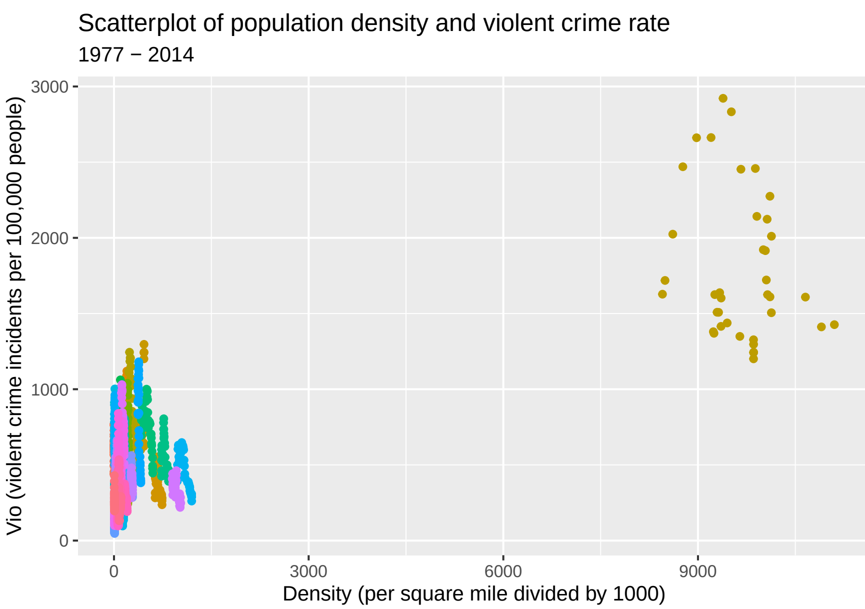
ECON-UB 251 Econometrics I Assignment 2

Kevin Song

1.1

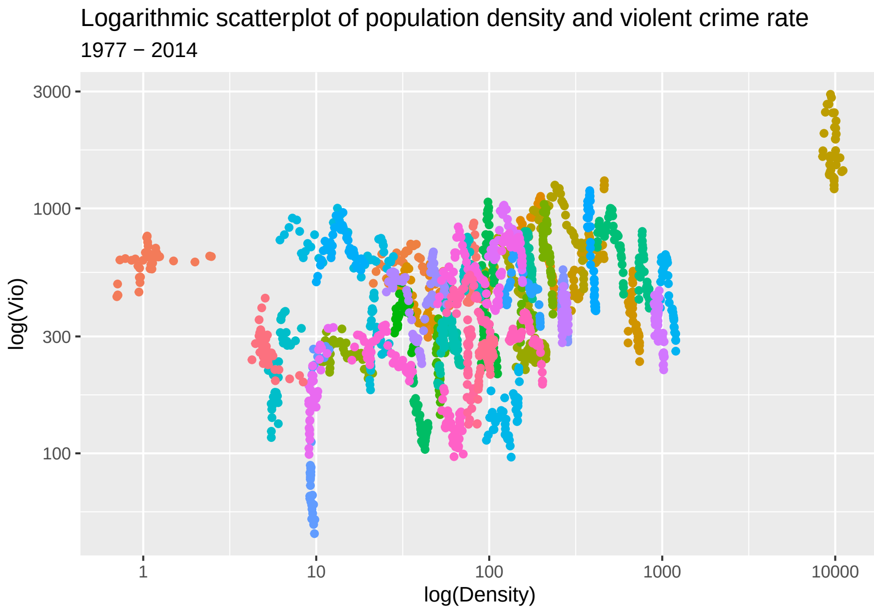
```
load("myWorkSpace.RData")
library(ggplot2)
library(readr)
data <- read_csv("assignment5_part1.csv")

ggplot(data, aes(x = density, y = vio, color = state))+
  geom_point() +
  theme(legend.position = "none")+
  labs(title = "Scatterplot of population density and violent crime rate",
       subtitle = "1977 - 2014",
       x = "Density (per square mile divided by 1000)",
       y = "Vio (violent crime incidents per 100,000 people)")
```



Almost all the states are below 1000 violent crime incidents per 100,000 citizens. The state that stands out is actually the District of Columbia.

```
ggplot(data, aes(x = density, y = vio, color = state))+
  geom_point() +
  theme(legend.position = "none") +
  scale_x_log10() + scale_y_log10() +
  labs(title = "Logarithmic scatterplot of population density and violent crime rate",
       subtitle = "1977 - 2014",
       x = "log(Density)",
```

