# NSDUH2015_codebook

*George Rhodes*

*November 12, 2017*

1) Study Design In the first section of your codebook, titled **Study Design**, describe the study design including the purpose of the study, the sponsor of the study, the name of the data collection organization, and the specific methodology used including the mode of data collection, the method of participant recruitment (if any), and the length of the field period. (12 points)

2) Sampling The second section, **Sampling** should clearly document all available sampling information. This includes a description of the population, the methods used to draw the sample, and any special conditions associated with the sample (i.e groups that were oversampled). (12 points)

3) Variable Index

```
# Setting working directory
setwd("/Users/george/Documents/School/UW/SOC321/honors_thesis/321_project")

# Loading libraries
library(memisc)
```

```
## Warning: package 'memisc' was built under R version 3.4.2
```

```
## Loading required package: lattice
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'memisc'
```

```
## The following objects are masked from 'package:stats':
##
##     contr.sum, contr.treatment, contrasts
```

```
## The following object is masked from 'package:base':
##
##     as.array
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:memisc':
##
##     collect, recode, rename
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
```

```
##     intersect, setdiff, setequal, union
```

```
#load National Survey on Drug Abuse and Mental Health and create object nsduh2015.
load("/Users/george/Documents/School/UW/SOC321/honors_thesis/honors_thesis/NSDUH-2015-survey-data.rda")

#change all variable names to all lowercase for ease of calling.
names(PUF2015_102016) <- tolower(names(PUF2015_102016))

#subset to relevant variables.
subset_nsduh2015 <- PUF2015_102016 %>%
  select(questid2, filedate, txevrrcvd, alclottm, catag6, irsex, newrace2, irmaritstat, eduhighcat,
         al30est, alcbng30d, irpinc3, irfamin3, poverty3, coutyp2)
#rename columns
colnames(subset_nsduh2015) <- c("ident", "date", "treatment", "alcohol", "age", "sex", "race", "marital
                                "drink30", "binge30", "income", "famincome", "poverty", "countytype")

#recovered_respondants set to those that have recieved treatment AND not drank in last 12 months.
recovered_respondants <- subset_nsduh2015 %>%
  filter(treatment == 1 & alcohol == 93)
  # 925 observations.

#selects ident, treatment, alcohol, edu, income, famincome, poverty, and county type.
project_variables <- recovered_respondants %>%
  dplyr::select(ident, treatment, alcohol, edu, income, famincome, poverty, countytype)
```

```
#I have some questions about translating recoded variables. Since some are not questions, but inferred

# Create data.set object from "data" object (tbl)
data_set <- as.data.set(project_variables)

# Look at new data.set object
data_set
```

```
##
## Data set with 925 observations and 8 variables
##
##       ident treatment alcohol edu income famincome poverty countytype
##  1 55015443         1      93   3      6         6       3           2
##  2 19892643         1      93   5      1         2       1           1
##  3 43267743         1      93   1      2         3       2           3
##  4 69210843         1      93   1      5         7       3           3
##  5 11374843         1      93   3      2         6       3           3
##  6 43230943         1      93   2      4         5       3           2
##  7 49340943         1      93   3      2         2       1           1
##  8 57111943         1      93   2      1         2       1           1
##  9 88912943         1      93   1      1         2       1           1
## 10 29323043         1      93   4      3         3       3           2
## 11 67422153         1      93   3      4         4       2           3
## 12 47709353         1      93   3      1         2       1           2
## 13 15784353         1      93   1      3         5       3           1
## 14 77205753         1      93   2      1         1       1           3
## 15 91953753         1      93   3      3         5       3           1
## 16 96879053         1      93   4      3         5       3           2
## 17 86024263         1      93   5      1         7       3           1
## 18 15124263         1      93   3      1         3       2           2
```

```
## 19 73389363         1      93   1      4         7       3          3
## 20 39670463         1      93   3      4         4       3          1
## 21 45380663         1      93   3      5         7       3          2
## 22 23784863         1      93   2      3         3       2          2
## 23 89578063         1      93   3      1         2       1          2
## 24 65399173         1      93   2      4         4       2          3
## 25 65570173         1      93   2      2         2       2          3
## .. ........ ......... ....... ... ...... ......... ....... ..........
## (25 of 925 observations shown)
```

```
# Creating variable index for data_set
data_set <- within(data_set,{
  # Description of the variables
  description(ident) <- "Unique person identification within given study year"
  description(treatment) <- "Ever recieved treatment"
  description(alcohol) <- "Used alcohol in last 12 months"
  description(edu) <- "Education Level"
  description(income) <- "Income level"
  description(famincome) <- "Family income"
