

INVESTIGATING TEMPORAL STRAIN DIVERSITY IN HUMAN *E. COLI* POPULATIONS USING PYROPRINTING: A NOVEL STRAIN IDENTIFICATION METHOD

February 25, 2012

Aldrin Montana *Alex Dekhtyar*
Emily Neal *Michael Black Chris Kitts*



Background

Why *E. coli*?

- Found in GI tract in many mammals, birds, and reptiles



- Indicator for fecal contamination
- **Microbial Source Tracking**

CURRENT METHODS FOR STRAIN IDENTIFICATION HAVE PROBLEMS

□ Reproducibility

- ▣ Randomly amplified polymorphic DNA (RAPD)
- ▣ Repetitive extragenic palindromic PCR (rep-PCR)
- ▣ Biochemical Profiling

□ Time Intensive/Laborious

- ▣ Pulsed-field gel electrophoresis (PFGE)
- ▣ Ribotyping

□ Expensive

- ▣ PFGE
- ▣ Multilocus enzyme electrophoresis (MLEE)

□ Sensitivity

- ▣ Antibiotic resistance profiles (ARP)
- ▣ Serotyping

Solution: Pyroprinting!

The Questions

- How do *E. coli* populations vary over time in human individuals?
 - ▣ How many different strains are detected over the total time period?
 - ▣ How frequently are strains detected?
 - ▣ How does population variation differ between individual hosts?

The Hypothesis



- Hosts will carry the same 1-2 dominant strains throughout the study
- Different minor strains will be detected throughout the study

The Method

Sampling

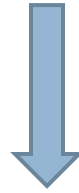
PCR

Pyroprinting

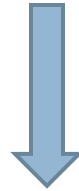
Clustering

Work Flow

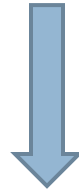
E. coli collection & isolation



Colony PCR



Pyroprinting
(Data Collection)

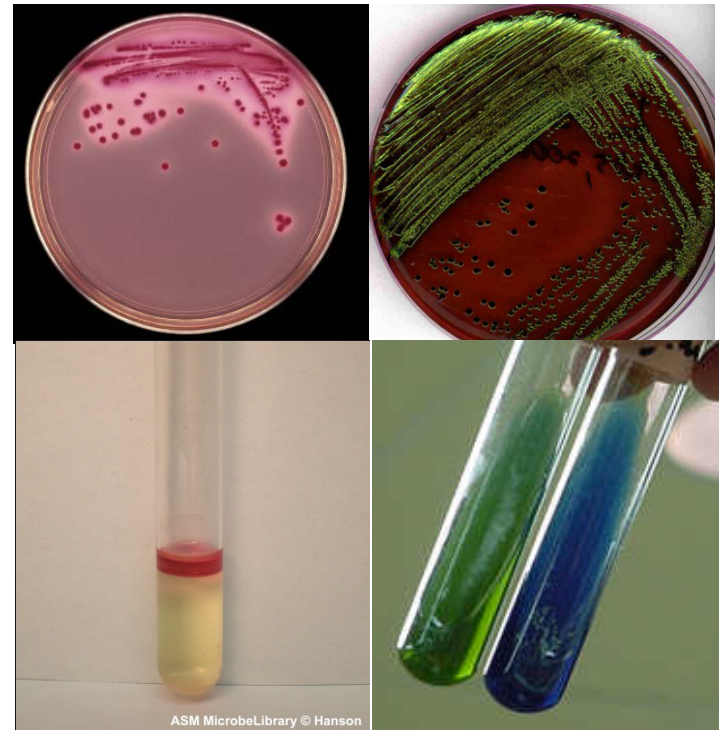


Match pyroprints = build strains

E. coli Collection & Isolation

- Collect & swab fecal samples
 - ▣ 3 individuals
 - ▣ 6 months
 - ▣ 15 isolates (colonies) once a month

- Metabolic confirmation



Colony PCR

Ribosomal RNA Locus

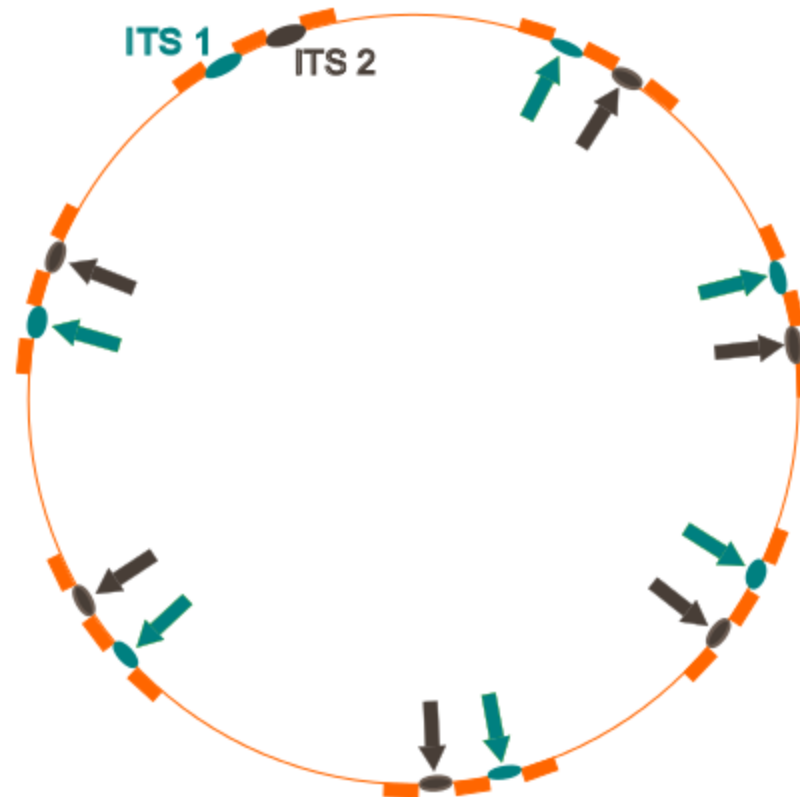


- ITS: Intergenic Transcribed Spacer
 - ▣ Non-coding
 - ▣ Increased accumulation of mutations

