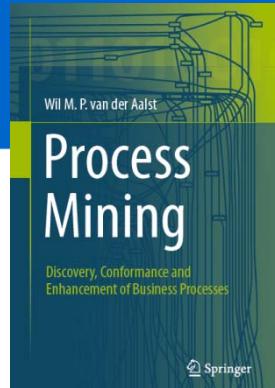


*Process Mining: Data Science in Action*

# Exploring Event Data



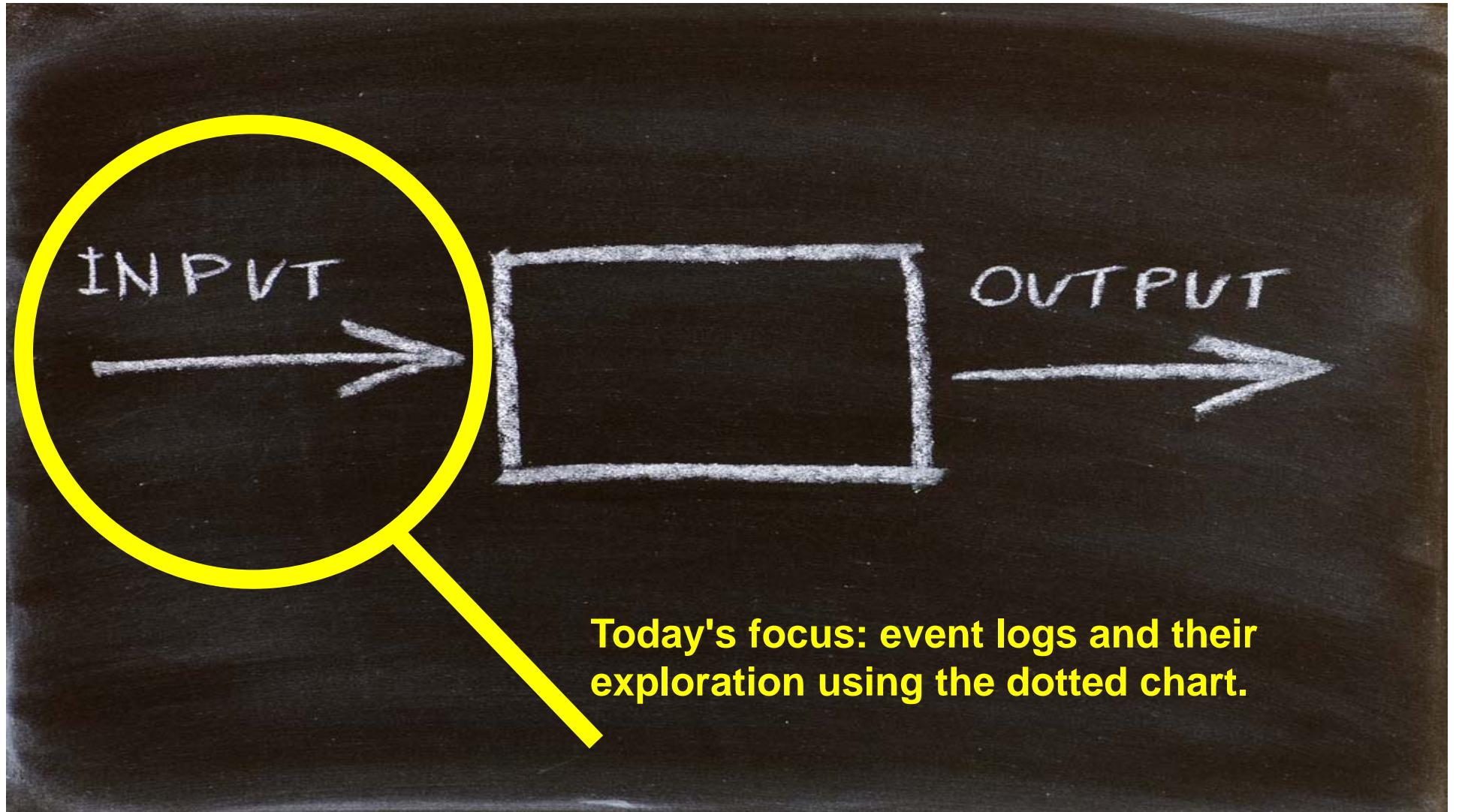
prof.dr.ir. Wil van der Aalst  
[www.processmining.org](http://www.processmining.org)



Technische Universiteit  
**Eindhoven**  
University of Technology

Where innovation starts

©Wil van der Aalst & TU/e (use only with permission & acknowledgements)

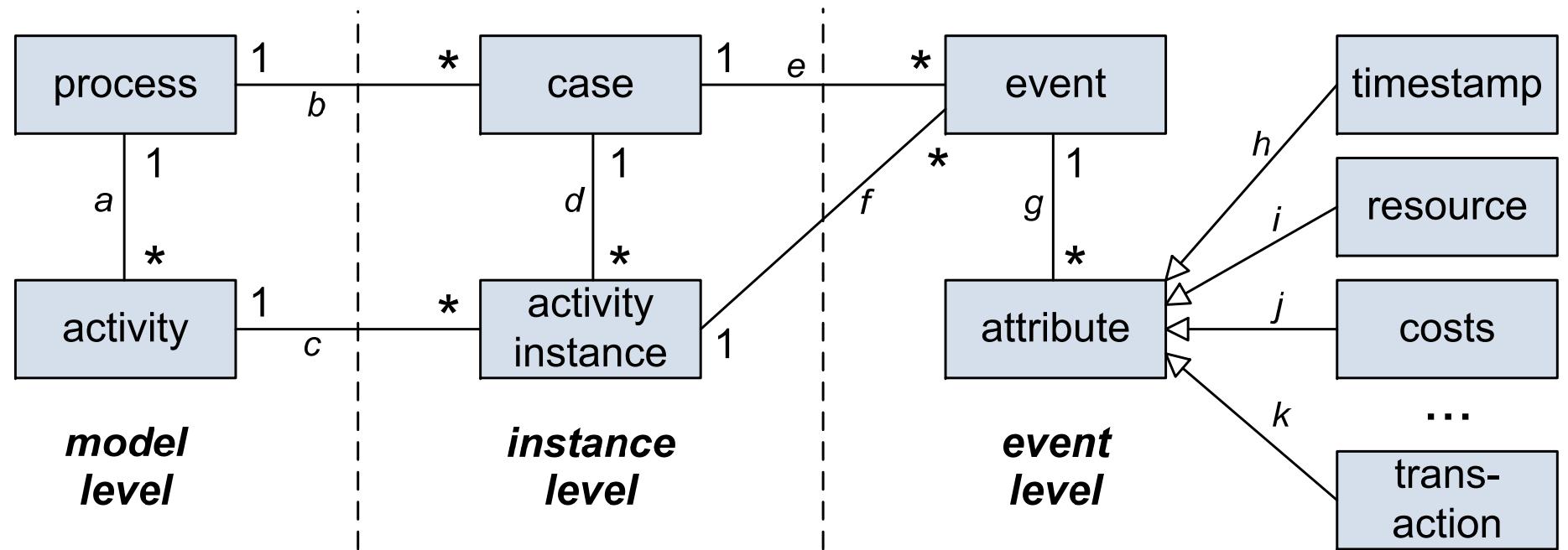


# Simplistic view on event data

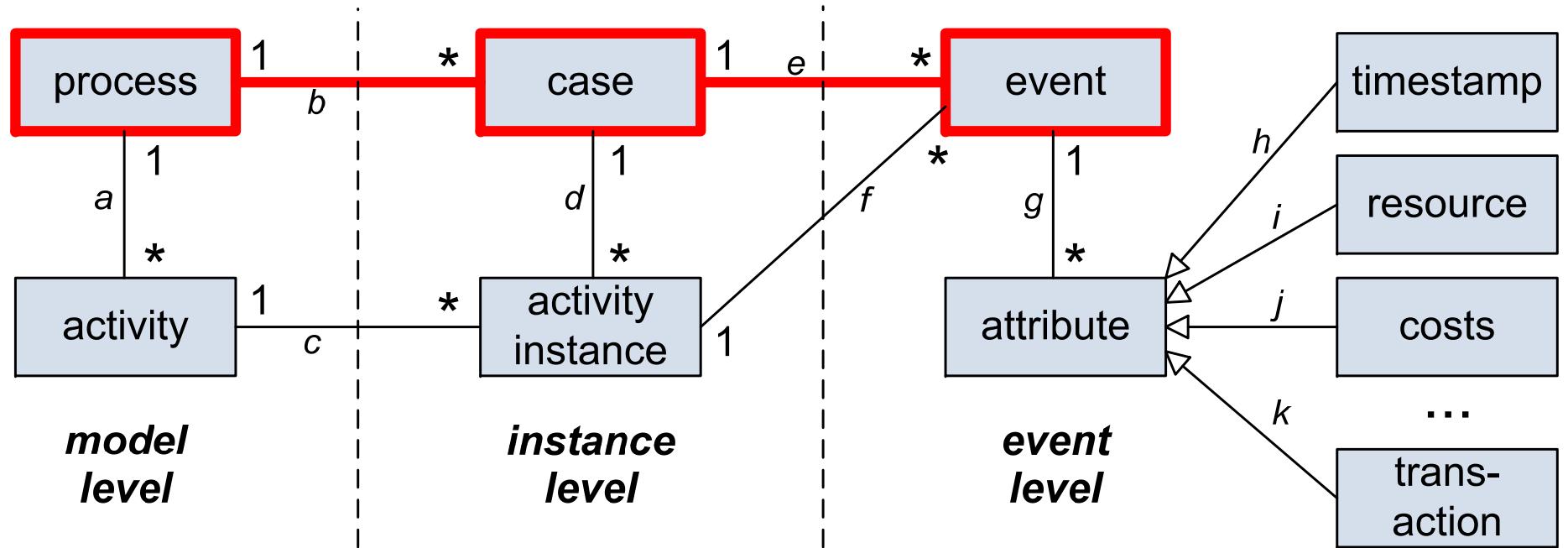
order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
...	...	...	...	...	...

**case id**    **activity name**    **timestamp**    **resource**    **other data**

# A more refined view



# Process – case - event

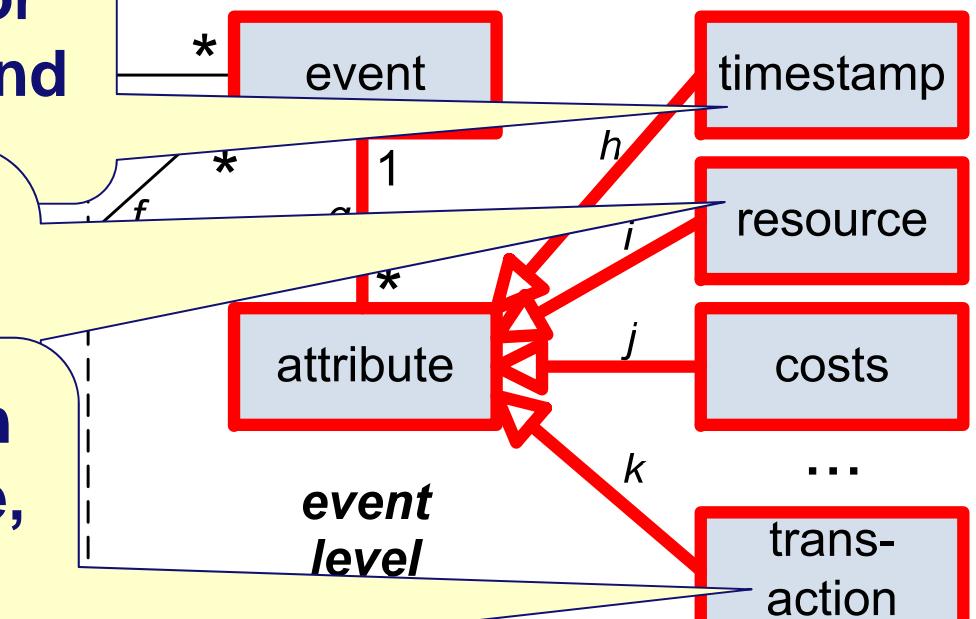


# Events have attributes

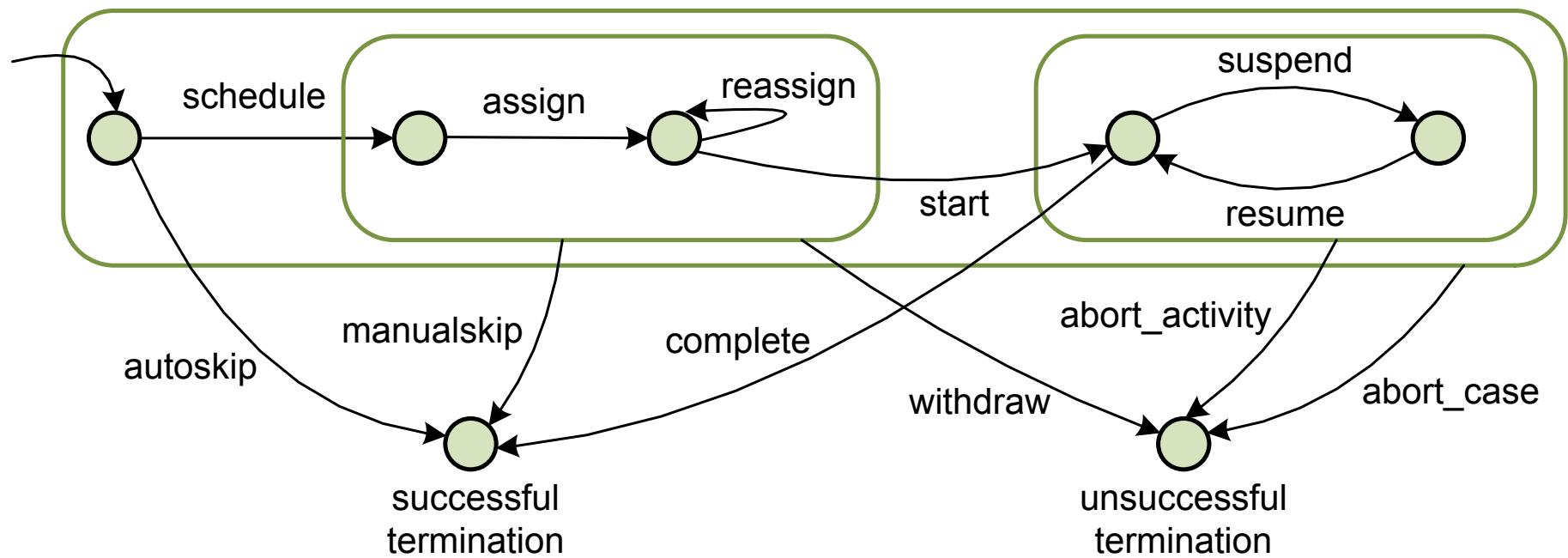
time of event: crucial for  
the ordering of events and  
performance analysis

resource associated to  
event (person, machine,  
department, organization)

