User's Guide and Codebook for the ANES 2020 Time Series Voter Validation Supplemental Data

American National Election Studies
The University of Michigan and Stanford University
September 2, 2022

Suggested Citation

ANES. 2022. User's Guide and Codebook for the ANES 2020 Time Series Voter Validation Supplemental Data. Ann Arbor, MI and Palo Alto, CA: the University of Michigan and Stanford University.

Acknowledgments

This study was funded by the National Science Foundation (www.nsf.gov) under grants SES-1835731 to the University of Michigan and SES-1835022 to Stanford University. Any opinions, findings and conclusions, or recommendations expressed in these materials are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The Principal Investigators for the 2020 ANES were Ted Brader (University of Michigan) and Shanto Iyengar (Stanford University). The Associate Principal Investigators were D. Sunshine Hillygus, Daron Shaw, and Nicholas Valentino. The directors of Stanford and Michigan operations, respectively, were Matthew DeBell and David Howell.

The 2020 vote validation process was directed by Matthew DeBell. The vote validation case matching for Vendor 1 was performed by NORC at the University of Chicago. We thank Jennifer Carter, David Dutwin, Dean Resnick, and Mark Watts for their contributions to the voter validation data processing. Matching for Vendor 2 was performed at Stanford University by Matthew Tyler. Manual review procedures were performed at the University of Michigan by Macey Owen and Trent Ingell and supervised by Jaime Ventura.

This documentation reprints some material from previous ANES documentation without explicit attribution.

Contact

The ANES website address is www.electionstudies.org
ANES sends occasional updates on Twitter @electionstudies
Any questions not answered on the ANES website or by this report may be directed to ANES staff by email at anes@electionstudies.org

Contents

Introduction and Purpose of the Dataset

This is documentation for the ANES 2020 Time Series Voter Validation Supplemental Data. The data file provides data on voter registration and turnout that were compiled from official voting records by commercial vendors and were then matched to the ANES 2020 Time Series Study's sample. The data are provided for purposes of methodological evaluation and for the analysis of voter registration and turnout status. This documentation describes the matching procedures and resulting data.

Several commercial vendors compile voter data for sale for use in campaigns. Working under a contract with the ANES, NORC at the University of Chicago obtained access to a national voter file from one of these vendors and used a probabilistic matching process to link ANES respondents and sampled addresses to their corresponding records on this voter file. These data are referred to as "Vendor 1." Separately, ANES obtained access to a national voter file from a different vendor and used a different probabilistic matching process to link ANES respondents and sampled addresses to their corresponding records on this second voter file. These data are referred to as "Vendor 2."

This file includes voter data of two distinct types: person records and address records. Address records apply to the addresses that were sampled for the ANES 2020 Time Series Study. Person records apply to individuals who were sampled at the sampled addresses. Because individuals were not sampled at all addresses due to nonresponse, person records do not exist for all sample units. Person records of turnout are intended for comparison to self-reported turnout. Address records of turnout, which indicate the number of voters associated with a sampled address, may be used to analyze nonresponse, the household context for sampled persons, or for other purposes.

ANES makes no representations about the quality of the voter data from its commercial sources. Users should be aware that evaluations of the quality of such data are controversial, in part because such vendors, as commercial political campaign operatives, do not build or document their datasets for research purposes and do not follow norms of scientific transparency. Errors from several sources may affect the accuracy of voter validation data. Source of errors in these data include data entry and transcription errors when individuals complete voter registration forms and government agencies enter data from registration forms into their databases, data entry errors when elections officials record voter turnout, imputation errors when data vendors use statistical models to predict and impute demographic characteristics for individuals, address matching errors when ANES and vendor personnel matched ANES sampled addresses to records on vendor files, person matching errors when ANES personnel matched the names, ages, and birth dates of ANES respondents to records on vendor files, and other data processing errors. Combinations of such errors may be common in these data, and one of the purposes of collecting the data was to evaluate its value.

Contents of the Dataset

The dataset has 18,430 cases, corresponding to the ANES sample, excluding cases from the General Social Survey and a small number of additional cases for which critical identification data were missing.

