

# Ch 7: Basic Data Modeling

Mike Lynott  
Adjunct Prof, CS  
Boise State University (v3)

# Agenda

- Introduction
- Why Data Modeling?
- How We'll Do Data Modeling
- What You'll Learn To Do
- What Can You Do With A Data Model?

# Introduction

- Born and raised in NE PA
- Education: Boston College, U of Kansas
- Worked/Studied in

West Australia	Canada
London, UK	former USSR

- Worked for

Computer HW companies	Non-IT Industries (e.g., Albertsons)
Computer SW companies (incl. Oracle)	Consulting (many industries)

# Why Data Modeling?

- To understand some subject
- To depict that understanding in graphical form
- To share that understanding with others
- To use in various business tasks

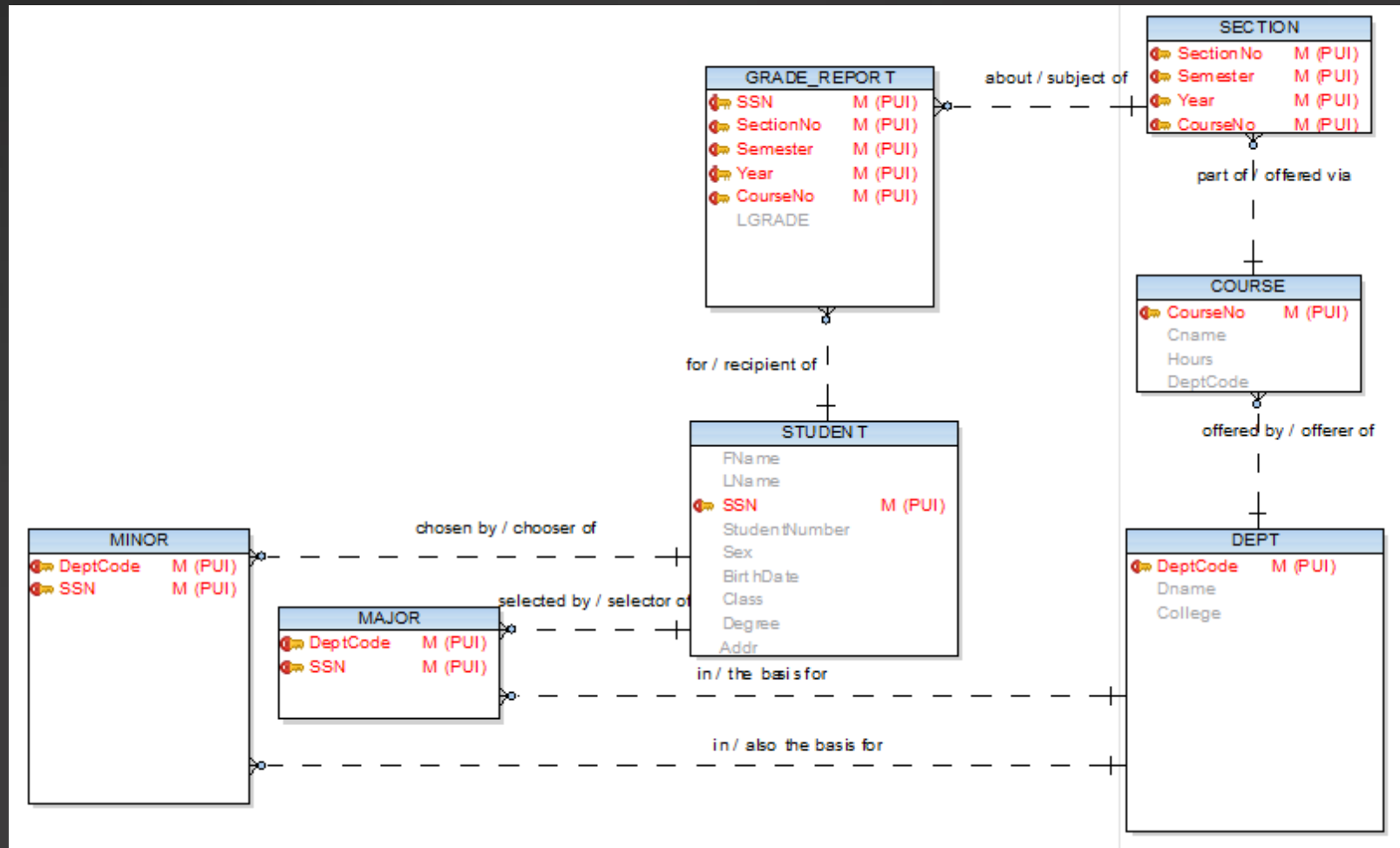
# Why Data Modeling?

