In [19]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import warnings
import re
import os
from IPython.core.display import display, HTML
from tqdm.auto import tqdm


plt.style.use('seaborn')
%matplotlib inline
warnings.filterwarnings('ignore')
display(HTML("<style>.container { width:90% !important; }</style>"))
```

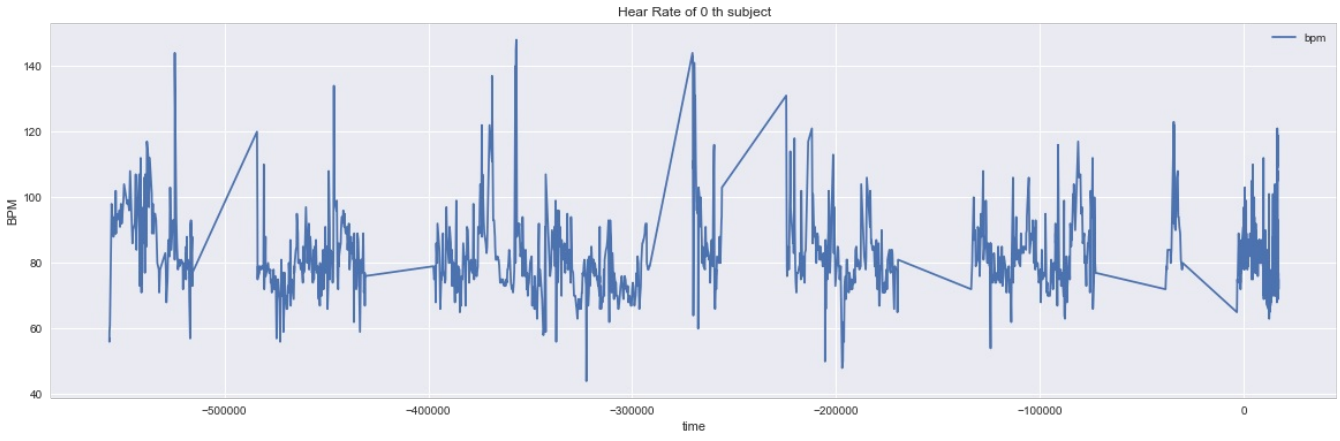
heart rate EDA

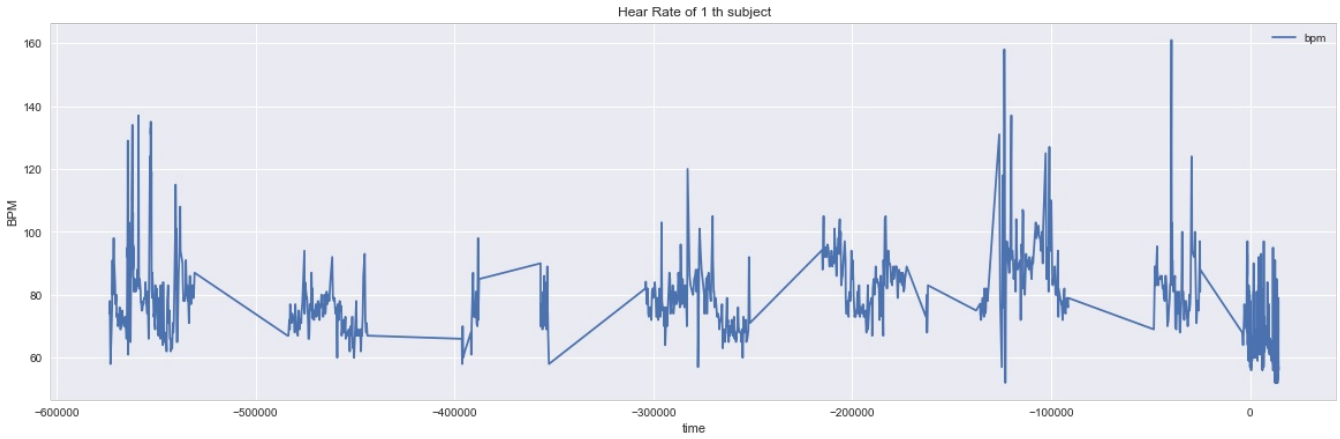
피험자 번호대로 0번부터 번호를 매긴 후 출력해본 결과,

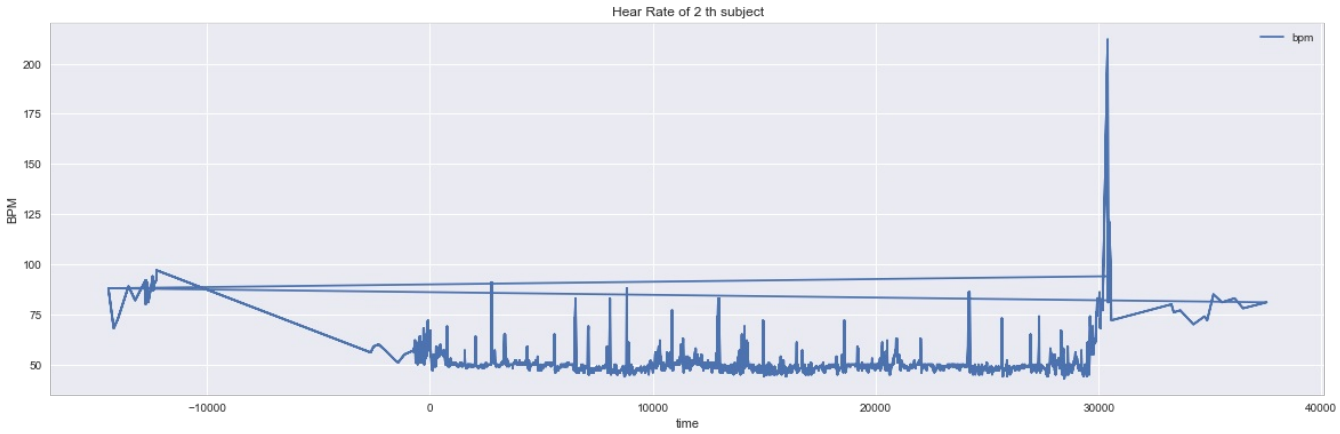
2, 4, 30번 피험자는 [시작 시간, 끝 시간] 구간이 두 번 중복해서 나왔음. 그래서 마지막 [시작 시간, 끝 시간] 구간만 저장했음.

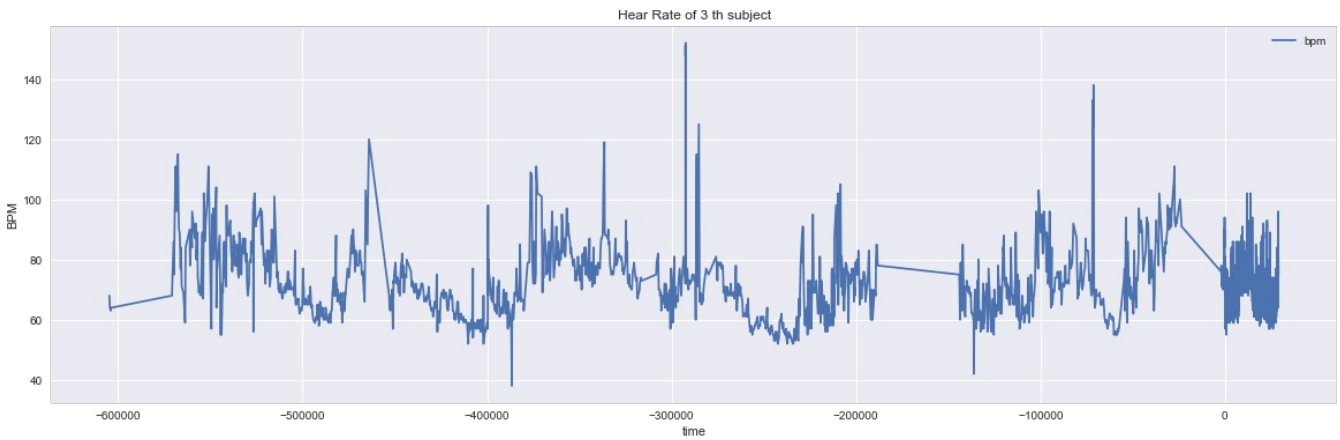
In [3]:
```
# heart rate 데이터 프레임 리스트
heart_rate = []

dir_name = '' # 파일 디렉토리 경로
path = '' # 데이터 파일 경로
file_list = os.listdir(path)

#  주어진 파일 이름에서 숫자 부분을 추출
# 파일 이름에서 정수 값을 추출하여 파일을 숫자순으로 정렬하는 데 사용
def get_number(filename):
    return int(re.search(r'\d+', filename).group())

# 리스트 파일 이름 순으로 정렬
file_list_txt = sorted([file for file in file_list], key=get_number)

# 텍스트 파일을 csv 파일로 저장
for i in file_list_txt:
    df = pd.read_csv(dir_name+i, header=None, names=['time', 'bpm'])
    heart_rate.append(df)
```

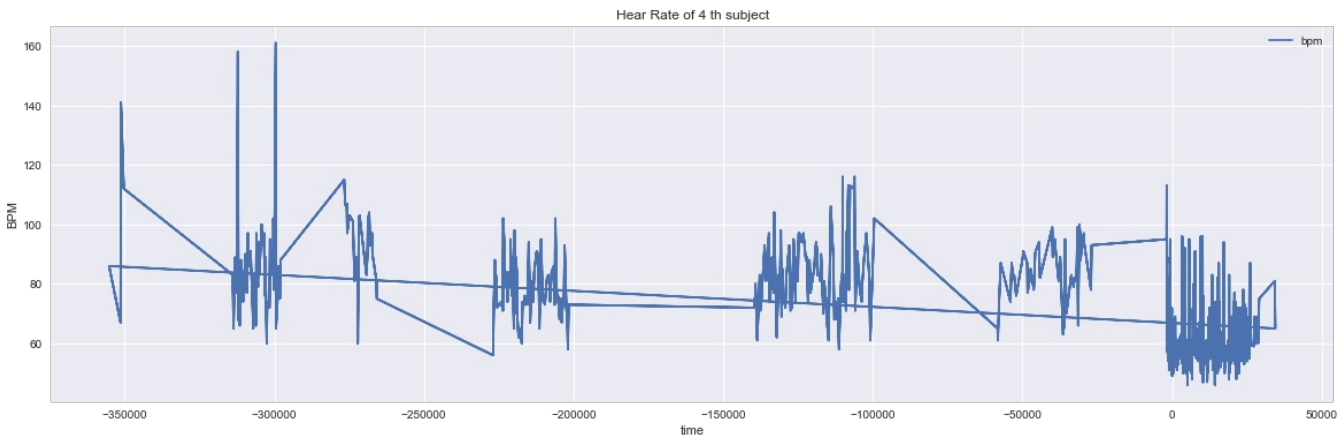
In [4]:
```
# 각 피험자별 heart rate 출력

for i in range(len(heart_rate)):
    f, ax = plt.subplots(1, 1, figsize=(20, 6))
    plt.title(f'Hear Rate of {i} th subject')
    plt.xlabel('Time')
    plt.ylabel('BPM')
    heart_rate[i].plot(x='time', y='bpm', kind='line', ax = ax)
    plt.show()
```

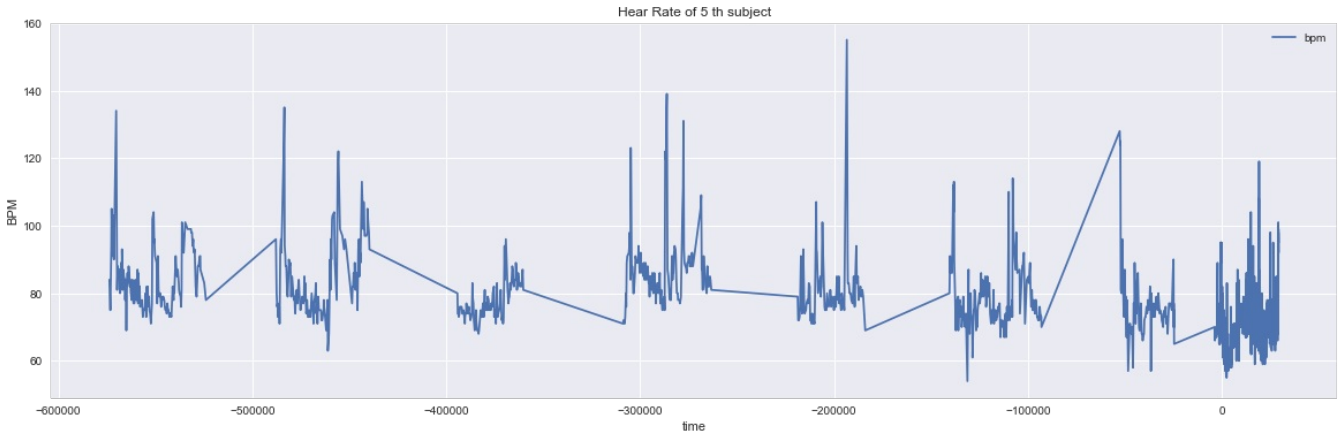


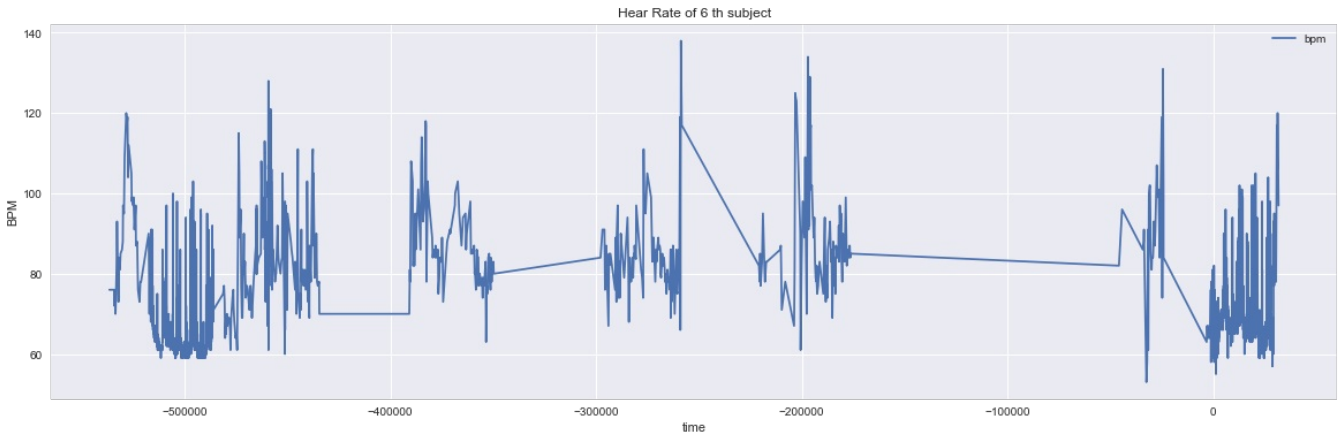

Hear Rate of 1 th subject
bpm
BPM
time
160
140
120
100
80
60
−600000
−500000
−400000
−300000
−200000
−100000
0

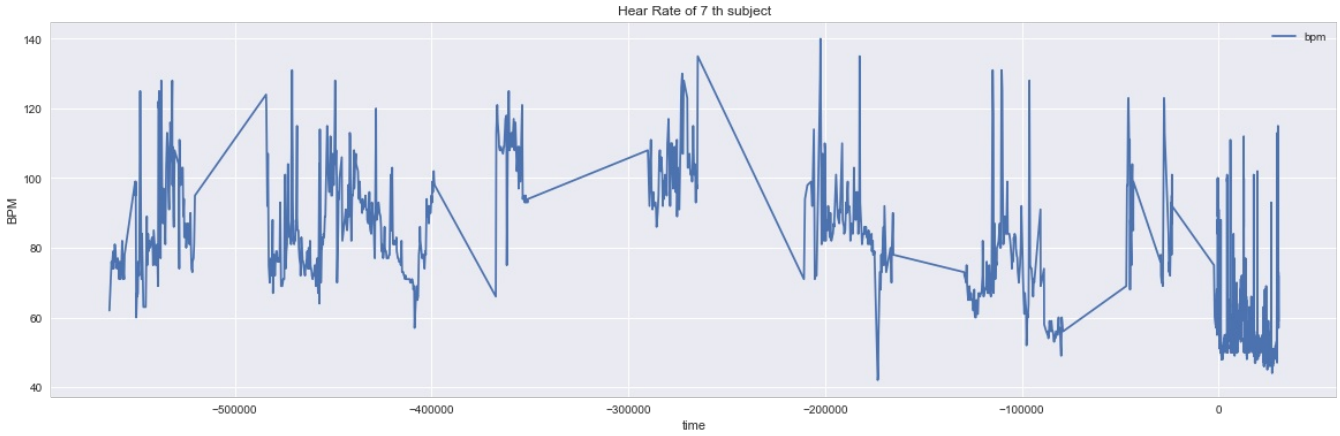

Hear Rate of 2 th subject
bpm
BPM
time
200
175
150
125
100
75
50
−10000
0
10000
20000
30000
40000

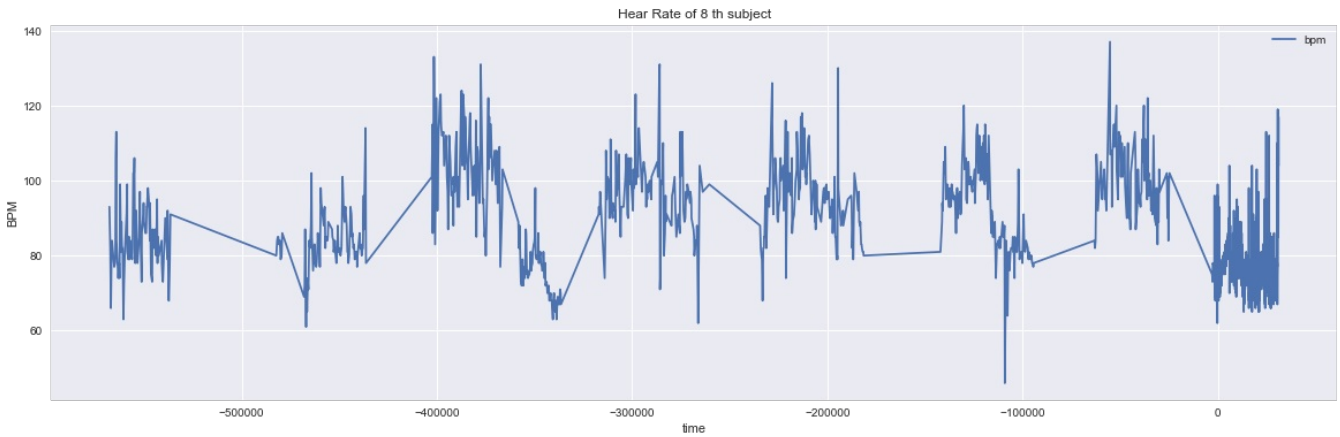

Hear Rate of 3 th subject
bpm
BPM
time
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

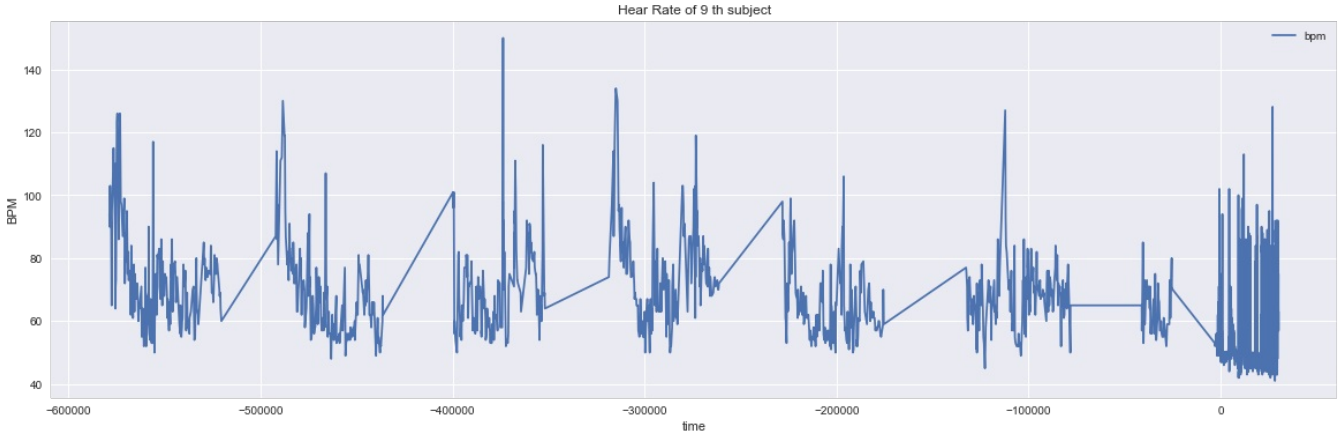

Hear Rate of 4 th subject
bpm
BPM
time
160
140
120
100
80
60
−350000
−300000
−250000
−200000
−150000
−100000
−50000
0
50000

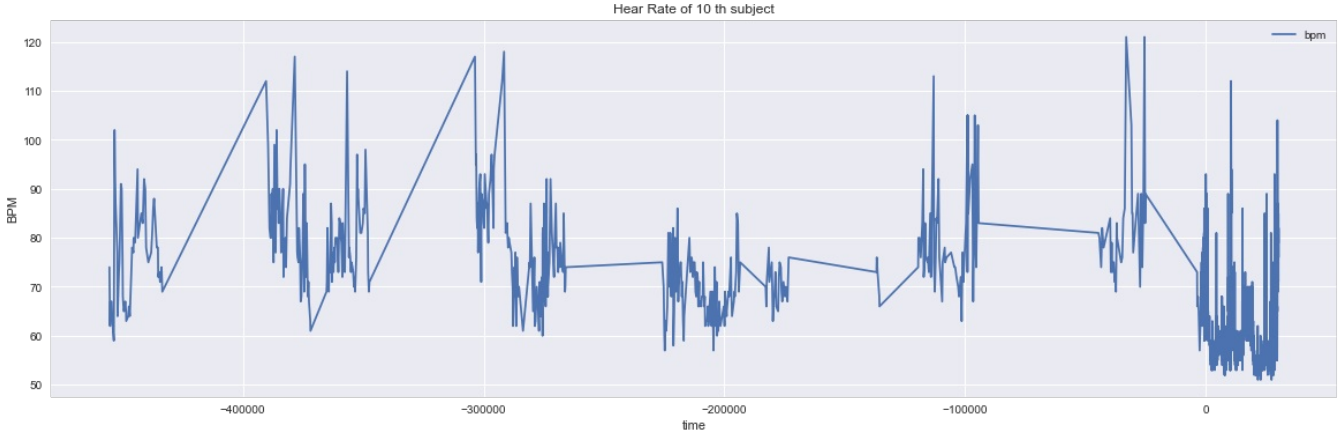

Hear Rate of 5 th subject
bpm
BPM
time
160
140
120
100
80
60
−600000
−500000
−400000
−300000
−200000
−100000
0

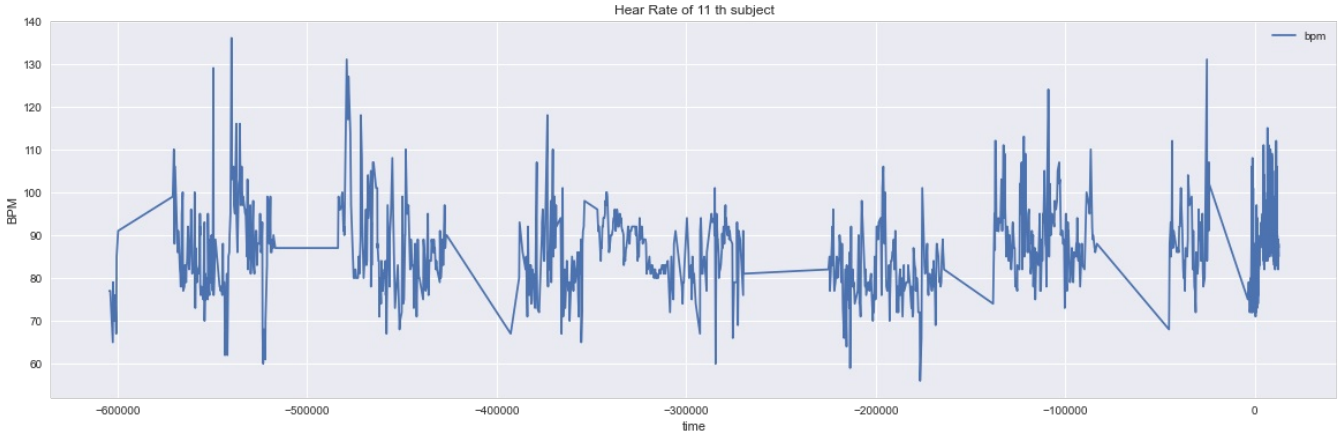

Hear Rate of 6 th subject
bpm
BPM
time
140
120
100
80
60
−500000
−400000
−300000
−200000
−100000
0

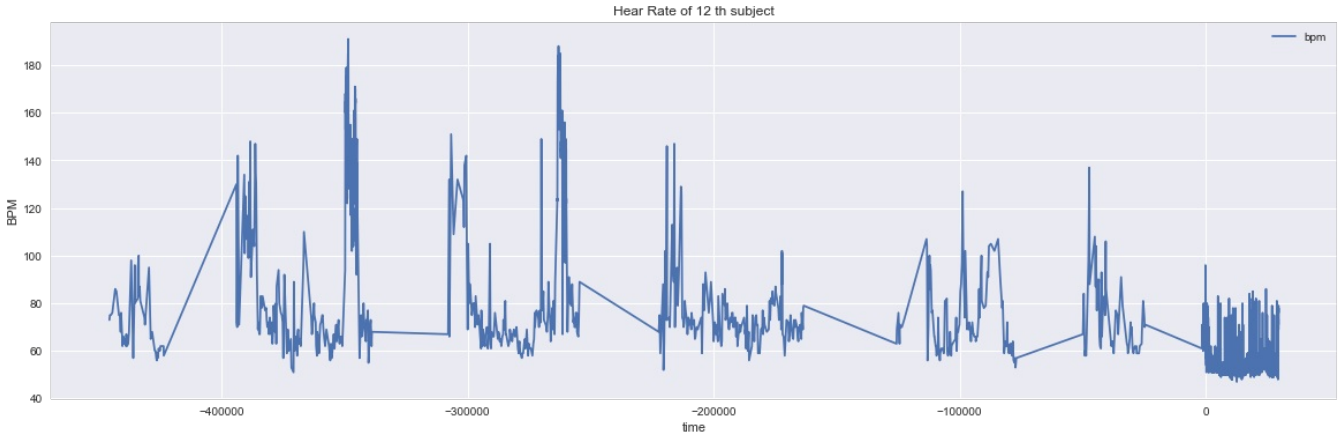

Hear Rate of 7 th subject
bpm
BPM
time
140
120
100
80
60
40
−500000
−400000
−300000
−200000
−100000
0

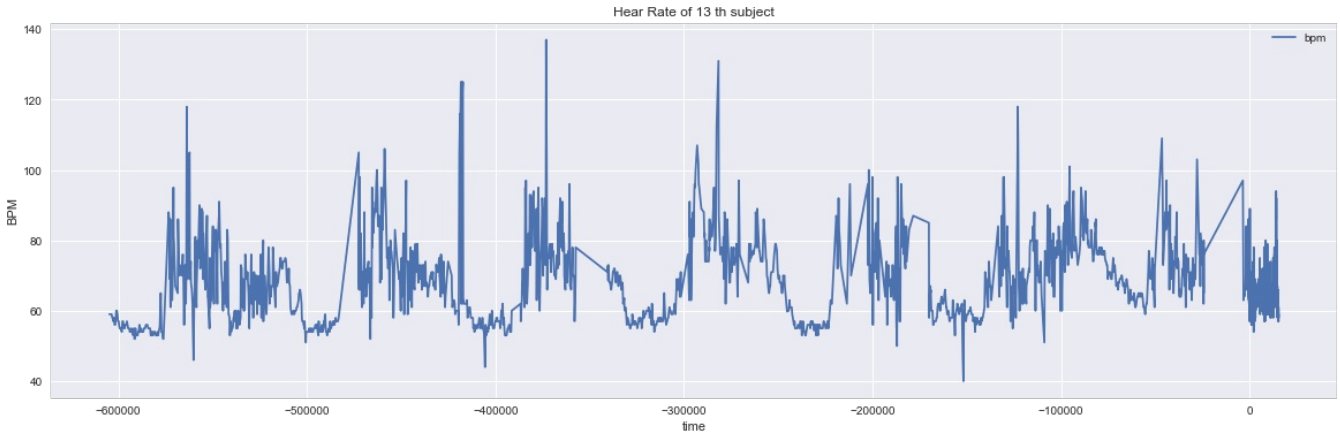

Hear Rate of 8 th subject
bpm
BPM
time
140
120
100
80
60
−500000
−400000
−300000
−200000
−100000
0

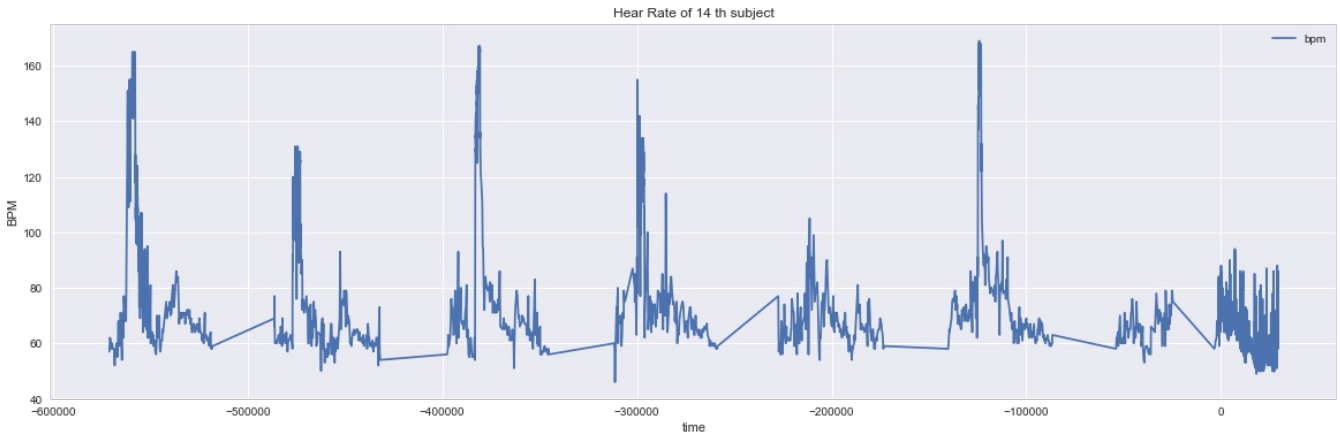

Hear Rate of 9 th subject
bpm
BPM
time
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

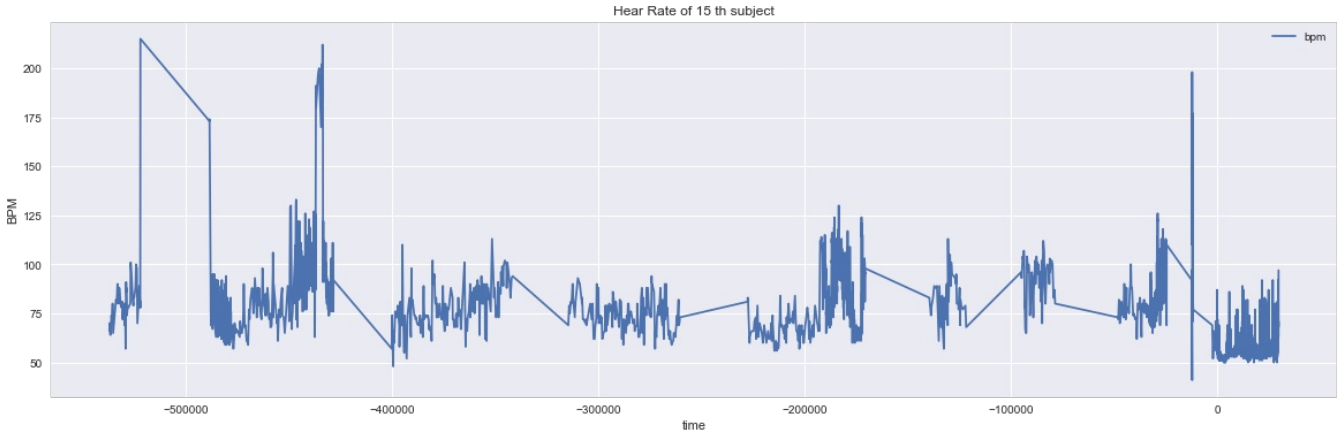

Hear Rate of 10 th subject
bpm
BPM
time
120
110
100
90
80
70
60
50
−400000
−300000
−200000
−100000
0

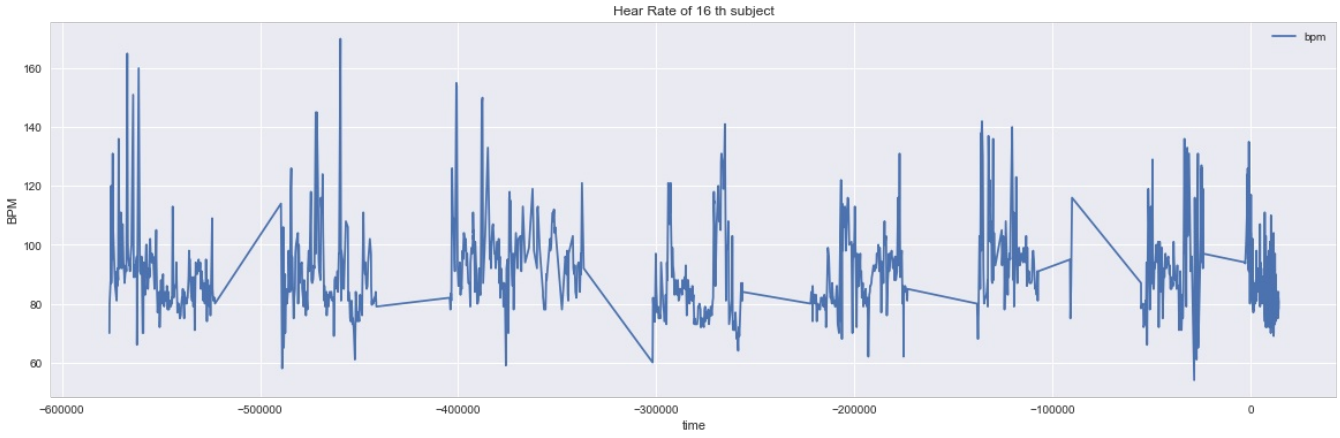

Hear Rate of 11 th subject
bpm
BPM
time
140
130
120
110
100
90
80
70
60
−600000
−500000
−400000
−300000
−200000
−100000
0

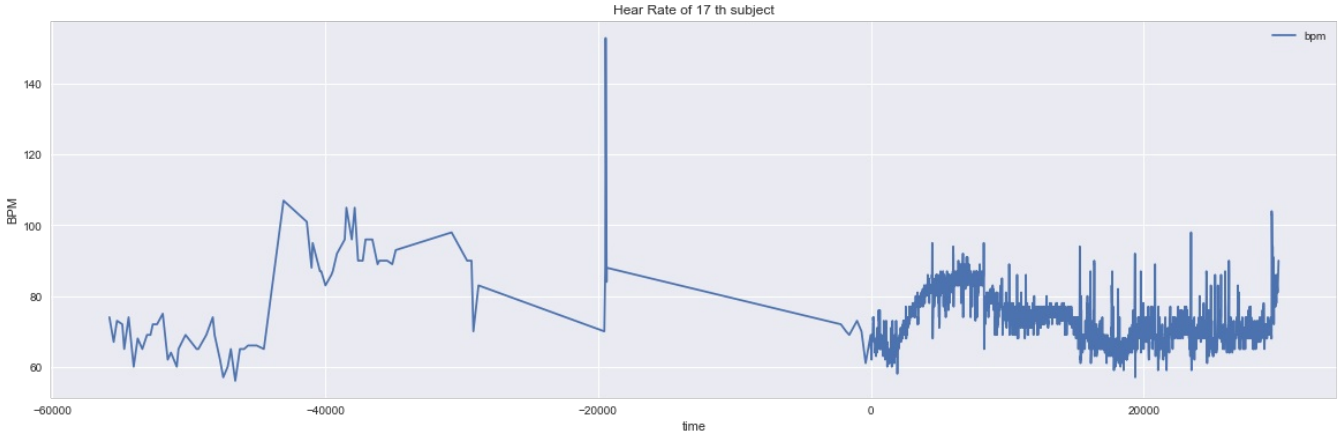

Hear Rate of 12 th subject
bpm
BPM
time
180
160
140
120
100
80
60
40
−400000
−300000
−200000
−100000
0

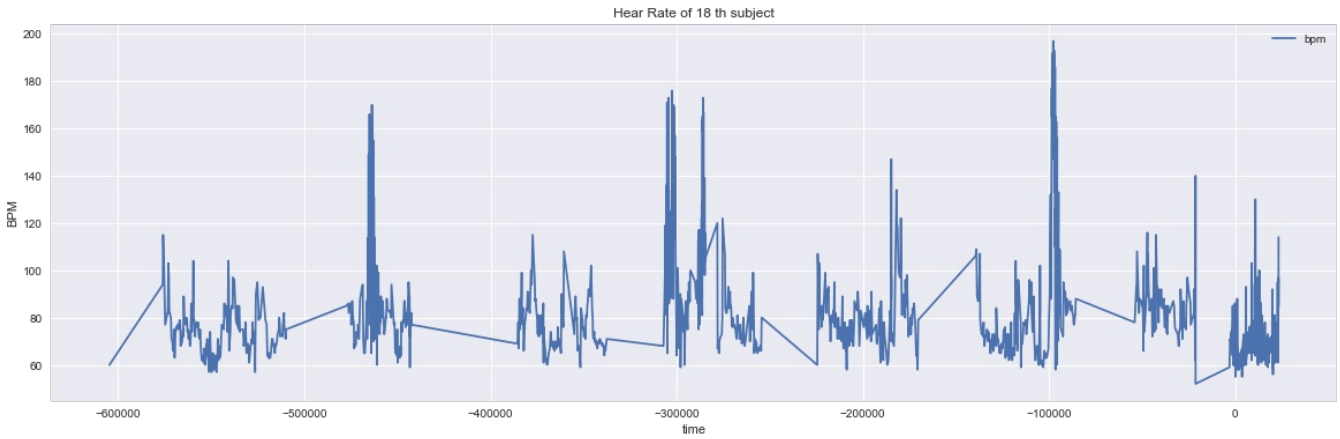

Hear Rate of 13 th subject
bpm
BPM
time
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

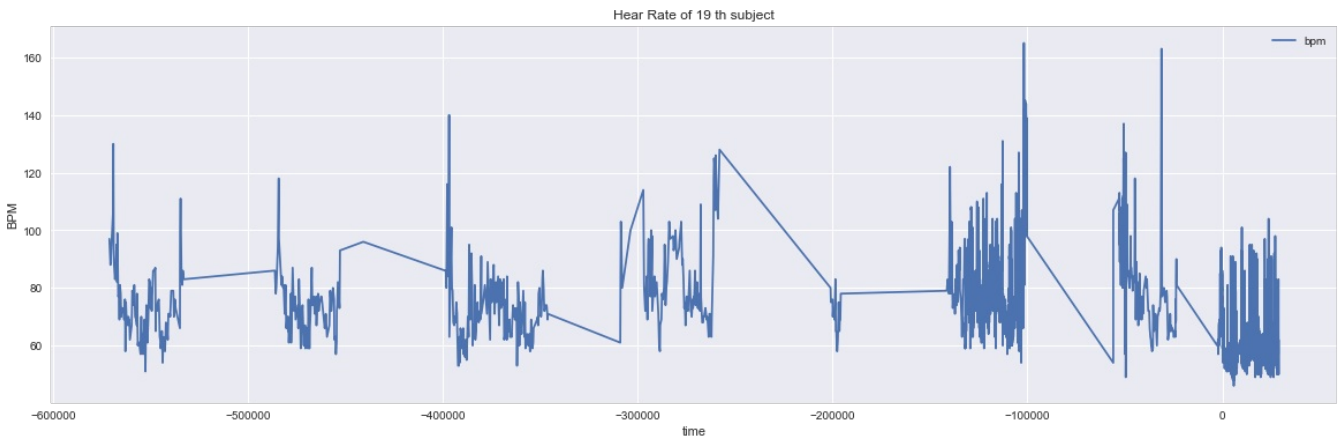

Hear Rate of 14 th subject
bpm
BPM
time
160
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

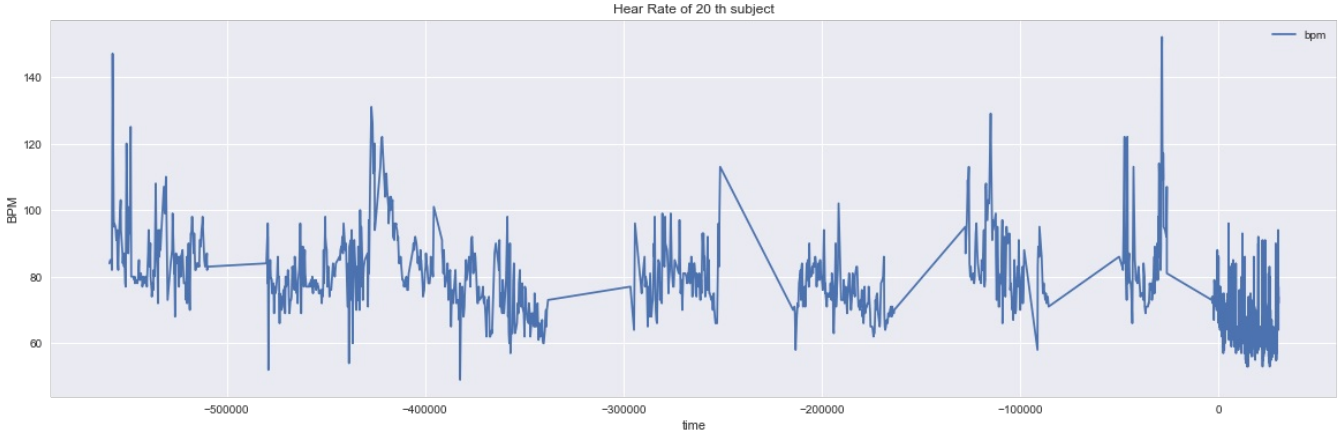

Hear Rate of 15 th subject
bpm
BPM
time
200
175
150
125
100
75
50
−500000
−400000
−300000
−200000
−100000
0

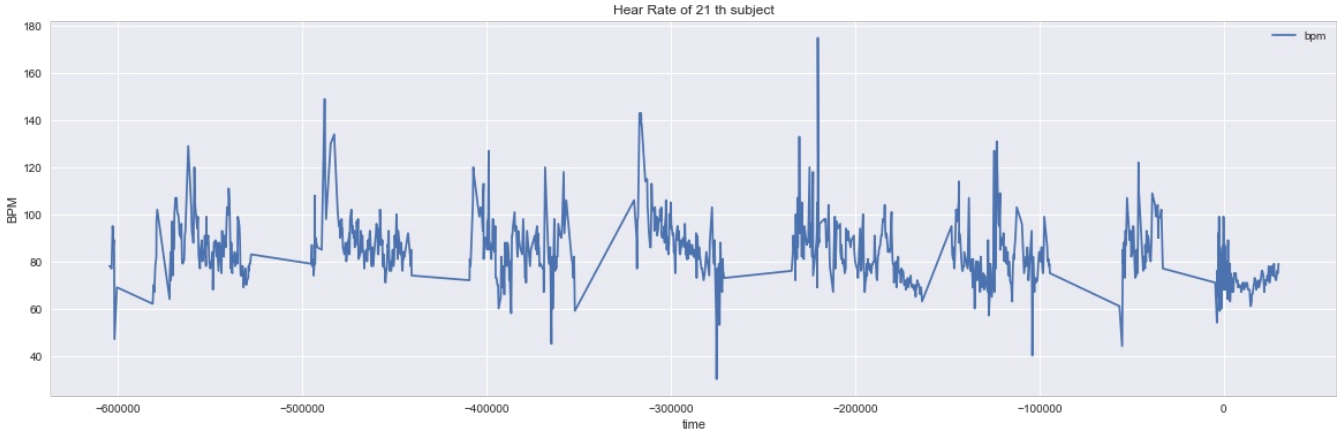

Hear Rate of 16 th subject
bpm
BPM
time
160
140
120
100
80
60
−600000
−500000
−400000
−300000
−200000
−100000
0

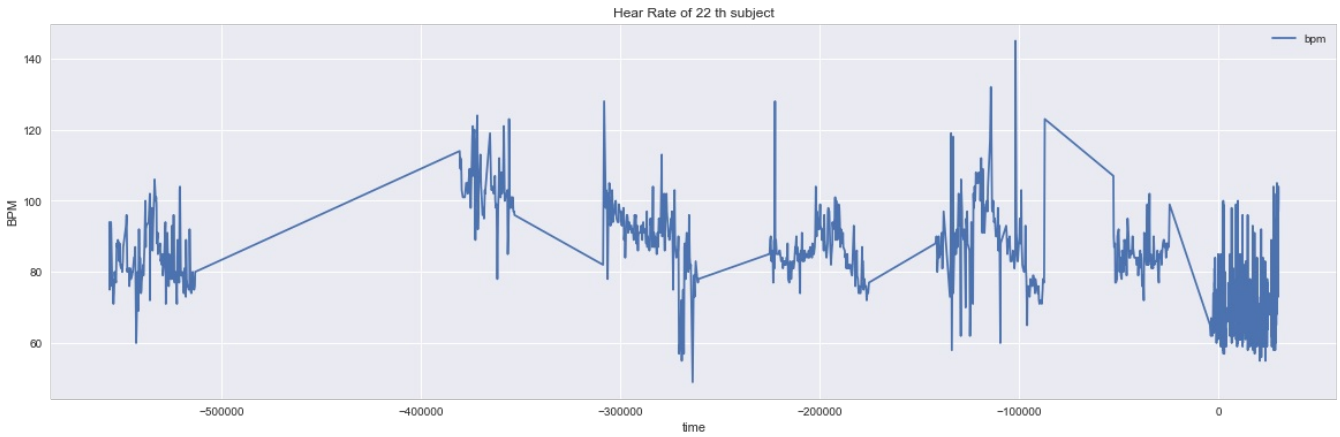

Hear Rate of 17 th subject
bpm
BPM
time
−60000
−40000
−20000
0
20000
60
80
100
120
140

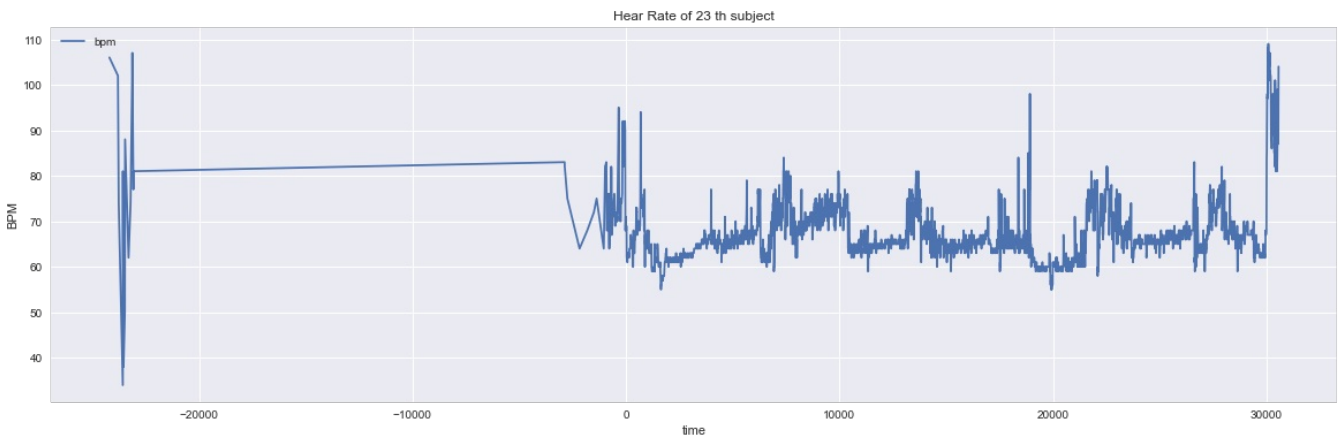

Hear Rate of 18 th subject
bpm
BPM
time
−600000
−500000
−400000
−300000
−200000
−100000
0
60
80
100
120
140
160
180
200

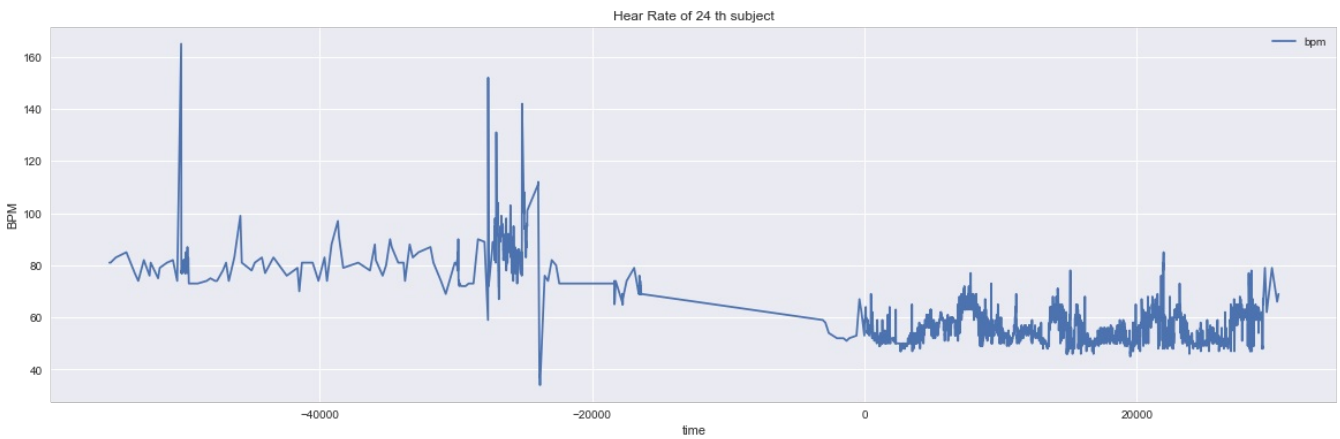

Hear Rate of 19 th subject
bpm
BPM
time
−600000
−500000
−400000
−300000
−200000
−100000
0
60
80
100
120
140
160

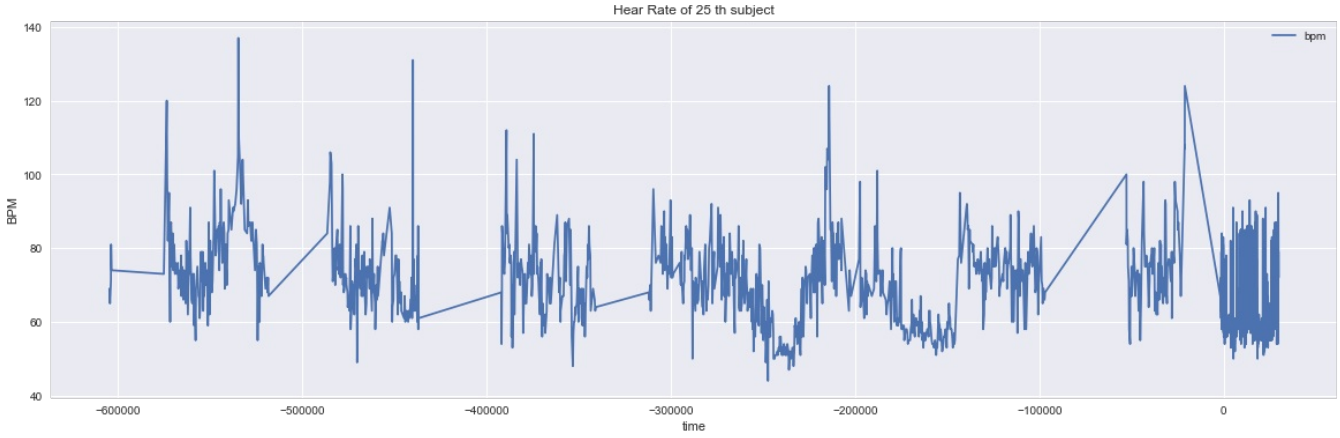

Hear Rate of 20 th subject
bpm
BPM
time
−500000
−400000
−300000
−200000
−100000
0
60
80
100
120
140

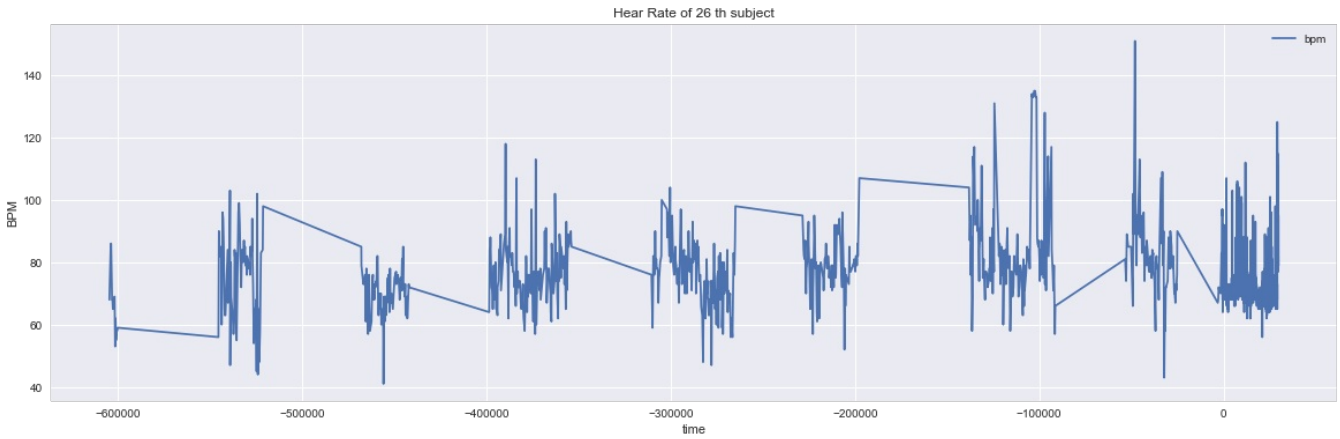

Hear Rate of 21 th subject
bpm
BPM
time
180
160
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

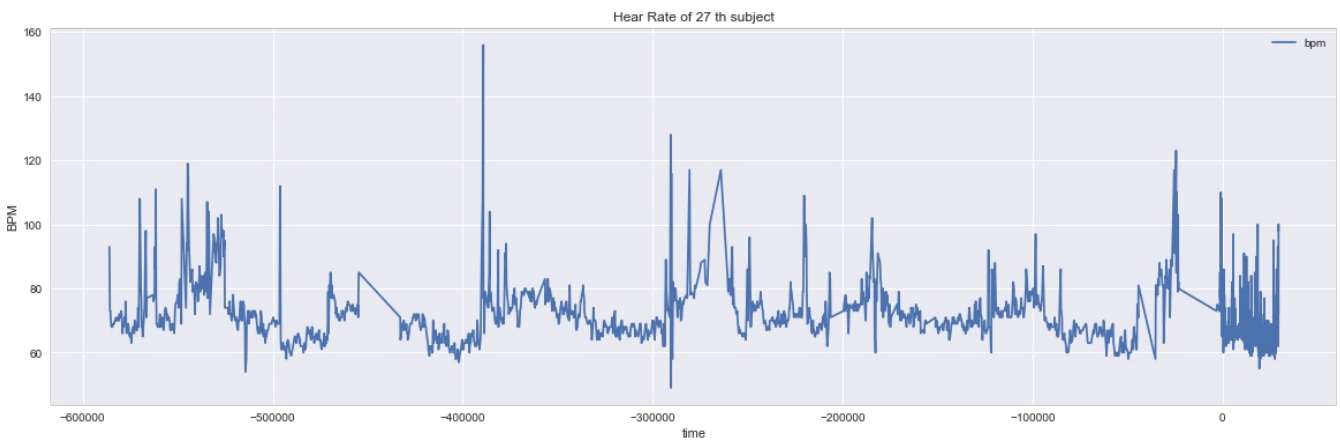

Hear Rate of 22 th subject
bpm
BPM
time
140
120
100
80
60
−500000
−400000
−300000
−200000
−100000
0

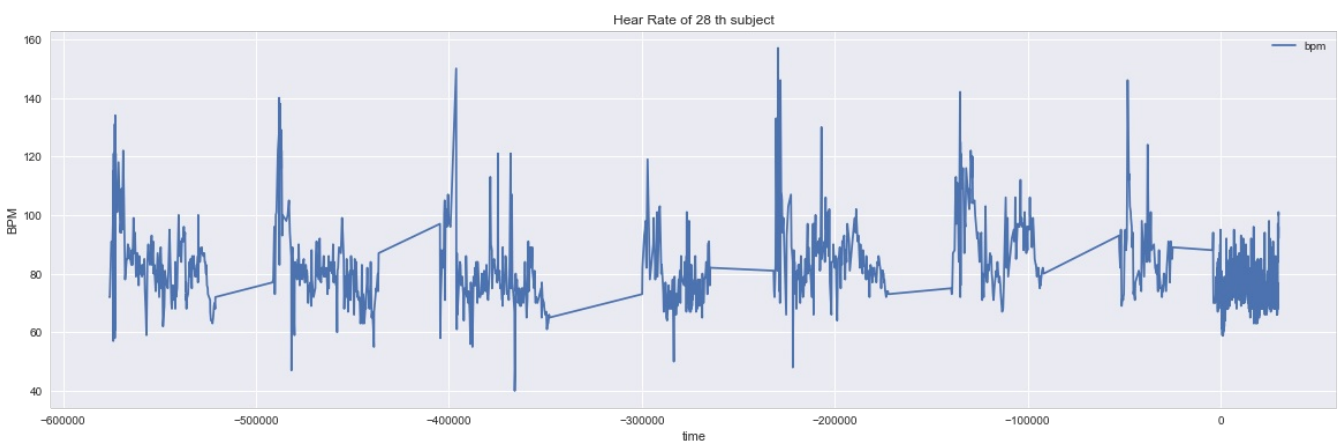

Hear Rate of 23 th subject
bpm
BPM
time
110
100
90
80
70
60
50
40
−20000
−10000
0
10000
20000
30000

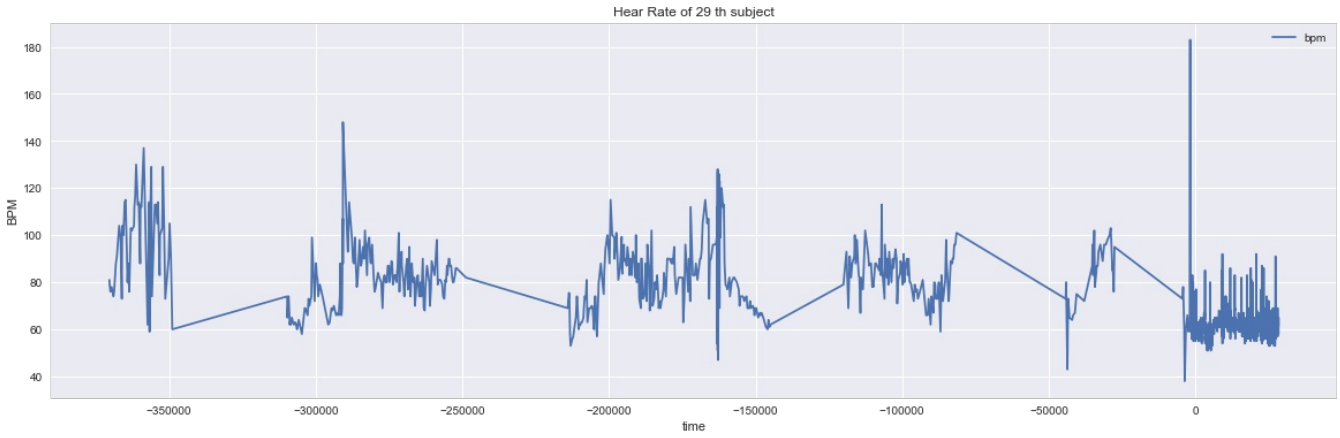

Hear Rate of 24 th subject
bpm
BPM
time
160
140
120
100
80
60
40
−40000
−20000
0
20000

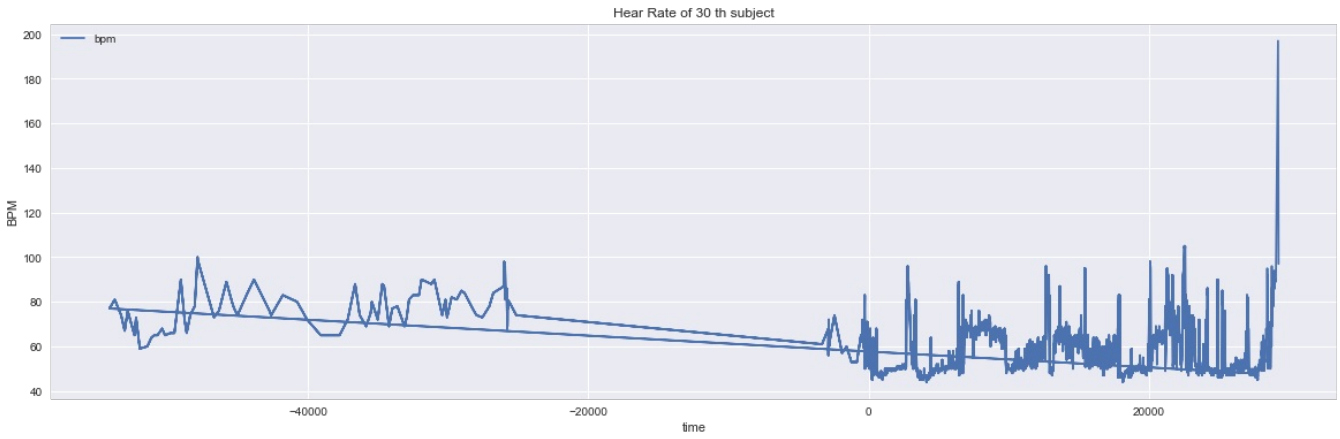

Hear Rate of 25 th subject
bpm
BPM
time
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0


Hear Rate of 26 th subject
bpm
BPM
time
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

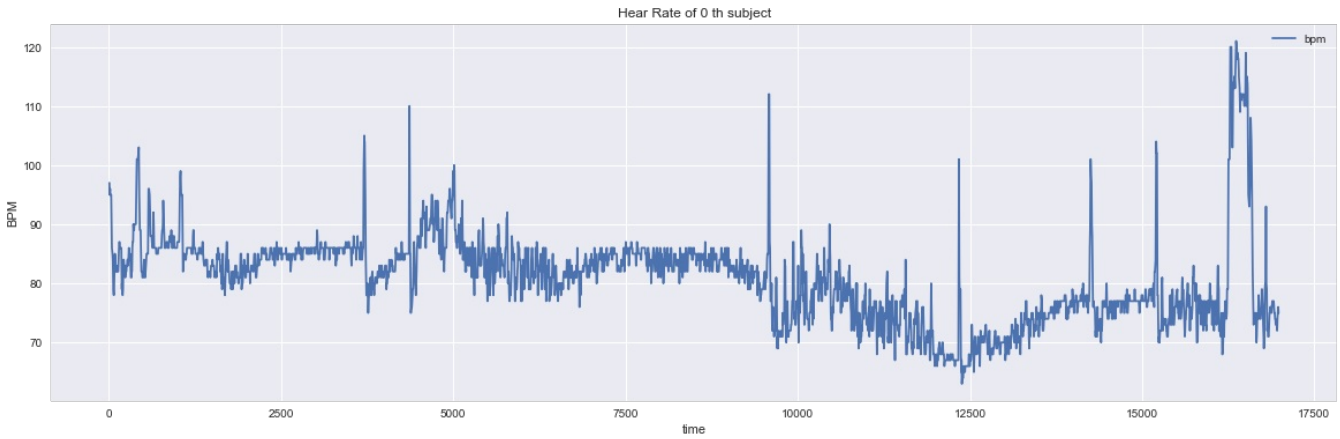

Hear Rate of 27 th subject
bpm
BPM
time
160
140
120
100
80
60
−600000
−500000
−400000
−300000
−200000
−100000
0

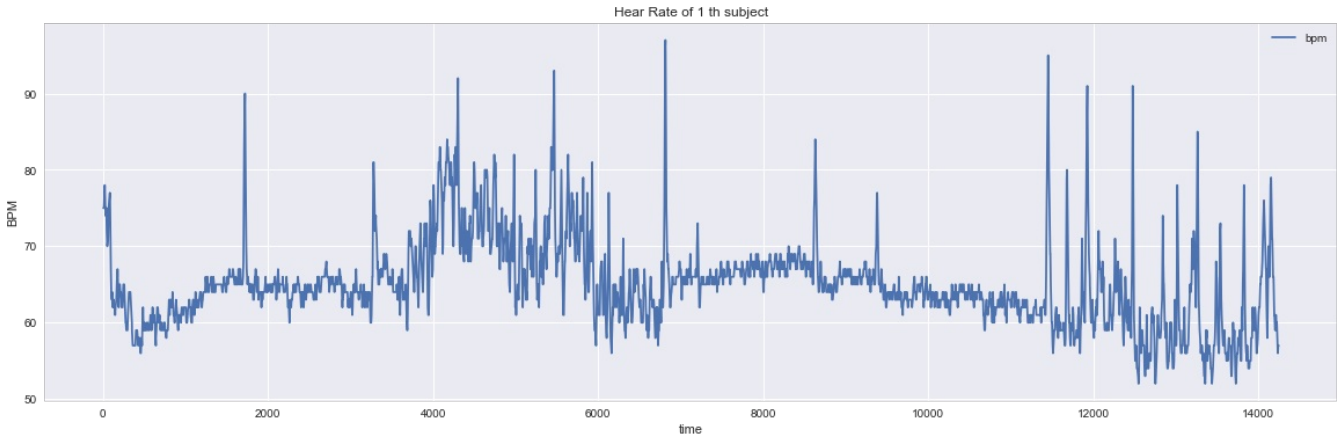

Hear Rate of 28 th subject
bpm
BPM
time
160
140
120
100
80
60
40
−600000
−500000
−400000
−300000
−200000
−100000
0

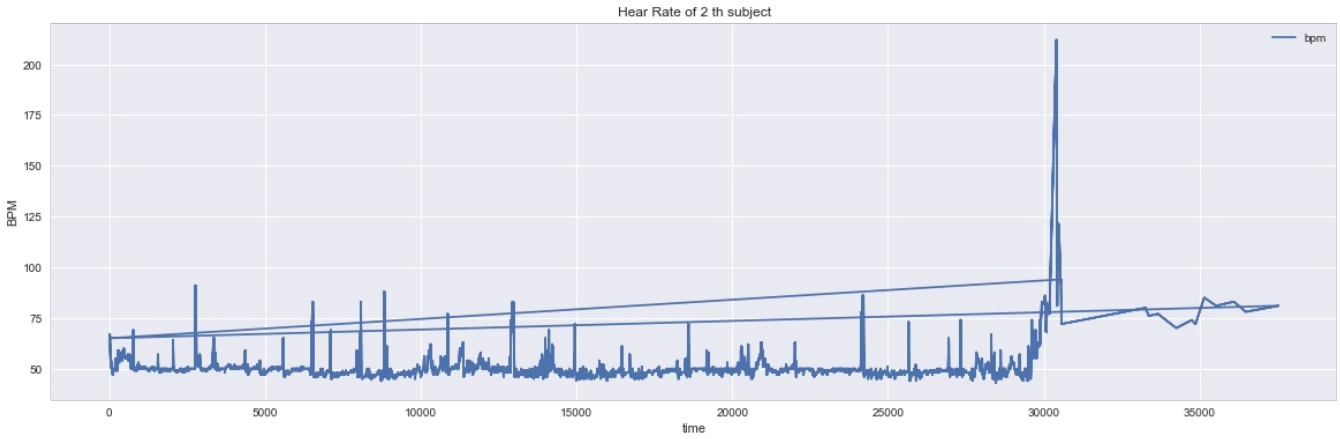


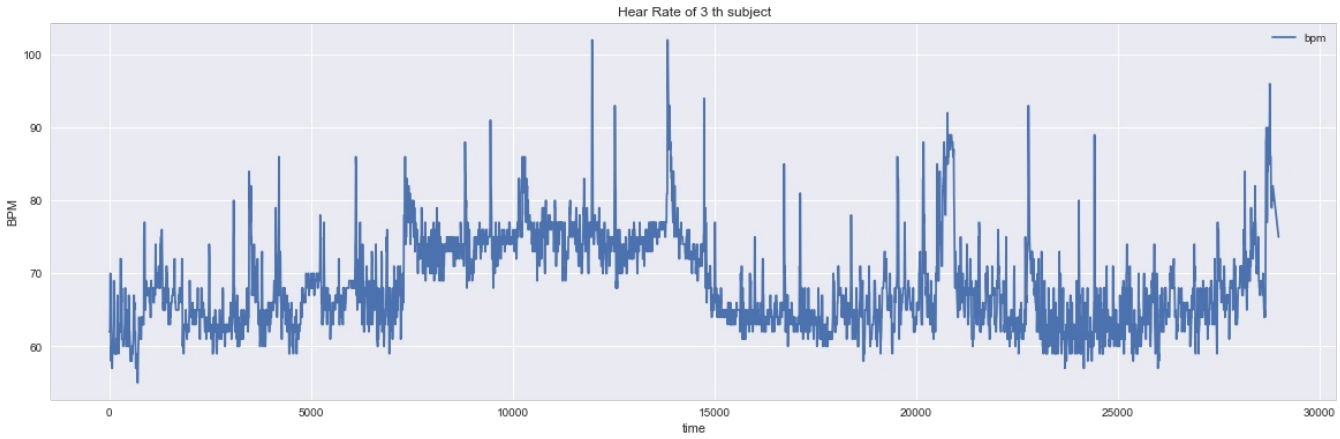


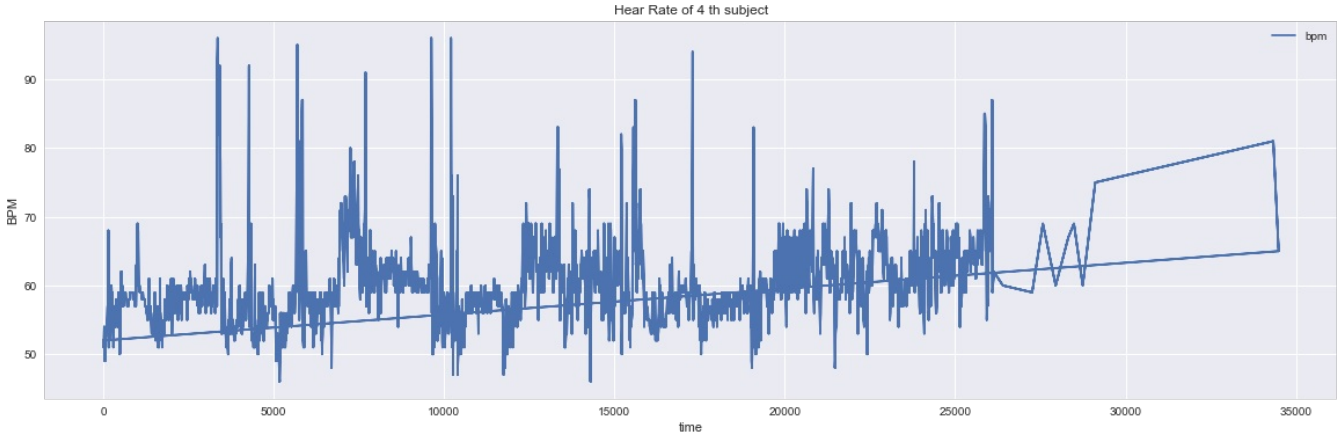
In [5]:
```
# PSG 측정 시작 시간 이후의 heart rate 확인

