Divorce study

Tyler Frankenberg

03/13/2022

Import packages

```
library(tidyverse)
library(ggcorrplot)
library(rms)
```

Import data

```
url <- "https://raw.githubusercontent.com/curdferguson/data621/main/datasets/divusa.txt"

divusa <- read_tsv(url, skip = 1, col_names = c("index", "year", "divorce", "unemployed", "femla
b", "marriage", "birth", "military"), show_col_types=FALSE)

divusa <- divusa[,2:8]
```

Glimpse dataset structure and each column's summary statistics

```
divusa %>% head(5)
```

```
## # A tibble: 5 x 7
##    year divorce unemployed femlab marriage birth military
##   <dbl>   <dbl>      <dbl>  <dbl>    <dbl> <dbl>    <dbl>
## 1  1920     8          5.2   22.7     92    118.     3.22
## 2  1921     7.2       11.7   22.8     83    120.     3.56
## 3  1922     6.6        6.7   22.9     79.7  111.     2.46
## 4  1923     7.1        2.4   23.0     85.2  110.     2.21
## 5  1924     7.2        5     23.1     80.3  111.     2.29
```

```
divusa %>% summary()
```

```
##       year         divorce        unemployed         femlab
##  Min.   :1920   Min.   : 6.10   Min.   : 1.200   Min.   :22.70
##  1st Qu.:1939   1st Qu.: 8.70   1st Qu.: 4.200   1st Qu.:27.47
##  Median :1958   Median :10.60   Median : 5.600   Median :37.10
##  Mean   :1958   Mean   :13.27   Mean   : 7.173   Mean   :38.58
##  3rd Qu.:1977   3rd Qu.:20.30   3rd Qu.: 7.500   3rd Qu.:47.80
##  Max.   :1996   Max.   :22.80   Max.   :24.900   Max.   :59.30
##     marriage          birth           military
##  Min.   : 49.70   Min.   : 65.30   Min.   : 1.940
##  1st Qu.: 61.90   1st Qu.: 68.90   1st Qu.: 3.469
##  Median : 74.10   Median : 85.90   Median : 9.102
##  Mean   : 72.97   Mean   : 88.89   Mean   :12.365
##  3rd Qu.: 80.00   3rd Qu.:107.30   3rd Qu.:14.266
##  Max.   :118.10   Max.   :122.90   Max.   :86.641
```

```
divusa <- divusa %>% mutate(year = year,
                            divorce = divorce / 1000,
                            unemployed = unemployed / 100,
                            femlab = femlab / 100,
                            marriage = marriage / 1000,
                            birth = birth / 1000,
                            military = military / 1000)

divusa %>% head(5)
```

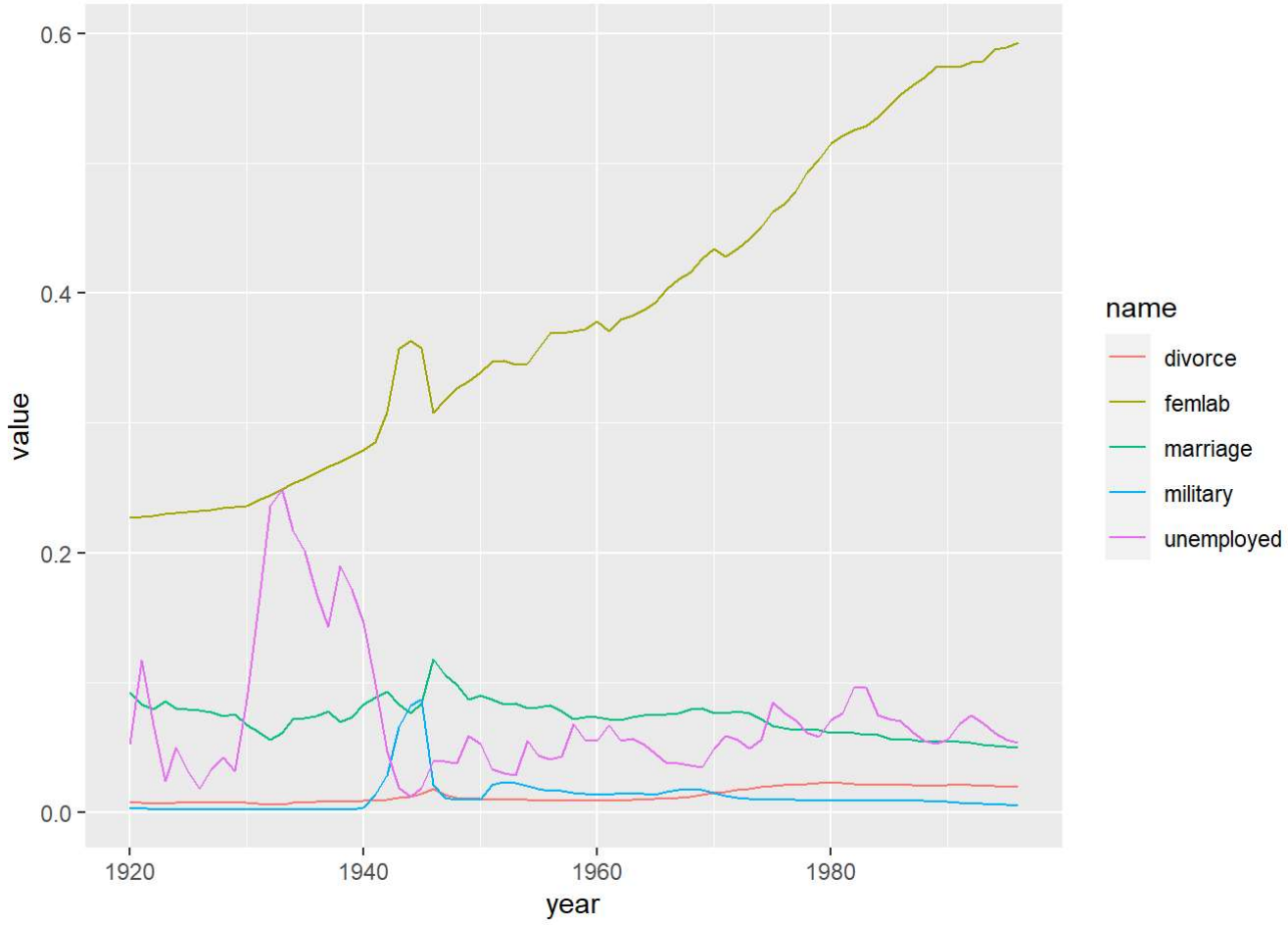
```
## # A tibble: 5 x 7
##    year divorce unemployed femlab marriage birth military
##   <dbl>   <dbl>      <dbl>  <dbl>    <dbl> <dbl>    <dbl>
## 1  1920  0.008       0.052  0.227   0.092  0.118  0.00322
## 2  1921  0.0072      0.117  0.228   0.083  0.120  0.00356
## 3  1922  0.0066      0.067  0.229   0.0797 0.111  0.00246
## 4  1923  0.0071      0.024  0.230   0.0852 0.110  0.00221
## 5  1924  0.0072      0.05   0.231   0.0803 0.111  0.00229
```

```
divusa %>% summary()
```

```
##       year         divorce          unemployed         femlab      
##  Min.   :1920   Min.   :0.00610   Min.   :0.01200   Min.   :0.2270  
##  1st Qu.:1939   1st Qu.:0.00870   1st Qu.:0.04200   1st Qu.:0.2747  
##  Median :1958   Median :0.01060   Median :0.05600   Median :0.3710  
##  Mean   :1958   Mean   :0.01327   Mean   :0.07173   Mean   :0.3858  
##  3rd Qu.:1977   3rd Qu.:0.02030   3rd Qu.:0.07500   3rd Qu.:0.4780  
##  Max.   :1996   Max.   :0.02280   Max.   :0.24900   Max.   :0.5930  
##     marriage           birth            military        
##  Min.   :0.04970   Min.   :0.06530   Min.   :0.001940  
##  1st Qu.:0.06190   1st Qu.:0.06890   1st Qu.:0.003469  
##  Median :0.07410   Median :0.08590   Median :0.009102  
##  Mean   :0.07297   Mean   :0.08889   Mean   :0.012365  
##  3rd Qu.:0.08000   3rd Qu.:0.10730   3rd Qu.:0.014266  
##  Max.   :0.11810   Max.   :0.12290   Max.   :0.086641
```

