

# RLMS-HSE:

От чего зависит  
заработная плата?



# Датасет

Российский мониторинг  
экономического положения и  
здоровья населения НИУ ВШЭ  
(RLMS-HSE)

Данные собраны за **2017** год

**3872** записи

**87** столбцов

**38** регионов РФ

# Характеристики индивидов

Количественные данные

| Название | Описание содержимого столбца       |
|----------|------------------------------------|
| Inwage   | Логарифм заработной платы индивида |
| age      | Возраст индивида в годах           |
| children | Количество детей                   |

# Характеристики индивидов

Качественные данные

| Название | Описание содержимого столбца                          |
|----------|-------------------------------------------------------|
| public   | Является ли место работы государственным предприятием |
| educ     | Уровень образования индивида                          |
| urban    | Место жительства индивида                             |
| male     | Пол индивида                                          |
| industry | Отрасль занятости индивида                            |

Доход

**IQR – межквартильный диапазон**

**q1 – первый квартиль (25%)**

**q3 – третий квартиль (75%)**

**Оставляем значения всех числовых переменных, принадлежащих промежутку :  
[q1 - 1,5 \* IQR ; q3 + 1,5 \* IQR]**

**Было:  
3872 ответа**

**Удалено:  
177 ответов**

**Стало:  
3695 ответов**

Возраст

**Не рассматриваем данные о клиентах  
моложе 23 лет**

# Преобразования БД

1

**status**

Тип населённого пункта

**status**

Тип населённого пункта в виде категориальной переменной, где 1 – село, а 4 – областной центр

