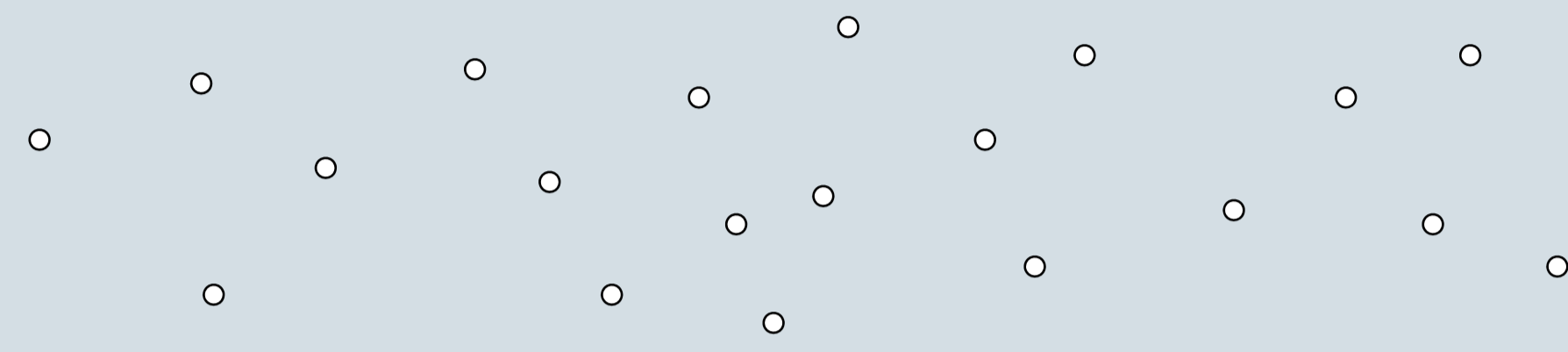




Range Median – Data Structures and Batched Queries

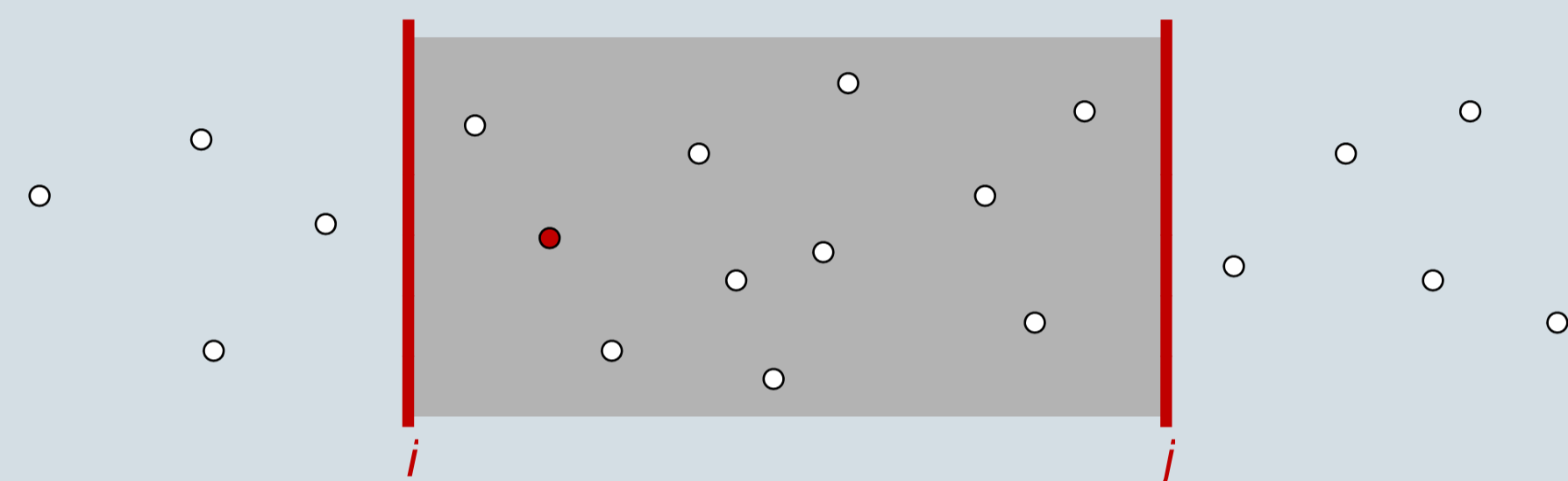
Range Median Data Structure

Maintain a set of n points in the plane.



