Study of French labour market and inequalities

L. Insolia, J. Kim and G. Yeghikyan

SNS

— Midterm results —

March 14, 2018

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

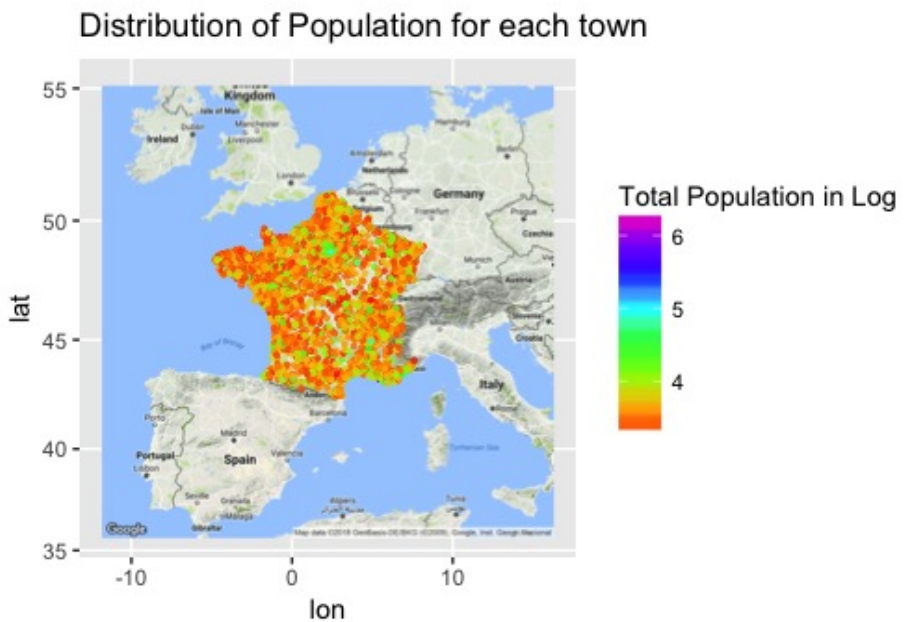
GitHub repo: https://github.com/LucaIns/TSL

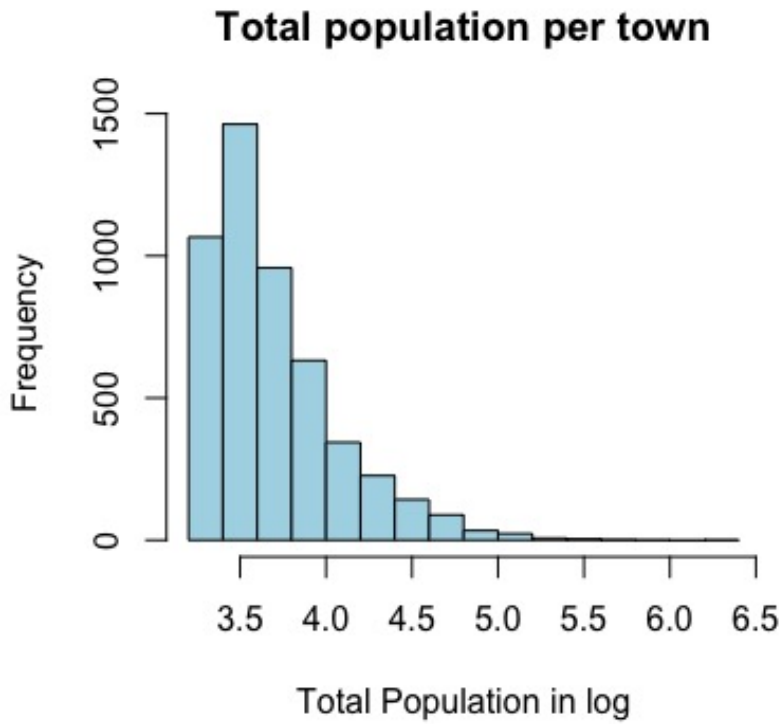
What has been done so far . . .

Pre-processing phase

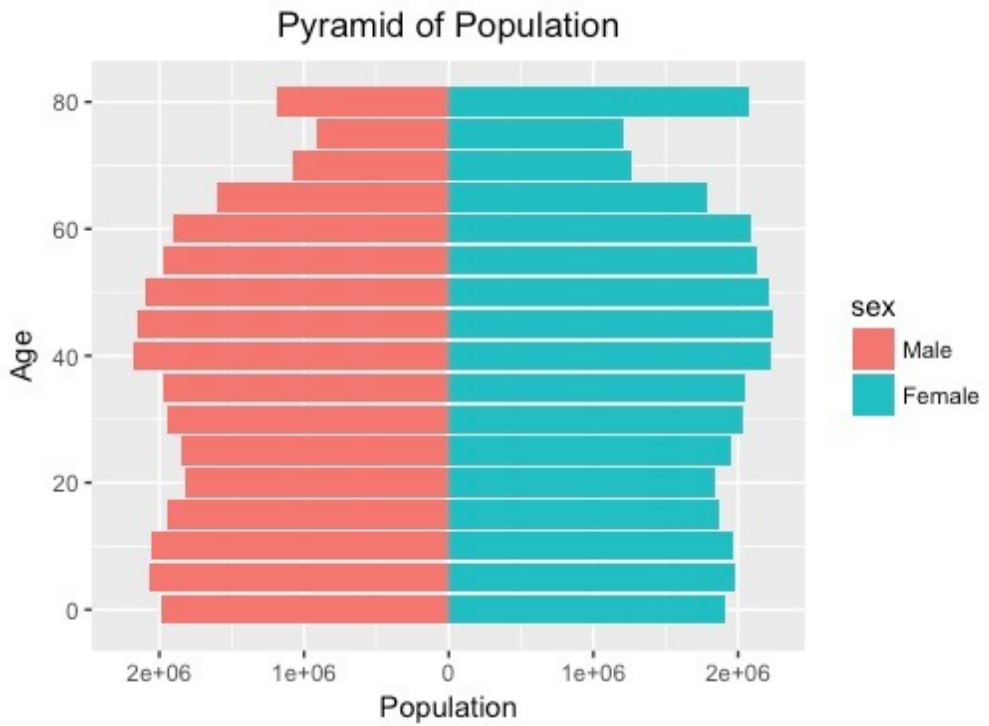
- Population: restructured the dataset and created new features
- Firms: categorized firms' sizes into 4 categories
- Geography: retrieved the missing data using Google API

