Classification of Land Types Through Clustering

Kyle Colton
Thomas Kwak

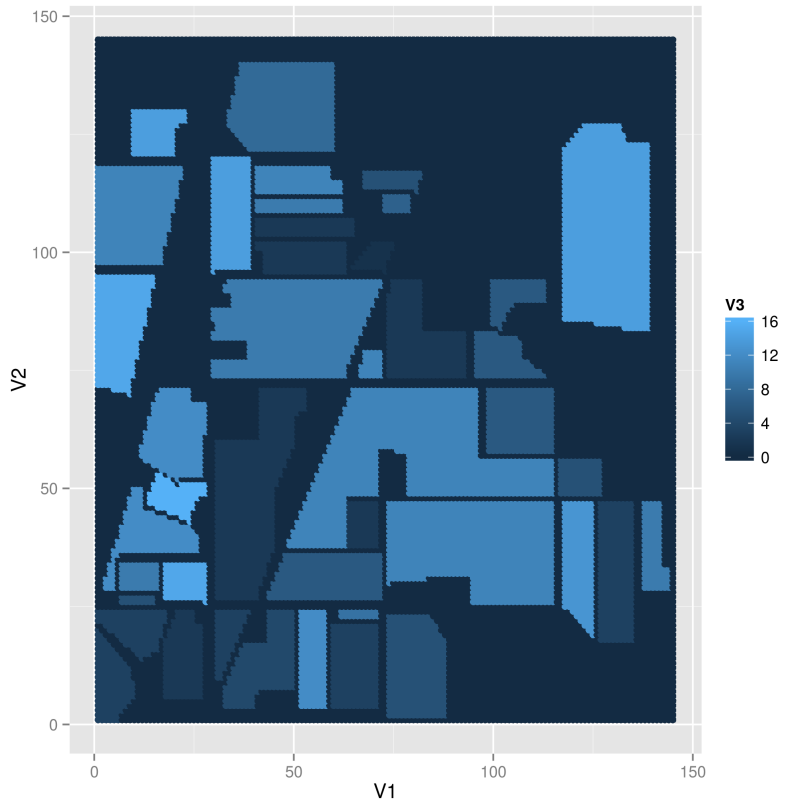
Math 191

June 2, 2015

Data Set

The `indian_pines` dataset consists of hyperspectral earth images of farmland

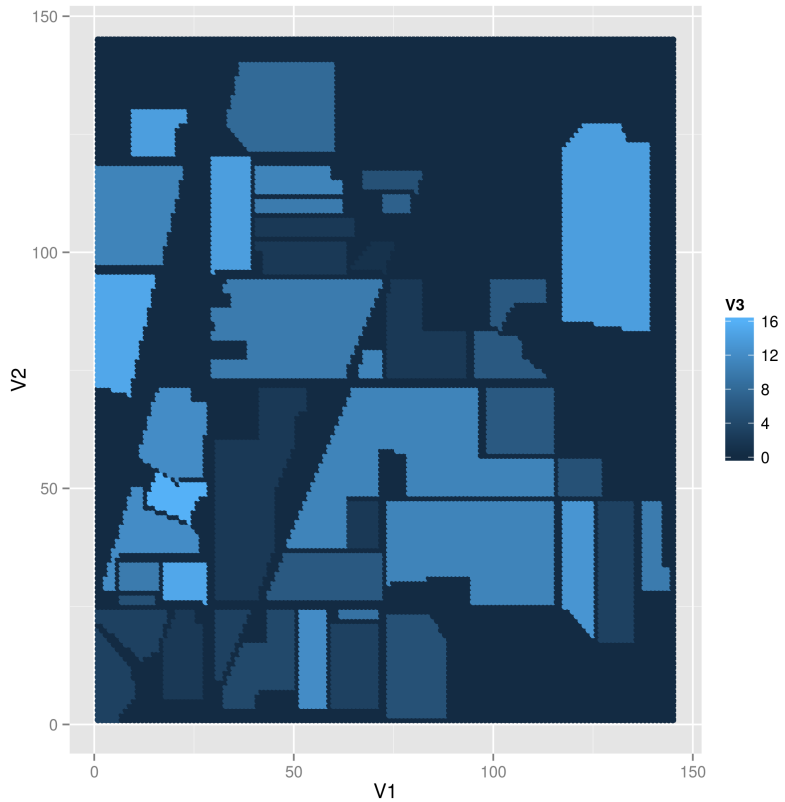
- Dimensions 145x145x200
- Each image is 145x145



