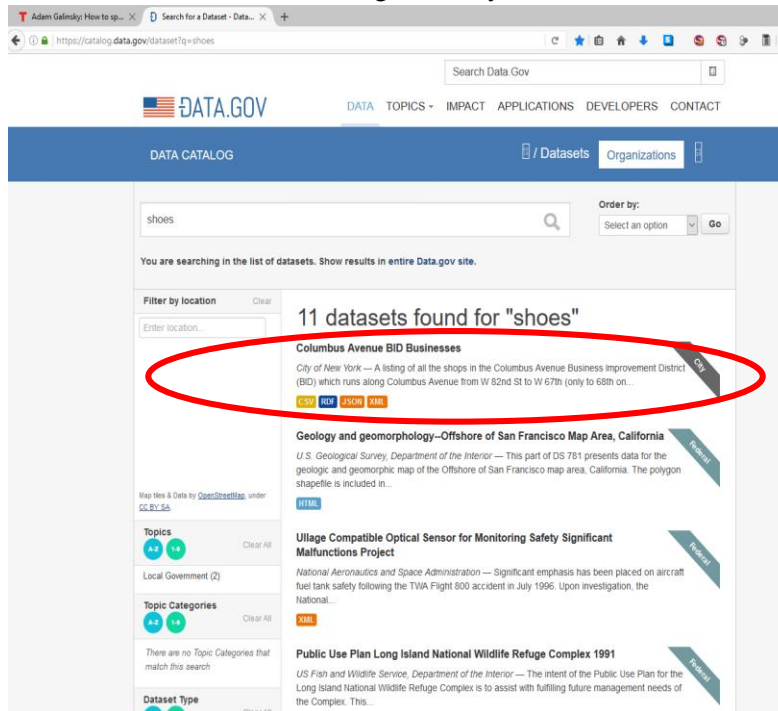# Student Self-Guided Worksheet to Working with Data in Excel

Finding Data and Importing Data into Excel

1. At Data.gov, search "shoes." Note that the available datasets are very limited, but note the format of the available datasets. The comma-separated value (CSV) format, a file format for organized data, most easily incorporates into Excel. Download the *Columbus Avenue BID Businesses* CSV file by clicking on the CSV icon below its description and opening the file with Excel. (Tip: You may have to drag the file to an Excel worksheet or change the format in the Excel open menu from *All Excel Files* to *Text Files*. You can also import the data using the menu *Data > From Text*.) Data is organized by columns. Columns can be resized to fit the data.

2. Obtain and import data using the following search engines and databases (use a new tab for each search): data.gov , konect, reddit, refdesk, Internet public library, iTools, Dmoz.org, Encyclopedia.com, Reference.com, Lifewire, and datahub.io (Tip: It is unlikely that a search will immediately give you formatted, packaged and relevant data. Try varying your search terms and exploring available data that is related to the search. For example, searching the term "shoe" does not immediately yield data related to shoe consumption or preferences. However, a search for "disposable income" may yield Census data with population demographics that may lead to pertinent and useful trends and results.)

3. Use the Knoema Data Finder plug-in to find data. Here, a direct search of "shoes" yields a table of planned monthly spending on shoes.

How to add Knoema to Excel:
1) Download the app > go to the File tab > click Options > click Add-Ins.
2) In the Manage box > click COM Add-ins > click Go.
3) In the Add-ins available box > select the check box next to Knoema (or Add it first, if it's downloaded but not listed)

