Sudoku Solver Analysis

Steven Kim

2025-11-14

Check for packages:

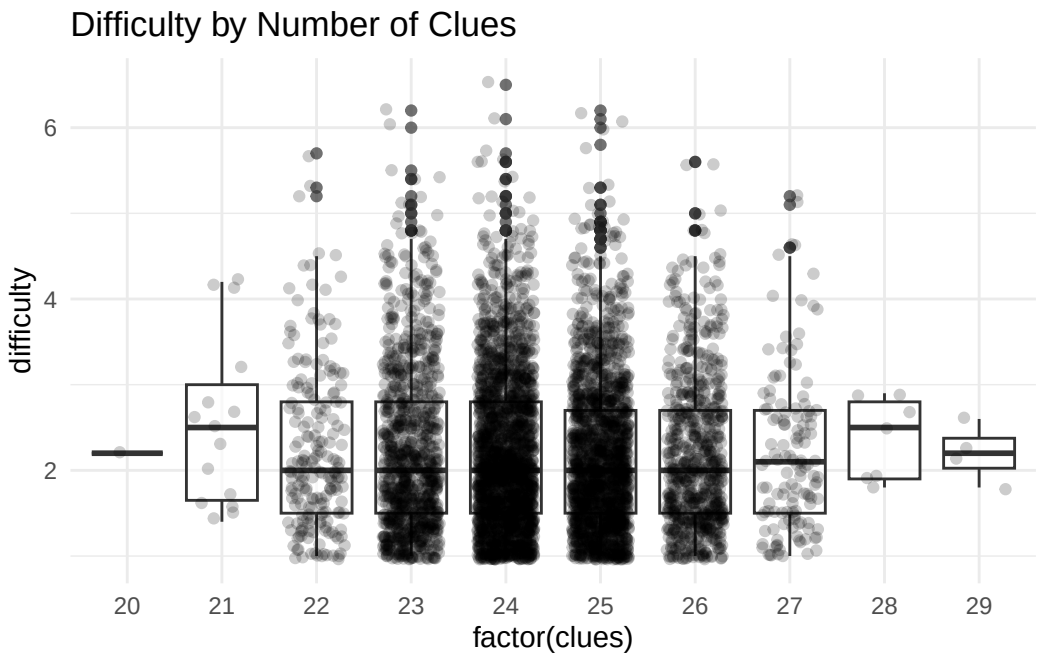
Read Data:

```
sudoku_data <- read_csv(here("data", "sample_10000.csv"))

sudoku_data_processed <- sudoku_data |>
  filter(is_solved == TRUE) |>
  filter(difficulty != 0)

glimpse(sudoku_data_processed)
```

