A3

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

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

```
library(dplyr)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(knitr)
library(devtools)
```

```
## Loading required package: usethis
```

1

```
library(ggplot2)
library(tidyr)
```
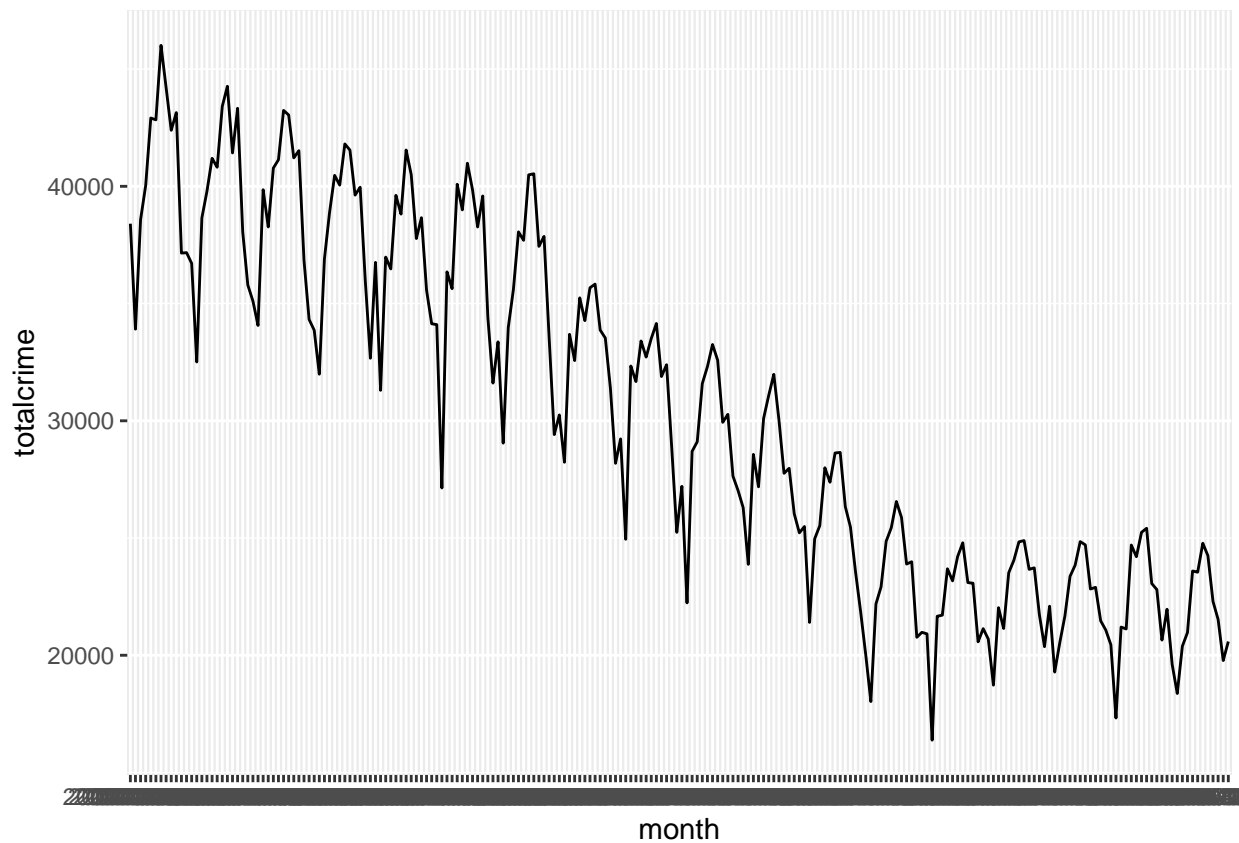
Exercise 1

