NCVS Accessibility & Publications

Department of Justice | Bureau of Justice Statistics

Miles Wang, Data Analytics Fellow | *UC Berkeley, Data Science & Statistics*

Keywords:

Data Transformation, Python Data Analysis, RStudio

Summary:

To aid in the Bureau of Justice Statistics (BJS)'s ongoing efforts to make National Crime Victimization Survey (NCVS) data more publicly accessible, Miles transformed the National Crime Victimization Survey (NCVS) database, encompassing over 6 million records, from SPSS to Excel format using Python and RStudio. This process involved adding descriptive labels and values to ensure the data was accessible and usable for public analysis. Additionally, conducted comprehensive data analysis using SPSS to extract meaningful insights from the NCVS, aiding ongoing research efforts.

Bureau of Justice Statistics

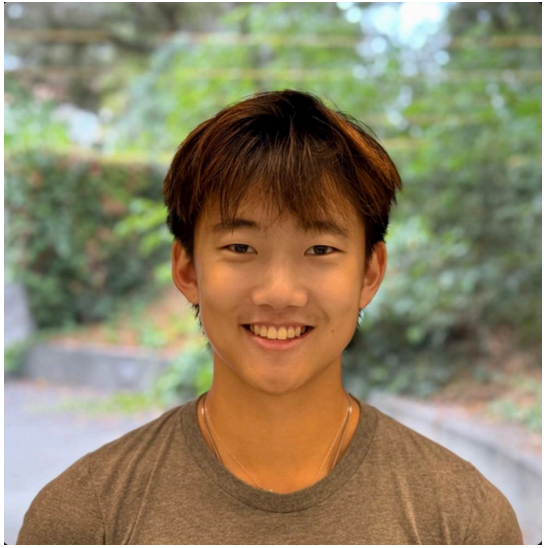
BJS Lunch & Learn
Coding It Forward Summer Presentation

Miles Wang
Coding it Forward 2024

Bureau of Justice Statistics

About Me

- Rising Junior at the University of California, Berkeley
- Majoring in Data Science, minoring in statistics.
- Grew up in Oakland and Berkeley, CA
- In my free time, I like to gym, thrift, and make jewelry



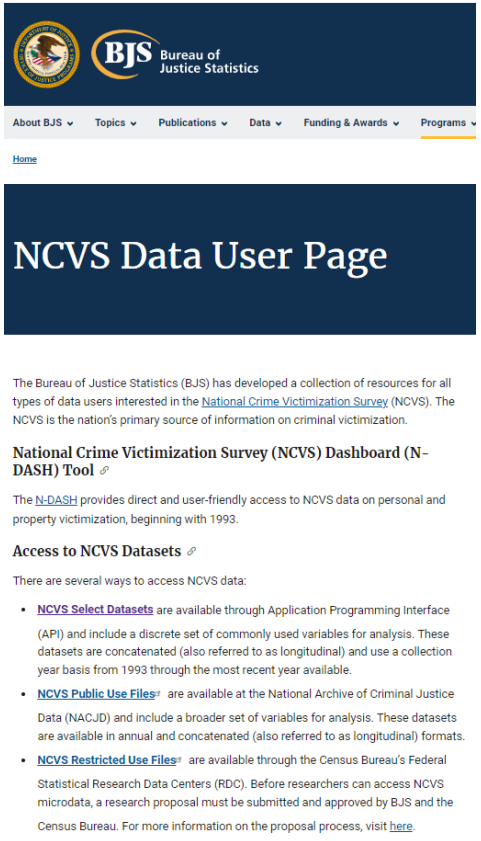
Project Overview

- NCVS Recode Shortlist
- NCVS Accessible Dataset Version
- Just The Stats: Urbanicity Publication
- Just The Stats: Victim/Offender Age Publication

Project Motivation

One of the primary objectives of my internship was to make the 'NCVS Data User Page' more accessible for new users.

