

Korea Income Distribution

minyoung-choi

Data

```
(income_kr <- read.table("../data/labor_income_kor.txt",
                          header = TRUE,
#                          encoding = "UTF-8",
                          row.names = 1))
```

```
##           Earners... Income...
## 0-5           19.1         1.7
## 5-10          12.3         3.6
## 10-20         22.8        12.8
## 20-30         14.4        13.6
## 30-40          9.8        13.0
## 40-60         12.0        22.5
## 60-80          5.4        14.1
## 80-100         2.3         7.8
## 100-200        1.6         7.7
## 200-300        0.1         1.2
## 300-500        0.1         0.8
## 500-1000       0.0         0.6
## 1000-         0.0         0.6
```

```
str(income_kr)
```

```
## 'data.frame':   13 obs. of  2 variables:
## $ Earners...: num  19.1 12.3 22.8 14.4 9.8 12 5.4 2.3 1.6 0.1 ...
## $ Income... : num  1.7 3.6 12.8 13.6 13 22.5 14.1 7.8 7.7 1.2 ...
```

```
names(income_kr) <- c("Earners(%)", "Income(%)")
income_kr
```

```
##           Earners(%) Income(%)
## 0-5           19.1         1.7
## 5-10          12.3         3.6
## 10-20         22.8        12.8
## 20-30         14.4        13.6
## 30-40          9.8        13.0
## 40-60         12.0        22.5
## 60-80          5.4        14.1
## 80-100         2.3         7.8
## 100-200        1.6         7.7
## 200-300        0.1         1.2
## 300-500        0.1         0.8
## 500-1000       0.0         0.6
## 1000-         0.0         0.6
```

```
rownames(income_kr) <- sub(pattern = "-",
                           replacement = " - ",
                           x = rownames(income_kr))

# kable(income_kr)
```

```
(r_names_split <- strsplit(rownames(income_kr),
                           split = " - "))
```

```
r_names_split[1]
r_names_split[1][[1]]
r_names_split[[1]]
r_names_split[[1]][1]
`[(r_names_split, 1)
`[[r_names_split, 1)
```

```
# (r_names_split_first <- sapply(r_names_split, function(x){x[1]}))
(r_names_split_first <- sapply(r_names_split,
                               FUN = `[(, 1))
```

```
## [1] "0"    "5"    "10"   "20"   "30"   "40"   "60"   "80"   "100"  "200"
## [11] "300"  "500"  "1000"
```

```
(income_breaks <- as.numeric(r_names_split_first))
```

```
## [1] 0 5 10 20 30 40 60 80 100 200 300 500 1000
```

```
(income_breaks <- c(income_breaks, 2000))
```

```
## [1] 0 5 10 20 30 40 60 80 100 200 300 500 1000 2000
```

```
(income_widths <- diff(income_breaks))
```

```
## [1] 5 5 10 10 10 20 20 20 100 100 200 500 1000
```

```
options(digits = 3)
(height_earners <- income_kr[, 1]/income_widths)
(height_earners_2 <- income_kr[, "Earners(%)"]/income_widths)
(height_earners_3 <- income_kr[[1]]/income_widths)
(height_earners_4 <- income_kr[1]/income_widths)
(height_earners_5 <- income_kr["Earners(%)"]/income_widths)
```

Probability Histogram with barplot()

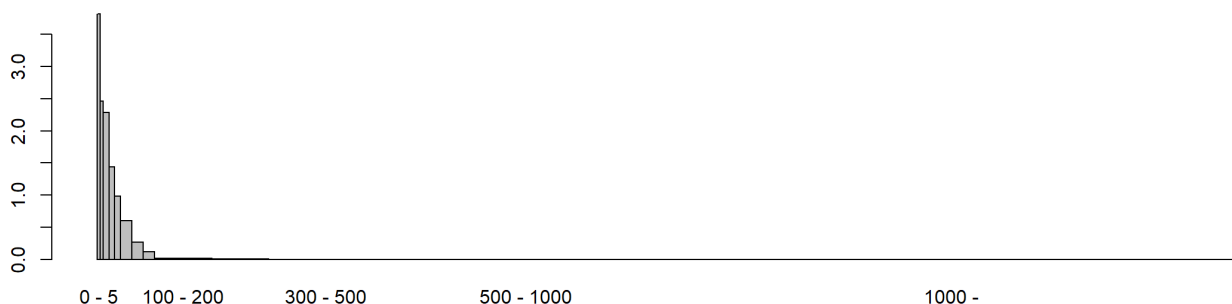
```
barplot(height_earners,
        width = income_widths)
```



```
(names_bar <- rownames(income_kr))
```

```
## [1] "0 - 5"      "5 - 10"     "10 - 20"    "20 - 30"    "30 - 40"
## [6] "40 - 60"    "60 - 80"    "80 - 100"   "100 - 200"  "200 - 300"
## [11] "300 - 500"  "500 - 1000" "1000 - "
```

```
barplot(height_earners,
        width = income_widths,
        space = 0,
        names.arg = names_bar)
```



```
income_kr_2 <- income_kr[1:11, ]
income_kr_2[11, ] <- apply(income_kr[11:13, ],
                           MARGIN = 2,
                           FUN = sum)

income_kr_2
```

##	Earners(%)	Income(%)
## 0 - 5	19.1	1.7
## 5 - 10	12.3	3.6
## 10 - 20	22.8	12.8
## 20 - 30	14.4	13.6
## 30 - 40	9.8	13.0
## 40 - 60	12.0	22.5
## 60 - 80	5.4	14.1
## 80 - 100	2.3	7.8
## 100 - 200	1.6	7.7
## 200 - 300	0.1	1.2
## 300 - 500	0.1	2.0

```
rownames(income_kr_2)
```

```
## [1] "0 - 5"      "5 - 10"      "10 - 20"     "20 - 30"     "30 - 40"     "40 - 60"
## [7] "60 - 80"     "80 - 100"    "100 - 200"   "200 - 300"   "300 - 500"
```

```
rownames(income_kr_2)[11] <- "300 - "
income_kr_2
```

##	Earners(%)	Income(%)
## 0 - 5	19.1	1.7
## 5 - 10	12.3	3.6
## 10 - 20	22.8	12.8
## 20 - 30	14.4	13.6
## 30 - 40	9.8	13.0
## 40 - 60	12.0	22.5
## 60 - 80	5.4	14.1
## 80 - 100	2.3	7.8
## 100 - 200	1.6	7.7
## 200 - 300	0.1	1.2
## 300 -	0.1	2.0

```
(income_breaks_2 <- income_breaks[1:12])
```