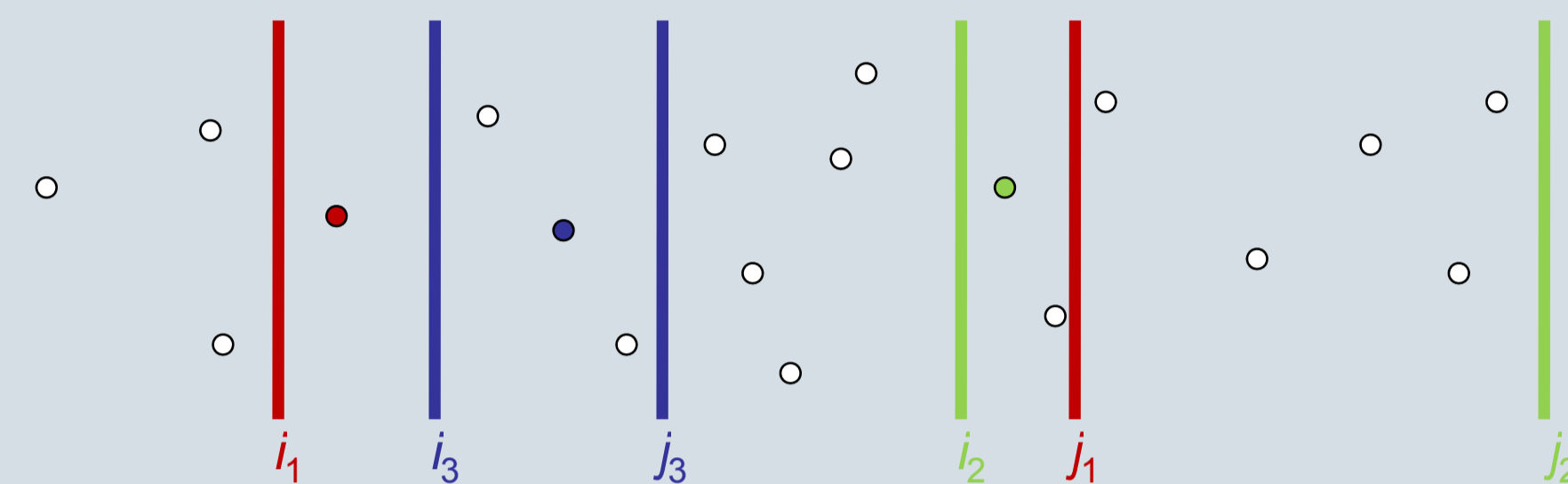
In the static case the point set is fixed, and in the dynamic case points can be inserted and deleted. The objective is to use small space and support efficient range median queries.

Median(i, j) Find the point with median y -value in $[i, j] \times [-\infty, \infty]$.



Batched Range Median

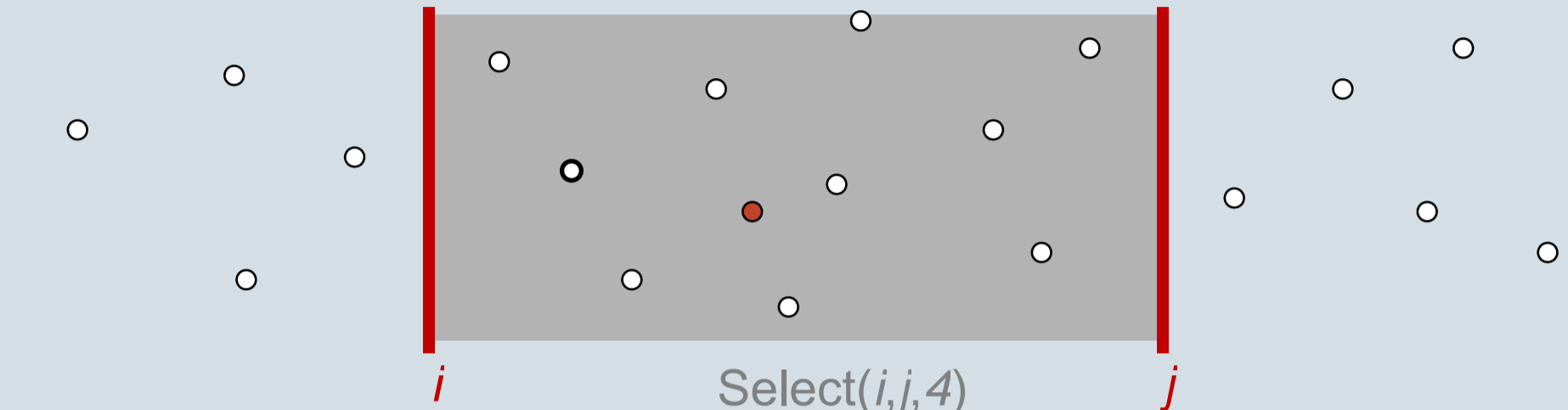
Given a set of n points and a set of k range median queries, return the answer to the k queries.



