Salary and Working Hours

Aleksandra Villepique

February 16, 2016

Yesterday I saw on the Facebook interesting bar graph that illustrated averaged annual working hours across countries included in OECD data collection. What intrigued me was that countries I knew to be poorer had more working hours than the richer countries.

Just to satisfy my curiosity, I decided to head to OECD site, collect some data and do the analysis myself, just to see is there a correlation between working hours and income.

I'm not really experienced with the OECD database so I could not find there median income per country, so I decided to use minimum income per country because average income is skewed towards larger values due to the huge income gap.

```
data=read.csv("RMW_16022016200632658.csv")
#looking at the data:
head(data)
```

```
##   COUNTRY Country SERIES                                          Series PERIOD
## 1     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
## 2     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
## 3     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
## 4     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
## 5     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
## 6     IRL Ireland    PPP In 2014 constant prices at 2014 USD PPPs      A
##   Pay.period TIME Time Unit.Code      Unit PowerCode.Code PowerCode
## 1     Annual 2000 2000       USD US Dollar              0     units
## 2     Annual 2001 2001       USD US Dollar              0     units
## 3     Annual 2002 2002       USD US Dollar              0     units
## 4     Annual 2003 2003       USD US Dollar              0     units
## 5     Annual 2004 2004       USD US Dollar              0     units
## 6     Annual 2005 2005       USD US Dollar              0     units
##   Reference.Period.Code Reference.Period Value Flag.Codes Flags
## 1                    NA               NA 18549         NA    NA
## 2                    NA               NA 18288         NA    NA
## 3                    NA               NA 18354         NA    NA
## 4                    NA               NA 18553         NA    NA
## 5                    NA               NA 19869         NA    NA
## 6                    NA               NA 20742         NA    NA
```

```
summary(data)
```

```
##     COUNTRY                Country     SERIES
##  AUS    :  60   Australia     :  60   EXR:750
##  BEL    :  60   Belgium       :  60   PPP:750
##  CAN    :  60   Canada        :  60
##  CHL    :  60   Chile         :  60
##  CZE    :  60   Czech Republic:  60
##  ESP    :  60   Estonia       :  60
##  (Other):1140   (Other)       :1140
##                                                  Series    PERIOD
##  In 2014 constant prices at 2014 USD exchange rates:750   A:750
##  In 2014 constant prices at 2014 USD PPPs          :750   H:750
##
##
##
##
##
##   Pay.period       TIME           Time      Unit.Code          Unit
##  Annual:750   Min.   :2000   Min.   :2000   USD:1500   US Dollar:1500
##  Hourly:750   1st Qu.:2003   1st Qu.:2003
##               Median :2007   Median :2007
##               Mean   :2007   Mean   :2007
##               3rd Qu.:2011   3rd Qu.:2011
##               Max.   :2014   Max.   :2014
##
##  PowerCode.Code PowerCode    Reference.Period.Code Reference.Period
##  Min.   :0      units:1500   Mode:logical          Mode:logical
##  1st Qu.:0                   NA's:1500             NA's:1500
##  Median :0
##  Mean   :0
##  3rd Qu.:0
##  Max.   :0
##
##      Value           Flag.Codes      Flags
##  Min.   :    0.610   Mode:logical   Mode:logical
##  1st Qu.:    5.788   NA's:1500      NA's:1500
##  Median :  637.490
##  Mean   : 6755.276
##  3rd Qu.:12877.750
##  Max.   :30615.000
##
```

