

Appendix 2: Phase 1 Pre-Processing

Jessala Grijalva

September 2023

Introduction

This R Markdown document provides a comprehensive script for data pre-processing, including data cleaning, missing value imputation, and feature engineering, for the LNS 2006 dataset.

SOMETHING

Load Libraries

```
library(pacman)
p_load(dplyr, tidyr, mice, forcats, factoextra, skimr, stringr, readr)
```

Load Data

```
# Full path directly
load("~/Desktop/Dissertation Analysis//Dissertation Analysis/Data/Raw/20862-0003-Data.rda")
```

Data Pre-Processing

Subsetting and Renaming Columns

```
# Subset and filter
# Rename columns
lns_subset <- da20862.0003 %>%
  dplyr::filter(BORNUS %in% c("(1) Mainland US", "(2) Puerto Rico") | NATUSCIT == "(1) YES") %>%
  dplyr::select(AGE, SEX, REDUC, HHINC, RELIGION, ANCESTRY, BORNUS, NATUSCIT, BIRTHPLC,
               PARBORN, GRANBORN, AMERICAN, RGIDENT, LAIDENT, KEEPSPAN, LEARNENG,
               BLEND, DISTINCT, IMPPOLICY, DREAMACT, IMMVIEW, IDEOLOGY, RACE1,
               FEELPART, PARTYID, PRIMEID, VOTEPRES, NONVPRES, SAYSO, GOVTRUST,
               INCSUPP, HEALTH, WT_NATION_REV) %>%
  dplyr::rename(WEIGHT = WT_NATION_REV)
```

Initial Data Structure

```
head(lns_subset, 5)
```

```
##   AGE      SEX      REDUC      HHINC
## 1  57 (2) FEMALE (1) Eighth grade or below (3) $25,000-34,999
## 2  69  (1) MALE  (4) High school graduate (4) $35,000-44,999
## 3  20  (1) MALE      (5) Some college (2) $15,000-24,999
## 4  57 (2) FEMALE      (2) Some high school (1) BELOW $15,000K
```

5 35 (2) FEMALE (1) Eighth grade or below (4) \$35,000-44,999

RELIGION ANCESTRY BORNUS NATUSCIT

1 (1) Catholic (12) Mexico (3) Some other country (1) YES

2 (1) Catholic (18) Spain (1) Mainland US <NA>

3 (1) Catholic (12) Mexico (3) Some other country (1) YES

4 (1) Catholic (17) Puerto Rico (2) Puerto Rico <NA>

5 (0) Jehovah's Witness (12) Mexico (3) Some other country (1) YES

BIRTHPLC PARBORN GRANBORN

1 (12) Mexico (3) Neither parent born in the U.S. (4) All

2 <NA> (2) Both parents born in the U.S. (5) Don't Know

3 (12) Mexico (3) Neither parent born in the U.S. (0) None

4 <NA> (3) Neither parent born in the U.S. (4) All

5 (12) Mexico (2) Both parents born in the U.S. (2) Two

AMERICAN RGIDENT LAIDENT

1 (2) Not very strongly (4) Very strongly (4) Very strongly

2 (3) Somewhat strongly (4) Very strongly (4) Very strongly

3 (2) Not very strongly (2) Not very strongly (4) Very strongly

4 (3) Somewhat strongly (3) Somewhat strongly (4) Very strongly

5 (3) Somewhat strongly (3) Somewhat strongly (3) Somewhat strongly

KEEPSAN LEARNENG BLEND

1 (4) Very Important (4) Very Important (2) Somewhat important

2 (4) Very Important (4) Very Important (3) Very important

3 (4) Very Important (4) Very Important (3) Very important

4 (3) Somewhat important (3) Somewhat important (3) Very important

5 (4) Very Important (4) Very Important (3) Very important

DISTINCT

1 (2) Somewhat important

2 (3) Very important

3 (3) Very important

4 (3) Very important

5 (3) Very important

IMPPOLICY

1 (1) Immediate legalization of current undocumented immigrants

2 (1) Immediate legalization of current undocumented immigrants

3 (1) Immediate legalization of current undocumented immigrants

4 (3) A guest worker program that permits immigrants to be in t...

5 (1) Immediate legalization of current undocumented immigrants

DREAMACT

1 (1) Strongly Oppose

2 (1) Strongly Oppose

3 (1) Strongly Oppose

4 (2) Oppose

5 (2) Oppose

IMMVIEW

1 (1) Immigrants today strengthen our country because of their hard work and talents

2 (1) Immigrants today strengthen our country because of their hard work and talents

3 (1) Immigrants today strengthen our country because of their hard work and talents

4 (1) Immigrants today strengthen our country because of their hard work and talents

5 (1) Immigrants today strengthen our country because of their hard work and talents

IDEOLOGY RACE1

1 (4) Don't think of self in these terms (6) Some other race (specify)

2 (1) Conservative (6) Some other race (specify)

3 (3) Middle of the Road (6) Some other race (specify)

4 (4) Don't think of self in these terms (1) White

```
## 5 (5) Don't know (1) White
## FEELPART PARTYID
## 1 (5) My feelings have not changed (1) Democrat
## 2 (3) I feel much closer to the Democrats than I used to (1) Democrat
## 3 (5) My feelings have not changed (4) Don't care
## 4 (5) My feelings have not changed (1) Democrat
## 5 (6) Don't Know/Other (4) Don't care
## PRIMEID VOTEPRES NONVPRES SAYSO
## 1 (4) None of the above <NA> <NA> (1) Strongly disagree
## 2 (1) ANSWERFROM(AQS4) (2) John Kerry <NA> (4) Strongly agree
## 3 (1) ANSWERFROM(AQS4) <NA> (2) John Kerry (1) Strongly disagree
## 4 (2) ANSWERFROM(AQB4B) <NA> <NA> (4) Strongly agree
## 5 (2) ANSWERFROM(AQB4B) <NA> (2) John Kerry (1) Strongly disagree
## GOVTRUST INCSUPP HEALTH WEIGHT
## 1 (2) Some of the time (1) Strongly Oppose (3) Support 1.16
## 2 (3) Most of the time (3) Support (4) Strongly Support 0.53
## 3 (2) Some of the time (4) Strongly Support (2) Oppose 2.77
## 4 (2) Some of the time (4) Strongly Support (3) Support 0.74
## 5 (2) Some of the time (4) Strongly Support (4) Strongly Support 0.55
```

