matplotlib Version 3.1.0

Seaborn

Matplotlib és Seaborn

Asztalos Áron Data Scientist

asztalosster@gmail.com

matplotlib Version 3.1.0

Seaborn

Matplotlib és Seaborn

Asztalos Áron Data Scientist

asztalosster@gmail.com

matplotlib Version 3.1.0

Seaborn

Miért fontos az adatvizualizáció?

Asztalos Áron Data Scientist

asztalosster@gmail.com

matplotlib Version 3.1.0

Seaborn

Miért fontos az adatvizualizáció?

Kinek a dolga? (Külön szakma?)

Más vet és más arat?

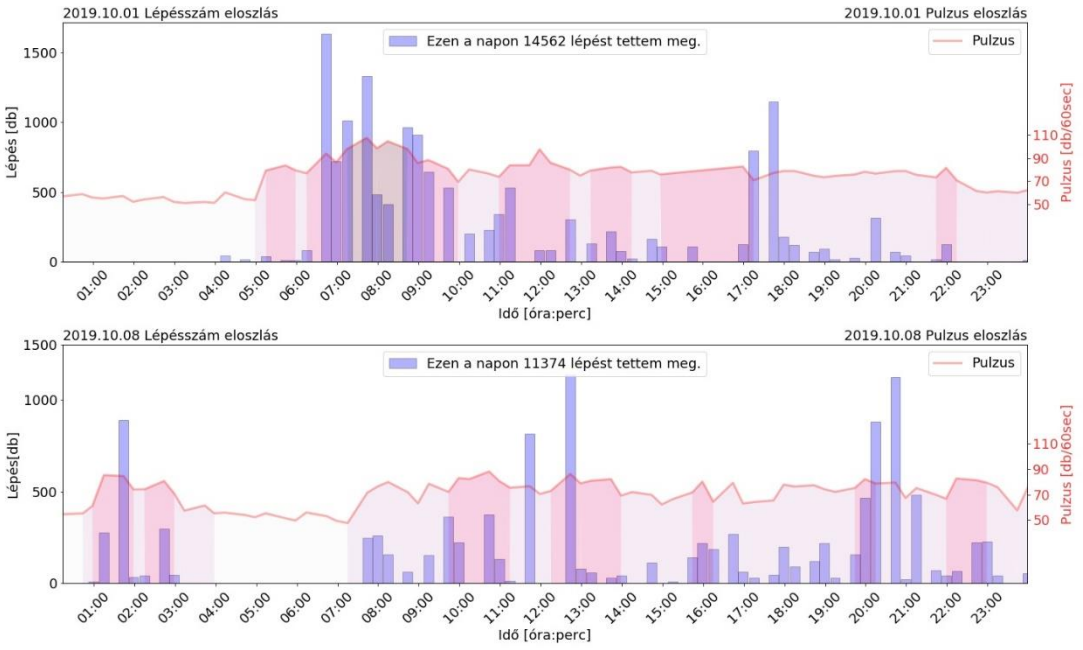
Ehhez tehetség kell?

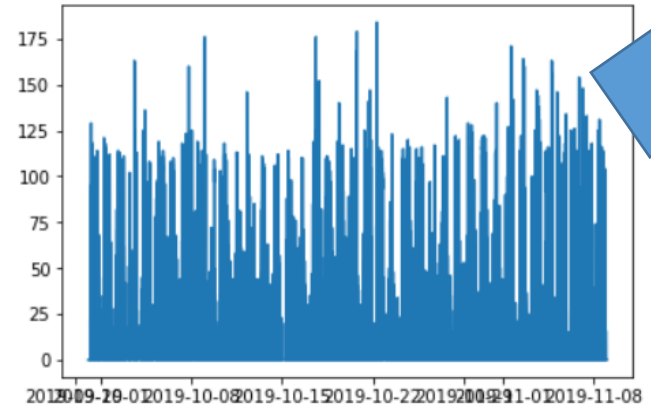
Asztalos Áron Data Scientist

asztalosster@gmail.com

matplotlib
Version 3.1.0

Alapok





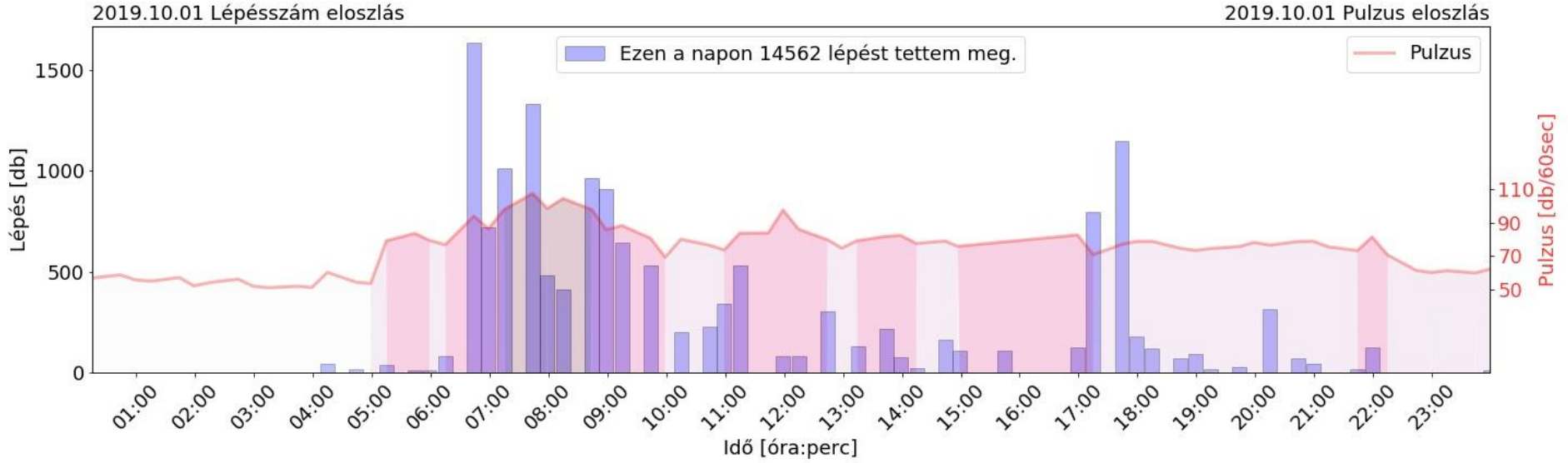
Tuti Paint!

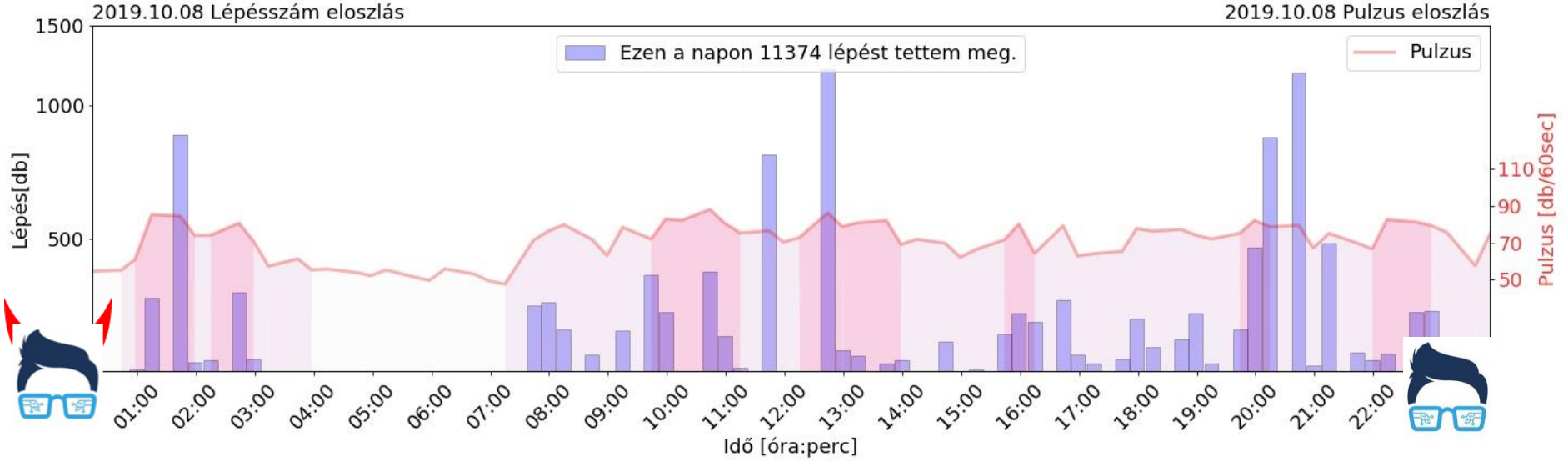
Deszép!

matplotlib
Version 3.1.0

Vajon milyen változás állt be az előadó életében?

