**Пакет «Пользовательский интерфейс 1.0»**

Март 29, 2020

**Тип** Пакет

**Название** Пользовательский интерфейс 1.0

**Версия** 1.0.0.1

**Дата** 29.03.2020

**Авторы** Зубков Илья, Копырин Максим, Каримов Далер, Ефимчик Борис

**Поддержка** R (>= 3.5.0)

**Используемые сторонние пакеты** data.table (чтение данных), dplyr (общие расчеты), lubridate (преобразование даты), ggplot2 (построение графиков), ggpubr (построение статистичеких графиков), rstatix (статистические тесты), aTSA (статистические тесты).

**Загрузочные файлы** Quotes.R, Discriptive\_stat.R, Plots\_and\_export.R, Test\_normality.R, Test\_stationarity.R.

**Описание** Позволяет провести фильтрацию данных с биржи, сформировать необходимый тайм-фрейм, рассчитать доходности, вывести описательную статистику, визуализировать и экспортировать полученные данные, а также проверить их на нормальность и стационарность.

**Содержание**

1. Установка необходимого ПО для работы 1

2. Загрузка исходных данных 1

3. Возможности фильтрации данных (выбор тикера, тайм-фрейма, временного интервала) 2

3.1 Функция quotes 2

4. Формирование дополнительных рядов данных (open-, low-, high-, close-цены, цепная, накопленная доходности) 3

5. Описательная статистика 3

6. Графическая часть и экспорт данных 4

6.1 Функция Plot\_and\_export 4

7. Тест на нормальность 5

7.1 Функция Normality\_test 5

8. Тест на стационарность 6

8.1 Функция Stationarity\_test 6

1. Установка необходимого ПО для работы

*Установка R и RStudio*

Подробно об установке Rstudio можно узнать по следующим ссылкам:

Загрузите последнюю версию языка R по ссылке: <https://cran.r-project.org/bin/windows/base/>

Загрузите RStudio по ссылке: <https://rstudio.com/products/rstudio/>

Воспользуйтесь инструкцией по установке R и RStudio: <https://bdemeshev.github.io/installation/r/R_installation.html>

1. Загрузка исходных данных

Открыть файл Quotes.R в R.Studio, загрузить и установить необходимые для работы дополнительные пакеты (data.table, dplyr, lubridate, ggplot2).

Для того, чтобы воспользоваться программой, необходим файл формата TradeLog с колонками TRADENO, SECCODE, TIME, BUYORDERNO, SELLORDERNO, PRICE, VOLUME в формате txt.

TRADENO – идентификационный номер записи (заявки).

SECODE – тикер компании.

TIME – время записи.

BUYORDERNO, SELLORDERNO – номер заявки на покупку или продажу.

PRICE – цена лота.

VOLUME – объем заявки.

