Murders in Philadelphia

Josef Fruehwald

April 3, 2012

Outline

The Data

The Philadelphia Inquirer has a Google Fusion table (link) where they have compiled publicly available data from the Philadelphia Police Department on every murder in Philadelphia County between 1988 and 2011.

The Data

```
nrow(philly)

## [1] 8991

colnames(philly)

##  [1] "date"      "lastname"  "firstname" "age"
##  [5] "race"      "sex"       "cause"     "motive"
##  [9] "weapon"    "time"
```

The Data

```
table(is.na(philly$cause))

##
## FALSE  TRUE
##  6931  2060

table(is.na(philly$motive))

##
## FALSE  TRUE
##  6931  2060

table(is.na(philly$time))

##
## FALSE  TRUE
##  2060  6931
```

The Data

```
library(reshape2)
dcast(philly, race ~ weapon)
```

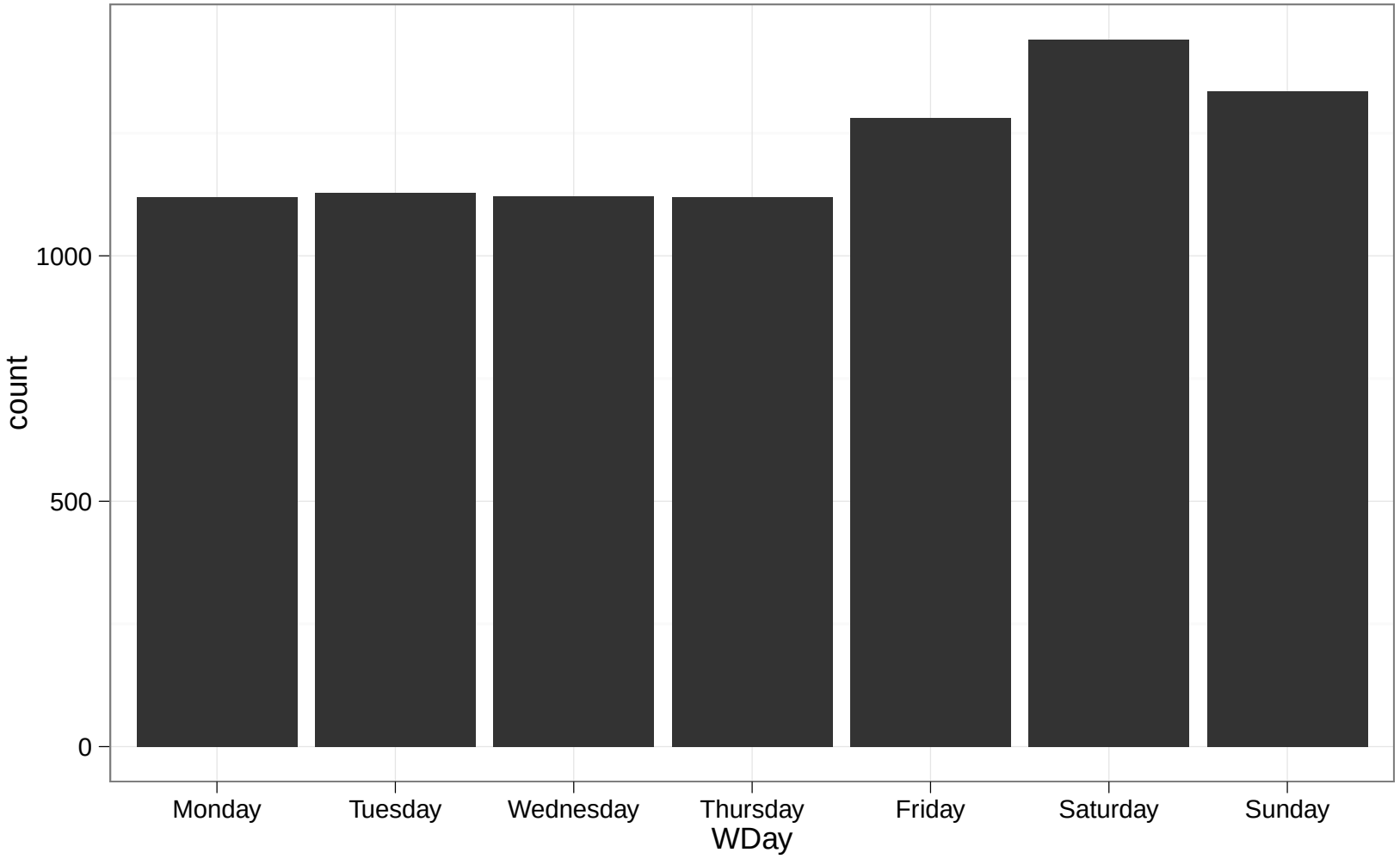
```
##   race ARSON FIREARM HANDS KNIFE OTHER UNKNOWN
## 1    A     1      67     9    17     1       0
## 2    B    97    5610   467   676   171      15
## 3    H     2     269    31    36     3       0
## 4    I     0       2     0     0     0       0
## 5    M     0       1     0     0     0       0
## 6    O     5      50    12     8     0       0
## 7    W    35     951   214   171    63       5
## 8 <NA>     0       1     0     1     0       0
```

