

Lorenz Curve and Gini Coefficient with Piketty Data

coop711

2015년 4월 24일

Data Description

Thomas Piketty가 쓴 **21세기 자본**에는 자본과 소득의 불평등에 대한 다양한 자료가 등장한다. 특히 점유율(share)이라는 관점에서 수집한 자료들이 매우 흥미를 자아낸다. 이 자료들은 Thomas Piketty - Capital in the 21st century (<http://piketty.pse.ens.fr/en/capital21c2>) 에서 다운로드받을 수 있다. 이 자료 중에서 불평등과 관련 부분을 가공하여 R에서 쉽게 사용할 수 있도록 하였다.

작업 파일에는 데이터 뿐 아니라 간단한 구조로 `lorenz curve`를 그리고, `gini`계수를 계산하는 함수가 저장되어 있다. 먼저 자료부터 확인하자.

```
load("piketty_gini.rda")
ls()
```

```
## [1] "gini"          "legend.gini"   "lorenz"        "piketty.co"
## [5] "piketty.df"    "piketty.labor" "piketty.total" "x"
```

이 중에서 `piketty.labor` 는 노동소득의 점유율, `piketty.co` 는 자본소득의 점유율, `piketty.total` 은 총소득의 점유율을 나타낸다.

```
piketty.labor
```

```
##      x  Low Medium High Very.High
## 1 0.00 0.00   0.00 0.00      0.00
## 2 0.50 0.35   0.30 0.25      0.20
## 3 0.90 0.45   0.45 0.40      0.35
## 4 0.99 0.15   0.18 0.23      0.28
## 5 1.00 0.05   0.07 0.12      0.17
```

노동소득의 경우 불평등이 낮은 경우(1970~1980년대 스칸디나비아 국가들)는 하위 50%가 전체 노동소득 중 35%를 가져가고 상위 10%는 20%를 가져가는데 그 중에서 상위 1%가 5%를 가져가고 있다. 불평등이 심해질수록 하위 50%에게 돌아가는 몫은 줄어들고 상위 1%가 가져가는 몫이 늘어난다는 점을 알 수 있다. 불평등이 매우 심한 경우(미국 2030년 예상) 하위 50%에게 돌아가는 몫은 20%에 불과하지만 상위 10%가 가져가는 몫은 무려 45%, 그 중에서도 상위 1%가 가져가는 몫은 17%에 달할 것으로 예상하고 있다. 이는 누적분포로부터 그 구조를 더 명확하게 살필 수 있다.

```
cbind(x = piketty.labor$x, apply(piketty.labor[, 2:5], 2, cumsum))
```

```
##          x  Low Medium High Very.High
## 1 0.00 0.00    0.00 0.00    0.00
## 2 0.50 0.35    0.30 0.25    0.20
## 3 0.90 0.80    0.75 0.65    0.55
## 4 0.99 0.95    0.93 0.88    0.83
## 5 1.00 1.00    1.00 1.00    1.00
```

자본소득의 집중을 살펴보자

```
piketty.co
```

```
##          x  Low Medium Med.High High Very.High
## 1 0.00 0.00    0.0    0.00 0.00    0.00
## 2 0.50 0.25    0.1    0.05 0.05    0.05
## 3 0.90 0.45    0.4    0.35 0.25    0.05
## 4 0.99 0.20    0.3    0.35 0.35    0.40
## 5 1.00 0.10    0.2    0.25 0.35    0.50
```

자본소득의 경우 불평등이 낮은 경우(존재한 적 없는 이상국가)는 하위 50%가 전체 노동소득 중 25%를 가져가고 상위 10%는 30%를 가져가는 데 그 중에서 상위 1%가 10%를 가져가고 있다. 불평등이 심해질수록 하위 50%에게 돌아가는 몫은 줄어들고 상위 1%가 가져가는 몫이 확연히 늘어난다는 점을 알 수 있다. 불평등이 매우 심한 경우(1910년대 유럽) 하위 50%에게 돌아가는 몫은 5%에 불과하지만 상위 10%가 가져가는 몫은 90%에 달하고, 그 중에서 상위 1%가 가져가는 몫이 무려 50%에 달한다. 이 또한 누적분포로부터 그 구조를 보다 더 명확하게 살필 수 있다. 일반적으로 노동소득 상위 10%가 전체 노동소득의 25~30%를 받는 반면, 자본소득 상위 10%는 항상전체 부의 50% 이상을 소유한다.

```
cbind(x = piketty.co$x, apply(piketty.co[, 2:6], 2, cumsum))
```

```
##          x  Low Medium Med.High High Very.High
## 1 0.00 0.00    0.0    0.00 0.00    0.00
## 2 0.50 0.25    0.1    0.05 0.05    0.05
## 3 0.90 0.70    0.5    0.40 0.30    0.10
## 4 0.99 0.90    0.8    0.75 0.65    0.50
## 5 1.00 1.00    1.0    1.00 1.00    1.00
```

총소득의 집중을 살펴보자.

```
piketty.total
```

```
##          x  Low Medium High Very.High
## 1 0.00 0.00    0.00 0.0    0.00
## 2 0.50 0.30    0.25 0.2    0.15
## 3 0.90 0.45    0.40 0.3    0.25
## 4 0.99 0.18    0.25 0.3    0.35
## 5 1.00 0.07    0.10 0.2    0.25
```

총소득의 불평등이 비교적 낮은 사회(1970~1980년대의 스칸디나비아 국가들)에서는 하위 50%가 전체 총소득의 30%를 가져가고 상위 10%는 25%를 가져가는 데 그 중에서 상위 1%는 7%를 가져간다. 불평등이 매우 심한 경우(2030년 미국 예상) 하위 50%에게 돌아가는 몫은 5%에 불과하지만 상위 10%가 가져가는 몫은 60%에 달하고, 그

