

# Анализ загрязнений окружающей среды в России

НИУ ВШЭ

30 ноября 2019

- 1 Поиск и обработка данных
- 2 Описание исследования
- 3 Подготовка
- 4 Предварительный анализ данных
- 5 Тестирование гипотез

# Поиск и обработка данных

## Данные

- Источник: Росстат и Росреестр
- 22 показателя
- 10 лет
- 82 региона

Out[48]:

	reg_id	year	population	revealed_pollution	spending	GRP_pc	water_pollution	manufactory	income	urban	ter_agricultun
int	738.000000	738.000000	648.000000	642.000000	410.000000	5.660000e+02	648.000000	1.640000e+02	646.000000	488.000000	711.000000
an	41.500000	2014.000000	1789.245370	58.332243	6657.223463	3.146248e+05	191.451235	4.493917e+05	23002.425697	69.693648	4644.353161
std	23.685653	2.58374	1766.458613	25.849043	<a href="#">9132.935136</a>	2.443094e+05	283.695671	7.120274e+05	8909.644454	13.232741	7428.956981
min	1.000000	2010.000000	50.000000	0.200000	54.900000	4.823920e+04	0.000000	3.700000e+02	7774.000000	27.700000	150.900000
5%	21.000000	2012.000000	787.500000	39.200000	1293.590000	1.818318e+05	37.000000	7.751100e+04	16892.750000	63.950000	1458.200000
7%	41.500000	2014.000000	1217.000000	62.400000	2977.355000	2.518391e+05	87.000000	1.915315e+05	21669.000000	70.800000	2314.500000
5%	62.000000	2016.000000	2422.500000	81.200000	8000.380000	3.562442e+05	216.500000	5.666228e+05	26067.250000	77.825000	4344.600000
max	82.000000	2018.000000	12507.000000	98.100000	59830.290000	1.716734e+06	3578.000000	5.589313e+06	70904.000000	100.000000	39760.900000

Целью исследования является выявить факторы, влияющие на экономическую обстановку в регионах. За целевую переменную брался уровень загрязнения воздуха за определенный год (pollution\_static).

## Гипотезы

- Деятельность государства по снижению загрязнения воздуха в регионах неэффективна
- Загрязнение воздуха зависит от экономических показателей и структуры регионов
- Характеристики населения (население, средний располагаемый доход и др.) влияют на экологические показатели

- Собрали панельные данные (22 переменные 820 наблюдений)
- Ограничили количество регионов
- Присвоили id каждому наблюдению (по региону и округу)
- Отобрали 22 признака для каждого наблюдения

Проблемы:

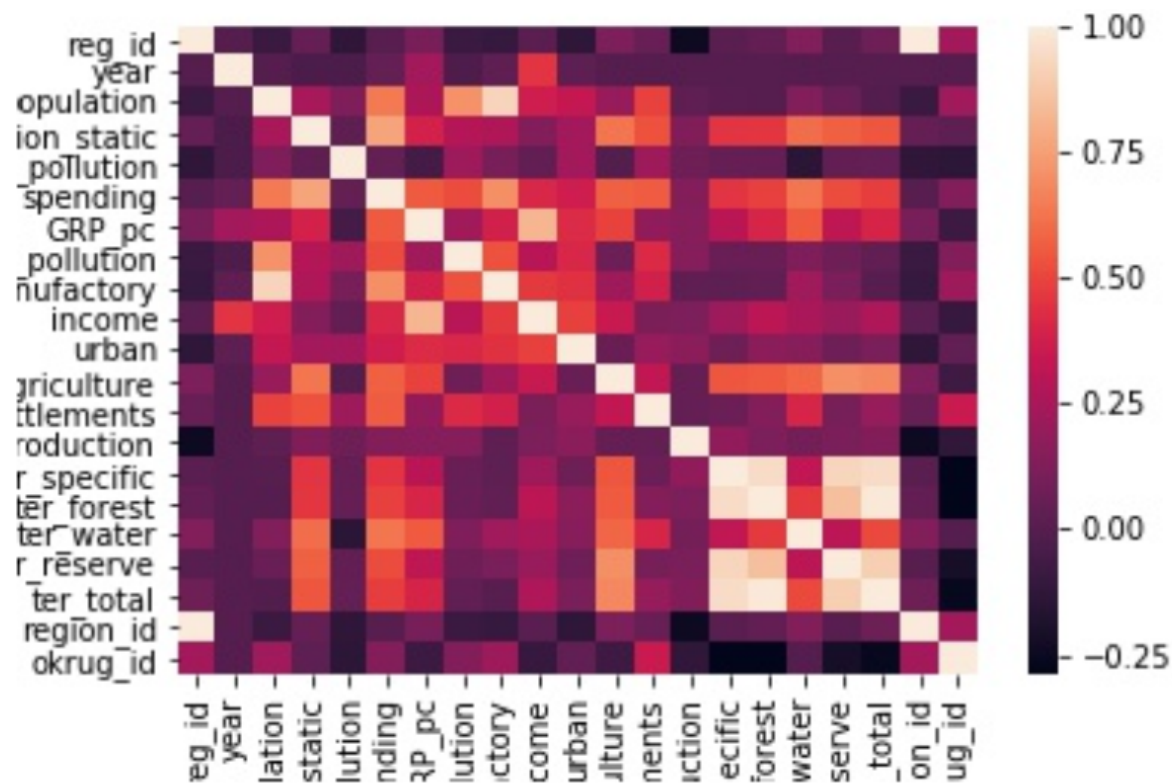
- 1 Разные качество данных для разных регионов.
- 2 Учитывали разнообразие регионов (нормирование показателей)