Colony PCR

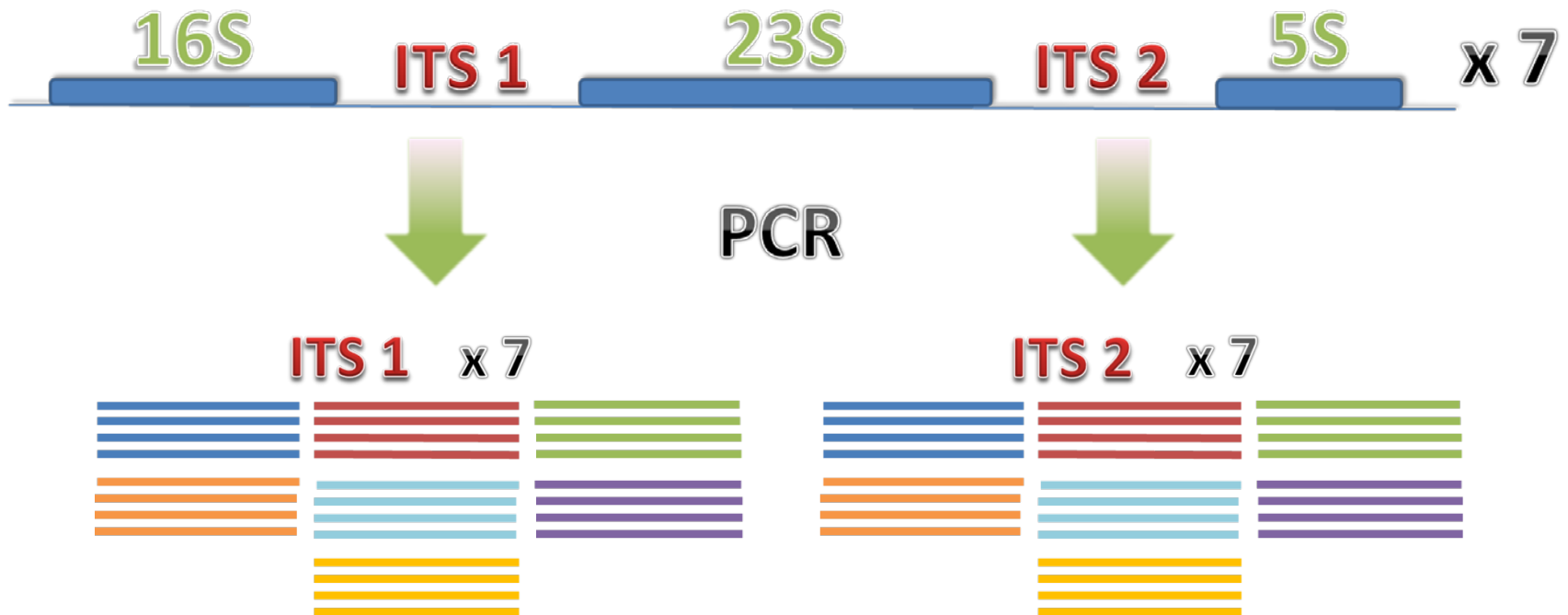
Ribosomal RNA Locus

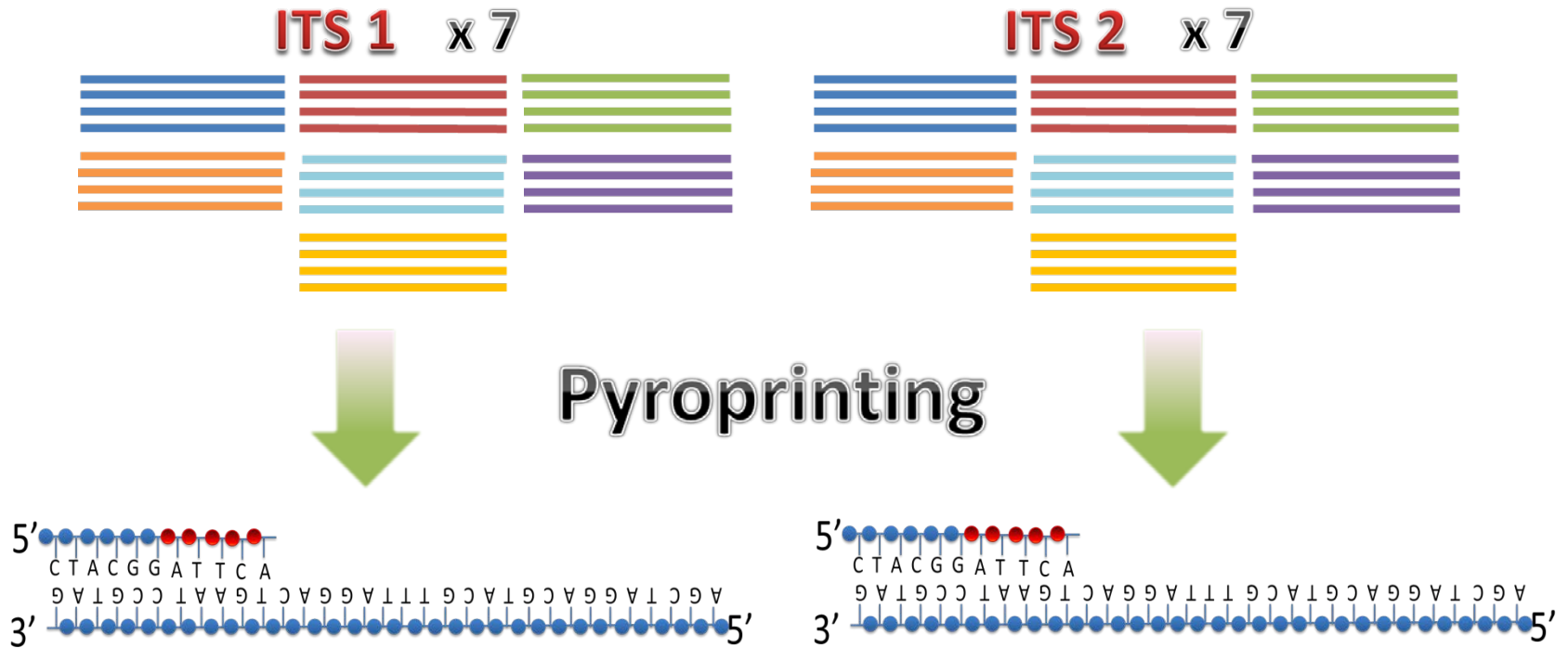
x 7 copies in *E. coli* genome



Colony PCR

Ribosomal RNA Locus



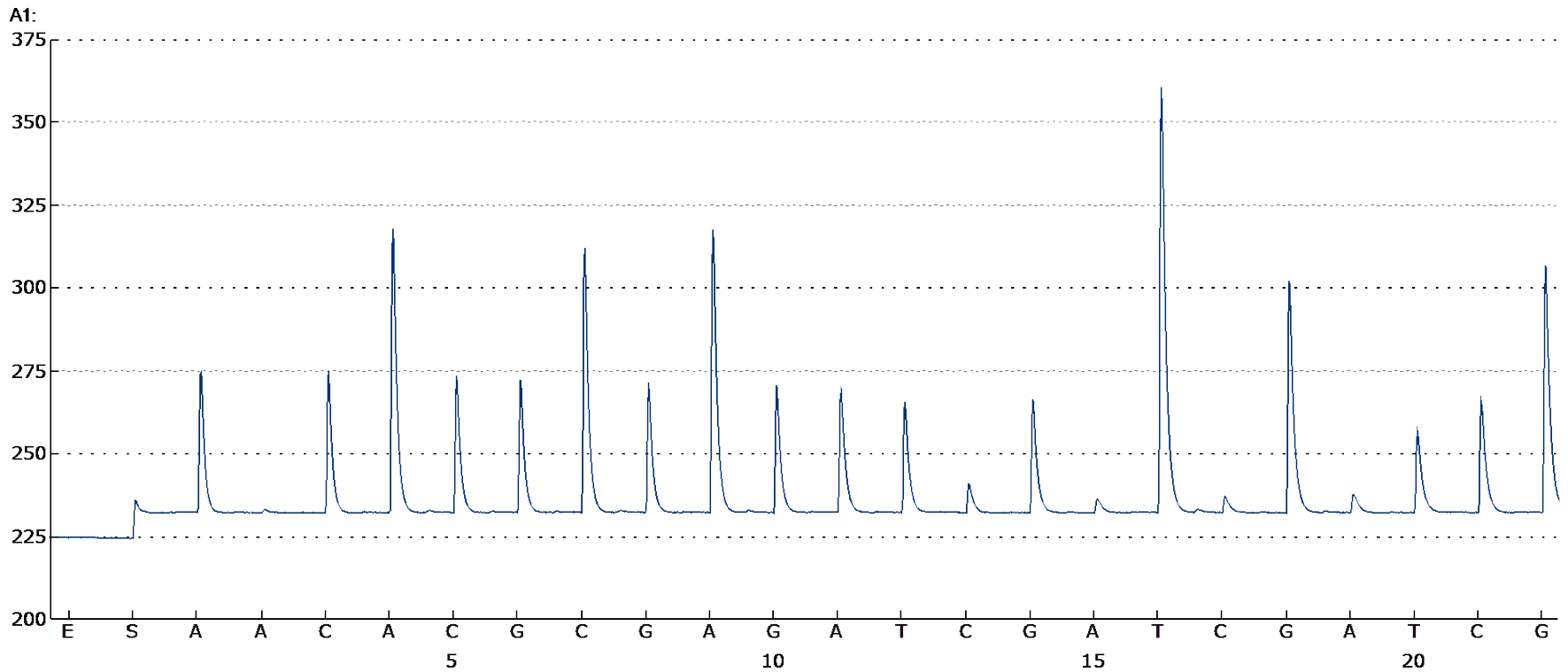


What is a Pyroprint?

- Obtained through pyrosequencing
