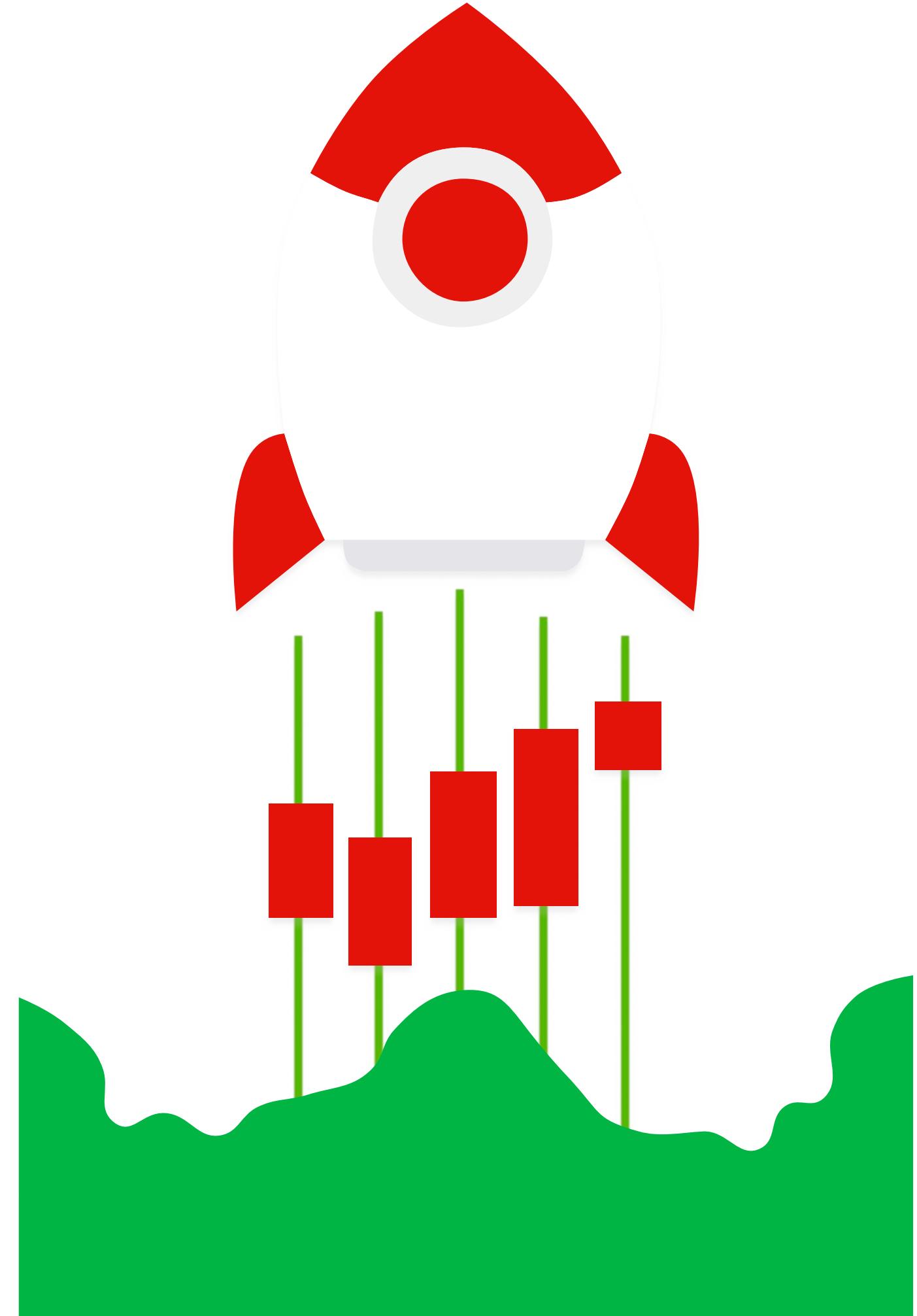


# Все идет по плану

Гайнетдинов Ян  
Шаклеин Владимир  
Редлих Алексей  
Бушин Андрей

TO  
THE  
MOON



## Структура данных

Данные из собственной системы  
управления растениеводством  
**ГК ЭкоНива - EkoCrop**

**14**

**Регионов**

От Алтая до СПб

**28**

**Месяцев**

С июля 22 по октябрь 24

**329**

**Тысяч строк**

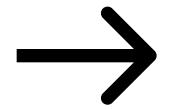
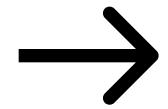
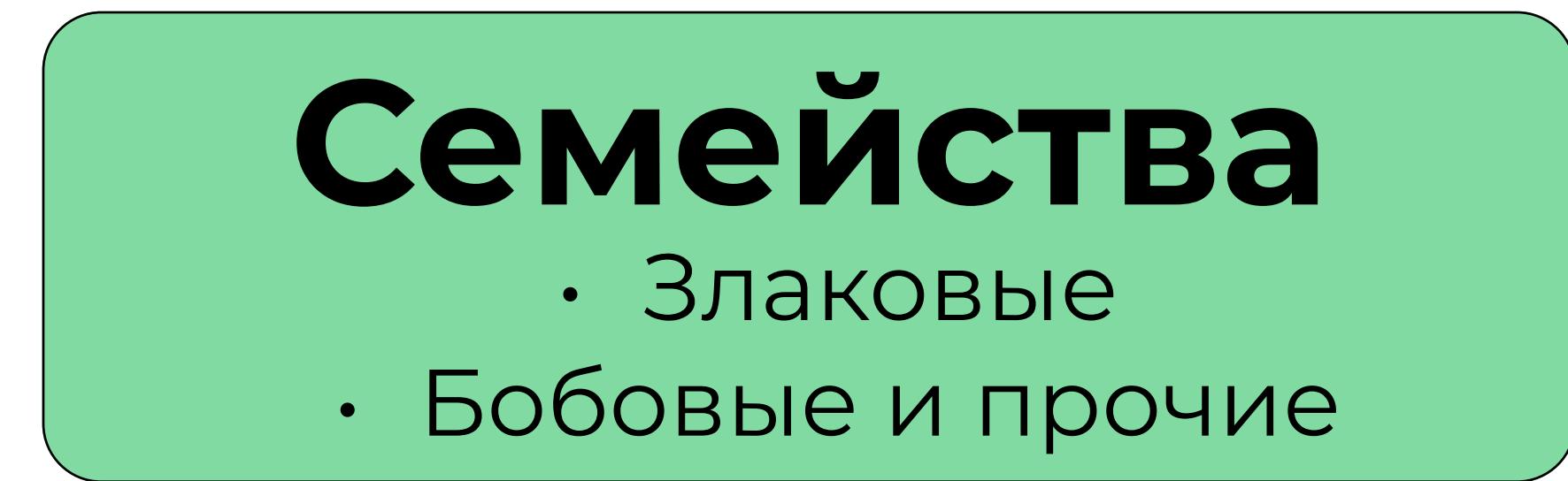
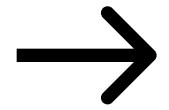
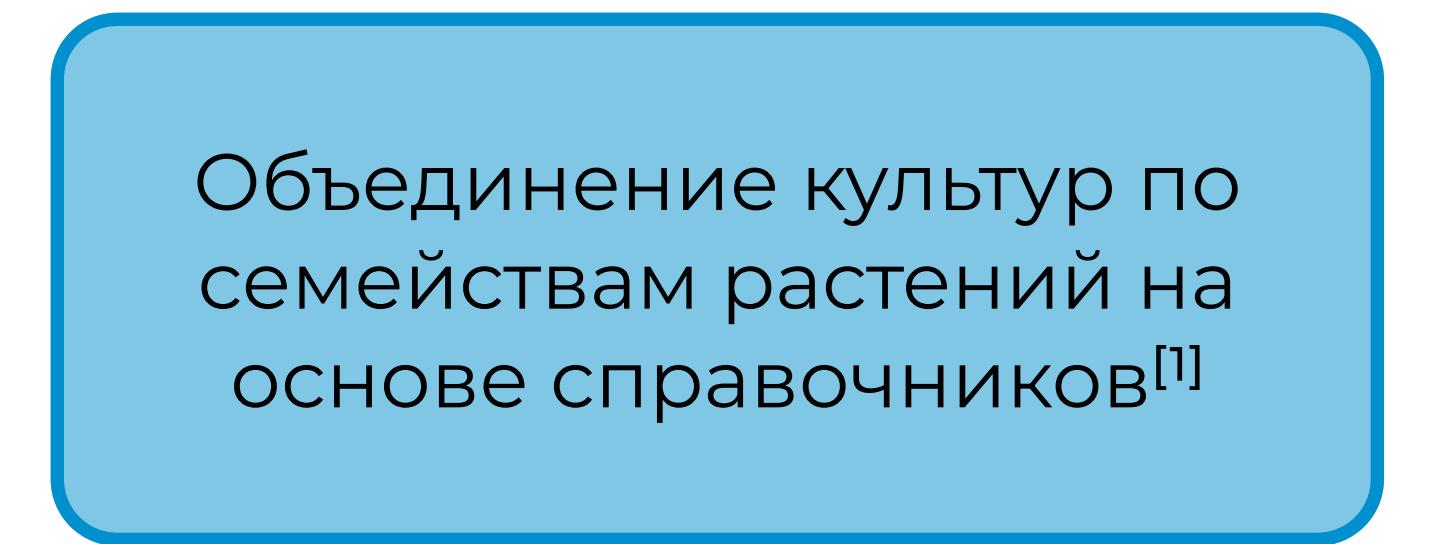
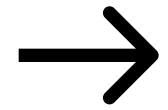
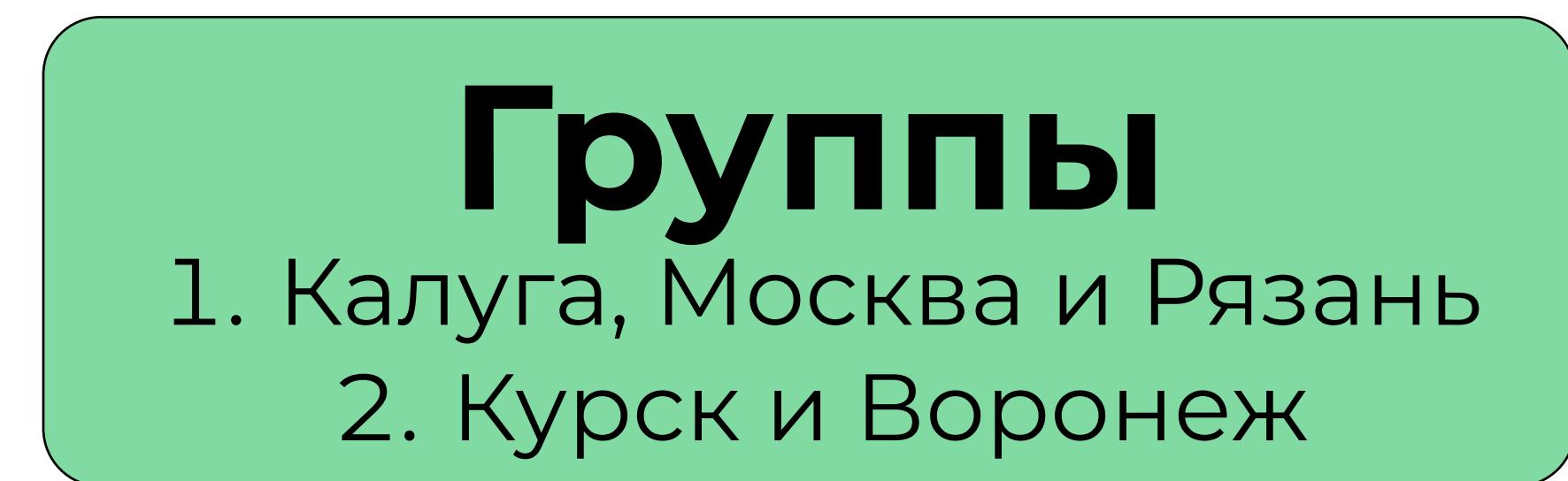
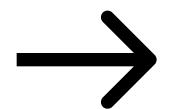
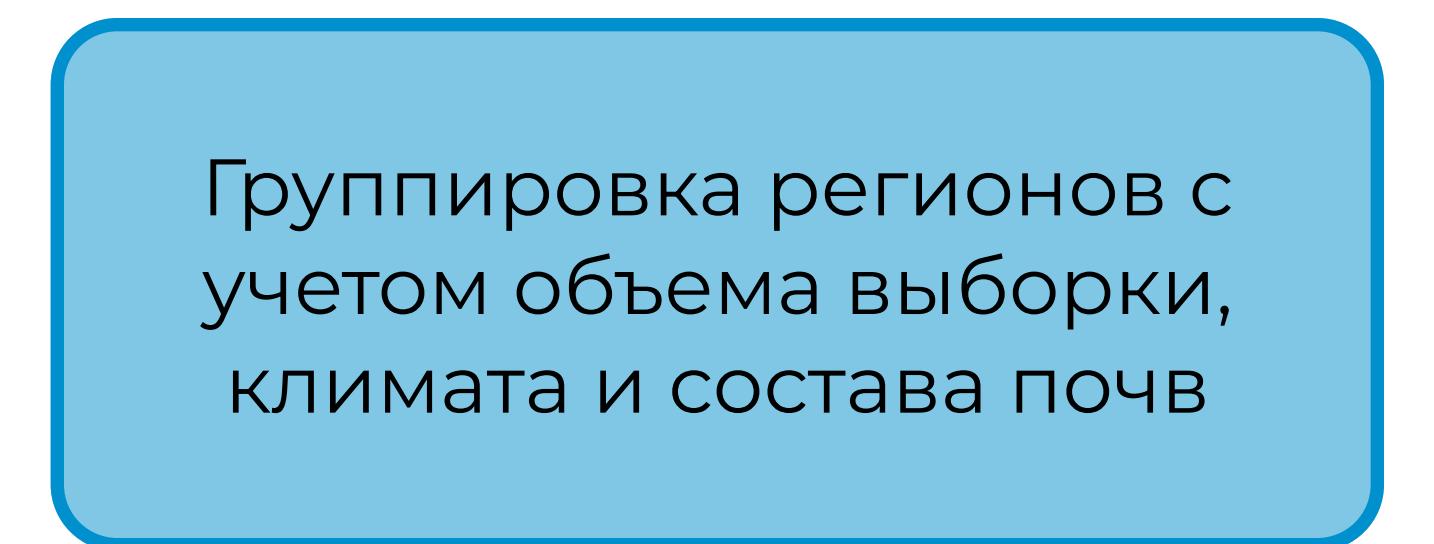
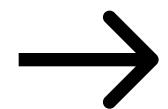
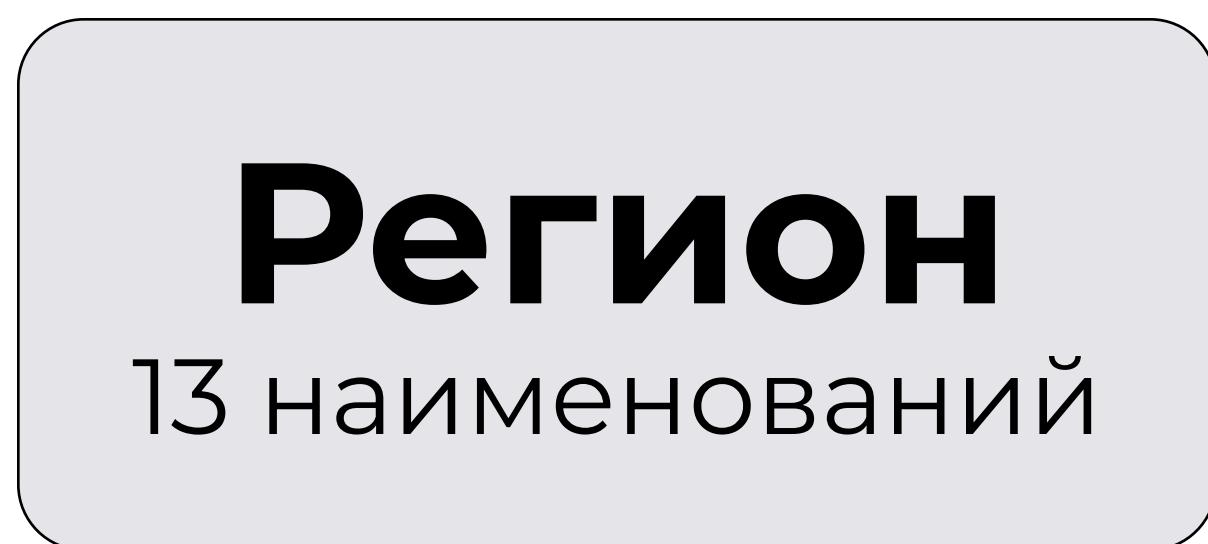
О полях, работах и ТМЦ