The Data

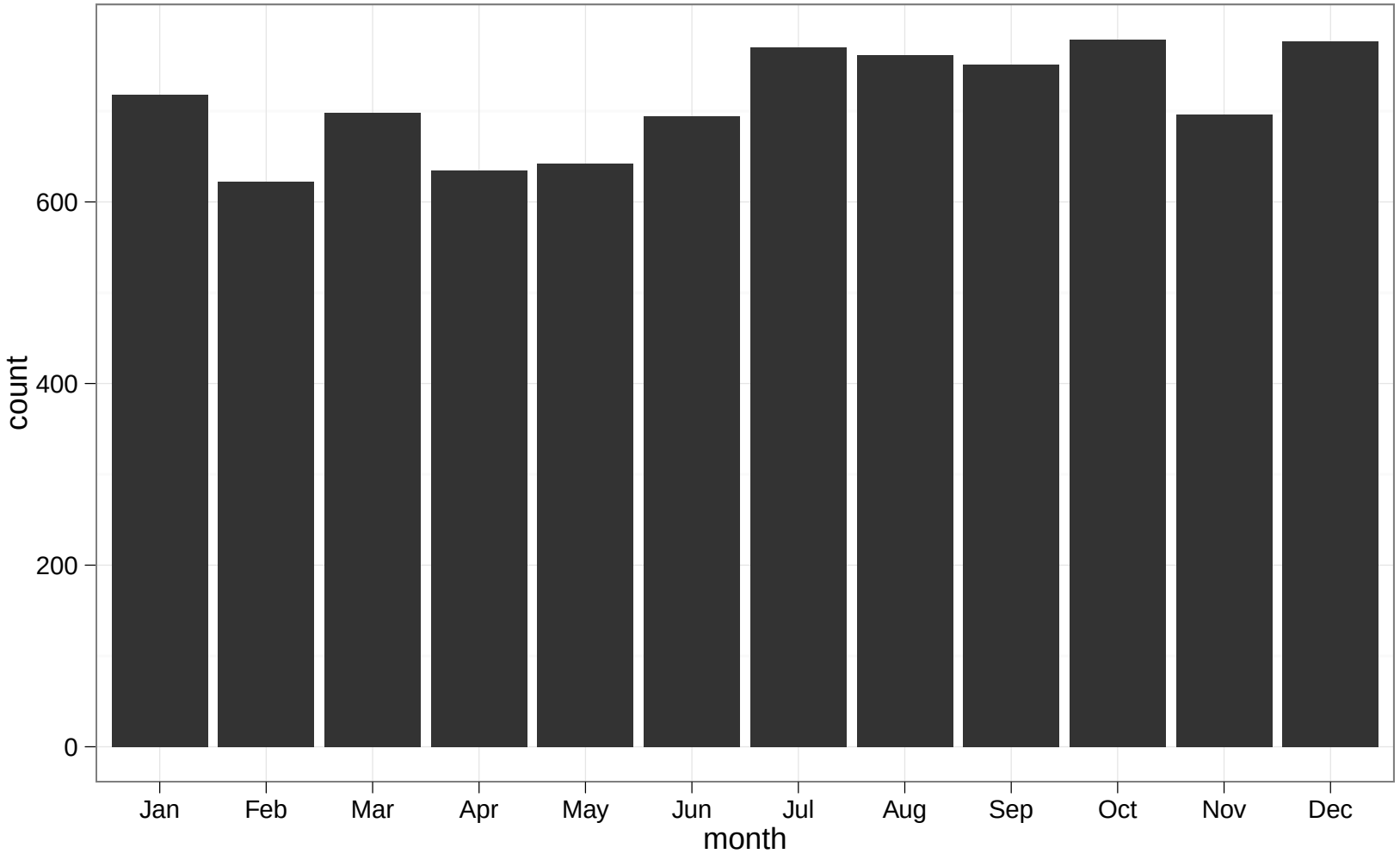
I've manipulated the data in my own way to include

