20 Questions Lesson Plan

Summary

20 Questions is a game that originated in the United States in the 19th century. The goal is to guess a hidden word using only twenty yes/no questions. By carefully choosing the questions, the answerer can quickly narrow down the possibilities to find the hidden word.

This lesson uses activities to teach the basics of decision trees and how they relate to data representation and pattern recognition.

Timeline

What	Time	Where
Introduction	5-10	Lesson plan
Activity 1 - Twenty Questions	10	
Discussion	10	DecisionTreeHelperSlides.pdf
Activity 2 - Worksheets	10	What Will I Do Today Identify the Animal
Wrap up Discussion	10	RealWorldDecisionTrees slides (optional)
Total time	50	

Materials

- Worksheet 1: Decision Trees. Need one copy per student.
- Worksheet 2: Animal Identification. Need one copy per *pair* of students.
- Slides of the diagrams in this lesson plan. This is helpful in order to minimize the time drawing on the board.

Purpose

This lesson helps students learn to solve problems using *data representation* and *pattern recognition*. First, this teaches students the importance of breaking down data in terms of simple questions. It also shows students how to identify patterns by finding defining characteristics between members in a set.

Some materials adapted with permission from CS Unplugged (csunplugged.org), Information Theory activity.

Introduction - Whole Class

There are two ways to introduce this topic. The first is a standard class discussion. In the second option, students will actually be moving around the room (could be good for informal settings, or if students need a break from sitting and listening; also relates well to cases studies on identifying people).

For either version, ask the students if they have ever played 20 questions. If yes, have a student explain the rules. If no, explain the rules yourself:

- One person has a "secret" that others need to guess
- Others can only ask yes/no questions
- The answer must be determined in 20 guesses, or the person with the "secret" wins

Version 1: Seated

Tell the students that you have picked an object in the classroom, and they need to figure out what it is. But they must play be the rules (i.e., only ask questions that require a "yes" or "no" response).

As students ask questions, write them up on the board and answer them.

Version 2: Moving around the room

Tell the students you want to change the goal. Instead of guessing a secret, you want to find a small number of questions that can be used to uniquely identify each one of them. First, brainstorm what makes people different. Some ideas are hair color, eye color, birthday (month, date). Now tell them that each question must be phrased as a yes/no.

With a small class size, have all the students get in one part of the room (e.g., the back). If you have a large class size, it will be easier if half the students remain seated, and half move around. The half that are seated could help devise questions.

Write a question on the board, such as: *Is your hair brown?*

All students with brown hair go to another part of the room, all students without brown hair stay where they are. Now pick another question to write on the board, such as *Do you have blue eyes?*

Again have the students separate based on their answers, but now you should have 4 groupings (brown hair + blue eyes, brown hair + other eyes, not brown hair + blue eyes, not brown hair + other eyes). Continue to ask questions and separate students until either all students are by themselves or a few are (depends on number of students and interest level).

Activity 1 - Twenty Questions

Now have the students play the 20 questions game with each other.

1. Ask the students to divide into groups of size 3, with possibly a group or two of size 4 depending on number of students.

    o   You may want to write the rules on the board, to help students remember.

2. One student in each group will be the secret keeper. That student should think of a person, place or thing. Option: you could provide some additional context or constraints as desired, such as think of a magical object, think of a famous person, etc.

3. The remaining students should take turns (round robin fashion) asking yes/no questions.

4. The secret keeper should count the number of questions. If the answer is not guessed within 20 questions, that's a win. Tell the other students the secret.

5. Then the next person becomes the secret keeper, and another "game" is played.

This activity should be engaging for the students. You should let the students play for a set amount of time, then go over the group discussion questions.

Discussion - Whole class

20 Questions Strategy

Ask the students about their experience with the 20 Questions game:

- Who guessed an object? What kinds of questions led you to the solution?

- How many guesses did it take? Which team had the minimum number? What was the secret?

- Did anyone "win" the game? i.e., were there secrets that no one was able to guess?