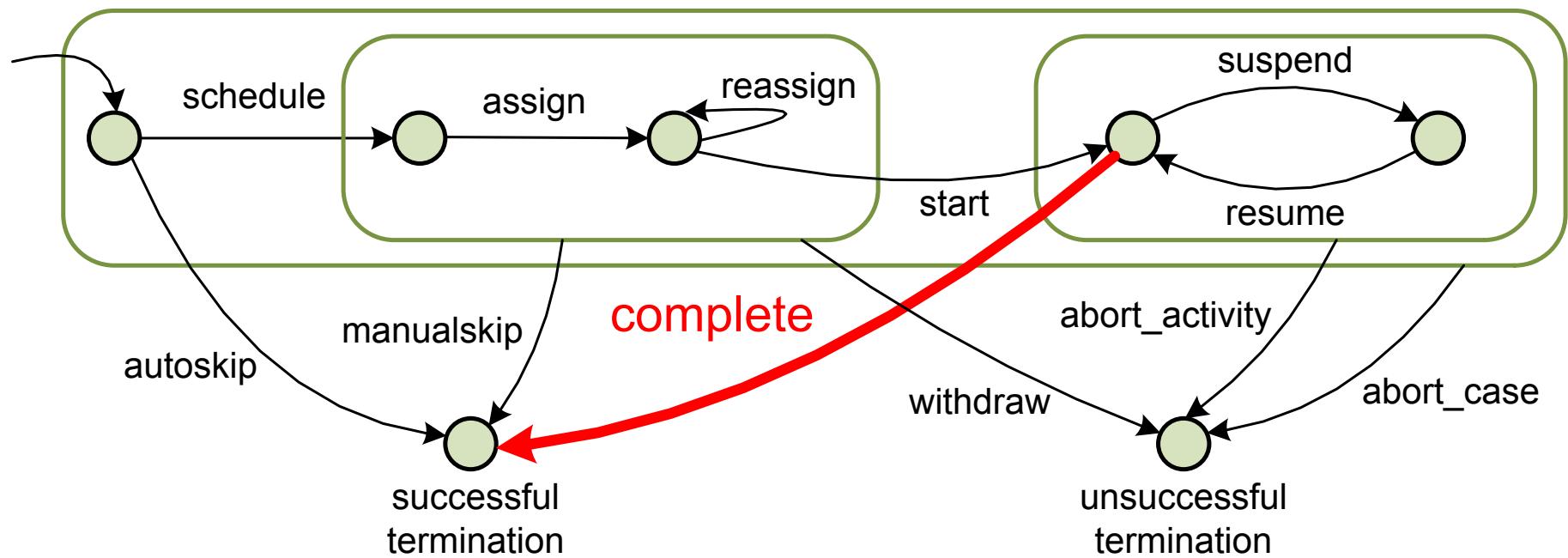
transactional information  
(start, suspend, complete,  
etc.): activities are often  
non-atomic



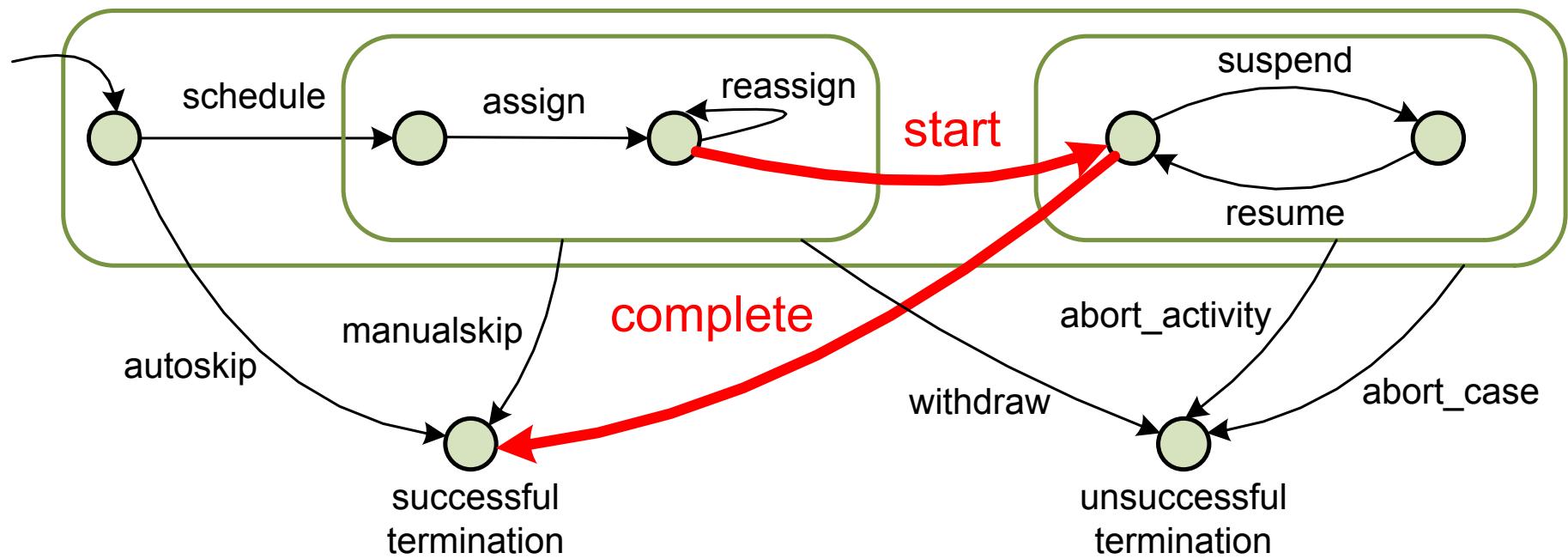
# Transactional model for activities



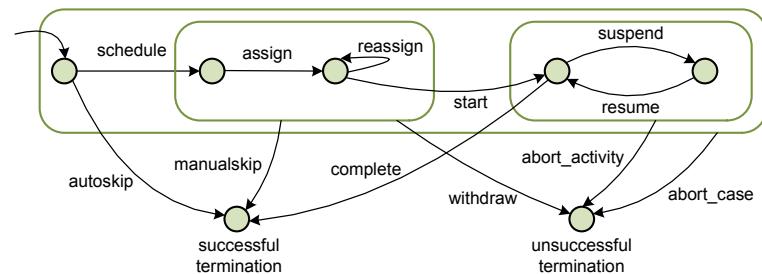
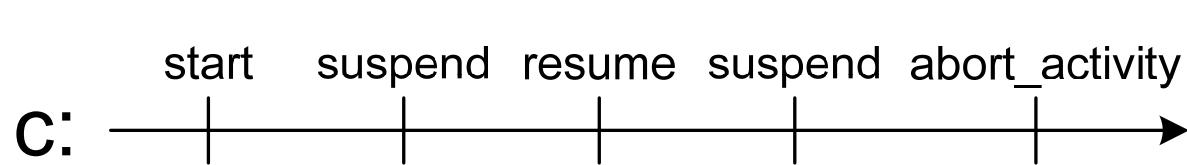
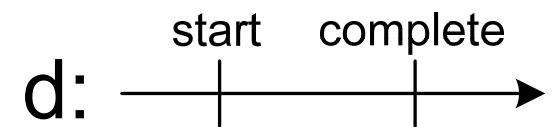
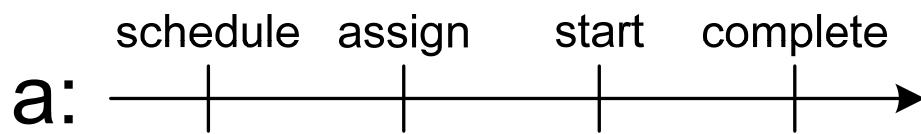
# Atomic activities



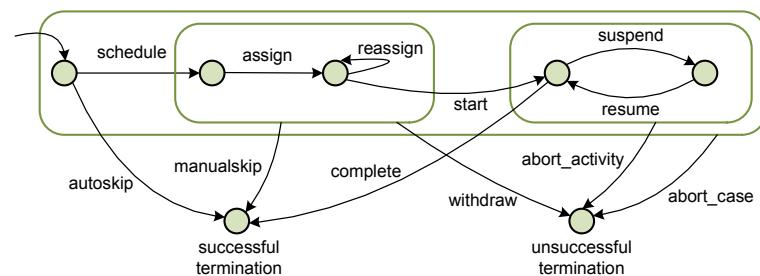
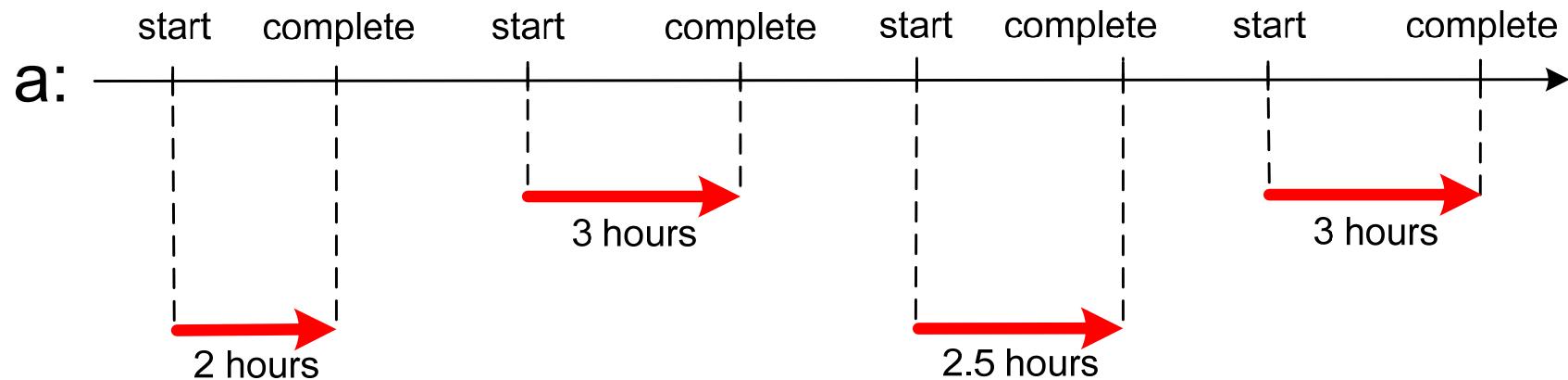
# Activities that have a duration



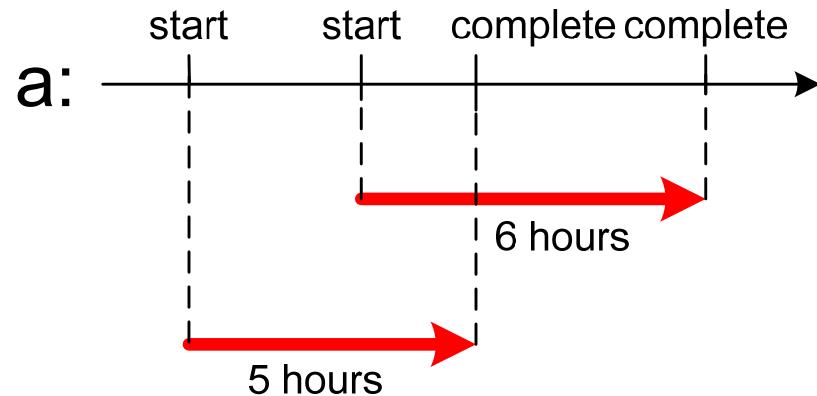
# Examples of activity instances



# Four instances of the same activity



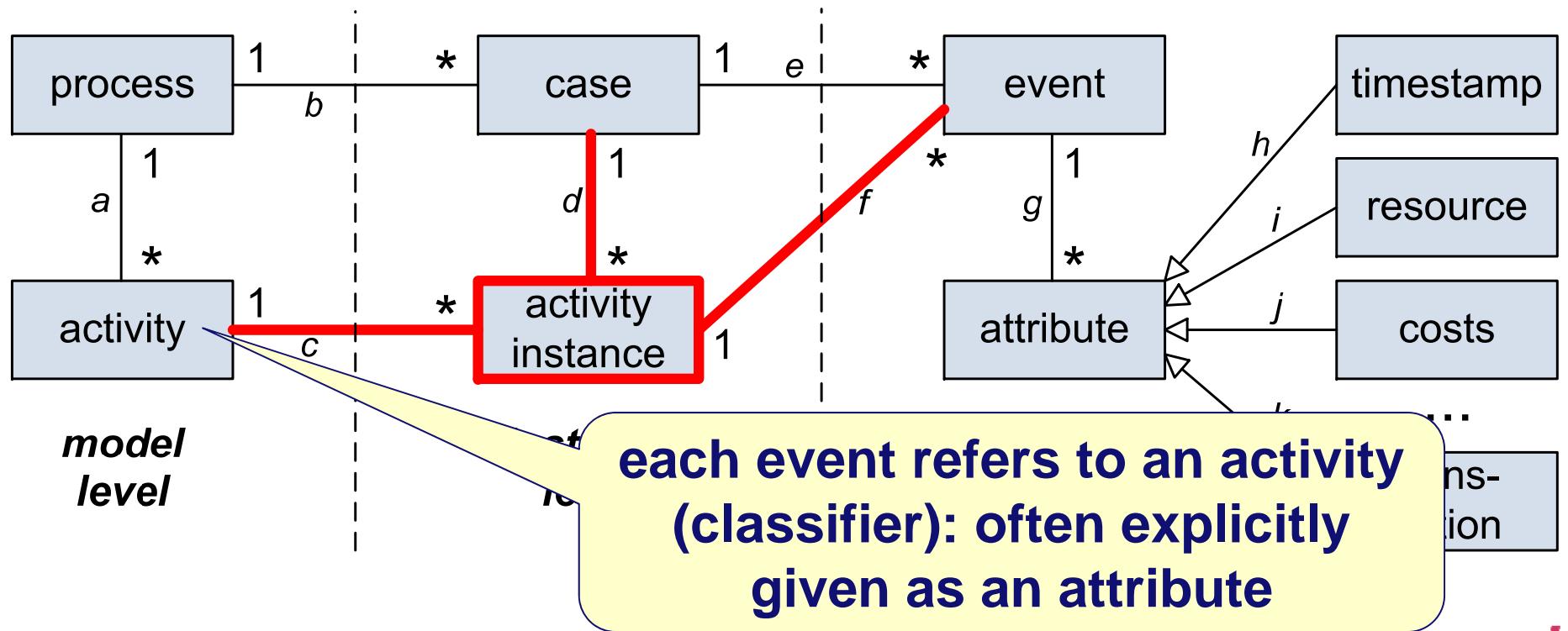
# Possible confusion



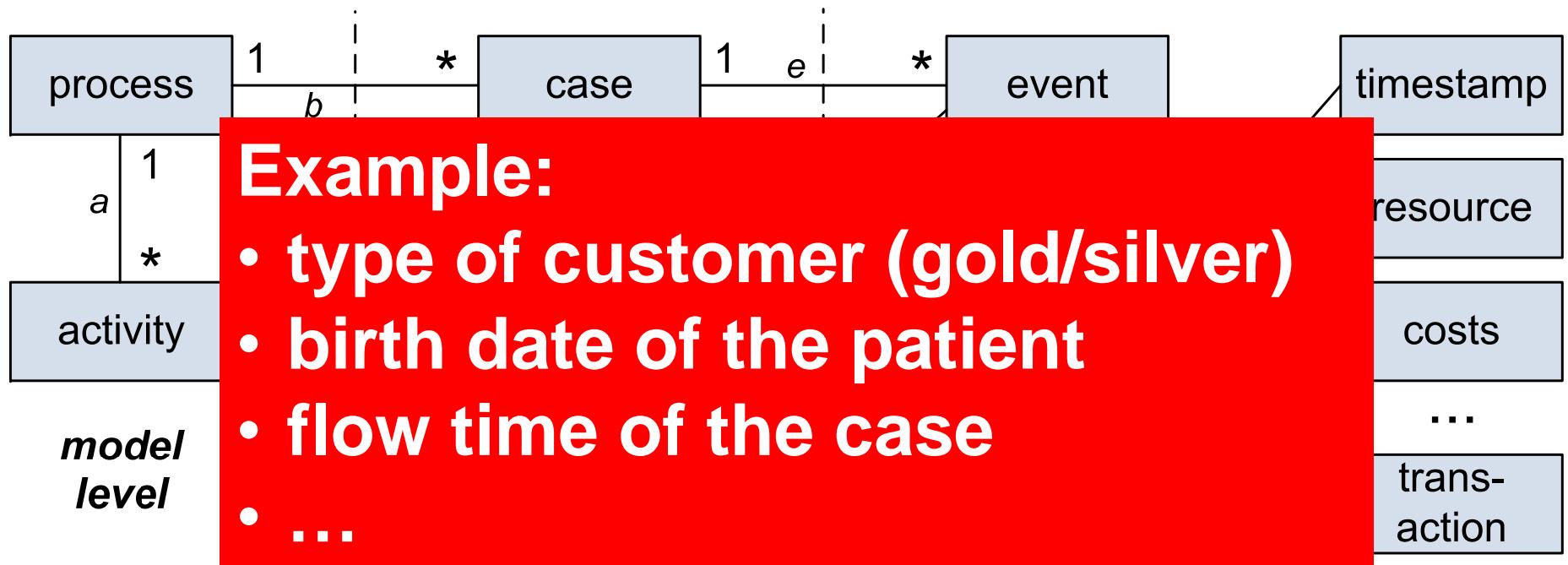
**Can be addressed in two ways:**

- a) Events explicitly refer to activity instances.**
- b) Activity instances are created/computed during replay (using heuristics).**

