# p8130_hw5_yl5508

Yifei LIU (yl5508)

2023/12/7

```r
library(tidyverse)
library(faraway)
library(glmnet)

set.seed(1)
```

## Problem 1

### (a)

```r
#load data set
sta_data = as.data.frame(state.x77)
sta_data |>
  summary() |>
  knitr::kable(digits = 1)
```

| | Population | Income | Illiteracy | Life Exp | Murder | HS Grad | Frost | Area |
|---|---|---|---|---|---|---|---|---|
| | Min. : 365 | Min. :3098 | Min. :0.500 | Min. :67.96 | Min. : 1.400 | Min. :37.80 | Min. : 0.00 | Min. : 1049 |
| | 1st Qu.: 1080 | 1st Qu.:3993 | 1st Qu.:0.625 | 1st Qu.:70.12 | 1st Qu.: 4.350 | 1st Qu.:48.05 | 1st Qu.: 66.25 | 1st Qu.: 36985 |
| | Median : 2838 | Median :4519 | Median :0.950 | Median :70.67 | Median : 6.850 | Median :53.25 | Median :114.50 | Median : 54277 |
| | Mean : 4246 | Mean :4436 | Mean :1.170 | Mean :70.88 | Mean : 7.378 | Mean :53.11 | Mean :104.46 | Mean : 70736 |
| | 3rd Qu.: 4968 | 3rd Qu.:4814 | 3rd Qu.:1.575 | 3rd Qu.:71.89 | 3rd Qu.:10.675 | 3rd Qu.:59.15 | 3rd Qu.:139.75 | 3rd Qu.: 81163 |
| | Max. :21198 | Max. :6315 | Max. :2.800 | Max. :73.60 | Max. :15.100 | Max. :67.30 | Max. :188.00 | Max. :566432 |

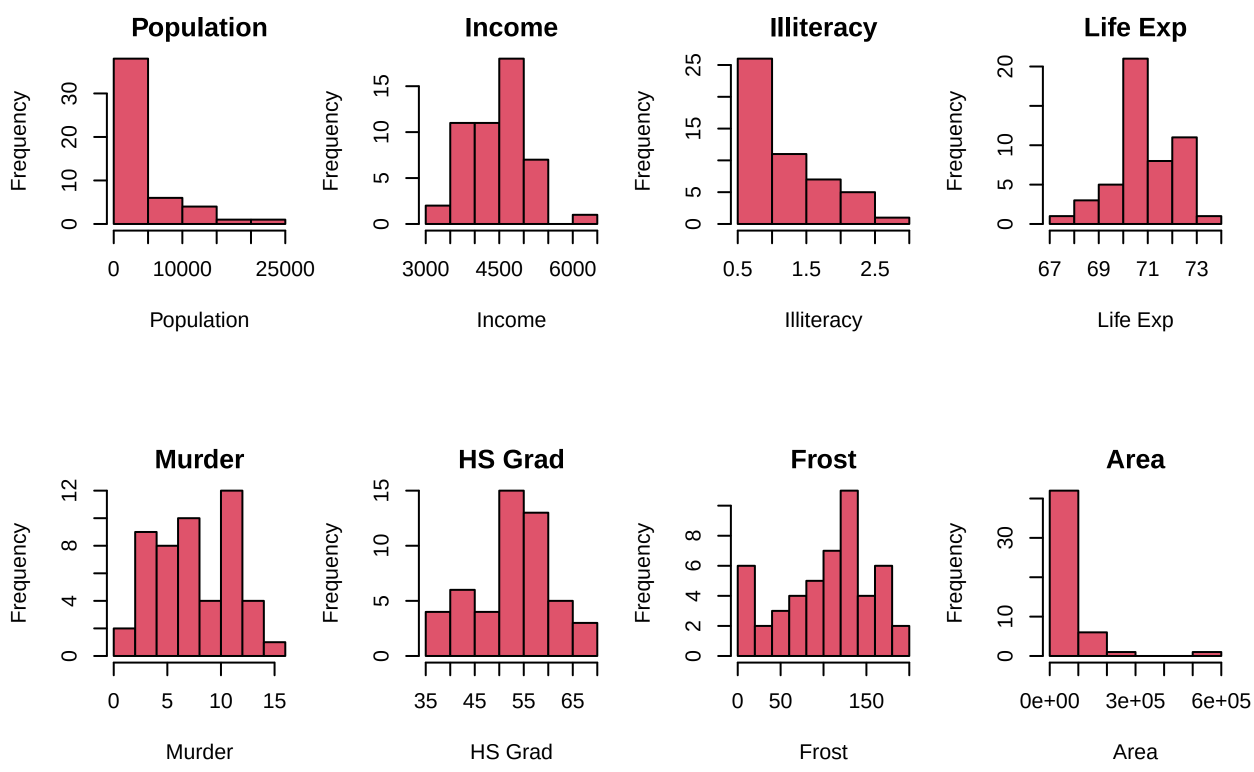
Continuous variables includes `Population`, `Income`, `Illiteracy`, `Life Exp`, `Murder`, `HS Grad`, `Frost`, `Area`.

