

Университет ИТМО

Практическая работа №4
по дисциплине «Визуализация и моделирование»

Автор: Власов Матвей Иванович

Поток: ВИМ 1.1

Группа: К3240

Факультет: ИКТ

Преподаватель: Чернышева А. В.

Санкт-Петербург, 2021 г.

Датасет

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

Описание датасета

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

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

Столбец	Описание	Тип	Шкала	Предобработка
PS_ID	Идентификатор избирательного 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, включенных в список избирателей	INT	Относительная	Не требуется
N_GIVEN	Число избирательных бюллетеней, полученных участковой избирательной комиссией	INT	Относительная	Удалить после 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 (избыточные данные)
N_OUT	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования	INT	Относительная	Не требуется
N_LEFT	Число погашенных избирательных бюллетеней	INT	Относительная	Удалить после подсчёта N_VOTED
N_PORTABLE	Число избирательных бюллетеней в переносных ящиках для голосования	INT	Относительная	Удалить (избыточные данные)
N_STATIC	Число бюллетеней в стационарных ящиках для голосования	INT	Относительная	Удалить (избыточные данные)
N_INVALID	Число недействительных избирательных бюллетеней	INT	Относительная	Не требуется
N_VALID	Число действительных избирательных бюллетеней	INT	Относительная	Удалить (избыточные данные)
N_LOST	Число утраченных избирательных бюллетеней	INT	Относительная	Удалить (незначительные данные)
N_UNUSED	Число избирательных бюллетеней, не учтенных при получении	INT	Относительная	Удалить (незначительные 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 Вольфович	INT	Относительная	Не требуется
PUTIN	Путин Владимир Владимирович	INT	Относительная	Не требуется
SOBCHAK	Собчак Ксения Анатольевна	INT	Относительная	Не требуется
SURAYKIN	Сурайкин Максим Александрович	INT	Относительная	Не требуется
TITOV	Титов Борис Юрьевич	INT	Относительная	Не требуется
YAVLINSKY	Явлинский Григорий Алексеевич	INT	Относительная	Не требуется

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.

In [4]: Таблица с информацией о каждом избирательном участке

Out[4]:

	PS_ID	REGION	SUBREGION	N_ALL	N_EARLY	N_OUT	N_INVALID	BABURIN
0	8140	Город Байконур (Республика Казахстан)	Город Байконур (Республика Казахстан)	2132	0	11	9	4
1	8141	Город Байконур (Республика Казахстан)	Город Байконур (Республика Казахстан)	2207	0	14	14	2
2	8142	Город Байконур (Республика Казахстан)	Город Байконур (Республика Казахстан)	2249	0	7	27	5
3	8143	Город Байконур (Республика Казахстан)	Город Байконур (Республика Казахстан)	1769	0	48	20	5
4	8144	Город Байконур (Республика Казахстан)	Город Байконур (Республика Казахстан)	1880	0	13	10	7
...
97689	1310	Ямало-Ненецкий автономный округ	Ямальская	2892	1000	0	0	0
97690	1311	Ямало-Ненецкий автономный округ	Ямальская	2867	921	0	0	0
97691	1312	Ямало-Ненецкий автономный округ	Ямальская	2895	927	0	0	1
97692	1313	Ямало-Ненецкий автономный округ	Ямальская	3397	1103	0	0	0
97693	1314	Ямало-Ненецкий автономный округ	Ямальская	2666	1022	0	0	0

97694 rows × 17 columns

In [5]: Таблица с информацией о каждом регионе

Out[5]:

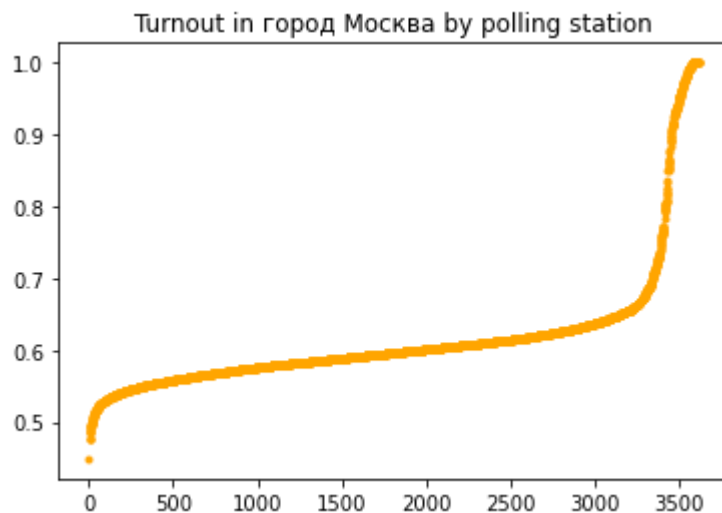
	REGION	N_ALL	N_EARLY	N_OUT	N_INVALID	BABURIN	GRUDININ	ZHIRINOV
0	Город Байконур (Республика Казахстан)	14575	0	168	104	32	1026	
1	Территория за пределами РФ	483957	53482	18026	5801	2010	23871	