# Структура данных

	Информация о полях	Информация о работах	Информация о ТМЦ
Качественные переменные	<ul style="list-style-type: none"> <li>• Регион</li> <li>• Вид с/х угодия</li> <li>• Текущая культура</li> <li>• Культура предшественник</li> <li>• Сорт / гибрид</li> <li>• Происхождение семян</li> <li>• Технология выращивания</li> </ul>	<ul style="list-style-type: none"> <li>• Тип работы</li> <li>• Подтип работы</li> <li>• Основная техника</li> <li>• Прицепная техника</li> <li>• Механизатор</li> <li>• Назначение готовой продукции</li> </ul>	<ul style="list-style-type: none"> <li>• Тип внесенного на поле ТМЦ</li> <li>• Название внесенного на поле ТМЦ</li> <li>• Единица измерения ТМЦ</li> </ul>
Количественные переменные	<ul style="list-style-type: none"> <li>• Балл плодородия</li> <li>• Сезон</li> <li>• Площадь</li> <li>• Урожайность плановая</li> <li>• Содержание фосфора</li> <li>• Содержания калия</li> <li>• Гумус</li> <li>• Кислотность почвы</li> </ul>	<ul style="list-style-type: none"> <li>• Сезон</li> <li>• Площадь выполненной работы</li> <li>• Глубина обработки</li> <li>• Вес убранной продукции</li> <li>• Содержание сухого вещества</li> <li>• Сорная примесь</li> <li>• Урожайность фактическая</li> </ul>	<ul style="list-style-type: none"> <li>• Вес внесенного на поле ТМЦ</li> <li>• Средняя доза внесения ТМЦ</li> <li>• Дата внесения ТМЦ на поле</li> <li>• ТМЦ - товарно-материальные ценности</li> </ul>