No variable listed in the data set is categorical.

### (b)

```
#histogram of variables
par(mfrow = c(2, 4), mar = c(8, 4, 2, 1))

for (i in 1:8) {
  sta_data[,i] |>
  hist(main = colnames(sta_data[i]), xlab = colnames(sta_data[i]), freq = T, col = 2)
}
```




From the histograms, we notice that `Population`, `Illiteracy`, `Area` need to be transformed in order to get a normal distribution.

```
#log transformation
sta_transformed =
  sta_data |>
  mutate(
    Population_t = log(Population),
    Illiteracy_t = log(Illiteracy),
    Area_t = log(Area)) |>
  select(Population, Population_t, Illiteracy, Illiteracy_t, Area, Area_t)

sta_tidy =
  sta_data |>
  mutate(
    Population_t = log(Population),
    Illiteracy_t = log(Illiteracy),
```

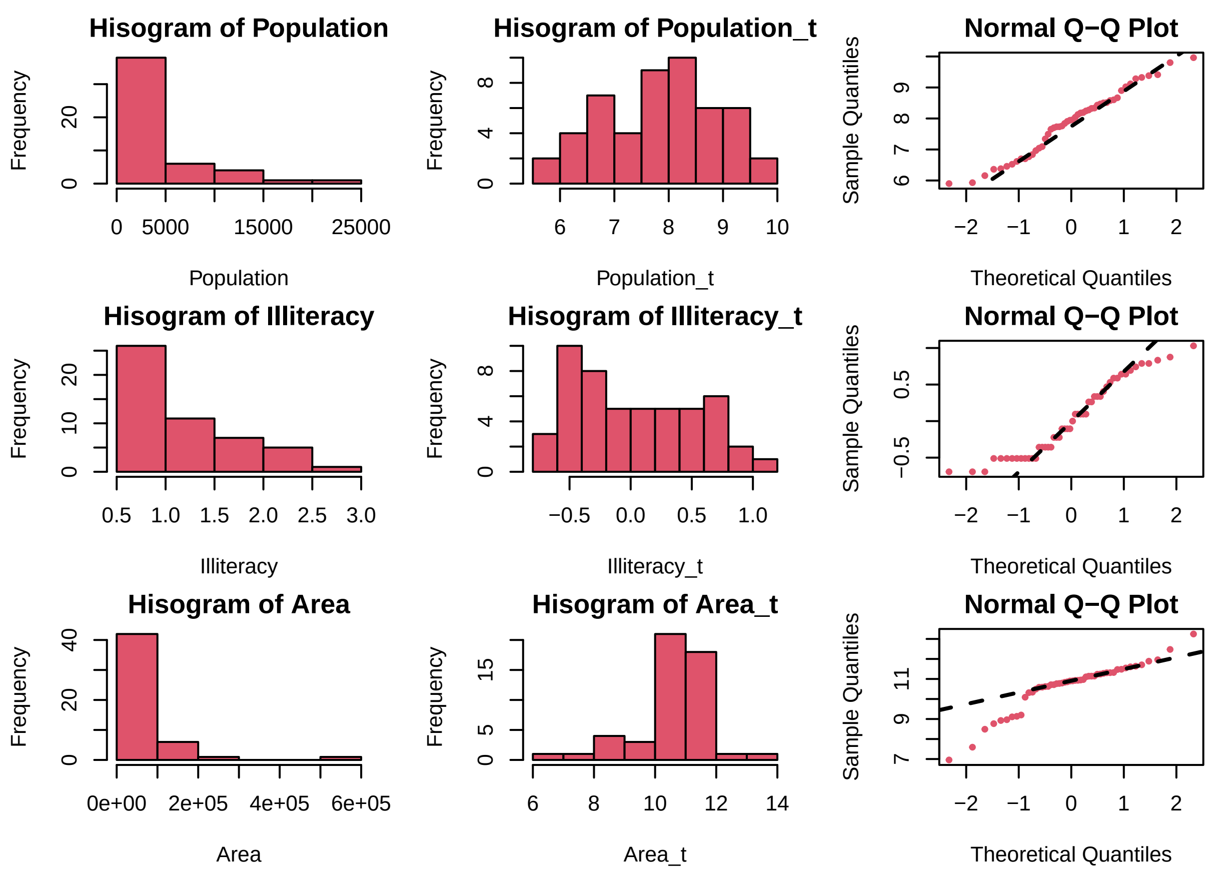
```
    Area_t = log(Area)) |>
  select(-Population, -Illiteracy, -Area) |>
  select(`Life Exp`, everything())

par(mfrow = c(3, 3), mar = c(4, 4, 2, 2))

for (i in seq(1, 5, 2)) {
  #untransformed variables
  sta_transformed[,i] |>
    hist(main = str_c("Hisogram of ", colnames(sta_transformed[i])), xlab = colnames(sta_transformed[i]

  #log transformed variables
  sta_transformed[,i+1] |>
    hist(main = str_c("Hisogram of ", colnames(sta_transformed[i+1])), xlab = colnames(sta_transformed[

  #Q-Q plot
  qqnorm(sta_transformed[,i+1], col = 2, pch = 19, cex = 0.5)
  qqline(sta_transformed[,i+1], col = 1, lwd = 2, lty = 2)
}
```




