Eurostat EDA

Sindre H. Øveraas, Alen Colacovic & Sebastian M. Fløysand

Eurostat EDA

Sub-National GDP

Country codes: BE - Belgium, BG - Bulgaria, HR - Croatia, IT - Italy, AT - Austria, SE - Sweden, RS - Serbia

```r
library(tidyverse)
library(vtable)
library(dineq)
library(dplyr)


library(readr)
GDP <- read.csv('nama_10r_3gdp__custom_3564935_linear.csv')
Population <- read_csv("demo_r_pjanaggr3__custom_3579517_linear.csv")
```

```
Rows: 5369 Columns: 10
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (7): DATAFLOW, LAST UPDATE, freq, unit, age, geo, OBS_FLAG
dbl (2): TIME_PERIOD, OBS_VALUE
lgl (1): sex

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
gdpdata <- GDP %>%
  rename(Year = TIME_PERIOD, GDP = OBS_VALUE, Region = geo)

populationdata <- Population %>%
```

```r
  rename(Year = TIME_PERIOD, Population = OBS_VALUE, Region = geo)

GDP_Per_Capita <- gdpdata %>%
  left_join(populationdata, by=c("Region", "Year")) %>%
  select(Region, Year, GDP, Population) %>%
  mutate(
    GDP_capita = (GDP * 1000000)/Population
  )
```

```
      GDP                 Population            GDP_capita
Min.   :     74.55   Min.   :   20320   Min.   : 1087
1st Qu.:   1738.28   1st Qu.:  164518   1st Qu.:17180
Median :   5614.05   Median :  273920   Median :25185
Mean   :  10238.24   Mean   :  406217   Mean   :24191
3rd Qu.:  10640.23   3rd Qu.:  429030   3rd Qu.:31351
Max.   : 181212.88   Max.   : 4355725   Max.   :72062
                     NA's   :771        NA's   :771
```

```r
GDP_Per_Capita$GDP_capita <- as.numeric(GDP_Per_Capita$GDP_capita)
GDP_Per_Capita$Population <- as.numeric(GDP_Per_Capita$Population)
```

```r
gini.wtd(GDP_Per_Capita$GDP_capita, weights = GDP_Per_Capita$Population)
```

```
[1] 0.2603924
```

```r
GDP_Per_Capita <- GDP_Per_Capita %>%
  mutate(NUTS2 = substr(GDP_Per_Capita$Region,1,4))

GDP_Per_Capita <- GDP_Per_Capita %>%
  mutate(NUTS = substr(GDP_Per_Capita$Region,1,2))

GDP_Per_Capita %<>%
  group_by(NUTS2) %>%
  na.exclude(GDP_Per_Capita) %>%
  mutate(GINI_NUTS2 = gini.wtd(GDP_capita, weights = Population)) %>%
  ungroup()
```
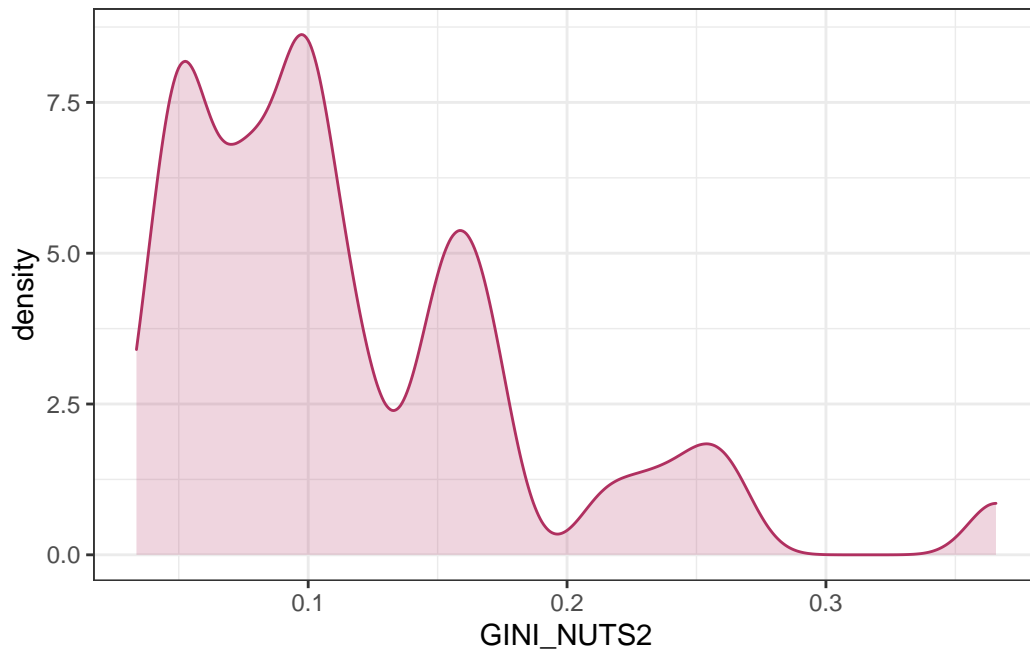
```r
summary(GDP_Per_Capita[8])
```

```
   GINI_NUTS2
Min.    :0.03367
1st Qu.:0.07065
Median :0.09839
Mean    :0.11800
3rd Qu.:0.15525
Max.    :0.36569
```


`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.