View change in each column over time

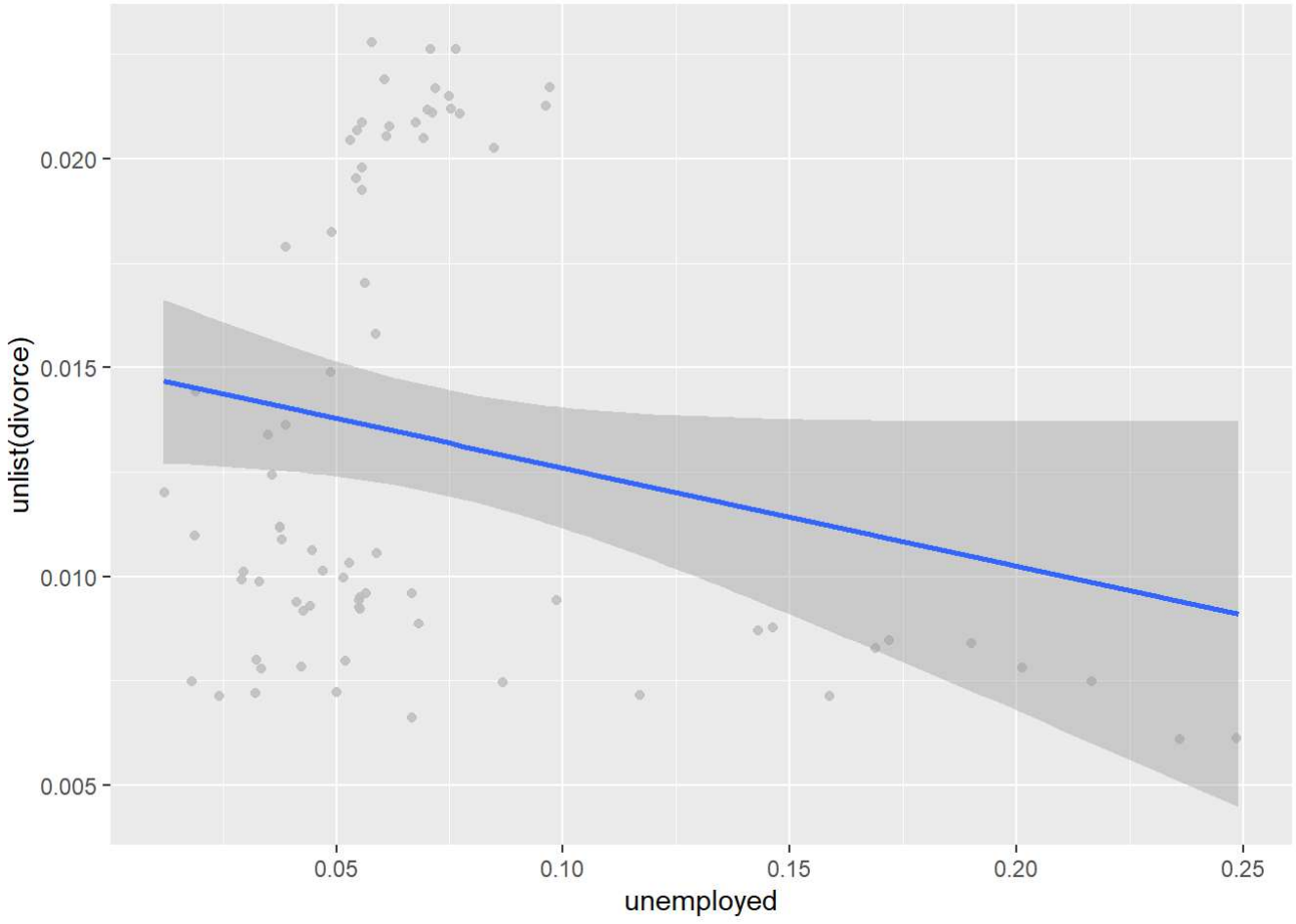
```
divusa_numeric_long <- divusa %>% select(!birth) %>% pivot_longer(colnames(select(divusa, !c(year, birth)))) %>% as.data.frame()

ggplot(data=divusa_numeric_long, aes(x=year, y=value, group=name, fill=name)) + geom_line(aes(col=name))
```



