

Mini-project 1

S681

Sumeet Mishra, Chun-Sheng Kimball Wu, Ling Tan, and Pei-Yi Cheng

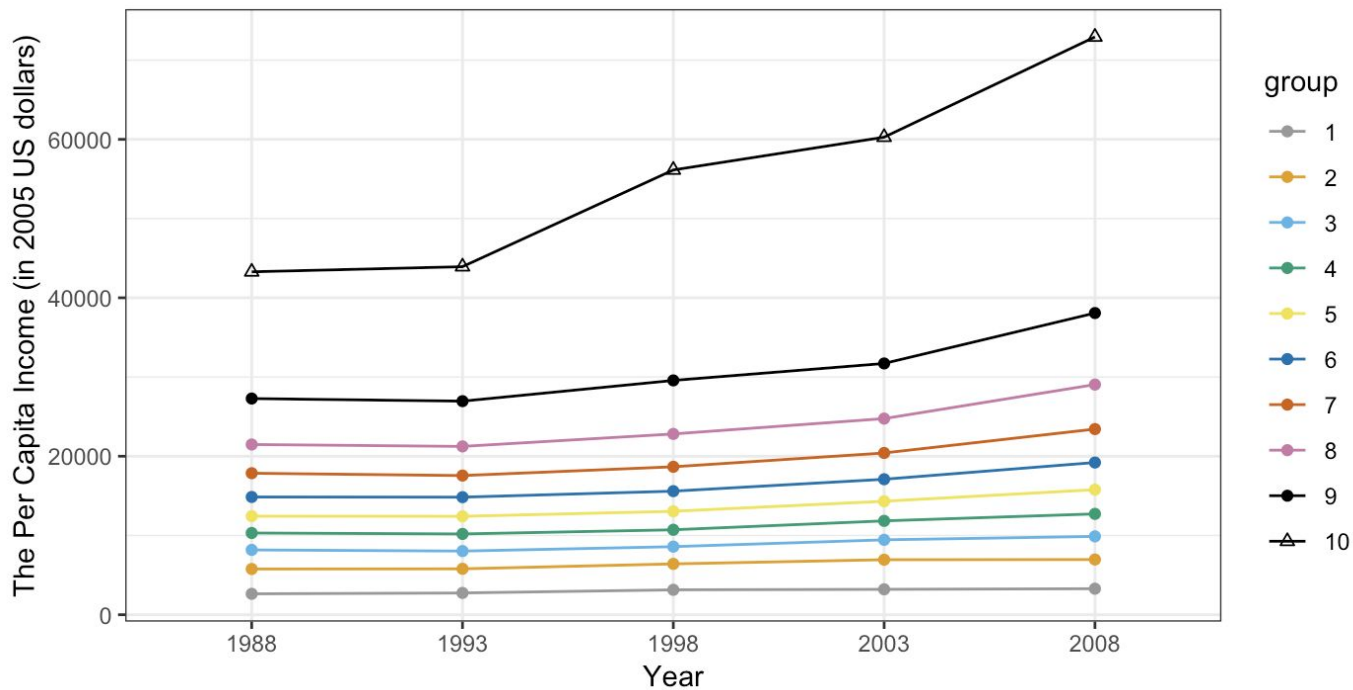
1. How has per capita income for each decile in the U.S. changed since 1988?

1988-2008 Per Capita Income for Each Decile in the U.S.

income decile group: a decile is 10% of the population.

"1": the 10% of individuals in the country with the lowest income

"10": the 10% of the individuals in the country with the highest income.



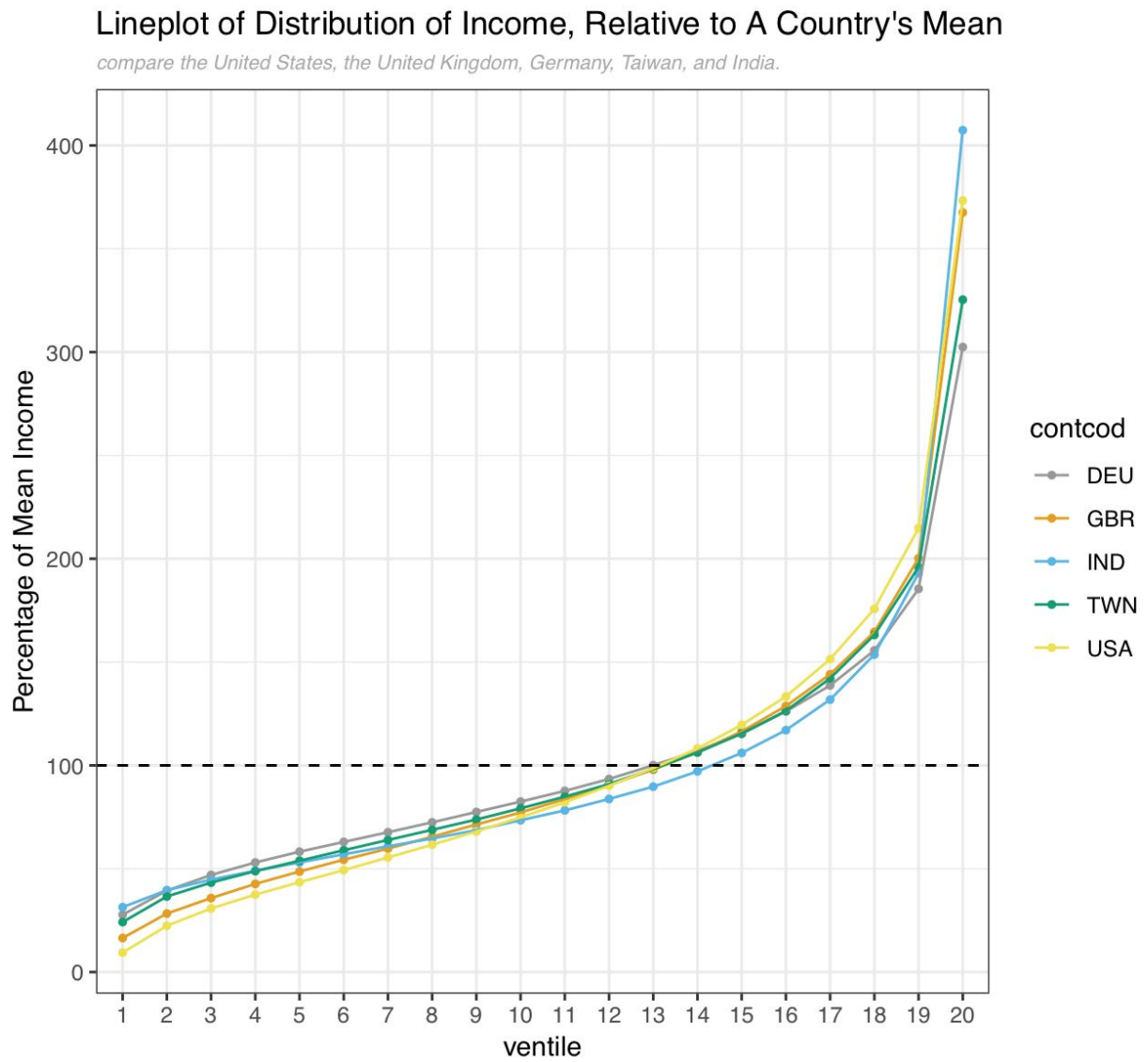
Group	1	2	3	4	5	6	7	8	9	10
(1). Total growth of per capita income until 2008, in percentage of 1988 income	24.5	20.7	23.1	24.8	27.1	29.5	33.4	36.7	41.3	68.5
(2). Average yearly growth rate of (1)	1.10	0.95	1.04	1.12	1.21	1.30	1.45	1.58	1.74	2.67

