Diffusion and Confusion

Claude Shannon, in one of the fundamental papers on the theoretical foundations of cryptography [“Communication theory of secrecy systems,” *Bell Systems Technical Journal* 28 (1949), 656 – 715], gave two properties that a good cryptosystem should have to hinder statistical analysis: **diffusion** and **confusion**.

Diffusion means that if we change a character of the plaintext, then several characters of the ciphertext should change, and similarly, if we change a character of the ciphertext, then several characters of the plaintext should change. We saw that the Hill cipher has this property. This means that frequency statistics of letters, [digraphs], etc. in the plaintext are diffused over several characters in the ciphertext, which means that much more ciphertext is needed to do a meaningful statistical attack.

Confusion means that the key does not relate in a simple way to the ciphertext. In particular, each character of the ciphertext should depend on several parts of the key. For example, suppose we have a Hill cipher with an $n \times n$ matrix, and suppose we have a plaintext-ciphertext pair of length n^2 with which we are able to solve for the encryption matrix. If we change one character of the ciphertext, one column of the matrix can change completely. Of course, it would be more desirable to have the entire key change. When a situation like that happens, the cryptanalyst would probably need to solve for the entire key simultaneously, rather than piece by piece.

The Vigenère and substitution ciphers do not have the properties of diffusion and confusion, which is why they are so susceptible to frequency analysis.
Introduction to Cryptography and Coding Theory, Wade Trappe and Lawrence C. Washington.