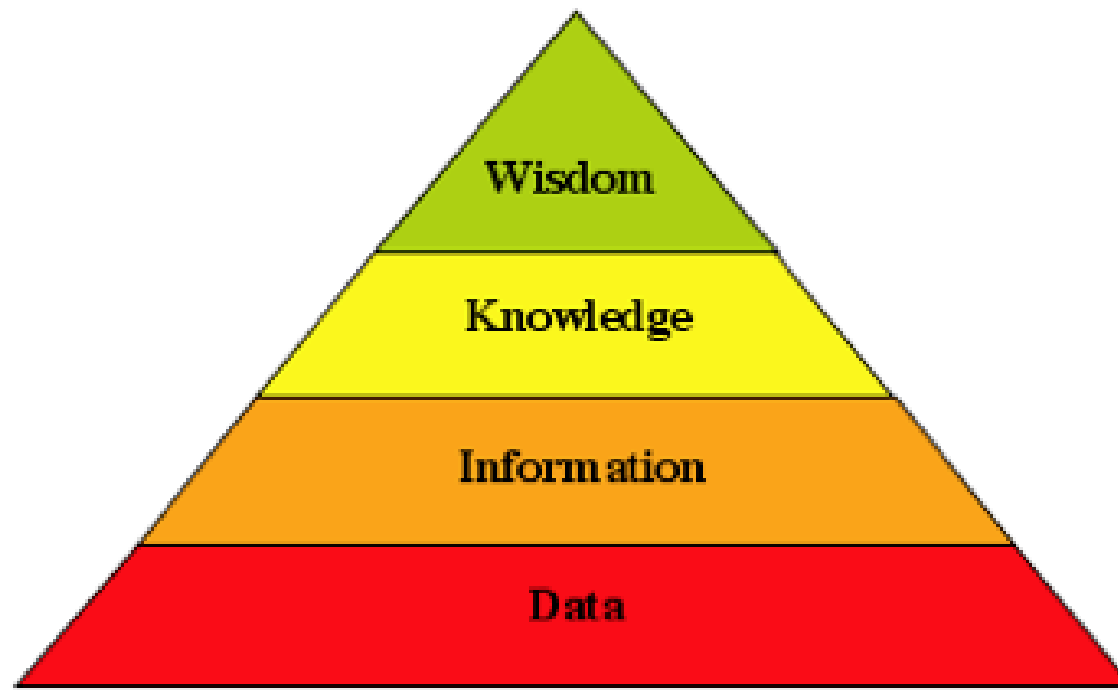


Data Analysis and Databases



Climbing the pyramid

The steps

- 1. What do you have? (the file)
 - Text file, database, JSON, XML, etc.?
- 2. What do you have? (the data)
 - What do the fields mean? What type of data is it?
- 3. What do you want to know?
 - This is the first step of data analysis.
 - Formulate questions you need the answers to.
 - Perform the steps needed to answer the question.

The basics

- Group and count
- Is it text?
 - Text analysis, word frequencies, etc.
- Is it discrete text?
 - Grouping, matching, counting.
- It is a number?
 - Averages, medians, percentages, rates, ratios, correlations.

Formulating questions

- Data analysis and the scientific method share many traits.
- State hypothesis = ask a question
- Get results
- Compare them to the hypothesis

Formulating questions

- A key question is always going to be this: As compared to what?
- [Comparing murder rates.](#)
- Am I using the right number? Average vs. median.
- Dependent vs independent variables.

```
SELECT [field or *] FROM [database  
name] WHERE [field] =/>/</<>  
[condition] GROUP BY [field]  
ORDER BY [field] (DESC)
```

Example: Parking on campus

Data analysis

- What questions did we ask of the parking data?
- What other questions can we ask of data?
- It depends on the data.

More SQL

New tools for the toolbox

New tools

- UPDATE – create new fields of data based on a formula.
- Joins – Connecting two datasets based on a common field. Your student ID number joins you to data on grades, bills, library books, parking tickets, etc.

UPDATE

- UPDATE tablename SET field = “something”
WHERE condition.
- UPDATE contributors SET state = ' GA ' WHERE
last_name = 'Cathy';
- Useful for calculating formulas by step.

Joins

- Inner join: Where both are equal.
- Outer join: Everything, really.
- Full outer join: Where things aren't equal.
- Left/Right join – everything from one table and only things that are equal in the second table.

JOINS

