

HA2__DA3

Bokonyi Gabor; Illes Gabor; Radics Gabor

26/11/2017

Problem #2.1

1. Data acquisition and cleaning

Download cross-county ‘GDP per capita’ data

```
GDP <- wb(indicator = "NY.GDP.PCAP.PP.KD", startdate = 2011, enddate = 2011)
```

Change column names

```
setnames(GDP, 'value', 'GDP_per_capita')
setnames(GDP, 'country', 'country_name')
```

Download cross-county ‘Life Expectancy’ data

```
LEX <- wb(indicator = "SP.DYN.LE00.IN", startdate = 2011, enddate = 2011)
```

Change column names

```
setnames(LEX, 'value', 'life_expectancy')
```

Merge the two datasets by ‘iso2c’ code

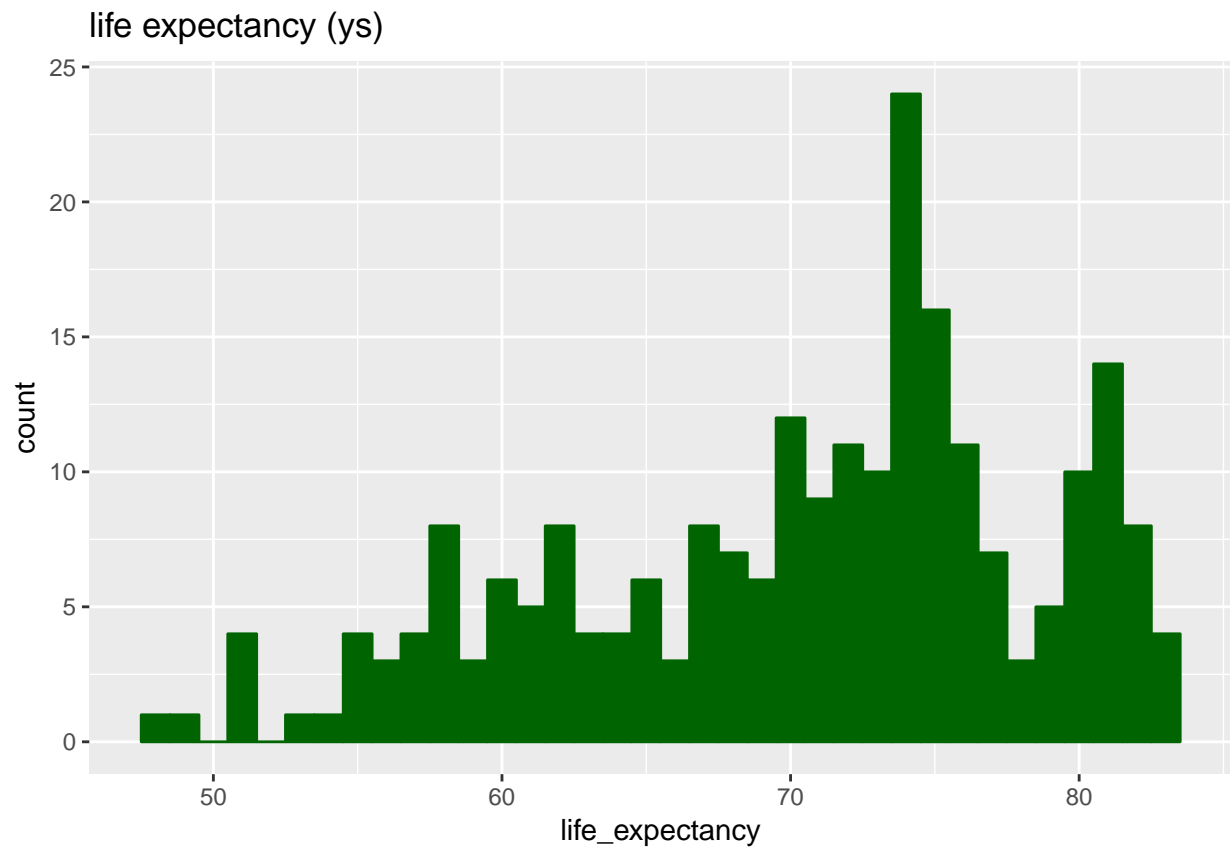
```
P21 <- data.table(merge(GDP[, c('iso2c', 'country_name', 'GDP_per_capita')], LEX[, c('iso2c', 'life_exp
```

