Exploring dataset "bilanciFVG"

This section is dedicated to load and preprocess financial statement data for the dataset *imprese-fvg*. The relevant file is "_DATA/imprese-fvg/bilanci-fvg.csv".

The relevant file is *bilanci-fvg.csv*. Each observation is a summary of balance sheet data (bsd) of a company (identified by *cf*) for a given year. Column labels need some improvement to remove whitespaces and possibly short english names.

```
bsd <- read_delim( paste0(pathRawData,"imprese/bilanci-fvg.csv") )
```

```
## Rows: 125617 Columns: 18

## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (16): cf, cia, Totale attivo, Totale Immobilizzazioni immateriali, Credi...
## dbl  (2): rea, anno

##
## 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.
```

```
spec(bsd) # tydiverse for str(companies)
```

```
## cols(
##   cf = col_character(),
##   cia = col_character(),
##   rea = col_double(),
##   anno = col_double(),
##   `Totale attivo` = col_character(),
##   `Totale Immobilizzazioni immateriali` = col_character(),
##   `Crediti esigibili entro l'esercizio successivo` = col_character(),
##   `Totale patrimonio netto` = col_character(),
##   `Debiti esigibili entro l'esercizio successivo` = col_character(),
##   `Totale valore della produzione` = col_character(),
##   `Ricavi delle vendite` = col_character(),
##   `Totale Costi del Personale` = col_character(),
##   `Differenza tra valore e costi della produzione` = col_character(),
##   `Ammortamento Immobilizzazione Immateriali` = col_character(),
##   `Utile/perdita esercizio ultimi` = col_character(),
##   `valore aggiunto` = col_character(),
##   tot.aam.acc.svalutazioni = col_character(),
##   `(ron) reddito operativo netto` = col_character()
## )
```

```
bsd <- bsd %>%
        rename(year = anno) %>%
        rename(totEquity = `Totale patrimonio netto`) %>%
        rename(totAssets = `Totale attivo`) %>%
        rename(totIntang = `Totale Immobilizzazioni immateriali`) %>%
        rename(staffCost = `Totale Costi del Personale`) %>%
        rename(turnover  = `Ricavi delle vendite`) %>%
        select(cf,year,turnover, totAssets, totIntang, staffCost  )
```

```
bsd <- bsd %>%
  mutate(across(everything(), gsub, pattern = "[.]", replacement = "")) %>%
  mutate(across(everything(), gsub, pattern = ",", replacement = ".")) %>%
  mutate(across(.cols = 2:6, .fns = as.numeric))
```

```
bsd %>% write_csv(paste0(pathTidyData,"bsd.csv"))
```

There are 18 columns but in this project we will use only 4, namely "cf", "year", revenues" and "staff cost". Data should be loaded as string and then converted taking into account some issues with format of numerical variables.

To convert bsd*revenuesandbsd*staffcost to numbers, we need to remove the "." used as thousand separators, and replace "," with "." as a decimal separator.

We will focus the analysis on a list of companies that are tenats at Area Science Park. The list is available in the file "data/imprese-fvg/area-tenants.csv" so we can load it in al list ("filter") and use it to subset *bsd*.

```
tenants <- read_delim( paste0(pathRawData,"area-science-park/tenants.txt") ) %>%
  select(cf)
```

```
## New names:
## * `` -> ...7
```

```
## Rows: 68 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (6): insediati, Ente/Azienda, cf, DENOMINAZIONEiifvg, Campus, Addetti (b...
## lgl (1): ...7
```

```
##
## 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.
```

