HW5 2

Autumn Li and UNI ql2280

Oct 31, 2016

i. We estimate a by minimizing

```
library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.0/Resources/library")
wtid.report <- read.csv("~/Desktop/wtid-report.csv", header=FALSE)
year = wtid.report$V2
P99 = wtid.report$V3
P99.5 = wtid.report$V4
P99.9 = wtid.report$V5
a=2
percentile_ratio_discrepancies =  function (P99,P99.5,P99.9,a=2){
  b = ((as.numeric(P99)/as.numeric(P99.9))^(1-a)-10)^2+
    ((as.numeric(P99.5)/as.numeric(P99.9))^(1-a)-5)^2+
    ((as.numeric(P99)/as.numeric(P99.5))^(1-a)-2)^2
  return(b)
}
percentile_ratio_discrepancies(P99=1e6, P99.5=2e6,P99.9=1e7,a)
```

```
## [1] 0
```

ii. Write a function, exponent.multi ratios est

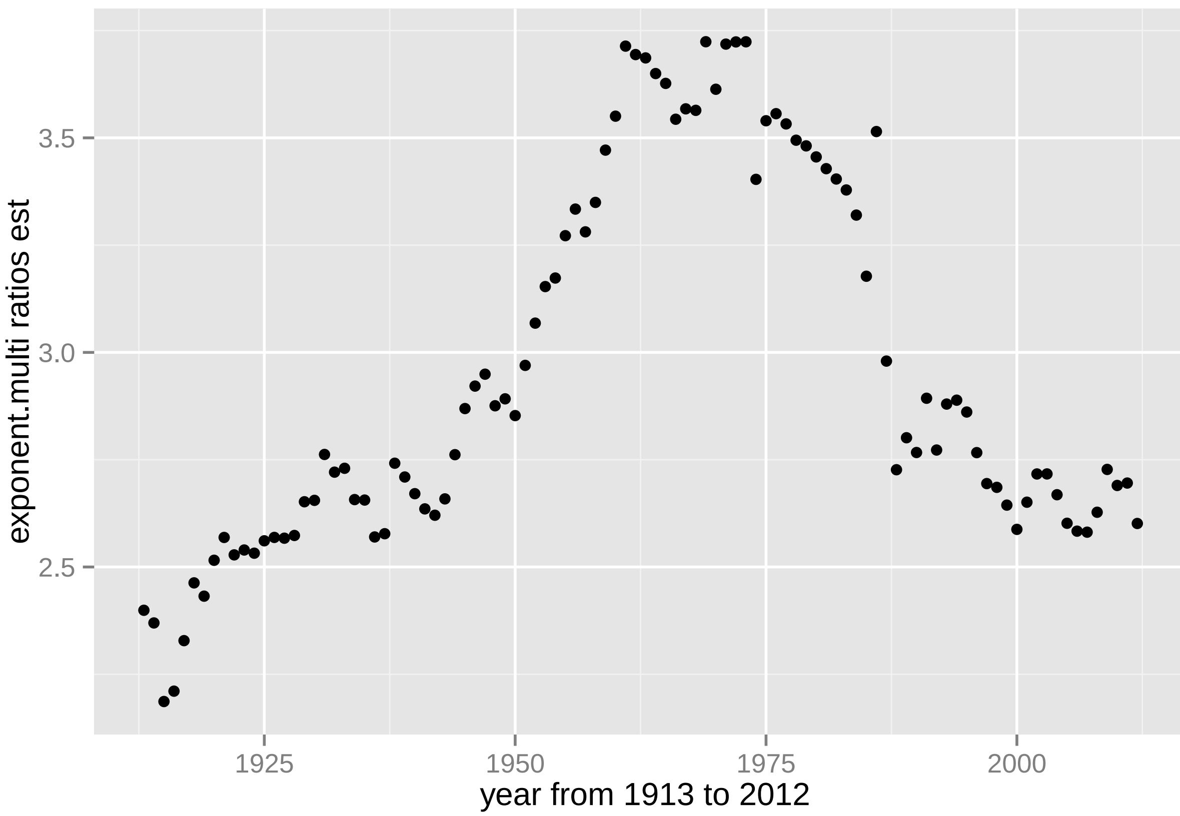
```
P99 = wtid.report$V3
P99.5 = wtid.report$V4
P99.9 = wtid.report$V5
exponent.multi_ratios_est = function (P99,P99.5,P99.9) {
  starting  = 1-log(10)/log(P99/P99.9)
  c = nlm(percentile_ratio_discrepancies, p=starting,P99 = P99,P99.5 =P99.5,P99.9 = P99.9)$estimate
  return(c)
}
exponent.multi_ratios_est(P99=1e6, P99.5=2e6, P99.9=1e7)
```

```
## [1] 2
```

iii. Write a function which uses exponent.multi ratios est to estimate a for the US for every year from 1913 to 2012.

```
year = wtid.report$V2
P99 = wtid.report$V3
P99.5 = wtid.report$V4
P99.9 = wtid.report$V5
us.estimate = matrix(NA,ncol = 100, nrow = 1)
for (i in 1913:2012){
  new.P99 = P99[year == i]
  new.P99 = as.numeric(levels(new.P99))[new.P99]
  new.P99.5 = P99.5[year == i]
```

```
  new.P99.5 = as.numeric(levels(new.P99.5))[new.P99.5]
  new.P99.9 = P99.9[year == i]
  new.P99.9 = as.numeric(levels(new.P99.9))[new.P99.9]
  d = exponent.multi_ratios_est(new.P99, new.P99.5, new.P99.9)
  us.estimate[i-1912] = d
  }
   us.estimate = as.numeric(us.estimate)
year1.3 = c(1913:2012)
  p1.3 = cbind(us.estimate,year1.3)
  p1.3 = data.frame(p1.3)
ggplot(data = p1.3)+ geom_point(mapping=aes(x=year1.3,y = us.estimate )) +
  labs(x = " year from 1913 to 2012", y = "exponent.multi ratios est")
```



iv.

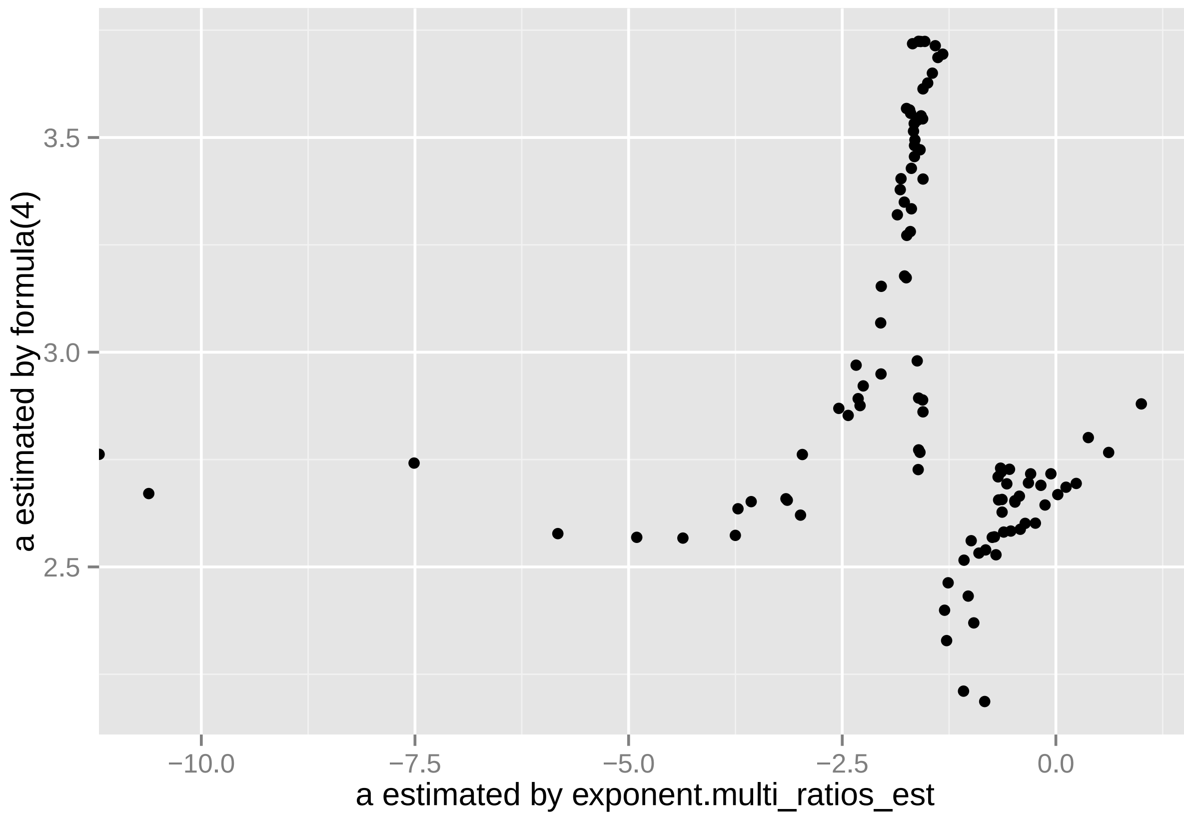
Use (4) to estimate a for the US for every year.

```
ab = 1-log(10)/(log(as.numeric(P99))/log(as.numeric(P99.9)))
length(ab)
```

```
## [1] 104
```

```
year = wtid.report$V2
P99 = wtid.report$V3
P99.5 = wtid.report$V4
P99.9 = wtid.report$V5
us.estimate = matrix(NA,ncol = 104, nrow = 1)
for (i in 1913:2015){
   new.P99 = P99[year == i]
```

```
  new.P99 = as.numeric(levels(new.P99))[new.P99]
  new.P99.5 = P99.5[year == i]
  new.P99.5 = as.numeric(levels(new.P99.5))[new.P99.5]
  new.P99.9 = P99.9[year == i]
  new.P99.9 = as.numeric(levels(new.P99.9))[new.P99.9]
  d = exponent.multi_ratios_est(new.P99, new.P99.5, new.P99.9)
  us.estimate[i-1912] = d
  }
   us.estimate = as.numeric(us.estimate)
  p1.4 = cbind(us.estimate,ab)
p1.4 = data.frame(p1.4)
ggplot(data = p1.4)+
  geom_point(mapping = aes(x = ab, y = us.estimate))+
  labs(y = "a estimated by formula(4)",x = "a estimated by exponent.multi_ratios_est")
```



