Semantrix: A Compressed Semantic Matrix

Nieves R. Brisaboa, Antonio Fariña, Gonzalo Navarro
and Tirso V. Rodeiro

citic

UNIVERSIDADE DA CORUÑA

Instituto Milenio
Fundamentos
de los datos

Route map

- Introduction
- The problem
- Our proposal
- Experiments
- Future work

Route map

- **Introduction**
- The problem
- Our proposal
- Experiments
- Future work

1st stop: Introduction

Introduction

Introduction

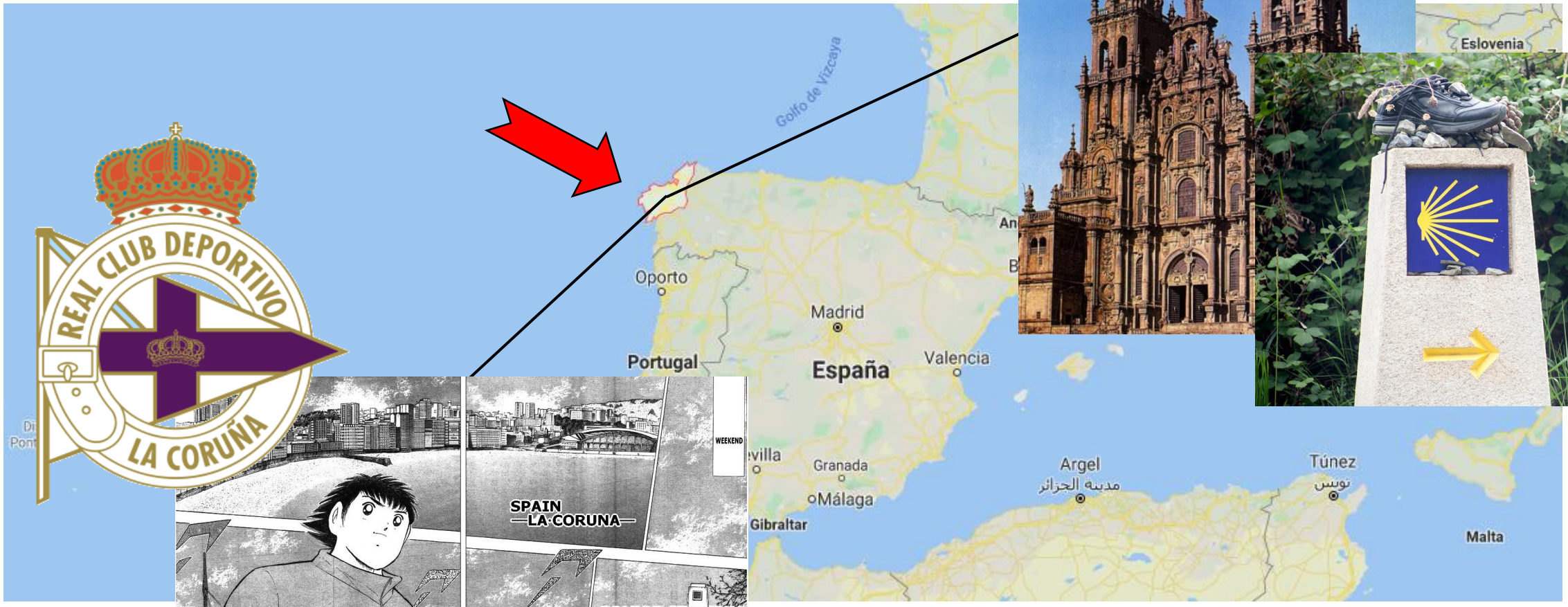
Introduction

Introduction

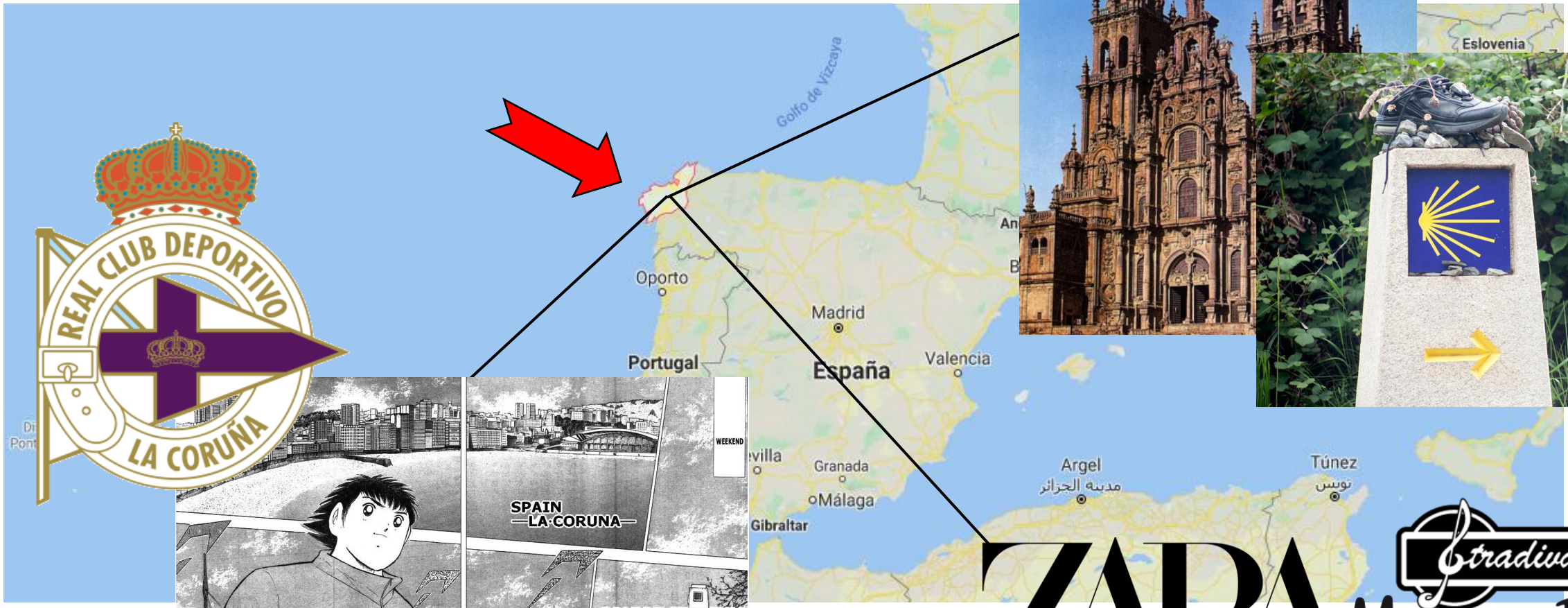
Introduction

Introduction

Introduction

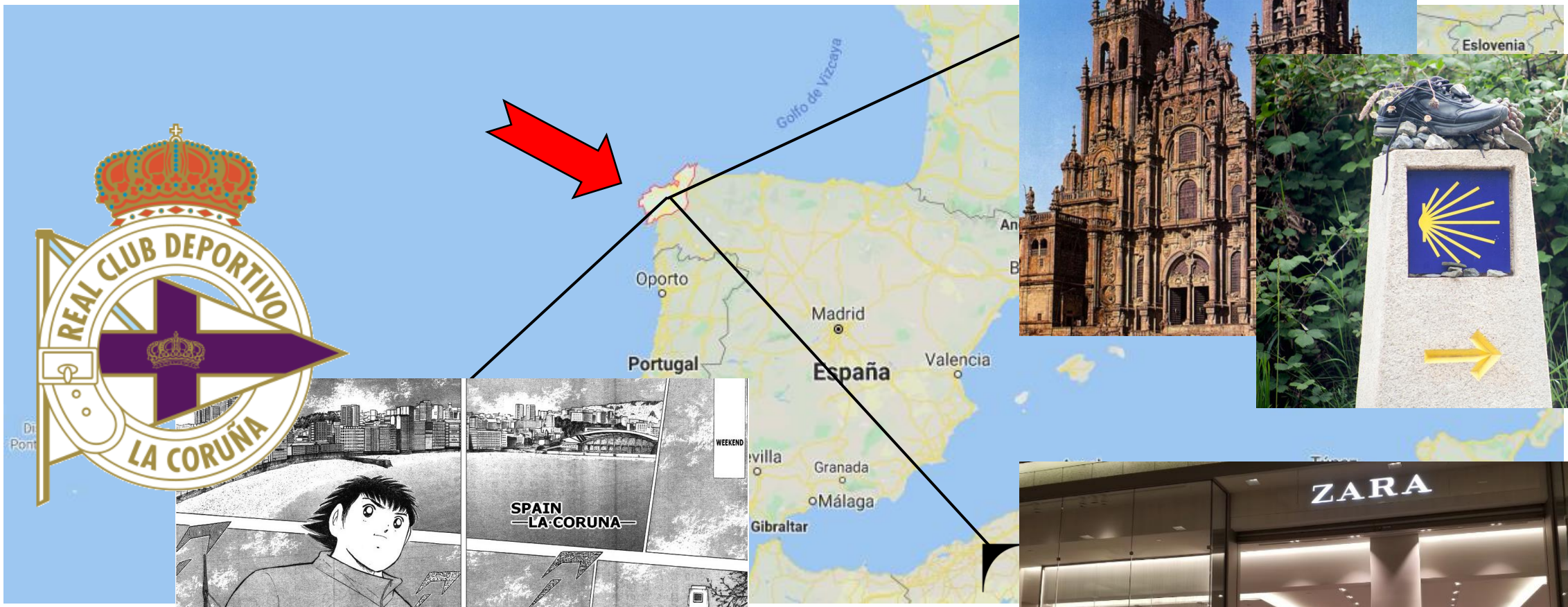


Introduction



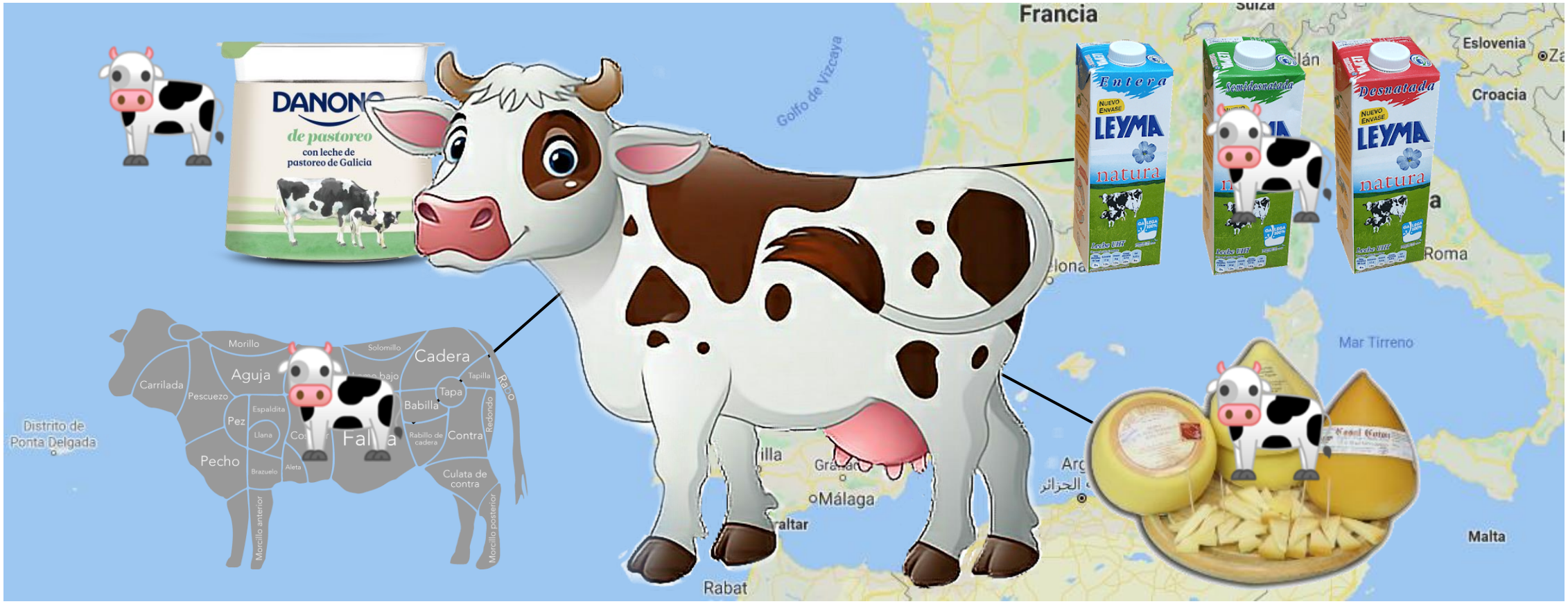
ZARA
Stradivarius
Massimo Dutti
oysho
Bershka
PULL&BEAR

Introduction

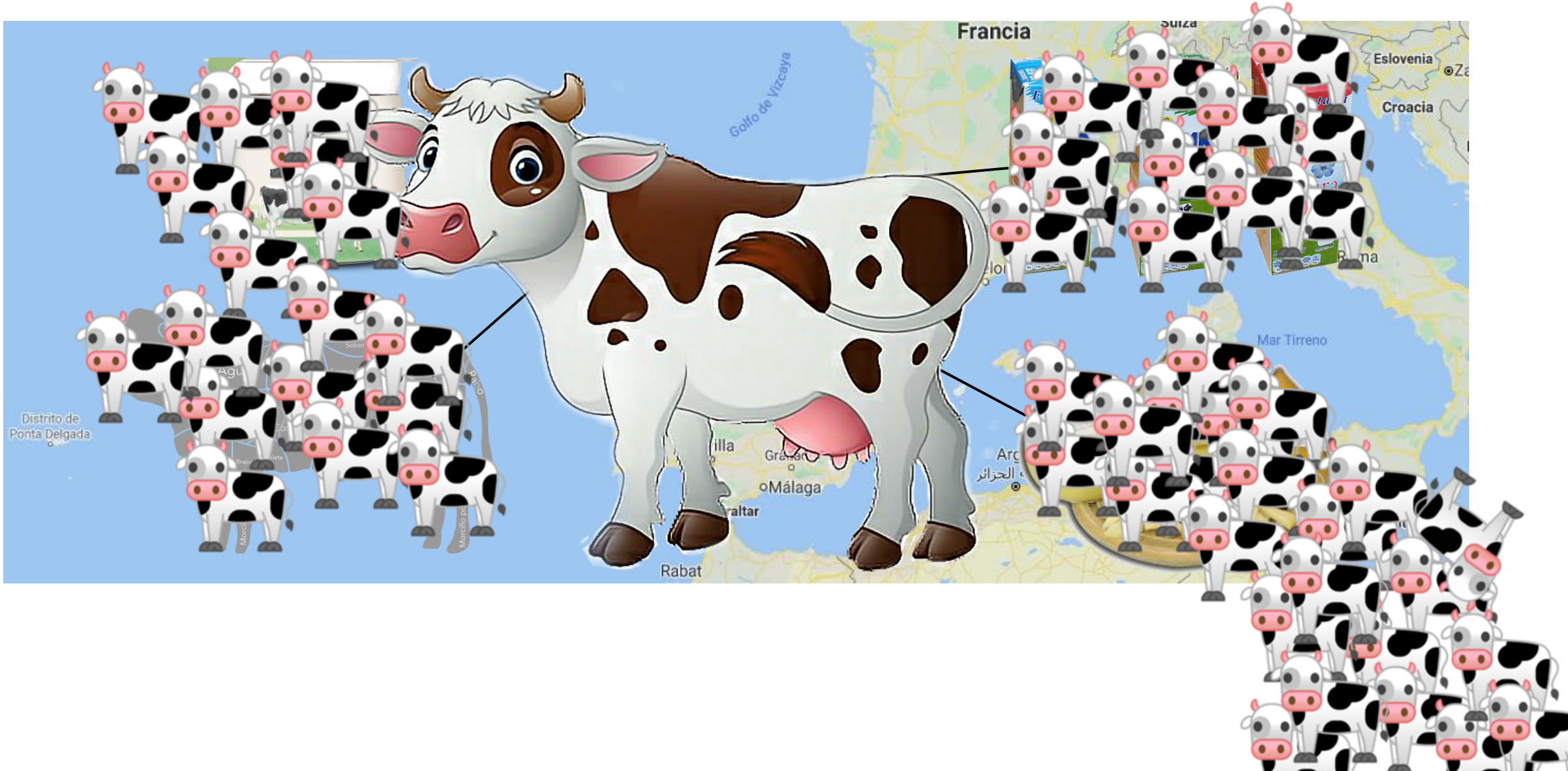


Introduction

Introduction



Introduction



Introduction

Introduction

- Early 2000s ➡ **Bovine spongiform encephalopathy** (**BSE**)

  [mad cow disease]
- Strict waste management protocol.
- Only official companies are allowed to do the job.
- Here is where we come in!
- Efficient process ➡ In this episode: Driver actions and trajectories.