for i in range(len(heart_rate)):
    f, ax = plt.subplots(1, 1, figsize=(20, 6))
    plt.title(f'Hear Rate of {i} th subject')
    plt.xlabel('Time')
    plt.ylabel('BPM')
    heart_rate[i][heart_rate[i]['time'] >=0].plot(x='time', y='bpm', kind='line', ax = ax)
    plt.show()
```

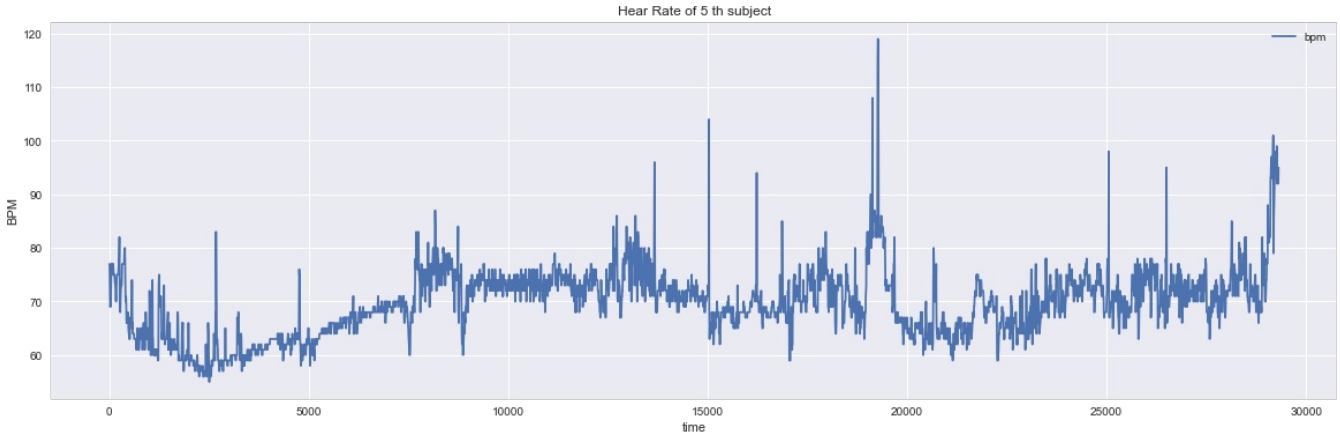


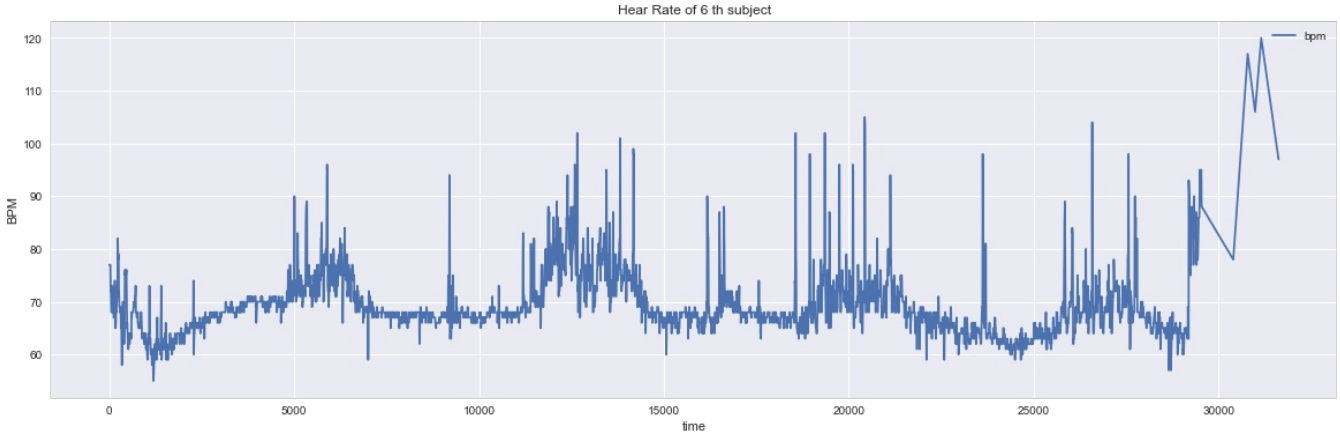


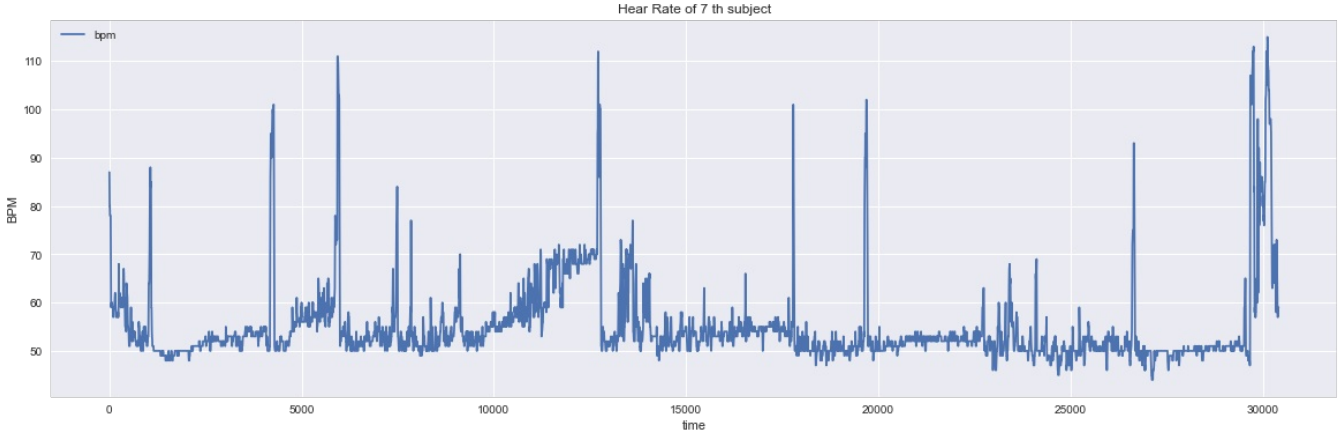

Hear Rate of 2 th subject
bpm
BPM
time

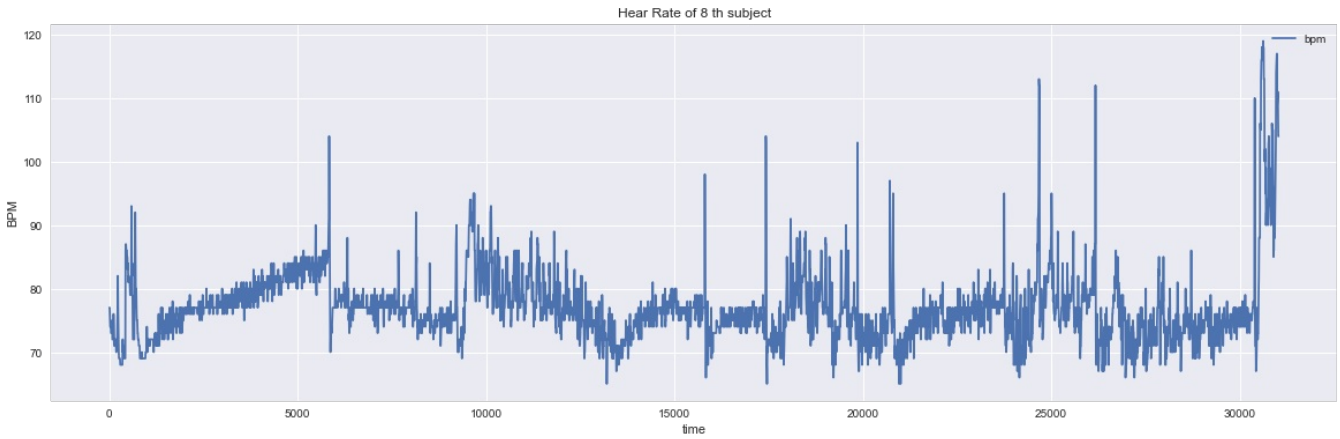

Hear Rate of 3 th subject
bpm
BPM
time

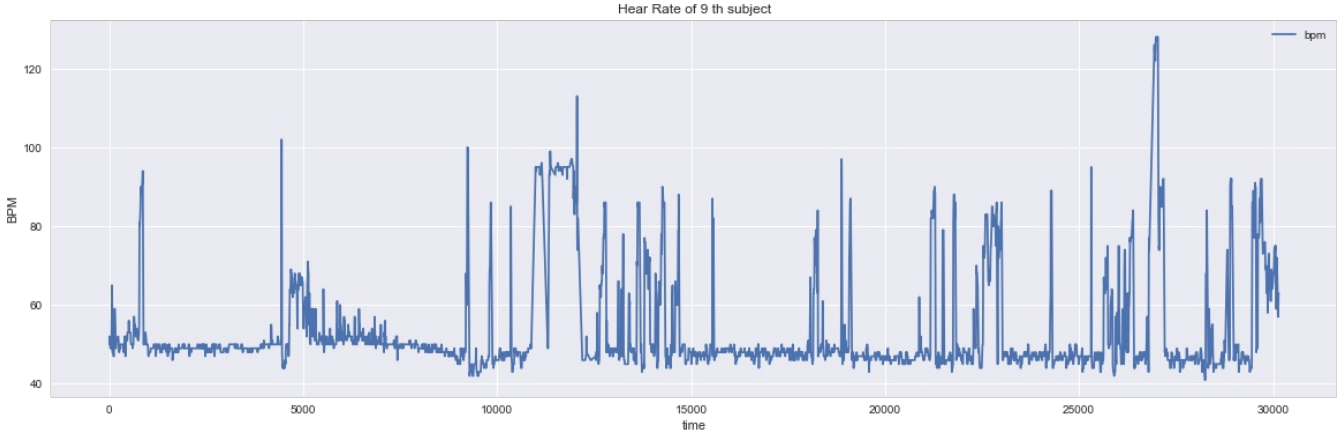

Hear Rate of 4 th subject
bpm
BPM
time

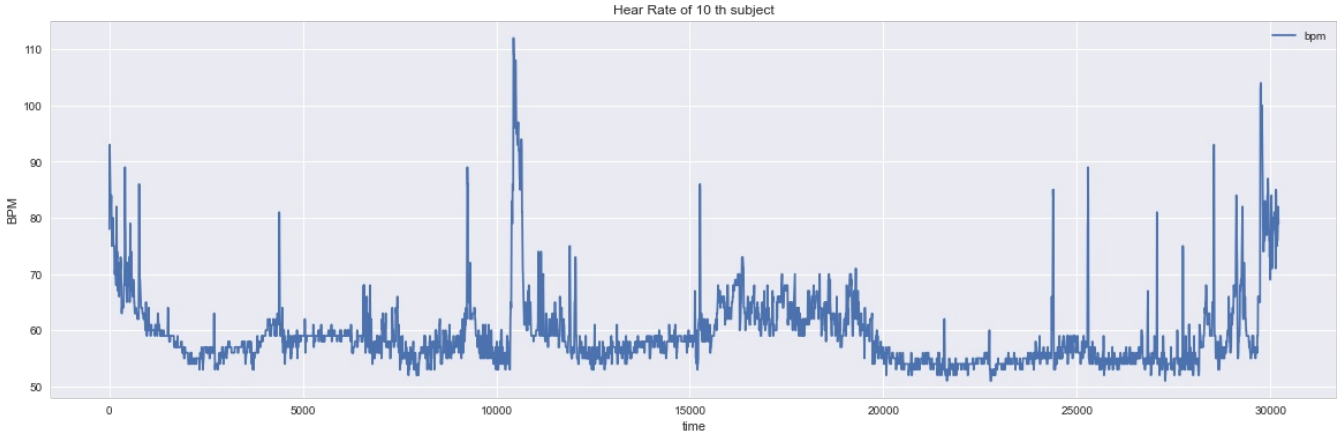

Hear Rate of 5 th subject
bpm
BPM
time

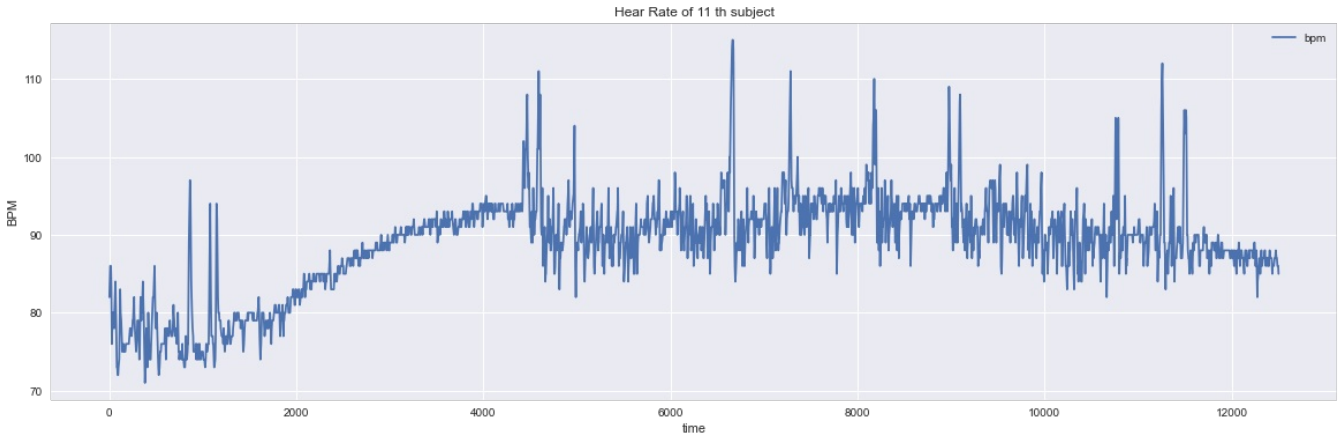

Hear Rate of 6 th subject
bpm
BPM
time
120
110
100
90
80
70
60
0
5000
10000
15000
20000
25000
30000

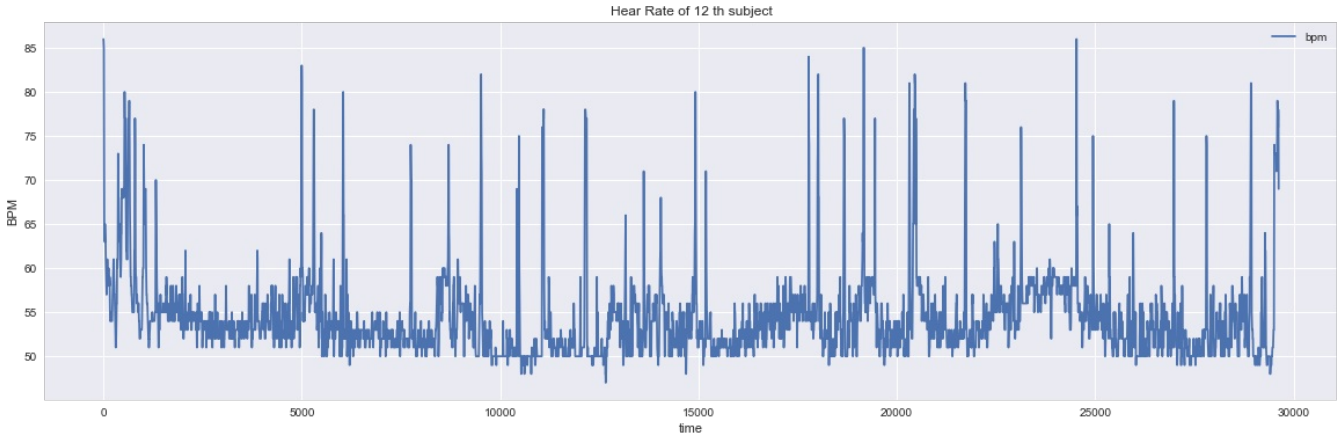

Hear Rate of 7 th subject
bpm
BPM
time
110
100
90
80
70
60
50
0
5000
10000
15000
20000
25000
30000

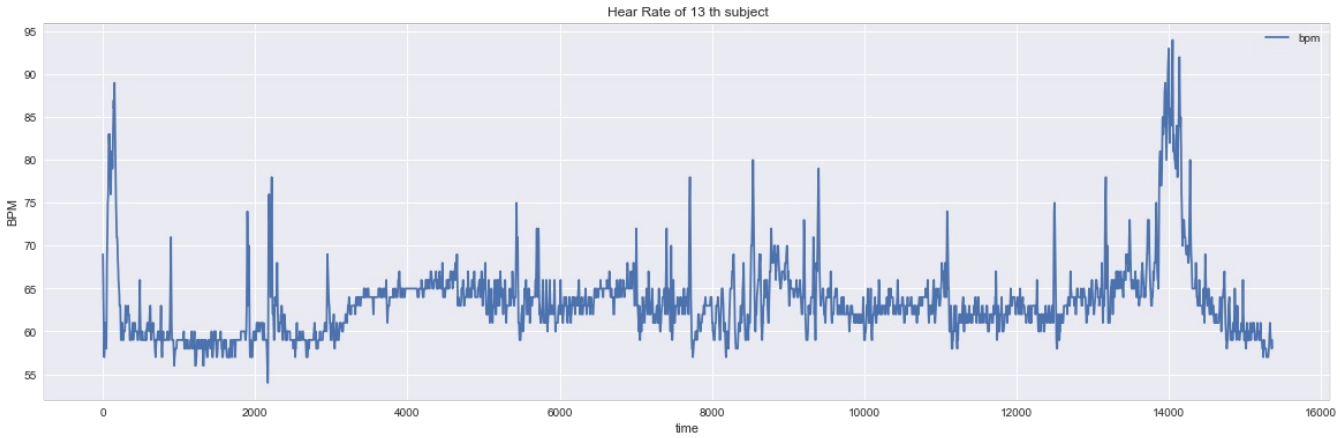

Hear Rate of 8 th subject
bpm
BPM
time
120
110
100
90
80
70
0
5000
10000
15000
20000
25000
30000

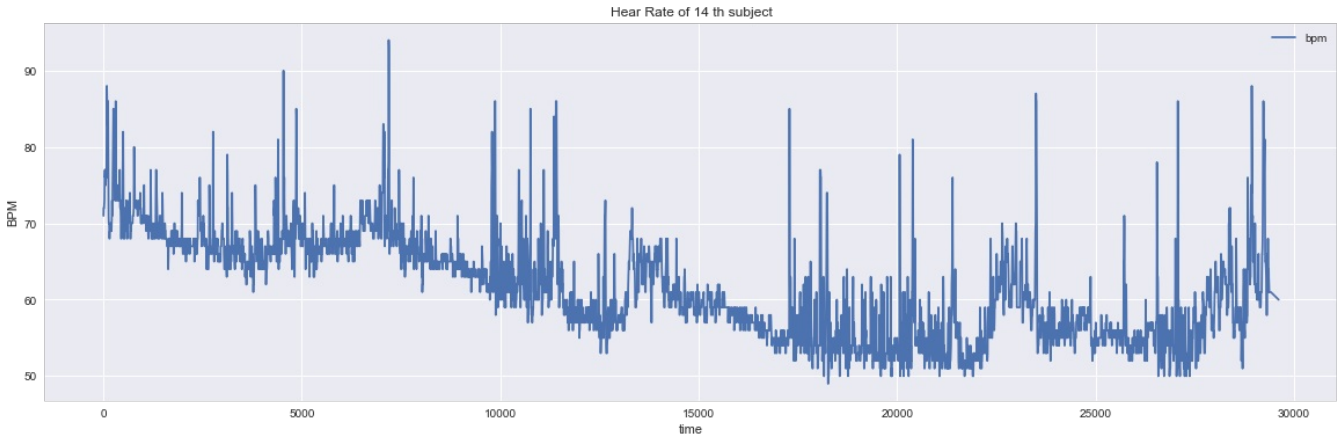

Hear Rate of 9 th subject
bpm
BPM
time
120
100
80
60
40
0
5000
10000
15000
20000
25000
30000

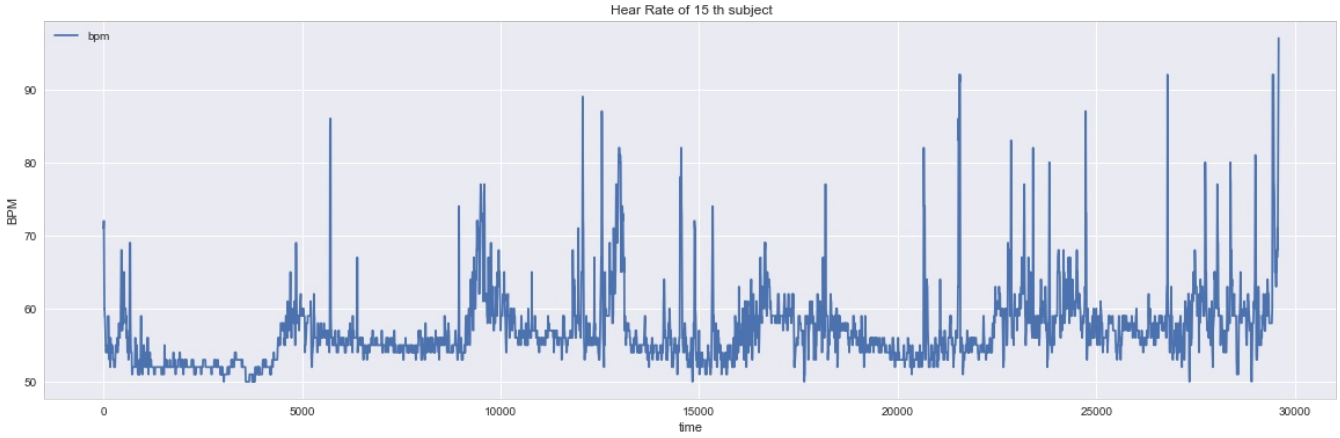

Hear Rate of 10 th subject
bpm
BPM
time
110
100
90
80
70
60
50
0
5000
10000
15000
20000
25000
30000

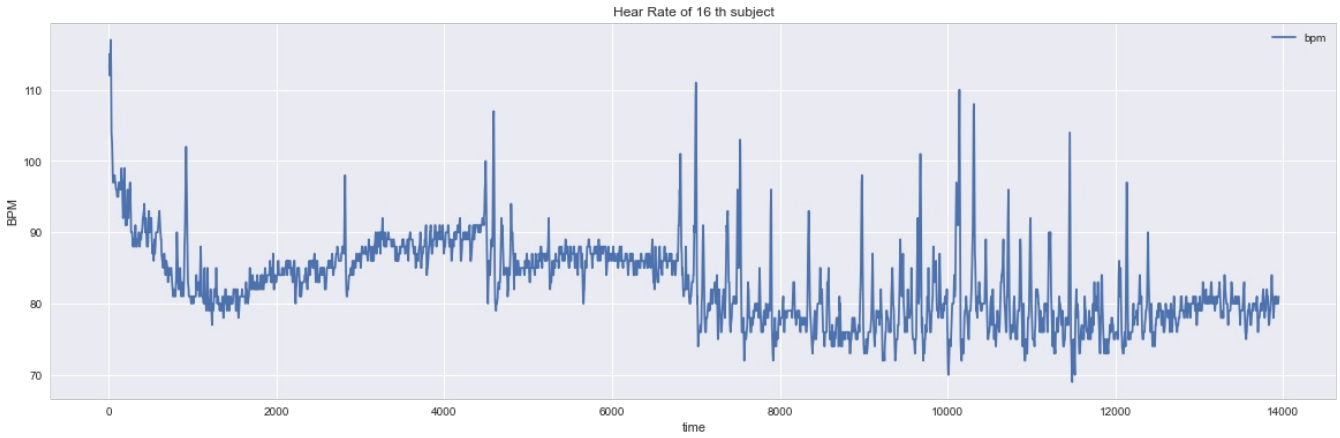

Hear Rate of 11 th subject
bpm
BPM
time
110
100
90
80
70
0
2000
4000
6000
8000
10000
12000

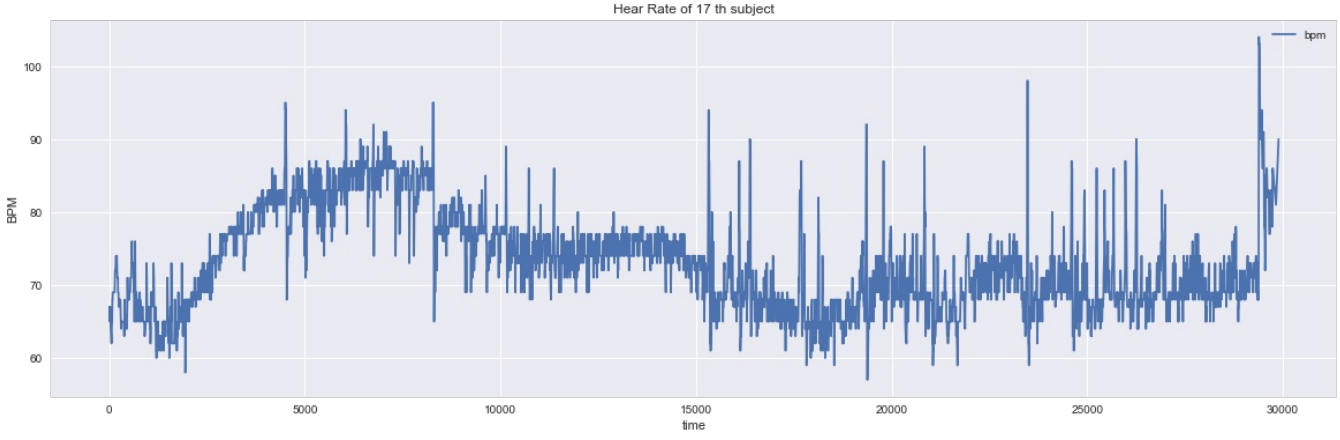

Hear Rate of 12 th subject
bpm
BPM
time
85
80
75
70
65
60
55
50
0
5000
10000
15000
20000
25000
30000

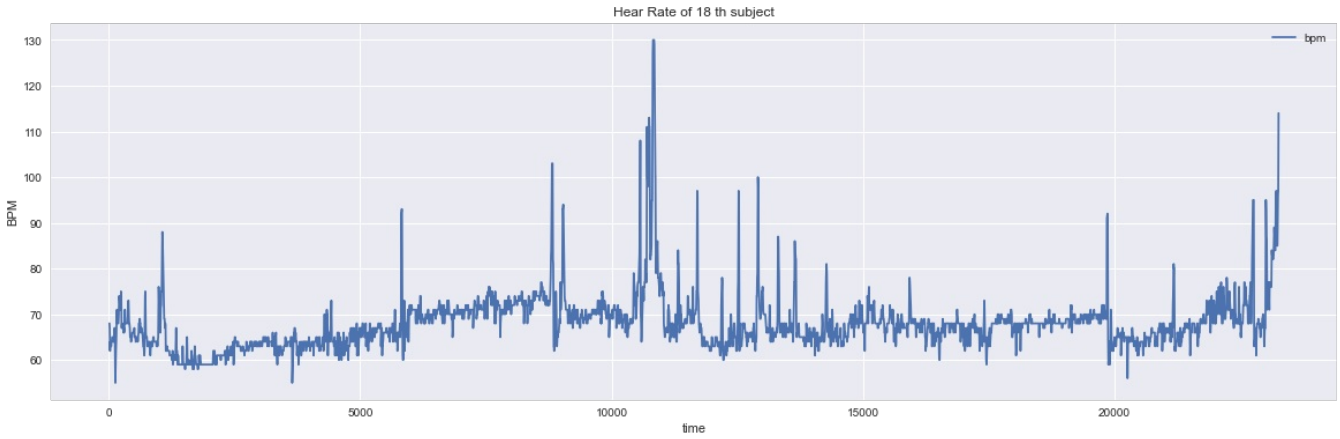

Hear Rate of 13 th subject
bpm
BPM
time
95
90
85
80
75
70
65
60
55
0
2000
4000
6000
8000
10000
12000
14000
16000

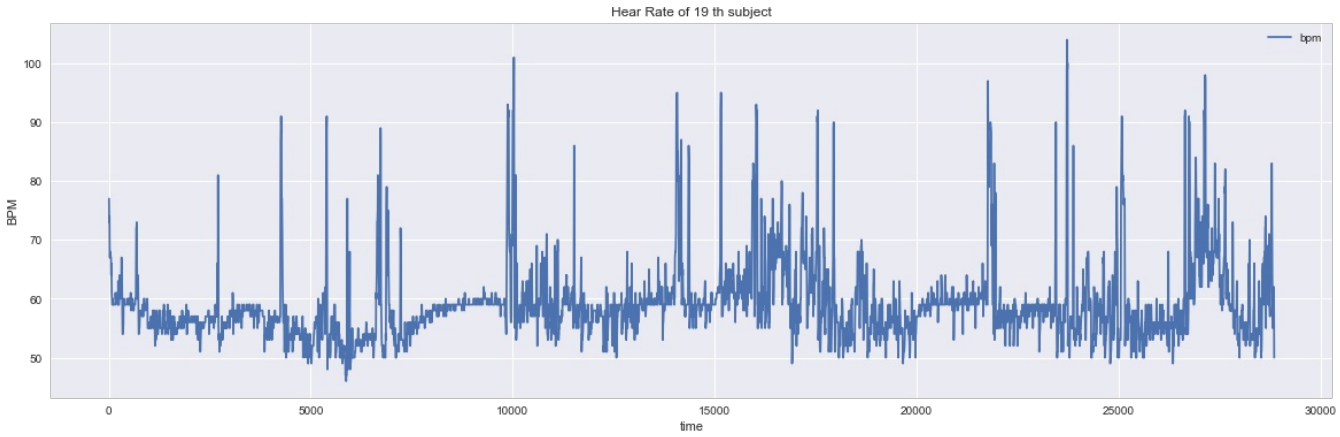

Hear Rate of 14 th subject
bpm
BPM
time
90
80
70
60
50
0
5000
10000
15000
20000
25000
30000

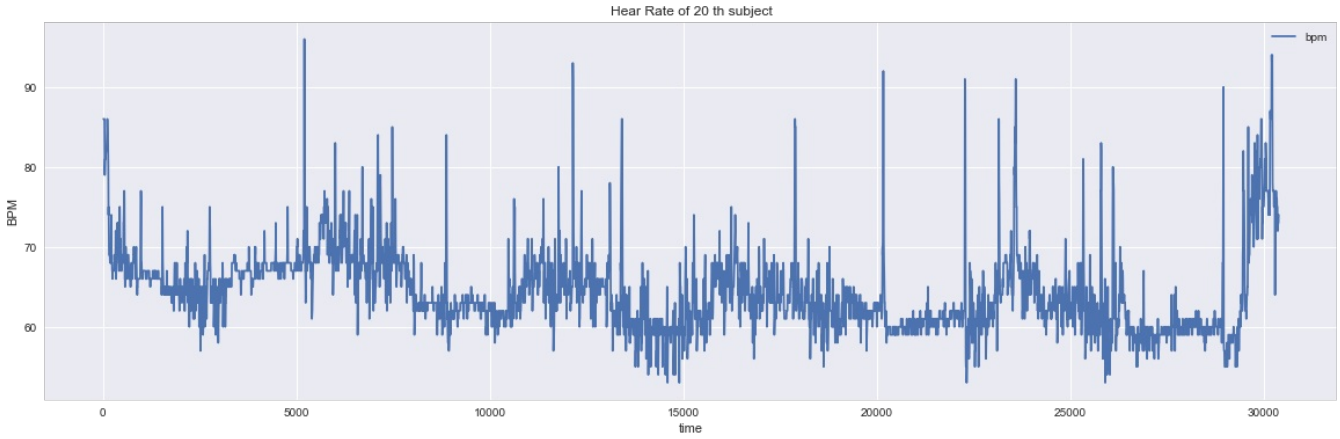

Hear Rate of 15 th subject
bpm
BPM
time
90
80
70
60
50
0
5000
10000
15000
20000
25000
30000

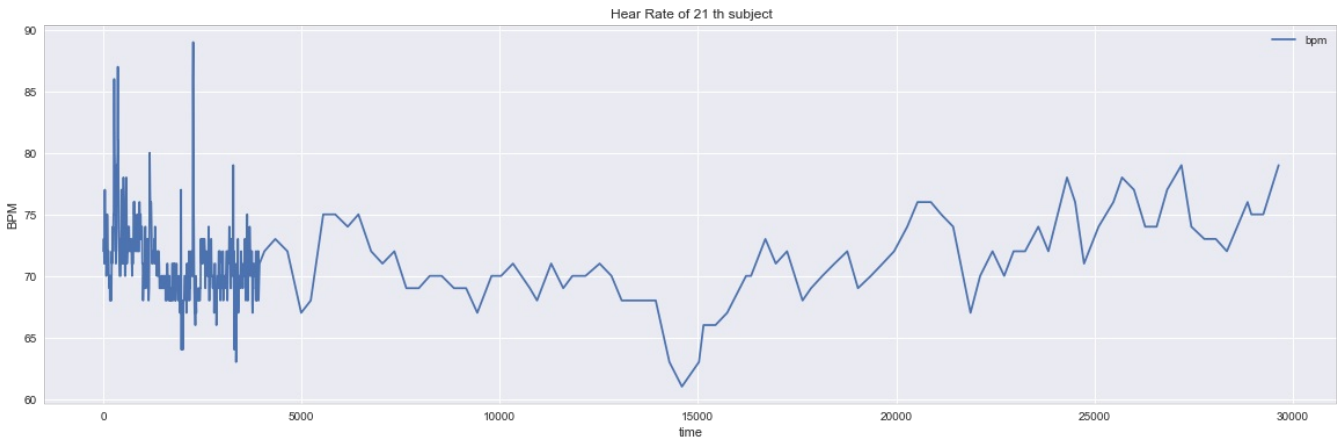

Hear Rate of 16 th subject
bpm
BPM
time
110
100
90
80
70
0
2000
4000
6000
8000
10000
12000
14000

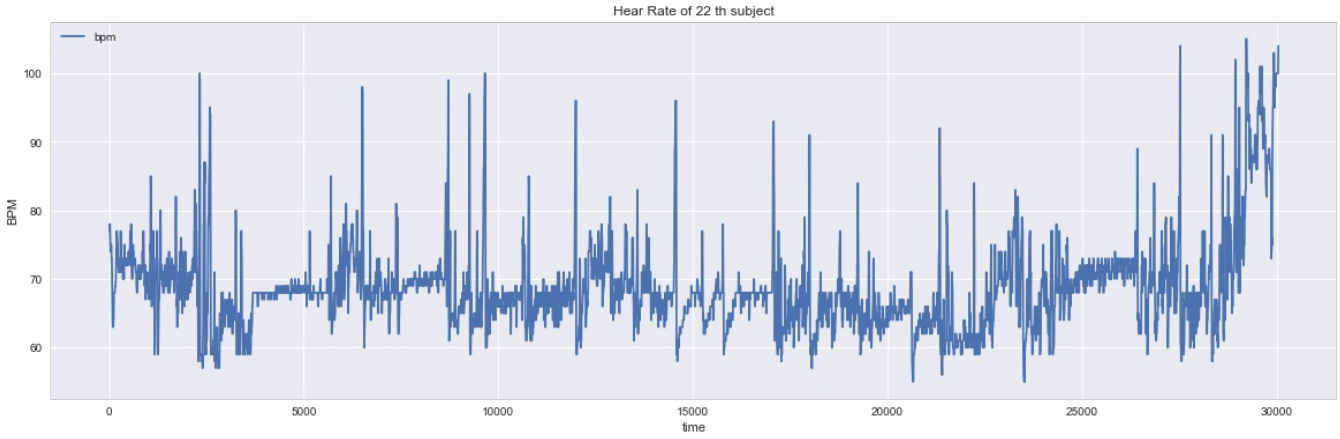

Hear Rate of 17 th subject
bpm
BPM
time
100
90
80
70
60
0
5000
10000
15000
20000
25000
30000

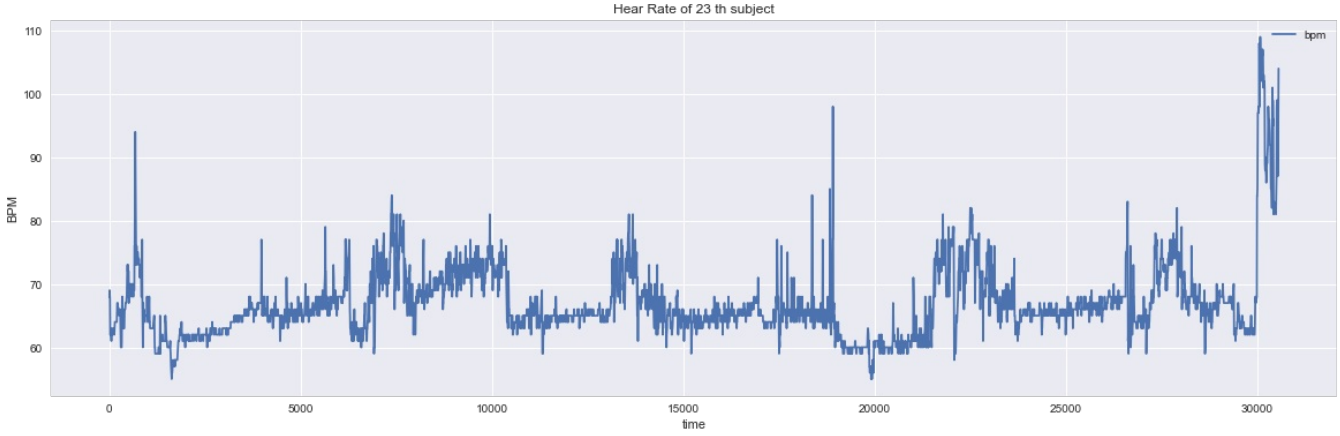

Hear Rate of 18 th subject
bpm
BPM
time
130
120
110
100
90
80
70
60
0
5000
10000
15000
20000

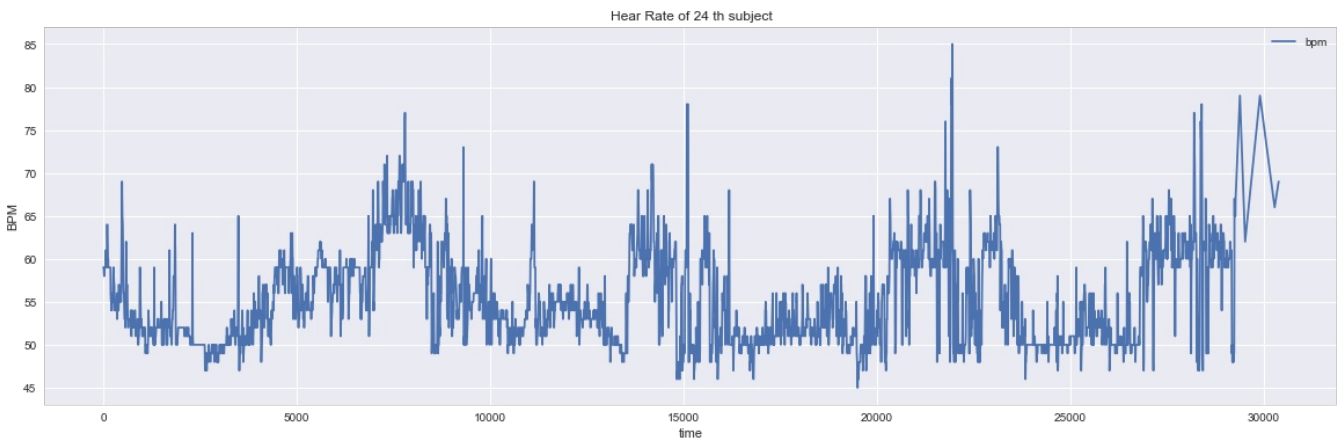

Hear Rate of 19 th subject
bpm
BPM
time
100
90
80
70
60
50
0
5000
10000
15000
20000
25000
30000

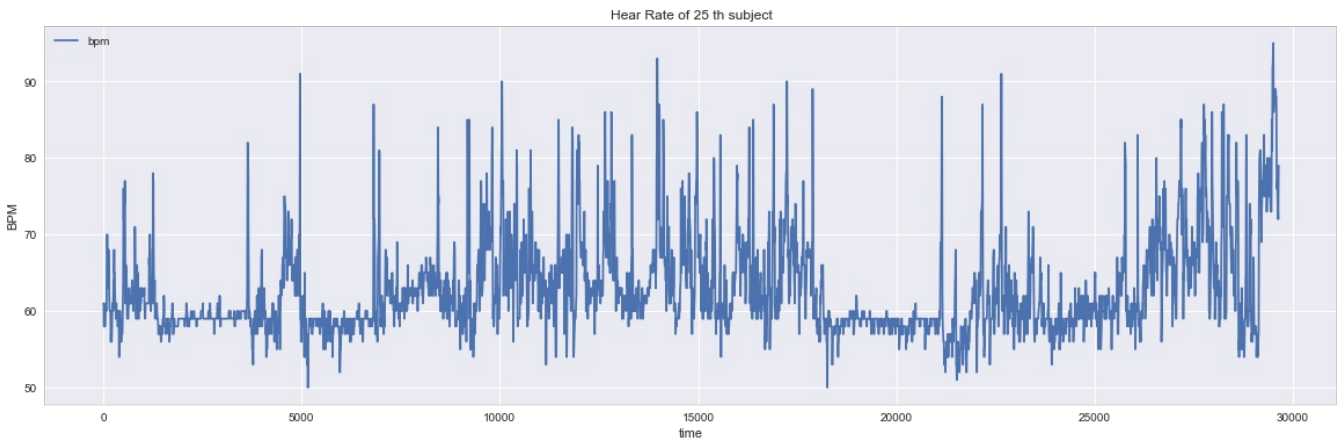

Hear Rate of 20 th subject
bpm
BPM
time
90
80
70
60
0
5000
10000
15000
20000
25000
30000

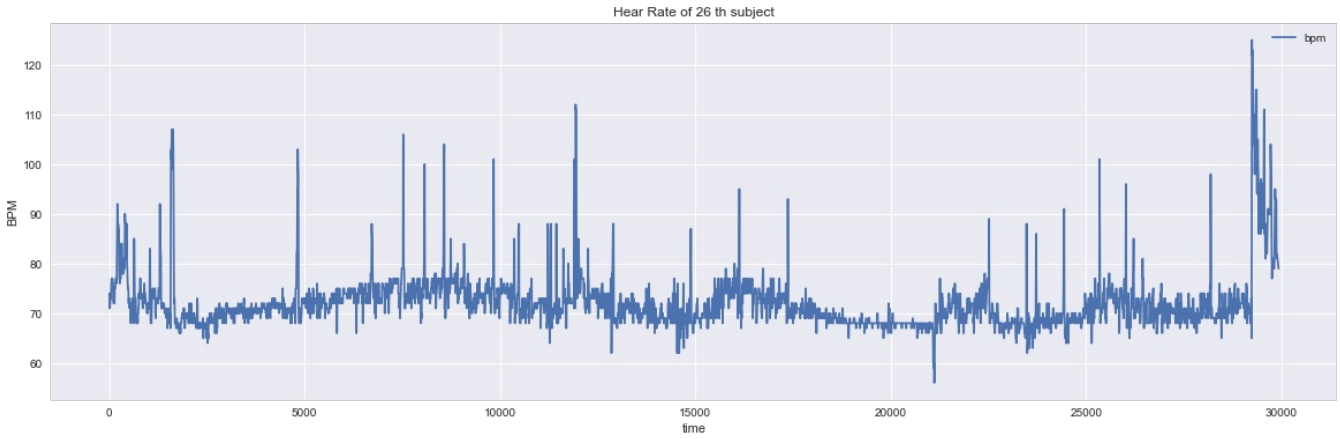

Hear Rate of 21 th subject
bpm
BPM
time
90
85
80
75
70
65
60
0
5000
10000
15000
20000
25000
30000

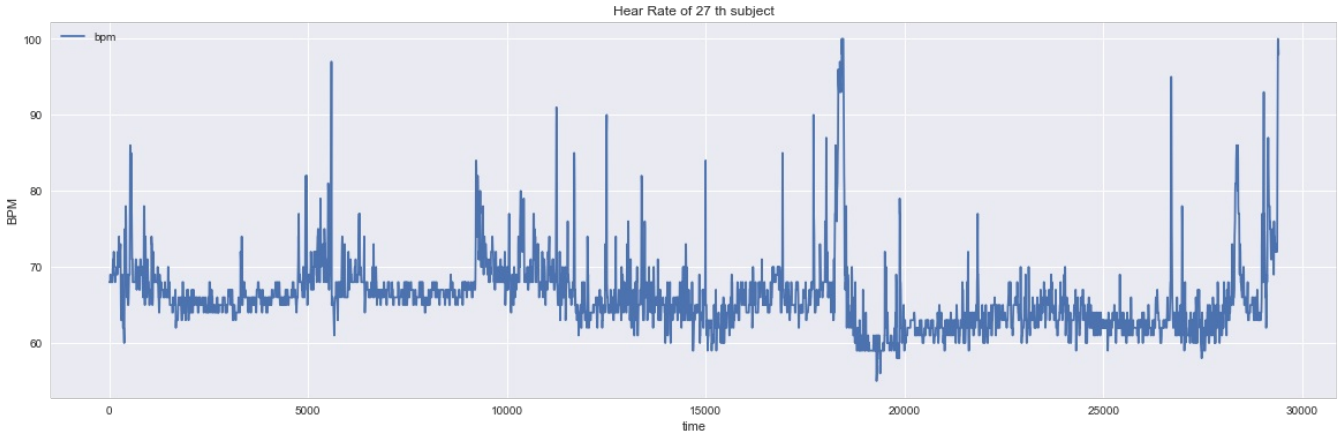

Hear Rate of 22 th subject
bpm
BPM
time

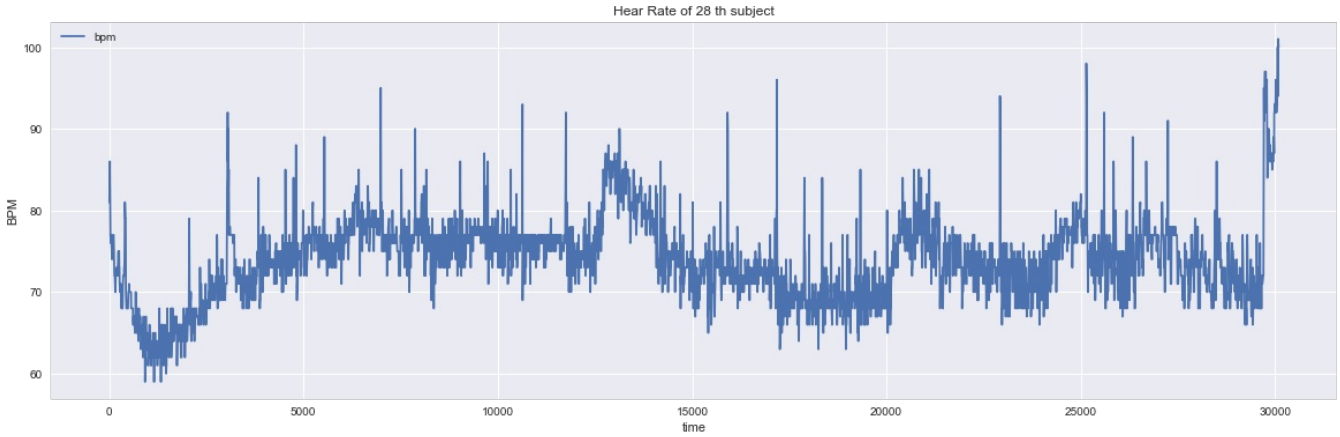

Hear Rate of 23 th subject
bpm
BPM
time

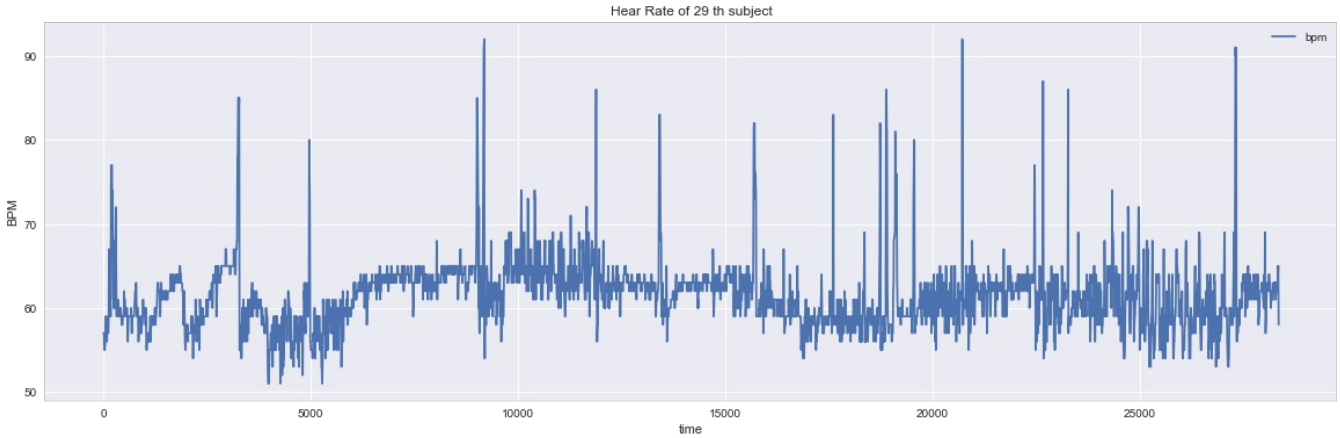

Hear Rate of 24 th subject
bpm
BPM
time

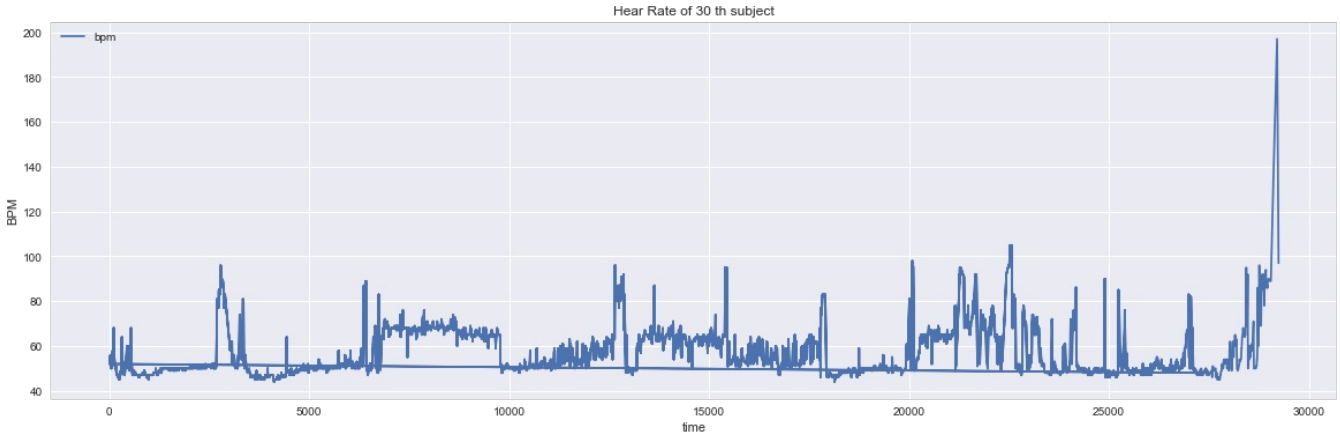

Hear Rate of 25 th subject
bpm
BPM
time


Hear Rate of 26 th subject
bpm
BPM
time


Hear Rate of 27 th subject
bpm
BPM
time

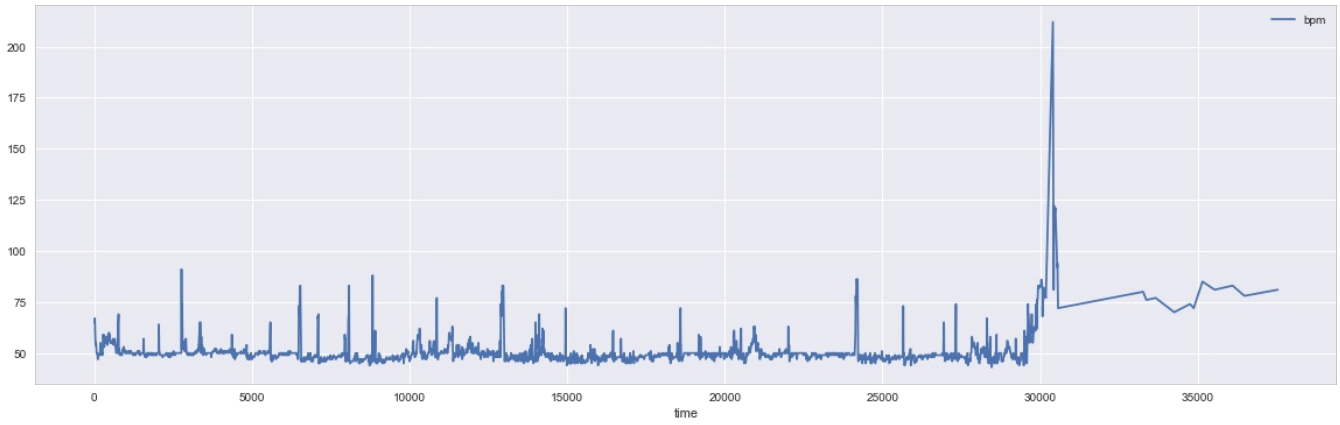

Hear Rate of 28 th subject
bpm
BPM
time

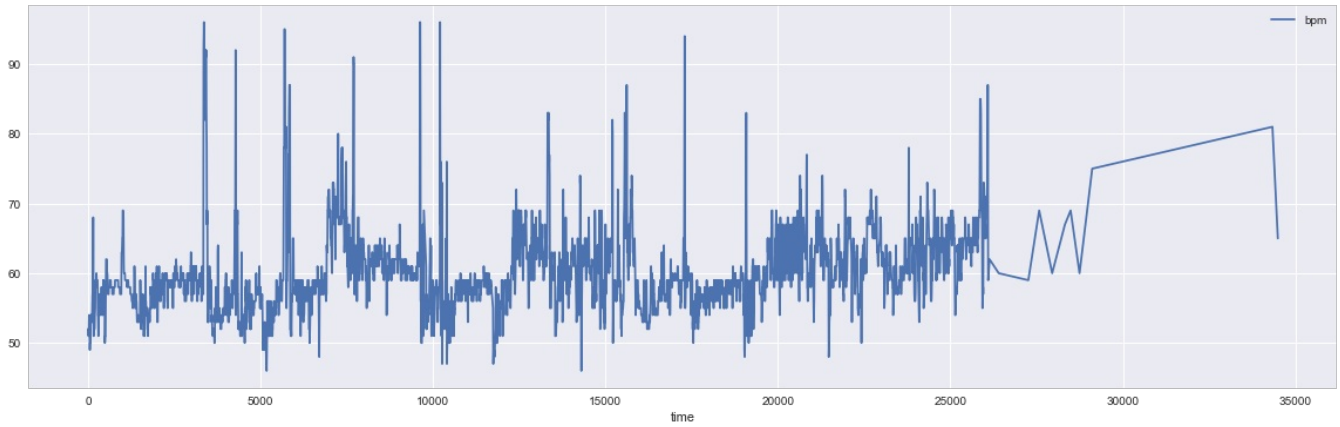

Hear Rate of 29 th subject
bpm
BPM
time

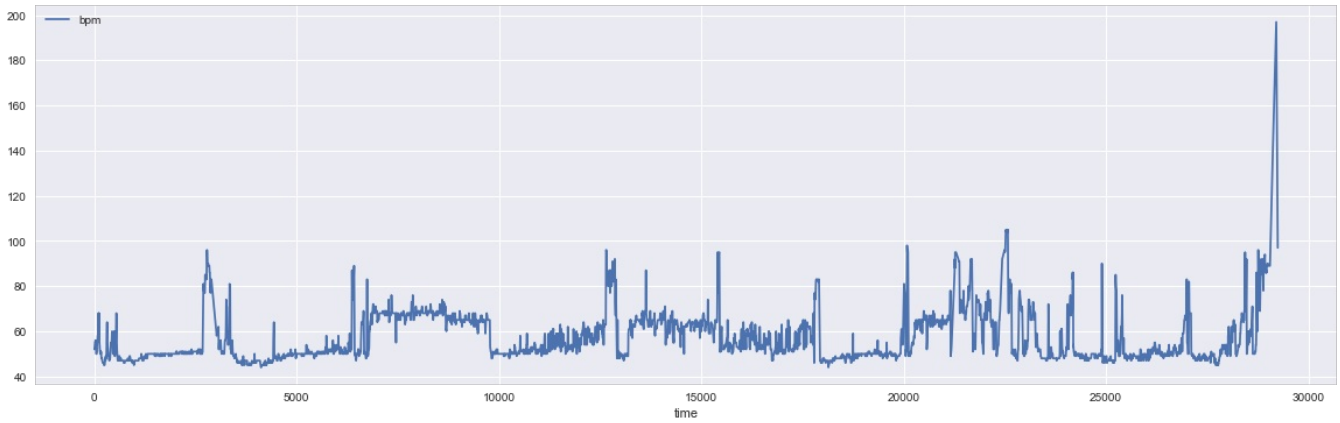


In [6]:

```
# 2, 4, 30번 피험자는 값이 중복해서 나오므로 정리 필요 ( 두 번 나옴 )
df_list = [heart_rate[2], heart_rate[4], heart_rate[30]]

for df in df_list:
    # 데이터프레임의 인덱스와 Time 열을 순회하면서 감소하는 지점을 찾음
    prev_val = -999999
    time_idx = []

    for index, row in df.iterrows():
        if row['time'] < prev_val:
            prev_val = row['time']
            time_idx.append(index)

        prev_val = row['time']

    df2 = df.iloc[max(time_idx):, :]
    f, ax = plt.subplots(1, 1, figsize=(20, 6))
    df2[df2['time'] >=0].plot(x='time', y='bpm', kind='line', ax = ax)
    plt.show()
```





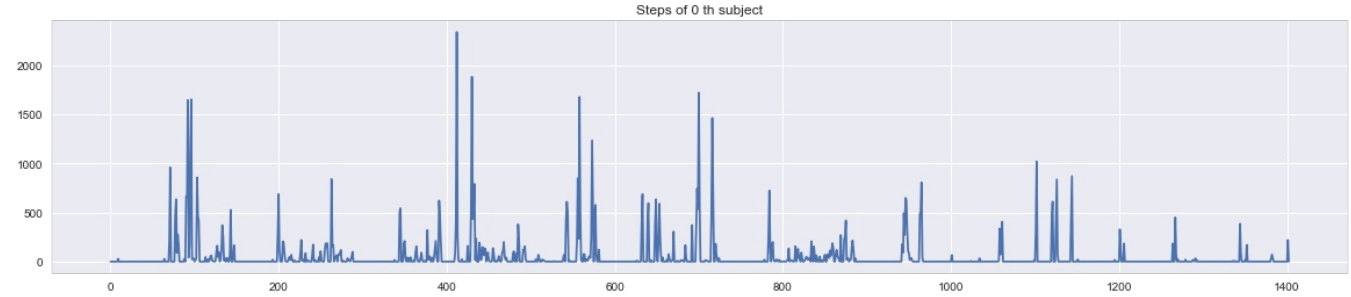


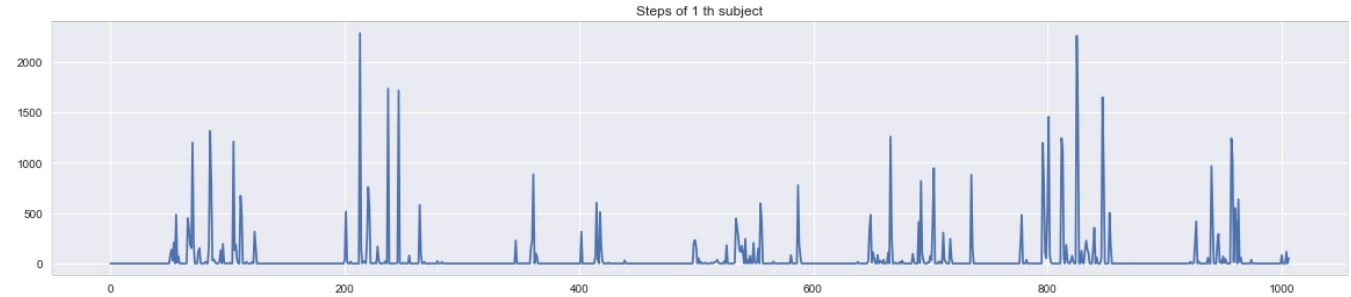
Steps EDA

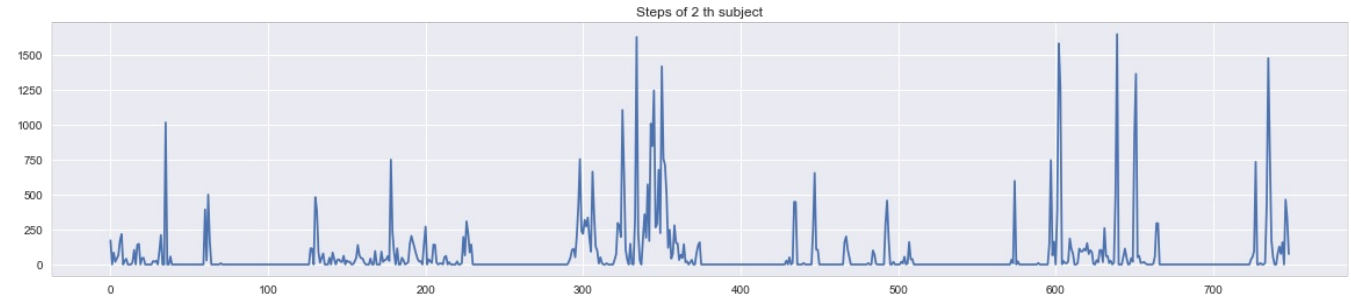
수면 이전의 데이터이므로 수면 분류를 할 때 사용하지 않을 것임

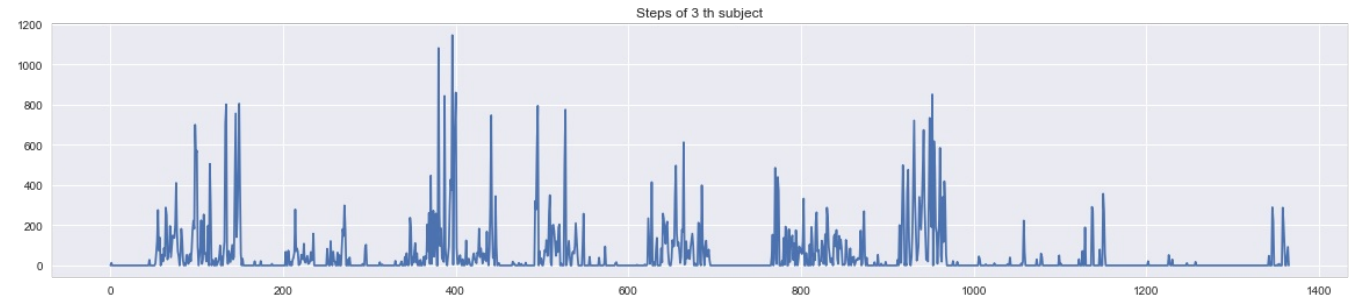
In [9]:
```
# steps 데이터 프레임 리스트
steps = []

dir_name = ''
path = ''
file_list = os.listdir(path)

#  주어진 파일 이름에서 숫자 부분을 추출
# 파일 이름에서 정수 값을 추출하여 파일을 숫자순으로 정렬하는 데 사용
def get_number(filename):
    return int(re.search(r'\d+', filename).group())

# 리스트 파일 이름 순으로 정렬
file_list_txt = sorted([file for file in file_list], key=get_number)

# 텍스트 파일을 csv 파일로 저장
for i in file_list_txt:
    df = pd.read_csv(dir_name+i, header=None, names=['time', 'step'])
    steps.append(df)
```

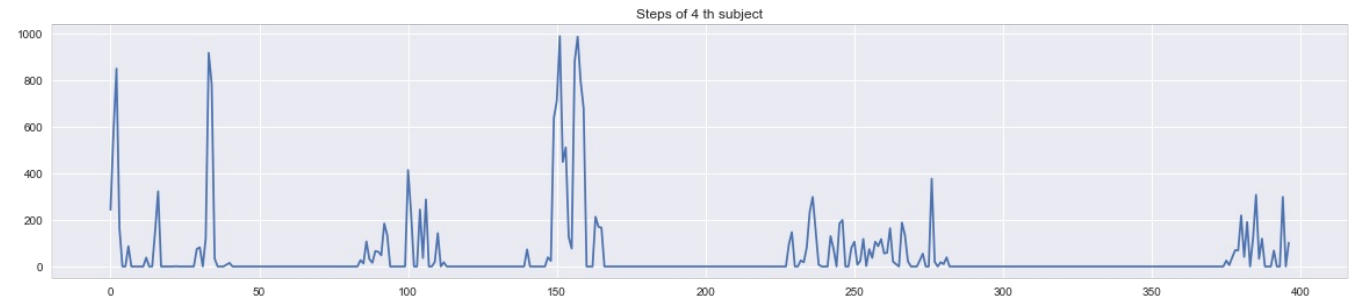
In [12]:
```
# 각 피험자별 steps 출력

for i in range(len(steps)):
    plt.figure(figsize=(20, 4))
    plt.title(f'Steps of {i} th subject')
    steps[i]['step'].plot()
    plt.show()
```

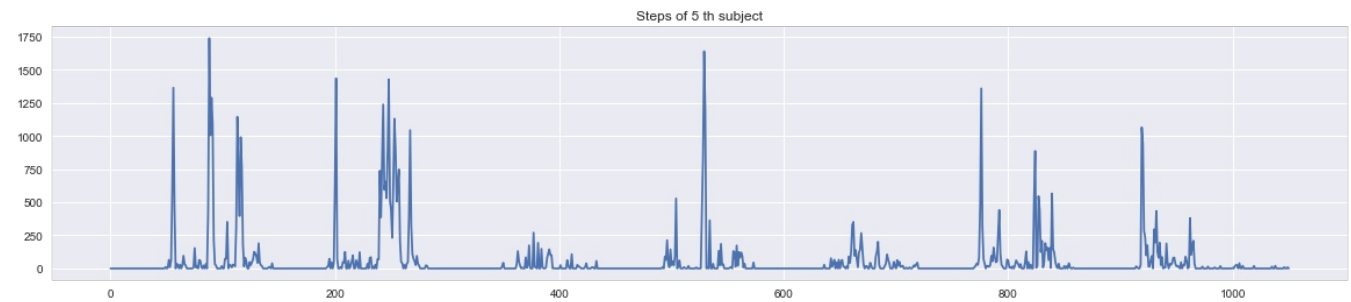

Steps of 0 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000
1200
1400

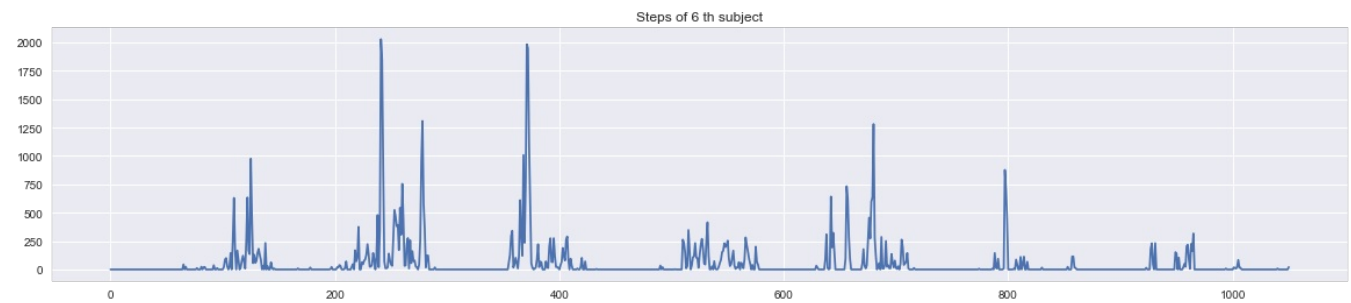

Steps of 1 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000

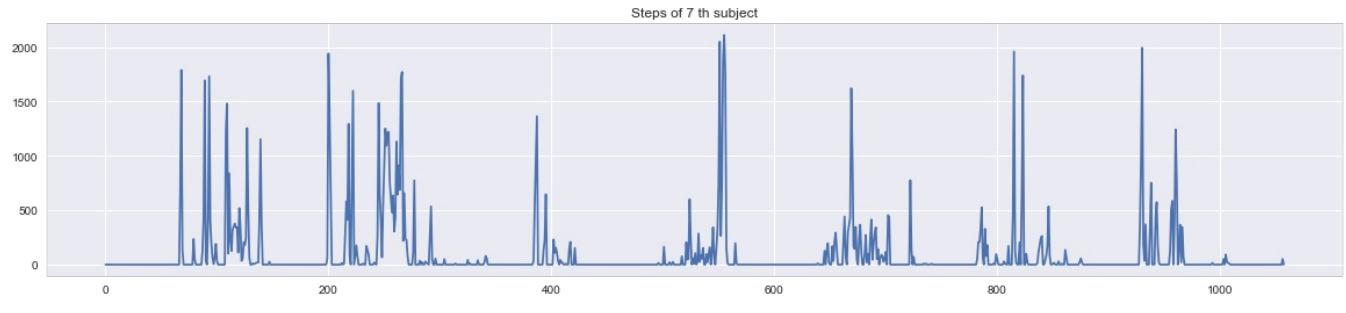

Steps of 2 th subject
1500
1250
1000
750
500
250
0
0
100
200
300
400
500
600
700

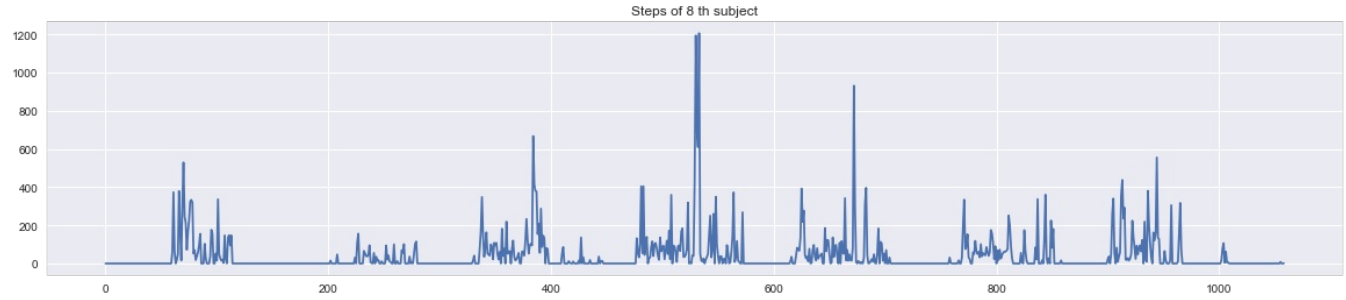

Steps of 3 th subject
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
1200
1400

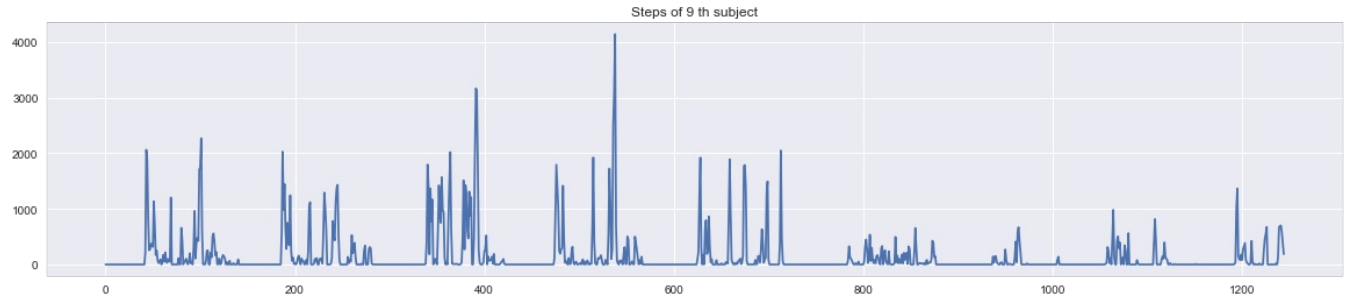

Steps of 4 th subject
1000
800
600
400
200
0
0
50
100
150
200
250
300
350
400

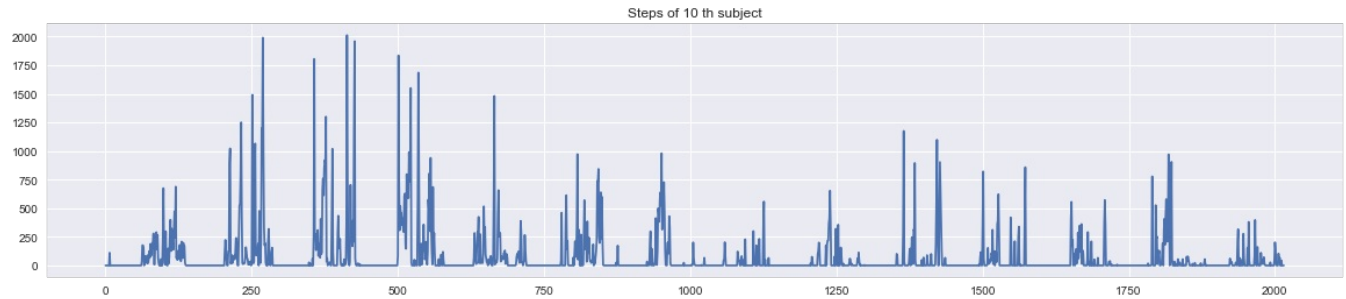

Steps of 5 th subject
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000

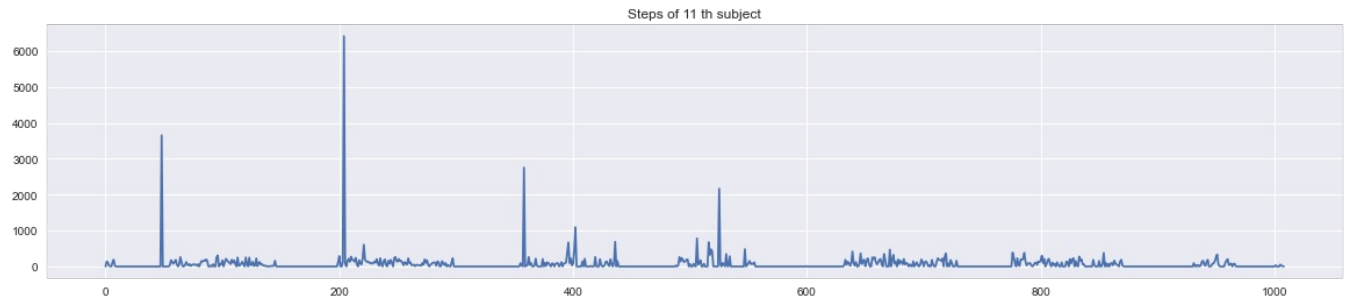

Steps of 6 th subject
2000
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000

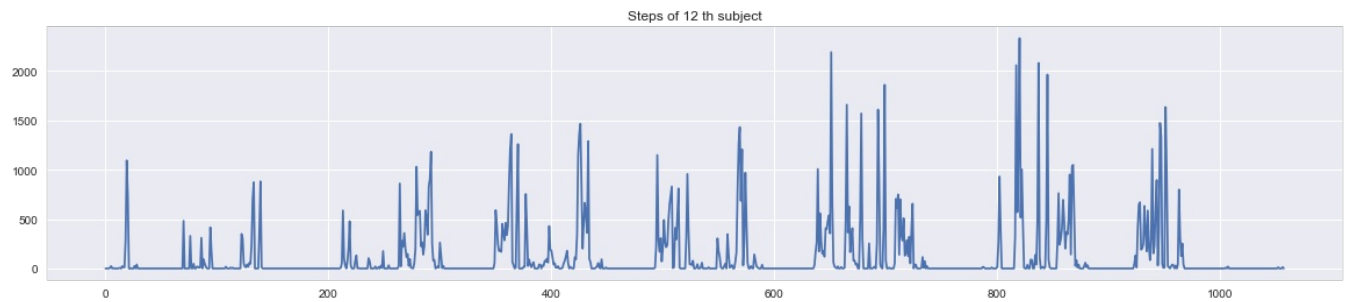

Steps of 7 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000

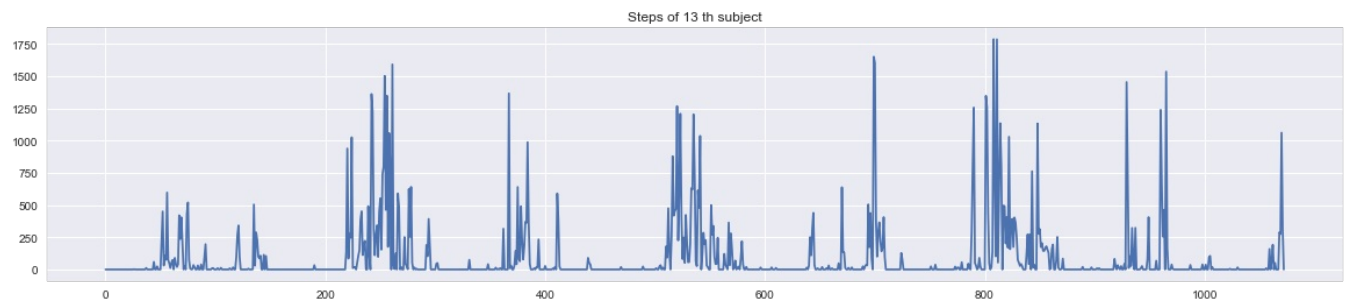

Steps of 8 th subject
1200
1000
800
600
400
200
0
0
200
400
600
800
1000

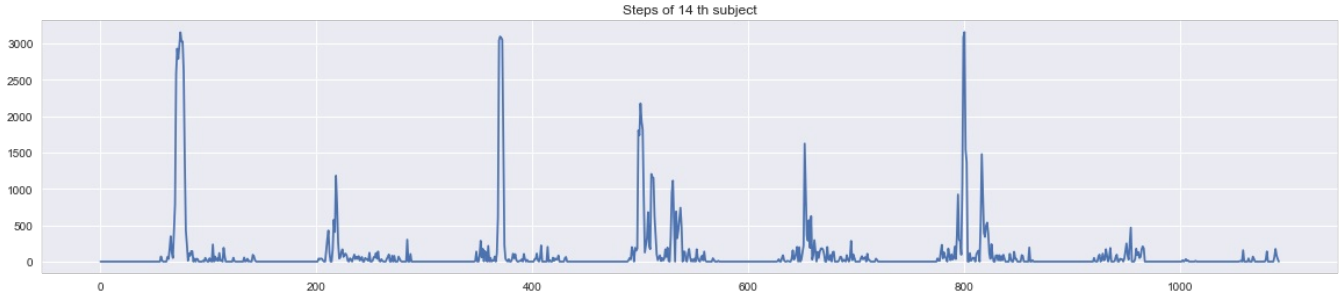

Steps of 9 th subject
4000
3000
2000
1000
0
0
200
400
600
800
1000
1200

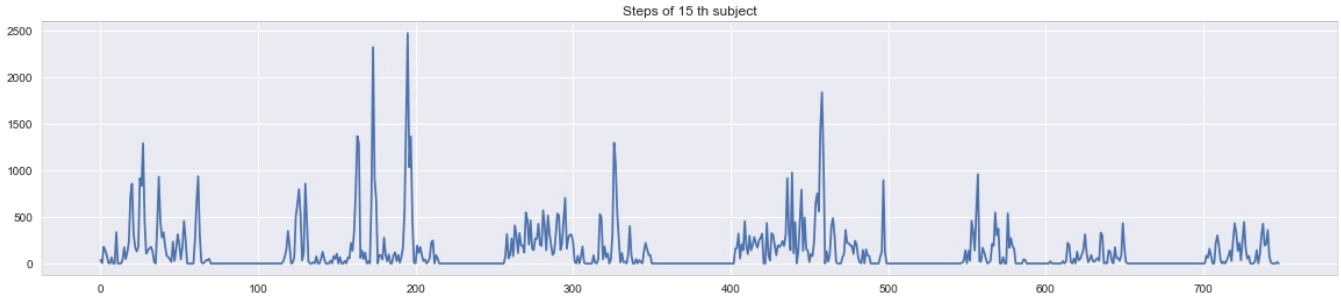

Steps of 10 th subject
2000
1750
1500
1250
1000
750
500
250
0
0
250
500
750
1000
1250
1500
1750
2000

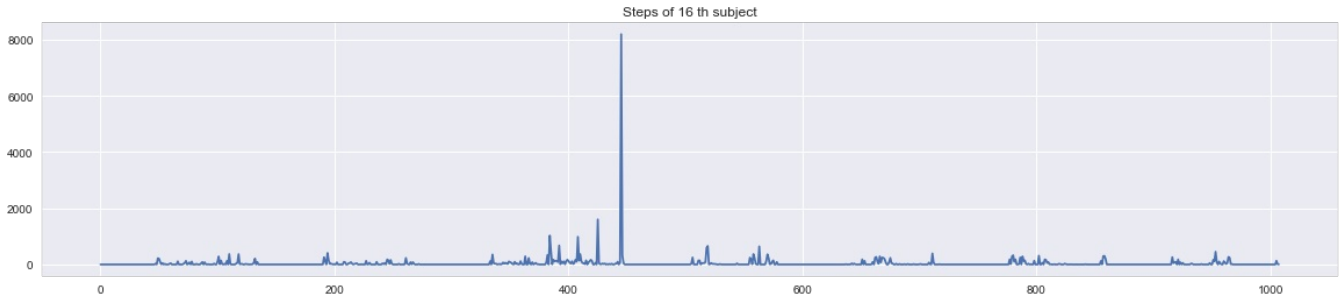

Steps of 11 th subject
6000
5000
4000
3000
2000
1000
0
0
200
400
600
800
1000

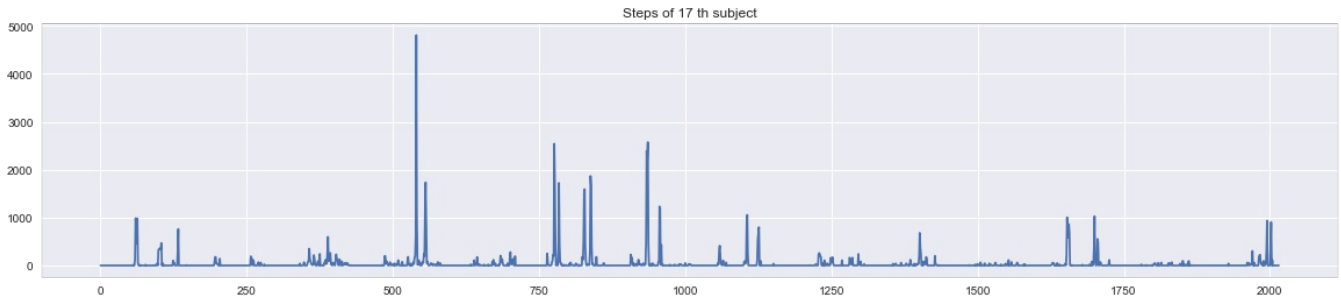

Steps of 12 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000

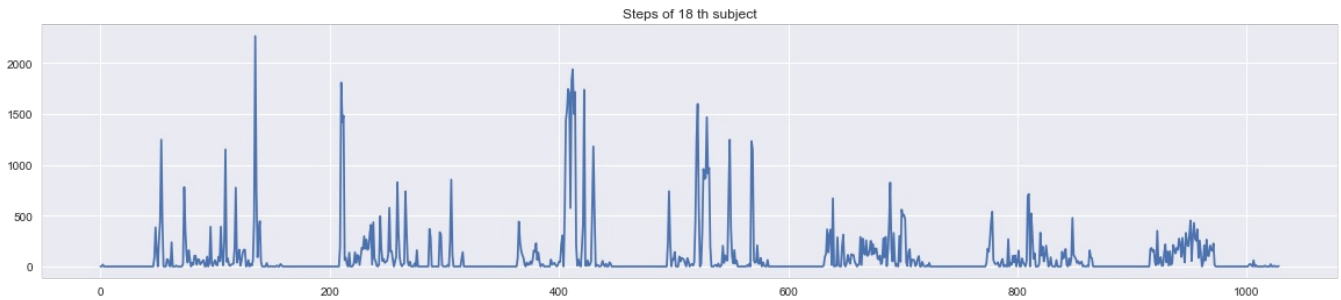

Steps of 13 th subject
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000

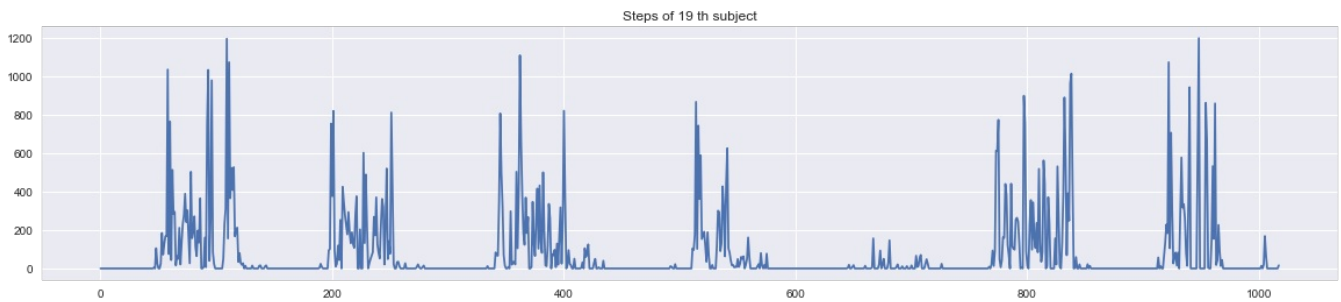

Steps of 14 th subject
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000

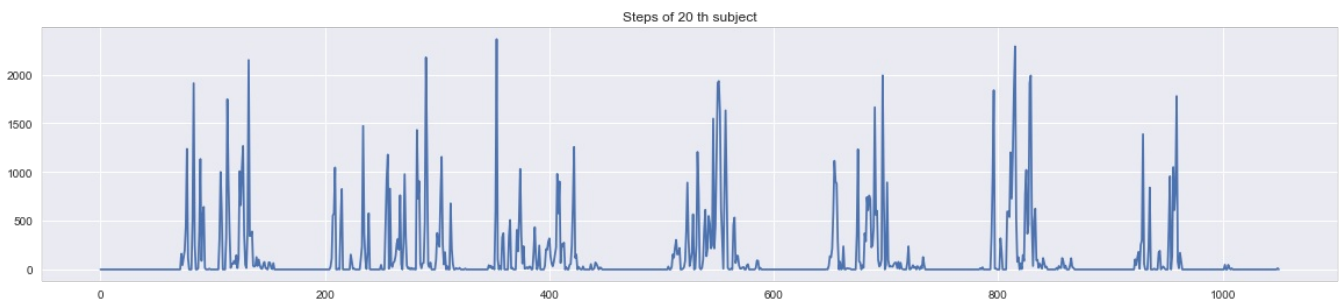

Steps of 15 th subject
2500
2000
1500
1000
500
0
0
100
200
300
400
500
600
700

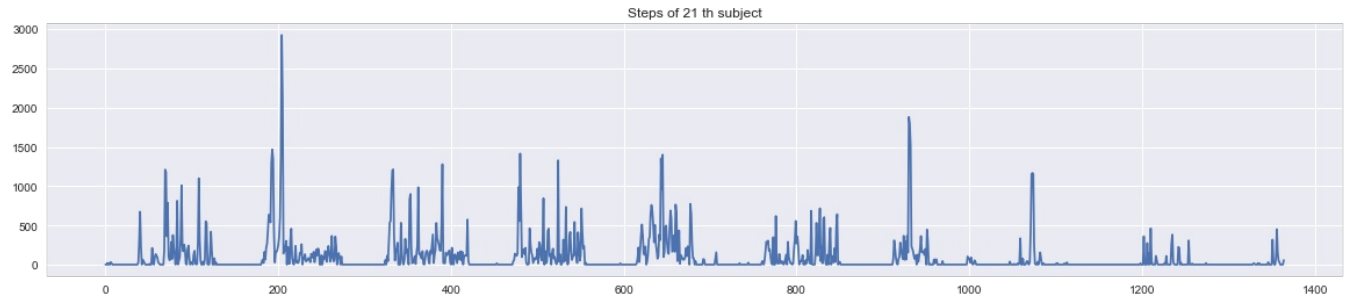

Steps of 16 th subject
8000
6000
4000
2000
0
0
200
400
600
800
1000

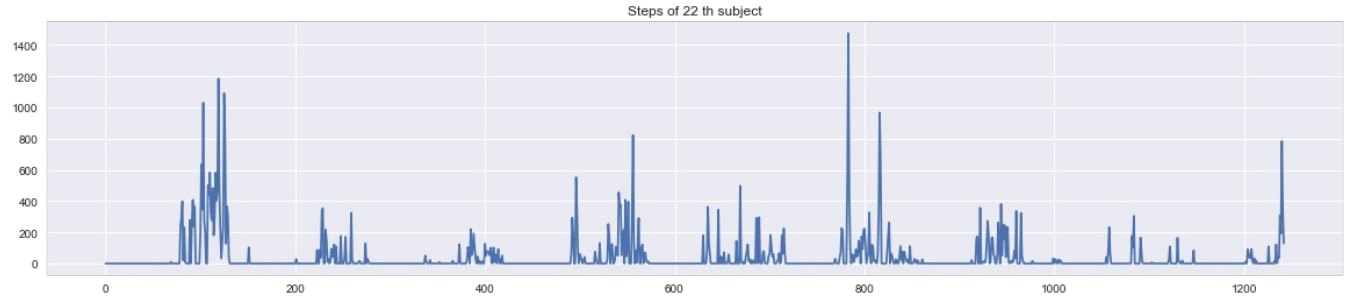

Steps of 17 th subject
5000
4000
3000
2000
1000
0
0
250
500
750
1000
1250
1500
1750
2000

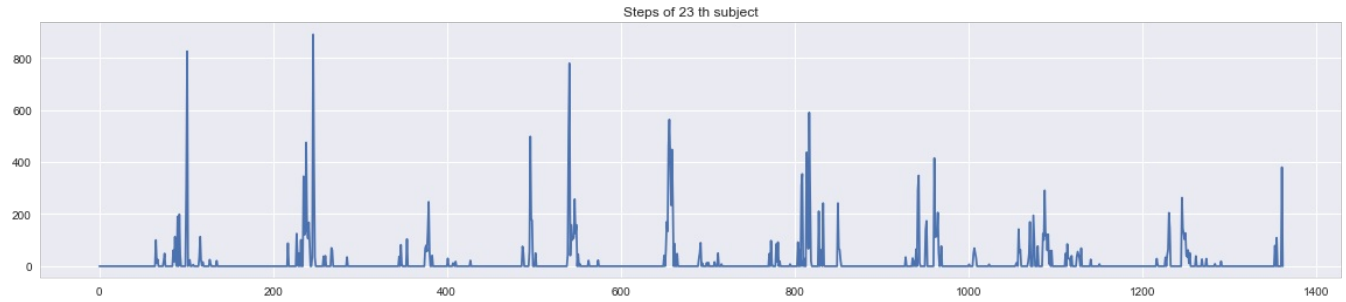

Steps of 18 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000

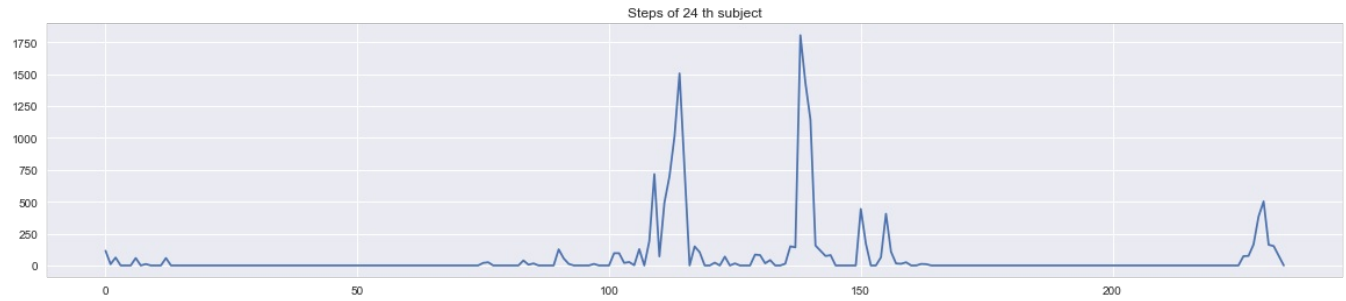

Steps of 19 th subject
1200
1000
800
600
400
200
0
0
200
400
600
800
1000

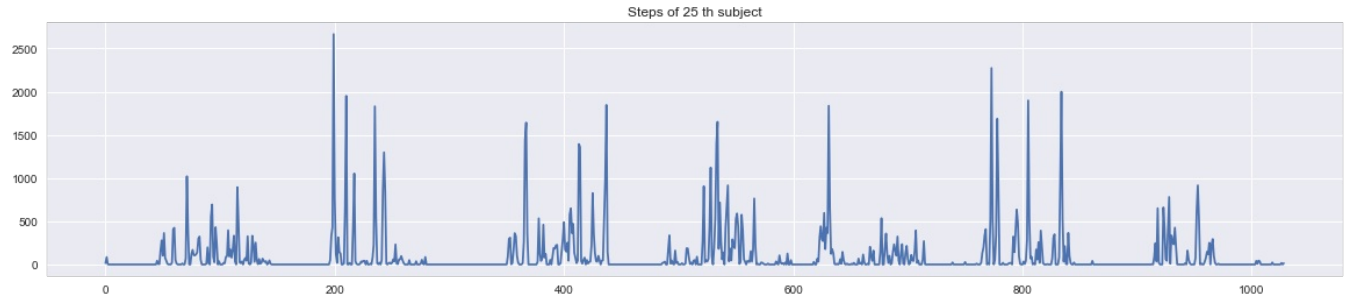

Steps of 20 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000

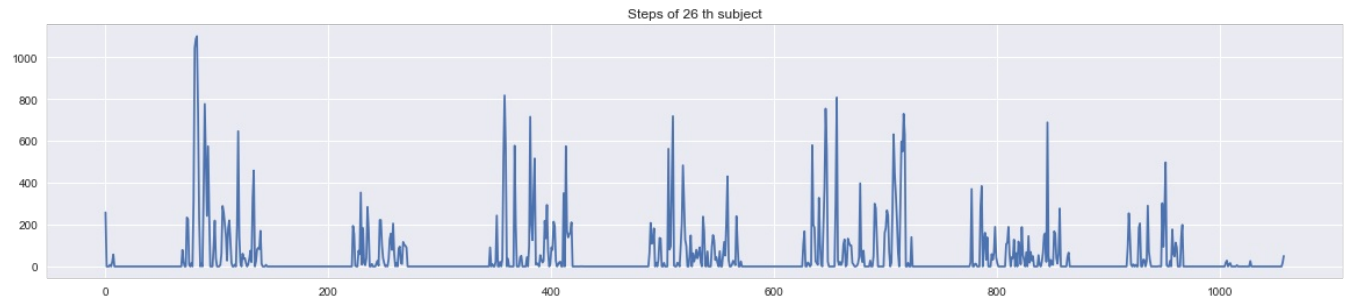

Steps of 21 th subject
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000
1200
1400

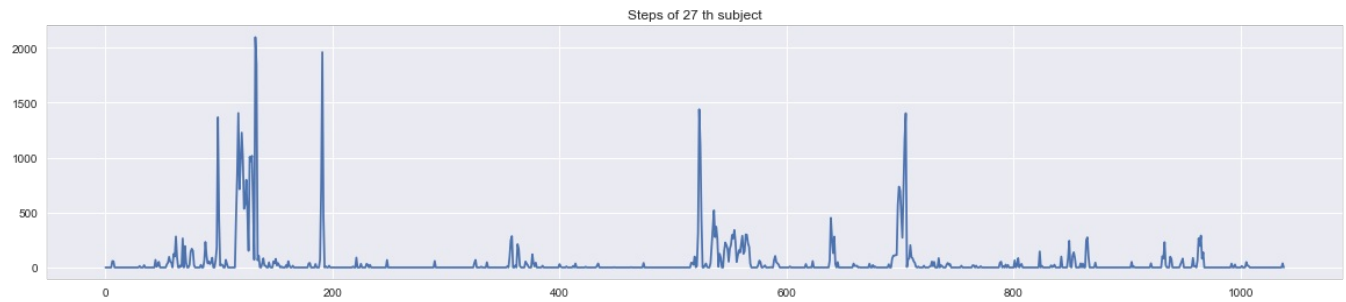

Steps of 22 th subject
1400
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
1200

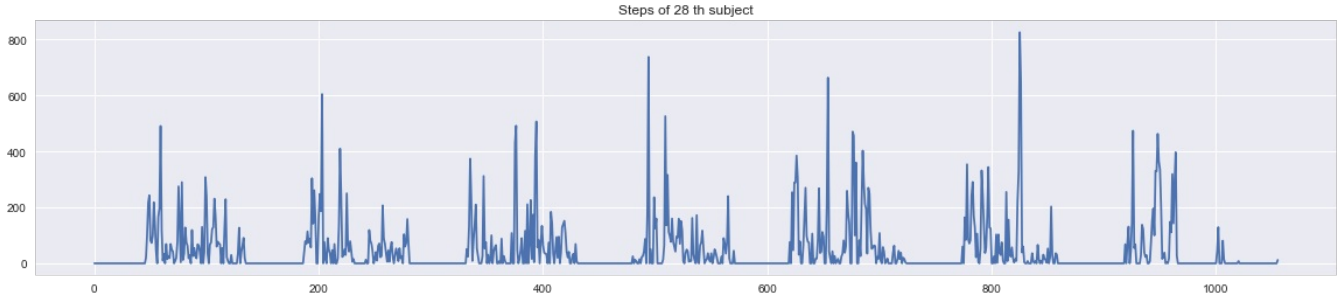

Steps of 23 th subject
800
600
400
200
0
0
200
400
600
800
1000
1200
1400

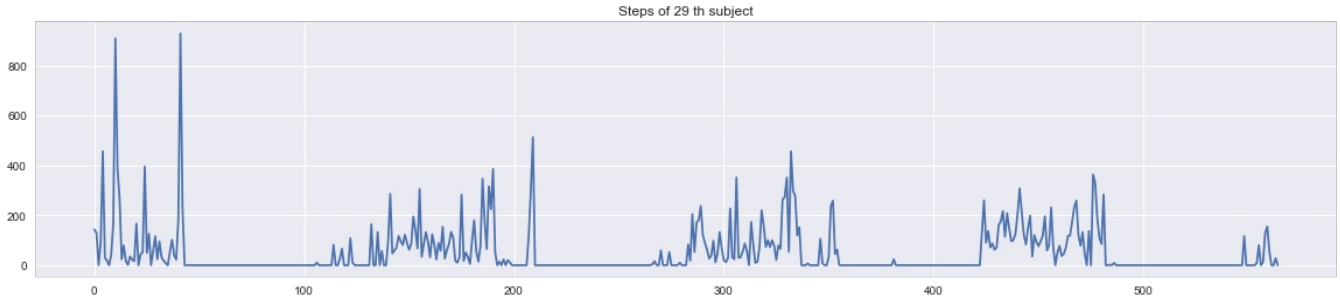

Steps of 24 th subject
1750
1500
1250
1000
750
500
250
0
0
50
100
150
200

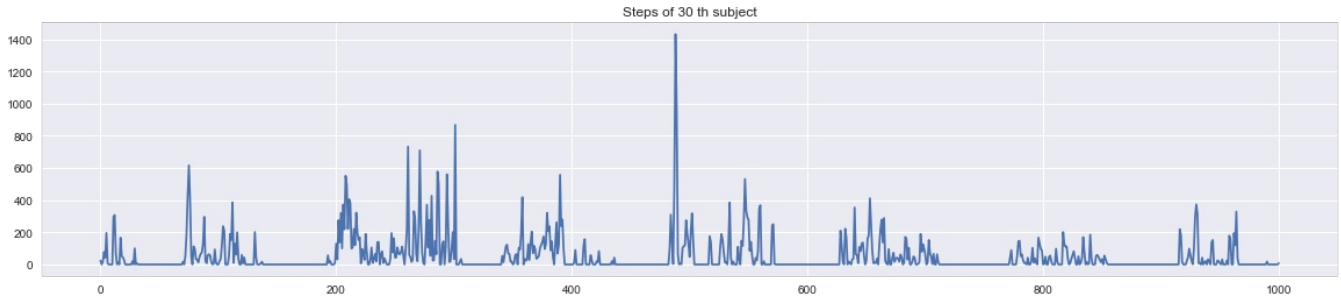

Steps of 25 th subject
2500
2000
1500
1000
500
0
0
200
400
600
800
1000


Steps of 26 th subject
1000
800
600
400
200
0
0
200
400
600
800
1000


Steps of 27 th subject
2000
1500
1000
500
0
0
200
400
600
800
1000



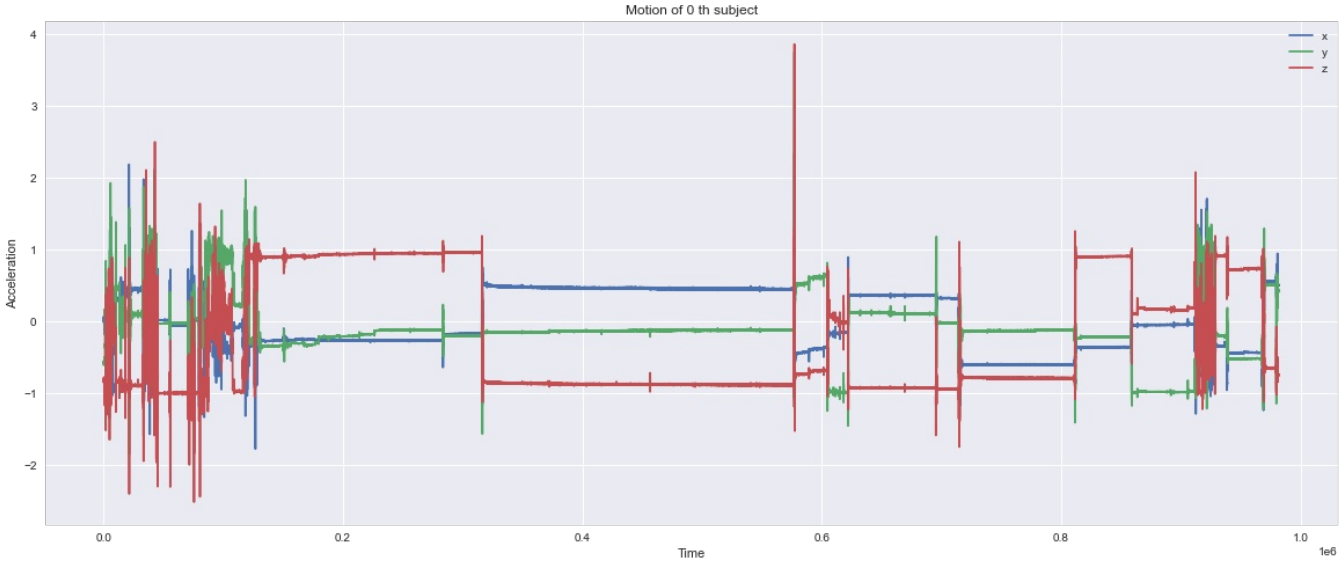


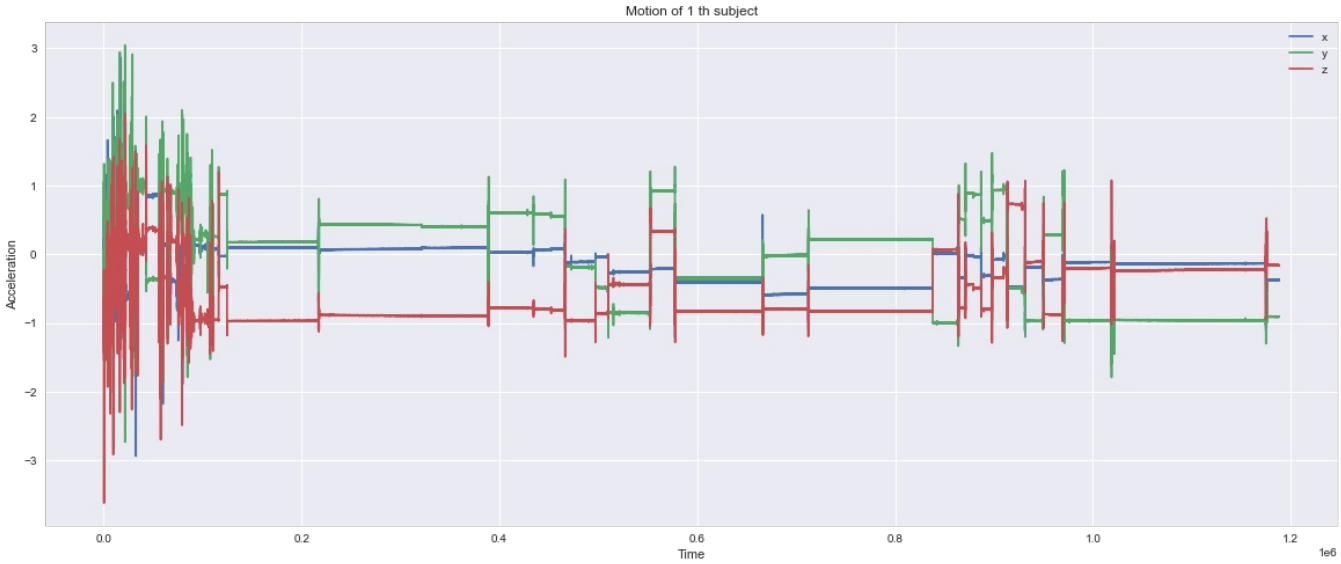


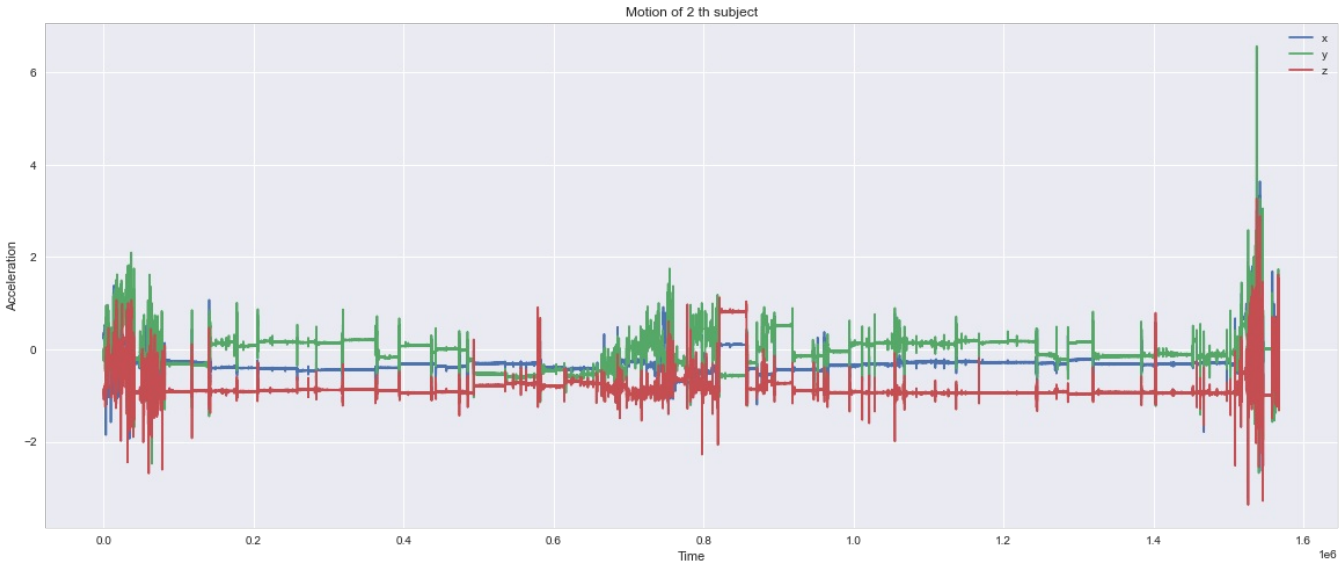
Motion EDA

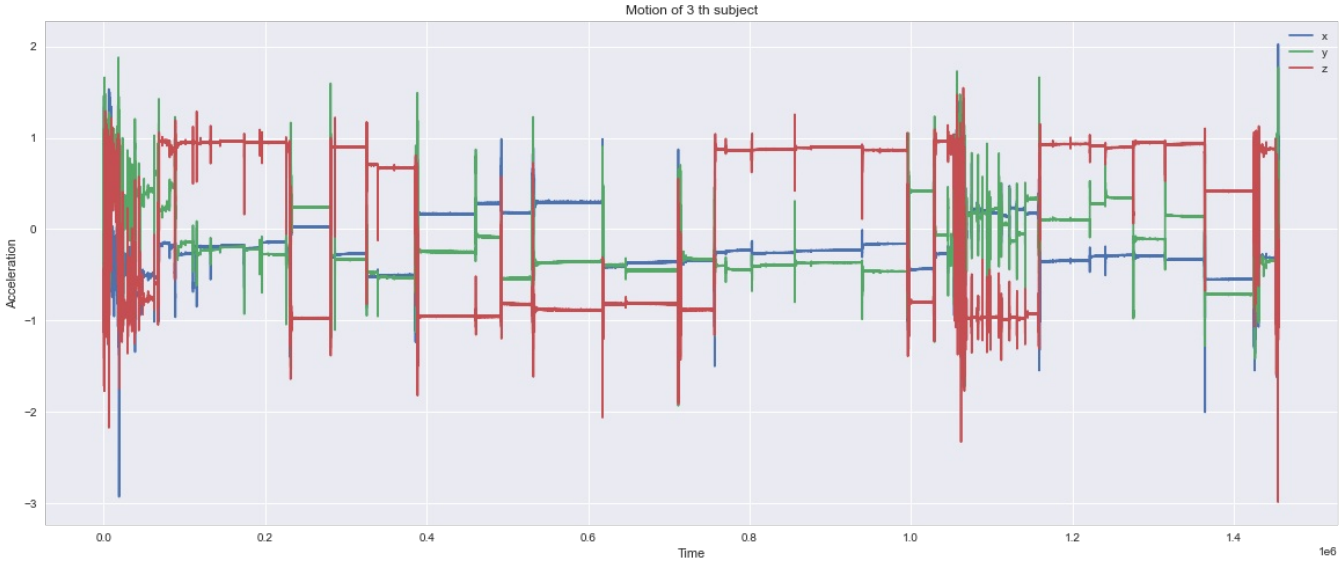
In [14]:
```
# motion 데이터 프레임 리스트
motion = []

