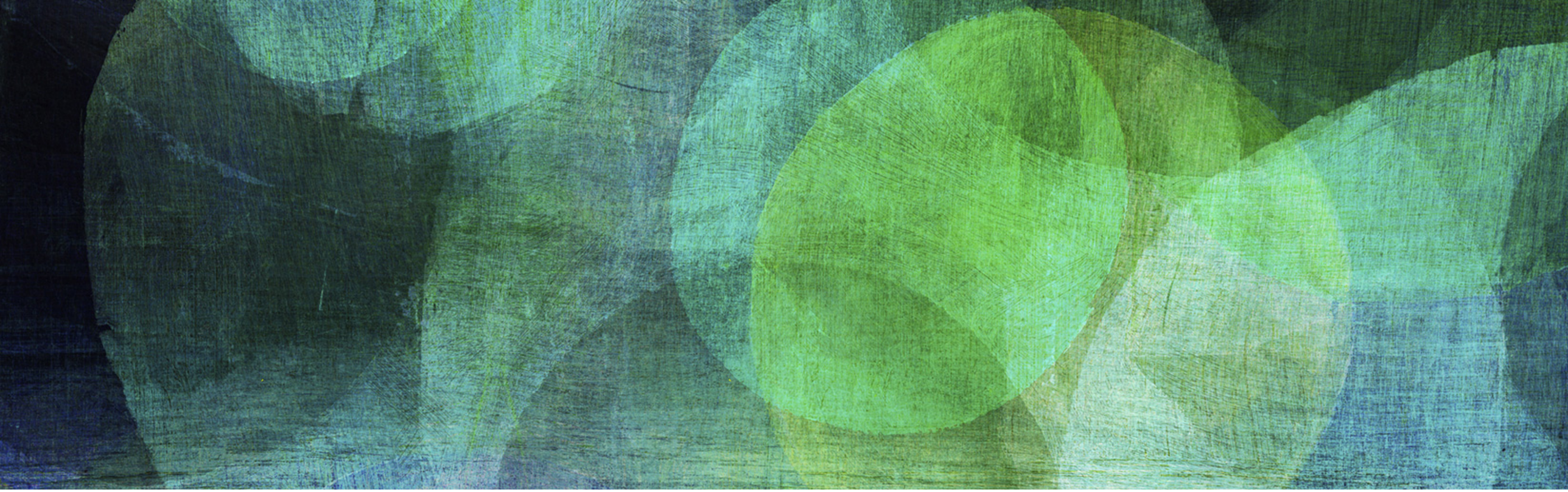
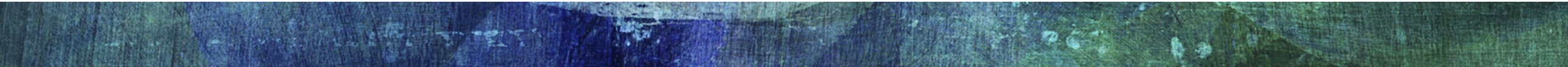


R-TREES: SOLUTION

NDBI007: Practical Class 6



EXERCISE 1

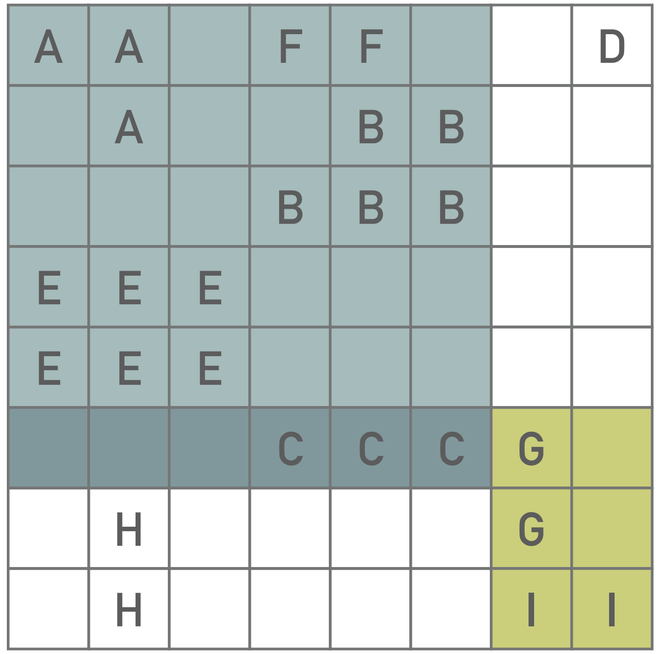
- Finish splitting of the overflown node
  - Continue with Guttman's method
  - The maximum number of items in a node is $M = 8$
  - The minimum number of items in a node is $m = 3$
- If there are more options to choose, explain the reason of yours choice

