USER GUIDE

Tactical Analysis using Accurint® Crime Analysis Workstation
# INDEX

**EXPORTING AND SAVING DATA** ................................................................. 24

**EXPORTING DATA** .................................................................................. 26
- Lesson 4: Converting a Microsoft Excel (.xls) Spreadsheet
to Microsoft Access (.mdb) ........................................................................ 24
- Lesson 5: Save a sub-set of your Data in an Excel (.xls) Format .......... 26
- Creating a New Workstation Database .................................................. 27
- Lesson 6: Create a new Workstation Database .................................... 27

**SAVING A WORKSTATION DATABASE** ............................................... 27
- Lesson 7: Save/Append Data into an Workstation Database ............... 28

**PREPARING DATA FOR ANALYSIS** ....................................................... 29
- Lesson 8: Preparing a Dataset for thorough Analysis ......................... 29
- Time Coding .......................................................................................... 31
- Sequencing ........................................................................................... 31
- Duration ................................................................................................. 31
- Interval .................................................................................................. 31
- Temporal Coordinates .......................................................................... 32
- Midpoint (Date/Time) ........................................................................... 32

**QUICK STAT** ......................................................................................... 32
- Lesson 9: Calculate Statistics – Frequency From Summary Statistics ..... 32
- The “IZE” Methodology ....................................................................... 33
- Categorize ............................................................................................. 33
- Generalize ............................................................................................. 33
- Organize ................................................................................................. 34
- Minimize ............................................................................................... 35
- Maximize ............................................................................................... 36

**LAYOUTS** ............................................................................................... 38
- ORGANIZE DATA USING LAYOUTS .................................................. 38
- Lesson 10: Create a New Layout ............................................................ 38
- Lesson 11: Save a Layout ...................................................................... 40
- Lesson 12: Organize Data by Sorting ..................................................... 43

**VARIABLE GROUPING** ........................................................................... 45
- Lesson 13: Organize Data by Grouping ................................................. 46

**DATA MINING (MINIMIZE)** ................................................................. 48

**FILTER BAR** .......................................................................................... 49
- Lesson 14: Minimize using the Filter Bar ............................................. 50
Lesson 15: Filter bar query to select by crime and day-of-week .......................... 52

VISUAL QUERIES (SQL STATEMENTS) .................................................................... 55
  SQL Query Types .................................................................................................. 55
Lesson 16: Query using the Visual Query Module – ‘all’ .................................... 57
Lesson 17: Query using the Visual Query Module – ‘all’ and ‘any’ ....................... 59
Lesson 18: Query using the Visual Query Module – ‘all’ and ‘any’ ....................... 61
Lesson 19: Saving/Loading Visual Queries .......................................................... 62
Exercises: ................................................................................................................. 64

THE QUERY WIZARD ............................................................................................... 65
  Lesson 20: Minimizing via the Query Wizard ...................................................... 65

DATA MINING – REGULAR EXPRESSIONS .............................................................. 68

KWIC EXPRESSIONS ................................................................................................. 69
  SIMPLE EXPRESSIONS .......................................................................................... 69
    Lesson 21: Query for Unique Terms using KWIC Expressions ....................... 69
  “OR” EXPRESSIONS ............................................................................................... 71
    Lesson 22: Query using the Either Operator ...................................................... 71

ORDER SPECIFIC EXPRESSIONS .......................................................................... 72
  Lesson 23: Identifying Records in a Specific Order .............................................. 73

COMBINED EXPRESSIONS ...................................................................................... 74
  Lesson 24: Identifying Records Regardless of Order .......................................... 74

PROXIMITY EXPRESSIONS ...................................................................................... 75
  Lesson 25: Query for Unique Terms Near One Another .................................... 75

CONCEPTS ............................................................................................................... 78
  CONCEPTS ............................................................................................................. 78
  FIND CONCEPT ...................................................................................................... 79
  REPLACE CONCEPT ............................................................................................... 79
  UPDATE CONCEPT ................................................................................................. 79
    Lesson 26: REPLACE: Analyzing data gathered from pirates around the world. .... 80

UPDATE CONCEPT ................................................................................................. 81
  Lesson 27: Creating a new field to house values from existing fields ............... 81
  Lesson 28: Identifying gang related assaults that occurred within Washington D.C. in 2008 ........................................................... 84

TREND HUNTER ...................................................................................................... 86
  MAXIMIZE USING THE TREND HUNTER ......................................................... 86
  Lesson 29: Create a New Scan ............................................................................ 86
TIME SERIES ANALYSIS .................................................................................................................. 89

THE TIME SERIES ANALYSIS MODULE .......................................................................................... 89
  Descriptive Analysis ......................................................................................................................... 89
  Temporal Topology ............................................................................................................................ 89
  Lesson 30: Create a Temporal Topology ......................................................................................... 89
  Predictive Analysis ........................................................................................................................... 92
  Lesson 31: Calculate Time Series Analysis ..................................................................................... 92
  Lesson 32: Next Event Prediction .................................................................................................... 96
  Calendar ........................................................................................................................................... 96
  Lesson 33: Applying Compstat Analysis to your data ................................................................. 97

SPATIAL ANALYSIS .......................................................................................................................... 98

  Workstation - Google Earth ............................................................................................................. 98
  Mapping Points ................................................................................................................................. 98
  Lesson 34: Placing points on a map ............................................................................................... 99
  Point to Point .................................................................................................................................... 99
  Lesson 35: Use the Point to Point function to identify auto thefts .............................................. 99

ALTITUDE ............................................................................................................................................ 100

  Lesson 36: Generating an altitude map within Google Earth ....................................................... 100
  Color Coding .................................................................................................................................. 101
  Lesson 37: Color coding records within the matrix ..................................................................... 101
  Color Coding with Google Earth .................................................................................................... 102
  Lesson 38: Pushing out our color coded records to Google Earth ............................................. 103
  Minimum Convex Polygon ............................................................................................................. 103
  Lesson 39: Using a minimum convex polygon to show the overall area related to a bank robbery series .................................................................................................................. 103
  Time Slider ..................................................................................................................................... 104
  Geographic Profiling ....................................................................................................................... 104
  Newton Swoope Buffer .................................................................................................................... 104

HOT SPOTS ....................................................................................................................................... 105

  Lesson 40: Generating a hot spot map to help demonstrate the pattern of arson's occurring throughout the D.C. area ........................................................................................................... 105
  Lesson 41: Hot spot map to highlight patterns related to residential burglaries .......................... 106

CHOROPLETH MAPS .......................................................................................................................... 107

  Lesson 42: Generating a Choropleth in Google Earth from Workstation ................................. 107
<table>
<thead>
<tr>
<th>Lesson 43: Generating a Choropleth using Non-Summarized Data</th>
<th>108</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Earth Exercise</td>
<td>109</td>
</tr>
<tr>
<td>Workstation Map</td>
<td>110</td>
</tr>
<tr>
<td>Lesson 44: Adding the data to Workstation Map</td>
<td>111</td>
</tr>
<tr>
<td>EXERCISE 1</td>
<td>113</td>
</tr>
<tr>
<td>CALCULATING STATISTICS</td>
<td>115</td>
</tr>
<tr>
<td>STATISTICS</td>
<td>115</td>
</tr>
<tr>
<td>Frequency</td>
<td>115</td>
</tr>
<tr>
<td>Lesson 45: Calculate Statistics – Frequency From Statistics Module</td>
<td>116</td>
</tr>
<tr>
<td>Chi Square</td>
<td>117</td>
</tr>
<tr>
<td>Lesson 46: Use Chi-Square to Analyze a Survey</td>
<td>118</td>
</tr>
<tr>
<td>Crosstab</td>
<td>120</td>
</tr>
<tr>
<td>Lesson 47: Use a Crosstab to Tabulate Total Officers by State</td>
<td>120</td>
</tr>
<tr>
<td>Regression &amp; Correlation</td>
<td>121</td>
</tr>
<tr>
<td>Lesson 48: Using Regression to predict Officer Volume based on Population</td>
<td>121</td>
</tr>
<tr>
<td>Goodness of Fit</td>
<td>123</td>
</tr>
<tr>
<td>Lesson 49: Use Goodness of Fit to Assess Workload</td>
<td>124</td>
</tr>
<tr>
<td>Threshold Analysis</td>
<td>125</td>
</tr>
<tr>
<td>Lesson 50: Use Threshold Analysis to determine volume of incidents for total timeframe</td>
<td>125</td>
</tr>
<tr>
<td>REPORTING</td>
<td>128</td>
</tr>
<tr>
<td>THE REPORTING MODULE</td>
<td>128</td>
</tr>
<tr>
<td>Lesson 51: Create a Report</td>
<td>128</td>
</tr>
<tr>
<td>TRENDS</td>
<td>131</td>
</tr>
<tr>
<td>Lesson 52: Trend Management</td>
<td>132</td>
</tr>
<tr>
<td>DESCRIBING TRENDS</td>
<td>132</td>
</tr>
<tr>
<td>Lesson 53: Create a New Trend</td>
<td>135</td>
</tr>
<tr>
<td>ASSIGNING CASES TO A TREND</td>
<td>136</td>
</tr>
<tr>
<td>Lesson 54: Assign Cases to a New Trend</td>
<td>136</td>
</tr>
<tr>
<td>TREND REPORTING</td>
<td>137</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>139</td>
</tr>
</tbody>
</table>
This workbook was designed to provide “hands on” instruction to pattern analysts and law enforcement officers. The authors of this workbook have designed lessons to follow a “typical” pattern analysis workflow using the Workstation software. Proceeding through the workbook, users will learn how to enter, organize and examine data (statistically, temporally, spatially and via chosen characteristics) and then look forward to predictions and forecasting. Finally, the workbook will highlight how Workstation provides a number of options to present analytic conclusions.

While Workstation is primarily designed to enable analysts to analyze and visualize their data statistically, geospatially and temporally, it has additional functionality to support metrics and planning. The workbook will also highlight and demonstrate how to leverage this additional functionality.

Workstation users are strongly encouraged to take advantage of the resources available in the Help Manual or the Help File within workstation. Workstation users are also encouraged to frequent the LexisNexis home page at www.accurint.com/workstation.

We hope the workbook serves as a valuable resource. We have endeavored to make the lessons comprehensive, while succinctly delivering the information users need to become efficient and successful in Pattern Analysis.

If you have comments/suggestions regarding the content of the workbook, we would welcome your input. Suggestions and comments should be sent to publicsafety.support@lexisnexisrisk.com.
When installing Workstation, users must decide how to best deploy Workstation. Workstation can operate in either stand-alone or networked environments. Under both deployment scenarios, Workstation’s core system files must be installed to all client machines.

**Installation**

**Stand-Alone**

In a stand-alone environment, Workstation is installed to the client’s machine. If Workstation is being loaded onto a Windows’ system, the .ini file will automatically install all the files required to run Workstation. All layouts and application files are stored on the client’s computer.

**Network**

In a networked environment, Workstation is typically installed to a network drive. All client computers point to the network and share one executable file (WORKSTATION.EXE). When using Workstation in a network environment, Workstation users can share layouts, the Tables database and the Admin database. When the client launches the Workstation executable from the network, Workstation looks to the directory where it is installed (the network drive) and uses all of the supporting administrative files. If multiple users are going to be accessing Workstation, a networked setup is the most advantageous because it allows for standardization of the tables used (variable values stored), increased security, and is easier to update. Under this scenario, only one update is necessary to update all clients using Workstation.

When installing Workstation to a network, users should utilize a client computer from which they will also want to install/use Workstation. The user should follow the setup’s “typical” installation option but change the default installation directory to the networked drive. That way, not only will the Workstation components be installed to the network drive, but all necessary Windows’ system files will be installed to the client computer. Subsequent client installations will only require the system files and data access components. It is recommended to follow the “custom” setup during installation and de-select the “Workstation Files” option when installing to a new client machine. This will ensure all the necessary drivers and DLL files are loaded without unnecessarily installing Workstation to the network drive or client’s hard drive. Doing so will avoid the possibility a user on the client computer will inadvertently use the Workstation installed to the hard drive thus negating the ability to share the tables, layouts, and adhere to the administrative privileges.

**Accounts**

There are two types of accounts in Workstation; Managers and Users. Those who have “Manager” accounts are allowed access to set a User’s preferences. Users with “Manager Rights” can tailor the use of Workstation to an individual
**Manager Controls**

- View individual User activity within Workstation
- Set individual User Account Preferences based on the following:

<table>
<thead>
<tr>
<th>Access to the Entry Module</th>
<th>Convert Heights</th>
<th>Export Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add in “View All About”</td>
<td>Create New Trends</td>
<td>Modify in “View All About”</td>
</tr>
<tr>
<td>Assign Cases to Trends</td>
<td>Delete Cases from Trend</td>
<td>Modify Tables Database</td>
</tr>
<tr>
<td>Change Default Databases</td>
<td>Delete in “View All About”</td>
<td>Modify Trend Query Values</td>
</tr>
<tr>
<td>Change IR Number</td>
<td>Delete Incident</td>
<td>Revert or Update Ages</td>
</tr>
<tr>
<td>Change Logo or Badge</td>
<td>Delete Multiple Records</td>
<td>Update Soundex</td>
</tr>
<tr>
<td>Change Password</td>
<td>Delete Trends</td>
<td></td>
</tr>
</tbody>
</table>

**User Controls**

Users can adjust a number of preferences that affect their day-to-day usage of Workstation. Set individual user preferences by selecting “Preferences” from the hub’s “Edit” menu. Preferences can be set based on the following:

<table>
<thead>
<tr>
<th>General</th>
<th>SQL Standards</th>
<th>Grid Formatting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Auto Mode</td>
<td>Display Relationally</td>
<td>Grid Colors</td>
</tr>
<tr>
<td>Utilize Sounds</td>
<td>Display Primary</td>
<td>Marquee Style</td>
</tr>
<tr>
<td>Highlight All Cases</td>
<td>Default Query</td>
<td>Heading Height</td>
</tr>
<tr>
<td>Related to each record</td>
<td>Default Sort</td>
<td>Font</td>
</tr>
<tr>
<td>Show Frequency Results</td>
<td>Temporal Preferences</td>
<td>Grid 3D</td>
</tr>
<tr>
<td>in Filter Bar Drop List</td>
<td>Earliest Permissible Date</td>
<td>Alternate Rows</td>
</tr>
<tr>
<td>Show Tips at Startup</td>
<td>Latest Permissible Date</td>
<td>Autosize Columns to</td>
</tr>
<tr>
<td>Perform Weekly Backup</td>
<td>Max Allowed Gap</td>
<td>Display Values</td>
</tr>
<tr>
<td>Anniversaries</td>
<td>Auto Mode Wait</td>
<td>Extend Grid's Right</td>
</tr>
<tr>
<td>Show Filter Bar</td>
<td>Midpoint</td>
<td>Column</td>
</tr>
<tr>
<td>Show Group By Bar</td>
<td>Exact</td>
<td>Wrap Text in All Columns</td>
</tr>
<tr>
<td>Confirm on Exiting</td>
<td>Voice Options</td>
<td>Show Row Selector</td>
</tr>
<tr>
<td>Workstation</td>
<td>Voice Speed</td>
<td>Assistant Preferences</td>
</tr>
<tr>
<td>Allow Workstation to</td>
<td>Variable Read</td>
<td>Available Assistants</td>
</tr>
<tr>
<td>communicate with Internet</td>
<td></td>
<td>Imagery</td>
</tr>
<tr>
<td>Popup Window Opacity</td>
<td></td>
<td>Default Image</td>
</tr>
</tbody>
</table>
Workstation provides several ways to learn the functions of the program (one of which is this workbook). The primary method is the Help File, but additional resources are in place to ensure users get the most out of Workstation.

**Help File**
Workstation is supported by an extensive help file. Press “fn-F1” within Workstation for searchable explanations, definitions, and lessons for Workstation’s features.

**Tool Tips**
Tool tips are gray bars that appear when the mouse pointer is located over an item.

**Status Bars**
Workstation displays information on the status of computations, database operations and analysis results through various status bars. Status bars are the embossed boxes that encompass the length of the bottom of a form. They are typically blank until an operation or function is being performed.
Workstation Databases

Workstation databases are designated by the “.atx” extension. Workstation databases are structured around a Microsoft Access 2000 database. However, LexisNexis has made modifications to the underlying structure, so care should be taken when opening or modifying a Workstation database in anything other than Workstation.

Workstation includes two administrative databases called TABLES.MDB and ADMIN.MDB. While minor system maintenance changes can be achieved by users via Microsoft Access 2000, this should rarely be necessary. Minor maintenance might include: compacting, repairing and changing value(s) in variable(s). Here are some guidelines that should be considered before modifying the Workstation tables.

- Never change the name of a table or variable.
- New values lengths should not exceed the length of the variable.
- Avoid using special characters in the value as they may interfere with SQL reserved syntax (e.g. % and * and _).  

There is a file named ATACCOPY.000 located in the directory installed with Workstation. This file is used when users save or export records. Workstation will copy this file to the new file users designate and append into it the records users selected. ATACCOPY.000 should never be relocated, modified, deleted or used to store data.

Tables Database

TABLES.MDB stores the values for every “drop list” variable in Workstation. Each table contained in the TABLES.MDB database contains a list of the possible values for that variable. For instance, the table “Crimes” contains a list of the available crime types in Workstation. Users can move the Tables.mdb database to a network drive to share a common set of drop list values. Designate the new location of the Tables.mdb database in the Accounts Manager.

Admin database

ADMIN.MDB stores each user’s information and Trend Scans. In a multi-user or network environment, the ADMIN.MDB file should be located on a shared drive. The ADMIN.MDB database is structured on a Microsoft Access 2000 database. When a user logs into Workstation, they are being authenticated through the ADMIN.MDB database.