dir_name = ''
path = ''
file_list = os.listdir(path)

#  주어진 파일 이름에서 숫자 부분을 추출
# 파일 이름에서 정수 값을 추출하여 파일을 숫자순으로 정렬하는 데 사용
def get_number(filename):
    return int(re.search(r'\d+', filename).group())

# 리스트 파일 이름 순으로 정렬
file_list_txt = sorted([file for file in file_list], key=get_number)

# 텍스트 파일을 csv 파일로 저장
for i in file_list_txt:
    df = pd.read_csv(dir_name+i,  sep=' ', header=None, names=['time', 'x', 'y', 'z'])
    motion.append(df)
```

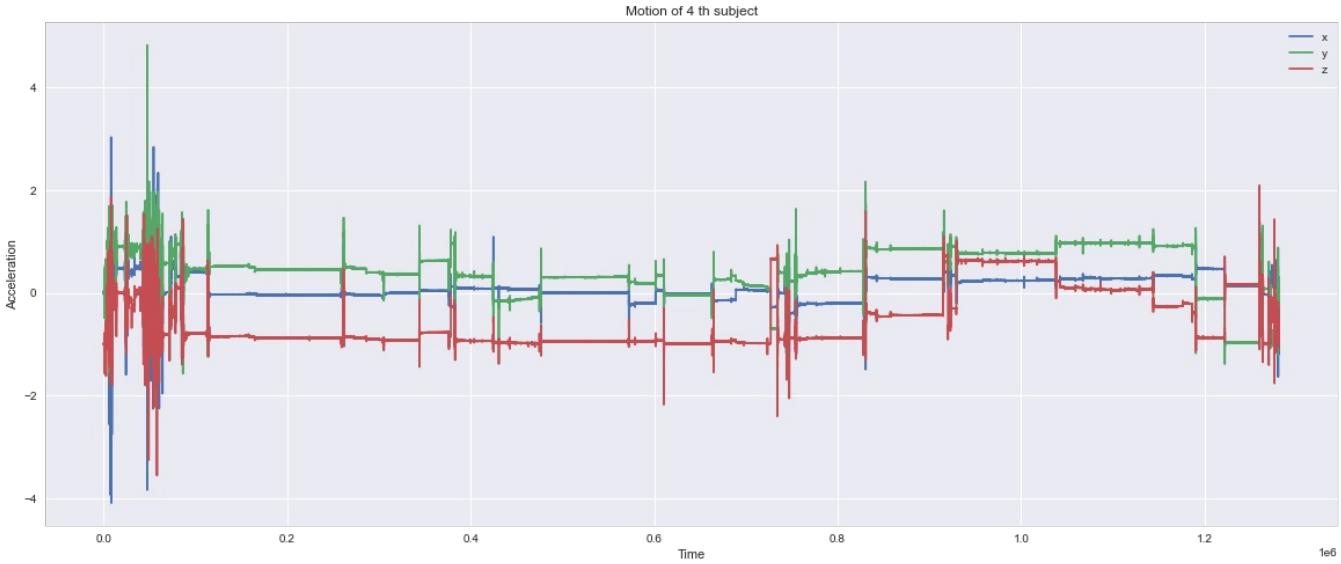
In [15]:
```
# 각 피험자별 motion 출력

for i in range(len(motion)):
    plt.figure(figsize=(20, 8))
    plt.title(f'Motion of {i} th subject')
    plt.xlabel('Time')
    plt.ylabel('Acceleration')
    motion[i]['x'].plot(label='x')
    motion[i]['y'].plot(label='y')
    motion[i]['z'].plot(label='z')

