MIS2502:
Review for Exam 2

JaeHwuen Jung
jaejung@temple.edu
http://community.mis.temple.edu/jaejung

Overview

- **Date/Time:** Wednesday, Mar 28, in class (50 minutes)
- **Place:** Regular classroom

Please arrive 5 minutes early!

- Multiple-choice and short-answer questions
- Closed-book, closed-note
- No computer

Coverage

Check the **Exam 2 Study Guide**

Not every item on this list may be on the exam, and there may be items on the exam not on this list.

SQL Out: LIMIT

- Example: Which order has the highest order amount? (Display the order ID and amount, and assume there is no tie)

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

OrderID	Amount
10308	100

```
SELECT OrderID, Amount
FROM MyDB.Order
ORDER BY Amount DESC
LIMIT 1;
```

SQL Joins

- Used to combine two or more tables, based on the common fields between them.
- Suppose we have….

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

MyDB.Customer

CustomerID	LastName	FirstName	Country
1	Futterkiste	Alfreds	Germany
2	Trujillo	Ana	US
3	Moreno	Antonio	Mexico

A Correct, Simple Join

SELECT * FROM MyDB.Order, MyDB.Customer

WHERE Order.CustomerID=Customer.CustomerID;

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

MyDB.Customer

CustomerID	LastName	FirstName	Country
1	Futterkiste	Alfreds	Germany
2	Trujillo	Ana	US
3	Moreno	Antonio	Mexico

OrderID	CustomerID	OrderDate	Amount	CustomerID	LastName	FirstName	Country
10308	2	1996-09-18	100	2	Trujillo	Ana	US
10309	1	1996-09-19	10	1	Futterkiste	Alfreds	Germany
10310	1	1996-09-20	63	1	Futterkiste	Alfreds	Germany

What If We Don't Have the WHERE condition?

SELECT * FROM MyDB.Order, MyDB.Customer

~~WHERE Order.CustomerID=Customer.CustomerID~~;

It will fetch **every possible combination (pair)** of records from the two tables

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

MyDB.Customer

CustomerID	LastName	FirstName	Country
1	Futterkiste	Alfreds	Germany
2	Trujillo	Ana	US
3	Moreno	Antonio	Mexico

3×3 = 9 rows

OrderID	CustomerID	OrderDate	Amount	CustomerID	LastName	FirstName	Country
10308	2	1996-09-18	100	1	Futterkiste	Alfreds	Germany
10308	2	1996-09-18	100	2	Trujillo	Ana	US
10308	2	1996-09-18	100	3	Moreno	Antonio	Mexico
10309	1	1996-09-19	10	1	Futterkiste	Alfreds	Germany
10309	1	1996-09-19	10	2	Trujillo	Ana	US
10309	1	1996-09-19	10	3	Moreno	Antonio	Mexico
10310	1	1996-09-20	63	1	Futterkiste	Alfreds	Germany
10310	1	1996-09-20	63	2	Trujillo	Ana	US
10310	1	1996-09-20	63	3	Moreno	Antonio	Mexico

More Variations to Join

Q: What is the total order amount by country?

Step 1: We start with a simple join

```
SELECT *
FROM MyDB.Order, MyDB.Customer
WHERE Order.CustomerID=Customer.CustomerID;
```

OrderID	CustomerID	OrderDate	Amount	CustomerID	LastName	FirstName	Country
10308	2	1996-09-18	100	2	Trujillo	Ana	US
10309	1	1996-09-19	10	1	Futterkiste	Alfreds	Germany
10310	1	1996-09-20	63	1	Futterkiste	Alfreds	Germany

Step 2: We then end up with -

```
SELECT Customer.Country, SUM(Order.Amount)
FROM MyDB.Order, MyDB.Customer
WHERE Order.CustomerID=Customer.CustomerID
GROUP BY Customer.Country;
```

Country	SUM(Amount)
US	100
Germany	73

SQL (Subselects)

- Subselect query can return
  - One single value (one column, one row)
  - A temporary table (one or multiple columns, one or multiple rows)

One single value: Used With Comparison Operators

- SELECT column_name1
  FROM schema_name.table_name1
  WHERE column_name2 comparison_operator
  (SELECT column_name3
  FROM schema_name.table_name2
  WHERE condition);

Treated as a single value

comparison_operator could be equality operators such as =, >, <, >=, <=, <>.

Subselect as One Single Value: Example 1

Q: What are the order IDs with order amount below average?