```
tail(data)
```

```
##      COUNTRY  Country SERIES
## 1495     PRT Portugal    EXR
## 1496     PRT Portugal    EXR
## 1497     PRT Portugal    EXR
## 1498     PRT Portugal    EXR
## 1499     PRT Portugal    EXR
## 1500     PRT Portugal    EXR
##                                                    Series PERIOD Pay.period
## 1495 In 2014 constant prices at 2014 USD exchange rates      A     Annual
## 1496 In 2014 constant prices at 2014 USD exchange rates      A     Annual
## 1497 In 2014 constant prices at 2014 USD exchange rates      A     Annual
## 1498 In 2014 constant prices at 2014 USD exchange rates      A     Annual
## 1499 In 2014 constant prices at 2014 USD exchange rates      A     Annual
## 1500 In 2014 constant prices at 2014 USD exchange rates      A     Annual
##      TIME Time Unit.Code      Unit PowerCode.Code PowerCode
## 1495 2009 2009       USD US Dollar              0     units
## 1496 2010 2010       USD US Dollar              0     units
## 1497 2011 2011       USD US Dollar              0     units
## 1498 2012 2012       USD US Dollar              0     units
## 1499 2013 2013       USD US Dollar              0     units
## 1500 2014 2014       USD US Dollar              0     units
##      Reference.Period.Code Reference.Period Value Flag.Codes Flags
## 1495                    NA                NA  9028         NA    NA
## 1496                    NA                NA  9398         NA    NA
## 1497                    NA                NA  9258         NA    NA
## 1498                    NA                NA  9008         NA    NA
## 1499                    NA                NA  8983         NA    NA
## 1500                    NA                NA  9101         NA    NA
```

```
str(data)
```

```
## 'data.frame':    1500 obs. of  17 variables:
##  $ COUNTRY              : Factor w/ 25 levels "AUS","BEL","CAN",..: 12 12 12 12 12 12 12
12 12 12 ...
##  $ Country              : Factor w/ 25 levels "Australia","Belgium",..: 10 10 10 10 10 1
0 10 10 10 10 ...
##  $ SERIES               : Factor w/ 2 levels "EXR","PPP": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Series               : Factor w/ 2 levels "In 2014 constant prices at 2014 USD exchan
ge rates",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ PERIOD               : Factor w/ 2 levels "A","H": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Pay.period           : Factor w/ 2 levels "Annual","Hourly": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TIME                 : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
##  $ Time                 : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
##  $ Unit.Code            : Factor w/ 1 level "USD": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Unit                 : Factor w/ 1 level "US Dollar": 1 1 1 1 1 1 1 1 1 1 ...
##  $ PowerCode.Code       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PowerCode            : Factor w/ 1 level "units": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Reference.Period.Code: logi  NA NA NA NA NA NA ...
##  $ Reference.Period     : logi  NA NA NA NA NA NA ...
##  $ Value                : num  18549 18288 18354 18553 19869 ...
##  $ Flag.Codes           : logi  NA NA NA NA NA NA ...
##  $ Flags                : logi  NA NA NA NA NA NA ...
```

```
#getting data from 2010- present
test <- subset(data, data$TIME >= 2010)
myVar<-c("COUNTRY", "Country", "TIME", "Time", "Value")
minSalary <- test[myVar]
```

