

UNDARK SKIES

AMAYA LING, ARYA SRIVASTAVA, SMRITHI PANUGANTI

THE TEAM

ABOUT US

(From L to R)

- Smrithi
- Arya
- Amaya

We chose this project because we wanted to explore how light pollution and areas that do not have dark skies affect different species and other environmental factors so we can have a broader look at how light pollution affects our environment.

Expert: Professor Aparna Venkatesan from USF Physics



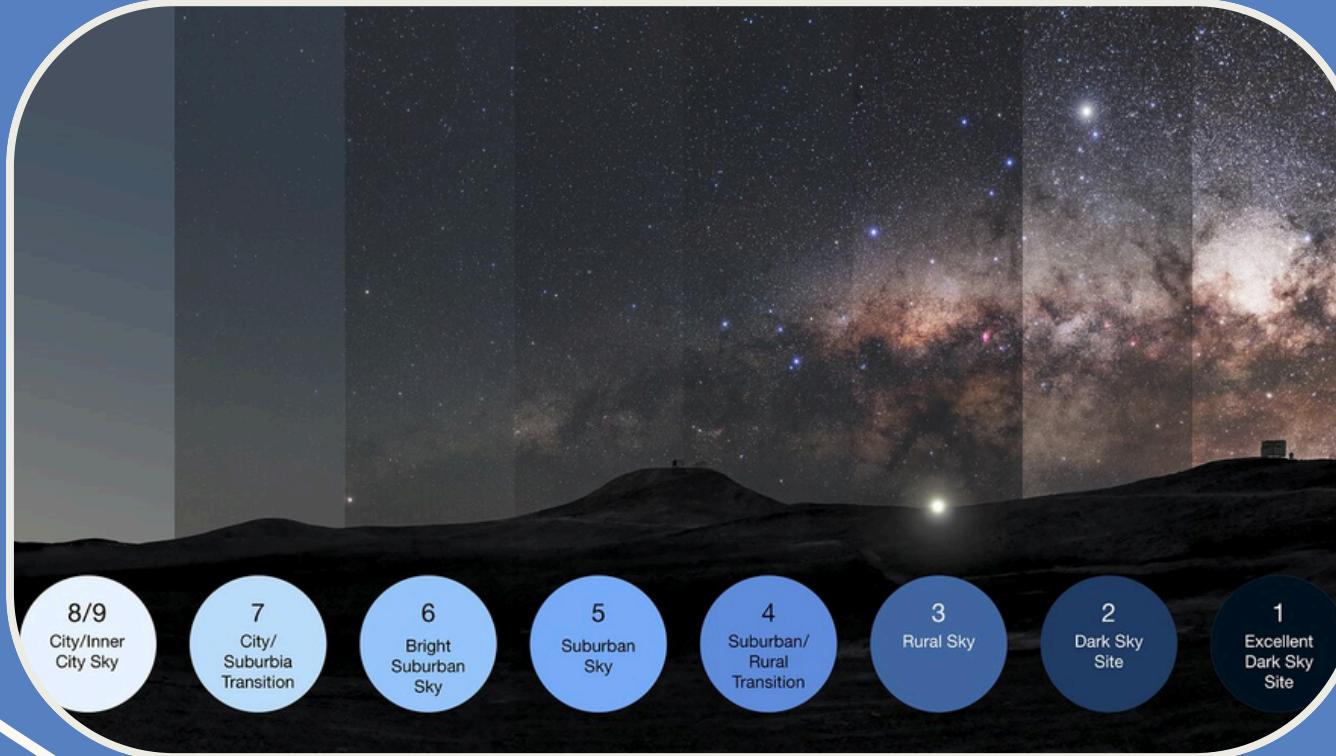
DESCRIPTION

- Reason for project
- Scope of project
 - Want to explore how ALAN affects different environmental factors



[From DarkSky.org](http://FromDarkSky.org)