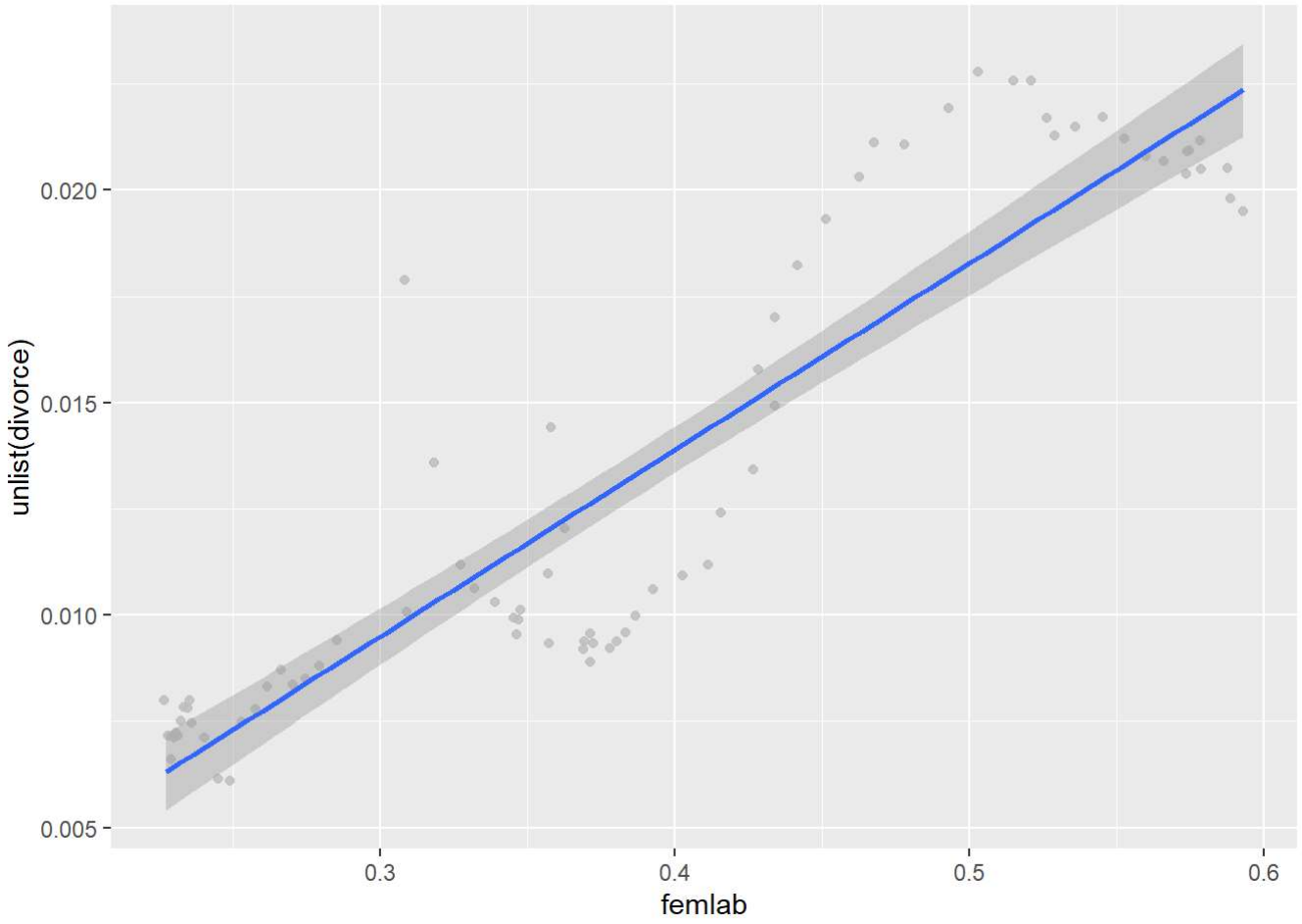
View scatter plots of each column

```
plot_list <- lapply(colnames(divusa[, 3:7]), function(c) {
  ggplot(divusa, aes(unlist(divusa[, c]), unlist(divorce))) +
    geom_jitter(color="gray", alpha=0.85) +
    geom_smooth(method = "lm") + xlab(c)
 })


plot_list
```

