



MASTER 1 OF COMPUTER SCIENCE
TER PROJECT REPORT

Analysis of daily activities and
establishment of processes models

Students:

Mohamed Ali SIDI YAHYA
Marwan BENABDALLAH

Supervisor:

Yougourthene MEDDOURI

College year: 2018-2019

Abstract

Process mining is an important sub-field of Data Science, currently with the explosion of significant amounts of data and the trend to analyze these data, companies and organizations are interested into using process-based techniques such as process mining to analyze their working processes in order to optimize , generalize and predict models that can improve their efficiency and effectiveness. Our project is about applying process mining in R using Bupar integrated suite of R-packages for the handling and the analysis of business process data.To optimize the daily activities processes and to help the company SAYNOVA to work with these tools, we have also used ProMTools to complete functionalities of Bupar suite.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 5 |
| 2 | Data set | 6 |
| 3 | Related Work | 7 |
| 4 | Preliminaries: Process Mining Techniques | 8 |
| 4.1 | Structuredness | 9 |
| 5 | Process Discovery | 9 |
| 5.1 | Comparisons between individuals | 10 |
| 5.1.1 | Entry and Exit points | 10 |
| 5.1.2 | Precedence Matrix | 11 |
| 5.1.3 | Process Variants | 12 |
| 5.1.4 | Performance Analysis | 13 |
| 5.1.5 | Time Analysis | 17 |
| 5.1.6 | Filtering | 20 |
| 5.2 | Comparisons between Working days and weekends | 20 |
| 5.2.1 | Entry and Exit points | 20 |
| 5.2.2 | Process Variants | 22 |
| 5.2.3 | Performance Analysis | 24 |
| 5.2.4 | Time Analysis | 28 |
| 5.3 | Generalization of all processes models | 30 |
| 6 | Conformance Checking | 30 |
| 7 | Conclusion | 31 |
| A | Annexe | 33 |

List of Figures

| | | |
|----|--|----|
| 1 | Optimization of the daily routine to achieve a healthier life. | 5 |
| 2 | Collector and labeling framework: Wear App (smart-watch, 1) and Hand App (smart-phone, 2). The positions of the devices may vary | 7 |
| 3 | Start point for hh102 | 10 |
| 4 | Start point for hh104 | 10 |
| 5 | Exits points for hh102 | 11 |
| 6 | Trace explorer hh110 | 12 |
| 7 | Trace length of hh102 , hh104 and hh110 | 13 |
| 8 | The dotted chart of hh104 for activities linked to eating | 14 |
| 9 | The dotted chart of hh102 for activities linked to eating | 15 |
| 10 | Dotted chart sleeping habit of hh110 | 16 |
| 11 | Dotted chart sleeping habit of hh104 | 17 |
| 12 | Throughput_time for hh102 , hh110 and hh104 | 18 |
| 13 | Processing time for hh102 | 18 |
| 14 | Processing time for hh104 | 19 |
| 15 | Idle time for hh102 , hh104 and hh110 | 19 |
| 16 | Starting activities working days | 21 |
| 17 | Starting activities weekends days | 21 |
| 18 | Ending activities | 22 |
| 19 | Trace length of working days and weekends | 23 |
| 20 | Activities presence in working days | 23 |
| 21 | Activities presence on weekends | 24 |
| 22 | The dotted chart of hhwork for activities linked to eating | 25 |
| 23 | The dotted chart of hhweekends for activities linked to eating | 26 |
| 24 | Dotted chart sleeping habit of working days | 27 |
| 25 | Dotted chart sleeping habit of weekends | 28 |
| 26 | Throughput_time for hhwork and hhweekends | 28 |
| 27 | Processing time for hhwork | 29 |
| 28 | Processing time for hhweekends | 29 |
| 29 | Idle time for hhwork and hhweekends | 30 |
| 30 | Example of fitness analysis in ProM of an individual with respect to a reference model: places with yellow background (X) represent situations where the individual deviates from the process model. Transitions without a label denote silent events not appearing in the event log[] | 31 |

1 Introduction

Physical inactivity is a major risk factor for certain types of diseases. Indeed, physical activity does not only prevent or relieve diseases but also improves public health and well being [1]. In this context, personalized health solutions and lifestyle monitoring can help to ensure that individuals are doing the right activity at the right time. However, the regular use of such methods is critical to achieve the desired result. Barriers for the adoption must be low, and using both software and devices should be as comfortable as possible.

The goal of our work is the development of an approach that monitors and analyzes the personal lifestyle of users and the provision of insightful visualizations. In this paper, we focus on deriving and analyzing personal process models, also providing a structured approach to do this for SAYNOVA Company, that can help them to analyze their log files to draw interesting conclusions and insights.

SAYNOVA (who proposed this project) a startup studio partner of companies, conceives and develops digital platforms to take advantage of the revolution of the digital uses, it creates a platform that helps people make the most of their work and life, based in 75 ALLÉE DES PARFUMEURS-92000 NANTERRE FRANCE.

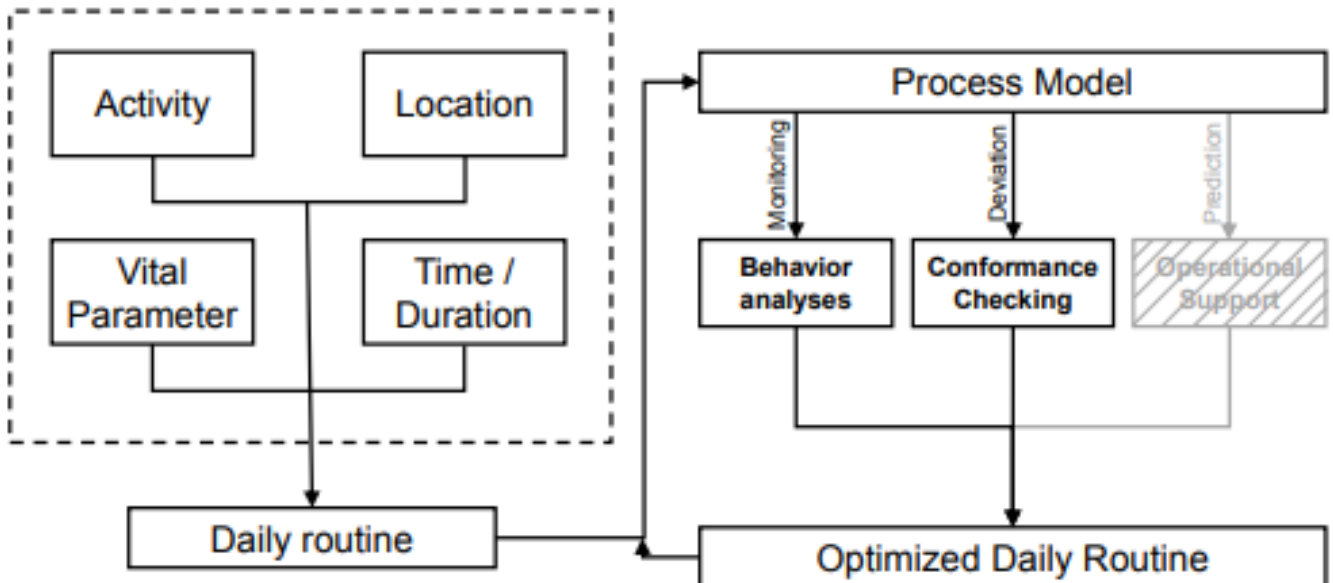


Figure 1: Optimization of the daily routine to achieve a healthier life.

The framework illustrates the interaction between the individual and considered components of the activity recognition (left) and the process mining domain (right). As our project is about the process

models so we do not focus on daily activities recognition. process mining can help us to elicit and analyze these processes. It allows discovering a process model from an event log focused on personal activity, and combined with, e.g., conformance checking, to explore deviations with respect to reference models. The results could be useful in the context of monitoring to provide meaningful feedback but also to create recommendations. As a result, the user learns to optimize the daily routine in respect of a healthy lifestyle. can then be used for the operational support based on the individual's process models, enabling predictions and recommendations in order to accomplish certain goals. We do not deal with activity recognition.

Commonly, machine learning techniques are used for activity recognition [3]. Therefore, the data set can be used to build or evaluate activity recognition systems, but in the following, we want to use the result of such a system in combination with process mining to create personal processes by using the manually created activity labels. The resulting personal process models should allow benefiting the user's health by making visualizations, recommendations, and predictions.

The paper is structured as follows: In Section 2 we present the data set we acquired. In Section 3, the related work concerning health care, activity recognition, and process mining is summarized. Section 4 introduces background knowledge regarding process mining that is considered in the following sections. Section 5 describes the possibilities of discovering personal processes and extracting meaningful patterns and rules. Based on this, Section 6 outlines how to analyze and compare these processes to detect deviations and optimize the behavior of the related person. Finally, Section 7 covers the future work of this paper and a conclusion. We can not put all the results in this paper so that it wouldn't be so long, but we will highlight the relevant result and the way to achieve it.

2 Data set

The importance of data in Data science forces us to dedicate a whole section to talk about it, our data set is composed of six tables created by the 4TU Centre For Research Data , and they did this by building a framework to record this data manually using sensors , acquiring these data automatically can be done using Machine Learning recognition algorithms, but it's hard to implement and requires having an enormous amount of data, training such algorithms use features like the geographical position for cooking, sleeping, working ...



Figure 2: Collector and labeling framework: Wear App (smart-watch, 1) and Hand App (smart-phone, 2). The positions of the devices may vary

To keep it simple the data sets contains generic activities like Eating for (Breakfast, Lunch and Dinner) ... with no sub-field.

There are three individuals : each one has two tables that consists of working days and weekends, these log files have a total of 61 cases and 8830 events.

Dealing with non-clean data is a part of Data scientist job, we developed a method for log files cleaning, it structures the data to respond to XES norm and then use the dplyr package to correct the value of activity instance id, bind tables if necessary and remove attributes that don't contribute to the analysis .

3 Related Work

Daily activity analyses has always interested scientists in different fields from psychology to Medicine even Computer science. Therefore, detecting wrong behavior or anormal activities may help to prevent undesirable consequences [13]. Accurate information on people's behavior and their daily routine allows to support them [4] or provide feedback . The event log which describes the daily routine, can be transformed to a Probabilistic approaches such as Markov Logic Networks or Hidden Markov Models are used to determine the performed activity or to predict an unobserved state, e.g., the next activity [5].

An enormous work in this field is done by Kathrin Kirchner ,Nico Herzberg and Andreas Rogge-Solti M [6] explaining how to apply P.M. techniques like conformance checking to computes fitness of models,

Also Timo Sztyler and Josep Carmona on the development of self tracking framework [7].

4 Preliminaries: Process Mining Techniques

In this section we provide the necessary background to understand the techniques which we consider in the following sections. We will focus on two main process mining disciplines: process discovery and conformance checking, which represent the core of process mining [8].

Process discovery is challenging because the derived model has to be fitting, precise, general, and simple. Conformance checking techniques are meant to verify these criteria to assess the quality of a model in representing the information contained in a log.

Generally Data Science techniques try to respond to these questions :

WHY Did It Happened? : can be respond with the analysis of bottlenecks.

WHAT Happened ? : What has happened in the process (Steps of the process) can be respond using Process Discovery.

WHAT WILL HAPPENED ? : predict the future and make recommendations using operational support .

The process discovery consist of three steps :

- 1) **Extraction**: transform raw data into event data
- 2) **Preprocessing**: enrich and filter event data

- **Aggregation**: remove redundant details

- **Filtering**: focusing analysis

- **Enrichment**: add useful data attributes

- 3) **Analysis**: gain useful insights in the process

- **Organizational**: Focusing on the actors of a process and how do they work together (this one does not concern our case).

- **Control-flow or Structuredness**: Focuses on the flow and the structuredness of the process e.g (The journey of a patient threw the emergency Room). Control flow refers to the different successions of activities , each case can be expressed as a sequence of activities , each unique sequence is called a trace or process variant .

- **Performance**: Focusses on time and efficiency e.g (How long does it take before a patient can leave the emergency department ? Or in which area or at what time of day are trains most delayed ?)

