# Pigeonholes and Two-way Counting[1]

*Anne Penfold Street*

*Anne Penfold Street is Professor of Mathematics at the University of Queensland in Brisbane, Australia. She gained her Ph D at the University of Illinois.*
*She has been on the Problems Committee of the Australian Mathematics Competition, the Australian Mathematical Olympiad Committee and the Board of the Australian Mathematics Trust. She has been involved with enrichment programs for secondary students. Her research interests are mostly in combinatorics.*

In this paper, we look at two techniques which are useful in solving problems. Although the underlying idea is very simple in both cases, they often allow us to solve quite complicated looking problems; see also [1],[11].

First of all, we state the **Pigeonhole Principle** in several of its different forms.

Suppose that $N$ objects are placed in $k$ pigeonholes. Then:

- if $N > k$, some pigeonhole contains more than one object;

[1]This paper was presented as a keynote address at WFNMC Congress-4, Melbourne August 2002

- if $N > mk$, some pigeonhole contains more than $m$ objects.

If the average number of objects per pigeonhole is $a$, then:

- some pigeonhole contains at least $a$ objects;
- some pigeonhole contains at most $a$ objects.

Next, we consider a few examples where this idea is applied.

Example 1: **Birthdays**
*Among* 367 *people, at least two share a birthday.*

In this case, the pigeonholes are the 366 possible dates in a year, where we are allowing for a leap year.

Note here what the pigeonhole principle *doesn't* tell us: we have no idea which two people share a birthday, nor what day the shared birthday might be, nor whether several days of the year are shared birthdays among these particular people, nor even whether all 367 of them were born on Leap Day.

Example 2: **Choosing numbers**
*Suppose we choose some* 19 *of the* 34 *numbers*

```
1, 4, 7, 10, 13, 16, ..., 97, 100.
```

*Then, among our chosen numbers, there are two which sum to* 104.

To see this, look at the following table.

The number 103 does not belong to our given set, so the number 1 is not part of a pair that sum to 104. Next we have 16 pairs of numbers such that each pair sums to 104. Finally we have the number 52, again not

```
        1
  4          100
  7           97
 10           94
 13           91
 16           88
 19           85
 22           82
 25           79
 28           76
 31           73
 34           70
 37           67
 40           64
 43           61
 46           58
 49           55
       52
```

one of a pair in this set, summing to 104. These give us 18 pigeonholes: two single numbers and 16 pairs.

Thus we could choose one number from each pigeonhole, a total of 18 numbers, without having any two that sum to 104. But to choose 19 numbers from this set, we must take both numbers from at least one of the 16 pairs. See [3].

Example 3: **Aspirins**

*A man takes at least one aspirin a day for* 30 *days. If he takes* 45 *aspirins altogether, then in some sequence of consecutive days, he takes exactly* 14 *aspirins.*

Let $a_i$ be the number of aspirins he takes in the first $i$ days. Since he takes at least one aspirin a day, we know that $a_i < a_{i+1}$ for $i = 1, \cdots, 29$.

Thus

$$1 \le a_1 < a_2 < a_3 < \cdots < a_{30} = 45.$$

When does he take exactly 14 aspirins? We also know that

$$15 \le a_1 + 14 < a_2 + 14 < \cdots < a_{30} + 14 = 59.$$

Then the 60 numbers $\{a_i | i = 1, \cdots, 30\}$ and $\{a_i + 14 | i = 1, \cdots, 30\}$ take at most 59 distinct values. By the pigeonhole principle, some two of these numbers must be equal.
Since the inequalities given above are all strict, we cannot have $a_i = a_j$ for distinct values of $i$ and $j$, so for some $i$ and $j$, we have $a_i = a_j + 14$, and he takes exactly 14 aspirins on days $j + 1, \cdots, i$.

Example 4: **Decimal representation**
*In the decimal representation of the number* $n = 5 \times 7^{34}$, *some digit must occur at least four times.*

By the Pigeonhole Principle, every integer with more than 30 digits must have at least one of the 10 decimal digits occurring at least four times in its decimal representation.

But $\log n = \log 5 + 34.\log 7 = 29.4324 < 30$. So $n$ has 30 digits.
If none of the 10 decimal digits occurs more than three times, then each of them must occur exactly three times. Then the sum of the digits must be divisible by 3, so that $n$ itself would be divisible by 3.
This is not the case, since the only primes dividing $n$ are 5 and 7. So some digit occurs at least four times. See [6].