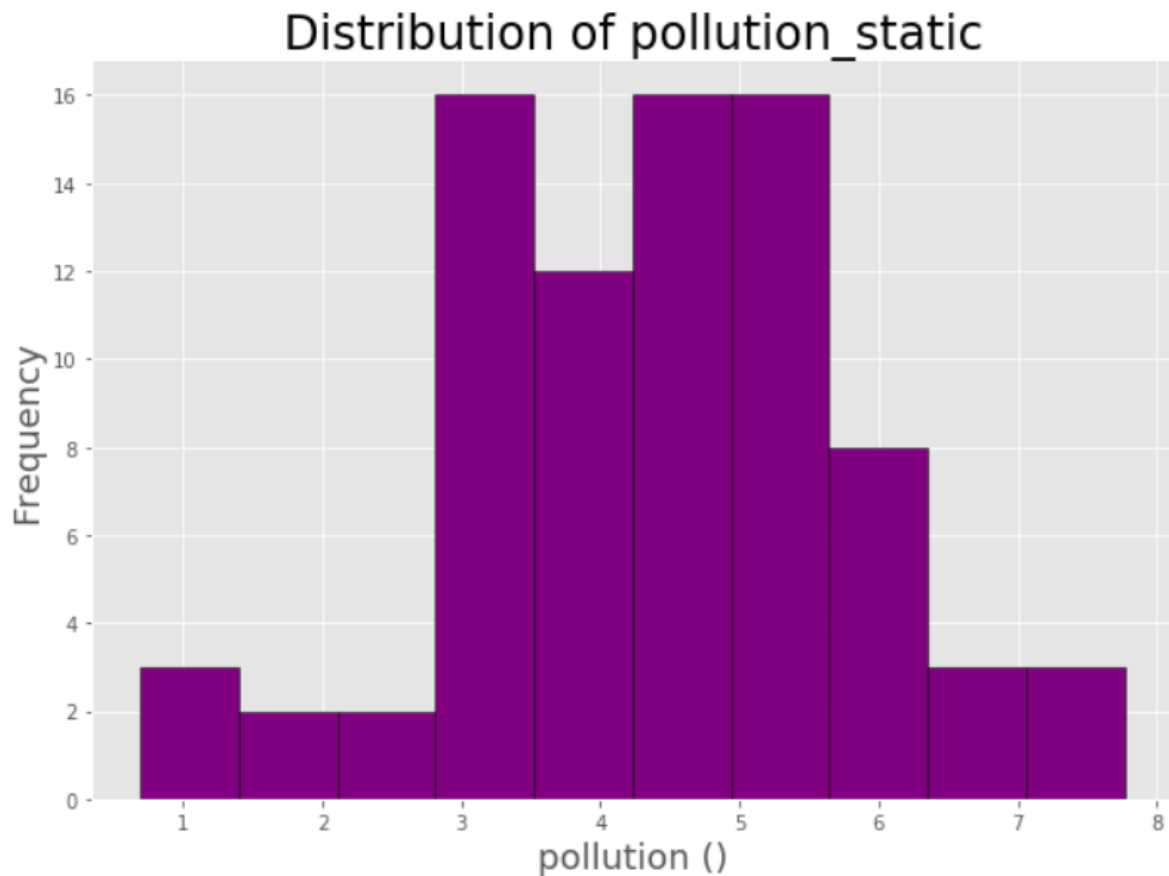
Основная проблема - различие названий регионов и их индексов на сайте Росстата (нет единого стандарта оформления). Для приведения названий регионов к единому виду использовалось расстояние Левенштейна.

```
In [6]: for reg1 in maindata['region'].unique():
        for reg2 in maindata['region'].unique():
            if reg1 > reg2:
                if jellyfish.levenshtein_distance(reg1, reg2) < 4:
                    print(reg1, reg2)
```