Drop observations with missing values and write csv file

```
P21 <- P21[(GDP_per_capita != 'NA' & life_expectancy != 'NA'),]
write.csv(P21[, c('country_name', 'life_expectancy', 'GDP_per_capita')], 'P21.csv')
```

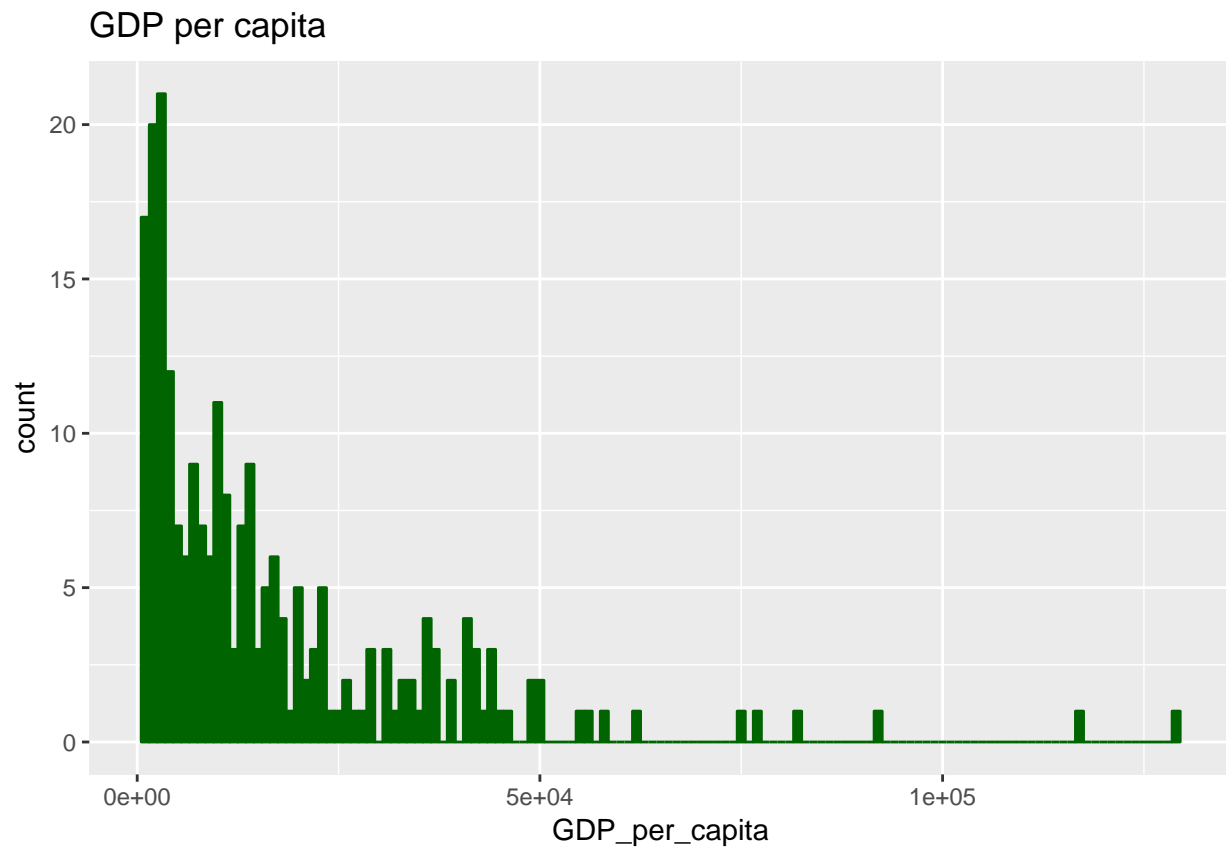
2. Estimate regressions

Histogram of ‘Life Expectancy’