4.1 Structuredness

The structuredness correspond to two phases

1) **Metrics** : looking to a one specific aspects

- **Entry and exit points** : starts and ends activity
- **Length of cases** : the distribution of the case length
- **Presence of activities** : which activities are always present in a case and which are exceptional
- **Rework** : for each of these topics , a metric similar to that resource_frequency can be calculated

2) **Visuals** :

- **Process map**
- **Trace explorer**
- **Precedence matrix**

5 Process Discovery

Due to the variability in personal activity data, there is not a simple process model that represents all possible paths for an individual, even for the reduced number of individuals monitored in this paper. In this section, we focus on the most frequent paths taken by each individual.

To illustrate the potential of a personal process model with respect to analyzing tons of raw data, we focus on two simple aspects: the difference in activity between work and weekend days on the one hand, and the differences across the individuals on the other hand.

During the week vs. weekend. Shows the main activity models during the working week and the weekend, respectively. The process models depicted in the figures have a very different structure. This clearly denotes a variation in the personal activity during the week and weekend, when considering the main activity by individuals. For instance, while in the week days the main behavior is centered towards "Work" which is also the most frequent activity, for further more see section comparison between working days and weekends.

Personal activity across users. Shows each main activity models. This aim to see the difference between these individuals and the common behavior between them to generalize a process model that fit the most.

5.1 Comparisons between individuals

5.1.1 Entry and Exit points

1) ENTRY POINTS :

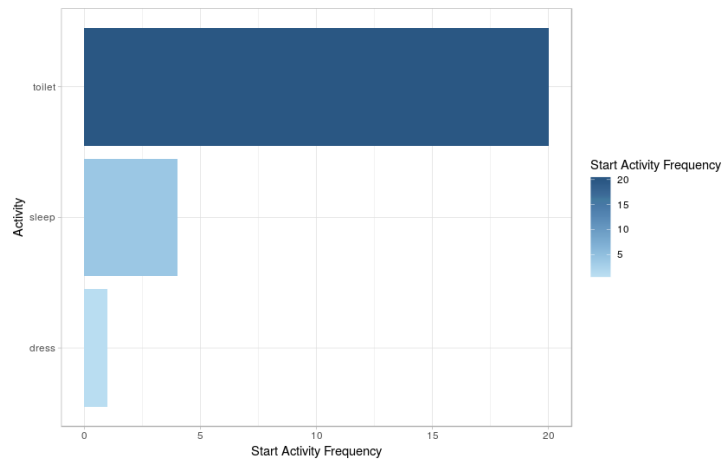


Figure 3: Start point for hh102

This figure shows that almost all people start their journey with going to the toilet and then sleep except for hh104 he may begin his journey with only going to bed directly.

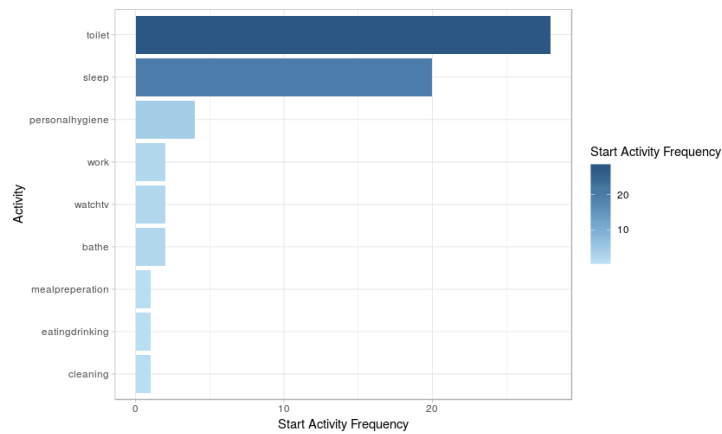


Figure 4: Start point for hh104

2) EXITS POINTS :

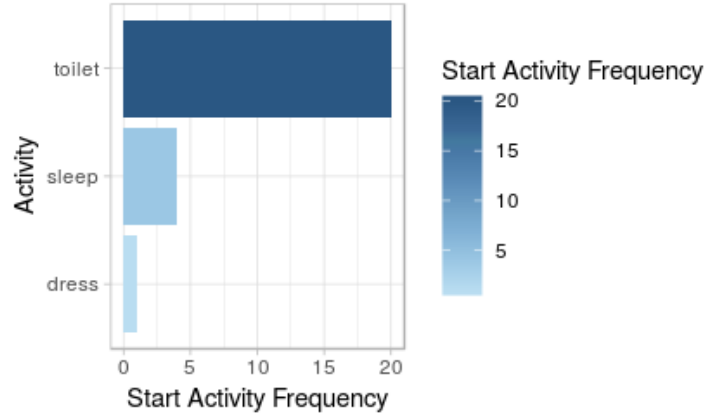


Figure 5: Exits points for hh102

Something very evident is that they all finish their journey with sleeping, or they go to the toilet just before sleeping but sometimes, like for hh104, he finish his day with activities such as works or watchtv. This behavior may not be healthy but we will talk about its impact later on.

5.1.2 Precedence Matrix

This matrix shows the flows between activities in a more structured way compares to process maps, it does show which activities preceded each activity, which one is at the start or the end and which are the most important flows in the process, this matrix shows for (see Annex):

-hh102: Eatingdrinking is commonly followed by cleaning or relaxing, it might be evident that Eatingdrinking also comes after meal preparation 75% of the time, but sometimes after meal-preparation he takes a bath or relax. Medication is followed by sleeping, outdoors is followed by relaxing and personal hygiene is sometimes followed by outdoors or relaxing. . .

-hh110 : this guy don't eat snacks (we will see why later) but he always eat just after mealpreparation and taking medication represent an important activity in his activities-flow.

5.1.3 Process Variants

- **Trace Explorer** : show the most frequent traces . A common trace begin with sleeping or toilet and finish by sleeping ,dressing and relaxing ... something annoying is that two of these individuals alternate between sleeping and toilet more than 5 times a night and they also takes medicament. It may be a sign of illness like diabetes or that the medicament causes Insomnia (cf hh110 traces extract in Annexe).

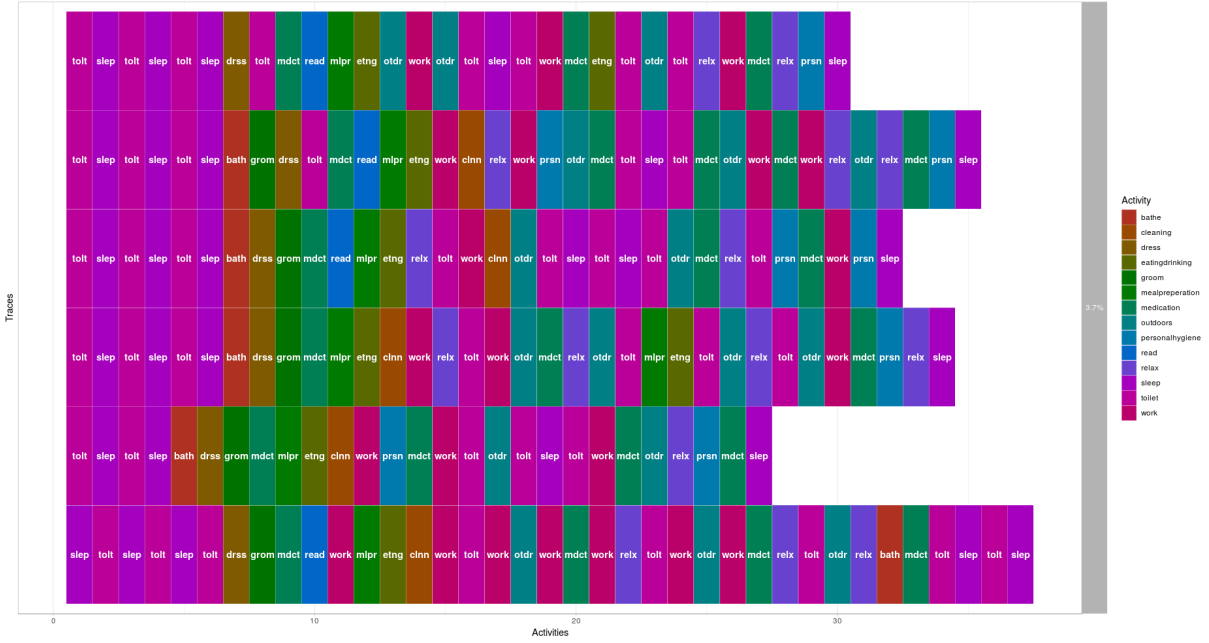


Figure 6: Trace explorer hh110

- **Trace_length** : Find the average number of activities per day. The average number of activities per day is 35 activities. At minimum 8 activities and at maximum 65 activities are done. This figure illustrate the 5 summary number.

```

> trace_length(hh102)
      min      q1    median      mean      q3      max    st_dev      iqr
21.000000 25.000000 30.000000 29.440000 34.000000 39.000000  5.362524  9.000000
> trace_length(hh104)
      min      q1    median      mean      q3      max    st_dev      iqr
26.000000 43.000000 46.000000 46.590164 52.000000 65.000000  7.655885  9.000000
> trace_length(hh110)
      min      q1    median      mean      q3      max    st_dev      iqr
 8.000000 29.500000 32.000000 31.000000 35.000000 43.000000  7.184385  5.500000

```

Figure 7: Trace length of hh102 , hh104 and hh110

- **Activity Presence** : check the presence of activities per day , we can see that activities like sleeping ,toilet , cleaning , relax , outdoors ...are always present (Very normal) , but taking medicament is 100% present in hh104 and hh110 journeys but not frequent for hh102. Also, hh110 read more often than hh102 but hh104 don't read.

5.1.4 Performance Analysis

- **Performance process map** : special type of process that does not show frequencies on the arcs and nodes but duration of activities and the times between activities .

process maps can be a good method to see different way of performance measure , all of our process maps figures are spaghetti like (technical term that refers to be a little hard to read) so we dedicate a place to them at the end .

For processing time : we can see the real time investment at activities nodes of the graph and how much it takes to go from each node to the connecting nodes (activities) see Fig.hh102 process map of processing time (Annex),process map's can be simplified using filters (see filter section).

- **Dotted Chart** : while the performance map focuses on the duration of activities , the dotted chart shows the distribution of activities over time , x-axes : time and y-axes : cases , lines can show that the same activity always occur at time and also can be seen that some patterns fade away.
 - **EATING HABITS**: We extract certain pattern to examine the eating habits for individuals. The dotted chart shows us that hh102 tend to replace lunch with snacks and to eat dinner earlier , we can say that if he eat breakfast earlier in the day he often eat the lunch between 11:00 and 1PM but if he eat breakfast late in the day he tend to skip lunch and eat snacks .

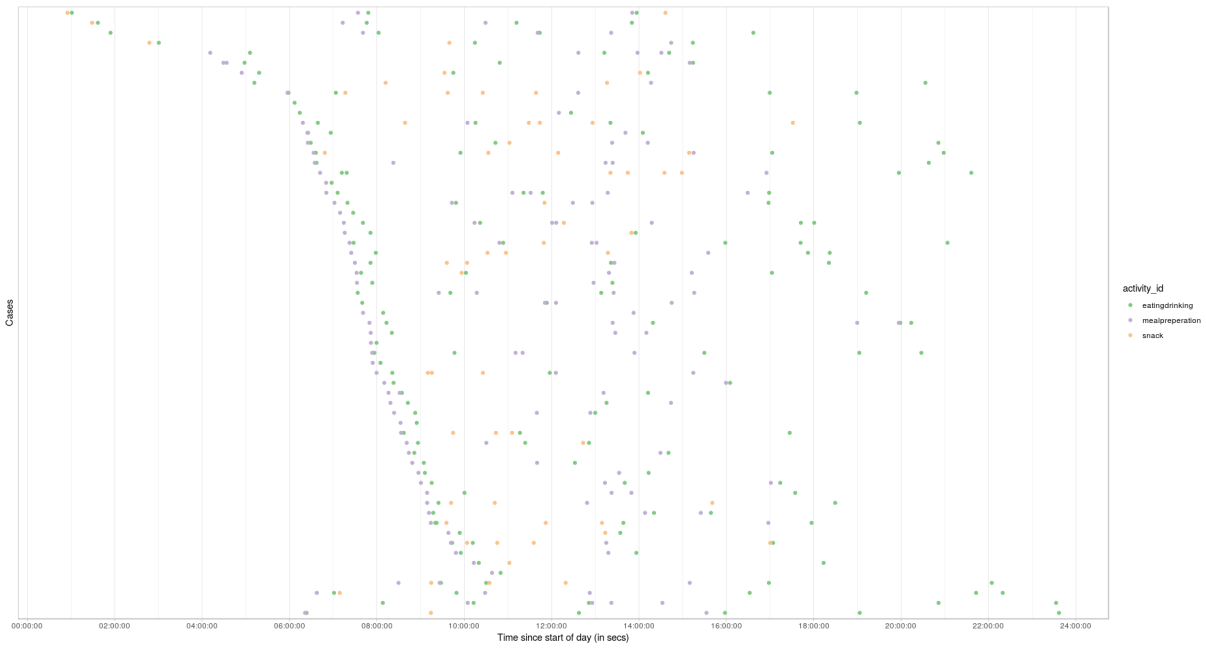


Figure 8: The dotted chart of hh104 for activities linked to eating

For hh104 things seems to be normal. Sometimes, he eats lunch multiple times and skips the dinner, an obvious pattern showed in the graph is that Breakfast is often taken between 6 AM and 10 AM.