- Month of year,
- month in Date format,
- year,
- a few different representations of the hour of day
- and the weekday.

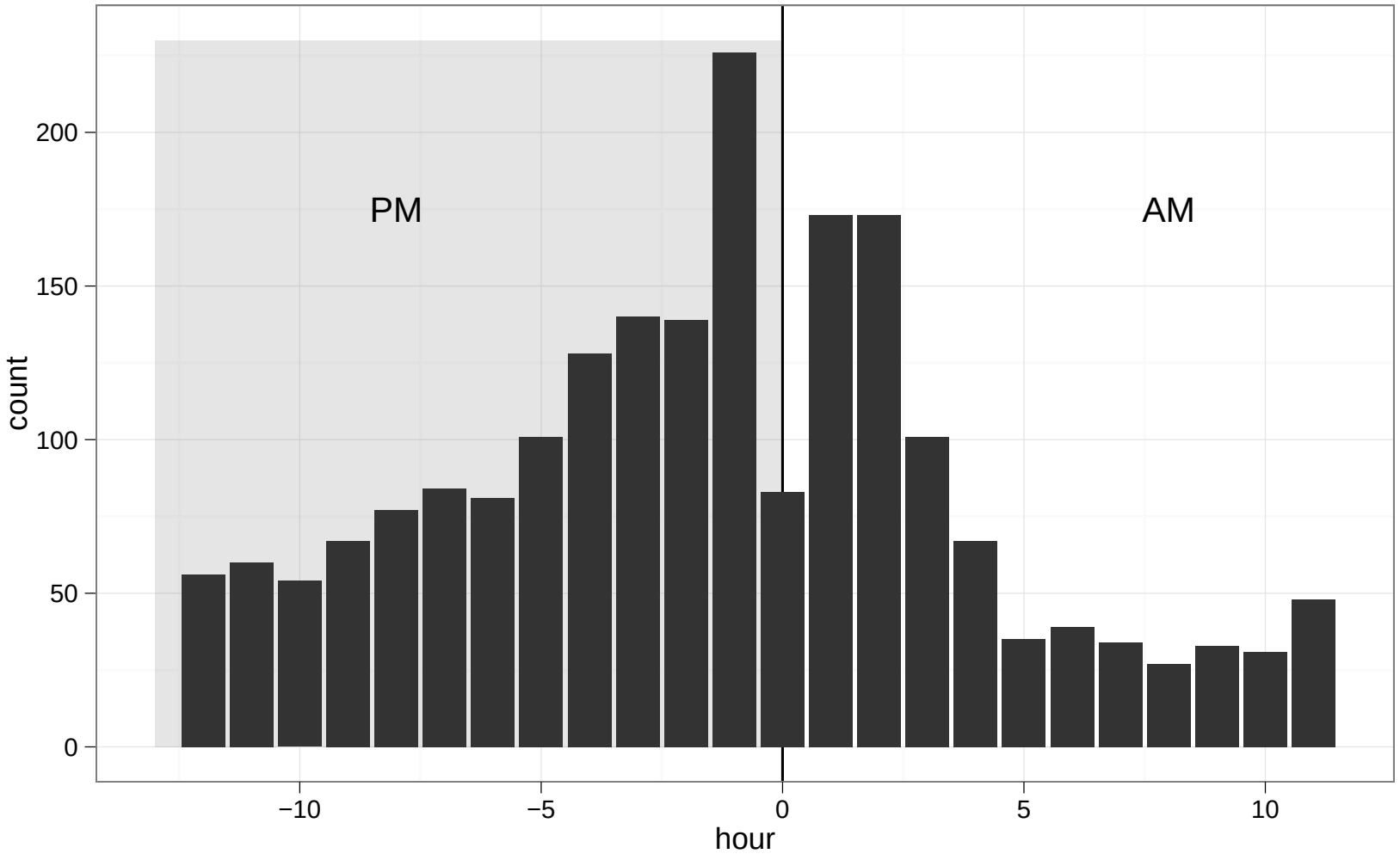
By Weekday


1000
500
0
count
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
WDay

By Month


count
600
400
200
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month

By Hour



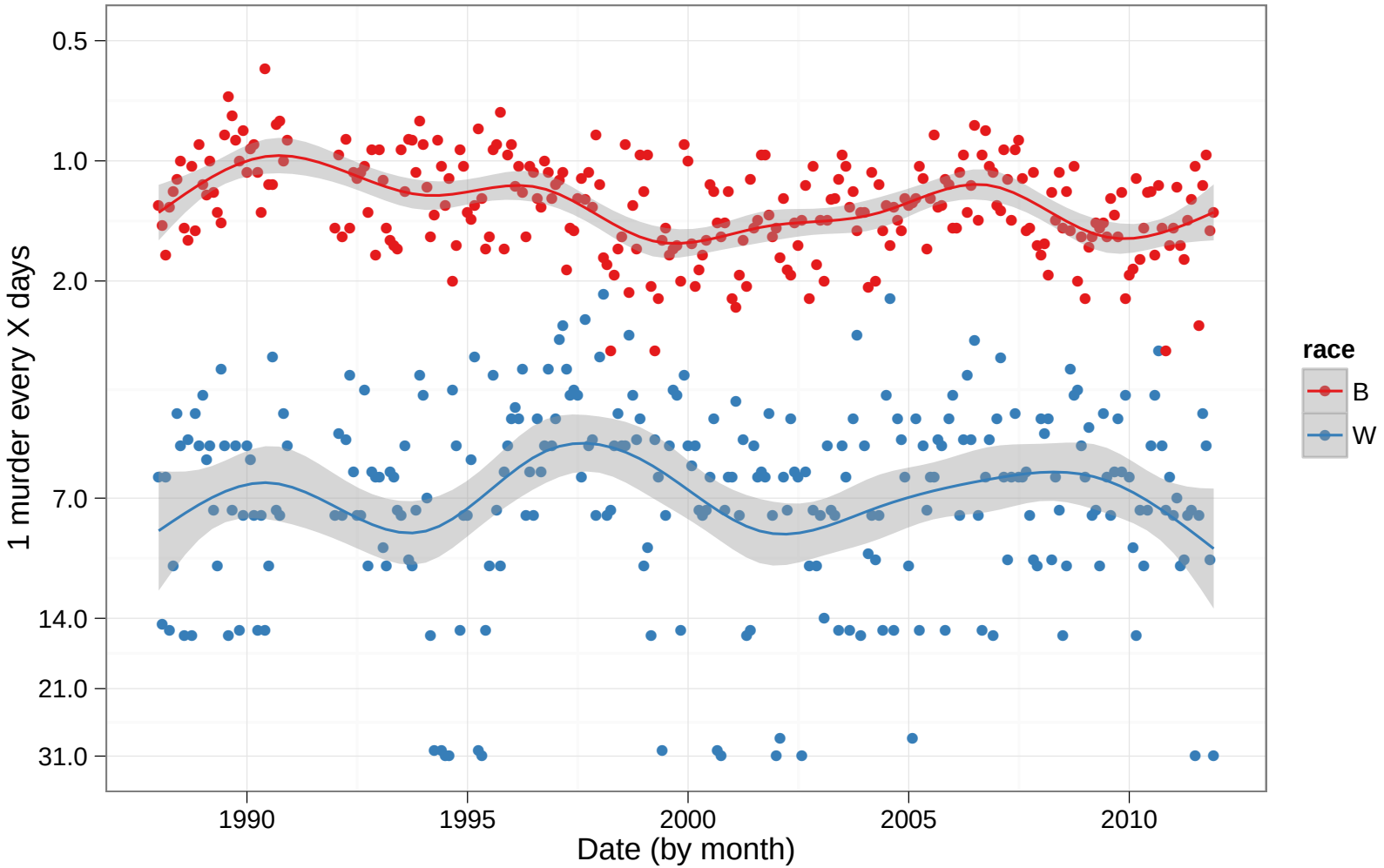
Outline

Two Philadelphias

```
philly.bw <- subset(philly, race %in%
    c("B", "W") & !is.na(month))
philly.bw.count <- count(philly.bw,
    c("month.date", "month", "race", "ndays"))
head(philly.bw.count)
```

```
##   month.date month race ndays freq
## 1 1988-01-01   Jan    B    31   24
## 2 1988-01-01   Jan    W    31    5
## 3 1988-02-01   Feb    B    29   20
## 4 1988-02-01   Feb    W    29    2
## 5 1988-03-01   Mar    B    31   18
## 6 1988-03-01   Mar    W    31    5
```

