



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Data Analysis National Olympiad - DANO

ТИПЫ ДАННЫХ. КАКИМИ ОНИ БЫВАЮТ?

Елена Семерикова,
Валерия Стефаненко

Москва, 2021



ВРЕМЕННОЙ РЯД

(time series)



Цена на акции ПАО Газпрома	
Дата	Цена, \$
Jul-01-2021	283,60
Jul-02-2021	285,45
Jul-05-2021	288,88
Jul-06-2021	295,60
Jul-07-2021	296,60
Jul-08-2021	291,74
Jul-09-2021	294,35

Показатели по России			
Год	ВВП, млрд	Зарубежные инвестиции, % от ВВП	Инфляция, % (GDP deflator)
2016	1 276	2,5	2,8
2017	1 574	1,8	5,3
2018	1 657	0,5	9,99
2019	1 687	1,89	3,08
2020	1 483	-	0,9

Виды

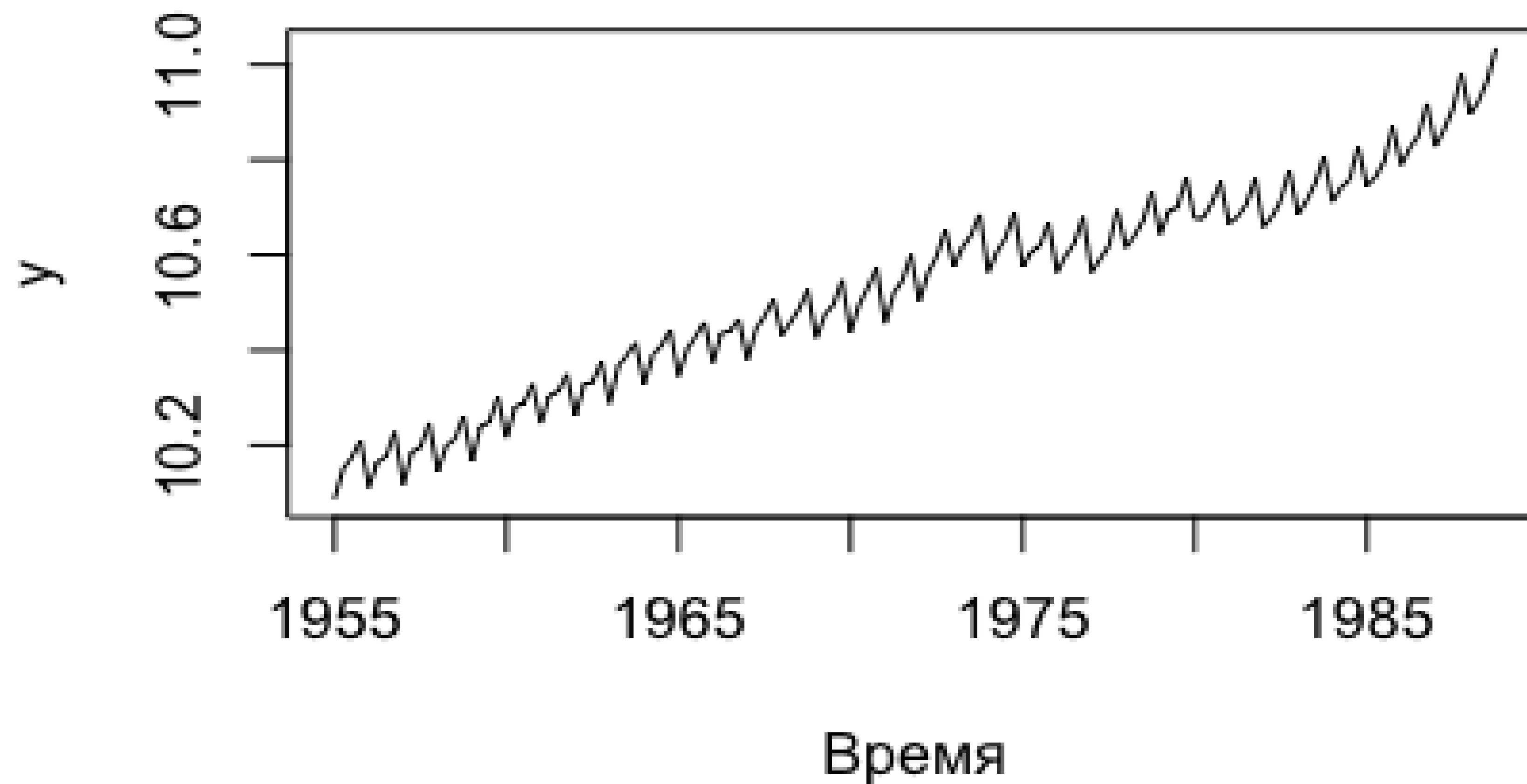
Одномерные
Многомерные

Низкочастотные
Высокочастотные

ВРЕМЕННОЙ РЯД

(time series)