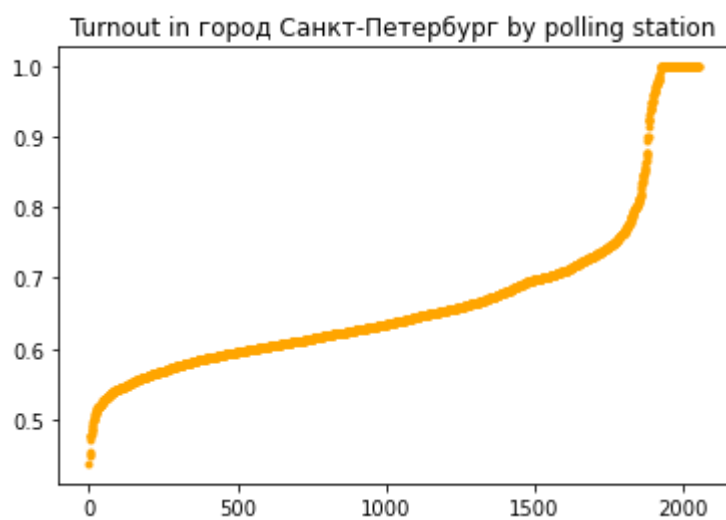
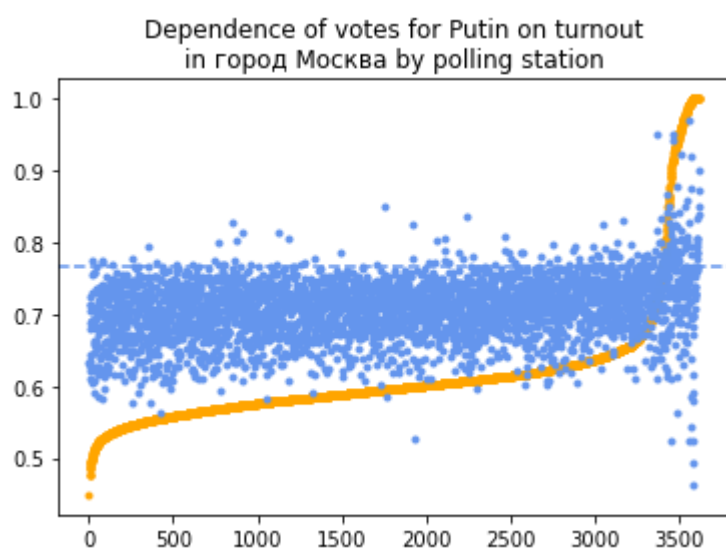
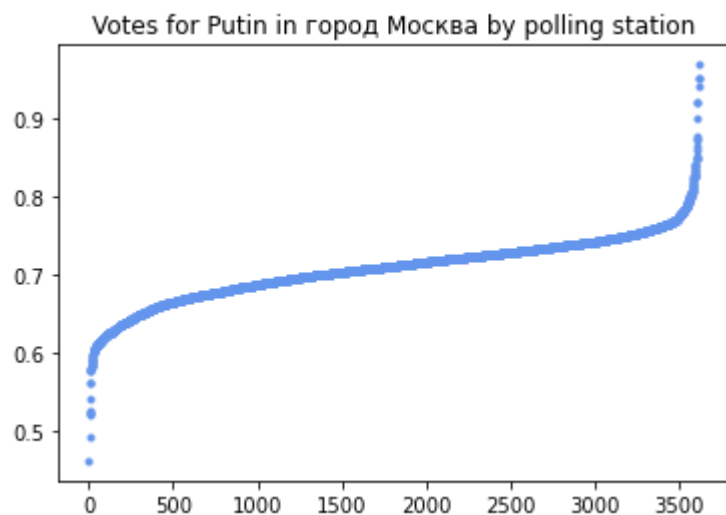
2	Республика Адыгея (Адыгея)	336720	0	15183	2647	1165	28711	
3	Республика Алтай	158891	1186	4630	974	446	21259	
4	Республика Башкортостан	3045698	0	111258	18506	15845	277797	11
...	
82	Еврейская автономная область	128844	106	4123	1331	501	14066	
83	Ненецкий автономный округ	39470	6116	579	267	185	3397	
84	Ханты-Мансийский автономный округ - Югра	1130343	26633	15804	10824	3916	94785	5
85	Чукотский автономный округ	33541	2531	874	293	115	1616	
86	Ямало-Ненецкий автономный округ	370823	41135	4008	2616	1052	19488	1

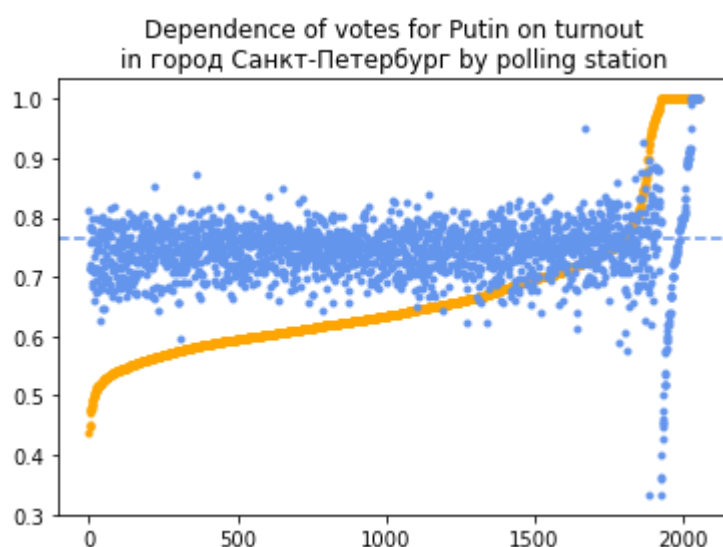
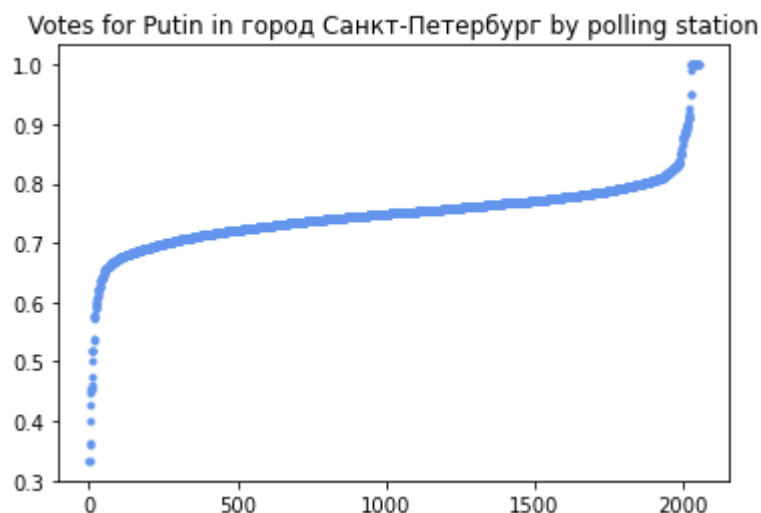
87 rows × 15 columns

