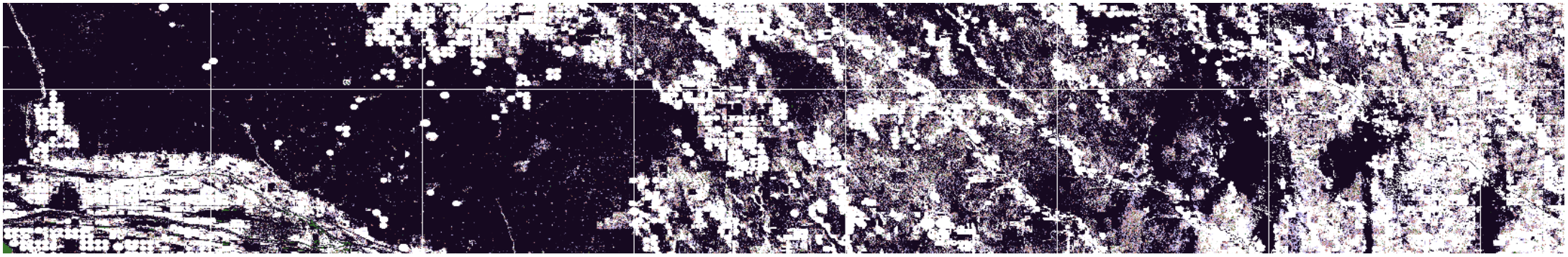
Filling the Land Conversion Gap with Machine Learning

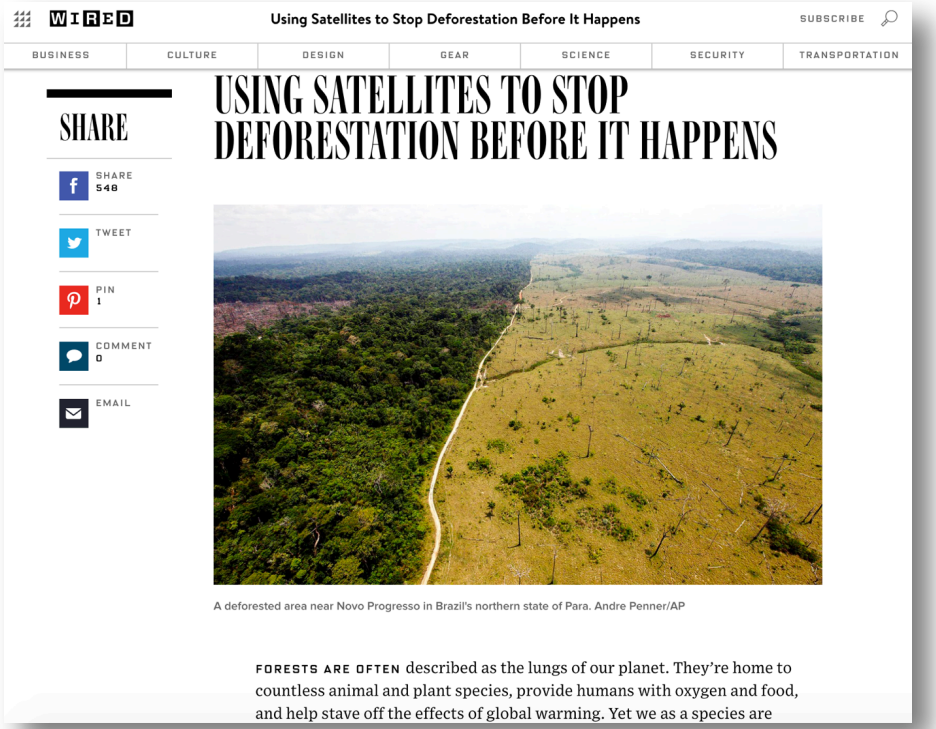
Byron Allen

Data Science Student at General Assembly

22 November 2016



Lack of Land Conversion Data



WIRED Using Satellites to Stop Deforestation Before It Happens SUBSCRIBE

BUSINESS CULTURE DESIGN GEAR SCIENCE SECURITY TRANSPORTATION

SHARE

SHARE 548

TWEET

PIN 1

COMMENT 0

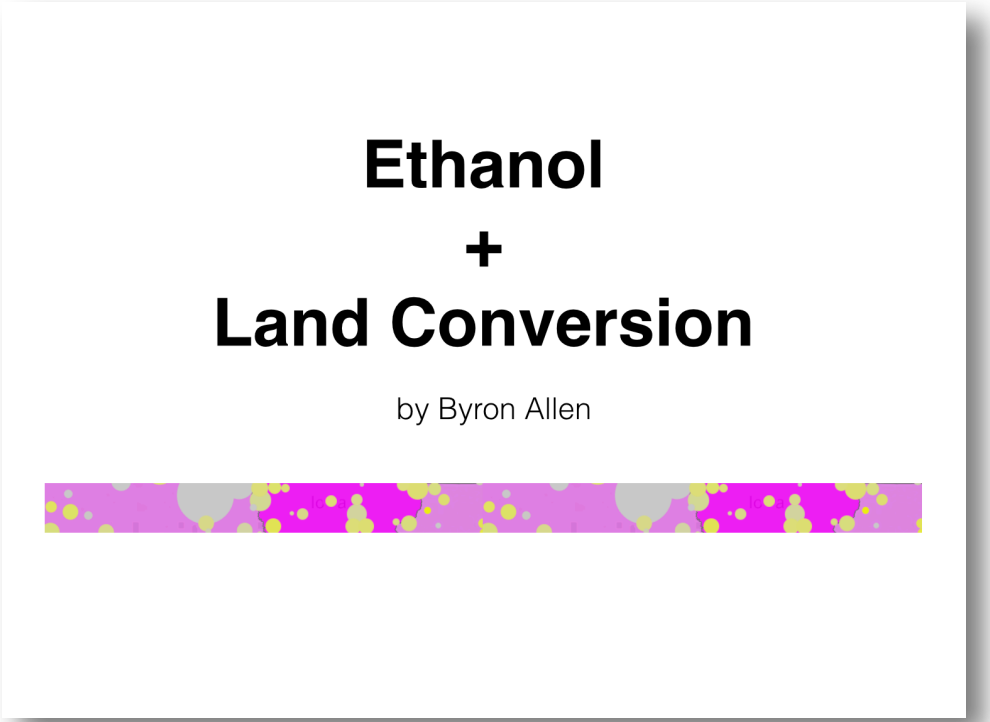
EMAIL

USING SATELLITES TO STOP DEFORESTATION BEFORE IT HAPPENS

A deforested area near Novo Progresso in Brazil's northern state of Para. Andre Penner/AP

FORESTS ARE OFTEN described as the lungs of our planet. They're home to countless animal and plant species, provide humans with oxygen and food, and help stave off the effects of global warming. Yet we as a species are

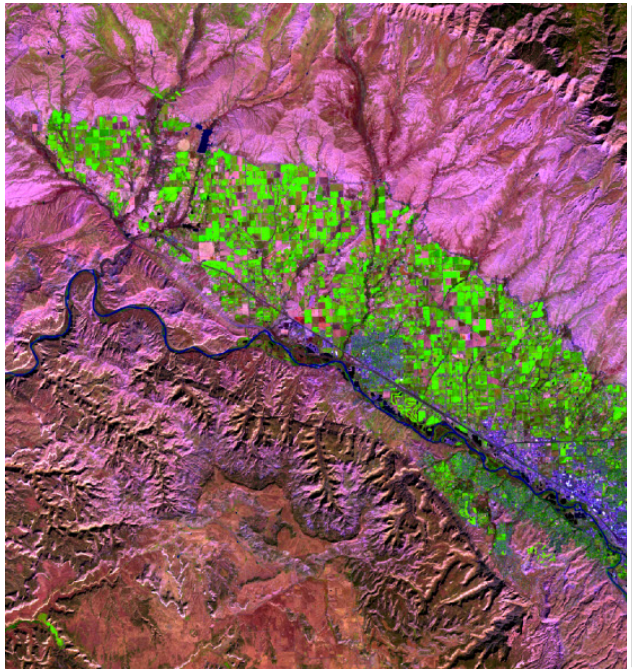
https://www.wired.com/2015/04/using-smart-satellites-to-monitor-deforestation-from-space/



PDF link
Video Presentation link
Tableau Dashboard link

How do we fill the gap?

