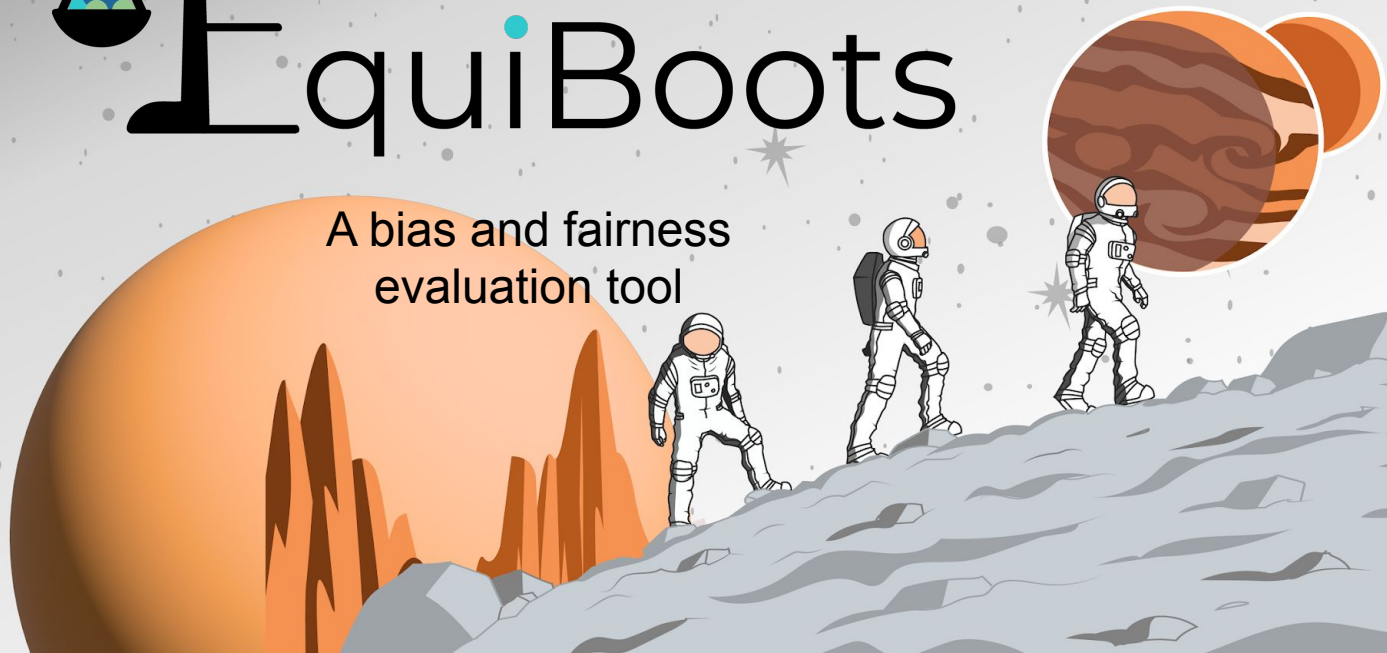
UCLA Health

Leon Shpaner, Arthur Funnell

jupytercon

EquiBoots

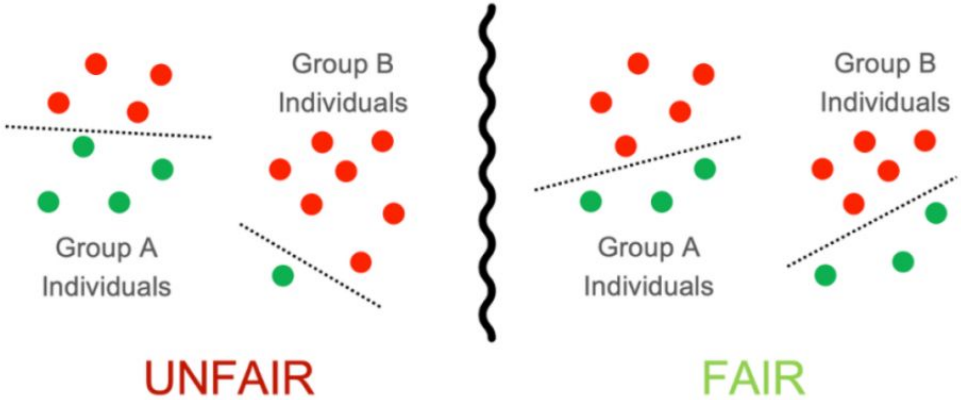
A bias and fairness evaluation tool



#JupyterCon

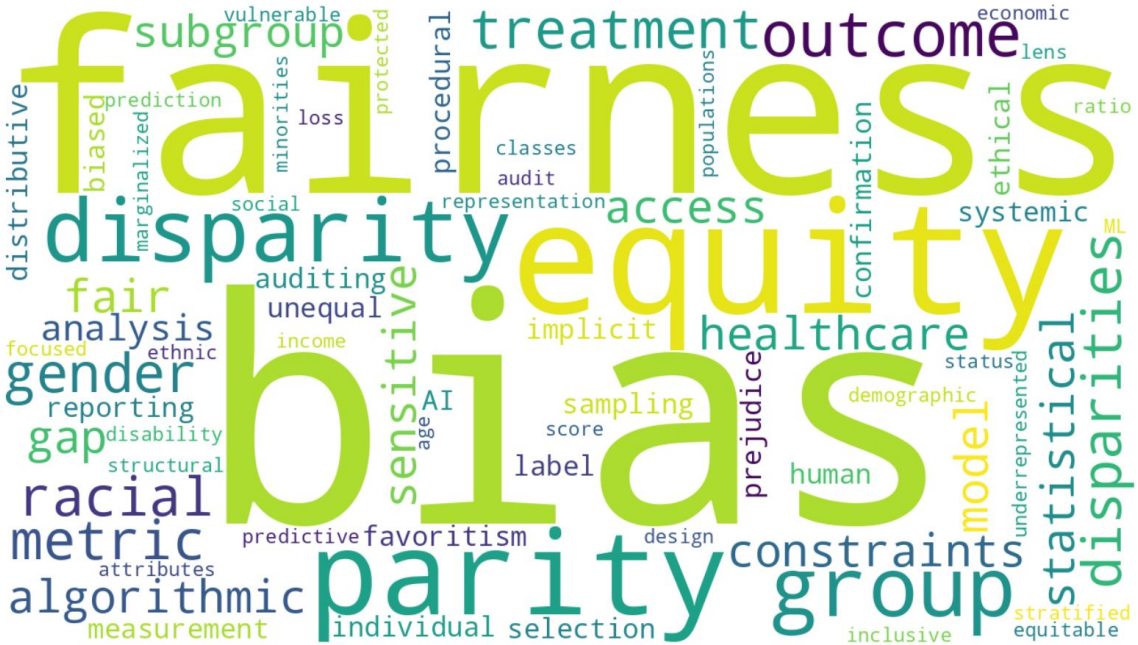
EquiBoots

- Identifies Biases
- A Range of Informative Plots
- Bootstrapping
- Statistical Tests
- Model Healing



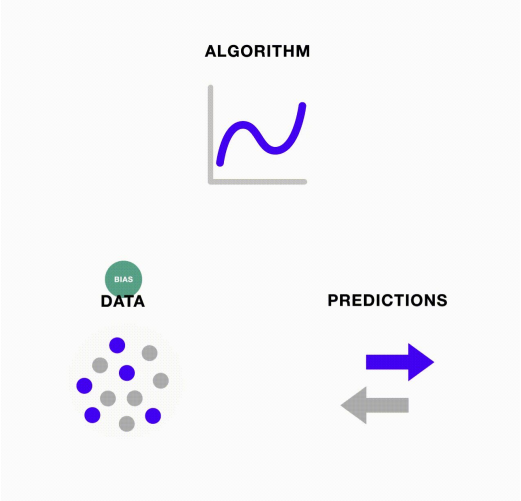
https://www.amazon.science/research-awards/success-stories/algorithmic-bias-and-fairness-in-machine-learning

What is bias and fairness?