Talk with the students about general strategies for good questions. Useful information can be found:
http://www.wikihow.com/Play-20-Questions

- Start with general questions. If first questions are too specific (e.g., is it the chalkboard) it will "use up" a lot of questions.
- Build on prior answers. For example, if the first question is whether the object is in the back of the room, a follow on question might be whether it's on the right or left side of the room (i.e., use location to narrow down possible objects).

Building a Decision Tree for 20 Questions

We first want to link the ideas of decision trees to the game of 20 Questions that the students just played. First ask the students what a good first question would be. They might say something like, "Is it on the wall?" Write that down at the top of the board.
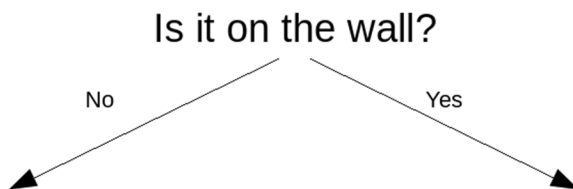
Then ask the students what other questions they might ask if the question is yes or no. Add those questions to the decision tree. For example, here's what the decision tree would look like if they thought a good question to ask after no would be "Is it a person", and after yes, "Does it have words on it?"
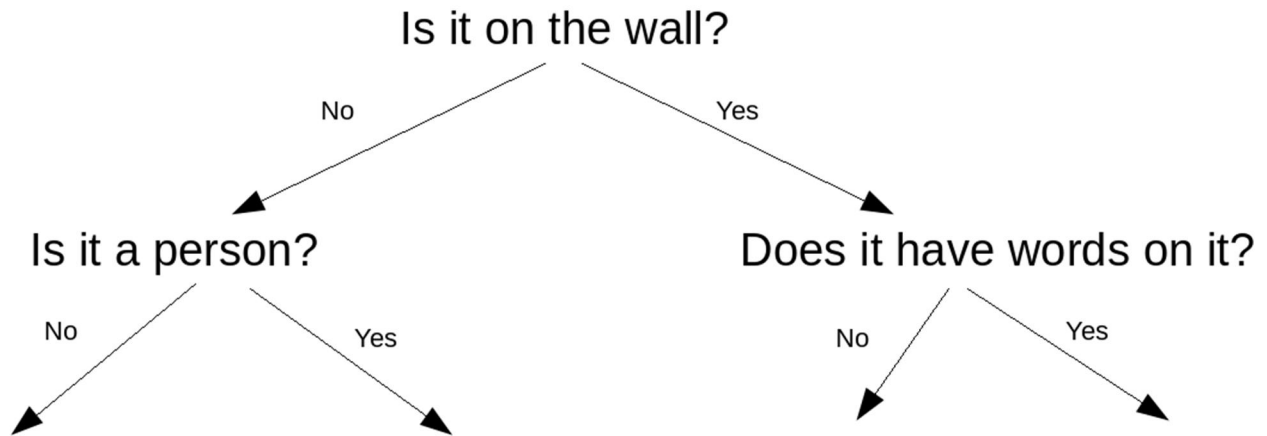
Then ask the students for more response in order fill in the rest of the tree to three levels. For example, this is a completed tree:
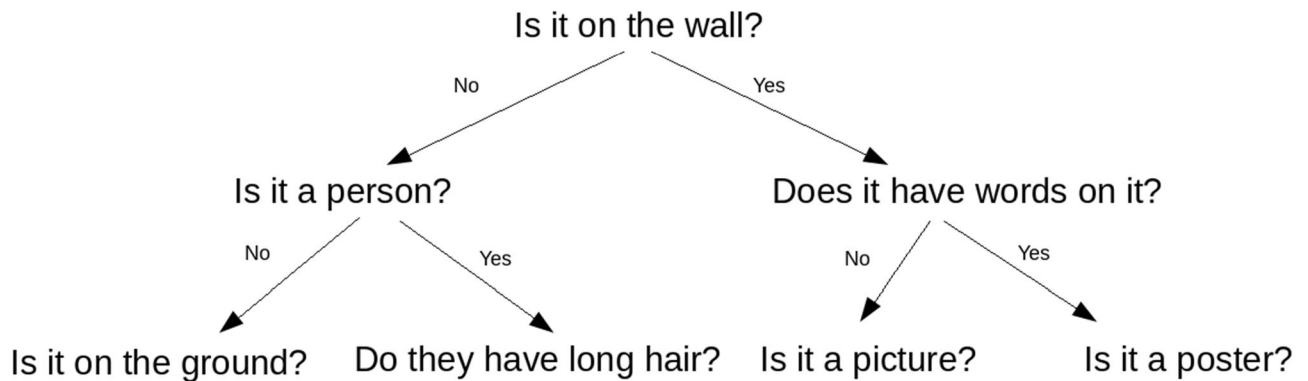
The point of this exercise is so the students have a good understanding of the format of decision trees, as well as understanding how their knowledge from the 20 questions game applies to the next exercise in making number guessing trees.

