LaurentiuMandocescu_hw5

Laurentiu Mandocescu

2025-03-15

a. Data gathering and integration

```r
# Load necessary libraries
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2

## Warning: package 'ggplot2' was built under R version 4.4.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
library(ggplot2)
library(skimr)
```

```
## Warning: package 'skimr' was built under R version 4.4.3
```

```r
# Load the dataset
anes_data <- read.csv("anes_timeseries_2024_csv_20250219.csv")

# Check the dimensions of the dataset
dim(anes_data)
```

```
## [1] 3349  708
```

```r
# Look at the first few rows
head(anes_data, 3)
```

```
##                      version V240001 V200001 V160001_orig V240002 V240003
## 1 ANES2024TimeSeries_20250219  200061      NA           NA       3       2
## 2 ANES2024TimeSeries_20250219  200078      NA           NA       2       2
## 3 ANES2024TimeSeries_20250219  200096      NA           NA       2       2
##   V240101a V240101c V240101d  V240102a V240102c V240102d  V240103a V240103c
## 1       NA       NA       NA        NA       NA       NA        NA       NA
## 2       NA       NA       NA 0.5414451        1        1 0.3772640        1
## 3       NA       NA       NA 0.7229589        2        1 0.5342351        2
##   V240103d  V240104a V240104c V240104d  V240105a V240105c V240105d V241001
## 1       NA 3.1652254        1        1 2.2790785        1        1      -1
## 2        1 0.8018586        2        1 0.5399182        2        1       1
## 3        1 0.6025312        3        1 0.4035259        3        1       1
##   V241002 V241003 V241004 V241005 V241006 V241007 V241008x V241009 V241010
## 1      -1      -1       4      -1      -1      -1       -1      -2      -2
## 2      -1      -1       3       1       1       1        1      -2      -2
## 3      -1      -1       2       1       1       1        1      -2      -2
##   V241011 V241012 V241013 V241014 V241015 V241016 V241017 V241018 V241019
## 1      -2      -1       3      -1       1      -3      -1      -2      -2
## 2      -2       1      -1       1      -1      -3      -1      -2      -2
## 3      -2       1      -1       2      -1      -3      -1      -2      -2
##   V241020 V241021 V241022 V241023 V241024 V241025 V241026 V241027 V241028
## 1      -2      -3      -2      -2      -3      -1      -3      -2      -2
## 2      -2      -3      -2      -2      -3       4      -3      -2      -2
## 3      -2      -3      -2      -2      -3       1      -3      -2      -2
##   V241029 V241030 V241031 V241032 V241033 V241034 V241035 V241036 V241037
## 1      -2      -1      -1      -1      -1       2      -1      -1      -1
## 2      -2      -1       3      -1      -1      -1       2      -1      -1
## 3      -2      -1       3      -1      -1      -1       2      -1      -1
##   V241038 V241039 V241039z V241040 V241041 V241042 V241043 V241043z V241044
## 1      -1      -1       -2      -1      -1      -1      -1       -2      -1
## 2      -1      -1       -2      -1      -1       1       1       -2       1
## 3      -1      -1       -2      -1      -1       1       1       -2       1
##   V241045 V241046 V241046z V241047 V241048 V241049 V241050 V241051 V241051z
## 1      -1      -1       -2       2       1       2      -1      -1       -2
## 2      -1      -1       -2      -1      -1       1      -1      -1       -2
## 3      -1      -1       -2      -1      -1       1      -1      -1       -2
##   V241052 V241053 V241053z V241054 V241055 V241055z V241056x V241057 V241058
## 1      -1      -1       -2      -1      -1       -2       -2      -1      -1
## 2       2      -1       -2       2      -1       -2       -2      -1      -1
## 3       1       1       -2      -1      -1       -2       -2      -1      -1
##   V241058z V241059 V241060 V241060z V241061 V241062 V241062z V241063 V241064
## 1       -2      -1      -1       -2      -1      -1       -2      -1      -1
## 2       -2       2      -1       -2       2      -1       -2      -1      -1
## 3       -2       1       1       -2      -1      -1       -2      -1      -1
##   V241064z V241065 V241066 V241066z V241067 V241068 V241068z V241069 V241070
## 1       -2      -1      -1       -2      -1      -1       -2      -1      -1
## 2       -2      -1      -1       -2      -1      -1       -2      -1      -1
## 3       -2      -1      -1       -2      -1      -1       -2      -1      -1
##   V241070z V241071 V241072 V241072z V241073 V241074 V241074z V241075x V241076x
## 1       -2      -1      -1       -2      -1      -1       -2       30       -1
## 2       -2      -1      -1       -2      -1      -1       -2       20       -2
## 3       -2      -1      -1       -2      -1      -1       -2       20       20
##   V241077x V241078x V241100 V241101 V241102x V241103 V241104 V241104z V241105
## 1       -1       -1      -1       2        2      -1      -1       -2       3
```

```
## 2       -2      -1       1       -1       -2       2      -1       -2      -1
## 3       20      -1       1       -1        2       1       1       -2      -1
##   V241106x V241107 V241108 V241108z V241109 V241110 V241111 V241112 V241113
## 1        3      -1      -1       -2      -1      -2      -1      -2      -1
## 2        1       2      -1       -2       1      -2       2      -2       2
## 3        2       1       1       -2       1      -2       2      -2       2
##   V241114 V241115 V241116 V241117 V241118 V241119 V241120 V241121 V241122
## 1      -2      -1      -2      -1      -1      -1      -1      -1      -1
## 2      -2       1      -2       2       4       2       2       3       3
## 3      -2       1      -2       2       4       4       4       3       1
##   V241123 V241124 V241125 V241126 V241127 V241128 V241129x V241130 V241131
## 1      -1      -1      -1      -1      -1      -1       -1      -1      -1
## 2       3       3       2       2       1       2        2       2       1
## 3       4       1       3       4       2       1        4       2       1
##   V241132 V241133x V241134 V241135 V241136 V241137x V241138 V241139 V241140x
## 1       3        3      -1      -1       2        2      -1      -1       -1
## 2      -1        4       1       2      -1        2       1       1        1
## 3      -1        4       1       2      -1        2       1       2        2
##   V241141 V241142 V241143x V241144 V241145 V241146x V241147 V241148 V241149x
## 1      -1      -1       -1      -1      -1       -1      -1      -1       -1
## 2       1       2        2       1       2        2       1       2        2
## 3       1       2        2       1       1        1       1       1        1
##   V241150 V241151 V241152x V241153 V241154 V241155x V241156 V241157 V241158
## 1      -1      -1       -1      -1      -1       -1      60       0      60
## 2       1       2        2       1       2        2      95       0      70
## 3       1       2        2       1       2        2      85       0      40
##   V241159 V241160 V241161 V241162 V241163 V241164 V241165 V241166 V241167
## 1      60      -1      -1      -1      -1       0      40      60      15
## 2       0     100       0      50      50       0      -1      95       0
## 3      40      85      50      50      50       0      -1      85       0
##   V241168 V241169 V241170 V241171 V241172 V241173 V241174 V241175 V241176
## 1       0      -1      -2      -1      -2      -1      -2      -1      -2
## 2      -1       1      -2       2      -2       2      -2       1      -2
## 3      -1       1      -2       2      -2       1      -2       1      -2
##   V241177 V241178 V241179 V241180 V241181 V241182 V241183 V241184 V241185
## 1       2      -1       2       7      -1      -1      -1      -1      -1
## 2       3      -1       2       7       9       9       2       7       1
## 3       2      -1       2       6       2       6       1       7       1
##   V241186 V241187x V241200 V241201 V241202 V241203 V241204 V241205 V241206
## 1      -1       -1      -1      -1      -1      -1      -1      -1      -1
## 2       1        1       1       1       1       1       1       5       5
## 3       1        1       1       2       2       2       2       5       5
##   V241207 V241208 V241209 V241210 V241211 V241211z V241212 V241213 V241213z
## 1      -1      -1      -1      -1      -1       -2      -1      -1       -2
## 2       5       5       5       1       1       -2       1       1       -2
## 3       4       5       3       1       1       -2       1       1       -2
##   V241214 V241215 V241216 V241217 V241218x V241219 V241220 V241221 V241221z
## 1      -1      -1      -1      -1       -1      -1      -1      -1       -2
## 2       2       2      -1       3        5       2       1       3       -2
## 3       2       1       1      -1        1       1       1       1       -2
##   V241222 V241223 V241224 V241225 V241226 V241227x V241228 V241229 V241230
## 1      -1      -1      -1      -1       3        3      -1       5      -1
## 2      -1       3      -1      -1      -1        3       3       2       2
## 3       1      -1      -1      -1      -1        1       2       3       3
```

```
##   V241231 V241232 V241233 V241234 V241235 V241236 V241237 V241238 V241239
## 1      -1      -1      -1       4      -1       4       2      -1       6
## 2       2       2       4       2       1       2       2       1       4
## 3       1       1       3       3       1       3       3       1      99
##   V241240 V241241 V241242 V241243 V241244 V241245 V241246 V241247 V241248
## 1       4       1      -1      -1      -1      -1      -1      -1      -1
## 2       6       1       3       2       7       4       2       7       2
## 3       4       4       5       4       7       3       5       5       1
##   V241249 V241250 V241251 V241252 V241253 V241254 V241255 V241256 V241257
## 1       1      -1      -1      -1      -1      -1       2       4       7
## 2      -1       2       7       4       2       7       2       1       6
## 3      -1       1       7       2       4       2       2       2       5
##   V241258 V241259 V241260 V241261 V241262 V241263x V241264 V241265 V241266x
## 1      -1      -1      -1      -1      -1       -1      -1      -1       -1
## 2       1       1       7       3      -1        3       1       1        1
## 3       1       1       7       3      -1        3       1       1        1
##   V241267 V241268 V241269x V241270 V241271 V241272x V241273 V241274 V241275x
## 1      -1      -1       -1      -1      -1       -1      -1      -1       -1
## 2       3      -1        3       1       1        1       1       2        2
## 3       3      -1        3       1       2        2       1       1        1
##   V241276 V241277 V241278x V241279 V241280 V241281x V241282 V241283 V241284x
## 1      -1      -1       -1      -1      -1       -1      -1      -1       -1
## 2       3      -1        3       1       1        1       1       1        1
## 3       3      -1        3       1       2        2       1       1        1
##   V241285 V241286 V241287x V241288 V241289 V241290x V241291 V241292 V241293
## 1      -1      -1       -1      -1      -1       -1      -1      -1       5
## 2       2       2        6       1       1        1       2      -1      -1
## 3       3      -1        4       1       1        1       1       2      -1
##   V241294x V241295 V241296 V241297x V241298 V241299 V241300x V241301 V241302
## 1        5      -1      -1       -1      -1      -1       -1      -1      -1
## 2        3       1       2        2       2      -1        3       3       4
## 3        2       2      -1        3       2      -1        3       2       4
##   V241302z V241303 V241304 V241305 V241306 V241307 V241308x V241309 V241310
## 1       -2      -1      -1      -1      -1      -1       -1      -1      -1
## 2       -2       4       4       1       2       2        3       2       2
## 3       -2       5       4       2       2       1        4       2       2
##   V241311 V241312x V241313 V241314 V241315 V241316 V241317 V241318 V241319x
## 1      -1       -1      -1       3      -1      -1      -1      -1       -1
## 2       1        4       4       4       5       2       1       2        2
## 3       2        3       3       2       4       5       1       2        2
##   V241320 V241321 V241322x V241323 V241324 V241325 V241326 V241327 V241328
## 1      -1      -1       -1      -1      -1      -1      -1      -1      -1
## 2       1       2        2       3       5       5       5       5       2
## 3       1       2        2       1       4       4       5       5       2
##   V241329 V241330x V241331 V241332 V241333x V241334 V241335 V241336 V241337
## 1      -1       -1      -1      -1       -1      -1      -1      -1      -1
## 2       1        7       2       1        7       5       5       1       3
## 3       3        5       3      -1        4       2       3       2       2
##   V241338x V241339 V241340 V241341 V241342 V241343 V241344x V241345 V241346
## 1       -1      -1      -1      -1      -1      -1       -1      -1      -1
## 2        3       1       1       1       2       1        7       1       1
## 3        5       1       1       1       3      -1        4       1       1
##   V241347x V241348 V241349 V241350x V241351 V241352 V241353x V241360 V241361
## 1       -1      -1      -1       -1      -1      -1       -1      -1      -1
```

```
## 2        1       2       2        5       1       1        1       1       1
## 3        1       2       1        6       1       1        1       1       1
##   V241362 V241363x V241364 V241365 V241366x V241367 V241368 V241369x V241370
## 1      -1       -1      -1      -1       -1      -1      -1       -1      -1
## 2      -1        1       1       1        1       1       1        1       1
## 3      -1        1       1       1        1       1       1        1       1
##   V241371 V241372x V241373 V241374 V241375x V241376 V241377 V241378x V241379
## 1      -1       -1      -1      -1       -1      -1      -1       -1      -1
## 2       2        2       3      -1        4       1       1        1       1
## 3       1        1       2       1        7       1       1        1       1
##   V241380 V241381x V241382 V241383 V241384 V241385x V241386 V241387 V241388
## 1      -1       -1      -1      -1      -1       -1      -1      -1      -1
## 2       1        1       1      -1      -1       -1       3       3      -1
## 3       1        1      -1       1       1        1       2       2       3
##   V241389x V241390 V241391 V241392x V241393 V241394 V241395x V241396 V241397
## 1       -1      -1      -1       -1      -1      -1       -1      -1      -1
## 2        4       2       1        6       3      -1        4       1       5
## 3        5       2       2        5       3      -1        4       2       1
##   V241398 V241399 V241400x V241401 V241402 V241403x V241404 V241405 V241406x
## 1      -1      -1       -1      -1      -1       -1      -1      -1       -1
## 2       1       1        1       1       1        1       1       1        1
## 3       1       3        3       1       1        1       1       1        1
##   V241407 V241408 V241409x V241410 V241411 V241412x V241420 V241421 V241421z
## 1      -1      -1       -1      -1      -1       -1       5      -1       -2
## 2       2       2        6       2       2        6       3       3       -2
## 3       1       3        3       2       3        7       3       5       -2
##   V241422 V241423 V241424 V241424z V241425 V241425z V241426 V241426z V241427
## 1      -1      -1      -3       -3      -3       -3      -3       -3      -3
## 2       5      -1      -3       -3      -3       -3      -3       -3      -3
## 3       5      -1      -3       -3      -3       -3      -3       -3      -3
##   V241427z V241428 V241428z V241429 V241429z V241430 V241430z V241431 V241431z
## 1       -3      -3       -3      -3       -3      -3       -3      -3       -3
## 2       -3      -3       -3      -3       -3      -3       -3      -3       -3
## 3       -3      -3       -3      -3       -3      -3       -3      -3       -3
##   V241432 V241432z V241433 V241434 V241434z V241435 V241436 V241437 V241438
## 1      -3       -3      -3      -3       -3      -3      -3      -3      -3
## 2      -3       -3      -3      -3       -3      -3      -3      -3      -3
## 3      -3       -3      -3      -3       -3      -3      -3      -3      -3
##   V241438z V241439 V241440 V241441 V241442 V241443a V241443b V241443c V241443d
## 1       -3      -1      -1      -1      -1       -1       -1       -1       -1
## 2       -3       2      -1      -1      -1        0        0        0        0
## 3       -3       1       4      -1      -1        0        0        0        0
##   V241443e V241443f V241443g V241443h V241444x V241445x V241450 V241451 V241452
## 1       -1       -1       -1       -1       -2       -2      -1      -1      -1
## 2        0        0        1        0       -2       -2       0       3       3
## 3        0        1        0        0       -2       -2       0       2       2
##   V241453 V241454 V241455 V241456 V241457 V241458x V241459 V241460 V241461x
## 1      -3      -3      -3      -3      -4       -2      -1       5        5
## 2      -3      -3      -3      -3      -1       40       6      -1        5
## 3      -3      -3      -3      -3      -1       29       6      -1        5
##   V241462 V241463 V241463z V241464 V241465x V241466 V241467 V241467z V241468
## 1      -1      -1       -3       4        4      -1      -1       -3      -1
## 2       2      14       -3      -1        5      -1      -1       -3      -1
## 3       2      13       -3      -1        4      -1      -1       -3      -1
```

```
##   V241469x V241470 V241471 V241472 V241473 V241474 V241475 V241476 V241477
## 1       -1      -1      -1      -1      -1      -1      -1      -1      -1
## 2       -1       3       1      -1      -1      -1      -1      -1      -1
## 3       -1       3       1      -1      -1      -1      -1      -1      -1
##   V241478 V241479 V241480 V241481 V241482 V241483 V241484 V241485 V241486
## 1      -1      -1      -1      -1      -1      -1      -3      -3      -3
## 2      -1       1      -1      40       3       1      -3      -3      -3
## 3      -1       1      -1      60       3       1      -3      -3      -3
##   V241487x V241488x V241489a V241489b V241489c V241489d V241489e V241490
## 1       -2       -2        0        0        0        0        1      -1
## 2       -2       -2        0        0        0        0        1      -1
## 3       -2       -2        0        0        0        0        1      -1
##   V241491 V241492 V241493 V241494 V241495a V241495b V241495c V241495d V241495e
## 1      -1      -1      -1      -1        0        0        0        0        0
## 2       2       1       2       2       -1       -1       -1       -1       -1
## 3       2       2       2       2       -1       -1       -1       -1       -1
##   V241495g V241495f V241496 V241497 V241498a V241498b V241498c V241499 V241500a
## 1        0        0      -3      -1        0        0        0       2       -3
## 2       -1       -1      -3       2       -1       -1       -1       2       -3
## 3       -1       -1      -3       2       -1       -1       -1       2       -3
##   V241500b V241500c V241500d V241500e V241501x V241502 V241503 V241504 V241505
## 1       -3       -3       -3       -3        1      -1      -3      -1      -3
## 2       -3       -3       -3       -3        1       2      -3      -1      -3
## 3       -3       -3       -3       -3        1       2      -3      -1      -3
##   V241506 V241507 V241507z V241508 V241509 V241510 V241510z V241511 V241512x
## 1      -1      -1       -3       1      -1      -3       -3      -3       -2
## 2       1       1       -3      -1       2      -3       -3      -3       -2
## 3       2       4       -3      -1       4      -3       -3      -3       -2
##   V241513 V241514 V241515 V241516 V241517 V241518a V241518b V241518c V241518d
## 1      -1      -3      -3      -1      -1       -3       -3       -3       -3
## 2      -1      -3      -3      -1      -1       -3       -3       -3       -3
## 3      -1      -3      -3      -1      -1       -3       -3       -3       -3
##   V241518e V241519x V241520 V241521 V241522a V241522b V241522c V241522d
## 1       -3       -1      -1      -1        0        0        0        0
## 2       -3       -1      -1       0       -1       -1       -1       -1
## 3       -3       -1      -1       0       -1       -1       -1       -1
##   V241522e V241523 V241524 V241525 V241526 V241527 V241528 V241529 V241529z
## 1        0      -1      -1      -1      -1      -1      -1      -1       -3
## 2       -1       1      -1       2       1      -1       1      36       -3
## 3       -1       1      -1       2       1      -1       1      25       -3
##   V241530 V241531 V241532 V241533 V241534 V241535 V241536 V241537 V241538
## 1       2      -1      -1      -3      -1      -3      -1      -1      -1
## 2       2      -1       2      -3       1      -3       1       1      -1
## 3       2      -1       2      -3       1      -3       1       1      -1
##   V241539 V241540 V241541 V241550 V241551 V241551z V241552 V241553 V241553z
## 1      -1      -1      -1       2      -1       -2      -1      -1       -2
## 2       4       2       2       2       2       -2       2       1       -2
## 3       5       1       2       2       2       -2       2       1       -2
##   V241554 V241555 V241556 V241557 V241558 V241559 V241560 V241561 V241562
## 1      -1      -3      -3      -3      -3      -3      -3      -3      -3
## 2       1      -3      -3      -3      -3      -3      -3      -3      -3
## 3       1      -3      -3      -3      -3      -3      -3      -3      -3
##   V241563 V241564 V241565 V241566x V241567x V241568a V241568b V241568c V241568d
## 1      -3      -3      -3       -1        5        0        0        0        0
```

```
## 2      -3      -3      -3      24       5       0       0       0       1
## 3      -3      -3      -3       1       1       0       0       0       0
##   V241568e V241568f V241569 V241570 V241571 V241572 V241573 V241574a V241574b
## 1        0        1      -1      -1      -1      -1      -1       -1       -1
## 2        0        0       0       3       1       1       1        0        0
## 3        0        1       0       5       1       1       1        0        0
##   V241574c V241574d V241574e V241574f V241574g V241574h V241574i V241574j
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        0        0        0        0        0        0        0        1
## 3        0        0        0        0        0        0        0        1
##   V241575 V241576 V241577 V241578 V241579 V241580 V241581 V241582x V241583
## 1      -1      -1      -1      -1      -1      -1      -1       -1      -1
## 2       2       2       2       2       1       2       2        6       0
## 3       1       1       2       2       1       2       3        5       0
##   V241600a V241600b V241600c V241600d V241600e V241601a V241601b V241601c
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        1        1        1        0        0        0        0        0
## 3        1        0        0        1        0        0        0        0
##   V241601d V241601e V241601f V241601g V241601h V241601i V241601j V241601k
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        0        0        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0        0
##   V241601m V241601n V241601p V241601q V241601r V241601s V241602a V241602b
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        0        0        0        0        0        1        0        0
## 3        0        0        0        0        0        0        0        0
##   V241602c V241602d V241602e V241602f V241602g V241602h V241602i V241602j
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        0        0        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0        0
##   V241602k V241602m V241602n V241602p V241602q V241602r V241603a V241603b
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2        0        0        0        0        0        0       -1       -1
## 3        0        0        0        0        0        0       -1       -1
##   V241603c V241603d V241603e V241603f V241603g V241603h V241603i V241603j
## 1       -1       -1       -1       -1       -1       -1       -1       -1
## 2       -1       -1       -1       -1       -1       -1       -1       -1
## 3       -1       -1       -1       -1       -1       -1       -1       -1
##   V241603k V241603m V241603n V241603p V241603q V241603r V241604 V241605 V241606
## 1       -1       -1       -1       -1       -1       -1      -2      -2      -2
## 2       -1       -1       -1       -1       -1       -1      -2      -2      -2
## 3       -1       -1       -1       -1       -1       -1      -2      -2      -2
##   V241607 V241608 V241609 V241610 V241611 V241612 V241613 V241614 V241615
## 1      -1      -1      -1      -1      -1      -1      -1      -1      -1
## 2       5       5       1    1950      -1       4       1       2       2
## 3       5       5       1    1990      -1       6       2       2       1
##   V241616 V241617 V241618 V241619 V241620 V241621 V241622 V243001    V243050
## 1      -1      -1      -1      -1      -1      -1      -2                  .
## 2      -1      -1       2       5       5       3      -2      NY   20240908
## 3      -1      -1       4       5       5       1      -2      NY   20240826
##      V243051    V243052
## 1          .   20241103
## 2   20240908          .
## 3   20240826          .
```

```r
# Get a summary of the dataset
summary_stats <- skim(anes_data)
summary_stats
```

Table 1: Data summary

Name	anes_data
Number of rows	3349
Number of columns	708
Column type frequency:	
character	5
logical	2
numeric	701
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
version	0	1	27	27	0	1	0
V243001	0	1	1	2	0	52	244
V243050	0	1	1	10	0	97	1
V243051	0	1	1	10	0	97	1
V243052	0	1	8	10	0	21	0

Variable type: logical

skim_variable	n_missing	complete_rate	mean	count
V200001	3349	0	NaN	:
V160001_orig	3349	0	NaN	:

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V240001	0	1.00	261879.91	64605.06	200061.00	213893.00	229162.00	327462.00	399909	
V240002	0	1.00	1.83	0.65	1.00	1.00	2.00	2.00	4	
V240003	0	1.00	2.31	0.46	2.00	2.00	2.00	3.00	3	
V240101a	2307	0.31	1.00	0.98	0.04	0.36	0.66	1.27	5	
V240101c	2307	0.31	1.47	0.50	1.00	1.00	1.00	2.00	2	
V240101d	2307	0.31	16.73	8.51	1.00	9.00	18.00	24.00	32	
V240102a	1286	0.62	1.00	0.88	0.06	0.45	0.73	1.23	5	
V240102c	1286	0.62	50.07	79.43	1.00	10.00	24.00	38.00	372	
V240102d	1286	0.62	26.47	16.62	1.00	12.00	29.00	40.50	52	
V240103a	244	0.93	1.00	1.02	0.04	0.37	0.63	1.20	5	
V240103c	244	0.93	33.76	68.69	1.00	2.00	10.00	31.00	372	
V240103d	244	0.93	40.65	24.62	1.00	23.00	41.00	62.00	84	
V240104a	1042	0.69	1.00	0.85	0.07	0.47	0.74	1.20	5	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V240104c	1042	0.69	57.50	91.66	1.00	12.00	26.00	43.00	425	
V240104d	1042	0.69	26.49	16.71	1.00	12.00	29.00	41.00	52	
V240105a	0	1.00	1.00	1.05	0.04	0.36	0.61	1.18	5	
V240105c	0	1.00	40.07	80.38	1.00	2.00	13.00	35.00	425	
V240105d	0	1.00	39.63	24.44	1.00	22.00	40.00	60.00	84	
V241001	0	1.00	0.86	0.56	-9.00	1.00	1.00	1.00	2	
V241002	0	1.00	-0.37	0.95	-1.00	-1.00	-1.00	1.00	2	
V241003	0	1.00	-0.22	1.22	-8.00	-1.00	-1.00	1.00	2	
V241004	0	1.00	2.54	1.15	-9.00	2.00	2.00	4.00	5	
V241005	0	1.00	1.50	1.00	-9.00	1.00	2.00	2.00	3	
V241006	0	1.00	1.09	1.52	-9.00	1.00	1.00	2.00	2	
V241007	0	1.00	1.09	0.85	-9.00	1.00	1.00	2.00	2	
V241008x	0	1.00	2.10	1.61	-2.00	1.00	2.00	4.00	4	
V241009	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241010	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241011	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241012	0	1.00	0.92	0.71	-9.00	1.00	1.00	1.00	2	
V241013	0	1.00	-0.65	1.34	-9.00	-1.00	-1.00	-1.00	6	
V241014	0	1.00	0.90	0.73	-9.00	1.00	1.00	1.00	2	
V241015	0	1.00	-0.87	0.59	-9.00	-1.00	-1.00	-1.00	2	
V241016	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241017	0	1.00	-0.93	1.65	-1.00	-1.00	-1.00	-1.00	51	
V241018	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241019	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241020	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241021	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241022	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241023	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241024	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241025	0	1.00	0.33	1.81	-9.00	-1.00	-1.00	1.00	5	
V241026	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241027	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241028	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241029	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241030	0	1.00	-0.78	0.75	-8.00	-1.00	-1.00	-1.00	2	
V241031	0	1.00	2.07	1.43	-9.00	1.00	3.00	3.00	3	
V241032	0	1.00	-0.46	1.63	-9.00	-1.00	-1.00	-1.00	9	
V241033	0	1.00	-0.58	1.21	-9.00	-1.00	-1.00	-1.00	9	
V241034	0	1.00	-0.77	0.96	-9.00	-1.00	-1.00	-1.00	6	
V241035	0	1.00	1.72	0.82	-8.00	2.00	2.00	2.00	2	
V241036	0	1.00	-0.88	0.54	-9.00	-1.00	-1.00	-1.00	2	
V241037	0	1.00	-0.89	0.54	-1.00	-1.00	-1.00	-1.00	3	
V241038	0	1.00	-0.92	0.40	-1.00	-1.00	-1.00	-1.00	2	
V241039	0	1.00	-0.90	0.58	-9.00	-1.00	-1.00	-1.00	6	
V241039z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241040	0	1.00	-0.91	0.43	-1.00	-1.00	-1.00	-1.00	2	
V241041	0	1.00	-0.79	1.24	-8.00	-1.00	-1.00	-1.00	7	
V241042	0	1.00	0.82	0.96	-9.00	1.00	1.00	1.00	2	
V241043	0	1.00	0.88	1.97	-9.00	1.00	1.00	2.00	6	
V241043z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241044	0	1.00	0.65	1.04	-9.00	1.00	1.00	1.00	2	
V241045	0	1.00	-0.75	0.86	-9.00	-1.00	-1.00	-1.00	2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241046	0	1.00	-0.90	0.56	-8.00	-1.00	-1.00	-1.00	6	
V241046z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241047	0	1.00	-0.79	0.96	-9.00	-1.00	-1.00	-1.00	7	
V241048	0	1.00	-0.77	0.78	-9.00	-1.00	-1.00	-1.00	2	
V241049	0	1.00	1.48	1.33	-9.00	1.00	2.00	2.00	3	
V241050	0	1.00	-0.93	0.51	-8.00	-1.00	-1.00	-1.00	2	
V241051	0	1.00	-0.93	0.51	-9.00	-1.00	-1.00	-1.00	5	
V241051z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241052	0	1.00	0.86	1.31	-9.00	1.00	1.00	1.00	2	
V241053	0	1.00	0.59	1.63	-9.00	-1.00	1.00	2.00	5	
V241053z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241054	0	1.00	-0.48	1.45	-9.00	-1.00	-1.00	-1.00	2	
V241055	0	1.00	-0.97	0.33	-9.00	-1.00	-1.00	-1.00	5	
V241055z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241056x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241057	0	1.00	-0.94	0.41	-8.00	-1.00	-1.00	-1.00	2	
V241058	0	1.00	-0.93	0.49	-9.00	-1.00	-1.00	-1.00	5	
V241058z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241059	0	1.00	0.38	1.25	-9.00	-1.00	1.00	1.00	2	
V241060	0	1.00	0.13	1.93	-9.00	-1.00	-1.00	1.00	5	
V241060z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241061	0	1.00	-0.61	1.13	-9.00	-1.00	-1.00	-1.00	2	
V241062	0	1.00	-0.98	0.30	-9.00	-1.00	-1.00	-1.00	5	
V241062z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241063	0	1.00	-1.00	0.00	-1.00	-1.00	-1.00	-1.00	-1	
V241064	0	1.00	-1.00	0.00	-1.00	-1.00	-1.00	-1.00	-1	
V241064z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241065	0	1.00	-0.99	0.15	-1.00	-1.00	-1.00	-1.00	2	
V241066	0	1.00	-0.99	0.14	-1.00	-1.00	-1.00	-1.00	2	
V241066z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241067	0	1.00	-1.00	0.09	-1.00	-1.00	-1.00	-1.00	2	
V241068	0	1.00	-1.00	0.00	-1.00	-1.00	-1.00	-1.00	-1	
V241068z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241069	0	1.00	-0.98	0.19	-1.00	-1.00	-1.00	-1.00	2	
V241070	0	1.00	-0.98	0.21	-1.00	-1.00	-1.00	-1.00	5	
V241070z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241071	0	1.00	-0.71	0.76	-8.00	-1.00	-1.00	-1.00	2	
V241072	0	1.00	-0.75	0.95	-9.00	-1.00	-1.00	-1.00	5	
V241072z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241073	0	1.00	-0.93	0.45	-1.00	-1.00	-1.00	-1.00	2	
V241074	0	1.00	-0.99	0.12	-1.00	-1.00	-1.00	-1.00	2	
V241074z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241075x	0	1.00	19.05	7.80	-2.00	20.00	20.00	21.00	32	
V241076x	0	1.00	13.68	10.26	-2.00	-1.00	20.00	21.00	32	
V241077x	0	1.00	9.32	11.16	-2.00	-2.00	10.00	20.00	32	
V241078x	0	1.00	1.42	6.80	-1.00	-1.00	-1.00	-1.00	31	
V241100	0	1.00	1.47	1.50	-1.00	1.00	1.00	2.00	5	
V241101	0	1.00	-0.77	0.90	-4.00	-1.00	-1.00	-1.00	6	
V241102x	0	1.00	2.67	1.33	-4.00	2.00	2.00	3.00	6	
V241103	0	1.00	1.04	1.01	-9.00	1.00	1.00	1.00	2	
V241104	0	1.00	0.62	1.47	-9.00	-1.00	1.00	2.00	5	
V241104z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241105	0	1.00	-0.77	0.87	-9.00	-1.00	-1.00	-1.00	4	
V241106x	0	1.00	2.02	0.90	-4.00	1.00	2.00	3.00	4	
V241107	0	1.00	0.97	1.48	-9.00	1.00	1.00	2.00	2	
V241108	0	1.00	0.55	1.63	-9.00	-1.00	1.00	2.00	5	
V241108z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241109	0	1.00	1.28	1.04	-9.00	1.00	1.00	2.00	2	
V241110	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241111	0	1.00	1.32	1.06	-9.00	1.00	2.00	2.00	2	
V241112	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241113	0	1.00	1.32	1.05	-9.00	1.00	2.00	2.00	2	
V241114	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241115	0	1.00	1.17	0.97	-9.00	1.00	1.00	2.00	2	
V241116	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241117	0	1.00	1.44	1.41	-9.00	1.00	2.00	2.00	2	
V241118	0	1.00	2.21	1.50	-9.00	1.00	2.00	3.00	5	
V241119	0	1.00	2.97	1.68	-9.00	2.00	3.00	4.00	5	
V241120	0	1.00	2.82	1.80	-9.00	2.00	3.00	4.00	5	
V241121	0	1.00	2.84	1.75	-9.00	2.00	3.00	4.00	5	
V241122	0	1.00	1.76	1.37	-9.00	1.00	2.00	3.00	5	
V241123	0	1.00	3.19	1.76	-9.00	3.00	3.00	4.00	5	
V241124	0	1.00	1.78	1.48	-9.00	1.00	2.00	3.00	5	
V241125	0	1.00	3.16	1.78	-9.00	2.00	3.00	4.00	5	
V241126	0	1.00	3.09	1.72	-9.00	2.00	3.00	4.00	5	
V241127	0	1.00	1.31	1.89	-9.00	1.00	2.00	2.00	2	
V241128	0	1.00	1.10	1.02	-9.00	1.00	1.00	2.00	2	
V241129x	0	1.00	2.85	1.67	-2.00	2.00	4.00	4.00	4	
V241130	0	1.00	1.18	1.84	-9.00	1.00	2.00	2.00	2	
V241131	0	1.00	1.08	0.92	-9.00	1.00	1.00	2.00	2	
V241132	0	1.00	-0.76	1.07	-9.00	-1.00	-1.00	-1.00	4	
V241133x	0	1.00	2.78	1.44	-2.00	2.00	3.00	4.00	4	
V241134	0	1.00	1.31	1.29	-9.00	1.00	2.00	2.00	2	
V241135	0	1.00	1.09	0.86	-9.00	1.00	1.00	2.00	2	
V241136	0	1.00	-0.72	1.09	-9.00	-1.00	-1.00	-1.00	4	
V241137x	0	1.00	2.80	1.34	-2.00	2.00	3.00	4.00	4	
V241138	0	1.00	1.13	1.61	-9.00	1.00	1.00	2.00	2	
V241139	0	1.00	1.10	0.91	-9.00	1.00	1.00	2.00	2	
V241140x	0	1.00	2.28	1.65	-2.00	1.00	2.00	4.00	4	
V241141	0	1.00	1.31	1.34	-9.00	1.00	2.00	2.00	2	
V241142	0	1.00	1.08	0.83	-9.00	1.00	1.00	2.00	2	
V241143x	0	1.00	2.54	1.64	-2.00	1.00	3.00	4.00	4	
V241144	0	1.00	1.27	1.44	-9.00	1.00	2.00	2.00	2	
V241145	0	1.00	1.11	0.79	-1.00	1.00	1.00	2.00	2	
V241146x	0	1.00	2.47	1.64	-2.00	1.00	3.00	4.00	4	
V241147	0	1.00	0.98	1.97	-9.00	1.00	1.00	2.00	2	
V241148	0	1.00	1.05	0.87	-9.00	1.00	1.00	2.00	2	
V241149x	0	1.00	2.12	1.72	-2.00	1.00	2.00	4.00	4	
V241150	0	1.00	1.26	1.63	-9.00	1.00	2.00	2.00	2	
V241151	0	1.00	1.10	0.82	-8.00	1.00	1.00	2.00	2	
V241152x	0	1.00	2.61	1.66	-2.00	2.00	3.00	4.00	4	
V241153	0	1.00	1.08	1.88	-9.00	1.00	2.00	2.00	2	
V241154	0	1.00	1.11	0.89	-9.00	1.00	1.00	2.00	2	
V241155x	0	1.00	2.35	1.69	-2.00	1.00	3.00	4.00	4	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241156	0	1.00	46.93	37.02	-9.00	0.00	50.00	85.00	100	
V241157	0	1.00	40.08	38.61	-9.00	0.00	40.00	75.00	100	
V241158	0	1.00	41.80	33.82	-9.00	2.00	40.00	70.00	100	
V241159	0	1.00	39.03	29.35	-9.00	12.00	50.00	60.00	100	
V241160	0	1.00	31.43	32.62	-9.00	-1.00	50.00	50.00	100	
V241161	0	1.00	29.69	34.27	-9.00	-1.00	30.00	50.00	100	
V241162	0	1.00	32.94	34.27	-9.00	-1.00	50.00	50.00	100	
V241163	0	1.00	21.99	30.44	-9.00	-9.00	0.00	50.00	100	
V241164	0	1.00	35.22	35.67	-9.00	0.00	30.00	60.00	100	
V241165	0	1.00	1.74	13.60	-1.00	-1.00	-1.00	-1.00	100	
V241166	0	1.00	45.35	32.91	-9.00	15.00	50.00	70.00	100	
V241167	0	1.00	43.33	32.40	-9.00	15.00	45.00	70.00	100	
V241168	0	1.00	1.88	14.42	-1.00	-1.00	-1.00	-1.00	100	
V241169	0	1.00	1.22	1.13	-9.00	1.00	1.00	2.00	2	
V241170	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241171	0	1.00	1.15	1.18	-9.00	1.00	1.00	2.00	2	
V241172	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241173	0	1.00	1.25	1.17	-9.00	1.00	1.00	2.00	2	
V241174	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241175	0	1.00	1.12	1.21	-9.00	1.00	1.00	2.00	2	
V241176	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241177	0	1.00	19.39	35.04	-9.00	3.00	4.00	6.00	99	
V241178	0	1.00	-0.35	1.82	-9.00	-1.00	-1.00	1.00	2	
V241179	0	1.00	3.31	11.30	-9.00	1.00	2.00	3.00	99	
V241180	0	1.00	6.38	11.10	-9.00	5.00	6.00	7.00	99	
V241181	0	1.00	1.96	4.22	-9.00	1.00	2.00	4.00	9	
V241182	0	1.00	3.86	4.79	-9.00	4.00	6.00	6.00	9	
V241183	0	1.00	1.98	2.37	-9.00	1.00	2.00	3.00	7	
V241184	0	1.00	4.74	3.20	-9.00	4.00	6.00	7.00	7	
V241185	0	1.00	0.88	1.11	-9.00	1.00	1.00	1.00	2	
V241186	0	1.00	0.99	0.80	-9.00	1.00	1.00	1.00	2	
V241187x	0	1.00	1.22	1.09	-2.00	1.00	1.00	1.00	4	
V241200	0	1.00	2.92	2.16	-9.00	2.00	3.00	5.00	5	
V241201	0	1.00	2.91	2.04	-9.00	2.00	3.00	5.00	5	
V241202	0	1.00	2.72	2.07	-9.00	1.00	3.00	5.00	5	
V241203	0	1.00	2.89	2.14	-9.00	2.00	3.00	5.00	5	
V241204	0	1.00	2.37	1.86	-9.00	1.00	2.00	4.00	5	
V241205	0	1.00	2.87	2.03	-9.00	1.00	3.00	5.00	5	
V241206	0	1.00	3.28	2.06	-9.00	2.00	4.00	5.00	5	
V241207	0	1.00	3.07	2.00	-9.00	2.00	3.00	5.00	5	
V241208	0	1.00	3.48	1.99	-9.00	2.00	4.00	5.00	5	
V241209	0	1.00	2.70	1.87	-9.00	2.00	3.00	4.00	5	
V241210	0	1.00	1.52	1.37	-9.00	1.00	1.00	2.00	5	
V241211	0	1.00	1.07	1.89	-9.00	1.00	1.00	2.00	5	
V241211z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241212	0	1.00	1.05	1.00	-9.00	1.00	1.00	1.00	2	
V241213	0	1.00	1.15	1.57	-9.00	1.00	1.00	2.00	5	
V241213z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241214	0	1.00	1.31	1.17	-9.00	1.00	2.00	2.00	2	
V241215	0	1.00	1.37	1.09	-9.00	1.00	1.00	2.00	3	
V241216	0	1.00	0.11	1.20	-1.00	-1.00	-1.00	1.00	3	
V241217	0	1.00	-0.12	1.27	-1.00	-1.00	-1.00	1.00	3	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241218x	0	1.00	3.16	2.64	-2.00	1.00	2.00	6.00	7	
V241219	0	1.00	1.99	1.46	-9.00	1.00	2.00	3.00	4	
V241220	0	1.00	1.52	1.63	-9.00	1.00	2.00	2.00	3	
V241221	0	1.00	1.76	1.52	-9.00	1.00	2.00	3.00	5	
V241221z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241222	0	1.00	0.39	1.21	-1.00	-1.00	1.00	1.00	2	
V241223	0	1.00	-0.03	1.54	-9.00	-1.00	-1.00	1.00	3	
V241224	0	1.00	-0.04	1.50	-9.00	-1.00	-1.00	2.00	2	
V241225	0	1.00	-0.76	1.13	-9.00	-1.00	-1.00	-1.00	3	
V241226	0	1.00	-0.63	1.53	-9.00	-1.00	-1.00	-1.00	7	
V241227x	0	1.00	3.79	2.53	-9.00	2.00	4.00	6.00	7	
V241228	0	1.00	2.64	1.88	-9.00	2.00	3.00	4.00	5	
V241229	0	1.00	3.49	1.28	-9.00	3.00	4.00	4.00	5	
V241230	0	1.00	2.97	1.61	-9.00	2.00	3.00	4.00	5	
V241231	0	1.00	0.90	1.29	-9.00	1.00	1.00	1.00	2	
V241232	0	1.00	1.14	1.02	-9.00	1.00	1.00	2.00	3	
V241233	0	1.00	2.55	1.74	-9.00	2.00	3.00	4.00	5	
V241234	0	1.00	2.83	1.15	-9.00	2.00	3.00	4.00	5	
V241235	0	1.00	1.66	1.18	-9.00	1.00	2.00	2.00	3	
V241236	0	1.00	3.07	1.88	-9.00	2.00	3.00	5.00	5	
V241237	0	1.00	3.18	1.92	-9.00	2.00	3.00	5.00	5	
V241238	0	1.00	2.01	2.07	-9.00	1.00	2.00	3.00	5	
V241239	0	1.00	18.95	34.07	-9.00	4.00	5.00	7.00	99	
V241240	0	1.00	6.55	12.83	-9.00	4.00	6.00	7.00	99	
V241241	0	1.00	4.33	13.04	-9.00	1.00	3.00	4.00	99	
V241242	0	1.00	16.85	32.66	-9.00	3.00	5.00	7.00	99	
V241243	0	1.00	2.73	3.17	-9.00	1.00	3.00	5.00	7	
V241244	0	1.00	4.17	3.36	-9.00	3.00	5.00	7.00	7	
V241245	0	1.00	11.87	27.63	-9.00	1.00	4.00	6.00	99	
V241246	0	1.00	2.04	2.74	-9.00	1.00	2.00	4.00	7	
V241247	0	1.00	4.65	3.31	-9.00	4.00	6.00	7.00	7	
V241248	0	1.00	9.12	23.78	-9.00	1.00	3.00	5.00	99	
V241249	0	1.00	-0.69	1.31	-9.00	-1.00	-1.00	-1.00	7	
V241250	0	1.00	1.67	2.37	-9.00	1.00	1.00	3.00	7	
V241251	0	1.00	4.62	3.12	-9.00	4.00	6.00	7.00	7	
V241252	0	1.00	13.28	29.05	-9.00	2.00	4.00	6.00	99	
V241253	0	1.00	2.11	2.67	-9.00	1.00	2.00	4.00	7	
V241254	0	1.00	4.74	3.28	-9.00	4.00	6.00	7.00	7	
V241255	0	1.00	17.02	33.17	-9.00	2.00	4.00	7.00	99	
V241256	0	1.00	3.76	13.01	-9.00	1.00	2.00	3.00	99	
V241257	0	1.00	6.62	12.75	-9.00	4.00	6.00	7.00	99	
V241258	0	1.00	16.02	33.27	-9.00	1.00	3.00	6.00	99	
V241259	0	1.00	1.86	2.69	-9.00	1.00	2.00	3.00	7	
V241260	0	1.00	4.64	3.41	-9.00	4.00	6.00	7.00	7	
V241261	0	1.00	1.44	1.45	-9.00	1.00	1.00	3.00	3	
V241262	0	1.00	0.44	1.22	-9.00	-1.00	1.00	1.00	2	
V241263x	0	1.00	1.79	1.30	-2.00	1.00	2.00	3.00	5	
V241264	0	1.00	1.28	1.23	-9.00	1.00	1.00	2.00	3	
V241265	0	1.00	0.62	1.13	-9.00	-1.00	1.00	1.00	2	
V241266x	0	1.00	1.64	1.34	-2.00	1.00	1.00	3.00	5	
V241267	0	1.00	1.32	1.26	-9.00	1.00	1.00	2.00	3	
V241268	0	1.00	0.61	1.19	-9.00	-1.00	1.00	1.00	2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241269x	0	1.00	1.72	1.39	-2.00	1.00	1.00	3.00	5	
V241270	0	1.00	1.41	1.31	-9.00	1.00	1.00	3.00	3	
V241271	0	1.00	0.53	1.22	-9.00	-1.00	1.00	1.00	2	
V241272x	0	1.00	1.78	1.32	-2.00	1.00	2.00	3.00	5	
V241273	0	1.00	1.70	1.62	-9.00	1.00	2.00	3.00	3	
V241274	0	1.00	0.36	1.30	-8.00	-1.00	1.00	2.00	2	
V241275x	0	1.00	2.52	1.64	-2.00	1.00	3.00	3.00	5	
V241276	0	1.00	1.47	1.34	-9.00	1.00	1.00	3.00	3	
V241277	0	1.00	0.47	1.27	-9.00	-1.00	1.00	2.00	2	
V241278x	0	1.00	1.83	1.26	-2.00	1.00	2.00	3.00	5	
V241279	0	1.00	1.48	1.56	-9.00	1.00	1.00	3.00	3	
V241280	0	1.00	0.44	1.24	-1.00	-1.00	1.00	1.00	2	
V241281x	0	1.00	1.98	1.44	-2.00	1.00	2.00	3.00	5	
V241282	0	1.00	1.44	1.31	-9.00	1.00	1.00	3.00	3	
V241283	0	1.00	0.54	1.20	-9.00	-1.00	1.00	1.00	2	
V241284x	0	1.00	1.92	1.45	-2.00	1.00	2.00	3.00	5	
V241285	0	1.00	1.84	1.67	-9.00	2.00	2.00	3.00	3	
V241286	0	1.00	0.45	1.42	-8.00	-1.00	1.00	2.00	3	
V241287x	0	1.00	4.19	2.35	-2.00	4.00	4.00	6.00	7	
V241288	0	1.00	1.62	1.77	-9.00	1.00	2.00	3.00	3	
V241289	0	1.00	0.50	1.43	-9.00	-1.00	1.00	2.00	3	
V241290x	0	1.00	3.46	2.38	-2.00	2.00	4.00	5.00	7	
V241291	0	1.00	2.07	1.32	-9.00	1.00	3.00	3.00	3	
V241292	0	1.00	0.70	1.23	-9.00	-1.00	1.00	2.00	2	
V241293	0	1.00	-0.67	1.30	-9.00	-1.00	-1.00	-1.00	5	
V241294x	0	1.00	3.66	1.40	-4.00	3.00	4.00	5.00	5	
V241295	0	1.00	1.51	1.65	-9.00	1.00	2.00	2.00	3	
V241296	0	1.00	0.36	1.37	-9.00	-1.00	1.00	2.00	2	
V241297x	0	1.00	2.52	1.59	-2.00	2.00	3.00	3.00	5	
V241298	0	1.00	1.67	1.53	-9.00	1.00	2.00	3.00	3	
V241299	0	1.00	0.34	1.45	-9.00	-1.00	1.00	2.00	2	
V241300x	0	1.00	2.68	1.61	-2.00	2.00	3.00	4.00	5	
V241301	0	1.00	1.58	1.52	-9.00	1.00	2.00	2.00	3	
V241302	0	1.00	2.82	1.77	-9.00	2.00	3.00	4.00	5	
V241302z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241303	0	1.00	3.37	1.74	-9.00	3.00	4.00	5.00	5	
V241304	0	1.00	2.93	2.43	-9.00	3.00	4.00	4.00	4	
V241305	0	1.00	1.37	2.14	-9.00	1.00	2.00	2.00	4	
V241306	0	1.00	1.00	1.59	-9.00	1.00	1.00	2.00	2	
V241307	0	1.00	1.18	0.91	-9.00	1.00	1.00	2.00	2	
V241308x	0	1.00	1.84	1.48	-2.00	1.00	2.00	3.00	4	
V241309	0	1.00	1.33	1.27	-9.00	1.00	1.00	2.00	3	
V241310	0	1.00	1.37	1.41	-9.00	1.00	2.00	2.00	2	
V241311	0	1.00	1.25	0.88	-9.00	1.00	1.00	2.00	2	
V241312x	0	1.00	2.59	1.54	-2.00	2.00	3.00	4.00	4	
V241313	0	1.00	2.80	1.78	-9.00	3.00	3.00	4.00	5	
V241314	0	1.00	3.24	1.71	-9.00	3.00	4.00	4.00	5	
V241315	0	1.00	3.04	2.00	-9.00	2.00	3.00	4.00	5	
V241316	0	1.00	1.97	2.07	-9.00	2.00	2.00	3.00	5	
V241317	0	1.00	1.24	1.10	-9.00	1.00	1.00	2.00	3	
V241318	0	1.00	0.84	1.14	-1.00	1.00	1.00	1.00	3	
V241319x	0	1.00	2.08	1.95	-2.00	1.00	1.00	4.00	7	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241320	0	1.00	1.37	1.38	-9.00	1.00	1.00	2.00	3	
V241321	0	1.00	0.84	1.34	-1.00	-1.00	1.00	2.00	3	
V241322x	0	1.00	2.64	2.11	-2.00	1.00	2.00	4.00	7	
V241323	0	1.00	1.51	1.22	-9.00	1.00	1.00	3.00	3	
V241324	0	1.00	3.21	2.08	-9.00	3.00	4.00	5.00	5	
V241325	0	1.00	3.81	2.00	-9.00	3.00	5.00	5.00	5	
V241326	0	1.00	4.00	1.88	-9.00	4.00	5.00	5.00	5	
V241327	0	1.00	3.54	2.00	-9.00	3.00	4.00	5.00	5	
V241328	0	1.00	1.68	1.61	-9.00	2.00	2.00	2.00	3	
V241329	0	1.00	0.76	1.28	-1.00	-1.00	1.00	2.00	3	
V241330x	0	1.00	4.55	2.57	-2.00	4.00	6.00	7.00	7	
V241331	0	1.00	1.74	1.59	-9.00	2.00	2.00	2.00	3	
V241332	0	1.00	0.70	1.30	-9.00	-1.00	1.00	2.00	3	
V241333x	0	1.00	4.64	2.51	-2.00	4.00	6.00	7.00	7	
V241334	0	1.00	3.11	2.00	-9.00	3.00	3.00	4.00	5	
V241335	0	1.00	2.16	1.57	-9.00	1.00	2.00	3.00	5	
V241336	0	1.00	1.11	1.33	-9.00	1.00	1.00	2.00	2	
V241337	0	1.00	1.49	1.04	-9.00	1.00	2.00	2.00	3	
V241338x	0	1.00	2.86	2.18	-2.00	1.00	2.00	5.00	6	
V241339	0	1.00	2.24	2.12	-9.00	1.00	2.00	4.00	5	
V241340	0	1.00	2.22	1.96	-9.00	1.00	2.00	3.00	5	
V241341	0	1.00	2.05	1.96	-9.00	1.00	2.00	3.00	5	
V241342	0	1.00	1.69	1.76	-9.00	1.00	2.00	3.00	3	
V241343	0	1.00	0.29	1.43	-9.00	-1.00	-1.00	1.00	3	
V241344x	0	1.00	3.02	2.13	-2.00	1.00	4.00	4.00	7	
V241345	0	1.00	0.83	1.99	-9.00	1.00	1.00	2.00	2	
V241346	0	1.00	1.18	1.03	-9.00	1.00	1.00	2.00	3	
V241347x	0	1.00	2.44	2.37	-2.00	1.00	2.00	5.00	6	
V241348	0	1.00	1.21	1.75	-9.00	1.00	2.00	2.00	2	
V241349	0	1.00	1.26	1.00	-9.00	1.00	1.00	2.00	3	
V241350x	0	1.00	3.72	2.48	-2.00	2.00	5.00	6.00	6	
V241351	0	1.00	0.93	1.83	-9.00	1.00	1.00	2.00	2	
V241352	0	1.00	1.65	1.26	-9.00	1.00	2.00	3.00	3	
V241353x	0	1.00	2.73	2.12	-2.00	1.00	3.00	4.00	6	
V241360	0	1.00	1.03	1.28	-9.00	1.00	1.00	1.00	3	
V241361	0	1.00	0.67	1.04	-9.00	1.00	1.00	1.00	2	
V241362	0	1.00	-0.90	0.53	-8.00	-1.00	-1.00	-1.00	2	
V241363x	0	1.00	1.33	1.14	-2.00	1.00	1.00	2.00	5	
V241364	0	1.00	1.35	1.75	-9.00	1.00	1.00	3.00	3	
V241365	0	1.00	0.62	1.32	-9.00	-1.00	1.00	2.00	3	
V241366x	0	1.00	2.58	2.19	-2.00	1.00	2.00	4.00	7	
V241367	0	1.00	1.20	1.24	-9.00	1.00	1.00	1.00	3	
V241368	0	1.00	0.83	1.18	-9.00	1.00	1.00	2.00	3	
V241369x	0	1.00	1.97	1.80	-2.00	1.00	1.00	3.00	7	
V241370	0	1.00	1.64	1.46	-9.00	1.00	2.00	2.00	3	
V241371	0	1.00	0.63	1.22	-9.00	-1.00	1.00	1.00	3	
V241372x	0	1.00	4.16	2.68	-2.00	2.00	4.00	7.00	7	
V241373	0	1.00	1.48	1.48	-9.00	1.00	1.00	2.00	3	
V241374	0	1.00	0.60	1.24	-9.00	-1.00	1.00	1.00	3	
V241375x	0	1.00	2.97	2.45	-2.00	1.00	3.00	4.00	7	
V241376	0	1.00	0.87	1.35	-9.00	1.00	1.00	1.00	2	
V241377	0	1.00	1.03	0.81	-9.00	1.00	1.00	1.00	2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241378x	0	1.00	1.36	1.25	-2.00	1.00	1.00	2.00	4	
V241379	0	1.00	0.83	1.59	-9.00	1.00	1.00	1.00	2	
V241380	0	1.00	1.24	1.01	-9.00	1.00	1.00	2.00	3	
V241381x	0	1.00	1.98	1.99	-2.00	1.00	1.00	3.00	6	
V241382	0	1.00	0.05	1.57	-9.00	-1.00	-1.00	1.00	3	
V241383	0	1.00	0.16	1.48	-9.00	-1.00	-1.00	1.00	3	
V241384	0	1.00	-0.14	1.17	-1.00	-1.00	-1.00	1.00	3	
V241385x	0	1.00	0.20	1.94	-2.00	-1.00	-1.00	1.00	6	
V241386	0	1.00	2.15	1.67	-9.00	1.00	3.00	3.00	4	
V241387	0	1.00	1.72	1.39	-9.00	1.00	2.00	3.00	3	
V241388	0	1.00	0.63	1.32	-9.00	-1.00	1.00	2.00	3	
V241389x	0	1.00	4.00	2.54	-2.00	2.00	4.00	6.00	7	
V241390	0	1.00	1.51	1.44	-9.00	2.00	2.00	2.00	2	
V241391	0	1.00	1.38	1.01	-9.00	1.00	1.00	2.00	3	
V241392x	0	1.00	4.37	2.20	-2.00	4.00	5.00	6.00	6	
V241393	0	1.00	1.52	1.32	-9.00	1.00	2.00	2.00	3	
V241394	0	1.00	0.72	1.20	-1.00	-1.00	1.00	1.00	3	
V241395x	0	1.00	3.37	2.60	-2.00	1.00	4.00	6.00	7	
V241396	0	1.00	1.52	1.20	-9.00	1.00	1.00	2.00	4	
V241397	0	1.00	12.29	28.34	-9.00	1.00	4.00	6.00	99	
V241398	0	1.00	1.47	1.55	-9.00	1.00	1.00	2.00	3	
V241399	0	1.00	0.72	1.37	-9.00	-1.00	1.00	2.00	3	
V241400x	0	1.00	3.09	2.37	-2.00	1.00	3.00	4.00	7	
V241401	0	1.00	1.56	1.57	-9.00	1.00	2.00	3.00	3	
V241402	0	1.00	0.66	1.39	-8.00	-1.00	1.00	2.00	3	
V241403x	0	1.00	3.30	2.40	-2.00	1.00	4.00	5.00	7	
V241404	0	1.00	1.40	1.63	-9.00	1.00	1.00	3.00	3	
V241405	0	1.00	0.65	1.38	-9.00	-1.00	1.00	2.00	3	
V241406x	0	1.00	2.66	2.15	-2.00	1.00	2.00	4.00	7	
V241407	0	1.00	2.27	2.03	-9.00	1.00	3.00	4.00	4	
V241408	0	1.00	0.32	1.75	-9.00	-1.00	-1.00	2.00	3	
V241409x	0	1.00	4.83	2.85	-2.00	3.00	5.00	8.00	8	
V241410	0	1.00	1.77	1.62	-9.00	1.00	2.00	3.00	3	
V241411	0	1.00	0.95	1.76	-1.00	-1.00	1.00	3.00	3	
V241412x	0	1.00	4.00	2.43	-2.00	2.00	4.00	6.00	7	
V241420	0	1.00	2.72	1.71	-9.00	1.00	3.00	4.00	5	
V241421	0	1.00	1.80	1.69	-9.00	1.00	2.00	3.00	5	
V241421z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241422	0	1.00	5.89	5.18	-9.00	1.00	5.00	11.00	12	
V241423	0	1.00	-0.18	1.75	-1.00	-1.00	-1.00	-1.00	4	
V241424	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241424z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241425	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241425z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241426	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241426z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241427	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241427z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241428	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241428z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241429	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241429z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241430	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241430z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241431	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241431z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241432	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241432z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241433	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241434	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241434z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241435	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241436	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241437	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241438	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241438z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241439	0	1.00	1.23	1.13	-9.00	1.00	1.00	2.00	2	