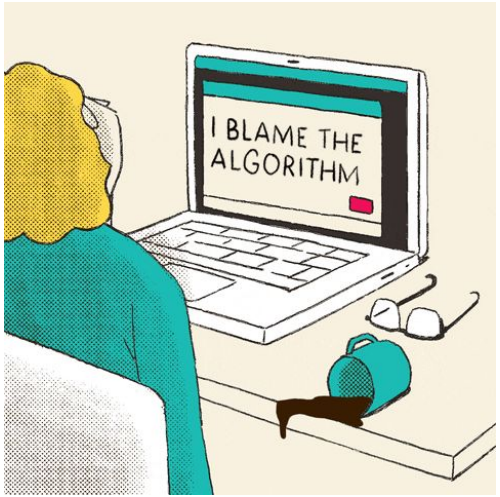


What is bias and fairness?

- Data can be unrepresentative, unfair, and biased.
  - **Data bias**: When the training data reflects existing societal prejudices
  - **Algorithmic bias**: When the model or algorithm generates unfair outcomes for different groups



https://urban-institute.medium.com/the-myth-of-the-impartial-machine-9fecb291abe0



https://media.gifdb.com/i-blame-the-algorithm-data-spilled-coffee-f41c8wldih0kqoc5.gif

- Machine learning models trained on these data can be biased in the same way.



https://tenor.com/view/you-are-so-biased-vicky-jain-pinkvilla-your-view-seems-one-sided-gif-4460185384567877902

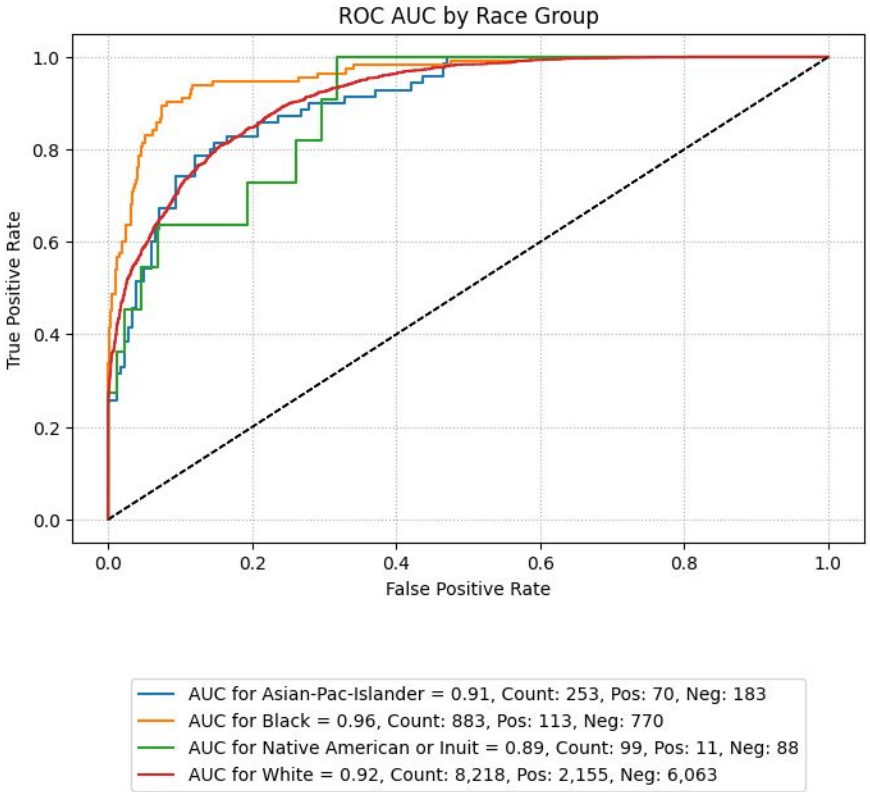
Real World Examples In Healthcare

1. *Dissecting racial bias in an algorithm used to manage the health of populations:*
   Less likely to recommend black patients based on cost.

