Given episode Sets:

One episode like this: $s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, r_3, s_3, \cdots, r_{T-1}, s_{T-1}, a_{T-1}, r_T, s_T$

$$R_s = R(s) = R(S_t) = E[R_{t+1} \mid S_t = s]$$

$$\mathrm{G}_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots, \quad \lambda \in [0,1]$$

$$v(s) = v(S_t) = E[G_t \mid S_t = s]$$

$$v(s) = E[R_{t+1} + \gamma \times v(S_{t+1}) \mid S_t = s]$$

$$v(s) = R_s + \gamma \sum_{s' \in S} P_{ss'} v(s')$$

$$v = R + \gamma P v$$

$$v = (I - \gamma P)^{-1} R$$

$$\begin{bmatrix} v(1) \\ \vdots \\ v(n) \end{bmatrix} = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} + \gamma \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} v(1) \\ \vdots \\ v(n) \end{bmatrix}$$

$$P_{ss'}^a = P[S_{t+1} = s' \mid S_t = s, A_t = a]$$

$$R_s^a = E[R_{t+1} \mid S_t = s, A_t = a]$$

$$\pi(a \mid s) = P[A_t = a \mid S_t = s]$$

$S_1,S_2,S_3,\ldots$ is a Markov process <S,P^{π}>
$S_1,R_2,S_2,R_3,S_3,\ldots$ is a Markov reword process <S,P^{π},R^{π}, γ >
Given a Markov Decision Process M = <S,A,P,R, γ > and a policy π
Where:

$$P_{ss'}^{\pi} = \sum_{a \in A} \pi(a \mid s) p_{ss'}^a$$

$$R_s^{\pi} = \sum_{a \in A} \pi(a \mid s) R_s^a$$

State_value function $v_{\pi}(s)$

$$v_{\pi}(s) = E_{\pi}[G_t \mid S_t = s]$$

Action_value function $q_{\pi}(s,a)$

$$q_{\pi}(s,a) = E_{\pi}[Gt \mid S_t = s, A_t = a]$$

* $v_{\pi}(s) = E_{\pi}[R_{t+1} + \gamma v_{\pi}(S_{t+1}) \mid S_t = s]$

* $q_\pi(s,a) = E_\pi[R_{t+1} + \gamma q_\pi(S_{t+1}, A_{t+1}) \mid S_t = s, A_t = a]$

* $v_\pi(s) = \sum_{a \in A} \pi(a \mid s) q_\pi(s,a)$

* $q_\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_\pi(s')$

put above two function together

** $v_\pi(s) = \sum_{a \in A} \pi(a \mid s)[R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_\pi(s')]$

** $q_\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \sum_{a' \in A} \pi(a' \mid s') q_\pi(s',a')$

** $v_\pi = (I - \gamma P^\pi)^{-1} R^\pi$

Optimal Value Function

*** $v_*(s) = \max_\pi v_\pi(s) \quad \forall \pi$

*** $q_*(s,a) = \max_\pi q_\pi(s,a) \quad \forall \pi$

$$\pi \geq \pi' \quad if \quad v_\pi(s) \geq v_{\pi'}(s), \forall s$$

$$\pi_* \geq \pi, \quad \forall \pi$$

$$v_{\pi_*}(s) = v_*(s)$$

$$q_{\pi_*}(s,a) = q_*(s,a)$$

$$\pi_*(a \mid s) = \begin{cases} 1 \quad if \quad a = \arg\max_{a \in A} q_*(s,a) \\ 0 \qquad\qquad\qquad otherwise \end{cases}$$

$$v_*(s) = \max_a q_*(s,a)$$

$$q^*(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_*(s')$$

$$v^*(s) = \max_a R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_*(s')$$

$$q^*(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \max_{a'} q_*(s',a')$$