## Подготовка базы данных

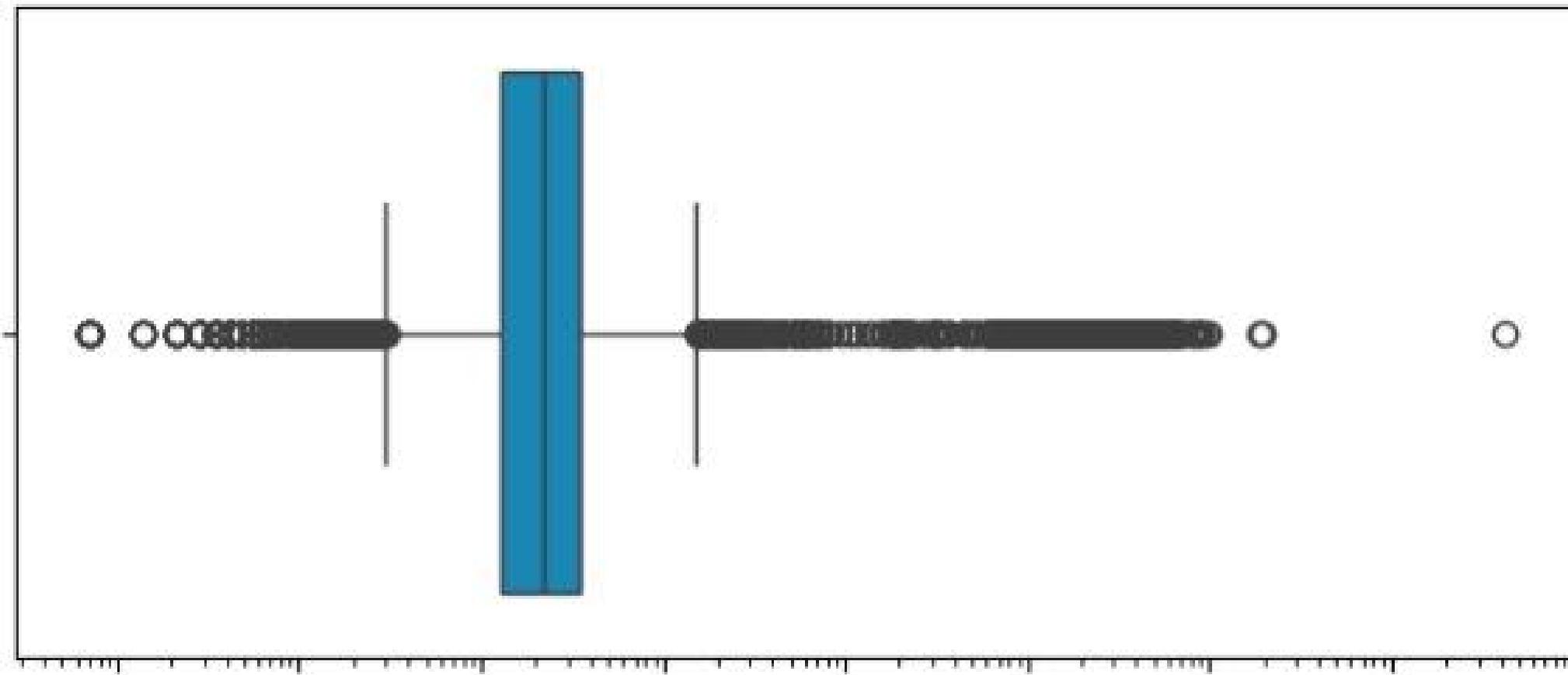


1. Каталог основных сельскохозяйственных культур

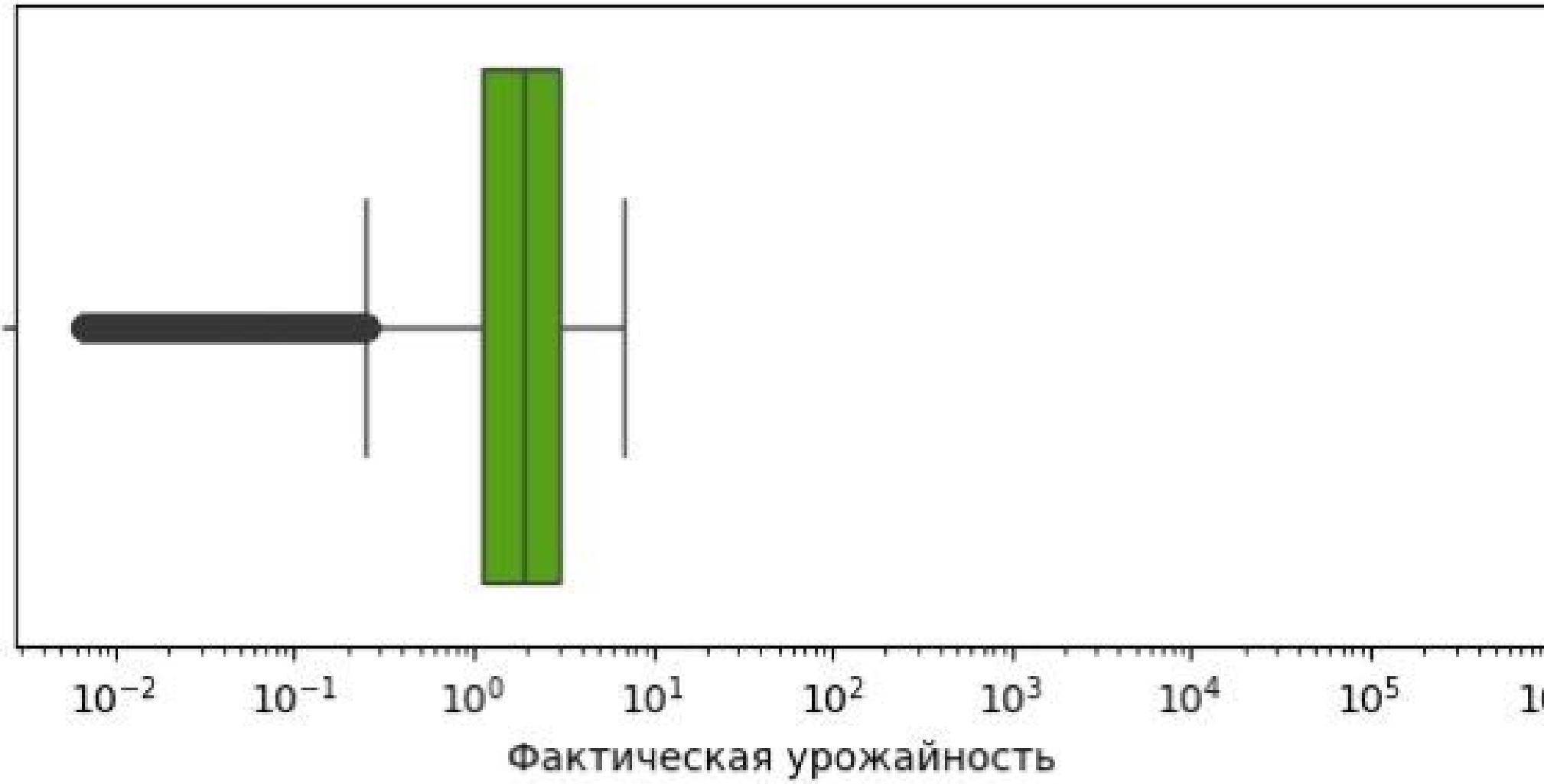
2. Способы уборки зерновых культур

## Подготовка базы данных: фактическая урожайность

До нормализации



После нормализации

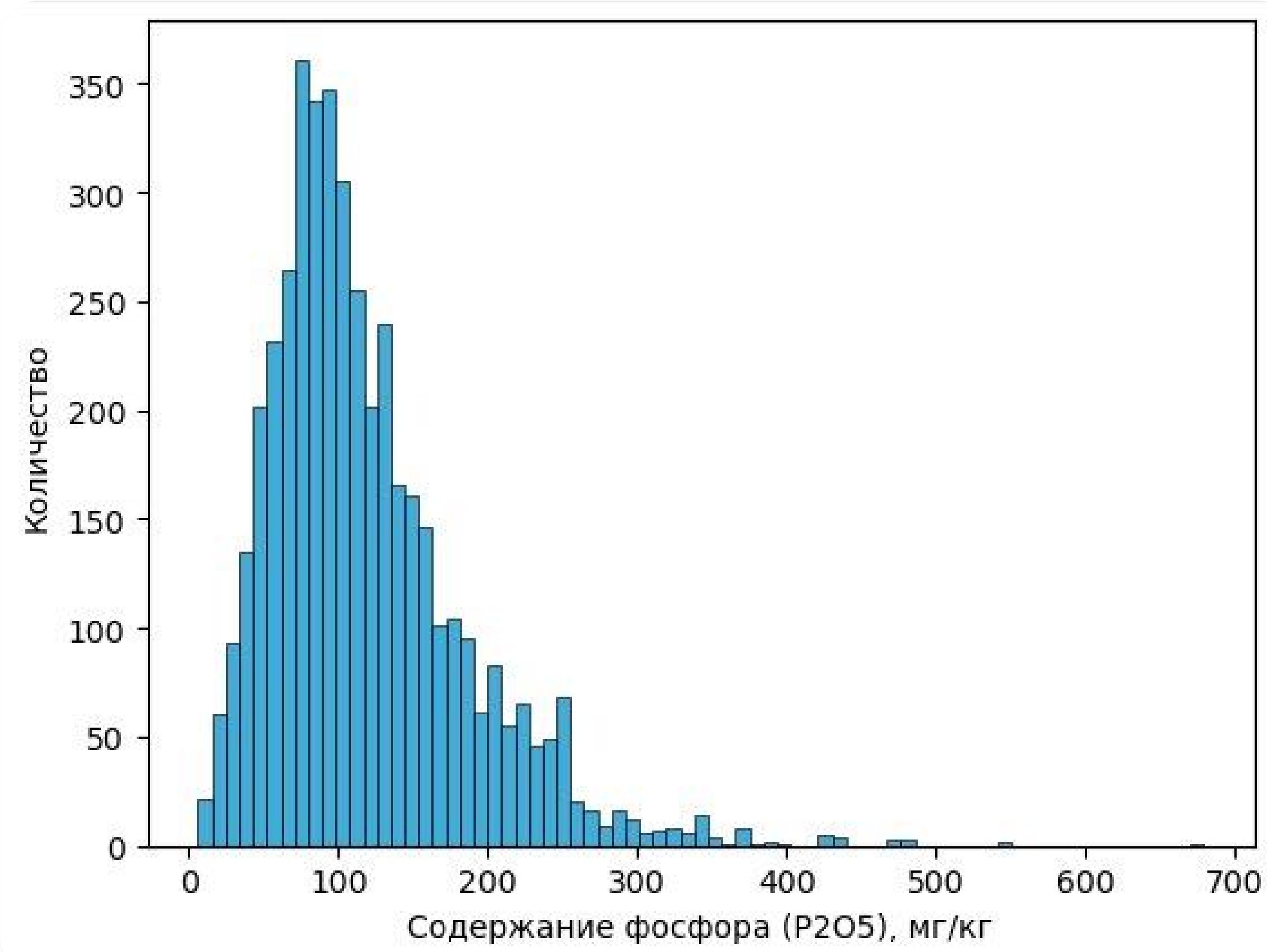
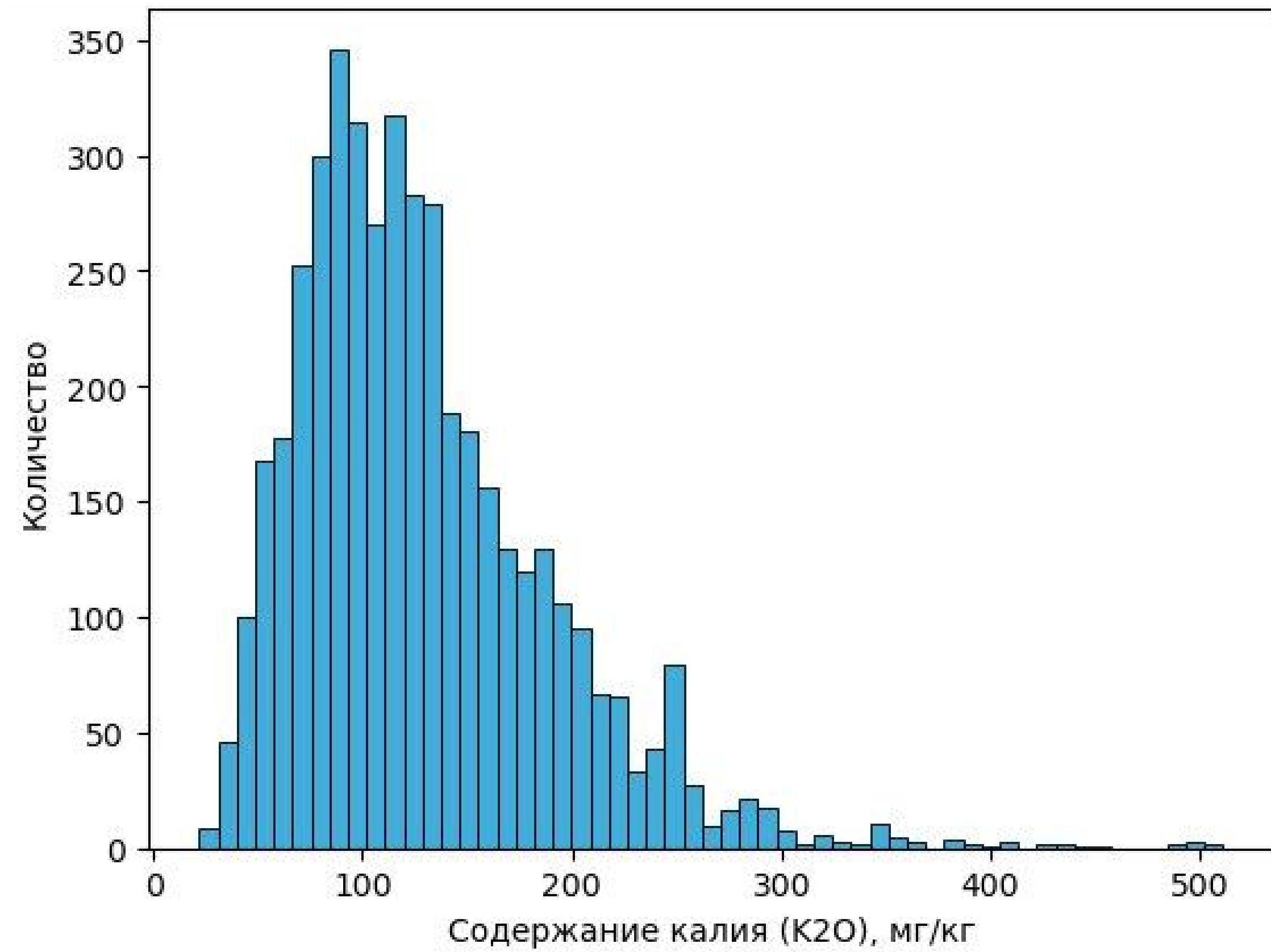


