

# The Unintended Benefits of Applying Your Data Scientists to Your EDW

Hannah Marshall  
Junior Data Scientist  
Mission Health

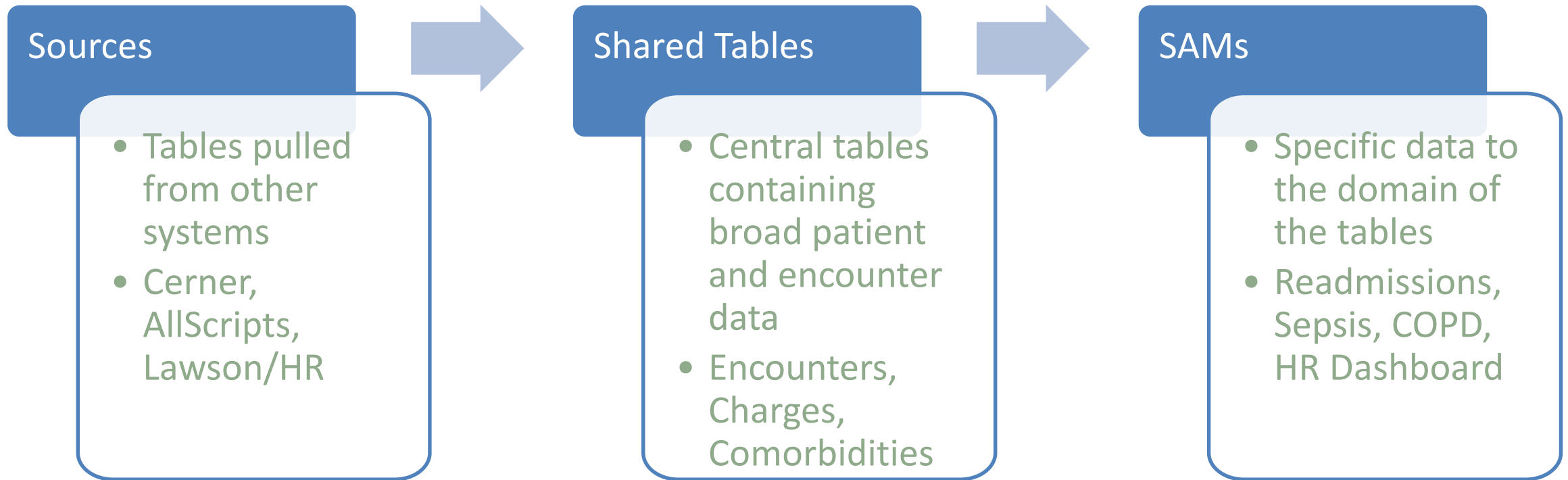
**Himss**

**NORTH CAROLINA** *Chapter*

# Outline

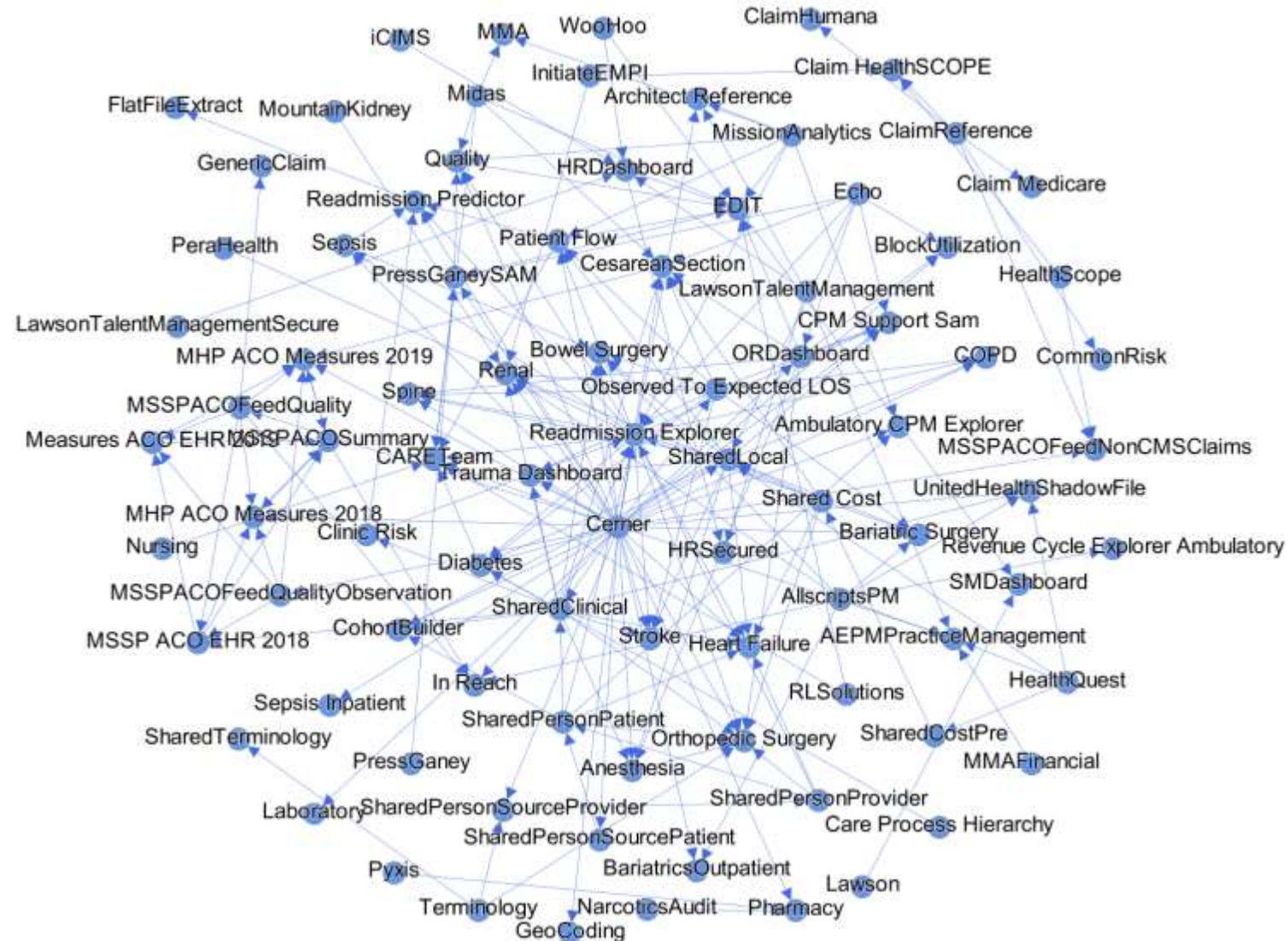
- Discuss Mission Health's
  - EDW
  - Roles in Analytics
- What is a Data Scientist?
  - Role at Mission Health
- Situations where data science has proven useful to the EDW

# Enterprise Data Warehouse (EDW)



The EDW is a sequence of transformations on data.