Alapok

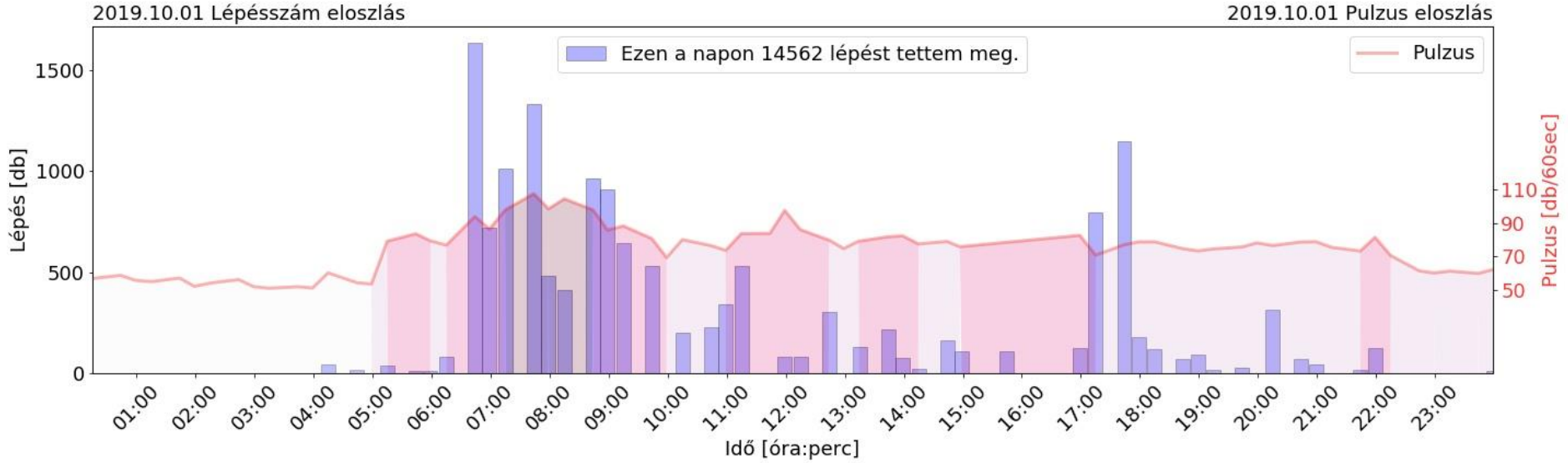


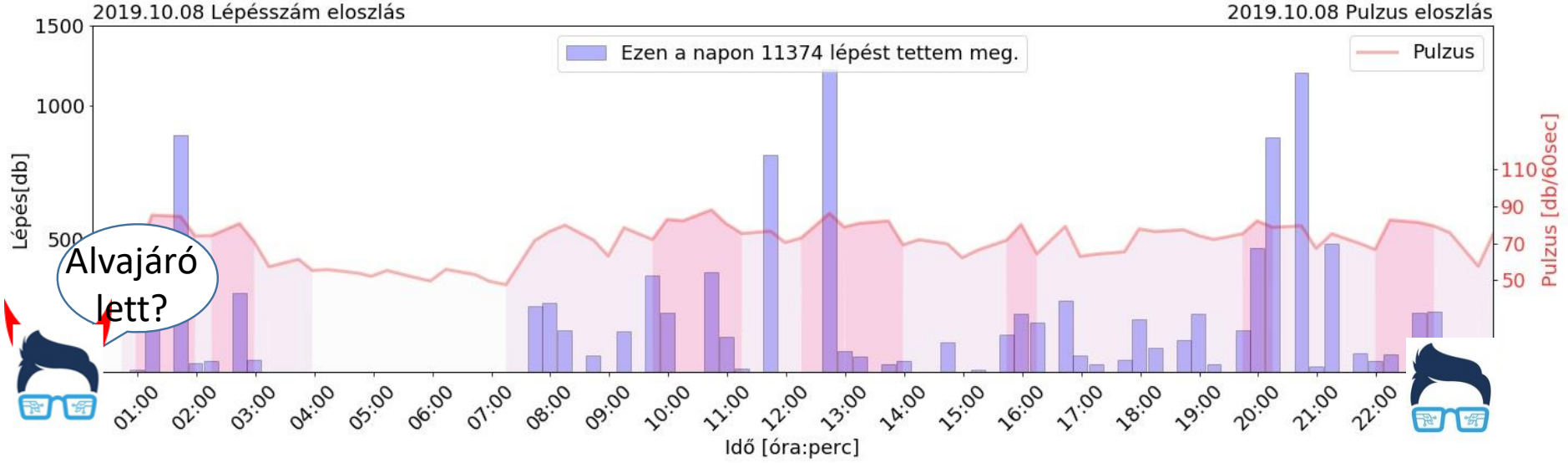


matplotlib
Version 3.1.0

Vajon milyen változás állt be az előadó életében?

