

# NC-NET Employability Skills Resource Toolkit

## Module 6: Information Processing

Overview .....	2
Presentation Materials .....	3
Teaching Resources.....	8
Activity: Acquiring Information.....	9
Activity: Evaluating Information .....	17
Activity: Organizing Information.....	22
Activity: Managing and Presenting Information.....	28
Activity: Interpreting Information.....	39
Assessment Tools/Strategies .....	49
Rubrics for Instructor Assessment .....	50
Rubric for Self-Assessing Information Processing Skills.....	52
Rubric for Assessing Information Processing Skills.....	55
Videos and Links.....	56
Videos.....	56
Weblinks.....	56

## Overview

The ability to manage information is essential in today's technological workplace. Beyond the obvious exchange of emails and text messages, there is the ever-growing flow of documents, training materials, photography and videography, and data of all types (financial, medical, legal, statistical, scientific, and so forth). While physical libraries still contain vast repositories of information and reading materials in print, more and more of our knowledge and information is being handled electronically. The world's total capacity to store information is growing exponentially, leaving behind the exabyte ( $10^{18}$  bytes) and moving into the zettabyte ( $10^{21}$  bytes). A zettabyte is equivalent to about 250 billion DVDs. According to one estimate, the world will contain 44 zettabytes of digital data by 2020 (*The Economist*, 2015 [<http://geloookahead.economist.com/storage/>]).



In today's workplace, workers in almost any career area must manage a large flow of information. This activity will equip students to process some of this information so that they will not be overwhelmed. To that end, we've broken this module into five categories and will consider each in turn:

1. **Acquiring:** locating and using sources of information; assessing their validity and reliability; searching libraries and the Internet; and citing sources to avoid plagiarism and copyright infringement
2. **Evaluating:** assessing the validity of sources and data (e.g., blogs versus scientific journals) and reliability of Internet data; considering periodicals, infomercials, advertising as potential sources; and citing sources for future research
3. **Organizing:** dealing with large volumes of information; using spreadsheets and databases; and making a table of contents or an index
4. **Managing/Presenting:** processing large volumes of information; avoiding data distortions (e.g., out-of-scale graphs and incorrect data subsets); and selecting and preparing charts (e.g., pie charts, line charts)
5. **Interpreting:** using basic statistics to analyze and graph data; understanding linear vs. logarithmic scales; and making historical comparisons

Clearly, each of these topics could be expanded into an entire course. The suggested activities will serve as an introduction to the wealth of subject material available for further study and exploration—a supplement for your existing coursework.

In this module we have provided

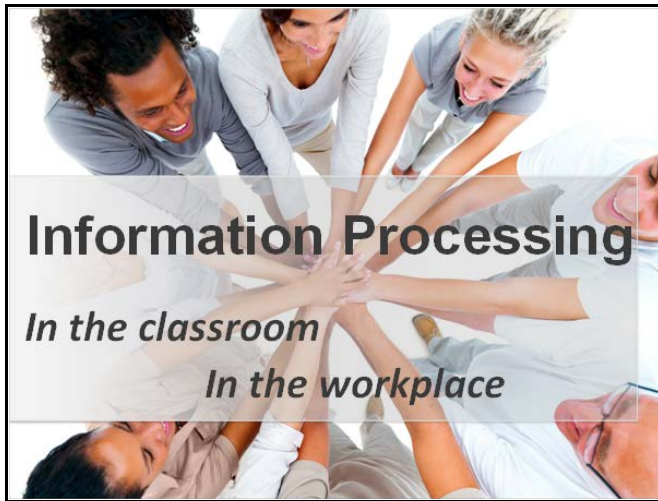
- **Instructor Presentation Materials**—A set of slides to help organize the topics for discussion, and to provide talking points that introduce the activities
- **Classroom Activities**—A set of activities requiring minimal material and preparation, addressing several facets of information processing
- **Assessment Tools**—Rubrics for use by both instructor and student to assist in gauging progress throughout the course
- **Videos and Links**—A collection of links that enable the user to take advantage of the abundance of electronic resources available

# Presentation Materials

SLIDE

TEACHER NOTES

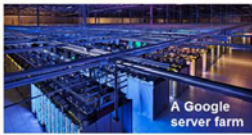
1



2

**Information is growing exponentially!**

- One recently developed SSD (solid-state drive) can hold 60 TB. (That’s ~60,000 GB or 60,000,000 MB!)
- According to one estimate, the world will contain 44 zettabytes of digital data by 2020.
- What’s a zettabyte?



3

**Units of Measure**

- 1 Bit = Binary Digit
- 8 Bits = 1 Byte =  $10^0$  B
- 1024 Bytes = 1 Kilobyte =  $10^3$  B
- 1024 Kilobytes = 1 Megabyte =  $10^6$  B
- 1024 Megabytes = 1 Gigabyte =  $10^9$  B
- 1024 Gigabytes = 1 Terabyte =  $10^{12}$  B
- 1024 Terabytes = 1 Petabyte =  $10^{15}$  B
- 1024 Petabytes = 1 Exabyte =  $10^{18}$  B
- 1024 Exabytes = 1 Zettabyte =  $10^{21}$  B
- 1024 Zettabytes = 1 Yottabyte =  $10^{24}$  B
- 1024 Yottabytes = 1 Brontobyte =  $10^{27}$  B
- 1024 Brontobytes = 1 Geopbyte =  $10^{30}$  B

- A zettabyte is a trillion gigabytes! So,  $2.5 \text{ ZB} \sim 2.5 \times 10^{21} \text{ bytes} = 2.5 \times 10^{12} \text{ GB}$ , that is 2.5 trillion gigabytes—quite incomprehensible.

- A review of those metric prefixes is on the next slide.


- Here’s a review of those metric prefixes. Strictly speaking, 1 kilobyte = 1,024 B, 1 megabyte = 1,024 kB, and so forth as shown on the slide here, but it’s common to round the value “1,024” to just “1000” and thus say 1 megabyte = 1 thousand kilobytes = 1 million bytes, and so forth.

## SLIDE

## TEACHER NOTES

4

**Handling Information in a Modern World**




- **Acquiring information:** locate and cite sources, assess source validity, use search tools
- **Evaluating information:** current, valid, reliable source
- **Organizing:** use tables, spreadsheets, databases, index
- **Managing/Presenting:** make graphs and charts, type of graph (x-y, pie chart, bar chart), don't distort scale
- **Interpreting:** statistics, graphs, linear/log scales, timelines

- **Acquiring:** Locating sources of information; assessing their reliability; techniques for searching (library, web browsers); citing sources to avoid plagiarism and copyright infringement
- **Evaluating:** Assessing the validity of sources and data (e.g., "Bigfoot" versus scientific); reliability of Internet data; considering periodicals, infomercials, advertising; citing sources for future research
- **Organizing:** Dealing with large volumes of information; using spreadsheets and databases; making a table of contents or an index
- **Managing/Presenting:** Dealing with large volumes of information; avoiding data distortions (e.g., out-of-scale graphs, data subsets); selecting and preparing charts
- **Interpreting:** Using basic statistics to analyze and graph data; linear versus logarithmic; making historical comparisons and perspectives

5

**Acquiring Information: A Scavenger Hunt**



- Given a list of items: **Search, Find, and Cite.**
- Internet search engines are powerful.
- Citation information:
  - **Text:** Last name, First name. Title of Book. Place of Publication: Publisher, Year of Publication.
  - **Web:** Page source (organization or business): Page title. Date of publication. Date of collection, URL.
- A citation enables a reader to verify your find.

- The activity for "Acquiring" will be like a scavenger hunt, in which students are given a list of questions to research and answer. In addition to the answers, they must cite the sources so that a reader can verify them.
- Students will likely be quite familiar with and comfortable using Internet search engines like Google, Bing, and Yahoo.
- The search topics can all be found online, but a couple can also be found in a traditional reference book (e.g., CRC Handbook). Encourage students to remember that printed materials are sometimes the best sources of information.
- Common citation formats are MLA, ALA, and Chicago. If you wish to teach a particular citation style, the Purdue University Online Writing Lab has an excellent resource: <http://owl.english.purdue.edu/owl/section/2/>.



SLIDE

TEACHER NOTES

6

### Evaluating Information

- Printed publications are edited, proofed, typeset, and printed: **usually reliable**.
- Web content can go from keyboard to web page in a matter of seconds: **trustworthy?**
- “Let the *Reader* beware.”
  - Who is responsible?
  - What is the source?
  - How accurate and reliable?


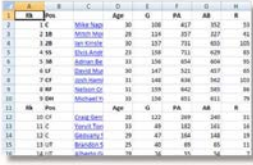



- Students who routinely surf the web quickly develop a sense about the reliability of the information presented on a website but still may be fooled by a slickly presented page that offers something that sounds too good to be true. This activity encourages students to question claims and “statements of fact” they find on the web. The handout offers sites to visit and evaluate. The instructor can add to the list. Another handout gives a checklist for assessing each website.

7

### Organizing Information: Using a Spreadsheet

- Data, data everywhere!
- Copy/paste data into a spreadsheet so you can:
  - Reorder it
  - Label it
  - Calculate with it
  - Sort it
- Example supply of data: sports statistics from several teams

- In this activity students select data and work a few exercises that demonstrate how easy it can be to harvest and manipulate large amounts of data in a few minutes.
- We’ve selected the career-neutral area of sports statistics to demonstrate the skills in this activity. But as noted (and suggested with examples) in the accompanying module, one can find sources of data in every career field. By using sports statistics, the data is current, voluminous, and reliable. Most Americans are acquainted with the basics of baseball, so much of the terminology should be understandable.
- The activity will give copy/paste instructions to accurately transfer the data from the web page to a spreadsheet file.

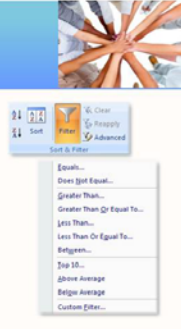
SLIDE

TEACHER NOTES

8

### Managing Information

- **Sort** the data by various characteristics and combinations of characteristics.
- **Filter** data to examine a subset of all the data.
- **Calculate** various statistics and parameters to interpret the data.




- Sorting (to find the largest or smallest, compare groupings, etc.) is one of the things computers do easily with large lists of data, but is hard for people to do manually.
- This is not a lesson about filtering, but exposes students to what can be accomplished by filtering. As illustrated by the Excel dropdown captured on this slide, filtering can use any of several basic value comparisons, as well as custom filters.
- Of course, every spreadsheet is equipped with a whole host of functions and presentation tools to analyze and further “crunch” the data. The activity guidelines and accompanying handouts offer some introduction to these capabilities.
- Just as Excel spreadsheets are good at calculating, they are also good at making graphs and charts.

9

### Presenting Information

- **Present** the data in various ways (charts, graphs) and groupings (pairings).



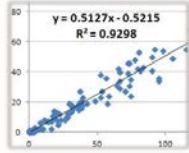
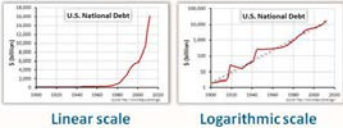
SLIDE

TEACHER NOTES

10

**Interpreting Information**

- Calculate statistics
  - Correlation
  - Regression
  - Statistical significance
- Consider transformations
  - Logarithmic
  - Exponential

Linear scale      Logarithmic scale

- In some ways the final activity is a continuation of the preceding activity since it involves graphing. But in this case, we are more interested in the analysis of the data, and in many cases a graph is the start of that analysis.
- Many data relationships we encounter are linear, so linear regression in which we are able to interpret the slope as a trend/rate, or a parameter in a relationship, is applied.
- The activity guidelines include a brief exposure to statistical analysis (t-test) using the elementary functions included in Excel. A more extensive treatment of statistical analysis would require special software and is well beyond the scope of this activity.
- Humans aren't very good at handling exponential relationships, but by applying a logarithmic scale to the graph they become linear. The final activity in this module directs students to the historical data showing the rapid growth of our national debt: is it truly "growing exponentially?"

- <http://www.brainyquote.com/quotes/quotes/a/alvintoff1389092.html>

11

**Quotable Quotes**

*You can use all the quantitative data you can get, but you still have to distrust it and use your own intelligence and judgment.*

Alvin Toffler

## Teaching Resources

### ACTIVITIES

The following activities can be completed in class to emphasize, teach, and practice Information Processing. They can be used “as-is” or they may be tailored to fit a specific course. Suggestions for adaptation precede each activity, with examples from several different subject areas/career pathways. The suggested modifications provide instructors with ideas for adapting the activity to fit content they are already teaching. Modifying the activities allows employability skills to be infused in subject area content more easily.



## ACTIVITY: ACQUIRING INFORMATION

### Instructor Preparation

In any career area, information is used to make decisions. But one must know where to find that information, how to identify helpful content and ignore the useless parts, and how to properly credit the source. In this activity, students receive exposure to the huge world of available information and learn to use some tools to retrieve it.

Searching for data is a lot like a scavenger hunt. In the handout for this activity, we offer a sample list of searches. You are encouraged to replace some or all these search tasks with some that resonate more with your course area. Try to list items that require a variety of information sources including some that will probably not be found on the Internet. For the younger generation, finding things outside of the Internet will be quite challenging, and time-consuming, yet often is the best option. While you could consider a variety of non-Internet information sources, such as library indices and catalogs, old newspaper archives on microfilm, encyclopedias, volumes of legal records and government proceedings, and so forth, it will be simpler to confine the recommended sources to those generally available on the Internet, and possibly one or two well-regarded reference manuals or handbooks that would be commonly found on a workplace desktop in your career area.

### Objectives

Students will:

1. Find specific facts in print and electronic forms.
2. Use an Internet search engine to find specific facts.
3. Cite sources of information, sufficient for future identification.

### Materials

- Access to an example reference book or handbook for a career area of your choice (e.g., a shop operations manual, a medical reference book such as the *Physicians' Desk Reference*, etc.)
- Access to the Internet and a search engine (i.e., in a computer lab or via smartphone or other device)
- Access to a photocopier (to record reference material found in books)
- Access to a printer (to print a sample of a webpage)
- Access to electronic storage media (e.g., network hard drive or USB memory stick, to save electronic data or images)
- Optional: access to screen capture software program (e.g. Jing, Windows printscreen command)

### Activity Guidelines

As suggested above, this activity can be treated as a homework assignment, a classroom activity in the computer lab, or as a scavenger hunt with student teams of two to three students competing against one another, earning points and rewards, and finally comparing/contrasting their findings for the benefit of all. The goals, the contest timing and other constraints, and the award or recognition given to

the winning team are left to your imagination. For the best result, modify the handout to customize the search items for your class and career area. Here are some ideas:

- In Manufacturing, searches could include recommended setup parameters found in a reference manual or handbook for a certain machining operation, safety regulations and MSDS literature, past and present comparisons for industrial raw material consumption, production levels, and labor input/output.
- In Health Science, searches could include results of recent research, new treatments, statistics on effectiveness of various therapies, salaries and wages, and the impacts of insurance and rising patient ages on healthcare costs.
- In Agriculture, searches could include effectiveness of various soil treatments on crop yields, best management practices for livestock management, the effects of weather on yields, influence of water table levels, and market pricing of various foods.

The central focus of the activity—especially if competition is involved—will be the list of search items you distribute to the class. (See the handout for this activity.) Spend time preparing it and be careful that the answers are not divulged prematurely. This activity will work best with your students if the list is customized to be interesting and relevant for your class/career area.

Note that for the information to qualify as being “found,” a cited reference (sufficient to enable anyone to return to the sources) is required. For graphical or pictorial results, a printed copy of a graphic or screen capture of a recognizable portion of the find is required. Instruct students to print only one (or two, at most) page of each find, as that will be sufficient evidence for this activity. Later activities in this Information Processing module may reuse the resources found through this activity, and for that reason, the accurate and reliable source citation (e.g., the web address) will be very valuable. In addition, students can digitally save the entire page, file, image, etc. (e.g., to a network drive or USB memory stick) for future use. Of course, for data that is constantly changing (e.g., stock prices or sports statistics), saving the data electronically may be unproductive since the data is soon out-of-date.