The dataset contains 23 variables, as follows.

version. ANES 2020 Voter Validation dataset version. This variable identifies the version of the vote validation dataset release. The first release version is 20220902 (produced September 2, 2022).

V200001. ANES 2020 case ID. This case identification number corresponds to the case IDs on the ANES 2020 Time Series Study and Methodology data files.

val1_addmatch. Address match found (vendor 1). This variable indicates whether or not a match for the ANES sample address was found on the vendor's voter file.

val1_number. Number of linking persons (vendor 1). This variable indicates the number of individual persons associated with a linked address on the vendor's voter file. Addresses with large number of linked individuals may indicate the residences of multiple households. When this occurs, it is frequently a result of matching all of the housing units at a basic street address, that is, of including everyone in an apartment building instead of the residents of a single sampled apartment. Cases with more linked persons than there are adults in a typical household (approximately 2 to 4) may warrant analysis with extra care.

val1_matchprob. Estimated person match probability (vendor 1). This indicates the estimated probability that the matched individual from the vendor's file is person identified on the ANES file. Match probabilities below approximately .900 to .990 warrant caution. Cases may be weighted by this match probability.

val1_turnout16. Voting status, 2016 general election (vendor 1). This indicates match status and turnout status for the 2016 general (presidential) election.

val1_turnout20. Voting status, 2020 general election (vendor 1). This indicates match status and turnout status for the 2020 general (presidential) election.

val1_hh_turnout16. Household turnout status 2016 (vendor 1). This differs from the other 2016 turnout status by indicating the status for the sampled address and the number of indicated voters at the sampled address.

val1_hh_turnout20. Household turnout status 2020 (vendor 1). This differs from the other 2020 turnout status by indicating the status for the sampled address and the number of indicated voters at the sampled address.

The following seven variables are for the data from Vendor 2 and correspond to the variables for Vendor 1.

val2_addmatch. Address match found (vendor 2).

val2_number. Number of linking persons (vendor 2).

val2_matchprob. Estimated person match probability (vendor 2).

val2_turnout16. Voting status, 2016 general election (vendor 2).

val2_turnout20. Voting status, 2020 general election (vendor 2).

val2_hh_turnout16. Household turnout status 2016 (vendor 2).

val2_hh_turnout20. Household turnout status 2020 (vendor 2).

The following seven variables are unique to vendor 2 and describe the outcomes of the clerical review process. The review process is described later in this guide.

val2_match. Final adjudicated outcome of manual review of the person matches (vendor 2). This variable is recommended to distinguish between valid matches (coded 1) and invalid matches (coded 0). Data users may wish to exclude the invalid matches from analysis.

val2_c1_minor. For coder 1, tally of 'minor' differences between the ANES sample record and the best match from vendor 2.

val2_c1_major. For coder 1, tally of 'major' differences between the ANES sample record and the best match from vendor 2.

val2_c1_match. For coder 1, subjective judgment of whether the best match from vendor 2 is correctly matched to the ANES record.

val2_c2_minor. For coder 2, tally of 'minor' differences between the ANES sample record and the best match from vendor 2.

val2_c2_major. For coder 2, tally of 'major' differences between the ANES sample record and the best match from vendor 2.

val2_c2_match. For coder 2, subjective judgment of whether the best match from vendor 2 is correctly matched to the ANES record.

Probabilistic Linkage Procedure for Vendor 1

The following documentation under this heading was provided by NORC to describe the linkage of the ANES sample records to the voter file vendor's database.

To conduct this linkage, we used a foundational methodology to this field formalized mathematically by Ivan Fellegi and Alan Sunter in 1969 from previous work of other researchers. This method works in the following manner:

- For two files being compared, comparison variables (i.e., from names, date-of-birth fields, and other identification fields) shared on the two files are identified.
- For each comparison variable (e.g. year-of-birth or phone number), the probability of agreement of the two values (i.e., one from each file) being compared are estimates for two distinct sets of pairs.
    - Pairs that are matched: i.e. they (truly) represent the same person
    - Pairs that are unmatched: i.e., they do not represent the same person.

