L.A. homeless arrests analysis

Christine Zhang

```
#install.packages('dplyr','feather','ggplot2')
library('dplyr')
library('feather')
library('ggplot2')
```

```
unzip("arrests.zip")
```

```
#load data as feather file type
data <- read_feather('arrests.feather')
```

```
# view column headers
names(data)
```

```
##  [1] "booking_num" "homeless"    "arrest_year" "arrest_ymd"  "booking_ymd"
##  [6] "gender"      "race"        "age"         "occupation"  "charge_code"
## [11] "charge_desc"
```

```
# view first few rows of data set
head(data)
```

```
## # A tibble: 6 x 11
##   booking_num homeless arrest_year arrest_ymd booking_ymd gender race
##         <int>    <dbl>       <dbl> <date>     <date>      <chr>  <chr>
## 1     2497688        0        2011 2011-01-01 2011-01-01  M      W
## 2     2497689        0        2011 2011-01-01 2011-01-01  M      H
## 3     2497690        0        2011 2011-01-01 2011-01-01  M      W
## 4     2497697        0        2011 2011-01-01 2011-01-01  F      W
## 5     2497698        0        2011 2011-01-01 2011-01-01  M      H
## 6     2497699        0        2011 2011-01-01 2011-01-01  F      H
## # ... with 4 more variables: age <dbl>, occupation <chr>,
## #   charge_code <chr>, charge_desc <chr>
```

Finding 1: The LAPD made 14,000 arrests of homeless people last year, a 30% increase over 2011

Group the data by arrest year and homeless and sum the total number of arrests

```
arrest.totals <- data %>%
  group_by(arrest_year, homeless) %>%
  distinct(booking_num) %>%
  summarize(arrests_number = n())
```

Filter to homeless arrests

```
homeless.totals <- arrest.totals %>% filter(homeless == 1)
homeless.totals
```

```
## # A tibble: 6 x 3
## # Groups:   arrest_year [6]
##   arrest_year homeless arrests_number
```

```
##         <dbl>    <dbl>          <int>
## 1        2011     1.00          10496
## 2        2012     1.00          11837
## 3        2013     1.00          12237
## 4        2014     1.00          12622
## 5        2015     1.00          13418
## 6        2016     1.00          14011
```

```
print(paste0("The *raw* increase in homeless arrests between 2011 and 2016 is ",
            round((homeless.totals[homeless.totals$arrest_year == 2016,]$arrests_number /
            homeless.totals[homeless.totals$arrest_year == 2011,]$arrests_number - 1) * 100), "%"))
```

```
## [1] "The *raw* increase in homeless arrests between 2011 and 2016 is 33%"
```

Fix the missing data issue:

```
# extract the booking dates as a vector
booking.dates <- data %>% select(booking_ymd)
booking.dates <- booking.dates %>% distinct(booking_ymd) %>% select(booking_ymd) %>% arrange(booking_ymd
booking.dates$has.data = 1

# get the time period (minimum date and maximum date) of the data set
time.min <- booking.dates$booking_ymd[1]
time.max <- booking.dates$booking_ymd[length(booking.dates$booking_ymd) - 1]

# create a dataframe of all the days spanning that time period
all.dates.frame <- data.frame(list(booking_ymd = seq(time.min, time.max, by="day")))

# merge this dataframe with the vector of booking dates to find the missing dates
merged.data <- merge(all.dates.frame, booking.dates , all=T)
missing.dates <- merged.data %>% filter(is.na(has.data) == T)
```

```
#view missing.dates
missing.dates
```

```
##   booking_ymd has.data
## 1  2011-02-20       NA
## 2  2011-03-17       NA
## 3  2011-04-17       NA
## 4  2011-05-18       NA
## 5  2011-06-11       NA
## 6  2011-08-04       NA
```

Pro-rate the 2011 figure to account for the missing six days

```
prorated.homeless.2011 <- homeless.totals[homeless.totals$arrest_year == 2011,]$arrests_number/(365 - 6
prorated.homeless.2011
```

```
## [1] 10671.42
```

```
print(paste0("The *prorated* change in homeless arrests between 2011 and 2016 is ",
            round((homeless.totals[homeless.totals$arrest_year == 2016,]$arrests_number /
            prorated.homeless.2011 - 1) * 100), "% (rounded down to 30% in the story)"))
```

```
## [1] "The *prorated* change in homeless arrests between 2011 and 2016 is 31% (rounded down to 30% in
```