2

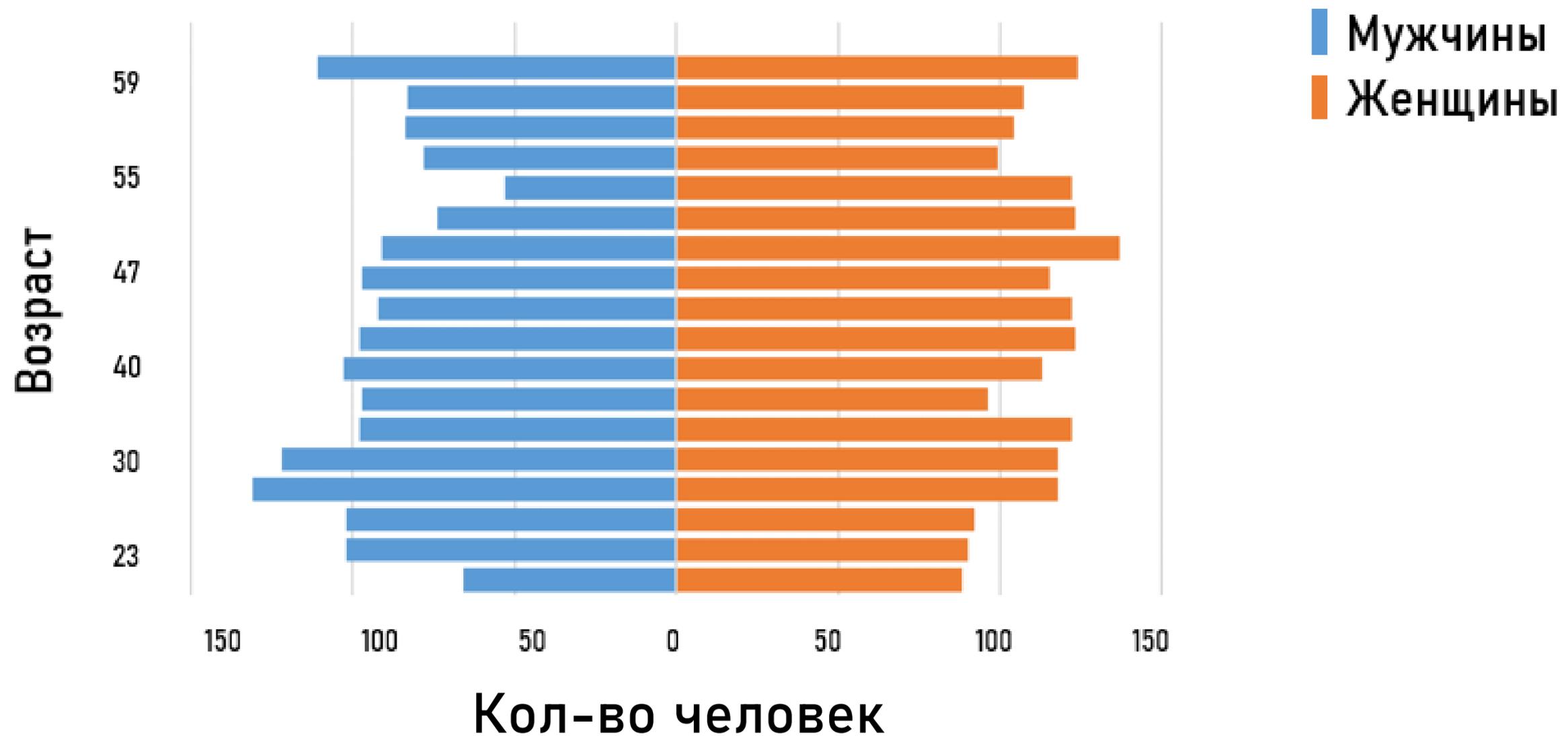
**industry**

Неполные наименования

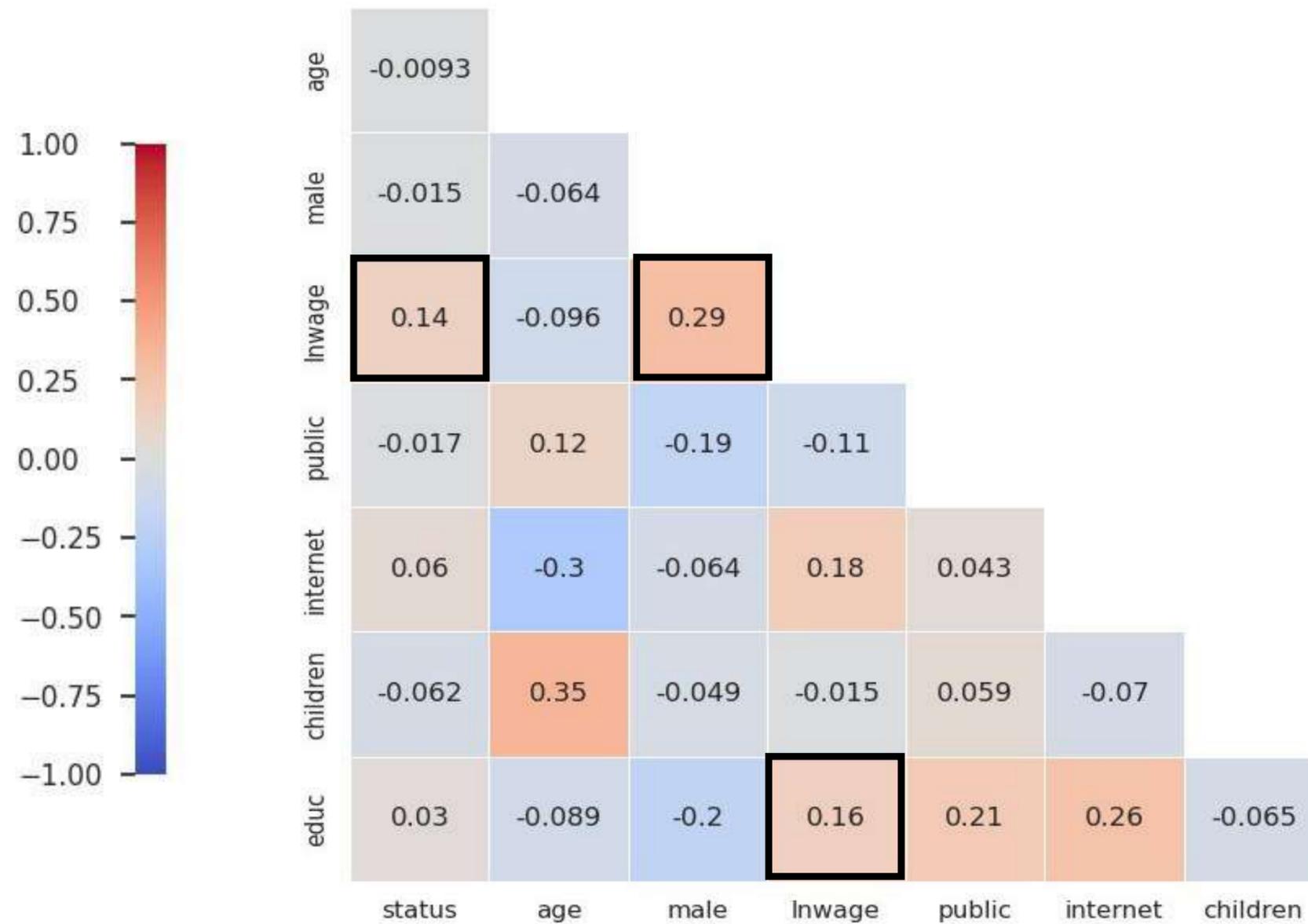
**industry**

Восстановленные названия из столбцов укрупнённых отраслей занятости (id64-id75)

