Introduction to R

Greg Ridgeway (gridge@upenn.edu)
Ruth Moyer (moyruth@upenn.edu)

October 07, 2018

PDF YAML header Ctrl-Shift-C to comment/uncomment

Introduction

Published by the FBI, the Uniform Crime Reports (UCR) is a yearly collection of official data on crime in the United States. Approximately 18,000 individual city, university and college, county, state, tribal, and federal law enforcement agencies report data to the FBI. Some agencies are more consistent than other agencies with reporting their data to the FBI. For example, you may find a handful of very rural counties that have no data in, e.g,. 2012, but have data in 2014.

Earlier, we worked with one dataset from the UCR, Law Enforcement officers Killed and Assaulted (LEOKA). Now, we're working with a different portion of the UCR, "Offenses Known and Clearances by Arrest." The University of Michigan converts this data into a format that is easy-to-use in R. Download the 2012 UCR data to your computer. As of October 2018, the Offenses Known and Clearances by Arrest ("OKCA") Data is available for up to 2017. We have called our subfolder `UCR2012` and here are the files in our subfolder. Yours should look the same.

```
list.files("UCR2012/", recursive=TRUE)
```

```
[1] "35021-0001-Setup.sas"             "35021-descriptioncitation.pdf"
[3] "35021-manifest.txt"               "35021-related_literature.txt"
[5] "DS0001/35021-0001-Codebook.pdf"   "DS0001/35021-0001-Data.rda"
[7] "factor_to_numeric_icpsr.R"        "series-57-related_literature.txt"
[9] "TermsOfUse.html"
```

For each year, there will be a zip file to download. There are two main files within the zip file that we will be using: (1) a .rda file (which contains your data in R format) and (2) the codebook (which is in PDF format).

The data file has monthly data on the number of Crime Index offenses reported and the number of offenses cleared by arrest or other means. Thus, the data isn't "incident-level" data - in contrast to the Chicago data that we've worked with. The data instead includes crime counts for each agency. You can get helpful summaries about the UCR OKCA data here and here.

Loading the UCR data

In this section, we are going to work with the 2012 UCR data. The first step when working with this UCR data is loading it into R. As with loading any data, it's very important that your path is correctly set using `setwd()`. Set it to wherever you have unzipped your 2012 UCR data. Again,

since our subfolder is called `UCR2012` we use the following `load()` to import the data into R. If you called your folder something other than `UCR2012` then you will have to edit this `load()` function to match your subfolder's name.

```
load("UCR2012/DS0001/35021-0001-Data.rda")
```

Let's check what R now has in its memory.

```
ls()
```

```
[1] "da35021.0001"
```

Now we see that a new object is in R's memory, `da35021.0001`. We can ask R what this is...

```
is(da35021.0001)
```

```
[1] "data.frame" "list"       "oldClass"   "vector"
```

...and, sure enough, it is a new R data frame, presumably containing the 2012 UCR data. Let's take a look inside the dataset to see what we have.