중에서 상위 1%가 가져가는 몫 또한 25%에 달한다. 이 또한 누적분포로부터 그 구조를 보다 더 명확하게 살필 수 있다. `x` 는 앞으로도 많이 출현하기 때문에 따로 R객체로 저장한다.

```
cbind(x = piketty$total$x, apply(piketty$total[, 2:5], 2, cumsum))
```

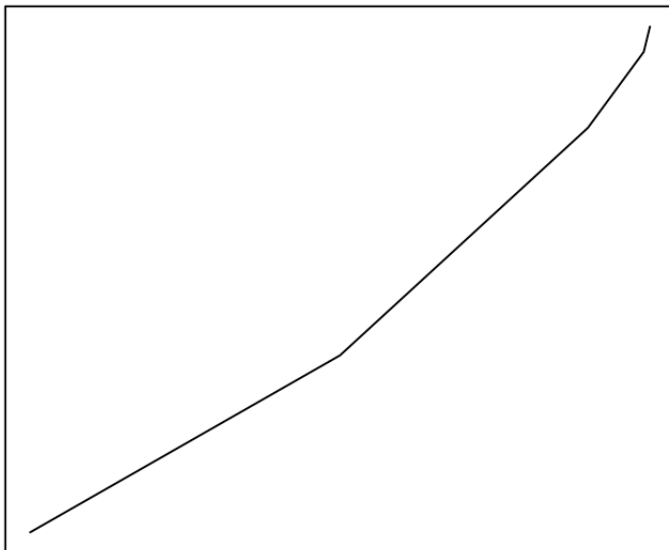
```
##      x Low Medium High Very.High
## 1 0.00 0.00   0.00  0.0      0.00
## 2 0.50 0.30   0.25  0.2      0.15
## 3 0.90 0.75   0.65  0.5      0.40
## 4 0.99 0.93   0.90  0.8      0.75
## 5 1.00 1.00   1.00  1.0      1.00
```

```
x<-piketty$total$x
```

Plots

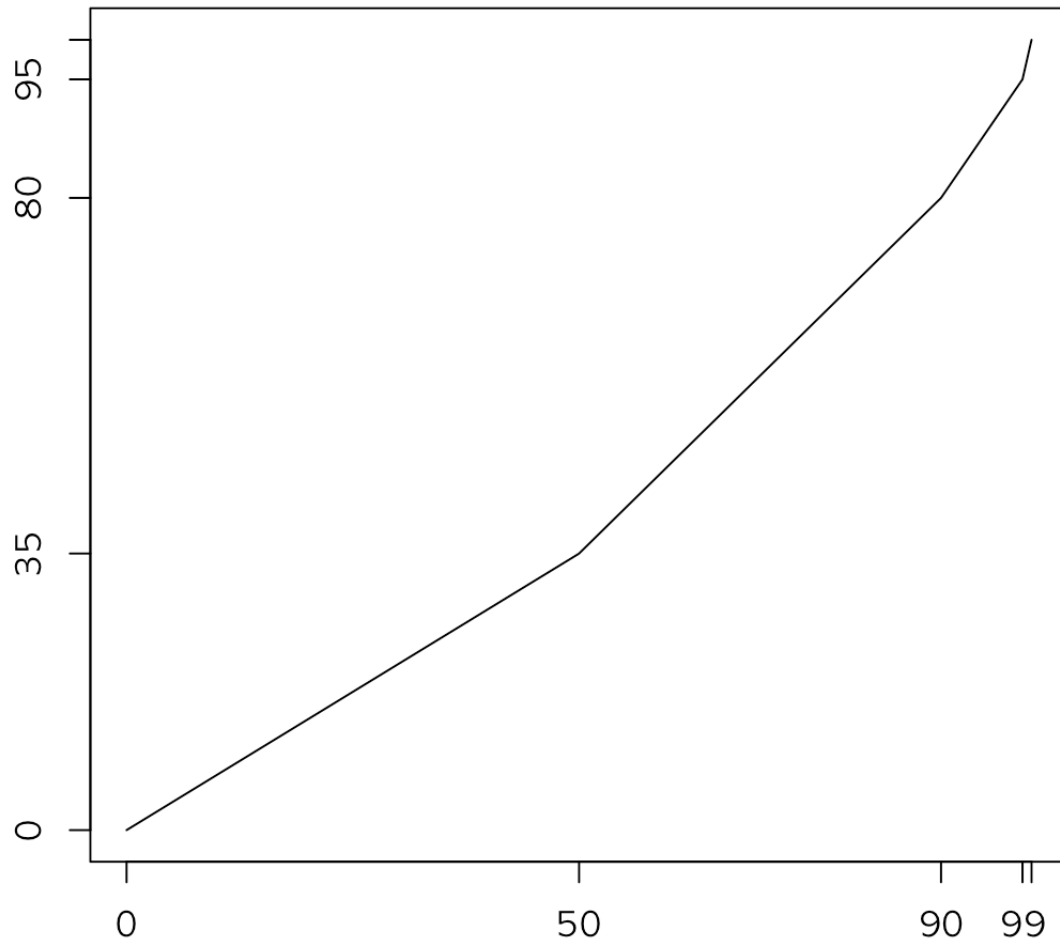
이 상황을 그래프로 옮겨보자. `x`-축에는 소득 순서대로 누적된 사람의 비율을 표시하고 `y`-축에는 그 사람들이 얼마정도 소득을 축적했는지 누적시켜간다. 따라서 첫 단계는 소득의 누적분포를 구하는 것이다. 우선 노동소득의 `Low` 열에 대하여 `plot()` 을 이용하여 누적도표를 그린다. `cumsum(piketty$labor$Low)` 등이 반복해서 나오게 되므로 별도의 R 객체로 저장한다.

```
piketty.labor.cum<-apply(piketty$labor[, 2:5], 2, cumsum)
y<-piketty.labor.cum[, 1]
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
```