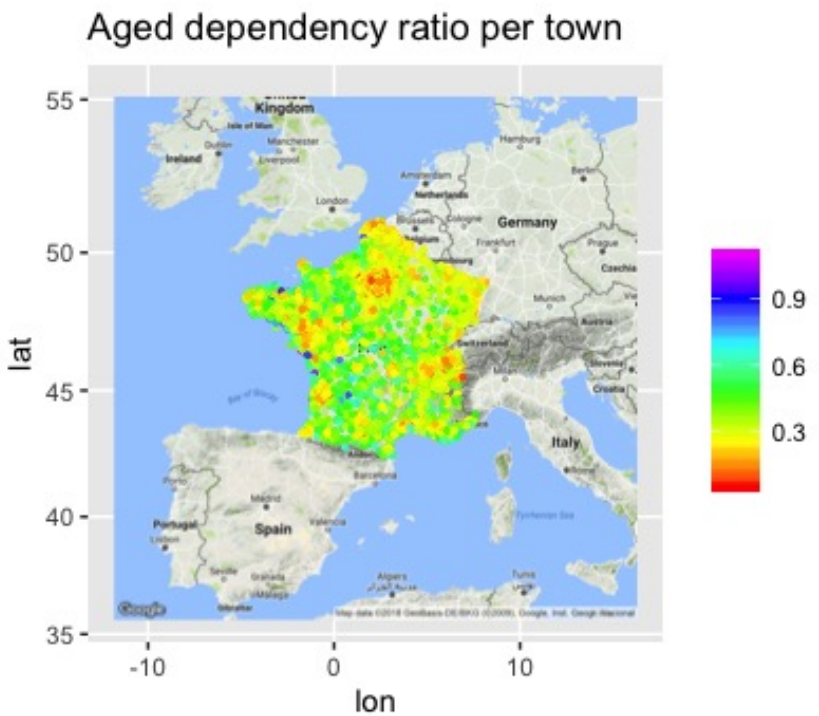
Distribution of population per town



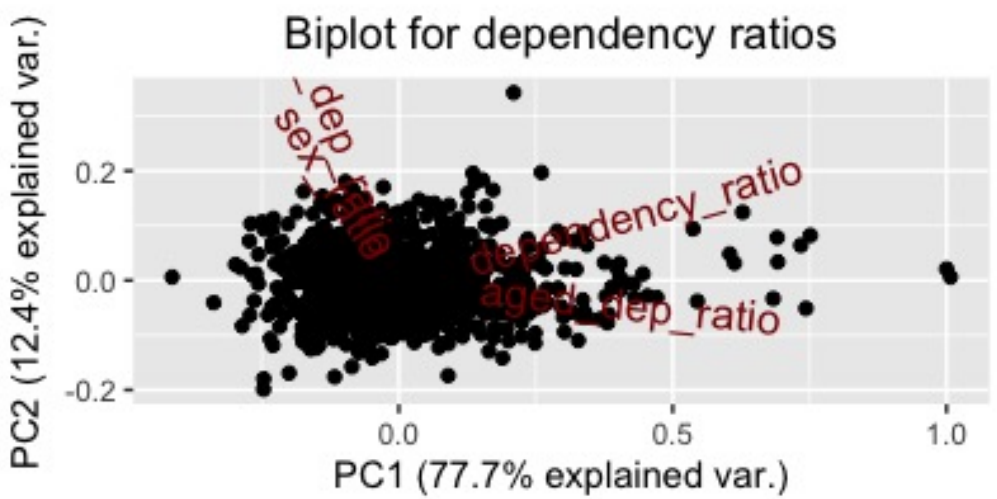


Population demographics

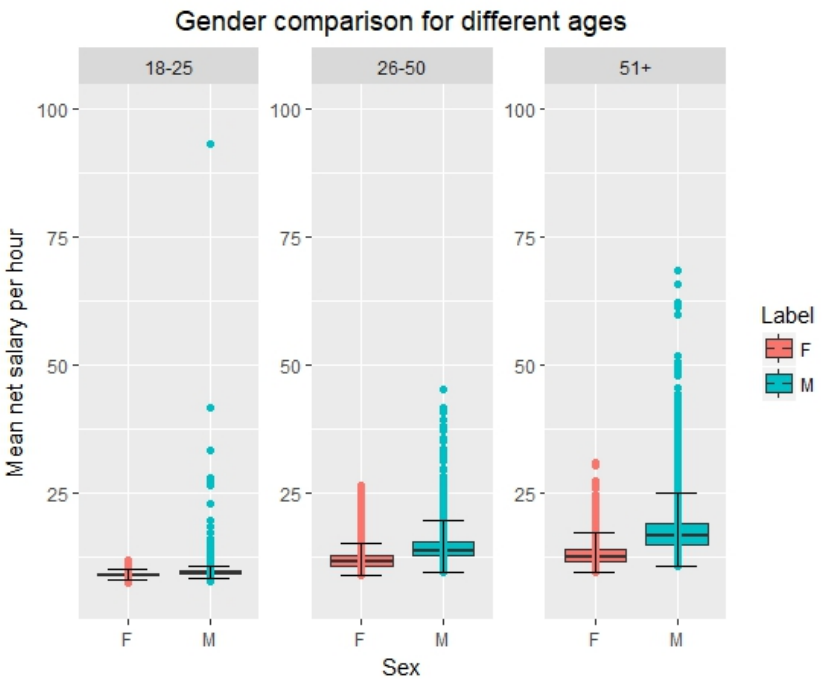


