Study of French labour market and inequalities

L. Insolia, J. Kim and Y. Yeghikyan

SNS

March 14, 2018

— Midterm results —

Objectives

- Structure of French labour market
- Inequalities (in terms of salary):
  - ages
  - gender
  - job categories
  - spatial distribution
- Firms' distribution
- Exploratory analyses

Methodology

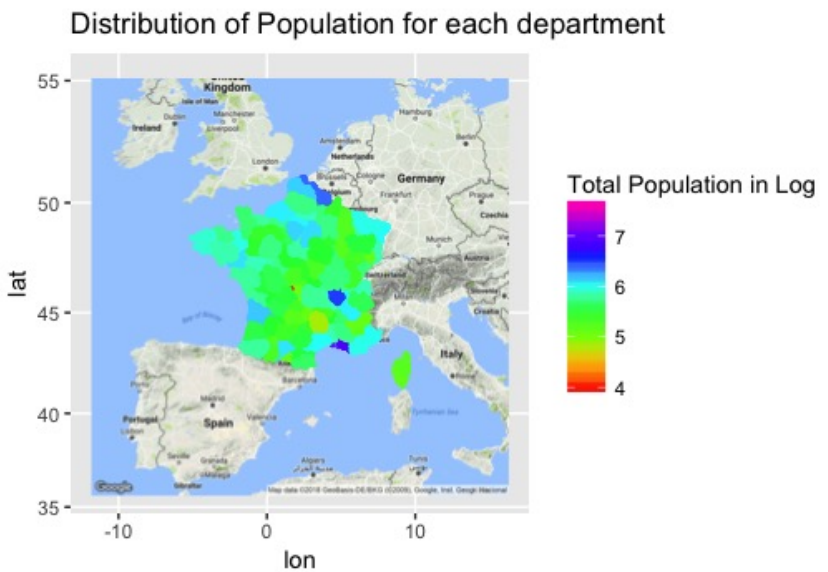
INSEE data

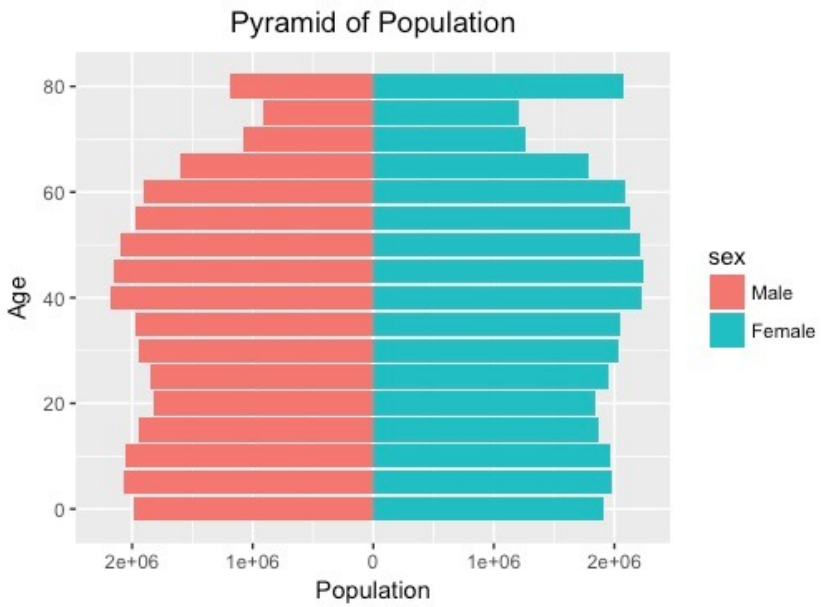
- Population: age, sex and cohabitation mode
- Salary: job categories, age and sex (mean net salary per hour in €)
- Firms: number of firms for each size
- Geography: GPS location