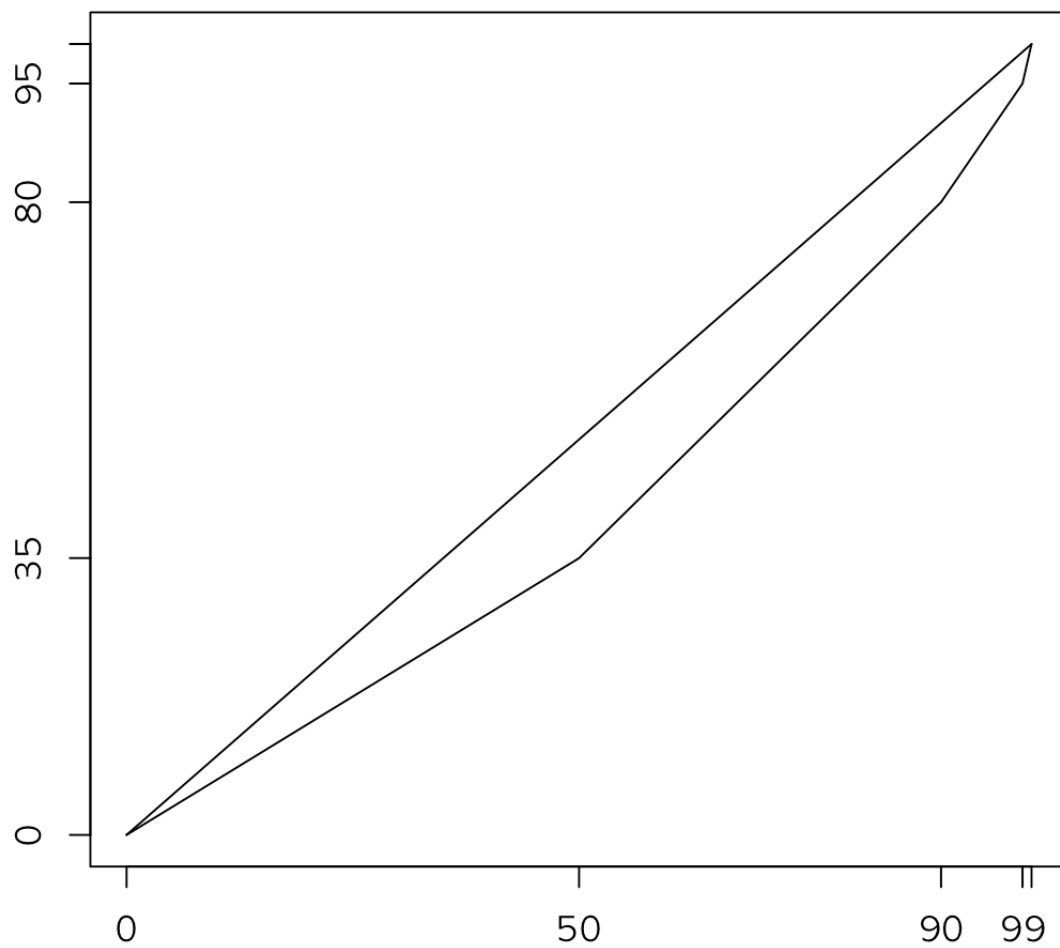
`x`축과 `y`축의 눈금을 표시한다.

```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
```



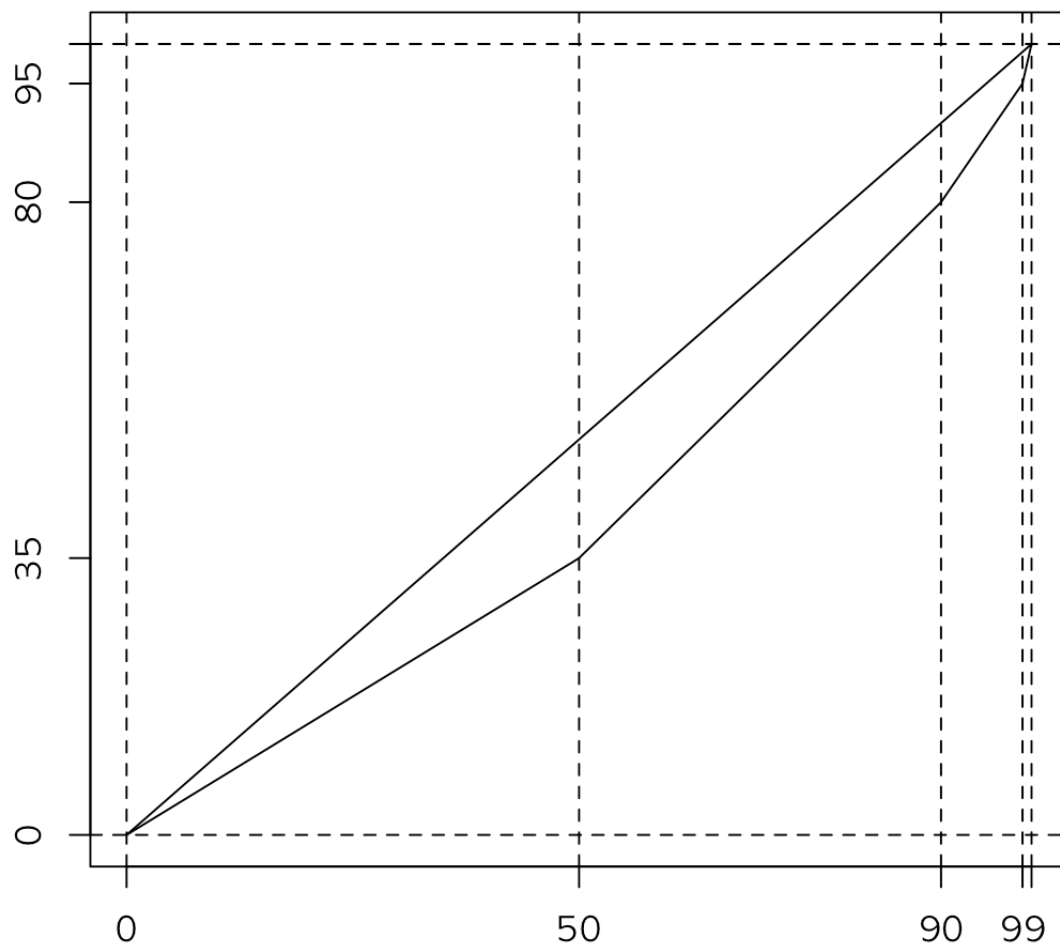
완전 평등선이라고 할 수 있는 $y = x$ 를 같은 영역에 그린다.

```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
```



