5206hw5_br2498

Bo Rong br2498

October 27, 2016

```
#part1
#i.
percentile_ratio_discrepancies<-function(vec=c(P99,P99.5,P99.9,a,na.rm=T)){
  value<-((vec[1]/vec[3])^(-vec[4]+1)-10)^2+((vec[2]/vec[3])^(-vec[4]+1)-5)^2+((vec[1]/vec[2])^(-vec[4]+
  return(value)
}
percentile_ratio_discrepancies(c(1e6,2e6,1e7,2))
```

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

```
#ii.
a=1-log(10)/log(1e6/1e7)
exponent.multi_ratios_est<-function(vec,na.rm=T){
  est<-nlm(percentile_ratio_discrepancies,vec)$estimate[4]
  return(est)
}
exponent.multi_ratios_est(c(1e6,2e6,1e7,1-log(10)/log(1e6/1e7)))
```

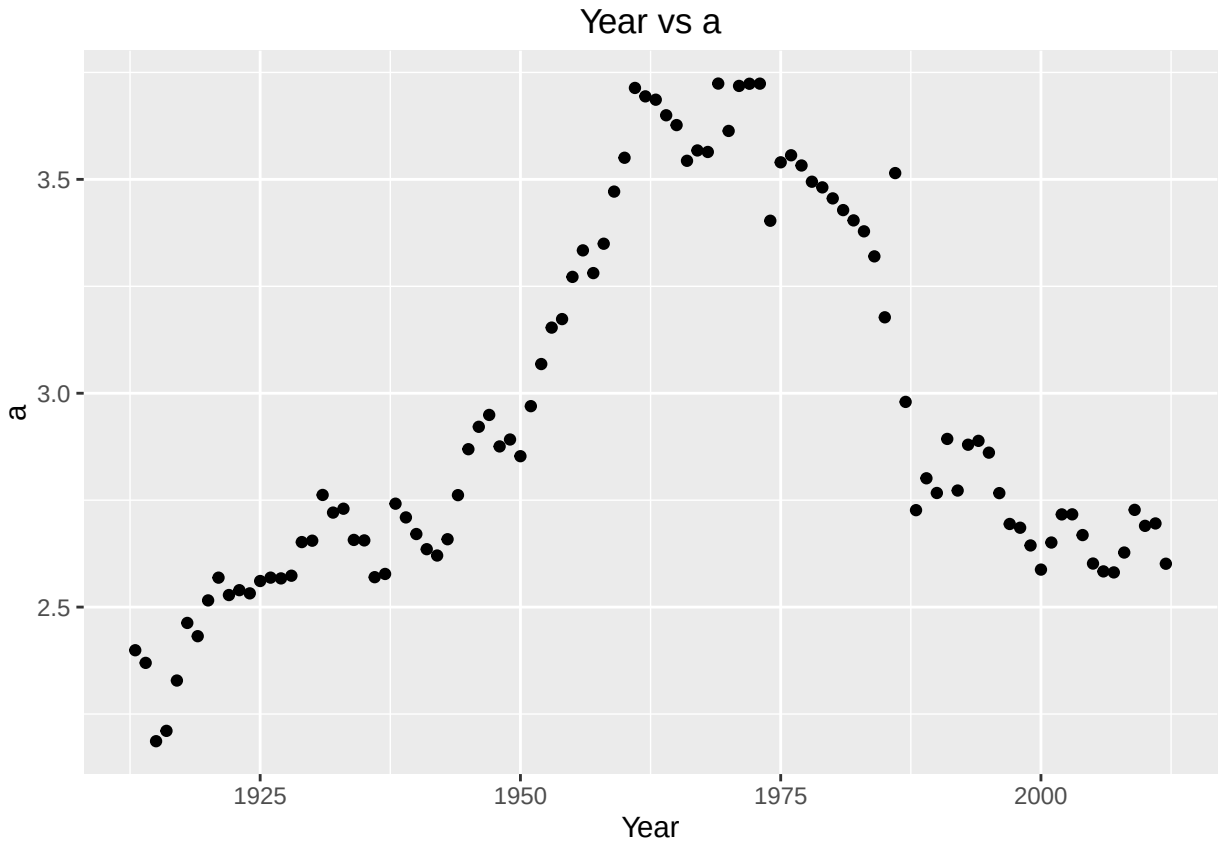
```
## [1] 2
```

```
#iii.
setwd("/Users/brong/Downloads")
wtid_report<- read.csv("wtid-report.csv", header = TRUE)
wtid_reportnew<-data.frame(wtid_report$Year,wtid_report$P99.income.threshold,wtid_report$P99.5.income.t
colnames(wtid_reportnew)<-c("year","P99","P99.5","P99.9")
wtid_reportnew$est.a<-1-log(10)/log(wtid_reportnew$P99/wtid_reportnew$P99.9)
wtid_reportnew$a<-apply(wtid_reportnew[,c(2,3,4,5)],1,exponent.multi_ratios_est)


library(ggplot2)
```

