

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Codebook for “Creating a UCR Utility”

Author(s): Michael D. Maltz ; Harald E. Weiss

Document No.: 215342

Date Received: August 2006

Award Number: 2004-IJ-CX-0083

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

CODEBOOK FOR “CREATING A UCR UTILITY”

Part of the Final Report to the National Institute of Justice

Prepared by

Michael D. Maltz
and
Harald E. Weiss

Criminal Justice Research Center
and
Department of Sociology
The Ohio State University

June, 23, 2006

Table of Contents

Running the Plotting Utility	1
Worksheet Definitions in State Data Files	3
State Abbreviations	4
Missingness Codes and Colors	5
Population Worksheet	6
Months Reported Worksheet	7
Covering ORIs Worksheet	8
Group Worksheet	9
SMSA Worksheet	10
Crime Category Worksheets	11
Revisions Worksheet	14

Running the Plotting Utility

Page 1

The charting utility is an Excel workbook, UCRPlot.xls. A few important points:

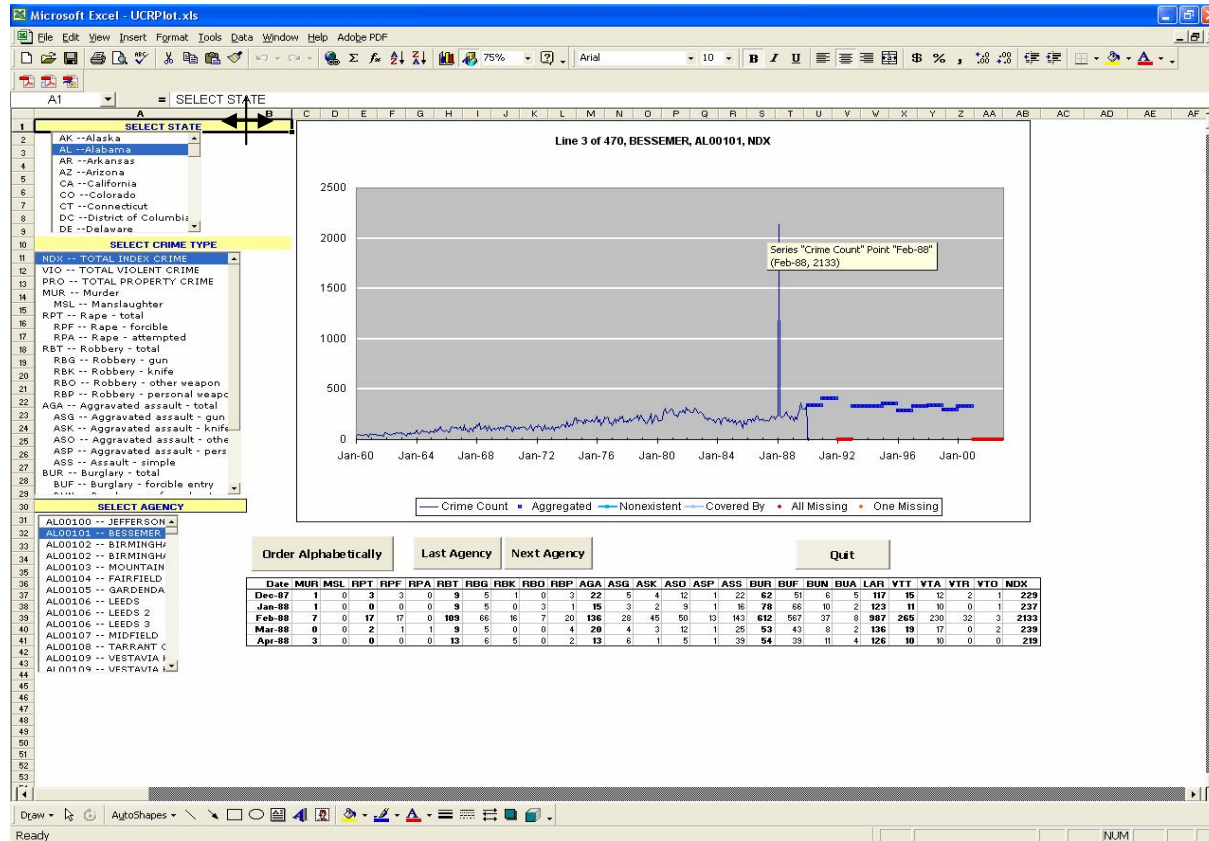
1. **Macros must be enabled.** To enable the macros, press Tools on the menu bar, then choose Macros, then Security, and set the security level to Medium. Then press the “Quit” button and reopen the program.
2. **UCRPlot.xls must be in the same directory as the state data.**
3. Different computers display the worksheet in different ways. You may want to adjust the figure to some extent. One easy way to do this is to change the width of column A. To change its width, move the cursor to the line separating columns A and B (at the top of the column, between the boxes that display “A” and “B”). A black cross will then appear, with arrowheads pointing left and right (see figure below). Hold the left button down and move the cursor either left or right to change the width of column A.