for different geographical levels (communes, departments, towns) in 2014

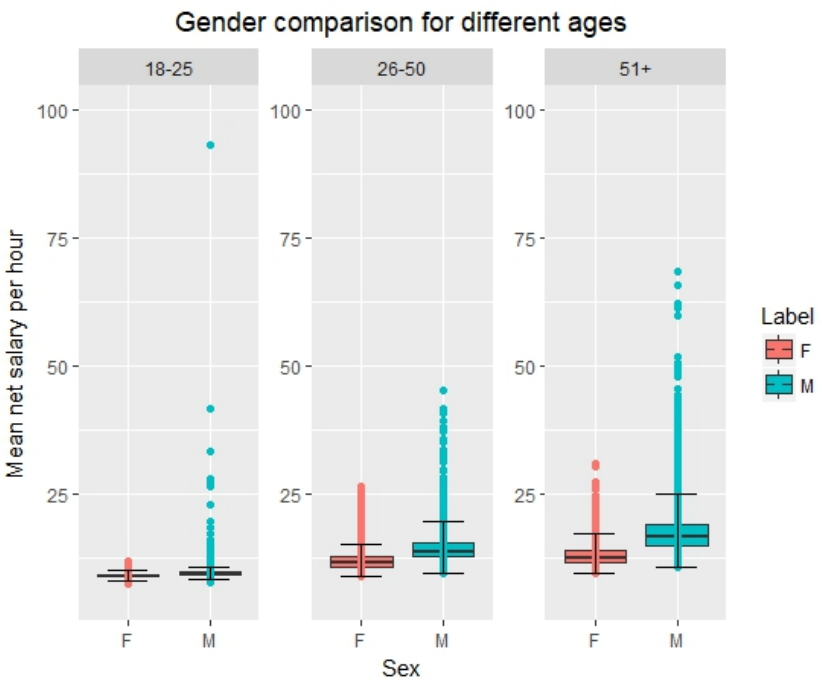
What has been done so far . . .