A	A		F	F			D
	A			B	B		
			B	B	B		
E	E	E					
E	E	E					
			C	C	C	G	
	H					G	
	H					I	I

EXERCISE 1: SOLUTION

- Next, iteratively add such an object into a node which will maximise the difference in the node areas if the object was inserted into the first or second node

Object	ABEF	GI	Difference
C	6x6-30=6	5x3-6=9	|6-9|=3
D	8x5-30=10	2x8-6=10	|10-10|=0
H	6x8-30=18	7x3-6=15	|18-15|=3

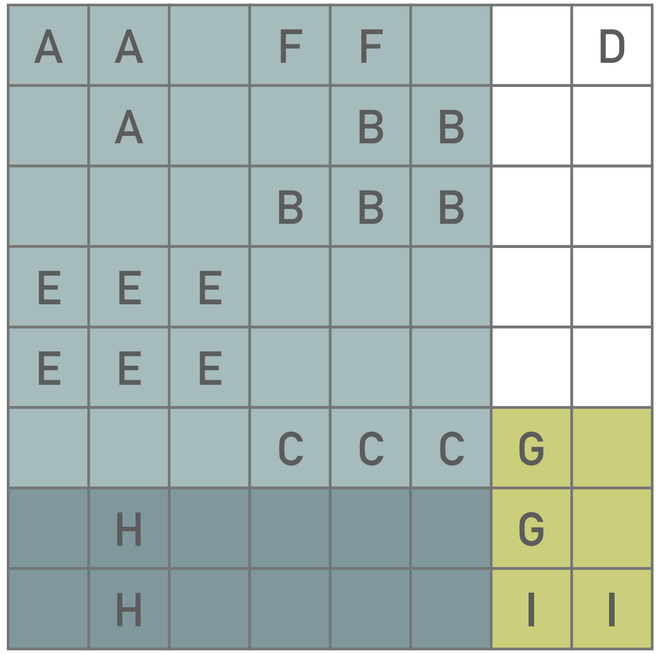


- The biggest difference shows the objects C and H, but we choose C so it will be inserted into the node which is closer, i.e., ABEF
- So now we have nodes ABCEF and GI

EXERCISE 1: SOLUTION

- Next, iteratively add such an object into a node which will maximise the difference in the node areas if the object was inserted into the first or second node

Object	ABCEF	GI	Difference
D	8x6-36=12	2x8-6=10	|12-10|=2
H	6x8-36=12	7x3-6=15	|12-15|=3



- The biggest difference shows the object H so it will be inserted into the node which is closer, i.e., ABCEF
- So now we have nodes ABCEFH and GI

