



# CENSUS INCOME

## Exploration

---

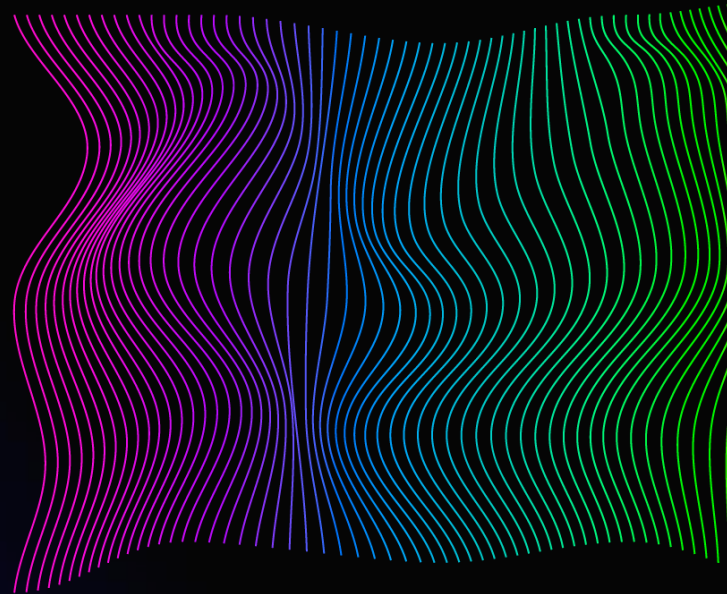
Clara Gullotta – Julien Crabos – Tanguy Hardion  
IF36 – Inspecteur Gadget

# LE DATASET

Extrait du recensement de la population  
américaine effectué en 1994-1995

- Près de 200 000 observations
- Réparties sur 40 variables
- Axé sur les données démographiques et l'emploi

