THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

CohortGen: Promoting Collaboration Through Automatic Cohort Generation

Billy Barnett MA
Charisse Madlock-Brown PhD

Have an EHR warehouse…what next?

- What is a Data Warehouse?
  - An integrated, granular, historical single point of reference for data
  - Data warehousing in healthcare
    - Millions of patient records collected
    - Billions of data points
  - Extracting data is challenging
    - Requires special skill
    - Many researchers are unclear on what is available

Promoting Research And Collaboration

- Need for cohort generation
  - Challenges associated with promoting collaboration through profiling systems
  - Difficulty identifying meaningful subsets from Big Data EHR data warehouses

Primary Data Source

- Health Facts Data Warehouse
  - Started in 2001 by Cerner
    - Over 70 million patient records
  - Types of data
    - Demographics
    - Prescriptions
    - Laboratory tests
    - Billing
    - Diagnoses
    - Procedures
    - Clinic locations
    - Hospitals

Elsevier Developer API

- REST API to retrieve research article metadata:
  - Title
  - Authors and affiliations
  - Abstract
  - Keywords

- With a developer key, we were able to get all UTHSC articles

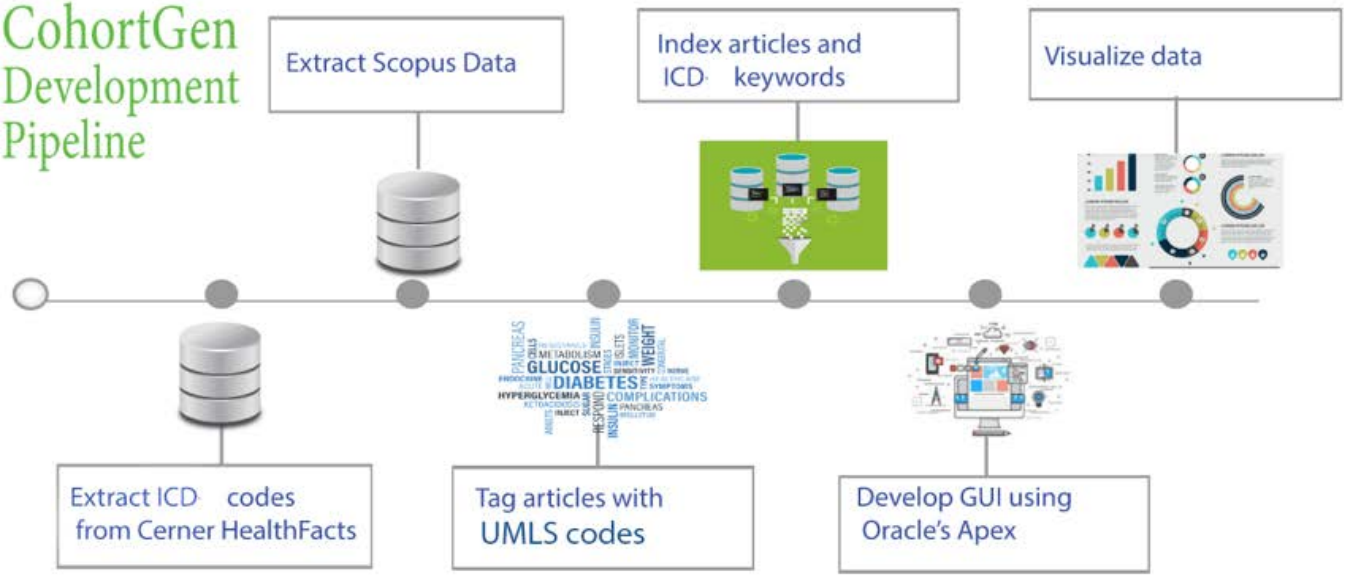
cTakes

- Clinical Text Analysis and Knowledge Extraction System (cTAKES)
- Apache NLP system for extracting information from EMR
- Can be used to tag documents with UMLS concepts

