

# How to Structure a L<sup>A</sup>T<sub>E</sub>X Document

Alex Kiar  
Department of Physics  
Western University  
akiar@uwo.ca

May 27, 2016

## Abstract

Space-based astronomical observatories generate vast quantities of data, and efficient means of analyzing those data are needed. The purpose of this research is to apply machine-learning methods to classification of point sources of light emission in nearby galaxies. An objects light emission over different wavelengths is the key data for classification as it indicates the composition of the object, along with its other physical attributes. The mean-shift, k-means, and minimum spanning tree clustering methods were applied to observations of point sources in the M83 galaxy, to identify objects that emit similar combinations of light over multiple wavelength bands. The data was collected by the Wide Field Camera 3 on the Hubble Space Telescope. To identify which combination of bands was the best at separating different classes of objects, the strength of the clustering was tested using a silhouette score. This metric measures an objects distance from a cluster outside the one it was originally assigned to. The clustering results were also compared with the results of independent classification, to determine if each object was correctly identified. The results of this work will allow astronomers to plan observations that can be used to automatically classify objects in nearby galaxies, leading to a stronger understanding of how stars, and star clusters, form and evolve.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Galactic Astronomy . . . . .	1
1.2	Clustering . . . . .	1
<b>2</b>	<b>Data</b>	<b>1</b>
2.1	WFC3 Survey . . . . .	1
2.2	Preprocessing . . . . .	1
<b>3</b>	<b>Methods</b>	<b>1</b>
3.1	Meanshift . . . . .	1
3.2	KMeans . . . . .	2
<b>4</b>	<b>Results</b>	<b>2</b>

## 1 Introduction

Introduction to galactic astronomy and clustering methods.

### 1.1 Galactic Astronomy

Describe the main features of galaxies.

### 1.2 Clustering

Describe at a high level what clustering is.

## 2 Data

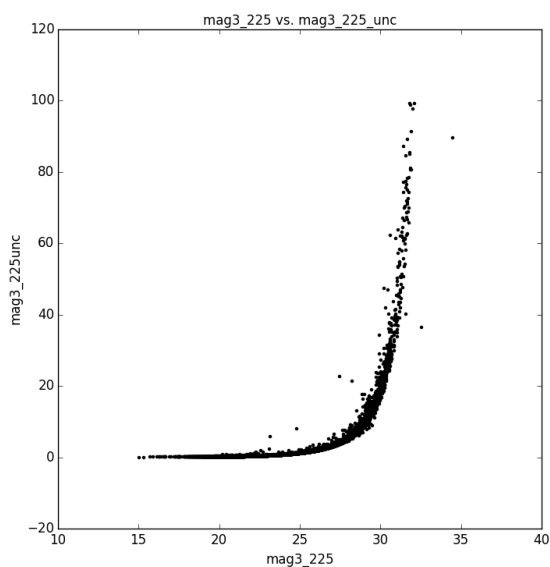
### 2.1 WFC3 Survey

Describe the survey.

### 2.2 Preprocessing

Describe the filters applied to the data.

Figure 1: Uncertainty

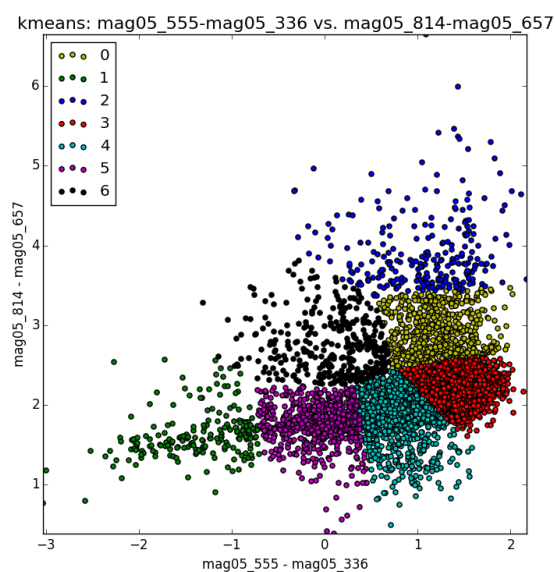


### 3.4 KMeans

Describe kmeans clustering.

Number of Clusters	Bandwidth	Score
3	6	3
4	65	4
5	7	8

Figure 2: Kmeans Clustering



## 3 Methods

### 3.1 Meanshift

Describe meanshift clustering.

### 3.2 DBSCAN

Describe DBscan clustering.

### 3.3 Affinity Propagation

Describe Affinity Propagation clustering.

## 4 Results

What did you learn.

1. Silhouette Score
2. Bandwidth
3. Clustering Variability
4. classification

Figure 3: MeanShift Clustering