Alapok

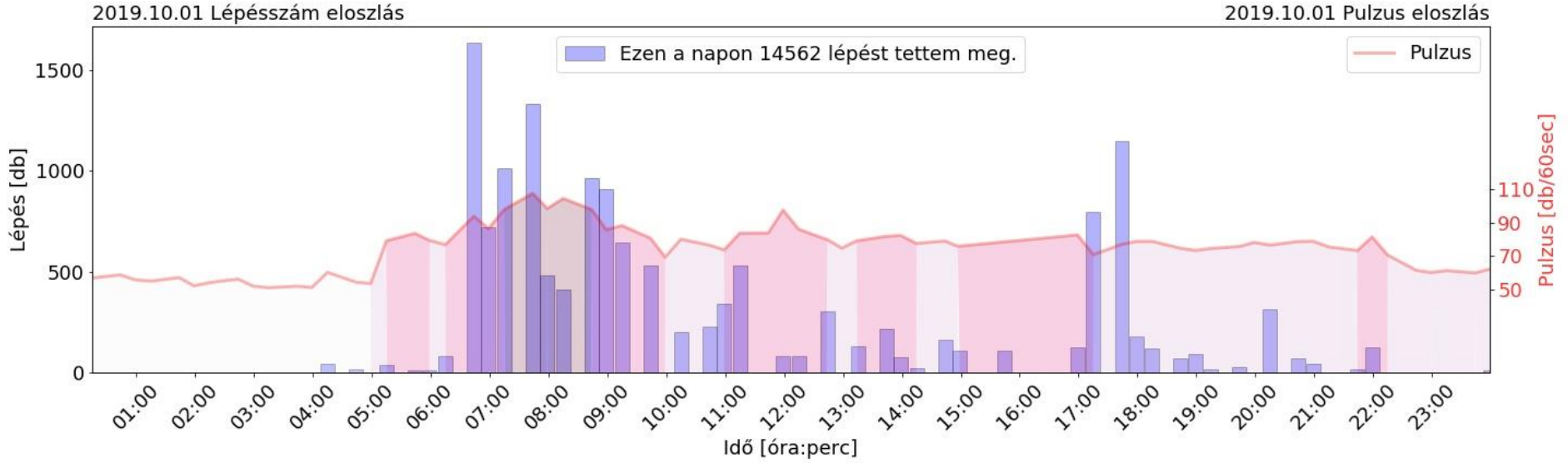


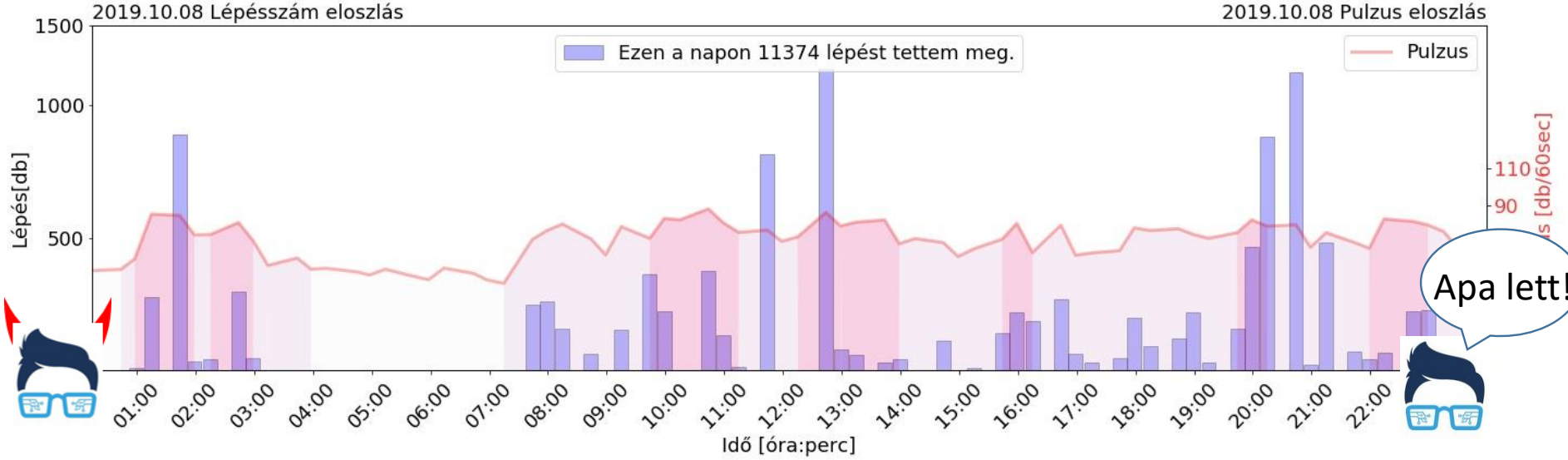


matplotlib
Version 3.1.0

Alapok

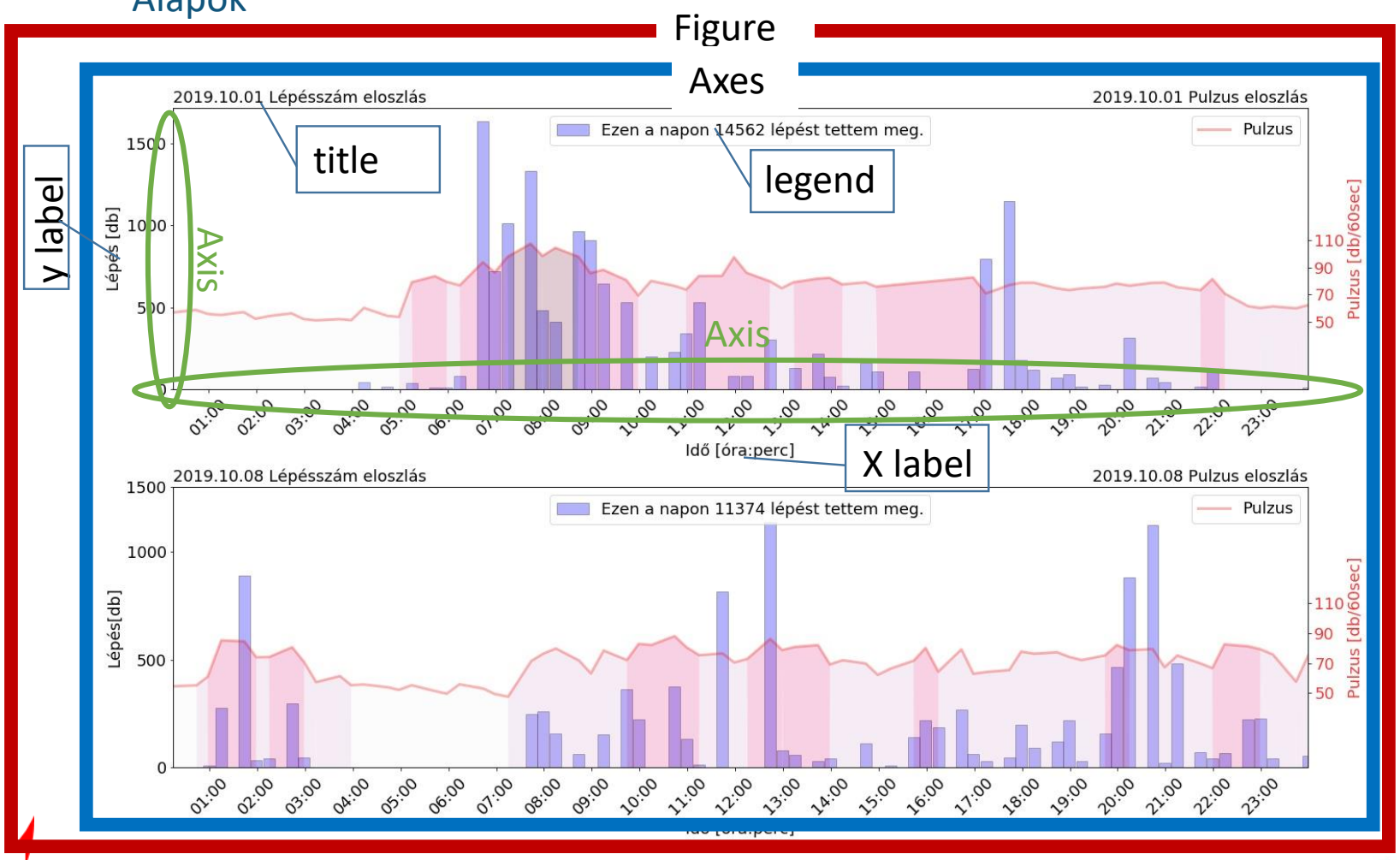
Vajon milyen változás állt be az előadó életében?





Az ábra anatómiája

Alapok

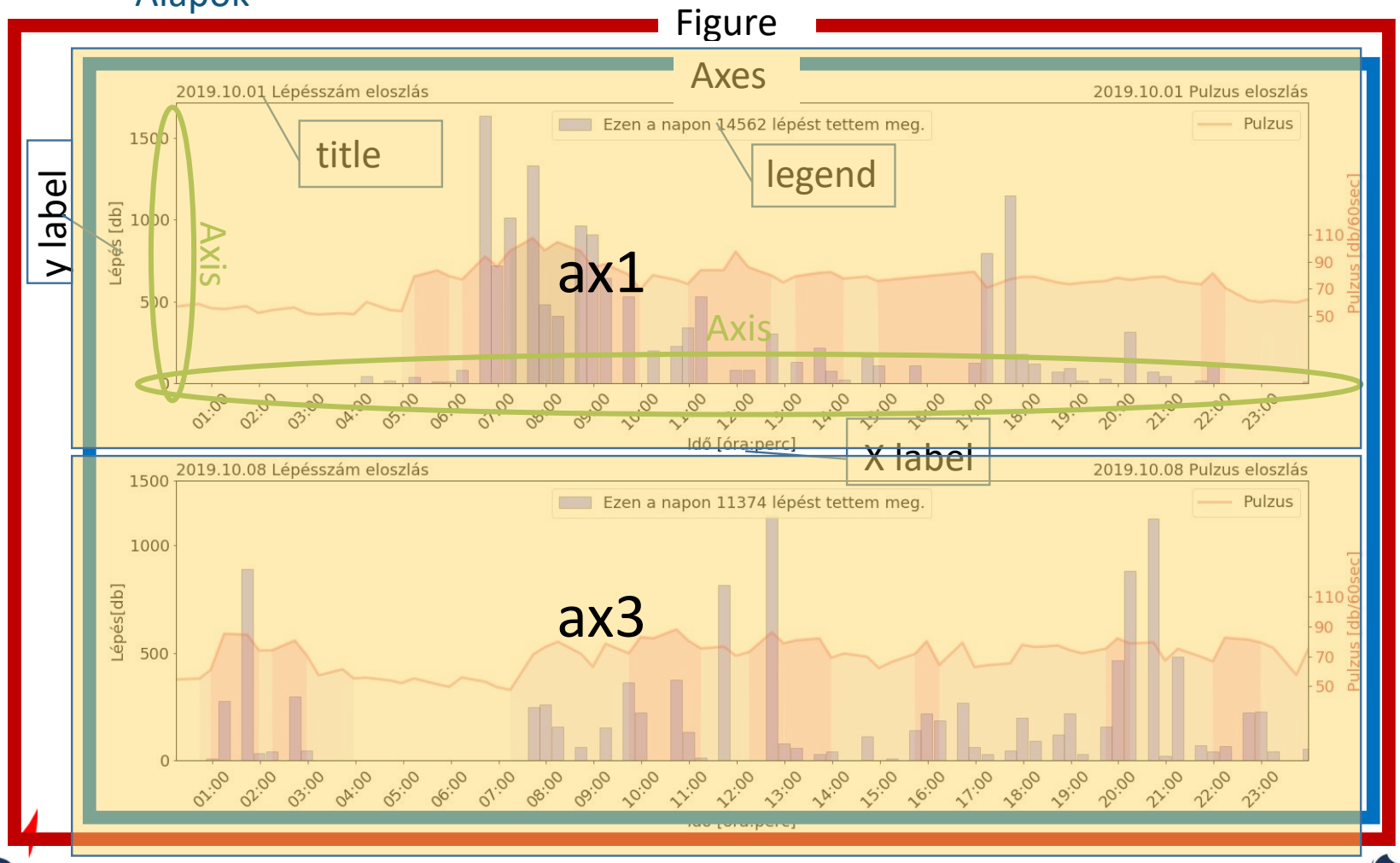


matplotlib
Version 3.1.0

Az ábra anatómiája

```
fig, (ax1,ax3) = plt.subplots(nrows=2, ncols=1, figsize=(20, 12))
```

Alapok

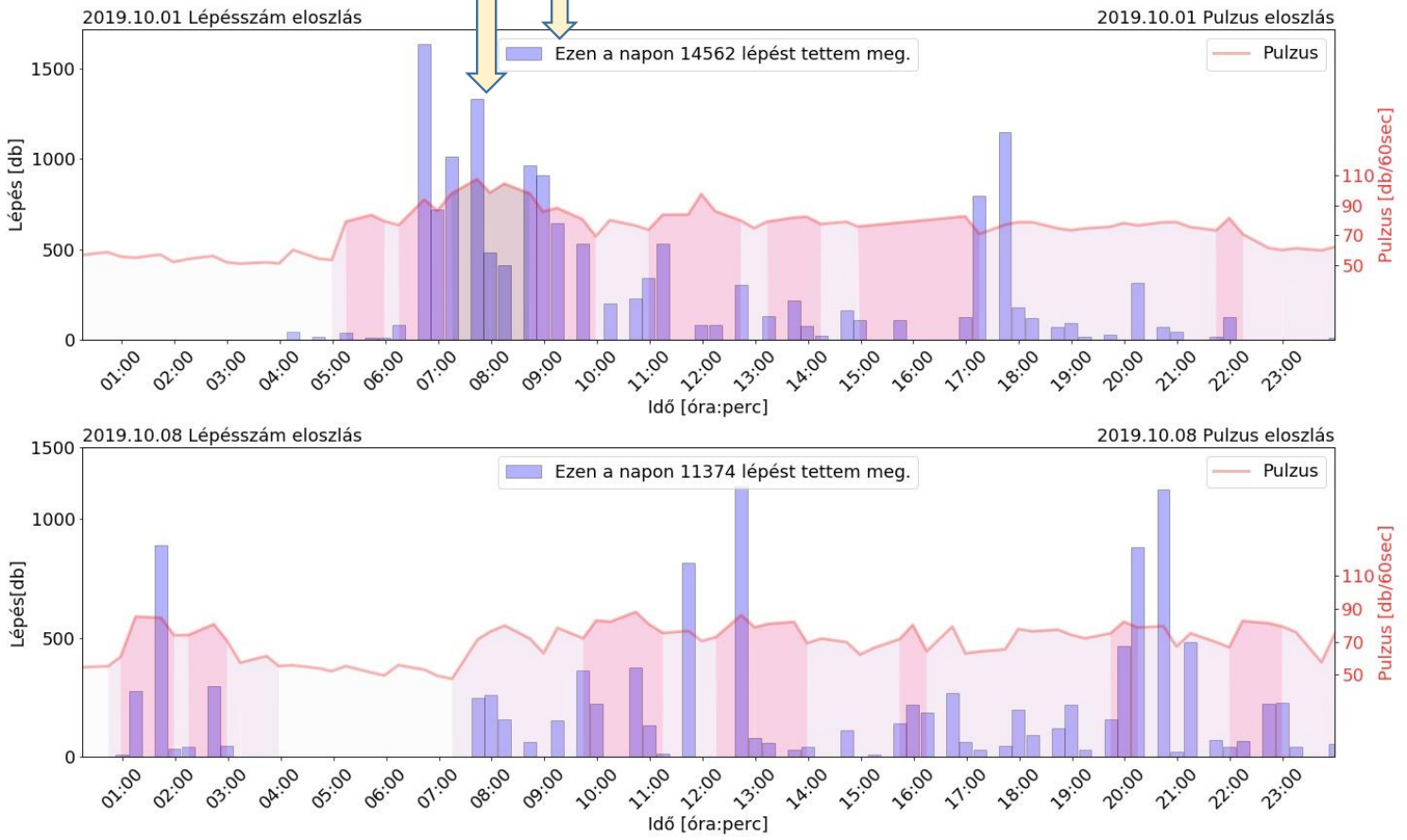


matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

```
ax1.bar(before_baby_steps_Tstamp_x1,before_baby_steps_y1,
    width=0.01,color = 'blue',edgecolor='black',alpha=0.3,
    label="Ezen a napon {} lépést tettem meg.".format(sum(before_baby_steps_y1)))
```