# Our EDW



# Clinical and Business Analytics

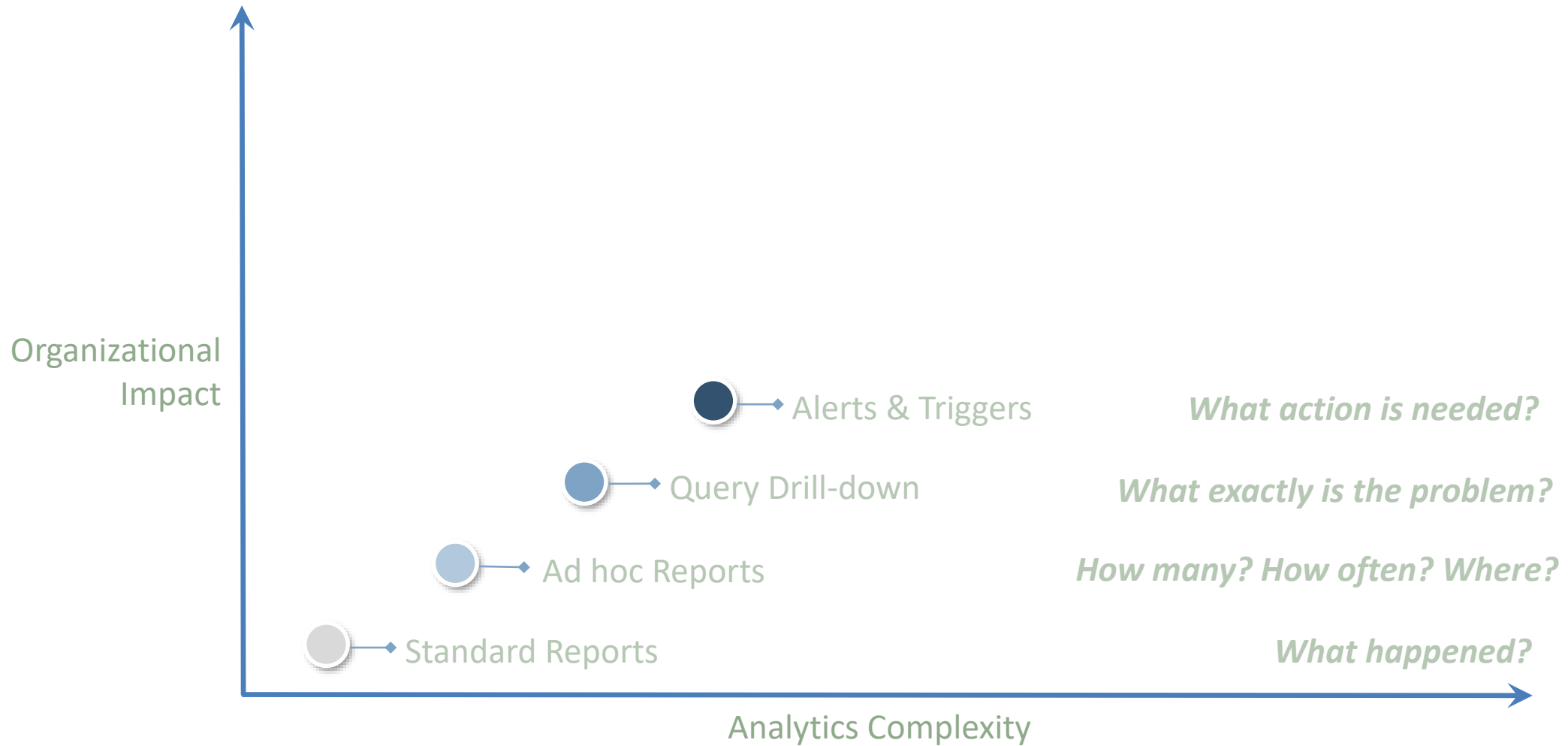
- 24 people in 3 Teams
  - Data Architects
  - Analytics Insight Consultants
  - Data Scientists

# Data Architects

- Work mostly in SQL
- Build the SAMs (domain specific tables)
- Manage the batch load so that tables update daily
- Build dashboards for end users



# Complexity of Analytics Questions



# Analytics Insight Consultants

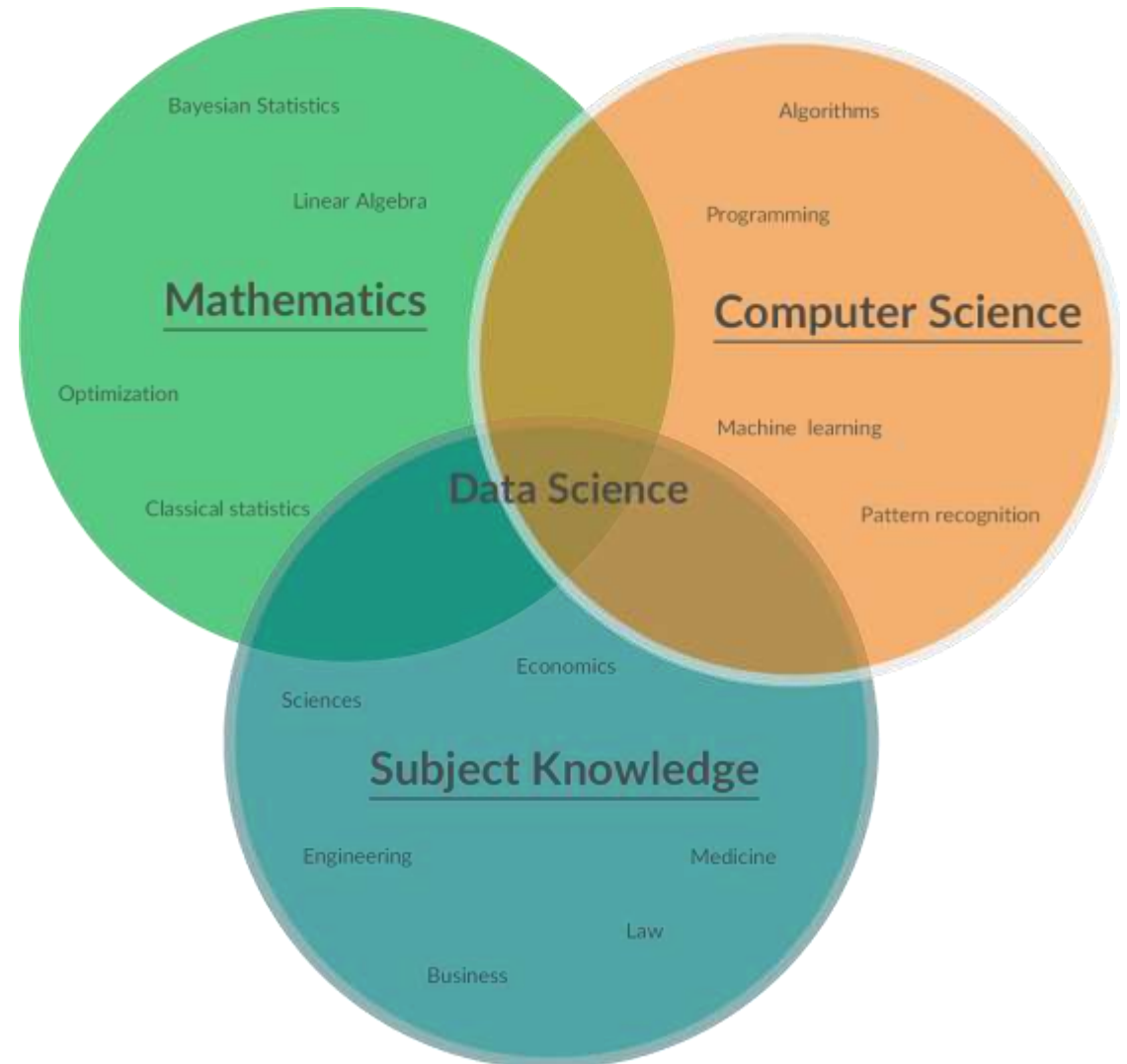
- Maintain relationship with users and stakeholders
- Interface with stakeholders to gather requirements
- Monitor quality checks of dashboards
- Maintain and communicate release schedule