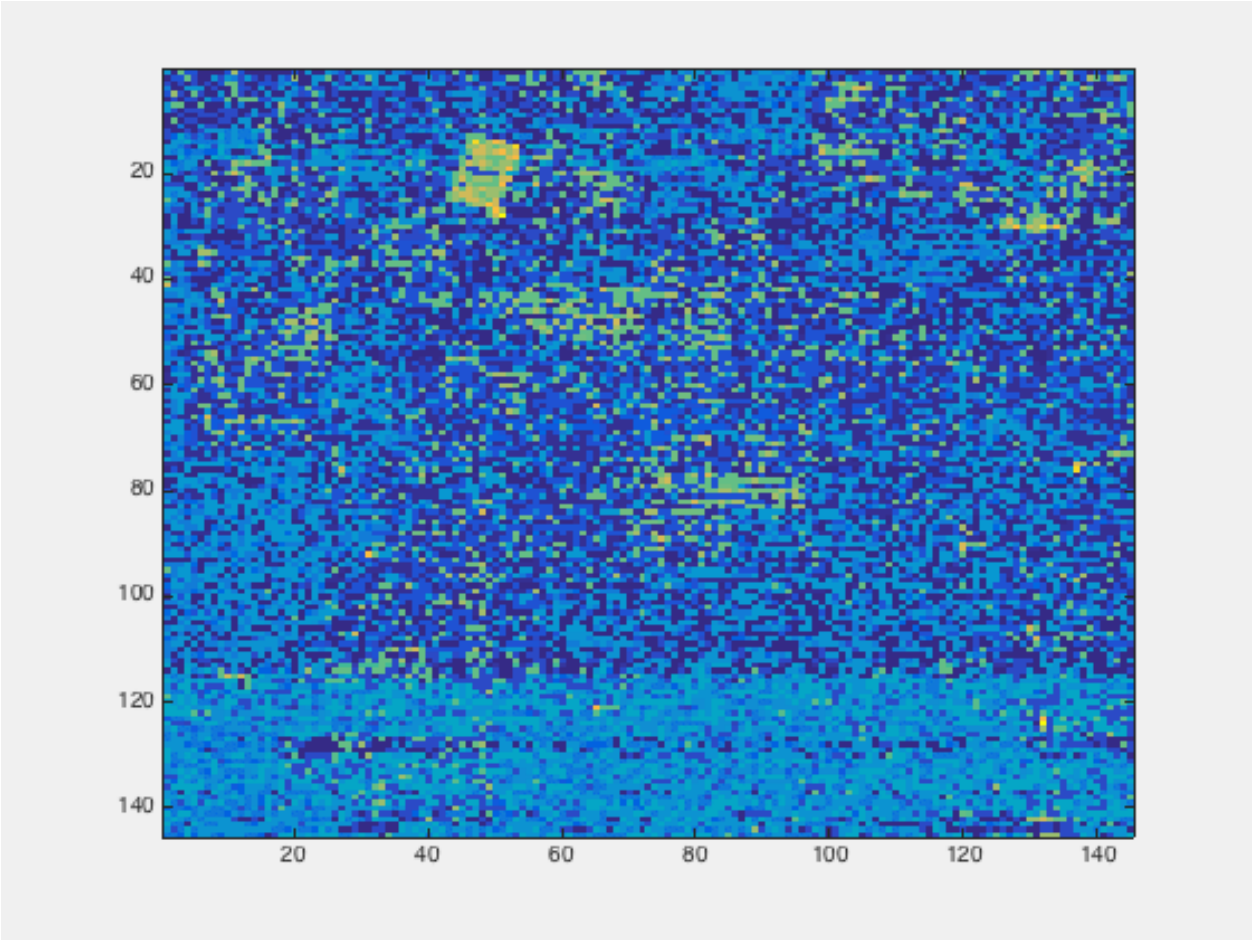
Data Set

The `indian_pines` dataset consists of hyperspectral earth images of farmland

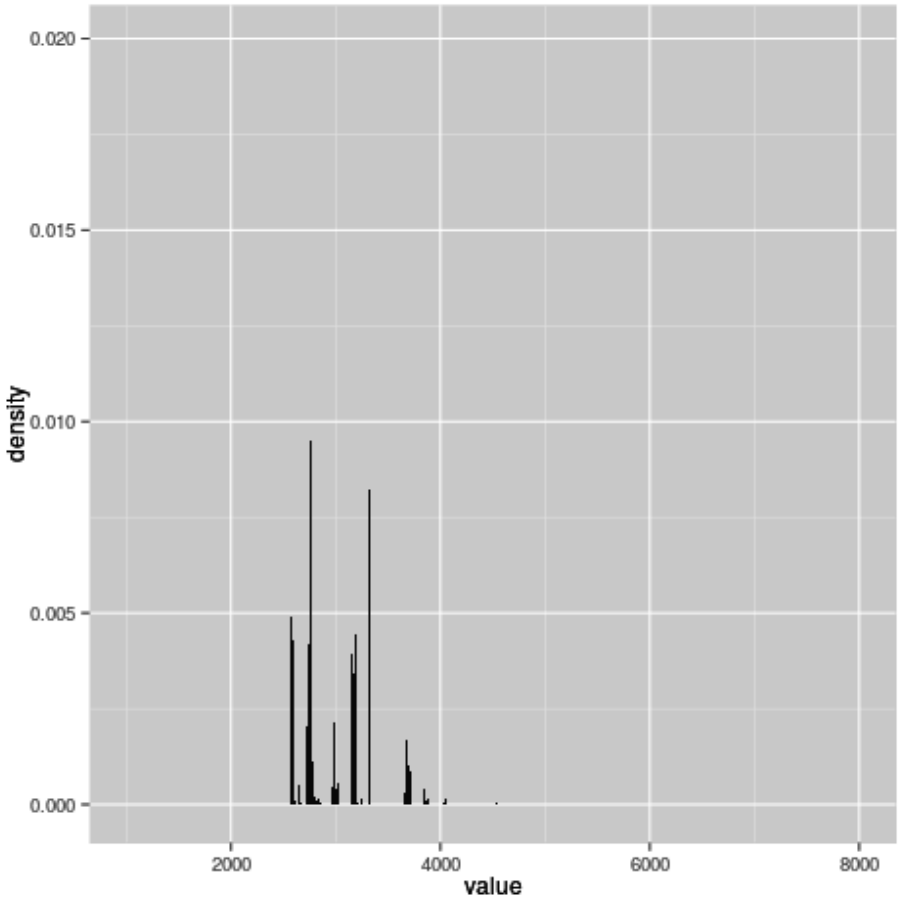
- Images were captured at 200 different wavelengths
- A total of 16 crops were used in the data set



