- HW Due Today!
  o Questions, Clarification?
- HDSC: [https://eng.uber.com/dsw/](https://eng.uber.com/dsw/)
- Data Representation Models
  o Structure vs. Semantics
  o Model vs. Schema
  o Data Model:
    § Constructs
    § Constraints
    § Operations
  o Example Data Models
  o Data independence vs. Physical data independence
- Entity-Relationship (ER) and Relational Model
  o ER Model:
    § Entities and their Attributes
    § Relationship between entities
    § Relationships also have attributes
  o Types of relationships
  o Relations as representation of E's and R's
- Keys
  o Super Key
  o Candidate Key
  o Primary Key
  o Foreign Key
    § Constraints
  o Examples
- Tidy Data
  o Nycflights13 is a tidy dataset
- SQL: operations over relations
  o SFW
  o Group-by summarize
  o Subqueries
Model vs. Schema

Entities: Flights
Attributes: 
- Arr Delay: Discrete numeric
- Dep Delay:
- Origin: Categorical Ordered

---

Model:
- Constructs
- Constraints
- Operations

→ XML, JSON, Protocol Buffers, Avro
Entity - Relationship Model

Relational Model

→ Physical Independence

Entity - Relationship Model

- Entities: Attributes
- Relationship: Attributes

D: entities
**Lesson Plan 2/8**
Thursday, February 8, 2018
2:01 PM

**HW Due Today! Questions, Clarification?**

- HDSC: [https://eng.uber.com/dsw/](https://eng.uber.com/dsw/)

**Data Representation Models**
- Structure vs. Semantics
- Model vs. Schema

**Data Model:**
- Constructs
- Constraints
- Operations

**Example Data Models**

**Data independence vs. Physical data independence**

- **Entity-Relationship (ER) and Relational Model**
  - **ER Model:**
    - Entities and their Attributes
    - Relationship between entities
    - Relationships also have attributes
  - **Types of relationships**
  - **Relations as representation of E’s and R’s**

- **Keys**
  - Super Key
  - Candidate Key
  - Primary Key
  - Foreign Key

- **Constraints**
  - Examples

**Tidy Data**
- Nycflights13 is a tidy dataset

**SQL:**
- [SFW](https://eng.uber.com/dsw/)
  - Group - by summarize
  - Subqueries
  - Baseball examples

---

**Professor (ssn, name, area, rank)**

**Chair (prof, dept)**

**One-to-one:**
HW Due Today!

Questions, Clarification?

HDSC: https://eng.uber.com/dsw/

Data Representation Models

Structure vs. Semantics

○ Model vs. Schema

Data Model:

§ Constructs

§ Constraints

§ Operations

○ Example Data Models

○ Data independence vs. Physical data independence

Entity-Relationship (ER) and Relational Model

ER Model:

Entities and their Attributes

§ Relationship between entities

§ Relationships also have attributes

○ Types of relationships

○ Relations as representation of E’s and R’s

Keys

Super Key

○ Candidate Key

○ Primary Key

○ Foreign Key

Constraints

○ Examples

Tidy Data

Nycflights13 is a tidy dataset

SQL: operations over relations

SFW

○ Group — by summarize

○ Subqueries

○ Baseball examples

Lesson Plan 2/8

Thursday, February 8, 2018

2:01 PM
Relation - sets of entities
- Entity: attribute values

Professor (123-456-7890, "Joe", "Acct")

Attributes

Keys

<table>
<thead>
<tr>
<th>name</th>
<th>category</th>
<th>price</th>
<th>name</th>
</tr>
</thead>
</table>
Entities: How entities are defined
- Unique within a relation
- Set of attributes that define entities uniquely

Key
- Super Key: set of attributes which contain a subset of attributes that define entities uniquely

- Minimal Key: Super key that can't be minimized (can't remove any attribute)

- Candidate Key: Any minimal key

- Primary Key: A single arbitrarily chosen candidate key

Why keys?)

- Performance: Independent entities
Define relationships

Relationship

\[ R : E_1 \times E_2 \] (cross product)

- primary key
- foreign key
- primary key

Foreign key: primary key that appears in another relation

Foreign key constraint: primary key must appear in relationship
Data Representation Models

Structure vs. Semantics

Model vs. Schema

Data Model:
- Constructs
- Constraints
- Operations

Example Data Models

Data independence vs. Physical data independence

- Entity-Relationship (ER) and Relational Model
  - ER Model: Entities and their Attributes
  - Relationship between entities
  - Relationships also have attributes
  - Types of relationships
  - Relations as representation of E's and R's

- Keys
  - Super Key
  - Candidate Key
  - Primary Key
  - Foreign Key

Constraints

Examples

Tidy Data
Nycflights13 is a tidy dataset

SQL: Structured Query Language

Declarative

```
SELECT * FROM team WHERE salary > 1000000;
```

```
CREATE TABLE Player (
  id INT,
  name VARCHAR(255),
  teamID INT,
  salary DECIMAL
);
```

```
ALTER TABLE Player ADD CONSTRAINT fk_team FOREIGN KEY (teamID) REFERENCES Team(id);
```

```
CREATE VIEW TeamView AS
  SELECT * FROM Team;
```

Lesson Plan 2/8
Thursday, February 8, 2018
2:01 PM
Data Representation Models
Structure vs. Semantics
- Model vs. Schema
  - Data Model: Constructs
    - Constraints
    - Operations
  - Example Data Models
- Data independence vs. Physical data independence
  - Entity-Relationship (ER) and Relational Model
    - ER Model: Entities and their Attributes
      - Relationship between entities
      - Relationships also have attributes
  - Types of relationships
    - Relations as representation of E's and R's
- Keys
  - Super Key
  - Candidate Key
  - Primary Key
  - Foreign Key
- Constraints
  - Examples
- Tidy Data
  - Nycflights13 is a tidy dataset
  - SQL: operations over relations
    - Group by summarize
    - Subqueries
    - Baseball examples
  - Describes the resulting relation