HW7-Econometrics-96109674

Alireza Darvishi

2020-05-11

1

```r
library(tidyverse)
library(readxl)
library(haven)
library(knitr)
D_R97p3s03 <- read_dta("raw/D_R97p3s03.dta")
rwo_names = c("hhid" , "Product_Code" , "Method_of_preparation" , "Value")
names(D_R97p3s03) = rwo_names
D_R97p3s03[, 1:4] = apply(D_R97p3s03[, 1:4], 2 , function(x)
  as.numeric(x))
write_csv(D_R97p3s03, path = "processed/D_R97p3s03.csv")
```

2

```r
D_R97p3s03 %>% group_by(hhid) %>% summarise(n = n()) -> repeat_numbers
kable(repeat_numbers[1:10, ])
```

hhid	n
20001383905	2
20001383908	2
20001383914	2
20001383917	4
20001384014	2
20001384102	2
20001384105	1
20001384108	3
20001384111	2
20001384114	1

```r
max(repeat_numbers$n)
```

```
## [1] 22
```

This dataframe shows how many purchases a HH had in that mounth and maximum number of purchases in a mounth was 22

1

3

```r
read_and_clean_file = function (file) {
  dta = read_dta(file)
  names(dta) = c("hhid" , "Product_Code" , "Method_of_preparation" , "Value")
  dta[, 1:4] = apply(dta[, 1:4], 2 , function(x)
    as.numeric(x))
  write_csv(dta, path = paste0("processed/", substr(
    file, start = 5, stop = 14
  ), ".csv"))
  return(dta)
}
files <-
  list.files(
    path = "raw",
    pattern = "*p3s03.dta",
    full.names = TRUE,
    recursive = FALSE
  )
data_list = lapply(files, read_and_clean_file)
```

4

```r
full_dataset = data.frame()
for (i in 1:length(files)) {
  year = 1300 + as.numeric(substr(files[i], 8, 9))
  urban = ifelse(substr(files[i], 7, 7) == "U", yes = 1 , no = 2)
  data_list[i] %>% data.frame() %>% group_by(hhid) %>%
    summarise(
      total_spending = 12 * sum(Value) ,
      year = year ,
      urban = urban
    ) %>% rbind(full_dataset , .) -> full_dataset
}
full_dataset = drop_na(full_dataset)
attr(full_dataset, "variable.labels") =
  c("Household ID",
    "total money spent on clothing in that year",
    "Year" ,
    "Urban = 1 Rural = 2")
```