Exploring the Data

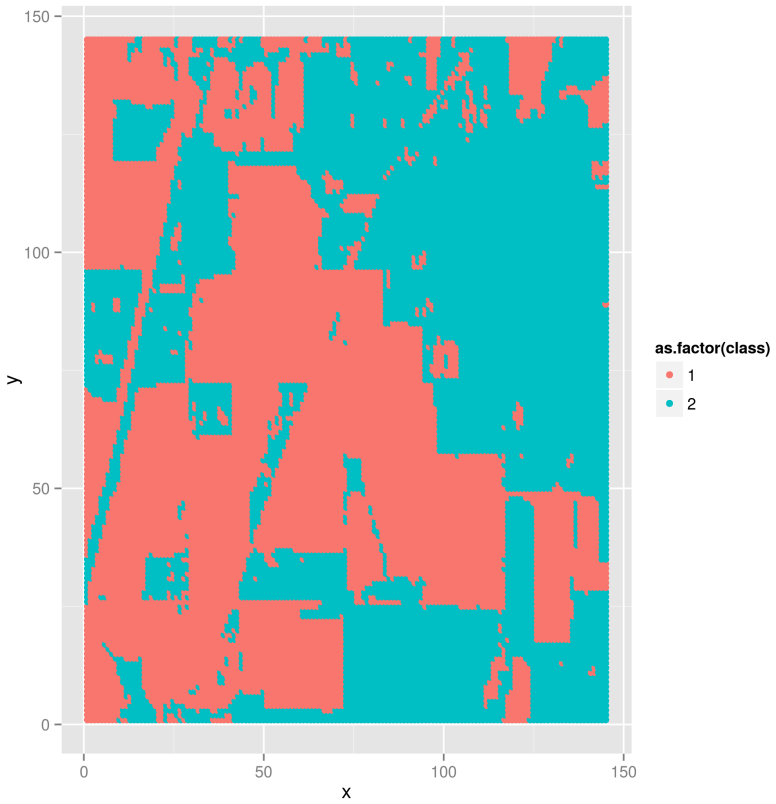


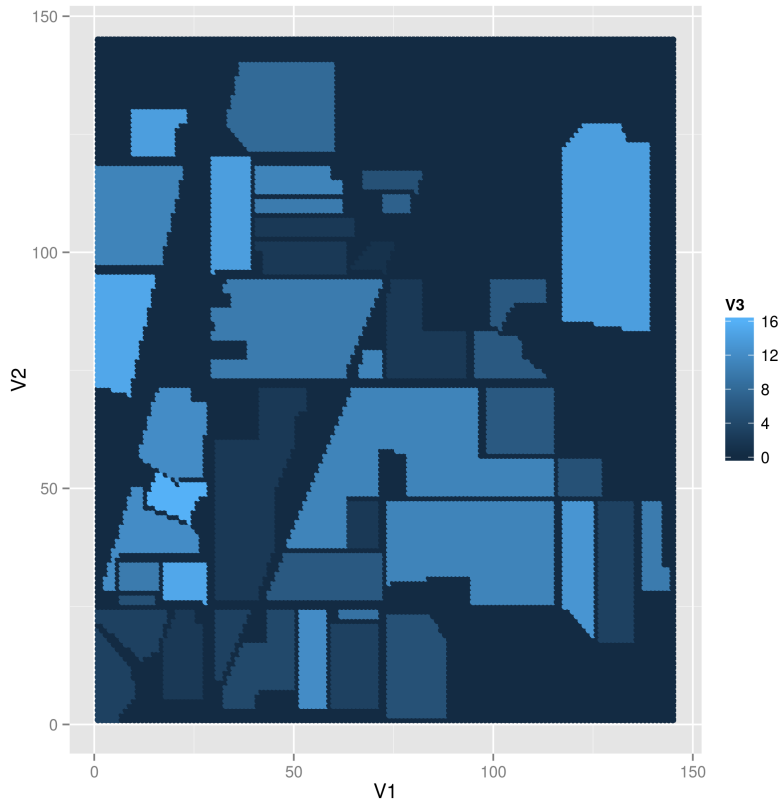
Historgrams



K-Means

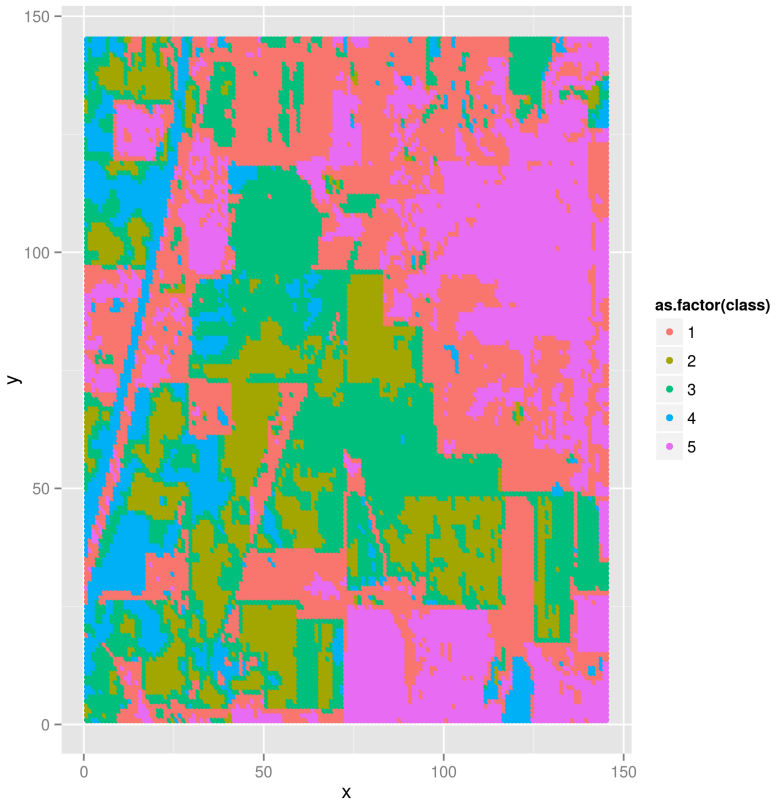