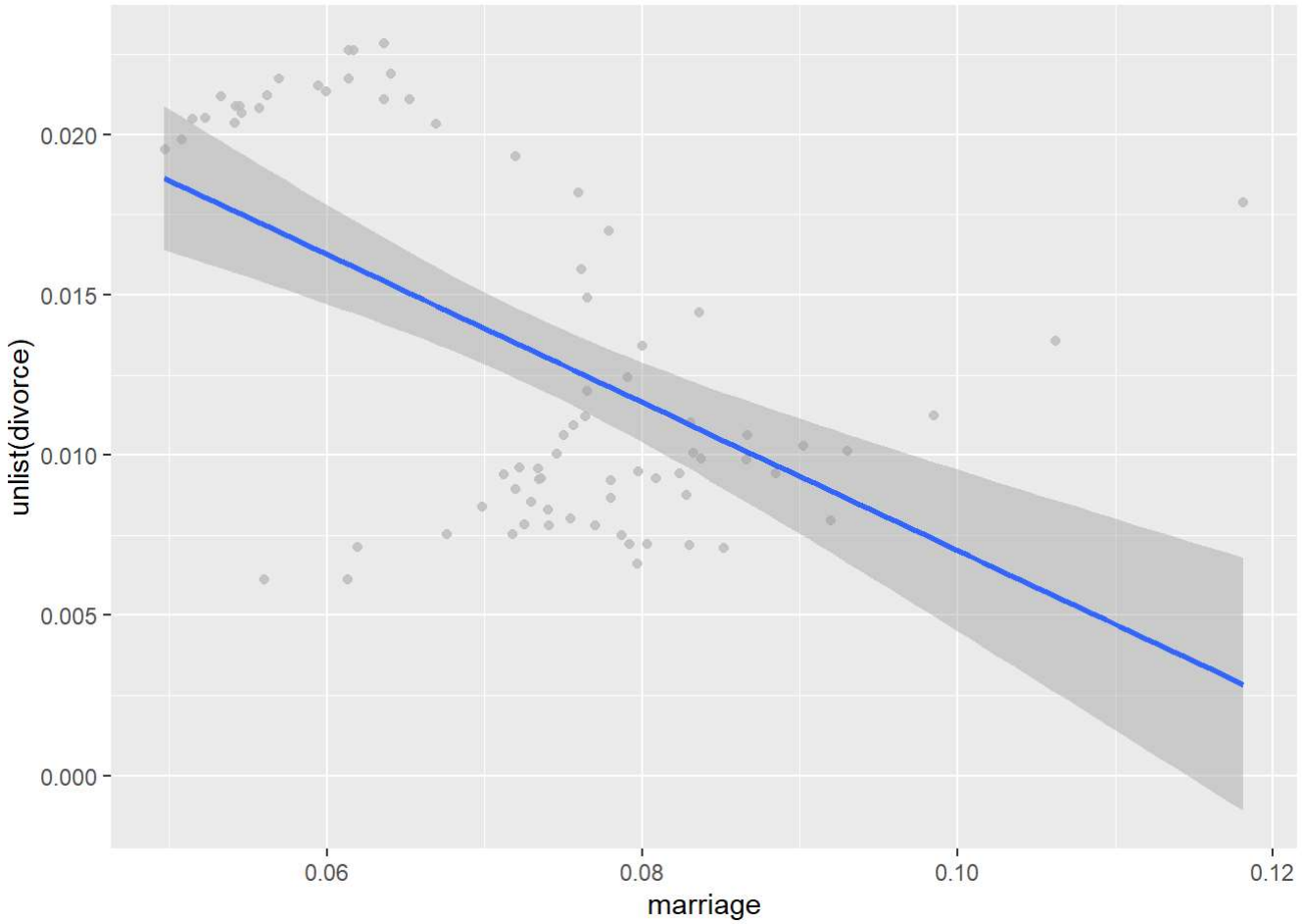
```
## [[1]]
```



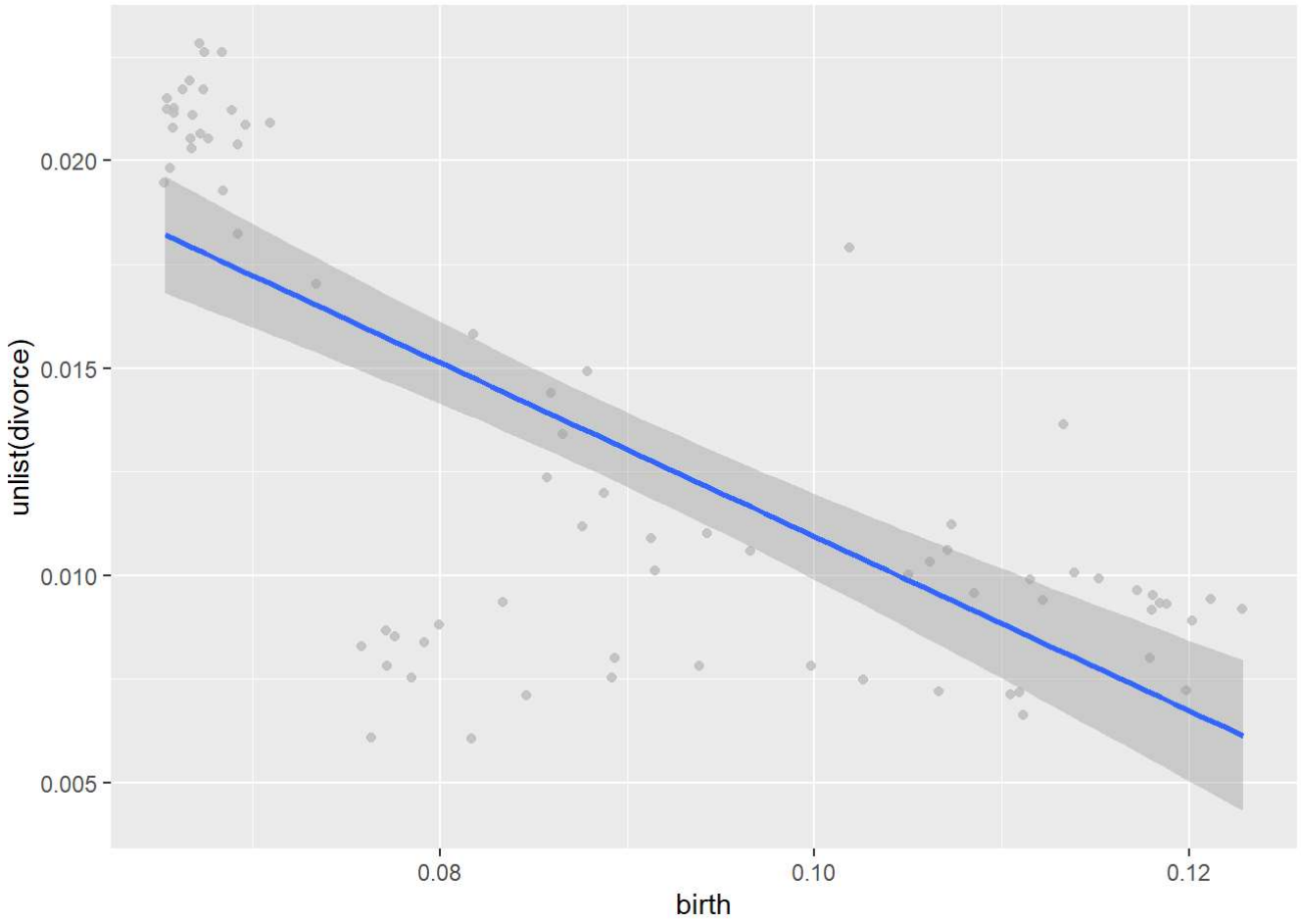

```
##
## [[2]]
```



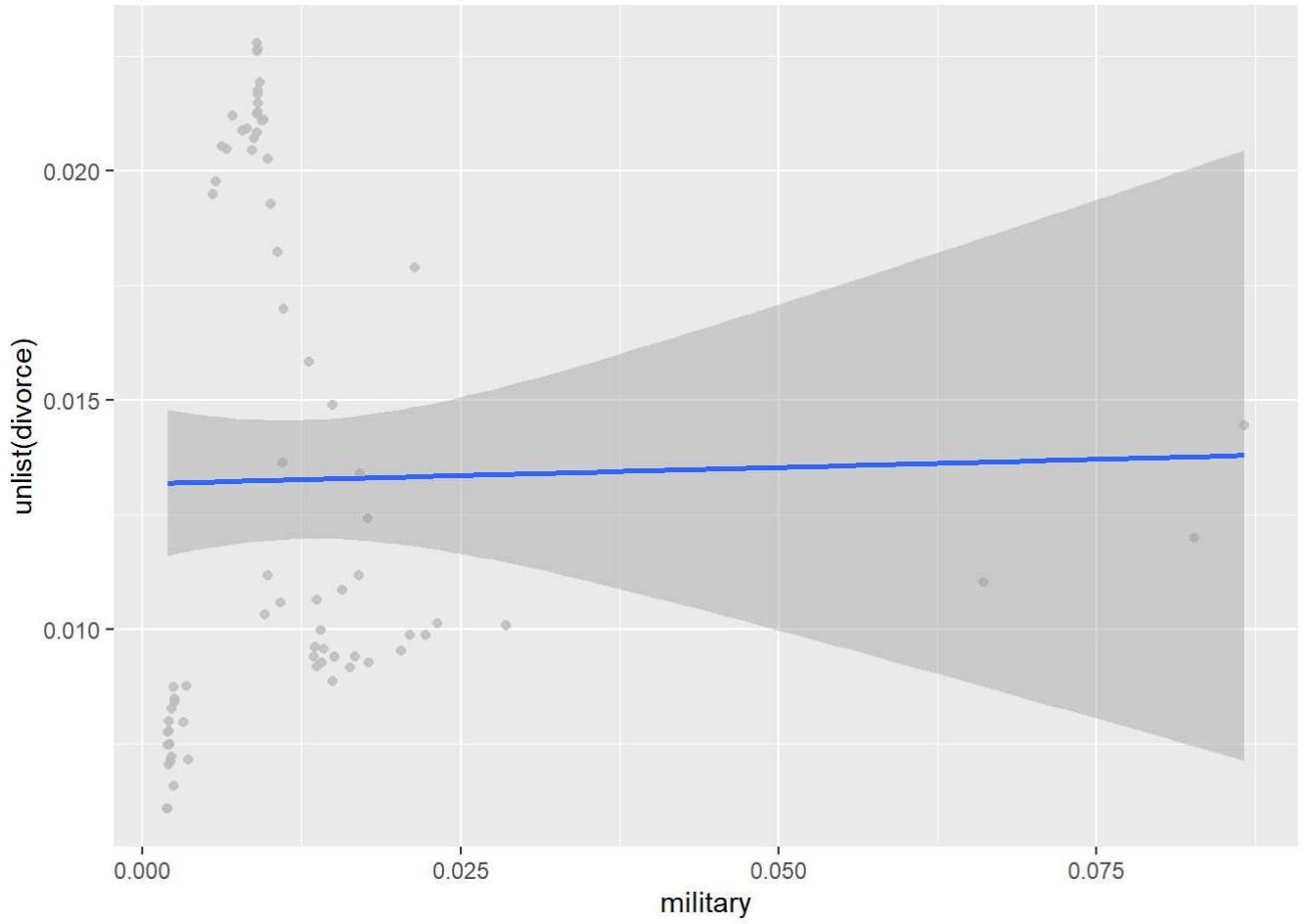

```
##
## [[3]]
```


0.020
0.015
0.010
0.005
0.000
unlist(divorce)
0.06
0.08
0.10
0.12
marriage

```
##
## [[4]]
```




```
##
## [[5]]
```

