1 Report

Name : Alexander Unger, StudentID = 801223

- All the scripts can be found in the R/api folder

2 Task 2f

Considering the limitations of the API, where only 10,000 search results can be accessed, an alternative strategy to obtain all job postings from Baden-Württemberg would be through web scraping. This approach circumvents the API restrictions by directly interacting with the website itself. In order to implement this first a careful exploration and analysis of the website is needed. This has to be done in order to understand the structure and the naviagtion patterns of the website. Then using R packages such as httr could be used to handle HTTP requests. A script will be written to fetch the web page using URLs constructed based on observed navigation patterns. The script will parse the HTML content to extract necessary data such as job title, employer, job location, publication date, and description. Effective handling of pagination is crucial to iterate through all available job listings. Regular checks and updates on the scraping script will be necessary to adapt to any changes in the website's layout or functionality. So actually by employing R for web scraping one can effectively bypass the API's limitations and create a comprehensive dataset of job postings from Baden-Württemberg that exceed 10000 search results.
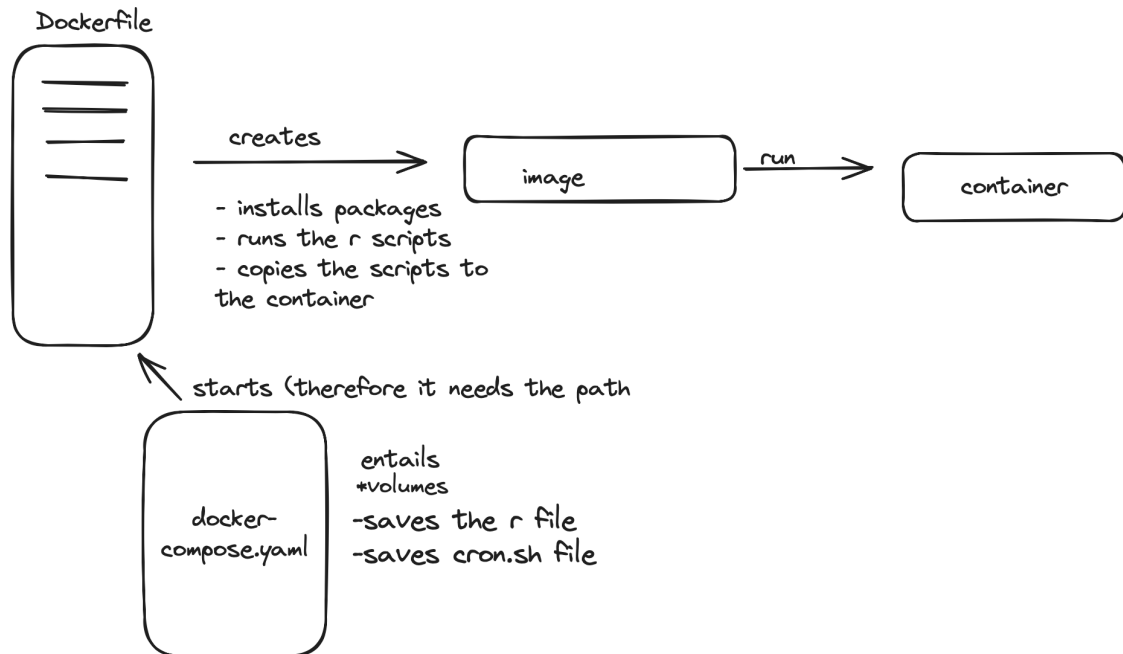
2.1 Task 3d

