

US Top Income Share 2014

coop711

2015년5월3일

Data Preparation

준비한 자료는 E. Saez 교수의 홈페이지에 있는 TabFig2014prel.xls 인데 이를 약간의 작업을 거쳐 R에서 불러들이기 편하게 고쳐놓았다.

```
options(digits=2)
library(xlsx)
```

```
## Loading required package: rJava
## Loading required package: xlsxjars
```

```
load("US_top_income_share_2014.rda")
#US.top.income.shares.14 <- read.xlsx("../data/TabFig2014prel.xlsx", sheetIndex
= 9, sheetName = "Table A3", startRow = 6, endRow = 107, colIndex = c(1:7, 9:1
3), header = FALSE)
# v.names <- read.xlsx("./data/TabFig2014prel.xlsx", sheetName = "Table A3", st
artRow = 4, endRow = 4, colIndex = c(2:7, 9:14), colClasses = character, header
= FALSE)
#str(US.top.income.shares.14)
v.names <- c("Year", "P90_100", "P95_100", "P99_100", "P99.5_100", "P99.9_100",
"P99.99_100", "P90_95", "P95_99", "P99_99.5", "P99.5_99.9", "P99.9_99.99")
names(US.top.income.shares.14) <- v.names
ls()
```

```
## [1] "data.1_10"                 "data.1_10.melt"
## [3] "US.top.income.shares.14" "v.names"
```

작업을 마친 자료파일은 US.top.income.shares.14 이며, 이 자료의 구조와 앞의 몇 열의 값은 다음과 같다.

```
str(US.top.income.shares.14)
```

```
## 'data.frame':   102 obs. of  12 variables:
## $ Year      : num  1913 1914 1915 1916 1917 ...
## $ P90_100    : num  NA NA NA NA 40.5 ...
## $ P95_100    : num  NA NA NA NA 30.6 ...
## $ P99_100    : num  18 18.2 17.6 19.3 17.7 ...
## $ P99.5_100  : num  14.7 15.1 14.6 16.4 14.3 ...
## $ P99.9_100  : num  8.62 8.6 9.22 10.51 8.4 ...
## $ P99.99_100 : num  2.76 2.73 4.36 4.78 3.37 ...
## $ P90_95     : num  NA NA NA NA 9.87 ...
## $ P95_99     : num  NA NA NA NA 12.9 ...
## $ P99_99.5   : num  3.23 3.08 3 2.94 3.39 ...
## $ P99.5_99.9 : num  6.11 6.48 5.36 5.86 5.94 ...
## $ P99.9_99.99: num  5.86 5.87 4.86 5.73 5.04 ...
```