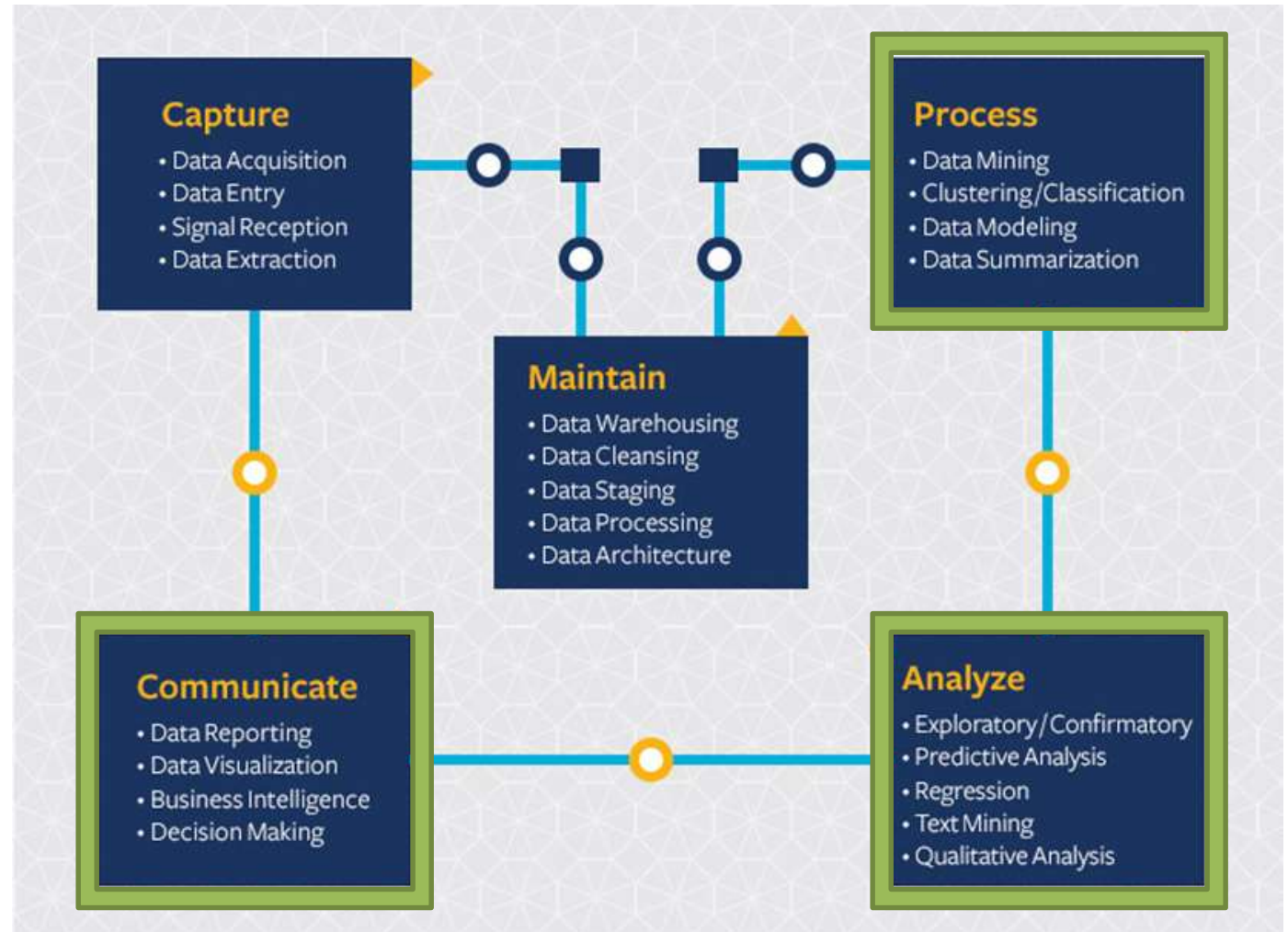
# What is a Data Scientist?

- Title first appeared in 2008
- Blends traditional statistics with other fields

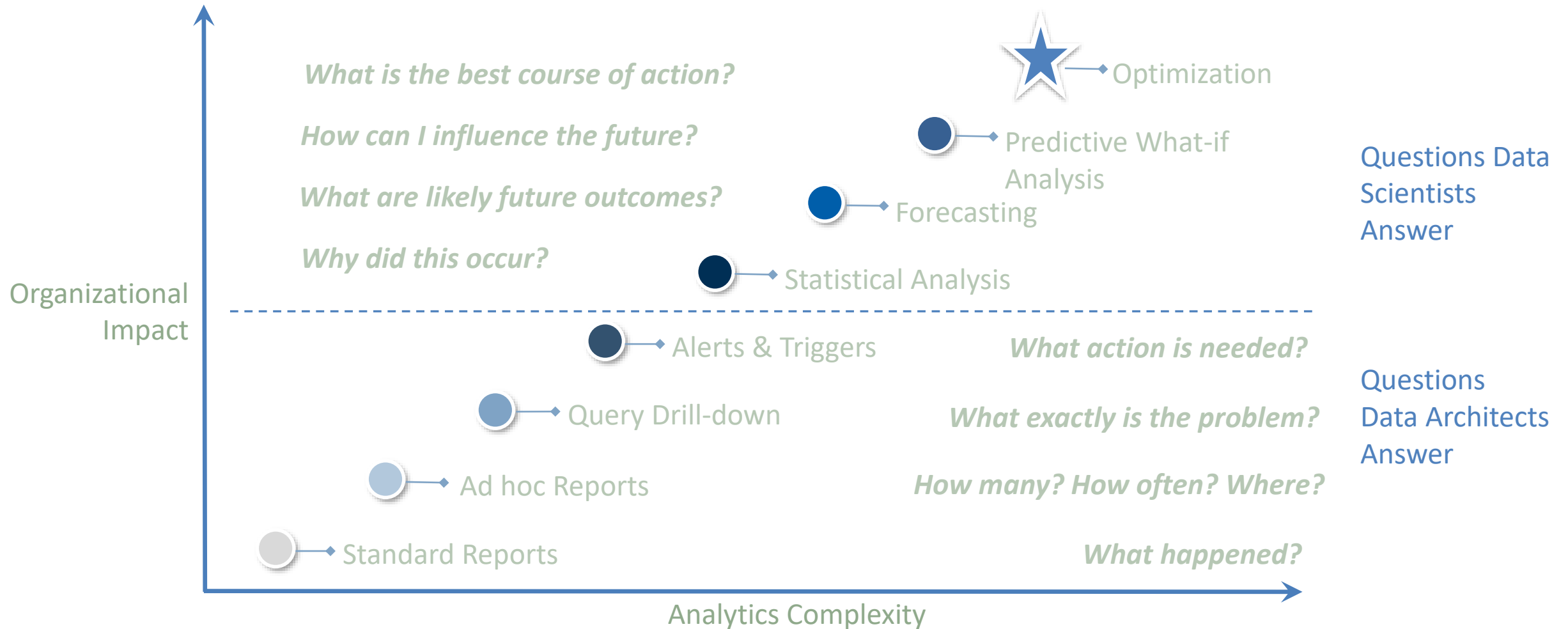


# What is a Data Scientist?

- Title first appeared in 2008
- Blends traditional statistics with other fields
- Roles defined differently at different locations