Satellite Images (Landsat 8 EVI & RAW)

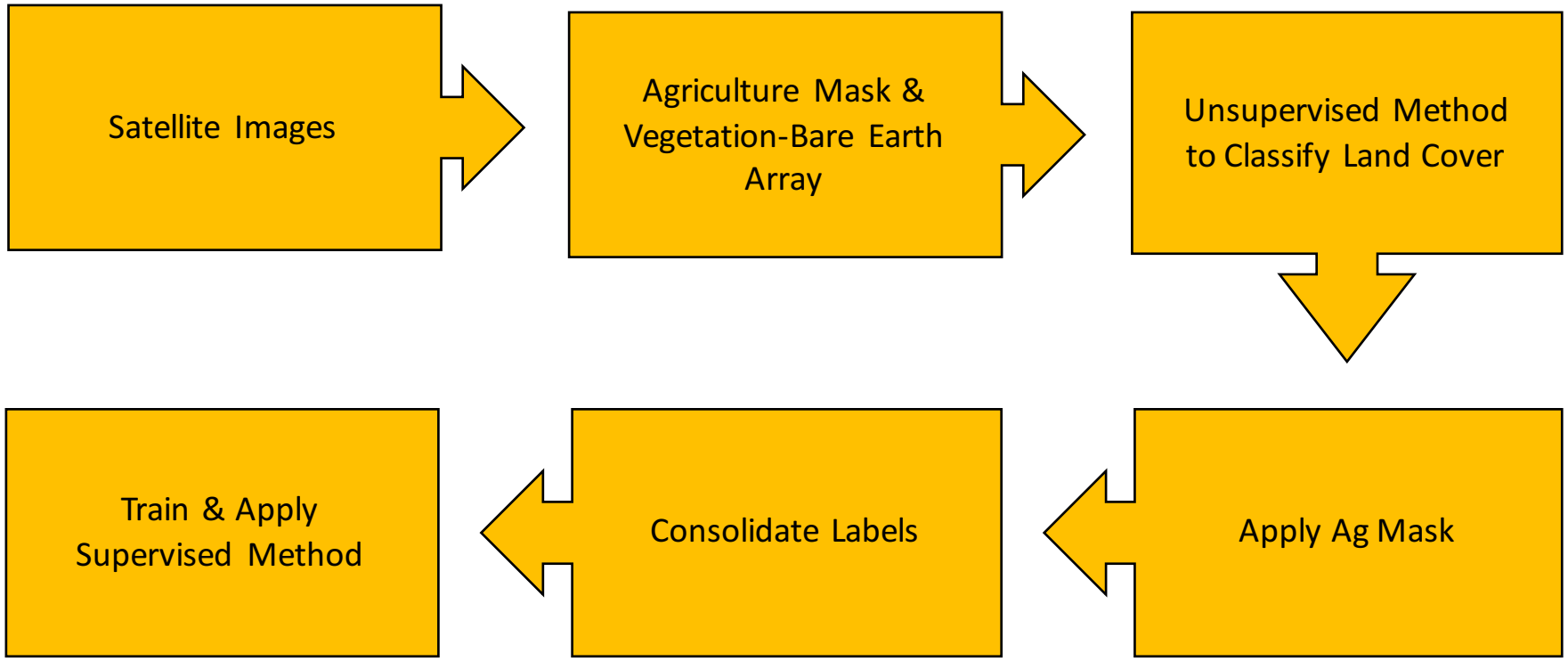


Band	Wavelength (micrometers)	Resolution (meters)
Band 1 – Coastal Aerosol	0.43 – 0.45	30
Band 2 – Blue	0.45 – 0.51	30
Band 3 – Green	0.53 – 0.59	30
Band 4 – Red	0.64 – 0.67	30
Band 5 – Near Infrared (NIR)	0.85 – 0.88	30
Band 6 – SWIR 1	1.57 – 1.65	30
Band 7 – SWIR 2	2.11 – 2.29	30
Band 8 – Panchromatic	0.50 – 0.68	15
Band 9 – Cirrus	1.36 – 1.38	30
Band 10 – Thermal Infrared (TIRS) 1	10.60 – 11.19	100
Band 11 – Thermal Infrared (TIRS) 2	11.50 – 12.51	100

Share Image

http://www.harrisgeospatial.com/Home/NewsUpdates/TabId/170/ArtMID/735/ArticleID/14305/The-Many-Band-Combinations-of-Landsat-8.aspx

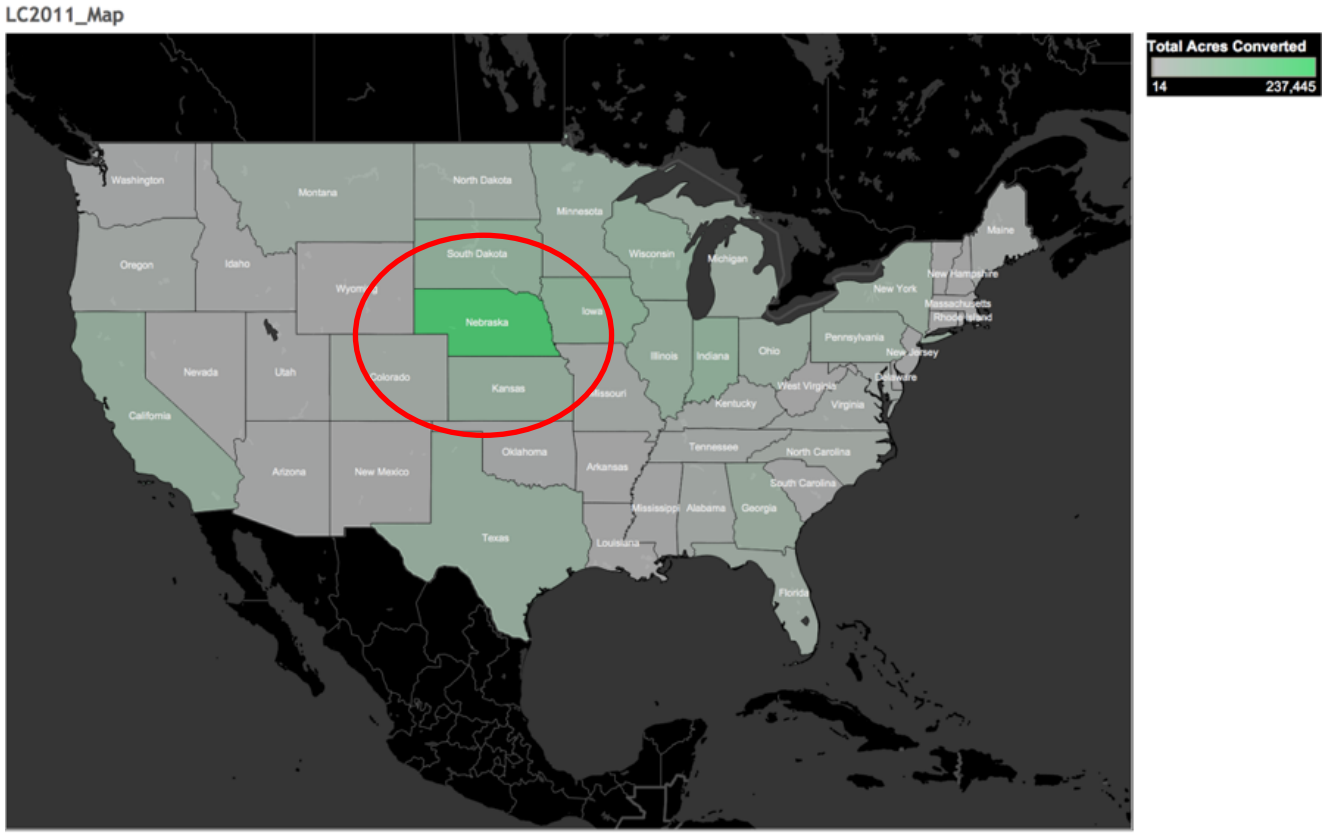
Workflow


Satellite Images
Agriculture Mask & Vegetation-Bare Earth Array
Unsupervised Method to Classify Land Cover
Apply Ag Mask
Consolidate Labels
Train & Apply Supervised Method

Assumptions & Limitations

- Geospatial sits outside the scope of GA
- Clouds were avoided as much as possible
- Importance placed on the workflow overall
- Some models still running as of today
- Need for mixed land type to train on
- Data comes from composite images ranges from 2013 – 2015

