# Regression

*Rita Alvarez and Rafael Schmuziger*

*November 11, 2016*

In our basic model we want to see what is the relationship between youth unemployment and voter turnout in General Election in Spain.For this we use the following equation:

$$Turnout_{it} = \beta_0 + \beta_1 Youth\ Unemployment\ Rate_{it} + \epsilon_{it}$$

where $i$ represents each of the 17 regions plus two Autonomous cities of Spain and $t$ represents each of the years we have data for $t = 2004, 2008, 2011, 2015$

plot(Reg1,which=1)

Call: lm(formula = Turnout ~ Unempl_rate_less25, data = Spain_data)

Residuals: Min 1Q Median 3Q Max -0.16960 -0.03690 0.01063 0.04226 0.09288

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.7992339 0.0171272 46.665 < 2e-16 ***Unempl_rate_less25 -0.0022214 0.0004454 -4.988 3.92e-06*** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 '' 1

Residual standard error: 0.05698 on 74 degrees of freedom Multiple R-squared: 0.2516, Adjusted R-squared: 0.2415 F-statistic: 24.88 on 1 and 74 DF, p-value: 3.916e-06

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: do., nov. 27, 2016 - 20:36:25

Table 1: Basic Regression

| | *Dependent variable:* |
|---|---|
| | Voter Turnout in General Elections |
| Youth Unemloyment | $-0.002^{***}$ |
| | (0.0004) |
| Constant | $0.799^{***}$ |
| | (0.017) |
| Observations | 76 |
| $R^2$ | 0.252 |
| Adjusted $R^2$ | 0.241 |
| Residual Std. Error | 0.057 (df = 74) |
| F Statistic | $24.877^{***}$ (df = 1; 74) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Our analysis can also go one step further and iclude regional fixed effects

$$Turnout_{it} = \beta_0 + \beta_1 Youth\ Unemployment\ Rate_{it} + \delta_i + \epsilon_{it}$$

```
Reg2 <- lm(Turnout ~ Unempl_rate_less25 + Region, data = Spain_data)
```

```r
summary(Reg2)
```

```
##
## Call:
## lm(formula = Turnout ~ Unempl_rate_less25 + Region, data = Spain_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.070112 -0.008874 -0.000297  0.008067  0.057838
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.770601   0.013073  58.945  < 2e-16 ***
## Unempl_rate_less25          -0.001099   0.000187  -5.879 2.39e-07 ***
## RegionAragon                 0.011345   0.014922   0.760 0.450292
## RegionAsturias              -0.021044   0.014765  -1.425 0.159627
## RegionBaleares              -0.075798   0.014822  -5.114 3.98e-06 ***
## RegionCanarias              -0.071002   0.014701  -4.830 1.10e-05 ***
## RegionCantabria              0.025146   0.014866   1.692 0.096287 .
## RegionCastilla_La_Mancha     0.051866   0.014745   3.518 0.000872 ***
## RegionCastilla_y_Leon        0.031933   0.014800   2.158 0.035262 *
## RegionCatalunya             -0.025340   0.014830  -1.709 0.093046 .
## RegionCeuta                 -0.115844   0.014845  -7.804 1.63e-10 ***
## RegionComunitat_Valenciana   0.041221   0.014737   2.797 0.007055 **
## RegionExtremadura            0.036837   0.014701   2.506 0.015154 *
## RegionGalicia               -0.019814   0.014826  -1.336 0.186822
## RegionLa_Rioja               0.040443   0.014875   2.719 0.008708 **
## RegionMadrid                 0.042142   0.014869   2.834 0.006377 **
## RegionMelilla               -0.149735   0.014800 -10.117 2.99e-14 ***
## RegionMurcia                 0.030969   0.014760   2.098 0.040414 *
## RegionNavarra               -0.008258   0.015049  -0.549 0.585357
## RegionPais_Vasco            -0.038778   0.014889  -2.604 0.011763 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02079 on 56 degrees of freedom
## Multiple R-squared:  0.9246, Adjusted R-squared:  0.8991
## F-statistic: 36.16 on 19 and 56 DF,  p-value: < 2.2e-16
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: do., nov. 27, 2016 - 20:36:26

```r
Reg3 <- lm(Turnout ~ Unempl_rate_less25 + eight + eleven + fifteen, data = Spain_data)
```