Figure 1: The function of the docker-compose.yaml

In our case docker-compose.yaml file is essential for creating and configuring containers from a Docker image. It requires a Dockerfile to specify how the image should be set up. In our project, the Dockerfile is responsible for installing necessary R libraries, copying project files, and defining the working directory for file execution. First, the Dockerfile prepares the Docker image by installing R libraries and copying the necessary files into the image. Next, the docker-compose.yaml specifies the path to the Dockerfile and the context for building the image, ensuring that the Dockerfile's instructions are executed. Once the image is built, the docker-compose.yaml uses this image to create a container. This container acts as a flexible and repeatable environment for running our R scripts. This setup is particularly effective for our needs, such as automatically fetching data from an API daily to update job listings.
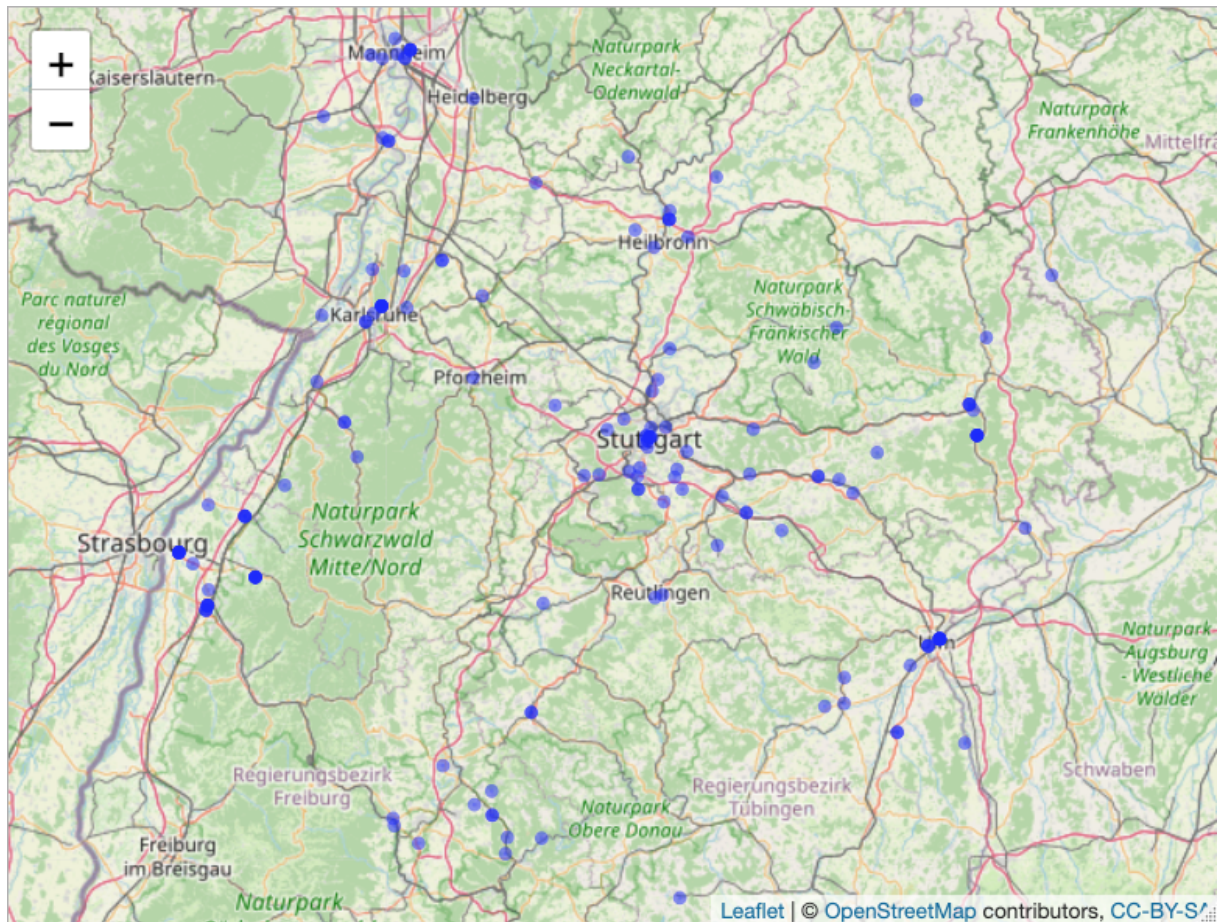
2.2 Task 4b

Figure 2: The active jobs for this week