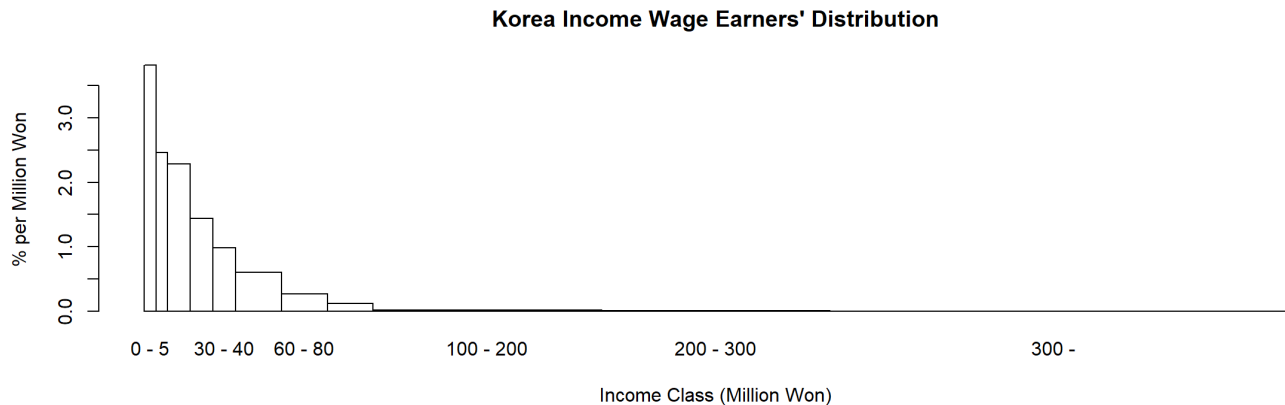
```
## [1] 0 5 10 20 30 40 60 80 100 200 300 500
```

```
income_widths_2 <- diff(income_breaks_2)
height_earners_2 <- income_kr_2[, 1]/income_widths_2
names_bar_2 <- rownames(income_kr_2)
```

```

title_1 <- "Korea Income Wage Earners' Distribution"
xlab_1 <- "Income Class (Million Won)"
ylab_1 <- "% per Million Won"
barplot(height_earners_2,
        width = income_widths_2,
        names.arg = names_bar_2,
        space = 0,
        col = "white")
title(main = title_1,
      xlab = xlab_1,
      ylab = ylab_1)

```



```

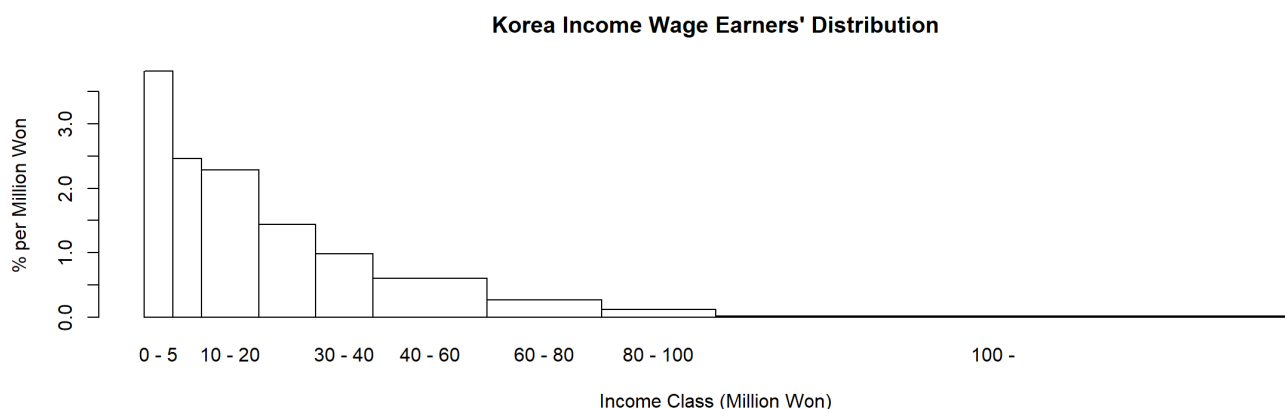
income_kr_3 <- income_kr_2[1:9, ]
income_kr_3[9, ] <- apply(income_kr_2[9:11, ], 2, sum)
rownames(income_kr_3)[9] <- "100 -"
income_breaks_3 <- income_breaks_2[-(11:12)]
income_widths_3 <- diff(income_breaks_3)
height_earners_3 <- income_kr_3[, 1]/income_widths_3
names_bar_3 <- rownames(income_kr_3)

```

```

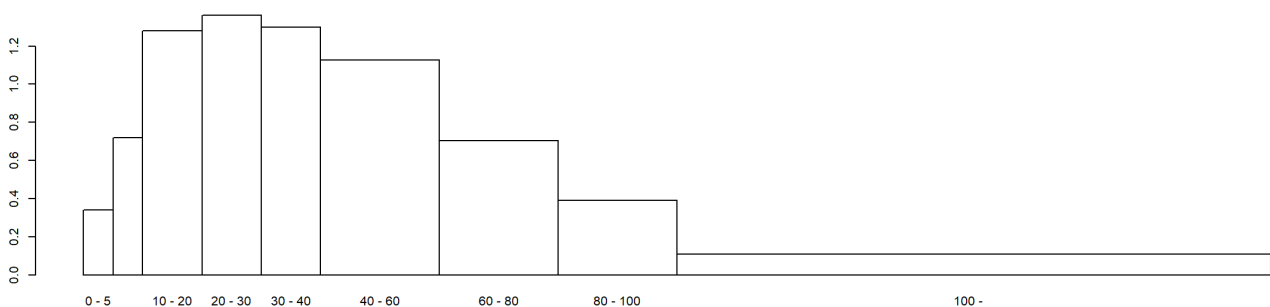
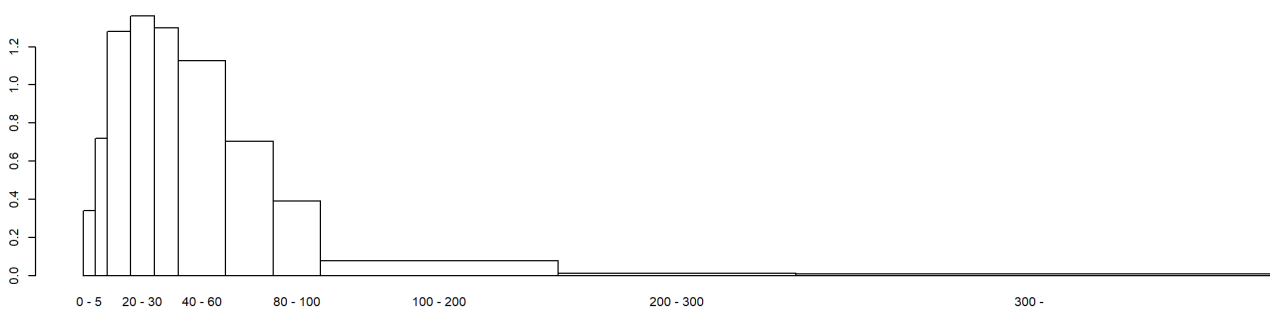
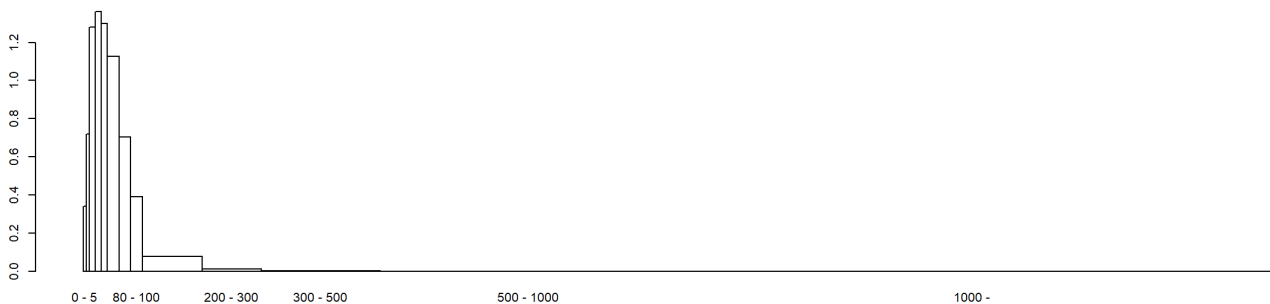
barplot(height_earners_3,
        width = income_widths_3,
        names.arg = names_bar_3,
        space = 0,
        col = "white")
title(main = title_1,
      xlab = xlab_1,
      ylab = ylab_1)