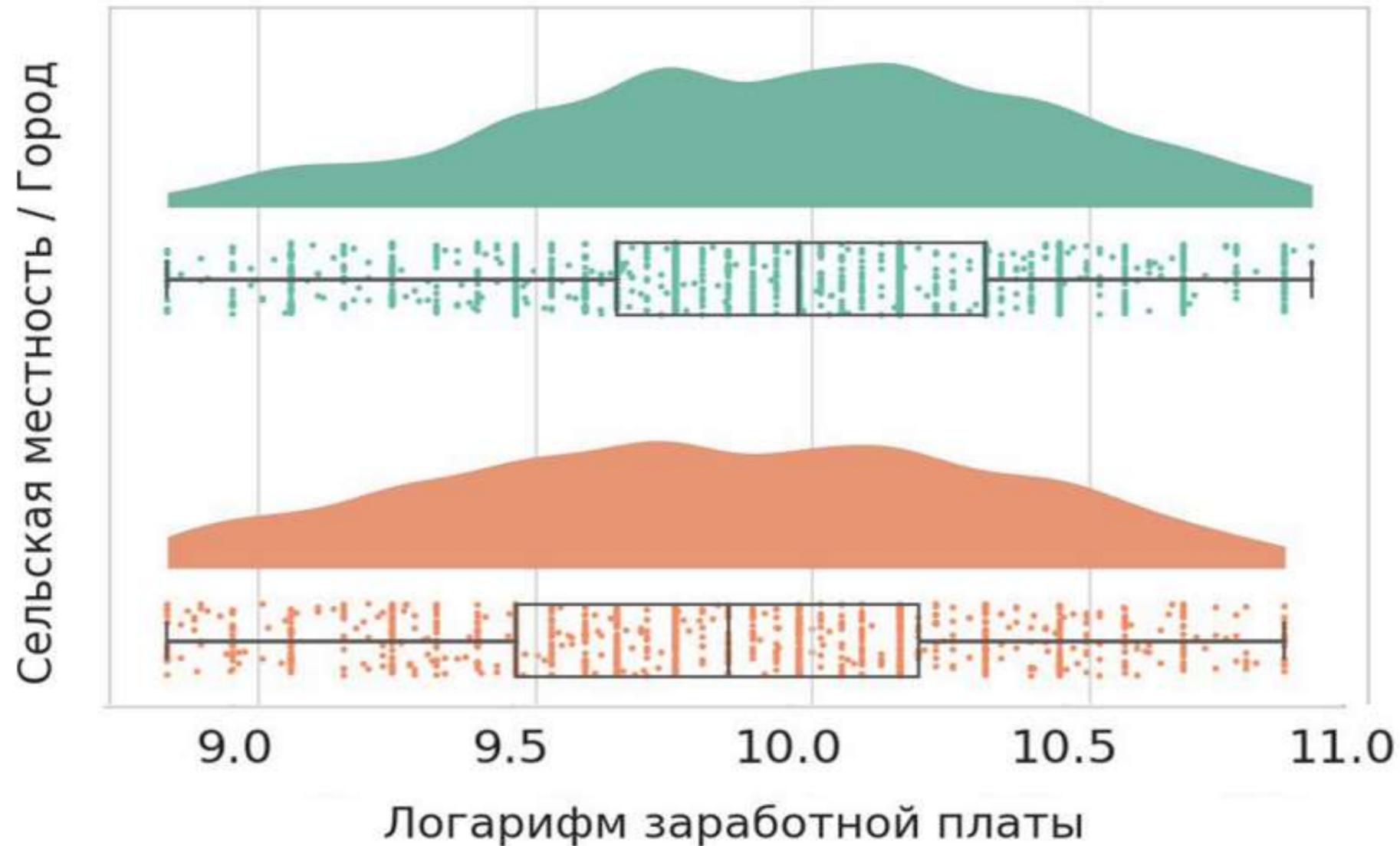
# Возрастно-половая пирамида



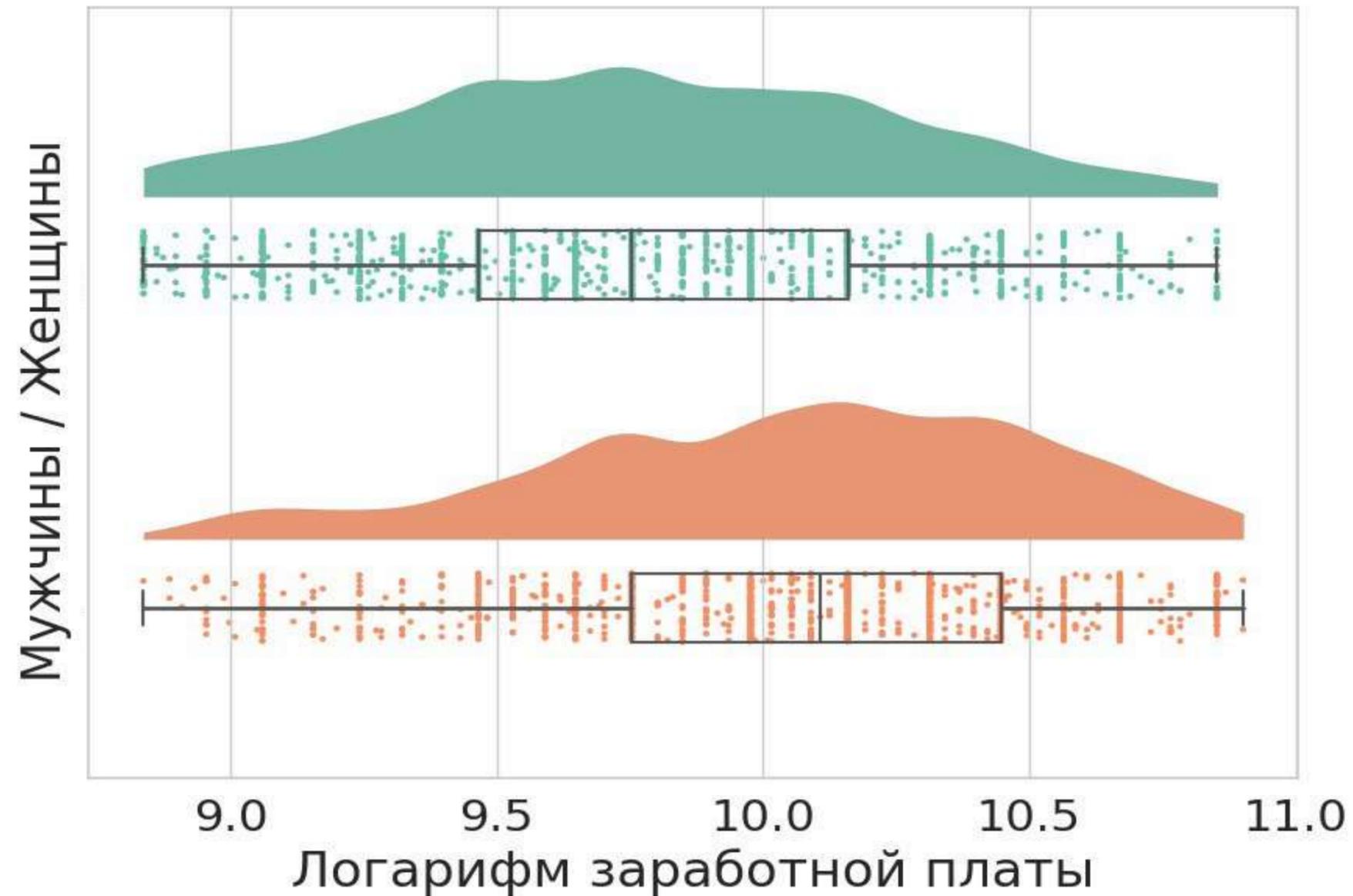
# Матрица корреляций



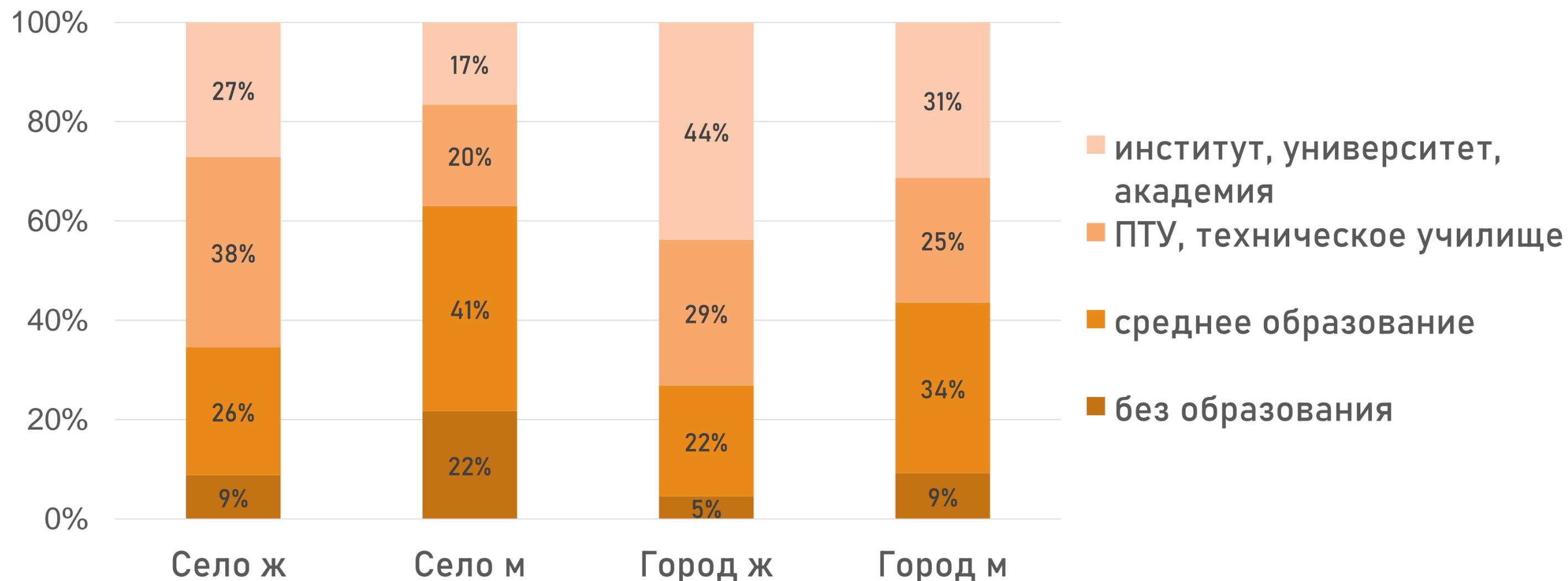
# Распределение заработной платы в городе и в сельской местности



# Распределение заработной платы у мужчин и у женщин



## Гистограмма долей видов полученного образования среди женщин и мужчин, проживающих в городе и в сельской местности



# Исследовательский вопрос

Как влияет получение образования на уровень дохода мужчин и женщин, проживающих в сельской местности?

# Гипотеза

**Получение образования мужчинами в сёлах имеет меньшее влияние на их уровень дохода, чем у женщин**

# Механизм

Мужчины способны на тяжёлый физический труд, который более распространён и высоко оплачивается в сёлах

Получение образования для высокой заработной платы не требуется

Основные отрасли занятости для женщин в сёлах – образование и здравоохранение

Для этих профессий получение образования необходимо

# Математическая модель

$H_0$  — Получение образования мужчинами в сёлах не имеет влияния на их уровень дохода.

$H_1$  — Получение образования мужчинами в сёлах имеет влияние на их уровень дохода.

Принимаемый уровень значимости: 0,05

1

Тест Шапиро-Уилка  
p-value > 0,05

2

T-тест зависимости дохода от образования  
p-value < 0,01

$H_0$   
отвергается

# Линейная регрессия

$H_0$  — Получение профессионального образования мужчинами в сёлах имеет **большее** влияние на их уровень дохода, чем у женщин

