DATABASE MANAGEMENT SYSTEM

What is a Database?

A database is a collection of information that's related to a particular subject or purpose, such as tracking customer orders or maintaining a music collection. If your database isn't stored on a computer, or only parts of it are, you may be tracking information from a variety of sources that you have to coordinate and organize yourself.

For example, suppose the phone numbers of your suppliers are stored in various locations: in a card file containing supplier phone numbers, in product information files in a file cabinet, and in a spreadsheet containing order information. If a supplier's phone number changes, you might have to update that information in all three places. In a database, however, you only have to update that information in one place — the supplier's phone number is automatically updated wherever you use it in the database.

Indeed a database is a collection of logically related data files which is integrated and organized so as to provide s single comprehensive file system. Database system design is concerned with the problem of organizing the database, accessing data on it, using it, updating it and providing output (reports or responses to file interrogations) to a variety of users; e. g. a students' file, Marks file, Payment's file. The Marks' file will comprise records of marks scored by the students.

Elements of a Database File

A file is a collection of logically related records; e.g students file, stock file.

A record is a collection of logically related data fields; e. g Data relating to students in students file. In a database table records are usually in rows.

A field is consecutive storage position of values. It is a unit of data within a record e. g student's number, Name, Age. In a database table fields are usually in columns.

S/NO	NAME	ADDRESS	AGE
1	ABDULLAHI YUSUF	13, JOS ROAD	45
2	OLADELE AYINDE	5, IKEJA ROAD	35
3	NWACHUKWU OJI	24, ENUGU ROAD	40

4 Fields

3 Records

Figure 6.1: A Database table showing fields and records

Database Management System refers to software that 'builds, manages and provides access to a database'. Examples are Microsoft Access, FoxPro, Clipper, Oracle, Foxbase

What is a Relational Database?

A relational database is a powerful tool which can help you store and manage large amounts of information of various kinds. Databases enable you to organize and reorganize information, quickly and flexibly retrieve information, and print quite a variety of reports.

The beauty of a relational database system is that you can store related data in separate tables. Then through defining relationships between the tables you can retrieve the information in the tables to use for making queries and writing reports. The four main elements of a relational database include

- **TABLES** for storing data
- **FORMS** for entering and viewing data on the screen
- **QUERIES** for extracting specific information from the data (asking questions)
- **REPORTS** for printing query results

Microsoft Access

Microsoft Access is a relational database product shipped with some of the more expensive versions of Microsoft Office. It comes with an impressive list of features that make it relatively easy for non-programmers to create little database-enabled applications.

Access databases are stored in files that use the `.mdb` extension (for Microsoft DataBase). When the Access program is running, you will also see a temporary file with the same name as the database but with the extension `.ldb`. You can safely ignore that file.

One thing you'll notice when you use Access is that you do not save the file as you might in Word or Excel. Instead, every change made to the data in the database is saved immediately (although changes to the objects in Access sometimes require you to click the Save icon).

Databases themselves are made up of files. To make the database, records made up of fields are entered by the database owner or by people specially trained to do data entry. Fields are discrete pieces of information, such as a name, a journal name or publisher name. Fields are made up of types of data which limits the information which can go into them. Fields may be text, an autonumber, a number, a date, a memo and so on. Data in the database can be sorted and the order of fields can be re-arranged without affecting the actual data. Reports and labels can be generated through generating an ordered list of fields and records.

One more important concept to keep in mind is that when you are creating a database, you must strive for complete uniqueness. In other words, no data in your database should be redundant. Redundant information lends itself to errors, plus it adds to the size of the database unnecessarily and is hard to keep up to date.

Planning Your Access Database

Before we begin the design process, we need to know what kind of database we are going to be creating in today's workshop and we need to have a clear idea of what we are going to use the serials database for.
We are going to be designing an Access database to track the ordering and management of the periodicals we receive in our resource centres. The database we are going to be creating is a relatively simple database but it does have a enough interesting features to make the design process interesting.
Planning a database is a lot harder and more time consuming than creating a PowerPoint presentation for an eight hour workshop. But planning is something which must be done carefully and with much consideration. Do not assume you can start creating tables and their relationships without having done some serious thinking ahead of time. It is much better to thoroughly plan the database and make your mistakes ahead of time than to have to redo the database later on because the design is poor. It is not very difficult to change the structure of a table once you have entered data into it, but it is better to avoid making any changes once you have entered data into a table.
It is inevitable that you will need to make some changes. Be sure to back up your database before you make changes to the table structures just to be on the safe side. By the way, we will use several terms when we speak of journals. We may use the word journal, periodical or serial. All three terms pretty much mean the same thing, although the format might be slightly different.
Another term used in public libraries is magazine. Examples of journals include the New England Journal of Medicine, Science, Nature, Milbank Quarterly, Medical Care, and so on.
Journals appear in a number of formats including hard copy, CD-ROM, and virtual (on the World Wide Web).

Common Mistakes

A common mistake people make is to create one large table with all the fields in it. Use Access to create many tables and the relationships between the tables. Another common mistake is to create a table in which the same information is repeated in each table. For example, you do not need to repeat the name of the person in each table. Use the relationships between the person and any other information to link information in two tables together. Remember not to repeat information and avoid redundant data. Do not type in anything more than you have to.

General Questions to Ask Yourself Before You Begin

There are a number of questions you need to ask yourself before you start.
Why do I think I want a database? Does the information I have lend itself to fields and records? (Most information lends itself to some kind of structure). What kind of data do I have? What is the best way to organize it? (These questions help you define your tables).
If I need to enter data into a table, how can I simplify the process to make it as easy as possible for someone to do data entry? (This question helps you define the forms you may need).
When I have the data entered into the database, what kinds of questions do I think I'll want to ask about it later on? (This question will help you define the queries and reports you will want to make based on your data).

Collecting Information on the Database

Be sure to identify and talk to potential users, even if you spend your days working in the area of the database you're going to create. You would be surprised as how often you forget important elements. Be certain to record the suggestions. Professionals who create databases for a living generally use Requirement Collection Forms to capture requests for data elements from the people they interview.

The Design Process

- Determine the purpose of our database
- Determine the tables we need for the database
- Determine which fields we will need
- Determine how each of the tables is related
- Refine and redefine the design

Determine the Purpose of our Database

Before we do anything else, we need to ask the following questions: What do we need to know from our database and what will this information allow us to do?

What do we need to know from our database?

We'll need to know a lot. This database is going to track a lot of information and should be able to answer any serials-related questions we have.

Relationships

Once you have built tables, you need some method of bring the information in the tables back together. You need a relationship between the tables and those relationships need to be defined. Once you get the relationships built, you can create queries, and forms and reports from several tables at once.
Building relationships is frustrating even if you understand the theory of building entity relationship diagrams. (We will not be covering the theory for this material). For those who have not been trained in computer science, it can sometimes be a nightmare. As mentioned elsewhere, I spend a good deal of my time gnashing my teeth, muttering under my breath, and even crying tears of frustration as I try to figure out what the relationships are. (They are as difficult as human relations and about as complex).

Microsoft Help says that a "relationship works by matching data in key fields, usually a field with the same name in both tables. In most cases, these matching fields are the primary key from one table, which provides a unique identifier for each record, and a foreign key in the other table".

Types of Relationships

You can create several types of relationships between tables:

- one to many
- one to one
- many to many

One to Many Relationship

A one-to-many relationship is the most common type of relationship. In a one-to-many relationship, a record in Table A can have many matching records in Table B, but a record in Table B has only one matching record in Table A. For example, one publisher can publish many journals, but it's highly unlikely that a journal would have more than one publisher. Similarly, one bindery could receive many volumes each year to be bound, but each volume would only go to one bindery.

One to One Relationship

In a one-to-one relationship, each record in one table can have only one matching record in a second table, and each record in the second table can have only one matching record in the first table. This type of relationship is fairly uncommon, because most information related in this way would be in one table. We do not have any examples of a one to one relationship in our serials database.

Many to Many Relationship

In a many-to-many relationship, a record in one table (table 1) can have many matching records in a second table (table 2), and a record in table 2 can have many matching records in table 1. This type of relationship is only possible by defining a third table (called a junction table) whose primary key consists of two fields - the foreign keys from both Tables 1 and 2. A many-to-many relationship is really two one-to-many relationships with a third table.
We have one example of a many to many relationship in our serials database.

ACCESS WORKSPACE

Using the Menu bar

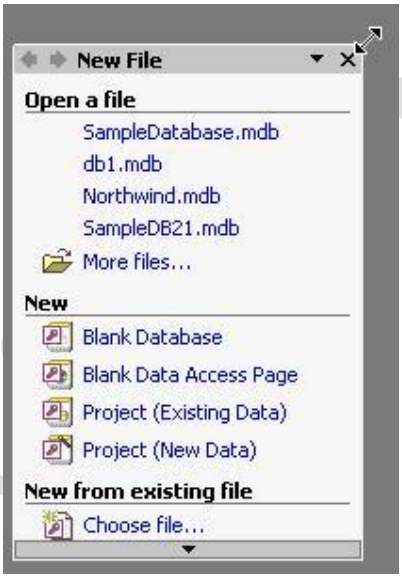
- The **Menu bar** contains all of the **Access** menu options. With these menu options, you can access most of the **Access** features. The options available in each menu often depend on the object you are working with and the mode of **Access** .
- By default, the **Menu bar** appears at the very top of the screen. However, you can place this bar wherever you want on the screen.
- To move the **Menu bar**, find the small gray lines at the far left of the menu bar. Place your mouse on these lines. The mouse indicator changes to a four-arrow symbol. This means you can drag the bar to your desired location:

Using the Database toolbar

- The **Database toolbar** is a bar of buttons that act as shortcuts to many of the Access features that are most often used. This toolbar is very similar to other applications' **Standard toolbar**.
- By default, the **Database toolbar** appears at the top of the screen, just below the **Menu bar**. Like the **Menu bar**, though, you can place this bar wherever you want it on the screen.
- To move the **Database toolbar**, find the small gray lines at the far left of the Database toolbar. Place your mouse on these lines; the mouse indicator changes to a four-arrow symbol. This means you can drag the bar to your desired location:

Using the Task Pane

- The **Task pane** is a window that acts like a dialog box, except you can keep it on the screen as you work. Depending on the task it relates to, it provides access to different features.
- Access provides three **Task panes**: **New File**, **Clipboard**, and **Search**. You can change the type of task pane by clicking the down arrow in the top right corner of the **Task pane**, the three types are then available on a drop-down menu.
- One example of a **Task pane** is when you create a new database file (by choosing **File > New**). The various options for creating a new database file are then provided in the pane.
- As with toolbars, you can move the **Task pane** around the screen. Do this by dragging its title bar.
- If the **Task pane** is too small for you, you can resize it. Do this by positioning the mouse indicator over a side edge or corner of the **Task pane**. The mouse indicator changes to a two-sided arrow, which means you can drag it to make the pane bigger or smaller:



Using the Status bar

- The **Status bar** appears at the bottom of the screen. It provides status information. For example, if you are inputting information into a table, the status bar might tell you what mode you are in or what column you are currently applying changes to.

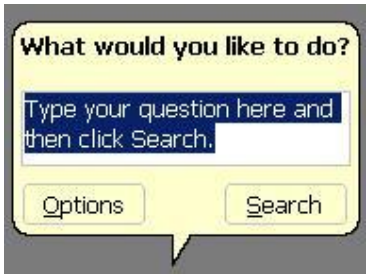
- The following image shows the **Status bar** when you have the Database window open:

- The following image shows the **Status bar** when you have a Design view open:

- Unlike the other toolbars, you cannot move the **Status bar** around the screen.

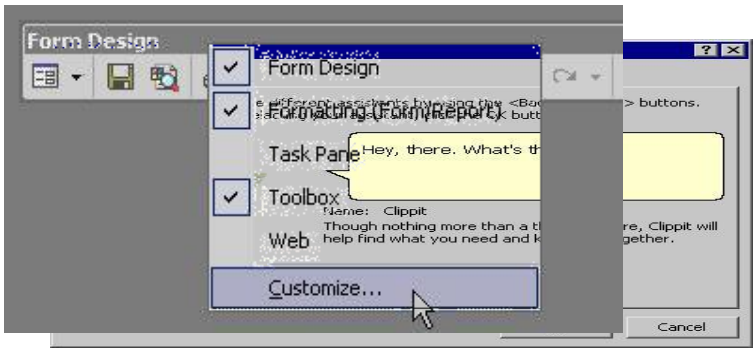
Using the Office Assistant

To open the Office Assistant:
From the menu bar, choose **Help > Show the Office Assistant**. The **Office Assistant** is a **Microsoft Help** feature. It works by asking you questions. If you need help with a certain task, or if you need an explanation on a particular feature or concept, you can simply type your request in the **Office Assistant** dialog box:



To ask the Office Assistant a question:
- In the text box of the **Office Assistant**, type a phrase.
- Click **Search**.
- If the **Office Assistant** finds information related to that phrase, it shows you the location in Access 's Help file. If it finds no matches, it tells you, and you can type a different phrase:

To change Office Assistant options:
- In the **Office Assistant** dialog box, click **Options**.
- In the resulting dialog box, use the **Gallery** tab to change the image for the Office Assistant:



- Use the **Options** tab to configure the behavior settings for the **Office Assistant**:

- Click **OK** to apply the changes and close the dialog box.

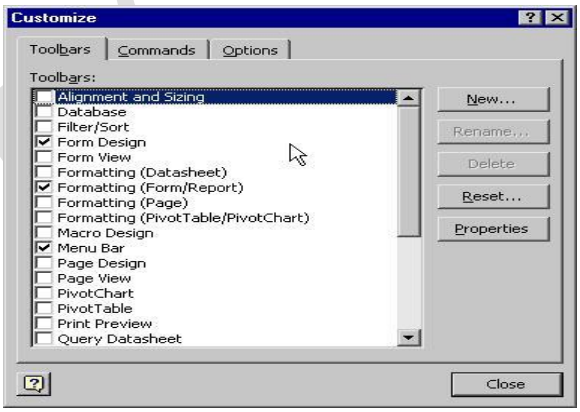
To hide the Office Assistant:

- From the **Help** menu, choose **Hide the Office Assistant**. You can also right-click the **Office Assistant** image and choose **Hide** on the context menu.

Customizing the toolbars

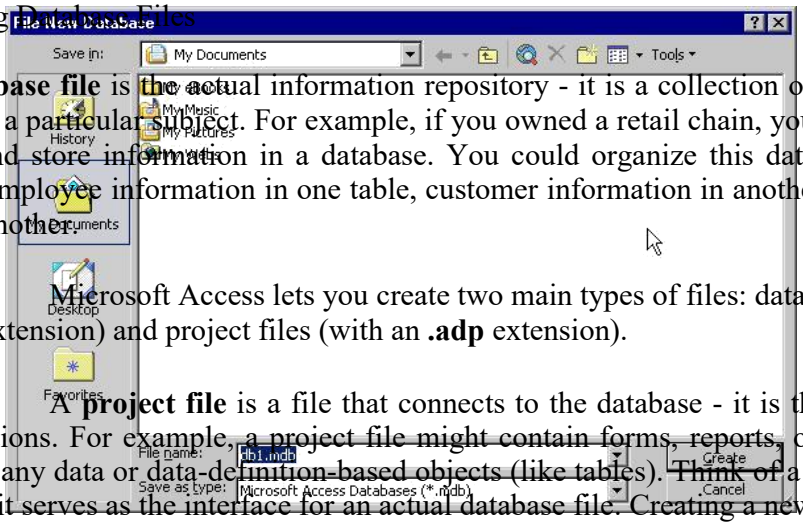
- You can right-click any toolbar and choose **Customize** on the context menu to customize the appearance and behavior of Access 's toolbars:

In the **Customize** dialog box, use the three tabs to specify what toolbars should appear and how you want the commands to work:



- When the **Customize** dialog box is showing, you can right-click any button or menu option to customize names, button images, properties, and so on:

Creating Database Files



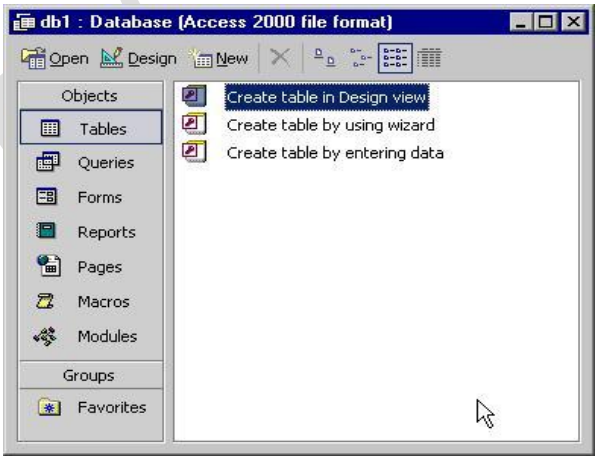
A **database file** is the actual information repository - it is a collection of related data that has to do with a particular subject. For example, if you owned a retail chain, you might want to place all sales and store information in a database. You could organize this data into tables; you could group employee information in one table, customer information in another, and store information in yet another.

- Microsoft Access lets you create two main types of files: database files (with an **.mdb** extension) and project files (with an **.adp** extension).

- A **project file** is a file that connects to the database - it is the basis for client/server applications. For example, a project file might contain forms, reports, or modules; it would not contain any data or data-definition-based objects (like tables). Think of a project as an "in-betwee n" file: it serves as the interface for an actual database file. Creating a new database file
- From the menu bar, choose **File > New *OR*** press the **Ctrl + N** key combination to open the **New File** task pane which appears on the right of your screen:

· Click **Blank Database**. In the **File New Database** dialog box, specify a location and a name for your new database file. Then click **Create**:

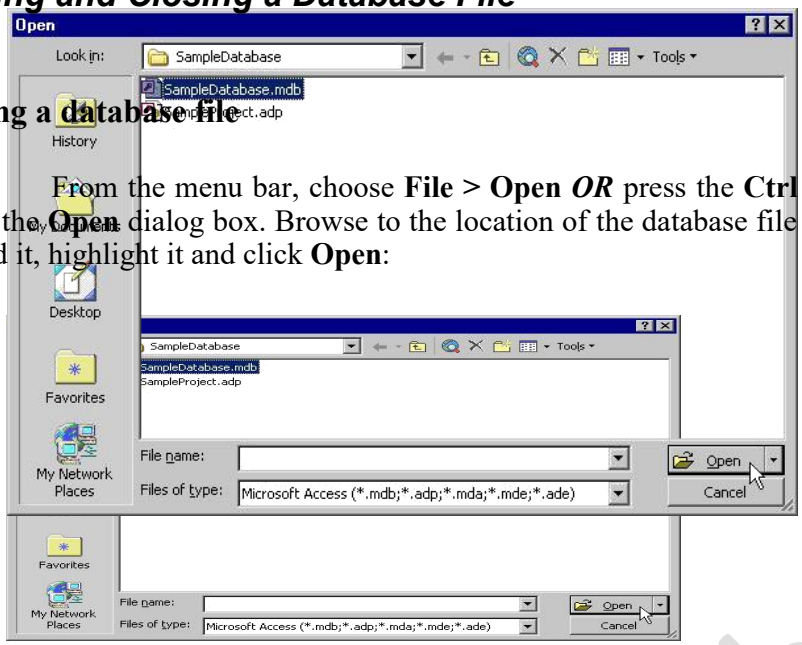
- The **Database** dialog box appears to let you start adding objects and information to your database. We'll deal with these features in a future section of this course:



Opening and Closing a Database File

Opening a database file

- From the menu bar, choose **File > Open** ***OR*** press the **Ctrl + O** key combination to display the **Open** dialog box. Browse to the location of the database file you want to open. When you find it, highlight it and click **Open**:



Note: You can also double-click the file to open it.

Closing a database file

- From the menu bar, choose **File > Close**.

Using the Open a File Area of the New File Task Pane

Opening the New File task pane

- From the menu bar, choose **File > New**.
- The **New File** task pane opens as a window on the right side of the screen.

Viewing and opening recent files

- The **Open a File** area of the **New File** task pane shows recently opened Access files:

- If the list does not include the file that you want to open, click **More Files**.
- The **Open** dialog box opens, which lets you browse to the location of a file, specify its name and type, and open it:

To open a file in the list, simply double-click its name.

Using the New Area of the New File Task Pane

Using the Blank Database option

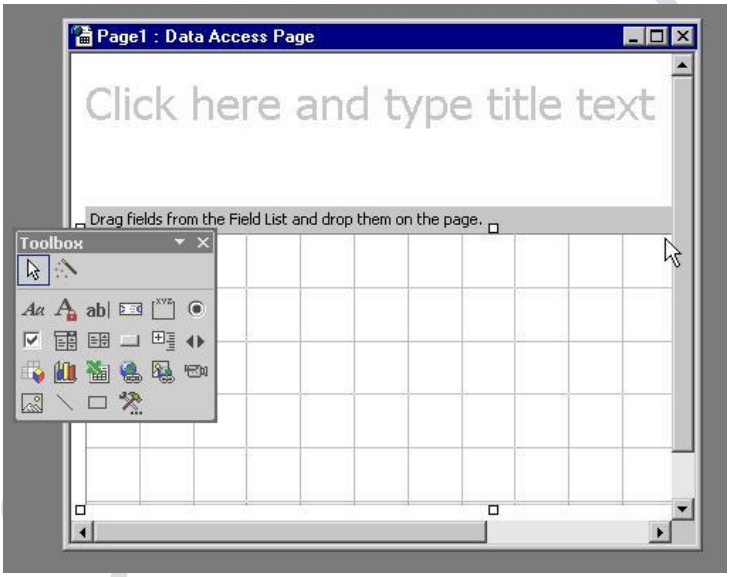
- The **Blank Database** option lets you create a new database file from scratch. You'll then be able to add new tables and other database objects, also from scratch:

- To use this option, click **Blank Database**. The **File New Database** dialog box opens.
- In the **File Name** box, type a name for the new database, and from the **Save As Type** list, choose a file format (usually **.mdb** for a Microsoft Access database).
- Use the **Save In** list to browse to the location in which you want to place the new database file. You can double-click folders in the large pane to open them.
- Click **Create** when you have entered all the required information.