V241440	0	1.00	0.72	2.03	-9.00	-1.00	-1.00	2.00	5	
V241441	0	1.00	-0.65	0.87	-1.00	-1.00	-1.00	-1.00	2	
V241442	0	1.00	0.56	1.40	-9.00	-1.00	1.00	2.00	2	
V241443a	0	1.00	-0.19	1.17	-9.00	0.00	0.00	0.00	1	
V241443b	0	1.00	0.08	1.29	-9.00	0.00	0.00	1.00	1	
V241443c	0	1.00	-0.20	1.17	-9.00	0.00	0.00	0.00	1	
V241443d	0	1.00	-0.16	1.19	-9.00	0.00	0.00	0.00	1	
V241443e	0	1.00	-0.16	1.19	-9.00	0.00	0.00	0.00	1	
V241443f	0	1.00	-0.18	1.17	-9.00	0.00	0.00	0.00	1	
V241443g	0	1.00	0.02	1.27	-9.00	0.00	0.00	1.00	1	
V241443h	0	1.00	-0.05	1.24	-9.00	0.00	0.00	0.00	1	
V241444x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241445x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241450	0	1.00	1.24	1.66	-9.00	0.00	1.00	2.00	11	
V241451	0	1.00	2.80	1.70	-9.00	2.00	3.00	4.00	5	
V241452	0	1.00	2.32	1.78	-9.00	2.00	3.00	3.00	5	
V241453	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241454	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241455	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241456	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241457	0	1.00	2.86	15.18	-4.00	-1.00	-1.00	-1.00	80	
V241458x	0	1.00	48.40	21.17	-2.00	33.00	50.00	66.00	80	
V241459	0	1.00	2.76	2.39	-9.00	1.00	1.00	6.00	6	
V241460	0	1.00	-0.76	1.05	-9.00	-1.00	-1.00	-1.00	5	
V241461x	0	1.00	2.53	1.72	-2.00	1.00	2.00	5.00	5	
V241462	0	1.00	0.35	1.46	-9.00	-1.00	-1.00	2.00	2	
V241463	0	1.00	11.21	9.40	-9.00	9.00	11.00	13.00	95	
V241463z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241464	0	1.00	-0.72	1.13	-9.00	-1.00	-1.00	-1.00	5	
V241465x	0	1.00	3.17	1.49	-9.00	2.00	3.00	4.00	5	
V241466	0	1.00	-0.61	0.86	-1.00	-1.00	-1.00	-1.00	2	
V241467	0	1.00	5.97	9.33	-9.00	-1.00	8.00	12.00	95	
V241467z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241468	0	1.00	-0.78	0.67	-1.00	-1.00	-1.00	-1.00	2	
V241469x	0	1.00	1.23	2.40	-9.00	-1.00	1.00	3.00	5	
V241470	0	1.00	2.59	1.15	-9.00	3.00	3.00	3.00	3	
V241471	0	1.00	1.21	0.95	-9.00	1.00	1.00	2.00	2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241472	0	1.00	0.12	1.46	-9.00	-1.00	-1.00	2.00	2	
V241473	0	1.00	0.08	1.43	-1.00	-1.00	-1.00	2.00	2	
V241474	0	1.00	0.03	1.45	-9.00	-1.00	-1.00	2.00	2	
V241475	0	1.00	-0.98	0.24	-1.00	-1.00	-1.00	-1.00	2	
V241476	0	1.00	-0.01	1.40	-9.00	-1.00	-1.00	2.00	2	
V241477	0	1.00	-0.92	0.43	-1.00	-1.00	-1.00	-1.00	2	
V241478	0	1.00	0.22	1.73	-9.00	-1.00	-1.00	2.00	4	
V241479	0	1.00	0.40	1.15	-9.00	-1.00	1.00	1.00	2	
V241480	0	1.00	3.06	12.02	-9.00	-1.00	-1.00	-1.00	52	
V241481	0	1.00	24.13	22.79	-9.00	-1.00	30.00	40.00	168	
V241482	0	1.00	1.11	1.80	-9.00	-1.00	1.00	3.00	3	
V241483	0	1.00	2.18	2.86	-9.00	1.00	1.00	3.00	9	
V241484	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241485	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241486	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241487x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241488x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241489a	0	1.00	0.11	0.58	-9.00	0.00	0.00	0.00	1	
V241489b	0	1.00	0.22	0.65	-9.00	0.00	0.00	1.00	1	
V241489c	0	1.00	0.02	0.52	-9.00	0.00	0.00	0.00	1	
V241489d	0	1.00	0.01	0.50	-9.00	0.00	0.00	0.00	1	
V241489e	0	1.00	0.50	0.71	-9.00	0.00	1.00	1.00	1	
V241490	0	1.00	498.27	870.04	-9.00	-1.00	-1.00	-1.00	2024	
V241491	0	1.00	0.62	1.53	-9.00	-1.00	2.00	2.00	2	
V241492	0	1.00	0.53	1.57	-9.00	-1.00	1.00	1.00	5	
V241493	0	1.00	0.64	1.52	-9.00	-1.00	2.00	2.00	2	
V241494	0	1.00	0.95	2.13	-9.00	-1.00	1.00	2.00	7	
V241495a	0	1.00	-0.11	1.00	-9.00	-1.00	0.00	1.00	1	
V241495b	0	1.00	-0.43	0.71	-9.00	-1.00	0.00	0.00	1	
V241495c	0	1.00	-0.41	0.74	-9.00	-1.00	0.00	0.00	1	
V241495d	0	1.00	-0.31	0.85	-9.00	-1.00	0.00	0.00	1	
V241495e	0	1.00	-0.41	0.74	-9.00	-1.00	0.00	0.00	1	
V241495g	0	1.00	-0.39	0.76	-9.00	-1.00	0.00	0.00	1	
V241495f	0	1.00	-0.43	0.72	-9.00	-1.00	0.00	0.00	1	
V241496	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241497	0	1.00	1.59	1.15	-9.00	2.00	2.00	2.00	2	
V241498a	0	1.00	-0.74	0.64	-9.00	-1.00	-1.00	-1.00	1	
V241498b	0	1.00	-0.77	0.58	-9.00	-1.00	-1.00	-1.00	1	
V241498c	0	1.00	-0.80	0.51	-9.00	-1.00	-1.00	-1.00	1	
V241499	0	1.00	1.83	0.80	-9.00	2.00	2.00	2.00	2	
V241500a	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241500b	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241500c	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241500d	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241500e	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241501x	0	1.00	1.53	1.75	-9.00	1.00	1.00	2.00	6	
V241502	0	1.00	1.61	1.41	-9.00	2.00	2.00	2.00	2	
V241503	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241504	0	1.00	-0.33	1.04	-1.00	-1.00	-1.00	1.00	3	
V241505	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241506	0	1.00	1.09	1.14	-9.00	1.00	1.00	1.00	3	
V241507	0	1.00	1.08	1.26	-9.00	1.00	1.00	1.00	4	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241507z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241508	0	1.00	-0.86	0.57	-9.00	-1.00	-1.00	-1.00	3	
V241509	0	1.00	0.75	1.98	-9.00	0.00	0.00	2.00	4	
V241510	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241510z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241511	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241512x	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241513	0	1.00	-0.95	0.44	-9.00	-1.00	-1.00	-1.00	2	
V241514	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241515	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241516	0	1.00	-0.63	1.12	-1.00	-1.00	-1.00	-1.00	5	
V241517	0	1.00	1.03	1.51	-9.00	-1.00	2.00	2.00	2	
V241518a	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241518b	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241518c	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241518d	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241518e	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241519x	0	1.00	0.75	1.77	-9.00	-1.00	1.00	1.00	6	
V241520	0	1.00	0.75	1.31	-9.00	-1.00	1.00	2.00	3	
V241521	0	1.00	0.45	1.31	-9.00	0.00	0.00	1.00	9	
V241522a	0	1.00	-0.44	0.82	-9.00	-1.00	-1.00	0.00	1	
V241522b	0	1.00	-0.62	0.55	-9.00	-1.00	-1.00	0.00	1	
V241522c	0	1.00	-0.63	0.53	-9.00	-1.00	-1.00	0.00	1	
V241522d	0	1.00	-0.63	0.54	-9.00	-1.00	-1.00	0.00	1	
V241522e	0	1.00	-0.60	0.60	-9.00	-1.00	-1.00	0.00	1	
V241523	0	1.00	0.86	0.80	-9.00	1.00	1.00	1.00	2	
V241524	0	1.00	-0.89	0.57	-9.00	-1.00	-1.00	-1.00	2	
V241525	0	1.00	1.53	1.03	-9.00	1.00	2.00	2.00	2	
V241526	0	1.00	0.86	0.74	-9.00	1.00	1.00	1.00	2	
V241527	0	1.00	-0.95	0.37	-1.00	-1.00	-1.00	-1.00	2	
V241528	0	1.00	0.80	0.81	-9.00	1.00	1.00	1.00	2	
V241529	0	1.00	26.91	17.91	-9.00	12.00	26.00	42.00	59	
V241529z	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241530	0	1.00	1.22	1.36	-9.00	1.00	1.00	2.00	3	
V241531	0	1.00	0.44	1.32	-9.00	-1.00	1.00	1.00	2	
V241532	0	1.00	1.40	1.28	-9.00	1.00	2.00	2.00	2	
V241533	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241534	0	1.00	2.08	2.11	-9.00	1.00	2.00	4.00	5	
V241535	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241536	0	1.00	0.91	1.00	-9.00	1.00	1.00	1.00	2	
V241537	0	1.00	1.20	1.23	-9.00	1.00	1.00	2.00	2	
V241538	0	1.00	-0.81	0.67	-1.00	-1.00	-1.00	-1.00	2	
V241539	0	1.00	3.28	1.91	-9.00	3.00	4.00	5.00	5	
V241540	0	1.00	0.74	1.71	-9.00	-1.00	1.00	2.00	5	
V241541	0	1.00	1.46	1.24	-9.00	1.00	2.00	2.00	2	
V241550	0	1.00	1.44	1.11	-9.00	1.00	2.00	2.00	2	
V241551	0	1.00	1.30	1.21	-9.00	1.00	1.00	2.00	4	
V241551z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241552	0	1.00	1.69	1.18	-9.00	2.00	2.00	2.00	2	
V241553	0	1.00	0.87	1.41	-9.00	1.00	1.00	1.00	4	
V241553z	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241554	0	1.00	0.59	2.60	-9.00	1.00	1.00	2.00	2	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241555	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241556	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241557	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241558	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241559	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241560	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241561	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241562	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241563	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241564	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241565	0	1.00	-3.00	0.00	-3.00	-3.00	-3.00	-3.00	-3	
V241566x	0	1.00	14.14	10.87	-9.00	5.00	16.00	24.00	28	
V241567x	0	1.00	2.90	3.25	-9.00	2.00	4.00	5.00	6	
V241568a	0	1.00	0.03	1.17	-9.00	0.00	0.00	0.00	1	
V241568b	0	1.00	0.33	1.25	-9.00	0.00	0.00	1.00	1	
V241568c	0	1.00	0.15	1.21	-9.00	0.00	0.00	1.00	1	
V241568d	0	1.00	0.23	1.23	-9.00	0.00	0.00	1.00	1	
V241568e	0	1.00	0.07	1.18	-9.00	0.00	0.00	0.00	1	
V241568f	0	1.00	0.20	1.22	-9.00	0.00	0.00	1.00	1	
V241569	0	1.00	7197.89	31439.28	-9.00	-1.00	0.00	0.00	850000	
V241570	0	1.00	2.92	2.08	-9.00	2.00	3.00	4.00	5	
V241571	0	1.00	0.83	1.21	-9.00	1.00	1.00	1.00	2	
V241572	0	1.00	1.12	1.49	-9.00	1.00	1.00	2.00	5	
V241573	0	1.00	1.73	1.88	-9.00	1.00	2.00	3.00	5	
V241574a	0	1.00	-0.10	1.32	-9.00	0.00	0.00	0.00	1	
V241574b	0	1.00	-0.15	1.29	-9.00	0.00	0.00	0.00	1	
V241574c	0	1.00	-0.19	1.27	-9.00	0.00	0.00	0.00	1	
V241574d	0	1.00	-0.25	1.24	-9.00	0.00	0.00	0.00	1	
V241574e	0	1.00	-0.24	1.25	-9.00	0.00	0.00	0.00	1	
V241574f	0	1.00	-0.22	1.26	-9.00	0.00	0.00	0.00	1	
V241574g	0	1.00	-0.21	1.26	-9.00	0.00	0.00	0.00	1	
V241574h	0	1.00	-0.26	1.23	-9.00	0.00	0.00	0.00	1	
V241574i	0	1.00	-0.21	1.26	-9.00	0.00	0.00	0.00	1	
V241574j	0	1.00	0.28	1.43	-9.00	0.00	1.00	1.00	1	
V241575	0	1.00	1.41	1.83	-9.00	1.00	1.00	2.00	5	
V241576	0	1.00	1.31	1.78	-9.00	1.00	1.00	2.00	5	
V241577	0	1.00	2.10	2.12	-9.00	1.00	2.00	3.00	4	
V241578	0	1.00	1.90	1.81	-9.00	1.00	2.00	3.00	5	
V241579	0	1.00	1.01	1.62	-9.00	1.00	1.00	1.00	5	
V241580	0	1.00	1.25	1.74	-9.00	1.00	1.00	3.00	3	
V241581	0	1.00	0.70	1.54	-9.00	-1.00	1.00	2.00	3	
V241582x	0	1.00	2.25	1.82	-2.00	1.00	2.00	4.00	7	
V241583	0	1.00	0.80	5.80	-9.00	0.00	0.00	1.00	99	
V241600a	0	1.00	0.51	1.30	-9.00	0.00	1.00	1.00	1	
V241600b	0	1.00	0.06	1.22	-9.00	0.00	0.00	1.00	1	
V241600c	0	1.00	0.39	1.29	-9.00	0.00	1.00	1.00	1	
V241600d	0	1.00	0.19	1.26	-9.00	0.00	0.00	1.00	1	
V241600e	0	1.00	-0.18	1.11	-9.00	0.00	0.00	0.00	1	
V241601a	0	1.00	-0.01	0.36	-1.00	0.00	0.00	0.00	1	
V241601b	0	1.00	-0.01	0.37	-1.00	0.00	0.00	0.00	1	
V241601c	0	1.00	-0.01	0.36	-1.00	0.00	0.00	0.00	1	
V241601d	0	1.00	-0.01	0.37	-1.00	0.00	0.00	0.00	1	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241601e	0	1.00	-0.03	0.34	-1.00	0.00	0.00	0.00	1	
V241601f	0	1.00	0.00	0.38	-1.00	0.00	0.00	0.00	1	
V241601g	0	1.00	-0.03	0.34	-1.00	0.00	0.00	0.00	1	
V241601h	0	1.00	-0.02	0.35	-1.00	0.00	0.00	0.00	1	
V241601i	0	1.00	-0.02	0.36	-1.00	0.00	0.00	0.00	1	
V241601j	0	1.00	0.00	0.38	-1.00	0.00	0.00	0.00	1	
V241601k	0	1.00	-0.02	0.36	-1.00	0.00	0.00	0.00	1	
V241601m	0	1.00	0.02	0.41	-1.00	0.00	0.00	0.00	1	
V241601n	0	1.00	0.05	0.44	-1.00	0.00	0.00	0.00	1	
V241601p	0	1.00	0.05	0.44	-1.00	0.00	0.00	0.00	1	
V241601q	0	1.00	-0.04	0.32	-1.00	0.00	0.00	0.00	1	
V241601r	0	1.00	0.02	0.41	-1.00	0.00	0.00	0.00	1	
V241601s	0	1.00	0.00	0.39	-1.00	0.00	0.00	0.00	1	
V241602a	0	1.00	0.01	0.39	-1.00	0.00	0.00	0.00	1	
V241602b	0	1.00	-0.04	0.33	-1.00	0.00	0.00	0.00	1	
V241602c	0	1.00	0.09	0.48	-1.00	0.00	0.00	0.00	1	
V241602d	0	1.00	0.03	0.42	-1.00	0.00	0.00	0.00	1	
V241602e	0	1.00	0.05	0.44	-1.00	0.00	0.00	0.00	1	
V241602f	0	1.00	0.00	0.38	-1.00	0.00	0.00	0.00	1	
V241602g	0	1.00	0.00	0.38	-1.00	0.00	0.00	0.00	1	
V241602h	0	1.00	-0.01	0.37	-1.00	0.00	0.00	0.00	1	
V241602i	0	1.00	0.05	0.44	-1.00	0.00	0.00	0.00	1	
V241602j	0	1.00	0.03	0.42	-1.00	0.00	0.00	0.00	1	
V241602k	0	1.00	-0.03	0.34	-1.00	0.00	0.00	0.00	1	
V241602m	0	1.00	-0.03	0.34	-1.00	0.00	0.00	0.00	1	
V241602n	0	1.00	-0.03	0.34	-1.00	0.00	0.00	0.00	1	
V241602p	0	1.00	-0.05	0.31	-1.00	0.00	0.00	0.00	1	
V241602q	0	1.00	0.01	0.39	-1.00	0.00	0.00	0.00	1	
V241602r	0	1.00	0.04	0.42	-1.00	0.00	0.00	0.00	1	
V241603a	0	1.00	-0.88	0.34	-1.00	-1.00	-1.00	-1.00	1	
V241603b	0	1.00	-0.88	0.34	-1.00	-1.00	-1.00	-1.00	1	
V241603c	0	1.00	-0.89	0.32	-1.00	-1.00	-1.00	-1.00	1	
V241603d	0	1.00	-0.88	0.34	-1.00	-1.00	-1.00	-1.00	1	
V241603e	0	1.00	-0.89	0.32	-1.00	-1.00	-1.00	-1.00	1	
V241603f	0	1.00	-0.88	0.35	-1.00	-1.00	-1.00	-1.00	1	
V241603g	0	1.00	-0.89	0.33	-1.00	-1.00	-1.00	-1.00	1	
V241603h	0	1.00	-0.88	0.33	-1.00	-1.00	-1.00	-1.00	1	
V241603i	0	1.00	-0.89	0.32	-1.00	-1.00	-1.00	-1.00	1	
V241603j	0	1.00	-0.89	0.33	-1.00	-1.00	-1.00	-1.00	1	
V241603k	0	1.00	-0.87	0.37	-1.00	-1.00	-1.00	-1.00	1	
V241603m	0	1.00	-0.87	0.37	-1.00	-1.00	-1.00	-1.00	1	
V241603n	0	1.00	-0.89	0.33	-1.00	-1.00	-1.00	-1.00	1	
V241603p	0	1.00	-0.88	0.35	-1.00	-1.00	-1.00	-1.00	1	
V241603q	0	1.00	-0.89	0.32	-1.00	-1.00	-1.00	-1.00	1	
V241603r	0	1.00	-0.89	0.32	-1.00	-1.00	-1.00	-1.00	1	
V241604	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241605	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241606	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	
V241607	0	1.00	2.52	2.29	-9.00	2.00	3.00	4.00	5	
V241608	0	1.00	2.90	2.45	-9.00	2.00	3.00	5.00	5	
V241609	0	1.00	0.76	1.32	-9.00	1.00	1.00	1.00	2	
V241610	0	1.00	1502.96	832.20	-9.00	1776.00	1965.00	1997.00	2024	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
V241611	0	1.00	-1.00	0.54	-5.00	-1.00	-1.00	-1.00	2	
V241612	0	1.00	4.55	5.96	-9.00	3.00	4.00	6.00	99	
V241613	0	1.00	2.00	2.30	-9.00	1.00	2.00	4.00	4	
V241614	0	1.00	1.15	1.82	-9.00	1.00	2.00	2.00	2	
V241615	0	1.00	0.93	1.81	-9.00	1.00	1.00	2.00	2	
V241616	0	1.00	-1.04	0.41	-5.00	-1.00	-1.00	-1.00	-1	
V241617	0	1.00	-0.48	1.48	-9.00	-1.00	-1.00	1.00	7	
V241618	0	1.00	2.43	2.27	-9.00	1.00	2.00	4.00	7	
V241619	0	1.00	4.00	2.25	-9.00	4.00	5.00	5.00	5	
V241620	0	1.00	1.37	2.52	-9.00	-1.00	1.00	4.00	5	
V241621	0	1.00	2.03	1.73	-9.00	1.00	2.00	3.00	5	
V241622	0	1.00	-2.00	0.00	-2.00	-2.00	-2.00	-2.00	-2	

```r
# Get more info about the variables
colnames(anes_data)[1:20]  # Look at the first 20 column names
```

```
##  [1] "version"      "V240001"      "V200001"      "V160001_orig" "V240002"
##  [6] "V240003"      "V240101a"     "V240101c"     "V240101d"     "V240102a"
## [11] "V240102c"     "V240102d"     "V240103a"     "V240103c"     "V240103d"
## [16] "V240104a"     "V240104c"     "V240104d"     "V240105a"     "V240105c"
```

```r
# Check for variables related to demographics, political opinions, and voting
demographic_vars <- grep("V241[45][0-9]{2}", names(anes_data), value = TRUE)
political_vars <- grep("V2414[0-9]{2}", names(anes_data), value = TRUE)
voting_vars <- grep("V2410[0-9]{2}", names(anes_data), value = TRUE)

# Print these variable names
cat("Demographic variables:\n")
```

```
## Demographic variables:
```

```r
print(demographic_vars)
```

```
##   [1] "V241400x" "V241401"  "V241402"  "V241403x" "V241404"  "V241405"
##   [7] "V241406x" "V241407"  "V241408"  "V241409x" "V241410"  "V241411"
##  [13] "V241412x" "V241420"  "V241421"  "V241421z" "V241422"  "V241423"
##  [19] "V241424"  "V241424z" "V241425"  "V241425z" "V241426"  "V241426z"
##  [25] "V241427"  "V241427z" "V241428"  "V241428z" "V241429"  "V241429z"
##  [31] "V241430"  "V241430z" "V241431"  "V241431z" "V241432"  "V241432z"
##  [37] "V241433"  "V241434"  "V241434z" "V241435"  "V241436"  "V241437"
##  [43] "V241438"  "V241438z" "V241439"  "V241440"  "V241441"  "V241442"
##  [49] "V241443a" "V241443b" "V241443c" "V241443d" "V241443e" "V241443f"
##  [55] "V241443g" "V241443h" "V241444x" "V241445x" "V241450"  "V241451"
##  [61] "V241452"  "V241453"  "V241454"  "V241455"  "V241456"  "V241457"
##  [67] "V241458x" "V241459"  "V241460"  "V241461x" "V241462"  "V241463"
##  [73] "V241463z" "V241464"  "V241465x" "V241466"  "V241467"  "V241467z"
##  [79] "V241468"  "V241469x" "V241470"  "V241471"  "V241472"  "V241473"
##  [85] "V241474"  "V241475"  "V241476"  "V241477"  "V241478"  "V241479"
##  [91] "V241480"  "V241481"  "V241482"  "V241483"  "V241484"  "V241485"
##  [97] "V241486"  "V241487x" "V241488x" "V241489a" "V241489b" "V241489c"
```

```
## [103] "V241489d" "V241489e" "V241490"  "V241491"  "V241492"  "V241493"
## [109] "V241494"  "V241495a" "V241495b" "V241495c" "V241495d" "V241495e"
## [115] "V241495g" "V241495f" "V241496"  "V241497"  "V241498a" "V241498b"
## [121] "V241498c" "V241499"  "V241500a" "V241500b" "V241500c" "V241500d"
## [127] "V241500e" "V241501x" "V241502"  "V241503"  "V241504"  "V241505"
## [133] "V241506"  "V241507"  "V241507z" "V241508"  "V241509"  "V241510"
## [139] "V241510z" "V241511"  "V241512x" "V241513"  "V241514"  "V241515"
## [145] "V241516"  "V241517"  "V241518a" "V241518b" "V241518c" "V241518d"
## [151] "V241518e" "V241519x" "V241520"  "V241521"  "V241522a" "V241522b"
## [157] "V241522c" "V241522d" "V241522e" "V241523"  "V241524"  "V241525"
## [163] "V241526"  "V241527"  "V241528"  "V241529"  "V241529z" "V241530"
## [169] "V241531"  "V241532"  "V241533"  "V241534"  "V241535"  "V241536"
## [175] "V241537"  "V241538"  "V241539"  "V241540"  "V241541"  "V241550"
## [181] "V241551"  "V241551z" "V241552"  "V241553"  "V241553z" "V241554"
## [187] "V241555"  "V241556"  "V241557"  "V241558"  "V241559"  "V241560"
## [193] "V241561"  "V241562"  "V241563"  "V241564"  "V241565"  "V241566x"
## [199] "V241567x" "V241568a" "V241568b" "V241568c" "V241568d" "V241568e"
## [205] "V241568f" "V241569"  "V241570"  "V241571"  "V241572"  "V241573"
## [211] "V241574a" "V241574b" "V241574c" "V241574d" "V241574e" "V241574f"
## [217] "V241574g" "V241574h" "V241574i" "V241574j" "V241575"  "V241576"
## [223] "V241577"  "V241578"  "V241579"  "V241580"  "V241581"  "V241582x"
## [229] "V241583"
```

```r
cat("\nPolitical opinion variables:\n")
```

```
##
## Political opinion variables:
```

```r
print(political_vars)
```

```
##   [1] "V241400x" "V241401"  "V241402"  "V241403x" "V241404"  "V241405"
##   [7] "V241406x" "V241407"  "V241408"  "V241409x" "V241410"  "V241411"
##  [13] "V241412x" "V241420"  "V241421"  "V241421z" "V241422"  "V241423"
##  [19] "V241424"  "V241424z" "V241425"  "V241425z" "V241426"  "V241426z"
##  [25] "V241427"  "V241427z" "V241428"  "V241428z" "V241429"  "V241429z"
##  [31] "V241430"  "V241430z" "V241431"  "V241431z" "V241432"  "V241432z"
##  [37] "V241433"  "V241434"  "V241434z" "V241435"  "V241436"  "V241437"
##  [43] "V241438"  "V241438z" "V241439"  "V241440"  "V241441"  "V241442"
##  [49] "V241443a" "V241443b" "V241443c" "V241443d" "V241443e" "V241443f"
##  [55] "V241443g" "V241443h" "V241444x" "V241445x" "V241450"  "V241451"
##  [61] "V241452"  "V241453"  "V241454"  "V241455"  "V241456"  "V241457"
##  [67] "V241458x" "V241459"  "V241460"  "V241461x" "V241462"  "V241463"
##  [73] "V241463z" "V241464"  "V241465x" "V241466"  "V241467"  "V241467z"
##  [79] "V241468"  "V241469x" "V241470"  "V241471"  "V241472"  "V241473"
##  [85] "V241474"  "V241475"  "V241476"  "V241477"  "V241478"  "V241479"
##  [91] "V241480"  "V241481"  "V241482"  "V241483"  "V241484"  "V241485"
##  [97] "V241486"  "V241487x" "V241488x" "V241489a" "V241489b" "V241489c"
## [103] "V241489d" "V241489e" "V241490"  "V241491"  "V241492"  "V241493"
## [109] "V241494"  "V241495a" "V241495b" "V241495c" "V241495d" "V241495e"
## [115] "V241495g" "V241495f" "V241496"  "V241497"  "V241498a" "V241498b"
## [121] "V241498c" "V241499"
```

```r
cat("\nVoting variables:\n")
```

```
##
## Voting variables:
```

```r
print(voting_vars)
```

```
##  [1] "V241001"  "V241002"  "V241003"  "V241004"  "V241005"  "V241006"
##  [7] "V241007"  "V241008x" "V241009"  "V241010"  "V241011"  "V241012"
## [13] "V241013"  "V241014"  "V241015"  "V241016"  "V241017"  "V241018"
## [19] "V241019"  "V241020"  "V241021"  "V241022"  "V241023"  "V241024"
## [25] "V241025"  "V241026"  "V241027"  "V241028"  "V241029"  "V241030"
## [31] "V241031"  "V241032"  "V241033"  "V241034"  "V241035"  "V241036"
## [37] "V241037"  "V241038"  "V241039"  "V241039z" "V241040"  "V241041"
## [43] "V241042"  "V241043"  "V241043z" "V241044"  "V241045"  "V241046"
## [49] "V241046z" "V241047"  "V241048"  "V241049"  "V241050"  "V241051"
## [55] "V241051z" "V241052"  "V241053"  "V241053z" "V241054"  "V241055"
## [61] "V241055z" "V241056x" "V241057"  "V241058"  "V241058z" "V241059"
## [67] "V241060"  "V241060z" "V241061"  "V241062"  "V241062z" "V241063"
## [73] "V241064"  "V241064z" "V241065"  "V241066"  "V241066z" "V241067"
## [79] "V241068"  "V241068z" "V241069"  "V241070"  "V241070z" "V241071"
## [85] "V241072"  "V241072z" "V241073"  "V241074"  "V241074z" "V241075x"
## [91] "V241076x" "V241077x" "V241078x"
```

```r
# Check data types and missing values for these subsets
summary(anes_data[c(sample(demographic_vars, min(5, length(demographic_vars))),
                    sample(political_vars, min(5, length(political_vars))),
                    sample(voting_vars, min(5, length(voting_vars))))])
```

```
##     V241421z      V241577           V241529         V241580     
##  Min.   :-2   Min.   :-9.000   Min.   :-9.00   Min.   :-9.000  
##  1st Qu.:-2   1st Qu.: 1.000   1st Qu.:12.00   1st Qu.: 1.000  
##  Median :-2   Median : 2.000   Median :26.00   Median : 1.000  
##  Mean   :-2   Mean   : 2.099   Mean   :26.91   Mean   : 1.248  
##  3rd Qu.:-2   3rd Qu.: 3.000   3rd Qu.:42.00   3rd Qu.: 3.000  
##  Max.   :-2   Max.   : 4.000   Max.   :59.00   Max.   : 3.000  
##     V241495d           V241438z     V241428z     V241427z     V241432z 
##  Min.   :-9.0000   Min.   :-3   Min.   :-3   Min.   :-3   Min.   :-3  
##  1st Qu.:-1.0000   1st Qu.:-3   1st Qu.:-3   1st Qu.:-3   1st Qu.:-3  
##  Median : 0.0000   Median :-3   Median :-3   Median :-3   Median :-3  
##  Mean   :-0.3076   Mean   :-3   Mean   :-3   Mean   :-3   Mean   :-3  
##  3rd Qu.: 0.0000   3rd Qu.:-3   3rd Qu.:-3   3rd Qu.:-3   3rd Qu.:-3  
##  Max.   : 1.0000   Max.   :-3   Max.   :-3   Max.   :-3   Max.   :-3  
##     V241428      V241046           V241068z     V241053z     V241011  
##  Min.   :-3   Min.   :-8.0000   Min.   :-2   Min.   :-2   Min.   :-2  
##  1st Qu.:-3   1st Qu.:-1.0000   1st Qu.:-2   1st Qu.:-2   1st Qu.:-2  
##  Median :-3   Median :-1.0000   Median :-2   Median :-2   Median :-2  
##  Mean   :-3   Mean   :-0.9015   Mean   :-2   Mean   :-2   Mean   :-2  
##  3rd Qu.:-3   3rd Qu.:-1.0000   3rd Qu.:-2   3rd Qu.:-2   3rd Qu.:-2  
##  Max.   :-3   Max.   : 6.0000   Max.   :-2   Max.   :-2   Max.   :-2  
##     V241030       
##  Min.   :-8.0000  
```

```
##   1st Qu.:-1.0000
##   Median :-1.0000
##   Mean   :-0.7793
##   3rd Qu.:-1.0000
##   Max.   : 2.0000
```