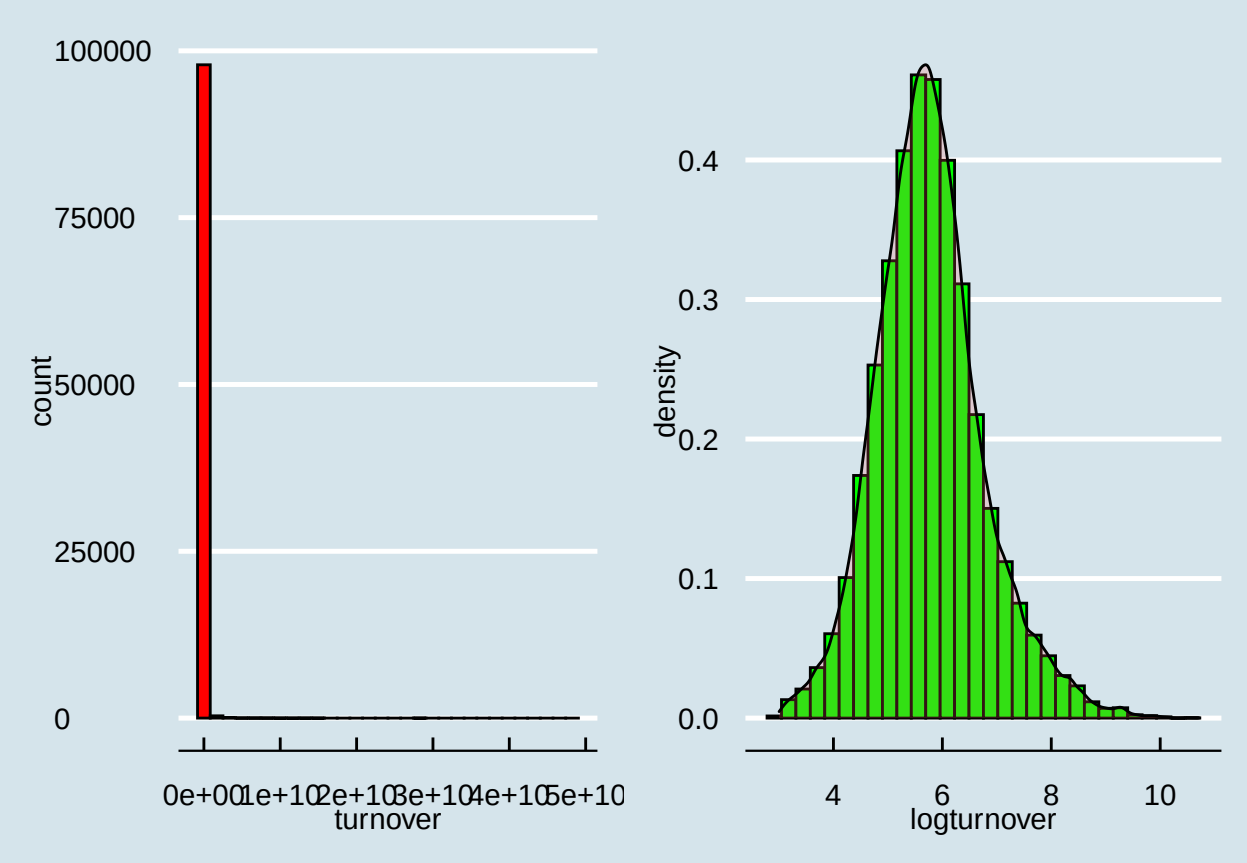
```
tens = c(tenants$cf)
bsd_tenants <- bsd %>% subset(cf %in% tenants$cf) %>%
                mutate(cf = as.factor(cf)) %>% drop_na()
```

The variable bsd$revenues spans from 0 to 1e9, so it is more convenient to work with log10

```
library(ggplot2)
library(ggpubr)
bsd3 <- bsd %>% subset(turnover > 1000) %>% subset(year = 2019)
bsd3$logturnover <- log10(bsd3$turnover)
# hist(bsd$turnover)
# hist(bsd$logturnover)
h1 <- ggplot(bsd3, aes(x=turnover)) + geom_histogram(color="black", fill="red")
h2 <- ggplot(bsd3, aes(x=logturnover)) + geom_histogram(color="black", fill="green", aes(y=..density..))
figure <- ggarrange(h1, h2)#labels = c("linear", "log"), ncol = 2, nrow = 1)
```