Let's look at the first two rows. Each row describes the features of one law enforcement agency. You will see that there are a lot of features recorded in the data, 1448 features to be precise. Unfortunately, the variable names are not very helpful. Several you can figure out by looking at the data within the column. Clearly, `V2` indicates the state where the law enforcement agency is located. `V10` looks like some kind of date. `V26` gives the name of the agency. But most of the rest are rather mysterious.

```
da35021.0001[1:2,]
```

```
                  V1           V2      V3 V4                             V5
1 (1) Offenses known (01) Alabama ALAST00 8D (6) East South Central States
2 (1) Offenses known (01) Alabama ALDI003 7  (6) East South Central States
    V6    V7 V8      V9      V10  V11                     V12
1 2012 70402  N          6231986 3490 (00) No months reported
2 2012 49797  N         10142009 3490  (12) Dec last reported
                     V13 V14 V15 V16 V17 V18 V19 V20 V21 V22 V23
1 (1) All other agencies   0  NA  NA   0  NA  NA   0  NA  NA   Y
2 (1) All other agencies   0  NA  NA   0  NA  NA   0  NA  NA   Y
                                     V24 V25                       V26
1 (9) Agency-contributor, not on mail list   N ALABAMA HIGHWAY PATROL
2 (9) Agency-contributor, not on mail list   N POARCH CREEK TRIBAL
     V27                             V28                            V29
1 ALA    CHIEF OF HIGHWAY PATROL        AL HIGHWAY PARTOL HEADQUARTERS
2 ALA    CHIEF OF POLICE                POARCH CREEK TRIBAL PD
                              V30                              V31   V32
1 500 DEXTER AVE                  MONTGOMERY, AL                  36130
2 5811 JACK SPRINGS RD            ATMORE, AL                      36502
                        V33     V34               V35               V36
1                      <NA>      NA              <NA>              <NA>
2 (00) Jan not w oth month 4182013 (5) Normal return (5) Normal return
              V37             V38             V39 V40 V41 V42 V43 V44 V45
1            <NA>            <NA>            <NA>                  0   0
```

```
2 (0) Not updated (0) Not updated (0) Not updated   P   P           0   0
  V46 V47 V48 V49 V50 V51 V52 V53 V54 V55 V56 V57 V58 V59 V60 V61 V62 V63
1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
2   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
  V64 V65 V66 V67 V68 V69 V70 V71 V72 V73 V74 V75 V76 V77 V78 V79 V80 V81
1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
2   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0
  V82 V83 V84 V85 V86 V87 V88 V89 V90 V91 V92 V93 V94 V95 V96 V97 V98 V99
1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
2   0   1   0   0   5   1   4   0  15   1   1   0   0  22   0   0   0   0
  V100 V101 V102 V103 V104 V105 V106 V107 V108 V109 V110 V111 V112 V113
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V114 V115 V116 V117 V118 V119 V120 V121 V122 V123 V124 V125 V126 V127
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V128 V129 V130 V131 V132 V133 V134 V135 V136 V137 V138 V139 V140 V141
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V142 V143 V144 V145 V146 V147 V148 V149 V150                     V151
1    0    0    0    0    0    0    0    0    0                     <NA>
2    0    0    0    0    0    0    0    0    0 (00) Feb not w oth month
     V152              V153              V154            V155
1      NA              <NA>              <NA>            <NA>
2 4182013 (5) Normal return (5) Normal return (0) Not updated
            V156            V157 V158 V159 V160 V161 V162 V163 V164 V165
1           <NA>            <NA>                      0    0    0    0
2 (0) Not updated (0) Not updated    P    P           0    0    0    0
  V166 V167 V168 V169 V170 V171 V172 V173 V174 V175 V176 V177 V178 V179
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V180 V181 V182 V183 V184 V185 V186 V187 V188 V189 V190 V191 V192 V193
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V194 V195 V196 V197 V198 V199 V200 V201 V202 V203 V204 V205 V206 V207
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    1    0    0    0    0    1    2    1    1    0
  V208 V209 V210 V211 V212 V213 V214 V215 V216 V217 V218 V219 V220 V221
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2   15    1    1    0    0   19    0    0    0    0    0    0    0    0
  V222 V223 V224 V225 V226 V227 V228 V229 V230 V231 V232 V233 V234 V235
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V236 V237 V238 V239 V240 V241 V242 V243 V244 V245 V246 V247 V248 V249
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V250 V251 V252 V253 V254 V255 V256 V257 V258 V259 V260 V261 V262 V263
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
```

```
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V264 V265 V266 V267 V268                    V269    V270
1    0    0    0    0    0                    <NA>      NA
2    0    0    0    0    0 (00) Mar not w oth month 4182013
               V271              V272            V273            V274
1              <NA>              <NA>            <NA>            <NA>
2 (5) Normal return (5) Normal return (0) Not updated (0) Not updated
            V275 V276 V277 V278 V279 V280 V281 V282 V283 V284 V285 V286
1           <NA>                      0    0    0    0    0    0    0
2 (0) Not updated    P    P           0    0    0    0    0    0    0
  V287 V288 V289 V290 V291 V292 V293 V294 V295 V296 V297 V298 V299 V300
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V301 V302 V303 V304 V305 V306 V307 V308 V309 V310 V311 V312 V313 V314
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V315 V316 V317 V318 V319 V320 V321 V322 V323 V324 V325 V326 V327 V328
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    5    0    5    0   14    0    0
  V329 V330 V331 V332 V333 V334 V335 V336 V337 V338 V339 V340 V341 V342
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0   19    0    0    0    0    0    0    0    0    0    0    0
  V343 V344 V345 V346 V347 V348 V349 V350 V351 V352 V353 V354 V355 V356
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V357 V358 V359 V360 V361 V362 V363 V364 V365 V366 V367 V368 V369 V370
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V371 V372 V373 V374 V375 V376 V377 V378 V379 V380 V381 V382 V383 V384
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V385 V386                    V387    V388              V389
1    0    0                    <NA>      NA              <NA>
2    0    0 (00) Apr not w oth month 4182013 (5) Normal return
               V390            V391            V392            V393 V394
1              <NA>            <NA>            <NA>            <NA>
2 (5) Normal return (0) Not updated (0) Not updated (0) Not updated    P
  V395 V396 V397 V398 V399 V400 V401 V402 V403 V404 V405 V406 V407 V408
1                   0    0    0    0    0    0    0    0    0    0    0
2    P              0    0    0    0    0    0    0    0    0    0    0
  V409 V410 V411 V412 V413 V414 V415 V416 V417 V418 V419 V420 V421 V422
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V423 V424 V425 V426 V427 V428 V429 V430 V431 V432 V433 V434 V435 V436
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V437 V438 V439 V440 V441 V442 V443 V444 V445 V446 V447 V448 V449 V450
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
```

```
2    0    0    0    3    1    2    0   10    2    2    0    0   15    0
  V451 V452 V453 V454 V455 V456 V457 V458 V459 V460 V461 V462 V463 V464
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V465 V466 V467 V468 V469 V470 V471 V472 V473 V474 V475 V476 V477 V478
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V479 V480 V481 V482 V483 V484 V485 V486 V487 V488 V489 V490 V491 V492
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V493 V494 V495 V496 V497 V498 V499 V500 V501 V502 V503 V504
1    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0
                     V505    V506              V507              V508
1                    <NA>      NA              <NA>              <NA>
2 (00) May not w oth month 4182013 (5) Normal return (5) Normal return
            V509            V510            V511 V512 V513 V514 V515 V516
1           <NA>            <NA>            <NA>                        0
2 (0) Not updated (0) Not updated (0) Not updated    P    P              0
  V517 V518 V519 V520 V521 V522 V523 V524 V525 V526 V527 V528 V529 V530
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V531 V532 V533 V534 V535 V536 V537 V538 V539 V540 V541 V542 V543 V544
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V545 V546 V547 V548 V549 V550 V551 V552 V553 V554 V555 V556 V557 V558
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    1    0    0    0    1    0    0    0    0    0    0    5
  V559 V560 V561 V562 V563 V564 V565 V566 V567 V568 V569 V570 V571 V572
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    5    0    0   16    1    1    0    0   23    0    0    0    0    0
  V573 V574 V575 V576 V577 V578 V579 V580 V581 V582 V583 V584 V585 V586
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V587 V588 V589 V590 V591 V592 V593 V594 V595 V596 V597 V598 V599 V600
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V601 V602 V603 V604 V605 V606 V607 V608 V609 V610 V611 V612 V613 V614
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V615 V616 V617 V618 V619 V620 V621 V622                     V623
1    0    0    0    0    0    0    0    0                     <NA>
2    0    0    0    0    0    0    0    0 (00) June not w oth month
     V624              V625              V626            V627
1      NA              <NA>              <NA>            <NA>
2 4182013 (5) Normal return (5) Normal return (0) Not updated
            V628            V629 V630 V631 V632 V633 V634 V635 V636 V637
1           <NA>            <NA>                        0    0    0    0
```

```
2 (0) Not updated (0) Not updated    P    P              0    0    0    0
  V638 V639 V640 V641 V642 V643 V644 V645 V646 V647 V648 V649 V650 V651
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V652 V653 V654 V655 V656 V657 V658 V659 V660 V661 V662 V663 V664 V665
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V666 V667 V668 V669 V670 V671 V672 V673 V674 V675 V676 V677 V678 V679
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    3    0    0    0    0    3   13    1   12    0
  V680 V681 V682 V683 V684 V685 V686 V687 V688 V689 V690 V691 V692 V693
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2   19    0    0    0    0   35    0    0    0    0    0    0    0    0
  V694 V695 V696 V697 V698 V699 V700 V701 V702 V703 V704 V705 V706 V707
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V708 V709 V710 V711 V712 V713 V714 V715 V716 V717 V718 V719 V720 V721