Number Guessing Game

Now you can erase the previous decision tree, and tell the students you want to play a variation of 20 questions where the object to be guessed is a number between 0 and 7. The same rule applies, students can only ask yes or no questions (but "is the number 0?" is a valid type of question). Take questions from the students.

After students have guessed your answer, tell them there's a really effective way to represent and guide this type of game. If students were asking questions like "Is the number 6?" you can point out that with that type of random guess it might take 8 questions to find the number. Ask them if they think there's a more effective strategy.

Tell them you think it should always be possible to guess the answer with 3 questions. Tell them you will use a tool called a *decision tree* to help with this task. You may want to write *Decision Tree* in a corner of the white board.

Figure 1 shows a complete *decision tree* for guessing a number between 0 and 7. Walk through the steps below to engage students in creating this chart.
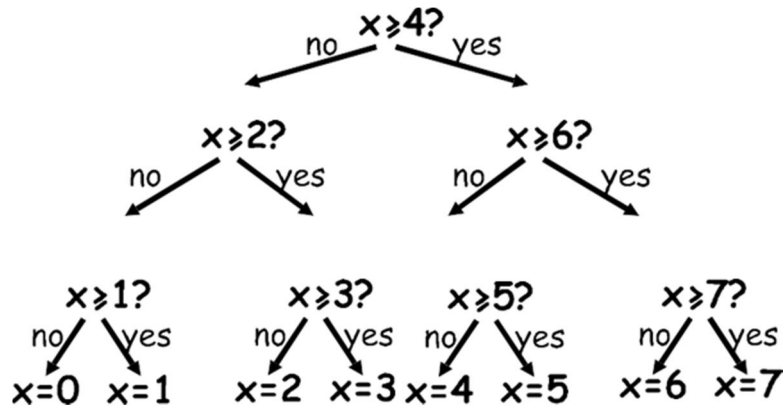

Figure 1: Decision Tree for 0-7

First, off to the side write:

- x is the secret number

Then draw the start of the tree, as shown in figure 2. **NOTE**: The *DecisionTreeHelperSlides* have these figures, your choice whether to use the board or the slides, or some combination.
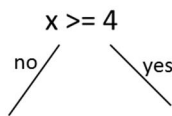

Figure 2: Root of decision tree

Get the students to tell you what numbers would be along each branch. Maybe update your drawing, as shown in figure 3:
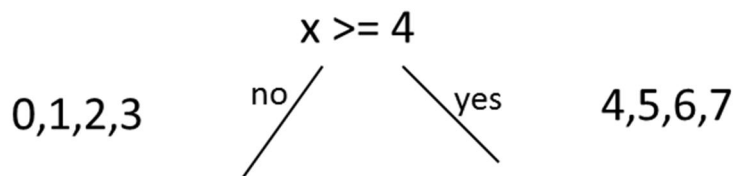

