HANDOUT 1: Data and Excel

What is Data?

Data is the plural of *datum*, which is:
- "A single piece of information, as a fact, statistic, or code"

You can think about data as being structured observations. We want to record a set of facts about a set of observations, and we want to be able to know which facts go with which observations. For example, let's say we're interested in how far people can throw a ball. Someone can observe me throwing a ball, and the observer can record all kinds of facts about that observation (such as the weather, the time of day, my height, the distance I threw the ball, etc.). Someone can also observe my friend throw a ball. This would be a *new* observation—it's a new event that needs to be observed. The same set of facts can be recorded about this new observation. We need to be able to compile all this information in such a way that we can make sense of it later on.

So data is really just an organized way of keeping track of which facts go with which observations. There are many ways to organize this information, which is why data is more general than the software we use to analyze data. For example, Excel is just *one* way to look at how data is organized. But there are many more. Check out this flower data, shown in several different formats below. In this data, we are observing each flower once, and we're recording five facts about each observation:

Excel
(SPSS looks very similar)

A1		:	✕	✓	*fx*	Sepal.Length	

	A	B	C	D	E
1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5	3.6	1.4	0.2	setosa
7	5.4	3.9	1.7	0.4	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5	3.4	1.5	0.2	setosa
10	4.4	2.9	1.4	0.2	setosa
11	4.9	3.1	1.5	0.1	setosa
12	5.4	3.7	1.5	0.2	setosa

R

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
>
```

Comma Separated Values (CSV); a popular way of representing data to store on a hard drive

```
"Sepal.Length","Sepal.Width","Petal.Length","Petal.Width","Species
5.1,3.5,1.4,0.2,"setosa"
4.9,3,1.4,0.2,"setosa"
4.7,3.2,1.3,0.2,"setosa"
4.6,3.1,1.5,0.2,"setosa"
5,3.6,1.4,0.2,"setosa"
5.4,3.9,1.7,0.4,"setosa"
4.6,3.4,1.4,0.3,"setosa"
5,3.4,1.5,0.2,"setosa"
4.4,2.9,1.4,0.2,"setosa"
4.9,3.1,1.5,0.1,"setosa"
5.4,3.7,1.5,0.2,"setosa"
```

JavaScript Object Notation (JSON); a popular way of representing data online

```
[
    {
        "Petal.Length": "1.4",
        "Sepal.Length": "5.1",
        "Petal.Width": "0.2",
        "Sepal.Width": "3.5",
        "Species": "setosa"
    },
    {
        "Petal.Length": "1.4",
        "Sepal.Length": "4.9",
        "Petal.Width": "0.2",
        "Sepal.Width": "3",
        "Species": "setosa"
    },
```

Going back to the ball-throwing example, how would you choose to organize that data?

Basics of Excel

Exercise 1: Enter the following data into Excel

```json
{
  "Name": "Chris",
  "Gender": "M",
  "Age": 26,
  "Played.Sports": "Y",
  "Distance.Thrown": 26
},
{
  "Name": "Jane",
  "Gender": "F",
  "Age": 24,
  "Played.Sports": "Y",
  "Distance.Thrown": 28
},
{
  "Name": "Carl",
  "Gender": "M",
  "Age": 22,
  "Played.Sports": "N",
  "Distance.Thrown": 15
},
{
  "Name": "Betty",
  "Gender": "F",
  "Age": 29,
  "Played.Sports": "Y",
  "Distance.Thrown": 22
},
{
  "Name": "Lauren",
  "Gender": "F",
  "Age": 21,
  "Played.Sports": "N",
  "Distance.Thrown": 12
},
{
  "Name": "Roger",
  "Gender": "M",
  "Age": 17,
  "Played.Sports": "N",
  "Distance.Thrown": 26
}
```

Importing Data

Often times you will want to import into Excel data that has already been collected. For example, in my laboratory experiments, we have separate software that can collect thousands of rows of data for each subject. We need to know how to import this kind of data into Excel so we can summarize and analyze it.

- Go to the data tab
- In the "Get External Data" section, select "From Text"

Excel wants to know how your data is organized. Another word for something that separates values in your data (for example, commas) is *delimiter.* The last couple prompts will ask you about how your data is delimited and where in the spreadsheet you want to put your data.

Exercise 2: Import the flower data into a new spreadsheet.

Summarizing Data With Functions

Excel is a powerful tool that can perform many data summaries and analyses. Excel uses simple functions for most of these abilities. If you want to perform a data summary:

- Click on an empty cell
- Type an equals sign (=); this tells Excel that what you're about to type next will be a function and not just ordinary text
- Type the name of the function you want to use, followed by an open parenthesis
- This next part varies depending on the function you're using, but in the simplest case you can just select with your mouse the data you want to summarize and press "Enter"

For example, you can select an empty cell and type: "=max("
Then select a bunch of numbers and press "Enter"
Can you explain why a number appeared in the cell once you pressed "Enter"? Why was it *that* specific number that appeared and not some other number?

If you're unsure what the functions are called, you can click on an empty cell and press the little "fx" button above the spreadsheet:

Exercise 3: Find the averages of the first four variables in the flower dataset.

WHAT TO TURN IN

- Enter into Excel all of your survey data from the questionnaire, email [yourlastname].xls file to your TA (Dave – dab414@lehigh.edu).
- **Hand in the hard copy of your class survey at the end of class.**