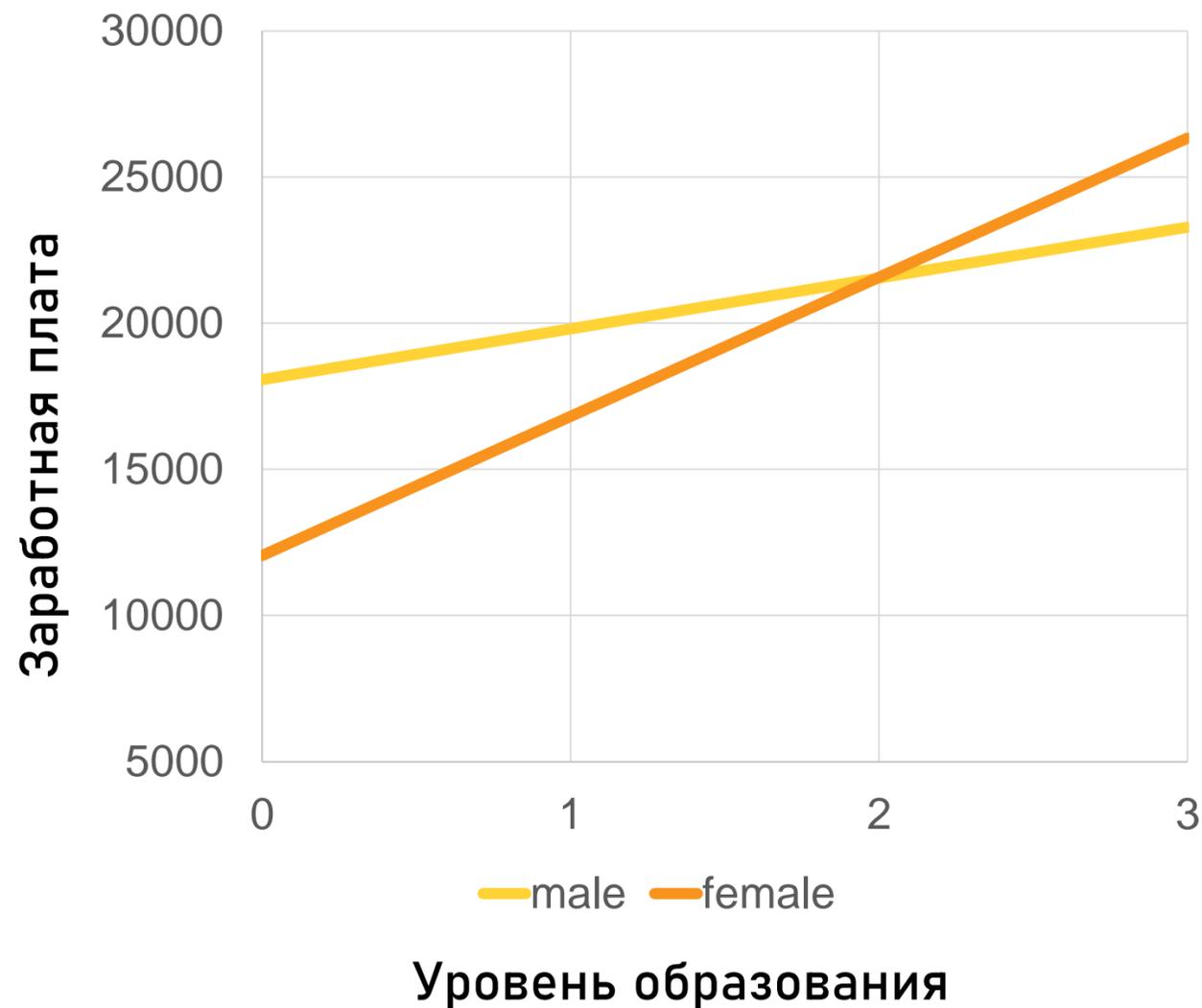
$H_1$  — Получение профессионального образования мужчинами в сёлах имеет **меньшее** влияние на их уровень дохода, чем у женщин

|             | Коэф.   | P-value |
|-------------|---------|---------|
| Пол         | 4760,95 | < 0,001 |
| Образование | 1736,13 | < 0,001 |
| Пол *       | 1260,13 | < 0,003 |
| Образование |         |         |

$H_0$  отвергается

# Интерпретация

Линейная регрессия



|         | Пол     | Образование | Пол*<br>Образование |
|---------|---------|-------------|---------------------|
| p-value | < 0,001 | < 0,001     | < 0,003             |

Гипотеза подтвердилась

Прирост зарплаты сельских мужчин в зависимости от их уровня образования ниже, чем у женщин

# Выводы

## Ограничения

- Нет информации о том, совпадает ли выбранная работа с полученным образованием
- В БД нет пожилых людей, а значит мы не можем обобщить исследования для них

## Перспективы

- Увеличение выборки; сбор данных в динамике (за 10 лет)
- Данные об опыте работы
- Данные о рисках для здоровья