BACKGROUND



From European Southern Observatory

① Coastal Habits

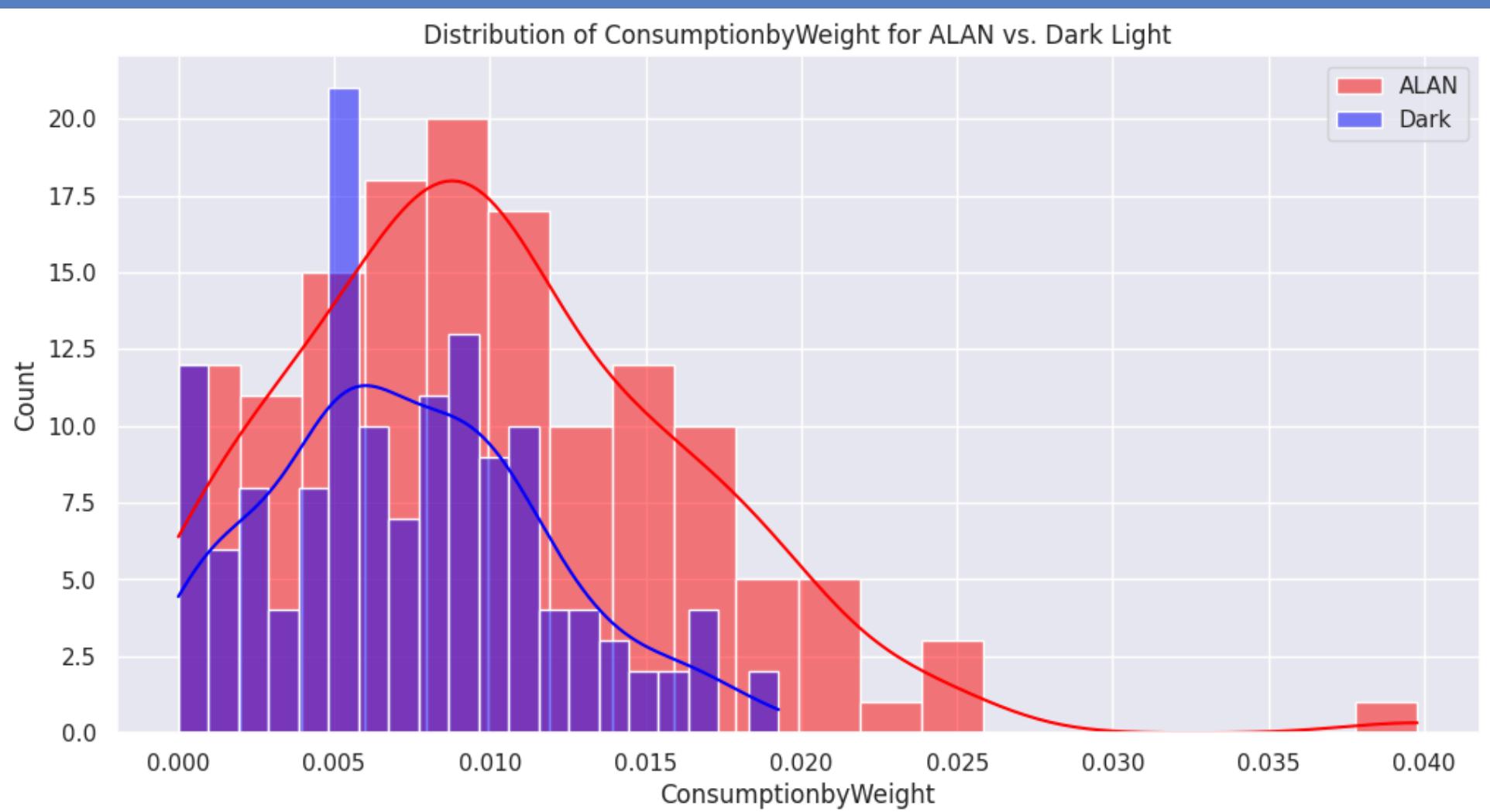
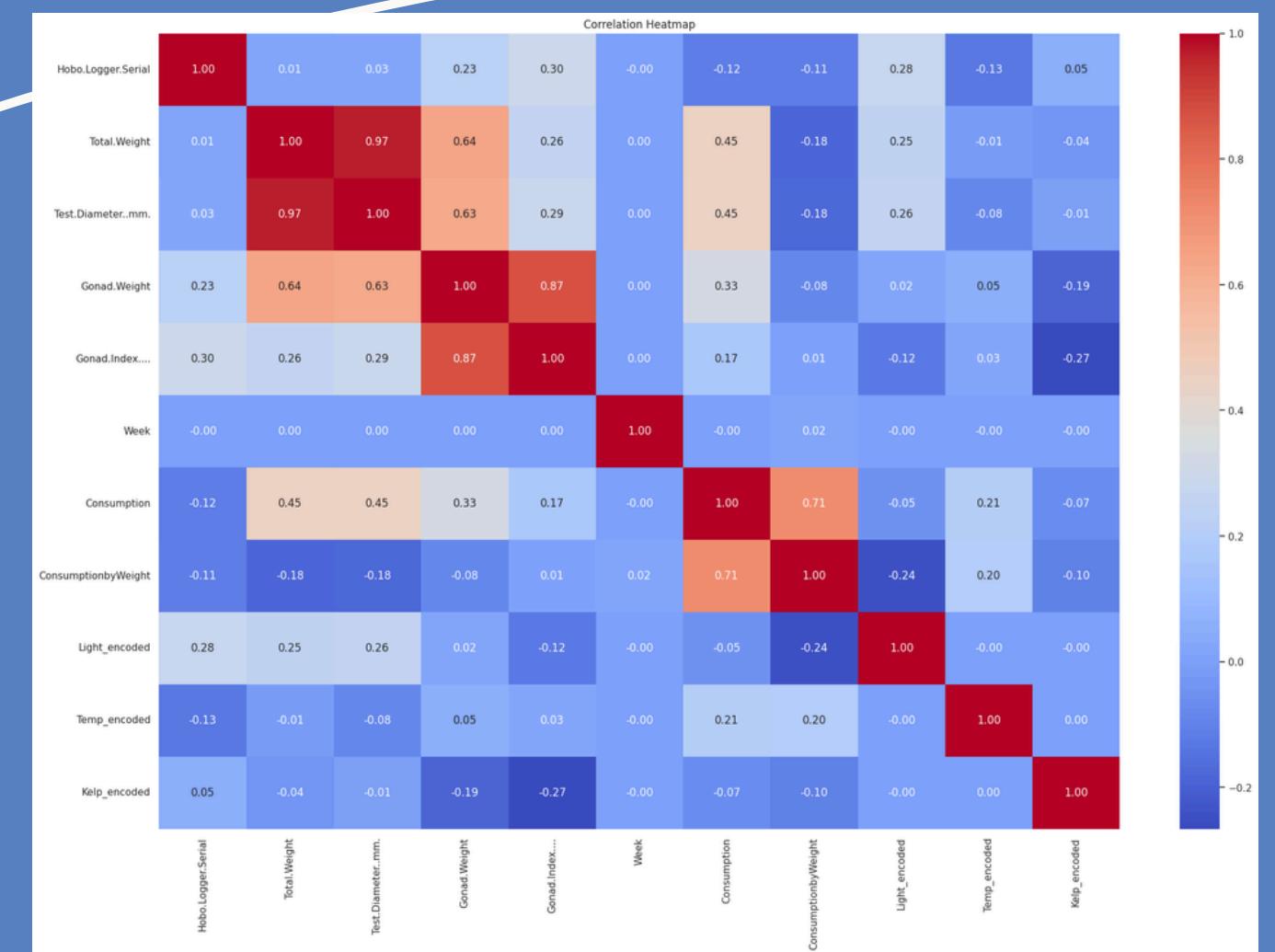
② Plant Diversity

