Working with 2012-2013 BJS National Crime Victimization Survey Data

Pauline I. Alvarado

Fall 2018

This coding exercise was from University of Pennsylvania's Criminal Justice Data Science Course taught by Dr. Gregory Ridgeway. Script has been modified from course examples to fit personal notation conventions. Data was taken from the National Archive of Criminal Justice Data (NACJD).

Load and view data

List files

```
# 2012 Files
  list.files("NCVS2012/", recursive = TRUE)

# 2013 Files
  list.files("NCVS2013/", recursive = TRUE)
```

```
##  [1] "34650-Codebook.pdf"               "34650-descriptioncitation.html"
##  [3] "34650-manifest.txt"               "34650-related_literature.txt"
##  [5] "DS0001/34650-0001-Data.rda"       "DS0002/34650-0002-Data.rda"
##  [7] "DS0003/34650-0003-Data.rda"       "DS0004/34650-0004-Data.rda"
##  [9] "DS0005/34650-0005-Data.rda"       "factor_to_numeric_icpsr.R"
## [11] "series-95-related_literature.txt" "TermsOfUse.html"
##  [1] "35164-Codebook.pdf"               "35164-descriptioncitation.html"
##  [3] "35164-manifest.txt"               "35164-related_literature.txt"
##  [5] "DS0001/35164-0001-Data.rda"       "DS0002/35164-0002-Data.rda"
##  [7] "DS0003/35164-0003-Data.rda"       "DS0004/35164-0004-Data.rda"
##  [9] "DS0005/35164-0005-Data.rda"       "factor_to_numeric_icpsr.R"
## [11] "series-95-related_literature.txt" "TermsOfUse.html"
```

Load and categorize data

```
# Address record-type (DS1)
  load("NCVS2012/DS0001/34650-0001-Data.rda")
  load("NCVS2013/DS0001/35164-0001-Data.rda")

  addr12 <- da34650.0001
  addr13 <- da35164.0001
```

```
# Household information (DS2)
  load("NCVS2012/DS0002/34650-0002-Data.rda")
  load("NCVS2013/DS0002/35164-0002-Data.rda")

  household12 <- da34650.0002
  household13 <- da35164.0002

# Personal information (DS3)
  load("NCVS2012/DS0003/34650-0003-Data.rda")
  load("NCVS2013/DS0003/35164-0003-Data.rda")

  person12 <- da34650.0003
  person13 <- da35164.0003

# Incident information (DS4)
  load("NCVS2012/DS0004/34650-0004-Data.rda")
  load("NCVS2013/DS0004/35164-0004-Data.rda")

  incident12 <- da34650.0004
  incident13 <- da35164.0004

# Remove objects from the environment
  rm(da34650.0001, da34650.0002, da34650.0003, da34650.0003, da34650.0004, da35164.0001, da35164.0002
```

View month and year of each incident

```
with(incident12, table(V4014, V4015))
with(incident13, table(V4014, V4015))
```

```
##                 V4015
## V4014            2011 2012
##   (01) January      0  728
##   (02) February     0  658
##   (03) March        0  705
##   (04) April        0  751
##   (05) May          0  768
##   (06) June         0  825
##   (07) July       159  670
##   (08) August     296  560
##   (09) September  366  426
##   (10) October    492  298
##   (11) November   608  139
##   (12) December   766    0
##   (98) Residue      0    0
##                V4015
## V4014           2012 2013
##   (1) January      0  566
##   (2) February     0  580
##   (3) March        0  615
##   (4) April        0  526
##   (5) May          0  688
```

```
##   (6) June         0  649
##   (7) July       144  580
##   (8) August     245  474
##   (9) September  306  306
##   (10) October   440  238
##   (11) November  557  116
##   (12) December  697    0
##   (98) Residue     0    0
```

Create 2012 incident dataframe

Merge 2012 and 2013 incident data

```
# Bind rows
  incident2012 <- rbind(incident12, incident13)

# View years
  table(incident2012$V4015)

# Subset only 2012 information
  incident2012 <- subset(incident2012, V4015 == 2012)
```

```
##
## 2011 2012 2013
## 2687 8917 5338
```

Exclude crimes occurring outside US or have unknown location

```
incident2012 <- subset(incident2012, V4022 != "(1) Outside U.S." | is.na(V4022))
```

Only include up to 10 occurrences in a series crime (BJS convention)

```
i <- with(incident2012, which((V4019 == "(2) No (is series)") & (V4016 >= 11) & (V4016 <= 996)))
incident2012$V4016[i] <- 10
incident2012$V4016[incident2012$V4016 >= 997] <- NA
```

Create a "date year" weight

```
i <- which(incident2012$V4019 == "(2) No (is series)")
incident2012$WGTVICDY <- incident2012$WGTVICCY
incident2012$WGTVICDY[i] <- with(incident2012, WGTVICDY[i] * V4016[i])
```

Standardize coding of crime type

```
incident2012$V4529 <- gsub("\\(([1-9]))", "(0\\1)", incident2012$V4529)
```