Variable Creation

Control and Descriptive Variables

```
# Combine NATUSCIT and BIRTHPLC into BIRTH_ORIGIN
lms_subset <- lms_subset %>%
  mutate(BIRTH_ORIGIN = case_when(
    !is.na(NATUSCIT) ~ as.character(NATUSCIT),
    !is.na(BIRTHPLC) ~ as.character(BIRTHPLC),
    TRUE ~ "Missing"
  )
)

# Create NATIONALITY Variable from BORNUS and BIRTHPLC
lms_subset <- lms_subset %>%
  mutate(
    NATIONALITY = case_when(
      BORNUS %in% c("(1) Mainland US", "(2) Puerto Rico") ~ "US Born",
      BIRTHPLC == "(01) Mexico" ~ "Mexico",
      BIRTHPLC == "(02) Puerto Rico" ~ "Puerto Rico",
      BIRTHPLC == "(03) Cuba" ~ "Cuba",
      BIRTHPLC %in% c("(04) Central American Countries") ~ "Central American",
      is.na(BORNUS) & is.na(BIRTHPLC) ~ "Missing",
      TRUE ~ "Other Latin Countries"
    )
  )

# Recode REDUC to EDUCATION
lms_subset <- lms_subset %>%
  mutate(
    EDUCATION = case_when(
      !is.na(REDUC) ~ as.character(REDUC), # Keep original value if REDUC is not NA
      TRUE ~ "Missing" # Assign "Missing" if REDUC is NA
    )
  )
```

```

)

# Modify the dataset to create the GEN_STATUS variable
lms_subset <- lms_subset %>%
  mutate(
    GEN_STATUS = case_when(
      BORNUS == "(3) Some other country" ~ "First Generation",
      BORNUS %in% c("(1) Mainland US", "(2) Puerto Rico") & PARBORN != "(2) Both parents born in the U.S." ~ "Second Generation",
      BORNUS %in% c("(1) Mainland US", "(2) Puerto Rico") & PARBORN == "(2) Both parents born in the U.S." ~ "Third Generation Plus",
      TRUE ~ "Other"
    )
  )

# Drop rows where GEN_STATUS is "Other"
lms_subset <- lms_subset %>%
  filter(GEN_STATUS != "Other")

# Modify the dataset to create the GEN_STATUS_NUM variable
lms_subset <- lms_subset %>%
  mutate(
    GEN_STATUS_num = as.numeric(case_when(
      GEN_STATUS == "First Generation" ~ 1,
      GEN_STATUS == "Second Generation" ~ 2,
      GEN_STATUS == "Third Generation Plus" ~ 3,
      TRUE ~ NA_real_ # for missing or other categories, if any
    ))
  )

# Recode RELIGION into Christians, Other, Non-Secular categories
lms_subset <- lms_subset %>%
  mutate(
    RELIGION = case_when(
      RELIGION %in% c("(1) Catholic", "(0) Jehovah's Witness", "(3) Southern Baptist",
                     "(5) Other Protestant", "(4) Pentecostal", "(2) Assemblies of God") ~ "Christian",
      RELIGION == "(8) Don't identify with any religious denomination" ~ "Non-Secular",
      RELIGION %in% c("(9) Other", "(6) Mormon", "(7) Jewish") ~ "Other",
      TRUE ~ as.character(RELIGION) # Keep original value if not recoded
    )
  )

```

Feature Engineering for Independent Variables

```

# Convert AMERICAN, LEARNENG, and BLEND to numeric
# Convert 'AMERICAN' to numerical
lms_subset$AMERICAN_num <- as.numeric(factor(lms_subset$AMERICAN,
      levels = c("(1) Not at all",
                  "(2) Not very strongly",
                  "(3) Somewhat strongly",
                  "(4) Very strongly")))

# Convert 'LEARNENG' to numerical
lms_subset$LEARNENG_num <- as.numeric(factor(lms_subset$LEARNENG,
      levels = c("(1) Not at all important",

```

```

        "(2) Not very important",
        "(3) Somewhat important",
        "(4) Very Important")))

# Convert 'BLEND' to numerical
lms_subset$BLEND_num <- as.numeric(factor(lms_subset$BLEND,
    levels = c("(1) Not at all important",
        "(2) Somewhat important",
        "(3) Very important")))

# Check if the conversion has worked as expected by looking at the first few rows
head(lms_subset[, c("AMERICAN", "AMERICAN_num", "LEARNENG", "LEARNENG_num", "BLEND", "BLEND_num")])

##          AMERICAN AMERICAN_num          LEARNENG LEARNENG_num
## 1 (2) Not very strongly          2          (4) Very Important          4
## 2 (3) Somewhat strongly          3          (4) Very Important          4
## 3 (2) Not very strongly          2          (4) Very Important          4
## 4 (3) Somewhat strongly          3 (3) Somewhat important          3
## 5 (3) Somewhat strongly          3          (4) Very Important          4
## 6   (4) Very strongly          4          (4) Very Important          4
##          BLEND BLEND_num
## 1 (2) Somewhat important          2
## 2   (3) Very important          3
## 3   (3) Very important          3
## 4   (3) Very important          3
## 5   (3) Very important          3
## 6 (2) Somewhat important          2

# Map the original RGIDENT and LAIDENT columns to numerical values
# Using factor levels for the mapping
lms_subset$RGIDENT_num <- as.numeric(factor(lms_subset$RGIDENT,
    levels = c("(1) Not at all", "(2) Not very strongly",
        "(3) Somewhat strongly", "(4) Very strongly")))

lms_subset$LAIDENT_num <- as.numeric(factor(lms_subset$LAIDENT,
    levels = c("(1) Not at all", "(2) Not very strongly",
        "(3) Somewhat strongly", "(4) Very strongly")))

# Calculate the row-wise mean for CULTURAL_IDENTITY
lms_subset$CULTURAL_IDENTITY <- rowMeans(cbind(lms_subset$RGIDENT_num, lms_subset$LAIDENT_num), na.rm =

# Checking the first few rows to see if CULTURAL_IDENTITY has been created accurately
head(lms_subset[, c("RGIDENT_num", "LAIDENT_num", "CULTURAL_IDENTITY")])

##  RGIDENT_num LAIDENT_num CULTURAL_IDENTITY
## 1           4           4              4.0
## 2           4           4              4.0
## 3           2           4              3.0
## 4           3           4              3.5
## 5           3           3              3.0
## 6           4           3              3.5

# Convert DISTINCT to numeric
lms_subset$DISTINCT_num <- as.numeric(lms_subset$DISTINCT)