Two Philadelphias

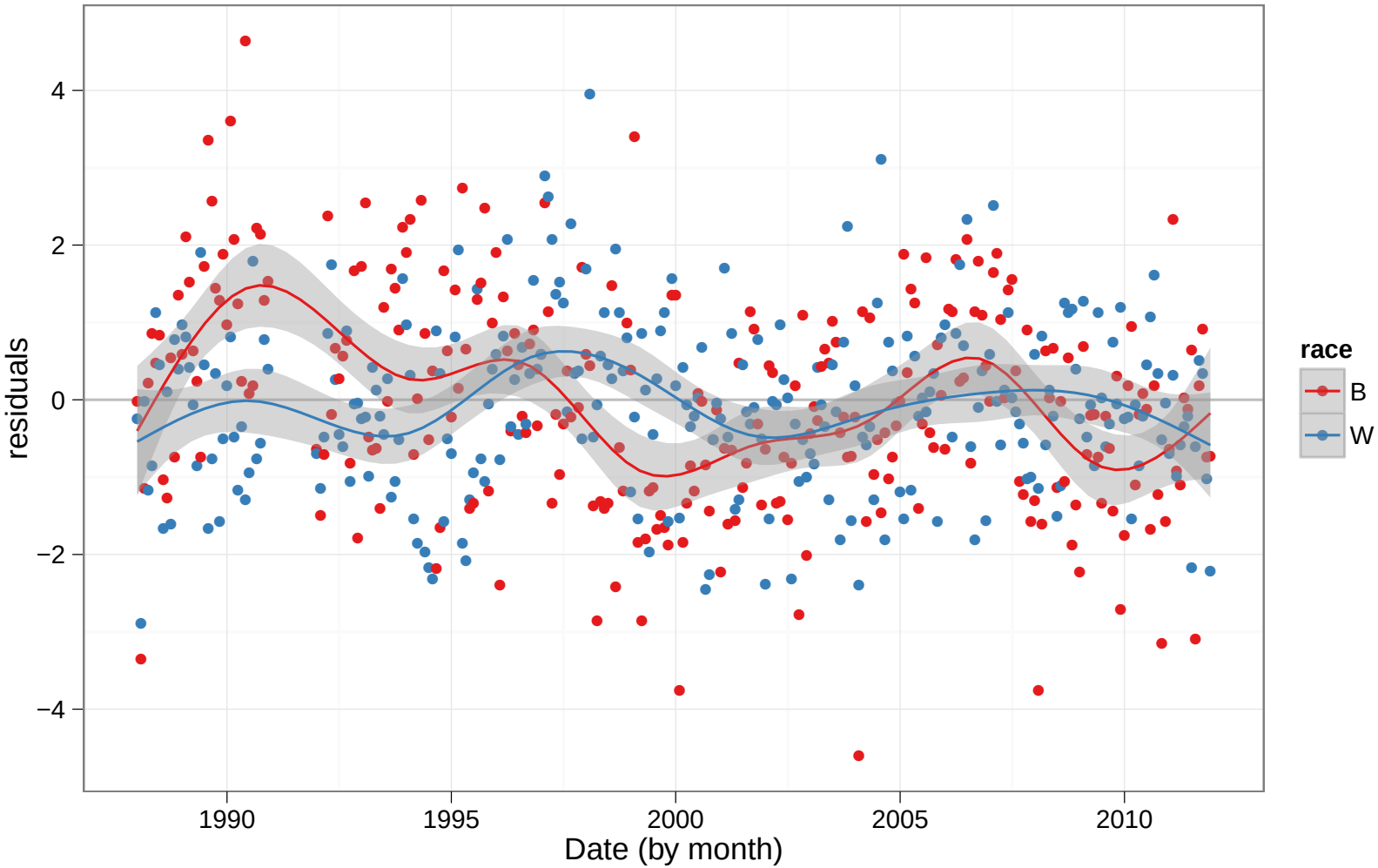


Two Philadelphias

```
model1 <- glm(freq ~ month * race,
    offset = ndays, family = poisson, data = philly.bw.count)
anova(model1, test = "Chisq")

## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: freq
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                         549       7593
## month      11     2913        538       4680   <2e-16 ***
## race        1     3795        537        886   <2e-16 ***
## month:race 