    So, for first name, the m-probability is the probability that two records representing the same person have the same value for it. Reasonably this probability is quite high, but

disagreement can occur because of misspellings, nicknames, substitution of middle names, etc.

Again for first name, the u-probability is the probability that two records not representing the same people agree on names. This probability is going to be quite low but dependent on the commonness of the names.

- Using the m- and u- probabilities, agreement and disagreement weights are assigned to each variable based on a formula developed even prior to work of Fellegi and Sunter (based on the statistical model structure).
    - The agreement weight (for each of comparison variable) is computed as $AW = log_2\left(\frac{M}{U}\right)$.
        - M – estimated m-probability; U – estimated u-probability
    - The disagreement weight is computes as $DW = log_2\left(\frac{(1-M)}{(1-U)}\right)$
- For each pair being analyzed, we analyze each comparison field (i.e., year of birth, last name, phone number):
    - For a given pair, if the values of the field agree on the two records composing the pair, we assign a variable weight equal to the agreement weight.
    - If the values of the field disagree, we assign a variable weight equal to the disagreement weight.
    - If it is not possible to compare the values (because either are missing or invalid) then a variable weight of 0 is assigned.
- Then for each pair, we sum all the variable weights to generate a pair weight for it.
- In the most basic use of the Fellegi-Sunter technique all pairs scoring above a set cut-off value are linked (i.e., imputed to be true matches). Clearly then, the cut-off value has to be set carefully and one possible way is to set it based on human judgment: looking at pairs ranked by pair score and seeing where it seems the best place to draw the line between what is imputed a match and what is imputed a non-match.

For this analysis, we convert the pair weight into a probability of being a true match. This conversion is based on the assumptions that

- Agreement statuses for the various variables are statistically independent of each other given matched/unmatched status.
- m- and u- probabilities can be accurately estimated
- The proportion of pairs in the set being analyzed can be accurately estimated.

If these assumptions hold then match probability is estimated as follows:

1. An adjustment factor applicable to a set of pairs under analysis is computed as $Adj = log_2\left(\frac{N_M}{N_U}\right)$, where N_M is the number of matched pairs and N_U is the number of unmatched pairs.
2. We compute the adjusted pair weight: $PW_{Adj} = PW + Adj$, where PW is the pair weight computed by summing the variable weights computed for a specific pair and PW_{Adj} is the adjusted pair weight.

3. We compute the pairs' odds of being a match: $Odds_M = 2^{PW_{Adj}}$ where $Odds_M$ is the Odds that the pair is a match. A value of $Odds_M = 3$ says that the odds that a pair is match is 3:1.
4. We compute the match probability as $P(M) = \frac{Odds_M}{1+Odds_M}$

   $P(M)$ is the match probability for a pair. So if the odds were 3 (or 3:1), then $P(M) = \frac{3}{4} = 75\%$.

Again this computation is only accurate to the degree that the assumptions stated above are being met. We will note that some variable agreements are typically correlated, such as sex/gender and first name: when first name agrees, there is a much higher probability that sex/gender will agree (i.e., these agreement statuses are not statistically independent). It is for this reason that we try not to use sex/gender as a comparison variable.

There are several reasons why it is best to use a match probability than a raw pair weight to assess match status:

- Can be used to make error estimates
- Allow pairs developed in different linkage passes to be compared
- Makes understanding accuracy of match status estimates more understandable.

For this analysis, for each sample person record, we keep the pair with the highest match probability assuming it is greater than 97.5%.

To conduct the linkage, we use SAS code (NORCLink) that NORC developed internally to apply the Fellegi-Sunter methodology.

There are several extensions to this methodology that have been implemented with the NORC code:

- Additional pre-processing is performed to clean the data fields being compared
- Nickname substitution are made for commonly used nicknames to generate alternate records to be included in analysis
- For first and last names (but only for exact agreement), we substitute name frequencies computed within the data for the estimated u-probabilities. This means that the u-probability for a common name will be much higher than for an unusual name, and in turn that means that the agreement weight computed for it will be substantially higher. The idea is that agreement on an uncommon name is much more indicative of match status than agreement on a common name.
- For name fields consisting only of an initial, we conduct analysis based on agreement of that initial but the m- and u- probabilities and agreement and disagreement weights are specific to name-initial comparison.
- Name comparisons are made based on measured string similarity. We use the Jaro-Winkler similarity score which ranks similarity from 0 – no letters are in common in the values (for name) being compared to 1 – the values for the name are exactly the same.
- Name comparisons are made according to several levels of similarity:
    - Jaro-Winkler similarity ≥ .85
    - Jaro-Winkler similarity ≥ .90

