Search project

Nisheeth

# Browser-based personalization

- Export your own browser history from Chrome
- Instructions
  - Copy the sqllite3.exe file I've given you into your working folder
  - Copy the 'History' file from your local Chrome cache into the same folder
  - At the command prompt, enter the following line
  - sqlite History "select datetime(last_visit_time/1000000-11644473600,'unixepoch'),url from urls order by last_visit_time desc" > history_export.txt

- You should have your history exported to a text file in the same folder now.

Baseline code

- You'll find it zipped up on the webpage
- Finds URL → Domain mapping
- Processes domain names and meta-data as text to construct a document corpus
- Runs a topic model on unique documents seen in the corpus
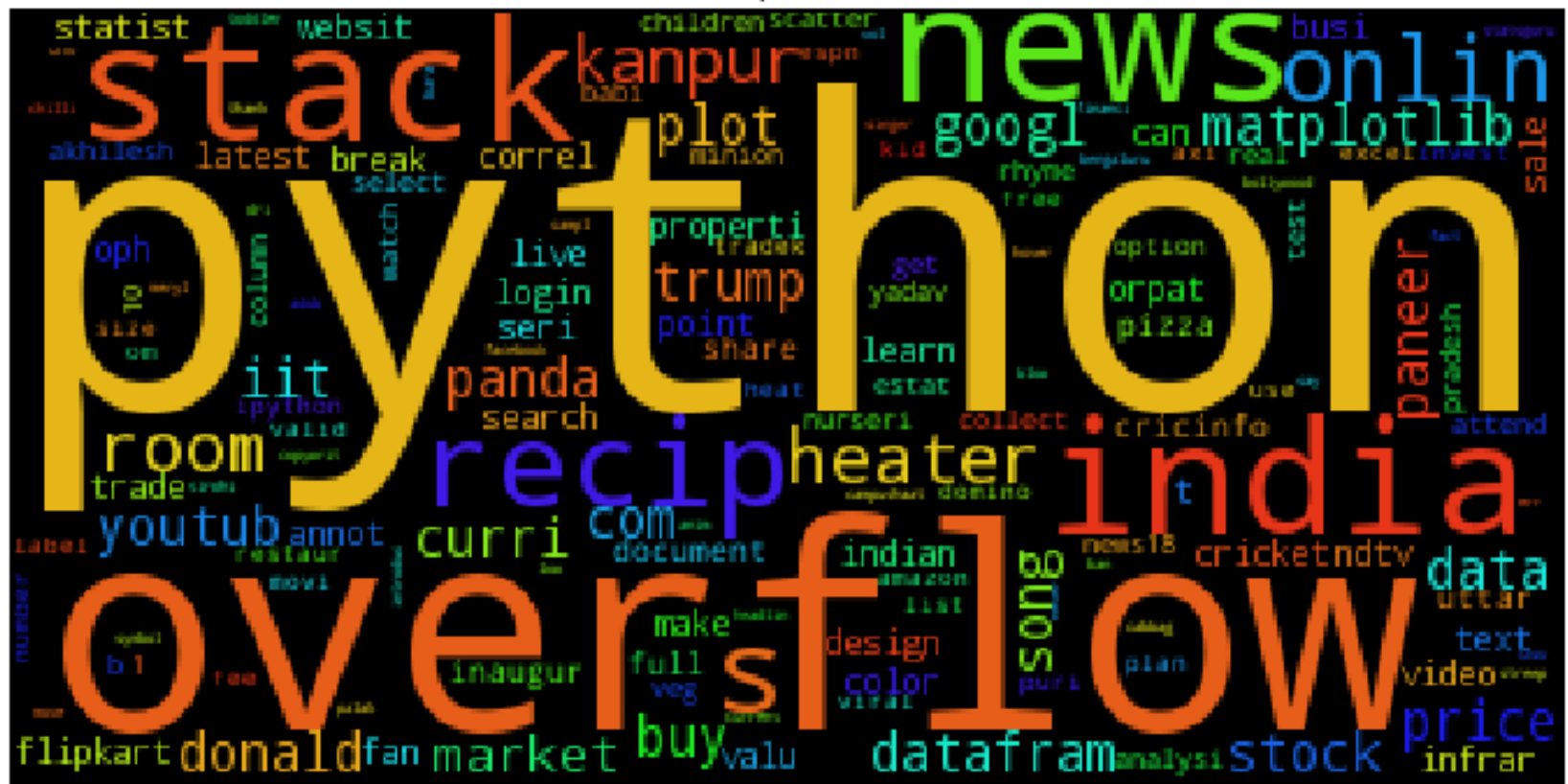- Shows basic results
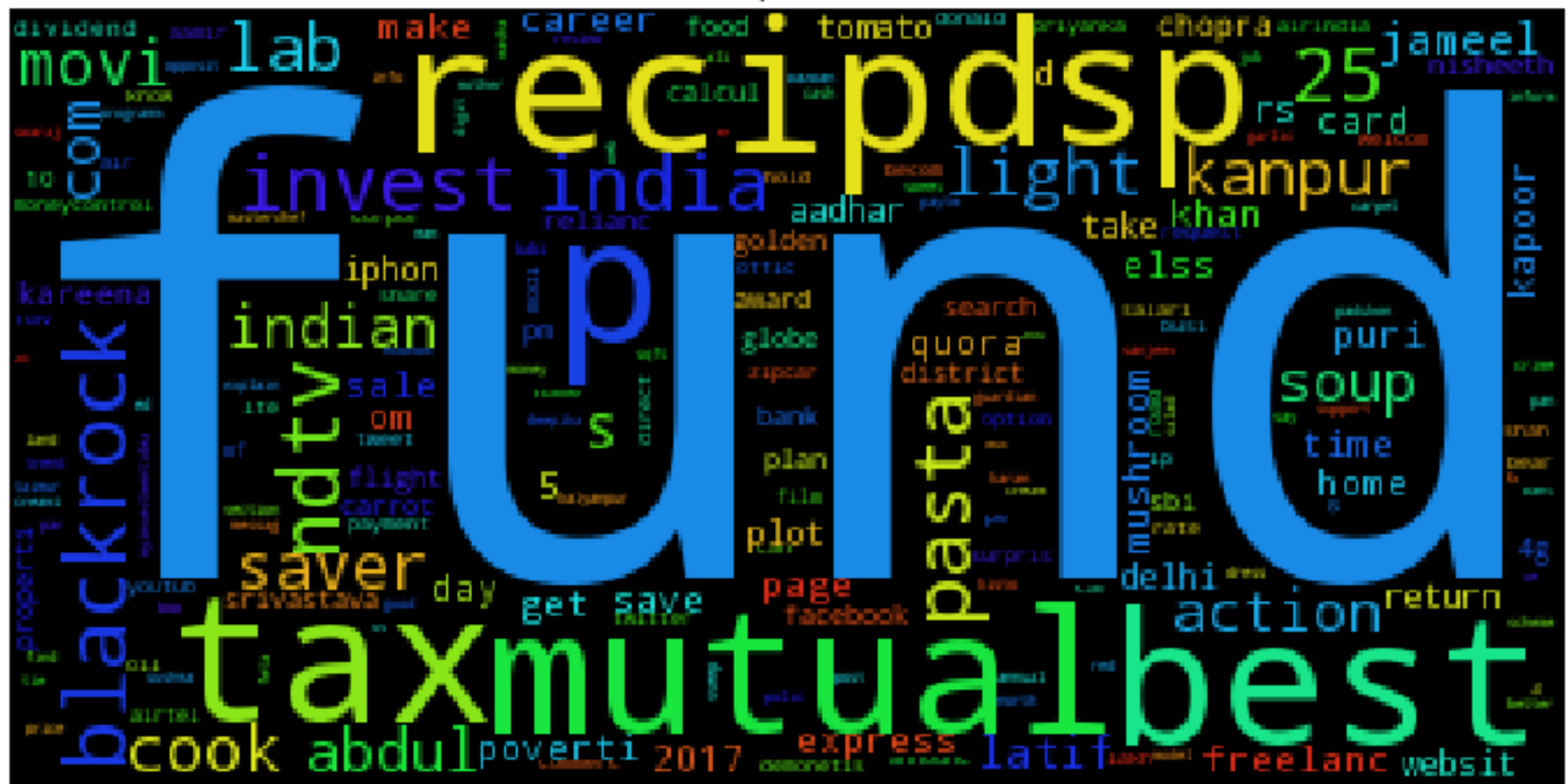
Sample topics