- Similar to a “fingerprint”
 - ▣ Used for identification—not information
- Creates a unique pattern of peaks based on DNA sequences from all 7 ITS copies



Data Output



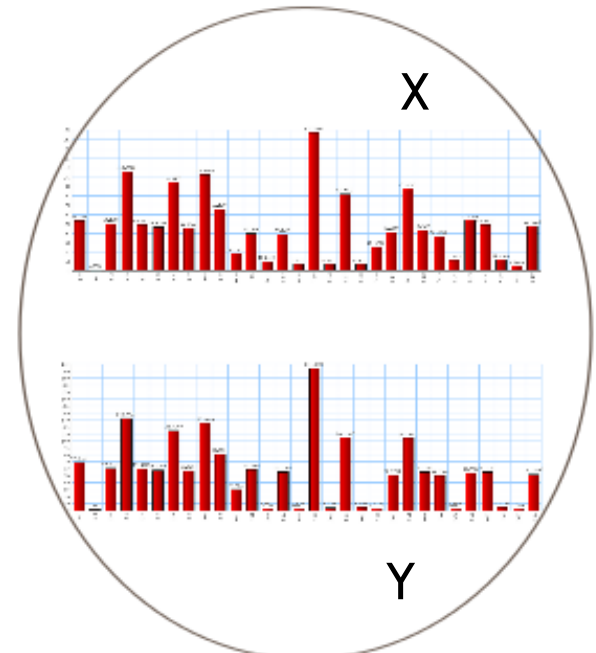
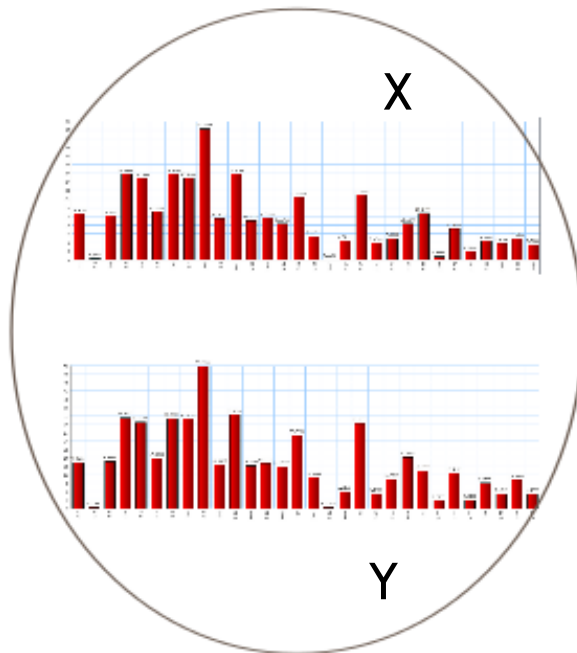
□ Extract Peak Heights

