The Optimal Number of People for Indoor and Outdoor Physical Distancing using Pigeonhole Principle

1) Jeonghoo Hyun, Torrey Pines Highschool, San Diego, The United States, jayh80961@gmail.com

Keywords: Physical distancing, Pigeonhole principle, Simulation, Minimum distance, Geometric shapes

1. Introduction

Maintaining safe physical distances between individuals is a fundamental consideration in both public health and event management. While physical distancing gained global attention during the COVID-19 pandemic as a key strategy to reduce the spread of infectious diseases, the importance of spacing is not limited to pandemics. For example, in crowded concerts or music festivals, ensuring a minimum distance between attendees can reduce the risk of trampling, panic-related injuries, and logistical hazards. Similarly, in emergency evacuation planning or fire safety scenarios, appropriate distancing is crucial to prevent congestion and facilitate swift movement to exits. Whether to reduce viral transmission, prevent crowd-related injuries, or ensure general safety, determining the optimal number of people that can occupy a given space while maintaining safe distances is a problem with broad and lasting relevance.

The pigeonhole principle is a fundamental concept in combinatorial mathematics that states that if you distribute more objects into fewer containers, at least one container must contain more than one object. This principle, also known as the Dirichlet box principle or the drawer principle, has its roots in ancient mathematics but was formally recognized and named after Peter Gustav Lejeune Dirichlet in 1834 (Miller et al., 2021). Early in the 17th century, Jean Leurechon indirectly referenced the same principle as well. (Leurechon, 1622; Mersenne, 1625; Rittaud & Heeffer. 2014; Jacquet & Baratgin, 2023).

The history of the pigeonhole principle can be traced back to ancient civilizations, where it was implicitly understood and used in various contexts. For example, an Indian text, the Mahabharata, describes a situation where five arrows are shot at a target, and each arrow ends up in a different part of the target (Murthy, 2003). This example demonstrates that if there are more arrows than distinct target regions, at least two arrows must land in the same region. The principle's underlying idea is intuitive and can be observed in everyday situations, such as assigning students to seats in a classroom or organizing items in drawers. However, its formalization and mathematical proof came much later. The pigeonhole principle was formally named and popularized by the German mathematician Peter Gustav Lejeune Dirichlet in the early 19th century. Dirichlet used the principle in number theory and analysis to prove the existence of solutions to various mathematical problems. His work laid the foundation for the development of the pigeonhole principle as a powerful tool in combinatorics. Since then, the pigeonhole principle has found numerous applications in various branches of mathematics, computer science, and other fields. It is frequently used to prove the existence of patterns, repetitions, or constraints in different settings. The principle has been applied in areas such as graph theory, cryptography, data analysis, scheduling problems, and more.

In this research paper, I will introduce the pigeonhole principle through several examples to highlight its utilization and showcase its practical application across different scenarios, including a proposed research problem related to physical distancing. There will be two main methods to solve the research problem. First, I will showcase the theoretical approach to calculate the critical distance in different segmentations. Secondly, I will conduct a computational approach to simulate the random plotting process in a simulation using Java.

2. Theory

2.1. Pigeonhole principle

The pigeonhole principle can be explained using the analogy of pigeons and pigeonholes. If there are N pigeons and M pigeonholes, and N is greater than M, then at least one pigeonhole must contain more than one pigeon. For instance, consider five pigeons and four pigeonholes as shown below.

Figure 1. The illustration of the pigeonhole principle in the case of N = 5, M = 4.

After distributing the four pigeons equally among the four pigeonholes, one pigeon will remain like the figure below. This pigeon can only be placed in one of the four pigeonholes, which explains how if there are N pigeons and M pigeonholes, then at least one pigeonhole must contain more than one pigeon. The following section explains the two example problems that the pigeonhole principle can be applied.

2.1.1. Example 1: Colors of socks problem

Consider a set of socks with

four different colors: orange, yellow, green, and blue. If one sock is randomly pulled out of the drawer each time, how many socks must be pulled out to guarantee a matching pair? The pigeonhole principle can be applied to this problem by treating the socks as pigeons and each color as a pigeonhole.

Figure 2. Two example problems can be solved by the pigeonhole principle. (a) colors of socks problem, and (b) the subset problem.

This demonstrates that once the four socks are distributed to each of the pigeonholes if another pigeon is distributed there must be at least one pigeonhole with more than one pigeon. Therefore, if five socks are drawn, there must be at least one pair of the same color. This conclusion is supported by the pigeonhole principle, and thus, the answer is 5.

2.1.2. Example 2: Subset problem

Another problem involves selecting a subset from the set of numbers from 1 to 100 such that no two chosen numbers add up to a multiple of 7. What is the largest possible size of this subset? To solve this problem, I must look at the remainders when an integer chosen from 1 through 100 is divided by 7. To begin with, the total number of integers with a remainder of 0, is 14. This is because it starts off with 7, which is the smallest integer that leaves a remainder of 0 when divided by 7 and ends with 98 which is the largest. By utilizing the same logic, the smallest integer with a remainder of 1 is 1 and the largest it 99. This implies the total is 15 integers. By repeatedly using the same logic on integers that leave a remainder of 2 through 6, there are 15 integers with remainder of

2, 14 integers with remainder of 3, 14 integers with remainder of 4, 14 integers with remainder of 5, and 14 integers with remainder of 6.

Since I want to find the largest set which no two integers add up to a multiple of 7, I must consider the remainder. From this, it can be interpreted that numbers with remainder of 1 and 6 cannot coexist in a set of integers for this problem. Additionally, applying the same logic, 2 and 5 cannot be both in the set of integers. Also, 3 and 4 cannot be in the same set as well. This means that in order to get the largest possible number of integers in a set, I have to get one of each pair. So, I get 15 integers from remainder of 1, 14 from either 2 or 5, and 14 from either 3 or 4.

This means that the largest possible size of this subset which has no two chosen numbers add up to a multiple of 7, is 15+14+14+1= 44. I add 1 because I can have 1 multiple of 7 as a multiple of 7 and a non-multiple of 7 added is not divisible by 7. If another multiple of 7 is added, then they add up in a multiple of 7.

2.2. Research problem: distancing points in each geometry

These simple problems serve as the foundation for utilizing the pigeonhole principle to solve more complex problems. The same principle can be applied to tackle more challenging problems, such as the triangle problem, which is the research problem of this study: “There is an equilateral has sides of length 1cm. Show that for any configuration of five points on this triangle (on the sides or in the interior), there is at least one pair of from these five points such that the distance between the two points in the pair is less than or equal to 0.5 cm”. This problem was found in Stanford University Mathematics Camp (SUMaC) 2019 Admission Exam. The proof of this problem is based on the pigeonhole principle, and I can suggest several follow-up questions such as whether the 0.5 cm can be replaced by a smaller number, or what if there are 6, 7, 8 points or so.

2.3. Proof

2.3.1. Classic proof

I can solve the research problem by dividing the 1 cm equilateral triangle, to 4 smaller triangles like Figure 3 (a). Since the four triangles are connecting the three mid points, each side of the smaller triangle is 0.5 cm. Thus, I have 4 small triangles and 5 points. The pigeonhole principle for N = 5 and M = 4 suggests that there exists at least one small triangle which contains more than two points. Because the maximum distance inside a small triangle is 0.5 cm, there is always a pair of points which have distance less than or equal to 0.5 cm.

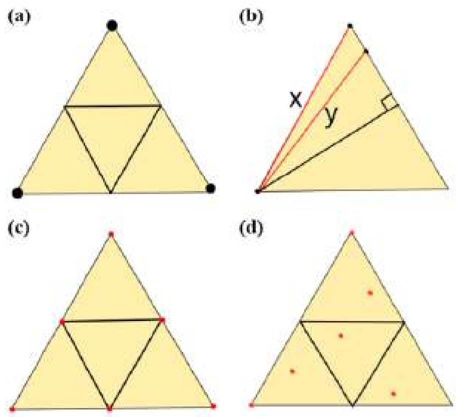


Figure 3. Classical segmentation of an equilateral into 4 smaller triangles. (a) The segmentation. (b) The distance between vertices (x) is always longer than the distance associated non-vertex points (y). (c) The optimal arrangement of 6 points to maximize the minimum distance. (d) An example of randomly placed 6 points.

