

Социально-экономические факторы электорального поведения в США: пространственный анализ результатов президентских выборов 2012 и 2016 гг.

Милицкая Алиса Ростиславна

*Аспирант департамента политики и управления, Национальный
исследовательский университет «Высшая школа экономики», Москва, Россия
alisamiletskaya@gmail.com*

Аннотация

В статье проводится пространственный анализ социально-экономических факторов, влияющих на электоральное поведение в США на президентских выборах 2012 и 2016 годов. Используя метод географически взвешенной регрессии (GWR), автор выявляет территориальные кластеры взаимосвязей между различными социально-экономическими показателями и результатами голосования. Для сравнения моделей регрессии также использовался расчет индекса пространственной автокорреляции (Moran's I) и локальный индекс пространственной автокорреляции Гетиса-Орда. Внимание уделяется изучению влияния демографических характеристик, социальных и экономических условий на электоральные предпочтения

американских избирателей. Исследование показывает, что воздействие данных факторов является географически нестационарным, а использование локальных моделей регрессии позволяет получить более точные объяснения в сравнении с глобальными моделями. В статье также рассматриваются кластеры, образованные взаимодействием различных факторов, и анализируется их пространственное распределение. Результаты работы подчеркивают значимость пространственной неоднородности и демонстрируют пересечения кластеров в ряде регионов США, что открывает новые перспективы для дальнейших исследований в области электоральной географии.

Ключевые слова

электоральное поведение, пространственный анализ, географически взвешенная регрессия, выборы в США, политическая география

Геотеги

Северная Америка, США

Дополнительные материалы

<https://clck.ru/3GPkn7>

Socio-Economic Factors of Electoral Behavior in the USA: Spatial Analysis of the 2012 and 2016 Presidential Elections

Alisa Miletskaya

*PhD Student at the Department of Politics and Governance, National Research University Higher School of Economics, Moscow, Russia
alisamiletskaya@gmail.com*

Abstract

The article conducts a spatial analysis of socio-economic factors influencing electoral behavior in the USA during the 2012 and 2016 presidential elections. Using the geographically weighted regression (GWR) method, the author identifies territorial clusters of relationships between various socio-economic indicators and voting results. Moran's I spatial autocorrelation index calculation and HotSpot Analysis of residuals were also used to compare regression models. The study focuses on the influence of demographic characteristics, social, and economic conditions on the electoral preferences of

American voters. The research demonstrates that the impact of these factors is geographically non-stationary, and the use of local regression models provides more accurate explanations compared to global models. The article also examines clusters formed by the interaction of various factors and analyzes their spatial distribution. The findings highlight the significance of spatial heterogeneity and demonstrate intersections of clusters in several regions of the USA, opening new perspectives for further research in electoral geography.

KEY WORDS

electoral behavior, spatial analysis, geographically weighted regression, US elections, political geography

GEO TAGS

North America, USA

SUPPLEMENTARY

<https://clck.ru/3GPkn7>

Введение

Что заставляет людей по-разному голосовать в различных частях страны? Известно, что электоральный процесс никогда не происходит в вакууме, а человеческое сообщество — это также явление пространственное, в рамках которого индивиды воспринимают особенности внешней среды, а также конструируют реальность в ответ [Hillier et al., 1984]. Избиратели, живущие в разных географических частях региона, неизбежно подвержены влиянию различных социальных и экономических факторов, которые могут иметь решающую роль в их политическом поведении.

Электоральное поведение подвержено влиянию среды, и поведения избира-

телей способно изменяться в зависимости от географии [Sui, Hugill, 2002]. Пространственная неоднородность возникает, когда один и тот же стимул вызывает разную реакцию в разных частях региона. Локация, на которой проживает потенциальный избиратель, может являться значимым предиктором того, как субъект будет голосовать даже с учётом рассмотрения и индивидуальных характеристик [Johnston et al., 2007].

При условии, что эти различные факторы играли решающую роль в развитии страны, мы не можем утверждать, что политическое поведение на современном этапе будет гомогенно в пространстве. Именно это предположение даёт стимул для исследования пространственной неоднородности значимости социально-экономических факторов на различия в голосовании. Мы также не можем утверждать, что особенности пространственной неоднородности стационарны во времени. Поэтому в исследовании предполагается рассмотрение случаев двух годов выборов: 2012 и 2016.

В данном исследовании хотелось бы сосредоточиться на социально-экономических факторах. Такие факторы не раз становились фокусом множества исследований электорального поведения на выборах в США [Campbell, 1980]. В частности, отмечалась значимая зависимость между предпочтениями избирателей и уровнем образования [Tenn, 2007], гендерной и расовой принадлежностью [Hajnal, Lee 2011], а также уровнем дохода [Evans, 2000].

Данные

Данные были взяты из открытого источника “GeoDa Data and Lab”⁵⁴. Независимые переменные, используемые для модели предпочтений избирателей, отражены в таблице [Таблица 1]. Переменные можно разделить на три основные категории: демографические показатели, которые отвечают за расселение возрастных групп, социальные категории, расовые и этнические группы, а также переменные, отвечающие за уровень образования, и экономические переменные, которые связаны со стоимостью жилья, уровнем бедности и другими монетарными категориями.

Табл. 1. Расшифровка независимых переменных

Table 1. Independent variables

Переменная	Описание
AGE295214	Процент населения до 18, 2014
AGE775214	Процент населения старше 65, 2014

⁵⁴ GeoDa Data and Lab [Электронный ресурс]: URL: https://geodacenter.github.io/data-and-lab/country_election_2012_2016-variables/ (accessed: 26.03.2025).