Trends Database

Once users begin using and identifying trends in Workstation, best practice is to build a tailored Trends database, be it for an individual user, a department or for periods (last year’s trends may be kept for reference purposes) of time. The
currently used trends database will be displayed on the Trend Manager’s status bar at the bottom of the form. The default location for the Trend information is the Sample Trends.mdb database (located in the directory where Workstation is installed). The table names are tblTrends (stores related Trend information), tblCases (stores information about each case in a trend) and tblSolutions (which stores information related to the trend’s offender). These tables should not be modified nor deleted.

**Concepts Database**

Any Concepts are stored in the Workstation root directory. Users create and merge Concept databases via the Concept toolbar. Users can specify a default Concept database in the Accounts Manager. Storing the Concepts locally enables agencies and users to share Concepts.

**Anniversaries Database**

The Anniversaries database enables users to keep track of significant events in history or in the future. Store such important information as religious holidays, terrorist activities, significant events etc. Anniversaries created are stored in the Anniversaries.mdb database in the root directory where Workstation is installed. Users open, create and merge anniversary databases via the Anniversaries module toolbar. Different users can specify a default anniversaries database in the Accounts Manager.

**Layouts**

Layouts are templates that allow users to adapt the placement of variables within the matrix, based on the needs of the task. The layouts for Workstation are located in the \LAYOUTS\ directory where Workstation is installed. Each distinct layout is given a .lyt extension. Users can share layouts by placing the layout file in the layout directory.

**Tables Structure**

Workstation is a relational database made up of three unique tables; MO, PERSON and VEHICLE. Through two common variables (the IR Number and STAMP variable), Workstation is able to display the data as a continuous record.

**How Workstation Displays Data**

Workstation displays data in a way that will not exclude any potential combination of the crime’s framework. Workstation’s relational view allows users to view multiple records associated with a single Incident Number. While there is a need to view all records associated with a single crime event, when performing statistical analysis, duplication of events can inflict an undesired bias. To overcome this,
Workstation has the primary view, which presents a single record containing just the primary MO, person and vehicle.

**IR Number**
Those variables common to all tables are the IR Number and the table specific “STAMP” field (preceded by its appropriate table name e.g. MOSTAMP). The IR Number represents the crime as a whole, all its unique MOs, potential persons and possible vehicles. There can be more than one occurrence of an IR Number in any one table. This will occur when a crime has more than one MO (Modus Operandi), vehicle, or person. In contrast, the STAMP variable contains a unique identifier for every individual table’s record. The STAMP variable is a combination of the date, time and the user’s ID number.
Non-Workstation data refers to any data type that Workstation can ingest, but is not in .atx format. Not only can Workstation ingest most all data types, but it allows users to export into different formats. For data types such as Excel, which only allows editing in Excel, this allows users to export to a format that can then be used to enable data for attribute, spatial and temporal analysis purposes.

More common data types that Workstation can ingest/export to include:

- Microsoft Access
- Microsoft Excel
- dBASE
- Text

Users can manually upload data into Workstation, which is compatible with both SQL Server and ODBC.

When opening non-Workstation databases, the import Wizard guides users through the process to ensure maximum functionality of the data. There are two critical fields when ingesting non-Workstation data, the Table or Query field and Case Number field. To load non-Workstation data sources, both of these fields must be populated.

By default, only the first 1,000 records will be returned to the Matrix when opening data. This is to improve performance and enable users to begin querying and sorting the data faster. To capture the entire dataset, the user must refresh the matrix via the retrieve all records button.

There are a small number of functions and features that may not be available for non-Workstation data types. The primary ones are:

- In non-writable databases, such as Excel, users will not be able to highlight records or add fields.
- The Query Wizard only works with .atx databases.
- The ability to switch from Primary to Relational view

Another possible limitation of not converting your data to .atx relates to the Time Series related options. Time Series analysis requires the date/time related fields be in a set format. Users can reformat their date/time data fields and LexisNexis provides a service to reformat data (for further information about this service please contact our support team via publicsafety.support@lexisnexisrisk.com).

# Primary/Relational view in Workstation allows users to view all records that have a common case number.
INGESTING DATA

Workstation provides a variety of ways to ingest data. Data input options include voice recognition and dictation, text input by keyboard or value lists, address verification, and canned phrases. By “capturing data,” we are referring to the creation of new data records directly in Workstation – not the use of existing data from other sources (that is covered in the subsequent section, “Opening Data”).

Adding and Validating Addresses
Workstation allows users to automatically add new addresses into an address table for later use when entering data. The “Add New Addresses to Quick List” option in the Entry Module will automatically add new addresses to the “Addresses” table in the TABLES.MDB database.

AutoSearch for Existing Addresses
The “AutoSearch for Existing Addresses” feature will search through the “Addresses” table in the TABLES.MDB database for existing addresses beginning with the characters the user enters as an address. As the user continues to enter more characters, the Addresses matching will update. Once a match has been identified, users can simply click on the desired address in the table to enter its information. Not only will the address be entered, but the address name, RD, Beat, Location Type, X Coordinate and Y Coordinate will also be entered. Using this feature will greatly improve the validity of addresses being entered as well as allow users to standardize the information accompanying the address.
The image below describes the main elements of the Workstation Window.
For the purposes of this workbook, we will use the following terms:

- **Records**: Referred to in some applications as rows
- **Fields**: Referred to in some applications as columns or variables
- **Cells**: Cells refer to the individual cells within a table
In most cases, analysts start by having some data. The format and amount of data will vary enormously, but most have a start point. The data may, or may not be in a format conducive to analysis beyond just attributes and basic statistical calculations. Analysts now all understand the need to view their data in a variety of ways to ensure thorough understanding of the data. The major ones are:

• Attribute
• Temporally
• Spatially
• Statistical

One of the first challenges for most analysts is to enable all of the above. For example, if the data is in Excel format, it can only be edited within Excel. If the data has an address, but doesn’t have XY’s, or those XY’s aren’t in the correct format, then how does the user enable that? Likewise, if you have the dates and times, but are simply eyeballing them in a spreadsheet and not able to view them in a more visually effective way (sequencing for example) it is not a question of if you might miss an important opportunity, but rather when you will miss it.

Workstation enables all of the above. Workstation ingests most data formats, provides the mechanisms to export to other formats as required and supplies the necessary solutions and tools to enable all four of the analytic methods listed above.

Most any type of data can be opened in Workstation and the majority of the matrix manipulation and organization features are still available for trend non-Workstation databases.

While most of Workstation’s functions are built to work with any type of data, certain features require a specific format. Those functions will be discussed more fully later in the workbook, but some modules are built for use with Workstation databases:

Modules
• Entry Module
• View Current Incident
• Relational/Primary view

In the following lessons, we will open a variety of data types beginning with Microsoft Access (.mdb).

Lesson 1: Opening Data

For this lesson: Examples.mdb | Table: CFS
1. To open a non-Workstation database, select **Open Data** from the **File** menu.

2. Change the **Files of Type** to Microsoft Access 2000.

3. Select “**Examples.mdb**” from the directory installed Workstation.

4. Click the **Open** button. The Database Wizard will display and select the first table in the database.

5. Select **CFS** from the ‘Table or Query’ drop list. The fields contained in the CFS table will display in the data grid. Use the **data grid** to verify key fields and the values contained therein.

![Data Grid Image]

Note: There are two mandatory fields for ingesting data – “**Table or Query**” which table holds the data required and “**Case Number**” – unique identifier within the table.

6. Click the **Automatically Search for Required Fields** checkbox to automatically search and populate the required fields. Workstation will recognize most field values commonly associated with those required. If key fields are not identified, users have the option to manually select a field from the drop down box. Ensuring all possible fields are recognized by Workstation ensures users are able to leverage the maximum amount of functionality.

7. Specify the “**INITIAL_TYPE**” field as the main variable, the “**RECEIVED_TIME**” field as first date as well as the first time.
8. Click “Save Field Selections” to automatically save the current field configuration, which will then be applied the next time the database is opened.

9. Select the Open button to open the database. If the “Save Field Selections” button was checked, the next time the user opens this database all the required fields will automatically be populated.
Lesson 2: Opening a Microsoft Excel (.xls) Spreadsheet

For this lesson: excelsample.xls

Microsoft Excel is the most popular spreadsheet program in the world. Many public safety agencies use Excel to store tabular data, perform statistics, calculate time series, and even perform Geographic Profiling. While Excel is a very common format, it restricts users from editing Excel externally which in turn restricts some of the Workstation functionality. To overcome this, the easiest solution is to export the excel file to a format that can be manipulated, such as Microsoft Access (.mdb) prior to loading in to Workstation, or the user can export the file to .mdb from Workstation and then reload (both options are illustrated – see lessons 5 and 6)

1. Select Open Data from the File menu.
2. Change the Files of Type to Microsoft Excel 97/2003.
3. Select excelsample.xls from the Workstation/data folder.
4. Click the Open button. Depending upon the structure of the Excel file, a sheet may not automatically populate the “Table or Query” combo box. In this case, there is only one worksheet associated with the file.
5. Select the “Automatically Search for Required Fields” checkbox to automatically search and populate Workstation's required fields. If key fields were not identified by Workstation, manually select a field by selecting it from the drop down combo box. While Workstation will attempt to identify the correct fields for the user, it is still the responsibility of the analyst to verify the fields are correct.
6. Select “Save Field Selections” to automatically save the current field configuration for the next time this database is opened in Workstation.
7. Select the Open button to open the spreadsheet. If the “Save Field Selections” button was checked the next time this spreadsheet is opened, all the required fields will automatically be populated.

Lesson 3: Opening an Workstation (.atx) Database

For this lesson: Sample.atx

In this lesson, we will open the Sample.atx database.

1. Select Open Data from the File menu or click the Open Database button located on the toolbar.
2. In the Open Data dialogue box, ensure the file type is “Workstation Databases (*.atx)”.
3. Select “Sample.atx” from the Workstation - Data folder.

4. Click the Open button. The database should open almost instantly and users will see records in the matrix window. The number of records loaded will be displayed in the status bar. The maximum number of records Workstation will load initially is 1,000. For databases with a record greater than 1,000, the user must select the “Retrieve all Records” button.

5. Once the data is loaded, use the scroll bar in the matrix window to scroll across.

6. To select an individual record manually, double click on the tab to the left of the record of interest.

7. Once selected the entire row is highlighted.

Exercise 1: Open Your Agency’s Data

Open your own data and answer the following questions:

What data format was your data in?

How many of the Database windows were filled in automatically?

How many additional fields were you able to add?

Did your data appear in the Matrix?

How many records are in your file?

Note: Users can limit the date range of your data by using the “Set Date Query Constraints” button on the toolbar. Using the “Set Date Query Constraints” feature will limit future queries to a certain time frame within your data. For crime events with no fixed event date (crime occurred during a week when the occupants of the premises were away, so users have a first and last date covering a 7-day period) users need to be wary. If users use ‘First’ and ‘Last Dates’ with the date query tool, it will look for events that fall on the date range the user stipulated in either of the fields, potentially skewing your data. To overcome this, users should change both values to reflect First Date. This will ensure discrete events are only appearing once in the stats.
User Notes:
Exporting Data

It is often necessary to export one set of data into another format in order for another application to be able to use it. For instance, users may want to export an Access database that is open in Workstation into an Excel spreadsheet to utilize Excel's extensive graphing and charting capabilities. Workstation provides the ability to export any data opened into another format. Alternatively, if a user wants to edit data that is in MS Excel format it has to be exported because Excel doesn't allow for editing via any other product than Excel.

Lesson 4: Converting a Microsoft Excel (.xls) Spreadsheet to Microsoft Access (.mdb)

With MS Excel being such a popular storage solution, the restriction in editing is problematic. There are several ways to overcome this, but all involve the user reformatting the data into one that can be edited. There are several ways to achieve this:

• Import the Excel worksheet into Workstation, export as a .mdb and then reopen in Workstation.
• While in Excel, highlight the records to be put into an Access database, copy and then paste into an empty Access database.
• Import the Excel spreadsheet into Access. Users can only import one worksheet at a time, but this is a very straightforward process and once done, users can save the wizard template to make it even easier.

Both MS Access and MS Excel provide clear instructions on how to import/export data between the two; the example below walks through how to complete the process using Workstation.

1. Open MO.xls as described in the previous lesson.
2. Ensure all the records are loaded (check record count and, if necessary, ‘Retrieve all records’).

3. File – Export Data
4. Navigate to the Student Workfolder
5. File name: ‘Accesstest’
6. Change the “Save as type” to ‘Microsoft Access (*.mdb)’ via the dropdown
7. Click on ‘Save’ or ‘Enter’
8. “Enter a table name” Sample
9. OK
10. Go to Edit and try to “Start Editing”. Note it is greyed out. This is the easy way to see the restriction placed on Workstation by using Excel. Not only does it affect manual editing, but other functions, such as sequencing are not available.


12. If necessary, import your data via the Wizard.

13. **Sanity check:** Go to Edit and look at the “Start Editing” option. It should no longer be greyed out.

**Exercise:** Open the following excel worksheets in Workstation and export them out as .mdb’s:

<table>
<thead>
<tr>
<th>Excel File</th>
<th>Save as $ in your Work Folder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maritime_Piracy</td>
<td>Pirate</td>
</tr>
<tr>
<td>CAD</td>
<td>CAD1</td>
</tr>
</tbody>
</table>

Once opened, review the new .mdb files by ensuring the edit function is available and provide a record count for each table.

Open CFS.dbf and create two new fields for X and Y coordinates.

**User Notes:**
Lesson 5: Save a sub-set of your Data in an Excel (.xls) Format

For this lesson: Examples.mdb | Table: Series_Mailbox Bomber

The Export to Excel function in Workstation provides for two options: exporting the current Database/Query to separate sheets for MO, persons and vehicle, or, exporting the data into one single sheet. Only Workstation databases will provide the option to export to separate sheets. If using a non-Workstation database, the default is to export the data as queried.

If users wish to take the contents of the Matrix into Excel to chart the data and make it more aesthetically appealing, we suggest copying the cells of interest by using Workstation’s copy cells function.

In this lesson, we will export all the records in the Examples.mdb database to Excel:
1. Open the Examples.mdb database and select the Series – Mailbox Bomber table.
2. Set the Field Selections to best match the data and then click Open.
3. From the File menu, choose Export Data and Navigate to the temp/Workstation/Work folder.
4. Select Microsoft Excel from the Files of Type drop list.
5. Type “Excel Example” into the File name text box.
6. Click the Save button.
7. Select “No” when asked “Would you like to Export the Matrix as Formatted?”
8. Enter “Mailbox Bomber” when prompted to enter a worksheet name.
9. Open the newly created excel spreadsheet and review.
Creating a New Workstation Database

Lesson 6: Create a new Workstation Database

It is unlikely users will need to create a new empty Workstation database, but Workstation enables users full flexibility to create and edit databases for all situations. For this lesson, we will create a new temporary Workstation database (.atx), which we will store in the C:\temp\Workstation\exercise.

Process

Step 1: From the File menu select Create Workstation Database.
Step 2: Navigate to the C:\temp\Workstation\exercise folder and enter “MyData.atx” in the File name text box.
Step 2: Click the Save button.

Notice that the name of the database is now on the main form’s title bar.

Also, the status bar at the bottom of the form displays that there are no records in this database.

Adding Incidents to the Database

Step 1: From the Edit menu select Entry Module
Step 2: Select the Add Incident button
Step 3: Once the desired information has been entered, select Save Incident

Saving a Workstation Database

Workstation can save a copy of the active database with a different name or in a different location. Use the ‘Save Database As’ feature to save queried and sorted data from existing databases. For instance, if a user selected a group of Robberies from the main database, they could then save them out to their own database for future analysis. Using the Save ‘Workstation Database As’ function will not delete
the records that have been saved to another database from the existing database.
If users select a database to “Save As” to which already exists, Workstation will append only those records from the database that don’t already exist in the target database.

Lesson 7: Save/Append Data into an Workstation Database

For this lesson: Sample.atx

1. Open the Sample.atx database from the C:\temp\Workstation\data folder.
2. From the File Menu, select Save Workstation Database As.
3. Navigate to C:\temp\Workstation\exercise and select the MyData.atx database.
4. Click the Save button. If the database already exists, Workstation will ask if the user wishes to append the current selection of records to the target database. Choose “Yes” to append the records to the existing database.
5. Open MyData.atx. Workstation has appended all the records from Sample.atx into MyData.atx.
Data comes in many formats and how users manipulate the data often depends upon the limitations of the particular product and/or data format. Workstation enables users to manipulate and interact with their data to facilitate maximum analytic functionality.

Lesson 8: Preparing a Dataset for thorough Analysis
For this lesson, we will be importing an Excel file into Workstation. From there, we will export the file into a format that can be edited (mdb.) and enhanced, in this case by adding X/Y fields and geocoding the addresses for Workstation and ArcGIS. Once formatted correctly, we will have a fully functional database where we will have the ability to alter it however we choose.

For this lesson: Beltway Sniper Series.xls

2. Locate the ‘Beltway Sniper Series’ file (making sure the file type is set to .xls) – Open it.
3. When prompted by the ‘Field Selection’ box, select ‘Open’.
5. Save file as Beltway Sniper Series, but this time save it as an .mdb file (Microsoft Access).
7. Locate the ‘Beltway Sniper Series’ file (this time, making sure the file type is set to .mdb) – Open it.
8. Again, when prompted by the ‘Field Selection’ box, Select ‘Open’.

We now have an editable database within Workstation, which allows users to customize the data. By formatting our data using Microsoft Access we have also prepared our data to be used by ArcGIS as well.

9. Select Edit – Create Fields.

The next step in developing a flexible database is to geocode the addresses so that we can map out our records.