11        9        526        876     0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two Philadelphias



Two Philadelphias

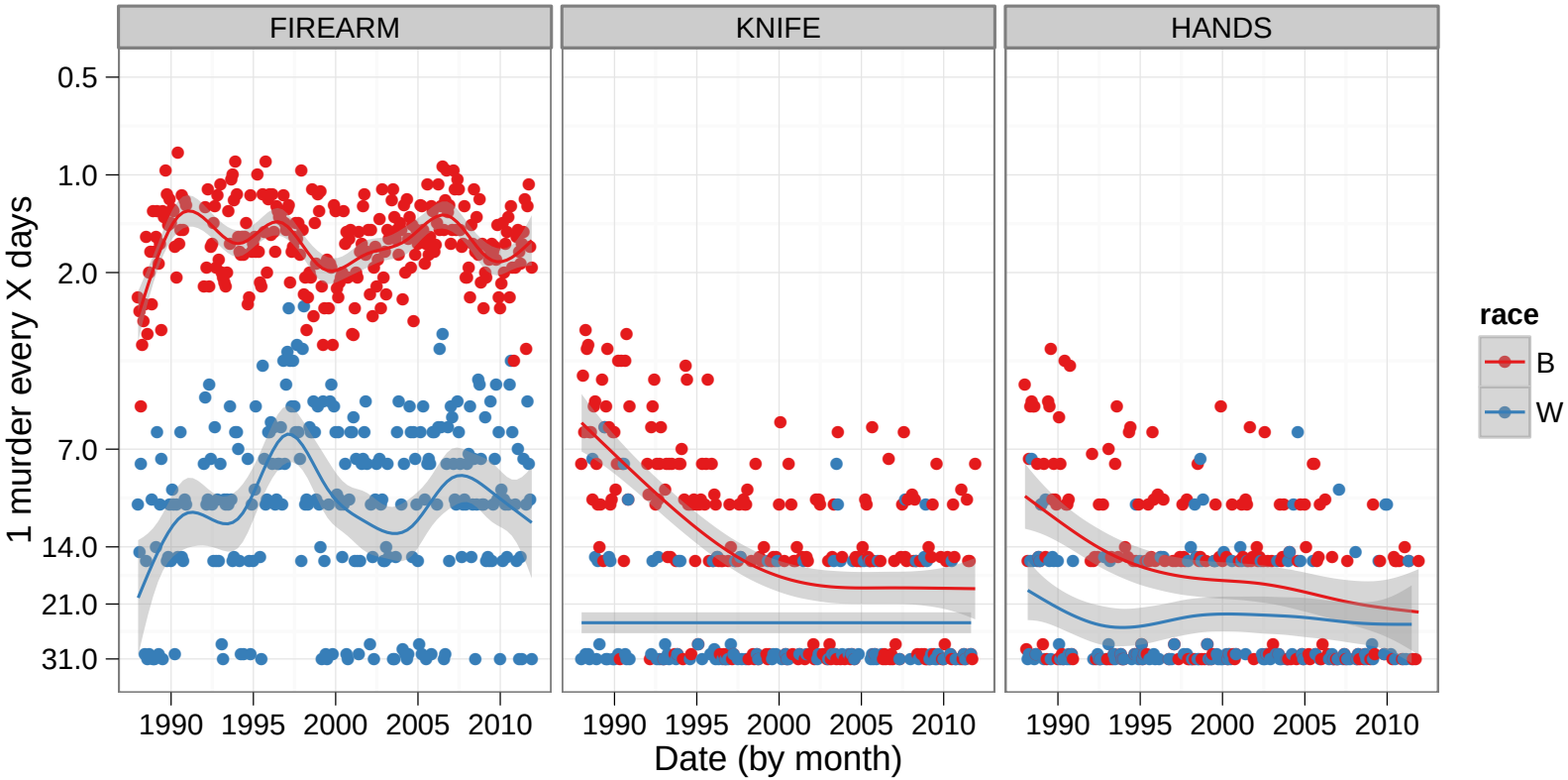
```
philly.bw.count.w <- count(philly.bw,
    c("month.date", "month", "race", "weapon",
        "ndays"))
philly.bw.count.w <- transform(philly.bw.count.w,
    weapon = reorder(weapon, -freq, sum))
philly.bw.count.w <- subset(philly.bw.count.w,
    weapon %in% c("FIREARM", "KNIFE", "HANDS"))
head(philly.bw.count.w)