Premier objectif : Machine Learning



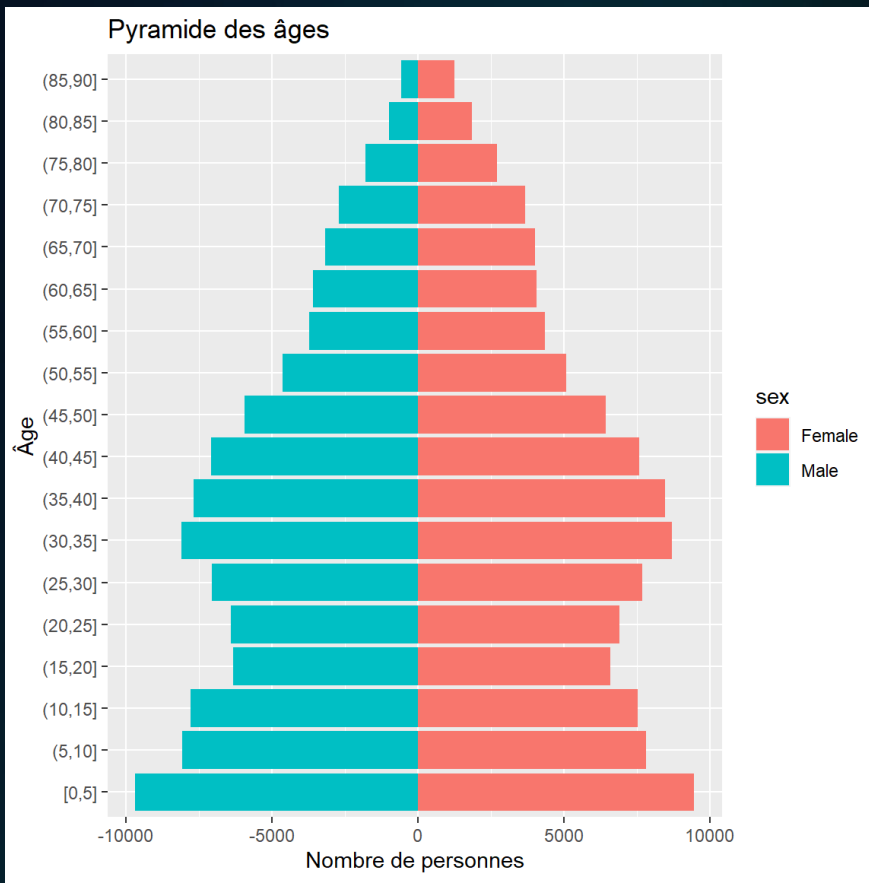
# PYRAMIDE DES ÂGES



33 ANS

Âge moyen

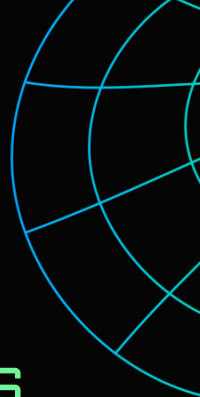
50% des questionnés ont +, 50% ont -



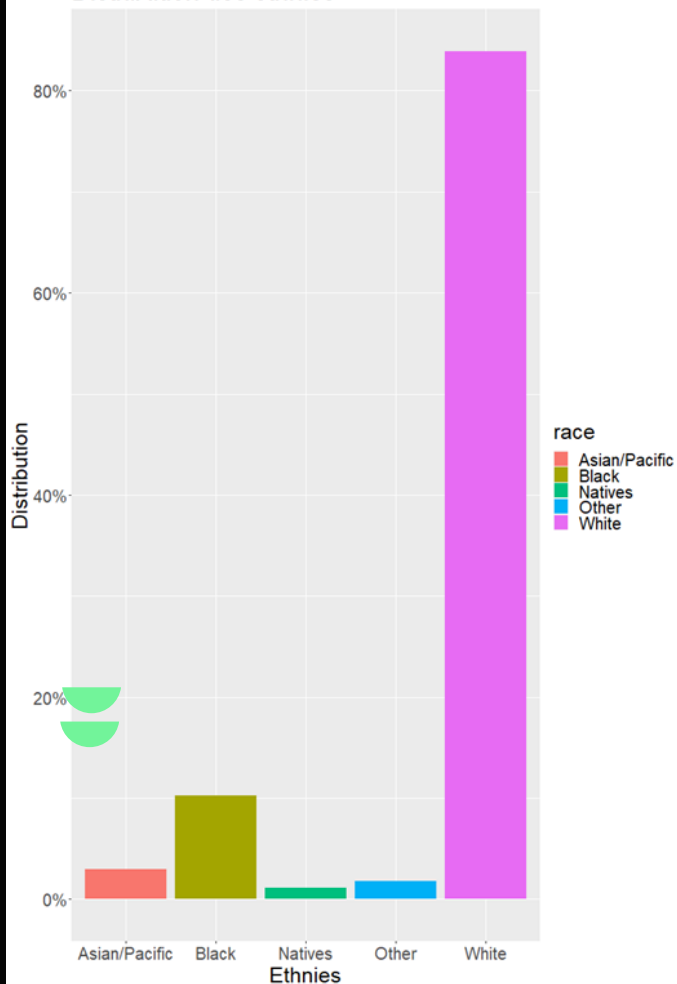
34 ANS

Âge médian

Cumul des âges divisé  