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V722 V723 V724 V725 V726 V727 V728 V729 V730 V731 V732 V733 V734 V735
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V736 V737 V738 V739 V740                      V741    V742
1    0    0    0    0    0                      <NA>      NA
2    0    0    0    0    0 (00) July not w oth month 4182013
              V743              V744             V745            V746
1             <NA>              <NA>             <NA>            <NA>
2 (5) Normal return (5) Normal return (0) Not updated (0) Not updated
            V747 V748 V749 V750 V751 V752 V753 V754 V755 V756 V757 V758
1           <NA>                      0    0    0    0    0    0    0
2 (0) Not updated    P    P           0    0    0    0    0    0    0
  V759 V760 V761 V762 V763 V764 V765 V766 V767 V768 V769 V770 V771 V772
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V773 V774 V775 V776 V777 V778 V779 V780 V781 V782 V783 V784 V785 V786
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    1    1    0    0
  V787 V788 V789 V790 V791 V792 V793 V794 V795 V796 V797 V798 V799 V800
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0   16    3   13    0   16    1    1
  V801 V802 V803 V804 V805 V806 V807 V808 V809 V810 V811 V812 V813 V814
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0   34    0    0    0    0    0    0    0    0    0    0    0
  V815 V816 V817 V818 V819 V820 V821 V822 V823 V824 V825 V826 V827 V828
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V829 V830 V831 V832 V833 V834 V835 V836 V837 V838 V839 V840 V841 V842
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
```

```
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V843 V844 V845 V846 V847 V848 V849 V850 V851 V852 V853 V854 V855 V856
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V857 V858                     V859    V860              V861
1    0    0                     <NA>      NA              <NA>
2    0    0 (00) Aug not w oth month 4182013 (5) Normal return
               V862            V863            V864            V865 V866
1              <NA>            <NA>            <NA>            <NA>
2 (5) Normal return (0) Not updated (0) Not updated (0) Not updated    P
  V867 V868 V869 V870 V871 V872 V873 V874 V875 V876 V877 V878 V879 V880
1                   0    0    0    0    0    0    0    0    0    0    0
2    P              0    0    0    0    0    0    0    0    0    0    0
  V881 V882 V883 V884 V885 V886 V887 V888 V889 V890 V891 V892 V893 V894
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V895 V896 V897 V898 V899 V900 V901 V902 V903 V904 V905 V906 V907 V908
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    1    1    0    0    0    1    0    0
  V909 V910 V911 V912 V913 V914 V915 V916 V917 V918 V919 V920 V921 V922
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    1    9    2    7    0   21    1    0    0    1   33    0
  V923 V924 V925 V926 V927 V928 V929 V930 V931 V932 V933 V934 V935 V936
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V937 V938 V939 V940 V941 V942 V943 V944 V945 V946 V947 V948 V949 V950
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V951 V952 V953 V954 V955 V956 V957 V958 V959 V960 V961 V962 V963 V964
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0    0    0
  V965 V966 V967 V968 V969 V970 V971 V972 V973 V974 V975 V976
1    0    0    0    0    0    0    0    0    0    0    0    0
2    0    0    0    0    0    0    0    0    0    0    0    0
                    V977    V978              V979              V980
1                   <NA>      NA              <NA>              <NA>
2 (00) Sep not w oth month 4182013 (5) Normal return (5) Normal return
            V981            V982            V983 V984 V985 V986 V987 V988
1           <NA>            <NA>            <NA>                      0
2 (0) Not updated (0) Not updated (0) Not updated    P    P              0
  V989 V990 V991 V992 V993 V994 V995 V996 V997 V998 V999 V1000 V1001 V1002
1    0    0    0    0    0    0    0    0    0    0    0     0     0     0
2    0    0    0    0    0    0    0    0    0    0    0     0     0     0
  V1003 V1004 V1005 V1006 V1007 V1008 V1009 V1010 V1011 V1012 V1013 V1014
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1015 V1016 V1017 V1018 V1019 V1020 V1021 V1022 V1023 V1024 V1025 V1026
1     0     0     0     0     0     0     0     0     0     0     0     0
```

```
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1027 V1028 V1029 V1030 V1031 V1032 V1033 V1034 V1035 V1036 V1037 V1038
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     9     2     7     0    22     1     1     0     0
  V1039 V1040 V1041 V1042 V1043 V1044 V1045 V1046 V1047 V1048 V1049 V1050
1     0     0     0     0     0     0     0     0     0     0     0     0
2    32     0     0     0     0     0     0     0     0     0     0     0
  V1051 V1052 V1053 V1054 V1055 V1056 V1057 V1058 V1059 V1060 V1061 V1062
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1063 V1064 V1065 V1066 V1067 V1068 V1069 V1070 V1071 V1072 V1073 V1074
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1075 V1076 V1077 V1078 V1079 V1080 V1081 V1082 V1083 V1084 V1085 V1086
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1087 V1088 V1089 V1090 V1091 V1092 V1093 V1094                    V1095
1     0     0     0     0     0     0     0     0                     <NA>
2     0     0     0     0     0     0     0     0 (00) Oct not w oth month
    V1096             V1097             V1098           V1099
1      NA              <NA>              <NA>            <NA>
2 4182013 (5) Normal return (5) Normal return (0) Not updated
           V1100           V1101 V1102 V1103 V1104 V1105 V1106 V1107
1            <NA>            <NA>                              0     0
2 (0) Not updated (0) Not updated     P     P                  0     0
  V1108 V1109 V1110 V1111 V1112 V1113 V1114 V1115 V1116 V1117 V1118 V1119
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1120 V1121 V1122 V1123 V1124 V1125 V1126 V1127 V1128 V1129 V1130 V1131
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1132 V1133 V1134 V1135 V1136 V1137 V1138 V1139 V1140 V1141 V1142 V1143
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     1     0
  V1144 V1145 V1146 V1147 V1148 V1149 V1150 V1151 V1152 V1153 V1154 V1155
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     1     4     2     2     0    16     0     0     0
  V1156 V1157 V1158 V1159 V1160 V1161 V1162 V1163 V1164 V1165 V1166 V1167
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0    21     0     0     0     0     0     0     0     0     0     0
  V1168 V1169 V1170 V1171 V1172 V1173 V1174 V1175 V1176 V1177 V1178 V1179
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1180 V1181 V1182 V1183 V1184 V1185 V1186 V1187 V1188 V1189 V1190 V1191
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1192 V1193 V1194 V1195 V1196 V1197 V1198 V1199 V1200 V1201 V1202 V1203
1     0     0     0     0     0     0     0     0     0     0     0     0
```

```
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1204 V1205 V1206 V1207 V1208 V1209 V1210 V1211 V1212
1     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0
                     V1213   V1214              V1215              V1216
1                     <NA>      NA               <NA>               <NA>
2 (00) Nov not w oth month 4182013 (5) Normal return (5) Normal return
            V1217           V1218           V1219 V1220 V1221 V1222 V1223
1            <NA>            <NA>            <NA>
2 (0) Not updated (0) Not updated (0) Not updated     P     P
  V1224 V1225 V1226 V1227 V1228 V1229 V1230 V1231 V1232 V1233 V1234 V1235
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1236 V1237 V1238 V1239 V1240 V1241 V1242 V1243 V1244 V1245 V1246 V1247
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1248 V1249 V1250 V1251 V1252 V1253 V1254 V1255 V1256 V1257 V1258 V1259
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1260 V1261 V1262 V1263 V1264 V1265 V1266 V1267 V1268 V1269 V1270 V1271
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     3     0     3     0    27     3
  V1272 V1273 V1274 V1275 V1276 V1277 V1278 V1279 V1280 V1281 V1282 V1283
1     0     0     0     0     0     0     0     0     0     0     0     0
2     3     0     0    33     0     0     0     0     0     0     0     0
  V1284 V1285 V1286 V1287 V1288 V1289 V1290 V1291 V1292 V1293 V1294 V1295
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1296 V1297 V1298 V1299 V1300 V1301 V1302 V1303 V1304 V1305 V1306 V1307
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1308 V1309 V1310 V1311 V1312 V1313 V1314 V1315 V1316 V1317 V1318 V1319
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1320 V1321 V1322 V1323 V1324 V1325 V1326 V1327 V1328 V1329 V1330
1     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0
                     V1331   V1332              V1333              V1334
1                     <NA>      NA               <NA>               <NA>
2 (00) Dec not w oth month 4182013 (5) Normal return (5) Normal return
            V1335           V1336           V1337 V1338 V1339 V1340 V1341
1            <NA>            <NA>            <NA>
2 (0) Not updated (0) Not updated (0) Not updated     P     P
  V1342 V1343 V1344 V1345 V1346 V1347 V1348 V1349 V1350 V1351 V1352 V1353
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1354 V1355 V1356 V1357 V1358 V1359 V1360 V1361 V1362 V1363 V1364 V1365
1     0     0     0     0     0     0     0     0     0     0     0     0
```

```
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1366 V1367 V1368 V1369 V1370 V1371 V1372 V1373 V1374 V1375 V1376 V1377
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1378 V1379 V1380 V1381 V1382 V1383 V1384 V1385 V1386 V1387 V1388 V1389
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     3     0     3     0    16     1
  V1390 V1391 V1392 V1393 V1394 V1395 V1396 V1397 V1398 V1399 V1400 V1401
1     0     0     0     0     0     0     0     0     0     0     0     0
2     1     0     0    20     0     0     0     0     0     0     0     0
  V1402 V1403 V1404 V1405 V1406 V1407 V1408 V1409 V1410 V1411 V1412 V1413
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1414 V1415 V1416 V1417 V1418 V1419 V1420 V1421 V1422 V1423 V1424 V1425
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1426 V1427 V1428 V1429 V1430 V1431 V1432 V1433 V1434 V1435 V1436 V1437
1     0     0     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0     0     0
  V1438 V1439 V1440 V1441 V1442 V1443 V1444 V1445 V1446 V1447 V1448
1     0     0     0     0     0     0     0     0     0     0
2     0     0     0     0     0     0     0     0     0     0
```