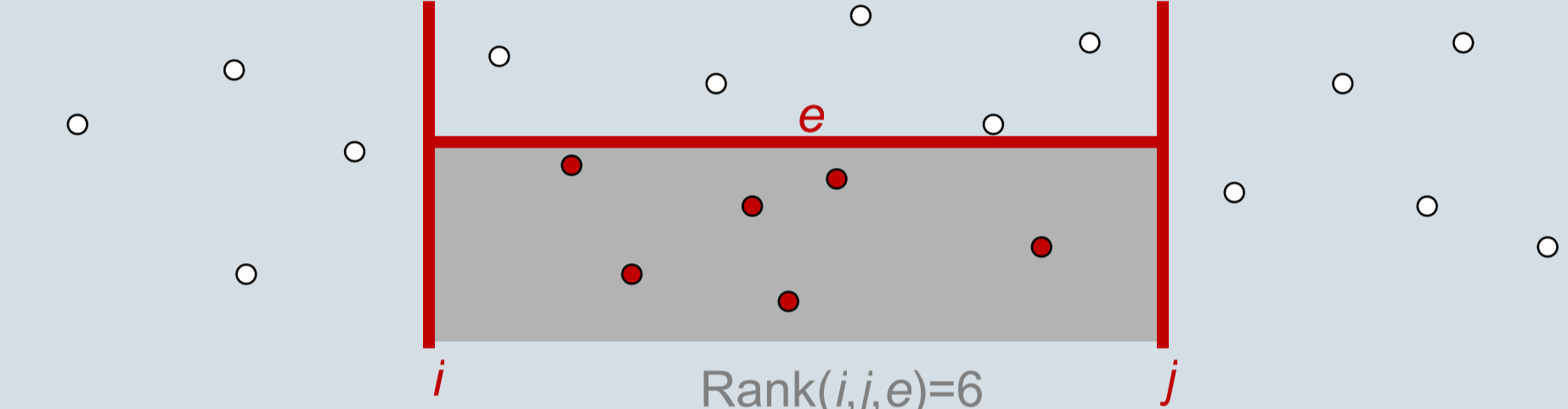
A batched range median query consisting of three queries: $[i_1, j_1], [i_2, j_2], [i_3, j_3]$. Each query and its answer is shown with a distinct color.

Extensions

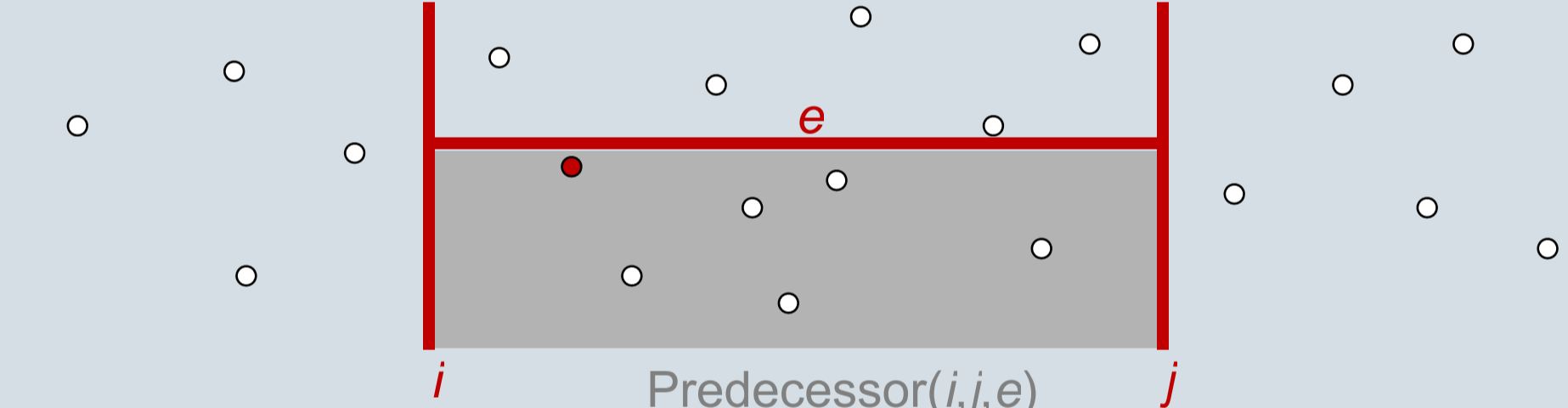
Select(i, j, s) Find the point with s 'th smallest y -value in $[i, j] \times [-\infty, \infty]$.



