1. For each part, let L refer to the language in question and let $x \sim_L y$ if and only if x and y are indistinguishable by suffix with respect to L. Myhill-Nerode states that L is regular if and only if $L/\sim_L$ is finite. To show that each language is not regular, we show that there exists an infinite set F such that the canonical projection $\pi_L : L \to L/\sim_L$ restricted to F is injective. This means if $\pi_L(x) = \pi_L(y)$, then $x = y$. This can be interpreted as no two distinct elements in F get mapped to the same equivalence class, or no two distinct elements in F are indistinguishable.

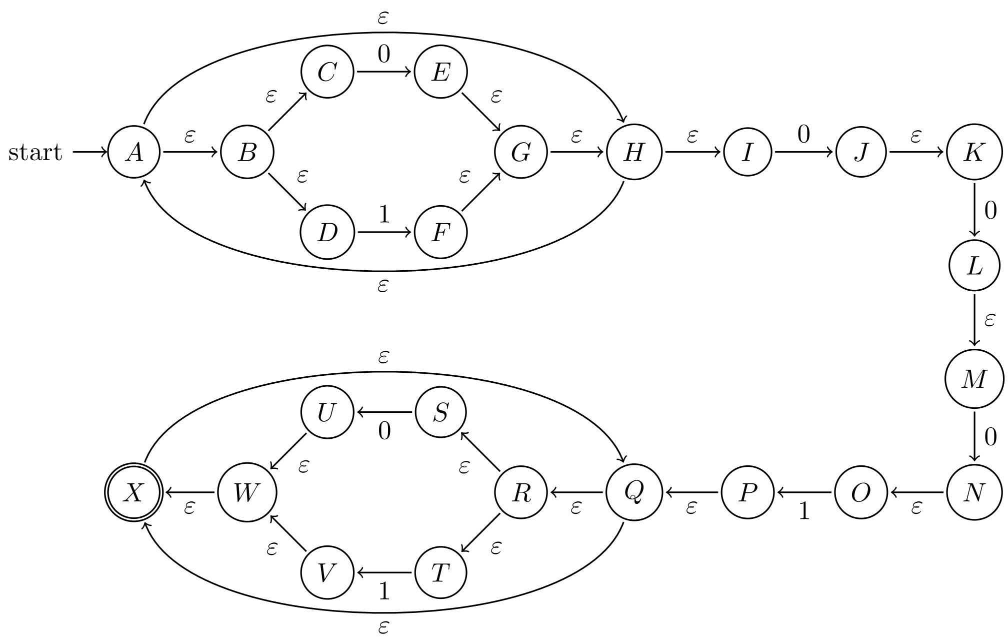
   (a) Consider the set $F = \left\{10^b 1 : b \in \mathbb{Z}_0^+\right\}$. Let $x, y \in F$ be distinct. Namely, $x = 10^{b_1}1, y = 10^{b_2}1$ with $b_1 \neq b_2$. Consider the suffix $z = 0^{b_1}$. We compute $xz = 0^0 10^{b_1} 10^{b_1}$ and $yz = 0^0 10^{b_2} 10^{b_1}$, where xz is in the language because $0 + b_1 = b_1$ and yz is not in the language because $0 + b_2 \neq b_1$. Thus $L/\sim_L$ is infinite so this language is not regular.

   (b) Consider the set $F = \left\{0^{2k} : k \in \mathbb{Z}_0^+\right\}$. Let $x, y \in F$ be distinct, so we can write $x = 0^{2a}, y = 0^{2b}, a \neq b$. Assume without loss of generality that $a < b$. Then consider the suffix $z = 1^a$. Compute $xz = 0^{2a}1^a$, satisfying $\#(0, xz) = 2\#(1, xz)$, so $xz \in L$. Also compute $yz = 0^{2b}1^a$, satisfying $\#(0, yz) > \#(0, xz) = 2\#(1, xz) = 2\#(1, yz)$, so $y \notin L$. Thus z is a distinguishing suffix, and $L/\sim_L$ is infinite so this language is not regular.

   (c) Consider the set $F = \{0^p : p \text{ is prime}\}$. Let $x, y \in F$ be distinct so we can write $x = 0^p, y = 0^q, p \neq q$. Consider the suffix $z = 1^p$. Then compute $xz = 0^p 1^p$, and $\gcd(p, p) = p$, so $xz \notin L$. Also compute $yz = 0^q 1^p$, and $\gcd(p, q) = 1$ because p and q are distinct primes, so $yz \in L$. Thus z is a dinstinguishing suffix and $L/\sim_L$ is infinite so this language is not regular.

