

Assignment3

Research Question

Twenty Years since the end of Apartheid: Did the collapse of Apartheid play a significant role in reducing racial and social inequality in South Africa? Is post-apartheid South Africa better off or worse off than during the apartheid era?

Definition of Racial and social inequality

Before answer the research question, we need to clarify what the racial and social inequality actually is. In this article, we defined it as a differences among races in terms of unemployment rate, education disparity, and income distribution. The reason why we defined it by these three index is that in capitalism society, income level is the most fundamental index, which estimates the quality of life of an individual. In addition, we try to identify the driver of inequality of income level by investigate possible factors such as unemployment rate and education level.

Literature review

Before starting investigation, we need to look around previous researches which has been written by various ambitious researchers.

According to Leibbrandt, (see Leibbrandt (n.d.)) 1, Since the fall of Apartheid (1993~2008), overall (include all races) income inequality increased. The same is true among four major racial groups. 2, However, the major driver of inequality increase is intra-African inequality in South-Africa.

The reason why we choose this article as the first reference article for this article is that this is the most cited work in the South-African Inequality Study.

Data Gathering

According to Leibbrandt, inequality has been increased since the fall of Apartheid. We will test this assumption by using other data which is not used in the article.

We found the data of monthly earnings among races and gender. We tried to scraping the data from the website.

```
URL <- 'http://businesstech.co.za/news/wealth/131524/this-is-the-average-salary-in-south-africa-by-race'

RaceEarningsTable <- URL %>% read_html() %>%
  html_nodes('#container > div.content_holder > div.content > div.post_single > div.post_content') %>%
  html_table() %>%
  as.data.frame
```

RaceEarningsTable

		X1	X2	X3	X4	X5	X6	X7
## 1		Median	Median	Median	Mean	Mean	Mean	
## 2	Race	2003	2012	Increase	2003	2012	Increase	
## 3	White	14 468	16 581	15%	11 249	11 991	7%	
## 4	Asian/Indian	7 825	11 701	50%	5 264	8 993	60%	
## 5	Coloured	4 241	7 058	66%	2 437	3 897	60%	
## 6	Black African	4 059	5 445	34%	2 437	2 998	23%	

```
URL <- 'http://businesstech.co.za/news/wealth/131524/this-is-the-average-salary-in-south-africa-by-race'

GenderEarningsTable <- URL %>% read_html() %>%
  html_nodes('#container > div.content_holder > div.content > div.post_single > div.p
  html_table() %>%
  as.data.frame

GenderEarningsTable
```

```
##      X1      X2      X3      X4      X5      X6      X7
## 1      Median Median  Median Mean  Mean  Mean
## 2 Race  2003  2012 Increase 2003 2012 Increase
## 3 Male 5 963 8 299      39% 3 375 4 317      28%
## 4 Female 4 849 6 399      32% 2 435 3 118      28%
```

Data Cleaning and Merging

In this section, we will try to clean the data so that they can be statistical analysed.

Firstly, we use command “summary” to investigate the structure (class of variables, number of vectors) of data frames we got in the previous section.

```
summary(RaceEarningsTable)
```

```
##      X1      X2      X3
## Length:6      Length:6      Length:6
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##      X4      X5      X6
## Length:6      Length:6      Length:6
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##      X7
## Length:6
## Class :character
## Mode :character
```

```
summary(GenderEarningsTable)
```

```
##      X1      X2      X3
## Length:4      Length:4      Length:4
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##      X4      X5      X6
## Length:4      Length:4      Length:4
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##      X7
## Length:4
## Class :character
## Mode :character
```

As shown, every variables has a class of “characters” even though it represents numerical data.

The data we want to have is the mean of earnings among races and gender in 2003, 2012.

Firstly, we make TimeVector and IndividualVector to labeling the data.

```
TimeVector <- c("2003","2012")
IndividualVector <- c("White","Asian/Indian","Coloured","BlackAfrican","Male","Female")
```

Then, we make DataVector for each year.

```
unique(as.numeric(unlist(strsplit(gsub("[^0-9]", "", unlist(RaceEarningsTable$X5[3])), ""))))
```

```
## [1] 1 2 4 9
```

```
data2003 <- c(as.numeric(RaceEarningsTable$X5[3]),as.numeric(RaceEarningsTable$X5[4]),as.numeric(RaceEa
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
data2003
```

```
## [1] NA NA NA NA NA NA
```

```
#2003
```

```
GenderEarningsTable$X5[3] #2003male
```

```
## [1] "3 375"
```

```
GenderEarningsTable$X5[4] #2003female
```

```
## [1] "2 435"
```

```
#2012
```

```
GenderEarningsTable$X6[3] #2012male
```

```
## [1] "4 317"
```

```
GenderEarningsTable$X6[4] #2012female
```

```
## [1] "3 118"
```

```
#2003
```

```
RaceEarningsTable$X5[3]
```

```
## [1] "11 249"
```

```
RaceEarningsTable$X5[4]
```

```
## [1] "5 264"
```

```
RaceEarningsTable$X5[5]
```

```
## [1] "2 437"
```

```
RaceEarningsTable$X5[6]
```

```
## [1] "2 437"
```

```
#2012
```

```
RaceEarningsTable$X6[3]
```

```
## [1] "11 991"
```

```
RaceEarningsTable$X6[4]
```

```
## [1] "8 993"
```

```
RaceEarningsTable$X6[5]
```

```
## [1] "3 897"
```

```
RaceEarningsTable$X6[6]
```

```
## [1] "2 998"
```

Conduct basic descriptive statistics

The data we gathered in previous section partially statistically analysed (mean and median are already calculated). In this section, we try to figure out the trend of inequality graphically by using descriptive statistics.

Briefly describing

References

Leibbrandt, et al., <!--// --> M. n.d. "Trends in South African Income Distribution and Poverty Since the Fall of Apartheid." OECD Publishing. doi:<http://dx.doi.org/10.1787/5kmms0t7p1ms-en>.