SEX255214	Процент населения (женщин), 2014
RHI125214	Процент белого населения, 2014
RHI325214	Процент представителей американо-индейской группы и уроженцы Аляски, 2014
RHI625214	Процент двух и более расовых групп, 2014
POP715213	Процент живущих в одном доме более года, 2009–2013
POP645213	Процент количества рожденных иностранцев по происхождению, 2009–2013
EDU635213	Процент населения старше 25 лет с полным школьным образованием, 2009–2013
EDU685213	Процент населения старше 25 лет с высшим образованием (бакалавриат), 2009–2013
HSG495213	Стоимость жилья, занимаемого владельцами, 2009–2013
HSD310213	Количество людей на домохозяйство, 2009–2013
PVY020213	Процент людей живущих за чертой бедности, 2009–2013

Источник: 2012 and 2016 Presidential Elections [Электронный ресурс]: URL: <https://clck.ru/3Jd38C> (accessed: 26.03.2025).

Электоральное поведение и экономические переменные

Существует ряд теорий, которые стремятся объяснить причину поведения избирателя с точки зрения географии. Пространственные модели голосования можно разделить по двум направлениям: композиционным и контекстуальным эффектам [Forest, 2018]. Композиционные эффекты относятся к моделям, возникающим из-за того, что сам состав населения различается в зависимости от региона. К таким исследованиям можно отнести работы, в которых в качестве первопричины выбора избирателя авторы относят этноконфессиональное расселение или исторические размежевания внутри региона [Steenbergen, 2010].

Напротив, контекстуальные эффекты проявляются в том, что внешняя среда и окружение избирателей формируют их политические предпочтения. Такие исследования соотносятся с теорией соседства Кокса, согласно которой расселение и социальные связи, которые приобретаются путём физической близости избирателей друг к другу, имеют непосредственное влияние на политические предпочтения [Cox, 1970]. Наше исследование предполагает выявление контекстуальных эффектов. Напомним, что мы стремимся обнаружить кластеры влияния отдельных факторов, поэтому рассмотрение географической близости занимает центральное внимание в исследовании.

Экономические факторы электорального поведения имеют достаточно вы-

сокую важность для выборов. Электоральные предпочтения избирателей имеют тенденцию к изменению в связи с их индивидуальным уровнем дохода [Lind, 2007], при котором близость людей с высоким доходом друг к другу определяет их электоральное поведение. В том числе это подтверждается в рамках многочисленных исследований выборов в США [Kramer, 1971; Lewis-Beck, Michael, 1988].

По некоторым исследованиям известно, что показатели национальной экономики сильно влияют на поведение избирателей [Kinder, Kiewiet, 1979], в то время как экономические показатели по штатам, такие как безработица, расовое и этническое распределение, имеют значимость на локальном уровне по выборам за 2008 год [Cho, Gimpel, 2009].

Гипотезы

- н1: Данные по социально-экономическим факторам по голосованию на президентских выборах за 2012 г. и 2016 г. являются географически нестационарными.
- н2: Для предполагаемых взаимосвязей географически взвешенная регрессия является более подходящей в сравнении с линейной регрессией.
- н3: Переменные по социально-экономическим факторам кластеризуются в пространстве, что оказывает значимое влияние на результаты выборов.
- н4: Кластеры, обнаруженные в ходе построения географической регрессии, будут иметь пересечения в пространстве.

Метод

В рамках линейной регрессии, построенной методом наименьших квадратов (МНК), предполагается, что наблюдаемый процесс протекает одинаково на всей исследуемой территории [Paez, 2004]. Хотя такой анализ может дать некоторые открытия относительно гипотез, выявив связи между предикторами и результатами, остается неясно, является ли пространственная нестационарность важным аспектом. Неоднократно было доказано, что географически взвешенная регрессия (GWR) является полезным методом в изучении электоральной географии в контексте исследования локальных взаимосвязей поведения избирателей, так как данная модель позволяет исследовать географическую неоднородность [Brunsdon et al., 1996].

При использовании данного метода предполагается, что глобальная модель (регрессия МНК), имеет меньший объяснительный потенциал в пользу локальной модели (GWR), которая позволяет выявить отличающиеся зависимости в разных

точках пространства. Известно, что глобальная регрессия тогда эффективна и может быть верно проинтерпретирована, когда остатки модели не коррелируют друг с другом и имеют постоянную дисперсию [Hamilton, 1992]. Однако существует предположение, что при возникновении пространственной автокорреляции в модели глобальной регрессии, локальная регрессия способна уменьшить её [Tu, Xia, 2008]. Это предположение мотивирует нас провести сравнительный анализ двух моделей.

Пространственную нестационарность можно определить как ситуацию, когда характер и значимость взаимосвязей между переменными различаются в зависимости от местоположения [Fotheringham et al., 2009]. В нашем случае это означает, что мы сможем обнаружить области с низким/высоким показателем голосования за Республиканскую или Демократическую партию и лучше понять, какие предикторы связаны с электоральными результатами в конкретном месте.

Во время анализа получившихся регрессий необходимо будет сравнить их путём интерпретации показателя информационного критерия Акаике (Akaike info criterion, AIC) [Akaike, 1974], а также показателя коэффициента детерминации для того, чтобы сделать вывод относительно того, насколько пространственная модель способна улучшить наш вывод.

Для того, чтобы сравнить результаты GWR и линейной регрессии, обе модели были построены с использованием процентной доли голосов, отданных Демократической партии, от числа голосов, отданных либо Демократической партии, либо Республиканской партии в каждом графстве в качестве зависимой переменной, и 13 социально-экономических факторов в качестве независимых переменных.

Уравнение регрессии, построенной методом наименьших квадратов будет выглядеть следующим образом:

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \varepsilon,$$

где y — зависимая переменная, β_0 — константа, β_i — оценка параметра для независимой переменной x_i , p — число независимых переменных, а ε — ошибка регрессии.

Уравнение географически взвешенной регрессии, дополненное координатами наблюдений:

$$y_j = \beta_0(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j) x_{ij} + \varepsilon_j,$$

где u_j и v_j — координаты местоположения наблюдения j , $\beta_0(u_j, v_j)$ — точка пересечения для наблюдения j , а $\beta_i(u_j, v_j)$ — оценка локального параметра для

независимых переменных x_i в данном местоположении.