```

```
# Convert KEEPSPAN to numeric
lms_subset$KEEPSPAN_num <- as.numeric(lms_subset$KEEPSPAN)

# Verify the first few rows
head(lms_subset)
```

```
##      AGE      SEX      REDUC      HHINC  RELIGION
## 1  57 (2) FEMALE (1) Eighth grade or below (3) $25,000-34,999 Christian
## 2  69 (1) MALE (4) High school graduate (4) $35,000-44,999 Christian
## 3  20 (1) MALE (5) Some college (2) $15,000-24,999 Christian
## 4  57 (2) FEMALE (2) Some high school (1) BELOW $15,000K Christian
## 5  35 (2) FEMALE (1) Eighth grade or below (4) $35,000-44,999 Christian
## 6  29 (2) FEMALE (5) Some college <NA> Other
##      ANCESTRY      BORNUS NATUSCIT      BIRTHPLC
## 1 (12) Mexico (3) Some other country (1) YES (12) Mexico
## 2 (18) Spain (1) Mainland US <NA> <NA>
## 3 (12) Mexico (3) Some other country (1) YES (12) Mexico
## 4 (17) Puerto Rico (2) Puerto Rico <NA> <NA>
## 5 (12) Mexico (3) Some other country (1) YES (12) Mexico
## 6 (12) Mexico (1) Mainland US <NA> <NA>
##      PARBORN      GRANBORN      AMERICAN
## 1 (3) Neither parent born in the U.S. (4) All (2) Not very strongly
## 2 (2) Both parents born in the U.S. (5) Don't Know (3) Somewhat strongly
## 3 (3) Neither parent born in the U.S. (0) None (2) Not very strongly
## 4 (3) Neither parent born in the U.S. (4) All (3) Somewhat strongly
## 5 (2) Both parents born in the U.S. (2) Two (3) Somewhat strongly
## 6 (2) Both parents born in the U.S. (4) All (4) Very strongly
##      RGIDENT      LAIDENT      KEEPSPAN
## 1 (4) Very strongly (4) Very strongly (4) Very Important
## 2 (4) Very strongly (4) Very strongly (4) Very Important
## 3 (2) Not very strongly (4) Very strongly (4) Very Important
## 4 (3) Somewhat strongly (4) Very strongly (3) Somewhat important
## 5 (3) Somewhat strongly (3) Somewhat strongly (4) Very Important
## 6 (4) Very strongly (3) Somewhat strongly (4) Very Important
##      LEARNENG      BLEND      DISTINCT
## 1 (4) Very Important (2) Somewhat important (2) Somewhat important
## 2 (4) Very Important (3) Very important (3) Very important
## 3 (4) Very Important (3) Very important (3) Very important
## 4 (3) Somewhat important (3) Very important (3) Very important
## 5 (4) Very Important (3) Very important (3) Very important
## 6 (4) Very Important (2) Somewhat important (3) Very important
##      IMPPOLICY
## 1 (1) Immediate legalization of current undocumented immigrants
## 2 (1) Immediate legalization of current undocumented immigrants
## 3 (1) Immediate legalization of current undocumented immigrants
## 4 (3) A guest worker program that permits immigrants to be in t...
## 5 (1) Immediate legalization of current undocumented immigrants
## 6 (3) A guest worker program that permits immigrants to be in t...
##      DREAMACT
## 1 (1) Strongly Oppose
## 2 (1) Strongly Oppose
## 3 (1) Strongly Oppose
## 4 (2) Oppose
## 5 (2) Oppose
```

```

## 6 (1) Strongly Oppose
##
## 1 (1) Immigrants today strengthen our country because of their hard work and talents
## 2 (1) Immigrants today strengthen our country because of their hard work and talents
## 3 (1) Immigrants today strengthen our country because of their hard work and talents
## 4 (1) Immigrants today strengthen our country because of their hard work and talents
## 5 (1) Immigrants today strengthen our country because of their hard work and talents
## 6 (1) Immigrants today strengthen our country because of their hard work and talents
##
##                                IDEOLOGY                                RACE1
## 1 (4) Don't think of self in these terms (6) Some other race (specify)
## 2                                (1) Conservative (6) Some other race (specify)
## 3                                (3) Middle of the Road (6) Some other race (specify)
## 4 (4) Don't think of self in these terms                                (1) White
## 5                                (5) Don't know                                (1) White
## 6 (4) Don't think of self in these terms                                (1) White
##
##                                FEELPART                                PARTYID
## 1                                (5) My feelings have not changed (1) Democrat
## 2 (3) I feel much closer to the Democrats than I used to (1) Democrat
## 3                                (5) My feelings have not changed (4) Don't care
## 4                                (5) My feelings have not changed (1) Democrat
## 5                                (6) Don't Know/Other (4) Don't care
## 6                                (5) My feelings have not changed (1) Democrat
##
##                                PRIMEID                                VOTEPRES                                NONVPRES                                SAYSO
## 1 (4) None of the above                                <NA>                                <NA> (1) Strongly disagree
## 2 (1) ANSWERFROM(AQS4) (2) John Kerry                                <NA> (4) Strongly agree
## 3 (1) ANSWERFROM(AQS4)                                <NA> (2) John Kerry (1) Strongly disagree
## 4 (2) ANSWERFROM(AQB4B)                                <NA>                                <NA> (4) Strongly agree
## 5 (2) ANSWERFROM(AQB4B)                                <NA> (2) John Kerry (1) Strongly disagree
## 6 (1) ANSWERFROM(AQS4)                                <NA>                                <NA> (4) Strongly agree
##
##                                GOVTRUST                                INCSUPP                                HEALTH WEIGHT
## 1 (2) Some of the time (1) Strongly Oppose (3) Support 1.16
## 2 (3) Most of the time (3) Support (4) Strongly Support 0.53
## 3 (2) Some of the time (4) Strongly Support (2) Oppose 2.77
## 4 (2) Some of the time (4) Strongly Support (3) Support 0.74
## 5 (2) Some of the time (4) Strongly Support (4) Strongly Support 0.55
## 6 (1) Never (3) Support (3) Support 1.14
##
## BIRTH_ORIGIN                                NATIONALITY                                EDUCATION
## 1 (1) YES Other Latin Countries (1) Eighth grade or below
## 2 Missing US Born (4) High school graduate
## 3 (1) YES Other Latin Countries (5) Some college
## 4 Missing US Born (2) Some high school
## 5 (1) YES Other Latin Countries (1) Eighth grade or below
## 6 Missing US Born (5) Some college
##
##                                GEN_STATUS GEN_STATUS_num AMERICAN_num LEARNENG_num BLEND_num
## 1 First Generation 1 2 4 2
## 2 Third Generation Plus 3 3 4 3
## 3 First Generation 1 2 4 3
## 4 Second Generation 2 3 3 3
## 5 First Generation 1 3 4 3
## 6 Third Generation Plus 3 4 4 2
##
## RGIDENT_num LAIDENT_num CULTURAL_IDENTITY DISTINCT_num KEEPSPAN_num
## 1 4 4 4.0 2 4
## 2 4 4 4.0 3 4
## 3 2 4 3.0 3 4

```

## 4	3	4	3.5	3	3
## 5	3	3	3.0	3	4
## 6	4	3	3.5	3	4

Feature Engineering for Dependent Variables

```
# Recode the 'IDEOLOGY' variable and convert it to numerical
lms_subset <- lms_subset %>%
  mutate(IDEOLOGY = case_when(
    IDEOLOGY == "(1) Conservative" ~ 1,
    IDEOLOGY == "(3) Middle of the Road" ~ 2,
    IDEOLOGY == "(4) Don't think of self in these terms" ~ 2,
    IDEOLOGY == "(2) Liberal" ~ 3,
    IDEOLOGY == "(5) Don't know" ~ 2
  ))

# Convert the recoded 'IDEOLOGY' variable to numeric (although it should already be numeric)
lms_subset$IDEOLOGY_num <- as.numeric(lms_subset$IDEOLOGY)

# Recode the 'FEELPART' variable based on the coding scheme and convert to numeric
lms_subset <- lms_subset %>%
  mutate(FEELPART = case_when(
    FEELPART == "(5) My feelings have not changed" ~ 3,
    FEELPART == "(3) I feel much closer to the Democrats than I used to" ~ 5,
    FEELPART == "(6) Don't Know/Other" ~ 3,
    FEELPART == "(2) I feel somewhat closer to the Republicans than I used to" ~ 2,
    FEELPART == "(4) I feel somewhat closer to the Democrats than I used to" ~ 4,
    FEELPART == "(1) I feel much closer to the Republicans than I used to" ~ 1,
    TRUE ~ NA_real_ # Set to NA for any other unexpected categories
  ))

# Convert the recoded 'FEELPART' variable to numeric
lms_subset$FEELPART_num <- as.numeric(lms_subset$FEELPART)

# Reverse code the 'IMMVIEW' variable and convert to numeric
lms_subset <- lms_subset %>%
  mutate(IMMVIEW = case_when(
    IMMVIEW == "(1) Immigrants today strengthen our country because of their hard work and talents" ~ 2
    IMMVIEW == "(2) Or immigrants today are a burden on our country because they take our jobs, housing
  ))

# Convert the reversed 'IMMVIEW' variable to numeric
lms_subset$IMMVIEW_num <- as.numeric(lms_subset$IMMVIEW)

# Recode the 'PARTYID' variable based on the new coding scheme and convert to numeric
lms_subset <- lms_subset %>%
  mutate(PARTYID = case_when(
    PARTYID == "(1) Democrat" ~ 3,
    PARTYID == "(2) Republican" ~ 1,
    PARTYID == "(3) Independent" ~ 2,
    PARTYID == "(4) Don't care" ~ 2,
    PARTYID == "(5) Don't know/other party" ~ 2,
  ))
```



```

# Convert the recoded 'PARTYID' variable to numeric
lns_subset$PARTYID_num <- as.numeric(lns_subset$PARTYID)

# Recode 'IMMPOLICY' based on the new coding scheme and partial string matching
lns_subset <- lns_subset %>%
  mutate(IMMPOLICY = case_when(
    str_detect(IMMPOLICY, "Immediate legalization of current undocumented immigrants") ~ 5,
    str_detect(IMMPOLICY, "A guest worker program that permits immigrants to be in t") ~ 4,
    str_detect(IMMPOLICY, "None of these") ~ 3,
    str_detect(IMMPOLICY, "A guest worker program leading to legalization eventually") ~ 2,
    str_detect(IMMPOLICY, "An effort to seal or close off the border to stop illegal") ~ 1,
    TRUE ~ NA_real_ # Set to NA for any other unexpected categories
  ))

# Convert the recoded 'IMMPOLICY' variable to numeric
lns_subset$IMMPOLICY_num <- as.numeric(lns_subset$IMMPOLICY)

# Convert the recoded 'IMMPOLICY' variable to numeric
lns_subset$GOVTRUST_num <- as.numeric(lns_subset$GOVTRUST)

# First, recode VOTEPRES and NONVPRES according to the new coding scheme
lns_subset <- lns_subset %>%
  mutate(
    RECODED_VOTEPRES = case_when(
      VOTEPRES == "(1) George W. Bush" ~ 1,
      VOTEPRES == "(2) John Kerry" ~ 3,
      VOTEPRES == "(3) Ralph Nader" ~ 2,
      VOTEPRES == "(4) Other" ~ 2,
      TRUE ~ NA_real_
    ),
    RECODED_NONVPRES = case_when(
      NONVPRES == "(1) George W. Bush" ~ 1,
      NONVPRES == "(2) John Kerry" ~ 3,
      NONVPRES == "(3) Ralph Nader" ~ 2,
      NONVPRES == "(4) Other" ~ 2,
      TRUE ~ NA_real_
    )
  )

# Now, combine RECODED_VOTEPRES and RECODED_NONVPRES into VOTE_PREF
lns_subset <- lns_subset %>%
  mutate(VOTE_PREF = ifelse(!is.na(RECODED_VOTEPRES), RECODED_VOTEPRES, RECODED_NONVPRES))

# Convert the combined VOTE_PREF variable to numeric
lns_subset$VOTE_PREF_num <- as.numeric(lns_subset$VOTE_PREF)

# Reverse code DREAMACT and convert to numeric
lns_subset$DREAMACT_num <- as.numeric(factor(lns_subset$DREAMACT,
  levels = c("(4) Strongly Support", "(3) Support",
    "(2) Oppose", "(1) Strongly Oppose")))

# Reverse code SAYSO and convert to numeric
lns_subset$SAYSO_num <- as.numeric(factor(lns_subset$SAYSO,