Uploading working hours:

```
data=read.csv("ANHRS_16022016200952408.csv")
#looking at the data:
head(data)
```

```
##   COUNTRY   Country EMPSTAT Employment.status FREQUENCY Frequency TIME
## 1     AUS Australia      TE  Total employment         A    Annual 2000
## 2     AUS Australia      TE  Total employment         A    Annual 2001
## 3     AUS Australia      TE  Total employment         A    Annual 2002
## 4     AUS Australia      TE  Total employment         A    Annual 2003
## 5     AUS Australia      TE  Total employment         A    Annual 2004
## 6     AUS Australia      TE  Total employment         A    Annual 2005
##   Time Unit.Code  Unit PowerCode.Code PowerCode Reference.Period.Code
## 1 2000      HOUR Hours              0     units                    NA
## 2 2001      HOUR Hours              0     units                    NA
## 3 2002      HOUR Hours              0     units                    NA
## 4 2003      HOUR Hours              0     units                    NA
## 5 2004      HOUR Hours              0     units                    NA
## 6 2005      HOUR Hours              0     units                    NA
##   Reference.Period  Value Flag.Codes Flags
## 1               NA 1779.2         NA    NA
## 2               NA 1735.5         NA    NA
## 3               NA 1731.3         NA    NA
## 4               NA 1735.9         NA    NA
## 5               NA 1735.5         NA    NA
## 6               NA 1729.7         NA    NA
```

```
summary(data)
```

```
##     COUNTRY                Country    EMPSTAT              Employment.status
##  AUT    : 30   Austria        : 30   DE:463   Dependent employment:463
##  CAN    : 30   Canada         : 30   TE:582   Total employment    :582
##  CHL    : 30   Chile          : 30
##  CZE    : 30   Czech Republic: 30
##  DEU    : 30   Finland        : 30
##  ESP    : 30   France         : 30
##  (Other):865   (Other)        :865
##  FREQUENCY  Frequency          TIME           Time      Unit.Code
##  A:1045    Annual:1045   Min.   :2000   Min.   :2000   HOUR:1045
##                          1st Qu.:2003   1st Qu.:2003
##                          Median :2007   Median :2007
##                          Mean   :2007   Mean   :2007
##                          3rd Qu.:2011   3rd Qu.:2011
##                          Max.   :2014   Max.   :2014
##
##     Unit       PowerCode.Code PowerCode    Reference.Period.Code
##  Hours:1045   Min.   :0      units:1045   Mode:logical
##               1st Qu.:0                   NA's:1045
##               Median :0
##               Mean   :0
##               3rd Qu.:0
##               Max.   :0
##
##  Reference.Period     Value      Flag.Codes      Flags
##  Mode:logical     Min.   :1278   Mode:logical   Mode:logical
##  NA's:1045        1st Qu.:1584   NA's:1045      NA's:1045
##                   Median :1760
##                   Mean   :1758
##                   3rd Qu.:1912
##                   Max.   :2512
##
```

```
tail(data)
```

```
##      COUNTRY    Country EMPSTAT    Employment.status FREQUENCY Frequency
## 1040     CRI Costa Rica      DE Dependent employment         A    Annual
## 1041     CRI Costa Rica      TE     Total employment         A    Annual
## 1042     CRI Costa Rica      TE     Total employment         A    Annual
## 1043     CRI Costa Rica      TE     Total employment         A    Annual
## 1044     CRI Costa Rica      TE     Total employment         A    Annual
## 1045     CRI Costa Rica      TE     Total employment         A    Annual
##      TIME Time Unit.Code  Unit PowerCode.Code PowerCode
## 1040 2014 2014      HOUR Hours              0     units
## 1041 2010 2010      HOUR Hours              0     units
## 1042 2011 2011      HOUR Hours              0     units
## 1043 2012 2012      HOUR Hours              0     units
## 1044 2013 2013      HOUR Hours              0     units
## 1045 2014 2014      HOUR Hours              0     units
##      Reference.Period.Code Reference.Period Value Flag.Codes Flags
## 1040                    NA               NA  2297         NA    NA
## 1041                    NA               NA  2321         NA    NA
## 1042                    NA               NA  2354         NA    NA
## 1043                    NA               NA  2287         NA    NA
## 1044                    NA               NA  2223         NA    NA
## 1045                    NA               NA  2216         NA    NA
```

```
str(data)
```

```
## 'data.frame':    1045 obs. of  17 variables:
##  $ COUNTRY              : Factor w/ 40 levels "AUS","AUT","BEL",..: 1 1 1 1 1 1 1 1 1 1
...
##  $ Country              : Factor w/ 40 levels "Australia","Austria",..: 1 1 1 1 1 1 1 1
1 1 ...
##  $ EMPSTAT              : Factor w/ 2 levels "DE","TE": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Employment.status    : Factor w/ 2 levels "Dependent employment",..: 2 2 2 2 2 2 2 2
2 2 ...
##  $ FREQUENCY            : Factor w/ 1 level "A": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Frequency            : Factor w/ 1 level "Annual": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TIME                 : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
##  $ Time                 : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
##  $ Unit.Code            : Factor w/ 1 level "HOUR": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Unit                 : Factor w/ 1 level "Hours": 1 1 1 1 1 1 1 1 1 1 ...
##  $ PowerCode.Code       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PowerCode            : Factor w/ 1 level "units": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Reference.Period.Code: logi  NA NA NA NA NA NA ...
##  $ Reference.Period     : logi  NA NA NA NA NA NA ...
##  $ Value                : num  1779 1736 1731 1736 1736 ...
##  $ Flag.Codes           : logi  NA NA NA NA NA NA ...
##  $ Flags                : logi  NA NA NA NA NA NA ...
```

```
#getting data from 2010- present
test <- subset(data, data$TIME >= 2010)
myVar<-c("COUNTRY", "Country", "TIME", "Time", "Value")
workHours <- test[myVar]
```