How to use Knoema Data Finder:
1) Click the Add-Ins tab
2) Click the Find data icon
3) Type your query in the box in place of Search for data and statistics.
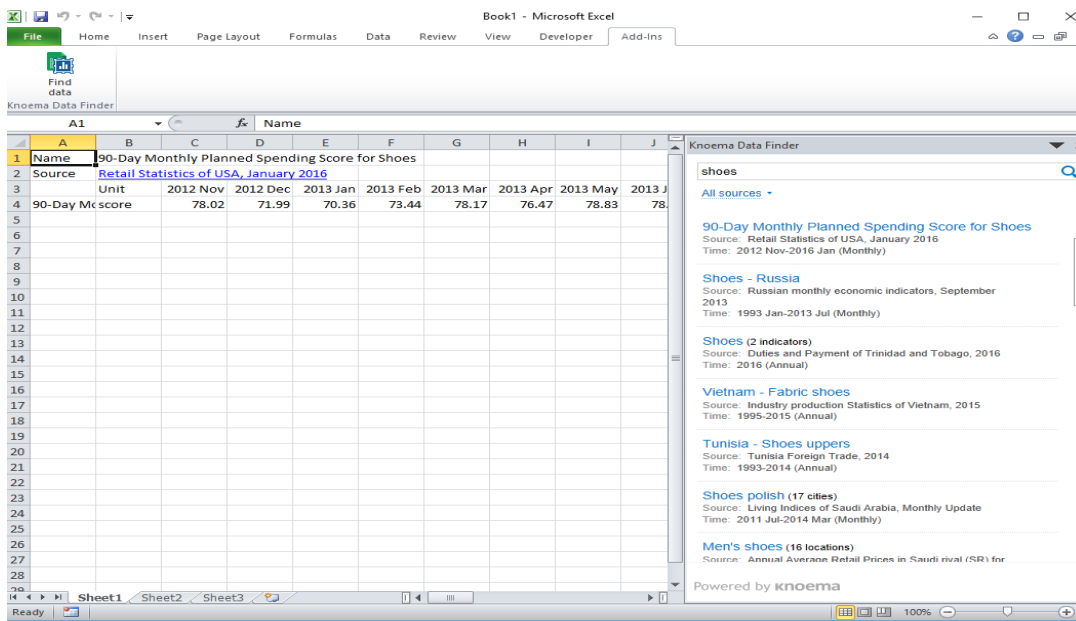
How to obtain data via Excel's *Data > From Web* command.
1) Use an Internet browser to find a website that hosts numeric data in tabular form. One example, shown below, is ESPN's MLB Player Batting Stats – 2017 page.
2) Copy the website's URL from the browser address bar.
3) In Excel > go to the Data tab > select From Web (left of the menu, shown below)
4) Paste the URL into the *Address* in the *New Web Query* form > click Go.
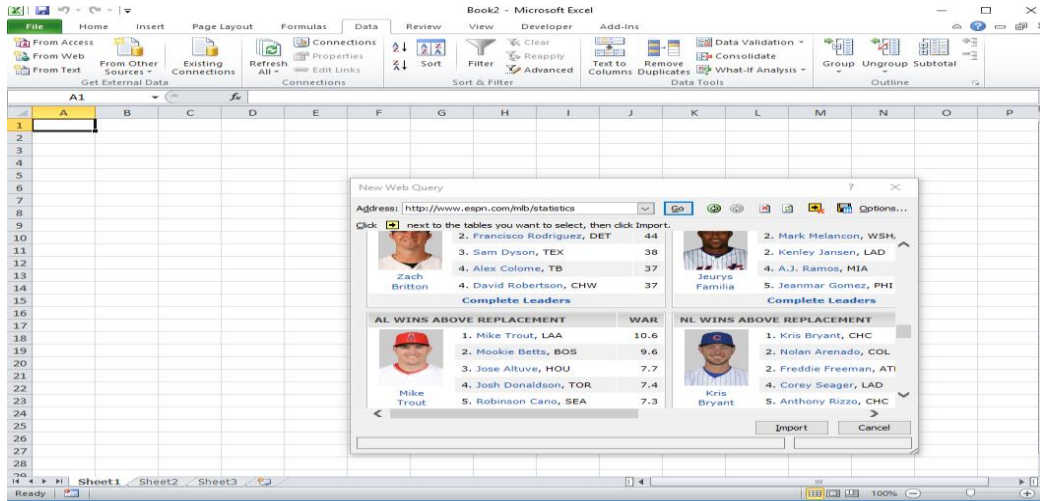
5) Click *OK* to import the data into the existing Excel worksheet. Notice that the data includes lots of extraneous information, as shown in the next image:

6) You can "Clean" the data by hiding (or deleting, with caution) rows or columns that repeat or are irrelevant. Go to the *Data* tab > click the *Remove Duplicates* button > click *Select All* > click *OK*. Manually hide remaining irrelevant text and information by selecting the rows or columns: right-click > select *Hide*. Save your progress for the next section.