③ Bird Migration Stop-overs

COASTAL HABITS

<https://datadryad.org/dataset/doi:10.5061/dryad.wpzgmsbt9>

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 280 entries, 0 to 279
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Tank             280 non-null    object  
 1   Species          280 non-null    object  
 2   Light            280 non-null    category
 3   Temp             280 non-null    category
 4   Kelp             280 non-null    category
 5   Hobo.Logger.Serial 195 non-null    float64
 6   Total.Weight     280 non-null    float64
 7   Test.Diameter..mm. 280 non-null    float64
 8   Gonad.Weight     280 non-null    float64
 9   Gonad.Index..... 280 non-null    float64
 10  Treatment         280 non-null    object  
 11  Week              280 non-null    int64  
 12  Consumption       280 non-null    float64
 13  Group             280 non-null    object  
 14  DateGroup         280 non-null    object  
 15  ConsumptionbyWeight 280 non-null    float64
 16  TempLight         280 non-null    object  
 17  Light_encoded     280 non-null    int8   
 18  Temp_encoded      280 non-null    int8   
 19  Kelp_encoded      280 non-null    int8  
dtypes: category(3), float64(7), int64(1), int8(3), object(6)
```



MODELS USED

- Random Forest Regressor
- Linear
- LASSO
- Ridge

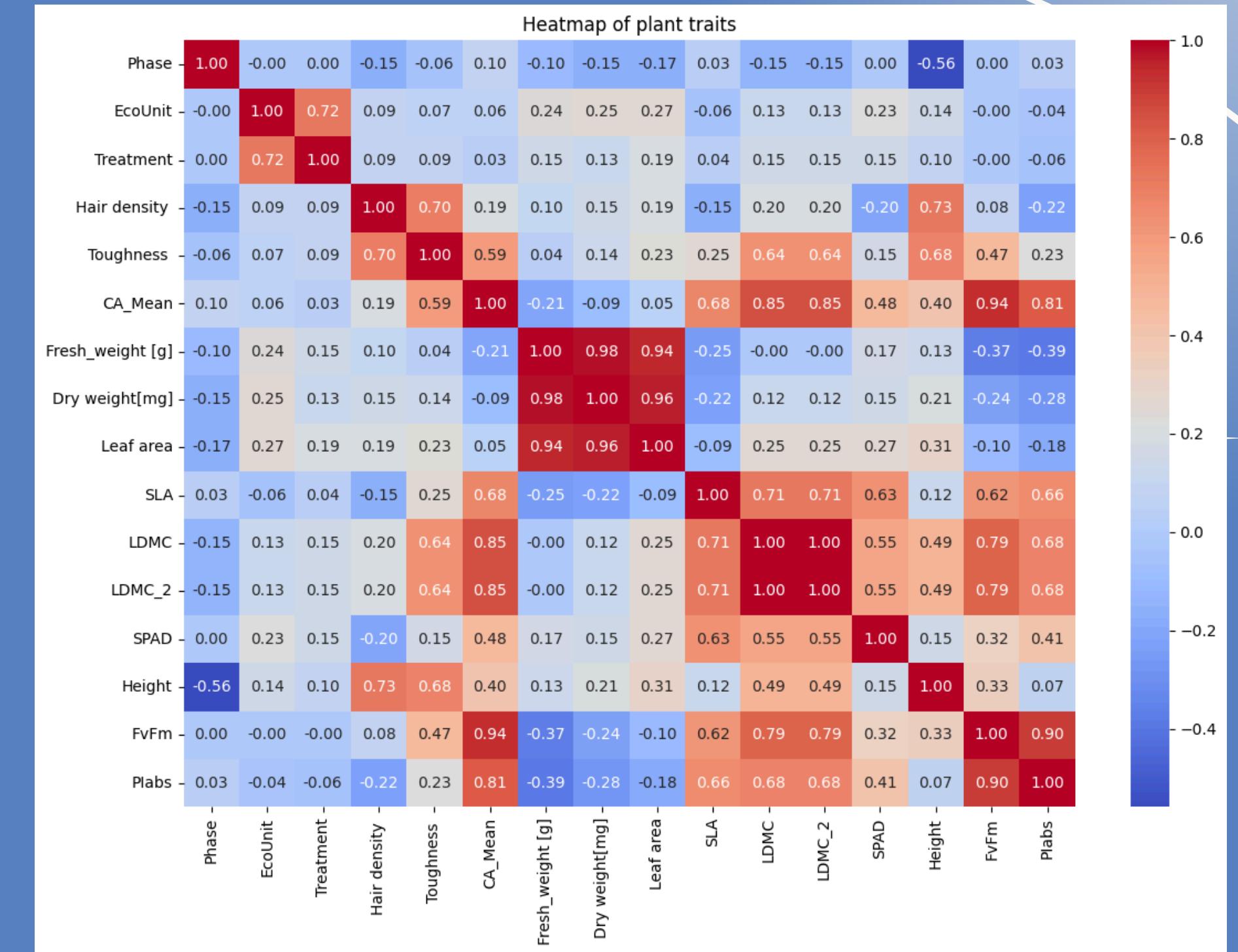
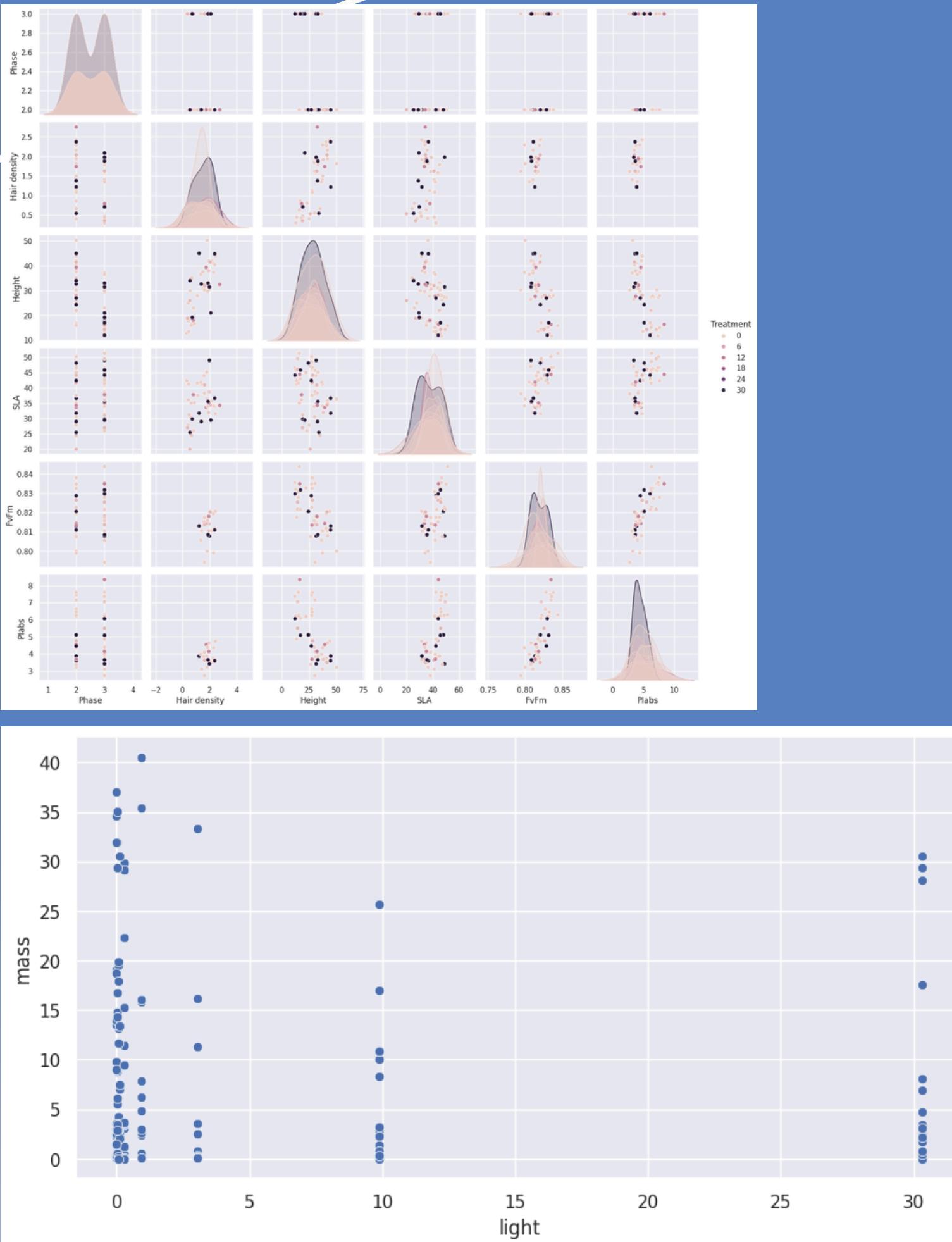
	Model	RMSE	R ²
0	Linear Regression	1.000000e-02	0.0639
1	LASSO	1.495600e+00	0.0541
2	Ridge	1.503000e+00	0.0447
3	Random Forest Regression	5.045627e-07	0.9825

PLANT DIVERSITY

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   ID               24 non-null      object  
 1   biomass          24 non-null      float64 
 2   Unit             24 non-null      int64   
 3   harvest          24 non-null      int64   
 4   Treatment [lux]  24 non-null      float64 
 5   Treatment [lux]_ln 24 non-null      float64 
dtypes: float64(3), int64(2), object(1)
memory usage: 1.3+ KB
```