- Jaro-Winkler similarity ≥ .95
- Jaro-Winkler similarity = 1.00 (exactly the same)

The use of multiple levels is meant to enhance the precision of name comparisons. Note that when name values have similarity < .85 they are treated as though they are completely dissimilar. The m- and u- probabilities and agreement and disagreement weights are computed identically to other comparison variable agreements. If the values have similarity ≥ .85 they result in the assigned variable weight being the agreement weight computed for this level based on the corresponding m- and u- probabilities. So they m-probability for the .85 level is the estimated probability that records for the same person have a name similarity of .85 or greater.

For the .90 level, m- and u- probabilities and agreement and disagreement weights are computed contingent on (i.e., conditionally) on having met the .85 level. This is to say that the m-probability for the .90 level is the probability of having a similarity score ≥ .85 or, mathematically, $m = P(JW \geq .90 | JW \geq .85)$. If for the pair being analyzed, the similarity score < .85, then no variable weights are assigned for the levels higher than it.

The logic used for the .90 is shared for the similarity levels above it. Note analysis for each level above .85 (i.e., .90, .95, 1.00) is applied contingent on having met the immediately lower level.

Specification for this linkage analysis
There were two methods used to pull Vendor 1 voter registration records to include in the linkage analysis:

- Based on agreement on street-level address (NORC sent Vendor 1 every address we have ever had on file for each panelist; NORC did not include apartment/unit information; Vendor 1 returned every record they had for any of these addresses, including every apartment/unit record for a given main address)
- Based on first and last name agreement and several other non-geographic variable agreements (NORC sent Vendor 1 the name, gender, and age of every panelist; Vendor 1 returned any record with a proprietary fuzzy name match plus gender (match or missing) and age (within 5 year range)

The linkage process was run separately for these two files as they are expected to have different linkage characteristics (particularly matching parameters).

For each of these runs, we used one or more blocking passes. Blocking is a commonly used strategy for record linkage. It indicates that only records falling in the same block are formed into pairs for analysis. For instance for the address-based run, blocking was made by sex and ZIP code. Within these blocks (defined by specific combinations of sex and ZIP code), every sample record included in the analysis was paired with every record in the street-level address-returned Vendor records: this is known as a Cartesian product. Blocking is used to reduce the computational load that would result from producing and analyzing all possible pairs across the files being linked. Note that records with different values for sex are never compared so a transcription error on this field (or persons undergoing sex change) will not have pairs returned for them.

These were the blocking passes that were conducted by run:

- Vendor 1 street-level address pull run
    1. Sex and ZIP Code
- Name agreement pull run
    1. Sex, State, ZIP code (first 3 digits)
    2. Sex, Year-of-Birth, Last Name (1st character, only)
    3. Sex, First Name (1st character, only), Month-of-birth, Day of Birth

For each blocking pass, m- and u- probabilities and agreement and disagreement weights are estimated or set separately. However, essentially the same comparison variables were used in each, except the street-level address pull run did not use email address as this was not available:

- First Name (or initial if only this was available)
    - Based on Jaro-Winkler comparison score, compared to levels
- Last Name (or initial if only this was available)
    - Based on Jaro-Winkler comparison score, compared to levels
- Day of Birth (i.e., the numerical day of the month)
- Month of Birth
- Year of Birth
- Phone Number (full 10 digits)
- email Address (only for name agreement pull run)

The results of all these runs were combined, and for each sample person record, we retained only the pairing with the highest estimated match probability if that probability was greater than 97.5%.

Within each pass, parameters (m- and u- probabilities and estimated number of true matches) were estimated used a machine learning methodology known as the Expectation Maximization Algorithm (EM, for short) which is a well-developed and researched method of obtaining these estimates under the Fellegi-Sunter paradigm.