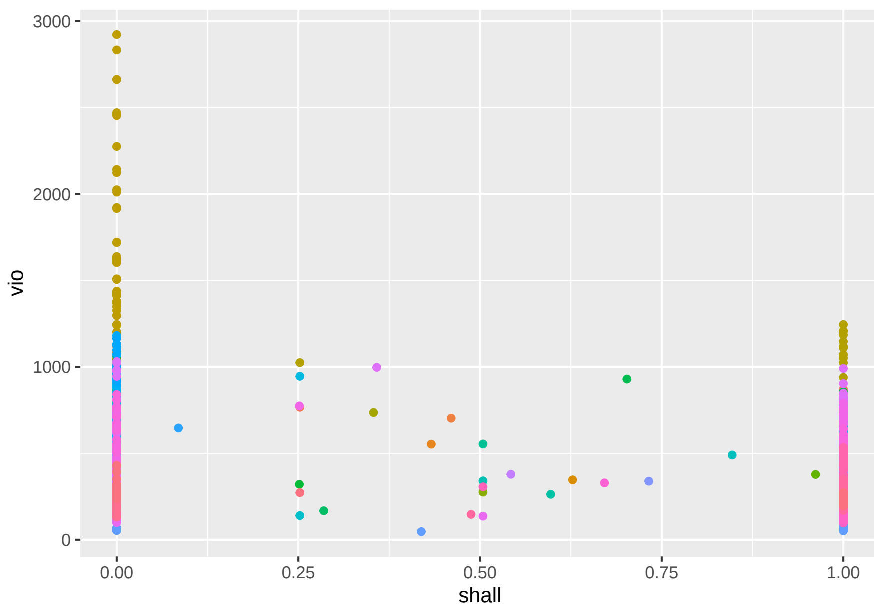
```
    y = "log(Vio)")
```



Using a logarithmic scale definitely improves clarity of the states and helps reveal other potential outliers with extremely low violent crime incident rates that otherwise were too difficult to see.

1.2

```
ggplot(data, aes(shall, vio, color = state)) + geom_point() + theme(legend.position = "none")
```



By observing shall v vio in each state, it can be more clearly seen that there is a slight negative relationship, meaning shall actually may lower violence rate overall.

1.3

```r
library(plm)

vio.pd <- pdata.frame(data, index=c("state","year"),
                      drop.index=TRUE, row.names=TRUE)
vio.pool <- plm((log(vio))~shall, data = vio.pd, model = "pooling")
summary(vio.pool)
```

```
Pooling Model

Call:
plm(formula = (log(vio)) ~ shall, data = vio.pd, model = "pooling")

Balanced Panel: n = 51, T = 38, N = 1938

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.139114 -0.383608  0.038437  0.406258  1.858520

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  6.121434   0.017676 346.315 < 2.2e-16 ***
shall       -0.314943   0.026351 -11.952 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    684.78
Residual Sum of Squares: 637.73
R-Squared:      0.068716
Adj. R-Squared: 0.068235
F-statistic: 142.85 on 1 and 1936 DF, p-value: < 2.22e-16
```

The estimate coefficient for shall is -0.3149. Since it is negative, having a shall-issue law lowers the violent crime rate by 31.5% relative to not having one. The p-value is less than 0.01, so this coefficient seems very strong. This supports the view that gun control laws are effective at lowering violent crime.

1.4

```r
vio.pd <- pdata.frame(data, index=c("state","year"),
                      drop.index=TRUE, row.names=TRUE)
vio.fe <- plm((log(vio))~shall, effect = "individual", data = vio.pd, model = "within")

summary(vio.fe, vcov = vcovHC)
```

```
Oneway (individual) effect Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = (log(vio)) ~ shall, data = vio.pd, effect = "individual",
    model = "within")

Balanced Panel: n = 51, T = 38, N = 1938
```

```
Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.7044334 -0.1435709 -0.0032302  0.1474221  1.0757576

Coefficients:
       Estimate Std. Error t-value Pr(>|t|)
shall -0.033673   0.038858 -0.8666   0.3863

Total Sum of Squares:    99.675
Residual Sum of Squares: 99.362
R-Squared:      0.0031396
Adj. R-Squared: -0.023817
F-statistic: 0.750942 on 1 and 50 DF, p-value: 0.39032
```

Beta is now -0.03367. This is 10 times smaller than calculated previously. The p-value is also much higher now, with it being insignificant at the 1% level. What this means is that differences between states are a large determinant in violent crime rates themselves regardless if shall law is passed in that state or not.

1.5

```r
vio.pd <- pdata.frame(data, index=c("state","year"),
                             drop.index=TRUE, row.names=TRUE)
vio.fe <- plm((log(vio))~shall, effect = "twoways", data = vio.pd, model = "within")

summary(vio.pool, vcov = vcovHC)
```

```
Pooling Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = (log(vio)) ~ shall, data = vio.pd, model = "pooling")

Balanced Panel: n = 51, T = 38, N = 1938

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.139114 -0.383608  0.038437  0.406258  1.858520

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  6.121434   0.084082 72.8028 < 2.2e-16 ***
shall       -0.314943   0.107348 -2.9338  0.003387 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    684.78
Residual Sum of Squares: 637.73
R-Squared:      0.068716
Adj. R-Squared: 0.068235
F-statistic: 8.60746 on 1 and 50 DF, p-value: 0.0050429
```

Beta is now 0.028157. It is positive now, which is significant different than both previous estimates which were negative. Now the model shows shall-issue laws actually increase violence, having a shall law increases violent crime rate by 2.8%. However, the p value is greater than 0.5, so the coefficient is not significant at all.