    plt.legend()
    plt.show()
```

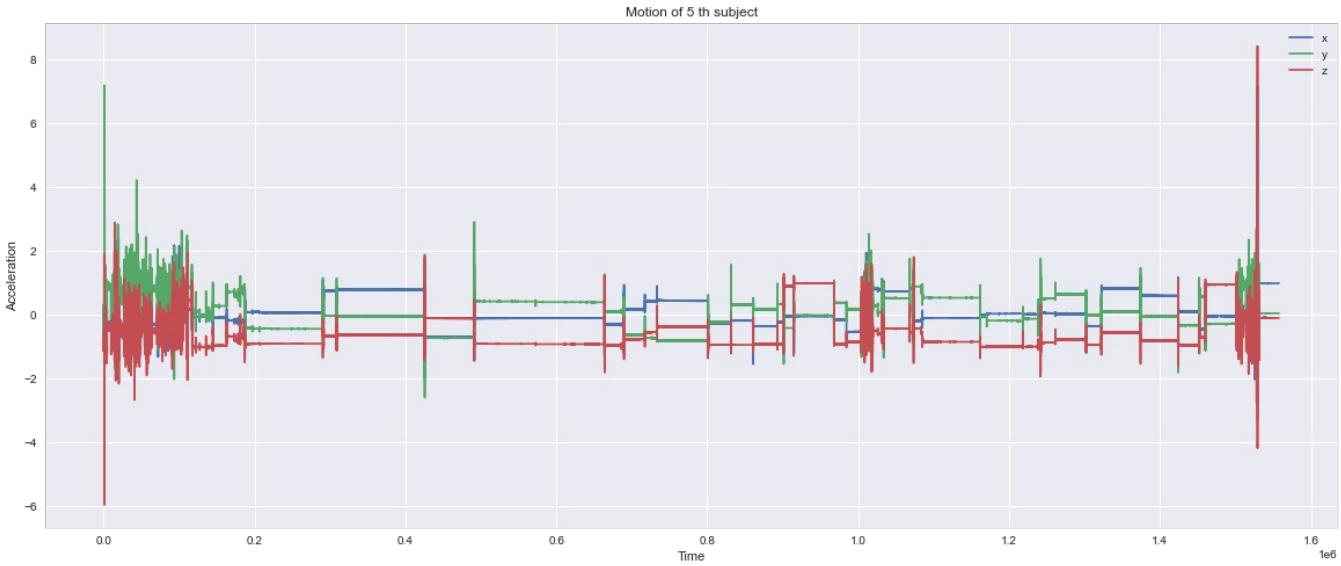

Motion of 0 th subject
x
y
z
Acceleration
Time
1e6

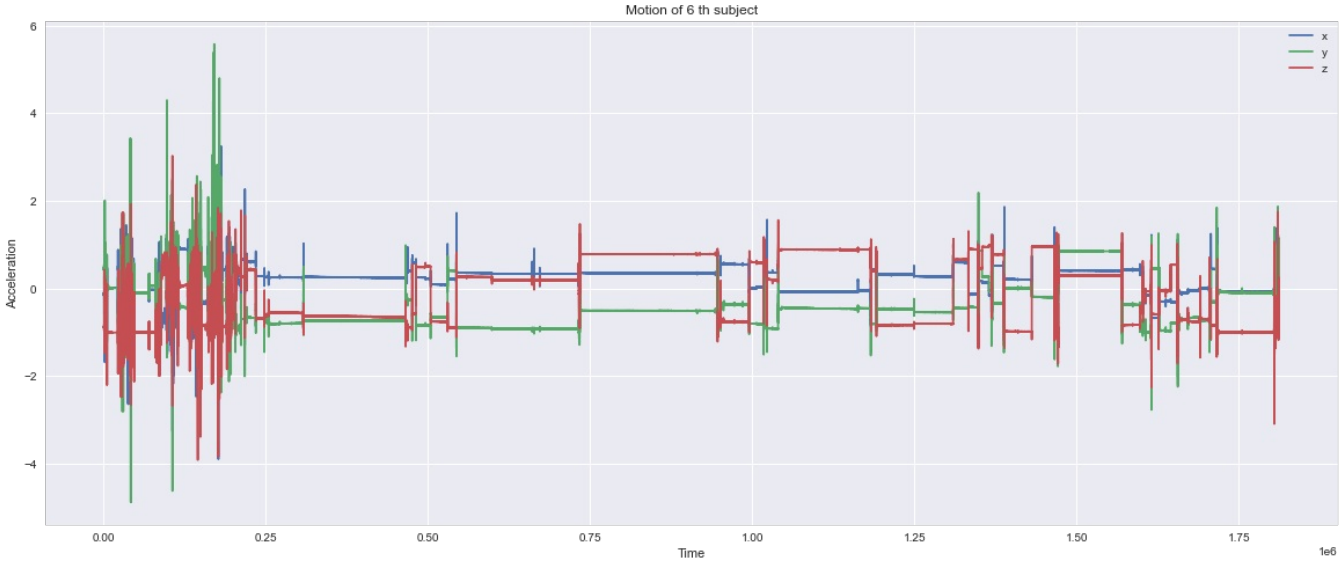

Motion of 1 th subject
x
y
z
Acceleration
Time
1e6

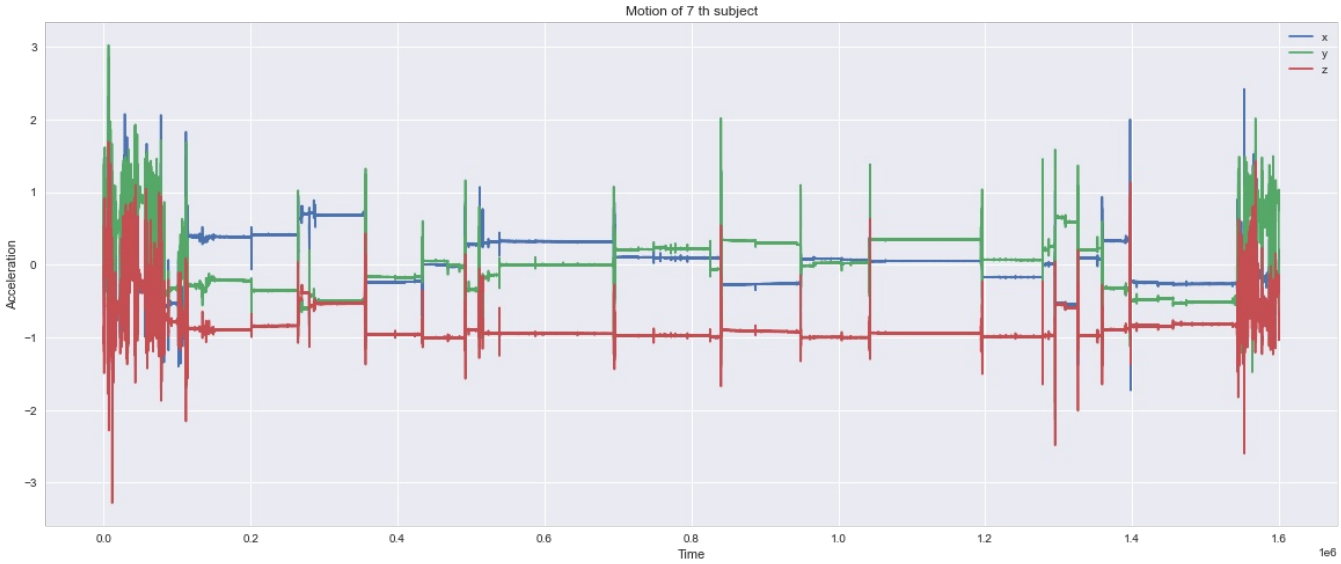

Motion of 2 th subject
x
y
z
Acceleration
Time
1e6

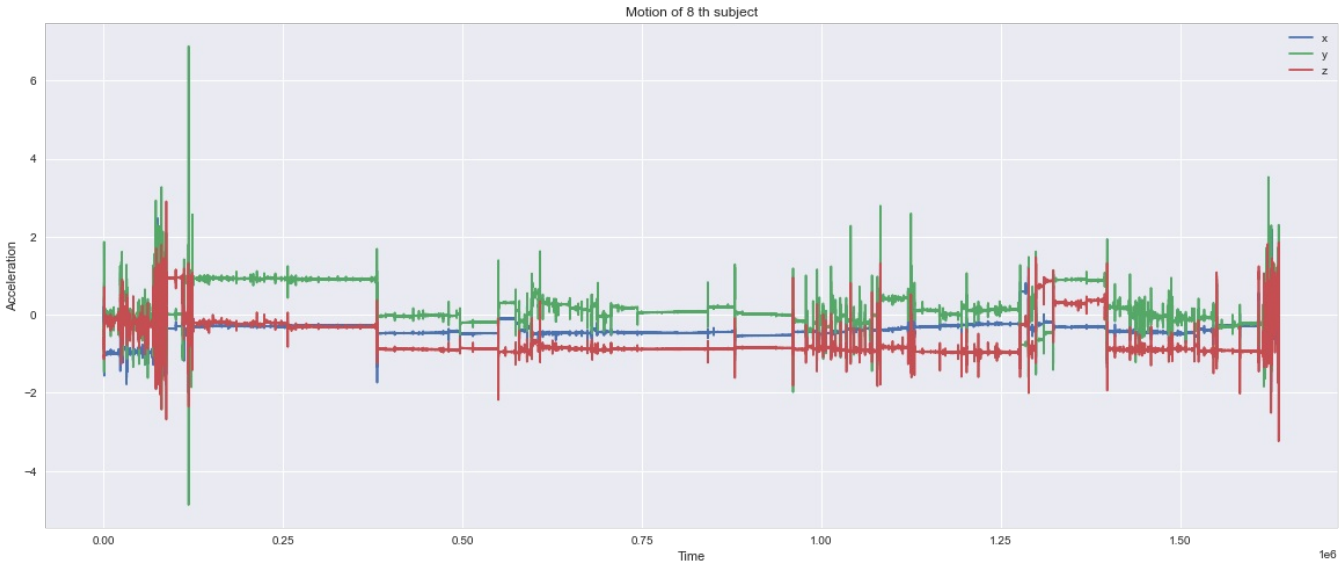

Motion of 3 th subject
x
y
z
Acceleration
Time
1e6

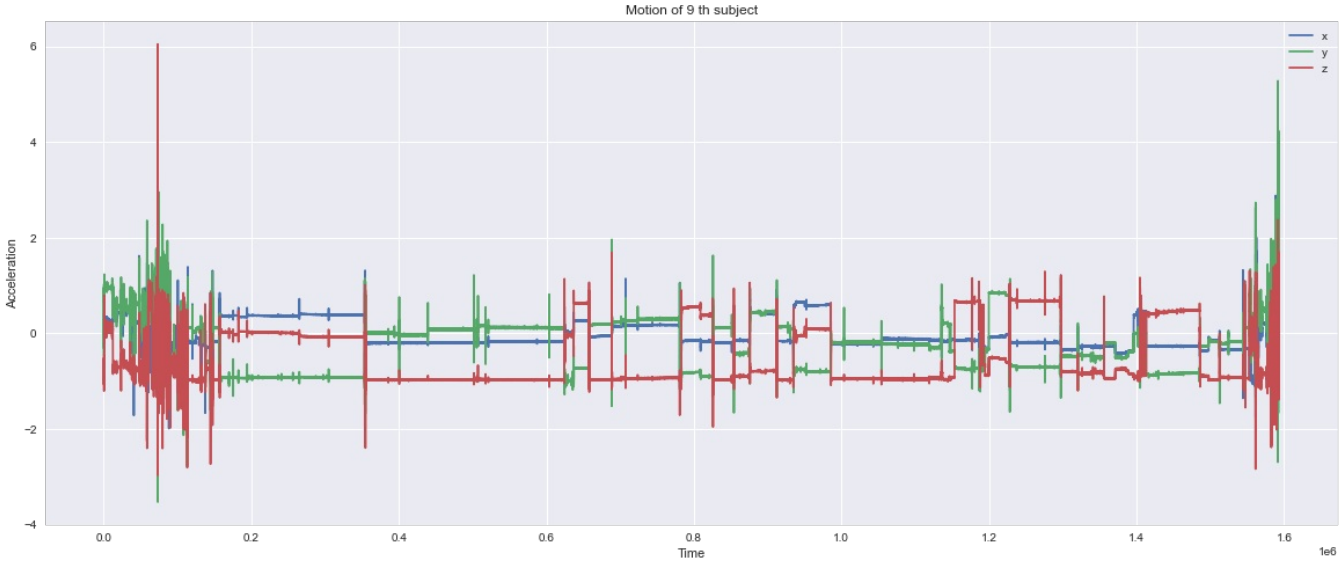

Motion of 4 th subject
x
y
z
Acceleration
Time
1e6

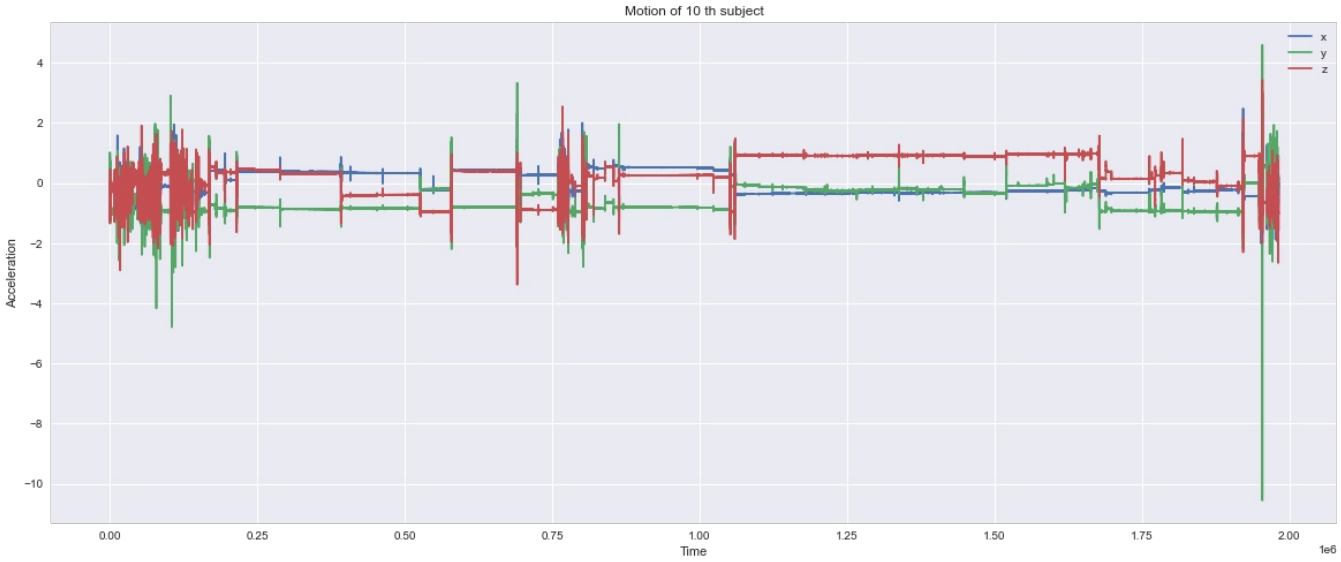

Motion of 5 th subject
x
y
z
Acceleration
Time
1e6

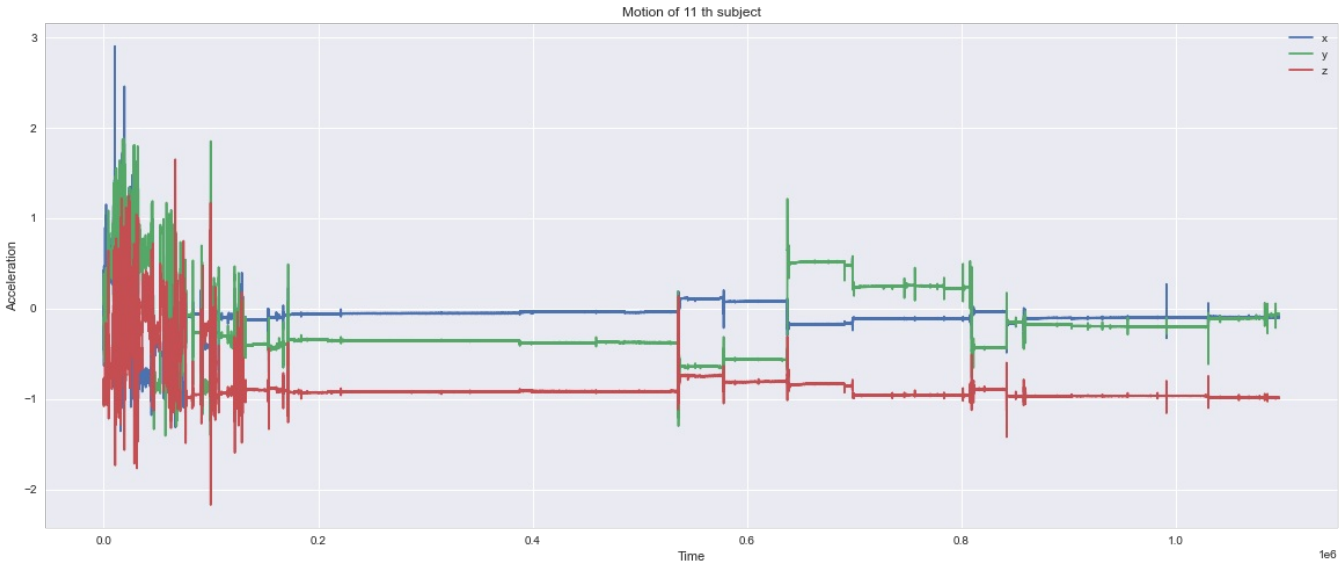

Motion of 6 th subject
x
y
z
Acceleration
6
4
2
0
−2
−4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Time
1e6

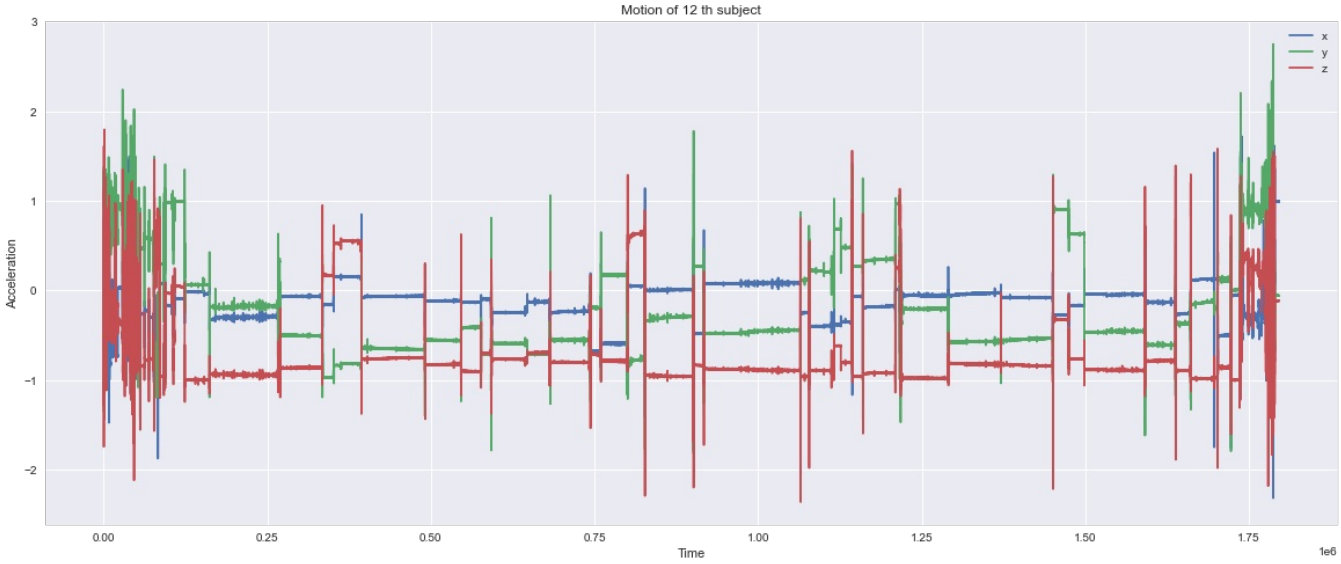

Motion of 7 th subject
x
y
z
Acceleration
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Time
1e6

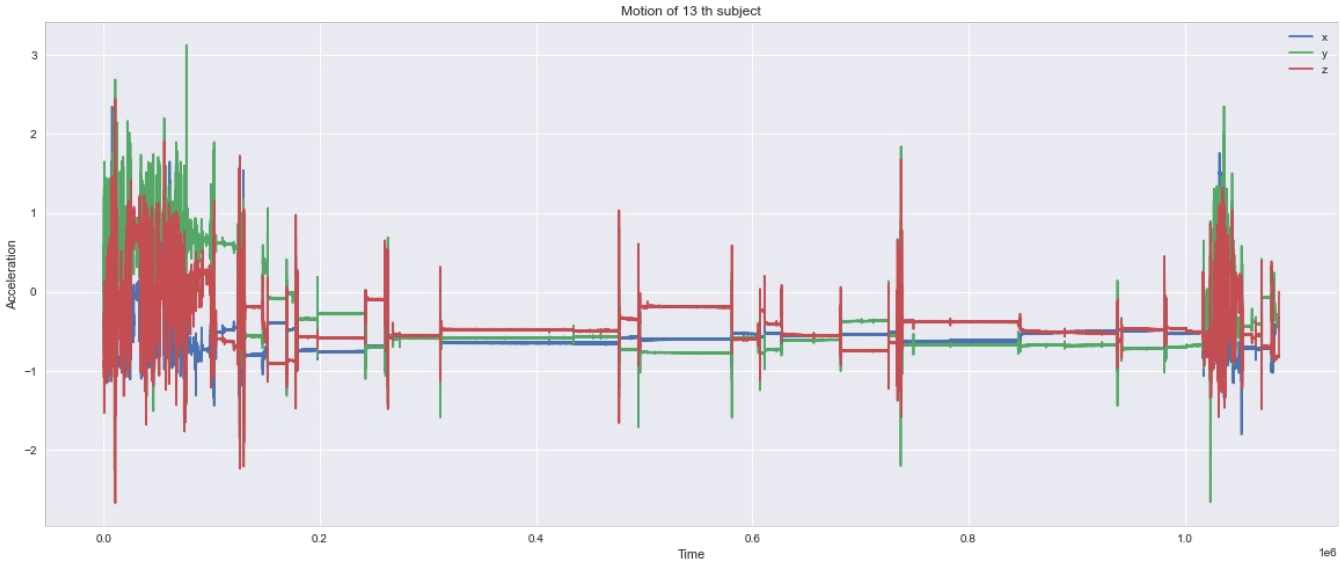

Motion of 8 th subject
x
y
z
Acceleration
6
4
2
0
−2
−4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Time
1e6

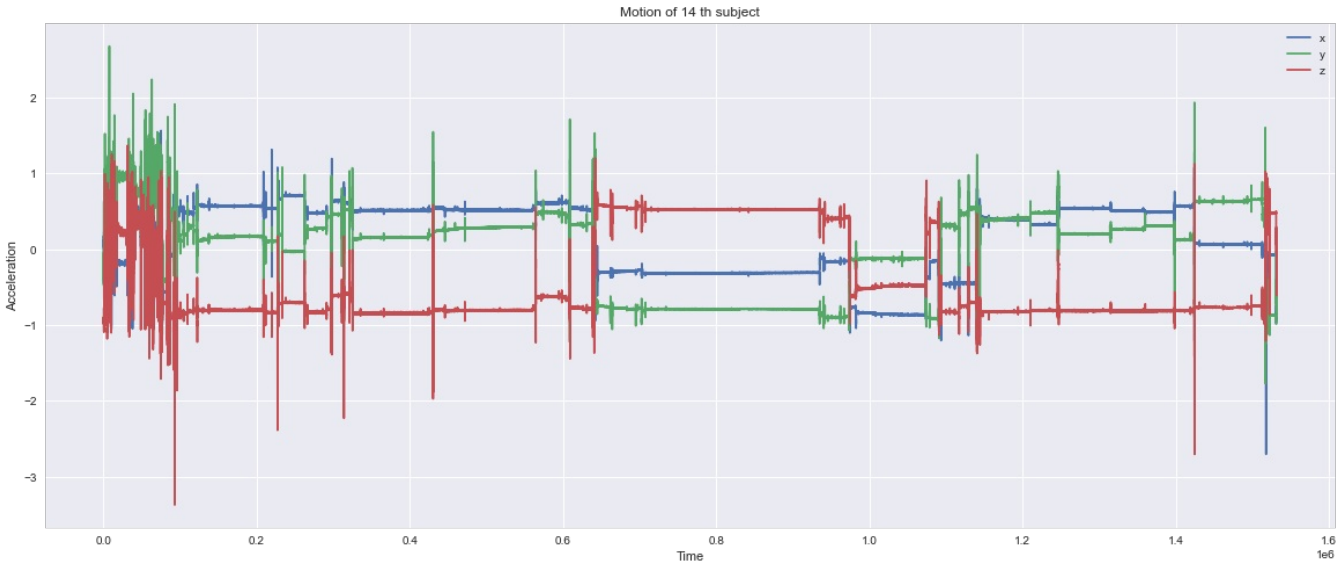

Motion of 9 th subject
x
y
z
Acceleration
Time
1e6

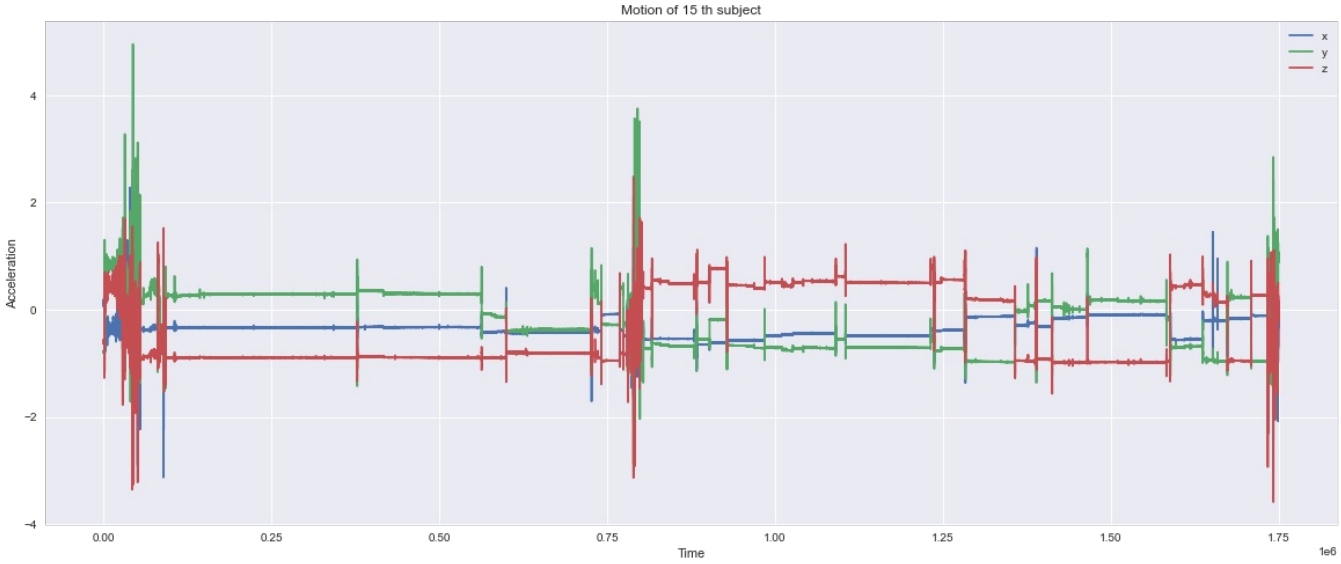

Motion of 10 th subject
x
y
z
Acceleration
Time
1e6

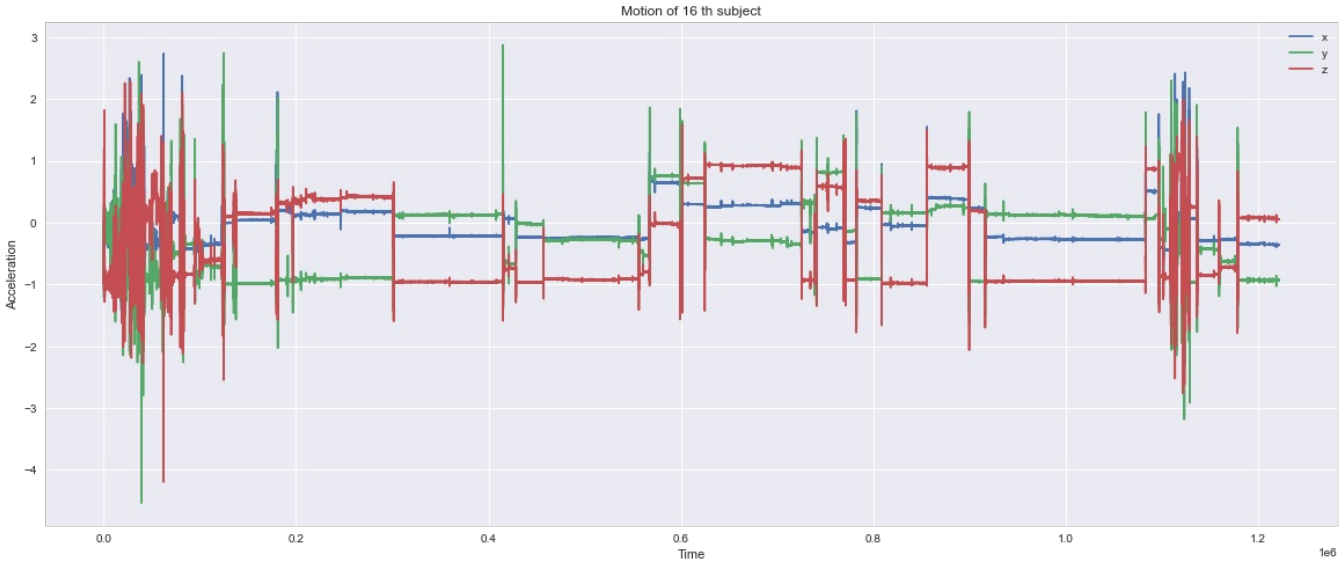

Motion of 11 th subject
x
y
z
Acceleration
Time
1e6

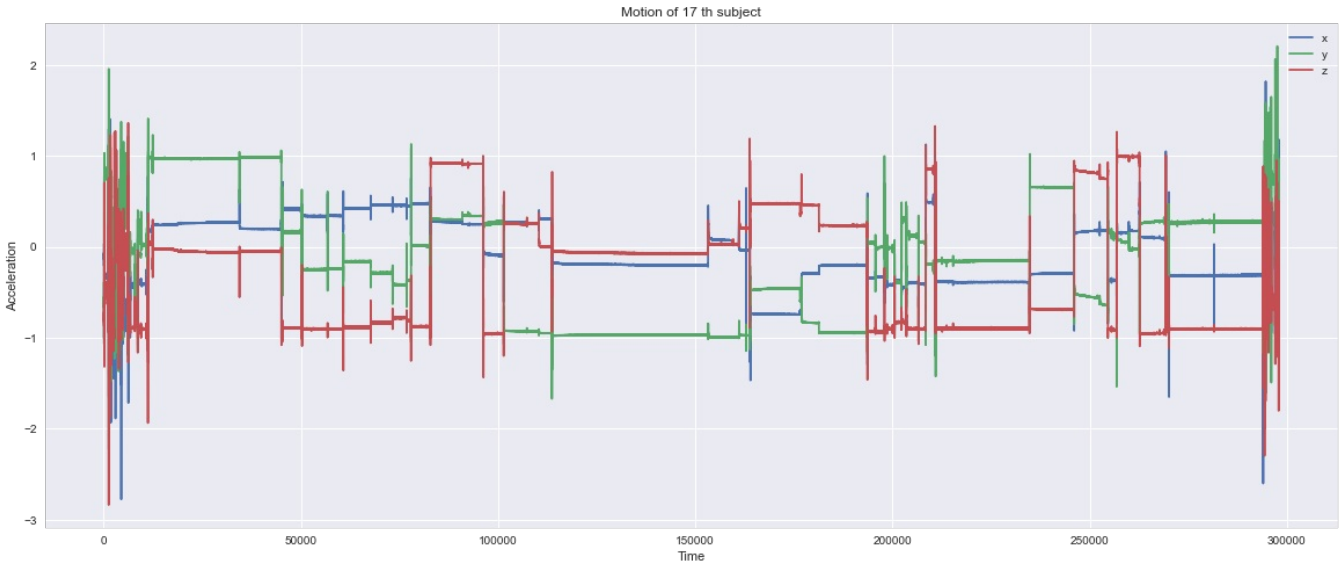

Motion of 12 th subject
x
y
z
Acceleration
Time
1e6

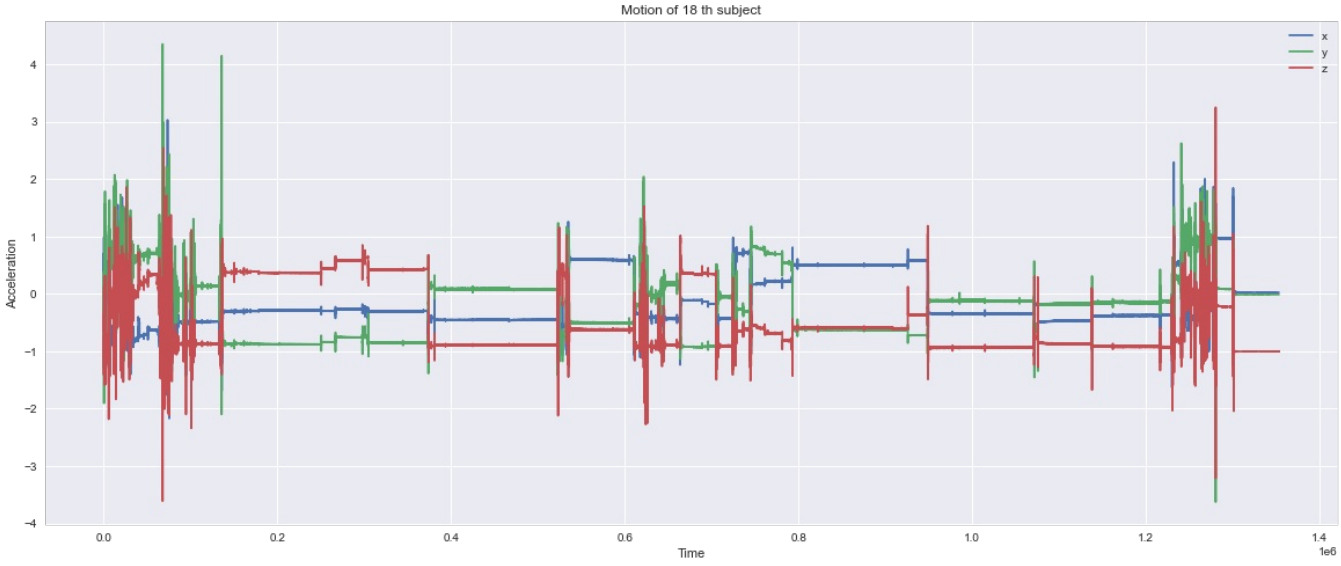

Motion of 13 th subject
x
y
z
Acceleration
Time
1e6

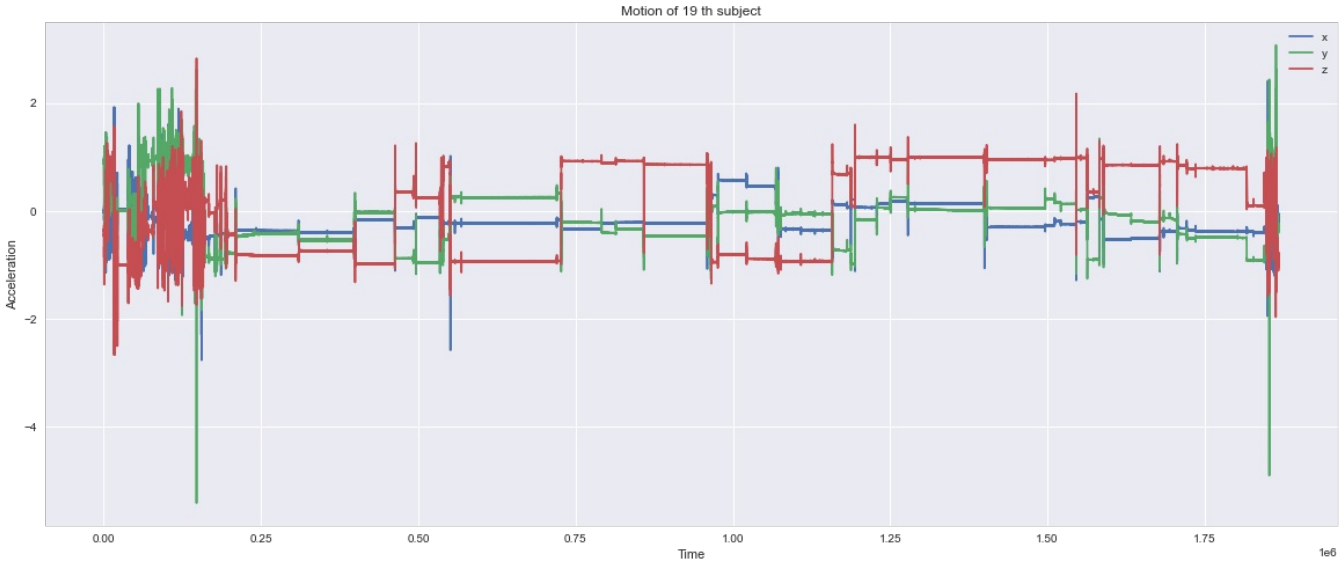

Motion of 14 th subject
x
y
z
Acceleration
Time
1e6

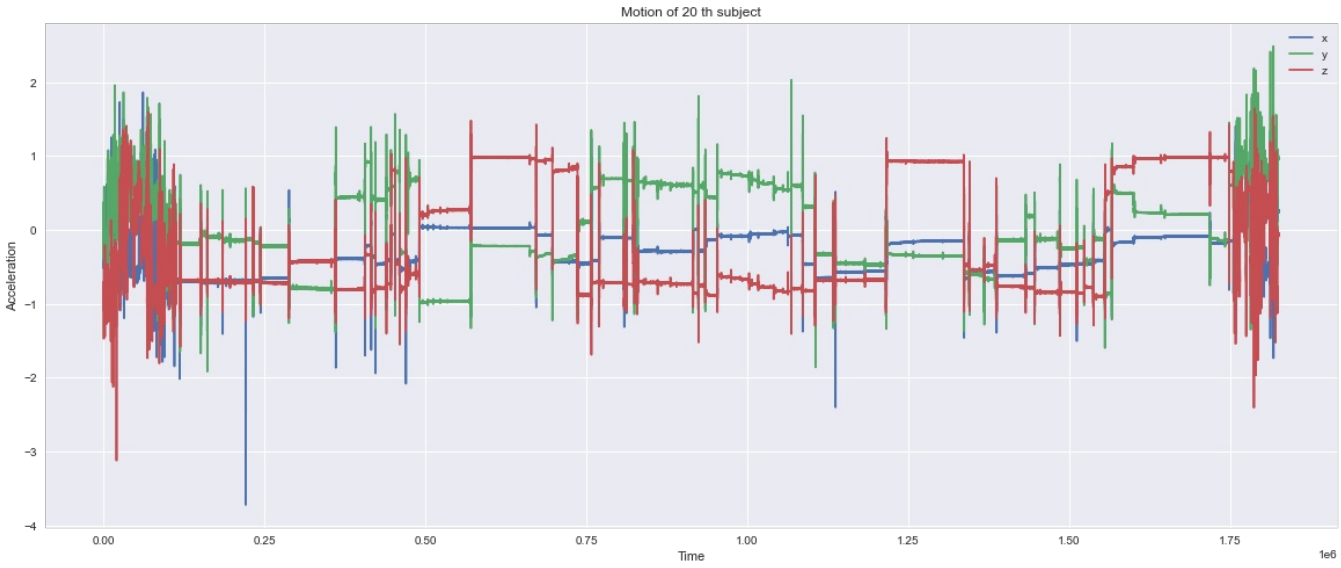

Motion of 15 th subject
x
y
z
Acceleration
Time
1e6

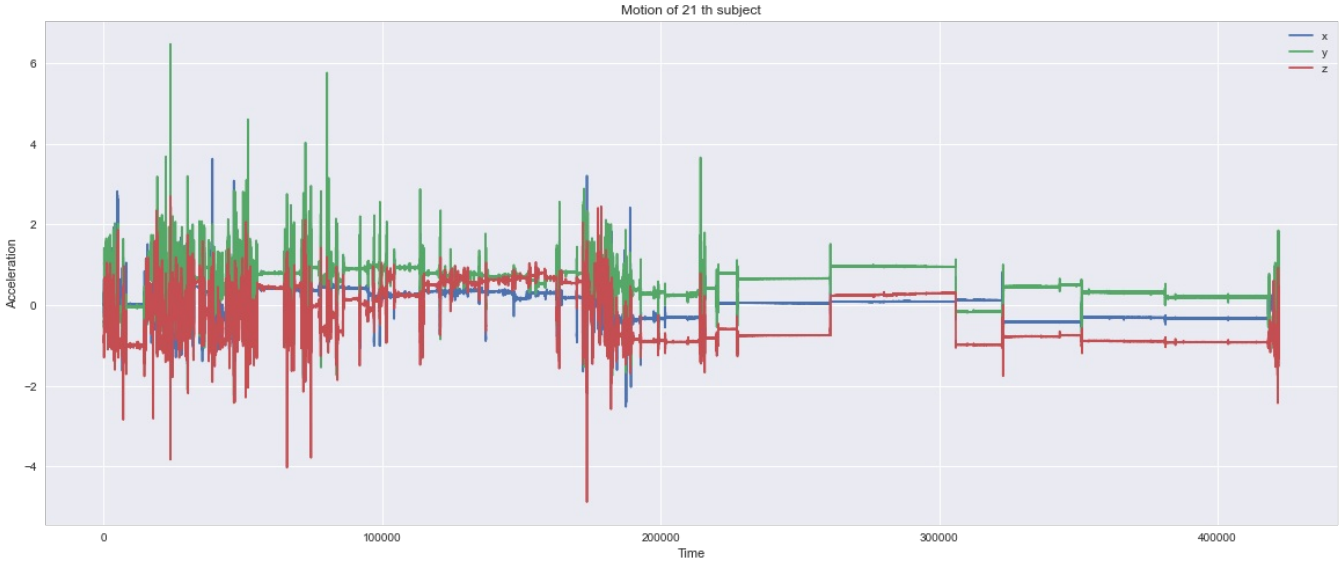

Motion of 16 th subject
x
y
z
Acceleration
Time
1e6

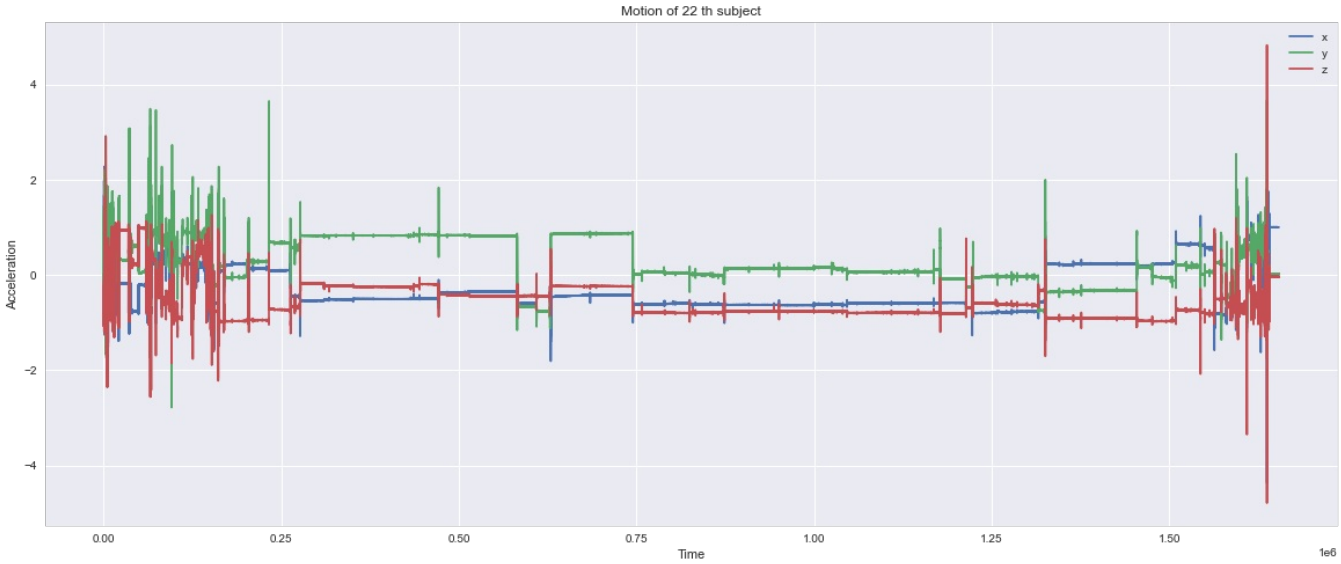

Motion of 17 th subject
x
y
z
Acceleration
Time

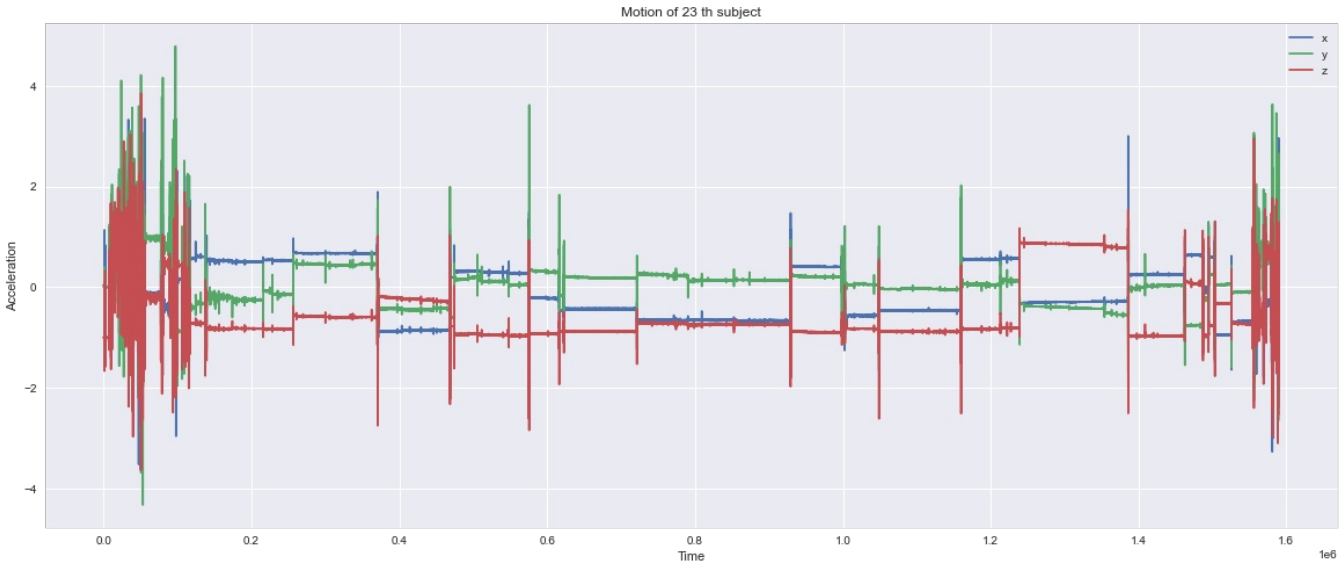

Motion of 18 th subject
x
y
z
Acceleration
Time
1e6

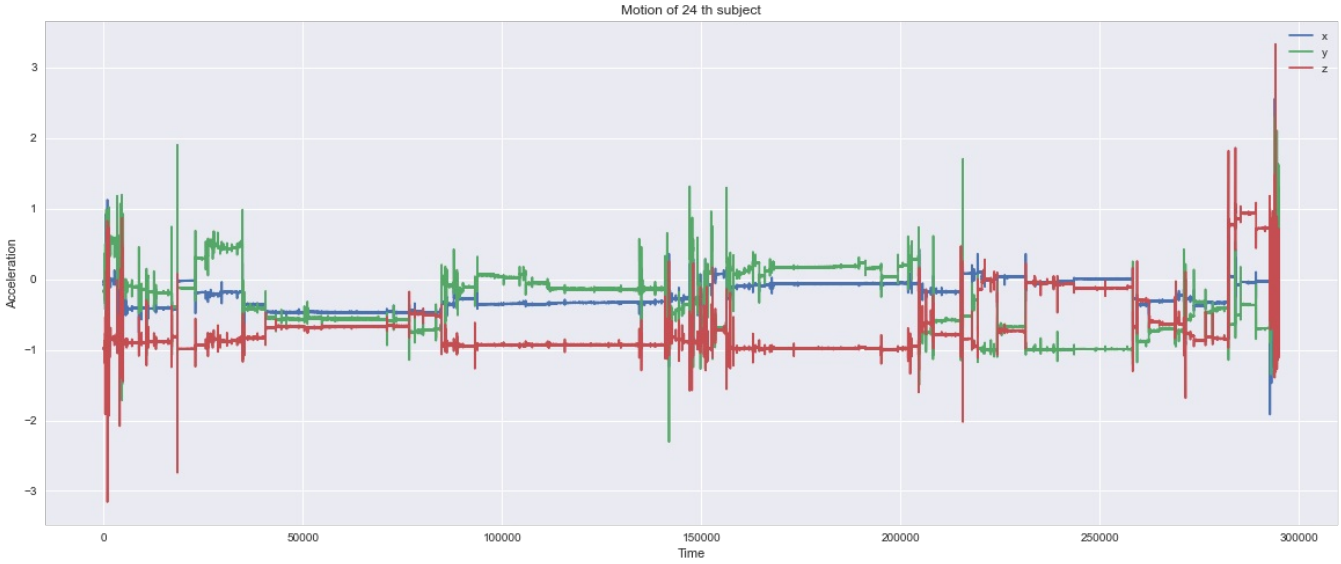

Motion of 19 th subject
x
y
z
Acceleration
Time
1e6

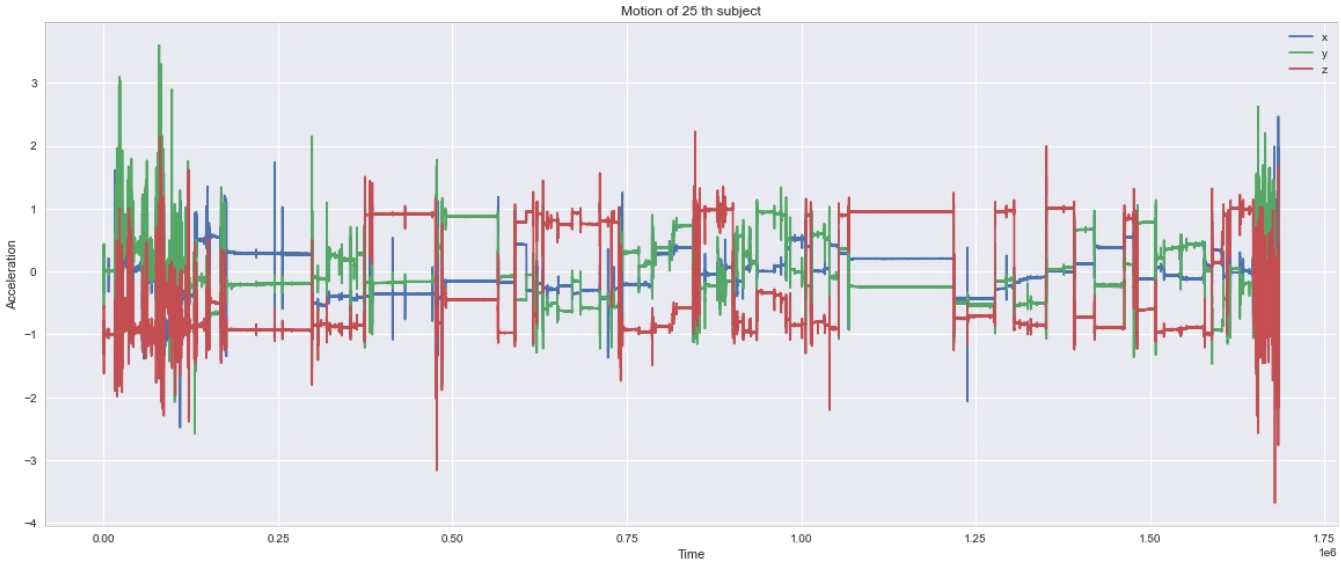

Motion of 20 th subject
x
y
z
Acceleration
Time
1e6

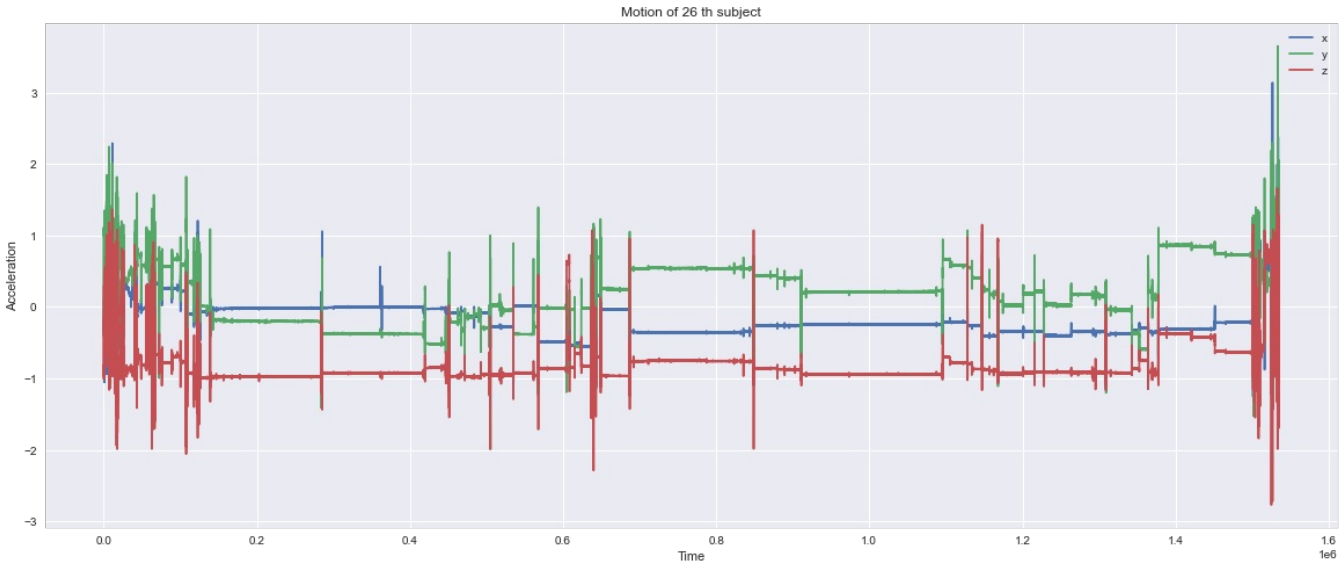

Motion of 21 th subject
x
y
z
Acceleration
6
4
2
0
−2
−4
0
100000
200000
300000
400000
Time

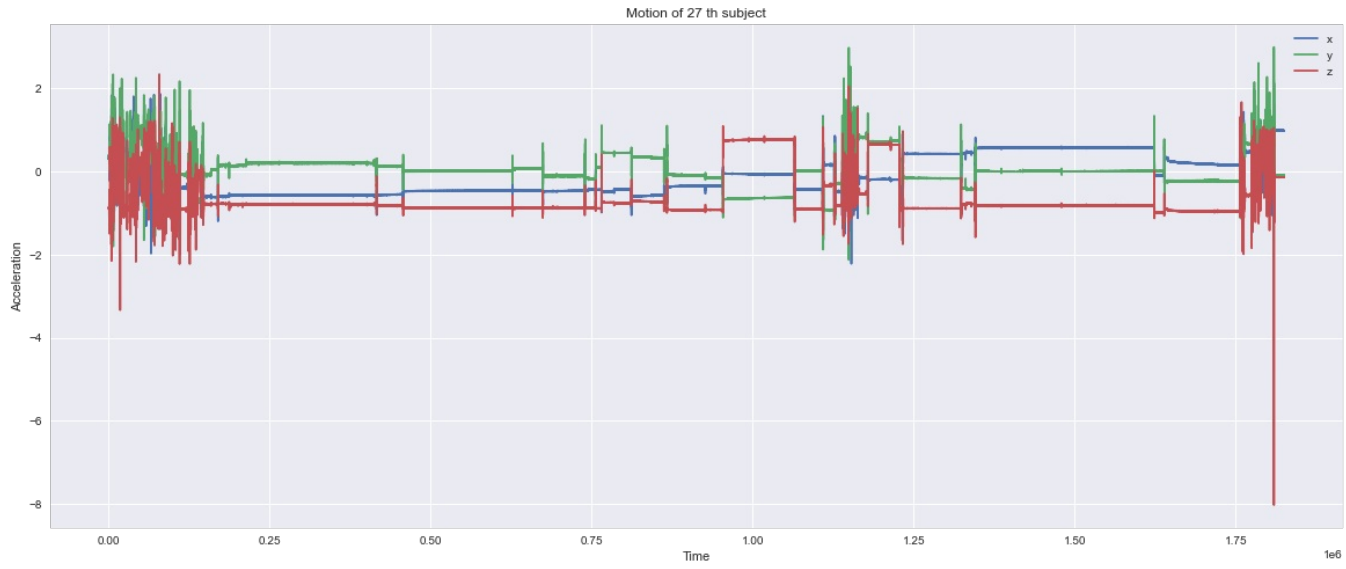

Motion of 22 th subject
x
y
z
Acceleration
4
2
0
−2
−4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Time
1e6

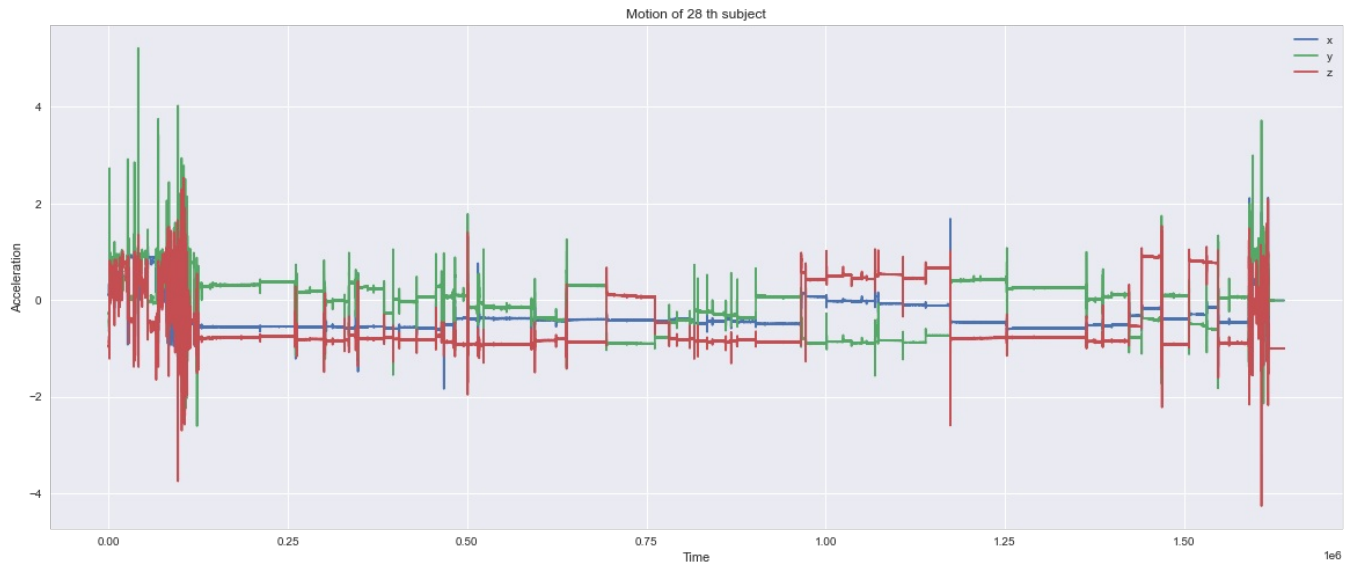

Motion of 23 th subject
x
y
z
Acceleration
4
2
0
−2
−4
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Time
1e6

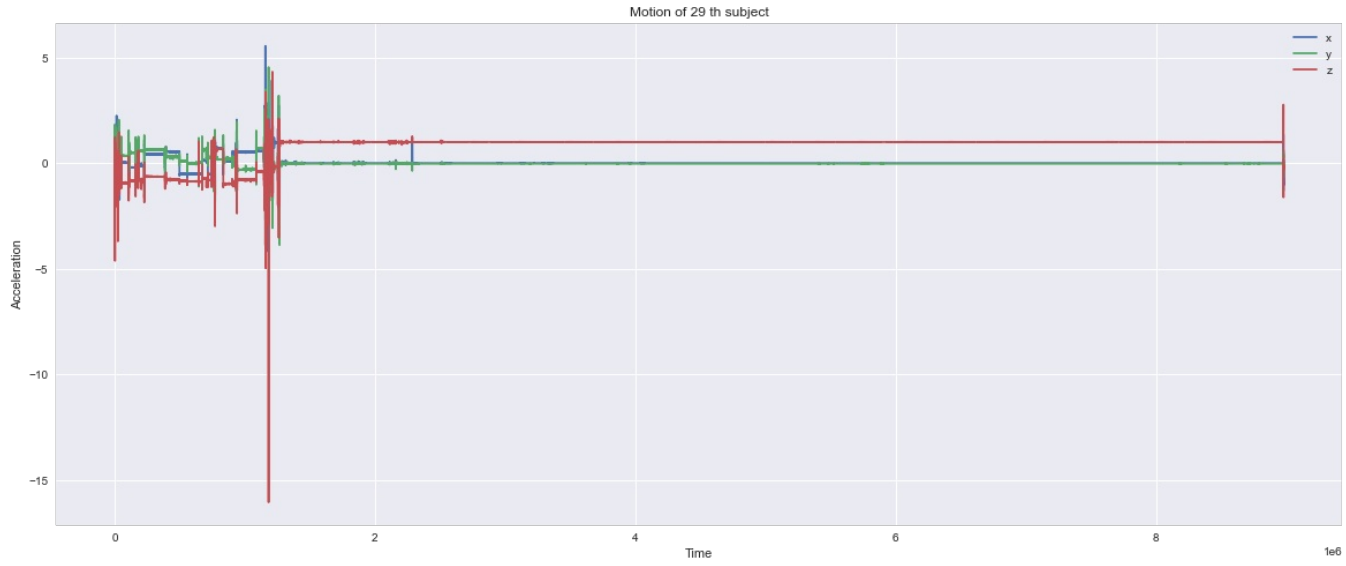

Motion of 24 th subject
x
y
z
Acceleration
Time

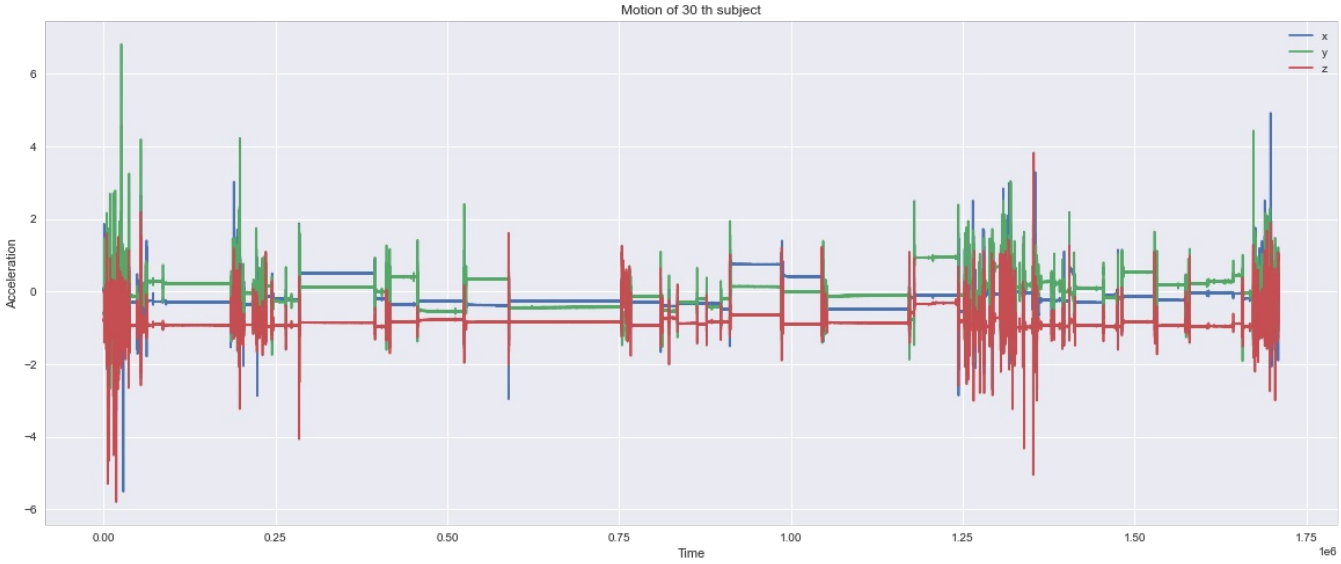

Motion of 25 th subject
x
y
z
Acceleration
Time
1e6


Motion of 26 th subject
x
y
z
Acceleration
Time
1e6


Motion of 27 th subject
x
y
z
Acceleration
2
0
−2
−4
−6
−8
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Time
1e6


Motion of 28 th subject
x
y
z
Acceleration
4
2
0
−2
−4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Time
1e6


Motion of 29 th subject
x
y
z
Acceleration
5
0
−5
−10
−15
0
2
4
6
8
Time
1e6



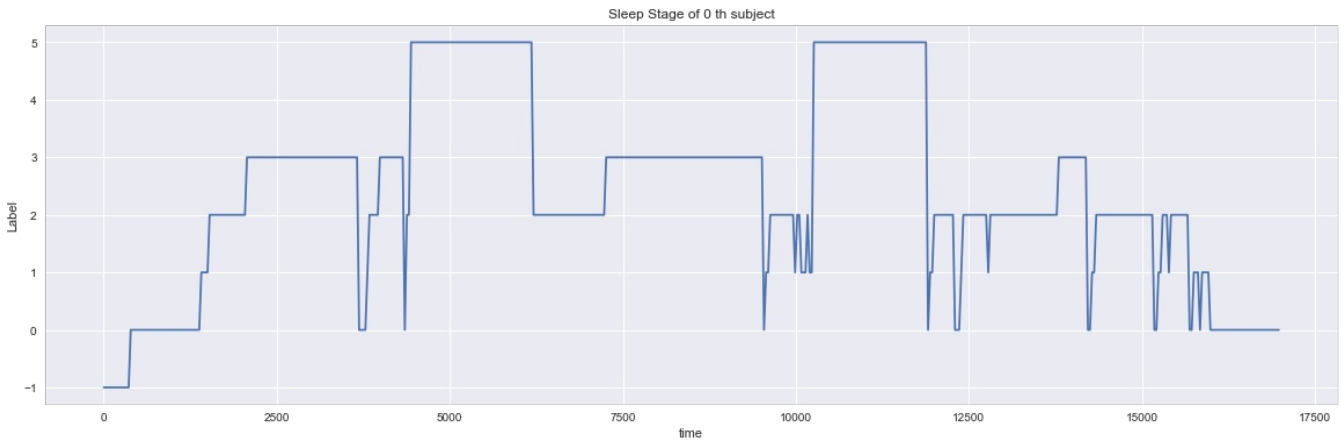
Label EDA

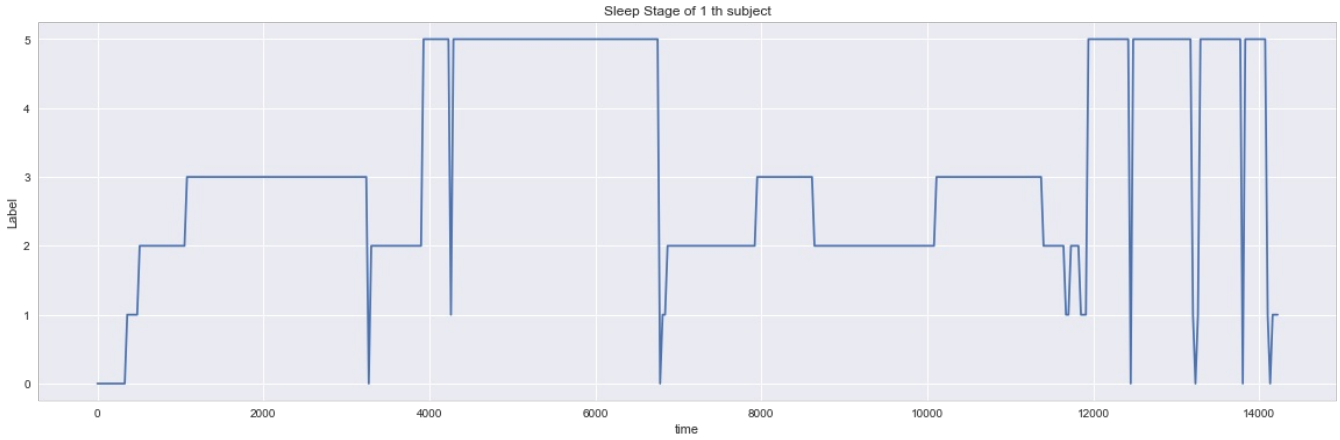
6, 17, 19, 22, 24, 25번 피험자는 중간에 -1(unscored)이 있으므로 수면 분류에 제외할 예정임.

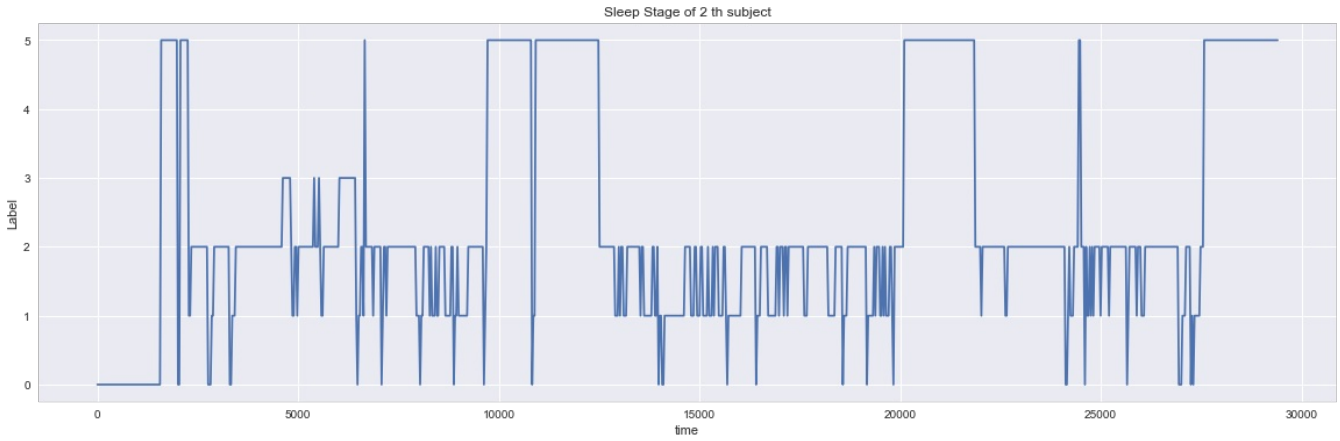
17, 24번 피험자의 수면 라벨링에 설명에 나와있지 않은 4가 있으므로 수면 분류에 제외할 예정임.

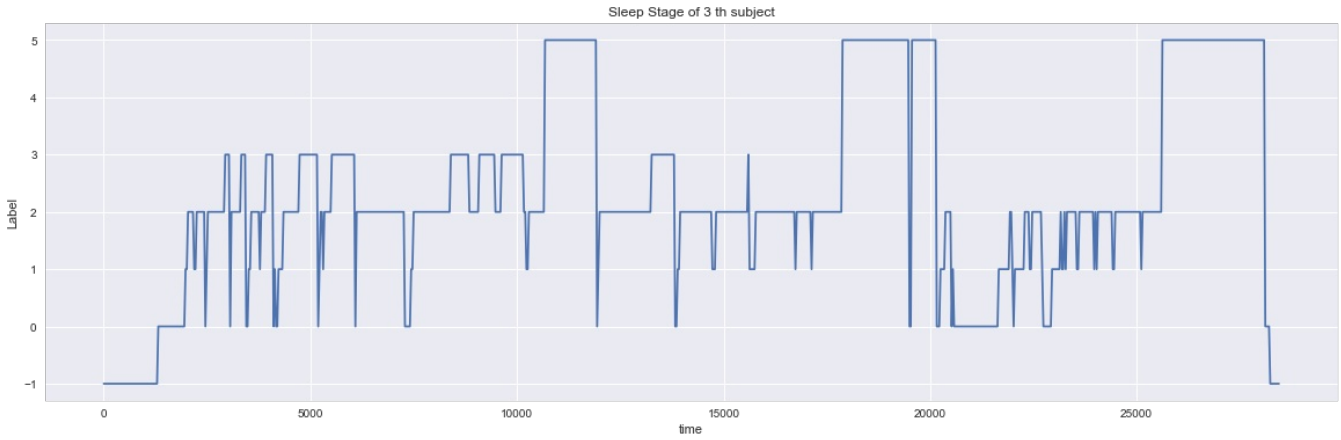
In [16]:
```
# label 데이터 프레임 리스트
label = []