Finding 2: LAPD arrests overall went down 15% from 2011 to 2016

```
all.totals <- arrest.totals %>%
  group_by(arrest_year) %>%
  summarize(arrests_number = sum(arrests_number))
```

```
print(all.totals)
```

```
## # A tibble: 6 x 2
##   arrest_year arrests_number
##         <dbl>          <int>
## 1        2011          96701
## 2        2012         101234
## 3        2013          98126
## 4        2014          94077
## 5        2015          87067
## 6        2016          83608
```

```
print(paste0("The *raw* change in overall arrests between 2011 and 2016 is ",
            round((all.totals[all.totals$arrest_year == 2016,]$arrests_number /
            all.totals[all.totals$arrest_year == 2011,]$arrests_number - 1) * 100), "%"))
```

```
## [1] "The *raw* change in overall arrests between 2011 and 2016 is -14%"
```

```
# Again, we need to pro-rate to take into account the six missing days in 2011.
prorated.arrests.2011 <- all.totals[all.totals$arrest_year == 2011,]$arrests_number/(365 - 6) * 365
prorated.arrests.2011
```

```
## [1] 98317.17
```

```
print(paste0("The *prorated* change between 2011 and 2016 is ",
            round((all.totals[all.totals$arrest_year == 2016,]$arrests_number /
            prorated.arrests.2011 - 1) * 100), "%"))
```

```
## [1] "The *prorated* change between 2011 and 2016 is -15%"
```

Finding 3: Two-thirds of those arrested were black or Latino

```
arrests.race <- data %>%
  group_by(arrest_year, homeless, race)  %>%
  distinct(booking_num)
```

```
# Create a variable, race.grp to represent racial/ethnic grouping, where W = white, B = black, H = Lati

table(arrests.race$race)
```

```
##
##      A      B      C      F      H      I      J      K      O      P
##     41 171726    431    541 253265     81    102      4  32248    104
##      W
## 102270
```

```
#rename variables
arrests.race$race.grp <- ifelse(arrests.race$race == 'W', "White",
                                ifelse(arrests.race$race == 'B', "Black",
```