```



```
height_income <- income_kr[, 2]/income_widths
height_income_2 <- income_kr_2[, 2]/income_widths_2
height_income_3 <- income_kr_3[, 2]/income_widths_3
```

```
par(mfrow = c(3, 1))
barplot(height_income,
        width = income_widths,
        names.arg = names_bar,
        space = 0,
        col = "white")
barplot(height_income_2,
        width = income_widths_2,
        names.arg = names_bar_2,
        space = 0,
        col = "white")
barplot(height_income_3,
        width = income_widths_3,
        names.arg = names_bar_3,
        space = 0,
        col = "white")
```



Cumulative distribution

```
income_kr_cum <- apply(income_kr,
                      MARGIN = 2,
                      FUN = cumsum)
```

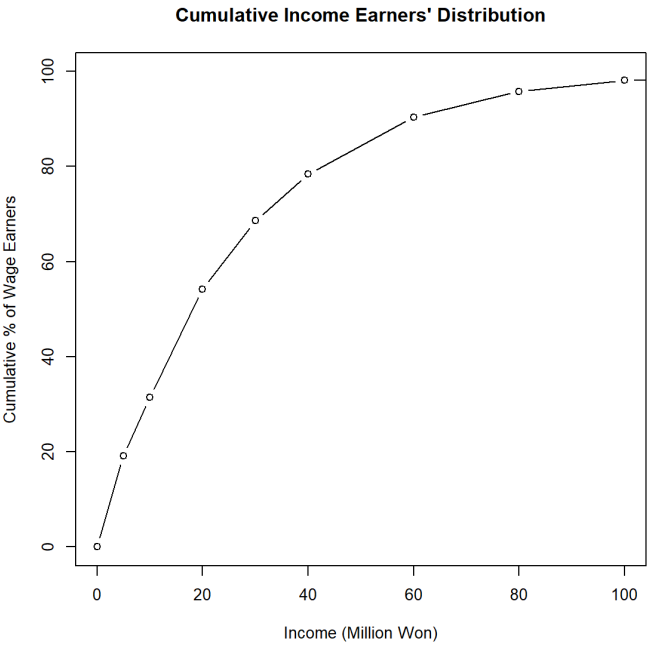
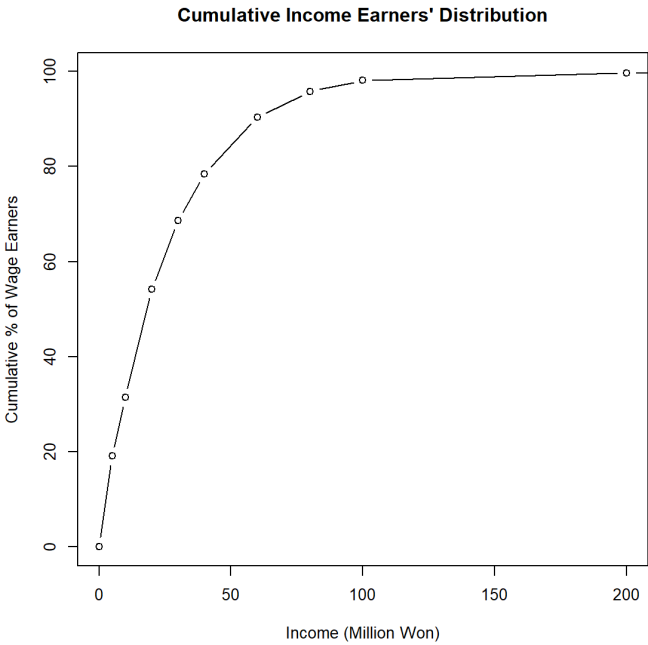
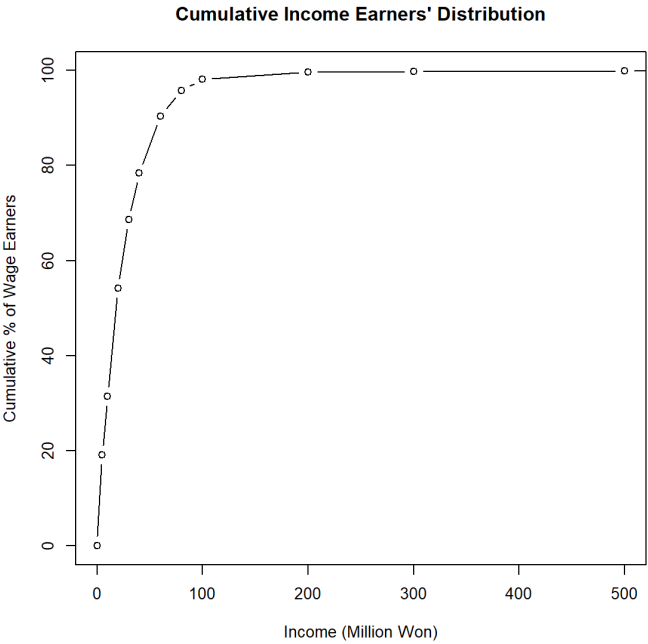
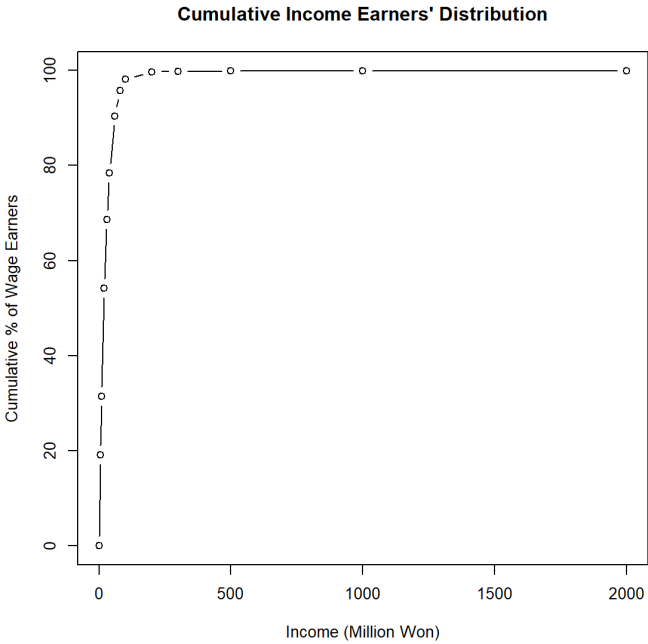
```
income_kr_cum <- rbind(rep(0, 2), income_kr_cum)
```

```
income_class_cum <- strsplit(rownames(income_kr_cum),
                             split = " - ")
income_class_cum <- sapply(income_class_cum,
                          FUN = function(x){x[2]})
income_class_cum <- paste("0 ~", income_class_cum)
income_class_cum[c(1, 14)] <- c("~ 0", "0 ~ 2000")
rownames(income_kr_cum) <- income_class_cum
colnames(income_kr_cum) <- c("Cumulated Wage Earners (%)", "Cumulated Income (%)")
income_kr_cum
```

##	Cumulated Wage Earners (%)	Cumulated Income (%)
## ~ 0	0.0	0.0
## 0 ~ 5	19.1	1.7
## 0 ~ 10	31.4	5.3
## 0 ~ 20	54.2	18.1
## 0 ~ 30	68.6	31.7
## 0 ~ 40	78.4	44.7
## 0 ~ 60	90.4	67.2
## 0 ~ 80	95.8	81.3
## 0 ~ 100	98.1	89.1
## 0 ~ 200	99.7	96.8
## 0 ~ 300	99.8	98.0
## 0 ~ 500	99.9	98.8
## 0 ~ 1000	99.9	99.4
## 0 ~ 2000	99.9	100.0

