HW5_hw2570

Hongbo Wang hw2570

October 29, 2016

Part 1: Data for the US

i.

```r
rm(list = ls())
library(ggplot2)
Data <- read.csv("~/Desktop/R data/wtid-report.csv", header = TRUE)
colnames(Data) <- c("Country", "Year", "P99", "P99.5", "P99.9")
percentile_ratio_discrepancies <- function(a, P99, P99.5, P99.9){
  p_r_d <- ((((P99 / P99.9) ^ (1 - a)) - 10) ^ 2) +
    ((((P99.5 / P99.9) ^ (1 - a)) - 5) ^ 2) +
    ((((P99 / P99.5) ^ (1 - a)) - 2) ^ 2)
  return(p_r_d)
}
percentile_ratio_discrepancies(2, 1e6, 2e6, 1e7)
```

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

ii.

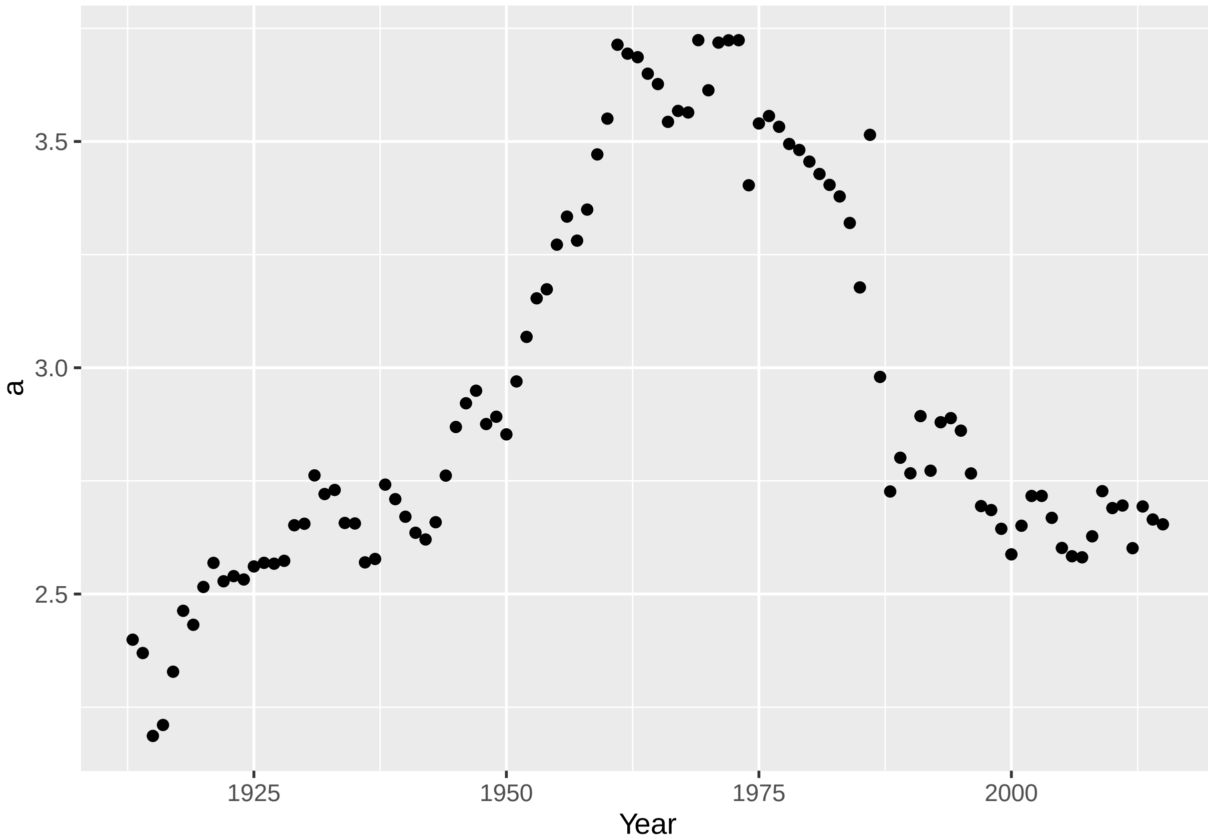
```r
exponent_multi_ratios_est <- function(P99, P99.5, P99.9){
  start_a <- 1 - (log(10) / log(P99 / P99.9))
  min_a <- nlm(percentile_ratio_discrepancies, start_a,
               P99 = P99, P99.5 = P99.5, P99.9 = P99.9)$estimate
  return(min_a)
}
exponent_multi_ratios_est(1e6, 2e6, 1e7)
```

