A.A. The determinant of A is 0, by the square matrix theorem.

A.B. The column space of A is not all of $\mathbb{R}^n$ (n-dimensional space), by the square matrix theorem. In other words, A has rank less than n.

A.C. Because A is $n \times n$, the SVD writes A as $A = U\Sigma V^\top$, where U, Σ, and V are all $n \times n$, U and V are orthogonal, and Σ is diagonal. Moreover, because the rank of A is less than n, we know that Σ has at least one 0 along its diagonal.

A.D. The transpose $A^\top$ is also $n \times n$ and non-invertible. The non-invertibility can be seen from $\det A^\top = \det A$, from $(A^\top)^{-1} = (A^{-1})^\top$, etc.

B.A. Let $\vec{x} = \begin{bmatrix} j & a \end{bmatrix}^\top$ be a vector that records the numbers of juvenile and adult females. Let $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ be the values of $\vec{x}$ as the years progress. Then we have $\vec{x}_{k+1} = A\vec{x}_k$, where

$$A = \begin{bmatrix} 0 & 1/2 \\ 1/5 & 9/10 \end{bmatrix}.$$

B.B. To find the eigenvalues of A, we compute

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} -\lambda & 1/2 \\ 1/5 & 9/10 - \lambda \end{bmatrix} \\ &= (-\lambda)(9/10 - \lambda) - 1/10 \\ &= \lambda^2 - \frac{9}{10}\lambda - \frac{1}{10} \\ &= (\lambda - 1)(\lambda + 1/10). \end{aligned}$$

So the eigenvalues are 1 and $-1/10$. As time progresses, the part of the system captured by the second eigenvector will shrink (because it is smaller than 1 in absolute value), and the system will be attracted to the first eigenvector. To find that eigenvector, we solve $(A - I)\vec{x} = \vec{0}$:

$$\begin{bmatrix} -1 & 1/2 & 0 \\ 1/5 & -1/10 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So we find that $j = a/2$. In other words, there are twice as many adult females as juvenile females. In other words, two thirds of females are adult, and one third are juvenile. Because the eigenvalue is 1, the population is not growing or shrinking. It is somehow perfectly in equilibrium with its habitat.

C.A. The three entries of $\vec{x}$ are the probabilities that a random web crawler is at the first, second, or third pages, respectively, at the time in question. (Equivalently, if we imagine a huge number of web crawlers, then the three entries are the fraction of those crawlers that are at the first, second, or third pages at the time in question.) Then $\vec{x}_{k+1} = P\vec{x}_k$, where

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1 \\ 1/2 & 0 & 0 \end{bmatrix}.$$

Because the entries of each $\vec{x}$ sum to 1, and each of the columns of P sum to 1, we have a Markov chain.

C.B. (By the way, P^5 is entirely positive, and P^4 is not entirely positive.) Because some power of P is entirely positive, we know that the Markov chain has a unique steady state, which is a certain eigenvector associated to eigenvalue 1. To find that eigenvector, we solve $(P - I)\vec{x} = \vec{0}$:

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1/2 & -1 & 1 & 0 \\ 1/2 & 0 & -1 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & -1/2 & 1 & 0 \\ 0 & 1/2 & -1 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The third variable is free, so we set it to 1 to obtain an eigenvector

$$\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

The entries of this vector sum to 5. We scale the vector by 1/5, to obtain a non-negative vector that sums to 1 and hence is a valid probability vector:

$$\vec{q} = \begin{bmatrix} 2/5 \\ 2/5 \\ 1/5 \end{bmatrix}.$$

This $\vec{q}$ is the steady state of the Markov chain.

C.C. (I accepted either of the following two answers. The first one is literal, while the second one demonstrates that we have seen this application already.)

If we set a bunch of random web crawlers loose on the Web, then come back at a much later time, we will find that about 40% of them are at the first page, 40% are at the second, and 20% are at the third. If we watch what happens from there, we will see the web crawlers maintaining those percentages even as they move around.

Practically, what we're doing here is the Markov chain interpretation of the PageRank algorithm. According to this algorithm, the first and second pages are each twice as valuable as the third page.