**Выбросы не в  
размерности<sup>[3]</sup>**

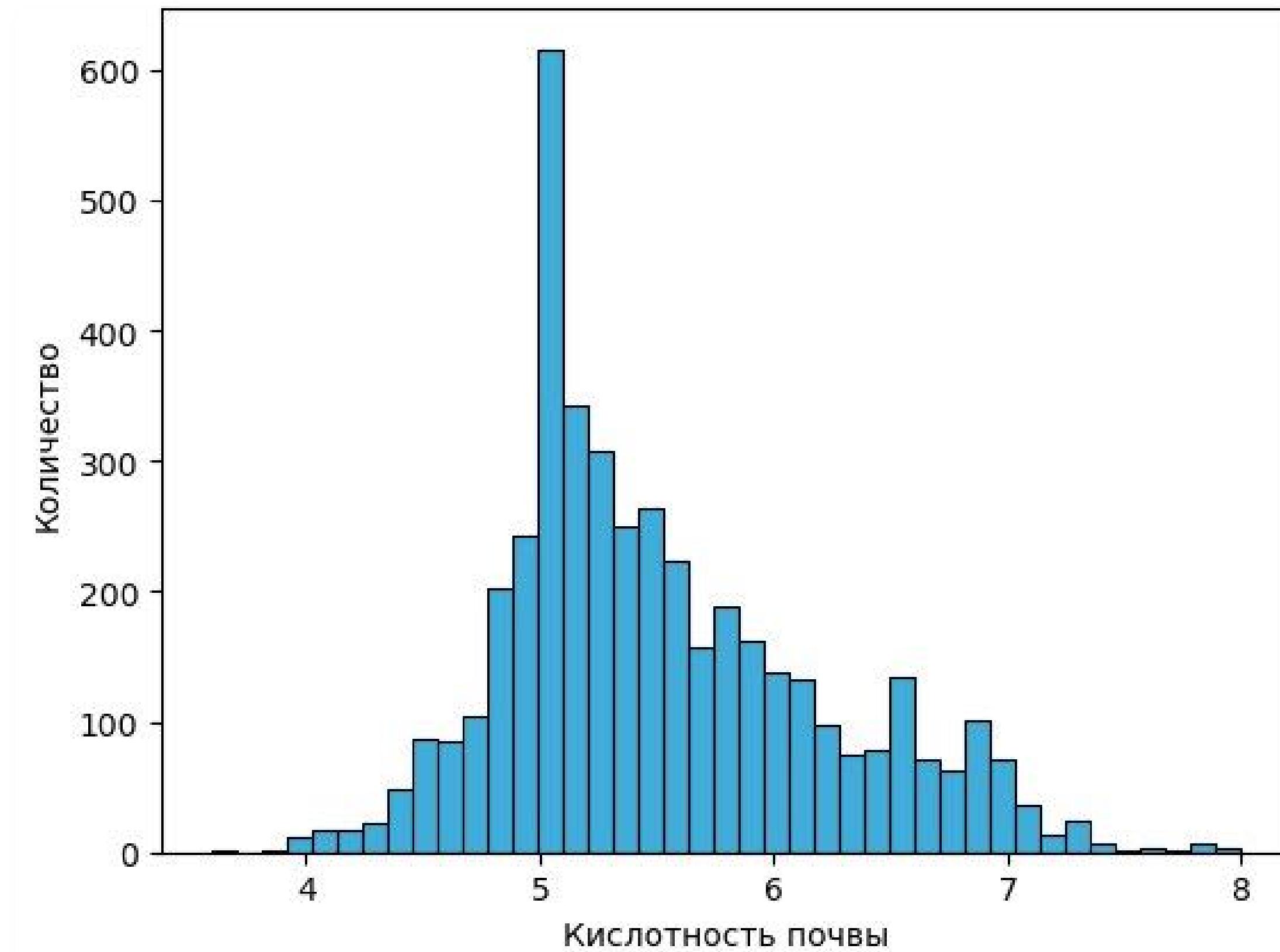
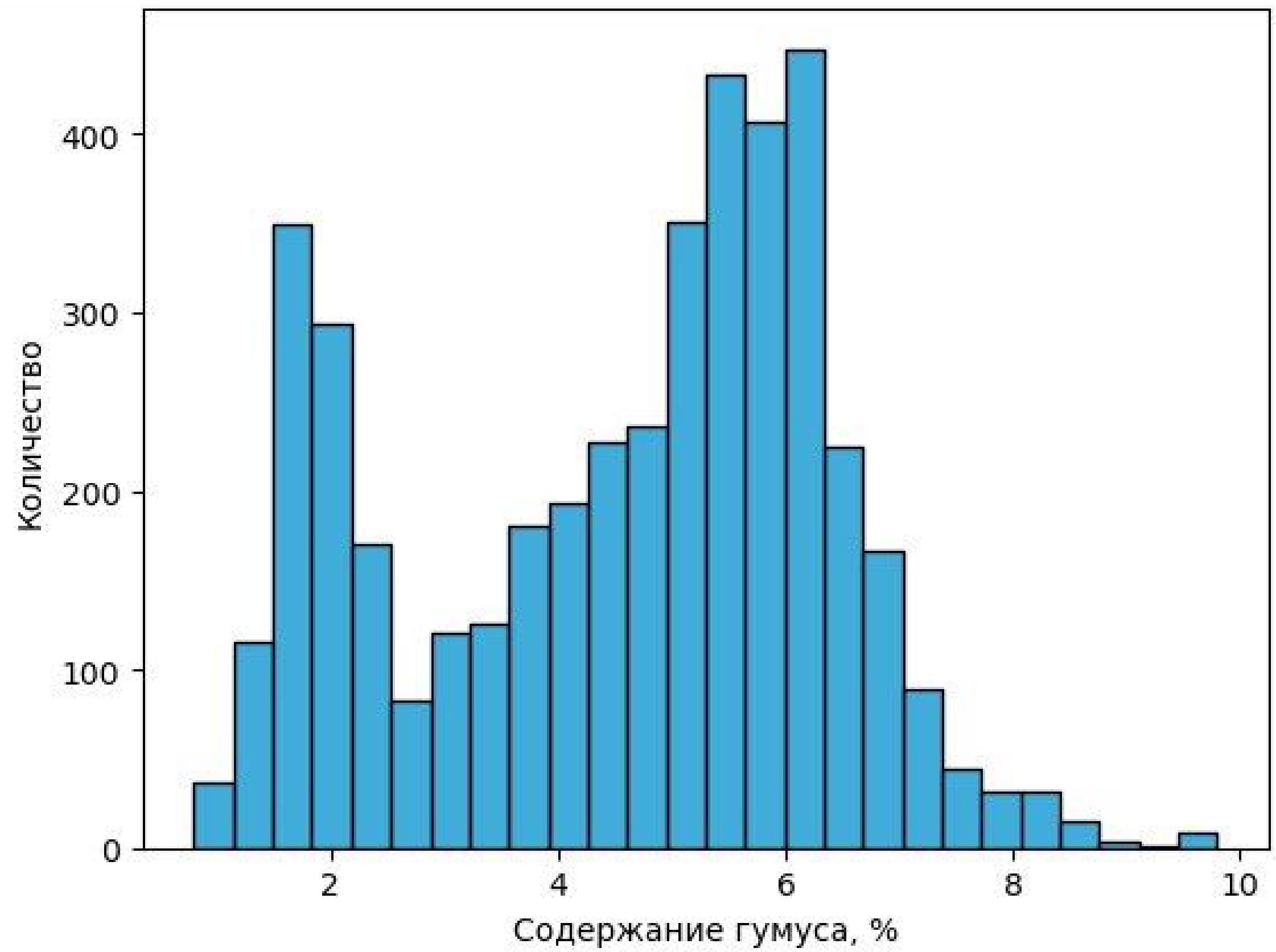
Вероятно, ошибка вноса данных

Находим выбросы методом  
межквартильного размаха и  
переводим их из **кг/га** в **т/га**