Route map

- Introduction
- **The problem**
- Our proposal
- Experiments
- Future work

2nd stop: The problem

The problem: Introduction

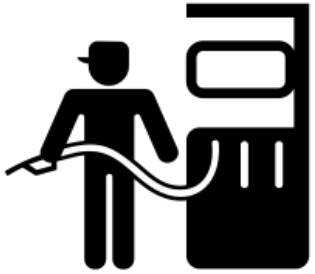
Where do we come from?

- Nieves R. Brisaboa, Miguel R. Luaces, Cristina Martínez Pérez, Ángeles Saavedra Places: ***Semantic Trajectories in Mobile Workforce Management Applications***. W2GIS 2017: 100-115

- They are able to break a trajectory into several segments and identify what activity was doing the truck driver during each segment.

The problem: What can a driver do?

1) Being at headquarters.
2) Working at a customer place.
3) Normal transit on planned route.
4) Slow transit on planned route.
5) Normal transit out of planned route.
6) Slow transit out of planned route.
7) Taking a break.
8) Unknown activity.
9) Inactive.



The problem: Where are we going?

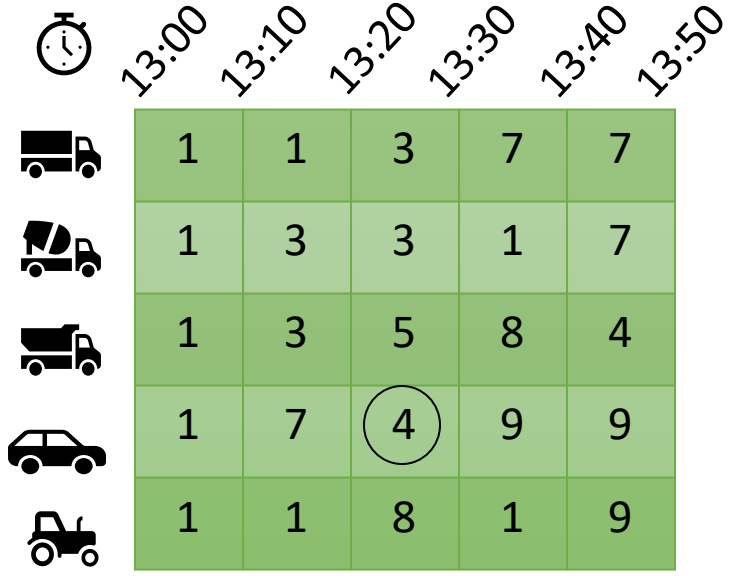
Where do we come from?

- Nieves R. Brisaboa, Miguel R. Luaces, Cristina Martínez Pérez, Ángeles Saavedra Places: ***Semantic Trajectories in Mobile Workforce Management Applications***. W2GIS 2017: 100-115

Where are we going?

- Create a compressed representation of a set of semantic trajectories/activities in such a way that we could still answer different relevant queries efficiently.

The problem: A naïve approach

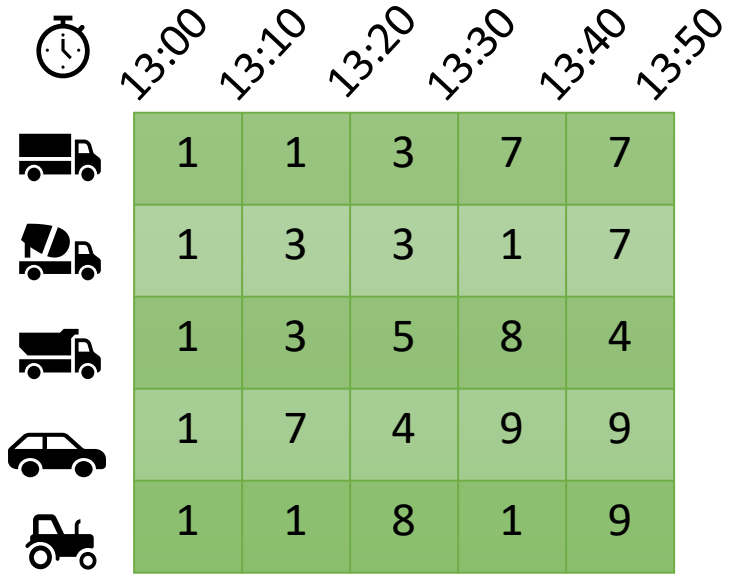


- **Columns**: Discretization of the time.
- **Rows**: Moving objects.
- **Cell:** ID of an activity.

- **Example**: the car was performing the activity with ID 4 from 13:20h to 13:30h (*slow transit on planned route*).

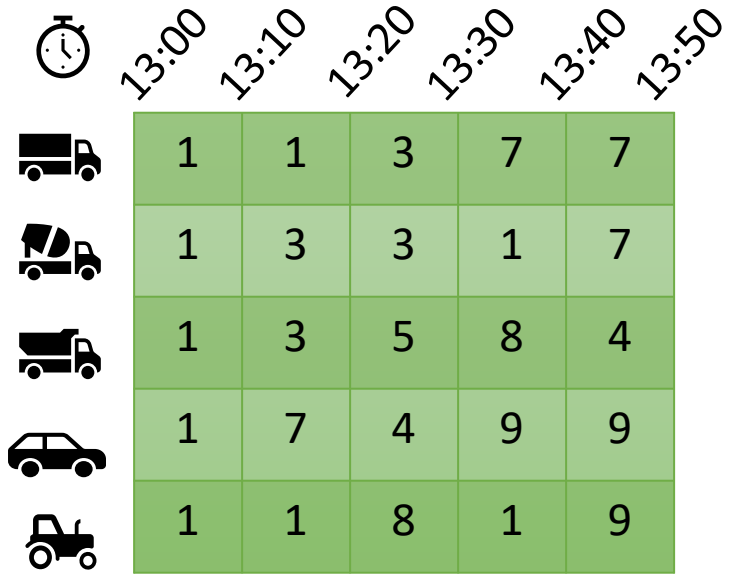
The problem: Queries

- **Individual queries**
- **Pattern queries**
- **Aggregated queries**



The problem: Queries

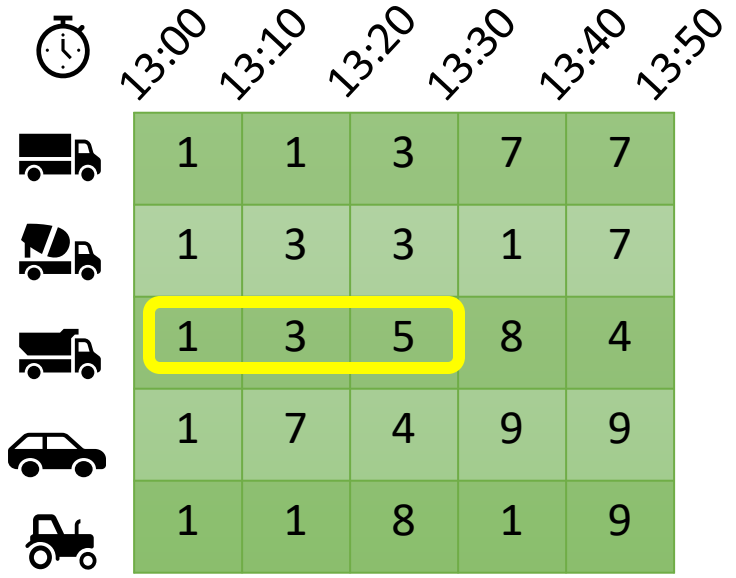
- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries**
- **Aggregated queries**



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	

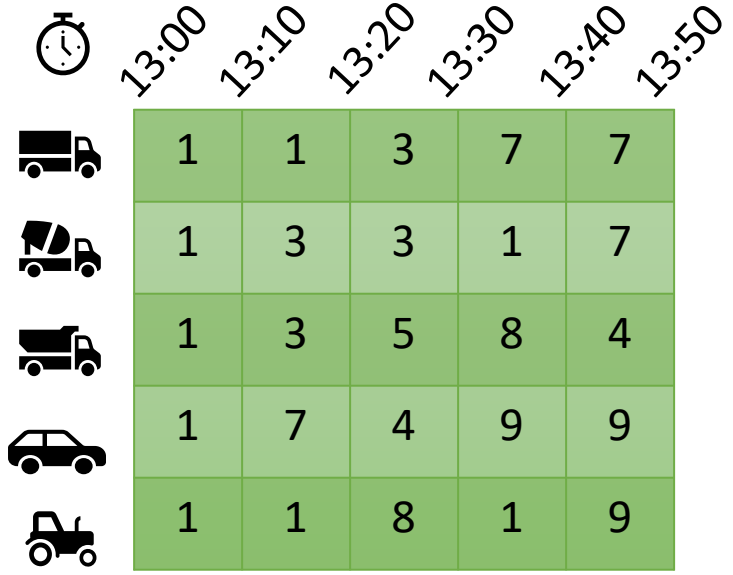
The problem: Queries

- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries**
- **Aggregated queries**



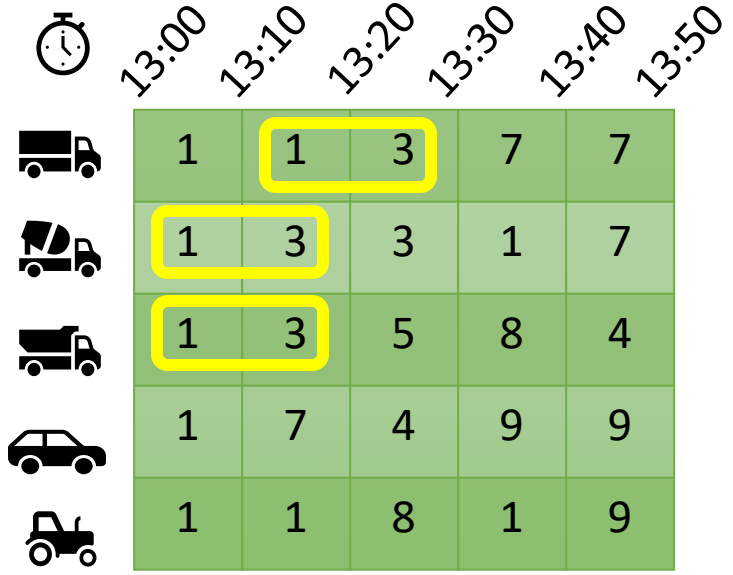
The problem: Queries

- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries:** *How many times activity 1 was followed by activity 3?*
- **Aggregated queries**



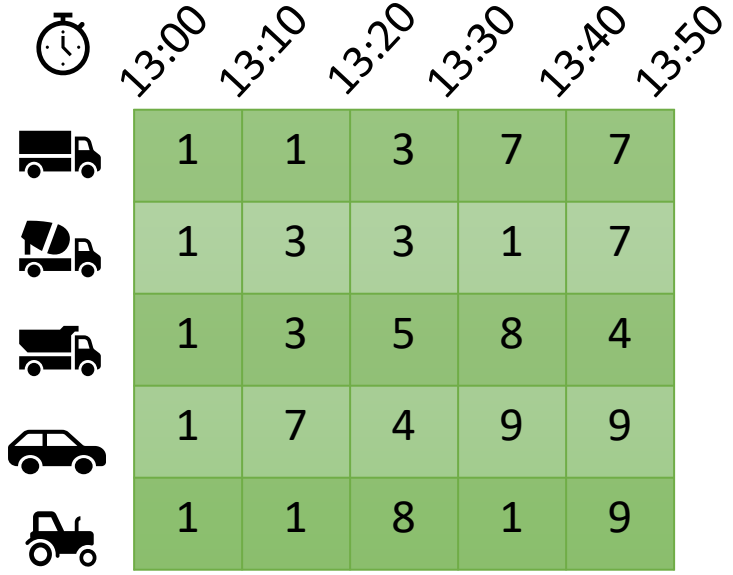
The problem: Queries

- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries:** *How many times activity 1 was followed by activity 3?*
- **Aggregated queries**



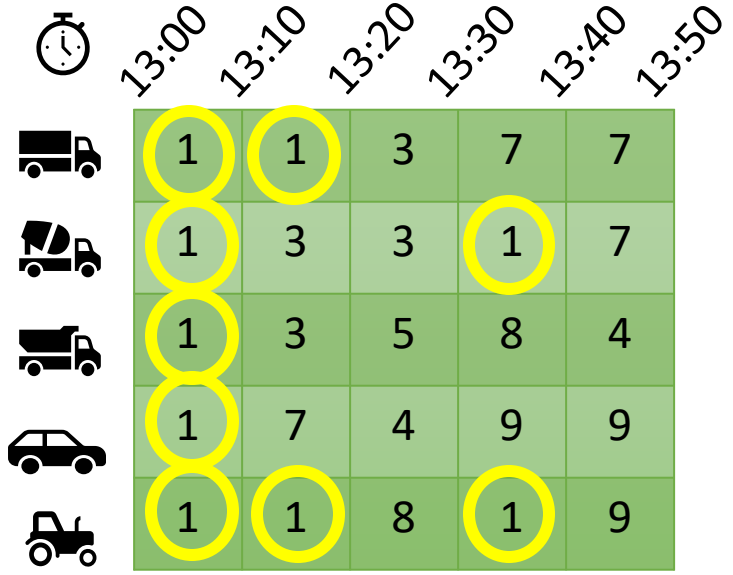
The problem: Queries

- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries:** *How many times activity 1 was followed by activity 3?*
- **Aggregated queries:** *How much time was actually spent at headquarters (ID: 1) by all the mobile objects?*



The problem: Queries

- **Individual queries:** *Which is the list of activities performed by a given driver between 13:00 and 13:30?*
- **Pattern queries:** *How many times activity 1 was followed by activity 3?*
- **Aggregated queries:** *How much time was actually spent at headquarters (ID: 1) by all the mobile objects?*