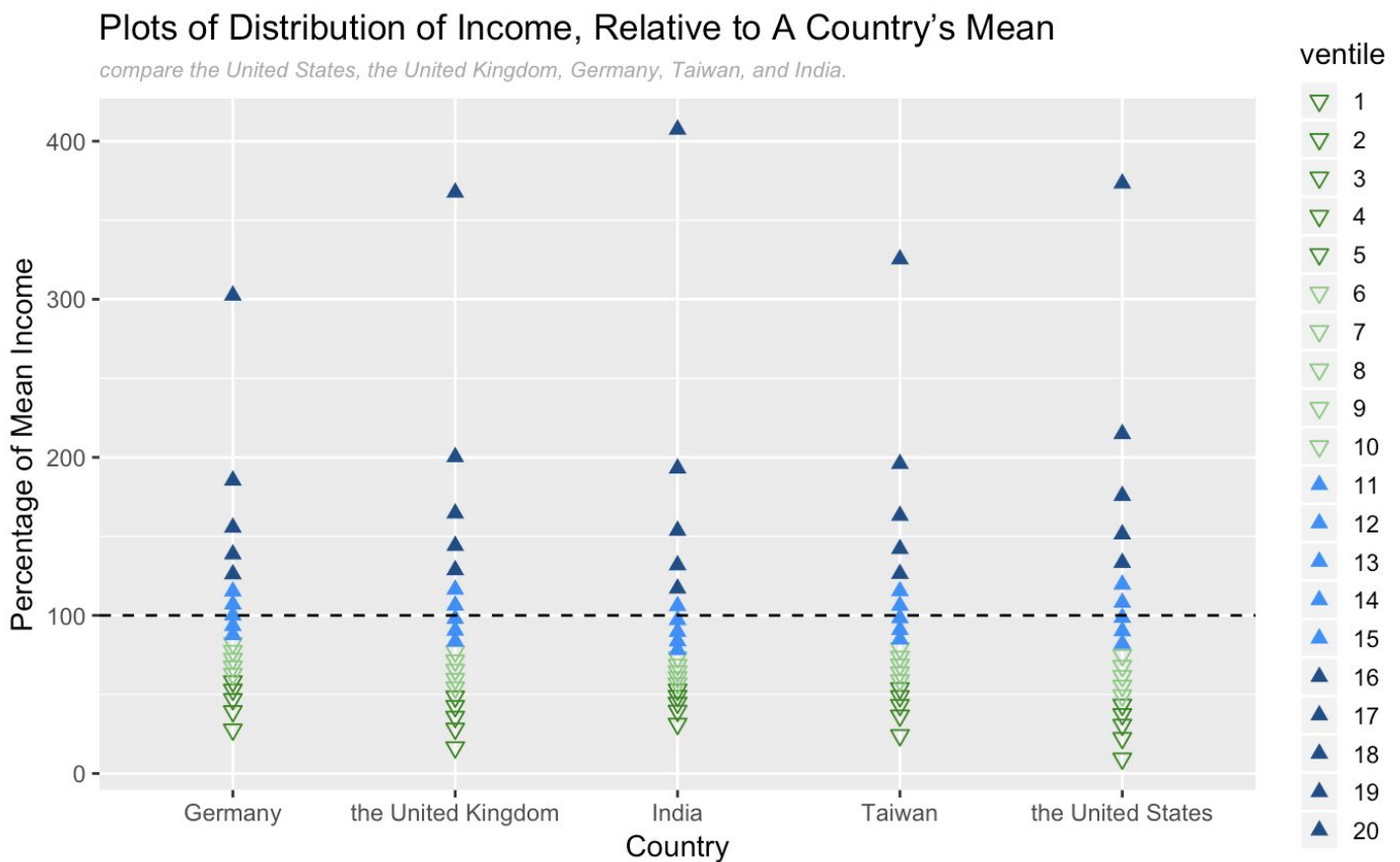
As shown in the graph, the total growth and the yearly growth rate in per capita income of correspondent decile group increases over the years and the growth rate of higher income decile group is more than the lower income decile group with an exception of the first 2 deciles of income groups.

Also we observed that, the per capita income of the top 10% income group increased nearly 70 percent from 1988 to 2008. However, in the same period, the per capita income for last 10% income group increased by only 25 percent.

To sum up: up to year 2008, even though the rich are getting richer and the poor are getting less poor, but the richest 5% contributes more than 30 percent of the US total income.