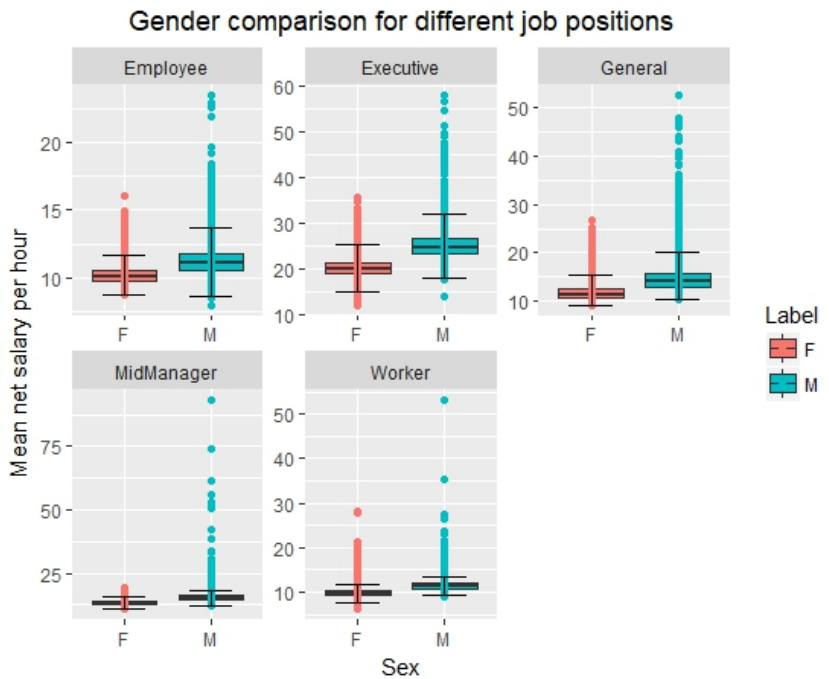


PCA

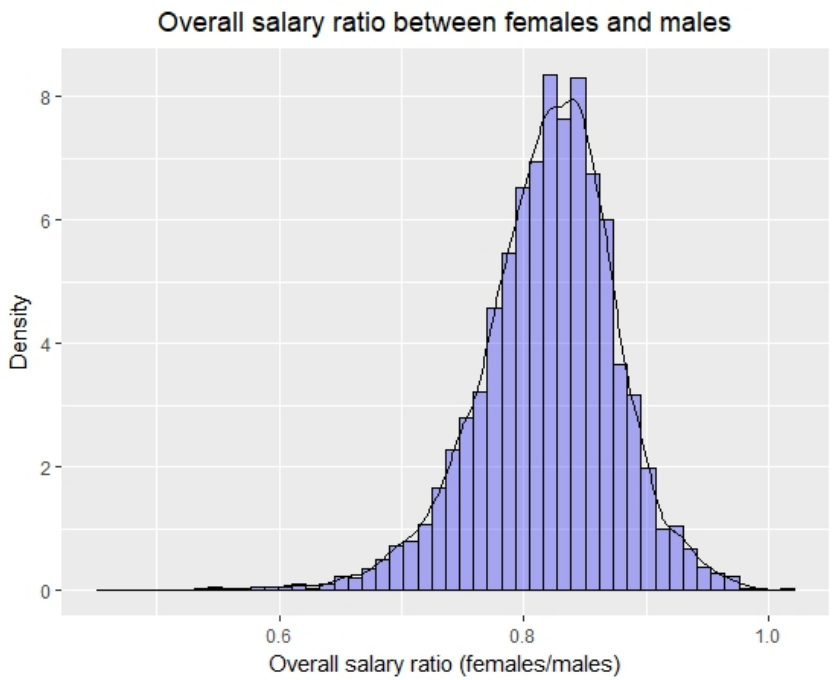

Biplot for dependency ratios
PC2 (12.4% explained var.)
PC1 (77.7% explained var.)
0.2
0.0
-0.2
0.0
0.5
1.0
dependency_ratio
aged_dep_ratio