4. Explore your own searches using the given tools.

Name: _____ Date: _____ Class: _____

Formulate Product Plan and Gather Data

This includes: Identify a product. Determine at least five variables related to the product's production, marketing and sales. Search for data using a variety of search resources and search terms. Import relevant data into Excel. Keep records of the data source URLs.

Analyzing Data with Excel

Practice inserting a scatter plot of related columns of data in order to analyze their relationship by inserting a trend line and viewing the coefficient of determination (R^2). These tools enable a visual analysis and provide a statistic of the correlation between the data, which show the strength of their interrelation.

1. Open the Excel file from the earlier importing data exercise (or ask the teacher for access to the *ESPN Example* file). Select the first column (or row) of data by clicking and highlighting just the data values, or by selecting the row or column header at the top. Then hold down the control key (*Ctrl*, usually near the space bar on the keyboard) while selecting the second row or column of data. As an example, select columns D (AVG) and I (GB) > go to the *Insert* menu > select *Scatter* > select the *Scatter with only Markers* option > move the chart to a blank, open area in the Excel sheet.
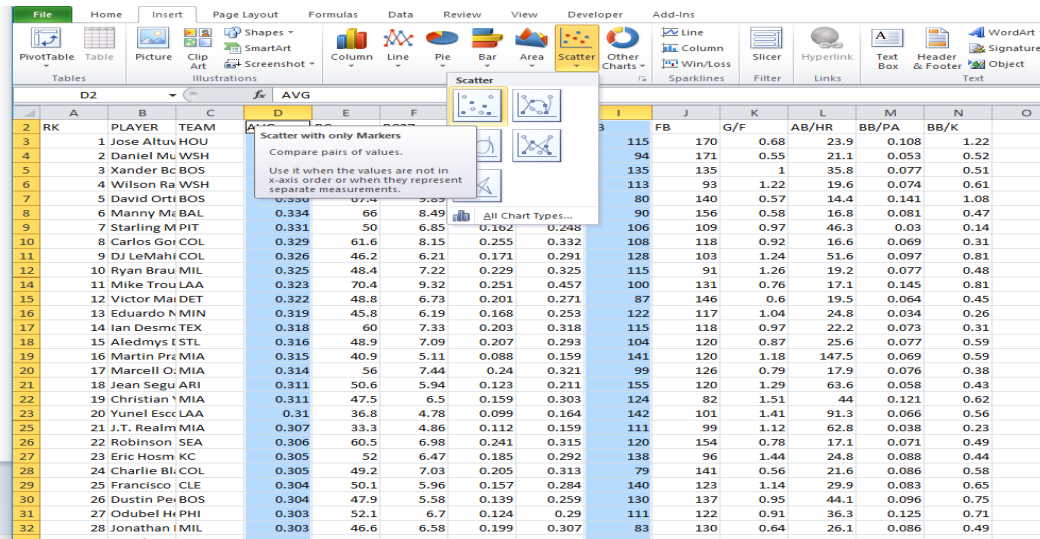
2. The chart auto-scales the axes. To customize the graph, double click on the values along one of the axes to bring up a menu with options.
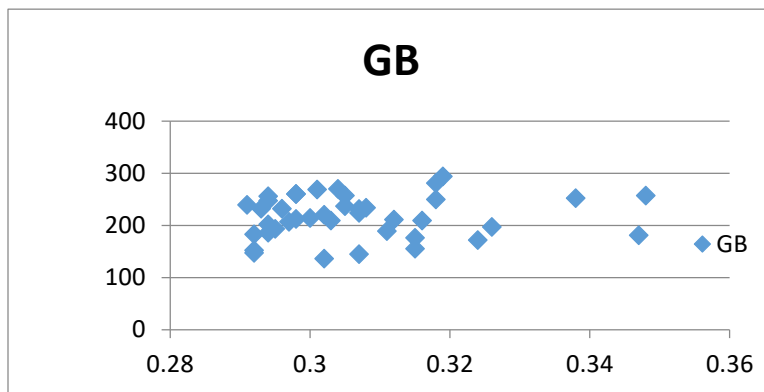
3. To customize the output, double click on part of the graph to bring up a menu box with options. Right click on the data points themselves to add a trend line, as shown below. In the *Format Trendline* box that appears, check the boxes *Display Equation* on chart and *Display R-squared value* on chart. (The R-squared