Rank(i, j, e) Return the number of points within $[i, j] \times [-\infty, e]$.

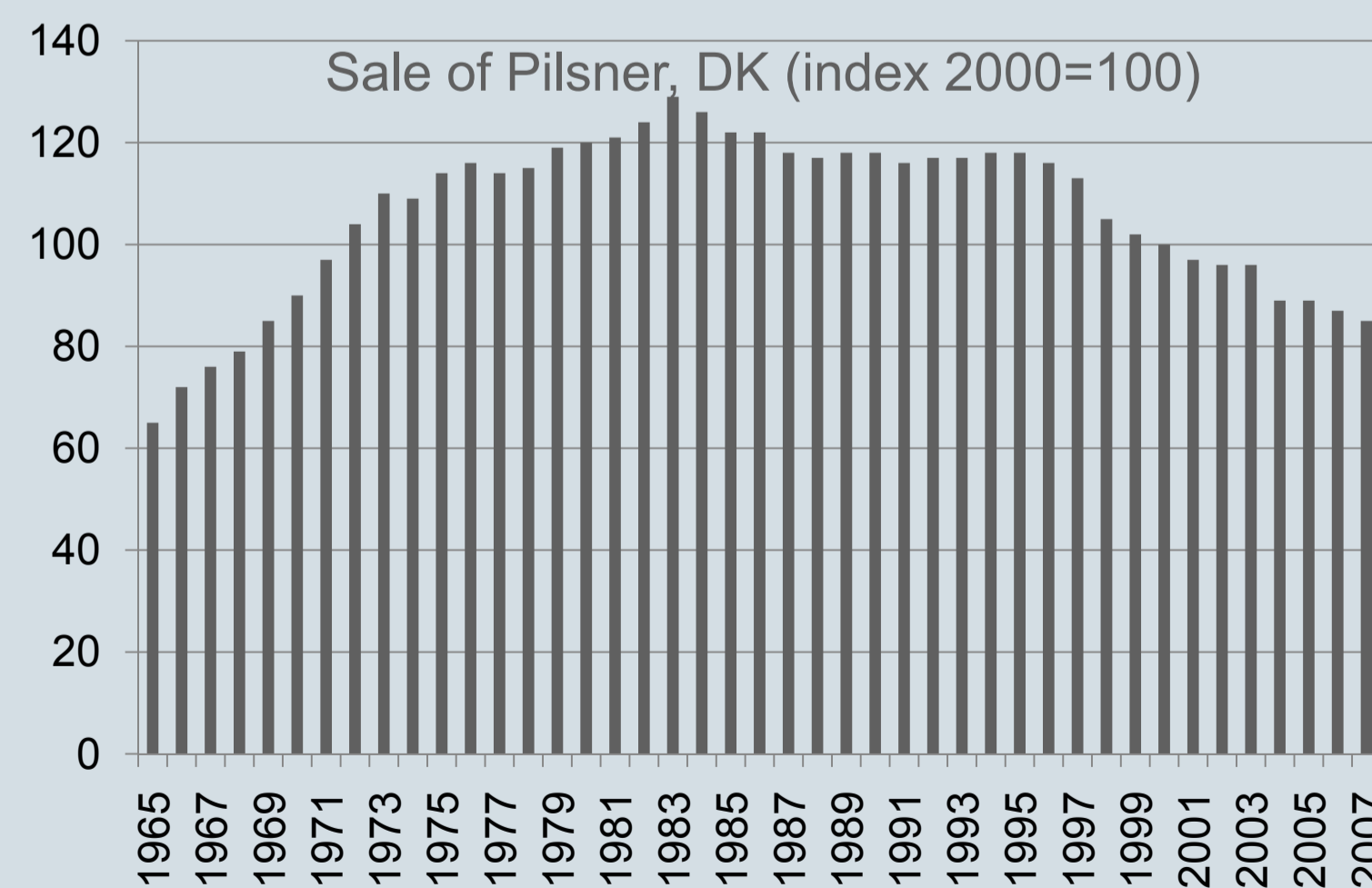


Predecessor(i, j, e) Find point with maximum y -value in $[i, j] \times [-\infty, e]$.



Application: Maintaining Histogram

A range median data structure presents a way of storing a histogram such that interesting queries can be answered efficiently.