dir_name = ''
path = ''
file_list = os.listdir(path)

#  주어진 파일 이름에서 숫자 부분을 추출
# 파일 이름에서 정수 값을 추출하여 파일을 숫자순으로 정렬하는 데 사용
def get_number(filename):
    return int(re.search(r'\d+', filename).group())

# 리스트 파일 이름 순으로 정렬
file_list_txt = sorted([file for file in file_list], key=get_number)

# 텍스트 파일을 csv 파일로 저장
for i in file_list_txt:
    df = pd.read_csv(dir_name+i, sep=' ', header=None, names=['time', 'label'])
    df = df.set_index('time') # 인덱스를 시간으로 지정
    label.append(df)
```

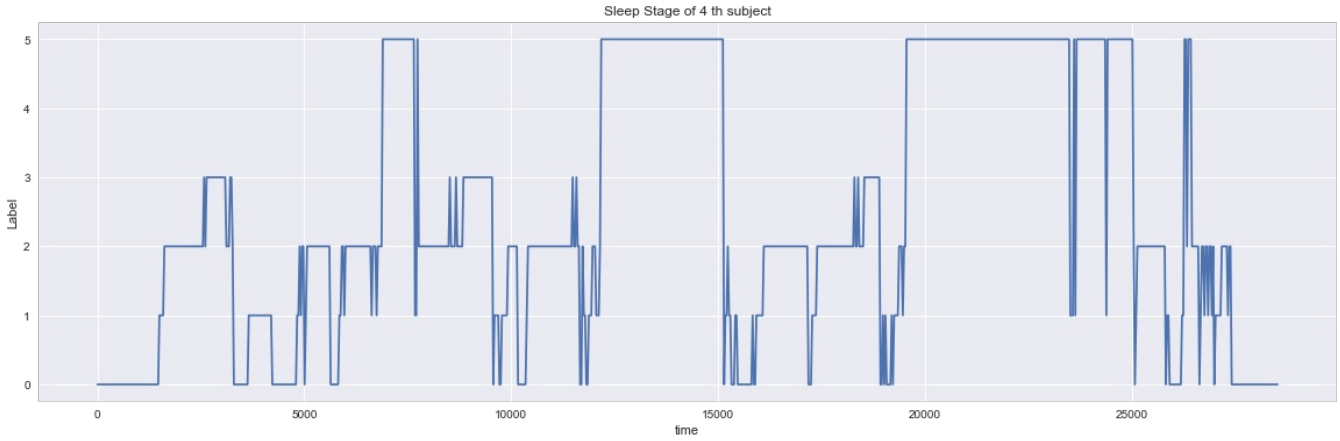
In [17]:
```
# 6, 17, 19, 22, 24, 25 는 중간에 -1이 있으므로 제외

for i in range(len(label)):
    plt.figure(figsize=(20, 6))
    plt.title(f'Sleep Stage of {i} th subject')
    plt.xlabel('Time')
    plt.ylabel('Label')