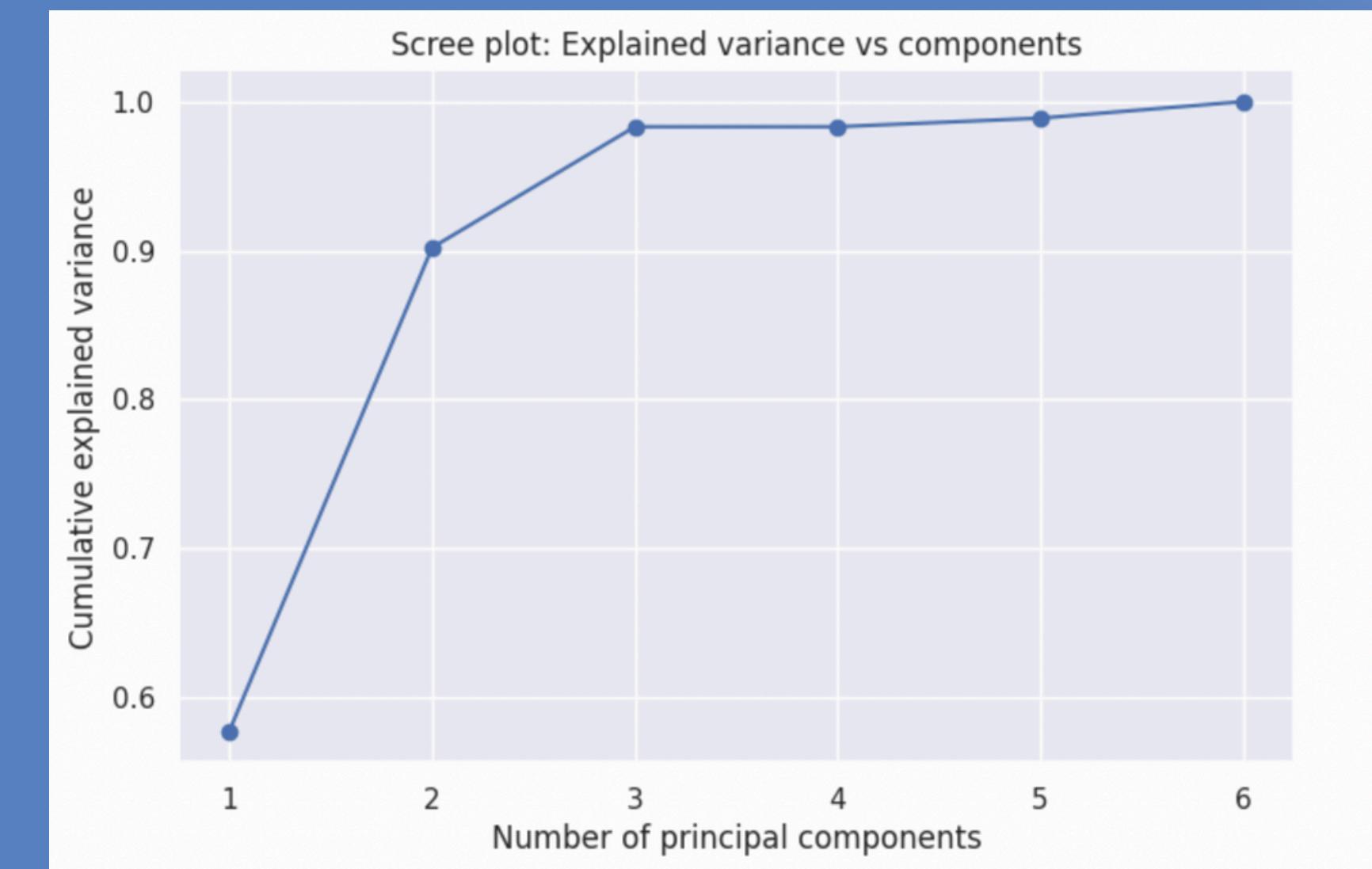
```
↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 72 entries, 0 to 71
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   ID               72 non-null      object  
 1   Species          72 non-null      object  
 2   Phase             72 non-null      int64   
 3   ID_neu            72 non-null      object  
 4   EcoUnit           72 non-null      int64   
 5   Treatment         72 non-null      float64 
 6   Hair density     45 non-null      float64 
 7   Toughness         69 non-null      float64 
 8   CA_Mean           69 non-null      float64 
 9   Fresh_weight [g]  68 non-null      float64 
 10  Dry weight [mg]  68 non-null      float64 
 11  Leaf area        68 non-null      float64 
 12  SLA              68 non-null      float64 
 13  LDMC              68 non-null      float64 
 14  LDMC_2            68 non-null      float64 
 15  SPAD              69 non-null      float64 
 16  Height            69 non-null      float64 
 17  FvFm              48 non-null      float64 
 18  PIabs             48 non-null      float64 
dtypes: float64(14), int64(2), object(3)
memory usage: 10.8+ KB
```

```
↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   species_ID       156 non-null      object  
 1   species          156 non-null      object  
 2   unit              156 non-null      int64   
 3   light             156 non-null      float64 
 4   mass              156 non-null      float64 
 5   harvest           156 non-null      int64  
dtypes: float64(2), int64(2), object(2)
memory usage: 7.4+ KB
```



