Data Mining

(Don't worry, I am not presenting these slides; just for your reading pleasure)

- There has been an evolution in how we look the relationship between the user and the archive.

- Shortly after HST launch, the user gave simple requests like "give me the output of my observations" and the archive provided data that was relatively unprocessed.

- As time progressed more users looked at the archive as a whole, and were able to query more complex metadata (e.g. give me F218W data of M33 taken between these dates) and the archive provided much more completely reduced data, along with access to the raw data.

- Hubble Source Catalog has extended this trend, allowing users to make queries on the characteristics of objects (e.g. SExtractor outputs) rather than metadata.

- What does the future of this trend look like?

- One could imagine more advanced queries like: "I have this SED; give me all the HSC sources that are consistent with it that have observations in these filters. This requires a better understanding of, e.g., non-detections.

- One could extend these advanced queries to the spectral database: "I have a spectrum, find me more that that are similar to it". We could enable this by applying algorithms on our end (*clustering, unsupervised learning*) to the existing spectral database.

- What does the future of this trend look like? *continued*

- We could provide a richer set of qualities (*feature vectors)* for objects larger than the PSF that would allow the users to better sort for the kinds of extended objects they are interested in.

- We could provide server-side classification services: users provide input to train our algorithms to find certain kinds of sources within the data (*supervised learning*).

- We could provided trainable/teachable image search (*segmentation*) algorithms to find complex sources that are not found by SExtractor, e.g. shocks, shells, filaments, arcs, etc.

- What does the future of this trend look like? *continued*

- Your idea here!