- Add more accessible tools to explore the NCVS dataset using SPSS
- Add additional methods to explore the NCVS without the need for SPSS or other special software



NCVS Recode Shortlist

- In order to parse through the NCVS dataset (using SPSS software), one must use ‘recodes’ in order to parse through the numerical SPSS data and retrieve descriptive variable names and values.
- Appropriate recodes should be posted to the NCVS data user page to accommodate public users who are curious about the NCVS.

```
*Education recode that separates high school graduates and advanced degrees.
*Variable available from 2003 forward. Categories changed in 2003 to include advanced degrees and high
recode v3020     (lo thru 00 =1)                          /*No schooling
                            (1 thru 5=2)                  /*Grade school
                             (6 thru 8=3)                 /*Middle school
                             (9 thru 12, 27=4)            /*High school
                             (28=5)                       /*High school graduate
                             (21 thru 26, 40, 41=6)       /*Some college and associate's degree
                             (42=7)                       /*Bachelor's degree
                             (43 thru 45=8)               /*Advanced degree
                             (98=98)                      /*Residue
                             into educate.
if (year le 2002) educate=-1.                             /*Invalid prior to 2003.
execute.
*******************************************************************************
*Add variable and value labels to all newly created variables.
VARIABLE LABELS
 gender         'GENDER'
 hispanic       'HISPANIC'
 age            'AGE OF VICTIM'
 marital        'MARITAL STATUS'
 marital2       'MARITAL STATUS'
 race           'RACE'
 race1          'RACE1'
 hisgen         'HISPANIC ORIGIN & GENDER OF VICTIM'
 ethnic         'RACE & HISPANIC ORIGIN'
 ethnic1        'RACE & HISPANIC ORIGIN'
 ethgen         'ETHNICITY & GENDER'
 racegndr       'RACE & GENDER'
 hincome        'ANNUAL HOUSEHOLD INCOME'
 ed             'LEVEL OF EDUCATION'
 educatn        'GRADE COMPLETED'
 raceEX          'victim race recode splitting out Asian and NHOPI'
 raceethEX      'race and Hispanic origin of victim splitting out Asian and NHOPI'
  educate            'education recode separating out high school graduate and advanced degrees'.
```

NCVS Recode Shortlist, cont.

Goal: recommend a list of SPSS recodes that public users can utilize to easily recreate BJS findings.

- Edit/revise existing recodes
  - Omit redundant information
- Recommend the most useful existing recodes to publish

```
*Education recode that separates high school graduates and advanced degrees.
*Variable available from 2003 forward. Categories changed in 2003 to include advanced degrees and high
recode v3020     (lo thru 00 =1)                          /*No schooling
                            (1 thru 5=2)                  /*Grade school
                             (6 thru 8=3)                 /*Middle school
                             (9 thru 12, 27=4)            /*High school
                             (28=5)                       /*High school graduate
                             (21 thru 26, 40, 41=6)       /*Some college and associate's degree
                             (42=7)                       /*Bachelor's degree
                             (43 thru 45=8)               /*Advanced degree
                             (98=98)                      /*Residue
                             into educate.
if (year le 2002) educate=-1.                             /*Invalid prior to 2003.
execute.
*******************************************************************************
*Add variable and value labels to all newly created variables.
VARIABLE LABELS
 gender         'GENDER'
 hispanic       'HISPANIC'
 age            'AGE OF VICTIM'
 marital        'MARITAL STATUS'
 marital2       'MARITAL STATUS'
 race           'RACE'
 race1          'RACE1'
 hisgen         'HISPANIC ORIGIN & GENDER OF VICTIM'
 ethnic         'RACE & HISPANIC ORIGIN'
 ethnic1        'RACE & HISPANIC ORIGIN'
 ethgen         'ETHNICITY & GENDER'
 racegndr       'RACE & GENDER'
 hincome        'ANNUAL HOUSEHOLD INCOME'
 ed             'LEVEL OF EDUCATION'
 educatn        'GRADE COMPLETED'
 raceEX          'victim race recode splitting out Asian and NHOPI'
 raceethEX      'race and Hispanic origin of victim splitting out Asian and NHOPI'
  educate            'education recode separating out high school graduate and advanced degrees'.
```