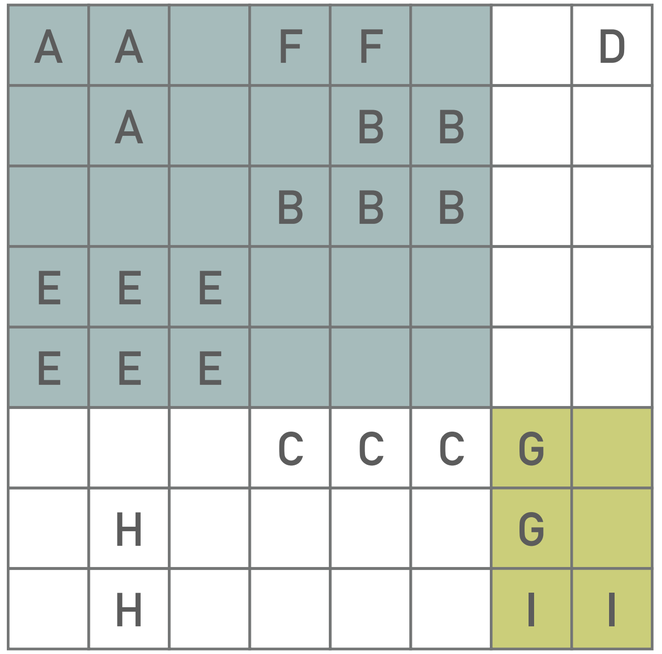
EXERCISE 1: SOLUTION

- Finally, object D must be placed in the node GI because the minimum number of items per node is $m = 3$ and $|GI| = 2$, i.e., $|GI| < m$
  - As a result, we have nodes ABCEFH and DGI

A	A		F	F			D
	A			B	B		
			B	B	B		
E	E	E					
E	E	E					
			C	C	C	G	
	H					G	
	H					I	I

EXERCISE 2

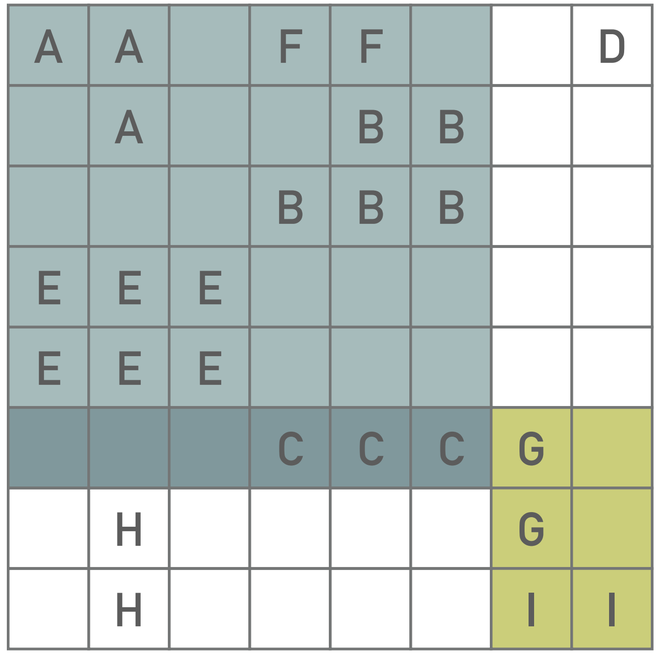
- Finish splitting of the overflown node
  - Continue with Guttman's method
  - The maximum number of items in a node is $M = 8$
  - This time, the minimum number of items in a node is $m = 4$, i.e., $m = M/2$
- If there are more options to choose, explain the reason of yours choice
- Compare and comment the results of exercises 1 and 2



EXERCISE 2: SOLUTION

- Next, iteratively add such an object into a node which will maximise the difference in the node areas if the object was inserted into the first or second node

Object	ABEF	GI	Difference
C	6x6-30=6	5x3-6=9	|6-9|=3
D	8x5-30=10	2x8-6=10	|10-10|=0
H	6x8-30=18	7x3-6=15	|18-15|=3



- The biggest difference shows the objects C and H, but we choose C so it will be inserted into the node which is closer, i.e., ABEF
- So now we have nodes ABCEF and GI

EXERCISE 2: SOLUTION

- Finally, objects H and D must be placed in the node GI because the minimum number of items per node is $m = 4$ and $|GI| = 2$, i.e., $|GI| < m$
  - As a result, we have nodes ABCEF and DGHI
  - There is a smaller death space in ABCEF node but for a price of a huge overlapping area, therefore it is already better to use smaller value of m in this particular case