K-Means using 2 classes

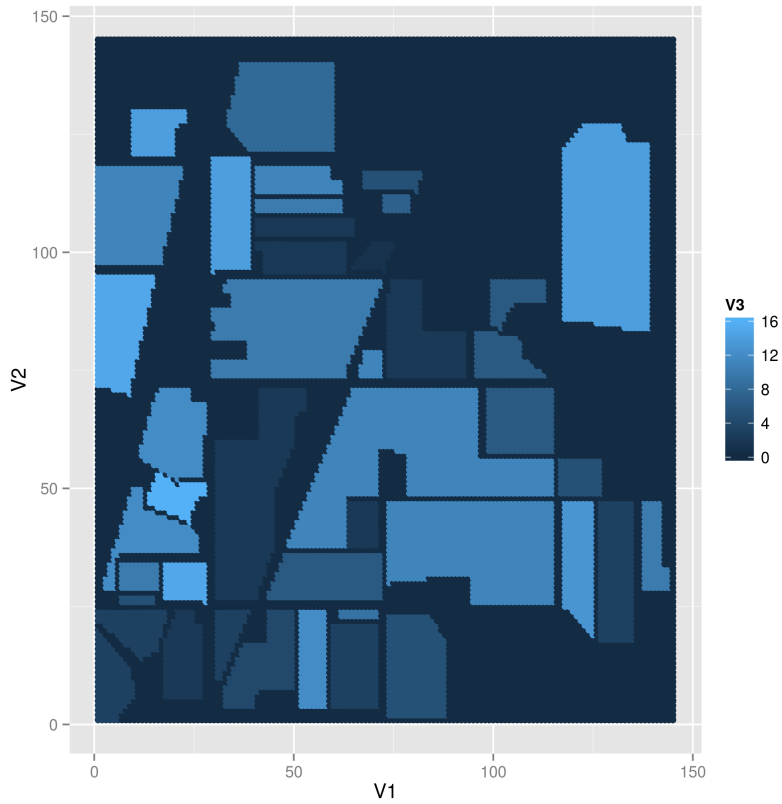




K-Means

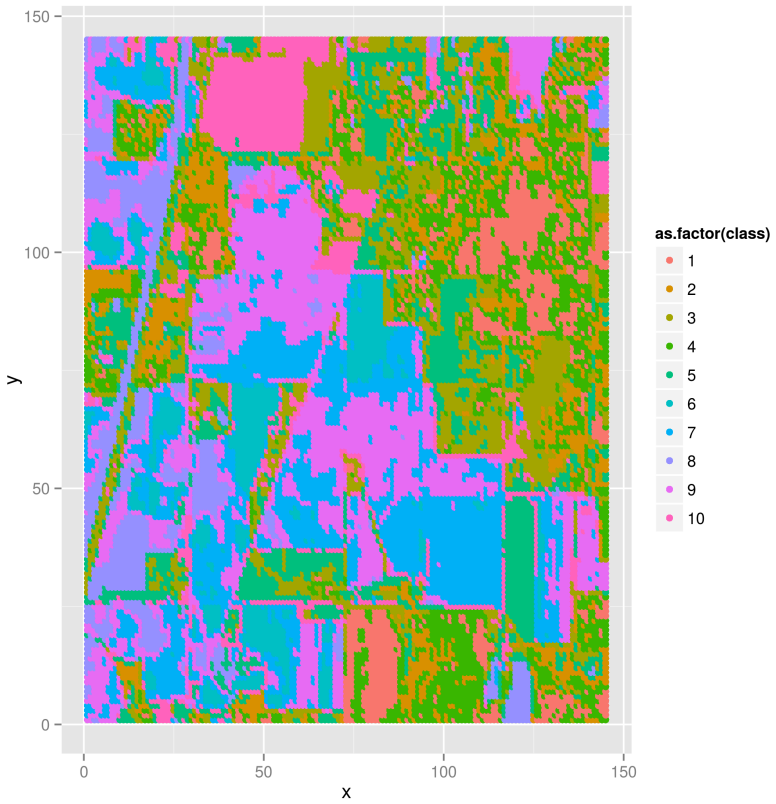
K-Means using 5 classes

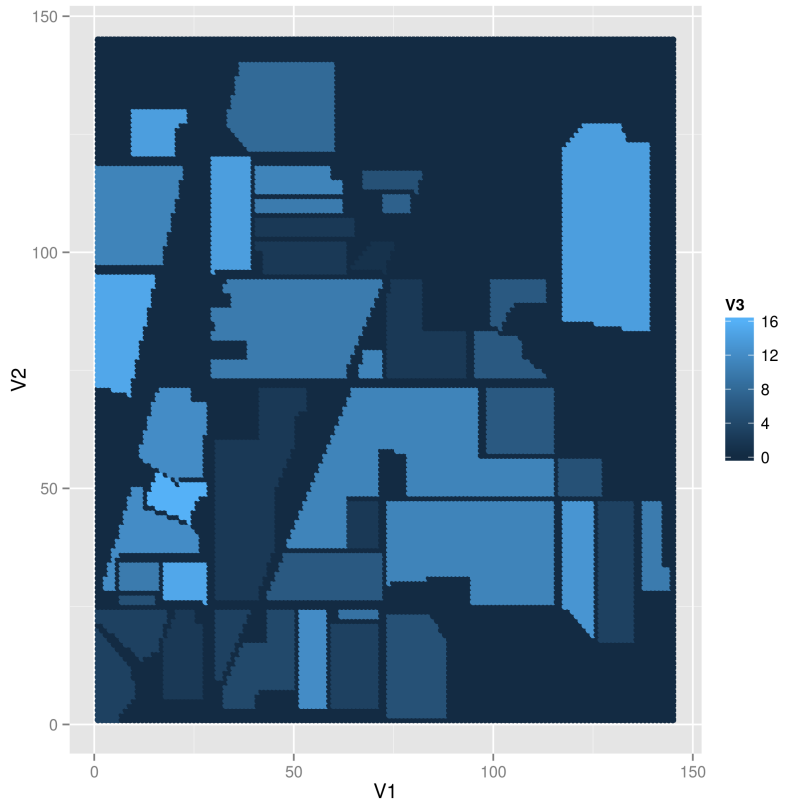