Nebraska had the Greatest LC in 2011

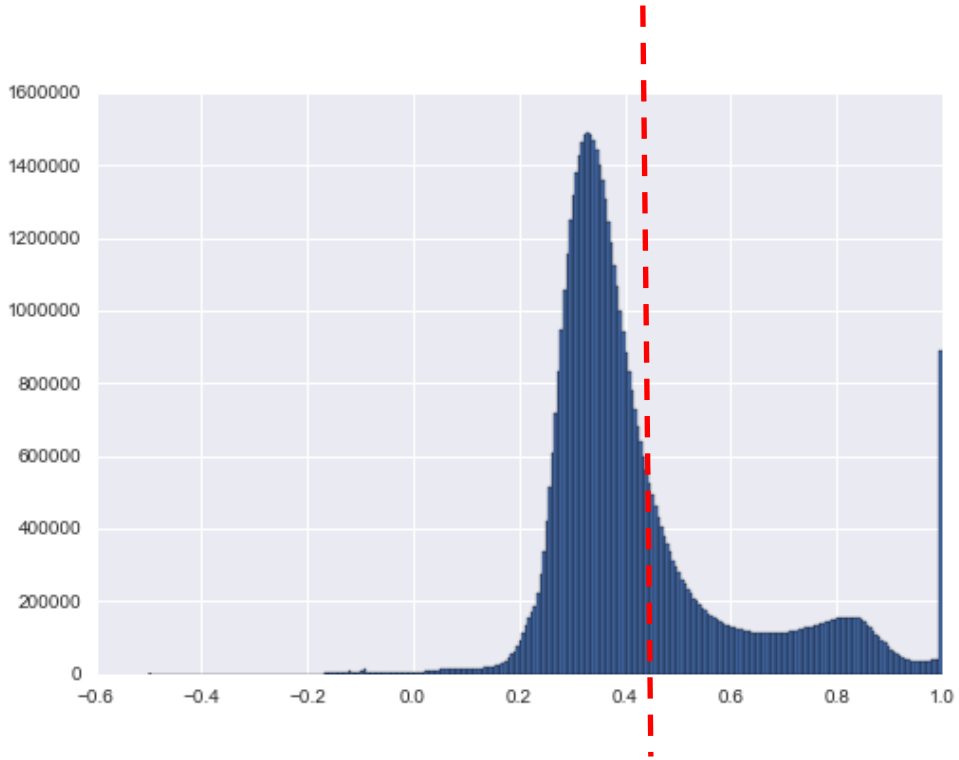


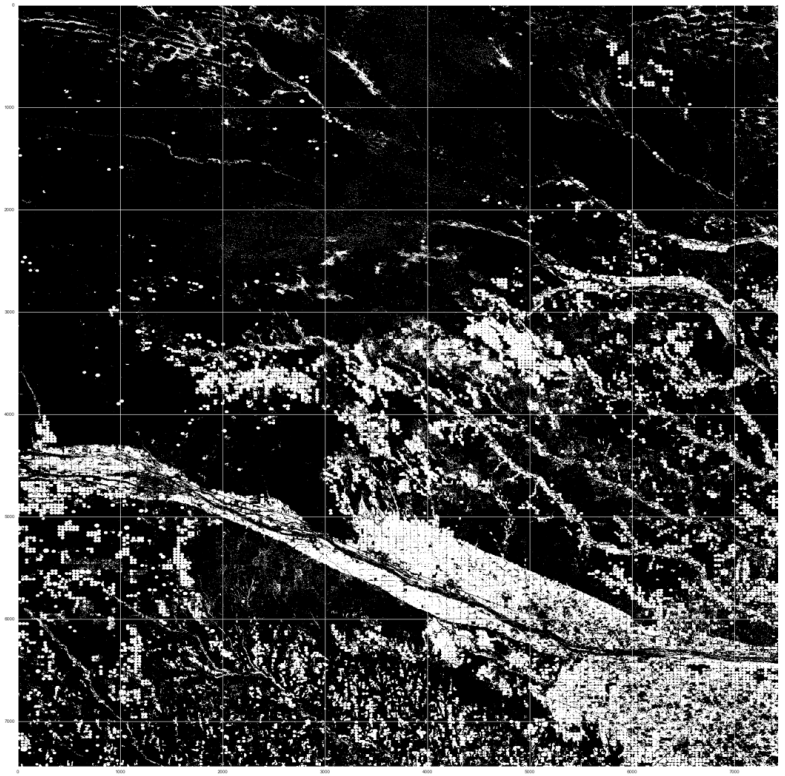
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Acres Converted. The marks are labeled by State. The data is filtered on Year, which keeps 2011.

Shape of Data | A Strong Dataset to Build On

- GTiff file downloaded from Google Earth Engine
- Middle of Nebraska, U.S.
- Resolution: 30m/pixel
- Height: 7439
- Width: 7422
- Pixel Count: 55,212,258
- Over 50,000km^2 of Nebraska's total area 200,360km^2
- For reference, VIC is 237,629 km^2

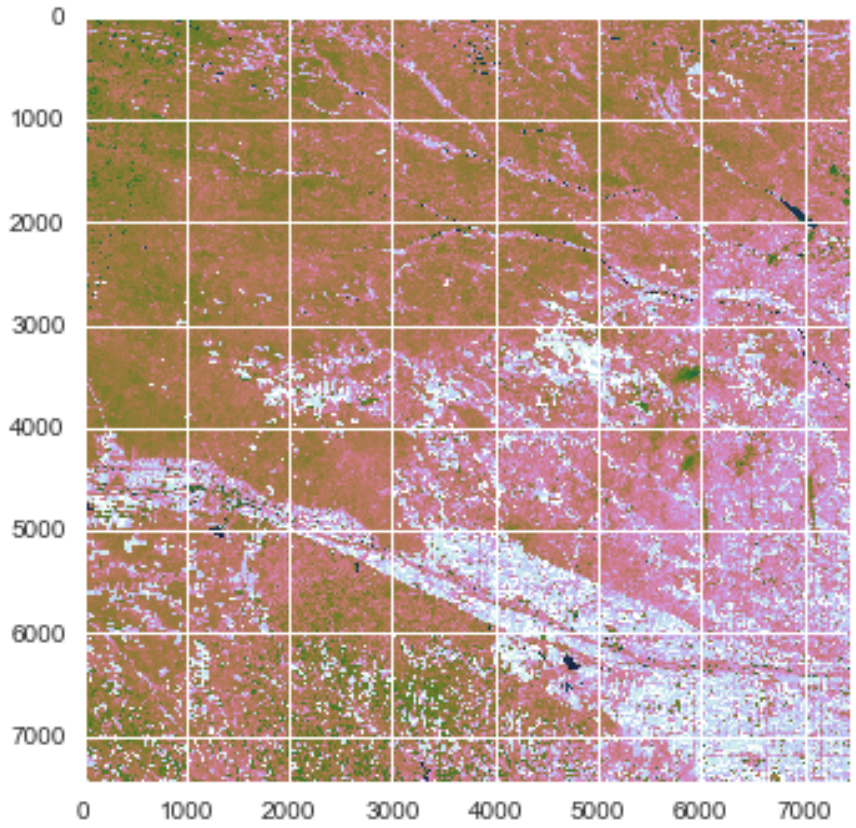
Bimodal Distribution to Create Ag Mask





Create Optimal Training Area

(Vegetation Array = (band_6 + band_5 + band_2) / 3)
-
(Bare Earth Array = (band_6 + band_3 + band_2) / 3)

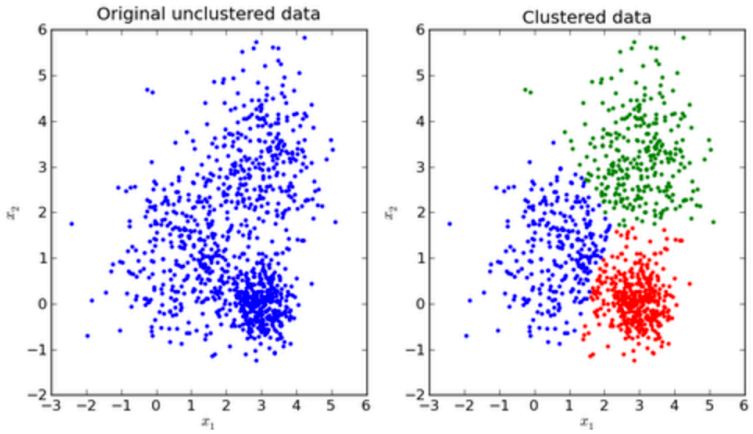


Unsupervised Model Selection

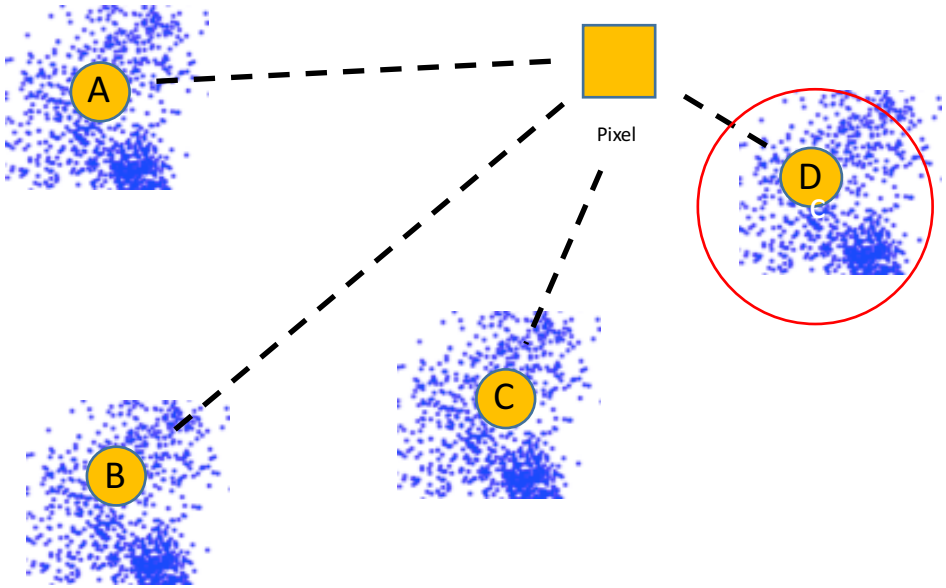
Key in defining underlying features of unlabeled data