Так как данная модель имеет под собой географическую составляющую, взвешенная функция уменьшения расстояния используется для наблюдения путём взвешивания наблюдения вокруг него. Такая функция может выражена в экспонентной форме:

$$w_{ij} = \exp\left(\frac{-d_{ij}^2}{b^2}\right),$$

где w_{ij} — вес наблюдения j для наблюдения i , d_{ij} — расстояние между наблюдениями i и j , а b — пропускная способность ядра. Такая модель предполагает, что вес быстрее приближается к нулю, когда расстояние превышает пропускную способность ядра. При построении модели необходимо учитывать, что пропускная способность ядра является ключевым управляющим параметром модели и может быть задана либо фиксированно (постоянная полоса пропускания), либо адаптивно (размер полосы пропускания меняется в зависимости от плотности данных) [Fotheringham et al., 2009].

В программном обеспечении ArcGIS, в котором как был проведён анализ, нет выбора одного из двух типов пропускной способности, а вместо этого существует функция “Golden Search”, которую мы будем использовать. С помощью неё программа самостоятельно просчитывает расстояние пропускной способности и строит модель регрессии с наилучшей.

Также стоит обратить внимание на показатель остатков модели. Показатели для чётко определённых моделей регрессии будут нормально распределены и пространственно случайны, без кластеризации значений. Для того, чтобы проверить это, мы воспользуемся инструментом пространственной автокорреляции (Индекс I Морана) [Morgan, 1948]. Он сравнивает значение одной переменной в определённом местоположении со значением этой переменной в других близлежащих местоположениях и принимает значения от -1 до 1. Значение «1» означает идеальную позитивную пространственную автокорреляцию (значения кластеризуются в одном месте), «-1» свидетельствует об идеальной негативной корреляции, а значение «0» наблюдается в случаях идеальной пространственной рандомизации [Ishizawa, Stevens, 2007].

Описательная статистика и результат глобальной модели

В таблице 2 приведены статистические данные о проценте голосования за демократическую партию США на выборах 2012 и 2016 гг.. В среднем результаты голосования за демократов в 2012 и 2016 гг. принимают схожие значения: в 2012 г. этот пока-

затель — 38,45%, в 2016 г. — 31,67%, а процент по округам варьируется от 3% до 93%.

Табл. 2. Статистические данные о проценте голосования за Демократическую партию США.

Table 2. Statistical data on the percentage of votes for the Democratic Party of the USA.

Field Name	Minimum	Maximum	Mean	Standard Deviation	Count
AGE295214	0	40,5	22,531081	3,360081	3108
AGE775214	0	52,9	17,635393	4,400768	3108
EDU635213	45	99	84,50888	6,912412	3108
EDU685213	3,2	74,4	19,735746	8,831593	3108
HSD310213	1,85	4,7	2,523662	0,242429	3108
HSG495213	19900	929700	129962,162162	76591,991198	3108
pct_dem_12	0,034483	0,933546	0,384502	0,147606	3108
pct_dem_16	0,031447	0,928466	0,316667	0,153265	3108
POP645213	0	51,3	4,477896	5,508712	3108
POP715213	50,8	99,8	86,432014	4,39421	3108
PVY020213	0	53,2	16,711519	6,488519	3108
RHI125214	0	99,3	85,454022	15,739722	3108
RHI325214	0	92,2	1,976093	6,550791	3108
RHI625214	0	10,3	1,824775	1,074612	3108
SEX255214	0	56,8	49,938771	2,380668	3108

Источник: составлено автором.

Процент голосов, отданных за Демократическую партию на двух президентских выборах также представлен на картах, построенных по интервалам, заданным вручную [Приложение 1]. Округа, окрашенные синим цветом, являются территориями, на которых за демократов проголосовало большинство избирателей (>50%). Округа с наибольшим количеством голосов превалируют в северо-восточном регионе США, захватывая штаты Мэн, Массачусетс, Нью-Гемпшир, Род-Айленд и Вермонт. В том числе высокие показатели доминируют в части Нью-Мексико, на западном побережье, а также на юге Техаса. Наблюдения основываются на картах по обоим годам выборов, так как пространственный паттерн за выбранный период не менялся.

Что касается статистических показателей, начнем с возраста. В среднем, 22,5% населения младше 18 лет, а 17,6% — старше 65 лет. Среди образовательных пока-

зателей: 84% имеют полное школьное образование, 19,7% — высшее. Средний размер домохозяйства составляет 2,5 человека, средняя стоимость жилья — \$130 000. Процент рождённых иностранцев в среднем — 4,5%, максимум — 50%. 86% семей живут в одном доме более года, а 16,7% находятся за чертой бедности. В расовом/этническом составе: 85,45% — белое население, 1,97% — американо-индейцы и уроженцы Аляски, 1,82% — представители двух и более рас.

Табл. 3. Модель линейной регрессии, предсказывающая процент голосов, отданных за Демократическую партию.

Table 3. OLS model predicting the percentage of votes cast for the Democratic Party.

