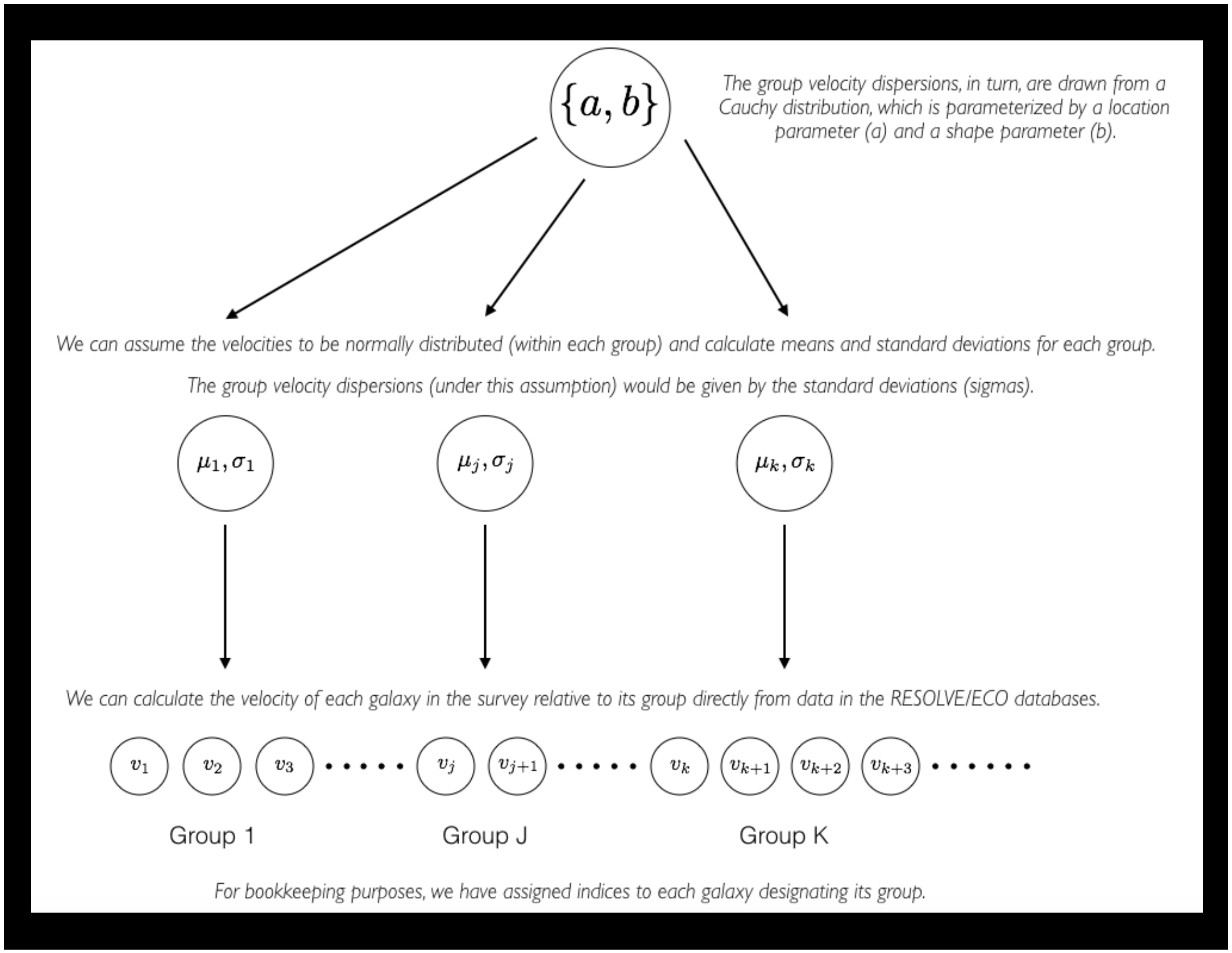
An Excursion into Hierarchical Bayesian Modeling with RESOLVE/ECO Group Velocity Dispersions



