1. (**50 points**) For each language below, determine whether or not the language is context-free, and justify your answer. For example, if the language **is** context-free, you can give a CFG or PDA for the language. If it is not context-free, you can use the pumping lemma for CFLs and/or other properties of CFLs to prove it. You may use relationships to other languages that we have shown to be context-free or not, and you can use the results from previous parts of this problem when solving subsequent parts.

   (a) **(10 points)** $A_1 = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid \exists x, y \in \{a, \mathtt{b}\}^*. \ (w = x\mathtt{c}y) \wedge (x \text{ is a substring of } y)\}$
   Solution:
   A_1 is not context-free.
   Let p be a proposed pumping lemma constant, and let $w = a^p b^p c a^p b^p \in A_1$. If $w = uvxyz$ such that $1 \leq |vy| \leq p$, then if v and y appear before the c, then $uv^2xy^2z \notin A_1$ because the string before c is longer than the string following c and is therefore not a substring. If either of v or y contain c, then $uv^0xy^0z \notin A_1$ because it does not contain a c. If v and y appear folowing the c, then $uv^0xy^0z \notin A_1$ the string following c is shorter than the string before c, and is therefore not a substring. If v appears before c and y appears after, then $uv^0xy^0z = a^p b^{p-|v|} c a^{p-|y|} b^p \notin A_1$ because more a's appear before the c than after if $|y| > 0$. Else, $|v| > 0$ and $uv^2xy^2z = a^p b^{p+|v|} c a^{p+|y|} b^p \notin A_1$ because more b's appear before the c than after.

   (b) **(10 points)** $A_2 = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid \exists i, j \in \mathbb{Z}^{\geq 0}. \ w = \mathtt{a}^i \mathtt{b}^i c^j$.
   Solution:
   A_2 is context-free.
   $S \rightarrow TC$
   $T \rightarrow aTb|\epsilon$
   $S \rightarrow cC|\epsilon$

   (c) **(10 points)** $A_3 = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}^* \mid \exists i, j \in \mathbb{Z}^{\geq 0}. \ w = \mathtt{a}^i \mathtt{b}^i \mathtt{c}^j \mathtt{d}^j\}$.
   Solution:
   A_3 is context-free.
   $S \rightarrow TU$
   $T \rightarrow aTb|\epsilon$
   $U \rightarrow cUd|\epsilon$

   (d) **(20 points)** $A_6 = \overline{E}$, where $E = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid \exists n \in \mathbb{Z}^{\geq 0}. \ w = \mathtt{a}^n \mathtt{b}^n \mathtt{c}^n\}$.
   Solution:
   A_6 is context-free.
   We view A_6 as the union of three languages A, B and C. With $\Sigma = \{a, b, c\}$, let $A = L((\Sigma^*(ba)\Sigma^*) \cup (\Sigma^*(ca)\Sigma^*) \cup (\Sigma^*(ca)\Sigma^*))$, that is, A is all strings that do not have symbols appearing in the order of a's followed by b's followed by c's. A is regular, and is therefore context-free. Let $B = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid \exists i, j, k \in \mathbb{Z}^{\geq 0}. \ (i \neq j) \wedge w = \mathtt{a}^i \mathtt{b}^j \mathtt{c}^k\}$, and let $C = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid \exists i, j, k \in \mathbb{Z}^{\geq 0}. \ (j \neq k) \wedge w = \mathtt{a}^i \mathtt{b}^j \mathtt{c}^k\}$.
   B is given by the following CFG:
   $S \rightarrow ATU|TBU$
   $A \rightarrow aA|a$
   $T \rightarrow aTb|\epsilon$
   $U \rightarrow cU|\epsilon$
   $B \rightarrow bB|b$

   C is given by the following CFG:
   $S \rightarrow ATU|ABT$
   $A \rightarrow aA|\epsilon$
   $T \rightarrow bTc|\epsilon$

$U \to cU|c$
$B \to bB|b$

Since $A_6 = A \cup B \cup C$ and A, B and C are context-free and context-free languages are closed under union, A_6 is context-free.

2. (**20 points**) Let $G = (V, \Sigma, R, S)$ be a CFG in Chomsky normal form. Show that if G generates any string, w, with $|w| > 2^{|V|-1}$, then $L(G)$ is an infinite set.
Solution:
Following the proof strategy from the October 3rd slides:
There is a string w with $|w| > 2^{|V|-1}$, which implies that $|w| \geq 2^{|V|-1} + 1$. From property 1 of $height$, $height(S_0, w) \geq \lceil \log_2 (2^{|V|-1} + 1) \rceil + 1 = |V| + 1$. From property 2 of $height$, there must be a path from S_0 to a terminal in w that passes through at least $|V| + 1$ variables. By the pigeon-hole principle, two of these parse-tree nodes are interchangable, and we may pump as with the pumping lemma for context free languages to generate strings in the language of length that is arbitrarily long. This implies that $L(G)$ is an infinite set.

3. **(20 points)** Draw the state diagram for a Turing machine that recognizes the language of all strings in $\{\texttt{a}, \texttt{b}\}^*$ for which the number of a's in the string is a perfect square.
Solution:
See the diagram in the seperate pdf file, hw6q3.pdf. Note that denotes the blank tape symbol, and that transition labels of the form $\{a, b\}, c \to R$ means "if symbol a or symbol b is read, write symbol c and move the tape head to the right". Here's a description of how the TM works. We use symbols s and o to track the current square to check, and the next odd number to add to s to generate the next square. The state labeled A is the accepting state. The path from 0 to 9 checks if the input string length is equal to 0, 1, 4 or 9. The path from 9 to q_2 repeatedly finds an a to the left of o, marks it as a', and then finds an a to the right of s and marks it as a'. This repeats until no a exists to the left of o. At this point, the TM takes the path from q_3 to q_4. This path does 3 things: (1) advances the o symbols by two positions to the right, (2) moves the s symbol to the tape location of the next square, which by construction is the cell to the right of the rightmost a', and (3) cleans up the tape by changing all marked a' symbols back to a. Then the TM returns to state 9 where we check for a blank to the right. If there is a blank, the TM accepts and if not, we repeat the procedure of adding o to s by looping from q_1 to q_2.