Assault/Homicide Mortality for young Black adult males in USA

Jon Minton

23 April 2016

# Introduction

This document will present analyses for death rate trends due to Assault/Homicide for young adult African Americans (aged 15 to 34 years), and the relationship, or otherwise, between this and whether Republicans or Democrats are in power.

# Data Management

First to load the packages and the data. (Not displayed)

## Warning: package 'knitr' was built under R version 3.1.3

## Warning: package 'readr' was built under R version 3.1.3

## Warning: package 'plyr' was built under R version 3.1.3

## Warning: package 'tidyr' was built under R version 3.1.3

## Warning: package 'dplyr' was built under R version 3.1.3

## Warning: package 'stringr' was built under R version 3.1.3

## Warning: package 'car' was built under R version 3.1.3

## Warning: package 'xts' was built under R version 3.1.3

## Warning: package 'zoo' was built under R version 3.1.3

## Warning: package 'mFilter' was built under R version 3.1.3

## Warning: package 'lubridate' was built under R version 3.1.3

## Warning: package 'ggplot2' was built under R version 3.1.3

## [1] "tbl\_df" "tbl" "data.frame"

The dataset looks as follows

dta

## Source: local data frame [30,144 x 8]  
##   
## icd\_class year age sex race  
## (chr) (int) (chr) (chr) (chr)  
## 1 icd\_08 1968 < 1 female black  
## 2 icd\_08 1968 < 1 female black  
## 3 icd\_08 1968 < 1 female black  
## 4 icd\_08 1968 < 1 female black  
## 5 icd\_08 1968 < 1 female black  
## 6 icd\_08 1968 < 1 female black  
## 7 icd\_08 1968 < 1 female black  
## 8 icd\_08 1968 < 1 female black  
## 9 icd\_08 1968 1-4 female black  
## 10 icd\_08 1968 1-4 female black  
## .. ... ... ... ... ...  
## Variables not shown: cause (chr), death\_count (int), population\_count  
## (dbl)

Now to create the subgroup of interest:

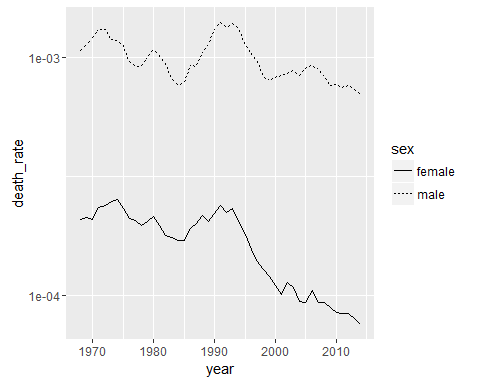
dta\_subset <- dta %>%   
 filter(race =="black") %>%   
 filter(age %in% c("15-19", "20-24", "25-34")) %>%   
 filter(cause == "assault") %>%  
 select(year, sex, death\_count, population\_count) %>%   
 group\_by(year, sex) %>%   
 summarise(  
 death\_count = sum(death\_count),   
 population\_count = sum(population\_count)  
 ) %>%   
 mutate(rep = ifelse(year %in% c(2010:2014, 1994:2001, 1978:1981, 1968:1969), 1, 0))   
  
dta\_subset

## Source: local data frame [94 x 5]  
## Groups: year [47]  
##   
## year sex death\_count population\_count rep  
## (int) (chr) (int) (dbl) (dbl)  
## 1 1968 female 710 3406000 1  
## 2 1968 male 3217 3027000 1  
## 3 1969 female 755 3538000 1  
## 4 1969 male 3566 3153000 1  
## 5 1970 female 760 3651372 0  
## 6 1970 male 3900 3267088 0  
## 7 1971 female 900 3840000 0  
## 8 1971 male 4514 3468000 0  
## 9 1972 female 956 4002000 0  
## 10 1972 male 4750 3640000 0  
## .. ... ... ... ... ...

# Figures

Here are some figures showing how the mortality rate due to this cause changed over time

dta\_subset %>%   
 mutate(death\_rate = death\_count / population\_count) %>%   
 ggplot(.) +   
 geom\_line(aes(x = year, y = death\_rate, group = sex, linetype = sex)) +   
 scale\_y\_log10()



We can see from this evidence of a very long-term fall throughout the period, but with a clear rise from around 1985 to the early 1990s, followed by a quick decline. For males the assault/homicide rates only fell back to mid 1980 levels by the late 1990s, whereas for females the decline since the early 1990s have been as follows. A table showing the same results is as follows:

dta\_subset %>%   
 mutate(deaths\_per\_100000 = 100000 \* death\_count / population\_count) %>%   
 select(year, sex, deaths\_per\_100000) %>%   
 spread(sex, deaths\_per\_100000) %>%   
 kable(., digits = 0)