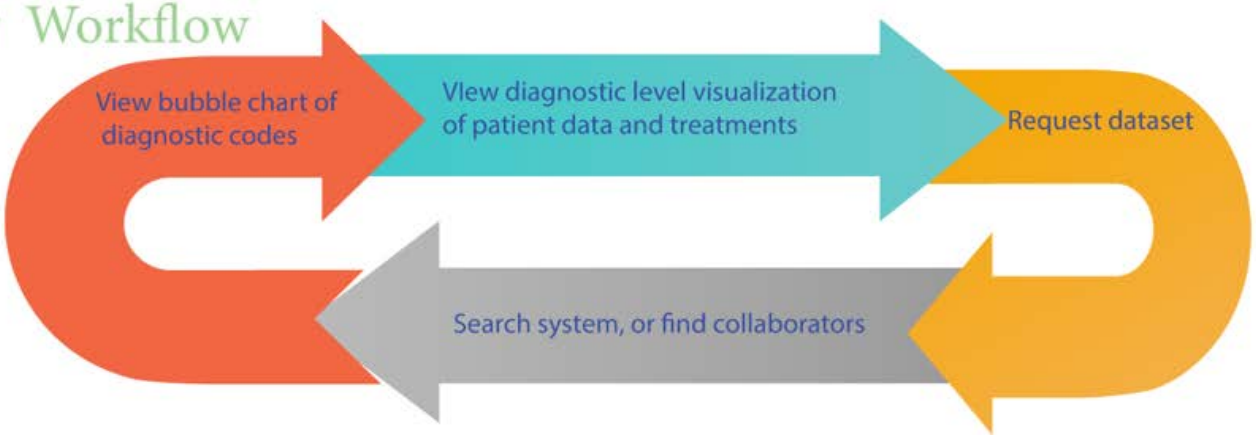
Other Data Sources

- Unified Medical Language System (UMLS)
  - Produced by the National Library of Medicine
  - Metathesaurus database
  - Useful for informatics research
- Faculty extract from human resources