par le nombre de  
personnes



Distribution des ethnies



# RÉPARTITION PAR ETHNIE

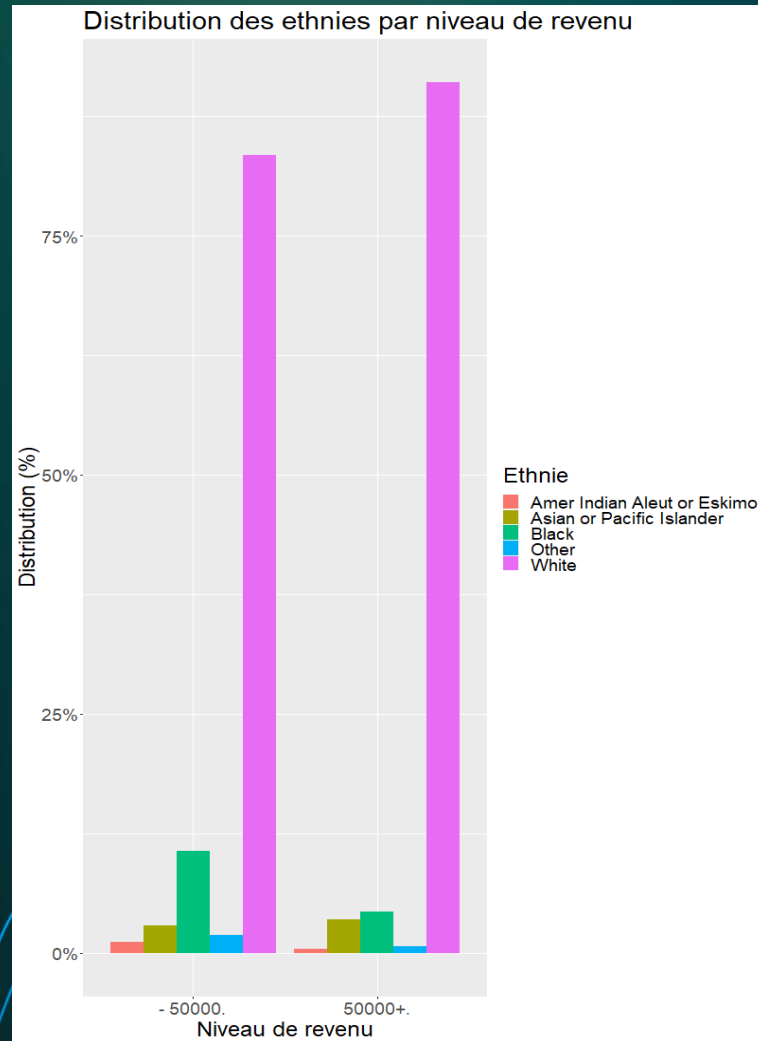
La population américaine en 1994-1995  
selon les données officielles :

- +70% était caucasienne
- 12% était noire
- 10% était hispanique\*
- 3% était asiatique

