

Домашнее задание 2 по базовой визуализации данных в R

Информационная часть

Дата выдачи: 27.09.2025

Дата дедлайна: **19.10.2025** 23:59 по Мск (GMT +3) – при всех расхождениях верны

эти дата и время

Максимально возможный балл: 25

При всех расхождениях верной считается информация из данного pdf-файла.

В задании вы продолжите работать с датасетом hogwarts. Данные, заполненные qmd-файлы с лекций-семинаров, а также пустой qmd для выполнения домашнего задания можно найти в папке на гугл-диске.

Полученный qmd-файл и результат render в формате html загрузите на свой гитхаб. Папка, в которой лежат файлы с результатами домашнего задания на гитхабе должна называться basic_datavis_hw2_2025. Скрипт с домашним заданием будет называться BI_retarining_datavis_hw_2.qmd (архив с оным доступен для скачивания на гугл-диске и в гугл-классах). Все пути внутри проекта должны быть относительными. Папка data с данными должна лежать внутри папки, в которой находится ваш R-проект с домашним заданием (пути для чтения должны выглядеть таким образом: "data/kakoy_to_dataset.csv"). Чтобы сдать задание, приложите ссылку на гитхаб в гугл-классы.

```
basic_datavis_hw2_2025
├── BI_retarining_datavis_hw_2.qmd
├── data
│   ├── hogwarts_2025.csv
│   └── podskazka2.csv
├── images
│   └── task_2_4.png
└── Домашнее задание 2 по базовой визуализации данных в R.pdf
```

Пример структуры папки с домашним заданием.

Для этой домашней работы почти соблюдаются принципы классицизма: одно задание = не менее одного графика (и не менее одного чанка), если не указано иное. Это справедливо даже для ситуаций, когда следующее задание просит вас обратиться к предыдущему графику. В то же время, вы всегда можете расширить и дополнить свою интерпретацию или предложить дополнительные способы визуализировать то же самое другим, более подходящим, на ваш взгляд, способом, создав дополнительные графики для задания.

Часть заданий требуют самостоятельного поиска информации и использования пакетов, не упоминавшихся на лекции.

Если у вас возникают трудности технического, коммуникативного или иного характера – можете писать в чат курса в телеграме или в личные сообщения @asklepiad (Богдан) @JuliN_Di (Юлия).

Задания

Во всех заданиях, пожалуйста, выполняйте несколько обязательных правил, даже если они не упомянуты в тексте задания явно.

- Делайте отображаемый текст на рисунке настолько крупным, насколько это позволяет сделать график и ваше чувство меры. Читаемые надписи – важный (но не очень часто встречающийся) элемент графиков.
- Не используйте в работе фон ggplot2 по умолчанию.
- У каждого графика должен быть заголовок, отражающий его суть, корректно называющиеся оси.
- Старайтесь, чтобы графики отображались в достаточном крупном для читабельности размере не только в срендеренном html, но и в процессе интерактивного исполнения внутри qmd.

Для выполнения некоторых заданий вам потребуется пакет patchwork, который мы уже загрузили ранее.

Если при выполнении задания и интерпретации графика у вас возникает гипотеза, не стесняйтесь ее проверить – сделать еще один (или несколько) графиков для того же задания, провести дополнительные исследования и т.д.

Собственно задания

1. Изучите [сайт со списком расширений ggplot2](#).

Подберите ранее не использовавшийся в лекциях и домашних заданиях пакет, опции которого вам показались любопытными для визуализации данных hogwarts.

Используйте возможности данного пакета, проинтерпретируйте получившийся график и опишите, почему использованный инструментарий пакета показался вам интересным и полезным. (3 б.)

2. Изучите информацию о графике treemap. Как вы полагаете, в чем его отличие от мозаик-плота?

Исходя из структуры графика, предположите, в каких ситуациях лучше будет подходить treemap, а в каких -- мозаик-плот. Создайте treemap на данных hogwarts. Вы вольны использовать любые дополнительные пакеты в этом задании. (3 б.)