A	A		F	F			D
	A			B	B		
			B	B	B		
E	E	E					
E	E	E					
			C	C	C	G	
	H					G	
	H					I	I

EXERCISE 3

- Split the following overflown node with Greene's split method
  - The maximum number of items in a node is $M = 9$
  - The minimum number of items in a node is $m = 3$

G			A			I	I				J
G		A	A	A							
			A					C			
	F							C	C		
	F				H	H		C			
	F										
				E	E	E		B	B	B	
D	D	D							B		

- I.e., execute the following methods:
  - PickSeeds
  - ChooseAxis
  - Distribute (ordering and placement)

EXERCISE 3: SOLUTION

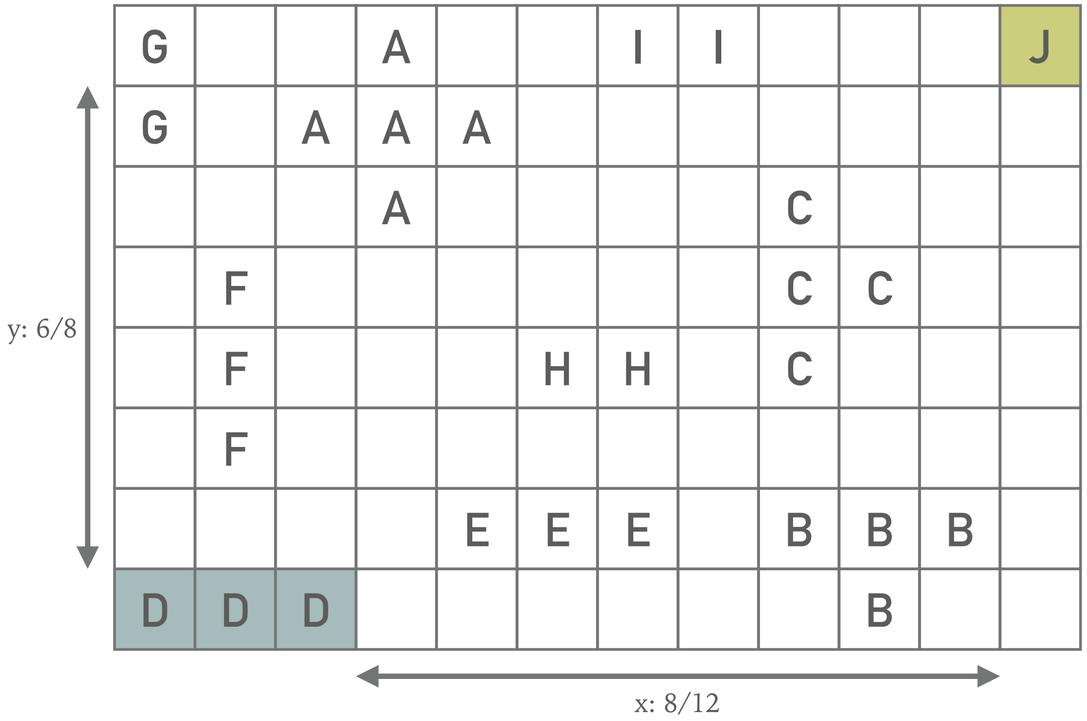
- PickSeeds
  - The largest dead space has DJ thus those will be the seeds of the splitting method

Pair	Overall area	Area of the objects	Dead space
AB	9x8=72	5+4=9	72-9=63
AC	8x5=40	5+4=9	40-9=31
…			
BG	11x8=88	4+2=6	88-6=82
…			
DJ	12x8=96	3+1=4	96-4=92
…			
IJ	6x1=6	2+1=3	6-3=3

G			A			I	I				J
G		A	A	A							
			A					C			
	F							C	C		
	F				H	H		C			
	F										
				E	E	E		B	B	B	
D	D	D							B		