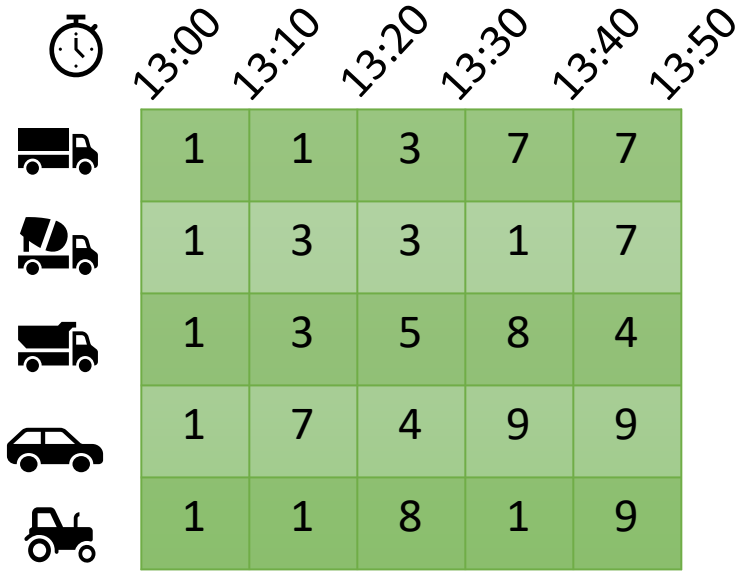


Route map

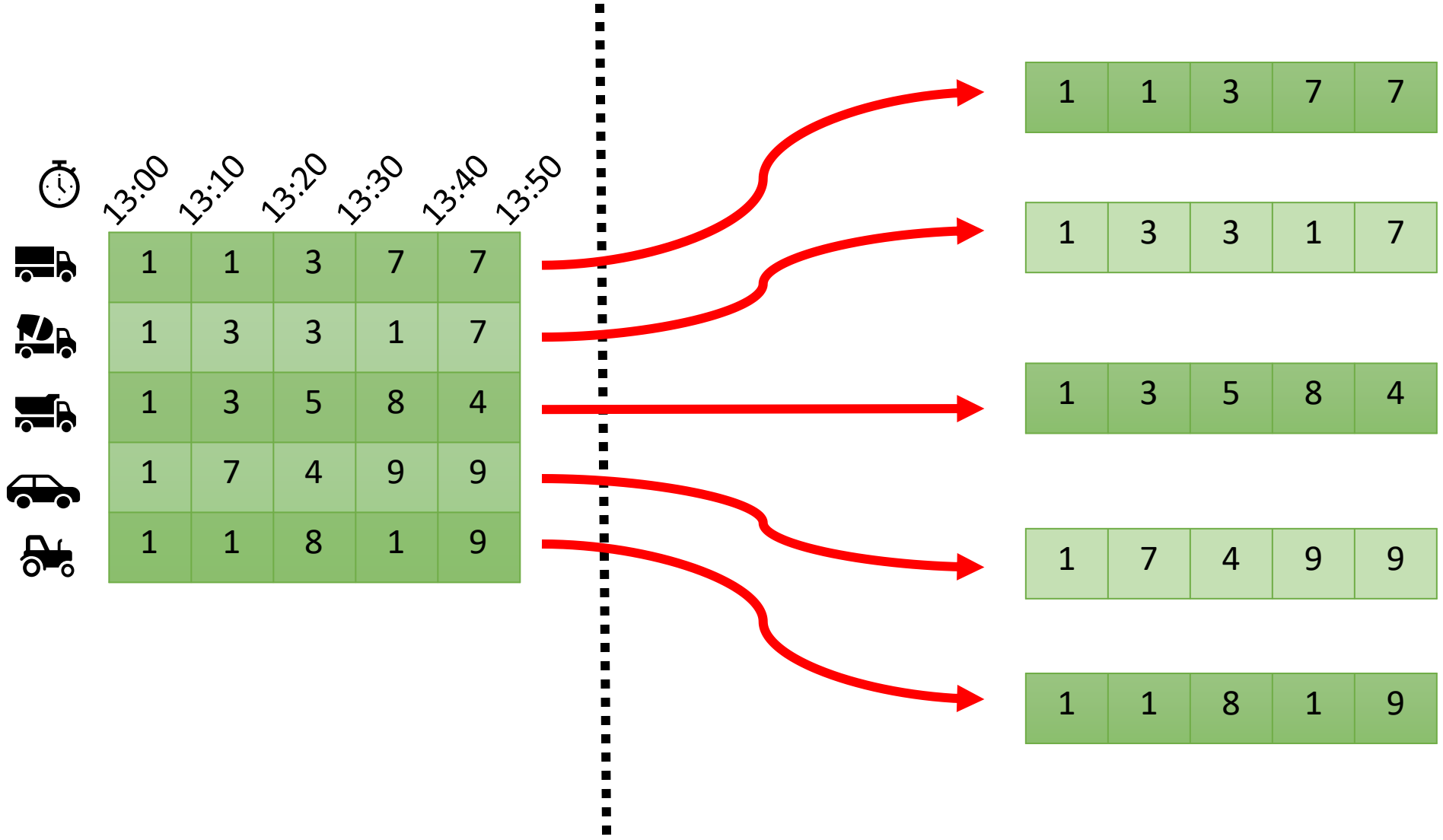
- Introduction
- The problem
- **Our proposal**
- Experiments
- Future work

3rd stop: Our proposal

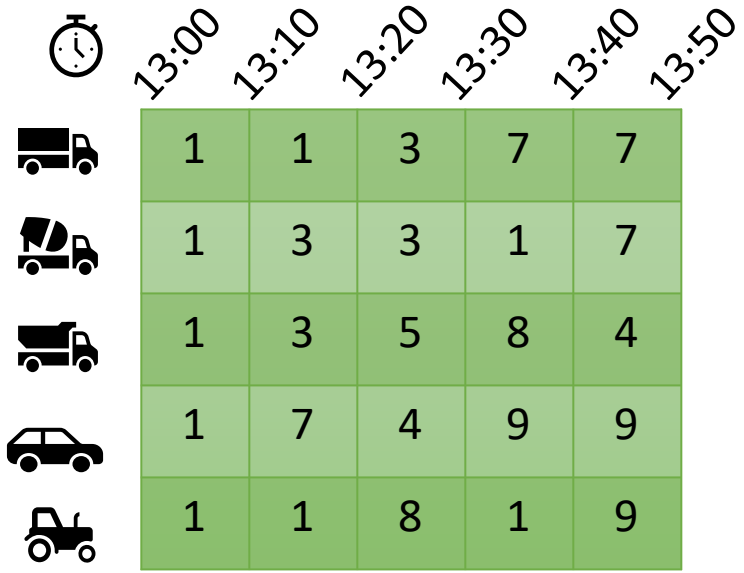
Our proposal: Improving the naïve approach



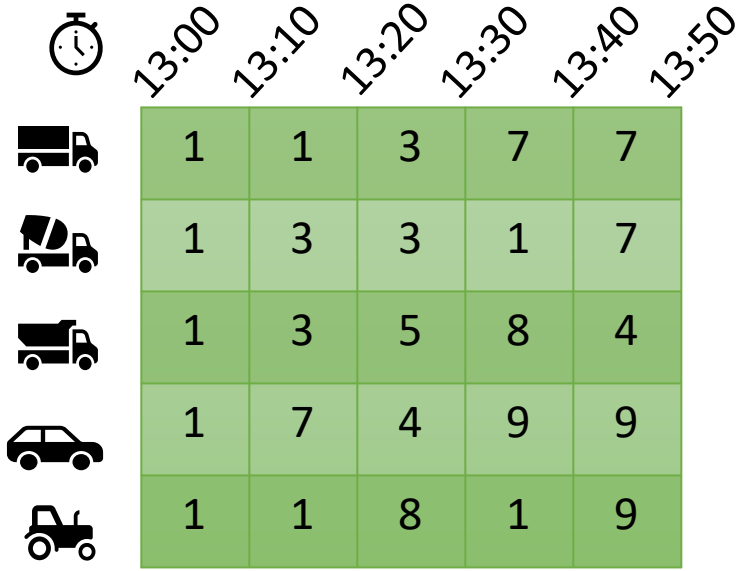
Our proposal: Improving the naïve approach

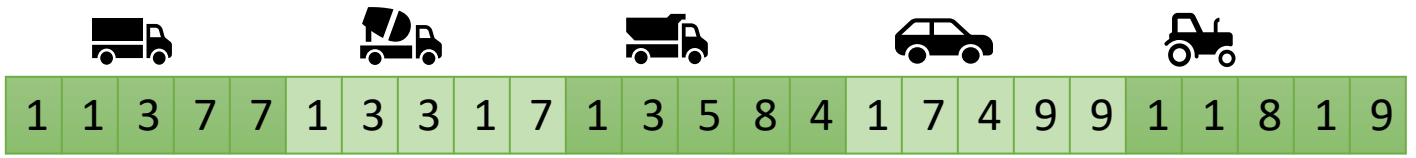


Our proposal: Improving the naïve approach

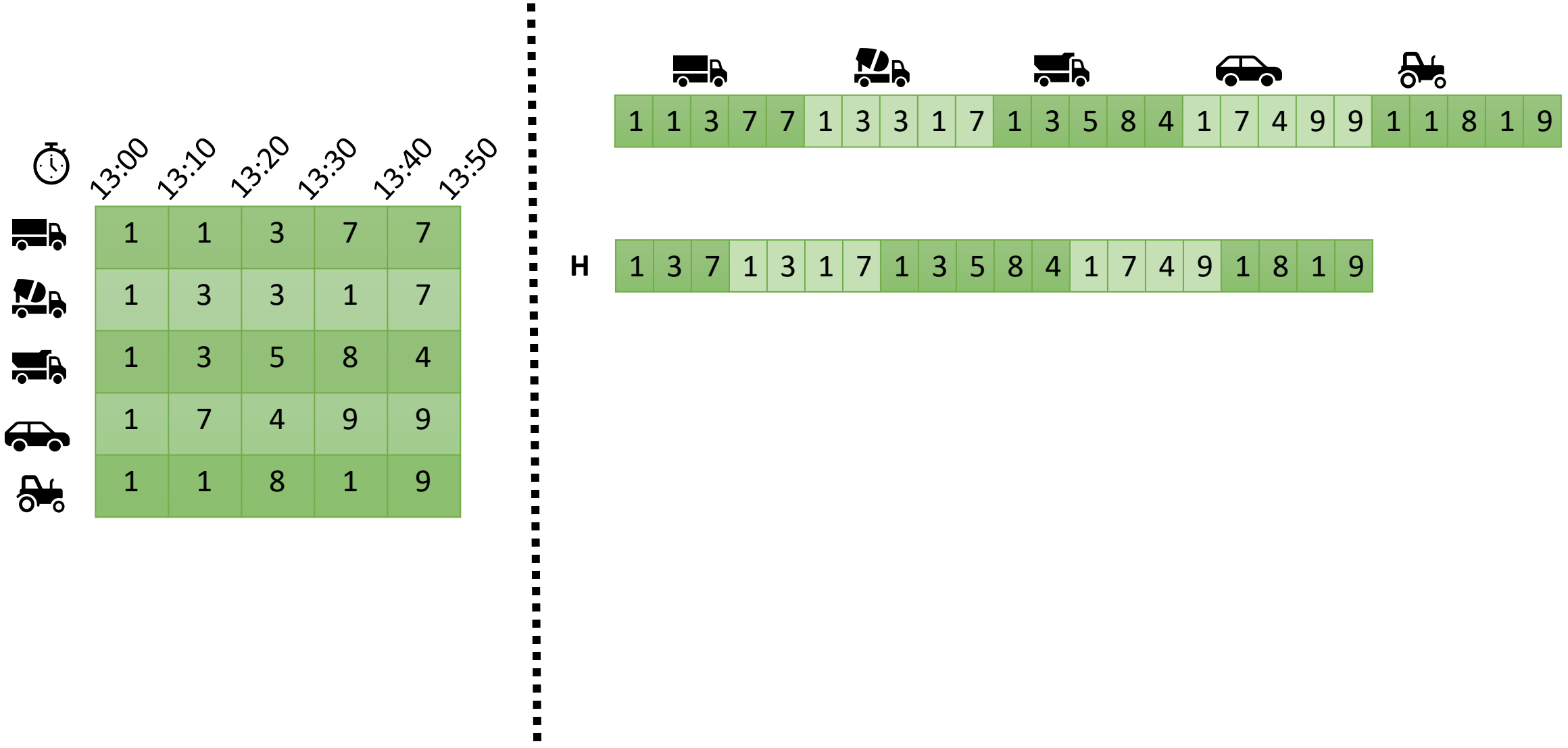


Our proposal: Improving the naïve approach

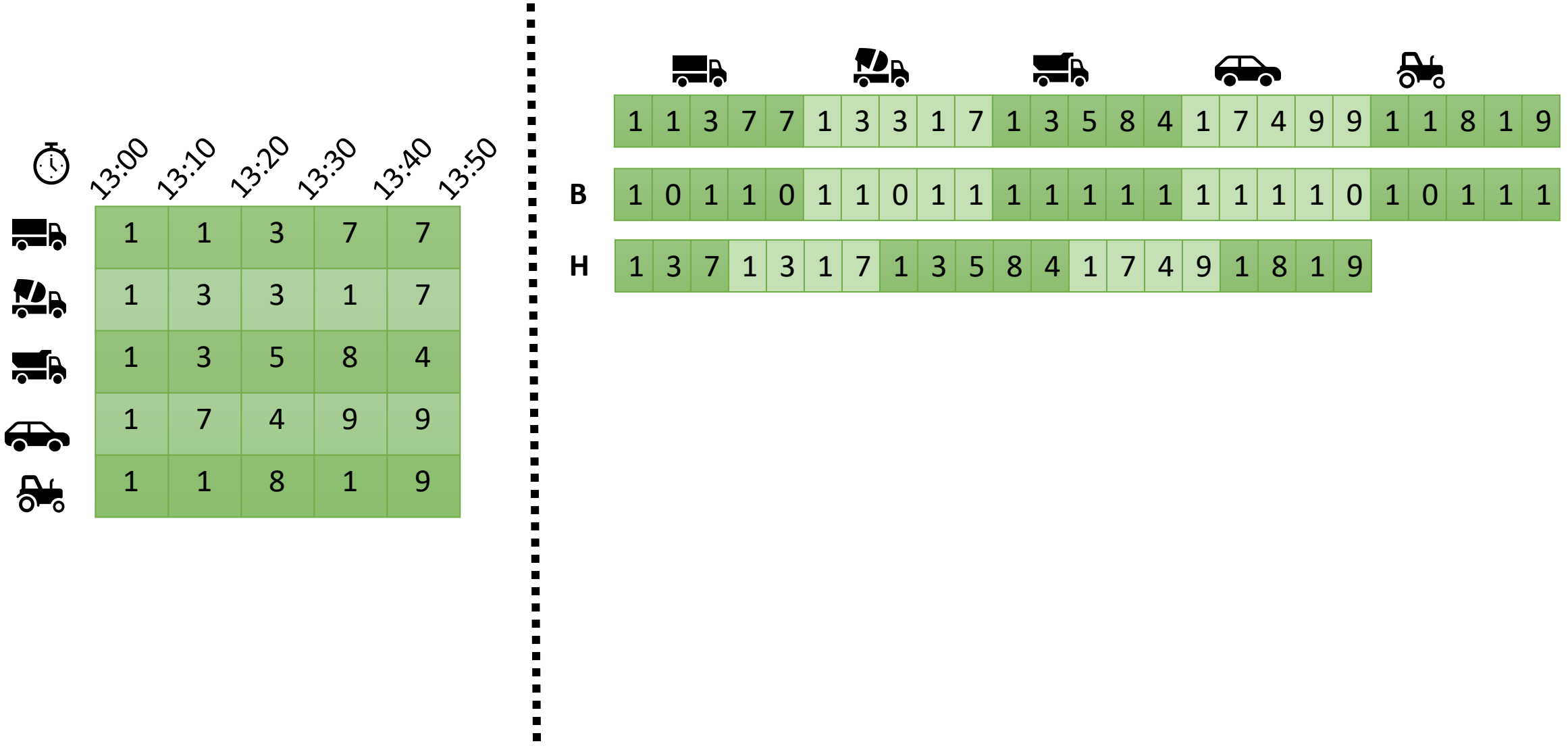




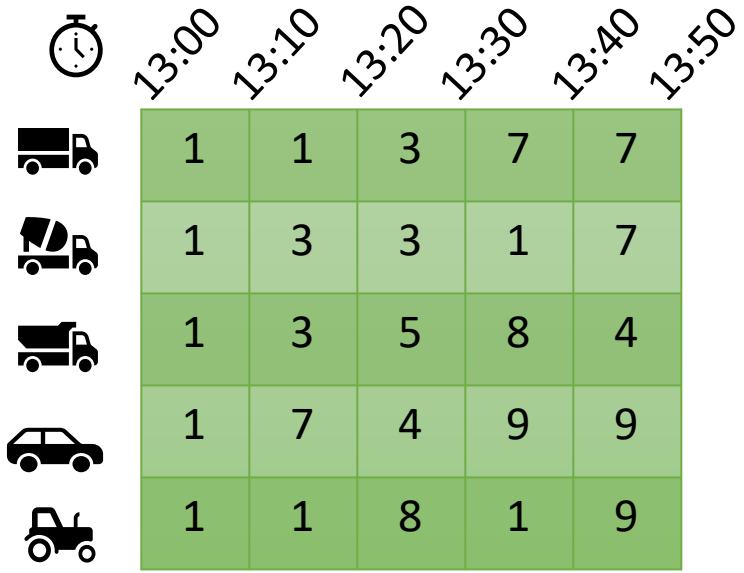
Our proposal: Improving the naïve approach