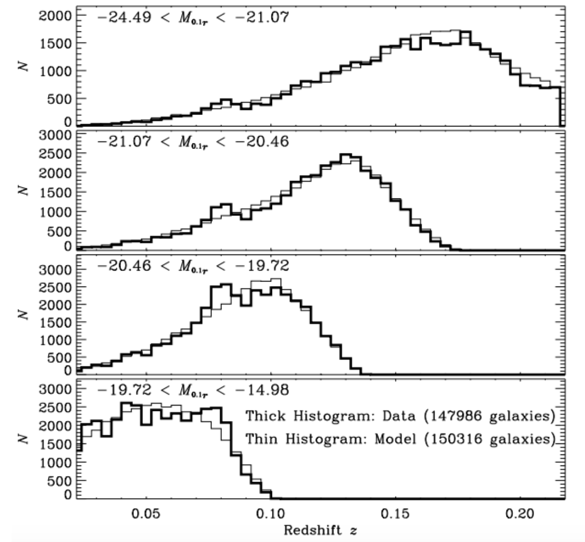
Charlie Bonfield
SAMSI Astrostatistics Course

November 30th, 2016

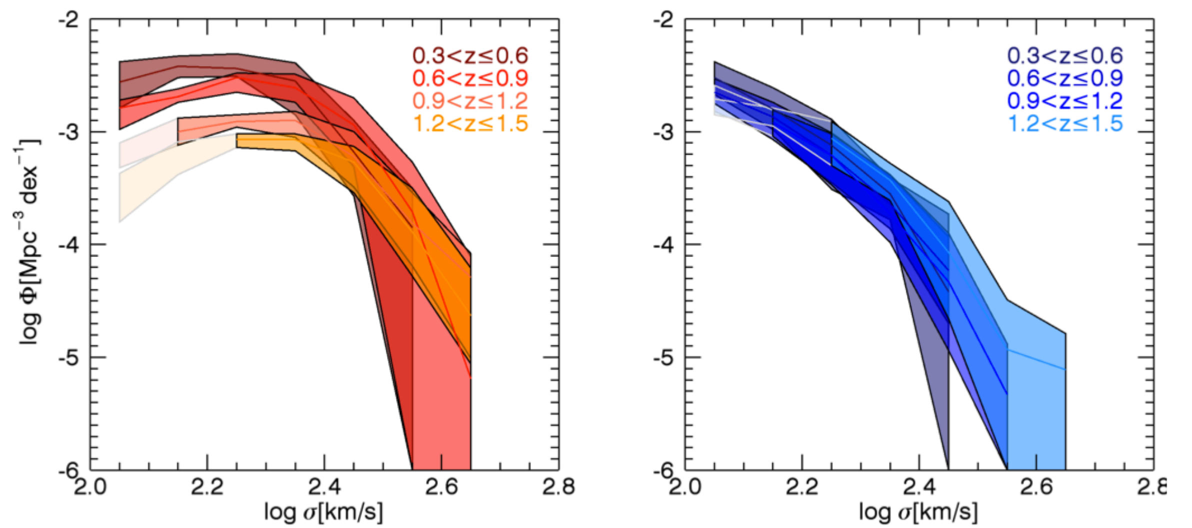
Introduction

- Astronomers like to calculate sets of quantities for collections of objects, bin them up, and throw them on histograms or plots similar to the ones at right.

- The ultimate goal of this enterprise is to make comparisons between members of this population, teasing out physical/astronomical information about the evolution of the population (and by extension, the universe!).



Blanton, M.R. et al., AJ 592, 819 (2003).



Bezanson, R. et al., ApJ 760, 62 (2012).

Velocity Dispersions

- For a group of galaxies, one of the most interesting quantities to measure is the *velocity dispersion*.

- Velocity dispersions crop in an array of different contexts in astronomy, and some believe them to be the best predictor of galaxy properties.

    - Examples: Faber-Jackson relation, M-σ relation, fundamental plane of elliptical galaxies

- For my project, I wanted to explore the distribution of group velocity dispersions for galaxies in RESOLVE/ECO.

