

## Introduction to Access and SQL

### Announcement: CIMBA (10/16)

Location: Rome Ballroom

Time: 10/16/2018, Tuesday 10/16, 5:30 – 8:00 pm.

Dress: Business Casual

As usual, for students who cannot come, please write a report on “The Importance of Information Technology at Today’s Workplace” (2-3 pages) to get the credits.

### What else can Excel do?

A VBA Example:

Use the “player - report.xlsx”

- 1) Select Cell A1
- 2) Choose Developer → “Record Macro”, give the macro a name, whatever, and keep other options default and click OK
  - a. if you do not have “Developer Tab”, go to File → Options → Customize Ribbon → Click “Developer”
- 3) Select “Use Relative References” under “record macro”<sup>1</sup>
- 4) Select the whole first row and copy it
- 5) Select the whole third row and “insert copied cells”
- 6) Select Cell A3
- 7) Click “Stop Recording”
- 8) Open the macro, select “step into”, you can find the codes such that:

```
Sub your_macro_name ()
'
' your_macro_name Macro
'
'
'
ActiveCell.Rows ("1:1") .EntireRow.Select
Selection.Copy
ActiveCell.Offset (2, 0) .Rows ("1:1") .EntireRow.Select
Selection.Insert Shift:=xlDown
ActiveCell.Select
End Sub
```



You can run this macro by click

<sup>1</sup> On previous versions of the EXCEL for MAC you could easily create a Macro by clicking on the "Record Macro" button. After this, a button was available to allow you to use either RELATIVE or ABSOLUTE reference. On the Excel 2016 for MAC, this button is no longer available, what makes creating a Macro a hard task for the ones that are not used to VBA programming. From this perspective, EXCEL 2016 for MAC seems to be LESS USER FRIENDLY than the previous version.

one hundred times.

Or you can revise the codes by adding a loop such that

```
Sub your_macro_name ()
'
' your_macro_name Macro
'
'
For i = 1 To 100
    ActiveCell.Rows("1:1").EntireRow.Select
    Selection.Copy
    ActiveCell.Offset(2, 0).Rows("1:1").EntireRow.Select
    Selection.Insert Shift:=xlDown
    ActiveCell.Select
Next i
End Sub
```

### The importance of coding

Almost all software (e.g., Excel, Stata, SPSS, JMP) has integrated a set of commands (or a combination of manipulations, or a series of actions, i.e. macro). However, you can only run the macro that the software provides. Suppose you want to use some prediction model in SPSS, but you want to make some small changes on the model, then you cannot do this with the embedded macros from the software. You have to create your own macros. That's why coding is very important. If you only know how to use the software with some "clicks", it is as with you are learning a language from a "sentence level" but not "word level". You can only speak the "existing" sentences but cannot create your own.

Excel Skill	Education Level
Input Data	Elementary
Simple Functions: if, vlookup	Junior
Pivot table, Complicated figures, combination of different functions	High School Level
VBA	College level

VBA is just a language like R, python, C++.

[https://en.wikipedia.org/wiki/Visual\\_Basic\\_for\\_Applications](https://en.wikipedia.org/wiki/Visual_Basic_for_Applications)

It is a language embedded in Excel.

What will be covered in the second half of this class?

Access: SQL

- 1) Database (Access, SQL)
- 2) Data manipulation (SQL, R)
- 3) How to do some simple data analysis (optional)

How will I teach?

- Case study

One Assignment for credits, other assignments for practice but no need to submit

Final Exam, Access SQL, R (could be optional)

### How to describe a table

Two main elements 1) Column name (header, head) 2) Contents, Records (type of columns, length of column)

Question:

Which column can uniquely define a row?