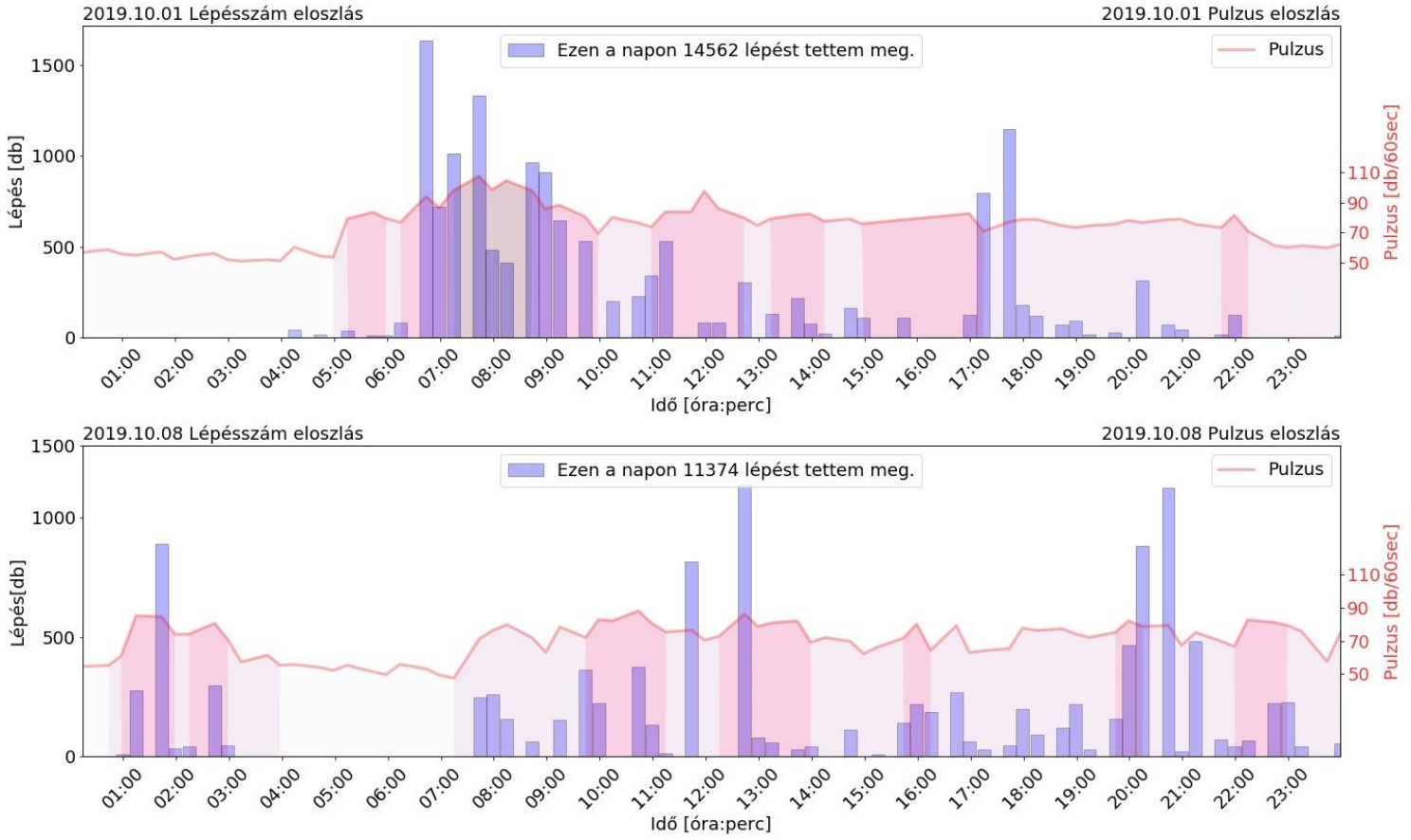


matplotlib
Version 3.1.0

Alapok

Az ábra anatómiája

```
ax1.set_title('2019.10.01 Lépésszám eloszlás',loc='left',fontsize = 18)
```