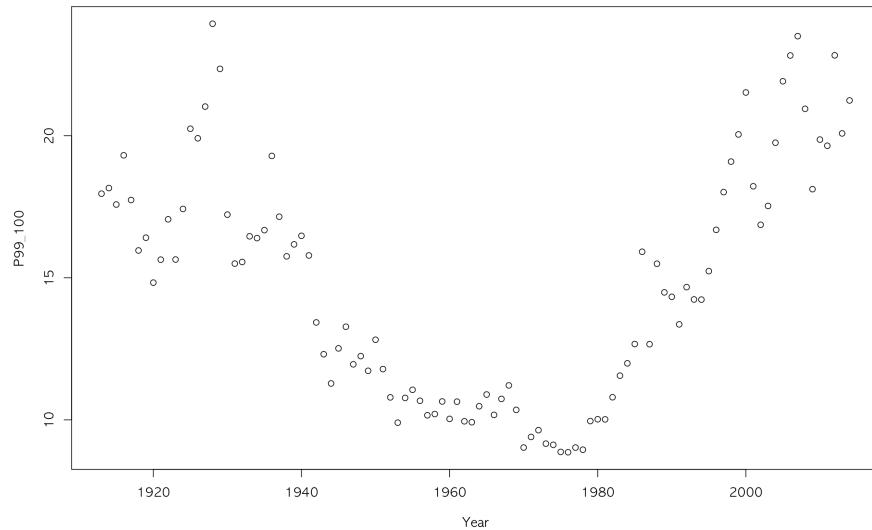
```
US.top.income.shares.14[, 1:8]
```

##	Year	P90_100	P95_100	P99_100	P99.5_100	P99.9_100	P99.99_100	P90_95
## 1	1913	NA	NA	18.0	14.7	8.6	2.76	NA
## 2	1914	NA	NA	18.2	15.1	8.6	2.73	NA
## 3	1915	NA	NA	17.6	14.6	9.2	4.36	NA
## 4	1916	NA	NA	19.3	16.4	10.5	4.78	NA
## 5	1917	41	31	17.7	14.3	8.4	3.37	9.9
## 6	1918	40	29	16.0	12.4	6.7	2.45	10.6
## 7	1919	40	30	16.4	12.6	6.6	2.29	10.1
## 8	1920	39	28	14.8	11.1	5.4	1.66	10.7
## 9	1921	43	31	15.6	11.7	5.6	1.69	12.4
## 10	1922	44	32	17.1	13.1	6.6	2.27	11.8
## 11	1923	41	30	15.6	11.9	5.9	2.00	11.7
## 12	1924	44	32	17.4	13.4	6.8	2.32	12.3
## 13	1925	46	35	20.2	15.9	8.5	3.31	11.3
## 14	1926	46	35	19.9	15.5	8.5	3.36	11.1
## 15	1927	47	36	21.0	16.6	9.3	3.75	11.0
## 16	1928	49	39	23.9	19.4	11.5	5.02	10.7
## 17	1929	47	36	22.4	18.1	10.9	4.99	10.2
## 18	1930	44	32	17.2	13.2	7.1	2.84	11.8
## 19	1931	45	31	15.5	11.6	5.9	2.25	13.3
## 20	1932	46	33	15.6	11.6	6.0	1.99	13.7
## 21	1933	46	33	16.5	12.5	6.6	2.34	12.4
## 22	1934	46	34	16.4	12.3	6.1	2.07	12.1
## 23	1935	44	32	16.7	12.6	6.4	2.19	12.2
## 24	1936	47	35	19.3	14.9	7.6	2.54	12.0
## 25	1937	44	32	17.1	13.0	6.5	2.17	12.0
## 26	1938	44	31	15.8	11.8	5.9	2.19	12.7
## 27	1939	46	32	16.2	12.1	5.9	1.96	13.2
## 28	1940	45	32	16.5	12.3	6.0	2.04	13.1
## 29	1941	42	30	15.8	11.9	5.8	1.98	11.9
## 30	1942	36	26	13.4	10.1	4.8	1.55	10.3
## 31	1943	34	24	12.3	9.1	4.3	1.24	9.6
## 32	1944	33	23	11.3	8.3	3.8	1.16	9.7
## 33	1945	34	25	12.5	9.1	4.2	1.26	9.6
## 34	1946	37	27	13.3	9.6	4.4	1.47	9.9
## 35	1947	34	25	12.0	8.6	3.9	1.30	9.7
## 36	1948	35	25	12.2	8.9	4.1	1.31	10.0
## 37	1949	35	25	11.7	8.5	3.8	1.24	10.2
## 38	1950	36	26	12.8	9.4	4.4	1.22	10.0
## 39	1951	34	24	11.8	8.5	3.9	1.28	10.0
## 40	1952	33	23	10.8	7.7	3.4	1.09	10.1
## 41	1953	32	22	9.9	7.0	3.1	0.97	10.3
## 42	1954	34	23	10.8	7.7	3.5	1.17	10.3
## 43	1955	34	24	11.1	8.0	3.7	1.32	10.3
## 44	1956	33	23	10.7	7.7	3.5	1.20	10.3
## 45	1957	33	23	10.2	7.2	3.2	1.05	10.4
## 46	1958	34	23	10.2	7.3	3.2	1.08	10.6
## 47	1959	34	23	10.6	7.7	3.4	1.19	10.6
## 48	1960	33	23	10.0	7.1	3.2	1.17	10.9
## 49	1961	34	24	10.6	7.7	3.7	1.38	10.8
## 50	1962	34	23	9.9	7.1	3.2	1.16	10.9
## 51	1963	34	23	9.9	7.0	3.1	1.15	10.9
## 52	1964	34	24	10.5	7.4	3.4	1.30	10.9

## 53	1965	35	24	10.9	7.7	3.7	1.49	10.9
## 54	1966	34	23	10.2	7.2	3.4	1.29	10.8
## 55	1967	34	24	10.7	7.7	3.7	1.42	10.7
## 56	1968	35	24	11.2	8.1	4.0	1.61	10.7
## 57	1969	34	23	10.4	7.4	3.7	1.56	10.8
## 58	1970	33	22	9.0	6.3	2.8	1.00	11.0
## 59	1971	33	22	9.4	6.6	3.0	1.11	11.1
## 60	1972	34	23	9.6	6.8	3.1	1.18	11.1
## 61	1973	33	22	9.2	6.3	2.8	0.94	11.1
## 62	1974	33	22	9.1	6.3	2.7	0.88	11.2
## 63	1975	33	22	8.9	6.1	2.6	0.85	11.5
## 64	1976	33	22	8.9	6.1	2.6	0.86	11.4
## 65	1977	34	22	9.0	6.2	2.7	0.92	11.5
## 66	1978	33	22	9.0	6.2	2.6	0.86	11.5
## 67	1979	34	23	10.0	7.1	3.4	1.37	11.3
## 68	1980	35	23	10.0	7.1	3.4	1.28	11.5
## 69	1981	35	23	10.0	7.2	3.6	1.37	11.5
## 70	1982	35	24	10.8	8.0	4.2	1.73	11.5
## 71	1983	36	25	11.6	8.6	4.6	1.88	11.5
## 72	1984	37	25	12.0	9.0	5.0	2.15	11.4
## 73	1985	38	26	12.7	9.6	5.3	2.24	11.4
## 74	1986	41	29	15.9	12.6	7.4	3.34	11.1
## 75	1987	38	27	12.7	9.4	4.9	1.91	11.7
## 76	1988	41	29	15.5	12.1	6.8	2.86	11.3
## 77	1989	40	29	14.5	11.1	6.0	2.45	11.5
## 78	1990	40	28	14.3	10.9	5.8	2.33	11.6
## 79	1991	40	28	13.4	10.0	5.1	1.96	11.8
## 80	1992	41	29	14.7	11.2	6.0	2.46	11.8
## 81	1993	41	29	14.2	10.8	5.7	2.32	11.9
## 82	1994	41	29	14.2	10.7	5.7	2.29	11.9
## 83	1995	42	30	15.2	11.6	6.2	2.46	11.9
## 84	1996	43	32	16.7	12.9	7.2	3.06	11.7
## 85	1997	45	33	18.0	14.2	8.2	3.53	11.5
## 86	1998	45	34	19.1	15.2	9.0	3.92	11.3
## 87	1999	46	35	20.0	16.0	9.6	4.21	11.3
## 88	2000	48	37	21.5	17.5	10.9	5.07	11.0
## 89	2001	45	33	18.2	14.3	8.4	3.70	11.5
## 90	2002	44	32	16.9	13.0	7.3	3.14	11.8
## 91	2003	45	33	17.5	13.7	7.9	3.49	11.8
## 92	2004	46	35	19.8	15.7	9.5	4.34	11.4
## 93	2005	48	37	21.9	17.8	11.0	5.13	11.2
## 94	2006	49	38	22.8	18.6	11.6	5.46	11.2
## 95	2007	50	39	23.5	19.3	12.3	6.04	11.1
## 96	2008	48	37	20.9	16.9	10.4	5.03	11.7
## 97	2009	47	34	18.1	14.2	8.3	3.89	12.4
## 98	2010	48	36	19.9	15.8	9.7	4.78	12.2
## 99	2011	48	36	19.6	15.5	9.3	4.32	12.2
## 100	2012	51	39	22.8	18.6	11.7	5.81	11.8
## 101	2013	49	37	20.1	15.9	9.5	4.48	12.3
## 102	2014	50	38	21.2	17.0	10.3	4.89	12.1