I decided to look at the variable documentation to better understand what these variables represent

I will try to identify some key variables related to demographics, political opinions, and voting behavior

```r
# Check for some specific demographic variables
demographic_vars <- c(
  "V241458x",  # Age
  "V241550",   # Gender
  "V241501x",  # Race/ethnicity
  "V241465x",  # Education level
  "V241567x",  # Income
  "V241227x"   # Party ID
)

# Political opinion variables
opinion_vars <- c(
  "V241117",   # Country direction
  "V241137x",  # Approval of president
  "V241133x",  # Approval of Supreme Court
  "V241129x",  # Approval of Congress
  "V241177",   # Liberal-conservative self-placement
  "V241229"    # Trust in government
)

# Voting behavior variables
voting_vars <- c(
  "V241075x",  # Presidential vote/intent/preference
  "V241102x",  # How likely to vote
  "V241106x"   # Past vote in 2020
)

# Create a subset of the data with these variables
selected_vars <- c(demographic_vars, opinion_vars, voting_vars)

# Check if all these variables exist in the dataset
existing_vars <- selected_vars[selected_vars %in% names(anes_data)]
missing_vars <- selected_vars[!selected_vars %in% names(anes_data)]

cat("Existing variables:", length(existing_vars), "\n")
```

```
## Existing variables: 15
```

```r
print(existing_vars)
```

```
##  [1] "V241458x" "V241550"  "V241501x" "V241465x" "V241567x" "V241227x"
##  [7] "V241117"  "V241137x" "V241133x" "V241129x" "V241177"  "V241229"
## [13] "V241075x" "V241102x" "V241106x"
```

```r
cat("\nMissing variables:", length(missing_vars), "\n")
```

```
##
## Missing variables: 0
```

```r
print(missing_vars)
```

```
## character(0)
```

```r
# Create our dataset with the existing variables
if(length(existing_vars) > 0) {
  analysis_data <- anes_data[, existing_vars]
  # Look at the structure of our selected dataset
  str(analysis_data)
}
```

```
## 'data.frame':    3349 obs. of  15 variables:
##  $ V241458x: int  -2 40 29 39 35 80 62 32 62 30 ...
##  $ V241550 : int  2 2 2 2 2 2 1 2 1 1 ...
##  $ V241501x: int  1 1 1 1 3 1 6 1 2 1 ...
##  $ V241465x: int  4 5 4 5 5 5 3 5 2 4 ...
##  $ V241567x: int  5 5 1 5 5 5 5 1 2 5 ...
##  $ V241227x: int  3 3 1 3 4 1 3 1 1 1 ...
##  $ V241117 : int  -1 2 2 2 2 2 2 2 -1 2 ...
##  $ V241137x: int  2 2 2 3 4 2 3 2 1 3 ...
##  $ V241133x: int  3 4 4 3 4 4 4 3 4 4 ...
##  $ V241129x: int  -1 2 4 3 4 4 4 3 -1 3 ...
##  $ V241177 : int  2 3 2 99 1 4 3 2 2 2 ...
##  $ V241229 : int  5 2 3 2 4 3 4 3 2 3 ...
##  $ V241075x: int  30 20 20 -2 22 20 20 20 30 20 ...
##  $ V241102x: int  2 2 2 6 2 2 2 2 2 2 ...
##  $ V241106x: int  3 1 2 1 2 2 2 2 3 2 ...
```

b. Data Exploration

```r
# Get a summary of our selected variables
summary(analysis_data)
```

```
##     V241458x       V241550          V241501x         V241465x
##  Min.   :-2.0   Min.   :-9.000   Min.   :-9.000   Min.   :-9.000
##  1st Qu.:33.0   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.: 2.000
##  Median :50.0   Median : 2.000   Median : 1.000   Median : 3.000
```

```
##  Mean   :48.4   Mean   : 1.436   Mean   : 1.533   Mean   : 3.175
##  3rd Qu.:66.0   3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.: 4.000
##  Max.   :80.0   Max.   : 2.000   Max.   : 6.000   Max.   : 5.000
##    V241567x         V241227x         V241117          V241137x
##  Min.   :-9.000   Min.   :-9.000   Min.   :-9.000   Min.   :-2.000
##  1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 2.000
##  Median : 4.000   Median : 4.000   Median : 2.000   Median : 3.000
##  Mean   : 2.899   Mean   : 3.788   Mean   : 1.437   Mean   : 2.798
##  3rd Qu.: 5.000   3rd Qu.: 6.000   3rd Qu.: 2.000   3rd Qu.: 4.000
##  Max.   : 6.000   Max.   : 7.000   Max.   : 2.000   Max.   : 4.000
##    V241133x         V241129x         V241177         V241229
##  Min.   :-2.000   Min.   :-2.000   Min.   :-9.00   Min.   :-9.000
##  1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.: 3.00   1st Qu.: 3.000
##  Median : 3.000   Median : 4.000   Median : 4.00   Median : 4.000
##  Mean   : 2.781   Mean   : 2.849   Mean   :19.39   Mean   : 3.492
##  3rd Qu.: 4.000   3rd Qu.: 4.000   3rd Qu.: 6.00   3rd Qu.: 4.000
##  Max.   : 4.000   Max.   : 4.000   Max.   :99.00   Max.   : 5.000
##    V241075x        V241102x         V241106x
##  Min.   :-2.00   Min.   :-4.000   Min.   :-4.000
##  1st Qu.:20.00   1st Qu.: 2.000   1st Qu.: 1.000
##  Median :20.00   Median : 2.000   Median : 2.000
##  Mean   :19.05   Mean   : 2.671   Mean   : 2.021
##  3rd Qu.:21.00   3rd Qu.: 3.000   3rd Qu.: 3.000
##  Max.   :32.00   Max.   : 6.000   Max.   : 4.000
```

