Oxygen-Hemoglobin Dissociation: Physiologic Principles to Bedside Care

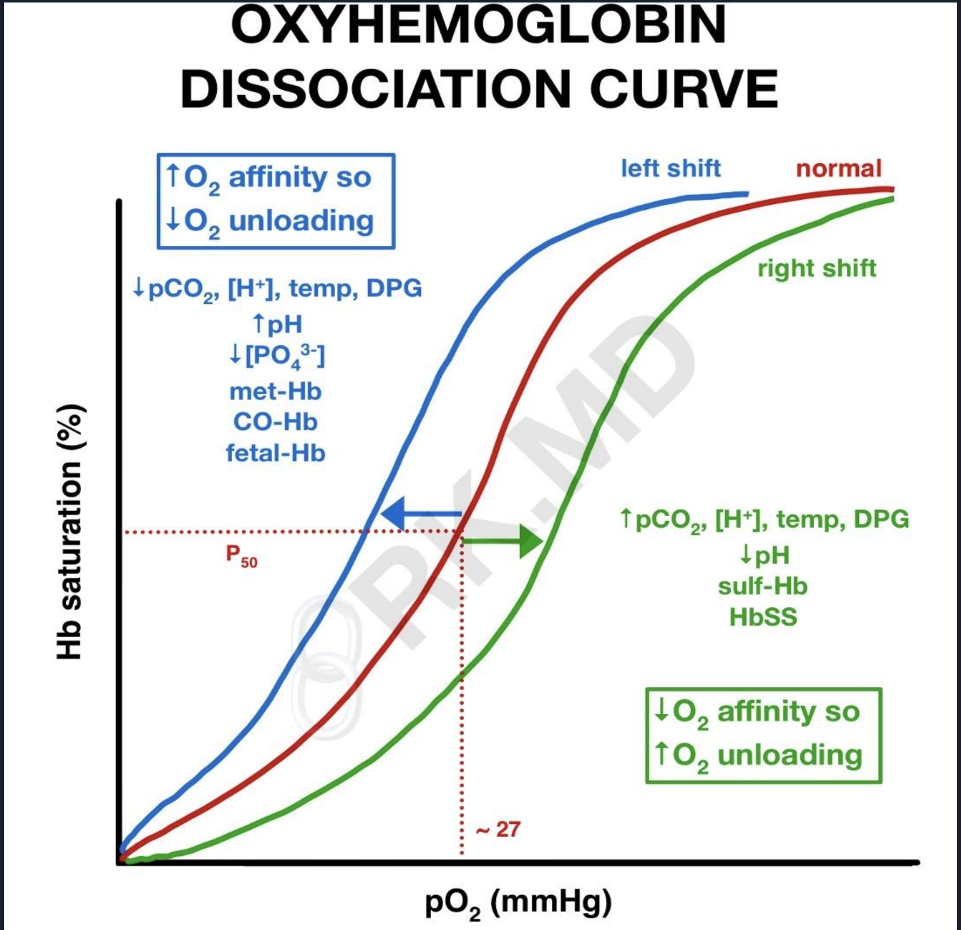
By team 'SIGMOID':

Jonathon Schwartz MD, Sumit Kapoor MD, Vinai M. Modem MD

Donghao Li, Aya El Mir, Fredrik Willumsen Haug, Susannah Oster

What are our Research Questions?

1. How does the oxygen-hemoglobin dissociation curve shifts for variables like changes in pH, temperature, pCo2, lactate, HCO3 etc?
2. Do we see any association of oxygen-hemoglobin dissociation curve shifts with factors like gender, race/ethnicity, socio-economic status , mortality etc?
3. Create an unsupervised machine learning clustering algorithm for p50 based on various variables.

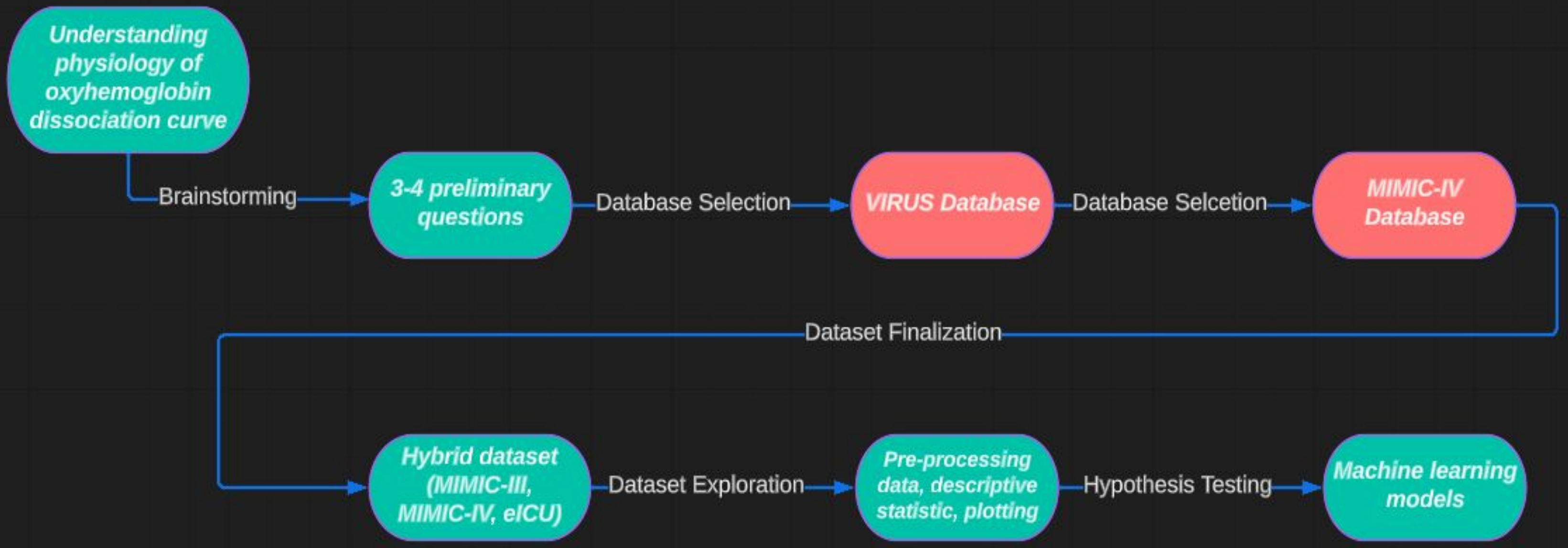


$$p50 = ((100x((pO2)^3)/SpO2)-(pO2)^3)^{1/3}$$

P50 Categories

- <22 mmHg- leftward shift of Oxy-Hb dissociation curve
- 22-30 mmHg- normal Oxy-Hb dissociation curve
- >30 mmHg- rightward shift of Oxy-Hb dissociation curve

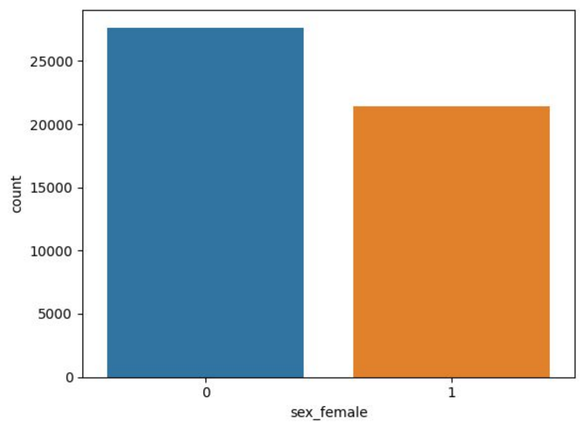
Project Roadmap



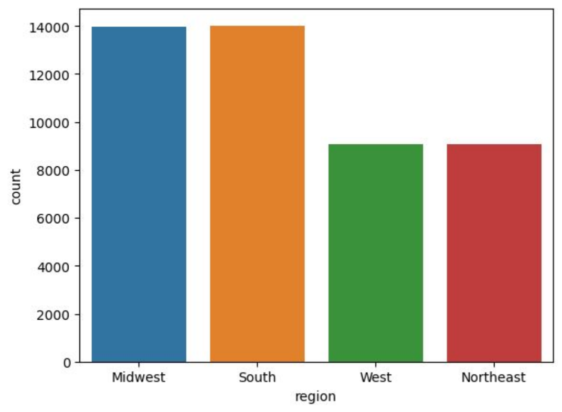
Summary Statistics

Variable	Mean (SD)
Age (years)	64.42 (15.8)
Admission Weight (kg)	85.6 (27.8)
ICU LOS (days)	5.2 (7.1)
Hospital LOS (days)	11.5 (13.6)