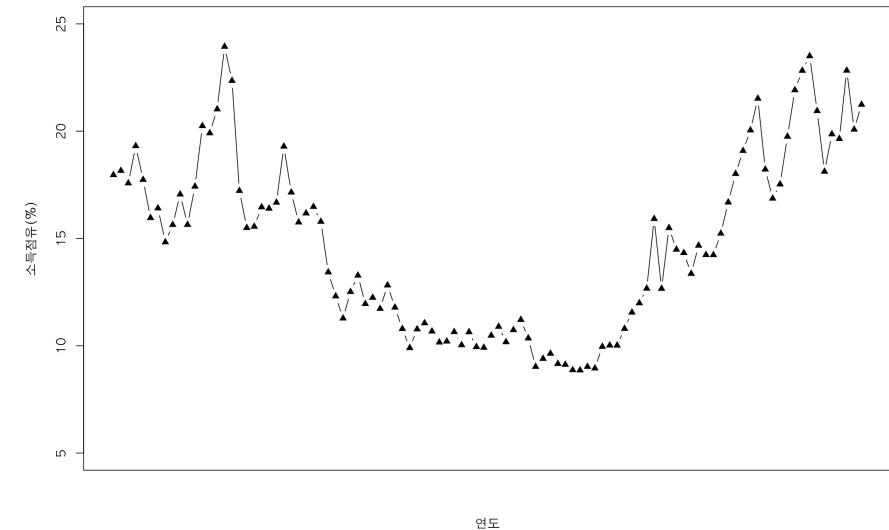
이 중에서 소득 상위 10%(`p90_100`)를 상위 1%(`p99_100`), 차상위 4%(`p95_99`), 차차상위 5%(`p90_95`)로 나누어 그 끝의 변화를 살펴보자. 우선 깊숙하게 상위 1%의 소득 점유율 변화만 그려보면,

```
plot(P99_100~Year, data=US.top.income.shares.14)
```



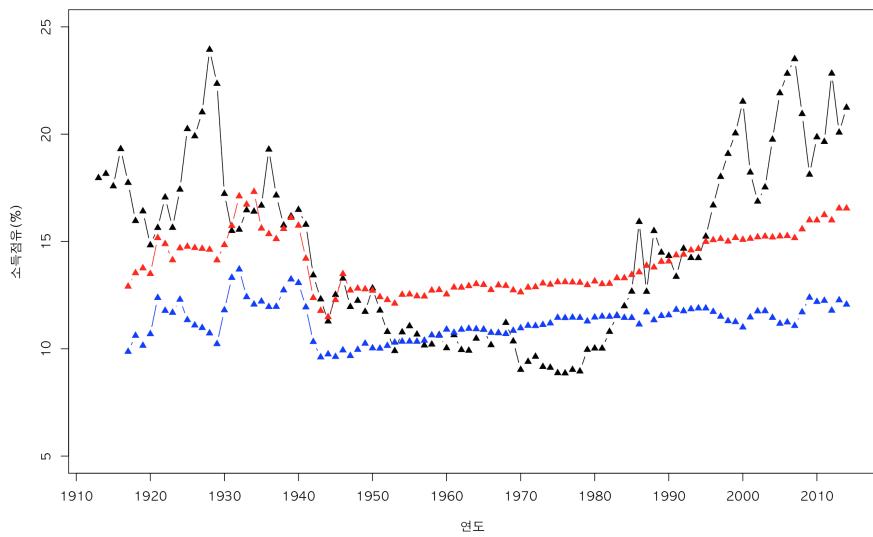
최소한의 정보를 주기 위하여 각 자료의 최대값, 최소값을 고려하여 y-축의 범위를 `ylim=c(5,25)`로 설정하고, x-축의 눈금은 비워둔다. 점들을 선으로 이어주고, 점은 다시 삼각형 모양으로 바꾼다.

```
plot(P99_100 ~ Year, data = US.top.income.shares.14, xlab="연도", ylab = "소득점  
유(%)", ylim = c(5, 25), xaxt = "n", type = "b", pch = 17)
```