Probabilistic Linkage Procedure for Vendor 2

Record linkage for Vendor 2 was performed in two stages. First, an automatic probabilistic matching procedure was performed. Second, results from the probabilistic procedure were manually reviewed to classify each case as correctly matched or not matched.

Overview of the automatic matching procedure

Probabilistic linkage for Vendor 2 was performed by ANES using the R package *fastLink*. Prior to matching, parsers were applied to Vendor 2 and ANES name and address variables. Person matching was performed within state-gender blocks based on first name, last name, age, street name, house number, and ZIP code. Address matching was performed within state blocks based on street name, house number, and ZIP code; this matched at the household level, allowing matches to one or more voters at the same address. Finally, all match pairs with match probabilities less than 10% (0.10) were discarded.

Details of the automatic matching procedure

- We used the R package fastLink 0.6.0.
- Full names were parsed into first name, middle name, and last name using the HumanName() function from the Python package nameparser 1.0.06. All names were converted to lower case for future comparisons.
- Age in years for the ANES respondents was computed based on the reference date March 3, 2021, since this corresponded to timestamp of the typical Vendor 2 record. Age from the vendor file was already computed by the vendor.
- Addresses from the ANES and Vendor 2 files were parsed into house number, street name, and zip code using with the tag() function from the Python package usaddress 0.5.10. This procedure keeps only the essential address information and drops directional indicators, apartment numbers, etc.
- Zip codes were truncated to 5-character strings where applicable.
- Matching occurs on households (1+ voters at the same street name + house number + zip code). Households were constructed using exact matching on street name, house number, and zip code.
- There were two mutually exclusive and exhaustive gender categories: self-reported male or other.
- Person-matching was attempted for ANES respondents whenever in cases where an individual's name was recorded.
- Person-matching was performed in two rounds: first, a within-state round blocking on state and gender. Second, an out-of-state round blocked only on gender.
- Within-state person-matching: following the 2016 vote validation, we "use three levels of agreement for the string valued variables (first name, last name, and street name) based on the Jaro-Winkler distance with 0.85 and 0.94 as the thresholds. We also use three levels of agreement for age based on the absolute distance between values, with 1 and 2.5 years as the thresholds for separate agreements, partial agreements, and disagreements, respectively. For the remaining variables (i.e., house number and postal code), we utilize a binary comparison based on exact matching, indicating whether they have an identical value." Thus, the fastLink call is the following:

```
fastLink(dfA = dfA, dfB = dfB,
   varnames = c("first_name", "last_name", "age",
                "house_number", "street_name", "zip"),
   stringdist.match = c("first_name", "last_name", "street_name"),
   numeric.match = c("age"),
   partial.match = c("first_name", "last_name", "street_name", "age"),
   cut.a = 0.94,
   cut.p = 0.85,
   cut.a.num = 1,
   cut.p.num = 2.5,
   return.all = TRUE,
   dedupe.matches = FALSE)
```

- Out-of-state person-matching: following the 2016 vote validation, we took the ANES records with observed names that failed to match in the within-state round and performed an out-of-state match on first name, middle name, last name, and age. We restricted the potential match pairs to those cases where first name, last name, and age were observed in both the ANES and vendor files to guard against false matches (middle name was allowed to be missing). As before, the out-of-state fastLink call is identical to the 2016 fastLink call:

```
fastLink(dfA = dfA, dfB = dfB,
        varnames = c("first_name", "middle_name", "last_name", "age"),
```

```
        stringdist.match = c("first_name", "last_name"),
        numeric.match = c("age"),
        cut.a = 0.94,
        cut.a.num = 1,
        return.all = TRUE,
        dedupe.matches= FALSE)
```

- Address-matching --- which was not attempted in 2016 --- was attempted for all records, even those without observed names.
- Address-matching: since gender is unobserved, we only block on state. The fastLink call --- which is identical to the person-matching call but drops name/age data --- is the following:

```
fastLink(dfA = dfA, dfB = dfB,
        varnames = c("house_number", "street_name", "zip"),
        stringdist.match = c("street_name"),
        partial.match = c("street_name"),
        cut.a = 0.94,
        cut.p = 0.85,
        cut.a.num = 1,
        cut.p.num = 2.5,
        return.all = TRUE,
        dedupe.matches = FALSE)
```

- In all cases, the same ANES record was allowed to match to multiple records in the vendor file (hereafter, "match pairs"). However, to guard against false matches, we immediately discarded all match pairs with match probabilities (obtained from the fastLink algorithm) less than 10%.