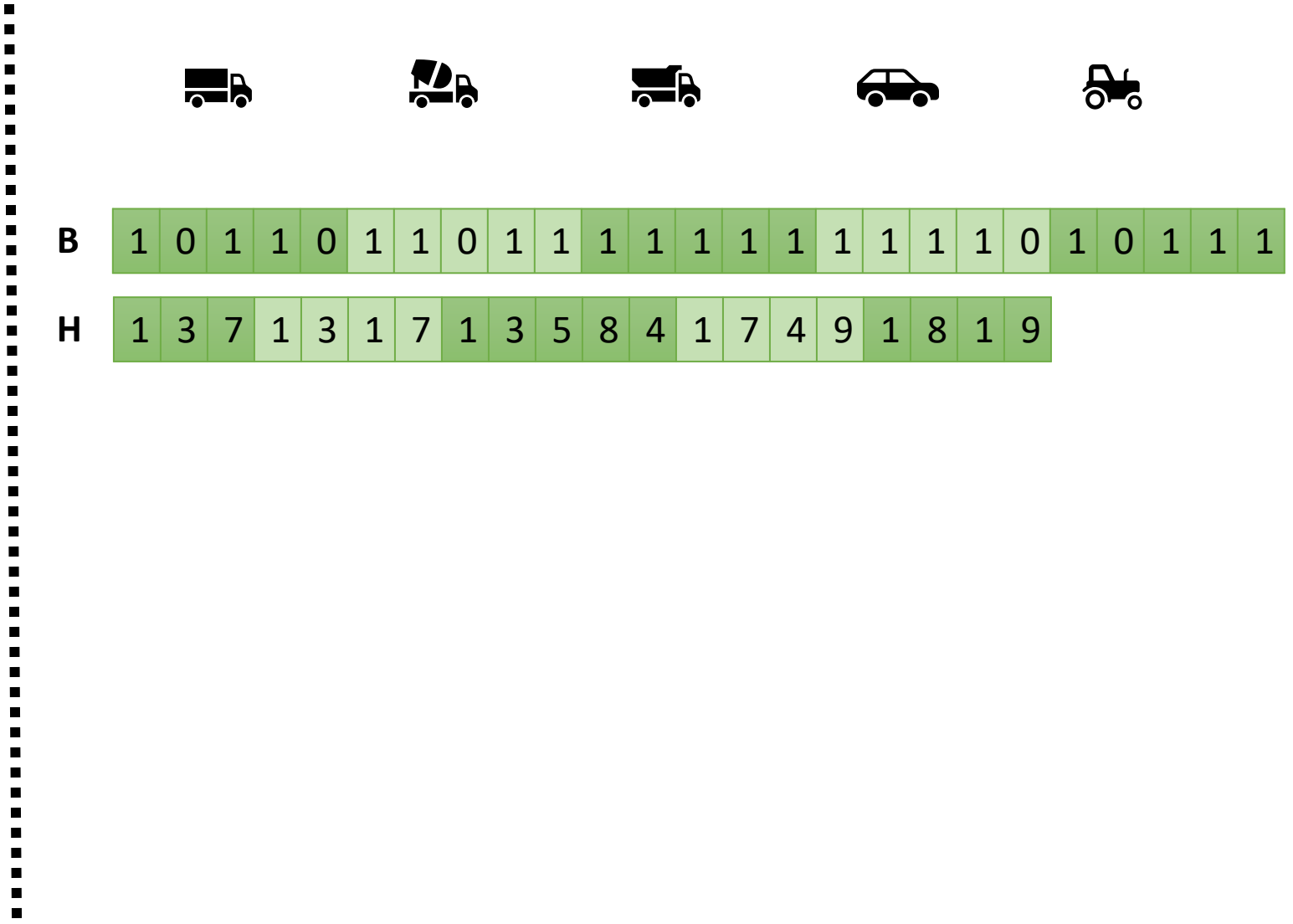


Our proposal: Improving the naïve approach

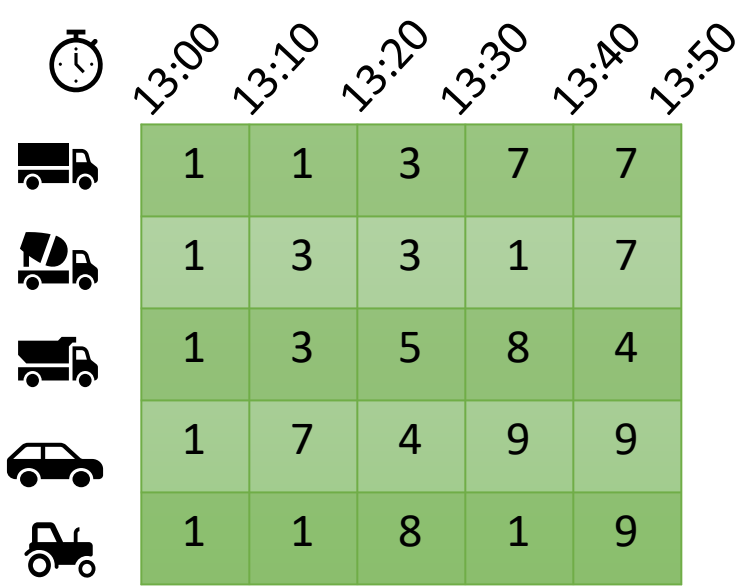


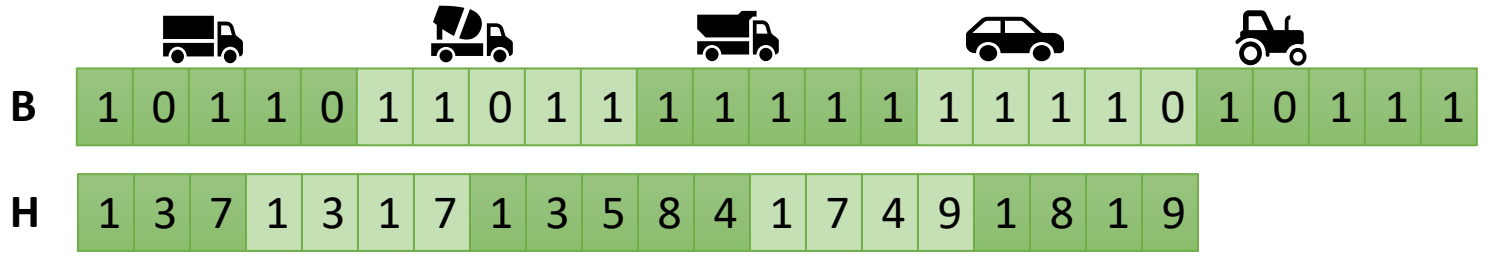
Our proposal: Improving the naïve approach





Our proposal: Improving the naïve approach





Our proposal: A brief digression

- Franklin C Crow, ***Summed-area tables for texture mapping***, ACM SIGGRAPH computer graphics, vol.18, no.3, pp. 207-212, 1984.

y \ x	1	2	3	4	5	6	7	8
1	2	4	3	2	1	3	4	2
2	1	3	2	1	3	1	3	1
3	2	1	3	2	2	1	3	1
4	2	3	3	5	2	1	1	1
5	2	3	1	1	3	1	1	1

(a) Original Matrix (A)

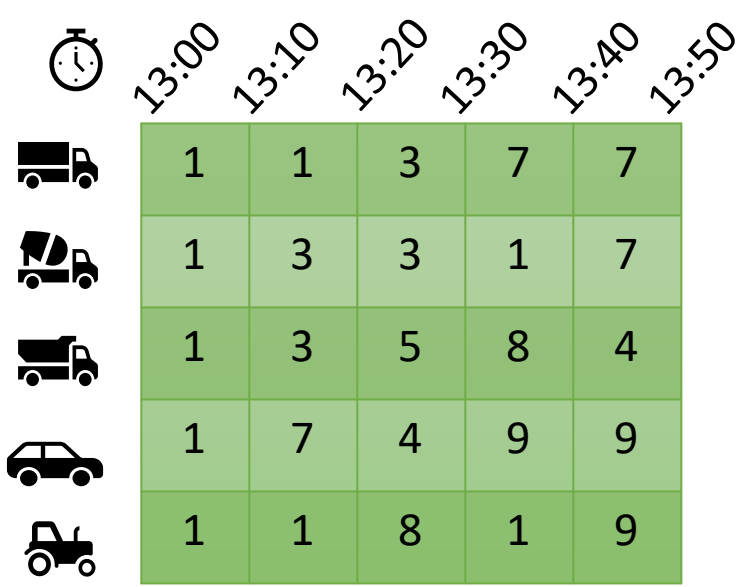
y \ x	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	2	**6**	9	11	12	15	**19**	21
2	0	3	10	15	18	22	26	33	36
3	0	5	13	21	26	32	37	47	51
4	0	7	**18**	29	39	47	53	**64**	69
5	0	9	23	35	46	57	64	76	82

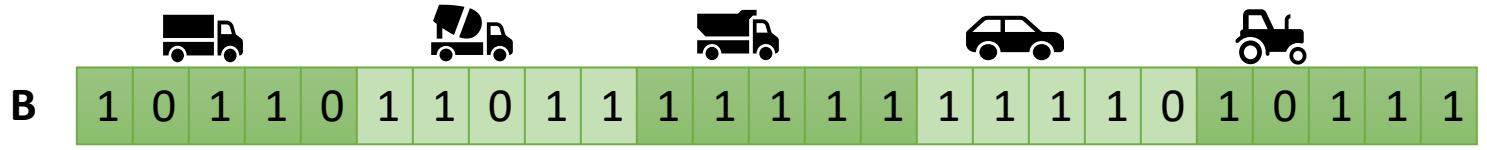
(b) Summed Area Table: Matrix (M)

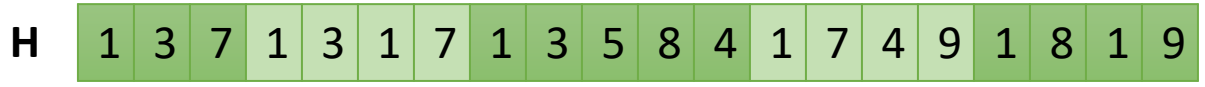
y \ x	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	2	**6**	9	11	12	15	**19**	21
2	0	3	10	15	18	22	26	33	36
3	0	5	13	21	26	32	37	47	51
4	0	7	**18**	29	39	47	53	**64**	69
5	0	9	23	35	46	57	64	76	82

(c) Summed Area Table: Matrix (M)

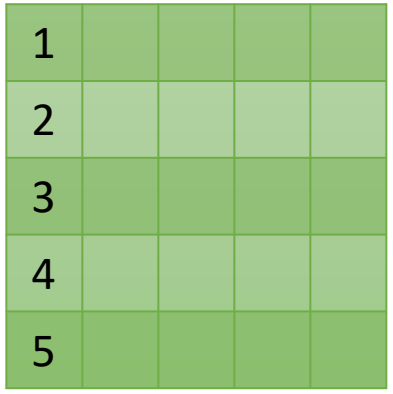
Our proposal: Improving the naïve approach