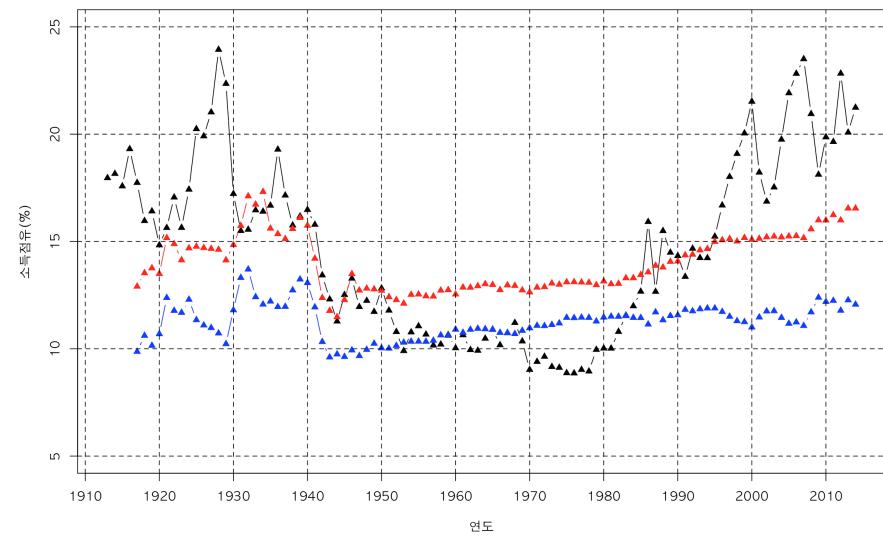
이제 x-축에는 연도를 10년 단위로 표시하고, `lines()`를 이용하여 차상위4%와 차차상위5%의 소득 점유율 변화를 함께 그린다.

```
plot(P99_100~Year,data=US.top.income.shares.14, xlab="연도", ylab="소득점  
유(%)", ylim=c(5,25), xaxt="n", type="b",pch=17)  
axis(side=1, at=seq(1910, 2010, by=10), labels=seq(1910, 2010, by=10))  
lines(P95_99~Year,data=US.top.income.shares.14,type="b", pch=17, col="red")  
lines(P90_95~Year,data=US.top.income.shares.14,type="b", pch=17, col="blue")
```



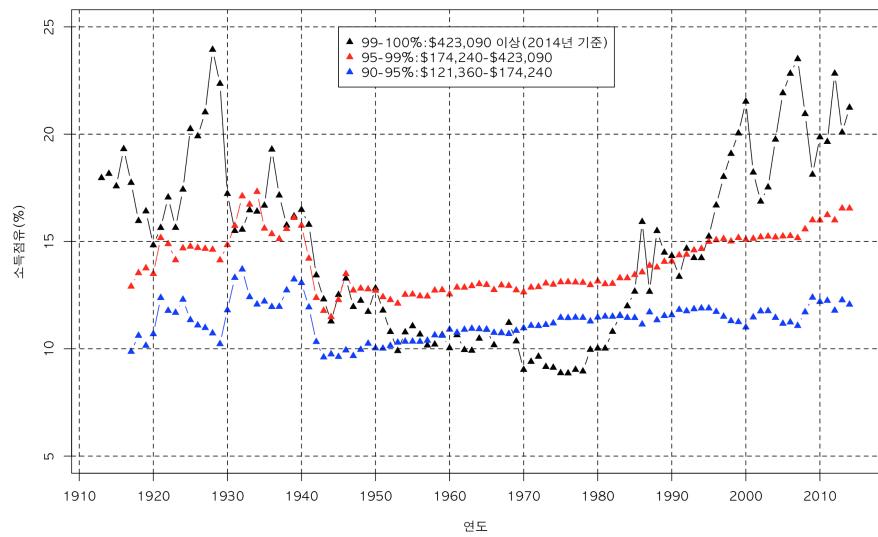
도표의 가독성을 높이기 위하여 x-축과 y-축을 가로지르는 격자를 설치한다.

```
plot(P99_100~Year,data=US.top.income.shares.14, xlab="연도", ylab="소득점유(%)", ylim=c(5,25), xaxt="n", type="b",pch=17)
axis(side=1, at=seq(1910, 2010, by=10), labels=seq(1910, 2010, by=10))
lines(P95_99~Year,data=US.top.income.shares.14,type="b", pch=17, col="red")
lines(P90_95~Year,data=US.top.income.shares.14,type="b", pch=17, col="blue")
abline(h=seq(5,25, by=5), lty=2)
abline(v=seq(1910, 2010, by=10), lty=2)
```



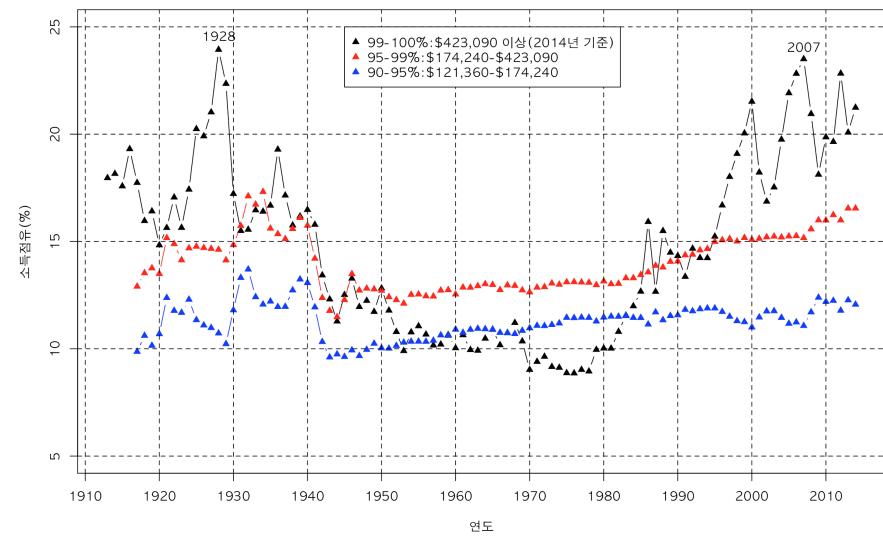
TabFig2014prel.xls 의 Table 0 시트에 나와 있는 정보를 활용하여 범례를 만든다.

```
plot(P99_100~Year,data=US.top.income.shares.14, xlab="연도", ylab="소득점유(%)", ylim=c(5,25), xaxt="n", type="b",pch=17)
axis(side=1, at=seq(1910, 2010, by=10), labels=seq(1910, 2010, by=10))
lines(P95_99~Year,data=US.top.income.shares.14,type="b", pch=17, col="red")
lines(P90_95~Year,data=US.top.income.shares.14,type="b", pch=17, col="blue")
abline(h=seq(5,25, by=5), lty=2)
abline(v=seq(1910, 2010, by=10), lty=2)
legend(x=1945, y=25, legend=c("99-100%:$423,090 이상(2014년 기준)", "95-99%:$174,240-$423,090", "90-95%:$121,360-$174,240"), pch=17, col=c("black", "red", "blue"))
```