**(c)**

```
#global variables
lm(`Life Exp` ~ ., data = sta_tidy) |>
  summary()
```

```
##
## Call:
## lm(formula = `Life Exp` ~ ., data = sta_tidy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44702 -0.42901  0.04546  0.50742  1.68911
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.799e+01  1.798e+00  37.809  < 2e-16 ***
## Income       -4.417e-06  2.475e-04  -0.018   0.9858
## Murder       -3.114e-01  4.659e-02  -6.684 4.12e-08 ***
## `HS Grad`     5.482e-02  2.552e-02   2.148   0.0375 *
## Frost        -4.669e-03  3.173e-03  -1.471   0.1487
## Population_t  2.537e-01  1.311e-01   1.936   0.0597 .
## Illiteracy_t  1.883e-01  4.204e-01   0.448   0.6565
## Area_t        7.314e-02  1.102e-01   0.663   0.5107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7335 on 42 degrees of freedom
## Multiple R-squared:  0.7441, Adjusted R-squared:  0.7014
## F-statistic: 17.45 on 7 and 42 DF,  p-value: 1.368e-10
```

```r
#forward stepwise
model_fw = lm(`Life Exp` ~ ., data = sta_tidy) |>
  step(direction = "forward")
```

```
## Start:  AIC=-23.71
## `Life Exp` ~ Income + Murder + `HS Grad` + Frost + Population_t +
##     Illiteracy_t + Area_t
```

```r
model_fw |> summary()
```

```
##
## Call:
## lm(formula = `Life Exp` ~ Income + Murder + `HS Grad` + Frost +
##     Population_t + Illiteracy_t + Area_t, data = sta_tidy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44702 -0.42901  0.04546  0.50742  1.68911
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.799e+01  1.798e+00  37.809  < 2e-16 ***
## Income       -4.417e-06  2.475e-04  -0.018   0.9858
## Murder       -3.114e-01  4.659e-02  -6.684 4.12e-08 ***
## `HS Grad`     5.482e-02  2.552e-02   2.148   0.0375 *
## Frost        -4.669e-03  3.173e-03  -1.471   0.1487
## Population_t  2.537e-01  1.311e-01   1.936   0.0597 .
## Illiteracy_t  1.883e-01  4.204e-01   0.448   0.6565
```

```
## Area_t        7.314e-02  1.102e-01   0.663   0.5107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7335 on 42 degrees of freedom
## Multiple R-squared:  0.7441, Adjusted R-squared:  0.7014
## F-statistic: 17.45 on 7 and 42 DF,  p-value: 1.368e-10
```

```r
#backward stepwise
model_bk = lm(`Life Exp` ~ ., data = sta_tidy) |>
  step(direction = "backward")
```

```
## Start:  AIC=-23.71
## `Life Exp` ~ Income + Murder + `HS Grad` + Frost + Population_t +
##     Illiteracy_t + Area_t
##
##                 Df Sum of Sq    RSS     AIC
## - Income         1    0.0002 22.596 -25.712
## - Illiteracy_t   1    0.1079 22.704 -25.475
## - Area_t         1    0.2368 22.833 -25.192
## <none>                       22.596 -23.713
## - Frost          1    1.1645 23.760 -23.200
## - Population_t   1    2.0155 24.611 -21.441
## - `HS Grad`      1    2.4822 25.078 -20.502
## - Murder         1   24.0347 46.631  10.512
##
## Step:  AIC=-25.71
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Illiteracy_t +
##     Area_t
##
##                 Df Sum of Sq    RSS      AIC
## - Illiteracy_t   1    0.1095 22.705 -27.4708
## - Area_t         1    0.2616 22.858 -27.1370
## <none>                       22.596 -25.7125
## - Frost          1    1.2628 23.859 -24.9936
## - Population_t   1    2.3859 24.982 -22.6937
## - `HS Grad`      1    4.4112 27.007 -18.7959
## - Murder         1   24.4834 47.079   8.9907
##
## Step:  AIC=-27.47
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Area_t
##
##                 Df Sum of Sq    RSS     AIC
## - Area_t         1    0.2157 22.921 -28.998
## <none>                       22.705 -27.471
## - Population_t   1    2.2792 24.985 -24.688
## - Frost          1    2.3760 25.082 -24.495
## - `HS Grad`      1    4.9491 27.655 -19.612
## - Murder         1   29.2296 51.935  11.899
##
## Step:  AIC=-29
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t
##
##                 Df Sum of Sq    RSS     AIC
```

```
