Университет ИТМО

Практическая работа $N_{2}5$

по дисциплине «Визуализация и моделирование»

Автор: Власов Матвей Иванович

Поток: ВИМ 1.1 Группа: К3240 Факультет: ИКТ

Преподаватель: Чернышева А. В.

Датасет

Для дальнейшей работы выбран датасет: "Russian Presidental Elections 2018 Voting Data": https://www.kaggle.com/valenzione/russian-presidental-elections-2018-voting-data

Описание датасета

В нашем датасете содержится информация об итогах выборов 2018 года, полученная с официального сайта ЦИК РФ.

Названия большинства стобцов исходного датасета представим в графе "Описание а сами названия сократим для удобства в дальнейшем:

Столбец	Описание	Тип	Шкала	Предобработка
PS_ID	Идентификатор избирательно- го участка	INT	Номинальная	Не требуется
REGION	Название региона	STRING	Номинальная	Убрать цифры в начале строки (если есть)
SUBREGION	Название округа	STRING	Номинальная	Убрать цифры в начале строки (если есть)
N_ALL	Число избирателей, включен- ных в список избирателей	INT	Относительная	Не требуется
N_GIVEN	Число избирательных бюллетеней, полученных участковой избирательной комиссией	INT	Относительная	Удалить по- сле подсчёта N_VOTED
N_EARLY	Число избирательных бюллетеней, выданных избирателям, проголосовавшим досрочно	INT	Относительная	Не требуется
N_IN	Число избирательных бюллетеней, выданных в помещении для голосования в день голосования	INT	Относительная	Удалить (избы- точные данные)
N_OUT	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования	INT	Относительная	Не требуется
N_LEFT	Число погашенных избиратель- ных бюллетеней	INT	Относительная	Удалить по- сле подсчёта N_VOTED
N_PORTABLE	Число избирательных бюллетеней в переносных ящиках для голосования	INT	Относительная	Удалить (избы- точные данные)
N_STATIC	Число бюллетеней в стационар- ных ящиках для голосования	INT	Относительная	Удалить (избы- точные данные)
N_INVALID	Число недействительных избирательных бюллетеней	INT	Относительная	Не требуется
N_VALID	Число действительных избира- тельных бюллетеней	INT	Относительная	Удалить (избы- точные данные)
N_LOST	Число утраченных избиратель- ных бюллетеней	INT	Относительная	Удалить (незначитель- ные данные)
N_UNUSED	Число избирательных бюллетеней, не учтенных при палучении	INT	Относительная	Удалить (незначитель- ные данные)

Столбец	Описание	Тип	Шкала	Предобработка
BABURIN	Бабурин Сергей Николаевич	INT	Относительная	Не требуется
GRUDININ	Грудинин Павел Николаевич	INT	Относительная	Не требуется
ZHIRINOVSKY	Жириновский Владимир Воль-	INT	Относительная	Не требуется
	фович			
PUTIN	Путин Владимир Владимиро-	INT	Относительная	Не требуется
	вич			
SOBCHAK	Собчак Ксения Анатольевна	INT	Относительная	Не требуется
SURAYKIN	Сурайкин Максим Александро-	INT	Относительная	Не требуется
	вич			
TITOV	Титов Борис Юрьевич	INT	Относительная	Не требуется
YAVLINSKY	Явлинский Григорий Алексее-	INT	Относительная	Не требуется
	вич			

Задачи, решаемые при помощи датасета

- 1. Визуализация результатов выборов.
- 2. Анализ данных на предмет возможных фальсификаций.
- 3. Выявление особенностей голосования в различных регионах.