CohortGen Development Pipeline

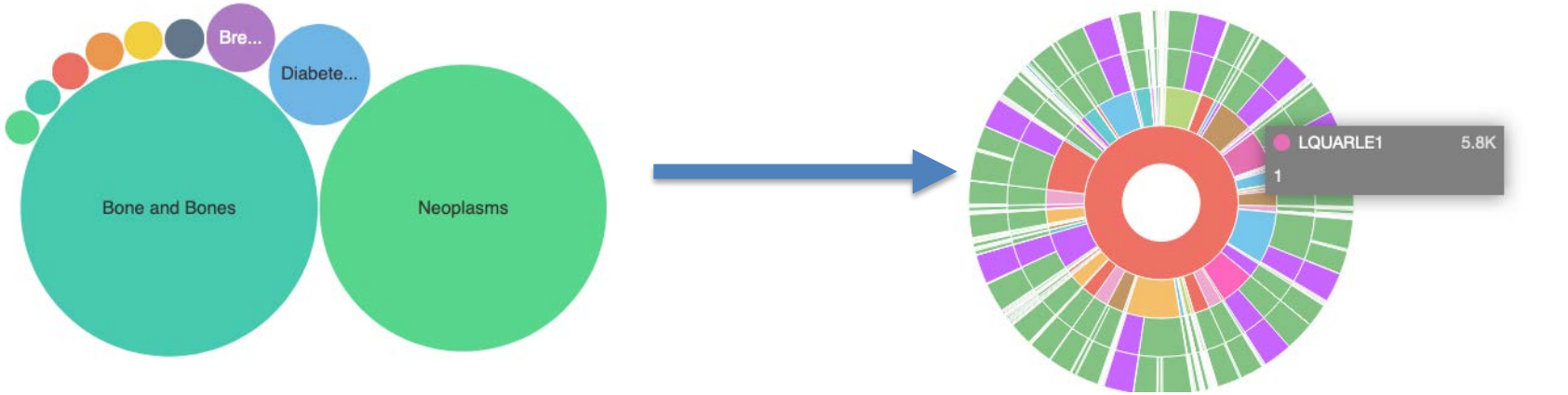


User Workflow



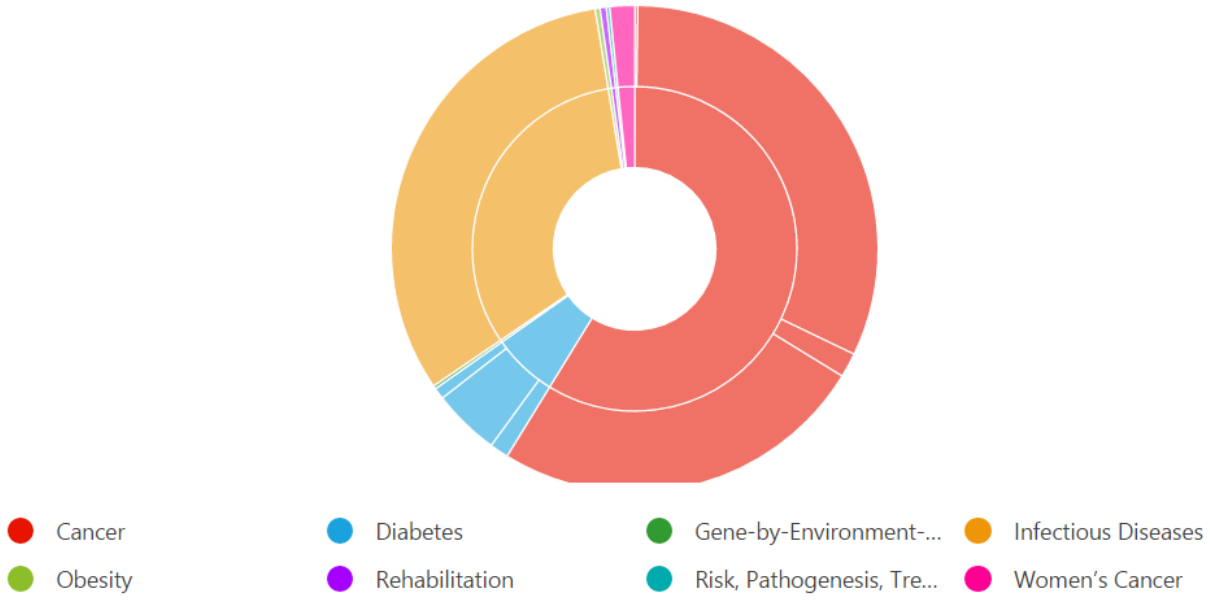
Sample Navigation of UTHSC Areas of Excellence