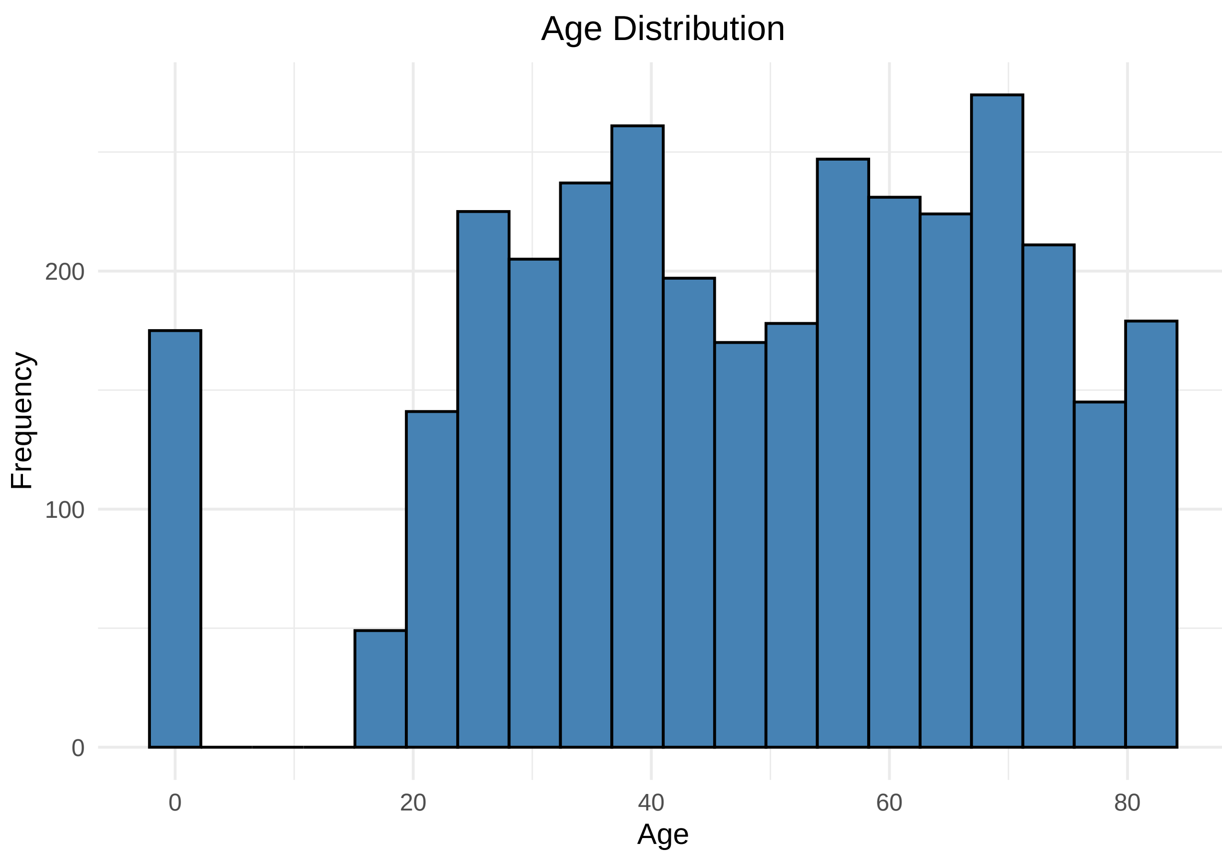
```r
# Check for missing values in each variable
missing_values <- colSums(is.na(analysis_data) | analysis_data < 0)
print(paste("Missing values per variable:"))
```

```
## [1] "Missing values per variable:"
```

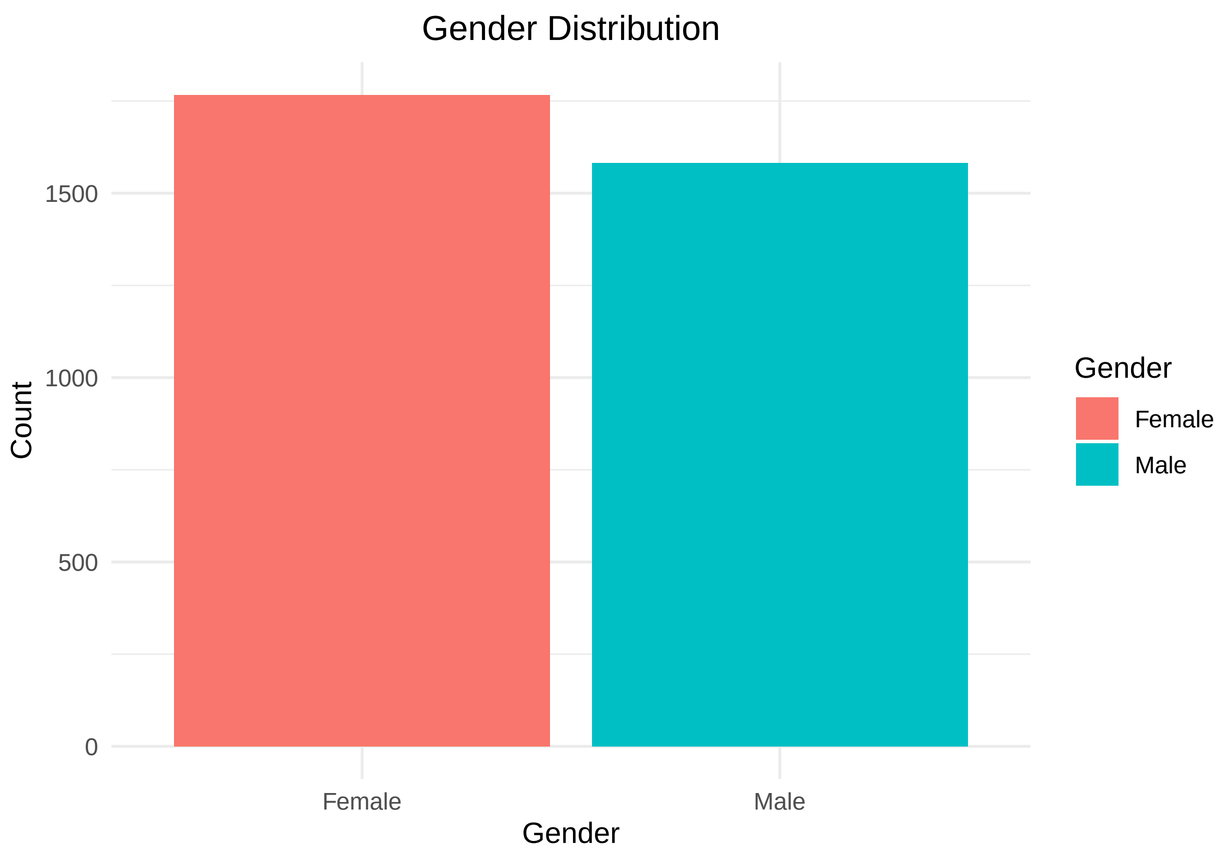
```r
print(missing_values)
```

```
## V241458x  V241550 V241501x V241465x V241567x V241227x  V241117 V241137x
##      175       32       47       51      203       31      288       40
## V241133x V241129x  V241177  V241229 V241075x V241102x V241106x
##      110      355       24       17      318        1       42
```

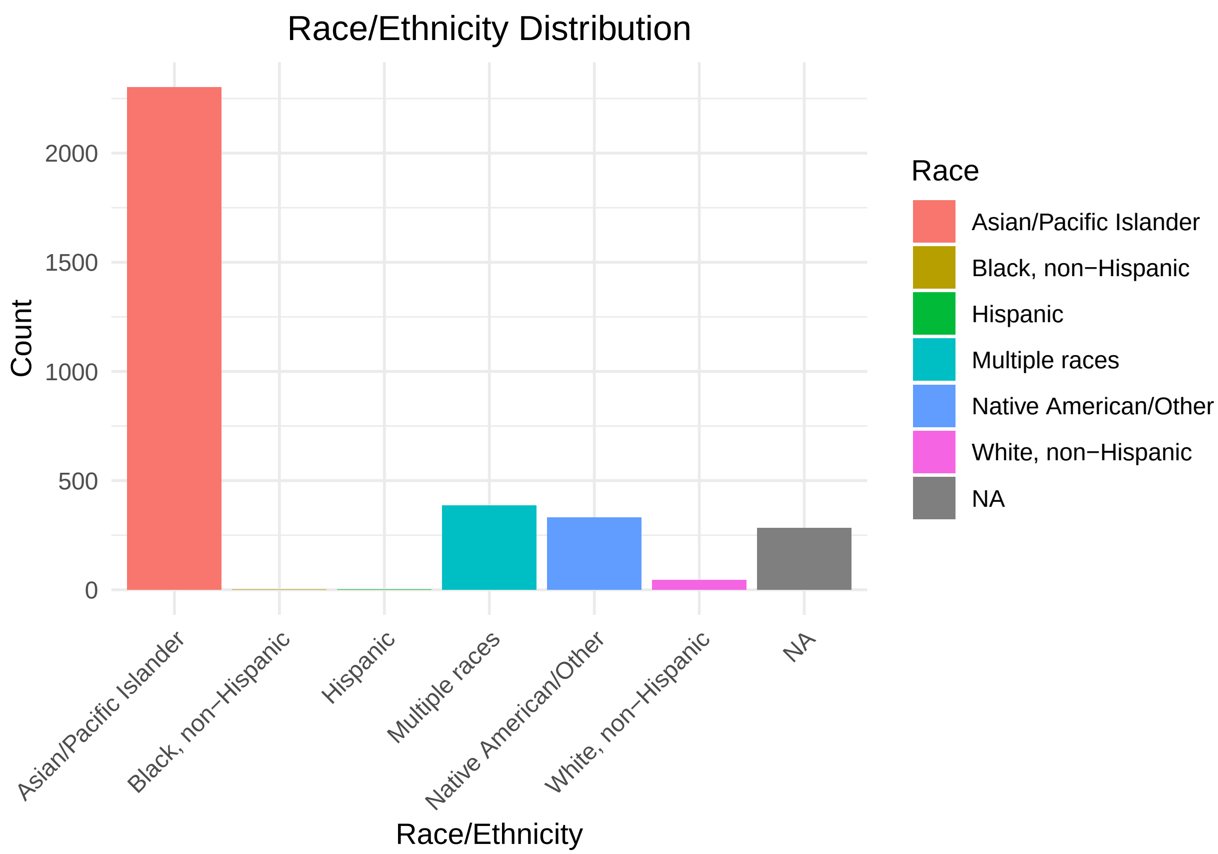
```r
# Explore the distributions of key variables

# Age distribution
ggplot(analysis_data, aes(x = V241458x)) +
  geom_histogram(bins = 20, fill = "steelblue", color = "black") +
  labs(title = "Age Distribution", x = "Age", y = "Frequency") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



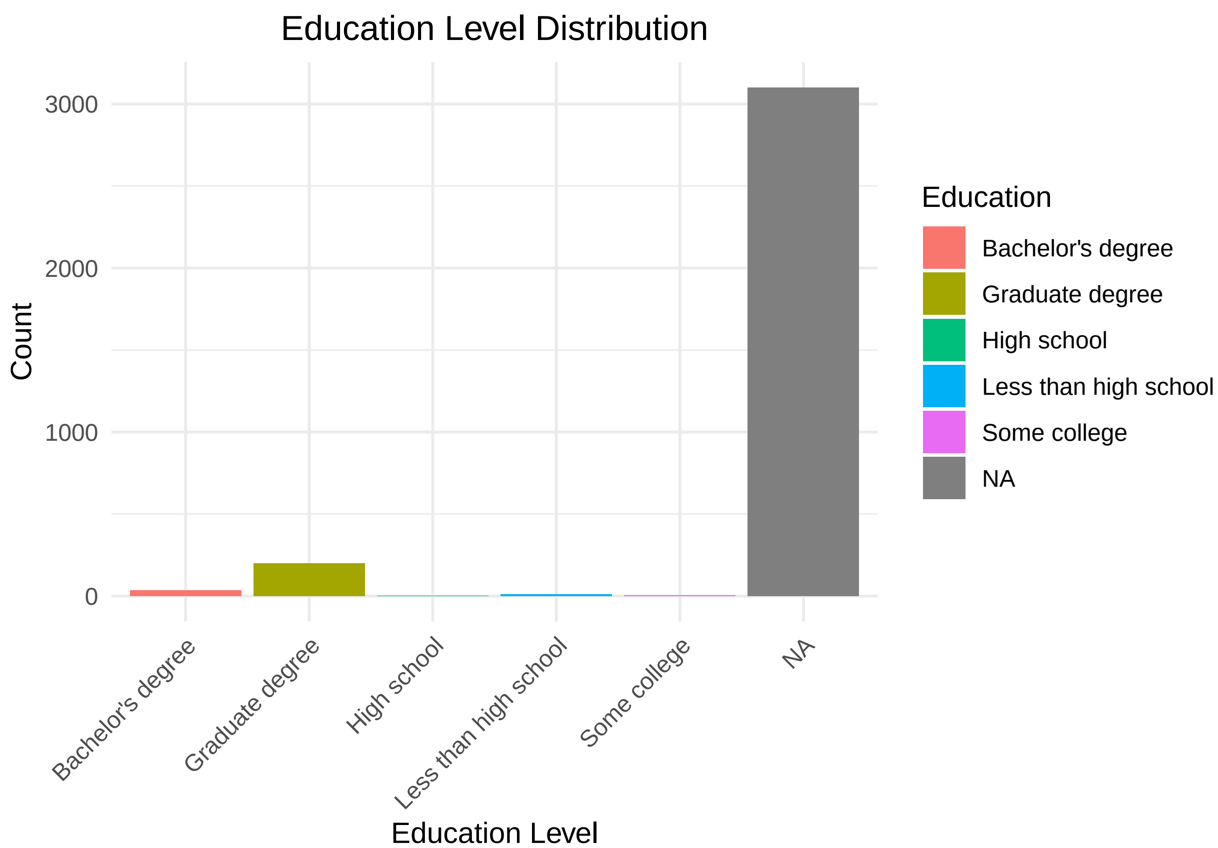

```r
# Gender distribution
gender_counts <- table(analysis_data$V241550)
gender_df <- data.frame(Gender = c("Male", "Female"), Count = as.numeric(gender_counts))
ggplot(gender_df, aes(x = Gender, y = Count, fill = Gender)) +
  geom_bar(stat = "identity") +
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



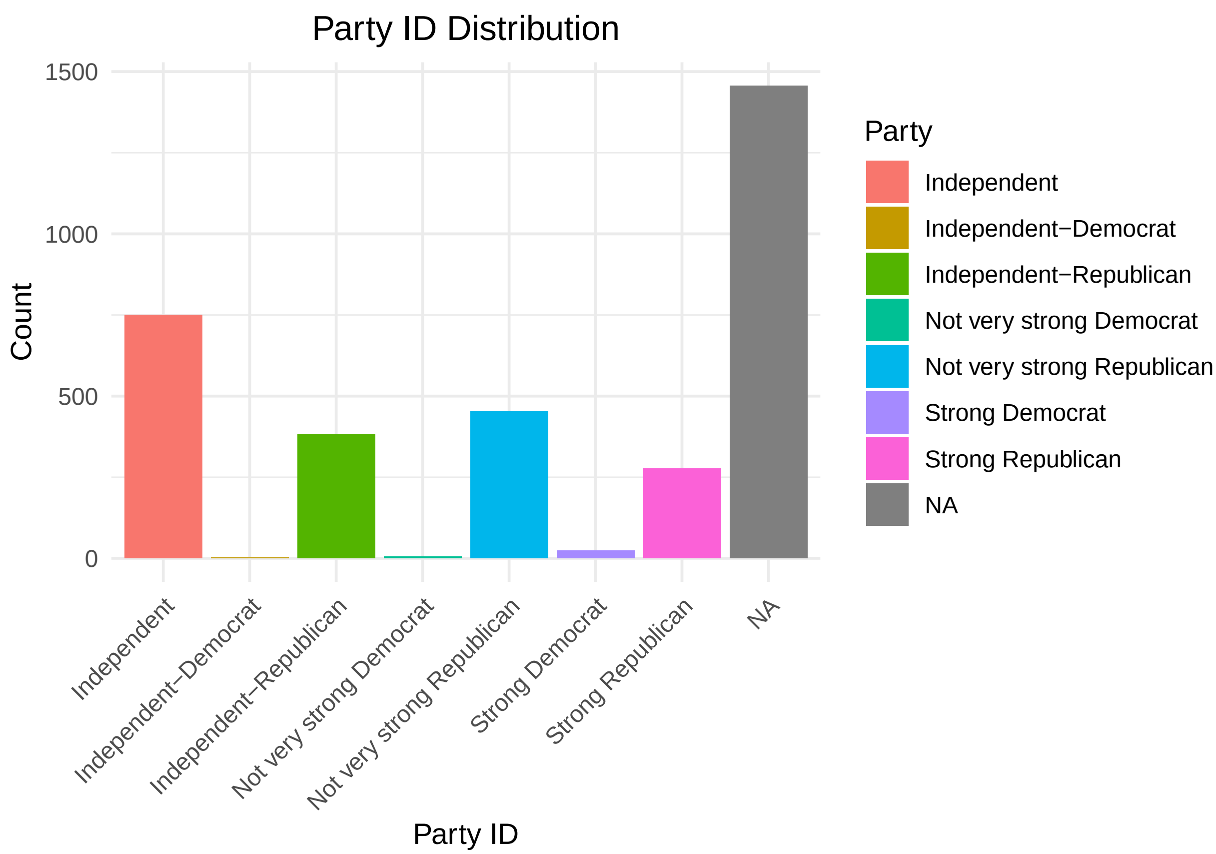

```
# Race/ethnicity distribution
race_counts <- table(analysis_data$V241501x)
race_labels <- c("White, non-Hispanic", "Black, non-Hispanic", "Hispanic",
                 "Asian/Pacific Islander", "Native American/Other", "Multiple races")
race_df <- data.frame(Race = race_labels[1:length(race_counts)], Count = as.numeric(race_counts))
ggplot(race_df, aes(x = Race, y = Count, fill = Race)) +
  geom_bar(stat = "identity") +
  labs(title = "Race/Ethnicity Distribution", x = "Race/Ethnicity", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```




```r
# Education level distribution
education_counts <- table(analysis_data$V241465x)
education_labels <- c("Less than high school", "High school", "Some college", "Bachelor's degree", "Grad
education_df <- data.frame(Education = education_labels[1:length(education_counts)],
                          Count = as.numeric(education_counts))
ggplot(education_df, aes(x = Education, y = Count, fill = Education)) +
  geom_bar(stat = "identity") +
  labs(title = "Education Level Distribution", x = "Education Level", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```




```r
# Party ID distribution
party_counts <- table(analysis_data$V241227x)
party_labels <- c("Strong Democrat", "Not very strong Democrat", "Independent-Democrat",
                  "Independent", "Independent-Republican", "Not very strong Republican", "Strong Republi
party_df <- data.frame(Party = party_labels[1:length(party_counts)], Count = as.numeric(party_counts))
ggplot(party_df, aes(x = Party, y = Count, fill = Party)) +
  geom_bar(stat = "identity") +
  labs(title = "Party ID Distribution", x = "Party ID", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```




```r
# relationships between variables

# Relationship between party ID and vote preference
vote_party_table <- table(analysis_data$V241227x, analysis_data$V241075x)
print("Party ID vs Vote Preference:")
```

```
## [1] "Party ID vs Vote Preference:"
```

```r
print(vote_party_table)
```

```
##
##        -2  10  11  12  20  21  22  30  31  32
##   -9   16   0   0   0   1   2   2   1   2   0
##   -8    1   0   0   0   2   0   0   2   0   0
##   -4    0   0   0   0   0   0   0   1   1   0
##   1    11  48   0   2 611   9   9  59   0   1
##   2    27  10   1   0 283  21  17  17   4   2
##   3    42  15   0   1 287  24  42  33   4   5
##   4   109   2   0   2  48  40  43  10  18   5
##   5    58   2   8   2  28 241  52   4  43   5
##   6    34   0   9   0  33 196  23   2  33   7
##   7    20   2  28   0  11 528   8   3  76   0
```

```r
# Relationship between education and party ID
edu_party_table <- table(analysis_data$V241465x, analysis_data$V241227x)
print("Education vs Party ID:")
```

```
## [1] "Education vs Party ID:"
```

```r
print(edu_party_table)
```

```
##
##       -9  -8  -4   1   2   3   4   5   6   7
##   -9   4   0   0   2   0   2   0   1   0   3
##   -8   1   0   0   0   0   0   0   0   0   0
##   -4   0   0   0   4   0   0   0   0   0   0
##   -2   0   0   0   5   4   6   2   3   3  11
##   1    2   1   1  40  15  23  29  29  17  41
##   2    5   2   0 122  54  69  73 102  68 164
##   3    6   0   0 182 114 136 106 144 106 244
##   4    3   2   0 205 103 144  31  95  89 141
##   5    3   0   1 190  92  73  36  69  54  72
```

```r
# Correlation matrix for numeric variables
# remove missing values and convert categorical variables
clean_data <- analysis_data
clean_data[clean_data < 0] <- NA  # Convert negative values to NA
correlation_matrix <- cor(clean_data, use = "pairwise.complete.obs")
print("Correlation Matrix:")
```

```
## [1] "Correlation Matrix:"
```

```r
print(correlation_matrix)
```

```
##              V241458x      V241550     V241501x     V241465x     V241567x
## V241458x  1.00000000 -0.036027706 -0.180419600  0.010814849 -0.035932554
## V241550  -0.03602771  1.000000000  0.008954665 -0.003059053 -0.098617007
## V241501x -0.18041960  0.008954665  1.000000000 -0.083407329 -0.119996936
## V241465x  0.01081485 -0.003059053 -0.083407329  1.000000000  0.435879263
## V241567x -0.03593255 -0.098617007 -0.119996936  0.435879263  1.000000000
## V241227x  0.01573130 -0.085655506 -0.105352252 -0.149886333 -0.009368483
## V241117  -0.14447191  0.027832695  0.029461964 -0.170955975 -0.092981651
## V241137x -0.15226942 -0.024938342 -0.022937908 -0.180708240 -0.045505376
## V241133x -0.06233311  0.126689934 -0.012633764  0.099106828  0.023966718
## V241129x  0.12830833 -0.040819658 -0.106050371  0.038142409  0.065485344
## V241177  -0.03370729  0.065232223  0.089799792 -0.306789048 -0.265290620
## V241229  -0.17743128 -0.002003807 -0.037280751 -0.082312304 -0.024896452
## V241075x -0.03604039 -0.024234308  0.012255338 -0.151623658 -0.073404374
## V241102x -0.27114112  0.014449425  0.133598864 -0.279688516 -0.204432626
## V241106x  0.28493349 -0.042960333 -0.187917486  0.170752829  0.158027611
##              V241227x     V241117     V241137x     V241133x    V241129x
## V241458x  0.015731301 -0.14447191 -0.15226942 -0.062333106  0.12830833
## V241550  -0.085655506  0.02783269 -0.02493834  0.126689934 -0.04081966
## V241501x -0.105352252  0.02946196 -0.02293791 -0.012633764 -0.10605037
```

```
## V241465x -0.149886333 -0.17095598 -0.18070824  0.099106828  0.03814241
## V241567x -0.009368483 -0.09298165 -0.04550538  0.023966718  0.06548534
## V241227x  1.000000000  0.41773505  0.71172545 -0.462834470  0.09951491
## V241117   0.417735049  1.00000000  0.54806383 -0.201463819  0.17922544
## V241137x  0.711725445  0.54806383  1.00000000 -0.378200246  0.16824714
## V241133x -0.462834470 -0.20146382 -0.37820025  1.000000000  0.23572641
## V241129x  0.099514910  0.17922544  0.16824714  0.235726410  1.00000000
## V241177   0.015249261  0.09348785  0.05789793  0.005025075 -0.03947146
## V241229   0.296651243  0.33688325  0.41677723 -0.067899514  0.25347618
## V241075x  0.174820317  0.15999899  0.19487789 -0.117597313  0.03248988
## V241102x  0.059135668  0.17588338  0.15718520 -0.009718066 -0.07407644
## V241106x  0.303697225  0.09526456  0.19799861 -0.228558327  0.10270879
##               V241177      V241229    V241075x     V241102x    V241106x
## V241458x -0.033707287 -0.177431285 -0.03604039 -0.271141124  0.28493349
## V241550   0.065232223 -0.002003807 -0.02423431  0.014449425 -0.04296033
## V241501x  0.089799792 -0.037280751  0.01225534  0.133598864 -0.18791749
## V241465x -0.306789048 -0.082312304 -0.15162366 -0.279688516  0.17075283
## V241567x -0.265290620 -0.024896452 -0.07340437 -0.204432626  0.15802761
## V241227x  0.015249261  0.296651243  0.17482032  0.059135668  0.30369723
## V241117   0.093487850  0.336883245  0.15999899  0.175883379  0.09526456
## V241137x  0.057897929  0.416777228  0.19487789  0.157185196  0.19799861
## V241133x  0.005025075 -0.067899514 -0.11759731 -0.009718066 -0.22855833
## V241129x -0.039471459  0.253476180  0.03248988 -0.074076445  0.10270879
## V241177   1.000000000  0.070446465  0.11220841  0.295278278 -0.18704794
## V241229   0.070446465  1.000000000  0.07818879  0.161667730  0.03932393
## V241075x  0.112208412  0.078188791  1.00000000  0.372275389 -0.08642692
## V241102x  0.295278278  0.161667730  0.37227539  1.000000000 -0.46400873
## V241106x -0.187047939  0.039323928 -0.08642692 -0.464008727  1.00000000
```

Some insights:

Demographics: The age distribution shows two peaks: one around 35-40 and another around 65-70, with a median age of 50. The gender distribution shows slightly more females (56.4%) than males (43.6%). Race/ethnicity distribution appears to have labeling issues, as Asian/Pacific Islander shows an unusually high count. The education level distribution shows potential issues with NA values dominating the chart. Party ID distribution shows a mix of affiliations with many NA values.

Missing Values: Several variables have significant numbers of missing values: V241129x (Approval of Congress): 355 missing values (10.6%) V241075x (Presidential vote/intent/preference): 318 missing values (9.5%) V241567x (Income): 203 missing values (6.1%) V241458x (Age): 175 missing values (5.2%)

Correlations: Strong positive correlation (0.71) between party ID (V241227x) and presidential approval (V241137x), suggesting political alignment. Moderate positive correlation (0.44) between education (V241465x) and income (V241567x). Moderate negative correlation (-0.46) between party ID (V241227x) and Supreme Court approval (V241133x). Moderate negative correlation (-0.46) between voting likelihood (V241102x) and past voting behavior (V241106x).

Party ID vs. Vote Preference: There's a clear pattern showing Democrats (values 1-3) predominantly supporting Democratic candidates (code 10, 20, 30) and Republicans (values 5-7) predominantly supporting Republican candidates (code 11, 21, 31). Independents (value 4) show more varied voting preferences.

c. Data Cleaning