Data Access Page option

- The **Blank Data Access Page** option lets you create a new data access page from scratch. A data access page is a Web page that is connected directly to the data in your database. You can use a data access page to view, edit, update, delete, filter, group, and sort information in the database:

- A data access page can contain components such as a spreadsheet, a PivotTable list, or a chart.
- Data access pages only work in **Microsoft Internet Explorer 5** or later.
- To use this option, click **Blank Data Access Page**. A view of the blank page appears:



- Create the page as desired. Use the toolbox that appears along with the view to add text and other objects to the page.

Note: If the toolbox doesn't appear, you can view it by choosing **View > Toolbox**.

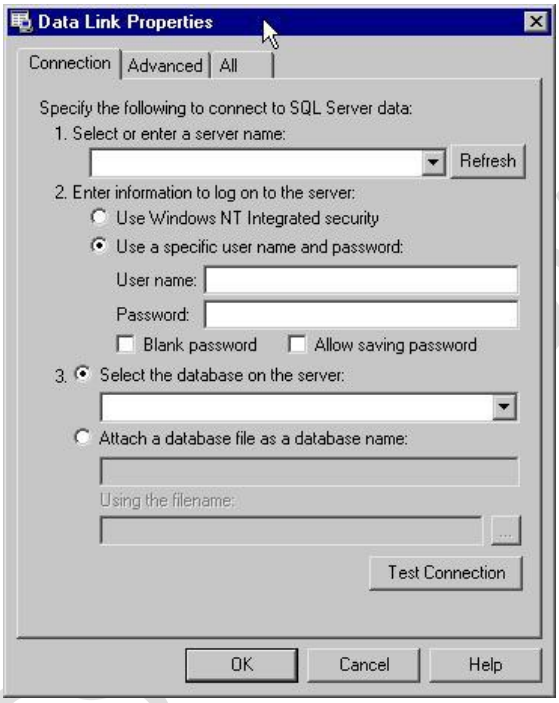
To make a page available to other users over the World Wide Web:

- You need to save the page to a Web folder or Web server.
- You also need to make the database (the one connected to the data access page) available to the users.

Using the Project (Existing Data) option

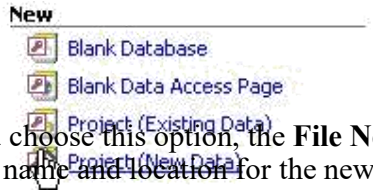
- The **Project (Existing Data)** option lets you create a new project file based on an existing database file. A project file is a file that connects to the database; it is the basis for client/server applications. For example, a project file might contain forms, reports, or modules; it would not contain any data or data-definition-based objects (like tables):

- When you choose this option, the **File New Database** dialog box opens. This dialog box lets you specify a name and location for the new project file.
- When you click **Create** in the **File New Database** dialog box, the **Data Link Properties** dialog box opens. This dialog box lets you provide the necessary information for Access to connect the new project file with an existing database:



Using the Project (New Data) option

- The **Project (New Data)** option lets you create a new project file that is not yet based on any database file:



- When you choose this option, the **File New Database** dialog box opens. This dialog box lets you specify a name and location for the new project file.
- When you click **Create** in the **File New Database** dialog box, the **Microsoft SQL Server Database Wizard** opens. This wizard lets you provide Access with information about the database you will attach to the project. You can provide a name for this database as well as connection information:

Using the New From Existing File Area of the New File Task Pane

Using the Choose File option

- The **Choose File** option lets you create a new database file based on the data in an existing file:

- When you select **Choose File**, the **New From Existing File** dialog box appears. This dialog box lets you specify the name and location of the file you want to base the new database file on.
- You can base a new database file on any of the following file types: another **Access database** file (*.**mdb**), an **Access project** file (*.**adp**), or an **HTML** file (*.**htm** or *.**html**).

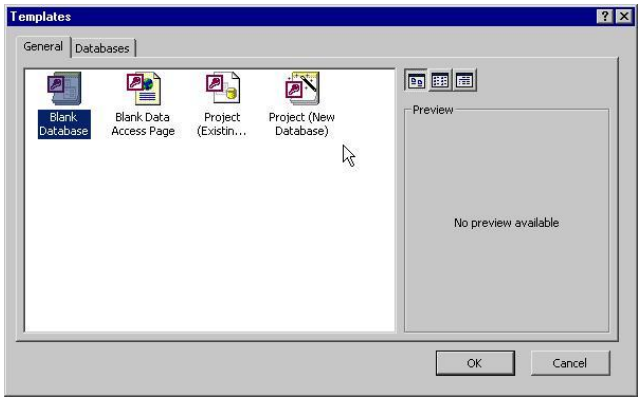
Using the New From Template Area of the New File Task Pane

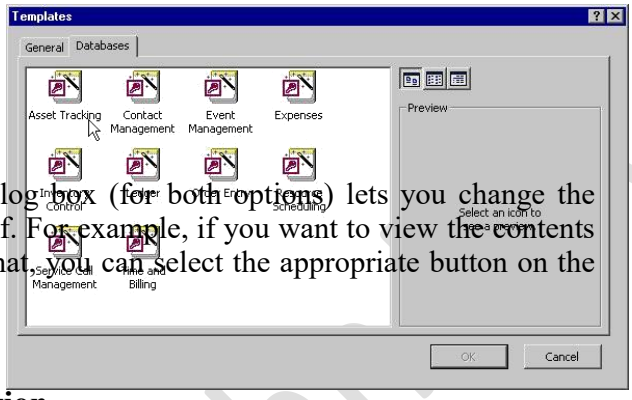
Using the General Templates option

- The **General Templates** option lets you create a new file (either a database file or a project file) based on one of Access's predefined templates:

- When you choose **General Templates**, the **Templates** dialog box opens with two tabs: the **General** tab and the **Databases** tab.
- Use the **General** tab to create a new blank database, blank data access page, project (based on an existing file), or project (not based on an existing file). These are the same options as the ones listed in the **New** area of the **New File** task pane:

· Use the **Databases** tab to view a list of predefined Access databases. To model your new database file on one of these, select it and click **OK**:





- The right half of the **Templates** dialog box (for both options) lets you change the manner in which the files are listed in the left half. For example, if you want to view the contents of the left half in list format instead of icon format, you can select the appropriate button on the right half.

Using the Templates On Microsoft.com option

- The **Templates on Microsoft.com** option lets you create a new file based on a Microsoft predefined template located on Microsoft web site:

- This option is a link that launches your Internet browser and brings you to a Microsoft Office web page. Follow the instructions on this page to view the available templates.

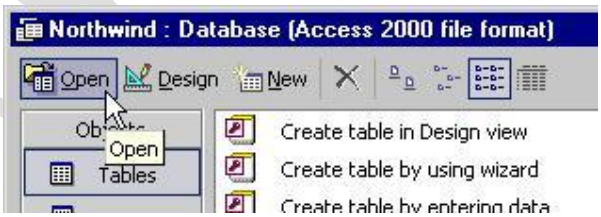
The Database Menu Bar

Using the Database Menu Bar

- The Database Window Menu bar allows you to open an object, access Design view for the selected object, create a new object, delete an object, and arrange the objects in the Database window.

Using the Open button

- The **Open** button of the Database window menu bar allows you to view an object in your database:



- The **Open** button works in conjunction with the object you have currently selected in the window. For example, if you have a particular table selected and click the **Open** button, a window containing that table opens.

Using the Design button

- The **Design** button allows you to view an object in Design view. Design view is a mode of Access that lets you create or modify the structure of tables and queries, create forms or pages through which you can display data, and format reports:

- Like the **Open** button, the **Design** button works in conjunction with the object you have currently selected in the window.
- The options within the Design view depend on the object you are modifying. For example, Design view for tables differs from that for forms.

Using the New button

- The **New** button allows you to create a new object using a wizard:

- The **New** button works in conjunction with the type of object button selected in the toolbar along the left side of the Database window. For example, to create a new table, select the **Tables** button and click **New**.

Using the Delete button

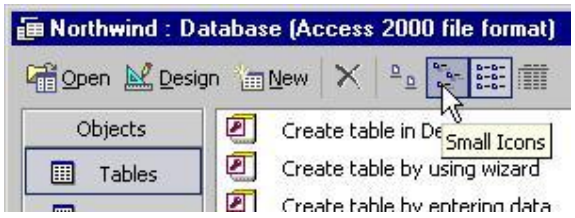
- The **Delete** button allows you to delete an object:

- To use this feature, select the object you want to delete and click this button. As a safeguard, Access will always prompt you to continue before it actually deletes the object.

Using the Icon buttons

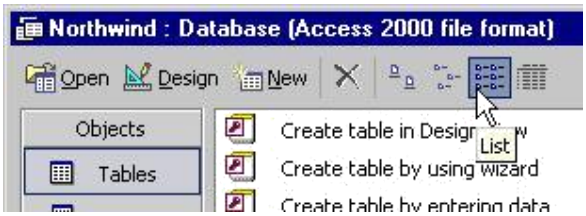
- The **Large Icons** and **Small Icons** buttons allow you to change the size of the graphics in the Database window.
- Click **Large Icons** to make the graphics bigger:

- Click **Small Icons** to make the graphics smaller:



Using the List button

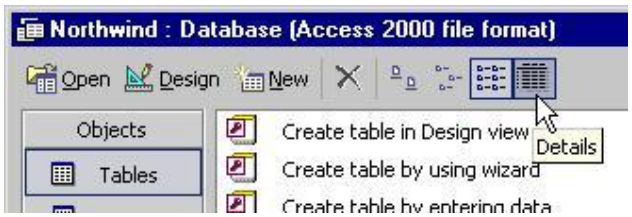
- The **List** button allows you to view the database objects in the Database window in a list format:



- In this list, only the names of the objects are listed.

Using the Details button

- The **Details** button is similar to the **List** button (it also lets you view the database objects in list format), but it provides additional information to just the name, such as a description of the object, the date it was last modified, and its type:

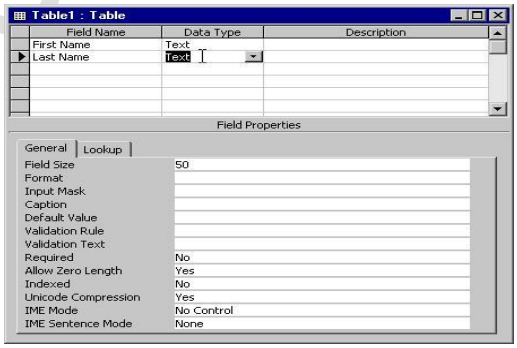


- To add a description to an object (so it appears under the **Description** column when you click the **Details** button), right-click an object in the Database window and choose **Properties**. In the **Properties** dialog box, you can type the description in the box provided. After you close this dialog box, the description will appear when you have the list details showing:

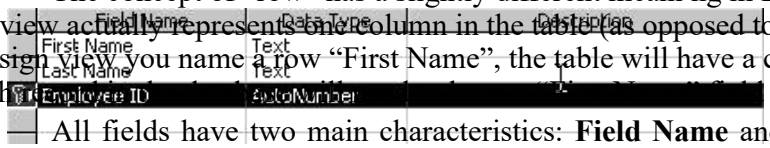
Create a Table

About creating a table in Design view

- You can create a table in Design view or with a wizard.
- **Design view** is a window that lets you configure the table columns (also called fields). Think of this view as a blueprint for the table: it doesn't show you actual data in each row, but it does define how the rows need to look:



- The concept of "row" has a slightly different meani ng in Design view: a "row" in Design view actually represents one column in the table (as opposed to one record). For example, if in Design view you name a row "First Name", the table will have a column called "First Name" and each entry.

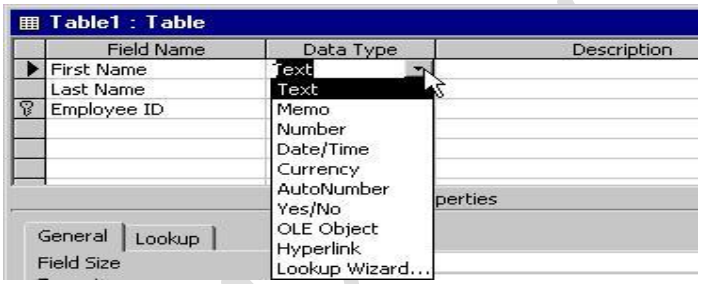


- All fields have two main characteristics: **Field Name** and **Data Type**. **Field Name** refers to the name of the column. **Data Type** refers to the type of information a field can contain (e.g., text instead of numbers).
- For each field, you can also define other properties, such as how many characters they can hold, or how they should be presented to users (e.g., as a drop-down list instead of a text box).
- It's usually a good idea to define one field in the table as a **primary key**. A primary key is a field whose value uniquely identifies each record in the table. Access provides an **AutoNumber** feature that can assign a unique number to each record for this purpose. For example, if you are creating an Employees table, you can create an Employee ID field, define it as the primary key, and use the AutoNumber feature to give each employee record a unique ID. Primary keys are used as the basis for relationships to foreign keys of other tables. A primary key cannot allow Null values:

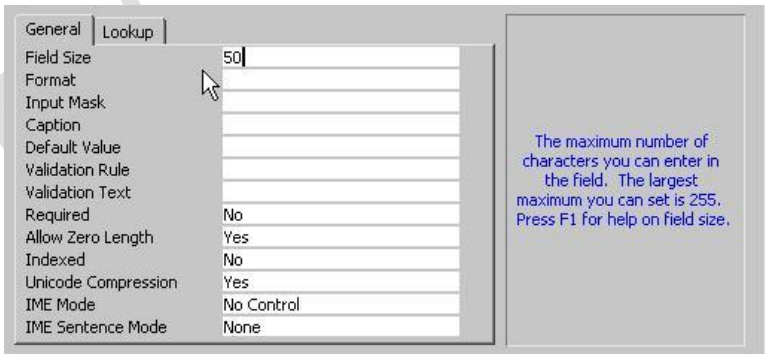
You can use the **Tab** key to move from field to field.

Creating a Table in Design View

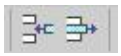
- Click the **Tables** button along the left side of the Database window.
- In the right pane, double-click **Create table in Design view**. A blank design view window opens.
- Begin creating fields: input a name and type for each. When you place the cursor in the **Data Type** field, a drop-down arrow lets you choose a type:



- For each field, use the tabs at the bottom of the window to further define field properties. If you need more information about these properties, place your cursor in each; the lower right corner of the window tells you how that property works:



- Use the **Insert Rows** or **Delete Rows** buttons on the toolbar at the top of the Access window to add new rows above your current row or delete your current row:



- Use the **Primary Key** button on the toolbar at the top of the Access window to make the current row the primary key. If you want to remove primary key status from that row, select it and click the **Primary Key** button again.

Note: You can also right-click a row and choose **Primary Key** from the context menu:

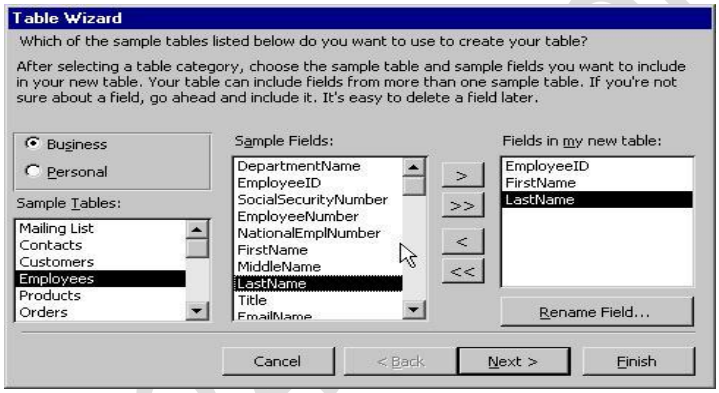
- When you are finished creating fields and setting up the table as desired, choose **File > Save** and type a name for your table.

About using the Table Wizard to create a table

- The other method for creating a table is by using a wizard. This wizard provides samples of common fields in different types of personal or business database tables. Note that there is no actual data attached to these fields; they are merely models that you can choose to incorporate. You can always rename the fields if you want a different name.

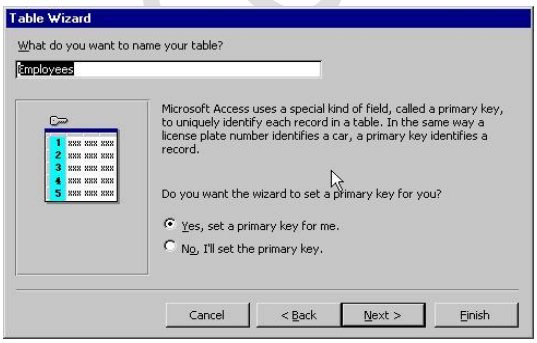
Using the Table Wizard to create a table

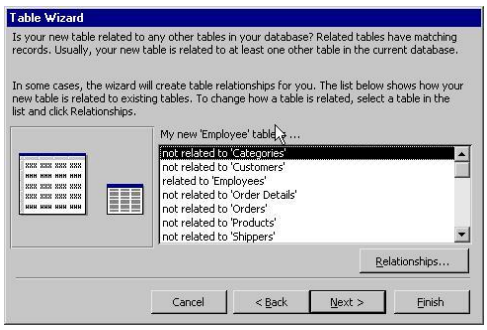
- Click the **Tables** button along the left side of the Database window.
- In the right pane, double-click **Create table by using wizard**. The wizard opens.
- On the first page, use the **Business** or **Personal** options to filter the list of sample tables and fields to the desired type. Clicking different tables in the **Sample Tables** list makes the contents of the **Sample Fields** list change accordingly. When you see a field name that matches a field you want in your new table, select it and click the **>** arrow button:



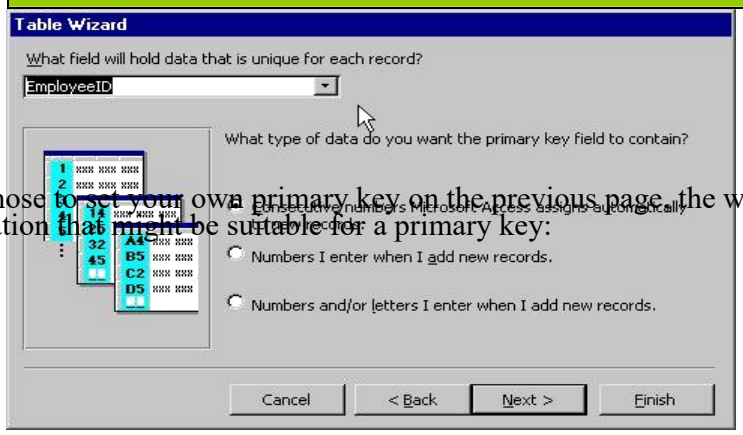
Fields you have added stay in the list on the right even if you change the options in the left half of the dialog box. If you want to remove a field from the list select it from the list on the right, and click the **<** arrow button. If you would like to rename a field, select it and click **Rename Field**.

- Click **Next**. Type a name for your table and choose whether you want the wizard to set a primary key. If you choose **No**, you'll need to do it in Design view after the table is created. If you choose **Yes**, you can always change the primary key after the table is created:





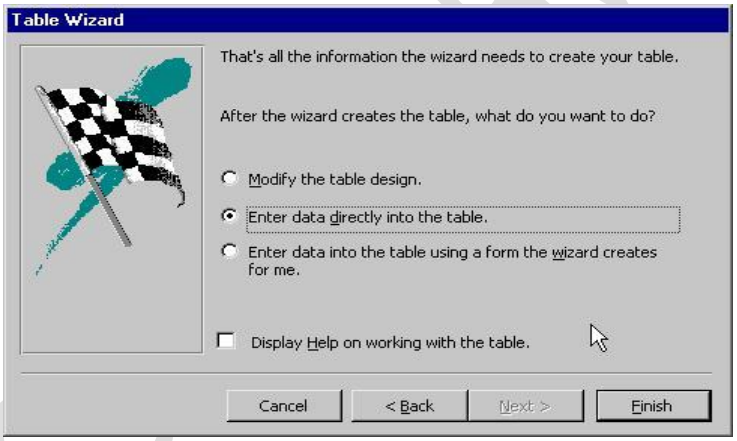
- Click **Next**. If you chose to let the wizard set a primary key for you on the previous page, you can now create relationships between this table and other existing tables if there is common information between the tables. The wizard lists all other existing tables; if you select one and click **Relationships**, the wizard asks you the type of relationship (and gives an explanation of each):



f you chose to set your own primary key on the previous page, the wizard asks you for information that might be suitable for a primary key:

- Click **Next**. If you chose to manually set your primary key, you are now prompted to specify relationships (described above). Otherwise, the final wizard screen appears.
- Choose an option for when the wizard finishes: you can have the wizard open Design view for you to modify the table design, start entering data directly into the table

OR have the wizard open a form for data entry:



- Click **Finish**.