Data Processing

$$\text{Sim}(X, Y) = \text{Pearson}(X, Y)$$

Hu-404

Hu-700



$$M_{\text{ITS1}}[\text{Hu-404}, \text{Hu-700}] = \text{Sim}(X, Y)$$

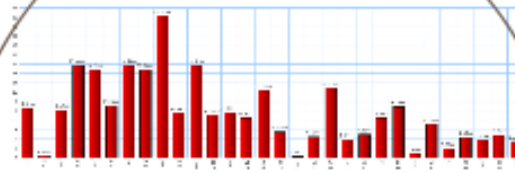
$$M_{\text{ITS2}}[\text{Hu-404}, \text{Hu-700}] = \text{Sim}(X, Y)$$

Data Processing

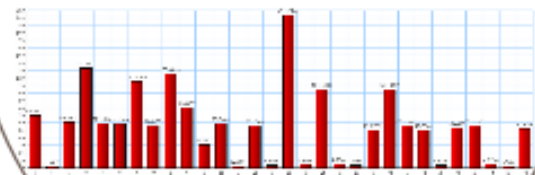
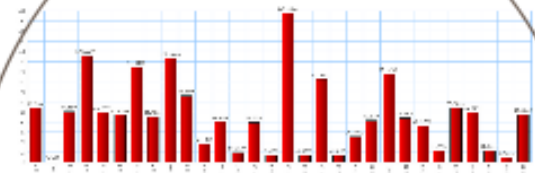
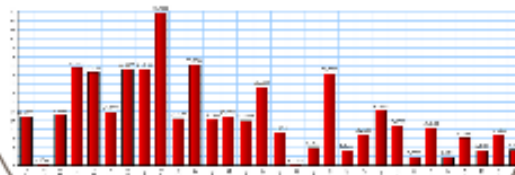
If $\text{thr}(\text{Sim}_{\text{ITS1}}(X, Y)) > \alpha$ and $\text{Sim}_{\text{ITS2}}(X, Y) < \beta$

$$\text{Sim}(\text{Sim}(C_m, C_n)) = \text{Avg}(\text{Sim}_{\text{ITS1}}, \text{Sim}_{\text{ITS2}}) := 0$$

Hu-404



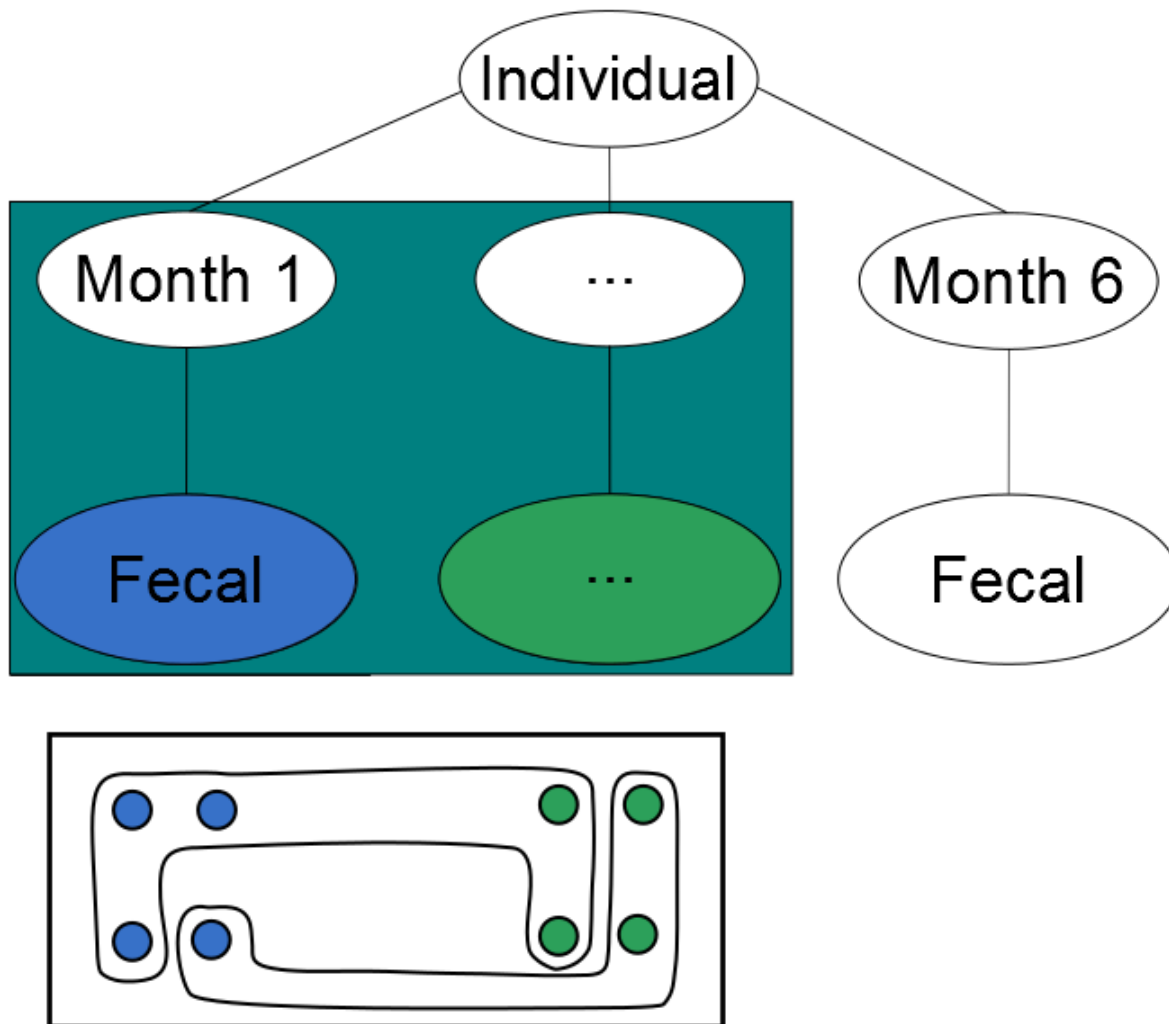
Hu-700



Novel Clustering Algorithm

- Aldrin Montana, Alexander Dekhtyar, Emily Neal, Michael Black, and Chris Kitts. 2011. Chronology-Sensitive Hierarchical Clustering of Pyrosequenced DNA Samples of *E. coli*: A Case Study. In *Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM '11)*. IEEE Computer Society, Washington, DC, USA, 155-159.