We need to refer to the codebook. Open up the codebook and flip through the pages. Starting on page 11 you can see what each variable is and clicking on the variable name in the pdf will give you more details on the agency attribute.

It would be more helpful if we had better variable names in our dataset. Transcribing variable names piecemeal isn't the most efficient - or error-proof method. Fortunately, our data has variable definitions stored in the dataset's attributes. Attributes provide information about the data that isn't explicitly in the dataset itself. It's a bit like a "footnote" that elaborates about something within the main "text" (the data). As we can see, the first 32 variables in our data are about agency and reporting.

```
var.lookup <- attributes(da35021.0001)$variable.labels
var.lookup[1:32]
```

```
                      V1                          V2
               "ID CODE"       "NUMERIC STATE CODE"
                      V3                          V4
              "ORI CODE"            "GROUP NUMBER"
                      V5                          V6
              "DIVISION"                    "YEAR"
                      V7                          V8
  "CITY SEQUENCE NUMBER"     "CORE CITY INDICATION"
                      V9                         V10
       "COVERED BY CODE"             "LAST UPDATE"
                     V11                         V12
          "FIELD OFFICE" "NUMBER OF MONTHS REPORTED"
```

```
                    V13                     V14
          "AGENCY COUNT"          "POPULATION 1"
                    V15                     V16
              "COUNTY 1"                "MSA 1"
                    V17                     V18
          "POPULATION 2"              "COUNTY 2"
                    V19                     V20
                "MSA 2"          "POPULATION 3"
                    V21                     V22
              "COUNTY 3"                "MSA 3"
                    V23                     V24
 "FOLLOW-UP INDICATION"  "SPECIAL MAILING GROUP"
                    V25                     V26
"SPECIAL MAILING ADDRESS"           "AGENCY NAME"
                    V27                     V28
    "AGENCY STATE NAME"  "MAILING ADDRESS-LINE 1"
                    V29                     V30
 "MAILING ADDRESS-LINE 2"  "MAILING ADDRESS-LINE 3"
                    V31                     V32
 "MAILING ADDRESS-LINE 4"              "ZIP CODE"
```

Importantly, the variable `V12` indicates which months have been reported.

```
table(da35021.0001$V12)
```

```
  (00) No months reported    (01) Jan last reported
                     5482                        25
   (02) Feb last reported  (03) March last reported
                       42                        37
 (04) April last reported    (05) May last reported
                       30                        50
  (06) June last reported   (07) July last reported
                       35                        51
(08) August last reported    (09) Sep last reported
                       49                        63
   (10) Oct last reported    (11) Nov last reported
                       95                       228
   (12) Dec last reported
                    15862
```

As you can see, most agencies report in December - the end of the year. Also, you'll see a sizable amount (5,482) that don't report for any months.

Also, not all "large" agencies report in December.

```
with(subset(da35021.0001,V12!="(12) Dec last reported"),
     summary(V14,digits=10)) # population of jurisdiction
subset(da35021.0001,V14==468609)[,1:30]
subset(da35021.0001,
```

```
      (V12!="(12) Dec last reported") & (V14>50000))[,c("V27","V29","V14","V13","V12")]
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       0     393    2167    1500  468609
                     V1            V2      V3 V4
12567 (1) Offenses known (31) New York NY05106 1C
                             V5   V6   V7 V8      V9      V10  V11
12567 (2) Middle Atlantic States 2012 7212  N NY05101 12201985 3540
                          V12                         V13    V14 V15 V16
12567 (00) No months reported (0) US Park & State Police 468609  52 603
      V17 V18 V19 V20 V21 V22 V23
12567   0  NA  NA   0  NA  NA   N
                                          V24 V25
12567 (7) Agency-non-contributor, not sent form   N
                           V26    V27                               V28
12567 BROOKHAVEN               N Y    CHIEF OF POLICE
                                 V29                               V30
12567 BROOKHAVEN POLICE DEPARTMENT
        V27                              V29    V14
303   ALA    LEE COUNTY                       57712
4615  ILL    OAK PARK POLICE DEPARTMENT       52128
5517  IND    HENDRICKS COUNTY                 76655
5551  IND    KOSCIUSKO COUNTY                 58570
5687  IND    VANDERBURGH COUNTY               62676
6404  KANS   MANHATTAN POLICE DEPARTMENT      53802
7740  MD     GAITHERSBURG POLICE DEPARTMENT   61634
7743  MD     ROCKVILLE POLICE DEPARTMENT      63936
8089  MASS   MEDFORD POLICE DEPARTMENT        57243
8881  MICH   PONTIAC POLICE DEPARTMENT        59930
9622  MISS   LAMAR COUNTY                     50141
12385 N Y    CLAY TOWN POLICE DEPARTMENT      53677
12563 N Y    BABYLON TOWN POLICE DEPARTMENT  166161
12567 N Y    BROOKHAVEN POLICE DEPARTMENT    468609
12568 N Y    HUNTINGTON POLICE DEPARTMENT    191852
12569 N Y    ISLIP POLICE DEPARTMENT         335350
12611 N Y    TOMPKINS COUNTY                  62208
14554 OHIO   ELYRIA POLICE DEPARTMENT         54578
20004 UTAH   TAYLORSVILLE CITY POLICE DEPT    60575
20861 WASH   SNOHOMISH COUNTY                312856
                            V13                         V12
303       (1) All other agencies   (07) July last reported
4615      (1) All other agencies    (10) Oct last reported
5517      (1) All other agencies   (00) No months reported
5551      (1) All other agencies (08) August last reported
5687      (1) All other agencies   (00) No months reported
6404  (0) US Park & State Police   (00) No months reported
7740  (0) US Park & State Police   (00) No months reported
```

```
7743  (0) US Park & State Police   (00) No months reported
8089      (1) All other agencies   (00) No months reported
8881  (0) US Park & State Police   (00) No months reported
9622      (1) All other agencies    (11) Nov last reported
12385 (0) US Park & State Police   (00) No months reported
12563 (0) US Park & State Police   (00) No months reported
12567 (0) US Park & State Police   (00) No months reported
12568 (0) US Park & State Police   (00) No months reported
12569 (0) US Park & State Police   (00) No months reported
12611     (1) All other agencies   (00) No months reported
14554     (1) All other agencies  (03) March last reported
20004 (0) US Park & State Police   (00) No months reported
20861     (1) All other agencies  (04) April last reported
```