12. Make sure that the rest of the fields are accurately filled in (i.e. City = City, X = X Coordinate, etc.).

13. Make sure that ‘Write X and Y Values into Database’ is checked.


You now have a fully functional database within Workstation. From here you are able to map your data via Workstation Map or Google Earth, as well as run any type of query imaginable.
**Time Coding**

Workstation provides a means to calculate the Sequence, Duration, Interval, and assign a Temporal Coordinate (T-Coordinate) to each record in the current database. This is useful for performing more detailed analysis of the data. It can also be used to perform spatio-temporal analysis given the data now contains integers that reflect the temporal extent as well as the spatial extent.

![Time Coding](image)

**Sequencing**

The Sequencing option enables the user to assign a numeric number beginning at 1 to each case in the current query based on the case's date and time relative to all other cases. The oldest case in the query will be given the number 1 whereas the newest or most recent case in the query will have the highest sequence number. Workstation assigns the sequencing number in the “Sequence” field based on the First Date, First Time, Last Date and Last Time.

Use this utility in conjunction with your GIS to determine sequencing and direction as well as distance intervals.

**Duration**

The Duration feature in the Time Coding module will allow the user to calculate for the duration of each incident in seconds, minutes, hours, days, weeks, months or years between each case's first date and time and last date and time.

Use this feature in conjunction with your GIS to conduct powerful 3D modeling of space-time continua. You may also want to use it in Workstation to perform queries based on the results (e.g. query all records where the incident duration is less than 1 hour) or to perform statistics against the results.

**Interval**

The Interval feature in the Time Coding module enables the user to calculate for each incident the interval between the first case and the next case in a series in seconds, minutes, hours, days, weeks, months or years.
Temporal Coordinates
The Temporal Coordinates (T-Coordinate) feature is primarily geared toward analyzing serial crimes; this function is extremely helpful for creating Lund Trajectories in your GIS or for studying the tempo and interval of a series.

Upon initialization, it will first order the crimes in a query according to which incident was first, second, third and so on. Based on this ordering, it then calculates the duration in the temporal units (which was previously specified by the user) from each case to the next in the series. The resulting number for each case represents the temporal units until the next case in the series.

The first case in the series will always have zero for its T-Coordinate value because the duration is calculated from the next case back. Therefore, the last case in the series will have a T Coordinate and it will reflect the duration from its time to the penultimate (second to last) case in the series. This method for calculating is important when converting Temporal Coordinates to spatial values in your GIS. (You will want to have your ‘base’ case number equal to zero.)

Midpoint (Date/Time)
The Midpoint feature enables the user to calculate the midpoint of the date and time of the incidents. This is used to find the likely average date and time within the dataset.

Quick Stat
Once your data is opened in Workstation, you can calculate some basic statistics by simply selecting the column of data you want and clicking the “Quick Stat” button from the Secondary Tool Bar. The results provided are determined based on the variable(s) selected by the user. For example, if you select a nominal variable (one containing text data) then the results only show the counts or frequency of the data. If the variable is of interval/ratio type (numbers), then a variety of summary statistics will be provided including: mean, mode, maximum, minimum, and standard deviation.

Lesson 9: Calculate Statistics – Frequency From Summary Statistics

For this lesson: DC Crime

1. Using the DC Crime database, click the column header for the UCR variable.
2. Click the **Perform Statistics on Highlighted Column** button from the toolbar. Review the output in the Summary Statistics window.

![Summary Statistics Window](image)

3. **Close** the Summary Statistics window and scroll to the **Property Value** field.

4. Select the **Property Value** column and click the **Perform Statistics on Highlighted Column** button. Note the different results based on a different variable type.

---

The “IZE” Methodology

Identifying crime trends is all a matter of organizing data so that similarities in records become apparent. If you were to try to identify a set of identical twins in a group of a thousand persons, it would be very difficult, if not impossible, to do so by simply scanning the crowd trying to uncover the match. However, by efficiently grouping the thousand people into categories by sex, race, age, etc. you would eventually and more efficiently identify the set of twins. They may in fact be standing next to each other after only a small number of categorizations.

How do you find a crime trend? It involves using your “ize” – Categorize, Generalize, Organize, Minimize and Maximize; so named because each step ends in “ize.” It follows the shape of an hour glass, narrowing results down until a pattern becomes apparent, and then broadening the search to encompass all of the cases in the pattern.
Categorize

Create variables conducive to finding crime trends (Hair, Race, Sex, Point of Entry, Weapon Type, Vehicle Make, etc.).

Categorize your data to create “variables” for the different types of data that you wish to capture. For instance, in looking for trends in data you would want to look at Race, Sex, Point of Entry, Weapon Type, etc. These are your categories. You most likely wouldn’t want “Date of Last Medical Exam” as a category about a person because it wouldn’t necessarily help you in matching crimes or people responsible for the same crime.

Workstation (.atx) has been designed to capture the main categories required for crime analysis as fields.

Generalize

Create general values for your categories (Handgun, Rifle, Male, Female, Brown, Black, Blonde, etc.).

Once you have identified a set of categories, generalize the values that you store in those variables. In other words, if you wanted to find commonalities in a “Property Taken” variable, it would not be a good idea to have every conceivable property type listed, as in the following example:

- Diamond ring
- Heart-shaped diamond band
- Diamond ring – marquee

Instead, you would want to “generalize” the values in your variables to describe a category, in this case, “Jewelry.”

Most often, agencies already have some form of generalization in place, based on their discrete needs/requirements. While the generalizations currently in place might serve the needs of the entire agency, it is possible, if not probable, that for the purposes of crime analysis they are not a good fit. As the Workstation analyst, you have control over how the information is generalized. Workstation comes with a comprehensive list of the basic categories used by most law enforcement agencies, but the list can be tweaked to ensure it best fits the needs of discrete agencies.

Other ways to generalize data might include: Crime (Part I/Part II), Location (Residential/Commercial).

For larger jurisdictions, visually depicting all Part I and all Part II crimes can be confusing to the customer. An alternative might be to depict all Part I crimes, but
use a heat map to summarize the Part II crimes. This would illustrate the general trend from a geospatial perspective, without distracting the customer from the Part I crimes. From the analytic standpoint, there are several reasons why analysts might want to generalize by crime type: Are there different hot spots for Part I crimes when compared to Part II? If yes then why? What is the ratio of Part I to Part II crimes for your jurisdiction and how does this reflect against the allocation of law enforcement resources?

For the location type, discriminating between house type (Single Family, duplex, town houses) should be part of the analytic, but dependent upon the question the analyst is trying to answer, it may be useful to generalize.

**Organize**

Group certain MO variables and person categories together; Organize data.

Organizing data is the primary means by which to identify patterns in your data. There are several ways to organize data. Most simply sort and group relevant values together (attribute based). This is an excellent means of identifying clusters in data. Once you determine which variable to sort, based on the type of trend you are attempting to identify, you simply sort by that variable looking for like values. Organizing by location is becoming more prevalent and involves identifying clusters by proximity, migration path etc. When used alongside sorting and grouping by variable this is an incredibly powerful way of organizing data. And lastly, organizing data, such that the most relevant information is grouped and prioritized ensures the follow on analysis is effectively directed/focused.

**Sorting and grouping relevant values**

If you need to analyze residential burglaries, looking at weapon type descriptors would prove less fruitful than looking at point of entry, method of entry, etc. Conversely, if analyzing the crime of robbery, you would focus more on the suspect's race, height, weight, age, eye color, hair color, weapon type, actions against persons, and so on. In order to analyze different types of crimes and the variables conducive to identifying patterns in those crimes, we need to organize variables (fields) to ensure the most significant are co-located.

**Sorting by location**

Location is a critical part of analysis. Everything happens somewhere and the location, proximity to bus stops, train stations, sporting venues, residential areas, schools, commercial areas, interstates, cul-de-sacs is a way of organizing events. Related crimes may all occur within a set distance of x, or may be moving in a North-South direction through a residential area. Professional shoplifting activities may be moving roughly in an E-W direction across a city following the interstate entrance/exits. Illegal border crossings will occur in predictable areas where the terrain lends itself to this type of activity.
Organizing the visual representation of data

As enrichment continues to burgeon, the data sets available to analysts are increasing in size exponentially. The growth in size of datasets is great news for analysis, but effective manipulation of data is now critical. Analysts can reduce the number of variables/fields captured, but excluding information that may later lend value is not ideal. A more efficient method is to organize the variables, allowing the analyst to tailor their data in response to the question. Workstation enables analysts to determine which variables are visible, how the variables are ordered within the table and basic grouping. Organizing data this way ensures critical variables are prominent, while unnecessary variables – not relevant to the task in hand, but that may be pertinent/critical to others – are not distracting.

Once organized, you begin to see patterns emerge in your data. You see particular data values or salient features in your data that you want to query on to further analyze a subset of data. This brings us to the fourth “ize” in finding crime trends...

Minimize

Query for clusters in your data by organizing it.

Minimize your data by reducing the data through queries. This is by far the most common and quickest way of focusing analytic endeavors. The process is straightforward – identify a variable that distinguishes data of interest from the rest. Once you have queried this variable, the dataset is reduced in size enabling better investigation. The process is repeated however many times is necessary to identify a trend. For example if looking for a trend related to residential burglaries:

Against the Crimes dataset – which holds information for all crimes – querying for ‘Residential Burglaries’ might reduce the Crimes dataset to 20% of the original size.

Against the Residential Burglaries subset, a query for ‘Dog Door’ as Point of Entry might reduce the Residential Burglaries subset by 90%.

So, a simple two-step query has reduced the original dataset by 98%. At this point, the effect of further queries will be less significant in regard to discarding large amounts of data, but with 2% of the original dataset now left to interrogate the dataset is now more manageable. From this point, minimizing data can continue as necessary using whatever variables the user deems relevant e.g. Time – residential burglaries where the point of entry was dog door and that took place Monday-Friday and between 0900 and 1600.

Workstation provides several ways for analysts to achieve minimization of their data, dependent upon the task. Analysts can interact with their queries (generate them and review results step-by-step) or can leverage tried and trusted queries created, either by themselves, or by other analysts.

While minimizing is an effective way of identifying trends, it isn’t ‘the’ solution. To assume a burglar will only target houses with dog doors is clearly flawed. The core
reason for minimizing is to provide focus; once that focus has been achieved and
the core variables have been determined, expanding to look for similar – but not
exact - events is absolutely vital for thorough and complete analytic. This leads us
to the next “ize”.

Maximize

Query features salient to the identified crime series.

The purpose of maximizing data after identifying the core variables is to identify
additional events that might relate to the trend found by minimizing. The
minimize process relies on implementing AND clauses – Residential Burglaries AND
Dog door AND between 0900 and 1600 AND that occur between Monday and Friday.
If the burglar entered a house where the homeowner didn’t return until 8pm
the following day and the crime was reported as taking place between 0600 (first
time and when the occupant left the house) and 2000 (last time and the time the
occupant returned the next day), a simple AND-based query with tight constraints
would fail to place this event in the trend.

Maximizing the dataset after identifying the core variables allows analysts to use
individual - or multiple – variables to identify additional events for assessment and
potential inclusion in the trend.

The next chapters of this workbook will focus on demonstrating how Workstation
enables users to effectively implement all the elements of the IZE methodology.
Organize Data using Layouts

Layouts are an integral part of Workstation and can be of enormous value to analysts. The matrix holds all variables available and this will almost always be greater than the visible area within the matrix. The purpose of layouts is to present variables within the matrix in such a way that lends itself to specific questions/crime types – effectively prioritizing the variables of greatest interest to the left of the matrix. Some variables, such as the date and incident number, are required to be easily visible for all crime types, but others will vary dependent upon the crime type/Workstation comes loaded with a number of layouts that can be used, but we recommend users a) review the variable used for each layout to ensure it adequately meets the unique needs of their department/agency and b) at the earliest convenience build their own layouts and tailor the matrix to meet the need of the individual analyst.

Lesson 10: Create a New Layout

For this lesson: DC Crime

In this lesson we will create a layout that will help users identify sexual assault patterns.


2. Select the Remove all Variables button from the Layout Organizer’s toolbar.
3. **Double-click** a variable from the “Available Variables” list to the left to move it to the Matrix Variable list. Users can also select multiple variables from the Available window and drag them over to the Matrix window. Variables will be displayed in the Workstation matrix in the order they are set in the Matrix variable window.

4. Select the following variables and add them to the Matrix Variable field: First Date, Address of Crime, Object of Attack_1, Point of Entry_1, Suspect Actions Against Persons_1 and Weapon Type_1.

   **Note:** If users need to reorder any variables in the Matrix Variable list, simply click the variable and drag it to a new location in the Matrix Variable list.

5. When all the variables are visible in the Matrix Variables window, select the **Apply** button from the toolbar.

   **Note:** The Refresh Database’s Original Variable Configuration features (bottom leftmost button) will restore the original order of variables as stored in the database. All layout configurations will be removed and hidden variables revealed.
Lesson 11: Save a Layout

For this lesson: DC Crime

This lesson will demonstrate how, having generated a layout, users can save the layout for future use. The image below shows the layout applied in the previous exercise using the Layout Organizer. The layout has already been applied.

1. From the Layouts hub, select the Save Layout button
2. The default location will be the root directory for Workstation/Layouts.

3. File name: Sexual Assaults

4. Save as type: **Layouts** (.lyt)

5. Click **Save**. Notice that your new layout is now listed in under the Layouts tab.
Exercise 2: Create layouts: 45 mins.
From DC Crime.mdb select the UCR field and create a layout suitable for Robberies. Review the variables (fields), determine the format, apply, review the results by applying the layout and then save the layout with an appropriate name.

Using the same dataset, create a layout suitable for Auto Theft. Review the variables (fields), determine the format, apply, review the results by applying the layout and then save the layout with an appropriate name.

**Question:** When applying a layout to the hub, remember the layout was configured against a specific data file type. **Why does this matter?**

**Your answer:**

**Summary:** Analysts are mostly human. By nature we look for the easiest way to do things. Having our data layout tailored to the needs of the task in hand helps maximize the precious commodity of time. Having the variables (fields) laid out in a logical format also helps users identify patterns and relationships that would be far more difficult to see otherwise. The Layout Organizer is a great enabler on a day-to-day basis, but it doesn’t come without negatives. If all the information analysts typically used for a particular question is kept to the left of the matrix, users are less likely to review other fields that rarely provide valuable insight. However, rarely does not equate to never. When users discard/hide variables all together it exacerbates the potential problem – even if users do scroll, they still won’t see the data.

**Exercise:** Open DC Crime
Query for Residential Burglaries in the UCR field, and then build a layout using the following variables:
I.R.Number, Location Type, Point of Entry (1), Method of Entry (1), Property Taken (1), Property Taken (2), Property Taken (3), Beat, RD, Property Value, First Date, First Time, Last Date, Last Time, Marked, Address of Crime.

Apply the layout and save it as ‘AAResidential’ to your working folder

Review the fields. **Are you** satisfied that all the necessary fields are there for review analysis purposes?
Lesson 12: Organize Data by Sorting

For this lesson: Examples.mdb | Table: Maritime Piracy

Layouts allow analysts to organize the data horizontally within the matrix (variable (field) order from left to right, visible or hidden). The next way to organize data is vertically. Users may wish to have the data in date order, by incident number, type of location, type of crime etc. To complete the next step in the IZE method – “Categorize”, users organize data vertically. Workstation has various ways to allow users to do this, but the most basic is via ‘Sorting’ the variables in ascending or descending order. This lesson will demonstrate how, using the sort option, users can sort their data via two discrete variables – ship type and ship name. The sort will sort in descending order by ship type and then by ship name.

1. Load the Maritime Piracy table from the Examples.mdb database.
2. Select the Sort tab located on the lower left of the Hub.

3. Select the Apply Descending Sort button located at the top of the sort tab.

By selecting ‘Apply Descending’ users are setting the sort direction to apply against the variable.
4. Double click on the Ship Type variable from the list of available fields. Notice that the variable was removed from the list and added to the “Sort Order” list, located in the window directly under the variables.

5. Select the **Apply Descending Sort** button, then double click on the **Ship Name** variable from the list to add it to the Sort Order List.

6. Select the **Execute Sort** button located on the Sort tab toolbar.

7. Go to the Matrix window and view the results of the sort.
   - What type of ‘ship’ is at the top of the matrix?
   - What’s the first ship name and are they in alphabetical order?
   - How many barges were attacked?

**Quick tip:** Left click on the Ship Type variable button and select the ‘Perform Statistics on Highlighted Column(s)’ button from the secondary tool bar.

Like layouts, sorts can be saved and applied as necessary. Manual sorting of variables is achieved by selecting a variable (field) and right click – sort ascending or descending. As soon as a user resorts the order based on a criteria the sort accomplished by using the Sort hub is lost (it doesn’t sort based on the new variable and then the preset variable).

1. Select (left click) the “Type” variable in the matrix.
2. Right click on “Type” and select sort ascending.
3. Review the matrix window.
   • What value is in the first Type record?
   • Are the Ship Type and Ship names still in descending/ascending order?

After reviewing the data, the decision is to look at the Ship Name in ascending order as the primary, rather than having the ship type as the primary.

How would you want to be able to resort? Try it out.

Select the Clear Sort button at the top of the Sort hub.

**Question:** How might having data sorted before starting the analysis process be useful for you?

**User Notes:**

**Variable Grouping**

Workstation provides the ability to group data by a particular column or columns to display how other variables relate. By simply grabbing and dragging a column to the grouping area, users can sort and display how a variable relates to all other variables. Variables placed in the grouping area are automatically sorted in ascending order. The “Show Group by Bar” preference must be set in order to display and allow grouping. The “Show Group by Bar” is located in the General tab in the Preferences dialogue box when you select Edit from the toolbar.
Lesson 13: Organize Data by Grouping

**For this lesson:** Examples.mdb | **Table:** Maritime Piracy

1. Left click on the ‘Region’ variable header to select the variable.
2. Left click and hold the left mouse button down on the Region variable header and Drag the variable to the area above the variable headers and release the mouse button.