The example list of search items on the handout is presented in increasing order of difficulty. That is, the first few items can be easily found.

Divide the students into teams of two or three each. Announce to the students a time limit for the hunt. If students are relatively inexperienced with Internet surfing, finding a half dozen items could take one to two hours to complete. But do not allow the search to go on that long. The last two items on the list are difficult and may require locating printed reference books to find the answer. If not already available on your classroom bookshelf, that sort of background research could itself consume one hour.

### Group Reflection Questions

When the allotted time has expired, bring the students back together and give them a chance to describe their searching experiences.

- Which items did you find using a one- or two-word search?
- Which items were more difficult to find?

- How did you decide which resources to use?
- Would you be able to find the same resources in the future, based on your citations?

The more difficult items in the list required more than simple word searches and also an interpretation of the resources found. The students should discuss such questions and how they applied the information, as well as report the results of their research. For students accustomed to web searches, the citation requirement probably seemed burdensome, so emphasize how results should now be easier to replicate and verify.

Check to be sure that everyone has good references for each of their finds. The references should allow you to verify students' answers.

### **Homework**

If the searching activity proceeds too slowly, those who have Internet access at home or in the school facilities after class or can complete the activity as homework. A follow-up to this activity could be to ask students to analyze assignment mistakes (misinformation) and do the research to correct them.

## Handout—Sample Search List

Search and Find	Citation and Location Info/ Screen Capture/Photocopy
What is the Big Mac Index? What does it show?	
What are the target dates for the United States to phase out household incandescent light bulbs? What authority (person or law) is making this happen?	
What major league baseball game had the fewest number of fans in the stands?	
What was last month's average daily high temperature in your city, compared to the same month a year ago?	
What alloys, if any, of Copper (Cu) are superconductive above 20°K?	
Value of the binomial coefficient, $\binom{45}{16}$	

Suggestions for minimum requirements of citations:

**Text:** Last name, First name. Title of Book. Place of Publication: Publisher, Year of Publication.

**Example:** Smith, James. Engineering in Metal. Chicago: Cenwich Publishers, 2010.

**Web:** Page source: Page title. Date of publication. Medium of publication. Date of collection, URL.

**Example:** United States Government Accountability Office. "FDA Regulation of Dietary Supplements (HRD-93-28R)", 2 Jul 1993. Web. 19 Mar 2013. <<http://www.gao.gov/assets/90/83193.pdf>>.

**For more detailed citation guidelines, see the Purdue Online Writing Lab**

<http://owl.english.purdue.edu/owl/section/2/>

**Suggested Solutions to Sample Search List (DO NOT DISTRIBUTE THIS)**

**Search and Find**

What is the Big Mac Index?  
What does it show?

**Citation and Location Info/Photo/Capture/Photocopy**

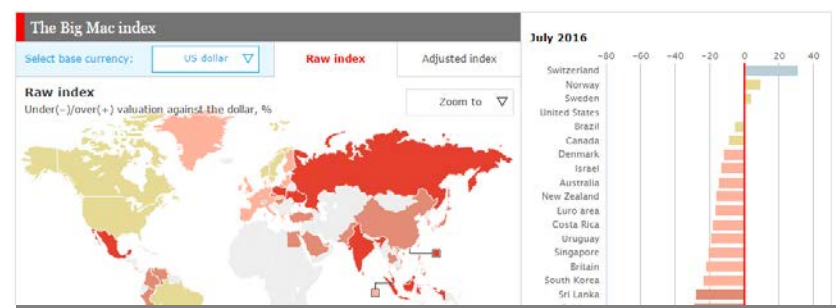
*Interactive currency-comparison tool: The Big Mac index.* 2016.  
<http://www.economist.com/content/big-mac-index> (July 2016).

**Interactive currency-comparison tool**  
**The Big Mac index**  
Global exchange rates, to go  
JULY 21ST 2016, BY D.H. & R.L.W.

THE Big Mac index was invented by *The Economist* in 1986 as a lighthearted guide to whether currencies are at their "correct" level. It is based on the theory of purchasing-power parity (PPP), the notion that in the long run exchange rates should move towards the rate that would equalise the prices of an identical basket of goods and services (in this case, a burger) in any two countries. For example, the average price of a Big Mac in America in July 2016 was \$5.04; in China it was only \$2.79 at market exchange rates. So the "raw" Big Mac index says that the yuan was undervalued by 45% at that time.


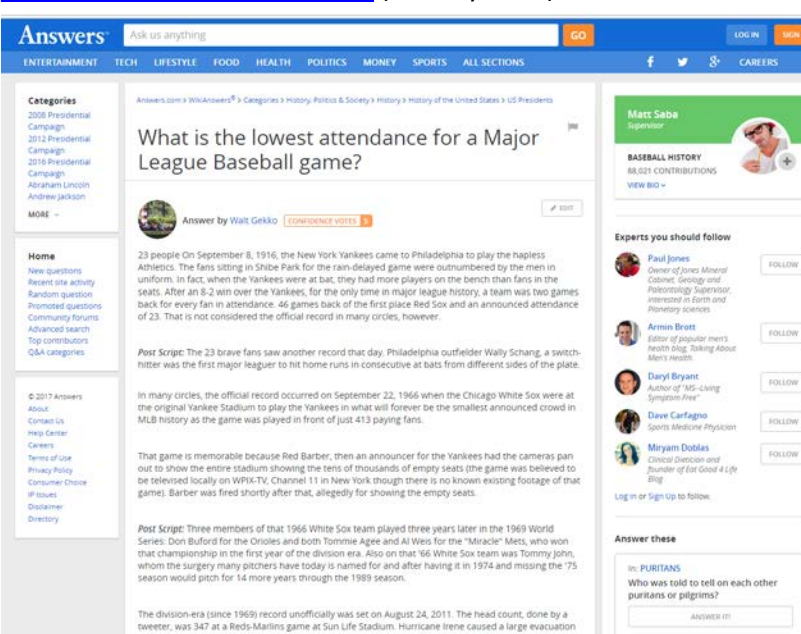
Burgernomics was never intended as a precise gauge of currency misalignment, merely a tool to make exchange-rate theory more digestible. Yet the Big Mac index has become a global standard, included in several economic textbooks and the subject of at least 20 academic studies. For those who take their fast food more seriously, we have also calculated a gourmet version of the index.

This adjusted index addresses the criticism that you would expect average burger prices to be cheaper in poor countries than in rich ones because labour costs are lower. PPP signals where exchange rates should be heading in the long run, as a country like China gets richer, but it says little about today's equilibrium rate. The relationship between prices and GDP per person may be a better guide to the current fair value of a currency. The adjusted index uses the "line of best fit" between Big Mac prices and GDP per person for 48 countries (plus the euro area). The difference between the price predicted by the red line for each country, given its income per person, and its actual price gives a supersized measure of currency under- and over-valuation.



**Answer:** The "Big Mac Index" is a comparison, promoted by *The Economist* magazine, of the price of a Big Mac around the world, normalized to the average price across the United States. It purports to be a comparison of world currency values using a comparable food item (a fairly standard fast-food hamburger) in different countries. A graphical interpretation quickly indicates countries with weak currencies (index > 50%), and those with strong currencies (index < 50%), compared to the US Dollar.

**Suggested Solutions to Sample Search List (DO NOT DISTRIBUTE THIS)**

Search and Find	Citation and Location Info/Photo/Capture/Photocopy
<p>What were the target dates for the United States to phase out household incandescent light bulbs? What authority (person or law) is making this happen?</p>	<p><i>Energy Independence and Security Act of 2007.</i>  <a href="http://en.wikipedia.org/wiki/Energy_Independence_and_Security_Act_of_2007">http://en.wikipedia.org/wiki/Energy_Independence_and_Security_Act_of_2007</a> (January 2017).</p>  <p><b>Answer:</b> 100-watt bulbs in January 2012, and 40-watt bulbs in January 2014. The Energy Independence and Security Act of 2007 (Pub.L. 110-140, originally named the Clean Energy Act of 2007) authorized this action, as well as many other energy-saving measures.</p>
<p>What major league baseball game had the fewest number of fans in the stands?</p>	<p><i>What is the lowest attendance for a major league baseball game?</i>  <a href="http://www.answers.com/Q/What_is_the_lowest_attendance_for_a_major_league_baseball_game">http://www.answers.com/Q/What_is_the_lowest_attendance_for_a_major_league_baseball_game</a> (January 2017).</p>  <p><b>Answer:</b> A probable answer: On September 8, 1916, the New York Yankees played the Philadelphia Athletics with 23 fans watching.</p>

**Suggested Solutions to Sample Search List (DO NOT DISTRIBUTE THIS)**

**Search and Find**

What was last month's average daily high temperature in your city, compared to the same month a year ago?

**Citation and Location Info/Photo/Capture/Photocopy**

For example, for Dallas, TX in December ("last month"): *Weather History for Dallas Texas, December.* [http://weather-warehouse.com/WeatherHistory/PastWeatherData\\_DallasRedbirdAirport\\_Dallas\\_TX\\_December.html](http://weather-warehouse.com/WeatherHistory/PastWeatherData_DallasRedbirdAirport_Dallas_TX_December.html) (January 2017).

**Weather Warehouse**  
Your Resource for Weather Information

Home About Us Articles Help/FAQ Products Weather Explorer Support

If you are using Internet Explorer, this page requires at least version 8 or higher to provide information

Click for More Historical Weather Data

Past Monthly Weather Data for Dallas, TX [Texas] ("Dallas Redbird Arpt") : DECEMBER, 2016

All months for this station: January February March April May June July August September October November December

Year	Lowest Temperature (F)	Highest Temperature (F)	Warmest Minimum Temperature (F)	Coldest Maximum Temperature (F)	Average Minimum Temperature (F)	Average Maximum Temperature (F)	Mean Temperature (F)	Total Precipitation (In)
2017	NR	NR	NR	NR	NR	NR	NR	NR
2016	15	82	68	29	41.4	59.4	50.4	0.59
2015	30	79	67	40	41.4	63.6	52.5	5.02
2014	28	79	62	40	45.0	59.8	52.4	1.23
2013	18	80	52	27	33.7	55.3	45.4	2.76
2012	21	82	67	33	40.7	62.7	51.7	2.09

**Answer:** 2016, 59.4 °F versus 2015, 63.6 °F

What alloys, if any, of Copper (Cu) are superconductive above 20°K?

Found in "Properties of Superconductors with  $T_c$  above 10K" in: Chemical Rubber Company. *Handbook of Chemistry and Physics 2007.* Cleveland, Ohio, 2008. (pg 12-66). (This might be discoverable online, too.)

12-66 Properties of Superconductors

B. Superconductors with  $T_c > 10K$

Substance	$T_c, K$	Crystal structure type
CTa	10.3	B1
CTa <sub>1-04</sub> W <sub>0-06</sub>	8.5-10.5	B1
C <sub>0.66</sub> Th <sub>0.15</sub> Y <sub>0.21</sub>	17	(C <sub>3</sub> Pu <sub>2</sub> )
C <sub>3</sub> Y <sub>2</sub>	11.5	(C <sub>3</sub> Pu <sub>2</sub> )
CW	10	B1
(Ca,Lu) <sub>2</sub> CuO <sub>4</sub>	16	(K <sub>2</sub> NiF <sub>4</sub> )
Cu(La,Sr) <sub>2</sub> O <sub>4</sub>	39	
Cu <sub>1-x</sub> Bi <sub>x</sub> Sr <sub>2</sub> O <sub>4</sub>	10.8	(Mo <sub>6</sub> PbS <sub>8</sub> )
Cr <sub>0.33</sub> SiV <sub>27</sub>	11.3	A15
GaNb	14.5	A15 (Cr-Si)

**Answer:** Yes, Cu(La,Sr)<sub>2</sub>O<sub>4</sub> has a critical temperature of 39K.

Suggested Solutions to Sample Search List (DO NOT DISTRIBUTE THIS)

Search and Find

Value of the binomial coefficient,  $\binom{45}{16}$ ?

Citation and Location Info/Photo/Capture/Photocopy

U.S. Department of Commerce. *Handbook of Mathematical Functions by the National Bureau of Standards*. Washington, D.C: U.S. Government Printing Office, November 1970 (Table 24.1, Pg 830). (This might be discoverable online, too.)

830 COMBINATORIAL

Table 24.1 BINOMIAL COEFFICIENTS

n\m	14	15	16
14	1		
15	15	1	
16	120	16	1
17	680	136	17
18	3060	816	153
19	11628	3876	969
20	38760	15504	4845
21	1 16280	54264	20349
22	5 19770	1 70544	74613
23	8 17190	4 90314	2 45157
24	19 61256	13 07504	7 35471
25	44 57400	32 68760	20 42975
26	96 57700	77 26160	53 11735
27	200 58300	173 83860	130 37895
28	401 16600	374 42160	304 21755
29	775 58760	775 58760	678 63915
30	1454 22675	1551 17520	1454 22675
31	2651 82525	3005 40195	3005 40195
32	4714 35600	5657 22720	6010 80390
33	8188 09200	10371 58320	11668 03110
34	13919 75640	18559 67520	22039 61430
35	23199 59400	32479 43160	40599 28950
36	37962 97200	55679 02560	73078 72110
37	61070 86800	93641 99760	1 28757 74670
38	96695 54100	1 54712 86560	2 22399 74430
39	1 50845 04396	2 51408 40660	3 77112 60990
40	2 32069 29840	4 02253 45056	6 28521 01650
41	3 52401 52720	6 34322 74896	10 30774 46706
42	5 28402 29080	9 86724 27616	16 45097 21602
43	7 83789 60360	15 15326 56696	26 51833 48218
44	11 49558 08528	22 99116 17056	41 67148 05914
45	16 68713 34960	34 48674 25584	64 66264 22970

Answer:  $\binom{45}{16} = 646,626,422,970$



## ACTIVITY: EVALUATING INFORMATION

### Instructor Preparation

The purpose of this activity is to encourage students to stop and think when they encounter information on the Internet. In the era of printed media, when material went through the process of editing, typesetting, proofing, and printing, it was likely that the published information was factual and trustworthy—as much as reasonably possible. So, the fruits of such research—probably from libraries and archives—were generally reliable as well. However, with the advent of the internet, hundreds of webpages are created daily, often with impressive text and graphics, made instantly available to the world, and subjected to little or no verification of veracity. In this activity, the goal is to suggest some guidelines and best practices to help students discern between fact and fiction.

**Note:** When searching the internet and visiting unfamiliar sites, up-to-date virus/spyware protection and caution are essential. Even legitimate websites may be supported by advertising links and unrelated imagery which should be ignored.

### Objectives

Students will:

1. Consider a website's URL and assess its legitimacy or potential maliciousness.
2. Evaluate the content of a webpage to assess its reliability and trustworthiness.

### Materials

- Handout 1—Evaluating Web-Based Information
- Handout 2—Sample Webpages to Review
- Handout 3—Checklist for Evaluating Website Reliability
- Computer lab with Internet access

### Activity Guidelines

People who routinely surf the web quickly develop a sense about the relative reliability of a visited website. Those who are not as savvy or who need practice evaluating will benefit from this activity.

Handout 1 can be distributed and discussed. Visit a few websites as a class and discuss the sites' features. The list of suggested websites in Handout 2 can serve as a starting point. Consider adding course-specific links related to the class subject or career field. (Some examples are provided.)

If students have access to the computer lab, Handouts 2 and 3 can be used for the hands-on portion of this activity. It would be helpful to have a webpage (or electronic file) prepared with the list of links for Handout 2 so that students don't have to type in the URLs.

Have students create a reliability score assigning points to each item on the checklist on Handout 3. They can then use those criteria to evaluate the content of each URL on Handout 2.

## Group Reflection Questions

Following the surfing activity, students should compare and discuss the findings (scores) and conclusions about each of the webpages or websites listed.

- Which of the criteria do you believe is most important in determining the truth and/or value of website content? Why?
- What evaluation criteria would you add to the list?
- In addition to finding misleading or false information, can you think of other reasons to be cautious about clicking on suspicious-looking URLs?

