

# Языки для анализа данных

## Сравнительный анализ python и R

Драпак Степан

ВМК МГУ. Кафедра ММП

Курс «Практикум на ЭВМ» для 317-ой группы



- 1 История
- 2 Плюсы и минусы R и Python
- 3 Выводы




- R
  - Создан в 1995 году в Оклендском университете
  - Создавался как альтернатива языку S
  - Изначально проектировался как язык для статистического анализа данных
  - Большая часть стандартной реализации написана на C, Fortran и R
- Python
  - Создан в 1991 сотрудником голландского института CWI
  - Проектировался как скриптовый язык для распределенной ОС Amoeba
  - Основная реализация CPython написана на C

- + Отличные возможности для визуализации. Пакеты `ggplot2`, `ggvis`
- + Прекрасная IDE RStudio. Есть возможность сразу, при загрузке данных, посмотреть на них в табличном виде, построить графики, импортировать все это в PDF и многое другое
- + Большое количество современных, развивающихся пакетов и активное сообщество. В рамках проекта Rexer Analytics 2013 Data Miner Survey 70% участников использовали R
- + Язык писался людьми изучающими статистику, для людей, изучающих статистику. Он интуитивно понятен для тех, кто погружен в эту тему. Помимо этого он расширяется на неакадемическую среду
- Язык достаточно медленный. Однако появляется все больше пакетов, повышающих его производительность: `rqr`, `renjin`...
- Язык достаточно сложен для изучения

- + Универсальный язык. Он объединяет людей из очень разных сфер. Дает возможности написать многи вещи, затратив при этом не очень много времени
- + Прост в изучение. Порог вхождения очень низок. Не требует никаких специальных знаний
- + IPython Notebook - очень удобный инструмент для командной(и не только) работы. Позволяет писать документированный код за достаточно сжатое время
- Уступает R в простоте и наглядности визуализации
- Поскольку Python не заточен под анализ данных, есть достаточно много важных пакетов в R, альтернативы которым нет в Python

- Наилучшим вариантом было бы знание обоих языков, по крайней мере, одного из них на хорошем уровне, а другого - на достаточном для решения задач, которые значительно проще решаются на нем
- При выборе основного языка полезно подумать, какие инструменты вам будут требоваться чаще всего и какие для них есть альтернативы

-  Курс по языку R на coursera(в рамках специализации Data Science)  
<https://www.coursera.org/course/rprog>
-  Официальный сайт <https://cran.r-project.org/>

-  Martijn Theuwissen, R vs Python for Data Science <http://www.kdnuggets.com/2015/05/r-vs-python-data-science.html>
-  wikipedia R(язык программирования) [https://ru.wikipedia.org/wiki/R\\_\(programming\\_language\)](https://ru.wikipedia.org/wiki/R_(programming_language))
-  wikipedia Python(Язык программирования) [https://en.wikipedia.org/wiki/Python\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/Python_(programming_language))



Спасибо за внимание

Вопросы?