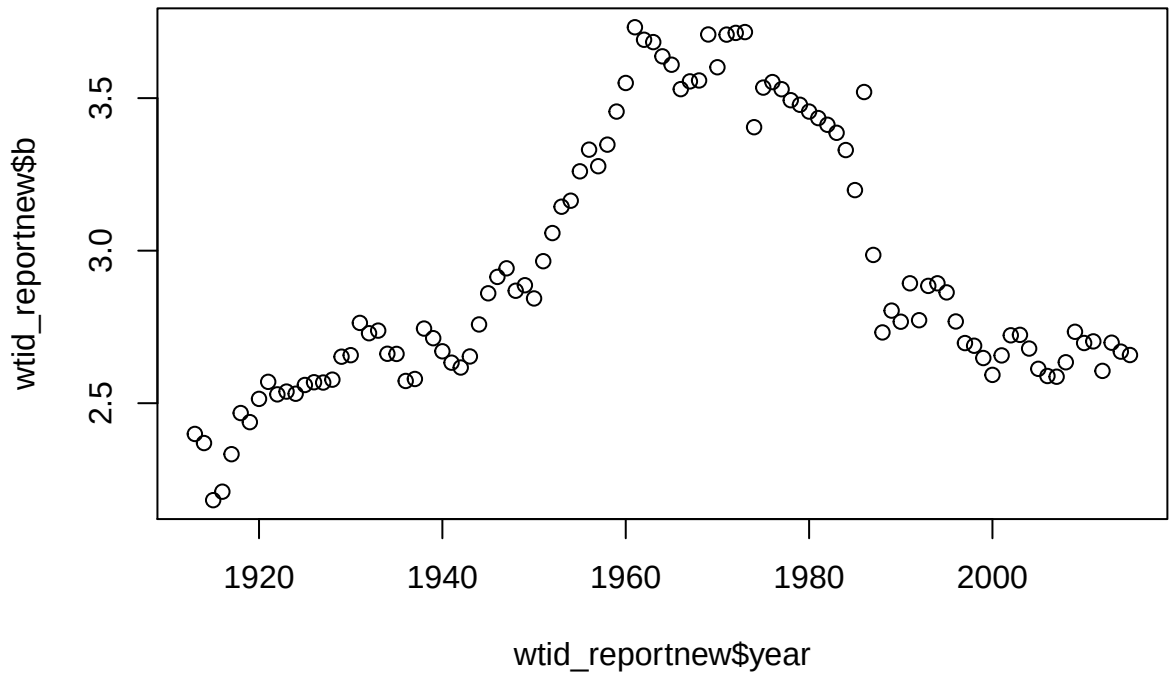
```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```
ggplot(data = wtid_reportnew)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Year vs a", x = "Year", y = "a")+
xlim(1913, 2012)
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```




```
#iv.
wtid_reportnew$b<-wtid_reportnew$est.a
plot(wtid_reportnew$year,wtid_reportnew$b)
```



```r
#THese two plots are very similar.
```

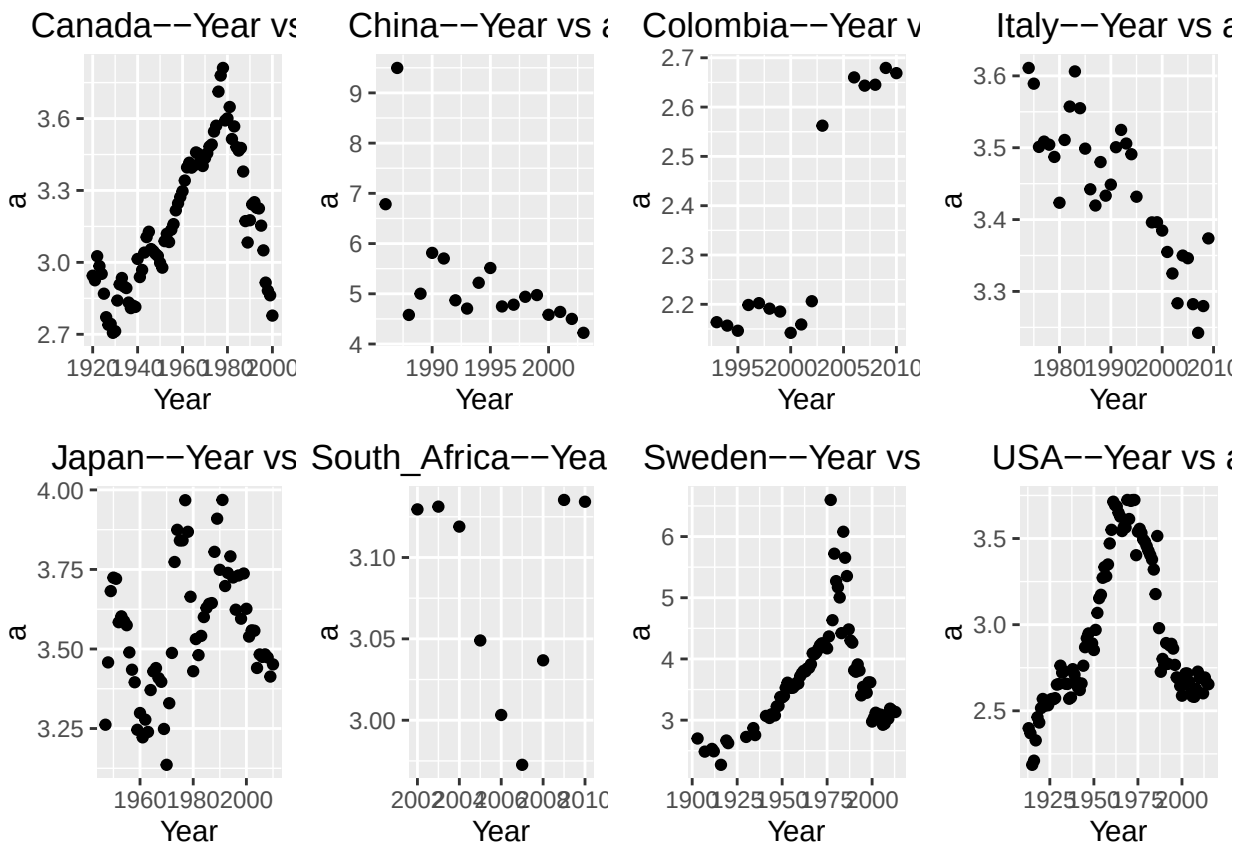
```r
#v.
setwd("/Users/brong/Downloads")
Canada_report<- read.csv("Canada.csv", header = TRUE)
Canada_report$est.a<-1-log(10)/log(Canada_report$P99/Canada_report$P99.9)
Canada_report$a<-apply(Canada_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)

China_report<- read.csv("China.csv", header = TRUE)
China_report$est.a<-1-log(10)/log(China_report$P99/China_report$P99.9)
China_report$a<-apply(China_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)


Colombia_report<- read.csv("Colombia.csv", header = TRUE)
Colombia_report$est.a<-1-log(10)/log(Colombia_report$P99/Colombia_report$P99.9)
Colombia_report<-na.exclude(Colombia_report)
Colombia_report$a<-apply(Colombia_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)




Italy_report<- read.csv("Italy.csv", header = TRUE)
Italy_report$est.a<-1-log(10)/log(Italy_report$P99/Italy_report$P99.9)
Italy_report<-na.exclude(Italy_report)
Italy_report$a<-apply(Italy_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)