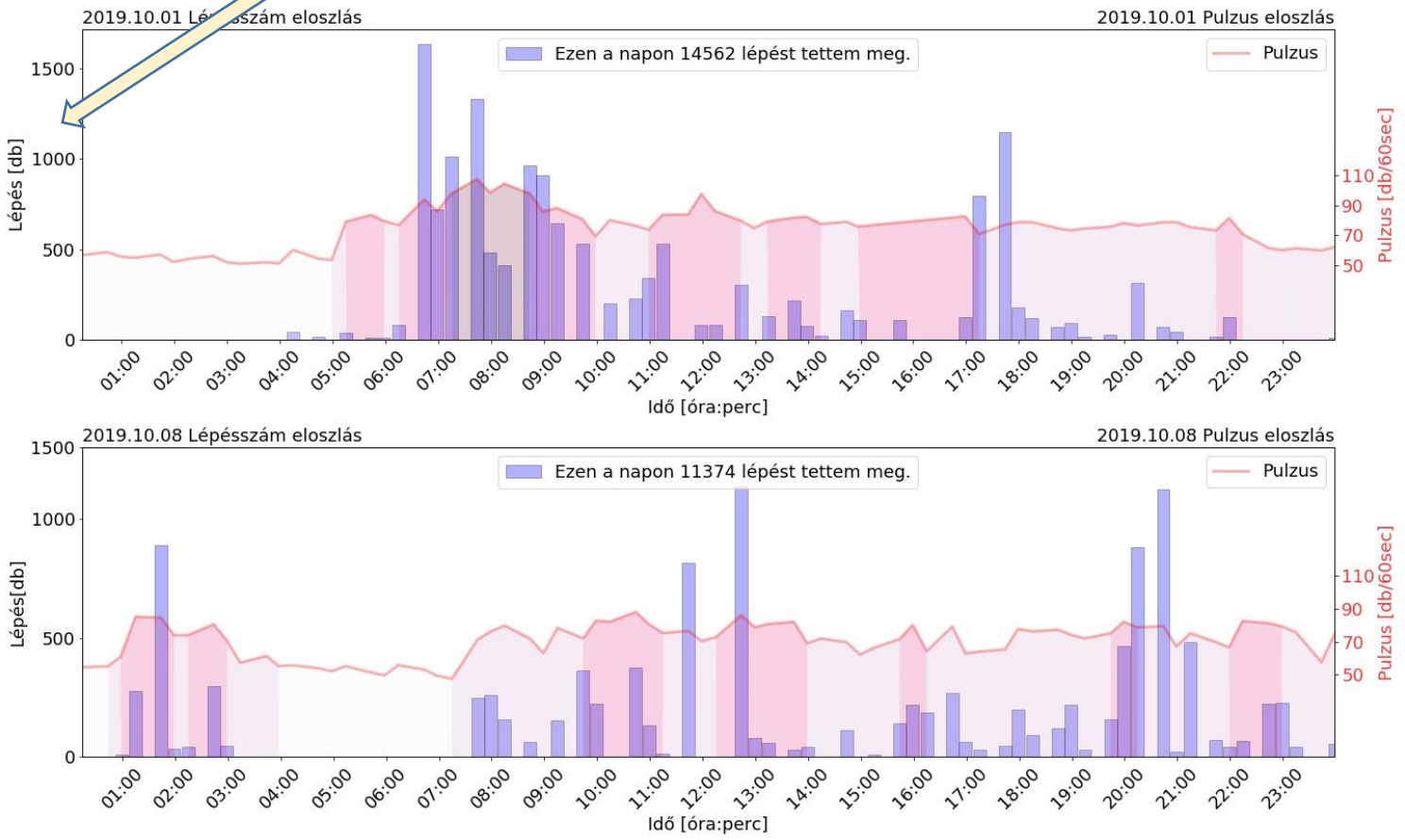


matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

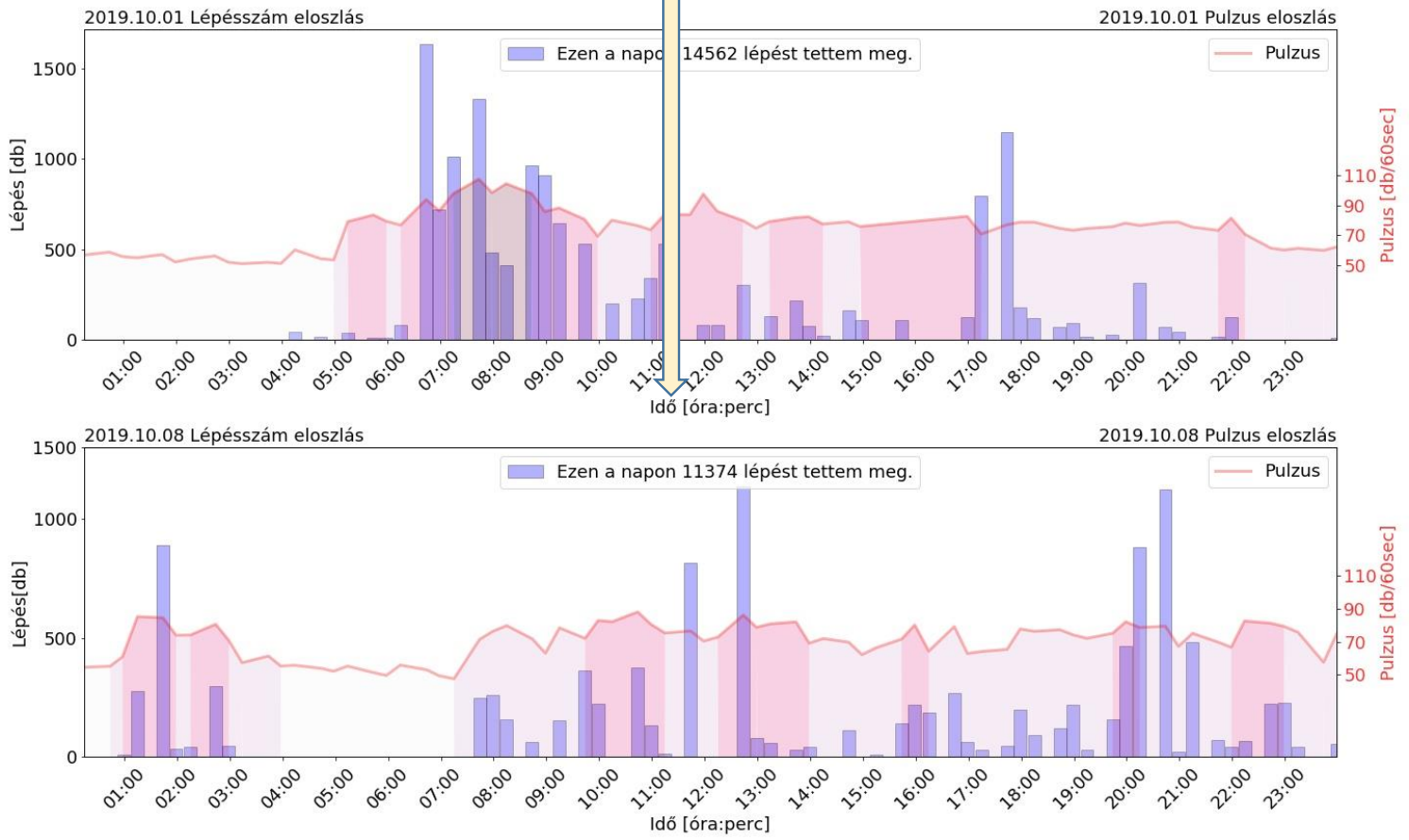
```
ax1.set_ylabel('Lépés [db]',fontsize = 18)
```



Az ábra anatómiája

Alapok

```
ax1.set_xlabel('Idő [óra:perc]',fontsize = 18)
```

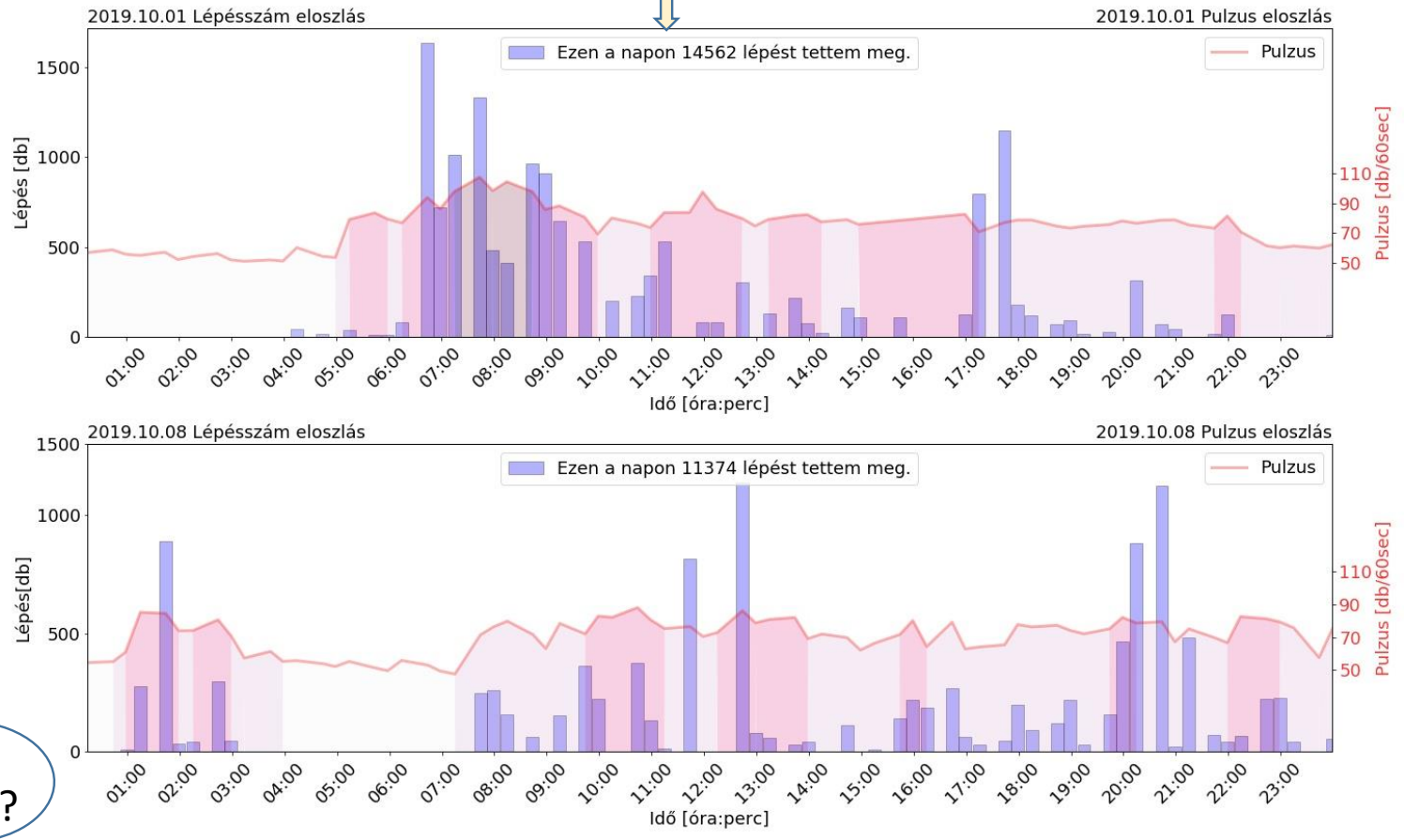


matplotlib
Version 3.1.0

Alapok

Az ábra anatómiája

```
ax1.legend(loc='upper center',fontsize = 18)
```



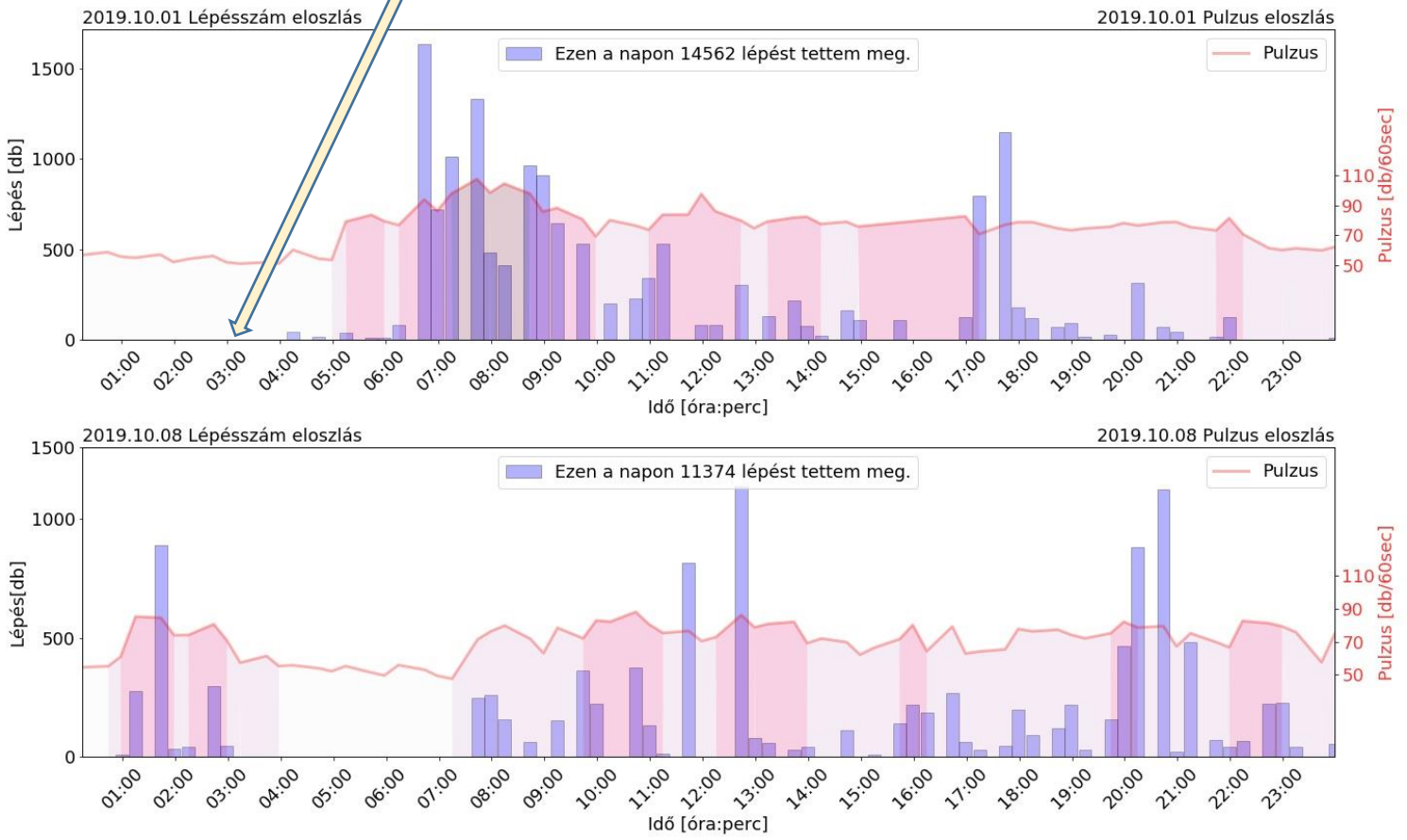
Van még?

matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

ax1.set_xticks(ticks=before_baby_xticks)

