

# SMACC

## SDSS MOC4 Asteroid Color Classification

Ben Montgomery

University of Southern Maine

June 8, 2019



## From Last Time...

- The Sloan Digital Sky Survey is a multi-spectral survey of celestial objects, most notably of asteroids.
- Records the UGRIZ color bands.
- Has 471,569 observations in it.
- Assorted people thought there were 2, 4, and 16 distinct asteroid compositions in the SDSS.
- Carvano [1] made that 16 class claim, so I'm trying to check his results using unsupervised learning.



# Methods

- Do the same data corrections that Carvano did.



# Methods

- Do the same data corrections that Carvano did.
- Apply the K-Means, K-Medoids, HDBSCAN\*, and Agglomerative unsupervised clustering algorithms to come up with a classification scheme.



# Methods

- Do the same data corrections that Carvano did.
- Apply the K-Means, K-Medoids, HDBSCAN\*, and Agglomerative unsupervised clustering algorithms to come up with a classification scheme.
- Determine how good the classifications are.



# Methods

- Do the same data corrections that Carvano did.
- Apply the K-Means, K-Medoids, HDBSCAN\*, and Agglomerative unsupervised clustering algorithms to come up with a classification scheme.
- Determine how good the classifications are.
- ???
- Profit.



# Data Correction

Carvano's corrections were limited to removing solar contribution from the color data. in an error-correcting manner. For the  $i^{th}$  band  $\lambda_i$  in UGRIZ, the reflectance color is calculated as

$$C_{\lambda_i} = -2.5 (\log_{10} R_{\lambda_i} + \log_{10} R_{\lambda_{ref}})$$

in order to produce the color reflectance gradient

$$\gamma_j = -0.4 \frac{C_{\lambda_{i+1}} - C_{\lambda_i}}{\lambda_{i+1} - \lambda_i}$$

First downside: only four columns to make predictions from.



# Data Correction

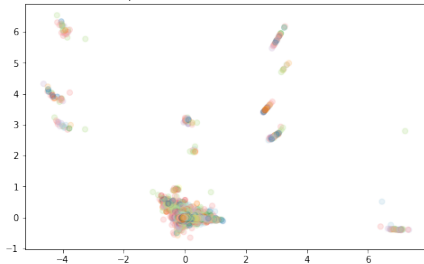
There is all kinds of error in the SDSS, such as phase reddening and zenith-induced error. It is expected that not correcting for these factors will produce significantly errant results [2, 3].



