

Database Design

Hands-On: R Basic and Database Using R

BMI701 Introduction of Biomedical Informatics
Lab Session 2

Wei-Hung Weng

September 14, 2016

HMS DBMI — MGH LCS



Programming?

- What computer can do for us?
- **Calculation**
 - `3 + 4`
 - `2 ** 5`
 - `exp(log(5 ** 4))`
- **Storage**
 - `x <- 4 * 9`
 - `y <- c('Harvard', 'MIT', 'Beaver')`
 - `write.csv(x, 'aloha.csv')`
- Writing papers...!?

Think Computationally

- Abstraction
 - Hide details, formulate interface
 - Functions
- Automation
 - Mechanize your abstraction
 - Use machine to interpret your notations and models
- Algorithm
 - Describe automated process
 - Recipes

MIT 6.00.1

MySQL Database in R

- Using MySQL to filter out the data
- ```
SELECT * FROM sitka WHERE tree = 1
INTO OUTFILE 'filename.csv'
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY 'BACKSLASH_n'
```
- Using RMySQL or RPostgreSQL with SQL syntax

- `library(RMySQL)`
- `con <- dbConnect(MySQL(), user="root",  
password="", dbname="gwas", host="127.0.0.1")`
- `dbListTables(con)`
- `dbGetQuery(con, "select * from gwas")`

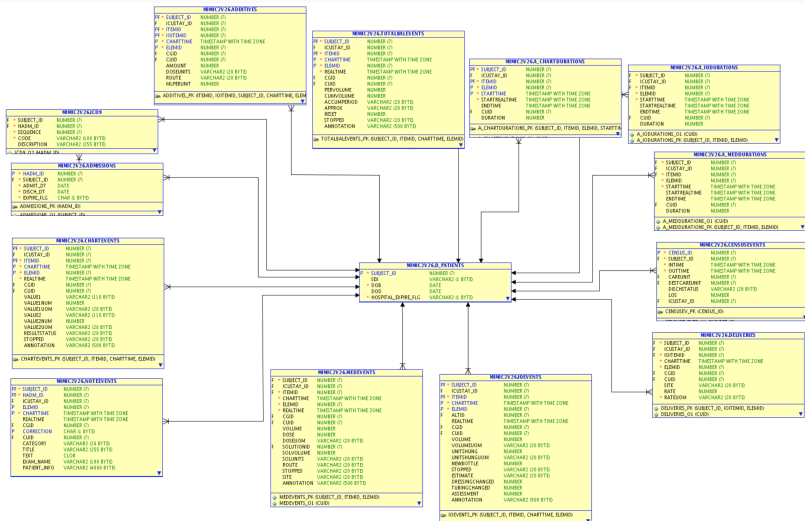
- Coding style is important for others to read your code
- Google's R style
- [github.com/ckbjimmy/bmi701lab/blob/master/lab02.R](https://github.com/ckbjimmy/bmi701lab/blob/master/lab02.R)

# Steps of Database Design

1. Create a data model
2. Define an attribute model
3. Define physical model
4. Review your design

Courtesy by Dr. Mujeeb Basit

## Revisiting ER Diagram

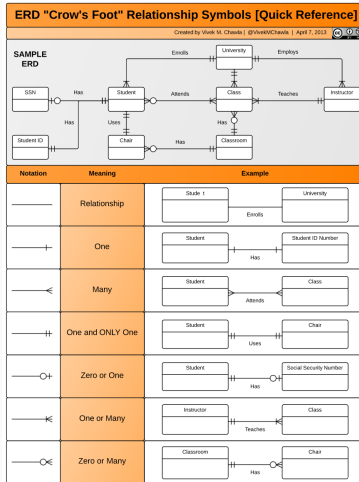




# Data Model - Entity and Relationship

- Identify candidate entities
  - Types of information
- Identify relationships:
  - Between the entities
  - Must be  $N \times (N - 1)$  relationships
- Define entities / relationships
  - one-to-one, one-to-many, many-to-many
- Review entity-relationship model
- Identify domains and ontologies

# Crawfoot Notation



Courtesy by VM Chawla

# Attribute Model

- Attribute is the property of entity
- List candidate attributes for each entity
- Add keys (PK, FK) to model
  - PK: uniquely identify the entity
  - Composite key: A key with  $>1$  attributes
  - FK: PK in other tables
- Define attributes data types and default values
  - Text (char, varchar, text), numbers (int, float), others

## Physical Model (Schema)

- Select appropriate DBMS
- Name tables / columns
- Name / Define indecies
- Define columns
- Verify / Update triggers

# Checklist

- Verify entities / definitions
- Verify relationships / definitions
- Verify attributes / definitions
- Verify constraints
- Approve database schema design

- ERD symbols
- [draw.io](#)

# Preparation for Next Week

- Register UTS Service
- Download SNOMED
- Load SNOMED into your MySQL database (try)
  1. Create database
  2. Set environment
  3. Load data (remember to replace the correct path to your SNOMED files!)

# Take Home Message

- Simple R syntax and coding style
- RMySQL with SQL syntax
- DB design: entity, relationship, attribute, constraints, normalization, schema
- Contact
  - [Github repository](#)
  - [ckbjimmy@gmail.com](mailto:ckbjimmy@gmail.com)
  - [Linkedin: Wei-Hung Weng](#)