Finding researchers



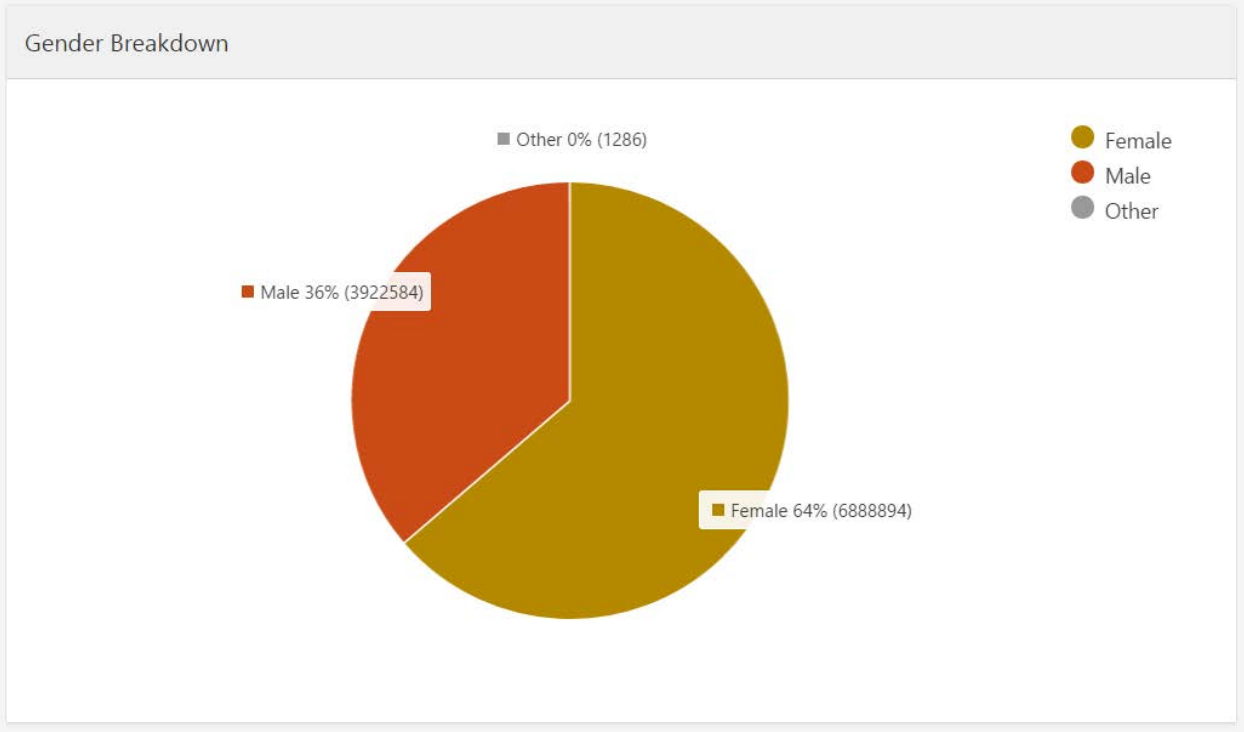
Researcher Profile

- Sunburst chart summary of research areas



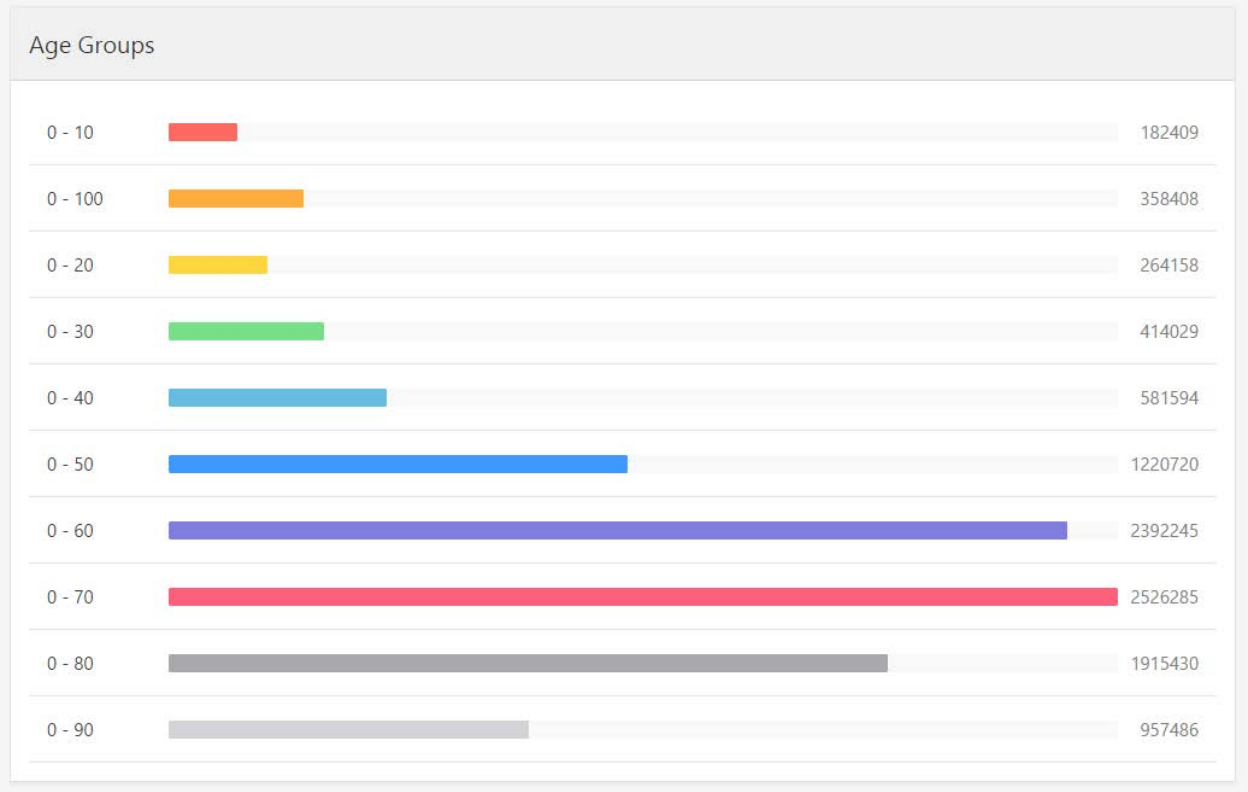
HEALTH Facts information for Area

- Gender breakdown summary



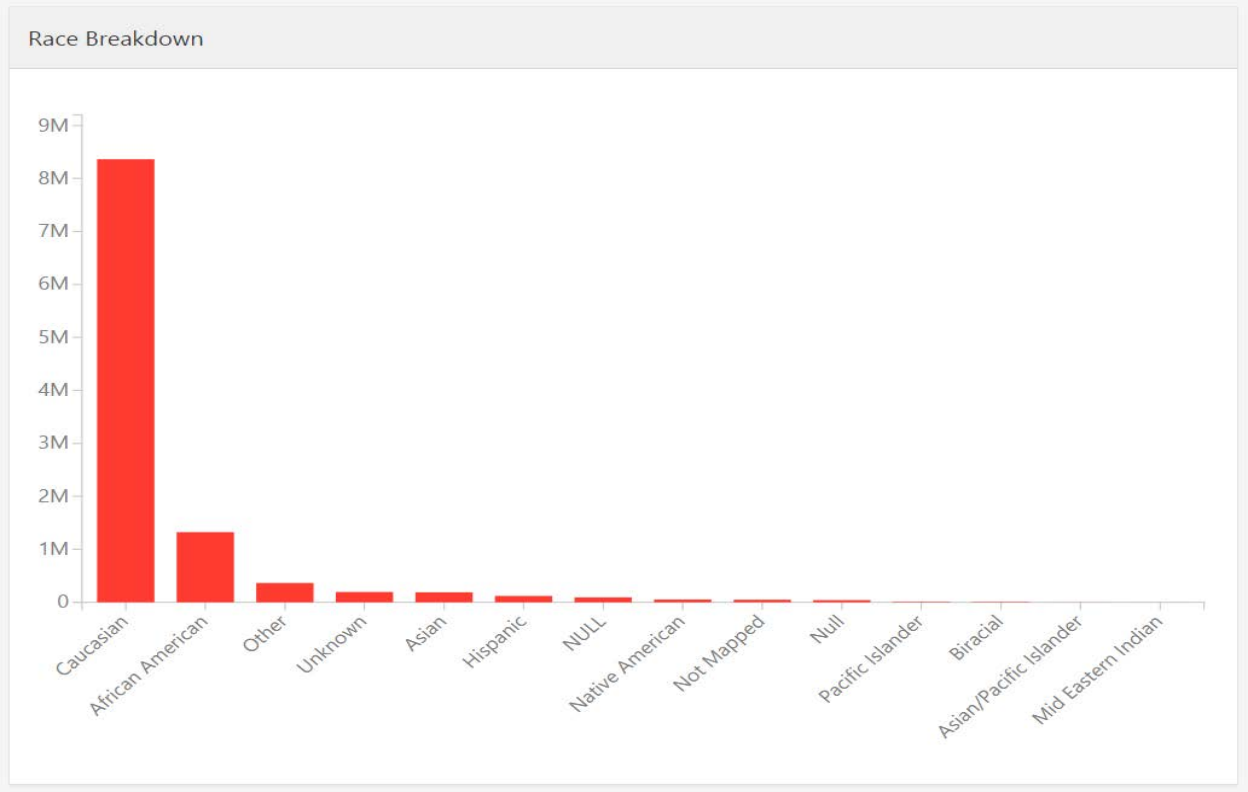
Health Facts information for Area

- Age group summary



Health Facts information for Area

- Race breakdown summary



Small User Review

- Group of three researchers found our system helpful
- Need for search option
  - We have created solr indexes for scholarly articles and diagnostic codes.
- Want to view articles in system
- Not interested in drill-down view
- Want collaborator suggestions when reaching Health Facts page
- Need emphasis on current research

CohortGen Fundamentals

- Generating personalized cohorts based an automatically identified faculty interest
- Data visualization of aggregated data gives users simple view of cohorts
- Research data can be linked to team opportunities

References

- Bennett LM, Gadlin H, Levine-Finley S. *Collaboration & Team Science: A Field Guide*. NIH Office of the Ombudsman, Center for Cooperative Resolution Bethesda …; 2010.
- 2. Shin S-Y, Kim WS, Lee J-H. Characteristics Desired in Clinical Data Warehouse for Biomedical Research. *Healthc Inform Res*. 2014;20(2):109. doi:10.4258/hir.2014.20.2.109