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

- Step 1. We start by write the subselect query that returns the average order amount:

```
SELECT AVG(Amount) FROM MyDB.Order
```

Amount
57.6667

- Step 2. We treat the subselect query as **a single value**:

```
SELECT OrderID, Amount FROM MyDB.Order
        WHERE Amount <
                (SELECT AVG(Amount) FROM MyDB.Order);
```

OrderID	Amount
10309	10

Subselect as One Single Value: Example 2

Q: Which German customer placed the order(s) with the highest order amount?

MyDB.Order

OrderID	CustomerID	OrderDate	Amount
10308	2	1996-09-18	100
10309	1	1996-09-19	10
10310	1	1996-09-20	63

MyDB.Customer

CustomerID	LastName	FirstName	Country
1	Futterkiste	Alfreds	Germany
2	Trujillo	Ana	US
3	Moreno	Antonio	Mexico

Recall the simple join query…

```
SELECT * FROM MyDB.Order, MyDB.Customer
WHERE
Order.CustomerID=Customer.CustomerID;
```

OrderID	CustomerID	OrderDate	Amount	CustomerID	LastName	FirstName	Country
10308	2	1996-09-18	100	2	Trujillo	Ana	US
10309	1	1996-09-19	10	1	Futterkiste	Alfreds	Germany
10310	1	1996-09-20	63	1	Futterkiste	Alfreds	Germany

Subselect as One Single Value: Example 2

Q: Which German customer placed the order(s) with the highest order amount?

- Step 1. We start by write the subselect query that returns the highest order amount for German customers:

```
SELECT MAX(Order.Amount)
FROM MyDB.Order, MyDB.Customer
WHERE Order.CustomerID=Customer.CustomerID
AND Customer.Country='Germany';
```

Amount
63

- Step 2. We treat the subselect query as **a single value**:

```
SELECT Customer.LastName, Customer.FirstName, Order.Amount
FROM MyDB.Order, MyDB.Customer
WHERE Order.CustomerID=Customer.CustomerID
AND Customer.Country='Germany'
AND Amount=      (SELECT MAX(Order.Amount)
FROM MyDB.Order, MyDB.Customer
WHERE Order.CustomerID=Customer.CustomerID
AND Customer.Country='Germany');
```

LastName	FirstName	Amount
Futterkiste	Alfreds	63

As a temporary table

- SELECT column_name(s)
  FROM

    (SELECT column_name(s)
    FROM schema_name.table_name2
    WHERE condition)

  WHERE condition;

Treated as a table

Subselect as Temporary table

Q: How many states are there in the customer table?

Recall the Customer table:

CustomerID	FirstName	LastName	City	State	Zip
1001	Greg	House	Princeton	NJ	09120
1002	Lisa	Cuddy	Plainsboro	NJ	09123
1003	James	Wilson	Pittsgrove	NJ	09121
1004	Eric	Foreman	Warminster	PA	19111

Subselect as Temporary table

Q: How many states are there in the customer table?

- Step 1. We start by write the subselect query that returns the unique states in the table :

```
SELECT DISTINCT State FROM  orderdb.Customer;
```

State
NJ
PA

- Step 2. We treat the subselect query as **a temporary table**:

```
SELECT COUNT(*)
FROM (SELECT DISTINCT State FROM
orderdb.Customer) AS tmp1;
```

2

SQL (CREATE TABLE)

If there is no foreign key:

```
CREATE TABLE schema_name.table_name (
columnName1 datatype [NULL][NOT NULL],
columnName2 datatype [NULL][NOT NULL],
PRIMARY KEY (KeyName);
```

If there is foreign key:

```
CREATE TABLE schema_name.table_name (
columnName1 datatype [NULL][NOT NULL],
columnName2 datatype [NULL][NOT NULL],
PRIMARY KEY (KeyName),
FOREIGN KEY (ForeignKeyName) REFERENCES
schema_name.reference_table_name(ReferenceKeyName));
```

When to use NOT NULL?

DROP TABLE, ALTER TABLE

DROP TABLE schema_name.table_name;

Drops a table

ALTER TABLE schema_name.table_name
ADD COLUMN column_name datatype
[NULL][NOT NULL];

Adds a column to the table

or

ALTER TABLE schema_name.table_name
DROP COLUMN column_name;

Removes a column from the table

or

ALTER TABLE schema_name.table_name
CHANGE COLUMN old_column_name
new_column_name datatype
[NULL][NOT NULL];

Changes a column in the table

INSERT INTO, UPDATE, DELETE FROM

- To insert a record into a table:

```
INSERT INTO schema_name.table_name
(columnName1, columnName2, columnName3)
VALUES (value1, value2, value3);
```

- To change data in a row:

```
UPDATE schema_name.table_name
SET columnName1=value1, columnName2=value2
WHERE condition;
```

What to use in the WHERE condition?

