# Appendix A: Data Sources

## Overview

Almost all the data tables used for chapter examples and for Application Scenarios contain real data. This appendix provides the citations and initial source of each table. Most are from public-domain Web sites, and those that are not also cite the original author or compiler who granted permission for the data table to be used in this book. Those generous individuals are also listed in About this Book. The bibliography includes citations of publications based on datasets as relevant. Items in boldface have been updated since the Second Edition.

The University of California at Irvine Knowledge Discovery in Databases (KDD) Archive has proved an especially valuable source of data (http://kdd.ics.uci.edu), as noted in the table beginning on the next page. Readers are encouraged to visit the KDD site for additional data tables. Similarly, the Data and Story Library (lib.stat.cmu.edu/DASL), maintained at Carnegie Mellon University, is a very rich source of data sets.

## Data Tables and Sources

| Table Name | Source | URL |
|---|---|---|
| **Airline delays** | US Department of Transportation, Bureau of Transportation Statistics. | https://www.transtats.bts.gov/DL_SelectFields.asp |
| **Airline delays 2** | US Department of Transportation, Bureau of Transportation Statistics. | https://www.transtats.bts.gov/DL_SelectFields.asp |
| **Airline delays 3** | US Department of Transportation, Bureau of Transportation Statistics. | https://www.transtats.bts.gov/DL_SelectFields.asp |
| Anscombe | JMP sample data tables. | |
| Binge | Based on tabulations reported by R.W Hingson, E.M. Edwards, T. Heeren, and D. Rosenblum | |

| Table Name | Source | URL |
|---|---|---|
| | (2009). | |
| **Birthrate 2017** | United Nations and World Bank | http://unstats.un.org/unsd/default.htm |
| **Cell Subscribers** | United Nations Millennium Development Goals Indicators. | http://unstats.un.org/unsd/mdg/Default.aspx |
| Concrete | Prof. I-Cheng Yeh (cf. Bibliography) | http://www.sigkdd.org/kddcup/index.php |
| Concrete 28 | Subset of Concrete table. | |
| Diameter | JMP sample data tables. | |
| Dolphins | JMP sample data tables. | |
| **Earthquakes** | U.S. Geologic Survey. | http://neic.usgs.gov/neis/epic/epic_global.html |
| **FAA Bird Strikes** | Federal Aviation Administration. National Wildlife Strike Database. | https://wildlife.faa.gov/databaseSearch.aspx |
| **FAA Bird Strikes CA** | Subset of FAA Bird Strikes. | |
| **Fertility** | United Nations DESA/Population Division | https://population.un.org/wpp/Download/Standard/Fertility/ |
| **Global Temperature and CO2** | Earth Policy Institute, NASA and NOAA. | http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts+dSST.txt and www.esrl.noaa.gov/gmd/ccgg/trends/co2_data_mlo.html |
| Gosset's Corn | Based on tables in W.S. Gosset, "The Probable Error of a Mean," Biometrika, 6 (1908). | |
| Haydn | Subset of Sonatas data. | |
| Hubway | Hubway and Metropolitan Area Planning Council, Boston, MA. | http://hubwaydatachallenge.org/trip-history-data/ |
| **India Industrial Production** | Government of India Ministry of Statistics and Programme Implementation. | http://mospi.nic.in/iip |
| Insulation | JMP Software: ANOVA and Regression course notes, SAS Institute. Used with permission. | |

| Table Name | Source | URL |
|---|---|---|
| **Life Expectancy** | World Development Indicators. | http://data.worldbank.org/data-catalog/world-development-indicators |
| **Life Expectancy 2017** | World Development Indicators. | http://data.worldbank.org/data-catalog/world-development-indicators |
| Maine SW | George Aronson. Used with permission. | |
| **MCD** | Yahoo! Finance. | http://www.finance.yahoo.com |
| Michelson 1879 | Data and Story Library; based on table in S.M. Stigler (1977). | lib.stat.cmu.edu/DASL |
| Military | Diez et. al, OpenIntro Stats; U.S. Department of Defense. | http://www.openintro.org/stat/extras.php and data.gov |
| Mozart | Subset of Sonatas data. | |
| NC Births | Diez et. al, OpenIntro Stats; North Carolina State Center for Health Statistics. | http://www.openintro.org/stat/extras.php |
| **NHANES 2016** | National Health and Nutrition Examination Survey (20015-2016). | http://www.cdc.gov/nchs/nhanes.htm |
| **NIKKEI225** (replaces FTSE100) | Yahoo! Finance. | http://www.finance.yahoo.com |
| Parkinsons | Prof. Max Little (cf. Bibliography) | http://www.sigkdd.org/kddcup/index.php |
| **Pipeline Safety** | US Dept of Transportation, Pipeline and Hazardous Materials Safety Program (PHMSA) Pipeline Safety Program. | https://www.phmsa.dot.gov/data-and-statistics/pipeline/pipeline-incident-flagged-files |
| Planets | Smithsonian Institution National Air and Space Museum. | http://www.nasm.si.edu/etp/ss/ss_planet data.html |
| Popcorn | Artificial data inspired by experimental data from Box, Hunter, and Hunter (1978). Factors are brand (popcorn), amount of oil (oil amt), and batch size (batch). | |
| **SDG_Technolog** | United Nations Dept. of | https://unstats.un.org/sdgs/indicators/dat |

| Table Name | Source | URL |
|---|---|---|
| **y** (replaces MDG Technology) | Economic and Social Affairs, Statistics Division | abase |
| **SDGCO2** (replaces MDGCO2) | United Nations Dept. of Economic and Social Affairs, Statistics Division | https://unstats.un.org/sdgs/indicators/database |
| Sleeping Animals | JMP sample data tables. | |
| Sonatas | Prof. Jesper Rydén (see Bibliography). | |
| **States** | Demographic and economic data about the 50 U.S. states and the District of Columbia, compiled from several sources | http://quickfacts.census.gov/qfd/index.html<br>http://www.infochimps.com<br>http://www-fars.nhtsa.dot.gov/<br>http://www.bls.gov/web/laus/laumstrk.htm |
| Stock Weekly Index | Yahoo! Finance. | http://www.finance.yahoo.com |
| **TimeUse** | Extract from the 2003, 2007, and 2017 American Time Use Survey. | http://www,atus.org |
| Tobacco Use | United Nations. | http://data.un.org/Browse.aspx?d=WHO&f=inID:RF17 |
| **USA Counties** | US Census Bureau. | www.census.gov |
| **WDI** | World Development Indicators. | http://data.worldbank.org/data-catalog/world-development-indicators |
| **World Nations** | World Development Indicators. | http://data.worldbank.org/data-catalog/world-development-indicators |

# Appendix B: Data Management

## Overview

Before any data analysis can be undertaken in JMP, one must make the data accessible to JMP either in the form of one or more JMP data tables, or at least in a format that JMP can read. Depending on the goals of a study, it is often desirable to integrate data from multiple sources into a single data table. This appendix presents some foundational ideas of data management, including techniques for entering data from a keyboard, reading data from an Excel worksheet, and transferring data from Internet sites. Additionally, it introduces some concepts and methods for combining data from two or more tables into a single table.