some explanation here

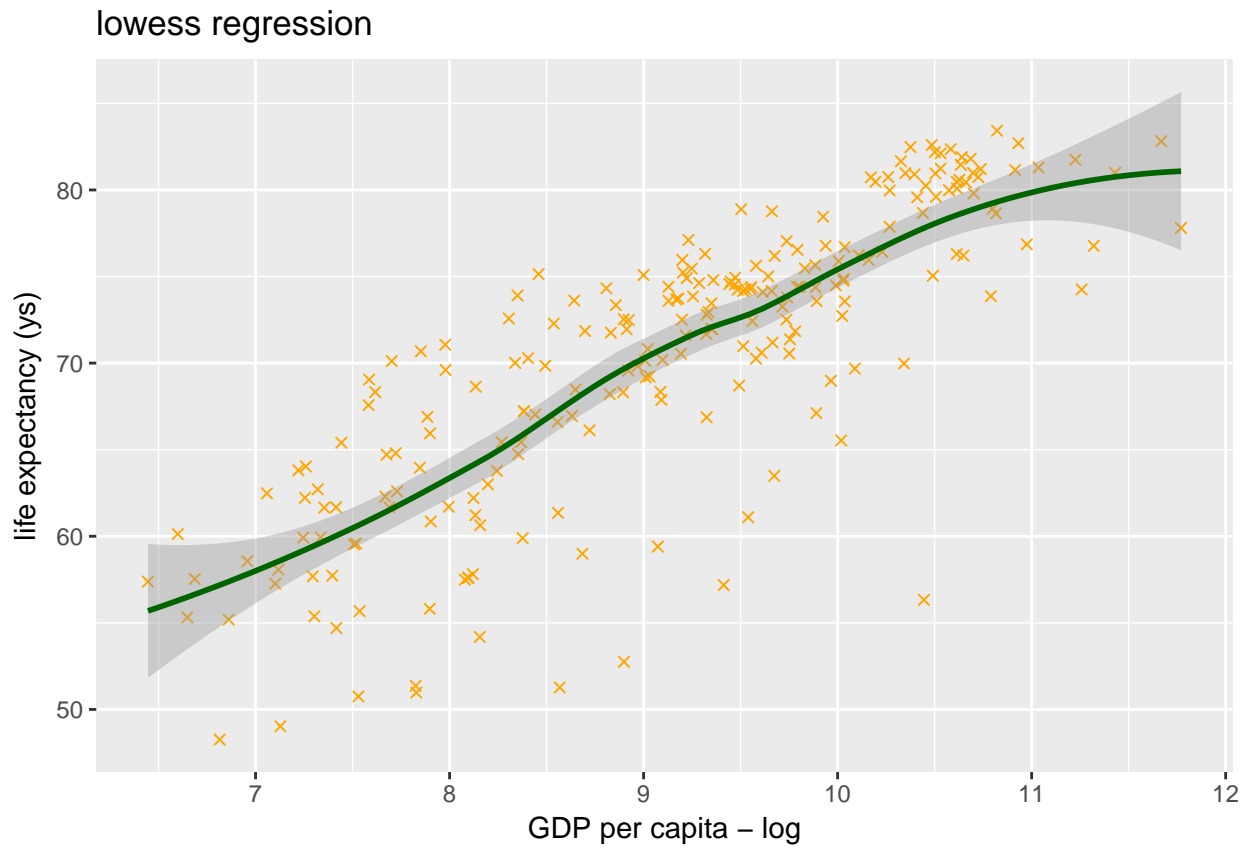
Histogram of 'GDP per capita'



some explanation here

Lowess regression of 'Life Expectancy' on ln 'Gdp per capita'