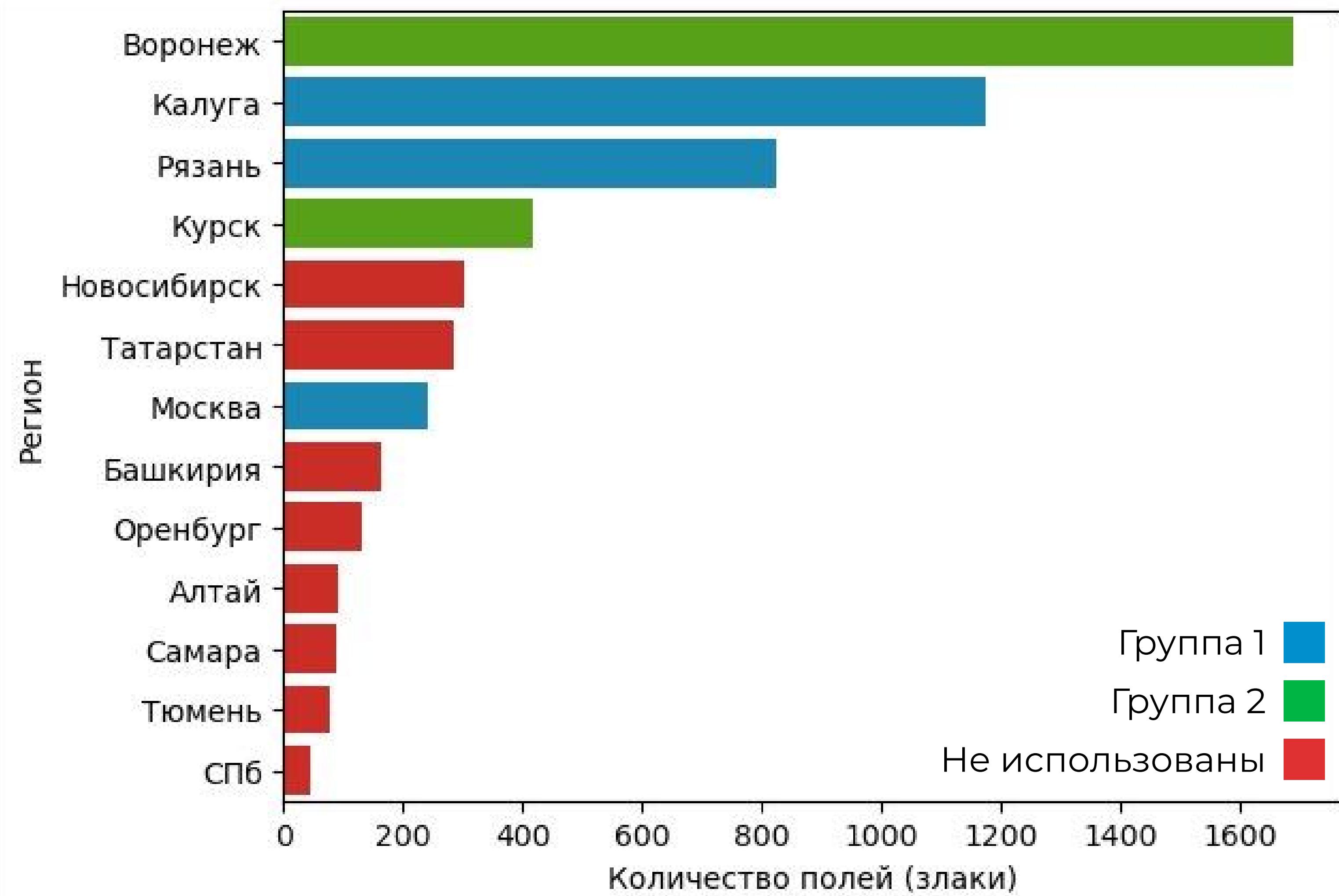
# Распределение калия и фосфора



# Распределение гумуса и кислотности почвы (Ph)



# Распределение злаковых полей по регионам

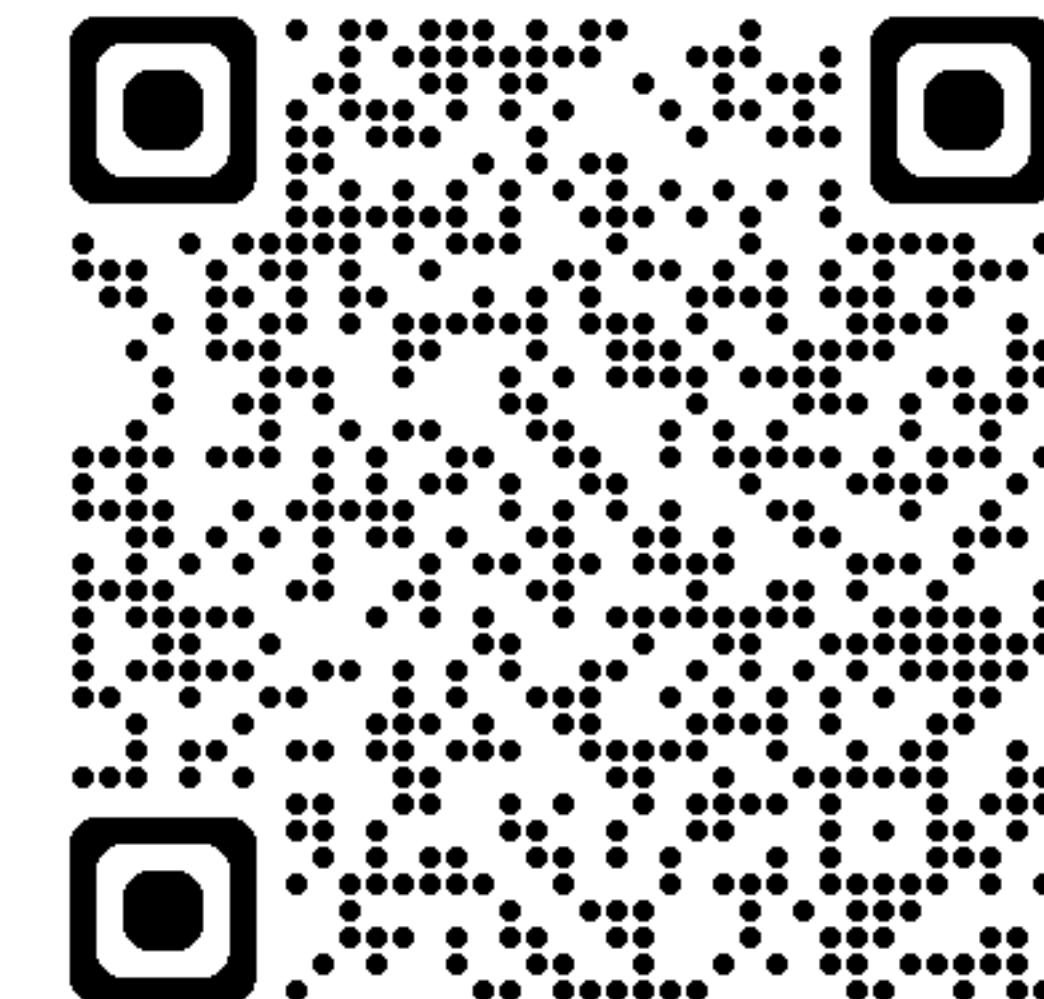


## 1. Калуга, Москва и Рязань

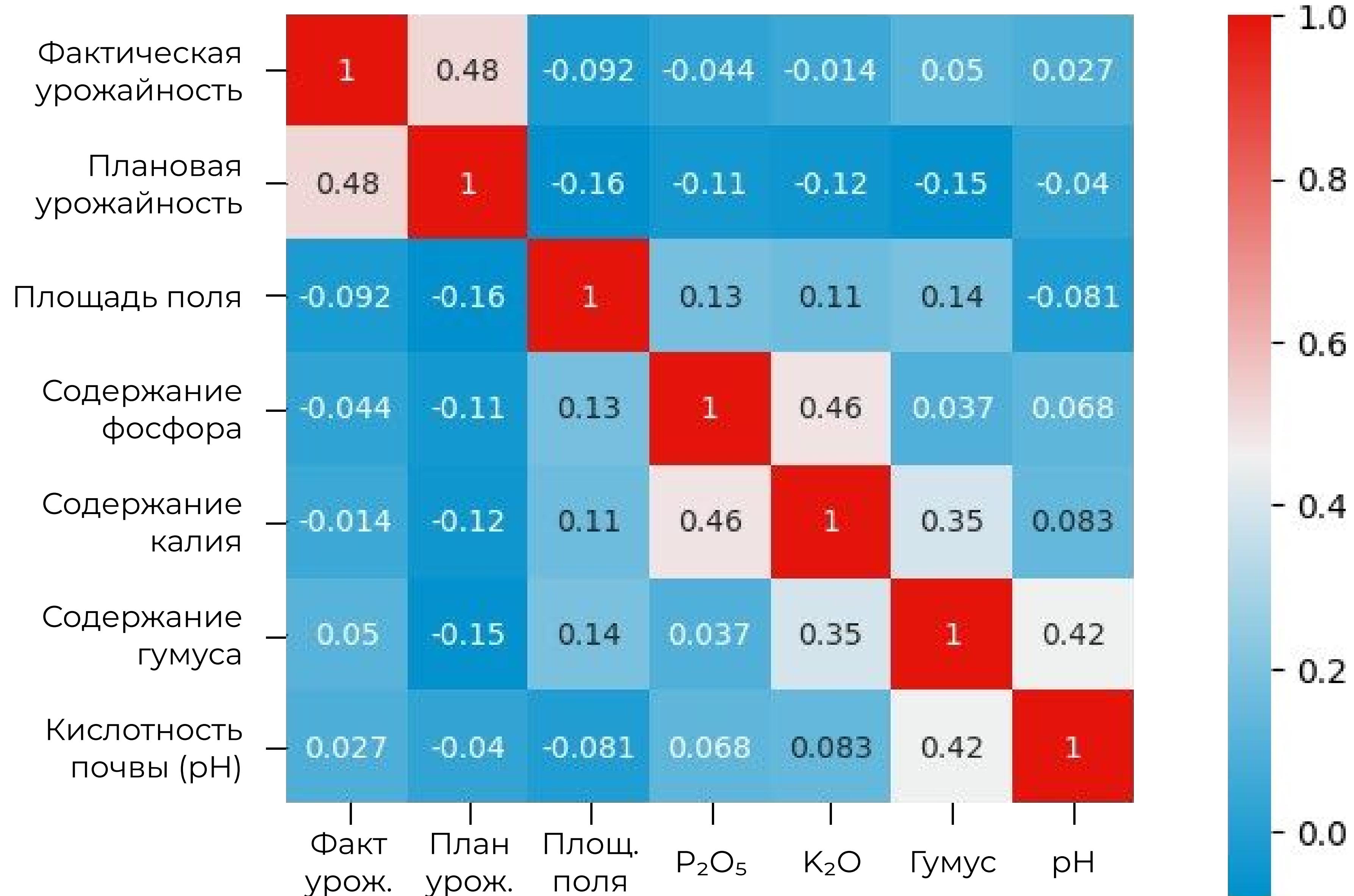
- Дерново-подзолистые
- Серые

## 2. Курск и Воронеж

- Черноземы



Карта почв России



## Очистка от выбросов

### 1 Химический состав почвы

- NaN в P<sub>2</sub>O<sub>5</sub>, K<sub>2</sub>O, Гумус, pH
- Выбросы по квартилям