Figure 3: Root of tree with numbers for each branch

Discuss with the students what to do next. They *could* just guess at this point x=0, x=1, etc. But in that case it could take as many as 5 guesses, and we want to get the answer in only 3. You might want to update the drawing (figure 4) to show a *less effective* tree, to help lead students to the right answer.
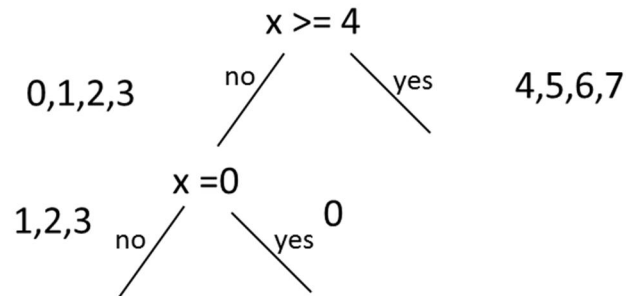
$$x >= 4$$

0,1,2,3   no   yes   4,5,6,7

$$x = 0$$

1,2,3   no   yes   0

Figure 4: Incorrect addition to decision tree

Point out that there are 3 numbers on one side of the tree, but only 1 on the other. See if any students have a better idea. Either way, erase "x = 0" and add the correct branches to the left side of the tree, as shown in figure 5.

$$x >= 4$$

0,1,2,3   no   yes   4,5,6,7

$$x >= 2$$

no   yes

$$x >= 1$$ $$x >= 3$$

no   yes   no   yes
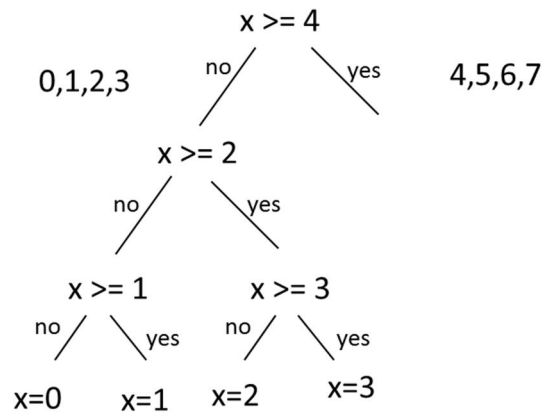
x=0   x=1   x=2   x=3

Figure 5: Half the decision tree

Ensure the students understand that with just 3 questions, this tree shows how to guess 0, 1, 2 or 3. Ask if the same technique could be used for 4, 5, 6 or 7 (yes, of course). Draw the tree with blanks, as shown in figure 6, and ask students to help fill in the numbers (result should be figure 1).
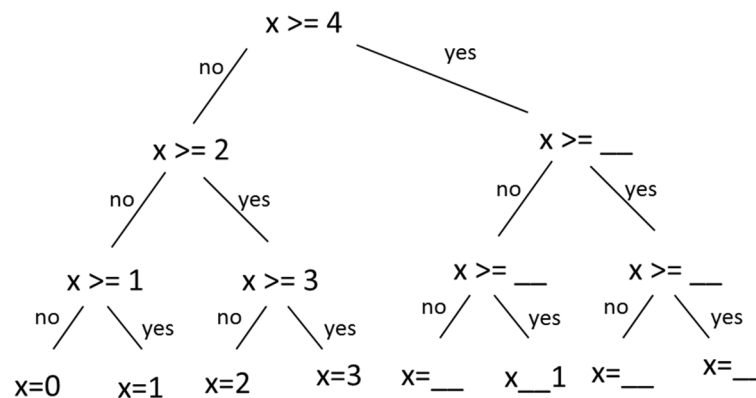

Figure 6: Full tree with blanks

To ensure students understand, ask questions like:

- What are the yes/no decisions needed to 'guess' the number 5? [**answer**: *yes, no yes*]

- How many yes/no guesses do you need to make to work out any number? [**answer**: *3*]

(**optional**) If you have covered the CS Unplugged Binary Numbers lesson, this is a chance to review. If no = 0 and yes = 1, then each set of "guesses" corresponds to the binary number. For example, 1 would be no, no, yes or 001. 7 would be yes, yes, yes or 111. So the decision tree also "encodes" the binary number (this also explains why we're guessing 0 to 7, rather than 1 to 8.)

Extending the decision tree

Ask the students if they know how many more guesses would we need if the number is doubled? So if the secret number could be 0 to 15.

