2018-05-10 Meeting with Steve Simon, Xing Song, and Mei Liu

Brainstormed on topics for the learning tutorials for researchers using HERON for data mining

Materials to be shared on github: /kumc-bmi/heron-i2b2-analytics/

**Basic Analytics for Beginners**

HERON data extraction topics for Beginners:

* How to connect to an oracle database and count the number of unique patients with SQL
* How to pull patient diagnostic codes and with date restrictions with SQL
* Documentation on the meta-data such as what is a concept\_path? What do you need it for? This is essential for hypothesis driven research where researchers need specific concepts to describe their cohorts.
* How to transform long-skinny data to sparse matrix format for R analysis

Programming Language:

* R, SAS, Python, SQL
* What is each language good for?

**Advanced Analytics**

Data Mining and direct access to full database allow us to do complex analyses with case studies such as breast cancer, acute kidney injury, diabetic kidney disease, and sepsis.

Each case study contains the following modules:

* Cohort construction using SQL with direct access to database
* EMR data extraction for the cohort in R
* Data dictionary creation for use in cohort characterization, data curation, and etc.
* Data cleaning
* Validation experiment set up, for example random sampling for cross validation.
* Modeling with machine learning models

Cluster at KU Center for Research Computing:

* What is it?
* Why?
* How to use?

**Identify Testers for the Learning Tutorial**

* Informatics background 🡪 Computer Science / Bioengineering (KU-Lawrence)
* Medical background 🡪KUMC students