Japan_report<- read.csv("Japan.csv", header = TRUE)
Japan_report$est.a<-1-log(10)/log(Japan_report$P99/Japan_report$P99.9)
Japan_report<-na.exclude(Japan_report)
Japan_report$a<-apply(Japan_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)



South_Africa_report<- read.csv("South_Africa.csv", header = TRUE)
South_Africa_report$est.a<-1-log(10)/log(South_Africa_report$P99/South_Africa_report$P99.9)
South_Africa_report<-na.exclude(South_Africa_report)
South_Africa_report$a<-apply(South_Africa_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)



Sweden_report<- read.csv("Sweden.csv", header = TRUE)
Sweden_report$est.a<-1-log(10)/log(Sweden_report$P99/Sweden_report$P99.9)
Sweden_report<-na.exclude(Sweden_report)
Sweden_report$a<-apply(Sweden_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)

USA_report<- read.csv("United_states.csv", header = TRUE)
USA_report$est.a<-1-log(10)/log(USA_report$P99/USA_report$P99.9)
USA_report$a<-apply(USA_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
```

```
Germany_report<- read.csv("Germany.csv", header = TRUE)
```

```
#vi.
library(ggplot2)
library(grid)
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.2.4
```

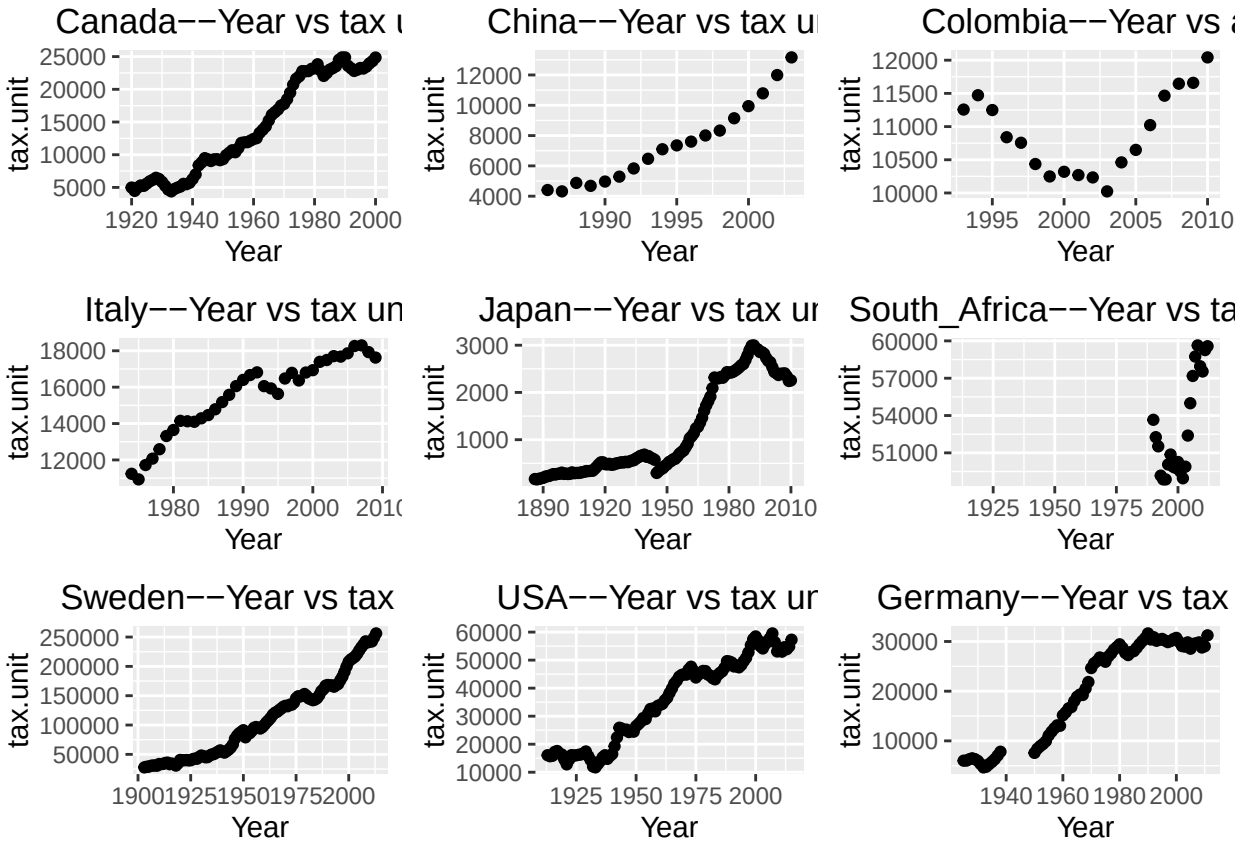
```
plot1<-ggplot(data =Canada_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Canada--Year vs a", x = "Year", y = "a")

plot2<-ggplot(data =China_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "China--Year vs a", x = "Year", y = "a")


plot3<-ggplot(data =Colombia_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Colombia--Year vs a", x = "Year", y = "a")

plot4<-ggplot(data =Italy_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Italy--Year vs a", x = "Year", y = "a")

plot5<-ggplot(data =Japan_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Japan--Year vs a", x = "Year", y = "a")

plot6<-ggplot(data =South_Africa_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "South_Africa--Year vs a", x = "Year", y = "a")

plot7<-ggplot(data =Sweden_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "Sweden--Year vs a", x = "Year", y = "a")

plot8<-ggplot(data =USA_report)+
geom_point(mapping = aes(x = year, y = a))+
labs(title = "USA--Year vs a", x = "Year", y = "a")