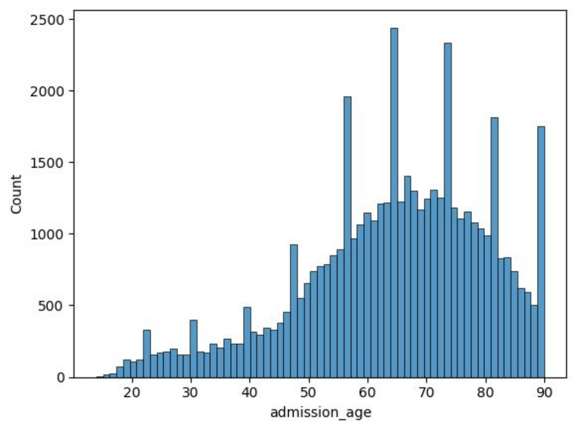
Variable	Count
Gender	Male-27660 Female- 21433
Teaching Status	Non-Teaching-30270 Teaching- 18823
Race/Ethnicity	White- 37380 NonWhite-11713
Region	South- 14018 Midwest- 13979 West- 9062 Northeast- 9052
Number of Beds	>=500- 25986 250-499-10112 100-249-7516 <100- 1554



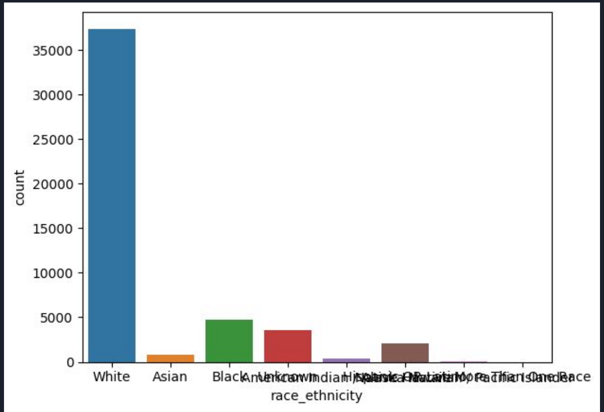
Gender distribution



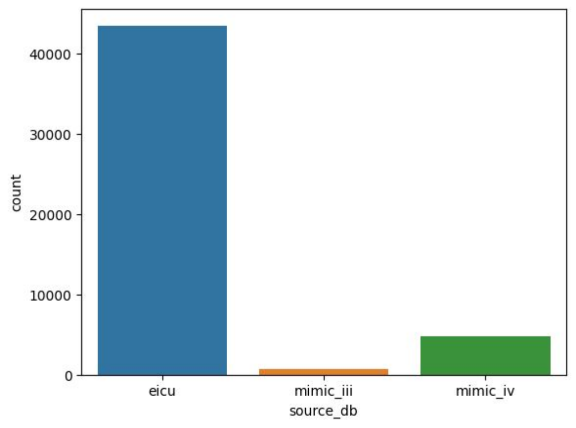
Region distribution



Age distribution

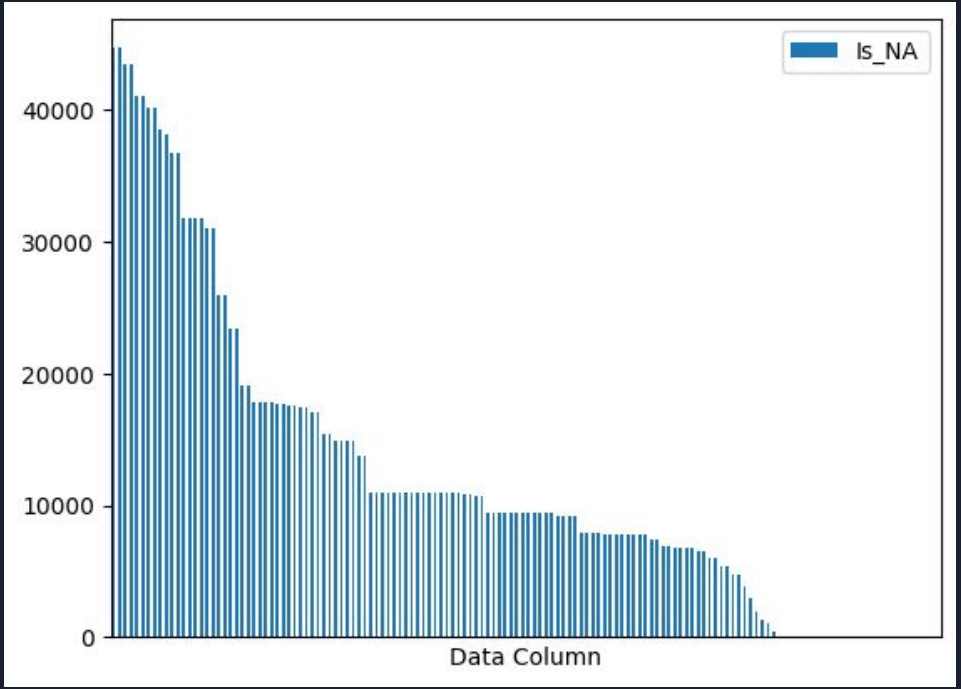


R Race/Ethnicity distribution



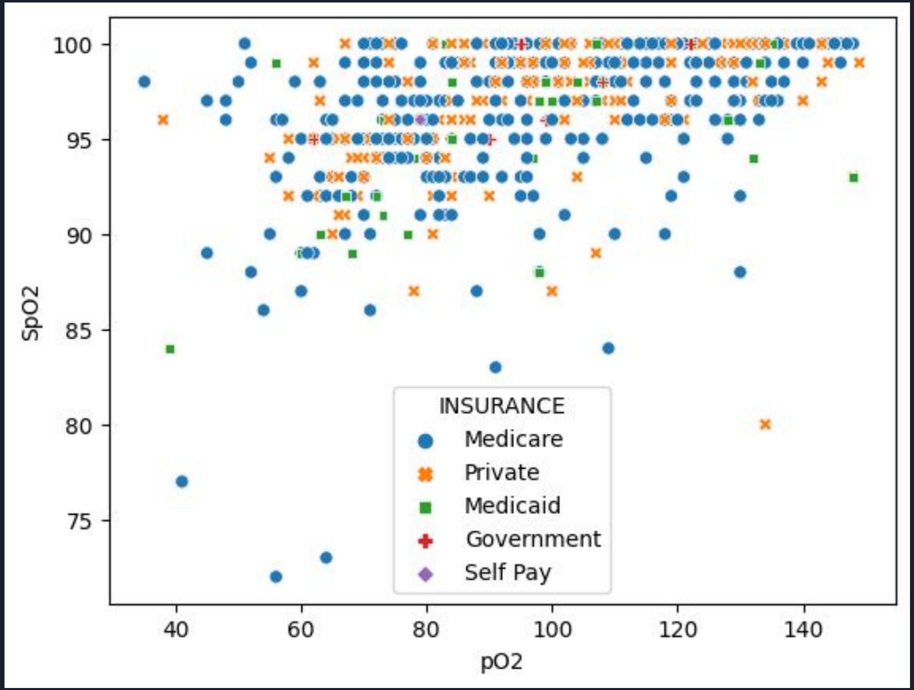
Source of data

Count of Missing Values in Data Columns in Raw Dataset

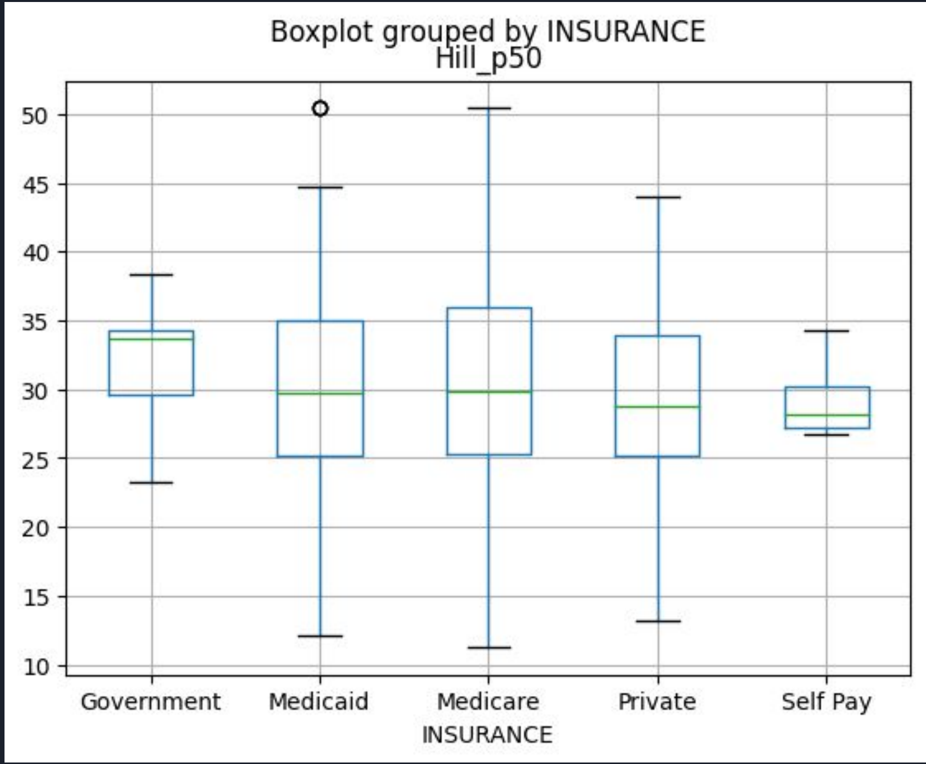


source_db	0
numbedscategory	3925
teachingstatus	0
region	2975
admission_age	2
sex_female	0
weight_admission	1104
height_admission	1284
BMI_admission	1985
los_hospital	0
los_ICU	0
comorbidity_score_name	0
comorbidity_score_value	2
in_hospital_mortality	404
race_ethnicity	0
pH	0
pCO2	0
pO2	0
SaO2	0
SpO2	0
Carboxyhemoglobin	38133
Methemoglobin	38517
vitals_heart_rate	5450
vitals_resp_rate	6876
vitals_sbp_ni	14952
vitals_dbp_ni	14956