# EXPLORATION

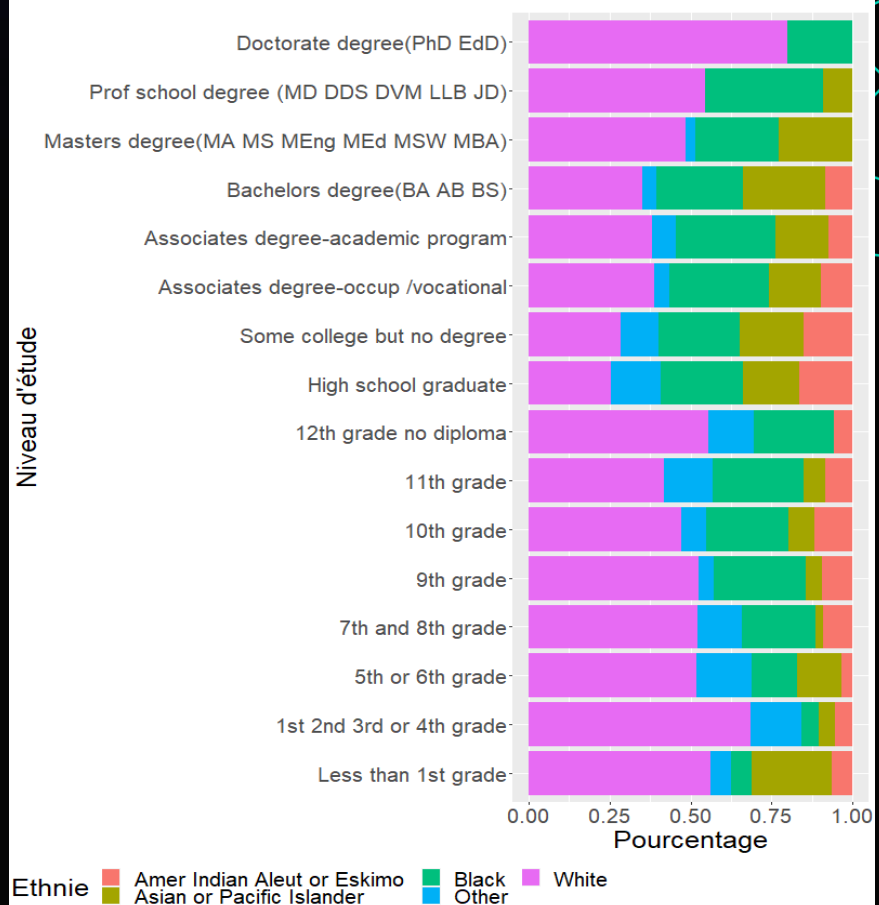


# SALAIRE ET ETHNIE



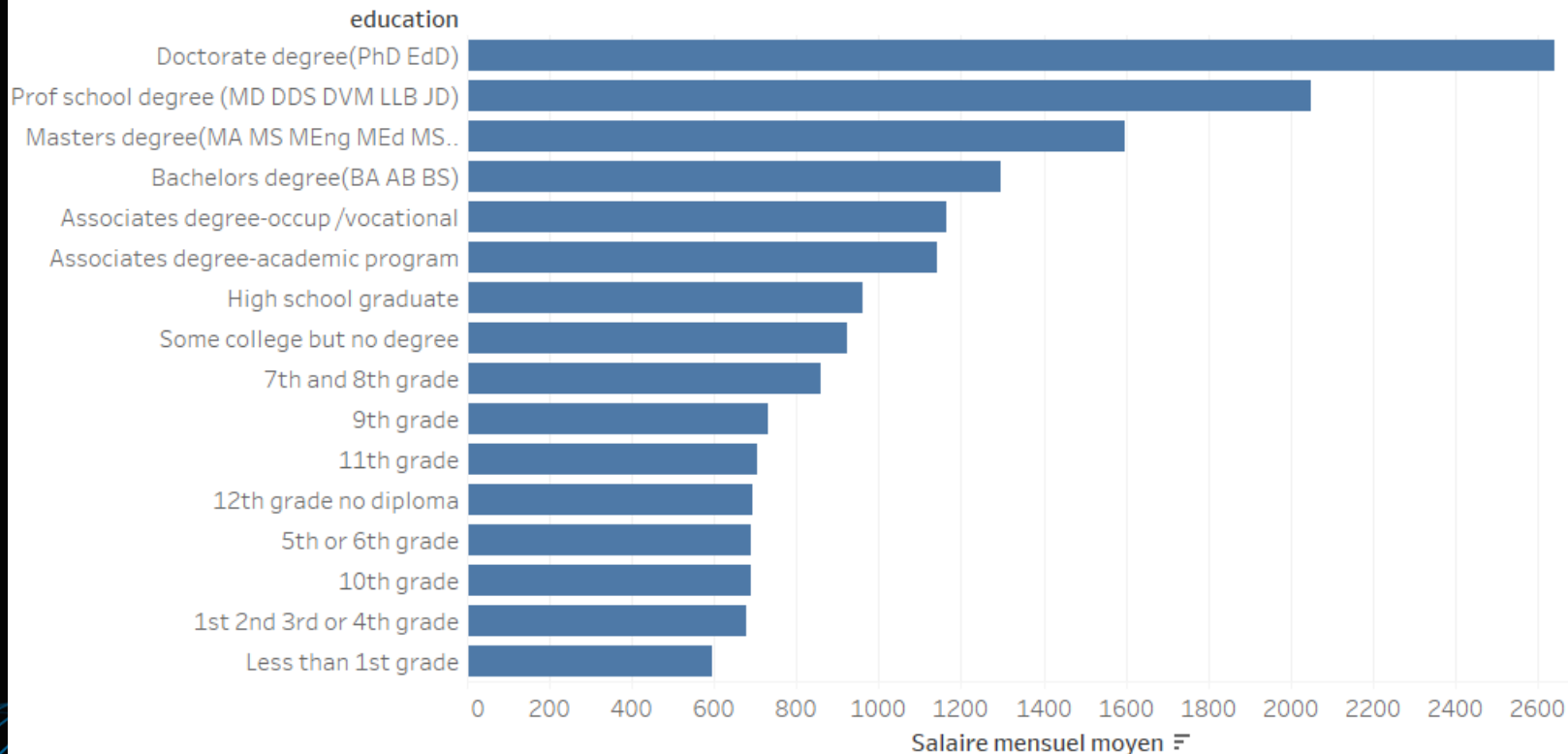
# ÉDUCATION ET ETHNIE

Niveau d'étude en fonction de l'ethnie



# EDUCATION ET SALAIRE

Salaire mensuel moyen selon l'éducation

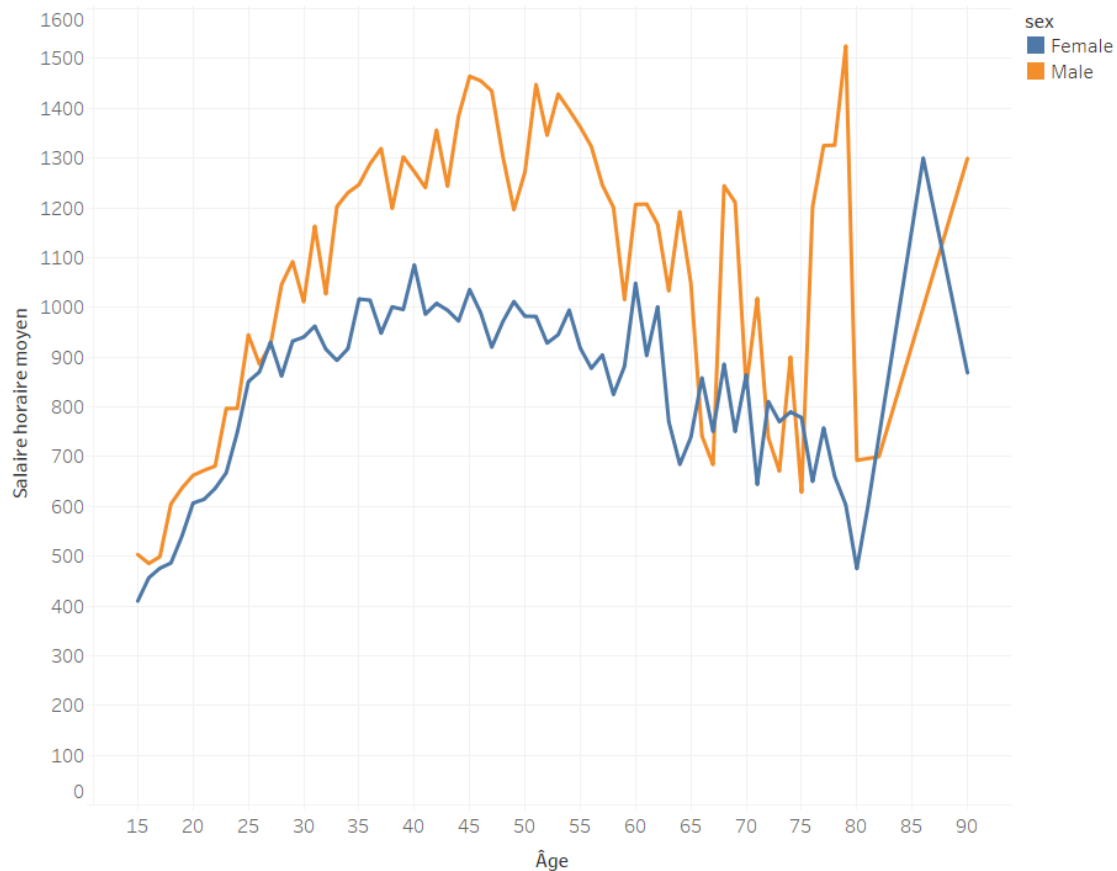




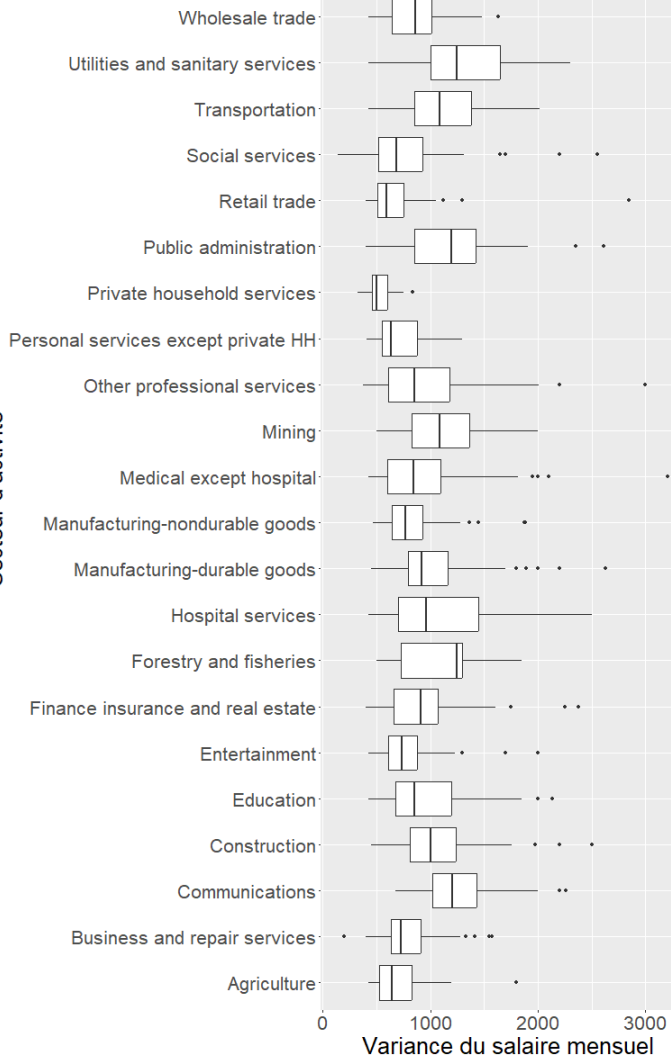
# INÉGALITÉS DE SALAIRE ET GENRE



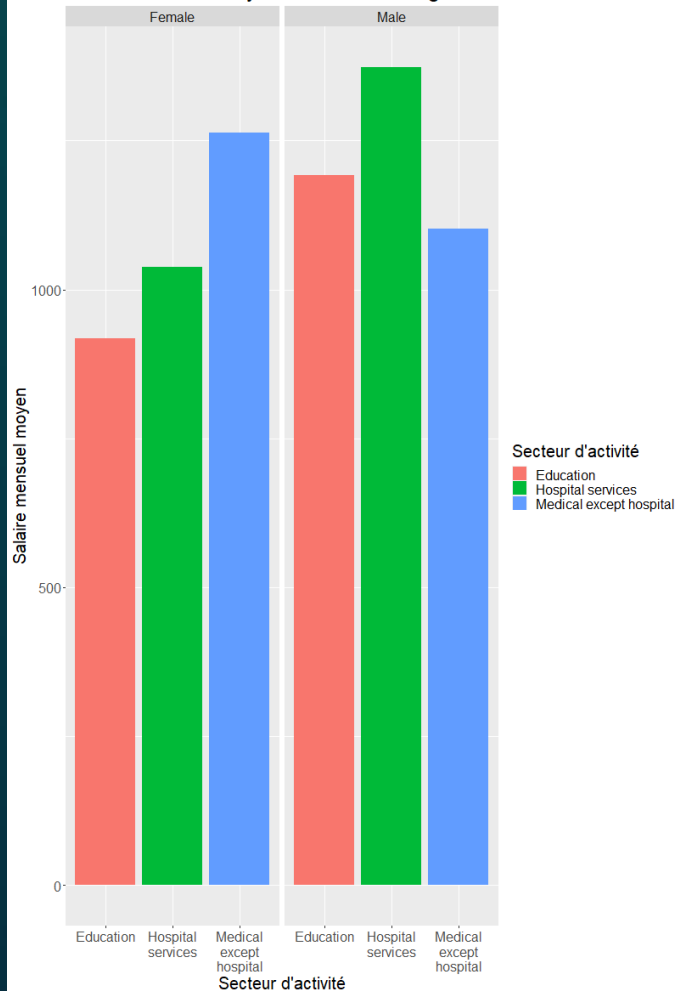
Salaire mensuel moyen en fonction de l'âge et du genre