grid.arrange(plot1, plot2,plot3,plot4,plot5,plot6,plot7,plot8,nrow=2,ncol=4)
```



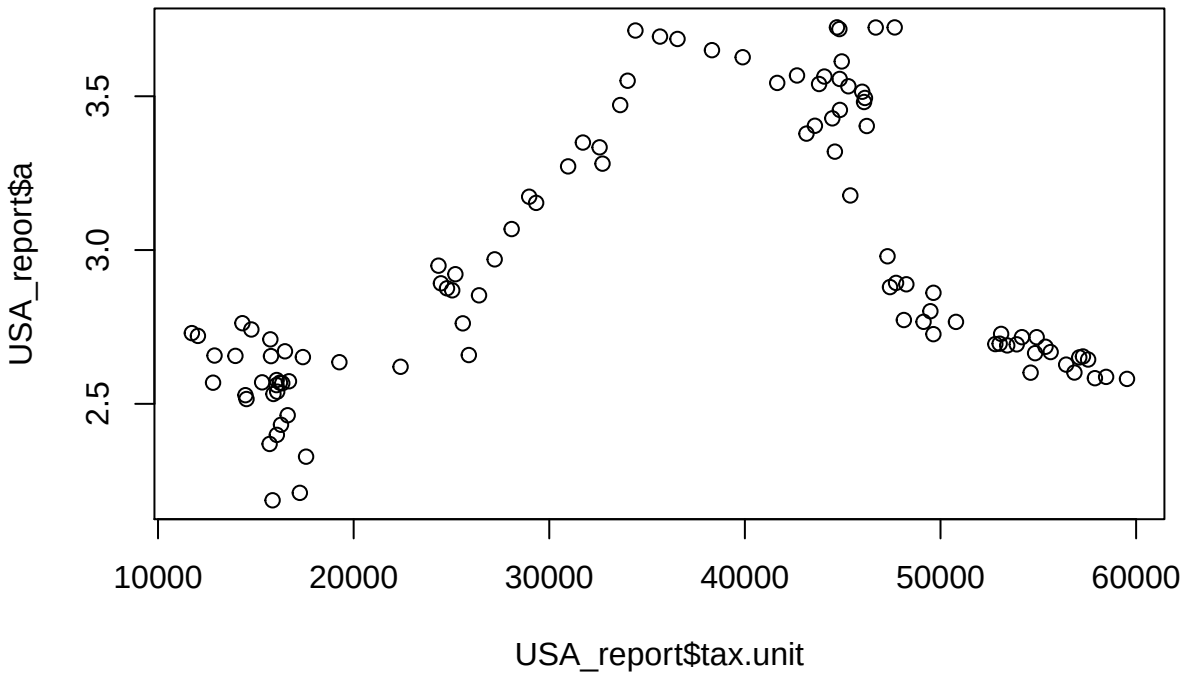

```
#vii.
setwd("/Users/brong/Downloads")
Canada_report<- read.csv("Canada.csv", header = TRUE)
China_report<- read.csv("China.csv", header = TRUE)
Colombia_report<- read.csv("Colombia.csv", header = TRUE)
Italy_report<- read.csv("Italy.csv", header = TRUE)
Japan_report<- read.csv("Japan.csv", header = TRUE)
South_Africa_report<- read.csv("South_Africa.csv", header = TRUE)
Sweden_report<- read.csv("Sweden.csv", header = TRUE)
Germany_report<- read.csv("Germany.csv", header = TRUE)


plot1<-ggplot(data =Canada_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "Canada--Year vs tax unit", x = "Year", y = "tax.unit")

plot2<-ggplot(data =China_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "China--Year vs tax unit", x = "Year", y = "tax.unit")


plot3<-ggplot(data =Colombia_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "Colombia--Year vs a", x = "Year", y = "tax.unit")

plot4<-ggplot(data =Italy_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
```

```
labs(title = "Italy--Year vs tax unit", x = "Year", y = "tax.unit")

plot5<-ggplot(data =Japan_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "Japan--Year vs tax unit", x = "Year", y = "tax.unit")

plot6<-ggplot(data =South_Africa_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "South_Africa--Year vs tax unit", x = "Year", y = "tax.unit")

plot7<-ggplot(data =Sweden_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "Sweden--Year vs tax unit", x = "Year", y = "tax.unit")

plot8<-ggplot(data =USA_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "USA--Year vs tax unit", x = "Year", y = "tax.unit")

plot9<-ggplot(data =Germany_report)+
geom_point(mapping = aes(x = year, y = tax.unit))+
labs(title = "Germany--Year vs tax unit", x = "Year", y = "tax.unit")

grid.arrange(plot1, plot2,plot3,plot4,plot5,plot6,plot7,plot8,plot9,nrow=3,ncol=3)
```

```
## Warning: Removed 77 rows containing missing values (geom_point).