SpO2-pO2 Relationship by Insurance Status



p50 vs. Insurance Status in MIMIC III subset

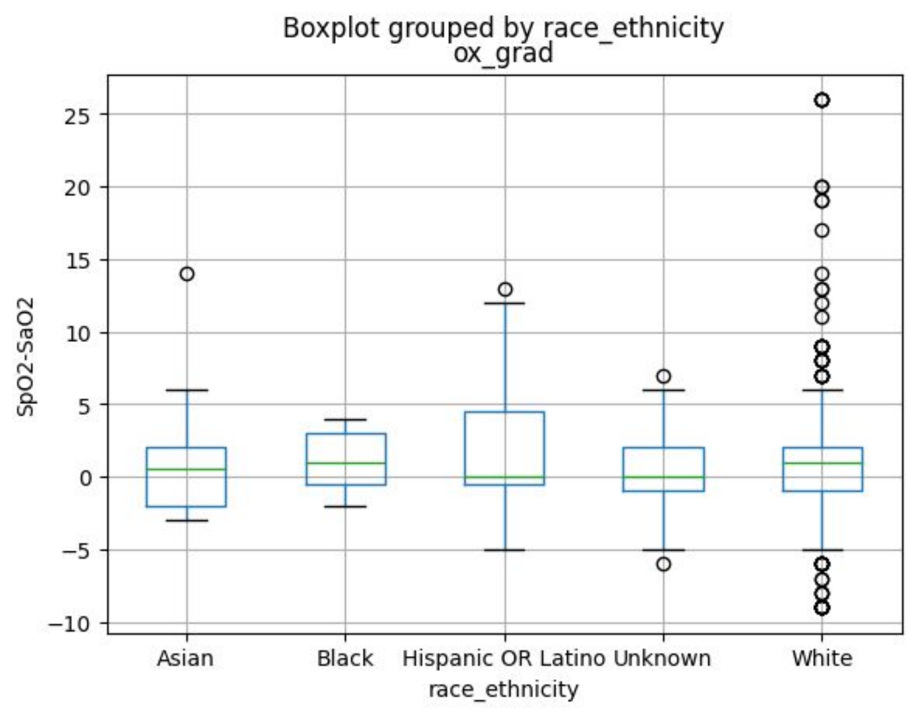


Example Analysis via Pandas

Mann-Whitney Medicaid vs. Private p50:

pvalue=0.32317439805614645

Association between SpO2-SaO2 discrepancy and race_ethnicity in MIMIC III subset of Dataset

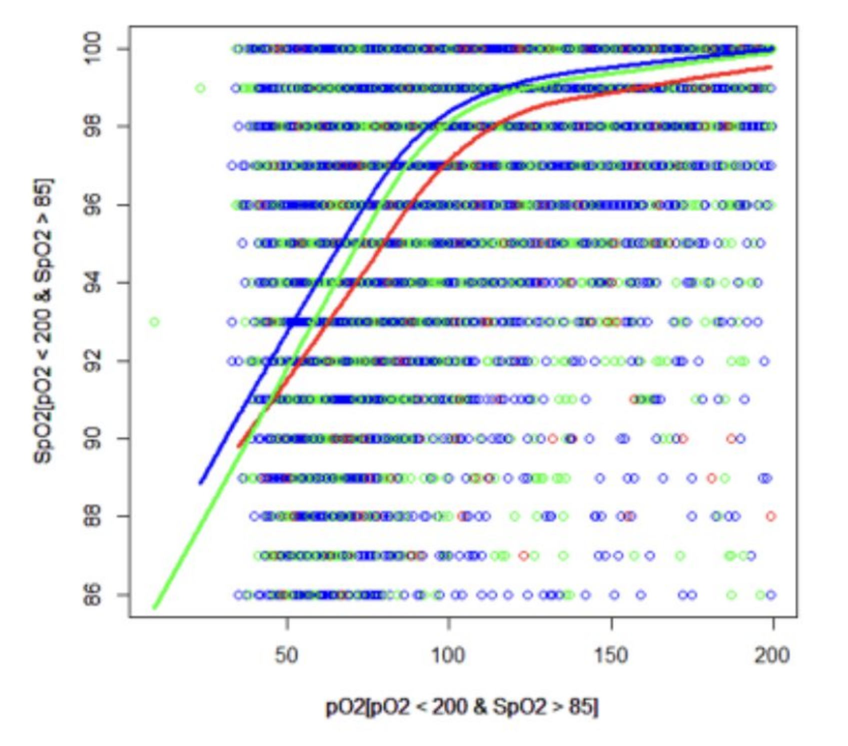


Example Analysis via Pandas

Mann-Whitney Black vs. Non-black population:

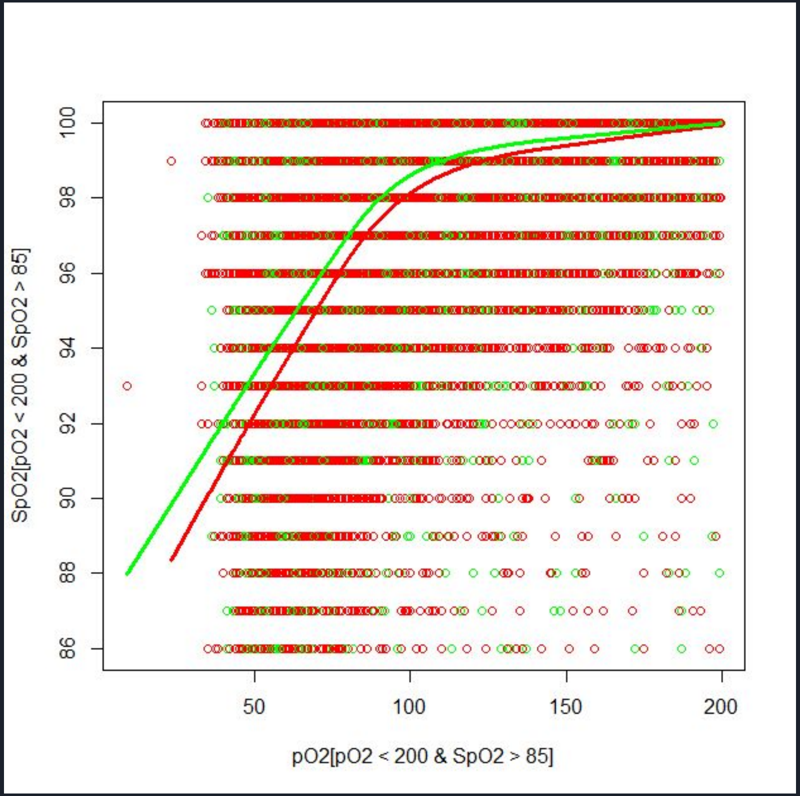
p-value=0.03790193327049529)

OxyHemoglobin dissociation curve with pH categories



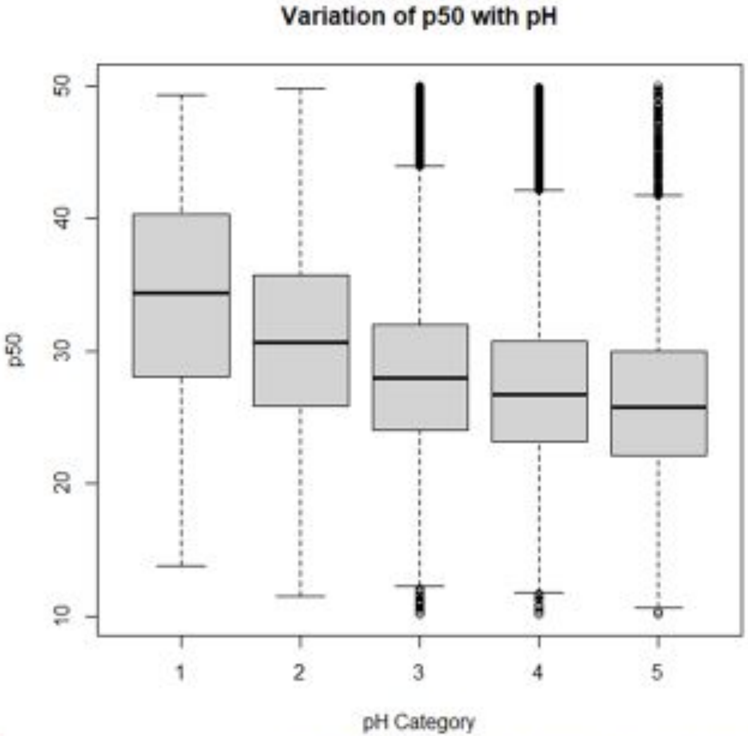
<u>pH Groups</u>
Red - pH < 7.2
Green - pH 7.2-7.35
Blue - pH >7.35

Oxyhemoglobin dissociation curve with Race (white versus non white)