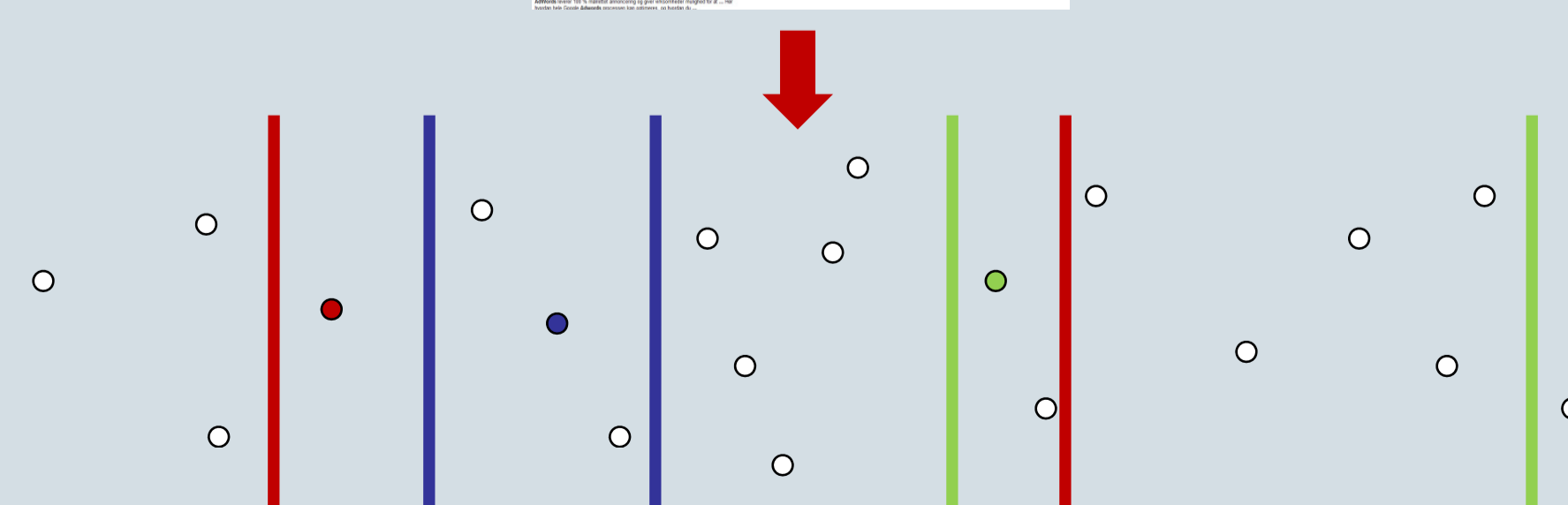
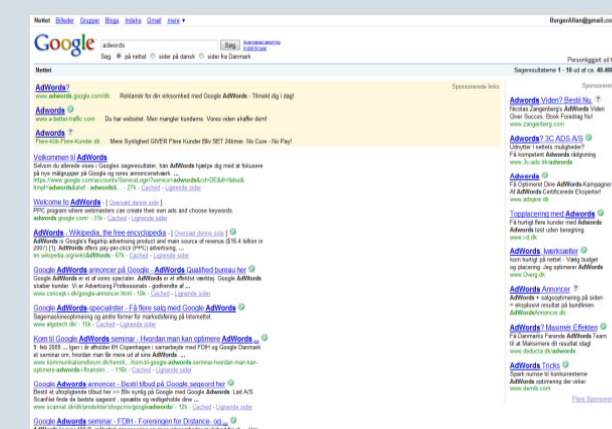


Source: Danmarks Statistik (dst.dk)

In this histogram you can compare your beer consumption with that of the general public over any period of years. You could determine the median beer consumption in your college years and compare with the beer consumption in the college years of for instance your kids.

Application: Analyzing Web Data

A point set can be made from logs of clicks on internet ads. Records contain the time of the click and the price paid by the advertiser. Each advertiser runs several ad campaigns spread over different intervals of time, and wish to compare the price to the general ad market during the campaign. A typical comparison is with the medium price for clicks during the time intervals. This computation corresponds to the **batched range median** problem.



Results

Median*	Space	Query	Updates
Static	$O(n)$	$O(\log n / \log \log n)$	-
Dynamic	$O(n \log n / \log \log n)$	$O((\log n / \log \log n)^2)$	$O((\log n / \log \log n)^2)$

* Bounds also hold for Select/Rank/Predecessor

	Space	Time
Batched Range Median	$O(n)$	$O(n \log k + k \log n / \log \log k)$

References

- [1] Gerth S. Brodal and Allan G. Jørgensen. Range Median – Data Structures and Batched Queries. 2009.