Language Definitions and Notations

See Lecture 2
Here are a bunch of basic definitions that we will use all semester. There is nothing exciting here but you won't be able to follow much until you get these in your head.

\( \Sigma \) is a finite set of symbols called our \textit{alphabet}. This could be the set \{0,1\} of binary digits, or the set of lower-case letters 'a' to 'z'. Don't let the term "alphabet" confuse you. \( \Sigma \) could also be the set of valid Java keywords and identifiers up to length 64 (so it is finite). Any finite set of atomic elements will do.

A string or word over \( \Sigma \) is any finite sequence of elements of \( \Sigma \).

\( \varepsilon \) represents the \textit{empty string}: the string of length 0
\( \Sigma^n \) is the set of strings over \( \Sigma \) of length \( n \) (exactly \( n \)).

\( \Sigma^* \) is the set of all strings over \( \Sigma \), including the empty string.

\( \Sigma^+ \) is the set of all strings with positive length over \( \Sigma \).

Obviously, \( \Sigma^* = \Sigma^+ \cup \{ \varepsilon \} \)

Kozen avoids this terminology, but most people say that a language over \( \Sigma \) is any subset of \( \Sigma^* \).
Question 1: How big is $\Sigma^*$?
   Well, if $\Sigma$ is the empty set then $\Sigma^*$ is $\{\varepsilon\}$. If $\Sigma$ is not empty then $\Sigma^*$ is countable -- it is a countable union of finite sets.

Question 2: How many languages are there over $\Sigma$?
   If $\Sigma$ is empty there are two, both trivial: {} and {$\varepsilon$}.
   If $\Sigma$ is not empty there are uncountably many languages over it (for if you could number the subsets of $\Sigma^*$ you could create a new subset that wasn't in any of them.)