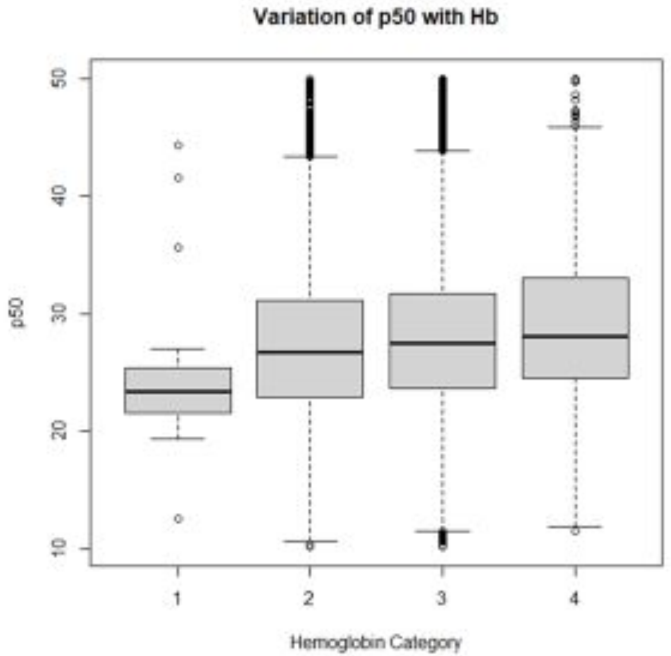
<u>Race</u>
Red - White
Green - Non-White

p50 variability with pH Groups



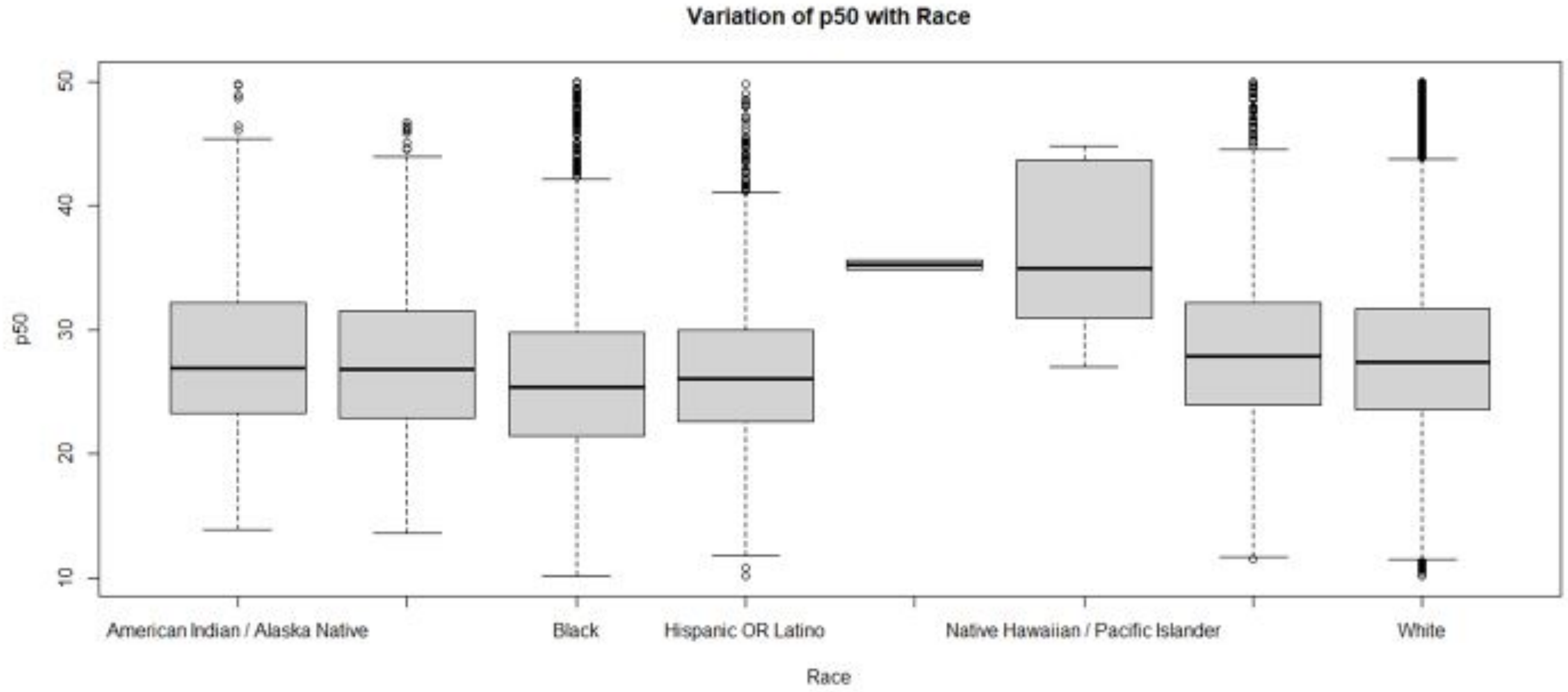
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
pHGroups[valid == 1]	1	31229	31229	685	<2e-16	***
Residuals	28749	1310637	46			

p50 variability with Hemoglobin




```
                     Df  Sum Sq Mean Sq F value Pr(>F)
HbGr[valid == 1]      1    3792    3792   81.33 <2e-16 ***
Residuals         24715 1152290      47
---
```

p50 variability with Race



Association between p50 and Mortality

	Non-Survivors	Survivors
Right-shifted oxygen dissociation curve (p50>30)	1876 (20.5%)	7257
Normal oxygen dissociation curve (p50≤30)	3115 (16%)	16275

p-value < 0.001, Odds Ratio 1.35

ML Process

- Data Preprocessing
  - Handling missing values
    - Unnecessary rows and columns
  - Variable Encoding
  - Computing p50
  - Group exclusion based on non physiological values

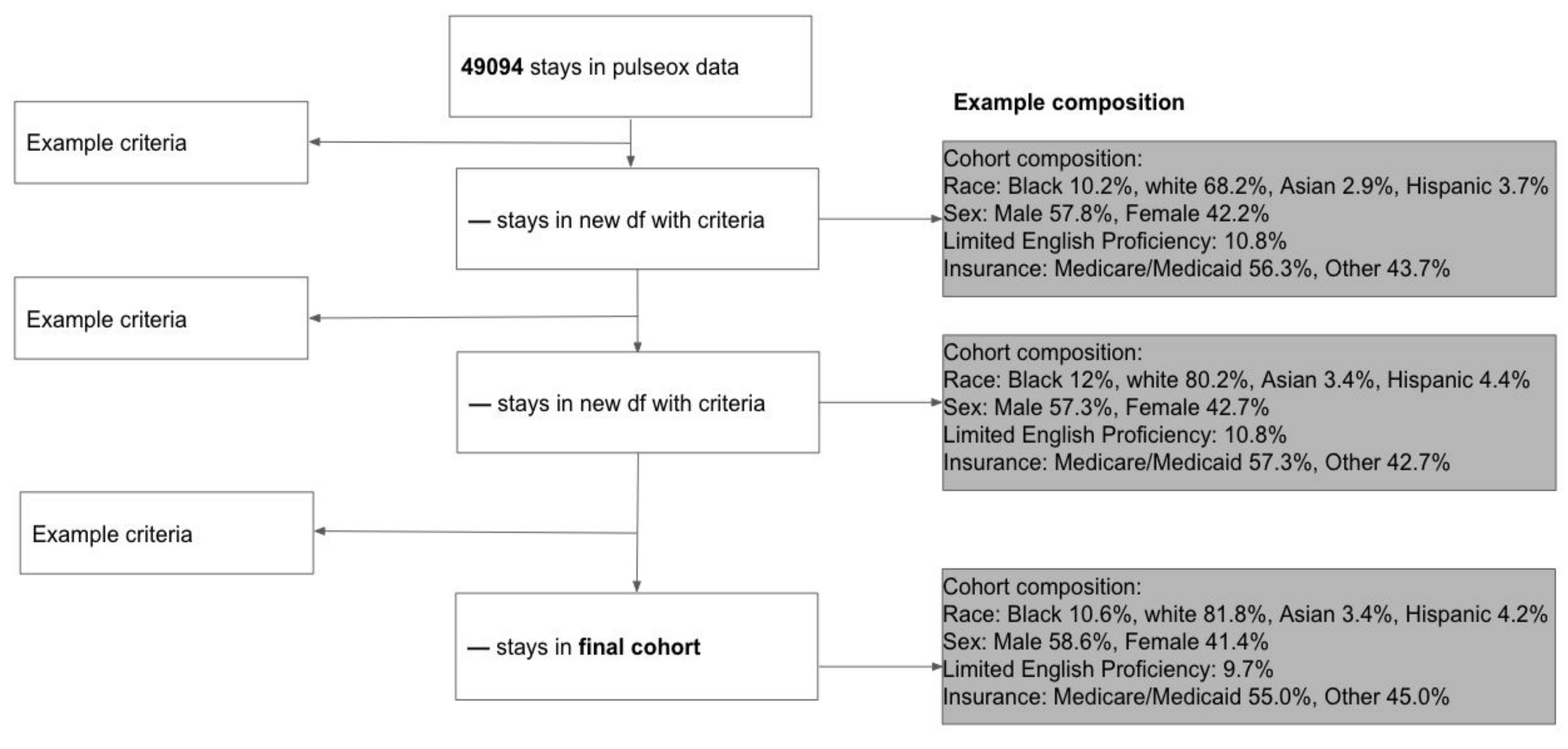
(49093, 142) -> (23036, 67)

```
Unique values in source_db:
['eicu' 'mimic_iii' 'mimic_iv']

Unique values in numbedscategory:
['100 - 249' '>= 500' '<100' '250 - 499' nan]

Unique values in teachingstatus:
[False  True]

Unique values in region:
['Midwest' nan 'South' 'West' 'Northeast']

Unique values in comorbidity_score_name:
['Charlson' 'Elixhauser']

Unique values in race_ethnicity:
['White' 'Asian' 'Unknown' 'American Indian / Alaska Native'
 'Hispanic OR Latino' 'Black' 'Native Hawaiian / Pacific Islander'
 'More Than One Race']

Unique values in SaO2_timestamp:
['2014-01-01 05:14:00' '2014-01-01 00:13:00' '2014-01-03 20:53:00' ...
 '2123-12-30 10:03:00+00:00' '2126-05-06 08:03:00+00:00'
 '2128-03-02 16:01:00+00:00']

Unique values in SpO2_timestamp:
['2014-01-01 05:14:00' '2014-01-01 00:13:00' '2014-01-03 20:50:00' ...
 '2123-12-30 10:00:00+00:00' '2126-05-06 08:00:00+00:00'
 '2128-03-02 16:00:00+00:00']
```