    label[i]['label'].plot()
    plt.show()
```

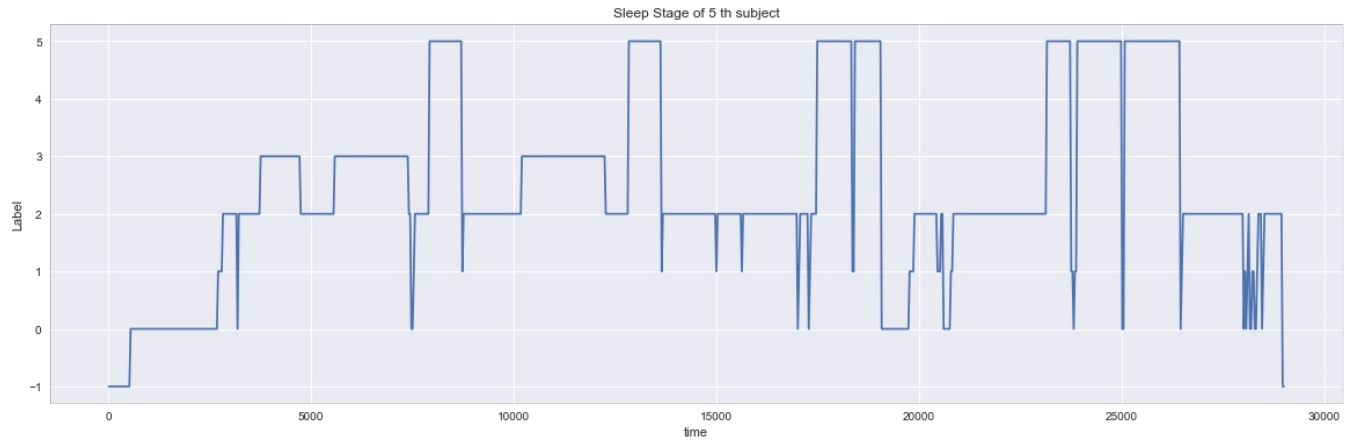


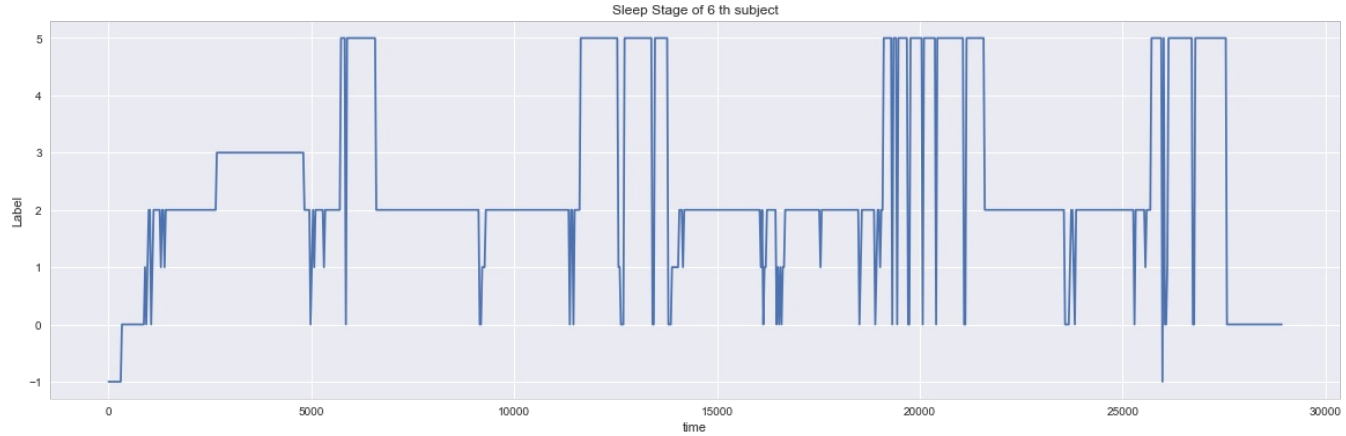

Sleep Stage of 1 th subject
Label
time
0
1
2
3
4
5
0
2000
4000
6000
8000
10000
12000
14000

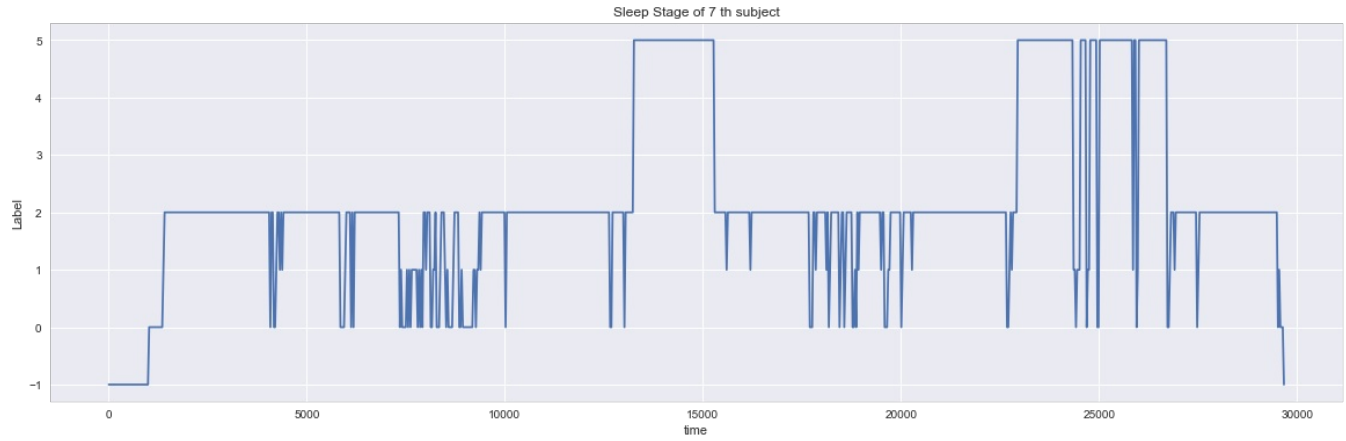

Sleep Stage of 2 th subject
Label
time
0
1
2
3
4
5
0
5000
10000
15000
20000
25000
30000

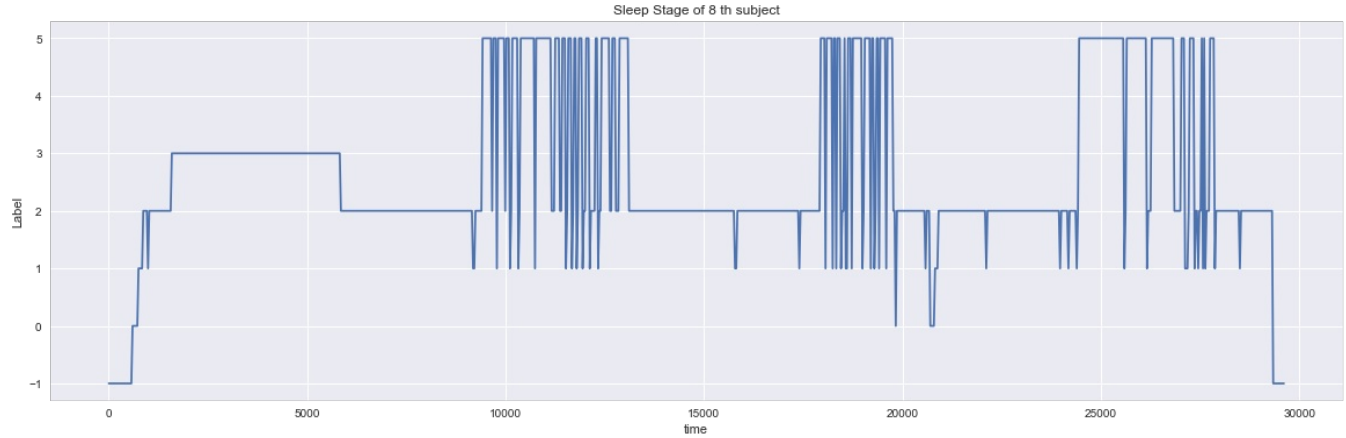

Sleep Stage of 3 th subject
Label
time
−1
0
1
2
3
4
5
0
5000
10000
15000
20000
25000

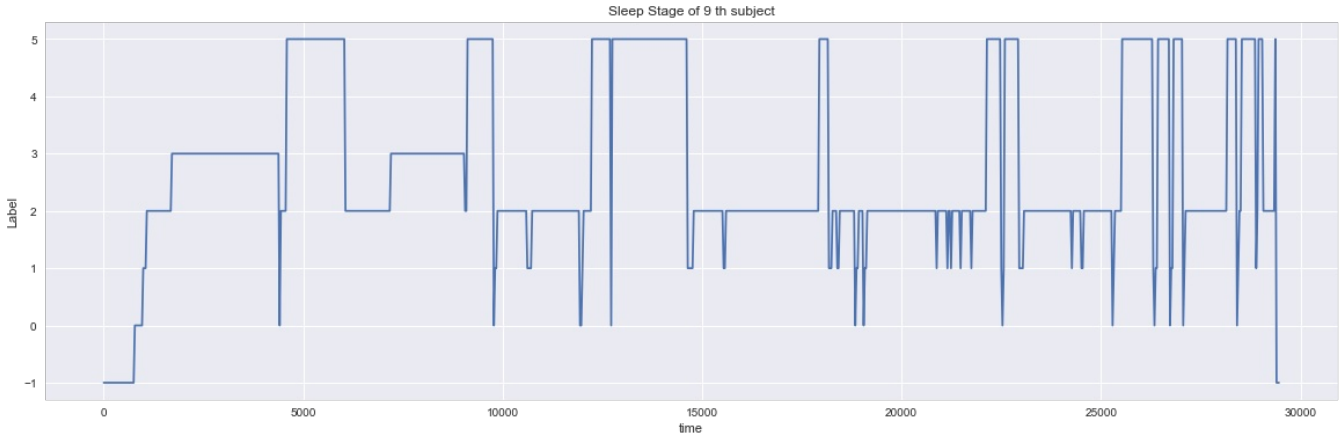

Sleep Stage of 4 th subject
Label
time
0
1
2
3
4
5
0
5000
10000
15000
20000
25000

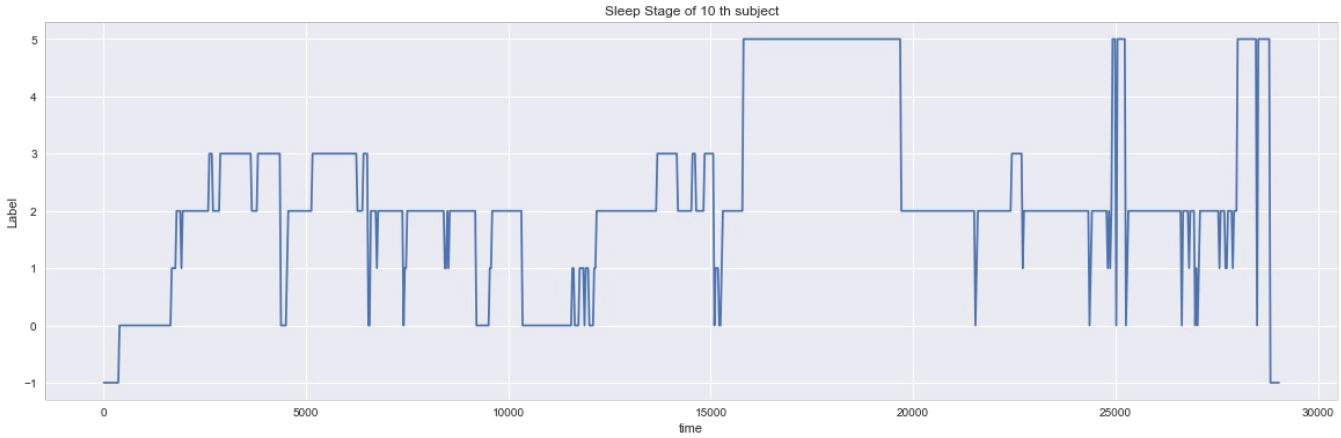

Sleep Stage of 5 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

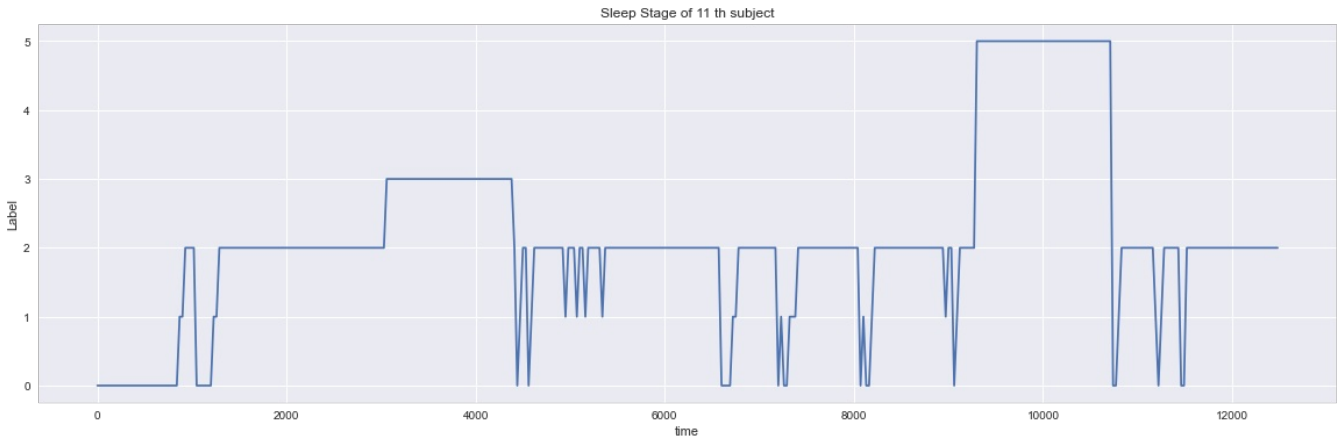

Sleep Stage of 6 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

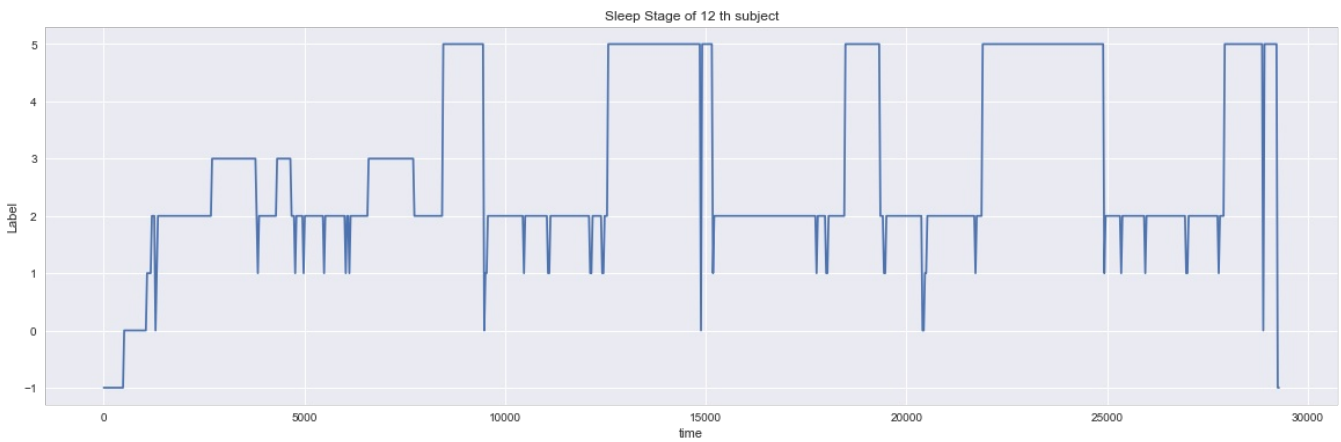

Sleep Stage of 7 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

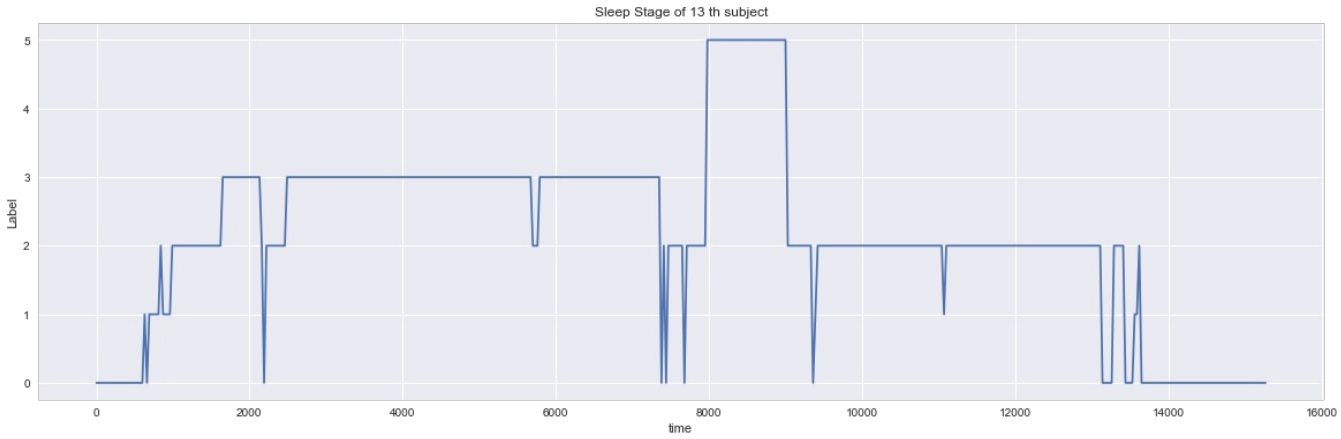

Sleep Stage of 8 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

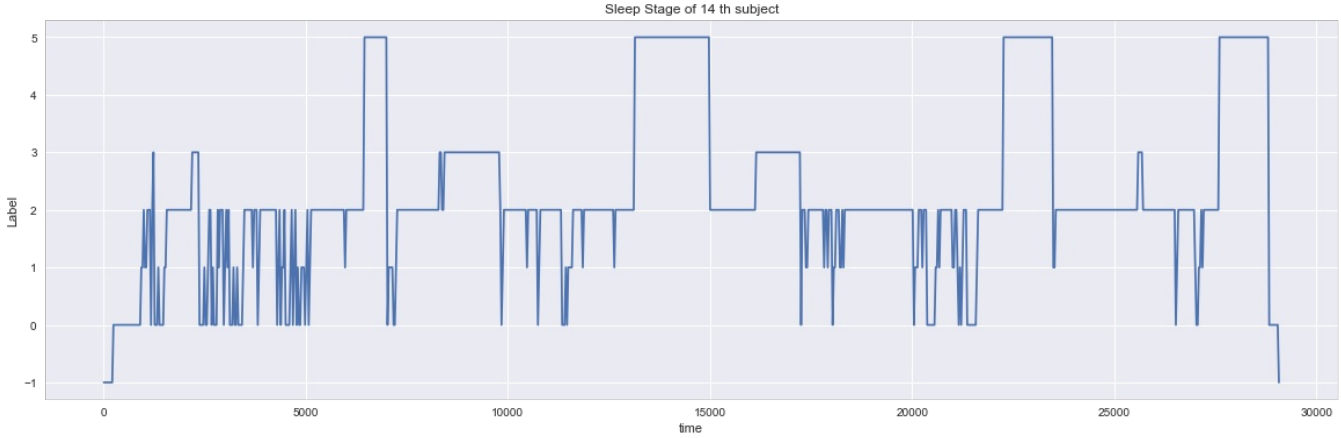

Sleep Stage of 9 th subject
Label
time

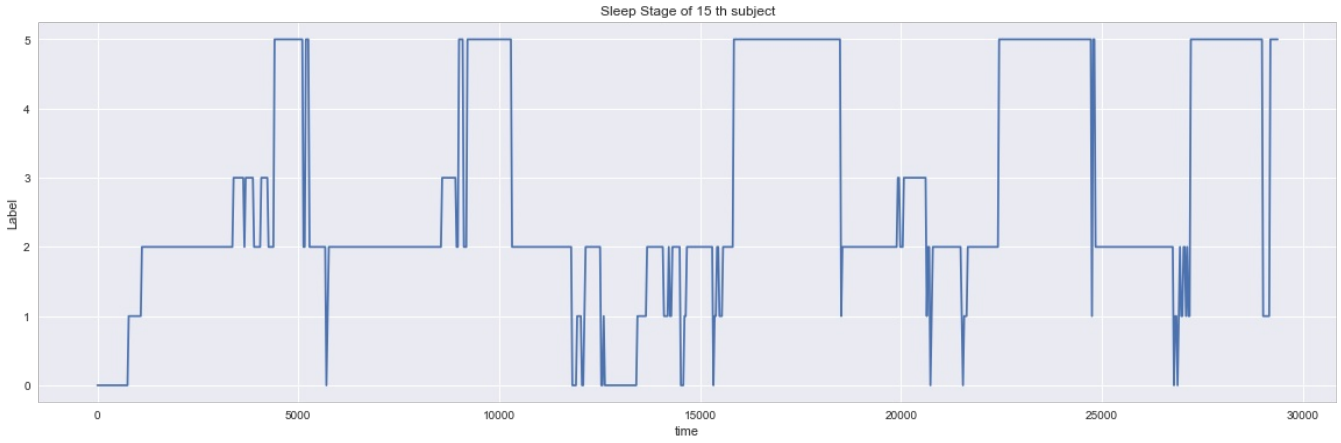

Sleep Stage of 10 th subject
Label
time

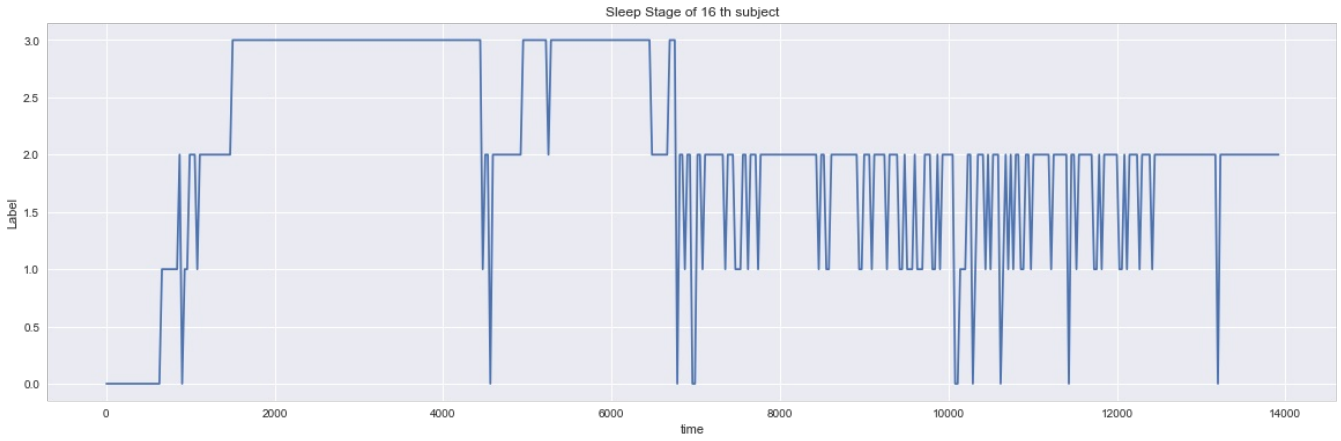

Sleep Stage of 11 th subject
Label
time

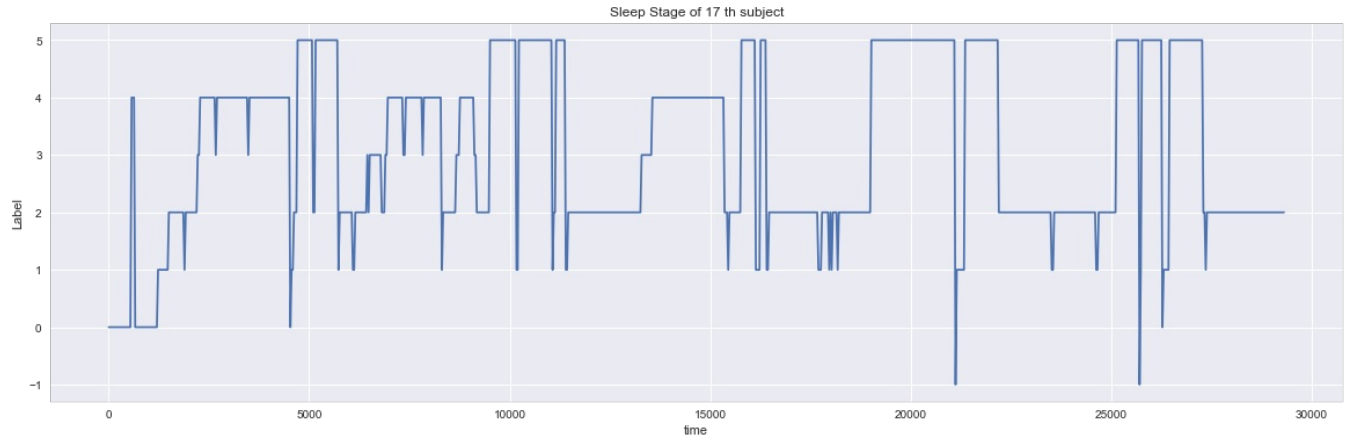

Sleep Stage of 12 th subject
Label
time

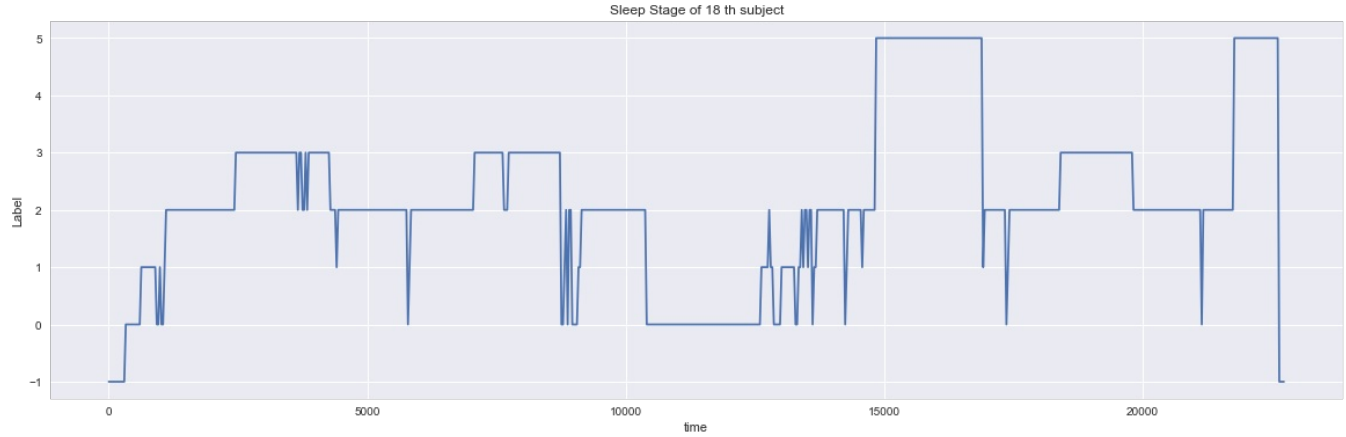

Sleep Stage of 13 th subject
Label
time
0
1
2
3
4
5
0
2000
4000
6000
8000
10000
12000
14000
16000

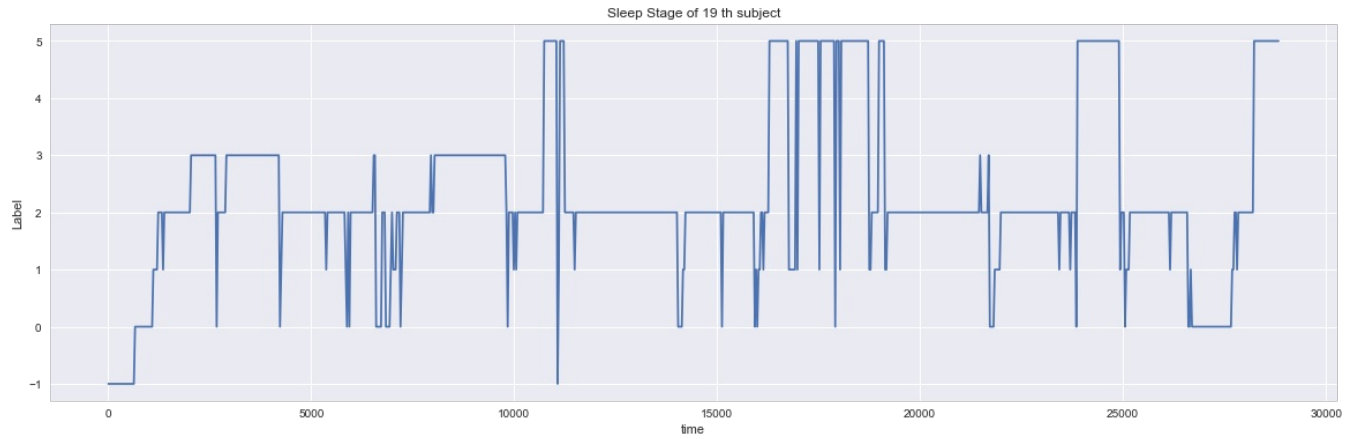

Sleep Stage of 14 th subject
Label
time
−1
0
1
2
3
4
5
0
5000
10000
15000
20000
25000
30000

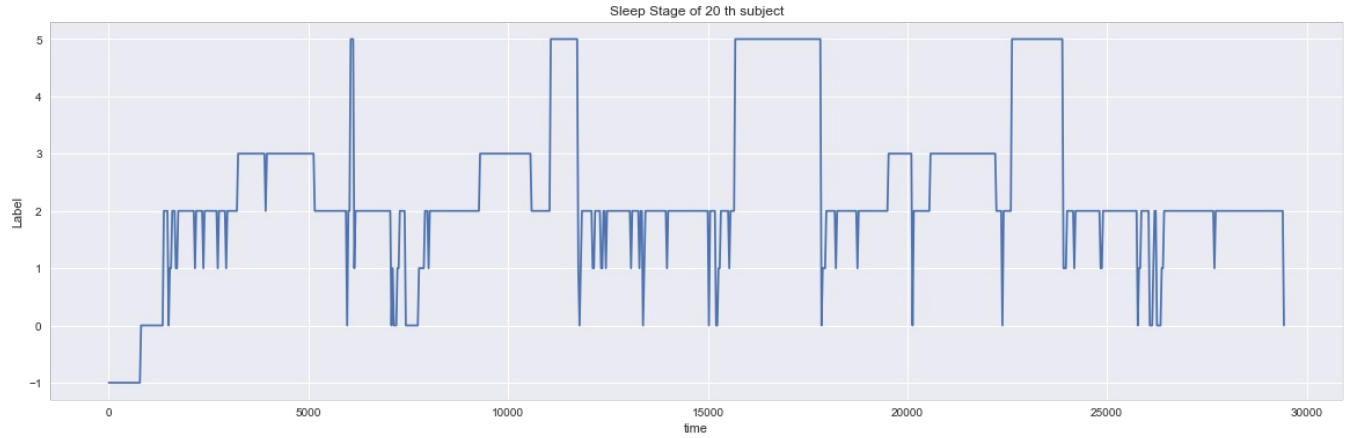

Sleep Stage of 15 th subject
Label
time
0
1
2
3
4
5
0
5000
10000
15000
20000
25000
30000

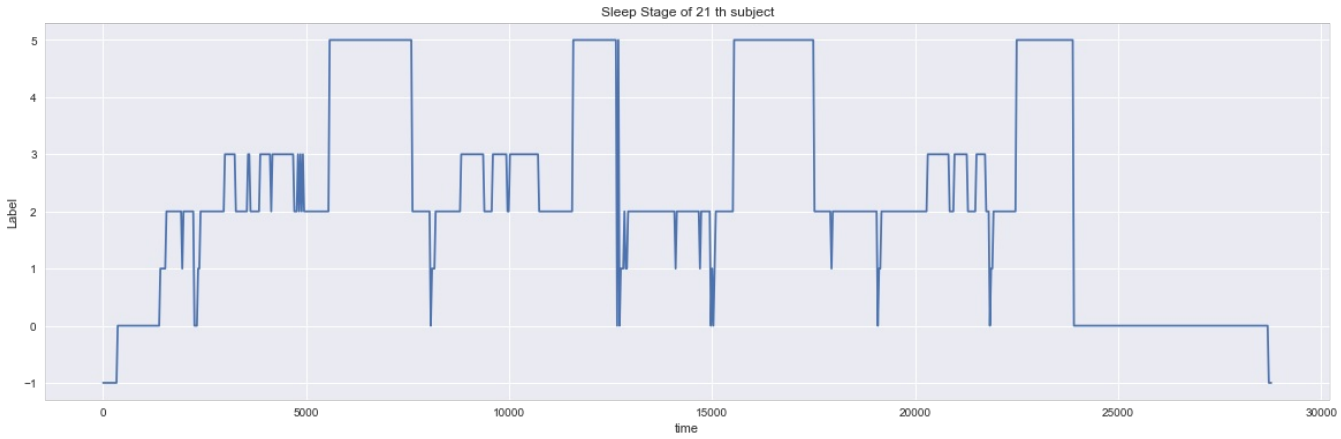

Sleep Stage of 16 th subject
Label
time
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
2000
4000
6000
8000
10000
12000
14000

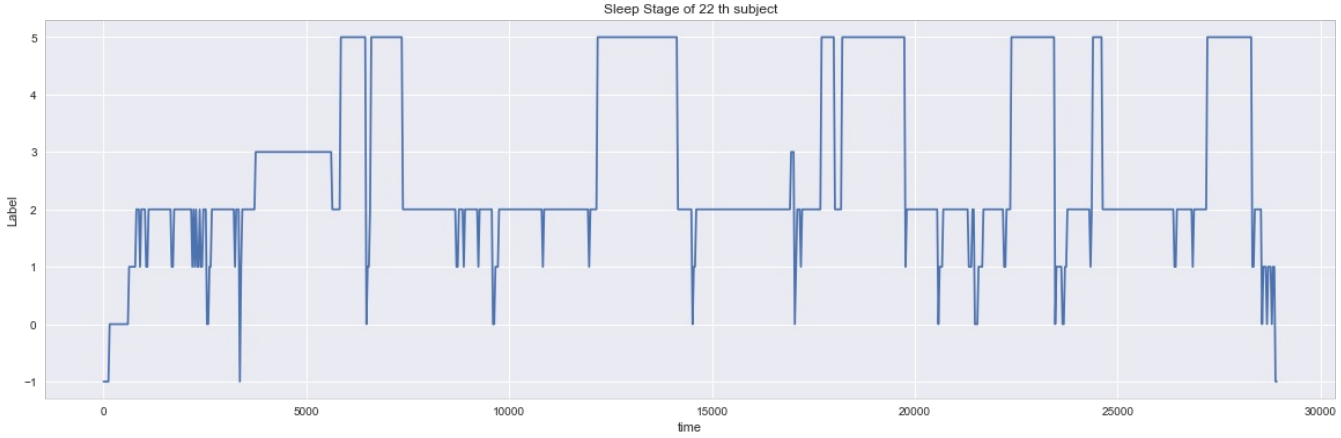

Sleep Stage of 17 th subject
Label
time

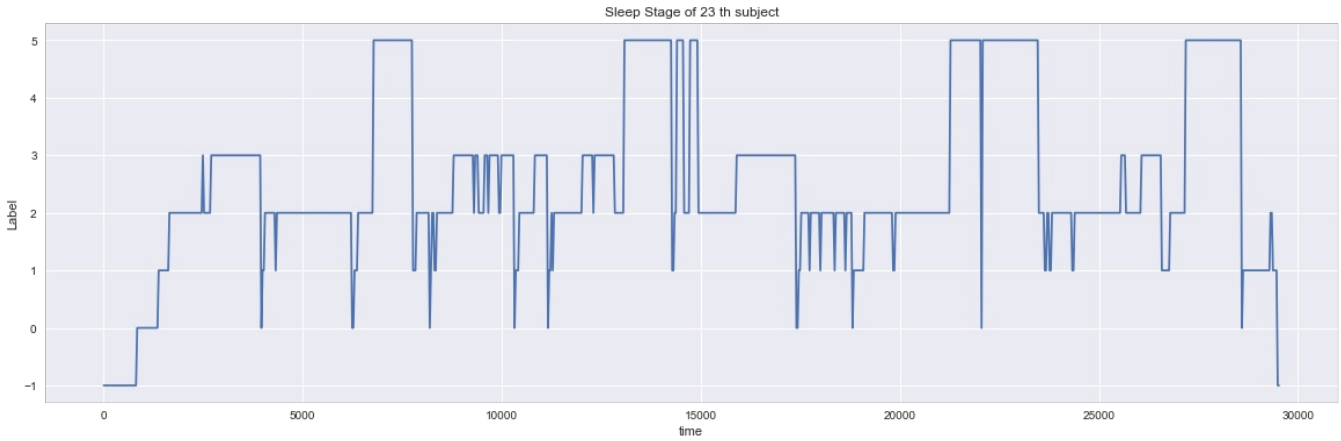

Sleep Stage of 18 th subject
Label
time

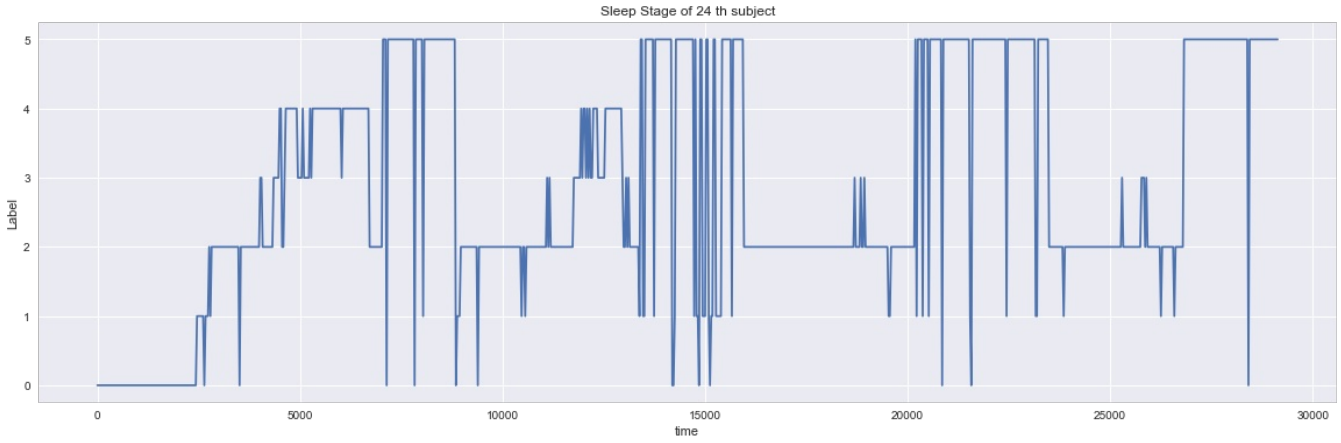

Sleep Stage of 19 th subject
Label
time

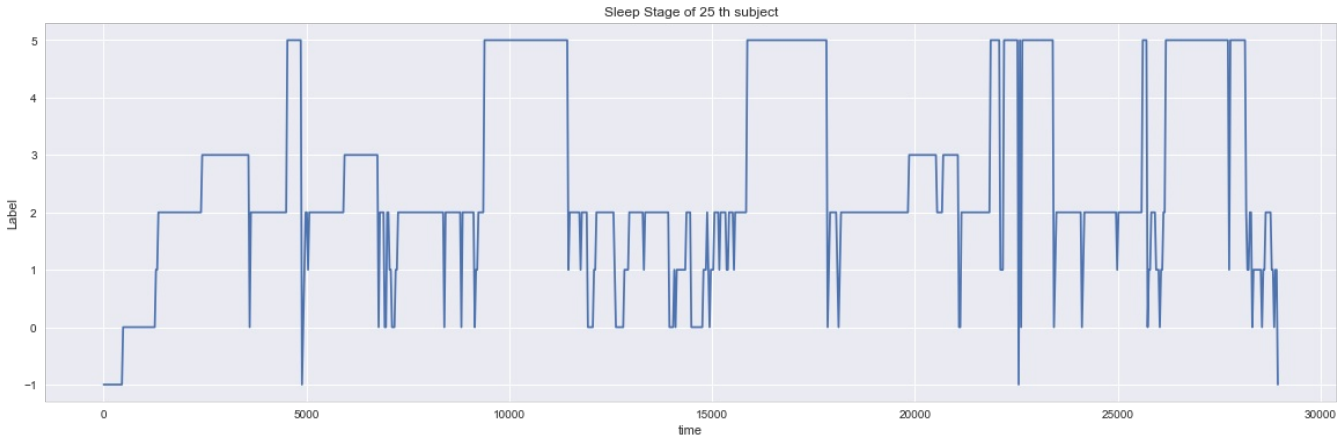

Sleep Stage of 20 th subject
Label
time

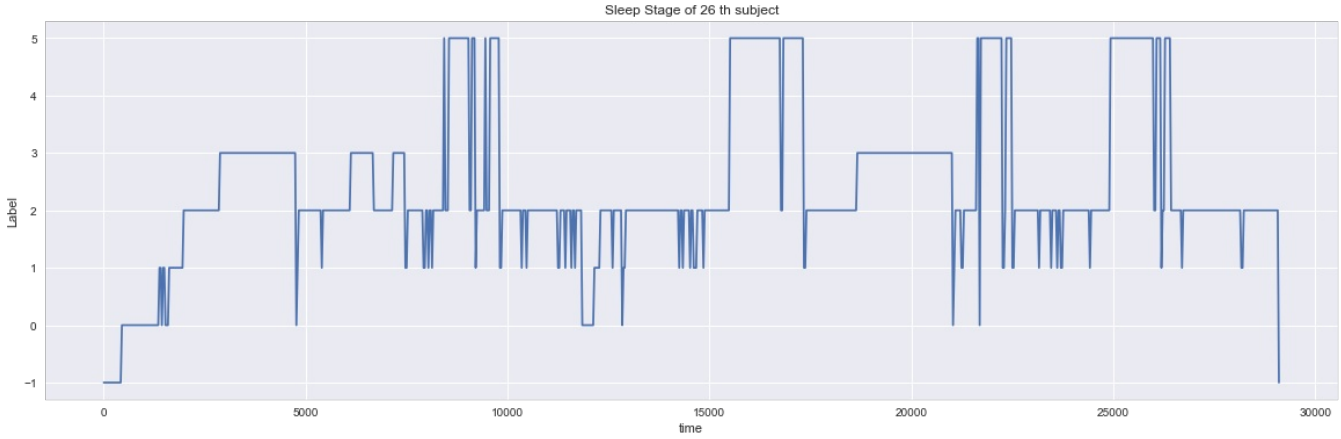

Sleep Stage of 21 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

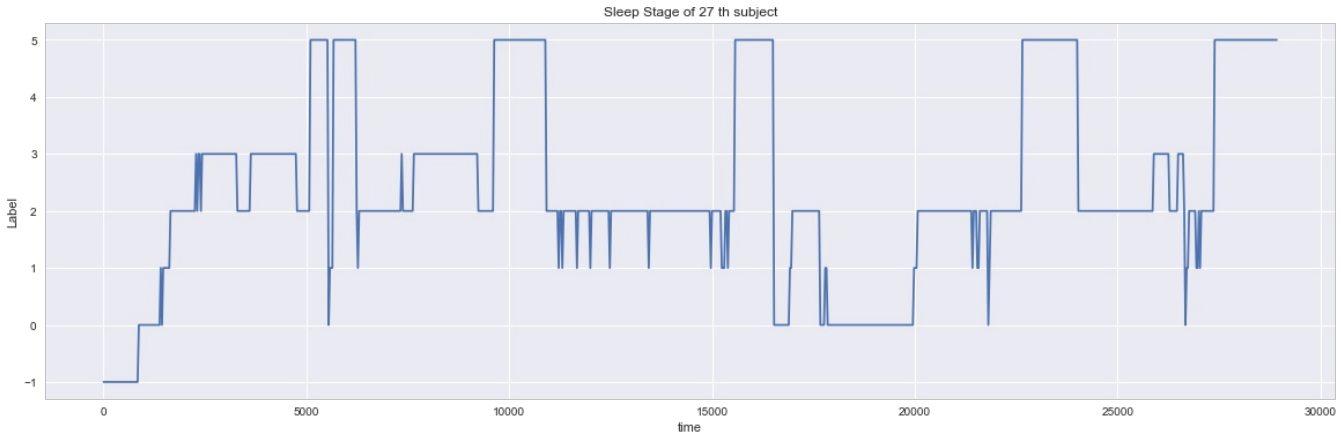

Sleep Stage of 22 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

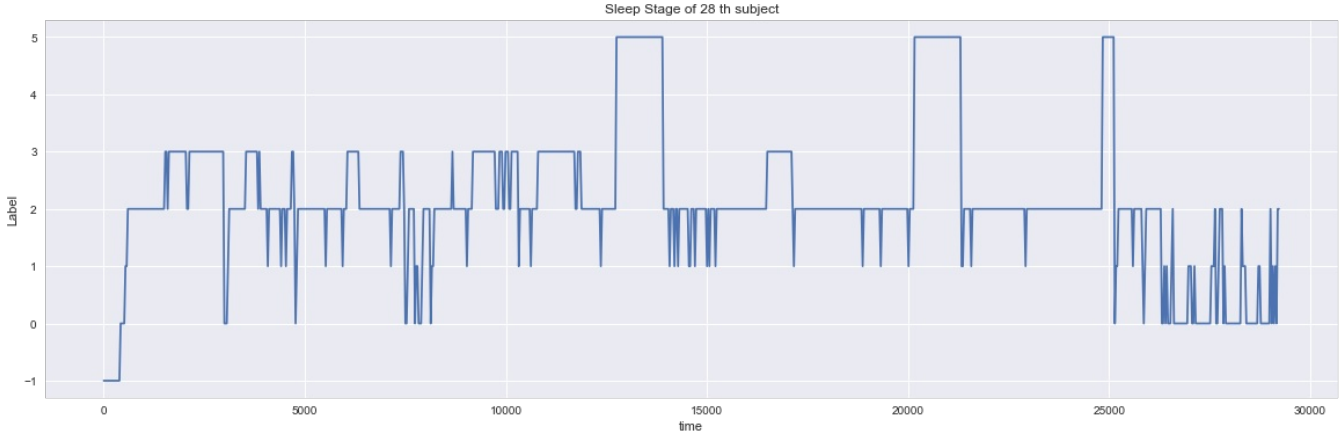

Sleep Stage of 23 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

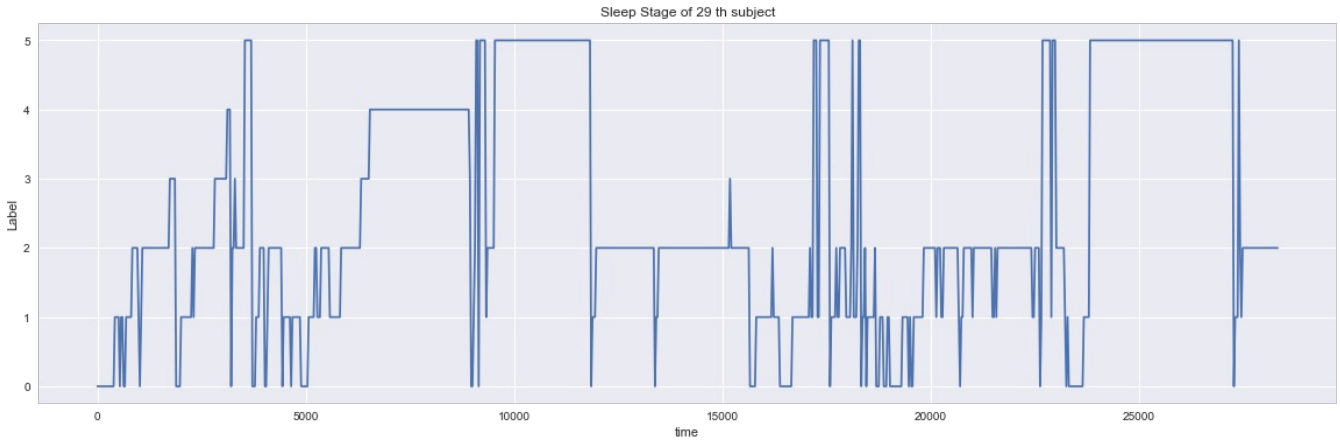

Sleep Stage of 24 th subject
Label
time
5
4
3
2
1
0
0
5000
10000
15000
20000
25000
30000

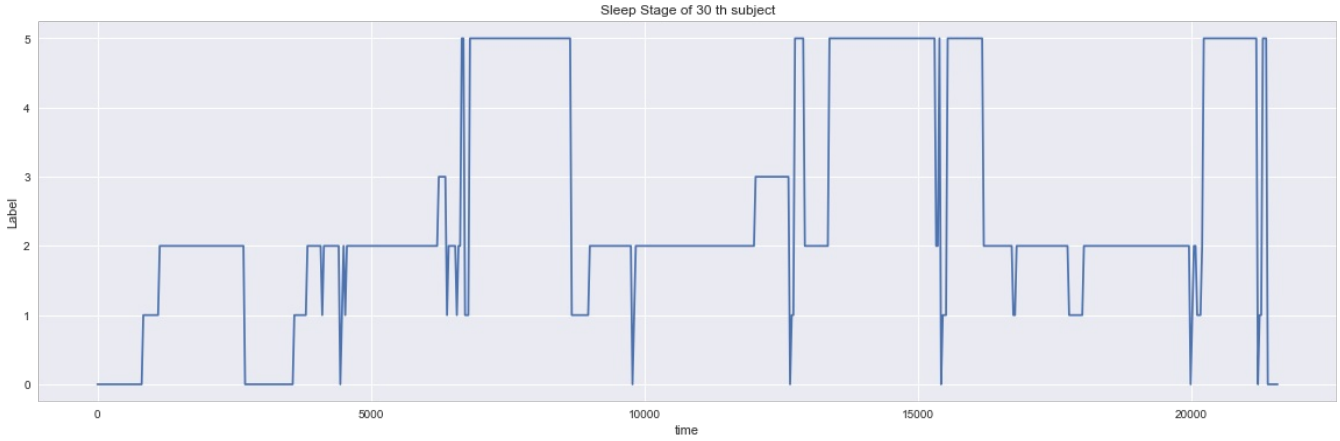

Sleep Stage of 25 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000


Sleep Stage of 26 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

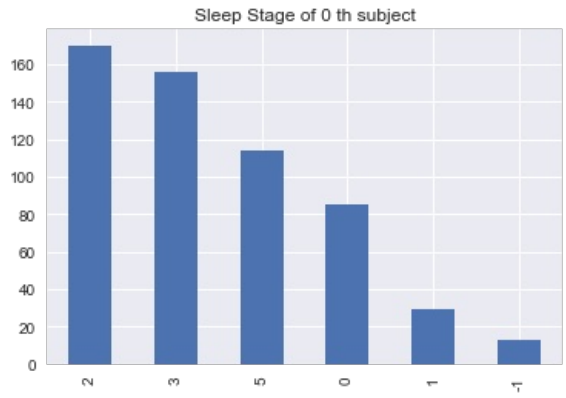

Sleep Stage of 27 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