Example 5: **Seven real numbers**
*Among any seven real numbers* $y_1, y_2, \ldots, y_7$, *there are two, say* $y_i$ *and* $y_j$, *such that*

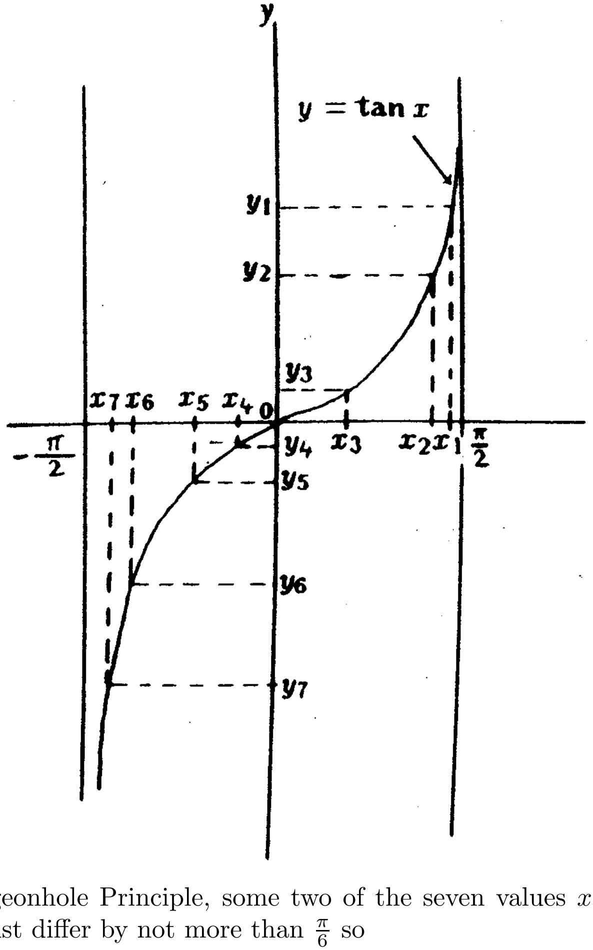
$$0 \le \frac{y_i - y_j}{1 + y_i y_j} \le \frac{1}{\sqrt{3}}.$$

This expression brings to mind the formula

$$\tan(x_i - x_j) = \frac{\tan x_i - \tan x_j}{1 + \tan x_i . \tan x_j},$$

especially since $\tan 0 = 0$ and $\tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}$.

For $n = 1, 2, \cdots, 7$, let's try $y_n = \tan x_n$, mapping the seven real numbers to seven images in the range between $-\frac{\pi}{2}$ and $+\frac{\pi}{2}$, as shown in the diagram.




By the Pigeonhole Principle, some two of the seven values $x_i$, for $i = 1, \cdots, 7$ must differ by not more than $\frac{\pi}{6}$ so

$$0 \leq x_i - x_j \leq \frac{\pi}{6}.$$

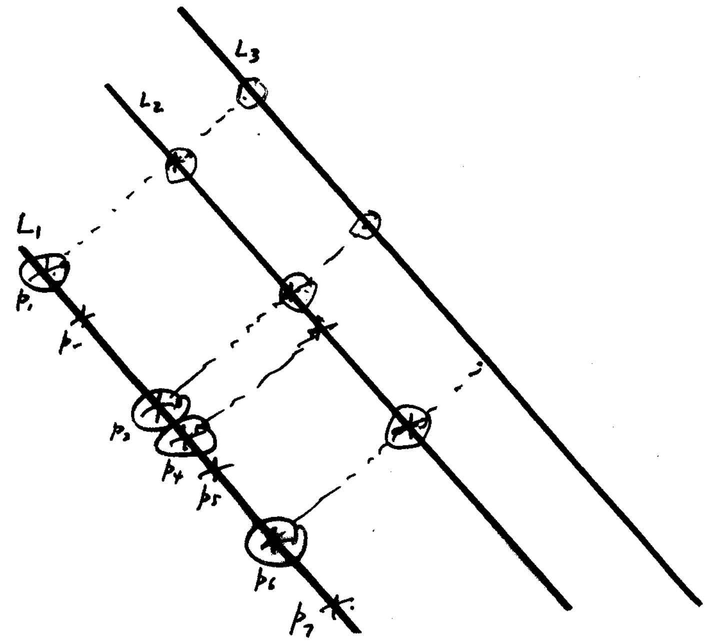
In this range, the tangent is strictly increasing, so we have exactly what we need, namely,

$$\tan 0 \leq \tan(x_i - x_j) \leq \tan\frac{\pi}{6}.$$

See [8].