## Policy Implication

Справочная информация

База для корректирования  
демографической политики

Усовершенствование  
профориентации



Датасет

Чистка БД

Пред. анализ

Исслед. вопрос

Гипотеза

Мат. модель

Выводы

Команда

# MONOVI



**Виктория  
Шакун**

Сортировка БД



**Лидия  
Северова**

Мат.  
Анализ



**Николай  
Волков**

Аналитик



**Полина  
Жуйкова**

Тимлид,  
Аналитик



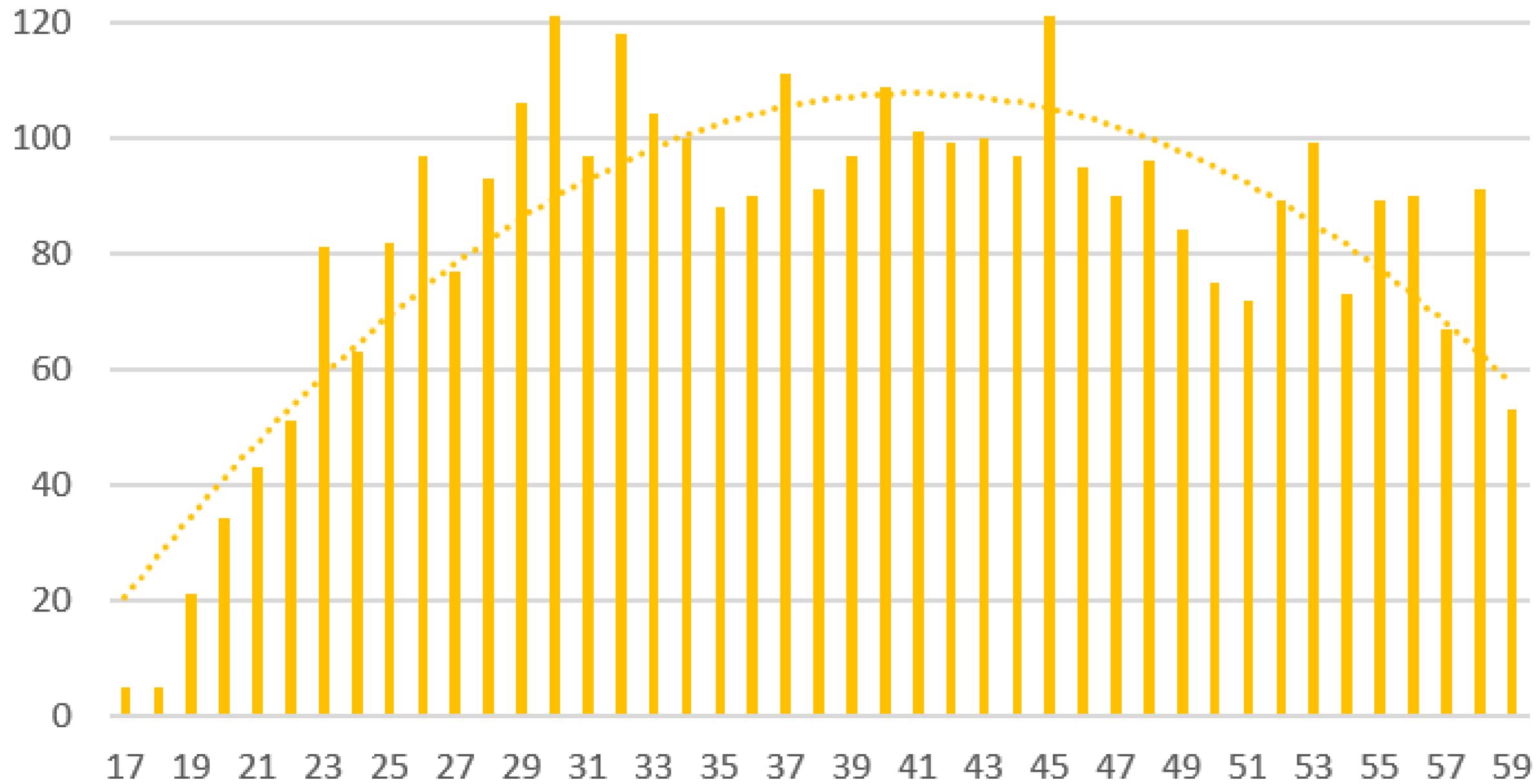
**Дарья  
Покацкая**

Сортировка БД