- To delete a row from a table:

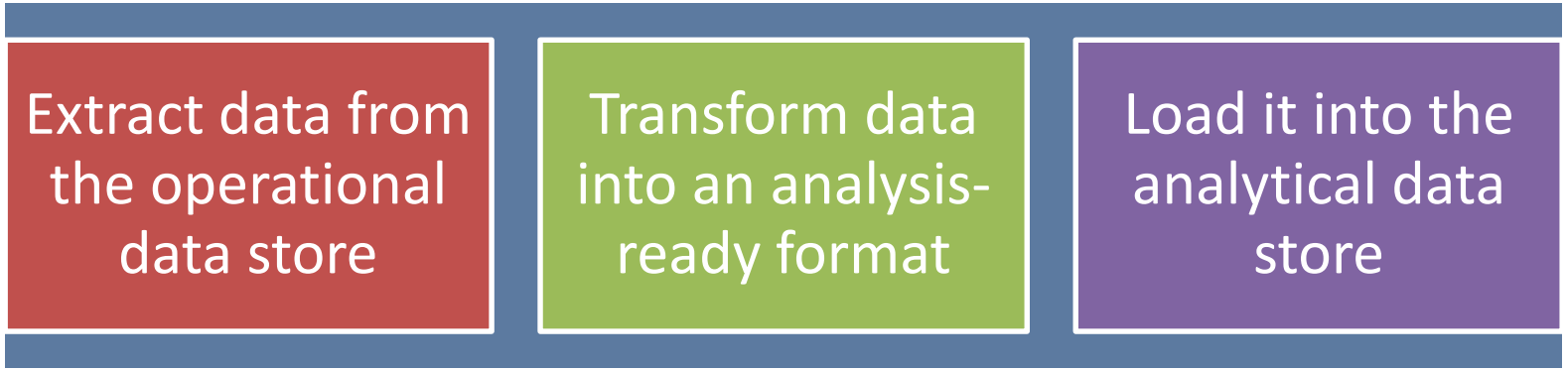
```
DELETE FROM schema_name.table_name
WHERE condition;
```

How about this WHERE condition?

Data Types

Data type	Description	Examples
INT	Integer	3, -10
DECIMAL(p,s)	Decimal. Example: decimal(5,2) is a number that has 3 digits before decimal and 2 digits after decimal	3.23, 3.14159
VARCHAR(n)	String (numbers and letters) with maximum length n	'Hello, I like pizza, MySQL!'
DATETIME, DATE	Date/Time, or just Date	'2011-09-01 17:35:00', '2011-04-12'
BOOLEAN	Boolean value	0 or 1

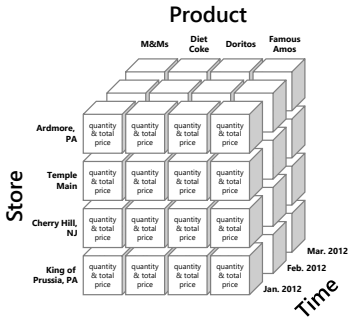
ETL and Assignment 4



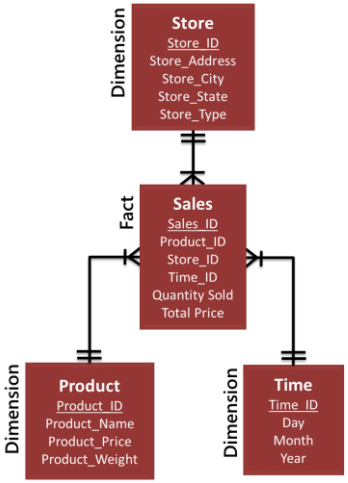
- What is it? Why is it important?
  - Data consistency
  - Data quality
- Explain the purpose of each component (Extract, Transform, Load)
- Excel functions **VLOOKUP** and **CONCATENATE**