To understand some subject

- An abstract concept, e.g., the “Standard Model” of nuclear physics
- An industry
- An enterprise (real or imagined)
- A (future) project
- An existing database (via automated reverse engineering)
- An existing application whose data structures aren't available (via manual reverse engineering)

# Why Data Modeling?

To depict your understanding in graphical form(1)



# Why Data Modeling?

To depict your understanding in graphical form(2)

- What are the parts of an ER diagram?
- A box: representing an entity (type)
- Terms within the box: representing attributes
- Lines between boxes: representing relationships

# Why Data Modeling?

To depict your understanding in graphical form (2)

- We will use formats available in the commonly – used modeling tools:
  - Toad Data Modeler (Free to students)
  - ERWin
  - Oracle Designer
- At work sites you may come across
  - Power Designer
  - System Architect
  - Rational Rose



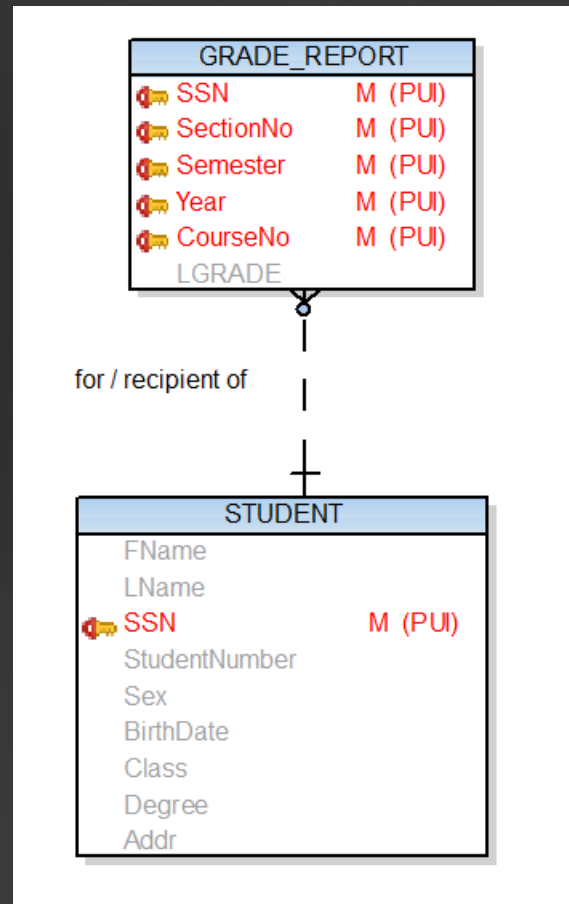
# Why Data Modeling?

To depict your understanding in graphical form (3)

- Why so many formats?
  - Attempt to build a “Unique Selling Proposition”
  - Big egos
  - Strongly-held opinions
  - Geography
  - The state of technology
- Expect to encounter some->all of these at work sites
- Anchor yourself in one tool/format then stay flexible and use the tools you’re provided

# Why Data Modeling?

To share your understanding with others (2)