Additionally, make note of variables like "MONTH INCLUDED IN." They indicate whether this month was reported in another month. The following lines make a table for each of these values

```
var.names <- grep("MONTH INCLUDED IN",var.lookup,value=TRUE)
var.names <- names(var.names) # grab just the "V" variable names
for(xj in var.names) print(table(da35021.0001[,xj]))
```

```
(00) Jan not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16567                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                       0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                       0                         0                         0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                       0                         0                         0
  (12) Reported with Dec
                       0

(00) Feb not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16542                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                       0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                       0                         0                         0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                       0                         0                         0
  (12) Reported with Dec
                       0

(00) Mar not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16500                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                       0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
```

```
                        0                         0                         0
  (09) Reported with Sep    (10) Reported with Oct    (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) Apr not w oth month    (01) Reported with Jan    (02) Reported with Feb
                    16463                         0                         0
  (03) Reported with Mar    (04) Reported with Apr    (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun    (07) Reported with Jul    (08) Reported with Aug
                        0                         0                         0
  (09) Reported with Sep    (10) Reported with Oct    (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) May not w oth month    (01) Reported with Jan    (02) Reported with Feb
                    16433                         0                         0
  (03) Reported with Mar    (04) Reported with Apr    (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun    (07) Reported with Jul    (08) Reported with Aug
                        0                         0                         0
  (09) Reported with Sep    (10) Reported with Oct    (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) June not w oth month    (01) Reported with Jan
                    16383                         0
   (02) Reported with Feb    (03) Reported with Mar
                        0                         0
   (04) Reported with Apr    (05) Reported with May
                        0                         0
   (06) Reported with Jun    (07) Reported with Jul
                        0                         0
   (08) Reported with Aug    (09) Reported with Sep
                        0                         0
   (10) Reported with Oct    (11) Reported with Nov
                        0                         0
   (12) Reported with Dec
                        0

(00) July not w oth month    (01) Reported with Jan
                    16348                         0
   (02) Reported with Feb    (03) Reported with Mar
                        0                         0
   (04) Reported with Apr    (05) Reported with May
```

```
                        0                         0
   (06) Reported with Jun    (07) Reported with Jul
                        0                         0
   (08) Reported with Aug    (09) Reported with Sep
                        0                         0
   (10) Reported with Oct    (11) Reported with Nov
                        0                         0
   (12) Reported with Dec
                        0

(00) Aug not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16297                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                        0                         0                         0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) Sep not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16248                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                        0                         0                         0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) Oct not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16185                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                        0                         0                         0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                        0                         0                         0
  (12) Reported with Dec
                        0

(00) Nov not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   16090                         0                         0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                        0                         0                         0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
```

```
                        0                           0                           0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                        0                           0                           0
  (12) Reported with Dec
                        0

(00) Dec not w oth month   (01) Reported with Jan   (02) Reported with Feb
                   15862                           0                           0
  (03) Reported with Mar   (04) Reported with Apr   (05) Reported with May
                        0                           0                           0
  (06) Reported with Jun   (07) Reported with Jul   (08) Reported with Aug
                        0                           0                           0
  (09) Reported with Sep   (10) Reported with Oct   (11) Reported with Nov
                        0                           0                           0
  (12) Reported with Dec
                        0
```

For now, we just want to work with agencies that have complete data.

```
da35021.0001 <- subset(da35021.0001, V12=="(12) Dec last reported")
```

Creating a dataframe

Of course, for most criminological questions we're interested in, we don't need all 32 columns. For example, we don't need to know Variable 25 - "SPECIAL MAILING ADDRESS." So, we'll create a smaller data frame with just the information we need.

Think about the structure of this dataframe. Why are we making some of the variables "as.character?" As a hint, note that we didn't use "as.character with the POP variable. What is the ORI variable? What is the AGENCY variable? Note that stringsAsFactors=FALSE states that we should simply treat all strings as characters - and not as factor variables.

What do you notice that is useful about this new dataframe? Hint - it has to do with the variable names. Each row in our dataset refers to an individual agency.

```
ucr <- data.frame(ORI   =as.character(da35021.0001$V3),
                  AGENCY=as.character(da35021.0001$V29),
                  AREA  =as.character(da35021.0001$V26),
                  POP   =da35021.0001$V14,
                  MONTHS=as.character(da35021.0001$V12),
                  STATE=as.character(da35021.0001$V2),
                  stringsAsFactors=FALSE)
```

Importantly, our UCR OKCA data covers only what the FBI calls "Part I" crimes. Part I crimes are Murder, Rape, Robbery, Larceny, Motor Vehicle Theft, Assaults (Aggravated), and Burglary.

So for each of the 15,862 agencies, we can count up the number of Murders as follows:

```
var.names <- grep("ACT NUM MURDER",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$murder <- rowSums(da35021.0001[,var.names])
```

Run the first three lines of the ucr dataset. You should see a new column called murder. Let's add a new column for rape count for each agency.

```
var.names <- grep("ACT NUM RAPE",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$rape <- rowSums(da35021.0001[,var.names])
```

And robbery

```
var.names <- grep("ACT NUM ROBBRY",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$robbery <- rowSums(da35021.0001[,var.names])
```

Let's add another column for Assault. Assault requires a bit more steps because we are interested in only what most jurisdictions would define as Aggravated Assault. We will subtract the number of simple assaults from the total number of assaults to obtain the total number of aggravated assaults.

```
var.names <- grep("ACT NUM ASSLT",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$assault <- rowSums(da35021.0001[,var.names])
var.names <- grep("ACT # SIMPLE ASSLT",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$assault <- ucr$assault - rowSums(da35021.0001[,var.names])
```

Let's add a column for burglary, a column for larceny, and a column for vehicle theft:

```
var.names <- grep("ACT # BURGLARY",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$burglary <- rowSums(da35021.0001[,var.names])

var.names <- grep("ACT # LARCENY",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$larceny <- rowSums(da35021.0001[,var.names])

var.names <- grep("ACT # VHC THEFT",var.lookup,value=TRUE)
var.names <- names(var.names)
ucr$gta <- rowSums(da35021.0001[,var.names])
```

Trying out our dataset

We can use this dataset to check crime counts for individual cities. First, take a look at http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/8tabledatadecpdf/table-8-state-cuts/table-8-pennsylvania. This provides some annotations about reporting involving any agency in Pennsylvania. Let's use our dataset to find out crime counts in Philadelphia in 2012. Note that if you use regular expressions to find "Philadelphia," you'll also