Dimensional Data Modeling

- Data warehouse vs data mart vs data cube
- Data Cube



- Star schema



- Kimball's four step process for data mart design

1. Choose the business process
2. Identify the fact
3. Decide on the level of granularity
4. Identify the dimensions

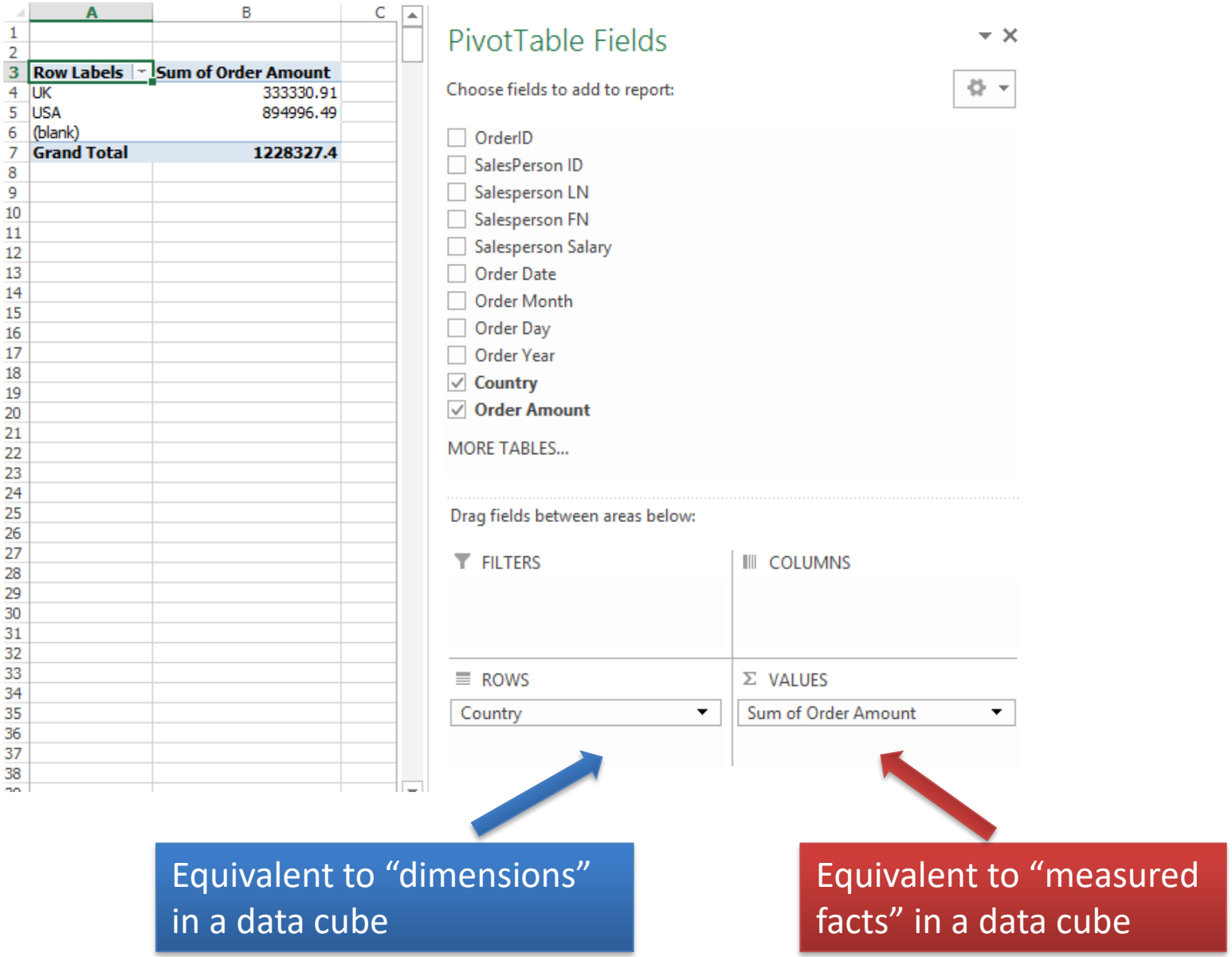
- "non-volatility" of data cubes

Pivot Tables

Given a question about a set of data, be able to identify the fields required to create a pivot table

- Identify which fields are assigned as VALUES and which ones are assigned as ROWS
- Identify the correct function for aggregation: i.e., SUM, COUNT, AVERAGE

Pivot Tables vs. data cubes



Row Labels	Sum of Order Amount
UK	333330.91
USA	894996.49
(blank)	
Grand Total	**1228327.4**

PivotTable Fields

Choose fields to add to report:

- OrderID
- SalesPerson ID
- Salesperson LN
- Salesperson FN
- Salesperson Salary
- Order Date
- Order Month
- Order Day
- Order Year
- **Country**
- **Order Amount**

MORE TABLES...