3. Repeat this process with additional variables. Add the Ship Type and Ship Name variables to the grouping area. To change the order of several grouped variables, simply grab the variable header and move it to a new position in the grouping window. Red arrows will indicate its new location prior to releasing the mouse button. To remove grouping, simply drag the grouped variable back into the grid.

4. File – Export Data – C:\temp\Workstation \Work – File type (.html) – File name ‘Pirate’ – Save
5. Open an Explorer window and navigate to the C:\temp\Workstation\Work folder.
6. Open Pirate.html and review
7. Workstation: File – Open Data – Examples.mdb – Table Pirate2
8. Place Flag as your first grouping and Position as the second.

**Summary:** By segregating and manipulating the way the information is presented, we have identified and highlighted activity of interest to our customer [in this case UK/US-registered vessels], without actually performing any analysis. The information (formatted data) becomes knowledge simply by capturing and presenting what we know to be of interest to our customers (understand your data – understand your customer).
Sorting and querying are the primary means used by analysts to identify patterns in data. As the previous lessons demonstrated, sorting is critical in enabling analysts to identify patterns/trends and the follow on representation of information/knowledge. Data mining has been described in numerous ways, but to put it simplistically, we mine mountains in search of treasure. In our case, the mountain is our dataset; it is formatted, has structure, but the relationships between so many discrete events can be hard to find. The treasure is those events that are more related than others and to accomplish this we query our data looking for common threads, systematically minimizing our dataset to help find the treasure.

There are a variety of ways to query data in Workstation, including the Query Wizard, Filter Bar, Visual Query and Color Cueing. Query Wizard, Filter Bar and Visual Query all use SQL to query the data. Workstation will track use of any combination of these SQL driven query options to generate a single SQL statement, which can be captured (repeatable). Color Cueing utilizes pattern matching within the matrix and allows users to highlight keywords, or strings within the entire dataset – allowing users to review in a broader context. This is particularly useful for highlighting trends and for exporting to Google Earth as it allows users to utilize discrete colors based on any given variable within the matrix.
The Filter Bar is a quick and easy way to query data. Users enter, or select values from the cell filter located directly under each variable name and refine the dataset by selecting only those records that include that value(s). The cell filter can be used against more than one variable, but when applying it against multiple values, the user needs to understand that they are building an AND clause (select values that equal this from variable A and that equal this value from variable B etc). The values placed in the filter bar can include wild cards and can be used against any data format.

Depending on the type of database the user opens and queries against, the selection values may be case sensitive (e.g. dBase IV).

If the Filter Bar is not visible, it can be activated via (Main Menu bar) Edit - Preferences - General and selecting the “Show Filter Bar”.

There are two buttons associated with the Filter Bar search function:

- **Execute Filter Bar Search** or . If the Execute Filter Bar button is Red, the filter bar is populated. If the button is Blue then the cells are empty.
- **Clear Filter Bar** .
Users can enter/apply values in a variety of ways:

- Manually enter values, with or without the use of wild cards.
- Select from the Filter Options, which will adapt based on the values within any given variable to reflect text, dates or numbers.
- Select from within the matrix and apply the value to the cell filter by right clicking on a string and adding, or appending to the cell filter.

** While the Execute Cell Filter button will work perfectly well, many analysts opt to use the enter key, which will also execute the query. **

To select a value from within the matrix, select the value of interest, right click and choose either Add Cell Text to Filter Bar, or, if the user wishes to apply multiple values (Mon, or Tues, or Weds etc.) add the first value and then right click on the next value to be added and right click Append Cell Text to Filter Bar.

Users can execute the filter for each variable they apply a value against each time they add a value, or they can choose to apply all the filter values and then execute. The preference is to do it for each step so the user can see what effect the filter has on the dataset, but users do have the choice.

**Lesson 14: Minimize using the Filter Bar**

**For this lesson: Examples.mdb | Table: CFS**

In this lesson users will use the Filter bar to identify all the calls for service (CFS) from 608 Virginia Drive that occurred after 11/14/2008 (effectively the last 30-days of the dataset).

**Requirement:** The Chief received a complaint from the tenant of 608 Virginia Drive. The tenant says she calls and either nobody shows up, they show up late, or they show up and disappear immediately without listening to her concerns or taking any notice of what she has to say.
Tasks:
• Review the CFS database and filter out just the activity related to 608 Virginia Drive.
• Determine how often the citizen calls and, on average, how long the officer spends on the call.

Process:

1. File Open – navigate to Examples.mdb
2. From the Tables or query dropdown list, choose the CFS dataset.
3. Review the default settings.
4. How many records are in the dataset? 1,000?? If so, use the ‘Retrieve all Records’ button.
5. In the empty cell directly underneath the ‘Address’ variable/field enter %608 virginia% in the text box.
6. Click the Filter button on the secondary tool bar or press Enter.
7. Review the number of records returned from the status bar (bottom of the Workstation window) – 58.
8. Click on the empty cell directly underneath the RECEIVED_TIME and select Filter Options.
9. Click on the month and change it to November 01, 2008 (note the criteria “is on or after”)
10. Leave the second calendar as is (on or before December 12 2008 that meet the address criteria).
11. Review the number of records returned that match the two conditions applied (wild card 608 Virginia wild card and any received time greater or equal to 11/14/2008): 5 records met the criteria.
12. To meet the second half of the Chiefs task, we will utilize the ‘Perform Statistics on Highlighted Column(s)’ button also referred to as Quick Statistics.

13. Click (highlight) the TIME_ON_CALL header to highlight the column.

14. Now click the Perform Statistics on Highlighted Column button to view a summary of the column. For this 30 day period, officers have spent an average of 46 minutes per call at this address.

15. Click the Clear Filter Bar button to remove any filters from the Filter Bar.

Note: Separate multiple query parameters in the same field using a comma. For example, to query for both Robbery and Sexual Assault, users would type: ROB%,SEX%.

Note: Using the ‘Perform Statistics on Highlighted Column(s) generates results in one of two formats:

For numeric values it will generate a summary with Sum, Count, Average, Minimum, Maximum, Mode, Standard Deviation, Standard Error and Variance.

For non-numeric values, it will generate a list of the values, along with a count showing the number of occurrences per value. This can be very useful for quickly identifying common values within a variable/field.

Lesson 15: Filter bar query to select by crime and day-of-week

For this lesson: DC Crime
This lesson demonstrates how to use the Filter Bar to select all Commercial Burglaries that took place on Saturdays or Sundays in the Crimes.mdb dataset.

1. File Open – DC Crime – review the fields that Workstation has populated for the format.

2. How many records are in the matrix? 1,000? If so – ‘retrieve all records’. There should be 78,978 records.

3. In the empty cell directly underneath the ‘UCR variable, use the dropdown in the cell and select Filter Options.

4. Click on the Select All box (it’s a toggle so will select/deselect all) to deselect all values.

5. Select ‘Burglary Commercial’ - OK.

6. Execute the filter query by either the Execute Filter button or hitting Enter.

7. Review the bottom status bar. There should be 5,048 records that meet the criteria ‘Burglary Commercial’.

8. Right click on the empty cell to the left of MO_IR_Number.

9. In the new cell, enter Firsts and as you enter the information, Workstation will present you with the variables that meet the criteria you are entering. Select First_Day and then select TAB.

10. Right click on the empty cell underneath First_Day and choose Filter Options.

11. Select from the options Sat and Sun.

12. OK.

13. Review the bottom status bar. There should be 1,019 records that meet the criteria ‘Burglary Commercial’ AND ‘First Day’ Sunday.
Summary: The Filter Bar search is very popular for quickly whittling down datasets. While very quick, the Filter Bar uses ‘AND’ clauses, making it less useful for capturing values that may be in any one of several variables/fields. If, for example, an analyst is looking for residential burglaries where the method of entry was a ‘pry tool’, there are two fields where the data might be held and the user can’t query both at the same time.

Exercises: Minimize using specific information using the Crimes.mdb dataset answer the following:

From the entire dataset, how many residential burglaries include ‘pry tool’ as the method-of-entry?
Answer:

From the entire dataset, how many Commercial burglaries were committed along ‘Benning Rd’?
Answer:

From the entire dataset we require all events that took place at Envoy Apartments: Place the string %envoy% in the ‘Address Name’ variable field. How many records did the query return?
Answer:

Did the query return only events that took place at Envoy Apartments and if not, why not?
Answer:

What string should have been used?
Answer:

How many residential burglaries took place with a First date greater than 30 June, 2008 on Massachusetts Av?
Answer:

User Notes:
Visual Queries (SQL Statements)

Workstation provides an interface for selecting from a list of variables, operators and query statements to visually build a query. Both simple and complex SQL statements can be constructed using the Visual Query tab. These queries can then be saved and later retrieved to execute again. While the Visual Query is slightly more time consuming to build, the queries can be far more focused than the more simplistic query solutions. Unlike some of the other query types, Visual Queries allow users to use both ‘AND’ and ‘OR’ clauses.

The Visual Query tab is located at the lower left of the Hub. The Visual Query function can be used with any type of data file.

SQL Query Types

Complex Condition

The first value in the Visual Query hub is a complex condition. A complex condition is the top level condition and sets the AND/OR rule for the following elementary conditions. A complex condition can be considered a precursor to one or more elementary conditions. By applying multiple complex conditions, each with one or more elementary conditions, users can build complex queries with both ‘AND’ and ‘OR’ clauses.
Users have only one variable to affect/change:

The variable can be any one of the following: ANY, ALL, NONE, NOT ALL.
ANY – If any one of the following elementary conditions are met – select them (OR clause).
ALL – Only select records that meet all the following elementary conditions (AND clause).
NONE – Select all records that don’t meet any of the following elementary conditions.
NOT ALL – Select any records where any of the following elementary conditions are met, but not all.
Users can have multiple complex conditions allowing for mixing of clauses and very complex queries.

Elementary Condition

The elementary condition is made up of three parts that the user controls via drop downs and text boxes:
• Which field the query is to be run against
• How the condition is to be applied (lesser than, greater than, equal to, contains, etc)
• The string/keyword value.
For example: If a user wanted to query all Commercial Burglaries from the Crime variable for the entire dataset the query could be structured as follows:
‘Records where Crime contains Commer’

Advanced Elementary Condition

The advanced elementary condition primarily allows users to compare multiple variables and constants. For instance, users could select all the crimes where the “first date” and “last date” variables were equal.
Lesson 16: Query using the Visual Query Module – ‘all’

For this lesson: DC Crime

For this lesson we will generate a visual query to capture all Sex Offenses that took place in apartments from the entire dataset.

With Visual Queries, it is always good to think of the question and break it down into parts. The question has two parts:

• Select all Sex Offenses from the Crimes field.
• Select all apartments from the Location field.

1. Open DC Crime.
2. Review the field allocation to make sure they are appropriate.
3. Ensure there are more than 1,000 records – ‘Retrieve all Records’ if necessary.
4. Select the Visual Query tab on the hub.
5. Make sure the Complex Condition variable is set to ‘all’

The above means that all the following elementary conditions must be met (AND clause).
6. Left click in the box to the left of the tick box in the field.

7. Add a new elementary condition.

8. From the first drop down select UCR
9. From the second drop down select ‘contains’
10. For the last variable add ‘sex’
    Records where UCR contains sex
11. Left click in the box to the left of the tick box by the first elementary condition.
12. ‘Add a new elementary condition’.
13. Use the drop downs for the first two variables and add ‘apartm’
    Records where Location_Type contains apartm

14. Click on the ‘Execute Query’ button
15. Review the number of records returned by the query (38).
Couple of things to note:
- By using ‘all’ in the complex condition, the following elementary conditions all had to be present in the records.
- Both strings were abbreviated; ‘sex’ was used for sexual assault and ‘apartm’ was used for apartment. Using abbreviated strings for key words reduces time when building queries, but requires users to know their data. Had there been aggravated assaults as a crime type, we may have got records that didn’t meet the criteria. It’s easy to see this kind of error with queries that return small amounts of records, less so with queries that return more than a small number of records.

Lesson 17: Query using the Visual Query Module – ‘all’ and ‘any’

For this lesson: DC Crime

For this lesson we will generate a visual query to capture all Sex Offenses that took place in apartments on Saturdays or Sundays.

This question has three parts:
• Select all Sex Offenses from the UCR variable
• Select all apartments from the Location variable
• Select events that occurred on Saturday or Sunday

The first two parts have to both be present in each record. For the third part either Saturday or Sunday must be present. This requires using a second complex condition to allow the user to apply an ‘OR’ clause, which is the ‘any’ option.

1. Open DC Crime.
2. Review the field allocation to make sure they are appropriate.
3. Ensure there are more than 1,000 records – ‘Retrieve all Records’ if necessary.
4. Select the Visual Query tab on the hub.
5. Make sure the complex condition variable is set to ‘all’.
6. Left click in the box to the left of the tick box and add a new elementary condition.
7. Use the drop downs for the first two variables and manually add ‘assault’.

Records where UCR contains sex
8. Add a new elementary condition.
9. Use the drop downs for the first two variables and add ‘apartm’.

Tactical Analysis using Accurint® Crime Analysis Workstation
Records where Location_Type contains apartm
10. Left click in the box to the left of the tick box by the second elementary condition.
11. Add a new complex condition.
12. Change the complex condition variable to ‘any’
13. Add a new elementary condition.
14. Use the drop downs for the first two variables and manually add ‘sun’.

Records where First Day contains sun
15. Add a new elementary condition.
16. Use the drop downs for the first two variables and manually add ‘sat’.

Records where First Day contains sat

The Visual Query should be the same as:

17. Execute query .
18. Review the results. There should be 7 records that meet the criteria above.

Notes:
Look at the structure of the visual query. In the lesson above we applied the ‘all’ first and the ‘any’ second. It didn’t really matter which way around we applied them, but it’s important to phrase the question in such a way that it makes sense.

The query can be executed at each individual stage or run once all the conditions have been applied. While learning and becoming familiar with building visual queries it is highly recommended users execute often. This will provide sanity checks along the way and make it easier to identify where errors occur.
Lesson 18: Query using the Visual Query Module – ‘all’ and ‘any’

For this lesson we will add a few more components to demonstrate, not only the strengths of Visual Queries, but the things analysts need to be aware of when building them.

The requirement: Identify all records from within the DC Crime database that meet the following criteria. Commercial Burglaries where the point of entry was window and that took place between Feb. 1, 2009 and July 31, 2009.

This question has three parts:
• Select all Burglaries from the UCR variable
• Select all records where the point of entry was window
• Select events that occurred between 2/1/2009 – 7/31/2009

Breaking it down one more level to include where we will find the data needed

Burglaries are all reported in the UCR variable/field
We want all events that are equal to, or greater than 2/1/2009 – FIRST TIME
We want all events that are equal to, or lesser than 7/31/2009 – FIRST TIME
We want all events that have window in the ‘point of entry’ field – POINT OF ENTRY (1) AND/OR (2)

1. Open DC Crime.
2. Select Visual Query.
3. Make sure the complex condition variable is set to ‘all’.
4. Add a new elementary condition.
5. Records where UCR contains comm.
6. Records where First Date is greater than or equal to 2/1/2009.
7. Records where First Date is less than or equal to 7/31/2009.
8. Add a new complex condition – ‘any’.
9. Add a new elementary condition.
10. Records where point of entry(1) contains wind
11. Records where point of entry(2) contains wind

![Visual Query Image]

12. Execute query

Review the results. There should be 142 records that meet the criteria above.

**Summary:** In lesson 14 the OR clause “any” was used to pick out Saturday or Sunday from a single field. In lesson 15, we had a single value, but with two variables/fields that the value could be found in. In cases where there are multiple locations where the value might be found, it may be necessary to search on the value before building the Visual Query. In DC Crimes.mdb there are five fields for ‘Suspect Actions against Persons’. If the user is building a query that will be repeated, the visual query should be built to capture all potential variables where the value could be found. For queries being built to answer a one-time question, users can use the Omnisearch to find out which variables house the value.

**Lesson 19: Saving/Loading Visual Queries**

The main benefit of using Visual Queries as opposed to others is that, once built, they can be called up and repeated. Building a Visual Query to separate out Part 1 crimes and another for Part 2 crimes by week may take a few minutes to build, but once built can be recalled, edited and rerun in seconds.

To demonstrate the save, we will build a visual query to identify thefts where the suspect used deceptive practices between 1/1/2010-11/1/2010.

We will build the query so it can be repeated week on week, but will use omnisearch to determine which fields the value has been seen to date.
1. Open DC Crime.
2. In the Omnisearch cell type ‘deceptive’

3. Omnisearch Current Combo Text.
4. Review the Suspects actions against Persons variable in the Matrix – all instances are in the first field(1).

We now know that – to date – all references to deceptive practices for thefts have been in Suspects Actions Against Persons(1). Unfortunately we can’t guarantee this will always be the case; if we want to re-use the query we have to assume it may be put into a different field at some point in the future and build the query accordingly.

5. Select Visual Query.
6. Make sure the complex condition variable is set to ‘all’.
7. Add a new elementary condition and build the query based on the parameters below.

8. Click on the Save Query to File button.
9. Name the query ‘Thefts’.
10. Click on Save.
11. Click on Clear Query.
12. ‘Retrieve all Records’.
13. ‘Load Query from File’.
14. Select and Open ‘Thefts’.
15. The visual query should now be back in the Visual Query hub.
17. Review results (21 records).
**Summary:** Visual Queries are very powerful. Unlike the Filter Bar, users can query against multiple fields with OR and AND clauses as opposed to a limited AND/OR option. They do require effort to become proficient with, but once built they are easy to save, are shareable and, if built correctly, easy to tweak. By creating queries that match key elements of your trends, you can routinely run the queries to identify a possible reemergence of a trend or a new case that may be related to the trend.