#Part 2 v. Use your function from problem (iii) to estimate a over time for each of them.

```
Series.layout.A.Table.1 <- read.csv("~/Desktop/report/Series-layout A-Table 1.csv", header=FALSE)
wid = Series.layout.A.Table.1
wid = wid[-c(1,2),]
country = wid$V1
per.unit = wid$V3
Year = wid$V2
per.unit.us = per.unit[country=="United States"]
# For Canada
country = wid$V1
year1 = Year[country=="Canada"]
wid.p99 = wid$V6[country=="Canada"]
```

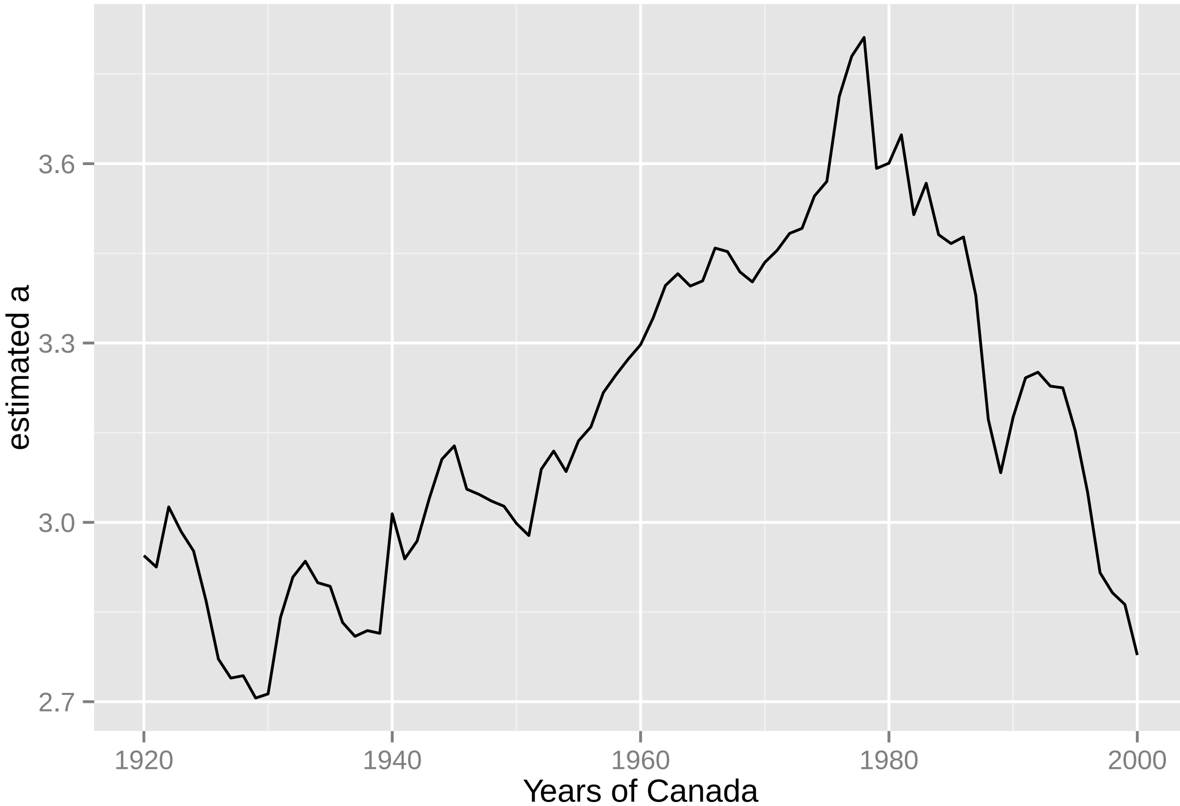
```r
wid.p99.5 = wid$V10[country=="Canada"]
wid.p99.9 = wid$V14[country=="Canada"]
estimate.1 = matrix(NA,ncol = 1, nrow =80)
for (i in 1920:2000){
  new.p99 = wid.p99[year1 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year1 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year1 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.1[i-1919] =g
  }
a_Canada= estimate.1
estimate.year1 = c(1920:2000)
estimate.Canada = cbind(estimate.year1,a_Canada)

# For China
country = wid$V1
Year = wid$V2
year2 = Year[country=="China"]
wid.p99 = wid$V6[country=="China"]
wid.p99.5 = wid$V10[country=="China"]
wid.p99.9 = wid$V14[country=="China"]
estimate.2 = matrix(NA,ncol = 1, nrow =16)
for (i in 1986:2003){
  new.p99 = wid.p99[year2 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year2 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year2 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.2[i-1985] = g
  }
a_China = estimate.2
estimate.year2 = c(1986:2003)
estimate.China = cbind(estimate.year2,a_China)

###For Columbia
country = wid$V1
Year = wid$V2
year3 = Year[country=="Colombia"]
wid.p99 = wid$V6[country=="Colombia"]
wid.p99.5 = wid$V10[country=="Colombia"]
wid.p99.9 = wid$V14[country=="Colombia"]
estimate.3 = matrix(NA,ncol = 1, nrow =10)
for (i in 1993:2002){
  new.p99 = wid.p99[year3 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year3 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year3 == i]
```

```
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.3[i-1992] =g
  }
a_Colombia = estimate.3
estimate.year3 = c(1993:2002)
estimate.Colombia = cbind(estimate.year3,a_Colombia)

#For Italy
country = wid$V1
Year = wid$V2
year4 = Year[country=="Italy"]
wid.p99 = wid$V6[country=="Italy"]
wid.p99.5 = wid$V10[country=="Italy"]
wid.p99.9 = wid$V14[country=="Italy"]
estimate.4 = matrix(NA,ncol = 1, nrow =32)
for (i in 1974:1995){
  new.p99 = wid.p99[year4 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year4 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year4 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.4[i-1973] =g
  }
for (i in 1998:2009){
  new.p99 = wid.p99[year4 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year4 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year4 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.4[i-1973] =g
  }
a_Italy = estimate.4
estimate.year4 = c(1974:2009)
estimate.Italy = cbind(estimate.year4,a_Italy)

#For Japan
country = wid$V1
Year = wid$V2
year5 = Year[country=="Japan"]
wid.p99 = wid$V6[country=="Japan"]
wid.p99.5 = wid$V10[country=="Japan"]
wid.p99.9 = wid$V14[country=="Japan"]
estimate.5 = matrix(NA,ncol = 1, nrow =63)
for (i in 1947:2010){
  new.p99 = wid.p99[year5 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year5 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
```