# Why Data Modeling?

To share your understanding with others

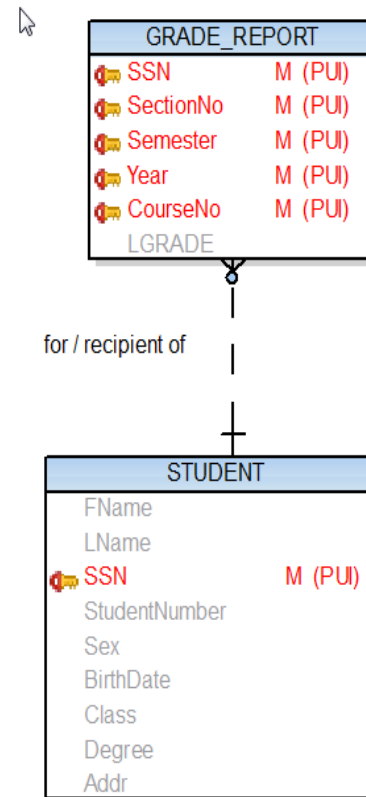
- The common (not recommended) options
  - The universal “has” / “have”
  - “There’s a M:1 relationship between A and B”
  - Plural sentence: “People rent cars”
- Recommended: a structured English sentence
  - Focus on the entity in the singular
  - Use English prepositions or prepositional phrases
  - State optionality (“must” or “may”)
  - State cardinality (“one” or “more than one”)
  - Make a strong, deniable assertion

# Why Data Modeling?

To share your understanding with others

“Each”

<Subject>	
“must be”	If there is a “ ” (or “—”) next to the Object
or	
“may be”	If there is an “o” next to the Object
<Relationship>	
“one-and-only-one”	If there is no crow’s foot (“>” or “v”) next to the Object
or	
“one or more”	If there is a crow’s foot (“>” or “v”) next to the Object
<Object(s)>	



# To Use In Various Business Tasks

- To develop a new application
  - Database design
  - Online transactions
  - Reports
  - Data Imports and Exports
- To divide a planned application into phases
- To explain an existing application
- To revise an existing application
- To support a “root cause” analysis

# How we We'll Do Data Modeling In This Class

## Objectives

- Use Information Engineering diagram format
- Read a data model
- Identify errors in a data model
- Modify a data model
  - Correct a relationship
  - Add an attribute
- Translate a data model to a db design
- Note: To learn to develop a complete data model, take the data modeling course!

# Exercise

- With a person sitting near you: read through (aloud) each of the relationships on the model of the university database.
- It may seem childish—it's not. I expect you to master this way of reading models, and reading them out loud is a good way to practice.

# Data Modeling Formats

- The text uses the NIAM format for data models; we will not be using this format.
- I have never encountered an organization that uses the NIAM format; few tools support it.
- Most books on data modeling use some variant of the IE format, or UML.
- My personal favorite is the Barker-Ellis format, but it is not widely supported. (Barker was my data modeling teacher and mentor.)



# Terms from the textbook

- The following terms from the textbook can be ignored
  - Complex attribute
  - Composite attribute
  - Entity Set
  - Multivalued attribute
  - Relationship Set
  - Weak entity

# What, then, is an entity?

- “A thing of significance about which information must be known or held.” (R. Barker)
  - A “thing”: often abstract
  - Which of these is an entity?
    - George Washington
    - US President
    - Government employee
    - Conservative
    - Columnist
    - Person

# More on entities

- Levels of abstraction

- Specific thing, or individual (George Washington)
- A role played by that individual (US President, government employee)
- A named, comprehensive set (Person)

- Categories

- Conservative
- Columnist

# More on entities (2)

- Very likely entities
  - Who
  - What
  - Where
  - When
- Often used as entities (often inappropriately)
  - Category: Student, SUV
  - Role: Dependant, Spouse
- Never entities: single person, place, or thing

# More on entities (3)

- So far we've looked at what are called "reference entities": often boxes with no crow's feet attached
- Now we add "associative entities": M:N relationships which include attributes, and usually have 2+ crow's feet attached
- On the university database
  - Reference entities: student, dept
  - Associative entity: grade\_report

# Quiz on entities

- Which of the following are entities?
  - Hotel Room Type
  - Person
  - Account
  - The MEC building
  - Boise State University
  - Professor
  - County
  - Ada County
  - The Oscar Mayer Wienermobile

# Quiz on entities

- Which of the following are entities?
  - Hotel Room Type Y
  - Person Y
  - Account Y
  - The MEC building N
  - Boise State University N
  - Professor ?
  - County Y
  - Ada County N
  - The Oscar Mayer Wienermobile (See next page)

# How Could This Be a Non-Entity?

## WIENERMOBILE





# Relationships

- “What one thing has to do with another; or, any significant way in which two things of the same or different type may be associated.” (R. Barker)
- The term “significant” has been used in the definitions of both an entity and a relationship.
  - Significant to whom? The “enterprise” that is the subject of the model, e.g., your employer or client
  - Significant how? Important enough that information about it needs to be retained by the enterprise.