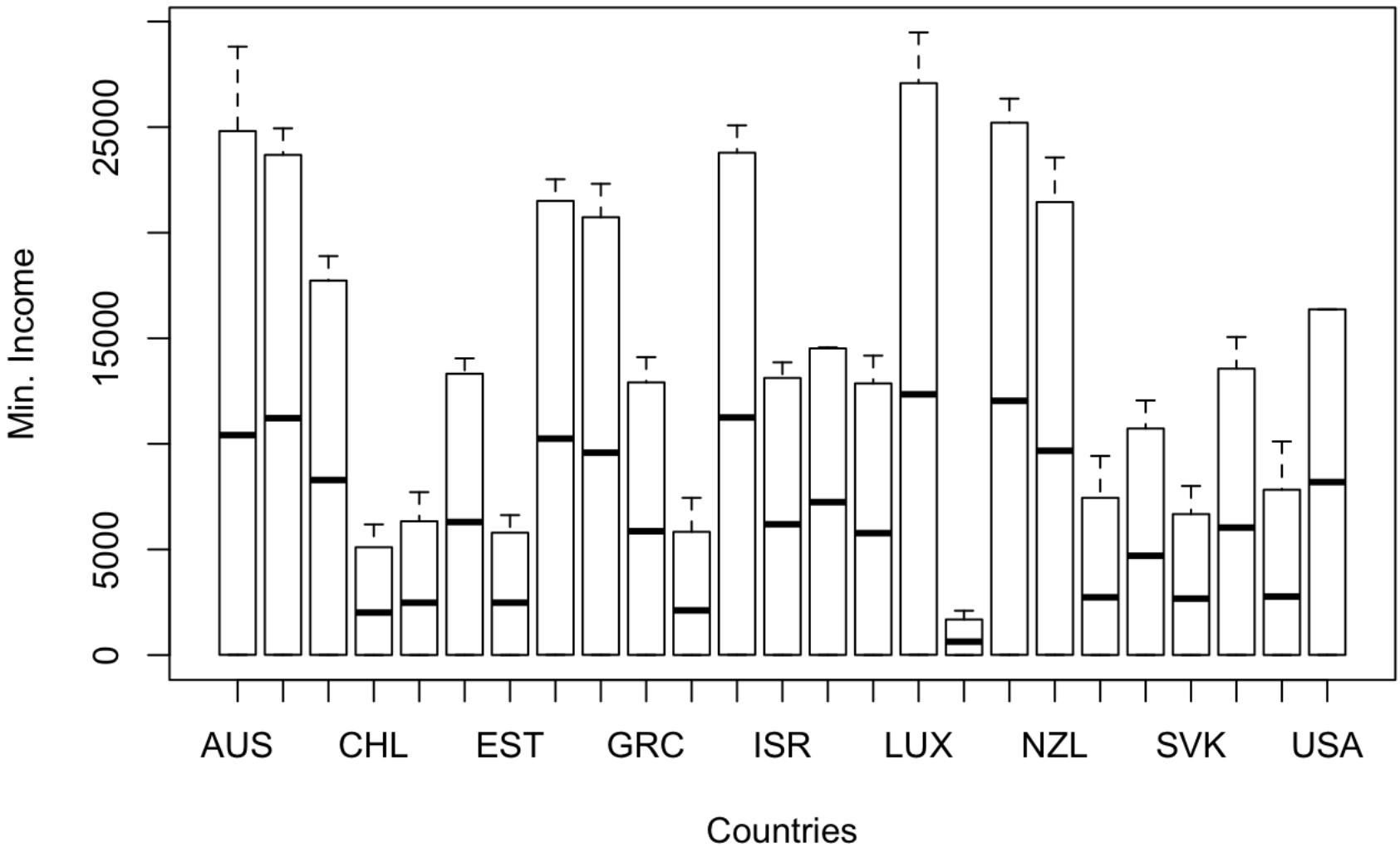
I picked only years 2010, 2011, 2012, 2013, and 2014. This period is similar, shows certain recovery in the economy. Yeah, I know we're now in the front another bubble, but that is not subject to this little analysis.

