Let \( L \) be the set of all strings over \( \{0, 1\}^* \) with exactly twice as many 0s as 1s.

1.A. Describe a CFG for the language \( L \).

(Hint: For any string \( u \) define \( \Delta(u) = \#(0,u) - 2\#(1,u) \). Introduce intermediate variables that derive strings with \( \Delta(u) = 1 \) and \( \Delta(u) = -1 \) and use them to define a non-terminal that generates \( L \).)

Solution:

\[
S \rightarrow \varepsilon \mid SS \mid 00S1 \mid 0S1S0 \mid 1S00
\]

1.B. Prove that your grammar \( G \) is correct. As usual, you need to prove both \( L \subseteq L(G) \) and \( L(G) \subseteq L \).

(Hint: Let \( u_{\leq i} \) denote the prefix of \( u \) of length \( i \). If \( \Delta(u) = 1 \), what can you say about the smallest \( i \) for which \( \Delta(u_{\leq i}) = 1 \)? How does \( u \) split up at that position? If \( \Delta(u) = -1 \), what can you say about the smallest \( i \) such that \( \Delta(u_{\leq i}) = -1 \)?)

Solution:

We separately prove \( L \subseteq L(G) \) and \( L(G) \subseteq L \) as follows:

**Claim 4.1.** \( L(G) \subseteq L \), that is, every string in \( L(G) \) has exactly twice as many 0s as 1s.

**Proof:** As suggested by the hint, for any string \( u \), let \( \Delta(u) = \#(0,u) - 2\#(1,u) \). We need to prove that \( \Delta(w) = 0 \) for every string \( w \in L(G) \).

Let \( w \) be an arbitrary string in \( L(G) \), and consider an arbitrary derivation of \( w \) of length \( k \). Assume that \( \Delta(x) = 0 \) for every string \( x \in L(G) \) that can be derived with fewer than \( k \) productions. There are five cases to consider, depending on the first production in the derivation of \( w \).

- If \( w = \varepsilon \), then \( \#(0,w) = \#(1,w) = 0 \) by definition, so \( \Delta(w) = 0 \).
- Suppose the derivation begins \( S \rightarrow SS \rightarrow^* w \). Then \( w = xy \) for some strings \( x, y \in L(G) \), each of which can be derived with fewer than \( k \) productions. The inductive hypothesis implies \( \Delta(x) = \Delta(y) = 0 \). It immediately follows that \( \Delta(w) = 0 \).
- Suppose the derivation begins \( S \rightarrow 00S1 \rightarrow^* w \). Then \( w = 00x1 \) for some string \( x \in L(G) \). The inductive hypothesis implies \( \Delta(x) = 0 \). It immediately follows that \( \Delta(w) = 0 \).
- Suppose the derivation begins \( S \rightarrow 1S00 \rightarrow^* w \). Then \( w = 1x00 \) for some string \( x \in L(G) \). The inductive hypothesis implies \( \Delta(x) = 0 \). It immediately follows that \( \Delta(w) = 0 \).
- Suppose the derivation begins \( S \rightarrow 0S1S1 \rightarrow^* w \). Then \( w = 0x1y0 \) for some strings \( x, y \in L(G) \). The inductive hypothesis implies \( \Delta(x) = \Delta(y) = 0 \). It immediately follows that \( \Delta(w) = 0 \).
In all cases, we conclude that $\Delta(w) = 0$, as required. ■

**Claim 4.2.** $L \subseteq L(G)$; that is, $G$ generates every binary string with exactly twice as many 0s as 1s.

**Proof:** As suggested by the hint, for any string $u$, let $\Delta(u) = \#(0, u) - 2\#(1, u)$. For any string $u$ and any integer $0 \leq i \leq |u|$, let $u_i$ denote the $i$th symbol in $u$, and let $u_{\leq i}$ denote the prefix of $u$ of length $i$.

Let $w$ be an arbitrary binary string with twice as many 0s as 1s. Assume that $G$ generates every binary string $x$ that is shorter than $w$ and has twice as many 0s as 1s. There are two cases to consider:

- If $w = \varepsilon$, then $\varepsilon \in L(G)$ because of the production $S \rightarrow \varepsilon$.
- Suppose $w$ is non-empty. To simplify notation, let $\Delta_i = \Delta(w_{\leq i})$ for every index $i$, and observe that $\Delta_0 = \Delta \big|_w = 0$. There are several subcases to consider:
  
  - Suppose $\Delta_i = 0$ for some index $0 < i < |w|$. Then we can write $w = xy$, where $x$ and $y$ are non-empty strings with $\Delta(x) = \Delta(y) = 0$. The induction hypothesis implies that $x, y \in L(G)$, and thus the production rule $S \rightarrow SS$ implies that $w \in L(G)$.
  
  - Suppose $\Delta_i > 0$ for all $0 < i < |w|$. Then $w$ must begin with 00, since otherwise $\Delta_1 = -2$ or $\Delta_2 = -1$, and the last symbol in $w$ must be 1, since otherwise $\Delta_{|w|-1} = -1$. Thus, we can write $w = 00x1$ for some binary string $x$. We easily observe that $\Delta(x) = 0$, so the induction hypothesis implies $x \in L(G)$, and thus the production rule $S \rightarrow 00S1$ implies $w \in L(G)$.
  
  - Suppose $\Delta_i < 0$ for all $0 < i < |w|$. A symmetric argument to the previous case implies $w = 1x00$ for some binary string $x$ with $\Delta(x) = 0$. The induction hypothesis implies $x \in L(G)$, and thus the production rule $S \rightarrow 1S00$ implies $w \in L(G)$.
  
  - Finally, suppose none of the previous cases applies: $\Delta_i < 0$ and $\Delta_j > 0$ for some indices $i$ and $j$, but $\Delta_i \neq 0$ for all $0 < i < |w|$.

  Let $i$ be the smallest index such that $\Delta_i < 0$. Because $\Delta_j$ either increases by 1 or decreases by 2 when we increment $j$, for all indices $0 < j < |w|$, we must have $\Delta_j > 0$ if $j < i$ and $\Delta_j < 0$ if $j \geq i$.

  In other words, there is a unique index $i$ such that $\Delta_{i-1} > 0$ and $\Delta_i < 0$. In particular, we have $\Delta_1 > 0$ and $\Delta_{|w|-1} < 0$. Thus, we can write $w = 0x1y0$ for some binary strings $x$ and $y$, where $|0x1| = i$.

  We easily observe that $\Delta(x) = \Delta(y) = 0$, so the inductive hypothesis implies $x, y \in L(G)$, and thus the production rule $S \rightarrow 0S1S0$ implies $w \in L(G)$.

In all cases, we conclude that $G$ generates $w$. ■

Together, Claim 1 and Claim 2 imply $L = L(G)$.

**Rubric:** 10 points:
• part (a) = 4 points. As usual, this is not the only correct grammar.
• part (b) = 6 points = 3 points for $\subseteq$ + 3 points for $\supseteq$, each using the standard induction template (scaled).