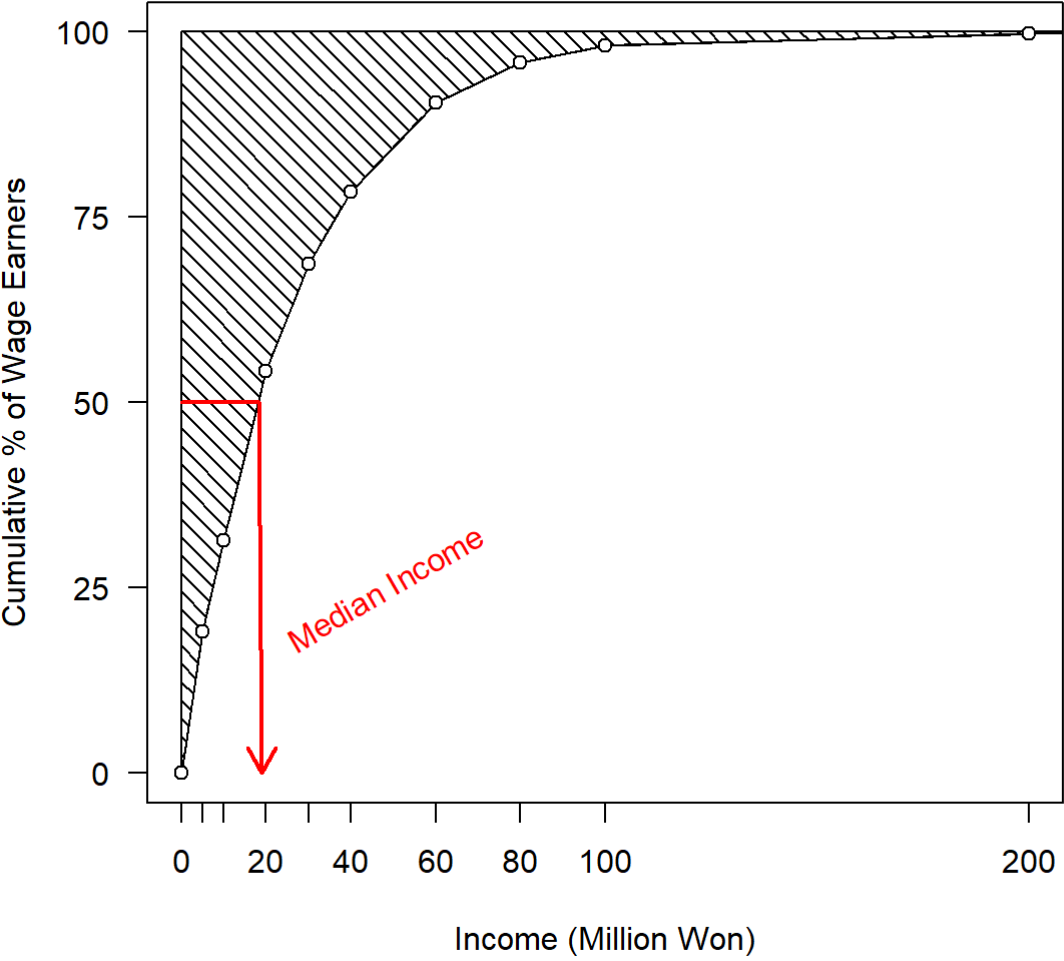
```
earners_kor_cum_df <- data.frame(x = income_breaks, y = income_kr_cum[, 1])
income_kr_cum_df <- data.frame(x = income_breaks, y = income_kr_cum[, 2])
```

```
par(mfrow = c(2, 2))
title_2 <- "Cumulative Income Earners' Distribution"
xlab_2 <- "Income (Million Won)"
ylab_2 <- "Cumulative % of Wage Earners"
plot(earners_kor_cum_df,
     type = "b",
     ann = FALSE)
title(main = title_2,
     xlab = xlab_2,
     ylab = ylab_2)
plot(earners_kor_cum_df,
     type = "b",
     xlim = c(0, 500),
     ann = FALSE)
title(main = title_2,
     xlab = xlab_2,
     ylab = ylab_2)
plot(earners_kor_cum_df,
     type = "b",
     xlim = c(0, 200),
     ann = FALSE)
title(main = title_2,
     xlab = xlab_2,
     ylab = ylab_2)
plot(earners_kor_cum_df,
     type = "b",
     xlim = c(0, 100),
     ann = FALSE)
title(main = title_2,
     xlab = xlab_2,
     ylab = ylab_2)
```