```
# Create a clean dataset for analysis
clean_data <- analysis_data

# 1. Handle missing values (negative values in this dataset)
# Convert negative values to NA
clean_data[clean_data < 0] <- NA

# 2. Check missing values after conversion
missing_after <- colSums(is.na(clean_data))
print(paste("Missing values after conversion:"))
```

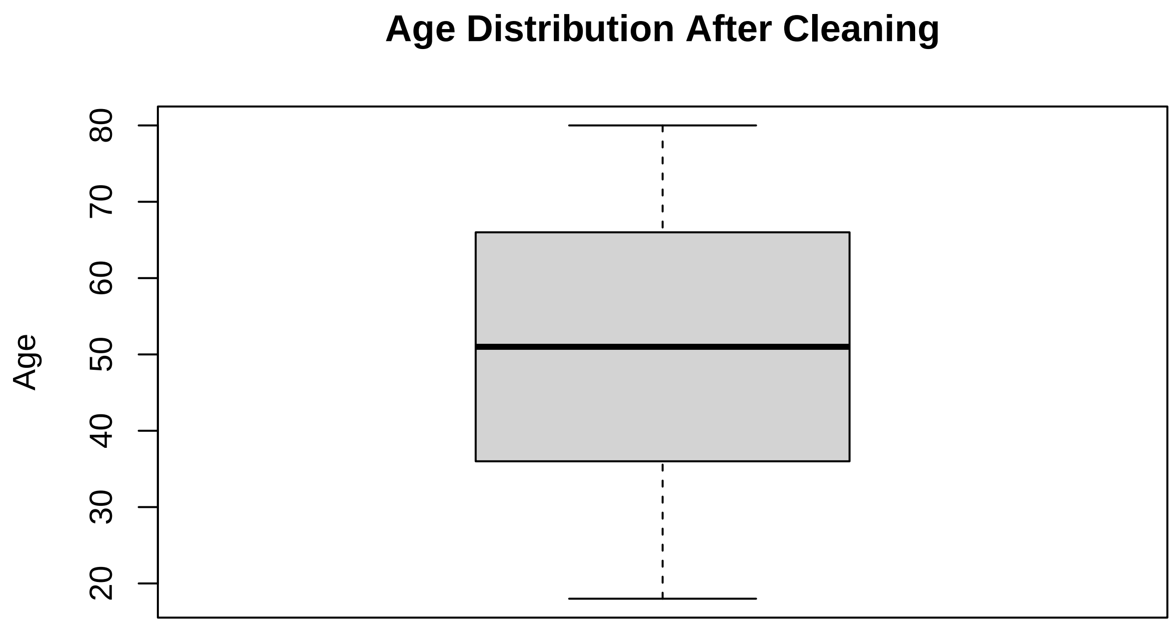
```
## [1] "Missing values after conversion:"
```

```
print(missing_after)
```

```
## V241458x  V241550 V241501x V241465x V241567x V241227x  V241117 V241137x
##      175       32       47       51      203       31      288       40
## V241133x V241129x  V241177  V241229 V241075x V241102x V241106x
##      110      355       24       17      318        1       42
```

```
# 3. Decide on strategy for each variable with missing values
# For demographic variables: impute with median/mode as appropriate
# For age: impute with median
clean_data$V241458x[is.na(clean_data$V241458x)] <- median(clean_data$V241458x, na.rm = TRUE)

# For gender: impute with mode
gender_mode <- as.numeric(names(sort(table(clean_data$V241550), decreasing = TRUE)[1]))
clean_data$V241550[is.na(clean_data$V241550)] <- gender_mode

# For race/ethnicity: impute with mode
race_mode <- as.numeric(names(sort(table(clean_data$V241501x), decreasing = TRUE)[1]))
clean_data$V241501x[is.na(clean_data$V241501x)] <- race_mode

# For education: impute with median
clean_data$V241465x[is.na(clean_data$V241465x)] <- median(clean_data$V241465x, na.rm = TRUE)

# For income: impute with median
clean_data$V241567x[is.na(clean_data$V241567x)] <- median(clean_data$V241567x, na.rm = TRUE)

# For party ID: impute with median
clean_data$V241227x[is.na(clean_data$V241227x)] <- median(clean_data$V241227x, na.rm = TRUE)

# For opinion variables: impute with median
clean_data$V241117[is.na(clean_data$V241117)] <- median(clean_data$V241117, na.rm = TRUE)
clean_data$V241137x[is.na(clean_data$V241137x)] <- median(clean_data$V241137x, na.rm = TRUE)
clean_data$V241133x[is.na(clean_data$V241133x)] <- median(clean_data$V241133x, na.rm = TRUE)
clean_data$V241129x[is.na(clean_data$V241129x)] <- median(clean_data$V241129x, na.rm = TRUE)
clean_data$V241177[is.na(clean_data$V241177)] <- median(clean_data$V241177, na.rm = TRUE)
clean_data$V241229[is.na(clean_data$V241229)] <- median(clean_data$V241229, na.rm = TRUE)
```

```r
# For voting variables: impute with median
clean_data$V241075x[is.na(clean_data$V241075x)] <- median(clean_data$V241075x, na.rm = TRUE)
clean_data$V241102x[is.na(clean_data$V241102x)] <- median(clean_data$V241102x, na.rm = TRUE)
clean_data$V241106x[is.na(clean_data$V241106x)] <- median(clean_data$V241106x, na.rm = TRUE)

# 4. Check for outliers in numeric variables
# For age, 99 might be a code for "don't know" rather than actual age
boxplot(clean_data$V241458x, main="Age Distribution After Cleaning", ylab="Age")
```




```r
# For V241177 (Liberal-conservative self-placement), 99 is likely a special code
table(clean_data$V241177)
```

```
##
##   1   2   3   4   5   6   7  99
## 149 482 301 778 339 563 195 542
```

```r
clean_data$V241177[clean_data$V241177 == 99] <- NA
clean_data$V241177[is.na(clean_data$V241177)] <- median(clean_data$V241177, na.rm = TRUE)

# 5. Verify no missing values remain
any_missing <- any(is.na(clean_data))
print(paste("Any missing values remaining:", any_missing))
```

```
## [1] "Any missing values remaining: FALSE"
```

```r
# 6. Check the structure of the cleaned data
str(clean_data)
```

```
## 'data.frame':    3349 obs. of  15 variables:
##  $ V241458x: num  51 40 29 39 35 80 62 32 62 30 ...
##  $ V241550 : num  2 2 2 2 2 2 1 2 1 1 ...
##  $ V241501x: num  1 1 1 1 3 1 6 1 2 1 ...
##  $ V241465x: num  4 5 4 5 5 5 3 5 2 4 ...
##  $ V241567x: num  5 5 1 5 5 5 5 1 2 5 ...
##  $ V241227x: num  3 3 1 3 4 1 3 1 1 1 ...
##  $ V241117 : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ V241137x: int  2 2 2 3 4 2 3 2 1 3 ...
##  $ V241133x: int  3 4 4 3 4 4 4 3 4 4 ...
##  $ V241129x: num  4 2 4 3 4 4 4 3 4 3 ...
##  $ V241177 : int  2 3 2 4 1 4 3 2 2 2 ...
##  $ V241229 : num  5 2 3 2 4 3 4 3 2 3 ...
##  $ V241075x: int  30 20 20 21 22 20 20 20 30 20 ...
##  $ V241102x: num  2 2 2 6 2 2 2 2 2 2 ...
##  $ V241106x: int  3 1 2 1 2 2 2 2 3 2 ...
```

```r
# 7. Summary statistics after cleaning
summary(clean_data)
```

```
##     V241458x        V241550         V241501x       V241465x       V241567x    
##  Min.   :18.00   Min.   :1.000   Min.   :1.00   Min.   :1.00   Min.   :1.000  
##  1st Qu.:36.00   1st Qu.:1.000   1st Qu.:1.00   1st Qu.:2.00   1st Qu.:3.000  
##  Median :51.00   Median :2.000   Median :1.00   Median :3.00   Median :4.000  
##  Mean   :51.17   Mean   :1.536   Mean   :1.67   Mean   :3.28   Mean   :3.663  
##  3rd Qu.:66.00   3rd Qu.:2.000   3rd Qu.:2.00   3rd Qu.:4.00   3rd Qu.:5.000  
##  Max.   :80.00   Max.   :2.000   Max.   :6.00   Max.   :5.00   Max.   :6.000  
##     V241227x        V241117         V241137x        V241133x    
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000  
##  Median :4.000   Median :2.000   Median :3.000   Median :3.000  
##  Mean   :3.904   Mean   :1.793   Mean   :2.858   Mean   :2.945  
##  3rd Qu.:6.000   3rd Qu.:2.000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :7.000   Max.   :2.000   Max.   :4.000   Max.   :4.000  
##     V241129x        V241177         V241229         V241075x    
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :10.00  
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:20.00  
##  Median :4.000   Median :4.000   Median :4.000   Median :21.00  
##  Mean   :3.411   Mean   :4.101   Mean   :3.554   Mean   :21.23  
##  3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:21.00  
##  Max.   :4.000   Max.   :7.000   Max.   :5.000   Max.   :32.00  
##     V241102x        V241106x    
##  Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:1.000  
##  Median :2.000   Median :2.000  
##  Mean   :2.672   Mean   :2.071  
##  3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :6.000   Max.   :4.000  
```

d. Data Preprocessing

```
# 1. Create variable names for better interpretability
names(clean_data) <- c("Age", "Gender", "Race", "Education", "Income", "PartyID",
                       "CountryDirection", "PresidentApproval", "SupremeCourtApproval",
                       "CongressApproval", "IdeologicalPlacement", "TrustGovernment",
                       "VotePreference", "VoteLikelihood", "PastVote2020")

# 2. Recode categorical variables for better interpretation
# Gender: 1 = Male, 2 = Female
clean_data$Gender <- factor(clean_data$Gender, levels = c(1, 2), labels = c("Male", "Female"))

# Race: 1 = White, 2 = Black, 3 = Hispanic, 4 = Asian/PI, 5 = Native Am/Other, 6 = Multiple
clean_data$Race <- factor(clean_data$Race,
                          levels = 1:6,
                          labels = c("White", "Black", "Hispanic", "Asian/PI", "Native Am/Other", "Multi

# Education: 1 = Less than HS, 2 = HS, 3 = Some college, 4 = Bachelor's, 5 = Graduate
clean_data$Education <- factor(clean_data$Education,
                               levels = 1:5,
                               labels = c("Less than HS", "HS", "Some college", "Bachelor's", "Graduate"

# Party ID: 1 = Strong Dem, 2 = Not strong Dem, 3 = Lean Dem, 4 = Independent, 5 = Lean Rep, 6 = Not st
clean_data$PartyID <- factor(clean_data$PartyID,
                             levels = 1:7,
                             labels = c("Strong Dem", "Not strong Dem", "Lean Dem",
                                        "Independent", "Lean Rep", "Not strong Rep", "Strong Rep"))

# Country Direction: 1 = Right direction, 2 = Wrong track
clean_data$CountryDirection <- factor(clean_data$CountryDirection,
                                      levels = 1:2,
                                      labels = c("Right direction", "Wrong track"))

# 3. Create a political leaning variable (simplified party ID)
clean_data$PoliticalLeaning <- "Independent"
clean_data$PoliticalLeaning[clean_data$PartyID %in% c("Strong Dem", "Not strong Dem", "Lean Dem")] <- "
clean_data$PoliticalLeaning[clean_data$PartyID %in% c("Strong Rep", "Not strong Rep", "Lean Rep")] <- "
clean_data$PoliticalLeaning <- factor(clean_data$PoliticalLeaning)

# 4. Recode VotePreference for easier interpretation
# Values 10, 20, 30 are Democratic; 11, 21, 31 are Republican; 12, 22, 32 are Other
clean_data$VoteChoice <- "Other"
clean_data$VoteChoice[clean_data$VotePreference %in% c(10, 20, 30)] <- "Democrat"
clean_data$VoteChoice[clean_data$VotePreference %in% c(11, 21, 31)] <- "Republican"
clean_data$VoteChoice <- factor(clean_data$VoteChoice)

# 5. Create a binary vote preference variable for classification
clean_data$VoteBinary <- ifelse(clean_data$VoteChoice == "Democrat", 1, 0)
clean_data$VoteBinary[clean_data$VoteChoice == "Other"] <- NA  # Remove "Other" votes for binary classi

# 6. Select variables for clustering and classification
# For clustering, we'll use demographic and opinion variables
cluster_vars <- c("Age", "Education", "Income", "IdeologicalPlacement",
```

```r
                 "PresidentApproval", "SupremeCourtApproval", "CongressApproval", "TrustGovernment")

# 7. Normalize numeric variables for clustering
# First, identify which variables need normalization
numeric_vars <- c("Age", "Income", "IdeologicalPlacement",
                 "PresidentApproval", "SupremeCourtApproval", "CongressApproval", "TrustGovernment")

# Create a new dataframe with normalized variables
normalized_data <- clean_data
for(var in numeric_vars) {
  normalized_data[[var]] <- scale(clean_data[[var]])
}

# 8. Check the preprocessed data
str(normalized_data)
```

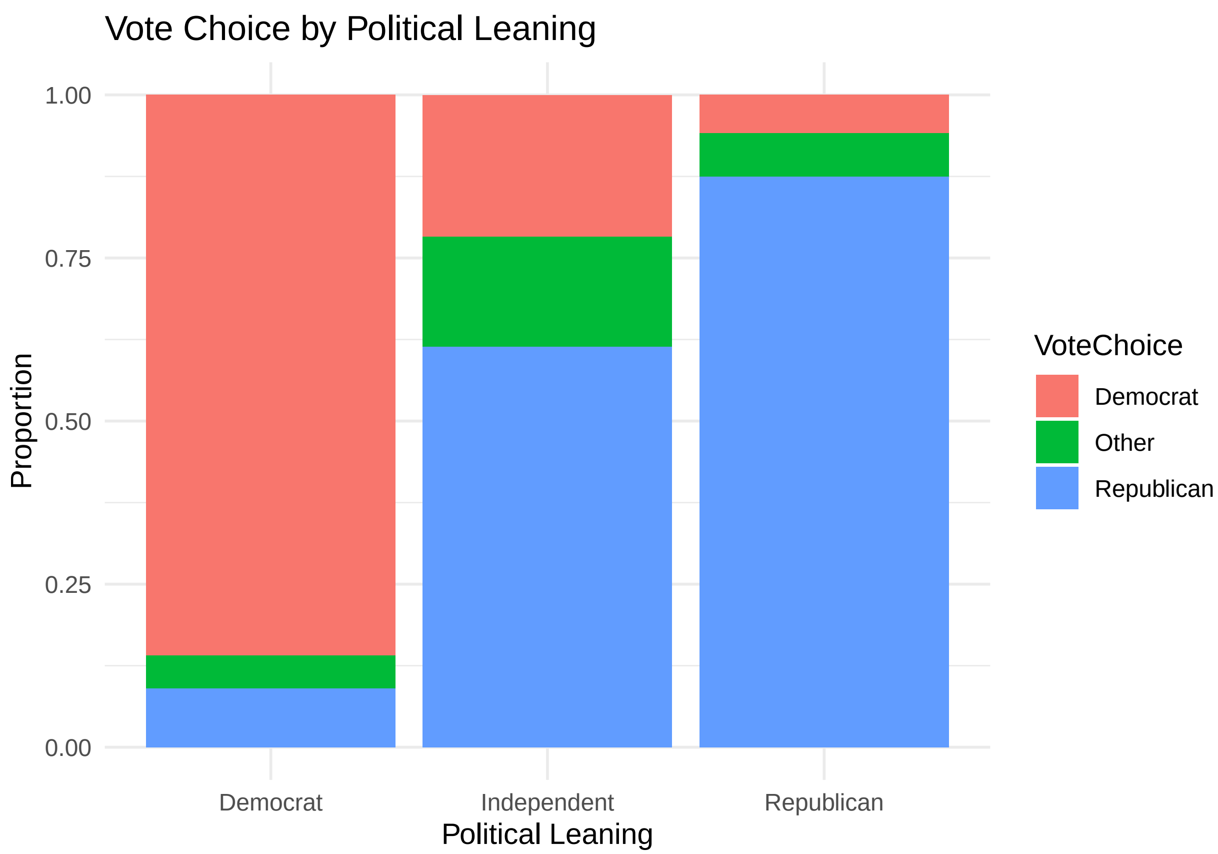
```
## 'data.frame':    3349 obs. of  18 variables:
##  $ Age                 : num [1:3349, 1] -0.00941 -0.63594 -1.26248 -0.6929 -0.92073 ...
##   ..- attr(*, "scaled:center")= num 51.2
##   ..- attr(*, "scaled:scale")= num 17.6
##  $ Gender              : Factor w/ 2 levels "Male","Female": 2 2 2 2 2 2 1 2 1 1 ...
##  $ Race                : Factor w/ 6 levels "White","Black",..: 1 1 1 1 3 1 6 1 2 1 ...
##  $ Education           : Factor w/ 5 levels "Less than HS",..: 4 5 4 5 5 5 3 5 2 4 ...
##  $ Income              : num [1:3349, 1] 0.96 0.96 -1.91 0.96 0.96 ...
##   ..- attr(*, "scaled:center")= num 3.66
##   ..- attr(*, "scaled:scale")= num 1.39
##  $ PartyID             : Factor w/ 7 levels "Strong Dem","Not strong Dem",..: 3 3 1 3 4 1 3 1 1 1 ..
##  $ CountryDirection    : Factor w/ 2 levels "Right direction",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ PresidentApproval   : num [1:3349, 1] -0.699 -0.699 -0.699 0.116 0.931 ...
##   ..- attr(*, "scaled:center")= num 2.86
##   ..- attr(*, "scaled:scale")= num 1.23
##  $ SupremeCourtApproval: num [1:3349, 1] 0.0481 0.9275 0.9275 0.0481 0.9275 ...
##   ..- attr(*, "scaled:center")= num 2.95
##   ..- attr(*, "scaled:scale")= num 1.14
##  $ CongressApproval    : num [1:3349, 1] 0.681 -1.634 0.681 -0.476 0.681 ...
##   ..- attr(*, "scaled:center")= num 3.41
##   ..- attr(*, "scaled:scale")= num 0.864
##  $ IdeologicalPlacement: num [1:3349, 1] -1.3693 -0.7175 -1.3693 -0.0658 -2.021 ...
##   ..- attr(*, "scaled:center")= num 4.1
##   ..- attr(*, "scaled:scale")= num 1.53
##  $ TrustGovernment     : num [1:3349, 1] 1.494 -1.606 -0.572 -1.606 0.461 ...
##   ..- attr(*, "scaled:center")= num 3.55
##   ..- attr(*, "scaled:scale")= num 0.968
##  $ VotePreference      : int  30 20 20 21 22 20 20 20 30 20 ...
##  $ VoteLikelihood      : num  2 2 2 6 2 2 2 2 2 2 ...
##  $ PastVote2020        : int  3 1 2 1 2 2 2 2 3 2 ...
##  $ PoliticalLeaning    : Factor w/ 3 levels "Democrat","Independent",..: 1 1 1 1 2 1 1 1 1 1 ...
##  $ VoteChoice          : Factor w/ 3 levels "Democrat","Other",..: 1 1 1 3 2 1 1 1 1 1 ...
##  $ VoteBinary          : num  1 1 1 0 NA 1 1 1 1 1 ...
```

```r
summary(normalized_data[, c(numeric_vars, "PoliticalLeaning", "VoteChoice")])
```

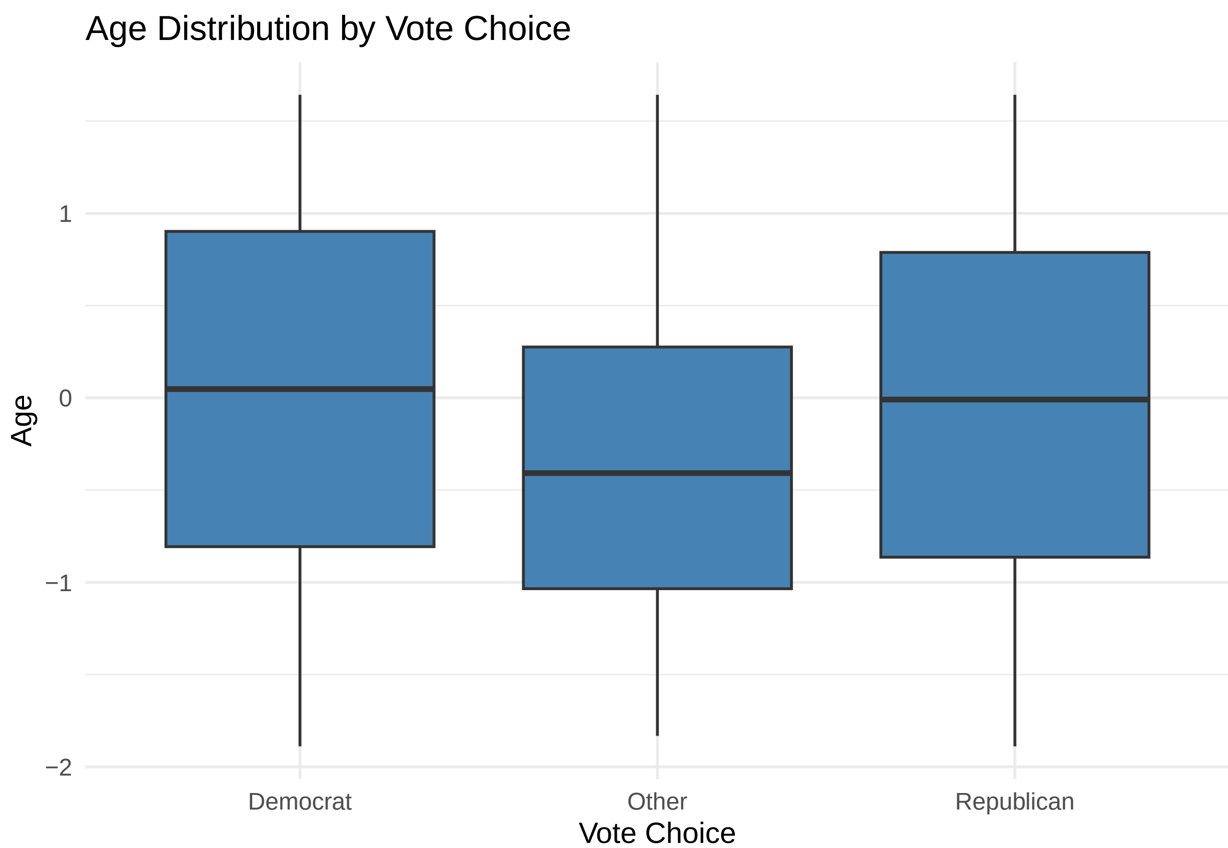
```
##         Age.V1              Income.V1       IdeologicalPlacement.V1
```