2. How does the present distribution of income, relative to a country's mean, differ between selected countries?



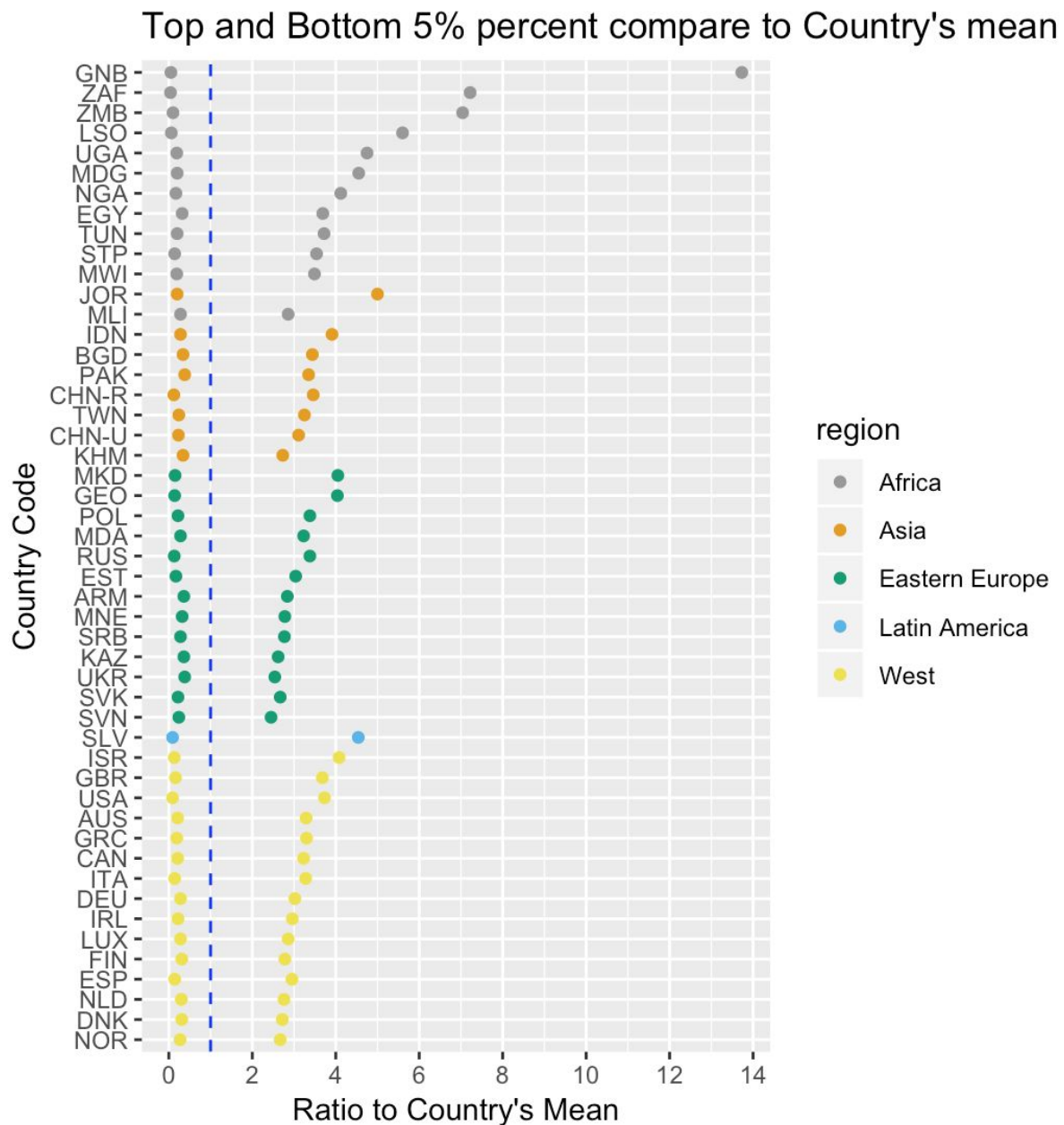


The line graph above shows the distribution of ventile income for each country related to country's mean. Since the selected five countries have different currency hence different income value. In order to remove this effect, we have normalized the income value of each ventile to percentage compare to this country's total income.

As we can observed from the graph above :

1. India is the country with most extreme income wealth gap, 70% of its people has income value lower than country's average. However, the top 5 percent has the most income amount compare to other four countries.
2. United State and United Kingdom has similar income wealth distribution throughout each ventile group. But, we observed that United State has slightly greater wealth gap than United Kingdom as the ventile groups for United State(USA) are lower than the country's mean and below United Kingdom(GBR), and vice versa for ventile groups above the country's mean.
3. Compare to other 4 countries, Germany (DEU) has the least income wealth gap; as the ventile groups below country's mean are generally above other 4 countries, and the top 5% ventile group has the least income value in comparison to other 4 countries. Moreover, Taiwan has similar trend with germany with slightly greater wealth gap. But has a lesser wealth gap than other four countries.

3. Is the percentage of income earned by the top 5 percent in a country related to mean income in that country? What about the percentage of income earned by the bottom 5 percent?



We approached this question by plotting all countries top 5 and bottom 5 percent income groups and their ratio to the world's average, since there are too many countries, we only picked the data from year 2010 to see whether we can detect some pattern from it.

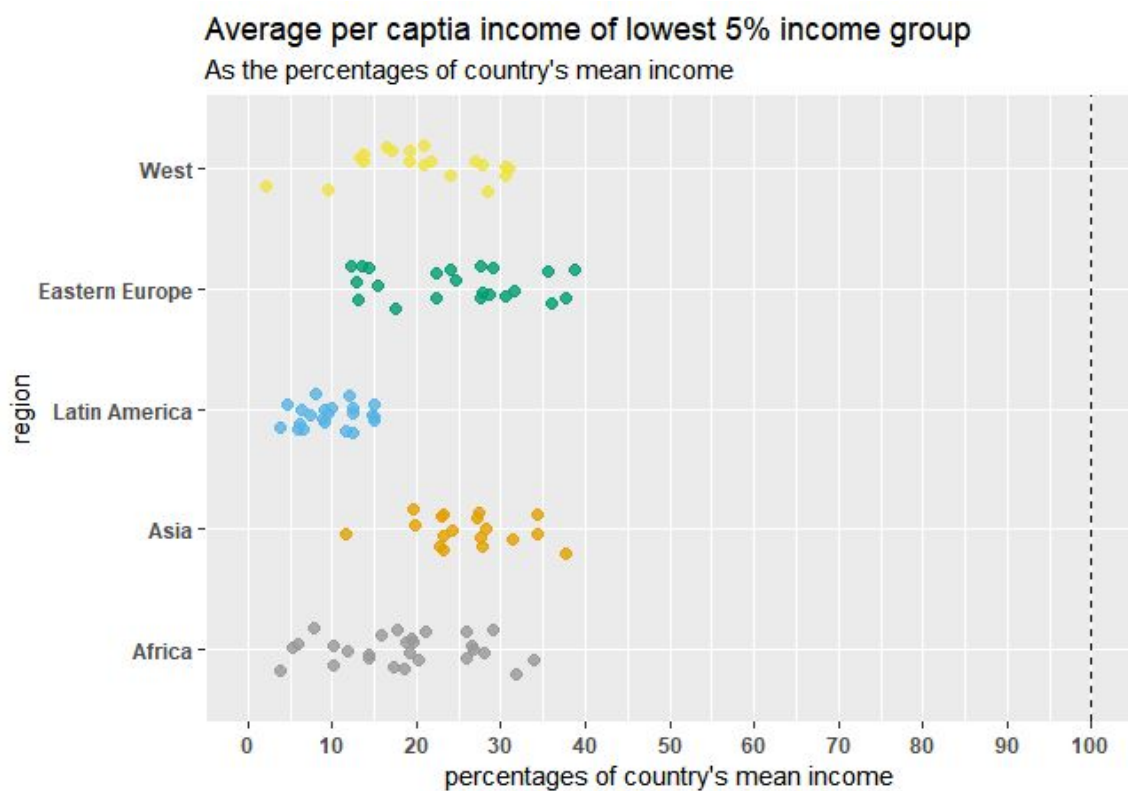
We observed from the above graph, the ratio is about 0.1-0.5 times the country's mean income for bottom 5 percent income group. Whereas for the top 5 percent, it varies from 2.5 times to almost 14 times of the country's mean income. Also, we noticed that, if a country belongs to a more developed regions, for instance, Eastern Europe and West (fully developed countries), the top 5 percent has income values that does not exceed 4 times of the country's mean. For countries in developing and underdeveloped regions, top 5 percent income group has much more various income ratios to mean. For example, for Africa, top 5 percent income ratio varies from Mali (MLI) to Guinea-Bissau (GNB) by 3 to 14 times the world mean income value.

In order to better understand the pattern, next we plotted the data from all the countries and grouped by regions.

For the lowest 5% income group, the per capita income were spread from 2.1% ~ 38.6% to the country's mean income.