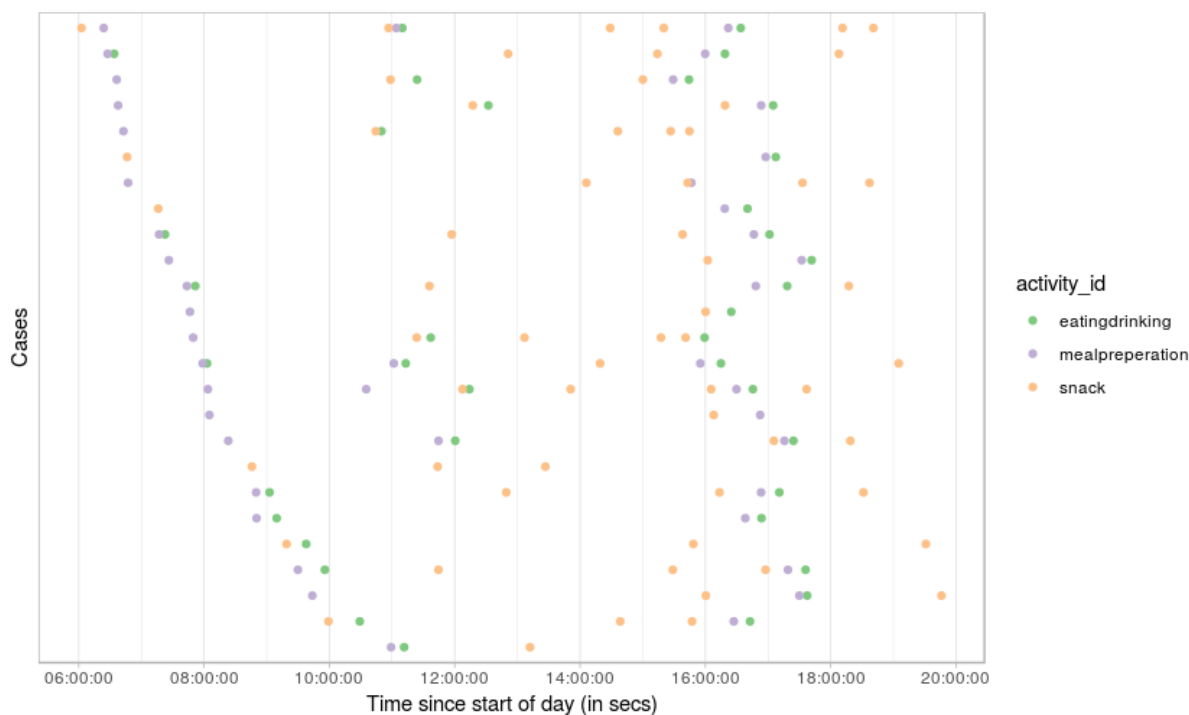


Figure 9: The dotted chart of hh102 for activities linked to eating

- **SLEEPING HABITS:** Both hh102 and hh110 are closer to have an organized sleeping habit than hh104, they tend to sleep around 10 PM or 11 PM and don't sleep multiple times during the day but hh104 don't have an organize sleeping habit, his dots are everywhere and it's difficult to see a common pattern except for going to bed around 10 PM or 11 PM like the others. He sleeps multiple times during the day (Multiple Naps) see next figures .

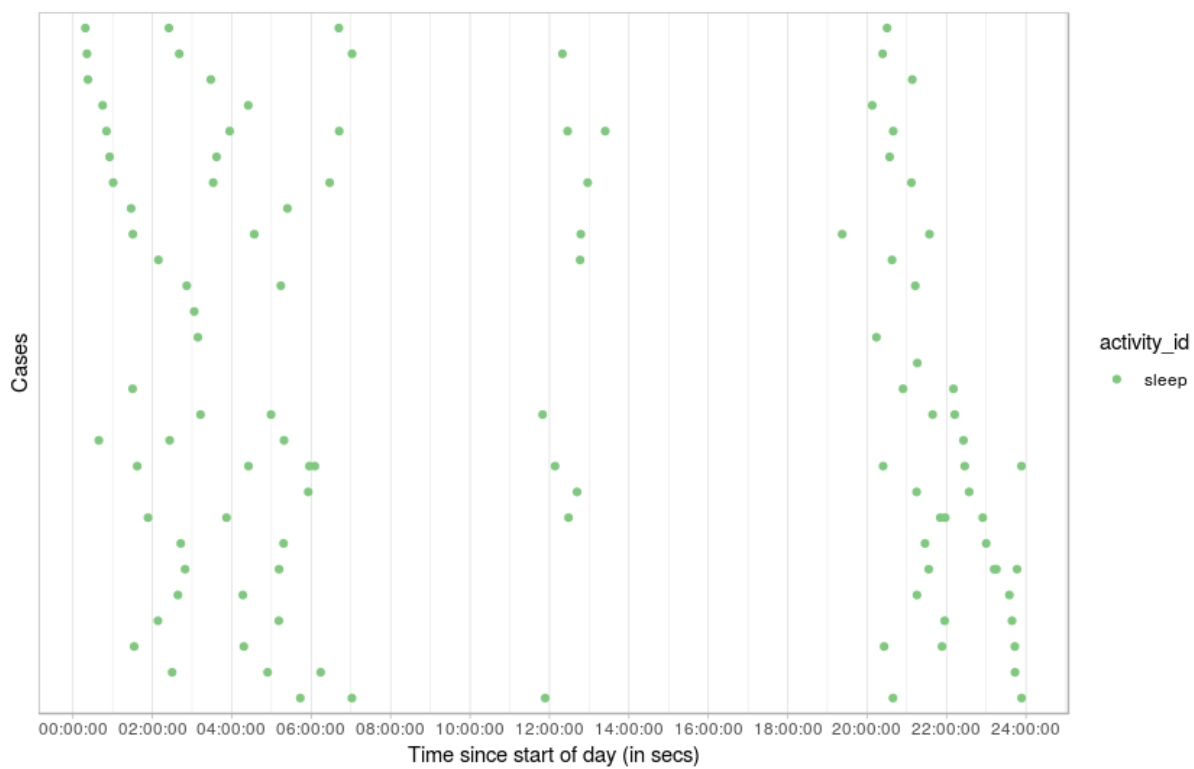


Figure 10: Dotted chart sleeping habit of hh110

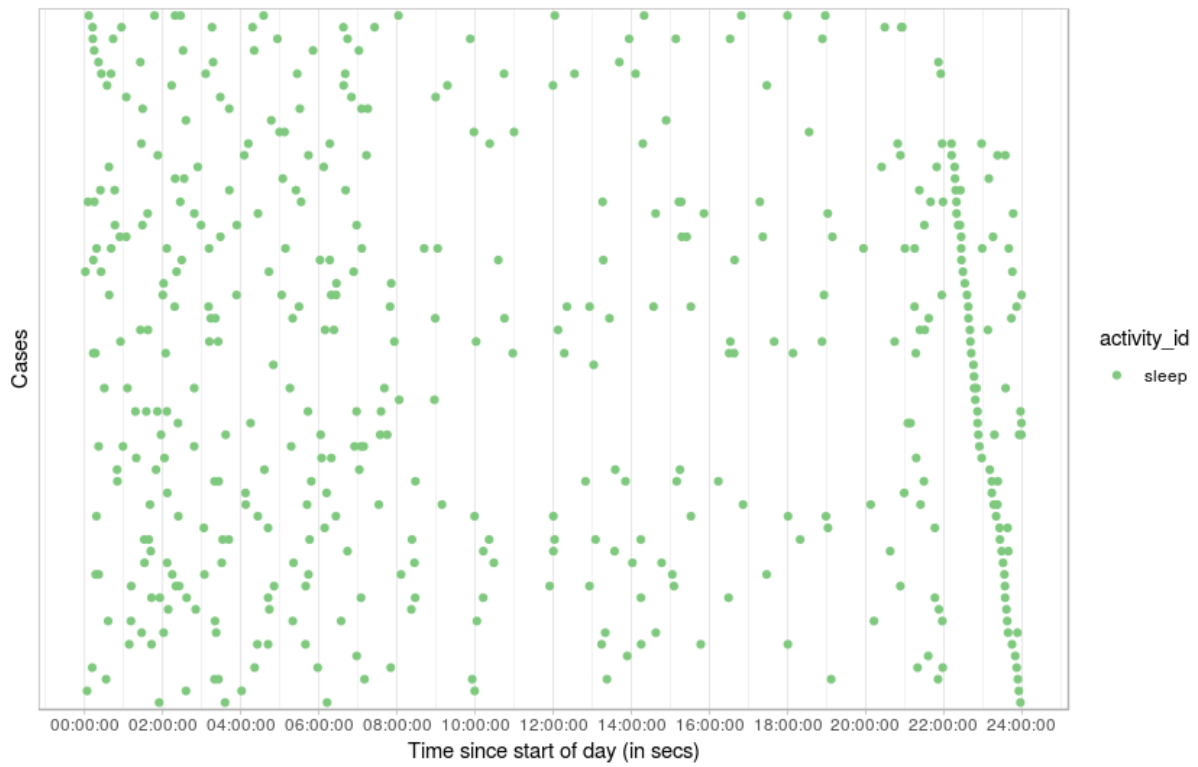


Figure 11: Dotted chart sleeping habit of hh104

5.1.5 Time Analysis

- **Troughput time** : it's the time since the start of a case until the end of the case, which includes both active time and idle time (How it takes to get something done ?)

```

> hh102 %>%
+   throughput_time(level="log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
13.145278 20.759444 23.626111 23.111044 26.961389 30.902222  4.923953  6.201944
attr(,"units")
[1] "hours"
> hh104 %>%
+   throughput_time(level="log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
14.775000 23.001111 23.665000 23.630378 24.257500 31.278889  1.997486  1.256389
attr(,"units")
[1] "hours"
> hh110 %>%
+   throughput_time(level="log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
 4.379167 22.377222 23.508056 26.784691 25.093750 134.238611 21.964821  2.716528
attr(,"units")
[1] "hours"

```

Figure 12: Throughput_time for hh102 , hh110 and hh104

This figure shows us that the median is around 23 hours , from the start of the day till it's end

- **Processing time** : it's the sum of the activity durations, which means it does not include the time in between different activities (it tells us about the real times investment) : these figures shows the processing time for these individuals :

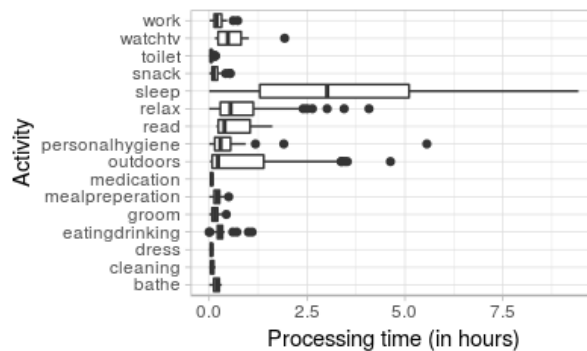


Figure 13: Processing time for hh102



Figure 14: Processing time for hh104

These figures shows us that hh102 is the closest to be normal for sleeping hours. Overall, it seems that they do not work (the real difference between working hours can be seen in the comparison of working days and weekends)

- **Idle Time:** it's the sum of the durations between the activities, in which no processing of the case takes place :

```
> hh102 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median     mean      q3     max    st_dev     iqr
0.6622222 1.2330556 1.4316667 2.0582889 1.7583333 12.4355556 2.3137264 0.5252778
attr(,"units")
[1] "hours"
> hh110 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median     mean      q3     max    st_dev     iqr
0.2463889 0.8663889 2.7652778 3.2912757 4.6661111 11.3080556 3.0647722 3.7997222
attr(,"units")
[1] "hours"
> hh104 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median     mean      q3     max    st_dev     iqr
0.4986111 5.3944444 6.7663889 7.5510929 9.7938889 16.8166667 3.4975107 4.3994444
attr(,"units")
[1] "hours"
```

Figure 15: Idl time for hh102 , hh104 and hh110

These results have an essential meaning about the quality of our data and how are these individuals ready to record the data, in another term it may also concern the productivity of these individuals because it tells us the duration that no one was doing nothing .