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

```
figure
```

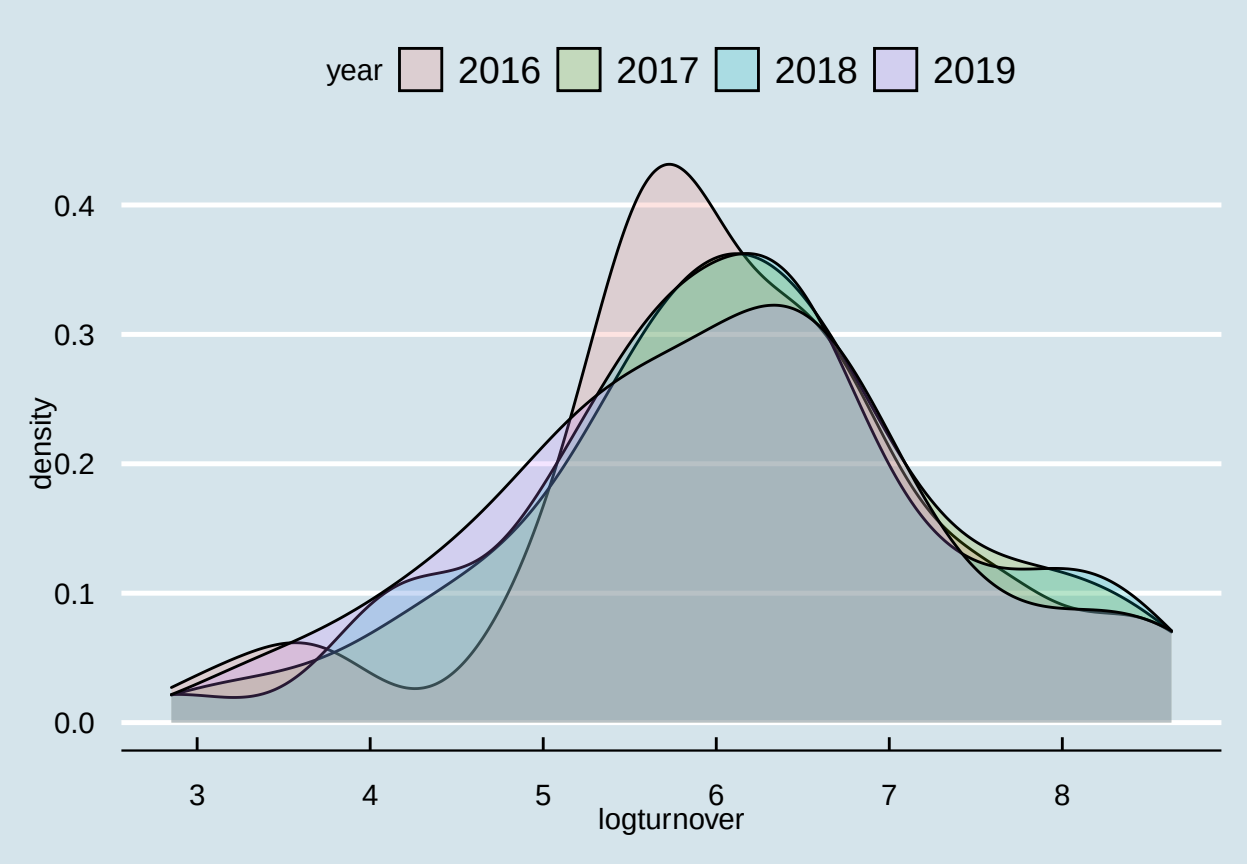



```
bsd_tenants$logturnover <- log10(bsd_tenants$turnover)

tmp <- bsd_tenants  %>%
              subset(year >= 2016) %>%
              mutate(year = as.factor(year))

figure <- ggplot(tmp, aes(x=logturnover,fill=year)) + geom_density(alpha=.2)
figure
```

```
## Warning: Removed 10 rows containing non-finite values (stat_density).
```


year
2016
2017
2018
2019
density
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
logturnover