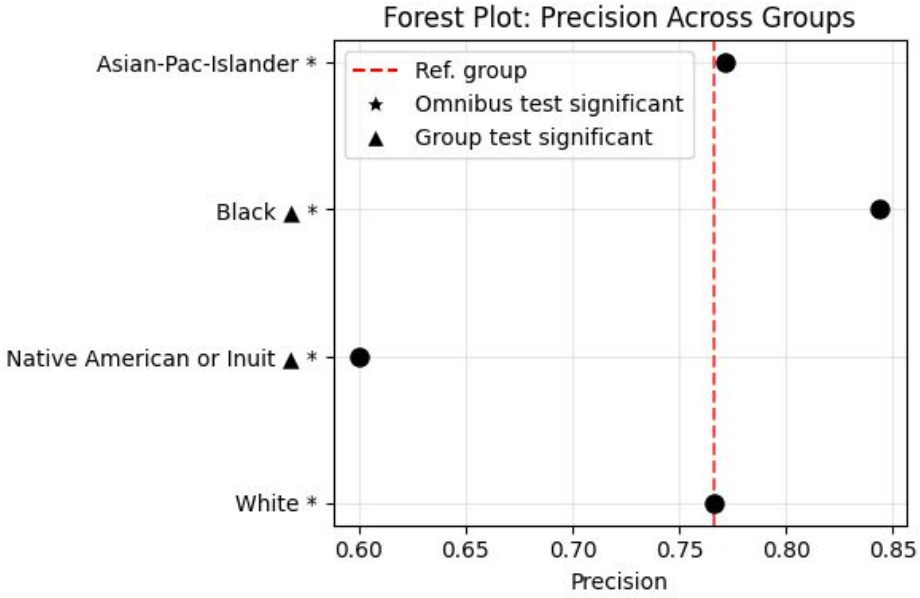
2. *Machine learning and health care disparities in dermatology:*
   Lack of diversity in dermatological datasets causing issues with classification tasks.

1. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342. PMID: 31649194.
2. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. JAMA Dermatol. 2018 Nov 1;154(11):1247-1248. doi: 10.1001/jamadermatol.2018.2348. PMID: 30073260.

Biased Model



Biased Threshold





https://tenor.com/view/i-am-kind-of-biased-andrew-baena-i'm-a-little-prejudiced-i-have-a-bias-sort-of-gif-8856442613648747536

EquiBoots: Tutorial

https://igit.me/equi



EquiBoots: Tutorial

Define Reference Group	Slice + Calculate Metrics	Plot + Statistical Tests
The most populous or groups that typically have bias in their favor. In our example: “White” “Male”	Split data into slices and calculate the metrics.	Plot metrics for comparison and conduct statistical tests to verify differences.

Set up: Initialise class and group

Point Estimates

```
sensitive_features = ['race', 'sex']

fairness_df = X_test[sensitive_features].reset_index(drop=True)

eq = eqb.EquiBoots(y_true=y_test,
                   y_pred=y_pred,
                   y_prob=y_prob,
                   fairness_df=fairness_df,
                   fairness_vars=sensitive_features)

eq.grouper(groupings_vars=sensitive_features)
```

[20] ✓ 0.0s

```
Groups created
```

Set up: Slice by sensitive variables

```
sliced_race_data = eq.slicer("race")
sliced_sex_data = eq.slicer('sex')

race_metrics = eq.get_metrics(sliced_race_data)
sex_metrics = eq.get_metrics(sliced_sex_data)
```