The user initially makes three choices, the state, the crime type, and the agency. Selection of a state brings up the list of agencies, arranged in order of their ORI identifier. The default crime type is Index crime and the default agency is the first listed agency. The agency listing can be changed from ordering by agency ID (default), which is essentially a geographical ordering, to alphabetical ordering, using the button immediately to the right of the agency list.

The next two buttons, “Last Agency” and “Next Agency,” permit the user to move through the list of agencies. The plot then shows the selected agency’s trajectory for the same crime type.

The next button, “Quit,” is self-explanatory; it closes whichever state file is open, as well as the charting program.

To get an indication of the crime count around a given point, the user can position the cursor directly on that point and left-click the mouse. The complete crime data for that date and two months before and after that date will now be displayed below the chart, with the Index crime counts in bold font. To show how this can be used, the user can select the agency Bessemer, AL (see figure below).



Worksheet Definitions in State Data Files

Page 3

Each state data file contains 83 worksheets, 78 (26 crime categories x 3 worksheets per category) containing crime count data and five with ancillary information. Each row in the data set corresponds to the data for a single agency, and each column consists of the data for a given time period.

Three worksheets are needed for listing the crime counts because each worksheet is limited to containing 256 columns, and there are 516 months of data between 1960 and 2002.

Some explanation is needed as to why we used Excel and why we need three worksheets to list the crime data. We had considered Microsoft Access as an alternative database instead of Excel, but it has a limit of 256 columns (as does Excel). Moreover, we felt that more people were familiar (and comfortable) with Excel than would be the case with Access or another database program.

We had also considered transposing the data set, so that columns held the monthly data (the limit on columns is 256 x 256, or 65536). The problem with this alternative is that agencies have very different numbers of agencies that report to the FBI (Pennsylvania tops the list at over 1500), so states would have different numbers of columns (and therefore worksheets). Moreover, there is considerable room for expansion, since the each worksheet for a crime category can contain twenty years of monthly data.

The tab labels for each of the crime category worksheets are listed on Pp. 10-11 and the contents of each worksheet are described on subsequent pages.

On all worksheets except the "Revisions" worksheet the first column (A) of each worksheet contains the agency identifier code (its ORI designation), and the first row of each worksheet describes the data in that column. The ORI of an agency situated in more than one county has multiple rows to contain its data; the ORI listing in the first column is colored yellow to indicate this circumstance, and the crime data are prorated by the population in each county. This is done to facilitate the calculation of county-level crime rates.

State Abbreviation: AK = Alaska
AZ = Arizona
AR = Arkansas
AS = American Samoa
AL = Alabama
CA = California
CO = Colorado
CT = Connecticut
CZ = Canal Zone
DC = Washington DC
DE = Delaware
FL = Florida
GA = Georgia
GU = Guam
HI = Hawaii
IA = Iowa
ID = Idaho
IL = Illinois
IN = Indiana
KS = Kansas
LA = Louisiana
MA = Massachusetts
MD = Maryland
ME = Maine
MI = Michigan
MN = Minnesota
MO = Missouri
MS = Mississippi
MT = Montana
NB = Nebraska
NC = North Carolina
ND = North Dakota
NH = New Hampshire
NJ = New Jersey
NM = New Mexico
NV = Nevada
NY = New York
OH = Ohio
OK = Oklahoma
OR = Oregon
PA = Pennsylvania
PR = Puerto Rico
RI = Rhode Island
SC = South Carolina
SD = South Dakota
TN = Tennessee
TX = Texas
UT = Utah
VA = Virginia
VI = Virgin Islands
VT = Vermont
WA = Washington State
WI = Wisconsin
WV = West Virginia
WY = Wyoming