Example 6: **Coloring the plane**
*If we colour the plane in just two colours, Red and Blue, then somewhere in the plane, there must be a rectangle with all four of its corners the same colour.*

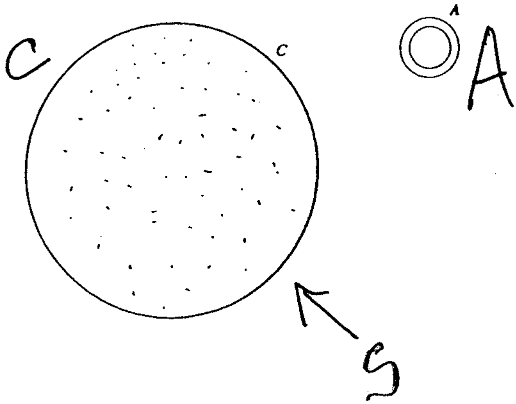
To see this, choose any three parallel lines, $L_1, L_2, L_3$, in the plane, and any seven points on one of these three lines, say, $p_1, \cdots, p_7$ on line $L_1$.. Colour each of these seven points either red or blue. By the Pigeonhole Principle, at least four of the points must be the same colour, say, $p_1, p_2, p_3, p_4$ are red.




Now drop perpendiculars from these four red points to the line $L_2$, and consider the points $q_1, q_2, q_3, q_4$ where the perpendiculars from $L_1$ intersect $L_2$. If we colour any two of these four points red, say, $q_i$ and $q_j$, then we have a rectangle with all its four corners red, namely, $p_i, q_i, q_j, p_j$. To avoid this happening, we must colour at least three of the chosen points on the line $L_2$ blue, say, $q_1, q_2, q_3$ are blue.

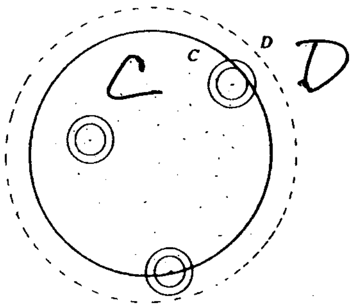
Finally we drop perpendiculars from these points to the line $L_3$, and consider the points $r_1, r_2, r_3$ where the perpendiculars from $L_2$ (and hence also from $L_1$) meet $L_3$. If two of these points, say, $r_i, r_j$, are colored blue, then we have a rectangle $q_i, r_i, r_j, q_j$ with all four of its corners blue. If at most one point is colored blue, then two points, say, $r_m, r_n$, are colored red, and we have a rectangle $p_m.r_m, r_n, p_n$, with all four of its corners red.

Example 7: **Circle and Annulus**
*Suppose that $C$ is a circle of radius 16 units, and that $A$ is an annulus with inner and outer radii 2 and 3 units respectively. Let a set $S$ of 650 points be chosen inside $C$. Then, no matter how these points are scattered over the circle $C$, the annulus $A$ can be placed so that it covers at least 10 of the points of $S$.*




Now suppose that a copy of $A$ is centered at each of the 650 points of $S$.

At a point near the edge, $A$ will stick out past the circumference of $C$. But since the centre of $A$ is inside $C$, a circle $D$, concentric with $C$ and of radius 19 units, will contain all 650 copies of $A$ in its interior.



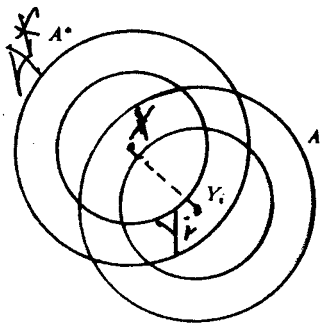

The area of $A$ is $\pi \times 3^2 - \pi \times 2^2 = 5\pi$. Hence 650 copies of $A$ must blanket $D$ with a total coverage of

$$650 \times 5\pi = 3250\pi.$$

If each point of $D$ is covered by at most 9 copies of $A$, then the total area covering $D$ is at most 9 times the area of $D$, that is,

$$9(\pi \times 19^2) = 9(361\pi) = 3249\pi.$$

Thus there must be some point $X$ of $D$ on which at least 10 copies of $A$ are piled up. If $Y_i$ is the centre of an annulus that covers such a point $X$, then the distance $XY_i$ must be between 2 and 3.