Creating and Using Data Entry Forms can be used for data entry, displaying data in a user-friendly format, and as a navigational tool.

Using Forms

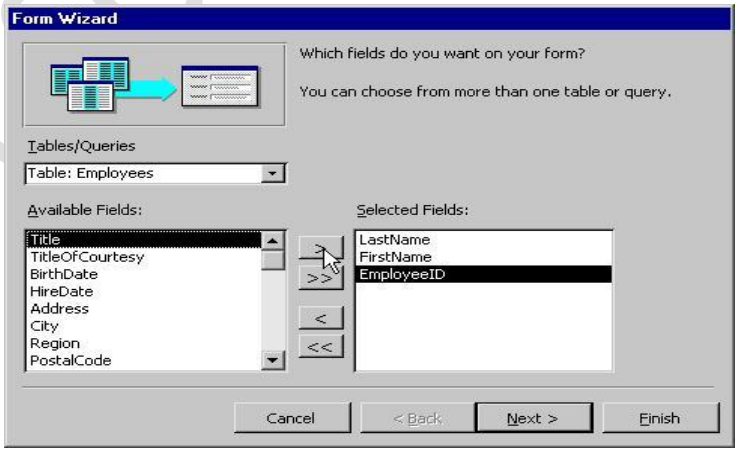
- A form is a type of interface you can create between you (or any user) and the database. Forms can be as simple or complicated as you want.
- The simplest type of form is a **data-entry** form. This type of form simply displays data from the database. For example, if you had a query that returned a set of values, you could display these values in a form.
- Another type of form is a **switchboard** form. This type of form can have links to (and open) other forms, and it can open reports.

- The third type of form is a **custom dialog box**. This type of form can accept user input and execute an action based on the input.
- A **bound** form is a form that is bound to one or more tables in the database. The data is stored in the underlying table (called the record source) and referenced into the form; the form's organizational information (like title, date, and page number) is stored in the form's design.
- You can change the form's appearance by adding graphics, lines, text, or other types of appearance-oriented features. For example, if you need to include a company logo on form to apply a uniform and standard look to your database, you can easily do so.
- The easiest way to create a form is to use the wizard that Access provides. This wizard lets you choose the fields or records from a table or a query that you have saved. Once the wizard places the appropriate data in the form, you can use Design view to customize the form's appearance. If you want, you can also create a form from scratch in Design view.
  - When the form is finished and you open it, it appears in its own window. You can click the arrow buttons in the bottom left of this window to move through various records.
- Forms can have different sections to them. Usually, the data (presented in text boxes or other user interface objects) appear in the Details section. You can also add a Header or Footer section to the form, perhaps to contain page numbers.

Creating forms with the wizard:

In the Database window, click the **Forms** button on the toolbar along the left side of the window:

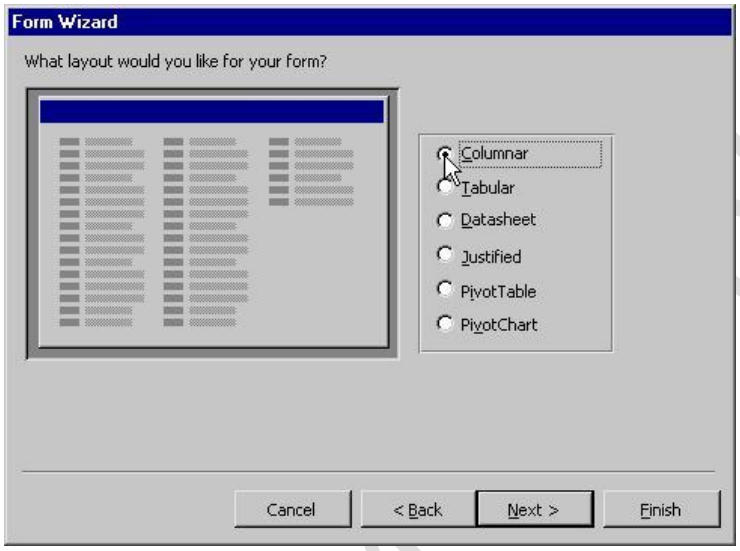
- Double-click **Create form by using wizard**. The **Form Wizard** opens:



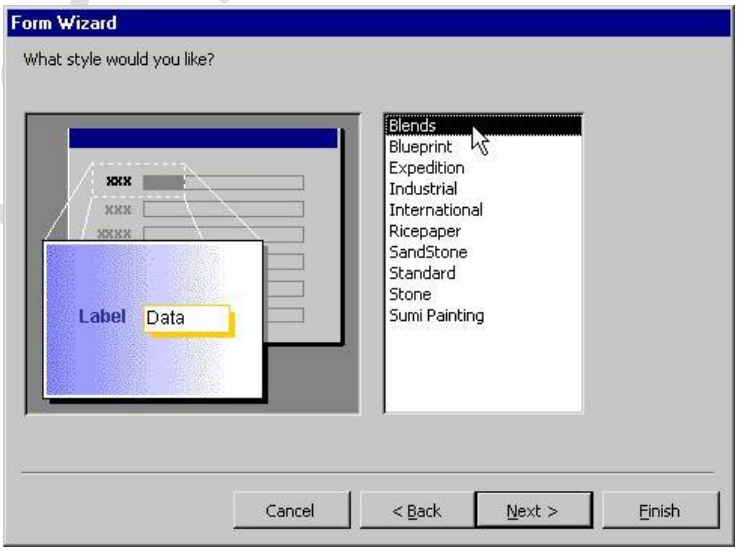
- On the first page of the form, specify where you want the data for the form to come from. The **Tables/Queries** drop-down list shows all table and queries you have created; when you select one of these, the fields from that table or query appear in the **Available Fields** list.
- Double-click fields in the **Available Fields** list to move them to the **Selected Fields** list.

Note: You can also select a field and use the arrow buttons to move them back and forth between the lists.

- When the **Selected Fields** list contains all the fields you want to be part of the form, click **Next**.
- On the next page of the wizard, choose one of the layout options. When you select the option, Access shows you a preview of that presentation option in the left side of the dialog box. When you are done, click **Next**:



- On the next page, choose one of the built-in styles to give your form attractive and professional look. Choose one of the styles listed. When you select an option, Access shows you a preview of that style in the left side of the dialog box. When you are done, click **Next**:



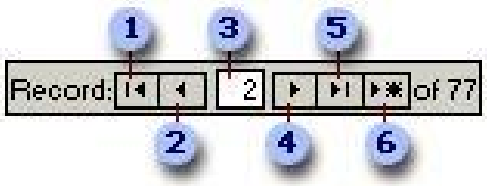
- On the last page of the wizard, specify a name for the form. After you are finished with the wizard, this name will appear in the Database window when you select the **Forms** button:

- Tell the wizard what to do when it closes. You can display the form immediately, or you can view the form's design in Design view.
- Click **Finish**.

Move between records or fields

Move between records by using navigation buttons in a form

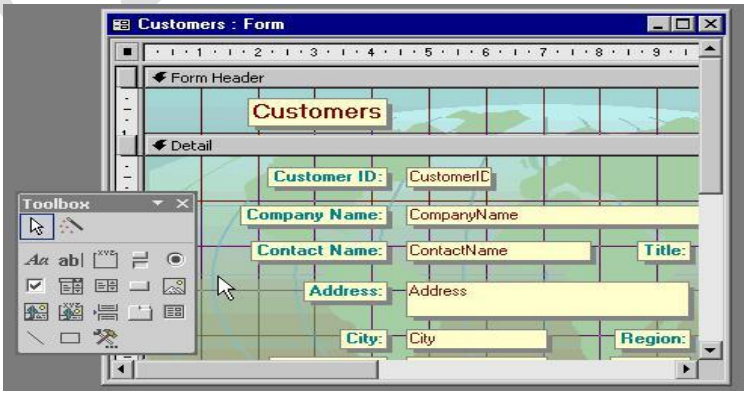
The navigation buttons are located at the bottom of the window in Datasheet or Form view. You can use these buttons to move quickly between records.



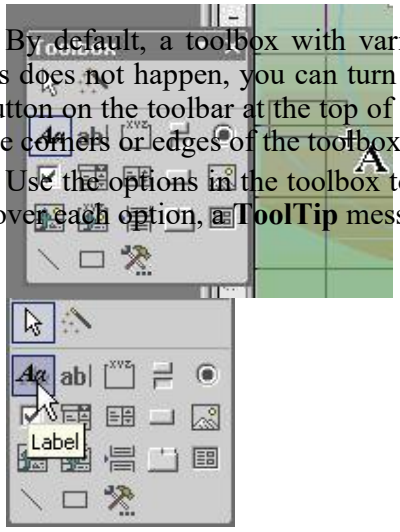
1 First record
2 Previous record
3 Record number
4 Next record
5 Last record
6 New record

Modifying a form in Design view

- If the form you want to modify is not already in Design view, do the following: in the Database window, with the **Forms** button selected, select the form and click the **Design** button on the Database window's toolbar. (You can also right-click the form and choose **Design View** from the context menu.) The form opens in Design view:



- By default, a toolbox with various formatting options should appear alongside the form. If this does not happen, you can turn the visibility of the toolbox on or off by clicking the **Toolbox** button on the toolbar at the top of the screen. You can resize the toolbox if you want by dragging the corners or edges of the toolbox to the new dimensions.
- Use the options in the toolbox to add elements of design to the form. When you hold the mouse over each option, a **ToolTip** message tells you what that option does:

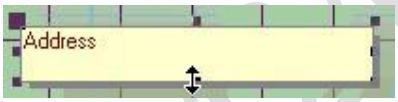


- After you select a toolbox item, move the mouse indicator over top of the form. You'll notice that the mouse indicator changes to show the tool that you have selected. Drag a box on the form to create the object.
- For example, if you want to add a label to the form, click the **Label** button in the toolbox. Move the mouse indicator over top of the form to the location where you want to type the label. Drag a box in the approximate shape of the label. As soon as you release the mouse button, a box will appear, with the cursor blinking in this

box: this means you can start typing text for the label in the box. As soon as you are done, press the **Enter** key, and the label is complete:

To resize an object in Design view:

- If you need to resize an object that you have added to the form, click the object. Access shows little black squares along the sides of the object. These black boxes are **resize handles**: click and drag one to resize the object. When you are done, click away from the object; the black boxes disappear, and the object appears with the new size you have configured:

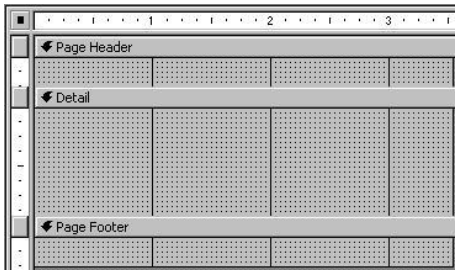


Creating Page Headers and Footers

- **Page headers** and **page footers** appear at the top and bottom of each page and can display information such as a column heading.

To create page headers and footers:

- Right-click in the **Detail** area of the form.
- From the pop-up menu, select **Page Header/Footer**. This opens the **Page Header** and **Page Footer** panes:

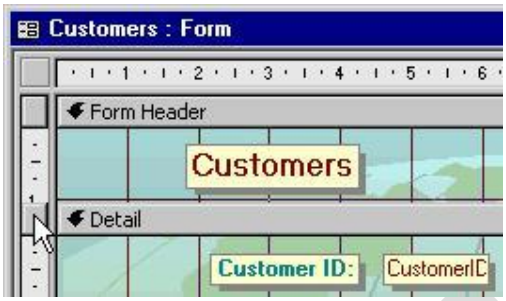


- Place the desired control on the **page header** or **page footer**.
- From the pop-up menu, select **Form Header/Footer**. This opens the **Form Header** and **Form Footer** panes.
- Place the desired control on the **form header** or **form footer**.

Note: Page headers and footers and **Form headers and footers** can both appear on the same form.

To add headers and footers to a form in Design view:

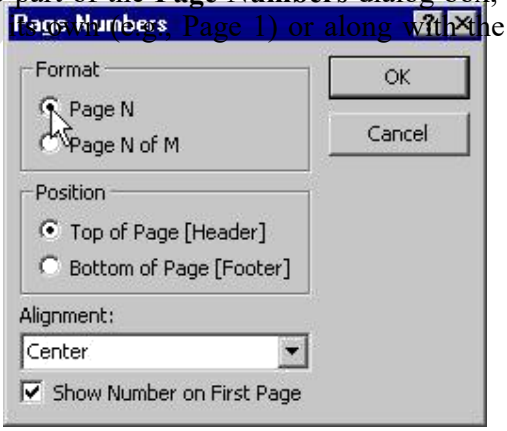
- You can add **headers** and **footers** to a form in Design view, similar to headers and footers in a Microsoft Word document. For example, you could place a page number in the header or footer, or insert a company logo.
- To create a header and/or footer, right-click anywhere in the form and choose **Form Header/Footer** from the context menu. You will notice that Access creates a **Form Header** section at the top of the form and a **Form Footer** section at the bottom of the form. Each of the section headings displays as a bar across the width of the form, and each of these bars has a small gray box at its left-most edge:



- To resize the sections, hold the mouse on the section bar's top or bottom edge. The mouse indicator changes to show arrows on the top and bottom. Drag up or down to make the section bar appear higher or lower. For precise positioning, use the ruler that's provided along the left edge of the form window.
- You can add objects from the toolbox to a header section or a footer section in the same way as you would for the Detail section.
- You can edit properties for any of the sections by right-clicking the section bar and choosing **Properties** from the context menu.

To insert page numbers in a header or footer:

- From the menu bar, choose **Insert > Page Numbers**.
- In the top part of the **Page Numbers** dialog box, choose whether you want the page number to appear on its own (e.g., Page 1) or along with the total number of pages in the form (e.g., Page 1 of 8):

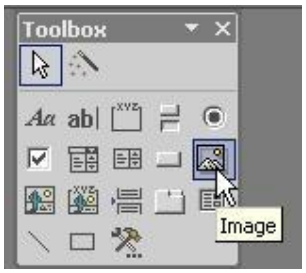


- In the middle part of the dialog box, choose whether you want the page number to appear in the header section or the footer section.
- In the bottom part of the dialog box, specify how you want the numbers to be aligned on the form and click **OK**.

Inserting an image that does not change from record to record

To insert an unbound image (which will be the same for all records):

- Open the form in Design view.
- Make sure that the toolbox is showing. If not, click the **Toolbox** button on the toolbar.
- Click the **Image** button in the toolbox:



- In the form itself, click where you want the image to appear. The **Insert Picture** dialog box then opens.
- Find and select the image that you want to insert. Then click **OK** to insert it.
- Right-click the image and choose **Properties** from the context menu.
- On the **Format** page of the **Properties** dialog box, click in the **Picture** cell and use the drop-down list to specify whether the picture should be linked or embedded.
- Use the other picture-related settings in the **Properties** dialog box (described in the previous procedure) to size and place the image appropriately. Then close the **Properties** dialog box.
- To move or resize the image manually, click the image. When you hold the mouse indicator over the image, the indicator turns into the image of a hand. You can then drag the image to a new location. If you select and drag one of the small black boxes along the edges of the image, you can resize it.

Replacing an image

To replace an unbound background image:

- Open the form in Design view.
- Right-click the form's title bar and choose **Properties** from the context menu.
- On the **Format** page, click in the **Picture** cell.
- Click the **Browse** button beside this cell.
- Browse to a different image and click **OK**.

Choosing an Appropriate Control

- Controls are objects that can be placed on forms, reports, or data access pages that display data, perform actions, or can be used for visual display. Controls shipped with ACCESS include:
- **Text Boxes:** Used to display data on a form. Text boxes can be **bound** to a data source or **unbound** to accept input or display a calculation.
- **Labels:** Used to display text on a form. Labels are useful for descriptive of instructional text. Labels are not associated with a data source.

- **List Boxes:** Used to display a pre-determined list of values. List boxes are useful when limiting selection choices or when enforcing data integrity.
- **Command Buttons:** Used to execute a pre-defined set of actions in a macro or event procedure written in visual basic.
- **Check Boxes:** Used to set **Yes/No** values from an underlying query or table.
- **Option Groups:** Used to select a limited list of choices. Sometimes interchangeable with **List Boxes** when selecting from only a few pre-defined options.
- **Toggle Buttons:** Used to set **Yes/No** values from an underlying query or table. Useful when combined with an **Option Group** to select values that are **Yes/No**.
- **Image:** Used when displaying a digital image on a form. An example would be the display of a company logo.
- **Tabbed Pages:** Used when presenting different sections of information as one logical set. Useful when separating levels of information related to a common topic such as employee, company, or personal information.
- **Rectangles and Lines:** Used when displaying limited geometric shapes. Useful when displaying breaks between fields or sections within a form.
- **Subform:** An advanced control that displays child data that is related to data on the parent form. **Subforms** are discussed later in this training manual.

Creating Queries

About queries

- A query is similar to the idea of filters discussed earlier: it is a way to search for data in your database. Unlike filters, though, which can only work within a single datasheet (or table) at a time, queries can let you search multiple tables at once.
- Access lets you create five main types of queries: **Select** queries, **Parameter** queries, **Crosstab** queries, **Action** queries, and **SQL** queries.
- A **Select** query is the most common type of query. You can use it to find and return data from one or several tables, displaying the found data in a datasheet. When creating a **Select** query, you can instruct Access how to group the records it retrieves and calculate sums, counts, averages, or other types of operations on the retrieved data.
- Think of a query as a set of instructions from you (the user) to the database: "Database, give me all records from the **Employees** table where the **Last Name** field of the record equals "Jones" and the **Phone Number** field contains the numbers '123'." It's possible that only one employe e named Jones will be returned, or maybe the query will find no matching data: either there are no "Jones" employees in the table, or the "Jones" empl oyees that were found do not have a '123' in their phone number.
- You can make the query as precise as you want, to return only very specific information, or you can make the query broad, to return multiple records.
- When creating the query, you can also tell Access how many fields of the records you want returned. For example, if you want to retrieve all employees with the last name "Jones", and you only want to see the **First Name** and **Last Name** fields of the records matching that criteria, you can specify only those two fields in the query. This lets you focus your results set on only meaningful data, so that the datasheet does not get too confusing.
- The data that Access finds in response to your query is called the **results**. Some other database applications call this returned data a **results set**.
- You can save queries to reuse them. This is a helpful feature if the data in the database continually changes, but you often search for the same type of information. For example, if you regularly need to search for sales in excess of $10,000 for any product, you can build and save a query to search the **Sales** fields when for values exceeding $10,000. You can then reuse this query each week or month; your search criteria doesn't change even though the data does.
- When working with queries, you should be aware of some presentation conventions. First, table columns (or fields) are sometimes presented as *tablename.fieldname*. For example, the **Last Name** field in the **Employees** table might be shown textually as *Employees.Last Name*. This can help to avoid confusion if two tables have the same field name. Second, an asterisk (*) can represent all fields in a table. For example, *Employees.** would mean all fields in the **Employees** table.

About creating a new query in Design view

- Access lets you create queries in one of two ways: in Design view or by using a wizard.
- For the Design view method, Access operates in a similar way to the Design view of creating a table: you build the structure of the query by working with the list of fields that Access provides.
- When you create a query in Design view, you can have the query span multiple tables. To do so, however, you need to create relationships between tables. The concept of relationships is a more advanced feature of Access, and it is not described in detail in this Foundation level. This level concentrates on teaching you to run a query on one table at a time.
- Design view lets you drag-and-drop items from a table into the grid in the lower pane.

Requirements for a query in Design view

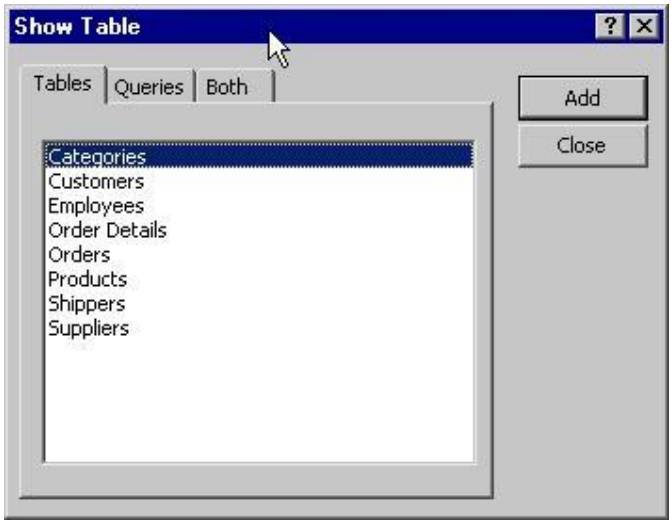
- A simple query in Design view consists of the following elements:

- **First**, you need to specify the field(s) you want involved in the query.
- **Second**, you need to specify the table each field belongs to.
- **Third**, you can optionally tell Access to sort the results based on a field. For example, if you have selected the Last Name and First Name fields and you want the results sorted alphabetically by Last Name, you can do an Ascending sort on the Last Name field.
- **Fourth**, you need to specify which fields you want to appear in the results. This is not the same thing as the second requirement. For example, you can build the query so that it looks for information in two different fields, but only shows one field in the results.
- **Fifth**, you need to give the actual criteria for the query. In other words, you need to tell Access what information you want it to search for.