[25] ✓ 0.1s

```
race_metrics['White']
```

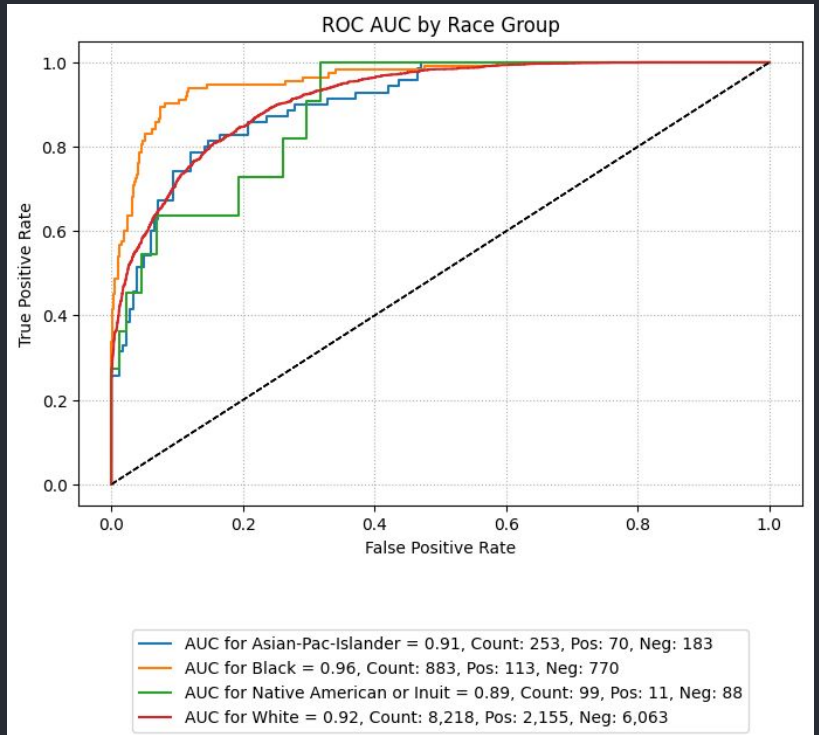
[28] ✓ 0.0s

```
{'Accuracy': 0.8551959114139693,
 'Precision': 0.7661334804191947,
 'Recall': 0.6445475638051044,
 'F1 Score': 0.7001008064516129,
 'Specificity': 0.9300676232888009,
 'TP Rate': 0.6445475638051044,
```

Group Comparison Graphs: ROC AUC

```
eqb.eq_plot_group_curves(
    sliced_race_data,
    curve_type="roc",
    title="ROC AUC by Race Group",
    exclude_groups=['Other']
)
```

✓ 0.1s



Statistical Tests: Running the tests

```
test_config = {
    "test_type": "chi_square",
    "alpha": 0.05,
    "adjust_method": "bonferroni",
    "confidence_level": 0.95,
    "classification_task": "binary_classification"
}
```

[29] ✓ 0.0s

```
stat_test_result_race = eq.analyze_statistical_significance(race_metrics,
                                                            "race",
                                                            test_config,)
```

[30] ✓ 0.0s

```
stat_test_result_sex = eq.analyze_statistical_significance(sex_metrics,
                                                           "sex",
                                                           test_config)
```