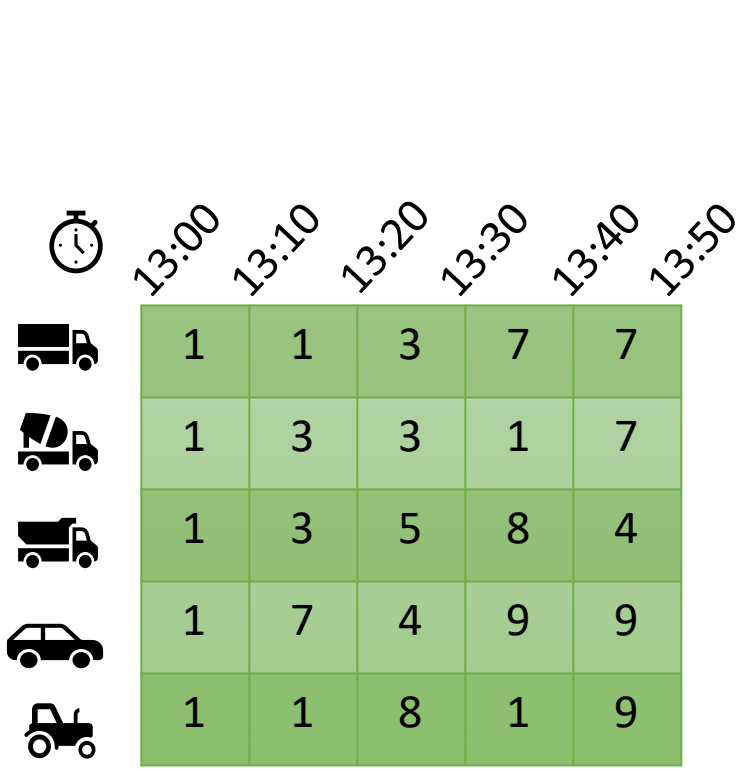




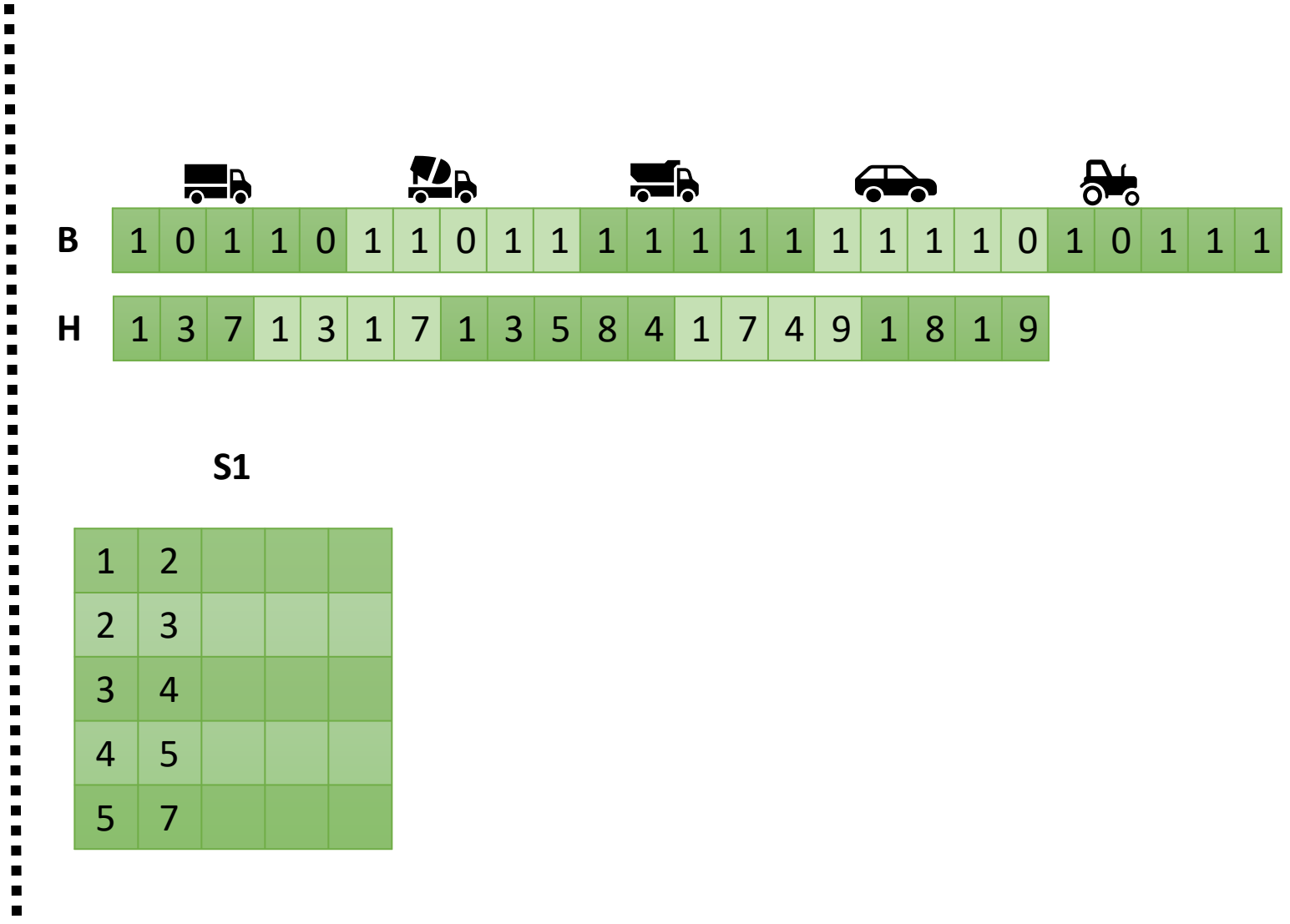
S1



Our proposal: Improving the naïve approach



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	

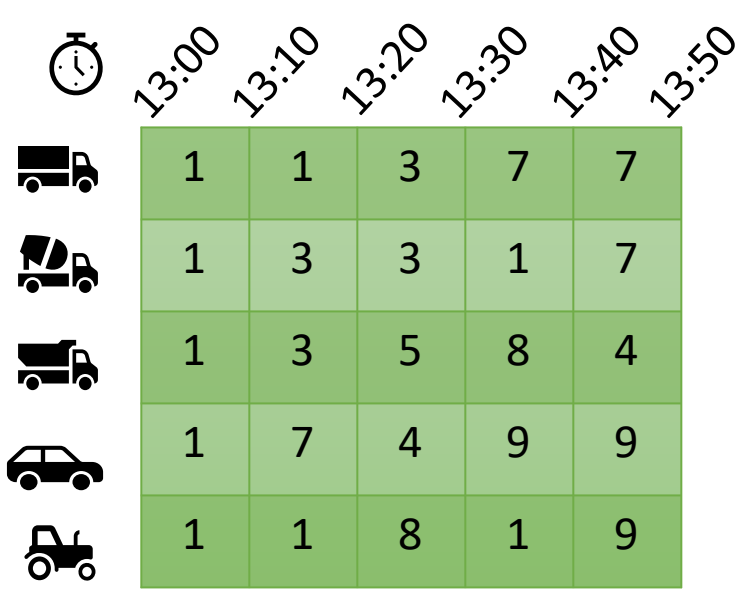


B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9				

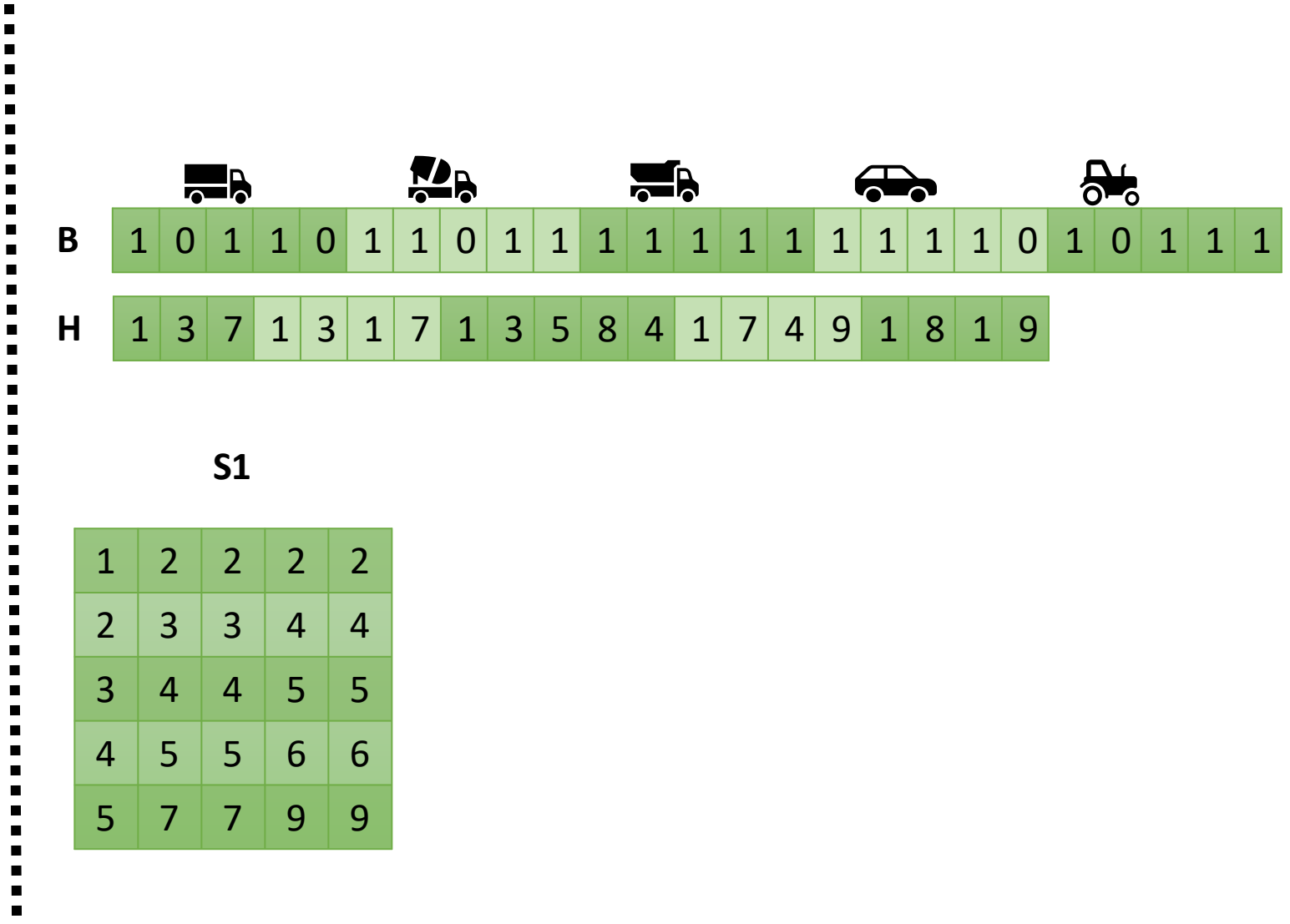
S1

1	2			
2	3			
3	4			
4	5			
5	7			

Our proposal: Improving the naïve approach



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B

1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1

H

1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

Our proposal: Improving the naïve approach

	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	

B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9					

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

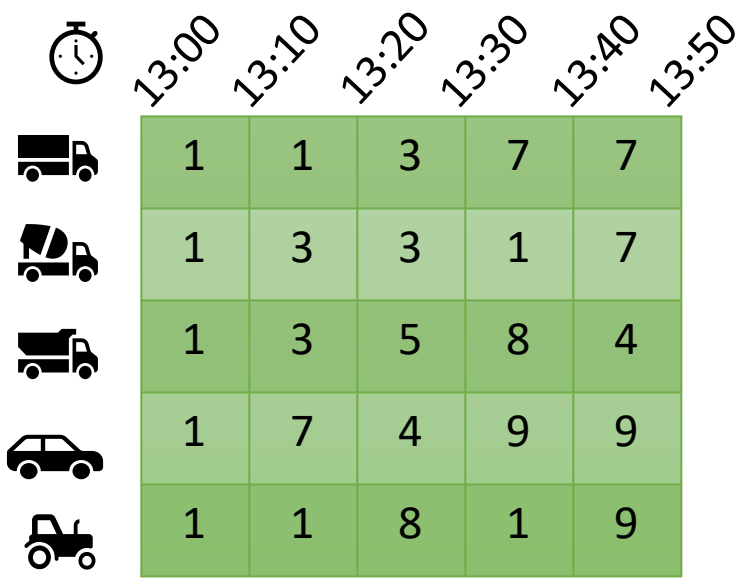
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

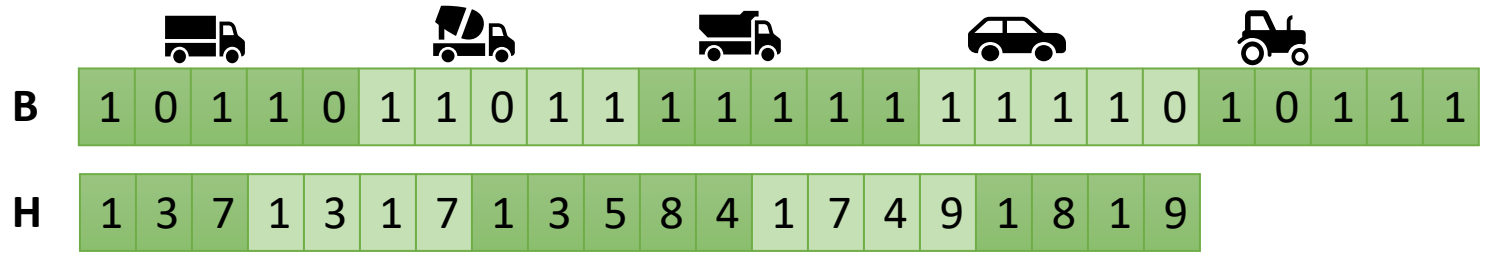
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Individual queries: *Which is the list of activities performed by a given driver between 13:00 and 13:30?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9					

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

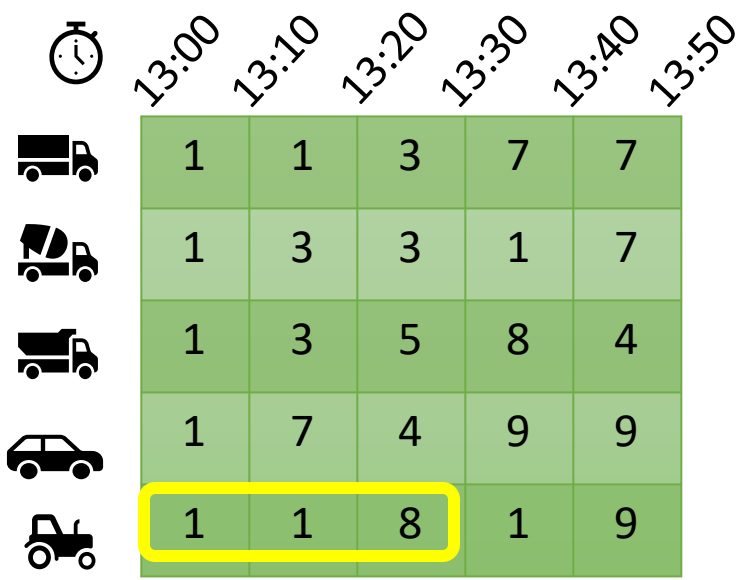
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

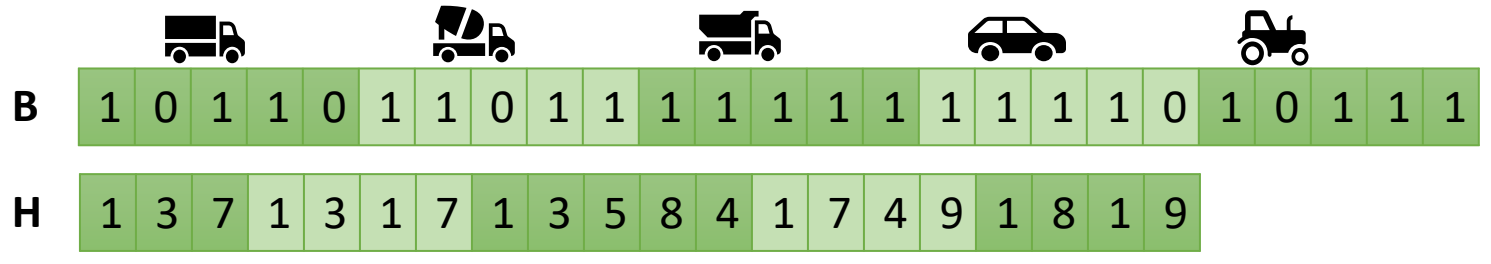
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Individual queries: *Which is the list of activities performed by a given driver between 13:00 and 13:30?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9					

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

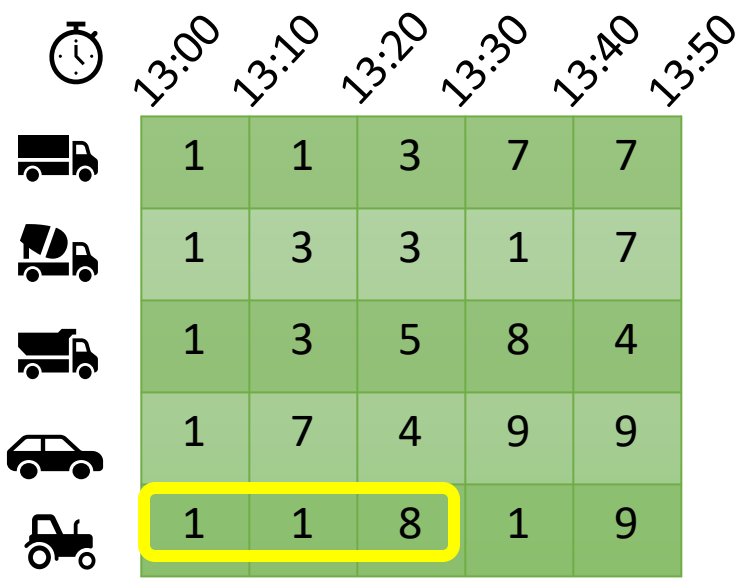
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