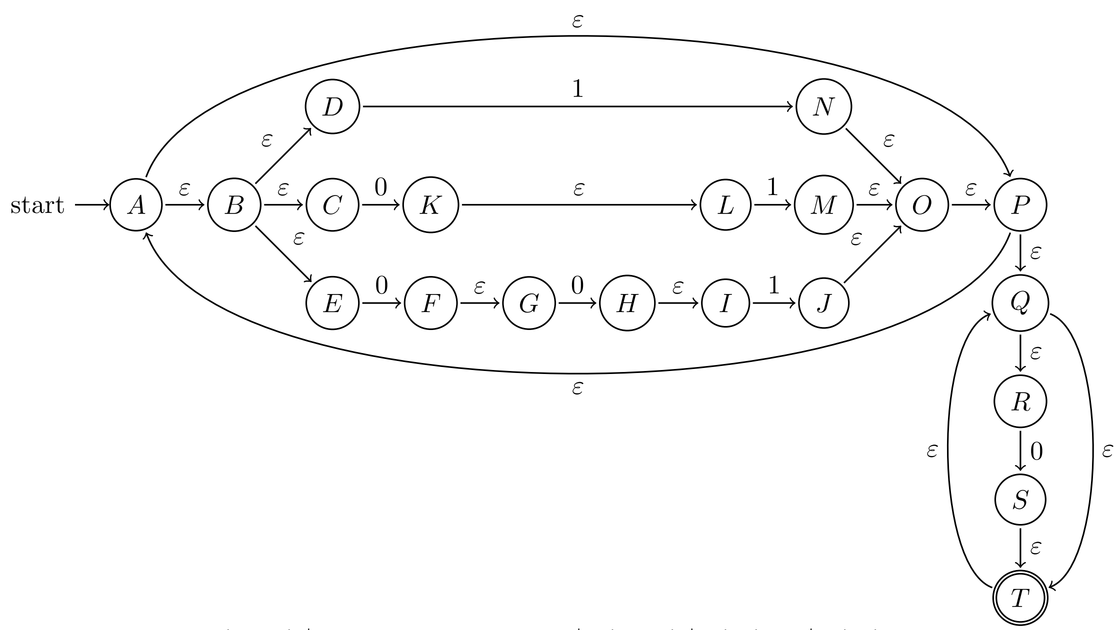
2. We give the NFA first and then the DFA in table form similar to the format used in the lecture notes. We also write out the elements in the ε-reach of the states in our NFA in string format in alphabetical order, to save space. e.g. the string $ABCDHI$ represents the ε-reach of $\{A, B, C, D, H, I\}$.

   (a)



$q' \in Q'$	ε-reach	$q' \in A'$	$\delta'(q', 0)$	$\delta'(q', 1)$
A	$ABCDHI$		EJ	F
EJ	$ABCDEGHIJK$		EJL	F
F	$ABCDFGHI$		EJ	F
EJL	$ABCDEGHIJKLM$		$EJLN$	F
$EJLN$	$ABCDEGHIJKLMNO$		$EJLN$	FP
FP	$ABCDFGHIPQRSTX$	✓	EJU	FV
EJU	$ABCDEGHIJKQRSTUWX$	✓	$EJLU$	FV
FV	$ABCDFGHIQRSTVWX$	✓	EJU	FV
$EJLU$	$ABCDEGHIJKLMQRSTUWX$	✓	$EJLNU$	FV
$EJLNU$	$ABCDEGHIJKLMNOQRSTUWX$	✓	$EJLNU$	FPV
FPV	$ABCDFGHIPQRSTVWX$	✓	EJU	FV

(b) For this particular NFA, we simply the ε-transitions generated from the two unions in $1+01+001$ to originate from the same state (B).



$q' \in Q'$	ε-reach	$q' \in A'$	$\delta'(q', 0)$	$\delta'(q', 1)$
A	$ABCDEPQRT$	✓	FKS	N
FKS	$FGKLQRST$	✓	HS	M
N	$ABCDENOPQRT$	✓	FKS	N
HS	$HIQRST$	✓	S	J
M	$ABCDEMOPQRT$	✓	FKS	N
S	$QRST$	✓	S	FAIL
J	$ABCDEJOPQRT$	✓	FKS	N
FAIL	$\varnothing$		FAIL	FAIL

FAIL is the unique state representing when all possible paths through the NFA have failed.