Variable	2012				2016			
	Coefficient	t-statistics	p-value	VIF	Coefficient	t-statistics	p-value	VIF
Intercept	-0,014723	-0,167732	0,9	-	0,056728	0,768389	0,6	-
AGE295214	-0,015376	-14,475359	<0,01	3,3	-0,012341	-13,813619	<0,01	3,3
AGE775214	-0,010421	-13,215601	<0,01	3,1	-0,007311	-11,022619	<0,01	3,1
SEX255214	0,011863	11,773741	<0,01	1,5	0,011145	13,151501	<0,01	1,5
RHI125214	-0,003692	-22,349624	<0,01	1,8	-0,004727	-34,023552	<0,01	1,8
RHI325214	0,002588	6,473667	<0,01	1,8	0,000707	2,10299	<0,01	1,8
RHI625214	-0,015814	-7,090434	<0,01	1,5	-0,014766	-7,871442	<0,01	1,5
POP715213	0,003357	5,725515	<0,01	1,7	0,001867	3,784871	<0,01	1,7
POP645213	0,007359	13,809738	<0,01	2,3	0,007753	17,298342	<0,01	2,3
EDU635213	0,005291	9,619371	<0,01	3,8	0,002872	6,209026	<0,01	3,8
EDU685213	-0,001626	-3,611846	<0,01	4,1	0,002239	5,911392	<0,01	4,1
HSG495213	0,0	8,160914	<0,01	3,0	0	9,598576	<0,01	3,0
HSD310213	-0,073155	-5,324136	<0,01	2,9	-0,023408	-2,025478	0,15	2,9
PVY020213	0,004553	8,833883	<0,01	2,9	0,004177	9,635312	<0,01	2,9

Источник: составлено автором.

Мы можем с уверенностью утверждать, что большая часть независимых переменных статистически значима. При этом самый высокий показатель теста на мультиколлинеарность (VIF) равняется 4,1, что сильно ниже общепризнанной предельной отметки в 10 [Menard, 2002], что свидетельствует об отсутствии избыточности среди переменных.

Проинтерпретируем результаты глобальной модели. Демографические переменные показывают, что каждое увеличение доли населения младше 18 лет на один процентный пункт связано с уменьшением на 1,53% голосов за Демократическую партию. Каждое увеличение численности людей старше 65 лет на один

процентный пункт связано с уменьшением на 1,04% голосов. Демографические особенности значимо ассоциированы с голосованием за демократов, как и гендерный состав: увеличение доли женщин связано с увеличением доли голосов за демократов на 1,18%.

Расовые/этнические показатели имеют меньшую значимость. Увеличение доли белого населения в округе на процентный пункт связано с уменьшением процента голосов за демократическую партию на 0,36%. При увеличении на один процентный пункт группы уроженцев Аляски процент голосов за демократов увеличивается на 0,26%. Процент наличия двух и более рас связан с уменьшением на 1,58%. С увеличением доли тех, кто родился, будучи иностранцем по происхождению, на каждый процентный пункт, доля голосования за демократов увеличивается на 0,74%, а при каждом увеличении процента живущих в одном доме более года голоса увеличиваются на 0,34%.

О переменных, отвечающих за образование, мы можем сказать следующее: с увеличением на один процентный пункт доли тех, у кого есть документ об окончании школьного образования, доля голосов за демократов увеличивается на 0,53% в 2012 году, в 2016 году значение составляет 0,29%. При каждом увеличении доли людей с высшим образованием процент голосов уменьшается на 0,16% в 2012 году, но в 2016 году становится положительным (0,22%).

Теперь проинтерпретируем экономические переменные. Стоимость жилья практически не влияет на результаты голосования, но переменная количества людей на домохозяйства имеет положительную связь с голосованием за демократическую партию. Каждое увеличение показателя количества людей на домохозяйство ведёт к увеличению голосов на 7,3% в 2012 году, и на 2,34% в 2016 году.

Стоит иметь в виду, что выбранная модель предполагает, что данные стационарны в пространстве. Можем предположить, что именно из-за этого показатель коэффициента детерминации отмечен на уровне 0,64 за 2012 год и 0,45 за 2016 год (см. Таблицу 4), что свидетельствует о малой пригодности модели. Упускается пространственное разнообразие, которое может быть важным аспектом при изучении выборов в США. На нынешнем этапе нам необходимо изучить пространственную неоднородность с помощью локальной модели.

Географически взвешенная регрессия

Интерпретацию стоит начать с пространственной визуализации локального показателя R^2 для каждого из округов США [Приложение 2]. Объяснительный потенциал модели колеблется между 33% и 97% в 2012 году, 63% и 98% в 2016 году. Модель локальной регрессии хорошо согласуется с данными по многим регионам США, в особенности на территории Западного побережья и юго-востока страны,

где наблюдаются кластеры значений выше 90%. В то же время территория Великих Озер имеет намного меньший показатель R_2 , который колеблется между 33% в 2012 году и 75% в 2016, как и часть юго-запада в 2012 году.

Для каждой переменной мы составили визуализацию, чтобы оценить значимость переменных по каждому округу, а также выявить кластеры значений. Статистически значимые локальные показатели коэффициентов по каждому округу были визуализированы с использованием градиентной заливки, где жёлтым были выделены самые высокие значения по каждому предиктору, а темно-фиолетовым — самые низкие. Серый цвет на карте показывает статистически незначимые пространства. В итоге у нас получилось 26 карт, было визуализировано 13 переменных по обоим годам выборов [Приложение 3].

Рассмотрим демографические переменные. На карте за показаны локальные оценки коэффициента независимой переменной, которая отвечает за процент населения до 18 лет. На выборах за оба года он связан с предпочтением голосовать за демократов как положительно, так и отрицательно. Отрицательная связь наблюдается в Сан-Диего и его соседних округах, а самая значимая положительная связь наблюдается на юге Техаса. Карта 3в представляет локальные оценки для переменной доли населения старше 65 лет, и её географический паттерн схож с предыдущей картой, однако связь имеет отрицательный характер по всей территории. Это свидетельствует о том, что две демографические переменные, отвечающие за возраст, кластеризуются в пространстве, поэтому в части о связности кластеров мы рассмотрим их более подробно.

Обратим внимание на гендерный фактор в голосовании. На рисунке 3с отображено локальное распределение по переменной с процентным показателем женщин. Данная переменная имеет сильно большую значимость в 2016 г. в сравнении с 2012 г. Ассоциация имеет положительную связь, однако на территории восточного побережья и центра страны значимость связи очень мала (0,43%). Несмотря на то, что фактор гендера был отмечен как пространственно однородный в одном из исследований [Forest, 2018: 5], по результату локальной модели мы можем отчётливо видеть, что процент женщин хоть и не имеет разного типа связи, но имеет неоднородный характер.