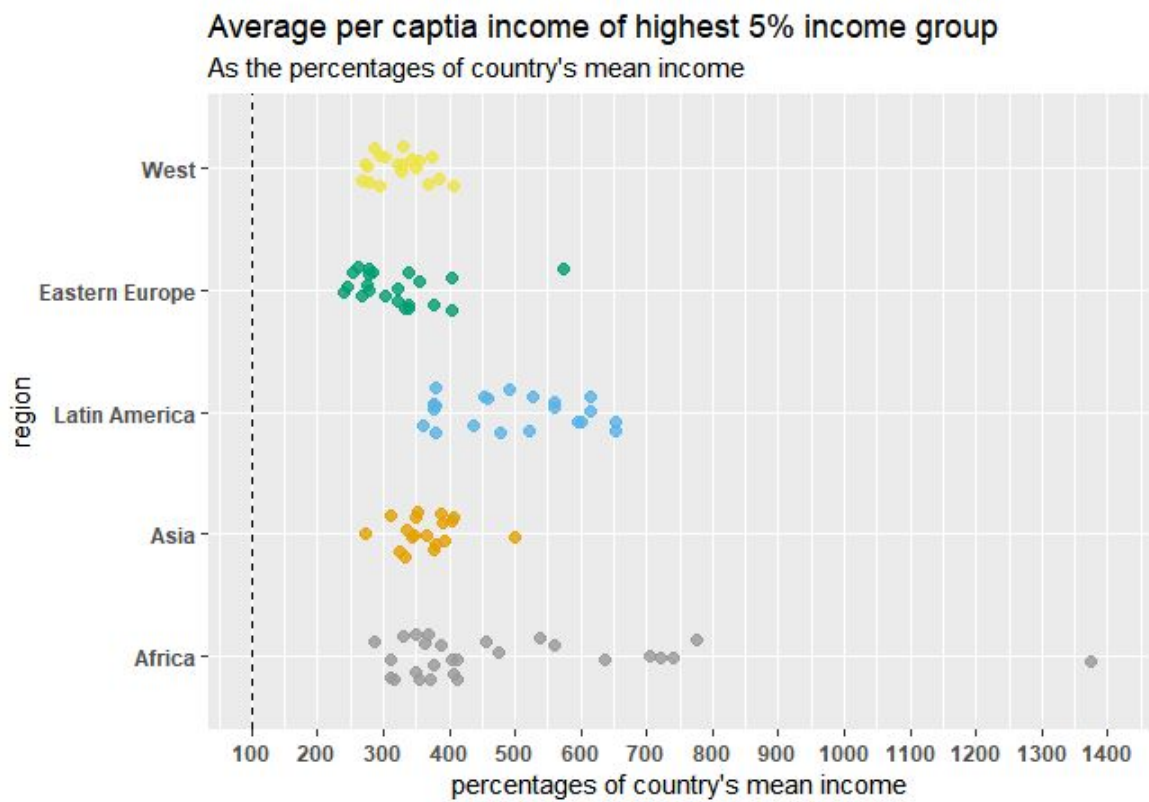
In view of regional difference,

all region except Latin America had the spread of country's income for nearly 30% to the mean income. For the Latin American countries, the per capita income of lowest 5% income group spread from 5% ~ 15% to the mean income. The lowest 5% income group in Eastern European and Asian countries tend to have higher income level (to the country's mean income) compare to the other regions and the lowest 5% income group in Latin American countries tend to have lower income level to the country's mean income.



The per capita income for the highest 5% income group spread more widely than the per capita income for the lowest 5% income groups. The whole range for the per capita income of highest 5% income group was 1134.2%. For most of the countries, the level per capita income of highest 5% income group was less than 400% and higher than 250% to the country's mean income level. For the countries which has highest income group whose per capita income is more than the 400% to

the country's mean income, most of them (30 out of 37) located in African and Latin American region. This also implies that the per capita income of highest 5% income group in Latin American and African regions spread more widely against other regions of the world.



=====END=====

R code for Q1=====

```
```{r}
```

```
library(tidyverse)
```

```
library(haven)
```

```
...
```

```
```{r}
```

```
WPID <- read_dta(file = "LM_WPID_web_2.dta")
```

```
w_var <- c("country", "contcod", "bin_year", "group", "RRinc", "RRmean")
```

```
WPID <- WPID[w_var]
```

```
WPID$bin_year <- as.factor(WPID$bin_year)
```

```
WPID$group <- as.factor(WPID$group)
```

```
summary(WPID)
```

```
str(WPID)
```

```
...
```

```
```{r}
```

```
Q1 <- subset(WPID, contcod == "USA", select = c(bin_year, group, RRinc))
```

```
summary(Q1)
```

```
q1_palette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00",
"#CC79A7", "#000000", "#000000")
```

```
p1_shape <- c(19, 19, 19, 19, 19, 19, 19, 19, 19, 2)
```

```
ggplot(Q1, aes(x = bin_year, y = RRinc, group = group, color = group)) + geom_line() + theme_bw() +
ggtitle('1988-2008 Per Capita Income for Each Decile in the U.S.') + labs(subtitle = "income decile group:
a decile is 10% of the population. \n“1”: the 10% of individuals in the country with the lowest income
\n“10”: the 10% of the individuals in the country with the highest income.") + xlab("Year") + ylab("The Per
Capita Income (in 2005 US dollars)") + theme(plot.subtitle=element_text(size=10, hjust=0, face="italic",
color="darkgrey")) + geom_point(aes(shape = group)) + scale_shape_manual(values=p1_shape) +
scale_color_discrete(name = "Income \nDecile \nGroup") + scale_colour_manual(values = q1_palette)
Q1_df <- Q1 %>% spread(Q1, key = group, value = c(RRinc))
...
```

R code for Q2=====