value, or the coefficient of determination, is a value between 0 and 1, in which, essentially, 0 indicates no relation, and 1 indicates a perfect relation.)
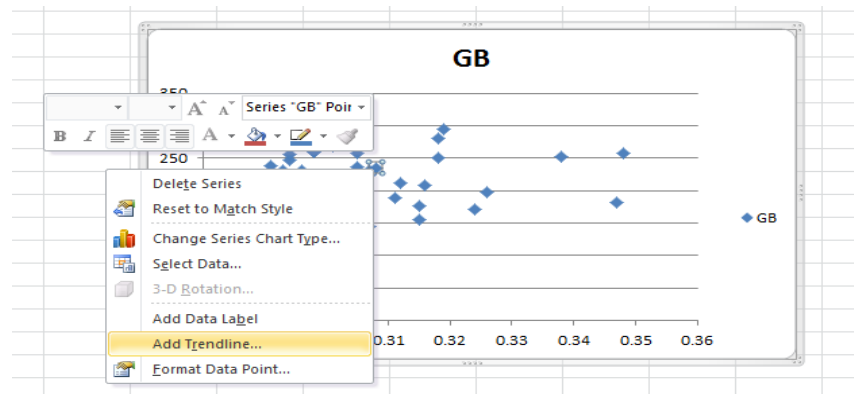
Note in this example, the R^2 value is 0.0089, which tells us that in this dataset, very little correlation exists between batting average and number of ground balls. Keep in mind that finding no correlation is useful knowledge! Experiment with other columns to note differences in plots, trends and R^2 values.

4. Load the *Analysis ToolPak* (included with Excel) by clicking on the *Developer* tab >selecting it from *Add-Ins Available*. After loading > go to the *Data* tab > click *Data Analysis*, then:
   - Explore *Histogram, Descriptive Statistics*, *Correlation* and others.
     See the online Microsoft Help for more details on the use of the Analysis ToolPak.
   - Experiment with variable pairs and graph customization, taking note of shape and trend line differences

5. Use the remaining time to continue to search for and import data.
   Remember to continually save your progress!

Research and Analyze Data

Focus on the goal of the project: To find relevant variables' effect on the data and make decisions as to planning and marketing.
- Create scatter plots, trend lines, and R2 values for collected data collected. Also, find descriptive statistics via the Analysis Toolpak.
- What do you learn from the data? For example, the low correlation between ground balls and batting average may suggest that batting average is more attributed to fly balls and line drives. Or, perhaps the best hitters hit ground balls as often for outs as they do for hits. These conjectures can lead to more research and discovery; however, each conjecture by itself is weak and might not be good advice without testing the data. Be careful not confuse correlation with causation.
- Note that the results may not reinforce your opinions and expectations. Look for creative responses and solutions; conduct further research as necessary.

Create a Presentation of Analysis

Create an electronic presentation of your research. Suggested outline:
   Page 1: Introduction with topic; include images, overview, why did you chose this topic?
   Page 2: Related variables with explanations; why was it chosen? how did it come up?
   Page 3: Graphs and relations with variables; explain patterns and/or lack of patterns, trends, inferences
   Page 4: Difficulties; where would you recommend more emphasis? What was lacking? What do you wish you would have located or discovered? What goals were not supported?
   Page 5: Recommendations based on research; give degree of risk of proposal, explain why you would recommend and what benefits may occur

Finish Work and Give Presentation

Complete presentations that summarize your research, graphs, interpretations and conclusions, and present to the class for assessment by the outline and rubric criteria.