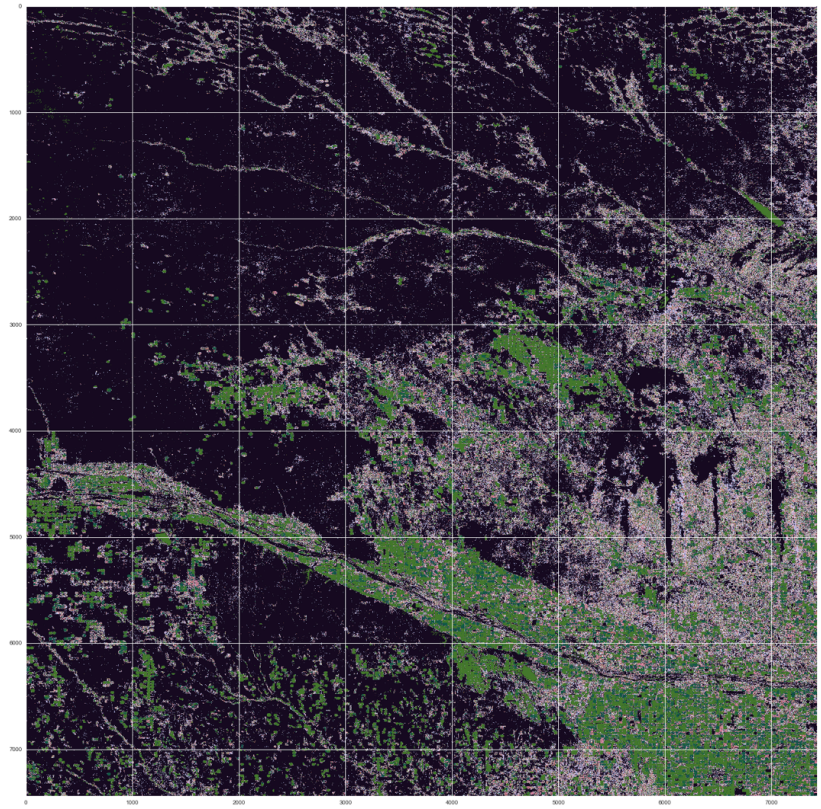
- KMeans
- Agglomerative Clustering
- Affinity Propagation

KMeans Example:

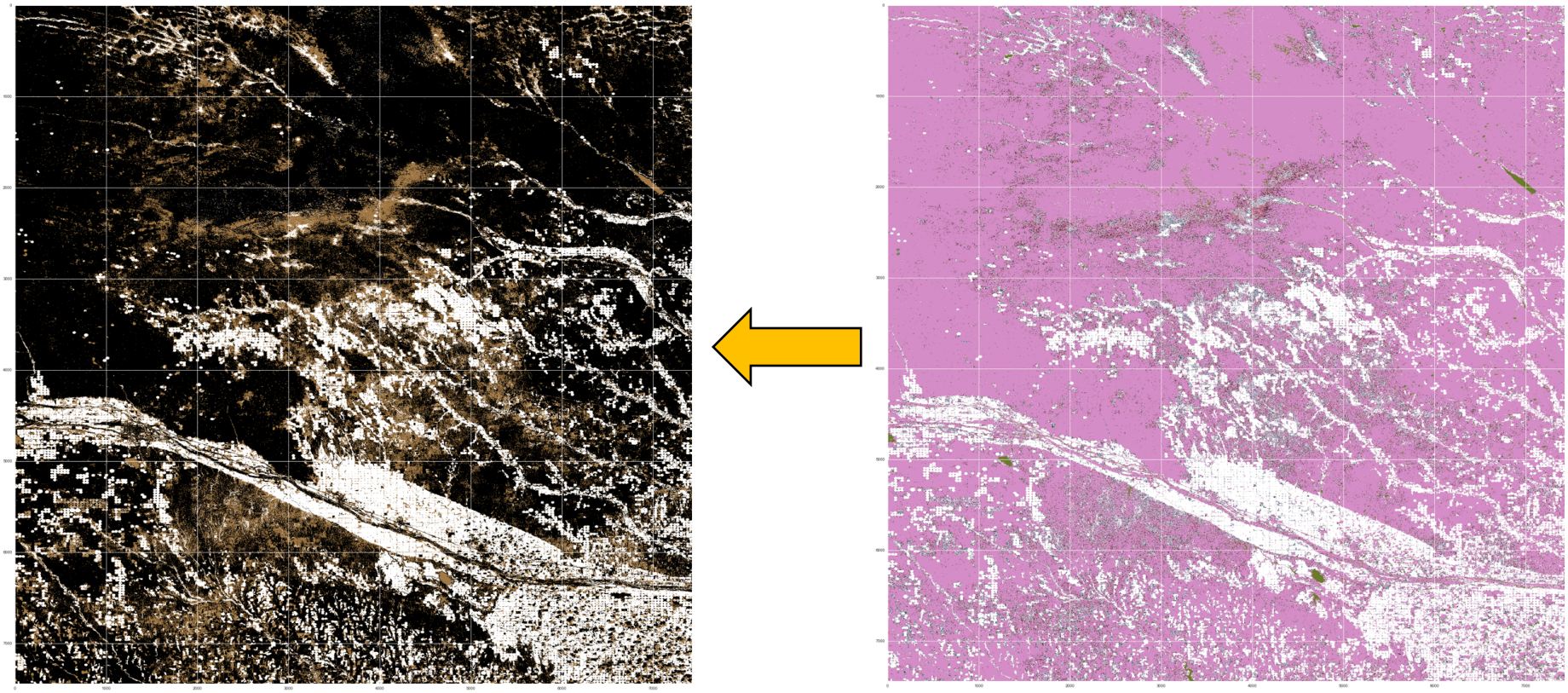


KMeans (k=30) Centroid-based Labels

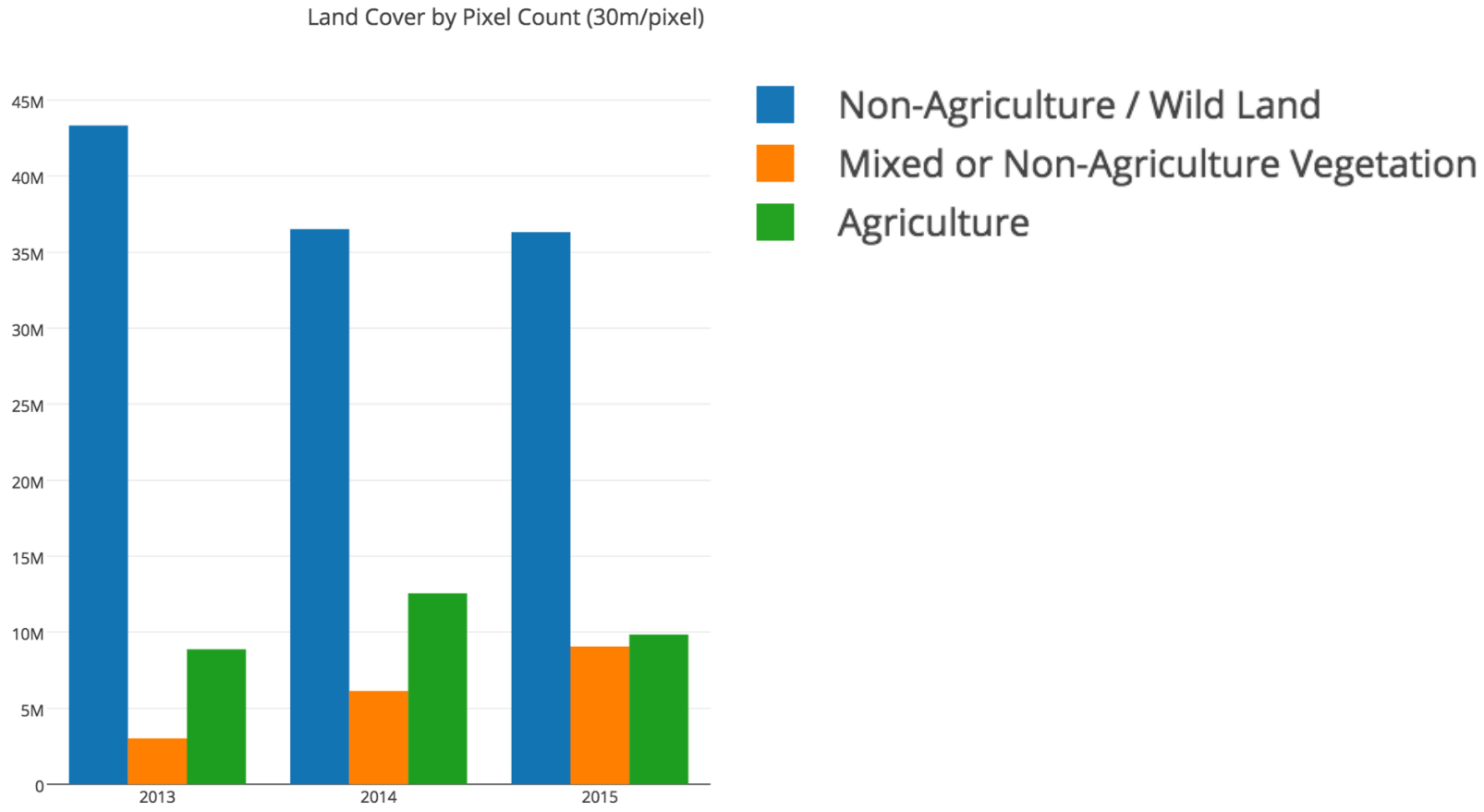




Apply Ag Mask and Consolidate Labels

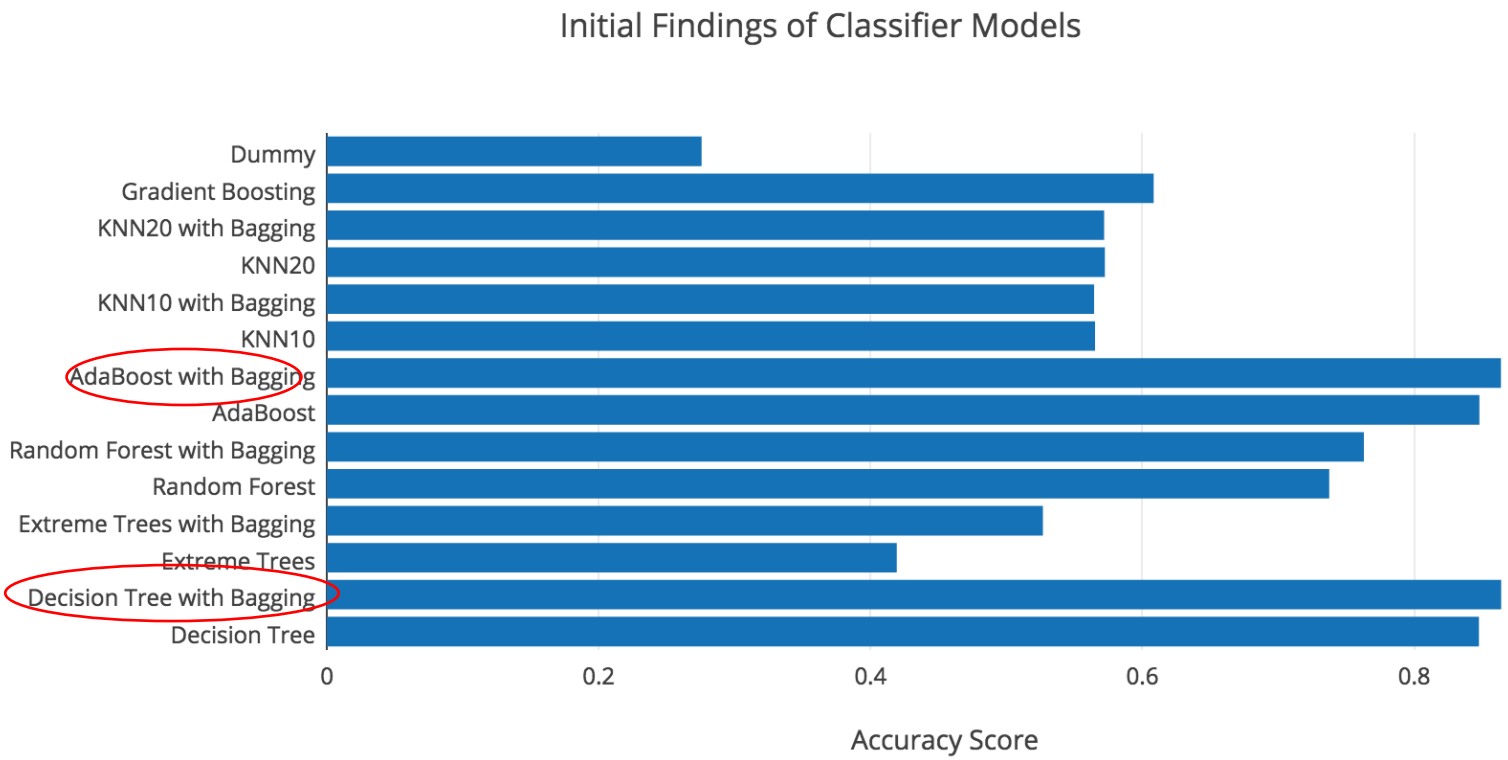


# Indicates Possible Land Conversion



Supervised Results

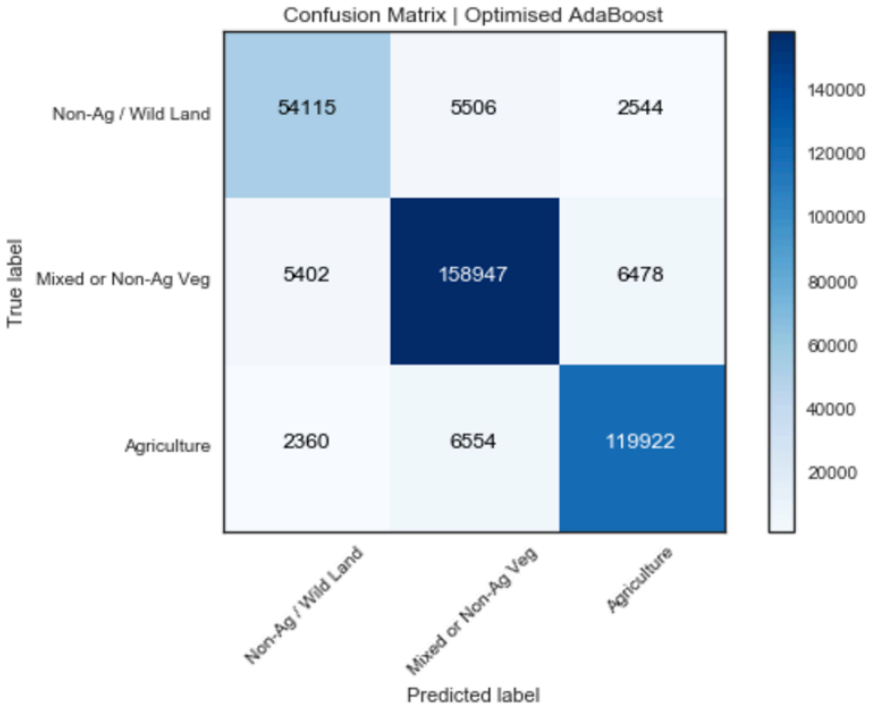
Initial EDA



AdaBoost + Decision Tree 92% Acc.

Initial EDA

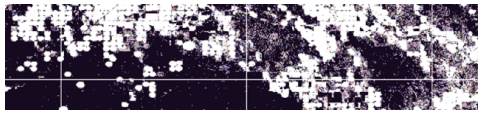
```
REPORT ON THE PERFORMANCE OF AN OPTIMISED AdaBoost MODEL

Accuracy:
0.920282565197

Classification Report:
             precision    recall  f1-score   support

        1.0       0.87      0.87      0.87     62165
        2.0       0.93      0.93      0.93    170827
        3.0       0.93      0.93      0.93    128836

avg / total       0.92      0.92      0.92    361828

{'algorithm': 'SAMME',
 'learning_rate': 0.5,
 'n_estimators': 30,
 'random_state': 42}
```



Recommendations/Next Steps

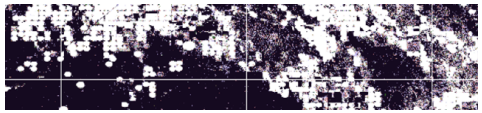
- Include final supervised results
- Improve ag masks
- Fine-tune the unsupervised method (e.g. mixed vegetation label)
- Sort out requirements to test hierarchical algorithms
- PCA across all bands
- Test a new 12-month workflow idea