  description(poverty) <- "Family income and local poverty level"
  description(countytype) <- "urban/rural"

  # Wording of survey item/interview questions (if applicable)
  wording(treatment) <- "Have you ever received treatment or counseling for your use of alcohol or any d
  wording(alcohol) <- "During the past 12 months, was there a month or more when you spent a lot of you
  wording(edu) <- "Re-coded education categories."
  wording(income) <- "Respondant total income (finer categories) - IMP REV."
  wording(famincome) <- "Recode - IMP Revised - Total family income."
  wording(poverty) <- "Recode - Poverty level (% of US Census poverty threshold)."
  wording(countytype) <- "County Metro/nonmetro status (3-level)"

  # type of measurement
  measurement(ident) <- "nominal"
  measurement(treatment) <- "nominal"
  measurement(alcohol) <- "nominal"
  measurement(edu) <- "ordinal"
  measurement(income) <- "ordinal"
  measurement(famincome) <- "ordinal"
  measurement(poverty) <- "ordinal"
  measurement(countytype) <- "nominal"

  # labels associated with underlying numeric
  labels(treatment) <- c(
    "Yes"                                                 =  1,
    "No"                                                  =  2,
    "BAD DATA Logically assigned"                         =  85,
    "Never used alcohol or drugs"                         =  91,
    "Don't know"                                          =  94,
    "Refused"                                             =  97,
    "blank (no answer)"                                   =  98)
  labels(alcohol) <- c(
    "Yes"                                                       =  1,
    "No"                                                        =  2,
    "Did not use alcohol past 12 months or used <6 days log asn" =  83,
```

```
    "BAD DATA Logically assigned"                                  = 85,
    "Never used alcohol"                                           = 91,
    "Did not use alcohol past 12 months or used <6 days"           = 93,
    "Don't know"                                                   = 94,
    "Refused"                                                      = 97,
    "blank (no answer)"                                            = 98)
  labels(edu) <- c(
    "Less than high school"                          = 1,
    "High school grad"                               = 2,
    "Some college/associates degree"                 = 3,
    "College graduate"                               = 4,
    "12 to 17 year old"                              = 5)
  labels(income) <- c(
    "Less than $10,000 (including loss)"             = 1,
    "$10,000 - $19,999"                              = 2,
    "$20,000 - $29,999"                              = 3,
    "$30,000 - $39,999"                              = 4,
    "$40,000 - $49,999"                              = 5,
    "$50,000 - 74,999"                               = 6,
    "$75,000 or more"                                = 7)
  labels(famincome) <- c(
    "Less than $10,000 (including loss)"             = 1,
    "$10,000 - $19,999"                              = 2,
    "$20,000 - $29,999"                              = 3,
    "$30,000 - $39,999"                              = 4,
    "$40,000 - $49,999"                              = 5,
    "$50,000 - 74,999"                               = 6,
    "$75,000 or more"                                = 7)
  labels(poverty) <- c(
    "Living in poverty"                              = 1,
    "Income up to 2x federal poverty threashold"     = 2,
    "Income more than 2x federal poverty threshold" = 3)
  labels(countytype) <- c(
    "Large Metro"                                    = 1,
    "Small Metro"                                    = 2,
    "Nonmetro"                                       = 3)

  # any annotation/notes you wish to add
  annotation(poverty)["Note"] <- "The poverty threshold was determined for each individual based on thei
  annotation(countytype)["Note"] <- "The definitions here are confusing and may not perfectly fit an url
  annotation(edu)["Note"] <- "Respondants are 18 years old and older unless they answered 5, in which ca

  # How are missing values coded? NEED!
  missing.values(treatment) <- c(85, 94, 97, 98)
  missing.values(alcohol) <- c(85, 94, 97, 98)
  missing.values(edu) <- c()
  missing.values(income) <- c()
  missing.values(famincome) <- c()
  missing.values(poverty) <- c()
  missing.values(countytype) <- c()

})
```

```
codebook(data_set)
```

```
## ------------------------------------------------------------------------
##
##    ident 'Unique person identification within given study year'
##
## ------------------------------------------------------------------------
##
##    Storage mode: character
##    Measurement: nominal
##
##    Values and labels     N     Percent
##
##         (unlab.vld.)   925   100.0 100.0
##
## ------------------------------------------------------------------------
##
##    treatment 'Ever recieved treatment'
##
##    "Have you ever received treatment or counseling for your use of alcohol
##    or any drug, not counting cigarettes?"