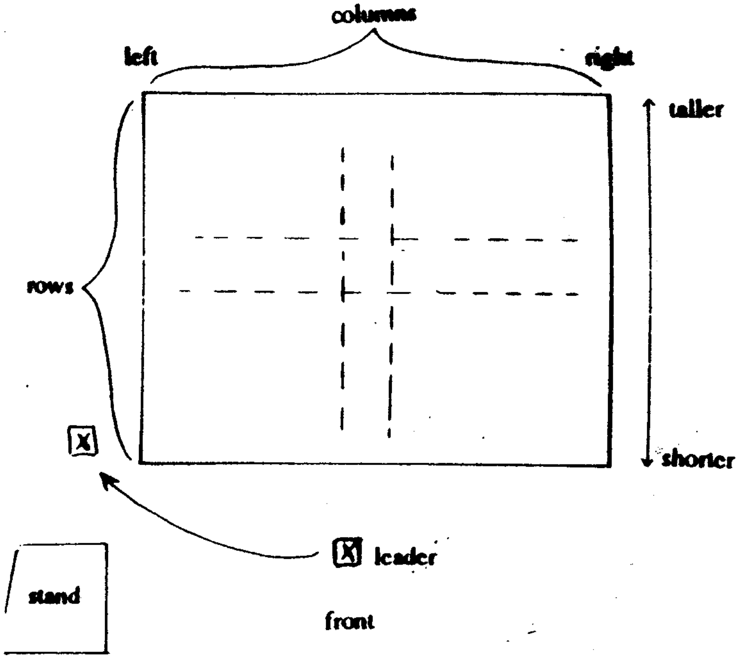


Now we can turn things around and centre a copy $A^*$ of $A$ at $X$ instead of at $Y_i$. Then $A^*$ covers $Y_i$.

Since at least 10 copies of $A$ cover $X$, the special annulus $A^*$ centered at $X$ covers their 10 (or more) centres $Y_1, Y_2, \cdots, Y_{10}$, each of which belongs to the set $S$. See [8].

Example 8: **Marching Band**

*When the leader of a marching band faced his musicians, he saw that some of the shorter people were hidden in the pack behind taller players. Keeping the columns intact, he brought the shorter ones forward till the people in each column stood in nondecreasing order of height from front to back.*

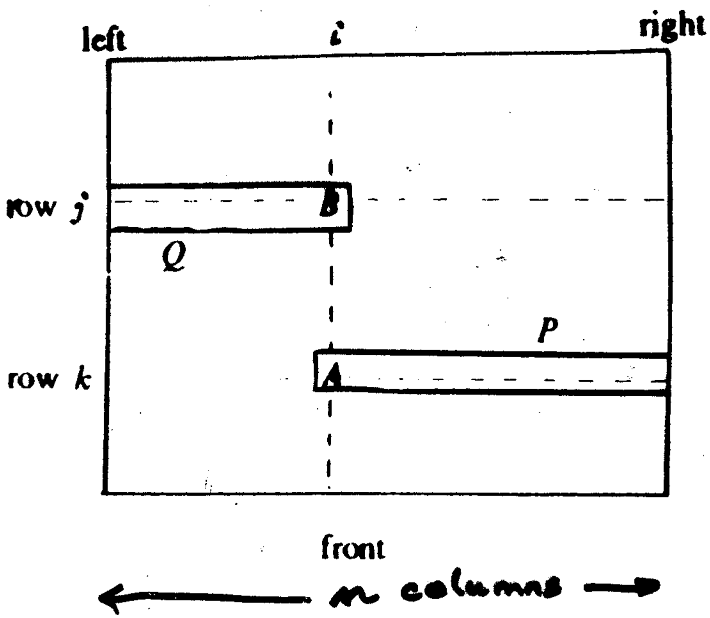



*Later, they were to salute the dignitaries in a reviewing stand which they*

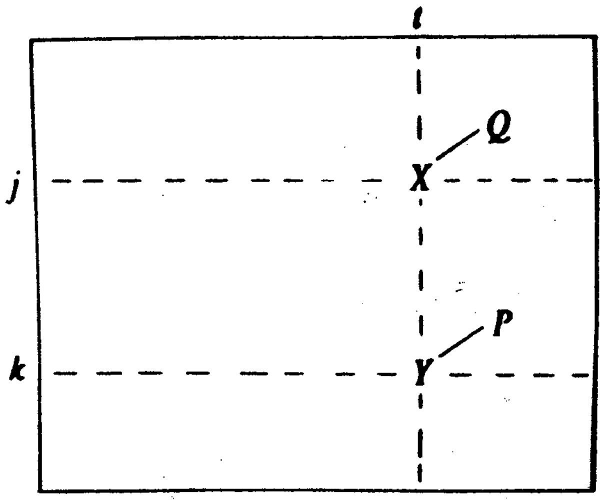
*would pass on their right, so the bandmaster went around to see how they looked from the side. He found that some of the shorter players were again blocked from view.*
*To correct this, he did to the rows what he had just done to the columns: keeping the rows intact, he arranged the players within each row in nondecreasing order of height from left to right (that is, from* his *left to right as he faced the troupe).*
*In fact, he had no need to worry that this shuffling about within the rows would foul up his carefully ordered columns.*