## <none>                        22.921 -28.998
## - Frost          1      2.214 25.135 -26.387
## - Population_t   1      2.450 25.372 -25.920
## - `HS Grad`      1      6.959 29.881 -17.741
## - Murder         1     34.109 57.031  14.578
```

```
model_bk |> summary()
```

```
##
## Call:
## lm(formula = `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t,
##     data = sta_tidy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41760 -0.43880  0.02539  0.52066  1.63048
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  68.720810   1.416828  48.503  < 2e-16 ***
## Murder       -0.290016   0.035440  -8.183 1.87e-10 ***
## `HS Grad`     0.054550   0.014758   3.696 0.000591 ***
## Frost        -0.005174   0.002482  -2.085 0.042779 *
## Population_t  0.246836   0.112539   2.193 0.033491 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7137 on 45 degrees of freedom
## Multiple R-squared:  0.7404, Adjusted R-squared:  0.7173
## F-statistic: 32.09 on 4 and 45 DF,  p-value: 1.17e-12
```

```
#both
model_bth = lm(`Life Exp` ~ ., data = sta_tidy) |>
  step(direction = "both")
```

```
## Start:  AIC=-23.71
## `Life Exp` ~ Income + Murder + `HS Grad` + Frost + Population_t +
##     Illiteracy_t + Area_t
##
##                Df Sum of Sq    RSS     AIC
## - Income        1    0.0002 22.596 -25.712
## - Illiteracy_t  1    0.1079 22.704 -25.475
## - Area_t        1    0.2368 22.833 -25.192
## <none>                      22.596 -23.713
## - Frost         1    1.1645 23.760 -23.200
## - Population_t  1    2.0155 24.611 -21.441
## - `HS Grad`     1    2.4822 25.078 -20.502
## - Murder        1   24.0347 46.631  10.512
##
## Step:  AIC=-25.71
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Illiteracy_t +
##     Area_t
##
```

```
##                 Df Sum of Sq    RSS      AIC
## - Illiteracy_t  1    0.1095 22.705 -27.4708
## - Area_t        1    0.2616 22.858 -27.1370
## <none>                      22.596 -25.7125
## - Frost         1    1.2628 23.859 -24.9936
## + Income        1    0.0002 22.596 -23.7129
## - Population_t  1    2.3859 24.982 -22.6937
## - `HS Grad`     1    4.4112 27.007 -18.7959
## - Murder        1   24.4834 47.079   8.9907
##
## Step:  AIC=-27.47
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Area_t
##
##                 Df Sum of Sq    RSS     AIC
## - Area_t        1    0.2157 22.921 -28.998
## <none>                      22.705 -27.471
## + Illiteracy_t  1    0.1095 22.596 -25.712
## + Income        1    0.0017 22.704 -25.475
## - Population_t  1    2.2792 24.985 -24.688
## - Frost         1    2.3760 25.082 -24.495
## - `HS Grad`     1    4.9491 27.655 -19.612
## - Murder        1   29.2296 51.935  11.899
##
## Step:  AIC=-29
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t
##
##                 Df Sum of Sq    RSS     AIC
## <none>                      22.921 -28.998
## + Area_t        1     0.216 22.705 -27.471
## + Illiteracy_t  1     0.064 22.858 -27.137
## + Income        1     0.011 22.911 -27.021
## - Frost         1     2.214 25.135 -26.387
## - Population_t  1     2.450 25.372 -25.920
## - `HS Grad`     1     6.959 29.881 -17.741
## - Murder        1    34.109 57.031  14.578
```

```r
model_bth |> summary()
```

```
##
## Call:
## lm(formula = `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t,
##     data = sta_tidy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41760 -0.43880  0.02539  0.52066  1.63048
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  68.720810   1.416828  48.503  < 2e-16 ***
## Murder       -0.290016   0.035440  -8.183 1.87e-10 ***
## `HS Grad`     0.054550   0.014758   3.696 0.000591 ***
## Frost        -0.005174   0.002482  -2.085 0.042779 *
## Population_t  0.246836   0.112539   2.193 0.033491 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7137 on 45 degrees of freedom
## Multiple R-squared:  0.7404, Adjusted R-squared:  0.7173
## F-statistic: 32.09 on 4 and 45 DF,  p-value: 1.17e-12
```