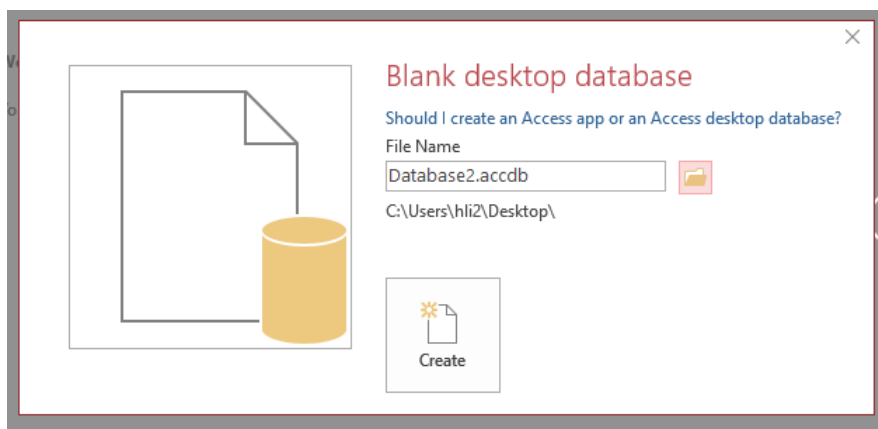
### **Primary Key**

A primary key (PK) is a column or several columns in the table, which allow to uniquely identify each record in that table. By having a PK, we make sure that there are no cases when two or more exact same records are present. Consequently, we need to make sure that the values of the PK in each table are unique (non-repeating) and never null (non-empty). There must be one and only one PK in each table.

- 1) There can never be two rows exactly the same
- 2) Examples for primary key: code, ID, driver license, social security number, serial code, PeopleSoft, NetID, telephone number, email (*can physical address be a PK?*)
- 3) Can be one column, can be a combination of several columns, can never be empty

### Import an Excel File to Access<sup>2</sup>

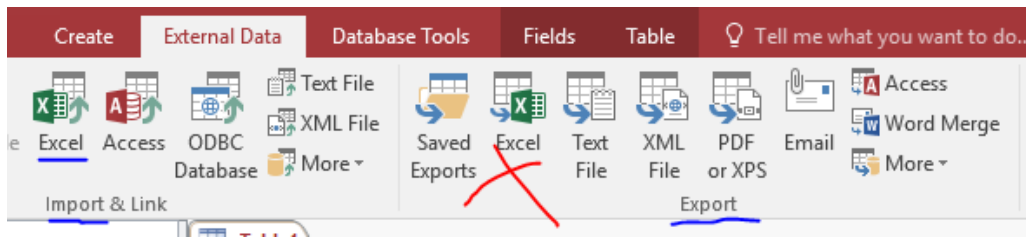
1) Launch Access and select “Blank Desktop Database”, then choose a location for your Access file. Choose a folder where it is convenient to find.