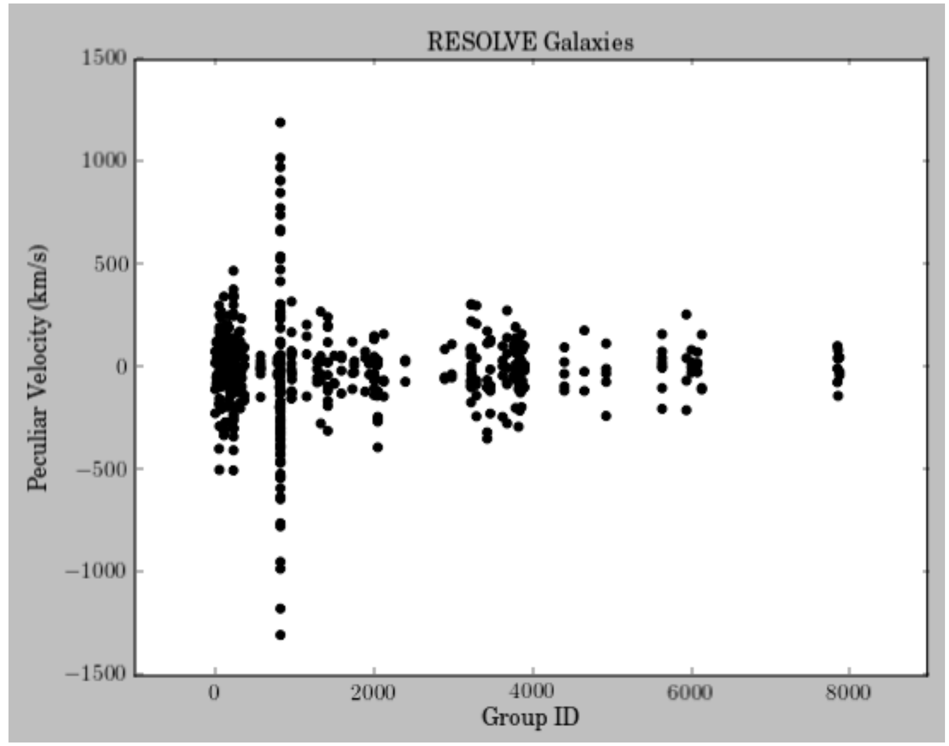
Data

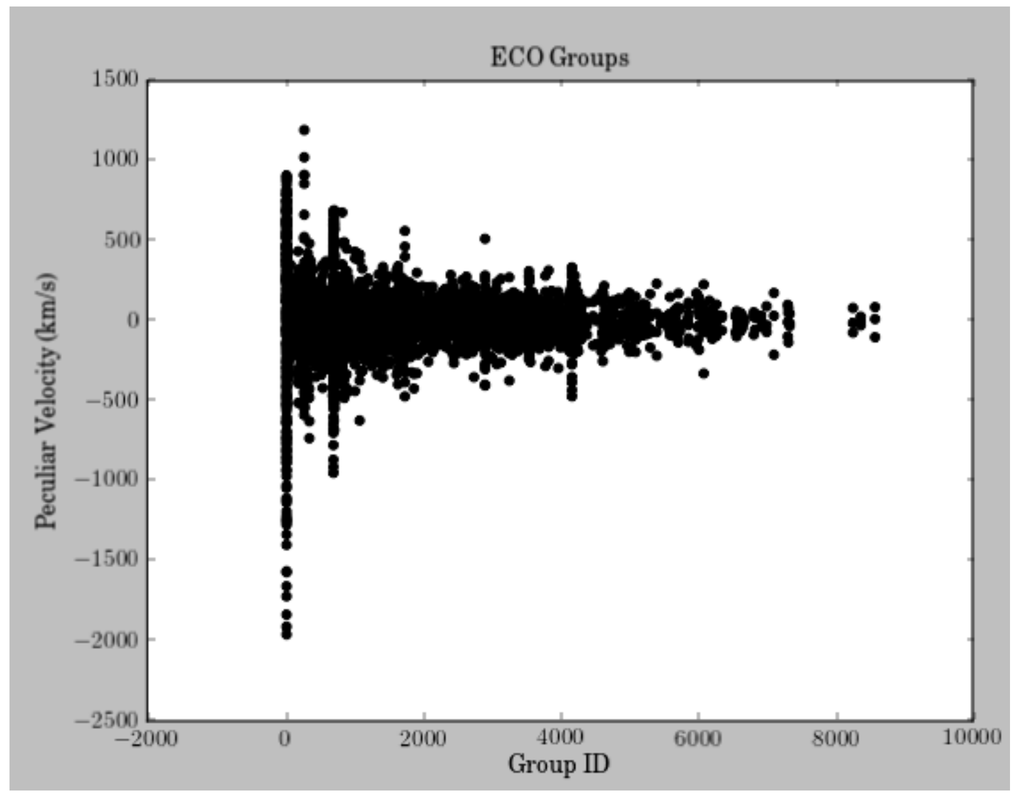
- I will use data that is readily available from the public RESOLVE/ECO databases*.