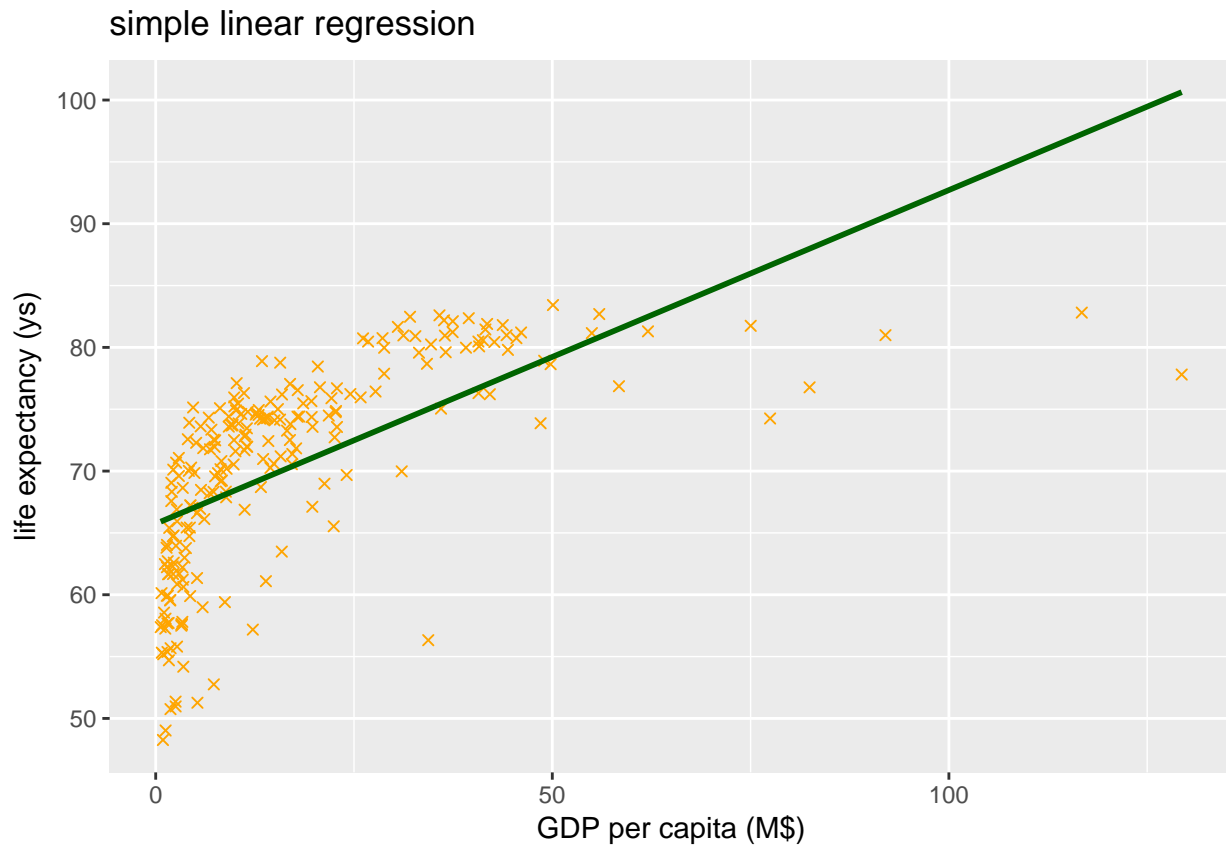
```
## `geom_smooth()` using method = 'loess'
```



some explanation here

Summary statistics of Lowess regression of 'Life Expectancy' on ln 'Gdp per capita'

Simple Linear Regression of 'Life Expectancy' on 'Gdp per capita'



some explanation here

Summary statistics of 'Life Expectancy' on 'Gdp per capita'

```
##
## Call:
## lm(formula = life_expectancy ~ GDP_per_capita_M, data = P21)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.834  -3.987   2.039   4.738   9.535
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    65.73825    0.55850   117.70  <2e-16 ***
## GDP_per_capita_M  0.26981    0.02172   12.42  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.362 on 229 degrees of freedom
## Multiple R-squared:  0.4025, Adjusted R-squared:  0.3999
## F-statistic: 154.3 on 1 and 229 DF, p-value: < 2.2e-16
```

some explanation here

Level-log linear regression of 'Life Expectancy' on 'Gdp per capita'