```
```{r fig.height=7, echo=FALSE}
```

```
vent <- read_dta(file = "ventile_2011_for_release_LCU.dta")
```

```
v_var = c("contcod", "ventile", "ventile_income")
vent = vent[v_var]
vent$ventile <- as.factor(vent$ventile)
summary(vent)
str(vent)
```

```
Q2 <- subset(vent, (contcod %in% c("USA","GBR","DEU", "TWN", "IND")))
Q2 <- Q2 %>%
  group_by(contcod) %>%
  mutate(avg = mean(ventile_income), percentage = (ventile_income/avg) * 100)
```

```
p2_shape = c(25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17)
p2_color = c("#999999", "#999999", "#999999", "#999999", "#999999", "#E69F00", "#E69F00",
"#E69F00", "#E69F00", "#E69F00", "#56B4E9", "#56B4E9", "#56B4E9", "#56B4E9", "#56B4E9",
"#009E73", "#009E73", "#009E73", "#009E73", "#009E73")
```

```
ggplot(Q2, aes(x = contcod, y = percentage, group = ventile, color = ventile)) + geom_point(aes(shape =
ventile), size=2) + scale_shape_manual(values=p2_shape) + theme_bw() + geom_hline(yintercept=100,
linetype="dashed", color = "black") + scale_colour_manual(values = p2_color) +
theme(legend.position="right",legend.key.size=unit(0.5,'cm')) + ylab("Percentage of Mean Income") +
xlab("Country") + scale_x_discrete(limit = c("DEU", "GBR", "IND", "TWN", "USA"), labels =
c("Germany","the United Kingdom","India", "Taiwan", "the United States")) + ggtitle('Dotplot of Distribution
of Income, \nRelative to A Country's Mean') + labs(subtitle = "compare the United States, the United
Kingdom, Germany, Taiwan, and India.") + theme(plot.subtitle=element_text(size=8, hjust=0, face="italic",
color="darkgrey"))
```

```
p21_color = c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442")
ggplot(Q2, aes(x = ventile, y = percentage, group = contcod, color = contcod)) + geom_point(size = 1) +
geom_line() + geom_hline(yintercept=100, linetype="dashed", color = "black") + ggtitle('Lineplot of
Distribution of Income, \nRelative to A Country's Mean') + labs(subtitle = "compare the United States
(contcod:USA), the United Kingdom (GBR), \nGermany (DEU), Taiwan (TWN), and India (IND).") +
theme_bw() + theme(plot.subtitle=element_text(size=8, hjust=0, face="italic", color="darkgrey")) +
ylab("Percentage of Mean Income") + scale_color_discrete(name = "Country", labels = c("Germany", "the
United Kingdom","India", "Taiwan", "the United States")) + scale_colour_manual(values = p21_color)
```

```
Q2_df <- Q2 %>% dplyr::select(contcod, ventile, percentage) %>% spread(Q2, key = ventile, value =
percentage)
```

```
...
```

```
# Code for reproduce graph in Q3
```

```
library(ggplot2)
library(foreign)
library(data.table)
```

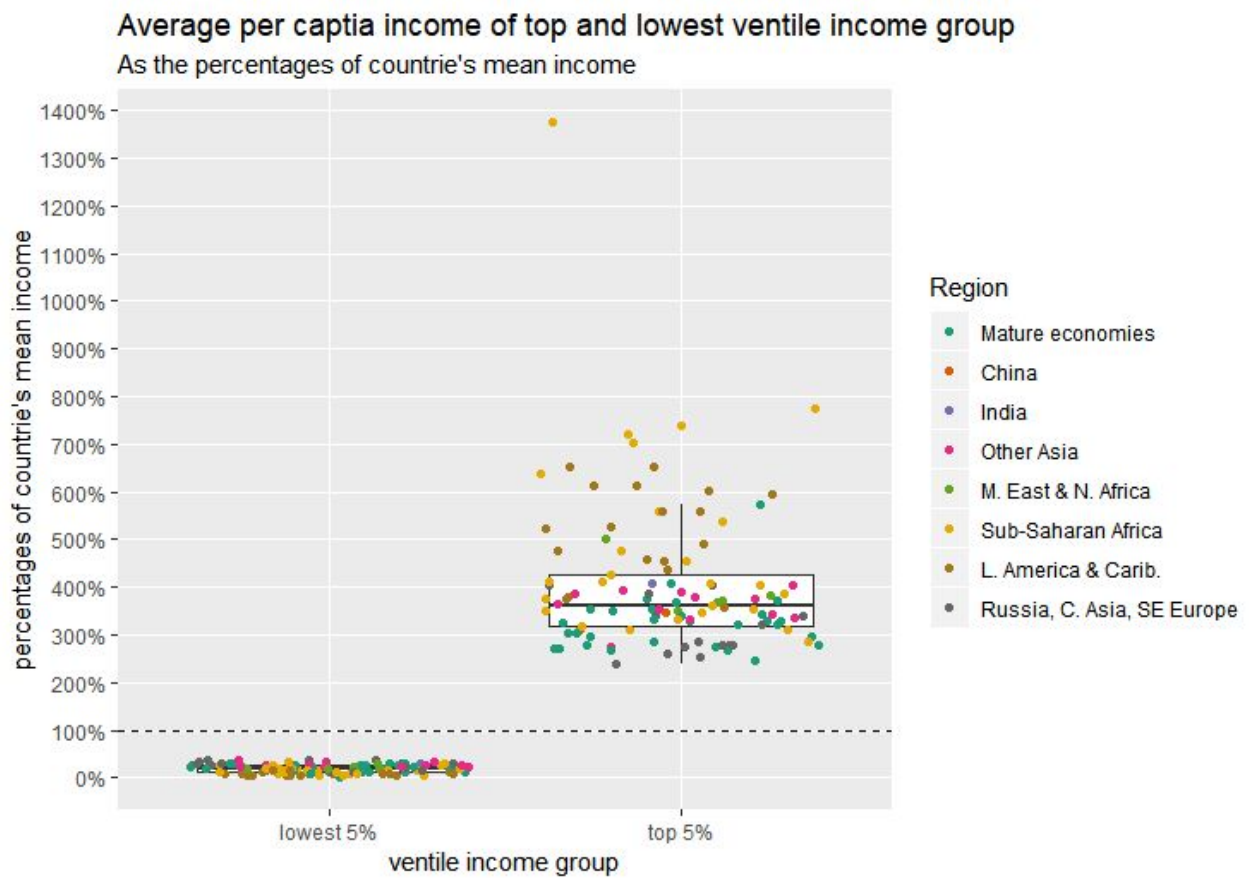


```
pals <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
ventile = read.dta('ventile_2011_for_release_LCU.dta')
wyd = read.dta('wyd_88_05_for_release.dta')
```

```
ventile = data.table(ventile)
ventile = merge(ventile, unique(wyd[,c('contcod','country','region')]), with=F, by='contcod')
ventile <- merge(ventile,
  ventile[,.(mean_income = mean(.SD[['ventile_income']]),
    sd_income = sd(.SD[['ventile_income']]), by=c('contcod')),
  by=c('contcod'))
ventile[,income:=ventile_income/mean_income*100]
ventile[,std_income:=(ventile_income-mean_income)/sd_income]
```

```
ggplot(ventile[ventile==1], aes(y=region, x=income, color=region))+
  geom_jitter(width=0, height=0.2, alpha=0.8, size=2)+
  geom_vline(xintercept=100, linetype='dashed')+
  scale_color_manual(values=pals, guide=F)+
  scale_x_continuous(limits = c(0,100), breaks = seq(0,100,10))+
  labs(title='Average per captia income of lowest 5% income group',
    subtitle = "As the percentages of country's mean income",
    x="percentages of country's mean income", y='region')+
  theme(axis.text = element_text(size=9,face='bold'))
```

```
ggplot(ventile[ventile==20], aes(y=region, x=income, color=region))+
  geom_jitter(width=0, height=0.2, alpha=0.8, size=2)+
  geom_vline(xintercept=100, linetype='dashed')+
  scale_color_manual(values=pals, guide=F)+
  scale_x_continuous(limits = c(100,1400), breaks = seq(100,1400,100))+
  labs(title='Average per captia income of highest 5% income group',
    subtitle = "As the percentages of country's mean income",
    x="percentages of country's mean income", y='region')+
  theme(axis.text = element_text(size=9,face='bold'))
```



Summary of bottom 5% income group-

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.147	12.930	19.903	20.064	27.615	38.648