# More on Relationships

- On university model these are some *possible* “insignificant” relationships
  - The grade report and the weather on the day the grade report was submitted
  - The location from which the professor submitted the grade report
  - The student’s vehicle’s license plate number
  - The number of scheduled days on which the section did not meet during the semester

# Review of relationships

- Re-examine the university model: are there any insignificant relationships on this model?

# Attributes

- “a quality ascribed to a person or thing; a characteristic quality.” (R. Barker)
- “a discrete, atomic piece of information that identifies, describes, classifies, or measures an entity.” (D. Hay)

# Alert!

- The next slides use culturally insensitive terms in order to demonstrate their cultural insensitivity.
- Apologies in advance to anyone who might be offended.

# Attribute Quiz

- First name
- Last name
- Christian name
- Middle name
- Confirmation name
- Preferred name
- Address
- ZIP Code
- State

# Attribute Quiz

- First name- okay
- Last name-okay, with warnings
- Christian name-culturally offensive
- Middle name-consider Prince Charles
- Confirmation name-culturally insensitive
- Preferred name-my friend “Boo”
- Address-actually many attributes
- ZIP Code- generically, Postal Code
- State- see ISO standard

# Is it ever okay...

- To combine two meanings in one attribute?
  - No
  - If you do, your colleagues and successors may hunt you down and work their revenge on you
- In one attribute, to have one meaning for one kind of row, and another meaning for another kind of row, e.g., one meaning where Sex = 'M' and one where Sex = 'F'?
  - No (for Yes, see above)



# Is it ever okay... (2)

- To combine multiple values in the same attribute? (e.g., names of children?)
  - No
- To reuse identifiers?
  - No
  - (I once explained to a CFO that doing so would materially affect the cost of operating the IT department for a decade. The word “materially” is a very meaningful word to accountants.)

# It's always a good idea...

- To check ISO (and related) standards to see whether an abbreviation or number has already been assigned.
- To check your industry's standards to see whether a standard has been established (e.g., the Diagnostics and Statistics Manual—DSM, in medicine).
- To check your enterprise's standards to see whether a standard has been established (e.g., building names).

# It's always a good idea... (2)

- To THINK BIG!: One client's staff had suffered through the agony of "running out of numbers." I made the ID for their main entity a huge number. I explained that there would be a number available for every millisecond since the Big Bang. Most seemed satisfied.

# It's ALMOST never a good idea

- To haggle over string and number lengths, e.g., trimming numbers or strings down to the smallest number possible.
  - We're not saving space on punched cards any more
  - Disk space used to be expensive—it isn't now!
  - The database will handle many space optimization issues.

# Resist!

- “Meaningful” codes—e.g., first four digits is the company, the next three are the product line, and the last five are the product. This is how UPCs were devised, and companies have been reusing UPC values for years.

# IBM Model 84 Card Sorter

THE system to handle positional values

- This is the device that handles “a C in column 1” and “the state code in columns 2-3”.