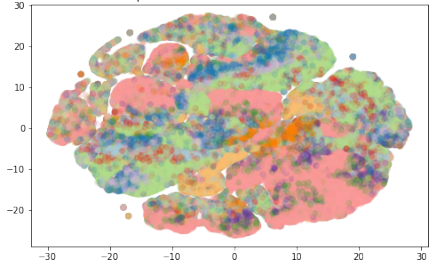


# Visualizing Carvano's Classifications

PCA Representation of the Carvano Classifications



TSNE Representation of the Carvano Classifications

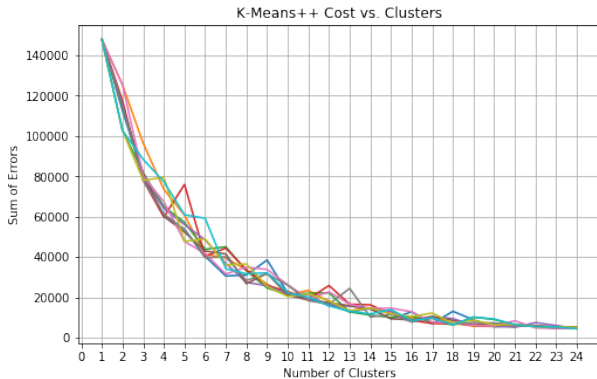


# K-Means++

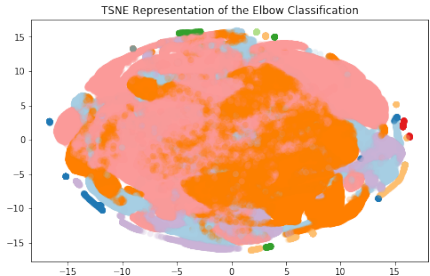
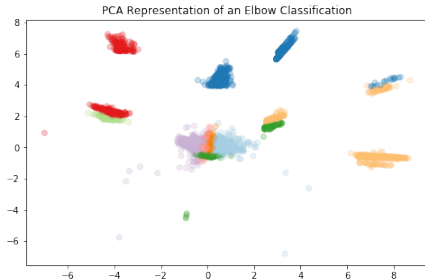
---



