Day 4 – Histogram Details

Distribution table

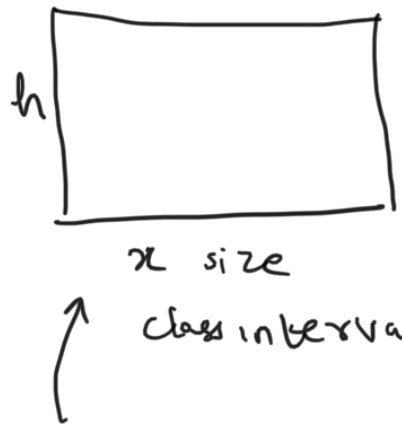
class interval	percentage of population in class interval
0 – 1000	1
1000 – 2000	2
2000 – 3000	3
3000 – 4000	4
4000 – 5000	5
5000 – 6000	5
6000 – 7000	5
7000 – 10,000	15

endpoint convention what to do with population on boundary points do class intervals ?

(eg) left end pant included, right excluded

→ $[a, b)$

First Approximation



$hx \leftrightarrow$ p percent of area

assume p % spread equally across the class interval range

divide up equally

p/x % area

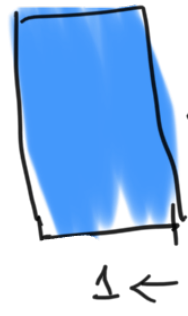
so table recomputed

0 – 1	p_1
1 – 2	p_2
⋮	
7 – 8	
8 – 9	
9 – 10	p_{10}

$\sum_i p_i = 100$

← area = p_i = height h_i

So i^{th} class interval (equal width)