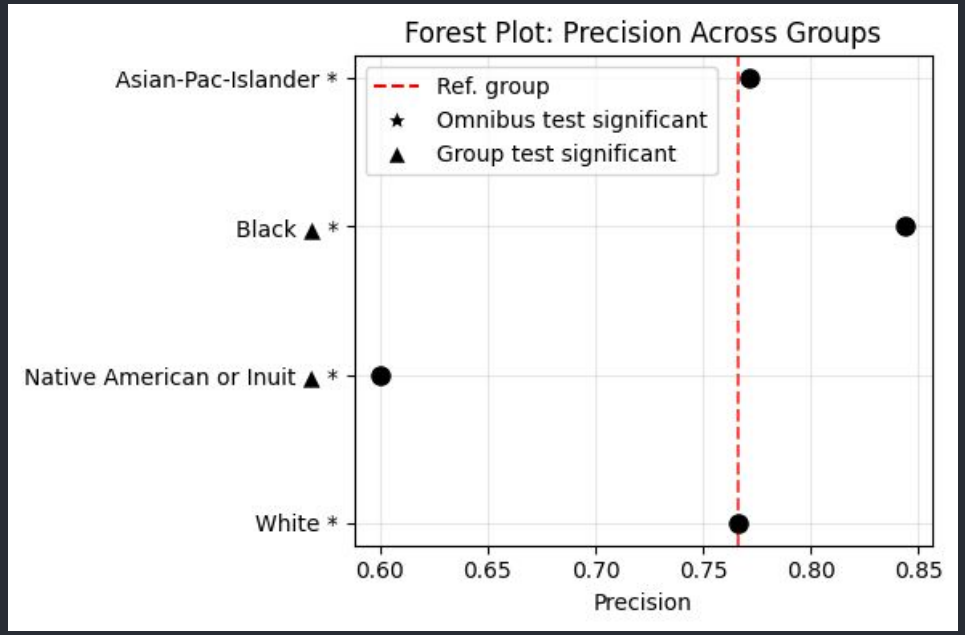
[31] ✓ 0.0s

Statistical Tests: Plotting Results

```
eqb.eq_plot_metrics_forest(
    group_metrics=race_metrics,
    metric_name="Precision",
    title="Forest Plot: Precision Across Groups",
    reference_group="White",
    statistical_tests=stat_test_results_race,

)
```

✓ 0.0s



BOOTSTRAPPING



Source: Back to the Future Part II

Bootstrapping: Initialising and Seeding

```
int_list = np.linspace(0, len(y_test), num=len(y_test), dtype=int).tolist()
```

[33] ✓ 0.0s

```
eqb_bootstrap = eqb.EquiBoots(
    y_true=y_test,
    y_pred=y_pred,
    y_prob=y_prob,
    fairness_df=fairness_df,
    fairness_vars=['race'],
    seeds=int_list,
    reference_groups=['White'],
    task='binary_classification',
    bootstrap_flag=True,
    num_bootstraps=5001,
    boot_sample_size=len(y_test),
    stratify_by_outcome=True
)
```

[34] ✓ 0.0s

Bootstrapping: Grouping and Slicing

```
eqb_bootstrap.grouper(groupings_vars=['race'])
```

[35] ✓ 1m 38.2s

```
Bootstrapping iterations:   0%|          | 0/5001 [00:00<?, ?it/s]/home/afunnell/Code/EquiBoo
  warnings.warn(
Bootstrapping iterations: 100%|██████████| 5001/5001 [01:38<00:00, 50.89it/s]
Groups created
```

```
bootstrap_race_data = eqb_bootstrap.slicer("race")
```

[36] ✓ 0.8s

```
/home/afunnell/Code/EquiBoots/equi_boots/src/equiboots/EquiBootsClass.py:331: UserWarning: Gr
  warnings.warn(
```