**Денис  
Игнатов**

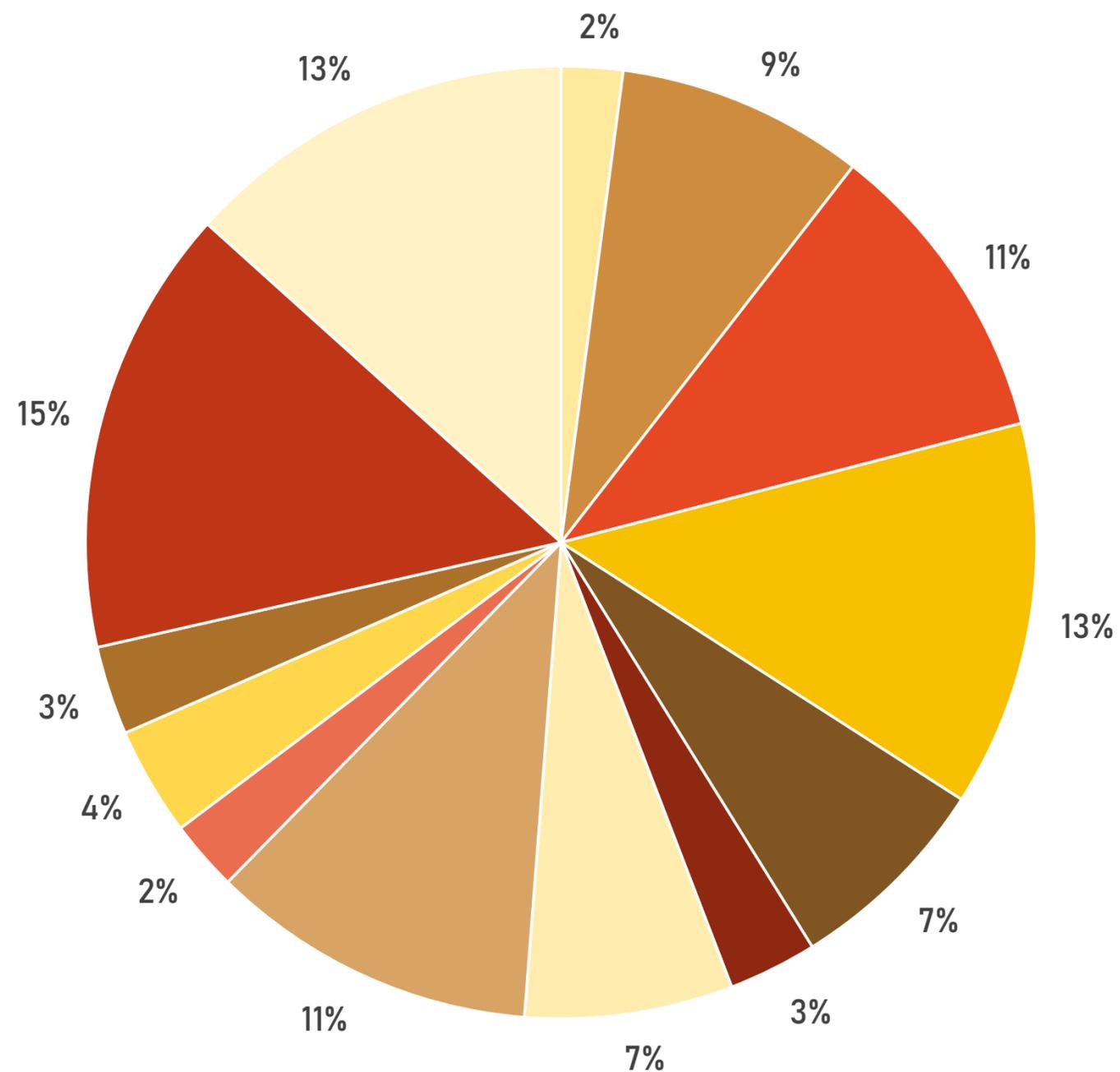
Программист



- Торговля
- Сельское хозяйство
- образование,
- наука
- Промышленность
- Здравоохранение
- ВПК
- Строительство
- Нефть

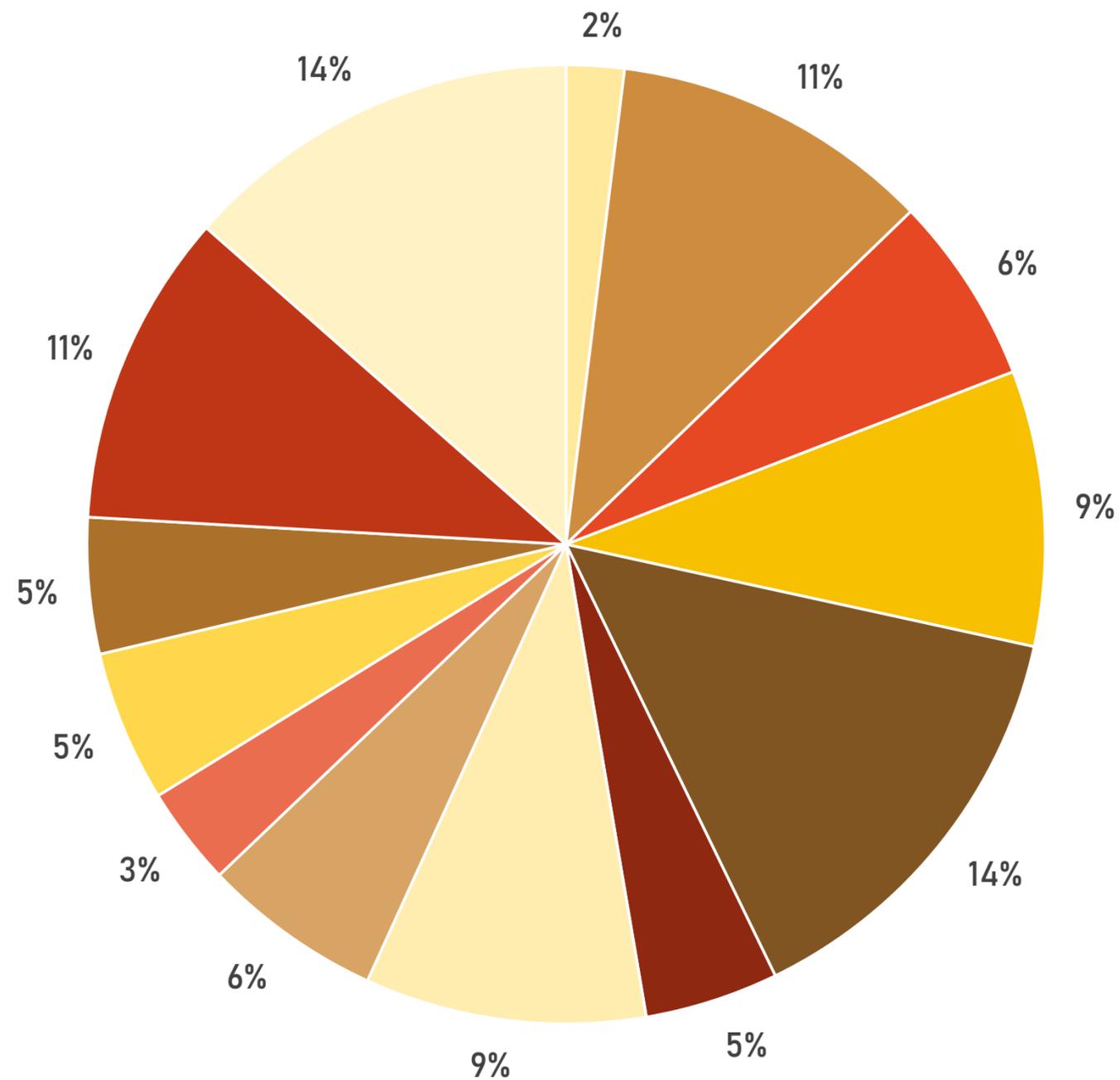


## Село



- ЖКХ
- Промышленность
- Образование и наука
- Здравоохранение
- Торговля и финансы
- ВПК
- Строительство
- Управление
- Добывающая пром-ть
- Легкая пром-ть
- Нефть, энергетика
- Образ., наука., здравоохранение
- С/х

