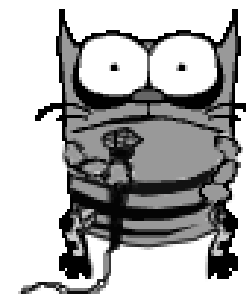


Анализ данных без питона

The ight Way

Александр Горбач

alex@gorbach.me



О чём мы вообще?

- ▮ Статистика – это важно
- ▮ Data Science – это круто
- ▮ Много вариантов – это жизнь
- ▮ Питон – это не наше всё
- ▮ Наше всё – это Пушкин



Что такое это R?

- ▮ Язык программирования
- ▮ Среда вычислений
- ▮ Опенсурс проект



Почему “R”?

- ▮ Основан на языке S
- ▮ Авторы Росс Айхэкой (***R**oss Ihaka*) & Роберт Джентлмен (***R**obert Gentleman*)



В чём соль?

- ▮ Рождён быть Open Source
- ▮ Создан математиками (факультет статистики Оклендского университета, Новая Зеландия)
- ▮ Для математиков
- ▮ Язык вдохновитель S создавался для обработки данных
- ▮ Поддерживается математиками



Как он работает?

- ▮ Интерактивный ввод
- ▮ Базовый тип – вектор
- ▮ Безумный оператор присваивания "<-"
- ▮ Все операции умеют понимать вектора
- ▮ Сила в Data Frame



Примеры: получение данных

- ▮ `read.csv("super_data.csv")`
- ▮ `read.xlsx("super_data.xlsx")`
- ▮ `download.file(url, "local_file.dat")`
- ▮ `socket <- url("http://something.website")`
- ▮ Базы данных



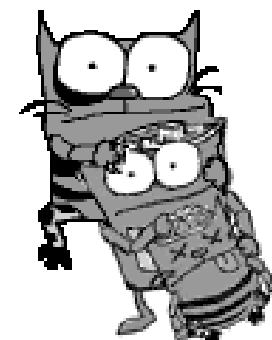
Примеры: выборки данных

- ❑ `cran %>%`
- ❑ `select(ip_id, country, package, size) %>%`
- ❑ `mutate(size_mb = size / 2^20) %>%`
- ❑ `filter(size_mb <= 0.5) %>%`
- ❑ `arrange(desc(size_mb))`



Примеры: трансформации

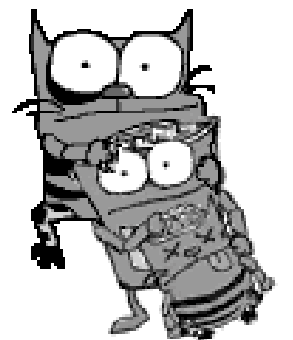
```
> students
  grade male female
1     A     1      5
2     B     5      0
3     C     5      2
4     D     5      5
5     E     7      4
```



Примеры: трансформации

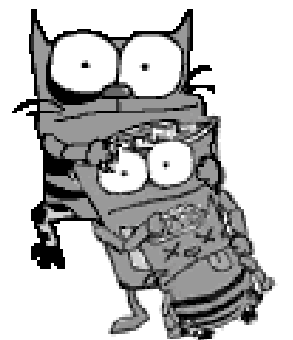
```
> gather(students, sex, count, -grade)
```

	grade	sex	count
1	A	male	1
2	B	male	5
3	C	male	5
4	D	male	5
5	E	male	7
6	A	female	5
7	B	female	0
8	C	female	2
9	D	female	5
10	E	female	4



Примеры: трансформации

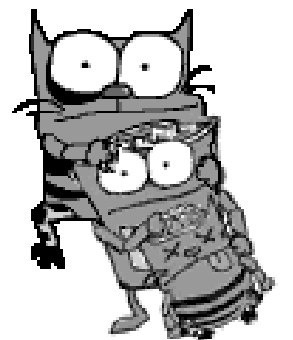
```
> students2
  grade male_1 female_1 male_2 female_2
1     A      3      4      3      4
2     B      6      4      3      5
3     C      7      4      3      8
4     D      4      0      8      1
5     E      1      1      2      7
```



Примеры: трансформации

```
> res <- gather(students2, sex_class, count, -grade)
```

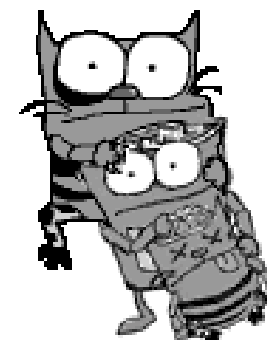
```
> res
  grade sex_class count
1     A   male_1     3
2     B   male_1     6
3     C   male_1     7
4     D   male_1     4
5     E   male_1     1
6     A   female_1    4
7     B   female_1    4
8     C   female_1    4
9     D   female_1    0
10    E   female_1    1
11    A   male_2     3
12    B   male_2     3
13    C   male_2     3
14    D   male_2     8
15    E   male_2     2
16    A   female_2    4
17    B   female_2    5
18    C   female_2    8
19    D   female_2    1
20    E   female_2    7
```



Примеры: трансформации

```
> separate(data = res, col = sex_class, into = c("sex", "class"))
```

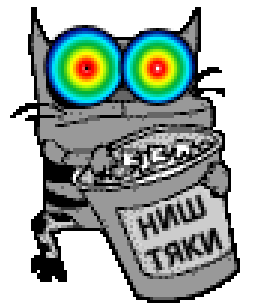
	grade	sex	class	count
1	A	male	1	3
2	B	male	1	6
3	C	male	1	7
4	D	male	1	4
5	E	male	1	1
6	A	female	1	4
7	B	female	1	4
8	C	female	1	4
9	D	female	1	0
10	E	female	1	1
11	A	male	2	3
12	B	male	2	3
13	C	male	2	3
14	D	male	2	8
15	E	male	2	2
16	A	female	2	4
17	B	female	2	5
18	C	female	2	8
19	D	female	2	1
20	E	female	2	7



Тонны
пакетов для
всего

CRAN

The Comprehensive R Archive Network



Вопросы?