## Entering Data from the Keyboard

Many readers are using this book in conjunction with a college course in statistics for which they have a principal textbook. Suppose that one such reader encounters a textbook problem based on the following small data set[1] from an online music seller that advertises at sporting events.

| Name | Age (yr) | Time Since Last Purchase (Days) | Area Code | Nearest Stadium | Internet Purchase? |
|------|----------|--------------------------------|-----------|-----------------|--------------------|
| Katharine H. | | 130 | 312 | Wrigley | Y |
| Samuel P. | 24 | 18 | 305 | Orange | N |
| Chris G. | 43 | 368 | 610 | Veterans | Y |

| Name | Age (yr) | Time Since Last Purchase (Days) | Area Code | Nearest Stadium | Internet Purchase? |
|------|----------|--------------------------------|-----------|-----------------|--------------------|
| Monique D. | | 5 | 413 | Fenway | Y |

Even with a small data set, we need a plan before we start tossing data into a table. This is a cross-sectional sample of customers containing both character and numeric data. "Name," "Nearest Stadium," and "Internet Purchase" are clearly categorical data, and "Age" and "Time Since Last Purchase" are continuous. What about "Area Code"? The digits are numeric, but the codes have no numerical meaning. Area codes are categorical labels identifying a geographic region.

We also have a choice to make about the "Internet Purchase" column. We can enter the data as Ys and Ns, as shown above. Alternatively, we can follow the NHANES example and use numbers to represent Yes and No. Either method is acceptable; for the sake of learning about column properties, we will use numeric codes.

1. Select **File ▶ New ▶ Data Table**. You will see an untitled blank table, as shown in Figure B.1. Initially, a JMP data table contains one empty numeric column.

**Figure B.1: A New Data Table**
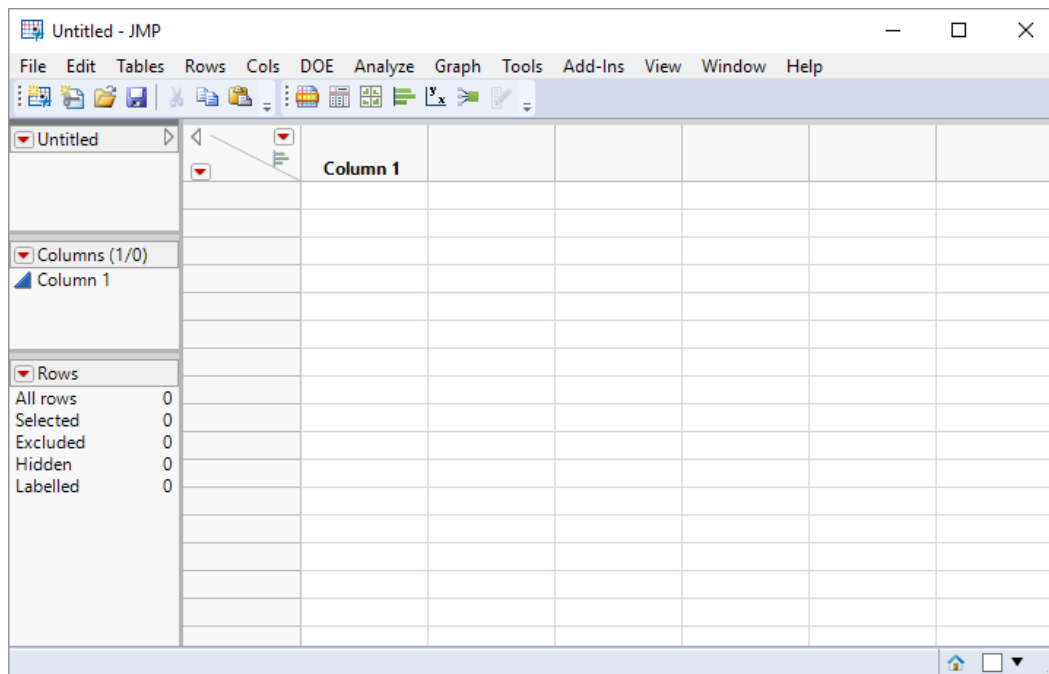


2. Before entering any data, let us name and document this data table. Click on the word **Untitled** in the **Table** panel and change the name of this table to **Customers**.

3. Click the red triangle in the upper left next to **Customers** and choose **New Table Variable**. A simple dialog box opens (not shown here). Next to **Name**, type **Background** and next to **Value,** type **Textbook example**. Then click **OK**.

We will first set up six columns, name them, and assign appropriate properties to each column in anticipation of typing in our data.

4. Move your cursor to the top of the first column over the cell labelled **Column 1** and right click once to bring up the menu shown in Figure B.2.

**Figure B.2: Context-Sensitive Column Menu**
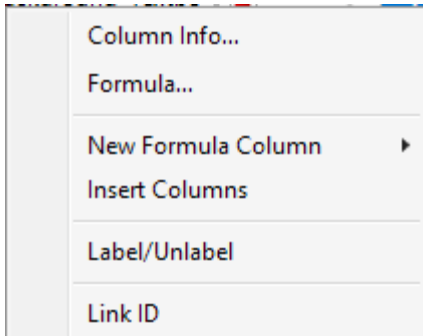


5. Choose **Column Info** from this menu to name the column and specify the properties that we want to set.

In this dialog, we specify the name, data type (numeric, character or row state), modeling type (continuous, ordinal or nominal), format, initial data values, and number of rows. We also have the option to select from many other column properties.