5.1.6 Filtering

Select processes based on some attributes or based on a process characteristic , Filter according to activity type or time period.

We filter by the 60% most common activities and it did solve the spaghetti like graphs : see fig .hh102 process map of processing time (annexe) .

An easy way to Visualize the happy path using a process map by setting frequency() value to "absolute_case" , see Fig. hh102 happy path process model (Annexe) .

Cases can be selected based on presence or absence of certain activities using the filter_activity_presence() function. In case of multiple activity labels, this shows us that filtering hh104 with medication presence, the amount of sleeping hours increases also for activities such as outdoors , watching TV and working .See Fig. hh104 process model without medication (annex), A big question is what is that medicine the it's more important than everyday life !

With these graphs we can see the impact of each activity on the everyday life of these individuals.

Trim to time period Trimming can be done either by defining a time period, or by defining the desired endpoints of cases and can show us e.g if there is a specific behavior during the day or during the night.

5.2 Comparisons between Working days and weekends

this section consists of repeating the same approach for working days and weekends to see if there is a common pattern or a different pattern .

5.2.1 Entry and Exit points

1) ENTRY POINTS : The figures shows us that the starting activities are little bit different , sleeping and toilet always at the beginning , but working and medication are specific to the working days .

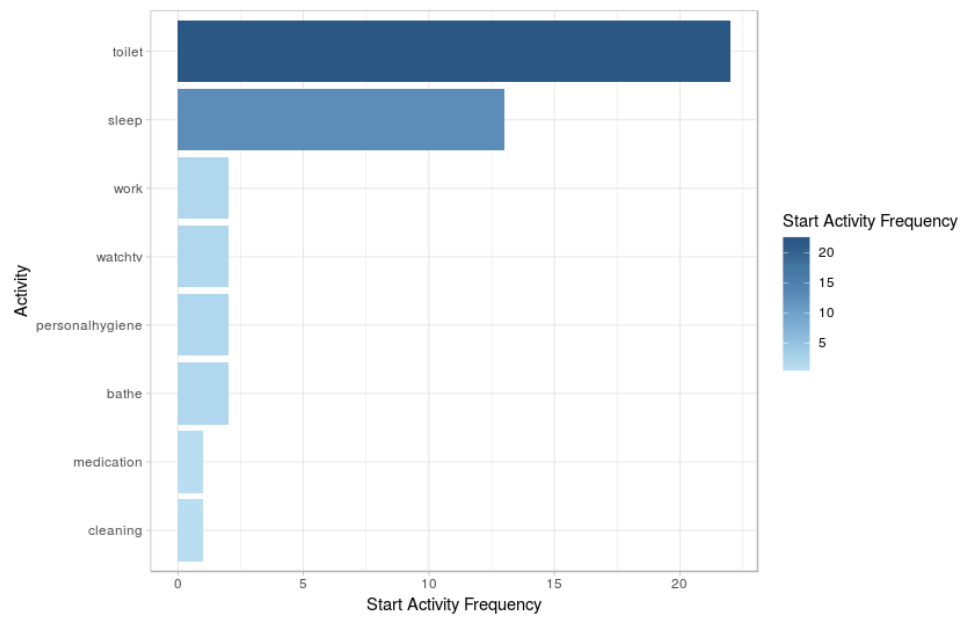


Figure 16: Starting activities working days

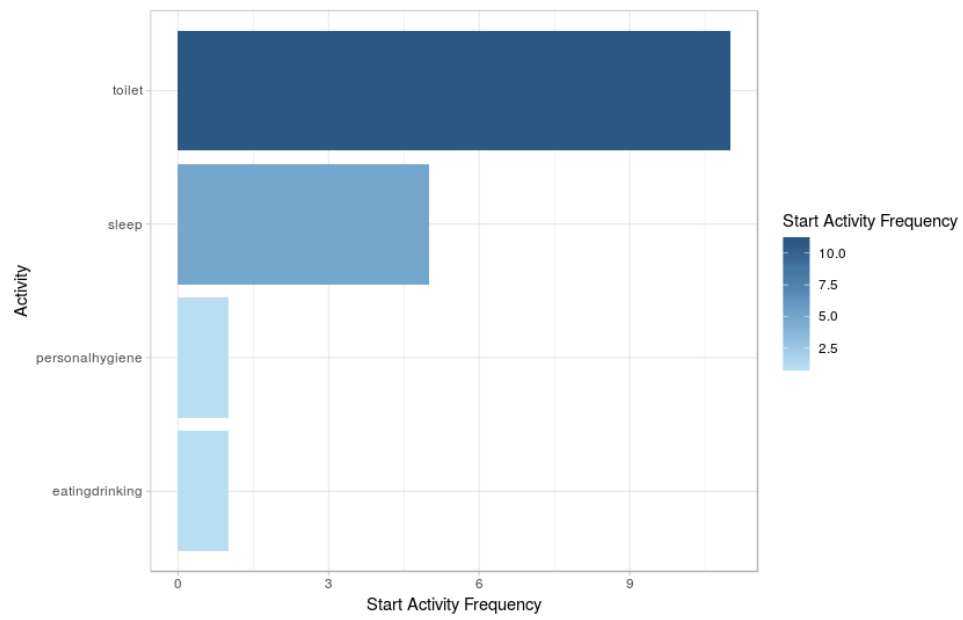


Figure 17: Starting activities weekends days

. 2) EXITS POINTS : Both working days and weekends have the same ends activities so individuals finishes their day often with sleeping or sometimes working on both working days and weekends (A Common pattern) , see figure :

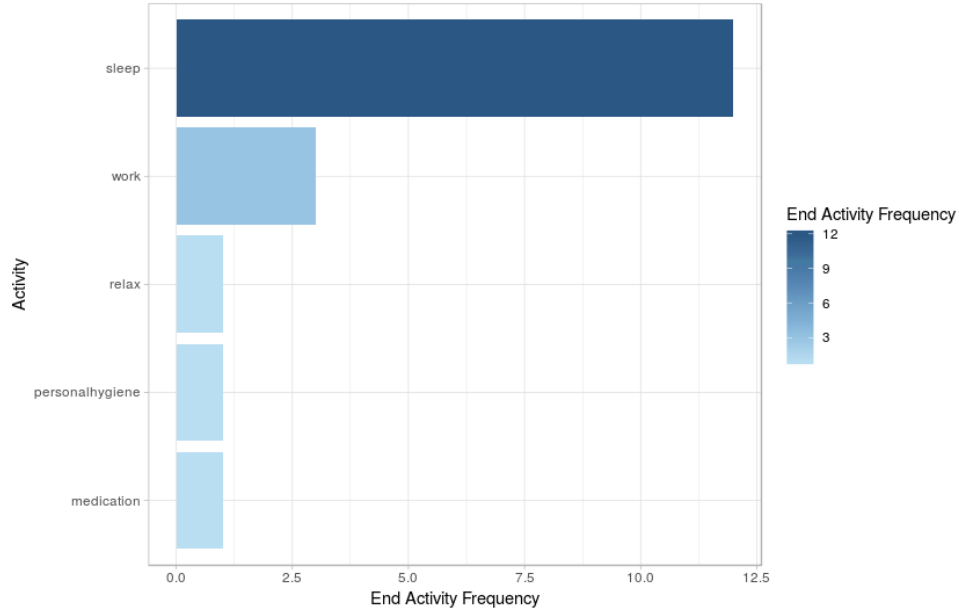


Figure 18: Ending activities

5.2.2 Process Variants

- **Trace_length** : In working days the average number of activities per day is 72 activity a minimum of 37 activities and a maximum of 127 activities , While in weekends the minimum is 31 activities the average is 67 and the maximum is 130 (normal because we can do a lot of things during the weekend because we do not have to go to work) , but both have more or less the same median , see figure :

```

> trace_length(hhweekends)
      min      q1    median      mean      q3     max    st_dev     iqr
31.00000 44.50000 52.00000 67.44444 94.50000 130.00000 31.07528 50.00000
> trace_length(hhwork)
      min      q1    median      mean      q3     max    st_dev     iqr
37.00000 45.00000 55.00000 72.88889 108.00000 127.00000 31.00529 63.00000

```

Figure 19: Trace length of working days and weekends

- **Activity Presence** : the amount of activities present in the day differs from working days to weekends , general activities will always be present but even working activity is 100% present during weekends but not with the same working duration also the amount of activities present in weekends is higher than working days ,see figures :