## Handout 1—Evaluating Web-Based Information

### Who is responsible?

- By its very nature, the Internet is a wide-open place. Nobody is responsible for the truthfulness, factuality, or reliability of any information or content..
- The success of any particular site is contingent on its acceptance and popularity, which, in turn, is generally dependent on its reliability and usefulness as a source of information, entertainment, communication, productivity, or for businesses profitability.
- Anybody can upload information to a webpage, often without accountability or responsibility.
- It is essential that reader critically evaluate the source, the possible bias, the evidence of accuracy and reliability, and the date of any information found and used.

### The Source

- The *Web Address* (uniform resource locator, i.e., URL) is a generally reliable source indicator. Read it carefully to discern the host computer and domain. The directory structure and file names are necessary details for returning to the page.
- The *Publisher or Sponsor* of the website, if well-known and reputable, generally lends credibility to the information found there.
- The *Author* of the webpage, if indicated, offers another indication of reliability.. Author/source contact information or a link to the author’s personal webpage or email allows further verification, assurances, and permissions for using content, if needed.
- The *Point of View or Bias* of the source can usually be determined with the help of the above information. For example, although business sites can often be good sources of information, be wary of sites that are promoting or selling products based on that information.



### Accuracy and Reliability

- Compared to books and other printed material, relatively little proofing and editing is applied to websites. Don’t be surprised to find typos, erroneous information, and even deception.
- Assess the quality of a site: spelling and grammar errors, citations/references for factual information, broken and out-of-date links.
- Try to confirm information from two or more independent sources. Multiple sites which merely copy each other do not qualify as confirmation.
- Blogs and discussion boards should never be treated as sources, as they are opinion-based rather than fact-based. If a blog links to an article in an online version of a printed source (e.g. newspaper, scientific journal, etc.), follow the link to the original source and evaluate it.
- Always ask, “How current is the webpage?” Information found on the web might be a few days old, reflecting the most current knowledge, or several decades old, representing very out-of-date information.
- Most text and imagery on the web is copyrighted. Copyright laws designed for written materials also apply to electronic publications. To avoid plagiarism (using someone else’s work as if it was your own), cite the source of material you use in your own works and seek permission from the owner to use it. Quality websites should do the same, as evidence that they respect copyright.

## Handout 2—Sample Webpages to Review

- **A Sample Webpage Demonstrating a Few HTML Features**  
[http://sheldonbrown.com/web\\_sample1.html](http://sheldonbrown.com/web_sample1.html)  
How reliable is this information?
- **Buying a telescope**  
<http://www.astronomy.com/Equipment/How-To/2004/07/Buying%20a%20telescope.aspx>  
Includes many historical facts and comparisons. Can these be trusted?
- **The Healthy Way Diet**  
<https://www.helpguide.org/articles/diet-weight-loss/healthy-weight-loss-and-dieting.htm>  
Nutritionally sound and medically researched information and advice to help you live healthier, happier lives. Sounds trustworthy, right?

## Handout 3—Checklist for Evaluating Website Reliability

Here is a checklist of criteria for evaluating websites to get you started. Create a reliability score based on these, assigning point values to each criterion. Use this checklist to evaluate the merits of the content at each URL on Handout 2.

- URL appears legitimate:** The server name is a reputable source, not merely similar to a reputable source. (For example, <http://cnn.com>, not <http://cnn-news.com>.)
- Sponsoring organization is included in the URL:** The sponsoring organization is reachable by truncating the directory information from the full-length URL. (For example, <http://www.cnn.com/2017/01/19/health/iyw-cancer-survivor-marathon-trnd/index.html> points to <http://www.cnn.com>.)
- Sponsor/author contact information:** Contact information is provided or can be verified; author or sponsor is well-known, and/or author can be reached if further information is needed. (For example, the page footer shows: © 2017 Cable News Network. Turner Broadcasting System, Inc. All Rights Reserved. )
- Author or source is reputable:** The information is from a well-known and respected source. (For example, “*Nearly one-third of U.S. households lack broadband service,*” found on the U.S. National Bureau of Standards website).
- Information is unbiased:** The source has no obvious reason to be biased to report dishonest information. On the other hand, commercial websites should be suspected of a bias intended to help sell their product. (For example, a table of recommended drill bit sizes found on a woodworking website is probably accurate information. A brand comparison of bits with links to purchase one particular brand should be investigated further, as the web author has a motive for posting promotional material—profit.)
- Information is verifiable:** Two or more independent and reliable sources confirm the information. Search engines make it easy to find and compare multiple sources of the same information. (For example, use a 13/64" drill bit for a #10 screw pilot hole in hard wood, as confirmed on several hardware and woodworking websites.)
- Information is current or up-to-date:** Information, especially about controversial topics, should be relatively current, as indicated by the reference information or copyright date. (For example, a presentation/report from the Berkley Lab titled “The Health and Environmental Benefits of Wind and Solar Energy in the United States, 2007-2015” [[https://eta.lbl.gov/sites/all/files/publications/health\\_and\\_environmental\\_benefits\\_wind\\_solar\\_2007-2015\\_jan2017.pdf](https://eta.lbl.gov/sites/all/files/publications/health_and_environmental_benefits_wind_solar_2007-2015_jan2017.pdf)] dated January 2017 is likely to be more current than a 20-year old article found elsewhere describing wind and solar energy.)
- Information includes references:** Information based on research or fact-finding should have source citations that could be used for independent verification. Include such citations as part of your own work in a footnote, bibliography, references-cited page.
- Website Quality Control:** Reliable websites generally will not include errors in spelling, grammar, and formatting, or obviously incorrect information or cross-references.
- Website ethics:** The links and advertisements contained on a website are usually a reflection of the quality of the content. Reputable websites usually link to other reputable sites.

## ACTIVITY: ORGANIZING INFORMATION

### Instructor Preparation

In this activity students will consider a few ways to collect and manage information gleaned from one or more websites. Collecting and managing a handful of pertinent facts and their citation information from print or electronic sources is not difficult. But when gathering hundreds or thousands of pieces of data—for sorting, analyzing, charting, or presenting—more data-handling skills are needed. This activity provides an opportunity for students to import data into a common spreadsheet (MS Excel recommended). (A similar strategy would also work for importing into a database program, although that is beyond the scope of this activity.) Rather than use contrived data to serve our purpose, this activity offers some guidelines for importing a collection of real-world sports statistics, which will interest many of your students, regardless of their career aspirations. Alternatively, you may use data sources from your own content or career area and modify the instructions to fit those sources. Here are some suggested data sources addressing specific fields:

- Corporate finance and valuation data for thousands of firms can be found online from many sources. Much of this data is already saved in spreadsheet format, but there is so much data that it will be necessary to extract a subset to make it more manageable. A good place to start is Google Finance (<https://www.google.com/finance>), which provides information on both United States and world markets.
- Government and public administration careers deal with large volumes of data routinely. For example, the U.S. Census Bureau makes available large amounts of data of all kinds on their website, <http://www.census.gov/sdc/>. Much of this is now available already saved to a spreadsheet format, making it very easy to incorporate into one’s own data accumulations.
- Science and technology careers yield data from all sorts of research, whether it is space (<http://eosweb.larc.nasa.gov/>), weather (<http://www.ncdc.noaa.gov/>), health (<http://www.cdc.gov/rdc/>), or almost any area of interest. Focus on a particular area of research and data collection. The activities that follow will build on the students’ imported data, so be sure they save their work (e.g., on a USB flash drive, hard drive, or other media), and save it at the end of the class.

All Races	Total	None	1st - 4th grade	5th - 6th grade
Both Sexes				
18 years and over	234,719	834	1,764	3,618
18 to 24 years	30,140	39	76	114
25 years and over	204,579	796	1,688	3,504
25 to 29 years	20,893	30	88	223
30 to 34 years	20,326	54	100	348
35 to 39 years	19,140	54	124	398

U.S. Department of Commerce  
National Oceanic & Atmospheric Administration  
National Environmental Satellite, Data, and Information Service

**Summary of Monthly Normals 1981-2010**

Station: WACO REGIONAL AIRPORT, TX US

Month	Mean						Temperature (°F)					
							Cooling Degree Days					
	Daily Max	Daily Min	Mean	Long Term Max Std. Dev.	Long Term Min Std. Dev.	Long Term Avg Std. Dev.	55	57	60	65	70	72
1	58.2	36.1	47.2	4.7	3.5	3.7	37	25	13	3	-7777	-7777
2	61.8	39.8	50.8	4.9	3.1	3.7	69	43	24	6	1	-7777
3	69.2	47.1	58.2	2.9	3.2	2.6	175	133	87	34	8	4
4	77.6	54.6	66.1	2.8	3.4	2.7	343	289	214	111	42	24
5	85.0	63.8	74.4	2.6	2.4	2.5	601	540	447	298	184	119
6	91.7	70.8	81.3	2.5	2.3	2.5	787	727	637	488	338	205

## Objectives

Students will:

Copy/paste tabular data from a webpage into a spreadsheet.

## Materials

- An internet-connected computer, equipped with a browser and spreadsheet program (e.g., MS Excel)—computer lab time is recommended
- Storage media for spreadsheet file (e.g., network drive, USB flash drive, or similar)
- Handouts 1 and 2
- Instructor computer with projector, for demonstrating Excel functions

## Activity Guidelines

Some websites will have their data already saved to a spreadsheet format. However, for this activity, students learn how to copy tabulated or listed data and save it to a tabular format in a spreadsheet. Students can practice by using a major league baseball website that posts game statistics (<http://www.baseball-reference.com/>) for this activity. Once there, find a favorite baseball team. Scroll down the page to find an abundance of batting, pitching, and fielding statistics. If the season has just begun, use the data from the previous season (see link to previous year). The goal is to copy the data for the team batting statistics—the first table on the page. Here’s an excerpt of a typical table:

Rk	Pos	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	
1	C	Mike Napoli	30	108	417	352	53	80	9	2	24	56	1	0	56	.227	.343	.469	.812	110	165	9	7	0	2	5	
2	1B	Mitch Moreland*	26	114	357	327	41	90	18	0	15	50	1	1	23	.275	.321	.468	.789	104	153	8	1	2	4	5	
3	2B	Ian Kinsler	30	157	731	655	105	168	42	5	19	72	21	9	60	.256	.326	.423	.749	95	277	14	10	1	5	0	
4	SS	Elvis Andrus	23	158	711	629	85	180	31	9	3	62	21	10	57	.286	.349	.378	.727	91	238	15	5	17	3	0	
5	3B	Adrian Beltre	33	156	654	604	95	194	33	2	36	102	1	0	36	.321	.359	.561	.921	137	339	8	5	0	9	8	
6	LF	David Murphy*	30	147	521	457	65	139	29	3	15	61	10	5	54	.304	.380	.479	.859	124	219	7	4	0	4	7	
7	CF	Josh Hamilton*	31	148	636	562	103	160	31	2	43	128	7	4	60	.285	.354	.577	.930	139	324	9	5	0	9	13	
8	RF	Nelson Cruz	31	159	642	585	86	152	45	0	24	90	8	4	48	.260	.319	.460	.779	101	269	7	5	0	4	2	
9	DH	Michael Young	35	156	651	611	79	169	27	3	8	67	2	2	33	.277	.312	.370	.682	78	226	26	1	0	6	3	
Rk	Pos	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	
10	CF	Craig Gentry	28	122	269	240	31	73	12	3	1	26	13	7	14	.304	.367	.392	.759	100	94	4	10	5	0	1	
11	C	Yorvit Torrealba	33	49	182	161	16	38	8	0	3	12	1	1	14	.236	.302	.342	.643	69	55	2	2	3	2	0	
12	C	Geovany Soto	29	47	164	148	19	29	6	0	5	25	1	0	11	.196	.253	.338	.591	53	50	6	1	2	2	0	
13	UT	Brandon Snyder	25	40	68	65	11	18	2	0	3	9	0	0	3	.26	.277	.309	.446	755	95	29	1	0	1	0	0

This data will be used in later activities so it will be helpful to note the meanings of the column headings by hovering the mouse cursor over each column heading to find G: Games played, PA: Plate appearances, and so forth. Handout 1 summarizes these descriptions. To simplify this activity we’ll focus only on the batters, i.e., the top half of this batting table. Ignore the bottom half of this batting table populated with the pitcher (Positions “Pos” denoted by “P”) data.

### Instructor demonstration:

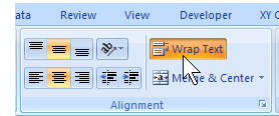
Click-and-drag with the mouse to highlight the entire row (from the top left to the far right cells) for batters (not pitchers) in the top half of the table. Release the mouse button and then use the menu (or right-mouse click) to copy the highlighted cells to the clipboard. If you have problems, click “away” and repeat this step, while clicking and dragging across slightly different locations on screen until all the batter data is obtained.

Instructor demonstration:

Now, open the spreadsheet program, and with the cursor positioned in cell A1 of a blank worksheet, paste the contents of the clipboard, (the batting table) into the Excel worksheet. This may be accomplished simply with “Edit, Paste,” or “right-click, Paste.” If you don’t get the desired result, use the Undo command and try again. The result should be that the columns of batting data values are in the corresponding columns of the spreadsheet, for example, as shown here.

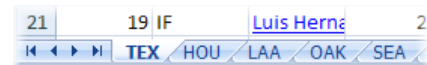
	A	B	C	D	E	F	G	H
1	Rk	Pos		Age	G	PA	AB	R
2	1	C	Mike Nap	30	108	417	352	53
3	2	1B	Mitch Mor	26	114	357	327	41
4	3	2B	Ian Kinsle	30	157	731	655	105
5	4	SS	Elvis Andr	23	158	711	629	85
6	5	3B	Adrian Be	33	156	654	604	95
7	6	LF	David Mur	30	147	521	457	65
8	7	CF	Josh Hami	31	148	636	562	103
9	8	RF	Nelson Cr	31	159	642	585	86
10	9	DH	Michael Yi	35	156	651	611	79
11	Rk	Pos		Age	G	PA	AB	R
12	10	CF	Craig Gen	28	122	269	240	31
13	11	C	Yorvit Tori	33	49	182	161	16
14	12	C	Geovany S	29	47	164	148	19
15	13	UT	Brandon S	25	40	69	65	11
16	14	1T	Alberto G	29	24	55	54	7

If long player names wrap and fill multiple lines, causing the rows to have different sizes, try this: With the pasted cells still selected, adjust the format of the selected cells to “NOT Wrap Text.” Look in the Alignment section/tab in Excel and un-check the box, or un-depress the button (as shown here). This should make the row height uniform for all rows, as shown above.



Student activity:

Since one team’s batting performance is not very interesting, return to the website and repeat the above process for the remaining four teams in the division selected. Paste each team’s data into a new worksheet. Identify each worksheet with the team name by renaming the worksheet tabs at the bottom of the page. (Hint: Double-click the tab name or right-click and choose Rename.) See the example tab names shown here.



When done, save the spreadsheet, using a meaningful file name. For example, save it as “Team\_Batting\_Stats” (or similar). This work will be used in the next activities.

**Group Reflection Questions**

- How long would it take to accumulate all the batting stats for the whole league? (Answer: Just a few minutes. And no typing required—just copy/paste. Pretty powerful.)
- Do all teams have the same number of batters? (Answer: No, they vary, and appear to have about 20 or so—some more, some less—depending on how many pitchers the team keeps.)
- What are some obvious advantages of using copy/paste, compared to retyping the stats from a printout? (Answer: Probably the most significant advantage is *quality control*—little or no chance for errors to creep in. Another answer is *speed*. A quick multiplication of **rows** × **columns** × **teams** will show that each student’s spreadsheet, created in just a few minutes, contains almost 3000 pieces of information—for just five teams. With just a little more effort, students could easily accumulate the batting stats for the whole league—about 10,000 pieces of data!).