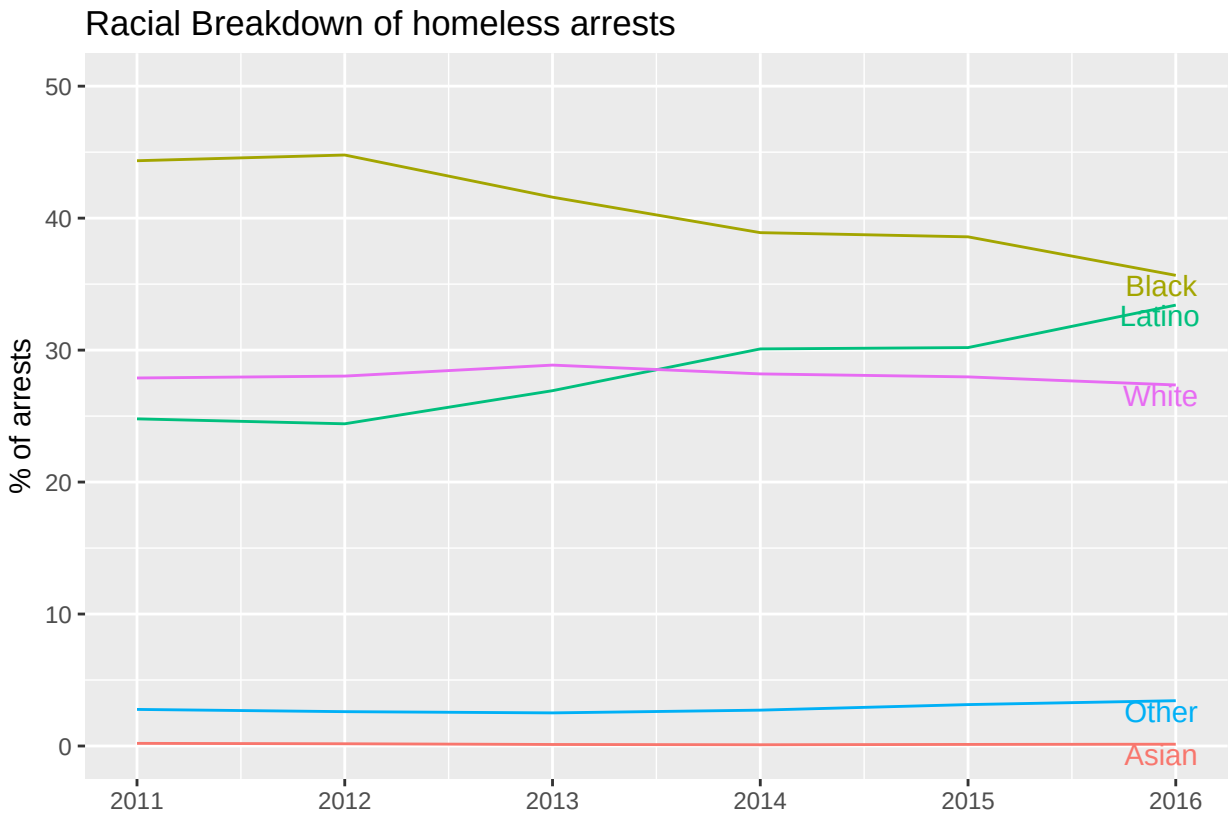
```
                                        ifelse(arrests.race$race == 'H', "Latino",
                                               ifelse(arrests.race$race == 'A' | arrests.race$race == 'C
                                                      'Other'))))
```

```
#Group by race.grp and calculate the total number and percentage of homeless arrests
arrests.race.yr <- arrests.race %>%
  group_by(arrest_year, homeless, race.grp) %>%
  summarize(arrests_number = n()) %>%
  mutate(arrests_percent = arrests_number / sum(arrests_number) * 100)

arrests.race.yr %>% filter(homeless == 1 & arrest_year == 2016) %>% arrange(desc(arrests_percent))
```

```
## # A tibble: 5 x 5
## # Groups:   arrest_year, homeless [1]
##   arrest_year homeless race.grp arrests_number arrests_percent
##         <dbl>    <dbl> <chr>             <int>           <dbl>
## 1        2016     1.00 Black              4997          35.7
## 2        2016     1.00 Latino             4681          33.4
## 3        2016     1.00 White              3833          27.4
## 4        2016     1.00 Other               481           3.43
## 5        2016     1.00 Asian                19           0.136
```

```
ggplot(arrests.race.yr %>% filter(homeless == 1 & arrest_year != 2017), aes(x = arrest_year,
                                                                             y = arrests_percent, color =
  geom_line() +
  geom_text(data = arrests.race.yr %>% filter(homeless == 1 &
                                               arrest_year == 2016), aes(label = race.grp), hjust = 0.7,
            vjust = 1) +
  scale_y_continuous(limits = c(0, 50)) +
  labs(x = "", y = "% of arrests", title = "Racial Breakdown of homeless arrests") +
  theme(legend.position = 'none')
```



Finding 4: In 2011, one in 10 people arrests citywide were of homeless people; in 2016, it was 1 in 6

```
# Use the grouped dataframe arrest.totals to calculate percentage of homeless arrests by year
arrest.totals %>%
    mutate(arrests_percent = arrests_number / sum(arrests_number) * 100) %>%
    filter(homeless == 1)
```

```
## # A tibble: 6 x 4
## # Groups:   arrest_year [6]
##   arrest_year homeless arrests_number arrests_percent
##         <dbl>    <dbl>          <int>           <dbl>
## 1        2011     1.00          10496            10.9
## 2        2012     1.00          11837            11.7
## 3        2013     1.00          12237            12.5
## 4        2014     1.00          12622            13.4
## 5        2015     1.00          13418            15.4
## 6        2016     1.00          14011            16.8
```