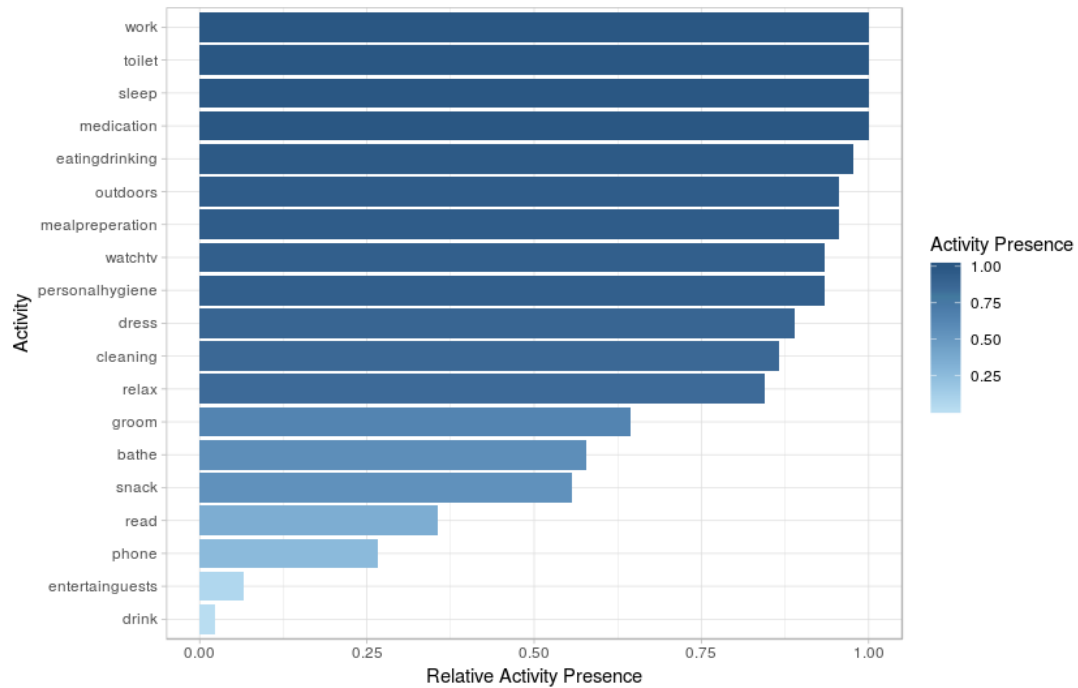


Figure 20: Activities presence in working days

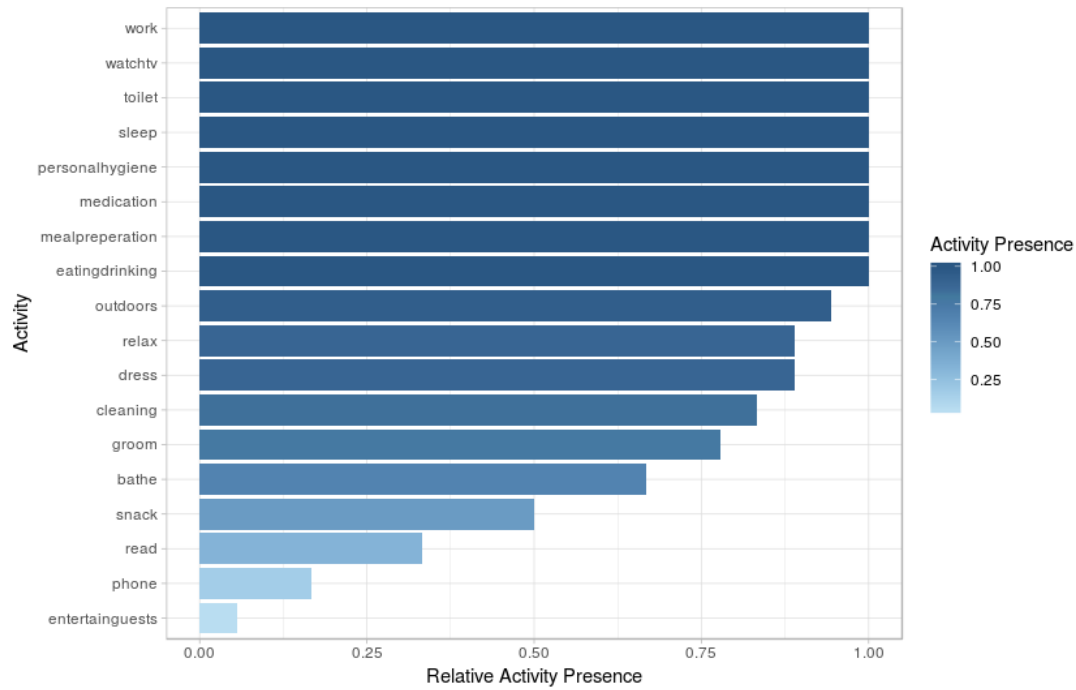


Figure 21: Activities presence on weekends

5.2.3 Performance Analysis

- **Performance process map** : the graphs shows the difference between week and weekends flow see Fig.work days and weekends process map of processing time (Annex), these working flows shows that the amount of time spent per activities like sleeping , outdoors and watchTV increases in weekends over working days .
- **Dotted Chart** :
 - **EATING HABITS**: Comparing eating habits on weekends and working days shows that eating habits seem to be more organized in working days with a strong pattern between 6-10 indicating the breakfast time, and that eating time often happens between 06:00 AM and 08:00 PM which is healthy, see next figures :



Figure 22: The dotted chart of hhwork for activities linked to eating

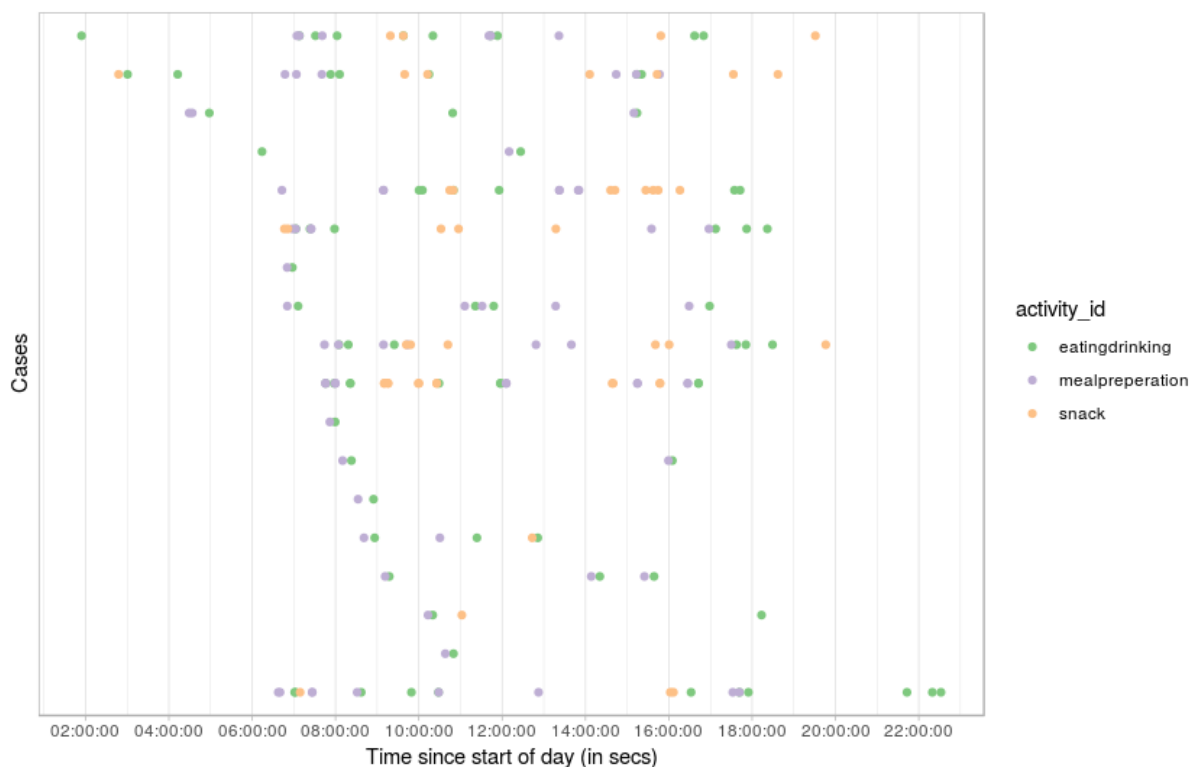


Figure 23: The dotted chart of hhweekends for activities linked to eating

- **SLEEPING HABITS:** these comparisons shows us that there is no significant difference between weekends and working days about sleeping habits for these individuals ,but they are more likely to sleep during the day in weekends than in working days, waking up multiple times during the night is common between these days, but on the working days they tend to go to bed between 09:00 PM and 00:00 AM ,but on the weekends it's a bit mess .

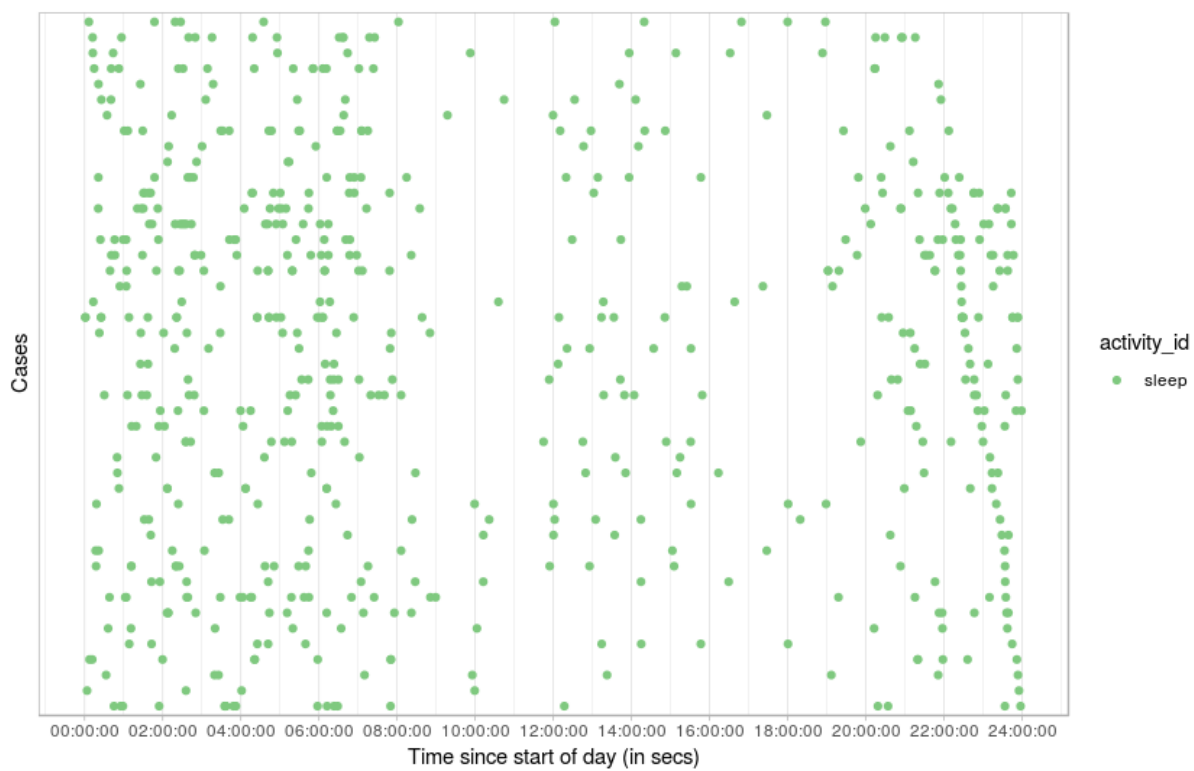


Figure 24: Dotted chart sleeping habit of working days

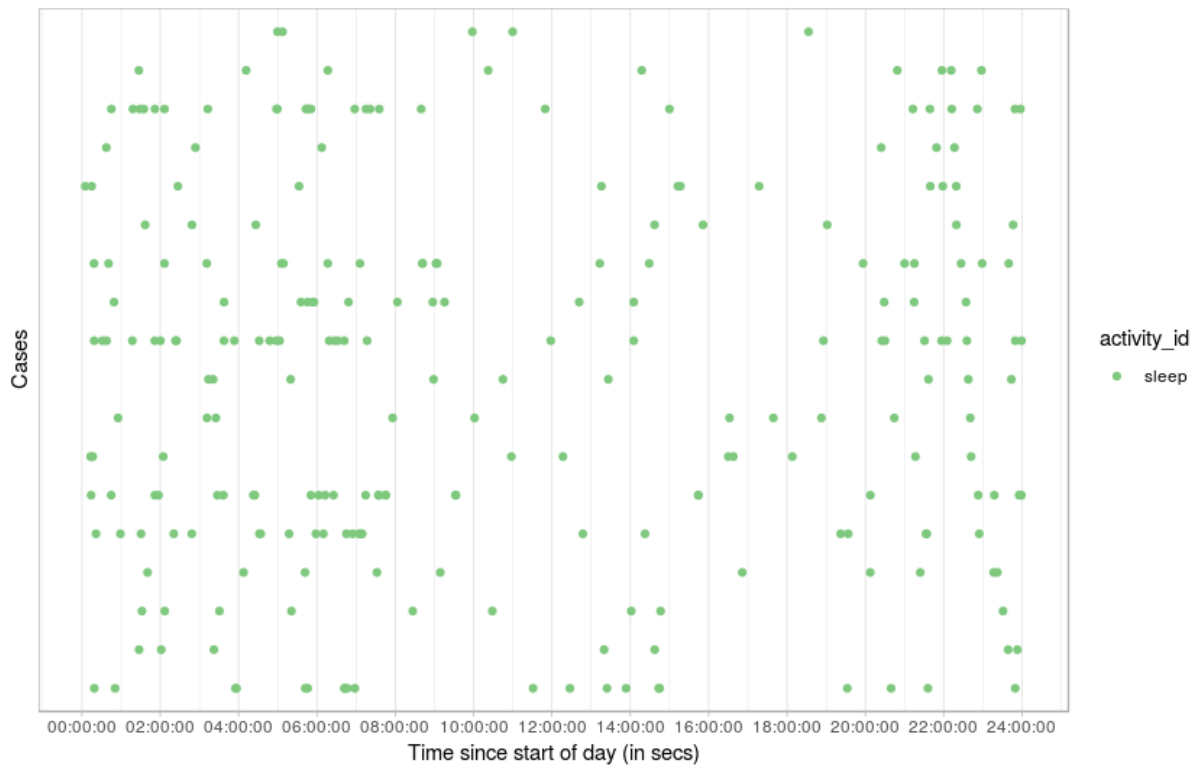


Figure 25: Dotted chart sleeping habit of weekends

5.2.4 Time Analysis

- Throughput time :

```
> hhwork %>%
+   throughput_time(level="log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
21.920278 23.307222 24.882222 31.811049 27.775000 135.678056 24.640848  4.467778
attr(,"units")
[1] "hours"
> hhweekends %>%
+   throughput_time(level="log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
17.287778 23.903681 24.589583 25.412299 26.584792 31.733333  3.471440  2.681111
attr(,"units")
[1] "hours"
```

Figure 26: Throughput_time for hhwork and hhweekends

This figure shows us that the median is around 24 hours , it's mean how much times it takes from the start of the day till it's end .