##   month.date month race  weapon ndays freq
## 1 1988-01-01   Jan    B FIREARM    31   13
## 2 1988-01-01   Jan    B   HANDS    31    7
## 3 1988-01-01   Jan    B   KNIFE    31    4
## 4 1988-01-01   Jan    W FIREARM    31    3
## 5 1988-01-01   Jan    W   KNIFE    31    1
## 7 1988-02-01   Feb    B FIREARM    29   11
```

```
## To capture 0 counts
months.grid <- expand.grid(month.date =
unique(philly.bw$month.date),
    race = c("B", "W"), weapon = c("FIREARM",
        "KNIFE", "HANDS"))
months.grid <- join(months.grid, ndays)
months.grid$month <-
month.abb[as.POSIXlt(months.grid$month.date)$mon +
    1]

philly.bw.count.w <- merge(months.grid,
    philly.bw.count.w, all.x = T)
philly.bw.count.w$freq[is.na(philly.bw.count.w$freq)] <- 0
```

Two Philadelphias



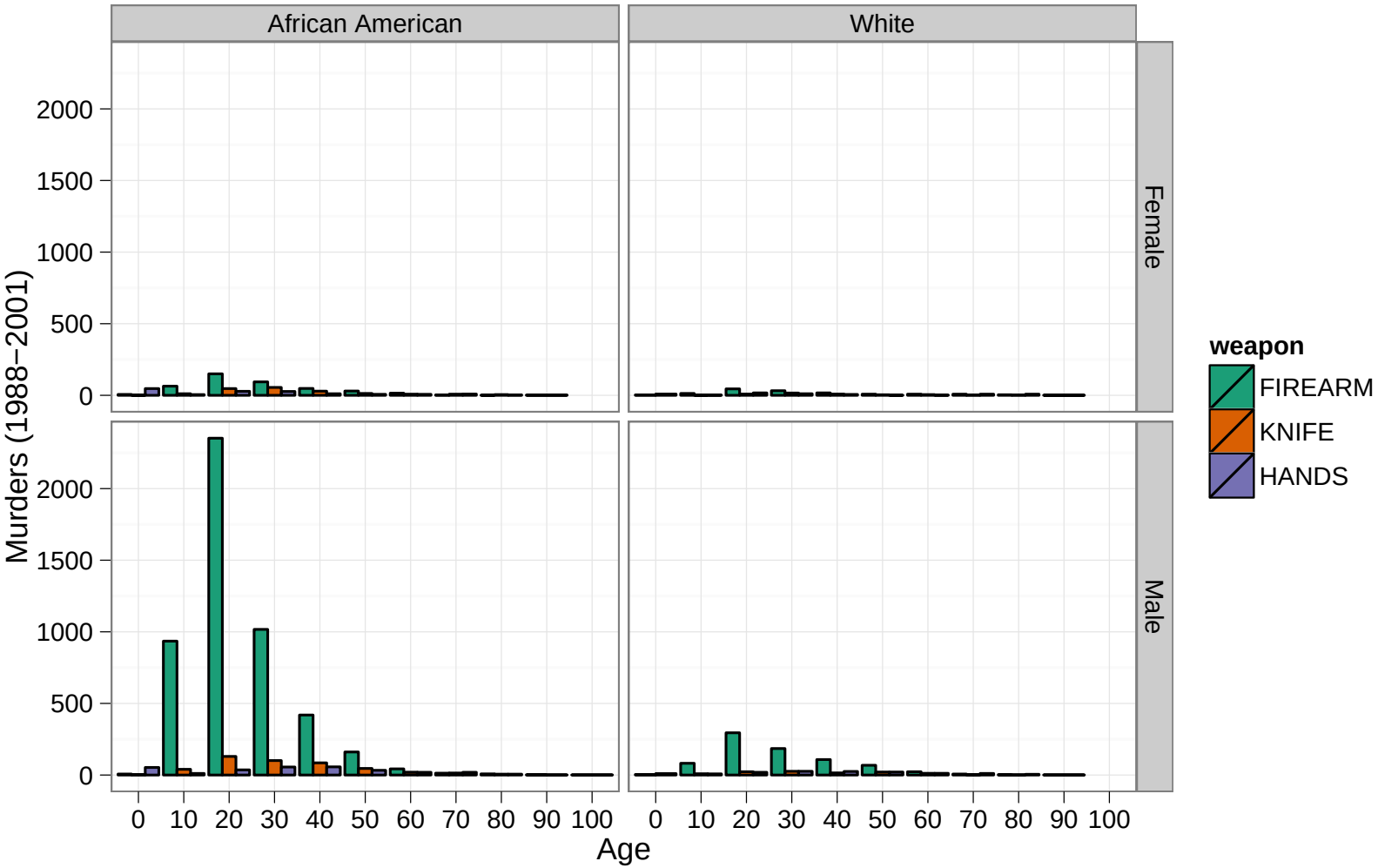
```
model2 <- glm(freq ~ month + weapon *
    race, offset = ndays, data = philly.bw.count.w,
    family = poisson)
anova(model2, test = "Chisq")

## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: freq
##
## Terms added sequentially (first to last)
##
##
##             Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                         1655      16831
## month       11     2841      1644      13990   <2e-16 ***
## weapon       2     7596      1642       6394   <2e-16 ***
## race         1     3766      1641       2628   <2e-16 ***
## weapon:race  2      116      1639       2512   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two Philadelphias

- White murder victims are 5.71$\times$ more likely to be shot than stabbed.
- African American murder victims are 8.62$\times$ more likely to be shot than stabbed.
- African American murder victims are 1.51$\times$ more likely to have been shot than White murder victims.

Two Philadelphias



Thanks