K-Means

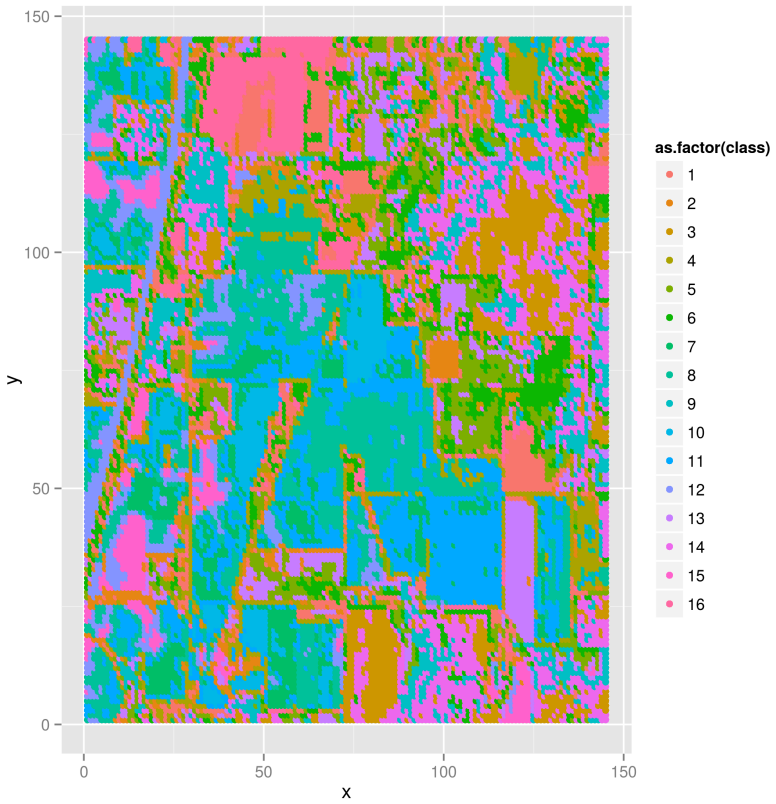
K-Means using 10 classes

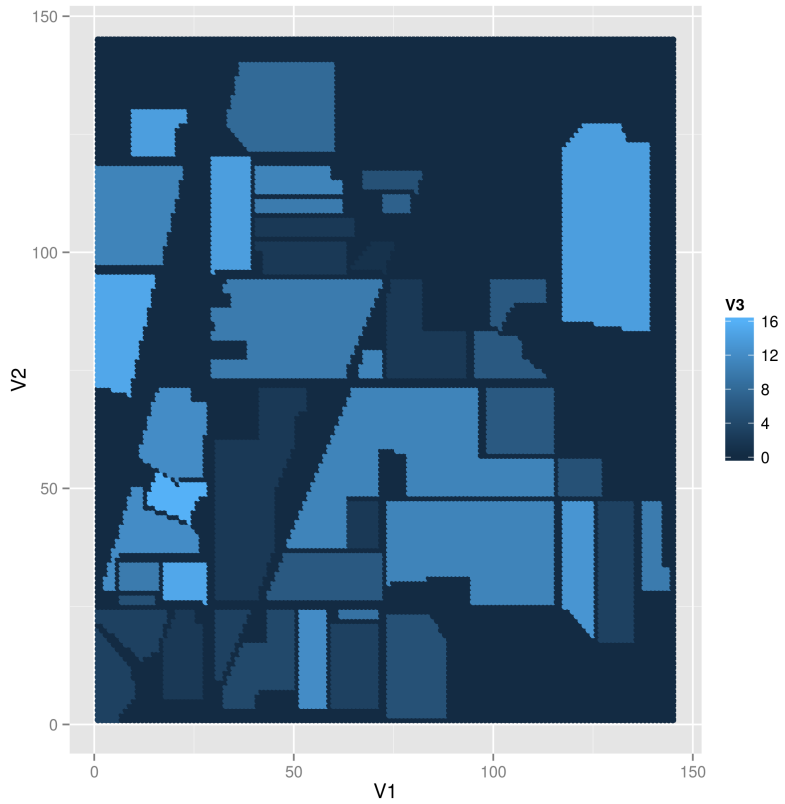




K-Means

K-Means using 16 classes

