



В США с 1787 года используется двухпартийная система, где кандидаты на выборах представляют либо Демократическую, либо Республиканскую партию. Президентские выборы проходят каждые четыре года. В этом датасете представлены данные о выборах президента США в 2016 году.

Голосование проходит в два этапа: первый — граждане выбирают «выборщиков» отдельно в каждом штате. Второй — «выборщики» голосуют за конкретного кандидата. В результате подсчета голосов в каждом штате тот кандидат, который набрал больше голосов, получает все голоса штата. В конечном счете голоса суммируются по всем штатам. Голосовать может любой гражданин, достигший возраста 18 лет.

Однако граждане на первом этапе голосования видят не некую коллегию избирателей, а конкретных кандидатов в президенты, и голосуют либо за одного кандидата, либо за всех кандидатов от одной партии — оба варианта равнозначны.

Таким образом, в нашей базе мы имеем данные о том, сколько человек проголосовали за партию в каждом округе штата и характеристики населения округа по демографии, финансам и географии. Округа штатов сильно различаются по этим характеристикам. Также различаются характеристики одного штата по годам. В датасете представлены данные за 2010, 2014 и 2016 годы.

Информация дана по ОКРУГАМ штатов, НЕ по штатам! Для некоторых показателей данные представлены за один год, для некоторых — за период (2009–2013).

| Переменная | Описание |
|------------|------------------------------------------------------------------------------------------------------|
| area_name | Название округа |
| state_abbr | Аббревиатура — обозначение штата |
| PST045214 | Численность населения округа на 2014 год |
| AGE135214 | Процент детей младше 5 лет, 2014 год |
| AGE295214 | Процент детей до 18 лет, 2014 год |
| AGE775214 | Процент людей старше 65 лет, 2014 год |
| SEX255214 | Кол-во женщин, 2014 год |
| RHI125214 | Количество белого населения, 2014 год |
| RHI225214 | Процент афроамериканцев, 2014 год |
| RHI325214 | Процент коренного населения (индейцы и коренные жители Аляски), 2014 год |
| RHI425214 | Процент азиатов, 2014 |
| RHI525214 | Процент коренного населения Гавайских островов и других Тихоокеанских островов, 2014 |
| RHI625214 | Процент метисов (чьи предки двух и более рас), 2014 |
| RHI725214 | Процент испанского и латиноамериканского населения, 2014 |
| RHI825214 | Процент белого населения (не испанцы и не латиноамериканцы), 2014 |
| POP715213 | Процент людей, проживающих в одном и том же доме год и более, 2009—2013 |
| POP645213 | Процент людей, рожденных не в США, 2009—2013 |
| POP815213 | Процент детей до 5 лет, у которых дома говорят более чем на одном языке, 2009—2013 |
| EDU635213 | Процент людей старше 25, окончивших как минимум старшую школу, 2009—2013 |
| EDU685213 | Процент людей старше 25, получивших минимум степень бакалавра, 2009—2013 |
| VET605213 | Численность ветеранов войны, 2009—2013 |
| LFE305213 | Среднее время в минутах на дорогу от дома для работы (данные по работникам старше 16 лет), 2009—2013 |
| HSG010214 | Число жилых помещений, 2014 |
| HSG445213 | Уровень домовладения (какой процент людей владеет собственной недвижимостью), 2009—2013 |

| Переменная | Описание |
|------------|----------------------------------------------------------------------------------------------------------------------|
| HSG096213 | Процент жилых помещений в многоквартирных домах, 2009—2013 |
| HSD410213 | Количество домашних хозяйств, 2009—2013 |
| HSD310213 | Сколько человек приходится на одно домашнее хозяйство, 2009—2013 |
| INC910213 | Денежный доход на душу населения за последние 12 месяцев (в долларах 2013), 2009—2013 |
| INC110213 | Медианный доход домашних хозяйств, 2009—2013 |
| PVY020213 | Процент людей, находящихся ниже уровня бедности, 2009—2013 |
| BZA010213 | Количество частных несельскохозяйственных предприятий, 2013 |
| BZA110213 | Количество занятых на частных несельскохозяйственных предприятиях, 2013 |
| BZA115213 | Процентное изменение количества занятых на частных несельскохозяйственных предприятиях, 2012—2013 |
| NES010213 | Кол-во предприятий, не являющихся работодателями, 2013 |
| BPS030214 | Кол-во выданных разрешений на строительство, 2014 |
| LND110210 | Площадь округа в квадратных милях, 2010 |
| Demvotes16 | Число голосов за кандидата от Демократической партии, 2016 |
| GOPvotes16 | Число голосов за кандидата от Республиканской партии, 2016 |
| total_2016 | Общее число голосов в 2016 году |
| pct_dem_16 | Процент голосов за Демократическую партию от общего числа голосов, 2016 |
| pct_gop_16 | Процент голосов за Республиканскую партию от общего числа голосов, 2016 |
| diff_2016 | Разница в числе голосов между Республиканской и Демократической партиями |
| pct_pt_16 | Разница (diff_2016) в процентах от общего числа. Отрицательная, если за Демократическую партию отдали меньше голосов |

В датасете много колонок. Чтобы было легче удержать их в голове, советуем:

- Сгруппировать данные по годам, например все переменные 2014 года.
- Сгруппировать данные и по областям знаний. Например, какой набор переменных отвечает за расу населения, какой — за материальное состояние, экономическое благополучие и так далее.

Так будет проще ориентироваться, в какой блок вы хотите побольше углубиться.

Также советуем визуализировать данные с помощью графиков, чтобы делать предположения на основании всей статистики, а не нескольких примеров.

Каким образом социально-демографическая и экономическая ситуация в стране повлияла на результаты выборов?

Какие факторы, по вашему мнению, могут оказывать наибольшее влияние на выбор президента? Определите для себя такие факторы и посмотрите, каким образом они повлияли на результаты выборов в Америке.

1. Какие признаки можно использовать для анализа причины различия мнений при выборах президента США по округам? Например, для одного округа будет более значимо расовое рассмотрение, а для другого — неравенство в доходах.
2. Какие характеристики избирателей могут быть связаны с их голосованием на выборах президента США по округам? Например, где-то влияет раса, а где-то — достаток.
3. По-разному ли голосуют разные округа? Что может быть причиной различий?
4. Какое распределение голосов было устроено на выборах президента США по округам? Есть ли какие-то особенности в этом распределении? Попробуйте построить график и посмотреть на разницу в голосах среди разных регионов.
5. Как можно классифицировать округа по результатам выборов президента США? Есть ли какие-то группы округов с похожими результатами голосования?
6. Какие факторы могут влиять на результаты выборов президента США по округам? Есть ли какие-то зависимости между этими факторами и результатами голосования? Возможно, и наиболее значимые факторы там будут схожи — или нет :) Проверьте эти закономерности.
7. Как можно использовать данные о результатах выборов президента США по округам для прогнозирования будущих выборов? На какие социально-демографические группы тогда стоит ориентироваться партиям при выстраивании предвыборной кампании?