Город



- ЖКХ
- Промышленность
- Образование и наука
- Здравоохранение
- Торговля и финансы
- ВПК
- Строительство
- Управление

$W = 0.86271$

$W = 0.87359$

data: df\$lnwage2

$W = 0.947$

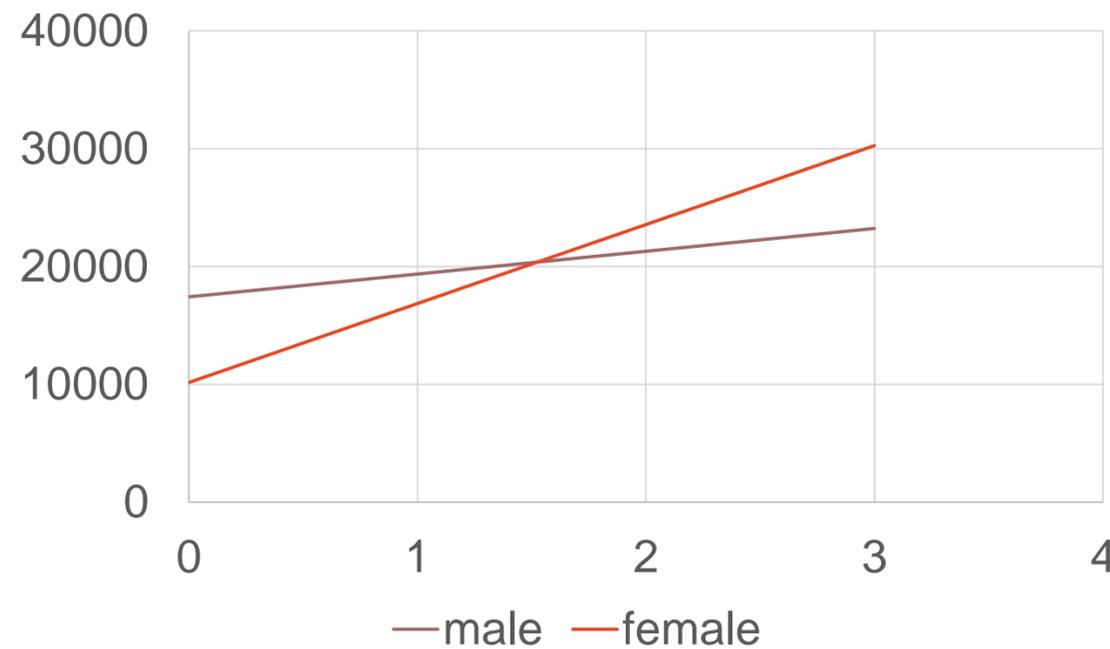
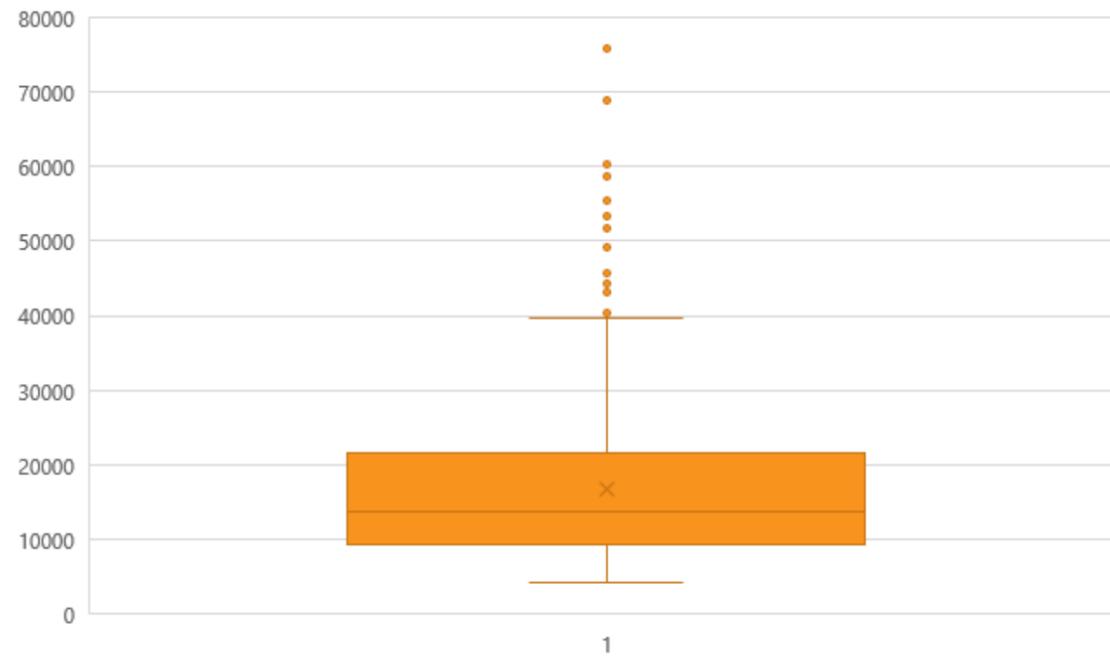
data: df\$aeugh

$W = 0.95857$

Тест Шапиро-Уилка

$p\text{-value} > 0,05$

### С выбросами



### Без выбросов