6. Complete the dialog box as shown below in Figure B.3. Click on the **Column Properties** button to add the **Notes** field to the column information. When you are done, click **OK**.

**Figure B.3: Entering Column Information**
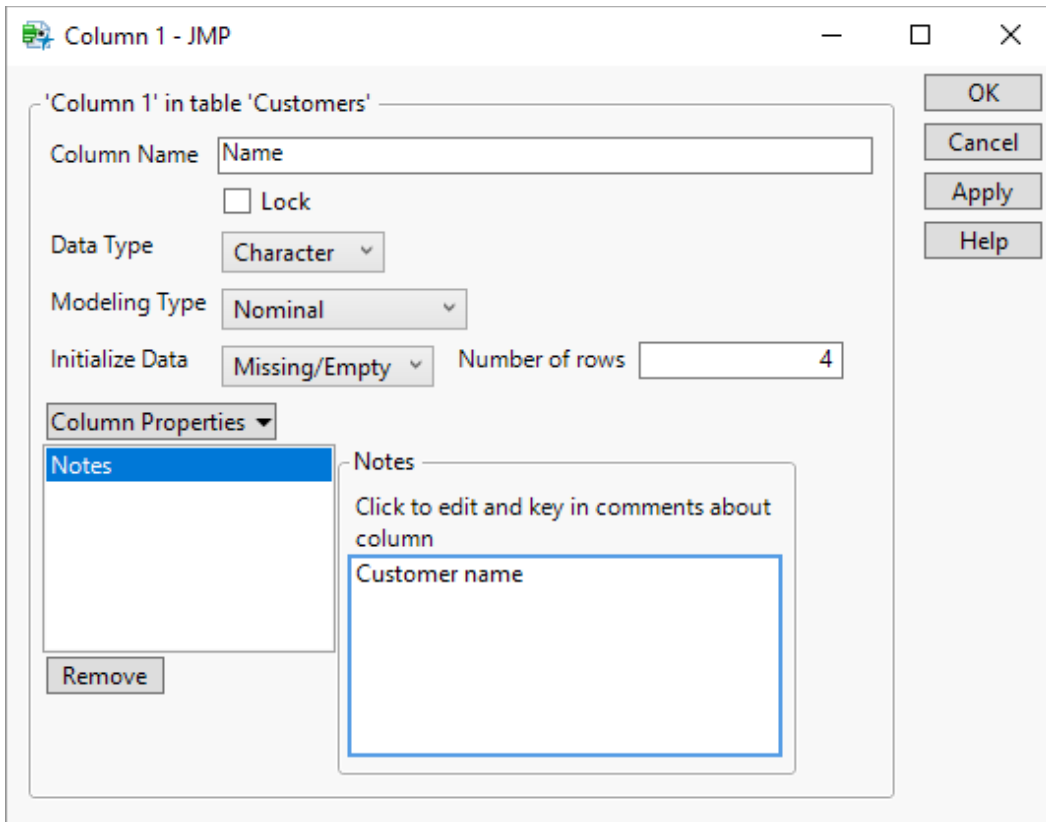


7. Now choose **Cols ▶ New Column** to add a second column for customer ages. Name the column **Age** and specify that it will contain numeric data. **Modeling Type** will be **Continuous**. Add a note to say that ages are in years and click **OK**.

> You are about to create five more columns of different data types. To save time double click at the top of the next blank column. This reserves the column for data entry and labels it as **Column 2**. Double click again at the top of the column and the **Column Info** dialog box opens.
>
> Alternatively, you can also create 5 new columns through the **New Column** dialog box, and then proceed to enter unique names and properties one at a time.

Continue to create new columns until we have six columns in all, with the following names and attributes:

| Col | Column Name | Data Type | Modeling Type | Note |
|-----|-------------|-----------|---------------|------|
| 1 | Name | Character | Nominal | Customer name |

| Col | Column Name | Data Type | Modeling Type | Note |
|-----|-------------|-----------|---------------|------|
| 2 | Age | Numeric | Continuous | Customer age in years |
| 3 | LastPurch | Numeric | Continuous | Days since last purchase |
| 4 | AreaCode | Numeric | Nominal | Area code |
| 5 | Stadium | Character | Nominal | Nearest stadium |
| 6 | Internet | Numeric | Nominal | Internet purchase? |

8. In the "Internet" column, we will establish a coding of 0 for No and 1 for Yes by choosing **Value Labels** from the column properties drop-down list.

9. Within the value labels portion of the dialog box (shown in Figure B.4), indicate that values of 0 should be labeled No and then click **Add**. Then associate the Yes label with values of 1 and click **OK**.

**Figure B.4: Specifying Value Labels**



10. With the columns fully defined, type the data into the cells one column or row at a time. See Figure B.5 for reference. For Katherine and Monique, just skip past their "Age" cells. In the "Internet" column, enter a 0 in row 2 and 1 in each of the other rows.

Figure B.5 shows the completed **Data Table** window for this example. When you have finished entering all the data, screen should look much like the one shown in Figure B.5. Having done the work to type in this set of data, you should save it.

**Figure B.5: The Completed Table for Customer Data**



11. Select **File ▶ Save As**. By default, JMP wants to save this file in the JMP data table format. Select a proper location (folder) in which to save it, and this new data table will be preserved under the name **Customers**.

# Moving Data from Excel Files into a JMP Data Table

There is a very good chance that you will want to begin an analysis with some data that you already have on site. Perhaps it is in a pdf format, a text file, or within an Excel workbook file. Because Microsoft Excel© is so prevalent, we will illustrate the import operation using Excel. There are several ways to transport data from Excel and other formats into JMP; Chapter 3 "Import Your Data" in *Using JMP* (**Help ▶ JMP Documentation Library ▶ Using JMP**) covers this topic in depth, but in this section, we will see two approaches.

## Importing an Excel File from JMP

This method works most simply if the Excel spreadsheet is already structured like a JMP data table: each column of the worksheet must represent a single variable, and each row should

represent the data from one case or observation. It is best if the top row is reserved for column names and the observations start in row 2. Initially We will assume there is just a single sheet within the workbook file, and that there are no merged cells. JMP can open multiple sheets in a workbook, but here we illustrate the simplest case.