**5797** - пустые  
**699** - выбросы

### 2 Фактическая урожайность

- NaN в yield\_fact
- Выбросы по квартилям

**11298** - пустые  
**96** - выбросы

### 3 Плановая урожайность

- NaN в yield\_plan
- Выбросы по квартилям

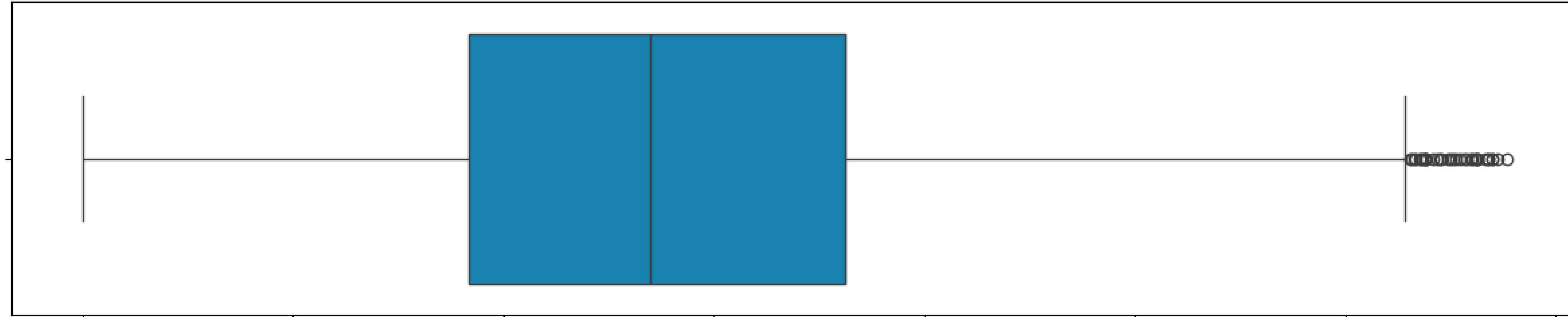
**5205** - пустые  
**191** - выбросы

**Было: 16060 полей / Стало: 3558 полей**

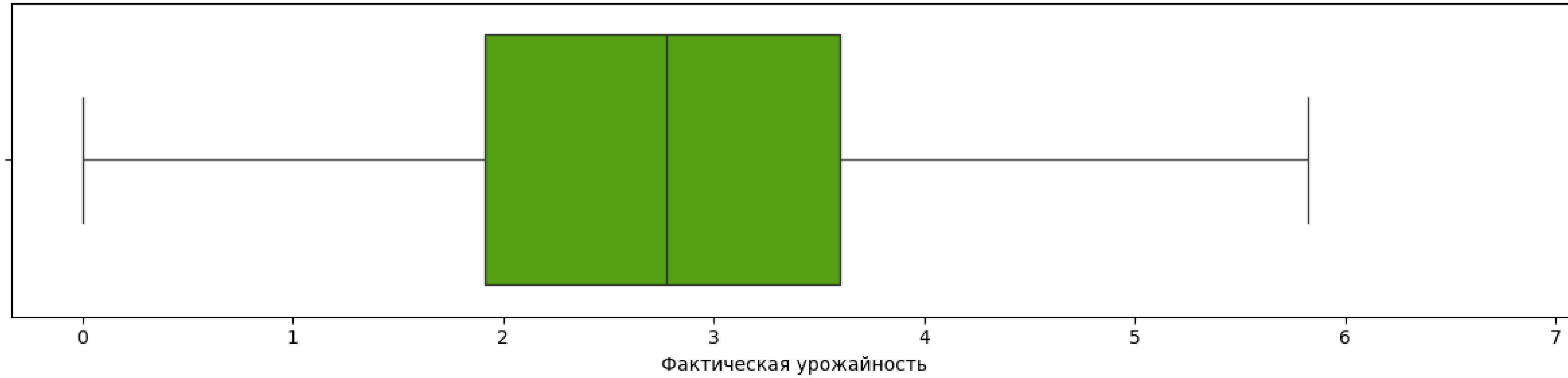
Из них по злакам **2203 поля**

# Сравнение выборок

Все значения

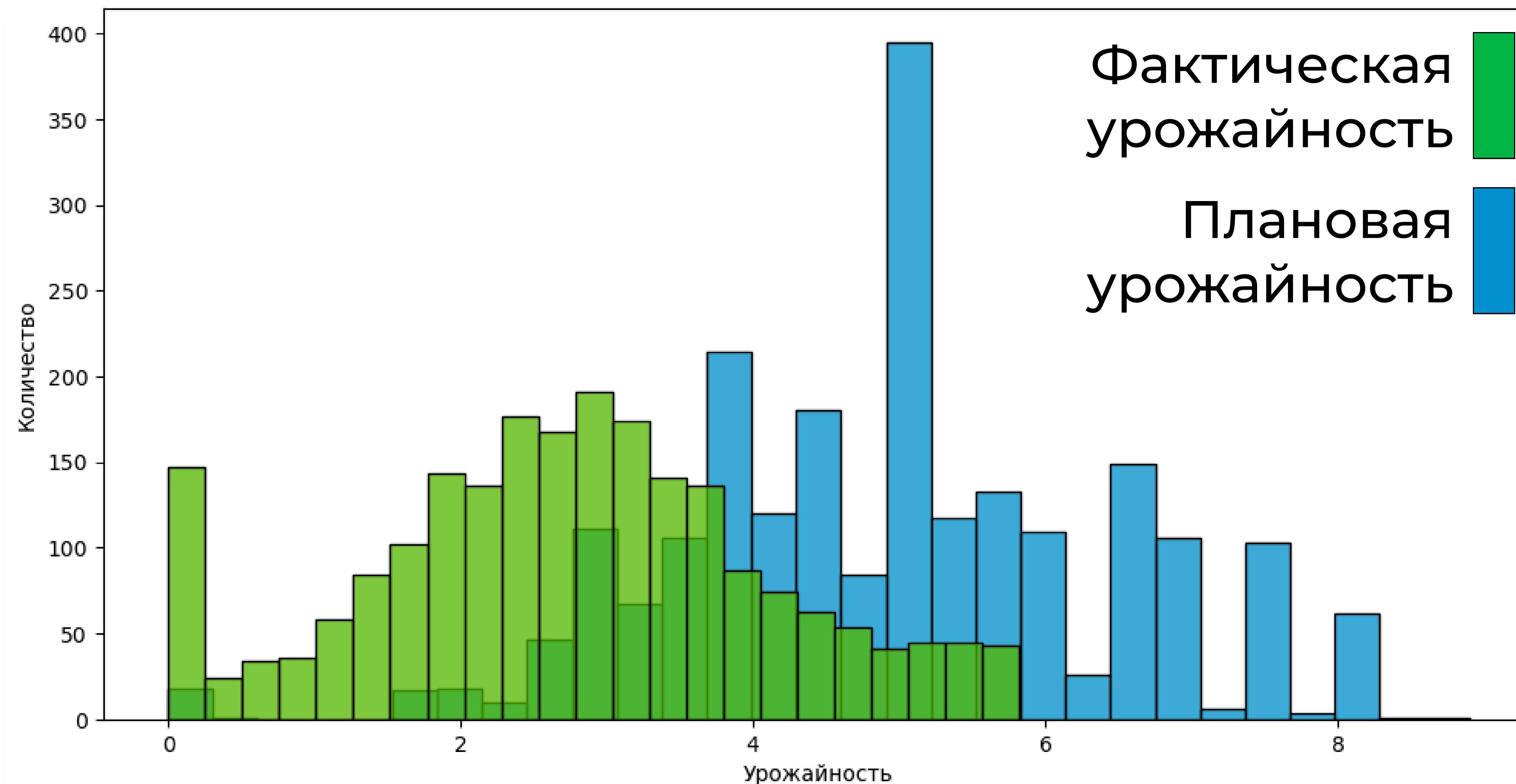


Без выбросов и аномалий



Фактическая урожайность

# Распределение фактической и плановой урожайности

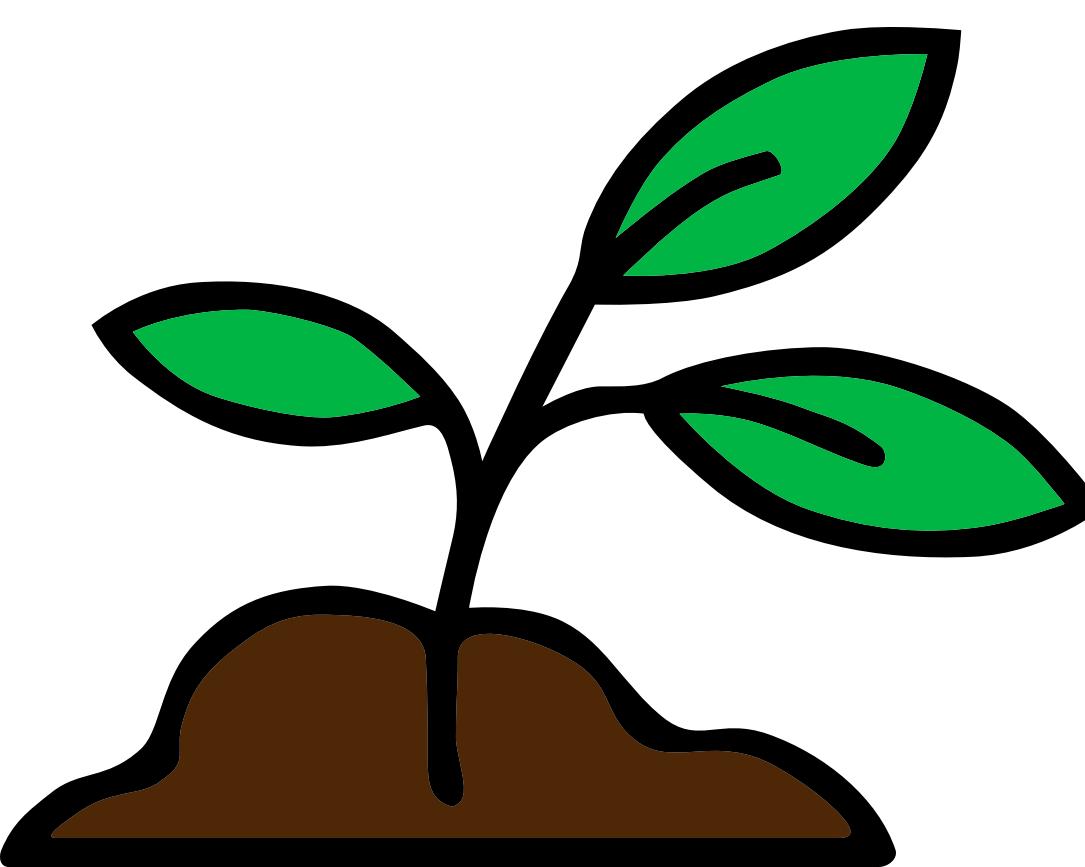


## Исследовательский вопрос

Как на основе химического состава почвы предсказать её урожайность?

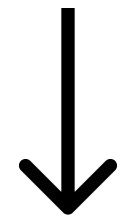
Гипотеза

Квадратичная зависимость между фактической урожайностью и химическим составом почвы позволяет делать релевантные прогнозы



## Механизм

Дефицит и профицит химических элементов в составе почвы негативно сказывается на урожайности<sup>[4]</sup>



Данная зависимость может быть представлена квадратичной функцией



Квадратичная зависимость между составом почвы и её урожайностью может быть использована для предсказаний



## Исследуемые показатели

Целевая переменная	Предикторы	Критерии подвыборок
<b>Фактическая урожайность</b>	<ul style="list-style-type: none"><li>• Калий (<math>K_2O</math>)</li><li>• Фосфор (<math>P_2O_5</math>)</li><li>• Гумус</li><li>• Кислотность почвы (рН)</li><li>• Группа регионов</li></ul>	<b>Семейство культур</b> (Злаковые)

## Математическая модель

**Цель:** Установить параметры зависимости между содержанием хим. элементов в почве и фактической урожайностью полей для различных групп регионов

**Метод:** Полиномиальная регрессия второго порядка

## Стандартное предположение

**Н<sub>0</sub>**

Полученная модель **не имеет статистически значимой нелинейной связи** между исследуемыми факторами и фактической урожайностью (коэффициенты не значимы)

**Н<sub>1</sub>**

Полученная модель **имеет статистически значимую нелинейную связь** между исследуемыми факторами и фактической урожайностью (коэффициенты значимы)



P-value

**<0.05**

## Разделение датасета

Для обучения база была разделена на 2 выборки:

1. 80% (train)
2. 20% (test)



Обучение модели по одной выборке и проверка точности по другой



# Feature engineering

Анализ регрессии

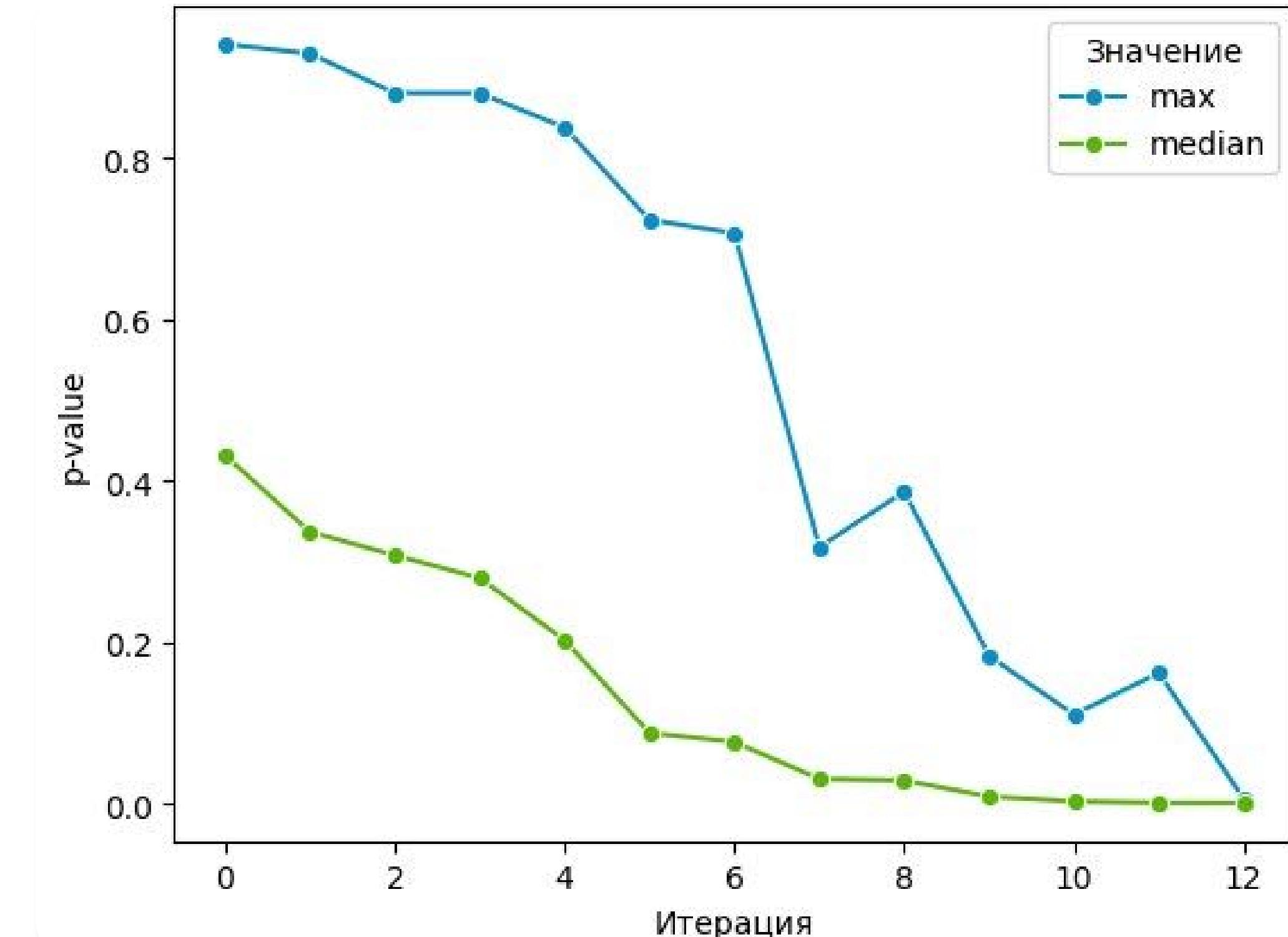
Есть ли фичи с  
 $p\text{-value} > 0.05?$

нет

да

Удалить самую  
статистически  
незначимую фичу

Все факторы  
статистически  
значимы



21 → 8  
фичей в  
регрессии

## Вывод по регрессии

R-squared: **0.317**

F-statistic: **6.358**

Prob (F-stat): < **0.01**

Проверка на  
**мультиколлинеарность**

показала ее отсутствие

x0 - Фосфор; x1 - Калий;

x2 - Гумус; x3 - pH; x4 -

is\_group2 (бинарный  
критерий региона)

Фичи предиктора	coef	std error	p-value
<b>const</b>	2.9830	0.174	< 0.05
<b>x0 * x1</b>	0.0000	0.001	< 0.05
<b>x0 * x4</b>	-0.0043	0.002	< 0.05
<b>x1 * x3</b>	-0.0011	0.001	< 0.05
<b>x1 * x4</b>	0.0045	0.001	< 0.05
<b>x2^2</b>	-0.0495	0.015	< 0.05
<b>x2 * x3</b>	0.0919	0.023	< 0.05
<b>x3^2</b>	-0.0303	0.010	< 0.05

# Проверка на устойчивость

Выборки	MAE модели	MAE агронома
23&24 train	<b>1,04</b>	2,40
23&24 test	<b>0,96</b>	2,31
23	<b>1,17</b>	2,32
24	<b>0,87</b>	2,45



## Интерпретация

**Гипотеза подтвердилась на семействе злаковых культур в группе 1 (Курск и Воронеж) и группе 2 (Калуга, Москва и Рязань)**

На других выборках модель не обучалась из-за малого числа данных

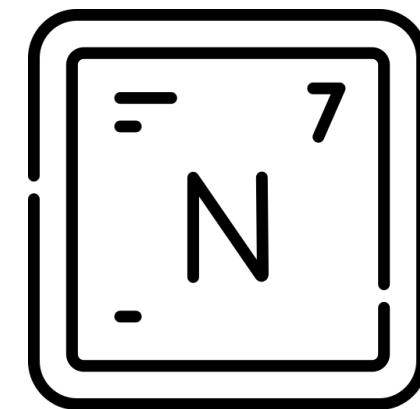
- Используя квадратичную зависимость, мы вывели все статистически значимые комбинации элементов
- Согласно максимальной абсолютной ошибки, модель оказалась в среднем в 2 раза точнее прогнозов агрономов по результатам 4 тестов



## Ограничения



Недостаточно данных для создания модели по другим регионам и культурам

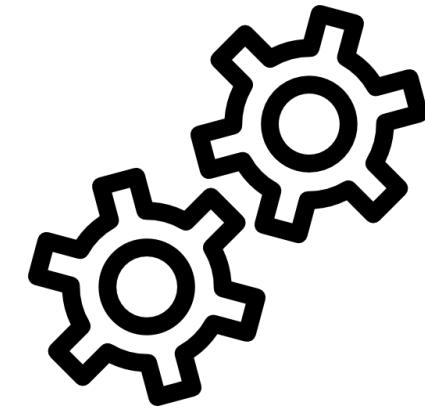


Химический состав почвы представлен не всеми основными элементами, что снижает точность предсказаний

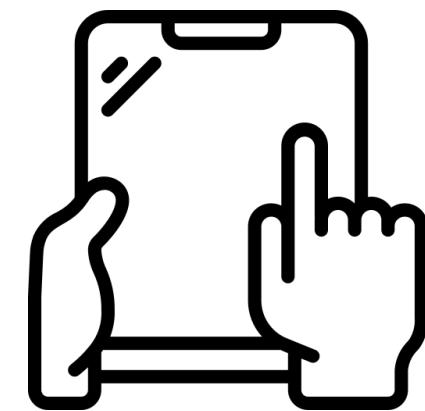


Отсутствие данных по внешним эффектам в рамках региона (количество осадков, средняя температура)

