

A COMPARATIVE STUDY OF FAIRNESS- ENHANCING INTERVENTIONS IN MACHINE LEARNING

Sorelle Friedler, **Carlos Scheidegger**, Suresh Venkatasubramanian,
Sonam Chaudhary, Evan P. Hamilton, Derek Roth

Haverford College
University of Arizona
University of Utah

Fairness, Accountability, Transparency
Atlanta, GA, 2019-01-31

PRE-FAT* HISTORY

“SOCIAL IMPACT? WHAT?”

“IT’S JUST MATH, OF COURSE
IT’S OBJECTIVE”

BECAUSE OF YOU, WE ARE
STARTING TO MOVE BEYOND
THAT

LAST YEAR: “21 FAIRNESS
DEFINITIONS AND THEIR
POLITICS”

TODAY: WHAT HAPPENS
WHEN YOU EXPERIMENTALLY
COMPARE THEM

WE RAN A COMPARISON ON A
NUMBER OF EXISTING TECHNIQUES,
DATASETS, AND MEASURES

TL;DR: **IT'S COMPLICATED**

1) **FAIRNESS/ACCURACY**
FIGURES ARE FRAGILE:
DEPEND ON PREPROCESSING

2) **FAIRNESS MEASURES**
APPEAR TO BE EXPERIMENTALLY
CORRELATED

3) ON ANY GIVEN DATASET,
**DIFFERENT INTERVENTIONS
PERFORM SIGNIFICANTLY
DIFFERENTLY**

DATASETS

- Ricci v. deStefano
- Adult Income
- German Credit
- Propublica Recidivism datasets
- *(your dataset here! see our tutorial notes)*