Manual review procedure for Vendor 2

ANES staff manually reviewed each case for which fastLink assigned a match probability of 10% or greater, by comparing the ANES record to the best matching record from Vendor 2. Two trained coders independently reviewed all 9,262 matched pairs of records. The two primary reviewers were trained undergraduate research assistants. Both primary reviewers practiced coding on several hundred cases and achieved inter-coder reliability exceeding 90% and exceeding Cohen's kappa = .90 before working on the primary coding.

Both coders independently made an objective count of the number of "minor" and "major" differences between the ANES sample and Vendor 2 records. Minor differences were specifically defined as differences such as an apparent typographical error, the use of a recognizable and appropriate nickname, omitting a name suffix such as Jr. or Sr. when it was given in the other record, omitting a unit or apartment number when it was given in the other record, or a difference in one field (month, day, or year) of the date of birth when the other two fields matched. Major differences were specifically defined as differences such as a different name, different address, or difference in two fields of the date of birth, or difference of more than 15 years in the date of birth. After scrutinizing such differences to tally them, coders then used their own subjective judgment to rate each match as a "correct match," "probably correct," "probably incorrect," or "clearly incorrect." These four categories were then collapsed into a dichotomous variable indicating the match was judged as likely correct or incorrect. Inter-coder

reliability for the dichotomous codes assigned to the 9,262 pairs of records was 98.3% agreement overall and Cohen's kappa = .953.

For each case where the two coders reached different subjective conclusions about the matching outcome, a third reviewer examined the cases and cast a tie-breaking vote. The third reviewer was a senior member of ANES staff. This procedure generated the adjudicated match decision variable: *val2_match*. This can be used to distinguish between correct and incorrect matches.

Codebook

```
-------------------------------------------------------------------------------------------
version                                        ANES 2020 Voter Validation dataset version
-------------------------------------------------------------------------------------------

                  Type: String (str53), but longest is str51

         Unique values: 1                          Missing "": 0/18,430

            Tabulation: Freq.  Value
                        18,430  "ANES 2020 Voter Validation dataset
                                version 20220902"

               Warning: Variable has embedded blanks.

-------------------------------------------------------------------------------------------
V200001                                                                  ANES 2020 Case ID
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)

                 Range: [200015,535551]               Units: 1
         Unique values: 18,430                    Missing .: 0/18,430

                  Mean:  370949
             Std. dev.: 98839.9

           Percentiles:     10%       25%       50%       75%       90%
                         218650    311083    361759    445986    511392

-------------------------------------------------------------------------------------------
val1_addmatch                                              Address match found (vendor 1)
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)
                 Label: val1_addmatch

                 Range: [0,1]                         Units: 1
         Unique values: 2                         Missing .: 0/18,430

            Tabulation: Freq.   Numeric  Label
                         1,628         0  No
                        16,802         1  Yes

-------------------------------------------------------------------------------------------
val1_number                                          Number of linking persons (vendor 1)
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)
```

```
                Range: [0,22316]                  Units: 1
        Unique values: 410                        Missing .: 0/18,430

                 Mean: 19.7233
            Std. dev.: 283.041

          Percentiles:     10%       25%       50%       75%       90%
                             1         1         2         3        19
```

---

val1_matchprob — Estimated person match probability (vendor 1)

```
                 Type: Numeric (double)

                Range: [1.936e-10,1]               Units: 1.000e-17
        Unique values: 706                        Missing .: 8,865/18,430

                 Mean: .771601
            Std. dev.: .388811

          Percentiles:     10%       25%       50%       75%       90%
                        .047446   .863424   .999943   .999994         1
```