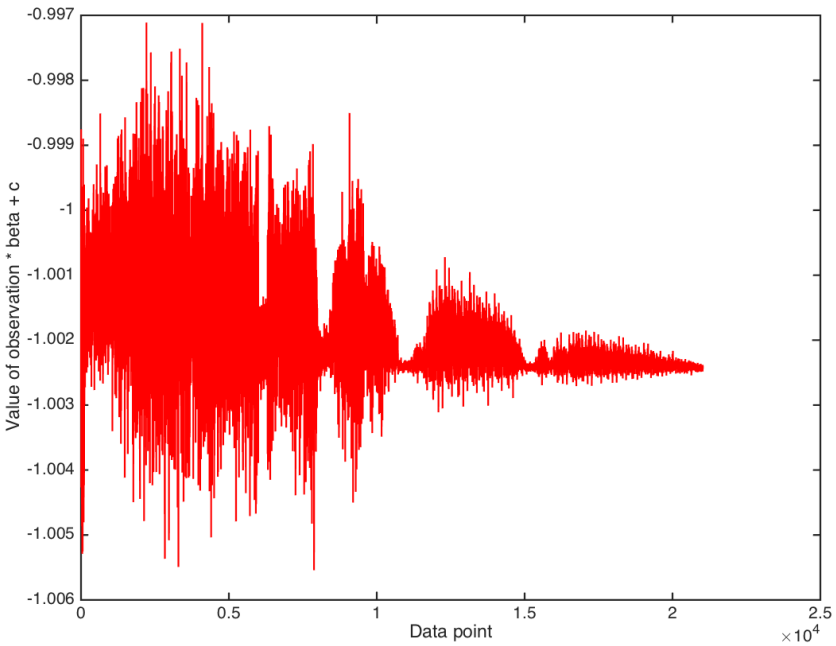




Support Vector Machine

- Separates data into distinct classes
- Constructs a hyperplane between classes
- Soft-margin SVM versus Hard-margin SVM