Bootstrapping: Metrics and Differences

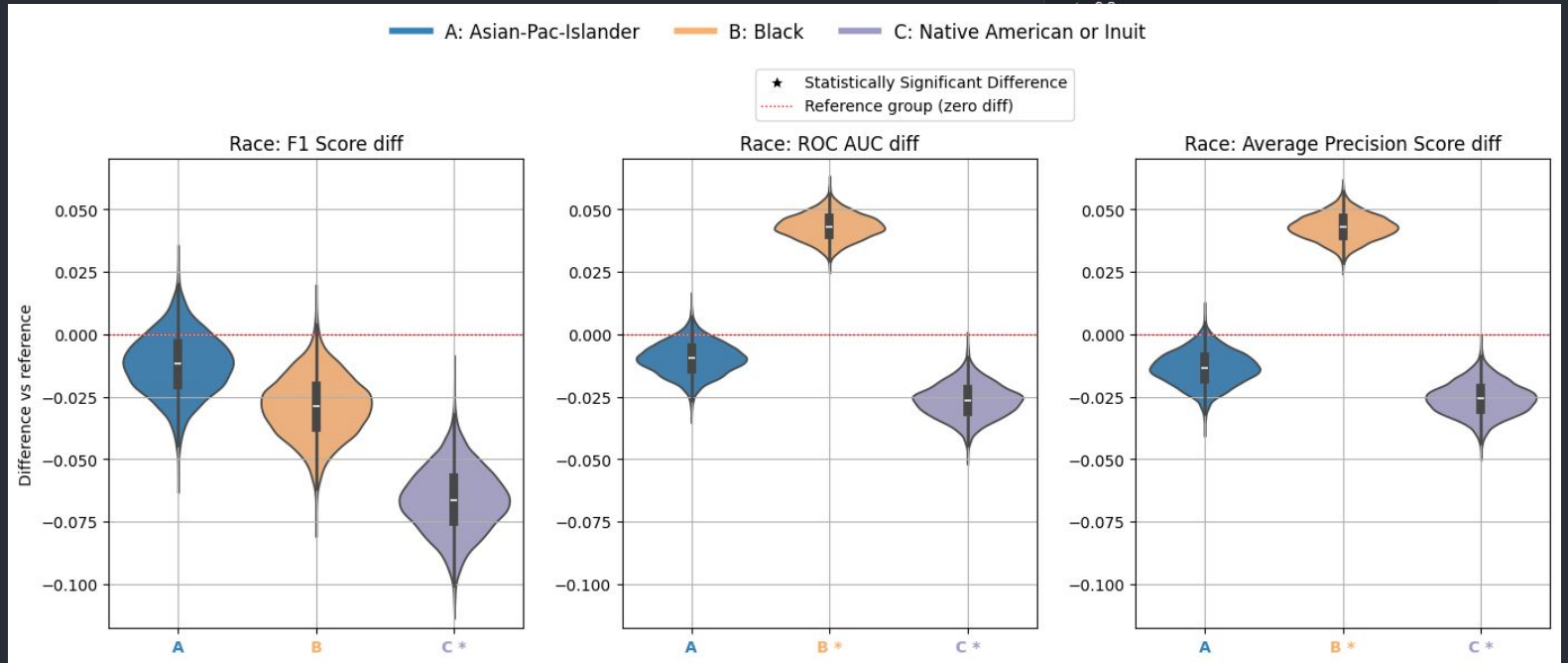
```
[37]  boots_race_metrics = eqb_bootstrap.get_metrics(bootstrap_race_data)
      ✓ 2m 37.2s

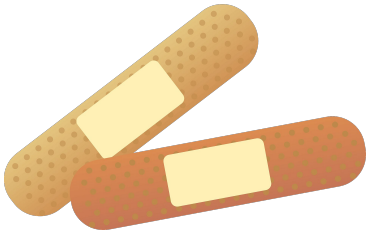
...   Calculating metrics for each bootstrap:
      100%|██████████| 5001/5001 [02:37<00:00, 31.81it/s]
```

```
[39]  diffs = eqb_bootstrap.calculate_differences(boots_race_metrics, "race")
      ✓ 0.1s
```

Bootstrapping: Differences

```
eqb.eq_group_metrics_plot(
    group_metrics=diffs,
    metric_cols=metrics_boot,
    name="race",
    categories="all",
    figsize=(14,5),
    plot_type="violinplot",
    color_by_group=True,
    show_grid=True,
    max_cols=3,
    statistical_tests=stat_test_results,
)
```





MODEL HEALING



Model Healing: Finding Thresholds

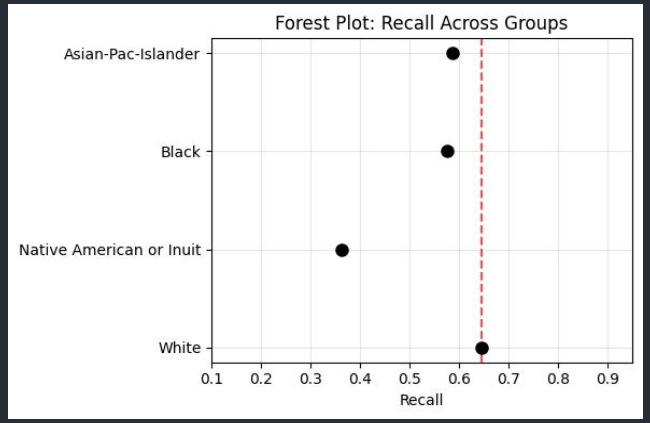
```
eqb.find_group_thresholds(y_true=y_test,
                          y_prob=y_prob,
                          reference_group='White',
                          group_vec=fairness_df['race'],
                          threshold_range=[0.1, 0.9],
                          n_steps=100,
                          default_threshold=0.5)
```