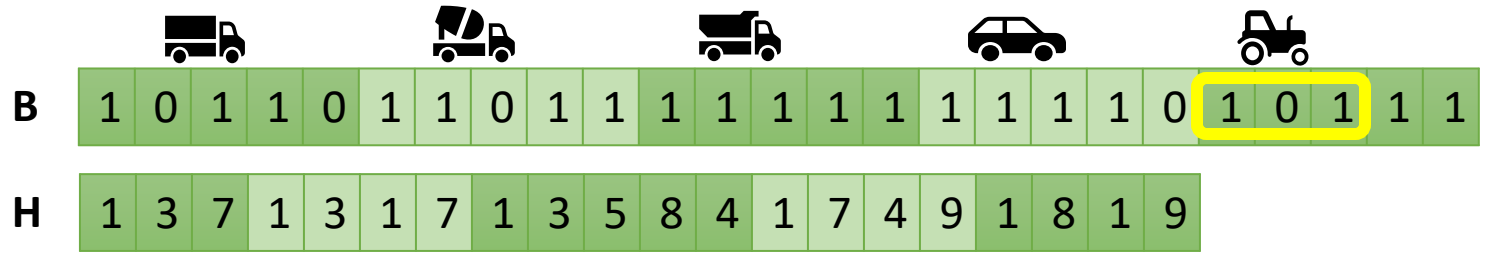
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Individual queries: *Which is the list of activities performed by a given driver between 13:00 and 13:30?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9				

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

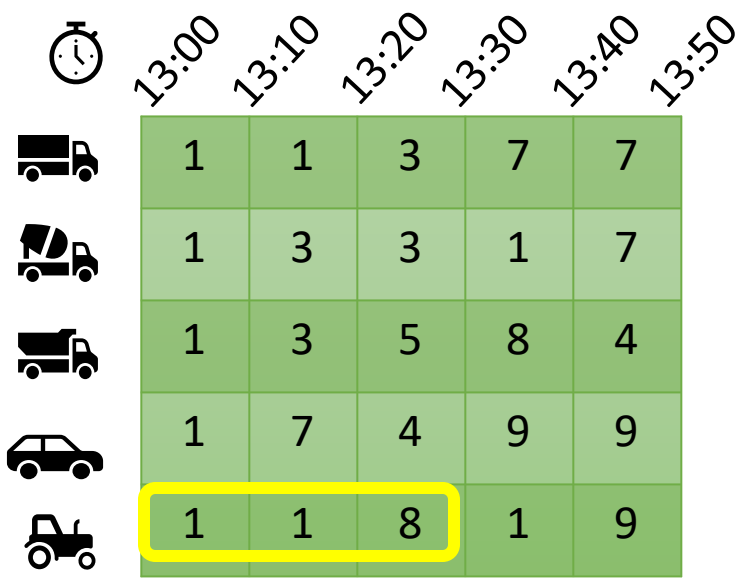
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

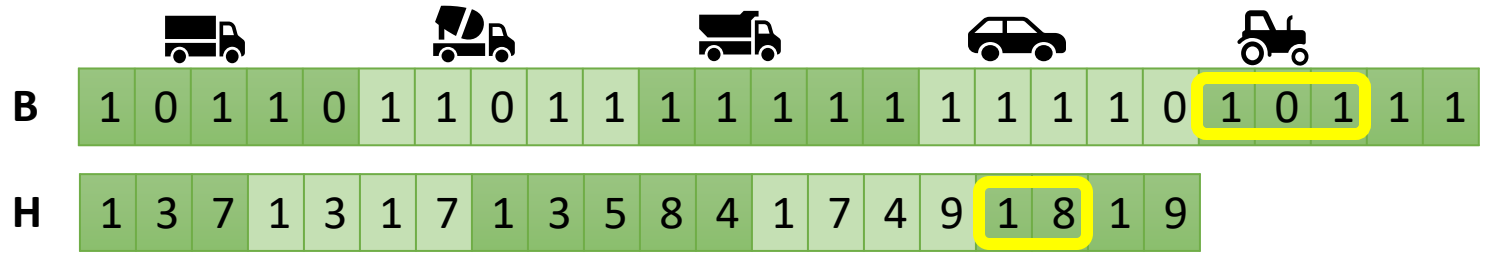
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Individual queries: *Which is the list of activities performed by a given driver between 13:00 and 13:30?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9					

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

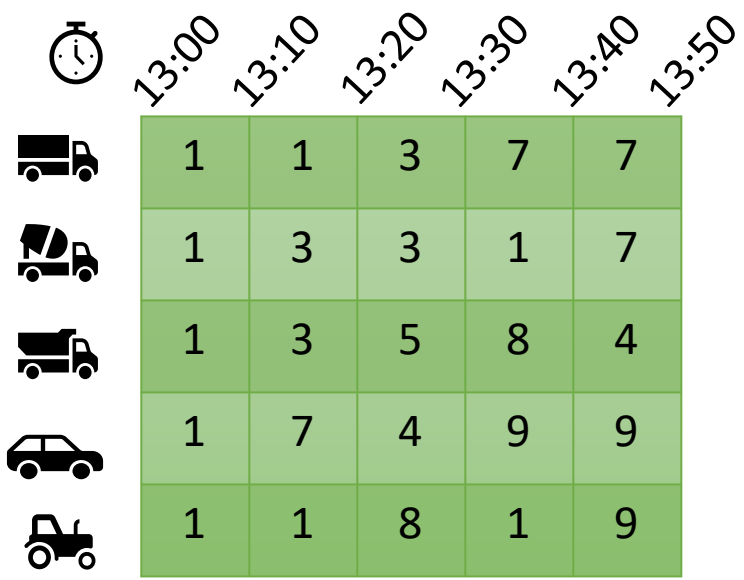
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

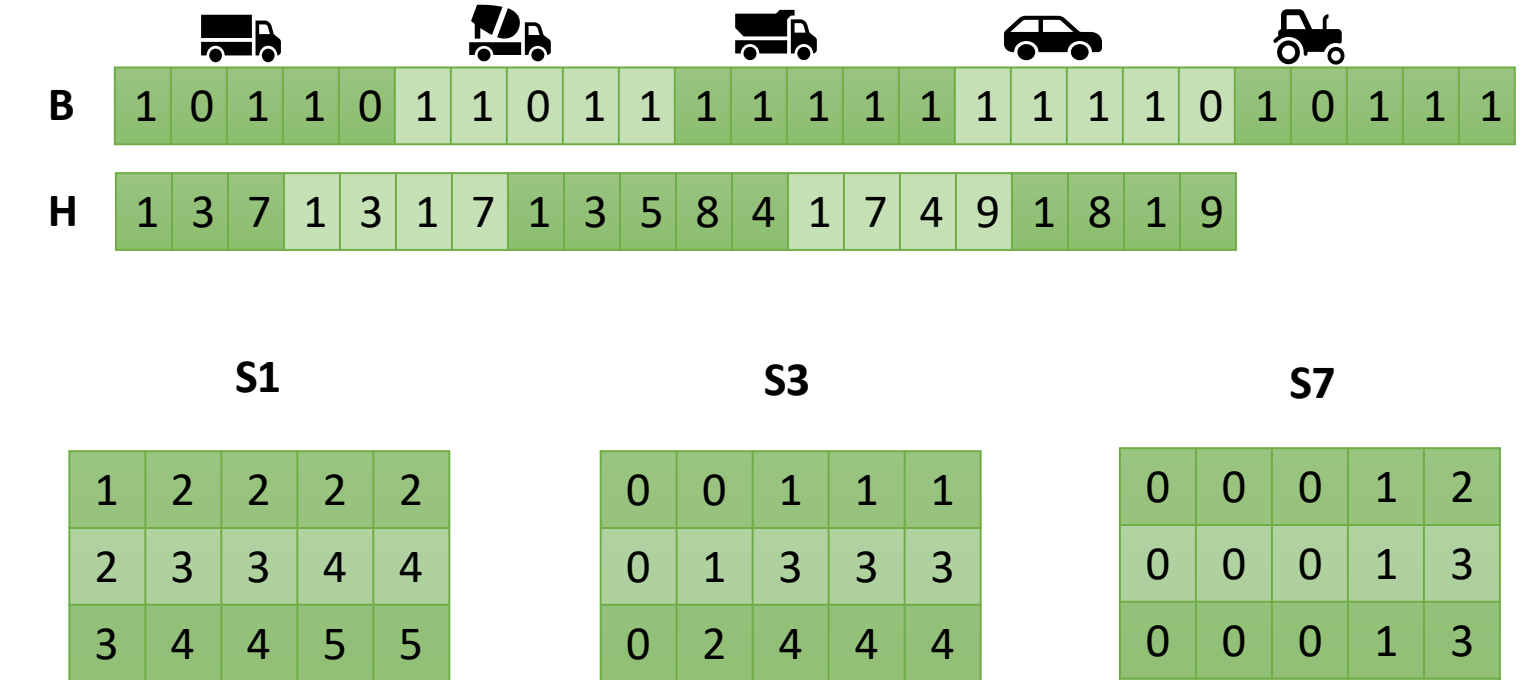
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Pattern queries: *How many times activity 1 was followed by activity 3?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9				

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

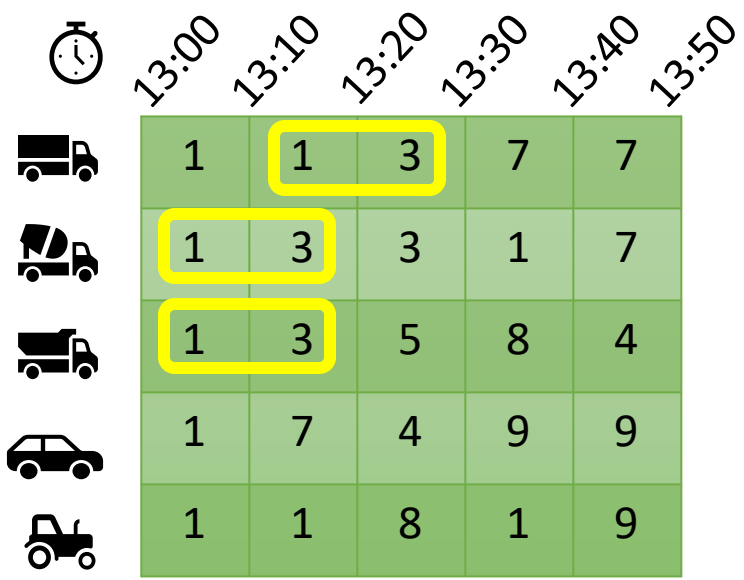
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

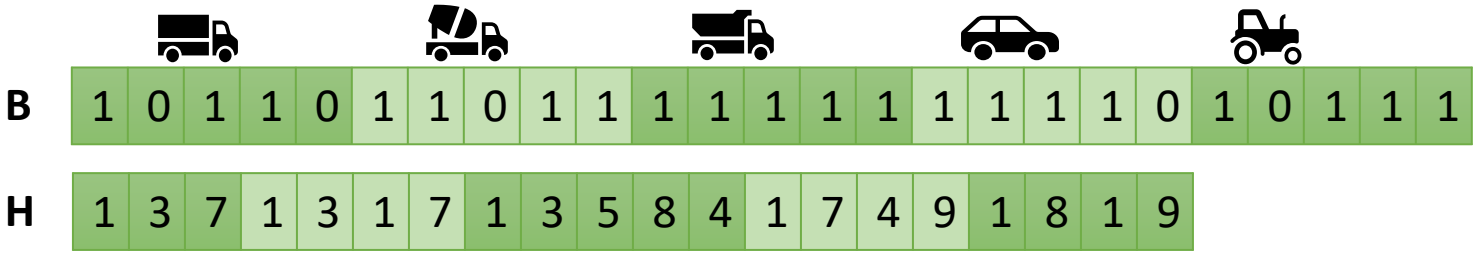
S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Pattern queries: *How many times activity 1 was followed by activity 3?*





S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

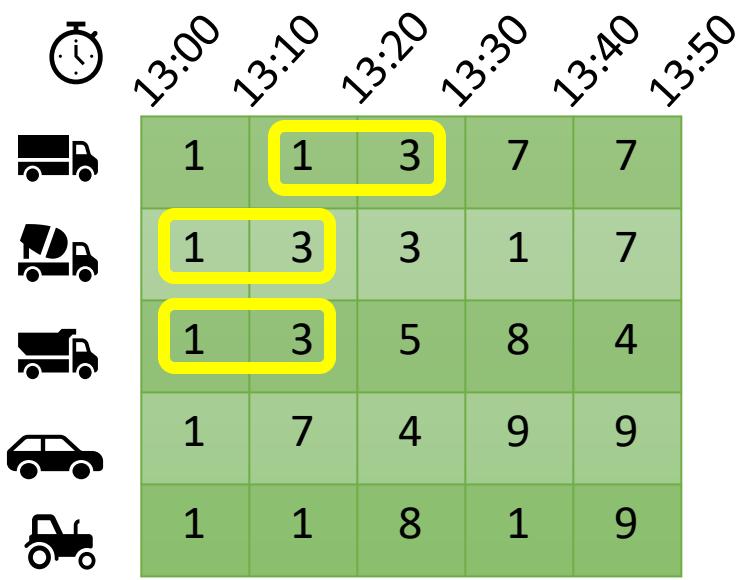
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

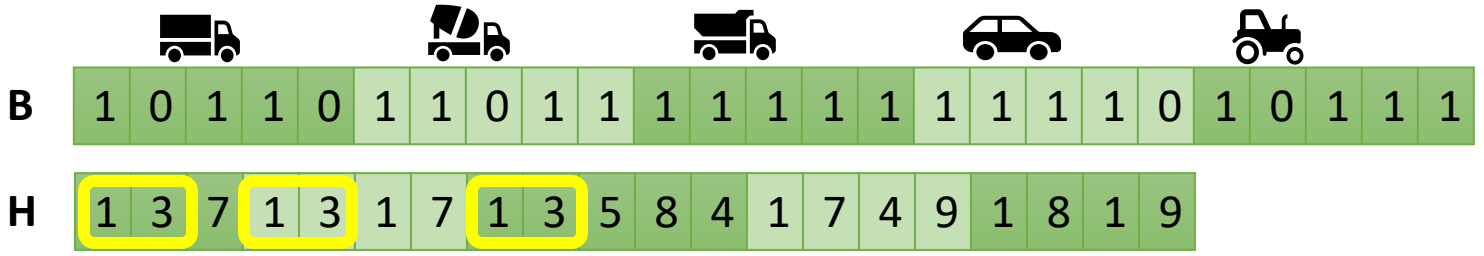
S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Pattern queries: *How many times activity 1 was followed by activity 3?*





