

# Barbiturates



innopolis  
UNIVERSITY

# Ideas

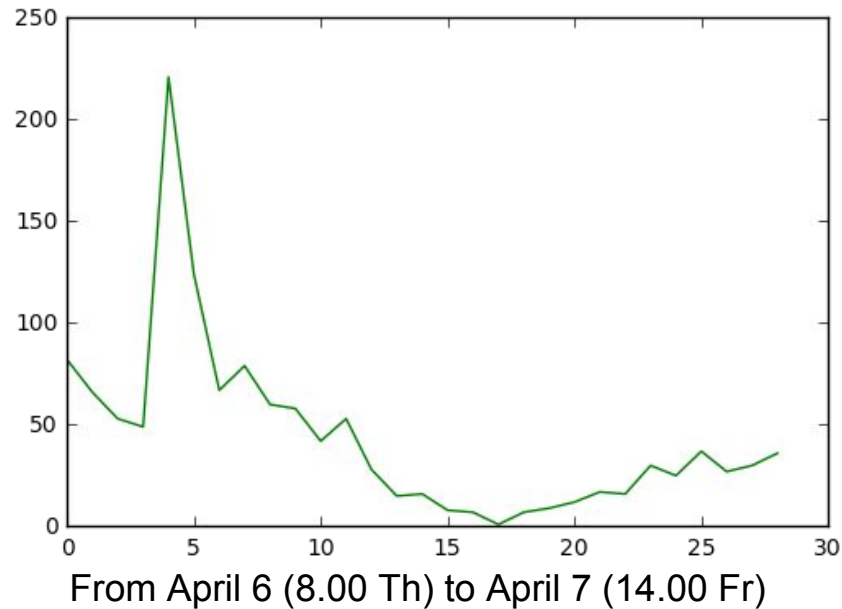
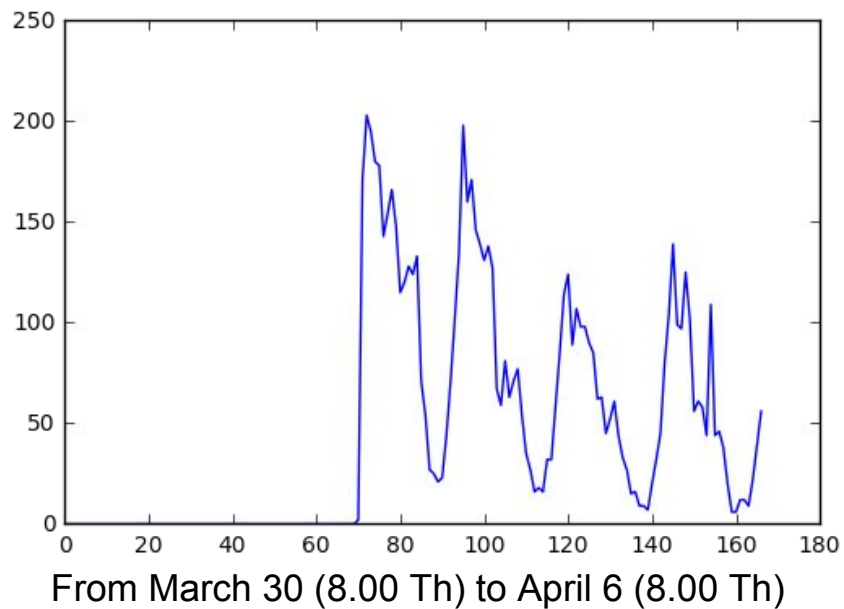
- Analyse time series distribution
- Clusterization by **users, tags, keywords**

Prediction =  $\alpha$  \* individ +  $(1 - \alpha)$  \* clustered

# Analysing time series

Take data from Thursday 8.00 am to a week LEFT

Frame: 1 hour



# Tags & NLP features

1. Choose best 25 tags
2. Encode them in a binary vector
3. Add it to our list of features
4. NLP text preprocessing:
  - a. Select 1000 most popular words from titles
  - b. Order by popularity and rename to numbers
  - c. word2vec

# Train

1. Extrapolation (SVR, polynomial, LinearRegression)
2. LSTM neural network
3. DecisionTreeRegressor