EXERCISE 3: SOLUTION

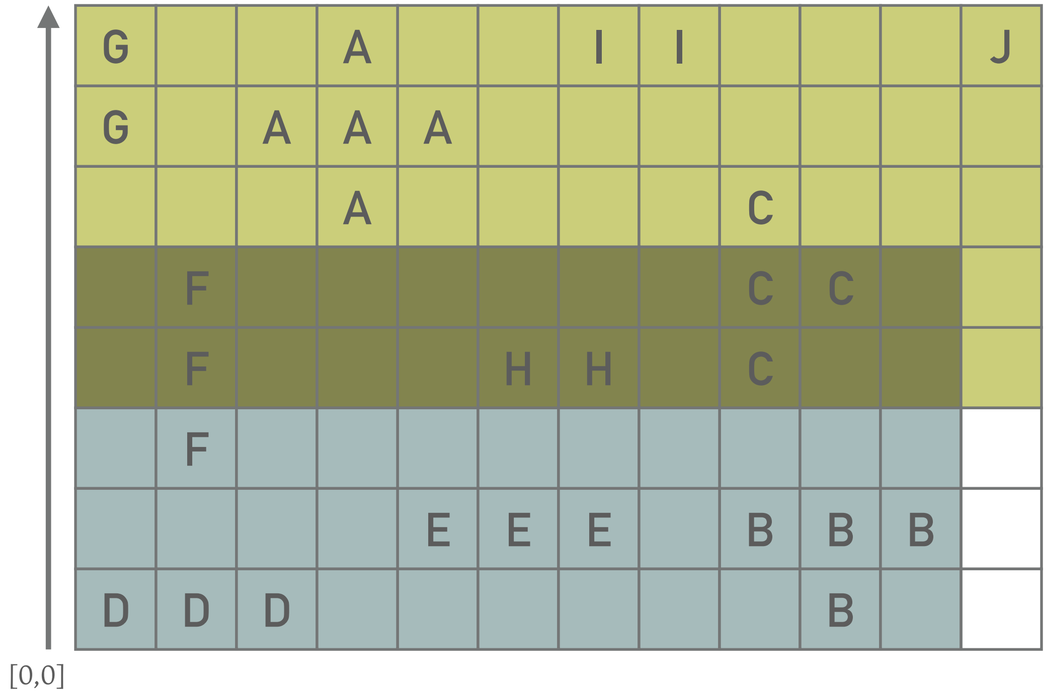
- ChooseAxis
  - x: 8/12 = 0.667
  - y: 6/8 = 0.750
  - In our case, the axis better separating D and J is y



EXERCISE 3: SOLUTION

- Distribute according to axis y

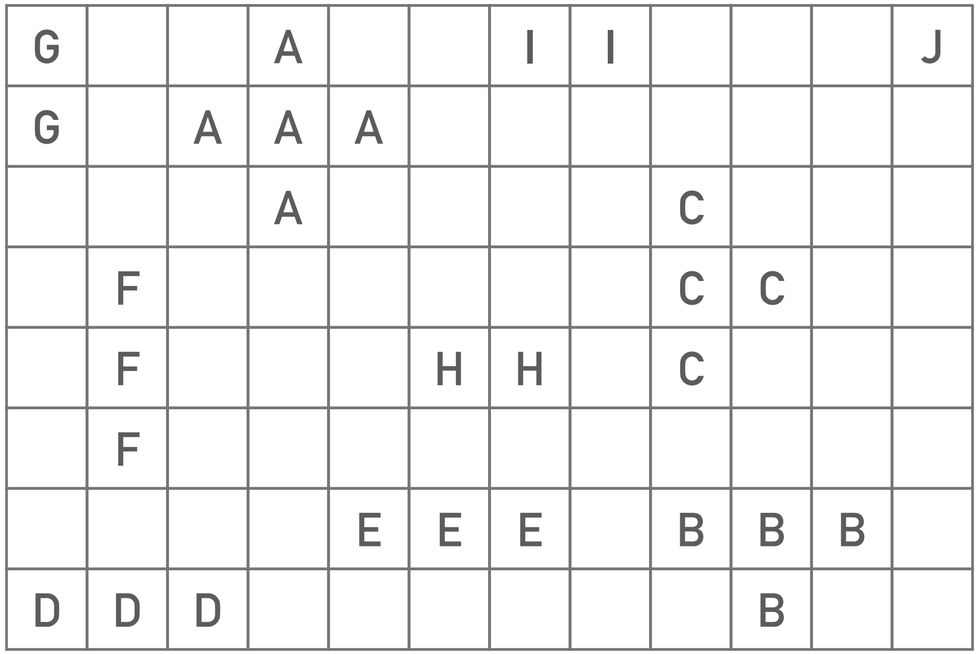
Object	D	B	E	F	H	C	A	G	I	J
Start	0	0	1	2	3	3	5	6	7	7
end	0	1	1	4	3	5	7	7	7	7