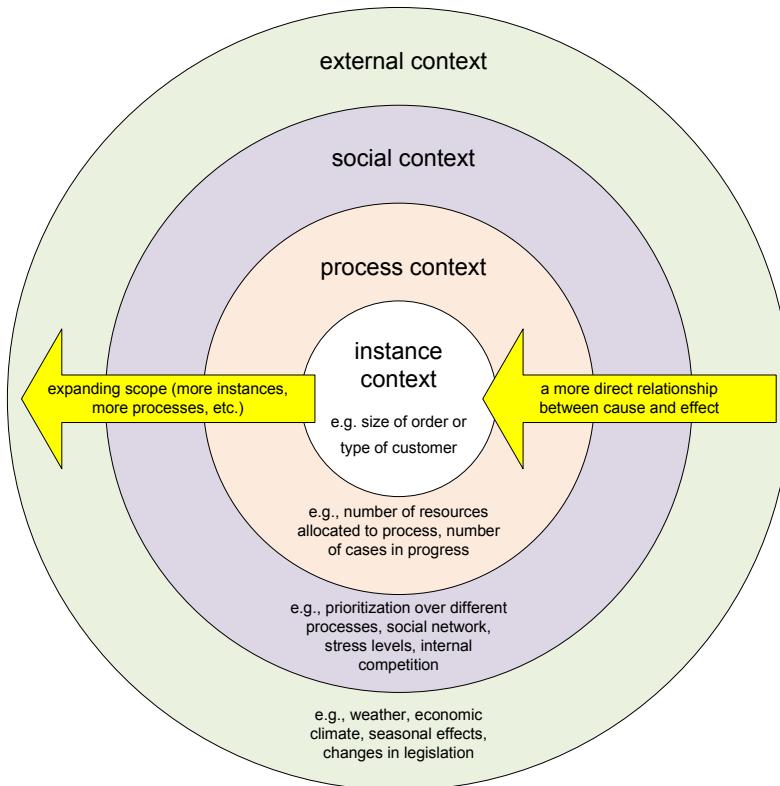
# Activity instances



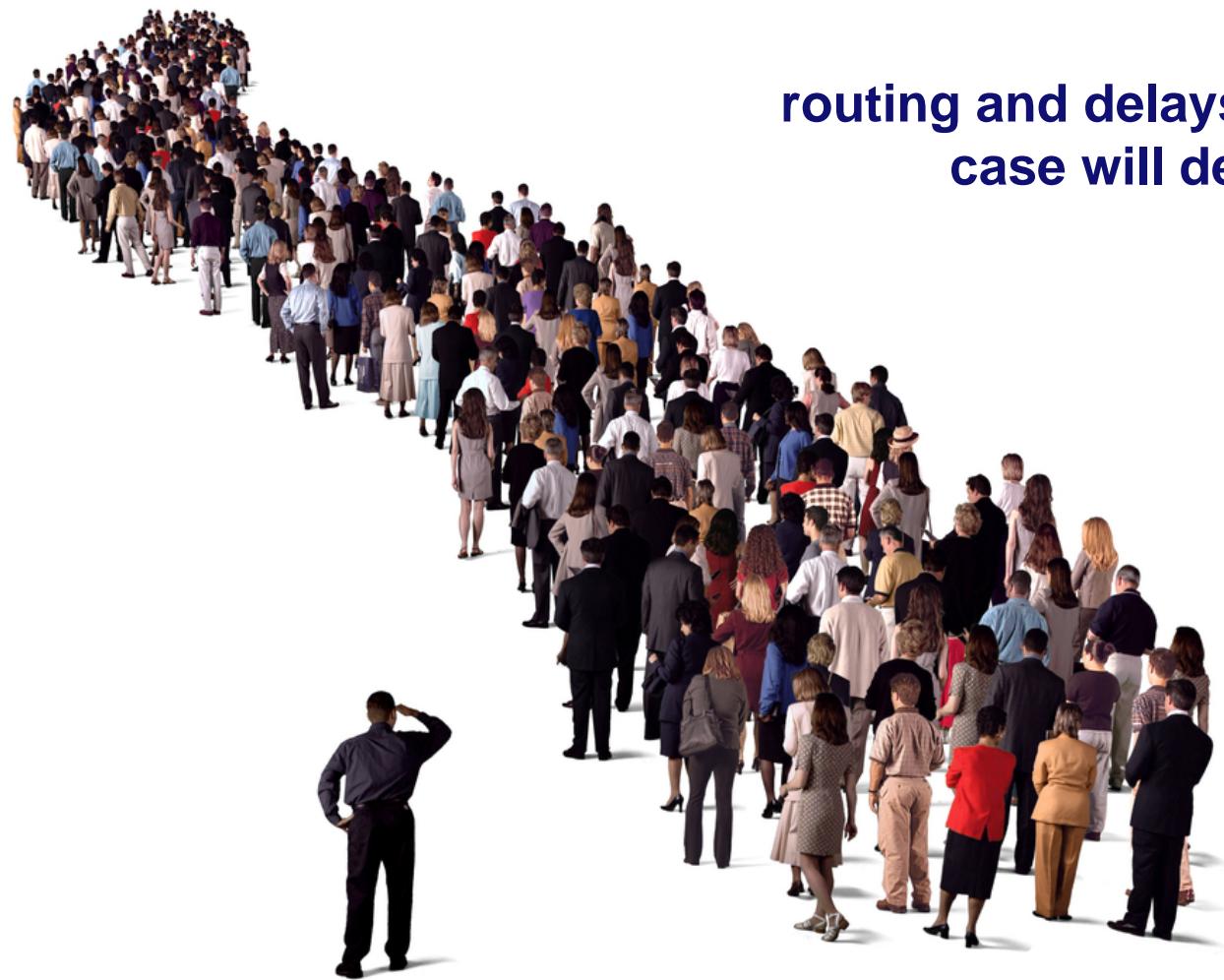
# Attributes: Also at case level or computed



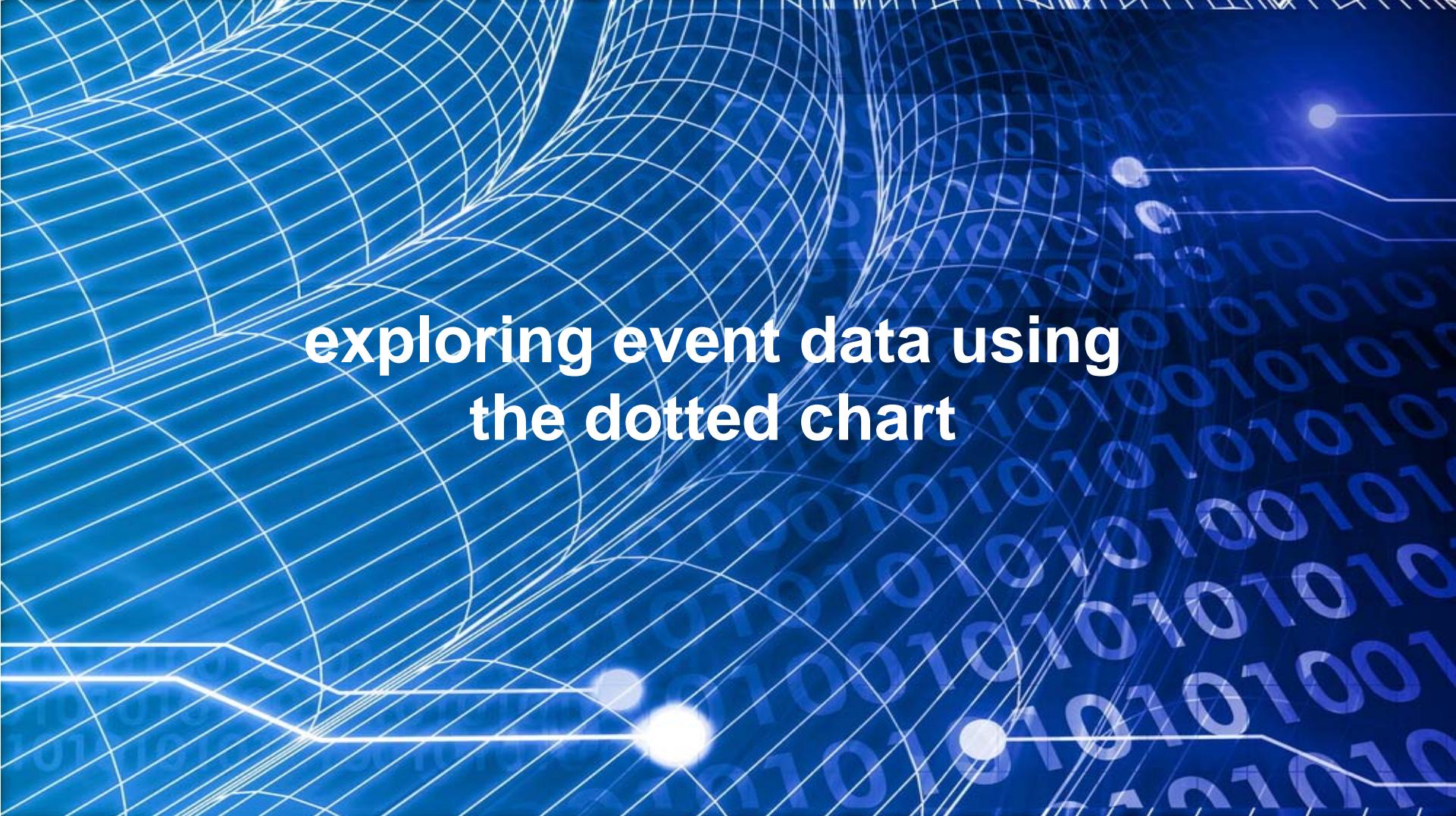
# Process mining in context



- **Cases and events cannot be seen in isolation:**
  - # pending cases
  - # resources available
  - weather
  - workload
- **Context can be computed or retrieved and added explicitly as derived attributes.**