In [7]: Посмотрим, как голосовали крупнейшие города - Москва и Санкт-Петербург

Note that the data about this region is incomplete

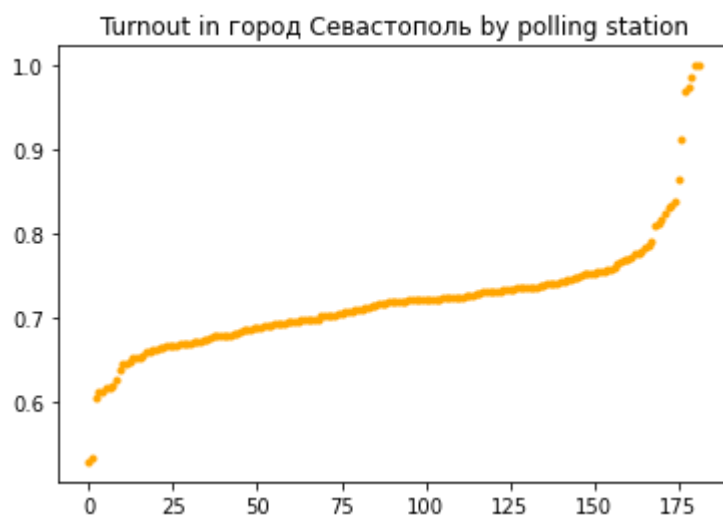




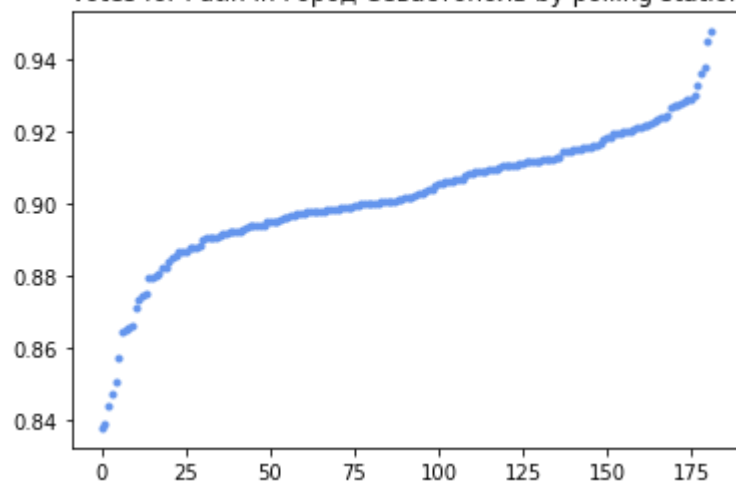


Видим, что явка на графиках находится на уровне 60 процентов, что ниже среднего по стране. Тем не менее тяжело не заметить большое количество участков со стопроцентной явкой (особенно в Санкт-Петербурге), что вызывает некоторые подозрения. За Путина в столицах голосовали менее активно, чем в регионах. Это заметно особенно в Москве, где на подавляющем большинстве участков рейтинг Путина не превысил средний по стране.

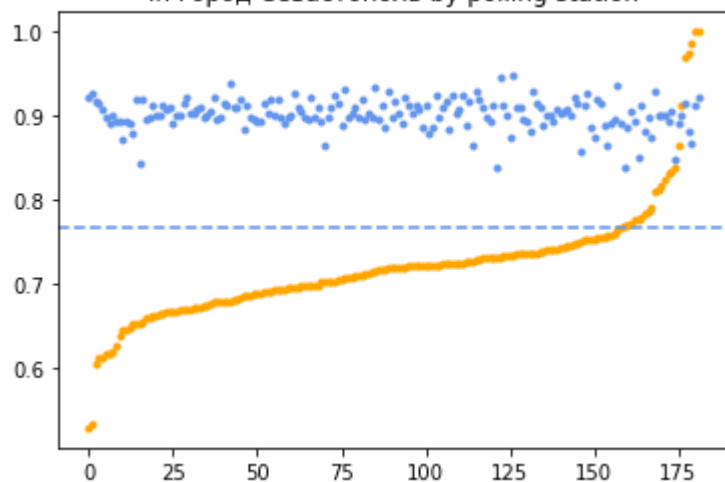
In [8]: Посмотрим, как прогосовали жители недавно присоединённой территории. Предполагаем, что поддержка власти там находится на высоком уровне



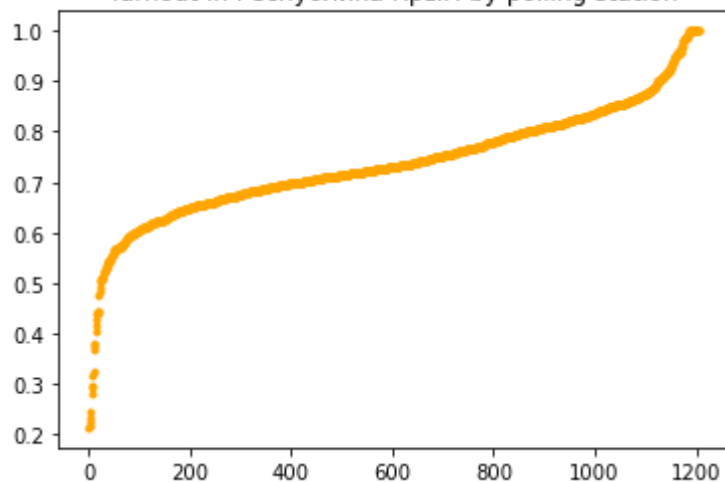
Votes for Putin in город Севастополь by polling station



Dependence of votes for Putin on turnout in город Севастополь by polling station



Turnout in Республика Крым by polling station

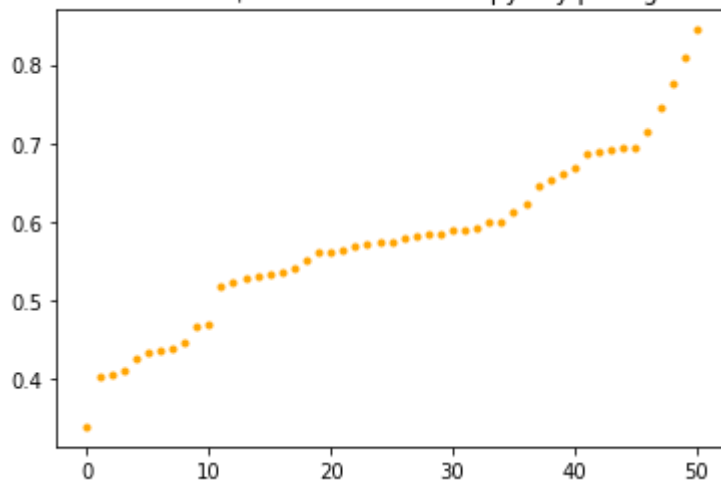


82	Новосибирская область	2159912	0	64418	16625	8473	213707	8
83	Омская область	1535403	0	51132	12487	10496	189303	8
84	Республика Татарстан (Татарстан)	2919482	0	104476	18045	13030	204618	8
85	Брянская область	978509	0	87912	7177	4472	68375	4
86	Город Байконур (Республика Казахстан)	14575	0	168	104	32	1026	