Clustering Approach



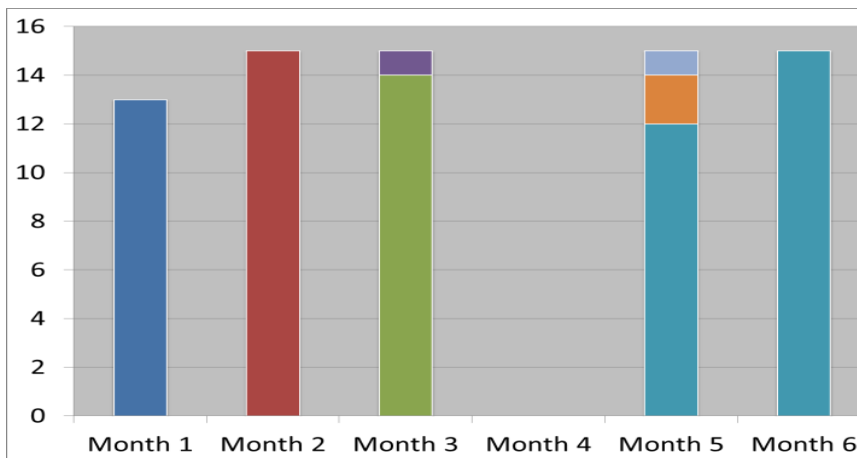


Results



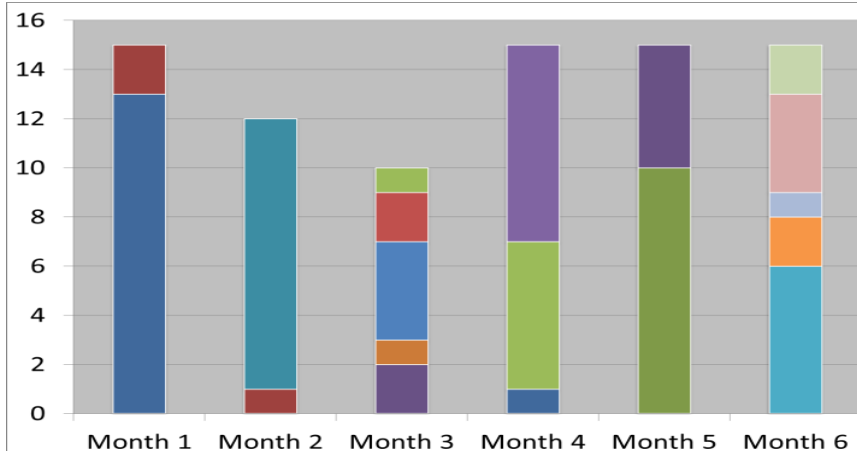
Person A

of
Isolates



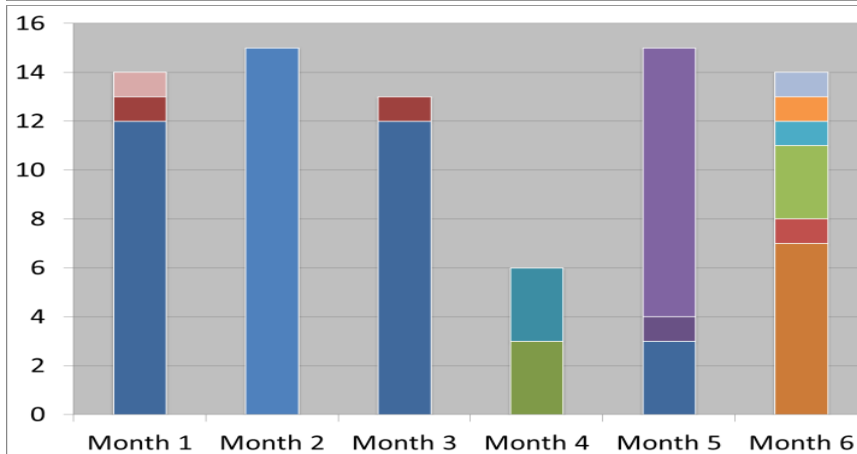
Person B

of
Isolates

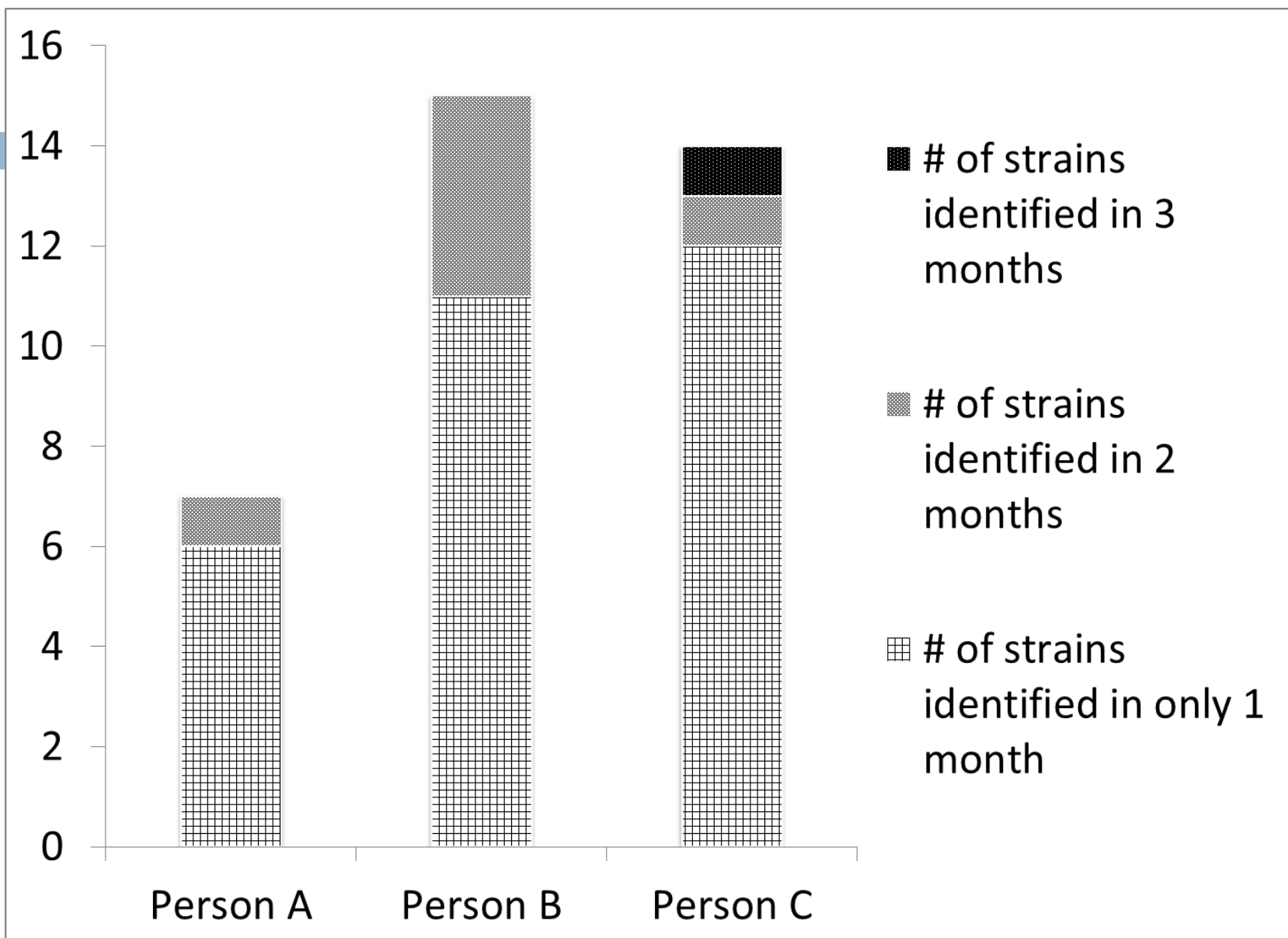


Person C

of
Isolates



of
Strains





Conclusions

What We Know So Far...

- Strain discrimination with pyroprinting
 - ▣ It's possible!