Answer: need just one more guess, so 4

To show this, draw a new "decision" of x >= 8 above the x>=4. Ask the students what label would go on the line from x >= 8 to x >= 4 [**answer**: no]. Ask the students what would be to the right. Students should see that it will be a decision tree with the same structure as the tree on the left, but with different values.

For your convenience, figure 7 shows the decision tree for 0 - 15. You probably do not want to spend time drawing the entire figure, but you might get students to guess what the first "decision" should be. The key point is that this decision will divide the remaining numbers in half. Since there are 8 numbers, this means adding half of 8 (i.e., 4) so the first decision is x >= 12.
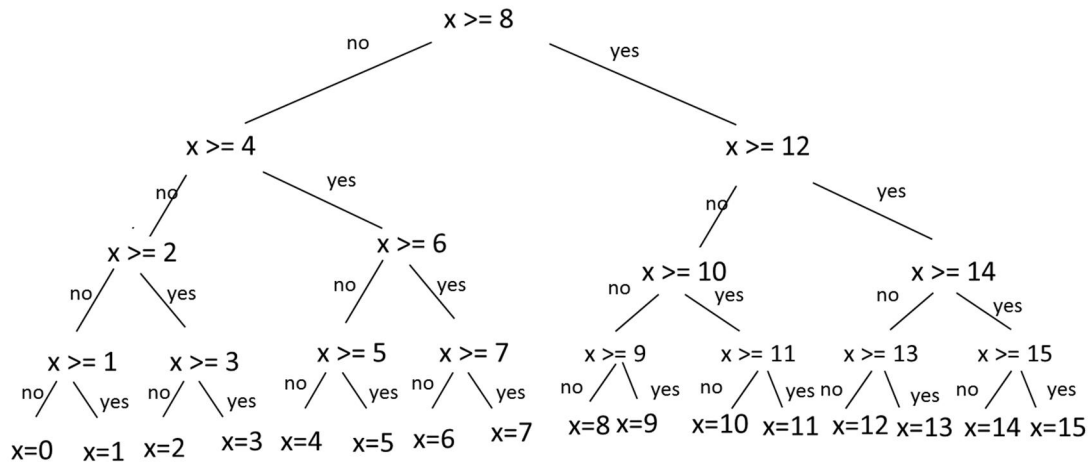

Figure 7: Decision Tree 0 to 15

Notice that we could double the number again, to guess from 0 to 32, with just 5 guesses. 0 to 64 would be 6 guesses, and 0 to 128 would be just 7 guesses. Much better than random! You might want to draw the table below on the board.

Range	# values	# guesses
0-7	8	3
0-15	16	4
0-31	32	5
0-63	64	6
0-1023	1024	10
	1 million	20
	1 billion	30
	1 trillion	40

Figure 8: Table of guesses

(*optional*) If you covered the CS Unplugged Binary Search lesson, this game reinforces the same concepts, but also includes a visual representation in the form of the decision tree.

If students have studied exponents, note that the number of values is a power of 2 (e.g., $2^3 = 8$, $2^4 = 16$, etc.)

Activity 2 - Worksheets

Ask students if they think decision trees would be useful for anything other than guessing numbers. Tell the students that they will do two worksheets to show some other uses.

What Will I Do Today

This is a simple worksheet that ensures students understand the basic format of a decision tree. It is probably best done individually. After students have time to complete, discuss the solutions with the students. As part of this discussion, point out that this type of tree only knows as much as we tell it. So in real life, for example, we might play chess after school, or even during school.

Solutions:

1. Follow the "no" path to determine the friend would play soccer.

2. Since the decision about whether to go to school is the first decision, it does not matter if it is raining, you always go to school if it is a school day

3. If your friend was playing chess, it must have been rainy because of the decision made in the previous step. (your friend would also have needed someone to play chess with, of course)

Identify the Animal

This is a creative worksheet where students try to come up with questions that could uniquely identify the animals.

*   Explain that students should fill in questions to identify the animals (i.e., all animals under "yes" have that trait, animals under "no" do not)

*   Tell the students there may be more than one right answer, and if they get stuck on a question they should move on.

*   Point out that one question is given (Does it live in a house?). House cats do, lions do not. Depending on the class, you may want to give one more question. For example, to distinguish a turtle from a snake the question might be "Does it hiss?"

*   Students may not finish. It will be best to allot an amount of time for this exercise, then gather

the class for a wrap-up discussion.

When the time is up, ask students what types of questions they used. A possible solution is given below for the top and mid-level questions.

Possible Solution:

Notice how questions like "Does it have four legs?" do not work as the top-most question, since the turtle has four legs as well as the rightmost 4 animals. It might take some revising of questions before a suitable set of questions is found.

Top question options:
- Is it a mammal?
- Can it make loud noises?
- Can it hear well?

Left side Second level question options:
- Does it live underwater?
- Is it a fish?

Right side Second level question options:
- Does it eat meat?
- Is it a feline?

What's It All About - Class Discussion

Where else might decision trees be used? See if students have any ideas, then cover some of the examples below. Decision trees are also available in *RealWorldDecisionTrees.pdf*.

Music Preferences and Possible Activity - Spotify

Spotify has an algorithm that suggests new music to listeners. Studies have shown that listeners prefer songs that are familiar. So knowing a bit about listeners' favorite songs could help Spotify suggest new songs that are more likely to be enjoyed. The decision tree in figure 9 shows a possible way that Spotify might approach this problem (this is not based on information from Spotify! Students might want to brainstorm other possible questions).
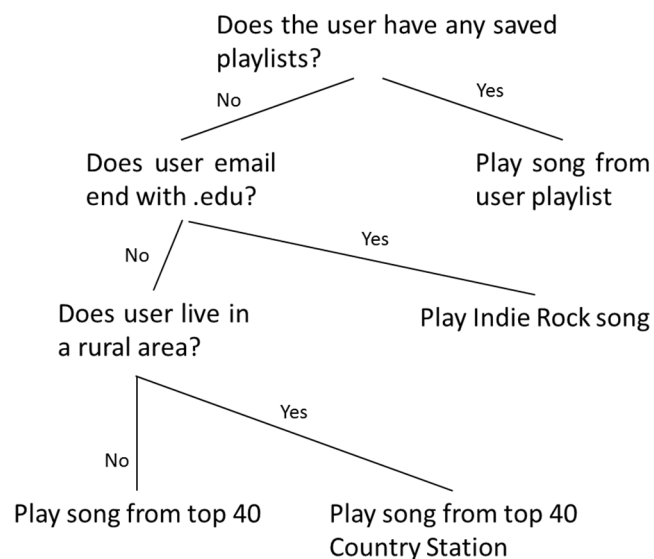
Figure 9: Possible Spotify Tree

If there is extra time in the class, ask the students to make their own decision tree of what to play if they were in charge of the algorithm at Spotify.

Predicting what you want to say

The guessing method can be used to build a computer interface that predicts what the user is going to type next. This can be very useful for physically disabled people who find it difficult to type. The computer suggests what it thinks they are likely to type next, and they just indicate what they want. A good system needs an average of only two yes/no answers per character, and can be of great assistance to someone who has difficulty making the fine movements needed to control a mouse or keyboard. This sort of system is also used in a different form to 'type' text on some cellphones. Figure 10 shows a possible tree that could be used when texting.
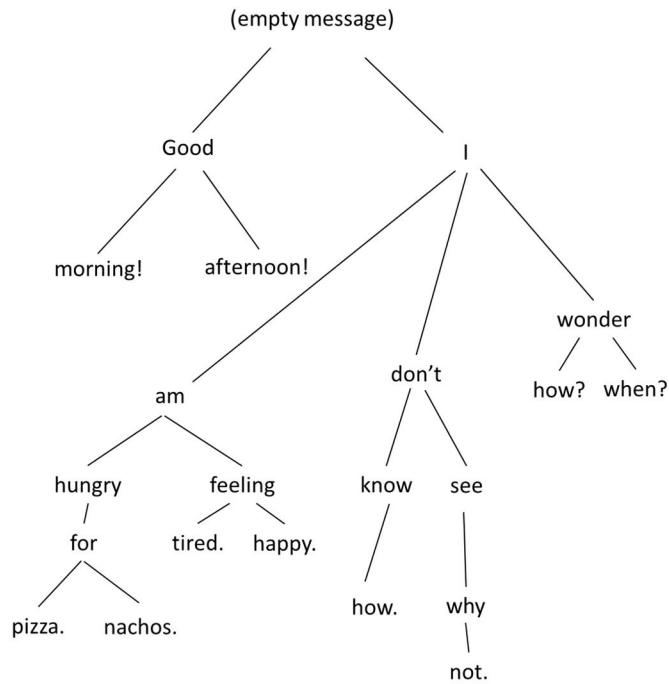
(empty message)