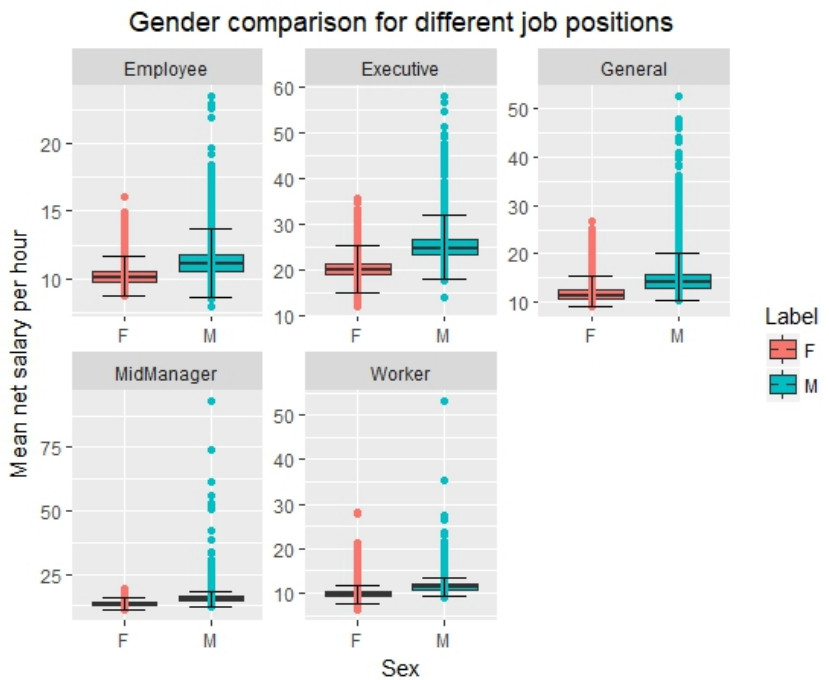
Demographic profiles



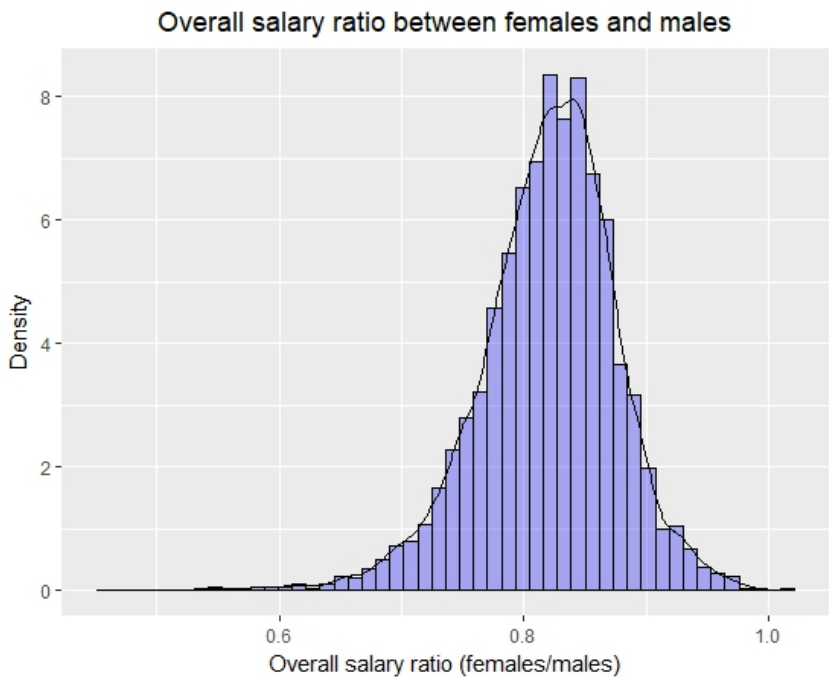


Inequality of salary



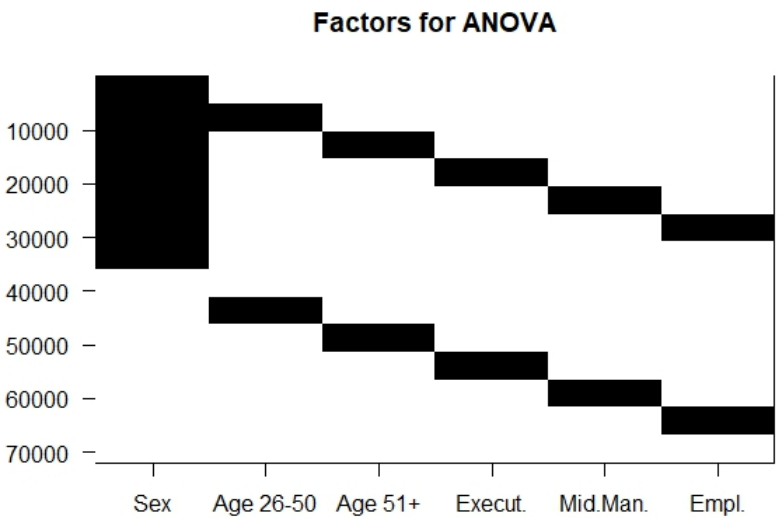


Inequality of salary



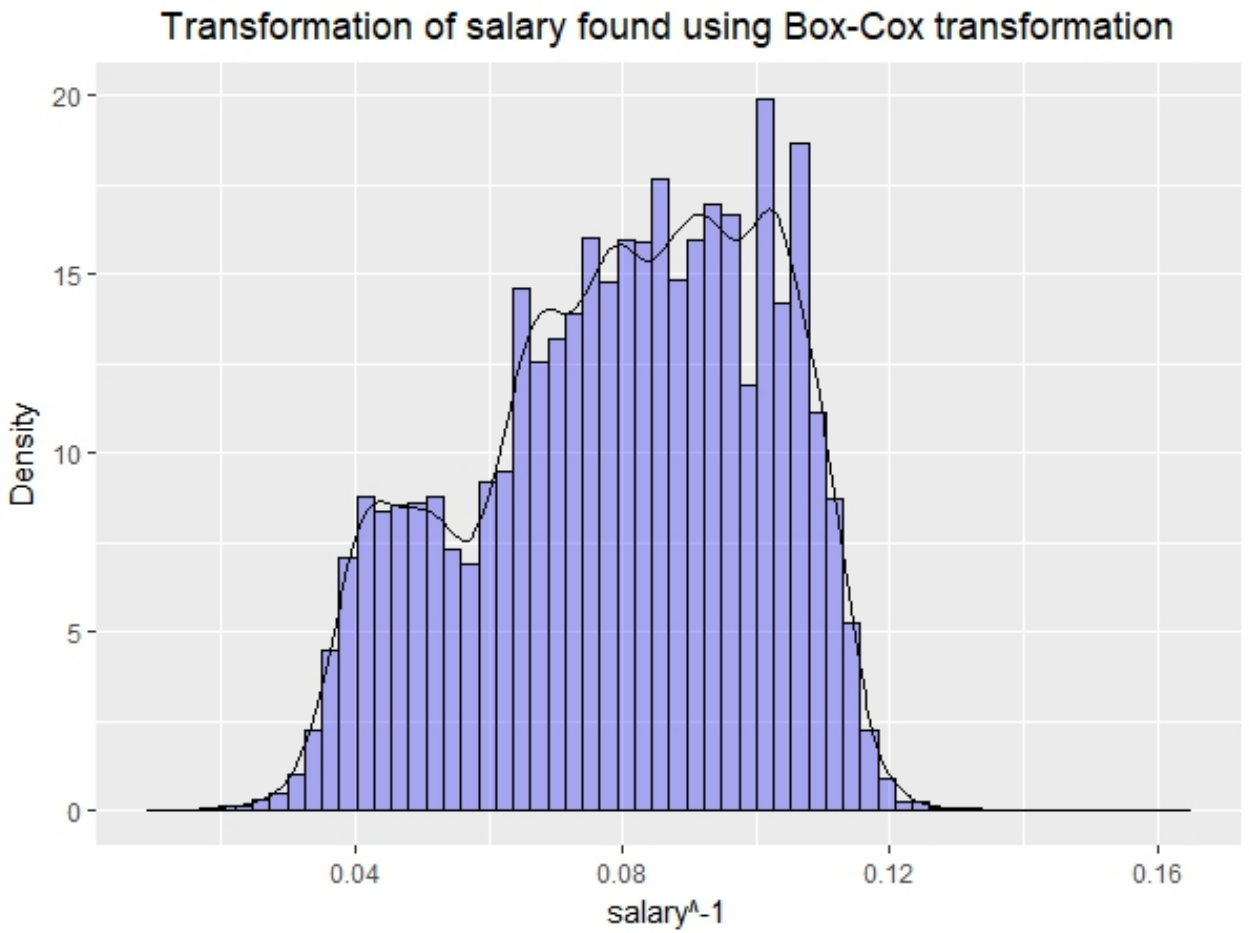
Inequality of salary

ANOVA using sex, job, age and interaction effects

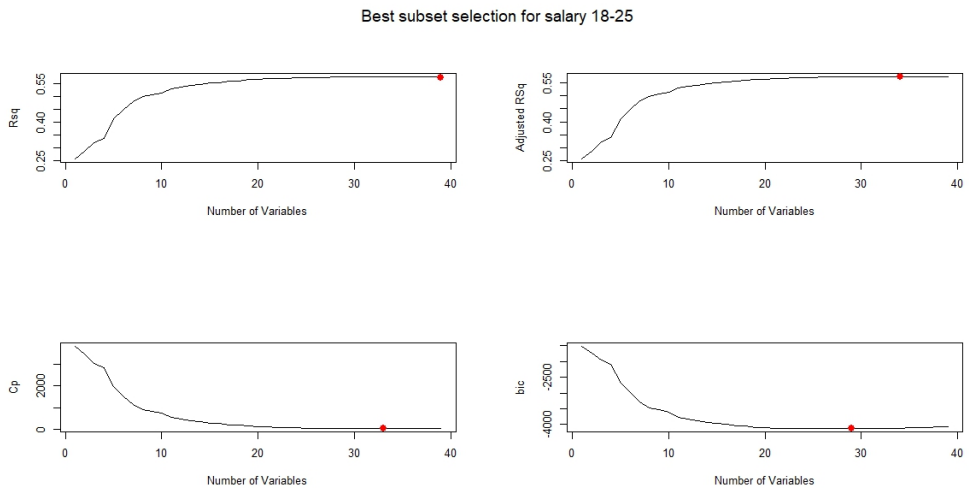


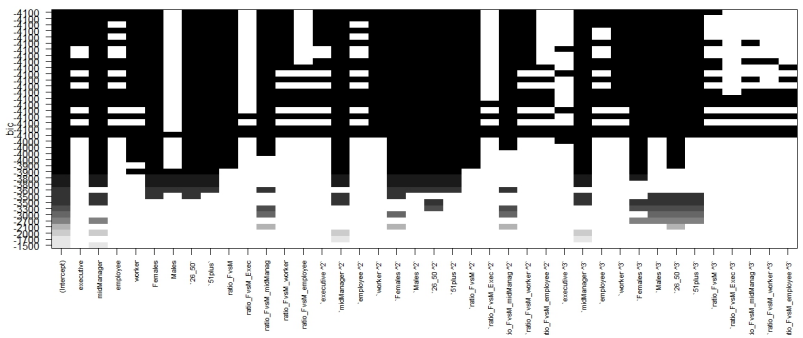

```
Call:
lm(formula = sal_y ~ sal_sex + sal_age + sal_job + sal_sex:sal_age +
    sal_sex:sal_job)

Residuals:
      Min        1Q    Median        3Q       Max
