1. We know both the matrix A, and its RREF:

$$A = \begin{bmatrix} 2 & 0 & 6 & 1 \\ 3 & 1 & 11 & 0 \\ -3 & 0 & -9 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(a) A basis for the image of A is $(2,3,-3)$, $(0,1,0)$, and $(1,0,1)$. The reason we know this is that we know that the linear relations among the column vectors of the RREF are the same as the linear relations among the column vectors of the original matrix A. For the RREF, the first, second, and fourth column vectors clearly can be used to write the third (i.e., they span) and they are also clearly linearly independent. We can therefore conclude the same things about the column vectors of A: the first, second, and fourth column vectors span the subspace of $\mathbb{R}^3$ generated by the columns (i.e., the image) and are linearly independent. They therefore form a basis for the image, by definition of basis.

(b) We want to figure out how to write $\vec{b} = (5,5,0)$ as a linear combination of $\vec{v}_1 = (2,3,-3)$, $\vec{v}_2 = (0,1,0)$ and $\vec{v}_3 = (1,0,1)$. We can do this in the usual way by row-reducing the correct matrix:

$$\left[\begin{array}{ccc:c} 2 & 0 & 1 & 5 \\ 3 & 1 & 0 & 5 \\ -3 & 0 & 1 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc:c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Which shows us that $\vec{b} = \vec{v}_1 + 2\vec{v}_2 + 3\vec{v}_3$.

(c) From the RREF, we see that the kernel is spanned by a single vector: $(3,2,-1,0)$. This fits with the "rank-nullity" theorem: the matrix A represents a linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$. From part (a), the image is 3-dimensional, so the kernel must be $4-3=1$ dimensional, and so spanned by a single vector.

(d) If $\vec{x}_1 = (1,2,0,3)$, then we already know that $T(\vec{x}_1) = \vec{b}$.

Where did the zero come from? It's because in part (b) we found out how to write $\vec{b}$ as a sum of the first, second, and third column vectors, and the zero is there simply to tell us to ignore the third column vector.

Now that we know a single vector $\vec{x}_1$ mapping to $\vec{b}$, we know how to find all the others: they're of the form $\vec{x} = \vec{x}_1 + \vec{v}$, with $\vec{v}$ in the kernel of T. From part (c), the kernel is all multiples of $(3,2,-1,0)$, and so the solutions are $\vec{x} = (1,2,0,3) + t(3,2,-1,0)$.

(e) Starting with the augmented matrix and row-reducing, we get

$$\left[\begin{array}{cccc:c} 2 & 0 & 6 & 1 & 5 \\ 3 & 1 & 11 & 0 & 5 \\ -3 & 0 & -9 & 1 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc:c} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 3 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right],$$

From which we see that the solutions are

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ -1 \\ 0 \end{bmatrix}.$$

(f) The answers in parts (d) and (e) are of course the same. This shows what the RREF algorithm for solving linear equations is really doing: from the point of view of linear transformations, it produces a vector $\vec{x}_1$ so that $T(\vec{x}_1) = \vec{b}$, along with all possible elements in the kernel of T. The free parameters in the solution represent all possible linear combinations of the basis for $\ker(T)$, which gives us all the vectors in the kernel.

2. If A is an $n \times m$ matrix, and B and invertible $n \times n$ matrix, then

(a) $\ker(A) = \ker(BA)$. The idea is just to follow a vector $\vec{v}$ through the composition. Suppose that $\vec{v}$ is a vector in the kernel of BA, and let $\vec{w} = A\vec{v}$. Then (since $\vec{v}$ is in the kernel of BA) $B\vec{w} = \vec{0}$. But since B is invertible, the only vector in its kernel *is* the zero vector, so we must have $\vec{w} = \vec{0}$. In other words, $\vec{v}$ is in the kernel of A. So, any vector in the kernel of BA is already in the kernel of A. The other way around is automatic: if $\vec{v}$ is in the kernel of A, then $A\vec{v} = \vec{0}$, so $BA\vec{v} = \vec{0}$ too. This shows that the kernels are equal.

(b) Both A and BA represent linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$. By the rank-nullity theorem, $\dim(\mathrm{im}(A)) + \dim(\ker(A)) = m$, and $\dim(\mathrm{im}(BA)) + \dim(\ker(BA)) = m$. Since $\ker(A) = \ker(BA)$ from part (a), they both have the same dimension. Using that equation above, that means that we must have $\dim(\mathrm{im}(A)) = \dim(\mathrm{im}(BA))$ too.

(c) One description of the rank of a matrix is that it is the dimension of the image the corresponding linear transformation. Since part (b) tells us that the dimensions of the image of the linear transformations given by A and BA are the same, their ranks must be the same too.