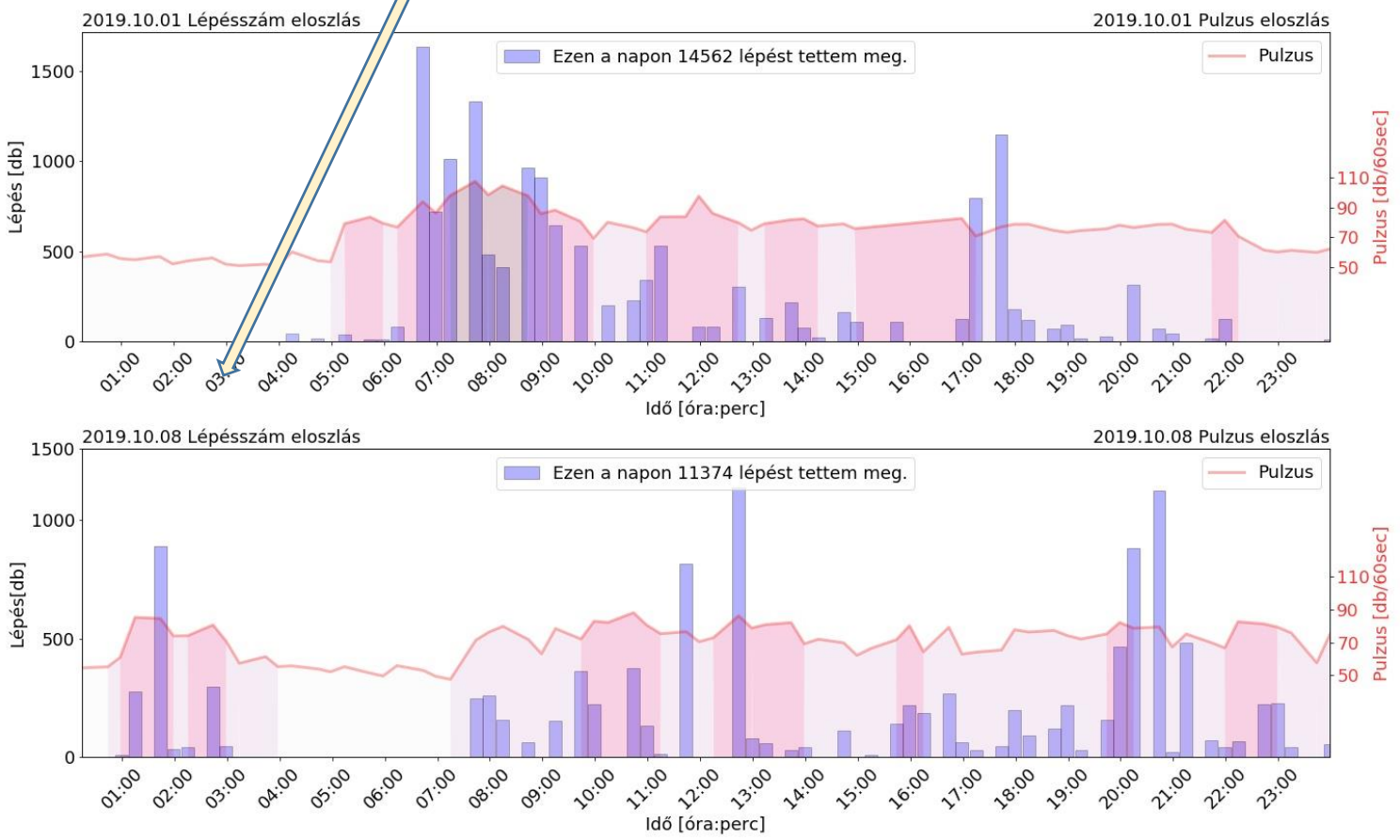


matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

```
ax1.set_xticklabels(before_baby_xtick_labels,fontsize = 18,rotation=45)
```

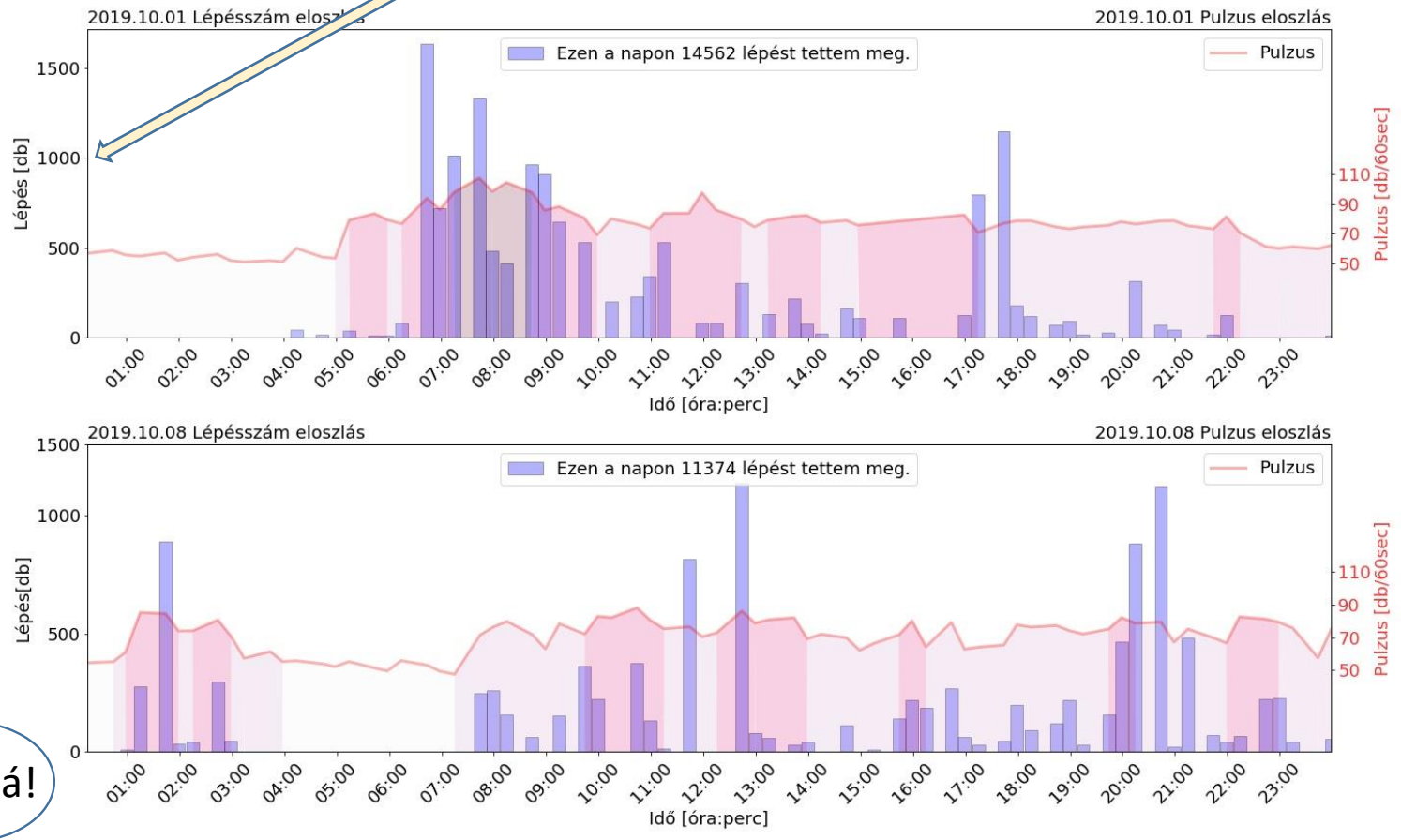


matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

```
ax1.set_yticks(ticks=before_baby_yticks)
```