Initially, the approach was to assume that there is a point on the three vertices of the triangle, as depicted with black dots in Figure 3 (a). Due to the proof that placing points on the vertices is the longest possible distance on two points chosen as Figure 3 (b), it is shown that any 5 points plotted on the big triangle, have a distance between one another, less than 0.5 cm. Detailed discussion on this is mentioned in the Appendix.

What if there are 6 points in the equilateral? In Figure 3 (c), the 6 points are all in the corner of the triangle, which maximize the minimum distance of two random points. Thus, it is proven that the points must be on the vertices of small triangles to have the distance of 0.5 cm. Especially, if points were to be randomly placed as Figure 3 (d), the distance between the points will be uneven and not fit the conditions, which will be a smaller number than 0.5. Therefore, the critical value for 6 points is also 0.5 cm. If I plot one more point in the arrangement of Figure 3 (c), I can deduce the minimum distance of a pair will be less than 0.5 cm for the case of 7 points.

What if there are 8, 9 or even 10 points? In other words, what if N equals to 8, 9, or 10? Since the pigeonhole principle can only be applied if there are at least two points in one region, the number of regions that the triangle can be divided into, can't exceed N-1. This means that the most regions that the triangle should be divided by is N-1, and the more regions divided, the smaller distance is created between the two points chosen. In the Appendix, more detailed proof is explained. In case of 10 points, I can divide the 1cm equilateral triangle into 9 smaller equilateral triangles with a side of 1/3 cm. The pigeonhole principle for N = 10 and M = 9 suggests there exists a pair of points which distance is less than or equal to 0.333 cm.

2.3.2. Alternative segmentations

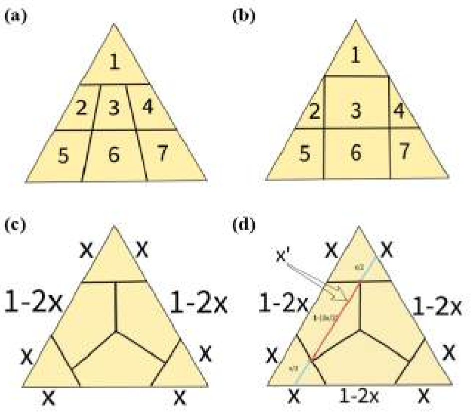


Figure 4. Segmentations for number of points more than 7. (a) One possible segmentation for 8 points. (b) Another possible segmentation for 8 points. (c) An alternative segmentation using a variable length, x. (d) The definition of x ', the longest length in the pentagons.

For 8 points case, Figure 4 (a) consists of 7 regions each can have two points on the vertices. The goal is to find out the maximum distance between two points that are possible in the diagram. There are five possible distances in this segmentation. Region 1 has 0.333, region 5,7 have 0.484, region 2,4 have 0.4, region 3 has 0.333, and region 6 has a distance of 0.4. From this observation, the maximum possible distance is region 5,7 which gives a value of 0.484. So, it can be said that this method has a 3.2% better value, compared to the 0.5 cm in the previous problem.

Another possible segmentation is Figure 4 (b) for the 8 points case. The maximum distance of regions 5,7 is about 0.44, regions 3,6 also have 0.44, and regions 2,4 are 0.33. Thus, the maximum possible distance in this segmentation is 0.44, which is 11% shorter than 0.5. Since the value of 0.44 is smaller than 0.484 of the first segmentation, it implies that the second method is better by approximately 8 %.

So far specific division methods with numbers were considered, but I can also approach the same problem with variables. In Figure 4 (c), the sides of the smaller equilateral triangle are set to x, so the remaining part would be 1-2x . There are two different regions on this diagram, one the triangle, and two the pentagon. The longest distance in the triangle is just x, while I need to consider the longest possible distance in the pentagon as illustrated in Figure 4 (d). The line of length x ', the longest line can be represented as 1-(3/2) x since the red points are defined as the midpoints of the small triangles. If I calculate all possible values, the longest distance turns out to be 0.4. The detailed derivation is explained in the Appendix.

3. Real World Simulations

3.1. Computation scheme

The simulation is designed to simulate the plotting situation of the given research problem. It creates a triangle and desired number of points inside the triangle (Figure 5 (a)). The imaginary circles have the radius of 0.5 cm; thus, the overlapped circles imply the existence of a pair which distance is less than 0.5 cm. I conducted simulations of the cases when the number of points is 2, 3, … 9. The points are randomly placed inside the triangle 10,000 times, and the minimum distance between two points chosen is stored in the memory. At the end, the average values and their standard deviations were printed out. The standard deviation σ was obtained using the definition of population standard deviation, assuming the data is being considered a population on its own. The equation is as below.

$$\sigma = \sqrt{\frac{\left[\left(x_1 - \mu\right)^2 + \left(x_2 - \mu\right)^2 + \ldots + \left(x_N - \mu\right)^2\right]}{N}}$$

where $\mu = \frac{\left(x_1 + \ldots + x_N\right)}{N}$ and is the average value for the data points $x_1, \ldots, x_N$. So, I express the measured value as $\mu \pm \sigma$.

Processing Java is suitable for simulating mathematical geometry visualization and simulation due to its robust object-oriented programming capabilities. Java's extensive libraries provide powerful tools for handling complex geometric calculations and rendering graphical representations. Additionally, Java's platform independence allows the simulation to be easily deployed across different operating systems, making it accessible to a wider range of users.

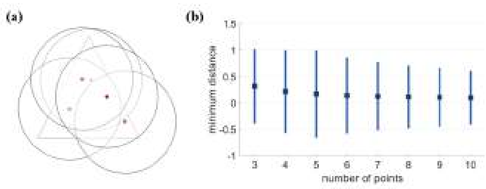
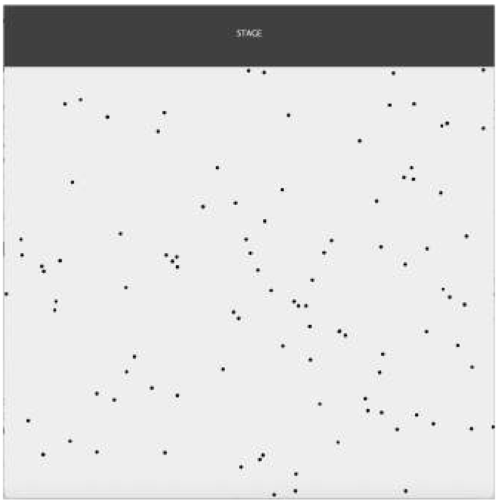


Figure 5. Processing Java Simulation of Plotting Points. (a) A screenshot of a random arrangement of 5 points. A point overlapped by other circles represent the pairs of points which distance is less than 0.5 cm. (b) Minimum distance of a pair of points for 3 to 10 points cases for each of 10,000 trials. The average values and their error bars represent the standard deviation.

3.2. Simulation & Test Results

To assess the real-world implications of the pigeonhole principle in spatial distancing scenarios, a series of simulations were conducted. Each model was designed to test how often randomly generated layouts could achieve a minimum safe distance between entities—whether human, animal, botanical, or mechanical. The results, taken from 1000 iterations per case, provide a probabilistic view of how viable spatial safety is under uncontrolled conditions. These simulations serve to bridge abstract mathematical theory with real-world applications in crowd control, ecological planning, disaster prevention, and transportation design.

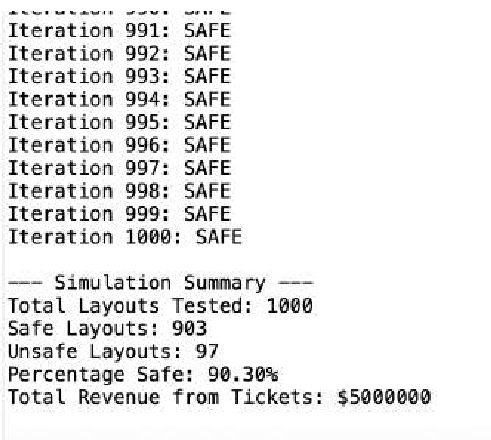
3.2.1 Concert Audience Distancing



In this simulation, 100 audience members begin at randomly initialized positions within a defined concert venue, with each person represented as a colored circle. A safe distance of approximately 2 pixels—equivalent to a real-world spacing target of around 1 feet—is enforced to reflect updated safety guidelines aiming for an 80–90% safety success rate. Rather than resetting positions entirely for each trial, individuals shift gradually by small, randomized pixel amounts from their original positions across 1000 iterations, simulating realistic micro-movements during an event.