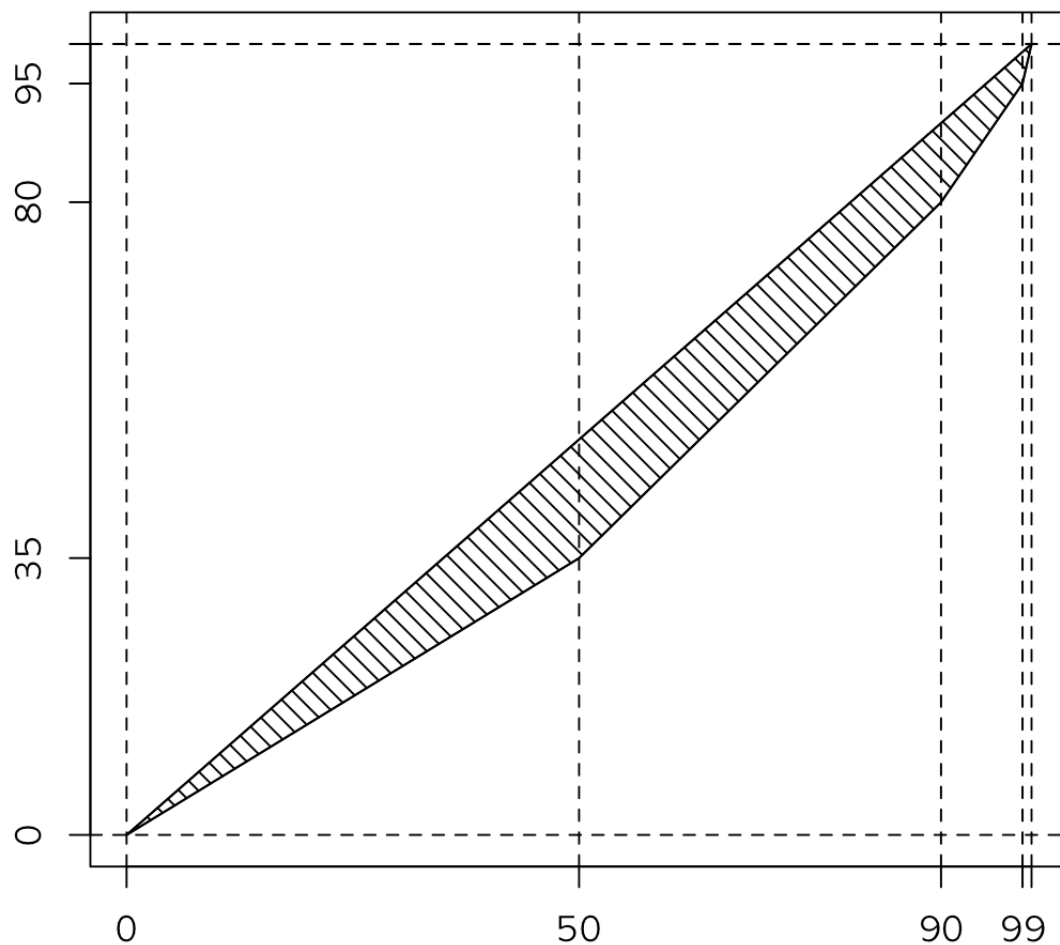
$x = 0, 0.5, 0.9, 0.99, 1$ 과 y 축의 아래 위에 점선으로 된 격자를 설치하여 윤곽을 쉽게 파악하도록 한다.

```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
```