**Exercises:**

Build a Visual Query to identify Residential Burglaries that include the theft of computer(s) for the month of March, 2010.

How many computers were stolen during the month?

Save the query as ‘computers’

Clear the query and retrieve all records

Load the ‘computers’ query from file

Adjust the ‘computers’ query to identify how many computers were stolen during residential burglaries during the months of January and February as well.

How many computers were stolen during those months?

Adjust the layout to show just the Crime, First Date, Address of Crime and the Property Stolen related fields.

Export the results to a new .html file named ‘computers’.

**User Notes:**
The *Query Wizard* allows users to query a database by selecting variables from the Query Wizard interface. Workstation uses the ‘Table Database’ to capture the variable values which users can then select from. In addition to providing easy focus and minimizing, Workstation generates SQL (Structured Query Language) statements, which can be saved and re-used as necessary. The Query Wizard is not case sensitive; making it very easy for users to quickly investigate variables with fixed values, but it is *only an option with .atx datasets*.

Note: If there are already parameters in the Query Wizard, the handle of the magnifying glass will turn RED to remind the user that a query is in effect.

*Query Wizard button without parameters set in the Query Wizard.*
*Query Wizard button with parameters set in the Query Wizard.*

**Lesson 20: Minimizing via the Query Wizard**

**For this lesson:** Sample.atx

1. Select **Query Wizard** from the **Analysis** menu or select the **Query Wizard** button from the main tool bar.
2. Select “**BURGLARY RESIDENTIAL**” from the Crime drop down list.

3. Click the **Execute Query** button or press the **Enter** key on your keyboard. The Query Wizard will automatically close. Note the number of records returned from the record counter indicator located on the bottom of the Hub’s status bar (518). This query returned only those incidents where “BURGLARY RESIDENTIAL” was contained in the Crime variable.

We will now use Query Wizard to retrieve ALL Burglaries.

1. Select the Query Wizard from the Analysis menu or select the Query Wizard button from the main tool bar and enter “BURG%” in the Crime drop down field.

2. Click the **Execute Query** button or press the **Enter** key on the keyboard. Note the number of records returned from the record count indicator located on the bottom of the Hub’s status bar (686). Also note that the Crime variable in the Hub’s data grid now contains both BURGLARY RESIDENTIAL and BURGLARY COMMERCIAL crimes.
By placing the wild card after Burg, and not before, the query only returns crimes that start with burg and then have additional characters. Multiple values can be queried by discrete variable, using the appropriate field name with a value, or by running strings alongside one another separated by a comma. For example: “BURG%,%SEX%” in the crime field.

**Summary:** Query Wizard provides a quick directed search function within Workstation, designed for use against .atx formatted datasets. It generates discrete SQL queries against the dataset. Whenever users run a query using Query Wizard, the query is against the entire dataset, so queries must be structured. Queries can be amended, as shown in the example above where burg%,%sex% were added to by adding another variable. Queries can be built against Modus Operandi, Persons or Vehicle, but they are not discrete. If users have burg% in the Modus Operandi window as the ‘Crime’ and then switch to the vehicle tab and select Toyota as ‘Make’, the query will look for the word burglary in the crime field and Toyota as the make of vehicle. Any queries built in the Query Wizard that search against more than one variable are ‘AND’ clauses.
In the previous lessons you learned about SQL (Structured Query Language) as the means to query/select data in various datasets. SQL is a powerful language which can be used against data, but as it names implies, it’s structured. You must designate fields to query and adhere to certain rules when passing query parameters to the data. Another language that can be used to perform data mining is Regular Expressions or RegEx. RegEx has a lot of the same syntax as SQL, but RegEx is geared more toward data retrieval and entity extraction. RegEx can be used to query for the word “copper” within any field, without designating the field and then extract that word and insert it into a field for future retrieval or analysis. RegEx is extremely powerful. Workstation provides several ways to leverage RegEx against any data by way of the KWIC (Key Word in Context) module and the Concepts module. Both will enable you to data mine any dataset as well as perform entity extraction again the data.
KWIC expressions provide a concise and flexible means for identifying strings of interest, such as particular characters, words, or patterns of characters within a dataset. The KWIC Expressions module is divided up into three main windows. The topmost window contains the expressions that you want to run, along with tips. The middle window area is the Test String, where users can validate expressions against a smaller subset of data. The bottom area is used to display the results.

If the test string window contains text, the KWIC Expression will only be run against that text. To execute the expression against the underlying data, remove any text from the sample text area.

Simple Expressions

A simple expression searches for a single word or phrase, presumably in a single variable.

Lesson 21: Query for Unique Terms using KWIC Expressions

For this lesson: DC Crime

1. Query for Burglary Commercial in the UCR field.
2. From the Hub, select the KWIC Expression option from the Analysis menu. Or, click CTRL + Q.
3. In the Expression text area at the top, enter “COPPER”.
4. Select Run KWIC Expression from the toolbar.

Note: Green Sphere indicates there have been items found that match the string ‘Copper’.
The results area will display the string queried along with a green or gray sphere. A green sphere indicates there are results; a gray sphere indicates the expression found no matches. To expand the results, users click the plus box located alongside the sphere in the **Expression History**.

The ‘Expression History’ provides an overall summary including: the expression/string, number of times the string was matched within the dataset, name of the dataset the string was run against and the SQL statement generated. The ‘Results’ tab provides a more detailed summary that includes the case number(s) the string was found in and where, within the record the string was matched.

5. In the Expression text area at the top, enter “**COPER**”.

6. Select **Run KWIC Expression** ![Run KWIC Expression Button](image) from the toolbar.

![KWIC Expression Window](image)

**Note:** Grey Sphere indicates there have not been items found that match the string ‘Coper’.

7. Exit the KWIC Expression window.

8. Open Concepts via the button ![Concepts Button](image) located on the primary tool bar or via the Analysis dropdown on the Main Menu bar.

9. Select the Results tab.

The workbook will go into more detail regarding the function of Concepts in subsequent lessons, but concepts sit alongside and complement KWIC Expressions. The results of KWIC Expression queries are reflected in the Concept results and provide an extra level of functionality to users. In addition to displaying the incident number where the expression/string was found, when users select an incident numbers from within the window, the matching record is highlighted and promoted to the top of the matrix.
10. Double-click the row with IR Number 0008000407. Workstation will move the incident to the top of the matrix.

<table>
<thead>
<tr>
<th>IRNumber</th>
<th>Crime</th>
<th>UCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0008000407</td>
<td>BURGLARY COMM</td>
<td>BURGLARY COMM</td>
</tr>
</tbody>
</table>


"Or" Expressions
Simple expressions can be created to look for variants of a word or phrase. The pipe character ( | ) is the reserved character that instructs the Quick Expression builder to search using “OR” operators. For example, “Jane | John” would return all those records where Jane “OR” John appeared in any text. Multiple words or phrases can be separated by pipes to create expressions.

Lesson 22: Query using the Either Operator

For this lesson: DC Crime

1. Query for Burglary Commercial in the UCR field.
2. From the Hub, select the KWIC Expression option from the Analysis menu. Or, click CTRL + Q.
3. In the Expression text area at the top, enter “CONSTRUCTION EQUIPMENT | APPLIANCE.”

4. Select **RUN KWIC Expression** from the toolbar. The Expression History will display the search results matching your expression.

Note: The ‘OR’ operator can be very useful when looking for strings in the dataset, but – as always – the analyst has to be part of the process. Typo’s happen. Had the user made a typo in ‘Construction equipment’, appliance would have returned 38 results. It isn’t sufficient to see results; users still have the responsibility to validate their work.

5. Review the results of the query.

6. Close out of the KWIC Expression window.

**Order Specific Expressions**

The benefit of using KWIC expressions, rather than directed queries against set variables is that the KWIC expression string runs against all the variables within the dataset. That said, some strings are going to be very common, making it important that some form of limiter be applied. In KWIC Expressions, one way to achieve this is via the ‘AND’ operator.

To combine two words with an “AND” statement you use the “.+” operator. For example: “robbery.+handgun” would recover all events where the word “robbery” AND “handgun” appeared. The ‘AND’ operator is order dependent, but proximity independent. For example:

- Robbery.+handgun will recover all records where robbery and handgun are found in that order.
- Handgun.+robbery will recover all records where handgun and robbery are found in that order.
Lesson 23: Identifying Records in a Specific Order

For this lesson: DC Crime

1. Query for **Homicide** within the UCR field.
2. From the Hub, select the KWIC Expression option from the Analysis menu. Or, click **CTRL + Q**.
3. In the **Expression** text area at the top, enter **HOMICIDE.+HANDGUN**.
4. Select **Run KWIC Expression** from the toolbar.

![KWIC Expression Result](image)

Note: Homicide and Handgun in this sequence results in 14 records being found.

5. In the **Expression** text area at the top, enter **HANDGUN.+HOMICIDE**
6. Select **Run KWIC Expression** from the toolbar.

![KWIC Expression Result](image)

Note: Putting Handgun before Homicide has changed our results significantly.
Using the AND operator this way can be an effective way of interrogating the datasets, but as can be seen above, the order of the words is critical.

**Combined Expressions**

When users apply the AND operator, the strings must be in the order defined in the syntax. If a user doesn’t know which order the strings will be within the data, they have the option to apply combined expressions – use both AND and OR operators. Using the example from the previous lesson; if a user wanted all instances where both handgun and homicide occur within a single record, regardless of order, they would need to query: all events containing Homicide AND Handgun OR all events containing Handgun AND Homicide.

Here is how the statement is expressed:

- homicide.+handgun|handgun.+homicide
- Homicide AND Handgun OR Handgun AND Homicide

**Lesson 24: Identifying Records Regardless of Order**

1. Query for Homicide within the UCR field.
2. From the Hub, select the KWIC Expression option from the Analysis menu. Or, click **CTRL + Q**.
3. In the Expression area at the top, enter HOMICIDE.+HANDGUN|HANDGUN.+HOMICIDE
4. Select **Run KWIC Expression** from the toolbar.

5. Right click on the green sphere in the expression history and **Clear List**.
Proximity Expressions

When the proximity of words isn’t important, the “and” expression can be very useful. However, you would not want to use the “and” expression when the proximity of the words does matter. For example, finding the words “black” and “SUV” using an “and” expression would return every record where the word black and SUV appeared. In this case, using an “and” statement could return the following unwanted results:

“A black male was seen leaving in a green SUV.”

In this case using proximity expressions is far more powerful.

Lesson 25: Query for Unique Terms Near One Another

1. Using the DC Crime database, query for Burglary Commercial within the UCR field.
2. Select KWIC Expression from the Analysis menu.
3. In the “Expression” text area at the top, enter “COPPER within 5 words of TUBING.”
4. Press the Enter key or select the Run KWIC Expression button on the toolbar to execute the search.
5. Click the plus sign next to the green ball to see some details regarding the search, expression, and number of records found.
6. Select the Results tab
In this example, the terms ‘Copper’ and ‘Tubing’ were alongside one another, but the expression would have brought back any records where copper and tubing were within five words of one another.

Exercise 10: Using the DC Crime dataset, query for Robbery within the UCR field and then identify how many crimes included the following values:
- Glock
- Pistol
- Handgun
- Firearm
- Revolver
Place the number of instances per value by the discrete string above.
The total number of records containing any of the above =

Exercise 11: Using the DC Crime dataset, query for %simple assault% in the Crime field and then identify how many crimes include the following values, regardless of order/sequence:
- Handgun and Glock =
- Black and Glock =
Exercise 12: Using the DC Crime dataset, identify how many crimes include the theft of firearms from the Property Taken fields.

User notes on KWIC Expressions:
Concepts are a collection of terms that use regular expressions. First, a concept is developed that describes an item or action. Then the analyst can combine concepts using Regular Expressions to uncover almost anything imaginable.

For example, a Concept might be used to find where a “green motorcycle” was used in a crime. In this case, the analyst might instruct Workstation to search for the Concept “green” within 20 words of the “motorcycle” Concept.

**Step 1.** The analyst first develops a concept for the color green by combining synonyms for “green” such as: “green”, “turquoise”, “teal”, “grn” and even “bluish”.

**Step 2.** Then the analyst compiles synonyms for motorcycle including “motorcycle”, “moped”, “chopper”, etc.

**Step 3.** Once the concepts are developed, the analyst simply tells the program to conduct a search using these concepts.

More complex and advanced concepts can be developed with a mastery of Regular Expressions. For example, the military might want to identify references to “tanks” in data. A concept that looks for tanks might instruct the program to look for “tank” not occurring within a certain number of words of “fish”, “gas”, or “propane.” Advanced Concepts can reduce the possibility of false positives and improve an analyst’s efficiency and effectiveness.

Concepts are stored in a database created and stored in the root directory of where Workstation is installed and used. Users can open, create and merge Concept databases easily from the Concept toolbar. Different users can specify a default Concept database in the Accounts Manager. Storing the Concepts in their own database enables agencies and users to share them with other users. Workstation’s default Concepts database is named “concepts.mdb” and is located in the Workstation program files folder.

Sharing Concepts is straightforward and our website hosts concepts that have been built by other users that new users can extract and tailor. Internally generated concepts display contact information, so the user will know who to contact if and when a watch item is identified in their data. The “reference” variable is used to store any information that would enable you to reference another agency’s concept, case, or incident number.

**Concepts**
The Concepts tool offers the capability to build complex, custom searches for patterns within your dataset and to edit/add variables at scale. There are three types of concepts that can be created within the Concepts tool:

- Find
- Replace
- Update
Find Concept
The ‘Find’ concept allows users to find strings that are related. The strings could be pseudonyms for popular drugs, names of discrete SUV types (large/mid/small SUV’s etc.) or strings/terms that, when seen within a certain proximity, make them relevant, but otherwise be innocuous.

Replace Concept
The ‘Replace’ concept enables standardization of data. Every agency enters information into their database uniquely, which can prove problematic to the analyst when combining data, as queries will lack effectiveness.

For example, if Police Department A entered “Door, dog” into their data as a point of entry; however, Police Department B entered “Doggy door” - then a query for “Doggy Door” throughout the combined data would not accurately reflect the data. In this case, the Replace tool within Concepts would be able to rectify this. An analyst would be able to create a ‘Replace’ Concept wherein they could highlight all records within the data that say “Door, dog” and replace it with “Doggy door.”

Update Concept
The “Update” modifies unstructured data into structured data. An example would be mining the synopsis field for references to gang related activities and creating/populating a new field to enable analysis. In doing so, the user can then run stats, visualize the influence of gangs, both collectively and by gang name.
Search all row variables combined and search each variable individually only apply to the concept type ‘Find’.

Using ‘Search all row variables combined’ results in the concept term being run against each entire record, with no regard to discrete variables. This type of search may result in false positives in unrelated fields especially when using colors.

Lesson 26: REPLACE: Analyzing data gathered from pirates around the world.

For this lesson: Examples.mdb | Table: Dirty Maritime Piracy

1. Open data.
2. Highlight ‘Status’ Field, then select the Quick Stats tool.

Executing this query has proven that some of our data is entered incorrectly. In order to fix our data, we will create a Replace Concept.

3. Ensuring that the ‘Status’ field is still highlighted - Create a new ‘Replace’ Concept – name it “Data Correction”.
4. In the ‘Replace With’ field – enter “Anchored”.
5. Under ‘Concept Development’ – enter “Achored” (as seen in the Summary Statistics box) – Build Concept.
6. Save Concept.
7. Ensure that only the ‘Data Correction’ concept is selected.
8. Select ‘Replace/Update in Highlight Variable’ from dropdown.
9. Highlight the ‘Status’ field, then select Quick Stats tool.

“Achored” has now been changed throughout our data to “Anchored” allowing for greater accuracy throughout our data.

**Exercise**

**For this exercise:** Examples.mdb | **Table:** Dirty Maritime Piracy

Using what you have learnt from this lesson, make corrections for the remaining issues in the Status Field. Change “Steamig” to “Steaming” and “Beaced” to “Beached”.

**Update Concept**

In the lessons that follow, we will be focusing on the Update type Concept, which modifies unstructured data into structured data.

**Lesson 27: Creating a new field to house values from existing fields**

In this lesson, the user will select all commercial burglaries that contain the word copper and populate a new variable called Copper Thefts to allow focused analysis and/or data manipulation. The update concept option populates the selected variable with the value used to build the concept, which in this case will be copper.

**For this lesson:** DC Crime

1. Ensure All Records are present.
2. Create a filter in the ‘First Date’ field to focus only on 2010.
3. Filter for **Burglary Commercial** within the UCR field.
4. In order to work on a smaller data set, we will now export our data to Microsoft Access and then re-open it into Workstation.

5. File – Export Data – Save to ‘C: temp' as DC Crime, 2010 (Leave Table Name as Data).


7. Retrieve All Records (You should have 1664 records).

8. Select Edit – Create Fields.

9. In the User Defined Field – Table: Data, User Defined: Enter- COPPER THEFT, Type: Text, Length: 255.

10. Then select Create Fields.

11. After the refresh select Cancel, and then scroll to the right to ensure COPPER THEFT is a field.


15. Concept Type: Update, Concept Name: Copper Thefts.
16. Select the Concept Development tab:

17. In the Simple Concept area: Word, Phrase, or Concept, insert ‘copper’.

18. Select Build Concept button.

19. Select the Concept tab.

20. From the ‘Update To’ dropdown option, select COPPER THEFT.

21. Save concept.

22. Ensure only the ‘Copper Thefts’ concept is selected (tick in the box).

23. Select ‘Replace/Update in Highlighted Variable’ from dropdown.

24. Highlight COPPER THEFTS field, select Quick Stats from toolbar.

You should now have 139 records populated within your COPPER THEFTS field.
Lesson 28: Identifying gang related assaults that occurred within Washington D.C. in 2008

For this lesson: DC Crime

1. Retrieve All Records.
2. Filter for ‘Assault’ within the UCR field.
3. Create a filter in the ‘First Date’ field to focus only on 2008.
6. Select Edit – Create Fields.
7. Title the new field: GANG.
10. Under the Concept Development tab, enter ‘bloods|crips|disciples’ – Build Concept.
11. From the ‘Update To’ dropdown option, select GANG.
12. Save concept.
13. Ensure that only the ‘Gang Assaults’ concept is selected.
14. Select ‘Replace/Update in Highlighted Variable’ from dropdown.
15. Highlight GANG field in your data, select Quick Stats tool.