```

```

                                levels = c("(4) Strongly agree", "(3) Somewhat agree",
                                              "(2) Somewhat disagree", "(1) Strongly disagree"))

# Checking the first few rows to see if the reverse coded variables are accurate
head(lns_subset[, c("DREAMACT", "DREAMACT_num", "SAYSO", "SAYSO_num")])

##           DREAMACT DREAMACT_num           SAYSO SAYSO_num
## 1 (1) Strongly Oppose           4 (1) Strongly disagree       4
## 2 (1) Strongly Oppose           4   (4) Strongly agree       1
## 3 (1) Strongly Oppose           4 (1) Strongly disagree       4
## 4           (2) Oppose           3   (4) Strongly agree       1
## 5           (2) Oppose           3 (1) Strongly disagree       4
## 6 (1) Strongly Oppose           4   (4) Strongly agree       1

# Convert INCSUPP to numeric
lns_subset$INCSUPP_num <- as.numeric(factor(lns_subset$INCSUPP,
                                           levels = c("(1) Strongly Oppose", "(2) Oppose",
                                                         "(3) Support", "(4) Strongly Support"))))

# Convert HEALTH to numeric
lns_subset$HEALTH_num <- as.numeric(factor(lns_subset$HEALTH,
                                           levels = c("(1) Strongly Oppose", "(2) Oppose",
                                                         "(3) Support", "(4) Strongly Support"))))

# Checking the first few rows to see if the numeric variables are accurate
head(lns_subset[, c("INCSUPP", "INCSUPP_num", "HEALTH", "HEALTH_num")])

```

```

##           INCSUPP INCSUPP_num           HEALTH HEALTH_num
## 1 (1) Strongly Oppose           1           (3) Support       3
## 2           (3) Support           3 (4) Strongly Support       4
## 3 (4) Strongly Support           4           (2) Oppose       2
## 4 (4) Strongly Support           4           (3) Support       3
## 5 (4) Strongly Support           4 (4) Strongly Support       4
## 6           (3) Support           3           (3) Support       3

```

Data Cleaning

Dropping Unnecessary Variables

```

# Drop variables that are not of interest
lns_subset <- lns_subset %>%
  dplyr::select(CULTURAL_IDENTITY, AMERICAN_num, LEARNENG_num, DISTINCT_num,
                KEEPSPAN_num, BLEND_num, VOTE_PREF_num, IDEOLOGY_num, FEELPART_num,
                PARTYID_num, SAYSO_num, GOVTRUST_num, INCSUPP_num, HEALTH_num, DREAMACT_num,
                IMMVIEW_num, IMPPOLICY_num,
                EDUCATION, NATIONALITY, BIRTH_ORIGIN, GEN_STATUS_num, AGE, SEX,
                RACE1, HHINC, RELIGION, WEIGHT)

```

Missing Data Handling

Missing Data Imputation

```

vars_to_impute <- c("CULTURAL_IDENTITY", "AMERICAN_num", "DISTINCT_num", "BLEND_num",
                    "VOTE_PREF_num", "SAYSO_num", "INCSUPP_num", "HEALTH_num",
                    "DREAMACT_num", "AGE", "RACE1", "HHINC")