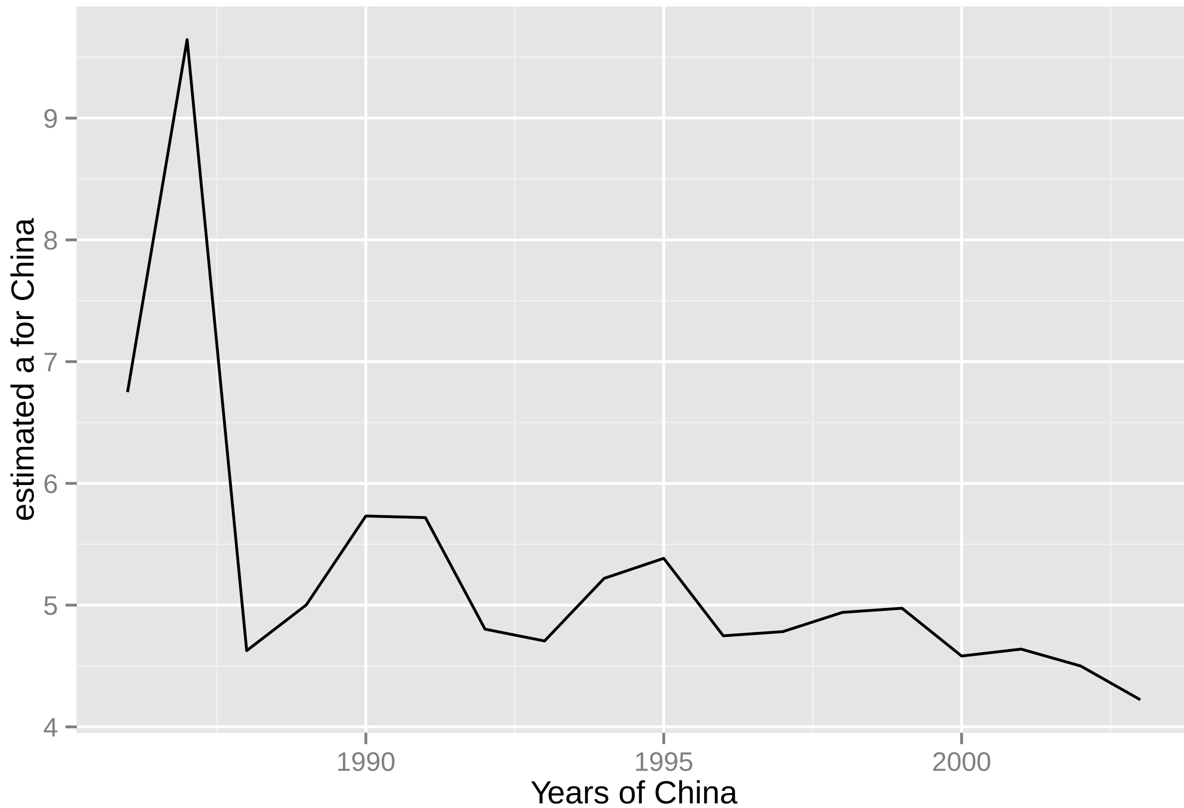
```
  new.p99.9 = wid.p99.9[year5 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.5[i-1946] =g
  }
a_Japan= estimate.5
estimate.year5 = c(1947:2010)
estimate.Japan = data.frame(estimate.year5,a_Japan)

#Sweden
country = wid$V1
Year = wid$V2
year6 = Year[country=="Sweden"]
wid.p99 = wid$V6[country=="Sweden"]
wid.p99.5 = wid$V10[country=="Sweden"]
wid.p99.9 = wid$V14[country=="Sweden"]
estimate.6 = matrix(NA,ncol = 1, nrow =69)
for (i in 1943:2012){
  new.p99 = wid.p99[year6 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year6 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year6 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.6[i-1942] =g
 }
a_Sweden = estimate.6
estimate.year6 = c(1943:2012)
estimate.Sweden = cbind(estimate.year6,a_Sweden)

#For U.S
country = wid$V1
Year = wid$V2
year7 = Year[country=="United States"]
wid.p99 = wid$V6[country=="United States"]
wid.p99.5 = wid$V10[country=="United States"]
wid.p99.9 = wid$V14[country=="United States"]
estimate.7 = matrix(NA,ncol = 1, nrow =99)
for (i in 1913:2012){
 new.p99 = wid.p99[year7 == i]
  new.p99 = as.numeric(levels(new.p99))[new.p99]
  new.p99.5 = wid.p99.5[year7 == i]
  new.p99.5 = as.numeric(levels(new.p99.5))[new.p99.5]
  new.p99.9 = wid.p99.9[year7 == i]
  new.p99.9 = as.numeric(levels(new.p99.9))[new.p99.9]
   g = exponent.multi_ratios_est(new.p99, new.p99.5, new.p99.9)
  estimate.7[i-1912] =g
 }
a_us = estimate.7
estimate.year7 = c(1913:2012)
estimate.us =cbind(estimate.year7,a_us)
```

vi. Plot your estimates of a over time for all the countries using ggplot.

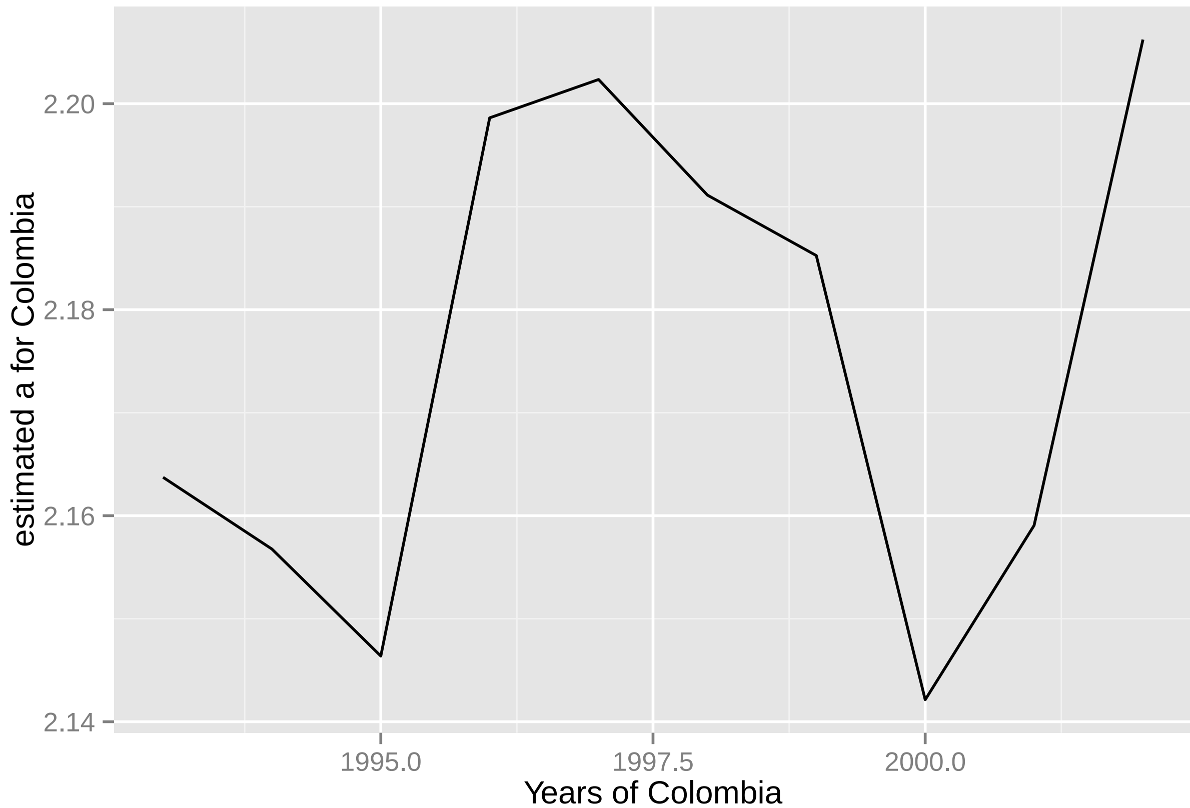
```
#Canada
estimate.Canada = cbind(estimate.year1,a_Canada)
estimate.Canada = data.frame(estimate.Canada)
  ggplot(data = estimate.Canada) + geom_line(mapping=aes(x = estimate.year1, y = a_Canada))+labs(x = "Ye
```



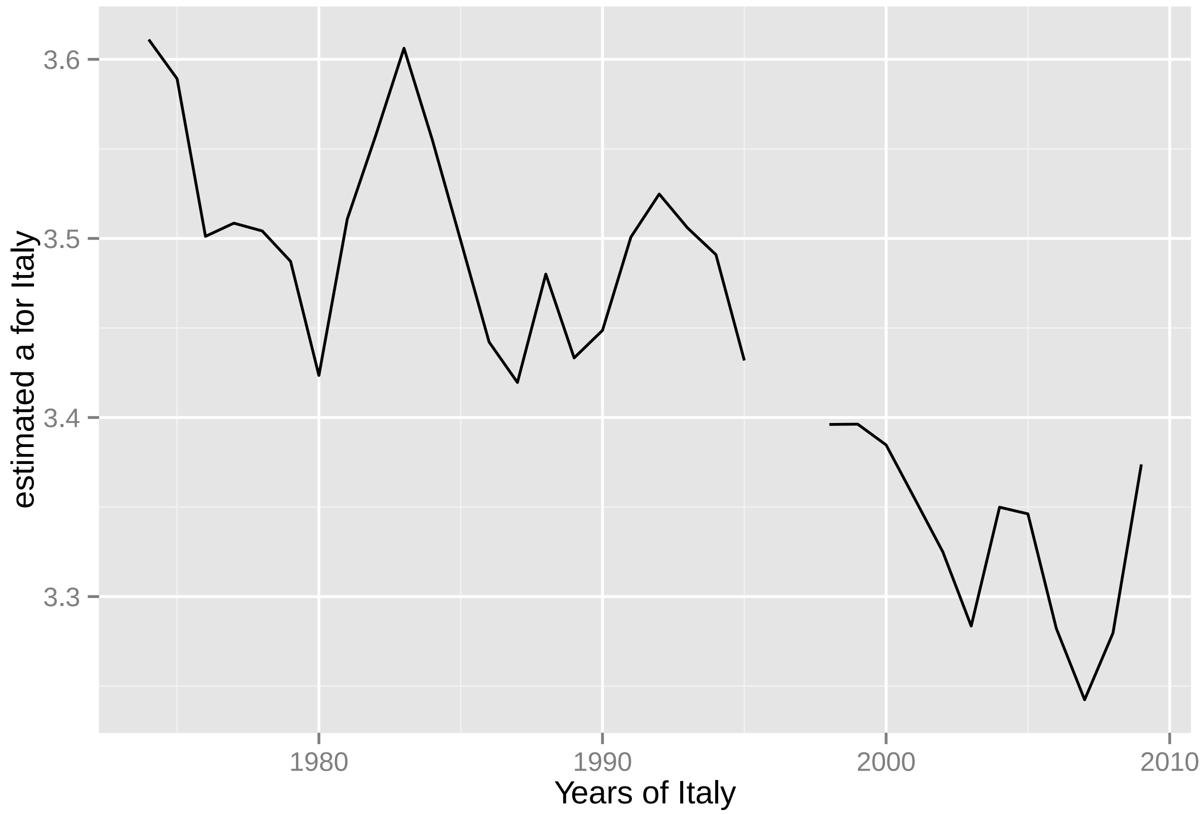

```
#China
a_China = estimate.2
estimate.year2 = c(1986:2003)
estimate.China = cbind(estimate.year2,a_China)
estimate.China = data.frame(estimate.China)
  ggplot(data = estimate.China) + geom_line(mapping=aes(x = estimate.year2, y = a_China ))+labs(x = "Yea
```