Inequality of salary





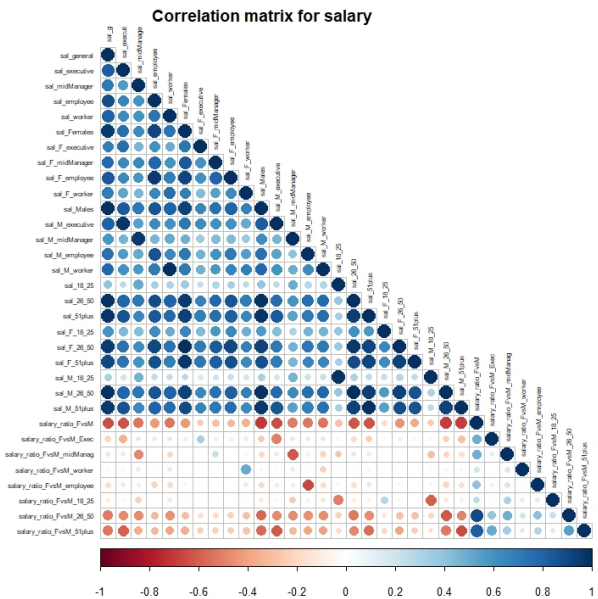
Inequality of salary

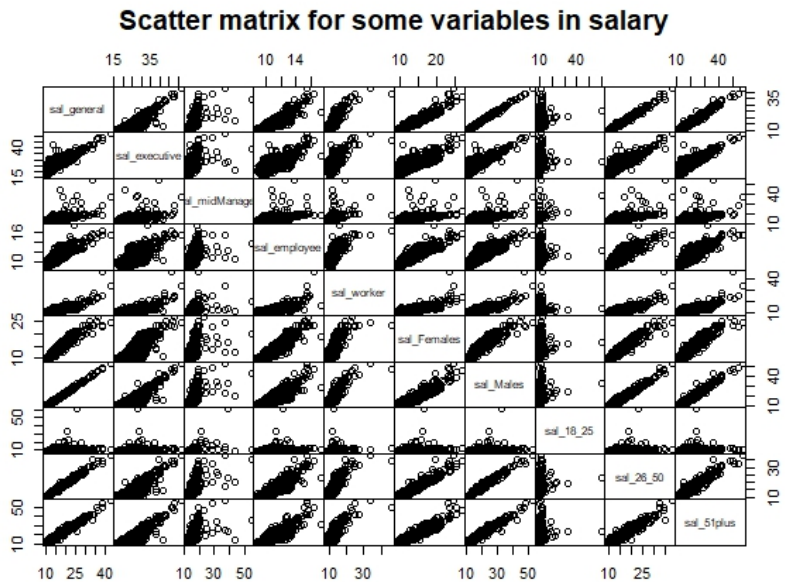


Inequality of salary

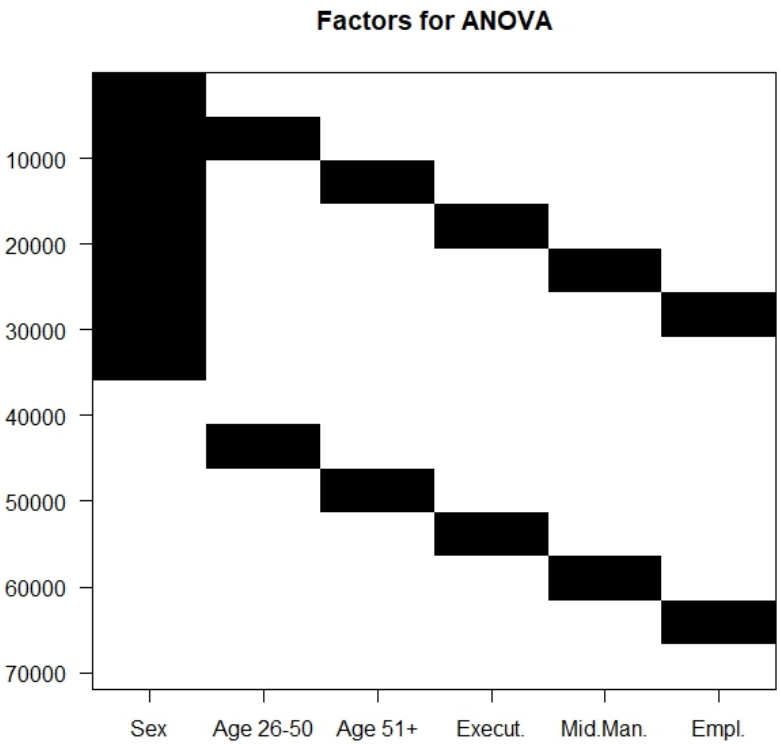
Inequality of salary

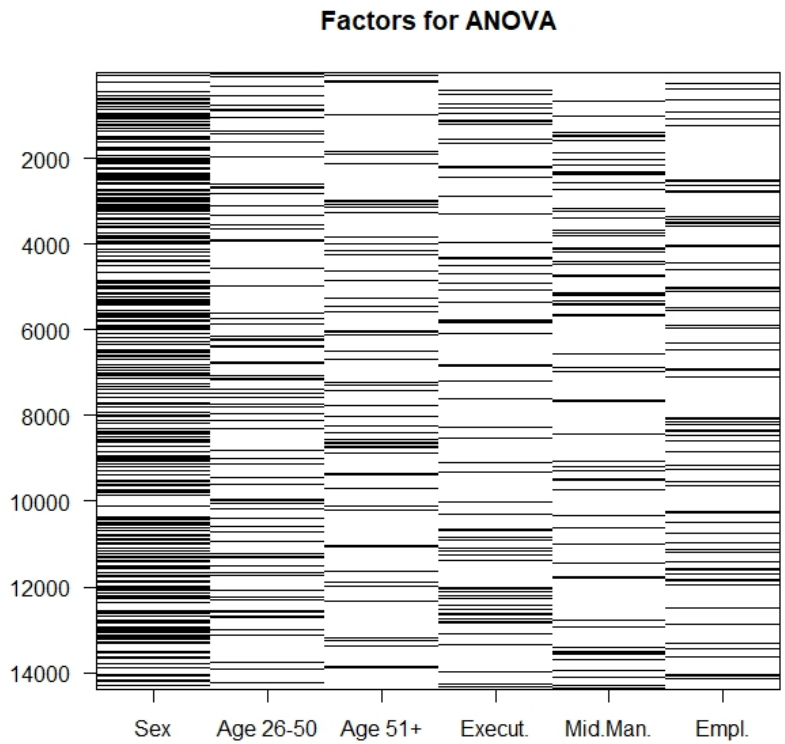