Creating a new query in Design view

- In the Database window, click the **Queries** button on the toolbar along the left side of the window:

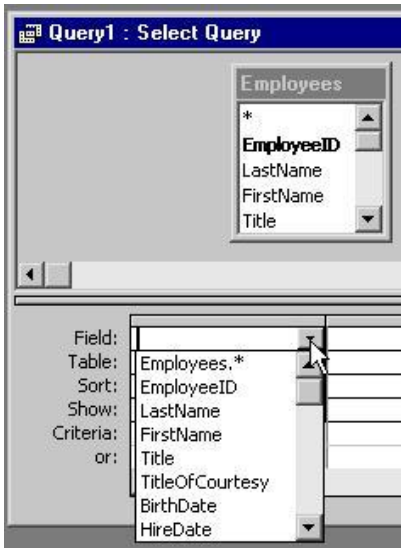
- Double-click **Create query in Design view**. A query window opens. If you have multiple tables in your database, the **Show Table** dialog box also opens:



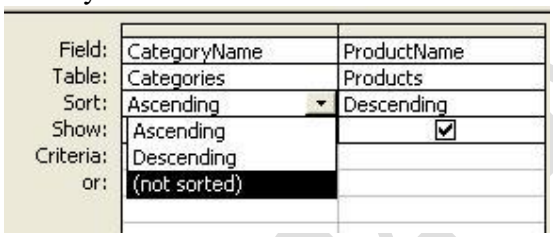
- If the **Show Table** dialog box opens, on the **Tables** tab, select the table(s) that you want to query and click **Add**. A box listing that table's columns (or attributes) then appears in the pane at the top of the query window.
- Close the **Show Table** dialog box when you are finished with it.
- Choose the fields you want involved in the query. You can do this in one of two ways:
- **First**, you can double-click a field listed in the table box in the top pane; the field then appears in the first empty column in the bottom pane:

- **Second**, you can place the cursor in a box in the **Field** row; a drop-down arrow appears, letting you select a field. If you use the second method, be aware of the following: if you have multiple tables open in the top pane, and if no information appears in the **Tables** row immediately below where you click, the drop-down list will contain all fields from all visible tables, in the form of *tablename.fieldname*.

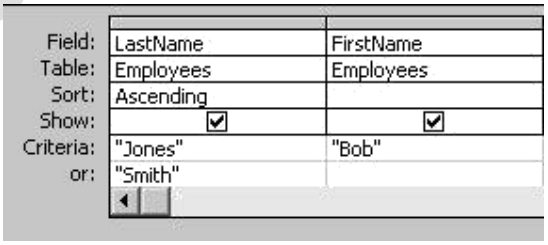
As well, the first option will list *tablename.**, which means all fields from that table:



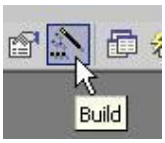
- In the **Field** row of the bottom pane, continue to choose fields you want involved in the query. You may need to specify the table that field belongs to in the **Table** row, right below the **Field** row.
- If you want to sort the results, in the **Sort** row of the bottom pane, click an entry under the field you want to sort by. For example, to sort by Last Name, click within the Last Name column. A drop-down list appears; choose **Ascending**, **Descending**, or **Not Sorted**. You can only sort by one field.



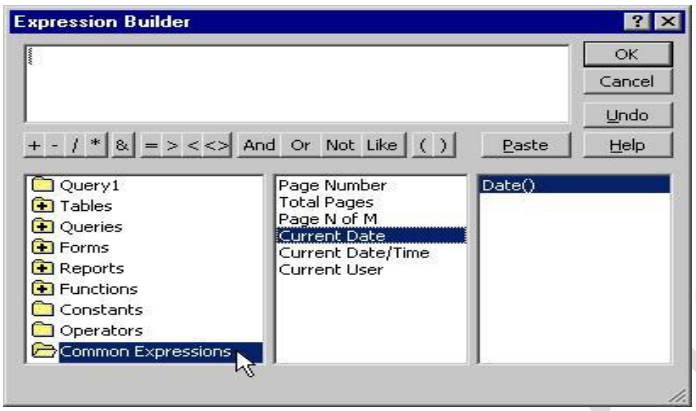
- In the fourth row (**Show**), click in the check box (so that a check mark appears) to make that field appear in the results. Remember that if you choose not to show a field, the query will still use the search criteria for that field but just won't show the field in the results. For example, if you build the query to search for First Name of "Bob" and Last Name of "Jones", but only sh ow the Last Name field, Access will still look for records containing "Bob" and "Jones" but will only show the Last Name field of matching records.
- In the fifth row (**Criteria**), type the data that Access should be looking for (e.g., "Jones"). You can specify criteria for multiple fi elds if you want. You can also specify multiple criteria for a single field by typing a second string in the **Or** row:



- If you want the criteria to involve a mathematical operator (e.g., sales greater than $10,000), place the cursor in a **Criteria** field and click the **Build** button on the toolbar:

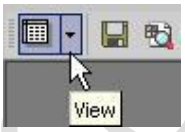


- The **Expression Builder** dialog box opens. In this dialog box, use the pane at the top to build the expression. You can add operators by clicking the buttons below this pane.
- You can have Access show you common expressions by clicking the folders in the bottom left corner of the dialog box and then double-clicking the entries that appear in the right panes to add them to the top pane:



- When you have built the query to the appropriate specifications, click the **Run** button on the toolbar. Access returns all matching records in a new window:

- To re-show the Design window, click the **View** button in the top left corner of the toolbar:



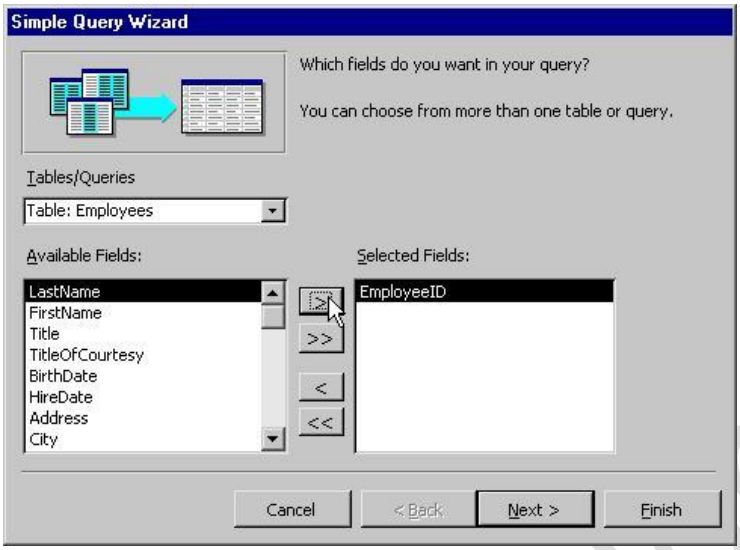
About creating a new query using a wizard

- The other method for creating a new query is by using a wizard. This wizard takes you through the various steps of building a simple query.
- The requirements for a building a query in this way are similar to those for building a query in Design view; you need to provide the same basic units of information.
- The screens of the wizard may vary, depending on the selections you make on an earlier screen. That's why the procedures below are broken up according to option selections.
- If you need to revise your design, you can always click the **Back** button in the wizard to return to an earlier step.

Creating a numerical query using the wizard

To create a query that includes a numerical field:

- In the Database window, click the **Queries** button on the toolbar along the left side of the window.
- Double-click **Create query by using wizard**. The Simple Query Wizard opens:



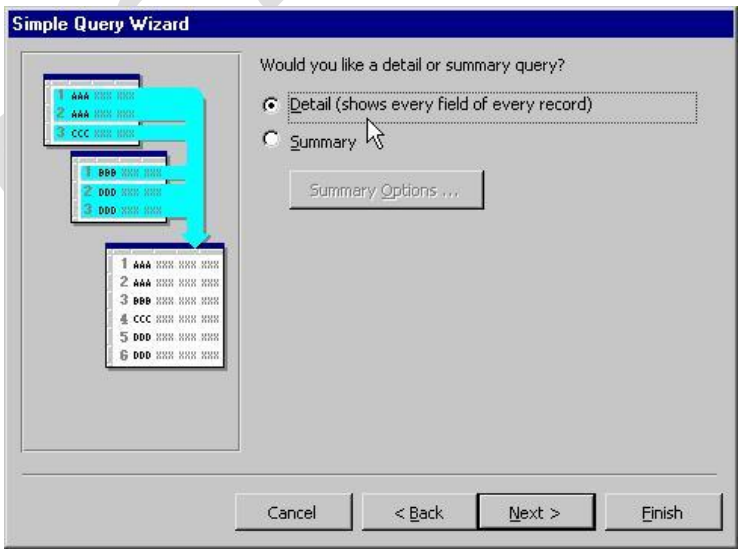
- On the first page, select the fields you want involved in the query. From the **Tables/Queries** drop-down list, choose a table. The available fields from that table then appear in the **Available Fields** list.Double-click a field to add it to the **Selected Fields** list

OR select one of the **Available Fields** and click the **arrow >** button.

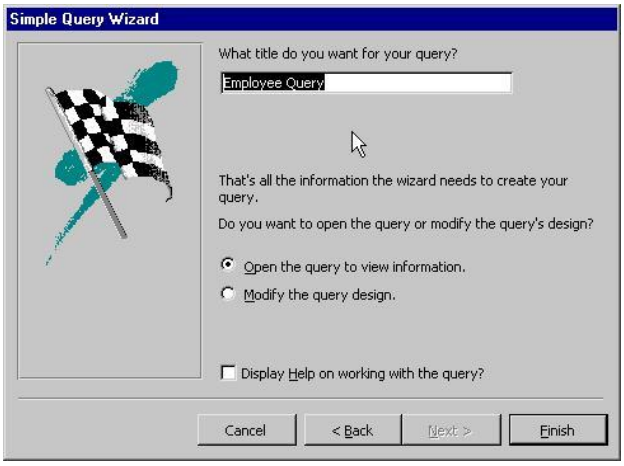
- You can remove a field from the **Selected Fields** list in the same way. Double-click one of the fields from the **Selected Fields** list

OR select one of the fields and click the ' **Back**' **arrow <** button.

- Make sure you include a numerical field.
- When you are finished, click **Next**. On the next page, tell the wizard how you want the information returned:



- Choose the **Detail** option if you want every field (of records that match the search criteria) returned. Choose the **Summary** option if you want Access to summarize the numerical data. You can specify the summary action by clicking **Summary Options** and providing the desired information in the resulting dialog box. If you want Access to show you a count of all the records it finds, select that check box and click **OK** to close the dialog box.
- Click **Next** to type a name for the query:

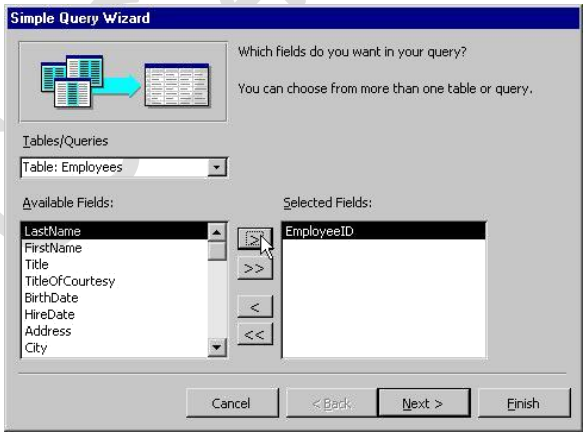


- To run the query and have the results returned to you, select **Open the query to view information** and click **Finish**.
- To hold off on running the query and to view the structure of the query in Design view before running it, select **Modify the query design** and click **Finish**. A query window containing the parameters you have specified opens; you can modify the query as desired. Click the **Run** button on the toolbar to run the query and see the results:

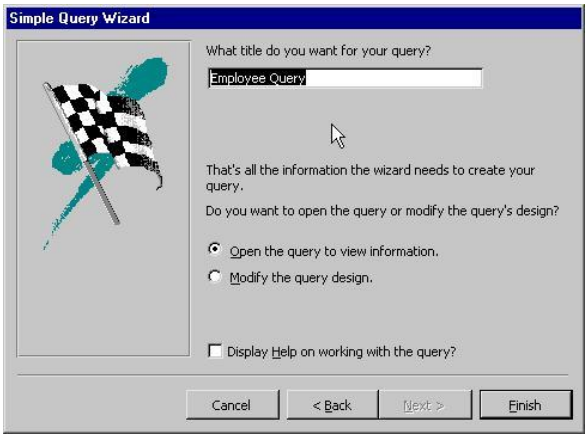
Creating a non-numerical query using the wizard

To create a query that does not include a numerical field:

- In the Database window, click the **Queries** button on the toolbar along the left side of the window.
- Double-click **Create query by using wizard**. The **Simple Query Wizard** opens:



- On the first page, select the fields you want involved in the query. From the **Tables/Queries** drop-down list, choose a table. The available fields from that table then appear in the **Available Fields** list.
- Double-click a field to add it to the **Selected Fields** list (or select one and click the arrow > button).
- You can remove a field from the **Selected Fields** list in the same way: double-click it or select it and click the 'Back' arrow < button.
- Do not include a numerical field. When you are finished, click **Next**.
- On the next page, type a name for the query:

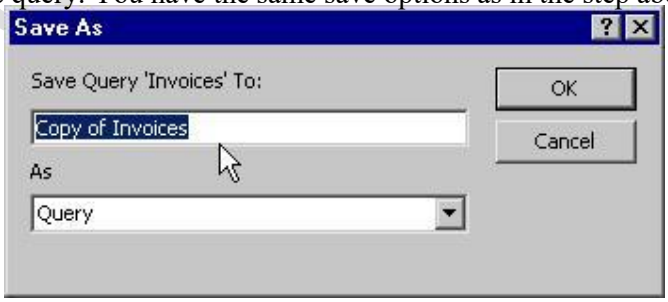


- To run the query and have the results returned to you, select **Open the query to view information** and click **Finish**.

- To hold off on running the query and to view the structure of the query in Design view before running it, select **Modify the query design** and click **Finish**.
- A query window containing the parameters you have specified opens. Here you can modify the query as desired.
- Click the **Run** button on the toolbar to run the query and see the results:

Saving queries

- You can save a query in multiple ways.
- If you are working in the Database window, you can click the **Queries** button on the left. Then right-click an existing query and choose **Save As** from the context menu. Type a new name to save the query with a different name. You can choose to save it as a query, form, report, or data access page.
- If you have the Results window open, from the menu bar, choose **File > Save As**. Type a name for the query. You have the same save options as in the step above:



- If you have the Query Design view open, from the menu bar, choose **File > Save As**. Type a name for the query. You have the same save options as in the step above.

Opening an existing query's result window

- Access shows all existing queries in the Queries view of the Database window.
- To open a query, click the **Queries** button on the toolbar on the left side of the Database window:

Double-click a query.

- The results of that query appear in a window

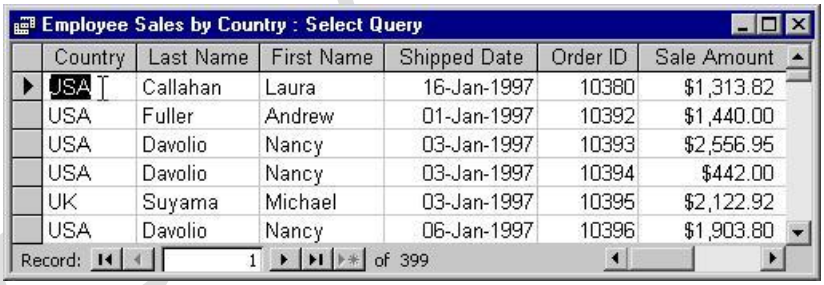
Open an existing query in Design view

- Click the **Queries** button on the toolbar on the left side of the **Database** window.
- Select a query.
- Click the **Design** button at the top of the **Database** window.

Note: You can also right-click a query and choose **Design View** from the context menu:

Refining the results of a query

- After you run a query, its results are displayed in a separate window. Take a moment to scan through the results; you may find that the information was not what you expected, or the wrong information appears. Often, an incorrect query is the result of errors in the criteria that you give Access when constructing the query:



- You can revise a query's construction in Design view. To refine results, take a look at all of the parameters you have specified in Design view. Ask yourself whether the fields, tables, and search criteria you have specified are related to the desired information.
- After you make some changes, you can re-run the query simply by clicking the **Run** button on the toolbar. If that result set is not sufficient, you can continue to modify the query in Design view and re-run the query until you get the results you want:

Creating and Using Reports

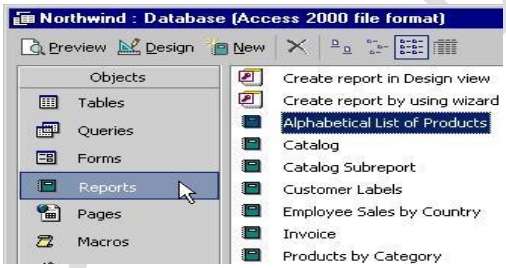
A report is a way to present information in a format that's suitable for printing. You can rearrange the information and configure the presentation so that it best suits your purposes.

- Reports are an effective tool for analyzing information. You can create a report that summarize data from the database, and then print the report to analyze numbers and trends.
- Reports are generally bound to one or several tables (or queries) in the database. The data that appears in the report is linked in from the fields in the underlying table or query. Other information on the report, like labels or titles or page numbers, are saved with the report itself (not with any underlying data source.)
- A report does not need to contain fields from each of the tables or queries it's based on.
- You can create links between a report and its underlying data source (e.g., tables or queries) by using graphical objects called **controls**. An example of a control is a text box: it displays data from a field in the underlying table.
- You can customize the report's design to make it more attractive.
- There are three ways to create a report.

○ **First**, you can create a report based on one table or query using a feature called AutoReport. AutoReport then creates a report that displays all the fields and records in the underlying table or query.
○ **Second**, you can create a report using a wizard. This wizard guides you through the process of selecting tables (or queries) and fields to serve as the basis for the report.
○ **Third**, you can create a report from scratch in Design view (similar to creating tables, queries, and forms in this way.)

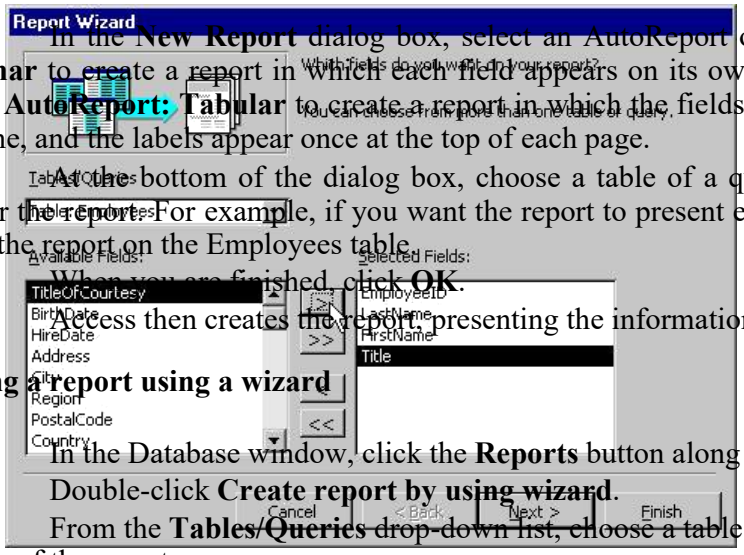
Creating a report using AutoReport

- In the Database window, click the **Reports** button along the left side of the window:



- On the toolbar at the top of the Database window, click the **New** button. The **New Report** dialog box opens:

- In the **New Report** dialog box, select an AutoReport option. Choose **AutoReport: Columnar** to create a report in which each field appears on its own line with a label to its left. Choose **AutoReport: Tabular** to create a report in which the fields of each record appear on the same line, and the labels appear once at the top of each page.
- At the bottom of the dialog box, choose a table of a query that you want to be the basis for the report. For example, if you want the report to present employee information, choose to base the report on the Employees table.
- When you are finished, click **OK**.
- Access then creates the report, presenting the information in the format you specified.