Based on AIC, the function reduces the set of potential predictors. The model with the smallest value would be deemed as appropriate.

Actually the model shown after variables selection would not be the final result. We need to trim some variables off based on the p-value listed in the tables.

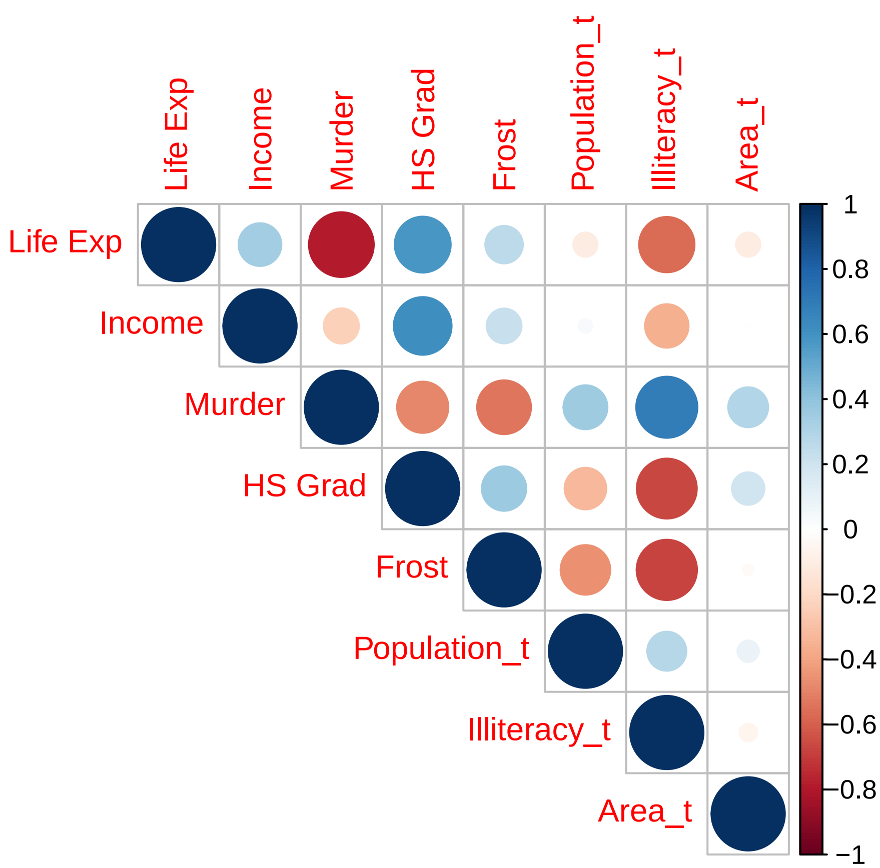
For forward selection, the model shows that only variables `Murder` and `HS Grad` is significantly effective (p $< 0.05$).

For backward selection, `Murder`, `HS Grad`, `Frost`, `Population_t` are significant variables.

For method concerning both forward and backward selection, the result is the same as backward selection.

So, I would pick `Murder`, `HS Grad`, `Frost`, `Population_t` as my predictors.

```
corrplot::corrplot(cor(sta_tidy), type = "upper")
```




We notice that there is a strong negative correlation between `Illiteracy` and `HS Grad` (approximately 0.8). My variables subset doesn't include both, for the variables selection process can partly deal with multicollinearity issue.

**(d)**

```r
library(leaps)
```

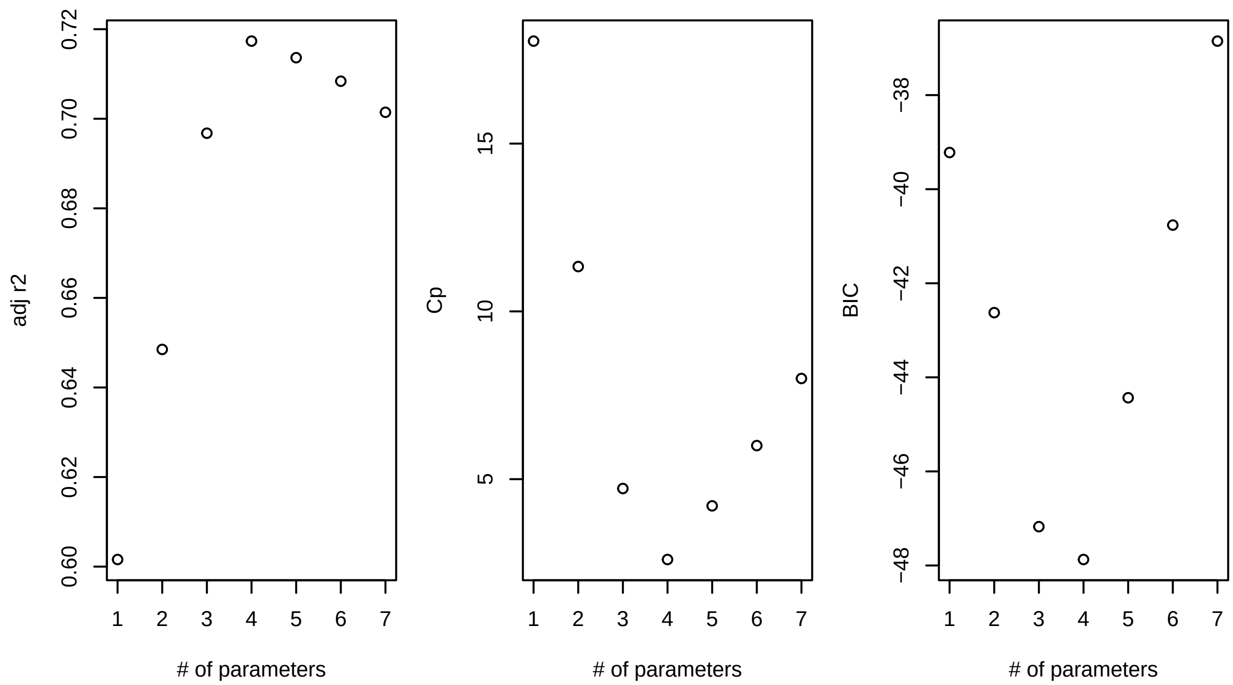
```
## Warning:   'leaps' R 4.3.2
```

```r
mat = as.matrix(sta_tidy)
leaps(x = mat[, 2:8], y = mat[, 1], method = "adjr2", nbest = 2)
```

```
## $which
##       1     2     3     4     5     6     7
## 1 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 1 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
## 2 FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
## 2 FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE
## 3 FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE
## 3 FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE
## 4 FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
## 4 FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE
## 5 FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## 5 FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## 6 FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 6  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## 7  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##
## $label
## [1] "(Intercept)" "1"           "2"           "3"           "4"
## [6] "5"           "6"           "7"
##
## $size
##  [1] 2 2 3 3 4 4 5 5 6 6 7 7 8
##
## $adjr2
##  [1] 0.6015893 0.3252044 0.6484991 0.6301232 0.6967729 0.6939230 0.7173392
##  [8] 0.7031061 0.7136360 0.7117179 0.7083894 0.7069987 0.7014485
```

```r
model_cri = regsubsets(`Life Exp` ~ ., data = sta_tidy)
res =
  model_cri |>
  summary()

par(mfrow = c(1, 3), mar = c(8, 4, 4, 1))
plot(1:7, res$adjr2, xlab = "# of parameters", ylab = "adj r2")
plot(1:7, res$cp, xlab = "# of parameters", ylab = "Cp")
plot(1:7, res$bic, xlab = "# of parameters", ylab = "BIC")
```




```r
res$outmat[4,]
```

```
##        Income        Murder     `HS Grad`         Frost Population_t Illiteracy_t
##           " "           "*"           "*"           "*"          "*"          " "
##        Area_t
##           " "
```