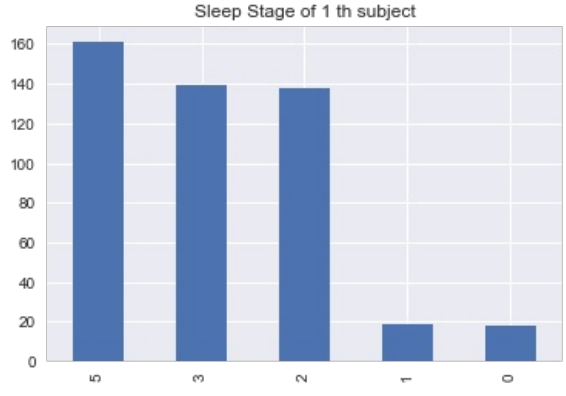

Sleep Stage of 28 th subject
Label
time
5
4
3
2
1
0
−1
0
5000
10000
15000
20000
25000
30000

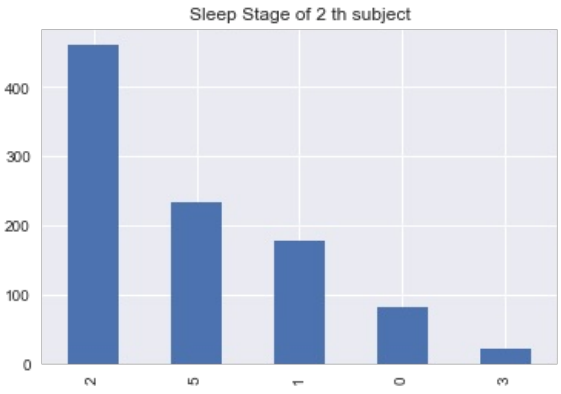


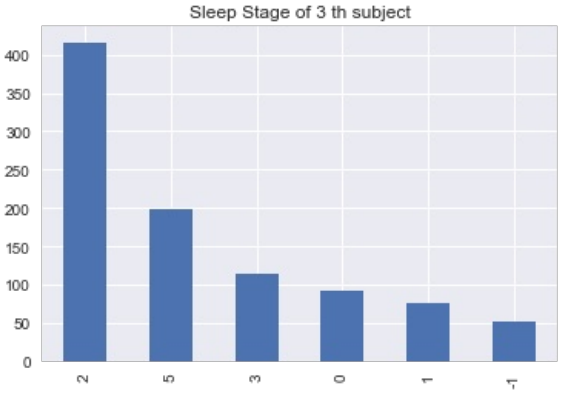


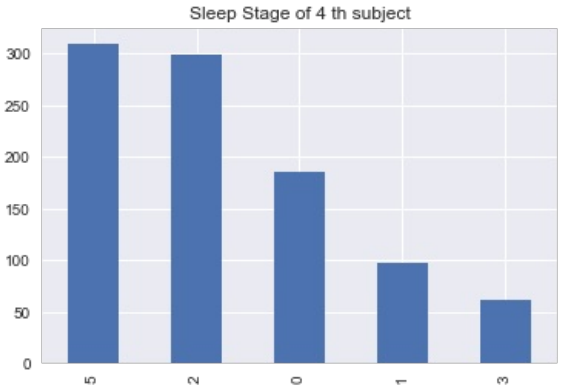
In [18]:
```
# 각 수면 단계가 얼마나 있는지 확인
# 17, 24번 피험자의 수면 라벨링에 설명에 나와있지 않은 4가 있으므로 수면 분류에 제외
for i in range(len(label)):
    plt.title(f'Sleep Stage of {i} th subject')
    label[i]['label'].value_counts().plot.bar()
    plt.show()
```

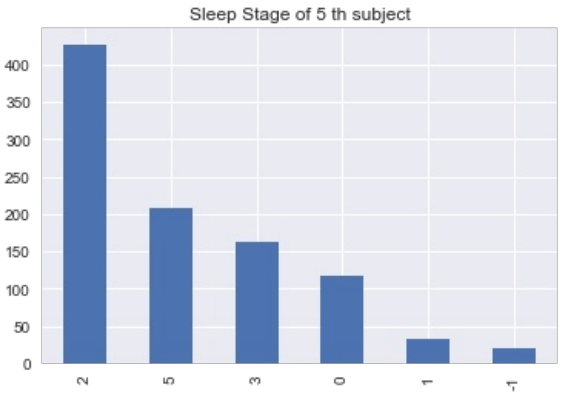


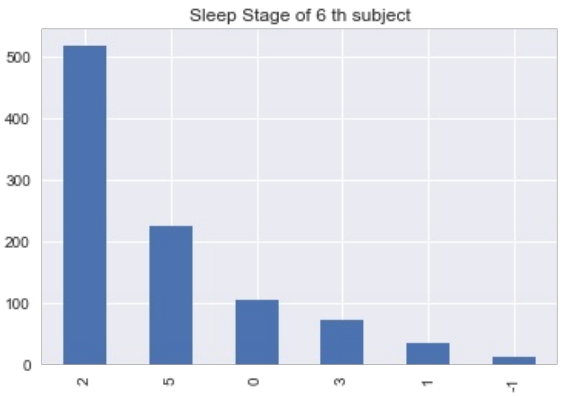


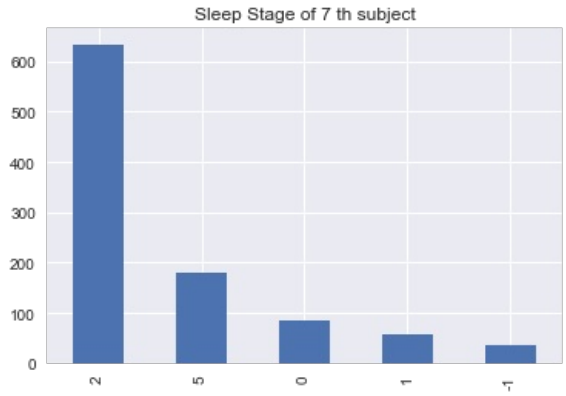

Sleep Stage of 2 th subject
400
300
200
100
0
2
5
1
0
3

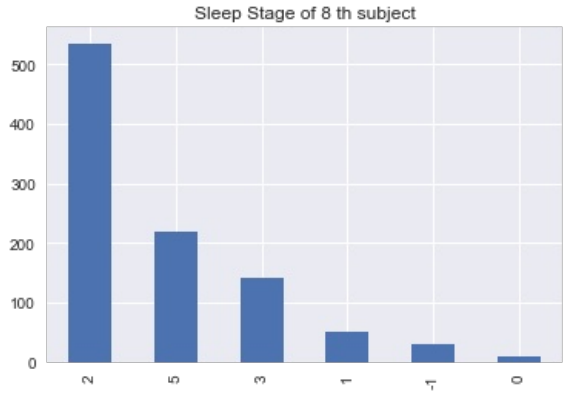

Sleep Stage of 3 th subject
400
350
300
250
200
150
100
50
0
2
5
3
0
1
-1

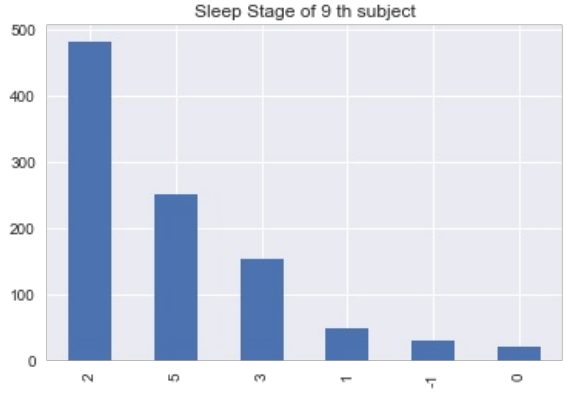

Sleep Stage of 4 th subject
300
250
200
150
100
50
0
5
2
0
1
3

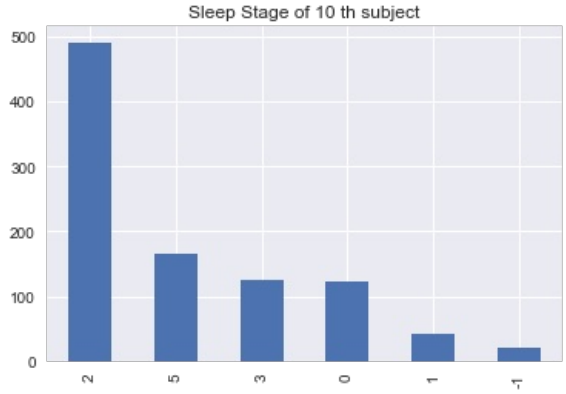

Sleep Stage of 5 th subject
400
350
300
250
200
150
100
50
0
2
5
3
0
1
-1

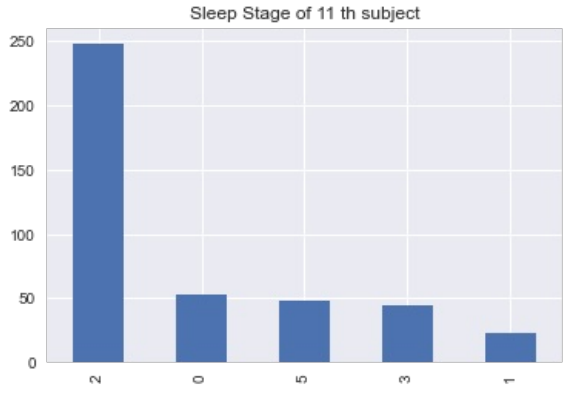

Sleep Stage of 6 th subject
500
400
300
200
100
0
2
5
0
3
1
-1

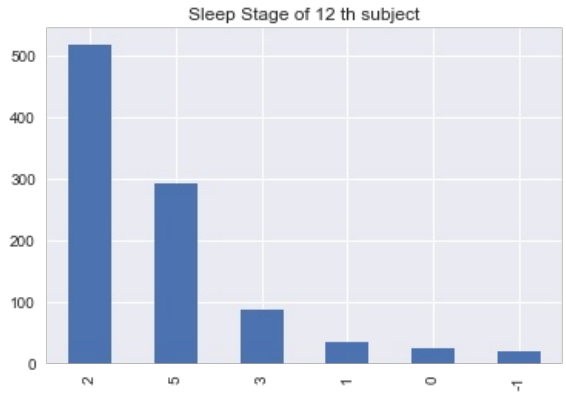

Sleep Stage of 7 th subject
600
500
400
300
200
100
0
2
5
0
1
-1

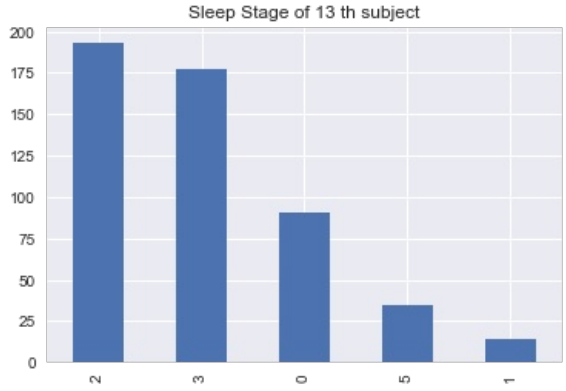

Sleep Stage of 8 th subject
500
400
300
200
100
0
2
5
3
1
-1
0

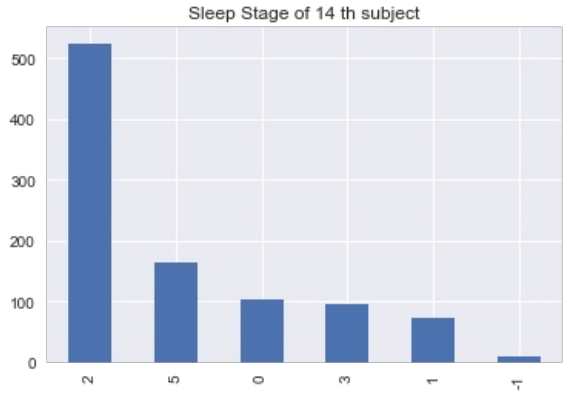

Sleep Stage of 9 th subject
500
400
300
200
100
0
2
5
3
1
-1
0

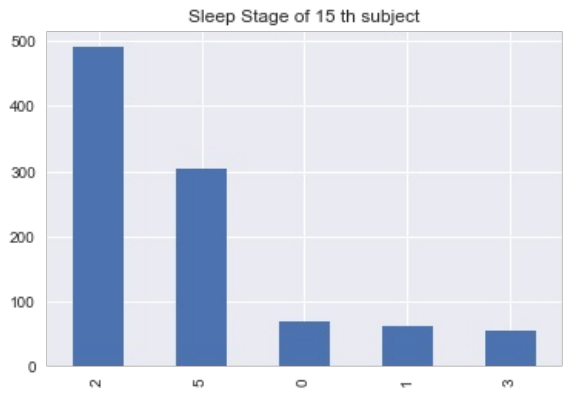

Sleep Stage of 10 th subject
500
400
300
200
100
0
2
5
3
0
1
-1

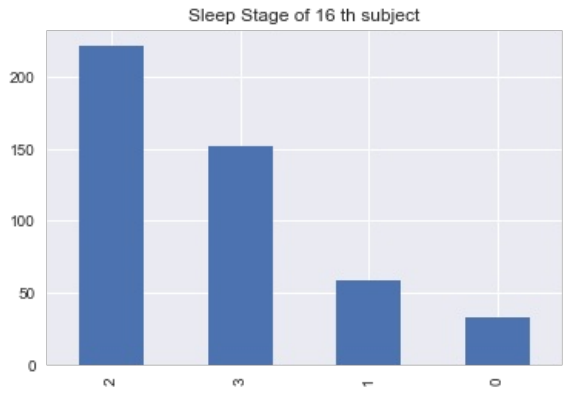

Sleep Stage of 11 th subject
250
200
150
100
50
0
2
0
5
3
1

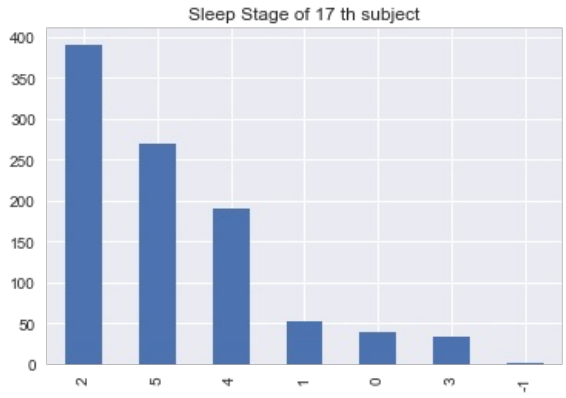

Sleep Stage of 12 th subject
500
400
300
200
100
0
2
5
3
1
0
-1

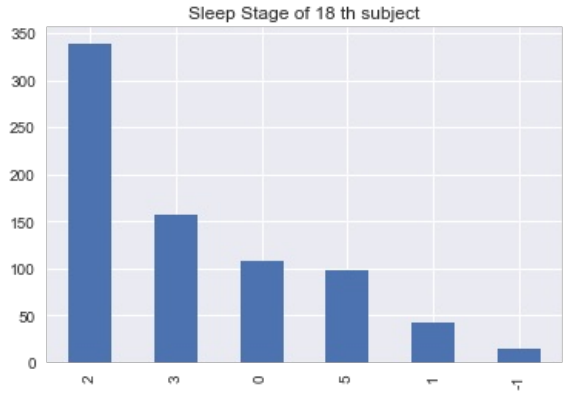

Sleep Stage of 13 th subject
200
175
150
125
100
75
50
25
0
2
3
0
5
1

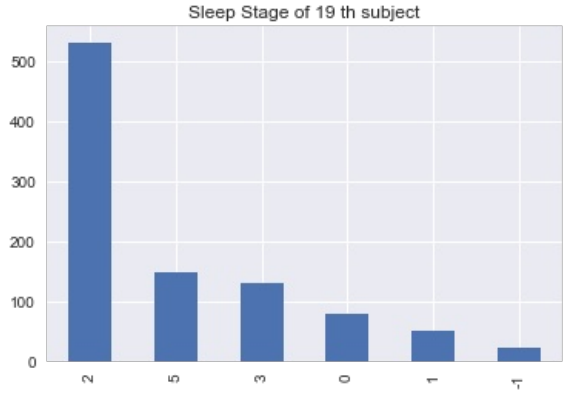

Sleep Stage of 14 th subject
500
400
300
200
100
0
2
5
0
3
1
-1

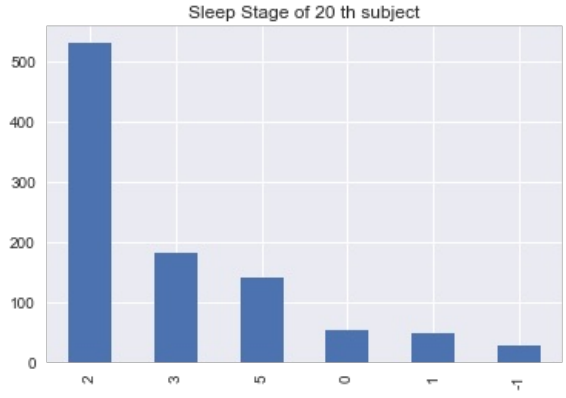

Sleep Stage of 15 th subject
500
400
300
200
100
0
2
5
0
1
3

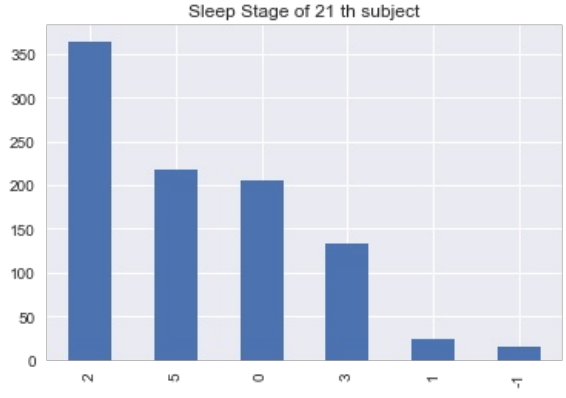

Sleep Stage of 16 th subject
200
150
100
50
0
2
3
1
0

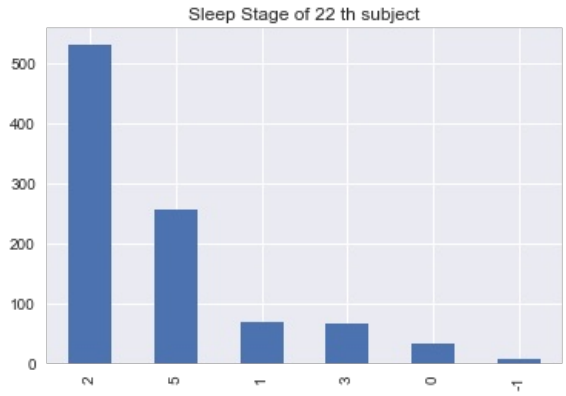

Sleep Stage of 17 th subject
400
350
300
250
200
150
100
50
0
2
5
4
1
0
3
-1

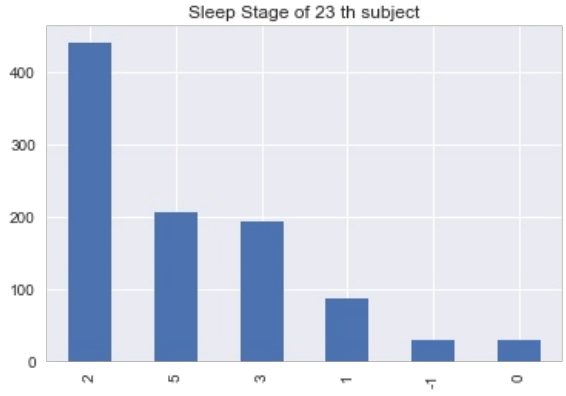

Sleep Stage of 18 th subject
350
300
250
200
150
100
50
0
2
3
0
5
1
-1

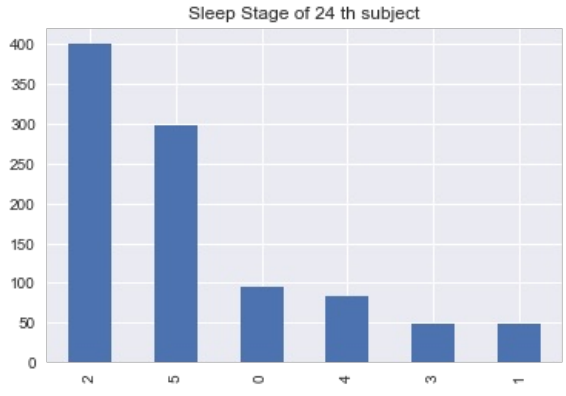

Sleep Stage of 19 th subject
500
400
300
200
100
0
2
5
3
0
1
-1

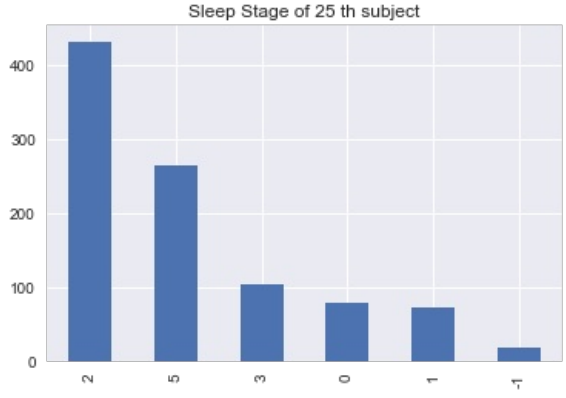

Sleep Stage of 20 th subject
500
400
300
200
100
0
2
3
5
0
1
-1

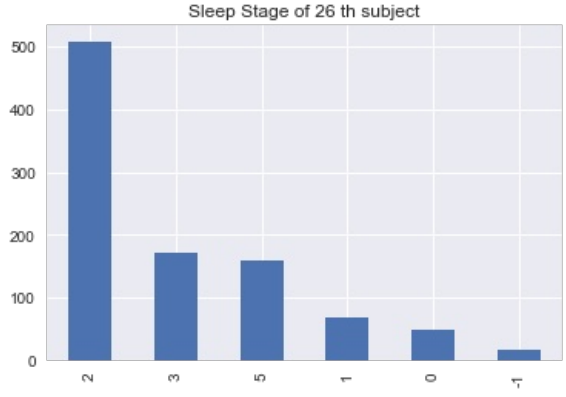

Sleep Stage of 21 th subject
350
300
250
200
150
100
50
0
2
5
0
3
1
-1


Sleep Stage of 22 th subject
500
400
300
200
100
0
2
5
1
3
0
-1


Sleep Stage of 23 th subject
400
300
200
100
0
2
5
3
1
-1
0


Sleep Stage of 24 th subject
400
350
300
250
200
150
100
50
0
2
5
0
4
3
1


Sleep Stage of 25 th subject
400
300
200
100
0
2
5
3
0
1
-1


Sleep Stage of 26 th subject
500
400
300
200
100
0
2
3
5
1
0
-1