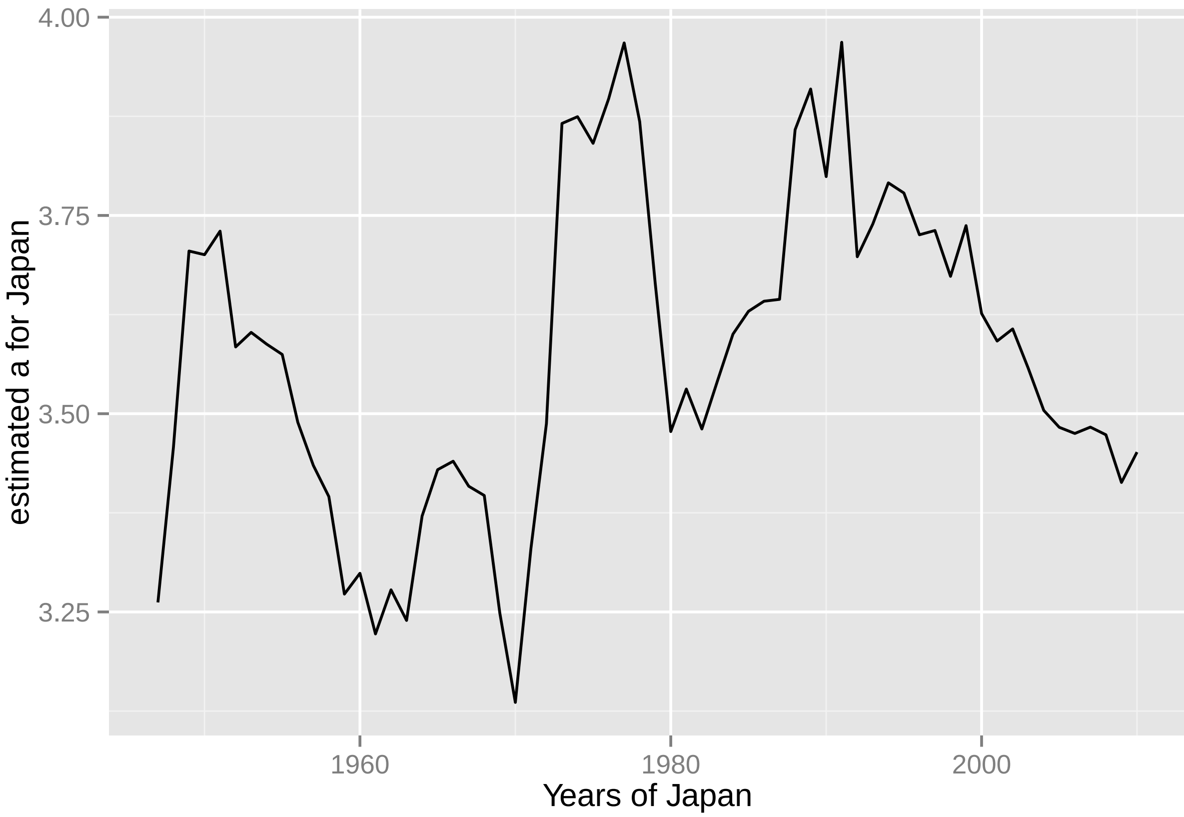

```
#Colombia
estimate.Colombia = cbind(estimate.year3,a_Colombia)
estimate.Colombia = data.frame(estimate.Colombia)
  ggplot(data = estimate.Colombia) + geom_line(mapping=aes(x = estimate.year3, y = a_Colombia ))+labs(x
```



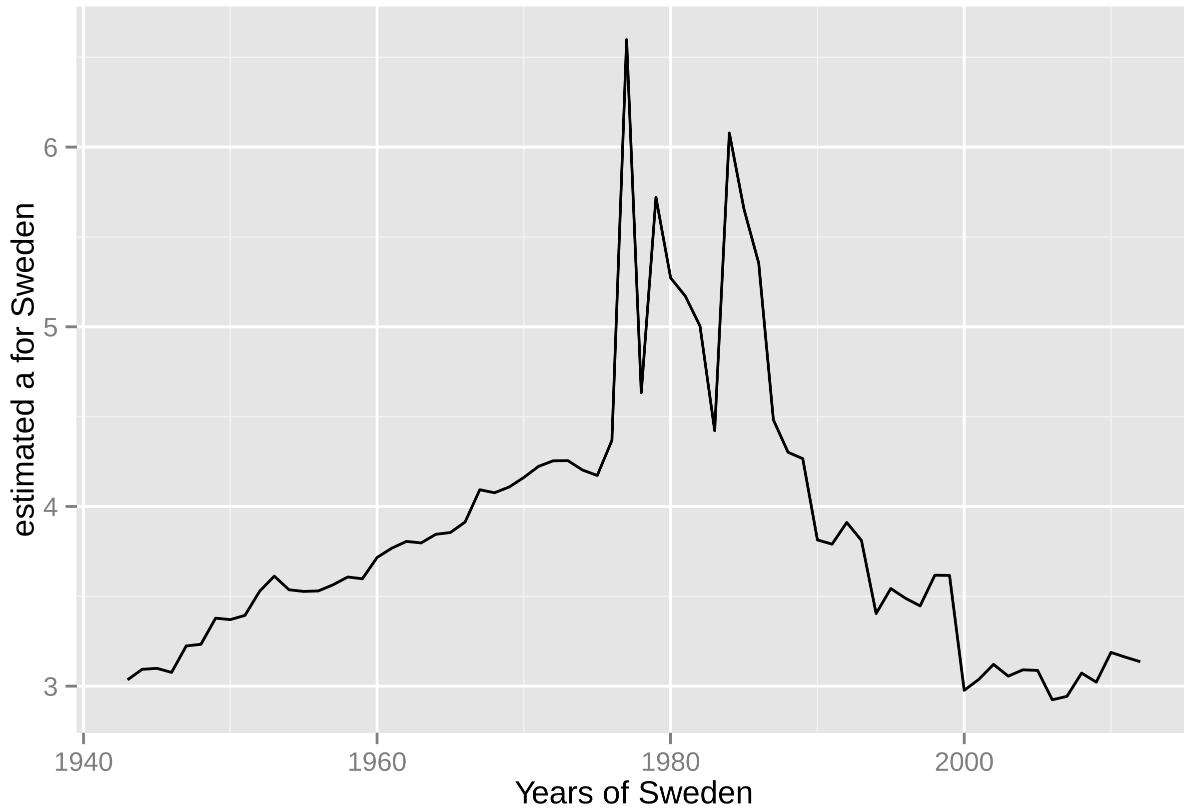

```
#Italy
estimate.Italy = cbind(estimate.year4,a_Italy)
estimate.Italy = data.frame(estimate.Italy)
  ggplot(data = estimate.Italy) + geom_line(mapping=aes(x = estimate.year4, y = a_Italy ))+labs(x = "Yea
```



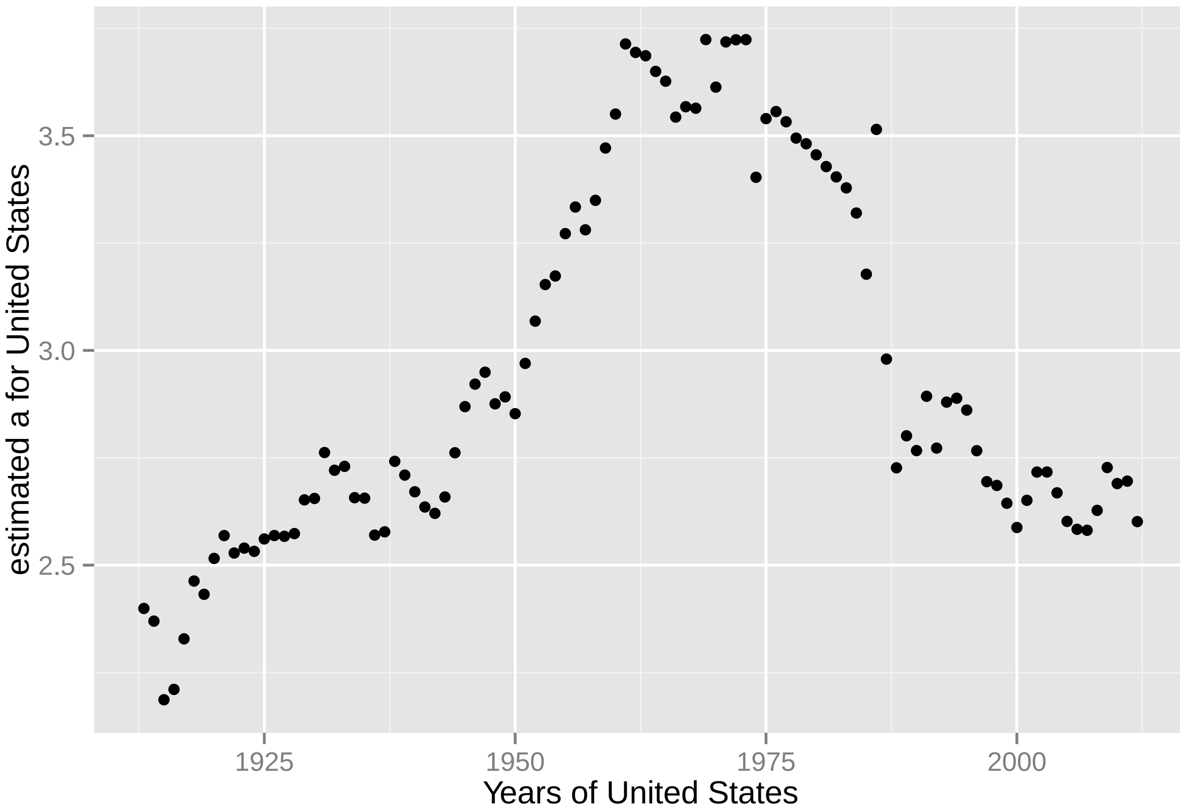

```
#Japan
estimate.Japan = cbind(estimate.year5,a_Japan)
estimate.Japan = data.frame(estimate.Japan)
  ggplot(data = estimate.Japan) + geom_line(mapping=aes(x = estimate.year5, y = a_Japan ))+labs(x = "Yea
```




```
#Sweden
estimate.Sweden = cbind(estimate.year6,a_Sweden)
estimate.Sweden = data.frame(estimate.Sweden)
  ggplot(data = estimate.Sweden) + geom_line(mapping=aes(x = estimate.year6, y = a_Sweden ))+labs(x = "Y
```