**routing and delays of an individual case will depend on context**

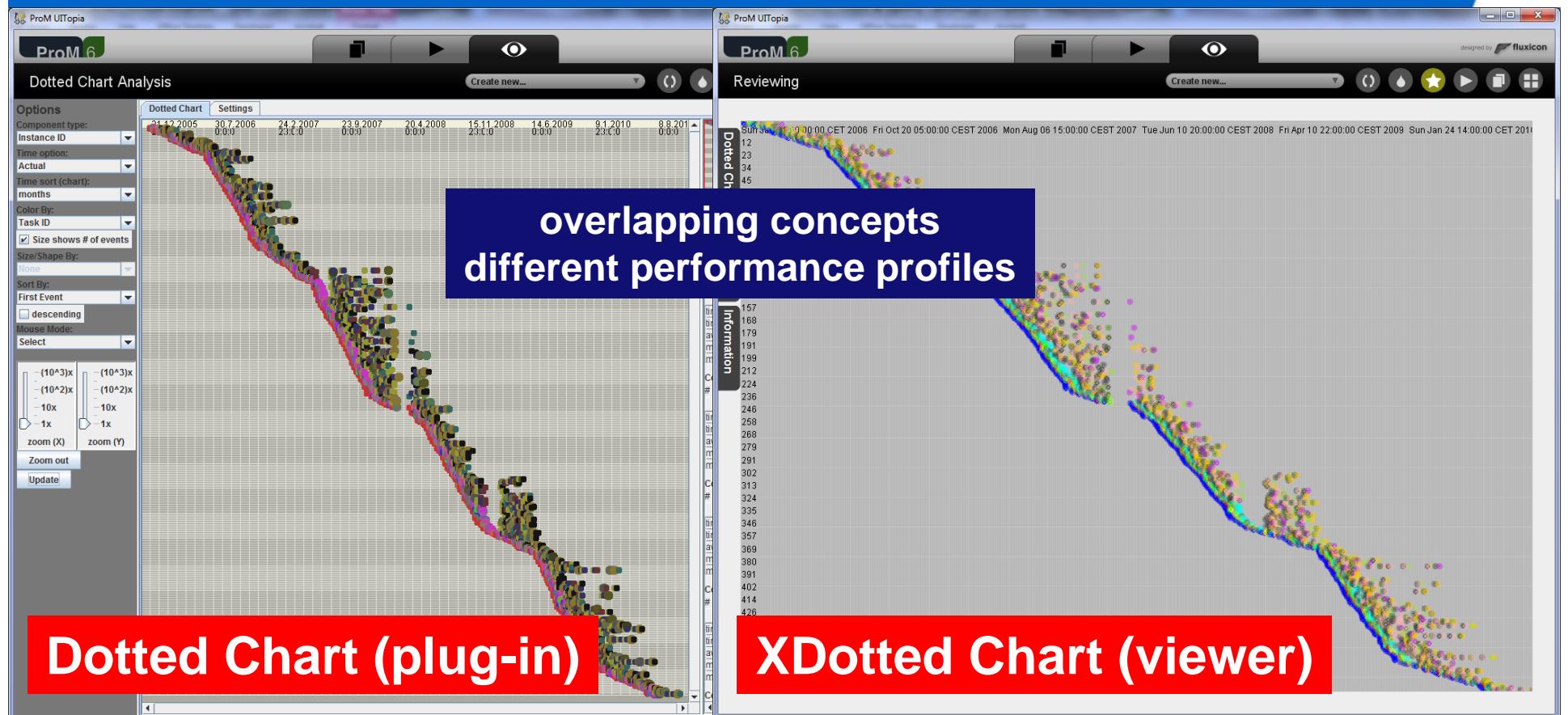


**exploring event data using  
the dotted chart**

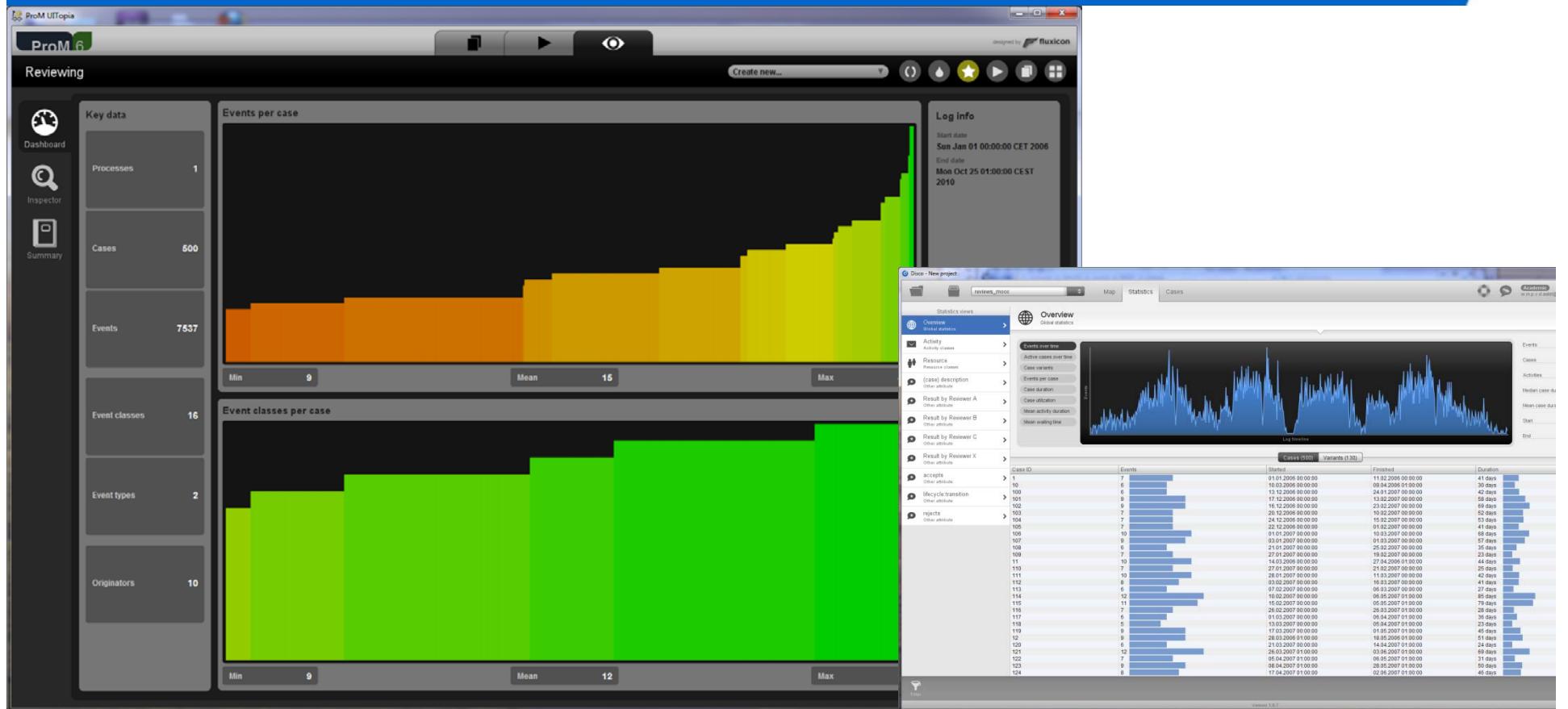
# Dotted chart: Main idea

- Helicopter view showing all events.
- There is a one-to-one correspondence between events and dots.
- Each event has a timestamp that is used to compute the X coordinate (absolute, relative, logical, ...).
- Another attribute is used for the Y coordinate, e.g., the case.
- Dots can be colored based on an attribute, e.g., resource or activity name.

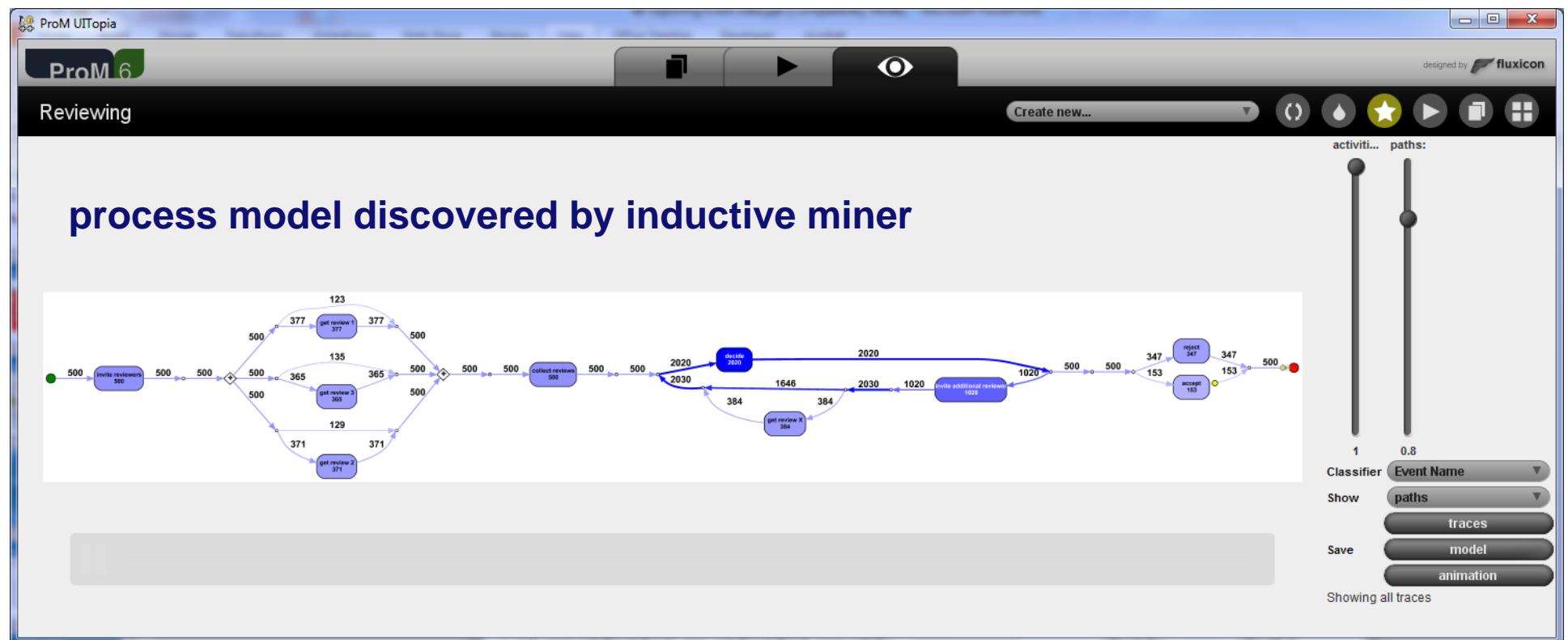
# Two variants in ProM



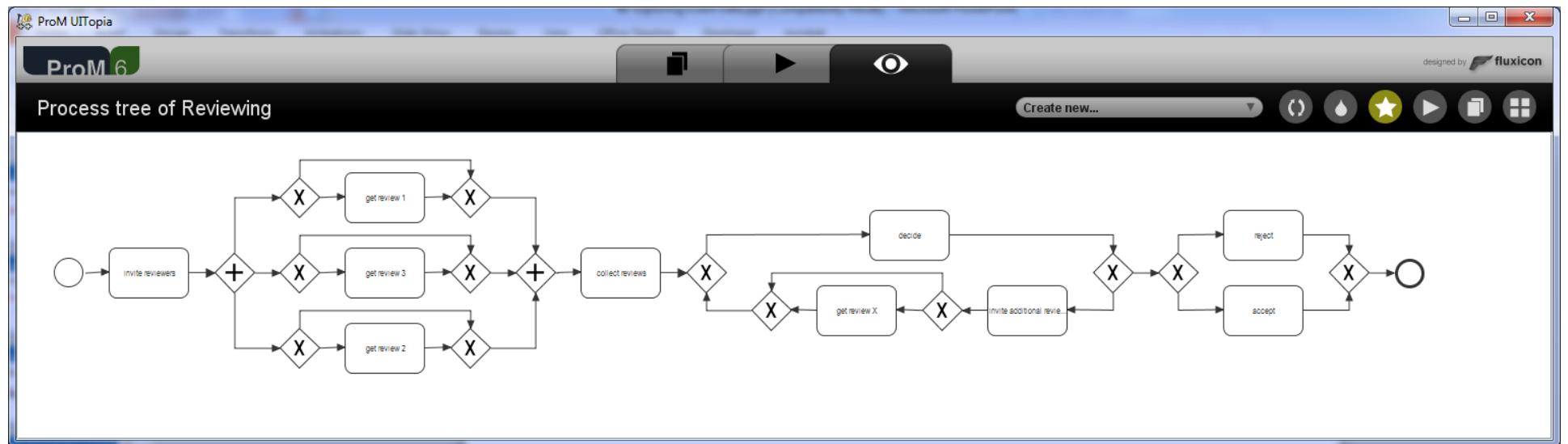
# Example event log (500 cases, 7537 events)