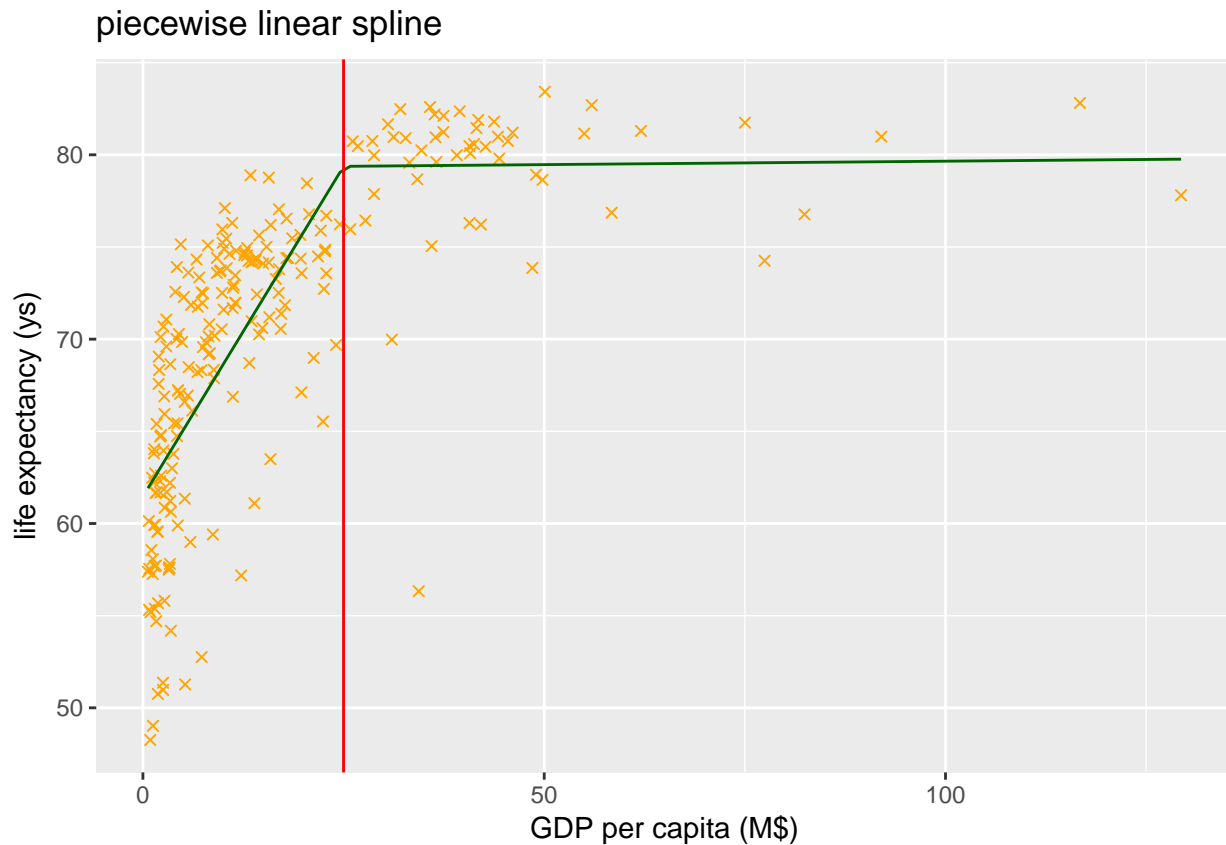
some explanation here

Summary statistics of level-log regression of ‘Life Expectancy’ on ‘Gdp per capita’

```
##
## Call:
## lm(formula = life_expectancy ~ GDP_per_capita_log, data = P21)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4554  -1.9496   0.8689   3.0881   8.7154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18.0699     2.4357   7.419 2.29e-12 ***
## GDP_per_capita_log  5.7169     0.2643  21.632 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.718 on 229 degrees of freedom
## Multiple R-squared:  0.6714, Adjusted R-squared:  0.67
## F-statistic:  468 on 1 and 229 DF, p-value: < 2.2e-16
```

some explanation here

Piecewise linear spline of ‘Life Expectancy’ on ‘Gdp per capita’ (lspline)