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Exercise 2

```
a<-aggregate(crime_long$crimes, by=list(month=crime_long$crime_month), FUN=sum)
colnames(a)<-c("month","totalcrime")
head(a)
```

```
##        month totalcrime
## 1 2002-01-01      38405
## 2 2002-02-01      33909
## 3 2002-03-01      38583
## 4 2002-04-01      40032
## 5 2002-05-01      42913
## 6 2002-06-01      42834
```

```
ggplot(a, aes(x = month, y = totalcrime, group = 1))+geom_line()
```

```
e2<-left_join(population,crime_long, by=c("month"="crime_month","district"="district"))
head(e2)
```

```
##         month period district tot_pop tot_white tot_black tot_hisp  p50_inc
## 1 2005-01-01      1        1   38472     22608      4953     2543 91084.91
## 2 2005-01-01      1        1   38472     22608      4953     2543 91084.91
## 3 2005-01-01      1        1   38472     22608      4953     2543 91084.91
## 4 2005-01-01      1        1   38472     22608      4953     2543 91084.91
## 5 2005-01-01      1        1   38472     22608      4953     2543 91084.91
## 6 2005-01-01      1        1   38472     22608      4953     2543 91084.91
##   crime_type crimes
## 1       drug      1
## 2       drug    188
## 3      other     62
## 4      other    302
## 5   property    624
## 6   property    160
```

```
e2n<-e2 %>% group_by(month,district, crime_type) %>%
        summarise(total_crime =sum(crimes))
```

```
## `summarise()` has grouped output by 'month', 'district'. You can override using the `.groups` argumen
```

```
head(e2n)
```

```
## # A tibble: 6 x 4
## # Groups:   month, district [2]
##   month      district crime_type total_crime
##   <chr>         <int> <chr>            <int>
## 1 2005-01-01        1 drug               189
## 2 2005-01-01        1 other              364
## 3 2005-01-01        1 property           784
## 4 2005-01-01        1 violent            212
## 5 2005-01-01        2 drug               209
## 6 2005-01-01        2 other              268
```

```
e2n<-spread(e2n, key = crime_type, value = total_crime)
head(e2n)
```

```
## # A tibble: 6 x 7
## # Groups:   month, district [6]
##   month      district  drug other property violent `<NA>`
##   <chr>         <int> <int> <int>    <int>   <int> <int>
## 1 2005-01-01        1   189   364      784     212    NA
## 2 2005-01-01        2   209   268      509     365    NA
## 3 2005-01-01        3   259   309      606     533    NA
## 4 2005-01-01        4   227   282      714     616    NA
## 5 2005-01-01        5   236   273      467     435    NA
## 6 2005-01-01        6   109   308      734     591    NA
```

```r
e2n2<-e2 %>% group_by(month,district) %>%
      summarise(total_crime =sum(crimes))
```

```
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
```

```r
head(e2n2)
```

```
## # A tibble: 6 x 3
## # Groups:   month [1]
##   month      district total_crime
##   <chr>         <int>       <int>
## 1 2005-01-01        1        1549
## 2 2005-01-01        2        1351
## 3 2005-01-01        3        1707
## 4 2005-01-01        4        1839
## 5 2005-01-01        5        1411
## 6 2005-01-01        6        1742
```

```r
e2n1<-e2 %>% group_by(month,district) %>%
  summarise(median_income=median(p50_inc),black_share=sum(tot_black)/sum(tot_pop),hispanic_share=sum(tot
```

```
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
```

```r
head(e2n1)
```

```
## # A tibble: 6 x 6
## # Groups:   month [1]
##   month      district median_income black_share hispanic_share white_share
##   <chr>         <int>         <dbl>       <dbl>          <dbl>       <dbl>
## 1 2005-01-01        1        91085.       0.129         0.0661       0.588
## 2 2005-01-01        2        29890.       0.947         0.0165       0.0166
## 3 2005-01-01        3        28048.       0.913         0.0174       0.0409
## 4 2005-01-01        4        39010.       0.621         0.284        0.0797
## 5 2005-01-01        5        33147.       0.942         0.0339       0.0113
## 6 2005-01-01        6        34672.       0.975         0.0101       0.00345
```

```r
panel<-left_join(e2n,e2n1, by=c("month"="month","district"="district"))
panel<-left_join(panel,e2n2, by=c("month"="month","district"="district"))
panel<-subset(panel, select=-c(other,drug))
panel<-subset(panel, select=-c(5))
head(panel)
```

```
## # A tibble: 6 x 9
## # Groups:   month, district [6]
##   month      district property violent median_income black_share hispanic_share
##   <chr>         <int>    <int>   <int>         <dbl>       <dbl>          <dbl>
## 1 2005-01-01        1      784     212        91085.       0.129         0.0661
## 2 2005-01-01        2      509     365        29890.       0.947         0.0165
## 3 2005-01-01        3      606     533        28048.       0.913         0.0174
## 4 2005-01-01        4      714     616        39010.       0.621         0.284
## 5 2005-01-01        5      467     435        33147.       0.942         0.0339
## 6 2005-01-01        6      734     591        34672.       0.975         0.0101
## # ... with 2 more variables: white_share <dbl>, total_crime <int>
```