# Example event log (500 cases, 7537 events)

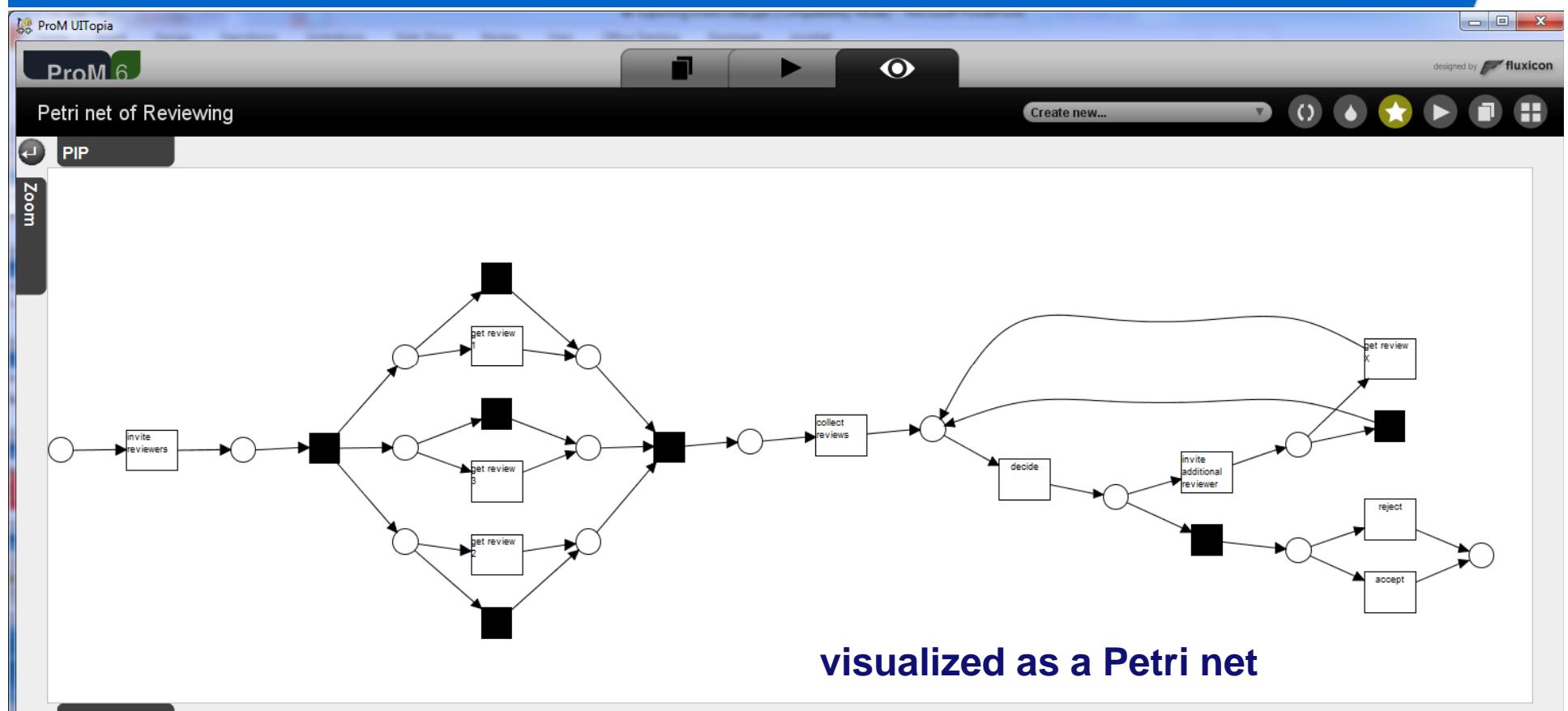


# Example event log (500 cases, 7537 events)

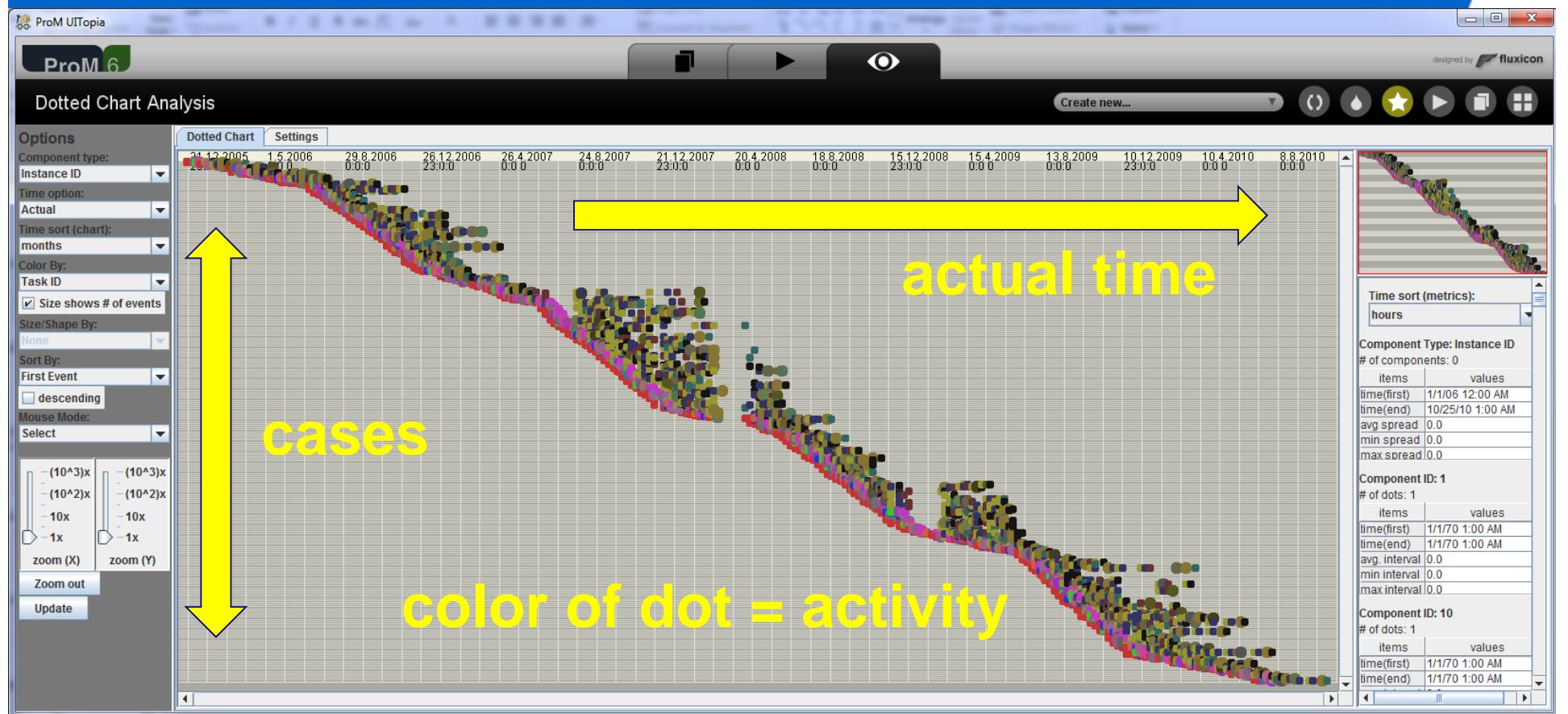


visualized as a BPMN model

# Example event log (500 cases, 7537 events)



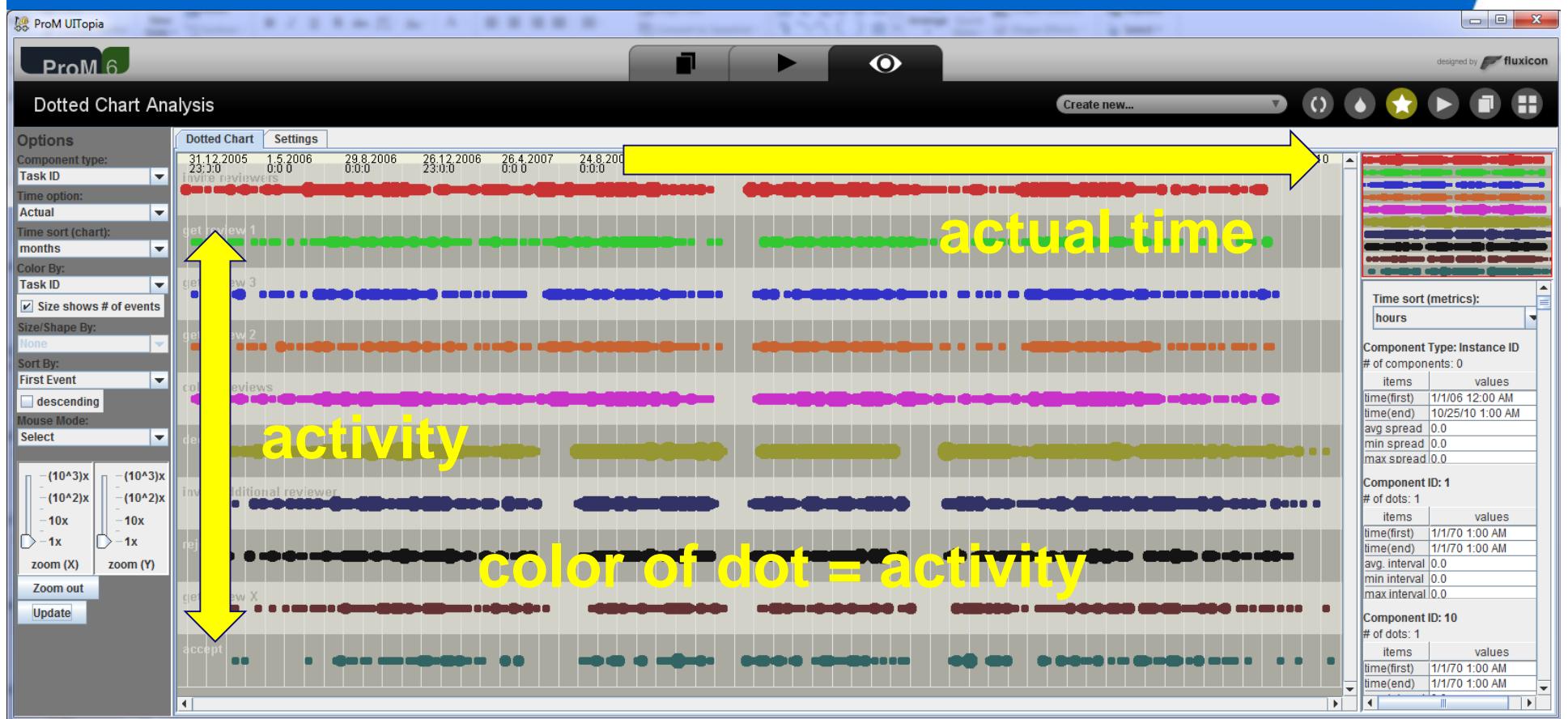
# Dotted chart: case id/absolute time



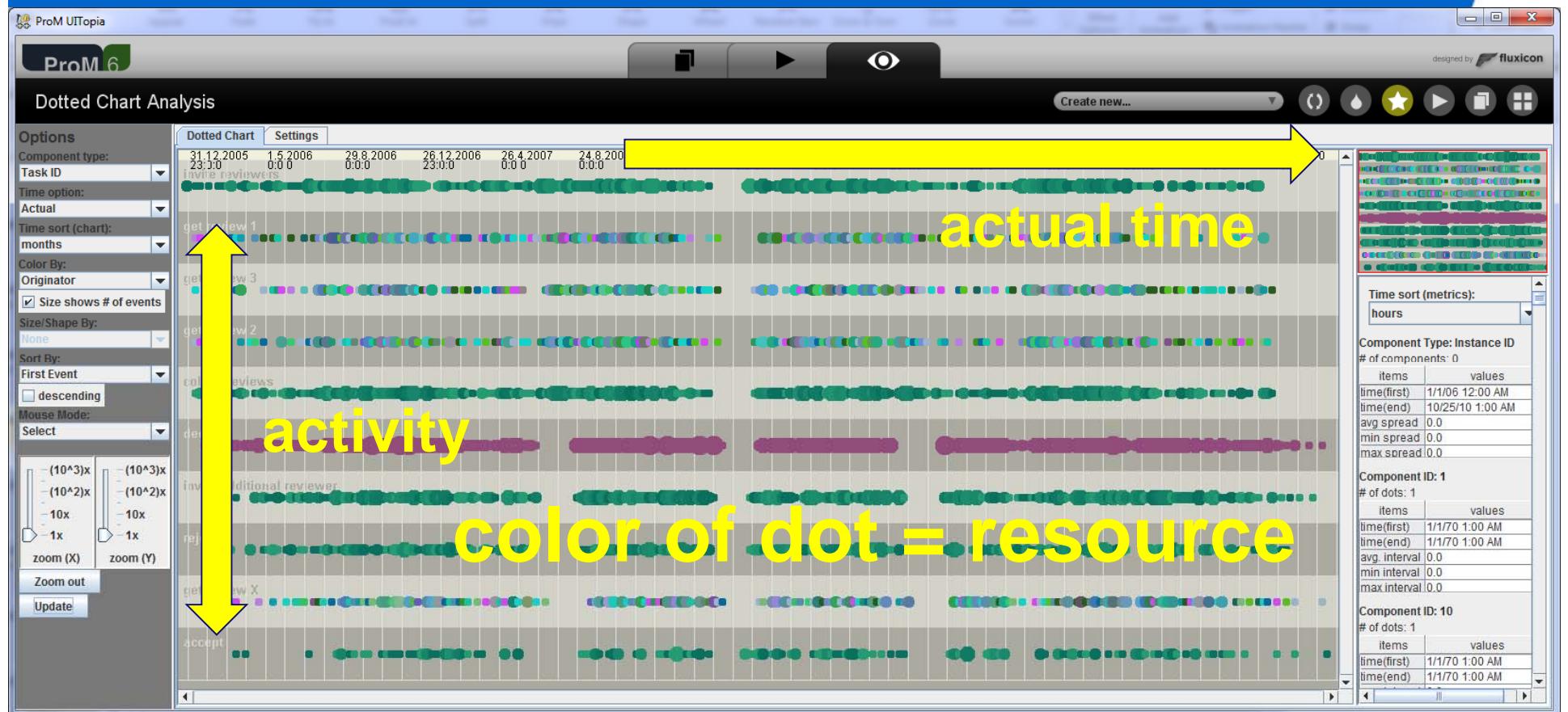
# One of 500 cases



# Dotted chart: activity/absolute time



# Dotted chart: activity/absolute time



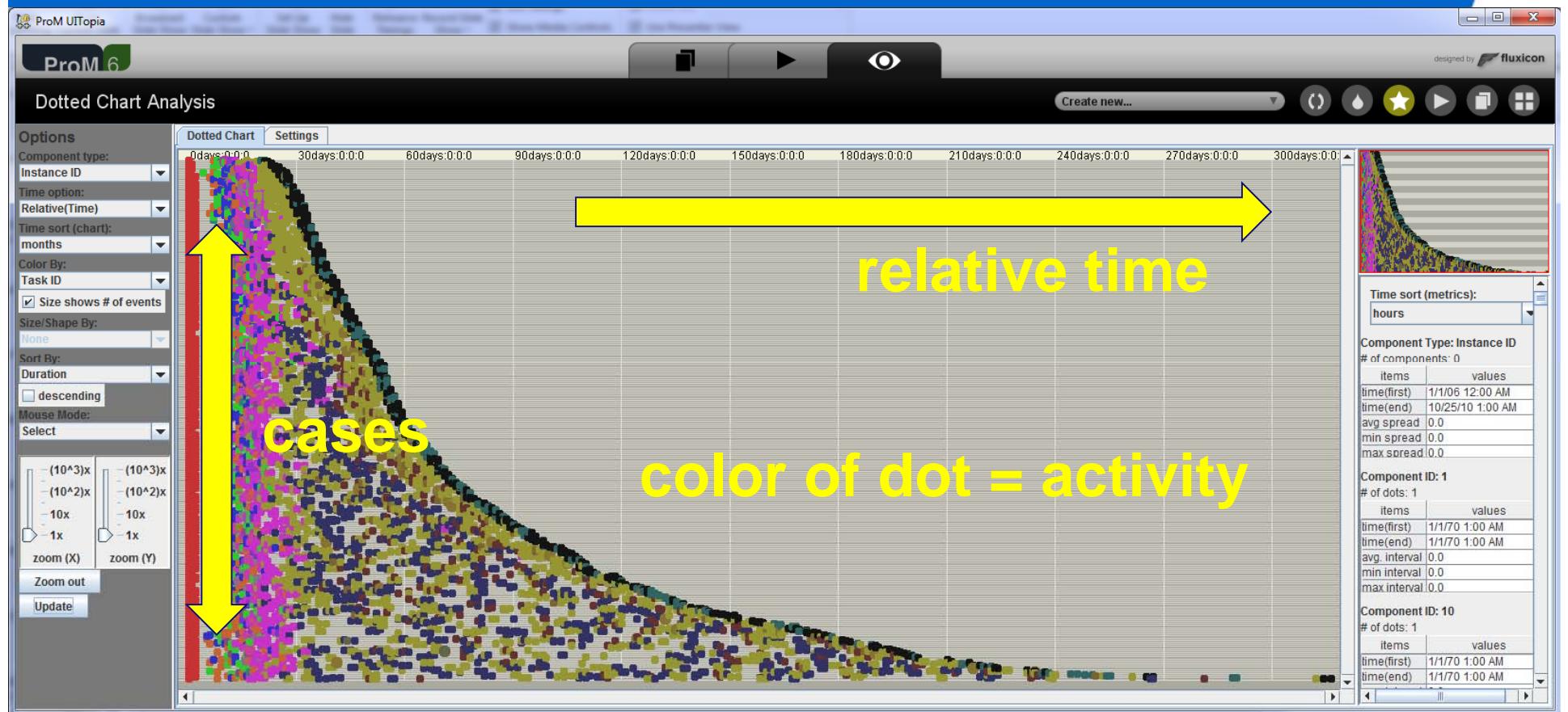
# Dotted chart: resource/absolute time



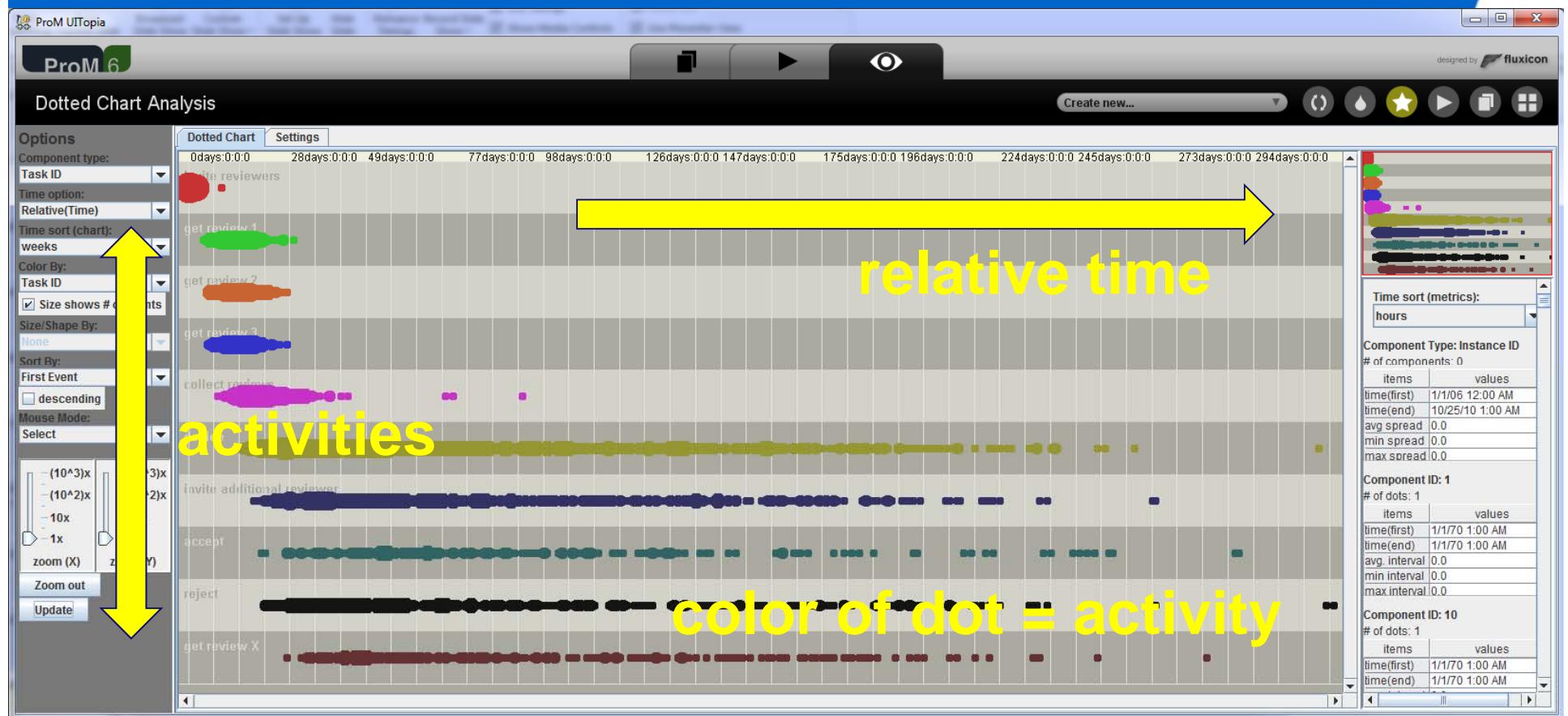
# Color of dots (activity or resource)