Code

Value

Color

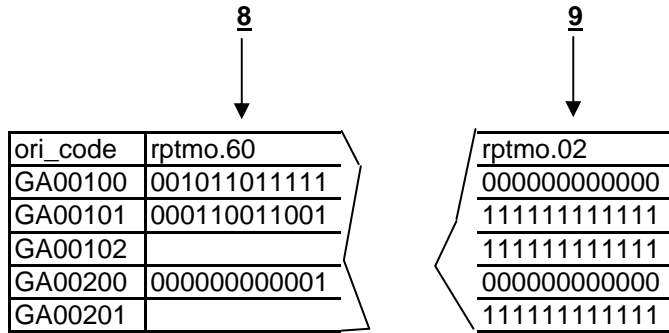
agency did not exist during this period	-80	blue (8)
ORI is covered by another agency	-85	blue(17)
we assign missing and record its value	-90	orange (45)
murder missing	-91	red (3)
rape missing	-92	red (3)
robbery missing	-93	red (3)
assault missing	-94	red (3)
burglary missing	-95	red (3)
larceny missing	-96	red (3)
motor vehicle theft missing	-97	red (3)
on CI page, more than 1 crime missing	-98	red (3)
no data for all crimes, for this month (true missing)	-99	red (3)
aggregated to February	-102	green (4)
aggregated to March	-103	green (4)
aggregated to April	-104	green (4)
aggregated to May	-105	green (4)
aggregated to June	-106	green (4)
aggregated to July	-107	green (4)
aggregated to August	-108	green (4)
aggregated to September	-109	green (4)
aggregated to October	-110	green (4)
aggregated to November	-111	green (4)
aggregated to December	-112	green (4)

POPULATION Tab

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
ori_code	County	No.Cnties	NAME	HISTORY	pop1.60	pop1.02
GA00100	1	1	APPLING	11	8978	13628
GA00101	1	1	BAXLEY	11	4268	4339
GA00102	1	1	SURRENCY	1		247
GA00200	2	1	ATKINSON	11	6188	4571
GA00201	2	1	PEARSON	11		1887

- 1** The **ORI** code is the **OR**iginating Agency Identifier: a seven-character string in the form of STccaa. The first two characters represent the state (see the two letter code on the abbreviations page); the next three characters are usually numbers that refer to the FBI sequence number for the county, starting with 001 (in some cases, however, they do not refer to county and in other cases they may be letters); and the next two characters usually correspond to a sequence number for the agency within the county (in some cases these two characters may be letters, like SP to designate the State Police barracks in that county). For example, the Ohio State University Police Department's ORI is OH02527, since it is in Franklin County whose FBI sequence number is 25, and is the 27th agency in that county.
- 2** FBI sequence number for the county
- 3** Number of counties the agency has jurisdiction over
- 4** Name of Agency
- 5** Annual reporting history of the agency. A value of 1 means that the agency reported, a blank field means that it did not. GA00102, for example, only reported 1 year between 1960 and 2002.
- 6** Population size for the year 1960 for the specific county (must be used in conjunction with field #2)
- 7** Population size for the year 2002 for the specific county (must be used in conjunction with field #2)

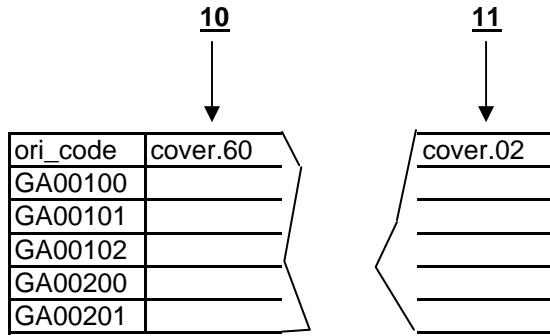
MONTHS REPORTED Tab



8 Months reported in 1960. A value of 1 means the agency reported for a month, a value of 0 means it did not. GA00100 reported March, May, June, August, September, October, November, and December of 1960

9 Months reported in 1960. A value of 1 means the agency reported for a month, a value of 0 means it did not. GA00100 did not report any crime data in 2002.

COVERING ORIs Tab



- 10 Coverage Data for 1960. If data of this agency was reported through a different agency the ORI code of the covering agency is recorded here.

- 11 Coverage Data for 2002. If data of this agency was reported through a different agency the ORI code of the covering agency is recorded here.

GROUP Tab

<u>12</u>		<u>13</u>	
ori_code	grp60		grp02
GA00100	8D		8C
GA00101	6		6
GA00102			7
GA00200	8D		8D
GA00201			7

12 and 13

The group the law enforcement agency is in denotes the type of area the agency have jurisdiction over. 0 represents U.S. possessions, 1-7 denotes cities, and 8-9 represent counties.