3. We repeat our restatement of Myhill-Nerode, because this is a different problem from 1.

   For each part, let L refer to the language in question and let $x \sim_L y$ if and only if x and y are indistinguishable by suffix with respect to L. Myhill-Nerode states that L is regular if and only if $L/\sim_L$ is finite. To show that each language is not regular, we show that there exists an infinite set F such that the canonical projection $\pi_L : L \to L/\sim_L$ restricted to F is injective. This means if $\pi_L(x) = \pi_L(y)$, then $x = y$. This can be interpreted as no two distinct elements in F get mapped to the same equivalence class, or no two distinct elements in F are indistinguishable.

   (a) Consider the set $F = \{0^k : k \geq 2\}$. The substring 00 occurs in 0^k precisely $k-1$ times (the $k-1$ substrings of length 2 are all 00). Let $x, y \in F$ be distinct, so we can write $x = 0^a, y = 0^b, a \neq b$. Consider the suffix $z = 1^a$. The substring 11 clearly occurs $a-1$ times in both xz and yz, but the substring 00 only occurs $a-1$ times in xz and $b-1$ times in yz. Thus $xz \in L$ but $yz \notin L$, so z is a dinstinguishing suffix and $L/\sim_L$ is infinte so this language is not regular.

   (b) We make the following observation. The number of occurrences of 10 and the number of occurrences of 01 never differ by more than 1. Recall that $(0+1)^* = 0^*(1^+0^+)^* + 1^*(0^+1^+)^*$. Every string is can either be grouped into blocks of 1^a0^b's or 0^a1^b's, with a choice of a preceeding block of characters that cannot be absorbed into the Kleene star. The number of alternations from 1 to 0 and 0 to 1 are exactly the same when the string starts and ends with the same character, e.g. $1^+(0^+1^+)^*$, so we have to exclude these with the transformation $1^+(0^+1^+)^* \mapsto (0^+1^+)^+$. L can then be represented by $(1^+0^+)^+ + (0^+1^+)^+$.

   (c) This language is clearly regular. It is clearly a subset of Σ^* because it consists entirely of strings concatenated from Σ^*. Every string $x \in \Sigma^*$ is also in L because $x = \varepsilon x \varepsilon \in L$. Thus $L = \Sigma^*$ and Σ^* is regular, represented by $(0+1)^*$.

   (d) Let $F = \{0^k1^k : k \in \mathbb{Z}^+\}$ and let x, y be two distinct strings in F so that we can write $x = 0^a1^a, y = 0^b1^b$ with $a \neq b$. Consider the suffix $z = 10^a1^a$. Assume WLOG that $a < b$. Clearly $xz = (0^a1^a)1(0^a1^a) \in L$. We claim $yz = (0^b1^b)1(0^a1^a) \notin L$ because none of the prefixes of yz match the suffixes of yz.

   Let $yz = uwv$ be a decomposition of yz into non-zero length prefixes and suffixes of the same length, with $w \neq \varepsilon$. We split into the following cases.

   $0 < |u| = |v| < b$. Then $\#(1, u) = 0$ because the first $a < b$ characters of u are necessarily 0's, and $\#(1, v) > 0$ because v is non-empty and ends with 1.

   $b \leq |u| = |v| \leq 2a$. Then $\#(0, u) = b$ but $\#(0, v) \leq a < b$, so this is also impossible. Such a decomposition may not always be possible, e.g. when $b > 2a$, but whenever this decomposition is possible, this is necessarily true.

   $2a < |u| = |v| \leq a + b$. Then u starts with 0 and v starts with 1, so this is also impossible.

   $|u| = |v| > a + b$ is an impossible decomposition because $|yz| = 2a + 2b + 1$ and $|u| + |v| + 1 > 2a + 2b + 1$ in this case.

   This shows that all possible decompositions of $yz = uwv$ with $|u| = |v|$ satisfy $u \neq v$, so $yz \notin L$, as desired. Thus our z distinguishing suffix and $L/\sim_L$ is infinite so this language is not regular.