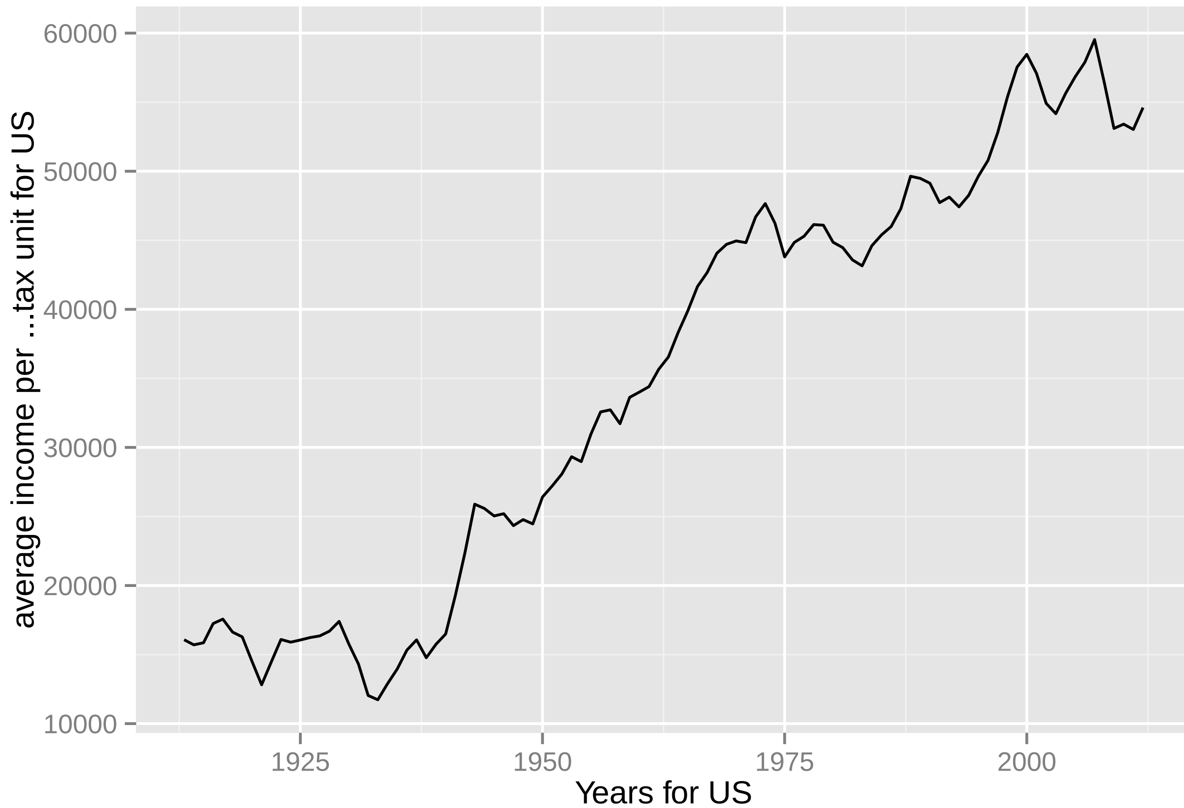



```
#U.S.
estimate.us =cbind(estimate.year7,a_us)
estimate.us = data.frame(estimate.us)
  ggplot(data = estimate.us) + geom_point(mapping=aes(x = estimate.year7, y = a_us ))+labs(x = "Years o
```

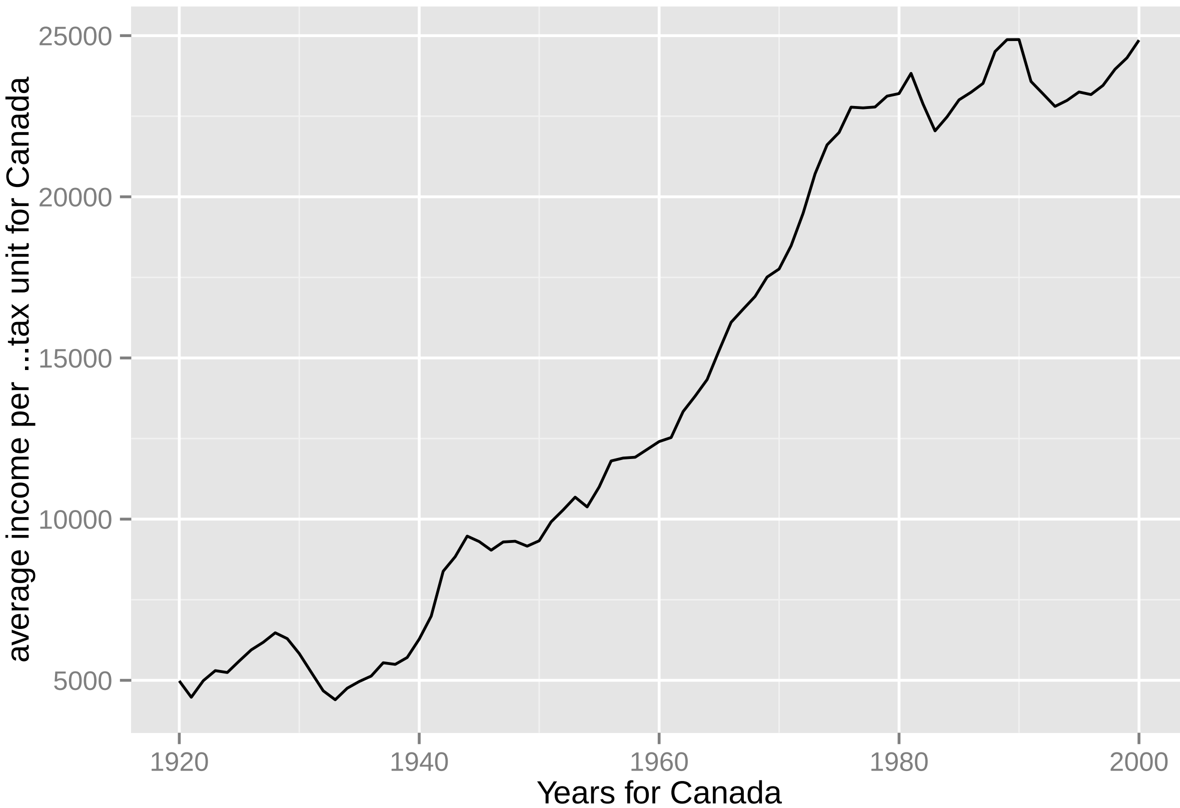


vii. Plot the series of average income per "tax unit" for the US and the countries against time in ggplot.