Create 2012 household dataframe

Bind 2012 and 2013 incident data

```
household2012 <- rbind(househld12, househld13)
household2012 <- subset(household2012, YEARQ >= 2012.1 & YEARQ <= 2013.2)
```

Standardize coding of "month allocated" variable

```
household2012$V2130 <- gsub("\\(0", "\\(", household2012$V2130)
table(household2012$V2130)
```

```
##
##   (1) January  (10) October (11) November (12) December  (2) February
##         21214         10692         10597         10630         21140
##     (3) March     (4) April       (5) May      (6) June      (7) July
##         21397         21334         21172         21262         10572
##    (8) August (9) September
##         10624         10678
```

Create 2012 person dataframe

Fix incompatible factor and numeric values between 2012 and 2013

2012 factor levels look like "(1) Yes" but coded only as "1" in 2013.

```
# gives factor levels for each variable
  i <- sapply(person12, levels)

# gives factor levels or each factor variable, non-factor variables return null result
  i <- i[!sapply(i, is.null)]

# store in i, variable where factor levels begin with "("
  i <- sapply(i, function(x) all(substring(x,1,1) == "("))
  var.fix <- names(i)[i]

# for-loop to fix names
  for(xj in var.fix)
  {
    person12[,xj] <- gsub("\\(([0-9]+)\\).*", "\\1", person12[,xj]) # remove words that follow the pa
    person12[,xj] <- as.numeric(person12[,xj])
  }
```

Stack 2012 and 2013 data

```
person2012 <- rbind(person12, person13)
person2012 <- subset(person2012, YEARQ >= 2012.1 & YEARQ <= 2013.2)
```

Merge incident and person dataframes

```
# Match and merge data
  a <- merge(incident2012,                          # indicent data
             person2012[, c("IDPER", "YEARQ",       # unique ID's of person
                            "V3014",                # age
                            "V3015",                # marital status
                            "V3018")],              # sex
             by = c("IDPER", "YEARQ"),              # variables to match/merge
             all.x = TRUE)                          # keep all incidents even if not matched

  incident2012 <- a

# Rename age, marital, status, and sex variables
  names(incident2012)[names(incident2012) == "V3014"] <- "age"
  names(incident2012)[names(incident2012) == "V3015"] <- "marital"
  names(incident2012)[names(incident2012) == "V3018"] <- "sex"

# View first line of incident data to see if the merge worked!
  incident2012[1, c("IDPER", "YEARQ", "age", "marital", "sex")]
```

```
##                          IDPER  YEARQ age marital sex
## 1 2501051210759582293728435O1 2012.3  28       3   1
```

Create new variable: age categories

```
incident2012$ageGroup <- cut(incident2012$age, breaks = c(0,16,21,35,45,60,110))
```

Assign number values to actual names to age and marital status variables

```
# Deal with missing value indicators
  incident2012$marital[incident2012$marital == 8] <- NA

# Add labels
  incident2012$marital <- factor(incident2012$marital, levels = 1:5,
                                 labels = c("married", "widowed", "divorced", "separated", "never mar

  incident2012$sex <- factor(incident2012$sex, levels = 1:2,
                             labels = c("male", "female"))
```

Reshape data to find out common crime types by sex (of the crime victim)

Estimated counts (weighted)

```
a <- aggregate(WGTVICDY ~ V4529 + sex, data = incident2012, FUN = sum)
a <- reshape(a, timevar = "sex", idvar = "V4529", direction = "wide")
a[is.na(a)] <- 0
names(a) <- c("crimeType", "male", "female")
a
```

```
##                   crimeType        male      female
## 1       (01) Completed rape    6318.130   67991.536
## 2       (02) Attempted rape   42077.861   17423.911
## 3    (03) Sex aslt w s aslt   38218.021    2994.590
## 4     (05) Rob w inj s aslt   62532.486   16810.786
## 5     (06) Rob w inj m aslt   30571.215   46993.672
## 6        (07) Rob wo injury  112476.912   63550.334
## 7     (08) At rob inj s asl   22330.349    6638.802
## 8     (09) At rob inj m asl   12200.917   14668.799
## 9        (10) At rob w aslt  104657.340   44199.671
## 10    (11) Ag aslt w injury  188925.090  196423.404
## 11    (12) At ag aslt w wea  185157.394   85898.556
## 12     (13) Thr aslt w weap  237527.692  183883.312
## 13     (14) Simp aslt w inj  448773.257  506208.479
## 14     (15) Sex aslt wo inj    3119.587   29460.740
## 15    (16) Unw sex wo force    2957.926   13034.133
## 16 (17) Asl wo weap, wo inj 1042741.375  962894.567
## 17     (18) Verbal thr rape   26408.008   13337.490
## 18    (19) Ver thr sex aslt    9298.262    6071.520
## 19     (20) Verbal thr aslt 1099721.249  919823.826
## 20      (23) Pocket picking   81230.111   45187.984
## 21     (31) Burg, force ent  609106.185  606180.810
## 22    (32) Burg, ent wo for  741492.194 1016552.357
## 23     (33) Att force entry  269383.309  441969.018
## 24     (40) Motor veh theft  256959.885  223318.276
## 25    (41) At mtr veh theft   87364.540   78632.297
## 26     (5) Rob w inj s aslt   18001.951   35466.007
## 27         (54) Theft < $10  444360.185  670778.978
## 28       (55) Theft $10-$49 1217450.179 1682478.881
## 29      (56) Theft $50-$249 2261589.762 2657037.634
## 30         (57) Theft $250+ 1825854.971 1964564.610
## 31      (58) Theft value NA  588405.556  781094.421
## 32     (59) Attempted theft  349959.481  336192.254
## 33     (6) Rob w inj m aslt    5039.391   59148.610
## 34        (7) Rob wo injury   38185.106   21174.398
## 38   (04) Sex aslt w m aslt       0.000    6515.781
## 44       (1) Completed rape       0.000   54822.944
## 55       (2) Attempted rape       0.000    1640.455
## 57     (21) Purse snatching       0.000   15990.538
## 58     (22) At purse snatch       0.000    7272.660
## 60    (3) Sex aslt w s aslt       0.000    5774.439
```

```
## 75     (9) At rob inj m asl       0.000   10626.371
```