A configuration is marked as "safe" if no two individuals are within the defined unsafe radius. The simulation outcomes show that a high safety rate is attainable only when careful spacing is maintained—supporting the pigeonhole principle: beyond a certain density threshold, avoiding overlap becomes mathematically constrained.

This approach is not only mathematically informative but also introduces a practical business perspective. With each audience member assigned a ticket price of $50, the simulation estimates potential revenue under safe and unsafe layout conditions. This dual focus allows event organizers to balance public health considerations with financial viability, emphasizing the need for spatial planning tools that combine crowd safety with economic forecasting.

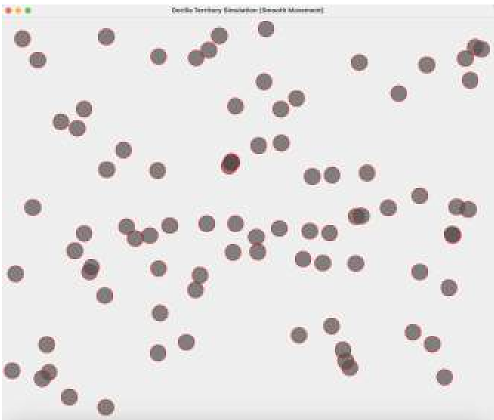


```
Iteration 991: SAFE
Iteration 992: SAFE
Iteration 993: SAFE
Iteration 994: SAFE
Iteration 995: SAFE
Iteration 996: SAFE
Iteration 997: SAFE
Iteration 998: SAFE
Iteration 999: SAFE
Iteration 1000: SAFE

--- Simulation Summary ---
Total Layouts Tested: 1000
Safe Layouts: 903
Unsafe Layouts: 97
Percentage Safe: 90.30%
Total Revenue from Tickets: $5000000
```

(Simulation result)

3.2.2 Wildlife Territorial Distribution



This simulation models territorial spacing for ten gorillas, each requiring an exclusive zone with a 100-foot radius to reflect natural avoidance behaviors observed in primate populations. To reduce the likelihood of conflict and ensure a buffer between neighboring territories, a safety distance was defined as half the territorial diameter—50 feet—based on established ethological standards that prioritize minimum approach distances in primate interactions. Rather than being randomly repositioned, the gorillas move smoothly across a 3000 ft × 3000 ft area with slight, randomized changes in direction, simulating natural wandering. Their territories are visualized as translucent circles, making it easy to detect overlaps as they interact.

Despite the expansive terrain, the results often reveal multiple overlaps in a single frame. This highlights a key insight: even when animals roam freely across what appears to be ample space, territorial conflicts emerge frequently. The pigeonhole principle is once again at play—only a finite number of non-overlapping zones can fit within any bounded region, and beyond that, overlaps become increasingly likely.

Such a simulation is valuable in conservation biology and habitat planning. By modeling movement dynamics rather than static placement, it offers a more realistic understanding of spatial requirements and potential conflict zones. This approach can inform the design of wildlife corridors, protected reserves, and other conservation strategies to ensure animal welfare and minimize territorial stress.

```
Frame 992: SAFE
Frame 993: SAFE
Frame 994: SAFE
Frame 995: SAFE
Frame 996: SAFE
Frame 997: SAFE
Frame 998: SAFE
Frame 999: SAFE
Frame 1000: SAFE

--- Simulation Summary ---
Total Frames Simulated: 1000
Safe Frames: 860
Unsafe Frames: 140
Percentage Safe: 86.00%
```

(Simulation result)

3.2.3 Traffic Following Distance

In this tunnel traffic simulation, up to 10 cars continuously travel through a 2500-foot one-lane tunnel at randomized speeds ranging from 40 to 60 miles per hour. The system is designed so that whenever a car exits the tunnel, a new one enters at the start, maintaining a consistent flow regulated by a virtual gate. Cars are placed at random initial positions and move smoothly forward. A minimum safe following distance of 50 feet is defined, and any moment where two cars are closer than this threshold is recorded as an unsafe condition. The simulation runs in real time, with safety violations evaluated once per simulated second and aggregated to assess overall tunnel safety.

The simulation reveals how even modest traffic density within a constrained space can frequently lead to spacing violations. Because each car moves at a different speed, faster cars may approach slower ones, creating unsafe proximity. The issue becomes more pronounced as new cars are introduced at regular intervals without sufficient space to accommodate changing speeds. This dynamic reinforces the cumulative impact of small safety buffers: even when each buffer is individually modest, their repeated application along a finite domain leads to a scenario where safe spacing becomes unsustainable.

This outcome exemplifies the pigeonhole principle in practice—only so many safe intervals can fit within a limited space, and once that capacity is exceeded, collisions or unsafe distances become inevitable. The model offers practical insights for real-world applications such as tunnel and highway design, where spatial constraints must be carefully balanced against throughput goals. It also holds significance for autonomous vehicle platooning algorithms, which must continuously manage inter-vehicle spacing, and for public policy decisions regarding road capacity. Ultimately, the simulation highlights how small changes in minimum required distance can have large ripple effects on the feasibility of maintaining safe, high-density traffic flow.

```
[Time 53s] UNSAFE (Unsafe frames this second: 60)
[Time 54s] UNSAFE (Unsafe frames this second: 60)
[Time 55s] UNSAFE (Unsafe frames this second: 60)
[Time 56s] UNSAFE (Unsafe frames this second: 60)
[Time 57s] UNSAFE (Unsafe frames this second: 60)
[Time 58s] UNSAFE (Unsafe frames this second: 60)
[Time 59s] UNSAFE (Unsafe frames this second: 60)
[Time 60s] UNSAFE (Unsafe frames this second: 60)

--- Simulation Complete ---
Total simulated time: 60 seconds
Safe seconds   : 7
Unsafe seconds : 53
Safety Rate    : 11.67%
```

(Simulation Result)

4. Discussion

Consider that the critical value for the physical distancing is l_c (in m) in a room of equilateral shape which length is 10 m. Then, I can define n as the floor function of 10/l_c. The floor function, which is equivalent to round down to an integer, takes as input a real number and gives as output the greatest integer less than or equal to the input. If I divide the given triangle into multiple smaller equilateral triangles in a n-staged pyramid-like segmentation just like Figure 3 (a), I have a total of $n^2 = 1+3+5+\cdots+(2n-1)$ small triangles. Thus, for (n^2+1) people, there exists at least two individuals which distance is closer than l_c. Thus, the maximum capacity based on the pigeonhole principle is n^2. For example, of 1 m distancing ($l_c=1$), the maximum capacity for an equilateral room of 10m side length is 100 people. The comparison between theoretical predictions derived from the pigeonhole principle and the outcomes of our simulations reveals a consistent trend: the number of safe configurations observed empirically tends to be lower than the theoretical maximum number of non-overlapping placements. This discrepancy illustrates that the pigeonhole principle, while useful for establishing hard constraints, often leads to overly conservative estimates of required spacing. In practice, this means that adhering strictly to the principle may result in more space being allocated than is strictly necessary, potentially reducing efficiency or feasibility in applied scenarios.

While this conservatism serves as a safeguard in sensitive environments (e.g., during a public health crisis), it is important to recognize that the trade-off between safety and utility varies by context. For instance, in concert audience planning, ensuring a three-foot distance between every attendee may significantly limit capacity and, by extension, ticket revenue. A rigid application of the pigeonhole principle in such a case could be financially unsustainable, particularly when other risk mitigation strategies (e.g., vaccination, ventilation, timed entry) are also in place. Instead, simulation data can be used to calibrate more flexible, evidence-based thresholds that still maintain a high level of safety.

In contrast, the application of this principle in wildlife territorial design layout carries fewer social constraints and potentially greater consequences for failure. For animals that rely on exclusive territories to avoid conflict or ecological stress, or in ecosystems vulnerable to rapid fire spread, a conservative spacing approach may be fully justified—even necessary. Here, the pigeonhole principle can offer strong preventative guidance by helping planners determine the minimum number of viable territory slots or safe tree placements within a finite area.

The traffic distancing model offers another nuanced case. While it might seem impractical to enforce rigid minimum spacing in real-world traffic, particularly in congested urban environments, such models can inform design and policy in automated traffic systems, self-driving vehicle logic, and highway lane density planning. In these domains, the conservative assumptions of the pigeonhole principle can be leveraged for safety without significantly compromising throughput, especially when spacing can be algorithmically maintained.