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

iii.

```r
est_a <- function(P99, P99.5, P99.9){
  est_a = c()
  for(i in 1:length(P99)){
    est_a[i] <- exponent_multi_ratios_est(P99[i], P99.5[i], P99.9[i])
  }
  return(est_a)
}
a <- est_a(Data$P99, Data$P99.5, Data$P99.9)
```

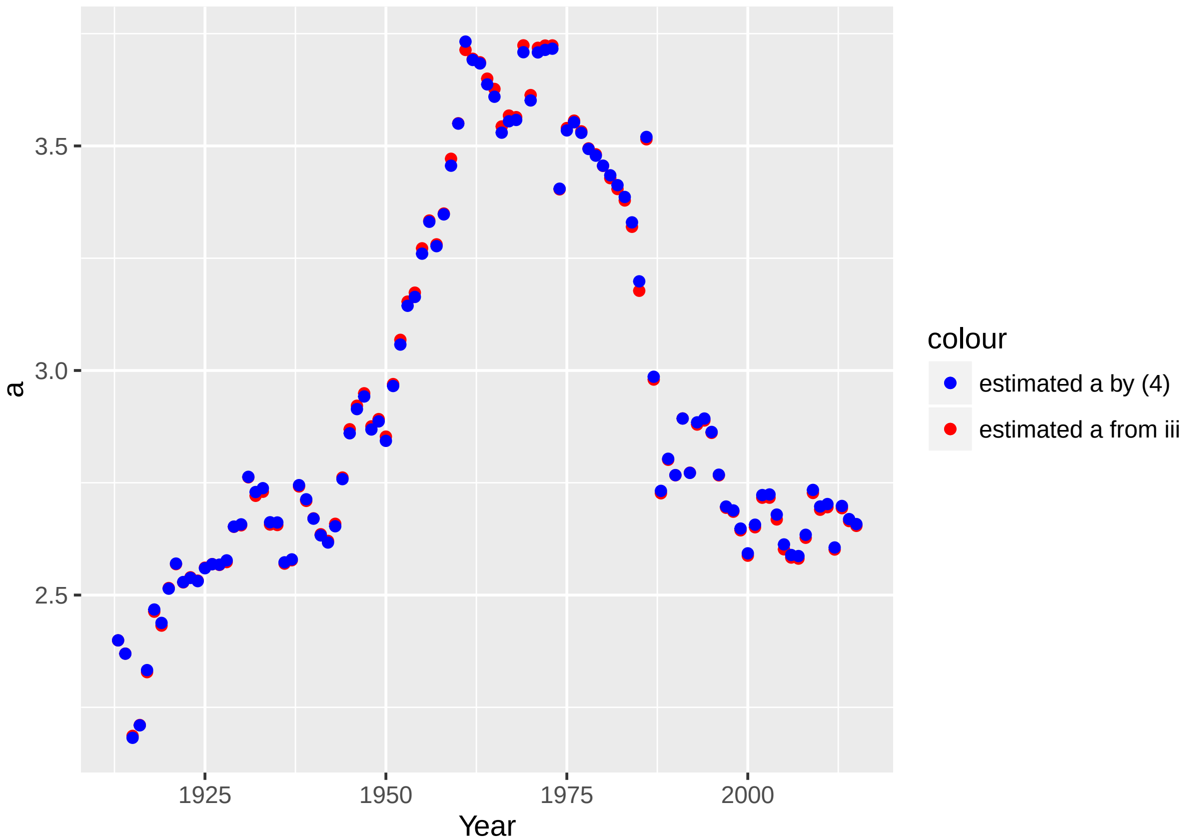
```r
ggplot(data = Data, aes(x = Year)) +
  geom_point(mapping = aes(y = a))
```