## Warning: Removed 11 rows containing missing values (geom_point).
```

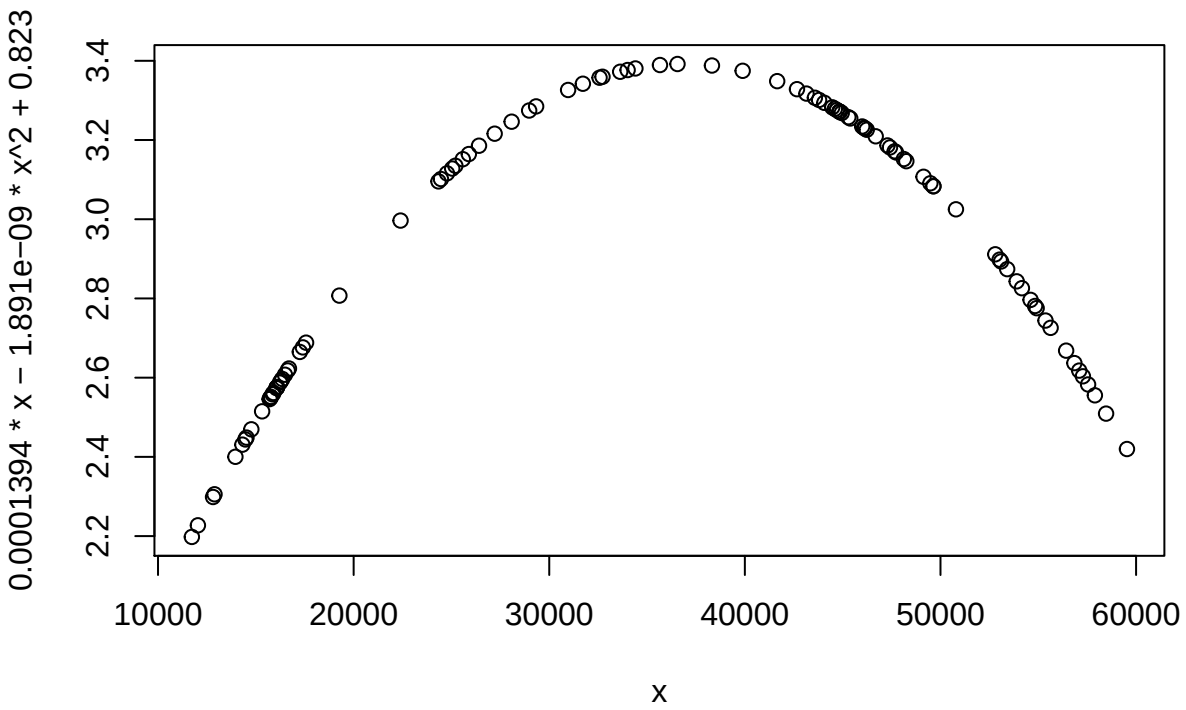



```
#viii.
plot(USA_report$tax.unit,USA_report$a)
```



```
#The plot has a "U curve" shape.
```

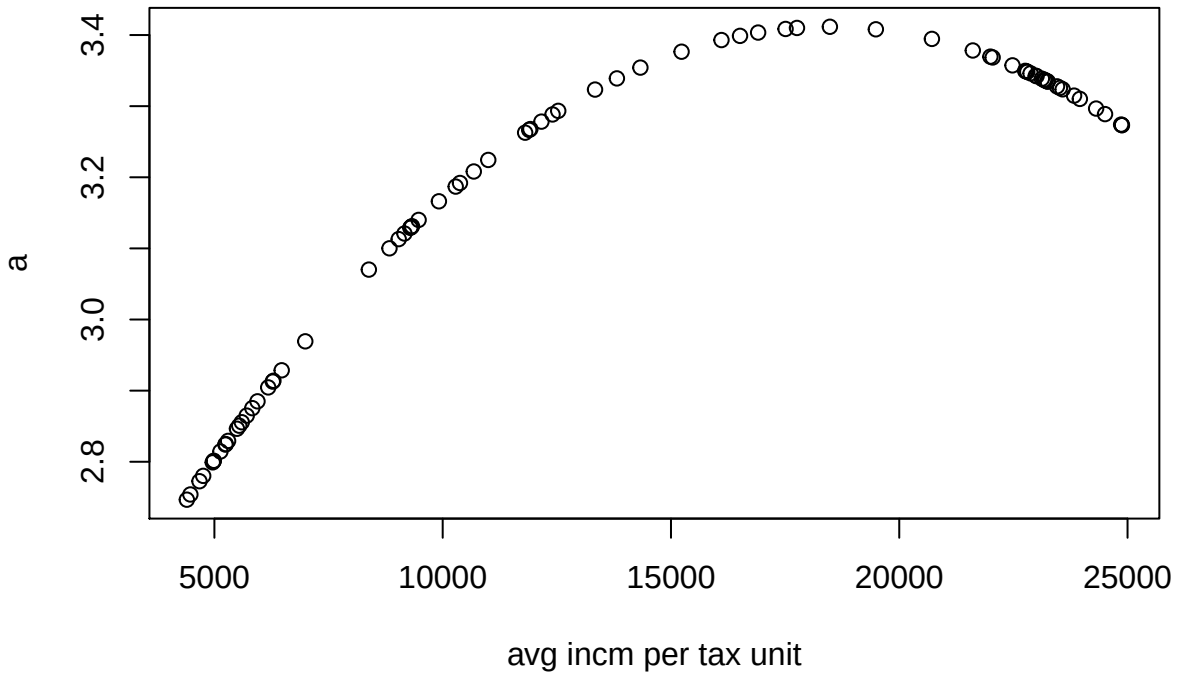
```
#ix.
x<-USA_report$tax.unit
y<-USA_report$a
fit<-lm(y~x+I(x^2))
#y=0.0001394x-0.000000001891x^2+0.823
plot(x,0.0001394*x-0.000000001891*x^2+0.823)
```



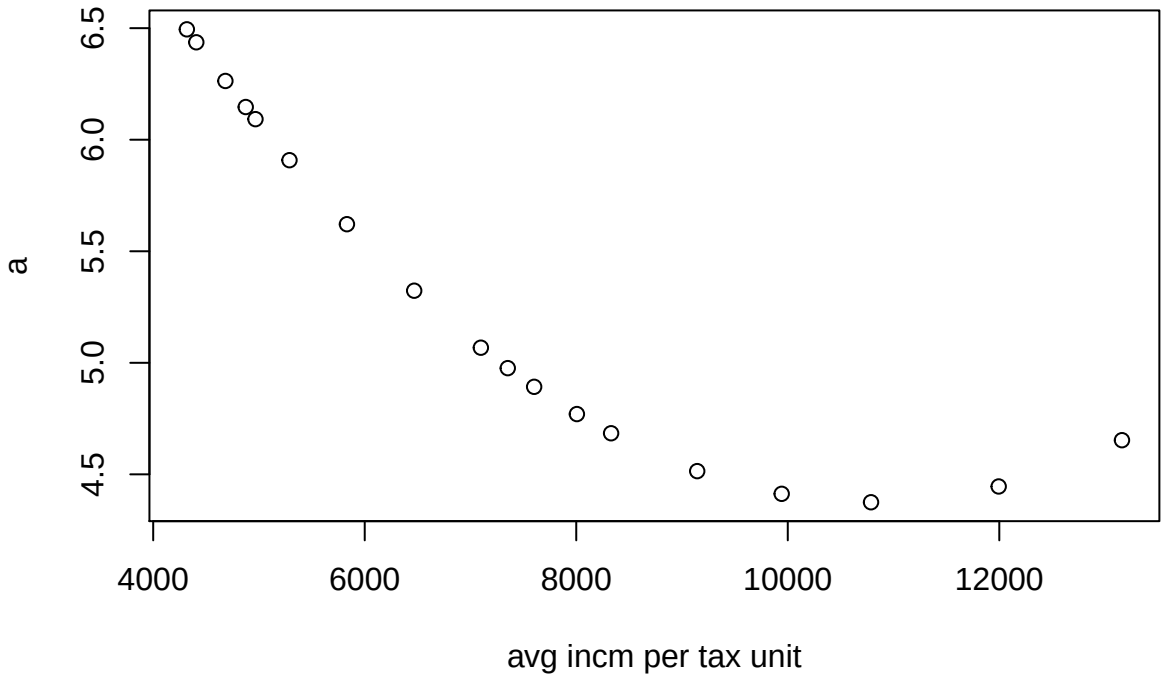