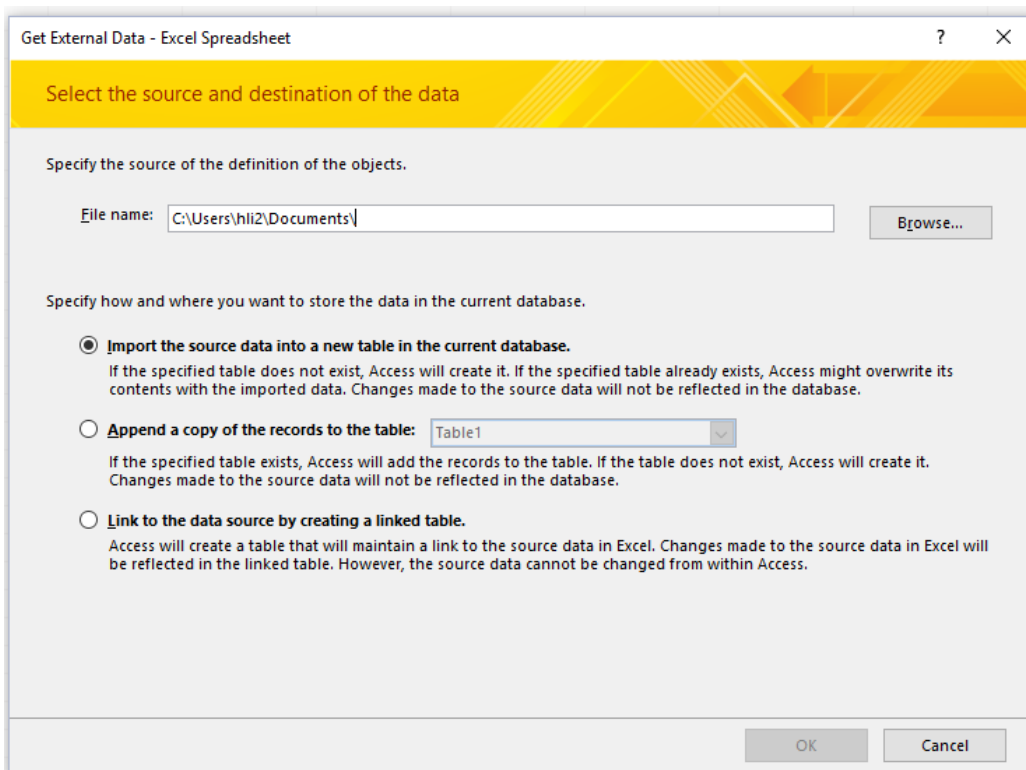
---

<sup>2</sup> Unfortunately, there is not Access for Mac. I guess this could be a competing strategy for Microsoft, making Mac uneasy to use in some software but not core software such as Word and Excel.

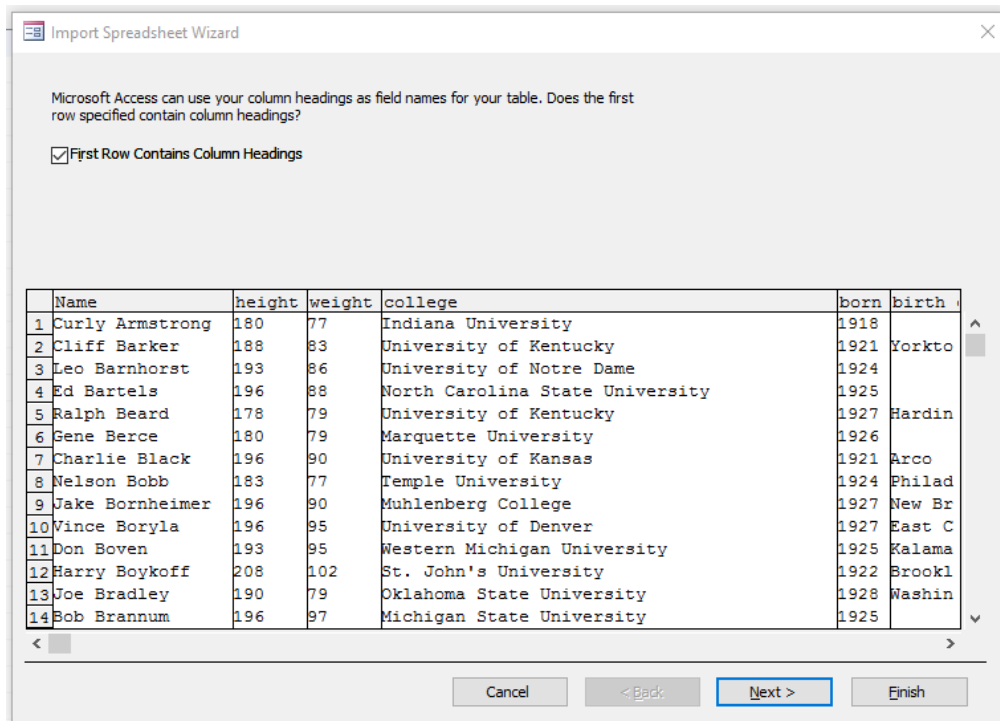
2) Then choose “External Data”, import Excel (do not choose Export)