Finally, it is important to understand that if there are formulas or functions in the Excel file, they will not be imported into JMP. JMP will read the *results* of a formula into a cell. If Column C of the spreadsheet equals the sum of Columns A and B, JMP will read in the sums but not the formula.

To illustrate this method, let us suppose we have the same customer data in a spreadsheet called **Customers.xlsx**. You should experiment in this section with any spreadsheet you already have; for this example, I have constructed an Excel worksheet containing the same data shown earlier.

Be aware that when you import from a spreadsheet file, JMP has no way of knowing whether some numeric columns are intended as nominal data (such as area codes); so, you will need to proofread the initial settings for the columns. You will also need to add any column notes or data value coding schemes that you want to establish.

1. Select **File ▶ Open…**. Within this dialog box, first navigate to the folder containing your properly constructed spreadsheet.
2. Find the file types drop down menu in the lower right of the dialog box. Click on the button and notice the many file types that JMP can read, including those from database and other statistical packages, as well as pdf format files. Select **Excel Files (*.xls, *.xlsx, *.xlsm)**. You will then see the list of all Excel files in the directory. Choose the file that you want and click **Open**.
3. This opens the **Excel Import Wizard**, as shown in Figure B.6. The Wizard allows considerable flexibility in working with multi-sheet workbooks, as well as with worksheets that have somewhat unusual layouts.

**Figure B.6: The Excel Import Wizard**
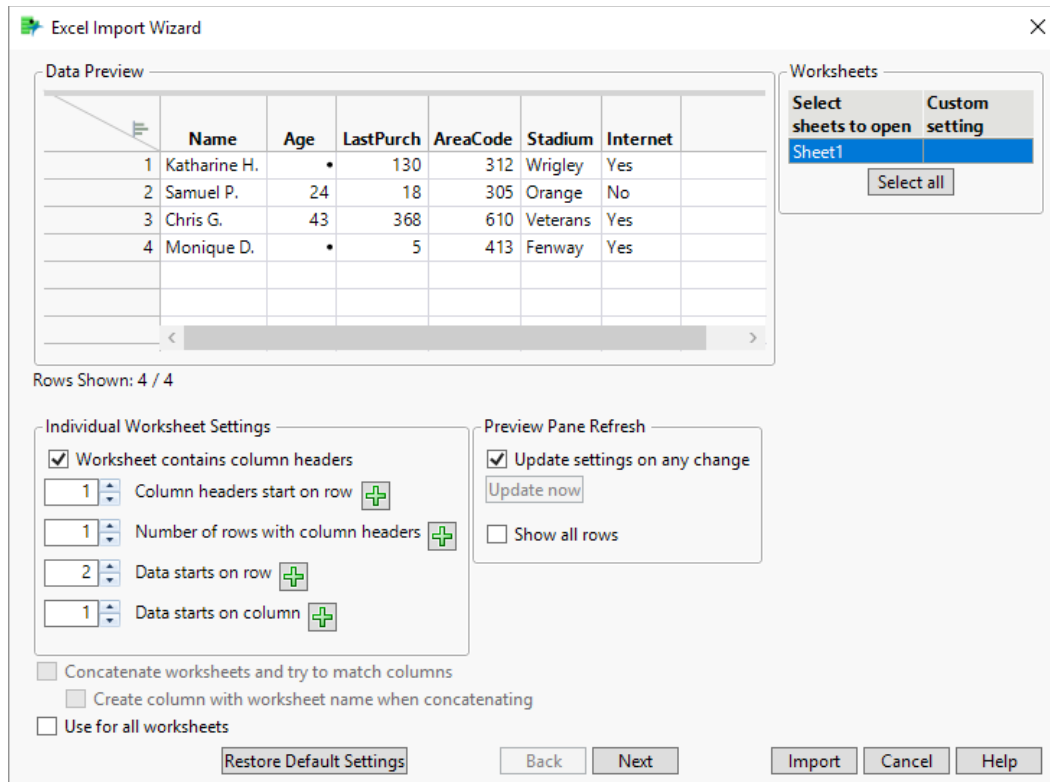


The upper portion of the Wizard displays a preview of the data as JMP will import it. In this example, the Excel file is organized exactly as discussed earlier. In other cases, you may find that an Excel file has column headings that occupy two rows, or that the data of interest begins in, say, column B of the worksheet rather than column A.

Notice that the character columns are left-justified, and the numeric columns are right-justified. This is almost what we want, except that **AreaCode** should be a nominal character column. We will need to fix this after the import by editing the **Column Info**.

Look in the lower left panel of the Wizard (see Figure B.7). This provides options to indicate precisely where the heading rows are and where the data rows and columns begin.

**Figure B.7:  How to Specify Location of Data Within the Worksheet**
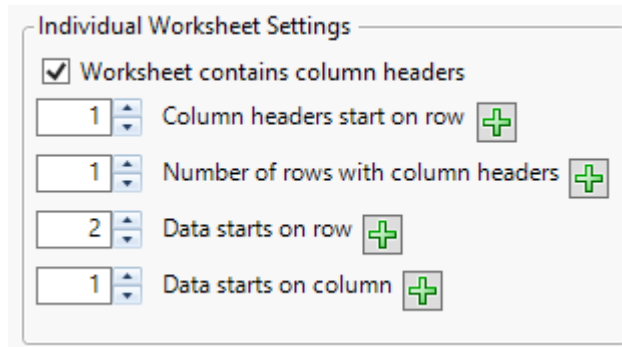


Additionally, if a workbook does contain several worksheets, you can specify either a single layout that applies to all worksheets or specify a different layout for each worksheet. The **Next** button in the Wizard opens a similar panel (not shown here) allowing further customization.

4.    Click the **Import** button in the Wizard, and a new JMP data table will open.

Particularly for a large worksheet, the Import Wizard is quite efficient, but as noted, it has no way to know that some numeric characters should be interpreted as nominal data, as in the case of area codes. Before starting to work with imported data, be sure to look closely at the **Columns** panel to check the modeling types and make any needed changes.

## The JMP Add-in for Excel

Suppose you are working in Excel and want to perform a simple analysis using a JMP platform, then continue your work in Excel. In the simplest case, let us say you have an Excel workbook and want to see the distribution of a numeric column of data. In Excel, several steps are needed to create a histogram and summary statistics.

