

(Fast) Introduction to R

Jump into a notebook

Miguel Portela

20 Março 2022

My beamer

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis et elit blandit, ultrices nulla a, volutpat lacus. Present arcu risus, varius id varius commodo, auctor ac purus. Nam quis turpis a eros tempus molestie lobortis sed quam. Maecenas volutpat lacus in enim iaculis, scelerisque aliquam arcu rhoncus. Cras arcu nisi, accumsan vel accumsan sed, imperdiet eu sem. Vivamus ultrices convallis mi, ut vestibulum nibh imperdiet eu. Ut ac nisi lectus. Aenean hendrerit, nisi at commodo vulputate, sem neque sodales ante, id facilisis nibh dui et odio.

Warning: package 'stargazer' was built under R version 4

Outline

- ① Motivation
- ② Data
- ③ Conceptual discussion

3. Import data (from an excel file)

Load your data using point and click

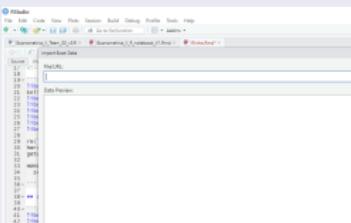


Figure 1: Point and click

which corresponds to the following code

```
nlswork <- as.data.frame(read_excel("nlswork.xlsx"))
# nlswork <- read_dta("nlswork.dta") # in case you have a Stata file
```

3.2. Define the object as a DataFrame

```
df <- as.data.frame(nlswork)
```

4. Data manipulation – check the pipe operator, %>%

4.1. Select a subset of variables

```
nlswork_s <- nlswork %>%
  select(idcode, ln_wage)
```

4.2. Rename variables

```
nlswork_r <- nlswork %>%
  rename(cae = ind_code)
```

4.3. Filter a subset of observations

```
nlswork_f <- nlswork %>%
  filter(age > 40)
```

4.4. Mutate: create variables

```
# nlswork_m <- nlswork %>%
#   mutate(ln_asd=ln(age))
```

5. Descriptive statistics

```
# install.packages("stargazer")
library(stargazer)
df %>%
  count()
```

```
##      n
## 1 28534
```

```
summary(df)
```

	idcode	year	birth_yr	age
## Min.	: 1	Min. :68.00	Min. :41.00	Min.
## 1st Qu.	:1327	1st Qu.:72.00	1st Qu.:46.00	1st Qu.
## Median	:2606	Median :78.00	Median :48.00	Median
## Mean	:2601	Mean :77.96	Mean :48.09	Mean
## 3rd Qu.	:3881	3rd Qu.:83.00	3rd Qu.:51.00	3rd Qu.
## Max.	:5159	Max. :88.00	Max. :54.00	Max.
##				NA's

Shorter statistics

Statistic	Min	Mean	St. Dev.	Max
-----------	-----	------	----------	-----

Collage Graduate	0	0.17	0.37	1
------------------	---	------	------	---

Experience	0.00	6.22	4.65	28.88	Hours	1	36.56	9.87	168
------------	------	------	------	-------	-------	---	-------	------	-----

5.1. Export descriptive statistics table to html, with 2 digits

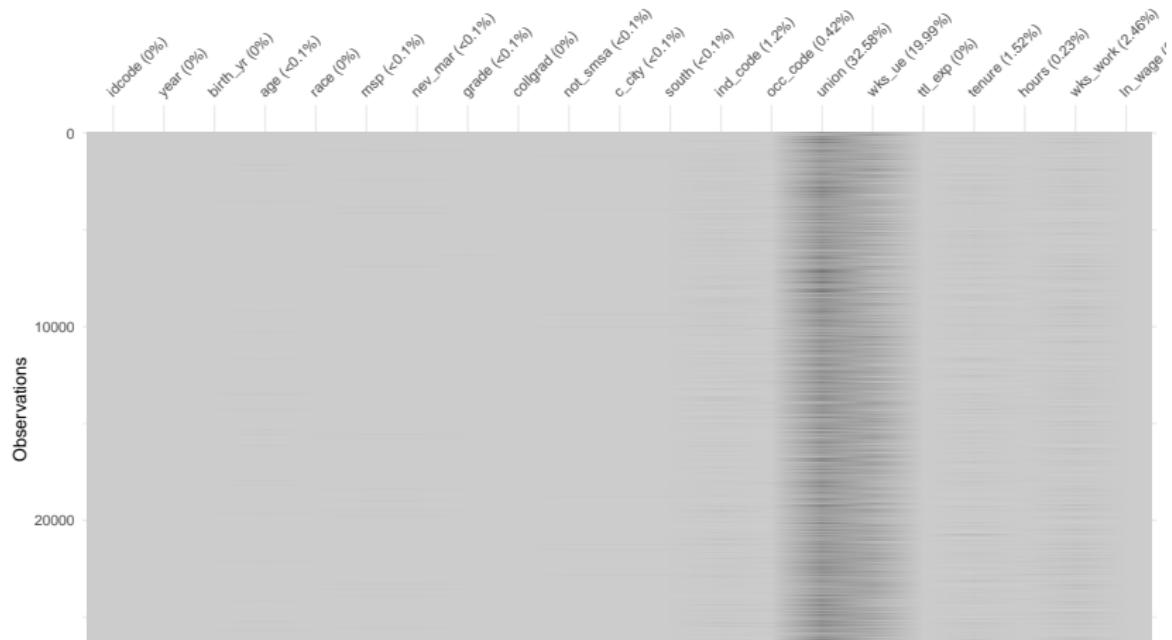
```
nlswork %>%
  dplyr::select(age, collgrad, ttl_exp, union, hours) %>%
  stargazer(title="Shorter statistics",
            type= "text", out = "Statistics_output.html",
            digits = 2)
```

```
##
## Shorter statistics
## =====
## Statistic   N      Mean    St. Dev.  Min   Max
## -----
```

Collage Graduate	0	0.17	0.37	1					
Experience	0.00	6.22	4.65	28.88	Hours	1	36.56	9.87	168

6. Visualize missing information:

```
# install.packages("visdat")
# install.packages("naniar")
library(naniar)
vis_miss(nlswork)
```



7. Visualisation to explore your data

7.1. Categorical variable

My histogram