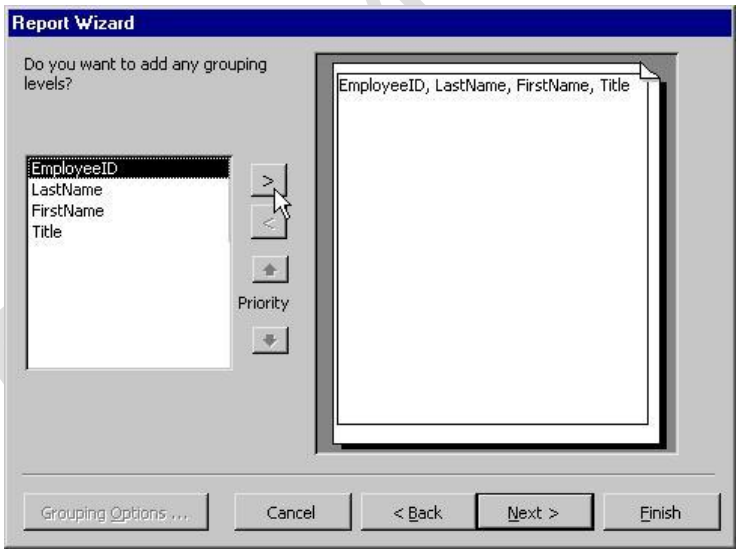


Creating a report using a wizard

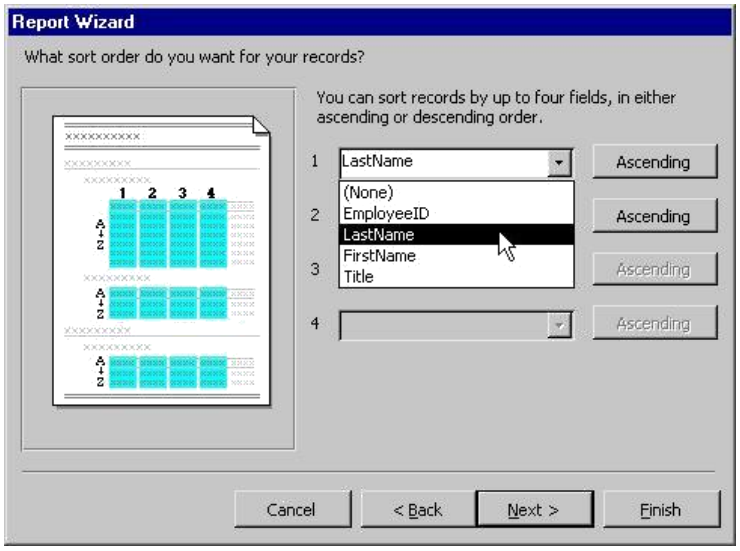
- In the Database window, click the **Reports** button along the left side of the window.
- Double-click **Create report by using wizard**.
- From the **Tables/Queries** drop-down list, choose a table or query that you want to be the basis of the report.
- The **Available Fields** list updates to show you the fields of the table or query you select in the **Tables/Queries** drop-down list. Double-click these fields to add them to the **Selected Fields** list.

Note: You can also select fields and use the arrow buttons to move fields to or from the **Selected Fields** list:
Continue to add fields to the **Selected Fields** list until you have all the fields you want for the report. Then click **Next**.

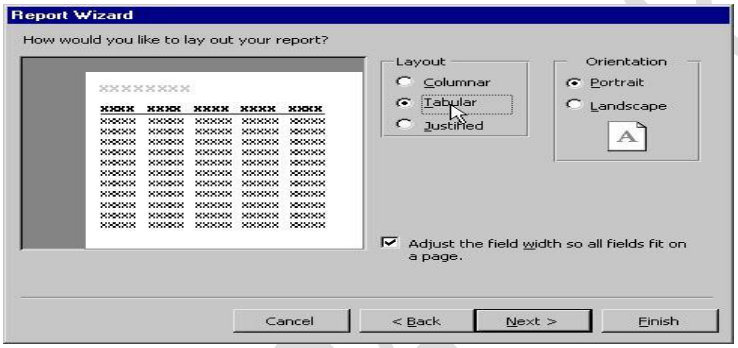
- On the next page of the wizard, specify whether you want to group your data according to certain fields. You can do this by selecting a field and clicking the arrow button to move it into the right half of the dialog box. To change the hierarchy of the group, click the **Priority** buttons. The right pane updates when you make changes, showing you how the grouping will affect the presentation of the data:



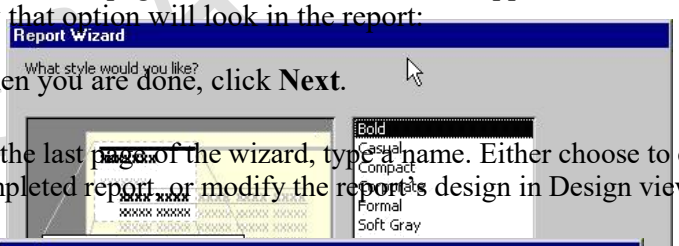
- When you are done, click **Next**.
- On the next page of the wizard, specify a sorting order. For every field you have included in the report, this page will have a drop-down list. Select an option from the first drop-down list to make the second drop-down list available, and so on. For fields you are sorting, you can toggle the **Ascending** button to **Descending** or vice versa:



- When you are done, click **Next**.
- On the next page of the wizard, specify a layout. When you click an option, the left pane shows how that option will look in the report. If you want the report to be sideways on the page, click the **Landscape** option instead of **Portrait**:



- When you are done, click **Next**.
- On the next page of the wizard, choose an appearance theme; the preview window shows you how that option will look in the report:



- When you are done, click **Next**.
- On the last page of the wizard, type a name. Either choose to close the wizard and preview the completed report, or modify the report's design in Design view first:

- When you are done, click **Finish**.

Creating a report from scratch in Design view

- In the Database window, click the **Reports** button along the left side of the window.
- Double-click **Create report in Design view**. A blank Design view window opens.
- Right-click the toolbar of the report window and choose **Properties** from the context menu. The **Properties** dialog box opens.
- Click the **Data** tab.
- Click inside the **Record Source** cell. A drop-down arrow then appears beside this cell, listing all of the available tables:

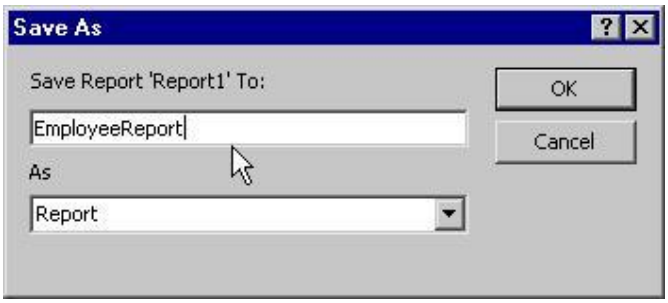
- Select the table that you want to use as the basis for the report. A small window listing the fields of that table then appears.
- Decide on the fields you want represented in the report, and drag a field from the window listing fields onto the report. Unless you want the field in a header or footer, you'll probably want to drag onto the **Detail** section of the report.
- Continue to drag fields onto the report until all of the desired fields are represented.
- If you are not happy with the placement of a field or label, click it to make the small black boxes appear around the object. You can then hold the mouse indicator over the object; it turns into a hand, which means you can drag the object to a new location.
- Use the options in the toolbox to add elements of design to the report. These are largely the same options as you learned to use in the Forms chapter. If the toolbox doesn't appear, click the **Toolbox** button on the toolbar.
- Add graphics (if desired) to the form in the same way as you did for forms. Like with forms, you can create an object frame on the report use the **Insert > Picture** option, or you can use the **Properties** dialog box (with the **Report** option selected from the drop-down list at the top) and click in the **Picture** cell to specify a background picture. These steps are all the same as you learned to do in the Forms chapter.

Saving, Maintaining, and Printing Reports

Saving a report

· Open the report. You can open it either in its regular view or in **Design** view.

- From the menu bar, choose **File > Save**. If you have saved the report before, Access will simply resave the report. If you have not yet saved the report, Access opens the **Save As** dialog box.
- In the **Save As** dialog box, type a name for the report. Then click **OK**:

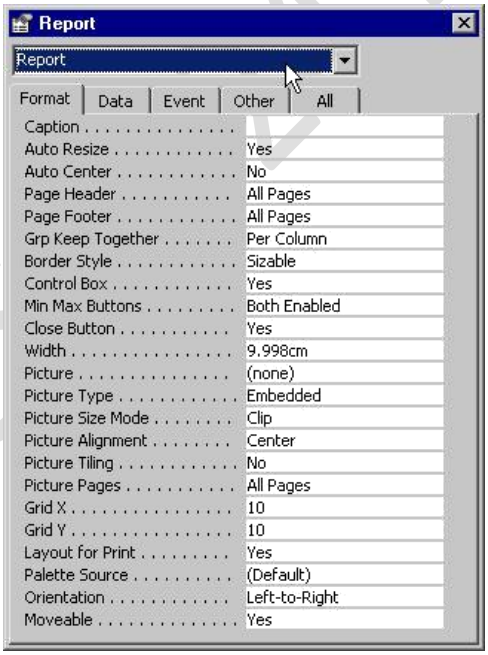


To save a report with a different name:

- Open the report. You can open it either in its regular view or in Design view.
- From the menu bar, choose **File > Save As**. Access opens the **Save As** dialog box.
- In the **Save As** dialog box, type a new name for the report and click **OK**.

Setting report properties

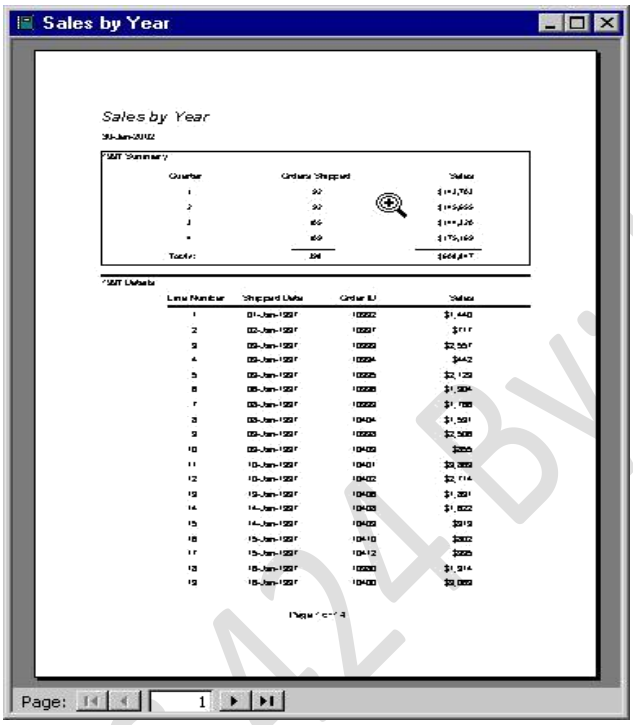
- In the **Database** window, click the **Reports** button along the left side of the window.
- Select the report you want to edit.
- Click the **Design** button on the toolbar of the **Database** window. The report opens in a design window.
- Right-click the title bar of the report window and choose **Properties** from the context menu. The **Report Properties** dialog box opens.
- Use the drop-down menu at the top of the dialog box to specify the portion of the report you want to set properties for. For example, if you want to set properties for the entire report, choose **Report** from this drop-down list:



- If you want to set properties just for a header or footer, or just for a particular field, choose that option from the drop-down list. The cells and properties in the lower half of the dialog box (on all of the tabs) may change depending on your selection in the top half of the dialog box. For example, the **Picture** options on the **Format** page may not be available for a particular field while they are available for the entire report.
- Close the **Properties** dialog box when you are finished and save the report.

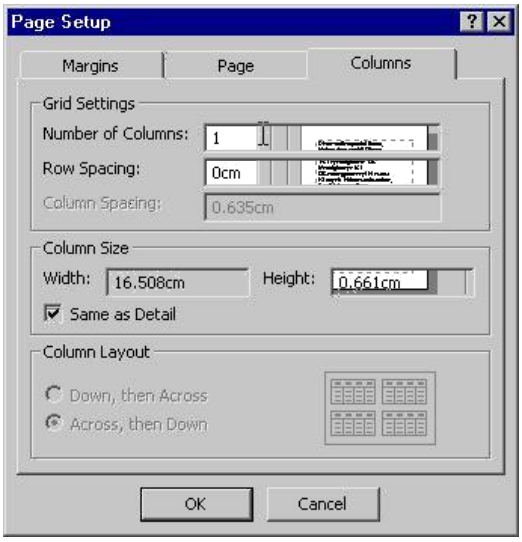
Previewing a report (before you print it)

- In the Database window, click the **Reports** button along the left side of the window.
- Right-click the report and choose **Print Preview** from the context menu.
- In the report window, you can see how the report will print. If the report spans multiple pages, you can use the arrow buttons at the bottom of the window to move to following or preceding pages:

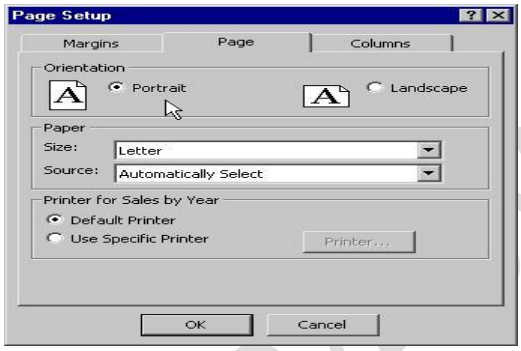


Printing a report

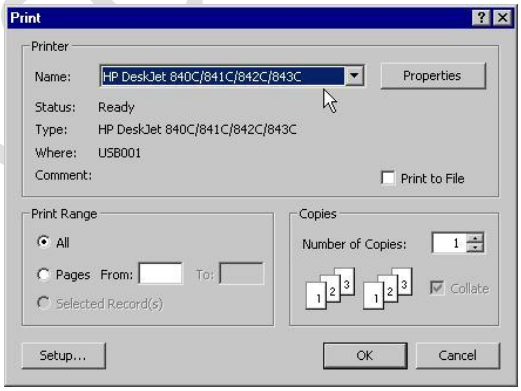
- In the **Database** window, click the **Reports** button along the left side of the window.
- Double-click the report to open it in its own preview window.
- From the menu bar, choose **File > Page Setup**. The **Page Setup** dialog box opens:



- On the **Margins** tab, make sure the margin settings are appropriate for the report.
- On the **Page** tab, make sure the page is set up properly. On this page, you can make the report print in a portrait manner (standard vertical view) or landscape (horizontal on the page).
- You can also specify the paper source for the printer and choose what printer you want to use:

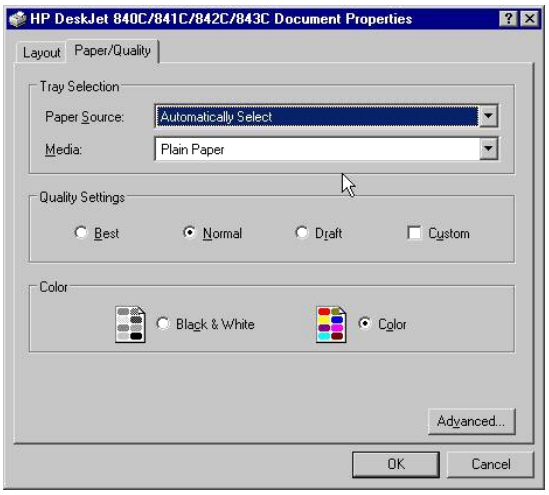


- On the **Columns** tab, specify whether you want to print the report in columns. If you specify a number higher than **1** in the **Number of Columns** field, the **Column Layout** options become available:
- Click **OK** when you are satisfied with the settings.
- From the menu bar, choose **File** > **Print**. The **Print** dialog box opens:



- This dialog box is the standard printing dialog box, which is the same for all other applications.

- You can configure the printer settings by clicking the **Properties** button. A **Properties** dialog box for your printer opens:



- Click the **Layout** tab to specify the paper orientation and the page order. For example, you can choose to print from the first page to the last, or vice versa.
- When the print job specifications are the way you want them to be, click **OK** to close the **Properties** dialog box.
- Check all of the settings in the **Print** dialog box, and click **OK** to start printing.

DESKTOP PUBLISHING DESIGN PRINCIPLE

Desktop publishing (also known as **DTP**) combines a personal computer and WYSIWYG (What You See Is What You Get) page layout software to create publication documents on a computer for either large scale publishing or small scale local multifunction peripheral output and distribution.

The term "desktop publishing" is commonly used to describe page layout skills. However, the skills and software are not limited to paper and book publishing. The same skills and software are often used to create graphics for point of sale displays, promotional items, trade show exhibits, retail package designs, and outdoor signs.

Desktop publishing began in 1984 with the introduction of MacPublisher, the first WYSIWYG layout program, which ran on the original 128K Macintosh computer. The DTP market exploded in 1985 with the introduction in January of the Apple LaserWriter printer, and later in July with the introduction of PageMaker software from Aldus which rapidly became the DTP industry standard software.

The ability to create WYSIWYG page layouts on screen and then print pages at crisp 300 dpi resolution was revolutionary for both the typesetting industry as well as the personal computer industry. Newspapers and other print publications made the move to DTP-based programs from older layout systems like Atex and other such programs in the early 1980s.

The term "desktop publishing" is attributed to Aldus Corporation founder Paul Brainerd, who sought a marketing catch-phrase to describe the small size and relative affordability of this suite of products in contrast to the expensive commercial phototypesetting equipment of the day.

Often considered a primary skill, increased accessibility to more user-friendly DTP software has made DTP a secondary skill to art direction, graphic design, multimedia development, marketing communications, administrative careers and advanced high school literacy in thriving economies. DTP skill levels range from what may be learned in a few hours (e.g. learning how to put clip art in a word processor) to what requires a college education and years of experience

(e.g. advertising agency positions). The discipline of DTP skills range from technical skills such as prepress production and programming to creative skills such as communication design and graphic image development.

By the standards of today, early desktop publishing was a primitive affair. Users of the PageMaker-LaserWriter-Macintosh 512K system endured frequent software crashes, the Mac's tiny 512 x 342 1-bit black and white screen, the inability to control letter spacing, kerning (the addition or removal of space between individual characters in a piece of typeset text to improve its appearance or alter its fit) and other typographic features, and discrepancies between the screen display and printed output. However, it was a revolutionary combination at the time, and was received with considerable acclaim.

Behind-the-scenes technologies developed by Adobe Systems set the foundation for professional desktop publishing applications. The LaserWriter and LaserWriter Plus printers included high quality, scalable Adobe PostScript-fonts built into their ROM memory. The LaserWriter's PostScript capability allowed publication designers to proof files on a local printer then print the same file at DTP service bureaus using optical resolution 600+ ppi PostScript-printers such as those from Linotronic. Later, the Macintosh II was released which was much more suitable for desktop publishing because of its larger, color screen, support for multiple displays, greater RAM capacity and its SCSI storage interface which allowed fast, high-capacity hard drives to be attached to the system.

Although Macintosh-based systems would continue to dominate the market, in 1986, the GEM-based Ventura Publisher was introduced for MS-DOS computers. While PageMaker's has a pasteboard metaphor closely simulated the process of creating layouts manually. Ventura Publisher automated the layout process through its use of tags/style sheets and automatically generated indices and other body matter. This made it suitable for manuals and other long-format documents. Desktop publishing moved into the home market in 1986 with Professional Page for the Amiga, Publishing Partner for the Atari ST, GST's Timeworks Publisher on the PC and Atari ST, Calamus for the Atari TT030, and even Home Publisher, Newsroom, and GEOPublish for 8-bit computers like the Apple II and Commodore 64.

During these early years, desktop publishing acquired a bad reputation from untrained users who created poorly-organized ransom note effect layouts — criticisms that would be levied again against early web publishers a decade later. However, some were able to realize truly professional results. For example, .info magazine became the very first desktop-published, full-color, newsstand magazine in the last quarter of 1986, using a combination of Commodore Amiga computers, Professional Page desktop publishing software, and an Agfa Graphics typesetter[1].

Page layout concepts

A page is a prefixed size of virtual printing material which can be viewed on the monitor in WYSIWYG format. Each page has full size and printable area. They are separated with margin guides. In most cases, the full size of page are set to international standard paper sizes such as A4, letter, etc.

Printing components

There are three main types of components to be laid out on a page. They are text, natural or scanned images, artificial or creative images.

Layout

Layout is the process by which the printing components are laid on the page aesthetically and precisely.

Comparisons with word processing

While desktop publishing software still provides extensive features necessary for print publishing, modern word processors now have publishing capabilities beyond those of many older DTP applications, blurring the line between word processing and desktop publishing.

In the early days of graphical user interfaces, DTP software was in a class of its own when compared to the fairly spartan word processing applications of the time. Programs such as WordPerfect and WordStar were still mainly text-based and offered little in the way of page layout, other than perhaps margins and line spacing. On the other hand, word processing software was necessary for features like indexing and spell checking, features that are today taken for granted.

As computers and operating systems have become more powerful, vendors have sought to provide users with a single application platform that can meet all needs. Software such as Open Office.org Writer and Microsoft Word offers advanced layouts and linking between documents, and DTP applications have added in common word processor features.

Comparisons with other electronic layout

In modern usage, DTP is not generally said to include tools such as TeX or troff, though both can easily be used on a modern desktop system and are standard with many Unix-like operating systems and readily available for other systems. The key difference between electronic typesetting software and DTP software is that DTP software is generally interactive and WYSIWYG in design, while older electronic typesetting software tends to operate in batch mode, requiring the user to enter the processing program's markup language manually without a direct visualization of the finished product. The older style of typesetting software occupies a substantial but shrinking niche in technical writing and textbook publication; however, since much software in this genre is freely available, it can be more cost-effective than the professionally-oriented DTP systems. It is also particularly suitable for corporate newsletters or other applications where consistent, automated layout is important.

There is some overlap between desktop publishing and what is known as Hypermedia publishing (i.e. Web design, Kiosk, CD-ROM). Many graphical HTML editors such as Microsoft FrontPage and Adobe Dreamweaver use a layout engine similar to a DTP program. However, some Web designers still prefer to write HTML without the assistance of a WYSIWYG editor and resort to such software, if at all, solely for complex layout that cannot easily be rendered in hand-written HTML code.