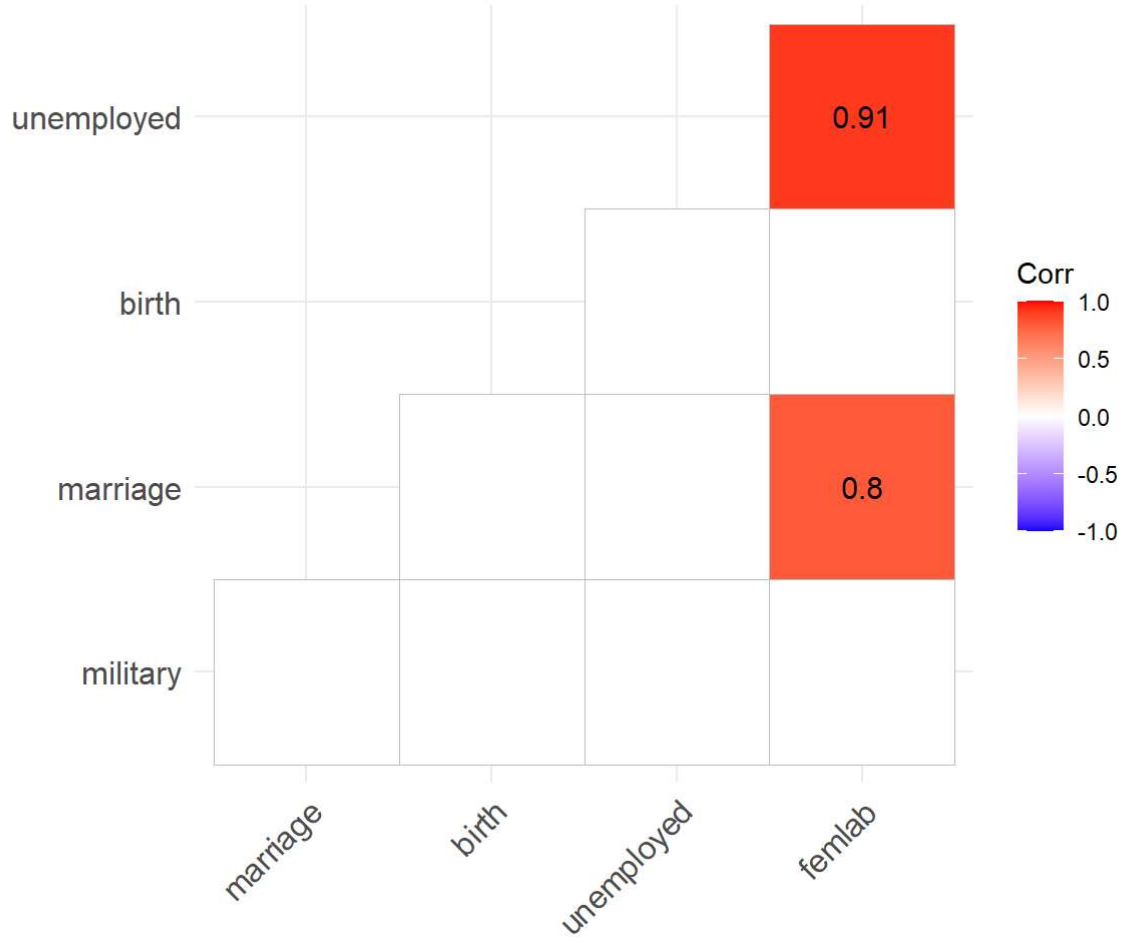


Fit linear model and address multicolinearity

```
div_lm1 <- lm(divorce ~ unemployed + femlab + marriage + birth + military, data=divusa)
div_lm1_sum <- summary(div_lm1, cor=TRUE)
div_lm1_sum
```

```
##
## Call:
## lm(formula = divorce ~ unemployed + femlab + marriage + birth +
##     military, data = divusa)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0038611 -0.0008916 -0.0000496  0.0008650  0.0038300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002488   0.003394   0.733   0.4659
## unemployed  -0.011125   0.005592  -1.989   0.0505 .
## femlab       0.038365   0.003059  12.543  < 2e-16 ***
## marriage     0.118674   0.024414   4.861 6.77e-06 ***
## birth       -0.129959   0.015595  -8.333 4.03e-12 ***
## military    -0.026734   0.014247  -1.876   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00165 on 71 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9152
## F-statistic: 165.1 on 5 and 71 DF,  p-value: < 2.2e-16
##
## Correlation of Coefficients:
##            (Intercept) unemployed femlab marriage birth
## unemployed -0.75
## femlab     -0.93        0.65
## marriage   -0.69        0.37       0.57
## birth      -0.55        0.47       0.51  -0.17
## military    0.02        0.24      -0.06  -0.21     0.07
```

```
# correlation matrix

corr1 <- cor(div_lm1_sum$correlation)[2:6, 2:6]
p.mat1 <- cor_pmat(div_lm1_sum$correlation)[2:6, 2:6]

ggcorr1 <- ggcorrplot(corr1, p.mat= p.mat1, hc.order = TRUE,
    type = "lower", insig = "blank", lab=TRUE)

ggcorr1
```




```
# Variance Inflation Factor
vif1 <- vif(div_lm1)
vif1
```

