



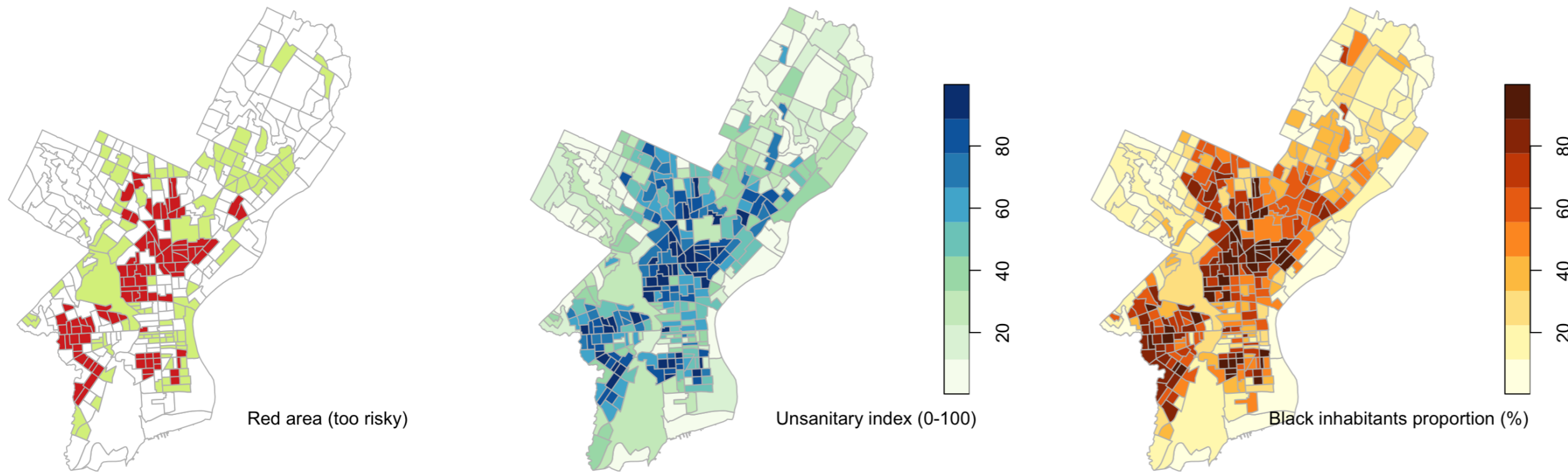
Insurance, discrimination and fairness

Agathe Fernandes Machado, Francois Hu, Philipp Ratz and **Arthur Charpentier**

Universite du Quebec a Montreal, <https://freakonometrics.github.io/>

"Technology is neither good nor bad; nor is it neutral," Kranzberg (1986)

Motivation: Redlining



(Fictitious maps, inspired by a Home Owners' Loan Corporation map from 1937)
Federal Home Loan Bank Board (FHLBB) "residential security maps" (for real-estate investments)

Discrimination in insurance as an "ill-posed problem"?

Discrimination and Insurance

"What is unique about insurance is that even statistical discrimination which by definition is absent of any malicious intentions, poses significant moral and legal challenges. Why? Because on the one hand, policy makers would like insurers to treat their insureds equally, without discriminating based on race, gender, age, or other characteristics, even if it makes statistical sense to discriminate (...). On the other hand, at the core of insurance business lies discrimination between risky and non-risky insureds. But riskiness often statistically correlates with the same characteristics policy makers would like to prohibit insurers from taking into account," Avraham (2017).

Regulation (Quebec & Colorado)

– Article 20.1 –

In an insurance or pension contract, a social benefits plan, a retirement, pension or insurance plan, or a public pension or public insurance plan, a distinction, exclusion or preference based on age, sex or civil status is deemed non-discriminatory where the use thereof is warranted and the basis therefor is a risk determination factor based on actuarial data

Charte des droits et libertes de la personne

– Section 5 (Estimating Race and Ethnicity) –

Insurers shall estimate the race or ethnicity of all proposed insureds that have applied for coverage on or after the insurer's initial adoption of the use of ECDIS, or algorithms and predictive models that use ECDIS, including a third party acting on behalf of the insurer that used ECDIS, or algorithms and predictive models that used ECDIS, in the underwriting decision-making process, by utilizing: BIFSG and the insureds' or proposed insureds' name and geolocation (...)