Bivariate relations



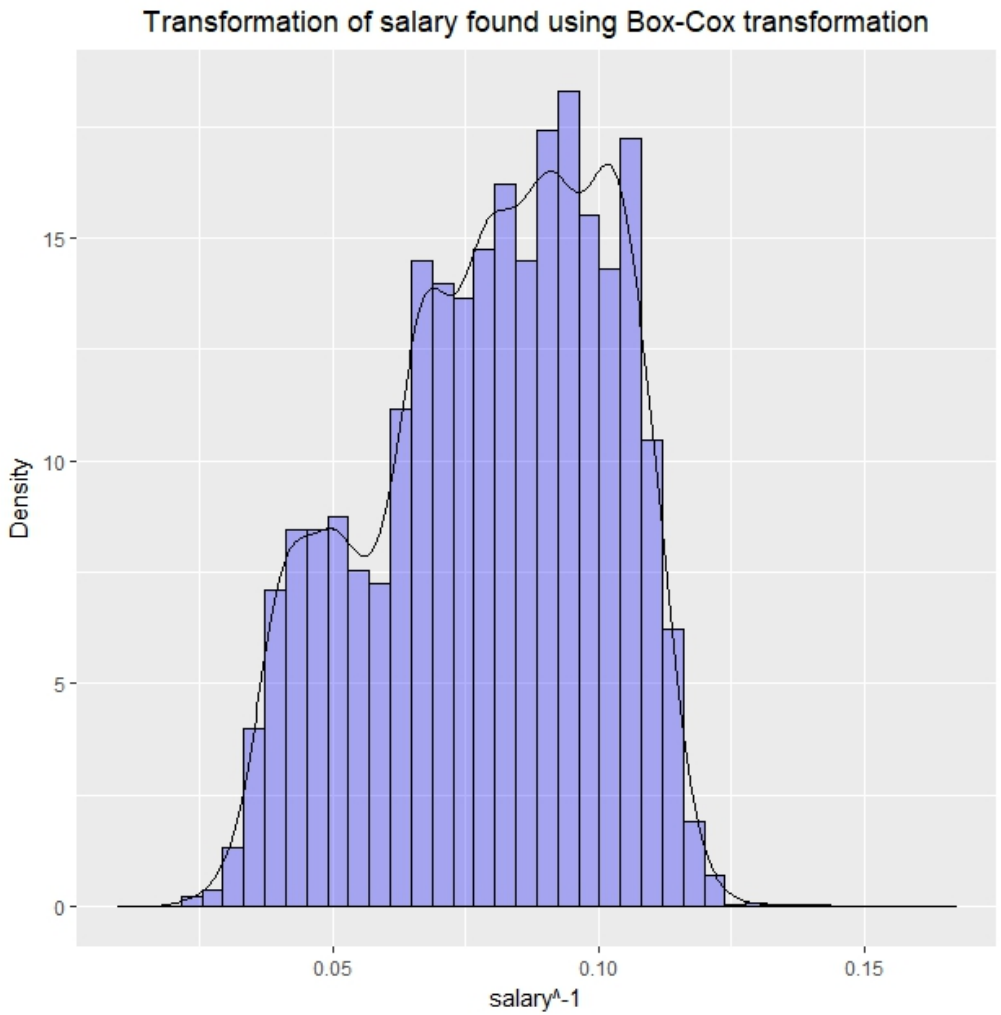


ANOVA using sex, job, age and interaction effects

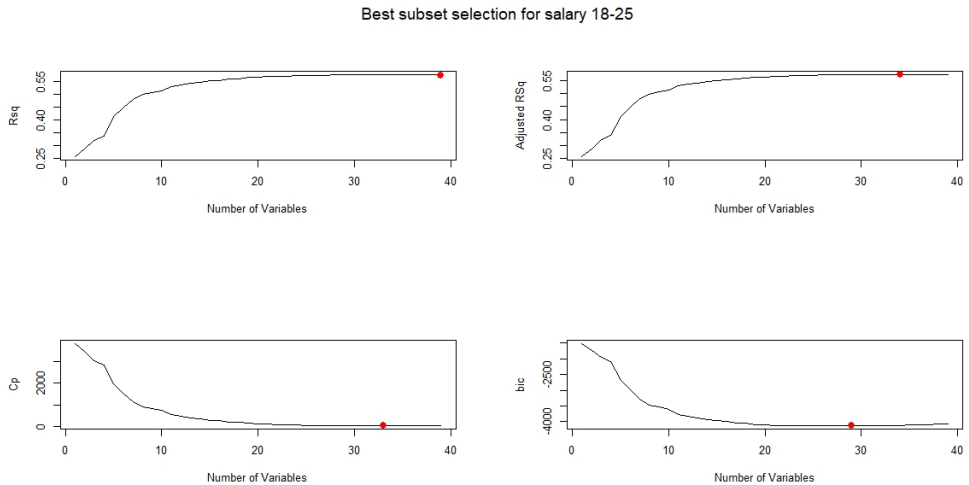




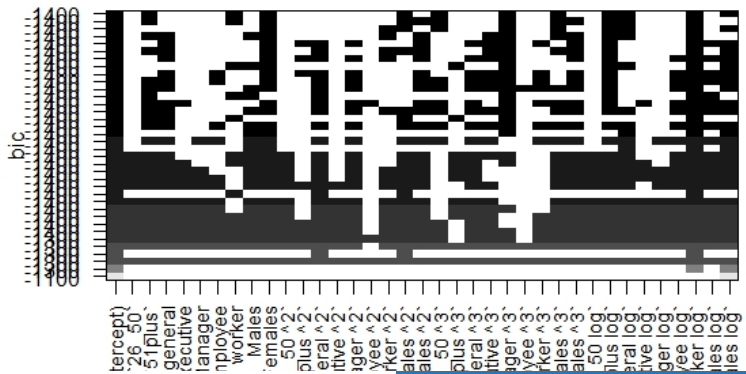
ANOVA using sex, job, age and interaction effects