Q&A

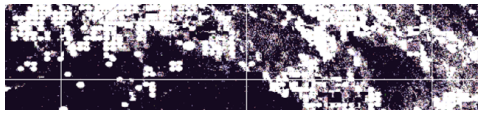


byron.a.allen@gmail.com

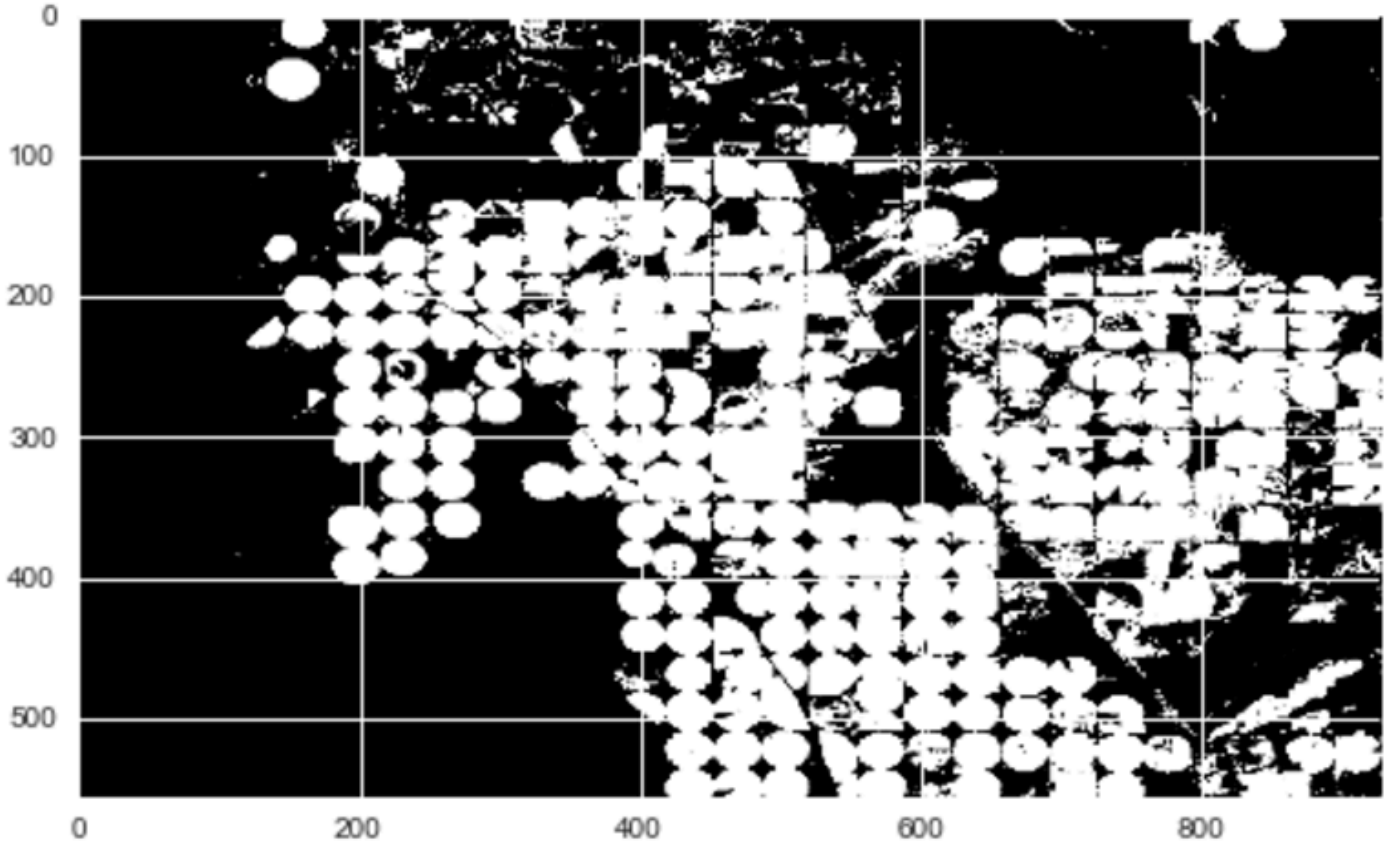
THANKS



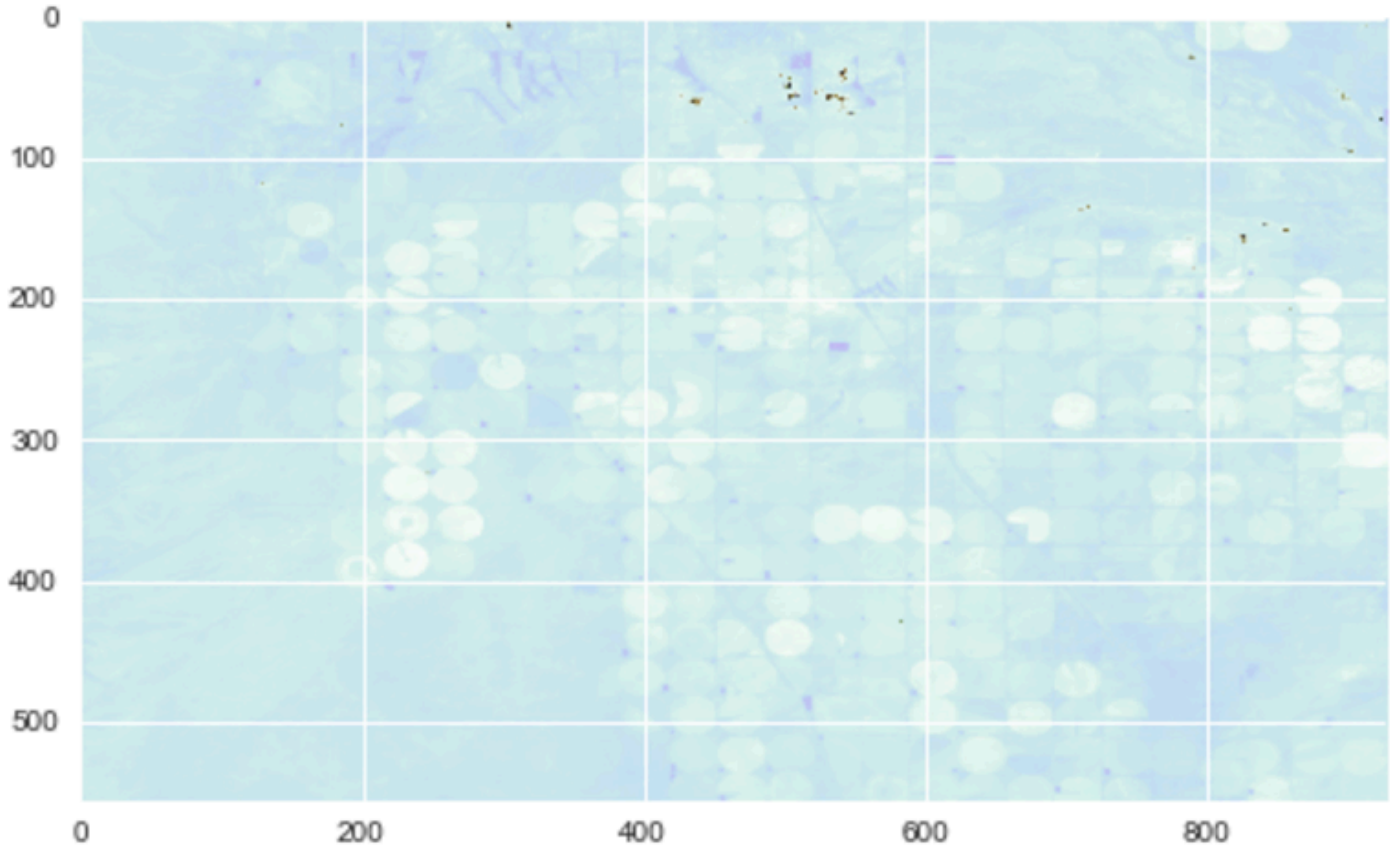
APPENDIX



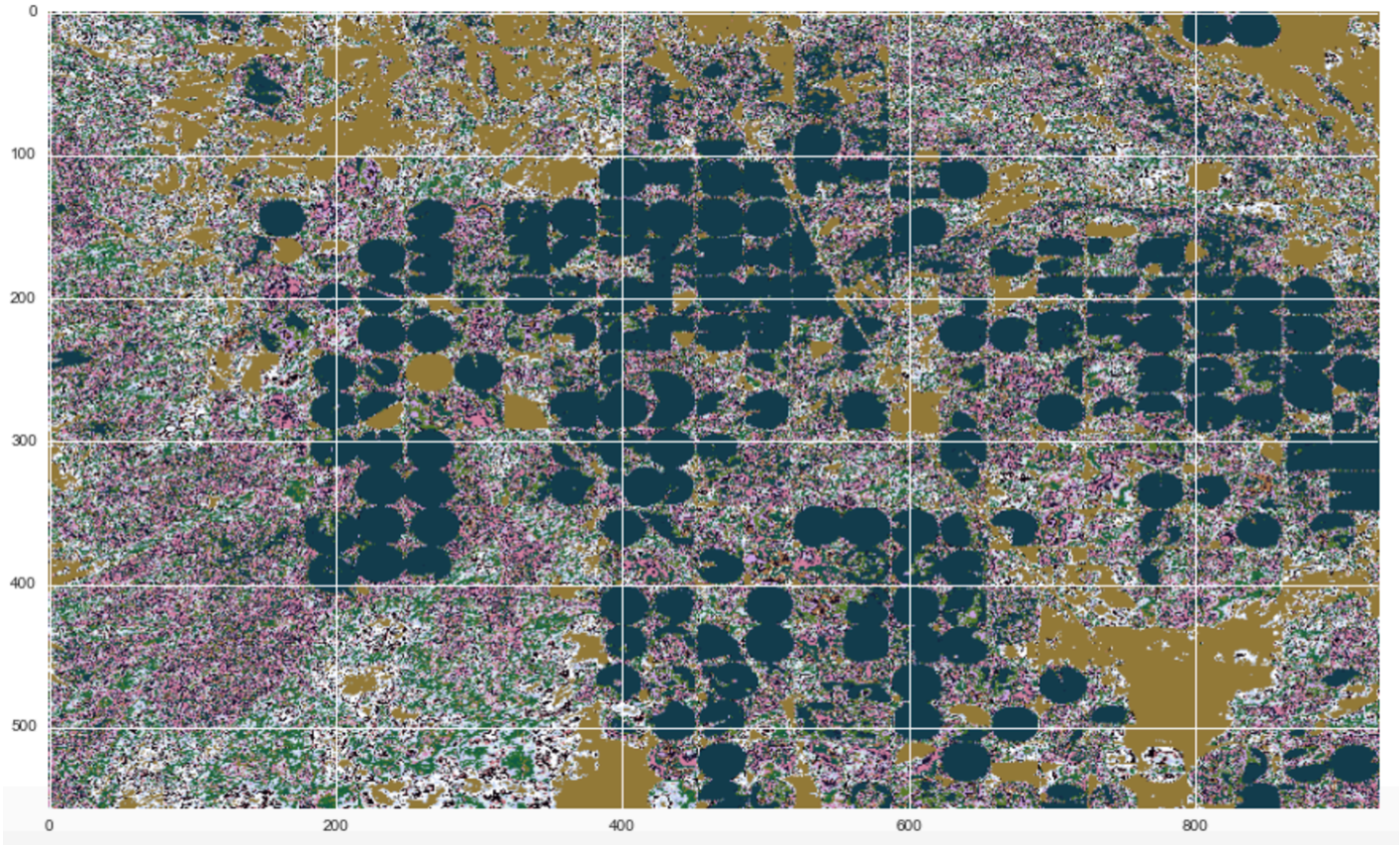
Agriculture Mask Example