```

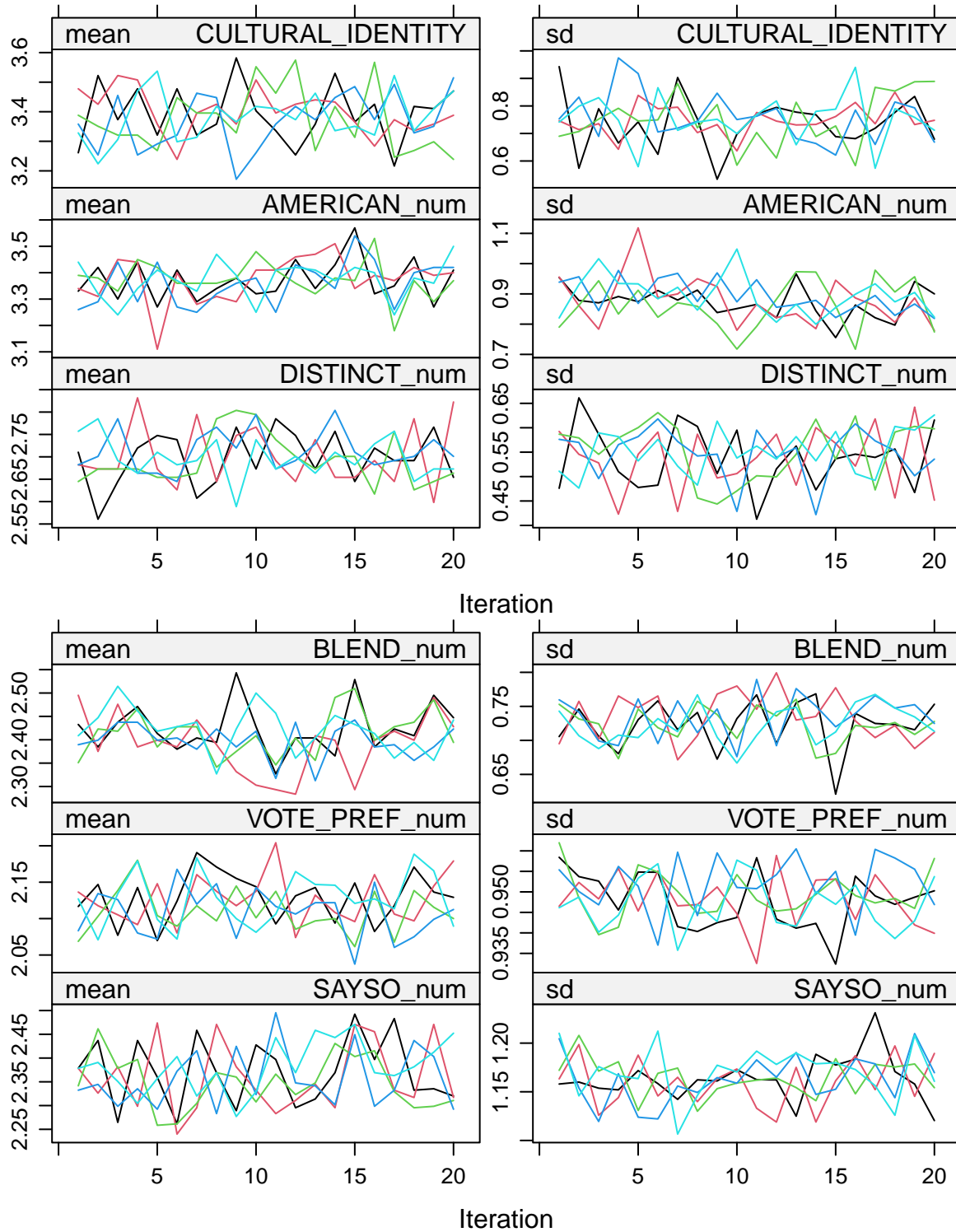
```
# Perform multiple imputation
imputed_data <- mice(lns_subset[, vars_to_impute], m = 5, maxit = 20, method = 'pmm', seed = 500)
```

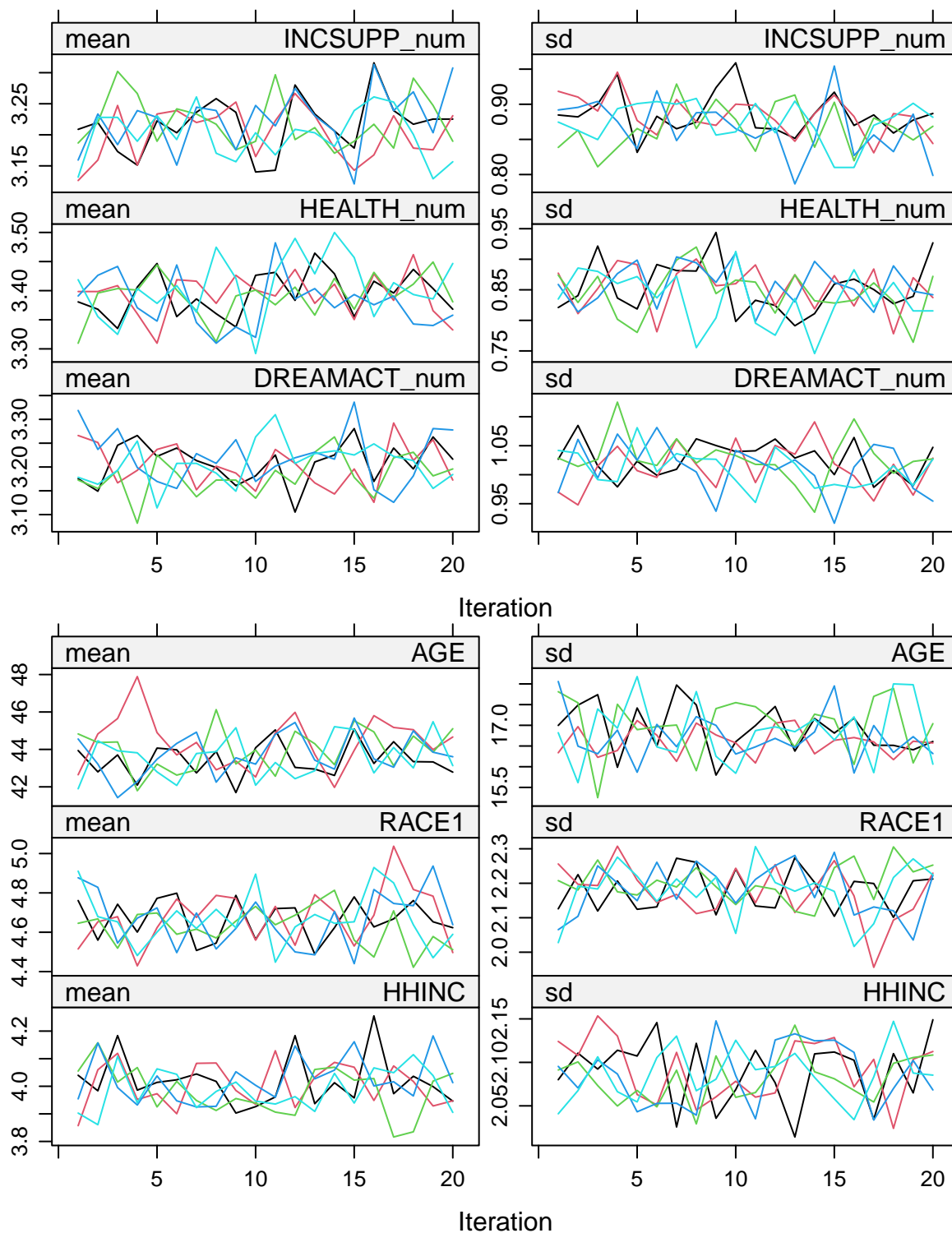
[illegible]

[illegible]

Convergence Diagnostics for Imputed Models

```
# Diagnostic plots to check convergence
plot(imputed_data)
```





```
# Obtain the first imputed dataset
single_completed_data <- complete(imputed_data, 1)

# Check dimensions to ensure they match
print(dim(lms_subset))
```

```
## [1] 4785 27
```

```

print(dim(single_completed_data))

## [1] 4785    12

# Replace the missing values in the original dataset
lms_subset[vars_to_impute] <- single_completed_data[vars_to_impute]

# Checking Convergence for the first imputation model
convergence_summary_1 <- summary(imputed_data)

## Class: mids
## Number of multiple imputations: 5
## Imputation methods:
## CULTURAL_IDENTITY    AMERICAN_num    DISTINCT_num    BLEND_num
##          "pmm"          "pmm"          "pmm"          "pmm"
##      VOTE_PREF_num    SAYSO_num    INCSUPP_num    HEALTH_num
##          "pmm"          "pmm"          "pmm"          "pmm"
##      DREAMACT_num    AGE    RACE1    HHINC
##          "pmm"          "pmm"          "pmm"          "pmm"
## PredictorMatrix:
##          CULTURAL_IDENTITY AMERICAN_num DISTINCT_num BLEND_num
## CULTURAL_IDENTITY          0          1          1          1
## AMERICAN_num          1          0          1          1
## DISTINCT_num          1          1          0          1
## BLEND_num          1          1          1          0
## VOTE_PREF_num          1          1          1          1
## SAYSO_num          1          1          1          1
##          VOTE_PREF_num SAYSO_num INCSUPP_num HEALTH_num DREAMACT_num
## CULTURAL_IDENTITY          1          1          1          1          1
## AMERICAN_num          1          1          1          1          1
## DISTINCT_num          1          1          1          1          1
## BLEND_num          1          1          1          1          1
## VOTE_PREF_num          0          1          1          1          1
## SAYSO_num          1          0          1          1          1
##          AGE RACE1 HHINC
## CULTURAL_IDENTITY  1    1    1
## AMERICAN_num      1    1    1
## DISTINCT_num      1    1    1
## BLEND_num         1    1    1
## VOTE_PREF_num     1    1    1
## SAYSO_num         1    1    1

# Display the summaries
print("Convergence Summary for first imputation model:")

## [1] "Convergence Summary for first imputation model:"
print(convergence_summary_1)

## Class: mids
## Number of multiple imputations: 5
## Imputation methods:
## CULTURAL_IDENTITY    AMERICAN_num    DISTINCT_num    BLEND_num
##          "pmm"          "pmm"          "pmm"          "pmm"
##      VOTE_PREF_num    SAYSO_num    INCSUPP_num    HEALTH_num
##          "pmm"          "pmm"          "pmm"          "pmm"