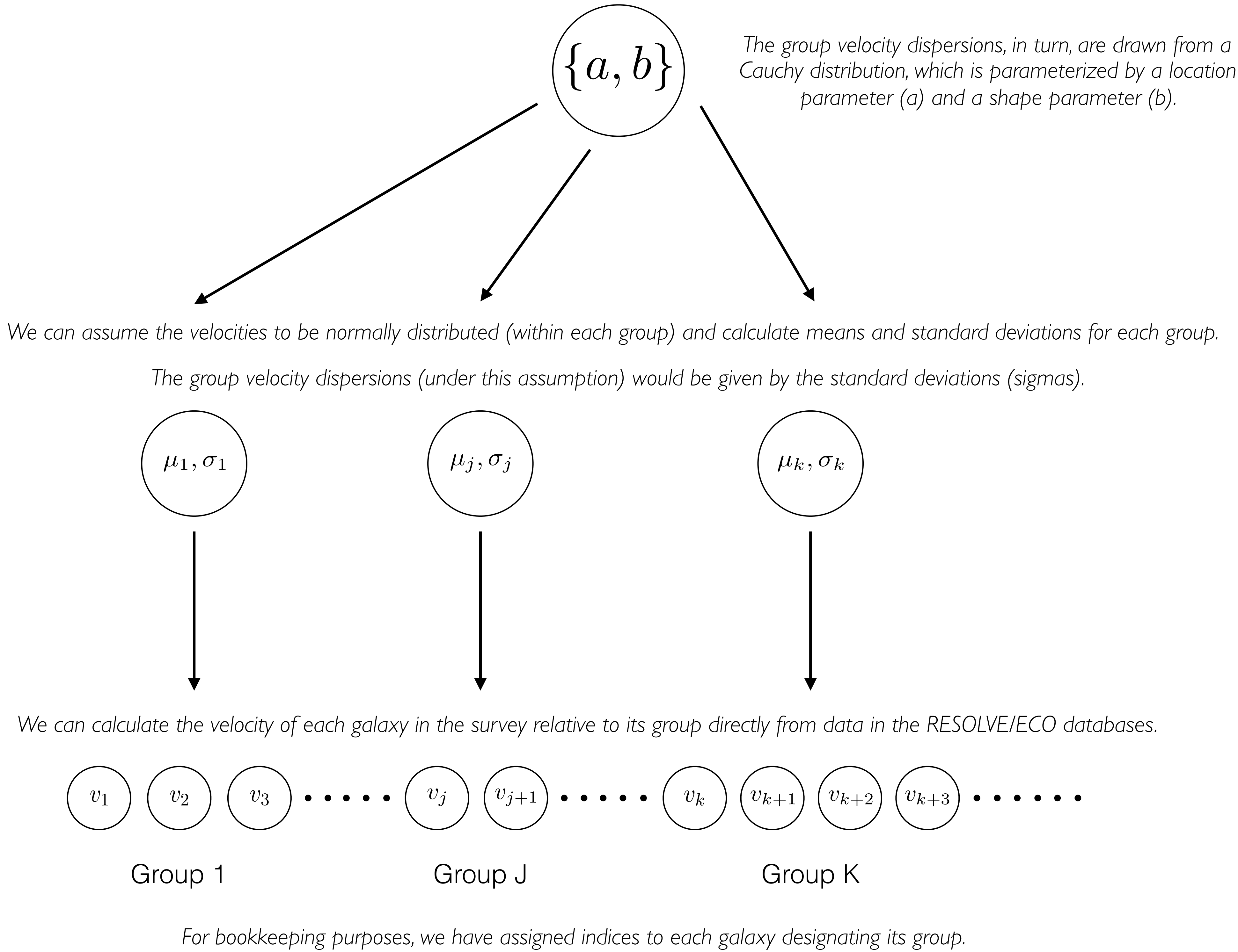
- Since the data contained in these databases is the result of the raw data being run through the code that does all of the heavy lifting, we do not have to do anything with the raw spectroscopy.