What emerges from these results is not a universal endorsement or rejection of the pigeonhole-based distancing model, but rather a spectrum of scenario-dependent viability. The principle proves highly applicable in safety-critical and spatially unconstrained environments (e.g., wildlife conservation), moderately applicable in structured systems like traffic design, and more challenging in high-density, revenue-sensitive contexts like concerts or indoor events.

Future research should aim to refine the balance between theoretical models and real-world flexibility. Increasing simulation run counts, incorporating irregular geometries (e.g., non-rectangular venues, natural landscapes), and applying agent-based modeling to simulate adaptive behavior could all improve our understanding of how distancing dynamics play out under different assumptions. In doing so, the pigeonhole principle can evolve from a static mathematical constraint into a dynamic tool for informed, context-sensitive spatial planning.

5. Next Steps

While the current simulations effectively demonstrate the conservative spatial estimates imposed by the pigeonhole principle, applying these insights in real-world decision-making will require scenario-specific refinement, stronger numerical fitting, and alignment with the operational constraints of each domain. Bridging the gap between abstract mathematical models and applied outcomes will demand both deeper simulation customization and stronger collaboration with stakeholders who manage space-critical environments.

For concert venue planners, the simulation highlights the tension between public safety and economic viability. Future work could include adaptive simulations that account for hybrid distancing policies—where only certain zones (e.g., near the stage or at entrances) enforce strict spacing while others allow for moderate clustering. By overlaying revenue optimization models onto these simulations, planners could quantify how distancing policies affect capacity, ticket pricing, and overall profitability. Integration with venue-specific blueprints and crowd flow models would further refine these predictions. This work could evolve into a software tool that planners use to test layout scenarios before each event, balancing safety thresholds with financial targets.

In the case of wildlife territorial management, the findings can be expanded with ecological data such as species-specific behavioral radii, migration tendencies, and landscape features. A more biologically-informed simulation could help conservationists model not just the physical spacing, but also the stress levels and movement patterns of animals in bounded habitats. Geographic information systems (GIS) could be integrated to map simulations directly onto real terrain, enabling planners to forecast sustainable population limits in a given wildlife reserve.

Traffic spacing models can be evolved by introducing variable vehicle sizes, speeds, and driver reaction profiles. Especially with the rise of autonomous vehicles, future simulations can test how strict distance enforcement algorithms impact highway throughput, accident probability, and fuel efficiency. These results could feed into DOT safety regulations, road design policies, or AV company testing protocols.

More broadly, one direction for future development is the creation of a generalized spatial planning simulator grounded in the pigeonhole framework but equipped with adjustable constraints, stakeholders' goals (e.g., revenue, ecological sustainability, public safety), and domain-specific physics. Such a tool could assist professionals from event organizers to city planners in quickly stress-testing layouts and making informed tradeoffs between capacity and safety.

Ultimately, the goal is to move from theoretical validation to practical utility: helping real-world stakeholders like public health officials, entertainment companies, transportation engineers, and environmental planners make decisions that are not only safer, but also smarter. By continuing to refine these models with more accurate physical parameters, behavioral logic, and business constraints, this research can become a foundation for more resilient and responsive spatial design across industries.

6. Conclusion

In conclusion, the pigeonhole principle emerges as a valuable tool for determining key parameters in physical distancing, aiding in the optimization of people capacity in various settings. The principle provides a framework to analyze and identify the appropriate minimum distance required between individuals to mitigate the transmission of infectious diseases effectively. Furthermore, the study findings demonstrate a strong agreement between the minimum distance data obtained from simulations and the theoretical predictions, validating the accuracy and reliability of the models employed. This correspondence further supports the efficacy of the pigeonhole principle in guiding decisions regarding physical distancing measures. Moving forward, future research can focus on exploring the application of the pigeonhole principle to different geometries and spatial configurations. Investigating the impact of various room shapes and layouts on optimal people capacity can enhance our understanding of how physical distancing guidelines can be adapted to diverse environments, thus enabling more tailored strategies for disease prevention and control.

6. References

[1] Beutels, P., & Faes, C., L. Verelst, F., (2020). Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006–2015). BMC Infectious Diseases, 16(1), 1-14.

[2] Baratgin, J., & Jacquet, B. (2023). The effect of cardinality in the pigeonhole principle. Thinking & Reasoning, 1-17.

[3] Cheng, H. Y., Huang, W. T., Jian, S. W., Lin, H. H., Liu, D. P., & Ng, T. C. (2020). Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Internal Medicine, 180(9), 1156-1163.

[4] J. Komlós (1990). A strange pigeon-hole principle. Order, 7, 107-113.

[5] J. Leurechon (1622). Selectæ propositiones in tota sparsim mathematica pulcherrimæ. Gasparem Bernardum.

[6] J. Miller (2015). Earliest known uses of some of the words of mathematics.

[7] M. Mersenne (1969). La vérité des sciences: contre les septiques ou Pyrrhoniens... chez Toussainct du Bray.

[8] S. S. N. Murthy (2003). The questionable historicity of the Mahabharata. Electronic Journal of Vedic Studies, 10(5), 1-15.

[9] A. Heeffer, B. Rittaud (2013). The pigeonhole principle, two centuries before Dirichlet.

[10] W. A. Trybulec (1990). Pigeonhole principle. Journal of Formalized Mathematics, 2(199), 0.

7. Appendix: Detailed Proof of the Research Problem

An equilateral has sides of length 1 cm.

(a) Show that for any

configuration of five points on this triangle (on the sides or in the interior), there is at least one pair of from these five points such that the distance between the two points in the pair is less than or equal to 0.5 cm.

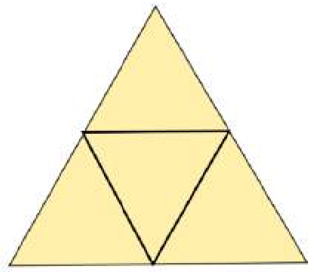
(b) Show that 0.5 cm cannot be replaced by a smaller number even if there are 6 points.

(c) If there are eight points, can 0.5 be replaced by a smaller number? Prove your answer.

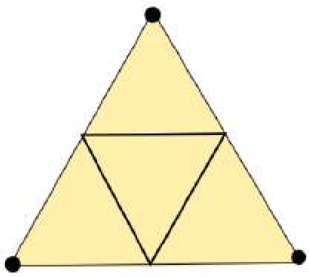
This problem was found in Stanford University Mathematics Camp (SUMaC) 2019 Admission Exam.

7.1 Part A

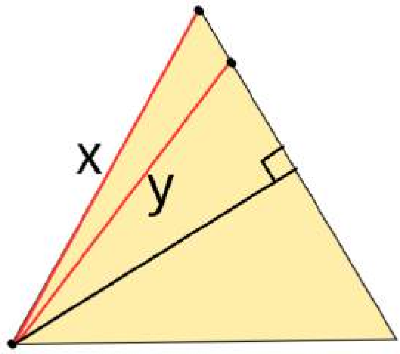
This first problem can be approached by dividing the 1 cm side equilateral triangle, to 4 smaller triangles like shown below. Since the four triangles are made by connecting the three mid points, each side of the smaller triangle is 0.5.



By doing this, each side of the smaller triangles becomes 0.5, which is the desired value. Initially, the approach was to assume that there is a point on the vertices of the triangle, as depicted below.



By doing this, the three points have been arranged in a way that maximizes their distance from each other within the triangle. By drawing a line X and Y as shown below, it can be proven that when one point is on a vertex, the other points must also be on vertices.

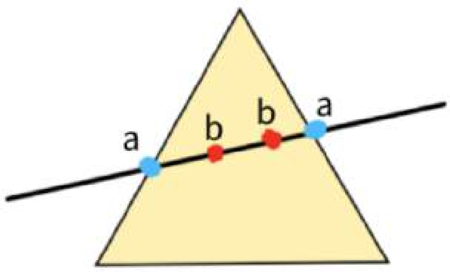


Alternatively, even if this proof is valid, it still rests on the assumption that one of the points must be located on a vertex of the triangle. Therefore, I must also establish the reasoning behind why a point must necessarily be situated on a vertex.

Since I do not have a proof of whether choosing points on vertices results in the longest length, I must consider and compare the points inside the triangle and on the edge of the triangle and figure out the maximum distance between the points.

To determine if the maximum distance between the points must be less than 0.5, it is necessary to examine the distance between two points on a small triangle and compare the values. The most effective approach is to demonstrate that the points must lie on the edge of the smaller triangle, and then explain why they must be situated at the corner of the edge.