```
##  Min.   :-1.8890150   Min.   :-1.9110954   Min.   :-2.0210124
##  1st Qu.:-0.8637732   1st Qu.:-0.4755771   1st Qu.:-0.7175226
##  Median :-0.0094051   Median : 0.2421821   Median :-0.0657778
##  Mean   : 0.0000000   Mean   : 0.0000000   Mean   : 0.0000000
##  3rd Qu.: 0.8449630   3rd Qu.: 0.9599413   3rd Qu.: 0.5859671
##  Max.   : 1.6423733   Max.   : 1.6777004   Max.   : 1.8894568
##  PresidentApproval.V1 SupremeCourtApproval.V1 CongressApproval.V1
##  Min.   :-1.5138795   Min.   :-1.7107959      Min.   :-2.7923165
##  1st Qu.:-0.6990318   1st Qu.:-0.8313706      1st Qu.:-0.4764502
##  Median : 0.1158159   Median : 0.0480546      Median : 0.6814829
##  Mean   : 0.0000000   Mean   : 0.0000000      Mean   : 0.0000000
##  3rd Qu.: 0.9306636   3rd Qu.: 0.9274798      3rd Qu.: 0.6814829
##  Max.   : 0.9306636   Max.   : 0.9274798      Max.   : 0.6814829
##   TrustGovernment.V1     PoliticalLeaning      VoteChoice
##  Min.   :-2.6385233   Democrat   :1585    Democrat  :1515
##  1st Qu.:-0.5724923   Independent: 308    Other     : 228
##  Median : 0.4605232   Republican :1456    Republican:1606
##  Mean   : 0.0000000
##  3rd Qu.: 0.4605232
##  Max.   : 1.4935387
```

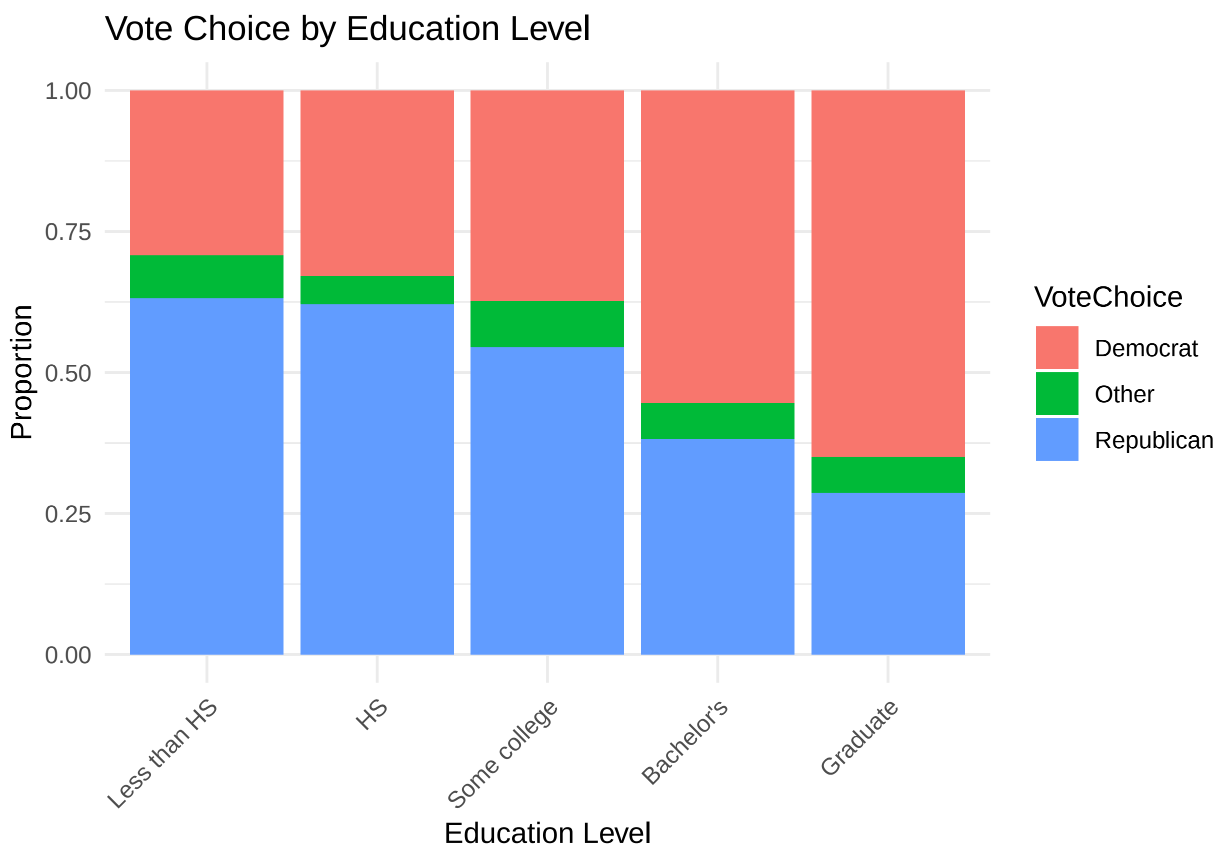
```r
# 9. Visualize key relationships in the preprocessed data
# Relationship between Political Leaning and Vote Choice
ggplot(normalized_data, aes(x = PoliticalLeaning, fill = VoteChoice)) +
  geom_bar(position = "fill") +
  labs(title = "Vote Choice by Political Leaning",
       y = "Proportion",
       x = "Political Leaning") +
  theme_minimal()
```




```r
# Relationship between Age and Vote Choice
ggplot(normalized_data, aes(x = VoteChoice, y = Age)) +
  geom_boxplot(fill = "steelblue") +
  labs(title = "Age Distribution by Vote Choice",
       y = "Age",
       x = "Vote Choice") +
  theme_minimal()
```




```r
# Relationship between Education and Vote Choice
ggplot(normalized_data, aes(x = Education, fill = VoteChoice)) +
  geom_bar(position = "fill") +
  labs(title = "Vote Choice by Education Level",
       y = "Proportion",
       x = "Education Level") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Some insights:

Variable Transformation: Successfully normalized numeric variables to enable fair comparisons in clustering. Created meaningful factor variables from numeric codes (Gender, Race, Education, PartyID). Developed derived variables like PoliticalLeaning and VoteChoice to simplify analysis.

Data Structure: We now have 18 variables, including both original and derived variables. The normalized variables have mean 0 and standard deviation 1, making them suitable for clustering.

Relationship Visualizations: Vote Choice by Political Leaning: Shows strong alignment between political leaning and voting behavior: Democrats overwhelmingly vote for Democratic candidates (~85%) Republicans overwhelmingly vote for Republican candidates (~90%) Independents show mixed voting patterns, with a slight Republican preference

Age Distribution by Vote Choice: "Other" voters tend to be slightly younger Democratic and Republican voters have similar age distributions

Vote Choice by Education Level: There's a clear educational gradient in voting patterns Higher education levels (Bachelor's and Graduate degrees) show stronger Democratic preferences Lower education levels show stronger Republican preferences

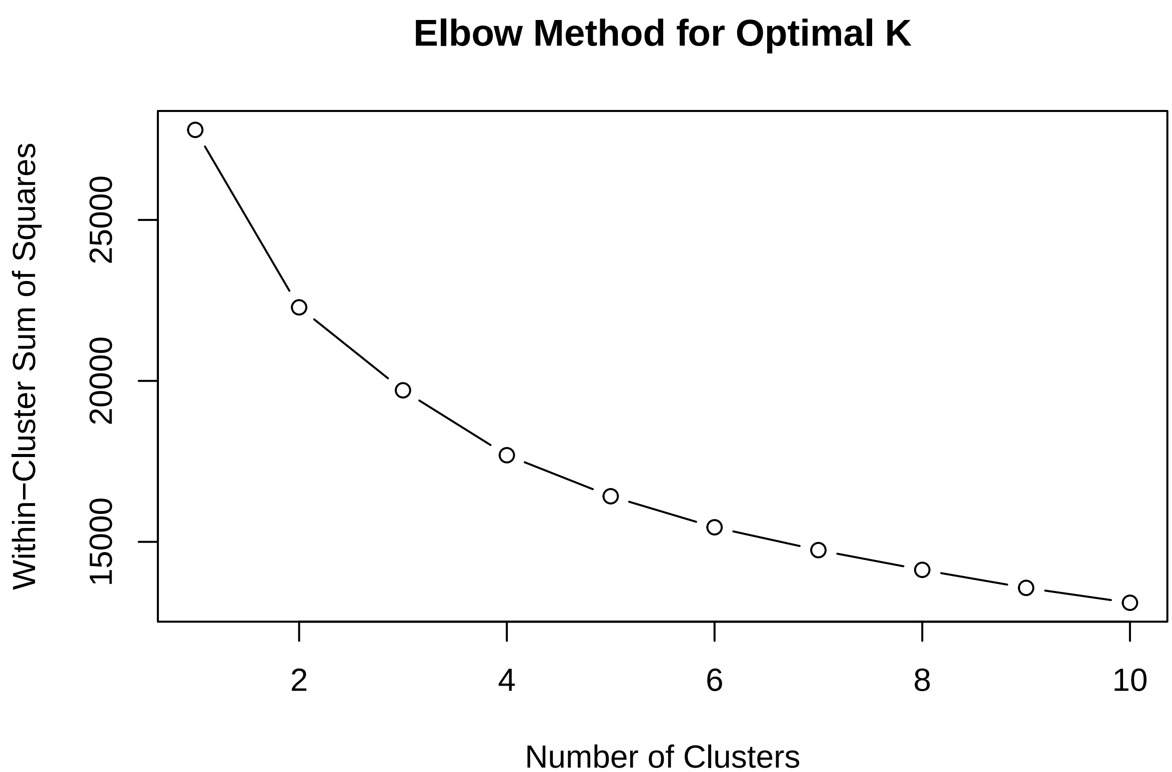
e. Clustering

```
# 1. Prepare data for clustering
# Select variables for clustering
```

```r
cluster_data <- normalized_data[, cluster_vars]

# Convert factor variables to numeric for clustering
cluster_data$Education <- as.numeric(cluster_data$Education)

# 2. Remove any remaining NA values
cluster_data <- na.omit(cluster_data)

# 3. Determine optimal number of clusters using the elbow method
set.seed(123) # For reproducibility
wss <- numeric(10)
for (i in 1:10) {
  kmeans_result <- kmeans(cluster_data, centers = i, nstart = 25)
  wss[i] <- kmeans_result$tot.withinss
}
```

```
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
```

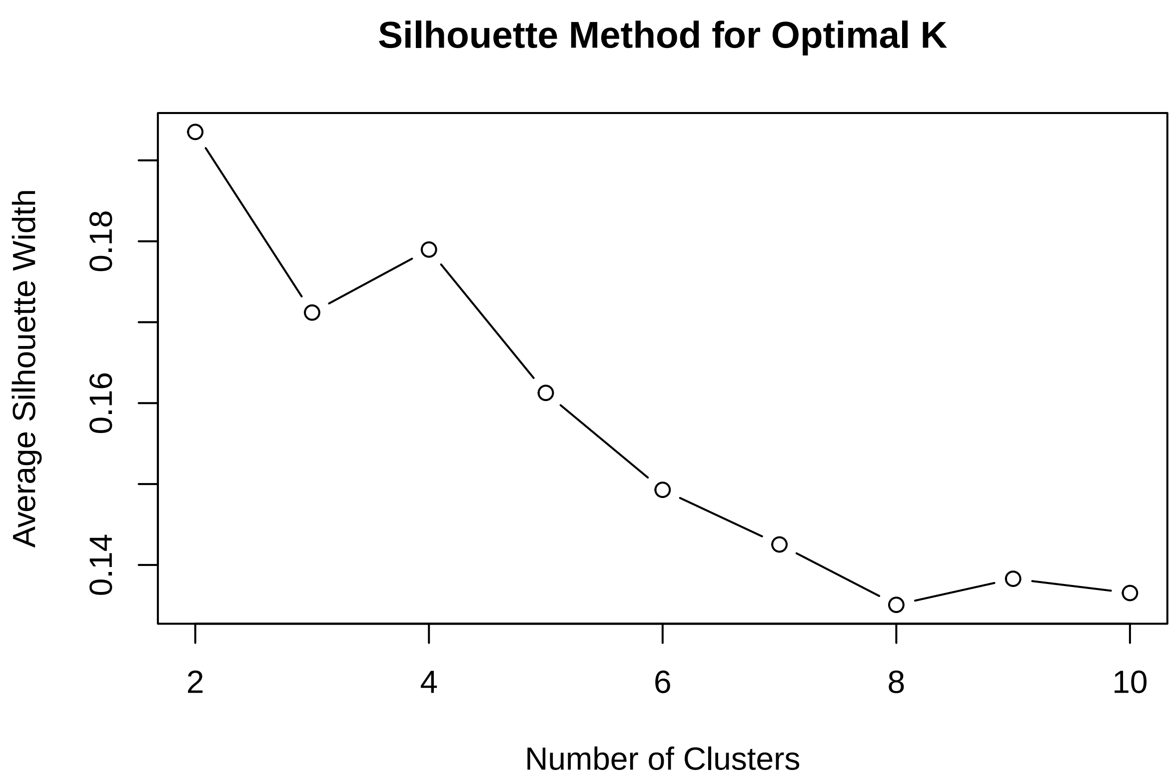
```r
# Plot the elbow curve
plot(1:10, wss, type = "b", xlab = "Number of Clusters",
     ylab = "Within-Cluster Sum of Squares",
     main = "Elbow Method for Optimal K")
```



```r
# 4. Use silhouette method to validate cluster number
library(cluster)
sil_width <- numeric(9)
for (i in 2:10) {
  kmeans_result <- kmeans(cluster_data, centers = i, nstart = 25)
  sil <- silhouette(kmeans_result$cluster, dist(cluster_data))
  sil_width[i-1] <- mean(sil[,3])
}
```

```
## Warning: did not converge in 10 iterations
```

```r
# Plot silhouette widths
plot(2:10, sil_width, type = "b", xlab = "Number of Clusters",
     ylab = "Average Silhouette Width",
     main = "Silhouette Method for Optimal K")
```




```r
# 5. Perform K-means clustering with the optimal number of clusters
# Based on the elbow and silhouette plots, let's choose k=3 for now
set.seed(123)
kmeans_result <- kmeans(cluster_data, centers = 3, nstart = 25)

# 6. Add cluster assignments back to the original data
normalized_data$cluster <- as.factor(kmeans_result$cluster)
```

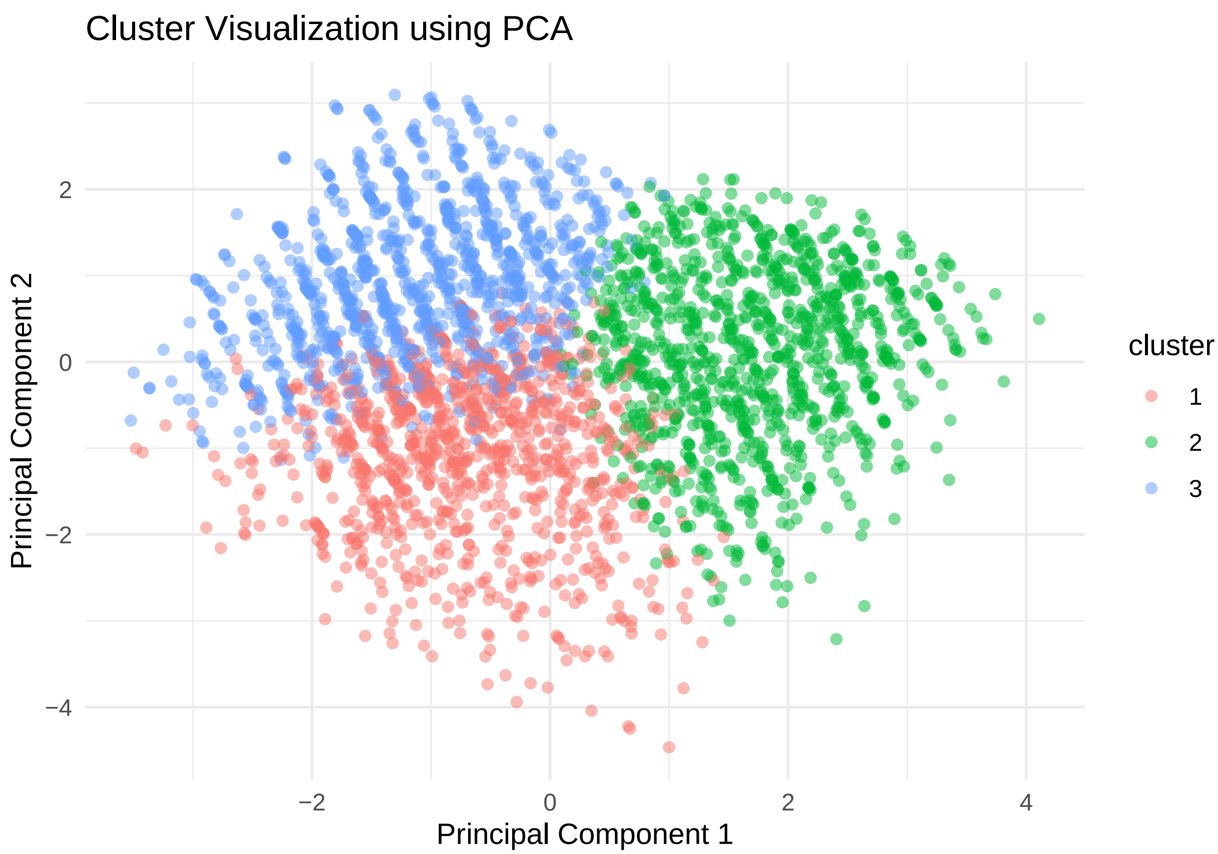
```r
# 7. Examine the clusters
# Cluster sizes
table(normalized_data$cluster)
```

```
##
##    1    2    3
## 1059 1169 1121
```

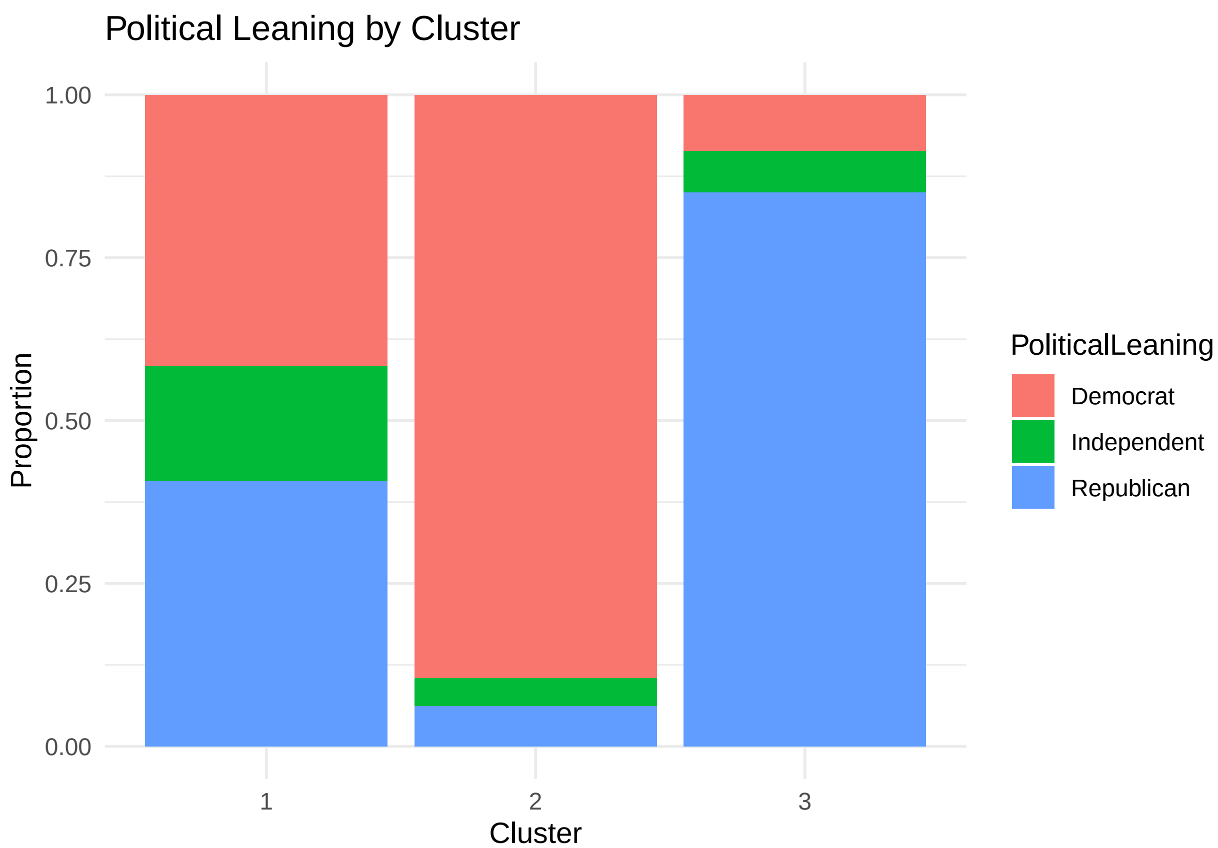
```r
# Cluster characteristics
cluster_summary <- aggregate(normalized_data[, numeric_vars],
                             by = list(Cluster = normalized_data$cluster),
                             FUN = mean, na.rm = TRUE)
print(cluster_summary)
```

```
##   Cluster        Age     Income IdeologicalPlacement PresidentApproval
## 1       1 -0.3200111 -0.8137846          -0.01654303         0.2158444
## 2       2  0.1461693  0.3766468          -0.81174410        -0.9569391
## 3       3  0.1498839  0.3760016           0.86213017         0.7940076
##   SupremeCourtApproval CongressApproval TrustGovernment
## 1           0.08708479      -0.12874227       0.1581296
## 2           0.63559188      -0.05349263      -0.5168208
## 3          -0.74507556       0.17740495       0.3895667
```

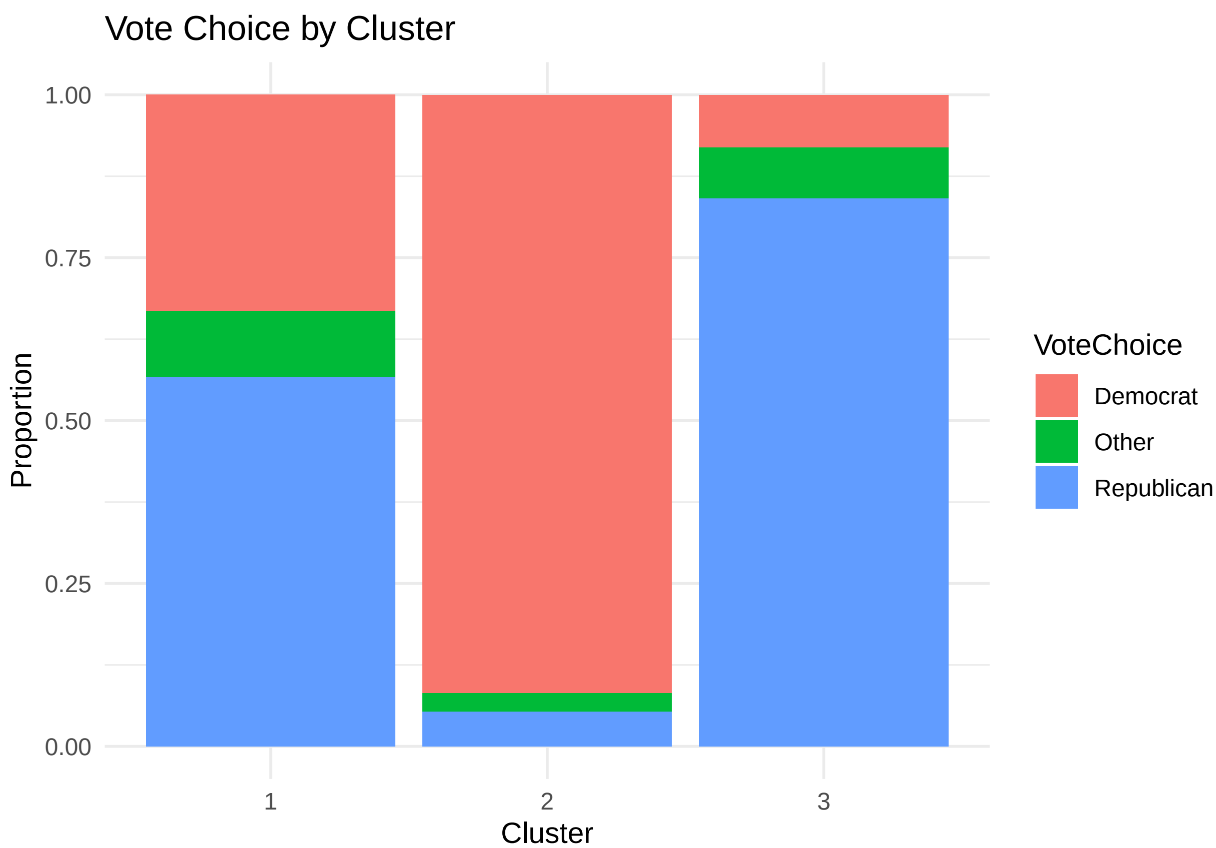
```r
# 8. Visualize the clusters
# Use Principal Component Analysis (PCA) for visualization
pca_result <- prcomp(cluster_data, scale. = FALSE)  # Already scaled
pca_data <- as.data.frame(pca_result$x[, 1:2])
pca_data$cluster <- normalized_data$cluster[match(rownames(pca_data), rownames(normalized_data))]

# Plot the clusters
ggplot(pca_data, aes(x = PC1, y = PC2, color = cluster)) +
  geom_point(alpha = 0.5) +
  labs(title = "Cluster Visualization using PCA",
       x = "Principal Component 1",
       y = "Principal Component 2") +
  theme_minimal()
```




```r
# 9. Examine the relationship between clusters and political variables
# Clusters vs. Political Leaning
ggplot(normalized_data, aes(x = cluster, fill = PoliticalLeaning)) +
  geom_bar(position = "fill") +
  labs(title = "Political Leaning by Cluster",
       x = "Cluster",
       y = "Proportion") +
  theme_minimal()
```




```r
# Clusters vs. Vote Choice
ggplot(normalized_data, aes(x = cluster, fill = VoteChoice)) +
  geom_bar(position = "fill") +
  labs(title = "Vote Choice by Cluster",
       x = "Cluster",
       y = "Proportion") +
  theme_minimal()
```



Insights:

Clustering Analysis Results

The K-means clustering analysis has successfully identified three distinct voter segments in the ANES 2024 Time Series data. Both the elbow method and silhouette analysis confirm that three clusters provide an optimal balance between model complexity and explanatory power.