```
#yes,the coefficients I get consistent with the hypothesis.
```

```
#x.
setwd("/Users/brong/Downloads")
Canada_report$est.a<-1-log(10)/log(Canada_report$P99/Canada_report$P99.9)
Canada_report$a<-apply(Canada_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
China_report$est.a<-1-log(10)/log(China_report$P99/China_report$P99.9)
China_report$a<-apply(China_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
Colombia_report$est.a<-1-log(10)/log(Colombia_report$P99/Colombia_report$P99.9)
Colombia_report<-na.exclude(Colombia_report)
Colombia_report$a<-apply(Colombia_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
Italy_report$est.a<-1-log(10)/log(Italy_report$P99/Italy_report$P99.9)
Italy_report<-na.exclude(Italy_report)
Italy_report$a<-apply(Italy_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
Japan_report$est.a<-1-log(10)/log(Japan_report$P99/Japan_report$P99.9)
Japan_report<-na.exclude(Japan_report)
Japan_report$a<-apply(Japan_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
South_Africa_report$est.a<-1-log(10)/log(South_Africa_report$P99/South_Africa_report$P99.9)
South_Africa_report<-na.exclude(South_Africa_report)
South_Africa_report$a<-apply(South_Africa_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
Sweden_report$est.a<-1-log(10)/log(Sweden_report$P99/Sweden_report$P99.9)
```

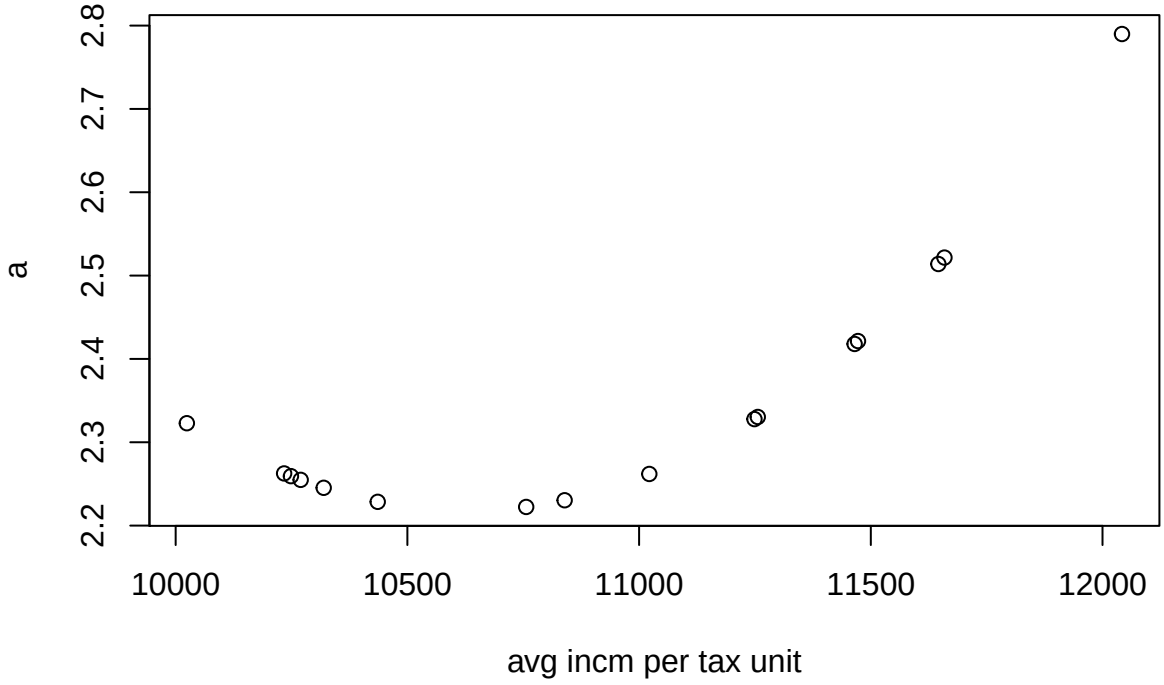
```
Sweden_report<-na.exclude(Sweden_report)
Sweden_report$a<-apply(Sweden_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
USA_report$est.a<-1-log(10)/log(USA_report$P99/USA_report$P99.9)
USA_report$a<-apply(USA_report[,c(4,5,6,7)],1,exponent.multi_ratios_est)
Germany_report<- read.csv("Germany.csv", header = TRUE)

coeffun<-function(A){
  n<-A$tax.unit
  m<-A$a
  t<-lm(m ~ n + I(n^2))
  t<-as.vector(coef(t))
  return(plot(n,t[1]+t[2]*n+t[3]*n^2,xlab ="avg incm per tax unit", ylab = "a"))
}