Good

morning!   afternoon!

I

am

hungry   feeling

for

pizza.   nachos.

tired.   happy.

don't

know   see

how.   why

not.

wonder

how?   when?

Figure 10: Predictive texting

Not really anonymous

The animal chart showed how questions can be used to identify animals. How easy is it to uniquely identify people?

1. Weld Medical Data

In 2010, the GIC (Group Insurance Commission) of Massachusetts released anonymized medical data. At the time GIC released the data, William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response, a graduate student named Latanya Sweeney (currently a professor at Harvard) started hunting for the Governor's hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date, only three of them men, and of them, only he lived in his ZIP code. In a theatrical flourish, Dr. Sweeney sent the Governor's health records (which included diagnoses and prescriptions) to his office.

From UCLA Law Review paper, "Broken Promises of Privacy" (Ohm 2010).

2. Simple Demographics Often Identify People Uniquely

Latanya Sweeney's of CMU analysis of datasets: Here are some surprising results using only three fields of information, even though typical data releases contain many more fields. It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}. About half of the U.S. population (132 million of 248 million or 53%) are likely to be uniquely identified by only {place, gender, date of birth}, where place is basically the city, town, or municipality in which the person resides. And even at the county level, {county, gender, date of birth} are likely to uniquely identify 18% of the U.S. population. In general, few characteristics are needed to uniquely identify a person.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000
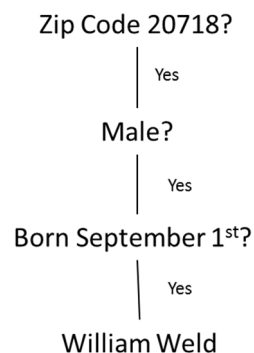
Possible decision tree

Zip Code 20718?

| Yes

Male?

| Yes

Born September 1st?

| Yes

William Weld

Figure 11: 5 pieces of data to indentify someone

Stylometry

By analyzing the patterns in data, it is possible to identify authorship of a work of literature or music just by comparing things like word usage frequency across works. Stylometry is the formal name for attributing authorship to works based solely on their stylistic usage. By gathering statistics based on small idiosyncrasies, an individual's work can stand out to computer analysis even when precautions are taken to hide authorship.

For example, a book called "The Cuckoo's Calling" written under a pseudonym was outed by computer scientist Patrick Juola as being written by J.K Rowling, author of the Harry Potter series of novels by using stylometric analysis. Similar analysis can be used to detect plagiarism.

Figuring Out Personality Traits using Facebook Likes

From: *http://www.cam.ac.uk/research/news/computers-using-digital-footprints-are-better-judges-of-personality-than-friends-and-family*

Psychometrics focuses on measuring personality traits. In the 1980s, researchers identified the "Big Five" in personality traits (OCEAN): openness to new experience, conscientiousness/perfectionism, extroversion, agreeableness and neuroticism. Knowing these traits can provide reasonable predictions about a person's needs, fears, and likely behavior. But it has been hard to assess OCEAN without lengthy questionnaires. Then came big data and Facebook likes:

> In the study, a computer could more accurately predict the subject's personality than a work colleague by analyzing just ten Likes; more than a friend or a cohabitant (roommate) with 70, a family member (parent, sibling) with 150, and a spouse with 300 Likes.

Data garnered in this fashion may be used for targeted marketing, which potentially includes political ad campaigns.