iv.

```r
a_2 <- 1 - (log(10) / log(Data$P99 / Data$P99.9))

ggplot(data = Data, aes(x = Year)) +
  geom_point(mapping = aes(y = a, col = "estimated a from iii")) +
  geom_point(mapping = aes(y = a_2, col = "estimated a by (4)")) +
  scale_color_manual(values = c("blue", "red"))
```



We can see some estimated values have small differences, but most of the estimated values by these two method are almost same.

Part 2: Data for Other Countries

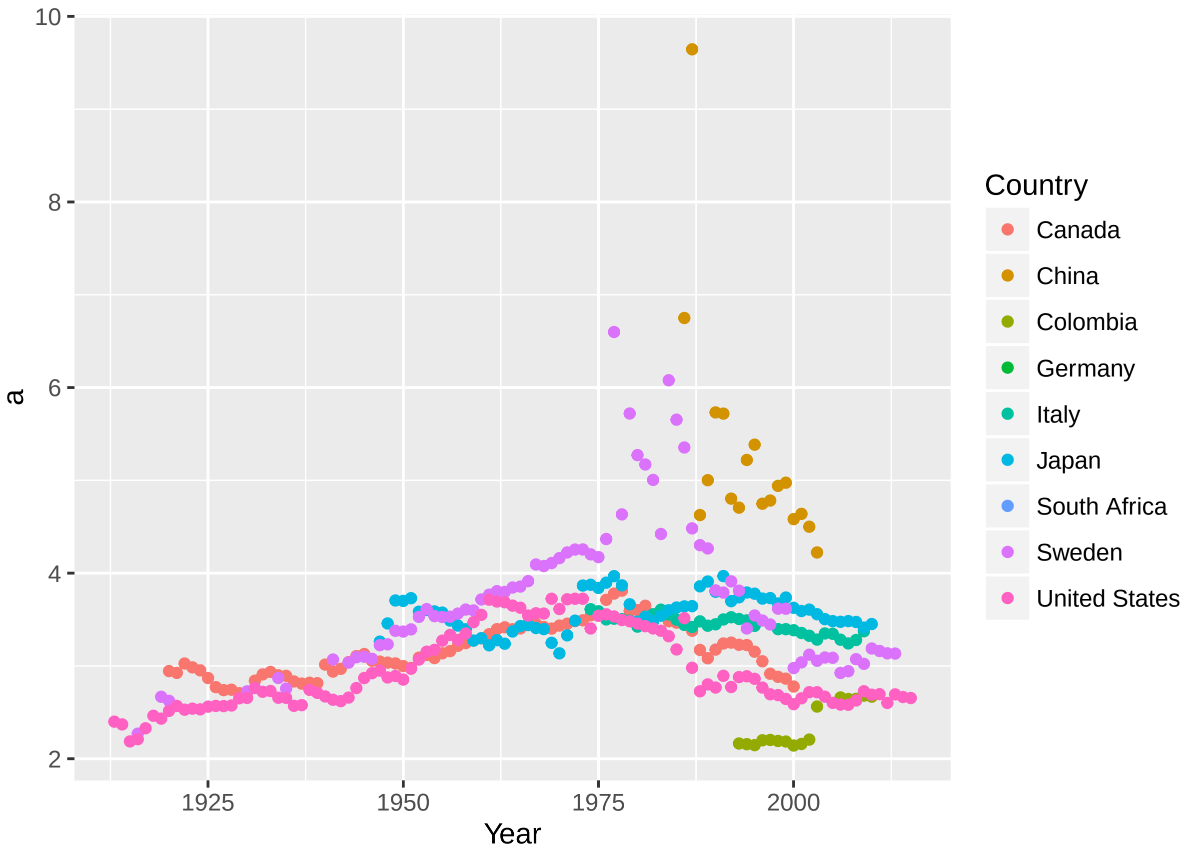
v.

```
Data <- read.csv("~/Desktop/R data/WTIDreport.csv", header = TRUE)
Data <- Data[, c("Country", "Year", "P99.income.threshold",
                 "P99.5.income.threshold", "P99.9.income.threshold",
                 "Average.income.per.tax.unit" )]
colnames(Data) <- c("Country", "Year", "P99", "P99.5", "P99.9", "Ave_Inc_PY")

a_all_Country <- c()
for(i in 1:dim(Data)[1]){
  if(is.na(Data$P99[i]) | is.na(Data$P99.5[i]) | is.na(Data$P99.9[i])){
    a_all_Country[i] <- NA
  }
  else{
    a_all_Country[i] <- est_a(Data$P99[i], Data$P99.5[i], Data$P99.9[i])
  }
}
Data$a <- a_all_Country
```