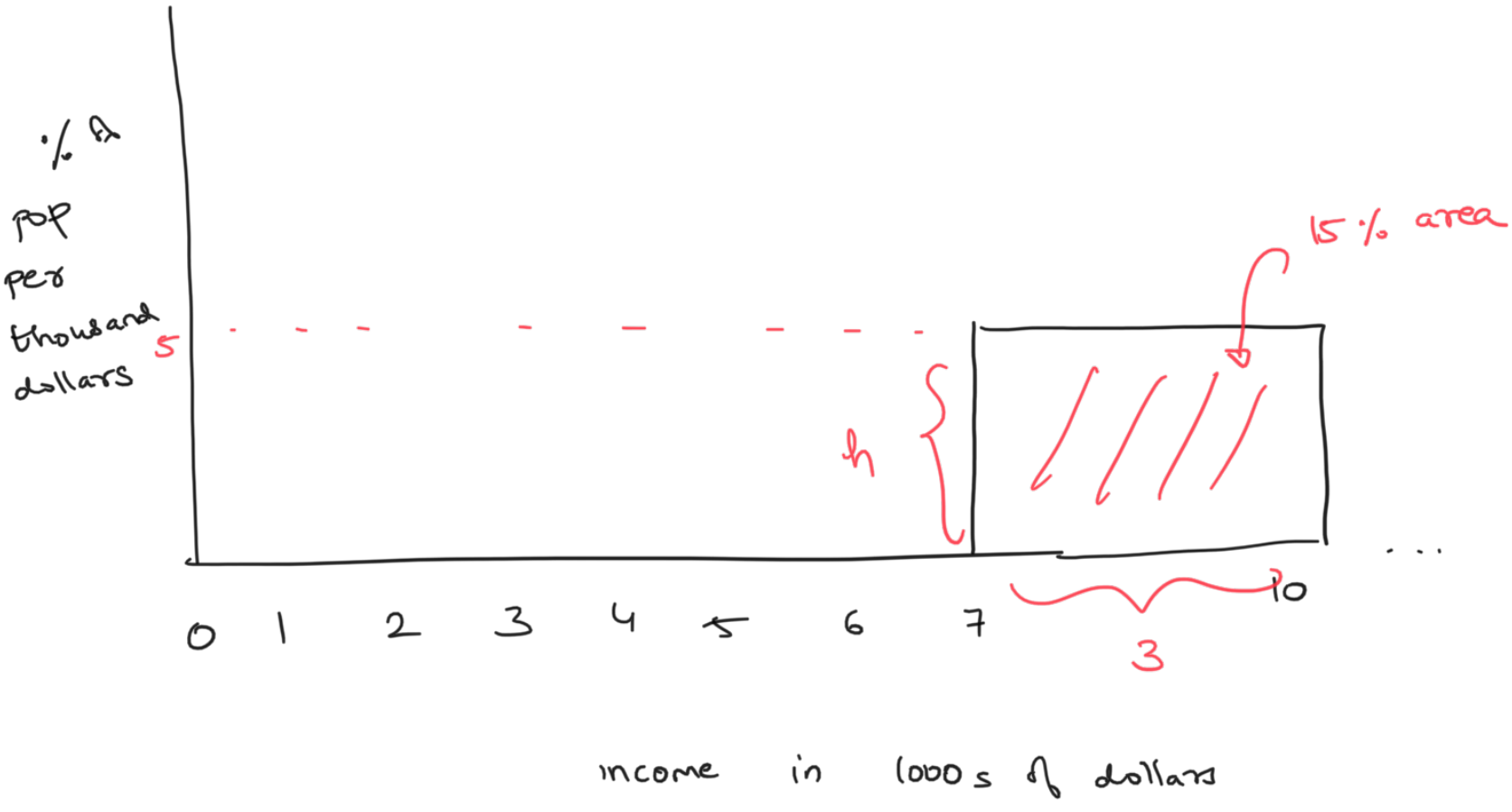


← $h_i = p_i$ %

← width = 1 unit

ht of a block / x-width class interval = $\frac{\text{original \%}}{x}$

in our example



$w = 3$ (1000\$) $\quad p = 15\%$ of pop

$wh = p$

$h = \frac{p}{w} =$ 5% pop / 1000 \$ (unit)

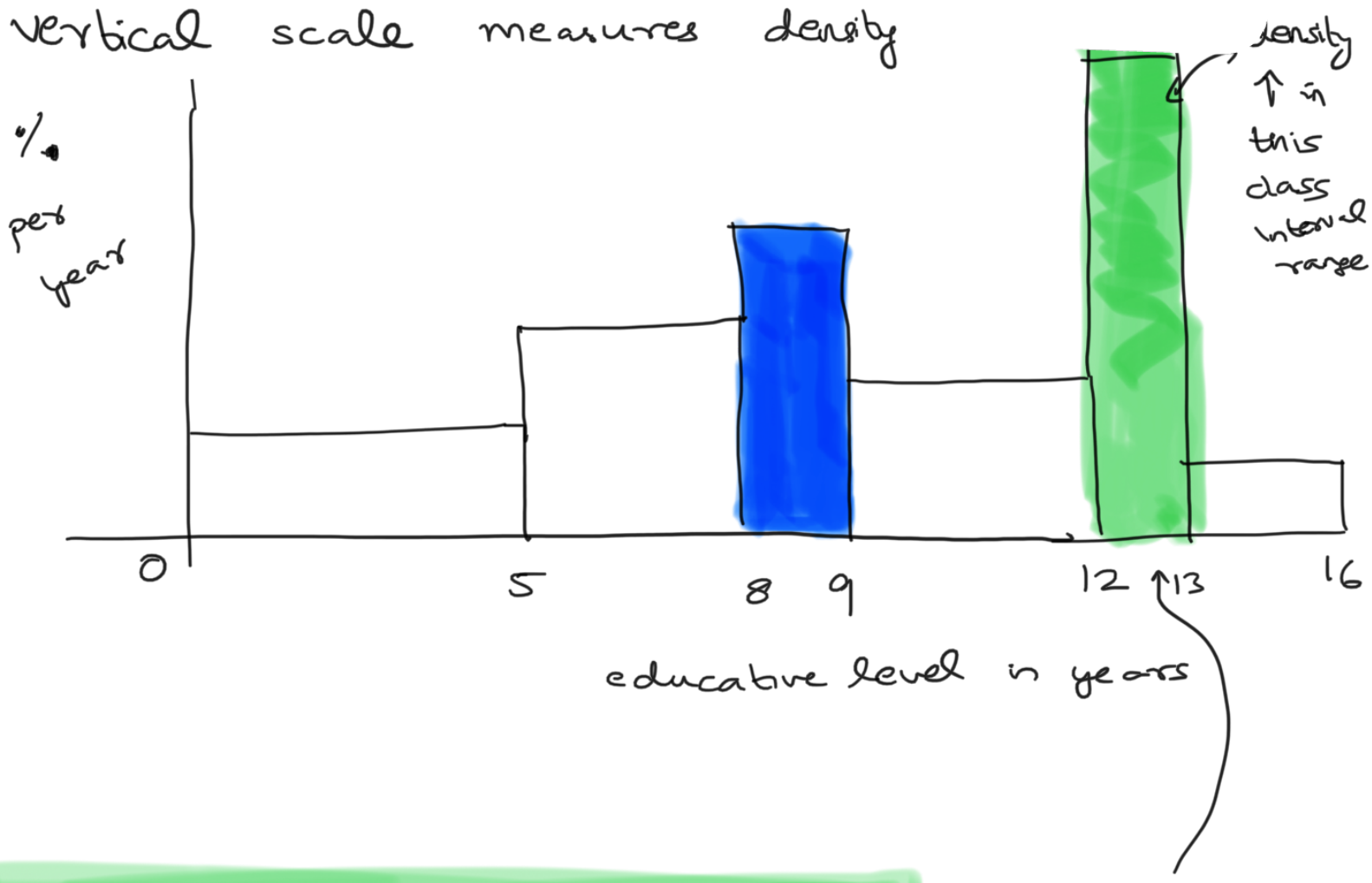
So ht of histogram block over class interval $[7, 10]$ (×1000)