## Перспективы



Добавить значения ключевых минералов (азот, железо), а также контрольные параметры (осадки, количество солнечных дней) для улучшения модели

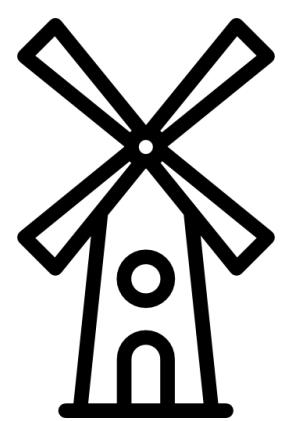


Провести тренинги для агрономов по использованию планшетов на полях для снижения количества пустых ячеек и ошибок в данных

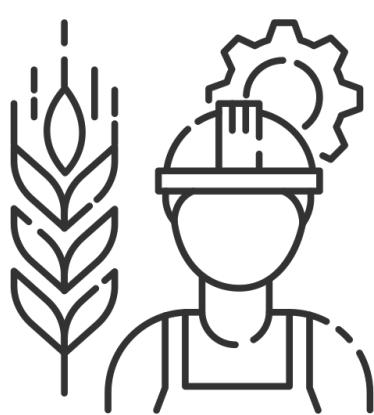
## Практическое применение



**Крупные фермы и компании:** снижение издержек на создание логистической инфраструктуры



**Малые и средние фермы:** использование модели для прогноза урожая, который пойдёт на продажу и останется для личного потребления, особенно когда невозможно провести полноценные исследования



**Агрономы:** более эффективный и точный прогноз урожайности позволит больше времени уделять экспериментам и уходу за полями

## Policy Implication



**Создание стабильных рабочих мест** в сельском хозяйстве за счет лучшего прогнозирования объемов урожая для государства

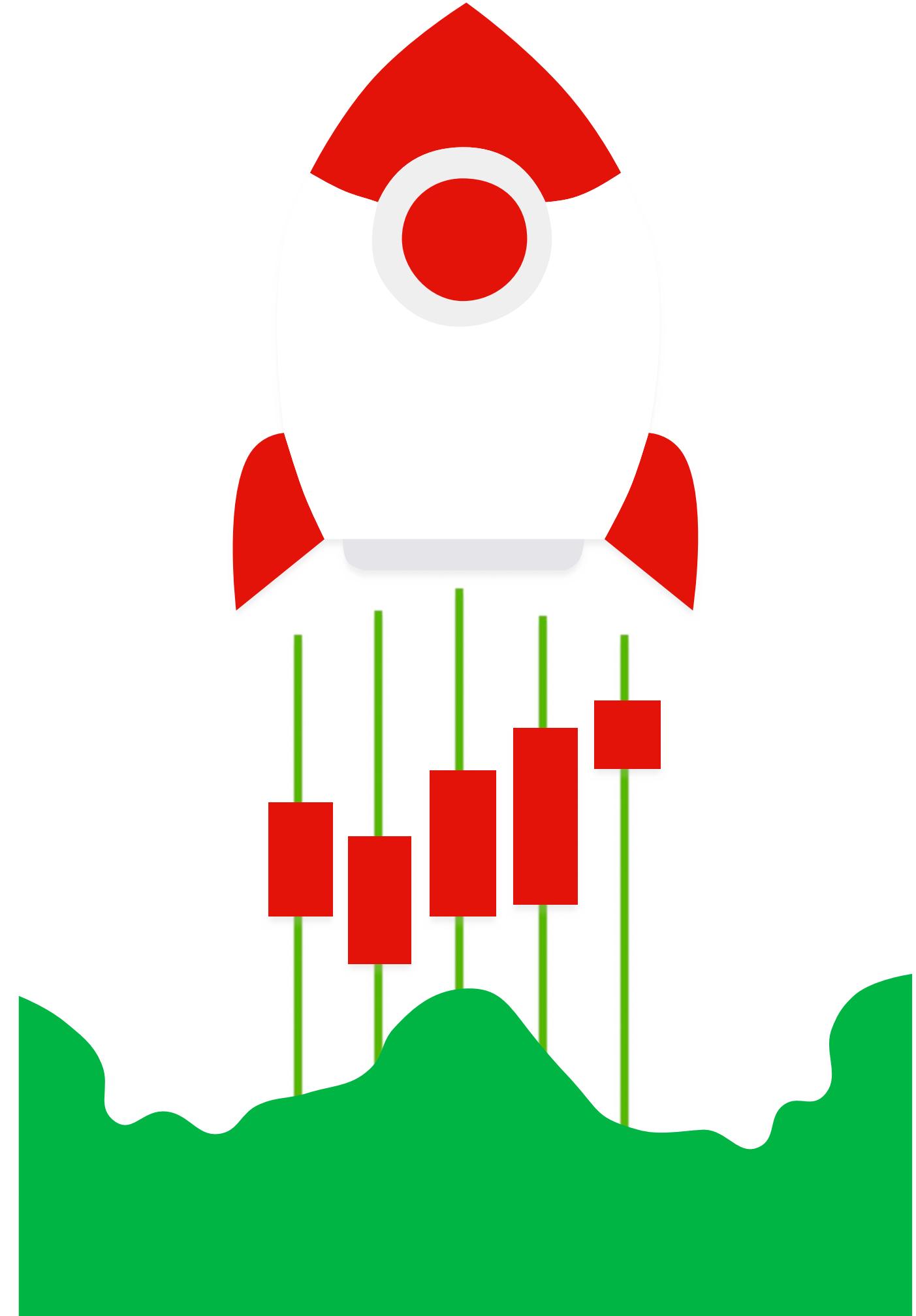


**Оценка уровня риска фьючерсов** на поставку продуктов, заключенных с ритейлерами в начале сезона для ГК “ЭкоНива” с возможностью снизить шанс невыполнения обязательств и увеличить объем поставок

# Все идет по плану

Гайнетдинов Ян  
Шаклеин Владимир  
Редлих Алексей  
Бушин Андрей

TO  
THE  
MOON



## Источники

1. Каталог основных сельскохозяйственных культур // baa.by URL: [https://baa.by/upload/agrarnye-klassy/katalog\\_plant.pdf](https://baa.by/upload/agrarnye-klassy/katalog_plant.pdf) (дата обращения: 13.12.2024).
2. Способы уборки зерновых культур // Б-Истокское РТПС URL: <https://istokrtps.ru/stati/sposobyi-uborki-zernovyix-kultur?ysclid=m4msfaokjn25221713> (дата обращения: 13.12.2024).
3. Рекорды Гиннесса по урожайности пшеницы и ячменя // [direct.farm](https://direct.farm/post/rekordy-ginnessa-po-urozhaynosti-pshenitsy-i-yachmenya-23237) URL: <https://direct.farm/post/rekordy-ginnessa-po-urozhaynosti-pshenitsy-i-yachmenya-23237> (дата обращения: 13.12.2024).
4. Влияние агрохимических свойств почвы и удобрений на урожайность культур в зернопропашном севообороте // cyberleninka URL: <https://cyberleninka.ru/article/n/vliyanie-agrohimicheskikh-svoystv-pochvy-i-udobreniy-na-urozhaynost-kultur-v-zernopropashnom-sevooborote?ysclid=m4tребе9ww749819922> (дата обращения: 13.12.2024).
5. Технология заготовки сенажа, силоса и кornажа // [studme.org](https://studme.org/404311/agropromyshlennost/tehnologiya_zagotovki_senazha_silosa_kornazha) URL: [https://studme.org/404311/agropromyshlennost/tehnologiya\\_zagotovki\\_senazha\\_silosa\\_kornazha](https://studme.org/404311/agropromyshlennost/tehnologiya_zagotovki_senazha_silosa_kornazha) (дата обращения: 13.12.2024).

# GitHub



## Приложение:

результаты регрессии

R-squared: **0.317**

Adj. R-squared: **0.295**

F-statistic: **6.358**

Prob (F-stat): **0.000**

Фичи предиктора	coef	std error	p-value
<b>const</b>	2.9830	0.174	0.000
<b>x0 * x1</b>	0.0000	0.001	0.001
<b>x0 * x4</b>	-0.0043	0.002	0.006
<b>x1 * x3</b>	-0.0011	0.001	0.000
<b>x1 * x4</b>	0.0045	0.001	0.002
<b>x2^2</b>	-0.0495	0.015	0.001
<b>x2 * x3</b>	0.0919	0.023	0.000
<b>x3^2</b>	-0.0303	0.010	0.003

## Приложение: проверка на мультиколлинеарность

	coef
1	2.9830
x0 x1	2.721e-05
x0 x4	-0.0043
x1 x3	-0.0011
x1 x4	0.0045
x2^2	-0.0495
x2 x3	0.0919
x3^2	-0.0303

	feature	VIF
1	x0 x1	5.613223
2	x0 x4	7.122009
3	x1 x3	6.807651
4	x1 x4	7.195287
5	x2^2	32.229487
6	x2 x3	47.105142
7	x3^2	4.988077

Удаляем x2 \* x3 →

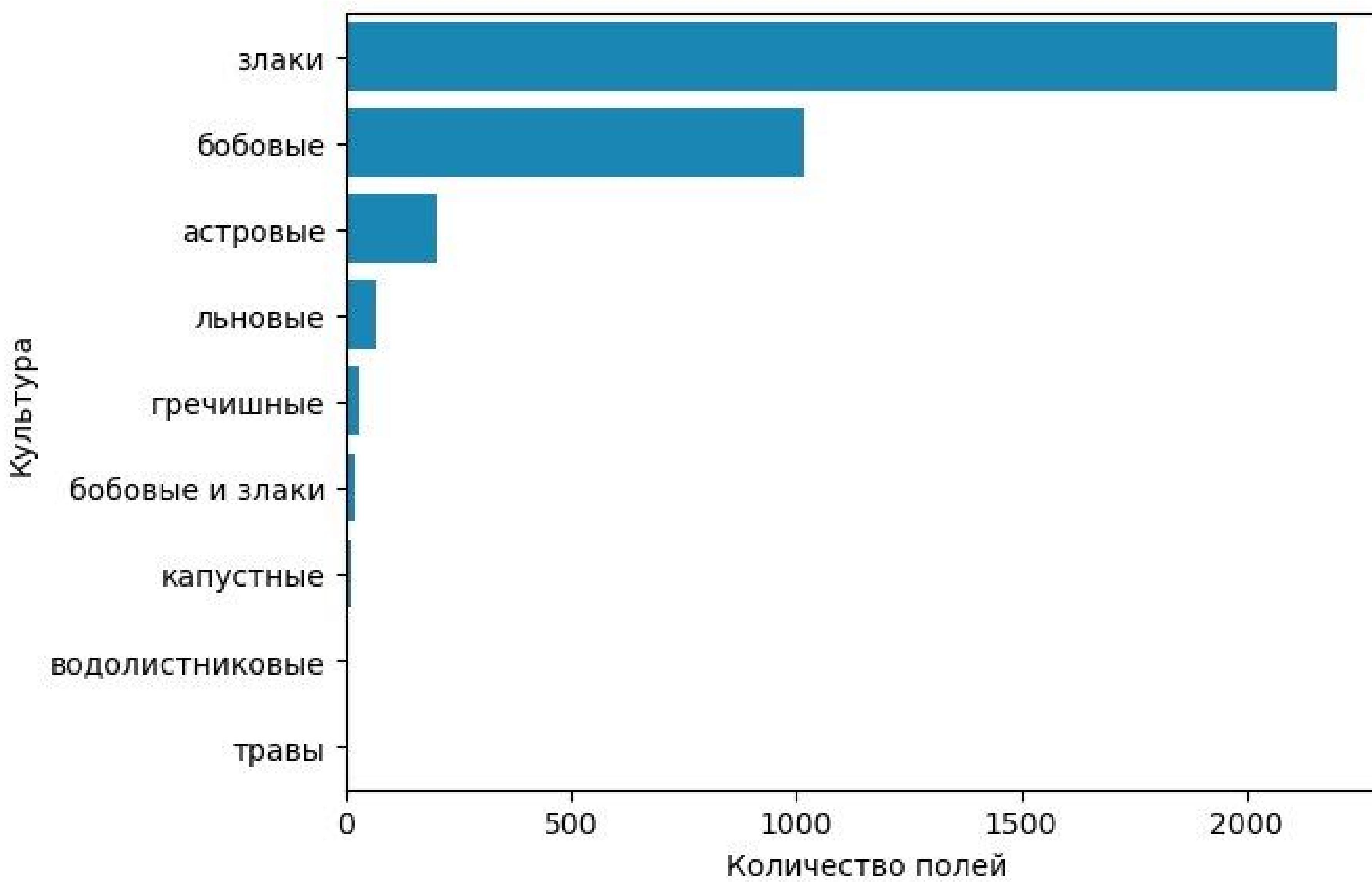
	coef
1	2.7640
x0 x1	2.255e-05
x0 x4	-0.0033
x1 x3	-0.0010
x1 x4	0.0045
x2^2	0.0096
x3^2	0.0044

	feature	VIF
1	x0 x1	5.478538
2	x0 x4	6.939547
3	x1 x3	6.654825
4	x1 x4	7.193079
5	x2^2	1.624052
6	x3^2	1.383000