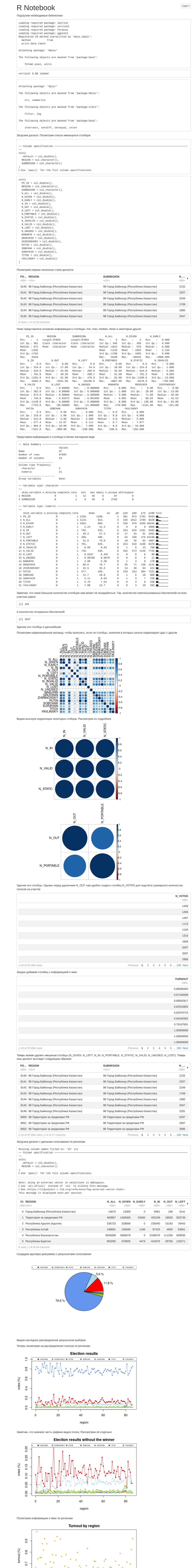
Гипотезы

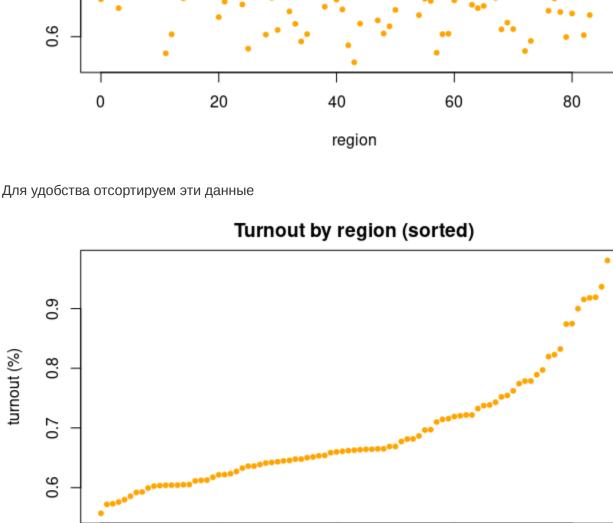
- 1. В Москве и Санкт-Петербурге ниже, чем в среднем, и процент за Путина, и явка (в больших городах более образованное население, а также большое количество наблюдателей, что затрудняет фальсификации).
- 2. В Крыму высокая явка и поддержка президента (из-за присоединения территории).
- 3. В регионах с большим количеством избирателей, проголосовавших досрочно, процент за Путина выше, чем в среднем (голоса, поданные досрочно, легче сфальсифицировать).
- 4. Есть регионы, где победил не Путин (у Грудинина в среднем больше 11 процентов вполне возможно, что где-то он набрал больше Путина).

Работа с датасетом

Обратите внимание, что часть таблиц отображается неполностью.

Для просмотра недостающей информации откройте файл с исходным кодом.





40

Видим, что основная явка - около 67 %, минимальная - чуть меньше 60 %, максимальная - почти 100 %

40

region

Histogram of turnout by polling station

region

Dependence of votes for Putin on turnout

60

60

Видим, что с увеличением явки увеличивается количество голосов за Путина. Отмечаем, что при явке меньше 70 % в большинстве