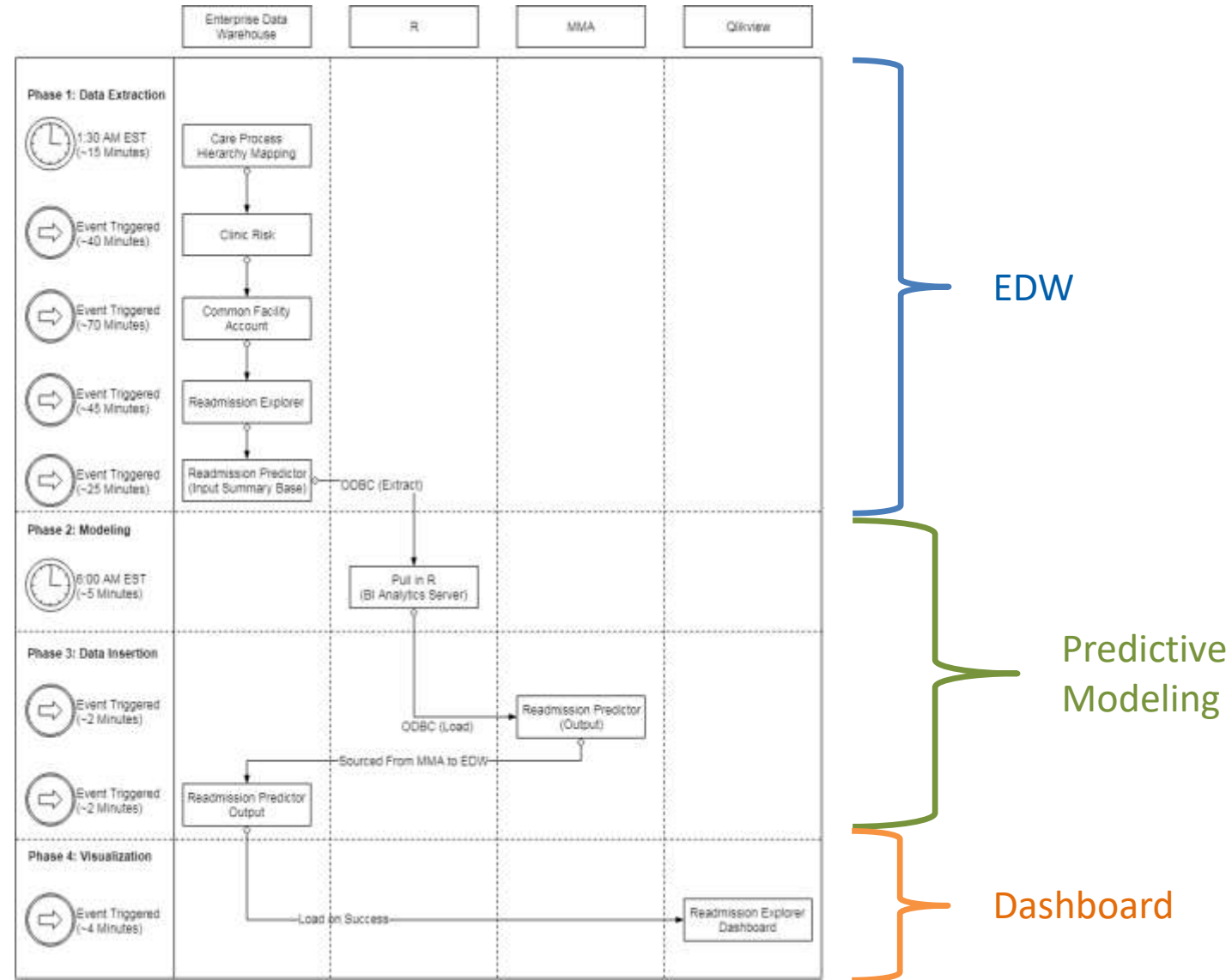
# Complexity of Analytics Questions



# Data Scientists at Mission Health

- Production Predictive Models
- Statistical Consulting
- Intervention Evaluation
- Increasing the value of EDW assets/data

# Production Data Science – Readmission Predictor



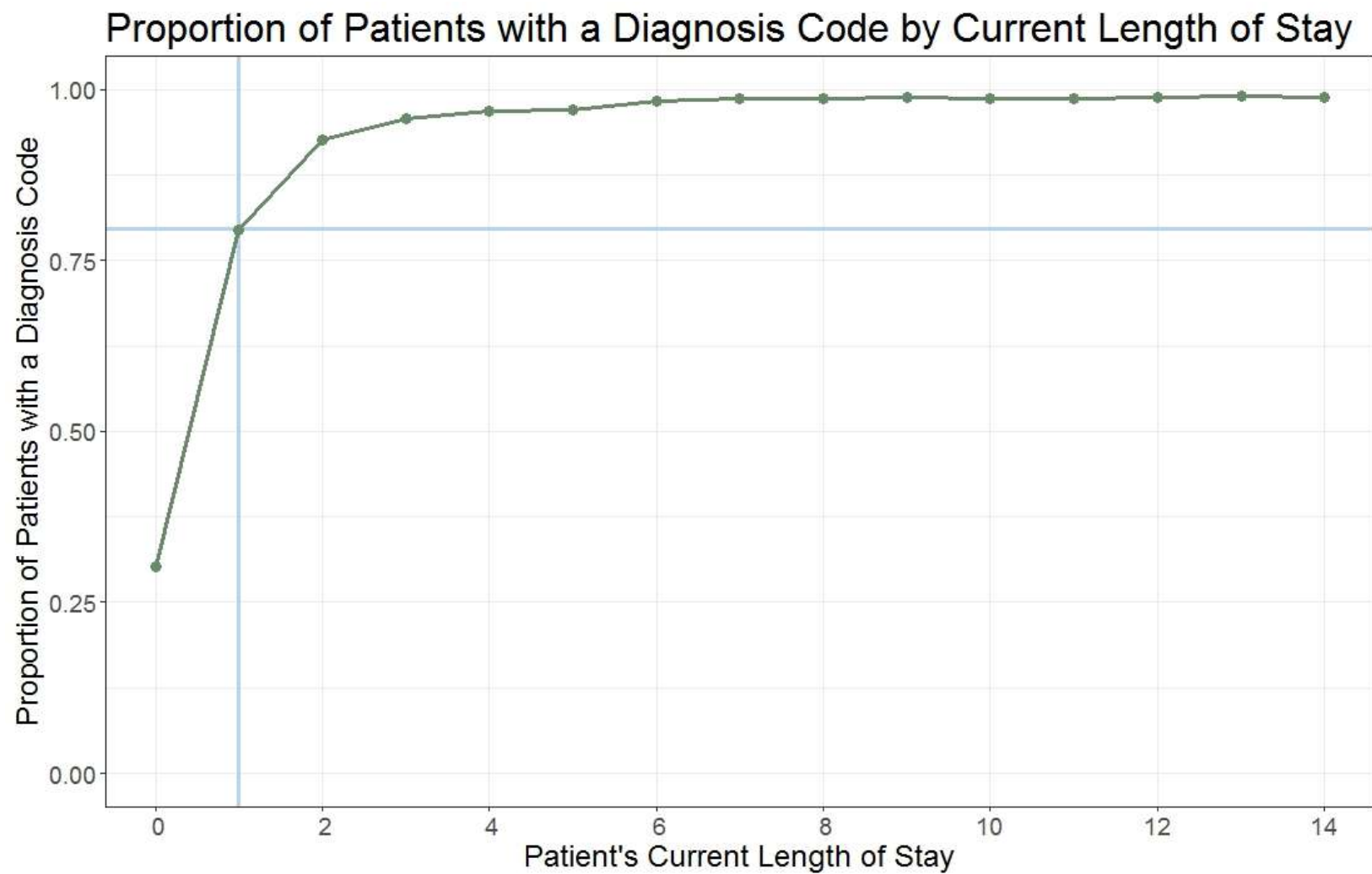
# Production Data Science

- To create Production Models to run automatically:
  - Cleaning data must occur automatically
  - You must anticipate issues in the data ahead of time

# Data Quality

- **Completeness:**
  - Does each patient have all of their data?
- **Timeliness:**
  - Does the data come in when you need it to?
- **Integrity:**
  - Is the data right?
- **Consistency:**
  - Does the data stay the same or change frequently?