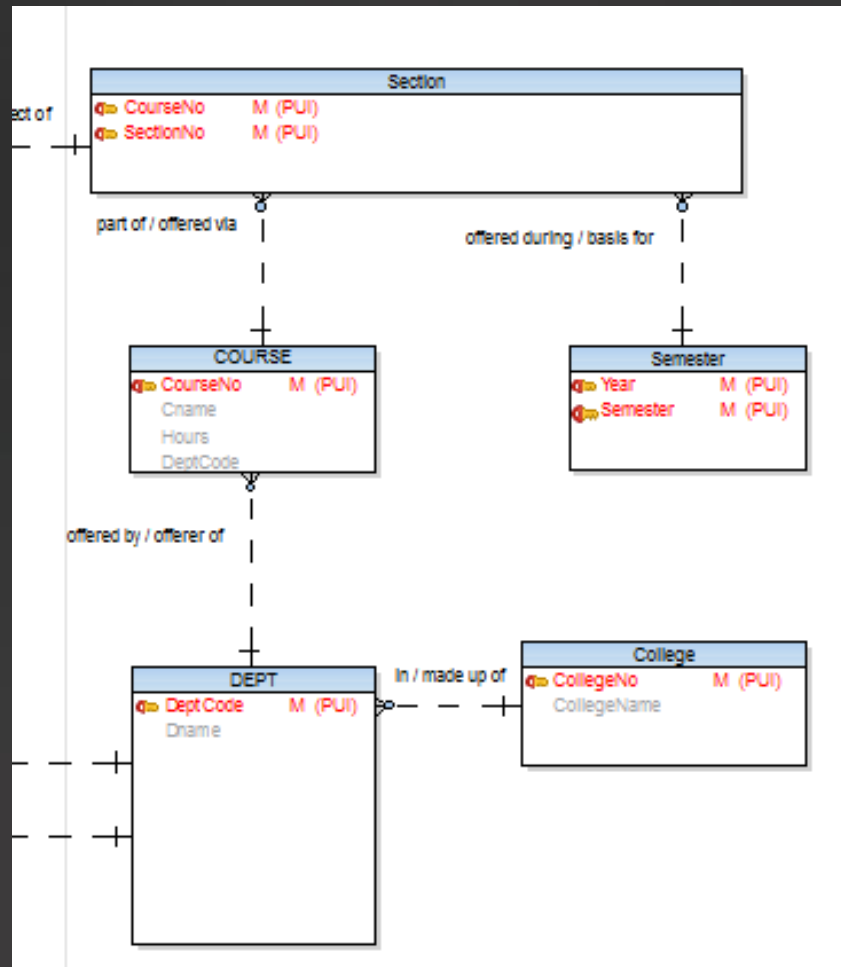


# Some “attributes” aren’t really

- There are “attributes” that hide
  - Entities
  - Relationships
- Are there any examples in the university model?

# A revised (partial) university model

May or may not be the “final” model

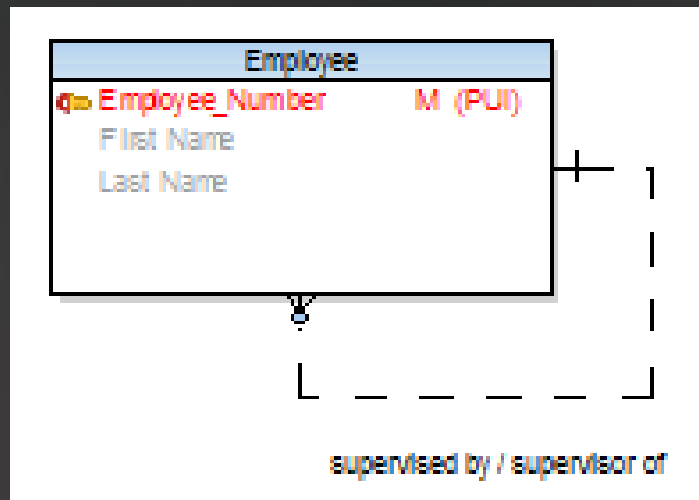




# Recursive relationships

Also known as “pig’s ears”

- Each employee may be supervised by one-and-only-one (other) employee
- Each employee may be the supervisor of one or more (other) employees



# Constraints

- Unique Identifier
  - Set of attributes
  - Added attribute
- Terms and abbreviations used
  - UI: Unique Identifier
  - UID: Unique Identifier
  - PUI: Primary Unique Identifier

# Examples of “UI”/“UID”/“PUI”

From the university model

- MAJOR: DeptCode +SSN
- MINOR: same as MAJOR
- GRADE\_REPORT: almost everything
- DEPT: DeptCode
- Student: SSN

# Mike's Preference

- Use an artificial key—for just about everything
  - Additional attribute
  - Numeric value
  - Meaningless number
  - Often generated by the dbms or os

# The “SSN” Issue

- There is a security and compliance issue with the use of SSN
- Several states (e.g., NY, CA) forbid the display of SSN (e.g., on an employee’s badge)
- Stolen SSN can be a factor in identity theft
- So don’t use SSN as a Primary Unique Identifier in a system—your organization should have a policy that states this specifically
- <https://www.privacyrights.org/my-social-security-number-how-secure-it>