## Handout 1—Copying and Pasting Tabular Data from a Webpage into a Spreadsheet

Some websites have their data already saved to a spreadsheet format. For this activity, however, you will learn how to copy tabulated or listed data and save it in an Excel spreadsheet. Practice by using a major league baseball website that posts game statistics (<http://www.baseball-reference.com/>) for this activity. Once there, find a favorite baseball team. Scroll down the page to find an abundance of batting, pitching, and fielding statistics. If the season has just begun, use the data from the previous season (see link to previous year). The goal is to copy the data for the team batting statistics—the first table on the page. Here’s an excerpt of a typical table from the website:

Rk	Pos	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB
1	C	Mike Napoli	30	108	417	352	53	80	9	2	24	56	1	0	56	.227	.343	.469	.812	110	165	9	7	0	2	5
2	1B	Mitch Moreland*	26	114	357	327	41	90	18	0	15	50	1	1	23	.275	.321	.468	.789	104	153	8	1	2	4	5
3	2B	Ian Kinsler	30	157	731	655	105	168	42	5	19	72	21	9	60	.256	.326	.423	.749	95	277	14	10	1	5	0
4	SS	Elvis Andrus	23	158	711	629	85	180	31	9	3	62	21	10	57	.286	.349	.378	.727	91	238	15	5	17	3	0
5	3B	Adrian Beltre	33	156	654	604	95	194	33	2	36	102	1	0	36	.321	.359	.561	.921	137	339	8	5	0	9	8
6	LF	David Murphy*	30	147	521	457	65	139	29	3	15	61	10	5	54	.304	.380	.479	.859	124	219	7	4	0	4	7
7	CF	Josh Hamilton*	31	148	636	562	103	160	31	2	43	128	7	4	60	.285	.354	.577	.930	139	324	9	5	0	9	13
8	RF	Nelson Cruz	31	159	642	585	86	152	45	0	24	90	8	4	48	.260	.319	.460	.779	101	269	7	5	0	4	2
9	DH	Michael Young	35	156	651	611	79	169	27	3	8	67	2	2	33	.277	.312	.370	.682	78	226	26	1	0	6	3
Rk	Pos	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB
10	CF	Craig Gentry	28	122	269	240	31	73	12	3	1	26	13	7	14	.304	.367	.392	.759	100	94	4	10	5	0	1
11	C	Yorvit Torrealba	33	49	182	161	16	38	8	0	3	12	1	1	14	.236	.302	.342	.643	69	55	2	2	3	2	0
12	C	Geovany Soto	29	47	164	148	19	29	6	0	5	25	1	0	11	.196	.253	.338	.591	53	50	6	1	2	2	0
13	UT	Brandon Snyder	25	40	69	65	11	18	2	0	3	8	0	0	3	.267	.277	.446	.755	85	28	1	0	1	0	0

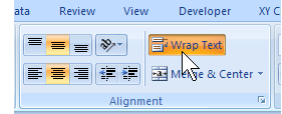
This data will be used in later activities so it will be helpful to note the meanings of the column headings by hovering the mouse cursor over each column heading to find G: Games played, PA: Plate appearances, and so forth. Handout 2 summarizes these descriptions. Focus only on the batters, i.e., the top half of this batting table. Ignore the bottom half of this batting table populated with the pitcher (Positions “Pos” denoted by “P”) data.

- Click-and-drag with the mouse to highlight the entire row (from the top left to the far right cells) for batters (not pitchers) in the top half of the table. Release the mouse button and then use the menu (or right-mouse click) to copy the highlighted cells to the clipboard. If you have problems, click away and repeat this step until all the batter data is obtained.

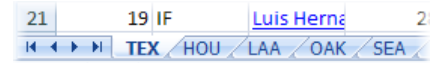
- Now, open the spreadsheet program, and with the cursor positioned in cell A1 of a blank worksheet, paste the contents of the clipboard, (the batting table) into the worksheet. This may be accomplished simply with “Edit, Paste,” or “right-click, Paste.” If you don’t get the desired result, use the Undo command and try again. The result should be that the columns of batting data values are in the corresponding columns of the spreadsheet, for example, as shown here.

	A	B	C	D	E	F	G	H
1	Rk	Pos		Age	G	PA	AB	R
2	1	C	Mike Napoli	30	108	417	352	53
3	2	1B	Mitch Moreland*	26	114	357	327	41
4	3	2B	Ian Kinsler	30	157	731	655	105
5	4	SS	Elvis Andrus	23	158	711	629	85
6	5	3B	Adrian Beltre	33	156	654	604	95
7	6	LF	David Murphy*	30	147	521	457	65
8	7	CF	Josh Hamilton*	31	148	636	562	103
9	8	RF	Nelson Cruz	31	159	642	585	86
10	9	DH	Michael Young	35	156	651	611	79
11	Rk	Pos		Age	G	PA	AB	R
12	10	CF	Craig Gentry	28	122	269	240	31
13	11	C	Yorvit Torrealba	33	49	182	161	16
14	12	C	Geovany Soto	29	47	164	148	19
15	13	UT	Brandon Snyder	25	40	69	65	11
16	14	UT	Alberto Gonzalez	29	24	55	54	7

If long player names wrap and fill multiple lines, causing the rows to have different sizes, try this: With the pasted cells still selected, adjust the format of the selected cells to “NOT Wrap Text.” Look in the Alignment section/tab in Excel and un-check the box, or un-depress the button (as shown here). This should make the row height uniform for all rows, as shown above.



- c. Since one team’s batting performance is not very interesting, return to the website and repeat the above process for the remaining four teams in the division selected. Paste each team’s data into a new worksheet. Identify each worksheet with the team name by renaming the worksheet tabs at the bottom of the page. (Hint: Double-click the tab name or right-click and choose Rename.) See the example tab names shown here.
- d. When done, save the spreadsheet, using a meaningful file name. For example, save it as “Team\_Batting\_Stats” (or similar). This work will be used in the next activities.



## Handout 2—Column Headings Key for Baseball Statistics

Label	Meaning
Rk	Rank. This is a count of the rows from top to bottom. It is recalculated following the sorting of a column.
Pos	Position
Age	Player's age at midnight of June 30th of that year
G	Games Played or Pitched
PA	Plate Appearances. When available, we use actual plate appearances from play-by-play game accounts. Otherwise estimated using $AB + BB + HBP + SF + SH$ , which excludes catcher interferences.
AB	At Bats
R	Runs Scored/Allowed
H	Hits/Hits Allowed
2B	Doubles Hit/Allowed
3B	Triples Hit/Allowed
HR	Home Runs Hit/Allowed
RBI	Runs Batted In
SB	Stolen Bases
CS	Caught Stealing
BB	Bases on Balls/Walks
SO	Strikeouts
BA	Hits/At Bats. For recent years, leaders need 3.1 PA per team game played. Bold indicates highest BA using current stats. Gold means awarded title at end of year.
OBP	$(H + BB + HBP)/(At\ Bats + BB + HBP + SF)$ . For recent years, leaders need 3.1 PA per team game played.
SLG	Total Bases/At Bats or $(1B + 2*2B + 3*3B + 4*HR)/AB$ . For recent years, leaders need 3.1 PA per team game played.
OPS	On-Base + Slugging Percentages For recent years, leaders need 3.1 PA per team game played.
OPS+	$OPS + 100 * [OBP / lg\ OBP + SLG / lg\ SLG - 1]$ . Adjusted to the player's ballpark(s).
TB	Total Bases. Singles + 2 x Doubles + 3 x Triples + 4 x Home Runs.
GDP	Double Plays Grounded Into. Only includes standard 6-4-3, 4-3, etc. double plays. First tracked in 1933. For gamelogs only in seasons we have play-by-play, we include triple plays as well. All official seasonal totals do not include GITP's.
HBP	Times Hit by a Pitch.
SH	Sacrifice Hits (Sacrifice Bunts)
SF	Sacrifice Flies. First tracked in 1954.
IBB	Intentional Bases on Balls. First tracked in 1955.
Positions	C: Catcher, 1B: first base, 2B: second base, 3B: third base, SS: short stop, LF: left field, CF: center field, RF: right field, DH: designated hitter, P: pitcher, UT: utility, OF: outfield, IF: infield, CI: corner infield, MI: middle infield

## ACTIVITY: MANAGING AND PRESENTING INFORMATION

### Instructor Preparation

In this activity students will consider ways to view and manage large volumes of data and to present it in an understandable way. Some obvious things to do with data are to sort it, extract selected subsets (filter), perform some basic calculations (e.g., average), and generate graphs. These processes for presenting and interpreting large data volumes used to be tedious and time-consuming, but are easy today with the modern spreadsheet, enabling calculations, charting, and analysis. MS Excel is used in this activity and instructors must have a strong working knowledge of the program's essential features.

Although we'll continue the discussion using the baseball statistics data from the previous activity, the techniques presented here will apply to any data sets that you may have opted to use for your own course topic. For example:

- Education and Training data sets will permit you to sort and filter state graduation data ([http://www.doe.virginia.gov/statistics\\_reports/graduation\\_completion/cohort\\_reports/index.shtml](http://www.doe.virginia.gov/statistics_reports/graduation_completion/cohort_reports/index.shtml)) or various kinds of national educational data (<https://nces.ed.gov/quicktables/>).
- Transportation and Distribution Logistics data presents the results of deliberate movements of millions of parcels and other material around the state, the nation, or the world. Watching for trends and patterns of disruption or smooth handling is greatly facilitated by a rapid analysis of timely data.
- Data reflecting Agriculture, Food, and Natural Resource production levels can fluctuate greatly with changes in weather, treatments, production methods, or even legislation and marketing pressures. Analyzing this data is facilitated by using spreadsheet charting features.

### Objectives

Students will:

1. Sort data in a spreadsheet.
2. Filter data in a spreadsheet.
3. Perform basic calculations on data, such as average, standard deviation, and frequency counts.
4. Generate simple graphs of data.

### Materials

- An internet-connected computer, equipped with a browser and spreadsheet program (e.g., MS Excel)—computer lab time is recommended
- Storage media for spreadsheet file (e.g., network drive, USB flash drive, or similar)
- Handouts 1 and 2
- Instructor computer with projector, for demonstrating Excel functions

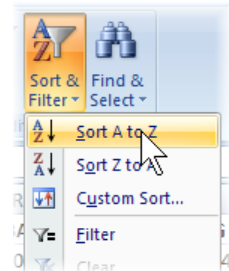
## Activity Guidelines

Returning to the spreadsheet created in the previous activity, this will be a quick activity to demonstrate how easy it is to manipulate the data to reveal various interpretations. Before doing anything, point out to students the relatively random appearance of this large volume of data. In its present state, it is virtually impossible to see any useful patterns. We want them to discover how easy it is to see the same data in an entirely different light with a few simple actions. Charting in Excel is a very powerful tool with hundreds of options and ways to customize a graph's appearance. These exercises demonstrate some of the power available to manage large amounts of information more easily than might be expected.

**NOTE:** The baseball statistics tables may include a duplicate header row following the first nine players. It will be helpful to delete that unneeded row on each worksheet. Also, once data order is changed or moved around, it's not always simple to return to the original order. However, the original order should not be required for these activities. The header row is important however, so it should remain on row 1 (or at the top of the table).

### Sorting

In any table, sorting is a natural step, and a readily applied feature in a spreadsheet. Simply place the cursor in a column, and from the *Home* tab, in the *Editing* group, click the *Sort and Filter* button, then *Sort A to Z* (i.e., ascending) order. In the table of batting statistics for one of the teams:



#### **Instructor demonstration:**

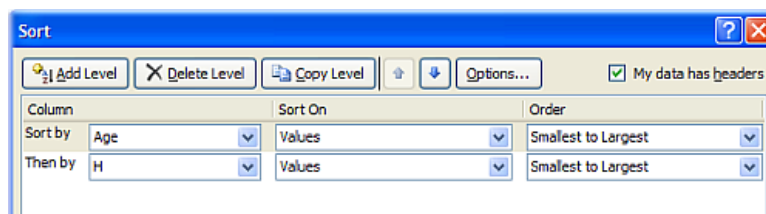
Sort the batters by Rank. Place the cursor in the Rank column (column A), and sort it in ascending order. Notice that 1) the whole table (all columns) is sorted, with the order of the rows being controlled by the Rank column; 2) the column headings may drop to the bottom of table (because the letters "Rk" are lower in the sorting order than digits).

#### **Student activity:**

Sort the batters by Age. Place the cursor in the Age column (column D), and sort it in ascending order. Examine the span of ages, from youngest to oldest. Are you surprised? Do all teams look the same in this respect?

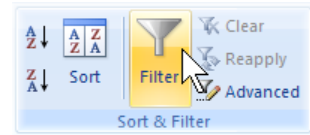
#### **Student activity:**

Sort the batters by both Age and Runs. Return to the spreadsheet state (Hint: "undo" the above sorts.) with the headers in row 1. Click *Custom Sort...*, and check the box for "My data has headers." Sort the data with two levels: first, sort by the Age column, and then by the H column (number of home runs), both from *Smallest to Largest*, as shown below. Do you see what this accomplishes? Does this work for all teams?



## Filtering

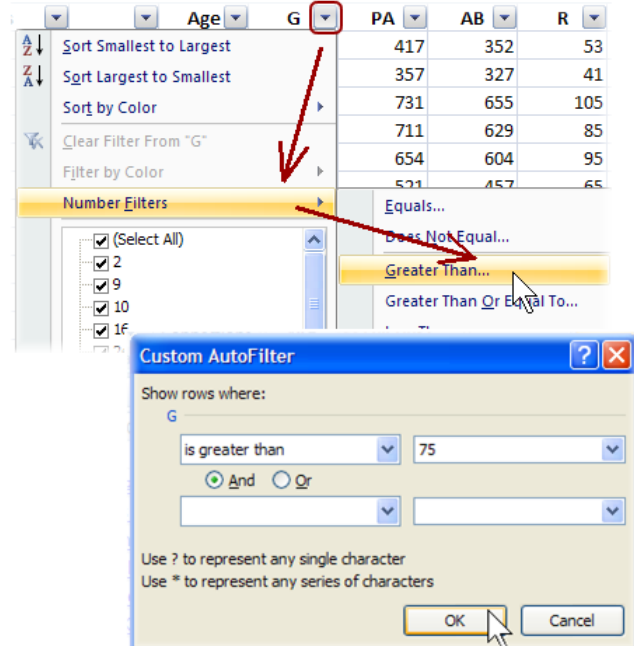
The *Filter* feature provides a powerful, almost database-like capability. It allows you to perform simple queries to extract rows that meet certain criteria or hide rows that do not meet the criteria. The *Filter* function is available on the same *Sort and Filter* button (see above), or under the *Data* tab in the *Sort and Filter* group, as the *Filter* button. Although we have not used it here, the *Advanced* function supports exporting the matching data to another part of the spreadsheet and/or using advanced criteria statements saved elsewhere in the spreadsheet.



Note: Place the cursor in the table of values before clicking *Filter*. If the “filter arrows” appear beside values rather than the column headings (as shown below, click and remove the *Filters*, then manually select the entire table, and click *Filter* again.

### Instructor demonstration:

View batters who played more than 75 games. Click the *Filter* button (or menu choice). Notice the new functionality added to each column heading. As shown, click the down-arrow for the column headed *G* (for “Games played”), then *Number Filters*, and then *Greater Than...* In the *Custom Autofilter* dialog box, in the blank box to the right of “G...is greater than...” type the value “75”, and click *OK*. Your table should now be showing only batters who played at least 76 games.

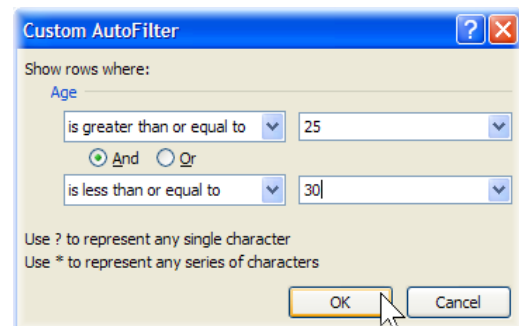


### Student activity:

View batters whose age is 30 years or older. Clear any existing filter by clicking the *Filter* function *Off*, and then back *On* again. Just as you applied the games-played filter, click the “down-arrow” for the *Age*, then *Number Filters*, and then *Greater Than or Equal To...* In the blank box to the right of “Age...is greater than or equal to...” type the value “30”, and click *OK*. Your table should be showing only the players age 30 and older.