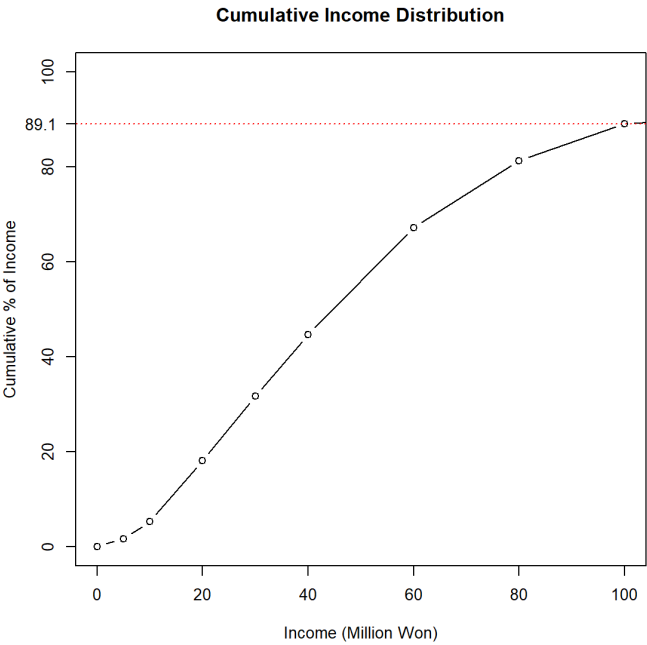
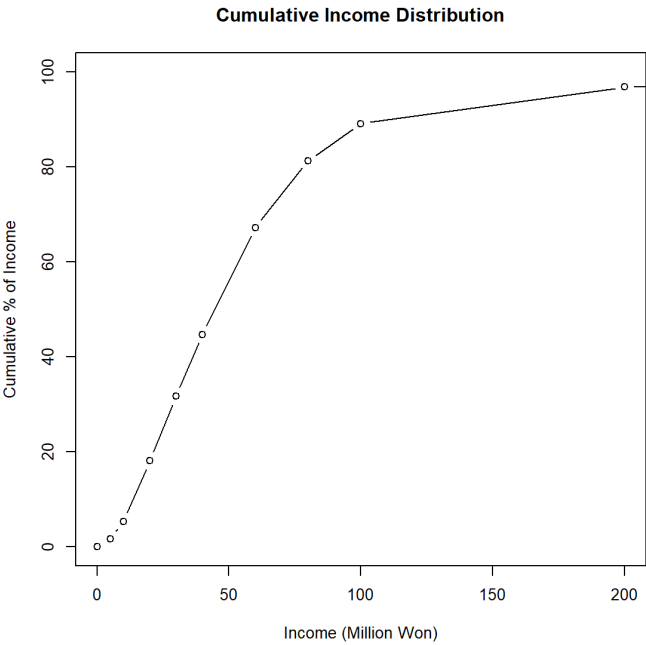
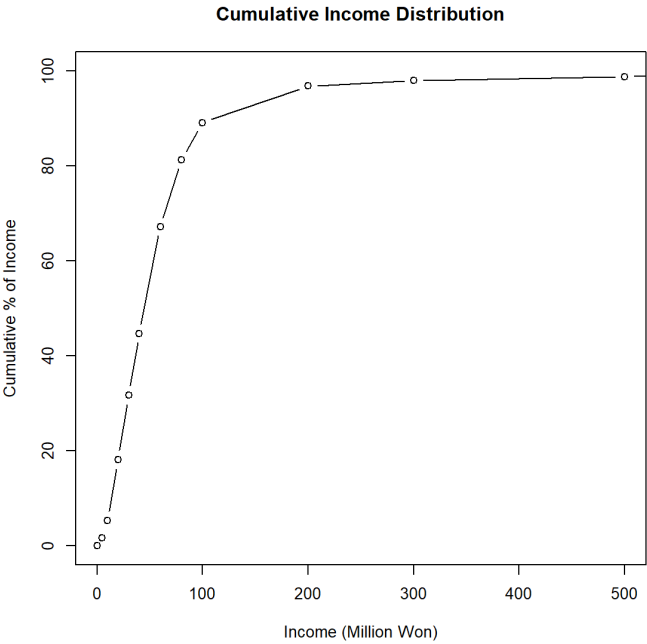
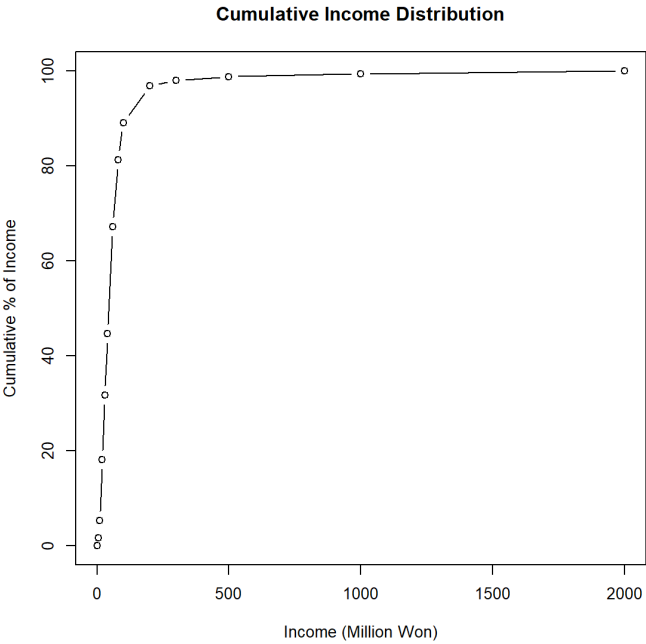



```
plot(earners_kor_cum_df,
     type = "b",
     xlim = c(0, 200),
     ann = FALSE,
     xaxt = "n",
     yaxt = "n")
axis(side = 1,
     at = income_breaks,
     labels = income_breaks)
axis(side = 2,
     at = seq(0, 100, by = 25),
     labels = seq(0, 100, by = 25),
     las = 1)
poly_df <- rbind(earners_kor_cum_df, c(0, 100))
polygon(poly_df,
        density = 15,
        angle = 135)
points(earners_kor_cum_df,
        pch = 21, col = "black", bg = "white")
lines(x = c(0, 18.2), y = rep(50, 2),
      col = "red", lwd = 2)
arrows(x0 = 18.2, y0 = 50, x1 = 19, y1 = 0,
       length = 0.15, col = "red", lwd = 2)
text(x = 48, y = 25,
     labels = "Median Income", srt = 30, col = "red")
title(main = title_2,
      xlab = xlab_2,
      ylab = ylab_2)
```

Cumulative Income Earners' Distribution



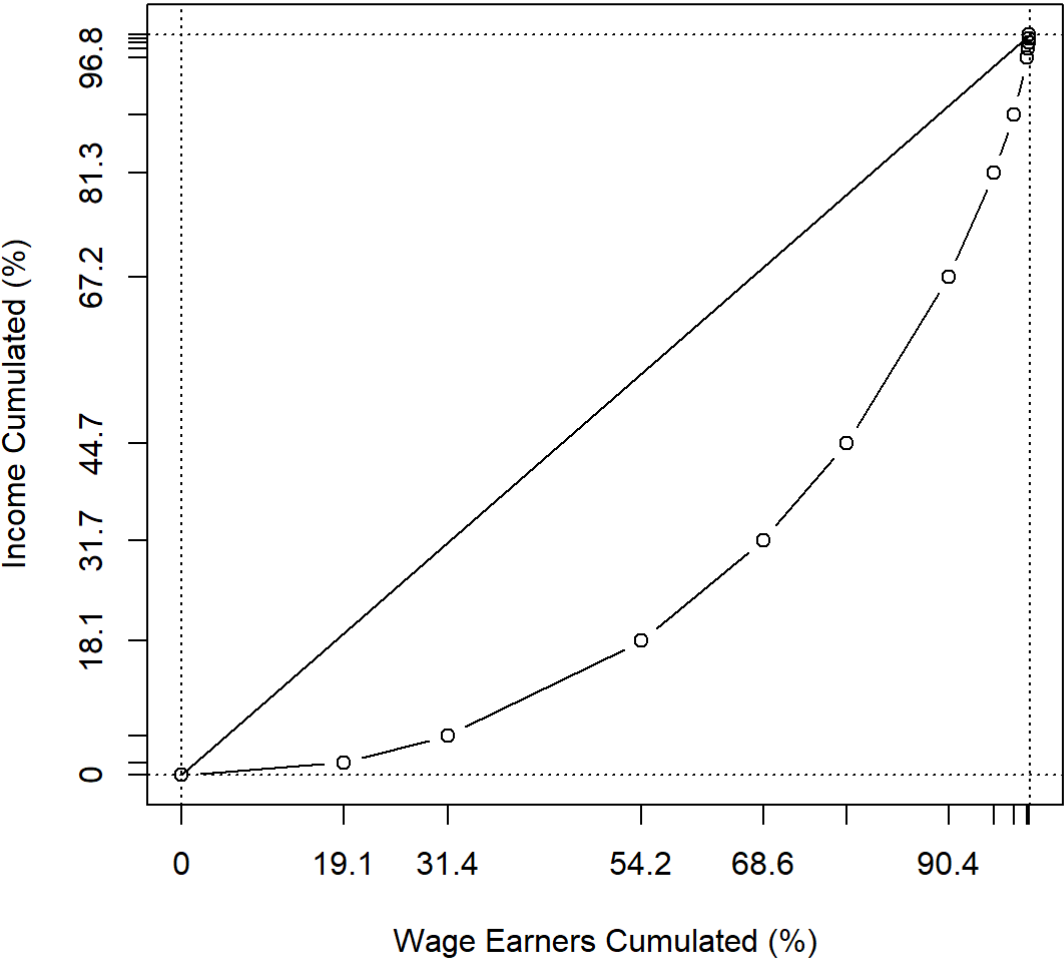
```
par(mfrow = c(2, 2))
title_3 <- "Cumulative Income Distribution"
ylab_3 <- "Cumulative % of Income"
plot(income_kr_cum_df,
     type = "b",
     ann = FALSE)
title(main = title_3,
     xlab = xlab_2,
     ylab = ylab_3)
plot(income_kr_cum_df,
     type = "b",
     ann = FALSE,
     xlim = c(0, 500))
title(main = title_3,
     xlab = xlab_2,
     ylab = ylab_3)
plot(income_kr_cum_df,
     type = "b",
     ann = FALSE,
     xlim = c(0, 200))
title(main = title_3,
     xlab = xlab_2,
     ylab = ylab_3)
plot(income_kr_cum_df,
     type = "b",
     ann = FALSE,
     xlim = c(0, 100))
abline(h = 89.1,
       lty = 3, col = "red")
axis(side = 2,
     at = 89.1,
     label = 89.1,
     las = 1)
title(main = title_3,
     xlab = xlab_2,
     ylab = ylab_3)
```