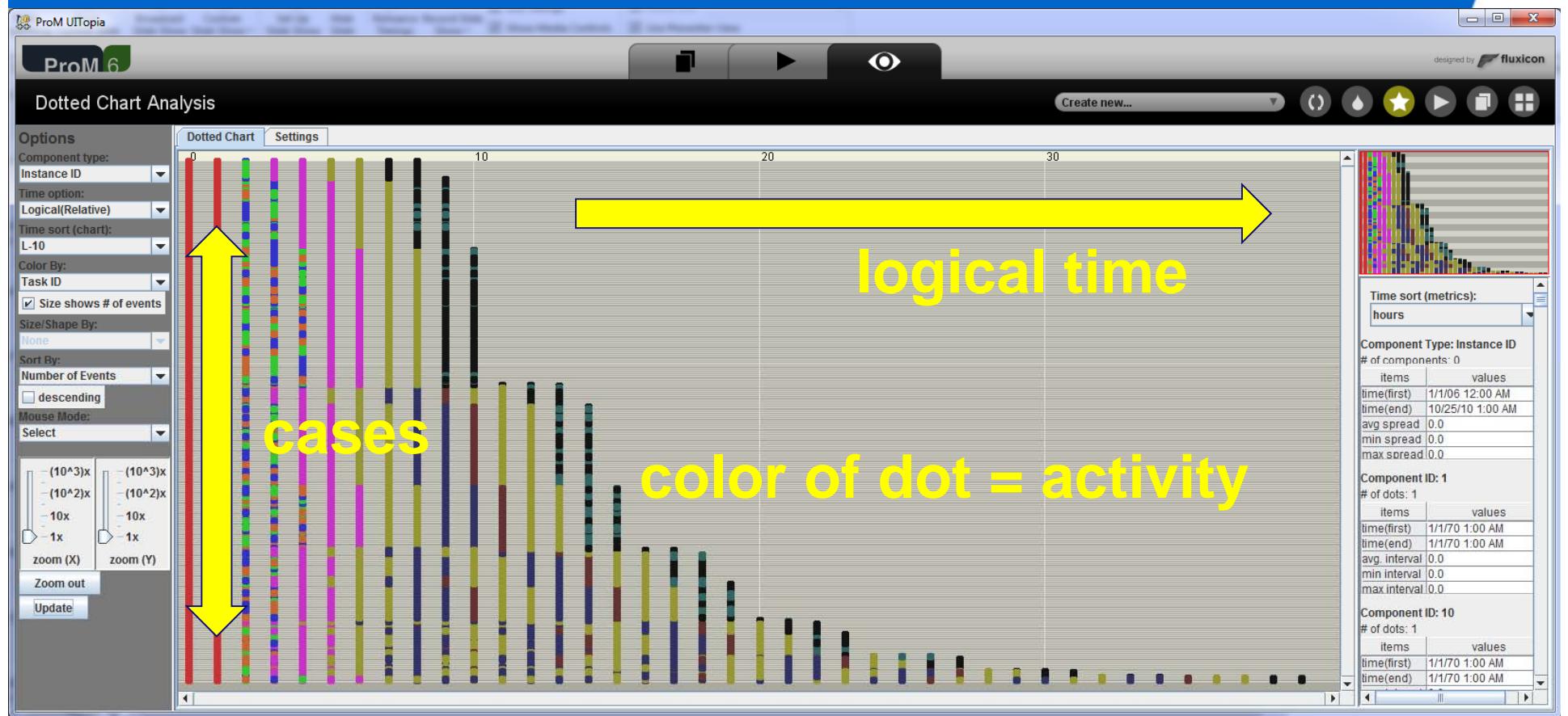
# Relative time (all cases start at time 0)



# Relative time (all cases start at time 0)

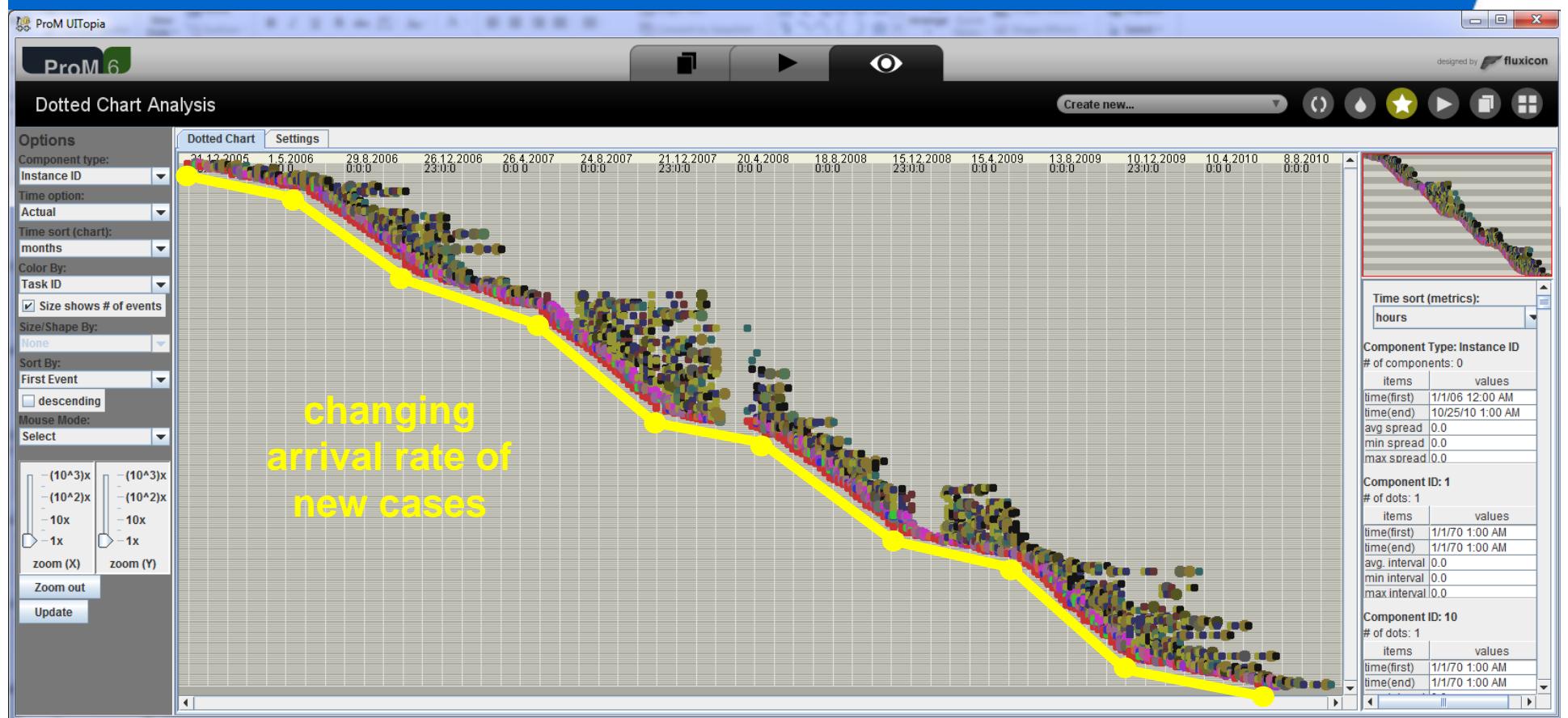


# Logical time (all cases start with step 1)

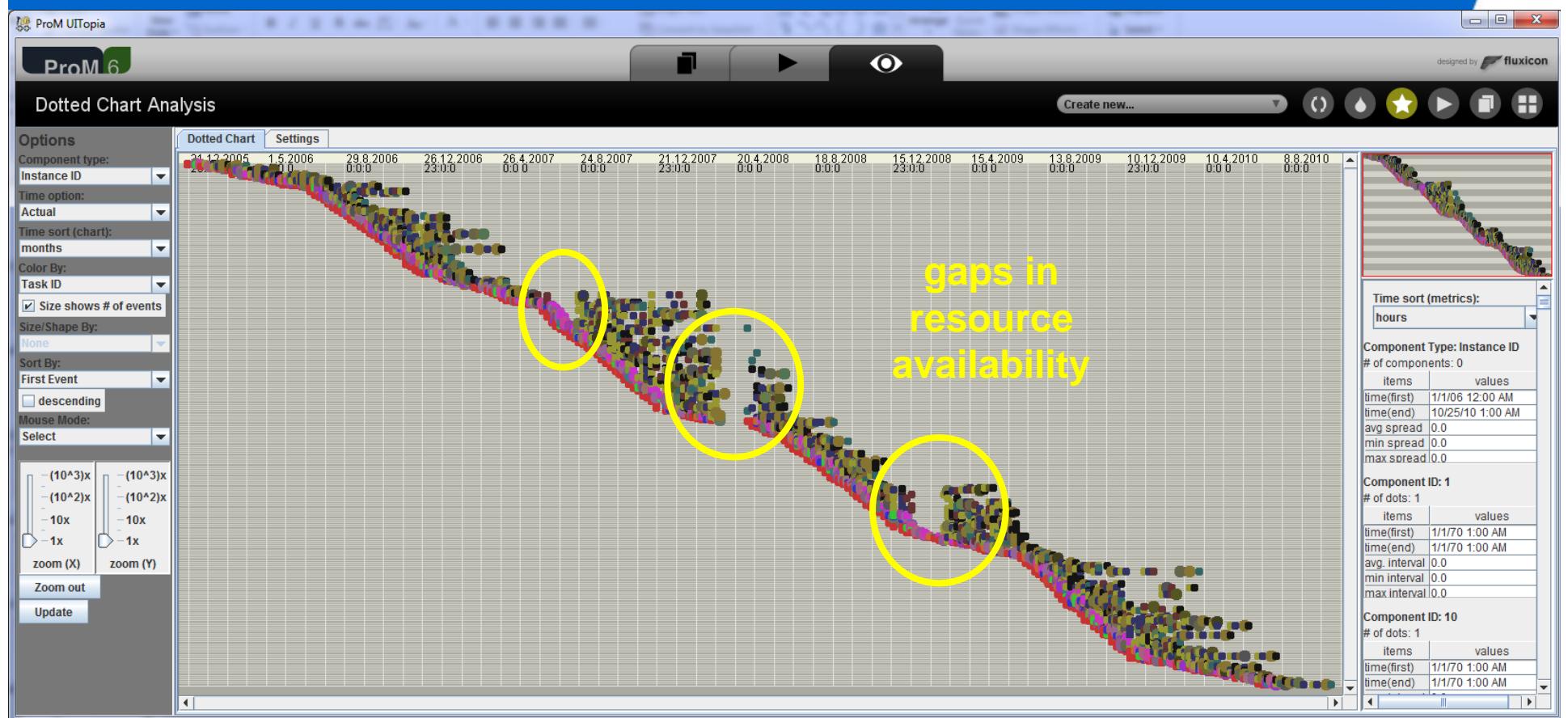


**so what can we see ...**

# What can you see?



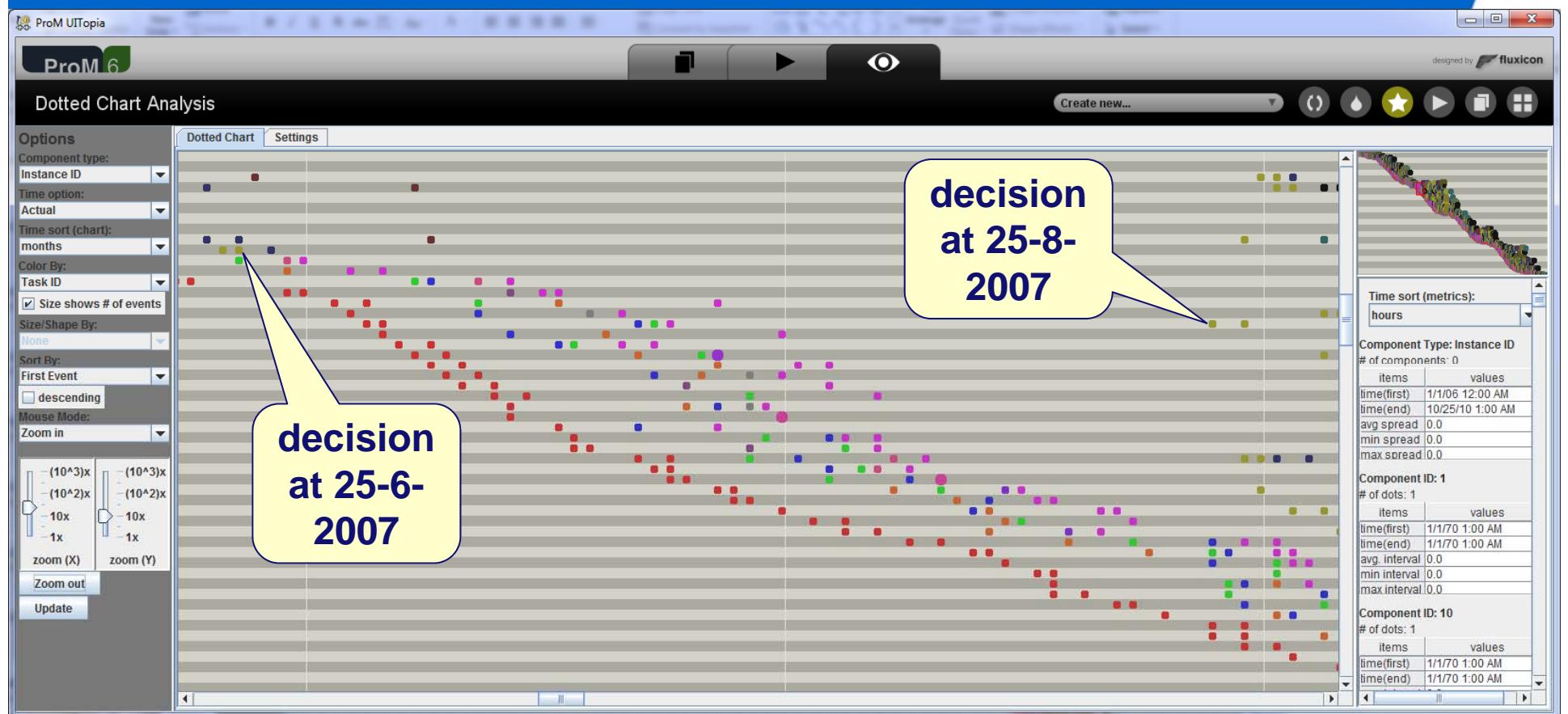
# What can you see?