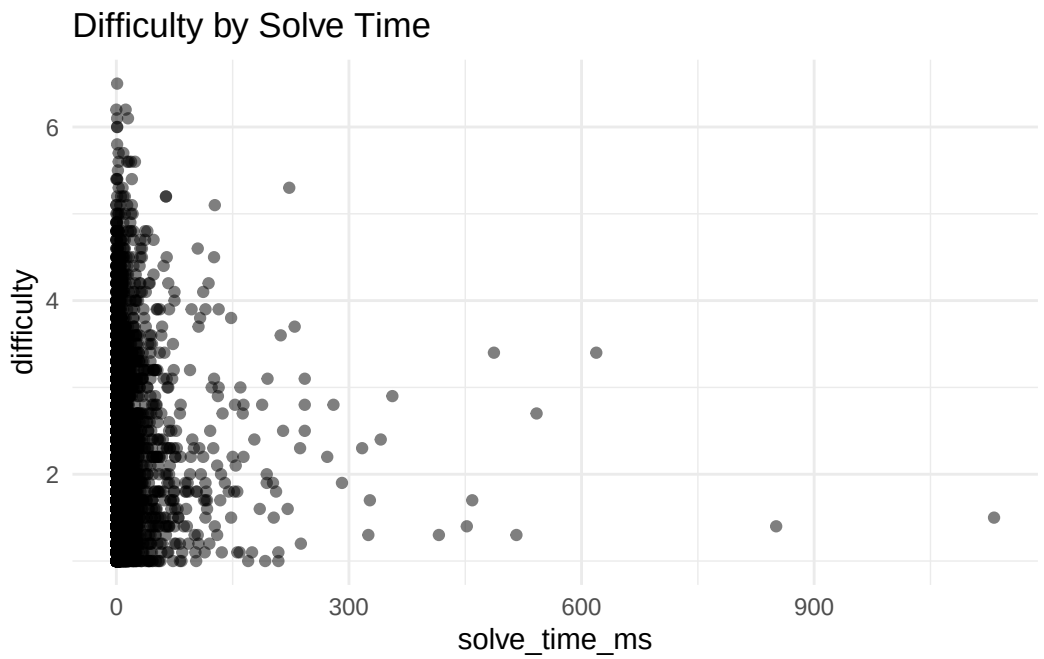
```
Rows: 5,585
Columns: 10
$ id                 <dbl> 1, 11, 41, 61, 71, 81, 101, 111, 161, 171, 181, 19~
$ puzzle             <chr> "1..5.37..6.3..8.9......98...1.......8761.........~
$ clues              <dbl> 27, 26, 24, 26, 24, 25, 27, 23, 24, 25, 25, 24, 24~
$ difficulty         <dbl> 2.2, 1.5, 3.8, 2.2, 2.0, 2.2, 4.6, 1.6, 2.4, 3.1, ~
$ solutions_found    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ nodes_explored     <dbl> 110300, 3073, 18266, 10832, 27909, 3575, 3556, 224~
$ max_recursion_depth <dbl> 53, 54, 56, 54, 56, 55, 53, 57, 56, 55, 55, 56, 56~
$ solve_time_ms      <dbl> 15, 0, 2, 1, 3, 0, 0, 3, 9, 6, 9, 1, 16, 5, 152, 1~
$ is_solved          <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TR~
$ leaves             <dbl> 33553, 855, 5611, 3011, 9476, 1195, 975, 5744, 217~
```

Preliminary Graphs:

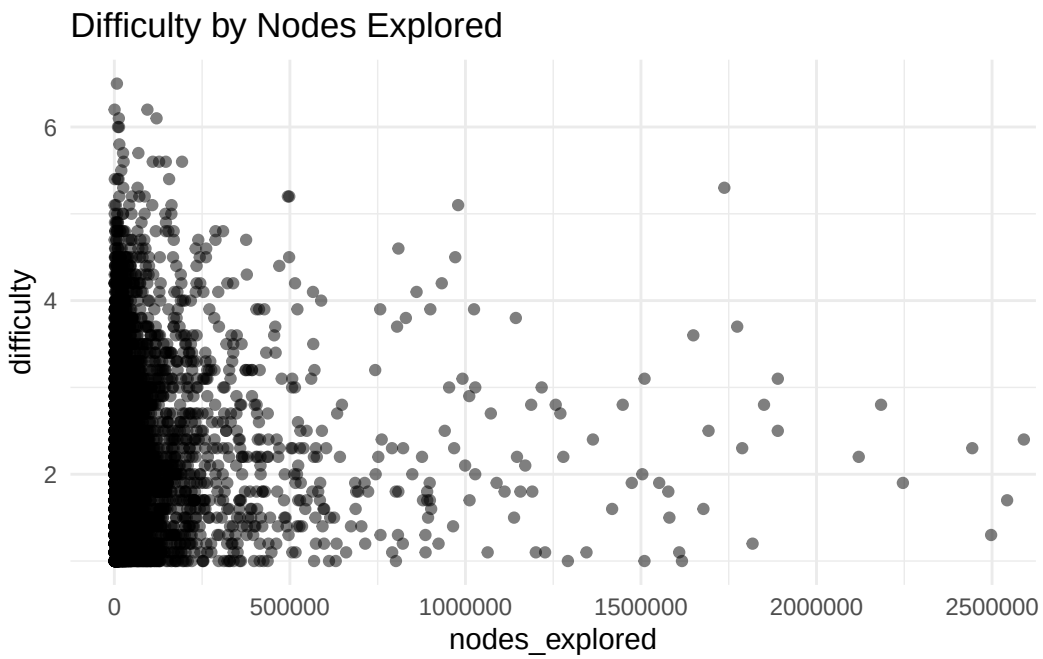
```r
ggplot(sudoku_data_processed, aes(x = factor(clues), y = difficulty)) +
  geom_boxplot(alpha = 0.7) +
  geom_jitter(alpha = 0.2, width = 0.3) +
  labs(title = "Difficulty by Number of Clues") +
  theme_minimal()
```