- **Processing time :**

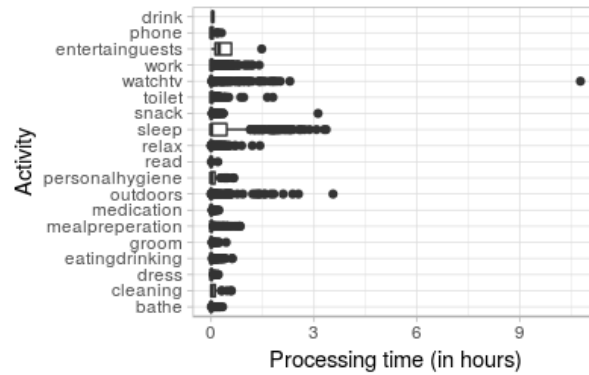


Figure 27: Processing time for hhwork

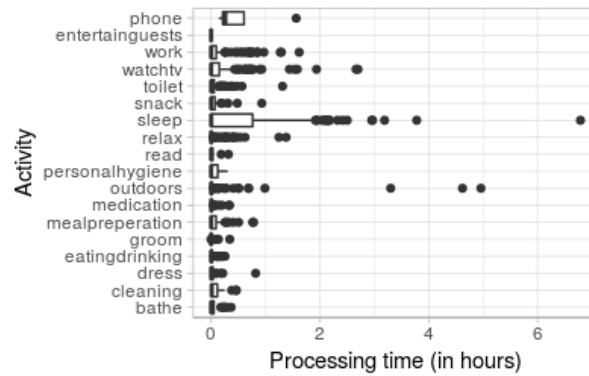


Figure 28: Processing time for hhweekends

- Idle Time:

```

> hh102 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
0.6622222 1.2330556 1.4316667 2.0582889 1.7583333 12.4355556 2.3137264 0.5252778
attr(,"units")
[1] "hours"
> hh110 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
0.2463889 0.8663889 2.7652778 3.2912757 4.6661111 11.3080556 3.0647722 3.7997222
attr(,"units")
[1] "hours"
> hh104 %>%
+   idle_time(level = "log", units = "hours")
      min      q1    median      mean      q3      max    st_dev      iqr
0.4986111 5.3944444 6.7663889 7.5510929 9.7938889 16.8166667 3.4975107 4.3994444
attr(,"units")
[1] "hours"

```

Figure 29: Idl time for hhwork and hhweekends

with this result we can see the difference on the level of productivity or the readiness of these individuals between working days and weekends , it's obvious with the median that they are more productive on working days than on weekends .

5.3 Generalization of all processes models

After seeing the two perspective weeks vs. individuals we try to generalize a process model that fit the best with previous models and filter it with the common activities to improve the readability of the graph , see Fig.General process model (Annex).

6 Conformance Checking

Conformance checking allows us to replay models on top of reality to see the gap between our models and the reality. In our case, seeing the gap between these individuals daily life habits and the reference model for a healthier life. Doing so requires having a reference model that follows the norm e.g sleeping for 8-9 hours and make physical activity that's correlated with how much calories we eat. Hence, this reference models has to describe precisely how activities should be carried out in order to satisfy a guideline. Thus, the reference model has to provide the opportunity to describe certain actions in a specific order (e.g., "Sport" should be followed by "Personal Grooming"), Reference models can be obtained in several ways.

One possibility would be to ask a domain expert to create manually the desired reference model for a given goal. A second option would be to collect event logs from successful individuals.

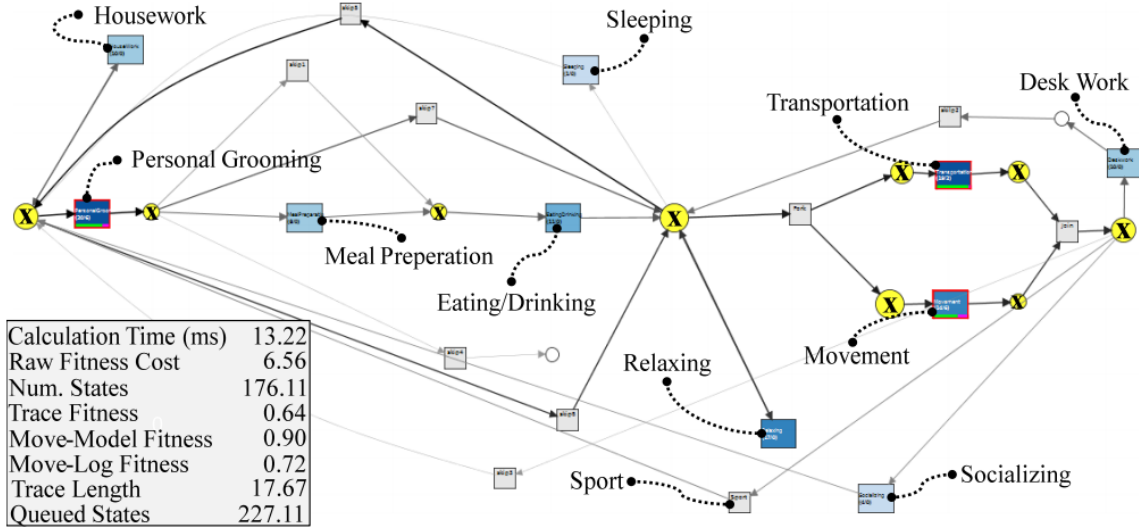


Figure 30: Example of fitness analysis in ProM of an individual with respect to a reference model: places with yellow background (X) represent situations where the individual deviates from the process model. Transitions without a label denote silent events not appearing in the event log[].

When a reference model is available, conformance checking techniques can be applied to assess the adequacy of the reference process model in representing the traces of individuals [13]. Since the reference model describes the ideal behavior, it is meaningful to focus the analysis on the fitness of the reference model with respect to the traces of individuals. A process model fits a given trace if it can reproduce it. If reference models are not available, simple rules can be used which should be satisfied by individuals on their daily routine. These rules may describe patterns that should satisfy an individual, e.g., "taking medicines" should be followed by "eating".

7 Conclusion

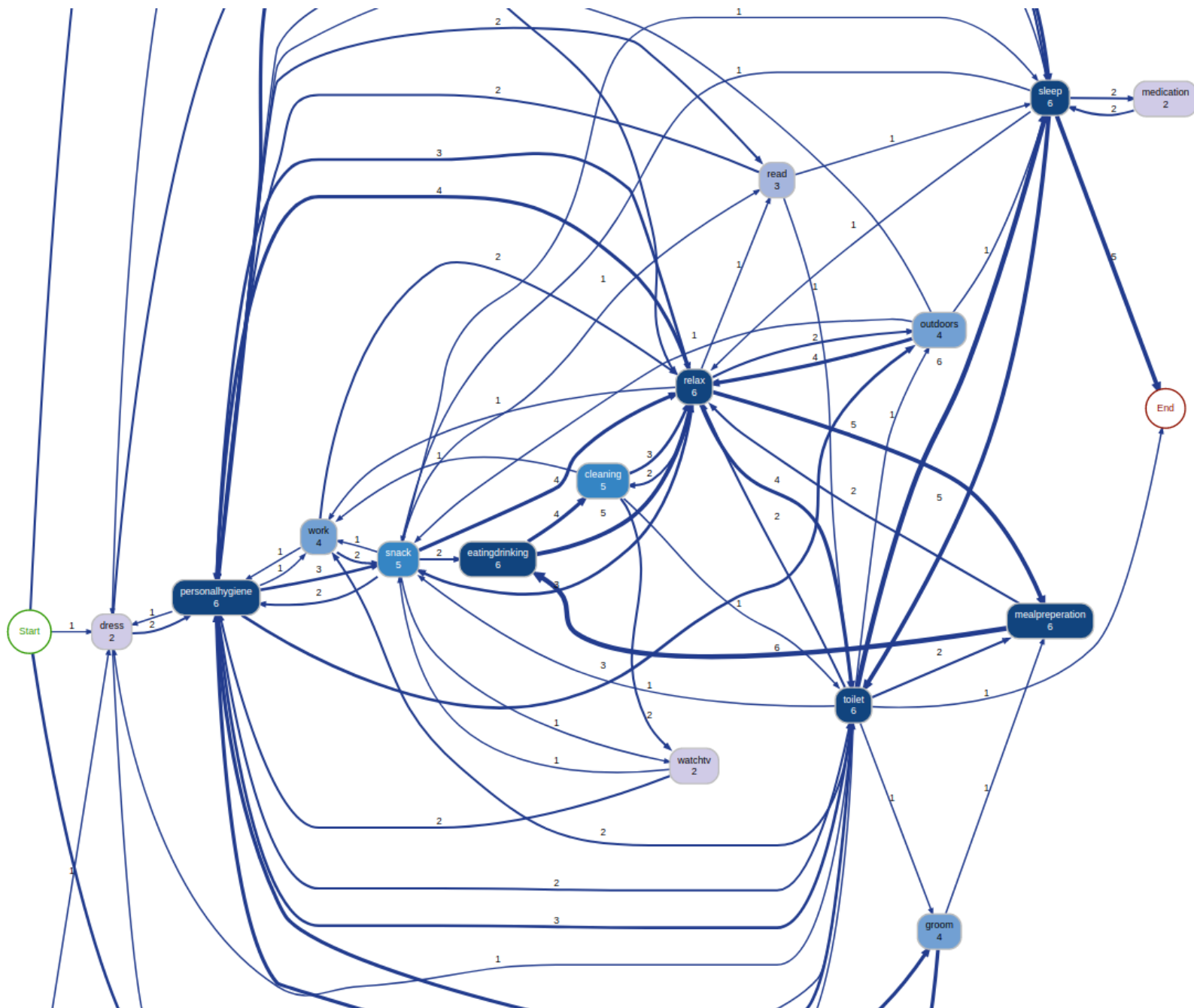
Process Mining techniques can help to analyze process models and compare them, combining these techniques with other Data Science techniques (Predicting, Model Fitness) provide a powerful tool to answer almost all process analyses questions , operational support allows us to predict next activities and make feedback about when the event will finish and what will happen.

References

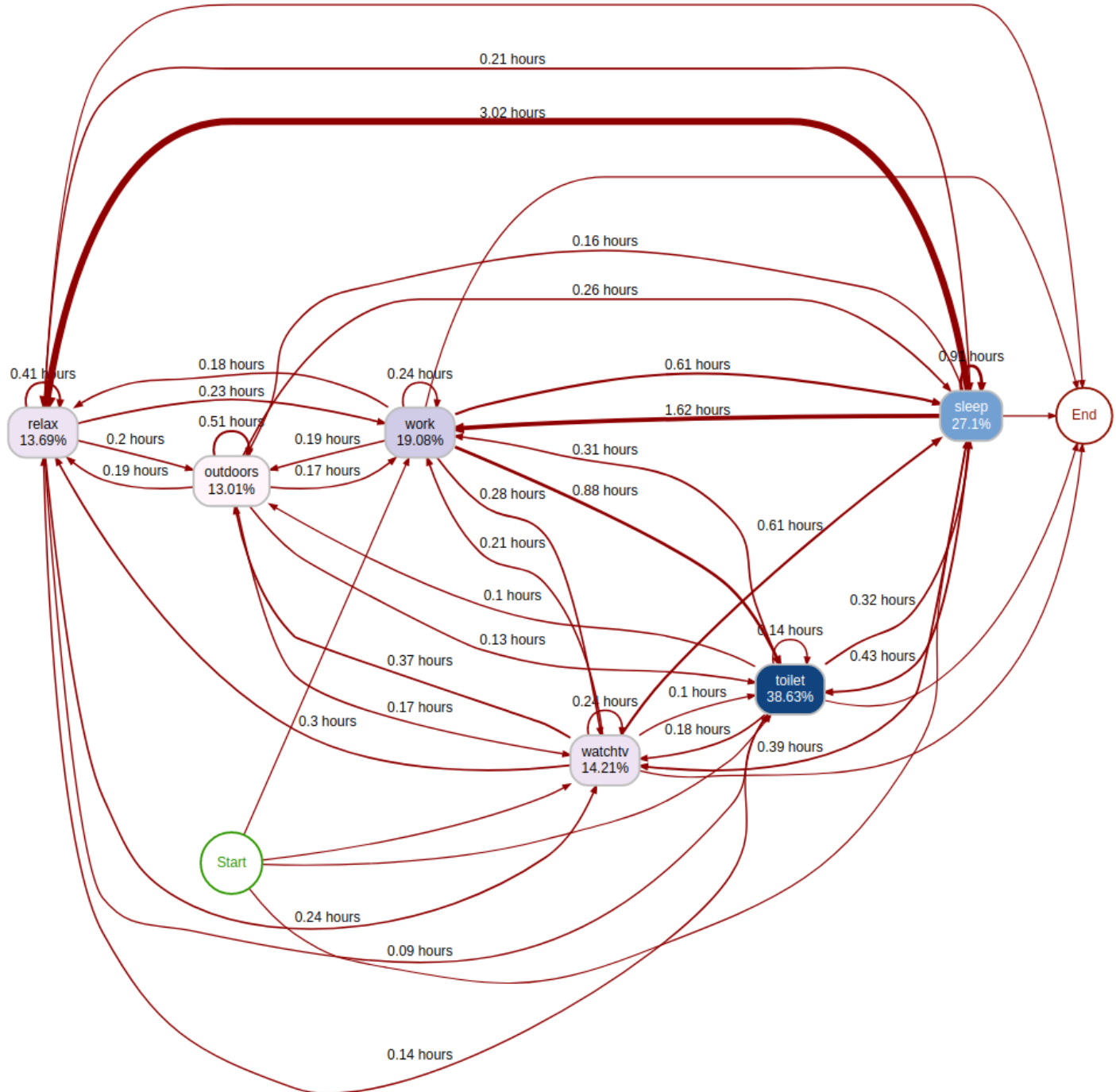
- [1] N. Sidorova A. Adriansyah and B.F. van Dongen. Cost-based fitness in conformance checking, 2011.
- [2] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data, 2004.
- [3] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data, 2004.
- [4] Steven N Blair and Tim S Church. The fitness . obesity, and health equation: is physical activity the common denominator?, 2004.
- [5] Steven N Blair and Tim S Church. The fitness . obesity, and health equation: is physical activity the common denominator?, 2004.
- [6] Nancy Brown. American heart association.
- [7] Sumi Helal Eunju Kim and Diane Cook. Human activity recognition and pattern discovery, 2010.
- [8] Boudewijn F. van Dongen Joos C. A. M. Buijs and Wil M. P. van der Aalst. Quality dimensions in process discovery: The importance of tness, precision, generalization and simplicity, 2014.
- [9] Lars Michael Kristensen Kurt Jensen and Lisa Wells. Coloured petri nets and cpn tools for modelling and validation of concurrent systems, 2014.
- [10] O.D. Lara and M.A. Labrador. A survey on human activity recognition using wearable sensors, 2013.
- [11] Miksch S. Peleg M. Reichert M. Riaño D. ten Teije Lenz, R. Process support and knowledge representation in health care, 2013.
- [12] Josep Carmona Timo Sztyler. Self-tracking reloaded: Applying process mining to personalized health care from labeled sensor data, 2015.
- [13] Wil M. P. van der Aalst. Process mining - discovery, conformance and enhancement of business processes, 2011.