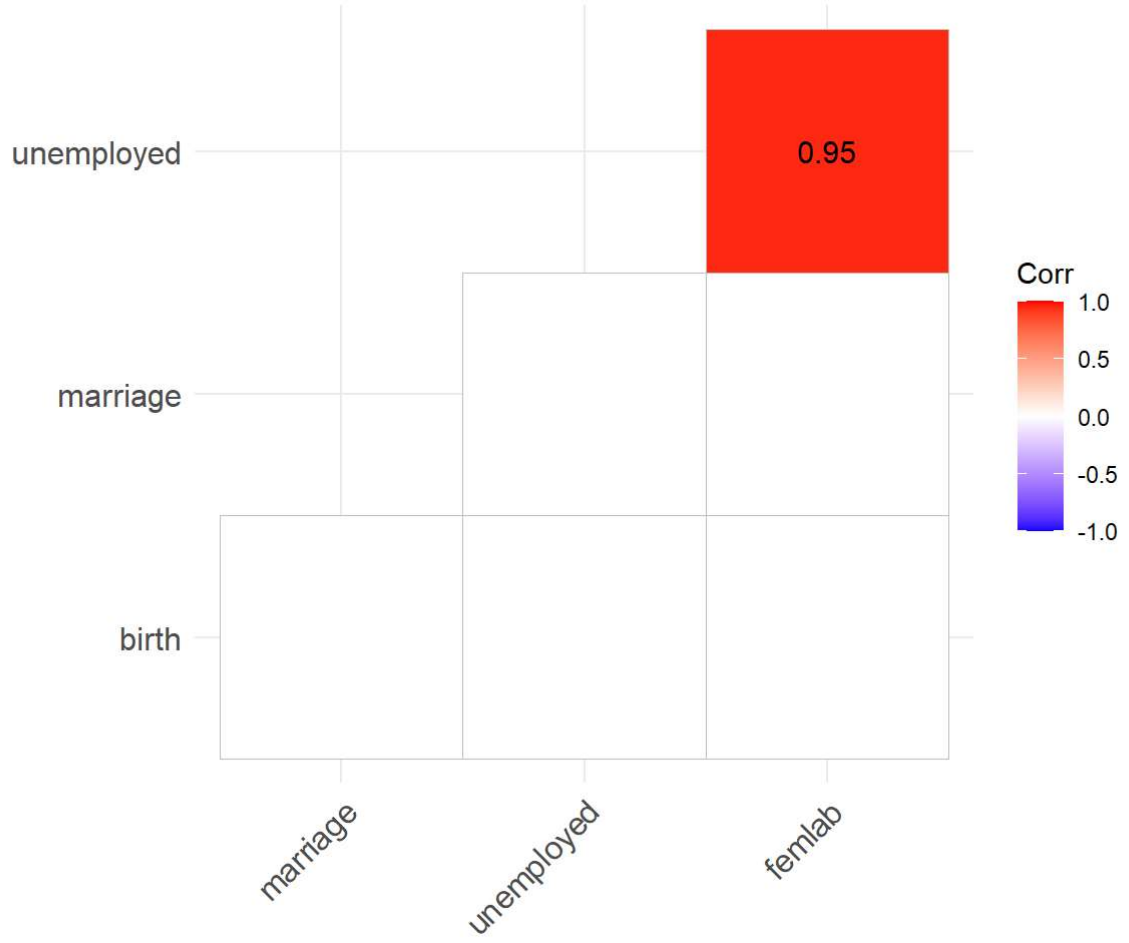
```
## unemployed     femlab   marriage      birth   military
##   2.252888   3.613276   2.864864   2.585485   1.249596
```

Remove an insignificant variable and re-fit model

```
#Fit 2nd model, removing military
div_lm2 <- lm(divorce ~ unemployed + femlab + marriage + birth, data=divusa)
div_lm2_sum <- summary(div_lm2, cor=TRUE)
div_lm2_sum
```

```
##
## Call:
## lm(formula = divorce ~ unemployed + femlab + marriage + birth,
##     data = divusa)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0038559 -0.0009217 -0.0000119  0.0009055  0.0040391
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002613   0.003452   0.757    0.451
## unemployed  -0.008613   0.005524  -1.559    0.123
## femlab       0.038004   0.003106  12.237  < 2e-16 ***
## marriage     0.109253   0.024307   4.495 2.61e-05 ***
## birth       -0.127819   0.015823  -8.078 1.10e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001679 on 72 degrees of freedom
## Multiple R-squared:  0.9169, Adjusted R-squared:  0.9123
## F-statistic: 198.6 on 4 and 72 DF,  p-value: < 2.2e-16
##
## Correlation of Coefficients:
##            (Intercept) unemployed femlab marriage
## unemployed -0.77
## femlab     -0.93        0.69
## marriage   -0.70        0.44       0.57
## birth      -0.56        0.47       0.51  -0.16
```

```
# correlation matrix

corr2 <- cor(div_lm2_sum$correlation)[2:5, 2:5]
p.mat2 <- cor_pmat(div_lm2_sum$correlation)[2:5, 2:5]

ggcorr2 <- ggcorrplot(corr2, p.mat= p.mat2, hc.order = TRUE,
    type = "lower", insig = "blank", lab=TRUE)

ggcorr2
```




```
# Variance Inflation Factor
vif2 <- vif(div_lm2)
vif2
```

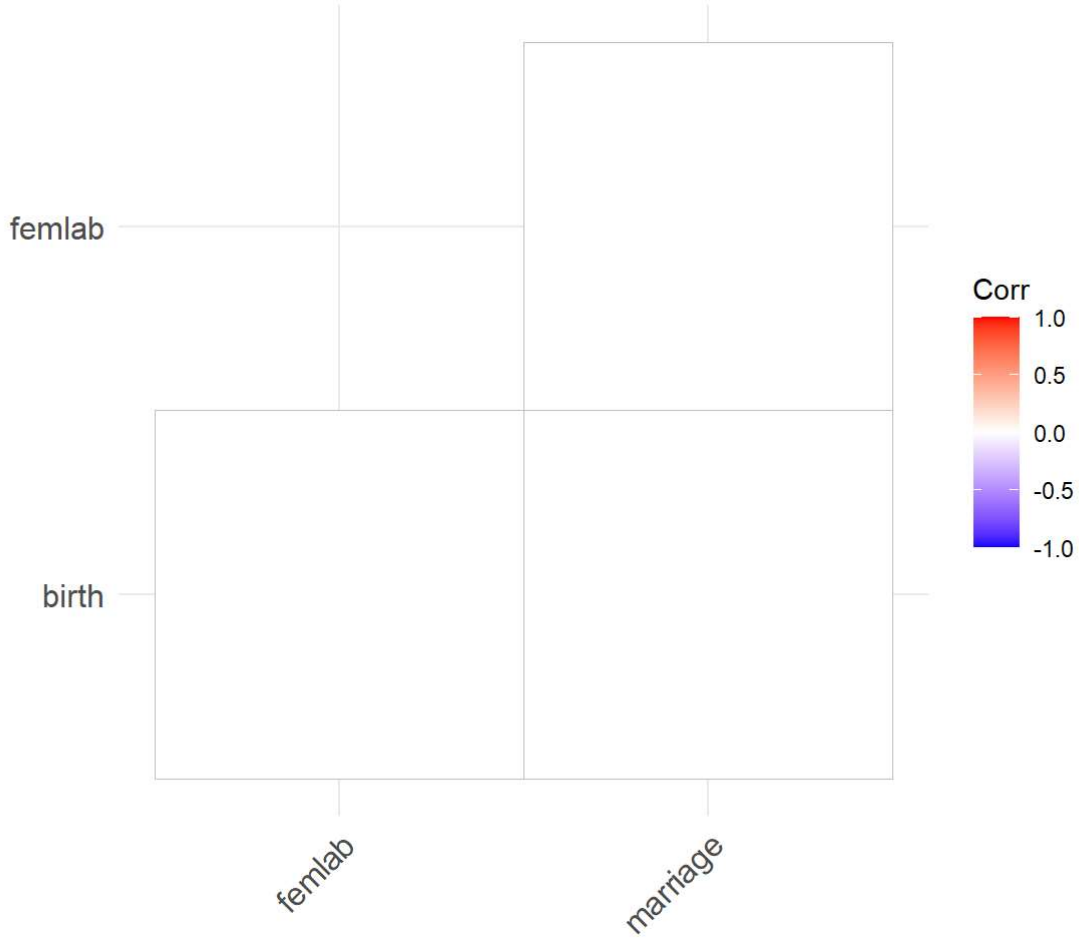
```
## unemployed     femlab   marriage      birth
##   2.123735   3.598997   2.743709   2.571650
```