To begin, the points can be located either on, or in the triangle. To show that the points have to be in the edges, the triangle below can be drawn.

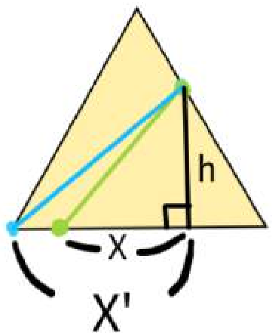


Here, two points labeled b are plotted inside the triangle. Drawing a line that passes through both points shows that there are two more sets of points labeled a on the edges. This demonstrates that any points chosen inside the triangle will have a greater distance than two points on the edge. Therefore, it is proved that the points must be on the edge to be the maximum distance apart.

To prove that the two points must be on the vertices, three cases need to be confirmed. These cases are illustrated below.

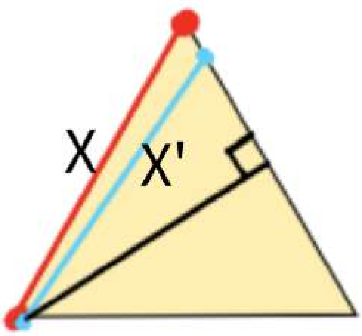
The three cases are both points being on vertices, one being on the corner and the other on edge, and both points being on edges. By comparing the second and the third triangle, then the first and the second triangle cases, the first triangle being the maximum case will be proven.

First, by comparing the case 2 and case 3, the diagram can be drawn like below.



To prove that the distance between two points as shown in case 2 is longer, I have to see that both the dots of the green line, is not in the corner; the base of the triangle that can be made is smaller than the blue line. From this, Pythagorean theory can be used by setting the base of the green line triangle as X and blue line triangle as X'. Applying the Pythagorean theorem, the length of the blue line is $\sqrt{(x')^2 + h^2}$ and the green line is, $\sqrt{(x)^2 + h^2}$. Since X < X', and X, X', h are all positive, $\sqrt{(x')^2 + h^2} > \sqrt{(x)^2 + h^2}$. This means that the blue line is longer than the green line, Although the blue line has one point that does not go to the corner, it can be seen that in order for a line to be longer, the point has to be on the edge, as closely as possible to the corner. There is no specific place the point is, but the points from the green line will meet with the point of the blue line.

Next, comparing the case of both points being in vertices, and one being in corner and the other in an edge, a diagram like the one below can be drawn.

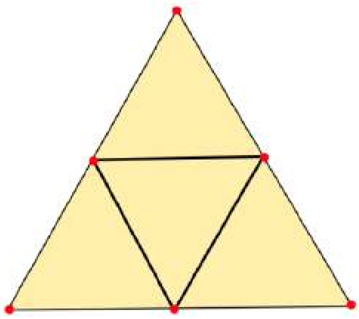


This diagram shows two lines that is made by connecting the two points. It is noticeable that the X' line can be drawn anywhere across the edge. But it can't overlap the line X. Here, the same logic as when the 2nd and 3rd case was proved, can be applied. By using the Pythagorean theorem, the same way, it is proven that the line X is longer than the line X'.

Now, case 1, where both points are in the vertices, is where the two points are plotted so that it has the maximum distance possible. This proof relates to the pigeonhole principle since now I can divide the triangle into many sections with the points on edges. Additionally, the proof can be now used to move on to the next step of, using this distance to what I originally wanted to do, prove that any 5 points plotted on the big triangle, have a distance between one another less than 0.5.

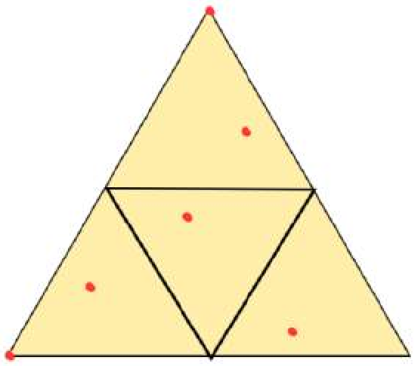
7.2 Part B

The second question to the problem was to show that .5 (in part (a)) cannot be replaced by a smaller number even if there are 6 points. In order to prove this, the drawing like below can be made.



Here, the 6 points are all in the corner of the triangle. It has to be drawn like this because, from above, it is proven that the points have to be on the vertices to have the desired 0.5. So, this means that the other 1 point must also be in the vertices. As the drawing shows, there is at least one pair of points that still have 0.5.

Specifically, if points were to be randomly placed like below, the distance between the points will be uneven and not fit the conditions, which will be a smaller number then 0.5.



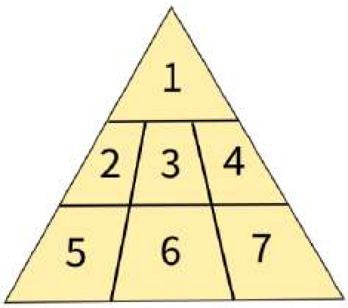
7.3 Part C

The final question is, if there are eight points, can 0.5 be replaced by a smaller number? Prove your answer.

Since, the pigeonhole principle can only be applied if there are at least two points in one region, the number of regions that the triangle can be divided into, can't exceed 7, because it is given that there is 8 points. This means that the most region that the triangle should be divided, to apply the pigeonhole principle, is 7, and the more regions divided, the smaller distance is created between two points chosen.

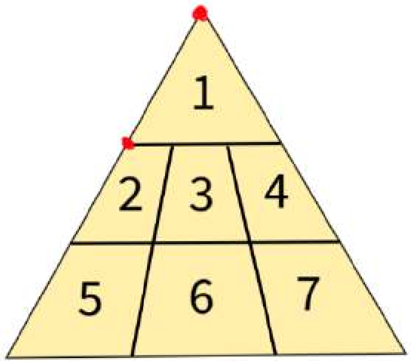
7.3.1 Approach 1

Since the maximum regions for the triangle to be divided to use pigeonhole is 7, there has to be some values that is given, in order to find the distance between two points chosen. The first approach is the drawing below.



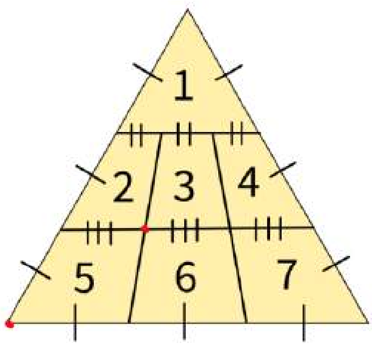
Here, the drawing consists of 7 regions that each can have two points on the vertices. The goal is to find out the maximum distance between two points that are possible in the diagram. There are 5 possible distances in this drawing.

First, the two dots at region 1 as drawn below, have the distance of 0.33 or $\frac{1}{3}$, since the side of the triangle is 1, and the drawing is drawn so that the lines are dividing the side into three equal parts.

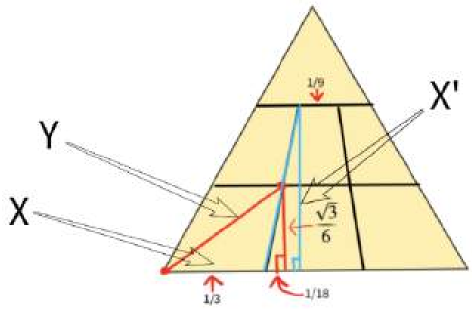


7.3.1.1 Regions 5,7

Next, another set of points can be drawn like below.



I don't calculate region 7 because it has a symmetry with region 5, which means they are the same.