SCREE PLOT

- ID
- biomass
- unit
- harvest
- Light treatment
- Light treatment (log transformed)



MODELS USED

- Random Forest Regression
- LASSO

LASSO Results:
 $R^2: 0.4568$
 $RMSE: 8.7658$

Random Forest Results:
 $R^2: 0.9992$
 $RMSE: 0.1163$

CHALLENGES

- Figuring out model works best
- Small dataset
- Feature importance

BIRD MIGRATIONS

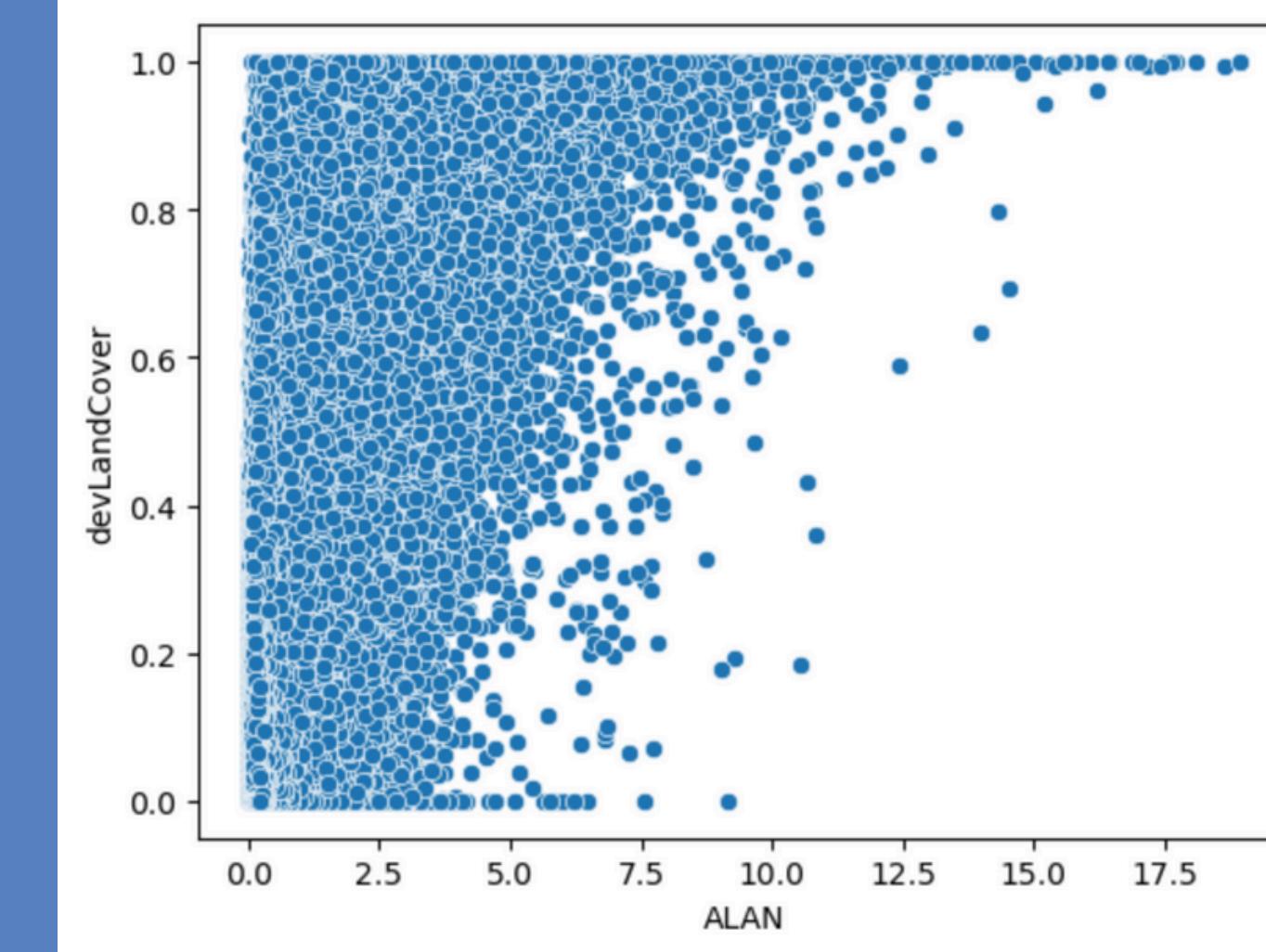
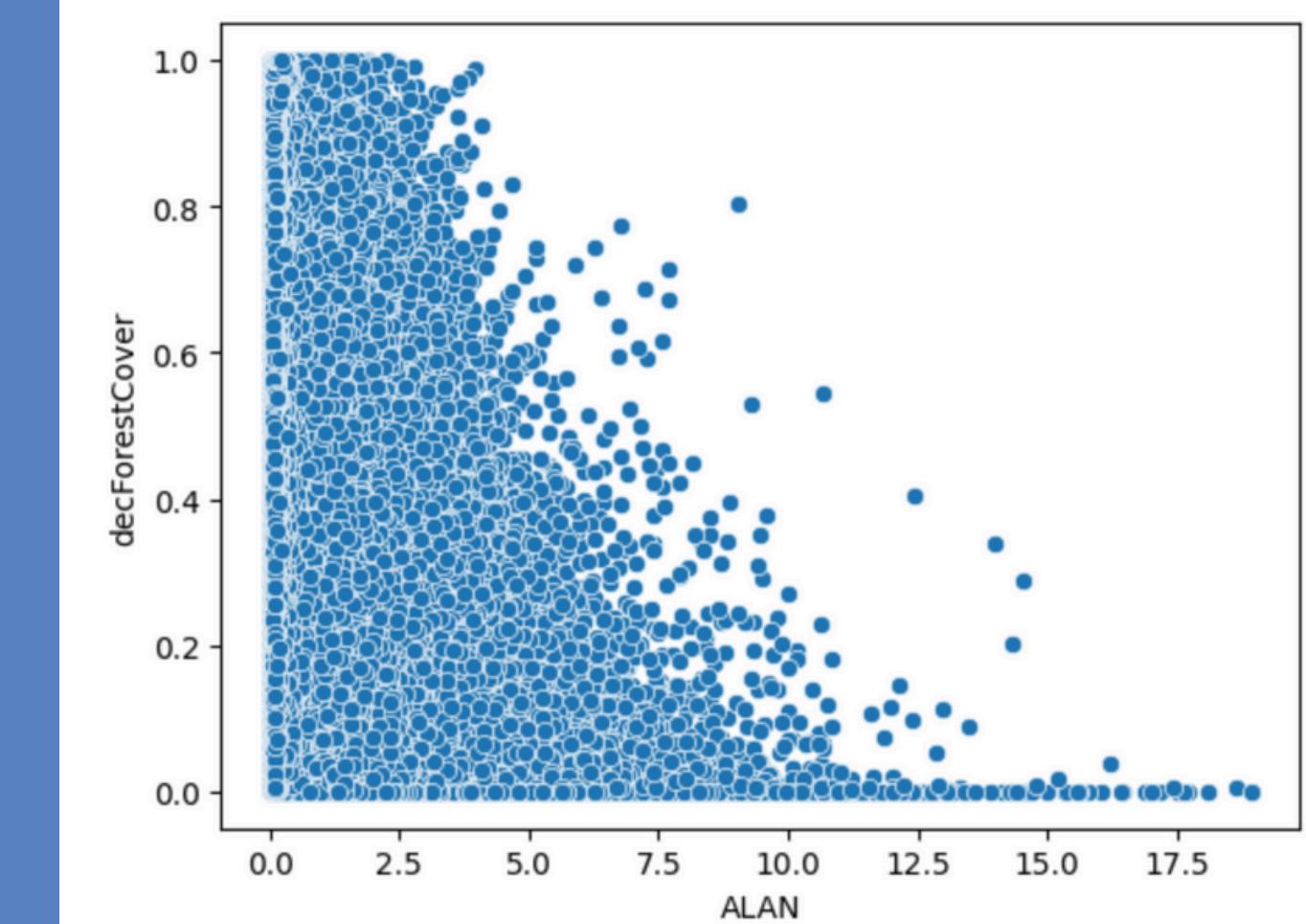
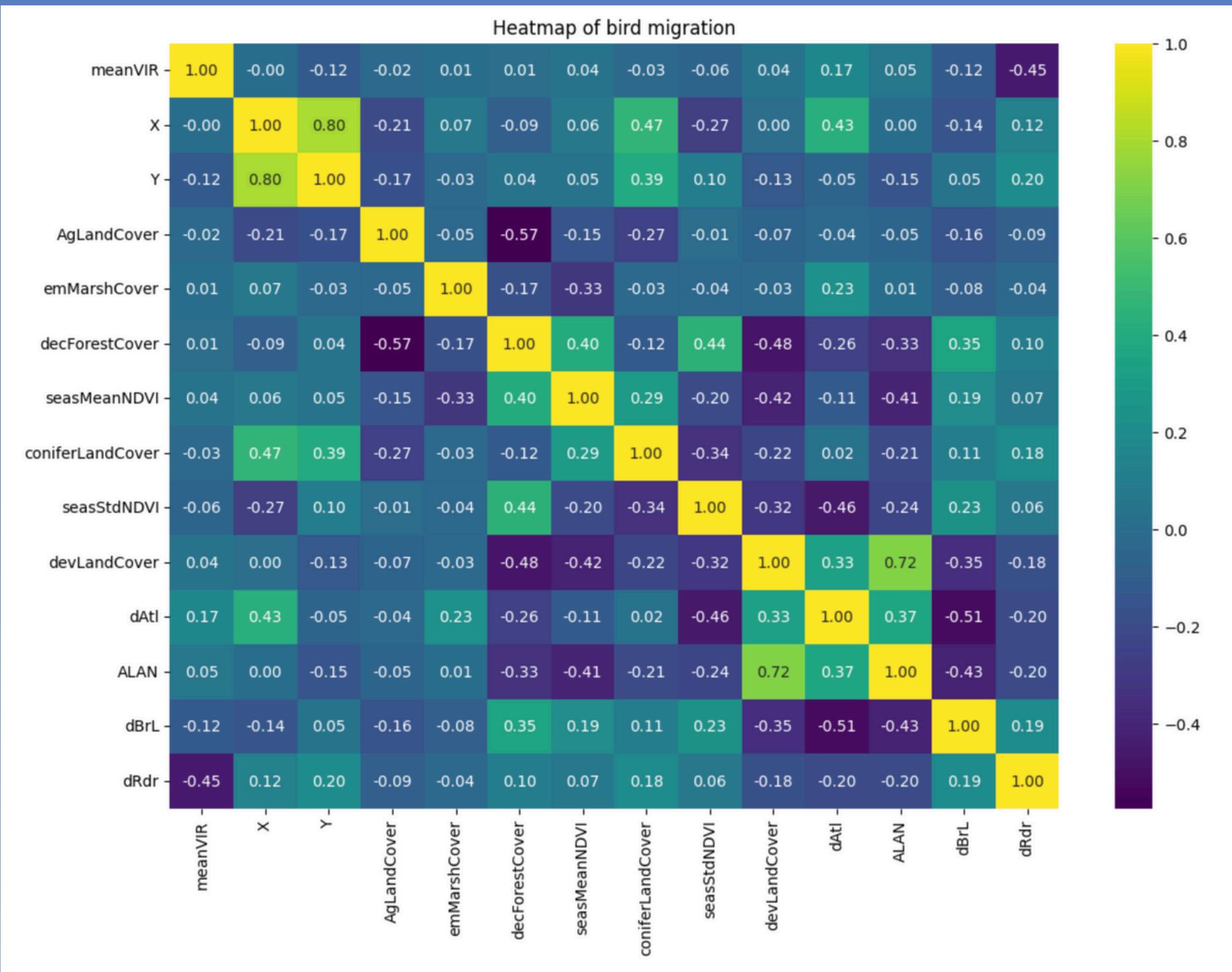
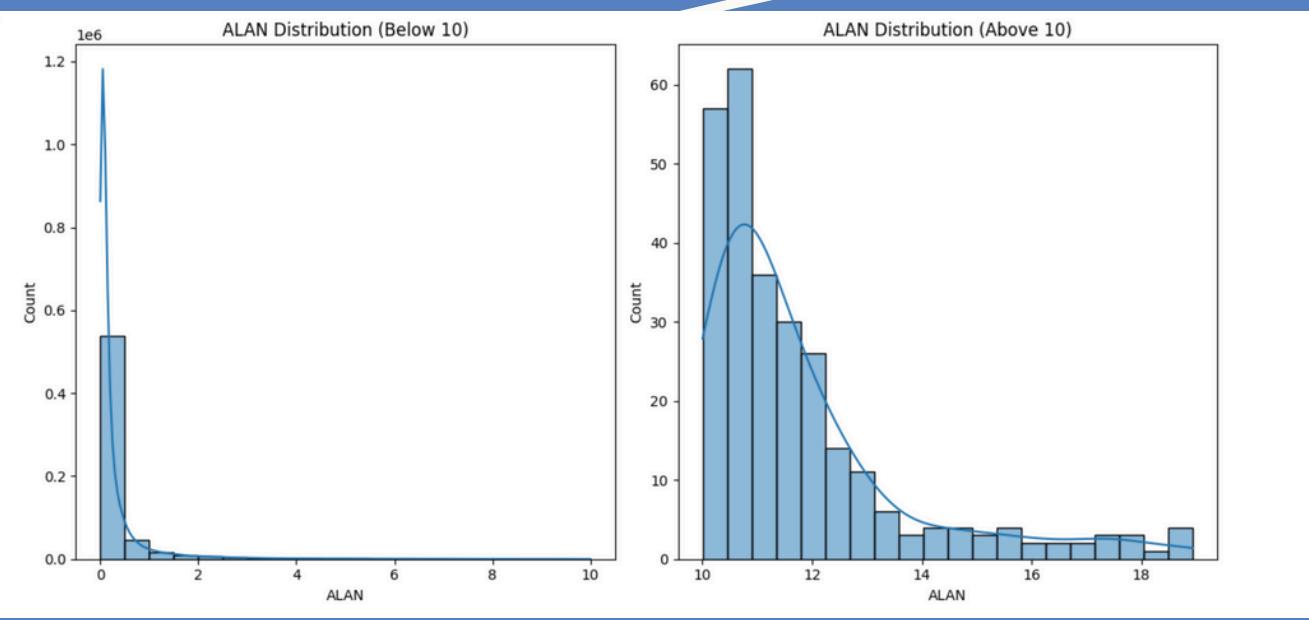
<https://datadryad.org/dataset/doi:10.5061/dryad.hhmqgnknt>

