

Визуализация биомедицинских данных

Домашняя работа №1

Вам предстоит выполнить задания ниже в RMarkdown документе. После чего результат (не просто сам .Rmd, но результат knitr'a) загрузить в ваш GitHub репозиторий¹. Домашнее задание сдаётся ссылкой на ваш репозиторий (проверьте настройки приватности).

Deadline: 23:59 21 октября 2022 г.

Домашнее задание оценивается по системе зачёт/незачёт. Зачёт ставится при выполнении любых 11 заданий. Любые спорные ситуации при оценке решаются в пользу студента.

Задания

1. Загрузите датасет `insurance_cost.csv` (лежит в папке домашнего задания). Это данные по базовым показателям здоровья индивида и сумме, которую страховая компания заплатила за его лечение в год. Обычно эти данные используют, чтобы потренироваться в предсказании того, как определённые характеристики индивида повышают траты страховой компании (и, соответственно, должны быть заложены в цену страховки).
2. Выведите гистограммы всех нумерических переменных.
3. Нарисуйте график плотности по колонке `charges`. Отметьте вертикальные линии средней и медианы на графике. Раскрасьте текст и линии средней и медианы разными цветами. Добавьте текстовые пояснения значения средней и медианы. Подберите тему для графика. Назовите оси.
4. Сделайте три `box_plot` по отношению переменных `charges` и (1) `sex` (2) `smoker` (3) `region`. Подберите тему для графика. Назовите оси.
5. Объедините графики из заданий 3 и 4 в один так, чтобы сверху шёл один график из задания 3, а под ним 3 графика из задания 4. Сделайте общее название для графика.
6. Сделайте фасет графика из задания 3 по колонке `region`.
7. Постройте `scatter plot` отношения переменных `age` и `charges`. Добавьте названия осей, название графика и тему. Сделайте так, чтобы числа по оси X отображались 14 шрифтом.
8. Проведите линию тренда для предыдущего графика.

¹ Есть два способа сделать это: [первый](#) лёгкий и не совсем корректный (но результат будет правильным), второй сложнее, зато поможет вам понять, как выстроить весь цикл работы в репозитории (он описан [здесь](#) и [здесь](#)). Во втором случае общая идея в том, что вы создаёте и клонируете свой репозиторий, а потом настраиваете R, чтобы делать коммиты удобнее).

9. Сделайте разбивку предыдущего графика по колонке smokers (у вас должно получиться две линии тренда для курящих и нет).
10. Сделайте график из заданий 7-9, но вместо переменной age используйте переменную bmi.
11. Самостоятельно задайте вопрос №1 к данным (вопрос должен быть про какую-то подвыборку данных). Ответьте на него построив график на подвыборке данных. График должен содержать все основные элементы оформления (название, подписи осей, тему и проч.). Аргументируйте выбор типа графика.
12. Самостоятельно задайте вопрос №2 к данным (вопрос должен быть про какую-то подвыборку данных). Ответьте на него построив график на подвыборке данных. График должен содержать все основные элементы оформления (название, подписи осей, тему и проч.). Аргументируйте выбор типа графика.
13. Самостоятельно задайте вопрос №3 к данным (вопрос должен быть про какую-то подвыборку данных). Ответьте на него построив график на подвыборке данных. График должен содержать все основные элементы оформления (название, подписи осей, тему и проч.). Аргументируйте выбор типа графика.
14. (это задание засчитывается за два) Приблизительно повторите график:

