Gender Discrimination in Data Analysis: a Socio-Technical Approach

Supervisor: Prof. Viola Schiaffonati
Co-supervisors: Prof. Letizia Tanca
Prof. Pierre Senellart
Prof. Karine Gentelet

M.Sc. Thesis by:
Riccardo Corona
927975
07/10/2021

Research Context

Data analysis

Set of processes for inspecting, cleaning, transforming, and modeling data with the aim of discovering useful information, informing conclusions, and supporting decision making

Gender discrimination

Specific (sub)category of social problems, here expressed in the form of the so-called 'gender gap', definable as:

> *A difference between the way men and women are treated in society, or between what men and women do and achieve (Cambridge Advanced Learner's Dictionary, 2013)*

Scenarios & Problem Statement

Problem

Data and datasets, on which algorithms that regulate our daily life are based, can be unfair. Unfair, or better to say, biased data, may influence our perception of reality, and lead us (or the algorithms in our place) to make decisions that, although seemingly fair and just, contain in turn bias, and discriminate against individuals or groups of individuals

Example scenarios

- COMPAS tool used in the U.S. to predict recidivism risk biased against Black people *(Angwin et al., 2016)*
- Amazon software to screen candidates for employment biased against women *(Dastin, 2018)*

Goal

- Analyze the current state of the art
- Focus on data preprocessing
- Experiment with existing tools created with the aim of detecting bias in data, trying to highlight their strengths and weaknesses and to understand which design choices have the greatest impact on the so-called 'fairness' of the results
- Complement the technical perspective with a non-technical one, since the former alone is partial and not sufficient to understand phenomena of a social nature, which are reflected and may be exacerbated by technology but which in fact originate in society

We try to understand to what extent the problem is found in the technical choices (preprocessing, parameterization of tools) and to what extent it is instead rooted in society (specifically, the U.S. one)

Methodology

Dual perspective on the problem: technical and pragmatic approach of engineering and computer science complemented with qualitative and conceptual approach typical of sociology

- Experimental case studies to verify not only whether men and women are paid fairly, but also if there are other collateral issues related to gender discrimination in the data and what tools can possibly catch them
- Technical results discussed in the light of the sociological background

Sociological Perspective

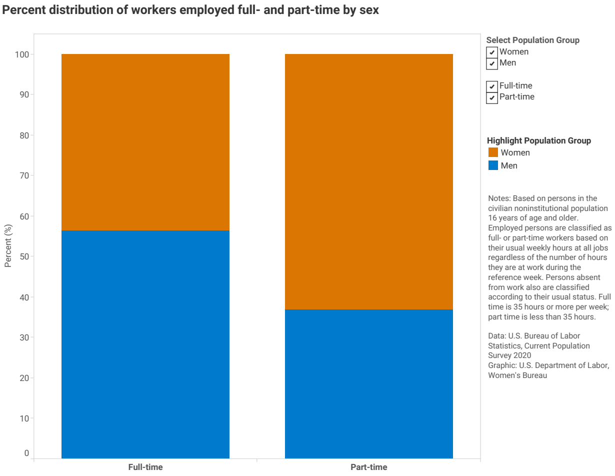
Global Gender Gap Index *(Schwab et al., 2017)*

Cumulative measure for ranking countries which benchmarks national gender gaps on economic, education, health, and political criteria

U.S. ranking: 49/144, overall score: 0.718 ($1 =$ gender parity)
$\rightarrow$ men more participatory in labor force; women tend to earn less than men, be employed part-time, or not to be paid for their work, and are underrepresented in managerial and higher-paying jobs

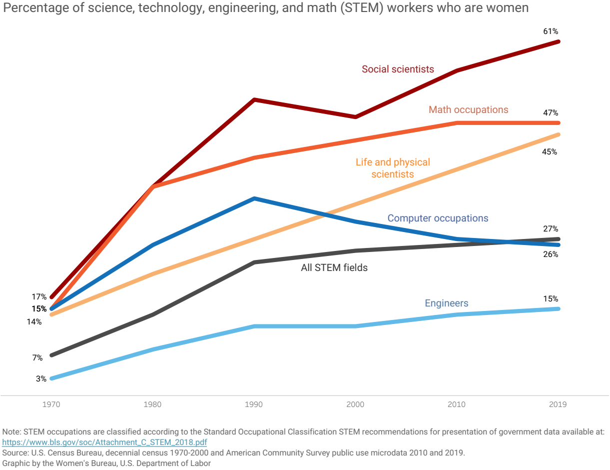
Some possible reasons:

- Statistical discrimination *(Tilcsik, 2021)*
- Institutional environment *(Beggs, 1995)*
- Unequal bargaining power *(Folbre, 2021)*



Percent distribution of workers employed full- and part-time by sex (2020).
U.S. Department of Labor.
Source: `https://www.dol.gov/agencies/wb/data`.

Sociological Perspective – Data & Statistics



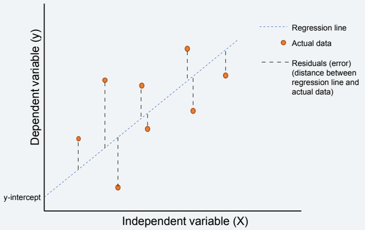
Percentage of Science, Technology, Engineering, and Math (STEM) workers who are women (1970-2019). U.S. Department of Labor. Source: `https://www.dol.gov/agencies/wb/data`.

- The 'Glassdoor Method': a framework for evaluating gender pay gap which relies on linear regression *(Chamberlain, 2017)*

Linear regression

Preprocessing technique used to smooth out noise or to find patterns within a dataset, which attempts to model the relationship between two or more variables by fitting data to a linear equation

$$y = \beta_0 + \beta_1 x + \epsilon$$



Simple linear regression graph.
Source: `https://www.reneshbedre.com/assets/posts/reg/mlr/residual.svg`.

Tools for Assessing Fairness: FAIR-DB

- FAIR-DB: an algorithm to detect bias in data based on functional dependencies and the related evaluation metrics *(Azzalini, Criscuolo, and Tanca, 2021)*

Functional Dependencies (FDs)

Constraint involving two (sets of) attributes of the same relation in which the first uniquely determines the second

Approximate Conditional Functional Dependencies (ACFDs)

FDs holding on a subset of tuples (Approximate) which use conditions on attribute values to specify the subset on which they hold (Conditional)

$$\mathit{Status} = \text{` F '}, \mathit{Gender} = \text{` female '} \rightarrow \mathit{AnnualSalaryBin} = \text{`} \leq 90\text{K '}$$

Tools for Assessing Fairness: Ranking Facts

- Ranking Facts: an application built on the idea of ranking which makes use of three statistical measures to evaluate fairness *(Yang et al., 2018)*
  - FA*IR: compares the number of protected elements in every prefix of the ranking (i.e., the top-i positions, with $i \in [1, k]$) with the expected number of protected elements if they were picked at random using Bernoulli trials with success probability p
  - Proportion: statistical measure based on the concept of (two-sample) z-test – particular type of hypothesis test which allows to compare two proportions to check whether they are the same
  - Pairwise: compares options in pairs and determines which is the preferred choice or has the highest level of importance based on defined criteria, ultimately ranking the options

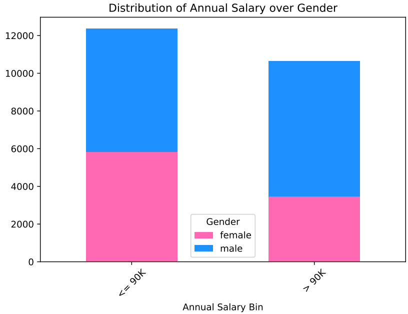
- Data preprocessing: 20,309 tuples, of which 16,146 males and 4,163 females, and with 35 distinct *Job Title* values and 20 distinct *Department* values

- The 'Glassdoor Method': 24.2% 'unadjusted' pay gap; 0.4% 'adjusted' pay gap → no evidence of a systematic gender pay gap
- FAIR-DB: 6 final functional dependencies; 11.4% of the dataset 'problematic' → dataset quite fair
- Ranking Facts: dataset fair for both males and females, for each statistical measure

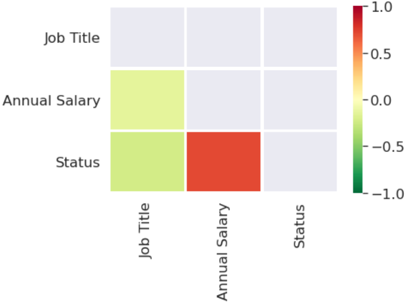
Case Study 2: San Francisco

- Data preprocessing: 22,996 tuples, of which 13,688 males and 9,308 females, and with 81 distinct *Job Title* values

- The 'Glassdoor Method': 30.4% 'unadjusted' pay gap; −5% 'adjusted' pay gap → no evidence of a systematic gender pay gap
- FAIR-DB: 10 final functional dependencies; 24.3% of the dataset 'problematic' → dataset slightly biased (at least 10% of women earn less than the male counterpart), and for higher-paying jobs men seem to have an economic advantage over women
- Ranking Facts: dataset fair for males and unfair for females, for each statistical measure → proportion of women in the top-k ranking effectively very low



(a) Distribution of the *Annual Salary* values for the San Francisco dataset (2 bins).



(b) Heatmap showing attribute correlations for the San Francisco dataset.

The Impact of Other Design Choices

- Part-time employees removal: most of the tuples removed related to women (Chicago); excessive amount of tuples removed (San Francisco)
- FAIR-DB: discretization using more bins: less and different final dependencies detected (Chicago and San Francisco)
- FAIR-DB: choice of different dependencies: 85.6% (Chicago) and 92.5% (San Francisco) of the dataset 'problematic'
- Grouping of job titles: overturning of the outcomes for Ranking Facts (Chicago dataset unfair for males and fair for females, for each statistical measure)
- Voluntary introduction of bias: results from each tool oriented toward unfair Chicago dataset, in which women are discriminated against (retaining 50%, 75%, and 90% of the *Annual Salary* value of female employees)

Outcomes

- Strengths and weaknesses of the tools highlight their non-exhaustiveness and complementarity
- Tools practically fail in capturing the several facets of equity

Equity

The idea that people should have access to resources (possibly of a different nature and to a different extent) in order to be able to reach the same condition, ultimately pointing at fairness

- Representation problem: disproportion in the percentage of women employed in different sectors
- Part-time problem: higher number of women employed in part-time jobs, typically less paid than full-time ones

Contributions

- Fairness is a multifaceted concept which cannot be exhausted by providing a single definition and pursuing it experimentally
  - Different tools provide different results (e.g., San Francisco)
- Tools are susceptible to decisional choices, and therefore users must be properly trained on the specific area of analysis
  - Grouping similar job titles or using more bins impacts on the results
- Sociological background provides the user with knowledge which can be used to make design decisions and parameterize the tools $\rightarrow$ dual perspective on the gender pay gap issue emphasized the importance of multidisciplinarity, especially when dealing with problems of an ethical and sociological nature
  - Even 'fair' datasets are not representative, and none of the tools addressed the issue

Future Work

- Combine all the tools in a unique, more complete instrument, possibly trying to encompass even more facets of equity, or different approaches to fairness
- Support analyses of this kind by sociological research
- Enrich sociological research by conducting an interview with workers and HR practitioners of the cities under study
- Retrieve further information in support of the mere data and create effective documentation, possibly pointing at context-awareness (provide the tools with knowledge on the context of use)

Thank you!

Any question?

References

Julia Angwin et al. “Machine Bias”. In: *ProPublica* (2016). https://www.propublica.org.

Fabio Azzalini, Chiara Criscuolo, and Letizia Tanca. “FAIR-DB: FunctionAl DependencIes to discoveR Data Bias”. In: *EDBT/ICDT Workshops*. http://ceur-ws.org/Vol-2841. 2021.

John J Beggs. “The Institutional Environment: Implications for Race and Gender Inequality in the U.S. Labor Market”. In: *American Sociological Review* 60.4 (1995), pp. 612–633. DOI: 10.2307/2096297.

Cambridge Advanced Learner’s Dictionary. ““Gender gap””. In: *Cambridge University Press* (2013). https://dictionary.cambridge.org. Accessed 28 June 2021.

Andrew Chamberlain. “How to Analyze Your Gender Pay Gap: An Employer’s Guide”. In: *Glassdoor Economic Research* (2017). https://www.glassdoor.com.

References

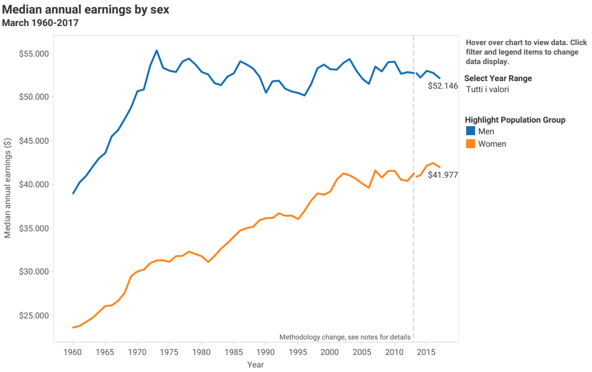
Jeffrey Dastin. "Amazon scraps secret AI recruiting tool that showed bias against women". In: *Reuters* (2018). https://www.reuters.com.

Nancy Folbre. "Gender inequality and bargaining in the U.S. labor market". In: *Economic Policy Institute* (2021). https://www.epi.org/209716.

Klaus Schwab et al. "The Global Gender Gap Report 2017". In: https://www.weforum.org. World Economic Forum. 2017. ISBN: 9781944835125.

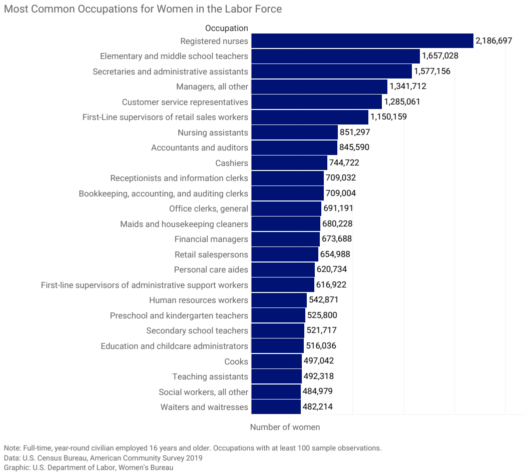
András Tilcsik. "Statistical Discrimination and the Rationalization of Stereotypes". In: *American Sociological Review* (2021). DOI: 10.31235/osf.io/gf7ry.

Ke Yang et al. "A Nutritional Label for Rankings". In: *Proceedings of the 2018 International Conference on Management of Data*. Association for Computing Machinery, 2018, pp. 1773–1776. DOI: 10.1145/3183713.3193568.



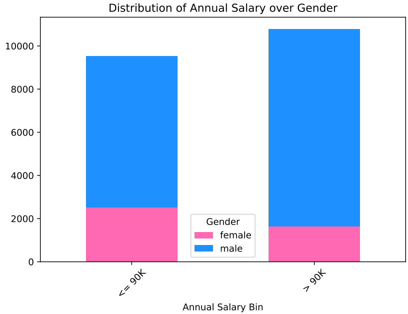
Notes: Earnings are based on median annual earnings of full-time, year-round workers, 15 years old and over beginning in March, 1980, and age 14 years old and over as of March of the following year for previous years. Before 1989 earnings are for civilian workers only.
The comparability of historical data has been affected at various times by methodological and other changes in the Current Population Survey. The 2014 CPS ASEC included redesigned questions for income and health insurance coverage for a subsample of the 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions, resulting in two estimates for 2013. Estimates based on the portion of the sample that received the redesigned income questions are the most appropriate for comparing estimates from ASEC 2014 with ASEC 2015 and beyond.
Earnings are in 2017 CPI-U-RS adjusted dollars.
Source: 1961-2018 Annual Social and Economic Supplements, Current Population Survey, U.S. Census Bureau.
Graph by the Women's Bureau, U.S. Department of Labor

Median annual earnings by sex (1960-2017).
U.S. Department of Labor.
Source: `https://www.dol.gov/agencies/wb/data`.

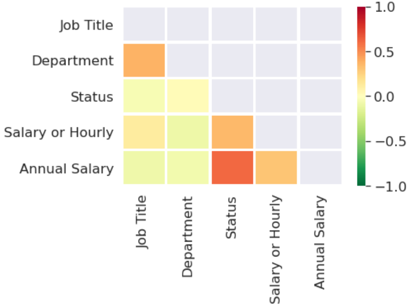


Most common occupations for women in the labor force (2019). U.S. Department of Labor. Source: `https://www.dol.gov/agencies/wb/data`.

[Extra] Case Study 1: Chicago



(a) Distribution of the *Annual Salary* values for the Chicago dataset (2 bins).



(b) Heatmap showing attribute correlations for the Chicago dataset.

[Extra] (Some) Limitations

- Sociological
  - Non-specific literature
  - No direct insights from U.S. workers
- Technical
  - Original datasets already partial or possibly grouped
  - `gender-guesser` (package to infer employees' gender)
- Design
  - Removal of job titles with less than 100 occurrences
  - FAIR-DB: parameter values; manual selection of rules; number of bins