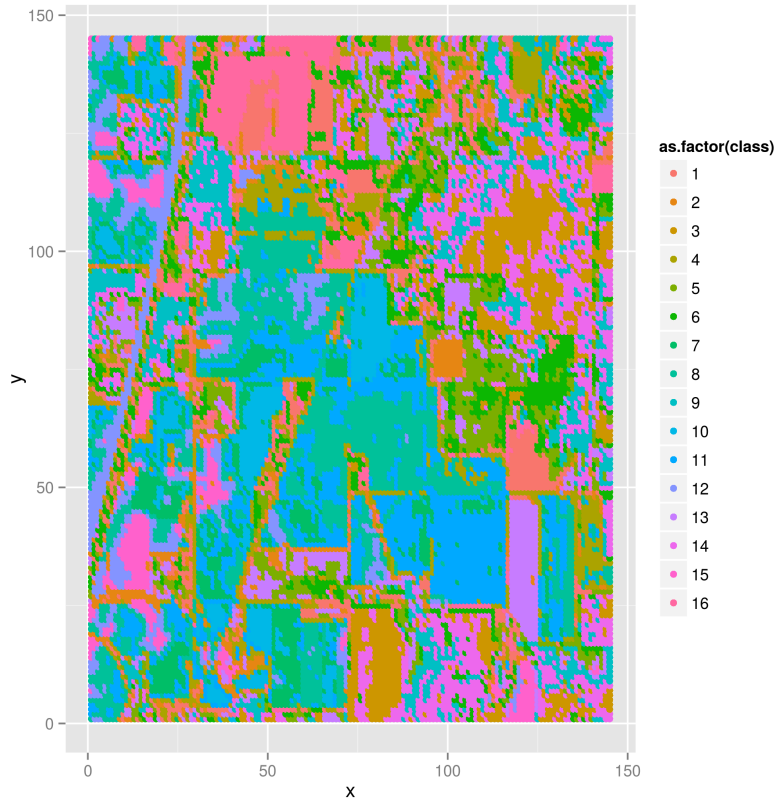
Support Vector Machine



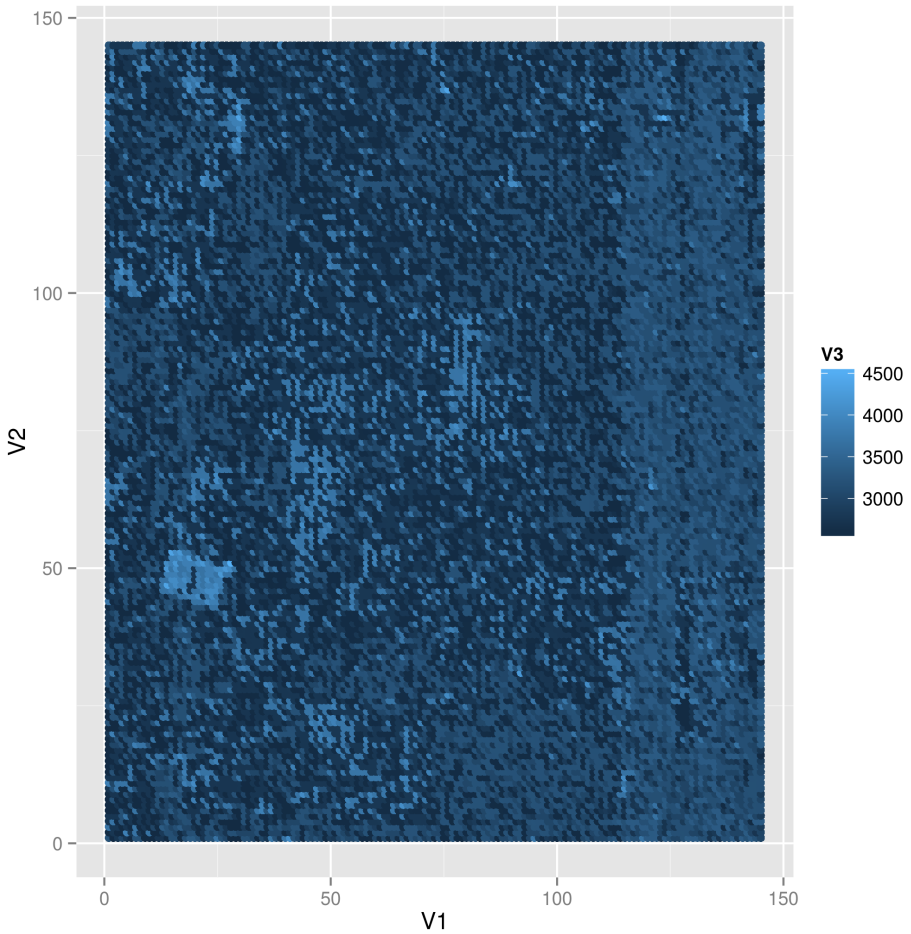
- Reduced data from 3D to 2D by appending rows (increasing the number of columns)
- Arbitrarily picked to separate between corn and not corn
- Soft SVM because of possibility of nonlinear separability and less influence by noise

Results

- We started to see land divisions in K-Means
- A soft SVM would very likely classify better
- K-Means: increase in the number of classes increases the noise level