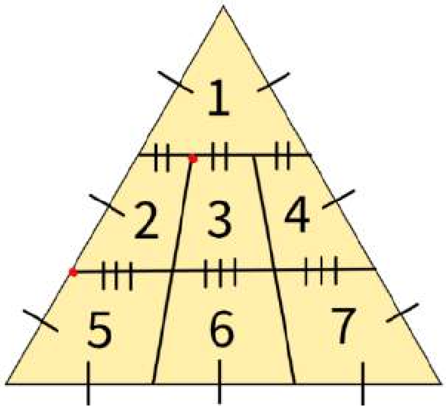


Here, to calculate the line Y, the two right triangles above have to be drawn. To begin, the height of the triangle X can be found by looking at the height of the entire equilateral triangle. The equilateral triangle has the height of $\frac{\sqrt{3}}{2}$ since the I have the hypotenuse of 1 and the base of $\frac{1}{2}$. So, by applying the Pythagorean theorem, it leads to $\frac{\sqrt{3}}{2}$. Then, the height of the triangle X, which is $\frac{1}{3}$ of the total height of the equilateral triangle, becomes $\frac{\sqrt{3}}{6}$. Next, to find the base, the triangle X' has to be used. The height ratio of the triangle X' to the small triangle inside the triangle X is 2:1. This means that the base of the small triangle is half the base of the triangle X'. Looking at the diagram, it can be seen that the line that contains both bases has the length of 1/3. The line that is parallel to it on the top, has the length of 1/9. This means that if I subtract 1/3 by 1/9, then divide it by 2, the base of the triangle X' is found. The value of the triangle X' base is 1/9, so if I divide it by 2 again, it becomes 1/18 which is the base value of the small triangle. Finally, the total value of the triangle X's base is 1/3 added to 1/18, which leads to $\frac{7}{18}$.

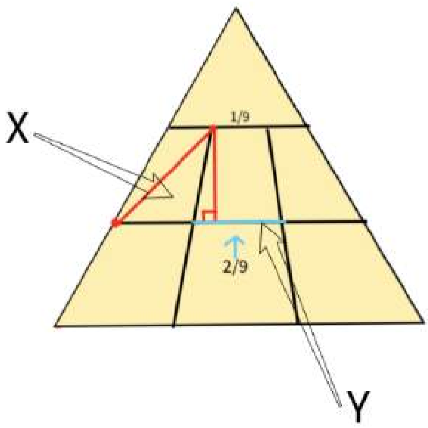
Since now the base and the height of the triangle X is known, the line segment's length can be calculated by the Pythagorean theorem. If I apply $\sqrt{a^2 + b^2} = c$, the value becomes $\frac{\sqrt{19}}{9}$, which is about 0.484. This means that the diagram drawn is a decent solution, but not the best since it slightly decreased from 0.5.

7.3.1.2 Regions 2, 4

Next, another set of possible distance is drawn below.



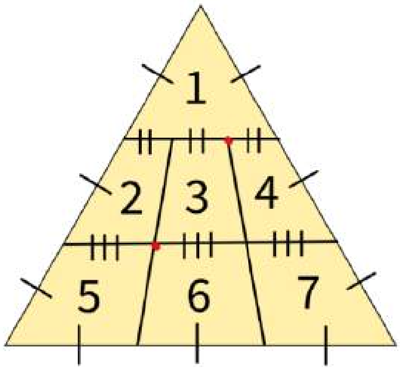
The region 4 is not calculated because of the symmetry in region 2. Region 2 can be calculated from the diagram below.



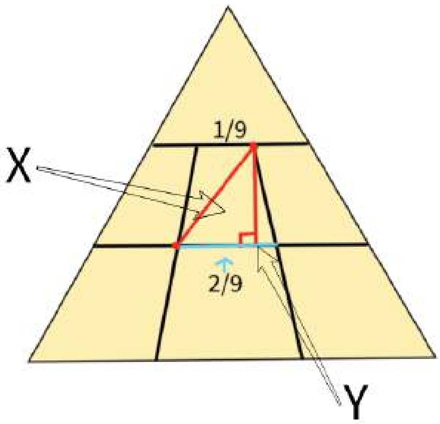
Here, the line Y is $\frac{2}{9}$ due to it being $\frac{1}{3}$ of the long line, which is a equilateral triangle with a side length of $\frac{2}{3}$. To find the base of the triangle X, two parts of the base, has to be found separately then added. The smaller part can be calculated by subtracting $\frac{1}{9}$ from $\frac{2}{9}$, then dividing it by 2. This leads to the value of the small part of the base of triangle X being $\frac{1}{18}$. Then to find the longer part, it is just $\frac{2}{9}$. So now, the value of the base of triangle X is $\frac{1}{18}, \frac{2}{9}$ which is $\frac{5}{18}$. The height of the triangle X is $\frac{\sqrt{3}}{6}$ since it was calculated in the previous part. When the Pythagorean theorem is applied, the distance between the two points become about 0.4, This means that region 4 and region 2 is possible, but not the best.

7.3.1.3 Region 3

To continue, the next possible region is region 3. The diagram can be drawn as below.



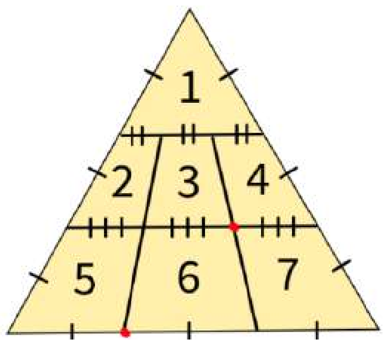
To calculate the region the diagram below must be drawn.



Here, the same values used before can be utilized to calculate the distance between the two points. The height of the triangle X is $\frac{\sqrt{3}}{6}$, as it has the same height as region 2, 5. The line Y was already calculated with the value of $\frac{2}{9}$. So, if the line Y is subtracted by $\frac{1}{9}$, then divided by 2, it will give the small part of the line that does not consist the red base. The value is $\frac{1}{18}$, so if the $\frac{2}{9}$ is subtracted by it, it gives $\frac{3}{18}$ or $\frac{1}{6}$, which is the value of the base. Now if Pythagorean theorem is applied, the distance between the two point is $\frac{1}{3}$. This means that region 3 is not possible, as I want the value to be bigger than 0.33.

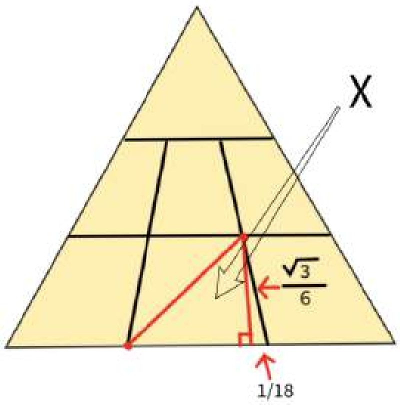
7.3.1.4 Region 6

Finally, the last region to be calculated is the region 6. The region 6's diagram can be drawn as below.



Here, in order to calculate the

distance between the two points, the drawing like below has to be created.

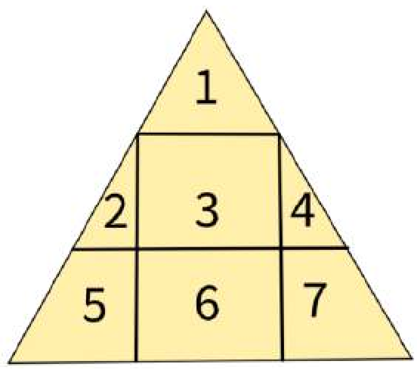


The distance between the two points can be calculated from, using the Pythagorean theorem. The height of the triangle X is just $\frac{\sqrt{3}}{6}$ as calculated in the previous regions. Since the small line that does not fit in the base of triangle X but is $\frac{1}{3}$ of the upper side that was calculated as $\frac{1}{9}$, is $\frac{1}{18}$, the base of the triangle X is $\frac{1}{3} - \frac{1}{18}$, which is $\frac{5}{18}$. So the hypotenuse, which is the desired value, is $\sqrt{\left(\frac{5}{18}\right)^2 + \left(\frac{\sqrt{3}}{6}\right)^2}$. This leads to the distance being about 0.4.

To sum up, the best distance that is possible out of the model, is region 5,7 which gives a value of 0.484. So it can be said that this method has given a 3.2% better value, then the 0.5 in the previous problem.

7.3.2 Approach 2

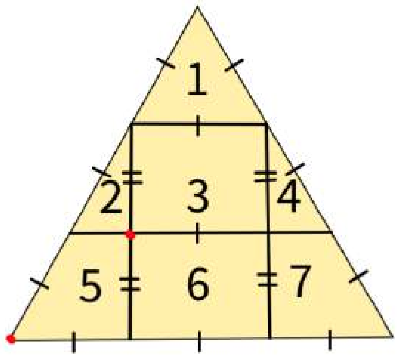
A diagram which can be drawn, to get a smaller value is a drawing like below which only has points that intersect the edges at 1/3 value points.