$= 5 \longleftrightarrow$ 5% per 1000 dollars

(ie) for every 1000 dollar interval bet [7000, 10000]

there are 5% of families in that interval.

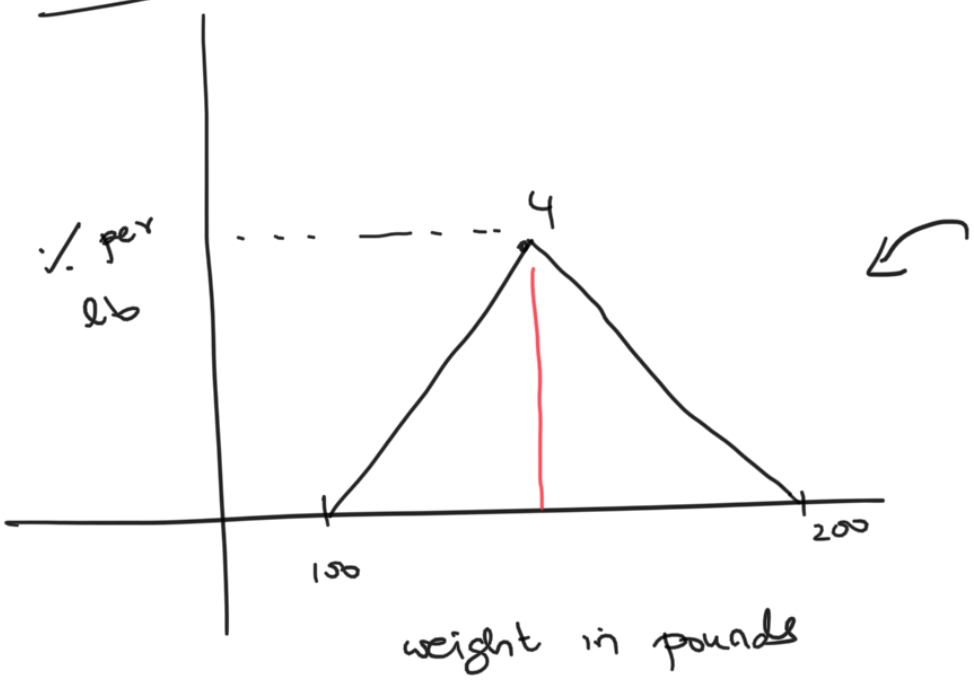
Density



ht of a block represents

crowding — % per horizontal unit

problem!