- The most important parameters under consideration are the galaxy group ID and galaxy velocity (along our line-of-sight).

*https://resolve.astro.unc.edu/pages/database.php






{a, b}
The group velocity dispersions, in turn, are drawn from a Cauchy distribution, which is parameterized by a location parameter (a) and a shape parameter (b).
We can assume the velocities to be normally distributed (within each group) and calculate means and standard deviations for each group.
The group velocity dispersions (under this assumption) would be given by the standard deviations (sigmas).
μ1, σ1
μj, σj
μk, σk
We can calculate the velocity of each galaxy in the survey relative to its group directly from data in the RESOLVE/ECO databases.
v1
v2
v3
vj
vj+1
vk
vk+1
vk+2
vk+3
Group 1
Group J
Group K
For bookkeeping purposes, we have assigned indices to each galaxy designating its group.

Jupyter Notebook

(Since I am using some data that has yet to be published, I cannot post this online.)

Known Issues/Questions

(Future Work)

- **Uncertainties**: Without uncertainties on the data, it is hard to quantify how confident we can be in our group velocity dispersions (and consequently, our hyperparameters). (*Answer: be more concerned about group finding algorithm*)

- **Raw Data**: Would there be any benefit to actually starting with the raw data (thereby adding another level to the hierarchy)? (*Answer: probably not!*)

- **Choice of top-level distribution**: Clearly, the choice of model is not sufficient for the data. Is there a better way to parse through the suite of analytic probability distributions if we have a general idea of the "true" form of the underlying distribution? Alternatively, can we construct custom distributions?

- **Notion of a "true" distribution**: How are we to ascertain the appropriateness/goodness of a given model? (This question sort of ties in with the previous one, and may be more astrophysical in nature.)

- **Choice of PyStan**: PyStan seemed to be the best option for HBM, but could this task be accomplished in a much more efficient/simpler fashion?

Known Issues/Questions

(Future Work)

- **Uncertainties**: Without uncertainties on the data, it is hard to quantify how confident we can be in our group velocity dispersions (and consequently, our hyperparameters). (*Answer: be more concerned about group finding algorithm*)

- **Raw Data**: Would there be any benefit to actually starting with the raw data (thereby adding another level to the hierarchy)? (*Answer: probably not!*)

- **Choice of top-level distribution**: Clearly, the choice of model is not sufficient for the data. Is there a better way to parse through the suite of analytic probability distributions if we have a general idea of the “true” form of the underlying distribution? Alternatively, can we construct custom distributions?

- **Notion of a “true” distribution**: How are we to ascertain the appropriateness/goodness of a given model? (This question sort of ties in with the previous one, and may be more astrophysical in nature.)

- **Choice of PyStan**: PyStan seemed to be the best option for HBM, but could this task be accomplished in a much more efficient/simpler fashion?

Questions?

Complementary Image