For suppose that after both rearrangements, we find, in column $i$, person $A$ who is both taller and closer to the front than person $B$. We know that in row $j$, everyone in segment $Q$ is no taller than person $B$, and in row $k$, everyone in segment $P$ is no shorter than person $A$. Since $A$ is taller than $B$, everyone in $P$ is taller than everyone in $Q$.

*It is important to note here that the total number of people in $P$ and $Q$ is $n+1$.*



Now we reverse the rearrangement of the rows, so that we are back to the stage where the columns were ordered but the rows still needed ordering.
In row $j$ the elements of $Q$ are put into their former places in the row, and in row $k$ the elements of $P$ are sorted into theirs. Since we have a total of $n+1$ people to replace in their original columns, the Pigeonhole Principle tells us that two of them must finish up in the same column, say, column $\ell$.




At this intermediate stage, the columns are properly sorted, with $X$ at least as tall as $Y$. But $X \in Q$ and $Y \in P$ so $Y$ must be taller than $X$.

Now we have a contradiction: it arose from assuming that the row rearrangement must have disturbed the column ordering. So, in fact, the row rearrangement caused no disturbance of the column ordering.

This example parallels one that arises in considering the properties of Young Tableaux; see [9].

Now we look at our second basic technique, namely, **Two-way Counting**, which simply says that counting the number of elements of a set in two different ways gives the same result, a fact which is often used in combination with the Pigeonhole Principle. A convenient statement of this idea is the following:
*Summing the entries in a matrix by rows or by columns gives the same result.*
(However, at the time this was written, most references on the web to two-way counting concerned Enron, Arthur Anderson, and the fact that counting the same thing in several different ways gives as many different results as convenient.)

Example 9: **Sums of numbers**
*If the nine non-negative real numbers*

$$a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9$$

*sum to* 90, *then there must be four of them with sum at least* 40.

We write the nine numbers four times, arranged as shown in the table. Thus each row of the table sums to 90, and the whole table must sum to 360. But now each of the nine columns must sum to at least $360/9 = 40$.

$$\begin{array}{ccccccccc}
a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 \\
a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_1 \\
a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_1 & a_2 \\
a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_1 & a_2 & a_3
\end{array}$$

Example 10: **Divisors of positive integers**
For every positive integer $n$, let $d(n)$ be the number of positive integers that divide $n$. For instance:

$$d(15) = 4, \text{ with divisors } 1,\ 3,\ 5,\ 15;$$
$$d(16) = 5, \text{ with divisors } 1,\ 2,\ 4,\ 8,\ 16;$$

$$d(17) = 2, \text{ with divisors } 1,\ 17;$$
$$d(18) = 6, \text{ with divisors } 1,\ 2,\ 3,\ 6,\ 9,\ 18.$$

Again $d(41) = 2, d(42) = 8, d(43) = 2$.
The function $d(n)$ keeps fluctuating, but we can use two-way counting to get an idea of its average behaviour, namely:
*The average number of divisors of the first $n$ positive integers,*

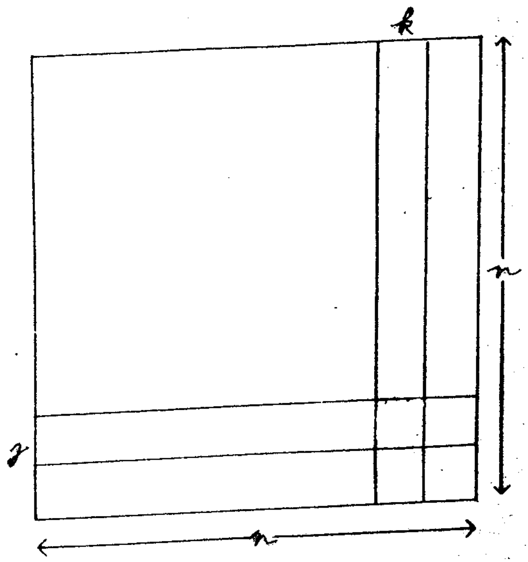
$$\frac{d(1) + d(2) + d(3) + \cdots + d(n)}{n}$$

*is approximately* $\log_e n$.