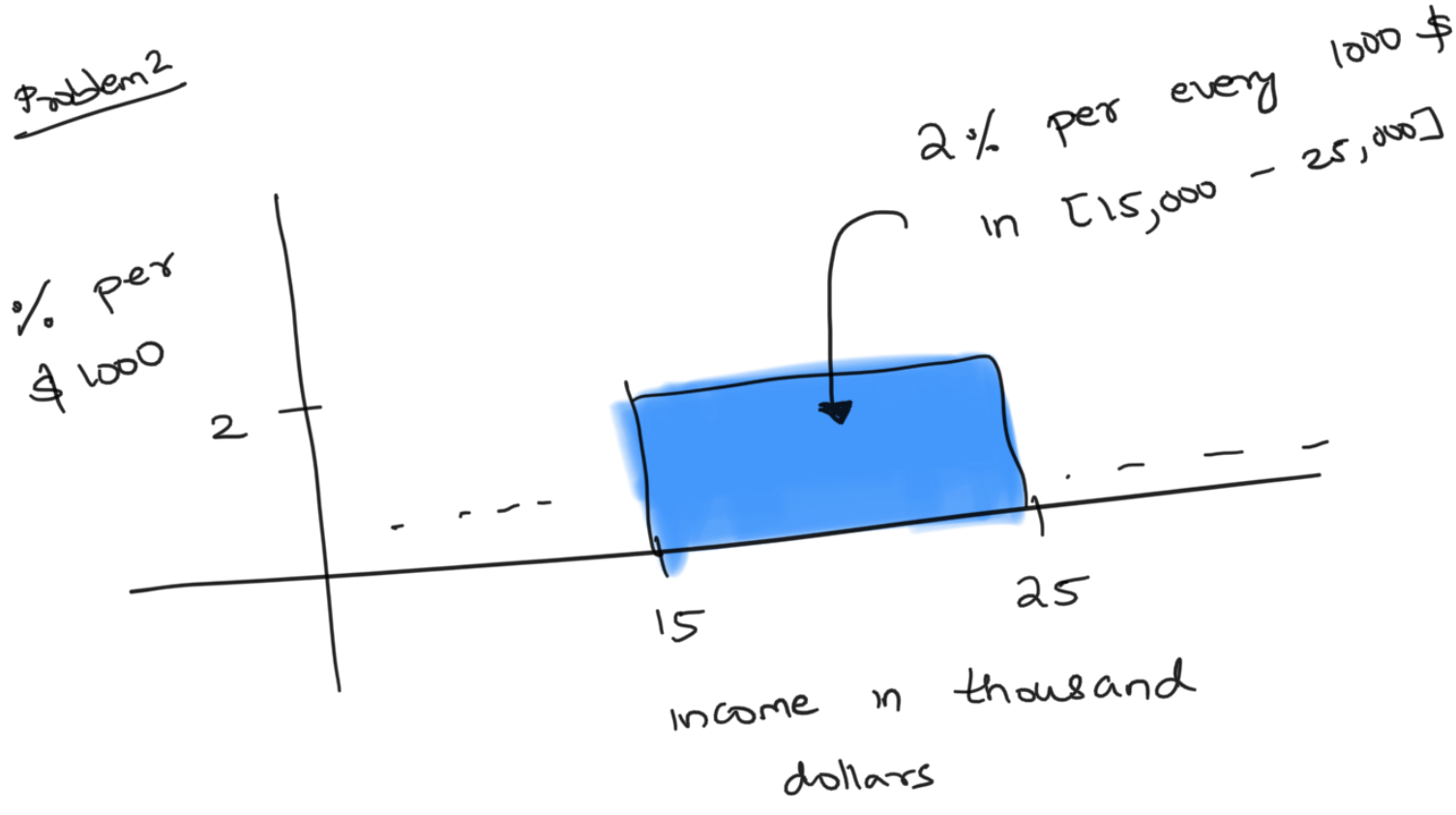
WRONG HISTOGRAM

as area

→ $100 \times \frac{1}{2} \times 4$

$= 200\ \%$

Problem 2



So total % pop in [15000, 25000]

$= 10 \times 2 = 20\%$

2% per $1000 × 10 × $1000

= **20%**