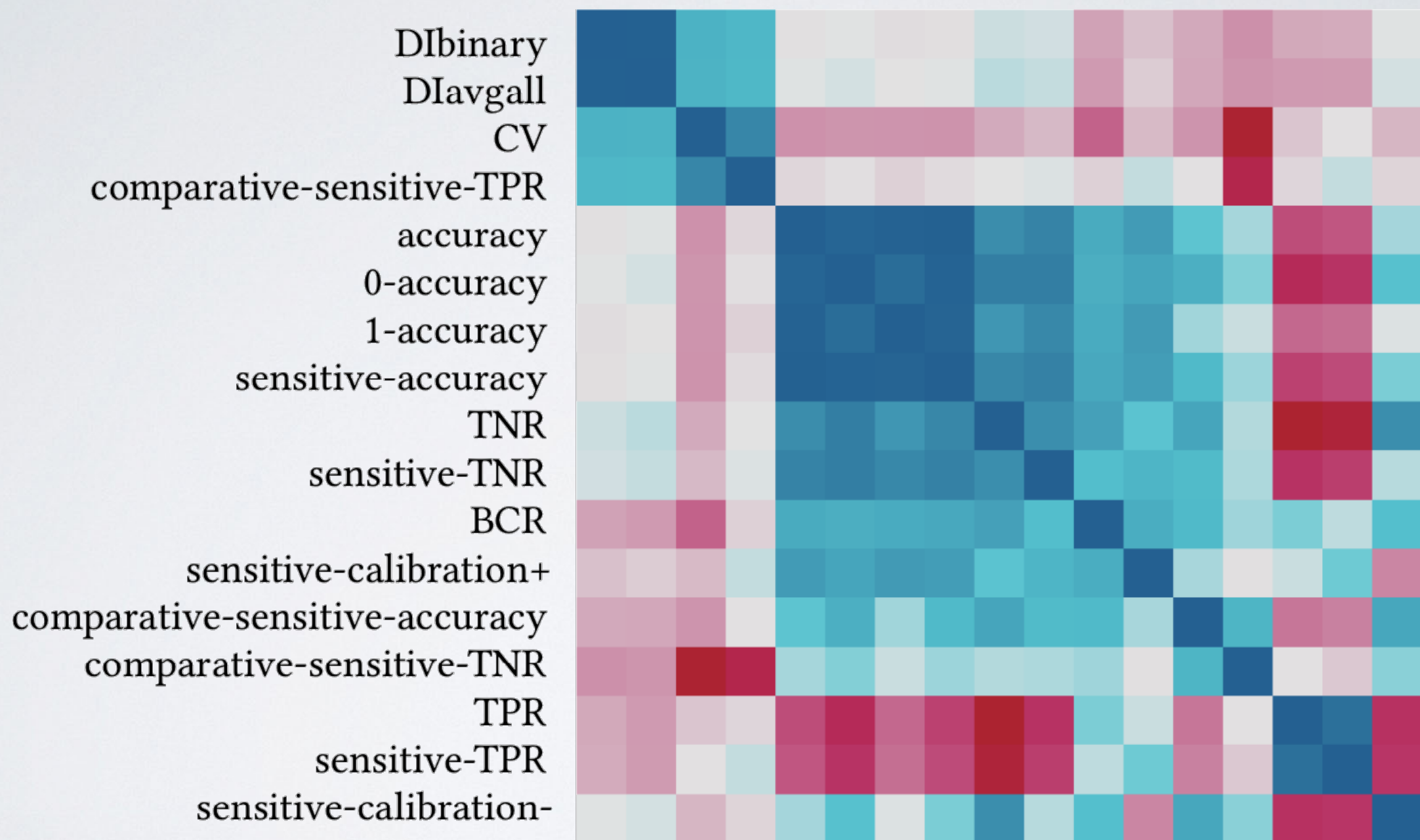
INTERVENTIONS

- Calders and Verwer's Two Naive Bayes
- Feldman et al.
- Kamishima et al.
- Zafar et al.
- *(your intervention here! see our tutorial notes)*

MEASURES

- Accuracy Measures
- Fairness Measures
- *(your measure here! see tutorial notes)*