 - ▣ More sensitive than other methods
- Hypothesis supported....partially
- Variation is variable!
 - ▣ Population structures in different humans may vary in different ways
- Change will occur
 - ▣ Supported by both this study and literature

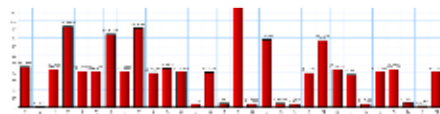
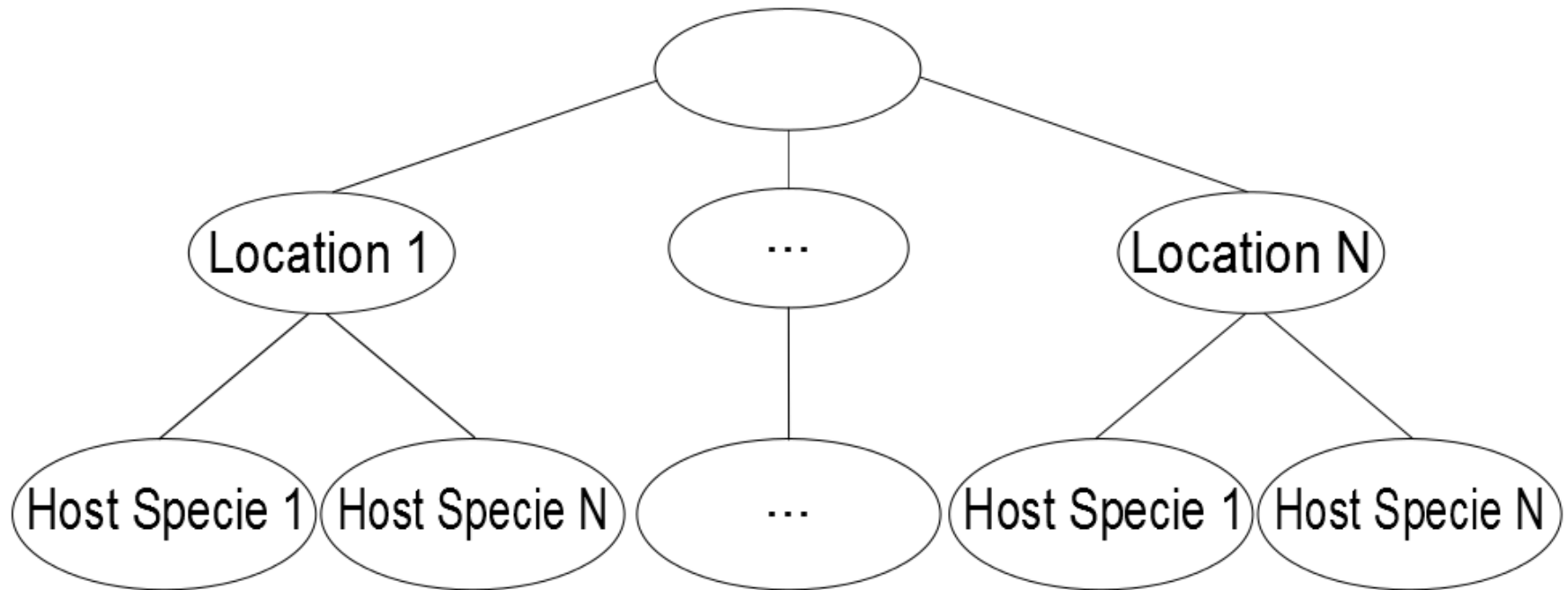
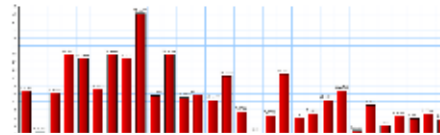
Future Work