Desktop Publishing Using MS Word

While Microsoft Word is still first and foremost a word processor, it is frequently used to create publications normally created by page layout programs. For some users, it may be the only desktop publishing tool they need. Designers may cringe at the thought of designing in a word processor, but within reason it is possible.

Designed primarily for text-focused documents, Word and other word processors can still be used for combining text and graphics into fliers, brochures, newsletters, or business cards. Today many businesses require that their everyday forms such as letterhead, fax sheets, internal and external forms be in the Word .doc format.

If you or a client requires .doc compatible digital files, don't dispair. Work within the program's capabilities to use Word for some desktop publishing. However, be aware of the limitations as well — especially when it com es to full-color or high-volume commercial printing.

- Use Microsoft Word for creating documents for printing to a desktop or network printer.
- Do not use Microsoft Word, WordArt, or other components of Microsoft Office for creating logos.
- If the document must be created in Word and must be commercially-printed, work with a service that specializes in printing Word documents, learn the limitations of the format, and learn how to prepare digital files for commercial printing and the proper use of TIFF and EPS graphics.

Samples of simple DTP documents produced using MS Word.

1. Greeting Card

The families of
Alh. Musa Hassan
And
Malam Zahraddeen Wahda
Cordially invite

To attend the
WEDDING CEREMONY OF
Jilani Musa Hassan
&
Zainab Zahraddeen Wahda
Schedule as follows

Date: 1st October, 2008
Time: 11:00am
Venue: Printing Department, Kaduna Polytechnic, Kaduna

Book cover

The Layout of an SPSS Data File

Cases

Most data files are rectangular in shape. They have three components:

1. **Cases** (typically the rows of the rectangular file). Cases are the individual participants in your study.
2. **Variables** (typically the columns of the file). Here are some examples of variables -
   - Subject ID
   - Demographic variables (e.g., age, sex, race, etc.)
   - Treatment variables (e.g., experimental conditions)
   - Response variables (e.g., the scores the dependent variables)
3. **Values** (the intersection of cases and variables).

Here is an example of some data from a survey about attitudes towards the death penalty. In this study information was collected about age, gender, and attitude towards the death penalty (the variables) for each of 4 research participants (the cases). Attitude towards the death penalty was measured on a 6-point continuum with the following labels: strongly opposed, opposed, slightly opposed, slightly approve, approve, and strongly approve.

Table 1. Original Data

Participant	Age	Gender	Death Penalty
Jones, W.	25	Male	strongly opposed
Anderson, S		Female	slightly opposed
Perez, C.	18	Female	
Smith, L.	41	Male	approve

The values that are entered into a data file are typically **numeric (numeric values contain only numbers)** rather than **alphanumeric (alphanumeric or 'string' values contain letters, or combinations of letters and numbers, rather than only numbers)**. One reason for using numeric values rather than alphanumeric values is that many SPSS procedures will only accept numeric values. For example, an analysis of variance can only be run using numeric values. Although some SPSS procedures will accept either numeric or alphanumeric values (e.g., frequencies) numeric values can be used by a wider range of procedures. Another reason to use numeric values is that it is easier to enter single digit to refer to a value (e.g., 1) than it is to enter a whole series of letters (e.g., strongly opposed).

In this example the values for Age are numeric , and the values for Gender and Death Penalty are alphanumeric. The alphanumeric values should be coded as numeric values prior to entering them into a raw data file. Any numeric values could be used to code the nominal variable of Gender. Lets, arbitrarily, decide to code females as "1" and males as "2". Death penalty can be considered to be an interval variable, and the values should range from 1 to 6. Strongly opposed could be coded as either "1" or "6". Suppose the responses to the death penalty scale were coded as follows: 1 = *strongly opposed*; 2 = *opposed*; 3 = *slightly opposed*; 4 = *slightly approve*; 5 = *approve*; and 6 = *strongly approve*. The values 1 through 6 would be entered into the data file. The codes for those values (e.g., strongly opposed) are called the *value labels*.

The APA ethical guidelines stipulate that the data collected from research participants are to be confidential. Rather entering the names of the participants into the raw data file you should create an ID variable and number each of the participants. Making those changes the data now look like this:

Table 2. Data with Assigned Values

ID	Age	Gender	Death Penalty
001	25	2	1
002		1	3
003	18	1	
004	41	2	5

You have probably noticed that the age value is missing for participant #002. and that the death penalty value is missing for participant #003. When you are entering data you can leave those values blank. SPSS will consider them to be **system missing** values and correctly handle them when running analysis. There are other options for how to enter missing values and SPSS offers several ways of dealing with missing values in each of its procedures. We will have much more to say about this topic in later sections.

In order to run a statistical analysis of your data you first need to create a data file. You can create a data file using the Data Editor within SPSS or you can create a data file using your favorite word processor, using some spreadsheet programs (e.g., EXCEL, LOTUS 1-2-3), or using some database programs (e.g., dBase). See *SPSS Base*, Chapter 3, Data Files for more information on how to read a data file that was not created by the SPSS data editor.

How to create a data file using the SPSS Data Editor is described in Entering Data Using the SPSS Data Editor. How to create a data file using a word processor is described in Overview of ASCII Data Files

Overview

There are three steps to creating a data file using the SPSS Data Editor -

For this discussion, lets assume that we have run an experiment to test the hypothesis that writing an essay in favor of doubling tuition at UCCS would make people more accepting of an increase in tuition. The participants were randomly divided into two groups, one group wrote an essay about why tuition should be doubled and the other group wrote an essay about the value of intercollegiate sports. We then administered a questionnaire to gather some demographic information about the participants. The questionnaire also measured attitudes towards several issues, including the issue about doubling tuition. These variables are summarized in Table 1. The responses from 4 participants are shown in Table 2.

BIRTHDAY	Date (mm/dd/yyyy)	Date of birth of the respondent(mm/dd/yyyy)
GENDER	String 1	Gender of respondent/ "F" "FEMALE" "M" "MALE"
CLASS	Numeric 1.0	Year in college/ 1 'Freshman' 2 'Sophomore' 3 'Junior' 4 'Senior'

TUITION	Numeric 1.0	Tuition at UCCS should be doubled/ -same value labels as SHOCK
PHDPSY	Numeric 1.0	UCCS should have a Ph.D. program in Psychology/ -same value labels as SHOCK

The numbers in the Variable Type column indicate: whether the variable is a numeric, date, dollar, or string variable; the width of the variable; and for numeric variables, how many decimal places. The first number after numeric, date, dollar, and string variables is the maximum width of the variable (including the number of whole digits, the decimal point, if any, and the number of decimal digits, if any. The number after the decimal point is the number of decimal places for the variable. Note that string variables and date variables do not have decimal places.

Table 2. Data from 4 Participants in the Tuition Study

ID	BIRTHDAY	GENDER	CLASS	INCOME	ESSAY	EXTRAV	SHOCK	TUITION	PHDPSY
1	3/20/1975	F	1	1823.62	1	50	5	2	6
2	5/32/1977	M	3	128.50	2		6	2	4
3	10/3/68	F	4	1239252	2	23	5	9	5

There are no hard and fast rules about how to organize the variables within the data file. I typically begin with an ID number followed by demographic data (age, gender, SES, etc.), the independent variable condition(s), individual difference measures, and finally the responses to the dependent variables. The data in Table 2 follow that organization. The variables in this example were chosen to illustrate some of the most common data types that students in the department have used over the years.

Parenthetically, you will normally be entering your data directly from the response sheets that you give to your participants. I don't recommend copying over the responses into a tabular format as in Table 2. The extra step of copying over the data onto a sheet to give to a data entry person who then enters the data into the computer will most surely cause errors to creep into the data.

Variable Definition

(**Recommendation**: Open SPSS in another window and create the data file described in these notes.)

Variable definition includes naming the variable (**Variable Name**), defining the type of variable, e.g., numeric or string (**Variable Type**), giving a long name for the variable (**Variable Label**), providing descriptions of the values that are entered into the data file (**Value Labels**), and defining missing values (**Missing Values**). Each of those elements will be described

ID (integer variable)

a. Variable Name

Let's enter the variable definition information for each of the variables in Table 1. First, double left click on the word "var" in the 1st column (the upper left corner) of the worksheet. The *Define Variable* dialogue box will open and the word "VAR00001" will be highlighted in the *Variable Name* text box. You could use the *default* name for the variable, VAR00001. But it is better to use a name that is descriptive of the variable. Enter **ID** as the variable name. You

should become familiar with the rules for naming variables (see the SPSS Help window under *variable naming rules*). If you use names that begin with a letter, that contain only letters, numbers and the symbols @, #, _, $, or period, and that are no longer than 8 characters long you should run into no problems.
[**Note:** To find the rules for naming variables press: Help, Topics, Index. Then enter the phrase *variable names:rules*. Then press the Display button.]
Click OK at the bottom of the Define Variable window and enter the values for ID for the first four cases. An easy way to do this is to move the cursor to the 1:ID cell, highlight the cell and then key in the value, 1. Then press the arrow key in the direction that you want to move, down in this instance. The value will be entered into the data file and the next cell, 2:ID, will be highlighted. Continue until all four values have been entered. Notice that the values are displayed with two decimal places, even though you only entered a whole number. Open the Define Variable window again.
In the Variable Description section of the Define Variable window the *Type* is defined as Numeric8.2, there is no *Variable Label*, there are no *Missing values*, and the *Alignment* of the data is *Right* justified within the space allotted to variable. These are the **default values** for every new variable. A default value is the value that is assigned by SPSS in the absence of any information provided by the user. Each of these elements are described in more detail below.
Variable names are not case sensitive. The following names are identical: ID, Id, and id. The SPSS Data Editor always dislpays the SPSS variable name in lower case letters.

b. Variable Type

Move the cursor the *Type...* button and left click. (if you press the ENTER key the cursor to the *Define Variable* box will close.) After pressing the *Type...* button several variable type options are presented. In psychology the most commonly used variable types are *numeric*, *string*, and *date*. Numeric variables can consist of the digits from 0 through 9 and an optional decimal point. String variables can contain any letters, numbers, and symbols. Date variables typically consist of a year, month, and day, but they can also include hours, minutes and seconds. Typical date variables are date of birth, date and time of testing, etc.

ID is a numeric variable. Note that the "numeric" box is already checked. The *width* of the variable refers how many spaces will be reserved for the variable when its values are displayed. *Decimal places* refers to how many of the *width* digits will be reserved for the decimal point and the decimal part of the number. The width does not refer to how many digits are stored in the data file, width refers to how many digits will be displayed in the data editor and in the output. For example, if you set the width at 2 digits, then you can still enter a value that is 3 or more digits wide into the data file. Values that are wider than the defined width are displayed by an asterisk (*).

The default width is 8 digits; the default number of decimal places is 2, resulting in the data type of Numeric8.2. Notice that the values you entered: 1, 2, 3, and 4 are displayed as 1.00, 2.00, 3.00, and 4.00. The optimal width of a numeric variable is determined by the range of values that are possible for the variable. If you have, say, between 10 and 99 cases, then the width of the ID variable should be set at 2. If you have between 100 and 999 cases then the width of the ID variable should be set at 3. Lets set the *Width* to 2 digits and the number of decimal places to 0. Press continue to close the Type... dialog box. Then click "OK" to close the Define Variable window. Note that the values are displayed as whole numbers rather than as decimal numbers. Try entering a decimal number. Note that decimals will be rounded to whole numbers in the display. Remember that the width and number of decimal places refers to the display of the values, not to the actual number that is stored in the data file.

In SPSS version 8.0 and 9.0 the assigning the width of a numeric variable seems to have no effect on how that that variable is either saved or displayed.

c. Variable Label

Next, press the *Labels..* button. Two options appear in the dialog box: *Variable Label* and *Value Labels*. A *variable label* is a longer description of the variable. Recall that the name of the variable can be no longer than eight characters. It is not mandatory to have a *variable label*. For example, ID is descriptive in itself, you probably do not need to add a longer *variable label* such as "Participant Identification Number."

Variable labels will preserve the case (upper and lower case) as entered.

d. Value Labels Value labels identifies the coding scheme for the values. *Value labels* are not mandatory and they would not be used for ID values or other interval type data such as temperature values or scores on tests (e.g., you wouldn't label each value of an IQ score). *Value labels* are typically used when the value refers to a specific category such as "male" and "female," or the scale values for a Likert-type response scale, e.g., "strongly agree." Lets leave the *Labels* section blank for the ID variable. Click the *Cancel* button to exit the dialog box.

e. Missing Values: Because you assign the values of ID there "no missing values."
f. Column Format: Column format refers to how the values are displayed **in the Data Editor**. We have

already altered how the values of ID are displayed by assigning values for the width and number of decimal places. Entering a value for *Column Width* will change the width of the display for the data editor only. The values you entered for the Variable Type will be in effect for any output involving those values. To see how this works change the column width to 2 and press continue and then OK to exit the Define Variable dialog box. The width of the ID column has been narrowed to two print columns columns. Any number that is wider than 2 digits is displayed as an asterisk, "**". Try it for yourself.

You change change the display width of a variable by moving the cursor to the edge of name of the variable and then dragging the column to make it wider or narrower.

Numeric variables are always aligned to the right.

g. Measurement

Measurement refers to scale of measurement: nominal, ordinal, interval, or ratio. SPSS allows you to assign one of three categories of measurement: nominal, ordinal, or scale. "Scale" refers to both interval and ratio scales. There is only one place in SPSS for windows where this information is used: in some chart (graphics) procedures that identify the measurement type. The help files also indicate that this information is used when you an SPSS data file with a program called "Answer Tree." Answer Tree is not a part of SPSS 8.0 or 9.0

For a review of the scales of measurement see Scales of Measurement.

BIRTHDAY (date variable)

a. Variable Name

Click on the 2nd variable column on the worksheet. Enter BIRTHDAY as the variable name.

b. Variable Type: Open *Type...* and click the button in front of *Date.* Select *mm/dd/yyyy* from the date options that appear and click the *Continue* button. Variable and value labels are not necessary or desirable for this variable so you can close the Define Variable dialog box.

Enter the first birthday value from Table 2 in the 1:BIRTHDAY cell (3/20/1975).

Enter the date for case #002 (5/32/1977). The Data Editor should beep and refuse to enter a value because the date is not possible. Reenter the date as 5/30/1977.

Enter the date for case #003 (10/3/68). Note that the year in the data file has been changed to 1968, while the display remains at 68. If you enter a 2-digit year, SPSS will automatically add "19" to the year. If you want to enter a date that is not in the 20th century, then you must enter all the digits of the date. Try entering a date from the year 2001.

Enter the date for case #004 (10/10/1582). The Data Editor beeps and does not enter the date. Why? SPSS stores the date as the number of seconds from October 14, 1582 (the beginning of the Julian calendar). As a consequence you cannot enter a date that is on or before October 14, 1582. Enter the date as 10/15/1582.

c. Variable Label

Not needed.

d. Value Label

Not needed.

e. Missing Values

The easiest way to deal with missing values for date variables is to just leave the value blank. A blank numeric value will be displayed as a single period(.). Try deleting one of the date values and note that the result is a period for that case. A blank numeric value or a blank date value is defined as a **system missing value**. SPSS will correctly handle system missing values.

f. Column Formats

Date variables are 10 digits wide (2 for the month, 2 for the day, 4 for the year, and 1 each for the two "/" separators. Try changing the the column format for this date variable from 10 digits to 8 digits. What happens?

g. Measurement

The scale of measurement for a data variable is "scale."

GENDER (string variable)

a. Variable Name: Click on the 3rd variable column of worksheet. Enter GENDER as the name for the third variable.

b. Variable Type

Click *string* as the variable type. Because only one letter is needed to enter the M and F codes, enter 1 as the number of characters. The term "string" variable is synonymous with "alphanumeric" variable.

The data editor will only allow you to enter as many characters as you have defined in this dialog box. Because the number of characters was defined as 1, you will only be allowed to enter single characters. The data editor will beep at you if you try to enter more than one character.

c. Variable Label

This is another instance where the SPSS name for the variable is very descriptive. You may or may not wish to enter a variable label such as "Gender of the Respondent." Here is rule of thumb, think about another person working on the research project after you are gone, will that person clearly understand the SPSS variable name? If there is any possibility for misinterpretation you should include a longer variable label. SPSS variable names can be no longer than 8 characters, variable labels can be up to 256 characters long.

d. Value Label

SPSS is sensitive to the case of string values. The value "F" is different from the value "f". The reason for this is that strings are coded according their ASCII code. For example, the ASCII code for a capital F is 070 while the ASCII code for a small f is
102. This adds a complication to entering values for string variables You must be consistent in the case that you are using. In this example gender is coded as upper case M's and F's. If you used both upper and lowercase F's and M's as values for gender then SPSS would think you had four different genders, F's, f's. M's, and m's.

Suppose you decide to enter the value M for males and the value F for females in the datafile. Then you can attach the value label "male" to the letter M and the value label "female" to the letter F. Enter M in the *Value:* box. Then enter "male" (without the quotes) in the *Value Label:* box. Then press the *Add* button to add the value label to the list of value labels. Do the same for F and "female."

e. Missing Values

There is no such thing as a **system missing value** for a string variable. Blank string values are considered to be valid values. If you have missing data for gender and happen to leave the value blank, SPSS will think you have three valid genders, M, F, and blank. Therefore you must set up a **user-defined missing value** for string variables.

To enter user-defined missing values click - Define Variable
Missing Values...
Discrete missing values

Your cursor should be in the first of the three boxes, press the space bar once and then click the Continue button. You have now defined a single blank as a user missing value.

f. Column Formats

Recall that the *column formats* dialog box refers to the display of values in the data editor. The column width was automatically set to the width of the string variable, 1 in this instance. The text allignment was autmatically set to left. By default, string or text variables are left justified. SPSS assumes that all string variables are left justified. Numeric variables are right justified.

g. Measurement

When you identified the variable type as "string," the measurement type was automatically set to "nominal." Nominal is the correct measurement type for gender.

Close the Define Variable dialog box.

Entering string data

Variables that have defined value labels can be entered in one of two ways: (a) by entering the value itself, or (b) by clicking on the label.

Note that only the first letter of the variable name, "g," is displayed in the data editor. In order to see what is happening when we enter the data, use your cursor to drag the column width wider so that you can read the whole variable name, "gender." As you expand the display width of the variable you probably see the value label for gender rather than the value itself. You can toggle between displaying values and value labels. The toggle switch is located in the **Value Labels icon** near the right end of the icon bar in the Data Editor, and in the

View pop-up menu
-Value Labels

Lets start by turning on the display of value labels.

Entering the value itself

Move the cursor to cell 1:gender and enter the value for the first case, "F." The value label that you defined for "F," Female, appears in the cell. Move the cursor to cell 2:gender and enter the value for the second case, "M." The value value label that you defined for "M," Male, appears in the cell. The values for string variables are stored in the SPSS system file as the ASCII value, the Data Editor can display either the value or its associated value label. Try toggling back and forth between displaying the ASCII values and the associated value labels. The third and fourth cases our example data are females. This time enter their values as lowercase "f"s and then switch back and forth between viewing the value and the value label. Note that "f" is not associated with a label. Only the value is displayed. You now have two different values in your dataset for females, "F" and "f". If you ran a frequencies on gender it would provide a separate count for all the "F"s and all the "f"s in the datafile. Remember to be consistent in the case that you use when entering string data.

(Note: My version of SPSS, 8.0.0, seems to have a bug. When the "column width" is set to 1 the value "M" does not appear in the cell, although the value "F" does.)

Using Value Labels to enter the value

The data editor will allow you to enter values by clicking on the value label itself. To try this out toggle the Value Labels switch to display value labels. Highlight the cell to enter data and press Ctrl-(left)click. A value labels box will open. You can double (left)click the appropriate value to enter it into the data file. The value that is coded into the data file is shown at the top in the cell editor window.

This seems like a very long way to go about entering the values M and F. However, I can see some value in this approach if you have a very long list of value labels. Suppose that you are classifying psychological traumas and have a list of 20 different codes. It might be helpful to be able to choose from a list that comes up.

Additional Considerations for String Variables

This string example was simplified by limiting the width of the string to a single character. Things get more complicated if your string values are longer than one character. Lets consider some of the complications that occur if we set the width of the gender variable to two characters rather than one character. There are now eight possible ways of entering upper and lower case M's and F's: M-space, space-M, m-space, space-m, F-space, space-F, f-space, and space-f. If you entered each of those variations, SPSS would identify eight different values for gender.

Suppose you enter M and then press Enter or an arrow key. Will SPSS interpret this as M-space or space-M? By default SPSS always left justifies string variables. If there are spaces left over then SPSS pads with blanks to right. So, M would be entered into the data file as M-space. (Numeric variables are always right justified.)

You need to remember those rules when you assign value labels. If you enter M without any spaces with value label will look like this, M = "Male". SPSS will pad to right with blanks and look for the value M-blank. That is, blank-M will not be assigned the value label of "Male."

What do you suppose happens if you define a single space as a user-missing value? If a string variable that is two characters wide (String 2) is left blank will that blank value be defined as missing?