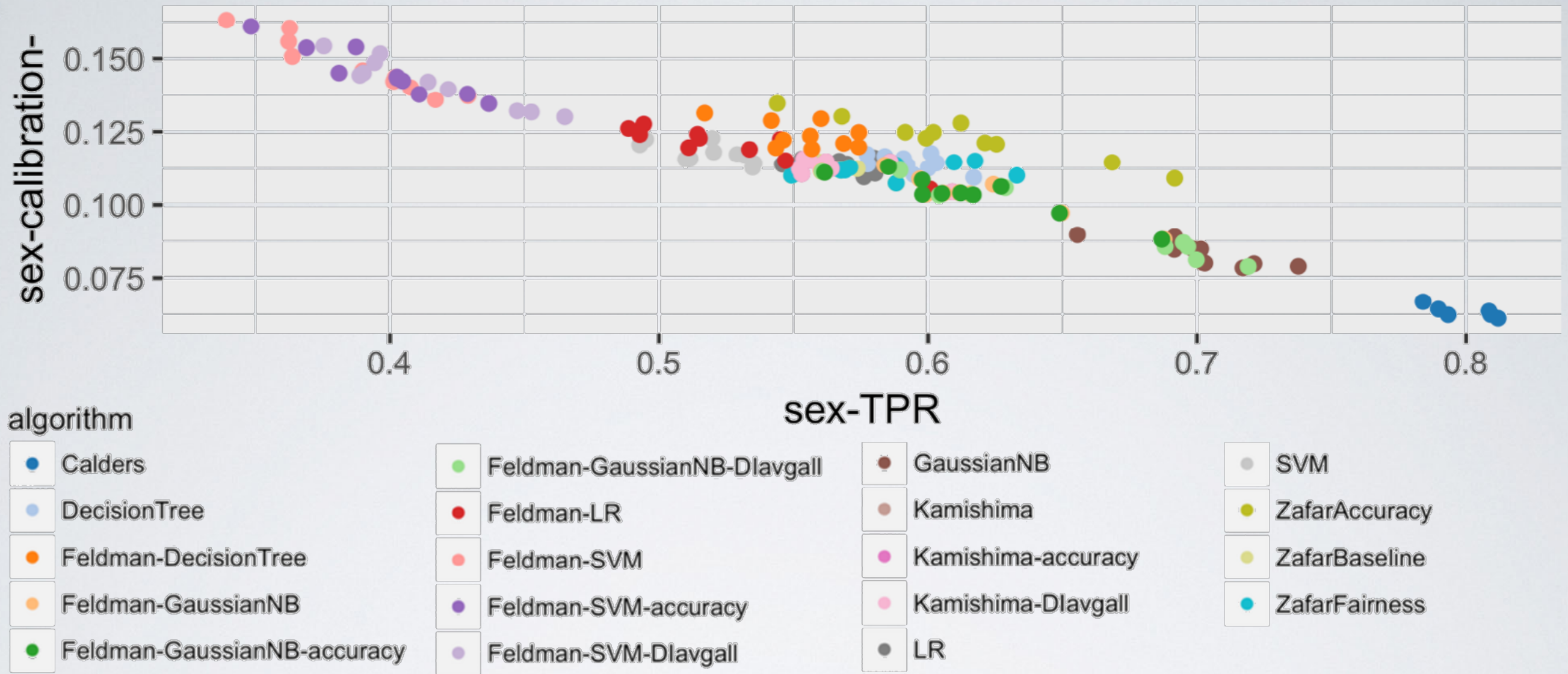
MEASURES CORRELATE



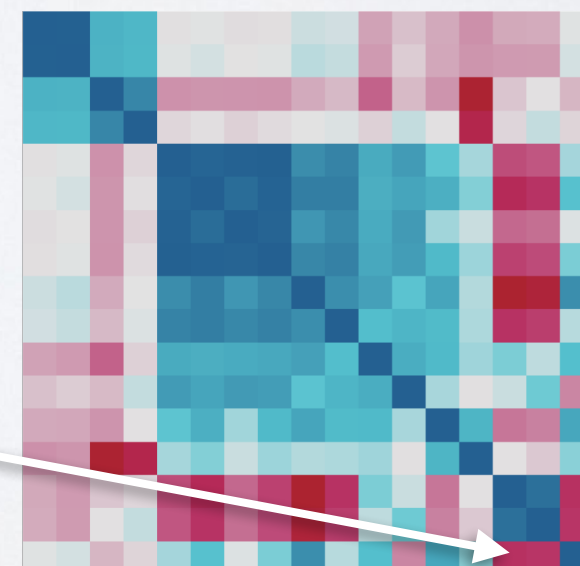
negative correlation

positive correlation

Adult dataset, Sex attribute

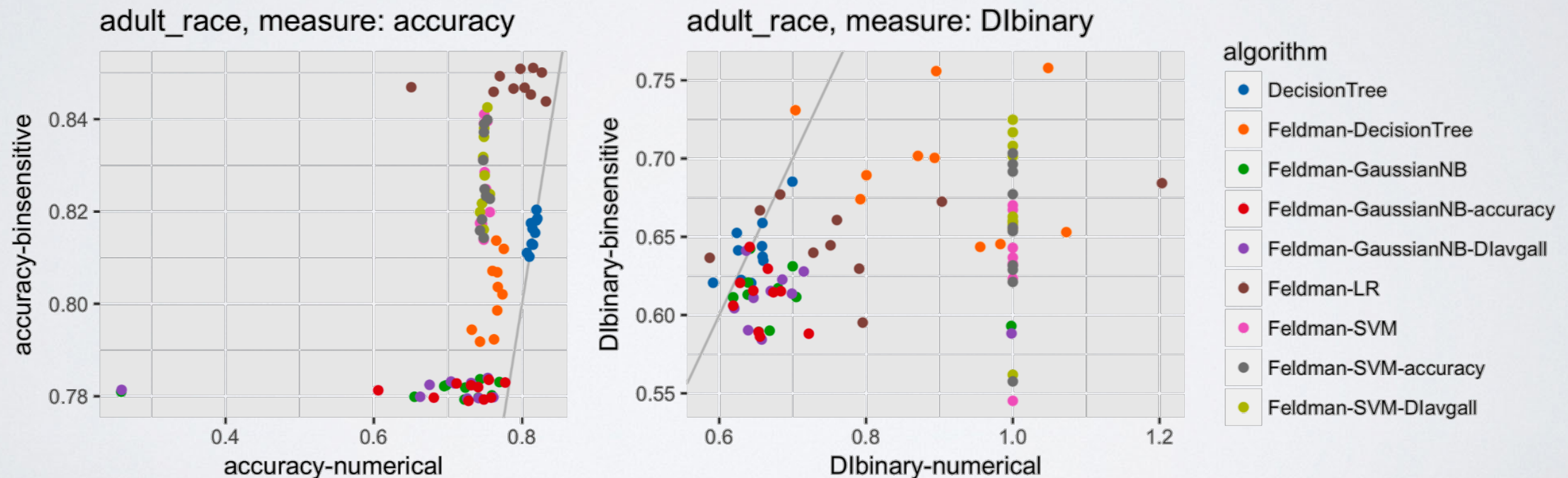


