LA_crime v. 0.1

Andrey Korotkiy

May 30, 2019

1. Data Wrangling

Librarys

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.1.1     v purrr   0.3.2
## v tibble  2.1.1     v dplyr   0.8.1
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df <- read_csv('~/Desktop/Data science/Datasets/LA Crime Data/Crime_Data_from_2010_to_Present.csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `Crime Code` = col_double(),
##   `Victim Age` = col_double(),
##   `Premise Code` = col_double(),
##   `Weapon Used Code` = col_double(),
##   `Crime Code 1` = col_double(),
##   `Crime Code 2` = col_double(),
##   `Crime Code 3` = col_double(),
##   `Crime Code 4` = col_logical()
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 93 parsing failures.
##    row          col           expected actual
##   6348 Crime Code 4 1/0/T/F/TRUE/FALSE    910 '~/Desktop/Data science/Datasets/LA Crime Data/Crime_Da
##   8833 Crime Code 4 1/0/T/F/TRUE/FALSE    998 '~/Desktop/Data science/Datasets/LA Crime Data/Crime_Da
##  74784 Crime Code 4 1/0/T/F/TRUE/FALSE    998 '~/Desktop/Data science/Datasets/LA Crime Data/Crime_Da
##  87094 Crime Code 4 1/0/T/F/TRUE/FALSE    998 '~/Desktop/Data science/Datasets/LA Crime Data/Crime_Da
## 105601 Crime Code 4 1/0/T/F/TRUE/FALSE    930 '~/Desktop/Data science/Datasets/LA Crime Data/Crime_Da
## ...... ............ .................. ...... ....................................................
## See problems(...) for more details.
```

Figure 1:

```
str(df)
```

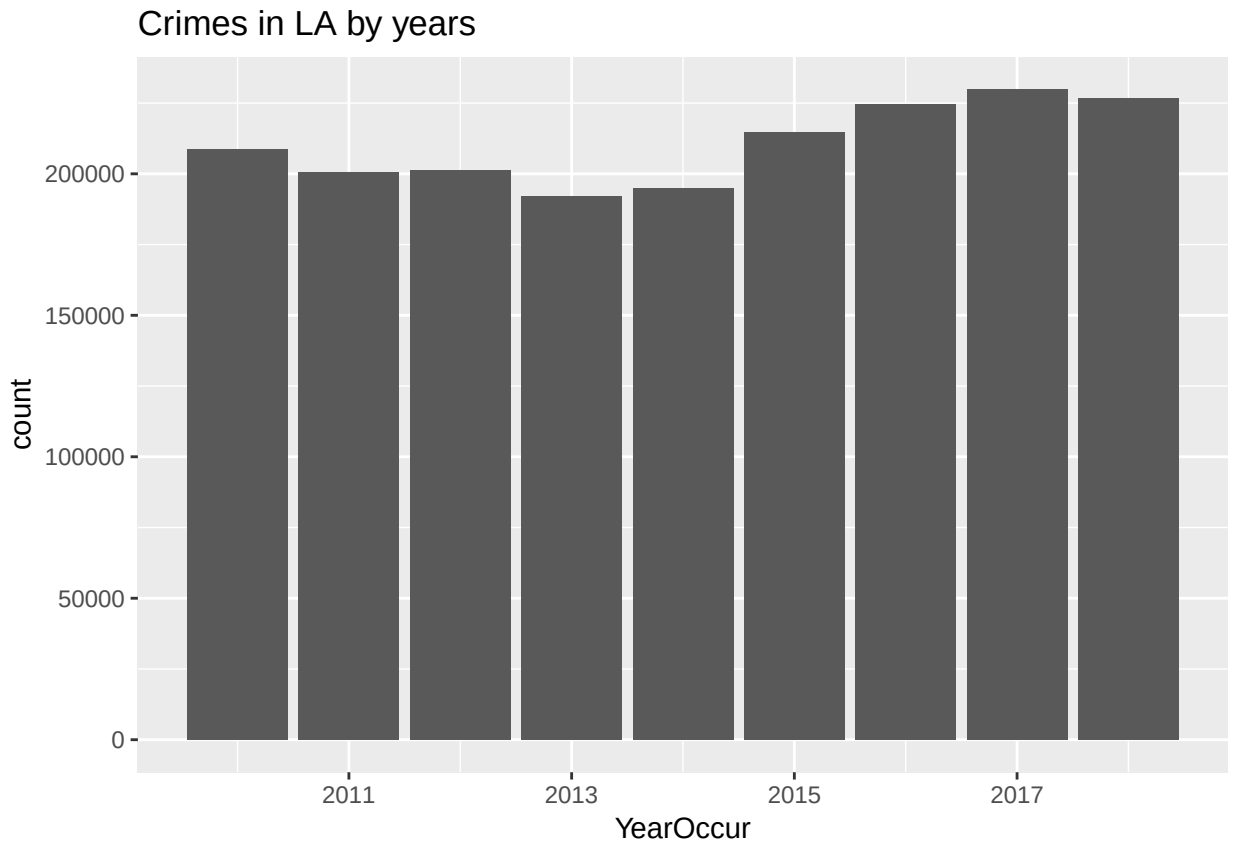
```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 1893487 obs. of  26 variables:
##  $ DR Number             : chr  "011401303" "070309629" "100100508" "100100511" ...
##  $ Date Reported         : chr  "09/13/2010" "08/09/2010" "01/08/2010" "01/09/2010" ...
##  $ Date Occurred         : chr  "09/12/2010" "08/09/2010" "01/07/2010" "01/06/2010" ...
##  $ Time Occurred         : chr  "0045" "1515" "2005" "2100" ...
##  $ Area ID               : chr  "14" "13" "01" "01" ...
##  $ Area Name             : chr  "Pacific" "Newton" "Central" "Central" ...
##  $ Reporting District    : chr  "1485" "1324" "0182" "0132" ...
##  $ Crime Code            : num  740 946 330 341 740 237 626 341 442 341 ...
##  $ Crime Code Description: chr  "VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)" "OTHER MIS
##  $ MO Codes              : chr  "0329" "0344" "0344" "0344 1402" ...
##  $ Victim Age            : num  0 0 46 55 24 11 28 36 23 0 ...
##  $ Victim Sex            : chr  "M" "M" "M" "M" ...
##  $ Victim Descent        : chr  "W" "H" "H" "W" ...
##  $ Premise Code          : num  101 103 101 710 102 502 108 710 404 102 ...
##  $ Premise Description   : chr  "STREET" "ALLEY" "STREET" "OTHER PREMISE" ...
##  $ Weapon Used Code      : num  NA NA NA NA NA NA 400 NA NA NA ...
##  $ Weapon Description    : chr  NA NA NA NA ...
##  $ Status Code           : chr  "IC" "IC" "IC" "IC" ...
##  $ Status Description    : chr  "Invest Cont" "Invest Cont" "Invest Cont" "Invest Cont" ...
##  $ Crime Code 1          : num  740 946 330 341 740 237 626 341 442 341 ...
##  $ Crime Code 2          : num  NA NA NA 998 NA NA NA NA 998 NA ...
##  $ Crime Code 3          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Crime Code 4          : logi  NA NA NA NA NA NA ...
##  $ Address               : chr  "SEPULVEDA                    BL" "1300 E  21ST
##  $ Cross Street          : chr  "MANCHESTER                   AV" NA "GRAND                          A
```