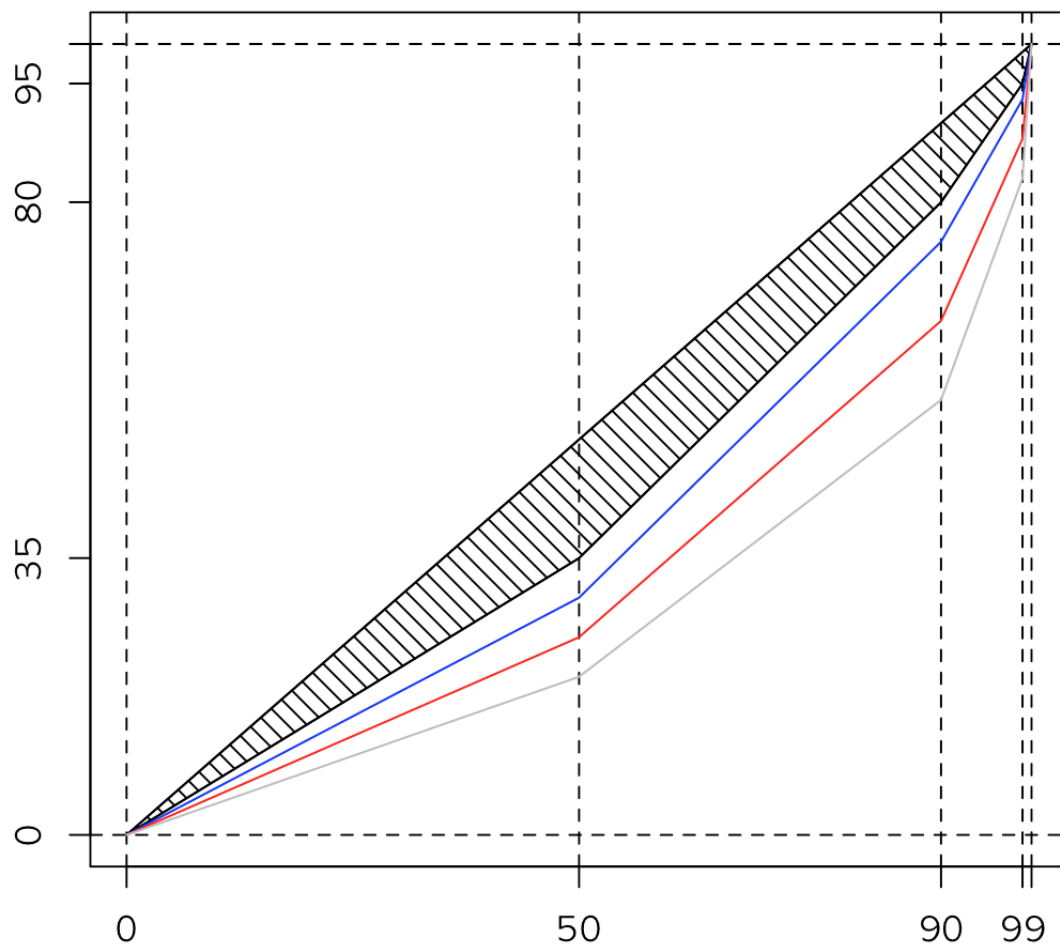
`polygon()` 을 이용하여 평등선과 소득누적분포의 차이를 빗금으로 표시한다.

```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
```



`lines()` 를 이용하여 나머지 소득 누적곡선을 그린다.

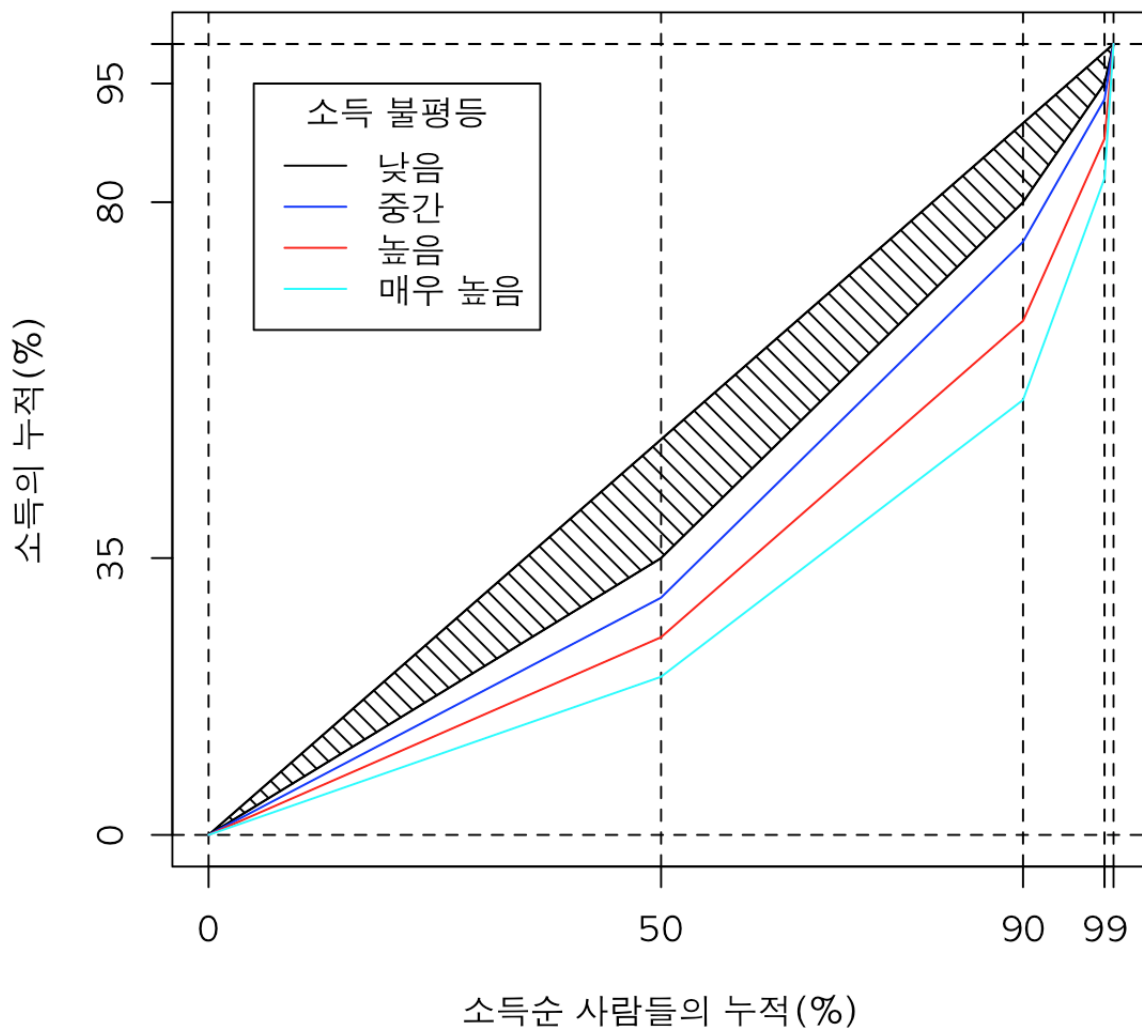
```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
lines(x = x, y = piketty.labor.cum[ , 2], col="blue")
lines(x = x, y = piketty.labor.cum[ , 3], col="red")
lines(x = x, y = piketty.labor.cum[ , 4], col="grey")
```



legend() 와 title() 로 마무리짓는다.

```
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
lines(x = x, y = piketty.labor.cum[ , 2], col="blue")
lines(x = x, y = piketty.labor.cum[ , 3], col="red")
lines(x = x, y = piketty.labor.cum[ , 4], col="cyan")
title(main = "노동 소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누
적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "높음", "매우 높음"), lty = 1, co
l = c("black", "blue", "red", "cyan"), title="소득 불평등")
```