grab "other" Philadelphias such as Philadelphia, Mississippi - so check your STATE column. Also, would we want to include "New Philadelphia?"

```
i <- grep("PHILADELPHIA",ucr$AGENCY)
ucr[i,]
```

```
          ORI                          AGENCY                      AREA
6865   MS05001 PHILADELPHIA POLICE DEPT       PHILADELPHIA
10221  OH07904 NEW PHILADELPHIA POLICE DEPT   NEW PHILADELPHIA
10770  PAPEP00 PHILADELPHIA POLICE DEPARTMENT PHILADELPHIA
11861  PA051BF PHILADELPHIA COUNTY            BF: PHILADELPHIA COUNTY
11862  PA051DE PHILADELPHIA COUNTY            BN: PHILADELPHIA COUNTY
11863  PA051SP ST POLICE: PHILADELPHIA COUNTY SP: PHILADELPHIA COUNTY
11898  PA05416 NEW PHILADELPHIA POLICE DEPT   NEW PHILADELPHIA
12167  PA51BCI INVESTIGATION:PHILADELPHIA CTY SP BCI: PHILADELPHIA CTY
          POP                  MONTHS             STATE murder rape robbery
6865     7529 (12) Dec last reported  (23) Mississippi      0    4      11
10221   17272 (12) Dec last reported         (34) Ohio      0    0       1
10770 1538957 (12) Dec last reported (37) Pennsylvania    331  880    7984
11861       0 (12) Dec last reported (37) Pennsylvania      0    0       0
11862       0 (12) Dec last reported (37) Pennsylvania      0    0       0
11863       0 (12) Dec last reported (37) Pennsylvania      1    1       0
11898    1082 (12) Dec last reported (37) Pennsylvania      0    0       0
12167       0 (12) Dec last reported (37) Pennsylvania      0    0       0
      assault burglary larceny  gta
6865       37      144     164   11
10221       3        6     227    1
10770    8658    12004   38592 6401
11861       0        0       0    0
11862       0        0       0    0
11863      11        0       2    0
11898       3        2       1    0
12167       0        0       1    0
```

Which rows pertain to Philadelphia, Pennsylvania? Why would Row 11863 have a POP of 0? Hint: that line refers to the Pennsylvania State Police who investigate/report crimes in Philadelphia County.

Let's try another large city - Chicago. First, look at this link: https://ucr.fbi.gov/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/8tabledatadecpdf/table-8-state-cuts/table_8_offenses_known_to_law_enforcement_by_illinois_by_city_2012.xls

As you can see, the data collection methodology for the offense of forcible rape used by Chicago, Illinois, does not comply with national Uniform Crime Reporting Program guidelines. Consequently, its figures for forcible rape and violent crime (of which forcible rape is a part) are not published in this table.

```
i <- grep("CHICAGO",ucr$AGENCY)
ucr[i,]
```

```
         ORI                        AGENCY                    AREA
```

```
3074 ILCPD00 CHICAGO POLICE DEPARTMENT      CHICAGO
3131 IL016AC NORTHWESTERN UNIV: CHICAGO     NORTHWESTERN UNIV:CHICAG
3151 IL01619 CHICAGO HEIGHTS POLICE DEPT    CHICAGO HEIGHTS
3157 IL01620 CHICAGO RIDGE POLICE DEPT      CHICAGO RIDGE
3204 IL0166A SOUTH CHICAGO HEIGHTS PD       SOUTH CHICAGO HEIGHTS
3206 IL0166C UNIVERSITY OF ILLINOIS:CHICAGO UNIV OF IL: CHICAGO
3207 IL0166L CHICAGO STATE UNIVERSITY       CHICAGO STATE UNIVERSITY
3299 IL02219 WEST CHICAGO POLICE DEPARTMENT WEST CHICAGO
3436 IL04915 NORTH CHICAGO POLICE DEPT      NORTH CHICAGO
3884 IN04503 EAST CHICAGO POLICE DEPARTMENT EAST CHICAGO
         POP                  MONTHS         STATE murder rape robbery
3074 2708382 (12) Dec last reported (12) Illinois    500    0   13476
3131       0 (12) Dec last reported (12) Illinois      0    0       0
3151   30422 (12) Dec last reported (12) Illinois      3   26     115
3157   14373 (12) Dec last reported (12) Illinois      1    5       8
3204    4160 (12) Dec last reported (12) Illinois      1    0       4
3206       0 (12) Dec last reported (12) Illinois      0    2       9
3207       0 (12) Dec last reported (12) Illinois      0    0       6
3299   27286 (12) Dec last reported (12) Illinois      0    8       5
3436   32697 (12) Dec last reported (12) Illinois      0    4      24
3884   29764 (12) Dec last reported  (13) Indiana      3    9      88
     assault burglary larceny   gta
3074   12272    22748   72717 17001
3131       0        4      36     0
3151     108      480     531    88
3157      15       50     539    36
3204       5       19     135     8
3206       4        6     488     4
3207       0        3      44     0
3299      11       61     264    24
3436      58       90     133     4
3884     120      560    1025   169
```

Certainly, Philadelphia had more than one agency; but, here's another large city has a large amount of agencies. So, if you were going to count up all of the crimes in Los Angeles County in 2012, you would have to (1) consider which agencies would apply and (2) then add up the counts for each of these agencies.

```
i <- grep("LOS ANGELES",ucr$AGENCY)
ucr[i,]
```

```
        ORI                          AGENCY                      AREA
857 CA0190D C/O SHERIFF LOS ANGELES COUNTY LA HABRA HEIGHTS
858 CA0190I LOS ANGELES COUNTY             UPRR: LOS ANGELES COUNTY
861 CA0190X C/O SHERIFF LOS ANGELES COUNTY WESTLAKE VILLAGE
862 CA01900 LOS ANGELES COUNTY             LOS ANGELES
865 CA01903 C/O SHERIFF LOS ANGELES COUNTY ARTESIA
866 CA01904 C/O SHERIFF LOS ANGELES COUNTY AVALON
870 CA01908 C/O SHERIFF LOS ANGELES COUNTY BELLFLOWER
```

```
873 CA0191F C/O SHERIFF LOS ANGELES COUNTY AGOURA HILLS
875 CA0191R C/O SHERIFF LOS ANGELES COUNTY SANTA CLARITA
876 CA0191W C/O SHERIFF LOS ANGELES COUNTY DIAMOND BAR
878 CA01911 C/O SHERIFF LOS ANGELES COUNTY BRADBURY
881 CA01914 C/O SHERIFF LOS ANGELES COUNTY COMMERCE
886 CA01919 C/O SHERIFF LOS ANGELES COUNTY CERRITOS
887 CA0192H C/O SHERIFF LOS ANGELES COUNTY CALABASAS
888 CA0192J C/O SHERIFF LOS ANGELES COUNTY MALIBU
889 CA0192U C/O LOS ANGELES COUNTY SHERIFF LA TRNSPRTN SERVICES BUR
891 CA01921 C/O SHERIFF LOS ANGELES COUNTY DUARTE
900 CA01930 C/O SHERIFF LOS ANGELES COUNTY HIDDEN HILLS
902 CA01932 C/O SHERIFF LOS ANGELES COUNTY INDUSTRY
905 CA01935 C/O SHERIFF LOS ANGELES COUNTY LAKEWOOD
906 CA01936 C/O SHERIFF LOS ANGELES COUNTY LA MIRADA
907 CA01937 C/O SHERIFF LOS ANGELES COUNTY LA PUENTE
909 CA01939 C/O SHERIFF LOS ANGELES COUNTY LAWNDALE
910 CA01940 C/O SHERIFF LOS ANGELES COUNTY LOMITA
912 CA01942 LOS ANGELES POLICE DEPARTMENT  LOS ANGELES
913 CA01943 C/O SHERIFF LOS ANGELES COUNTY LYNWOOD
919 CA01949 C/O SHERIFF LOS ANGELES COUNTY NORWALK
920 CA01950 C/O SHERIFF LOS ANGELES COUNTY PALMDALE
922 CA01952 C/O SHERIFF LOS ANGELES COUNTY PARAMOUNT
924 CA01954 C/O SHERIFF LOS ANGELES COUNTY PICO RIVERA
927 CA01957 C/O SHERIFF LOS ANGELES COUNTY ROLLING HILLS
928 CA01958 C/O SHERIFF LOS ANGELES COUNTY ROLLING HILLS ESTATES
929 CA01959 C/O SHERIFF LOS ANGELES COUNTY ROSEMEAD
930 CA01960 C/O SHERIFF LOS ANGELES COUNTY SAN DIMAS
938 CA01968 C/O SHERIFF LOS ANGELES COUNTY SOUTH EL MONTE
941 CA01971 C/O SHERIFF LOS ANGELES COUNTY TEMPLE CITY
944 CA01974 C/O SHERIFF LOS ANGELES COUNTY WALNUT
947 CA01977 C/O SHERIFF LOS ANGELES COUNTY CARSON
949 CA01980 C/O SHERIFF LOS ANGELES COUNTY RANCHO PALOS VERDES
952 CA01983 CA ST UNIVERSITY: LOS ANGELES  CA ST UN: LOS ANGELES
954 CA01990 C/O SHERIFF LOS ANGELES COUNTY LA CANADA FLINTRIDGE
955 CA01996 C/O SHERIFF LOS ANGELES COUNTY LANCASTER
956 CA01997 UNIVERSITY OF CA: LOS ANGELES  UN OF CA: LOS ANGELES
957 CA01999 HP: LOS ANGELES COUNTY         HP: LOS ANGELES COUNTY
```