Lorenz Curve

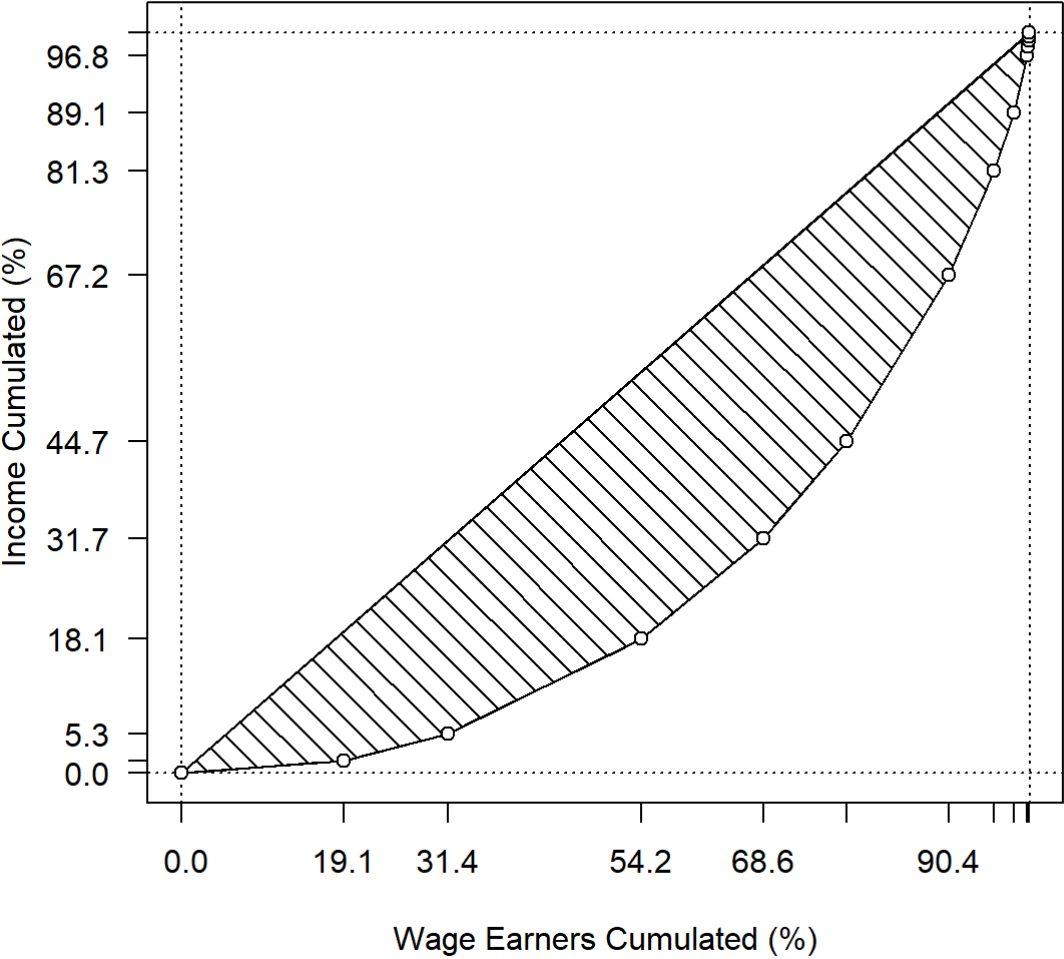
```
earners <- income_kr_cum[, 1]
income <- income_kr_cum[, 2]
earners_income_df <- data.frame(Earners = earners, Income = income)
plot(earners_income_df,
     type = "b",
     ann = FALSE,
     xaxt = "n",
     yaxt = "n")
# abline(a = 0, b = 1, xlim = c(0, 100), ylim = c(0, 100))
lines(x = c(0, 100), y = c(0, 100), type = "l")
axis(side = 1,
     at = earners,
     labels = earners)
axis(side = 2,
     at = income,
     labels = income)
abline(h = c(0, 100), lty = 3)
abline(v = c(0, 100), lty = 3)
title_4 <- "Lorenz Curve of Korea Wage Earners' Income"
xlab_4 <- "Wage Earners Cumulated (%)"
ylab_4 <- "Income Cumulated (%)"
title(main = title_4,
     xlab = xlab_4,
     ylab = ylab_4)
```

Lorenz Curve of Korea Wage Earners' Income



```
plot(earners_income_df,
     type = "b",
     ann = FALSE,
     xaxt = "n",
     yaxt = "n")
# abline(a = 0, b = 1, xlim = c(0, 100), ylim = c(0, 100))
lines(x = c(0, 100), y = c(0, 100), type = "l")
axis(side = 1,
     at = earners,
     labels = format(earners, nsmall = 1))
axis(side = 2,
     at = income[c(1:10, 14)],
     labels = format(income[c(1:10, 14)], nsmall = 1),
     las = 1)
abline(h = c(0, 100), lty = 3)
abline(v = c(0, 100), lty = 3)
title(main = title_4,
     xlab = xlab_4,
     ylab = ylab_4)
polygon(earners_income_df,
     density = 10,
     angle = 135)
points(earners_income_df,
     pch = 21, col = "black", bg = "white")
```