노동 소득의 로렌츠 곡선



지금까지의 과정을 잘 살펴보면 소득순 사람들의 누적분포와 이에 따른 소득의 누적분포를 각각 x 와 y 로 입력하면 로렌츠 곡선을 그리고 평등선과의 차이를 빗금으로 표시할 수도 있다는 것을 알 수 있다. 이를 함수로 나타낸 것이 `lorenz()` 이다. 코드를 살펴보면,

`lorenz`

```
## function (x,y)
## {
##     plot(x = x, y = y, type = "l", ann = FALSE, xaxt = "n", yaxt = "n")
##     axis(side = 1, at = x, labels = paste(x*100))
##     axis(side = 2, at = y, labels = paste(y*100))
##     lines(x = x, y = x)
##     abline(v = x, h = c(0, 1), lty = 2)
##     polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
## }
```

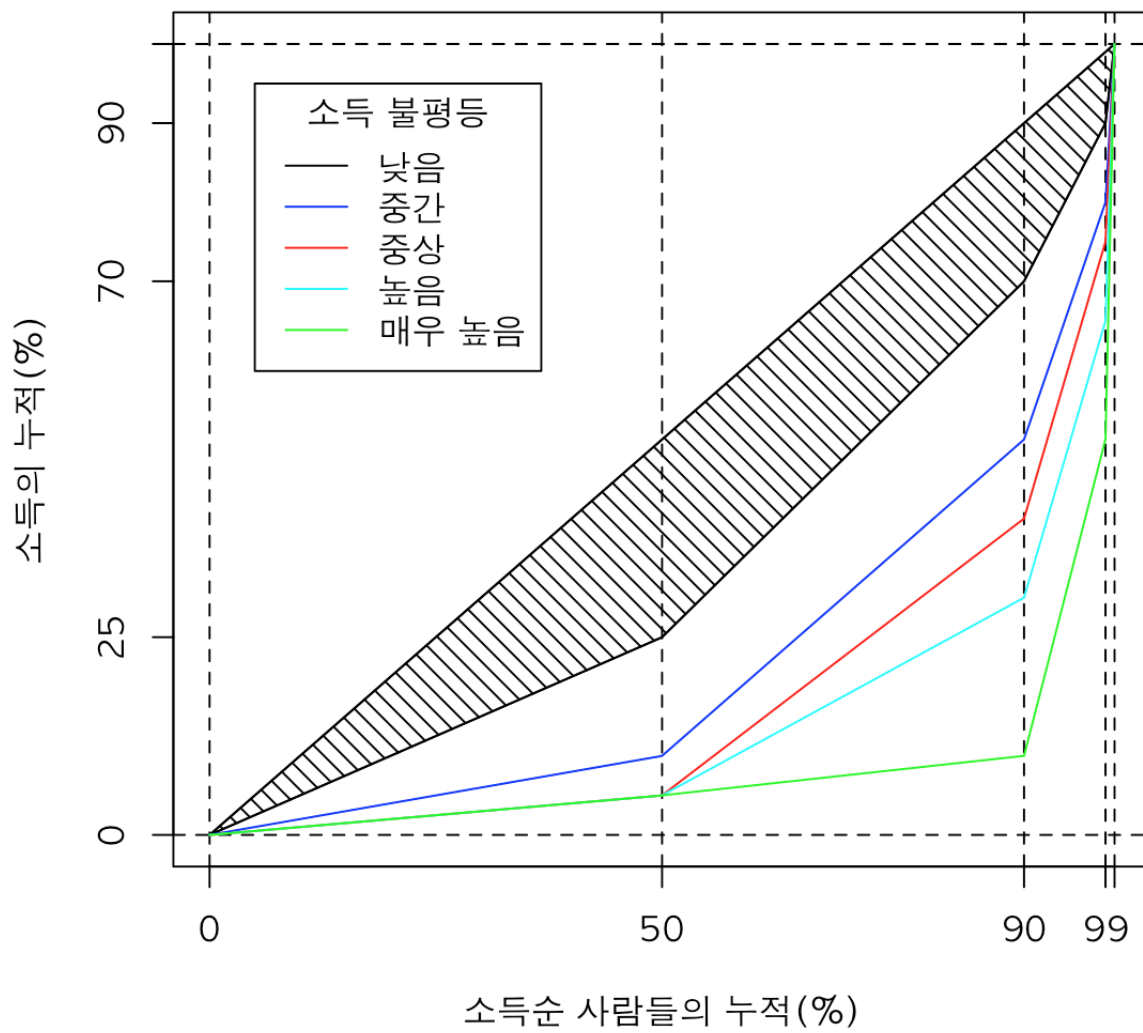
이제 `lorenz()` 함수를 이용하여 자본소득의 로렌츠 곡선들을 그려보자.

```
piketty.co.cum<-apply(piketty.co[, 2:6], 2, cumsum)
piketty.co.cum
```

```
##      Low Medium Med.High High Very.High
## 1 0.00    0.0    0.00 0.00    0.00
## 2 0.25    0.1    0.05 0.05    0.05
## 3 0.70    0.5    0.40 0.30    0.10
## 4 0.90    0.8    0.75 0.65    0.50
## 5 1.00    1.0    1.00 1.00    1.00
```

```
y<-piketty.co.cum[, 1]
lorenz(x, y)
lines(x = x, y = piketty.co.cum[, 2], col="blue")
lines(x = x, y = piketty.co.cum[, 3], col="red")
lines(x = x, y = piketty.co.cum[, 4], col="cyan")
lines(x = x, y = piketty.co.cum[, 5], col="green")
title(main = "자본 소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누
적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "중상", "높음", "매우 높음"), lty
= 1, col = c("black", "blue", "red", "cyan", "green"), title="소득 불평등")
```