```r
summary(Reg3)
```

```
##
## Call:
## lm(formula = Turnout ~ Unempl_rate_less25 + eight + eleven +
##     fifteen, data = Spain_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.132415 -0.036335  0.007654  0.037552  0.090087
##
```

Table 2:

| | *Dependent variable:* | |
|---|---|---|
| | Turnout | |
| | (1) | (2) |
| Unempl_rate_less25 | $-0.002^{***}$ | $-0.001^{***}$ |
| | (0.0004) | (0.0002) |
| RegionAragon | | 0.011 |
| | | (0.015) |
| RegionAsturias | | $-0.021$ |
| | | (0.015) |
| RegionBaleares | | $-0.076^{***}$ |
| | | (0.015) |
| RegionCanarias | | $-0.071^{***}$ |
| | | (0.015) |
| RegionCantabria | | $0.025^{*}$ |
| | | (0.015) |
| RegionCastilla_La_Mancha | | $0.052^{***}$ |
| | | (0.015) |
| RegionCastilla_y_Leon | | $0.032^{**}$ |
| | | (0.015) |
| RegionCatalunya | | $-0.025^{*}$ |
| | | (0.015) |
| RegionCeuta | | $-0.116^{***}$ |
| | | (0.015) |
| RegionComunitat_Valenciana | | $0.041^{***}$ |
| | | (0.015) |
| RegionExtremadura | | $0.037^{**}$ |
| | | (0.015) |
| RegionGalicia | | $-0.020$ |
| | | (0.015) |
| RegionLa_Rioja | | $0.040^{***}$ |
| | | (0.015) |
| RegionMadrid | | $0.042^{***}$ |
| | | (0.015) |
| RegionMelilla | | $-0.150^{***}$ |
| | | (0.015) |
| RegionMurcia | | $0.031^{**}$ |
| | | (0.015) |
| RegionNavarra | | $-0.008$ |
| | | (0.015) |

```
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.8437398  0.0199905  42.207  < 2e-16 ***
## Unempl_rate_less25 -0.0047546  0.0007289  -6.523 8.69e-09 ***
## eight                0.0029997  0.0169314   0.177 0.859881
## eleven               0.0757919  0.0242326   3.128 0.002555 **
## fifteen              0.1033705  0.0260240   3.972 0.000169 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05181 on 71 degrees of freedom
## Multiple R-squared:  0.4063, Adjusted R-squared:  0.3729
## F-statistic: 12.15 on 4 and 71 DF,  p-value: 1.411e-07
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: do., nov. 27, 2016 - 20:36:26

Table 3:

| | *Dependent variable:* | |
|---|---|---|
| | Turnout | |
| | (1) | (2) |
| Unempl_rate_less25 | $-0.002^{***}$ | $-0.005^{***}$ |
| | (0.0004) | (0.001) |
| eight | | 0.003 |
| | | (0.017) |
| eleven | | $0.076^{***}$ |
| | | (0.024) |
| fifteen | | $0.103^{***}$ |
| | | (0.026) |
| Constant | $0.799^{***}$ | $0.844^{***}$ |
| | (0.017) | (0.020) |
| Observations | 76 | 76 |
| $R^2$ | 0.252 | 0.406 |
| Adjusted $R^2$ | 0.241 | 0.373 |
| Residual Std. Error | 0.057 (df = 74) | 0.052 (df = 71) |
| F Statistic | $24.877^{***}$ (df = 1; 74) | $12.148^{***}$ (df = 4; 71) |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: do., nov. 27, 2016 - 20:36:26