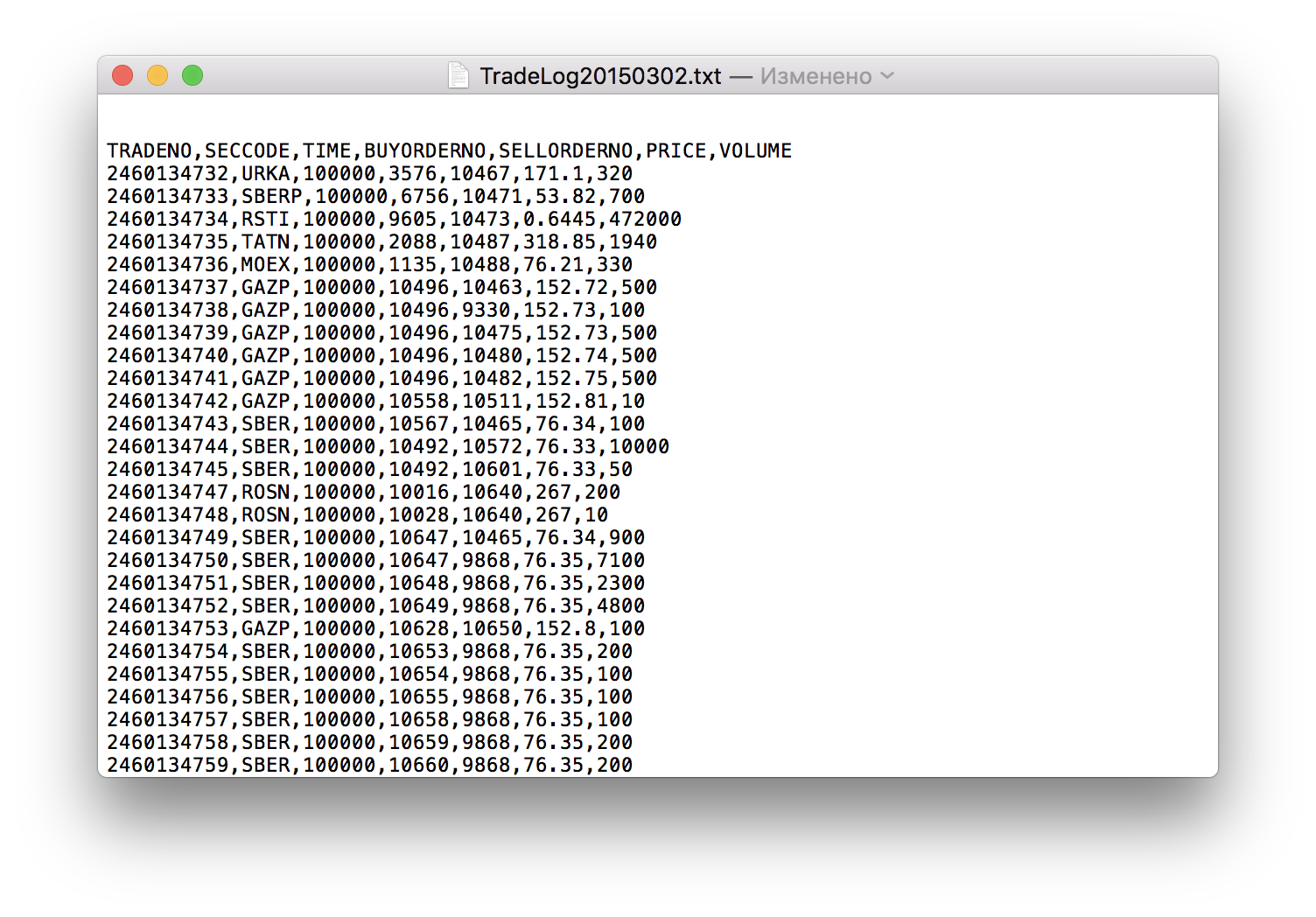


Рис.1. Пример текстового файла для работы с пакетом «Пользовательский интерфейс 1.0».

Затем необходимо поместить текстовый файл в рабочую директорию. Адрес рабочей директории можно узнать с помощью оператора *getwd()*, а установить новый путь к ней можно во вкладке «Session – Set Working Directory – Choose Directory)».

1. Возможности фильтрации данных (выбор тикера, тайм-фрейма, временного интервала)

Загрузить файл с исходными данными в R.Studio с помощью функции *fread*.

*Пример:*

OrderLog <- fread( ' OrderLog20150302.txt ' )

Необходимо определиться с тикером компании, тайм-фреймом и временным интервалом и ввести данные параметры в аргументы функции *quotes*.

Полный список доступных в исходном файле тикеров можно получить с помощью следующей функции:

Ticker\_list <- as.data.frame(unique(OrderLog$SECCODE))

Этот список всегда будет доступен в разделе Environment в правом верхнем окне.

3.1 Функция quotes

Функция *quotes* позволяет провести фильтрацию данных: выбрать компанию, временной интервал, тайм-фрейм.