To summarize: (a) string variables are left justified, (b) SPSS pads string variables to the right with blanks.

CLASS (integer numeric with value labels)

a. Variable Name

Click on the 4th variable column of the worksheet.

Define Variable
Variable Name:

Enter CLASS as the variable name.

b. Variable Type

Define Variable
Type...

Variable Type is Numeric1.0 (width = 1, decimal places = 0)

c. Variable Label

Define Variable
Labels...
Variable Label

The variable name "class" may not be readily understood by someone else so you should enter a variable label. How about "Year in school?"

d. Value Label

Define Variable
Labels...

Value Labels
Label:
Value Label:

Enter the four value labels: 1 = 'Freshman', 2 = 'Sophomore', 3 = 'Junior', and 4 = 'Senior.'

e. Missing Values.

Missing values should be left blank. SPSS will consider the blank numeric values to be **system missing values**.

f. Column Formats

The column format is set to the width of the variable, 1 in this instance. It can be left as is, or you could set it to a wider value.

g. Measurement

What scale of measurement is "year in school?" It is at least ordinal, a person has earned more credit hours if he or she is a senior than if he or she is a junior. Is the interval between freshman and sophomore the same as the interval between sophomore and junior?

INCOME (Decimal Numeric)

Income is saved in dollars and cents. The variable variable type is "dollar." Because INCOME has values both to the right and left of the decimal point it is a decimal numeric number. You could define the width as 11 and the decimal places as 2, indicating that the total number of digits displayed, including the decimal point and the dollar sign, is 11 and the number of digits displayed after the decimal place is 2. Remember that this format determines how the values will be displayed. The format does not affect how the values are stored, or how the values are used in computations.

a. Variable Name

Click on the 5th variable column of the worksheet.

Define Variable
Variable Name:

Enter INCOME as the variable name.

b. Variable Type

Define Variable
Type...

The *variable type* is Dollar.

c. Variable Label

Define Variable
Labels...
Variable Label:

Enter "Monthly income" as the *Variable Label*.

d. Value Label

There is no reason to label each individual value for the INCOME variable.

e. Missing Values

I recommend using **system missing values** for this variable. Simply leave the cell blank if the income data is missing.

f. Column Formats

The column format for the data editor has been set automatically to the width of the dollar variable.

Commas will be displayed in the data editor if the width of the variable will accommodate them, otherwise the display of commas is suppressed. Decimal places will be displayed if the width of the variable will accommodate them, otherwise the display of the decimal part of the number is suppression. The dollar sign will be displayed if the width of the variable will accommodate them, otherwise the display of the dollar sign will be suppressed. After entering the income data use your cursor to change the width of the income variable and note the changing display for commas, cents, and the dollar sign.

g. Measurement

The scale of measurement for the dollar variable is "scale." It is actually a ratio variable, dollar has a rational zero point. It is right justified, as are all numeric variables.

ESSAY (Integer Numeric with Value Labels)

a. Variable Name

Click on the 6th variable column of the worksheet.

Define Variable
Variable Name:

Enter ESSAY as the variable name.

b. Variable Type

Define Variable
Type...

It takes only a single digit to code the values of ESSAY, the independent variable.
The essay score is a one-digit, whole number. Therefore the *variable type* is Numeric.
Set the *width* to 1 and number of *decimal places* to 0.

c. Variable Label

Define Variable
Labels...
Variable Label:

The *Variable Label* could be "Essay Condition."

d. Value Label
Define Variable
Labels...
Value:
Value Label:

Enter the two value labels: 1 = "doubling tuition" 2 = "intercollegiate athletics."

e. Missing Values

Everyone should have been assigned to a treatment condition so there should be no missing values.

f. Column Formats

Column format has been set to a width of 1, the width of the variable set in variable type. This is a numeric variable so it is right justified.

g. Measurement

What is the scale of measurement for this variable?

EXTRAV (Integer Numeric without Value Labels)

a. Variable Name

Click on the 7th variable column of the worksheet.

Define Variable
Variable Name:

Enter EXTRAV as the variable name.

b. Variable Type

The extraversion score is a two-digit, whole number. Therefore the *variable type* is Numeric. Set the *width* to 2 and number of *decimal places* to 0.

c. Variable Label

Define Variable
Labels...
Variable Label:

The extraversion score is a subscale of the Eysenck Personality Inventory (EPI). The *Variable Label* might be "Eysenck Personality Inventory (EPI): Extraversion Subscale."

d. Value Label Individual values of scales (e.g., personality measures, IQ, etc.) are not labeled.

e. Missing Values

I recommend using **system missing values** for this variable. Simply leave the cell blank if the extraversion score is missing.

f. Column Formats

Changing the *Column Formats* will change change how the values are displayed in the Data Editor.

g. Measurement

The scale of measurement for nearly all personality scales is interval. This extraversion scale is measured as an interval scale so the measurement type should be set to "scale."

SHOCK, TUITION, PHDPSY (Integer Numeric with Common Value Labels)

SHOCK, TUITION, and PHDPSY have a common set of value labels. You can set up a template and then use that template to define the common value labels, missing values, and column format for the other variables. To define a template click

Data (on the top row of buttons)
Templates...
Define>>
Name:

Enter a name for the template, e.g., "6-PT SCALE" and click the *Add* button to add this to the list of templates.
Type...
Set the variable type as Numeric1.0 (width = 1, decimal places = 0)
Value Labels...
Enter the following value labels: 1 = "Strongly Disagree"
2 = "Disagree"
3 = "Slightly Disagree"
4 = "Slightly Agree"
5 = "Agree"
6 = "Strongly Agree"
9 = "No Opinion on This Issue"
Missing Values... The value "9" (no opinion on this issue) is an out-of-range value that should not be considered when finding the mean or other statistics for this variable. All 9s will be considered to be valid values unless you explicitly define them as missing. Because you define which values are to be considered missing those values are called **user-defined missing values**.
There is only a single missing value for this variable. Click the *discrete missing values* option, enter "9" (without the quotes) in the first box, and then press *Continue*.
See Missing Values Summary for additional information about user-defined missing values.
Column Format...
Enter a value for Column Format, e.g., 4 or 5 and press *Continue*.
To save the new template press *Close.*
The template can now be used to define new variables. First, define the variable name and variable label and then use the template to add the variable type, value labels, missing values, and column format.
SHOCK
a. Variable Name

Click on the 8th variable column of the worksheet.
Define Variable
Variable Name:
Enter SHOCK as the variable name.
b. Variable Label
Define Variable
Labels...
Variable Label:
Enter "Electric shock should not be used in experiments" and click *Continue*. Close the *Define Variable* dialog box by clicking the *OK* button.
Add the values from the 6-PT SCALE template.
Data
Templates...
Select the 6-PT SCALE template
Click all the options in the *Apply* section.
Click *OK* to add the template values and close the template.
You can reopen the Define Variable dialog box to verify that the template has been appriopriately applied.
The scale of measurement for this type of variable is a source of some controversy among statisticians. The most conservative statisticians argue that this 6-point Likert type scale is at most ordinal. The majority would agree that the scale is interval. I know of no case where a journal editor has rejected a paper where t-tests or analyses of variance (statistics for interval level of measurement) had been performed on Likert-type scales. I go with "interval" as the scale of measurement for this type of scale.

Follow the same steps to define TUITION and PHDPSY.

COMMENT: Templates are stored as a part of the SPSS program on your computer, not as a part of the data file that you have created. There is no way that I know of to save a template to a floppy disk. Unless you log onto the same computer again you will probably have to reenter the template again.

Concept of Missing Values

System missing values
-Leave the cell blank.
-No special action needs to be taken when system missing values are used in an SPSS data file.
User-defined missing values
-User-defined missing values should be clearly "out -of-range" of the regular values for the variable, e.g., "-9" or "-99" for age.
-If user-defined missing values are used then you must specify which values are to be considered missing. If you forget to define the user-missing values, they will be considered to be valid values by SPSS. For example, the -99s will be used as valid ages when computing the mean age for the participants.

Enter the Values

In this set of notes we have entered the values for most of the variables as we defined each variable. Normally you would enter the values on a case by case basis. If you have not done so please go back and enter the remainder of the data from Table 2. Key in the data and then use the arrow key to both enter the data into data file and move to the next cell.

Editing already entered values is as simple as going to the cell and changing the value. You can insert a new case or a new variable. Move the cursor to the row or column where you want to make the insertion and then click on *Data* and then click either *Insert Case* or *Insert Variable.*
You can go to a particular case by clicking on *Data* then *Go to Case* then enter the case number.
It is possible to cut, copy, paste, and clear entire rows, columns, or blocks of data. Highlight the data and use the editing tools under the Edit button or press the right mouse button to bring up the editing tools.

Save the Data File

You can run SPSS procedures using the data file you have just created. But if you exit SPSS without saving the file it will be lost. To save the file click on *File* (top row of buttons) then *Save*. Enter a file name and then click *Save*. The file will be saved as an SPSS for Windows 9.0 **system file** with the extension ".sav." A **system file** is a file created by SPSS that includes all the values plus all the variable definition information.

If you choose the *Save As* option you can save the file in other formats including several different SPSS formats, ASCII, and several spreadsheet formats. Of particular interest is the SPSS *portable* format. The SPSS portable format is a generic format that can be read by any version of SPSS (e.g., Unix, Vax, and IBM mainframe version, SPSS/PC+, and SPSS version 7.0). I used to recommend that when you leave UCCS you should save your files in the portable format, you will be able to read them in any version of SPSS. Most institutions now use SPSS for Windows rather than mainframe versions so this is no longer a problem.

IMPORTANT: DO NOT EDIT AN SPSS SYSTEMS FILE WITH A WORD OR TEXT PROCESSOR.

An SPSS system file is a specially formatted file that should not be edited with a word or text processor. If you try to edit a systems file outside of the SPSS Data Editor you may destroy structure of the file and you may not be able to open it again within SPSS. If you happen to open an SPSS systems file with a word processor close the file without making any changes to it.

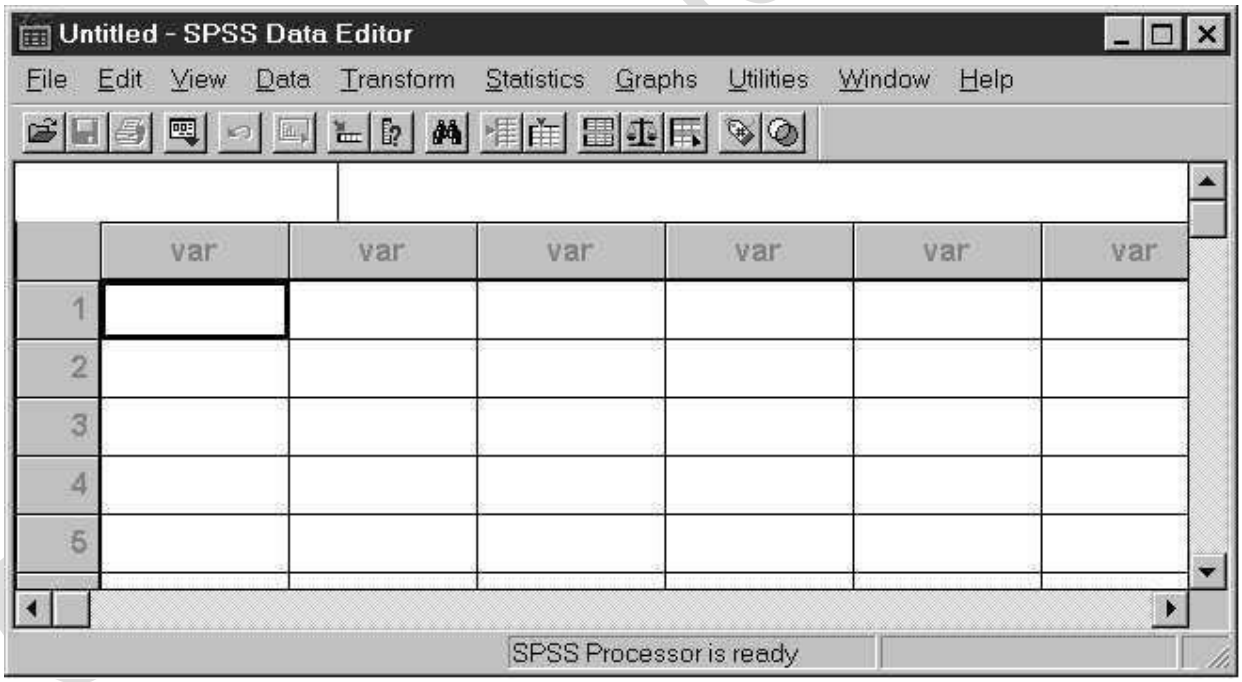
Recall that string variables are stored as ASCII values. If you open an SPSS systems file you would be able to read the values of the string variables. Numeric variables are stored as their binary equivalents. You will not be able to read the values of numeric variables because your text editor only reads ASCII values. A word processor will attempt to interpret the value of a binary, numeric variable as if it were an ASCII value. It would give you strange values. If you happen to change even one character in a systems file you could damage the values of one or more variables, or even make the file inaccessible to SPSS.

Caution: You should always back up your data files to a floppy disk.

How to input data into the SPSS data editor

Introduction

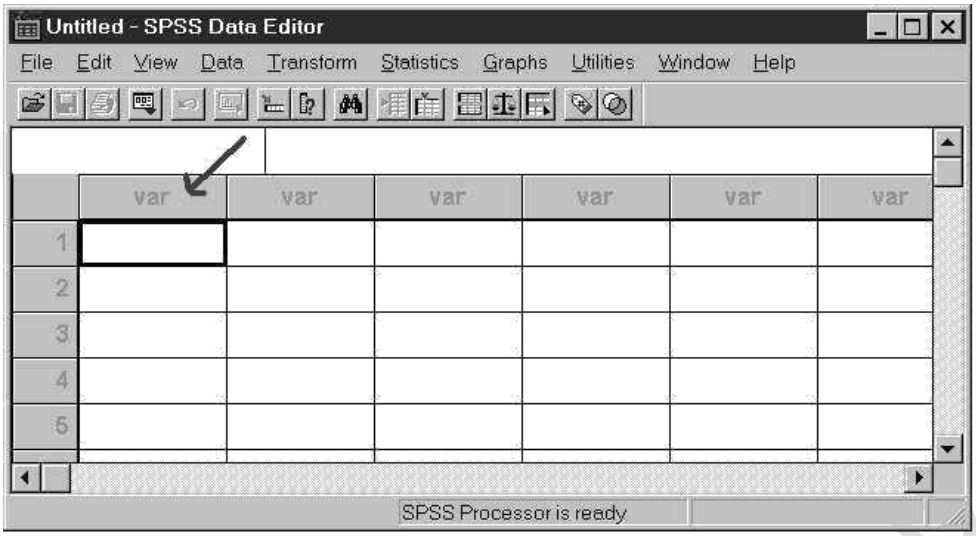
This page shows the basics of entering data into the SPSS data editor. The SPSS data editor can be a good choice for entering your data. It has a friendly interface that resembles an Excel spreadsheet and by entering the data directly into SPSS, you don't need to worry about converting the data from some other format into SPSS. For example, you might enter your data in Excel, and then try to convert it to SPSS and find out that you used the latest version of Excel, but your version of SPSS has trouble reading the latest Excel files.
Below is a screen snapshot of what the SPSS data editor looks like when you start SPSS. As you see, it does look like an Excel spreadsheet. In this editor, the columns will represent your variables, and the rows will represent your observations (sometimes called records, subjects or cases).



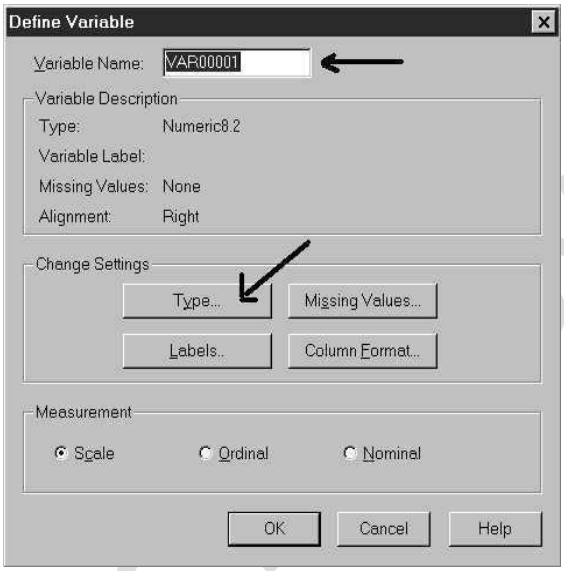
Let's illustrate how to enter data into the SPSS data editor using the data file below.

```
make          mpgweight price
AMC Concord    222930   4099
AMC Pacer      173350   4749
AMC Spirit     222640   3799
Buick Century 20 3250   4816
Buick Electra 15 4080   7827
```

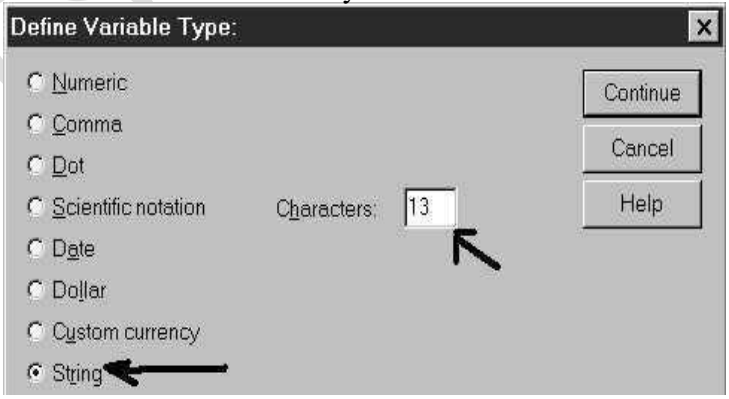
First, you need to tell SPSS the names of your variables. You can double click on the column heading (shown with the arrow below for the first column). That permits you to enter information about the variable for that column.



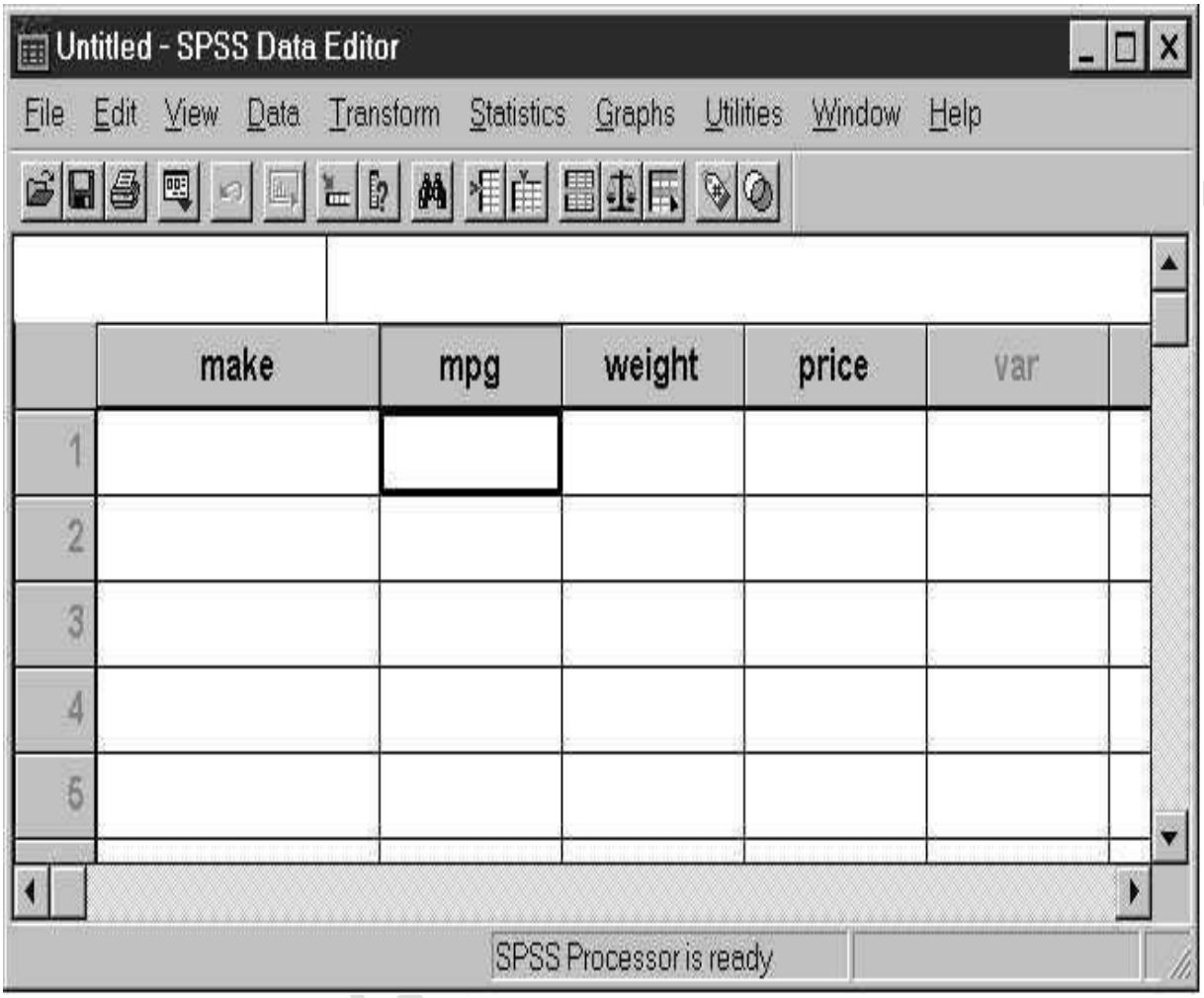
Below you see the dialog box that appears that allows you to enter information about your variable. For the first variable, let's change the Variable Name to be **mpg** (see arrow) and click on **Type** so we can tell SPSS that this is a string variable.