Spain_data[“unempl_rich”]<-Spain_data$Unempl_rate_less25*Spain_data$Rich

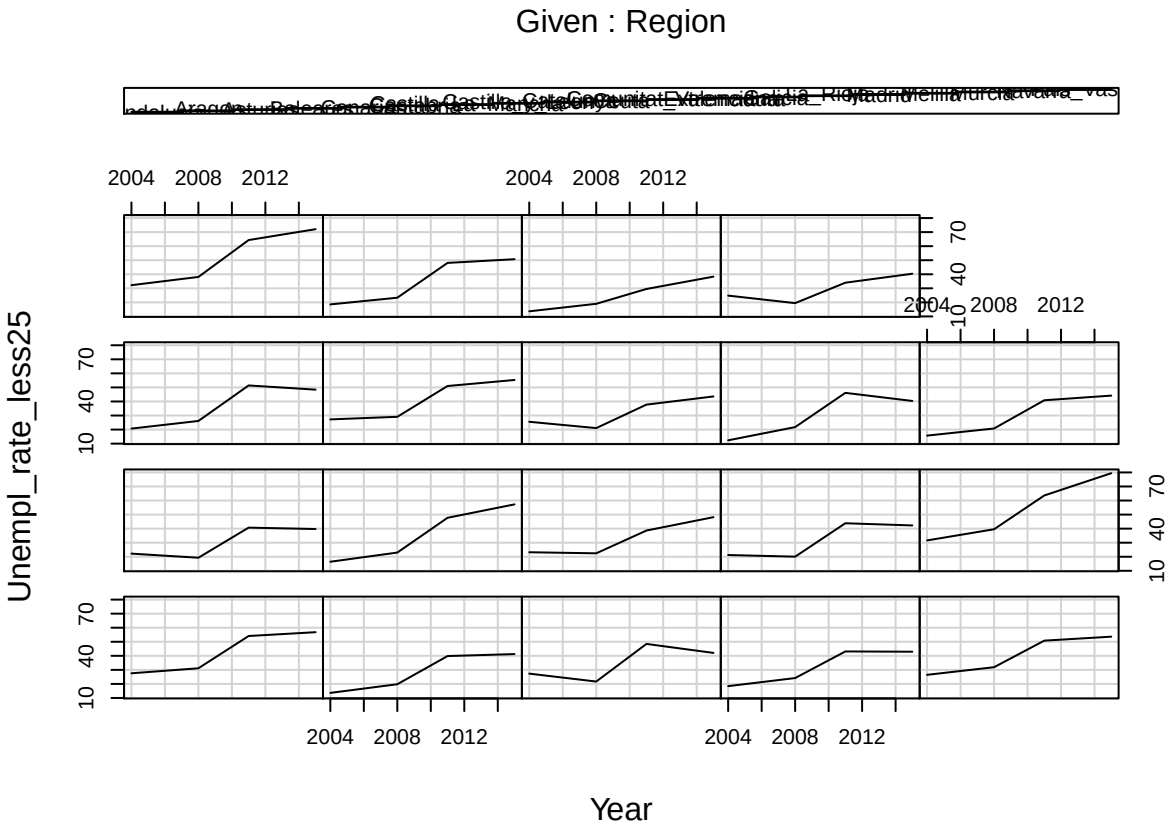
```
Spain_data["unempl_rich"]<-Spain_data$Unempl_rate_less25*Spain_data$Rich
Reg4 <- lm(Turnout ~ Unempl_rate_less25 + Rich + unempl_rich, data = Spain_data)
```

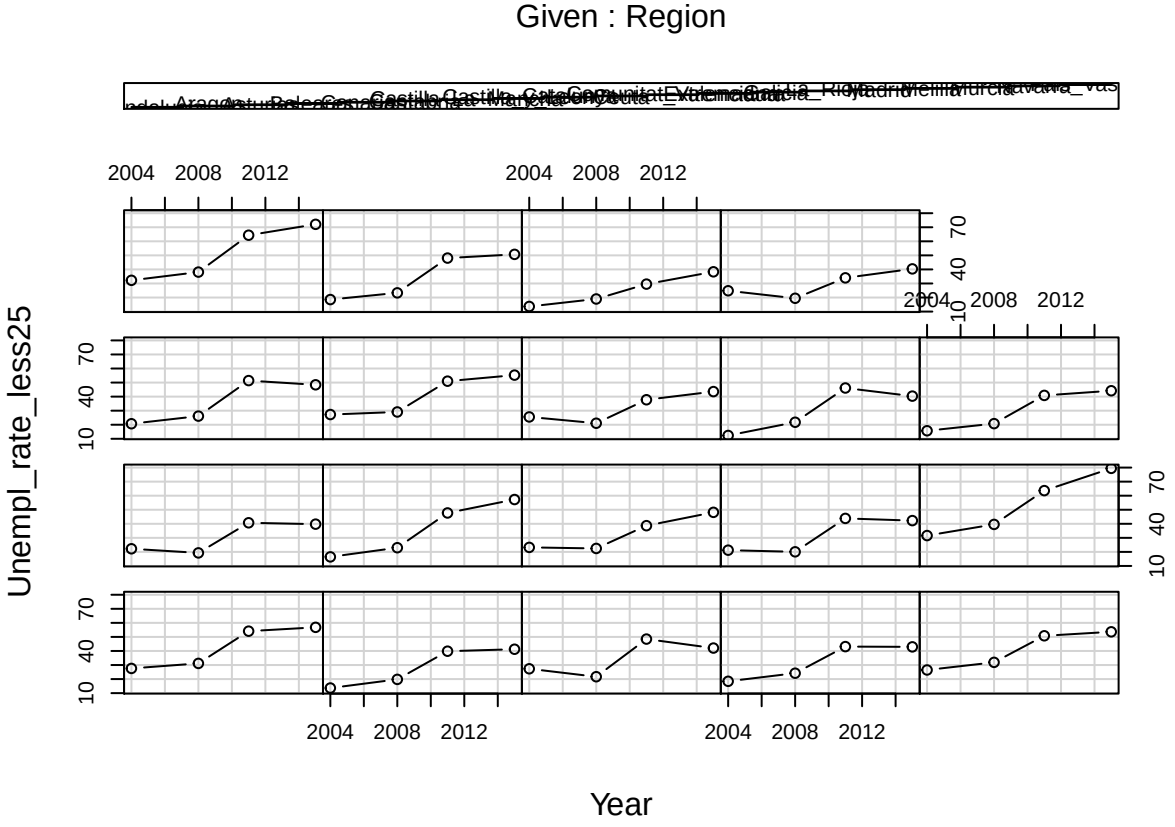
Table 4:

| | *Dependent variable:* | | |
|---|---|---|---|
| | Turnout | | |
| | (1) | (2) | (3) |
| Unempl_rate_less25 | −0.002*** | −0.001*** | −0.005*** |
| | (0.0004) | (0.0002) | (0.001) |
| RegionAragon | | 0.011 | |
| | | (0.015) | |
| RegionAsturias | | −0.021 | |
| | | (0.015) | |
| RegionBaleares | | −0.076*** | |
| | | (0.015) | |
| RegionCanarias | | −0.071*** | |
| | | (0.015) | |
| RegionCantabria | | 0.025* | |
| | | (0.015) | |
| RegionCastilla_La_Mancha | | 0.052*** | |
| | | (0.015) | |
| RegionCastilla_y_Leon | | 0.032** | |
| | | (0.015) | |
| RegionCatalunya | | −0.025* | |
| | | (0.015) | |
| RegionCeuta | | −0.116*** | |
| | | (0.015) | |
| RegionComunitat_Valenciana | | 0.041*** | |
| | | (0.015) | |
| RegionExtremadura | | 0.037** | |
| | | (0.015) | |
| RegionGalicia | | −0.020 | |
| | | (0.015) | |
| RegionLa_Rioja | | 0.040*** | |
| | | (0.015) | |
| RegionMadrid | | 0.042*** | |
| | | (0.015) | |
| RegionMelilla | | −0.150*** | |
| | | (0.015) | |
| RegionMurcia | | 0.031** | |
| | | (0.015) | |
| RegionNavarra | | −0.008 | |
| | | (0.015) | |

```
summary(Reg4)
```