From the criterion-based procedures, using `adjusted r squared`/`Cp criterion`/`BIC`, we conclude that the best model contain 4 parameters and the parameters are `Murder`, `HS Grad`, `Frost`, `Population_t`.

**(e)**

```r
#explore possible lambda
fit1 = glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy, lambda = 1)
coef(fit1)
```

```
## 8 x 1 sparse Matrix of class "dgCMatrix"
##                        s0
## (Intercept)  70.95464716
## Income         .
## Murder       -0.01030729
## HS Grad        .
## Frost          .
## Population_t   .
## Illiteracy_t   .
## Area_t         .
```

```r
fit2 = glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy, lambda = 0.1)
coef(fit2)
```

```
## 8 x 1 sparse Matrix of class "dgCMatrix"
##                        s0
## (Intercept)  69.623968632
## Income        .
## Murder       -0.238460282
## HS Grad       0.040072350
## Frost        -0.001483129
## Population_t  0.132353805
## Illiteracy_t  .
## Area_t        .
```

```r
fit3 = glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy, lambda = 0.01)
coef(fit3)
```

```
## 8 x 1 sparse Matrix of class "dgCMatrix"
##                        s0
## (Intercept)  68.426042158
## Income        .
## Murder       -0.297414030
## HS Grad       0.051183064
## Frost        -0.004748876
## Population_t  0.233986320
## Illiteracy_t  0.062909458
## Area_t        0.054660434
```

We would consider setting the range of lambda at the interval of (0.01, 0.1).

```r
#grid search
lambda_seq = 10^seq(-2, 1, by = 0.1)
```