	ColumnName	MissingPercentage
25	Carboxyhemoglobin	77.675025
26	Methemoglobin	78.457214
33	vitals_mbp_i	63.306378
34	vitals_sbp_i	64.685393
35	vitals_dbp_i	64.701689
46	coag_fibrinogen	83.698287
59	bmp_lactate	52.905710
64	hfp_bilirubin_direct	88.458640
66	others_ck_cpk	74.912920
67	others_ck_mb	81.848736
68	others_ld_ldh	91.012975

```
Column names: ['unique_subject_id', 'unique_hospital_admission_id', 'unique_icustay_id', 'subject_id', 'hospital_admi
ssion_id', 'icustay_id', 'source_db', 'hospitalid', 'numbedscategory', 'teachingstatus', 'region', 'admission_age',
'sex_female', 'weight_admission', 'height_admission', 'BMI_admission', 'datetime_hospital_admit', 'datetime_hospital_
discharge', 'datetime_icu_admit', 'datetime_icu_discharge', 'los_hospital', 'los_ICU', 'comorbidity_score_name', 'com
orbidity_score_value', 'in_hospital_mortality', 'race_ethnicity', 'SaO2_timestamp', 'pH', 'pCO2', 'pO2', 'SaO2', 'SpO
2', 'Carboxyhemoglobin', 'Methemoglobin', 'SpO2_timestamp', 'delta_SpO2', 'delta_vitals_heart_rate', 'vitals_heart_ra
te', 'delta_vitals_resp_rate', 'vitals_resp_rate', 'delta_vitals_mbp_ni', 'vitals_mbp_ni', 'delta_vitals_sbp_ni', 'vi
tals_sbp_ni', 'delta_vitals_dbp_ni', 'vitals_dbp_ni', 'delta_vitals_mbp_i', 'vitals_mbp_i', 'delta_vitals_sbp_i', 'vi
tals_sbp_i', 'delta_vitals_dbp_i', 'vitals_dbp_i', 'delta_vitals_tempc', 'vitals_tempc', 'delta_cbc_hemoglobin', 'cbc
_hemoglobin', 'delta_cbc_hematocrit', 'cbc_hematocrit', 'delta_cbc_mch', 'cbc_mch', 'delta_cbc_mchc', 'cbc_mchc', 'de
lta_cbc_mcv', 'cbc_mcv', 'delta_cbc_platelet', 'cbc_platelet', 'delta_cbc_rbc', 'cbc_rbc', 'delta_cbc_rdw', 'cbc_rd
w', 'delta_cbc_wbc', 'cbc_wbc', 'delta_coag_fibrinogen', 'coag_fibrinogen', 'delta_coag_inr', 'coag_inr', 'delta_coag
_pt', 'coag_pt', 'delta_coag_ptt', 'coag_ptt', 'delta_bmp_sodium', 'bmp_sodium', 'delta_bmp_potassium', 'bmp_potassiu
m', 'delta_bmp_chloride', 'bmp_chloride', 'delta_bmp_bicarbonate', 'bmp_bicarbonate', 'delta_bmp_bun', 'bmp_bun', 'de
lta_bmp_creatinine', 'bmp_creatinine', 'delta_bmp_glucose', 'bmp_glucose', 'delta_bmp_aniongap', 'bmp_aniongap', 'del
ta_bmp_calcium', 'bmp_calcium', 'delta_bmp_lactate', 'bmp_lactate', 'delta_hfp_alt', 'hfp_alt', 'delta_hfp_alp', 'hfp
_alp', 'delta_hfp_ast', 'hfp_ast', 'delta_hfp_bilirubin_total', 'hfp_bilirubin_total', 'delta_hfp_bilirubin_direct',
'hfp_bilirubin_direct', 'delta_hfp_albumin', 'hfp_albumin', 'delta_others_ck_cpk', 'others_ck_cpk', 'delta_others_ck_
mb', 'others_ck_mb', 'delta_others_ld_ldh', 'others_ld_ldh', 'delta_sofa_past_overall_24hr', 'sofa_past_overall_24h
r', 'delta_sofa_past_coagulation_24hr', 'sofa_past_coagulation_24hr', 'delta_sofa_past_liver_24hr', 'sofa_past_liver_
24hr', 'delta_sofa_past_cardiovascular_24hr', 'sofa_past_cardiovascular_24hr', 'delta_sofa_past_cns_24hr', 'sofa_past
_cns_24hr', 'delta_sofa_past_renal_24hr', 'sofa_past_renal_24hr', 'delta_sofa_future_overall_24hr', 'sofa_future_over
all_24hr', 'delta_sofa_future_coagulation_24hr', 'sofa_future_coagulation_24hr', 'delta_sofa_future_liver_24hr', 'sof
a_future_liver_24hr', 'delta_sofa_future_cardiovascular_24hr', 'sofa_future_cardiovascular_24hr', 'delta_sofa_future_
cns_24hr', 'sofa_future_cns_24hr', 'delta_sofa_future_renal_24hr', 'sofa_future_renal_24hr']
```

Exploratory Data Analysis



PhysioNet Find Share About News Projects fredwh Search

Challenge Credentialed Access

MIT Critical Datathon 2023: a MIMIC-IV Derived Dataset for Pulse Oximetry Correction Models

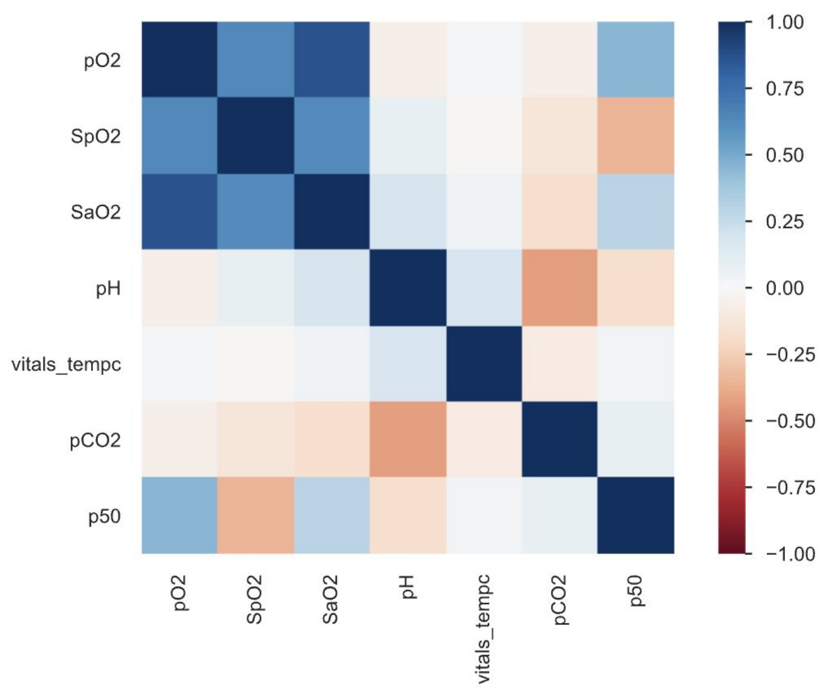
João Matos, Tristan Struja, David S Restrepo, Luis Filipe Nakayama, Jack Gallifant, Luca Weishaupt, Nikita Mullangi, Maria Loureiro, Skyler Shapiro, Adrien Carrel, Leo Anthony Celi