# Applying K-Means++



# Applying K-Means++

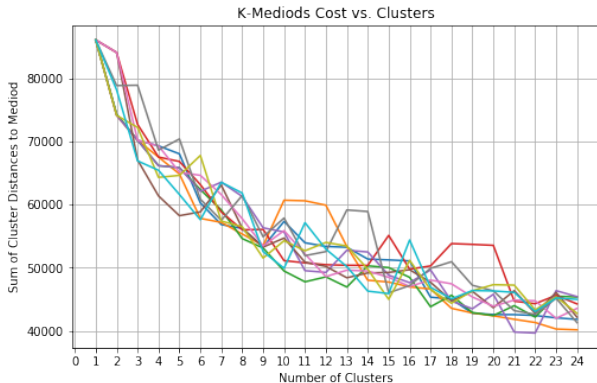


# K-Medoids

---

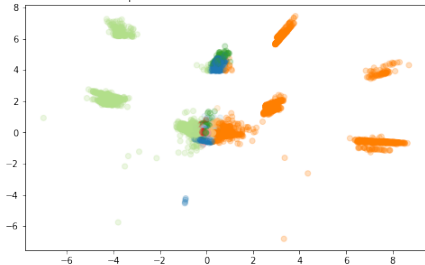


# Applying K-Medoids

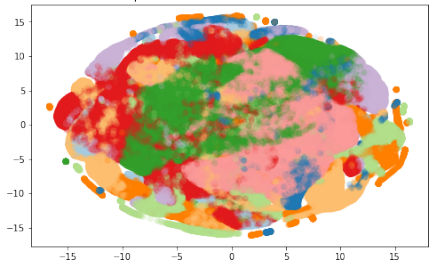


# Applying K-Medoids

PCA Representation of an K-Medoids Classification



TSNE Representation of the K Medoids Classification



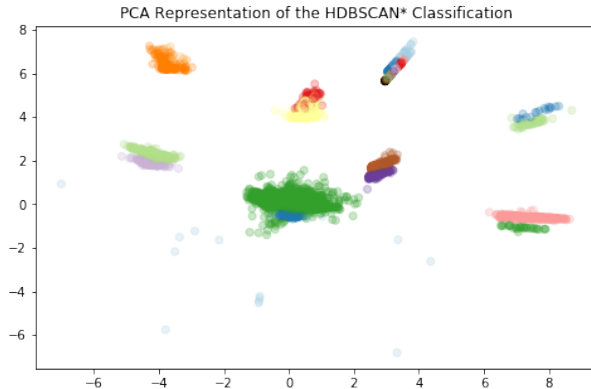
# HDBSCAN\*

---





# Applying HDBSCAN\*

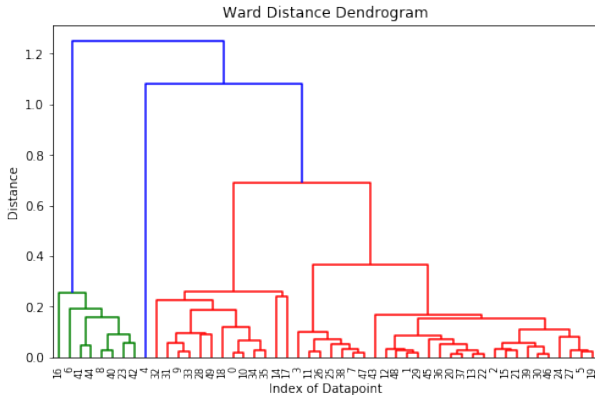


# Agglomerative Clustering

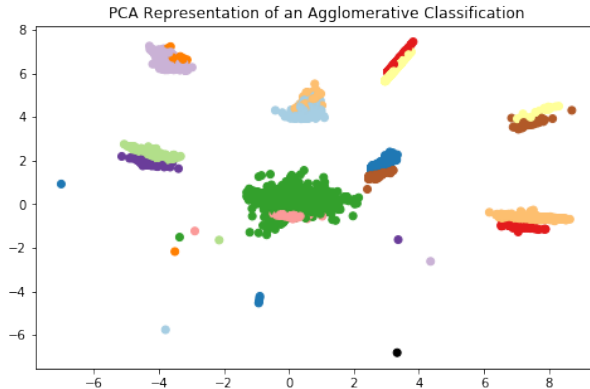
---



# Applying Agglomerative Clustering



# Applying Agglomerative Clustering



## How Good are the Results?

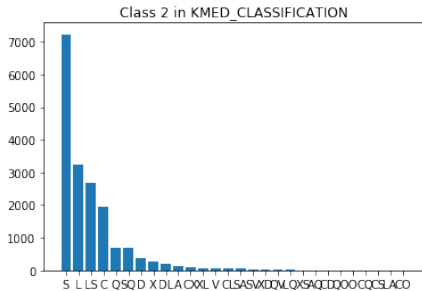
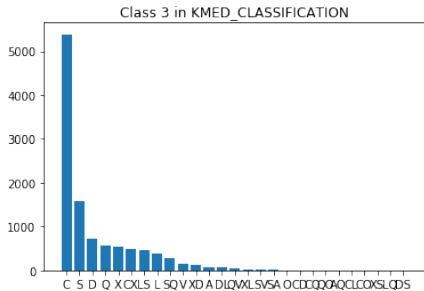
For a given asteroid  $a$  out of  $n$  total, having  $m_a$  observations, and the number of class occurrences  $c1_a, c2_a \dots cn_a$  for a given classification  $cn$ , consistency will be defined as:

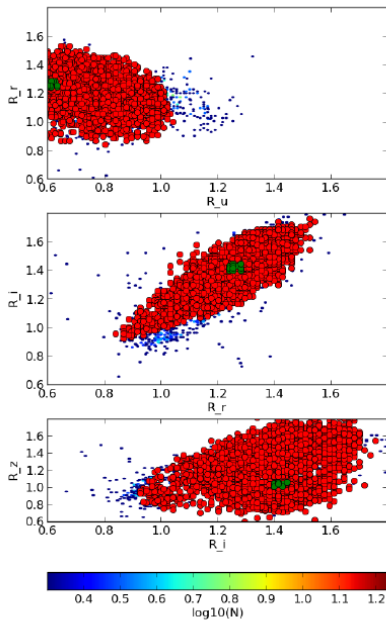
$$\sum_{a=0}^n \frac{\max(c1_a, c2_a \dots cn_a)}{m_a}$$

	K-Means++	K-Medoids	HDBSCAN*	Agglomerative
Consistency	87.8%	77.3%	97.2%	99.8%
Blind-Guess	60.0%	25.7%	97.2%	99.3%
# Classes	9	9	0 - 3000+ (14)	9






# How Good are the Results?





# References

-  J. M. Carvano, P. H. Hasselmann, D. Lazzaro, and T. Mothé-Diniz, "SDSS-based taxonomic classification and orbital distribution of main belt asteroids," *Astronomy and Physics*, 2009.
-  J. A. S. et al., "Phase reddening on near-earth asteroids: Implications for mineralogical analysis, space weathering and taxonomic classification," *Icarus*, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0019103512001376?via%3Dihub>
-  P. H. H. et al., "Adapted g-mode clustering method applied to asteroid taxonomy," *PROC. OF THE 12th PYTHON IN SCIENCE CONFERENCE*, 2013.