Compared to the state-fixed model, the coefficient does not change significantly, with it being less than 2 standard deviations away: 0.0281 - 2(0.0507) < -0.03367. Compared to the pooled model, the twoway model is over 6 standard deviations away. This could mean either state effects have a greater effect than time effects, or it does not mean anything since this coefficient is insignificant anyways.

```
vio.pd <- pdata.frame(data, index=c("state","year"),
                       drop.index=TRUE, row.names=TRUE)
vio.pool <- plm((log(vio))~shall + rpcpi + rpcui + rpcim + density + pbm1019 + pbm2029 + pwm1019 + pwm2

summary(vio.pool, vcov = vcovHC)
```

```
Twoways effects Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = (log(vio)) ~ shall + rpcpi + rpcui + rpcim + density +
    pbm1019 + pbm2029 + pwm1019 + pwm2029, data = vio.pd, effect = "twoways",
    model = "within")

Balanced Panel: n = 51, T = 38, N = 1938

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.6476782 -0.1057866  0.0049937  0.1052378  0.9846603

Coefficients:
           Estimate  Std. Error t-value Pr(>|t|)
shall   -3.1669e-02  4.8851e-02 -0.6483 0.516883
rpcpi   -1.6270e-05  1.9777e-05 -0.8227 0.410812
rpcui   -1.3795e-03  4.3411e-04 -3.1778 0.001508 **
rpcim    2.7479e-04  4.7227e-04  0.5818 0.560750
density -3.5316e-04  1.9396e-04 -1.8208 0.068806 .
pbm1019  6.2806e-02  1.0697e-01  0.5871 0.557189
pbm2029  2.1999e-01  1.0658e-01  2.0640 0.039158 *
pwm1019 -8.5756e-02  4.0139e-02 -2.1365 0.032770 *
pwm2029  1.0988e-01  4.3081e-02  2.5505 0.010837 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    76.326
Residual Sum of Squares: 66.014
R-Squared:      0.13511
Adj. R-Squared: 0.090005
F-statistic: 4.03258 on 9 and 50 DF, p-value: 0.00062891
```

Beta is now -0.03167. This is negative 3% compared to the original shall model's negative 31%. However, the standard error is 0.0488 which is very large and once again the beta is insignificant with a p-value > 0.5. This shows that after controlling for these other confounding variables, those variables may only have a marginal effect on violent crime rate. It may not at all considering shall does not even pass the significance test at any level in this final model.

2.1

```
library(dplyr)
library(openxlsx)
library(lmtest)
data2 <- read.xlsx("assignment5_part2.xlsx")

data2 = filter(data2, (employed + unemployed) !=0)

employ <- data2 %>%
          select(employed, unemployed) %>%
          tidyr::gather(status, value) %>%
          group_by(status) %>%
          summarize(N = sum(value, na.rm = T)) %>%
          mutate(Percentage = N / sum(N)) %>%
          knitr::kable()
employ
```

status	N	Percentage
employed	4738	0.9519791
unemployed	239	0.0480209

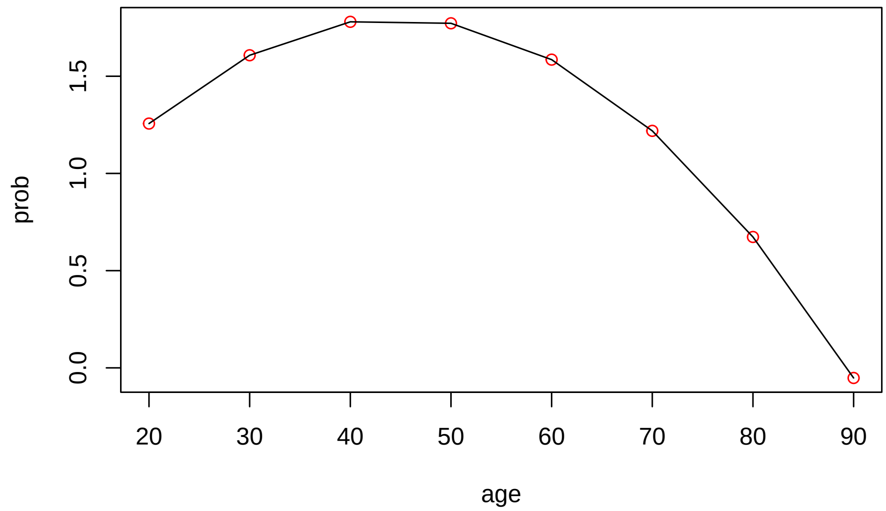
From the table, 239 people lost their job, about 5% of the sample.