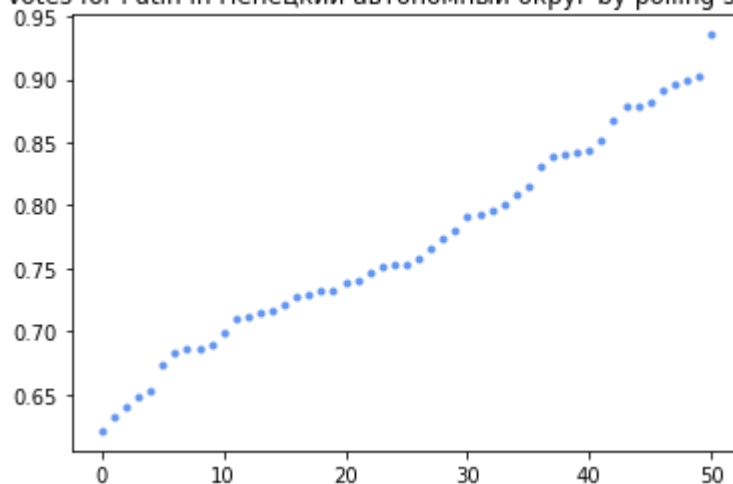
87 rows × 16 columns

In [10]: Некоторые регионы мы уже рассматривали в связи с высокой явкой. Рассмотрим Ненецкий и Чукотский автономные округа

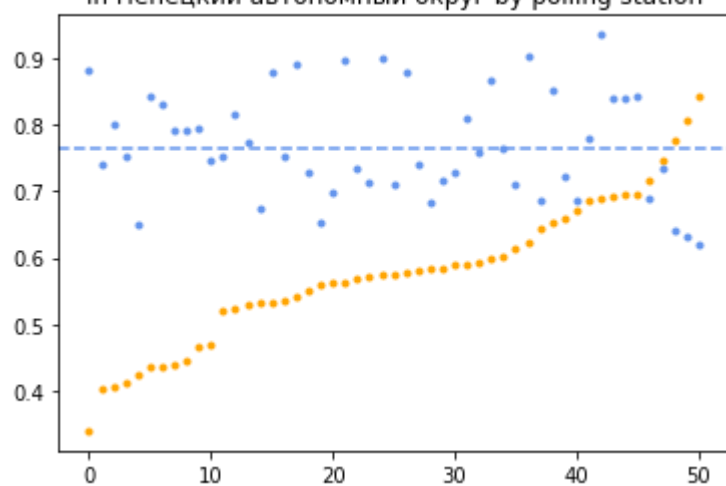
Turnout in Ненецкий автономный округ by polling station



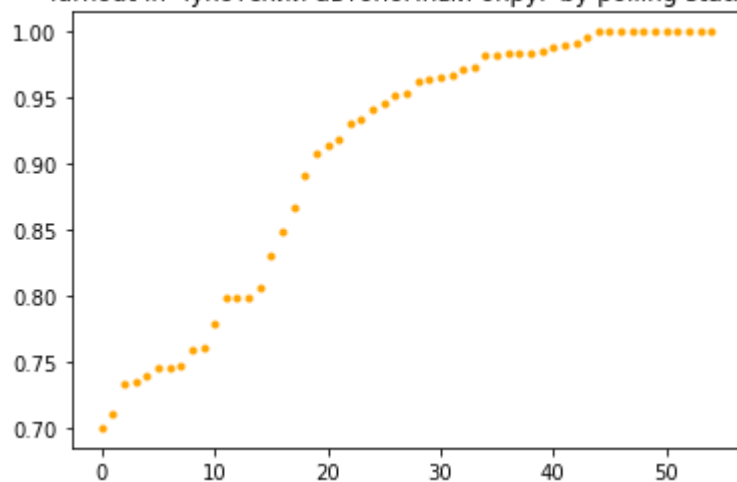
Votes for Putin in Ненецкий автономный округ by polling station



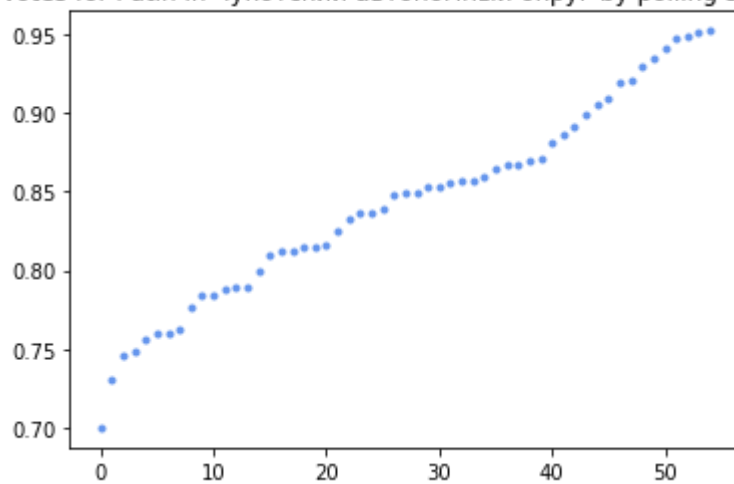
Dependence of votes for Putin on turnout
in Ненецкий автономный округ by polling station