some explanation here

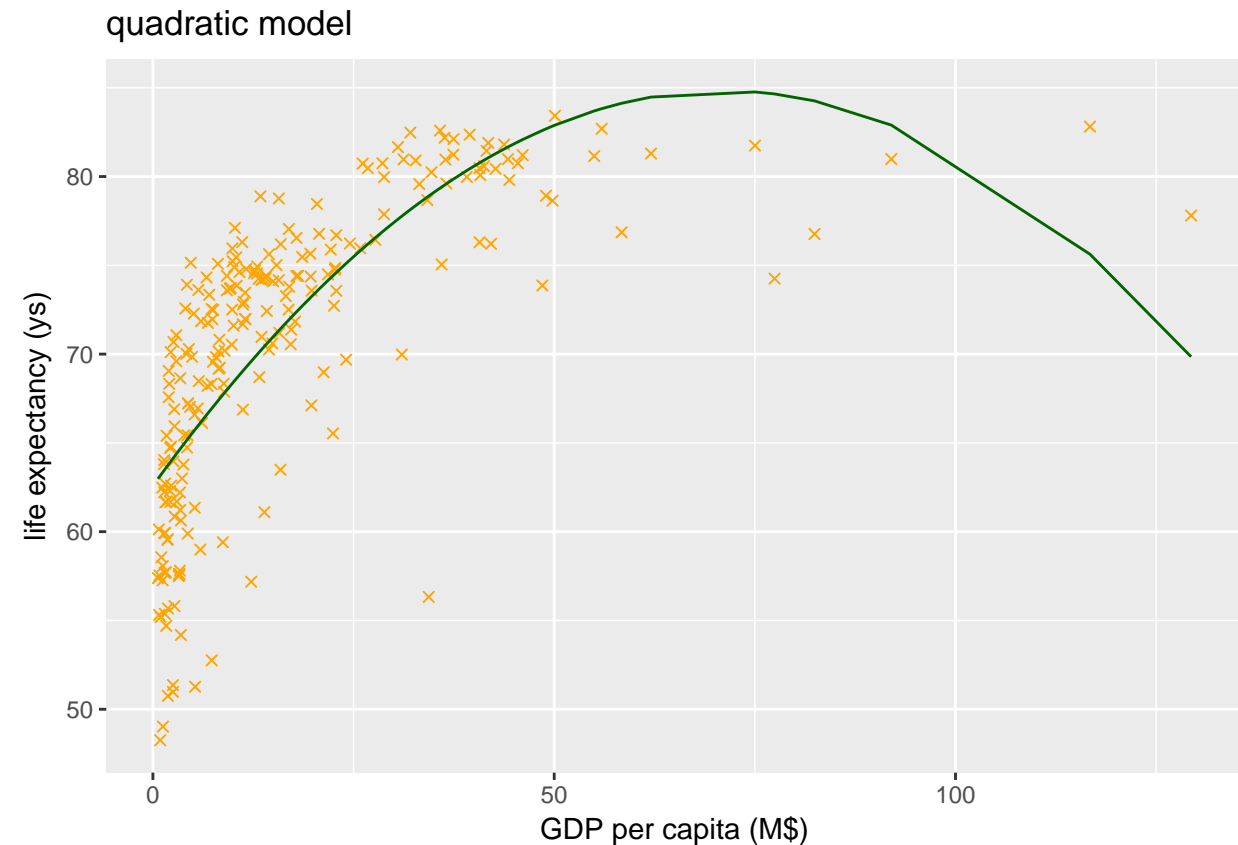
Summary statistics of linear spline of 'Life Expectancy' on 'Gdp per capita'

```
##
## Call:
## lm(formula = life_expectancy ~ lspline(GDP_per_capita_M, c(25)),
##     data = P21)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.084  -2.636   1.038   3.041  10.300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    61.464263    0.591897  103.843  <2e-16
## lspline(GDP_per_capita_M, c(25))1  0.716416    0.043775   16.366  <2e-16
## lspline(GDP_per_capita_M, c(25))2  0.003728    0.029635    0.126    0.9
##
## (Intercept) ***
## lspline(GDP_per_capita_M, c(25))1 ***
## lspline(GDP_per_capita_M, c(25))2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.132 on 228 degrees of freedom
## Multiple R-squared:  0.6129, Adjusted R-squared:  0.6095
```

```
## F-statistic: 180.5 on 2 and 228 DF, p-value: < 2.2e-16
```

some explanation here

Quadratic regression of 'Life Expectancy' on 'Gdp per capita'



some explanation here

Summary statistics of Quadratic regression of 'Life Expectancy' on 'Gdp per capita'