- The solution:
  - BDEFH || ACGIJ

EXERCISE 4

- Split the following overflown node with R*Tree split method
  - The maximum number of items in a node is $M = 9$
  - The minimum number of items in a node is $m = 3$

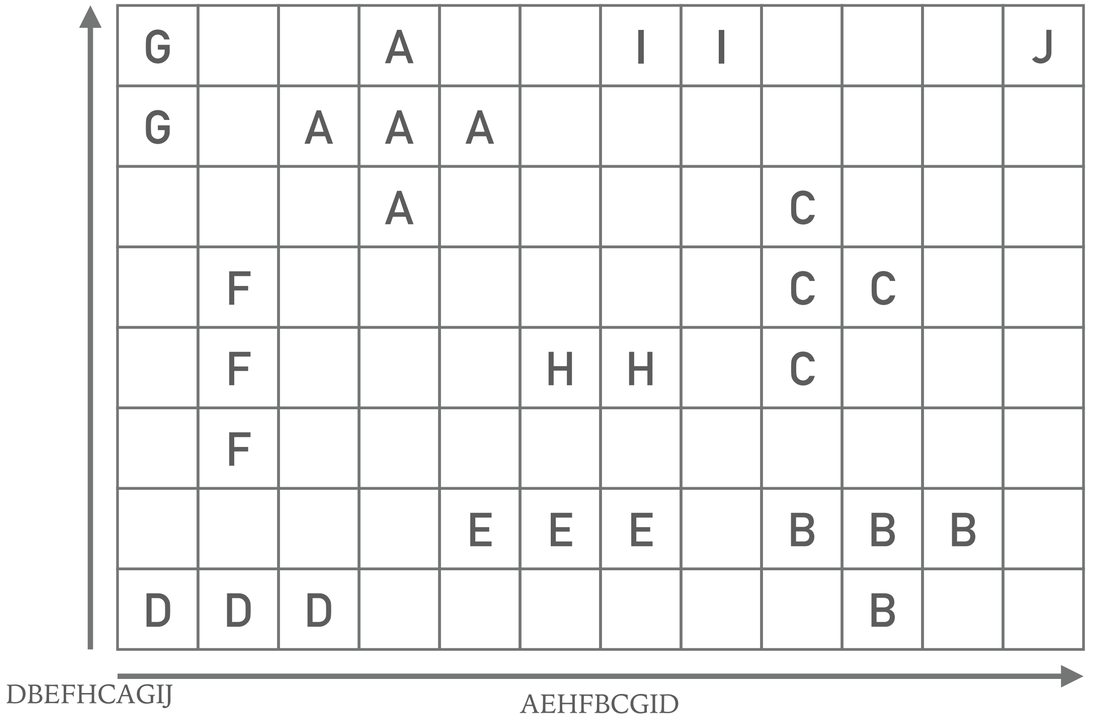


- I.e., execute the following methods:
  - ChooseSplitAxis
  - Distribute
  - Illustrate the result