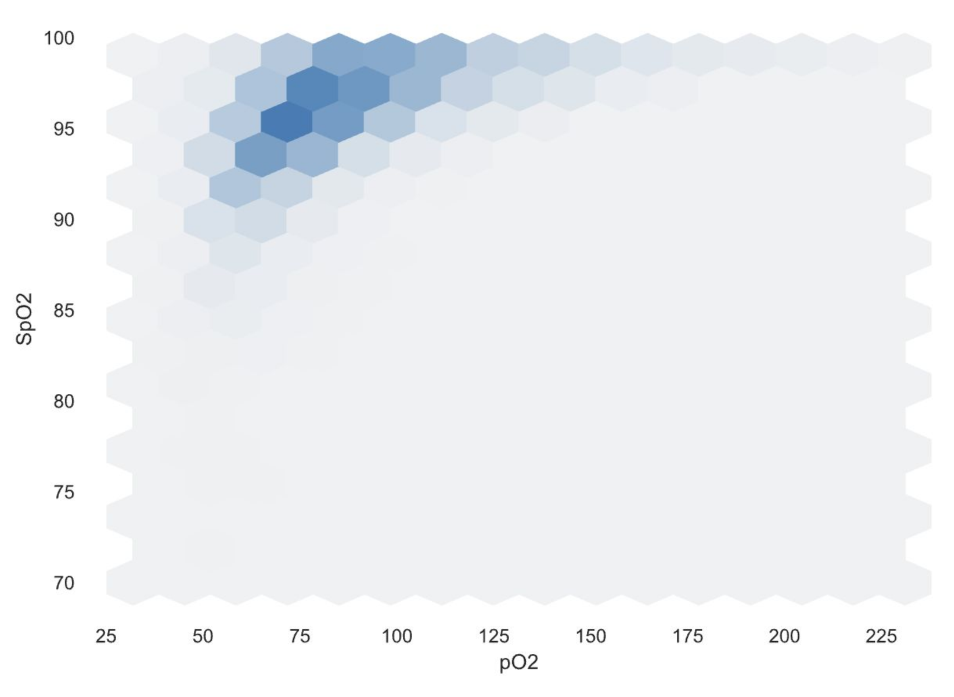
Published: May 8, 2023. Version: 1.0.0

SQL code: https://github.com/CriticalDatathon/data-prep/tree/main/MIMIC-IV

Dataset Distribution

Features of Interest





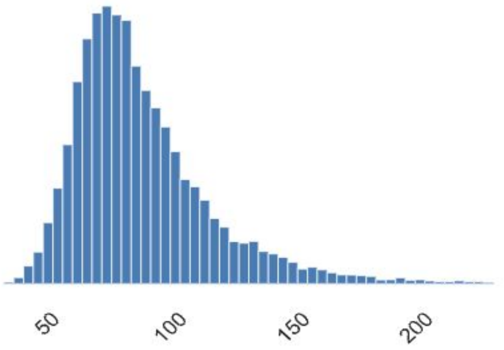
pandas-profiling 3.6.6

```
pip install pandas-profiling
```

pO2

Real number (ℝ)

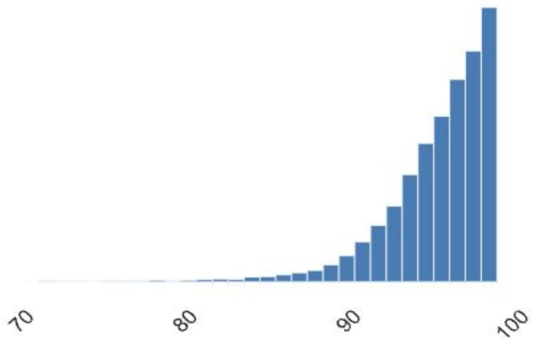
Distinct	962	**Minimum**	32
Distinct (%)	4.2%	**Maximum**	231
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	87.838618	**Memory size**	180.1 KiB



SpO2

Real number (ℝ)

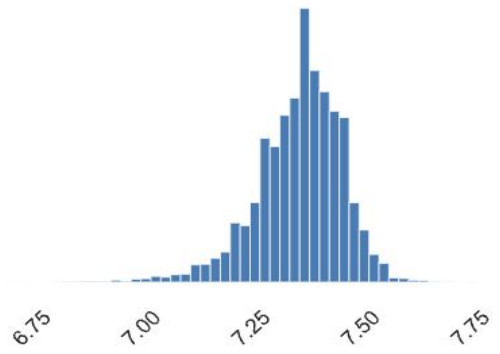
Distinct	30	**Minimum**	70
Distinct (%)	0.1%	**Maximum**	99
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	95.71423	**Memory size**	180.1 KiB



pH

Real number (ℝ)

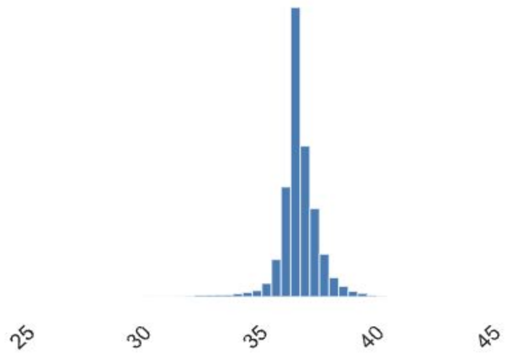
Distinct	622	**Minimum**	6.68
Distinct (%)	2.7%	**Maximum**	7.81
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	7.3579062	**Memory size**	180.1 KiB



vitals_tempc

Real number (ℝ)

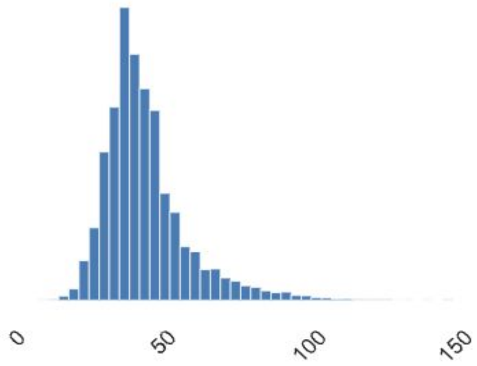
Distinct	413	**Minimum**	23.7
Distinct (%)	1.8%	**Maximum**	44.3
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	36.84159	**Memory size**	180.1 KiB



pCO2
Real number (ℝ)

Distinct	857
Distinct (%)	3.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	43.499687

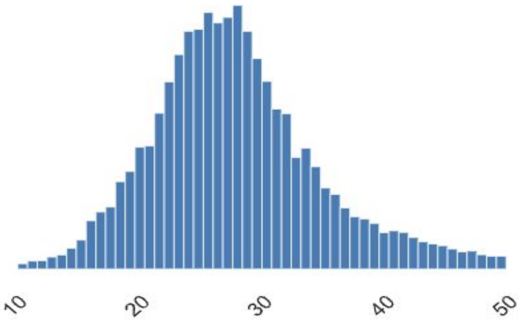
Minimum	0
Maximum	173
Zeros	1
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	180.1 KiB



p50
Real number (ℝ)

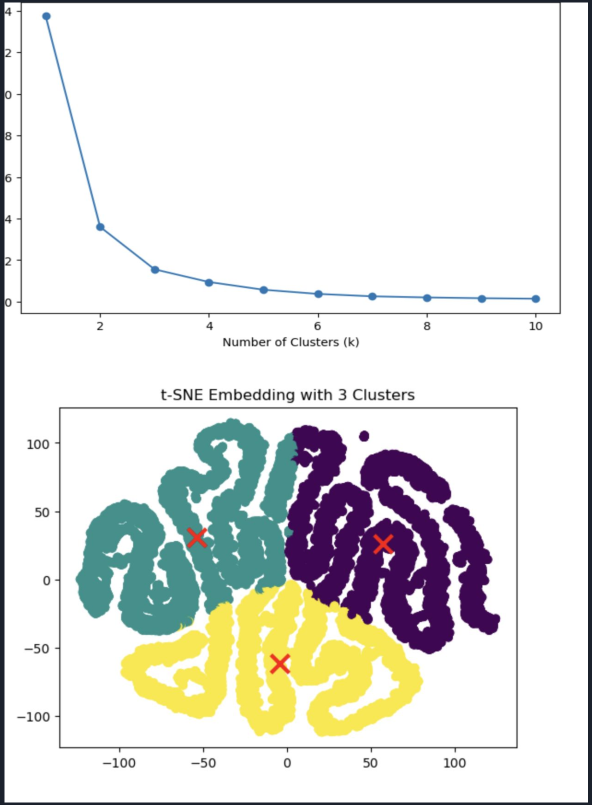
Distinct	3865
Distinct (%)	16.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	27.948443

Minimum	10.159823
Maximum	49.991348
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	180.1 KiB