# Let us focus on the first gap



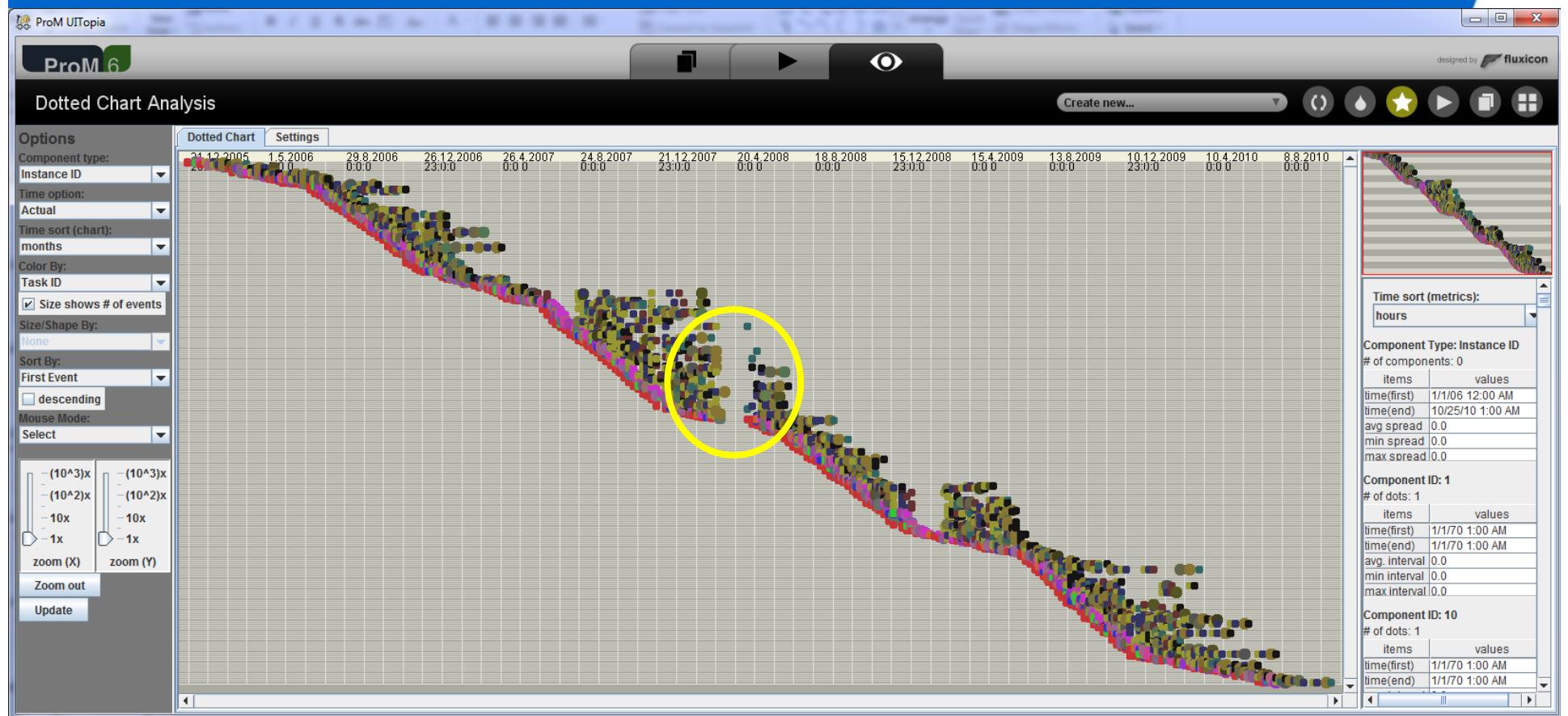
# Two month period without any decisions



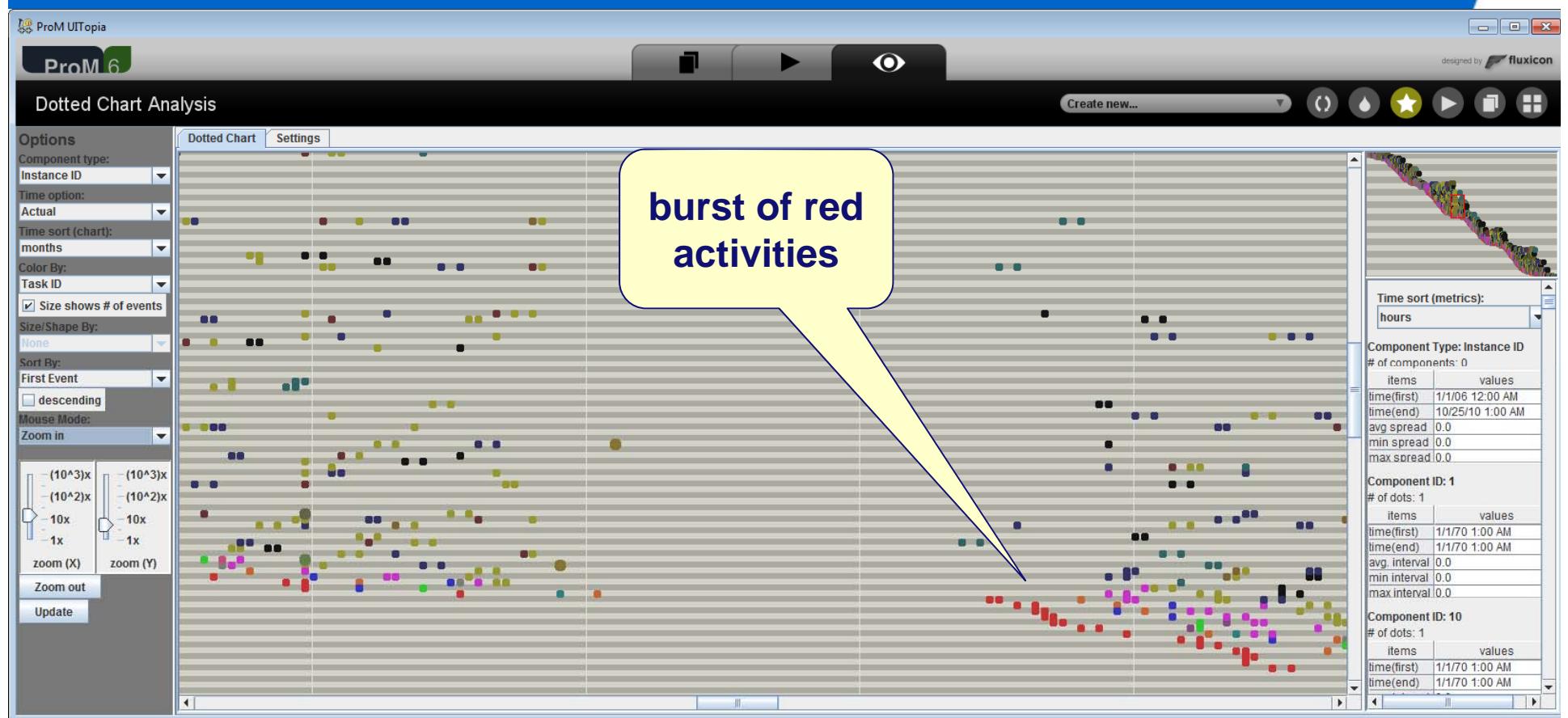
# Can be explained by absence of resource



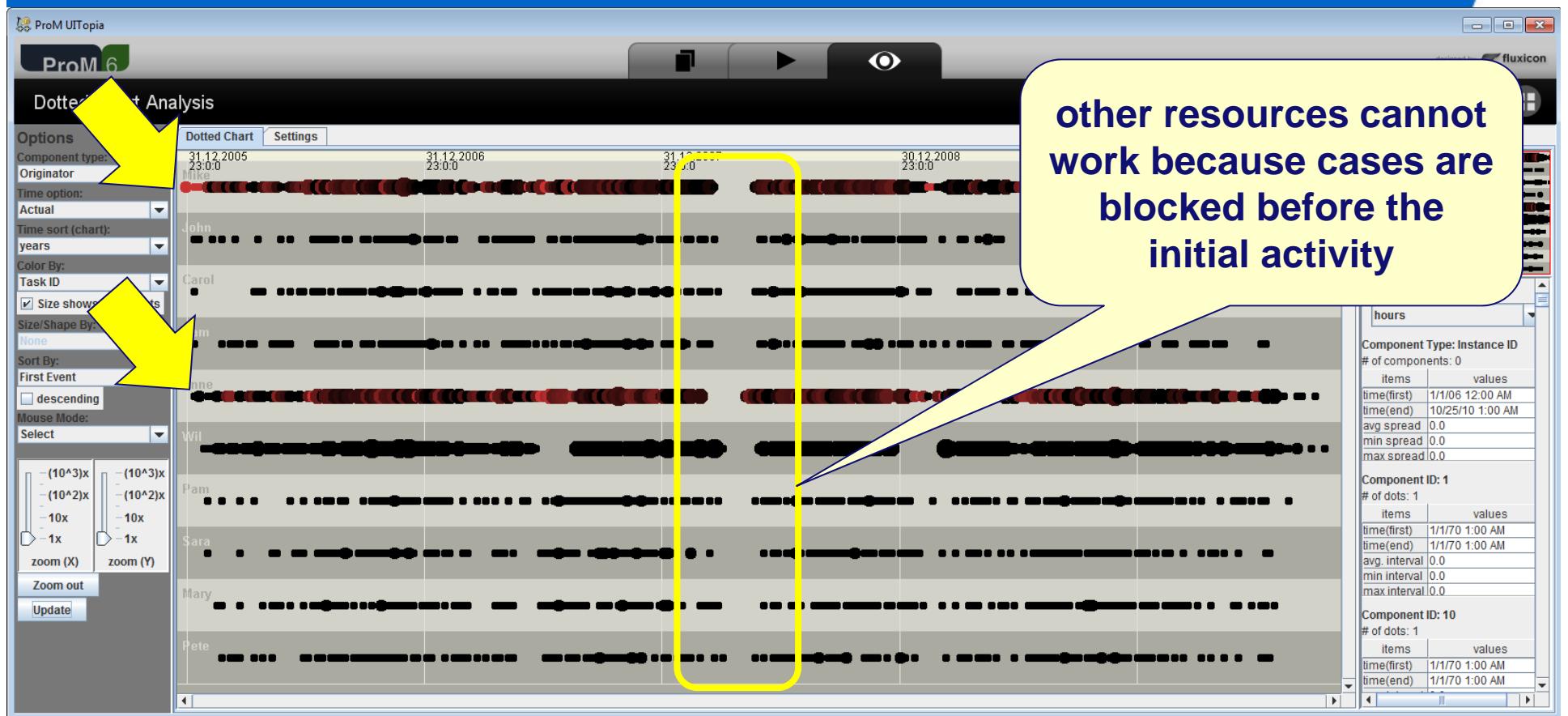
# Let us focus on the second gap



# Six week of limited activity



# Two resources execute the red initial activity



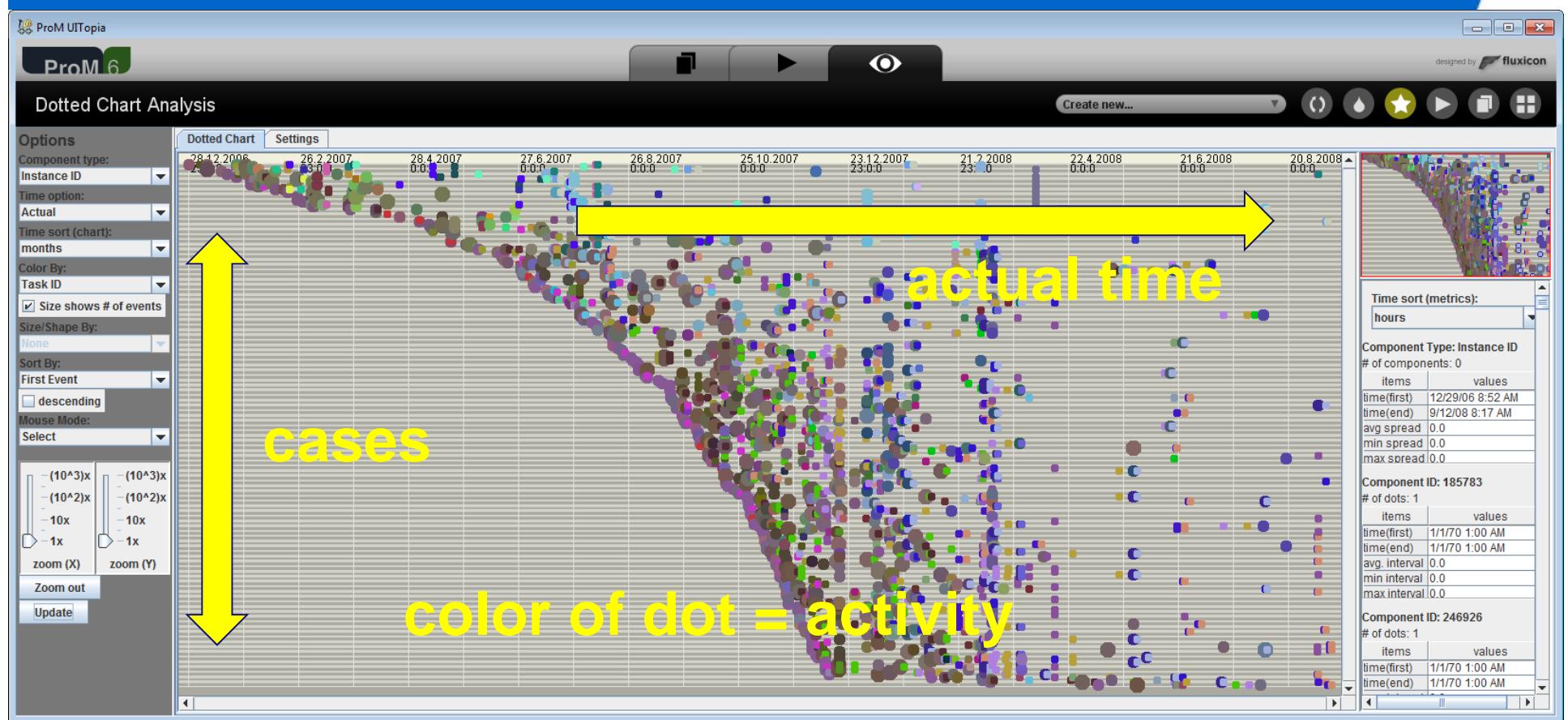
# Let us focus on the third gap



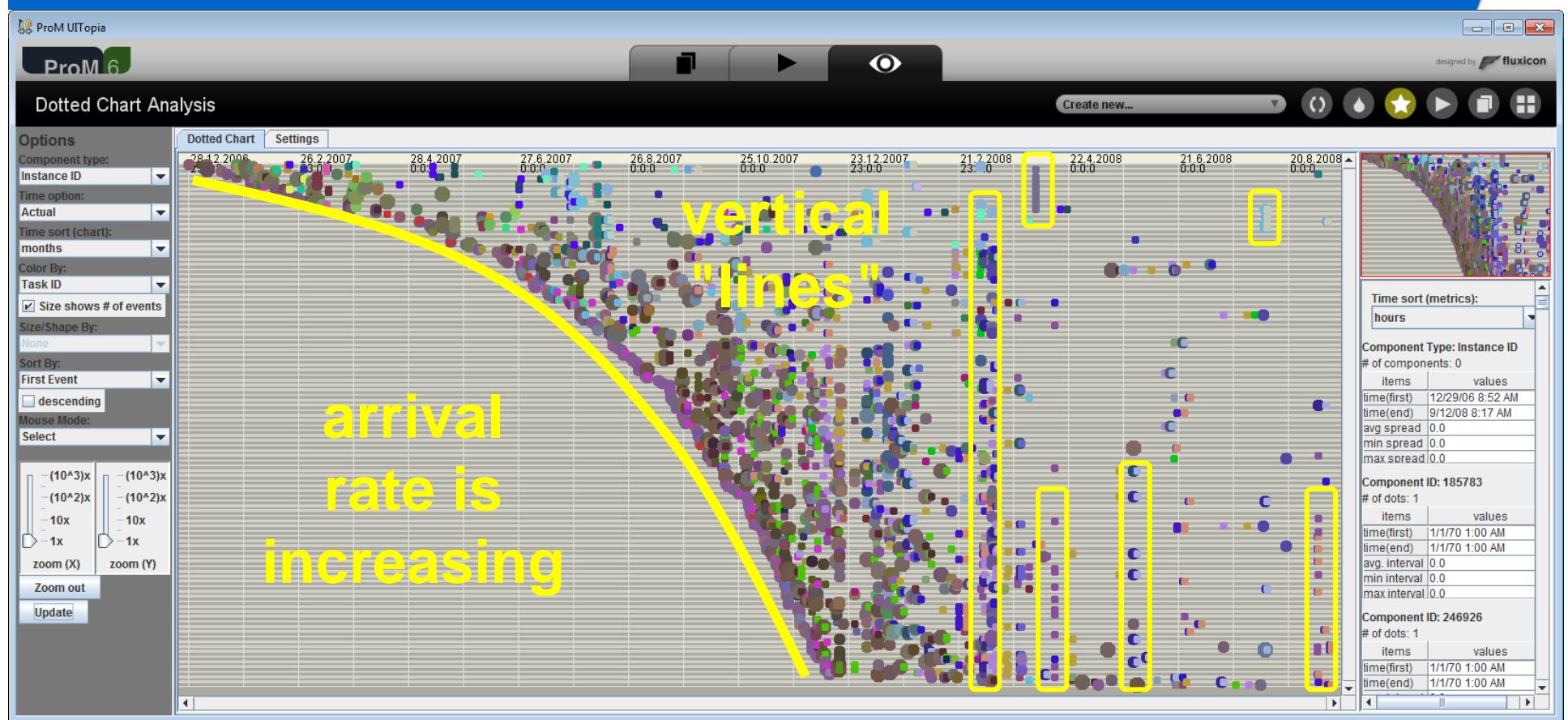
# Same problem as for first gap: longer period without decisions