### Student activity:

View batters whose age is between 25 and 30 who played more than 100 games. This filter will require several criteria that must be satisfied. Clear any existing filter, and then click the *Filter* function *On* again. Click the down-arrow for the *Age*, and specify a *Number Filters*, for *Age...is greater than or equal to...25*. Now restrict the ages. With *And* selected as shown here, click the drop-down list to choose *is less than or equal to*, enter the value 30, and click *OK*. Now add a Games played filter. As before, add a *Number Filter* for values of “G...is greater than...100.” How many players satisfy these criteria in the division?



### Calculating

Spreadsheets are very good at performing calculations on small or large quantities of data using simple or powerful functions.

**Instructor demonstration:**

Which is the “youngest team” in your division? Calculate the average age for each team using the *AVERAGE()* function. In a blank cell below the table, type “=AVERAGE(” without the quotes, and then use the mouse to select the entire range of age values for the team, and *Enter*. See the calculated average value. Repeat for the other teams. Which team in the division might be said to be the youngest?

	A	B	C	D	E	F
1	Rk	Pos		Age	G	PA
2	1	C	Mike Nap	30	108	417
3	2	1B	Mitch Mor	26	114	357
4	3	2B	Ian Kinsle	30	157	731
5	4	SS	Elvis Andr	23	158	711
6	5	3B	Adrian Be	33	156	654
7	6	LF	David Mur	30	147	521
8	7	CF	Josh Ham	31	148	636
9	8	RF	Nelson Cr	31	159	642
10	9	DH	Michael Y	35	156	651
11	10	CF	Craig Gen	28	122	269
12	11	C	Yorvit Tor	33	49	182
13	12	C	Geovany S	29	47	164
14	13	UT	Brandon S	25	40	69
15	14	UT	Alberto G	29	24	55
16	15	CF	Leonys M	24	24	52
17	16	UT	Mike Olt	23	16	40
18	17	C	Luis Marti	27	10	19
19	18	MI	Jurickson	19	9	17
20	19	IF	Luis Hern	28	2	2

**Student activity:**

What is the age of the oldest player on each team? The youngest player? Just as you calculated the average, use the *MAX()* and *MIN()* functions.

**Student activity:**

Which team scored the most runs all season? The fewest? Use the *SUM()* function on the “R” (runs scored) column.

**Instructor demonstration plus student activity:**

Create a Frequency Distribution of player ages. Use the *FREQUENCY()* function to create tallies for a histogram. Here are the steps:

Below the data, in some blank rows, create a column of ages: 15, 20, 25, 30, 35, 40, 45. These will be the “bins” for counting the number of players whose ages are in each age category, e.g., up to 15 years, from 16 up to 20, from 21 up to 25, etc.

Highlight the blank cells beside the bins. You will next enter what’s called an “array formula” in the highlighted cells.

With the blank cells still highlighted, type =FREQUENCY( and then 1) use the mouse to select the entire range of age values for the team (it’s OK to include the header row and blank rows),  
 2) type a comma,  
 3) use the mouse to select the range of bins, and  
 4) To complete array formulas, don’t press *Enter*, but instead press *Ctrl+Shift+Enter*.

To edit (or delete) this array formula, again highlight all the cells in the array, make your edit, and again press *Ctrl+Shift+Enter* (or *Delete*).

The array formula and resulting frequency values will fill the selected blank cells. These frequency values will be used to create a histogram (bar chart) below.

Steve Pea	29	21	75	63	2	16
Brandon L	24	17	37	35	2	9
Avg:	26.72					
Max:	36					
Min:	22					
Runs:	568					

### Presenting

Just as spreadsheets are good at calculating, they are also good at making graphs and charts.

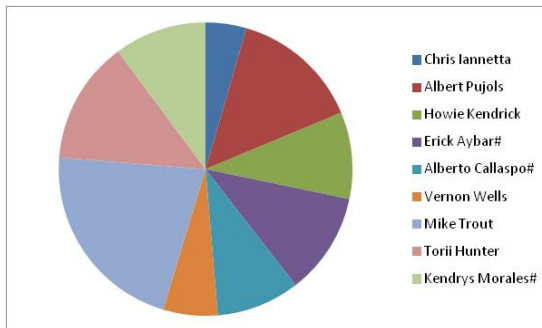
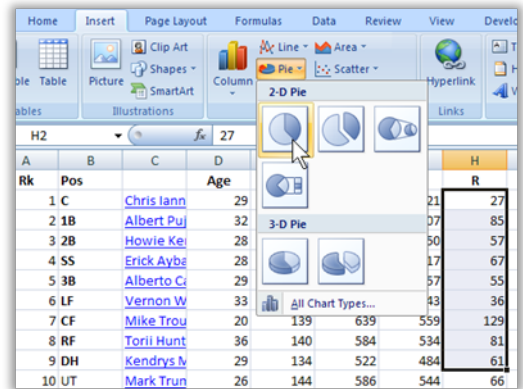
**Instructor demonstration plus student activity:**

Create a pie chart showing the starting players scoring performance.

Highlight the runs scored for the first 9 ranked players in the column headed by “R.”

On the *Insert* tab, click *Pie* (in the *Charts* group) and choose one of the pie chart formats. Excel calculates the percentages and creates all the sections of the pie chart for you.

Right-click the pie chart that appears and choose *Select Data...* Under *Horizontal (Category) Axis Labels*, click *Edit*. Then click/drag to and highlight the player names corresponding to the first nine positions. Click *OK*, and *OK*.



The resulting pie chart will now associate each player name with a colored section. You can further customize the chart appearance (title, shadows, labels, colors) as you wish.

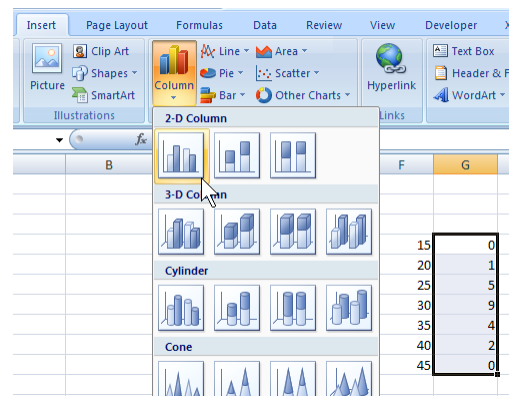
**Optional activity:**

Make a bar chart showing the distribution of player ages using the results of the frequency array formula calculated previously.

Just as for the pie chart, start by highlighting the data values to be charted: the column of Frequency calculations (beside the bins of age values).

On the *Insert* tab, click on *Column* (in the *Charts* group) and choose one of the formats. Excel creates the bar chart using the frequency counts.

Just as for the pie chart, right-click the new bar chart and choose *Select Data...* Under *Horizontal (Category) Axis Labels*, click *Edit*. Then click/drag to highlight the bin values corresponding to the frequency values. Click *OK*, and *OK*. Excel labels the bars with the bin values.



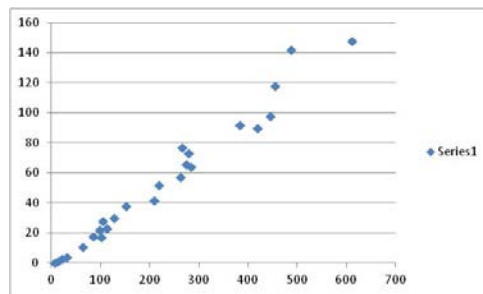
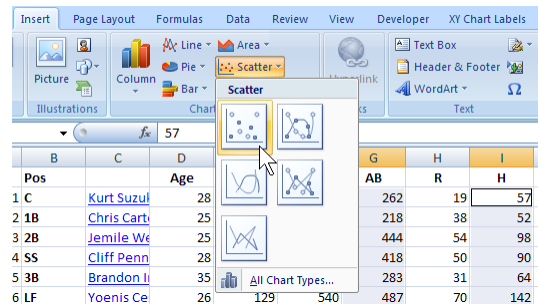


**Optional activity:**

Make a scatter graph of the number of times at bat versus the number of hits. You can probably anticipate the outcome here, but it will be a good check of graphing skills.

As for the other charts, start by highlighting the data values to be charted. But a scatter graph needs both *x*- and *y*-values. Start by dragging the mouse to highlight the *x*-values: the *At Bats* (headed by “AB”) column of values. Then, while holding down the Ctrl-key, drag the mouse to highlight the *y*-values—the *Hits* (headed by “H”) column of values. Release the Ctrl-key.

On the *Insert* tab, click on *Scatter* (in the *Charts* group), and choose the *Scatter with Only Markers* format.



The resulting scatter chart suggests a relationship between the number of times at bat (the horizontal axis) and the number of hits (the vertical axis).

You can further edit the characteristics of the chart’s titles, labels, and legend.

Explore other relationships to the *At Bats* values. Click the scatter chart and notice the blocks of data values selected on the spreadsheet are outlined. Carefully position the mouse over the outline around the *Hits*, mouse-down and drag the outline over to another column of values and release. The chart instantly updates with the new scatter comparison. Try out several columns—especially the *Batting Average* (headed by “BA”). Does a player’s batting average seem to improve when they have more times at bat?

If time permits, encourage students to replicate their work for all five teams in the division (or if using a different data set, as much data as is practical) to convey the power of using the spreadsheet to manage and present their data. This may not seem like a lot of data used in this activity, but it is far more than anyone would care to handle with a handheld calculator. Even the graphing was accomplished in a matter of a few minutes.

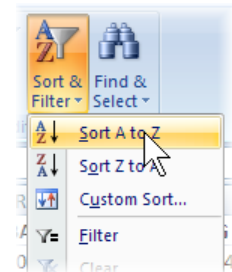
**Group Reflection Questions**

- Did the quick analysis done in this activity produce any surprising discoveries?
- What are some interesting comparisons or calculations that come to mind for this data?

## Handout 1—Sorting/Filtering Data, Performing Calculations, and Generating Graphs and Charts

### Sorting

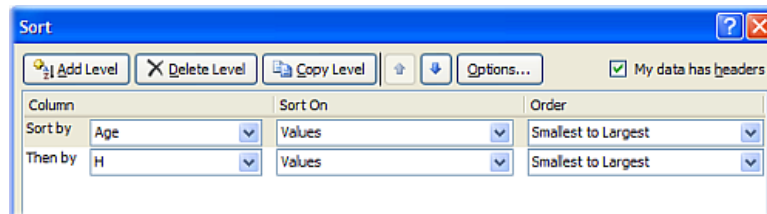
In any table, sorting is a natural step, and a readily applied feature in a spreadsheet. Simply place the cursor in a column, and from the *Home* tab, in the *Editing* group, click the *Sort and Filter* button, then *Sort A to Z* (i.e., Ascending) order. In the table of batting statistics for one of the teams:



**Sort the batters by Rank.** Place the cursor in the Rank column (column A), and sort it in ascending order. Notice that 1) the whole table (all columns) is sorted, with the order of the rows being controlled by the Rank column; 2) the column headings may drop to the bottom of table (because the letters “Rk” are lower in the sorting order than digits).

**Sort the batters by Age.** Place the cursor in the Age column (column D), and sort it in ascending order. Examine the span of ages, from youngest to oldest. Are you surprised? Do all teams look the same in this respect?

**Sort the batters by both Age and Runs.** Return to the spreadsheet state (Hint: “undo” the above sorts.) with the headers in row 1. Click *Custom Sort...*, and check the box for “My data has headers.” Sort the data with two levels: first, sort by the Age column, and then by the H column (number of home runs), both from *Smallest to Largest*, as shown below. Do you see what this accomplishes? Does this work for all teams?

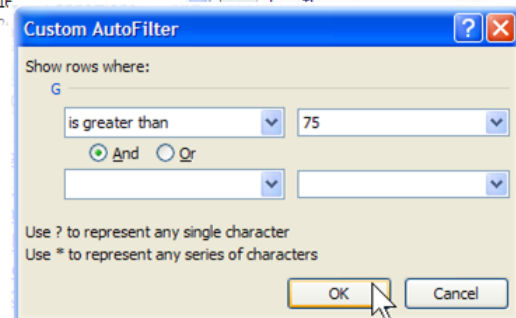
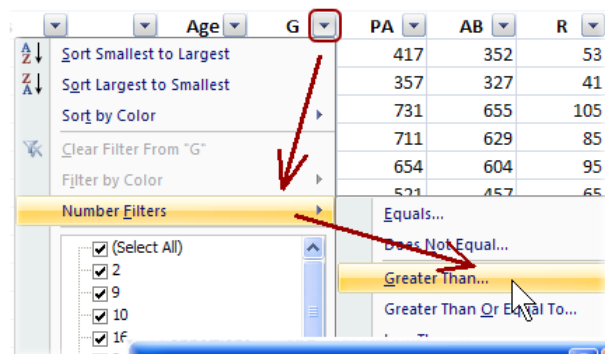


### Filtering

The *Filter* feature provides a powerful, almost database-like capability. It allows you to perform simple queries to extract rows that meet certain criteria or hide rows that do not meet the criteria.

The *Filter* function is available on the same *Sort and Filter* button (see above), or under the *Data* tab, in the *Sort and Filter* group, as the *Filter* button.

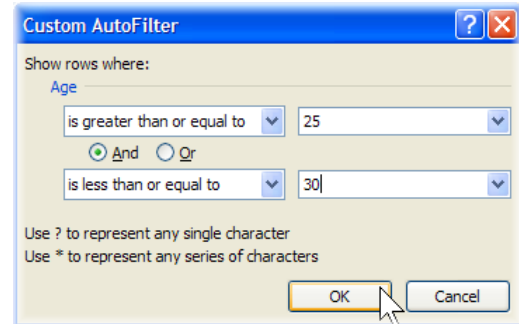
Note: Place the cursor in the table of values before clicking *Filter*. If the “filter arrows” appear beside values rather than the column headings (as shown), click and remove the *Filters*, then manually select the entire table, and click *Filter* again.



**View batters who played more than 75 games.** Click the *Filter* button (or menu choice). Notice the new functionality added to each column heading. As shown, click the down-arrow for the column headed *G* (for “Games played”), then *Number Filters*, and then *Greater Than...* In the *Custom Autofilter* dialog box, in the blank box to the right of “*G...is greater than...*” type the value “75”, and click *OK*. Your table should now be showing only batters who played at least 76 games.

**View batters whose age is 30 years or older.** Clear any existing filter by clicking the *Filter* function *Off*, and then back *On* again. Just as you applied the games-played filter, click the “down-arrow” for the *Age*, then *Number Filters*, and then *Greater Than or Equal To...* In the blank box to the right of “*Age...is greater than or equal to...*” type the value “30”, and click *OK*. Your table should be showing only the players age 30 and older.

**View batters whose age is between 25 and 30 who played more than 100 games.** This filter will require several criteria that must be satisfied. Clear any existing filter, and then click the *Filter* function *On* again. Click the down-arrow for the *Age*, and specify a *Number Filters*, for *Age...is greater than or equal to...25*. Now restrict the ages. With *And* selected as shown here, click the drop-down list to choose *is less than or equal to*, enter the value 30, and click *OK*. Now add a Games played filter. As before, add a *Number Filter* for values of “*G...is greater than...100*.” How many players satisfy these criteria in the division?



**Calculating**

Spreadsheets are very good at performing calculations on small or large quantities of data using simple or powerful functions.

**Which is the “youngest team” in your division?**

Calculate the average age for each team using the *AVERAGE()* function. In a blank cell below the table, type “*=AVERAGE()*” without the quotes, and then use the mouse to select the entire range of age values for the team, and *Enter*. See the calculated average value. Repeat for the other teams. Which team in the division might be said to be the youngest?