Another reason why I decided to pick several years is to make sure that results from one year are not some kind of fluke.

So here are the results for 2010:

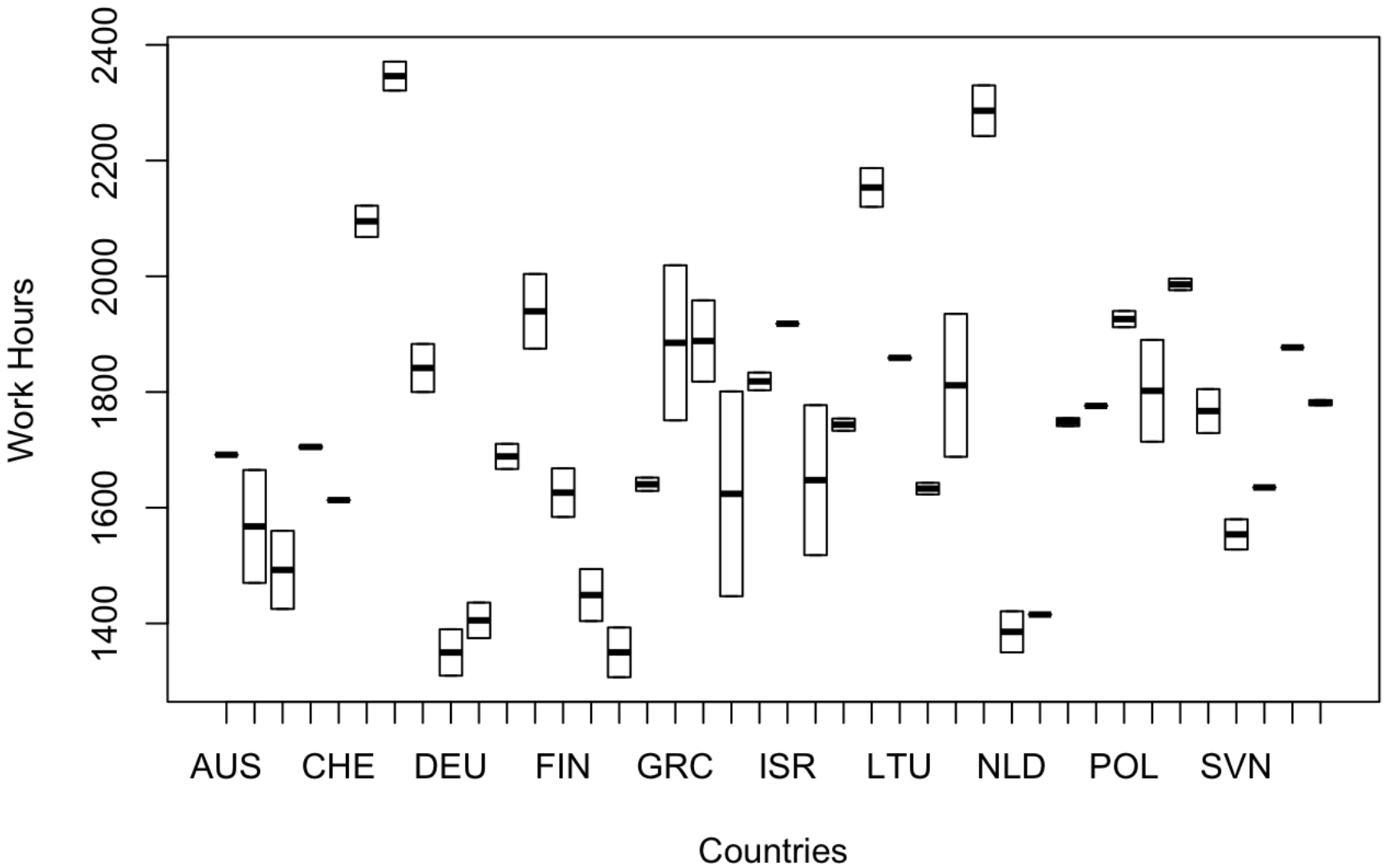
```
MS2010<-subset(minSalary, minSalary$TIME==2010)
```