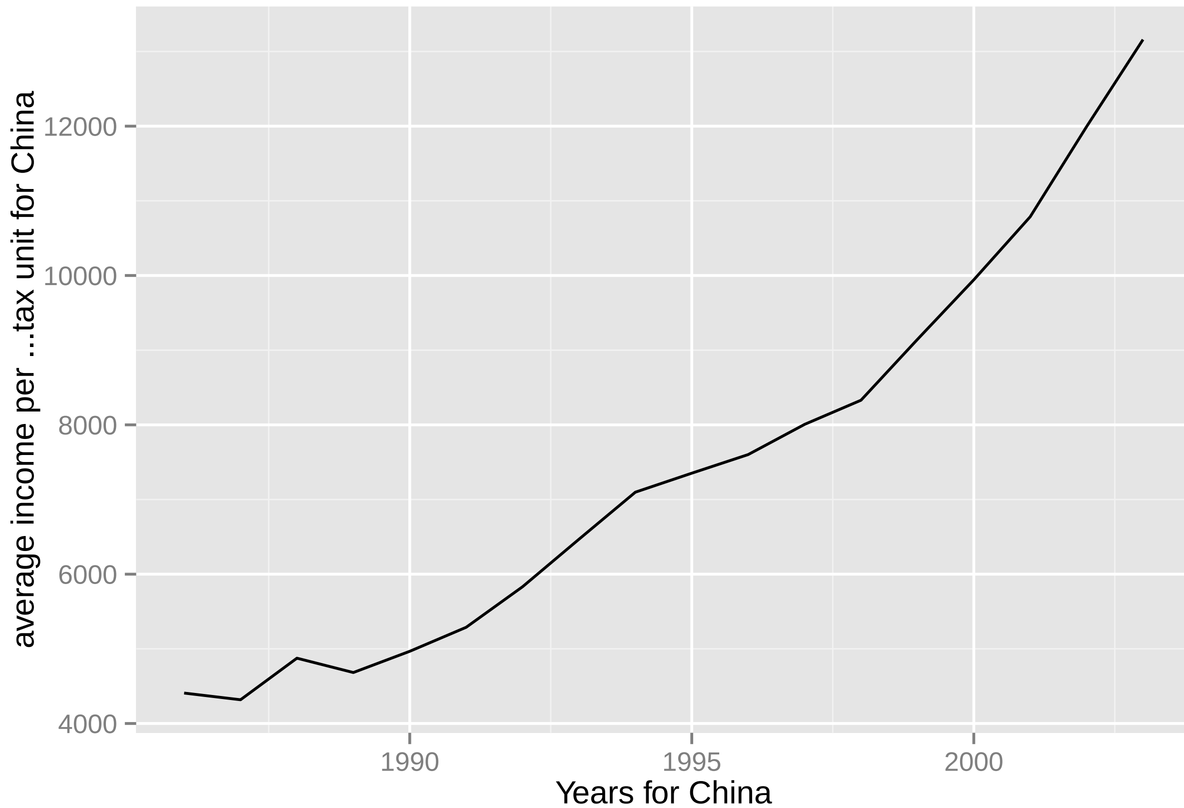
```
#For U.S.
per.unit = wid$V3
Year = wid$V2
per.unit2 = per.unit[country=="United States"]
per.unit2 = as.numeric(levels(per.unit2))[per.unit2]
year=c(1913:2012)
per.unit.US = cbind(per.unit2,year)
per.unit.US = data.frame(per.unit.US)
ggplot(data = per.unit.US)+geom_line(mapping = aes(y = per.unit2 ,x = year)) + labs(y="average income pe
```



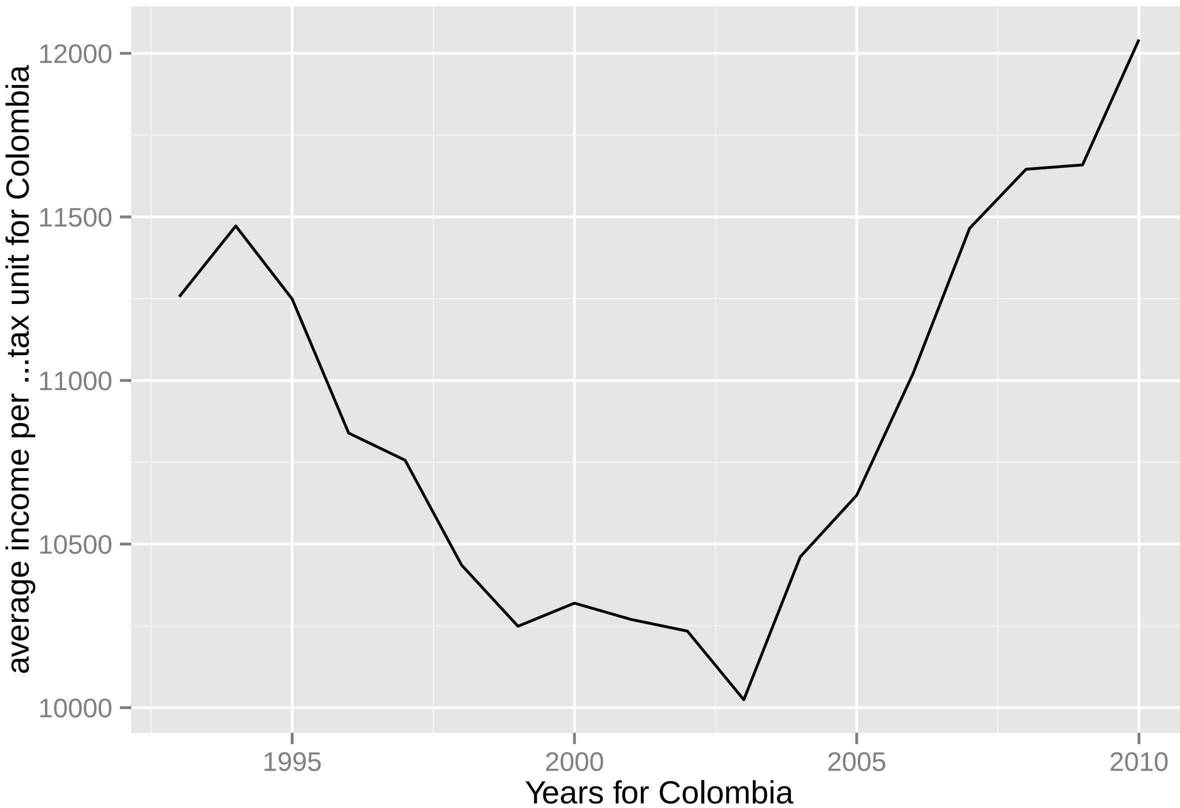

```
#For Canada
per.unit = wid$V3
Year = wid$V2
per.unit1= per.unit[country=="Canada"]
per.unit1 = as.numeric(levels(per.unit1))[per.unit1]
per.unit1= as.vector(per.unit1)
per.unit1 = per.unit1[-c(82:91)]
year=c(1920:2000)
per.unit.Canada= cbind(per.unit1,year)
per.unit.Canada = data.frame(per.unit.Canada)
ggplot(data = per.unit.Canada)+geom_line(mapping = aes(y = per.unit1 ,x = year)) + labs(y="average incor
```



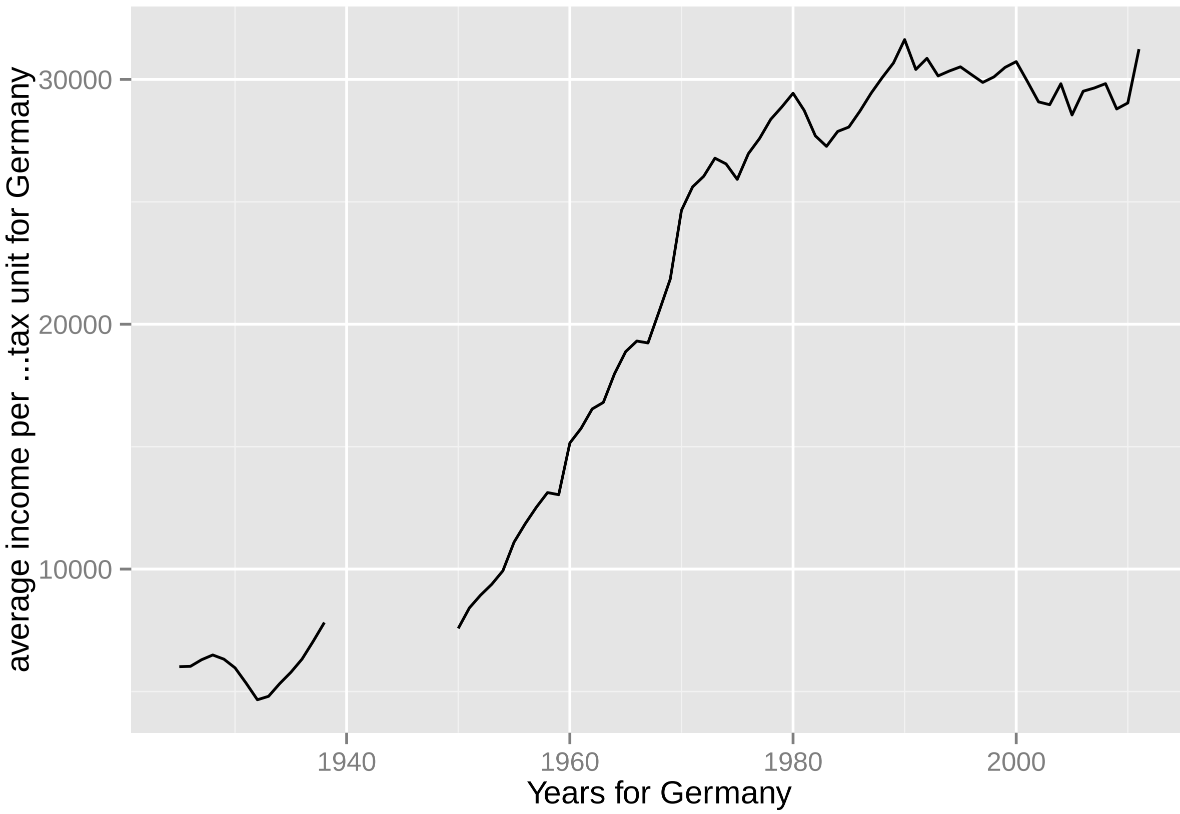

```
#China
per.unit = wid$V3
Year = wid$V2
per.unit3= per.unit[country=="China"]
per.unit3 = as.numeric(levels(per.unit3))[per.unit3]
year=c(1986:2003)
per.unit.China= cbind(per.unit3,year)
per.unit.China = data.frame(per.unit.China)
ggplot(data = per.unit.China)+geom_line(mapping = aes(y = per.unit3,x = year)) + labs(y="average income
```