```



```
##      DREAMACT_num      AGE      RACE1      HHINC
##      "pmm"      "pmm"      "pmm"      "pmm"
## PredictorMatrix:
##      CULTURAL_IDENTITY AMERICAN_num DISTINCT_num BLEND_num
## CULTURAL_IDENTITY      0      1      1      1
## AMERICAN_num      1      0      1      1
## DISTINCT_num      1      1      0      1
## BLEND_num      1      1      1      0
## VOTE_PREF_num      1      1      1      1
## SAYSO_num      1      1      1      1
##      VOTE_PREF_num SAYSO_num INCSUPP_num HEALTH_num DREAMACT_num
## CULTURAL_IDENTITY      1      1      1      1      1
## AMERICAN_num      1      1      1      1      1
## DISTINCT_num      1      1      1      1      1
## BLEND_num      1      1      1      1      1
## VOTE_PREF_num      0      1      1      1      1
## SAYSO_num      1      0      1      1      1
##      AGE RACE1 HHINC
## CULTURAL_IDENTITY      1      1      1
## AMERICAN_num      1      1      1
## DISTINCT_num      1      1      1
## BLEND_num      1      1      1
## VOTE_PREF_num      1      1      1
## SAYSO_num      1      1      1
```

Robustness Checks

```
# Define the variables to impute - make sure these are the column names with missing values
vars_to_impute <- c("CULTURAL_IDENTITY", "AMERICAN_num", "DISTINCT_num", "BLEND_num",
  "VOTE_PREF_num", "SAYSO_num", "INCSUPP_num", "HEALTH_num",
  "DREAMACT_num", "AGE", "RACE1", "HHINC")

# Perform multiple imputation using a linear model ('norm')
# This assumes that 'lms_subset' is your dataset and it has been properly loaded and prepared
imputed_data_linear <- mice(data = lms_subset[, vars_to_impute], m = 5, maxit = 20, method = 'norm', seed = 1234)

##
## iter imp variable
## 1 1
## 1 2
## 1 3
## 1 4
## 1 5
## 2 1
## 2 2
## 2 3
## 2 4
## 2 5
## 3 1
## 3 2
## 3 3
## 3 4
## 3 5
## 4 1
```


##	4	2
##	4	3
##	4	4
##	4	5
##	5	1
##	5	2
##	5	3
##	5	4
##	5	5
##	6	1
##	6	2
##	6	3
##	6	4
##	6	5
##	7	1
##	7	2
##	7	3
##	7	4
##	7	5
##	8	1
##	8	2
##	8	3
##	8	4
##	8	5
##	9	1
##	9	2
##	9	3
##	9	4
##	9	5
##	10	1
##	10	2
##	10	3
##	10	4
##	10	5
##	11	1
##	11	2
##	11	3
##	11	4
##	11	5
##	12	1
##	12	2
##	12	3
##	12	4
##	12	5
##	13	1
##	13	2
##	13	3
##	13	4
##	13	5
##	14	1
##	14	2
##	14	3
##	14	4
##	14	5

```
## 15 1
## 15 2
## 15 3
## 15 4
## 15 5
## 16 1
## 16 2
## 16 3
## 16 4
## 16 5
## 17 1
## 17 2
## 17 3
## 17 4
## 17 5
## 18 1
## 18 2
## 18 3
## 18 4
## 18 5
## 19 1
## 19 2
## 19 3
## 19 4
## 19 5
## 20 1
## 20 2
## 20 3
## 20 4
## 20 5
```

```
# Diagnostic plots to check convergence for the linear model imputation
# This will generate plots for each imputed variable across all iterations and imputations
plot(imputed_data_linear)
```

mean

sd

Iteration

```
# Obtain the first imputed dataset from the linear model imputation
single_completed_data_linear <- complete(imputed_data_linear, 1)

# Replace the missing values in the original dataset with the first imputed dataset from the linear model
lns_subset[vars_to_impute] <- single_completed_data_linear[vars_to_impute]

# Check dimensions to ensure they match
print(dim(lns_subset))

## [1] 4785  27

print(dim(single_completed_data_linear))

## [1] 4785  12

# Compare summary statistics for the dataset before and after linear model imputation
# You may need to replace 'single_completed_data' with your original imputed dataset if you intend to c
# with 'single_completed_data_linear'. This code assumes 'single_completed_data' is defined elsewhere a
summary(single_completed_data) # Replace 'single_completed_data' as needed

## CULTURAL_IDENTITY AMERICAN_num DISTINCT_num BLEND_num
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.000
## Median :3.500 Median :4.000 Median :3.000 Median :3.000
## Mean :3.441 Mean :3.368 Mean :2.742 Mean :2.392
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.000
## Max. :4.000 Max. :4.000 Max. :3.000 Max. :3.000
##
## VOTE_PREF_num SAYSO_num INCSUPP_num HEALTH_num DREAMACT_num
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.000
## Median :3.000 Median :2.00 Median :3.000 Median :4.000 Median :4.000
```

```
## Mean :2.111 Mean :2.39 Mean :3.222 Mean :3.422 Mean :3.264
## 3rd Qu.:3.000 3rd Qu.:3.00 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :3.000 Max. :4.00 Max. :4.000 Max. :4.000 Max. :4.000
##
## AGE RACE1
## Min. :18.00 (1) White :1257
## 1st Qu.:30.00 (2) Black, African American or Negro : 43
## Median :42.00 (3) American Indian or Alaskan Native : 65
## Mean :43.69 (4) Asian Indian : 4
## 3rd Qu.:56.00 (5) Native Hawaiian or Pacific Islander: 16
## Max. :97.00 (6) Some other race (specify) :3400
##
## HHINC
## (1) BELOW $15,000K: 610
## (2) $15,000-24,999: 751
## (3) $25,000-34,999: 757
## (4) $35,000-44,999: 681
## (5) $45,000-54,999: 510
## (6) $55,000-64,999: 398
## (7) ABOVE $65,000 :1078
```

```
summary(single_completed_data_linear)
```

```
## CULTURAL_IDENTITY AMERICAN_num DISTINCT_num BLEND_num
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.000
## Median :3.500 Median :4.000 Median :3.000 Median :3.000
## Mean :3.441 Mean :3.368 Mean :2.742 Mean :2.392
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.000
## Max. :4.000 Max. :4.000 Max. :3.000 Max. :3.000
##
## VOTE_PREF_num SAYSO_num INCSUPP_num HEALTH_num DREAMACT_num
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.000
## Median :3.000 Median :2.00 Median :3.000 Median :4.000 Median :4.000
## Mean :2.111 Mean :2.39 Mean :3.222 Mean :3.422 Mean :3.264
## 3rd Qu.:3.000 3rd Qu.:3.00 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :3.000 Max. :4.00 Max. :4.000 Max. :4.000 Max. :4.000
##
## AGE RACE1
## Min. :18.00 (1) White :1257
## 1st Qu.:30.00 (2) Black, African American or Negro : 43
## Median :42.00 (3) American Indian or Alaskan Native : 65
## Mean :43.69 (4) Asian Indian : 4
## 3rd Qu.:56.00 (5) Native Hawaiian or Pacific Islander: 16
## Max. :97.00 (6) Some other race (specify) :3400
##
## HHINC
## (1) BELOW $15,000K: 610
## (2) $15,000-24,999: 751
## (3) $25,000-34,999: 757
## (4) $35,000-44,999: 681
## (5) $45,000-54,999: 510
## (6) $55,000-64,999: 398
## (7) ABOVE $65,000 :1078
```