Lorenz Curve of Korea Wage Earners' Income

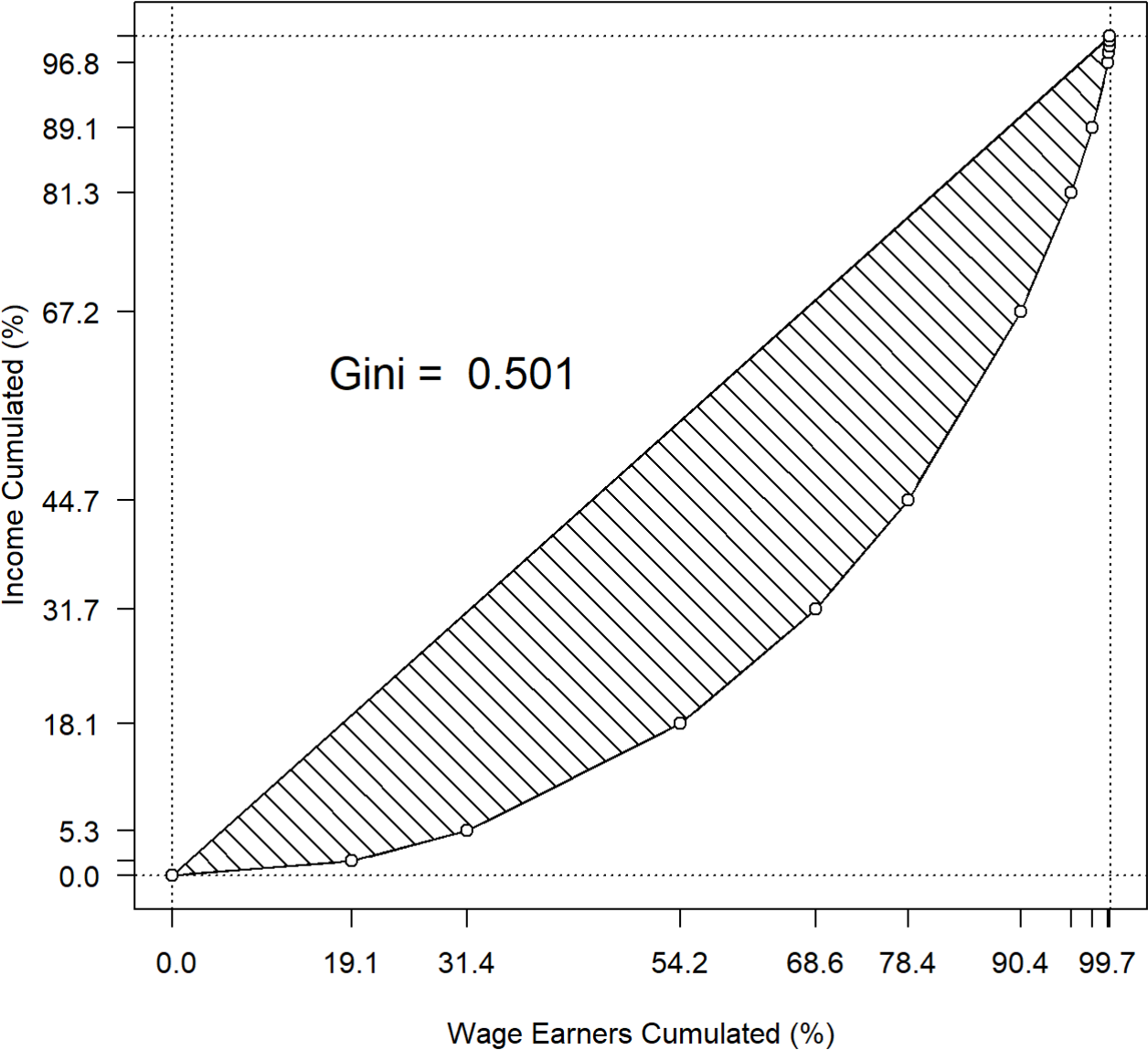


Gini coefficient

```
source("area.R")  
gini <- 2 * (1/2 - area_R(x = earners, y = income)/10000)
```

```
plot(earners_income_df,  
     type = "b",  
     ann = FALSE,  
     xaxt = "n",  
     yaxt = "n")  
lines(x = c(0, 100), y = c(0, 100), type = "l")  
axis(side = 1,  
     at = earners,  
     labels = format(earners, nsmall = 1))  
axis(side = 2,  
     at = income[c(1:10, 14)],  
     labels = format(income[c(1:10, 14)], nsmall = 1),  
     las = 1)  
abline(h = c(0, 100), lty = 3)  
abline(v = c(0, 100), lty = 3)  
title(main = title_4,  
     xlab = xlab_4,  
     ylab = ylab_4)  
polygon(earners_income_df,  
     density = 10,  
     angle = 135)  
points(earners_income_df,  
     pch = 21, col = "black", bg = "white")  
text(x = 30, y = 60,  
     labels = paste("Gini = ", round(gini, digits = 3)), cex = 1.5)
```

Lorenz Curve of Korea Wage Earners' Income



ggplot

Cumulative Distribution

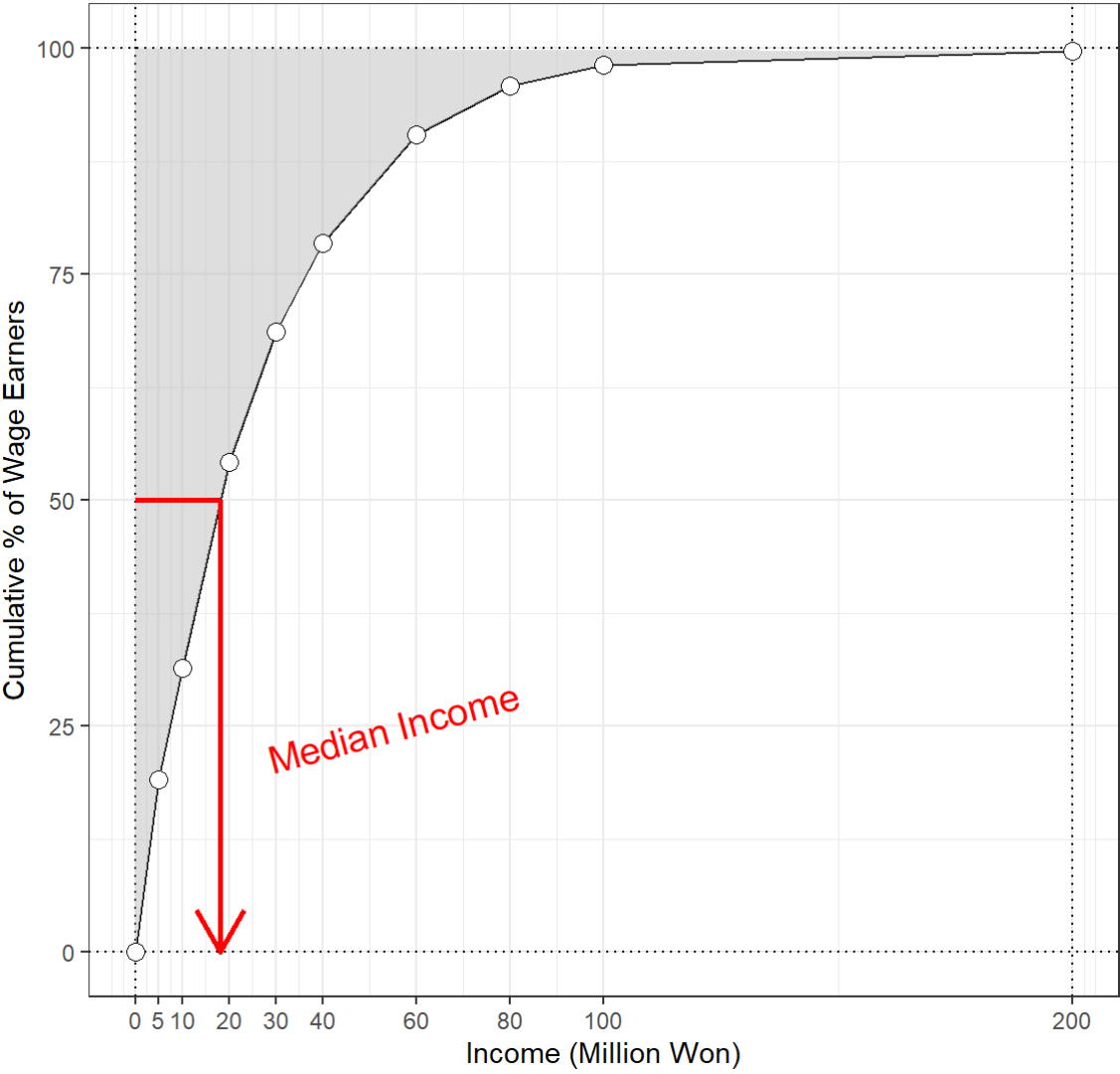
```
library(ggplot2)
(c1 <- ggplot() +
  geom_line(data = earners_kor_cum_df,
            mapping = aes(x = x, y = y), na.rm = TRUE))
(c2 <- c1 +
  scale_x_continuous(breaks = earners_kor_cum_df$x,
                    labels = earners_kor_cum_df$x,
                    limits = c(0, 200)))
(c3 <- c2 +
  geom_hline(yintercept = c(0, 100), linetype = "dotted"))
(c4 <- c3 +
  geom_vline(xintercept = c(0, 200), linetype = "dotted"))
(c5 <- c4 +
  geom_polygon(data = poly_df[-(11:14), ],
              mapping = aes(x = x, y = y),
              alpha = 0.5, fill = "grey"))
(c6 <- c5 +
  geom_point(data = earners_kor_cum_df,
            mapping = aes(x = x, y = y),
            shape = 21, fill = "white", size = 3,
            na.rm = TRUE))
(c7 <- c6 +
  ggtitle(title_2) + xlab(xlab_2) + ylab(ylab_2))
(c8 <- c7 +
  scale_y_continuous(breaks = seq(0, 100, by = 25), labels = seq(0, 100, by = 25)))
(c9 <- c8 +
  annotate("segment", x = 0, xend = 18.2, y = 50, yend = 50, colour = "red", size = 1))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ■ Please use `linewidth` instead.
```

```
(c10 <- c9 +
  geom_segment(data = data.frame(x1 = 18.2, x2 = 18.2, y1 = 50, y2 = 0),
              aes(x = x1, y = y1, xend = x2, yend = y2),
              arrow = arrow(),
              colour = "red",
              size = 1))
(c11 <- c10 +
  annotate("text", x = 55, y = 25,
            label = "Median Income", size = 5, color = "red", srt = 15))
(c12 <- c11 +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, size = 15)))
```

```
c12
```