We think of $d(k)$ as the number of pairs $(j, k)$ such that $j|k$. In an $n \times n$ array, suppose that

$$(j,k) = \begin{cases} 1 & \text{if } j|k, \\ 0 & \text{otherwise.} \end{cases}$$

Then the sum $d(1) + d(2) + \cdots + d(n)$ is the sum of the entries in the array, counted column by column.



Now we sum by rows instead. How many 1s are there in row $j$ ? We want the number of multiples of $j$ between 1 and $n$. This is approximately $n/j$ (or, more precisely, between $n/j$ and $(n/j)-1$). So the sum of the entries in the array is roughly

$$\frac{n}{1}+\frac{n}{2}+\cdots+\frac{n}{n}=n(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}).$$

But now

$$\frac{d(1)+d(2)+\cdots+d(n)}{n}$$

is roughly

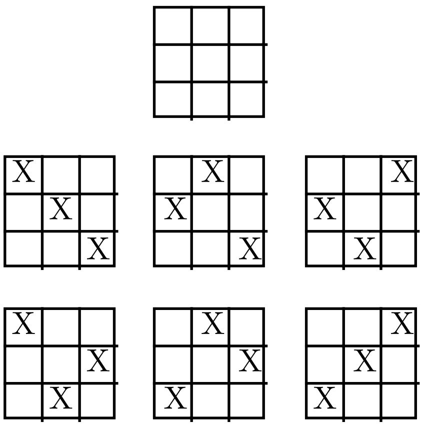
$$\frac{n\times\log_e n}{n}=\log_e n$$

since the sum by rows equals the sum by columns. See [14].

Our next example needs the idea of a **transversal**, that is, a set of $n$ cells, one in each row and one in each column, in an $n\times n$ square. In such a square there are

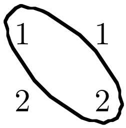
$$n\times(n-1)\times(n-2)\times\cdots\times 1=n!$$

transversals. The diagram shows the possible transversals in a $3\times 3$ square.



Example 11: **Transversals with distinct numbers**
*Let $n \geq 4$ be even. In any $n \times n$ square in which each of $n^2/2$ numbers appears twice, there is a transversal without duplication.*

Certainly $n = 2$ is an exception.




| | |
|---|---|
| 1 | 2 |
| 2 | 1 |

| | | | |
|---|---|---|---|
| 1 | 7 | 6 | 3 |
| 1 | 2 | 4 | 6 |
| 5 | 8 | 3 | 5 |
| 7 | 8 | 4 | 2 |

as shown in the top right. The $4 \times 4$ that shows a transversal without duplication. The main diagonal has a duplication but the transversal with cells circled does not.

*Note:* if the two cells occupied by a given number are always in the same row or the same column, then no transversal has any duplication. Trouble only starts when a pair of cells occupied by the same number are not in the same row or the same column.
We call a pair of cells in different rows and different columns, but containing the same number, a **singular pair**.

- *Any singular pair is contained in precisely* $(n-2)!$

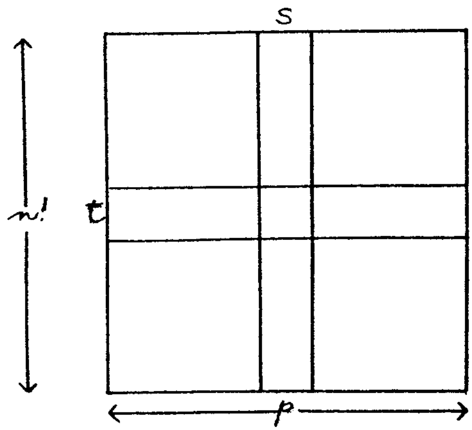
*transversals.* For if we choose a singular pair, then we can choose the next element in $n-2$, the next in $n-1$ ways, and so on.

- *Every transversal contains at least $n/2$ distinct numbers,* since no number occurs more than twice in the square.
- *If there are $p$ singular pairs, then* $0 < p < n^2/2$.

Suppose a given $n \times n$ square has $p$ singular pairs. We take an array with $n!$ rows, one for each transversal, and $p$ columns, one for each singular pair. If $t$ is a transversal and $s$ a singular pair, then:

$$(t, s) = \begin{cases} 1 & \text{if } t \text{ contains } s, \\ 0 & \text{otherwise} \end{cases}$$

and we sum the entries in two ways.