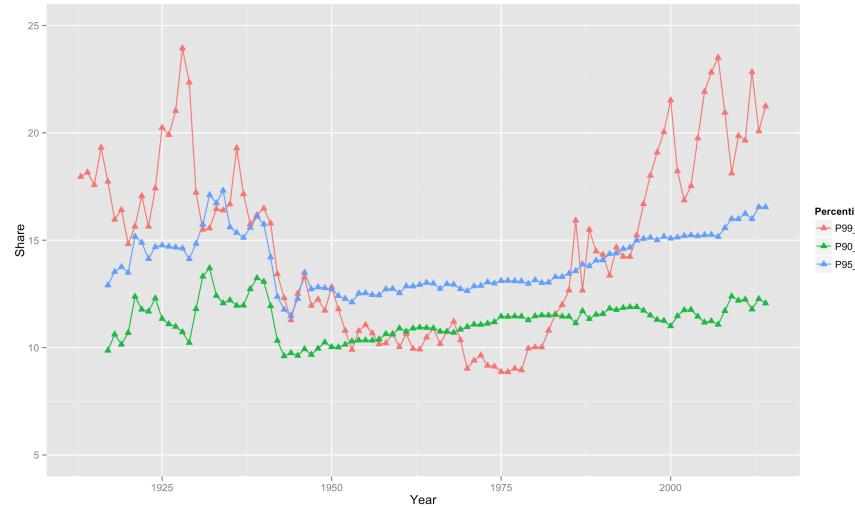
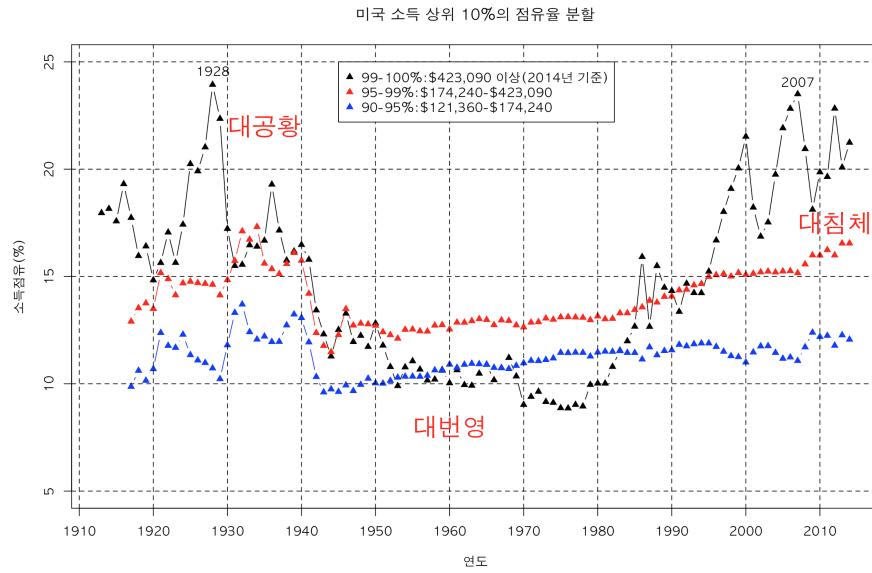
메인 타이틀을 입력하고, 상위1%의 소득점유율이 최고에 달했던 연도를 표시한다.

```
plot(P99_100~Year,data=US.top.income.shares.14, xlab="연도", ylab="소득점유(%)", ylim=c(5,25), xaxt="n", type="b", pch=17)
axis(side=1, at=seq(1910, 2010, by=10), labels=seq(1910, 2010, by=10))
lines(P95_99~Year,data=US.top.income.shares.14,type="b", pch=17, col="red")
lines(P90_95~Year,data=US.top.income.shares.14,type="b", pch=17, col="blue")
abline(h=seq(5,25, by=5), lty=2)
abline(v=seq(1910, 2010, by=10), lty=2)
legend(x=1945, y=25, legend=c("99-100%:$423,090 이상(2014년 기준)", "95-99%:$174,240-$423,090", "90-95%:$121,360-$174,240"), pch=17, col=c("black", "red", "blue"))
title(main="미국 소득 상위 10%의 점유율 분할")
text(x=c(1928, 2007), y=c(24, 23.5), labels=c("1928", "2007"), pos=3)
```