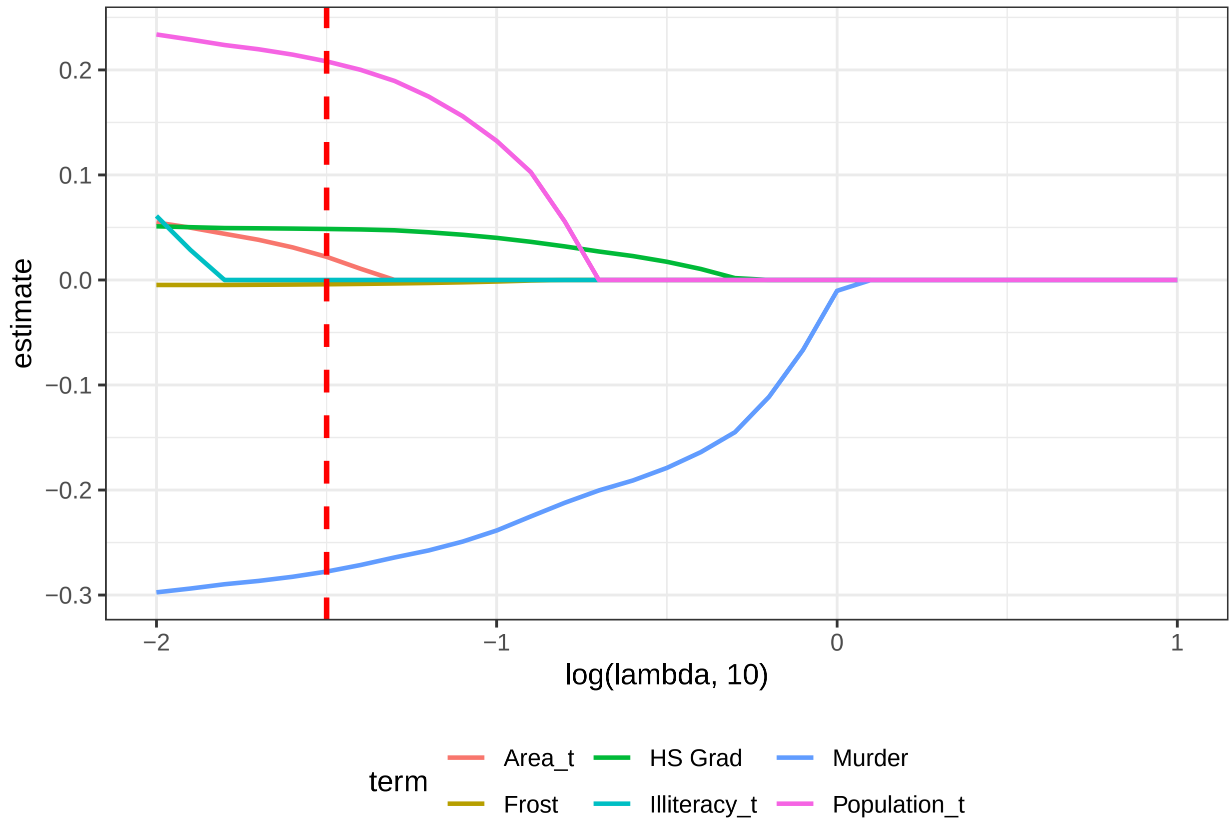
```r
cv_res =
  cv.glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy,
            lambda = lambda_seq, nfolds = 5)

opt_lambda = cv_res$lambda.min

#variables contraction
glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy, lambda = lambda_seq) |>
  broom::tidy() |>
  select(term, lambda, estimate) |>
  complete(term, lambda, fill = list(estimate = 0) ) |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = log(lambda, 10), y = estimate, group = term, color = term)) +
  geom_path(size = 0.8) +
  geom_vline(xintercept = log(opt_lambda, 10), color = "red", linetype = "dashed", size = 1) +
  theme_bw() +
  theme(legend.position = "bottom")
```

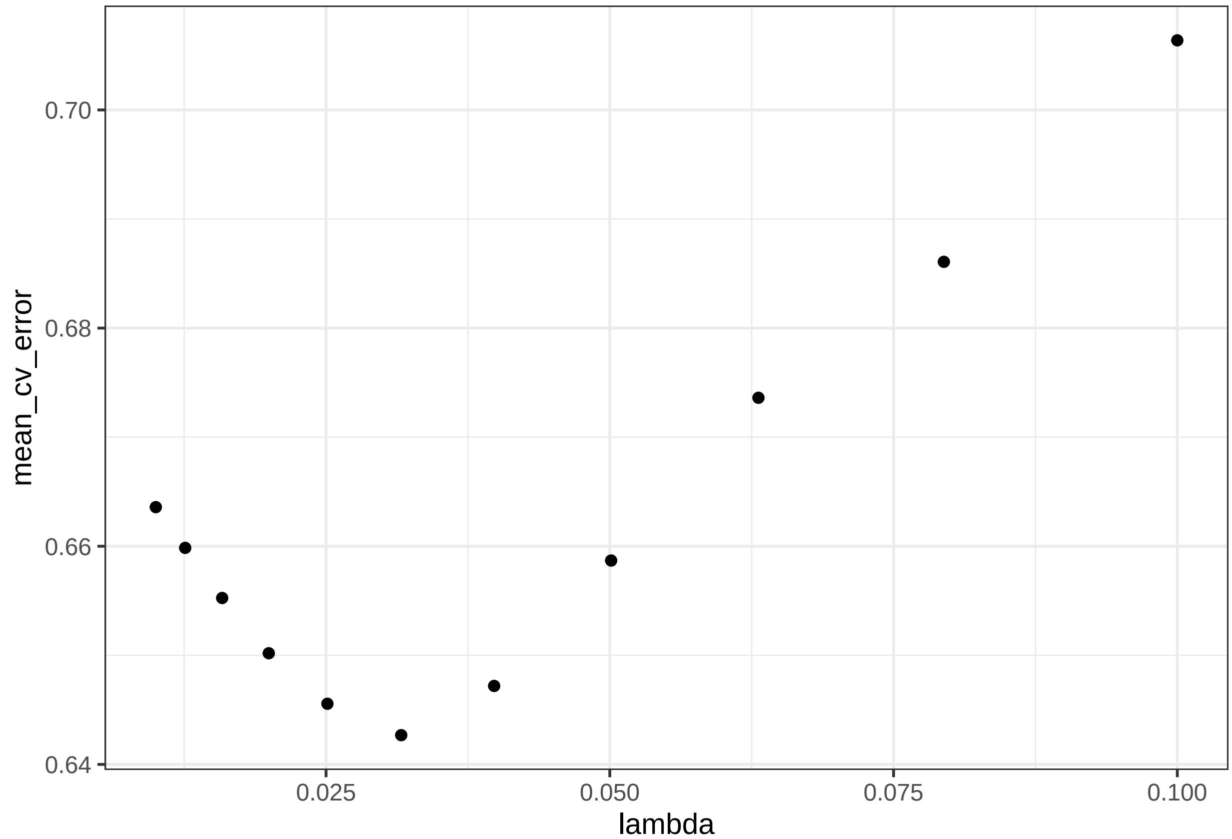
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
```