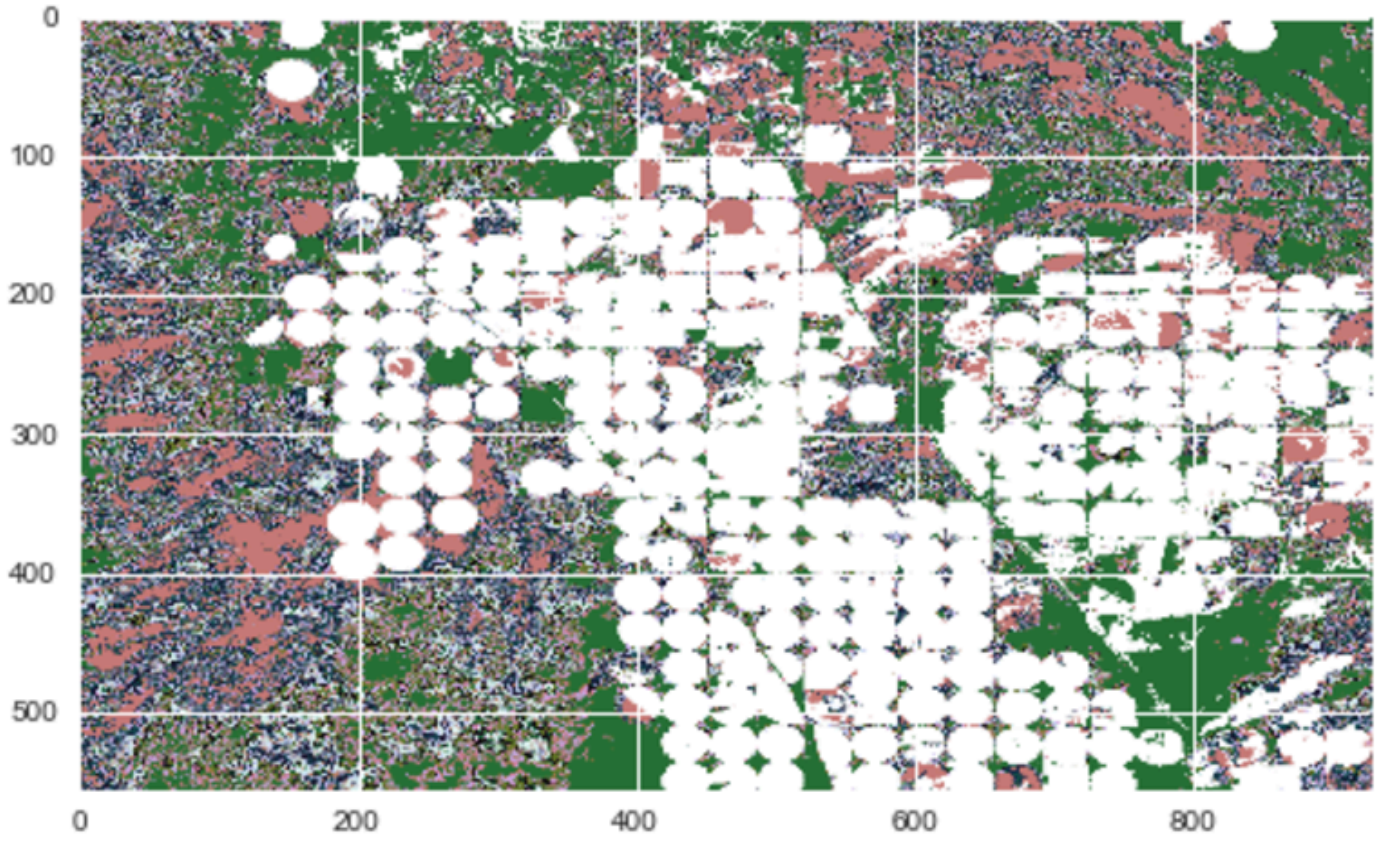
Vegetation-Bare Earth Array Example



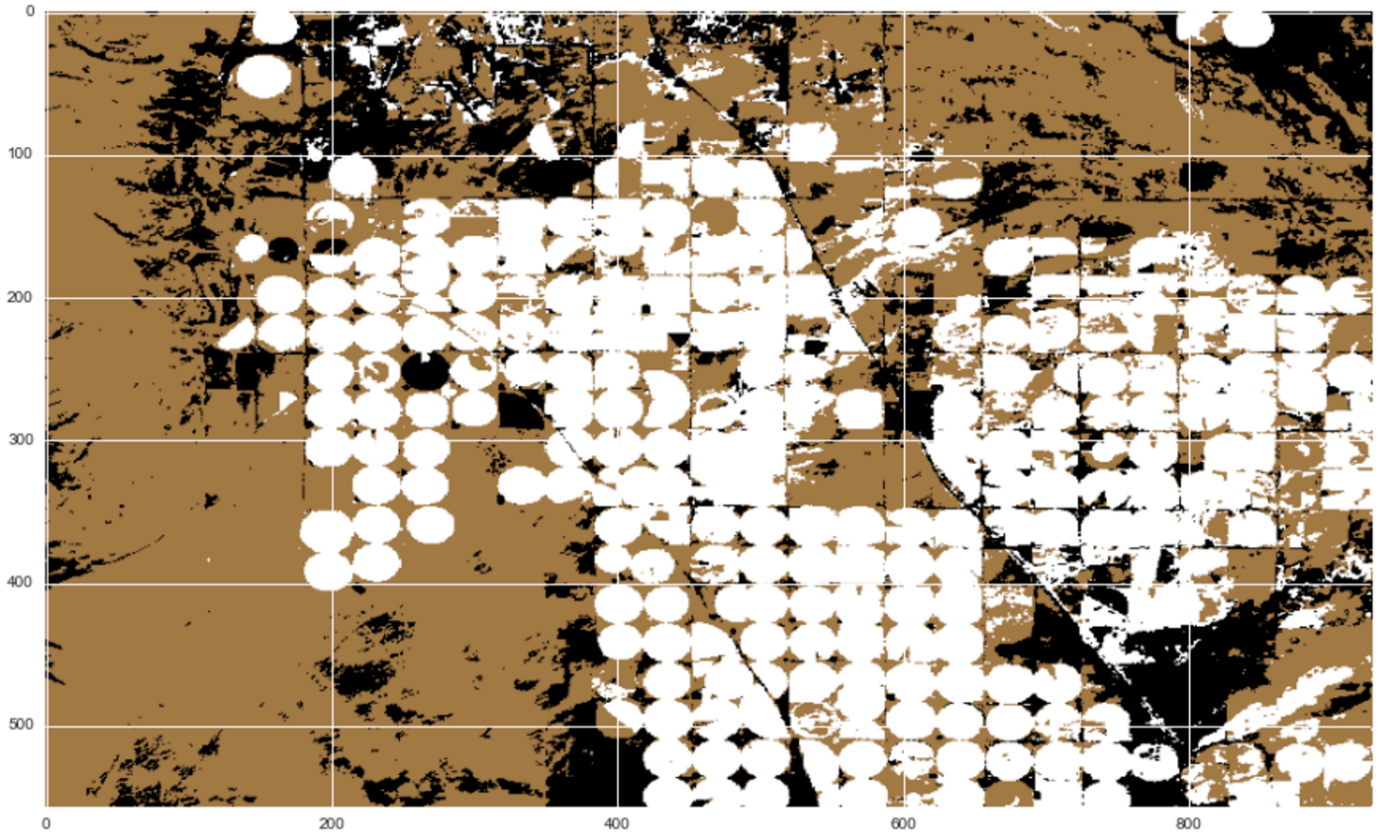
KMeans Centroid-based Labels Example



Applied Agriculture Mask Example



Consolidated Labels Example



GridsearchCV Results
(Best Supervised Models + Dummy)