자본 소득의 로렌츠 곡선



총소득의 로렌츠 곡선들도 같은 방법으로 그릴 수 있다.

```
piketty.total.cum<-apply(piketty.total[, 2:5], 2, cumsum)
piketty.total.cum
```

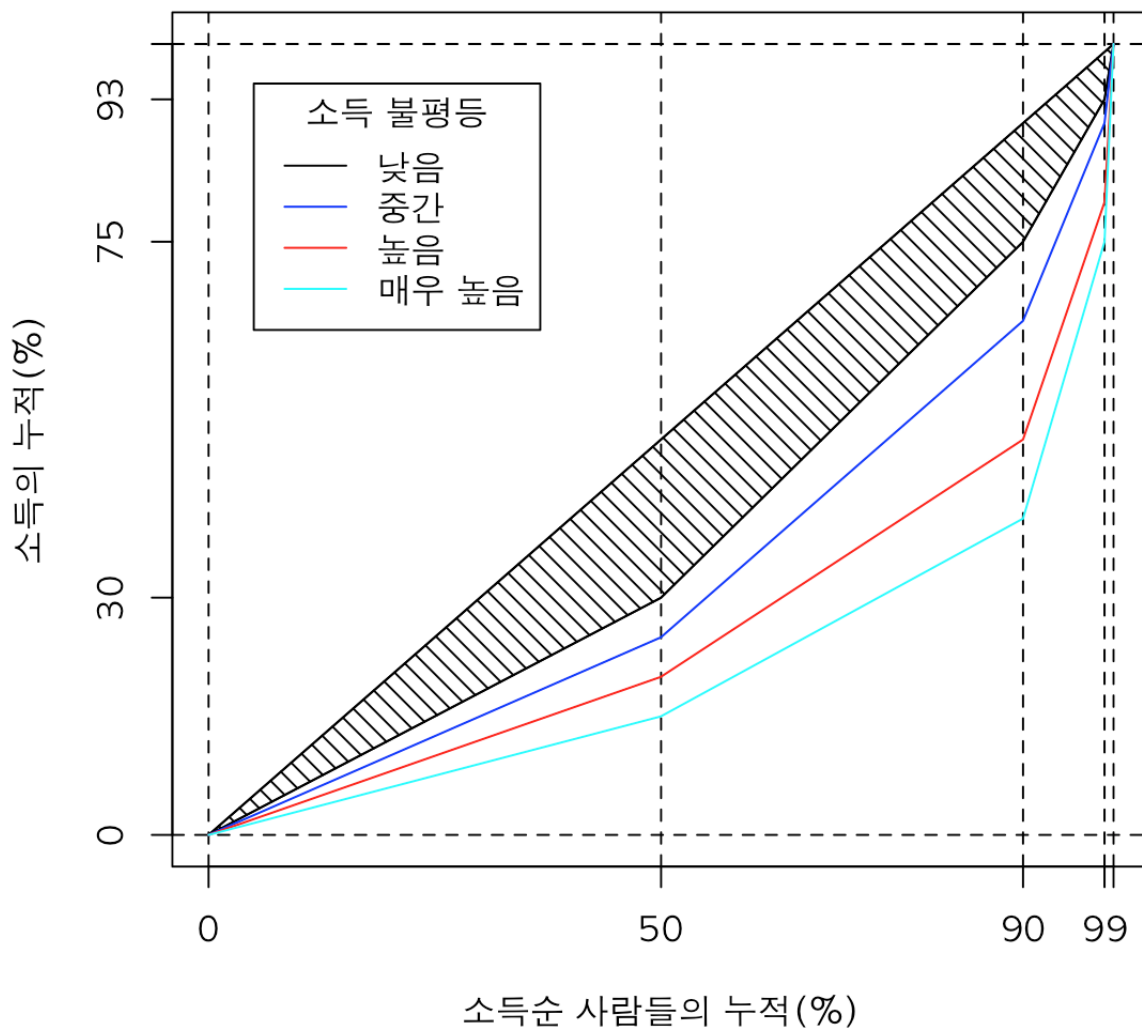
##	Low	Medium	High	Very.High
## 1	0.00	0.00	0.0	0.00
## 2	0.30	0.25	0.2	0.15
## 3	0.75	0.65	0.5	0.40
## 4	0.93	0.90	0.8	0.75
## 5	1.00	1.00	1.0	1.00

```

y<-piketty.total.cum[, 1]
lorenz(x, y)
lines(x = x, y = piketty.total.cum[ , 2], col="blue")
lines(x = x, y = piketty.total.cum[ , 3], col="red")
lines(x = x, y = piketty.total.cum[ , 4], col="cyan")
title(main = "총소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누
적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "높음", "매우 높음"), lty = 1, co
l = c("black", "blue", "red", "cyan"), title="소득 불평등")