- ✓ Тренд
- ✓ Сезонность
- ✓ Цикличность

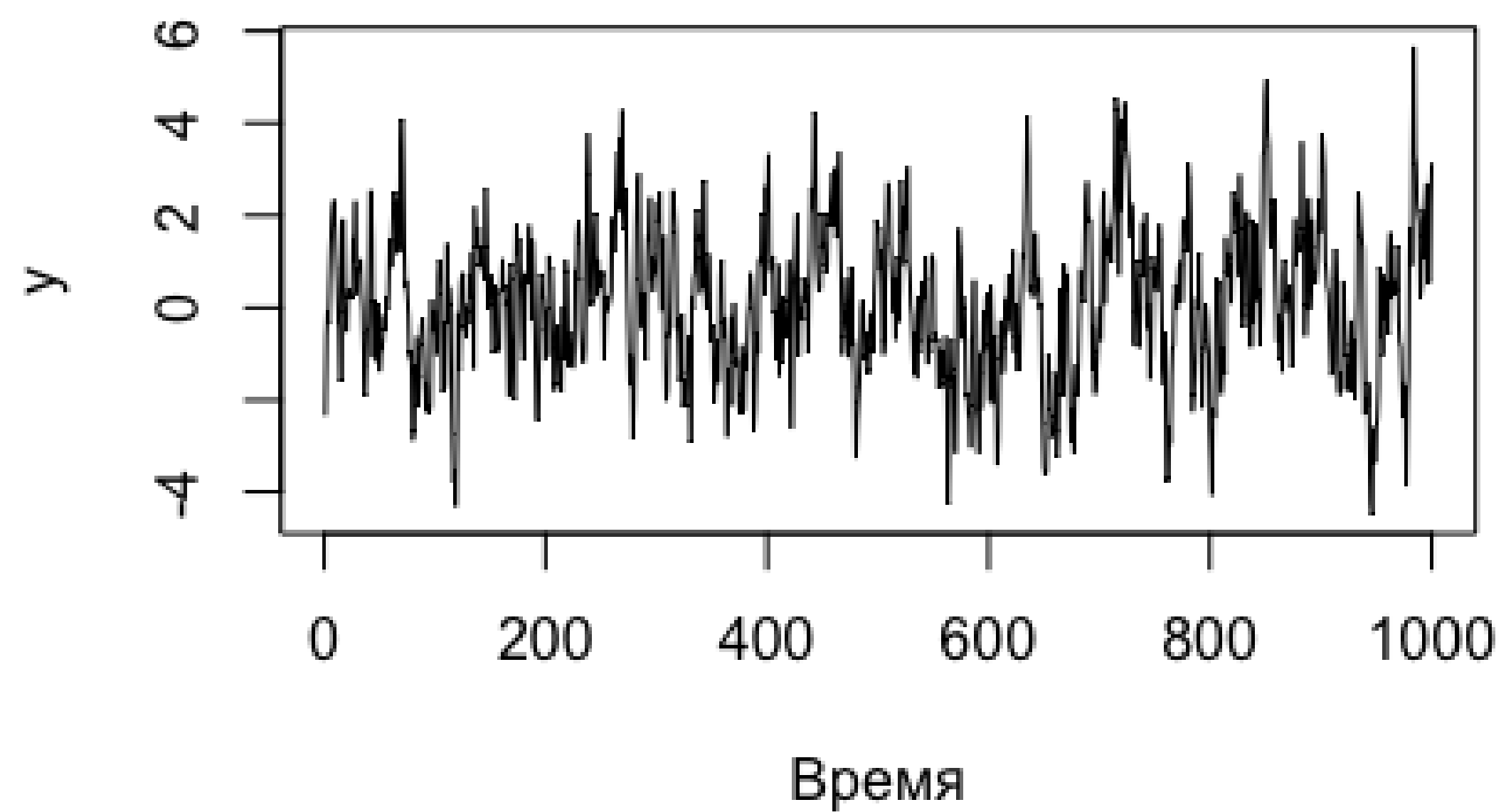




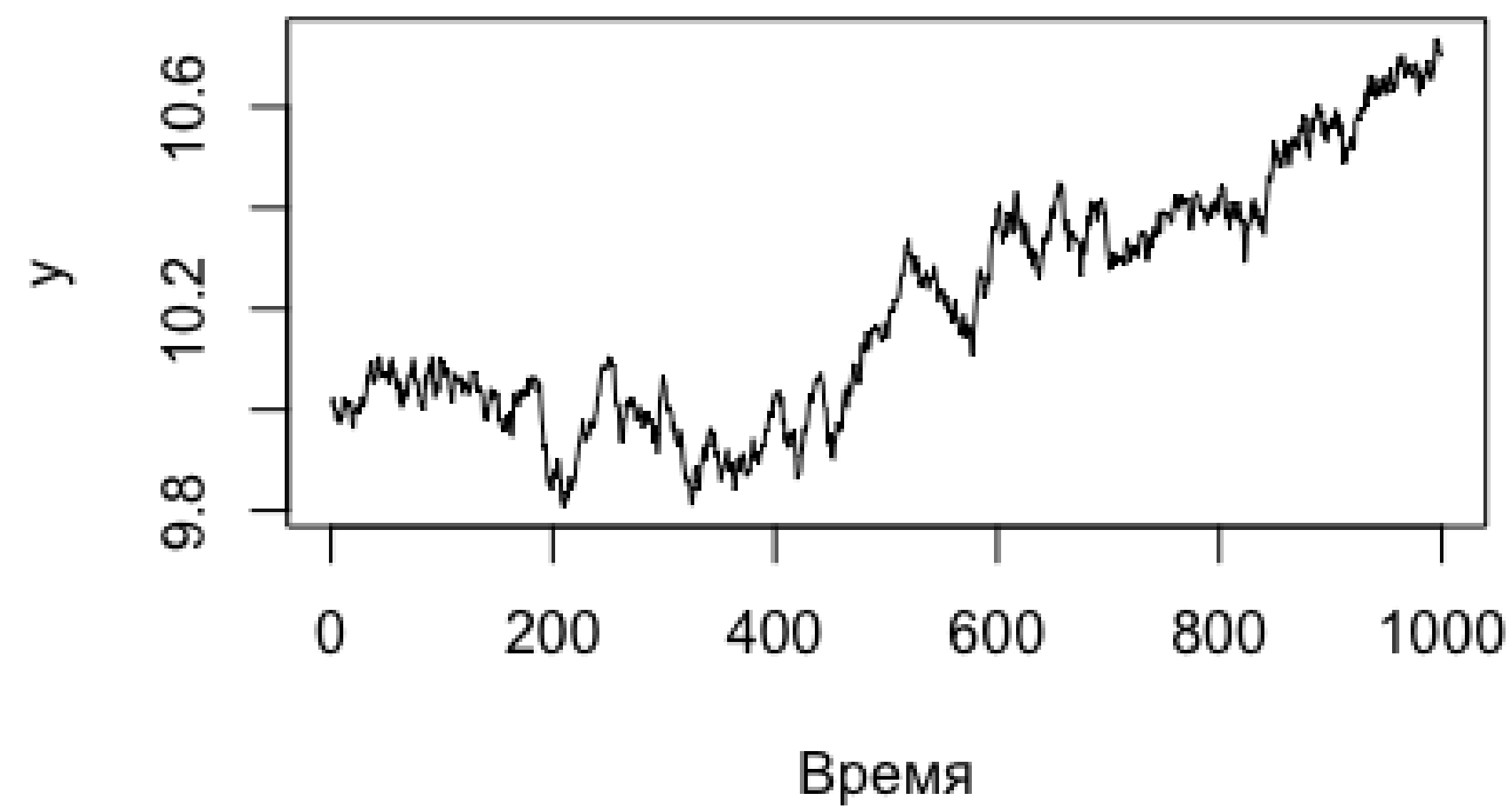
ВРЕМЕННОЙ РЯД

(time series)

СТАЦИОНАРНЫЙ



НЕСТАЦИОНАРНЫЙ



ПЕРЕКРЕСТНАЯ ВЫБОРКА

(cross-section data)



Регионы России	Выявлено случаев COVID-19	Новые за сутки	Выздоровело
Москва	1 490 707	2 629	1 336 194
Московская область	377 445	2 161	305 330
Санкт-Петербург	519 129	1 929	493 848
Ленинградская область	54 878	244	46 791
Краснодарский край	57 738	257	46 881



+



=

**ПАНЕЛЬНЫЕ
ДААННЫЕ**
(panel data)

**ВРЕМЕННОЙ
РЯД**
(time series)

**ПЕРЕКРЕСТНАЯ
ВЫБОРКА**
(cross-section data)

ПАНЕЛЬНЫЕ ДАнные

(panel data)



Пример 1

Покупатель	Оценка качества услуг
1	Хорошо
2	Затрудняюсь ответить
3	Отлично
4	Отлично
5	Можно лучше