|  |  |  |
| --- | --- | --- |
| year | female | male |
| 1968 | 21 | 106 |
| 1969 | 21 | 113 |
| 1970 | 21 | 119 |
| 1971 | 23 | 130 |
| 1972 | 24 | 130 |
| 1973 | 25 | 118 |
| 1974 | 25 | 118 |
| 1975 | 23 | 112 |
| 1976 | 21 | 96 |
| 1977 | 21 | 92 |
| 1978 | 20 | 91 |
| 1979 | 21 | 100 |
| 1980 | 22 | 107 |
| 1981 | 20 | 102 |
| 1982 | 18 | 94 |
| 1983 | 18 | 81 |
| 1984 | 17 | 76 |
| 1985 | 17 | 78 |
| 1986 | 19 | 92 |
| 1987 | 20 | 91 |
| 1988 | 22 | 105 |
| 1989 | 20 | 113 |
| 1990 | 22 | 131 |
| 1991 | 24 | 140 |
| 1992 | 22 | 133 |
| 1993 | 23 | 139 |
| 1994 | 21 | 132 |
| 1995 | 18 | 113 |
| 1996 | 16 | 104 |
| 1997 | 14 | 96 |
| 1998 | 13 | 83 |
| 1999 | 12 | 80 |
| 2000 | 11 | 82 |
| 2001 | 10 | 84 |
| 2002 | 11 | 86 |
| 2003 | 11 | 88 |
| 2004 | 9 | 84 |
| 2005 | 9 | 90 |
| 2006 | 11 | 92 |
| 2007 | 9 | 89 |
| 2008 | 9 | 83 |
| 2009 | 9 | 76 |
| 2010 | 9 | 77 |
| 2011 | 8 | 75 |
| 2012 | 8 | 76 |
| 2013 | 8 | 74 |
| 2014 | 8 | 70 |

# Knotted models

First to create a series of models using a rep dummy variable

## Source: local data frame [94 x 13]  
## Groups: year [47]  
##   
## year sex death\_count population\_count rep k1970 k1978 k1982 k1994  
## (int) (chr) (int) (dbl) (dbl) (dbl) (dbl) (dbl) (dbl)  
## 1 1968 female 710 3406000 1 0 0 0 0  
## 2 1968 male 3217 3027000 1 0 0 0 0  
## 3 1969 female 755 3538000 1 0 0 0 0  
## 4 1969 male 3566 3153000 1 0 0 0 0  
## 5 1970 female 760 3651372 0 0 0 0 0  
## 6 1970 male 3900 3267088 0 0 0 0 0  
## 7 1971 female 900 3840000 0 1 0 0 0  
## 8 1971 male 4514 3468000 0 1 0 0 0  
## 9 1972 female 956 4002000 0 2 0 0 0  
## 10 1972 male 4750 3640000 0 2 0 0 0  
## .. ... ... ... ... ... ... ... ... ...  
## Variables not shown: k2002 (dbl), k2010 (dbl), death\_rate (dbl), lg10mr  
## (dbl)

The summaries of the models are as follows:

llply(models, summary)

## $female  
##   
## Call:  
## lm(formula = lg10mr ~ year + k1970 + k1978 + k1982 + k1994 +   
## k2002 + k2010, data = x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.057364 -0.023080 -0.002725 0.016194 0.067359   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -63.585285 39.904606 -1.593 0.11914   
## year 0.030436 0.020263 1.502 0.14114   
## k1970 -0.035382 0.022663 -1.561 0.12655   
## k1978 -0.012739 0.009954 -1.280 0.20817   
## k1982 0.023098 0.008270 2.793 0.00805 \*\*   
## k1994 -0.047528 0.004763 -9.979 2.72e-12 \*\*\*  
## k2002 0.037737 0.006058 6.229 2.49e-07 \*\*\*  
## k2010 -0.013505 0.011567 -1.168 0.25006   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.03572 on 39 degrees of freedom  
## Multiple R-squared: 0.9619, Adjusted R-squared: 0.9551   
## F-statistic: 140.8 on 7 and 39 DF, p-value: < 2.2e-16  
##   
##   
## $male  
##   
## Call:  
## lm(formula = lg10mr ~ year + k1970 + k1978 + k1982 + k1994 +   
## k2002 + k2010, data = x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.082860 -0.035583 0.004911 0.033209 0.084179   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.416e+01 5.175e+01 -1.819 0.076540 .   
## year 4.633e-02 2.628e-02 1.763 0.085735 .   
## k1970 -6.140e-02 2.939e-02 -2.089 0.043271 \*   
## k1978 -3.581e-04 1.291e-02 -0.028 0.978010   
## k1982 3.006e-02 1.073e-02 2.803 0.007853 \*\*   
## k1994 -4.030e-02 6.177e-03 -6.525 9.7e-08 \*\*\*  
## k2002 2.803e-02 7.857e-03 3.568 0.000973 \*\*\*  
## k2010 -2.146e-02 1.500e-02 -1.431 0.160458   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04632 on 39 degrees of freedom  
## Multiple R-squared: 0.7436, Adjusted R-squared: 0.6976   
## F-statistic: 16.16 on 7 and 39 DF, p-value: 9.242e-10