- Soft-Margin SVM: data cannot be separated linearly. Multiclass SVM would provide better results

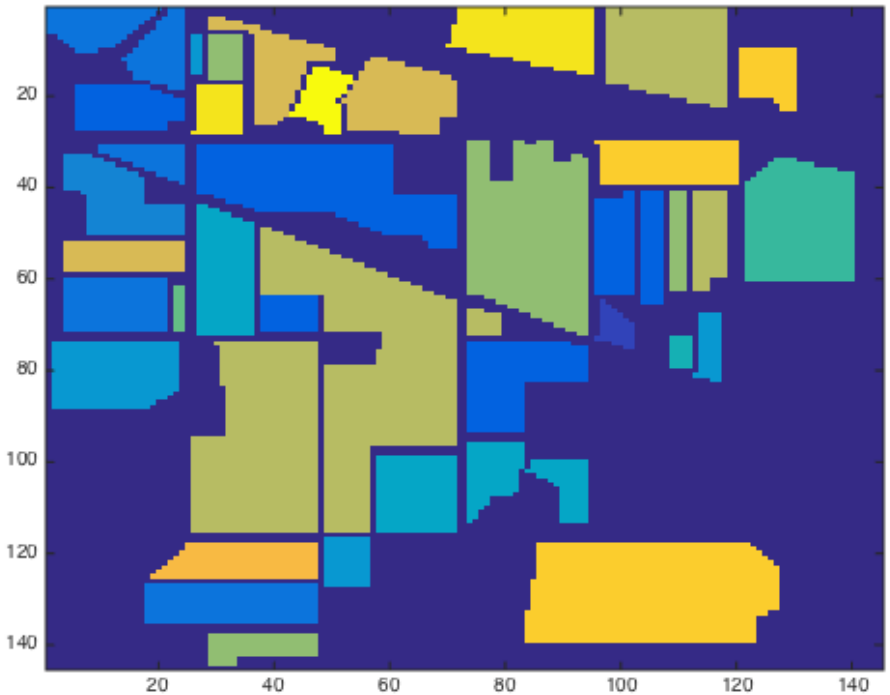


Future Ideas



- Some of the data appears noisy
- This could negatively impact classification
- Detecting noise and biasing the model towards "clean" data could help the prediction

Future Ideas



- Implement a multiclass SVM function
- Completely split the data into 16 classes instead of 2
- Increase accuracy of data separation