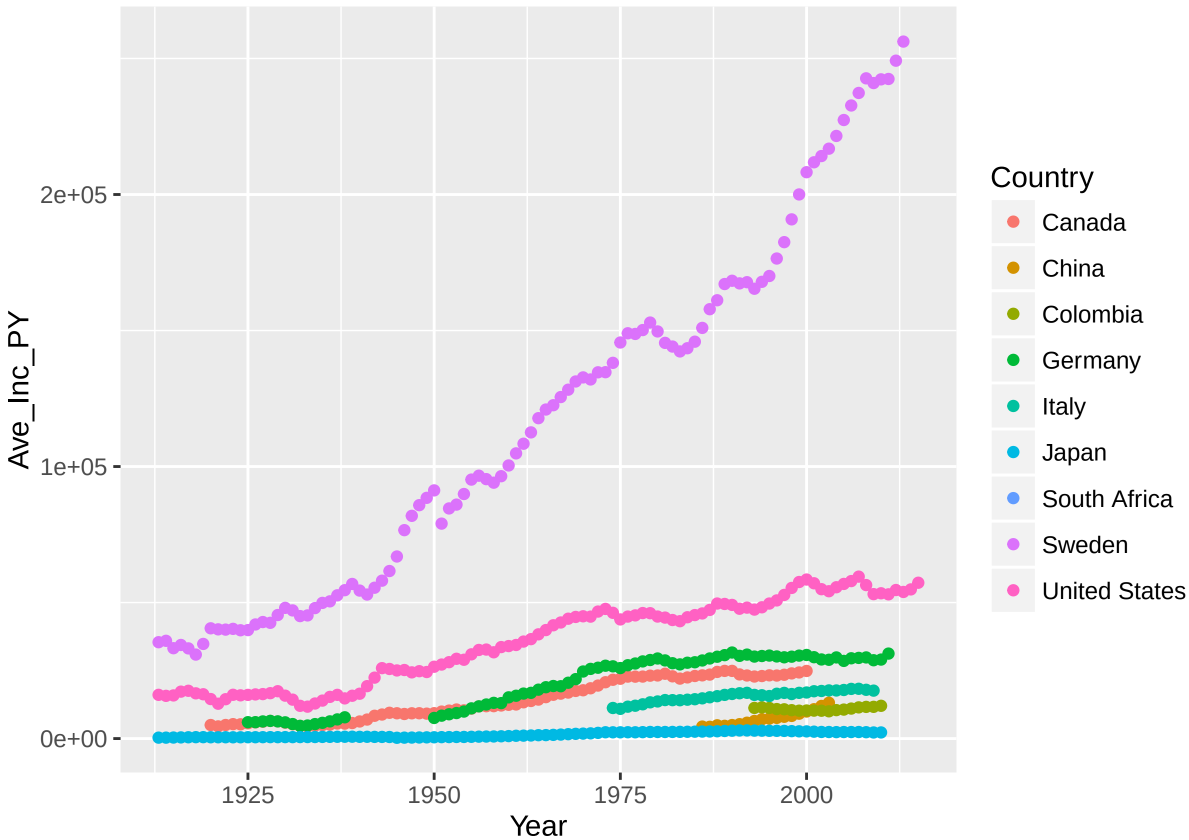
vi.

```r
ggplot(data = Data, aes(x = Year)) +
  geom_point(mapping = aes(y = a, color = Country), na.rm = TRUE)
```