Outlier Handling

```
# Define the lower and upper bounds for the AGE variable based on Tukey's method
lower_bound_age <- -9 # Lower bound for AGE
upper_bound_age <- 95 # Upper bound for AGE

# Filter the dataset to remove rows with outliers in the AGE variable
filtered_lns_subset <- lns_subset[lns_subset$AGE >= lower_bound_age & lns_subset$AGE <= upper_bound_age]

# Check the dimensions of the filtered dataset to confirm rows were removed
print(dim(filtered_lns_subset))

## [1] 4784 27
```

Final Housekeeping

```
# Convert EDUCATION and HHINC to numeric
# Define the mapping for EDUCATION
education_mapping <- c(
  "(0) None" = 0,
  "(1) Eighth grade or below" = 1,
  "(2) Some high school" = 2,
  "(3) GED" = 3,
  "(4) High school graduate" = 4,
  "(5) Some college" = 5,
  "(6) 4 year college degree" = 6,
  "(7) Graduate or professional degree" = 7
)

# Define the mapping for HHINC
hhinc_mapping <- c(
  "(1) BELOW $15,000" = 1,
  "(2) $15,000-24,999" = 2,
  "(3) $25,000-34,999" = 3,
  "(4) $35,000-44,999" = 4,
  "(5) $45,000-54,999" = 5,
  "(6) $55,000-64,999" = 6,
  "(7) ABOVE $65,000" = 7
)

# Apply the mappings to your variables
lns_subset$EDUCATION <- education_mapping[lns_subset$EDUCATION]
lns_subset$HHINC <- hhinc_mapping[lns_subset$HHINC]
```

Data Type Conversions

```
# Convert BIRTH_ORIGIN, SEX, NATIONALITY and RELIGION to factors
lns_subset <- lns_subset %>%
  mutate(BIRTH_ORIGIN = as.factor(BIRTH_ORIGIN),
         SEX = as.factor(SEX),
         NATIONALITY = as.factor(NATIONALITY))

lns_subset$RELIGION <- as.factor(lns_subset$RELIGION)
```

```
# Simplify variable names by removing '_num' suffix
names(lns_subset) <- gsub("_num", "", names(lns_subset))
```

