I would like to demonstrate one proof of a result of Sauer, Perles and Shelah, Vapnik and Chervonenkis from 1971,1972. The proof is due to Smolensky and is from 1997. There are a variety of proofs, basic induction works fine.

The first idea involves a vector space of polynomials. We say a polynomial in variables $x_1, x_2, \ldots, x_m$ is multilinear if it has no expressions containing $x_t^i$ for $t \geq 2$. Thus it is linear in each variable. The degree of such a polynomial is given by the usual definition. Define

$$V = \{\text{multilinear polynomials of degree } \leq 2\}$$

Then $\dim(V) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$ since we readily find a basis $\{1, x_1, x_2, \ldots, x_m, x_1x_2, x_1x_3, \ldots, x_{m-1}x_m\}$ of size $\binom{m}{2} + \binom{m}{1} + \binom{m}{0}$.

Our object of study are so called simple matrices whose entries are either 0 or 1 with the additional condition that no column is repeated. Thus if $A$ is an $m \times n$ simple matrix then it is ‘easy’ to see that $n \leq 2^m$ because there are only $2^m$ possible columns of 0’s and 1’s. We wish to impose an additional property on $A$ and obtain a good bound on $n$ as a function of $m$. Let

$$K_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

We want $A$ not to contain $K_3$ as a configuration, namely there is no $3 \times 8$ submatrix of $A$ which is a row and column permutation of $K_3$. The following $5 \times 16$ matrix has the desired property.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

This follows because each triple of rows of the above matrix avoids

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

There are many ways to create a matrix avoiding $K_3$ that are not so symmetric.

**Theorem 1** (Vapnik and Chervonenkis 1971, Sauer 1972, Perles and Shelah 1972)

Let $A$ be an $m \times n$ simple matrix with no configuration $K_3$ (with no $3 \times 8$ submatrix which is a row and column permutation of $K_3$). Then

$$n \leq \binom{m}{2} + \binom{m}{1} + \binom{m}{0}.$$ 

**Proof:** To prove this lovely result we need some polynomials, one per column, that are linearly independent and are multilinear and of degree at most 2. We will only be evaluating these polynomials on the values of columns of $A$. 

For the $i$th column $A_i$ of $A$ we create a polynomial as follows. Let $A_i = (a_1, a_2, \ldots, a_m)^T$. We form a polynomial \( f_i(x) = \prod_{i=1}^{m}(1 - x_1 - a_i) \)

We will only be evaluating this polynomial on columns of $A$ for which $x_i \in \{0,1\}$. As well $a_i \in \{0,1\}$. We note that $(1 - x_i - a_i) = 0$ for $x_i \neq a_i$, $(1 - x_i - a_i) = 1$ for $x_i = a_i = 0$ and $(1 - x_i - a_i) = -1$ for $x_i = a_i = 1$.

Thus $f_i(A_i) \neq 0$ while $f_i(A_j) = 0$ for $j \neq i$. Certainly this means the polynomials are linearly independent. Also they are multilinear but of degree $m$. So we have some work to do!

Now we use the fact that $A$ has no configuration $K_3$. For each triple of rows $i, j, k$ there must be some column of three elements missing say perhaps

\[
\begin{array}{ccc}
  i & c \\
  j & d \\
  k & e \\
\end{array}
\]

We can form a polynomial

\[
f_{ijk}(x) = (1 - x_i - c)(1 - x_j - d)(1 - x_k - e)
\]

which has the property that evaluated at any column $y$ of $A$, $f_{ijk}(y) = 0$. with $f_{ijk} = -x_i x_j x_k + (1 - e) x_i x_j + (1 - d) x_j x_k - (1 - c) (1 - d) x_k - (1 - c) (1 - d) x_j - (1 - e) x_j - (1 - d) (1 - e) x_i + (1 - c) (1 - d) (1 - e)$ we can use the identity

\[
x_i x_j x_k = (1 - e) (x_i x_j) + (1 - d) x_i x_k + (1 - c) x_j x_k - (1 - c) (1 - d) x_k
\]

\[-(1 - c) (1 - e) x_j - (1 - d) (1 - e) x_i + (1 - c) (1 - d) (1 - e),
\]

at least when evaluated on the columns of $A$. 

The right hand side has terms of degree at most 2. We use such an identity on our polynomials $f_i$ taking any term that contains the product $x_i x_j x_k$ and replacing $x_i x_j x_k$ by $(1 - c) (x_i x_j + (1 - d) x_i x_k + (1 - c) x_j x_k - (1 - c) (1 - d) x_k - (1 - c) (1 - d) x_j - (1 - e) x_j - (1 - d) (1 - e) x_i + (1 - c) (1 - d) (1 - e)$.

Repeat over and over again to get a polynomial $f'_i(x)$ that agrees with $f_i(x)$ on columns of $A$ and has degree at most 2. The $n$ polynomials $f'_i(x)$ are linearly independent and all in $V$ with $\dim(V) = {m \choose 2} + {m \choose 1} + {m \choose 0}$ and so

\[
n \leq \binom{m}{2} + \binom{m}{1} + \binom{m}{0}.
\]