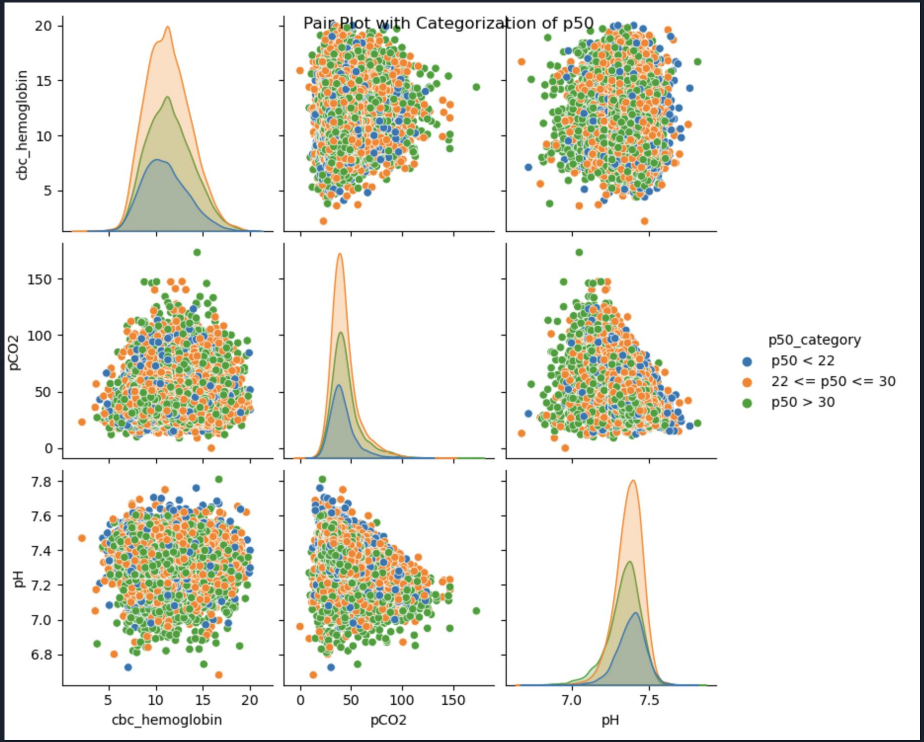


Clustering

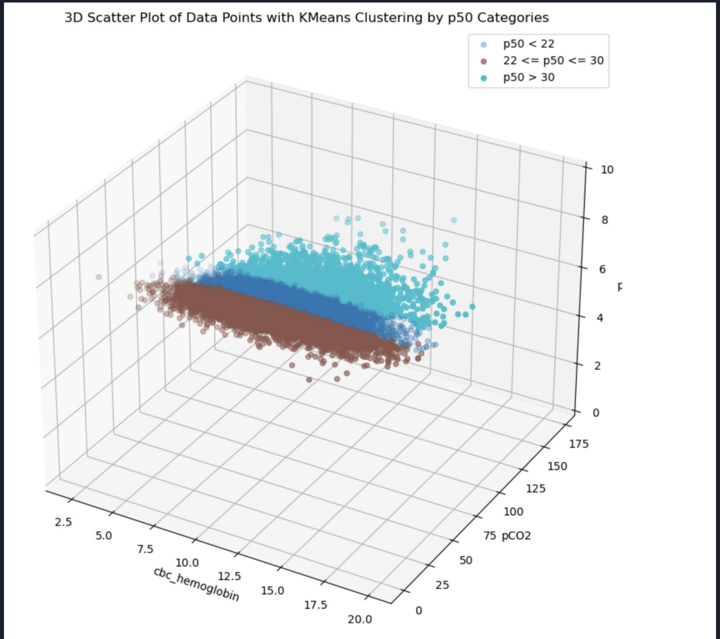
- Choosing the number of Clusters (Elbow)
- Different unsupervised ML models
    - K means
    - DBSCAN
    - tSNE

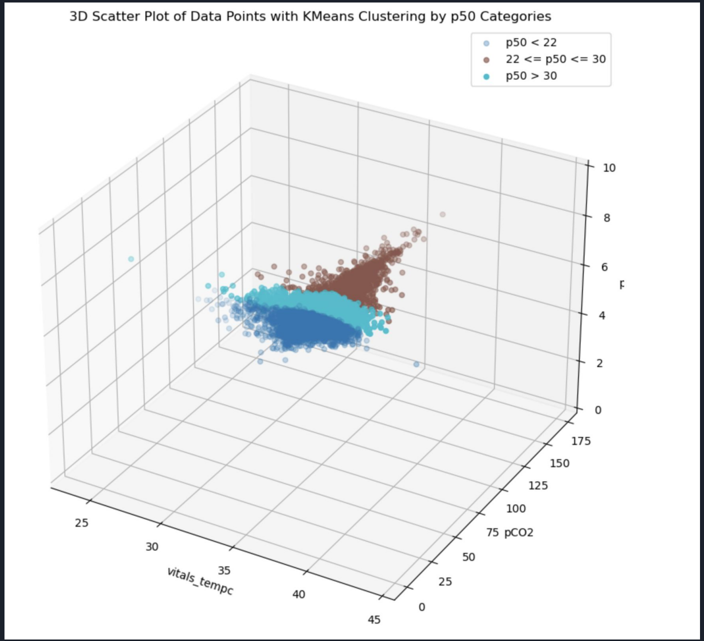


K means: Pairplot of p50 categories with variables (2D)

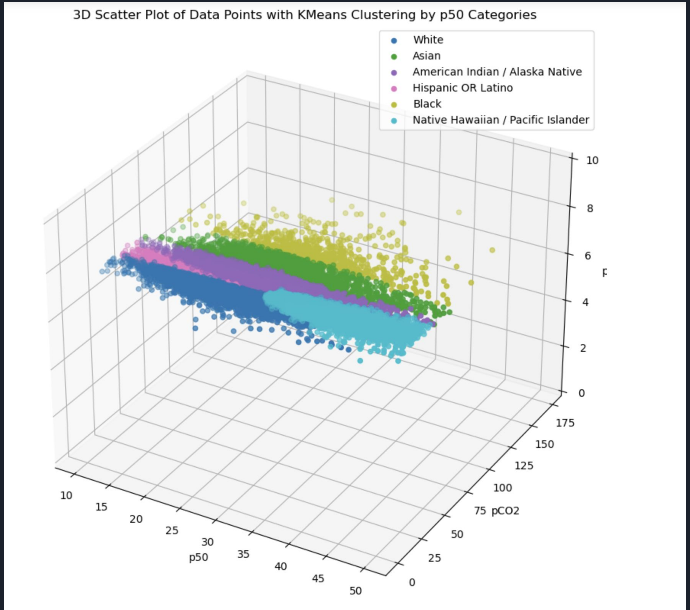


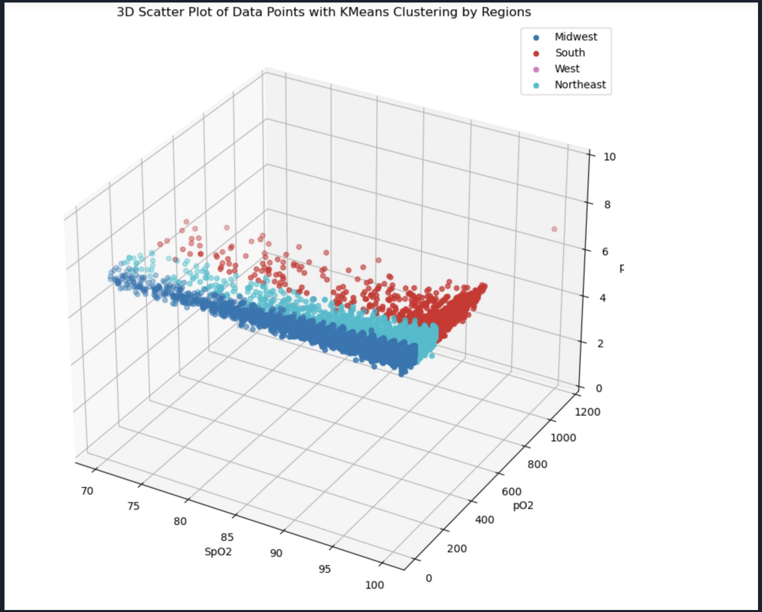
K means: Pairplot of p50 categories with variables (3D)





K-means clustering can assist in discovering race and geography related patterns in data





Conclusions

- Oxy-Hb dissociation curve worked well with changes in pH, HC03, lactate levels
- Non-White race had higher Sp02 levels compared to white race
- Rightward shift in Oxy-Hb dissociation curve is associated with higher mortality
- No differences in Oxy-Hb curves based on age categories, gender, insurance type
- P50 represents shift of Oxy-Hb dissociation curve
- Higher p50 (RIGHTWARD) shifts associated with increased mortality

Challenges we encountered

- Formulating a relevant clinical question.
- Understanding the strengths and limitations of Data Scientists and Clinicians and working as one team
- Finalizing the dataset to work on
- P50 as the outcome variable and understanding it from clinical standpoint
- Preprocessing data/ finalizing the relevant machine learning algorithm

Future Steps

- We will study relationship of oxyhemoglobin dissociation curve with changes in pH, pCO_2, lactate, SOFA scores, age etc.
- We will investigate any disparities based on gender, race, socio-economic status, age.
- Inspired by Dr. Celi's talk this morning we are curious to see what will happen if we run an unsupervised clustering model to look for unexplainable race classification based on clinical conditions alone.
- Incorporate MIMIC IV data for subgroup analysis on Insurance
- Create a prediction algorithm based on K-means clustering prediction

Limitations

- Class Imbalance and Confounding
    - Multiple admissions per patient → overrepresented patients bias data with hidden confounders
- Inconsistent Insurance Data across source Databases
- General data entry errors (e.g. incorrect gender entry for a subject, out of range values for variables)
- Temporal analysis for data quality
    - e.g. general trends in SpO2 and PaO2,
    - May reflect changes in instrument calibration)

Thank You!