When JMP 15 is installed on a Windows-based computer that also has Excel installed, JMP automatically installs an add-in within Excel, which creates a JMP tab in the Excel Ribbon. Figure B.8 shows the options available in the JMP add-in; as with the Excel Import Wizard, the add-in is extensively documented in Chapter 3 of *Using JMP*.

**Figure B.8: JMP Excel Add-In Commands**
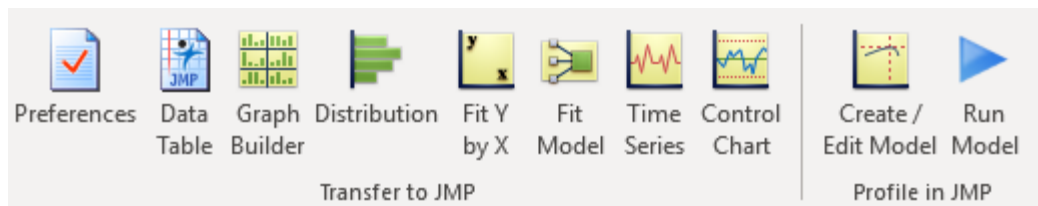


These buttons invoke JMP from within the Excel environment, and you can issue JMP commands beginning in Excel. If you select a data column in Excel and click the **Distribution** button, JMP will open and display the standard distribution report for that variable.

## Importing Data Directly from a Website

There is a large and growing amount of authoritative data available online. Many websites offer the option of downloading data tables in Excel or other formats; often, a website simply displays tabular data on a single page. One might be tempted to copy and paste such a table either into Excel or directly into a JMP data table, but JMP offers a simpler and more direct approach.

Throughout the book we have used data about nations of the world. As an illustrative example of a direct web import, we will locate a table listing several standard coding systems used to identify countries. In addition, this table includes small images of the flag of each nation.

1.   Open a browser and visit a Wikipedia page about the three coding systems established by the International Organization for Standardization, the ISO: https://en.wikipedia.org/wiki/ISO_3166-1.   Scroll down to locate the Current codes. Figure B.9 shows the first several rows of the table.

**Figure B.9:  Wikipedia Entry on Current ISO Country Codes**



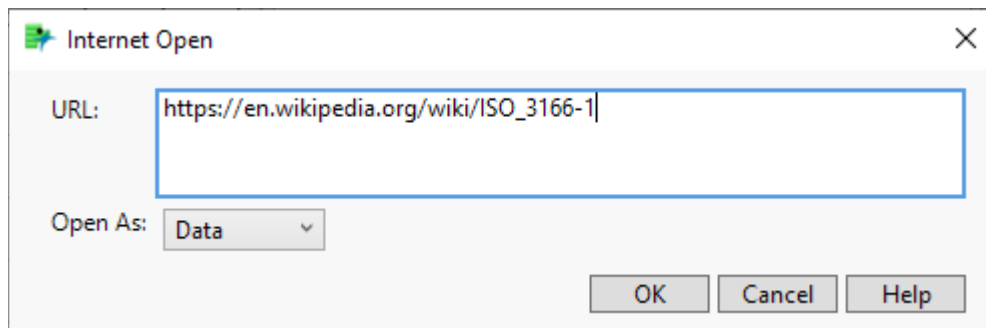Current codes  [edit]

**Officially assigned code elements**  [edit]

The following is a complete ISO 3166-1 encoding list of the countries which are assigned official codes.[17] It is listed in alphabetical order by the country's English short name used by 3166/MA
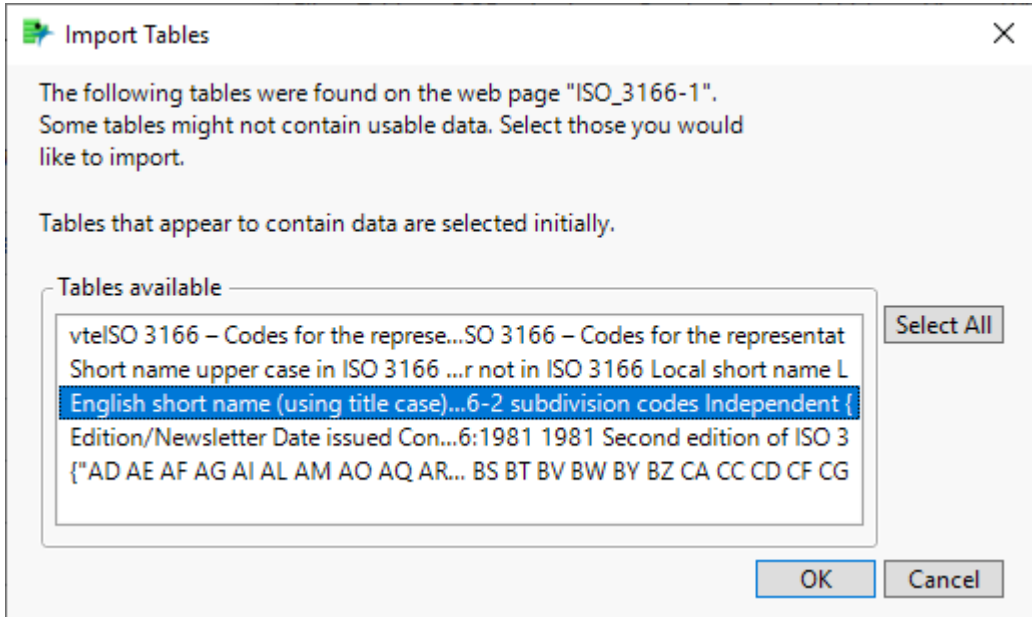
Note: Each country's *alpha-2 code* is linked to more information about the assignment of its code elements.

| English short name (using title case) ⬧ | Alpha-2 code ⬧ | Alpha-3 code ⬧ | Numeric code ⬧ | Link to ISO 3166-2 subdivision codes ⬧ | Independent ⬧ |
|---|---|---|---|---|---|
| Afghanistan | AF | AFG | 004 | ISO 3166-2:AF | Yes |
| Åland Islands | AX | ALA | 248 | ISO 3166-2:AX | No |
| Albania | AL | ALB | 008 | ISO 3166-2:AL | Yes |
| Algeria | DZ | DZA | 012 | ISO 3166-2:DZ | Yes |
| American Samoa | AS | ASM | 016 | ISO 3166-2:AS | No |
| Andorra | AD | AND | 020 | ISO 3166-2:AD | Yes |
| Angola | AO | AGO | 024 | ISO 3166-2:AO | Yes |
| Anguilla | AI | AIA | 660 | ISO 3166-2:AI | No |

2.   Move the cursor to the URL address box of your browser and copy the entire URL.
3.   Move back to JMP and choose **File ▶ Internet Open ▶ Web Page…** This will open the dialog box shown in upper panel of Figure B.10; paste the URL as shown.