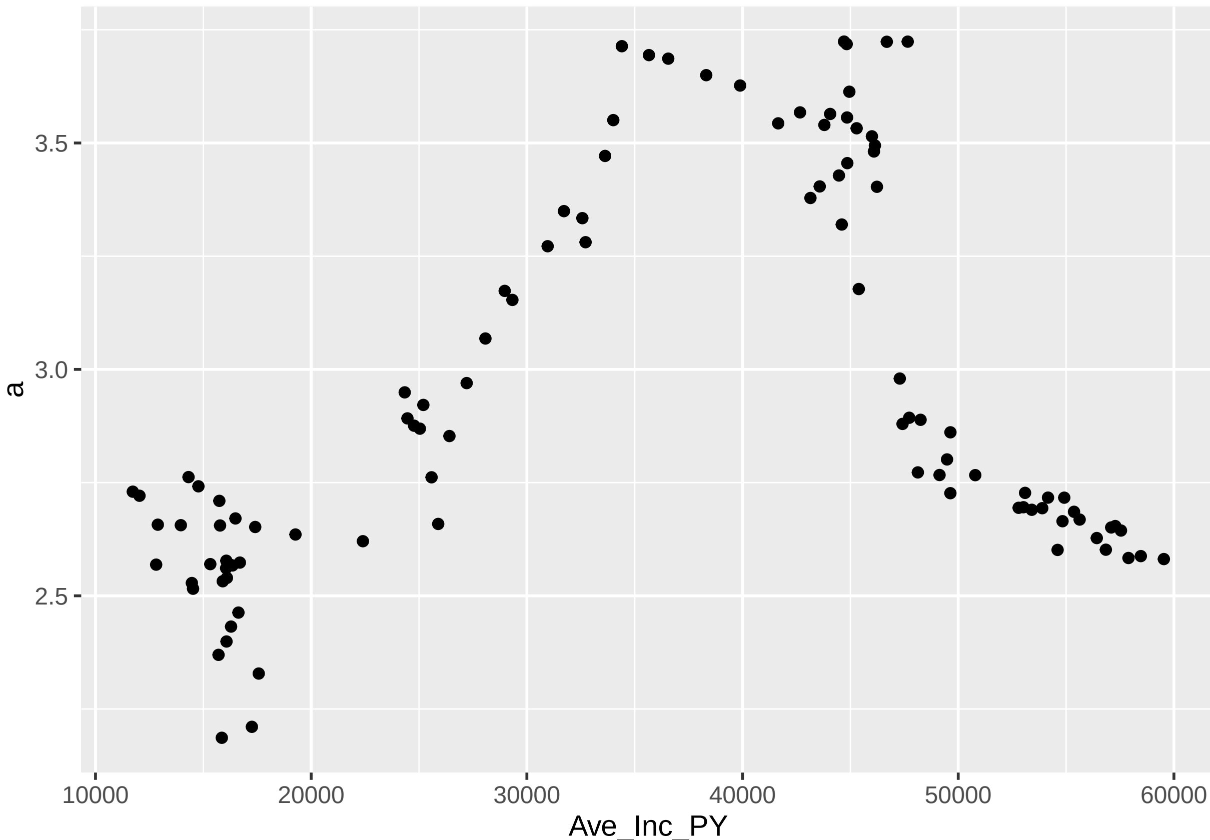
vii.

```r
ggplot(data = Data, aes(x = Year)) +
  geom_point(mapping = aes(y = Ave_Inc_PY, color = Country), na.rm = TRUE)
```



viii.

```
Data_US <- Data[Data$Country == "United States", ]
ggplot(data = Data_US) +
  geom_point(mapping = aes(x = Ave_Inc_PY, y = a))
```