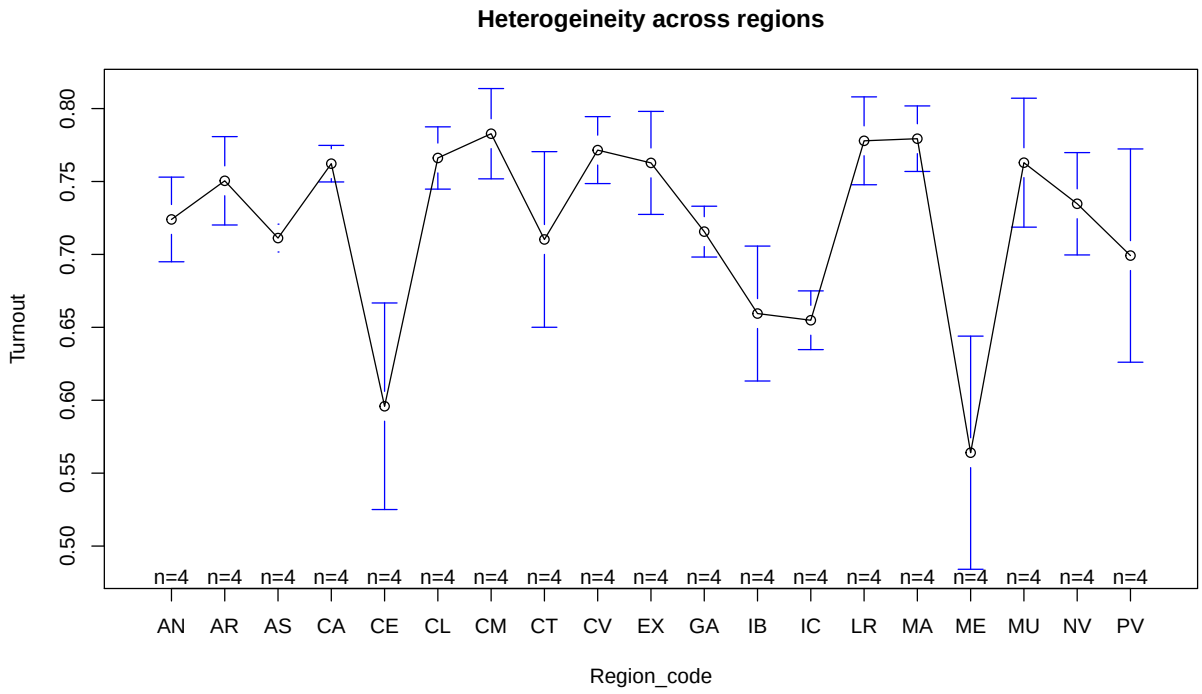
```
##
## Call:
## lm(formula = Turnout ~ Unempl_rate_less25 + Rich + unempl_rich,
##     data = Spain_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16602 -0.03755  0.01333  0.04190  0.08534
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.8109865  0.0248366  32.653  < 2e-16 ***
## Unempl_rate_less25 -0.0026969  0.0005793  -4.656 1.44e-05 ***
## Rich               -0.0446612  0.0352602  -1.267   0.2094
## unempl_rich         0.0018206  0.0009570   1.902   0.0611 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05577 on 72 degrees of freedom
## Multiple R-squared:  0.3024, Adjusted R-squared:  0.2733
## F-statistic:  10.4 on 3 and 72 DF,  p-value: 9.089e-06
```