Remove least significant variable and re-fit model

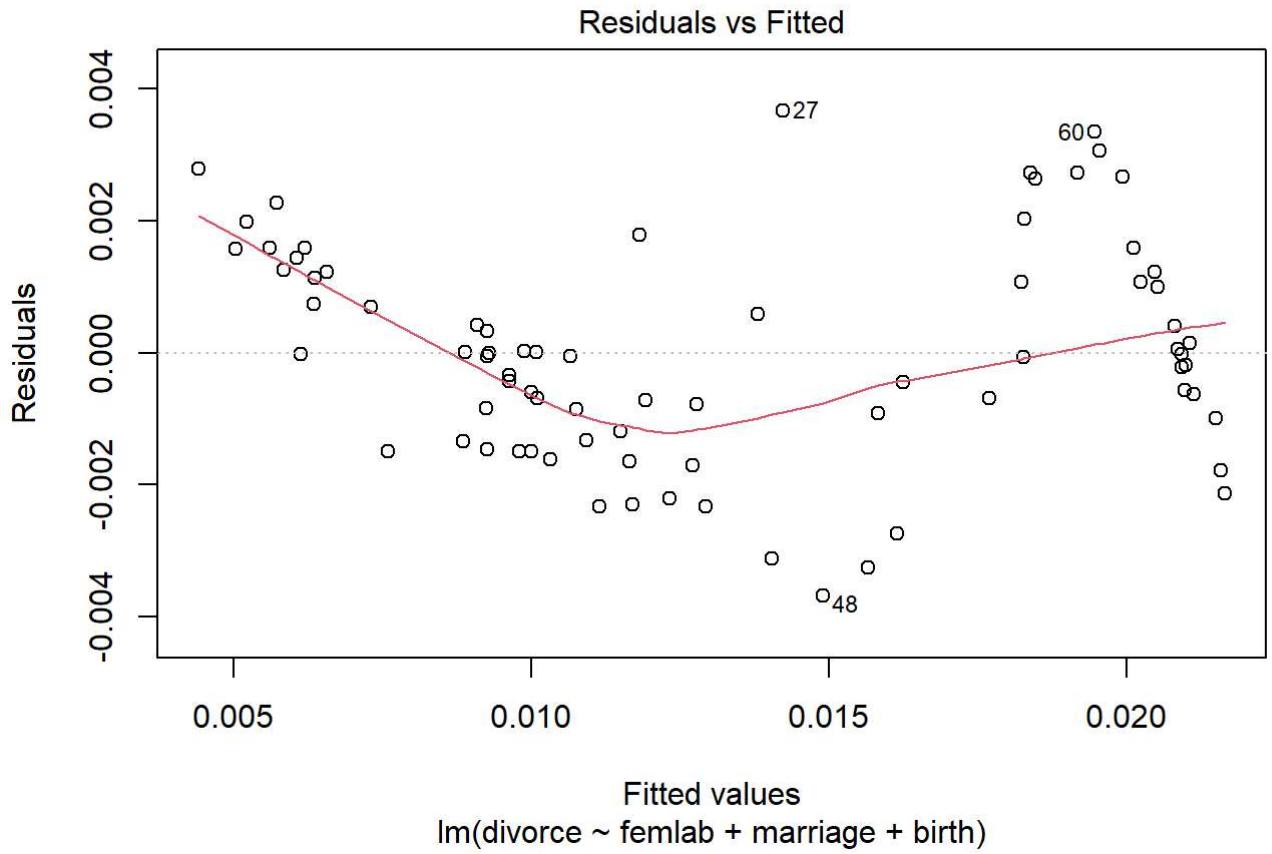
```
# Fit 3rd Model, removing unemployed
div_lm3 <- lm(divorce ~ femlab + marriage + birth, data=divusa)
div_lm3_sum <- summary(div_lm3, cor=TRUE)
div_lm3_sum
```

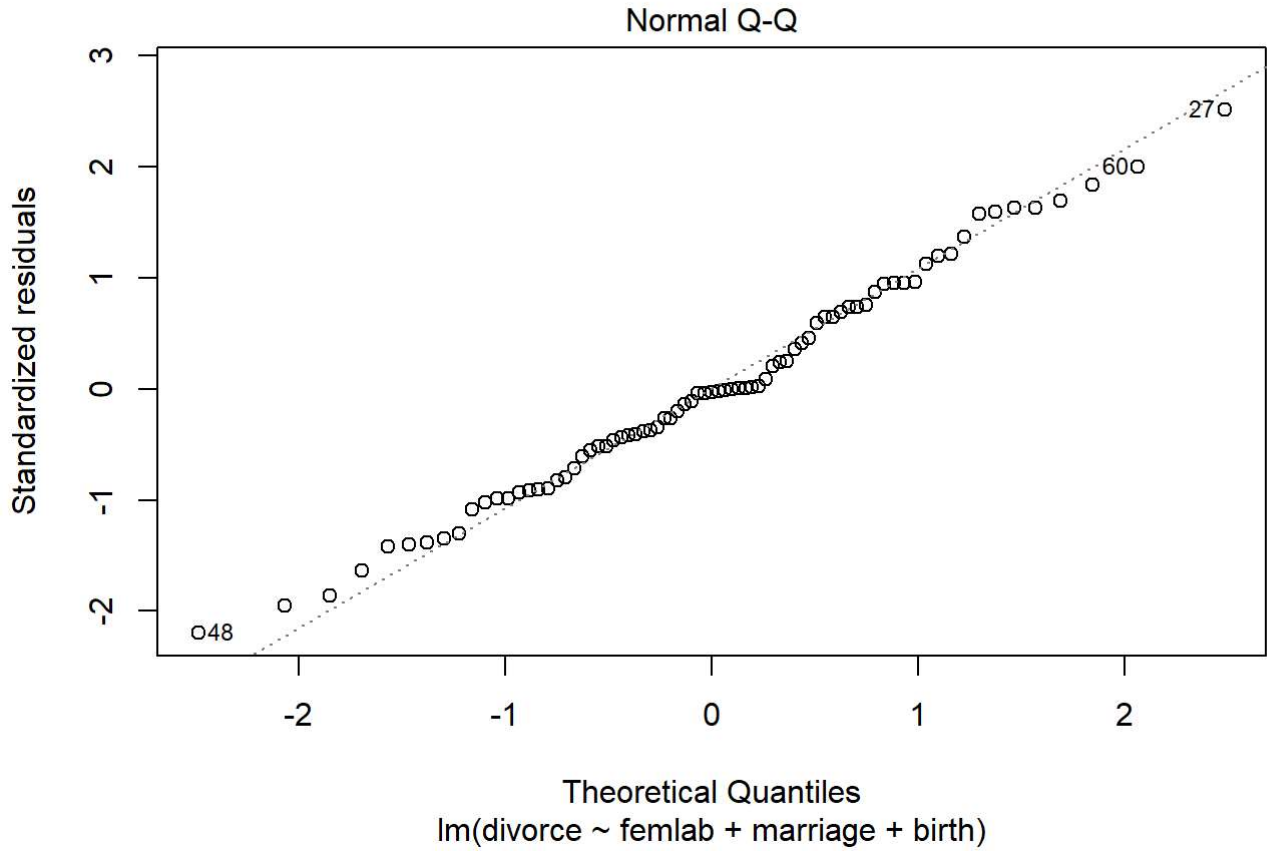
```
##
## Call:
## lm(formula = divorce ~ femlab + marriage + birth, data = divusa)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0036923 -0.0011934 -0.0000534  0.0012265  0.0036701
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.001545   0.002212  -0.699    0.487
## femlab       0.041337   0.002275  18.174  < 2e-16 ***
## marriage     0.126094   0.021988   5.735 2.07e-07 ***
## birth       -0.116274   0.014120  -8.235 5.10e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001695 on 73 degrees of freedom
## Multiple R-squared:  0.9141, Adjusted R-squared:  0.9106
## F-statistic: 258.9 on 3 and 73 DF,  p-value: < 2.2e-16
##
## Correlation of Coefficients:
##          (Intercept) femlab marriage
## femlab   -0.86
## marriage -0.62        0.41
## birth    -0.35        0.30  -0.46
```

```
# correlation matrix

corr3 <- cor(div_lm3_sum$correlation)[2:4, 2:4]
p.mat3 <- cor_pmat(div_lm3_sum$correlation)[2:4, 2:4]

ggcorr3 <- ggcorrplot(corr3, p.mat= p.mat3, hc.order = TRUE,
    type = "lower", insig = "blank", lab=TRUE)

ggcorr3
```