We’re would appreciate questions, constructive feedback, and further discussion!

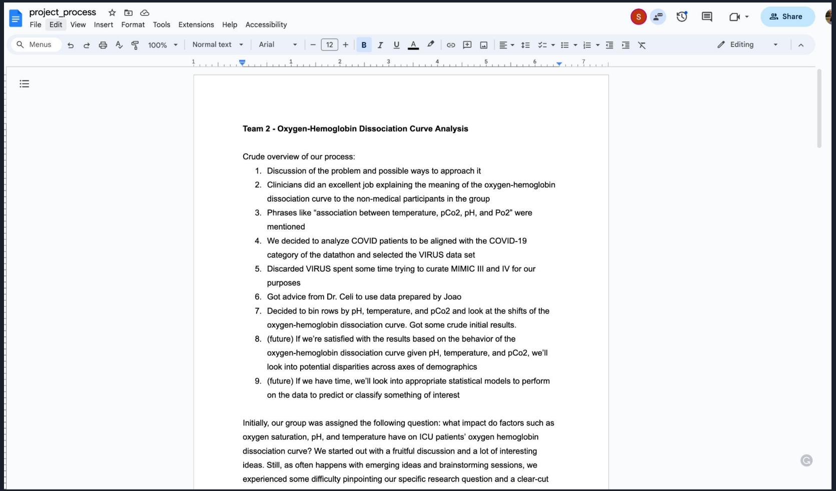
The Simoid Team

Code

Code: https://github.com/SCCMdatathon2023/team_02

Dataset used in our analysis

- Hybrid dataset comprised on MIMIC-III, MIMIC-IV and eICU from MIT Critical Data Consortium
- 109 columns, 49093 rows



Team 2 - Oxygen-Hemoglobin Dissociation Curve Analysis

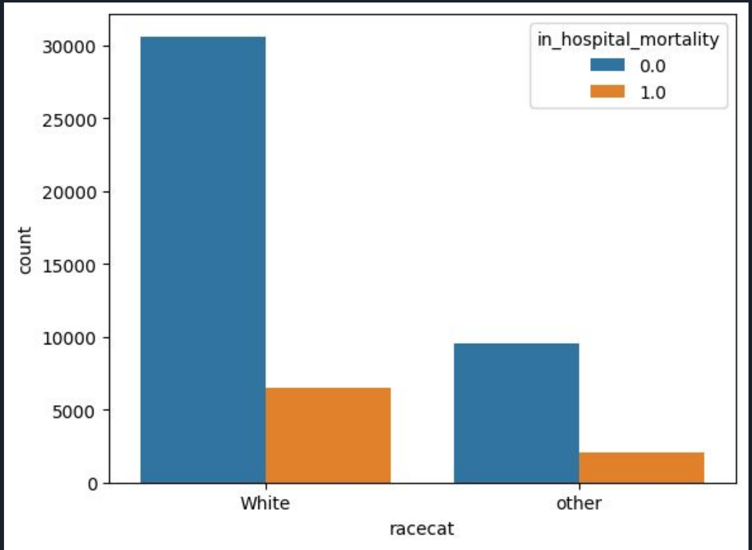
Crude overview of our process:

1. Discussion of the problem and possible ways to approach it
2. Clinicians did an excellent job explaining the meaning of the oxygen-hemoglobin dissociation curve to the non-medical participants in the group
3. Phrases like "association between temperature, pCo2, pH, and Po2" were mentioned
4. We decided to analyze COVID patients to be aligned with the COVID-19 category of the datathon and selected the VIRUS data set
5. Discarded VIRUS spent some time trying to curate MIMIC III and IV for our purposes
6. Got advice from Dr. Celi to use data prepared by Joao
7. Decided to bin rows by pH, temperature, and pCo2 and look at the shifts of the oxygen-hemoglobin dissociation curve. Got some crude initial results.
8. (future) If we're satisfied with the results based on the behavior of the oxygen-hemoglobin dissociation curve given pH, temperature, and pCo2, we'll look into potential disparities across axes of demographics
9. (future) If we have time, we'll look into appropriate statistical models to perform on the data to predict or classify something of interest

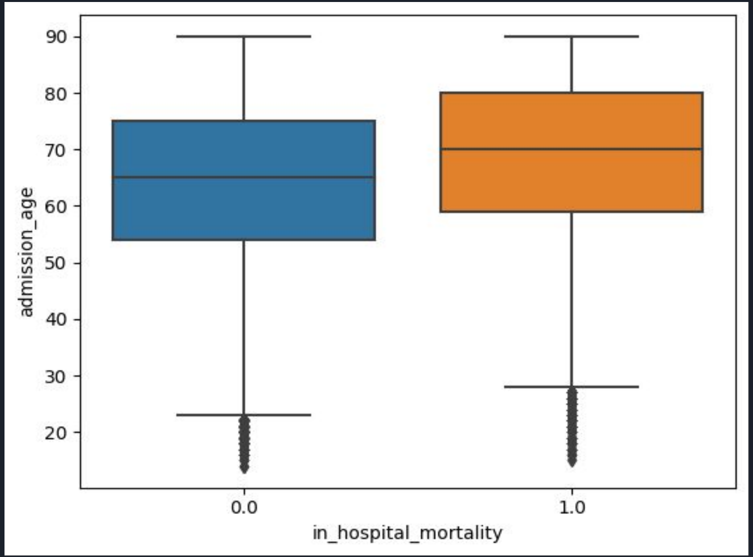
Initially, our group was assigned the following question: what impact do factors such as oxygen saturation, pH, and temperature have on ICU patients' oxygen hemoglobin dissociation curve? We started out with a fruitful discussion and a lot of interesting ideas. Still, as often happens with emerging ideas and brainstorming sessions, we experienced some difficulty pinpointing our specific research question and a clear-cut

In hospital Mortality versus…

Race/Ethnicity

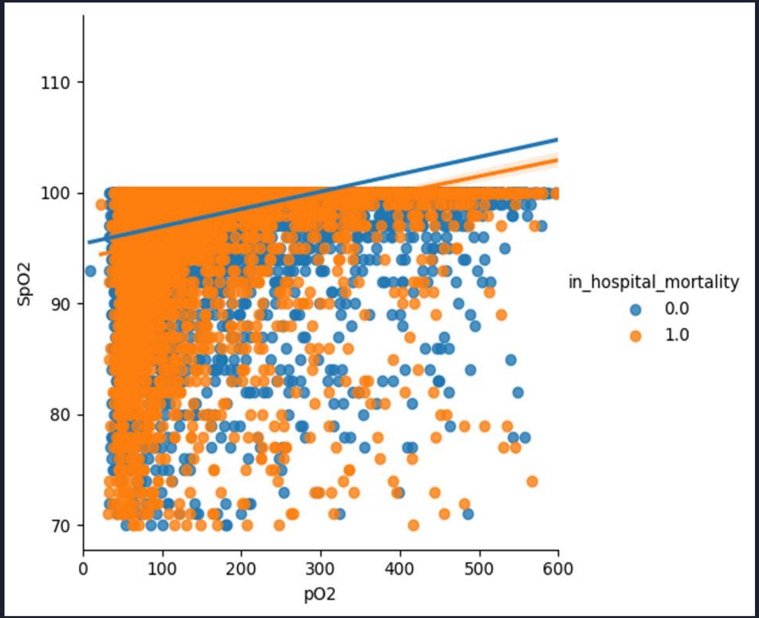


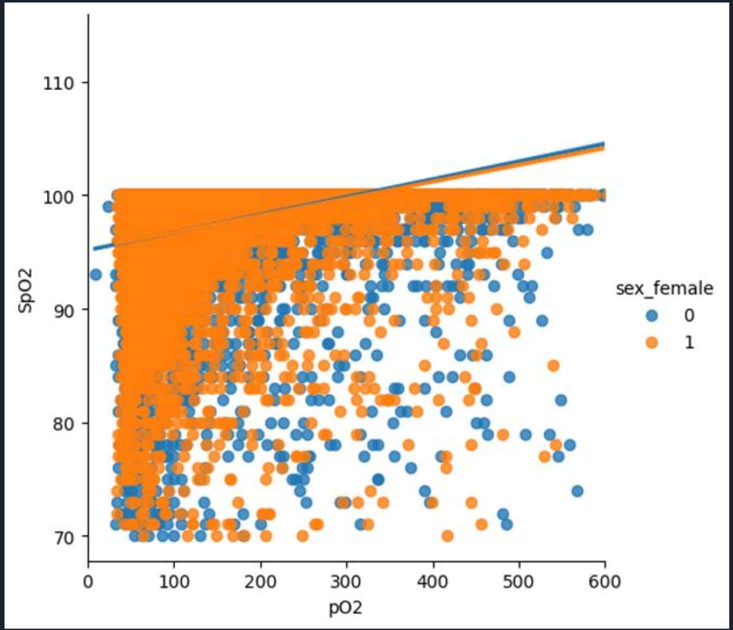
Age



Whites-21% mortality
Non-Whites- 21.5% mortality

Scatterplots of SpO2 versus pO2 by gender and mortality





Scatterplots of SpO2 versus pO2 with HCO3, lactate, Age, pCO2