Привет! Совсем недавно наша научная группа вернулась из экспедиции, в которой наконец-то собрали все необходимые данные для исследования. Спасибо еще раз большое, что согласился помочь со статической обработкой.

Сбором данных руководил аспирант Ярослав (очень рассеянный парень), а помогали ему 10 студентов. Из-за этого все данные сохранены в наборе файлов, что очень неудобно для анализа. Для каждого моллюска измерены такие показатели как число колец, пол (мужской, женский и детская особь), длина моллюска, диаметр, ширина и длина раковины и другие. В первую очередь нас будут интересовать следующие вопросы:

1. Нам надо как-то объединить наблюдения в единую таблицу. Пожалуйста, напиши пользовательскую функцию, благодаря которой мы сможем собрать все наши наблюдения воедино. Так как экспедици являются ежегодными, а данные всегда называются по разному, то функция должна объединять все файлы одного расширения из заданной папки. Суммарно у тебя должно получиться 4177 наблюдений.
(8 баллов)
2. Посмотри, действительно ли все данные корректны? Если найдешь, что что-то не так, то исправь это, пожалуйста. Только объясни нам, почему ты воспользовался именно этим подходом. Может быть у него есть альтернативы?
(6 балла)
3. Рассчитай среднее значение и стандартное отклонение переменной Length для моллюсков разного пола.
(4 балла)
4. У какого процента моллюсков значение переменной Height не превышает 0.165?
(4 балла)
5. Чему равняется значение переменной Length, которое выше чем у 92% от всех наблюдений?
(4 балла)
6. Создай новую переменную Lenght_z_scores и сохрани в нее значения переменной Length после её стандартизации.
(4 балла)
7. Сравни между собой диаметр моллюсков с числом колец 5 и 15. К каким выводам ты пришел? Пожалуйста, оформи результаты так, чтобы мы сразу могли использовать их для статьи.
(4 + 2 балла)
8. Нас особенно интересуют переменные Diametr и Whole_weight. Что ты можешь про них сказать? Есть ли у нас основания предполагать, что они могут быть взаимосвязаны? Как ты это определил?
(4 + 2 балла)
9*. Наверное, пока ты работал с данными, ты заметил в них что-нибудь интересное. Ты можешь выдвинуть парочку гипотез и проверить их. Мы будем рады получить свежий взгляд со стороны)

Наглядные графики будут плюсом)

Технические требования к отчёту (за базовое соблюдение - 8 баллов)

- Отчёт должен быть представлен в формате .rmd и скомпилированного html.
- Оба файла, а также данные должны быть помещены в отдельную ветку вашего гит репозитория по курсу.
- Отчёт НЕ должен содержать вашего имени и любых других явных опознавательных знаков, которые позволят вас вычислить.
- Если ваш файл .rmd не компилируется на другом компьютере (я готов переписать только строку с путем до файла и установить пакеты. Хотя и это можно автоматизировать), то работа оценивается в ноль баллов.
- Все разделы отчета, в особенности графики в отчёте должны быть оформлены в едином стиле. Подписи должны быть полными, логичными и читаемыми.
- Ваш отчёт должен быть универсальным. Так, чтобы при добавлении новых данных он работал корректно.
- Отчет должен иметь структуру: заголовки, логичное разбиение на чанки
- Ответы на задания, требующие работы только с кодом, представляют собой соответствующие объекты. На задания по статистике требуется развернутый ответ с указанием всех необходимых данных.
- 9 задание - со звездочкой. В нем предполагается провести EDA, выдвинуть и проверить несколько гипотез в рамках тех методов, которые мы освоили. Никакого шаблонного решения тут нет. Будет оцениваться ваше умение "общаться" с данными. Баллы за это задание не являются обязательными
- Помните, что верхний порог баллов не ограничен!