Concerning Quantitative Testing of External Consumer Data and Information Sources, Algorithms, and Predictive Models Used for Life Insurance Underwriting for Unfairly Discriminatory Outcomes, September 27, 2023

Bayesian Improved First Name Surname Geocoding, or "BIFSG" and External Consumer Data and Information Source, or "ECDIS"

Actuaries

"To be an actuary is to be a specialist in generalization, and actuaries engage in a form of decision making that is sometimes called actuarial. Actuaries guide insurance companies in making decisions about large categories that have the effect of attributing to the entire category certain characteristics that are probabilistically indicated by membership in the category, but that still may not be possessed by a particular member of the category," Schauer (2006).

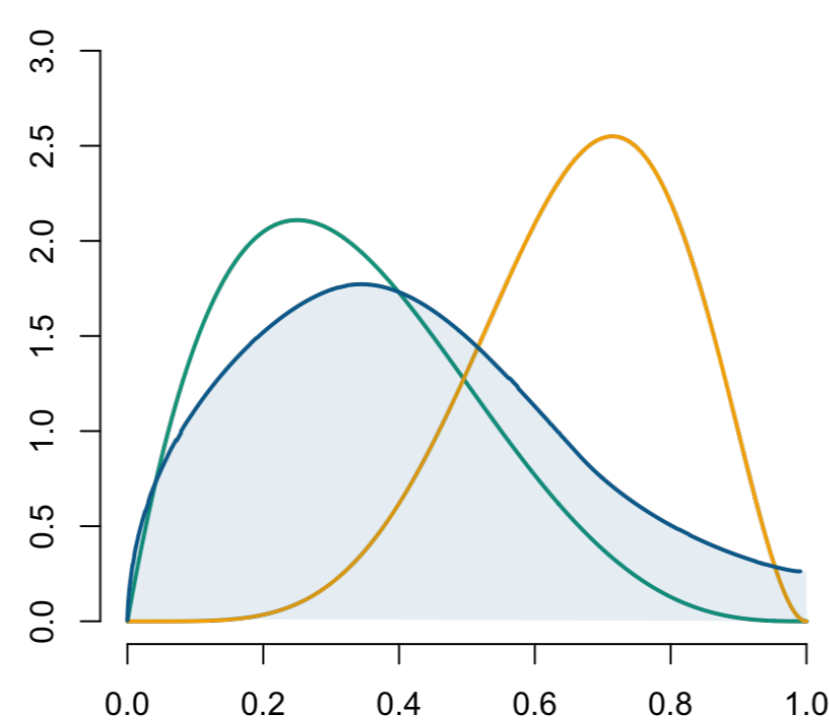
Brian Glenn coined "the myth of the actuary, a powerful rhetorical situation in which decisions appear to be based on objectively determined criteria when they are also largely based on subjective ones" or "the subjective nature of a seemingly objective process," Glenn (2000).

Mitigation

Wasserstein barycenter,

$$P_{m^*} = \underset{Q}{\operatorname{argmin}} \int_{\mathcal{S}} d_w^2(Q; P_s) \times \prod_{s \in \mathcal{S}} f_A; B_g$$

$$\begin{aligned} m^*(x; s = A) &= P[S = A] m(x; s = A) \\ &+ P[S = B] F_B^{-1}(F_A(m(x; s = A))) \\ m^*(x; s = B) &= P[S = A] F_A^{-1}(F_B(m(x; s = B))) \\ &+ P[S = B] m(x; s = B) \end{aligned}$$



Supreme Court Justice Harry Blackmun, 1978, "In order to get beyond racism, we must first take account of race. There is no other way. And in order to treat some persons equally, we must treat them differently."

John G. Roberts of the U.S. Supreme Court, 2007 "The way to stop discrimination on the basis of race is to stop discriminating on the basis of race."