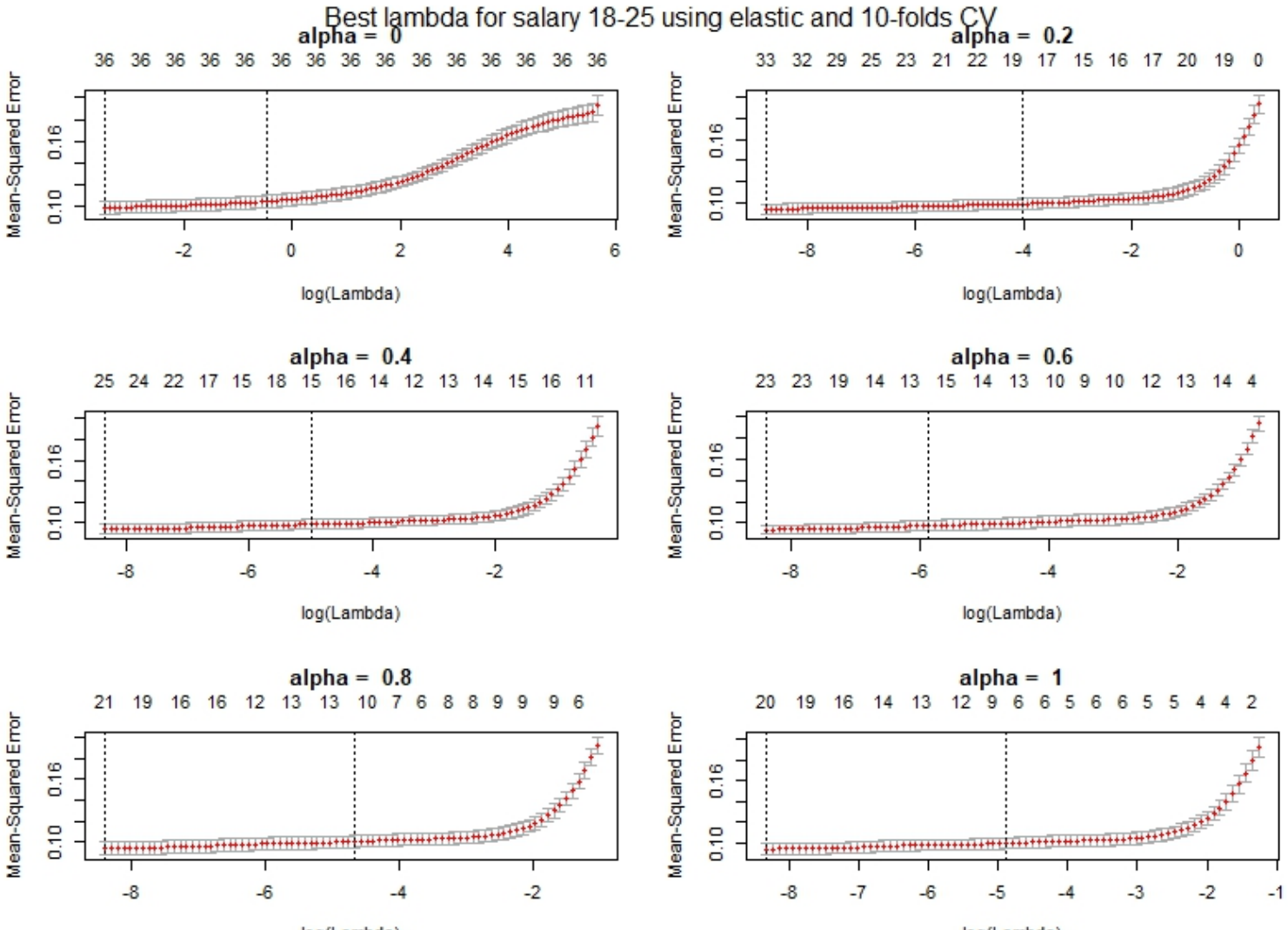
Prediction for young people using BSS



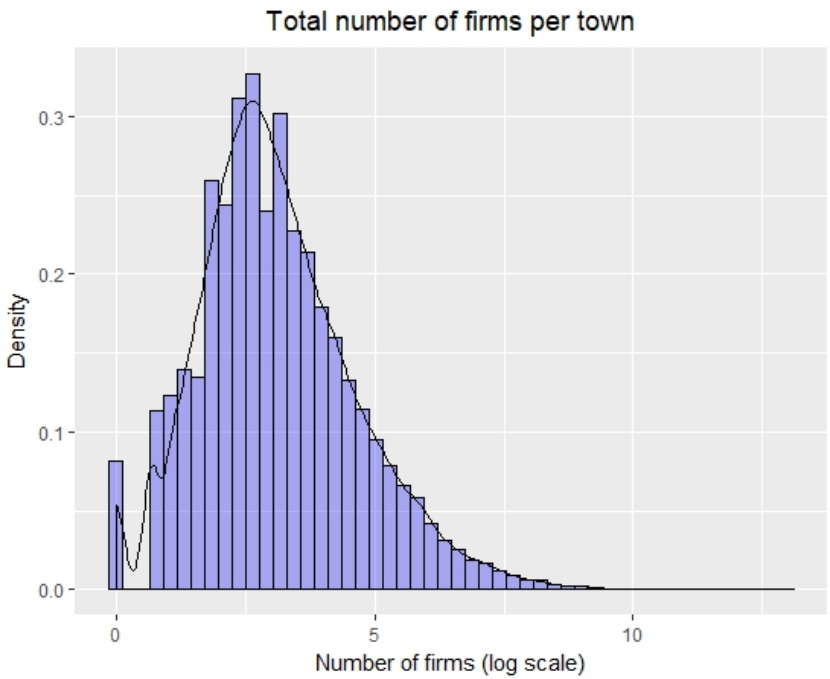
Best subset selection for salary 18-25 using BIC