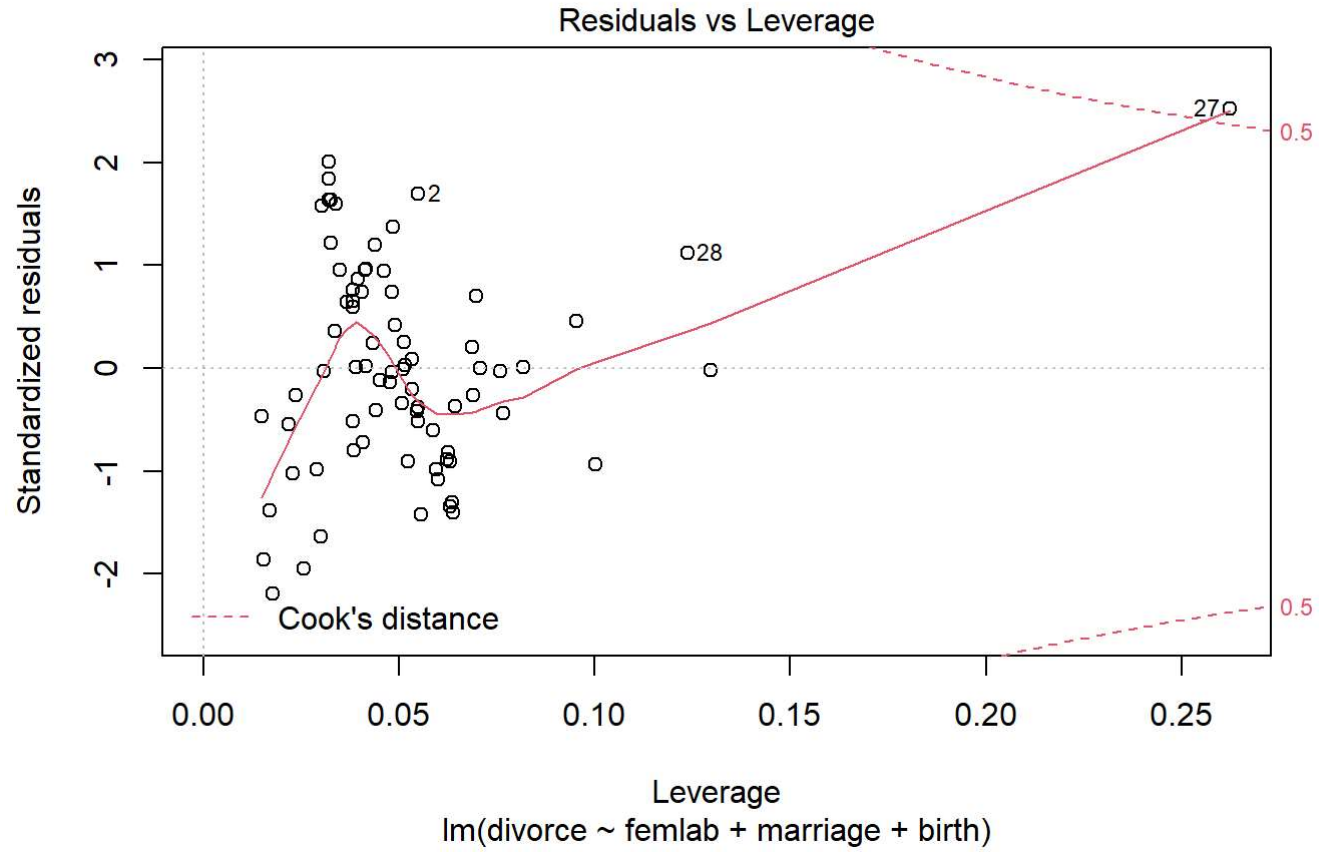



```
plot(div_lm3)
```


Residuals vs Fitted
Residuals
0.004
0.002
0.000
-0.002
-0.004
27
60
48
0.005
0.010
0.015
0.020
Fitted values
lm(divorce ~ femlab + marriage + birth)


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
27
60
48
-2
-1
0
1
2
Theoretical Quantiles
lm(divorce ~ femlab + marriage + birth)


Scale-Location
27
48
60
√|Standardized residuals|
1.5
1.0
0.5
0.0
0.005
0.010
0.015
0.020
Fitted values
lm(divorce ~ femlab + marriage + birth)


Residuals vs Leverage
27
2
28
Standardized residuals
3
2
1
0
-1
-2
0.5
0.5
Cook's distance
0.00
0.05
0.10
0.15
0.20
0.25
Leverage
lm(divorce ~ femlab + marriage + birth)

```
# Variance Inflation Factor
vif3 <- vif(div_lm3)
vif3
```

```
##   femlab marriage    birth
## 1.893390 2.201891 2.008469
```

```
```