Summary Statistics and Checks

Final Summary Statistics

```
# Check for missing data in each column
missing_data_summary <- colSums(is.na(lns_subset))

# Print the summary
print(missing_data_summary)
```

```
## CULTURAL_IDENTITY      AMERICAN      LEARNENG      DISTINCT
##              0              0              0              0
##      KEEPSPAN      BLEND      VOTE_PREF      IDEOLOGY
##              0              0              0              0
##      FEELPART      PARTYID      SAYSO      GOVTRUST
##              0              0              0              0
##      INCSUPP      HEALTH      DREAMACT      IMMVIEW
##              0              0              0              0
##      IMPPOLICY      EDUCATION      NATIONALITY      BIRTH_ORIGIN
##              0              0              0              0
##      GEN_STATUS      AGE      SEX      RACE1
##              0              0              0              0
##      HHINC      RELIGION      WEIGHT
##              0              0              0
```

```
str(lns_subset)
```

```
## 'data.frame':  4785 obs. of  27 variables:
## $ CULTURAL_IDENTITY: num  4 4 3 3.5 3 3.5 4 4 4 4 ...
## $ AMERICAN         : num  2 3 2 3 3 4 3 4 1 2 ...
## $ LEARNENG          : num  4 4 4 3 4 4 4 4 4 4 ...
## $ DISTINCT         : num  2 3 3 3 3 3 3 3 2 3 ...
## $ KEEPSPAN         : num  4 4 4 3 4 4 4 4 4 3 ...
## $ BLEND            : num  2 3 3 3 3 2 3 3 3 3 ...
## $ VOTE_PREF        : num  3 3 3 2 3 1 3 2 3 1 ...
## $ IDEOLOGY         : num  2 1 2 2 2 2 2 2 2 1 ...
## $ FEELPART         : num  3 5 3 3 3 3 5 5 3 2 ...
## $ PARTYID          : num  3 3 2 3 2 3 3 3 2 1 ...
## $ SAYSO            : num  4 1 4 1 4 1 3 4 1 2 ...
## $ GOVTRUST         : num  2 3 2 2 2 1 4 1 4 3 ...
## $ INCSUPP          : num  1 3 4 4 4 3 4 4 1 3 ...
## $ HEALTH           : num  3 4 2 3 4 3 3 4 4 3 ...
## $ DREAMACT         : num  4 4 4 3 3 4 3 4 4 3 ...
## $ IMMVIEW          : num  2 2 2 2 2 2 2 2 2 2 ...
## $ IMPPOLICY        : num  5 5 5 4 5 4 4 3 2 4 ...
## $ EDUCATION        : num  1 4 5 2 1 5 4 5 1 7 ...
## $ NATIONALITY      : Factor w/ 2 levels "Other Latin Countries",...: 1 2 1 2 1 2 1 2 1 1 ...
## $ BIRTH_ORIGIN     : Factor w/ 2 levels "(1) YES","Missing": 1 2 1 2 1 2 1 2 1 1 ...
## $ GEN_STATUS       : num  1 3 1 2 1 3 1 3 1 1 ...
## $ AGE              : num  57 69 20 57 35 29 53 74 30 62 ...
## $ SEX              : Factor w/ 2 levels "(1) MALE","(2) FEMALE": 2 1 1 2 2 2 1 2 1 1 ...
```

```
## $ RACE1          : Factor w/ 6 levels "(1) White","(2) Black, African American or Negro",...: 6 6 6
## $ HHINC          : num  3 4 2 1 4 1 2 3 3 6 ...
## $ RELIGION       : Factor w/ 3 levels "Christian","Non-Secular",...: 1 1 1 1 1 3 1 1 1 1 ...
## $ WEIGHT         : num  1.16 0.53 2.77 0.74 0.55 1.14 1.01 0.99 0.5 0.87 ...
## - attr(*, "variable.labels")= Named chr [1:427] "Sequential Record Identifier (Added by ICPSR)" "Re
## ..- attr(*, "names")= chr [1:427] "CASEID" "RESPID" "RSTATE" "GEOG" ...
## - attr(*, "codepage")= int 28591
```

```
summary(lns_subset)
```

```
## CULTURAL_IDENTITY    AMERICAN      LEARNENG      DISTINCT
## Min.   :1.000        Min.   :1.000    Min.   :1.000    Min.   :1.000
## 1st Qu.:3.000        1st Qu.:3.000    1st Qu.:4.000    1st Qu.:3.000
## Median :3.500        Median :4.000    Median :4.000    Median :3.000
## Mean   :3.441        Mean   :3.368    Mean   :3.883    Mean   :2.742
## 3rd Qu.:4.000        3rd Qu.:4.000    3rd Qu.:4.000    3rd Qu.:3.000
## Max.   :4.000        Max.   :4.000    Max.   :4.000    Max.   :3.000
##  KEEPSPAN            BLEND          VOTE_PREF      IDEOLOGY
## Min.   :1.000        Min.   :1.000    Min.   :1.000    Min.   :1.000
## 1st Qu.:4.000        1st Qu.:2.000    1st Qu.:1.000    1st Qu.:2.000
## Median :4.000        Median :3.000    Median :3.000    Median :2.000
## Mean   :3.761        Mean   :2.392    Mean   :2.111    Mean   :1.904
## 3rd Qu.:4.000        3rd Qu.:3.000    3rd Qu.:3.000    3rd Qu.:2.000
## Max.   :4.000        Max.   :3.000    Max.   :3.000    Max.   :3.000
##  FEELPART            PARTYID          SAYSO          GOVTRUST      INCSUPP
## Min.   :1.000        Min.   :1.000    Min.   :1.00    Min.   :1.000    Min.   :1.000
## 1st Qu.:3.000        1st Qu.:2.000    1st Qu.:1.00    1st Qu.:2.000    1st Qu.:3.000
## Median :3.000        Median :2.000    Median :2.00    Median :2.000    Median :3.000
## Mean   :3.219        Mean   :2.301    Mean   :2.39    Mean   :2.241    Mean   :3.222
## 3rd Qu.:4.000        3rd Qu.:3.000    3rd Qu.:3.00    3rd Qu.:3.000    3rd Qu.:4.000
## Max.   :5.000        Max.   :3.000    Max.   :4.00    Max.   :4.000    Max.   :4.000
##  HEALTH              DREAMACT          IMMVIEW          IMPPOLICY
## Min.   :1.000        Min.   :1.000    Min.   :1.000    Min.   :1.000
## 1st Qu.:3.000        1st Qu.:3.000    1st Qu.:2.000    1st Qu.:2.000
## Median :4.000        Median :4.000    Median :2.000    Median :3.000
## Mean   :3.422        Mean   :3.264    Mean   :1.865    Mean   :3.156
## 3rd Qu.:4.000        3rd Qu.:4.000    3rd Qu.:2.000    3rd Qu.:5.000
## Max.   :4.000        Max.   :4.000    Max.   :2.000    Max.   :5.000
##  EDUCATION            NATIONALITY        BIRTH_ORIGIN      GEN_STATUS
## Min.   :0.000        Other Latin Countries:1875    (1) YES:1875    Min.   :1.000
## 1st Qu.:3.000        US Born                      :2910    Missing:2910    1st Qu.:1.000
## Median :4.000                                     Median :2.000
## Mean   :4.163                                     Mean   :1.822
## 3rd Qu.:5.000                                     3rd Qu.:2.000
## Max.   :7.000                                     Max.   :3.000
##  AGE                  SEX
## Min.   :18.00        (1) MALE :2157
## 1st Qu.:30.00        (2) FEMALE:2628
## Median :42.00
## Mean   :43.69
## 3rd Qu.:56.00
## Max.   :97.00
##
## RACE1              HHINC
## (1) White          :1257    Min.   :1.000
## (2) Black, African American or Negro : 43    1st Qu.:2.000
```

```
## (3) American Indian or Alaskan Native : 65 Median :4.000
## (4) Asian Indian : 4 Mean :4.094
## (5) Native Hawaiian or Pacific Islander: 16 3rd Qu.:6.000
## (6) Some other race (specify) :3400 Max. :7.000
## RELIGION WEIGHT
## Christian :4026 Min. :0.3200
## Non-Secular: 331 1st Qu.:0.5200
## Other : 428 Median :0.6900
## Mean :0.9841
## 3rd Qu.:1.0700
## Max. :4.3400
```

```
# Getting dimensions of the final dataset
```

```
final_dimensions <- dim(lns_subset)
```

```
# Displaying the number of observations and variables
```

```
print(paste("The final dataset has", final_dimensions[1], "observations and", final_dimensions[2], "variables"))
```

```
## [1] "The final dataset has 4785 observations and 27 variables."
```

```
# Displaying summary statistics for the final dataset
```

```
final_summary_stats <- summary(lns_subset)
```

```
print("Summary Statistics for the final dataset:")
```

```
## [1] "Summary Statistics for the final dataset:"
```

```
print(final_summary_stats)
```

```
## CULTURAL_IDENTITY AMERICAN LEARNENG DISTINCT
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:4.000 1st Qu.:3.000
## Median :3.500 Median :4.000 Median :4.000 Median :3.000
## Mean :3.441 Mean :3.368 Mean :3.883 Mean :2.742
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:3.000
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :3.000
## KEESPAN BLEND VOTE_PREF IDEOLOGY
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:4.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000
## Median :4.000 Median :3.000 Median :3.000 Median :2.000
## Mean :3.761 Mean :2.392 Mean :2.111 Mean :1.904
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:2.000
## Max. :4.000 Max. :3.000 Max. :3.000 Max. :3.000
## FEELPART PARTYID SAYSO GOVTRUST INCSUPP
## Min. :1.000 Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:1.00 1st Qu.:2.000 1st Qu.:3.000
## Median :3.000 Median :2.000 Median :2.00 Median :2.000 Median :3.000
## Mean :3.219 Mean :2.301 Mean :2.39 Mean :2.241 Mean :3.222
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.00 3rd Qu.:3.000 3rd Qu.:4.000
## Max. :5.000 Max. :3.000 Max. :4.00 Max. :4.000 Max. :4.000
## HEALTH DREAMACT IMMVIEW IMPPOLICY
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:2.000
## Median :4.000 Median :4.000 Median :2.000 Median :3.000
## Mean :3.422 Mean :3.264 Mean :1.865 Mean :3.156
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:5.000
## Max. :4.000 Max. :4.000 Max. :2.000 Max. :5.000
```



```

##      EDUCATION                NATIONALITY    BIRTH_ORIGIN    GEN_STATUS
##  Min.      :0.000    Other Latin Countries:1875    (1) YES:1875    Min.      :1.000
##  1st Qu.:3.000    US Born                :2910    Missing:2910    1st Qu.:1.000
##  Median :4.000
##  Mean   :4.163
##  3rd Qu.:5.000
##  Max.   :7.000
##
##      AGE                SEX
##  Min.      :18.00    (1) MALE :2157
##  1st Qu.:30.00    (2) FEMALE:2628
##  Median :42.00
##  Mean   :43.69
##  3rd Qu.:56.00
##  Max.   :97.00
##
##                                RACE1                HHINC
##  (1) White                        :1257    Min.      :1.000
##  (2) Black, African American or Negro : 43    1st Qu.:2.000
##  (3) American Indian or Alaskan Native : 65    Median :4.000
##  (4) Asian Indian                  : 4    Mean   :4.094
##  (5) Native Hawaiian or Pacific Islander: 16    3rd Qu.:6.000
##  (6) Some other race (specify)        :3400    Max.      :7.000
##
##      RELIGION                WEIGHT
##  Christian :4026    Min.      :0.3200
##  Non-Secular: 331    1st Qu.:0.5200
##  Other      : 428    Median :0.6900
##
##                                Mean   :0.9841
##                                3rd Qu.:1.0700
##                                Max.    :4.3400

```

Save Processed Data

Save as RDA and CSV

```

# Save Processed Data
save(lns_subset, file = "~/Desktop/Dissertation Analysis/Dissertation Analysis/Data/Processed/lns_clean

```

End of Script