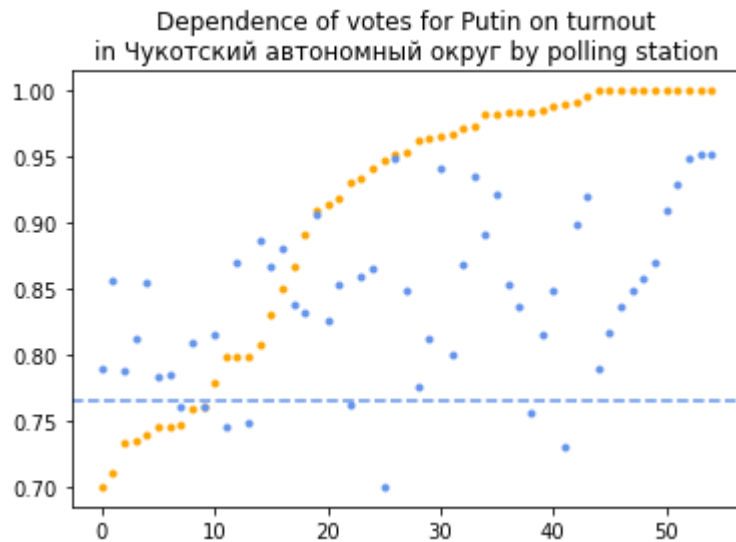


Turnout in Чукотский автономный округ by polling station



Votes for Putin in Чукотский автономный округ by polling station

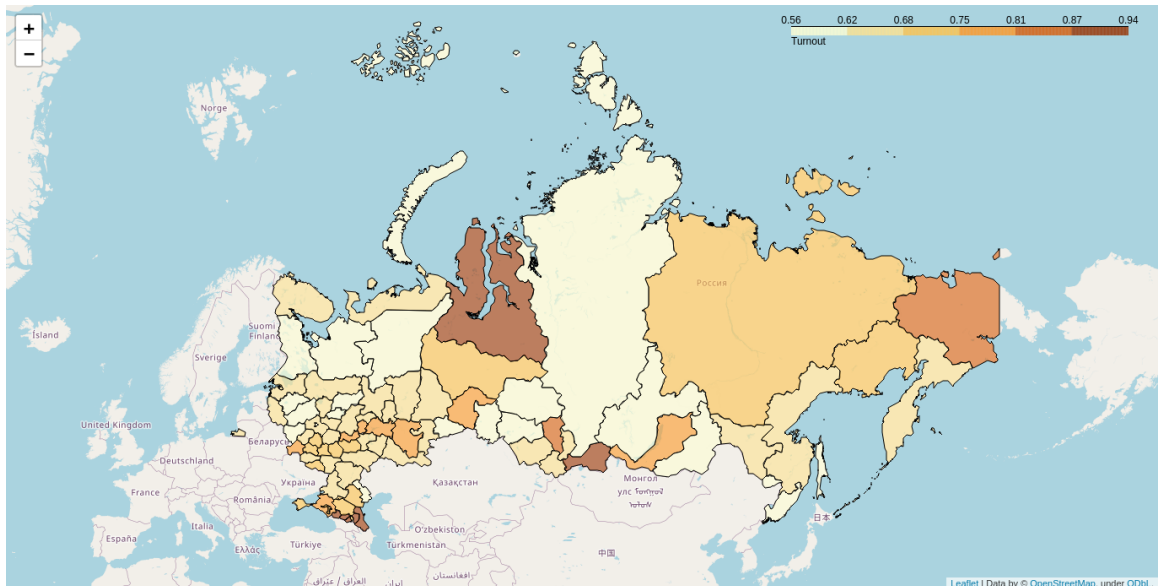




В Ненецком автономном округе результаты напоминают средние по стране, в то время как на Чукотке мы видим значительно превышающие средние значения явку и поддержку президента

In [17]: Посмотрим, как выглядит данные о явке по регионам на карте

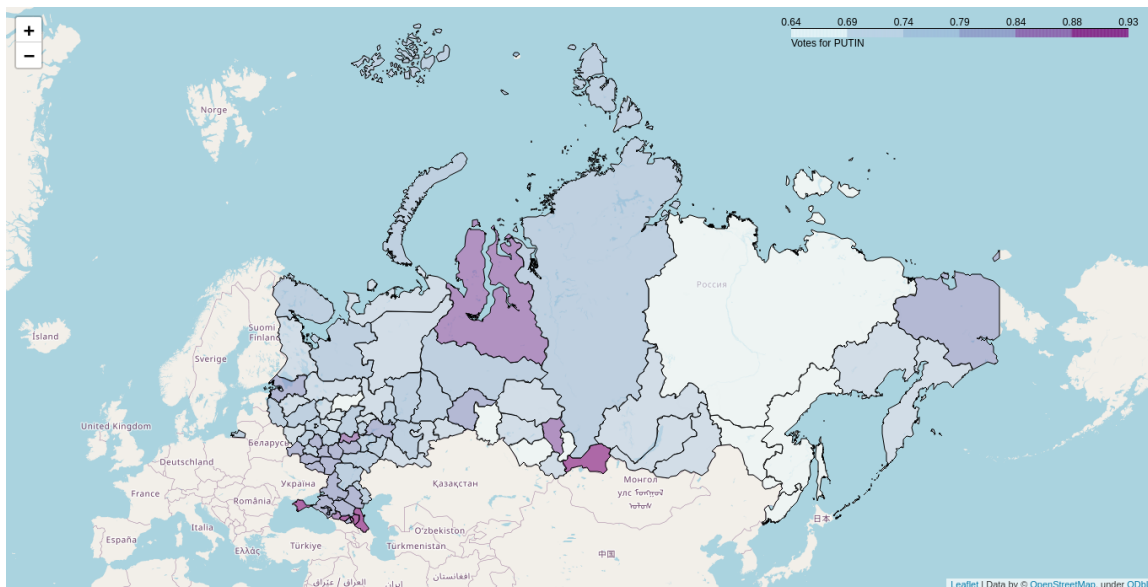
Out[17]:



В зависимости от региона явка на выборах 2018 года составила от 56 до 94 процентов. Видим, что наибольшая явка была в Тыве, Ямало-Ненецком автономном округе и на Северном Кавказе. Наиболее низкая явка на севере европейской части страны, в частности в Карелии и Архангельской области.

In [21]: Посмотрим на карту с голосами за Путина

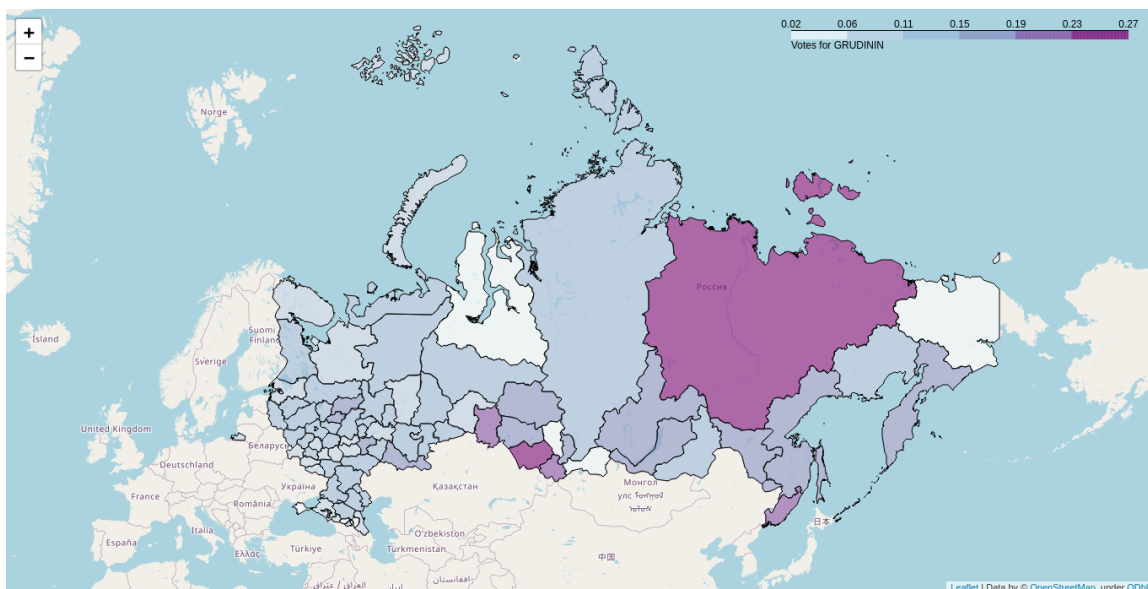
Out[21]:



В очередной раз замечаем, что регионы с высокой явкой (коричневые) совпадают с регионами с высокой поддержкой Путина (фиолетовые). Максимальный уровень поддержки - 93 процента. Меньше всего президенту доверяют в Якутии и в прилегающих к ней с юга территориях. Минимальный уровень поддержки - 64 процента

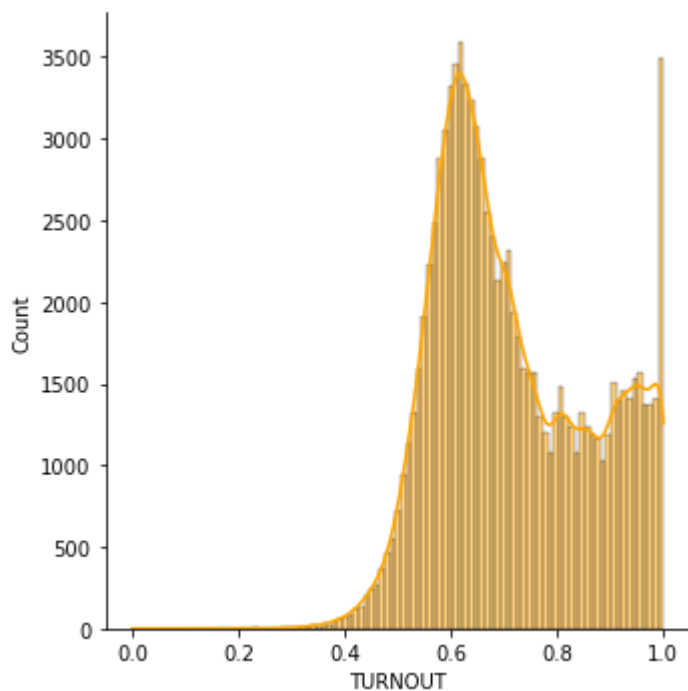
In [23]: Посмотрим на карту с голосами за Грудина

Out[23]:



Видим, что наибольшая поддержка Грудина в восточной части страны, особенно в тех регионах, где за Путина проголосовало относительно небольшое количество человек. Однако нет ни одного региона, где победу одержал Грудинин. Максимум - чуть больше четверти избирателей.

In [24]: С помощью облака слов посмотрим, какие регионы внесли больший вклад в победу Путина (в абсолютных цифрах)



Видим увеличение количества участков после 0.9, что не соответствует нормальному распределению. Кроме того, очень заметно большое количество участков с явкой 100 %. Посмотрим на них в таблице

In [22]: Замечаем увеличение количества участков после 0.9, что не соответствует нормальному распределению. Кроме того, очень заметно большое количество участков с явкой 100 %. Посмотрим на них в таблице

Out[22]:

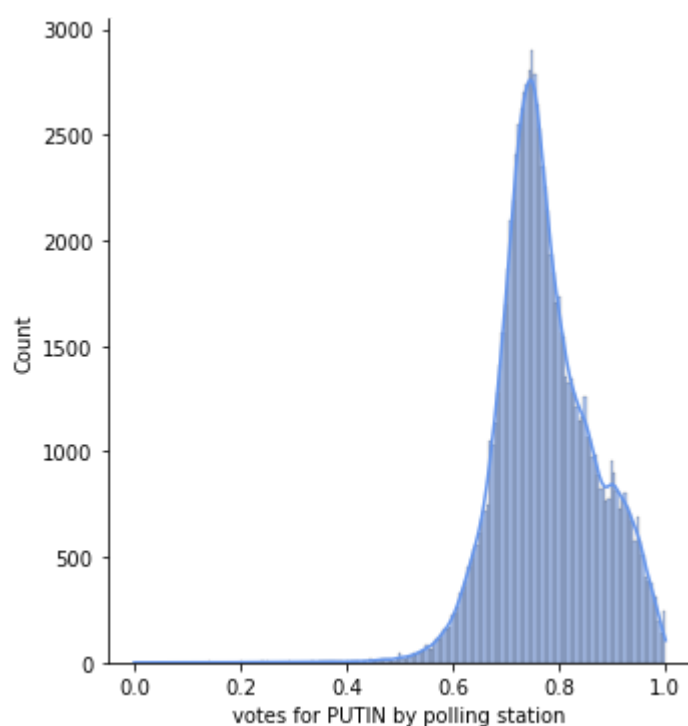
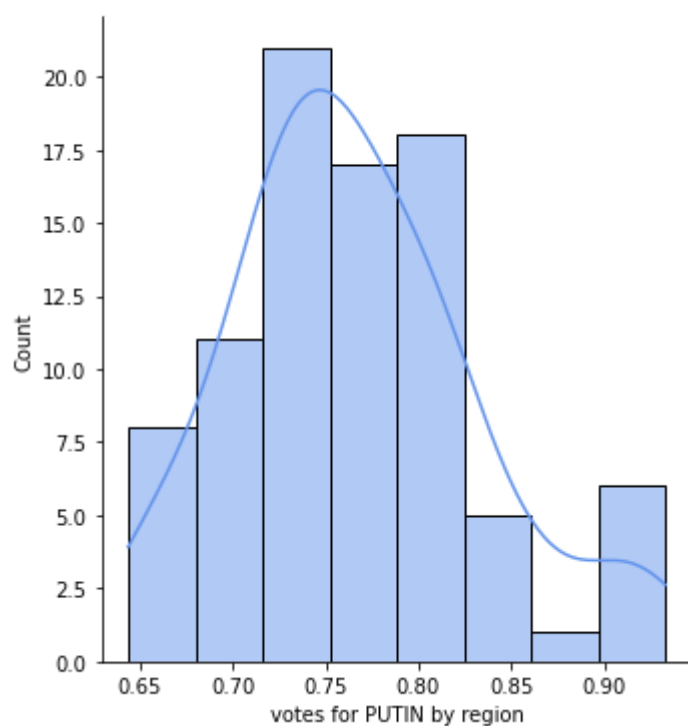
	PS_ID	REGION	SUBREGION	N_ALL	N_EARLY	N_OUT	N_INVALID	BAB
456	67	Республика Адыгея (Адыгея)	Красногвардейская	691	0	33	0	
461	72	Республика Адыгея (Адыгея)	Красногвардейская	290	0	9	1	
573	184	Республика Адыгея (Адыгея)	Майкопская городская	434	0	68	5	
574	185	Республика Адыгея (Адыгея)	Майкопская городская	430	0	152	4	
592	203	Республика Адыгея (Адыгея)	Тахтамукайская	593	0	16	6	
...
97524	520	Ямало-Ненецкий автономный округ	Надымская	235	78	6	0	
97525	521	Ямало-Ненецкий автономный округ	Надымская	552	187	8	1	
97535	601	Ямало-Ненецкий автономный округ	Новоуренгойская городская	166	0	96	1	