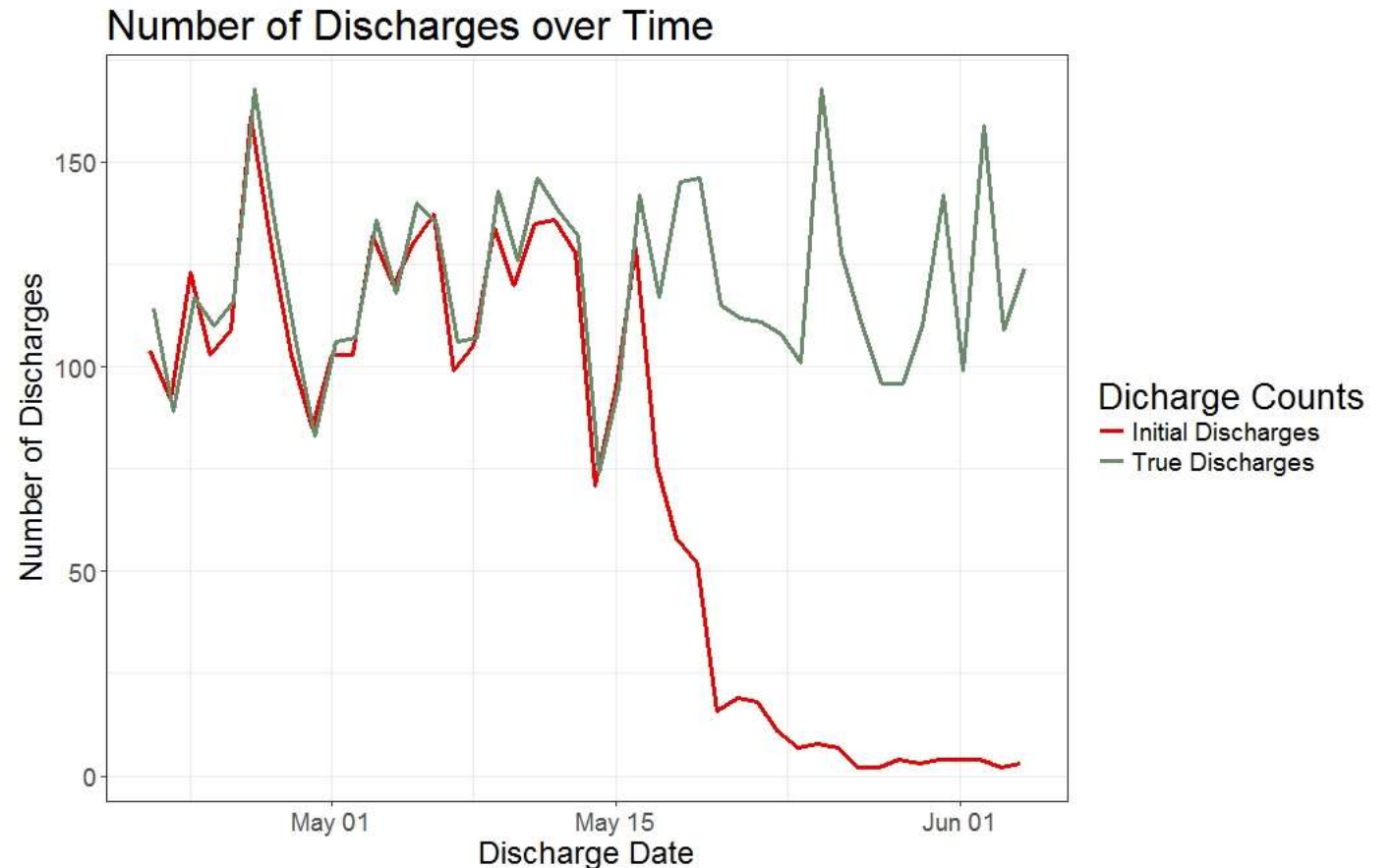
# When Does Data Come In? (Timeliness)





# The Case of the Disappearing Patients (Integrity)

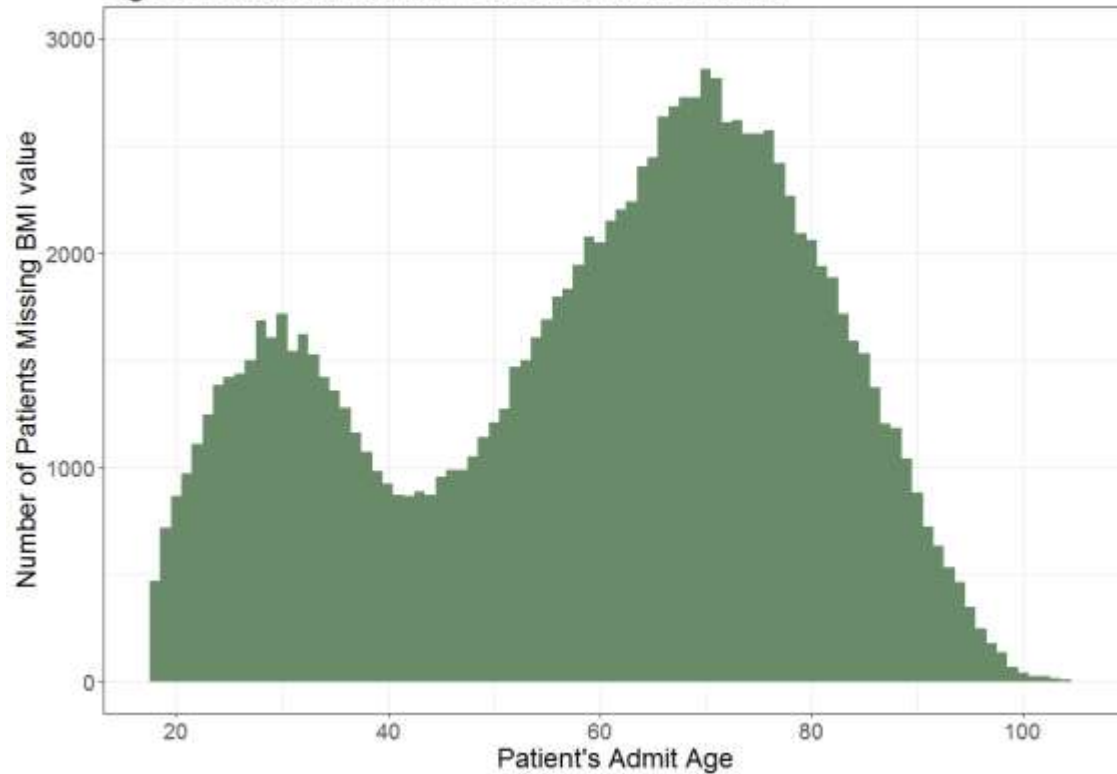
- Mid May 2017 we had patient flow decrease in app
- 2 source tables had been turned off