Exercise 3

```
panel2<-left_join(panel,officers,by=c("month"="month","district"="unit"))
panel2<-drop_na(panel2)
head(panel2)
```

```
## # A tibble: 6 x 12
## # Groups:   month, district [1]
##   month       district property violent median_income black_share hispanic_share
##   <chr>          <int>    <int>   <int>         <dbl>       <dbl>          <dbl>
## 1 2007-01-01         1      671     187        91085.       0.129         0.0661
## 2 2007-01-01         1      671     187        91085.       0.129         0.0661
## 3 2007-01-01         1      671     187        91085.       0.129         0.0661
## 4 2007-01-01         1      671     187        91085.       0.129         0.0661
## 5 2007-01-01         1      671     187        91085.       0.129         0.0661
## 6 2007-01-01         1      671     187        91085.       0.129         0.0661
## # ... with 5 more variables: white_share <dbl>, total_crime <int>, NUID <int>,
## #   tenure <int>, arrest <int>
```

```
model<-lm(arrest ~tenure+total_crime+median_income+black_share+hispanic_share+white_share,data=panel2)
summary(model)
```

```
##
## Call:
## lm(formula = arrest ~ tenure + total_crime + median_income +
##     black_share + hispanic_share + white_share, data = panel2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5017 -0.4993 -0.4981  0.5008  5.5025
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.067e-01  1.278e-02  39.644   <2e-16 ***
## tenure         -4.161e-06  8.354e-06  -0.498    0.618
## total_crime     2.229e-07  1.805e-06   0.124    0.902
## median_income   1.618e-08  9.186e-08   0.176    0.860
## black_share    -8.102e-03  1.340e-02  -0.604    0.546
## hispanic_share -5.363e-03  1.391e-02  -0.385    0.700
## white_share    -1.207e-02  1.632e-02  -0.740    0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7068 on 1077898 degrees of freedom
## Multiple R-squared:  2.032e-06,  Adjusted R-squared:  -3.535e-06
## F-statistic: 0.365 on 6 and 1077898 DF,  p-value: 0.9014
```