**data** – дата-фрейм с исходными данными.

**company\_ticker** – тикер компании (SECCODE).

**time\_interval** – тайм-фрейм – деление временной шкалы на отрезки; укажите значение от 1 до 60 минут.

**start** – начало временного интервала для фильтрации данных; к примеру «2017-01-01 10:00».

**finish** – окончание временного интервала для фильтрации данных; к примеру «2017-12-31 16:00».

*Пример:*

Quotes\_result <- quotes(data = OrderLog, company\_ticker = ' SBER ', time\_interval = 10, start = ' 2015-03-02 10:00 ', finish = ' 2015-03-02 15:00 ' )

Посмотреть тикеры компаний можно, выполнив следующий запрос:

1. Формирование дополнительных рядов данных (open-, low-, high-, close-цены, цепная, накопленная доходности)

В ходе исполнения скрипта будут рассчитаны Open-, Low-, High-, Close-цены, объемы (volume), а также цепная (return) и куммулятивная доходности (cumulative\_yield).

Open – цена открытия.

Close – цена закрытия.

High – максимальное значение цены за период.

Low – минимальное значение цены за период.

Доходность – разница между текущей и предшествующей ценой.

Куммулятивная доходность – суммирование доходностей за все периоды.

Данные будут записаны в новый дата-фрейм *Quotes\_result* (или любое другое имя).

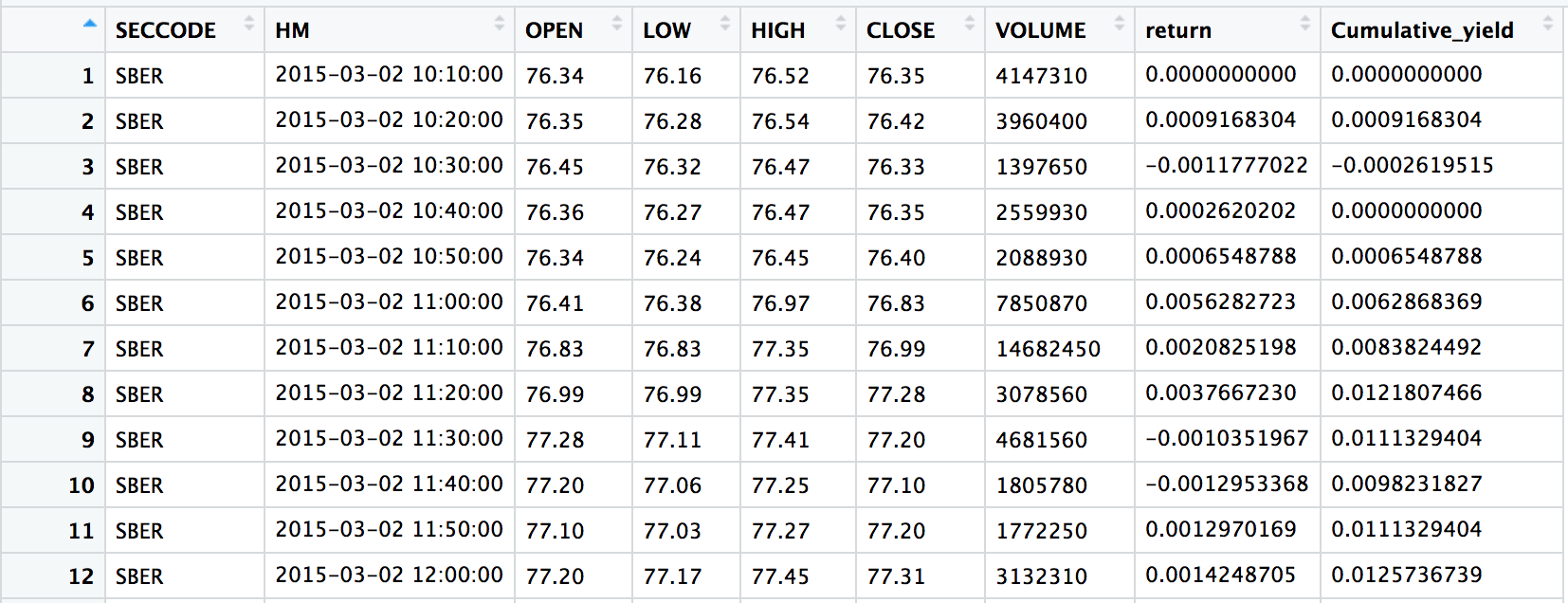


Рис.2. Дата-фрейм Quotes\_result.