Mathematical Formalism

"Machine learning won't give you anything like gender neutrality 'for free' that you didn't explicitly ask for," Kearns and Roth (2019)

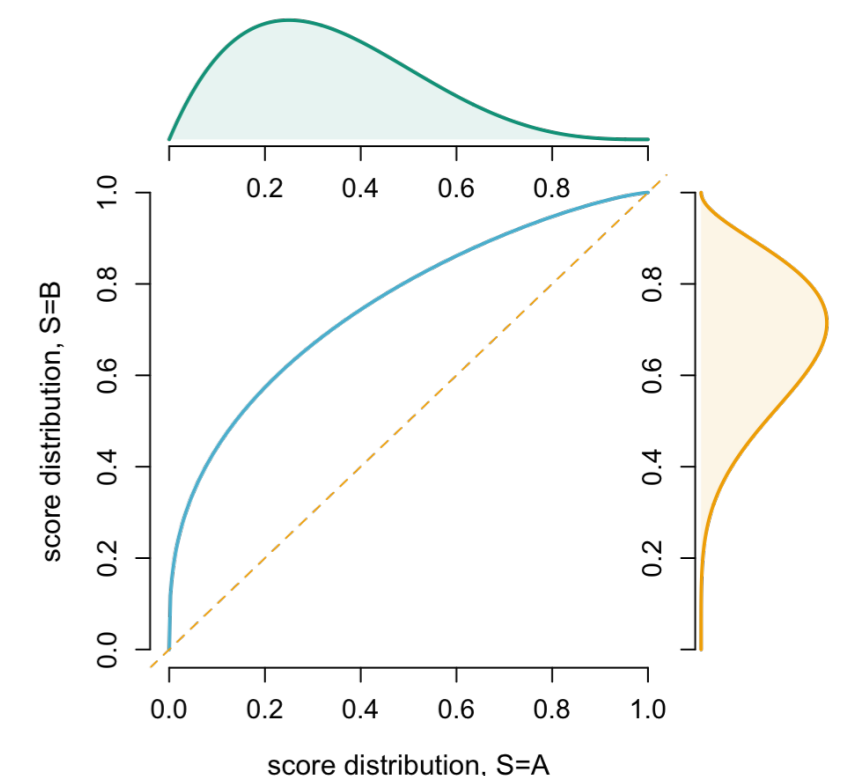
$\mathcal{X} \subset \mathbb{R}^d$: 'explanatory' variables
 $\mathcal{S} \subset \mathcal{A}; \mathcal{B}$: "sensitive variable"
 $\mathcal{Y} \subset \{0, 1\}$: classification problem
 $\hat{y} \in \{0, 1\}$: prediction, classically $\hat{y} = 1(m(x; s) > t)$

A model m satisfies the **independence property** if $m(\mathcal{X}; \mathcal{S}) \perp\!\!\!\perp \mathcal{S}$, with respect to the distribution P of the triplet $(\mathcal{X}; \mathcal{S}; \mathcal{Y})$ **demographic parity**

A model m satisfies the **separation property** if $m(\mathcal{X}; \mathcal{S}) \perp\!\!\!\perp \mathcal{S} \mid \mathcal{Y}$, with respect to the distribution P of the triplet $(\mathcal{X}; \mathcal{S}; \mathcal{Y})$ **equalized odds**

A model m satisfies the **sufficiency property** if $\mathcal{Y} \perp\!\!\!\perp \mathcal{S} \mid m(\mathcal{X}; \mathcal{S})$, with respect to the distribution P of the triplet $(\mathcal{X}; \mathcal{S}; \mathcal{Y})$ **calibration**

We have **counterfactual fairness** if "had the protected attributes (e.g., race) of the individual been different, other things being equal, the decision would have remained the same," Kusner et al. (2017)



References

Avraham, R. (2017). Discrimination and insurance. In Lippert-Rasmussen, K., editor, *Handbook of the Ethics of Discrimination*, pages 335{347.

Charpentier, A. (2024). *Insurance: biases, discrimination and fairness*. Springer Verlag.

Charpentier, A., Hu, F., and Ratz, P. (2023). Mitigating discrimination in insurance with Wasserstein barycenters. *BIAS, 3rd Workshop on Bias and Fairness in AI, International Workshop of ECML PKDD*.

Kearns, M. and Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kranzberg, M. (1986). Technology and history: "Kranzberg's laws". *Technology and culture*, 27(3):544{560.

Kusner, M. J., Loftus, J., Russell, C., and Silva, R. (2017). Counterfactual fairness. In *Neural Information Processing Systems*.