Summary of top 5% income group-

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
239.2	317.1	362.4	402.4	426.6	1373.4

(finished)

EXTRA

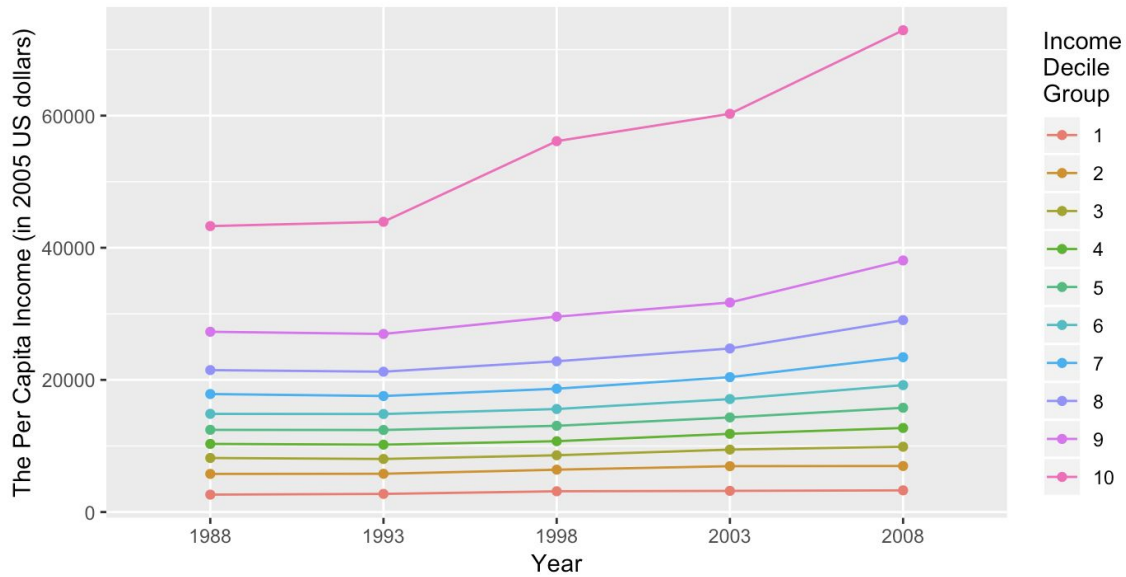
=====

1988-2008 Per Capita Income for Each Decile in the U.S.

income decile group: a decile is 10% of the population.

"1": the 10% of individuals in the country with the lowest income

"10": the 10% of the individuals in the country with the highest income.



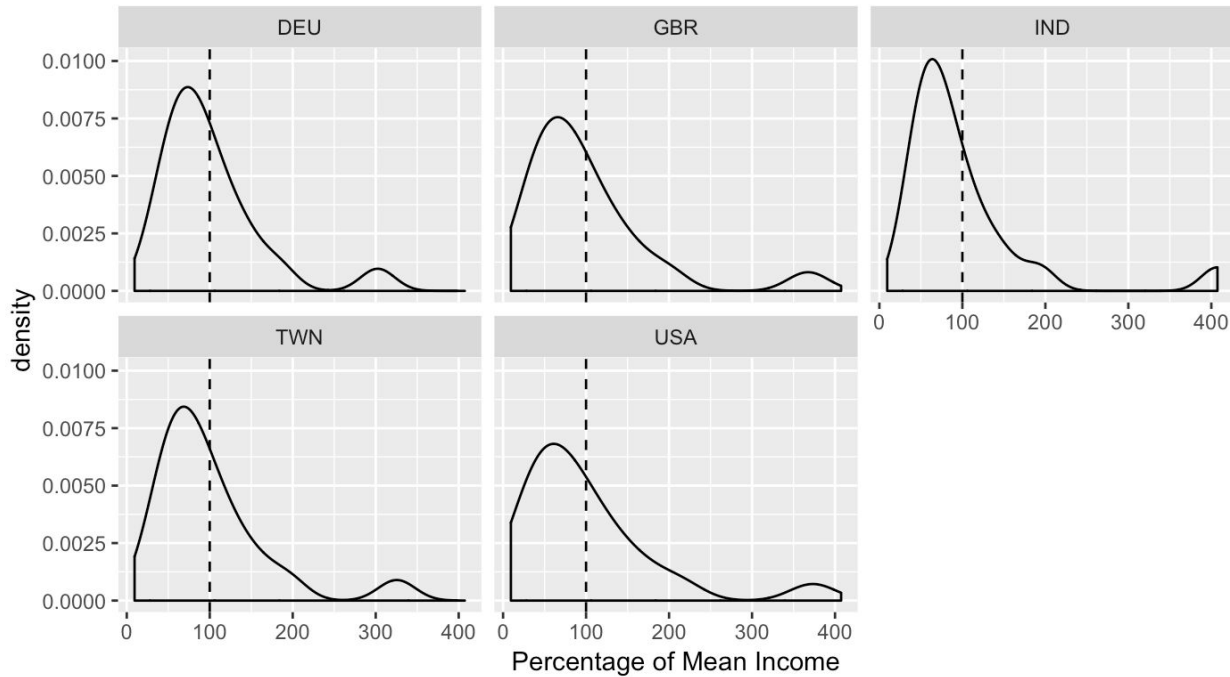
(As shown in the graph and the table above, per capita income for each ventile group has increased since 1988. However, the increasing amount for each ventil group various considerably. For ventil group 10, their per capita income increased the most amount; it has increased around 70 percentage. As the ventil group decreases, their increase amount almost decreases. Per capita income for ventil group 1 only increased by 25 percentage since 2008. To sum up: up to 2008, even though the rich are getting richer and the poor are getting less poor, but the richest 5% contributes more than 30 percent of the US total income)

Year	income decile group									
	1	2	3	4	5	6	7	8	9	10
1988	2636	5770	8180	10312	12441	14861	17860	21485	27285	43279
1993	2746	5792	8032	10194	12418	14836	17562	21242	26954	43930
1998	3142	6413	8592	10720	13053	15595	18670	22816	29571	56135
2003	3203	6941	9451	11843	14321	17098	20414	24753	31713	60274
2008	3283	6967	9885	12725	15782	19209	23430	29048	38077	72925

2. How does the present distribution of income, relative to a country's mean, differ between selected countries?

Density Plots of Distribution of Income, Relative to A Country's Mean

compare the United States (USA), the United Kingdom (GBR), Germany (DEU), Taiwan (TWN), and India (IND).



