2020 US CES Voter File Dataset

Julian Marrero

April 3, 2024

1 Motivation

Why was the dataset created?

The dataset was created to facilitate advanced analysis of voting behavior, political attitudes, and demographic correlates in the 2020 US elections. Its purpose is to aid researchers, policy-makers, and political analysts in understanding the complex dynamics of electoral participation and political polarization.

Who funded the creation of the dataset?

Funded by a collaborative grant from the Democracy Research Institute and the Public Engagement Foundation.

2 Composition

What does the dataset represent?

Individual-level data from registered US voters, combining voter file information with responses from the 2020 Cooperative Election Study.

What and who is included in the dataset?

It includes data on 200,000 registered voters across the United States who participated in the CES and whose records were matched with voter file information.

How many instances (e.g., interviews, profiles) are in the dataset?

200,000 voter profiles.

Are there any missing values?

Yes, approximately 5% of records have one or more missing values in non-critical fields. Critical fields have been imputed or flagged.

3 Collection Process

How was the data collected?

Voter file data were licensed from SecureVoteDataCorp. The CES data were collected via an online survey platform, targeting a stratified sample of registered voters to ensure representativeness.

Who was involved in the data collection process?

The National Center for Political Studies, SecureVoteDataCorp for voter files, and QualiSurvey Inc. for administering the CES.

Over what timeframe was the data collected?

Voter file data as of November 2020. CES data were collected in November 2020, immediately following the presidential election.

4 Preprocessing/Cleaning/Labeling

Was any preprocessing, cleaning, or labeling of the data done?

Yes, voter files were standardized to a common format, and personal identifiers were encrypted. CES responses were cleaned for inconsistencies and merged with voter files using a secure matching algorithm.

5 Post-stratification

How was post-stratification performed?

Based on demographic and geographic information from the voter file, weights were applied to CES responses to adjust for sampling biases and ensure representativeness of the national voting population.

6 Recommended Uses

For what tasks is the dataset suitable?

Suitable for academic research, policy analysis, and non-commercial public interest research focused on voting behavior, political engagement, and electoral democracy.

What tasks is the dataset not suitable for?

Not suitable for individual voter targeting, commercial purposes, or any activity that could compromise voter privacy or the integrity of the electoral process.

7 Ethical Considerations

What ethical considerations were taken into account?

Consent was obtained from CES participants, data anonymization was applied, and access is restricted to approved researchers. An ethics review was conducted to ensure compliance with privacy laws and ethical research standards.

Are there any legal concerns?

Complies with applicable privacy laws and regulations, including GDPR for any EU citizens included in the dataset.

8 Access

How can one access the dataset?

Access is granted on a case-by-case basis to researchers affiliated with accredited academic or non-profit institutions, subject to a data use agreement that stipulates ethical use and protection of privacy.

9 Citation

How should the dataset be cited?

National Center for Political Studies (2024). 2020 US CES Voter File Dataset, Version 1.0. [DOI/Publisher].

10 Usage Liscence

Under what license is the dataset distributed?

This dataset is distributed under the Academic Non-Commercial Use License, which allows for non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and used for academic or research purposes only.