coeffun(Canada_report)
```



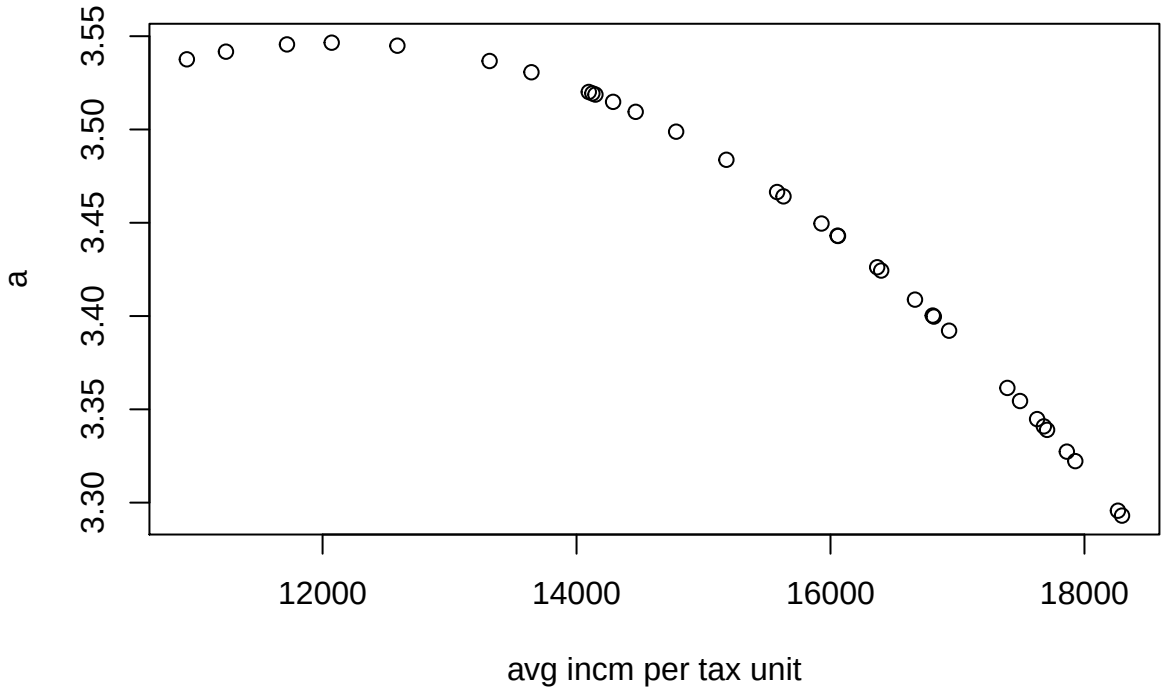

```
coeffun(China_report)
```



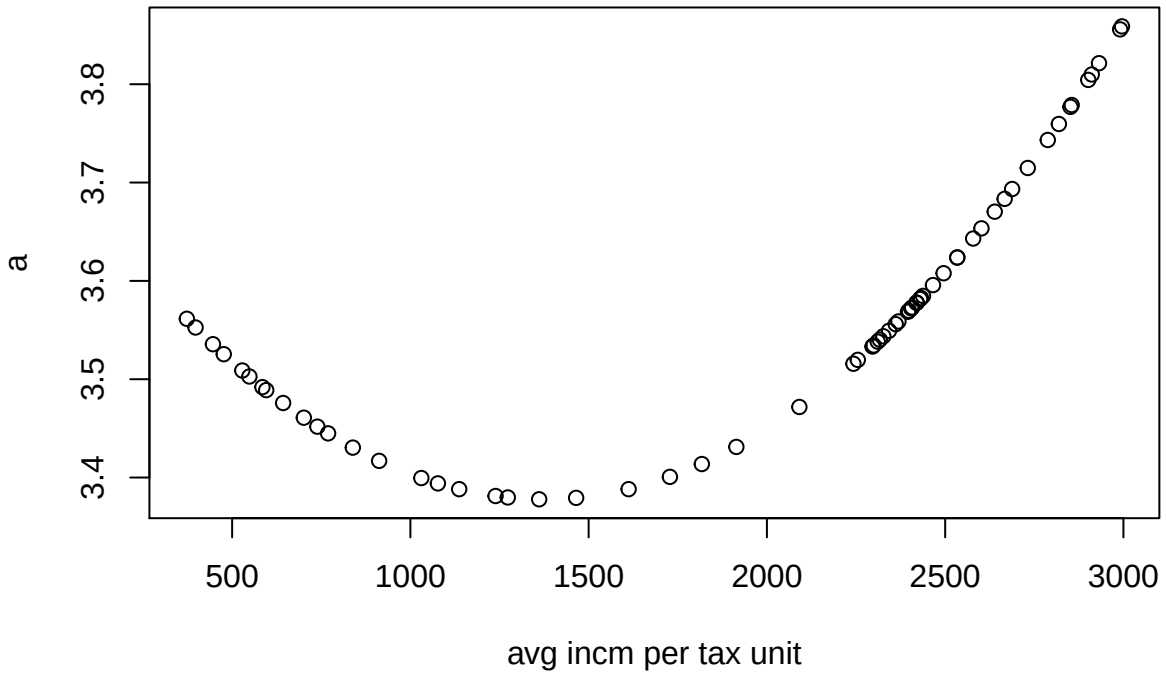

```
coeffun(Colombia_report)
```



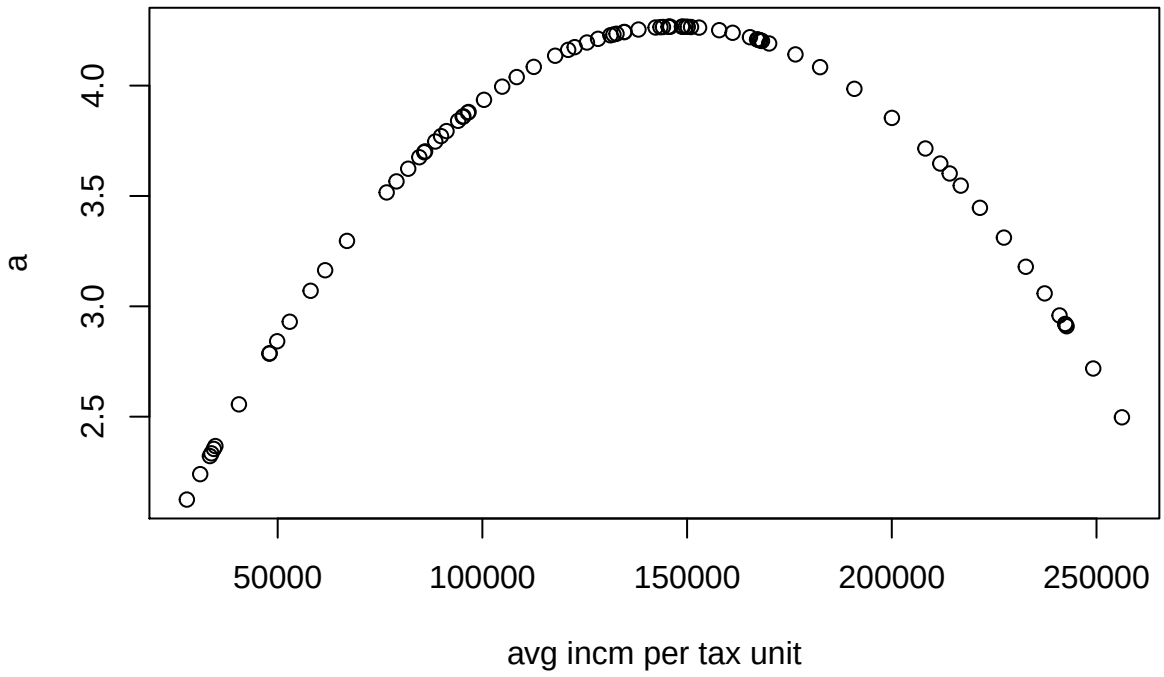

```
coeffun(Italy_report)
```



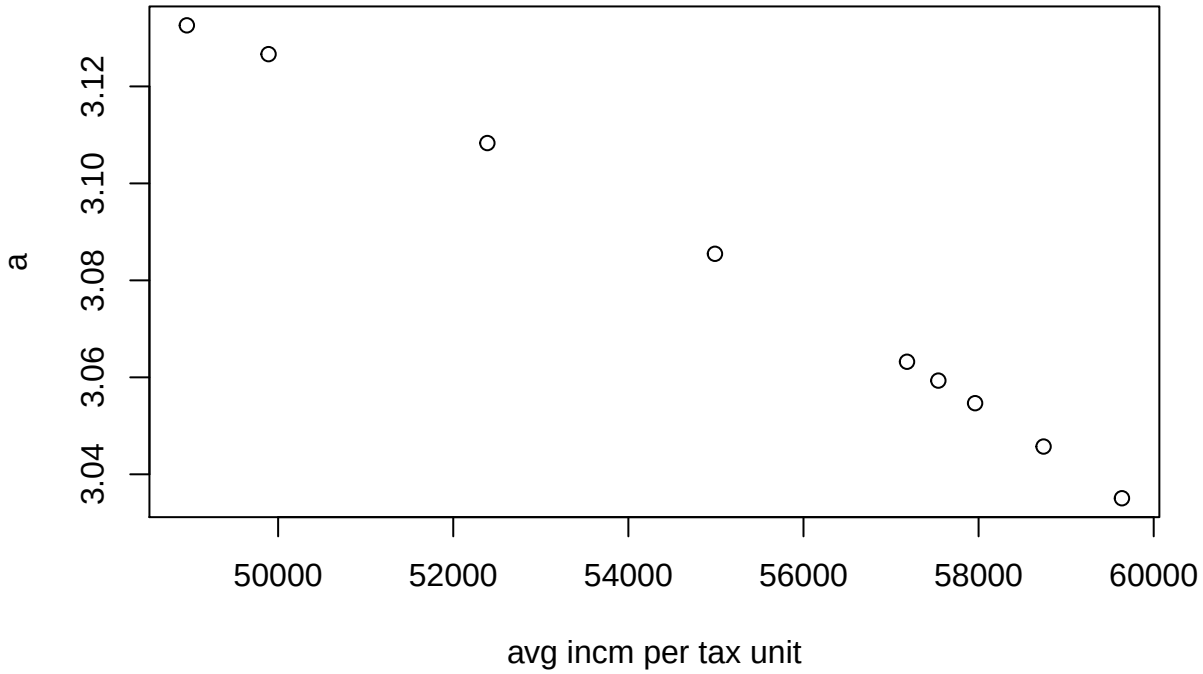