-0.084405 -0.004353  0.000683  0.005477  0.057842

Coefficients:
                  Estimate Std. Error  t value Pr(>|t|)
(Intercept)       1.061e-01  8.471e-05 1252.443  < 2e-16 ***
sal_sex          -1.097e-02  1.198e-04  -91.569  < 2e-16 ***
sal_age1         -2.160e-02  1.467e-04 -147.227  < 2e-16 ***
sal_age2         -2.838e-02  1.467e-04 -193.440  < 2e-16 ***
sal_job1         -5.601e-02  1.467e-04 -381.776  < 2e-16 ***
sal_job2         -3.036e-02  1.467e-04 -206.917  < 2e-16 ***
sal_job3         -8.621e-03  1.467e-04  -58.758  < 2e-16 ***
sal_sex:sal_age1 -2.502e-03  2.075e-04  -12.057  < 2e-16 ***
sal_sex:sal_age2 -7.572e-03  2.075e-04  -36.491  < 2e-16 ***
sal_sex:sal_job1  1.197e-03  2.075e-04    5.770 7.94e-09 ***
sal_sex:sal_job2  4.873e-04  2.075e-04    2.349   0.0188 *
sal_sex:sal_job3  3.059e-03  2.075e-04   14.742  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.008585 on 71892 degrees of freedom
Multiple R-squared:  0.841,     Adjusted R-squared:  0.841
F-statistic: 3.458e+04 on 11 and 71892 DF,  p-value: < 2.2e-16
```

