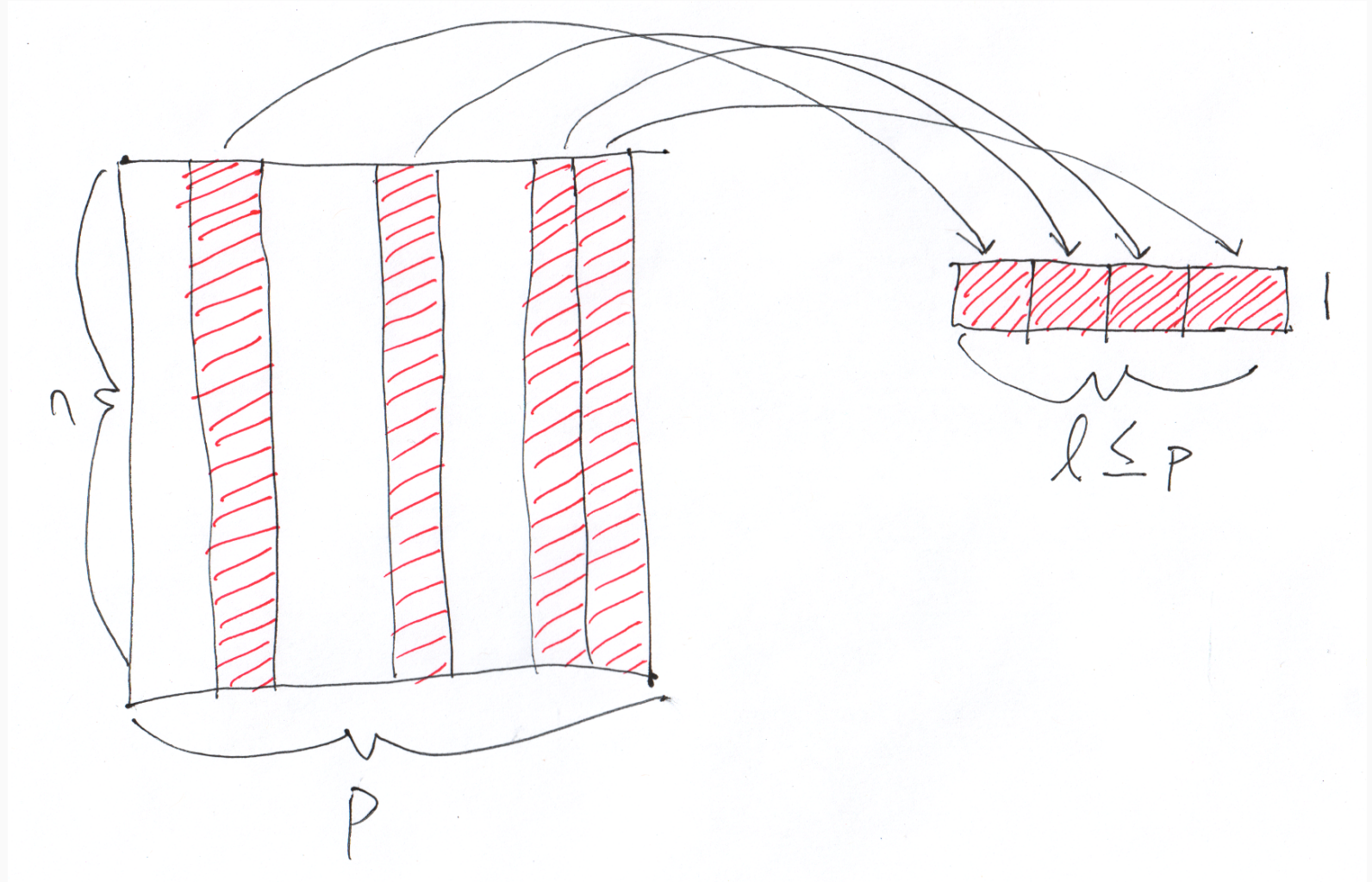


Data transformation

The `group_by` and `summarize` functions



group_by and summarize schematic



Structure of the group_by and summarize functions

```
name_of_dataset %>%  
  group_by(column_one, column_three) %>%  
  summarize(  
    count = n(),  
    summation = sum(column_two)  
  )
```

The above `summarize` operations on the right of the equals signs are just two examples of the types of data aggregations you can perform.

summarize operations

The following are a few examples of the types of data aggregation that you can perform using `summarize`.

- `n()`: Counts number of rows in a group
- `sum()`: For numerical variables, sums rows within a group
- **statistical:** `mean()`, `median()`, `sd()`, `min()`, `max()`
- Counts and proportions of logical values: `sum(x > 10)`, `mean(y == 0)`

group_by and summarize demo

Follow along in RStudio

Credits

License

[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](#)

Acknowledgments

Ideas and examples for the dplyr demos adapted from *Modern Data Science with R* by Benjamin Baumer, Daniel Kaplan, and Nicholas Horton, chapter 4.