1. Описательная статистика

Для получения описательной статистики данных, сохраненных в новом дата-фрейме *Quotes\_result*, используйте файл Discriptive\_stat.R.

Загрузите функцию *Descr\_stat*, выбрав данные для расчета. В нашем случае, это созданный ранее дата-фрейм *Quotes\_result.*

*Пример*

Descr\_stat(Quotes\_result)

Будут рассчитаны стандартное отклонение, среднее, медианное, максимальное и минимальное значение.

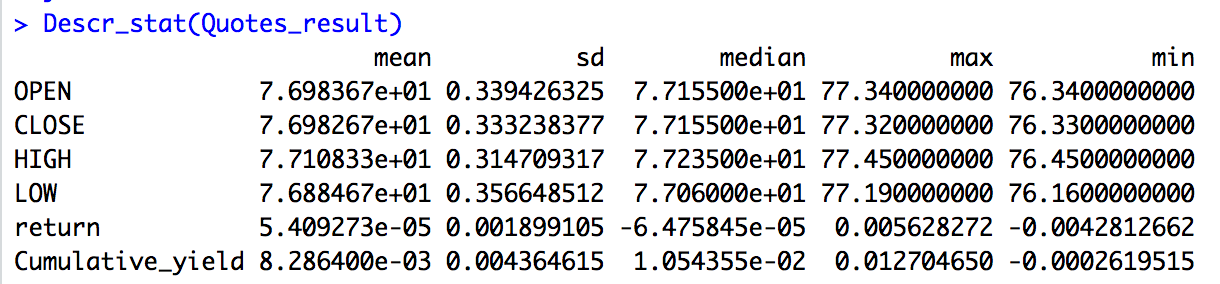


Рис.3. Вывод описательной статистики.

1. Графическая часть и экспорт данных

Для построения графика и экспорта данных откройте файл Plots\_and\_export.R, загрузите функцию *Plot\_and\_export*, указав необходимые аргументы функции.

6.1 Функция Plot\_and\_export

Функция *Plot\_and\_export* позволяет экспортировать данные в формате csv или xlsx, а также представлять данные в виде графика.

**data** – дата-фрейм с данными для расчета.

**line chart** – укажите T (TRUE) в случае необходимости графического представления данных, либо F (FALSE) в противоположном случае.

Если необходимо сохранить данные в формате **.csv**, укажите в аргументах функции «**export\_csv = T, export\_xlsx = F**».

Если необходимо сохранить данные в формате **.xlsx**, укажите в аргументах функции «**export\_csv = F, export\_xlsx = T**».

*Пример*

Plot\_and\_export(data = Quotes\_result, line\_chart = T, export\_csv = F, export\_xlsx = T)

В результате в рабочей директории будет создан файл Quotes.xlsx, в котором будут сохранены данные из дата-фрейма *Quotes\_result*, а также построен график.

