The Pumping Lemma

Some languages are not regular languages. The *pumping lemma* can be used to show this. It uses proof by contradiction and the pigeon-hole principle.

Application of pigeonhole principle: If a string is as long or longer than the number of states in a DFA, then some state is visited more than once.

Application of proof by contradiction: Assume $L$ corresponds to some DFA $M$. Let $m$ be the number of states in $M$. Create a long string $w \in L$. Show that if $M$ accepts $w$, then $M$ must accept some string $w' \notin L$. This contradicts assumption that $\mathcal{L}(M) = L$. 
Proof of the Pumping Lemma

Theorem: Let $L$ be a regular language. There exists a number $m$, for all $w$, if $|w| \geq m$ and $w \in L$, then there exists $x, y, z$ such that

$$w = xyz,$$

$$|xy| \leq m,$$

$$|y| \geq 1,$$

$$xy^i z \subseteq L,$$ i.e., $xy^i z \in L$ for all $i \geq 0$

Proof: Let $M$ be a DFA such that $L = \mathcal{L}(M)$. Let $m$ be the number of states in $M$. Suppose $|w| \geq m$ and $w \in L$. Then a repetition of states in first $m$ symbols. Let $w = xyz$, where $\delta^*(q_0, x) = \delta^*(q_0, xy)$.

Clearly, any $xy^i$ leads to the same state, from which $z$ leads to the final state.
Using the Pumping Lemma

$L_1 = \{a^nb^n : n \geq 0\}$
Suppose a DFA $M_1$ accepts $L_1$.
Let $m$ be the number of states in $M_1$.
Choose $w = a^mb^m$.
$M_1$ must repeat states reading $a^m$.

By the PL, if $M_1$ accepts $a^mb^m$, then $M_1$ accepts strings with more $a$’s without changing the number of $b$’s.

Contradicts assumption that $M_1$ accepts $L_1$.

$L_2 = \{a^lb^n : l \geq n\}$
Suppose a DFA $M_2$ accepts $L_2$.
Let $m$ be the number of states in $M_2$.
Choose $w = a^mb^m$.
$M_2$ must repeat states reading $a^m$.

By the PL, if $M_2$ accepts $a^mb^m$, then $M_2$ accepts one string with fewer $a$’s without changing the number of $b$’s.

Contradicts assumption that $M_2$ accepts $L_2$. 
\( L_3 = \{ww^R : w \in \{a, b\}^*\} \)
Suppose a DFA \( M_3 \) accepts \( L_3 \).
Let \( m \) be the number of states in \( M_3 \).
Choose \( w = a^m b b a^m \).
\( M_3 \) must repeat states reading \( a^m \).

By the PL, if \( M_3 \) accepts \( a^m b b a^m \), then \( M_3 \) accepts strings with more \( a \)'s on the left without changing the number of \( a \)'s on the right.
Contradicts assumption that \( M_3 \) accepts \( L_3 \).

\( L_4 = \{a^{2^k} : k \geq 0\} \)
Suppose a DFA \( M_4 \) accepts \( L_4 \).
Let \( m \) be the number of states in \( M_4 \).
Choose \( w = a^{2^n} \), where \( 2^n > m \).
\( M_4 \) must repeat states reading first \( m \) \( a \)'s

By the PL, if \( M_4 \) accepts \( a^{2^n} \), then \( M_4 \) accepts a string with 1 to \( m \) more \( a \)'s.
However \( 2^n < 2^n + m < 2^{n+1} \), i.e., the number of \( a \)'s won't be a power of 2.
Contradicts assumption that \( M_4 \) accepts \( L_4 \).
\[ L_5 = \{ a^n : n \text{ is not a power of } 2 \} \]
If \( L_5 \) was regular,
then \( \overline{L_5} = L_4 \) would be regular,
which is a contradiction.