# Sometimes small things show a bigger problem too...

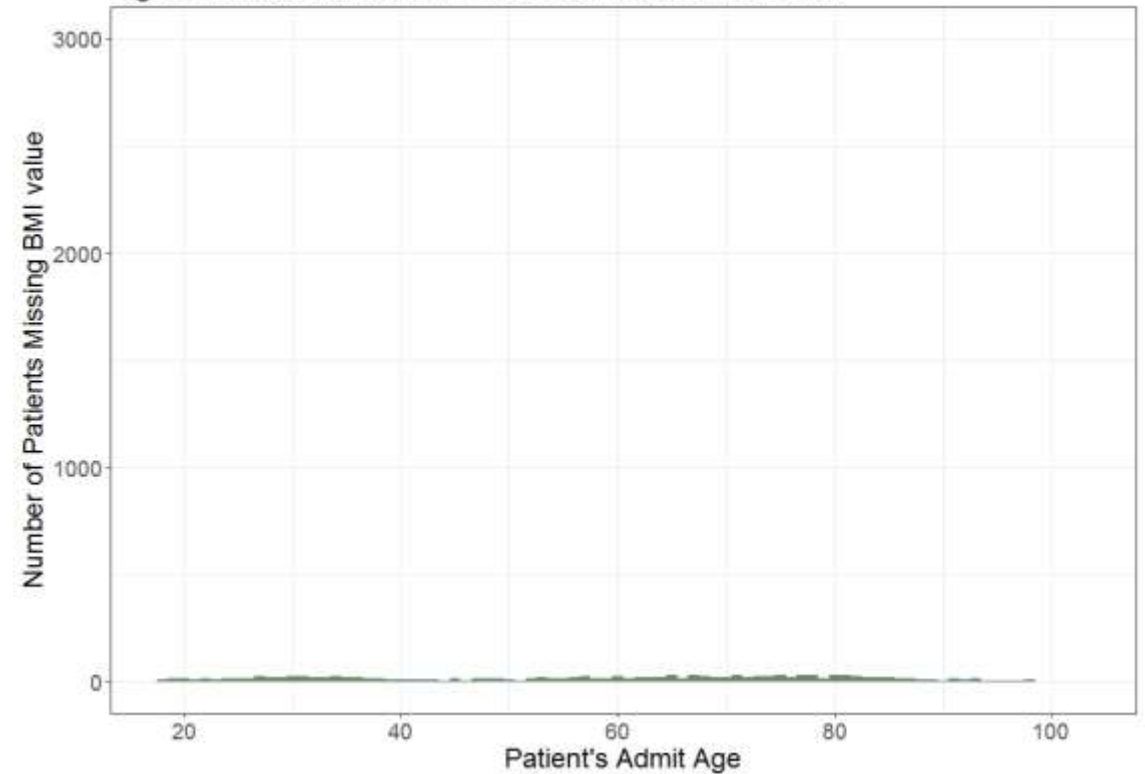
## 78% Missing BMI

Age Distribution of the First Version of BMI



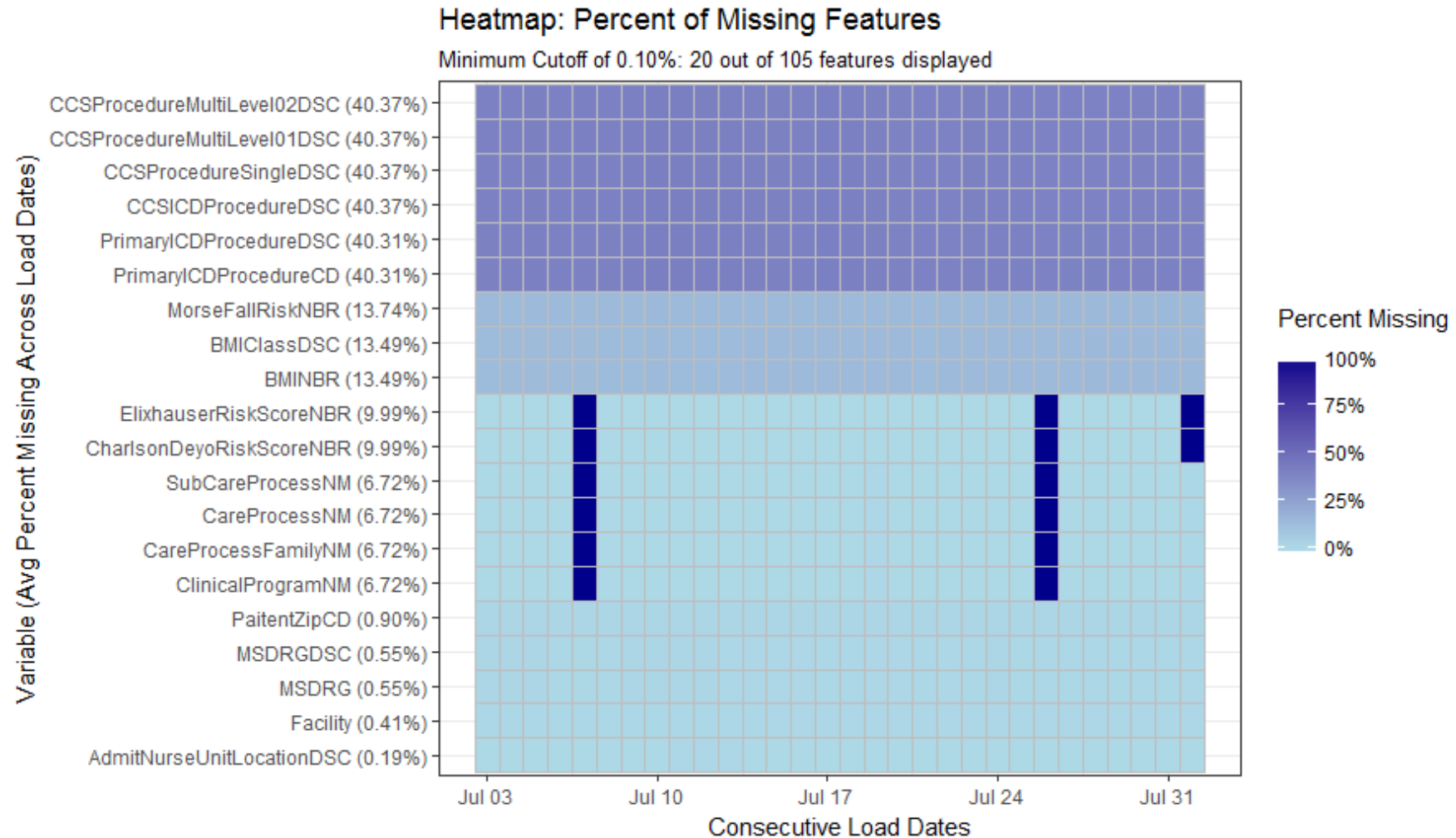
## .9% Missing BMI

Age Distribution of the Second Version of BMI



# Missing Data Fields (Consistency)

- Track Variables over multiple load dates
- Some days certain variables wouldn't load