3. Постройте “леденцовый график” (lollipop-plot) для количества набранных студентами 5-го курса баллов за весь учебный год (по оси ординат – id студента, по оси абсцисс – итоговый балл).

Отсортируйте студентов в порядке убывания итогового балла. Раскрасьте точки на “леденцах” в зависимости от сердцевины волшебной палочки. Палочки с сердечной жилой дракона должны быть красного цвета, с пером феникса – желтого, с волосом единорога – серого.

В этом задании вы можете использовать только основные пакеты tidyverse. (2 б.)

4. Подробно (останавливаясь на каждом пункте, а не в общем) опишите, какие стороны графика и интерпретации (приводится ниже) вы считаете верными, и сделанными в соответствии с хорошими практиками, а какие -- нет.

Сделайте свой график, отвечающий на тот же исследовательский вопрос, но лишенный технических и интерпретационных недостатков оригинала (если вы таковые нашли и упомянули). (7.5 б)

График лежит в папке images под псевдонимом task_2_4.png:

Интерпретация графика:

*В ходе исследования, проведенного в ФГБОУССО “Хогвартс” (далее Хогвартс) для оценки успешности образовательного процесса и психоэмоционального состояния участников вышеназванного процесса, были выявлены тревожные закономерности. А именно – при оценке среднего балла, полученного в течение недели совокупностью учащихся Хогвартса была выявлена тенденция к резкому снижению данного показателя с течением времени. Так, в начале учебного процесса студенты набирали в среднем 2 балла (максимальный показатель). В последующие недели балл достоверно снижался, достигая минимальных значений к концу года (см. рис. 1). Мы полагаем, что данный процесс связан с ухудшением психологического состояния преподавателей означенного учебного заведения. Таким образом, в течение года они от мотивирующих мер положительного характера (добавление баллов за правильные ответы на занятия и успехи во внеучебной деятельности) переходят к репрессивным действиям (учащийся и увеличивающийся по модулю отъем баллов факультета за предполагаемые нарушения дисциплины, недостаточную, по мнению преподавателя, подготовку к занятию и пр.). Это вызывает снижение у учащихся мотивации к получению знаний, что вызывает следующий шаг снижения средненедельного балла – т.о., поведение педагогов вызывает своего рода *circulus vitiosus*.*

По итогам исследования и с учетом полученных результатов предлагается следующее: произвести замену профессорско-преподавательского состава с заменой педагогов, наиболее подверженных выгоранию в течение года (проявляющемуся в виде систематического снижения баллов студентам). Издать декрет о недопущении избыточного штрафования учащихся, начиная с 18-й недели учебного процесса до конца года. К нарушителям применять взыскания материально-магического характера. Объявить выговор с занесением в

личное дело текущим представителям руководящего состава Хогвартса (далее дирекция).

5. Рубрика "вредные советы". А теперь воспроизведите график из задания выше. Подсказка к реализации элемента, который мы не упоминали на лекции, находится в датасете `podskazka2.csv` -- сделайте из него скаттерплот, отрисовав переменную `X` по горизонтали, а `Y` -- по вертикали.

Уточнение – не обязательно точно попасть в конкретный цвет. Задание не предполагает угадывание оттенков “на глаз”. (7.5 б)

6. Напишите функцию, которая визуализирует распределение балла студентов за весь год посредством гистограммы со столбиками определенного цвета.

Сделайте конвеер, который создаст графики 36-ти заданных цветов (сохранены в переменную `res_colours` в чанке).

Соберите все графики на одно полотно при помощи функционала пакета `patchwork`.

Если возникли проблемы с `qmd`-файлом – вот необходимый код для начала выполнения задания 6.

```
` `` {r}
set.seed(2025)

colours_ <- colours()
res_colours <- colours_[colours_ |>
str_detect("grey|gray|black|white", negate = TRUE)]|>
  sample(size = 36)
` ``

(2 б.)
```