Summing by columns, the total is

$$(p) \times (n-2)!$$

and summing by rows, the total is

$$(n!) \times \mathcal{S}$$

where $\mathcal{S}$ is the average number of singular pairs per transversal. Then

$$\mathcal{S} = \frac{p(n-2)!}{n!}$$

and, since $p \leq n^2/2$,

$$\mathcal{S} \leq \frac{n^2}{2} \times \frac{(n-2)!}{n!} = \frac{n}{2(n-1)}.$$

But $n \geq 4$ which means that

$$\mathcal{S} < 1$$

and there must be a transversal with less than one singular pair, that is, with no duplication.

This kind of argument, combining the Pigeonhole Principle with Two-way Counting, is sometimes referred to as 'existence by averaging' [14], [15], [10], and is the source of many interesting problems. We look briefly at examples involving a *latin square of order* $n$, that is, an $n \times n$ square based on the set $\{1, 2, \ldots, n\}$, in which each element appears precisely once in each row and in each column. We use the fact that

$$(1 - \frac{1}{q})^q \to \frac{1}{e} \approx 0.37.$$

An argument similar to that of Example 11 shows that the average number of distinct symbols in the transversals of a latin square of order $n$ must be

$$n(1 - \frac{1}{2!} + \frac{1}{3!} - \ldots \pm \frac{1}{n!}),$$

so there is a transversal with at least $(1 - \frac{1}{e})n \approx 0.63n$ elements.

Now suppose that $qm = n$. Then in a latin square of order $n$, there exist $q$ rows such that the union of the sets of integers in the first $m$ columns of these $q$ rows contains at least

$$n[1 - (1 - \frac{1}{q})^q]$$

distinct elements. For instance, if $m = 2$ and $q = 3$, then in a latin square of order 6, there are 3 rows such that the first 2 columns of these rows contain at least

$$6[1 - (1 - \frac{1}{3})^3] = 38/9$$

distinct elements. In other words, they contain at least 5 distinct elements; see [15].

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 3 | 1 | 5 | 6 | 4 |
| 3 | 1 | 2 | 6 | 4 | 5 |
| 4 | 6 | 5 | 2 | 1 | 3 |
| 5 | 4 | 6 | 3 | 2 | 1 |
| 6 | 5 | 4 | 1 | 3 | 2 |

In the $6 \times 6$ latin square, rows 1, 2 and 4 contain all the elements except 5 in columns 1 and 2.

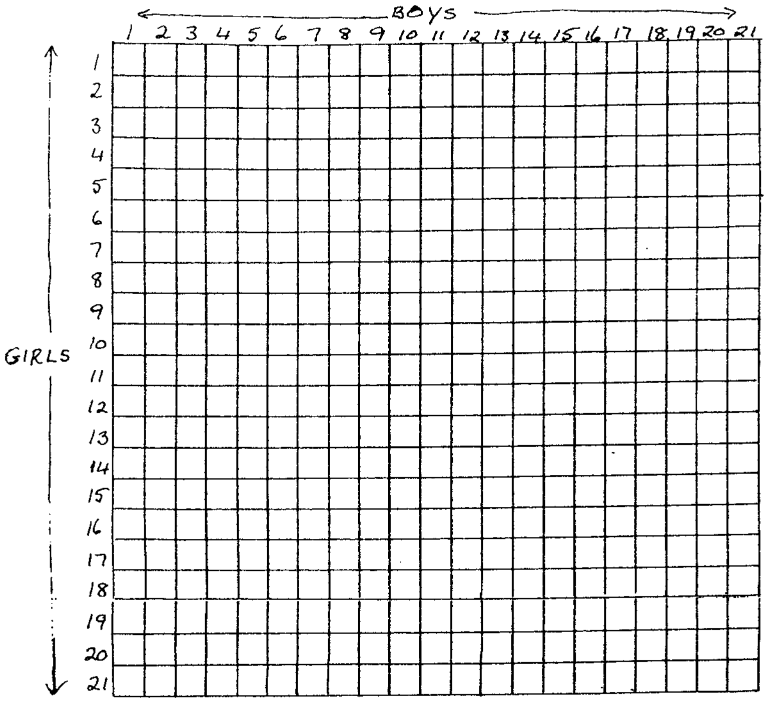
Even more interesting is the *Ryser conjecture:*
In a latin square of order $n$, there is a transversal with $n$ elements if $n$ is odd, and a transversal with $n-1$ elements, if $n$ is even. See also [2],[12], [13], [16].