**What is the age of the oldest player on each team?**

**The youngest player?** Just as you calculated the average, use the *MAX()* and *MIN()* functions.

**Which team scored the most runs all season? The fewest?** Use the *SUM()* function on the “*R*” (runs scored) column.

**Create a Frequency Distribution of player ages.**

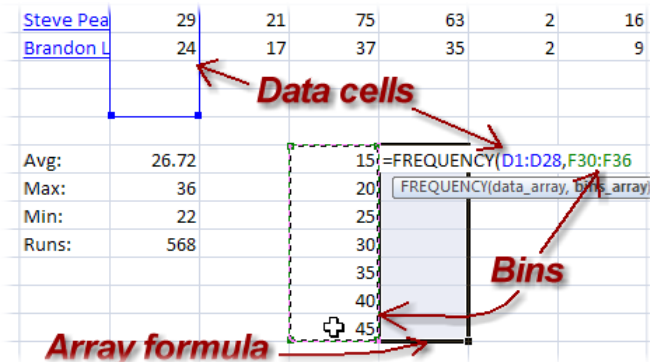
Use the *FREQUENCY()* function to create tallies for a histogram. Here are the steps:

Below the data, in some blank rows, create a column of ages: 15, 20, 25, 30, 35, 40, 45. These will be the “bins” for counting the number of players whose ages in each age category, e.g., up to 15 years, from 16 up to 20, from 21 up to 25, etc.

	A	B	C	D	E	F
1	Rk	Pos		Age	G	PA
2	1	C	Mike Nap	30	108	417
3	2	1B	Mitch Mo	26	114	357
4	3	2B	Ian Kinsle	30	157	731
5	4	SS	Elvis Andr	23	158	711
6	5	3B	Adrian Be	33	156	654
7	6	LF	David Mur	30	147	521
8	7	CF	Josh Ham	31	148	636
9	8	RF	Nelson Cr	31	159	642
10	9	DH	Michael Y	35	156	651
11	10	CF	Craig Gen	28	122	269
12	11	C	Yorvit Tor	33	49	182
13	12	C	Geovany S	29	47	164
14	13	UT	Brandon S	25	40	69
15	14	UT	Alberto G	29	24	55
16	15	CF	Leony's M	24	24	52
17	16	UT	Mike Olt	23	16	40
18	17	C	Luis Marti	27	10	19
19	18	MI	Jurickson	19	9	17
20	19	IF	Luis Hern	28	2	2
21						
22						
23						

Highlight the blank cells beside the bins. You will next enter what’s called an “array formula” in the highlighted cells.

- 1) With the blank cells still highlighted, type =FREQUENCY( and then 1) use the mouse to select the entire range of age values for the team (it’s OK to include the header row and blank rows),
- 2) type a comma,
- 3) use the mouse to select the range of bins, and
- 4) To complete array formulas, **don’t press Enter**, but instead press **Ctrl+Shift+Enter**. To edit (or delete) this array formula, again highlight all the cells in the array, make your edit, and again press **Ctrl+Shift+Enter** (or **Delete**).



The array formula and resulting frequency values will fill the selected blank cells. These frequency values will be used to create a histogram (bar chart) below.

### Presenting

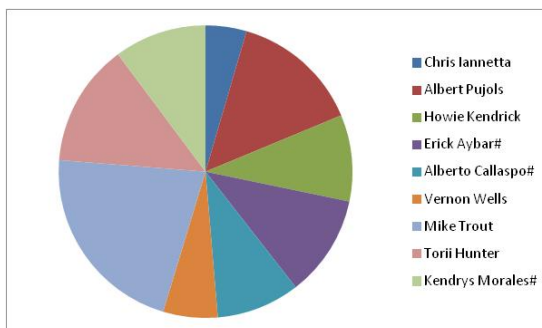
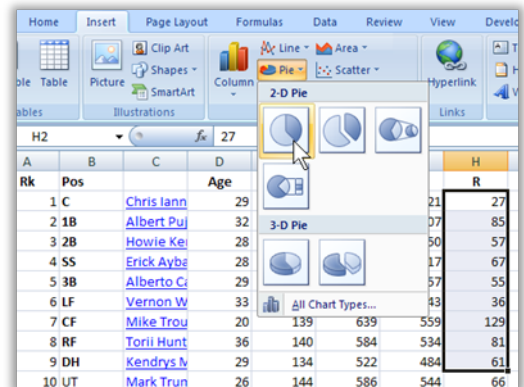
Just as spreadsheets are good at calculating, they are also good at making graphs and charts.

#### Create a pie chart showing the starting players scoring performance.

Highlight the runs scored for the first 9 ranked players in the column headed by “R.”

On the *Insert* tab, click *Pie* (in the *Charts* group) and choose one of the pie chart formats. Excel calculates the percentages and creates all the sections of the pie chart for you.

Right-click the pie chart that appears and choose *Select Data...* Under *Horizontal (Category) Axis Labels*, click *Edit*. Then click/drag to and highlight the player names corresponding to the first nine positions. Click *OK*, and *OK*. The resulting pie chart will now associate each player name with a colored section.



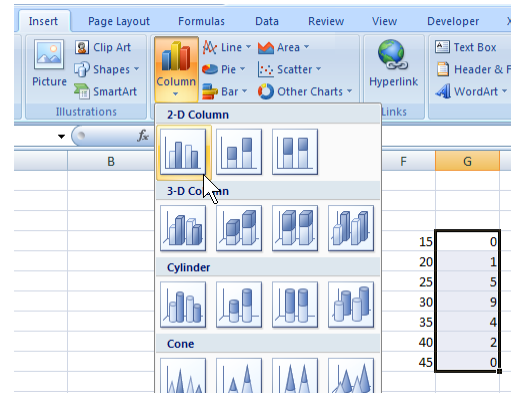
You can further customize the chart appearance (title, shadows, labels, colors) as you wish.

**Make a bar chart showing the distribution of player ages** using the results of the frequency array formula calculated previously.

Just as for the pie chart, start by highlighting the data values to be charted: the column of Frequency calculations (beside the bins of age values).

On the *Insert* tab, click on *Column* (in the *Charts* group) and choose one of the formats. Excel creates the bar chart using the frequency counts.

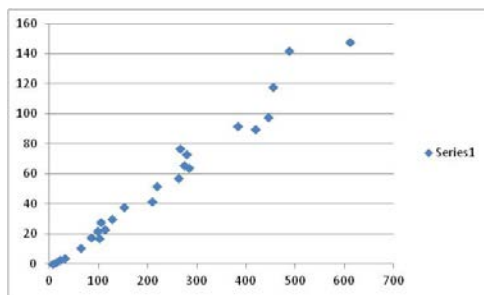
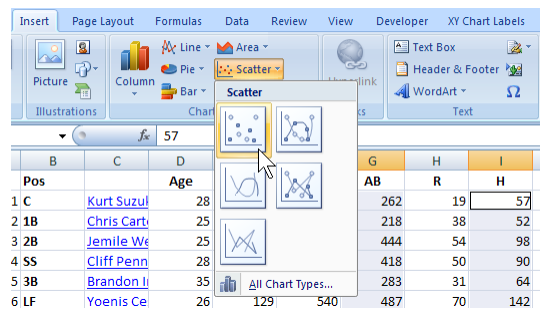
Just as for the pie chart, right-click the new bar chart and choose *Select Data...* Under *Horizontal (Category) Axis Labels*, click *Edit*. Then click/drag to highlight the bin values corresponding to the frequency values. Click OK, and OK. Excel labels the bars with the bin values.



**Make a scatter graph of the number of times at bat versus the number of hits.** You can probably anticipate the outcome here, but it will be a good check of graphing skills.

As for the other charts, start by highlighting the data values to be charted. But a scatter graph needs both *x- and y-values*. Start by dragging the mouse to highlight the *x-values*: the *At Bats* (headed by “AB”) column of values. Then, while holding down the Ctrl-key, drag the mouse to highlight the *y-values*—the *Hits* (headed by “H”) column of values. Release the Ctrl-key.

On the *Insert* tab, click on *Scatter* (in the *Charts* group), and choose the *Scatter with Only Markers* format.



The resulting scatter chart suggests a relationship between the number of times at bat (the horizontal axis) and the number of hits (the vertical axis).

You can further edit the characteristics of the chart’s titles, labels, and legend.

**Explore other relationships to the *At Bats* values.** Click the scatter chart and notice the blocks of data values selected on the spreadsheet are outlined. Carefully position the mouse over the outline around the *Hits*, mouse-down and drag the outline over to another column of values and release. The chart instantly updates with the new scatter comparison. Try out several columns—especially the *Batting Average* (headed by “BA”). Does a player’s batting average seem to improve when they have more times at bat?

NOTE: A chart with no labels could be misinterpreted, so before sharing any of these charts add chart and axis titles and adjust the font sizes and colors.

## Handout 2—Column Headings Key for Baseball Statistics

Label	Meaning
Rk	Rank. This is a count of the rows from top to bottom. It is recalculated following the sorting of a column.
Pos	Position
Age	Player's age at midnight of June 30th of that year
G	Games Played or Pitched
PA	Plate Appearances. When available, we use actual plate appearances from play-by-play game accounts. Otherwise estimated using $AB + BB + HBP + SF + SH$ , which excludes catcher interferences.
AB	At Bats
R	Runs Scored/Allowed
H	Hits/Hits Allowed
2B	Doubles Hit/Allowed
3B	Triples Hit/Allowed
HR	Home Runs Hit/Allowed
RBI	Runs Batted In
SB	Stolen Bases
CS	Caught Stealing
BB	Bases on Balls/Walks
SO	Strikeouts
BA	Hits/At Bats. For recent years, leaders need 3.1 PA per team game played. Bold indicates highest BA using current stats. Gold means awarded title at end of year.
OBP	$(H + BB + HBP)/(At\ Bats + BB + HBP + SF)$ . For recent years, leaders need 3.1 PA per team game played.
SLG	Total Bases/At Bats or $(1B + 2*2B + 3*3B + 4*HR)/AB$ . For recent years, leaders need 3.1 PA per team game played.
OPS	On-Base + Slugging Percentages For recent years, leaders need 3.1 PA per team game played.
OPS+	$OPS + 100 * [OBP / lg\ OBP + SLG / lg\ SLG - 1]$ . Adjusted to the player's ballpark(s).
TB	Total Bases. Singles + 2 x Doubles + 3 x Triples + 4 x Home Runs.
GDP	Double Plays Grounded Into. Only includes standard 6-4-3, 4-3, etc. double plays. First tracked in 1933. For gamelogs only in seasons we have play-by-play, we include triple plays as well. All official seasonal totals do not include GITP's.
HBP	Times Hit by a Pitch.
SH	Sacrifice Hits (Sacrifice Bunts)
SF	Sacrifice Flies. First tracked in 1954.
IBB	Intentional Bases on Balls. First tracked in 1955.
Positions	C: Catcher, 1B: first base, 2B: second base, 3B: third base, SS: short stop, LF: left field, CF: center field, RF: right field, DH: designated hitter, P: pitcher, UT: utility, OF: outfield, IF: infield, CI: corner infield, MI: middle infield

## ACTIVITY: INTERPRETING INFORMATION

### Instructor Preparation

The preceding activities called for information to be collected, evaluated for integrity and trustworthiness, and placed in an organized form. The data can now be interpreted. A truly intelligent interpretation depends on an understanding of the data. In the case of the baseball statistics, for example, the manager is already aware of the many variables and interactions that affect a player's batting performance and can spot notable variations from expected values. However, we are not in such a privileged position. In this activity we will explore some more ways to analyze and interpret abundant data to make meaningful comparisons.

Clearly, the statistical functions and methods we applied to the baseball statistics can also be applied to data from any other field of interest.

- In a Law and Public safety career area, an analysis of traffic fatality data helps us explore the correlation between speeding and traffic fatalities. For this, using data available from the U.S. Department of Transportation website can be used:  
<https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data>
- In an Information Technology course, a study of the growing use of personal technology could include a very interesting analysis of the changing demographics of home computers and internet access, iPads, cell phones, and so forth. For example, a wealth of data is available from the International Telecommunications Union, at: <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- A geology course could examine the trends in specific mineral production and use over the past several years. For example, a wide assortment of mineral summary data can be found at the USGS website: <http://minerals.er.usgs.gov/minerals/pubs/commodity/>

### Objectives

Students will:

1. Calculate the linear correlation coefficient between two sets of data with the aid of a spreadsheet.
2. Apply and interpret curve-fitting of various types (lines, parabolas, exponential) to graphs of data pairs with the aid of a spreadsheet.
3. Use student's t-test for statistical significance testing of experimental data.
4. Use logarithmic graphs to study exponential data.

### Materials

- An internet-connected computer, equipped with a browser and spreadsheet program (e.g., MS Excel)—computer lab time is recommended
- Storage media for spreadsheet file (e.g., network drive, USB flash drive, or similar)

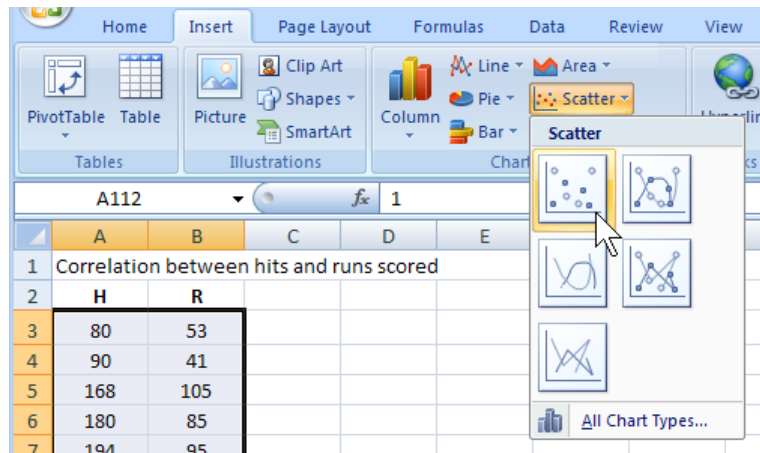
- Handout
- Instructor computer with projector, for demonstrating Excel functions

## Activity Guidelines

1. **Below are example calculations, graphs, and analyses that are available in Excel to assist in interpreting data. Of course, this is not an exhaustive treatise on data analysis, but is meant to whet students' appetite for further exploration.** Correlation is a common statistical measure of interest on one or more variables that one suspects of being related. One common calculation is known as the linear correlation coefficient,  $R^2$  (ranging in values from 0 to 1 [or 0% to 100%], is the degree to which two attributes show a tendency to vary in a linear fashion. For example, considering the baseball statistics, while many factors will affect the number of base hits a player makes and the number of runs scored, we expect a relationship. A linear relationship is especially easy to see when the two variables are graphed; a numerical measure is also easy to get.

- a. **Instructor demonstration plus student activity:**

In the baseball statistics spreadsheet, create a new worksheet. Copy the values from two columns of the team statistics to this worksheet: the Hits (H) and the Runs Scored (R). Hint: Put the *Hits* data in the first column and the *Runs Scored* data in the second column. This creates ordered pairs of (x, y) values.

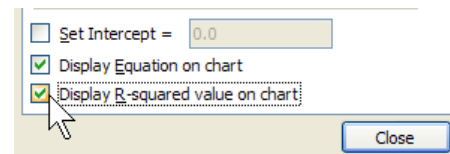


- b. **Student activity:**

Drag the mouse to highlight the two columns of data from all the divisions. On the *Insert* tab, in the *Charts* group select *Scatter*, and then *Scatter with only markers*. Observe the resulting graph. Titles to the graph and the axes can be added. Does there appear to be a relation between hits and runs? What would the graph look like if there were no correlation?

- c. **Student activity:**

The spreadsheet can easily tell about the linear correlation. Right-click the plotted data values on the chart (if something else is selected first, press Escape), and choose *Add Trendline...* from the pop-up menu. Verify that "Linear" is the type selected. Near the bottom of the *Trendline Options*, check the options for *Display Equation on chart* and *Display R-squared value on chart*. Click *Close*. Notice that Excel adds to the chart: 1) a line of best fit (the "trendline"), 2) the equation of the line, and 3) the value of the correlation coefficient,  $R^2$ . The closer  $R^2$  is to 1, the stronger the correlation. For two variables behaving randomly,  $R^2$  will be nearly zero.