- 0 = Possessions (Puerto Rico, Guam, Canal Zone, Virgin Islands, and American Samoa)
- 1A= Cities 1,000,000 or over
- 1B= Cities from 500,000 thru 999,999
- 1C= Cities from 250,000 thru 499,999
- 2 = Cities from 100,000 thru 249,000
- 3 = Cities from 50,000 thru 99,000
- 4 = Cities from 25,000 thru 49,999
- 5 = Cities from 10,000 thru 24,999
- 6 = Cities from 2,500 thru 9,999
- 7 = Cities under 2,500
- 8 = Non-MSA Counties:
 - 8A= Non-MSA counties 100,000 or over
 - 8B= Non-MSA counties from 25,000 thru 99,999
 - 8C= Non-MSA counties from 10,000 thru 24,999
 - 8D= Non-MSA counties under 10,000
 - 8E= Non-MSA State Police
- 9 = MSA Counties:
 - 9A= MSA counties 100,000 or over
 - 9B= MSA counties from 25,000 thru 99,999
 - 9C= MSA counties from 10,000 thru 24,999
 - 9D= MSA counties under 10,000
 - 9E= MSA State Police

SMSA Tab

<u>14</u>	
ori_code	smsa1.60
GA00100	
GA00101	
GA00102	
GA00200	
GA00201	

<u>15</u>	
	smsa1.02

14 and 15

The SMSA columns refer to the location of the law enforcement agency in a certain metropolitan statistical area.

CRIME Tabs

<u>16</u> ↓		<u>17</u> ↓	
ori_code	mur69.01	mur79.12	
GA00100	-99		1
GA00101	-99		0
GA00202	-80		-99
GA00301	-99		0
GA00400	-105		0

Describe the number of murders that were reported by the specific agencies in a specific month. The first three alphanumeric symbols refer to specific crime categories (see listing below). The first two numbers denote the year (arrow 16 is 1969), the last two numbers denote the month (arrow 16 is January, arrow 17 is December). The codes below are the most commonly found ones. Other codes are outlined on Page 1.

16 and 17

-99
-80
-105

data is missing

agency did not exist in this year/month yet.

aggregated crime, the agency summed the number of offenses for a number of months and then reported the overall number for this time period. The last two numbers of the -1## code denote the number of months that were aggregated. GA00400 had aggregated 5 months with January 1969 being part of this aggregation. The non -1## code that is highlighted green represents the aggregated number that was reported by the agency.

Codes for Crime Tabs

- MUR1 = Murder, January 1960 - December 1979
- MUR2 = Murder, January 1980 - December 1999
- MUR3 = Murder, January 2000 - December 2002 (to be expanded)
- MSL1 = Manslaughter, January 1960 - December 1980
- MSL2 = Manslaughter, January 1980 - December 2000

MSL3 = Manslaughter, January 2000 - December 2002 (to be expanded)
RPT1 = Rape: Total, January 1960 - December 1981
RPT2 = Rape: Total, January 1980 - December 2001
RPT3 = Rape: Total, January 2000 - December 2002 (to be expanded)
RPF1 = Rape: Forcible, January 1960 - December 1982
RPF2 = Rape: Forcible, January 1980 - December 2002
RPF3 = Rape: Forcible, January 2000 - December 2002 (to be expanded)
RPA1 = Rape: Attempted, January 1960 - December 1983
RPA2 = Rape: Attempted, January 1980 - December 2003
RPA3 = Rape: Attempted, January 2000 - December 2002 (to be expanded)
RBT1 = Robbery: Total, January 1960 - December 1984
RBT2 = Robbery: Total, January 1980 - December 2004
RBT3 = Robbery: Total, January 2000 - December 2002 (to be expanded)
RBG1 = Robbery: Gun, January 1960 - December 1985
RBG2 = Robbery: Gun, January 1980 - December 2005
RBG3 = Robbery: Gun, January 2000 - December 2002 (to be expanded)
RBK1 = Robbery: Knife, January 1960 - December 1986
RBK2 = Robbery: Knife, January 1980 - December 2006
RBK3 = Robbery: Knife, January 2000 - December 2002 (to be expanded)
RBO1 = Robbery: Miscellaneous, January 1960 - December 1987
RBO2 = Robbery: Miscellaneous, January 1980 - December 2007
RBO3 = Robbery: Miscellaneous, January 2000 - December 2002 (to be expanded)
RBP1 = Robbery: Strongarm / Unarmed, January 1960 - December 1988
RBP2 = Robbery: Strongarm / Unarmed, January 1980 - December 2008
RBP3 = Robbery: Strongarm / Unarmed, January 2000 - December 2002 (to be expanded)
AGA1 = Assault: Aggravated, January 1960 - December 1989
AGA2 = Assault: Aggravated, January 1980 - December 2009
AGA3 = Assault: Aggravated, January 2000 - December 2002 (to be expanded)
ASG1 = Assault: Gun, January 1960 - December 1990
ASG2 = Assault: Gun, January 1980 - December 2010
ASG3 = Assault: Gun, January 2000 - December 2002 (to be expanded)
ASK1 = Assault: Knife, January 1960 - December 1991
ASK2 = Assault: Knife, January 1980 - December 2011
ASK3 = Assault: Knife, January 2000 - December 2002 (to be expanded)
ASO1 = Assault: Other, January 1960 - December 1992
ASO2 = Assault: Other, January 1980 - December 2012
ASO3 = Assault: Other, January 2000 - December 2002 (to be expanded)