Пример 2

Покупатель	Месяц	Оценка качества услуг
1	Январь	Хорошо
2	Январь	Затрудняюсь ответить
3	Январь	Отлично
1	Февраль	Отлично
2	Февраль	Можно лучше
3	Февраль	Хорошо



ПАНЕЛЬНЫЕ ДАННЫЕ

(panel data)

Адрес магазина	Покупатель	Месяц	Оценка качества услуг
Калужская	1	Январь	Хорошо
Строгино	2	Январь	Затрудняюсь ответить
Арбат	3	Январь	Отлично
Строгино	1	Февраль	Отлично
Строгино	2	Февраль	Можно лучше
Калужская	3	Февраль	Хорошо

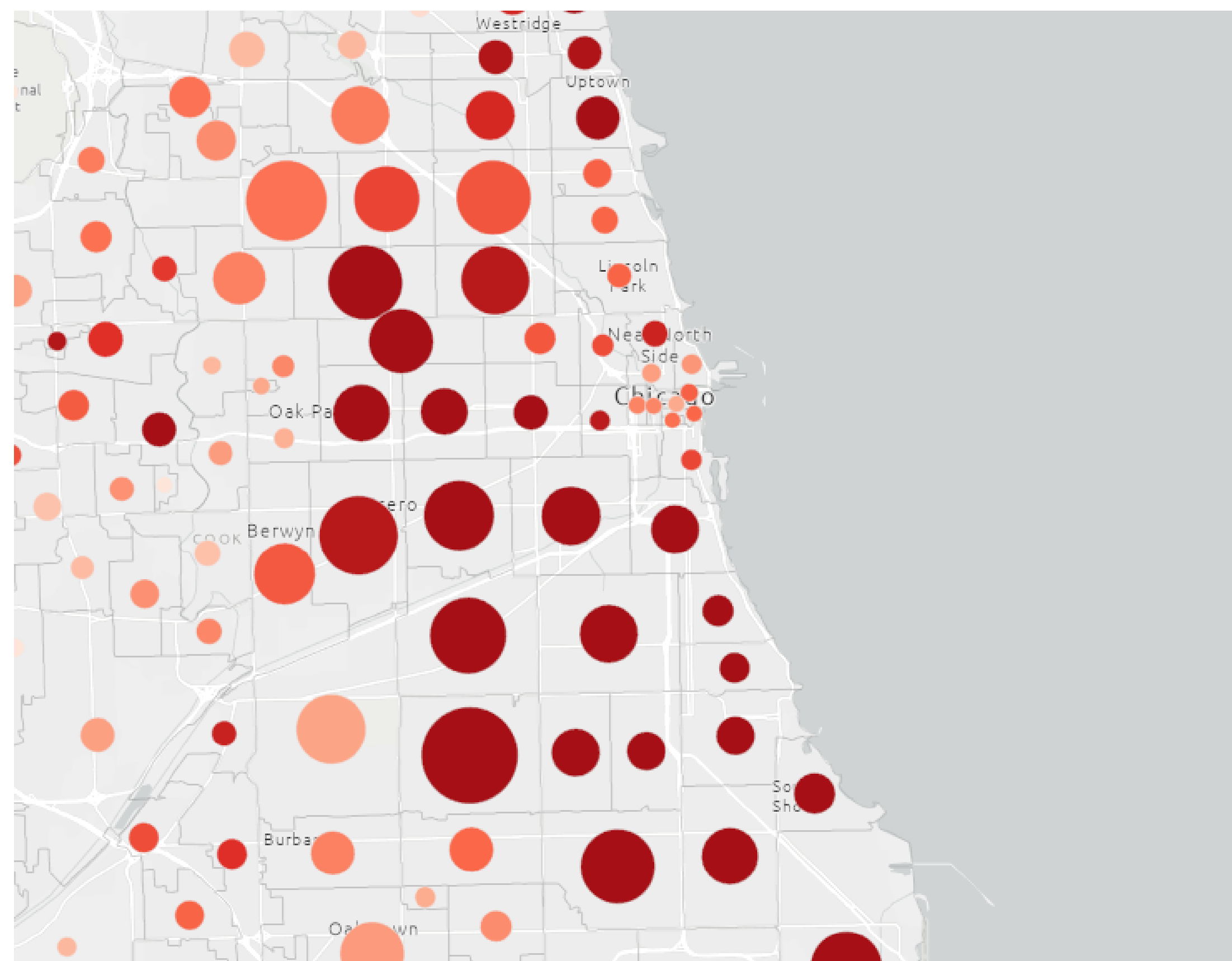
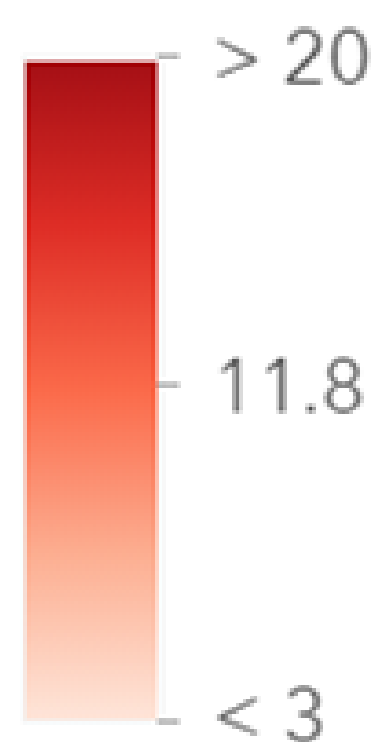
ПРОСТРАНСТВЕННЫЕ ДАННЫЕ