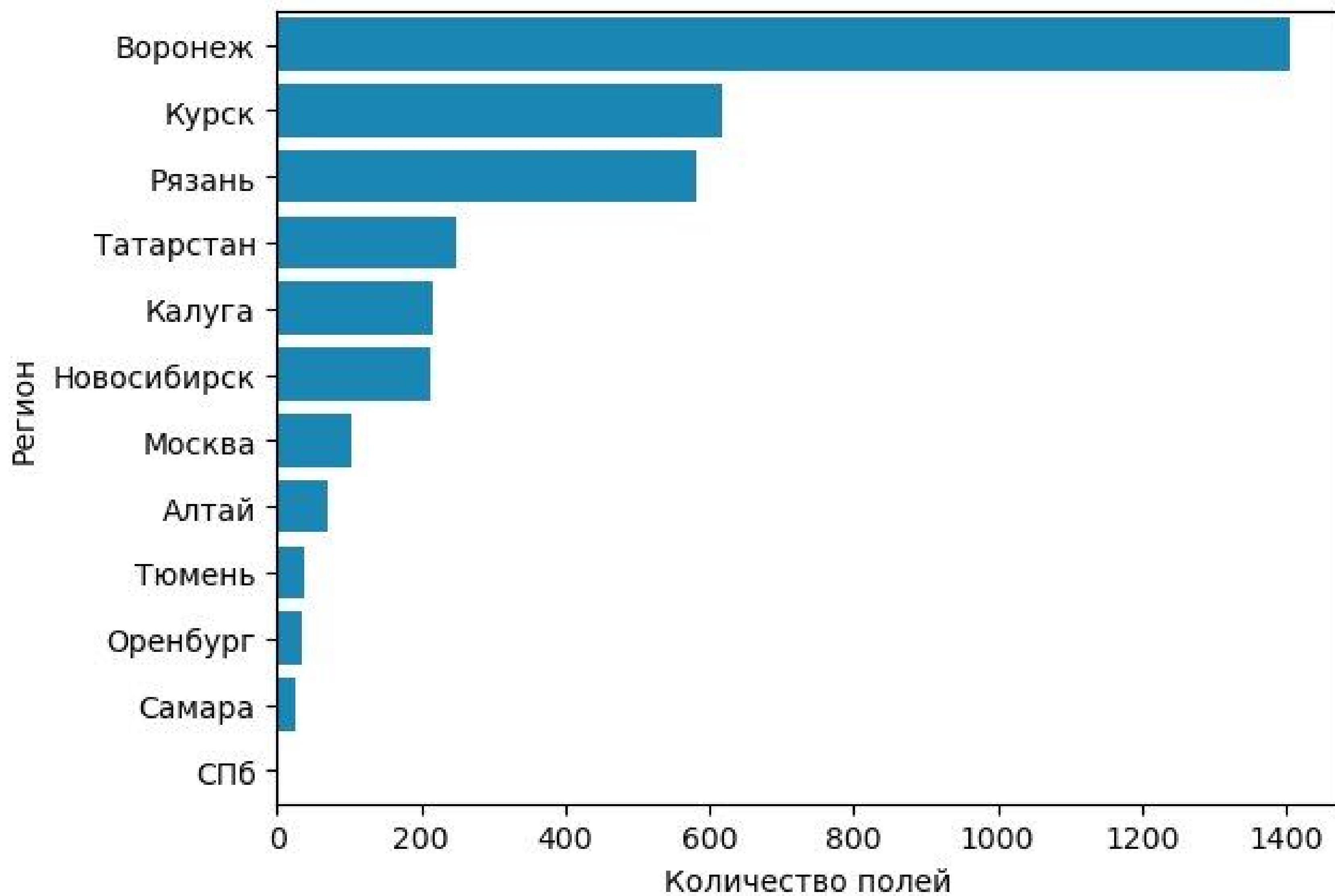
# Приложение:

## карта почв России



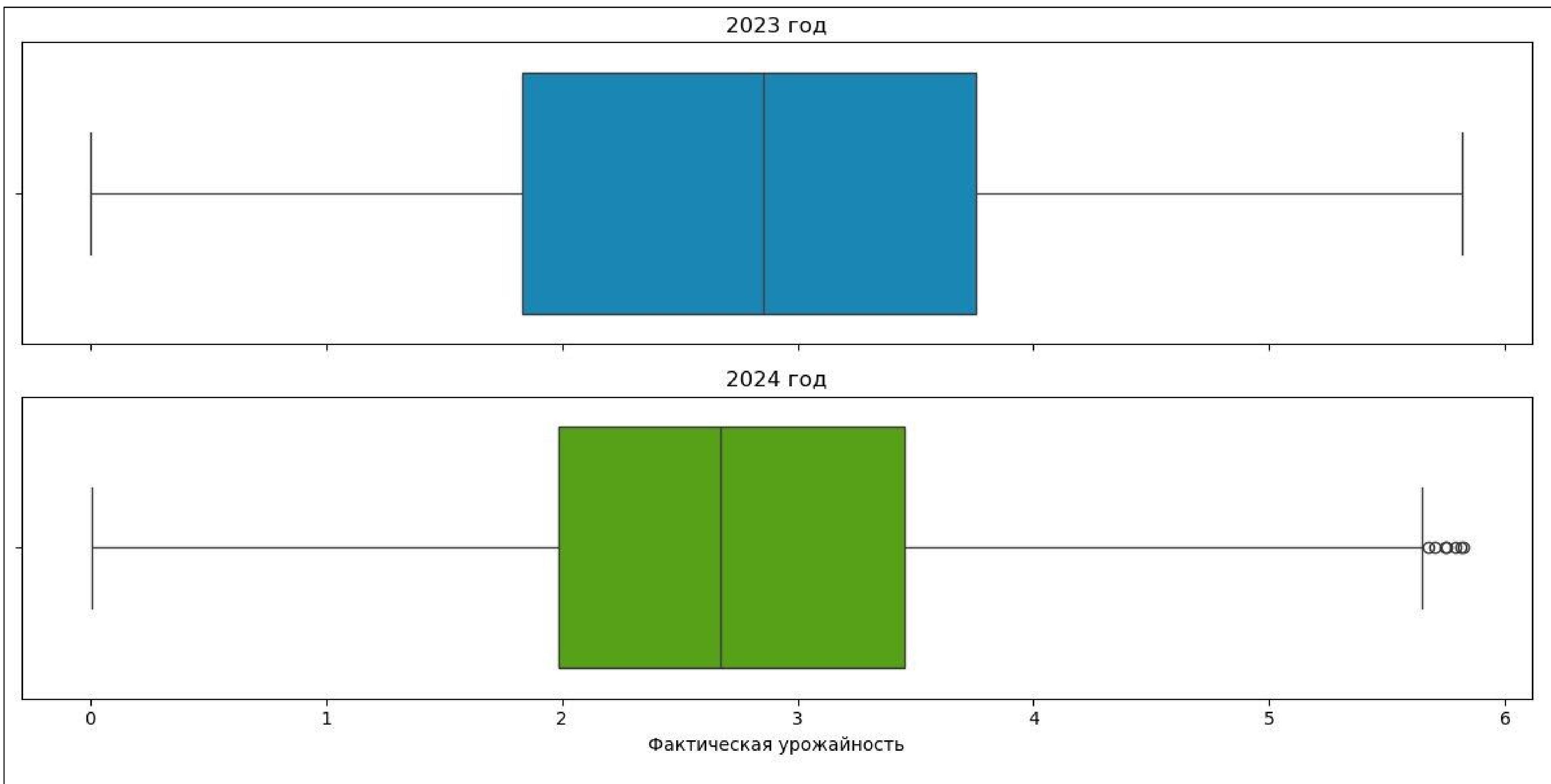


**Приложение:**  
количество полей по культурам

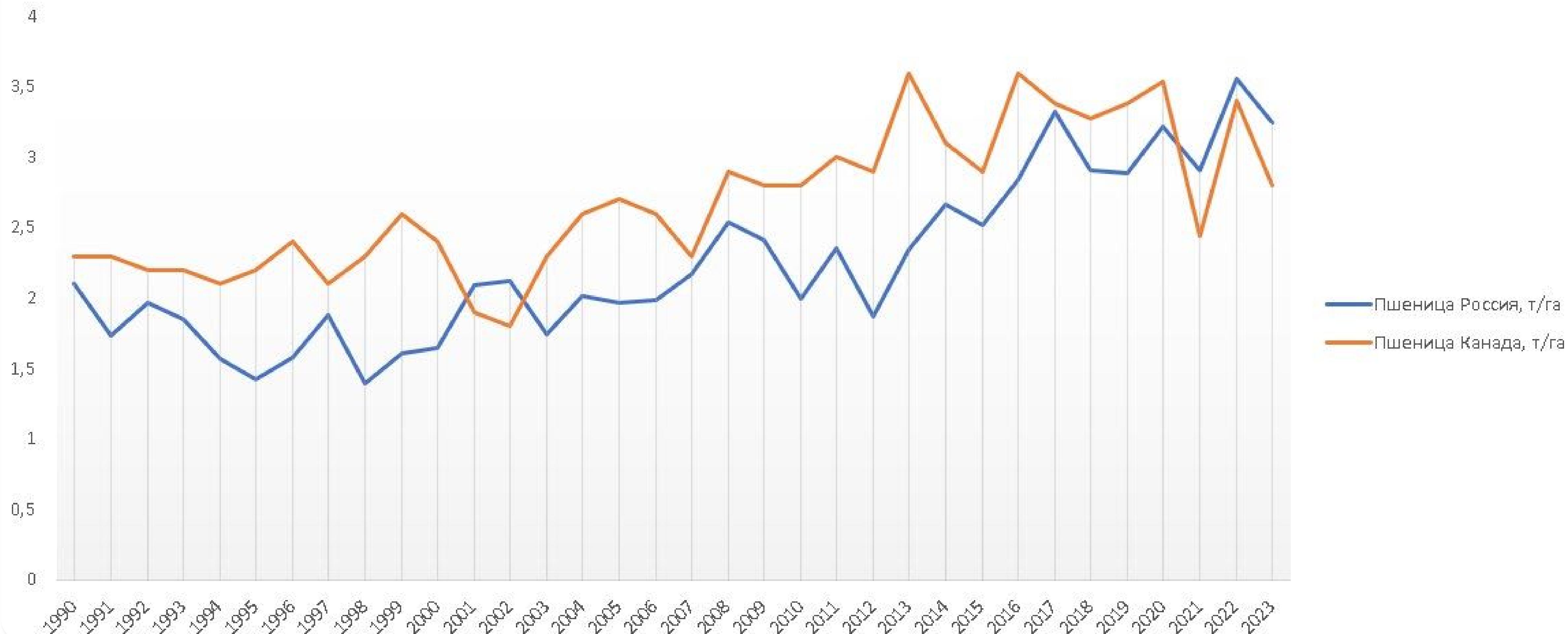


**Приложение:**  
количество полей по регионам

## Приложение: распределение фактической урожайности по сезонам



## Динамика урожайности пшеницы в Канаде и России



# Приложение: распределение химического состава почвы после очистки выбросов