Finding 5: The 14,000 arrests of homeless people in 2016 included more than 500 unique charges

```
#Filter the data to include homeless arrests in 2016 and calculate the number and percent of times each

arrest.reasons <- data %>% filter(homeless == 1 & arrest_year == 2016) %>%
                           group_by(charge_code, charge_desc) %>%
                           summarize(times_cited = n()) %>%
                           ungroup() %>%
```

```
                            mutate(percent_cited = times_cited/sum(times_cited) * 100)
```

```
#get the number of unique charges
length(unique(arrest.reasons$charge_code))
```

```
## [1] 523
```

Finding 6: The most common offense was failure to appear in court for unpaid petty or minor citations

```
#Sort by percent of the time the charge was cited to get the top charges

head(arrest.reasons %>% arrange(desc(percent_cited)))
```

```
## # A tibble: 6 x 4
##   charge_code charge_desc                      times_cited percent_cited
##   <chr>       <chr>                                  <int>         <dbl>
## 1 853.7PC     fta-after-written-promise               4447         21.1
## 2 11377(A)HS  possession-controlled-substance          924          4.39
## 3 459.5PC     ""                                       674          3.20
## 4 3000.08CPC  ""                                       613          2.91
## 5 3454(C)PC   ""                                       579          2.75
## 6 245(A)(1)PC adw-wo-firearmgbi                        530          2.52
```

Many codes did not come with charge descriptions in the data. Those that appear in the above table are described as follows:

- 459.5PC: shoplifting
- 3000.08CPC: parole warrant
- 3454(C)PC: flash incarceration