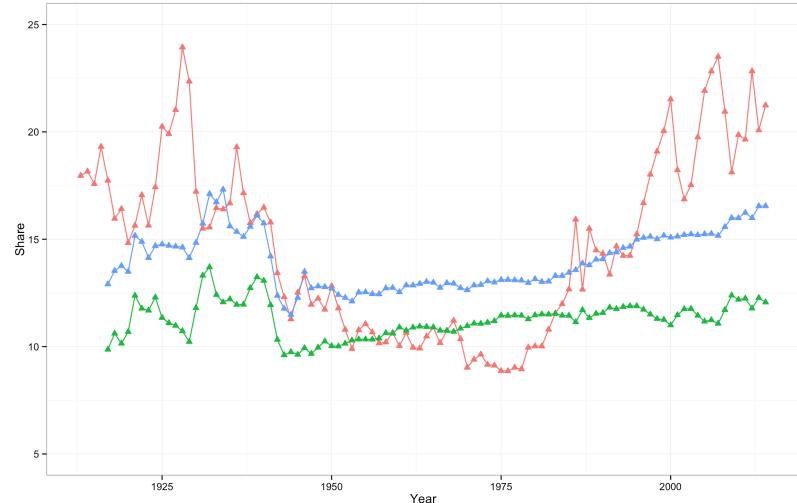
역사적으로 각 시기를 어떻게 부르고 있는지 텍스트를 추가한다.

```
plot(P99_100~Year,data=US.top.income.shares.14, xlab="연도", ylab="소득점유(%)", ylim=c(5,25), xaxt="n", type="b", pch=17)
axis(side=1, at=seq(1910, 2010, by=10), labels=seq(1910, 2010, by=10))
lines(P95_99~Year,data=US.top.income.shares.14,type="b", pch=17, col="red")
lines(P90_95~Year,data=US.top.income.shares.14,type="b", pch=17, col="blue")
abline(h=seq(5,25, by=5), lty=2)
abline(v=seq(1910, 2010, by=10), lty=2)
legend(x=1945, y=25, legend=c("99-100%:$423,090 이상(2014년 기준)", "95-99%:$174,240-$423,090", "90-95%:$121,360-$174,240"), pch=17, col=c("black", "red", "blue"))
title(main="미국 소득 상위 10%의 점유율 분할")
text(x=c(1928, 2007), y=c(24, 23.5), labels=c("1928", "2007"), pos=3)
text(x=1935, y=22, label="대공황", cex=2.0, col="red")
text(x=1960, y=8, label="대번영", cex=2.0, col="red")
text(x=2012, y=17.5, label="대침체", cex=2.0, col="red")
```



theme_bw() 적용

```
library(ggplot2)
(g1 <- g0 + theme_bw())
```



격자 설정

```
(g2 <- g1 + theme(panel.grid.major = element_line(linetype = "dashed", colour = "black")))
```

ggplot

Data Reshaping

- reshape2 패키지를 이용하여 wide format 을 long format 으로

```
library(reshape2)
data.1_10 <- US.top.income.shares.14[c("Year", "P99_100", "P90_95", "P95_99")]
data.1_10.melt <- melt(data.1_10, id.vars = "Year", measure.vars = c("P99_100",
"P90_95", "P95_99"), variable.name = "Percentiles", value.name = "Share")
str(data.1_10.melt)
```

```
## 'data.frame': 306 obs. of 3 variables:
## $ Year      : num 1913 1914 1915 1916 1917 ...
## $ Percentiles: Factor w/ 3 levels "P99_100","P90_95",...: 1 1 1 1 1 1 1 1 1 ...
## $ Share      : num 18 18.2 17.6 19.3 17.7 ...
```

골격 그리기

```
library(ggplot2)
(g0 <- ggplot(data.1_10.melt, aes(x = Year, y = Share, colour = Percentiles)) +
geom_line() + geom_point(shape=24, aes(fill = Percentiles)) + ylim(5, 25))
```



```
source("./theme_kr.R")
ls()
```

```
## [1] "data.1_10"
## [3] "g0"
## [5] "g2"
## [7] "theme.kr"
## [9] "v.names"
```

한글 테마 적용, x-축과 y-축의 라벨 설정

```
(g4 <- g3 + theme.kr + xlab("연도") + ylab("소득점유(%)"))
```



전체 제목 추가

```
(g5 <- g4 + ggtitle("미국 소득 상위 10%의 점유율 분할") + theme(plot.title=element_text(size=20)))
```

x-축 눈금 위치 설정

```
library(ggplot2)
(g3 <- g2 + scale_x_continuous(breaks = seq(1910, 2010, by = 10)))
```