##
## ------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: nominal
##    Missing values: 85, 94, 97, 98
##
##                     Values and labels     N     Percent
##
##     1   'Yes'                            925   100.0 100.0
##     2   'No'                               0     0.0   0.0
##    85 M 'BAD DATA Logically assigned'      0           0.0
##    91   'Never used alcohol or drugs'     0     0.0   0.0
##    94 M 'Don't know'                       0           0.0
##    97 M 'Refused'                          0           0.0
##    98 M 'blank (no answer)'                0           0.0
##
## ------------------------------------------------------------------------
##
##    alcohol 'Used alcohol in last 12 months'
##
##    "During the past 12 months, was there a month or more when you spent a
##    lot of your time getting or drinking alcohol?"
##
## ------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: nominal
##    Missing values: 85, 94, 97, 98
##
##                                                 Values and labels     N     Percent
##
##     1   'Yes'                                                        0     0.0   0.0
```

```
##     2   'No'                                                              0     0.0   0.0
##    83   'Did not use alcohol past 12 months or used <6 days log asn'     0     0.0   0.0
##    85 M 'BAD DATA Logically assigned'                                     0           0.0
##    91   'Never used alcohol'                                             0     0.0   0.0
##    93   'Did not use alcohol past 12 months or used <6 days'           925   100.0 100.0
##    94 M 'Don't know'                                                     0           0.0
##    97 M 'Refused'                                                        0           0.0
##    98 M 'blank (no answer)'                                              0           0.0
##
## ============================================================================
##
##    edu 'Education Level'
##
##    "Re-coded education categories."
##
## ----------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: ordinal
##
##                        Values and labels    N    Percent
##
##    1   'Less than high school'           171   18.5 18.5
##    2   'High school grad'                293   31.7 31.7
##    3   'Some college/associates degree'  281   30.4 30.4
##    4   'College graduate'                124   13.4 13.4
##    5   '12 to 17 year old'                56    6.1  6.1
##
##    Note:
##
##        Respondants are 18 years old and older unless they answered 5,
##        in which case they are 12-17 years old.
##
## ============================================================================
##
##    income 'Income level'
##
##    "Respondant total income (finer categories) - IMP REV."
##
## ----------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: ordinal
##
##                             Values and labels    N    Percent
##
##    1   'Less than $10,000 (including loss)'  353   38.2 38.2
##    2   '$10,000 - $19,999'                   212   22.9 22.9
##    3   '$20,000 - $29,999'                   110   11.9 11.9
##    4   '$30,000 - $39,999'                    77    8.3  8.3
##    5   '$40,000 - $49,999'                    54    5.8  5.8
##    6   '$50,000 - 74,999'                     62    6.7  6.7
##    7   '$75,000 or more'                      57    6.2  6.2
##
```

```
## ============================================================================
##
##    famincome 'Family income'
##
##    "Recode - IMP Revised - Total family income."
##
## ----------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: ordinal
##
##                            Values and labels    N    Percent
##
##    1   'Less than $10,000 (including loss)'  123   13.3 13.3
##    2   '$10,000 - $19,999'                   165   17.8 17.8
##    3   '$20,000 - $29,999'                   124   13.4 13.4
##    4   '$30,000 - $39,999'                   100   10.8 10.8
##    5   '$40,000 - $49,999'                    97   10.5 10.5
##    6   '$50,000 - 74,999'                    109   11.8 11.8
##    7   '$75,000 or more'                     207   22.4 22.4
##
## ============================================================================
##
##    poverty 'Family income and local poverty level'
##
##    "Recode - Poverty level (% of US Census poverty threshold)."
##
## ----------------------------------------------------------------------------
##
##    Storage mode: integer
##    Measurement: ordinal
##
##                                      Values and labels    N    Percent
##
##    1   'Living in poverty'                                252   27.2 27.2
##    2   'Income up to 2x federal poverty threashold'      245   26.5 26.5
##    3   'Income more than 2x federal poverty threshold'  428   46.3 46.3
##
##    Note:
##
##        The poverty threshold was determined for each individual based
##        on their age, family size, number of children in the
##        household, and total family income. This excludes persons aged
##        18-22 living in dorms, which was coded with a period.
##
## ============================================================================
##
##    countytype 'urban/rural'
##
##    "County Metro/nonmetro status (3-level)"
##
## ----------------------------------------------------------------------------
##
##    Storage mode: integer
```

```
##    Measurement: nominal
##
##    Values and labels    N    Percent
##
##    1   'Large Metro'  340   36.8 36.8
##    2   'Small Metro'  354   38.3 38.3
##    3   'Nonmetro'     231   25.0 25.0
##
##    Note:
##
##        The definitions here are confusing and may not perfectly fit
##        an urban/rural analysis.
```