D. Let A be the matrix on the left. We are trying to solve $A\vec{x} = \vec{0}$. In the solution $\vec{x} = \begin{bmatrix} x & y & z \end{bmatrix}^\top$, the matrix on the right shows that x and z are free and $y = -2z$. Setting $x = 1$ and $z = 0$, and then $x = 0$ and $z = 1$, produces two independent solution vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}.$$

These vectors constitute a basis for the null space of A.

E. To compute the SVD by hand, first we compute

$$A^\top A = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}.$$

Then we find the eigenvalues of $A^\top A$:

$$\begin{aligned} \det(A^\top A - \lambda I) &= \det \begin{bmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{bmatrix} \\ &= (6-\lambda)^2 - 1 \\ &= \lambda^2 - 12\lambda + 35 \\ &= (\lambda - 7)(\lambda - 5). \end{aligned}$$

The eigenvalues are $\lambda_1 = 7$ and $\lambda_2 = 5$. To find the eigenvector $\vec{v}_1$ for λ_1, we solve $(A^\top A - 7I)\vec{x} = \vec{0}$. The augmented matrices row reduce like this:

$$\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So we obtain $\vec{v}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$. Similarly, we find that the eigenvector $\vec{v}_2$ for λ_2 is $\vec{v}_2 = \begin{bmatrix} -1 & 1 \end{bmatrix}^\top$. The eigenvectors are orthogonal, as expected. Scaling them to have length 1 and putting them into the columns of a matrix V, we have

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

The Σ matrix is

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{bmatrix}.$$

The first two columns of the U matrix are

$$\vec{u}_1 = \frac{1}{\sqrt{7}} A\vec{v}_1 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{14} \\ 2/\sqrt{14} \\ 1/\sqrt{14} \end{bmatrix}$$

and

$$\vec{u}_2 = \frac{1}{\sqrt{5}} A\vec{v}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{10} \\ 0 \\ 3/\sqrt{10} \end{bmatrix}.$$

(A solution that made it this far was awarded 10 of the 12 points.) To find the third column of U, we need to find a unit vector perpendicular to both $\vec{u}_1$ and $\vec{u}_2$. So we want to solve the linear system $\vec{u}_1 \cdot \vec{x} = 0, \vec{u}_2 \cdot \vec{x} = 0$. In augmented matrices this system is

$$\begin{bmatrix} 3/\sqrt{14} & 2/\sqrt{14} & 1/\sqrt{14} & 0 \\ -1/\sqrt{10} & 0 & 3/\sqrt{10} & 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2/3 & 1/3 & 0 \\ 0 & 2/(3\sqrt{10}) & \sqrt{10}/3 & 0 \end{bmatrix}$$
$$\mapsto \begin{bmatrix} 1 & 2/3 & 1/3 & 0 \\ 0 & 1 & 5 & 0 \end{bmatrix}$$
$$\mapsto \begin{bmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & 5 & 0 \end{bmatrix}.$$

So a solution and a unit solution are

$$\begin{bmatrix} 3 \\ -5 \\ 1 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 3/\sqrt{35} \\ -5/\sqrt{35} \\ 1/\sqrt{35} \end{bmatrix}.$$

So we finally see that

$$U = \begin{bmatrix} 3/\sqrt{14} & -1/\sqrt{10} & 3/\sqrt{35} \\ 2/\sqrt{14} & 0 & -5/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & 1/\sqrt{35} \end{bmatrix}.$$

(And one can check that $A = U\Sigma V^\top$, as expected.)

F.A. The first column of P is, by convention, the eigenvector corresponding to the largest eigenvalue of the sample covariance. (The largest eigenvalue is the same as the greatest eigenvalue, because all eigenvalues of covariance are non-negative.) This eigenvector $\vec{v}$ tells us the direction of the greatest dispersion (spread) in the data set, relative to the sample mean $\vec{m}$. In other words, if we were trying to draw a line through the "middle" of the data cloud, then that line would pass through $\vec{m}$ in the direction of $\vec{v}$.

F.B. To compute a $\vec{w}$ from its corresponding $\vec{x}$, we translate by $-\vec{m}$ and then rotate/flip by $P^\top$, like this: $\vec{w}_{37} = P^\top(\vec{x}_{37} - \vec{m})$.

F.C. To visualize the data set, we plot the first two or three components of the $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n$ vectors. Because these vectors are organized to show the directions of greatest dispersion, key features of the data set, such as multimodality, are often revealed by looking at their first couple of components.