Exercise 4

```
model4<-lm(arrest ~tenure+total_crime+median_income+black_share+hispanic_share+white_share+c(month)+fact
summary(model4)
```

```
##
## Call:
## lm(formula = arrest ~ tenure + total_crime + median_income +
##     black_share + hispanic_share + white_share + c(month) + factor(district),
##     data = panel2, x = FALSE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5282 -0.5003 -0.4920  0.5008  5.5163
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.362e-01  1.061e-01   5.996 2.03e-09 ***
## tenure              -3.810e-06  8.525e-06  -0.447   0.6549
## total_crime         -6.320e-06  5.305e-06  -1.191   0.2335
## median_income       -4.910e-07  6.505e-07  -0.755   0.4504
## black_share         -9.201e-02  1.043e-01  -0.882   0.3775
## hispanic_share      -1.398e-01  2.028e-01  -0.690   0.4904
## white_share         -1.013e-01  1.818e-01  -0.557   0.5773
## c(month)2007-02-01   2.747e-03  1.112e-02   0.247   0.8049
## c(month)2007-03-01   5.425e-03  1.098e-02   0.494   0.6213
## c(month)2007-04-01  -4.433e-03  1.100e-02  -0.403   0.6868
## c(month)2007-05-01   9.642e-03  1.105e-02   0.872   0.3830
## c(month)2007-06-01  -1.518e-02  1.100e-02  -1.380   0.1675
## c(month)2007-07-01  -2.476e-03  1.107e-02  -0.224   0.8230
## c(month)2007-08-01  -9.283e-03  1.101e-02  -0.843   0.3992
## c(month)2007-09-01   2.832e-03  1.096e-02   0.258   0.7962
## c(month)2007-10-01   5.532e-03  1.098e-02   0.504   0.6145
## c(month)2007-11-01   3.270e-03  1.088e-02   0.300   0.7638
## c(month)2007-12-01  -8.522e-03  1.089e-02  -0.783   0.4339
## c(month)2008-01-01  -1.580e-02  1.090e-02  -1.450   0.1470
## c(month)2008-02-01  -7.936e-03  1.096e-02  -0.724   0.4691
## c(month)2008-03-01  -7.042e-03  1.090e-02  -0.646   0.5182
## c(month)2008-04-01   7.880e-03  1.089e-02   0.723   0.4695
## c(month)2008-05-01   1.890e-02  1.091e-02   1.732   0.0833 .
## c(month)2008-06-01   1.043e-03  1.088e-02   0.096   0.9237
## c(month)2008-07-01  -1.992e-05  1.092e-02  -0.002   0.9985
## c(month)2008-08-01   1.040e-03  1.093e-02   0.095   0.9242
## c(month)2008-09-01   9.368e-03  1.086e-02   0.863   0.3882
## c(month)2008-10-01  -3.356e-03  1.085e-02  -0.309   0.7571
## c(month)2008-11-01  -1.542e-03  1.083e-02  -0.142   0.8868
## c(month)2008-12-01   1.093e-02  1.088e-02   1.005   0.3149
## c(month)2009-01-01  -6.007e-03  1.086e-02  -0.553   0.5801
## c(month)2009-02-01  -5.783e-03  1.092e-02  -0.530   0.5964
## c(month)2009-03-01  -8.894e-03  1.085e-02  -0.820   0.4124
## c(month)2009-04-01  -4.911e-03  1.087e-02  -0.452   0.6515
## c(month)2009-05-01  -1.545e-03  1.086e-02  -0.142   0.8869
## c(month)2009-06-01   4.206e-03  1.086e-02   0.387   0.6986
## c(month)2009-07-01   5.471e-03  1.087e-02   0.503   0.6147
## c(month)2009-08-01  -4.667e-03  1.087e-02  -0.429   0.6678
```

```
## c(month)2009-09-01 -4.795e-03  1.087e-02  -0.441   0.6592
## c(month)2009-10-01 -1.586e-03  1.087e-02  -0.146   0.8840
## c(month)2009-11-01 -5.210e-03  1.088e-02  -0.479   0.6322
## c(month)2009-12-01 -9.886e-03  1.095e-02  -0.902   0.3668
## c(month)2010-01-01  1.815e-03  1.094e-02   0.166   0.8682
## c(month)2010-02-01  3.620e-03  1.116e-02   0.324   0.7457
## c(month)2010-03-01 -8.878e-03  1.099e-02  -0.808   0.4192
## c(month)2010-04-01  7.082e-03  1.100e-02   0.644   0.5198
## c(month)2010-05-01 -3.414e-03  1.101e-02  -0.310   0.7565
## c(month)2010-06-01 -5.868e-03  1.102e-02  -0.532   0.5944
## c(month)2010-07-01 -1.353e-02  1.102e-02  -1.228   0.2196
## c(month)2010-08-01  9.875e-03  1.103e-02   0.895   0.3708