---

val1_turnout16 — Voting status, 2016 general election (vendor 1)

```
                 Type: Numeric (double)
                Label: val1_turnout16

                Range: [1,6]                      Units: 1
        Unique values: 6                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       8,478        1  Linkage not attempted due to no
                                       PII
                         387        2  Linkage attempted, no record
                                       found
                       3,304        3  Linked, not voting
                       1,073        4  Linked, voted absentee
                         911        5  Linked, voted early
                       4,277        6  Linked, voted in person on
                                       election day
```

---

val1_turnout20 — Voting status, 2020 general election (vendor 1)

```
                 Type: Numeric (double)
                Label: val1_turnout20

                Range: [1,6]                      Units: 1
        Unique values: 6                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       8,478        1  Linkage not attempted due to no
                                       PII
                         387        2  Linkage attempted, no record
                                       found
```

```
                        2,220         3  Linked, not voting
                        3,320         4  Linked, voted absentee
                        1,104         5  Linked, voted early
                        2,921         6  Linked, voted in person on
                                            election day

-------------------------------------------------------------------------------------------
val1_hh_turnout16                               Household turnout status 2016 (vendor 1)
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)
                 Label: val1_hh_turnout16

                 Range: [-1,5]                        Units: 1
         Unique values: 7                          Missing .: 0/18,430

            Tabulation: Freq.   Numeric  Label
                        1,628        -1  No address match found by
                                            vendor 1
                        4,027         0  Linked address, no voters found
                        6,913         1  Linked address, 1 voter match
                        3,396         2  Linked address, 2 voter matches
                        1,249         3  Linked address, 3 voter matches
                          751         4  Linked address, 4 voter matches
                          466         5  Linked address, 5 or more voter
                                            matches

-------------------------------------------------------------------------------------------
val1_hh_turnout20                               Household turnout status 2020 (vendor 1)
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)
                 Label: val1_hh_turnout20

                 Range: [-1,5]                        Units: 1
         Unique values: 7                          Missing .: 0/18,430

            Tabulation: Freq.   Numeric  Label
                        1,628        -1  No address match found by
                                            vendor 1
                        2,808         0  Linked address, no voters found
                        7,146         1  Linked address, 1 voter match
                        3,847         2  Linked address, 2 voter matches
                        1,475         3  Linked address, 3 voter matches
                        1,004         4  Linked address, 4 voter matches
                          522         5  Linked address, 5 or more voter
                                            matches

-------------------------------------------------------------------------------------------
val2_addmatch                                         Address match found (vendor 2)
-------------------------------------------------------------------------------------------

                  Type: Numeric (double)
                 Label: val2_addmatch

                 Range: [0,1]                         Units: 1
         Unique values: 2                          Missing .: 0/18,430

            Tabulation: Freq.   Numeric  Label
                          320         0  No
                       18,110         1  Yes
```

```
-------------------------------------------------------------------------------
val2_number                                        Number of linking persons (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_number, but 395 nonmissing values are not labeled

                Range: [-1,1278]                     Units: 1
        Unique values: 396                        Missing .: 0/18,430

             Examples: 1
                       2
                       3
                       5

-------------------------------------------------------------------------------
val2_matchprob                          Estimated person match probability (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)

                Range: [.10013547,1]                 Units: 1.000e-10
        Unique values: 1,360                      Missing .: 9,168/18,430

                 Mean: .914625
            Std. dev.: .197573

          Percentiles:     10%       25%       50%       75%       90%
                        .606676   .978939         1         1         1

-------------------------------------------------------------------------------
val2_match                        Final (adjudicated) record match judgment (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_match

                Range: [-1,1]                        Units: 1
        Unique values: 3                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,168        -1  -1 Inapplicable, no person
                                        match found
                       2,276         0  No match or unlikely match
                       6,986         1  Match or likely match

-------------------------------------------------------------------------------
val2_turnout16                          Voting status, 2016 general election (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_turnout16

                Range: [1,7]                         Units: 1
        Unique values: 4                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       8,478         1  Linkage not attempted due to no
                                        PII
                         690         2  Linkage attempted, no record
```

```