# Social Security Administration

## Steps to avoid

- Never list an SSN when posting a paper record on a public site
- Never send SSNs via an [unencrypted] electronic format
- Never have a computer log-in system where a person has to use their SSN
- Never use SSNs on ID cards
- Never store SSNs on unprotected computer systems
- Never carry an SSN card with you

# SSN Issue: more information

- For more information, see
- <https://www.privacyrights.org/my-social-security-number-how-secure-it>

# Back to Constraints

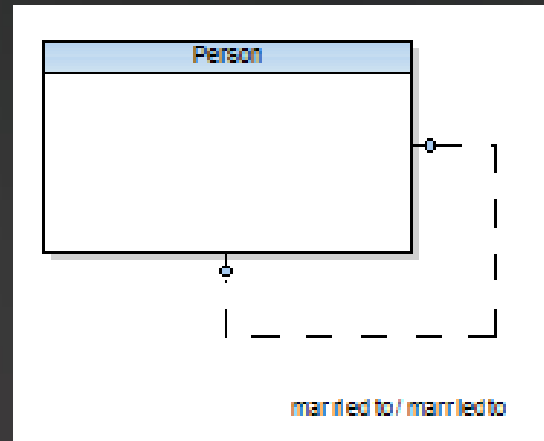
- Other Unique Identifiers (sometimes “AK”)
  - Combination of columns whose value must be unique (e.g., if you assign arbitrary employee numbers, the SSN still must be unique.)
- Other constraints
  - Combination of columns (one or more) whose values must meet certain rules (e.g., graduation date must be greater than enrolment date)



# Apparent 1:1 relationships

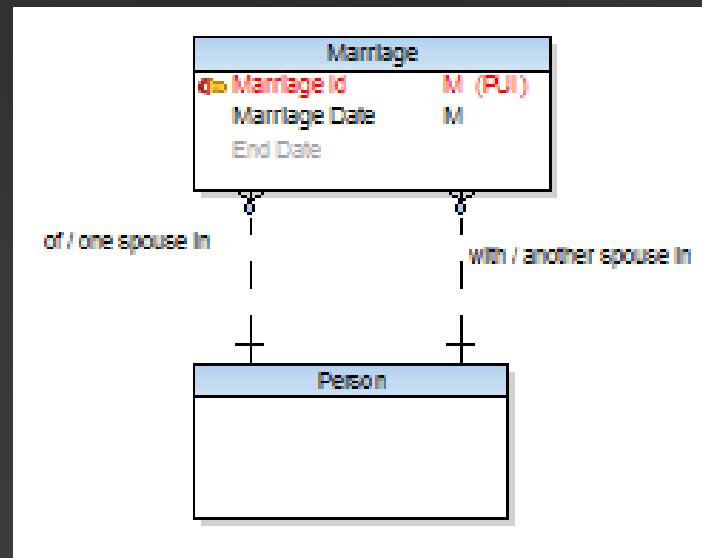
## That really aren't 1:1

- Each person (1) may be married to one and only one spouse.
- Each person (2) may be married to one and only one spouse.



# Apparent 1:1 relationships

- VERY rare—make that two “very”s
- Test each one on a diagram
- Ask about history: can a person be married more than once?



# Attributes and their domains

- Domain: “ A set of business validation rules, format constraints, and other properties that might apply to a group of attributes.” R. Barker
- The definition for “domain” in my dictionary has 8 options/variants. It’s a term used in many ways.
- Domains in E/R Modeling are not related to
  - the Domain Name System (DNS)
  - a domain of knowledge (e.g., cost accounting)

# Domain examples

- English three-letter abbreviations for days of the week
- Positive integers  $< 1000$
- Colors in the “2014 Royal Horticultural Society Large Colour Chart” (916 named colors)
- Degrees of latitude
- RGB color value (0-255)

# Domains and E/R Models

- Domains are not (usually) shown on an E/R diagram.
- They are usually available in modeling tools, where they can be very useful.

# In-class Exercise

- Finish the diagram in the handout (based on the diagram Figure 7.20 in the text)
  - Question entity names
  - Add relationships where needed
  - Revise relationships where appropriate
  - Question attributes—e.g., find “hidden” entities
  - Question Unique Identifiers

# Summary

- Entity
- Relationship
- Attribute
- Unique Identifier
- Domain