## c(month)2010-09-01  2.215e-03  1.103e-02   0.201   0.8408
## c(month)2010-10-01 -2.548e-03  1.103e-02  -0.231   0.8172
## c(month)2010-11-01 -9.175e-03  1.110e-02  -0.827   0.4084
## c(month)2010-12-01 -4.759e-03  1.122e-02  -0.424   0.6715
## c(month)2011-01-01 -8.207e-03  1.119e-02  -0.733   0.4633
## c(month)2011-02-01 -5.389e-03  1.145e-02  -0.471   0.6378
## c(month)2011-03-01 -4.549e-03  1.121e-02  -0.406   0.6849
## c(month)2011-04-01 -1.189e-03  1.120e-02  -0.106   0.9154
## c(month)2011-05-01 -5.307e-04  1.117e-02  -0.048   0.9621
## c(month)2011-06-01  1.207e-04  1.115e-02   0.011   0.9914
## c(month)2011-07-01  4.004e-03  1.113e-02   0.360   0.7191
## c(month)2011-08-01 -1.043e-04  1.114e-02  -0.009   0.9925
## c(month)2011-09-01 -1.135e-02  1.101e-02  -1.031   0.3027
## c(month)2011-10-01  1.049e-02  1.100e-02   0.954   0.3399
## c(month)2011-11-01 -6.974e-03  1.106e-02  -0.631   0.5283
## c(month)2011-12-01  2.879e-04  1.109e-02   0.026   0.9793
## c(month)2012-01-01 -1.045e-02  1.113e-02  -0.939   0.3477
## c(month)2012-02-01  2.973e-03  1.121e-02   0.265   0.7909
## c(month)2012-03-01  1.820e-03  1.105e-02   0.165   0.8692
## c(month)2012-04-01  4.576e-03  1.114e-02   0.411   0.6812
## c(month)2012-05-01  4.361e-03  1.107e-02   0.394   0.6936
## c(month)2012-06-01 -2.065e-03  1.107e-02  -0.187   0.8520
## c(month)2012-07-01  6.781e-04  1.107e-02   0.061   0.9512
## c(month)2012-08-01 -5.126e-03  1.111e-02  -0.461   0.6445
## c(month)2012-09-01  2.322e-03  1.117e-02   0.208   0.8353
## c(month)2012-10-01 -1.458e-02  1.117e-02  -1.305   0.1919
## c(month)2012-11-01  3.463e-03  1.125e-02   0.308   0.7582
## c(month)2012-12-01 -1.687e-02  1.130e-02  -1.494   0.1352
## c(month)2013-01-01  3.795e-03  1.131e-02   0.335   0.7373
## c(month)2013-02-01 -8.078e-03  1.156e-02  -0.699   0.4847
## c(month)2013-03-01 -1.585e-02  1.137e-02  -1.395   0.1631
## c(month)2013-04-01  3.512e-03  1.133e-02   0.310   0.7565
## c(month)2013-05-01  1.411e-02  1.123e-02   1.257   0.2088
## c(month)2013-06-01  7.059e-03  1.125e-02   0.628   0.5303
## c(month)2013-07-01  1.753e-02  1.122e-02   1.563   0.1180
## c(month)2013-08-01 -2.732e-03  1.122e-02  -0.244   0.8076
## c(month)2013-09-01  3.037e-03  1.130e-02   0.269   0.7882
## c(month)2013-10-01 -8.699e-03  1.134e-02  -0.767   0.4430
## c(month)2013-11-01 -5.810e-03  1.144e-02  -0.508   0.6117
## c(month)2013-12-01  3.319e-03  1.155e-02   0.287   0.7738
## c(month)2014-01-01 -1.496e-02  1.171e-02  -1.278   0.2011
## c(month)2014-02-01  2.499e-03  1.186e-02   0.211   0.8331
```

```
## c(month)2014-03-01 -4.441e-03  1.155e-02  -0.385   0.7005
## c(month)2014-04-01 -1.862e-02  1.148e-02  -1.622   0.1049
## c(month)2014-05-01 -5.901e-03  1.137e-02  -0.519   0.6038
## c(month)2014-06-01 -1.761e-03  1.130e-02  -0.156   0.8762
## c(month)2014-07-01  5.575e-03  1.126e-02   0.495   0.6206
## c(month)2014-08-01 -1.113e-03  1.130e-02  -0.098   0.9215
## c(month)2014-09-01  1.055e-03  1.134e-02   0.093   0.9259
## c(month)2014-10-01 -1.730e-03  1.133e-02  -0.153   0.8787
## c(month)2014-11-01 -8.539e-03  1.154e-02  -0.740   0.4591
## c(month)2014-12-01 -7.021e-03  1.153e-02  -0.609   0.5427
## c(month)2015-01-01 -4.922e-03  1.157e-02  -0.425   0.6705
## c(month)2015-02-01 -9.308e-03  1.196e-02  -0.778   0.4363
## c(month)2015-03-01  4.154e-03  1.153e-02   0.360   0.7187
## c(month)2015-04-01 -3.524e-03  1.156e-02  -0.305   0.7604
## c(month)2015-05-01 -6.722e-03  1.142e-02  -0.589   0.5561
## c(month)2015-06-01 -7.111e-03  1.143e-02  -0.622   0.5340
## c(month)2015-07-01 -8.631e-03  1.138e-02  -0.758   0.4483
## c(month)2015-08-01 -5.341e-03  1.136e-02  -0.470   0.6383
## c(month)2015-09-01 -1.012e-03  1.147e-02  -0.088   0.9297
## c(month)2015-10-01 -8.335e-03  1.147e-02  -0.727   0.4673
## c(month)2015-11-01  2.320e-03  1.162e-02   0.200   0.8418
## c(month)2015-12-01 -7.373e-03  1.161e-02  -0.635   0.5253
## c(month)2016-01-01 -5.344e-03  1.166e-02  -0.458   0.6467
## c(month)2016-02-01 -7.099e-03  1.180e-02  -0.601   0.5476
## c(month)2016-03-01 -1.453e-02  1.155e-02  -1.258   0.2084
## c(month)2016-04-01  9.818e-03  1.160e-02   0.846   0.3973
## c(month)2016-05-01  9.858e-03  1.145e-02   0.861   0.3891
## c(month)2016-06-01 -5.147e-03  1.141e-02  -0.451   0.6520
## c(month)2016-07-01 -1.082e-02  1.138e-02  -0.951   0.3417
## c(month)2016-08-01 -1.921e-02  1.134e-02  -1.694   0.0903 .
## c(month)2016-09-01 -1.542e-03  1.141e-02  -0.135   0.8925
## c(month)2016-10-01  3.118e-03  1.141e-02   0.273   0.7847
## c(month)2016-11-01 -1.398e-02  1.155e-02  -1.210   0.2261
## c(month)2016-12-01 -1.290e-02  1.163e-02  -1.109   0.2673
## c(month)2017-01-01  8.784e-05  1.159e-02   0.008   0.9940
## c(month)2017-02-01 -5.838e-03  1.179e-02  -0.495   0.6205
## c(month)2017-03-01  6.054e-03  1.171e-02   0.517   0.6050
## c(month)2017-04-01 -5.738e-03  1.164e-02  -0.493   0.6222
## c(month)2017-05-01  8.410e-03  1.154e-02   0.729   0.4660
## c(month)2017-06-01 -1.219e-02  1.151e-02  -1.059   0.2894
## c(month)2017-07-01 -6.123e-03  1.146e-02  -0.534   0.5930
## c(month)2017-08-01 -2.664e-03  1.148e-02  -0.232   0.8165
## c(month)2017-09-01 -8.493e-03  1.157e-02  -0.734   0.4628
## c(month)2017-10-01 -9.683e-03  1.155e-02  -0.839   0.4017
## c(month)2017-11-01 -1.685e-02  1.166e-02  -1.444   0.1486
## c(month)2017-12-01 -8.924e-03  1.166e-02  -0.765   0.4442
## factor(district)2  -2.438e-02  1.901e-02  -1.282   0.1998
## factor(district)3  -2.001e-02  2.128e-02  -0.940   0.3471
## factor(district)4  -3.633e-05  3.221e-02  -0.001   0.9991
## factor(district)5  -1.580e-02  2.106e-02  -0.750   0.4531
## factor(district)6  -1.386e-02  2.275e-02  -0.609   0.5424
## factor(district)7  -2.109e-02  2.271e-02  -0.928   0.3532
## factor(district)8   2.343e-02  6.598e-02   0.355   0.7226
## factor(district)9   3.110e-03  6.192e-02   0.050   0.9599
```

```
## factor(district)10   7.682e-03   6.920e-02    0.111   0.9116
## factor(district)11  -1.134e-02   2.276e-02   -0.498   0.6185
## factor(district)12   4.412e-04   4.463e-02    0.010   0.9921
## factor(district)13  -3.064e-03   4.251e-02   -0.072   0.9425
## factor(district)14   2.441e-02   5.880e-02    0.415   0.6781
## factor(district)15  -2.027e-02   2.122e-02   -0.956   0.3393
## factor(district)16   7.288e-03   5.178e-02    0.141   0.8881
## factor(district)17   1.080e-03   4.859e-02    0.022   0.9823
## factor(district)18   4.017e-03   2.594e-02    0.155   0.8770
## factor(district)19   8.935e-03   3.625e-02    0.247   0.8053
## factor(district)20  -1.545e-02   3.378e-02   -0.457   0.6475
## factor(district)21  -3.672e-02   2.291e-02   -1.603   0.1090
## factor(district)22  -6.345e-04   1.390e-02   -0.046   0.9636
## factor(district)23  -1.097e-02   4.180e-02   -0.262   0.7929
## factor(district)24  -1.540e-02   3.085e-02   -0.499   0.6177
## factor(district)25   2.203e-02   7.539e-02    0.292   0.7702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7068 on 1077743 degrees of freedom
## Multiple R-squared:  0.0001208,  Adjusted R-squared:  -2.854e-05
## F-statistic: 0.8089 on 161 and 1077743 DF,  p-value: 0.964
```