You should now have 45 records populated within your GANG field.
Exercise
For this exercise: DC Crime

Using what you have learned during this lesson, query for the total amount of homicides that have occurred in the last three years. Create a field called ‘FIREARM’ then generate an Update Concept in order to populate the field. Use the Quick Stats tool to confirm completion.

(Hint: search for handgun, gun, shotgun)
Maximize Using the Trend Hunter

When an analyst searches for patterns, series and trends in their data they use a mental methodology when comparing reports. Depending on the crime type being focused on there are certain characteristics that we place different values (or weights) upon. If searching for a burglary series for instance, the most important factors to match are probably the point of entry, method of entry and property taken. Person’s crimes such as robberies have observed suspects, so we place more importance on fields such as suspect descriptions and weapon usage. The scores come from our judgment as an analyst as we make the determinations of if cases are related or not. Workstation’s Trend Hunter module takes those scores and puts them in action for you.

The Trend Hunter scans reports in your dataset and awards points to fields that have matching values to your designed case. Reports accumulate points for each matching field. The amount of points each field is worth is determined by you when the scan is built. Each field can be assigned points individually, or they may be grouped. For example, if all three fields; age, weight, and gender will receive 25 points if they match your designated case, they may all be grouped into a “25 point” group. Each variable will still receive its own points however, giving the combination of a possible 75 point total if they all match.

Lesson 29: Create a New Scan

For this lesson: DC Crime

In this lesson we will create a layout that will help users identify sexual assault patterns.

2. Select the **Add Trend Scan** choice from the first dropdown.

3. Name your scan. Choose the primary crime type. Add any notes in the description fields for your reference at a later point.

4. Next we will start adding variables into groups. Choose “Add Scan Group” from the second dropdown.

5. Scan groups can contain multiple variables that will receive the same number of points. This is just for convenience, each value still operates independently.
After all the scan groups are complete, the inclusive scan will compare your designated case against ALL of the rest of the cases in the dataset. Each case will accumulate points for its matching fields and its final score written into the “commonalities” fields in the Workstation matrix. They will be ranked in descending order.
The Time Series Analysis Module

The Time Series Analysis module will enable you to perform most time series methods on all or a selection of your data. The Time Series module supports several statistical models including: Exact Time Analysis, Weighted Method, Equal Opportunity, Midpoint Analysis, Duration, and day of the week analysis. When the Time Series has completed its analysis, you can then choose individual or ranges of times to determine cumulative or hourly percentages. The Time Series module will display the results of the temporal distributions through an area, line, or bar graph. It will also calculate out to three standard deviations providing both potential beginning and ending times for each deviation.

Descriptive Analysis

Temporal Topology

The Temporal Topology function allows you to create a Virtual Temporal Topology (VTT) based on Hour of Day and Day of Week. A VTT is a simple way to clearly visualize complex data arising from cross tabulation of two types of temporal cycles. This module focuses on the most useful temporal cross tabulation; Hour of Day (00-24) and Day of Week (SUN – SAT).

Using the Weighted Method and Day of Week functions, you can determine which Days of the Week and which Hours of the Day have the highest and lowest activity, and you can derive some simple models of your data. Unfortunately, this simplistic picture can be misleading. Even if a given weekday and hour have the highest levels of activity individually, the combination of hour and weekday, together, might lead to some very different results. Using a VTT to display these data the analyst can make more accurate determinations as to how (and potentially why) activity might develop.

Lesson 30: Create a Temporal Topology

For this lesson: DC Crime

In this lesson, we will focus on Arsons that occurred in 2010 throughout Washington, D.C.

Using the DC Crime database, query for Arson within the UCR field, then filter down to view only those crimes that occurred in 2010.

1. Select Time Series Analysis button from the toolbar.
2. Select the **Weighted Method Day** graph from the list of analytics. Notice the distribution of events across the hours in the day.

3. Now select **Temporal Topology Weighted** from the list of analytics.

4. **Click and drag the graph** to view the Topology from various perspectives. Pay particular attention to the hotspots that show the times of days and days of week that have the most overlap.

5. Now, select the **Temporal Cityscape** analytic.
6. Modify the graph’s colors and labeling. Select File – ‘Set Default Color Scheme’ and change to ‘Green to Red’.

Every combination of 24-hour time of day and seven-day day-of-week will be used to count events. The number of events at each intersection is tabulated, and this is treated like a “Z” coordinate.

Imagine the topology as being like a map of terrain:

**X-Coordinate** (horizontal) indicates the hour of the day.

**Y-Coordinate** (vertical) indicates the day of the week.

**Z-Coordinate** is represented by color changes and height.

To identify the level of activity at a particular day and hour, merely cross-reference the Y and X coordinates representing the values you seek, and then compare the color at that point to the legend.

Although the model describes a three-dimensional surface, the default display is only two-dimensional. This display works most clearly with large data sets. When the analyst is attempting to model the interaction of a small number of cases (such as a crime series) it may be difficult to clearly visualize your surface in two dimensions.

Reading a Virtual Temporal Topology is simple. Colors show areas of high and low activity at the intersections of each day and hour. The more “hot” the color, the more active that day/hour value. This can be a useful way to visualize how your events are occurring in the context of a weekly and daily activity cycle, which is extremely common in tactical crime analysis. While in 3D mode, the analyst may use the Pitch and Elevation slider controls to manipulate the view of the topology to better effect.

As in reference to the robbery example, we can see that our initial analysis and hypothesis about the distribution of our events was correct. The temporal “hotspot” is around 2100 hours on Saturday. That is, the majority of our robberies occurred then. Now we have a tool that can truly analyze and display time and day simultaneously!
Note: Adjust graph colors and settings by right clicking on the graph to display the “3d Chart Control Properties” window.

Predictive Analysis

Lesson 31: Calculate Time Series Analysis

For this lesson: DC Crime

For the following lesson, we will look at robberies that occurred from 2008 – 2010.

1. Query for Robbery within the UCR field, then for the date range of January, 2008 – December, 2010.

2. Select the Time Series button  from the toolbar.

Note: When you load the Time Series module, Workstation will automatically update your query using only primary data (for an explanation of primary data, see the section titled “Query Preferences” in the Help File.) Workstation will calculate time series for each MO in an incident. Therefore, those incidents having several MO will have each MO factored into the calculation.
The Time Series module displays time data both graphically and numerically. The top portion of the screen is dedicated to displaying the hourly percentage results and the graph. The lower portion of the Time Series module is where the mean and standard deviations are displayed.

The Status Bar, located along the bottom of the Time Series screen is used to display several key elements of information. The first, leftmost region is used to display the count of those records included. The second box will display the number of cases excluded from the analysis; those whose difference between first and last time exceed 24 hours (or whatever you have your “Max Gap” temporal preference set). Those crimes that exceed your Max Allowable Gap, Earliest Permissible Date and Latest Permissible Date as set in your Temporal Preferences module will be excluded from all-time series analysis. Likewise, the Time Series module will not display those standard deviations that are greater than a 24 hour time period.

The Time Series Graph will display the results of your Time Series selection in either a line, bar, area, or Radar graph format. The Time Series graph’s display properties can be modified to a variety of formats and colors. Both horizontal and vertical lines can be added, the graph’s background, foreground, and text color, as well as the graph type can be changed to enhance your analysis or increase the readability when printed. Finally, the Time Series graphs can be exported to a JPG image format.

3. Select Weighted Method Day from the list of analytics on the left side of the dashboard.
4. Notice that each hour of the 24 hour day is represented by a percentage bar showing the percentage that hour accounted for all hours. Below the graph, each hour is again represented with the percentage that hour accounted specifically.

5. Select Equal Opportunity from the analytics list. The graph represents the 24 hours of a day.

6. Click the bar for the 21st hour of the day to select the bar. Notice that the color of the selected bar will change and the Cumulative Percent located in the lower left of the dashboard will adjust based on your selection.

7. Continue this process for hours 21, 22, and 23. The percentages from the selected bars will be added to the Cumulative Percent value.

8. As you can see, from 2100 hours to 2359 hours accounted for approximately 24% of the activity of all the hours of the day. In other words, 24% of the cases occurred within this three hour span of time.

9. From viewing the results of the Weighted Method and Equal Opportunity analysis, it appears that about a third of the cases fell between 8 p.m. and 12 p.m. That is, our bell curve tends to be slightly more toward the midnight hour than the noon hour. This is important to note for conducting Exact Time analysis and Midpoint Analysis, as you will now see. Also notice that the results of the
Weighted Method are basically identical to the Equal Opportunity results. This is because there were no cases that had “gaps” or spans of time in which the crime could have occurred. All the robberies in this series had exact times.

10. Select **Midpoint Analysis** from the analytics list. The dashboard will automatically switch to the Statistics tab to display the results of your analysis. The results include the Mean Time and three categories of standard deviation.

11. Select **Day of Week** from the list of analytics.

In this series, no cases had time spans. Therefore, the exact day the crime occurred was used in the calculation. The results of the Day of Week graph show that Saturday is the day with the most incidents.

**Note:** If you analyze cases that have time spans that crossed the midnight hour, the day that contains the most time will be used in the calculation.
Lesson 32: Next Event Prediction

For the following lesson, we will look at a commercial burglary series. In the Trend field, query for Cat Burglar.

12. Select **Value/Interval Regression** from the analytics list. This tool uses value as a measurement to judge the offender’s pattern. In this example, it relates Property Value taken from each crime to the date of the event, and based on the result predicts upon the next event.

13. Select **Next Event Prediction** from the list of analytics. This tool plots all events on a graph, comparing sequence against interval, then predicts upon when the next event is likely to occur.

**Calendar**

This feature enables the user the ability to analyze their data using a relational calendar. This tool is especially helpful as it allows for the analysis of crime while viewing all 365 days at one time. The colors used for each day refers to the volume of crime that occurred on a given day. Another helpful aspect of this tool is the
ability to not refresh your data at all. For example, if you selected May 10th and directly after September 7th, the matrix automatically refreshes the data for each day.

Lesson 33: Applying Compstat Analysis to your data

For this lesson: DC Crime

1. Filter down to 2010 data only.
2. Export it, and then re-open the new database. (Should have 23,698 records)
3. Select Time Series Analysis button from the toolbar.
4. Select Calendar.
Workstation - Google Earth

The results of your spatial analysis can be exported to Google Earth in KML format.

Note: Only data containing valid Decimal Degree Latitude and Longitude values can be properly displayed in Google Earth. Projected data will not display (e.g. NAD 27/83 or UTM data.)

To send data to Google Earth, right click on the Matrix and click “View in Google Earth”. Google Earth Options will appear allowing you to change a number of display options before loading the data in Google Earth.

The Google Earth entry wizard will open up as shown below. There is a great deal of functionality provided, but many of the options are for more advanced analytics. They are covered within the help files, but for the purposes of this class we will focus on the more standard options.

The wizard will automatically look for known fields linked to coordinates, such as Lat/Lon and X/Y.

When reviewing the wizard window, if the X Coordinate and Y Coordinate fields are not populated, then the user will need to enter them. The user is also required to enter a file name for output purposes (please add the .kml suffix to the file name).

Selecting how Google Earth will represent the data is achieved by selecting options from the titles on the right and completing the associated values needed from the labels to the left of the window. The symbology can be changed at this point, along with the balloon fill and text options (these are the labels that will be attached to the points when pushed to Google Earth.)

Mapping Points

The most basic tool within Google Earth is to map your points. This functionality enables the user to view where events have taken place in relation to one another, on a virtual globe. Mapping points is a simple, yet very necessary function within Google Earth.
Lesson 34: Placing points on a map

For this lesson: Examples.mdb  |  Table: Mailbox Bomber

1. Open data.
2. Right click on data – Select ‘View in Google Earth’
3. Check ‘Show Points’
4. Save Output file as: ‘... C: temp\MailboxBomber.kml.

Point to Point

The point to point tool within Google Earth allows the user to link relational data on a map.

For this example, we will look at auto thefts in our data. This tool enables the user to link where the vehicle was stolen with where it was recovered. For this map, the red points indicate where the vehicle was stolen; whereas the green points represent where the vehicle has been recovered.

Lesson 35: Use the Point to Point function to identify auto thefts.

For this lesson: Bentley Bandit.xls

1. Open Bentley Bandit.xls.
2. As learnt in previous lessons, export your data into a .mdb format.
3. Once data is formatted correctly, open Google Earth – Check ‘Show Points’ and ‘Point to Point Links’.
4. Select ‘Point to Point’ from list on left.
5. Ensure that Starting Point X and Y are set correctly, and then set Ending Point X and Y to Recover_X and Recover_Y.
6. Save as ‘. . . C: temp\BentleyBandit.kml’.

Altitude

The Altitude tool within Google Earth enables users to assign a value to a specific field, using height to represent the assigned value. For example, if you assign altitude to represent ‘Time on Call’ in your map, then the highest points would reflect the longest calls; while the shortest calls would be closest to the Earth.

Lesson 36: Generating an altitude map within Google Earth.

For this lesson: Examples.mdb | Table: Vegas Bank Robber

For this lesson we will be analyzing a series of bank robberies. We will assign ‘property value’ as the value to each incident on our map.
7. Open data.
8. Select Google Earth.
9. Check ‘Clear All Checkboxes’.
10. Check ‘Extrude Points based on Altitude’ and ‘Create Altitude as Bars Layer’.
11. Select Altitude from list on left.
12. Set Altitude to ‘Property Value’, Max Height to ‘2’ Miles (or 10,560 feet) and Opacity to ‘75’.
13. Save Output File as ‘ . . . C: \temp\BankRobber.kml’.
14. Select Save.

Color Coding
Color coding is a tool within Workstation that enables the user to visually assign each row a value indicated by color, allowing for greater visualization of your data while in the matrix. The tool enables the user to either color code their data by density or by value. Color coding by density will assign colors to your records comparable to a hot spot map. In our example, however, we will be examining the importance of color coding by value.

Lesson 37: Color coding records within the matrix.

For this lesson: DC Crime

1. Open data.
2. Use the filter bar to query for Sex Offenses in the UCR field.
3. Select the ‘Color Cue by Value’ button from the tool bar.
4. Set the ‘Field to Colorize’ to Location Type and the ‘Color Ramp’ to Thermal.

5. Select ‘Turn Color Values On/Off’

Color Coding with Google Earth

While it is very helpful to view our color coded records in the matrix, it is even more valuable to view on a map. The Google Earth tool within Workstation allows users to generate points based on how the records are color coded within the matrix.
Lesson 38: Pushing out our color coded records to Google Earth.
1. With the rows still color coded, select Google Earth.
2. Select ‘Show Points’, ‘Draw Legend’, and ‘Use Row Color for Point Color’

Minimum Convex Polygon
A minimum convex polygon is created using the outer position points of a cluster. Workstation enables users to create a regular MCP as well as a Time Enabled MCP, which allows for the observance of an offender’s pattern as it occurs.

Lesson 39: Using a minimum convex polygon to show the overall area related to a bank robbery series.

For this lesson: Examples.mdb | Table: Vegas Bank Robber

1. Open data.
2. Select Google Earth.
3. Check ‘Clear All Checkboxes’.
5. Save Output File as ‘...C: temp\BankRobberMCP.kml’.
6. Select Save.

**Time Slider**

The Time Slider tool allows you to view your points temporally. In Google Earth, you can use a time slider to set the range of dates for your points, show points based on when they occurred, and to play back a timeline of your events. Using the time slider in this case will alter the polygon as you progress.

**Geographic Profiling**

Geographic profiling is an investigative methodology that aims to identify the spatial relation of a serial criminal offender as it relates to the distribution of connected crimes and their anchor points. Workstation enables users to create geographic buffers using the Newton Swoope tool within Google Earth.

**Newton Swoope Buffer**

A Newton-Swoope Buffer is a buffer zone in which the serial offender is unlikely to commit a crime due to it being too closely located to their “haven”. Haven or anchor points can include such things as a residence, work, gym, etc.

A common approach to determine a criminal’s home space is to use the center of all crimes committed. By using the average crime location as a center, the buffer radius is devised from the use of scaled maximum distances of crimes in the x and y directions. This creates a circle centered at the criminal’s haven in which the criminal is unlikely to strike. The Newton Swoope Buffer begins creating buffer zones at the third crime, and continually updates as crimes continue occurring within the set.
Hot Spots
A Hot Spot is a cluster within a spatial distribution. Color, shape, and elevation illustrate the degree of separation amongst crime in a determined area.

Lesson 40: Generating a hot spot map to help demonstrate the pattern of arson’s occurring throughout the D.C. area.

For this lesson: DC Crime

1. Once data is prepared, filter for ‘Arson’ within the UCR column.
2. Select Google Earth from the toolbar.
4. Select ‘Density/HG’ from list on left.
5. Check ‘Extrude Density based on Altitude Field’.
6. Select ‘Altitude’ from list on left.
7. Set Max Height to ‘5’ miles.
8. Select ‘General’ from list on left.
10. Select Save.
Lesson 41: Hot spot map to highlight patterns related to residential burglaries.

For this lesson: DC Crime

1. Once data is prepared, filter for ‘Burglary Residential’.
2. Select Workstation Map from toolbar.
3. Select ‘Identify Hotspots’ from Workstation Map toolbar.
4. Select ‘Select Points’ from Workstation Map toolbar.
5. Create a box around the hotspot.