S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

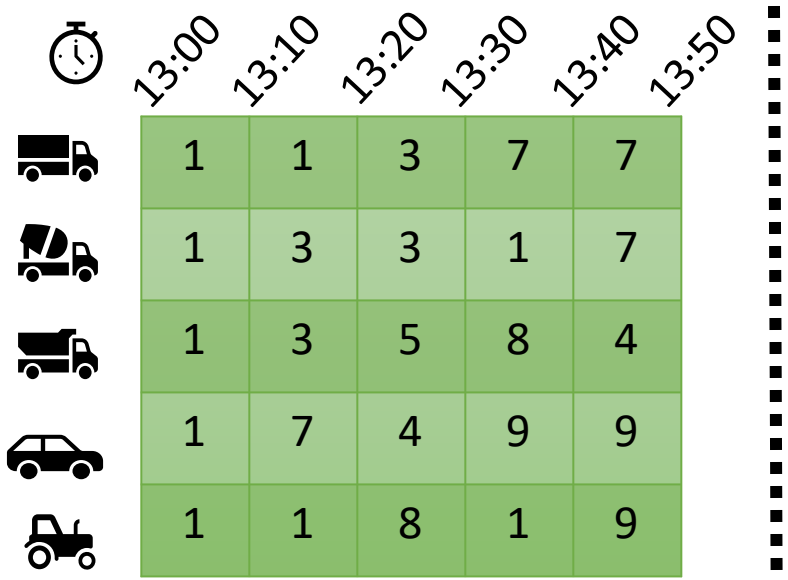
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

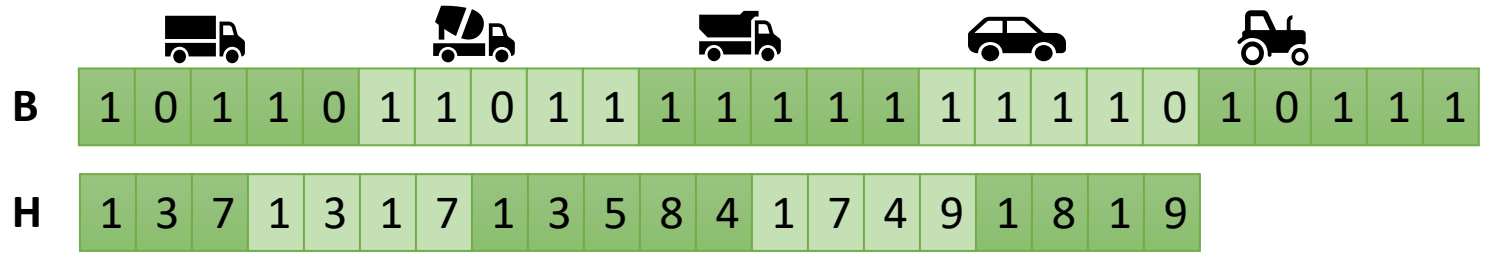
0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*



	13:00	13:10	13:20	13:30	13:40	13:50
	1	1	3	7	7	
	1	3	3	1	7	
	1	3	5	8	4	
	1	7	4	9	9	
	1	1	8	1	9	



B	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
H	1	3	7	1	3	1	7	1	3	5	8	4	1	7	4	9	1	8	1	9				

S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

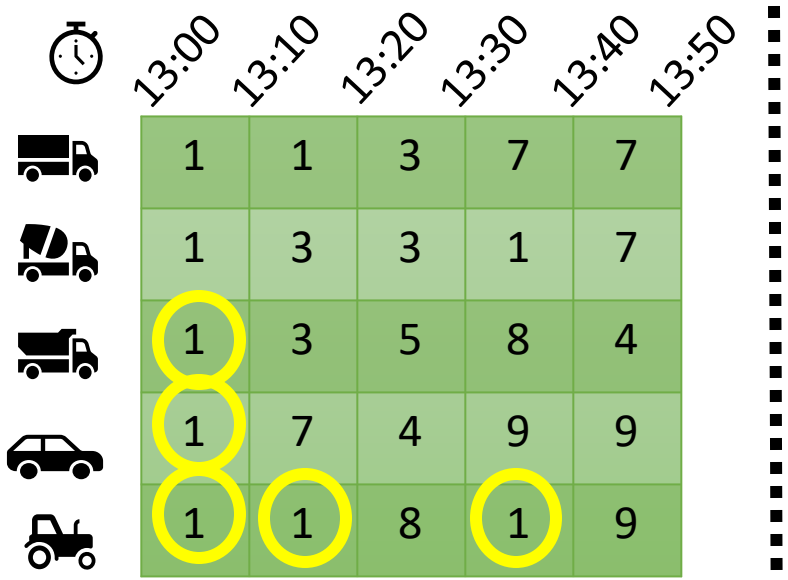
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

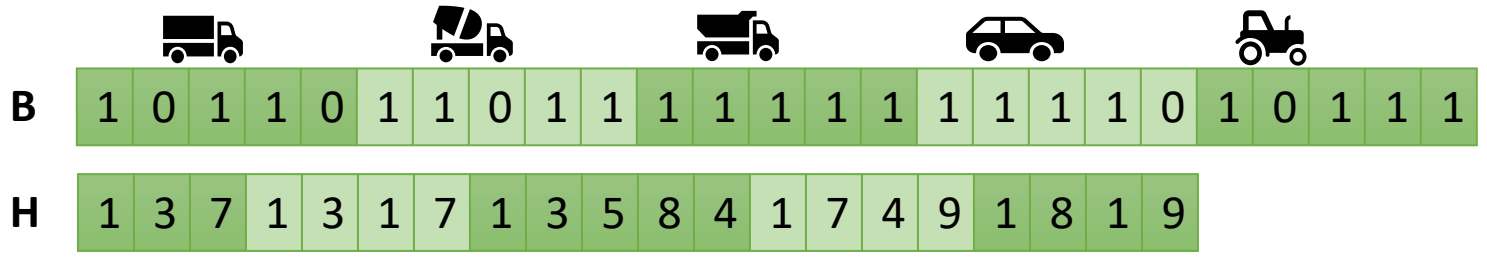
S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*





S1

1	2	2	2	2
2	3	3	4	4
3	4	4	5	5
4	5	5	6	6
5	7	7	9	9

S3

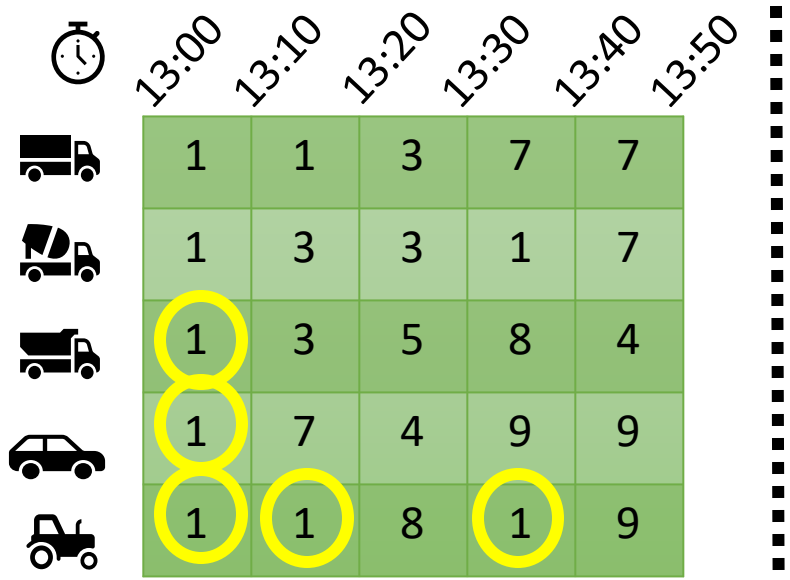
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

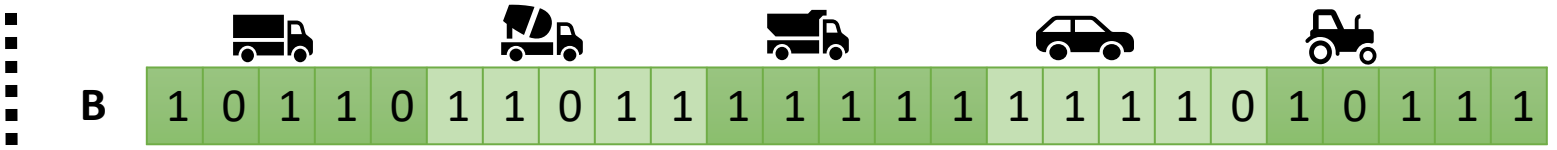
S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*





S1

0	1	2	2	2	2
0	2	3	3	4	4
0	3	4	4	5	5
0	4	5	5	6	6
0	5	7	7	9	9

S3

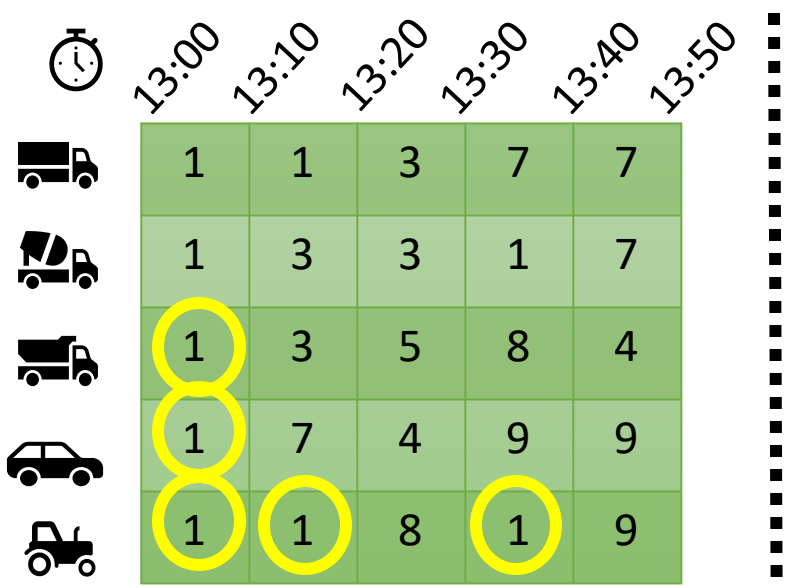
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

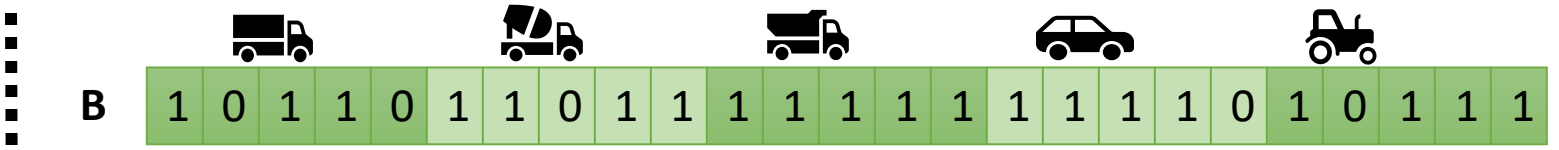
S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

Our proposal: Solving queries

Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*





S1

0	1	2	2	2	2
0	2	3	3	4	4
0	3	4	4	5	5
0	4	5	5	6	6
0	5	7	7	9	9

S3

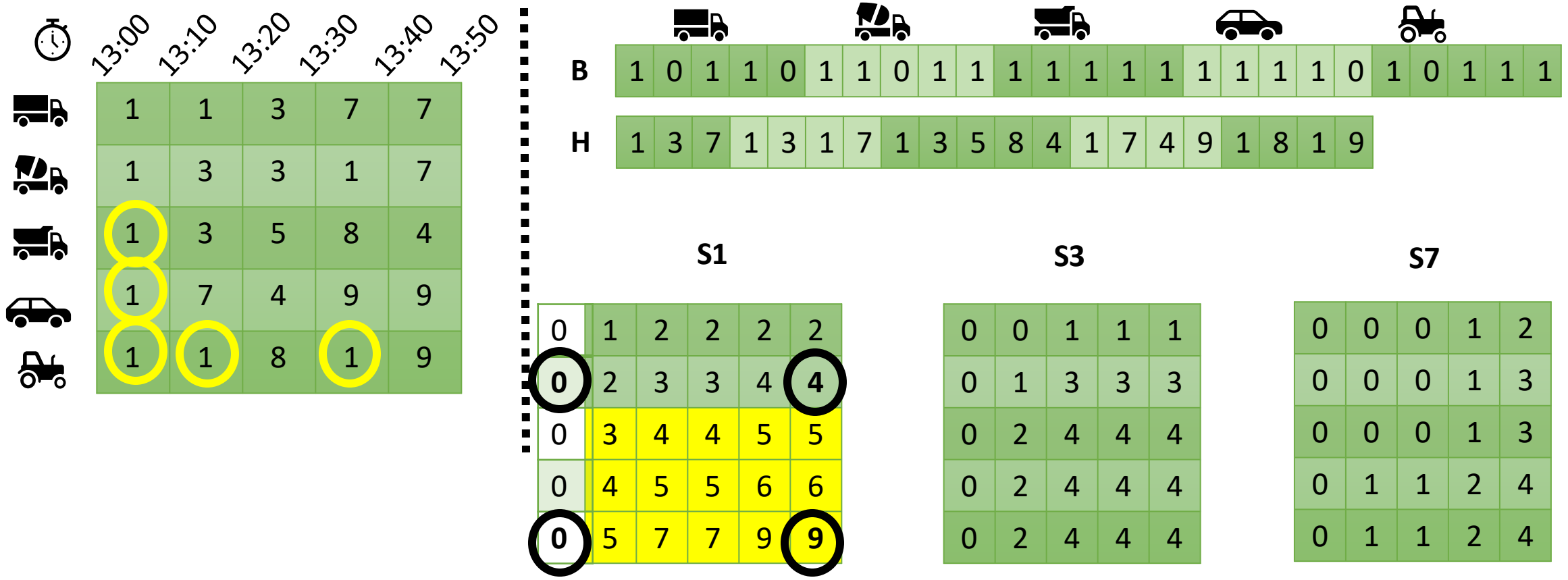
0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

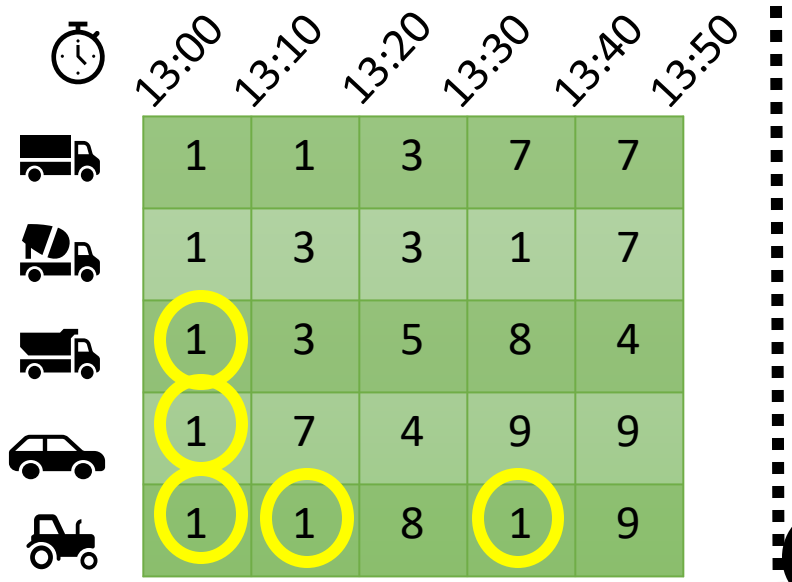
Our proposal: Solving queries

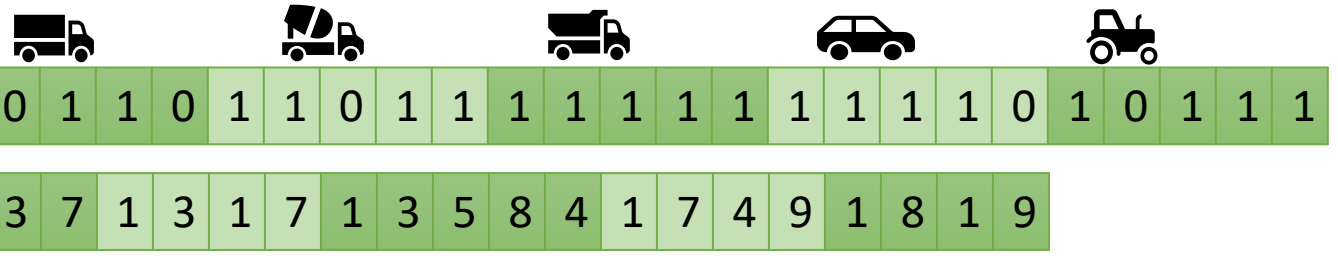
Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*