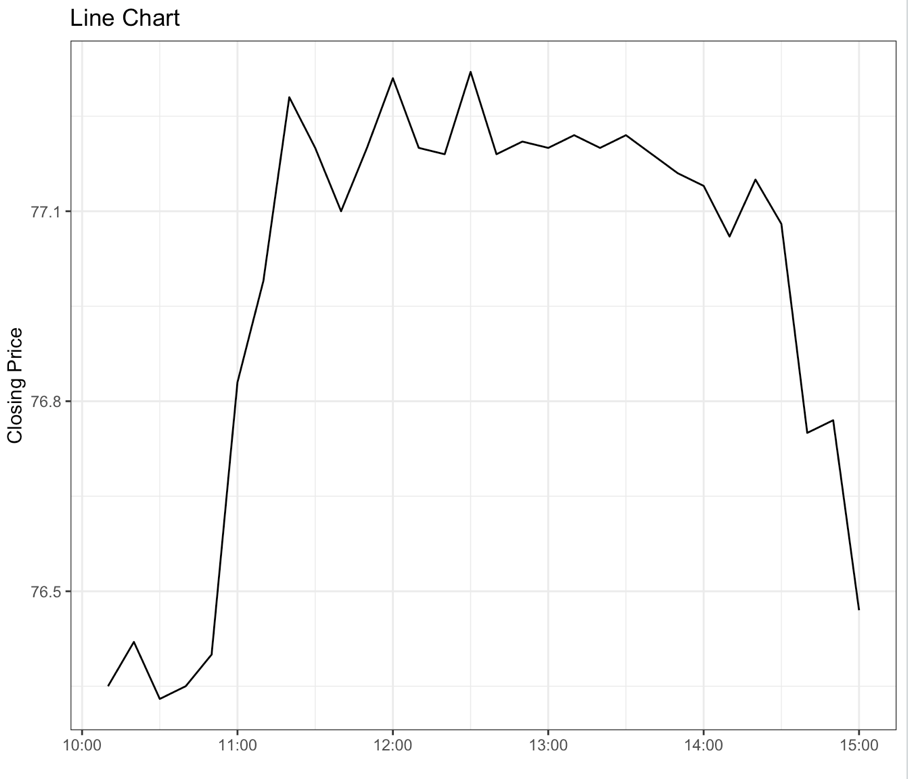


Рис.4. Представление данных Quotes\_result в графическом виде.

1. Тест на нормальность

Для проверки полученных данных на нормальность необходимо загрузить файл Test\_normality.R и установить дополнительные пакеты (ggpubr и rstatix).

Затем использовать функцию *Normality\_test*, установив необходимые аргументы.

7.1 Функция Normality\_test

Функция *Normality\_test* позволяет провести тест на нормальность данных.

**data** – дата-фрейм с данными для тестирования.

**sign\_level** – доверительный интервал; укажите значения от 0 до 1.

Если необходимо построить график квантиль-квантиль (q-q), то укажите «**qqplot = T, dplot = F**».

Если же необходимо построить график плотностного распределения, то укажите «**qqplot = F, dplot = T**».

Если необходимо отобразить на экране результаты теста, введите «**print\_statiscics = T**», в противном случае «**print\_statiscics = F**».

**qqplot** – график квантиль-квантиль (q-q) - графический метод определения того, исходят ли два набора данных из совокупностей с общим распределением.

**dplot** – плотностный график - графическое представление распределения числовой переменной.

Для проверки данных на нормальность используется [критерий Шапиро-Уилка](http://www.machinelearning.ru/wiki/index.php?title=Критерий_Шапиро-Уилка).

*Примеры*

1. Normality\_test(data = Quotes\_result, sign\_level = 0.05, qqplot = T, dplot = F, print\_statiscics = T)
2. Normality\_test(data = Quotes\_result, sign\_level = 0.05, qqplot = F, dplot = T, print\_statiscics = T)