It seems appropriate to conclude a collection of problems from mathematics competitions with a competition problem concerning a mathematics competition, namely, one from the 2001 IMO. The solution given here was due to Reid Barton, a gold medallist ([4]).

Example 12: **Mathematics Competition**
21 *girls and* 21 *boys took part in a mathematical contest. Each contestant solved at most six problems. For each girl and each boy (that is, for each boy–girl pairing) at least one problem was solved by both of them.*

*Then there must be a problem that was solved by at least three girls and at least three boys.*

To see this, take a $21 \times 21$ array, with one row for each girl and one column for each boy. In each cell of the array, place a letter representing a problem that was solved by the girl in the corresponding row and the boy in the corresponding column. Then every cell of the array will be filled.




No row can contain more than six different letters, so each row contains repeated letters. Look for letters that appear at least three times in a row, and colour every cell of the row containing those letters red.
How many cells in each row are colored red?
Each cell *not* colored red is filled with a letter that occurs at most twice

in the row. At most five letters can occur at most twice, so at most 10 cells are not colored red. Thus the number of red cells in each row is at least 11, so
**more than half the cells in each row are red.**

A similar argument works for the boys, where we colour cells in each column blue to represent problems solved by at least three girls. So
**more than half the cells in each column are blue.**

Since more than half the cells are red and more than half are blue, there must be at least one cell colored *both* red and blue.
The letter in this cell represents a problem that was solved by at least three girls and at least three boys.

## References


[1] Alexander Bogomolny,
http://www.cut-the-knot.com/pigeonhole/aspirin.shtml 1996-2002.

[2] A E Brouwer, A J de Vries and R M A Wieringa, A lower bound for the length of partial transversals in a Latin square, *Nieuw Arch. Wisk.* (3) **24** (1978), 330-332.

[3] Michael Evans and Bruce Henry, editors. *Euler student problems*, 2001 Enrichment Stage, Mathematics Challenge for Young Australians.

[4] Constance Holden (editor), Random Samples, *Science* vol 292, 3 August 2001 page 79. Recipe for Success: Reid Barton's solution of 2001 IMO problem (*Science,* 27 July 2001, page 597).

[5] Ross Honsberger (editor), *Mathematical Plums*, Dolciani Mathematical Expositions volume 4, 1979, The Mathematical Association of America (Incorporated), ISBN 0-88385-304-3.

[6] Ross Honsberger, *Mathematical Chestnuts from around the World* number 23, 32 Miscellaneous Problems, pages 186-187.

[7] Ross Honsberger, *More Mathematical Morsels* Morsel 17, The Circle and the Annulus, pages 70-72.

[8] Ross Honsberger, *More Mathematical Morsels*, pages 198–199 . Olympiad Corners–1984, Canadian Olympiad, 1984 (page 1810).

[9] Ross Honsberger, *More Mathematical Morsels*, Olympiad Corners–1980, Olympiad Corner 17, Canadian Olympiad 1980 (page 209), pages 45-48 in MMM. Also in his guest appearance in Martin Gardner's 'Mathematical games' column in *Scientific American,* August 1980.

[10] C C Lindner and N C Perry, Transversals of Latin Squares, 1970, Auburn University, mimeographed.

[11] A W Plank and N H Williams, editors. *Mathematical Toolchest*, 1996. Australian Mathematics Trust Enrichment Series ISSN 1326-0170, Mathematical Toolchest ISBN 0-85889-292-8.

[12] H J Ryser, Neure Probleme der Kombinatorik in Vorträge über Kombinatorik, Oberwolfach 24-29 Juli, 1967. *Mathematischen Forschgsinstitut*, Oberwolfach, 1968. existence of transversal — conjecture

[13] P W Shor, A lower bound for the length of a partial transversal in a Latin square, *Journal of Combinatorial Theory, Series A* **33**, 1-8 (1982).

[14] S K Stein, Existence out of Chaos, Chapter 4, pages 62-93, in *Mathematical Plums*, edited Ross Honsberger. Divisors of positive integers and Existence of transversals with no duplications.

[15] S K Stein, Transversals of Latin Squares and Their Generalizations, *Pacific Journal of Mathematics,* vol 59, (1975) 567 - 575. As above + CCL/NCP

[16] David E Woolbright, An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols, *Journal of Combinatorial Theory, Series A* **34**, 235-237 (1978).

*Anne Penfold Street*
*Centre for Discrete Mathematics and Computing*
*Department of Mathematics*
*The University of Queensland*
*AUSTRALIA*
*email: aps@maths.uq.edu.au*

* * *