Рассмотрим переменные, связанные с расовой и этнической принадлежностью. Первые две карты под нумерацией 3д отображают процент белого населения. Мы можем обнаружить существенную связь между этим показателем и значительной частью территории США. Переменная статистически значима на территории от центра страны до восточного побережья. Показатель гомогенен в пространстве, имея лишь один кластер с сильно отрицательной связью в регионе Великих равнин. Помимо этого, в 2016 году можем заметить кластер в Калифорнии с положительной связью, который в 2012 г. не наблюдался.

Коэффициент показателя процента представителей американо-индийской

группы и уроженцев Аляски представлен на карте 3е. По обоим годам связь непостоянная по своему направлению, так как на территории центра страны дугой с юга Южной Дакоты до Теннеси распространяется кластер округов с отрицательной связью, в то время как в юго-западной части Техаса и на территории Калифорнии связь приобретает положительный характер.

В то же время округа, в которых проживает две и более расовые группы, имеют несколько статистически значимых географических кластеров с положительной и отрицательной связями, что представлено на карте 3f. Один кластер с отрицательной связью наблюдается в регионе Великих равнин и распространяется на восток, а второй наблюдается неявно в 2012 году, однако становится более заметным в 2016 году. Он занимает территорию юга Техаса и всего Нью-Мексико.

На данном этапе стоит сделать вывод по всем переменным, относимым к расовым и этническим группам. Существуют кластеры, которые по своему территориальному расположению и характеру связи совпадают по всем трём переменным. К примеру, кластер с отрицательной связью в регионе Великих равнин и кластер с положительной связью на севере Калифорнии.

На карте 3g мы можем увидеть неоднородную ассоциацию между процентом семей, живущих в одном доме год и более, и голосованием за демократов. Можем заметить, что связь является неоднородной. Отметим кластер с положительной связью в штате Нью-Мексико, а также кластер с отрицательной связью на юге штата Техас.

Перейдём к части рассмотрения влияния образования на голосование. В приложении 3i мы можем видеть карту по переменной, отвечающей за процент людей со школьным образованием. На карте представлена неоднородная связь, кластер с негативной связью может быть замечен в штатах Техас, Колорадо и Нью-Мексико. Значимая положительная связь замечена на севере США.

Рассмотрим процент людей с высшим образованием 3j. В 2012 году мы можем заметить кластер со значимой положительной связью в южной части Техаса, а также с региона Скалистых гор до юго-западной части региона страны. Интересно, что процент школьного образования в регионе имел отрицательную связь, а процент высшего образования в практически том же кластере имеет положительную связь. Также мы можем заметить кластер в районе северо-востока, который в 2016 г. распространяется на юг.

Пространственная неоднородность связи между стоимостью жилья и процентом голосов отражена на карте 3k. По обоим циклам выборов связь носит положительный характер, с несколькими кластерами в штате Аризона. Если говорить о переменной, которая отвечает за среднее количество людей на домохозяйство на картах 3l, то кластер с положительной связью локализуется в центре Техаса и частично в Нью-Мексико, а отрицательная связь заметна в регионе Великих озер и в Нью-Йорке. На протяжении двух циклов выборов процент людей, живущих

за чертой бедности, имеет позитивную связь, что мы можем увидеть на картах 3м. Значимый кластер с такой связью наблюдается в Техасе и на территории всего Нью-Мексико.

Сравнение моделей регрессии

Проверим гипотезу о том, что данные по голосованию на президентских выборах за 2012 г. и 2016 г. являются географически нестационарными.

Карты пространственного распределения остатков двух моделей за 2012 и 2016 гг. отражены в приложении [Приложение 4]. Визуализация остатков была выполнена методом вручную настроенных интервалов, что позволило отобразить территории с относительно небольшим показателем остатков ($\pm 2,5$), и с показателем в более чем $\pm 2,5$.

Известно, что величина остатков уравнения регрессии является одним из показателей соответствия модели. Чем больше показатель остатков, тем менее подходящей можно считать модель. Показатель ошибок у модели географической регрессии намного меньше, а также они носят случайный характер, без кластеризации в какой-либо части исследуемого региона. Рассмотрим расчёт пространственной автокорреляции подробнее.

Значимая позитивная пространственная автокорреляция была найдена у глобальных моделей за 2012 г. (индекс Морана = 0,5358, $p\text{-value} < 0,001$) и 2016 г. (индекс Морана = 0,5119, $p\text{-value} < 0,001$). Из-за того, что результат пространственной автокорреляции статистически значим ($p\text{-value} = 0,0000$), а z -оценка положительна и относительно высока (95,4 за 2012 г. и 91,2 за 2016 г. значительно больше минимального показателя в 2,58), мы можем отклонить нулевую гипотезу о том, что значения распределены случайным образом. Индекс Морана положителен в каждом случае, что означает, что пространственное распределение сгруппировано, и остатки модели глобальной регрессии пространственно автокоррелированы.

Для того, чтобы попытаться сделать вывод о конкретных территориях, в которых наблюдается упущение переменных, мы выполним расчёт локального индекса пространственной автокорреляции Гетиса-Орда. Этот инструмент позволит нам определить статистически значимые пространственные кластеры.

В качестве результата мы получим карту кластеризации вместе с z -score и $p\text{-value}$, которые являются показателями статистической значимости. Коэффициенты показывают, является ли наблюдаемая пространственная кластеризация значений более выраженной, чем можно было бы ожидать при их случайном распределении [Ord, Getis 1995].