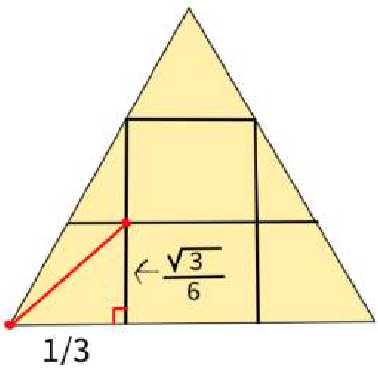
Here, there are three places that must be tested, in order to make sure there are no exceptions.

7.3.2.1 Region 5,7

First, the two points below has to be tested.



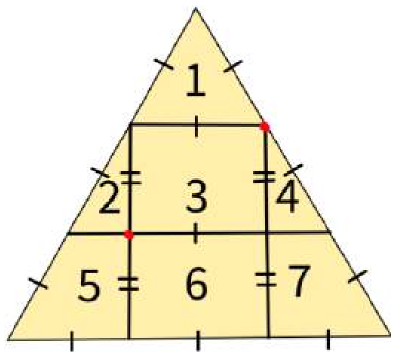
Here, only region 5 is calculated since 7 is symmetric to it. The diagram like below can be drawn to visualize the needed values to calculate the distances.



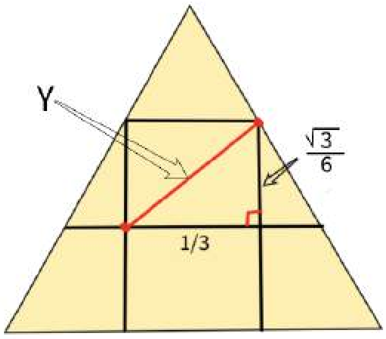
To figure out the distance, the value of height and the base has to be found to use the Pythagorean theorem. The base is 1/3 because the lines were divided that way, and the height is 1/3 of the height of the entire equilateral triangle. As explained in the previous method, the values is $\frac{\sqrt{3}}{6}$. After using the Pythagorean theorem, the value of the distance between two points is $\frac{\sqrt{7}}{6}$, which is about 0.44.

7.3.2.2 Region 3,6

Next, the second possible way, is the two points below.



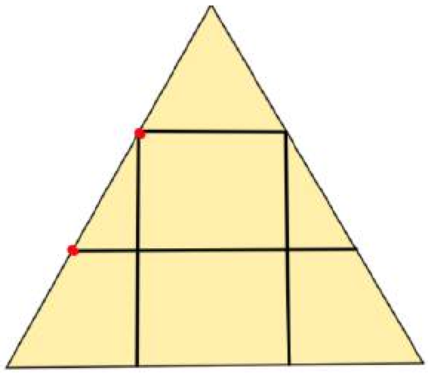
Only region 3 is calculated because region 6 has a symmetry with it. The distance between the two points can be figured out by making the diagram like below.



The line Y, which is what is to be found, is on a triangle that has the height of $\frac{\sqrt{3}}{6}$ and the base of $\frac{1}{3}$. Pythagorean theorem can be applied in the same way to find the hypotenuse, which becomes the value of $\frac{\sqrt{7}}{6}$, which is about 0.44.

7.3.2.3 Regions 2,4

Finally, the last regions to be tested are 2, 4. Since 4 is symmetric to region 2, I only calculate one of them. The drawing like below can be drawn to calculate the distance.



Region 2 has the longest line as it is the side. This means that the distance is just 0.33, which is smaller than the other regions.

To sum up, the best possible distance in this approach is 0.44, which is 11% better than 0.5.

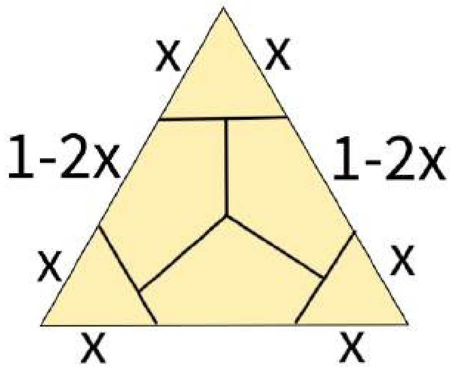
7.3.2.4 Comparison between two approaches

Since the value of 0.44 is smaller than 0.484, it means that this way is a better way than the first approach. The second approach is better by about 8 percent.

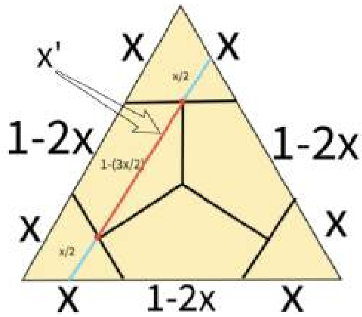
7.3 Approach 3 (with variables)

Another drawing that can be

drawn to get a good value is the diagram below.

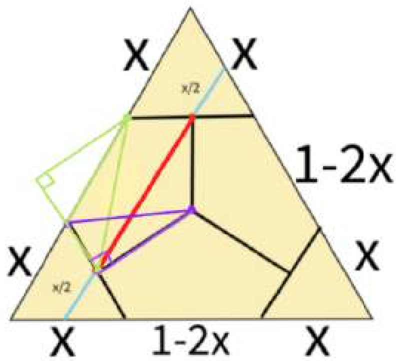


Here, each side of the equilateral triangle is 1. So, if the sides of the smaller equilateral triangle are set to x, the remaining part would be $1 - 2x$. There are two different regions on this diagram, one the triangle, and two the pentagon. The longest distance in the triangle is just x, while to calculate the longest distance in the pentagon, the drawing like below can be made.



The drawing above shows the longest possible distance in the pentagon.

Here, only the line X’ is calculated because it is the longest line. This can be proven by the drawing below.



First, comparing the green line and the purple line, the base and the altitude make the hypotenuse longer for the green triangle. This means that inside the pentagon, the green line is a longer line choice. Next, the red line contains the same altitude as the green line but has a longer base, leading to the red line being the longest choice inside the pentagon.

The red line, which is the longest distance in the region, is $1 - \frac{3x}{2}$. This is due to since, by connecting the blue and red lines, it creates a side of a new triangle. This triangle is an equilateral triangle with a side of either $x + (1 - 2x) + x$ or $\frac{x}{2} + (A) + \frac{x}{2}$. So if the A is the desired value, the equations have to be set to equal. Which means the equation of $x + (1 - 2x) + x = \frac{x}{2}$+(A)+$\frac{x}{2}$ is created. When calculated, the A becomes $1 - \frac{3x}{2}$. After the red line is calculated, this has to be applied to the main purpose. The focus on this method is, finding the value of x or $1 - \frac{3x}{2}$ so that the bigger one is the smallest possible. This means that when they are equal, it is the value desired.

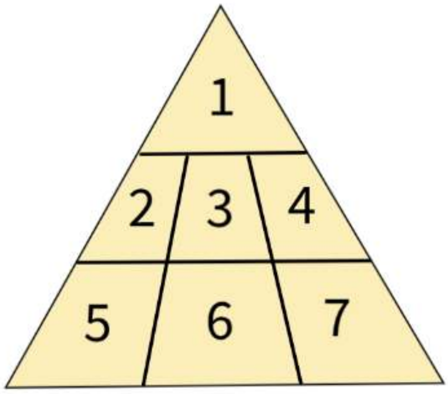
When they are set to equal, the equation of $x = 1 - \frac{3x}{2}$ is created. By adding $\frac{3x}{2}$to both sides, the equation becomes $\frac{5x}{2} = 1$. This leads to x being $\frac{2}{5}$. This means that the x has a value of 0.4. Looking at this value, it is better than the value of 0.5 by 20%.

7.3.1 Comparison to approaches before

Now it can be seen that the 3rd approach which gives a value of 0.4 is better than the other two by a significant amount of about 10% and 18%. This concludes how the last approach is the best approach to calculate the minimum distance that is over 0.33 but below 0.5.

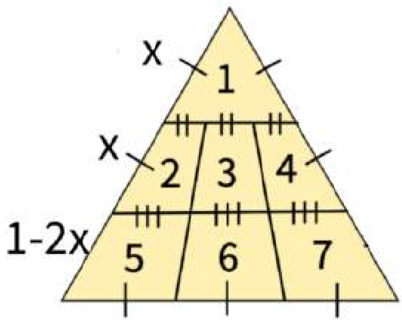
7.4 An addition to the first approach

Originally, the first approach had a set value of $\frac{1}{3}$ on each section like below.