We are looking at that specific pairwise comparison of measures

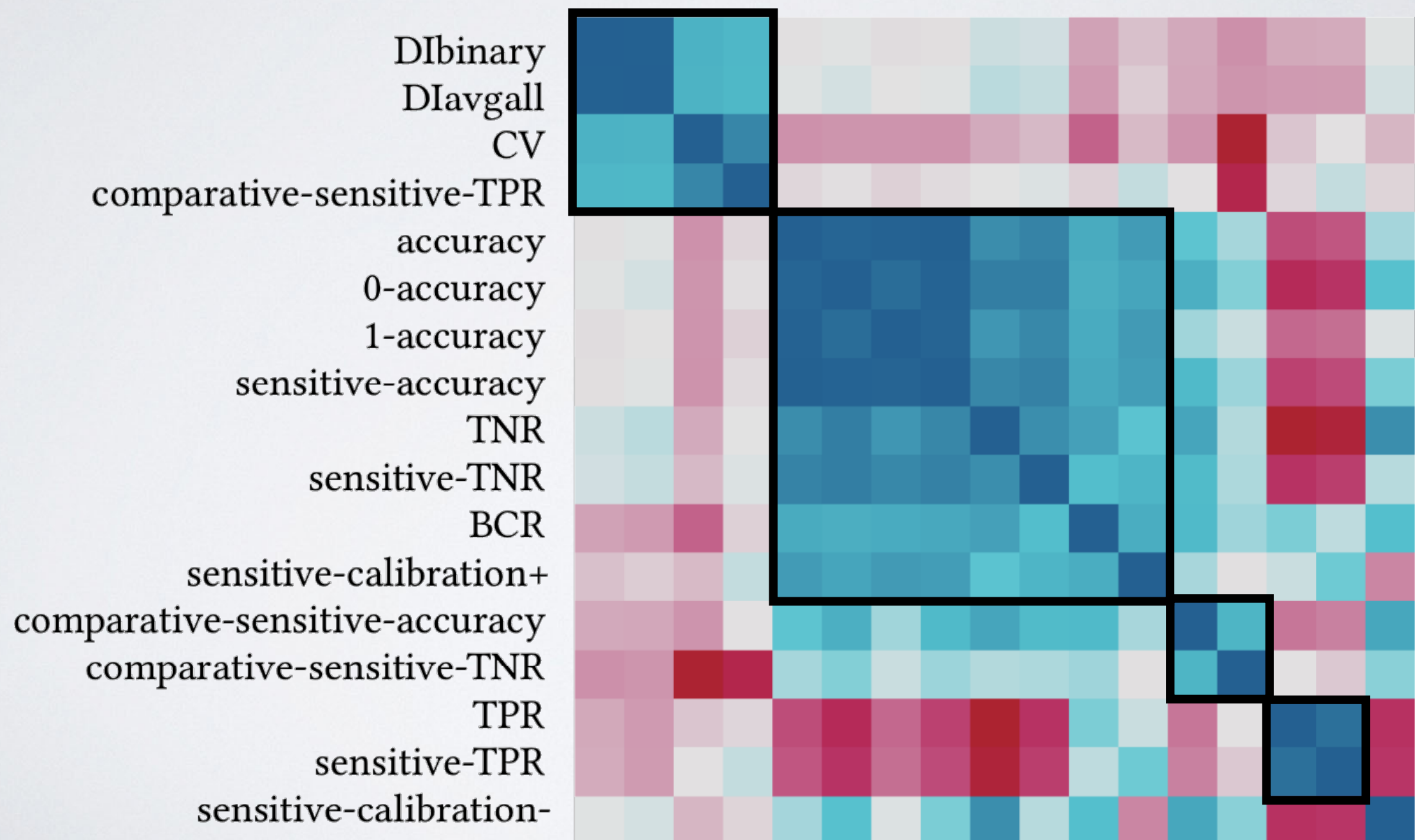


TAKEAWAYS

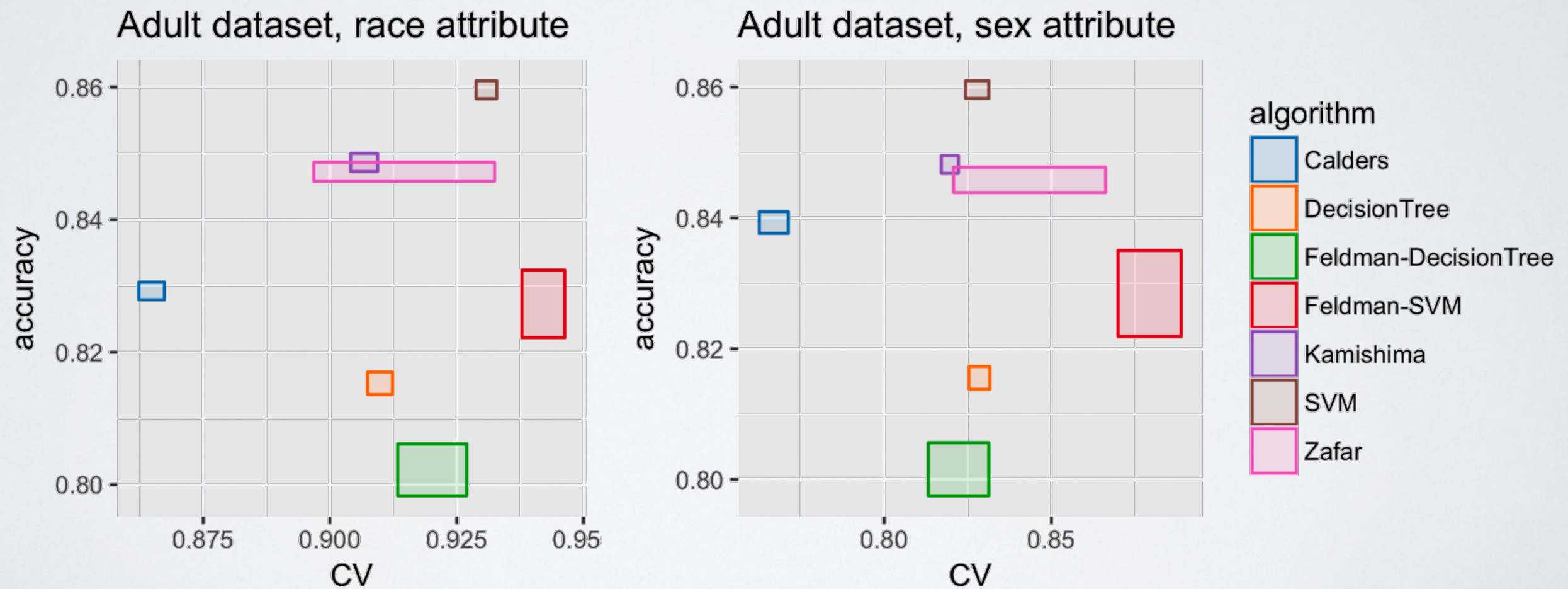
DON'T ABSTRACT OVER PREPROCESSING REQUIREMENTS



AVOID PROLIFERATION OF MEASURES (MAYBE)



EXPECT VARIABILITY



THANK YOU, AND JOIN US!

<https://github.com/algofairness/fairness-comparison>

Carlos Scheidegger
@scheidegger