Theorem[1]: The Schur decomposition of a square matrix A expresses it in the following form:

$$A = BRB^{-1} = BRB^H,$$

where B is an orthogonal matrix (i.e. its columns are orthonormal vectors) and R is upper triangular.

Preliminaries:
(P1) The fundamental theorem of algebra states that every non-constant single-variable polynomial with complex coefficients has at least one complex root. When applied to the characteristic polynomial coming from a matrix eigenvalue problem, this tell us that any square matrix (or a linear transformation in general) must have at least one complex root, and thus at least one nontrivial eigenvector.

(P2) Matrix corresponding to a linear transformation (LT): Let $T : V \to V$ be a LT. If we choose a the set of vectors $\{b\}_{i=1}^n$ as the basis for V, then the matrix A corresponding to this LT is generated by giving each of the basis vectors as input to T and expressing the output as a linear combination of the basis vectors

$$T([b_1 \dots b_n]) = [b_1 \dots b_n]A. \tag{1}$$

Here, each column of A can be interpreted as the coefficients of a linear combination. Two observations can be made:
(a) changing the basis leads to a *different* matrix for the *same* linear transformation, and
(b) the length of the basis vectors does not enter into the matrix size. For e.g. if we had a k-dimensional subspace of V, say V_k, spanned by the basis vectors $\{c_i\}_{i=1}^k$, then even though c_i is defined by n entries, a linear transformation of the type $T_k : V_k \to V_k$ requires a matrix A_k of size $k \times k$.

Proof:
(1) Consider a $n \times n$ matrix, A. By (P1), it can be said that there is atleast one eigenvalue (call it λ_1) and one nontrivial eigenvector corresponding to it. Let's generalize this and say that λ_1 has geometric multiplicity k, giving us k linear independent eigenvectors for this eigenvalue. Let's take matters further and apply Gram-Schmidt to these vectors to produce an orthonormal set and denote this set: $\{b_i\}_{i=1}^k$. Thus,

$$A[b_1 \dots b_k] = [b_1 \dots b_k] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_1 \end{bmatrix} \tag{2}$$

(2) The above set $\{b_i\}_{i=1}^k$ spanned a k dimensional subspace of $\mathbb{C}^n$, call it V_b. We can now construct a $n-k$ dimensional subspace of $\mathbb{C}^n$, call it V_c that is orthogonal to V_b, such that $V_b + V_c$ spans $\mathbb{C}^n$. As before, Gram-Schmidt can be used to generate this new basis; let's denote it by $\{c_i\}_{i=1}^{n-k}$. We can further add this set to both sides of Eq. (2) to get:

$$A[b_1 \dots b_k \; c_1 \dots c_{n-k}] = [b_1 \dots b_k \; c_1 \dots c_{n-k}] \left[\begin{array}{c|c} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_1 \end{bmatrix}_{k \times k} & (A_r)_{k \times n-k} \\ \hline 0_{n-k \times k} & (A_2)_{n-k \times n-k} \end{array} \right] \tag{3}$$

[1]Uday Khankhoje, Linear Algebra EE5120, July-Nov 2018, Electrical Engineering IIT Madras http://www.ee.iitm.ac.in/uday/2018b-EE5120/

(3) From the above, let's consider the action of A on c_i: the result is a linear combination of the basis sets $\{b_i\}_{i=1}^{k}$ and $\{c_i\}_{i=1}^{n-k}$, explicitly as

$$A\,c_i = [b_1 \dots b_k \; c_1 \dots c_{n-k}] \begin{bmatrix} (A_r)_i \\ (A_2)_i \end{bmatrix} \tag{4}$$

If we were to change the basis from the set $\{c_i\}$ to some new set $\{d_i\}$ (allowed since there are infinite possible bases for a vector space), the only thing that would change would the form of the matrix A_2, while A_r would remain unchanged. In other words, A_2 seems to be controlling/depicting what is happening within this $n-k$ dimensional subspace, V_c. In fact, we can think of A_2 as the matrix representation of a linear transformation of the kind: $T_2 : V_c \to V_c$. In the spirit of Eq. (1), we can say:
$T_2([c_1 \dots c_{n-k}]) = ([c_1 \dots c_{n-k}])A_2$.

(4) By invoking (P1), there must be atleast one eigenvalue (call it λ_2) and nontrivial eigenvector (call it d) for this LT. Using Gram Schmidt, we can take the set $\{c_i\}$ and create a new basis set for V_c as $[d\; c'_1 \dots c'_{n-k-1}]$, where the c'_is are all orthonormal, and along with d, span V_c. How will the matrix corresponding to this LT look now?

$$T_2([d_1\; c'_1 \dots c'_{n-k-1}]) = [d_1\; c'_1 \dots c'_{n-k-1}] \begin{bmatrix} \lambda_2 & A'_{1,1} & \dots & A'_{1,n-k-1} \\ \vdots & \dots & \ddots & \vdots \\ 0 & A'_{n-k,1} & \dots & A'_{n-k,n-k-1} \end{bmatrix} \tag{5}$$

where it is evident that original matrix A_2 has been modified from before.

(5) Having understood this concept, we can generalize and say that the eigenvalue λ_2 has geometric multiplicity l, and therefore an orthonormal eigenbasis set can be created as $\{d_i\}_{i=1}^{l}$. Supplementing this set with an additional $n-k-l$ vectors $\{c'_i\}$ in order to span V_c, we get a new matrix representation of T_2 similar to the above equation, except that the top-left block will be an $l \times l$ diagonal matrix with λ_2 on its diagonals (much like the block diagonal form of the RHS of Eq. (2)).

(6) We can now put all previous observation together and examine the action of A on the updated basis set: $B' = [b_1 \dots b_k\; d_1 \dots d_l\; c'_1 \dots c'_{n-k-1}]$. We get (with $p = n-k-l$):

$$A\,B' = [b_1 \dots b_k\; d_1 \dots d_l\; c'_1 \dots c'_p] \left[\begin{array}{c|c|c} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_1 \end{bmatrix}_{k\times k} & (A_{r1})_{k\times l} & (A_{r2})_{k\times p} \\ \hline 0_{l\times k} & \begin{bmatrix} \lambda_2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_2 \end{bmatrix}_{l\times l} & (A_{r3})_{l\times p} \\ \hline 0_{p\times k} & 0_{p\times l} & (A_3)_{p\times p} \end{array}\right] \tag{6}$$

(7) The above line of reasoning can now be applied to the linear transformation corresponding to A_3 and the basis set $\{c'_i\}_{i=1}^{p}$. Each time the set of basis vectors used are orthonormal and the resulting operations will ensure that the rightmost matrix in the above equation is upper triangular. In other words,

$$AB = BR \quad \implies \quad A = BRB^{-1} = BRB^{H} \tag{7}$$

where we have used the fact that for an orthogonal matrix , $B^{-1} = B^{H}$.
QED