- Target of study
- How this is related to ALAN and undark skies

DEFINITIONS

- ALAN
- dBrL
- devLandCover
- AgLandCover
- emMarshCover
- decForestCover
- seasMeanNDVI
- dAtl
- dRdr

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 628603 entries, 0 to 628602
Data columns (total 14 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   meanVIR          166296 non-null    float64
 1   X                628603 non-null    float64
 2   Y                628603 non-null    float64
 3   AgLandCover      628603 non-null    float64
 4   emMarshCover     628603 non-null    float64
 5   decForestCover   628603 non-null    float64
 6   seasMeanNDVI    628603 non-null    float64
 7   coniferLandCover 628603 non-null    float64
 8   seasStdNDVI     628603 non-null    float64
 9   devLandCover     628603 non-null    float64
 10  dAtl             628603 non-null    float64
 11  ALAN            628603 non-null    float64
 12  dBrL            628603 non-null    float64
 13  dRdr            628603 non-null    float64
dtypes: float64(14)
memory usage: 67.1 MB
```



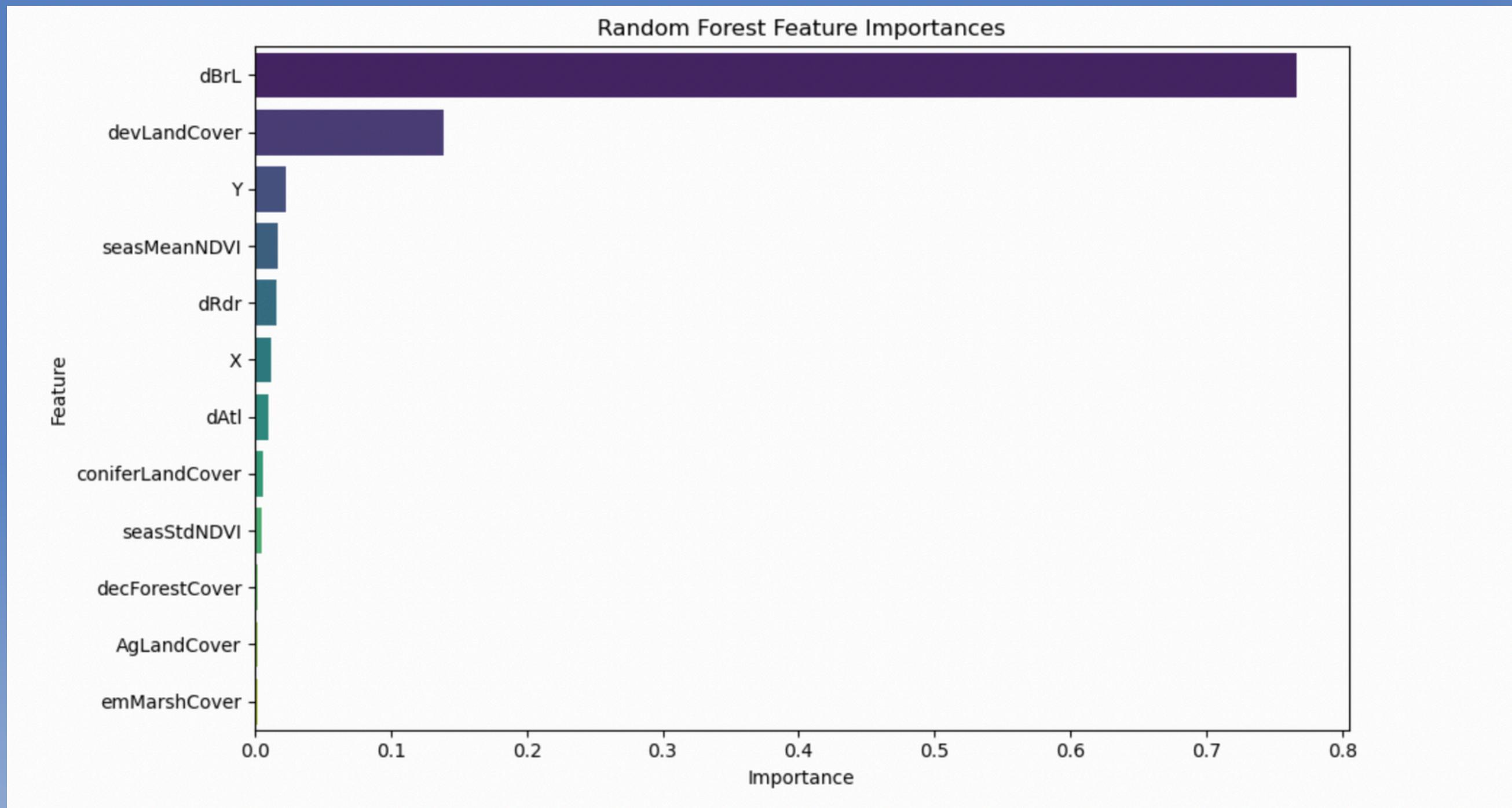
MODELS USED

- Linear Regression
- Ridge Regression
- LASSO Regression
- Random Forest Regression

	Model	R^2	RMSE
0	Linear Regression	0.585596	0.487451
1	Ridge Regression	0.585592	0.487453
2	LASSO Regression	0.239220	0.660463
3	Random Forest	0.943960	0.179254

	Model	R^2	RMSE
0	Linear Regression	0.585596	0.487451
1	Ridge Regression	0.585596	0.487451
2	LASSO Regression	0.529168	0.519579
3	Random Forest	0.943948	0.179274

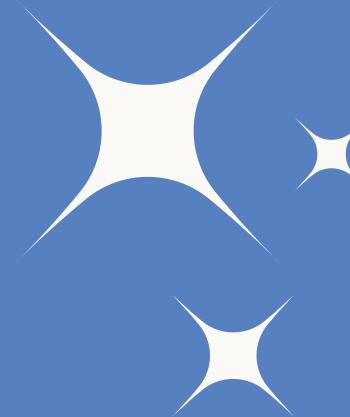
FEATURE IMPORTANCE



CHALLENGES

- 600,000+ rows of data
- Getting the same R² and RMSE values for Linear, Ridge, and LASSO at one point
- Random Forest wasn't working
- Confronting unexpected results

THANK YOU!



CONCLUSION

In conclusion, we found that by looking at a variety of environments and the effect of ALAN on them that ALAN has a negative effect on the environment as a whole.