Exercise 5

1. consider within estimator

```
panel3<-panel2 %>%
  group_by(NUID) %>%
  summarise(meanarrest=mean(arrest),meantenure=mean(tenure),meantotal_crime=mean(total_crime),
            meanmed=mean(median_income), meanblack=mean(black_share),meanhis=mean(hispanic_share),mean
head(panel3)
```

```
## # A tibble: 6 x 8
##     NUID meanarrest meantenure meantotal_crime meanmed meanblack meanhis
##    <int>      <dbl>      <dbl>           <dbl>   <dbl>     <dbl>   <dbl>
## 1      1      0.485       82.7           1060.  59407.     0.133   0.411
## 2      2      0.496       78.2           1270.  27940.     0.933  0.0370
## 3      6      0.333      157.             597.  52333.     0.107   0.183
## 4      7      0.477      199.             990.  57487.     0.111   0.336
## 5     16      0.463       63.4           1534.  51768.     0.715  0.0273
## 6     17      0.738      247.             486.  63652.    0.0998  0.0937
## # ... with 1 more variable: meanwhite <dbl>
```

```
between<-lm(meanarrest ~meantenure+meantotal_crime+meanmed+meanblack+meanhis+meanwhite,data=panel3)
summary(between)
```

```
##
## Call:
## lm(formula = meanarrest ~ meantenure + meantotal_crime + meanmed +
##     meanblack + meanhis + meanwhite, data = panel3)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50944 -0.06213 -0.00365  0.05341  2.49863
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.669e-01  2.686e-02  17.384   <2e-16 ***
## meantenure      1.203e-05  1.439e-05   0.835   0.4035
## meantotal_crime -7.837e-06  4.425e-06  -1.771   0.0766 .
## meanmed         1.429e-07  1.901e-07   0.752   0.4522
## meanblack       4.042e-02  2.837e-02   1.425   0.1543
## meanhis         4.905e-02  2.940e-02   1.668   0.0953 .
## meanwhite       2.217e-02  3.347e-02   0.662   0.5078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1466 on 13021 degrees of freedom
## Multiple R-squared:  0.0005331,  Adjusted R-squared:  7.253e-05
## F-statistic: 1.157 on 6 and 13021 DF,  p-value: 0.326
```