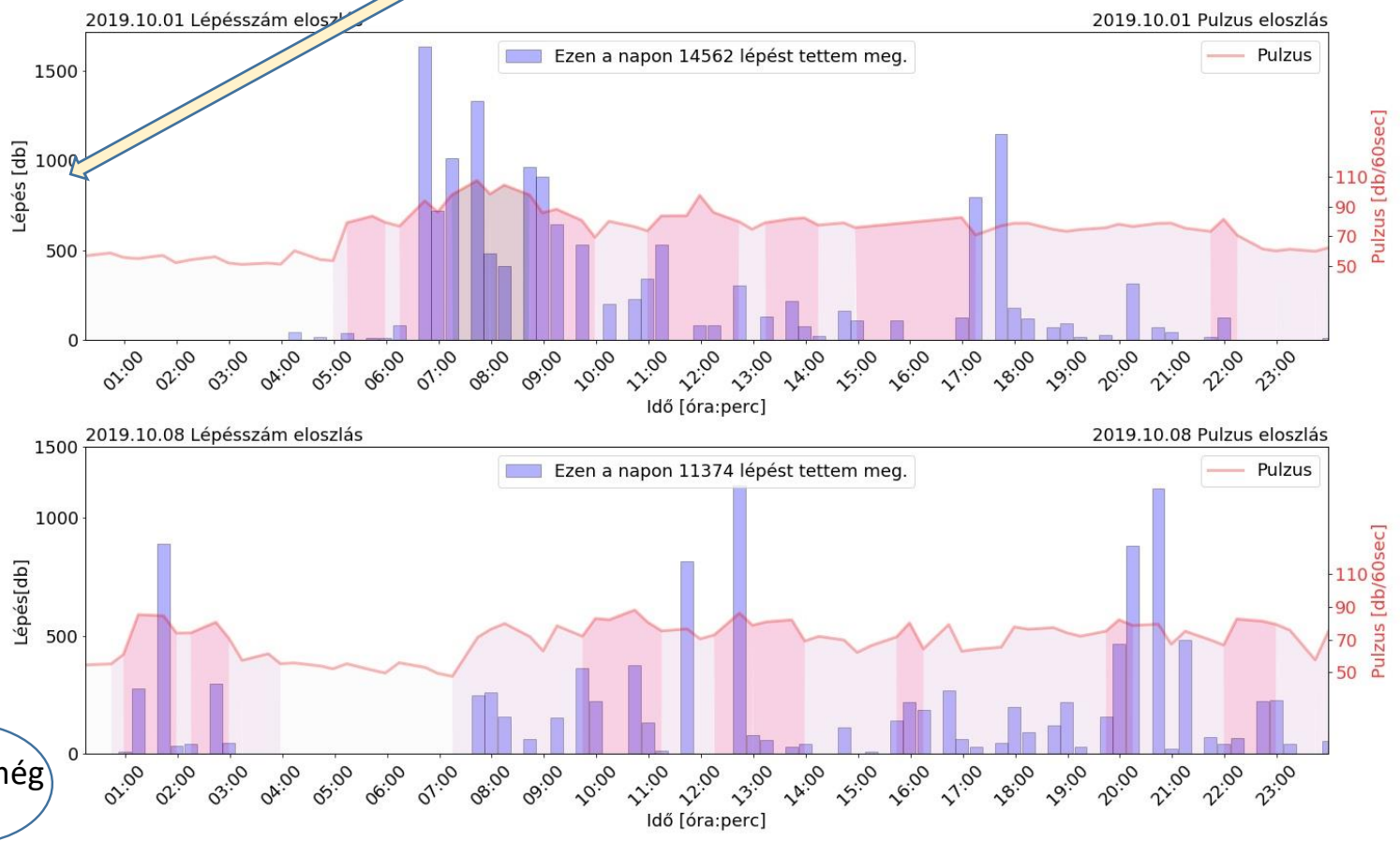


matplotlib
Version 3.1.0

Az ábra anatómiája

Alapok

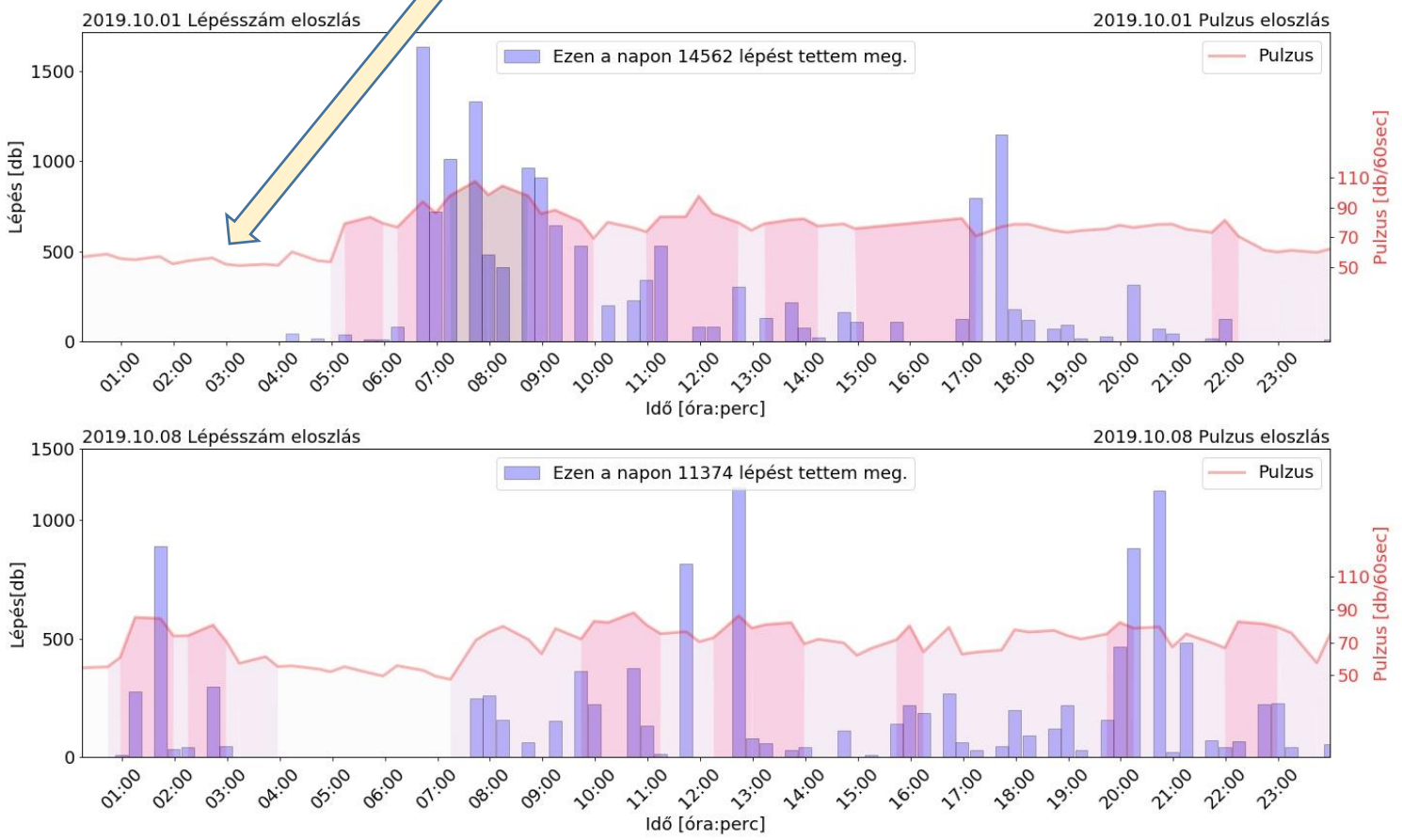
```
ax1.set_yticklabels(before_baby_ytick_labels,fontsize = 18,rotation=0)
```



matplotlib
Version 3.1.0

Alapok

```
ax2 = ax1.twinx()
"""Such axes are generated by calling the Axes.twinx method.
Likewise, Axes.twiny is available to generate axes that share a y axis but have different
top and bottom scales."""
ax2.plot(before_baby_heartrate_Tstamp_x1,before_baby_heartrate_hr_y1,color = 'red',
        label="Pulzus",alpha=0.3,linewidth=3)
```



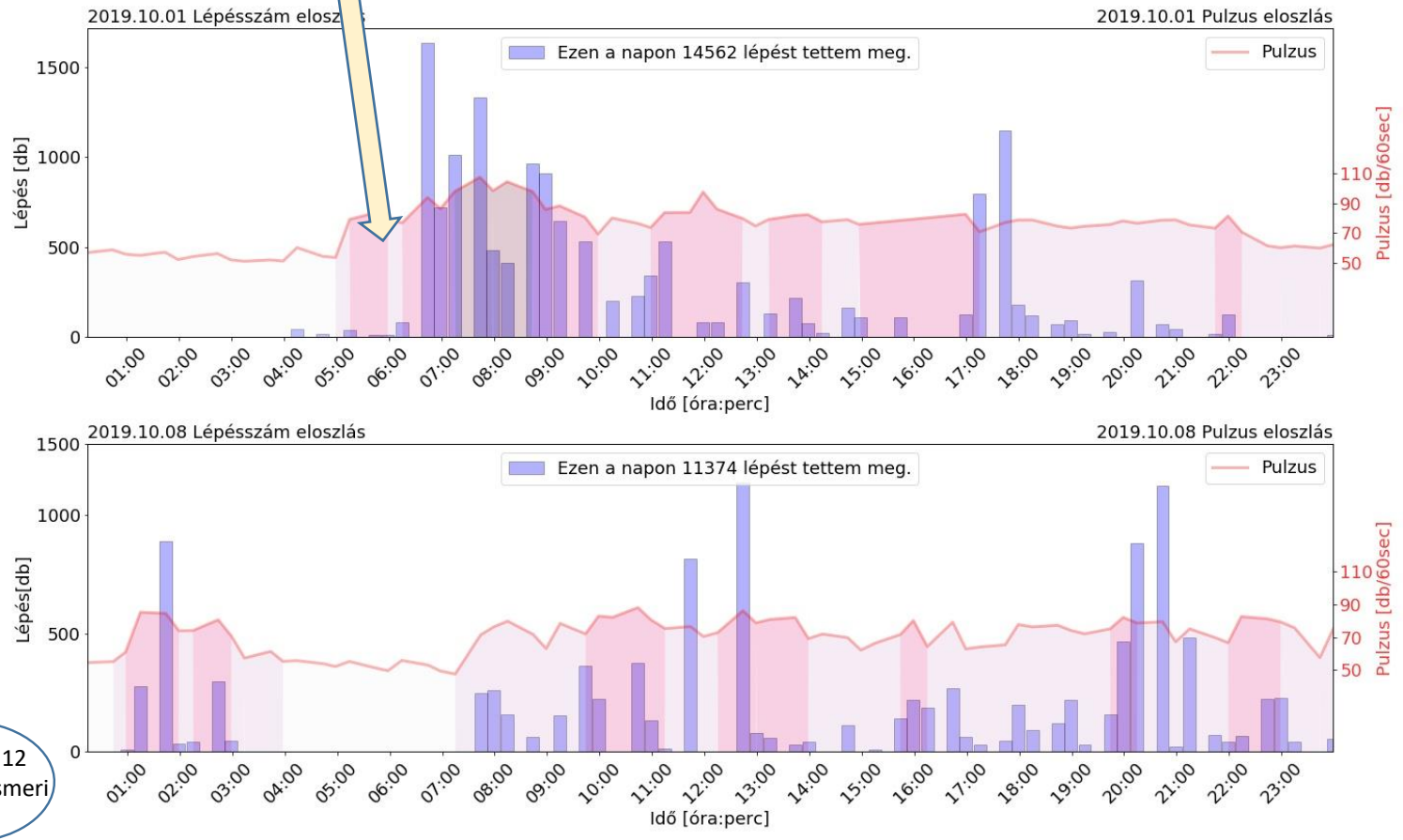
Trükkös rész!