2.2

```
save.image("myWorkSpace.RData")
probit <- glm(employed~ age +I(age^2), data = data2, family = binomial(link = "probit"))
coeftest(probit, vcov. = vcovHC, type = "HC1")
```

```
z test of coefficients:

               Estimate  Std. Error z value  Pr(>|z|)
(Intercept)  0.01560905  0.35474297  0.0440    0.9649
age          0.07998691  0.01826374  4.3795 1.189e-05 ***
I(age^2)    -0.00089708  0.00022022 -4.0736 4.630e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
probit.df <- select(data2, age, employed) %>%
              mutate(fit = predict(probit, type = "response"))

coeffs <- coefficients(probit)
prob <- coeffs[1] + coeffs[2]*age + coeffs[3]*(age^2)
plot(age,prob,col="red") + lines(age, prob)
```




```
integer(0)
```

From this parabola, it can be observed that age has a great effect on the probability of being employed in one's early years from 20 to 40, but after 50 it becomes harder to be employed. The p-values are very small, < 0.01, so this result is significant.

2.3

```
logit <- glm(employed ~ age + I(age^2) +earnwke + race + married + female + ne_states + so_states + ce_s
we_states + educ_lths + educ_hs + educ_somecol + educ_aa + educ_bac + educ_adv, data = data2,
             family = binomial(link = "logit"))
coeftest(logit, vcov. = vcovHC, type = "HC1")
```

```
z test of coefficients:

                Estimate  Std. Error z value  Pr(>|z|)
(Intercept)   0.72104961  0.84852631  0.8498 0.3954547
age           0.11636423  0.04240863  2.7439 0.0060718 **
I(age^2)     -0.00140855  0.00051074 -2.7578 0.0058184 **
earnwke       0.00013267  0.00017564  0.7554 0.4500271
race         -0.08358919  0.11331077 -0.7377 0.4606977
married       0.32913322  0.15408456  2.1361 0.0326749 *
female        0.42869060  0.14709624  2.9144 0.0035642 **
ne_states     0.39626091  0.20741357  1.9105 0.0560705 .
so_states     0.19912295  0.17980828  1.1074 0.2681132
ce_states     0.34904090  0.19699442  1.7718 0.0764225 .
educ_lths    -1.38592519  0.36324123 -3.8154 0.0001359 ***
educ_hs      -0.88401127  0.31642225 -2.7938 0.0052097 **
educ_somecol -0.45270426  0.33769129 -1.3406 0.1800549
educ_aa      -0.33141189  0.37245082 -0.8898 0.3735659
educ_bac     -0.44990116  0.32381786 -1.3894 0.1647218
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The types of workers that were most affected by the Great Recession seems to be people with less education, with it being worst off for people with less than high school education. The coefficient is the most negative of all regressors and has a significant z value of -3.8.

2.4

```
logit <- glm(unemployed ~ age + I(age^2) +earnwke + race + married + female + ne_states + so_states + ce
we_states + educ_lths + educ_hs + educ_somecol + educ_aa + educ_bac + educ_adv, data = data2,
             family = binomial(link = "logit"))
coeftest(logit, vcov. = vcovHC, type = "HC1")
```

```
z test of coefficients:

                Estimate  Std. Error z value  Pr(>|z|)
(Intercept)  -0.72104961  0.84852631 -0.8498 0.3954547
age          -0.11636423  0.04240863 -2.7439 0.0060718 **
I(age^2)      0.00140855  0.00051074  2.7578 0.0058184 **
earnwke      -0.00013267  0.00017564 -0.7554 0.4500271
race          0.08358919  0.11331077  0.7377 0.4606977
married      -0.32913322  0.15408456 -2.1361 0.0326749 *
female       -0.42869060  0.14709624 -2.9144 0.0035642 **
ne_states    -0.39626091  0.20741357 -1.9105 0.0560705 .
so_states    -0.19912295  0.17980828 -1.1074 0.2681132
ce_states    -0.34904090  0.19699442 -1.7718 0.0764225 .
educ_lths     1.38592519  0.36324123  3.8154 0.0001359 ***
educ_hs       0.88401127  0.31642225  2.7938 0.0052097 **
educ_somecol  0.45270426  0.33769129  1.3406 0.1800549
educ_aa       0.33141189  0.37245082  0.8898 0.3735659
educ_bac      0.44990116  0.32381786  1.3894 0.1647218
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this test I am looking for the most positive regressors instead of most negative. It is education level again, with not completing high school being hte most affected with a significant p-value. The signs are