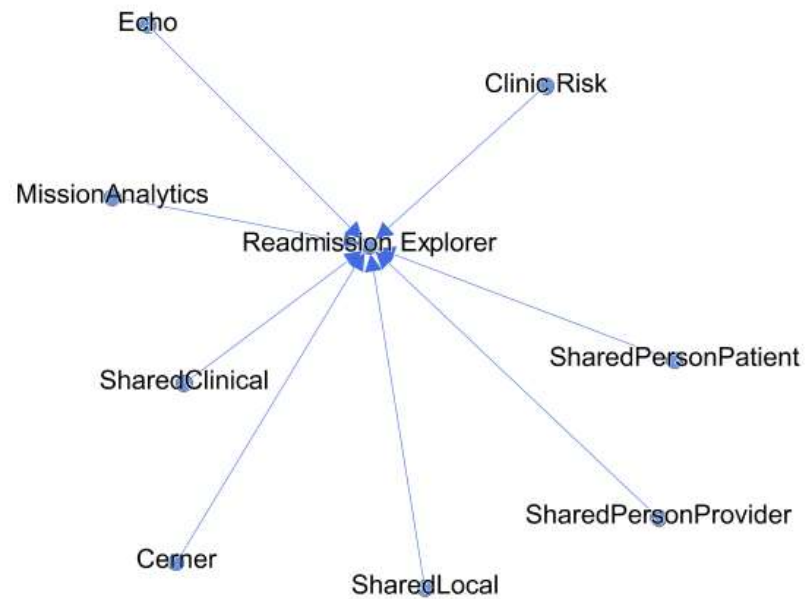
# Tracking Table Dependencies – Initial State

- Dependencies: What tables have to run before current table
- Only able to look 1 level up
  - What tables are called within the SQL used to create new table

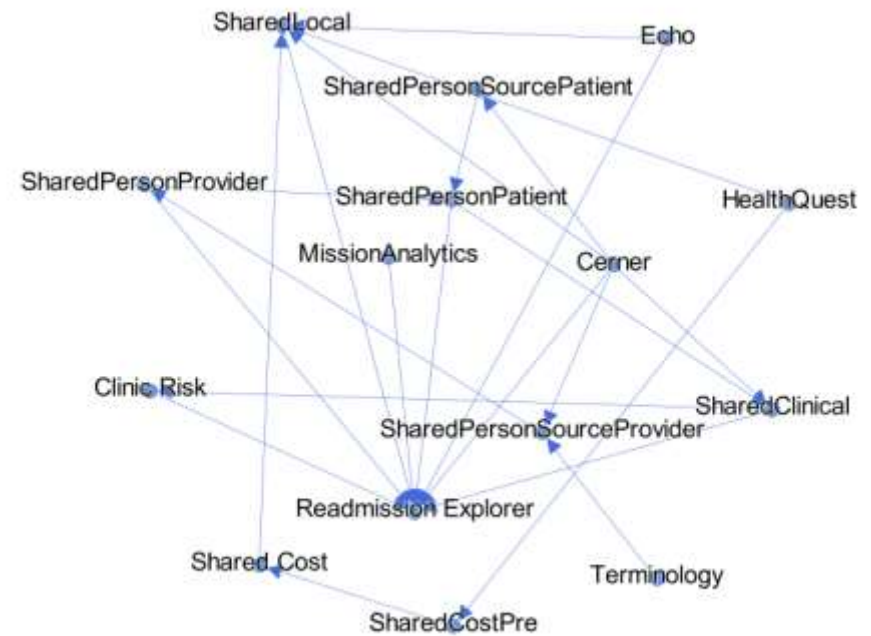
# Comparing Perspectives on Table Dependencies