Вологодская область Волгоградская область

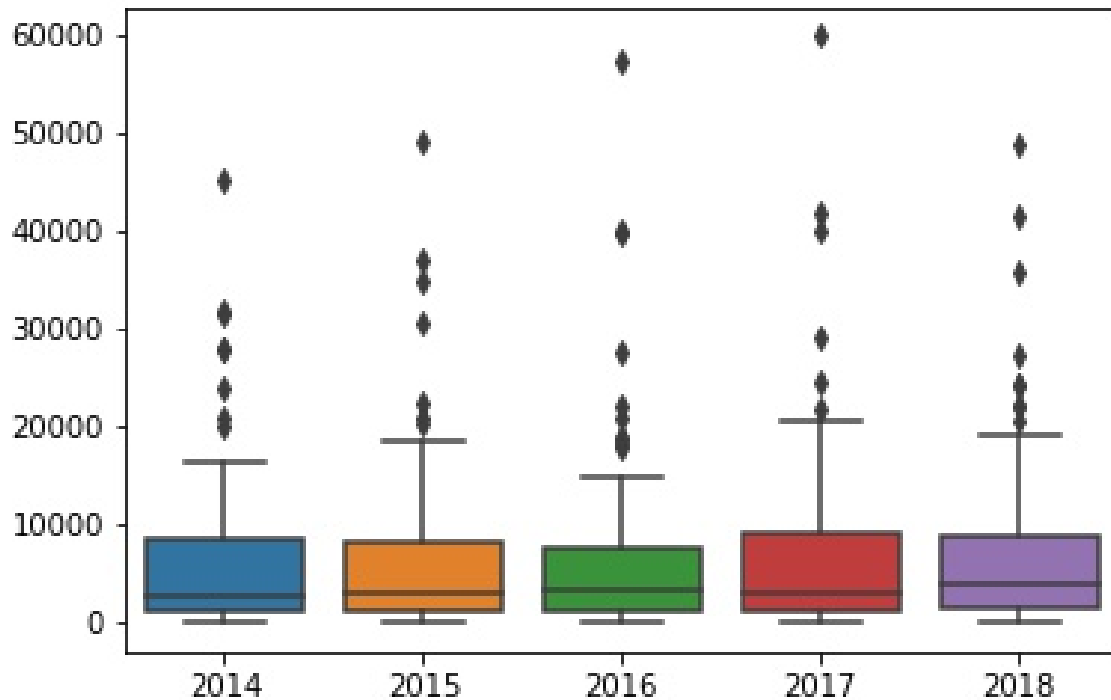
# Предварительный анализ данных







# Предварительный анализ данных



# Тестирование гипотез

## Гипотеза 1

	<i>Dependent variable:</i>
	log(pollution_static)
log(spendlagged)	0.929*** (0.206)
log(revpollagged)	0.167 (-0.544)
shareofproduction	1.016 (-2.741)
log(GRP_pc)	0.056 (-0.951)
Крымский федеральный округ	-0.329 (-1.003)
Приволжский федеральный округ	-0.170 (-0.891)
Северо-Западный федеральный округ	0.126 (-0.440)
Сибирский федеральный округ	0.474 (-0.617)
Уральский федеральный округ	0.606 (-0.844)
Центральный федеральный округ	-0.096 (-0.761)
Южный федеральный округ	-0.444 (-0.932)
Constant	-4.364 (13.349)
Observations	157
R <sup>2</sup>	0.836
Adjusted R <sup>2</sup>	0.823
F Statistic	66.968*** (df = 11; 145)

