

# Random Ordering Sudoku Solver Analysis

Steven Kim

2025-11-14

Check for packages:

## Read Data:

```
rand_1 <- read_csv(here("data", "rand_sample_1.csv"))

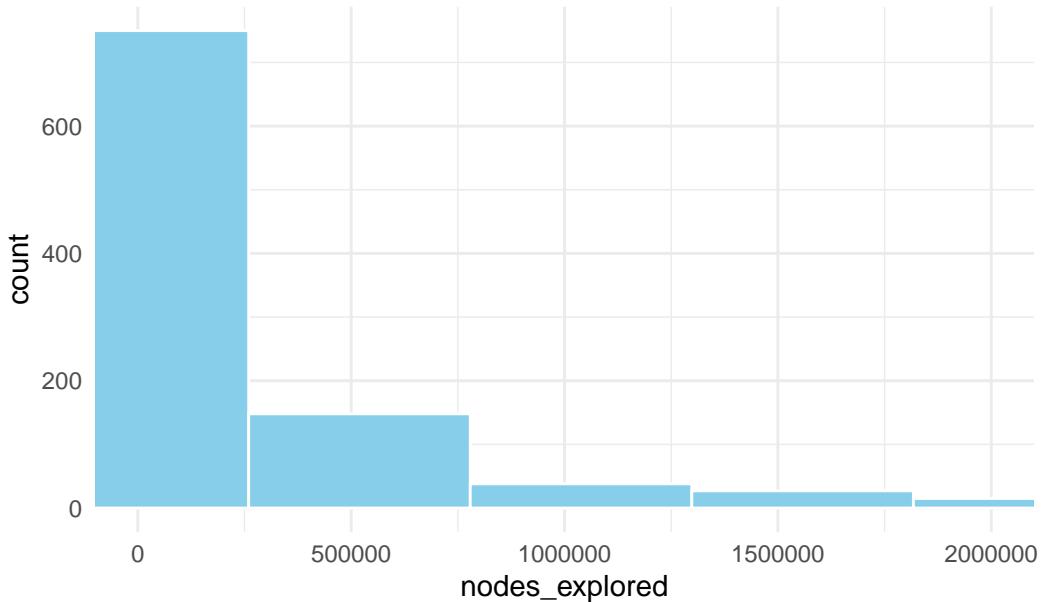
processed_rand_1 <- rand_1 |>
  filter(is_solved == TRUE)

glimpse(processed_rand_1)
```

```
Rows: 1,000
Columns: 12
$puzzle_id <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$puzzle <chr> "1..5.37..6.3..8.9.....98...1.....8761.....~
$clues <dbl> 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, ~
$difficulty <dbl> 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, ~
$run_id <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ~
$solutions_found <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$nodes_explored <dbl> 64805, 698410, 106409, 321338, 158532, 236683, 590~
$max_recursion_depth <dbl> 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53~
$solve_time_ms <dbl> 8, 94, 14, 43, 22, 32, 0, 0, 545, 6, 1, 0, 26, 12, ~
$is_solved <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, ~
$leaves <dbl> 18900, 229293, 38121, 111692, 41730, 83737, 1743, ~
$backtracks <dbl> 64751, 698356, 106355, 321284, 158478, 236629, 584~
```

```
ggplot(processed_rand_1, aes(x = nodes_explored)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "white") +
  coord_cartesian(xlim = c(0, 2000000)) +
  labs(title = "Distribution of Nodes Explored") +
  theme_minimal()
```

## Distribution of Nodes Explored



```
rand_10_100 <- read_csv(here("data", "rand_sample_10_1000.csv"))

processed_rand_10_100 <- rand_10_100 |>
  filter(is_solved == TRUE)

glimpse(processed_rand_10_100)
```

Rows: 10,000  
Columns: 12

\$ puzzle_id	<dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
\$ puzzle	<chr> "1..5.37..6.3..8.9.....98...1.....8761.....~
\$ clues	<dbl> 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27~
\$ difficulty	<dbl> 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, ~
\$ run_id	<dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ~
\$ solutions_found	<dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
\$ nodes_explored	<dbl> 1310, 33462, 976292, 23306, 2115, 16571, 35816, 61~
\$ max_recursion_depth	<dbl> 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53~
\$ solve_time_ms	<dbl> 0, 4, 127, 3, 0, 2, 4, 0, 27, 210, 6, 4, 10, 44, 3~
\$ is_solved	<lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TR~
\$ leaves	<dbl> 390, 10830, 326622, 7391, 632, 5332, 9126, 155, 60~
\$ backtracks	<dbl> 1256, 33408, 976238, 23252, 2061, 16517, 35762, 56~

```

summary_stats <- processed_rand_10_100 |>
  group_by(puzzle_id) |>
  summarise(
    mean_nodes = mean(nodes_explored),
    median_nodes = median(nodes_explored),
    sd_nodes = sd(nodes_explored),
    min_nodes = min(nodes_explored),
    max_nodes = max(nodes_explored),
  ) |>
  arrange(mean_nodes)

```

```
summary_stats
```

```

# A tibble: 10 x 6
  puzzle_id mean_nodes median_nodes sd_nodes min_nodes max_nodes
  <dbl>      <dbl>       <dbl>     <dbl>      <dbl>     <dbl>
1       4      85475.    34412.   154958.     135    2582152
2       3     126201.    67416.   181675.     408    1861811
3       7     142255.    45490.   328673.      66    3953269
4       5     222399.    68066.   402425.     114    3306088
5       1     297075.    89288.   685826.     150    9592845
6       8     333465.   138840.   533937.     268    5752925
7       6     507361.    87464.   1187861.    449    12799177
8       9     683555.    92756.   2324342.     81    38847691
9      10     689413.   224174.   1162391.    282    11473428
10      2     863406.   225328.   1938290.    1387   24943389

```

```

processed_rand_10_100 |>
  mutate(puzzle_id = fct_reorder(factor(puzzle_id), nodes_explored, .fun = mean)) |>
  ggplot(aes(x = puzzle_id, y = nodes_explored)) +
  geom_violin(alpha = 0.6) +
  geom_boxplot(aes(color = difficulty)) +
#  coord_cartesian(ylim = c(0, 20000000)) +
  labs(
    title = "Puzzles Ordered by Average Nodes Explored",
    x = "Puzzle ID (ordered by mean)",
    y = "Nodes Explored"
  ) +
  theme_minimal()

```