Finding 7: The top five charges were for non-violent or minor offenses

Some charge codes are grouped. For example. charge codes 40508(A)VC, 853.7PC, and 853.8PC all cover failure to appear.

```
arrest.reasons$failure <- ifelse(arrest.reasons$charge_code == '40508(A)VC'|
                                 arrest.reasons$charge_code == '853.7PC'|
                                 arrest.reasons$charge_code == '853.8PC', 1, 0)

arrest.reasons$trespass <- ifelse(arrest.reasons$charge_code == '419PC'|
                                  arrest.reasons$charge_code == '602(K)PC'|
                                  arrest.reasons$charge_code == '602(O)(2)PC'|
                                  arrest.reasons$charge_code == '602.5(A)PC'|
                                  arrest.reasons$charge_code == '555PC'|
                                  arrest.reasons$charge_code == '484F(A)PC'|
                                  arrest.reasons$charge_code == '602(L)(1)PC'|
                                  arrest.reasons$charge_code == '602(P)PC'|
                                  arrest.reasons$charge_code == '602.5(B)PC'|
                                  arrest.reasons$charge_code == '602PC'|
```

```
                                arrest.reasons$charge_code == '602(M)PC'|
                                arrest.reasons$charge_code == '602(Q)PC'|
                                arrest.reasons$charge_code == '602.8PC'|
                                arrest.reasons$charge_code == '602(A)PC'|
                                arrest.reasons$charge_code == 'A602(N)1PC'|
                                arrest.reasons$charge_code == '602(S)PC'|
                                arrest.reasons$charge_code == '626.8(A)1PC'|
                                arrest.reasons$charge_code == '602(D)PC'|
                                arrest.reasons$charge_code == '602(N)PC'|
                                arrest.reasons$charge_code == '602(U)(1)PC'|
                                arrest.reasons$charge_code == '647(E)PC'|
                                arrest.reasons$charge_code == '602(F)PC'|
                                arrest.reasons$charge_code == '602(O)PC'|
                                arrest.reasons$charge_code == '602.1(A)PC'|
                                arrest.reasons$charge_code == '647(H)PCLPP'|
                                arrest.reasons$charge_code == '602(J)PC'|
                                arrest.reasons$charge_code == '602(O)(1)PC'|
                                arrest.reasons$charge_code == '602.1(B)PC'|
                                arrest.reasons$charge_code == '369I(A)PC', 1, 0)

arrest.reasons$shoplift <- ifelse(arrest.reasons$charge_code == '18 1708'|
                                arrest.reasons$charge_code == '484PCTFMV'|
                                arrest.reasons$charge_code == '485PC'|
                                arrest.reasons$charge_code == '488PC'|
                                arrest.reasons$charge_code == '459.5PC'|
                                arrest.reasons$charge_code == '484F(A)PC'|
                                arrest.reasons$charge_code == 'A488PC'|
                                arrest.reasons$charge_code == '490PC'|
                                arrest.reasons$charge_code == 'A484PC'|
                                arrest.reasons$charge_code == '484E(D)PC'|
                                arrest.reasons$charge_code == '666PC'|
                                arrest.reasons$charge_code == '484PC'|
                                arrest.reasons$charge_code == '490.2PC'|
                                arrest.reasons$charge_code == '666(A)PC'|
                                arrest.reasons$charge_code == '484(A)PC'|
                                arrest.reasons$charge_code == '490.5(A)PC'|
                                arrest.reasons$charge_code == '537(A)(1)PC'|
                                arrest.reasons$charge_code == '666.5PC'|
                                arrest.reasons$charge_code == '484E(A)PC'|
                                arrest.reasons$charge_code == '587CPC'|
                                arrest.reasons$charge_code == '666.5(A)PC'|
                                arrest.reasons$charge_code == '484E(B)PC', 1, 0)

arrest.reasons$supervision_viol <- ifelse(arrest.reasons$charge_code == '1203.2PC'|
                                        arrest.reasons$charge_code == '3000.08CPC'|
                                        arrest.reasons$charge_code == '3454(C)PC'|
                                        arrest.reasons$charge_code == '3455(B)1PC'|
                                        arrest.reasons$charge_code == '1203.2(A)PC'|
                                        arrest.reasons$charge_code == '3056PC'|
                                        arrest.reasons$charge_code == '3455(A)4PC'|
                                        arrest.reasons$charge_code == '3455(C)PC'|
                                        arrest.reasons$charge_code == '3000.08FPC'|
                                        arrest.reasons$charge_code == '3454PC'|
```

```
                                        arrest.reasons$charge_code == '3455(A)PC'|
                                        arrest.reasons$charge_code == '18 3606US', 1, 0)

arrest.reasons$drug_poss <- ifelse(arrest.reasons$charge_code == '11377(A)HS'|
                                   arrest.reasons$charge_code == '11377(A)1HS'|
                                   arrest.reasons$charge_code == '11377HS'|
                                   arrest.reasons$charge_code == '11350(A)HS'|
                                   arrest.reasons$charge_code == '11350HS'|
                                   arrest.reasons$charge_code == '11357HS'|
                                   arrest.reasons$charge_code == '11357(A)HS'|
                                   arrest.reasons$charge_code == '11357(B)HS'|
                                   arrest.reasons$charge_code == '11357(C)HS'|
                                   arrest.reasons$charge_code == '4573.6PC'|
                                   arrest.reasons$charge_code == '11550(A)HS'|
                                   arrest.reasons$charge_code == '11375(B)2HS'|
                                   arrest.reasons$charge_code == '11351HS'|
                                   arrest.reasons$charge_code == '4060BP', 1, 0)

arrest.reasons$charge_desc_grouped <- ifelse(arrest.reasons$drug_poss == 1, 'drug_poss',
                                             ifelse(arrest.reasons$trespass == 1, 'trespass',
                                                    ifelse(arrest.reasons$shoplift == 1, 'shoplift',
                                                           ifelse(arrest.reasons$supervision_viol == 1,
                                                                  ifelse(arrest.reasons$failure == 1, '
                                                                         arrest.reasons$charge_code))))
```

Get top five offenses using charge_desc_grouped as the charge identifier

```
arrest.reasons %>% group_by(charge_desc_grouped) %>%
    summarise(times_cited = sum(times_cited)) %>%
    mutate(perc_cited = times_cited/sum(times_cited) * 100) %>% arrange(desc(times_cited)) %>% head(5)
```

```
## # A tibble: 5 x 3
##   charge_desc_grouped   times_cited perc_cited
##   <chr>                       <int>      <dbl>
## 1 failure to appear            4576      21.8
## 2 drug_poss                    2147      10.2
## 3 supervision violation        2085       9.91
## 4 shoplift                     1650       7.85
## 5 trespass                     1056       5.02
```