# CGS4144 Bioinformatics - Assignment 1

# Directions: Following the directions below, fill out this form then submit it on Canvas. Remove text surrounded by <> and replace with your information.

# Important information: Your scientific question and data must be approved by the instructor before you proceed to Assignment 2. You may check with the instructor/TA to ensure that both scientific question and data are appropriate for your project before submitting Assignment 1.

**Create your team:**

Add the full names, email addresses, and GitHub handles for each team member (3-4 people total) to this table:

| Name | Email | GitHub |
| --- | --- | --- |
| Ethan Fan | efan@ufl.edu | github.com/ethan12103 |
| Nathan Gilman | ngilman@ufl.edu | github.com/npgilman |
| Rama Janco | rjanco@ufl.edu | github.com/ramawama |
| Matheus Kunzler Maldaner | mkunzlermaldaner@ufl.edu | github.com/matheusmaldaner |

**Data:** <Link to the dataset you would like to use>

**Scientific Question:** <Question you will answer using the data you’ve selected>

**GitHub Repository for Project:** <Link to your team GitHub repository>

# Directions:

# Creating a Team:

# Find 2-3 other students in the class to create a semester-long group for your class projects.

# Add your names to the table above, one row per student.

# Creating a shared Github repository:

# Create the repository (<https://docs.github.com/en/get-started/quickstart/create-a-repo>)

# Add a README to the repository with information about the project including the data you will be using and the scientific question you will be investigating.

# Make sure the instructor and TAs have access to your repository.

# Selecting a DataSet:

# Navigate to the Refine.Bio website ([https://www.refine.bio/)](https://www.refine.bio/) and search for data.

# For example if I wanted to find mouse melanoma samples, I would enter “melanoma” (without quotes) in the search bar, then select Mus Musculus under the Organism filter and Rna-seq under the Technology filter.

# Make sure there are at least 50 samples in the dataset.

# Double check that you are choosing RNA-seq data. You will need a counts data matrix!

# Refine.Bio Guide: [docs.refine.bio/en/latest/](http://docs.refine.bio/en/latest/)

# Example dataset: [RNA-Seq of the rat pineal transcriptome](https://www.refine.bio/experiments/SRP049818/rna-seq-of-the-rat-pineal-transcriptome-with-in-vivo-and-in-vitro-samples-under-various-treatment-and-surgical-conditions)

# Asking a Scientific Question:

# Next, discuss with your group what would be interesting to know about this data. Are you interested in investigating differences in samples with a disease vs without? Disease severity? Another trait?

# Remember the class rules: be a good, respectful Gator. Ask questions that would not upset the patients

# Given the data you have selected, identify 2 or more distinct groups within that data. E.g. cancer vs healthy, drug administered vs not. Read the metadata included with the dataset for ideas.

# Fill in your question at the top of the document.

# Example questions:

# Can we identify the different tissue types based on gene expression?

# What are the expression differences between healthy and tumor samples?