EXERCISE 4: SOLUTION

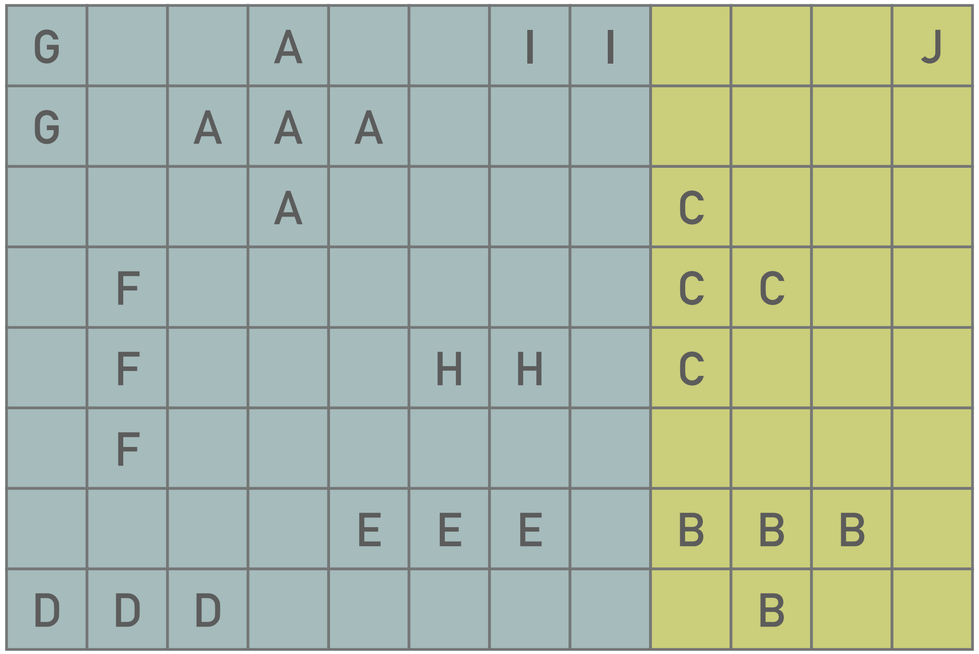
- Ordering* based on the x-axis: GDFAEHICBJ
  - margin-value (GDF||AEHICBJ) = (3+8)*2+(10+8)*2=22+36=58
  - margin-value (GDFA||EHICBJ) = (5+8)*2+(8+8)*2=26+32=58
  - margin-value (GDFAE||HICBJ) = (7+8)*2+(7+8)*2=30+30=60
  - margin-value (GDFAEH||ICBJ) = (7+8)*2+(6+8)*2=30+28=58
  - margin-value (GDFAEHI||CBJ) = (8+8)*2+(4+8)*2=32+24=56
  - Sum = 58+58+60+58+56=290

- Ordering* based on the y-axis: DBEFHCAGIJ
  - margin-value (DBE||FHCAGIJ) = (11+2)*2+(12+6)*2=26+36=62
  - margin-value (DBEF||HCAGIJ) = (11+5)*2+(12+5)*2=32+34=66
  - margin-value (DBEFH||CAGIJ) = (11+5)*2+(12+5)*2=32+34=66
  - margin-value (DBEFHC||AGIJ) = (11+6)*2+(12+3)*2=34+30=64
  - margin-value (DBEFHCA||GIJ) = (11+8)*2+(12+2)*2=38+24=62
  - Sum = 62+66+66+64+62=320



EXERCISE 4: SOLUTION

- We chose splitting along the x-axis (smaller sum)
  - overlap-value (GDF||AEHICBJ) = 8 (column AD)
  - overlap-value (GDFA||EHICBJ) = 8 (column AE)
  - overlap-value (GDFAE||HICBJ) = 16 (columns HE; IHE)
  - overlap-value (GDFAEH||ICBJ) = 8 (column IHE)
  - overlap-value (GDFAEHI||CBJ) = 0



- There is only one distribution having the smallest overlap, therefore the area-value does not have to be computed

- The result is: GDFAEHI||CBJ