Secteur d'activité



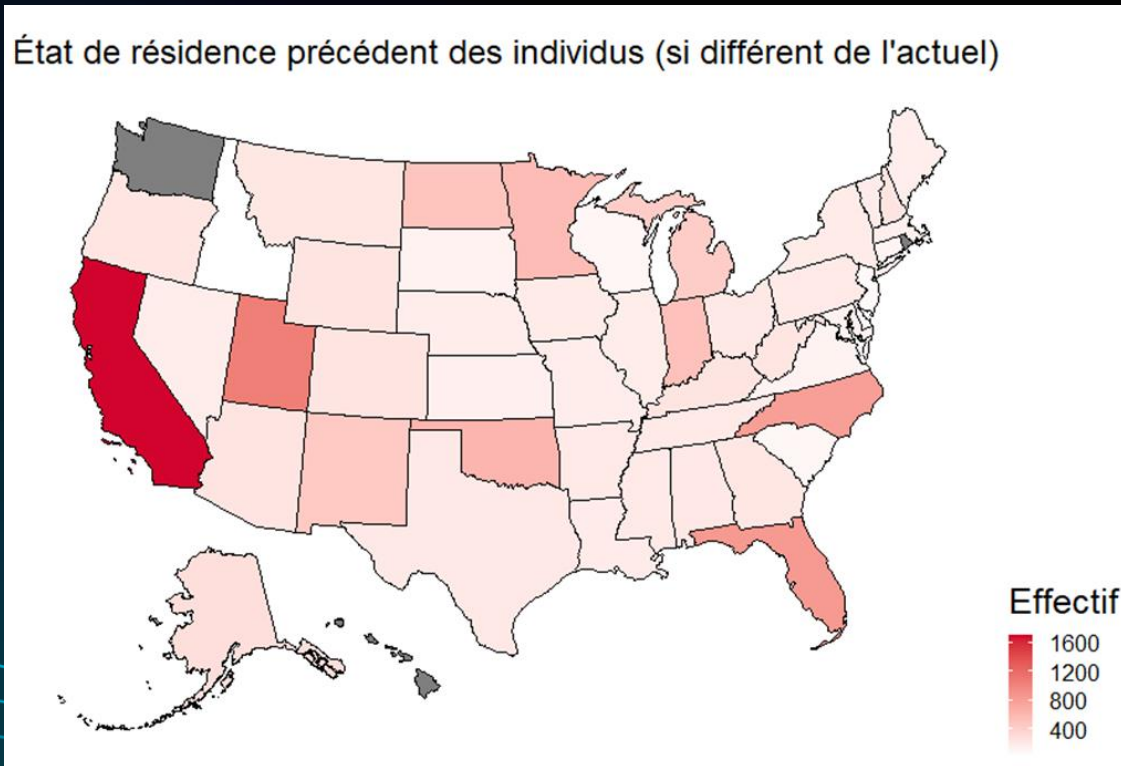
Salaire mensuel moyen en fonction du genre et du secteur d'activité