- Существенно эффекта от деятельности гос-ва нет. Государство тратит на те регионы, где ситуация заметно хуже.
- Эффекта от идентификация выбросов нет. Регион также не значен.
- Т. к. данных ограниченное количество, то возможно мы не в состоянии идентифицировать никакого эффекта.

## Переменные

- млн.руб затрат на охрану окр.среды (spend lagged)
- процент зафиксированных выбросов (revpollagged)
- GRP- валовый региональный продукт per capita
- Зависимая переменная - млн.тонн выбросов (log).

# Гипотеза 2

Гипотеза подтверждается, в регионе положительно влияет на уменьшение выбросов качество городской среды. Ошибки робастные - то есть коэффициент значим.

```


```

	Estimate	Std. Error	t-value	Pr(> t )	
(Intercept)	-10.4129139	1.0888904	-9.5629	< 2.2e-16	***
log(roads)	-0.4269226	0.0444877	-9.5964	< 2.2e-16	***
log(population)	1.2763176	0.0546848	23.3395	< 2.2e-16	***
log(GRP_pc)	0.5700155	0.0875076	6.5139	2.325e-10	***
shareofproduction	1.2115117	1.8317475	0.6614	0.508759	.
okrugкрымский федеральный округ	-0.4319535	0.4398733	-0.9820	0.326727	.
okrugПриволжский федеральный округ	0.3294818	0.1744656	1.8885	0.059718	.
okrugСеверо-Западный федеральный округ	0.8046813	0.1623135	4.9576	1.078e-06	***
okrugСибирский федеральный округ	0.7130898	0.1556192	4.5823	6.247e-06	***
okrugуральский федеральный округ	0.6836319	0.2028239	3.3706	0.000827	***
okrugЦентральный федеральный округ	0.3357180	0.1758222	1.9094	0.056961	.
okrugЮжный федеральный округ	-0.4236463	0.1960386	-2.1610	0.031317	*
urban	0.0055687	0.0041761	1.3335	0.183179	.

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    897.84
Residual Sum of Squares: 173.21
R-Squared:               0.80708
Adj. R-Squared: 0.80099
F-statistic: 132.48 on 12 and 380 DF, p-value: < 2.22e-16

```

# Гипотеза 3

Гипотеза про доход подтверждается. Увеличение людей с большим доходом влияет на качество среды.

```

                                Estimate Std. Error t-value Pr(>|t|)
(Intercept)                   -11.420224    1.165375  -9.7996 < 2.2e-16 ***
log(population)                 1.148827    0.044946  25.5603 < 2.2e-16 ***
log(GRP_pc)                    1.660359    0.129790  12.7927 < 2.2e-16 ***
shareofproduction              0.907809    1.683350   0.5393  0.589913
okrugКрымский федеральный округ -1.500775    0.546342  -2.7470  0.006218 **
okrugПриволжский федеральный округ -0.617044    0.147443  -4.1850  3.333e-05 ***
okrugСеверо-Западный федеральный округ 0.070246    0.142415   0.4933  0.622038
okrugСибирский федеральный округ  0.295284    0.146559   2.0148  0.044426 *
okrugУральский федеральный округ  0.106558    0.186782   0.5705  0.568580
okrugЦентральный федеральный округ -0.697034    0.135089  -5.1598  3.487e-07 ***
okrugЮжный федеральный округ    -1.317899    0.154547  -8.5275 < 2.2e-16 ***
log(income)                   -1.265105    0.203622  -6.2130  1.044e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1301.1
Residual Sum of Squares: 297.3
R-Squared:              0.7715
Adj. R-Squared: 0.76681
F-statistic: 164.517 on 11 and 536 DF, p-value: < 2.22e-16
```