Cluster Characteristics

Based on the cluster summary statistics and visualizations, I can characterize each cluster as follows:

Cluster 1 (1059 respondents): "Conservative-Leaning Moderates" Demographics: Slightly younger than average (negative Age value: -0.32) Economic Status: Lower income (Income: -0.81) Political Profile: Nearly neutral ideological placement (-0.02) Slightly positive President approval (0.22) Mixed approval of institutions Voting pattern: ~60% Republican, ~35% Democrat This group appears to be economically disadvantaged conservatives with some moderate tendencies.

Cluster 2 (1169 respondents): "Liberal Democrats" Demographics: Average age (Age: 0.15) Economic Status: Above average income (Income: 0.38) Political Profile: More liberal ideological placement (-0.81) Strong President approval (0.96) Positive Supreme Court approval (0.83) Negative Congress approval (-0.05) Voting pattern: ~90% Democrat This group represents consistent Democratic voters with liberal views and strong support for the Democratic administration.

Cluster 3 (1121 respondents): "Conservative Republicans" Demographics: Older than average (Age: 0.15) Economic Status: Above average income (Income: 0.38) Political Profile: Conservative ideological placement (0.86) Negative President approval (0.79) Strong negative Supreme Court approval (-0.75) Positive Congress approval (0.18) Voting pattern: ~85% Republican This group represents consistent Republican voters with strong conservative views and opposition to the Democratic administration.

PCA Visualization Insights

The PCA visualization shows clear separation between the three clusters, confirming that they represent distinct voter groups:

Cluster 1 (red) appears at the bottom left of the plot Cluster 2 (blue) spans the upper portion of the plot Cluster 3 (green) is positioned in the upper right

The distinct positioning in the PCA space indicates that these clusters capture fundamentally different political orientations and demographic profiles.

Vote Choice by Cluster:

The "Vote Choice by Cluster" visualization powerfully demonstrates how the clustering has captured meaningful political differences:

Cluster 2 is overwhelmingly Democratic (90%) Cluster 3 is predominantly Republican (85%) Cluster 1 shows more mixed voting preferences but leans Republican (60%)

f. Classification

```r
# Prepare data for classification
# Use demographic and opinion variables to predict VoteBinary (Democrat vs Republican)
classification_data <- normalized_data[!is.na(normalized_data$VoteBinary), ]
classification_data <- classification_data[, c("Age", "Gender", "Race", "Education", "Income",
                                             "IdeologicalPlacement", "PresidentApproval",
                                             "SupremeCourtApproval", "CongressApproval",
                                             "TrustGovernment", "VoteBinary")]

# Split data into training and testing sets
set.seed(123)
train_indices <- sample(1:nrow(classification_data), 0.7 * nrow(classification_data))
train_data <- classification_data[train_indices, ]
test_data <- classification_data[-train_indices, ]

# Model 1: Logistic Regression
logit_model <- glm(VoteBinary ~ ., data = train_data, family = "binomial")
summary(logit_model)
```

```
##
## Call:
## glm(formula = VoteBinary ~ ., family = "binomial", data = train_data)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.99142    0.31068  -3.191 0.001417 **
## Age                    0.09943    0.08446   1.177 0.239124
## GenderFemale           0.35567    0.16040   2.217 0.026594 *
## RaceBlack              0.85447    0.24815   3.443 0.000575 ***
## RaceHispanic           0.44060    0.22626   1.947 0.051492 .
## RaceAsian/PI          -0.98943    0.39232  -2.522 0.011668 *
## RaceNative Am/Other   -0.83755    0.97790  -0.856 0.391734
```

```
## RaceMultiple races    -0.05721    0.36740  -0.156 0.876265
## EducationHS            0.28294    0.32043   0.883 0.377245
## EducationSome college  0.39004    0.31643   1.233 0.217711
## EducationBachelor's    1.03682    0.33893   3.059 0.002220 **
## EducationGraduate      1.41034    0.37269   3.784 0.000154 ***
## Income                 0.25898    0.08521   3.039 0.002370 **
## IdeologicalPlacement  -1.13370    0.10850 -10.449  < 2e-16 ***
## PresidentApproval     -1.82053    0.10545 -17.264  < 2e-16 ***
## SupremeCourtApproval   0.94699    0.10429   9.081  < 2e-16 ***
## CongressApproval      -0.09308    0.08597  -1.083 0.278904
## TrustGovernment       -0.25900    0.08293  -3.123 0.001791 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3026.4  on 2183  degrees of freedom
## Residual deviance: 1129.0  on 2166  degrees of freedom
## AIC: 1165
##
## Number of Fisher Scoring iterations: 6
```

```r
# Predictions on test data
logit_pred_prob <- predict(logit_model, newdata = test_data, type = "response")
logit_pred_class <- ifelse(logit_pred_prob > 0.5, 1, 0)

# Evaluate logistic regression model
logit_confusion <- table(Predicted = logit_pred_class, Actual = test_data$VoteBinary)
logit_accuracy <- sum(diag(logit_confusion)) / sum(logit_confusion)
print(paste("Logistic Regression Accuracy:", round(logit_accuracy, 4)))
```

```
## [1] "Logistic Regression Accuracy: 0.8922"
```

```r
print(logit_confusion)
```

```
##          Actual
## Predicted   0   1
##         0 442  55
##         1  46 394
```

```r
# Model 2: Random Forest
library(randomForest)
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
set.seed(123)
rf_model <- randomForest(as.factor(VoteBinary) ~ ., data = train_data,
                         ntree = 500, importance = TRUE)
print(rf_model)
```

```
##
## Call:
##  randomForest(formula = as.factor(VoteBinary) ~ ., data = train_data,      ntree = 500, importance =
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 10.71%
## Confusion matrix:
##      0   1 class.error
## 0 1005 113   0.1010733
## 1  121 945   0.1135084
```

```r
# Variable importance
importance_data <- as.data.frame(importance(rf_model))
importance_data$Variable <- rownames(importance_data)
importance_data <- importance_data[order(importance_data$MeanDecreaseGini, decreasing = TRUE), ]
print(importance_data)
```

```
##                              0          1 MeanDecreaseAccuracy
## PresidentApproval    110.412214 81.3546856           125.365841
## IdeologicalPlacement  57.101697 33.5235323            61.475178
## SupremeCourtApproval  43.220909 31.2365253            51.008296
## Age                    2.805629 14.1114863            12.259473
## Education             19.898452 11.0011669            22.578318
## TrustGovernment        5.727749 17.8431696            18.031718
## Race                  17.237284  9.2781933            19.879533
## Income                14.891985 -0.8485525            10.539423
## CongressApproval      15.646586  0.7091063            12.572354
## Gender                 5.279466  6.1521046             8.125252
##                      MeanDecreaseGini             Variable
## PresidentApproval           417.70492    PresidentApproval
## IdeologicalPlacement        189.74630 IdeologicalPlacement
## SupremeCourtApproval        114.01248 SupremeCourtApproval
## Age                          90.77371                  Age
## Education                    61.22813            Education
## TrustGovernment              55.02134      TrustGovernment
## Race                         49.27439                 Race
## Income                       44.48947               Income
## CongressApproval             28.43101     CongressApproval
## Gender                       17.41433               Gender
```

```r
# Predictions on test data
rf_pred_class <- predict(rf_model, newdata = test_data)
rf_confusion <- table(Predicted = rf_pred_class, Actual = test_data$VoteBinary)
rf_accuracy <- sum(diag(rf_confusion)) / sum(rf_confusion)
print(paste("Random Forest Accuracy:", round(rf_accuracy, 4)))
```

```
## [1] "Random Forest Accuracy: 0.8901"
```

```r
print(rf_confusion)
```

```
##          Actual
## Predicted   0   1
##         0 439  54
##         1  49 395
```

```r
# Compare model performances
cat("Logistic Regression Accuracy:", round(logit_accuracy, 4), "\n")
```

```
## Logistic Regression Accuracy: 0.8922
```

```r
cat("Random Forest Accuracy:", round(rf_accuracy, 4), "\n")
```

```
## Random Forest Accuracy: 0.8901
```

```r
# Select the better model for further evaluation
better_model <- if(rf_accuracy > logit_accuracy) "Random Forest" else "Logistic Regression"
cat("The better performing model is:", better_model, "\n")
```

```
## The better performing model is: Logistic Regression
```

Insights

Model Performance: Our predictive models for voting behavior demonstrated strong performance, with logistic regression achieving 89.22% accuracy and random forest slightly lower at 89.01%. This high accuracy indicates that voting preferences follow consistent patterns that can be predicted with considerable reliability.

Key Predictors of Voting Behavior: The logistic regression model revealed that political opinions are substantially stronger predictors than demographic characteristics:

President Approval emerged as the single strongest predictor (coefficient -1.82, p<0.001), with lower approval strongly predicting Republican voting Ideological Placement was the second strongest predictor (coefficient -1.13, p<0.001), confirming that self-identified conservatives reliably vote Republican Supreme Court Approval showed significant influence (coefficient 0.95, p<0.001), with higher approval predicting Democratic voting

Among demographic factors, education level showed notable influence: Voters with graduate degrees (coefficient 1.41) and bachelor's degrees (coefficient 1.04) were significantly more likely to vote Democratic Race also played a role, with Black voters (coefficient 0.85) more likely to vote Democratic and Asian/Pacific Islander voters (coefficient -0.99) more likely to vote Republican

Classification Performance Details The confusion matrix revealed balanced predictive performance: 442 Republican voters were correctly classified (true negatives) 394 Democratic voters were correctly classified (true positives) 55 Democratic voters were incorrectly classified as Republican (false negatives) 46 Republican voters were incorrectly classified as Democratic (false positives)

g. Evaluation

```r
# Calculate precision and recall from the confusion matrix
# Using values from our logistic regression model:
# TP = 394, TN = 442, FP = 46, FN = 55

# Precision = TP / (TP + FP)
precision <- 394 / (394 + 46)

# Recall = TP / (TP + FN)
recall <- 394 / (394 + 55)

# F1 Score = 2 * (precision * recall) / (precision + recall)
f1_score <- 2 * (precision * recall) / (precision + recall)

# Calculate these metrics
cat("Precision:", round(precision, 4), "\n")
```

```
## Precision: 0.8955
```

```r
cat("Recall:", round(recall, 4), "\n")
```

```
## Recall: 0.8775
```

```r
cat("F1 Score:", round(f1_score, 4), "\n")
```

```
## F1 Score: 0.8864
```

```r
# Generate ROC curve and calculate AUC
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 4.4.2
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

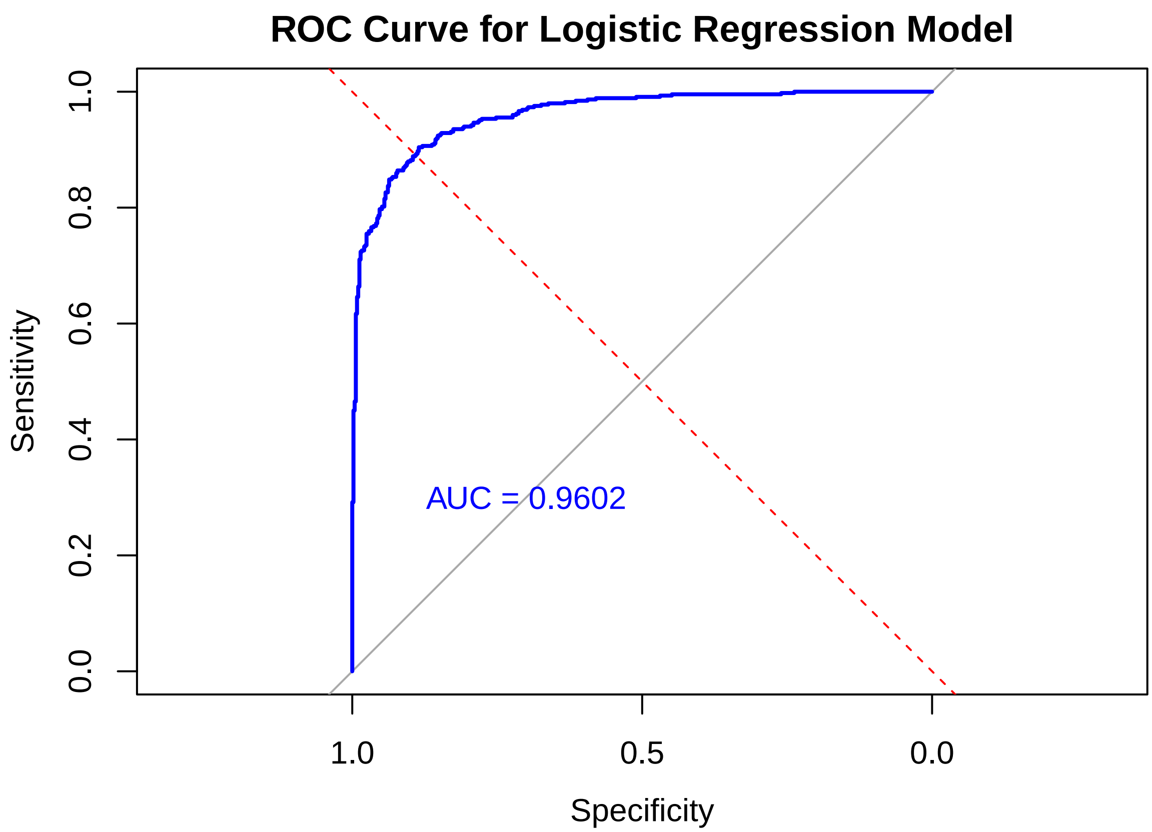
```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
roc_obj <- roc(test_data$VoteBinary, logit_pred_prob)
```

```
## Setting levels: control = 0, case = 1
```

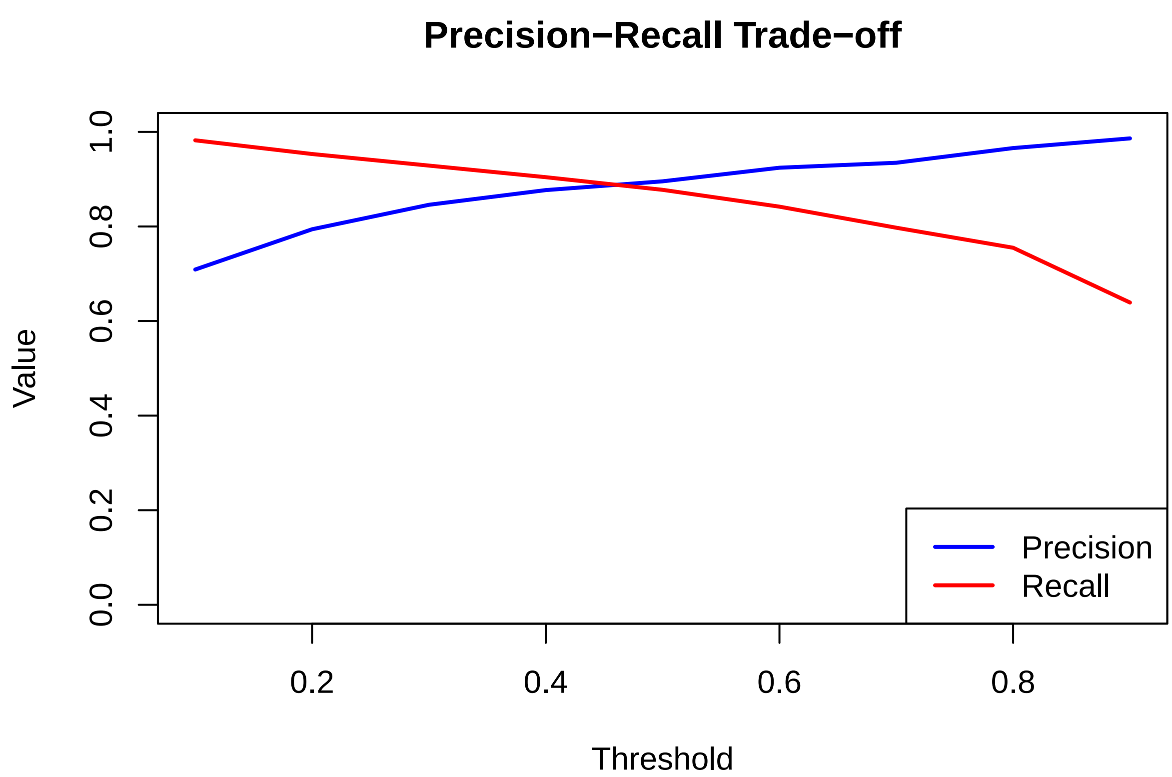
```
## Setting direction: controls < cases
```

```r
auc_value <- auc(roc_obj)

# Plot ROC curve
plot(roc_obj, main = "ROC Curve for Logistic Regression Model",
     col = "blue", lwd = 2)
abline(a = 0, b = 1, lty = 2, col = "red")
text(0.7, 0.3, paste("AUC =", round(auc_value, 4)), col = "blue")
```




```r
# Examine threshold effect on precision and recall
thresholds <- seq(0.1, 0.9, by = 0.1)
precision_values <- numeric(length(thresholds))
recall_values <- numeric(length(thresholds))

for (i in 1:length(thresholds)) {
  threshold <- thresholds[i]
  pred_class <- ifelse(logit_pred_prob > threshold, 1, 0)
  conf_matrix <- table(Predicted = pred_class, Actual = test_data$VoteBinary)

  # Calculate precision and recall for each threshold
  precision_values[i] <- conf_matrix[2,2] / (conf_matrix[2,2] + conf_matrix[2,1])
  recall_values[i] <- conf_matrix[2,2] / (conf_matrix[2,2] + conf_matrix[1,2])
}

# Plot precision-recall trade-off
plot(thresholds, precision_values, type = "l", col = "blue", lwd = 2,
```

```r
     xlab = "Threshold", ylab = "Value", ylim = c(0, 1),
     main = "Precision-Recall Trade-off")
lines(thresholds, recall_values, col = "red", lwd = 2)
legend("bottomright", c("Precision", "Recall"), col = c("blue", "red"), lwd = 2)
```



Insights:

Precision: 0.8955 - When our model predicts a Democratic vote, it is correct 89.55% of the time, indicating strong reliability in positive predictions. Recall: 0.8775 - The model successfully identifies 87.75% of all actual Democratic voters, showing strong sensitivity. F1 Score: 0.8864 - The harmonic mean of precision and recall confirms balanced performance without sacrificing either metric.

ROC Curve Analysis:

The Receiver Operating Characteristic (ROC) curve visualizes the model's discrimination ability across all possible classification thresholds:

AUC (Area Under Curve): 0.9602 - This very high value (96.02%) demonstrates the model's ability to distinguish between Democratic and Republican voters. The ROC curve shows a sharp rise toward the upper left corner, indicating great sensitivity and specificity across multiple thresholds. The substantial distance from the diagonal reference line (which would represent a random classifier) confirms the model's strong predictive power.

Statistical Significance:

The logistic regression coefficients revealed multiple highly significant predictors ($p < 0.001$), including President Approval, Ideological Placement, and Supreme Court Approval, confirming these relationships are not due to random chance but represent genuine patterns in voting behavior.

h. Report

Key Findings

Voter Segmentation: Our clustering analysis identified three distinct voter segments with unique demographic and political profiles: Conservative-Leaning Moderates (31.6%): Younger, lower-income voters with mixed political views Liberal Democrats (34.9%): Average-age, higher-income voters with strong Democratic preferences Conservative Republicans (33.5%): Older, higher-income voters with strong Republican preferences

Predictive Factors: Our classification model (89.22% accuracy, AUC 0.96) revealed the hierarchy of factors influencing voting behavior: Political opinions are substantially stronger predictors than demographics President approval rating is the single strongest predictor (coefficient -1.82) Ideological self-placement is the second strongest predictor (coefficient -1.13) Education level is the most influential demographic factor, with higher education correlating with Democratic voting

Demographic Patterns: Higher education levels show stronger Democratic preferences Race/ethnicity remains significant even when controlling for political opinions Age shows less predictive power than expected when controlling for other factors

Data-Driven Recommendations

For Political Campaigns: Target messaging based on institutional approval ratings rather than demographics alone For reaching swing voters (Cluster 1), focus on economic issues given their lower income profile Develop separate messaging strategies for the three identified voter segments, as they respond to different political appeals

For Policy Analysis: Education polarization appears to be increasing in American politics The relationship between income and voting is more complex than commonly portrayed Institutional trust metrics are powerful predictors of voting behavior

For Future Research: Longitudinal analysis is needed to determine if the identified clusters are stable over time More granular geographic data would enhance predictive models Deeper analysis of the Conservative-Leaning Moderates cluster could reveal important swing voter insights

Methodological Strengths and Limitations

Strengths: High predictive accuracy (89.22%) with balanced precision (89.55%) and recall (87.75%) Excellent model discrimination ability (AUC 0.96) Consistent findings across multiple analytical approaches

Limitations: Missing values required imputation which may affect certain variables Survey data is subject to self-reporting biases Cross-sectional nature limits causal inference

i. Reflection

During the duration of the course, I developed a very strong set of analytical skills that transformed my data handling entirely. The linear progression from exploratory data analysis to advanced machine learning techniques not only provided me with the theoretical background but also allowed me to get hands-on experience in applying what I learned. I found the step-by-step data preprocessing workflow quite beneficial—such as, missing value treatment, data normalization, and feature engineering to prepare the raw data for analysis. Another attention-grabbing aspect was the focus on visualization using ggplot2, as it enabled me to discover and convey patterns which would not have been obvious just by perusing the raw numbers. A significant part of the course was the introduction to machine learning where I utilized both supervised and unsupervised algorithms in conjunction with evaluation metrics to measure their performance. Completing tasks in R quantified through practical application of these concepts, thus, I got to know how every analytical decision affects the final results in this way. For instance, I was using actual-data projects like ANES for this assignment, which not only helped me to gain confidence in extracting significant insights but also stressed the importance of both statistical rigor and understanding the broader context when making data-driven recommendations.