In our setting, low "a" values imply more inequality. Hence the Kuznets should produce and an "upside-down U shape". Since the plot among estimated exponents against the average income in US is significantly an "upside-down U shape" curve, we can say the hypothesis is support by our data.

ix.

```r
fit_KH <- lm(a ~ Ave_Inc_PY + I(Ave_Inc_PY^2), data = Data_US)
fit_KH
```

```
##
## Call:
## lm(formula = a ~ Ave_Inc_PY + I(Ave_Inc_PY^2), data = Data_US)
##
## Coefficients:
##     (Intercept)       Ave_Inc_PY  I(Ave_Inc_PY^2)
##       8.230e-01        1.394e-04       -1.891e-09
```

In our setting, low "a" values imply more inequality. Hence the Kuznets should produce and an "upside-down U shape". So we just need to look at the sign of the quadratic term. If the sign is negative, it does support the hypothesis. In this question, the quadratic term is $-1.891e-09$, so it does support the hypothesis.

x.

```r
Data <- na.omit(Data)
lm_K <- function(x, data = Data, na.rm = TRUE){
  fit <- lm(data[data$Country == x, "a"] ~
              data[data$Country == x , "Ave_Inc_PY"] +
              I(data[data$Country == x , "Ave_Inc_PY"]^2))
  return(fit)
}
Data_coeff <- c()

for(i in unique(Data$Country)){
  cat("For country", i,"\n")
  print(Data_coeff <- lm_K(i)$coefficient)
  cat(" _____________________________________________________\n")
}
```

```
## For country Canada
##                                  (Intercept)
##                                 2.266054e+00
##      data[data$Country == x, "Ave_Inc_PY"]
##                                 1.240966e-04
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                                -3.360837e-09
##  _____________________________________________________
## For country China
##                                  (Intercept)
##                                 1.039781e+01
##      data[data$Country == x, "Ave_Inc_PY"]
##                                -1.126763e-03
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                                 5.257536e-08
##  _____________________________________________________
## For country Colombia
##                                  (Intercept)
##                                 3.461240e+01
##      data[data$Country == x, "Ave_Inc_PY"]
##                                -6.095234e-03
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                                 2.867133e-07
##  _____________________________________________________
## For country Italy
##                                  (Intercept)
##                                 2.582416e+00
##      data[data$Country == x, "Ave_Inc_PY"]
##                                 1.594300e-04
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                                -6.591048e-09
##  _____________________________________________________
## For country Japan
##                                  (Intercept)
##                                 3.729107e+00
##      data[data$Country == x, "Ave_Inc_PY"]
##                                -5.136191e-04
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
```

```
##                                1.889447e-07
##  -----------------------------------------------------
## For country Sweden
##                                 (Intercept)
##                                7.411214e-01
##      data[data$Country == x, "Ave_Inc_PY"]
##                                4.772962e-05
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                               -1.603479e-10
##  -----------------------------------------------------
## For country United States
##                                 (Intercept)
##                                8.230049e-01
##      data[data$Country == x, "Ave_Inc_PY"]
##                                1.394435e-04
## I(data[data$Country == x, "Ave_Inc_PY"]^2)
##                               -1.890556e-09
##  -----------------------------------------------------
```

From the output, we can see **Canada, Italy, Sweden and United States** have a negative quadratic output, so they are compatible with the hypothesis.