На картах [Приложение 5] мы можем увидеть несколько статистически значимых кластеров с концентрацией низких и высоких значений. Кластеры с кон-

центрацией высоких значений остатков в 2012 году наблюдаются в южной части Техаса, в северных частях Нью-Мексико и Миннесоты, на юго-западе Висконсина, а также имеется кластер, рассредоточенный по территории Массачусетса, Вермонта и Род-Айленда. Кластеры с низкими значениями более рассредоточены в пространстве и распределены скорее рандомно ($p\text{-value} < 0,1$ и $z\text{-score}$ около $-1,4$). В 2016 году паттерн в целом сохраняется, единственным отличием является пропажа кластера с низкими значениями на севере Небраски, а также появление кластеров с высокими значениями на западном побережье.

География кластеров с высокими значениями почти совпадает с географическим распределением высокого процента голосов за демократов. При этом значения с низким процентом голосов не коррелируют с низкими кластерами. Получается, что глобальная регрессия не так хорошо соответствует данным именно для тех частей страны, где наблюдается высокий процент голосования за демократов.

Результат индекса для локальной модели значительно разнится с глобальной моделью. У него наблюдается корреляция в 2012 г. (индекс Морана = 0,123, $p\text{-value} < 0,001$) и в 2016 г. (индекс Морана = 0,105, $p\text{-value} < 0,001$). Это очень низкий показатель, и пространственную автокорреляцию локальной модели можно считать случайной.

Табл. 4. Сравнение моделей глобальной и локальной регрессии.

Table 4. Comparison of global and local regression models.

	2012		2016	
	OLS	GWR	OLS	GWR
R^2	0,64	0,83	0,45	0,9
AIC	-4940,4	-8222,3	-6016,3	-9143,7
Moran's I	0,5358*	0,123*	0,5119*	0,105*
Spatial Pattern	кластер	случайно	кластер	случайно

* $p\text{-value} < 0,001$

Источник: составлено автором.

В таблице 4 отмечен ранее получившийся показатель индекса Морана для остатков модели, который мы отметили как более удачный у глобальной модели. Также в таблице представлен показатель R^2 , который значительно выше у глобальной модели за оба года выборов. Критерий AIC значительно ниже у глобальной модели: -8222,3 у локальной модели в сравнении с -4940,4 у глобальной модели за 2012 год.

Так, модель линейной регрессии менее пригодна для определения взаимосвязей, поскольку пространственные зависимости голосования не учитываются

должным образом в модели, в то время как GWR повышает надёжность найденных взаимосвязей за счет уменьшения пространственных автокорреляций в остатках. Ориентируясь на данный результат, мы можем однозначно сказать, что GWR уменьшает значение ошибок, как предполагалось на этапе выбора данной модели [Tu, Xia, 2008].

Эпицентры кластеров

Кластеры, которые нам удалось обнаружить в рамках визуализации локальной регрессии, имеют определённые пересечения. На данном этапе стоит проанализировать данную согласованность кластеров, визуализировав её на ранее созданных картах.

Для начала выделим переменные, которые по итогу географически взвешенной регрессии оказались совпадающими по кластерам:

- 1) Демографические переменные — кластер между процентом людей до 18 лет и старше 65 лет в 2016 году.
- 2) Социальные переменные — кластер между процентом белого населения, процентом американо-индейской группы и процентом двух и более рас в регионе в 2016 году; кластер по уровню образования между теми, кто окончил школу и имеет диплом о высшем образовании за 2016 год.
- 3) Смешанные переменные — кластер между процентом количества рождённых иностранцев и процентом женщин за 2016 год; процент людей, живущих за чертой бедности и людей с высшим образованием за 2016 год.

Таким образом, у нас получилось 5 пересекающихся кластеров. Все пять пересечений кластеров были визуализированы на картах [Приложение 6] следующим образом: на карту были нанесены два (и более) слоя с уровнем прозрачности в 50%. Это позволило увидеть, в каких территориях кластеры действительно пересекаются, так как вследствие пересечения одно цвета насыщенность цвета в целом на карте увеличивалась.

Начнём анализ с кластеров по демографическим показателям процентного соотношения возрастных групп, отражённых на карте 6а. В первую очередь обратим внимание на несколько пересекающихся кластеров с отрицательной связью. Один из них можем обнаружить на севере региона Скалистых Гор, он захватывает множество штатов, а также продолжается на юг данного региона, имея наиболее выраженные показатели в штатах Нью-Мексико и Аризона. В штате Орегон также наблюдается кластер с отрицательной связью. Можем предположить, что это связано с проводимой демографической политикой в выделенных областях со стороны государства, или с конкретной политикой, проводимой партией на местах.

Теперь рассмотрим кластер между расово-этническими переменными на карте

6b. На карте обнаруживается по одному кластеру с положительной и отрицательной связью. Первый локализуется в центральной части региона Великих равнин, а второй располагается в северной части Калифорнии. Рискнём предположить, что кластер с положительной связью в Калифорнии связан с этническим расселением, которое исторически сформировалось в регионе определенным образом [Forbes, 1968]. Однако это наблюдение требует дополнительных исследований.

Пересечение кластеров, которые относятся к влиянию уровня образования на голосование за Демократическую партию, отражено на карте 6c. Мы можем заметить один кластер с отрицательной связью с эпицентром в штате Колорадо. Связность образования в штате Колорадо выглядит весьма интересно, так как существуют не менее образованные штаты (например, Массачусетс), в которых подобной кластеризации не прослеживается. Анализ на конкретных примерах роли образования в сравнении поведения избирателей помог бы прояснить значимость этого вопроса в местном масштабе.

Перейдём к рассмотрению смешанных кластеров. Первый кластер, который мы рассмотрим, визуализирован на карте 6d, и показывает пересечение влияния процента женщин и процента рождённых иностранцев по происхождению. Можем заметить кластеризацию с положительной связью на западном побережье США в штате Орегон. На карте 6e мы видим кластер в юго-западной части Техаса и в регионе Новой Англии, который свидетельствует о пересечении процента людей, живущих за чертой бедности и процента людей с высшим образованием. Эти комбинации выглядят весьма нетривиально, так как рассматривают по две разные группы населения, и, тем не менее, они имеют общие уровни кластеризации, что может являться стимулом для дальнейших исследований.