(panel data)

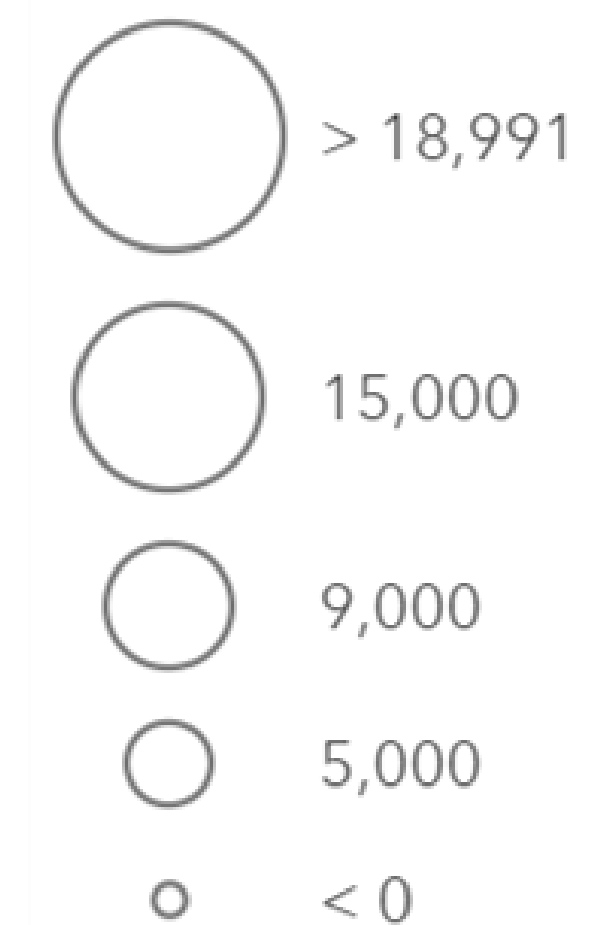
Legend

Poverty and Education

2009-2013 ACS Percent of Households with Income Below Poverty Level



High School Diploma



ПАНЕЛЬНЫЕ ДААННЫЕ в России

(panel data)



Мониторинг социально-экономического положения и состояния здоровья населения Российской Федерации (RLMS-HSE)

- проводится с 1994 года
- информация о структуре доходов, материальном благосостоянии, инвестиционном поведении, структуре занятости, миграции, состоянии здоровья и структуре питания, о планировании семьи и образовании и т.д.



ПОЛЕЗНЫЕ ССЫЛКИ

для изучения

Временные ряды. МФТИ

<https://www.coursera.org/lecture/data-analysis-applications/vriemiennyye-riady-EjNEV>

Ратникова Т.А. Введение в эконометрический анализ панельных данных. Лекция 1/ Экономический журнал НИУ ВШЭ, №2, 2006 http://ecsocman.hse.ru/data/2010/11/16/1214796647/10_02_06.pdf

Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. Изд. 8. М., Дело, 2008

<http://math.isu.ru/ru/chairs/me/files/books/magnus.pdf>

Лонгитюдное обследование домохозяйств РМЭЗ НИУ ВШЭ

<https://www.hse.ru/rlms/>



ЗАДАНИЯ

для самопроверки

Предположим, вы делаете опрос среди своих одноклассников о том, сколько времени они потратили на подготовку к ЕГЭ и какой балл получили. Определите тип данных.

Предположим, Ваша задача определить, влияет ли наличие репетитора на успехи в учебе по математике. Каждый год вы записываете оценки по математике у своих одноклассников и наличие занятий с репетитором. Определите тип данных.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