- More data to analyze!
- Building a database
- More temporal studies
- Broad human population studies

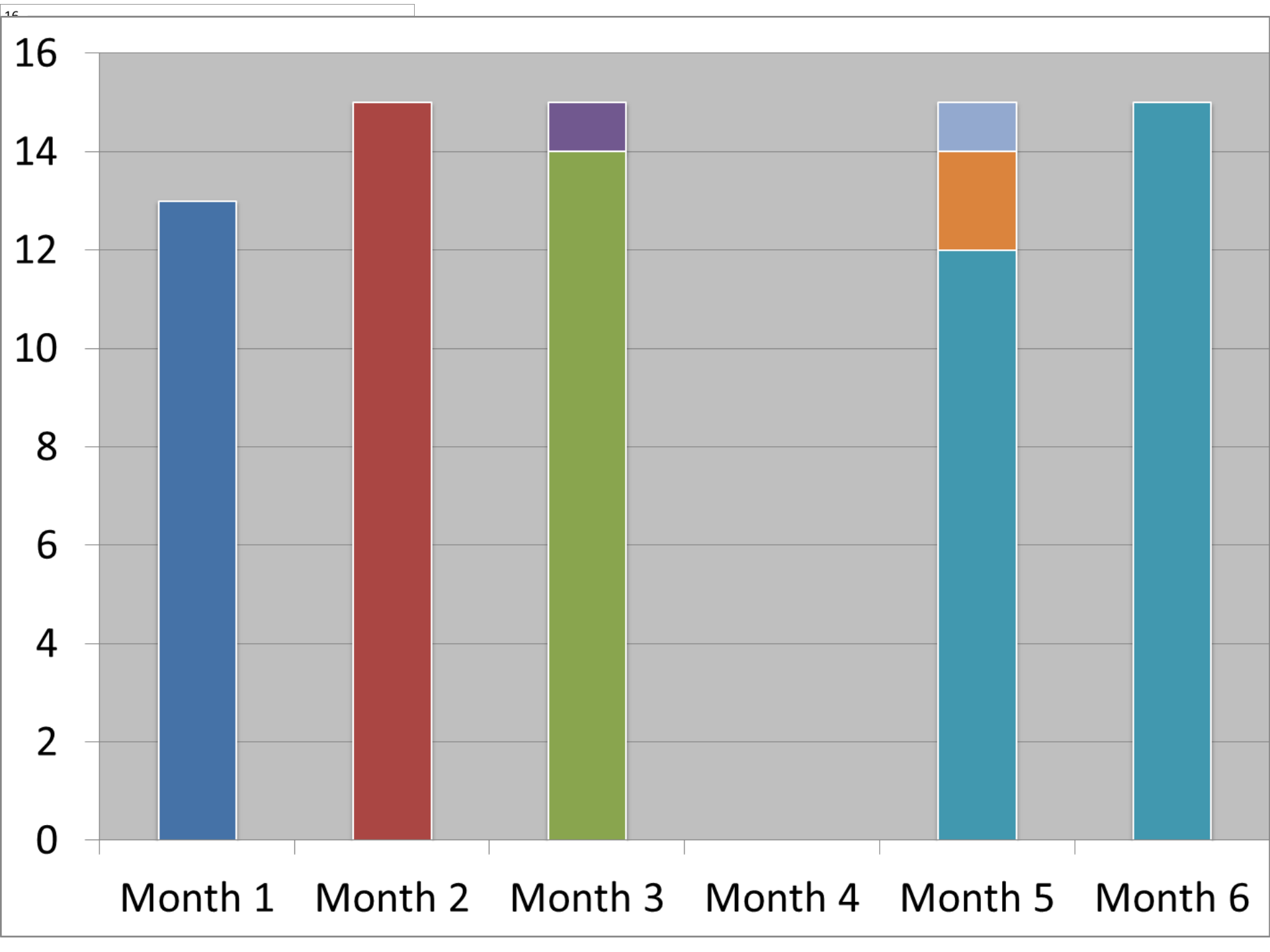