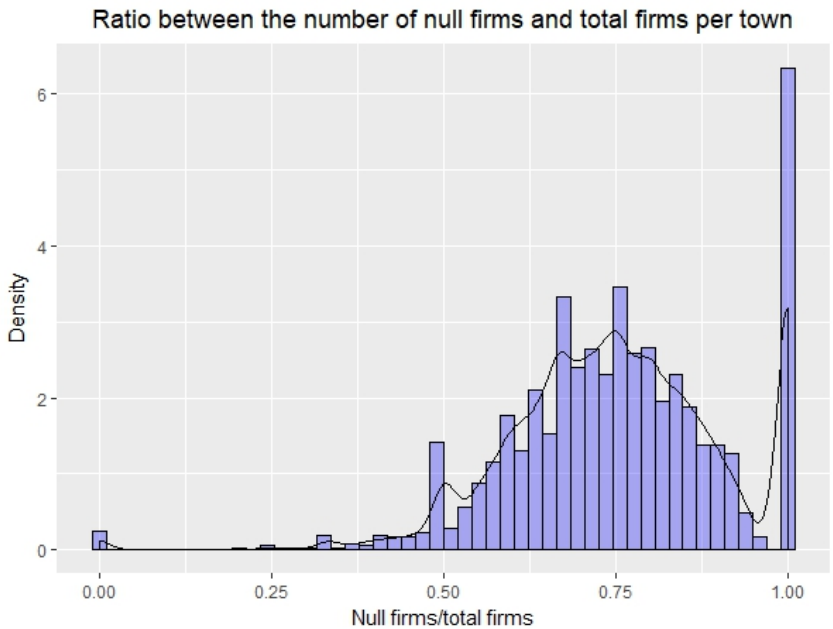


Elastic net and and 10-folds CV

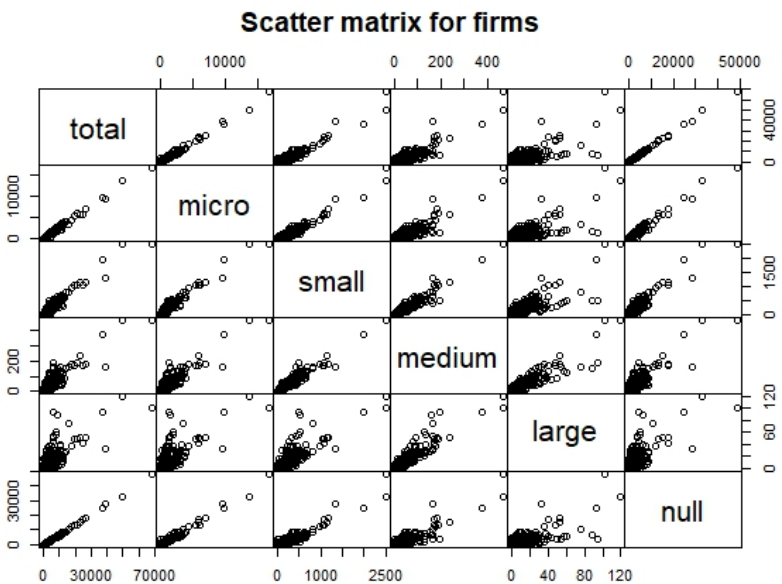


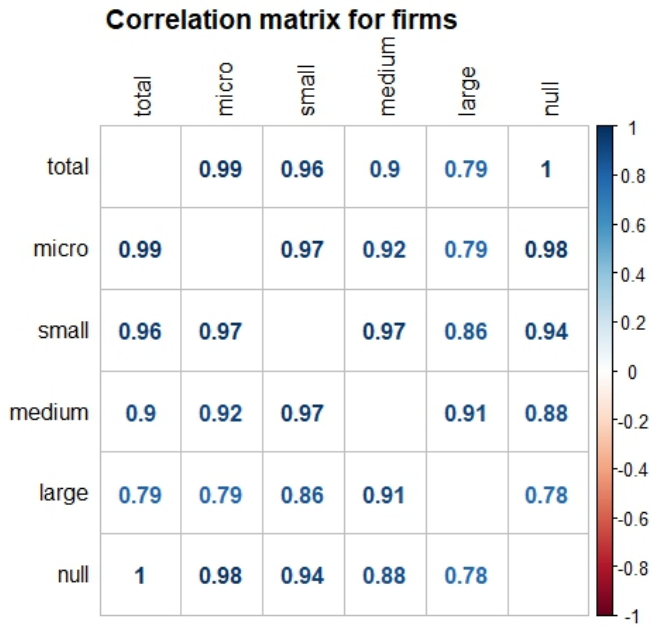
Distribution of firms per town





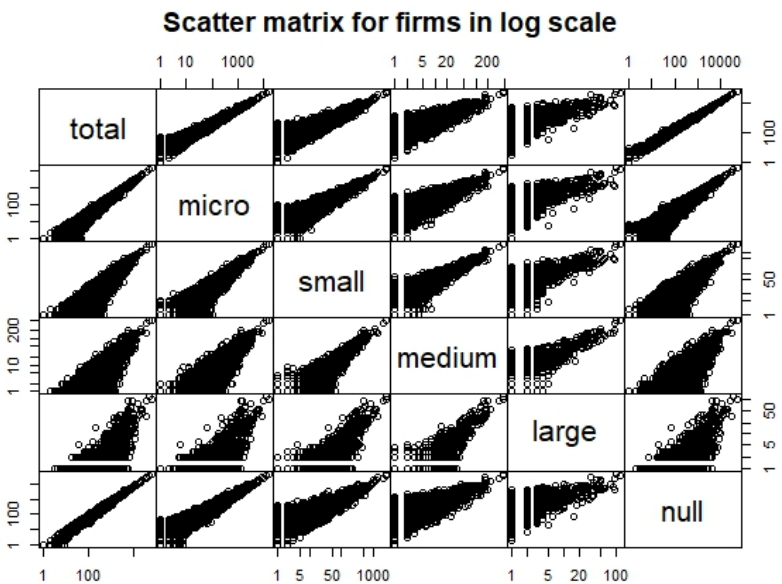
Bivariate relations

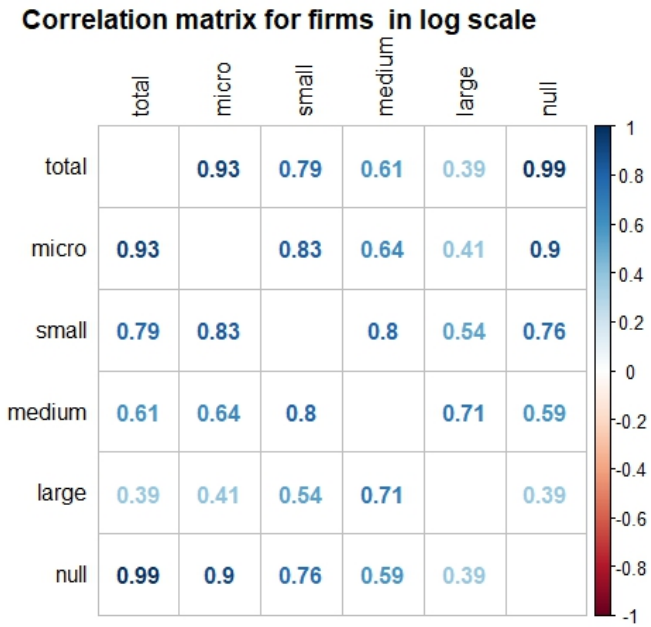




Excluding Paris

Bivariate relations

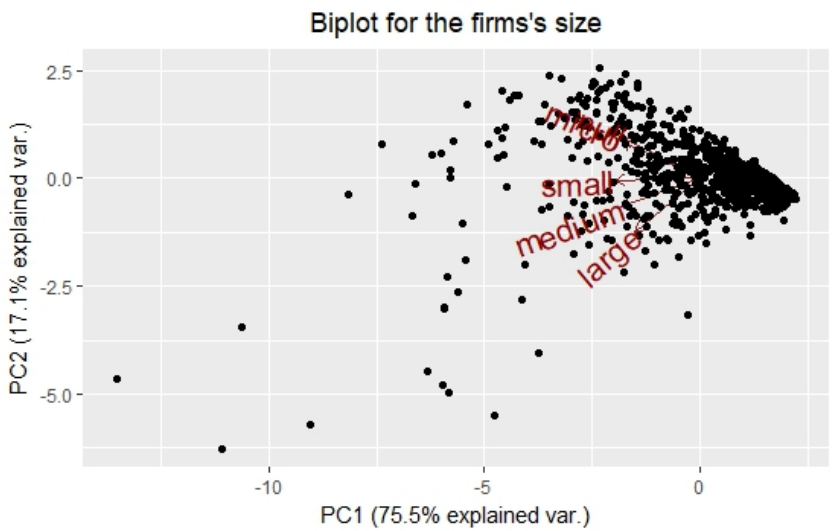