- d. **Optional student activity:**  
Repeat the above steps for each team in the division. Do all teams have nearly the same relationship between hits and runs scored? Do these graphs and statistics detect which team ultimately won the division?
- e. **Optional student activity:**  
Repeat the above steps for another pair of columns of baseball data (e.g., age vs hits, stolen bases versus caught stealing, and so forth).
2. **Curve fitting** is the process of finding the best equation parameters to fit the data. A good-fitting curve will have the smallest deviations from all the data points. The curve could be a line (like  $y = mx + b$ , finding the best values for  $m$  and  $b$ ), a parabolic shaped curve (like  $y = ax^2 + bx + c$ , finding the best values for  $a$ ,  $b$  and  $c$ ), or an exponential curve (like  $y = ae^{bx}$ , finding the best values for  $a$  and  $b$ ). In any case, the resulting equation provides some ability to predict and understand the behavior of the variables and also provides a smooth curve to help the eye see the trend of the data.
- a. **Student activity:**  
Follow the steps described in the correlation activity to again generate a (or return to the previously made) graph of the times at bat (AB on the  $x$ -axis) versus the number of hits (H) on the  $y$ -axis. Add a linear trendline. What do you think *the slope* of this line tells you? **Hint:** what would be the slope if every player had a hit every time at bat?
- b. **Student activity:**  
Make a new graph: Follow the steps described in the correlation activity to generate a graph of the relationship between the games played ( $G$  on the  $x$ -axis) and a player's on-base percentage ( $OBP$  on the  $y$ -axis). Add a linear trendline.
- c. **Student activity:**  
The data probably doesn't really appear to fit a linear curve well, since players who play more will naturally tend to perform better—or maybe players who perform better will play more. Right-click the trendline, *Format Trendline...* and try other trendline equations to obtain better fitting curves. Do the choices that seem to fit better also yield larger  $R^2$  values?
- d. **Optional student activity:**  
While considering the above graphs (or others) and the resulting curve-fits, apply one or more filters to the data. For example, does the slope of the curve for *At Bats* versus *Hits* change for younger players? Older players? What does this mean? **Hint:** Move the graphs down to blank rows below your table of values before applying filters, since the filtering will likely hide rows in which the graphs reside.
3. **Statistical significance testing** is the process of comparing two or more samples of data to assess the likelihood that they came from truly different sources. Manually, the process would be labor-intensive, but with the data in a spreadsheet, such computations become easy. The hardest part is interpreting the results.

One commonly used analysis, called the  $t$ -test, compares the average values of two samples. Here, for example, one can ask: Are the on-base-percentages (OBP) from Team 1 significantly different than the OBP values from Team 2? Gather the teams' OBP values into one worksheet and set up a table as a guide to all the possible team comparisons.

a. **Instructor demonstration:**

Copy the columns of on-base percentages (OBP) from the five teams to a new worksheet: Team 1 in column A, Team 2 in column B, and so forth. Calculate the *t*-test for all ten possible comparisons of these five teams. Make a table, similar to that shown here, using team names. The shaded cells indicate where formulas will be entered; “x” represents redundant pairings.

	Team 1	Team 2	Team 3	Team 4	Team 5
Team 1	x				
Team 2	x	x			
Team 3	x	x	x		
Team 4	x	x	x	x	
Team 5	x	x	x	x	x

b. **Instructor demonstration:**

Now use Excel’s *t*-test function to compare the on-base percentages (OBP) of Team 1 and Team 2. With the cursor positioned in the Team 1-Team 2 cell, type: =TTEST(. Now, specify the first array (the Team 1 OBP values): click and drag to highlight the cells for Team 1 OBP values and then type a comma. Next, specify the second array: highlight the cells for the Team 2 OBP values and type another comma. The *t*-test function needs two more parameters. First, choose a one-tail or two-tail test. Since the test is to see if the teams are different (not whether a certain one is larger than the other), this is a *two-tail test*, so type a “2” and another comma. Lastly, choose the function for a type-1, type-2, or type-3 test. See the summary of types shown here. Since there is no reason to expect equal variation in the OBP values from team to team, specify the safest (and least sensitive) test—type-3. Test by typing the value “3” and pressing *Enter*.

	Team 1	Team 2	Team 3	Team 4	Team 5
Team 1	x	=TTEST(A2:A26,B2:B26,2,3)			
Team 2	x	TTEST(array1, array2, tails, type)			
Team 3	x	x	x		
Team 4	x	x	x	x	
Team 5	x	x	x	x	x

t-test: Types of comparisons	
Type 1:	Two samples with paired values
Type 2:	Two samples, expecting equal variances
Type 3:	Two samples, probably unequal variances

c. **Instructor demonstration:**

The value from the formula is known as a *p*-value. Generally, if *p* is less than 0.05, the difference between the two samples is “significant.” If  $0.5 \leq p \leq 1.0$ , it might be slightly significant and bear further scrutiny. Otherwise, if  $p > 1$  there is generally no significant difference between the samples. Did the comparison of the first two teams yield “significantly different” OBP values? Hint: Format cells having significant *p* values with some obvious color. Advanced users: Apply conditional formatting to these cells, indicating small *p*-values with some color, for example.

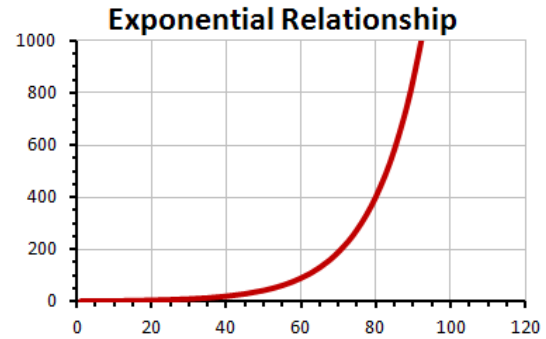
d. **Student activity:**

Carefully repeat the previous steps for the remaining nine combinations of team values. Do any of pairs of Teams have significantly different OBP values?

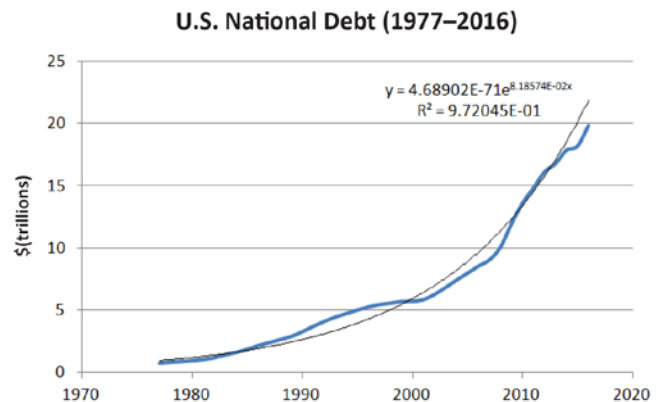
NOTE: There are many other tests of significance that can be performed on sample data (chi-square, ANOVA, and so forth), some accessible with Excel functions. However, although *Excel* is useful for basic data handling and calculating, for more complex statistical analyses it would be best to use a true statistical analysis package.

4. **Logarithmic and Exponential processes.** Not every dependent relationship is linear. In fact, most are not. Processes that can be measured and found to depend on the amount being measured are usually reflecting exponential relationships. Since it's not likely to find exponential behavior in the baseball statistics, look to another source that might be exponential: The U.S. National Debt. Annual data is publicly available at:

<http://www.treasurydirect.gov/govt/reports/pd/histdebt/histdebt.htm>



- a. **Instructor demonstration plus student activity:**  
Visit the above site and find the debt data for at least the past 40 years. Copy paste the columns of data into a blank spreadsheet. “Clean up” the data, if necessary, removing any footnotes or text.
- b. **Optional Instructor demonstration:**  
Assuming dates are in column A, and debt amounts in column B, you can translate the pasted values for easier handling. For example, add a column of just year values: in C1 enter “=YEAR(A1).” Add a fourth column to convert the debt values to trillion dollar amounts, dividing each debt amount by  $10^{12}$  (in Excel form that is 1E12) in D1, enter “=B1/1E12.” Copy these formulas down for each row of the original data.
- c. **Student activity:**  
Highlight the data (or the converted data) and create a smooth XY-graph (i.e., scatter graph).
- d. **Student activity:**  
Does the data seem to be linear? Or does it more closely resemble the exponential relationship shown above?
- e. **Student activity:**  
Right-click the curve and select “Add Trendline” to the data. Excel will default and apply a linear-fit trendline to the data. Right-click the trendline and add the  $R^2$  value (a measure of goodness-of-fit) to the chart. Values of  $R^2$  less than 0.9 indicate a rather weak fit to the data.
- f. **Student activity:**  
Right-click the trendline again and choose an exponential trendline. Does the new trendline seem to fit the data better and the  $R^2$  value improve?  
**Optional:** Select “Display Equation on Chart” to see the exponential equation that best fits the data. Right-click the equation, choose *Format Trendline Label*, and change the *Number* format to scientific notation with 5 or 6 decimal places. (See our result above.)



g. **Student activity:**

Using this equation, predict a little bit into the future. What might the debt be expected to reach in the next two or three years? (For the example above, the 2018 prediction is well over \$20 trillion, although the fitted curve is only suggesting the trend, not the exact amounts.)

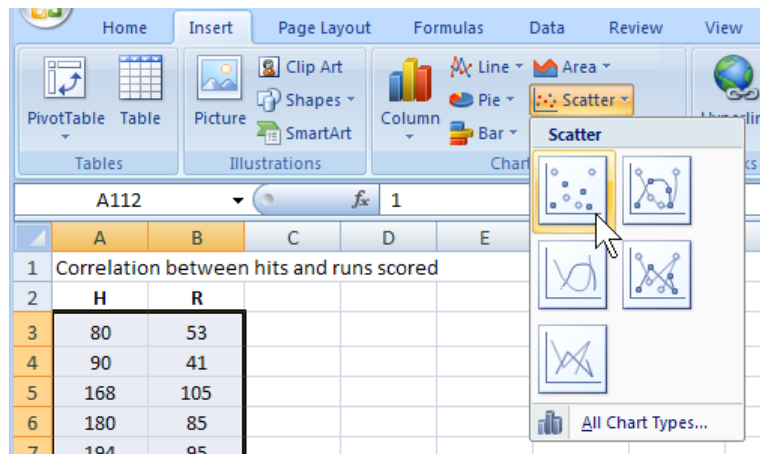
## Wrap-up

Data analysis and interpretation is a career unto itself, so it is quite beyond the scope of a single activity to adequately equip or provide practice for students. However, the abundance of available data, thanks to the internet, can make presentation or persuasive case in any career area stronger based on real-world data. When comparing sports statistics, there are many, many variables that affect each player's performance. The result is an almost random variation of dozens of statistics for the players—if not for the strong effect that certain skills and talents can have on the outcomes. For baseball aficionados, therein lies the entertainment value of the sport, as they enjoy predicting the consequences of those various skills and talents. While this activity has focused on historical data, the real excitement can come from using recent data to try to predict future data. Many hours of study, analysis, and discussion can come from such pursuits.

## Handout—Linear Correlations, Curve-Fitting, Statistical Significance, and Logarithmic Graphs

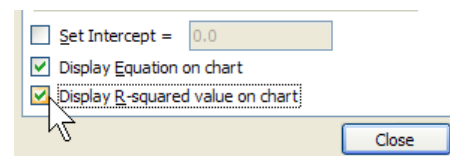
1. Correlation is a common statistical measure of interest on one or more variables that one suspects of being related. One common calculation is known as the linear correlation coefficient,  $R^2$  (ranging in values from 0 to 1 [or 0% to 100%], is the degree to which two attributes show a tendency to vary in a linear fashion. For example, considering the baseball statistics, while many factors will affect the number of base hits a player makes and the number of runs scored, we expect a relationship. A linear relationship is especially easy to see when the two variables are graphed; a numerical measure is also easy to get.

- a. In the baseball statistics spreadsheet, create a new worksheet. Copy the values from two columns of the team statistics to this worksheet: the Hits (H) and the Runs Scored (R). Hint: Put the *Hits* data in the first column and the *Runs Scored* data in the second column. This creates ordered pairs of (x, y) values.



- b. Drag the mouse to highlight the two columns of data from all the divisions. On the *Insert* tab, in the *Charts* group select *Scatter*, and then *Scatter with only markers*. Observe the resulting graph. Titles to the graph and the axes can be added. Does there appear to be a relation between hits and runs? What would the graph look like if there were no correlation?

- c. The spreadsheet can easily tell about the linear correlation. Right-click the plotted data values on the chart (if something else is selected first, press Escape), and choose *Add Trendline...* from the pop-up menu. Verify that "*Linear*" is the type selected. Near the bottom of the *Trendline Options*, check the options for *Display Equation on chart* and *Display R-squared value on chart*. Click *Close*. Notice that Excel adds to the chart: 1) a line of best fit (the "trendline"), 2) the equation of the line, and 3) the value of the correlation coefficient,  $R^2$ . The closer  $R^2$  is to 1, the stronger the correlation. For two variables behaving randomly,  $R^2$  will be nearly zero.



- d. Repeat the above steps for each team in the division. Do all teams have nearly the same relationship between hits and runs scored? Do these graphs and statistics detect which team ultimately won the division?
- e. Repeat the above steps for another pair of columns of baseball data (e.g., age vs hits, stolen bases versus caught stealing, and so forth).

2. **Curve fitting** is the process of finding the best equation parameters to fit the data. A good-fitting curve will have the smallest deviations from all the data points. The curve could be a line (like  $y = mx + b$ , finding the best values for  $m$  and  $b$ ), a parabolic shaped curve (like  $y = ax^2 + bx + c$ , finding the best values for  $a$ ,  $b$  and  $c$ ), or an exponential curve (like  $y = ae^{bx}$ , finding the best values for  $a$  and  $b$ ). In any case, the resulting equation provides some ability to predict and understand the behavior of the variables and also provides a smooth curve to help the eye see the trend of the data.
- Follow the steps described in the correlation activity to again generate a (or return to the previously made) graph of the times at bat (AB on the  $x$ -axis) versus the number of hits (H) on the  $y$ -axis. Add a linear trendline. What do you think *the slope* of this line tells you? **Hint:** what would be the slope if every player had a hit every time at bat?
  - Make a new graph: Follow the steps described in the correlation activity to generate a graph of the relationship between the games played ( $G$  on the  $x$ -axis) and a player's on-base percentage ( $OBP$  on the  $y$ -axis). Add a linear trendline.
  - The data probably doesn't really appear to fit a linear curve well, since players who play more will naturally tend to perform better—or maybe players who perform better will play more. Right-click the trendline, *Format Trendline...* and try other trendline equations to obtain better fitting curves. Do the choices that seem to fit better also yield larger  $R^2$  values?
  - While considering the above graphs (or others) and the resulting curve-fits, apply one or more filters to the data. For example, does the slope of the curve for *At Bats* versus *Hits* change for younger players? Older players? What does this mean? **Hint:** Move the graphs down to blank rows below your table of values before applying filters, since the filtering will likely hide rows in which the graphs reside.
3. **Statistical significance testing** is the process of comparing two or more samples of data to assess the likelihood that they came from truly different sources. Manually, the process would be labor-intensive, but with the data in a spreadsheet, such computations become easy. The hardest part is interpreting the results.

One commonly used analysis, called the  $t$ -test, compares the average values of two samples. Here, for example, one can ask: Are the on-base-percentages (OBP) from Team 1 significantly different than the OBP values from Team 2? Gather the teams' OBP values into one worksheet and set up a table as a guide to all the possible team comparisons.

- Copy the columns of on-base percentages (OBP) from the five teams to a new worksheet: Team 1 in column A, Team 2 in column B, and so forth. Calculate the  $t$ -test for all ten possible comparisons of these five teams. Make a table, similar to that shown here, using team names. The shaded cells indicate where formulas will be entered; "x" represents redundant pairings.