NCVS Accessible Dataset

- Goal: Provide public users a method to access and parse NCVS data without needing SPSS software or other knowledge
- Solution: rewrite the NCVS incident data file into a .csv or .xlsx format that contains descriptive values instead of numerical ones

year	ager	sex	hispanic	race
1993	6.00	2.00	2.00	1.00
1993	4.00	2.00	2.00	1.00
1993	3.00	2.00	2.00	2.00
1993	2.00	1.00	2.00	1.00
1993	2.00	1.00	2.00	1.00
1993	2.00	1.00	2.00	1.00
1993	5.00	2.00	1.00	1.00

year	age	sex	hispanic	race
1993	65 or older	Female	Non-Hispanic	White
1993	35-49	Female	Non-Hispanic	White
1993	25-34	Female	Non-Hispanic	Black
1993	18-24	Male	Non-Hispanic	White
1993	18-24	Male	Non-Hispanic	White
1993	18-24	Male	Non-Hispanic	White
1993	50-64	Female	Hispanic	White

NCVS Accessible Dataset, cont.

- How it's done: use Rstudio to permanently recode variable labels and values
- Annotated Rstudio scripts to provide documentation for future conversion needs

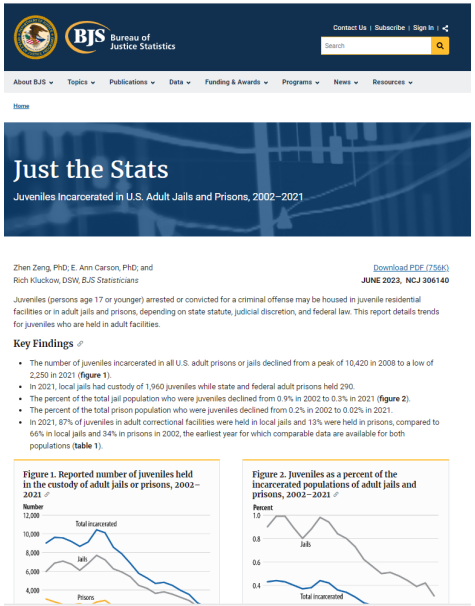
```
#create and run function to turn 'quarter' column into just quarters not including year data
keep_last <- function(x) {
  substr(x, nchar(x), nchar(x))
}
crime$quarter <- keep_last(crime$quarter)
#recode age column keys into values, so 1 -> 12-17, 2 -> 18-24, etc
crime <- crime %>%
  mutate(age = recode(age,
                      "1" = "12-17",
                      "2" = "18-24",
                      "3" = "25-34",
                      "4" = "35-49",
                      "5" = "50-64",
                      "6" = "65 or older"))
#continue recodes with other columns
crime <- crime %>%
  mutate(sex = recode(sex,
                      "1" = "Male",
                      "2" = "Female"))
crime <- crime %>%
  mutate(hispanic = recode(hispanic,
                      "1" = "Hispanic",
                      "2" = "Non-Hispanic",
                      "98" = "Residue"))
```

Just the Stats

Goal: write short, informative statistical publications on NCVS data

Topics:

- Percentage of crime reported to police by urbanicity
- Victim and Offender characteristics by age group



Just the Stats: Urbanicity

- Analyzed whether the percentage of victimizations reported to police varied by urbanicity (urban, suburban, and rural) and crime type.
- Includes information on both violent (rape/sexual assault, robbery, aggravated assault, and simple assault) and property crime types (burglary/trespassing, motor vehicle theft, and other household theft).
- Final report to be submitted to BJS leadership for publication approval to post the figure and associated text on the BJS website.

Just the Stats: Victim/Offender Characteristics

- Examined the percentage of violent incidents by victim and offender age.
- The age group's share of the population was compared to that group's share of violent incidents, as both victims and offenders.
- Final report to be submitted to BJS leadership for publication approval to post the figure and associated text on the BJS website.

Final Thoughts/What I've Learned

- Learned how the workflow and office environment of a federal government agency works
- Gained more insight into the purpose and ethical responsibilities of a federal statistical agency
- Gained experience in a data analytics role

Thank you!

Bureau of Justice Statistics

810 Seventh Street, NW, Washington, DC 20531 | Phone: +1 (202) 307-0765 | **bjs.ojp.gov**