5

```r
weights <- read_dta("raw/weights.dta")
CPI <- read_excel("raw/CPI.xlsx")
names(CPI) = c("year", "urban", "CPI90")
CPI$year = CPI$year + 1300
full_dataset_cpi = merge (full_dataset , CPI, by = c("year", "urban"))
full_dataset_cpi_weights = merge (full_dataset_cpi , weights[-c(4:6, 8:10)] ,
                                  by =c("year", "urban", "hhid"))
```

```
full_dataset_cpi_weights %>% mutate(
  real_spending = 100 * total_spending / CPI90 ,
  real_spending_per_capita = real_spending / hh_N) %>%
  group_by(urban, year) %>%
  summarise(real_spending_per_capita_weighted =
            sum( real_spending_per_capita * hh_weight)) %>%
  ungroup() %>% mutate(urban =ifelse(urban == 1, "Yes", "No")) -> real_spending_data
weights %>% group_by(year) %>%
  summarise(full_weight = sum(hh_weight)) -> final_weights
merge(real_spending_data , final_weights) %>%
  mutate(real_spending_per_capita =
         real_spending_per_capita_weighted /full_weight) %>%
  ggplot() + geom_col(position = "dodge" ,
                      aes(x = year , y = real_spending_per_capita , fill = urban)) +
  scale_x_continuous(breaks = 1384:1397)  +
  labs(y = "real spending per capita" ,
       title = "real spending per capita in iran years 1384-1397") +
  theme(plot.title = element_text(hjust = 0.5))
```
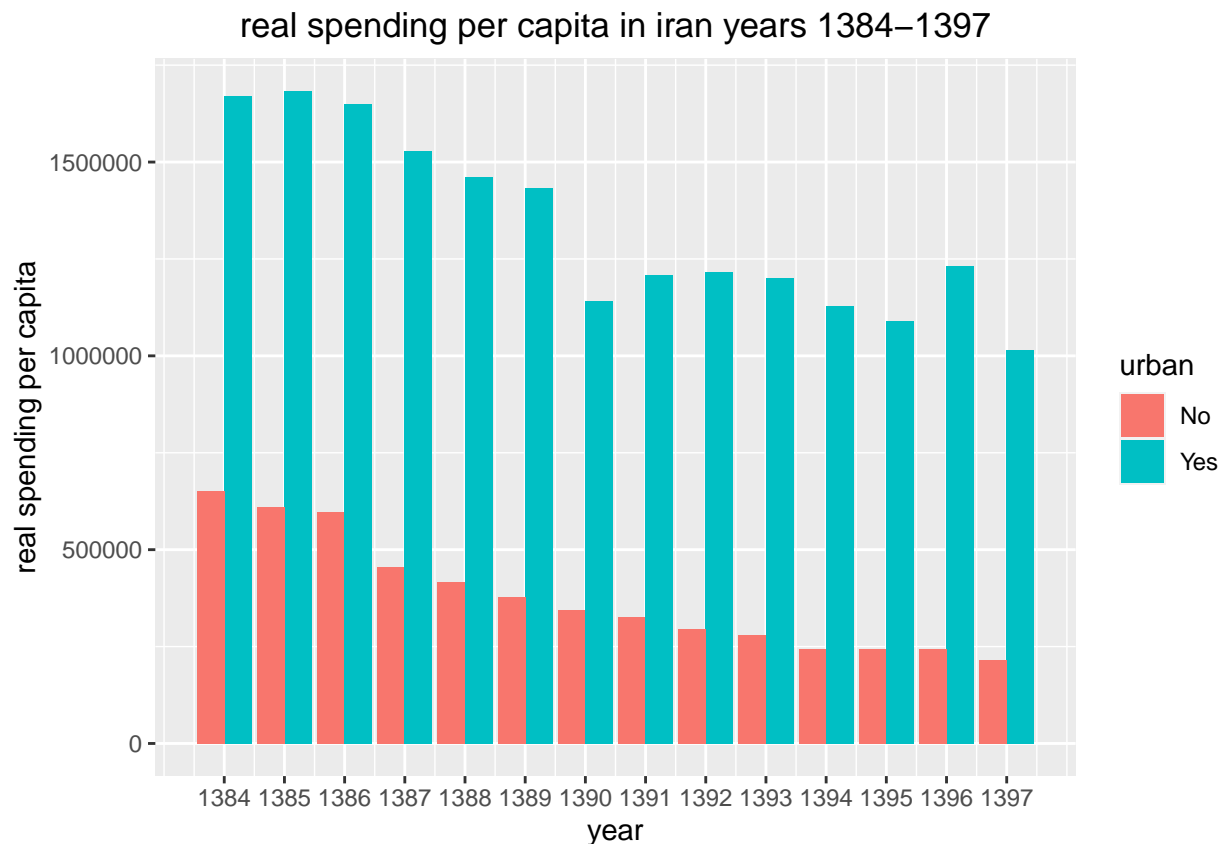

6

```
full_dataset %>% group_by(year) %>%
  summarise(spent_on_clothing = n()) -> spent_on_clothing
weights %>% group_by(year) %>% summarise(total_hh = n()) -> total_hh
merge(spent_on_clothing , total_hh , by = "year") %>%
```

```
mutate(percent = 100 * spent_on_clothing / total_hh) %>%
ggplot() + geom_col(aes(x = year , y = percent), fill = "blue") +
coord_cartesian(ylim =c(40, 80)) +
labs(title = "Percent of HHs spent on cloth") +
theme(plot.title = element_text(hjust = 0.5)) +
scale_x_continuous(breaks = 1384:1397)
```


Percent of HHs spent on cloth