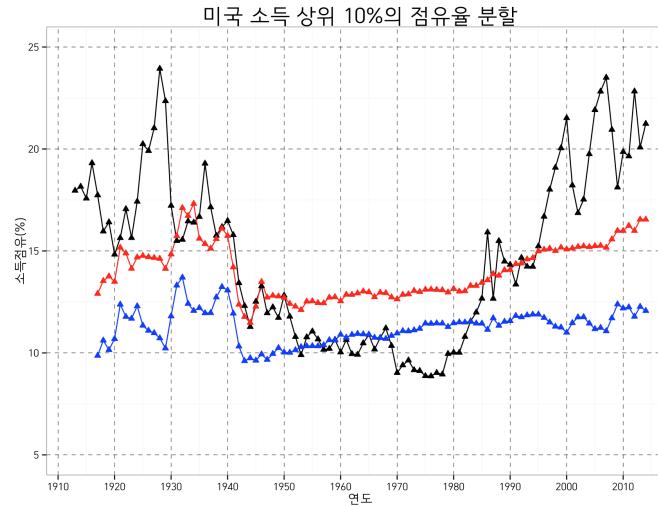


한글 정보 입력

한글 테마 sourcing

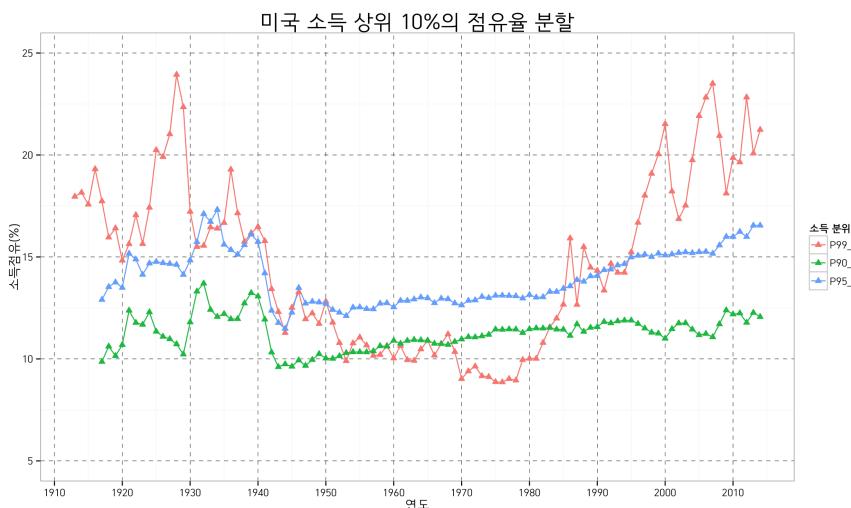


```
(g7 <- g6 + scale_colour_manual(name = "", values = c("black", "blue", "red"),
  labels = c("99-100%:$423,090 이상(2014년기준)", "95-99%:$174,240-$423,090", "90-95%:$121,360-$174,240")) + scale_fill_manual(name = "", values = c("black", "blue", "red"), labels = c("99-100%:$423,090 이상(2014년기준)", "95-99%:$174,240-$423,090", "90-95%:$121,360-$174,240")))
```



범례 제목 수정

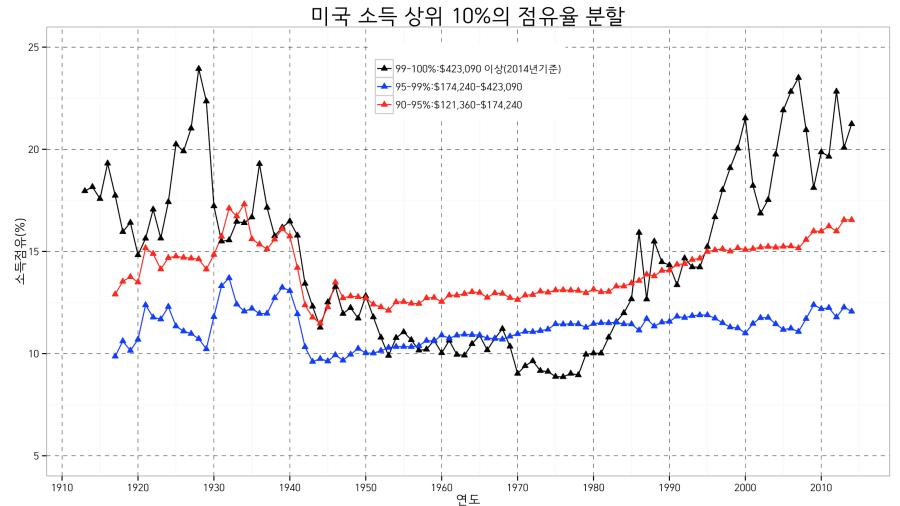
```
(g6 <- g5 + labs(colour = "소득 분위", fill = "소득 분위"))
```



범례와 색깔 수정, 범례 제목 없애기

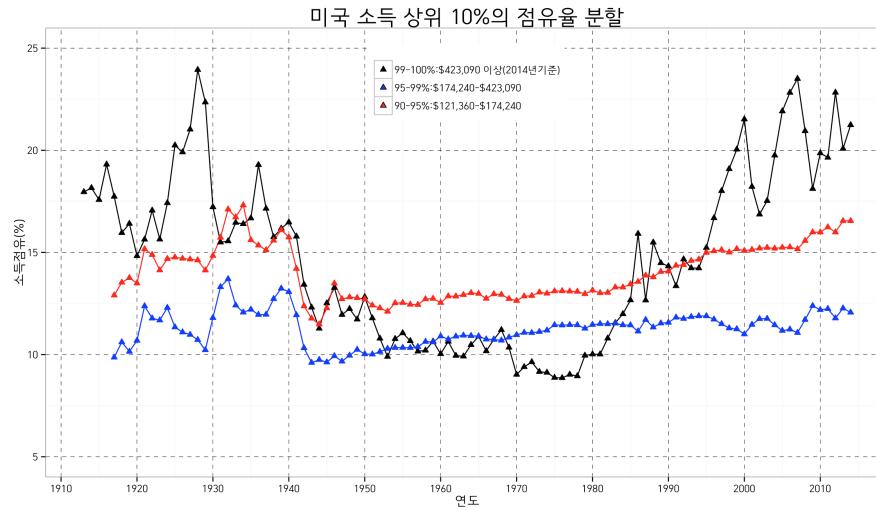
범례를 안쪽으로

```
(g8 <- g7 + theme(legend.position = c(0.5, 0.88)))
```