A Annexe

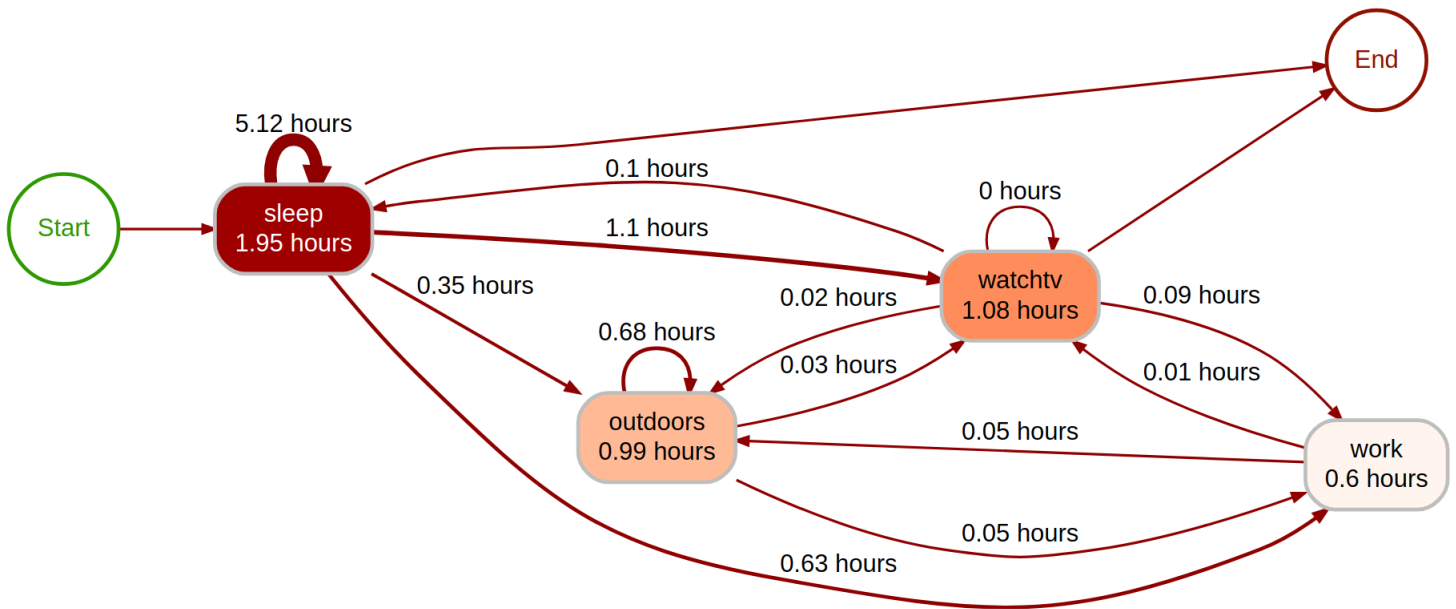
hh102 happy path process model



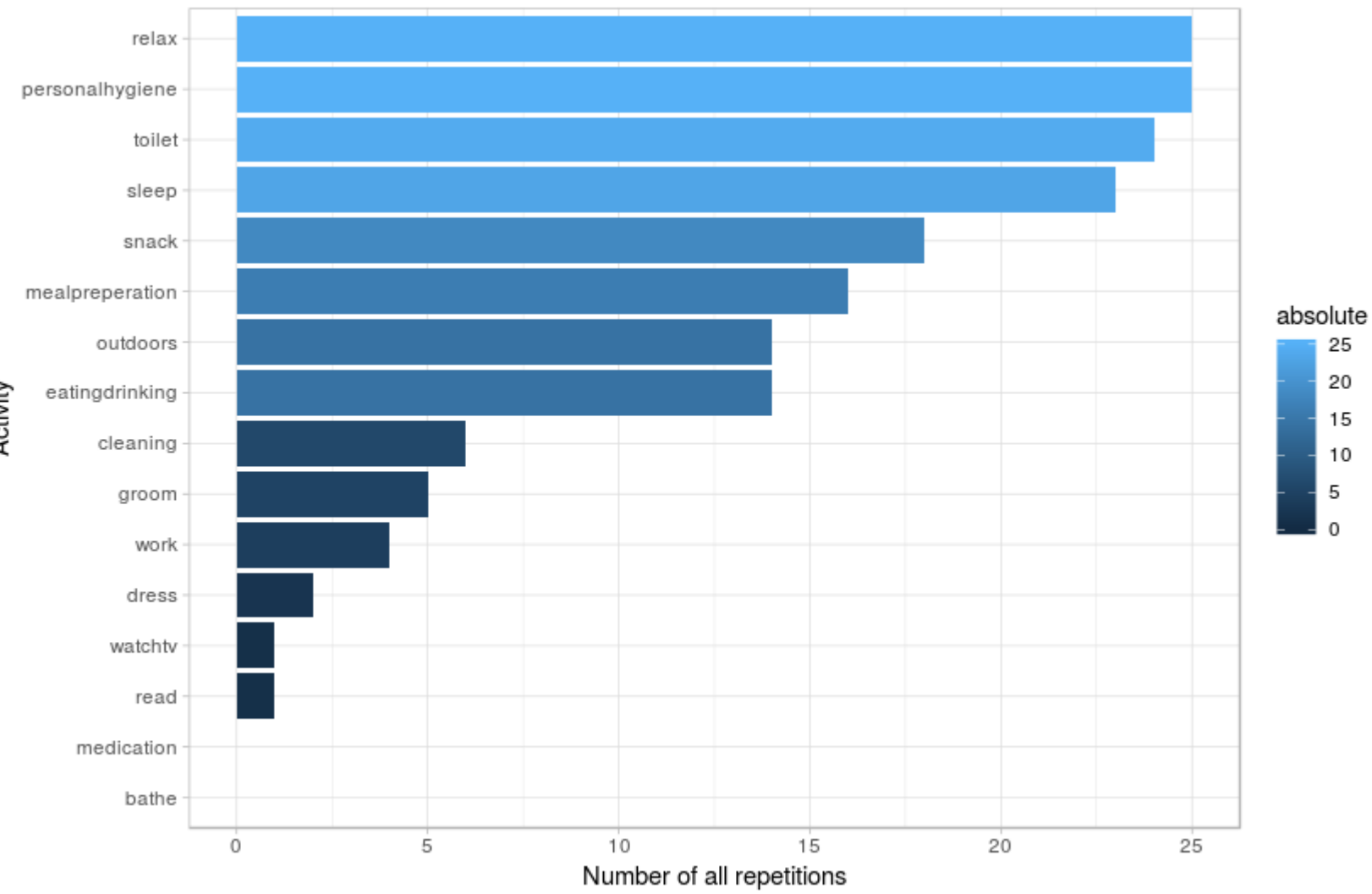
General process model



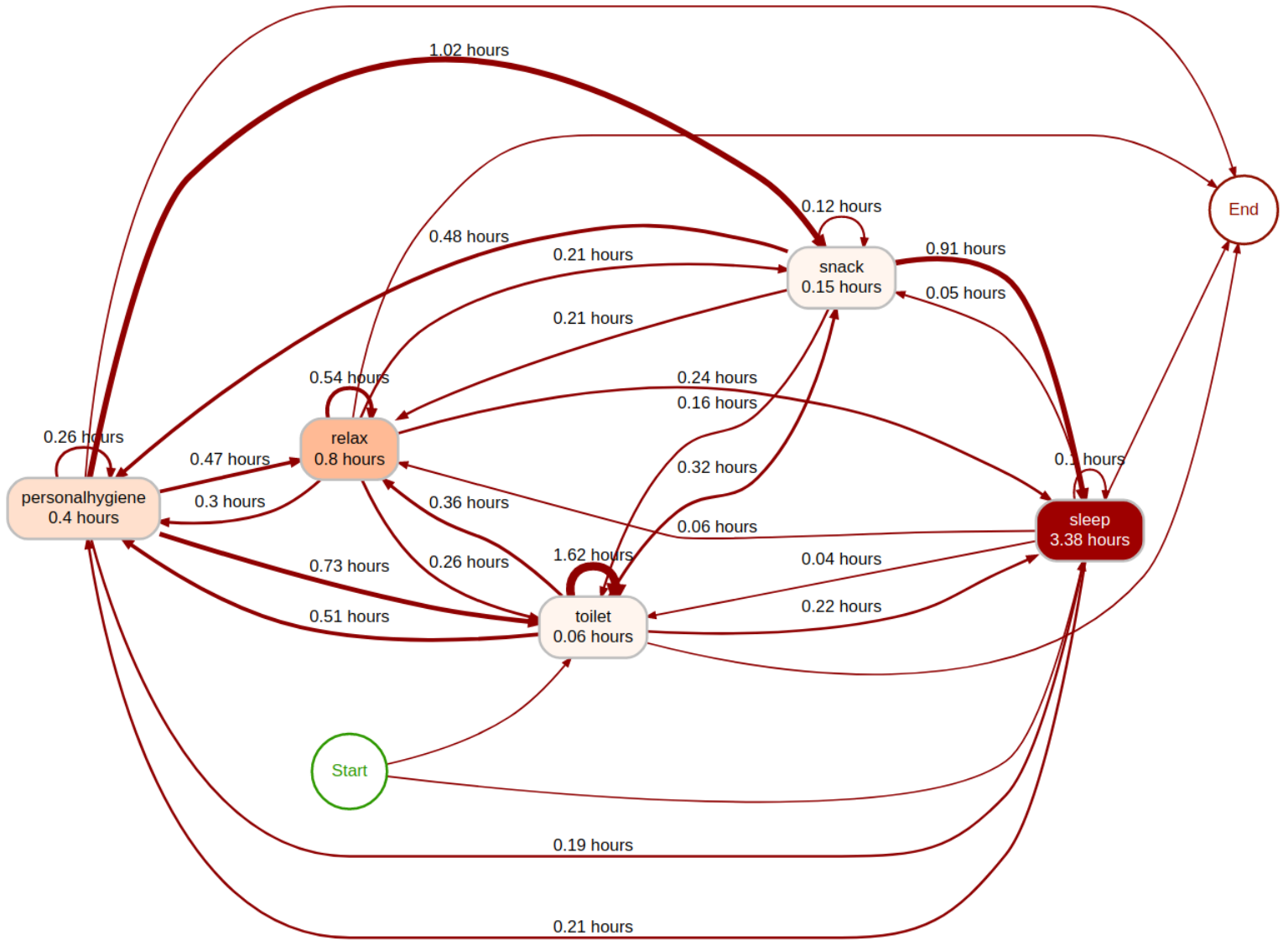
hh104 process model without medication



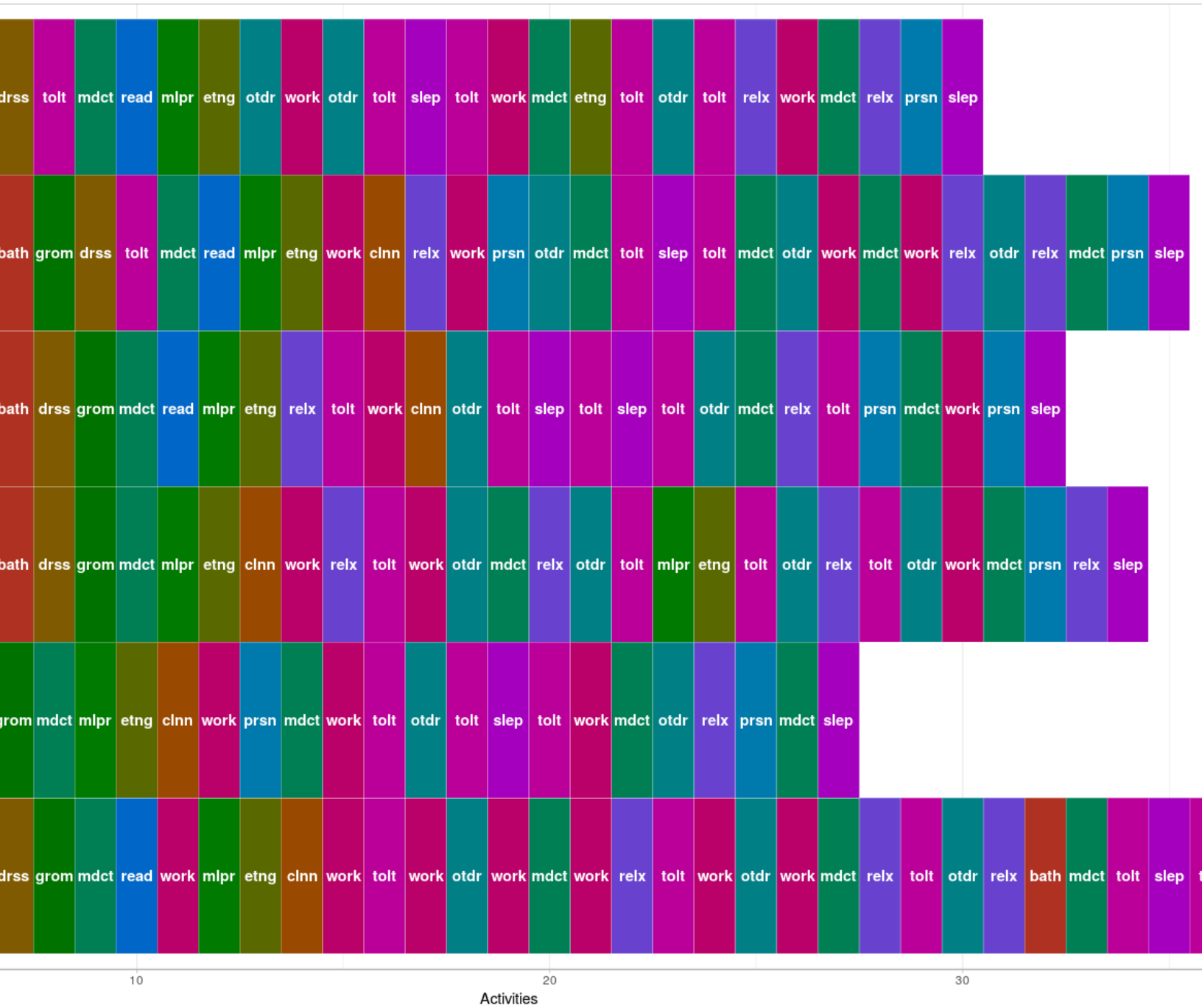