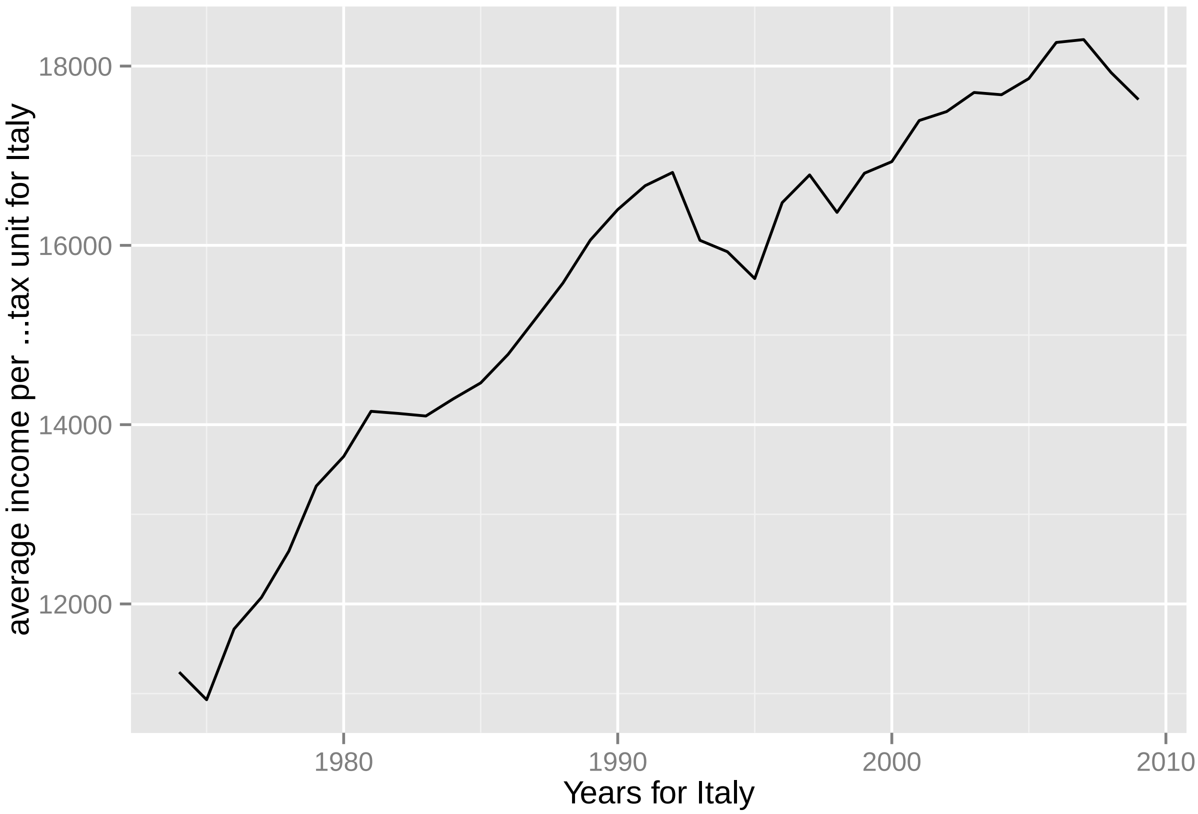

```
#Colombia
per.unit = wid$V3
Year = wid$V2
per.unit4= per.unit[country=="Colombia"]
per.unit4 = as.numeric(levels(per.unit4))[per.unit4]
year=c(1993:2010)
per.unit.Colombia= cbind(per.unit4,year)
per.unit.Colombia = data.frame(per.unit.Colombia)
ggplot(data = per.unit.Colombia)+geom_line(mapping = aes(y = per.unit4,x = year)) + labs(y="average inc
```



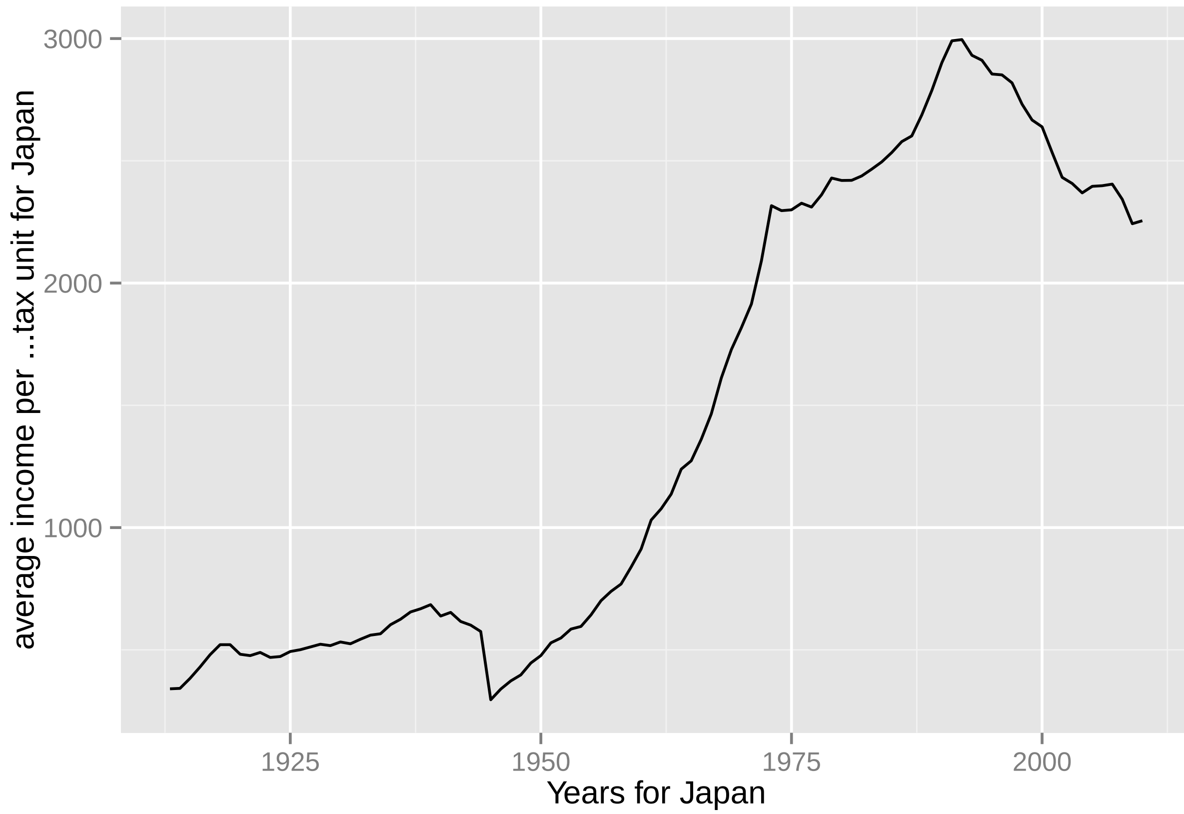

```r
#Germany
per.unit = wid$V3
Year = wid$V2
per.unit5= per.unit[country=="Germany"]
per.unit5 = as.numeric(levels(per.unit5))[per.unit5]
year=c(1925:2011)
per.unit.Germany= cbind(per.unit5,year)
per.unit.Germany = data.frame(per.unit.Germany)
ggplot(data = per.unit.Germany)+geom_line(mapping = aes(y = per.unit5,x = year)) + labs(y="average incon
```