Заключение

Проведённое исследование, рассматривающее кластеры влияния социально-экономических факторов на голосование на президентских выборах за 2012 и 2016 годы позволяет делать ряд важных выводов относительно важности пространственной неоднородности и избирательного поведения в США.

Первой гипотезой являлось то, что влияние социально-экономических факторов на голосование является географически нестационарным. С помощью метода географически взвешенной регрессии мы выяснили, что все независимые переменные оказывают разное влияние на голосование за Демократическую партию, формируя кластеры в пространстве.

В том числе путём построения регрессий мы выяснили, что географическая регрессия имеет намного больший объяснительный потенциал. Локальная регрессия действительно способна снижать количество и величину значения ошибок, и,

что самое интересное, с помощью неё нам действительно удалось снизить уровень ошибок в конкретных областях. Так, наша вторая гипотеза также подтвердилась.

В дальнейшем мы изучили географию независимых переменных и выяснили, что они кластеризуются в пространстве. Вся визуализация была выполнена в виде карт, на которых отображались округа с показателями коэффициента, и мы обнаружили статистически значимые кластеры по каждой переменной. Таким образом, мы путём визуализации многоуровневых карт обнаружили не только одиночные кластеры, но и связную концентрацию эпицентров социальных и смешанных факторов. Наши две последние гипотезы также подтвердились.

Продолжением нашего исследования может стать поиск причин возникновения пересечения кластеров методом кейс-стади, либо с помощью количественного исследовательского дизайна с адаптацией подхода, ранее используемого в исследовании нетипичных значений электорального поведения [Morrill et al., 2011]. Другим возможным подходом при дальнейшем исследовании может являться подход с поиском социальных расколов по выявленным кластерам, подавляющее число которых пересекает границы штатов, что обеспечивает использовании менее типичного подхода.

СПИСОК ЛИТЕРАТУРЫ:

1. Ahmed R., Pesaran M.H. (2020), *Regional heterogeneity and US presidential elections*, CESifo Working Paper, no. 8615, Munich: Munich Society for the Promotion of Economic Research, 74 p.
2. Akaike H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723.
3. Brunsdon C., Fotheringham A.S., Charlton M.E. (1996), Geographically weighted regression: a method for exploring spatial non-stationarity, *Geographical Analysis*, vol. 28, no. 4, pp. 281–298.
4. Campbell A. (1980), *The American Voter*, Chicago: University of Chicago Press, 576 p.
5. Cho W.K.T., Gimpel J.G. (2009), Presidential voting and the local variability of economic hardship, *The Forum*, vol. 7, no. 1.
6. Cox K.R. (1970), Residential relocation and political behavior: Conceptual model and empirical tests, *Acta Sociologica*, vol. 13, no. 1, pp. 40–53.
7. Evans G. (2000), The continued significance of class voting, *Annual Review of Political Science*, vol. 3, no. 1, pp. 401–417.
8. Forest B. (2018), *Electoral geography: From mapping votes to representing power*, *Geography Compass*, vol. 12, no. 1, e12352.
9. Fotheringham A.S., Brunsdon C., Charlton M.E. (2009), Geographically weighted regression, *The Sage Handbook of Spatial Analysis*, eds. Fotheringham A.S., Rogerson P.A., London: SAGE, pp. 243–254.
10. Hajnal Z.L., Lee T. (2011), *Why Americans don't join the party: Race, immigration, and the failure of political parties to engage the electorate*, Princeton: Princeton University Press, 320 p.
11. Hamilton L.C. (1992), *Regression with graphics: A second course in applied statistics*, Belmont, CA: Duxbury Press, 363 p.
12. Hillier B., Hanson J., Peponis J. (1984), What do we mean by building function?, *Designing for building utilisation*, eds. Powell J.D., Cooper I., Lera S., London: E & FN Spon Ltd, pp. 61–72.
13. Ishizawa H., Stevens G. (2007) Non-English language neighborhoods in Chicago, Illinois: 2000, *Social Science Research*, vol. 36, no. 3, pp. 1042–1064.
14. Iyengar S., Westwood S.J. (2015), *Fear and*

- loathing across party lines: New evidence on group polarization, *American Journal of Political Science*, vol. 59, no. 3, pp. 690–707.
15. Jana M., Sar N. (2016), Modeling of hotspot detection using cluster outlier analysis and Getis-Ord Gi* statistic of educational development in upper-primary level, *India, Modeling Earth Systems and Environment*, vol. 2, no. 2, pp. 1–10.
 16. Johnston R., Jones K., Propper C., Burgess S. (2007), Region, local context, and voting at the 1997 general election in England, *American Journal of Political Science*, vol. 51, no. 3, pp. 640–654.
 17. Kinder D.R., Kiewiet D.R. (1979), Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting, *American Journal of Political Science*, vol. 23, pp. 495–527.
 18. Kramer G.H. (1971), Short-term fluctuations in US voting behavior, 1896–1964, *American Political Science Review*, vol. 65, no. 1, pp. 131–143.
 19. Lewis-Beck M.S. (1988), *Economics and elections: The major Western democracies*, Ann Arbor: University of Michigan Press, 183 p.
 20. Lind J.T. (2007), Does permanent income determine the vote?, *The BE Journal of Macroeconomics*, vol. 7, no. 1.
 21. Menard S. (2002), *Applied logistic regression analysis*, Thousand Oaks: SAGE, 128 p.
 22. Moran P.A. (1948), The interpretation of statistical maps, *Journal of the Royal Statistical Society. Series B*, vol. 10, no. 2, pp. 243–251.
 23. Ord J.K., Getis A. (1995), Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, vol. 27, no. 4, pp. 286–306.
 24. Paez A. (2004), Anisotropic variance functions in geographically weighted regression models, *Geographical Analysis*, vol. 36, no. 4, pp. 299–314.
 25. Steenbergen M.R. (2010), Decomposing the vote: Individual, communal, and cantonal sources of voting behavior in Switzerland, *Swiss Political Science Review*, vol. 16, no. 3, pp. 403–424.
 26. Sui D.Z., Hugill P.J. (2002), A GIS-based spatial analysis on neighborhood effects and voter turnout: A case study in College Station, Texas, *Political Geography*, vol. 21, no. 2, pp. 159–173.
 27. Tenn S. (2007), The effect of education on voter turnout, *Political Analysis*, vol. 15, no. 4, pp. 446–464.
 28. Tu J., Xia Z.G. (2008), Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation, *Science of the Total Environment*, vol. 407, no. 1, pp. 358–378.
 29. Weber D., Englund E. (1992), Evaluation and comparison of spatial interpolators, *Mathematical Geology*, vol. 24, pp. 381–391.