**Figure B.10: Specifying an Internet Data Import**



Internet Open                                                                    ✕

URL:      https://en.wikipedia.org/wiki/ISO_3166-1|

Open As:   Data    ⌄

OK      Cancel      Help

JMP finds five "candidate" data tables on this web page. Each is identified by its column names. As such, the third is the one we want, with "English short name" as the label in the first column.

4. Just highlight the third choice and click OK. You will now have a new data table containing the data from the website, shown in Figure B.11.

**Figure B.11: ISO Country Code Data Table Import**

This does not look exactly like the web page, but it does contain the same column headings. The contents of the first column show the country name and a link to a Wikipedia Commons repository of country flag images.

5.  In the Table panel at the upper left of the data table, click the green arrow next to **Load pictures**... After a few moments, this will add a new column to the data table, displaying the flags, as shown in Figure B.12.

**Figure B.12: National Flags Unfurled**



6.  Open the Column Info and change the column name to **Flag**. Also notice that the **Data Type** is **Expression**. A JMP expression column contains JMP Scripting Language (JSL) code; in this case, the code is a reference to the image file for each flag.
7.  Select **Cols ▶ Use for Marker**. When we later make a graph using this data table, the points in the graph will be replaced by flags.
8.  Save this file as **Country_flags** in a directory of your choice for later use in this appendix.

With the list of country codes and flags we can add flag images to hover labels in graphs, but only after we combine this data table with one or more others. We will see how to do this in the next section.

# Combining Data from Two or More Sources: Query Builder

When conducting an observational study, it is common to assemble data from multiple sources into a single data table for analysis. For example, consider the data table that we just created,

**County_flags**, and the large **WDI** table that we have worked with throughout the book. The former table contains ISO codes and flag images for 249 countries; the latter contains detailed annual data for 215 countries. Suppose we want to revisit one of the visualizations from Chapter 5 to look at the relationship between life expectancy and per capita GDP, but use country flag images rather than simple points. One might be tempted to copy and paste the relevant columns from one table into the other, but it is not such a simple matter.

To see why, open the two data tables and look at the **Rows** panels. There are 6,235 rows, comprising 215 countries over a 29-year period listed in the **WDI** table. In contrast, the **Country_flags** table lists 249 countries. For this reason, the copy and paste approach would require considerable care to ensure accurately aligning of countries and data values. With more than 6,000 rows, the task would be fussy; with 60,000 rows, it would be entirely inefficient.

A better approach relies on database management, in which it is quite common to merge data residing in several different tables. The approach hinges on the idea of *key fields*, which are table columns that uniquely identify an observation and that are common to several tables.  In database terminology, we want to *join* data from two different tables, and we can accomplish this if and only if the two tables share a matching key column.

We can summarize the content of our two data tables as follows:

| Country_flags (n = 249) | WDI (n=6,235) |
|---|---|
| English short name… | Country Code |
| Alpha-2 code | Country Name |
| Alpha-3 code | Region |
| Numeric code | Lending Category |
| Link to ISO 3… | Year |
| Independent | air_pass |
| [3 metadata columns] | battle_deaths |
| Flag | [35 more columns] |

To join data from the two tables, we need a key field that appears in both and that contains matching information. Looking at the tables, there is one column that can serve as a key field: **Alpha-3-code** in **Country_flags** and **Country Code** in **WDI**.

To merge the data from two tables, we must designate one as the primary table. Because we will use data from WDI to create our graph, **WDI** is the primary table.  In the language of databases, we want to *query* the combined data. We combine, or *join* the two tables by linking records via their shared key columns. Note that we have more country flags than we have WDI countries. We want to include all the WDI country data and overlook unmatched flags. We will do this using the JMP Query Builder.

> Query Builder is a flexible and highly capable platform for database operations. This illustration leads you through essential steps to create and run a query. Interested readers should consult the JMP Documentation Library book *Using JMP*, specifically the Chapter 3 section titled "Build SQL Queries in Query Builder".

1.   Move your cursor to the **WDI** table and choose **Tables ▶ JMP Query Builder**. This opens the dialog box shown in Figure B.13. Because **WDI** was the active table when we invoke **Query Builder**, it appears as the primary table with the designation **(t1)**.

**Figure B.13: Selecting Tables to Build a Query**
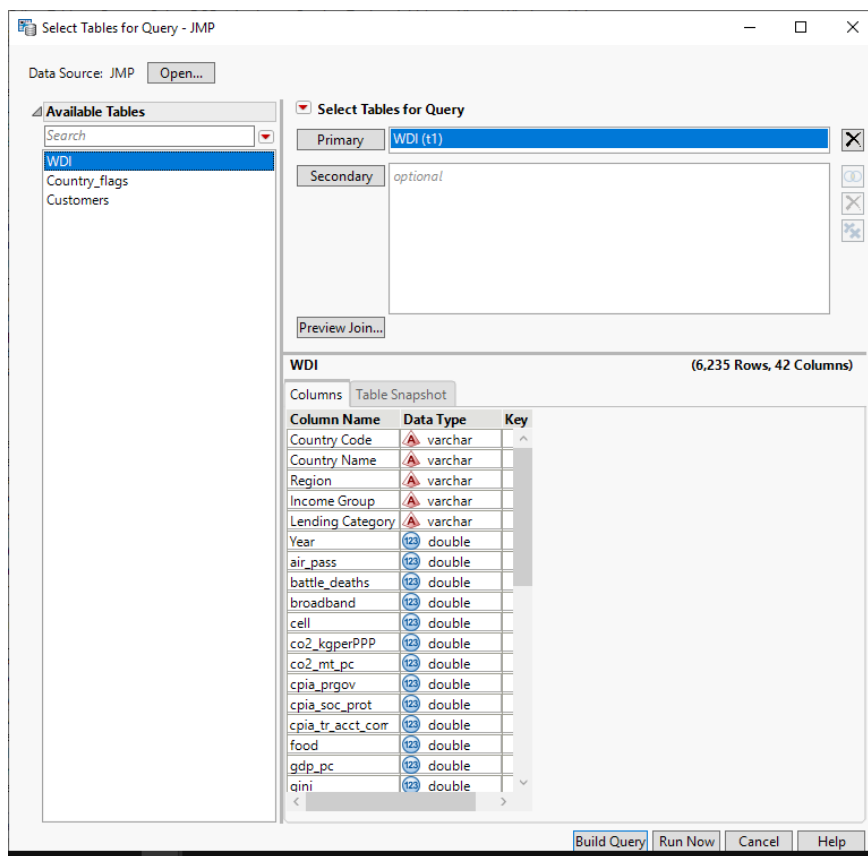
2. We can combine data from all open tables, but we only need two tables to create the visualization. In the upper left panel, below **WDI** select **Country_flags** and click the **Secondary** button to its right.