Initial EDA

REPORT ON THE PERFORMANCE OF AN OPTIMISED DECISION TREE

Accuracy:
0.851404534751

Classification Report:

	precision	recall	f1-score	support
1.0	0.83	0.78	0.80	62165
2.0	0.86	0.89	0.88	170827
3.0	0.84	0.84	0.84	128836
avg / total	0.85	0.85	0.85	361828

REPORT ON THE PERFORMANCE OF AN OPTIMISED AdaBoost MODEL

Accuracy:
0.920282565197

Classification Report:

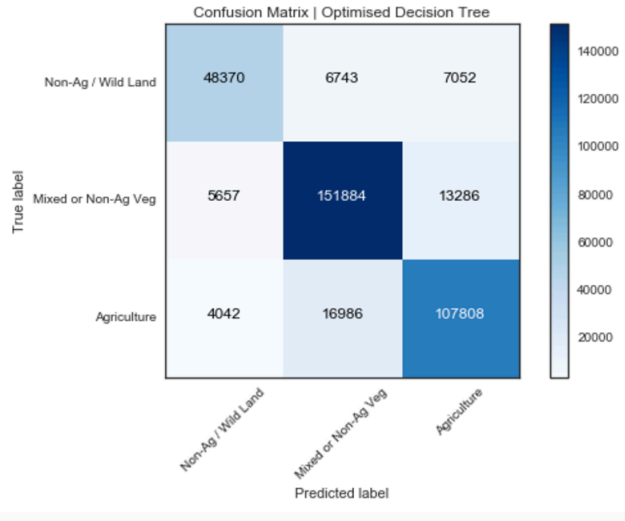
	precision	recall	f1-score	support
1.0	0.87	0.87	0.87	62165
2.0	0.93	0.93	0.93	170827
3.0	0.93	0.93	0.93	128836
avg / total	0.92	0.92	0.92	361828

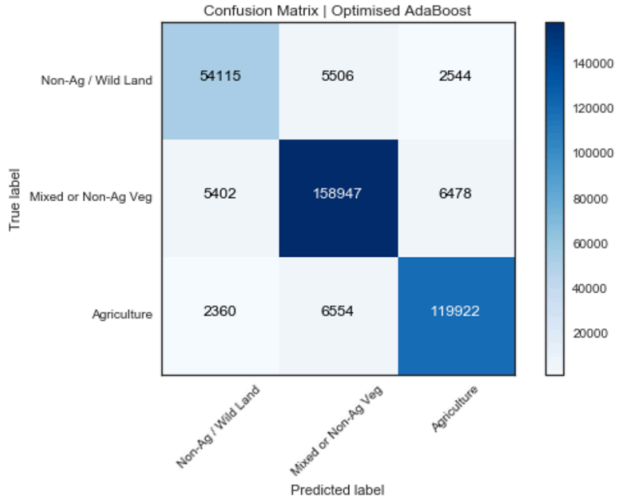
REPORT ON THE PERFORMANCE OF DUMMY

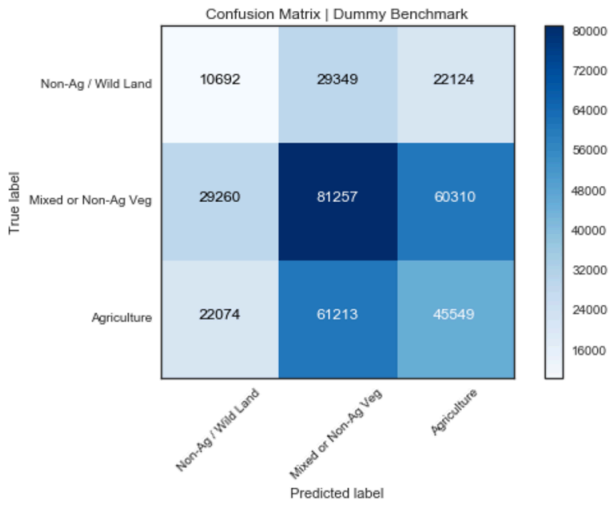
Accuracy:
0.38000928618

Classification Report:

	precision	recall	f1-score	support
1.0	0.17	0.17	0.17	62165
2.0	0.47	0.48	0.47	170827
3.0	0.36	0.35	0.35	128836
avg / total	0.38	0.38	0.38	361828







Libraries & Utilities Used

- Python:
  - Fiona (a GIS lib)
  - Rasterio (a GIS lib)
  - GeoPandas (a GIS lib)
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn
  - Datetime
- Google Earth Engine:
  - Javascript base
  - Landsat 8 32-day EVI Raster Images
  - Landsat 8 32-day RAW Raster Images
- My own functions:
  - ExploreRasterImage
  - CreateMaskArray
  - StoreBandsasArrays
  - KMeansCentroidBasedLabeling
  - ClassifierAccuracyScore

LC-related Articles & Papers

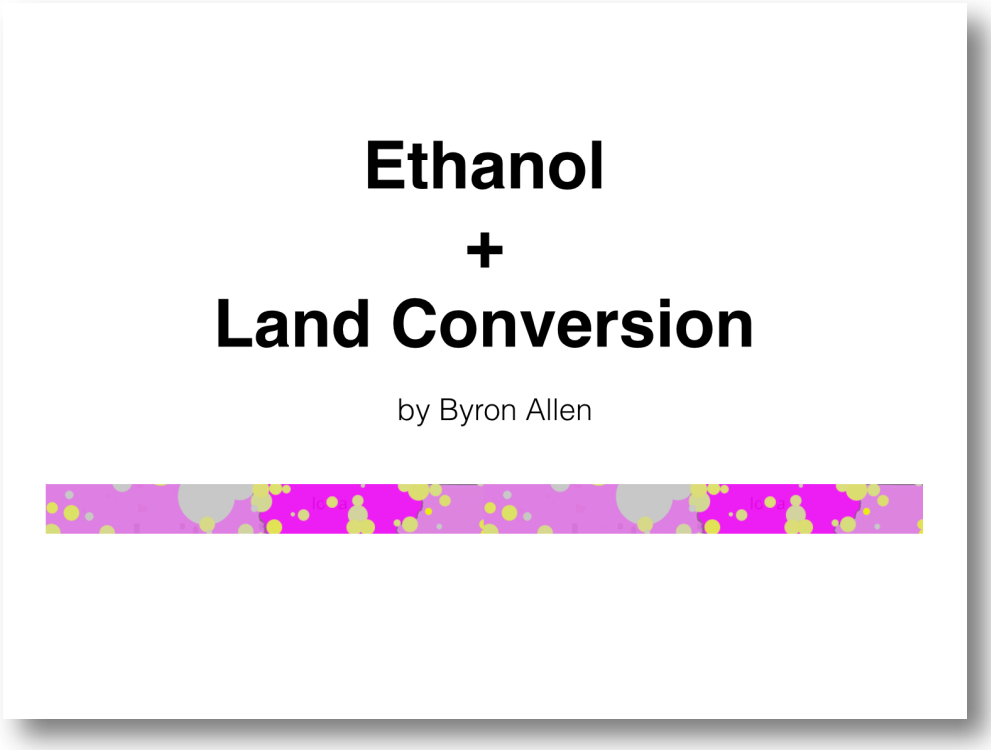
https://www.nwf.org/News-and-Magazines/Media-Center/News-by-Topic/Wildlife/2013/9-18-13-USDA-Data-Grasslands-Forests-Being-Converted-to-Cropland-at-Alarming-Rates.aspx

https://www.fsa.usda.gov/FSA/webapp?area=newsroom&subject=landing&topic=foi-er-fri-dtc

http://nationalaglawcenter.org/wp-content/uploads/assets/crs/RL33950.pdf

http://www.oecd.org/tad/sustainable-agriculture/44807867.pdf

My Previous Project on Land Conversion



PDF link
Video Presentation link
Tableau Dashboard link