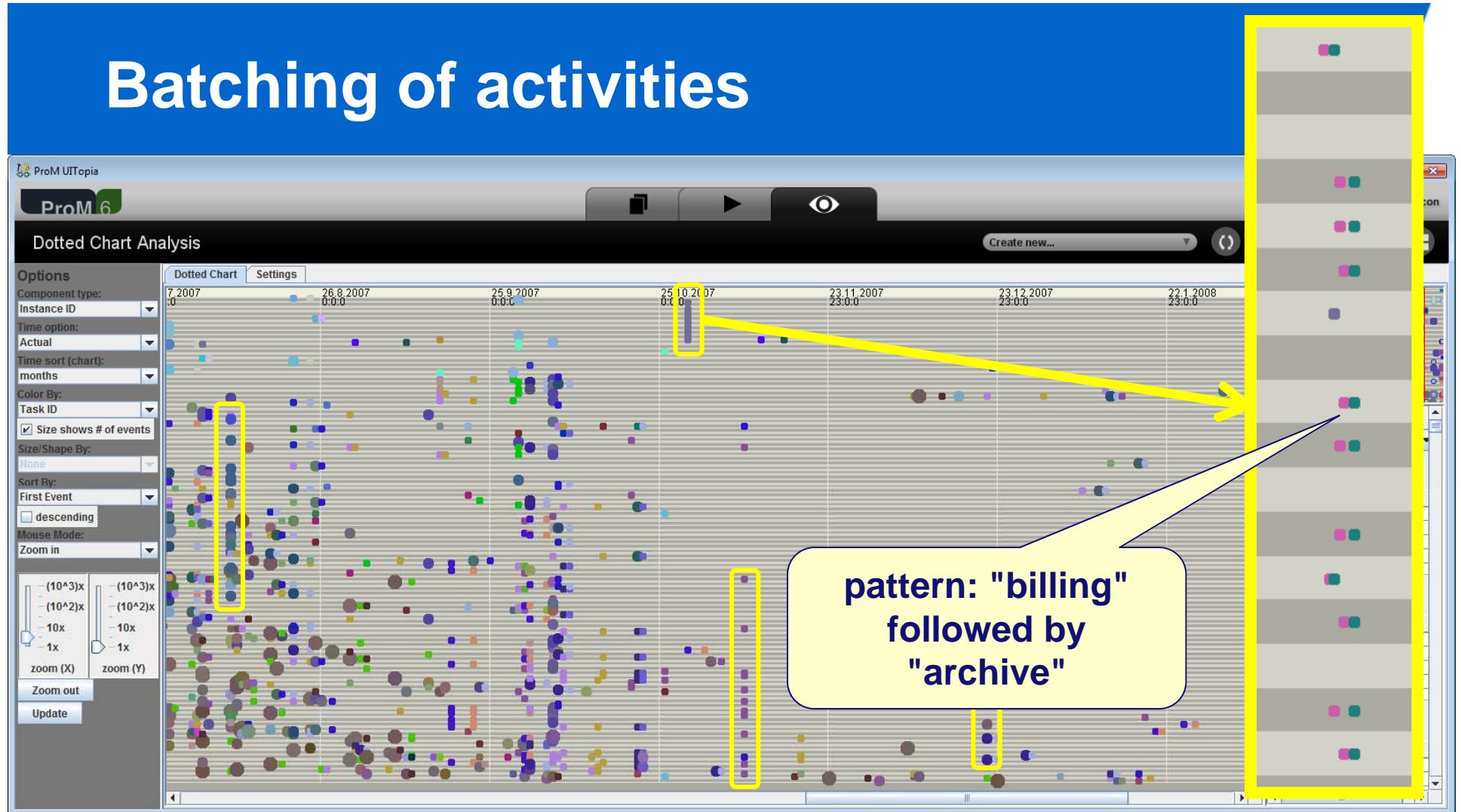
# Event log from Dutch housing agency

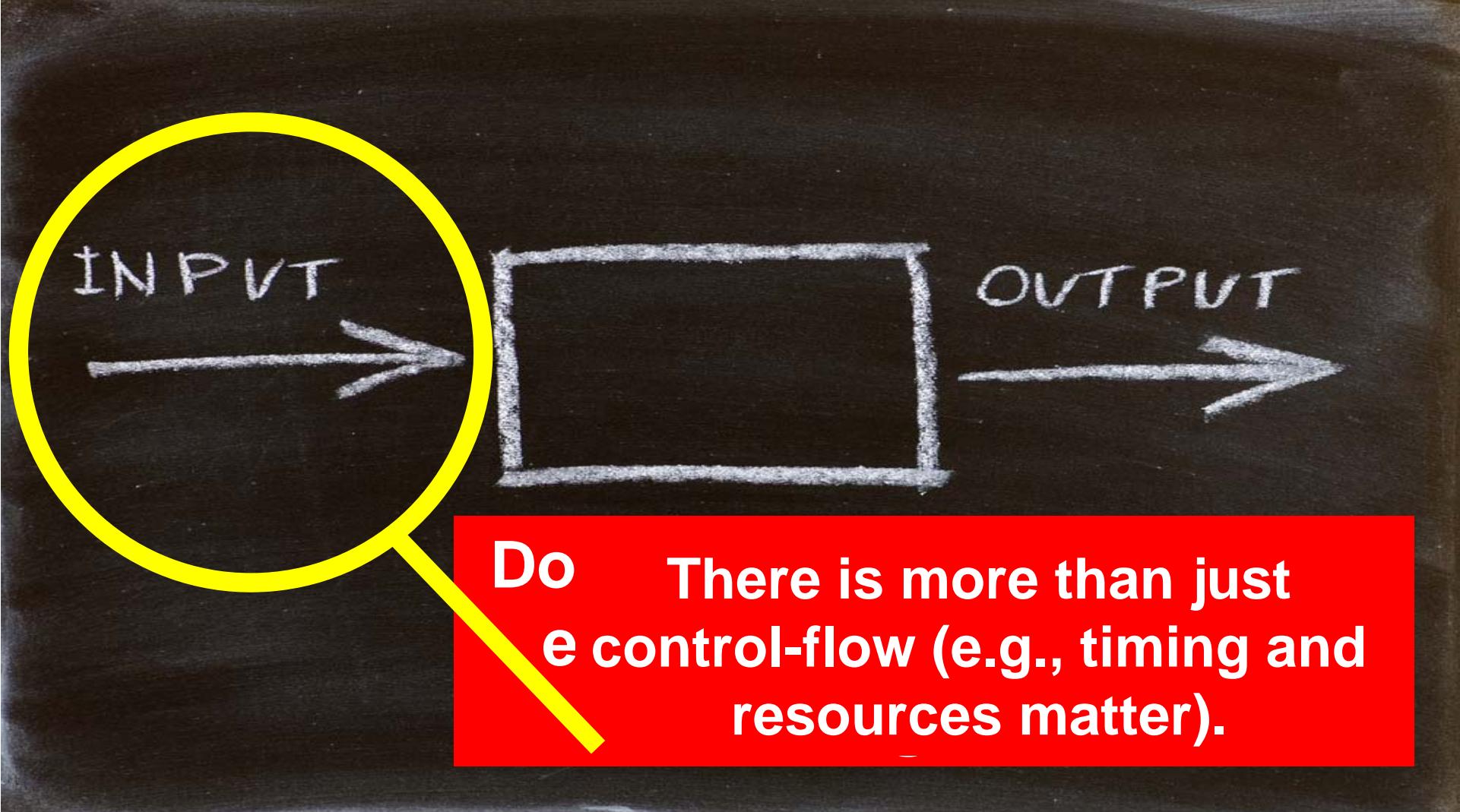


# What can you see?



# Batching of activities





INPUT



OUTPUT

Do There is more than just  
the control-flow (e.g., timing and  
resources matter).

<i>Part I: Preliminaries</i>		
<b>Chapter 1</b> Introduction	<b>Chapter 2</b> Process Modeling and Analysis	<b>Chapter 3</b> Data Mining
<i>Part III: Beyond Process Discovery</i>		
<b>Chapter 7</b> Conformance Checking	<b>Chapter 8</b> Mining Additional Perspectives	<b>Chapter 9</b> Operational Support

## *Part II: From Event Logs to Process Models*

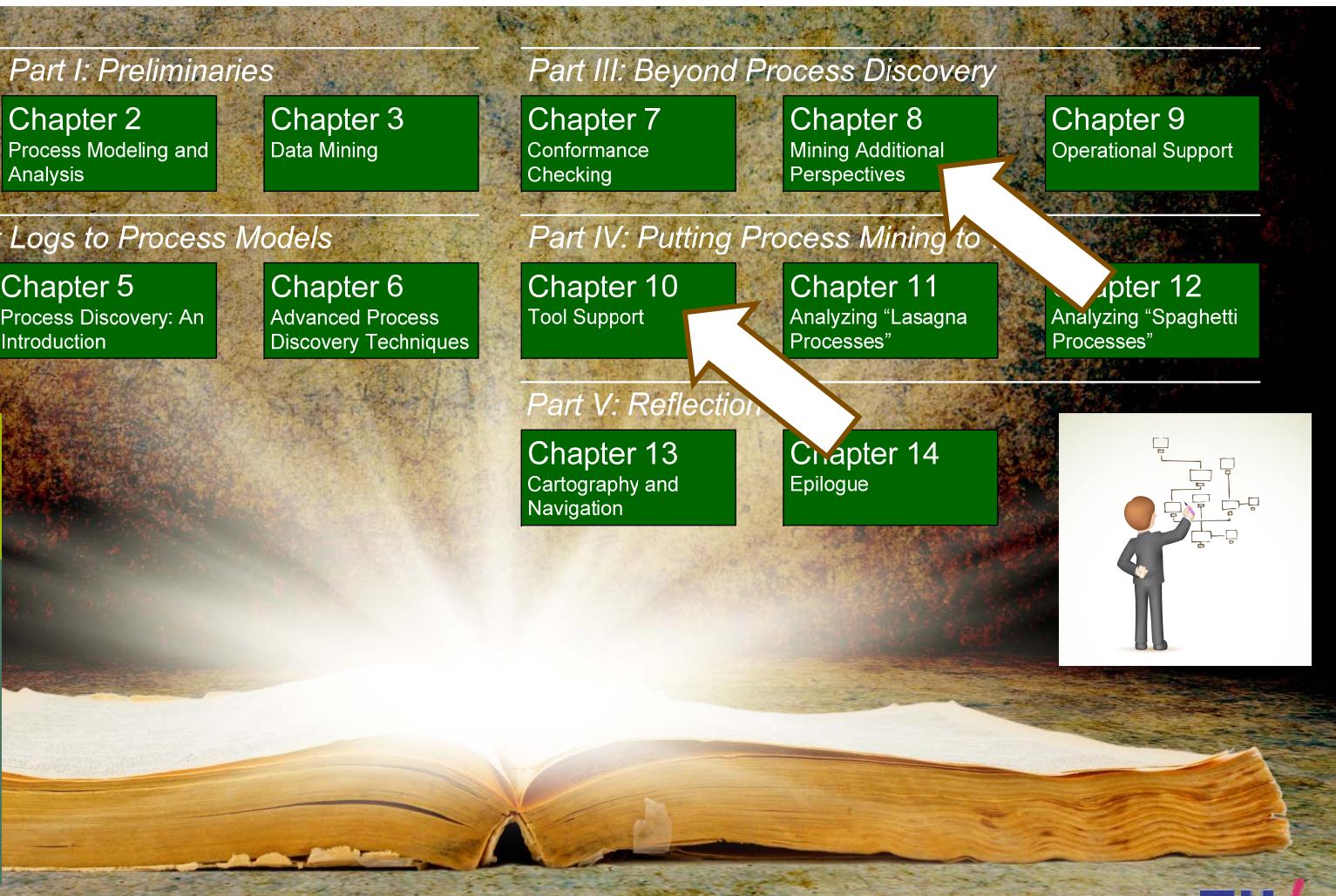
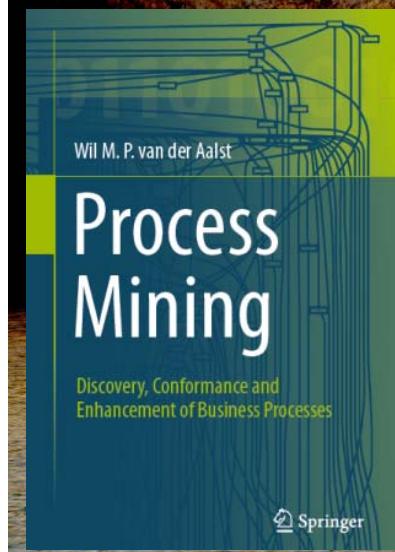
<b>Chapter 4</b> Getting the Data	<b>Chapter 5</b> Process Discovery: An Introduction	<b>Chapter 6</b> Advanced Process Discovery Techniques
--------------------------------------	--	---

## *Part IV: Putting Process Mining to Work*

<b>Chapter 10</b> Tool Support	<b>Chapter 11</b> Analyzing “Lasagna Processes”	<b>Chapter 12</b> Analyzing “Spaghetti Processes”
-----------------------------------	--	--

## *Part V: Reflection*

<b>Chapter 13</b> Cartography and Navigation	<b>Chapter 14</b> Epilogue
---	-------------------------------



©Wil van der Aalst & TU/e (use only with permission & acknowledgements)