ASP1 = Assault: Unarmed, January 1960 - December 1993
ASP2 = Assault: Unarmed, January 1980 - December 2013
ASP3 = Assault: Unarmed, January 2000 - December 2002 (to be expanded)
ASS1 = Assault: Simple, January 1960 - December 1994
ASS2 = Assault: Simple, January 1980 - December 2014
ASS3 = Assault: Simple, January 2000 - December 2002 (to be expanded)
BUR1 = Burglary: Total, January 1960 - December 1995
BUR2 = Burglary: Total, January 1980 - December 2015
BUR3 = Burglary: Total, January 2000 - December 2002 (to be expanded)
BUF1 = Burglary: Forced entry, January 1960 - December 1996
BUF2 = Burglary: Forced entry, January 1980 - December 2016
BUF3 = Burglary: Forced entry, January 2000 - December 2002 (to be expanded)
BUN1 = Burglary: No forced entry, January 1960 - December 1997
BUN2 = Burglary: No forced entry, January 1980 - December 2017
BUN3 = Burglary: No forced entry, January 2000 - December 2002 (to be expanded)
BUA1 = Burglary: Attempted, January 1960 - December 1998
BUA2 = Burglary: Attempted, January 1980 - December 2018
BUA3 = Burglary: Attempted, January 2000 - December 2002 (to be expanded)
LAR1 = Larceny, January 1960 - December 1999
LAR2 = Larceny, January 1980 - December 2019
LAR3 = Larceny, January 2000 - December 2002 (to be expanded)
VTT1 = Vehicle theft: Total, January 1960 - December 2000
VTT2 = Vehicle theft: Total, January 1980 - December 2020
VTT3 = Vehicle theft: Total, January 2000 - December 2002 (to be expanded)
VTA1 = Vehicle theft: Car, January 1960 - December 2001
VTA2 = Vehicle theft: Car, January 1980 - December 2021
VTA3 = Vehicle theft: Car, January 2000 - December 2002 (to be expanded)
VTR1 = Vehicle theft: Truck, January 1960 - December 2002
VTR2 = Vehicle theft: Truck, January 1980 - December 2022
VTR3 = Vehicle theft: Truck, January 2000 - December 2002 (to be expanded)
VTO1 = Vehicle theft: other, January 1960 - December 2003
VTO2 = Vehicle theft: other, January 1980 - December 2023
VTO3 = Vehicle theft: other, January 2000 - December 2002 (to be expanded)
NDX1 = Crime Index, January 1960 - December 2004
NDX2 = Crime Index, January 1980 - December 2024
NDX3 = Crime Index, January 2000 - December 2002 (to be expanded)

Revisions Tab

Page 14

Two sets of corrections to the data are listed on the “Revisions” tab page. The corrections listed in columns A-F represent one stage in the cleaning process, based on the appearance of zeroes in the crime data. The corrections listed in columns H-AJ represent a second stage.

Stage 1:

Three different types of data issues are listed in these columns. The first type is described as “No update indicated, but contains data.” This represents the fact that, for that ORI for that month, the “Date updated” field in the original data set was empty; this normally indicates that no data had been received for that month. The data set itself, however, did contain data for that month. We accepted the data as being valid, but point out the anomaly.

The second type of issue is the most common one. Between the two listed dates we changed a zero entry to -99 (data completely missing), -80 (ORI did not exist at the time), or some other value indicating a type of missingness – see the listing of missingness codes and colors on page XX.

The third type calls attention to a data point that we did not change, but that we feel may be incorrect. For the most part, these are unusually high crime counts (spikes in crime). In some cases they are explainable and due to known events (e.g., the homicide count in Oklahoma City OK in April 1995, due to the attack on the Alfred E. Murrah building), but in other cases they may be incorrect entries. Since we are generally not familiar with the local contexts of these crime counts, we merely noted their existence but made no changes in the data.

Stage 2:

These corrections represent the cases in which a string of zeroes was followed by a spike in the data, leading us to believe that the data for a number of consecutive months was aggregated and reported in a single month. These corrections contain a list of all of the crime data that was subsequently assumed to represent multiple months.