Our proposal: Solving queries

Aggregated queries: *How much time was actually spent at headquarters (ID: 1) by the three last vehicles?*





S1

0	1	2	2	2	2
0	2	3	3	4	**4**
0	3	4	4	5	5
0	4	5	5	6	6
0	5	7	7	9	**9**

S3

0	0	1	1	1
0	1	3	3	3
0	2	4	4	4
0	2	4	4	4
0	2	4	4	4

S7

0	0	0	1	2
0	0	0	1	3
0	0	0	1	3
0	1	1	2	4
0	1	1	2	4

- **9 – 0 – 4 + 0 = 5**

Route map

- Introduction
- The problem
- Our proposal
- **Experiments**
- Future work

4th stop: Experiments

Experiments: Baseline +

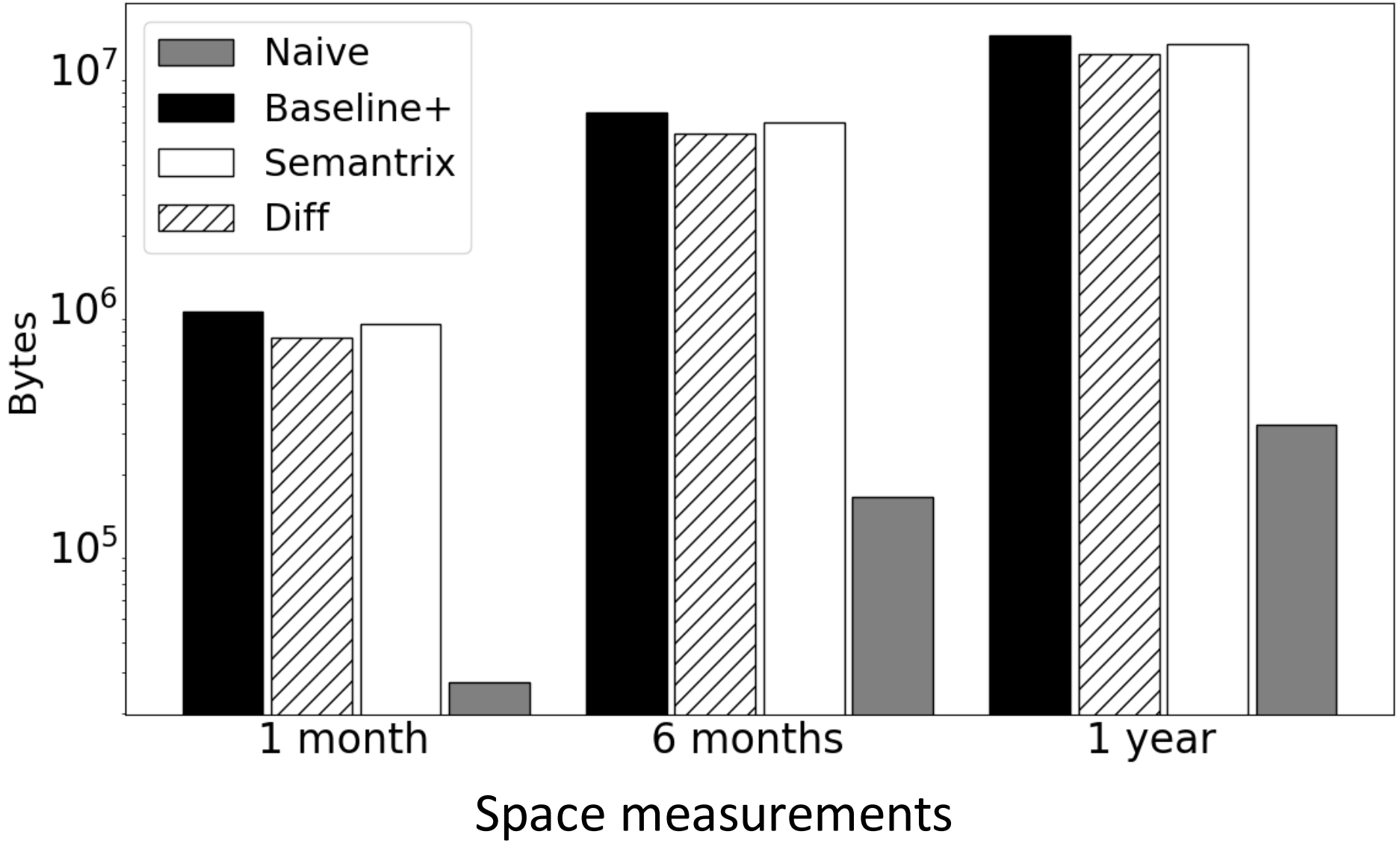
OS: sequence of activities

1	1	3	7	7	1	3	3	1	7	1	3	5	8	4	1	7	4	9	9	1	1	8	1	9

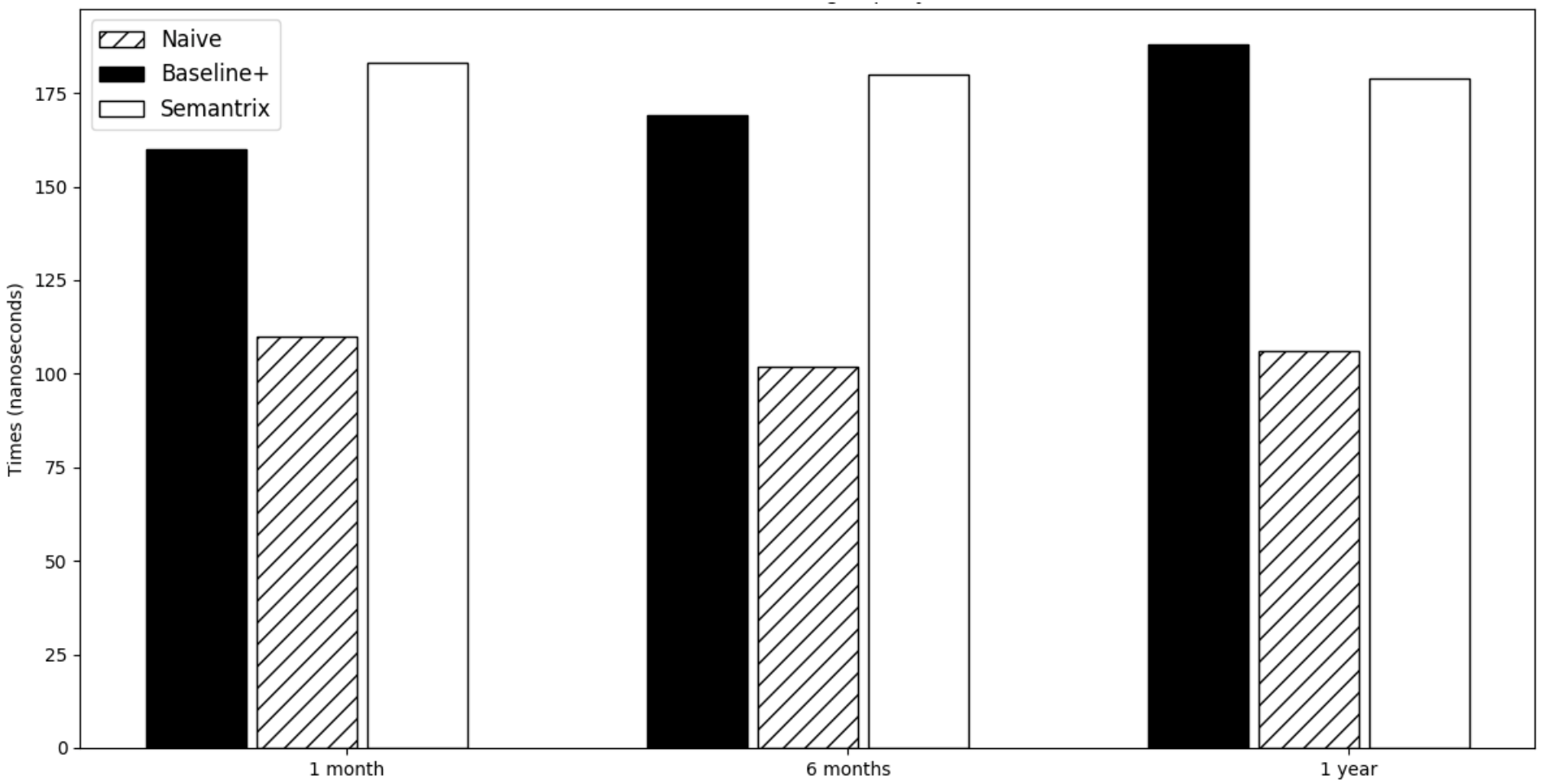
Accumulative sequences

A1	1	2	2	2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	7	8	8	9	10
A2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	1	1	1	1	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4
⋮													⋮												
A9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	2	3

Experiments: Space

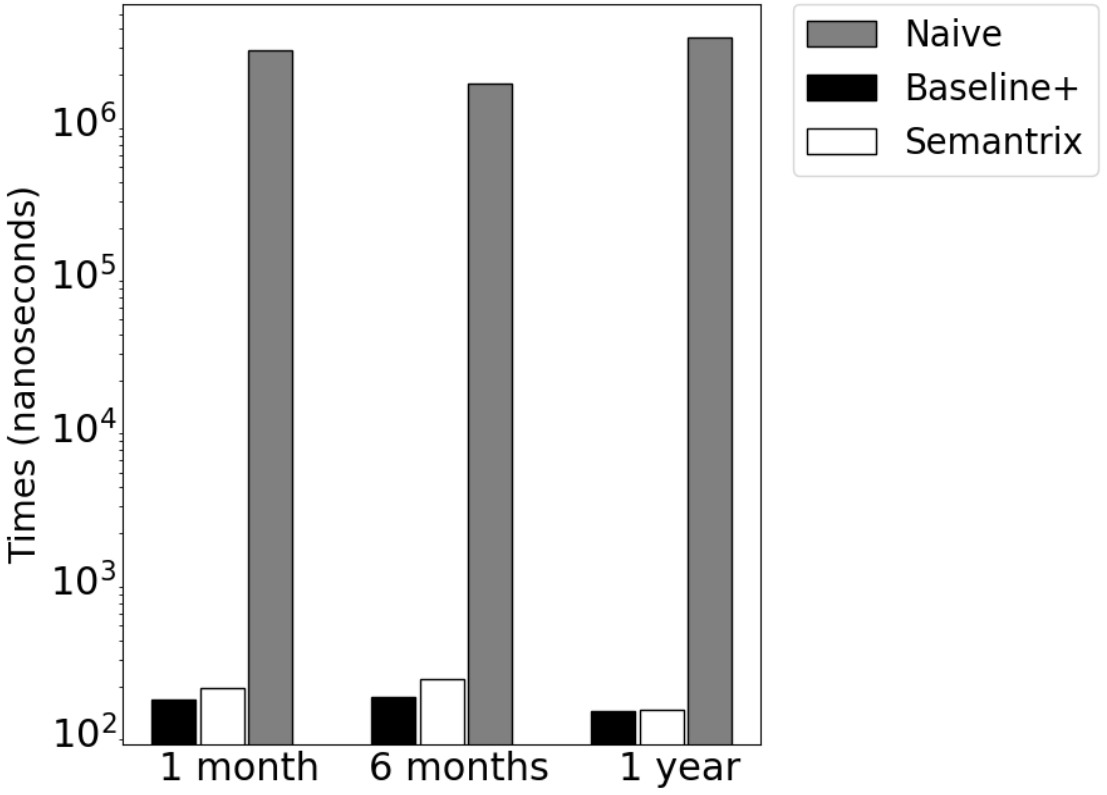


Experiments: Time I

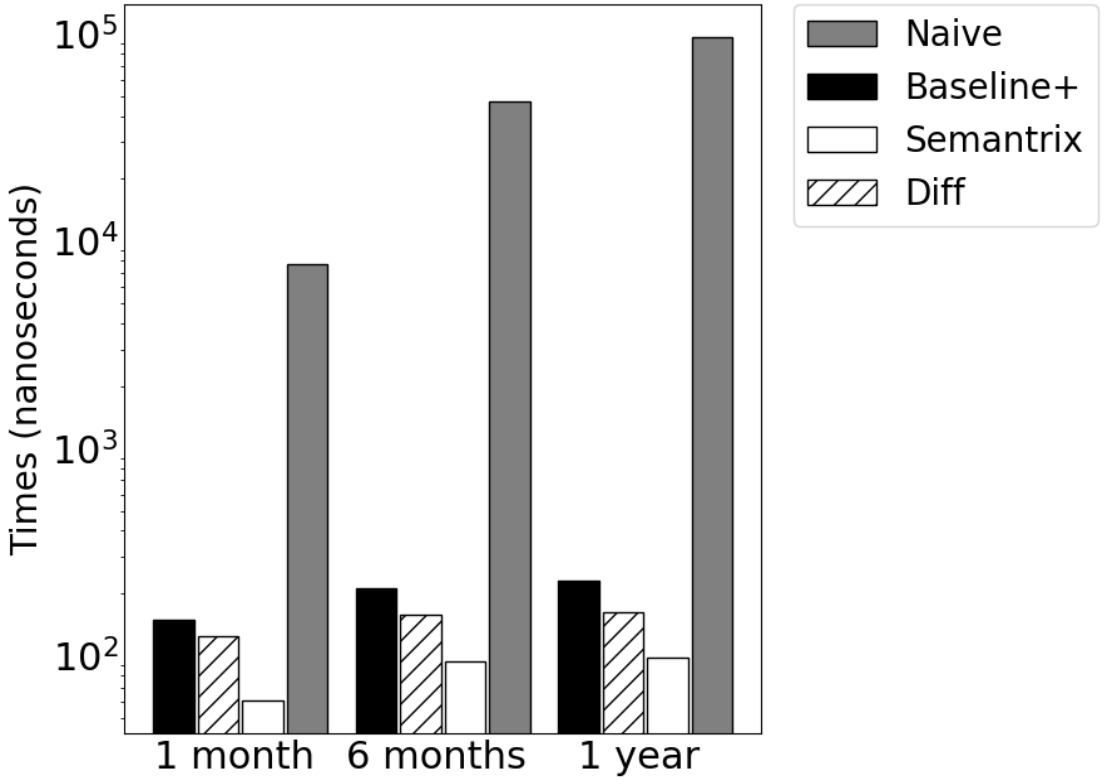


Times resolving individual queries

Experiments: Time II



Times resolving patter queries



Times resolving aggregation queries

Route map

- Introduction
- The problem
- Our proposal
- Experiments
- **Future work**

5th stop: Future work

Future work

- The first step will be to increase the scope of this work in order to represent in a compact way also the **geometry** of each semantically tagged segment or semantic trajectory.

- This idea opens a wide new field of possibilities to perform queries combining spatial, temporal, and semantic constraints.

Semantrix: A Compressed Semantic Matrix

Nieves R. Brisaboa
Antonio Fariña
Gonzalo Navarro
Tirso V. Rodeiro

tirso.varela.rodeiro@udc.es

Instituto Milenio Fundamentos de los datos

citic