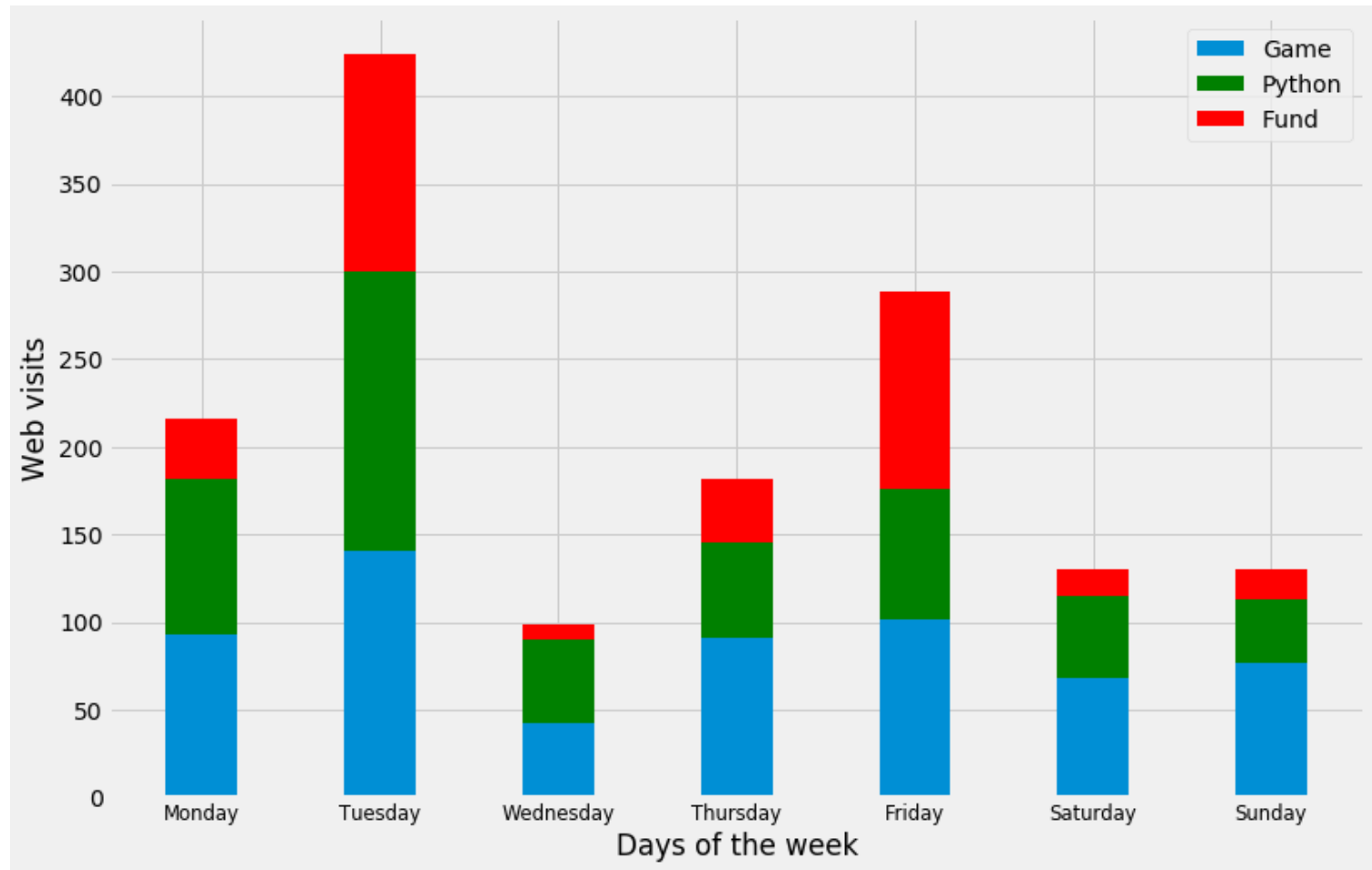Topic #0
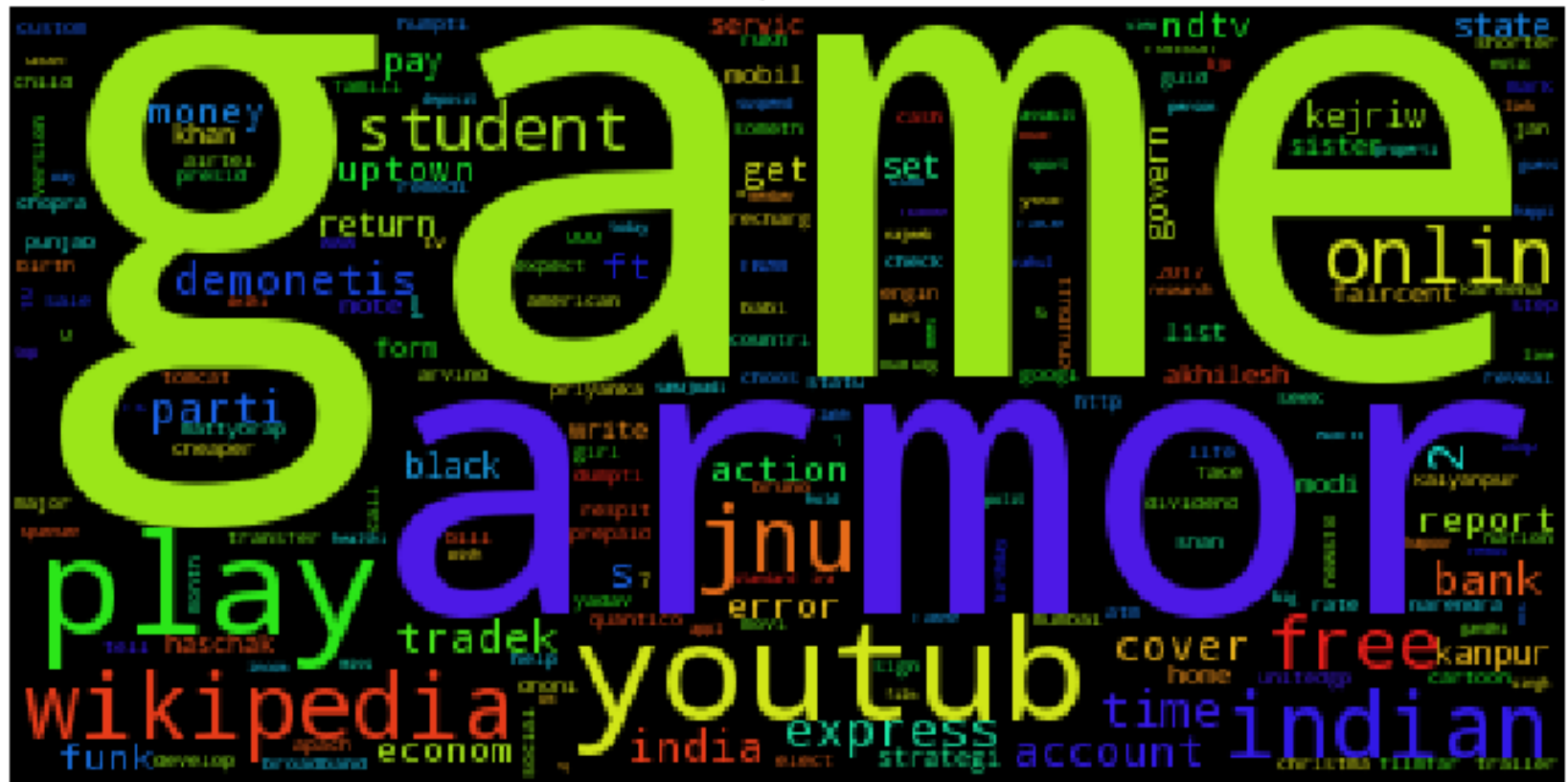
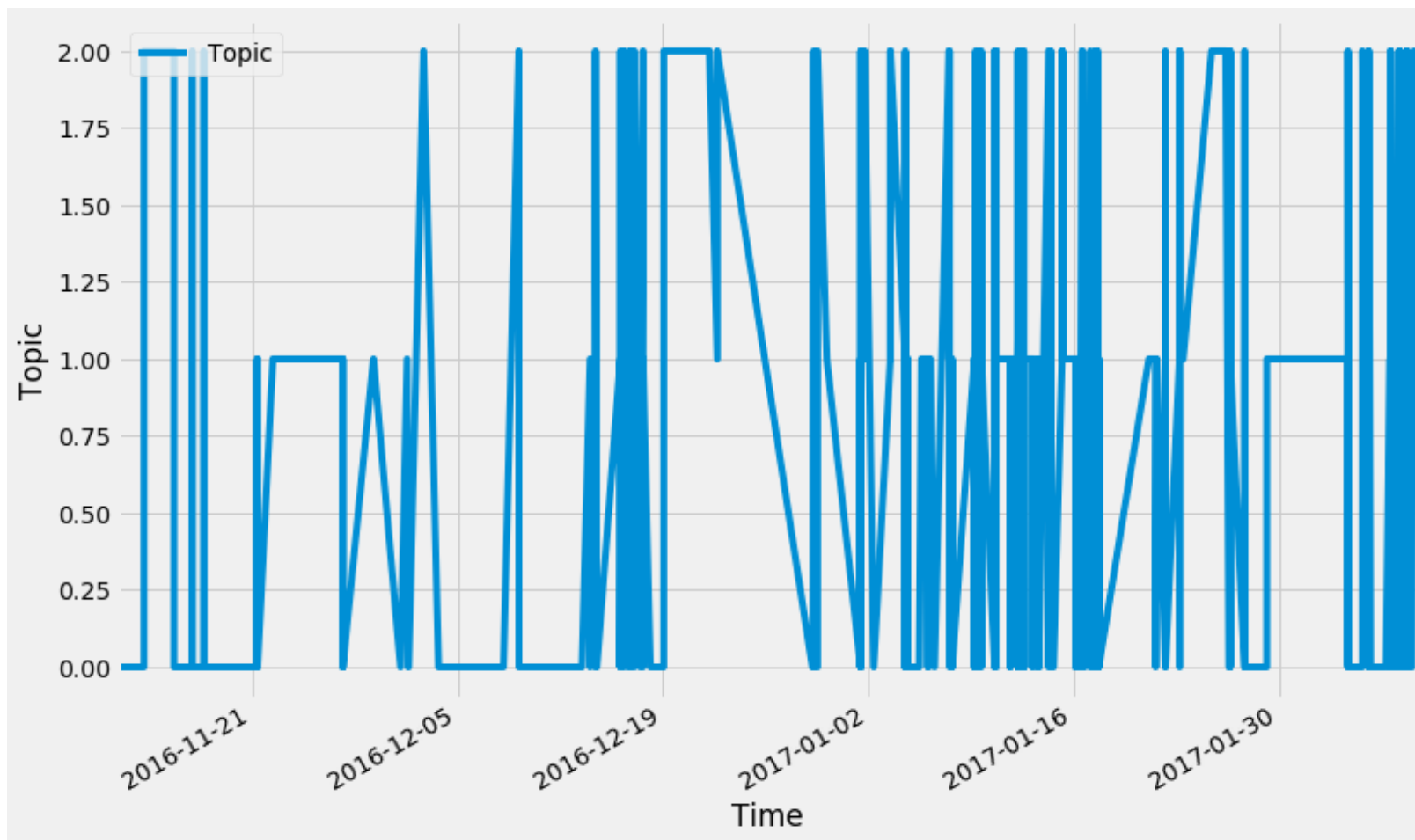# Sample topics

Topic #1

Sample topics

Topic #2

Elementary counts analysis

Elementary temporal analysis

Possibilities

- Option 1: Extend counts analysis to hours within the day
  - Instead of days within week I showed here
  - Generate within-day predictions for which topics are likely to be viewed in the coming hour
  - ***

- Option 2: Identify transition points within browsing history
  - Dig into the LDA code, or your favorite clustering algorithm's code, to probabilistically characterize the location of the user within the topic at the time of a transition
    - Bonus points for using dwell time as a marker
  - ****(*)

- Option 3: Predict topic transitions (at whatever time-scale)
  - Use HMMs to estimate transition probability as a function of hidden latent parameters, estimated via temporal and learned TP characteristics
  - *****

More information

- Deadline is March 8th
- Feel free to come to me if you get stuck
  - Optional deadline to tell me what you're doing Feb 21st
- Use python – it will make your like easier
  - My setup: Anaconda
    - Extra packages needed for this project
      - Stop_words
      - Gensim
      - NLTK
    - Can install easily from conda command line