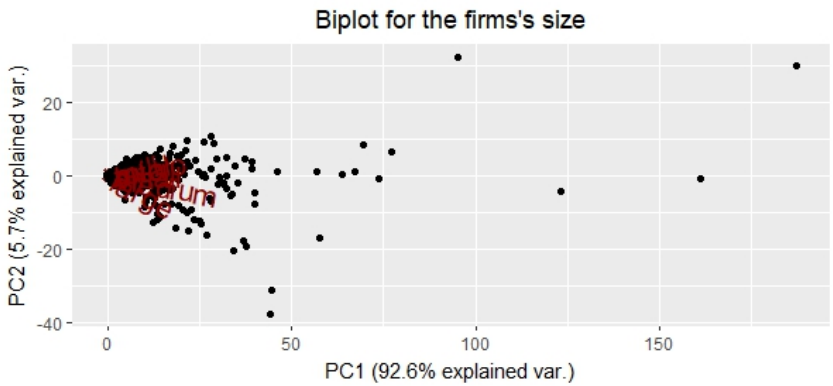


Including Paris

PCA

Using original data scaled (not logs)
Most typical vs. Excluding just Paris





Issues

- Unique code for salary data 1/7 of the total
- Loss of information when combining the separated datasets
- Missing additional information
- French DOM-TOM regions
- Outliers and spatial correlation

Future works

- 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 –