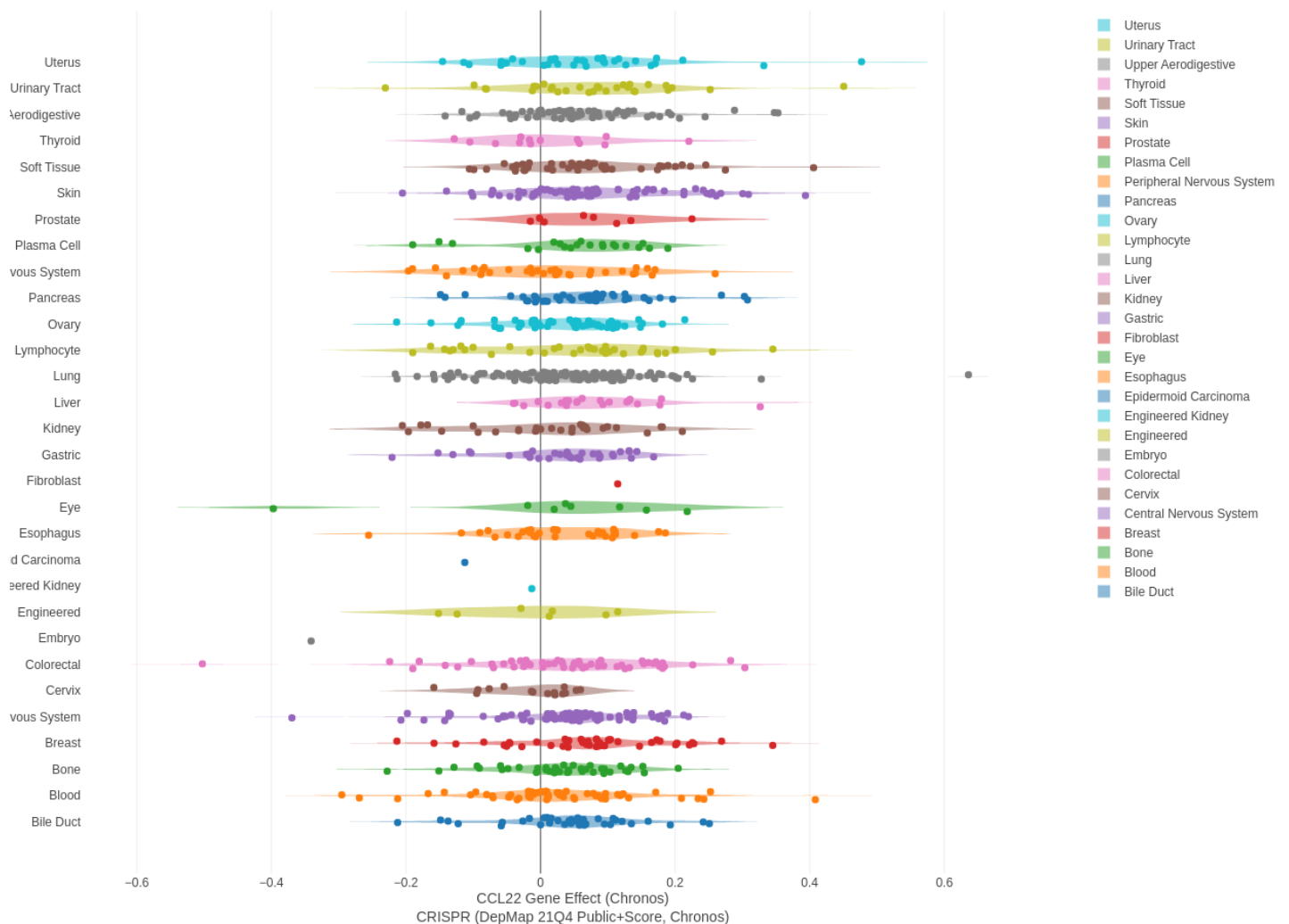
3. **Additional Features [OPTIONAL]**

Great! So, now that you have built the above 2 sections, if you think you can make use of other visualizations to summarize the data, feel free to add another section in the app with its implementation.

Bonus Section (OPTIONAL)

Now that you have a functioning app ready, let's share it publicly so that other people can use it as well.

  a. Dockerize your application

    (hint: https://stackabuse.com/dockerizing-python-applications/)

  b. Push the docker image to dockerhub

  c. Write an 'instructions.txt' file with two commands -

      i. How to pull the image from the dockerhub? (docker pull)

      ii. How to run the app in docker? (docker run)

Submission Guidelines

1. Create an 'info.txt' file and provide a list of all dependencies along with appropriate versions and other relevant environment information for your web app.
   Also, please include the command that runs/starts your application.
2. Include the 'instructions.txt' file if you have dockerized your app.
3. Put all your codes and text file(s) in a folder with your name and compress it
4. Send us the compressed file.

All the best!

Elucidata
www.elucidata.io