```

총소득의 로렌츠 곡선

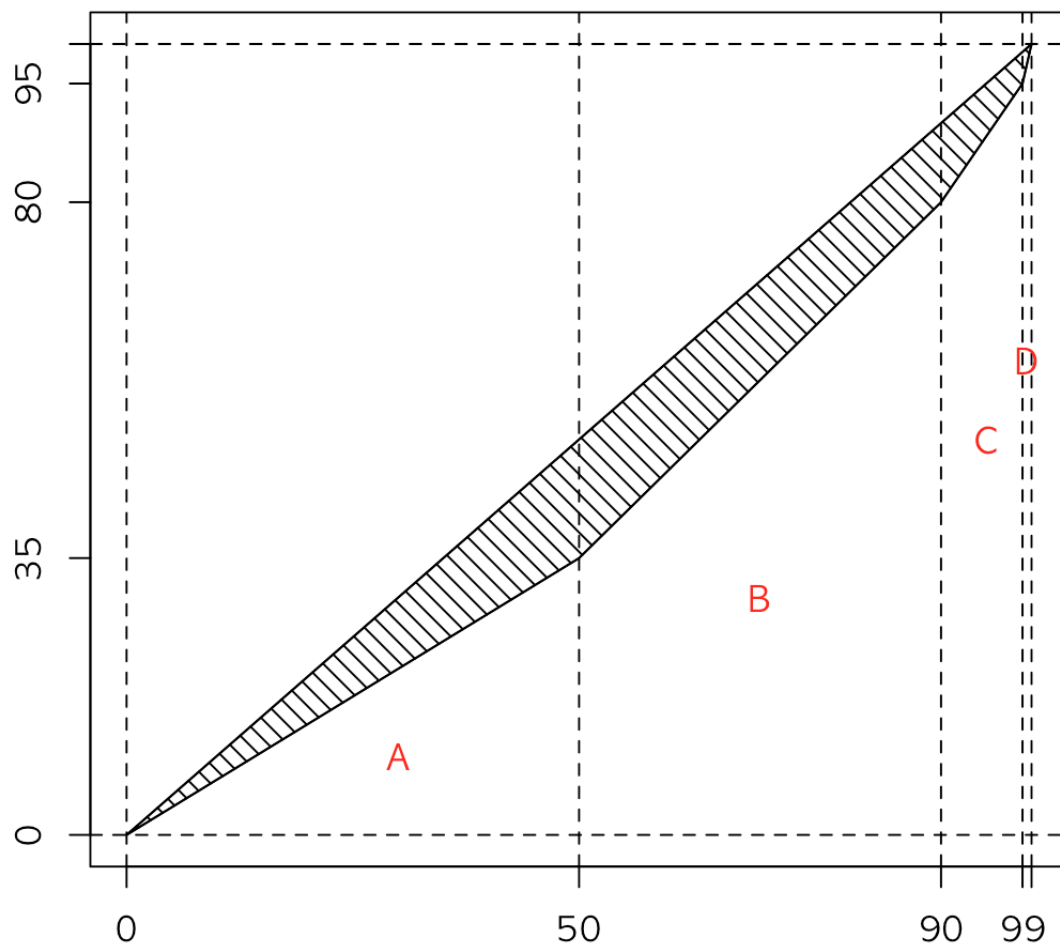


평등선과 소득누적곡선과의 차이를 나타내는 gini계수는 결국 면적 계산임을 알 수 있다. gini계수는 $\frac{\text{빈금친면적}}{1/2} = 2 \times \text{빈금친면적}$ 으로 주어지기 때문에 각 삼각형과 사다리꼴들의 면적을 계산하면서 2로 나누어 줄 필요가 없게 된다.

```

lorenz(x = x, y = cumsum(piketty.labor$Low))
text(x=c(0.3, 0.7, 0.95, 0.995), y=c(0.1, 0.3, 0.5, 0.6), labels=c("A", "B",
"C", "D"), col="red")

```



A, B, C, D의 면적을 합해서 2배를 해 주면 되기 때문에 2로 나눌 필요 없이 A면적의 2배는

$(y[1] + y[2]) * (x[2] - x[1])$, B면적의 2배는 $(y[2] + y[3]) * (x[3] - x[2])$, C면적의 2배는 $(y[3] + y[4]) * (x[4] - x[3])$, D면적의 2배는 $(y[4] + y[5]) * (x[5] - x[4])$ 로 주어진다.

$x[1]$ 과 $y[1]$ 이 모두 0이지만 규칙성을 파악하기 쉽도록 집어넣었다. 따라서 x -벡터의 길이보다 하나 적은 갯수의 사다리꼴들의 합을 구하는 문제로 귀착된다. 이 때 i -번째 사다리꼴 면적의 2배는

$(y[i] + y[i + 1]) * (x[i + 1] - x[i])$ 로 주어지게 된다. 이들의 합을 구해서 1에서 빼주면 된다.

따라서 다음 코드가 나온다.

```
gini
```

```
## function(x,y)
## {
##     n<-length(x)
##     i<-1:(n - 1)
##     s<-sum((y[i] + y[i + 1])*(x[i + 1] - x[i]))
##     s<-1 - s
##     return(s)
## }
```

`gini()` 함수를 이용하여 노동소득, 자본소득, 총소득에 나타난 각 경우에 대한 Gini계수는 다음과 같이 구할 수 있다. 노동소득의 각 경우에 지니계수들은,

```
apply(piketty.labor.cum, 2, gini, x = x)
```

##	Low	Medium	High	Very.High
##	0.1880	0.2595	0.3585	0.4575

로 계산되고, 자본소득의 각 경우에 지니계수들은,

```
apply(piketty.co.cum, 2, gini, x = x)
```

##	Low	Medium	Med.High	High	Very.High
##	0.332	0.575	0.674	0.733	0.846

이고, 총소득의 각 경우에 지니계수들은,

```
apply(piketty.total.cum, 2, gini, x = x)
```

##	Low	Medium	High	Very.High
##	0.2595	0.3565	0.4850	0.5840

로 계산되어 자본소득의 경우 불평등이 매우 심한 경우에는 지니계수가 0.8을 넘기도 한다는 것을 알 수 있다.

뒷 마무리