1. [Given nonsingular matrix A of size $n \times n$]

i) A has an inverse
ii) $\det(A) \neq 0$
iii) $\text{rank}(A) = n$
iv) For any vector $z \neq 0_n$, $Az \neq 0_n$

Show equivalence:

An invertible matrix is defined as being square and having a determinant that is not equal to 0. As such, we know that i) and ii) are equivalent. In addition, iii) is equivalent due to the fact that for a determinant to not be equal to 0 ii), there has to be nonzero rows in the matrix. As such, the rank has to be equal to the number of rows/columns in the matrix. Similarly, iv) is equivalent because the vector z cannot have a nonzero component, and that A is invertible i) due to being nonsingular, there are no zero components that are in the 0_n subspace, making it essentially equal to statements ii) and iii).

Note: Since A annihilates no nontrivial vector, the columns of A form a basis for an n-dimensional space, implying i) of being an invertible matrix.

Note: If the $\det(A)$ is taken with a row of zeros in A, the determinant is 0. Thus, the reduced echelon form of A needs to have pivots in every nonzero row, and thus would have a rank of n.

2. Prove that under the rules of the question, it's impossible to form more groups than n citizens. (use matrix properties)

In this case, we would consider defining matrix A of $m \times n$, with m groups and n citizens with $A_{ij} = 1$ if $j \in G_i$ and 0 o.w. We can set constraints that each row of A has an odd sum when added up, and that the dot product of any of the groups or rows, must be even. We would then use the pigeonhole principle to know that there must be 2 rows that are the same. When we take the dot product of these rows, it will be odd due to the fact that the number of members is odd when summing up in the row. This contradicts that the dot product is even, so it isn't possible to form more groups than n.

3. Assuming that A is an $m \times n$ matrix has $\text{rank}(A) = n$ where $m > n$. $[P = A(A^TA)^{-1}A^T]$

$$= \frac{1}{A^T} \cdot A^T$$
$$= 1$$

1.
$$P^T = (A(A^TA)^{-1}A^T)^T$$
$$= (A^T \cdot ((A^TA)^{-1})^T \cdot (A^T)^T)$$
$$= (A^T \cdot (\frac{1}{\cancel{A^T}A} \cdot \cancel{(A^T)})^T)$$
$$= (A^T \cdot \frac{1}{A^T})$$
$$= 1$$

$$P^2 = A(A^TA)^{-1}A^T \cdot A(A^TA)^{-1}A^T$$
$$= (\cancel{A}(\frac{1}{\cancel{A^TA}})\cancel{A^T})^2$$
$$= 1^2 \text{ or } 1$$

Both are equal to 1, which is equal to P.

2. Showing that $\|Pb\| \leq \|b\|$ for any vector p-norm:

We know that P is the orthogonal projector onto $\text{span}(A)$, with $b = Pb + P_\perp b$.
Thus, given $\|b\| \geq 0$ from the definition of the p-norm, $\|Pb\| \leq \|Pb + P_\perp b\|$.
$$\leq \|Pb + r\|.$$
$$\leq \|b\|.$$

In the context of linear-least squares, the projection of vector b onto the column space or rank, of A is a closer approximation to the solution of b. Pb is the best fit of b onto the column space of A.

3. $Ax = Pb$ is all about orthogonality:

$$Ax = (A(A^TA)^{-1}A^T)(y+r)$$
$$= \underbrace{(1)}(Ax + \underbrace{(b-Ax)})$$
$$= 1)$$

Geometrically, the vector x which is the solution to linear least-squares is the orthogonal projection of b onto the span (A). We find the point in the subspace closest to b via the Euclidean distance. P maps b onto that point.

4. https://colab.research.google.com/drive/1Zee2rwQJm-WrOafngu2eZgUQGLwFKFP7?usp=sharing