```
##  $ Location              : chr  "(33.9599, -118.3962)" "(34.0224, -118.2524)" "(34.0389, -118.2643)"
##  - attr(*, "problems")=Classes 'tbl_df', 'tbl' and 'data.frame': 93 obs. of  5 variables:
##   ..$ row     : int  6348 8833 74784 87094 105601 113403 130868 134535 158597 205980 ...
##   ..$ col     : chr  "Crime Code 4" "Crime Code 4" "Crime Code 4" "Crime Code 4" ...
##   ..$ expected: chr  "1/0/T/F/TRUE/FALSE" "1/0/T/F/TRUE/FALSE" "1/0/T/F/TRUE/FALSE" "1/0/T/F/TRUE/FA
##   ..$ actual  : chr  "910" "998" "998" "998" ...
##   ..$ file    : chr  "'~/Desktop/Data science/Datasets/LA Crime Data/Crime_Data_from_2010_to_Present
##  - attr(*, "spec")=
##   .. cols(
##   ..   `DR Number` = col_character(),
##   ..   `Date Reported` = col_character(),
##   ..   `Date Occurred` = col_character(),
##   ..   `Time Occurred` = col_character(),
##   ..   `Area ID` = col_character(),
##   ..   `Area Name` = col_character(),
##   ..   `Reporting District` = col_character(),
##   ..   `Crime Code` = col_double(),
##   ..   `Crime Code Description` = col_character(),
##   ..   `MO Codes` = col_character(),
##   ..   `Victim Age` = col_double(),
##   ..   `Victim Sex` = col_character(),
##   ..   `Victim Descent` = col_character(),
##   ..   `Premise Code` = col_double(),
##   ..   `Premise Description` = col_character(),
##   ..   `Weapon Used Code` = col_double(),
##   ..   `Weapon Description` = col_character(),
##   ..   `Status Code` = col_character(),
##   ..   `Status Description` = col_character(),
##   ..   `Crime Code 1` = col_double(),
##   ..   `Crime Code 2` = col_double(),
##   ..   `Crime Code 3` = col_double(),
##   ..   `Crime Code 4` = col_logical(),
##   ..   Address = col_character(),
##   ..   `Cross Street` = col_character(),
##   ..   Location = col_character()
##   .. )
```

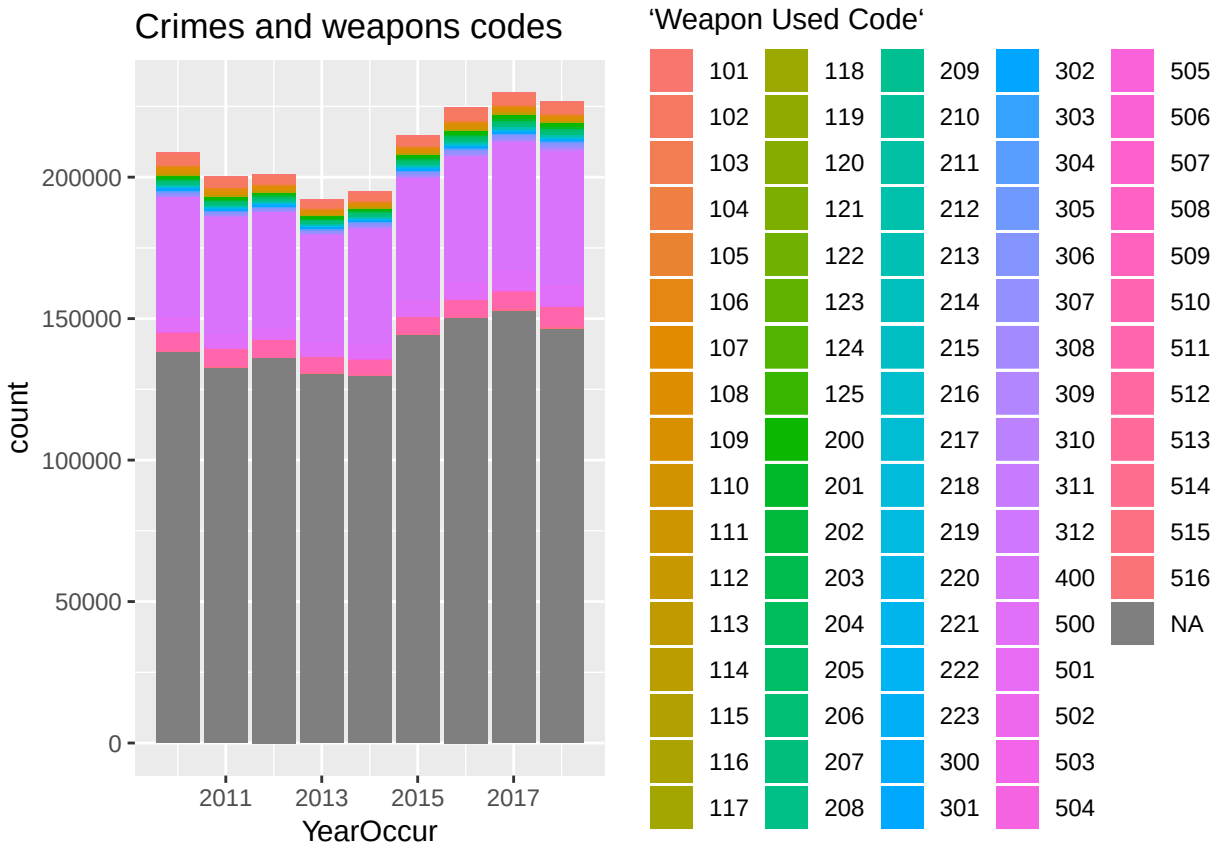
```
df$`Date Reported` <- as.Date(df$`Date Reported`, format = "%m/%d/%Y")
df$`Date Occurred` <- as.Date(df$`Date Occurred`, format = "%m/%d/%Y")
df$`Weapon Used Code` <- as.factor(df$`Weapon Used Code`)

df$YearOccur  <- as.numeric(format(df$`Date Occurred`,'%Y'))
```

```
ggplot(df, aes(x= YearOccur)) +
  geom_bar() +
  ggtitle("Crimes in LA by years")
```




```
ggplot(df, aes(x= YearOccur, fill = `Weapon Used Code`)) +
  geom_bar() +
  ggtitle("Crimes and weapons codes")
```


Crimes and weapons codes
count
200000
150000
100000
50000
0
2011
2013
2015
2017
YearOccur
'Weapon Used Code'
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
300
301
302
303
304
305
306
307
308
309
310
311
312
400
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
NA