Cumulative Income Earners' Distribution



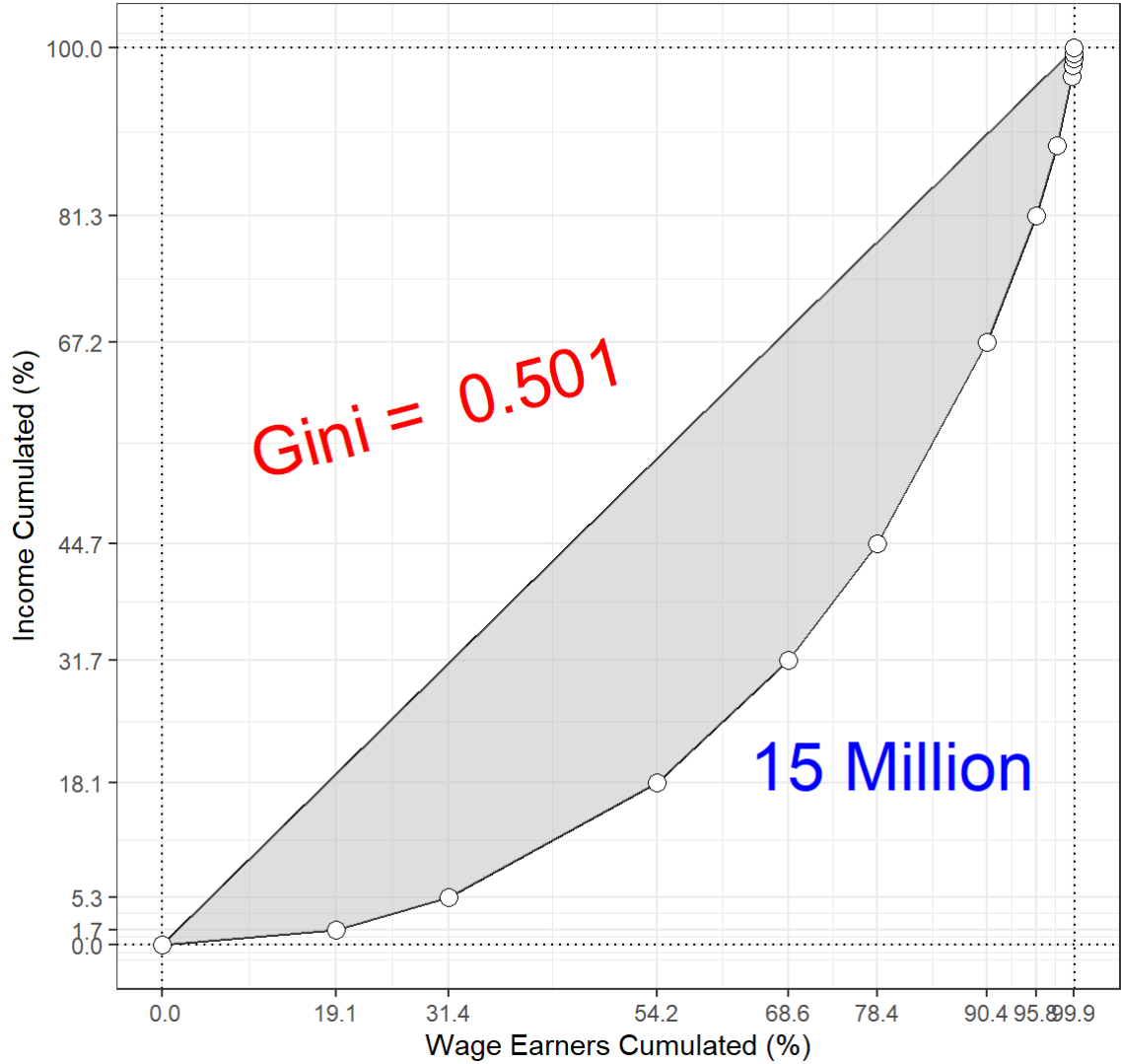
```
ggsave("../pics/cumulative_plot_wage_kr.png", width = 9, height = 9)
```

Lorenz Curve

```
(g1 <- ggplot() +
  geom_line(data = earners_income_df,
    mapping = aes(x = earners, y = income)))
(g2 <- g1 +
  geom_line(data = data.frame(x = c(0, 100), y = c(0, 100)),
    mapping = aes(x = x, y = y)))
(g3 <- g2 +
  geom_hline(yintercept = c(0, 100), linetype = "dotted"))
(g4 <- g3 +
  geom_vline(xintercept = c(0, 100), linetype = "dotted"))
(g5 <- g4 +
  geom_polygon(data = earners_income_df,
    mapping = aes(x = earners, y = income),
    alpha = 0.5, fill = "grey"))
(g6 <- g5 +
  geom_point(data = earners_income_df,
    mapping = aes(x = earners, y = income),
    shape = 21, fill = "white", size = 3))
(g7 <- g6 +
  labs(title = title_4, x = xlab_4, y = ylab_4))
(g8 <- g7 +
  scale_x_continuous(breaks = earners[c(1:8, 14)],
    labels = format(earners[c(1:8, 14)], nsmall = 1)))
(g9 <- g8 +
  scale_y_continuous(breaks = income[c(1:8, 14)],
    labels = format(income[c(1:8, 14)], nsmall = 1)))
# scale_y_continuous(breaks = seq(0, 100, by = 25)))
(g10 <- g9 +
  annotate("text", x = 30, y = 60,
    label = paste("Gini = ", format(gini, digits = 3, nsmall = 2)),
    size = 9, color = "red", srt = 15))
(g11 <- g10 +
  annotate("text", x = 80, y = 20,
    label = "15 Million",
    size = 9, color = "blue"))
(g12 <- g11 +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, size = 15)))
```

g12

Lorenz Curve of Korea Wage Earners' Income



```
ggsave("../pics/lorenz_curve_wage_kr.png", width = 9, height = 9)
```