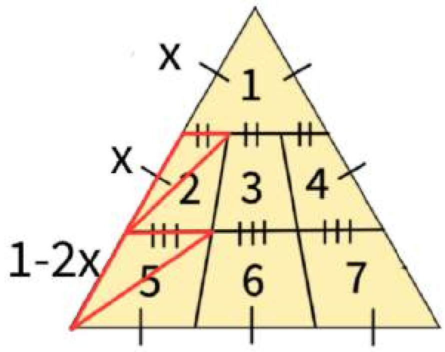
On the other hand, is each section having the side length of $\frac{1}{3}$ in this shape, really the best possible length?

To find that out, each section of the triangle must be labeled like below.



Here, two sides are given the value of x and the other has the value of $1 - 2x$. This is done this way because, since region 5 has to longest value if the side length of it is decreased and the other two sides are increased, then the value will be the smallest length p possible which is the best length out of the regions.

To figure out what value the x must be, the following diagram has to be drawn.



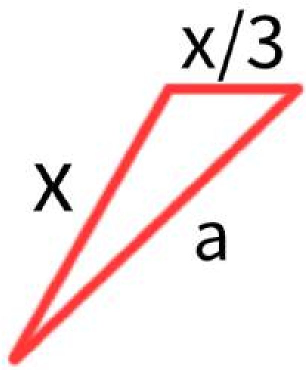
Here, by using the law of cosine,

$a^2 = b^2 + c^2 - 2bc \cos cos\,(angle)$

, the regions' distances can be represented in x. Region 1 is just x since it is just the side of the triangle. When the law of cosine is applied to region 2,4, the equation produced is

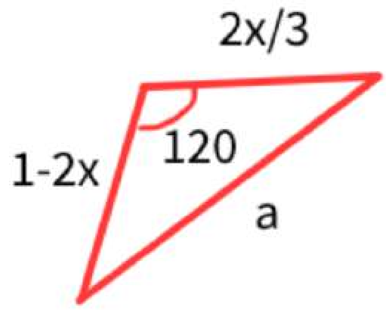
$a^2 = x^2 + \left(\frac{x}{3}\right)^2 - 2x\left(\frac{x}{3}\right) \cos cos\,(120)$

. This can be shown in the drawing below.



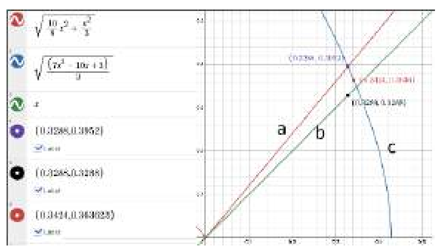
This leads to $a^2 = \frac{10}{9}x^2 + \frac{x^2}{3}$. Since the value of "a" is desired, the equation has to be square rooted on both sides which becomes $a = \sqrt{\frac{10}{9}x^2 + \frac{x^2}{3}}$.

Next, the region 5,7 can also be calculated by using the law of cosine. When I apply it, the equation of $a^2 = (1-2x)^2 + \left(\frac{2x}{3}\right)^2 - 2(1-2x$ is made. The law of cosine is from the drawing below.



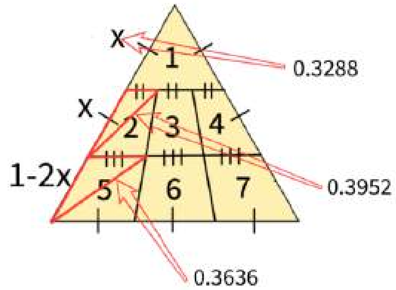
Now if both sides of the equation are square-rooted and simplified, it gives the equation of $a = \sqrt{\frac{7x^2 - 10x + 3}{3}}$.

Now that the longest distance of each region is calculated, the app "Desmos" can be used to draw graphs for each one. When each equation is inputted, the following graphs are drawn.



Here, graph b is region 1, graph a is region 2,4, and graph c is region 5,7. The x-coordinate of graphs b and c must be equal because the side of the section was both set as x. While on the other hand, the x-coordinate of graph a must be 1-2x.

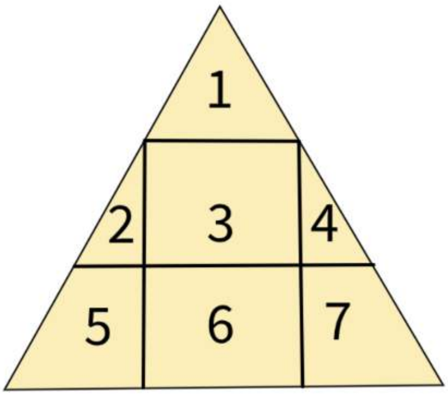
Now looking at the graph, the goal of this was to minimize the longest distance out of the regions. That distance would be represented as the y-value in the graphs. The longest distance would be 0.3952. This is because the x-value is increased from 0.3288, and the y-value of the graph b will continuously increase, which means a higher value. So, keeping the intersection of the graph a and c is the best option. Since the x-value has to be 0.3288, the 1-2x would be 0.3424. This leads to the x-coordinate of the green line being 0.3424. Since 0.3636, which is the y-value, does not exceed 0.3952, the largest distance would be 0.3952 in this case. To sum up, the following drawing can be made.



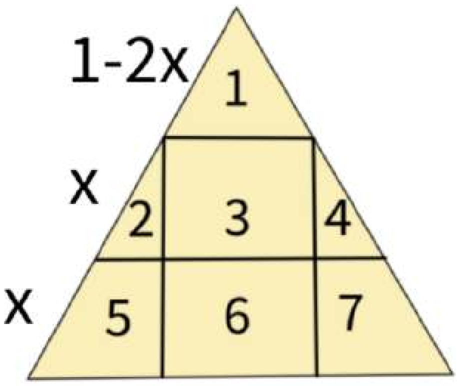
In conclusion, setting the two section's side lengths as x leads to the best possible distance being 0.3952 at region 2,4. This value is about 21% better than 0.5. Additionally, this method has a better distance than the same drawing method that had a side that was set as $\frac{1}{3}$.

7.5 An addition to the second approach

Originally, the second approach was the drawing below.



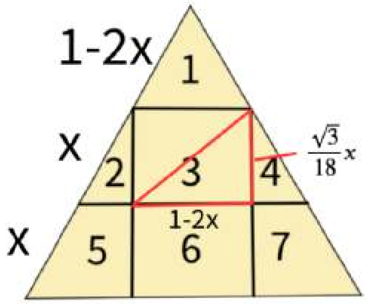
Since this method was also assumed with each section's side as $\frac{1}{3}$, the sides represented by x can be assigned as below.



The original method had region 1's side as the smallest, so if the value of x is decreased, the side of region 1 will increase which leads to the smaller distances of the overall region.

7.5.1 Region 3

To begin, region 3, which has the longest distance from the regions in its row, can be calculated by making a drawing like below.



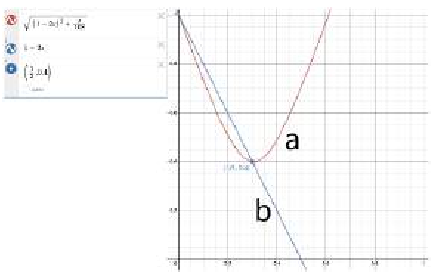
This is because since the 1-2x side creates a small equilateral triangle, the base of the triangle that needs to be found, has the same base as it. Next, $\frac{\sqrt{3}}{18}x$ comes from the previous method as, since when the side was $\frac{1}{3}$, it created the height of $\frac{\sqrt{3}}{6}$. Since the height to section side ratio is always constant (by similar triangles), when the section side becomes x, the height must be divided by 3 and then multi 3 and then multiplied.

If the Pythagorean theorem is applied, the distance between the two opposite vertices becomes, $\sqrt{(1 - 2x)^2 + \frac{x}{108}}$.

7.5.2 Region 5

Next, region 5 is the same as region 3, as when the method was done by using $\frac{1}{3}$, it was equal, which means when the side is x, it is still the same.

Now that all regions are calculated, it must be graphed in DESMOS to calculate the maximum distance out of the minimized value.



Here, graph a intersects with graph b at $\left(\frac{1}{3}, 0.4\right)$. If the value of x is decreased or increased, the y-value of either graph a or b will increase. This means that the x value of graph a, which is $\frac{1}{3}$, is the best x value.

This shows how the previous method had each section side as $\frac{1}{3}$, was the best option for minimizing the maximum distance. In conclusion, 0.4 is the smallest distance between two points.