7. Select Google Earth from toolbar.
8. Check ‘Density Grid’, then select ‘Density/HG’ from list on left.
9. Change the ‘Dense Opacity’ to 55.
10. Save Output File.
11. Select Save.
Choropleth Maps

In this lesson we will be creating a Choropleth map using summarized data. A Choropleth map shows shaded areas based on volumes. Choropleth mapping is very helpful in creating both quantitative and qualitative representations of data. For this map, we will look at military spending worldwide.

Lesson 42: Generating a Choropleth in Google Earth from Workstation

For this lesson: Examples.mdb | Table: GDP Military Spending

1. Open **GDP Military Spending** data.
2. Set Value Field as ‘GDPPercen’.
3. Select Google Earth from the toolbar.
4. Check ‘Create Choropleth’.
5. Select ‘Choropleth’ from list on left.
6. Check ‘Use Values and Not Counts’
7. Set ‘GDPPercen’ as Frequency Variable.
8. Set Boundary Variable to ‘Country’.
9. Set Min Height to 1000 and Max Height to 2000.

11. Select General from list.
Lesson 43: Generating a Choropleth using Non-Summarized Data

For this lesson: Examples.mdb – | Table: US Law Enforcement Agencies

1. Open **US Law Enforcement Agencies** data.
2. Set Value Field as ‘Officers’.
3. Select Google Earth from toolbar.
4. Check ‘Create Choropleth’.
5. Select ‘Choropleth’ from list on left.
6. Set Boundary Variable to ‘StateAbbreviation’.
7. Since this dataset is using non-summarized data, we will leave the Frequency Variable field blank.
8. Set Max Height to 1000.
10. Save file as: ‘... C: temp\UnitedStates.kml.'
Google Earth Exercise

The chief is very interested in analyzing auto thefts throughout Washington, D.C. He is especially interested in the Bentley Bandit and has requested a full report on the crimes committed.

For this exercise: DC Crime

1. Create a map pinpointing each theft committed.
2. In a new map, use the Point to Point function to find any potential patterns with where the vehicles are stolen, in relation to where they are recovered.
3. Create an altitude map, using Property Value as the altitude.
4. Produce a color coded map, using Method of Entry as the key value.
5. Create a Time-Enabled MCP. Explain what patterns, if any, evolve as you incorporate the time slider.
6. Next, generate a map displaying a Newton Swoope buffer with a Time Slider and explain why buffer zones change as more crimes are committed.
7. Using the hot spot function, portray the Bentley Bandit’s patterns.
8. Create a choropleth map, using beats as a boundary variable. The .kml for this map is _______________.

User Notes:
Workstation Map

Google Earth is a very powerful way to visualize data. However, it is limited to spherical x and y coordinate data (decimal degrees.) Workstation Map provides a way to analyze both spherical and projected data. Projected data is data that has been “flattened” to best represent a certain section of the world. When data is projected, it can then be more easily measured, analyzed and show on 2d surfaces; like a computer screen. Most GIS programs use projected data so many ESRI shapefiles are projected to either State Plane Feet or Universal Transverse Mercator formats. Both can be read in Workstation.

If you have projected data and wish to display the points in Google Earth, you can use Workstation’s “Convert Coordinates” feature to take projected, flattened data and re-project it back into a spherical format.

Workstation Map offers users a fully functional mapping software solution that is linked to, and from, the matrix window, providing a relational capability from spatial-attribute and vice versa.

Mapping layers (.shp files) can be added to raise overall awareness of the spatial factors affecting the data.

The tool suite is fairly self-explanatory and is well documented in the Help file, but one very useful option is the ‘Select Points’ tool which allows users to highlight specific points within the map display that may be related, or of interest and then view them in the matrix.

In this lesson, the user will be shown how to look at a crime series (Sexual Assaults) and look beyond just the attributes. The user will learn how to add data to Workstation Map (both crime data and mapping layers), sequence the records based on date/time and then how to review their data to look for possible trends within the series.

Data:
- Crime data - Examples.mdb – Sexual Assaults
- GIS data – Gothamnew.shp and streets.shp

Process:
- Load the Sexual Assault series into Workstation
- Review the data using attributes visible in the matrix and list any significant commonalities linking the cases together. Look particularly at day of week and note anything that stands out (or not).
- There is no sequence field in the data, so add one.
How to add the sequence:

- **Edit – Add Field - Sequence**

Using the layout tool, move the sequence so it sits before the First Date variable. Review the sequence and sort in Ascending order – what did the sequence tool do?

**Lesson 44: Adding the data to Workstation Map.**

**For this lesson:** Examples.mdb | **Table:** Sexual Assaults

1. Select the Workstation Map icon from the Main button bar.
   The points will now auto load to Workstation Map as shown below.

2. Select the Add Shapefile icon.
3. Navigate to temp/Workstation/GIS Data and add Gothamnew.shp
4. Navigate to temp/Workstation/GIS Data and streets.shp
There are two shapefiles in Workstation Map, but only one is visible (Gothamnew). Gothamnew is above streets and is hiding streets, which are underneath. To bring the streets on top, the user has to move streets above gothamnew.

5. Right click in the Layers area will bring up the preferences window.
   There are four areas in preferences.
   • Layers, selecting and dragging the layer name up/down moves the layer on the map
   • Source
   • Alias allows the user to change the name in the map
   • Point/Line/polygon attributes where the user can change the color, shape, type and width of the layers.

6. Select and drag the streets layer so it is above gothamnew.
7. Select gothamnew and change the alias name to Gotham.
8. Select streets and change the line color to burgundy (deep red).

Looking at the map there appears to be three spatial clusters. Is this relevant and if so, how?

9. From the button bar, click on the Select points button.
10. Select (click, hold down and box) all the events to the right side of the map as shown in blue above.
11. Go to the Workstation matrix window and select the Query Highlighted Options button - Promote Highlighted Records.
All the events that happen in the eastern cluster occur on Saturdays or Sundays and, because we left our other records visible we can see the assaults throughout the week.

What else can you see from the Saturday/Sunday looking at the map?

12. In the Workstation matrix window, select Query Highlighted Records from the highlight options button dropdown.
13. From the menu bar – Utilities – Time Coding Hours – Sequence - OK
14. In the matrix window go to the sequence field and sort in ascending order.
15. Go to Workstation Map and zoom to full extent.
16. Switch off (deselect by clicking in the boxes next to the values) the Gotham and Streets layers.
17. Select the Sequence button.

We can now see there is a directional element to our weekend sexual assaults. To determine if the direction is N-S or S-N, click on the top point and view it in the matrix to determine where it falls within the data. In this case the uppermost event is the earliest event and the movement is from north to south. Could this be seen in the matrix window from the start?

In fact it was visible. Had the user looked at just the weekend events and sorted them in ascending order and looked at the XY values there would have been a trend, but how easy would it have been to see compared to looking at it with the map.

There won’t always be interesting nuggets to find this way, but there will be cases where reviewing the map data alongside the attribute data helps build a broader understanding of events.

**Exercise 1**

**Requirement:** The chief has asked you to analyze the amount of auto thefts that occurred on weekends in September, 2010. He would like you to present a map that lays out the thefts by beat and include a count by beat for these incidents.

Data sources required:
- DC Crimes
- Beats (GIS Layer)
Workstation provides two ways to quickly perform a variety of statistics. Both means are available against any data you can open in Workstation. The quickest way to recover some summary statistics against a selection of data is using the “Summary Statistics” function located on the Secondary Tool Bar. If more advanced or detailed analysis are needed, the Stat Workstation module provides more functionality and flexibility.

**Statistics**

The Statistics module enables you to perform a multitude of statistical functions including:

- Frequency (often used for COMPSTAT)
- Chi-Square
- Goodness of Fit
- Cross-tabulations
- Descriptive (Mean, Median, Mode, Sum, Count, Minimum, Maximum, Standard Deviation of Population and Sample, and Variance)
- Regression (Linear, Logarithmic, Exponential and Power), and Solve for Dependent or Independent variables using Regression

**Frequency**

Workstation will enable you to perform frequencies on one or a combination of your variables. Workstation also provides you the option of including or excluding null values. You can also add a frequency column heading such as “Frequency”, “Count” or “Crimes” specific to the information you are displaying or wish to convey. The Frequency module is an excellent tool for executing queries on Person’s variables Last Name and First Name. This will return those individuals who have been contacted multiple times in your Workstation data and make it easy to identify those individuals whom are routinely contacted by police or who have been repeat victims.

Uses of Frequency in law enforcement include:

- Calculate the most frequent address. In Workstation, you can use Workstation’s ability to combine variables for a unique address if your data is parsed. That is, the street number, directional, name, and type are located in separate fields. In Workstation, simply calculate the frequency against all those variables for a unique address field.
- Find the most frequently contacted person last month for Field Interviews.
- Identify the most frequent beat, crime, and call type.
- Find which officer has the most arrests, traffic stops, and complaints.
Lesson 45: Calculate Statistics – Frequency From Statistics Module

For this lesson: DC Crime

1. Using the DC Crime database, select Statistics from the Analysis menu.
2. In the Statistics module, select the Frequency tab.

1. Under the list of Available Fields, double click UCR to move it to the Frequency Order list. Since we are trying to find the frequency across a wide array of data, we will choose the “UCR” variable. Then will calculate the frequency of each crime type within the database.
2. Click the Execute Statistic button. The results table will appear.
Chi Square

Chi Square is a statistical technique that can determine whether the groupings of cases on one variable are related to the groupings of cases on another variable. Say, for example, we want to know whether boys or girls get into trouble more often in school. The school resource officer runs a simple pivot table and finds these results.

<table>
<thead>
<tr>
<th></th>
<th>Trouble</th>
<th>No Trouble</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>46</td>
<td>71</td>
<td>117</td>
</tr>
<tr>
<td>Girls</td>
<td>37</td>
<td>83</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>83</td>
<td>154</td>
<td>237</td>
</tr>
</tbody>
</table>

This example is relatively straightforward in that we can fairly quickly tell that more boys than girls got into trouble in school. Calculating percentages, we find that 39 percent of boys got into trouble (46 boys got in trouble out of 117 total boys = 39%), as compared with 31 percent of girls (37 girls got in trouble out of 120 total girls = 31%). To reframe the issue, what if we wanted to test the hypothesis that boys get in trouble more often than girls in school? These figures are a good start to examining that hypothesis; however, the figures in the table are only descriptive. To examine the hypothesis, we need to employ a statistical test, the chi-square test.

We cannot reject the null hypothesis and conclude that boys are not significantly more likely to get in trouble in school than girls. The null hypothesis is that the two variables are independent - or, in this particular case that the likelihood of getting in trouble is the same for boys and girls. The alternative hypothesis to be tested is that the likelihood of getting in trouble is not the same for boys and girls.

Note: The chi-square test only tests whether two variables are independent. It cannot address questions of which is greater or lesser. Using the chi-square test, we cannot evaluate directly the hypothesis that boys get in trouble more than girls; rather, the test can only test whether the two variables are independent or not.

Uses of Chi-Square in law enforcement include:

- Calculate if the race of the officer and the race of the traffic offender are independent (Racial profiling? Fair policing?)
- Determine if the day of the week influences the prevalence of juvenile crime
- Find out if response time is related to the size or population of a beat

Using the Statistical Examples.xls file, you will analyze the results from a survey completed to determine adjustment into society for parolees. The survey measured how well a parolee adjusted back into society (unsatisfactory, fair, good, outstanding) and whether or not they went back to their hometown or someplace new (hometown or not hometown). Using Chi Square, we can determine if “it matters” where the parolees chooses to live when released from prison. This would help inform and guide a parole officer’s response to the individual.
Lesson 46: Use Chi-Square to Analyze a Survey

For this lesson: Statistical Examples.xls | Table: FC – Contingency Table

1. **Open** the “FC Survey – Contingency Table” from the Statistical Examples.xls workbook. Only a case number is required to continue. **Specify** the “Case” variable as the Case Number. Click **Open** to proceed.

2. From the Analysis menu on the Hub, select the Statistics menu or click the Statistics button on the toolbar.

3. Click the Chi-Square/Crosstab tab and verify you are on the Contingency Table tab.

4. From the Rows Variable drop list select “Adjustment.” From the Columns Variable drop list select “Residence.” The Significance Level is set at .05; as this represents a fairly typical standard value for Significance in this context.
5. **Click** the **Execute Statistic** button from the toolbar. The report tab is currently active and the results of the Chi-Square analysis are provided. Scroll down through the report until you see the last statement: “based on this you would conclude there is no relationship between adjustment and residence.”

Workstation not only provides the Chi-Square results, but also interprets the results and provides the conclusion! The conclusion provided is that there is no relationship between the two variables. In other words, where a parolee chooses to live (hometown or not hometown) has no relevance upon how well he or she will adjust back into society.

**Exercise 15: Expand on your Use of Chi-Square**

Using the Statistical Examples.xls workbook and the “Accidents – Contingency Table” spreadsheet survey, determine if someone using a cellular phone is more likely to be involved in a traffic accident using the Chi-Square test. How would you report these findings to command staff, an officer, or city council?

**User Notes:**
Crosstab

The cross tab function in Workstation allows for multiple variables to be taken into consideration in a group of data. When there are attributes of multiple records that you wish to compile, a crosstab can sum those results instead of merely counting them. An example would be if you wanted to know a sum of the value of property taken, based on an attribute such as method of entry, regardless of crime type.

**Lesson 47: Use a Crosstab to Tabulate Total Officers by State**

*For this lesson: Examples.mdb | Table: US Law Enforcement Agencies*

1. Open data.
2. Use the filter bar to query for States that begin with the letter “C”
3. Use the filter bar to query for Agencies with more than 500 officers.
4. Open the Statistics module and navigate to the Chi Square/Crosstab tab, and select the Crosstab tab.

5. In the “Columns Variable” box, enter “State Abbreviation”.
6. In the “Rows Variable” box, enter “Officers”.
7. Switch the bottom option to “Sum”, and enter “Officers”.

For this lesson: Examples.mdb | Table: US Law Enforcement Agencies

1. Open data.
2. Use the filter bar to query for States that begin with the letter “C”
3. Use the filter bar to query for Agencies with more than 500 officers.
4. Open the Statistics module and navigate to the Chi Square/Crosstab tab, and select the Crosstab tab.

5. In the “Columns Variable” box, enter “State Abbreviation”.
6. In the “Rows Variable” box, enter “Officers”.
7. Switch the bottom option to “Sum”, and enter “Officers”.

For this lesson: Examples.mdb | Table: US Law Enforcement Agencies
8. Click the **Execute Statistic** button. The results will calculate into a table with total officers for each state starting with “C” listed at the bottom of the table.

Exercise 16: Expand on your Use of Crosstab
Open the Examples.mdb | CAD table and tabulate the number of arrests by call description.

**Regression & Correlation**
Regression analysis examines the relation or strength of a dependent variable (response variable) to specified independent variables. Regression is often one of the more misunderstood statistics in the analyst's repertoire. It is often difficult to calculate and even more difficult to interpret. Workstation provides the means to conduct regression analysis, interprets the results, and, provides a means to solve for regression in an easy to use process.

Uses of regression in law enforcement include:

- Determine if there is a relationship across the United States between population and officer staffing level. If a relationship exists, calculate how many officers your department needs based on your jurisdiction's population.
- Determine if there is a relationship between call volume and arrests. If a relationship exists, calculate how many arrests an officer should have given their call volume last month.

**Lesson 48: Using Regression to predict Officer Volume based on Population**

**For this lesson:** Examples.mdb | **Table:** US Law Enforcement Agencies

1. Open the data.
2. Retrieve All Data.
3. From the **Analysis** menu on the Hub, select the **Statistics** menu or **click** the **Statistics** button on the toolbar.

4. From the Statistics module, click the **Correlation** tab and make certain you are on the **Regression** tab.

![Regression Tab](image)

5. We must now formulate our hypothesis. The Chief believes that the larger the city population will result in more staffed officers for that jurisdiction. The officers feel that there is not a correlation between the two variables. So, who’s right? Let’s first determine if there is a relationship between the two variables.

6. The **Independent variable** is the predictor variable. In this example, we are trying to determine if population predicts a higher officer rate. From the **Independent Variable** drop list, select **CityPopulation**.

7. The **Dependent variable** is the variable being predicted. We want to predict the amount of officers based on population. From the **Dependent Variable** drop list, select **Officers**.

![Dependent Variable](image)

8. The **Significance Level** is set at .05; we’ll leave that set for now as that is a good level for statistics such as these. We will also leave the regression type as **Linear**.
9. **Click** the **Execute Statistic** button from the toolbar. The report tab is currently active and the results of the Regression analysis are provided. Scroll through the report and note the “Literal Interpretation.”

The literal interpretation is the literal explanation of the R Squared score. The closer the “r” score is to 1, the stronger the relationship between the two variables. In our example, the score is .66, which is considered a moderate positive correlation. In other words, as city population's increase, the amounts of staffed officers moderately increase.