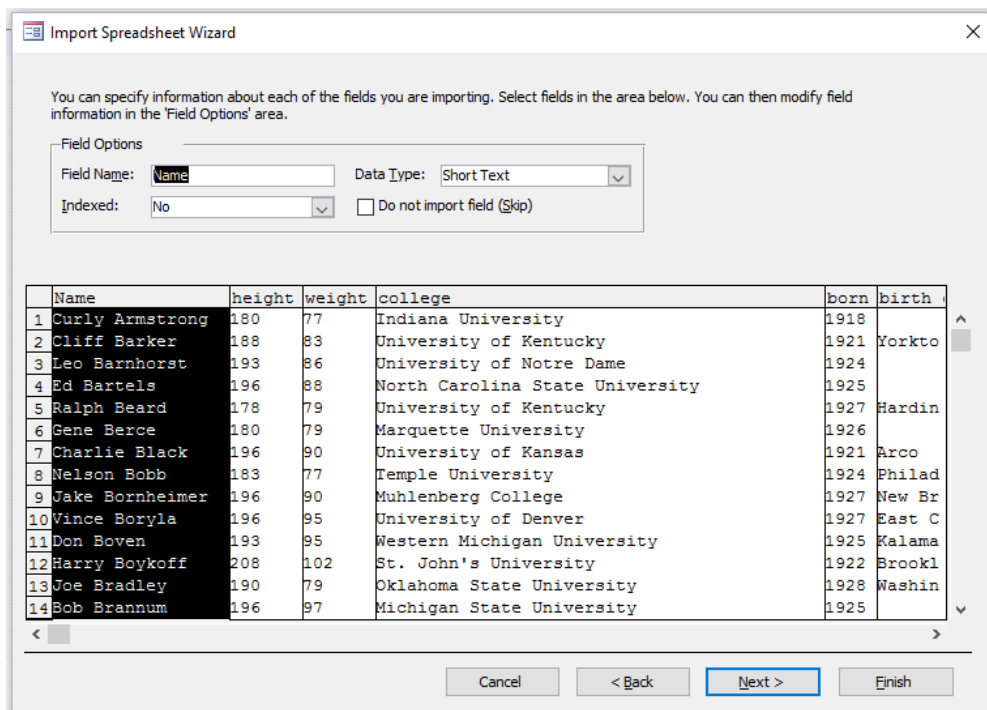
3) Browse the locations and find the Excel file you want to import, and keep other options default



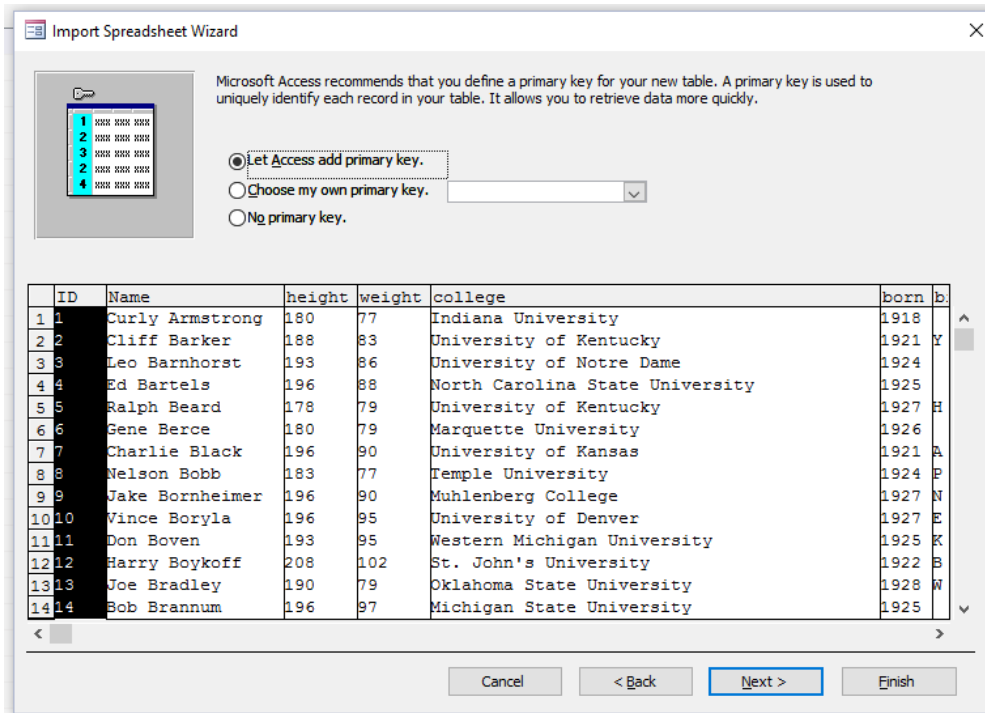
## 4) Keep “First Row Contains Column Headings” clicked



## 5) You can change the name and data type here and let's keep it default and move on.



6) Then we need to check the primary key. A safe way is to “Let Access add primary key”, which is always correct. You can also choose your own primary key if you know which column’s records are definitely unique.



7) Then you have successful imported an Excel file into Access.

