Theory of Computation, CSCI 438 spring 2022
Regular expressions describe regular languages, pg. 66-69, Jan. 28

From before:
Theorem 1.39 Every NFA has an equivalent DFA. (Page 55)
Corollary 1.40 A language is regular iff some NFA recognizes it. (Page 56)

Theorem: A language is regular iff some regular expression describes it.
(Theorem 1.54, page 66)

Proof ($\Leftarrow$) the "if" direction:

Given a regular expression R, it will be shown that the language described by R is regular. Since we know that given any NFA a DFA which recognizes the same language can be gotten, it suffices to show that there is an NFA that recognizes the language described by R.

Since R is a regular expression it was built according to the rules of definition 1.52. Since the rules are defined recursively, we can use induction. First show that for every regular expression in the basis, there is an NFA which recognizes it.

Basis:

- a where $a \in \Sigma$ is a regular expression
  can be described by the NFA $M = (\{q_0,q_1\}, \Sigma, \{((q_0,a),q_1)\}, q_0, \{q_1\})$

- ε is a regular expression
  can be described by the NFA $M = (\{q_0\}, \Sigma, \{\}, q_0, \{q_0\})$

- Φ is a regular expression
  can be described by the NFA $M = (\{q_0\}, \Sigma, \{\}, q_0, \{\})$

Next assume the inductive hypothesis, that for any regular expressions R_1 and R_2 there are NFAs, $M_1 = (Q_1, \sum, \delta_1, q_{1,0}, F_1)$ and $M_2 = (Q_2, \sum, \delta_2, q_{2.0}, F_2)$, respectively which recognize the languages described by R_1 and R_2.

- $R_1 \cup R_2$ is a regular expression
  can be described by the NFA $M_{union} = (Q_1 \cup Q_2 \cup \{q_{new}\}, \sum, \delta, q_{new}, F_1 \cup F_2)$ where
    - δ (q,x) is equal to the following:
        - δ_1 (q,x) for $q \in Q_1$
        - δ_2 (q,x) for $q \in Q_2$
        - $\{q_{1,0}, q_{2.0}\}$ for $q = \{q_{new}\}$ and $x = \varepsilon$

- $R_1 \circ R_2$ is a regular expression
  can be described by the NFA $M_{concat} = (Q_1 \cup Q_2, \sum, \delta, q_{1,0}, F_2)$ where
    - δ (q,x) is equal to the following:
        - δ_1 (q,x) for $q \in Q_1\text{-}F_1$ or $x \in \sum$
        - δ_1 (q,x) $\cup \{q_{2,0}\}$ for $q \in F_1$ and $x = \varepsilon$
        - δ_2 (q,x) for $q \in Q_2$

- R_1* is a regular expression
  can be described by the NFA $M_{Kleene} = (Q_1 \cup \{q_{new}\}, \sum, \delta, q_{new}, \{q_{new}\})$ where
    - δ (q,x) is equal to the following:
        - δ_1 (q,x) for $q \in Q_1\text{-}F_1$ or $x \in \sum$
        - δ_1 (q,x) $\cup \{q_{new}\}$ for $q \in F_1$ and $x = \varepsilon$

Thus we have shown that if the inductive hypothesis holds, there are NFAs which recognize the languages described by the three new regular expressions above. Thus by the Principle of Strong Mathematical Induction, given any regular expression R, there is an NFA which recognizes the language described by R.

Since, from a previous theorem, any NFA can be converted into a DFA and since any language which can be recognized by a DFA is regular, we have shown that the language described by any regular expression is regular.