Nonregular Languages
How Do We Know What We Are Missing?

Regular languages may be specified either by regular expressions or by deterministic or nondeterministic finite automata.

How do we show a language is not regular?
Bounded Memory

Since finite automata are not allowed to back up, the amount of memory required to determine whether or not a string is in the language must be bounded.

For example, consider \( L = \{0^n 1^n : n \geq 0\} \).
That is a Thought, Not a Proof!

\[ C = \{ w \mid w \text{ has an equal number of 0s and 1s} \}. \]

\[ D = \{ w \mid w \text{ has an equal number of occurrences of 01 and 10 as substrings} \}. \]
Look at it Another Way

From the specification point of view, a regular language is infinite if and only if its corresponding regular expression contains a Kleene star.

Kleene stars correspond to loops in finite automata.

Both Kleene stars and loops give rise to simple repetitive patterns in the language.
The Pigeonhole Principle

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting an infinite language $L$. Suppose $|Q| = p$.

Let $w = w_1w_2 \cdots w_p \in L$.

$w = w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5 \quad w_6 \quad w_7 \quad \ldots \quad w_p$

$q_0 \quad q_1 \quad q_2 \quad q_3 \quad q_4 \quad q_5 \quad q_6 \quad q_7 \quad q_{p-1} \quad q_p$
Let $q_j$ be the first repeated state, that is, $q_j = q_{j+k}$ for some $k$, $0 \leq j < j+k \leq p$.

Where $w = w_1 w_2 \cdots w_p = xyz$, 
$x = w_1 w_2 \cdots w_j$ \hspace{1cm} $y = w_{j+1} \cdots w_{j+k}$ \hspace{1cm} $z = w_{j+k+1} \cdots w_p$

We conclude that $xy^i z \in \mathcal{L}$ for all $i \geq 0$. 

Machine Loops
The Pumping Lemma

Theorem. If $A$ is a regular language, then there is a number $p$ (the pumping length) where if $w$ is a string of length at least $p$, then $w = xyz$, such that

1. for each $i \geq 0$, $xy^iz \in A$,
2. $|y| > 0$, and
3. $|xy| \leq p$. 
Deciding Regularity

Is $L = \{ 0^n 1^n : n \geq 0 \}$ regular? If so, then there are strings $x, y, \text{ and } z$ such that $xy^iz \in L$ for all $i \geq 0$. What does $y$ look like?

- String $y$ consists entirely of 0s?
- String $y$ consists entirely of 1s?
- String $y$ consists of both 0s and 1s?
Combining Results

Is $C = \{ w \mid w \text{ has an equal number of 0s and 1s} \}$ a regular language?
Picking the Right Substring to Pump

Is $PAL = \{ \ w \in \{0, 1\}^* : w \text{ is a palindrome} \}$ a regular language?