**Goodness of Fit**

The Goodness of Fit (GOF) test is a very widely used statistic. The purpose of GOF is to compare an observed distribution to the expected distribution. For instance, suppose we conducted a national survey and compiled, per State, the average salary of an entry level police officer. We could then use the GOF test to determine if there is a statistical difference between State salaries. Our hypothesis would be that we expect each state to be roughly equal in terms of starting salary. GOF would enable us to assess whether this was true or not. Another example might be in assessing officer’s workload through your jurisdiction. We would hope that each officer is working equally as hard having to respond to call and take reports. Hence, we would expect similar numbers of reports to be taken by each officer with a slight variation from each officer. What we would want to determine is if the differences in the number of reports each officer is taking is statistical and significantly different. If they were, as show through GOF, then we might know that the assignment or allocation of those officers may need to be reviewed to make a more equitable distribution of work.
Lesson 49: Use Goodness of Fit to Assess Workload

For this lesson: Statistical Examples.xls | Table: Officer Salaries -GOF

1. Open the “Officer Salaries – GOF” table from the Statistical Examples.xls workbook.
2. Specify the “State” variable as the Case Number. Click Open to proceed. Once the data is open, you’ll see each State and the average hourly starting wage of a police officer.
3. From the Analysis menu on the Hub, select the Statistics menu or click the Statistics button on the toolbar.
4. Click the “Chi-Square/Crosstab” tab and select the “Goodness of Fit” tab.
5. Select “Hourly Wage” from the Observed Variable drop list. Specify a .05 significant level and check the “Use Values and Not Frequency” checkbox as we have already aggregated our results.
6. Click the Execute Statistic button from the toolbar. The report tab is currently active and the results of the Goodness of Fit analysis are provided. The Goodness of Fit test enabled us to determine that there is a statistical difference between States in regards to a starting hourly range for each officer.
Threshold Analysis
Threshold analysis is used to find the normal or expected volume for specific incidents, in a specific area, in a specific time frame and how the current activity compares to what is “normal.” Threshold analysis mathematically compares data to determine if the volume is up or down and calculates the significance of those changes.

Lesson 50: Use Threshold Analysis to determine volume of incidents for total timeframe

For this lesson: Sample.atx

Use Threshold analysis with Law Enforcement data to look at data over the past 11 months and compare the 12th month to those 11 months.
• If the crimes are high for the 12th month, this data would then need to be looked at further to decide why they are higher.
• If the 12th month is “normal” then the incidents have stayed the same through the 11 months and compared to the 12th month.
• If the 12th month is low, some resource allocation, deployments and predictive policing may have had a positive impact to bring down the total numbers of that specific incident.

1. Using the Sample.atx database, select Statistics from the Analysis menu.
2. In the Statistics module, select the Threshold tab.
3. Identify the **Main Variable** which can be any field within your database. This main variable is the field that will be examined over the specified timeframe to see if the compare field meets the expected volume. For this lesson, we will use “Crime” as the main variable.

4. Identify the **Timeframe for Threshold Analysis Baseline** using the calendar. Knowing the total date range of your data is important. Use this information to then identify the timeframe for the threshold analysis. For this lesson the timeframe will be 1/1/2006 to 11/30/2006.

5. Identify the **Baseline/Interval** field which will be used to calculate the timeframe. For this lesson, we will use “Monthly.”

6. Identify the **Compare** field by selecting the correct number of weeks, months, etc. This compare field will look at the main variable of data over the timeframe using the baseline data to show whether or not this compare field is high, low or normal compared to the average number of incidents in the data. For this lesson, we will choose the 12th month.

7. Select Execute and the threshold analysis for the data will update the report, grid and graph tabs. The report and grid tabs will show the statistics used for the threshold and what incidents of the main variable field are low, high or normal. As you can see below, Criminal Trespass and Public Sexual Indecency are “Very High” compared to the previous 11 months. These are two crime types the agency should look at to determine why they have a higher count for the 12th month and can use this to put in a plan for resource allocation, predictive policing and deployments.
Threshold Analysis:

<table>
<thead>
<tr>
<th>Crime Category</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Normal Range</th>
<th>13th Month</th>
<th>% Change</th>
<th>Z-Score</th>
<th>Activity Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>127.4</td>
<td>15.78</td>
<td>15</td>
<td>100</td>
<td>6%</td>
<td>0.52</td>
<td>Normal</td>
</tr>
<tr>
<td>Residential Burglary</td>
<td>3.7</td>
<td>9.1</td>
<td>3</td>
<td>100</td>
<td>15%</td>
<td>0.52</td>
<td>Very Low</td>
</tr>
<tr>
<td>Commercial Burglary</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>100</td>
<td>10%</td>
<td>0.52</td>
<td>Low</td>
</tr>
<tr>
<td>Motor Vehicle</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>100</td>
<td>10%</td>
<td>0.52</td>
<td>Low</td>
</tr>
<tr>
<td>Public Order</td>
<td>0.1</td>
<td>0.5</td>
<td>0</td>
<td>100</td>
<td>10%</td>
<td>0.52</td>
<td>Normal</td>
</tr>
<tr>
<td>Assault</td>
<td>5.5</td>
<td>5.5</td>
<td>5</td>
<td>100</td>
<td>0%</td>
<td>0.52</td>
<td>Normal</td>
</tr>
</tbody>
</table>

*Activity level colors. Red = Very High Activity Level, Orange = Moderately High Activity Level, Light Blue = Moderately Low Activity Level, Blue = Low Activity Level and no color means the activity level is normal. The activity level is based off of the Z-Score.*
Workstation provides numerous, full-color reports which can be printed or exported to either Rich Text Format (rtf), HTML, or Adobe Acrobat Format (pdf).

Workstation also provides XML based reporting for expansive reporting functions and options. Workstation’s XML based reports can be modified using existing XML editing software. Workstation’s XML reports can even be exported to Text and a variety of HTML file options.

**The Reporting Module**

The Reporting Module will allow you to set the additional display parameters for your report. From this module, you will have control over the logo and badge used, security phrase displayed, report’s main heading, sub heading, details heading, left and center footer, display options and printer settings. Additionally, you can set a detailed narrative or synopsis for the report.

The Reports Module allows for modification to most of the reporting options contained on your report. The sample bulletin on the right side of the form provides a guide as to where elements will be displayed as well as allows for modification to the imagery. If you change the logo or badge the changes will only take effect until you close the Reports Module. The changes will not overwrite your default logo or badge settings in the User Accounts Module.

Let’s use the **Sample.atx** database for subsequent examples. Reopen the **Sample.atx** database and select IR Number ATACPD06001.

**Lesson 51: Create a Report**

For this lesson: Sample.atx

1. Open the **Sample.atx** database and have only the **WORKSTATIONPD06001** record selected. From the Hub, click the “Reporting Module” button located on the toolbar or select Reporting Module from the Reports menu.

2. Select the “**Wanted Poster.xml**” report from the “**Reports**” folder.
3. Type the following information into the various combo boxes and text boxes from this module:
   
   a. **Security Phrase:** Law Enforcement Use Only
   
   b. **Main Heading:** Wanted Poster
   
   c. **Sub Heading:** Dead or Alive
   
   d. **Details Heading:** Usama Bin Laden

4. From the Header Images tab, Double-click the area beneath the Left Image label. Navigate to your Workstation path to and select the *ShieldBadge.bmp* file therein.

5. Double-click the area beneath the Right Image label. Navigate to your Workstation path to and select the *Logo.bmp* file therein.

You can also type a brief summary for this report by clicking the **Summary** tab and entering your text in the box provided. Various font attributes and images can be included in this area. The Report Synopsis is also capable of using Rich Text Formatted text like those developed using Microsoft Word. For this bulletin, enter: “*Wanted in connection with the 9/11 terrorist attack on the World Trade Center.*”

At this point you could either print the report or use the export features under the Report menu to export this report to an RTF file for later editing and enhancing in a word processing or publishing application, or, export the report into a PDF (Portable Document Format) to World Wide Web distribution or as an attachment to an email.

---

**Exercise 25: Expand on Reporting**

Use the **Sample.atx** database and the Query Wizard; select only those individuals where a picture is available.

Run the Six Person Photo Lineup report.

Use the File menu and export the report to a PDF document.

Pull a Histogram and Tempogram into your reporting module and then export them.
A Trend is simply a way of putting a marker against a number of records that have been identified as likely being related. They remain within the larger dataset, allowing them to be considered alongside the overall pool of events, but are marked, making it easy to concentrate on them, or highlight possible relationships with other records. Trends can be allocated discrete colors to make visual discrimination easy. Trends are not queries; users identify events that fit the parameters of a particular Trend and then append them to the trend.

Trends are collected under a title, but users are encouraged to apply key words to help easily identify why records have been grouped. Records can easily be added and removed from the Trend, providing the analyst with the option to err on the side of caution and add possibly related records, rather than discarding them because of the additional workload associated with managing the data.

While many users generate Trends based on specific crime types/series, they can be used to highlight by geography too. All crimes in a given zip code, beat etc. could be put into a trend and then color coded within the larger dataset. Records can be attributed to more than one trend and the user will be able to highlight the records associated with trends, but a record can only be allocated one color and the trend name visible in the matrix can only be one. The last trend a record is attributed to, will be the color and name users will see in the matrix.

By selecting Trends from the hub list, users are presented a number of control options for Trends. For color cueing purposes, control is via the dropdown button highlighted below.

The available Trend Color Cueing functions are listed below:

**Color Cue All Trends**
Visually “marks” all rows in the current query using all Trends color cues.

**Color Cue Selected Trends**
Visually “marks” all rows associated with a specified Trend.

**Color Cue All Trend Keywords**
Visually “marks” all cells in the current query for cells containing any text stored in the “Keywords” variable for ALL Trends using that Trend’s color cue.

**Color Cue Selected Trends Keywords**
Visually “marks” all cells in the current query that contain “Keywords” listed in a specified Trend. One or multiple Trends can be selected.
Lesson 52: Trend Management

Workstation's Trend Manager provides the means to store and track the Trends that have been identified in a Workstation database or other databases. The left half of the Trend Manager allows the user to enter and display information relevant to the entire collection of cases grouped into a trend. It shouldn’t be used to refer to any specific case or feature of a case. Analysts should first compile a Trend; later they may append Case information onto it. The right half of the Trend Manager is shared between Case and Offender information.

The default location for the Trend information is the Trends.mdb database located in the directory where Workstation is installed.

Users can create additional databases to store archived trends, trends for individual years, units, departments, etc. The trends database’s name and location in current use will be displayed on the Trend Manager’s status bar.

Describing Trends
Trend Number

The navigation bar located at the top of the screen shows the current trend number where you are positioned. You can scroll through the list of trends by using the navigation arrows. Notice that as you do so, the list of cases on the right side changes to reflect the currently selected trend.

Trend Name

A unique title by which the trend may be referenced, such as: “Downtown Purse Snatches”, “Mobile Home Rapes”, “Bay Area Cat Burglar”, or “Streetwalker Homicides”.

Category

A general description of the trend which usually reflects the predominant crime involved (e.g. “Murder”). Choose the Category which most fully describes the nature of your trend. If a broad variety of crimes are involved, it is recommended analysts use the Category describing the worst crime in the group. The Uniform Crime Report is an excellent standard to use.

Type

Refers to the nature of the group of cases. There are five standard choices:

Series

A group of cases which share the same causality; usually, this means that the same person or group of persons is committing the crimes (e.g., a serial child molester who abducts children from downtown parks by asking them to help look for a lost puppy.).

Trend

A collection of cases which share specific characteristics, such that a strategy for dealing with one could be used to deal with others in the same group, but which do not necessarily share the same causality (e.g., fifteen cases of cars stolen from hotel valet parking areas).

Pattern

A group of cases sharing one or more traits, but that are not related to one another in a significant manner (e.g., dog-door residential burglaries).

Collection

Any group of cases the analyst has grouped together for further study. They may be unrelated to one another by crime type, but be within a confined geographic area, such as a business district.
Spree
Groups of cases sharing the same causality, but which occur so close to one another in time, space, and causality that they can be viewed as multiple actions within a single crime. For example, the theft of 8-vintage cars overnight in a city when only one has been stolen in the previous three years.

Trend Status
The Trend status is used to describe the current situation of the trend. There are four standard classifications:
- **Active** - The trend is ongoing.
- **Inactive** – While the perpetrators have not been arrested, the trend is no longer visible; it may reassert itself in the future.
- **Solved** - The trend has been successfully concluded.
- **Exceptionally Cleared** - The case lacks sufficient evidence to effectively pursue any further and is closed pending further information.

Keywords
Keywords are phrases, whole or partial words that have been attributed to a trend. If keywords are separated by commas, Workstation can later identify and mark those words anywhere in the matrix using the Trend’s unique Trend Color.

Trend Color
Trend color allows users to apply a unique color to each discrete trend, making it easier to distinguish the trend from other events.

User ID
The user ID number identifies the analyst responsible for adding the trend to the database.

Identified By
Describes the primary person responsible for identifying the trend. There are several standard options including “Detective,” “Patrol,” “Analyst,” and “Public” available via the drop down, but users can manually add others if needed. Choose the method that most accurately describes how your trend first came to light.

Date Identified
The date the user created the trend.
Lesson 53: Create a New Trend

In this lesson, we will create a new Trend called ‘Auto Theft Collection’.

1. Open data.
2. Select ‘Trends’ from the hub.
3. Use the Trend Manager button to launch Trend Manager.
4. Click the Add New Trend button.
5. Type **Auto Theft Collection** in the Trend’s name field.

6. For the remaining fields enter the following:
7. Click the **Save This Trend** button to complete the process of building a new trend.

**Assigning Cases to a Trend**

Records highlighted within the matrix can be assigned/appended to a trend. The case records will reflect the name of the last trend they were assigned to in the trend variable within the matrix. While not the norm, cases are occasionally attributed to more than one trend. When cases are attributed to more than one trend, the trend variable within the matrix will reflect the last trend it was attributed to; the record will still be assigned to any other trends it has been linked to, but won’t be as easily visible via the matrix. Likewise, each case (IR Number) will now be associated with a particular Trend in the Trend Manager. Once a case has been assigned to a trend, Trend Manager will auto populate many of the fields in the Case information text boxes including concatenating several of the MO fields together for inclusion into the Case Notes field.

**Lesson 54: Assign Cases to a New Trend**

**For this lesson: DC Crime**

In this lesson, we will select the ‘Bentley Bandit’ cases from the DC Crime database, using the Filter bar query, and assign all of them to our newly created ‘Auto Theft Collection’ trend.

1. Open data.
2. Scroll to the Trend variable and use the Filter Bar to select the ‘Bentley Bandit’ Series.
3. Select Trends from the hub.
4. Open the Trend Manager.
5. If the Trends are not visible, Open Trend Database and navigate to the Program Files - Workstation and select Sample Trends - Open.
6. Scroll through (using the left/right arrow button) to find the ‘Auto Theft Collection’ trend.

Another way to open Trend Manager and find a trend of interest is to double click on the trend name in the hub.
7. From the Trend Manager, Click the Assign Queried Records to Current Trend button.

8. Workstation will confirm that you wish to add your cases to the current Trend - Yes.

9. Review the Trend Events (20 records).

10. Step through the events and notice how Workstation has captured relevant fields and placed them in the event window for review.

Note: Some analysts are more comfortable working in a matrix-oriented display environment. Sometimes the ability to view and compare many trends or cases to one another simultaneously is highly desirable. To accommodate this need, the Trend Manager allows Grid views of both Trends and Cases. Users can change to Grid mode by clicking the Matrix tabs. These will invoke a matrix view of your Trends and/or Cases. Try clicking on the various tabs in various combinations and selecting from different trends. Notice how the Case information is updated and displayed when a new Trend is given the focus.

**Trend Reporting**

As users continue to identify new, or are adding to existing trends, they will want to begin to analyze interaction, overlap, types, categories, etc. Workstation provides users several tools designed to enable reporting and analysis of trends.
The first tool is the **Timeline Chart**. The Timeline Chart generates a chart illustrating the duration of each trend. The chart shows all trends in the chart, allowing users to easily detect potential temporal relationships between trends. While the default view for the Timeline Chart is all trends, users can tailor which trends are visible.

The chart can be exported for inclusion in reports as necessary.
You have just completed lessons for some of the most useful and widely used functions of Workstation. While we have covered the majority of the functionality, we have barely begun to cover how you might use the analytic options within your own agency. Additionally, there are other tools within Workstation that have not been covered in this workbook/class due to limitations of time.

The goal of this workbook/class was to provide you with the basic skills needed to return to your own workspace and understand how and where Workstation will be able to assist you in your day-to-day analysis workflows and extend into areas previously untested due to time/tool constraints.

While the class is over, the hope is that this workbook will continue to act as a guide and help refresh your memory.
Tactical Analysis using Accurint® Crime Analysis Workstation

**Bottom Toolbar**

- Revert to Original Layout
- Layout Organizer
- autosize columns
- Column Justification
- Delete Highlighted Column
- Copy Selected Cells to Clipboard
- Display the Data in Grid Mode
- Display the Data in Form Mode
- Primary View
- Relational View

**Queries / Searches**

- Filter Bar Search
- Kwic Expression
- Concept Query
- SQL Viewer
- Visual Query
- Proximity Query
- Query Wizard
- Color cue by...
- Omni Search
Analysis related

- Time Series Analysis
- Statistics
- Perform statistics on highlighted column
- Create Archetype from selected records
- Trend Hunter
- Time Coding

Layout related

- Revert to original configuration
- Layout Organizer
- Autosize columns to display values
- Pre-configured layout options

Layouts

- Column Arrangement for Burglary: Commercial
- Column Arrangement for Burglary: Residential
- Column Arrangement for Drunk and Drunk
- Column Arrangement for Robbery
- Column Arrangement for Sexual Assault
- Column Arrangement for Stabbing
- Column Arrangement for Suicide/Other
- Column Arrangement for Vehicle Information
About LexisNexis Risk Solutions

At LexisNexis Risk Solutions, we believe in the power of data and advanced analytics for better risk management. With over 40 years of expertise, we are the trusted data analytics provider for organizations seeking actionable insights to manage risks and improve results while upholding the highest standards for security and privacy. Headquartered in metro Atlanta, LexisNexis Risk Solutions serves customers in more than 100 countries and is part of RELX Group plc, a world-leading provider of information and analytics for professional and business customers across industries. For more information, please visit www.lexisnexisrisk.com.

For more information, go to
http://www.accurint.com/workstation/techsupport.html