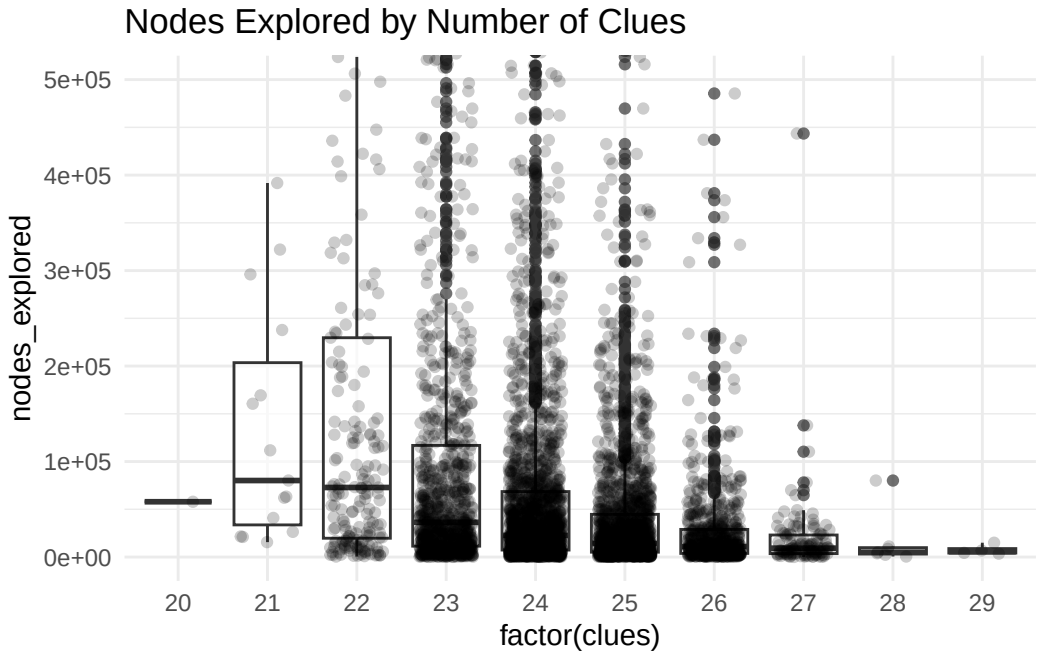

```r
ggplot(sudoku_data_processed, aes(x = solve_time_ms, y = difficulty)) +
  geom_point(alpha = 0.5) +
# coord_cartesian(xlim = c(0, 2500000)) +
  labs(title = "Difficulty by Solve Time") +
  theme_minimal()
```




```
ggplot(sudoku_data_processed, aes(x = nodes_explored, y = difficulty)) +
  geom_point(alpha = 0.5) +
  coord_cartesian(xlim = c(0, 2500000)) +
  labs(title = "Difficulty by Nodes Explored") +
  theme_minimal()
```




```
ggplot(sudoku_data_processed, aes(x = factor(clues), y = nodes_explored)) +
  geom_boxplot(alpha = 0.7) +
  geom_jitter(alpha = 0.2, width = 0.3) +
  coord_cartesian(ylim = c(0, 500000)) +
  labs(title = "Nodes Explored by Number of Clues") +
  theme_minimal()
```


Nodes Explored by Number of Clues
nodes_explored
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
20
21
22
23
24
25
26
27
28
29
factor(clues)