80

80

20

20

регионов средний процент за Путина меньше общероссийского

Построим диаграмму распределения явки по избирательным участкам

0.9

0.8

0.7

9.0

15000

25000

15000

5000

0

25

20

15

9

Diverging Bars Election turnout

Frequency

0.0

0.2

Видим, что на большинстве участков Путин получил около 75 %

0.4

0.6

Votes for Putin

Histogram of votes for Putin by region

8.0

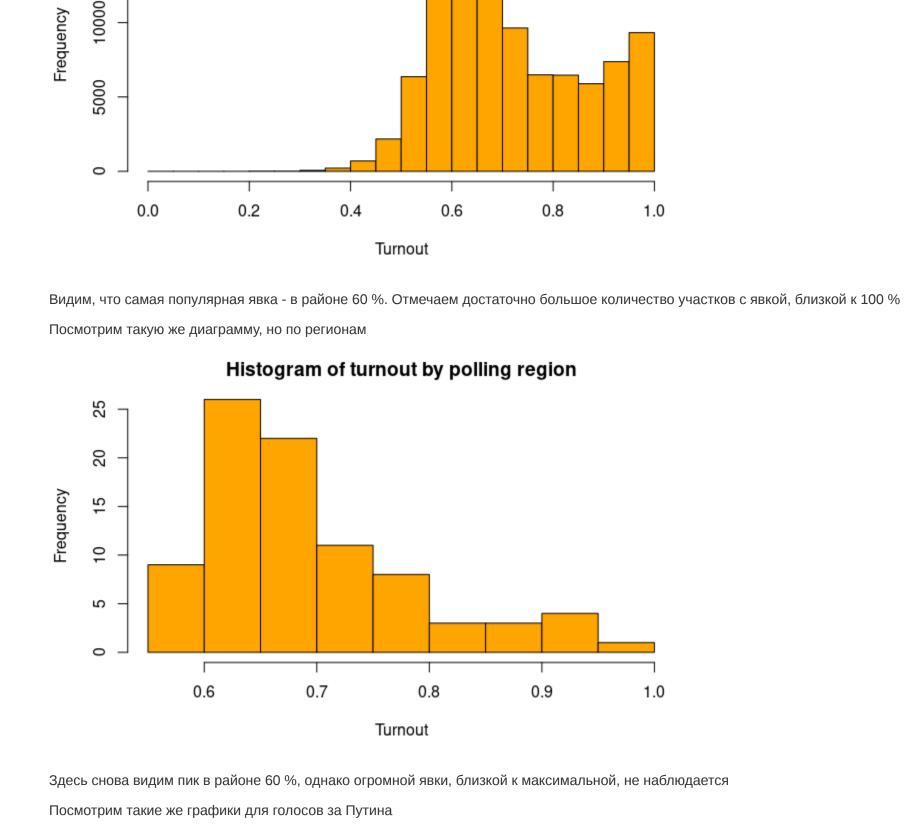
1.0

Frequency

0

turnout (%)

Посмотрим, как зависит процент голосов за Путина в зависимости от явки



Histogram of votes for Putin by polling station

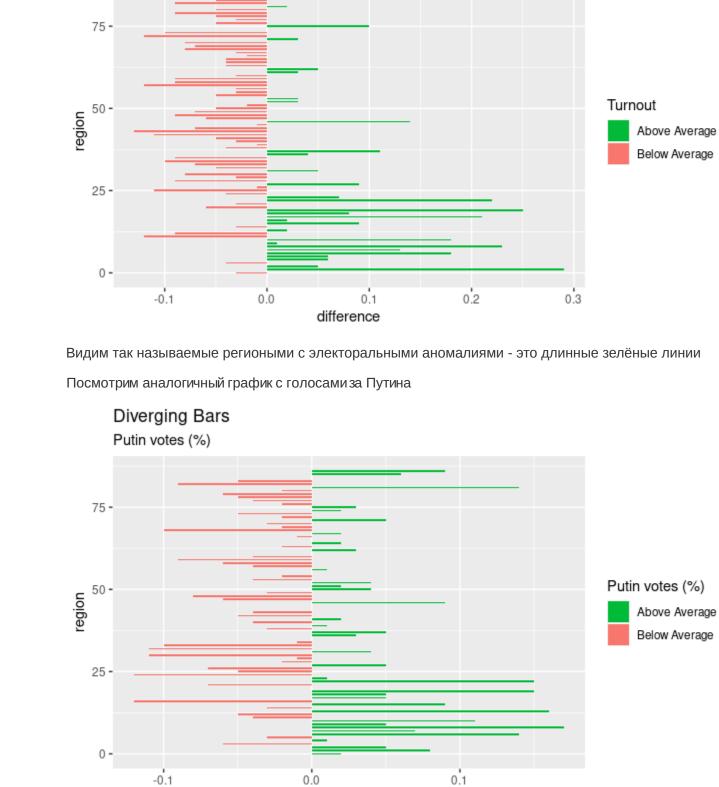
0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95

Votes for Putin

Посмотрим график с наглядным представлением разницы явки в каждом регионе от общероссийской

предыдущего, в то время как в этом графике выделяется последний столбец

В большинстве регионов - аналогично. Заметим, однако, что в предыдущем графике после пика каждый следующий столбец был меньше



difference

Здесь разница распределена более равномерно, колеблется в районе 10 %