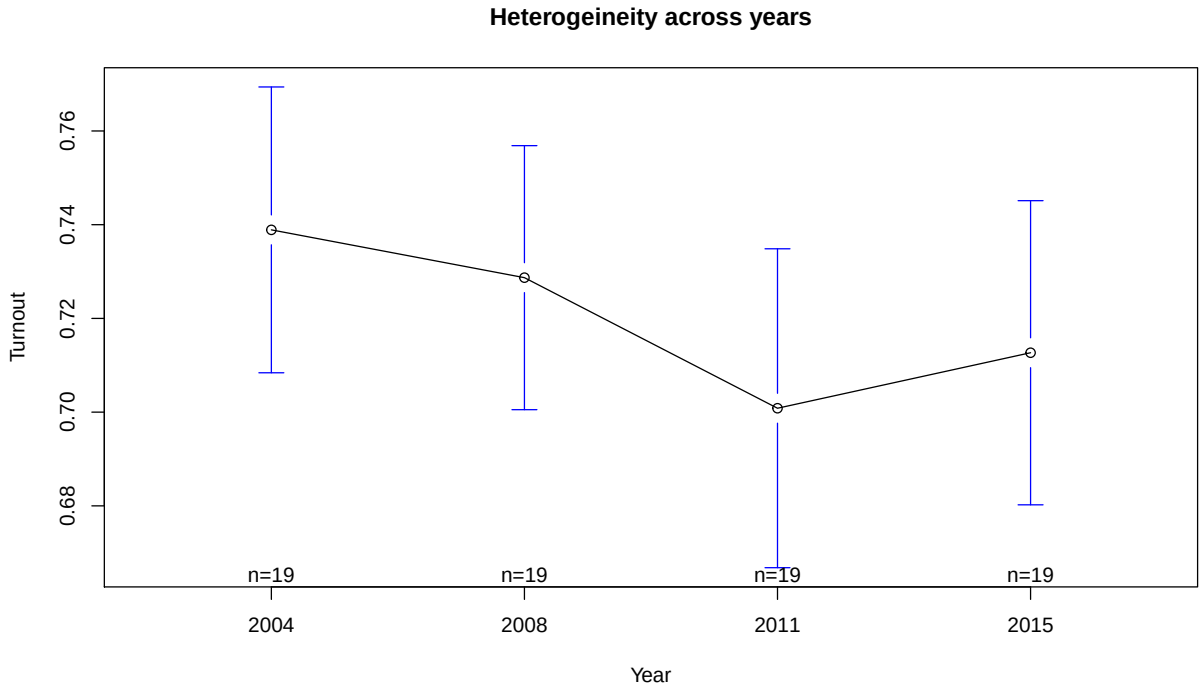


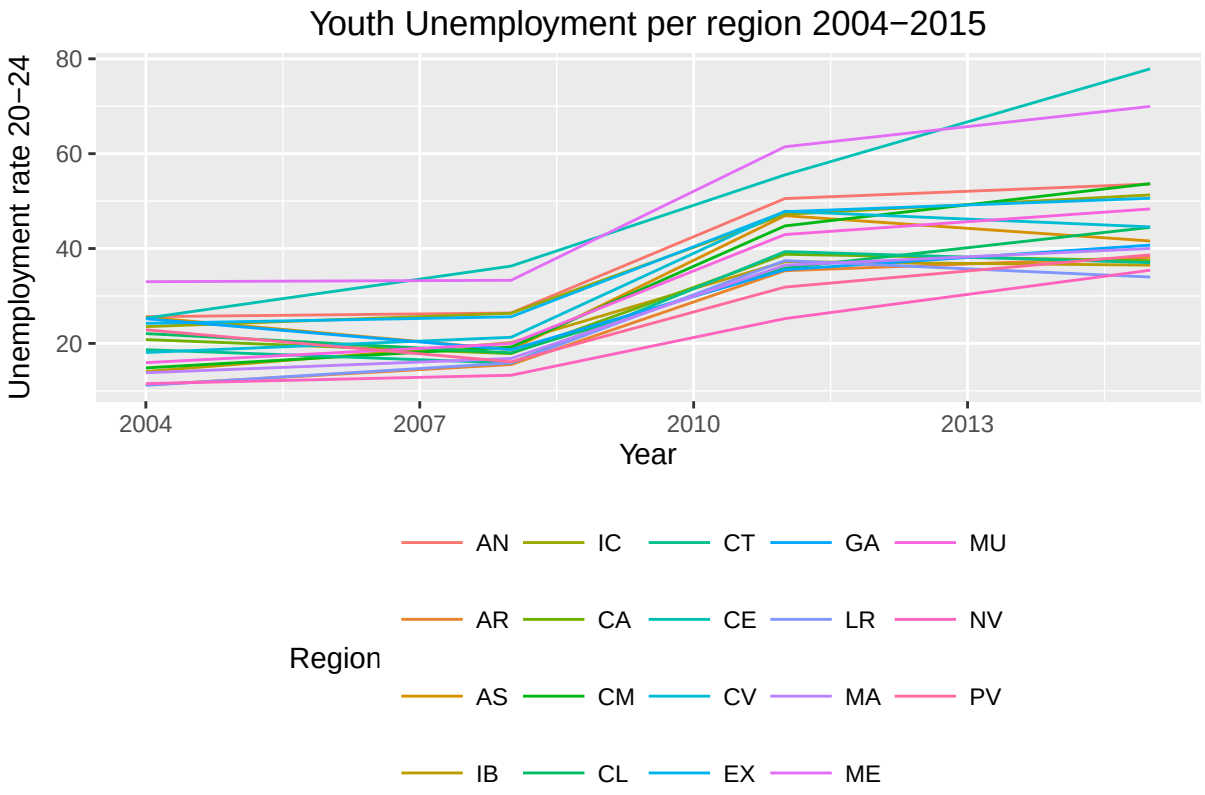



```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :
## zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :
## zero-length arrow is of indeterminate angle and so skipped
```


Heterogeineity across regions
Turnout
0.50
0.55
0.60
0.65
0.70
0.75
0.80
n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4 n=4
AN AR AS CA CE CL CM CT CV EX GA IB IC LR MA ME MU NV PV
Region_code


Heterogeineity across years
Turnout
0.68
0.70
0.72
0.74
0.76
n=19 n=19 n=19 n=19
2004 2008 2011 2015
Year




legend.key = element_rect(size = 5, fill = 'white', colour = 'white'), legend.key.size = unit(2.0, 'lines'), axis.title.y=element_text(margin=margin(10,10,0,10)))