ANOVA using sex, job, age and interaction effects

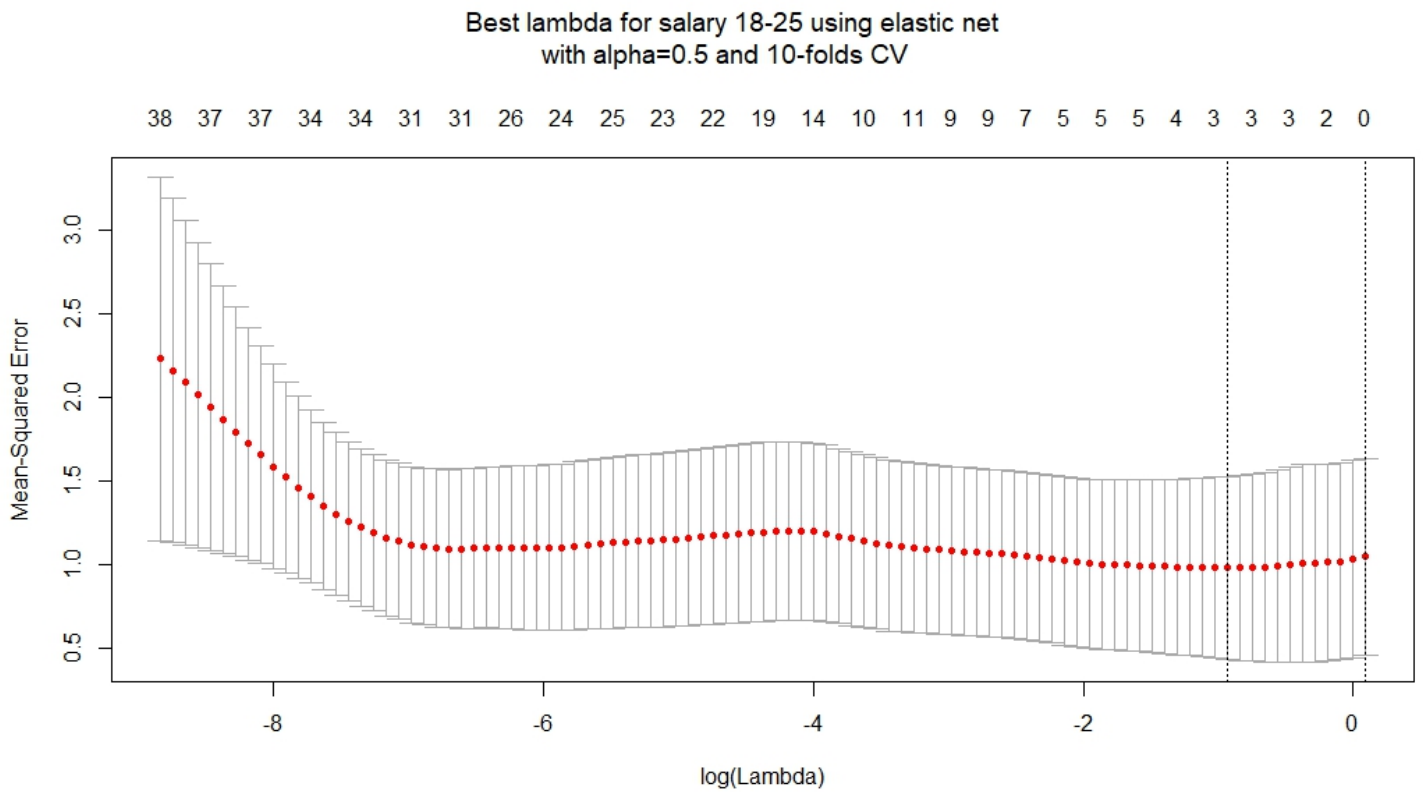


Prediction for young people using BSS

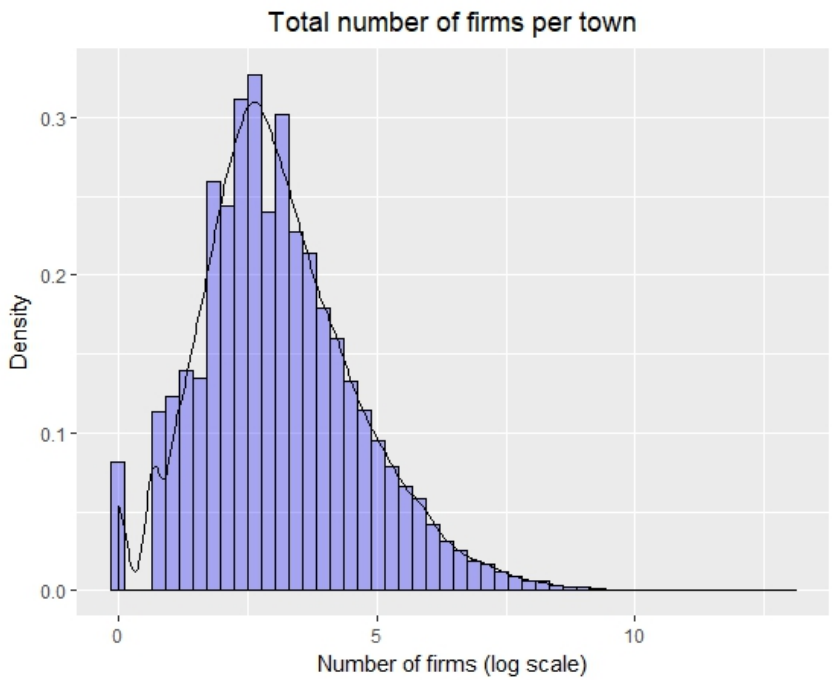


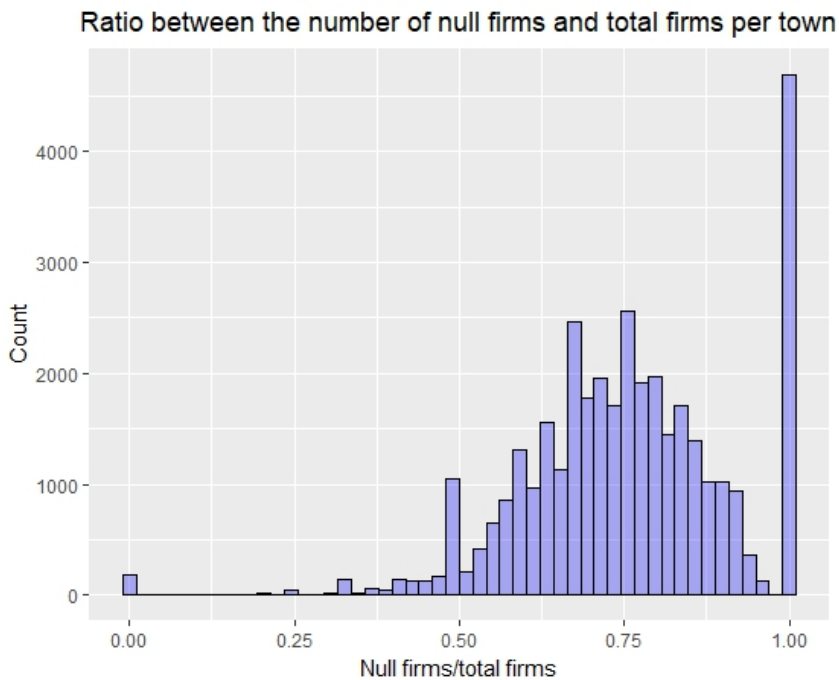


Prediction for young people using elastic net and and 10-folds CV

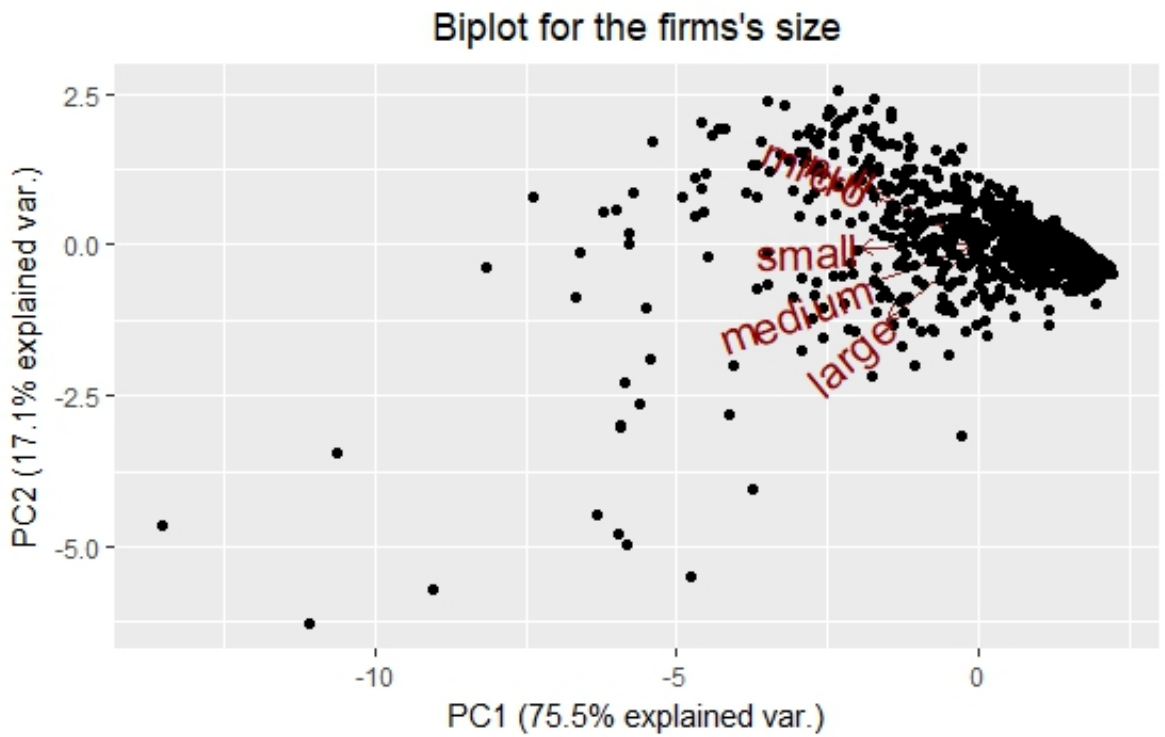


Distribution of firms per town





PCA


Biplot for the firms's size
micro
small
medium
large
2.5
0.0
-2.5
-5.0
-10
-5
0
PC2 (17.1% explained var.)
PC1 (75.5% explained var.)

Issues

- A lot of NA in geo locations (retrieved from Google API)
- Unique code for salary data 1/7 of the total
- Missing additional information
- French DOM-TOM regions
- Outliers and spatial correlation

Future works

- Combine the separated datasets
- Create meaningful indicators
- Take correlation into account (especially spatial)
- Perform clustering techniques to identify geographical clusters
- Perform groupwise lasso to predict salary data
- Verification/improvement of the obtained results
- Compare the methodologies used with robust ones
- Find complementary datasets

– Thank you –