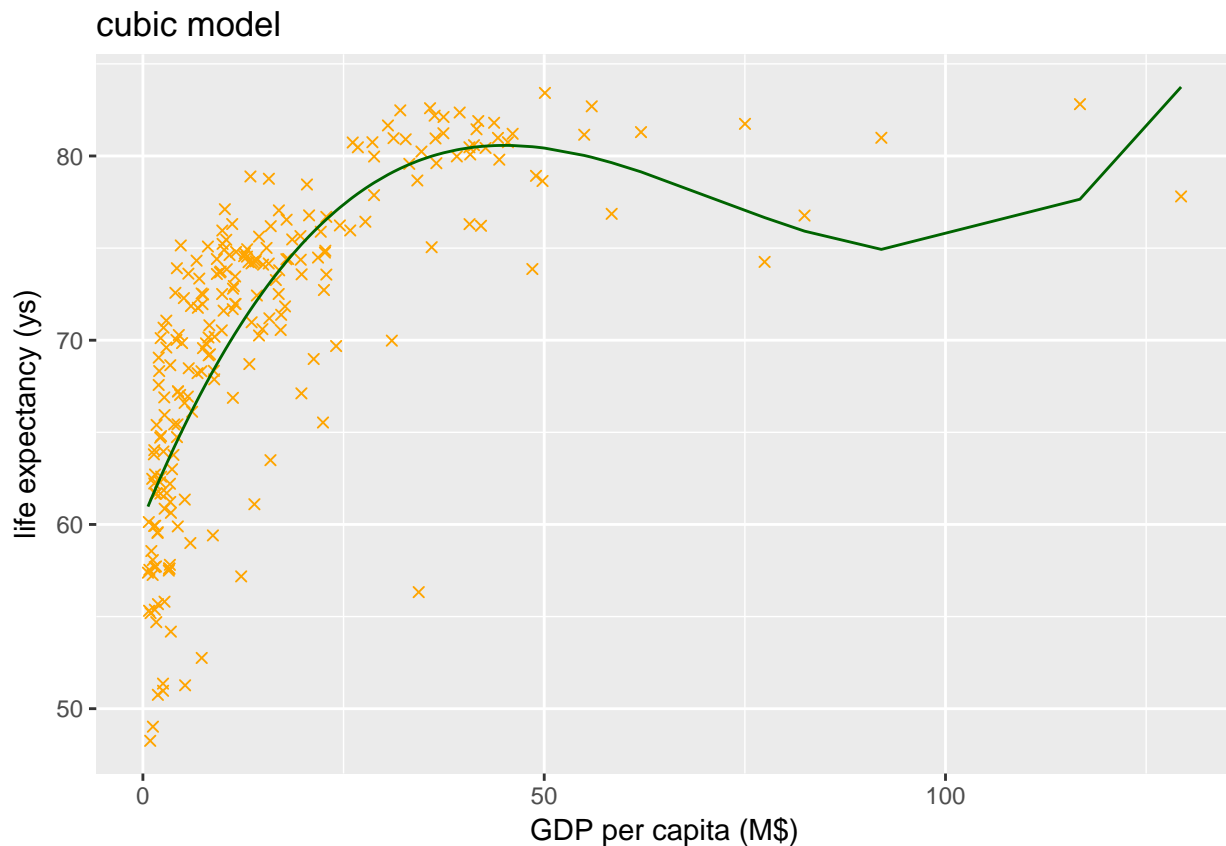
```
##
## Call:
## lm(formula = life_expectancy ~ GDP_per_capita_M + GDP_per_capita_M_sq,
##     data = P21)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5747  -2.8076   0.8192   3.8826   9.7030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    62.5844670   0.5710193  109.601  <2e-16 ***
## GDP_per_capita_M    0.6262626   0.0409124   15.307  <2e-16 ***
## GDP_per_capita_M_sq -0.0044072   0.0004525   -9.741  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.358 on 228 degrees of freedom
```



```
## Multiple R-squared:  0.5781, Adjusted R-squared:  0.5744
## F-statistic: 156.2 on 2 and 228 DF,  p-value: < 2.2e-16
```

some explanation here

Cubic regression of 'Life Expectancy' on 'Gdp per capita'



some explanation here

Summary statistics of Cubic regression of 'Life Expectancy' on 'Gdp per capita'

```
##
## Call:
## lm(formula = life_expectancy ~ GDP_per_capita_M + GDP_per_capita_M_sq +
##     GDP_per_capita_M_cub, data = P21)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.4213  -2.1337   0.8573   2.8635  10.1976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.031e+01  6.495e-01  92.845  < 2e-16 ***
## GDP_per_capita_M    1.063e+00  8.127e-02  13.079  < 2e-16 ***
## GDP_per_capita_M_sq -1.724e-02  2.152e-03  -8.011 5.94e-14 ***
## GDP_per_capita_M_cub  8.057e-05  1.325e-05   6.080 5.04e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.98 on 227 degrees of freedom
## Multiple R-squared:  0.6372, Adjusted R-squared:  0.6324
## F-statistic: 132.9 on 3 and 227 DF,  p-value: < 2.2e-16
```

some explanation here

3. Estimate a weighted regression

weighted level-log linear regression



some explanation here

```
##
## Call:
## lm(formula = life_expectancy ~ GDP_per_capita_log, data = PS21w,
##     weights = population_MM)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -193.942   -6.073    1.178   10.707  158.977
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.8855     2.0402   3.865 0.000145 ***
## GDP_per_capita_log  6.8373     0.2235  30.591 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  
## Residual standard error: 42.85 on 229 degrees of freedom  
## Multiple R-squared:  0.8034, Adjusted R-squared:  0.8025  
## F-statistic: 935.8 on 1 and 229 DF,  p-value: < 2.2e-16
```

some explanation here