Which looks like:




```
WH2010<-subset(workHours, workHours$TIME==2010)
```

Which looks like:



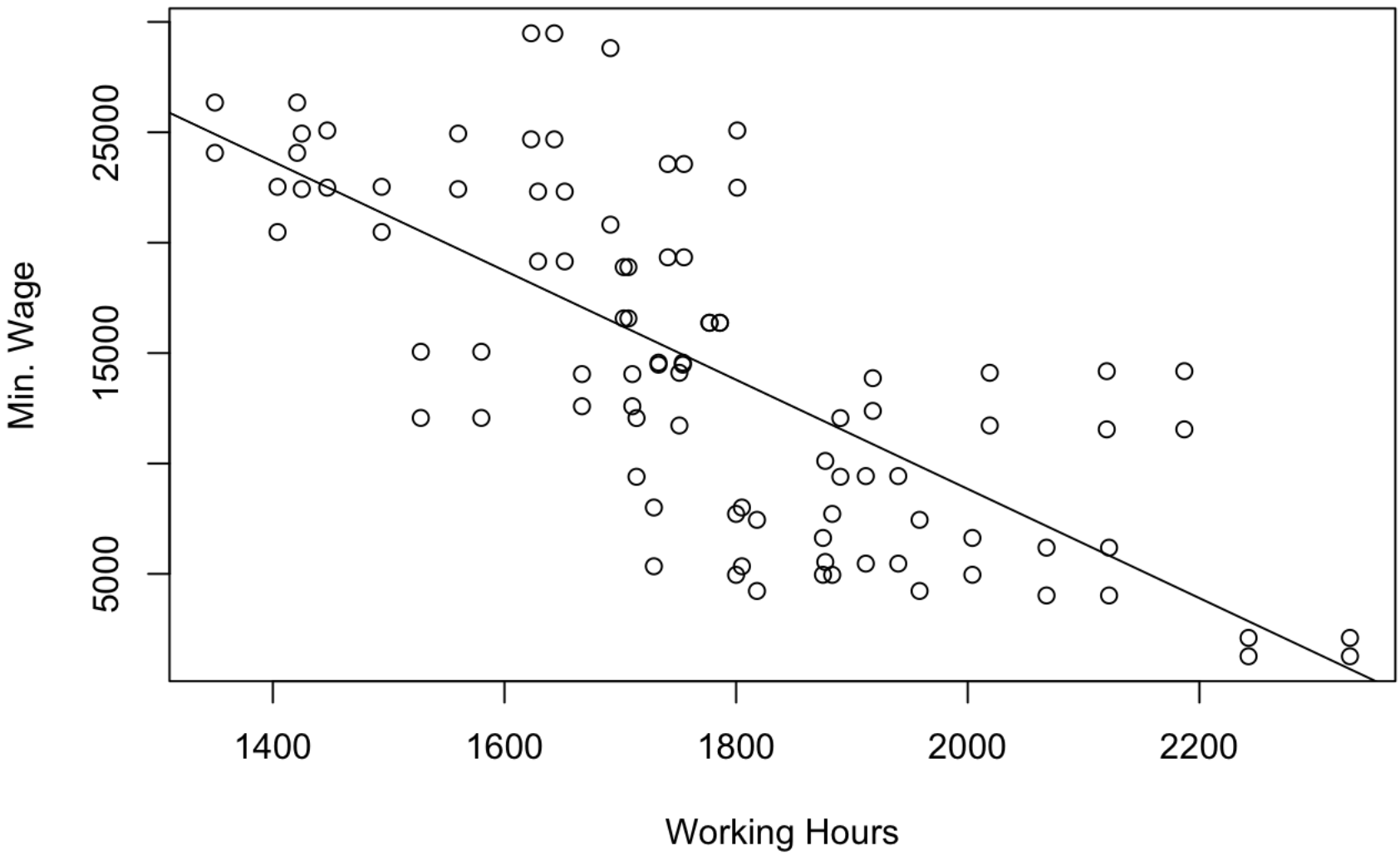
But now, how do they correlate?