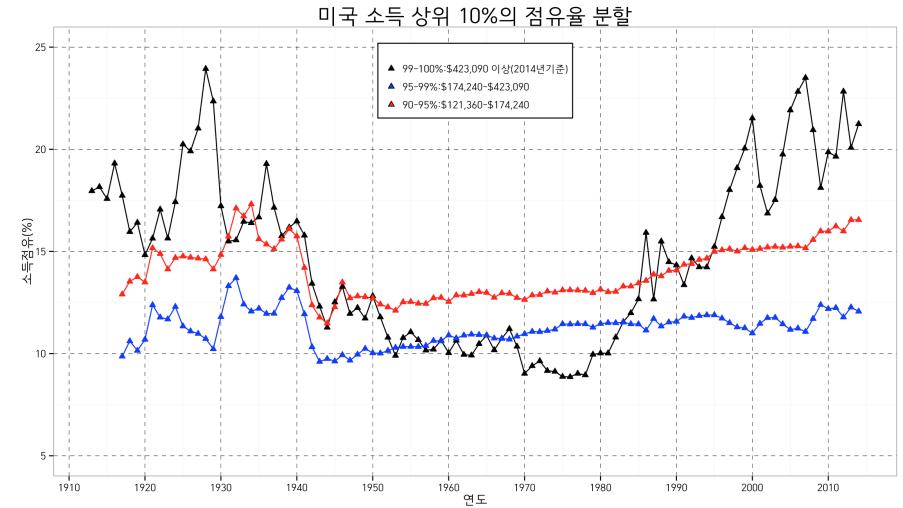
colour 범례 없애기(colour 설정으로 나타나는 범례의 선 없애는 효과)

```
#(g9 <- g8 + guides(colour = guide_legend(title=NULL), fill = guide_legend(title=NULL)))
(g9 <- g8 + guides(colour = "none"))
```



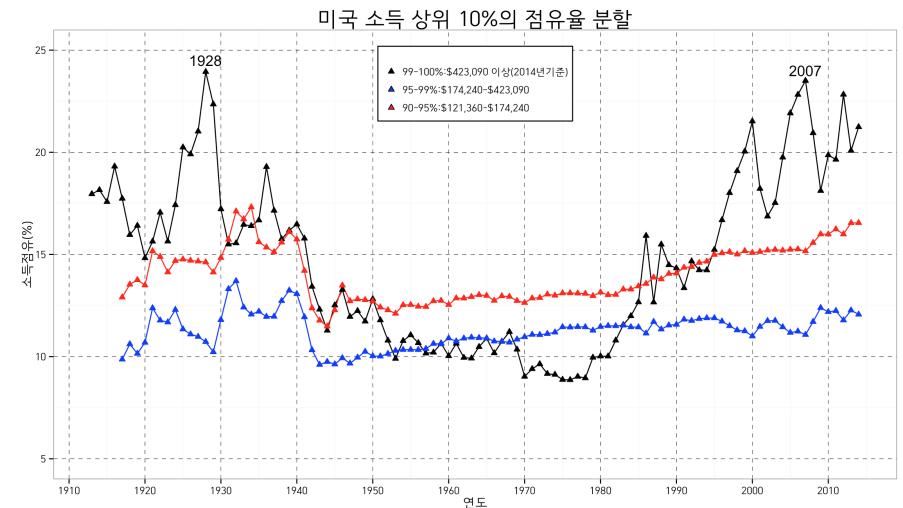
범례 항목 박스 없애기

```
(g11 <- g10 + theme(legend.key = element_blank()))
```



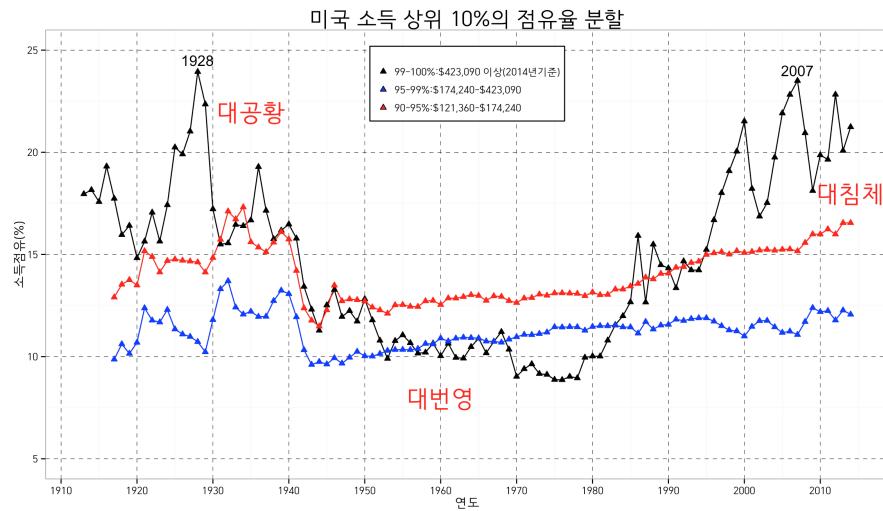
역사상 고점 시기

```
(g12 <- g11 + annotate("text", x=c(1928, 2007), y=c(24.5, 24), label = c(1928, 2007)))
```



시대적 특징 텍스트 입력

```
times.label <- c("대공황", "대번영", "대침체")
(g13 <- g12 + annotate("text", x=c(1935, 1960, 2014), y=c(22, 8, 18), label = times.label, colour="red", family="HCR Dotum LVT", size = 8))
```



뒷 정리

```
save.image(file="US_top_income_shares_2014_add.rda")
```