consider within estimator

```
panel4<-left_join(panel3,panel2,by=c("NUID"="NUID"))
```

```
panel4 <-panel4 %>%
  mutate(arrest=arrest-meanarrest,tenure=tenure-meantenure,black=black_share-meanblack,white=white_share
```

```
within<-lm(arrest~tenure+total_crime+median_income+black_share+hispanic_share+white_share,panel4)
summary(within)
```

```
##
## Call:
## lm(formula = arrest ~ tenure + total_crime + median_income +
##     black_share + hispanic_share + white_share, data = panel4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7496 -0.5075 -0.4281  0.4927  5.5120
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     7.684e-03  1.241e-02   0.619    0.536
## tenure          1.466e-05  2.190e-05   0.670    0.503
## total_crime     1.191e-06  1.846e-06   0.645    0.519
## median_income  -7.879e-08  1.041e-07  -0.757    0.449
## black_share    -9.614e-03  1.342e-02  -0.716    0.474
## hispanic_share -1.047e-02  1.393e-02  -0.752    0.452
## white_share    -9.638e-03  1.514e-02  -0.637    0.524
##
## Residual standard error: 0.7025 on 1077898 degrees of freedom
## Multiple R-squared:  1.329e-06,  Adjusted R-squared:  -4.237e-06
## F-statistic: 0.2388 on 6 and 1077898 DF,  p-value: 0.9638
```

consider first difference estimator

```
panel5<-panel2 %>%
  group_by(NUID) %>%
  mutate(beforetenure = lag(tenure),beforearrest=lag(arrest),before_crime=lag(total_crime),before_media
panel5 <-panel5 %>%
  mutate(tenurediff=tenure-beforetenure,arrestdiff=arrest-beforearrest,crime=total_crime-before_crime,me
panel5<-drop_na(panel5)
```

```
firstd<-lm(arrestdiff~tenurediff+median+crime+median+black+his+white,panel5)
summary(firstd)
```

```
##
## Call:
## lm(formula = arrestdiff ~ tenurediff + median + crime + median +
##     black + his + white, data = panel5)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0005 -0.9992  0.0000  0.9994  6.0005
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.569e-04  1.190e-03  -0.468    0.640
## tenurediff   5.004e-04  6.709e-04   0.746    0.456
## median       1.470e-07  1.162e-06   0.126    0.899
## crime       -2.721e-06  6.922e-06  -0.393    0.694
## black       -9.485e-02  1.590e-01  -0.597    0.551
## his         -1.119e-01  1.688e-01  -0.663    0.507
## white       -1.195e-01  2.154e-01  -0.555    0.579
##
## Residual standard error: 0.9995 on 1064870 degrees of freedom
## Multiple R-squared:  1.292e-06,  Adjusted R-squared:  -4.342e-06
## F-statistic: 0.2293 on 6 and 1064870 DF,  p-value: 0.9673
```

Comparing these three results, the estimator for beta in within and between are quite similiar, while coefficients for other varibales are quite different. While first difference estimator has a large diff. we can see that the first difference estimator can not handle time fixed effects, the failure for considering time also exists for withiin and between estimator. Also, first difference lack cross section information and fixed effects(since all of them are offset).

using GMM in one step: consider dataset, we should use panel2 the computation eats all my memory and I coould not get the result at all.

```
phi=length(unique(panel2$district))
chi=length(unique(panel2$month))
alpha=length(unique(panel2$NUID))
beta=0
gamma=5
```

initialize the coefficient&create intercept

```r
panel2$intercept=1
alpha1<-rep(0,alpha)
```

```r
library(fastDummies)
library(dummies)
```

```r
## dummies-1.5.6 provided by Decision Patterns
```

```r
X=panel2
X<-dummy_cols(X,select=c("month","district"))

y=panel2$arrest
```

```r
library(hash)
```

```r
## hash-2.2.6.1 provided by Decision Patterns
```

```r
h <- hash()
individual<-as.list(unique(X$NUID))
```

```r
coeff<-rep(1,1+1+5+phi+chi)
```

consider function (my pc cannot handle this computation)

result<-optim(runif(10,-0.1,0.1),gmm,method="BFGS")