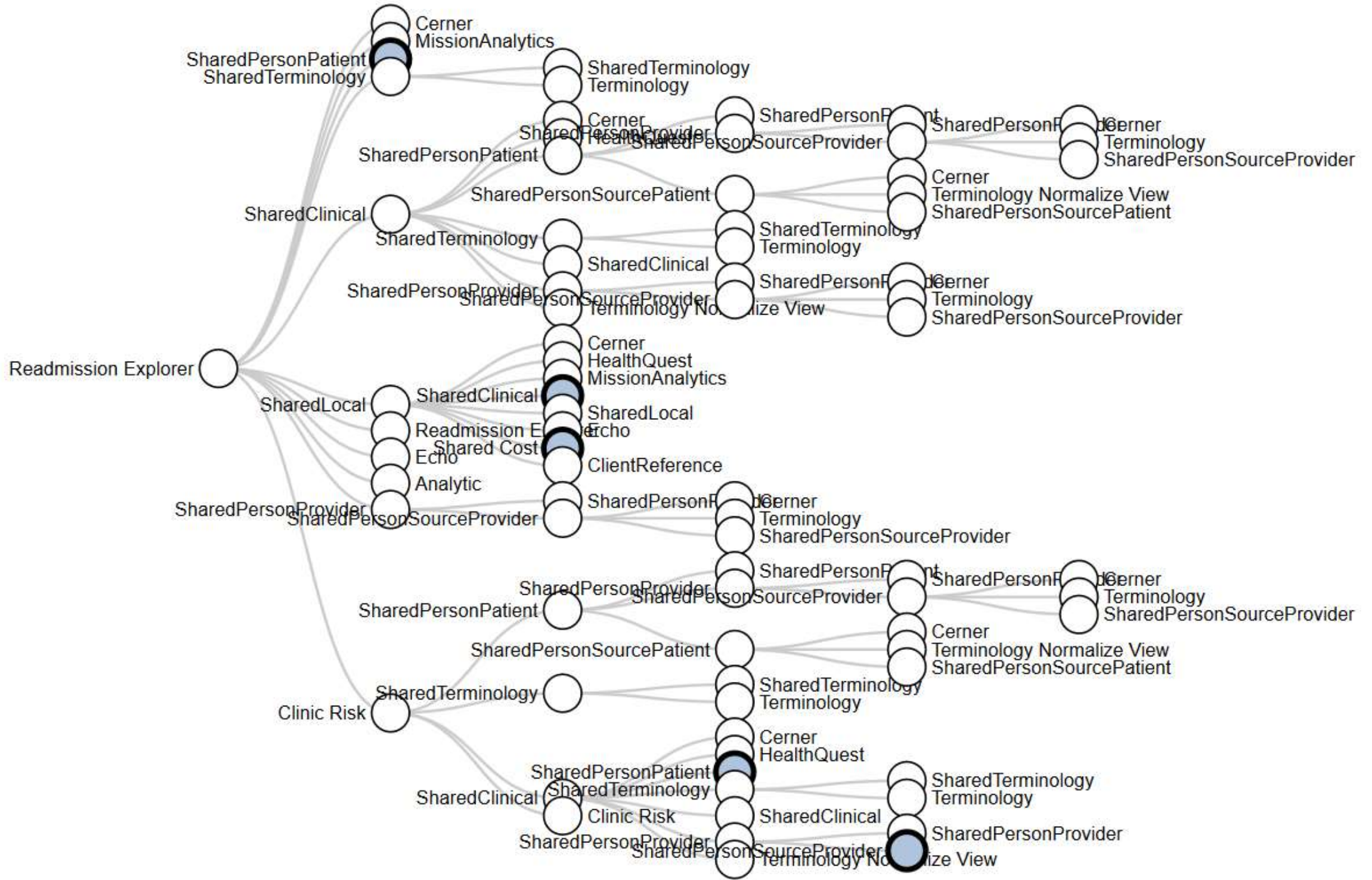
## Readmission Explorer

### Parents Only



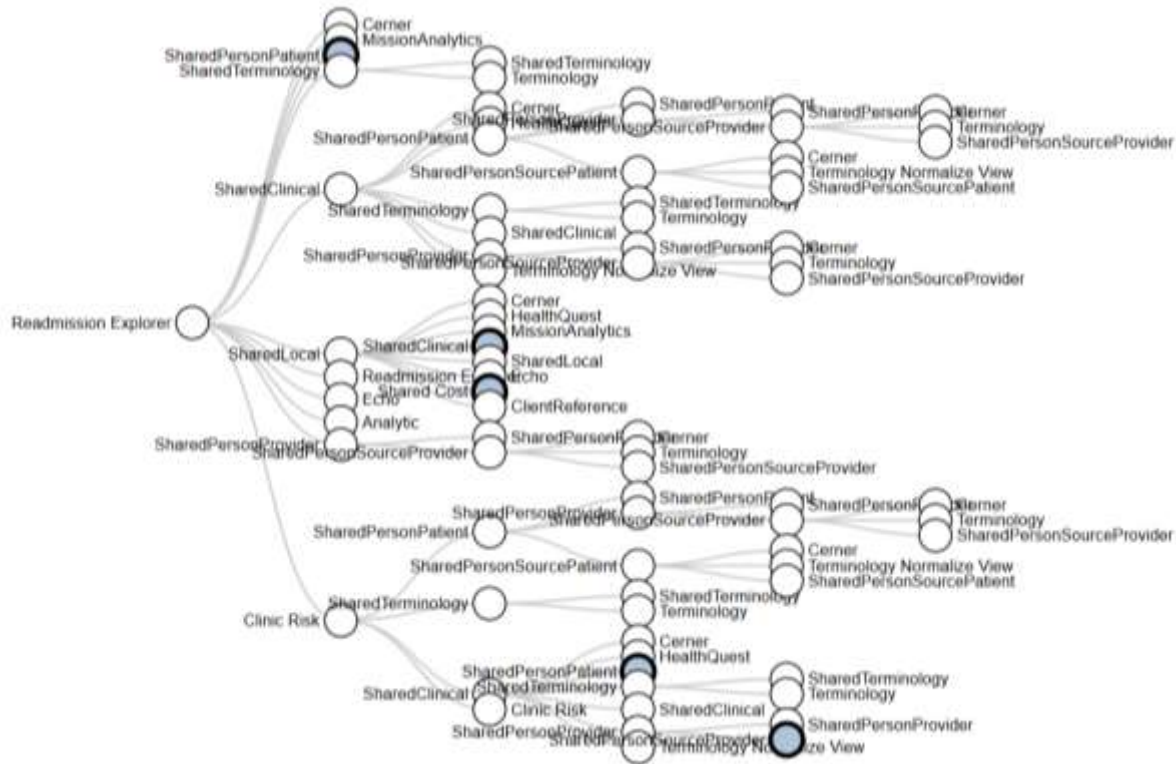
### Full Family Tree





# EDW Upgrade – Cleaning out the Closet

If you need to remove tables, you need to know if they are being used.

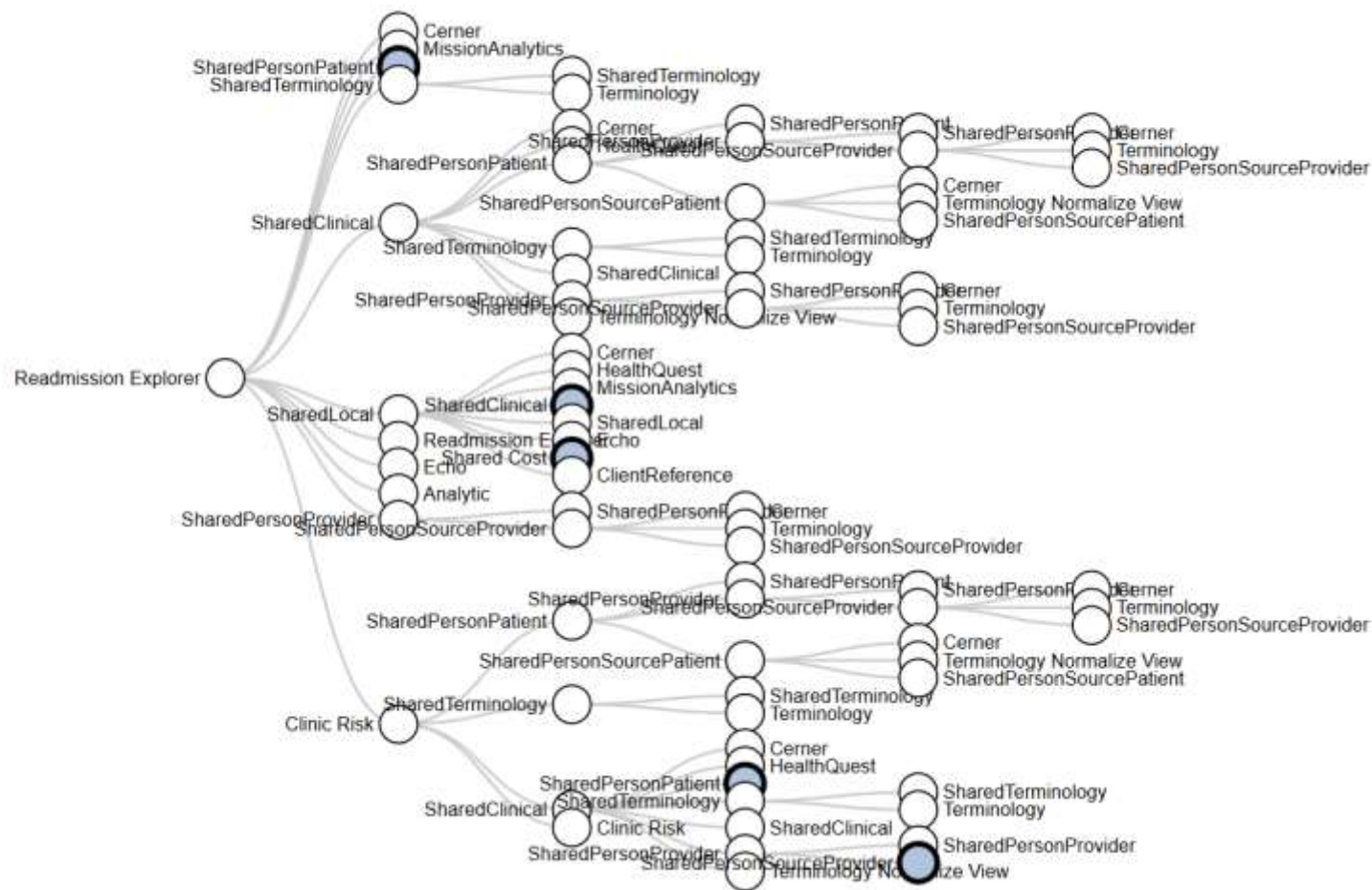


child	parent
Readmission Explorer	Cerner
Readmission Explorer	MissionAnalytics
Readmission Explorer	SharedPersonPatient
Readmission Explorer	SharedTerminology
Readmission Explorer	SharedClinical
Readmission Explorer	SharedLocal
Readmission Explorer	Readmission Explorer
Readmission Explorer	Echo
Readmission Explorer	Analytic
Readmission Explorer	SharedPersonProvider
Readmission Explorer	Clinic Risk

# Coming Soon

## Flexible application for entire department

Parent	Child
Cerner	Readmission Explorer
Clinic Risk	Readmission Explorer
Echo	Readmission Explorer
MissionAnalytics	Readmission Explorer
SharedClinical	Readmission Explorer
SharedLocal	Readmission Explorer
SharedPersonPatient	Readmission Explorer
SharedPersonProvider	Readmission Explorer
SharedCostPre	Shared Cost
Cerner	SharedClinical
SharedPersonPatient	SharedClinical
...	...





# Conclusion

- It takes a variety of roles to ensure the quality of EDWs, including data scientists.
- Data Scientists have unique tools and viewpoints to assess the completeness, timeliness, integrity, and consistency of the EDW data.