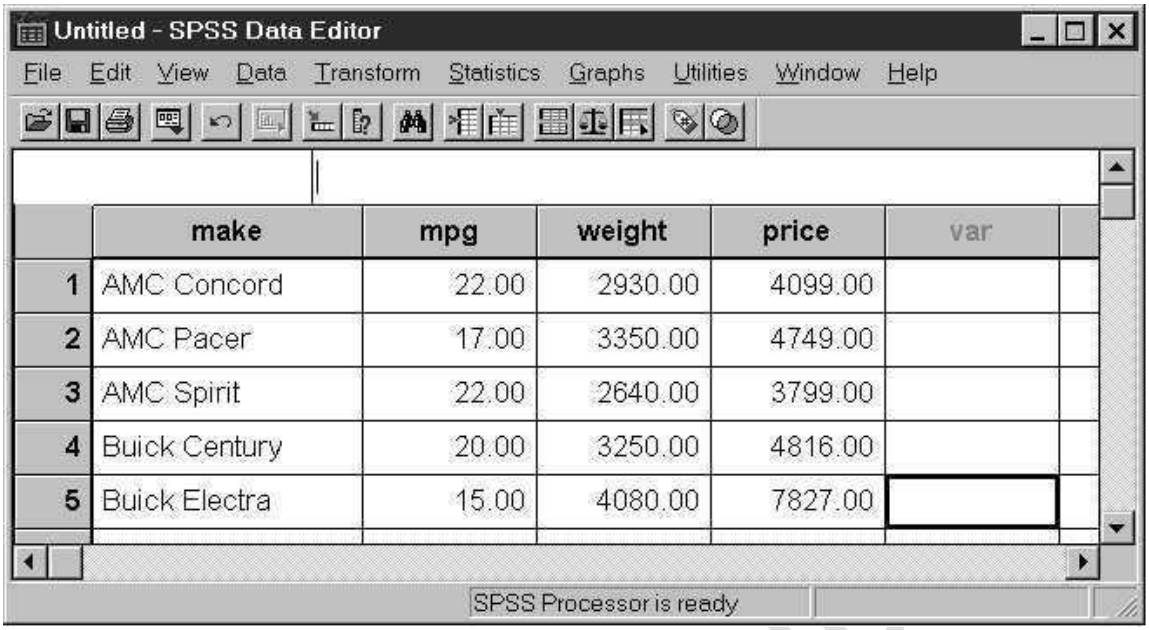
You can tell SPSS that **make** is a string by clicking **String** at the left (see left arrow) and then we should indicate that the length is up to 13 characters (see arrow at right).
You can then click **Continue**. When you return to the **Define Variable** dialog, click **OK**.



Click the column heading for column 2 and supply the name **mpg** and click **ok**. Click the column heading for column 3 and indicate the variable name is **weight**, and likewise for column 4 indicate that the variable name is **price**. Once you have done this, the Data Editor will look like below.



Once you have created the column headings (variable names) you are ready to enter the data. It is usually best to enter the data one observation at a time going from left to right. After you type in a entry for a variable, you can press the **Tab** key to move to the next variable on the right. Once you reach the last column (**race)** then use the arrow keys to move to the first column of the next observation. Once you have entered the sample data file, the SPSS Data Editor would look like this.



	make	mpg	weight	price	var
1	AMC Concord	22.00	2930.00	4099.00	
2	AMC Pacer	17.00	3350.00	4749.00	
3	AMC Spirit	22.00	2640.00	3799.00	
4	Buick Century	20.00	3250.00	4816.00	
5	Buick Electra	15.00	4080.00	7827.00	

You can save your data file by clicking **File** then **Save**. It would be wise to save your data about every 10-15 minutes. Imagine spending three hours typing in data, and then the power goes out, your computer stops responding, and then you have to enter the data all over again. The data are not saved when you type them in; rather, the file is saved when you choose **File** then **Save**.

Problems to look out for

- If you have character data, be sure to click **Type** in the **Data Definition** window and tell SPSS that your variable is a string variable. SPSS will not allow you to enter character data into a variable unless you tell it that the variable is a string variable.

For more information

- For more detailed information about the SPSS Data Editor, see the SPSS Manual.

The homework problems will state the data set that they are using, generally using a data set called GSS2000.SAV, a subset of cases and variables from the 2000 General Social Survey.

FREQUENCIES AND DESCRIPTIVE PROCEDURES

The FREQUENCIES and DESCRIPTIVE procedures have two uses. One important use is to make another check on whether you have entered the data correctly. You should make it a habit to run FREQUENCIES on all your variables before you start running other statistics or tests of your hypotheses. Carefully examine each variable looking for values that may have been entered by mistake. Check for out-of-range values. For example, are there 6's in a variable which should only have the values between 1 and 5? Check to make sure that data make sense. If you know that about 60% of your subjects were females and the percent females turns out to be only 20% then you know you have a problem in the date set or in the DATA DEFINITION.

A second important use is to provide descriptive statistics. Descriptive statistics are often used to describe the demographic characteristics of the sample. What percent of males and females are there? What was the range of ages and what was the median age? Survey data is often reported in terms of frequencies.

We have already used the FREQUENCIES procedure. Now we get to look at all the various possible options and statistics that are available within the FREQUENCIES procedure.

The output created by running the FREQUENCIES procedure is saved in the **Viewer** window. We will discuss how to manipulate data in the Viewer window, how to save data in the Output Navigator window, and how to print data from that window.

The data that we will be using is your Skills Survey data, *skills99.sav*. The discussion below assumes that the file skills99.sav has been opened in the SPSS Data Editor.

As always, you should open an SPSS window and follow along by actually running the SPSS procedures.

The general strategy for running any SPSS procedure is as follows:

Select the Procedure		Select the Variables		Select the Options	
Run the Procedure		Interpret the Output		Save or Print the Output	

2. Select the Frequencies Procedure

The frequencies dialog box is opened by clicking

Analyze
Descriptive Statistics
Frequencies

The dialog box shows the variable labels for each of the variables in the data set in the left box. The variables are organized according to their order in the active file. The label for the variable is displayed rather than the 8-character name of the variable. Variables that are preceded by the "#" symbol (within a diamond shape) are numeric variables. Variables that are preceded by the "A>" symbol (within a box) are string variables. You can see additional information about a variable by right clicking a variable and then selecting the **Variable Information** option. The variables to be analyzed will be moved to the empty box to the right.
Note: You have the option of displaying the names of the variables either in the order that they appear in the file, or alphabetically according to the name of the variable. You also can choose to display either the variable names or the variable labels. These options can be changed by in the SPSS Options dialog box.
Edit (in the top row of bttons)
Options...
General (Tab)
See the **Variable Lists** section
The **Statistics...**, **Charts...**, and output **Format...** options are accessed by the buttons at the bottom of the dialog window. The buttons to the far right will run the procedure (**OK**), **Paste** the commands into the syntax window, **Reset** the variables to be analyzed, **Cancel** the frequencies

procedure, and provide **Help**. You can turn the frequencies tables output on or off by checking the **Display frequencies tables** option.

3. Select One or More Variables

You can select a variable to be analyzed by double clicking the variable. The variable will move from the left pane to the **Variables:** pane on the right. Double clicking a variable in the **Variables:** pane will unselect the variable. You could also single click on a variable to highlight it and then press the right arrow button (located between the two panes) to move it to the **Variables:** pane.

Selecting Multiple Variables

Multiple variables can be selected in one of three ways.

a) Selecting another variable and moving it to the Variable(s): box by double clicking the variable or by highlighting the variable and pressing the right arrow button.

b) You can press and holding the **ctrl** key while selecting (highlighting) a set of variables and then move the entire set to the Variable(s): box. For example you could select all the knowledge of statistics items for analysis by holding down the ctrl key while selecting the variables for knowledge of analysis of variance (ANOVA), chi square, correlation, factor analysis, frequency distribution, multiple regression, repeated measures ANOVA, and t test and then press the right arrow key to move the entire set to the Variable(s): box.

c) If the variables you wish to select are contiguous to each other you can pressing and holding the **shift** key while selecting a set of variables. For example you could select all the variables for analysis by clicking the variable at the top of the list, pressing and holding the shift key, and then clicking the variable at the bottom of the list. Or you could hold down the shift key while moving the cursor down the variable list.

If you wanted to run frequencies on nearly all the variables you could use the shift key to select all the variables and then use the ctrl key to unselect the unwanted variables. For example, suppose that you wanted to run frequencies on all the variables except ID. First, use the shift key to select all the variables and then press and hold the cltr key while you click on the ID variable.

Run the Procedure

Once a variable is selected the **OK** and **Paste** buttons become available. Make sure the **Display frequency tables** option is checked and then press the **OK** button to run the frequencies procedure on the selected variable.

Interpret the Output

The results are shown in an SPSS Viewer window.. The left pane of the viewer window contains an outline of the information, the right pane contains the data.

The output is divided into four sections: title, notes, a statistics summary for all the variables selected, and frequencies tables for each of the selected variables. Each element in the output is an "object" that can be edited, saved, and or printed.

Title. The default title is the name of the procedure that was run. In this case the title is **Frequencies.** You edit the title by moving the cursor to the title and then double clicking the left mouse button. If

you selected all the statistics questions from the skills survey you could change the title to, say, Frequencies: Knowledge of Statistics.

Notes. The notes section contains information about when the statistics were run, some information about the data file, how missing values were handled, the syntax commands that were generated, and the resources used. The notes section is normally closed (the icon next to "Notes" is a closed book rather than an open book). You can open (display) the notes section by double clicking on the closed book icon in front of "Notes." Double click on the open book icon to close the Notes section.

Statistics. The statistics section includes statistics for all the selected variables. The default statistics include the number of valid cases and the number of missing cases for each variable analyzed.

Frequencies tables. The final section includes frequencies tables for each of the selected variables. The first column of the table lists the value labels for the **valid** values followed by the **missing** values and then the **total** number cases. Value labels for each non-null value (i.e., for each value that has a frequency of at least 1). The second column list the **frequencies** for each value. The third column lists the **percent** of cases for each value. This percent is based on the total number of cases. the fourth column lists the **valid percent** of cases for each value. The valid percent is based on the number of valid cases. If there are no missing values then the percent and valid percent columns will be identical. The last column lists the **cumulative percent**. The cumulative percent is based on the valid percent column.

4. Statistics Options

Click the **Statistics...** button at the bottom of the Frequencies dialog box to display the statistics that are available.

The statistics are grouped into four sections: measures of central tendency (mean, median, mode, and sum); measures of dispersion (standard deviation, variance, range, minimum, maximum, and standard error of the mean); measures of distribution (skewness and kurtosis), and percentile values.

Measures of central tendency (mean, median, mode, and sum). You should all recognize these central tendency measures.

Measures of dispersion (standard deviation, variance, range, minimum, maximum, and standard error of the mean). These measures are also very common. The standard error of the mean is found by dividing the standard deviation by the square root of the number of valid cases.

$$S.E.\ Mean = \frac{SD}{\sqrt{n}}$$

Measures of distribution (skewness and kurtosis). The terms skewness and kurtosis refer to distribution shapes that deviate from the shape of a normal distribution.

A skewed distribution is characterized as by a tail off towards the high end of the scale (a positive skew) or towards the low end of the scale (a negative skew). If the distribution has no skewness, then the skewness statistic will be zero. If the distribution has a positive skew, then the skewness statistic will be positive. If the distribution has a negative skew then the skewness statistic will be negative.

A distribution with kurtosis is characterized by the distribution being to narrow and peaked (a leptokurtic distribution) or too wide and flat (a platykurtic distribution). Again, if there is no kurtosis, the kurtosis statistic will be zero. If the distribution is leptokurtic, then the kurtosis statistic will be positive. If the distribution is platykurtic, then the kurtosis statistic will be negative.

A normal distribution has both no skewness and no kurtosis. As for any statistic, the actual values of the skewness and kurtosis statistics rarely turn out to be exactly zero. That is, if you randomly sampled a set of values from a population that was perfectly normal, it is unlikely that the skewness and kurtosis statistics would both be equal to zero. The question becomes, are the skewness or kurtosis scores so different from zero that we have to reject the hypothesis that they represent a normal distribution. We do this by setting up a 95% confidence interval (C.I.) around the skewness score and another 95% confidence interval around the kurtosis score. If the 95% confidence interval includes the value zero then we cannot reject the hypothesis that the distribution has no skewness (or no kurtosis).

The 95% confidence intervals are defined as

95% C.I. = skewness statistic ± 1.96 * (standard error of skewness)

and

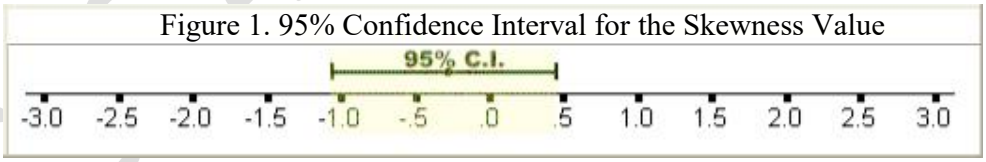
95% C. I. = kurtosis statistic ± 1.96 * (standard error of kurtosis).

For example, suppose the skewness statistic for the knowledge of correlations question was -.339 and the standard error of skewness was .388. Is the distribution for the correlation question negatively skewed? The 95% confidence interval is found as follows:

95% C. I. = skewness statistic ± 1.96 * (standard error of skewness)
= -.339 ± 1.96 * .388
= -.339 ± 0.761
= (-.339 - 0.761) to (-.339 + 0.761)
= -1.100 to 0.422

A graphic representation of the 95% confidence interval for this skewness value is shown in Figure 1.

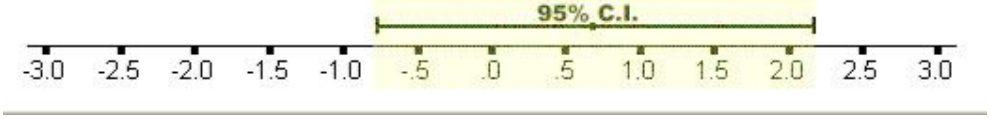
Figure 1. 95% Confidence Interval for the Skewness Value



The 95% confidence interval ranges from -1.100 (through zero) to 0.422. Because the 95% confidence interval includes zero we say that there is no evidence to reject the hypothesis that the distribution is not skewed. Or more simply, the distribution is not skewed.
Further suppose that the kurtosis statistic for the correlation question was .705 and that the standard error of kurtosis was .759. Is the distribution leptokurtic? The 95% confidence interval is found as follows:
95% C. I. = kurtosis statistic ± 1.96 * (standard error of kurtosis)
= .705 ± 1.96 * .759
= .705 ± 1.488
= (.705 - 1.488) to (.705 + 1.448)
= -0.783 to 2.193

A graphic representation of the 95% confidence interval for this kurtosis value is shown in Figure 2.

Figure 2. 95% Confidence Interval for the Kurtosis Value



The 95% confidence interval ranges from -0.783 (through zero) to 2.193. Because the 95% confidence interval includes zero we say that there is no evidence to reject the hypothesis that the distribution has no kurtosis. Or more simply, the distribution has no kurtosis. Because there no kurtosis and no skewness the correlation scores are said to be normally distributed.

The concept of a confidence interval is basic to understanding statistics. Confidence intervals are a standard part of the output of many SPSS procedures. Press this button if you would like a mini refresher course on confidence intervals

CONFIDENCE INTERVAL

Percentile values. The percentile values option will print the values at a given percentage. If you select quartiles the scores at the 25th percentile, the 50th percentile (the median), and the 75th percentile will be given. You can choose to find the cut points that divide the scores into n equal groups. For example if you choose 5 equal groups, then the following scores will be given: 20th percentile, 40th percentile, 60th percentile, and the 80th percentile. The percentile(s) option allows you to select any given percentile score.

The percentile values options are useful if you want to define groups of participants. For example, if you wish to divide your participants into the top, middle, and bottom third on the basis of their IQ scores you could find the cut points by selecting the option to find the cut points that divides the group into three equal groups.

5. Chart Options

The chart options include bar charts, pie charts, histograms, and histograms with a normal curve. The values for the charts can be expressed as either frequencies or percentages.

Bar charts and pie charts are commonly used when you have categorical data such as gender or race. Any value that is empty (no one selected that value) is not included in these charts.

Histograms are commonly used when you have interval data such as age or IQ scores. Histograms show the all the values from the lowest to the highest scores. Empty values between the lowest and highest scores are not excluded for histograms.

You can ask to display a normal distribution curve on top of the histogram chart. The normal distribution displayed is the what the histogram should look like if the data were normally distribution. You can visually compare the histogram data to the superimposed normal curve to get a visual sense about whether or not your data is normally distributed. If it looks like there is a problem with the data you could then compute the 95% confidence intervals for skewness and kurtosis to see if the shape is statistically different from a normal distribution.

6. Format Options

Order by. The format options refer to how the frequencies tables are formatted. The values for a variable are normally listed ascending order of the values themselves. For example, if you

displayed a frequencies table for the ID variable, the values would begin with ID = 1 and end with the highest ID value. You can also have the values listed in descending order. In the ID example the values would begin with the highest ID value and end with ID = 1.

The other two options allow you to order the output according to the counts for each value. Ordering by counts is very useful in some contexts. Suppose that you are doing a survey for one of the fast food chains in the area. One of the question is "When you think of fast food restaurants, which one comes to your mind first?" You then get a bunch of answers including Wendy's, Burger King, McDonalds, etc. It would be very helpful to have the FREQUENCIES for this question ordered by the descending counts.

Multiple Variables. The **Compare variables** option will display a statistics table that includes all the selected variables. Separate frequencies tables will be displayed for each variable even though the Compare variables option has been selected. The **Organize output by variables** option will display the statistics table and then the frequencies table for each of the selected variables.

Display frequency tables. This format option allows you to suppress the printing of frequencies tables based on the number of categories. This is typically used to suppress the printing for variables with a large number of categories. For example you may not wish to print out frequencies tables for the age variable, preferring instead to display a histogram of those values.

7. Saving the syntax commands

You can save the syntax commands that SPSS creates by clicking the **Paste** button on the right side of the Frequencies dialog box. The syntax editor window will open displaying all the commands for the current settings. You can save the commands in the normal way from the syntax editor window. Saving the commands will allow you to easily run the same set of commands again without having to select all the options from the various dialog boxes.

8. Save or Print the Output

The output displayed in the Output Navigator window can be printed or saved to a disk.

Print the output. Select the output to be printed by highlighting elements in the outline pane. The highlighted elements with the **open book** icon will be saved. You can open or close the book on an individual element by double clicking on the book icon at the front of each element.

You can preview how the document will look when printed by selecting

File
Print Preview

Save the output as an *.spo file. If you press the disk icon or

File
Save

then the highlighted elements in the outline pane will be saved. As with the print option, only those highlighted elements with the open book icon will be saved. The default file extension is *.spo (SPss Output). The *.spo file can be opened only in the SPSS Viewer window. Use the

File
Open

sequence in SPSS to open the *.spo file. You cannot open an *.spo file in your favorite word processor.

The **Save As** option gives you the option of saving the file as a Navigator file (*.spo) or as "all files (*.*). However, saving the output navigator file with the *.doc extension still will not allow the file to be opened successfully with MS Word or with MS WordPad.

Export the output as an .html or .txt file. Another possibility is to **Export** one or more objects. The export option saves objects in either .htm or .txt format. The .htm format can be used in web pages or read by MS Word 95. The .txt format can be read by any ASCII editor. You could insert either an .htm or .txt file into MS Word 97 but I do not recommend inserting a .txt file into Word 97. Table boundaries in .txt files are identified by dashes (----) and bars (|) rather than by solid lines and the .txt tables are cumbersome to edit. Use the

File
Export
option or right click on an object and select **Export** to open the **Export Output** dialog box. The default file name is OUTPUT, you will probably want to change the name to reflect the content that is being exported. Select the file type (*.htm or *.txt) as desired. When you have completed selecting the desired options press OK to export the file.

9. Save the Output in an MS Word File

You can save an SPSS Viewer object directly into a word processor such as MS Word by (a) right clicking the object, (b) press the **copy** option, (c) open an MS Word file and then **paste** the object into the open file. The object is created as a normal table in Word. You can edit the elements of the table within Word. (Note: this option does not work with objects created in SPSS 7.5.)

It is also possible to use the **copy objects** command rather than the copy command to copy and paste an object from the SPSS Viewer to a Word file. The object is created as an object in Word. Although it is possible to edit the contents of an object in WS Word, I have found the process to be tedious and unreliable. It is easier to do all your editing in the output navigator prior to copying the objects to MS Word.

In summary, I recommend using copy and paste to move objects from the SPSS Viewer to an MS Word file.