```
coeffun(Japan_report)
```



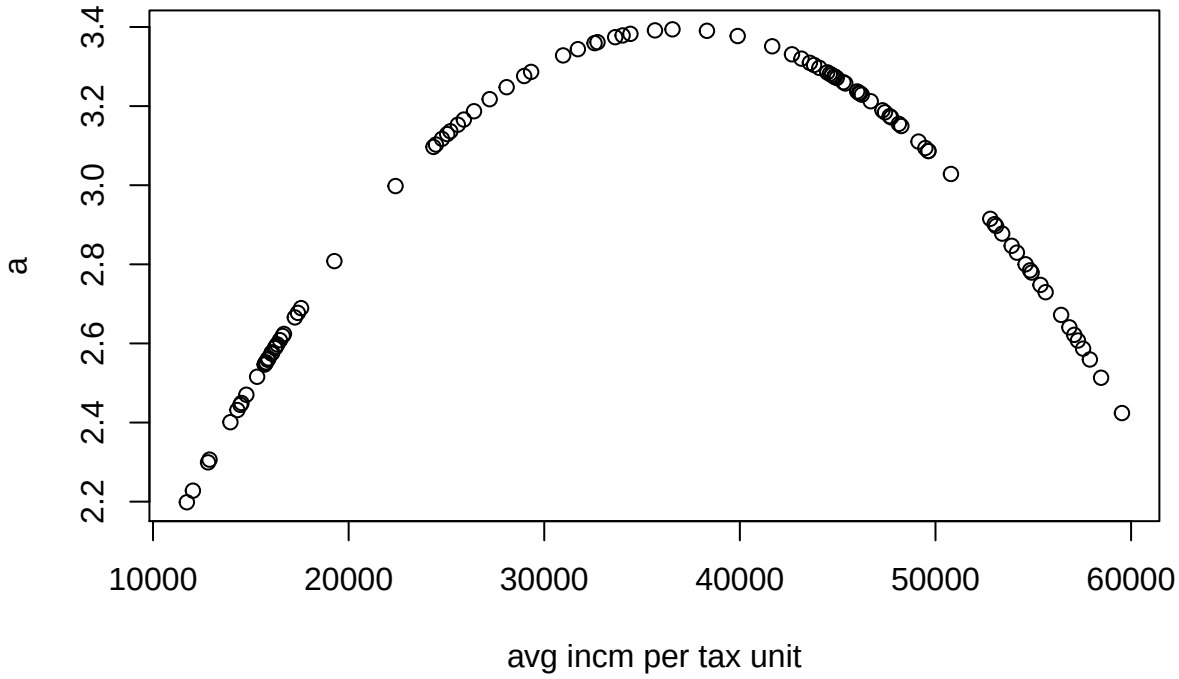

```
coeffun(Sweden_report)
```




```
coeffun(South_Africa_report)
```




```
coeffun(USA_report)
```




```
#based on the plots, Canada,USA,Italy and Sweden compatible with the hypothesis.
```