As shown in the facet density plots, all the selected countries ventile income distribution are heavily right skewed, and large group of people are making less than the country's mean, and there are small group of outliers for each country; there are small group of people making three to four times the country's mean.

income ventile group										
Country	1	2	3	4	5	6	7	8	9	10
DEU	27.7685	39.3802	46.9838	53.0025	58.2353	62.9572	67.6595	72.4273	77.444	82.4738
GBR	16.4715	28.295	35.7813	42.6458	48.6413	54.3577	59.7073	65.4795	71.3873	77.2153
IND	31.3841	39.6037	44.7367	49.0017	53.0196	56.9799	60.8311	64.5903	68.6656	73.4595
TWN	24.1885	36.5645	43.275	48.8407	53.8464	58.9229	63.8893	68.8148	73.7891	79.2051
USA	9.40577	22.3837	30.7534	37.4564	43.4611	49.3217	55.4301	61.6015	68.0386	74.8452

income ventile group										
Country	11	12	13	14	15	16	17	18	19	20
DEU	87.6542	93.3978	100.065	107.04	115.226	126.111	138.698	155.635	185.414	302.428
GBR	83.6493	90.4659	97.9283	106.363	116.449	128.695	144.074	164.588	200.234	367.572
IND	78.2131	83.7484	89.7148	97.1125	106.033	117.052	131.819	153.627	193.029	407.378

TWN	84.8589	90.92	98.3125	106.218	115.443	126.361	142.116	163.139	195.944	325.352
USA	82.1383	90.1341	98.6344	108.184	119.599	133.352	151.393	175.714	214.828	373.325

All five country had right skewed income distribution with roughly similar shape of distribution. However, in detail level, Germany and Taiwan were similar to each other in terms of the position of the lowest ventile group (located between 25~30% of mean income), the position of highest income group (located between 300~350% of mean income), and about 50% people gathered in the interval between 50~100% of mean income. In contrast, United Kingdom and United States were similar to each other in terms of higher income level for the highest ventile group (around 370%), lower income level for the lowest ventile group(<20%), and less people gathered around 50~100% income interval (about 40~45% of people). Compare to Germany-Taiwan, income distribution of UK-US were slightly widely spread. India's income distribution were different from all the others because of its more-centralized distribution and the ultra-high income level for the highest ventile group (>400%).

3. Is the percentage of income earned by the top 5 percent in a country related to mean income in that country? What about the percentage of income earned by the bottom 5 percent?

If so, what's the relationship, and does the relationship have a simple explanation, such as regional differences? Are there any outliers that require special explanation?

The Relationship between Percentage of Income Earned by the Top / Bottom 5 Percent and Mean Income in that Country

ventile:
a ventile is 5% of the population.
"1" means the 5% of individuals in the country with the lowest income
"20" means the 5% of the individuals in the country with the highest income.



```
lm(formula = avg ~ ventile_income, data = q3t)
      coef.est  coef.se
(Intercept)  -21199.81  97544.67
ventile_income    0.26    0.01
---
n = 111, k = 2
residual sd = 984186.49, R-Squared = 0.94
lm(formula = avg ~ ventile_income, data = q3b)
      coef.est  coef.se
(Intercept)  135435.46 136403.83
ventile_income    4.20    0.14
---
n = 111, k = 2
residual sd = 1387609.30, R-Squared = 0.89
```

We found a linear relationship between the percentage of income earned by the top 5 percent and the bottom 5 percent in a country related to mean income in that country.

Bottom 5 percent: *Average Per Capita Income* = $-2119.81 + 0.26 * bottom5income$

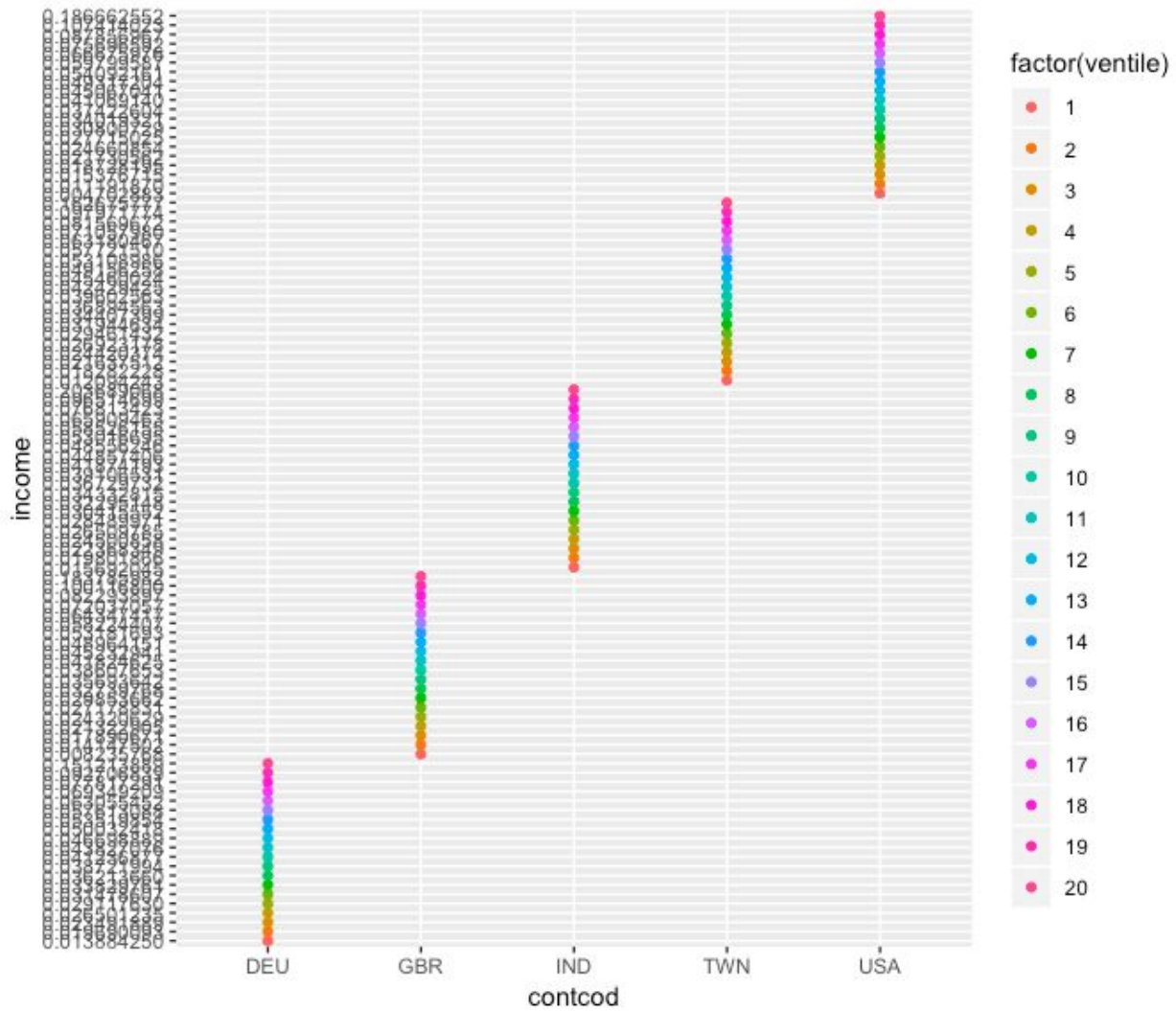
For the top 5 percent, for every 1 dollar increases in the income of the bottom 5 percent, the country's average per capita income goes up by 4.20 dollar. The model with 0.89 R squared value, meaning it explains 89% of the variation in the country's per capita income mean.