3. The list of properties is

(i) A is invertible

(vi) $\ker(A) = \vec{0}$.

(vii) The column vectors of A form a basis of $\mathbb{R}^n$.

(viii) The column vectors of A span $\mathbb{R}^n$

(ix) The column vectors of A are linearly independent.

Let's see why these are the same.

(a) Suppose that $\vec{v}_1, \ldots, \vec{v}_n$ are the column vectors of A, i.e., that

$$A = \left[\vec{v}_1 \,|\, \vec{v}_2 \,|\, \cdots \,|\, \vec{v}_n\right].$$

For any vector $\vec{x} = (x_1, \ldots, x_n)$, $A\vec{x}$ is a linear combination of the column vectors of A,

$$A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n.$$

If $\vec{x}$ is in the kernel of A, then $A\vec{x} = \vec{0}$. In terms of the equation above, that means we've found a linear combination of the column vectors which add up to $\vec{0}$.

But, if the column vectors of A are linearly independent, the only way we could have such a combination is if all the coefficients x_i are zero, i.e., that $\ker(A) = \vec{0}$, so this is why (ix) implies (vi). The argument is reversible: if $\ker(A) = \vec{0}$, then there are no linear combinations of the column vectors which give the zero vector other than the one where all the coefficients are zero, so the column vectors must be linearly independent, so (vi) implies (ix).

(b) The span of the column vectors is the image of A. The rank-nullity theorem applied to A says that $\dim(\mathrm{im}(A)) + \dim(\ker(A)) = n$. Condition (vi) is that $\dim(\ker(A)) = 0$. Condition (viii) is that $\dim(\mathrm{im}(A)) = n$. The equation coming from the rank-nullity equation above shows that (vi) is the same condition as (viii).

(c) By definition, a basis for $\mathbb{R}^n$ should consist of vectors which are linearly indepenent, and which span $\mathbb{R}^n$. Condition (ix) is that the column vectors are linearly independent, and condition (viii) is that they span $\mathbb{R}^n$. Both of them together are therefore equivalent to (vii).

(d) In order for a linear transformation to be invertible, it should be surjective (i.e., everything in $\mathbb{R}^n$ should be "hit"), which means that the image should be all of $\mathbb{R}^n$, i.e., that the column vectors should span, which is exactly (viii). But the linear transformation described by A should also be injective (i.e., no more than one $\vec{x}$ in $\mathbb{R}^n$ should get sent to the same $\vec{b}$), and this is the same as saying that the kernel is just the zero vector, i.e., (vi). Therefore, A being invertible is the same as (vi) and (viii) together.

(e) Yes, we have enough to show that if any one of these properties is true for an $n \times n$ matrix A, then they are all true. It's a little confusing to keep track of what we've established above, and perhaps this diagram will help:

$$\overbrace{\underbrace{\text{(viii)} \overset{\text{(b)}}{\longleftrightarrow} \text{(vi)}}_{\begin{array}{c}\updownarrow \text{(d)} \\ \text{(i)}\end{array}} \overset{\text{(a)}}{\longleftrightarrow} \text{(ix)}}^{\begin{array}{c}\text{(vii)} \\ \updownarrow \text{(c)}\end{array}}$$

The bottom bracket means that (viii) and (vi) are together the same as (i), by part (d). The top bracket is meant to mean that (viii) and (ix) together (but without needing (vi)) are the same as (vii), by part (c).

Here's how to put it all together:

Suppose that (i) is true. Then by part (d), both (vi) and (viii) are true too. But then since (vi) is true, part (a) shows that (ix) is true. But now both (ix) and (viii) are true, so part (c) shows that (vii) is true. That means they're all true.

Or, suppose that (vi) is true. By part (a), this means that (ix) is true too, and by part (b) that (viii) is true. But now (viii) and (vi) together mean that (i) is true by part (d), and finally, (viii) and (ix) together show that (vii) is true, so again they're all true.

If we suppose that (vii) is true, then part (c) shows us that (ix) and (viii) are both true. Applying part (b) to (viii) we see that (vi) is true, and then both of them together show that (i) is true by part (d).

If we suppose that (viii) is true, then part (b) gives us (vi), which then gives us (i) by part (d), then (ix) by part (a), and then finally (vii) by part (c).

Finally, supposing that (ix) is true, then (a) gives us (vi), then (b) gives us (viii), which then gives us (i) by part (d). Since (viii) and (ix) are now both true, part (c) gives us (vii), so they're all true again.

In other words, no matter which one of the properties we start with, we can show that all the others are also true, using the arguments in parts (a) through (d). This is what it means for the properties to be equivalent: if one is true then they all are, and if one is false then they all are too.