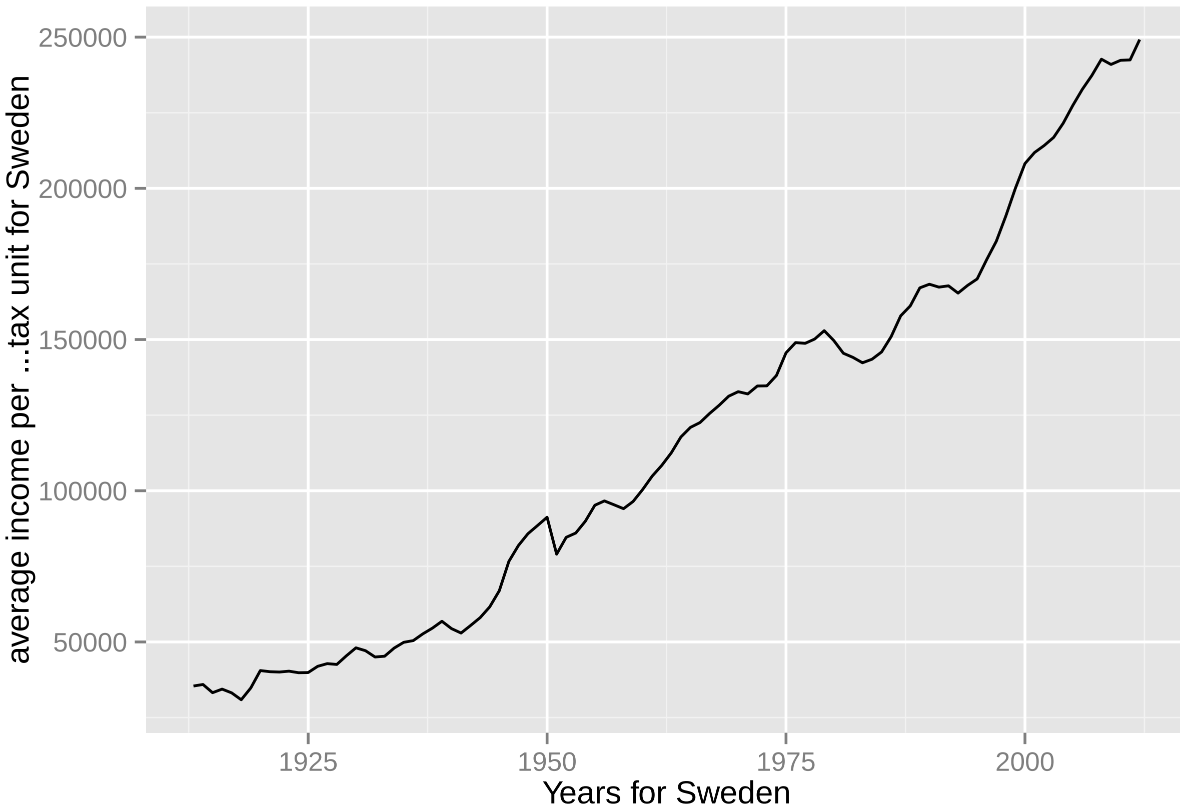

```
#Italy
per.unit = wid$V3
Year = wid$V2
per.unit6= per.unit[country=="Italy"]
per.unit6 = as.numeric(levels(per.unit6))[per.unit6]
year=c(1974:2009)
per.unit.Italy= cbind(per.unit6,year)
per.unit.Italy = data.frame(per.unit.Italy)
ggplot(data = per.unit.Italy)+geom_line(mapping = aes(y = per.unit6,x = year)) + labs(y="average income
```




```
##Japan
per.unit = wid$V3
Year = wid$V2
per.unit7= per.unit[country=="Japan"]
per.unit7 = as.numeric(levels(per.unit7))[per.unit7]
year7=c(1913:2010)
per.unit.Japan= cbind(per.unit7,year7)
per.unit.Japan = data.frame(per.unit.Japan)
ggplot(data = per.unit.Japan)+geom_line(mapping = aes(y = per.unit7,x = year7)) + labs(y="average income
```



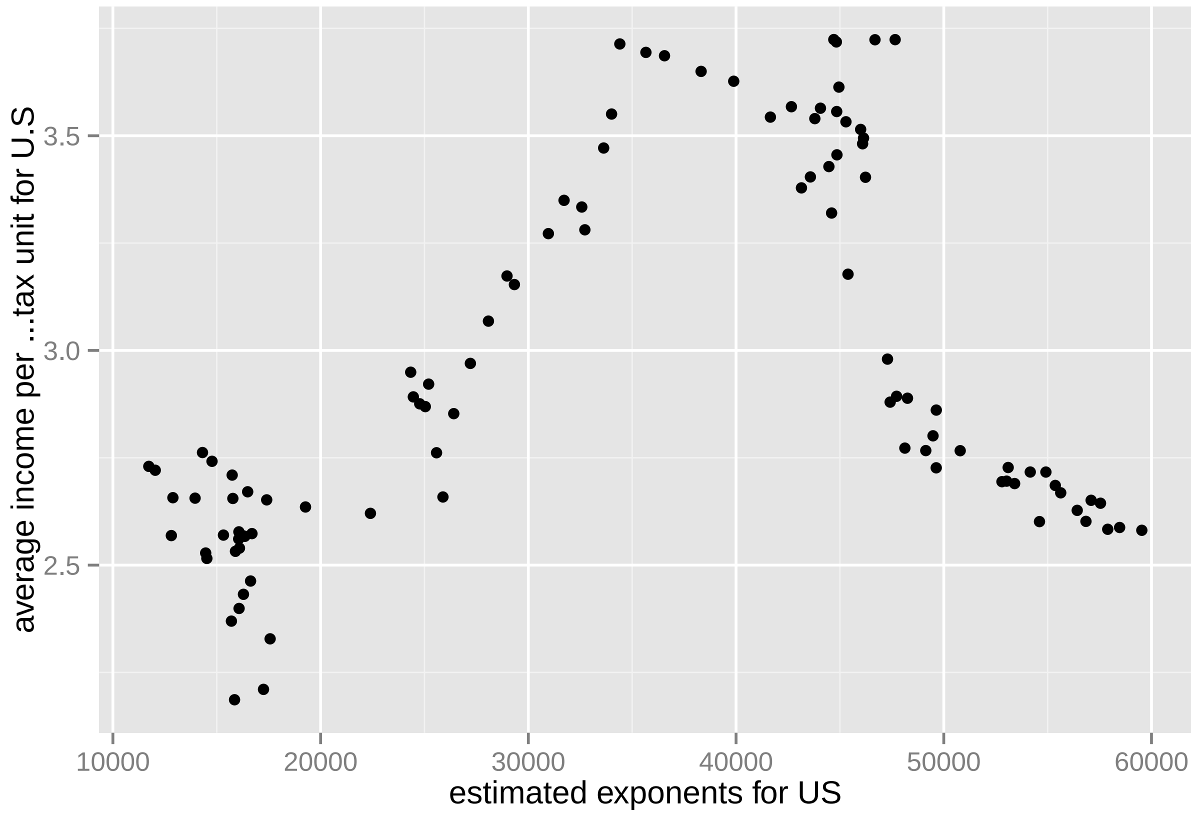

```
#Sweden
per.unit = wid$V3
Year = wid$V2
per.unit8= per.unit[country=="Sweden"]
per.unit8 = as.numeric(levels(per.unit8))[per.unit8]
year8=c(1913:2012)
per.unit.Sweden= cbind(per.unit8,year8)
per.unit.Sweden = data.frame(per.unit.Sweden)
ggplot(data = per.unit.Sweden)+geom_line(mapping = aes(y = per.unit8,x = year8)) + labs(main = " Sweden
```



viii.

Kuznets curve

```
per.unit = wid$V3
Year = wid$V2
per.unit2 = per.unit[country=="United States"]
per.unit2 = as.numeric(levels(per.unit2))[per.unit2]
a_us = estimate.7
us.curve = cbind(per.unit2,a_us)
us.curve = data.frame(us.curve)
ggplot(data = us.curve)+geom_point(mapping = aes(x = per.unit2,y = a_us)) + labs(main = " U.S",y="averag
```




```
#From the plot, we can see the in inequality rises during the early, #industrializing phases of economi
```

ix. For a more quantitative check on the Kuznets hypothesis,

```
summary(lm(a_us~per.unit2 + I(per.unit2^2)))
```

```
##
## Call:
## lm(formula = a_us ~ per.unit2 + I(per.unit2^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50594 -0.19791 -0.02468  0.20151  0.54519
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     8.300e-01  1.550e-01   5.355 5.73e-07 ***
## per.unit2       1.388e-04  1.044e-05  13.292  < 2e-16 ***
## I(per.unit2^2) -1.879e-09  1.505e-10 -12.485  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2495 on 97 degrees of freedom
## Multiple R-squared:  0.666,  Adjusted R-squared:  0.6591
## F-statistic: 96.69 on 2 and 97 DF,  p-value: < 2.2e-16
```

x. Do a separate quadratic regression for each country.

```r
#U.S
lm_us = lm(a_us~per.unit2 + I(per.unit2^2))
coefficients(lm_us)
```

```
##   (Intercept)     per.unit2 I(per.unit2^2)
##  8.300055e-01  1.388163e-04  -1.878718e-09
```

```r
#Canada
lm_Canada = lm(a_Canada~per.unit1 + I(per.unit1^2))
coefficients(lm_Canada)
```

```
##   (Intercept)     per.unit1 I(per.unit1^2)
##  2.266054e+00  1.240966e-04  -3.360837e-09
```

```r
#China
lm_China = lm(a_China~per.unit3 + I(per.unit3^2))
coefficients(lm_China)
```

```
##   (Intercept)     per.unit3 I(per.unit3^2)
##  1.039781e+01 -1.126763e-03   5.257536e-08
```

```r
#For Colombia
per.unit4 = per.unit4[-c(11:18)]
lm_Colombia = lm(a_Colombia~per.unit4 + I(per.unit4^2))
coefficients(lm_Colombia)
```

```
##   (Intercept)     per.unit4 I(per.unit4^2)
## -6.150952e+00  1.560704e-03  -7.300586e-08
```

```r
#For Italy
per.unit6= per.unit[country=="Italy"]
per.unit6 = as.numeric(levels(per.unit6))[per.unit6]
lm_Italy = lm(a_Italy~per.unit6 + I(per.unit6^2))
coefficients(lm_Italy)
```

```
##   (Intercept)     per.unit6 I(per.unit6^2)
##  2.582416e+00  1.594300e-04  -6.591048e-09
```

```r
#For Japan
per.unit7 = per.unit7[-c(1:34)]
lm_Japan = lm(a_Japan~per.unit7 + I(per.unit7^2))
coefficients(lm_Japan)
```

```
##   (Intercept)     per.unit7 I(per.unit7^2)
##  3.729107e+00 -5.136191e-04   1.889447e-07
```

```r
#For Sweden
per.unit8 = per.unit8[-c(1:30)]
lm_Sweden = lm(a_Sweden~per.unit8 + I(per.unit8^2))
coefficients(lm_Sweden)
```

```
##    (Intercept)      per.unit8 I(per.unit8^2)
##  -8.808640e-02   5.889961e-05  -1.950797e-10
```

```
# Summary: US, Canada,Italy,Sweden compatible with the hypothesis
```