Top 5 percent: *Average Per Capita Income* = $135435.46 + 4.2 * top5income$

For the lower 5 percent: for every 1 dollar increases in the income of the top 5 percent, the country's average per capita income goes up by 0.26 dollar. Also, from the R squared value, we can see this model explains 94% of the the variation in the country's per capita income mean.

To sum up:

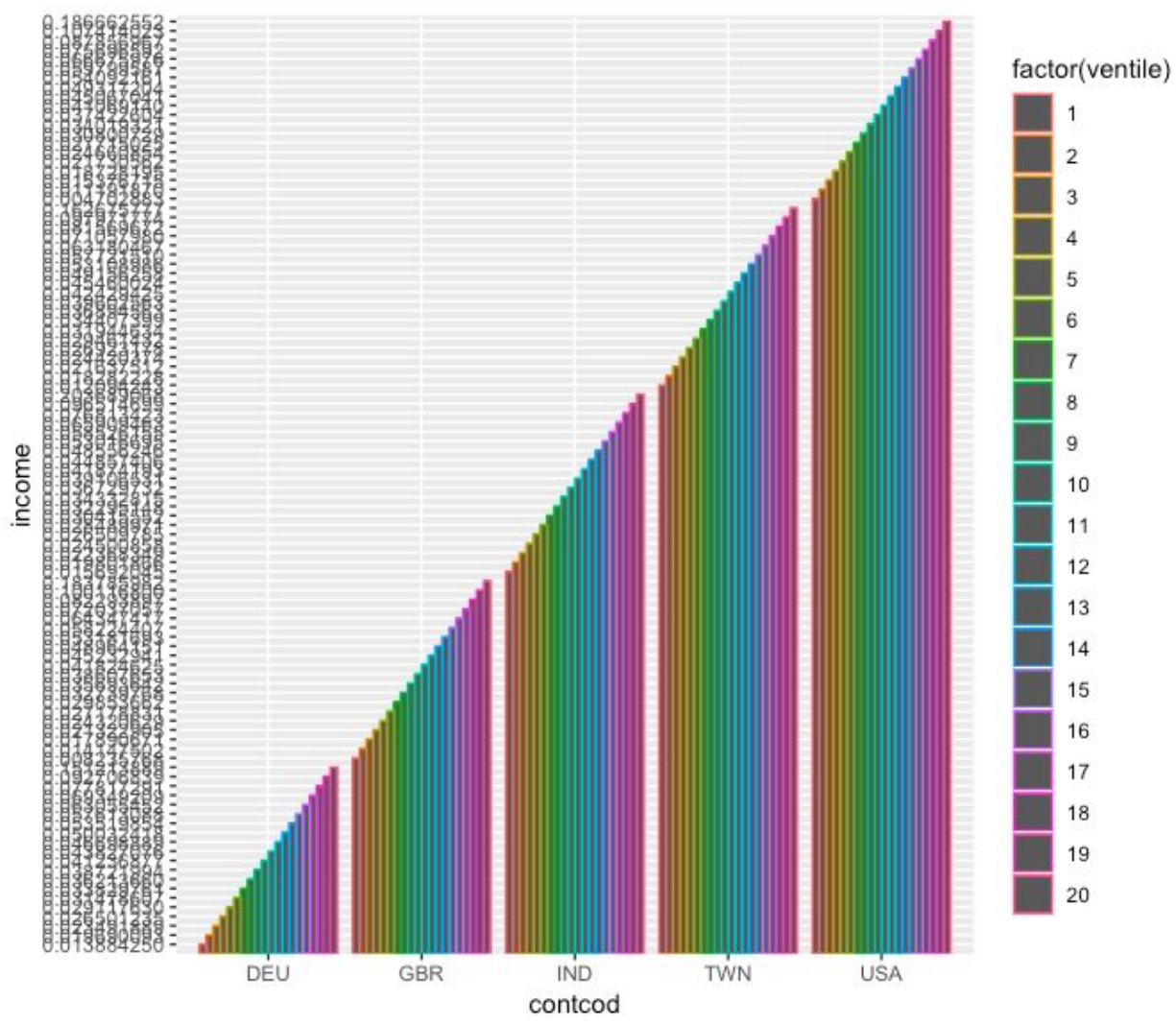
Question2

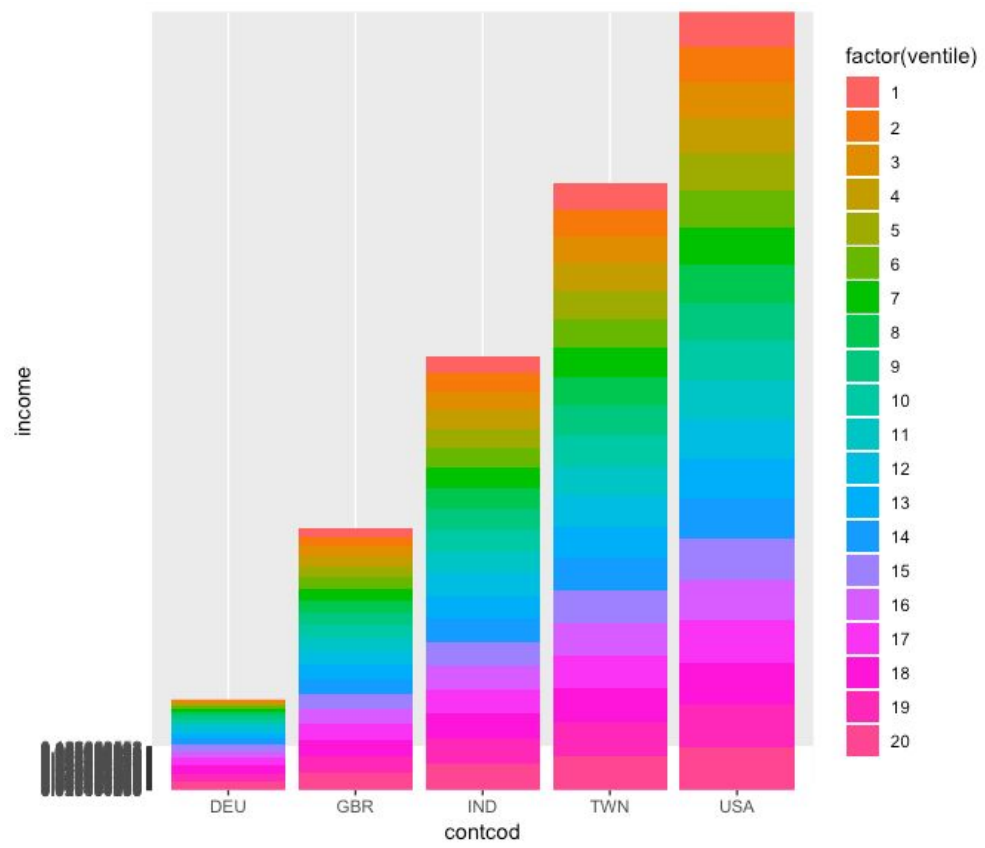


Is this graph good?

Which question is this graph for? Q2?

Or this one(Q2)





Or this one?(Q2)