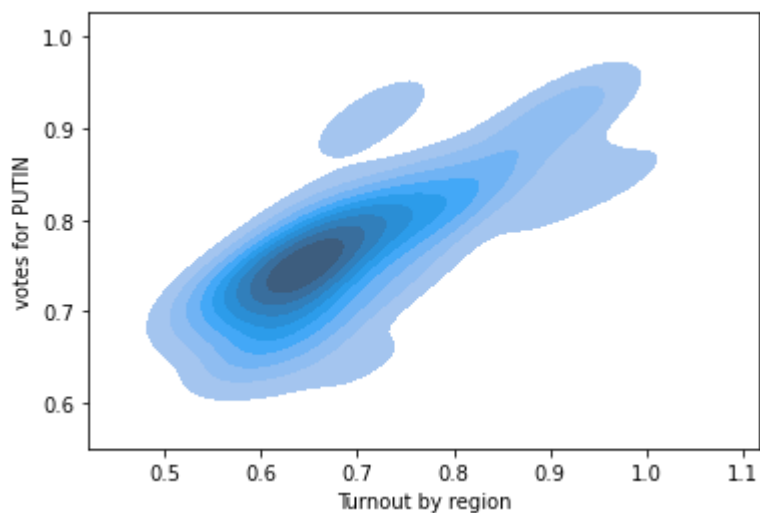
97671	1111	Ямало- Ненецкий автономный округ	Тазовская	2778	2727	0	7
97680	1301	Ямало- Ненецкий автономный округ	Ямальская	561	0	33	7

2290 rows × 17 columns

Более чем на 2000 участках явка составила 100 %. Это значительное количество

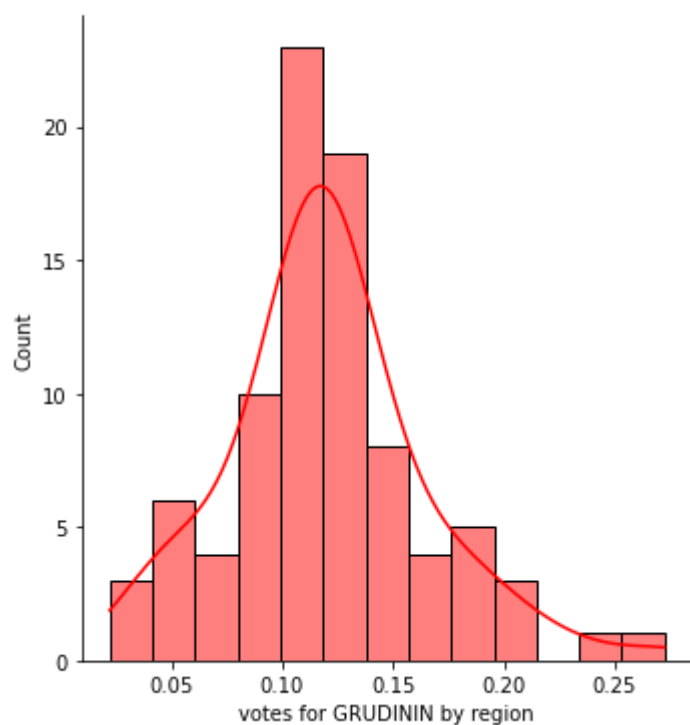
In [28]: Построим такие же диаграммы для Путина и добавим ещё одну, показывающую распределение голосов за кандидата в зависимости от явки

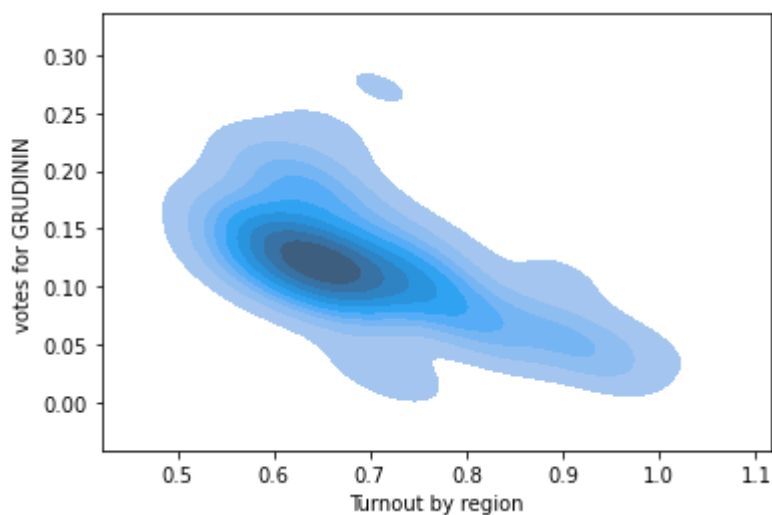
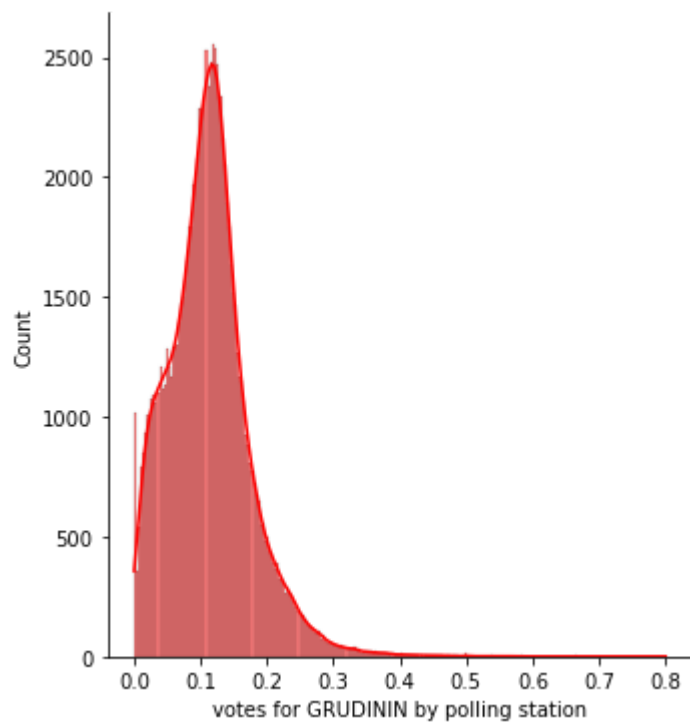