Compute row percentages to determine what percentage of each crime is male and female

```
temp <- a
row.total <- with(temp, male + female)
temp$male <- with(temp, 100 * male / row.total)
temp$female <- with(temp, 100 * female / row.total)
rowSums(temp[,-1])  # check that the rows sum to 100

temp$ratio <- temp$female / temp$male
temp[order(-temp$ratio),]
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
## 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  38  44  55  57  58  60
## 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
##  75
## 100
##                    crimeType      male     female       ratio
## 38   (04) Sex aslt w m aslt  0.000000 100.000000         Inf
## 44       (1) Completed rape  0.000000 100.000000         Inf
## 55       (2) Attempted rape  0.000000 100.000000         Inf
## 57     (21) Purse snatching  0.000000 100.000000         Inf
## 58     (22) At purse snatch  0.000000 100.000000         Inf
## 60    (3) Sex aslt w s aslt  0.000000 100.000000         Inf
## 75     (9) At rob inj m asl  0.000000 100.000000         Inf
## 33     (6) Rob w inj m aslt  7.850986  92.149014 11.73725303
## 1       (01) Completed rape  8.502433  91.497567 10.76133910
## 14     (15) Sex aslt wo inj  9.575063  90.424937  9.44379525
## 15    (16) Unw sex wo force 18.496217  81.503783  4.40651096
## 26     (5) Rob w inj s aslt 33.668671  66.331329  1.97012019
## 23     (33) Att force entry 37.869182  62.130818  1.64066964
## 5     (06) Rob w inj m aslt 39.413730  60.586270  1.53718690
## 27         (54) Theft < $10 39.847958  60.152042  1.50953888
## 28       (55) Theft $10-$49 41.982068  58.017932  1.38196939
## 22    (32) Burg, ent wo for 42.177099  57.822901  1.37095490
## 31      (58) Theft value NA 42.964992  57.035008  1.32747629
## 8     (09) At rob inj m asl 45.407688  54.592312  1.20227024
## 29      (56) Theft $50-$249 45.980099  54.019901  1.17485394
## 13     (14) Simp aslt w inj 46.992863  53.007137  1.12798272
## 30         (57) Theft $250+ 48.170260  51.829740  1.07596969
## 10    (11) Ag aslt w injury 49.027074  50.972926  1.03968935
## 21     (31) Burg, force ent 50.120357  49.879643  0.99519727
## 32     (59) Attempted theft 51.003220  48.996780  0.96066051
## 16 (17) Asl wo weap, wo inj 51.990561  48.009439  0.92342607
## 25    (41) At mtr veh theft 52.630244  47.369756  0.90004820
## 24     (40) Motor veh theft 53.502305  46.497695  0.86907836
## 19     (20) Verbal thr aslt 54.453910  45.546090  0.83641543
## 12     (13) Thr aslt w weap 56.364853  43.635147  0.77415526
## 18    (19) Ver thr sex aslt 60.497034  39.502966  0.65297359
```

```
## 6        (07) Rob wo injury 63.897444  36.102556  0.56500782
## 20      (23) Pocket picking 64.255130  35.744870  0.55629598
## 34        (7) Rob wo injury 64.328546  35.671454  0.55451983
## 17     (18) Verbal thr rape 66.442765  33.557235  0.50505477
## 11    (12) At ag aslt w wea 68.309658  31.690342  0.46392183
## 9        (10) At rob w aslt 70.307297  29.692703  0.42232748
## 2       (02) Attempted rape 70.716989  29.283011  0.41408737
## 7     (08) At rob inj s asl 77.083202  22.916798  0.29729950
## 4     (05) Rob w inj s aslt 78.812588  21.187412  0.26883284
## 3    (03) Sex aslt w s aslt 92.733801   7.266199  0.07835545
```