                                found
                2,791         3  Linked, not voting
                6,471         7  Linked, voted, timing not
                                 available

-------------------------------------------------------------------------------
val2_turnout20                    Voting status, 2020 general election (vendor 2)
-------------------------------------------------------------------------------

            Type: Numeric (double)
           Label: val2_turnout20

           Range: [1,7]                      Units: 1
   Unique values: 4                        Missing .: 0/18,430

      Tabulation: Freq.   Numeric  Label
                  8,478         1  Linkage not attempted due to no
                                   PII
                    690         2  Linkage attempted, no record
                                   found
                  1,380         3  Linked, not voting
                  7,882         7  Linked, voted, timing not
                                   available

-------------------------------------------------------------------------------
val2_hh_turnout16          Number of 2016 voters linked to sampled address (vendor 2)
-------------------------------------------------------------------------------

            Type: Numeric (double)
           Label: val2_hh_turnout16, but 240 nonmissing values are not labeled

           Range: [-1,904]                   Units: 1
   Unique values: 241                      Missing .: 0/18,430

        Examples: 0
                  1
                  2
                  3

-------------------------------------------------------------------------------
val2_hh_turnout20          Number of 2020 voters linked to sampled address (vendor 2)
-------------------------------------------------------------------------------

            Type: Numeric (double)
           Label: val2_hh_turnout20, but 306 nonmissing values are not labeled

           Range: [-1,997]                   Units: 1
   Unique values: 307                      Missing .: 0/18,430

        Examples: 1
                  2
                  2
                  4

-------------------------------------------------------------------------------
val2_c1_minor                        Coder 1 tally of minor differences (vendor 2)
-------------------------------------------------------------------------------

            Type: Numeric (double)
           Label: val2_c1_minor, but 5 nonmissing values are not labeled
```

```
                Range: [-1,4]                        Units: 1
        Unique values: 6                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,168        -1  -1 No clerical review (usually
                                        inapplicable)
                       2,622         0
                       5,451         1
                       1,120         2
                          66         3
                           3         4

-------------------------------------------------------------------------------------------
val2_c1_major                               Coder 1 tally of major differences (vendor 2)
-------------------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_c1_major, but 10 nonmissing values are not labeled

                Range: [-1,9]                        Units: 1
        Unique values: 11                         Missing .: 0/18,430

             Examples: -1    -1 No clerical review (usually inapplicable)
                       -1    -1 No clerical review (usually inapplicable)
                       0
                       1

-------------------------------------------------------------------------------------------
val2_c1_match                                         Coder 1 match judgment (vendor 2)
-------------------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_c1_match, but 4 nonmissing values are not labeled

                Range: [-1,3]                        Units: 1
        Unique values: 5                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,170        -1  -1 No clerical review (usually
                                        inapplicable)
                       6,071         0
                         875         1
                         252         2
                       2,062         3

-------------------------------------------------------------------------------------------
val2_c2_minor                               Coder 2 tally of minor differences (vendor 2)
-------------------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_c2_minor, but 5 nonmissing values are not labeled

                Range: [-1,4]                        Units: 1
        Unique values: 6                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,168        -1  -1 No clerical review (usually
                                        inapplicable)
                       4,286         0
                       4,099         1
                         838         2
```

```
                          37         3
                           2         4

-------------------------------------------------------------------------------
val2_c2_major                              Coder 2 tally of major differences (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_c2_major, but 6 nonmissing values are not labeled

                Range: [-1,5]                        Units: 1
        Unique values: 7                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,168        -1  -1 No clerical review (usually
                                        inapplicable)
                       5,567         0
                       1,478         1
                         408         2
                         726         3
                           9         4
                       1,074         5

-------------------------------------------------------------------------------
val2_c2_match                                    Coder 2 match judgment (vendor 2)
-------------------------------------------------------------------------------

                 Type: Numeric (double)
                Label: val2_c2_match, but 4 nonmissing values are not labeled

                Range: [-1,3]                        Units: 1
        Unique values: 5                          Missing .: 0/18,430

           Tabulation: Freq.   Numeric  Label
                       9,169        -1  -1 No clerical review (usually
                                        inapplicable)
                       6,401         0
                         671         1
                         301         2
                       1,888         3
```