The data sets were not equal, one had more countries than other, so I had to tweak the data frames before merging them. This gave me 25 different countries.

```
test<-unique(MS2010$COUNTRY)
test2<-subset(WH2010, WH2010$COUNTRY %in% test)
test3 <- merge(test2, MS2010, by.x="COUNTRY", by.y="COUNTRY")
d2010<- subset(test3, test3$Value.y >=20)
```

And graph with linear fit is:

Which looks like:



And statistical values of linear regression are:

```
d2010.lm<-lm(d2010$Value.y~d2010$Value.x)
summary(d2010.lm)$r.squared
```

```
## [1] 0.5316456
```

```
summary(d2010.lm)
```

```
##
## Call:
## lm(formula = d2010$Value.y ~ d2010$Value.x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10205.9  -3697.8   -548.8   2775.4  12333.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   58302.26    4334.16   13.45   <2e-16 ***
## d2010$Value.x   -24.73       2.42  -10.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5190 on 92 degrees of freedom
## Multiple R-squared:  0.5316, Adjusted R-squared:  0.5266
## F-statistic: 104.4 on 1 and 92 DF,  p-value: < 2.2e-16
```

```
cor(d2010$Value.y,d2010$Value.x)
```

```
## [1] -0.7291403
```

I was quite pleased to see that correlation is negative, as I suspected. The value of -0.73 shows quite good a negative relation between variables.

Meaning, more work does not bring more money to a human. How much money individual earns depends on other factors that form labor situation in the country. The factors I did not consider in this little analysis.

R-squared value is also quite decent; 0.53 shows that linear fit is a good approximation. And of course, the p-value is 2.2 e-16, adding another confirmation of the significance of this result.

The rest of the analyzed years gave similar results.

2011:

```
WH2011<-subset(workHours, workHours$TIME==2011)
MS2011<-subset(minSalary, minSalary$TIME==2011)

test<-unique(MS2011$COUNTRY)
test2<-subset(WH2011, WH2011$COUNTRY %in% test)
test3 <- merge(test2, MS2011, by.x="COUNTRY", by.y="COUNTRY")
d2011<- subset(test3, test3$Value.y >=20)

d2011.lm<-lm(d2011$Value.y~d2011$Value.x)
summary(d2011.lm)$r.squared
```

```
## [1] 0.5464418
```

```
summary(d2011.lm)
```

```
##
## Call:
## lm(formula = d2011$Value.y ~ d2011$Value.x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9953.4 -3481.2  -350.2  3025.9 12930.7
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57748.330   4163.915   13.87   <2e-16 ***
## d2011$Value.x   -24.508      2.328  -10.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5052 on 92 degrees of freedom
## Multiple R-squared:  0.5464, Adjusted R-squared:  0.5415
## F-statistic: 110.8 on 1 and 92 DF,  p-value: < 2.2e-16
```

```
cor(d2011$Value.y,d2011$Value.x)
```

```
## [1] -0.739217
```