Future Work

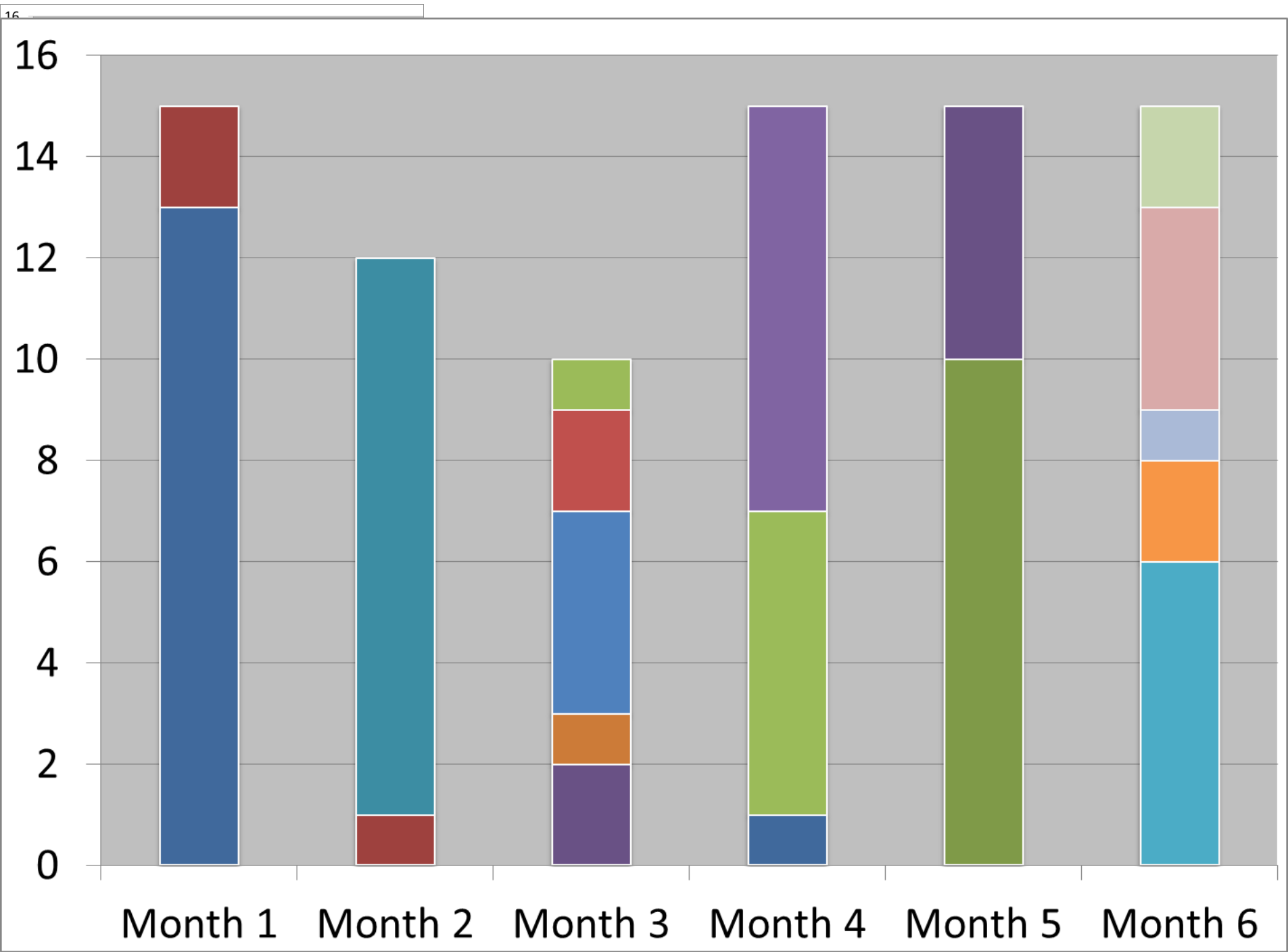
Hu-404

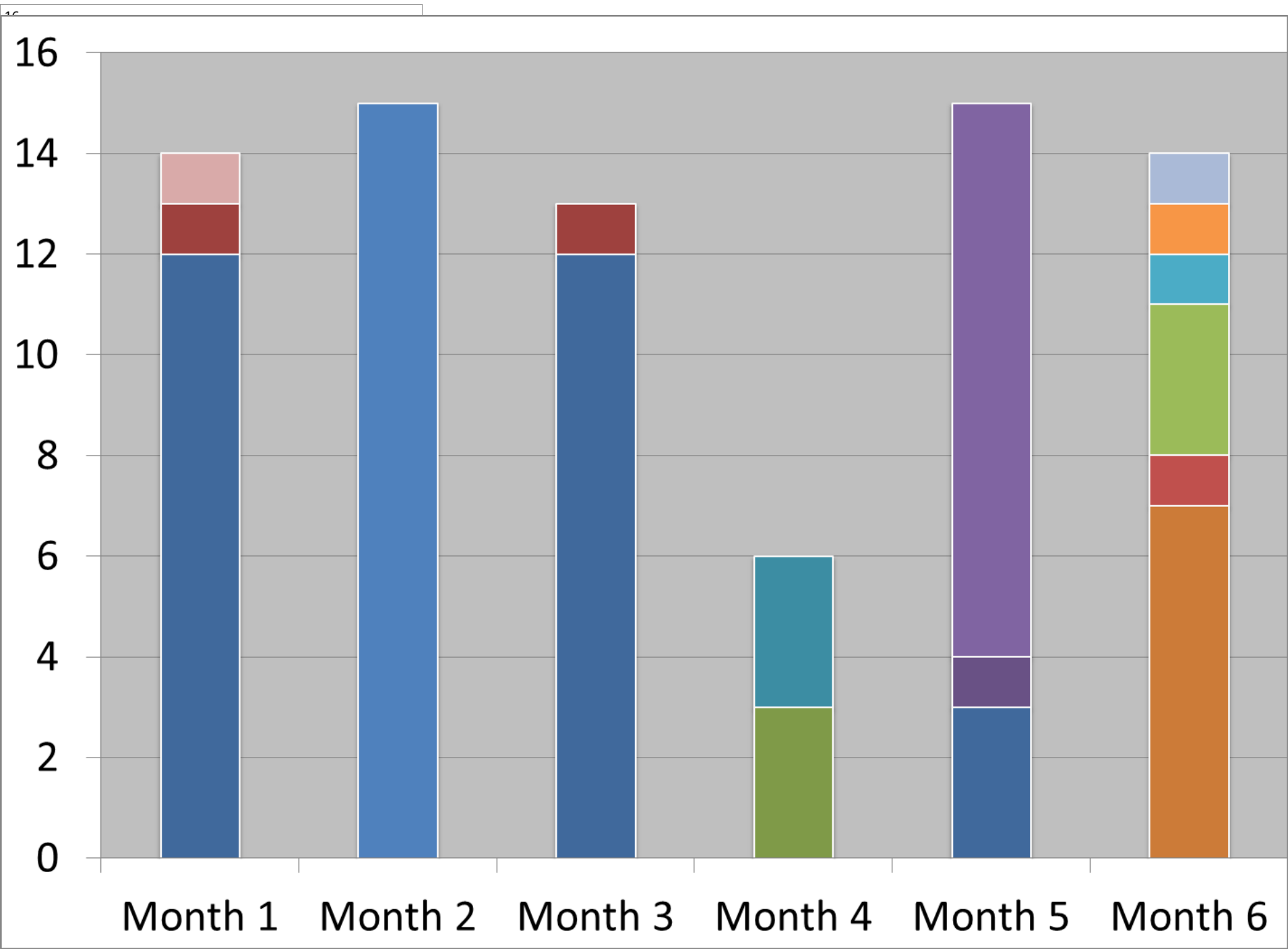


Questions?









Clustering Approach