	POP	MONTHS	STATE	murder	rape	robbery
857	5413	(12) Dec last reported (04)	California	0	0	0
858	0	(12) Dec last reported (04)	California	0	0	0
861	8406	(12) Dec last reported (04)	California	0	1	1
862	1074864	(12) Dec last reported (04)	California	75	161	1490
865	16793	(12) Dec last reported (04)	California	1	1	27
866	3795	(12) Dec last reported (04)	California	0	0	3
870	77886	(12) Dec last reported (04)	California	6	11	135
873	20667	(12) Dec last reported (04)	California	0	2	1
875	179248	(12) Dec last reported (04)	California	1	30	96
876	56470	(12) Dec last reported (04)	California	0	1	17

878	1067	(12) Dec last reported	(04) California	0	0	0
881	13035	(12) Dec last reported	(04) California	1	2	53
886	49856	(12) Dec last reported	(04) California	0	7	67
887	23442	(12) Dec last reported	(04) California	1	5	3
888	12854	(12) Dec last reported	(04) California	0	4	5
889	0	(12) Dec last reported	(04) California	1	4	379
891	21673	(12) Dec last reported	(04) California	0	5	17
900	1887	(12) Dec last reported	(04) California	0	0	0
902	222	(12) Dec last reported	(04) California	0	1	34
905	81382	(12) Dec last reported	(04) California	1	13	102
906	49312	(12) Dec last reported	(04) California	2	4	20
907	40479	(12) Dec last reported	(04) California	3	6	42
909	33312	(12) Dec last reported	(04) California	2	4	51
910	20591	(12) Dec last reported	(04) California	1	1	30
912	3855122	(12) Dec last reported	(04) California	299	936	8983
913	70908	(12) Dec last reported	(04) California	9	24	211
919	107295	(12) Dec last reported	(04) California	11	13	171
920	155294	(12) Dec last reported	(04) California	6	37	224
922	54997	(12) Dec last reported	(04) California	3	7	129
924	63988	(12) Dec last reported	(04) California	1	6	92
927	1891	(12) Dec last reported	(04) California	0	0	0
928	8202	(12) Dec last reported	(04) California	0	0	1
929	54656	(12) Dec last reported	(04) California	0	4	54
930	33923	(12) Dec last reported	(04) California	0	2	14
938	20452	(12) Dec last reported	(04) California	2	3	38
941	36148	(12) Dec last reported	(04) California	0	2	16
944	29658	(12) Dec last reported	(04) California	0	2	10
947	93233	(12) Dec last reported	(04) California	7	14	142
949	42335	(12) Dec last reported	(04) California	0	1	6
952	0	(12) Dec last reported	(04) California	0	0	0
954	20584	(12) Dec last reported	(04) California	0	1	4
955	159155	(12) Dec last reported	(04) California	8	54	313
956	0	(12) Dec last reported	(04) California	0	14	21
957	0	(12) Dec last reported	(04) California	0	0	0

	assault	burglary	larceny	gta
857	6	24	18	2
858	2	201	46	0
861	1	36	106	12
862	3732	4816	8223	3942
865	31	70	164	28
866	10	11	41	12
870	152	417	886	499
873	9	62	158	16
875	215	616	1822	304
876	37	354	525	73
878	0	6	4	0
881	56	130	616	264
886	46	441	1253	176

887	4	62	165	11
888	6	78	232	19
889	281	20	791	88
891	49	121	317	69
900	0	1	3	0
902	33	111	785	214
905	111	371	1381	310
906	72	170	503	103
907	70	110	253	158
909	110	135	164	98
910	63	81	264	46
912	8329	16388	56006	15084
913	297	350	503	520
919	238	539	1537	533
920	545	1035	2068	290
922	105	272	810	454
924	162	349	1063	368
927	0	9	17	1
928	8	36	90	3
929	85	184	511	218
930	35	141	471	56
938	45	104	184	111
941	20	116	201	37
944	25	141	209	32
947	357	601	1558	550
949	28	167	306	25
952	1	9	170	10
954	7	129	179	16
955	484	1043	2051	404
956	27	173	736	14
957	38	9	46	377

Exercises

1. How many car thefts reported to the police?
2. How many rapes reported to the police?
3. Pick out all agencies that have "New York" in their name?
4. Next, make a data frame of just "New Jersey". How many rows does it have?
5. How many murders occurred in New Jersey? How many robberies?
6. Find the burglary rate/person in New Jersey? per 100,000?
7. Add a new column to your dataset for the burglary rate for each agency. Sort your dataset by order of burglary rate.

Solutions to the exercises

1. How many car thefts reported to the police?

```
sum(ucr$gta)
```

```
[1] 714060
```

2. How many rapes reported to the police?

```
sum(ucr$rape)
```

```
[1] 83073
```

3. Pick out all agencies that have "New York" in their name?

```
i <- grep("NEW YORK",ucr$AGENCY)
ucr[i,]
```

```
         ORI                          AGENCY                      AREA
8306 NJ00912 WEST NEW YORK POLICE DEPT      WEST NEW YORK
8890 NY03030 NEW YORK CITY POLICE DEPT      NEW YORK
8905 NY03240 NEW YORK MILLS VILLAGE PD      NEW YORK MILLS VILLAGE
9182 NY330SS SP: NEW YORK COUNTY            SP: NEW YORK COUNTY
         POP                  MONTHS           STATE murder rape robbery
8306   50507 (12) Dec last reported (29) New Jersey      0    8      93
8890 8289415 (12) Dec last reported   (31) New York    419 1162   20201
8905    3339 (12) Dec last reported   (31) New York      0    0       0
9182       0 (12) Dec last reported   (31) New York      0    1       0
     assault burglary larceny  gta
8306      95      195     487   51
8890   31211    18635  115935 8190
8905       1        6      52    2
9182       2        0      70    0
```

Note what the `i` (indexing variable) does - it is a list of the row numbers in the `ucr` dataframe that have "NEW YORK" somewhere in the agency column. As you can see, there are 4 agencies that have "New York" in their name.