3. Next we need to specify the criteria for JMP to use when matching data rows. This requires three steps. First, highlight **Country_flags (t2)** under **Select Tables for Query**; refer to Figure B.14.

**Figure B.14:  Specifying Key Fields, Step 1**
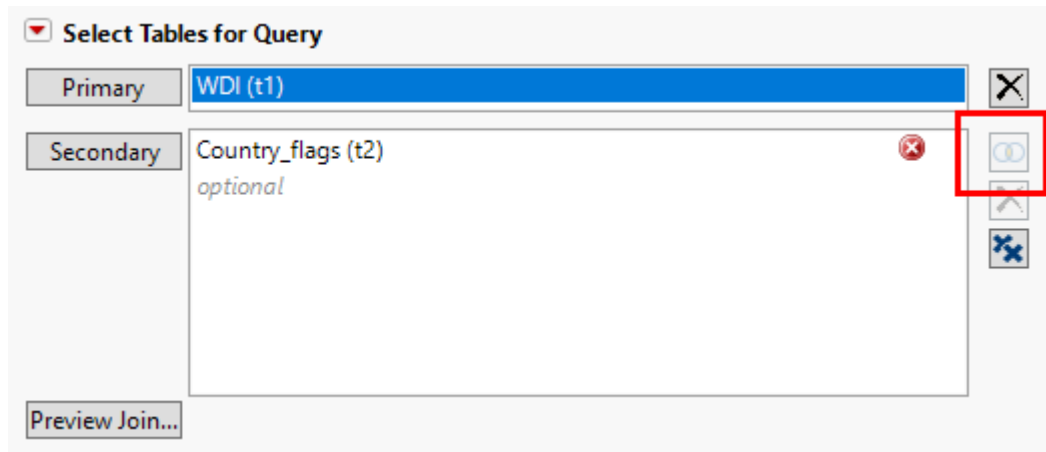


4. Once selected, the Venn diagram icon will become active. Click it to open the **Add Condition** dialog box in Figure B.15.

**Figure B.15:  Specifying Key Fields, Step 2**
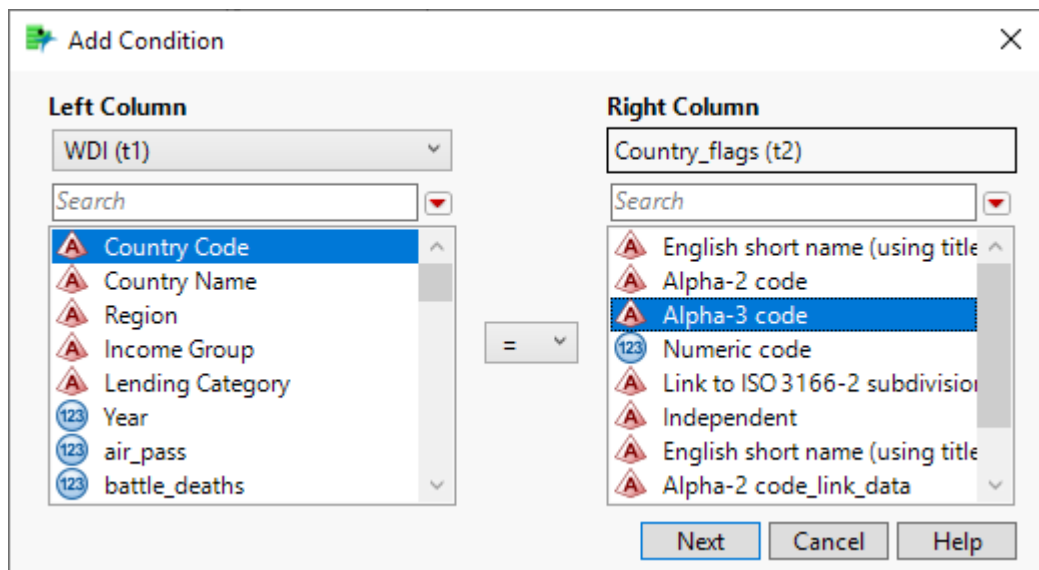
5. In this step we specify that the **Country Code** in **t1** should match the **Alpha-2 code** in **t2**. Highlight those column names as shown in Figure B.15 and click **Next**. Another dialog box (Figure B.16) opens to permit us to edit the join conditions. Click **OK**.