2012:

```
MS2012<-subset(minSalary, minSalary$TIME==2012)
WH2012<-subset(workHours, workHours$TIME==2012)

test<-unique(MS2012$COUNTRY)
test2<-subset(WH2012, WH2012$COUNTRY %in% test)
test3 <- merge(test2, MS2012, by.x="COUNTRY", by.y="COUNTRY")
d2012<- subset(test3, test3$Value.y >=20)
d2012.lm<-lm(d2012$Value.y~d2012$Value.x)
summary(d2012.lm)$r.squared
```

```
## [1] 0.52222
```

```
summary(d2012.lm)
```

```
##
## Call:
## lm(formula = d2012$Value.y ~ d2012$Value.x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9500.3 -4312.8    31.1  3254.7 13077.8
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57777.239   4381.109   13.19   <2e-16 ***
## d2012$Value.x   -24.684      2.462  -10.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5200 on 92 degrees of freedom
## Multiple R-squared:  0.5222, Adjusted R-squared:  0.517
## F-statistic: 100.6 on 1 and 92 DF,  p-value: < 2.2e-16
```

```
cor(d2012$Value.y,d2012$Value.x)
```

```
## [1] -0.7226479
```

2013:

```
MS2013<-subset(minSalary, minSalary$TIME==2013)
WH2013<-subset(workHours, workHours$TIME==2013)

test<-unique(MS2013$COUNTRY)
test2<-subset(WH2013, WH2013$COUNTRY %in% test)
test3 <- merge(test2, MS2013, by.x="COUNTRY", by.y="COUNTRY")
d2013<- subset(test3, test3$Value.y >=20)

d2013.lm<-lm(d2013$Value.y~d2013$Value.x)
summary(d2013.lm)$r.squared
```

```
## [1] 0.4982009
```

```
summary(d2013.lm)
```

```
##
## Call:
## lm(formula = d2013$Value.y ~ d2013$Value.x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9570.3 -4258.8   116.9  2804.5 12796.8
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57662.816   4571.638  12.613  < 2e-16 ***
## d2013$Value.x   -24.608      2.575  -9.557 1.95e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5299 on 92 degrees of freedom
## Multiple R-squared:  0.4982, Adjusted R-squared:  0.4927
## F-statistic: 91.34 on 1 and 92 DF,  p-value: 1.95e-15
```

```
cor(d2013$Value.y,d2013$Value.x)
```

```
## [1] -0.7058335
```

and 2014:

```
MS2014<-subset(minSalary, minSalary$TIME==2014)
WH2014<-subset(workHours, workHours$TIME==2014)

test<-unique(MS2014$COUNTRY)
test2<-subset(WH2014, WH2014$COUNTRY %in% test)
test3 <- merge(test2, MS2014, by.x="COUNTRY", by.y="COUNTRY")
d2014<- subset(test3, test3$Value.y >=20)
d2014.lm<-lm(d2014$Value.y~d2014$Value.x)
summary(d2014.lm)$r.squared
```

```
## [1] 0.4910365
```

```
summary(d2014.lm)
```

```
##
## Call:
## lm(formula = d2014$Value.y ~ d2014$Value.x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9852.0 -4248.4  -122.2  2433.7 13027.2
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   58315.737   4855.145  12.011  < 2e-16 ***
## d2014$Value.x   -24.789      2.721  -9.109 2.95e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5376 on 86 degrees of freedom
## Multiple R-squared:  0.491,  Adjusted R-squared:  0.4851
## F-statistic: 82.97 on 1 and 86 DF,  p-value: 2.954e-14
```

```
cor(d2014$Value.y,d2014$Value.x)
```

```
## [1] -0.70074
```

All of them had similar values for correlation, a coefficient of determination and p-value. I do not know much about the economy to conclude why there is this sad result. The only thing I can see is that it does not matter how much someone works, but in which country they live.