Уровень каких белков различается в мышиной модели синдрома Дауна

Проект должен быть представлен в виде отчета в формате .rmd (достаточно загрузить в отдельную ветку вашего приватного репозитория этот файл).
Технические требования к компиляции и оформлению отчета указаны в конце файла
Дедлайн сдачи задания 20 февраля 23:59. С 20 января и до этого времени можно свободно задавать вопросы и обсуждать возможные варианты решения

В 2015 году на мышиной модели были проведены исследования как синдром Дауна влияет на изменения уровней различных белков. Данные для проекта можно скачать по следующей ссылке -- https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression#
Doi статьи на которую можно ориентироваться -- 10.1371/journal.pone.0119491.
Для этих данных вам необходимо:

1. Сделать описание датасета (5 баллов)
-- сколько всего мышей было в эксперименте
-- какие группы вы можете можете выделить
-- насколько эти группы сбалансированы
-- какое количество полных наблюдений (речь про NA)

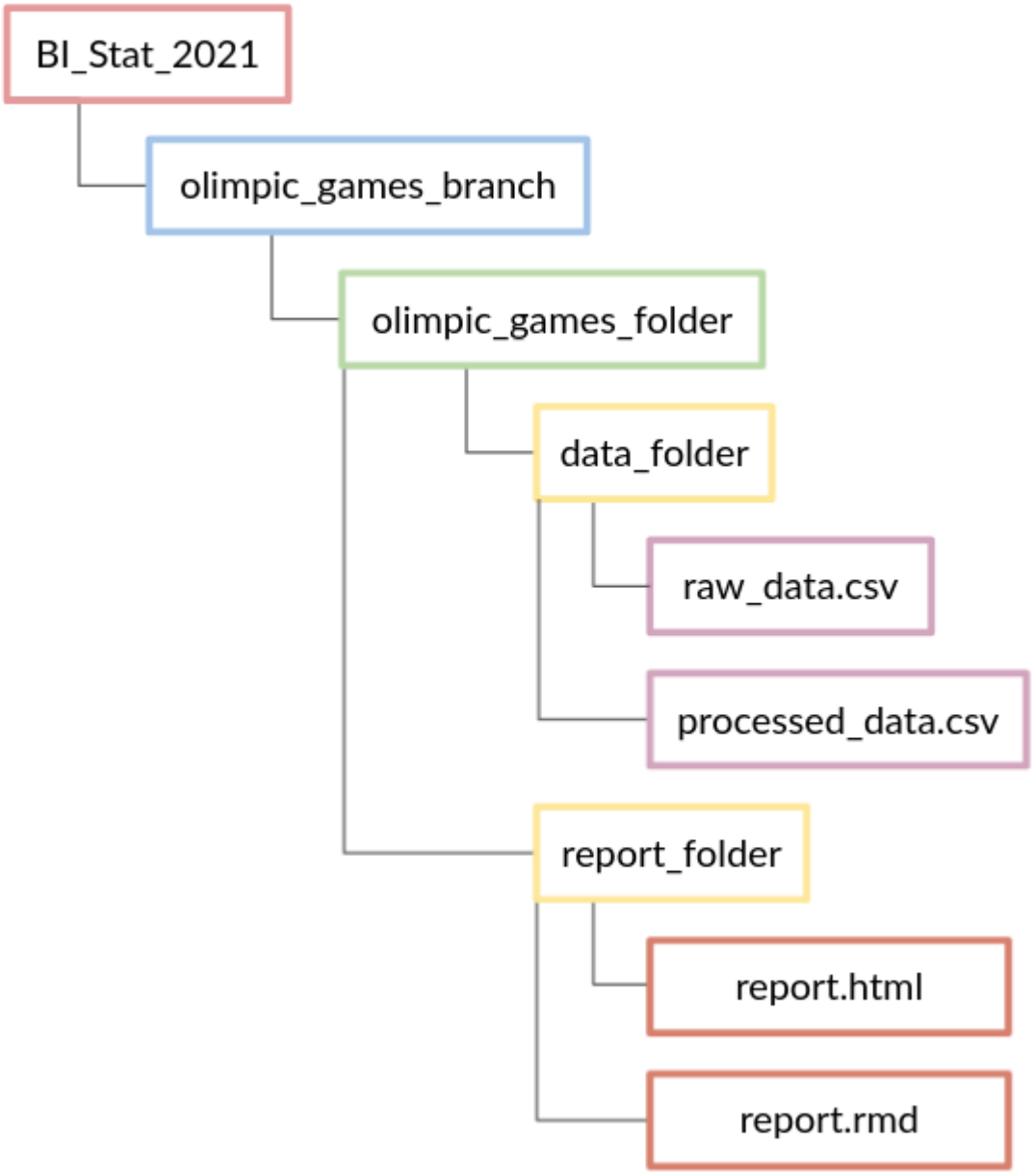
2. Есть ли различия в уровне продукции BDNF_N в зависимости от класса в эксперименте (10 баллов)
3. Попробовать построить линейную модель, способную предсказать уровень продукции белка ERBB4_N на основании данных о других белках в эксперименте (15 баллов)
-- провести диагностику полученной линейной модели
-- объяснить, почему это является хорошим/не хорошим решением
4. Сделайте PCA (15 баллов)
-- ординацию
-- постройте графики факторных нагрузок
-- определите, какой процент объясняет каждая компонента
-- постройте трехмерный график для первых 3-х компонент

5. Поиск дифференциальных белков -- творческая часть задания (15 баллов)
-- можно сделать реанализ из статьи, но предупреждаю сразу, что там машинное обучение
-- один из вариантов решения - использование методов направленной ординации
-- можно использовать limma/DeSeq2 (limma проще с осознания логики, также в limma и DeSeq2 немного по разному работают статистические тесты)
Дополнительные баллы по накопительной системе за каждую адекватную идею и её реализацию (до 15 баллов).

Технические требования к отчёту:

1. Отчёт должен быть представлен в формате rmd и скомпилированного html (можно в формате pdf).

2. Вам нужно будет создать репозиторий для нашего курса и назвать его *BI_Stat_2021*. Внутри него создать ветку для текущего проекта. В ветке проекта создайте папку (назовите ее также как и ветку). Внутри этой папки у вас должно быть две директории: data для данных, с которыми вы работаете и report для отчета в формате rmd и html (или pdf). В итоге у вас должно получиться нечто подобное:



Внутри папки olimpic_games_folder вы можете безболезненно создать отдельный файл Readme.md для данного проекта, чтобы коротко описать в нем то, чем вы занимались.

3. Ваш файл rmd должен компилироваться не только у вас на компьютере. За ошибки компиляции баллы будут сильно снижаться.

4. Все разделы отчета, в особенности графики должны быть оформлены в едином стиле. Подписи должны быть полными, логичными и читаемыми.

5. Ваш отчет должен быть универсальным. Так, чтобы при добавлении новых данных он работал корректно.

6. Отчет должен быть структурирован (например, загрузка данных -> краткий EDA -> задачи -> дополнительная часть (по желанию) -> выводы).
7. Ответы на задания должны быть выделены. Когда вы используете тот или иной статистический критерий, то нужно указывать условия его применимости, а в качестве результата приводите значения статистики, p value, число наблюдений и параметры, с которыми вы запускали функцию, если они отличались от параметров по умолчанию.
8. За отчет на английском языке можно получить **дополнительно** 5 баллов.
9. За хорошо оформленный Readme для проекта **дополнительно** 3 балла.