	Team 1	Team 2	Team 3	Team 4	Team 5
Team 1	x				
Team 2	x	x			
Team 3	x	x	x		
Team 4	x	x	x	x	
Team 5	x	x	x	x	x

- b. Now use Excel’s *t*-test function to compare the on-base percentages (OBP) of Team 1 and Team 2. With the cursor positioned in the Team 1-Team 2 cell, type: =TTEST(. Now, specify the first array (the Team 1 OBP values): click and drag to highlight the cells for *Team 1* OBP values and then type a comma. Next, specify the second array: highlight the cells for the *Team 2* OBP values and type another comma. The *t*-test function needs two more parameters. First, choose a one-tail or two-tail test. Since the test is to see if the teams are different (not whether a certain one is larger than the other), this is a *two-tail test*, so type a “2” and another comma. Lastly, choose the function for a type-1, type-2, or type-3 test. See the summary of types shown here. Since there is no reason to expect equal variation in the OBP values from team to team, specify the safest (and least sensitive) test— type-3. Test by typing the value “3” and pressing *Enter*.

	Team 1	Team 2	Team 3	Team 4	Team 5
Team 1	x	=TTEST(A2:A26,B2:B26,2,3)			
Team 2	x	TTEST(array1, array2, tails, type)			
Team 3	x	x	x		
Team 4	x	x	x	x	
Team 5	x	x	x	x	x

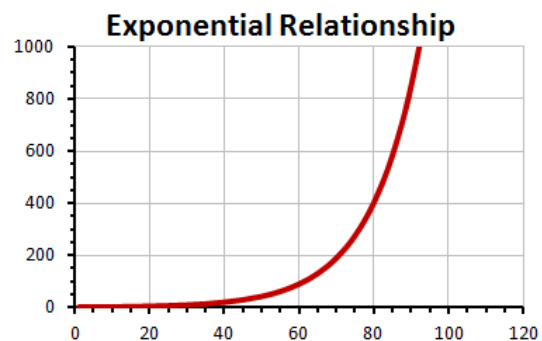
**t-test: Types of comparisons**

- Type 1: Two samples with paired values
- Type 2: Two samples, expecting equal variances
- Type 3: Two samples, probably unequal variances

- c. The value from the formula is known as a *p*-value. Generally, if *p* is less than 0.05, the difference between the two samples is “significant.” If  $0.5 \leq p \leq 1.0$ , it might be slightly significant and bear further scrutiny. Otherwise, if  $p > 1$  there is generally no significant difference between the samples. Did the comparison of the first two teams yield “significantly different” OBP values? Hint: Format cells having significant *p* values with some obvious color. Advanced users: Apply conditional formatting to these cells, indicating small *p*-values with some color, for example.
- d. Carefully repeat the previous steps for the remaining nine combinations of team values. Do any of pairs of Teams have significantly different OBP values?

NOTE: There are many other tests of significance that can be performed on sample data (chi-square, ANOVA, and so forth), some accessible with Excel functions. However, although *Excel* is useful for basic data handling and calculating, for more complex statistical analyses it would be best to use a true statistical analysis package.

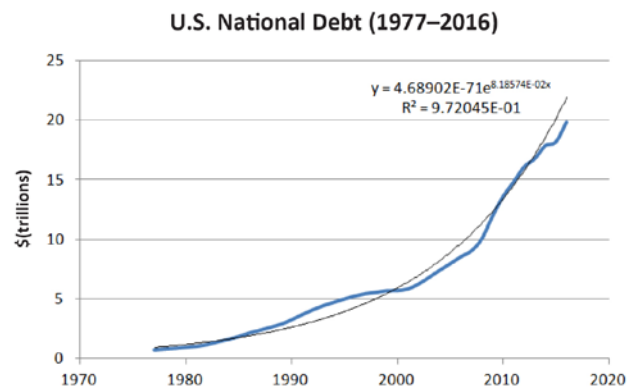
4. **Logarithmic and Exponential processes.** Not every dependent relationship is linear. In fact, most are not. Processes that can be measured and found to depend on the amount being measured are usually reflecting exponential relationships. Since it’s not likely to find exponential behavior in the baseball statistics, look to another source that might be exponential: The U.S. National Debt. Annual data is publicly available at:



<http://www.treasurydirect.gov/govt/reports/pd/histdebt/histdebt.htm>

- Visit the above site and find the debt data for at least the past 40 years. Copy paste the columns of data into a blank spreadsheet. “Clean up” the data, if necessary, removing any footnotes or text.
- Assuming dates are in column A, and debt amounts in column B, you can translate the pasted values for easier handling. For example, add a column of just year values: in C1 enter =YEAR(A1). Add a fourth column to convert the debt values to trillion dollar amounts, dividing each debt amount by  $10^{12}$  (in Excel form that is 1E12) in D1, enter =B1/1E12. Copy these formulas down for each row of the original data.
- Highlight the data (or the converted data) and create a smooth XY-graph (i.e., scatter graph).
- Does the data seem to be linear? Or does it more closely resemble the exponential relationship shown above?

- Right-click the curve and select “Add Trendline” to the data. Excel will default and apply a linear-fit trendline to the data. Right-click the trendline and add the  $R^2$  value (a measure of goodness-of-fit) to the chart. Values of  $R^2$  less than 0.9 indicate a rather weak fit to the data.
- Right-click the trendline again and choose an exponential trendline. Does the new trendline seem to fit the data better and the  $R^2$  value improve?



Optional: Select “Display Equation on Chart” to see the exponential equation that best fits the data. Right-click the equation, choose *Format Trendline Label*, and change the *Number* format to scientific notation with 5 or 6 decimal places. (See our result above.)

- Using this equation, predict a little bit into the future. What might the debt be expected to reach in the next two or three years? (For the example above, the 2018 prediction is well over \$20 trillion, although the fitted curve is only suggesting the trend, not the exact amounts.)



## Assessment Tools/Strategies

This section includes specific strategies and instruments for assessing students' knowledge, skills, and attitudes in regard to Information Processing.

### RUBRICS

Rubrics are valuable assessment tools. Students should be provided with the rubric by which they will be assessed before an activity begins so they will know from the beginning what behaviors and quality of work are expected. When time permits it is useful to have students contribute to the rubrics by brainstorming with the instructor on what a quality behavior or product looks like. For example, before assigning a team project to locate, analyze, and interpret data and prepare a group presentation, ask students to describe characteristics of the ideal presentation. Prompt students with specific components. Then have them describe a really poor analysis and interpretation of data and associated class presentation. These will be the descriptions of the characteristics for the highest and lowest ends of the Likert scale for each essential attribute. Instructors should add any required attributes that must be addressed to the rubric if the students do not come up with them on their own.

- The first two rubrics cover Information Gathering and Data Analysis, and include spaces where the instructor can insert additional attributes to tailor the rubric to a specific project or activity. It was developed to be used by the instructor or other observer who will be assessing a student.
- The third Information Processing rubric is a self-rating rubric that is for use by the students. The students indicate the degree to which they think they are performing each attribute. They can periodically return to the rubric to reassess and determine if they are improving those skills.
- The final type of rubric is the most complex. It is exemplified by the rubrics: Team Member Plan for Developing Interpersonal Skills and Team Member Plan for Contributing to the Team's Success. The student completes a rubric by providing examples of satisfactory or exemplary performance of the tasks/behaviors listed. Then the student meets with the instructor or peer observer and compares his/her reflections with their instructor's or peers' observations and formulates an action plan for improving attitudes, behaviors or skills.

This type of rubric most resembles the type of assessment an employee might receive on the job. It is also the most time-consuming. Ideally, this rubric would be used at least three times during a course:

- At the beginning of the course, to get a baseline and to give students suggestions for specific actions they might take to improve their performance,
- At the midpoint of the course, to check progress and refine the recommendations for improvement, and
- At the end of the course, to assess the progress made over the duration of the course. Additional suggestions can be made for students' continued growth beyond the end of the course.

## RUBRICS FOR INSTRUCTOR ASSESSMENT

**Outcome: Information Gathering**— Student demonstrates an ability to efficiently use an internet search engine, library catalog, workshop manual index, or similar to locate a desired collection of data or facts for an assigned task. The student must persevere when the specific answer to a specific question is not directly found. Once located, the student is able to record its source and save the data for later analysis.

Uses internet search tools to locate relevant and current information, from both electronic and printed resources.	1 2 3 4 5
Avoids malicious or unreliable web sources.	1 2 3 4 5
Places value on accuracy, style, plausibility of other points of view, and appropriateness for the given situation.	1 2 3 4 5
Collects data from websites such as capturing graphics, tabular data, and text lists.	1 2 3 4 5
Saves collected data to useful and recoverable format, such as spreadsheet or database table.	1 2 3 4 5
Saves sufficient source information from captured and collected data to enable verification, updating, and citation for referencing.	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5

5	Always	Excellent
4	Most of the Time	Good
3	Sometimes	Adequate
2	Occasionally	Fair
1	Never	Poor or None

**Outcome: Data Analysis** – Demonstrate the ability to manage raw data saved in a spreadsheet or database format, perform some basic analyses of the data such as count, average, standard deviation, frequency distribution, and graphical analysis to spot obvious trends and dependencies.

Performs and applies basic mathematical operations including calculations involving exponents, roots, logarithms, and powers.	1 2 3 4 5
Performs mathematical calculations with the aid of spreadsheet functions, including average, standard deviation, frequency distributions, t-test, and linear correlation.	1 2 3 4 5
Constructs and/or interprets graphs, charts, and tables to assist in solving mathematical problems or to illustrate findings, including histograms, scatter diagrams, and linear and logarithmic graphs.	1 2 3 4 5
Perform sorting and filtering of data in a spreadsheet as a means to analyze and interpret data.	1 2 3 4 5
Manipulates and names data by rows and columns, applying and copying formulas to connect cells or ranges of cells.	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5
	1 2 3 4 5

5	Always	Excellent
4	Most of the Time	Good
3	Sometimes	Adequate
2	Occasionally	Fair
1	Never	Poor or None

## RUBRIC FOR SELF-ASSESSING INFORMATION PROCESSING SKILLS

Use this instrument throughout the course to self monitor your information processing skills. Note: It is fairly common to overrate your skills during the initial assessment. You may rate yourself a bit lower as you get feedback from others and discover the complexities of each attribute. By the end of the course you should see improvements in your ratings if you focus on improving the quality of your information gathering, processing, analysis, and interpretation.

**Objective:** Collect data and other information from reliable sources.

**Essential Attribute**

I recognize and collect appropriate and reliable data from relevant sources (to address a particular research need).

I recognize and avoid malicious or unreliable sources encountered during the process of research.

I maintain an accurate log of where information was found and collected.

I use key words with an internet search engine, library search tool, or book index to efficiently locate desired information.

I use various media sources for needed data, such as internet sites, libraries, manuals, and so forth.

I assemble materials with needed information into organized and easily reviewed structures.

I am able to accurately save data collected from external sources to a spreadsheet or database program.

	I	II	III	IV
I recognize and collect appropriate and reliable data from relevant sources (to address a particular research need).	Seldom  ----- ----- ----- -----	Sometimes  ----- ----- ----- -----	Usually  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I recognize and avoid malicious or unreliable sources encountered during the process of research.	Seldom  ----- ----- ----- -----	Sometimes  ----- ----- ----- -----	Usually  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I maintain an accurate log of where information was found and collected.	Seldom  ----- ----- ----- -----	Sometimes  ----- ----- ----- -----	Usually  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I use key words with an internet search engine, library search tool, or book index to efficiently locate desired information.	Seldom  ----- ----- ----- -----	Occasionally  ----- ----- ----- -----	Often  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I use various media sources for needed data, such as internet sites, libraries, manuals, and so forth.	Seldom  ----- ----- ----- -----	Occasionally  ----- ----- ----- -----	Often  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I assemble materials with needed information into organized and easily reviewed structures.	Seldom  ----- ----- ----- -----	Sometimes  ----- ----- ----- -----	Usually  ----- ----- ----- -----	Consistently  ----- ----- ----- -----
I am able to accurately save data collected from external sources to a spreadsheet or database program.	Seldom  ----- ----- ----- -----	Sometimes  ----- ----- ----- -----	Usually  ----- ----- ----- -----	Consistently  ----- ----- ----- -----

**Objective:** Analyze and interpret data and collected information.

**Essential Attribute**

I organize or manipulate data into manageable sets and subsets for further processing, using labeled columns, rows, pages, folders, worksheets, and so forth.

I use mathematical formula, procedures, and functions to calculate statistics and other necessary analyses or conversions of raw data.

I interpret the results of my analysis.

I calculate linear correlation between variables, perform simple curve-fitting to data, and perform basic statistical testing for significance of data.

	I	II	III	IV
	Seldom	Sometimes	Usually	Consistently
	----- ----- ----- -----			
	Seldom	Sometimes	Usually	Consistently
	----- ----- ----- -----			
	Seldom	Sometimes	Usually	Consistently
	----- ----- ----- -----			
	Seldom	Sometimes	Usually	Consistently
	----- ----- ----- -----			

**Objective:** Present data and information.

**Essential Attribute**

I report the results of my data analysis using the complete and proper terminology.

I use charts, graphs, tables, and lists to convey the findings of my data analysis.

I create charts, graphs, and tables that are labeled and documented to convey the required supporting information.

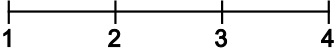
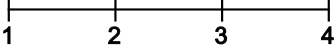
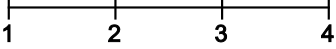
I report the results of my calculations of relevant statistics with sufficient explanation.

I create and interpret line graphs, pie charts, bar charts, and scatter graphs of diverse data.

	I	II	III	IV
I report the results of my data analysis using the complete and proper terminology.	Seldom	Sometimes	Usually	Consistently
I use charts, graphs, tables, and lists to convey the findings of my data analysis.	Seldom	Sometimes	Usually	Consistently
I create charts, graphs, and tables that are labeled and documented to convey the required supporting information.	Seldom	Sometimes	Usually	Consistently
I report the results of my calculations of relevant statistics with sufficient explanation.	Seldom	Sometimes	Usually	Consistently
I create and interpret line graphs, pie charts, bar charts, and scatter graphs of diverse data.	Seldom	Sometimes	Usually	Consistently

## RUBRIC FOR ASSESSING INFORMATION PROCESSING SKILLS

**Outcome: Spreadsheet skills** – Use the power of a spreadsheet program to assimilate and analyze large amounts of data by sorting, filtering, calculating, and graphing.

Performance Criteria		
<b>Reflection</b> Reflect on your actions during class or at a workplace and identify examples of when you:		<b>Personal Plan</b> Based on your examples and the feedback of your peers or instructor, describe the steps you might take to continue or improve your spreadsheet skills.
Applied one of the internet search-and-import data skills learned in this activity to another class or subject of personal interest.	Example:  Peer/instructor review:  Do not agree <span style="float: right;">Strongly agree</span> 	Steps:
Applied one of the spreadsheet analysis skills learned in this activity to another class or subject of personal interest.	Example:  Peer/instructor review:  Do not agree <span style="float: right;">Strongly agree</span> 	Steps:
Enthusiastically shared one or more of the newly learned skills from this activity with a friend or relative.	Example:  Peer/instructor review:  Do not agree <span style="float: right;">Strongly agree</span> 	Steps:

<p><b>Peer comments and suggestions:</b></p>
<p><b>Instructor comments:</b></p>

## Videos and Links

### VIDEOS

The following is an annotated list of videos that are available at the links provided. The instructor may choose to view this list as preparation for these activities, or to use these in class to give additional Information Processing skills, or to stimulate discussion.

<https://www.youtube.com/watch?v=ZpNQldtzrVo>

How to copy/paste tabular data from a PDF file into a spreadsheet

<https://www.youtube.com/watch?v=-aH56dxqLhg>

How to exchange data between MS Excel (spreadsheet) and MS Access (database).

<https://www.youtube.com/watch?v=FaLvuyeVen8>

Array formula basics for MS Excel. The first few minutes are enough to learn a little about using array formulas in MS Excel for this activity.

<https://www.youtube.com/watch?v=81LwaB17GTk>

How to insert a trendline, and get the regression equation and  $R^2$  value in MS Excel.

### WEBLINKS

Here are some links relevant to this module that may be useful.

<http://www.baseball-reference.com/>

Source of current major league baseball statistics that can be used in this activity.

See **Handout 2 of Activity: Evaluating Information**, which contains a collection of contrasting websites, including examples of sites of questionable value as examples of sites to avoid.

<http://www.data.gov/>

Clearinghouse website for accessing government data of all types.