Видим, что Путин в среднем получил 76 %: наибольшее количество участков находится вблизи данного значения. Отмечаем, что практически отсутствуют участки, где данный кандидат набрал меньше 55 процентов

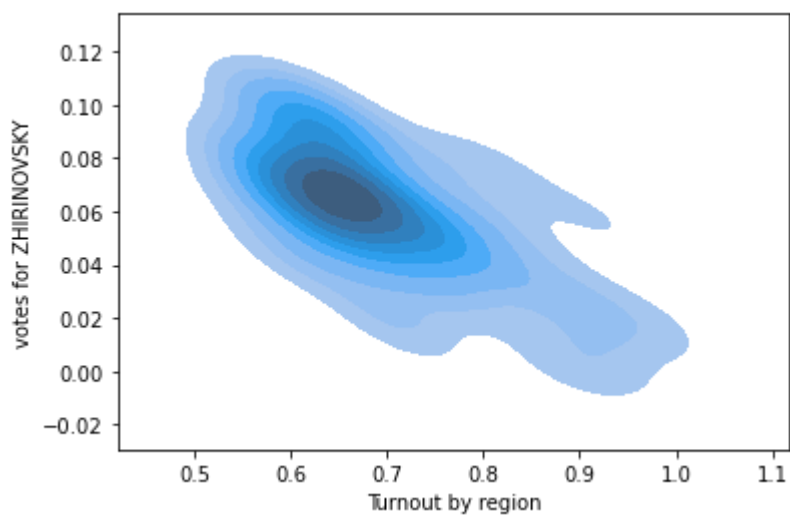
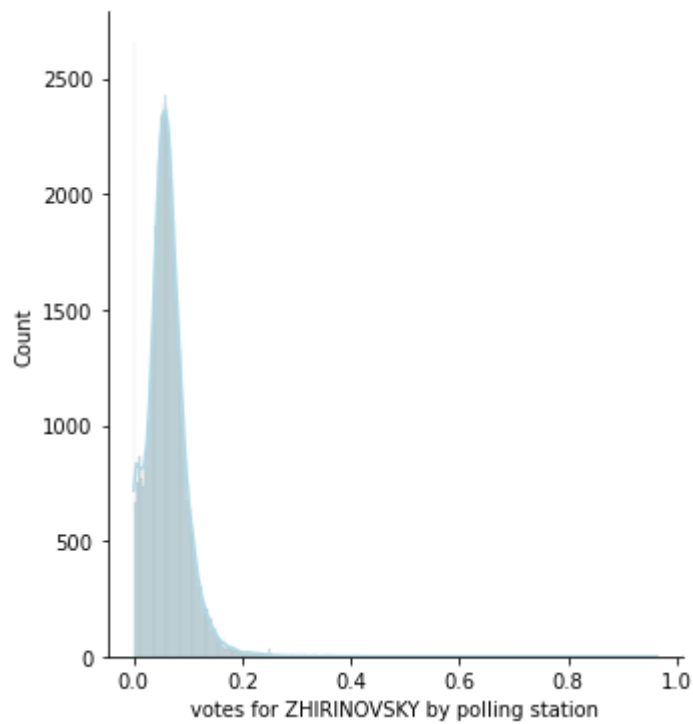
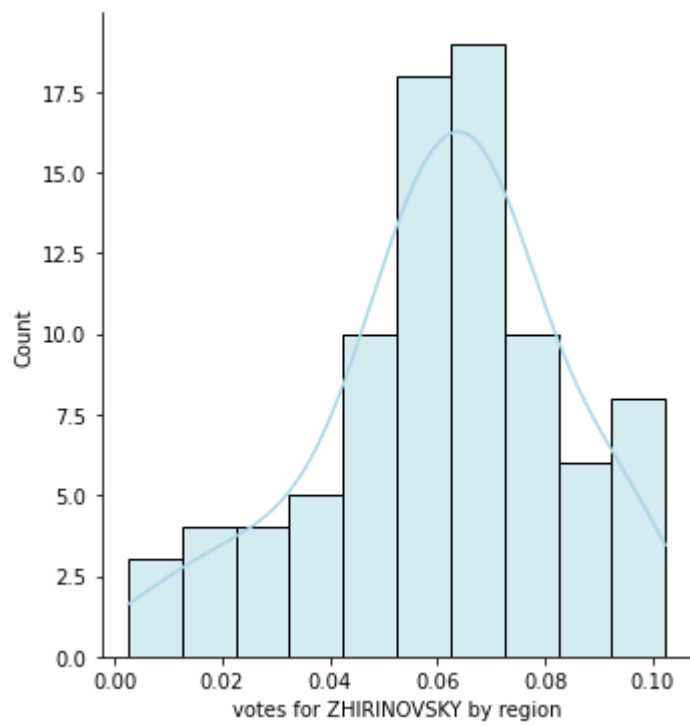
In [29]: для Грудинина





Видим, что Грудинин в среднем получил чуть больше 10 %: наибольшее количество участков находится вблизи данного значения. Отмечаем, что практически отсутствуют участки, где данный кандидат набрал больше 25 процентов

In [30]: и для Жириновского



Видим, что Жириновский в среднем получил около 6 %. Это заметнее на последнем графике, где находится жирный круг. Отмечаем, что практически отсутствуют участки, где данный кандидат набрал больше 10 процентов

Отмечаем, что целом распределение по регионам имеет похожую форму (причём не только у представленных выше кандидатов). На всех диаграммах мы видим рост последних столбцов, однако он незначителен у всех, кроме Путина, где последний столбец больше предыдущего в 5 раз. По последнему графику каждого из кандидатов мы видим, что с увеличением явки падает процент голосов за данного кандидата. У всех, но не у Путина: у него в этом месте наблюдается значительный рост