4. Next, make a data frame of just "New Jersey". How many rows does it have? Here are two different ways to create the dataframe

```
nj <- ucr[grep("New Jersey", ucr$STATE),]
nrow(nj)
```

```
[1] 567
```

Or,

```
nj2 <- subset(ucr, STATE=="(29) New Jersey")
nrow(nj2)
```

```
[1] 567
```

5. How many murders occurred in New Jersey? How many robberies?

```
sum(nj$murder)
sum(nj$robbery)
```

```
[1] 388
[1] 11345
```

6. Find the burglary rate/person in New Jersey? per 100,000?

```
sum(nj$murder)/sum(nj$POP)

sum(nj$murder)/sum(nj$POP)*100000
```

```
[1] 4.421895e-05
[1] 4.421895
```

7. Add a new column to your dataset for the burglary rate for each agency. Sort your dataset by order of burglary rate.

```
nj$burglaryrate <- (nj$burglary/nj$POP)*100000
head(nj[order(nj$burglaryrate),])
tail(nj[order(nj$burglaryrate),])
```

```
         ORI                           AGENCY                       AREA   POP
8217 NJ00503 C/O CHIEF CAPE MAY POLICE DEPT CAPE MAY POINT                290
8311 NJ01004 CALIFON POLICE DEPARTMENT      CALIFON                      1078
8418 NJ01347 LAKE COMO POLICE DEPARTMENT    LAKE COMO                    1769
8540 NJ01807 FAR HILLS POLICE DEPARTMENT    FAR HILLS                     928
8590 NJ02021 WINFIELD TOWNSHIP POLICE DEPT  WINFIELD TOWNSHIP           1487
8433 NJ01405 CHATHAM TOWNSHIP POLICE DEPT   CHATHAM TOWNSHIP           10560
                    MONTHS           STATE murder rape robbery assault
8217 (12) Dec last reported (29) New Jersey      0    0       0       0
8311 (12) Dec last reported (29) New Jersey      0    0       0       0
8418 (12) Dec last reported (29) New Jersey      0    0       0       1
8540 (12) Dec last reported (29) New Jersey      0    0       0       0
8590 (12) Dec last reported (29) New Jersey      0    0       0       0
8433 (12) Dec last reported (29) New Jersey      0    0       0       1
     burglary larceny gta burglaryrate
8217        0       7   0      0.00000
8311        0       6   1      0.00000
8418        0      26   2      0.00000
8540        0       9   0      0.00000
8590        0       6   0      0.00000
8433        2      21   2     18.93939
         ORI                           AGENCY                       AREA POP
8567 NJ01989 PROSECUTOR: SUSSEX COUNTY      PROSECUTOR: SUSSEX CNTY    0
8568 NJ020SP STATE POLICE: UNION COUNTY     SP: UNION COUNTY           0
8569 NJ02000 UNION COUNTY                   UNION                      0
8592 NJ02089 PROSECUTOR: UNION COUNTY       PROSECUTOR: UNION COUNTY   0
8594 NJ02100 WARREN COUNTY                  WARREN                     0
```

```
8608 NJ02189 PROSECUTOR: WARREN COUNTY      PROSECUTOR: WARREN CNTY    0
                    MONTHS           STATE murder rape robbery assault
8567 (12) Dec last reported (29) New Jersey      0    0       0       0
8568 (12) Dec last reported (29) New Jersey      0    0       3       4
8569 (12) Dec last reported (29) New Jersey      0    0       0       0
8592 (12) Dec last reported (29) New Jersey      0    0       0       0
8594 (12) Dec last reported (29) New Jersey      0    0       0       0
8608 (12) Dec last reported (29) New Jersey      0    0       0       0
     burglary larceny gta burglaryrate
8567        0       0   0          NaN
8568        0       2  15          NaN
8569        0       0   0          NaN
8592        0       0   0          NaN
8594        0       0   0          NaN
8608        0       0   0          NaN
```

We have a bit of a snag. Look at the burglary rate column. You'll notice some `NaN` and some `Inf` values too. Let's remove those from the dataset.

```
njNew <- subset(nj, !is.nan(burglaryrate) & !is.infinite(burglaryrate))
```

Now sort the dataset by burglary rate:

```
head(njNew[order(njNew$burglaryrate),])
tail(njNew[order(njNew$burglaryrate),])
```

```
         ORI                          AGENCY                      AREA   POP
8217 NJ00503 C/O CHIEF CAPE MAY POLICE DEPT CAPE MAY POINT             290
8311 NJ01004 CALIFON POLICE DEPARTMENT      CALIFON                   1078
8418 NJ01347 LAKE COMO POLICE DEPARTMENT    LAKE COMO                 1769
8540 NJ01807 FAR HILLS POLICE DEPARTMENT    FAR HILLS                  928
8590 NJ02021 WINFIELD TOWNSHIP POLICE DEPT  WINFIELD TOWNSHIP         1487
8433 NJ01405 CHATHAM TOWNSHIP POLICE DEPT   CHATHAM TOWNSHIP         10560
                    MONTHS           STATE murder rape robbery assault
8217 (12) Dec last reported (29) New Jersey      0    0       0       0
8311 (12) Dec last reported (29) New Jersey      0    0       0       0
8418 (12) Dec last reported (29) New Jersey      0    0       0       1
8540 (12) Dec last reported (29) New Jersey      0    0       0       0
8590 (12) Dec last reported (29) New Jersey      0    0       0       0
8433 (12) Dec last reported (29) New Jersey      0    0       0       1
     burglary larceny gta burglaryrate
8217        0       7   0      0.00000
8311        0       6   1      0.00000
8418        0      26   2      0.00000
8540        0       9   0      0.00000
8590        0       6   0      0.00000
8433        2      21   2     18.93939
         ORI                          AGENCY                      AREA  POP
8225 NJ00513 WEST WILDWOOD POLICE DEPT      WEST WILDWOOD             602
8222 NJ00509 SEA ISLE CITY POLICE DEPT      SEA ISLE CITY            2109
```

```
8477 NJ01509 HARVEY CEDARS POLICE DEPT      HARVEY CEDARS             341
8226 NJ00514 WILDWOOD POLICE DEPARTMENT     WILDWOOD                 5315
8204 NJ00433 C/O HADDONFIELD POLICE DEPT    TAVISTOCK                   5
8200 NJ00429 PINE VALLEY POLICE DEPARTMENT  PINE VALLEY                12
                    MONTHS           STATE murder rape robbery assault
8225 (12) Dec last reported (29) New Jersey      0    0       0       0
8222 (12) Dec last reported (29) New Jersey      0    0       4       1
8477 (12) Dec last reported (29) New Jersey      0    0       0       0
8226 (12) Dec last reported (29) New Jersey      0    4      20      69
8204 (12) Dec last reported (29) New Jersey      0    0       0       0
8200 (12) Dec last reported (29) New Jersey      0    0       0       0
     burglary larceny gta burglaryrate
8225       19      19   1     3156.146
8222       71     136   0     3366.524
8477       12      13   4     3519.062
8226      221     469  14     4158.043
8204        1       0   0    20000.000
8200        4       0   0    33333.333
```

Why do you think some of the agencies are at the top of the list? Hint: take a look at the population column.