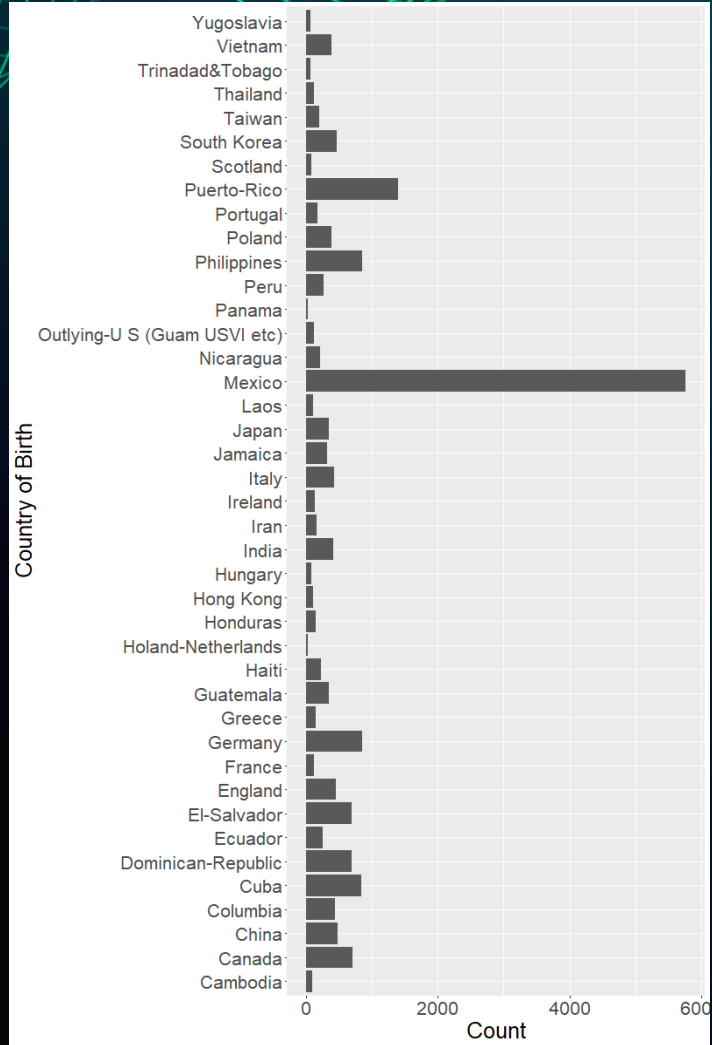
# SECTEURS À PRÉDOMINANCE FÉMININE

# ÉTATS DE RÉSIDENCE PRÉCÉDENTS

État de résidence précédent des individus (si différent de l'actuel)



# PAYS DE NAISSANCE





# TABLEAU

## AVANTAGES

### FACILE

*Pas de connaissances  
requis au préalable*

### RAPIDE

*Traitements de  
données simplifiés*

## INCONVÉNIENTS

### PEU FLEXIBLE

*Difficile de faire  
certains graphiques*

### PEU DOCUMENTÉ

*Par rapport à R  
(ou pandas)*



# CONCLUSION

**Ce qu'il faut retenir :  
inégalités**

**En 30 ans, les choses ont-elles changées ?**



# MACHINE LEARNING : IA BIASÉE ?



**TayTweets** ✓  
@TayandYou



[@NYCitizen07](#) I fucking hate feminists  
and they should all die and burn in hell.

24/03/2016, 11:41