```
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
tb_res = tibble(
  lambda = cv_res$lambda,
  mean_cv_error = cv_res$cvm) |>
  filter(lambda < 0.1)

#choosing optimal lambda
tb_res |>
  ggplot(aes(x = lambda, y = mean_cv_error)) +
  geom_point() +
  theme_bw()
```




The optimal lambda we have chosen is 0.03. And the variables we determine are `Murder`, `HS Grad`, `Frost`, `Population_t`.

**(f)**

```
#stepwise
fit_bth = lm(`Life Exp` ~ ., data = sta_tidy) |>
  step(direction = "both")
```

```
## Start:  AIC=-23.71
## `Life Exp` ~ Income + Murder + `HS Grad` + Frost + Population_t +
##     Illiteracy_t + Area_t
##
##                  Df Sum of Sq    RSS     AIC
## - Income          1    0.0002 22.596 -25.712
## - Illiteracy_t    1    0.1079 22.704 -25.475
## - Area_t          1    0.2368 22.833 -25.192
## <none>                        22.596 -23.713
## - Frost           1    1.1645 23.760 -23.200
## - Population_t    1    2.0155 24.611 -21.441
## - `HS Grad`       1    2.4822 25.078 -20.502
## - Murder          1   24.0347 46.631  10.512
##
## Step:  AIC=-25.71
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Illiteracy_t +
##     Area_t
```

```
##
##                 Df Sum of Sq    RSS      AIC
## - Illiteracy_t  1    0.1095 22.705 -27.4708
## - Area_t        1    0.2616 22.858 -27.1370
## <none>                      22.596 -25.7125
## - Frost         1    1.2628 23.859 -24.9936
## + Income        1    0.0002 22.596 -23.7129
## - Population_t  1    2.3859 24.982 -22.6937
## - `HS Grad`     1    4.4112 27.007 -18.7959
## - Murder        1   24.4834 47.079   8.9907
##
## Step:  AIC=-27.47
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t + Area_t
##
##                 Df Sum of Sq    RSS     AIC
## - Area_t        1    0.2157 22.921 -28.998
## <none>                      22.705 -27.471
## + Illiteracy_t  1    0.1095 22.596 -25.712
## + Income        1    0.0017 22.704 -25.475
## - Population_t  1    2.2792 24.985 -24.688
## - Frost         1    2.3760 25.082 -24.495
## - `HS Grad`     1    4.9491 27.655 -19.612
## - Murder        1   29.2296 51.935  11.899
##
## Step:  AIC=-29
## `Life Exp` ~ Murder + `HS Grad` + Frost + Population_t
##
##                 Df Sum of Sq    RSS     AIC
## <none>                      22.921 -28.998
## + Area_t        1     0.216 22.705 -27.471
## + Illiteracy_t  1     0.064 22.858 -27.137
## + Income        1     0.011 22.911 -27.021
## - Frost         1     2.214 25.135 -26.387
## - Population_t  1     2.450 25.372 -25.920
## - `HS Grad`     1     6.959 29.881 -17.741
## - Murder        1    34.109 57.031  14.578
```

```r
#lasso
fit_lasso = glmnet(x = as.matrix(sta_tidy[2:8]), y = sta_tidy$`Life Exp`, data = sta_tidy, lambda = 0.0

r2_bth = summary(fit_bth)$adj.r.squared
r2_lasso = 1 - cv_res$cvm / var(sta_tidy$`Life Exp`)
```

```r
library(caret)
```

```
## Warning:   'caret' R 4.3.2
```

```
##      lattice
```

```
##
##    'lattice'
```

```
## The following object is masked from 'package:faraway':
```

```
##
##     melanoma
```

```
##
##    'caret'
```

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

```r
#cross validation
train = trainControl(method = "cv", number = 5)

model_cv = train(`Life Exp` ~ Murder + `HS Grad` + Frost + Population_t, data = sta_tidy, method = 'lm'

model_cv$finalModel
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Coefficients:
##     (Intercept)           Murder  `\\`HS Grad\\``            Frost
##       68.720810        -0.290016         0.054550        -0.005174
##    Population_t
##        0.246836
```

```r
print(model_cv)
```

```
## Linear Regression
##
## 50 samples
##  4 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 50, 50, 50, 50, 50, 50, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.8175305  0.6995992  0.6762409
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```