Announcement: Cimba Event

The second half of this class will mainly include two sessions:

## 1 Database

- 1) the basic knowledge of database
- 2) How to write queries with SQL, how to use Access to generate queries

## 2 Data Manipulation

- 1) How to aggregate data by R, which is similar to SQL group by, you will enhance your understanding on database by this session
- 2) Data Analysis (Optional, depends on time)

We have been talking about arrangement of data for the whole semester, but how will we use data

- Regression
- Factor Analysis
- Simulation
- Forecasting

Of course, I want you can learn as much as possible, but teaching is not about how much I teach to you, but about how much you learn from me.

SQL/Access and R will take up most parts in Final exam.

There will be two assignments. You can submit either of them. If you submit both, I will choose the higher-grade one as your assignment grade.

### How will I teach?

Case Study – Learn by answering questions

I will give you the original table before class on HuskyCT. Then I will ask some questions and show what I want. Then let's do it together to get the results we want.

1. Access (Queries)
2. Database theories
3. Data clean, filter, manipulation (SQL, R Data.table)

### Incentive - Why is coding important?

NBA dataset <https://www.kaggle.com/drgilermo/nba-players-stats>

Example of Excel-Report.

```
Sub play_record()
```

```
ActiveSheet.Range("A1").Select
```

```
For i = 1 To 100
```

```
    ActiveCell.Rows("1:1").EntireRow.Select
```

```
    Selection.Copy
```

```
    ActiveCell.Offset(2, 0).Rows("1:1").EntireRow.Select
```

```
    Selection.Insert Shift:=xlDown
```

```
    ActiveCell.Select
```

```
Next i
```

```
End Sub
```

Try by yourself: Paint the odd number columns with yellow for the first 100 rows

Skill	Excel Education Level
Input data	Elementary
Simple Functions such as If, vlookup	Junior

Pivot Table, Complicated Graph, Combinations of Different Functions	High School
Excel VBA	College
Create Some Embedded Functions	Graduate

Essentially, software integrate a series of computer command. However, the number of commends is limited no matter how powerful the Excel is.

Think about a language. For English, you learn “words” and “grammar” so that you know the reasonable combinations. If you learn English only from sentences, you can never create your own language.

Due to the class limitation, we will not cover VBA. However, if you find a job which needs you to use Excel a lot, you need to spend some time to learn some VBA because most of the work in company is duplicated. If it is duplicated but not creative, computer can do much better than human being.

### Question:

Which field/column can be a unique identifier for a row? Which field makes this row different from all the other rows?

A **primary key**, also called a **primary** keyword, is a **key** in a relational database that is unique for each record. It is a unique identifier, such as a driver license number, telephone number (including area code), or vehicle identification number (VIN). A relational database must always have one and only one **primary key**.

Which one can be the primary key of this table?

Last Name	First Name	DoB	PeopleSoft	School	Department	Grade

More Examples: find the primary key for each table.

## Session 1 Basic Knowledge of Table and SQL-SELECT

### 1.1 Elements of a Table

See an example “Student” table.

Question: How to change a table to its design view? Generate two elements.

### Practice

We can describe every table as combined by two elements: 1) Field name (first row) 2) Data Type (other rows)



Field name	Data Type
Last Name	Text
First Name	Text
DoB	Date
PeopleSoft	Number
School	Text
Department	Text
Grade	Text

Usually the first row of a table express the field names and the second row and below include all the data.

Use “Player.xlsx”, answer these questions

1. Return all the information in this table
2. Return the first 10 rows of this table
3. Return only the “college” column of this table
4. Return the “college” column without repetition
5. Return the number of rows of this table

Practice:

1. Return only “Name” column
2. Return top 5 rows
3. Return birth\_state without repetition

SQL stands for Structured Query Language

### **1 Select Specific Columns**

Grammar: **SELECT** *column1, column2, ...***FROM** *table\_name*;

Select \* from players;

Select top 10 \* from players;

select distinct college from players;

**SELECT DISTINCT** *column1, column2, ...***FROM** *table\_name*;

**SELECT COUNT(DISTINCT** collage) **FROM** Customers;

**SELECT Count(\*) AS** DistinctColleges

**FROM (SELECT DISTINCT** Country **FROM** Customers);