Ebben az esetben az x tengelyek skálája megegyezik!

matplotlib

Alapok

```
state = pd.cut(pd.DataFrame(before_baby_heartrate_hr_y1).hr, bins=bins,labels=range(len(bins)-1))
cmap = plt.get_cmap('PuRd') #https://matplotlib.org/2.0.1/examples/color/colormaps_reference.html
for i,color in enumerate(cmap(np.linspace(start=0,stop=1,num = len(bins)-1))):
    ax2.fill_between(before_baby_heartrate_Tstamp_x1, before_baby_heartrate_hr_y1, -5000,
                     where=state==i,facecolor=color,interpolate=True,alpha=0.2)
```



Windows 12 alapszínét ismeri csak

-5000?

matplotlib Version 3.1.0

Alapok

Egy kép többet mond ezer szónál?

```
In [45]:  1  code_str = code_str.replace("."," ").repla
          2  print(len(code_str.split()))
```

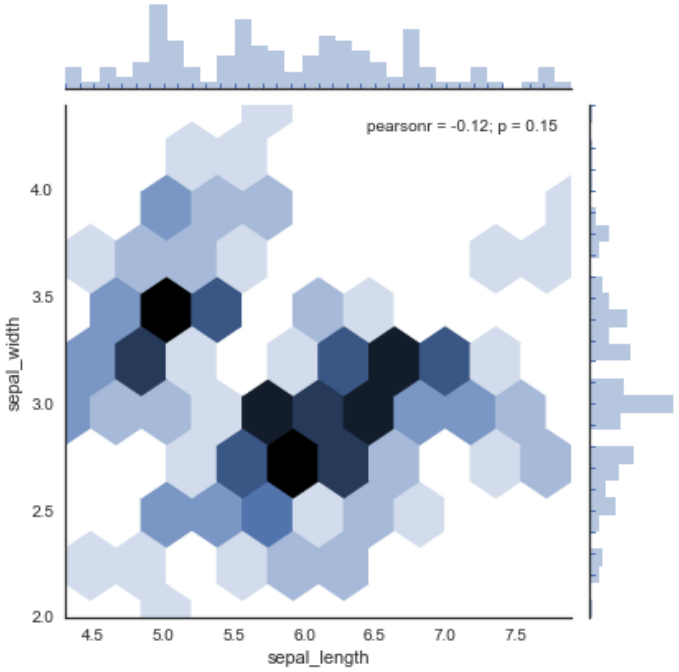
```
1600
```

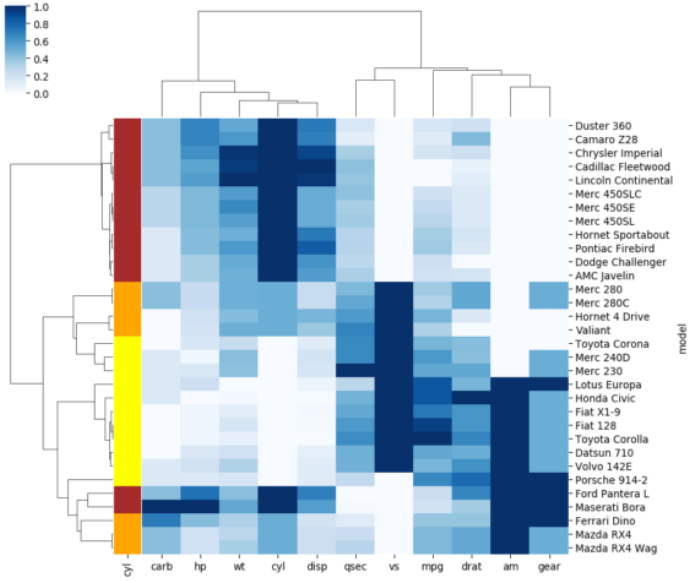
```
In [ ]:  1
```

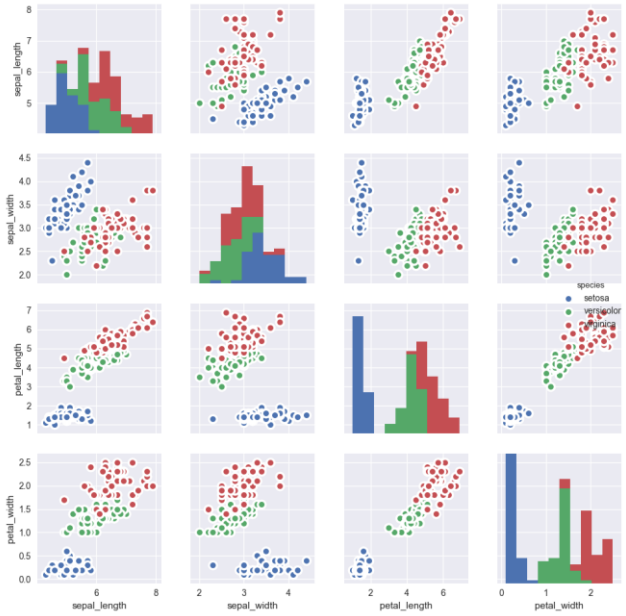
Seaborn

On Seaborn’s official website, they state:
If matplotlib “tries to make easy things easy and hard things possible”, seaborn tries to make a well-defined set of hard things easy too.

Seaborn


pearsonr = -0.12; p = 0.15
sepal_width
sepal_length
4.0
3.5
3.0
2.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5


1.0
0.8
0.6
0.4
0.2
0.0
Duster 360
Camaro Z28
Chrysler Imperial
Cadillac Fleetwood
Lincoln Continental
Merc 450SLC
Merc 450SE
Merc 450SL
Hornet Sportabout
Pontiac Firebird
Dodge Challenger
AMC Javelin
Merc 280
Merc 280C
Hornet 4 Drive
Valiant
Toyota Corona
Merc 240D
Merc 230
Lotus Europa
Honda Civic
Fiat X1-9
Fiat 128
Toyota Corolla
Datsun 710
Volvo 142E
Porsche 914-2
Ford Pantera L
Maserati Bora
Ferrari Dino
Mazda RX4
Mazda RX4 Wag
model
cyl
carb
hp
wt
cyl
disp
qsec
vs
mpg
drat
am
gear


sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica

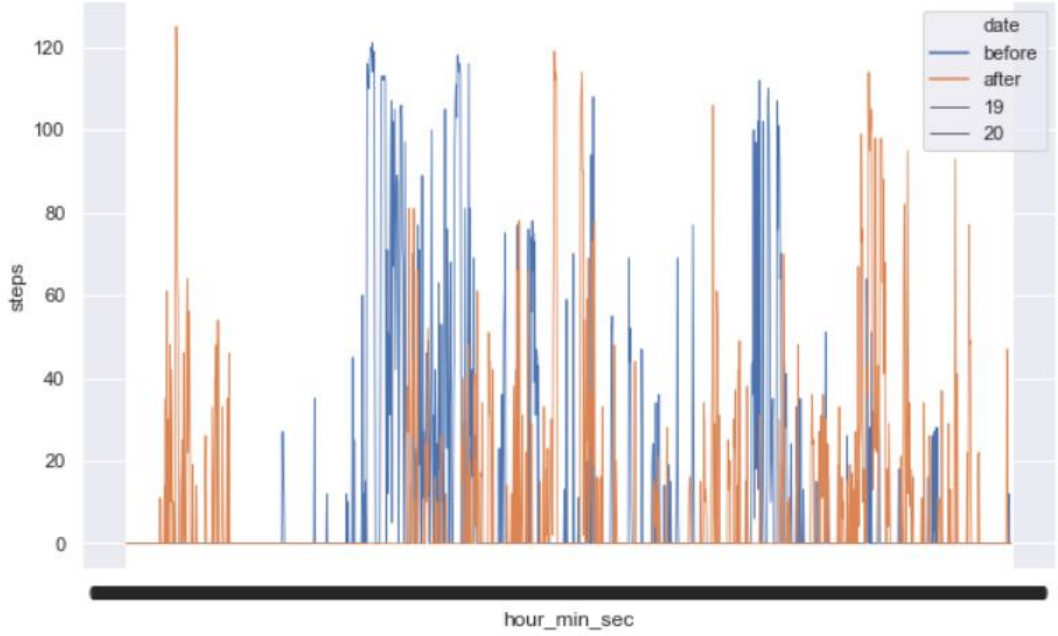

Excel!


Miben más?

Seaborn

```
In [77]:  1  plt.figure(figsize=(10,6))
          2  sns.lineplot(x="hour_min_sec", y="steps", hue="date", data=steps_sns_df,size=20)
```

```
Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x23b29daa898>
```



Hasznos linkek

Matplotlib alapjai I. (realpython)
Matplotlib alapjai II. (towardsdatascience)

Seaborn tutorial I. (medium.com)
Seaborn tutorial II. (elitedatascience)

https://github.com/asztalosaron/BudapestBI2019_Matplotlib_Seaborn

Asztalos Áron Data Scientist
asztalosster@gmail.com

What would you like to show?

COMPARISON

- Among Items
  - Two Variables per Item
    - VARIABLE WIDTH COLUMN CHART
  - Many Categories
    - TABLE WITH EMBEDDED CHARTS
  - Few Categories
    - BAR CHART HORIZONTAL
    - BAR CHART VERTICAL
- Over Time
  - Cyclical Data
    - CIRCULAR AREA CHART
  - Non-Cyclical Data
    - LINE CHART
  - Single or Few Categories
    - BAR CHART VERTICAL
  - Many Categories
    - LINE CHART
- One Variable per Item

RELATIONSHIP

- Two Variables
  - SCATTER PLOT
- Three or more Variables
  - SCATTER PLOT BUBBLE SAZE

DISTRIBUTION

- Single Variable
  - Few Data Points
    - BAR HISTOGRAM
  - Many Data Points
    - LINE HISTOGRAM
  - Two Variables
    - SCATTER PLOT

COMPOSITION

- Changing Over Time
  - Few Periods
    - Only Relative Differences Matter
      - STACKED 100% BAR CHART
    - Relative and Absolute Differences Matter
      - STACKED BAR CHART
  - Many Periods
    - Only Relative Differences Matter
      - STACKED AREA 100% CHART
    - Relative and Absolute Differences Matter
      - STACKED AREA CHART
- Static
  - Simple Share of Total
    - PIE CHART
  - Accumulation or Subtraction to Total
    - WATERFALL CHART
  - Components of Components
    - STACKED 100% BAR CHART WITH SUBCOMPONENTS
  - Accumulation to total and absolute difference matters
    - TREE MAP