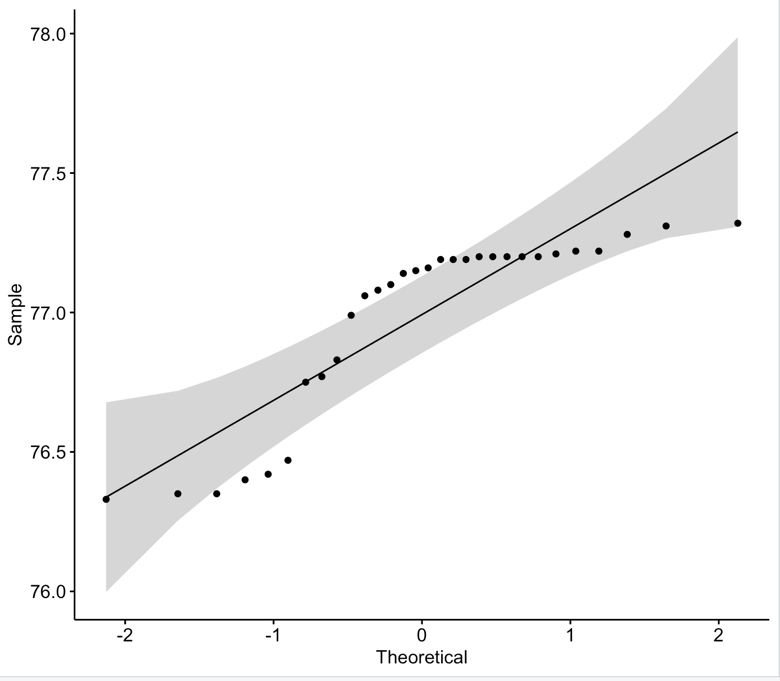


Рис.5. График квантиль-квантиль (qqplot) для анализируемого массива данных.

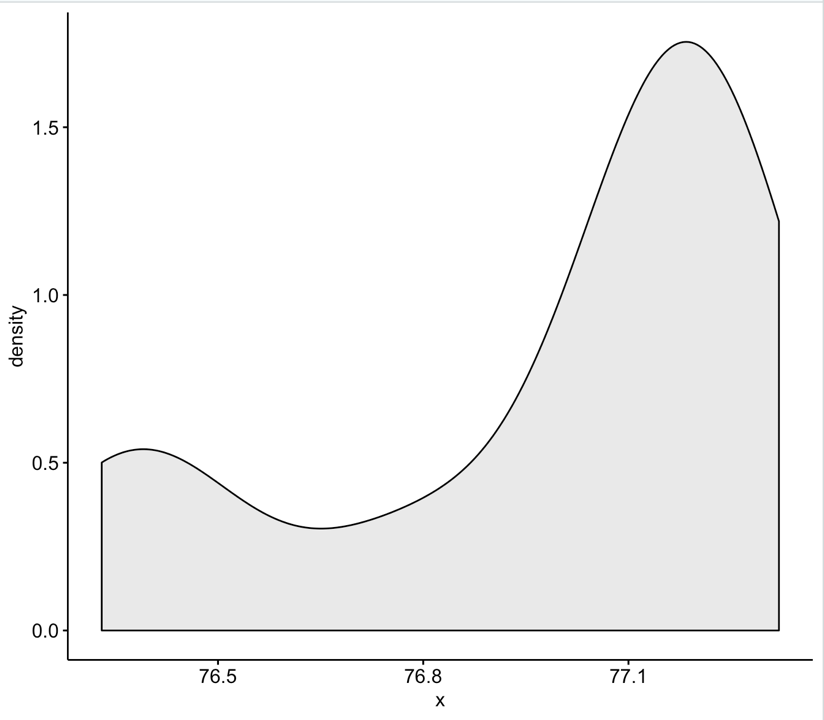


Рис.6. Плотностный график (dplot) для анализируемого массива данных.

1. Тест на стационарность

Для проверки полученных данных на стационарность необходимо загрузить файл Test\_stationarity.R и установить дополнительный пакет aTSA.

Затем использовать функцию *Stationarity\_test*, установив необходимые аргументы.

8.1 Функция Stationarity\_test

Функция *Stationarity\_test* позволяет проверить данные на стационарность.

**data** – дата-фрейм с данными для тестирования.

**sign\_level** – доверительный интервал; укажите значения от 0 до 1.

**lag** – временной лаг; укажите значение от 1 и более.

Если необходимо отобразить на экране результаты теста, введите «**print\_statiscics = T**», в противном случае «**print\_statiscics = F**».

Для проверки данных на стационарность используется [метод Дики-Фуллера](https://ru.wikipedia.org/wiki/Тест_Дики_—_Фуллера).

*Пример*

Stationarity\_test(data = Quotes\_result, sign\_level = 0.05, lag = 1, print\_statiscics = T)