[40] ✓ 0.9s

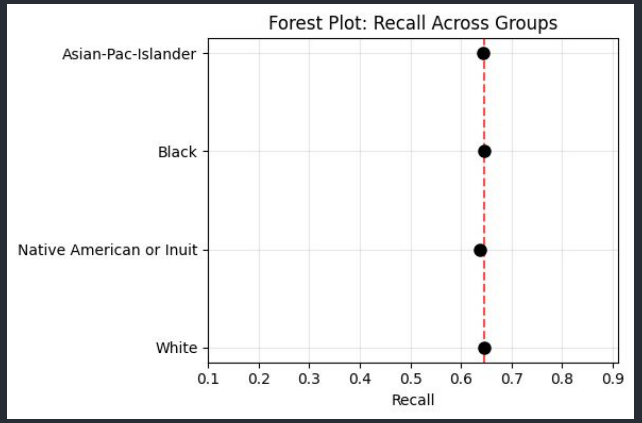
```
{'Asian-Pac-Islander': 0.48787878787879,
 'Black': 0.407070707070707,
 'Native American or Inuit': 0.3909090909090909,
 'Other': 0.17272727272727273,
 'White': 0.5}
```

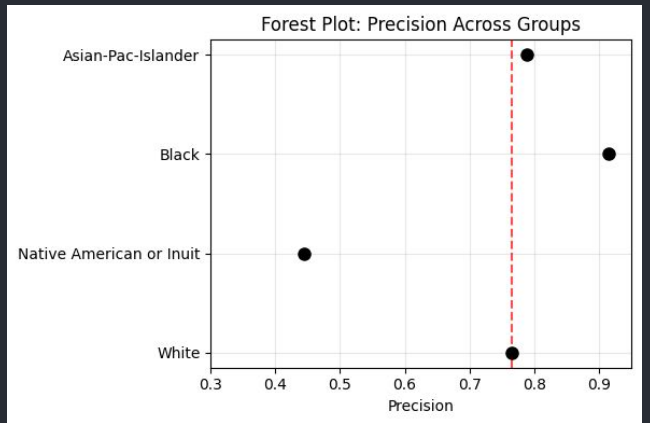
Model Healing: Finding Thresholds

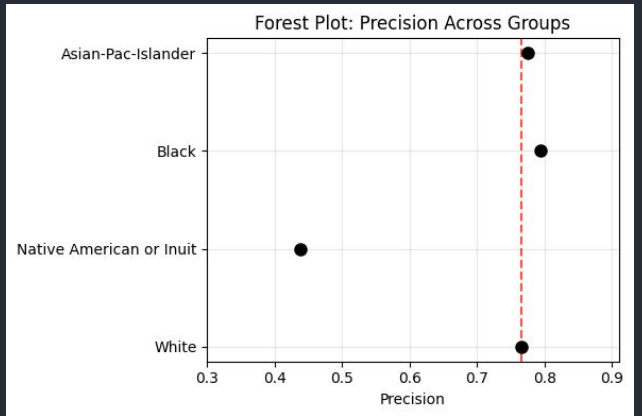
BEFORE



AFTER







EquiBoots: Toward Fairer Machine Learning

- Brings transparency
- Quantifies bias
- Empowers healing
- Bridges research and practice

Thank you!

Questions or collaborations welcome.

📧 [Lshpaner@mednet.ucla.edu; afunnell@mednet.ucla.edu]

Thank you to:
Panayiotis Petousis, Al Rahooh,
Colin Beam, Dr. Alex Bui

UCLA Clinical and
Translational Science Institute
(NIH/NCATS UL1 TR001881) grant

[https://github.com/uclamii/equi_boots]