REFERENCES:

1. Ahmed R., Pesaran M.H. (2020), Regional heterogeneity and US presidential elections, CESifo Working Paper, no. 8615, Munich: Munich Society for the Promotion of Economic Research, 74 p.
2. Akaike H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723.
3. Brunsdon C., Fotheringham A.S., Charlton M.E. (1996), Geographically weighted regression: a method for exploring spatial non-stationarity, *Geographical Analysis*, vol. 28, no. 4, pp. 281–298.
4. Campbell A. (1980), *The American Voter*, Chicago: University of Chicago Press, 576 p.
5. Cho W.K.T., Gimpel J.G. (2009), Presidential voting and the local variability of economic hardship, *The Forum*, vol. 7, no. 1.
6. Cox K.R. (1970), Residential relocation and political behavior: Conceptual model and empirical tests, *Acta Sociologica*, vol. 13, no. 1, pp. 40–53.
7. Evans G. (2000), The continued significance of class voting, *Annual Review of Political*

- Science, vol. 3, no. 1, pp. 401–417.
8. Forest B. (2018), Electoral geography: From mapping votes to representing power, *Geography Compass*, vol. 12, no. 1, e12352.
 9. Fotheringham A.S., Brunsdon C., Charlton M.E. (2009), Geographically weighted regression, *The Sage Handbook of Spatial Analysis*, eds. Fotheringham A.S., Rogerson P.A., London: SAGE, pp. 243–254.
 10. Hajnal Z.L., Lee T. (2011), Why Americans don't join the party: Race, immigration, and the failure of political parties to engage the electorate, Princeton: Princeton University Press, 320 p.
 11. Hamilton L.C. (1992), *Regression with graphics: A second course in applied statistics*, Belmont, CA: Duxbury Press, 363 p.
 12. Hillier B., Hanson J., Peponis J. (1984), What do we mean by building function?, *Designing for building utilisation*, eds. Powell J.D., Cooper I., Lera S., London: E & FN Spon Ltd, pp. 61–72.
 13. Ishizawa H., Stevens G. (2007) Non-English language neighborhoods in Chicago, Illinois: 2000, *Social Science Research*, vol. 36, no. 3, pp. 1042–1064.
 14. Iyengar S., Westwood S.J. (2015), Fear and loathing across party lines: New evidence on group polarization, *American Journal of Political Science*, vol. 59, no. 3, pp. 690–707.
 15. Jana M., Sar N. (2016), Modeling of hotspot detection using cluster outlier analysis and Getis-Ord Gi* statistic of educational development in upper-primary level, India, *Modeling Earth Systems and Environment*, vol. 2, no. 2, pp. 1–10.
 16. Johnston R., Jones K., Propper C., Burgess S. (2007), Region, local context, and voting at the 1997 general election in England, *American Journal of Political Science*, vol. 51, no. 3, pp. 640–654.
 17. Kinder D.R., Kiewiet D.R. (1979), Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting, *American Journal of Political Science*, vol. 23, pp. 495–527.
 18. Kramer G.H. (1971), Short-term fluctuations in US voting behavior, 1896–1964, *American Political Science Review*, vol. 65, no. 1, pp. 131–143.
 19. Lewis-Beck M.S. (1988), *Economics and elections: The major Western democracies*, Ann Arbor: University of Michigan Press, 183 p.
 20. Lind J.T. (2007), Does permanent income determine the vote?, *The BE Journal of Macroeconomics*, vol. 7, no. 1.
 21. Menard S. (2002), *Applied logistic regression analysis*, Thousand Oaks: SAGE, 128 p.
 22. Moran P.A. (1948), The interpretation of statistical maps, *Journal of the Royal Statistical Society. Series B*, vol. 10, no. 2, pp. 243–251.
 23. Ord J.K., Getis A. (1995), Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, vol. 27, no. 4, pp. 286–306.
 24. Paez A. (2004), Anisotropic variance functions in geographically weighted regression models, *Geographical Analysis*, vol. 36, no. 4, pp. 299–314.
 25. Steenbergen M.R. (2010), Decomposing the vote: Individual, communal, and cantonal sources of voting behavior in Switzerland, *Swiss Political Science Review*, vol. 16, no. 3, pp. 403–424.
 26. Sui D.Z., Hugill P.J. (2002), A GIS-based spatial analysis on neighborhood effects and voter turnout: A case study in College Station, Texas, *Political Geography*, vol. 21, no. 2, pp. 159–173.
 27. Tenn S. (2007), The effect of education on voter turnout, *Political Analysis*, vol. 15, no. 4, pp. 446–464.
 28. Tu J., Xia Z.G. (2008), Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation, *Science of the Total Environment*, vol. 407, no. 1, pp. 358–378.
 29. Weber D., Englund E. (1992), Evaluation and comparison of spatial interpolators, *Mathematical Geology*, vol. 24, pp. 381–391.