Drag fields between areas below:

FILTERS | COLUMNS

ROWS: Country | VALUES: Sum of Order Amount

Equivalent to “dimensions” in a data cube

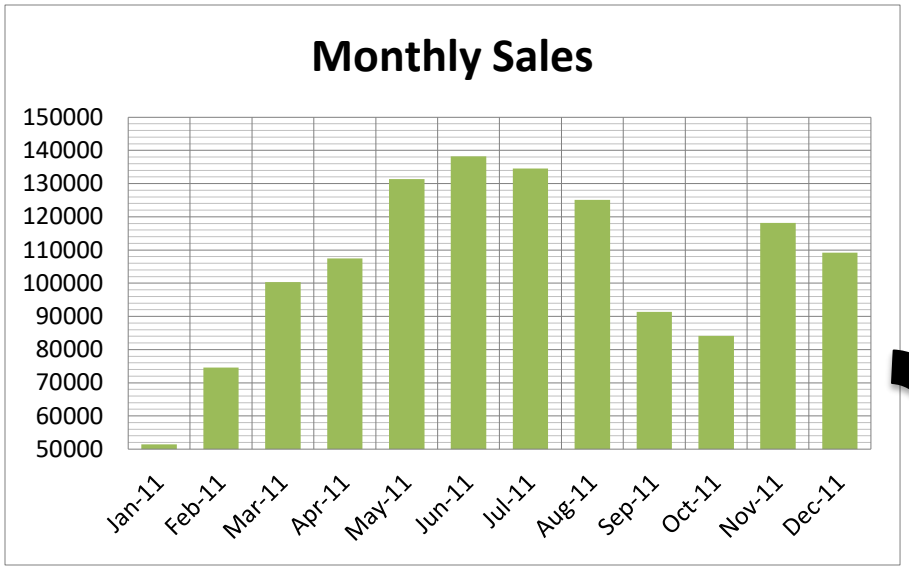
Equivalent to “measured facts” in a data cube

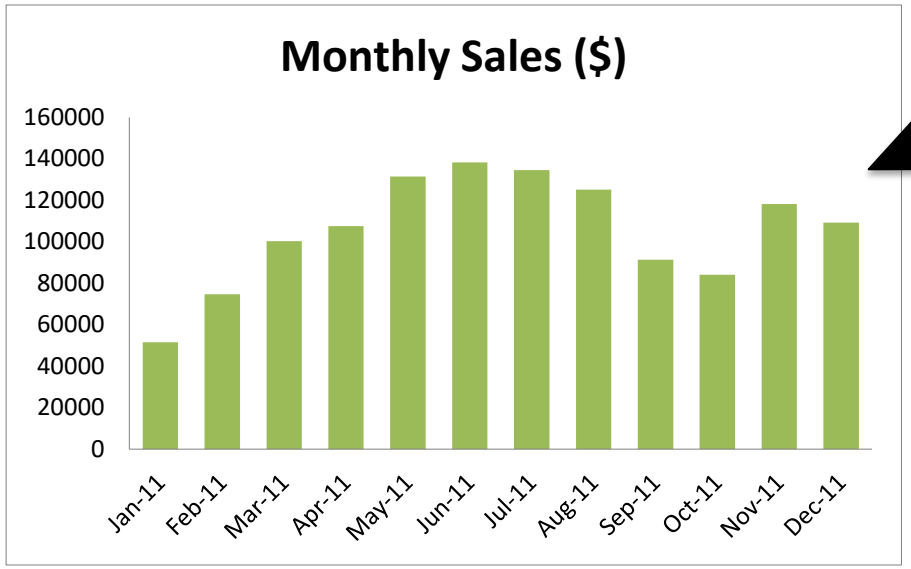
Data Visualization

- Data visualization principles.
  - Tell a story
  - Graphical integrity (lie factor)
  - Minimize graphical complexity (data ink, chartjunk, Moiré effects)

Issues with this chart?

Chart #1:





Issues:

- Tell a Story
  - The vertical axis isn't labeled. We don't know the unit.
- Graphical Integrity
  - The vertical axis does not start from zero
- Graphical Complexity
  - Horizontal and vertical lines are unnecessary (Chartjunk)

Good Luck!