number of repetitions of each activity for hh102



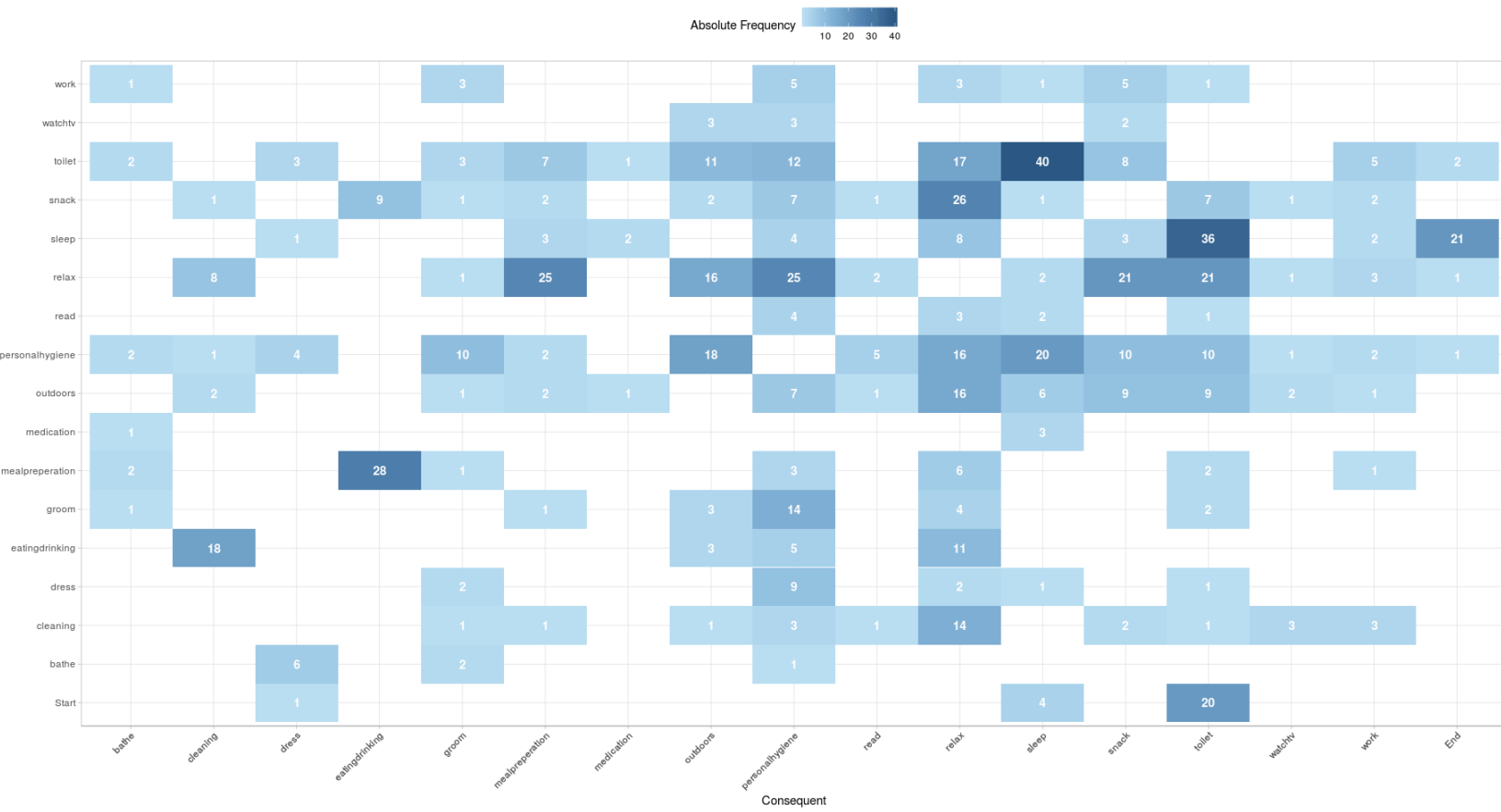
hh102 process map of processing time



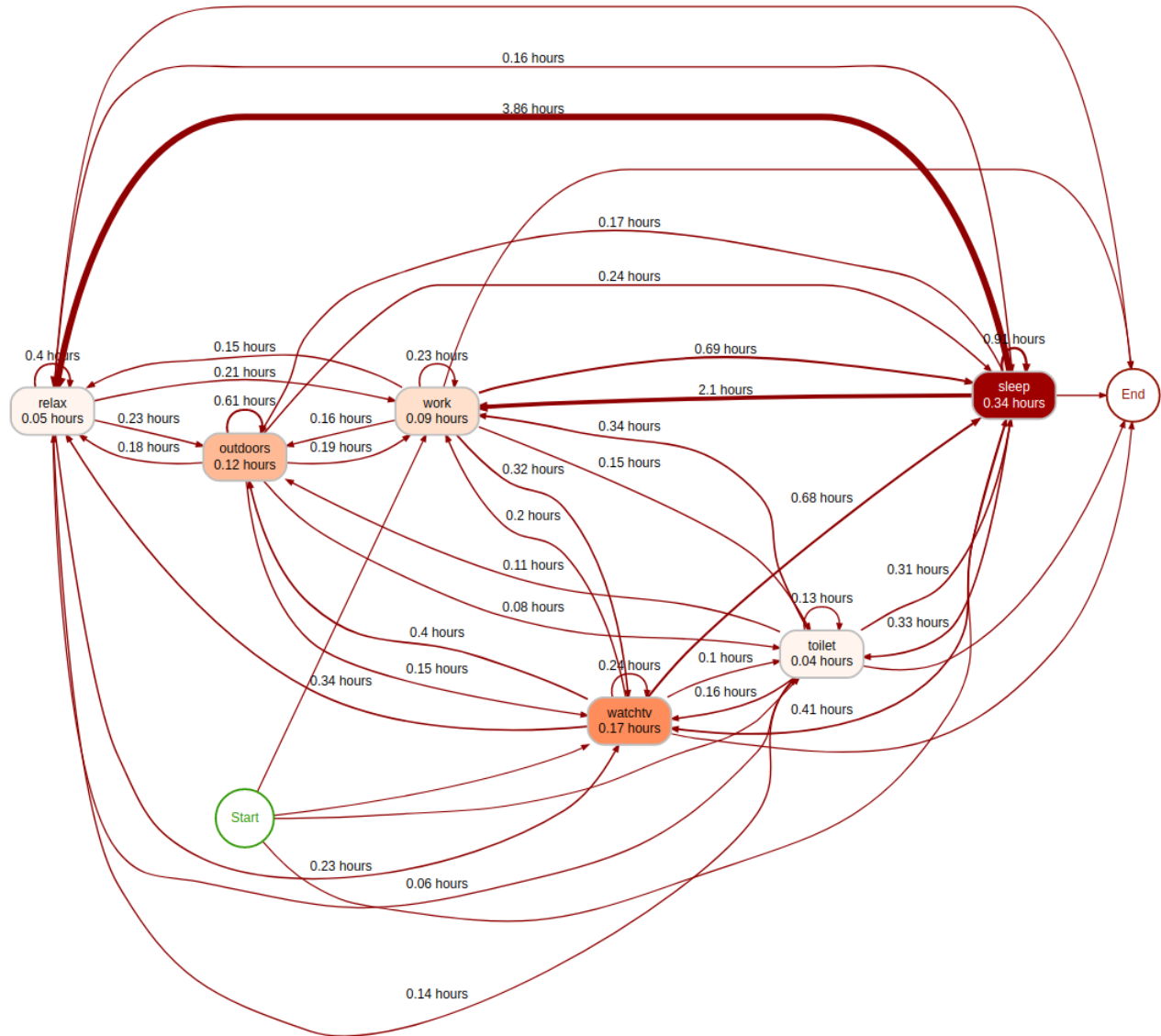
hh110 traces extract



hh102 precedence Matrix



working days process map



weekends process map