**Figure B.16: Specifying Key Fields, Step 3**
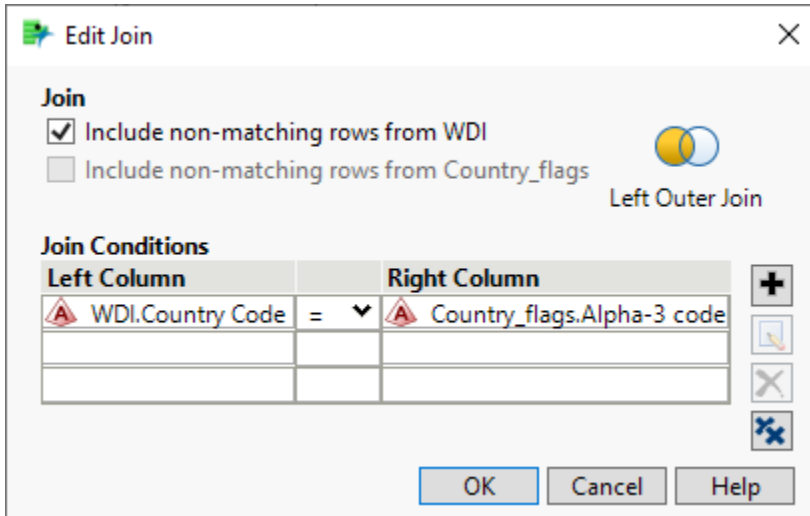


In the upper right corner of the dialog, notice the Venn diagram and the indication that JMP will perform a Left Outer Join. When combining data from two tables, there are four typical options for determining which rows will be preserved:

- Inner Join: Keep only those rows matching keys in both t1 and t2

- Left Outer Join: Keep all rows from t1 and any from t2 whose keys match t1.

- Right Outer Join: Keep all rows from t2 and any from t1 whose keys match t2.

- Full Outer Join: Keep all rows from both tables.

6. At this point, we could run the query and we would obtain a large table of 6,235 rows and all columns from the two source tables, or 52 columns. To create the visualization that we desire, we only need a few of the columns. Hence, we want to build a query in which we select only the columns we need. First, we specify the columns to include, and then we run the query. Click **Build Query** at the bottom of the window to open one last dialog box, shown in Figure B.17.

The left side of the window displays the currently chosen tables and their columns. Each column name has a prefix of either t1 or t2, indicating the source table.

**Figure B.17: Selecting Columns for a Query**
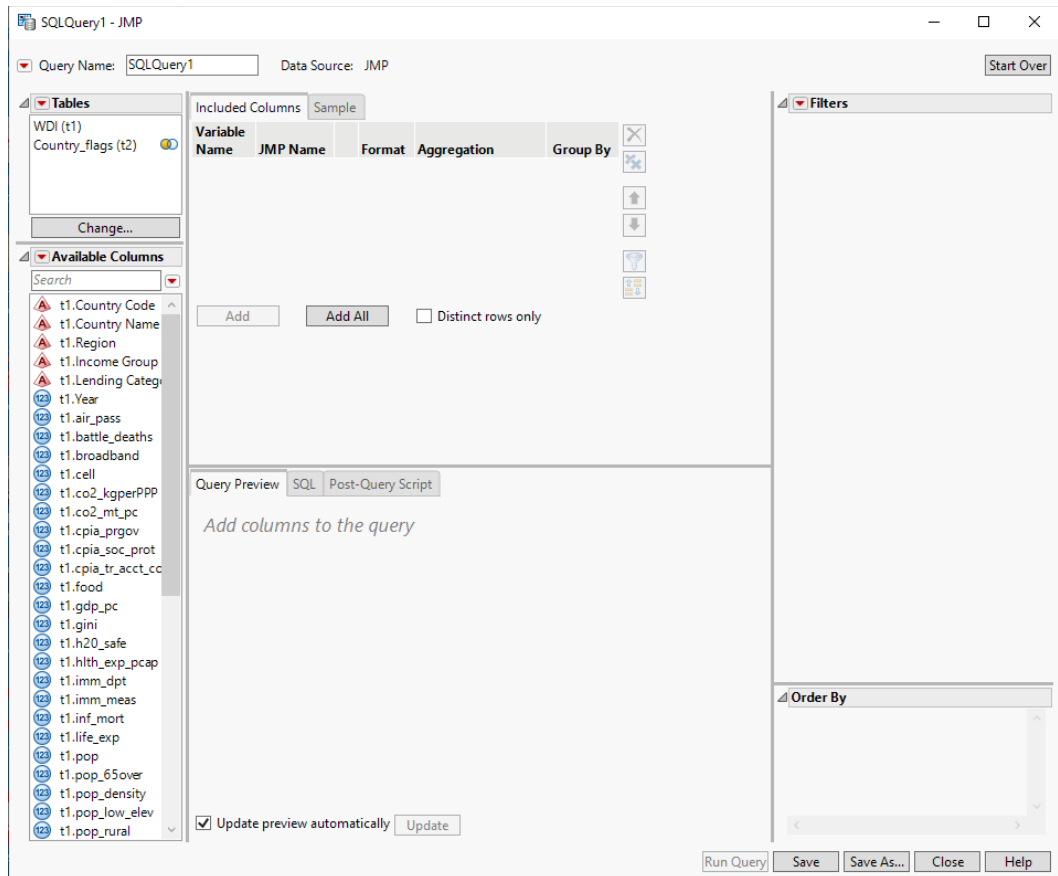


7.  Hold the Ctrl key and choose the columns **t1.Country Code**, **t1.Country Name**, **t1.Year**, **t1.gdp_pc**, **t1.life_exp**, and **t2.Flag**.
8.  In the upper center, click **Add**.

The Included Columns tab is now populated with metadata about the six columns we have selected, and the Query Preview tab displays the first several rows of the query results.

9.  Now click **Run Query** to produce a new data table containing the six columns we requested.
10. Finally, to demonstrate the use of images as point markers, choose **Graph ▶ Graph Builder**.
11. In the list of columns, right click on **gdp_pc** and Transform it to a Log scale.
12. Drag *Log[gdp_pc]* to the **X** drop zone and **life_exp** to the **Y** drop zone.
13. Click the red triangle next to **Graph Builder** and select **Local Data Filter**.
14. Select **Year** as the filtering column and click the red triangle next to **Local Data Filter**. Choose **Animation** and click. **Done** and run the animation.

There are myriad use cases for combining data from different sources. This application is admittedly on the trivial side, but it does illustrate the process.

## Further Considerations

Practical data analysis often calls for reorganization, consolidation, or other data management activities. JMP anticipates these needs and provides several tools to facilitate them. The **Tables** menu provides several other data management operations. We have seen some of them earlier in the book, such as the **Subset** and **Stack** functions. Though the objective in this section is to introduce the powerful **Query Builder** platform, be aware that the coverage has been limited and selective. As you undertake projects, revisit the JMP Documentation Library for other useful approaches.

---

[1] De Veaux, R. D., P. F. Velleman, and D. E. Bock. 2004. *Intro Stats*. 1st ed. Boston: Pearson Education, p. 7.