14 August 2015

ABHINAV

1. Looked up sources for various statistics. Important web-sites for the stats regarding Bills etc.
    a. http://www.prsindia.org/aboutus/what-we-do/
    b. http://rajyasabha.nic.in/
    c. http://mospi.nic.in/mospi_new/site/India_Statistics.aspx
    d. http://parliamentofindia.nic.in/ls/
    e. http://164.100.47.4/newbios_search/sessionreport3.aspx
    f. http://www.indiastat.com/default.aspx

2. Still to figure out how to fetch the data from these web-sites in a usable form.

3. Trying to get bokeh-server working on Windows as well. Will try to do it again tomorrow, else switch over to Linux. Bokeh-server is necessary for the construction of Dynamic and Interactive Plots.

4. **TODO** Need explore Pandas and Bokeh deeply. Bokeh - Done!   Pandas - Making Progress.

5. **QUESTION NIHARIKA** How do you plan to tackle the MAIN PROJECT?

6. Finalized the Book and the Toolkit for the Project [**For all Projects**] -

    a. BOOK - Python3 and Text Processing with NLTK3 Cookbook.

    b. http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/

    c. LIBRARIES
        i. Pandas
        ii. Requests
        iii. Beautiful Soup
        iv. Bokeh
        v. MatplotLib
        vi. NumPy

7. **NOTE** What do we need to represent is still on discussion-stage.
    a. The sources are still to be decided.
    b. The ***Visualization-scheme*** is yet to be finalized Done
    c. A ***timeline*** of introduction of bills Done

d. the proportion of for-against candidates.
    1. Candidates Education level
e. Public reaction viz fetched from the networking websites. It's parameters
    1. Age Done
    2. Gender Done
    3. State Economy and General Education Level Done

16 August 2015

ABHINAV

1. Need to look for CSV encoded data when we download statistical data regarding various parameters we are going to co-relate the sentiments with.Think More

2. Regarding the Tweets, the data can obviously be fetched regarding public sentiment from the API's provided by Twitter and other Social Platforms.

3. Programming Skills that need to be focused on . Think More
    a. Git
    b. Regular Expressions
    c. Data Visualization Techniques

4. Another book, apart from numerous manuals that is quite useful for the project. - **Python for Data Analysis**

17 August 2015

NIHARIKA

1. What about **Anaconda**? Everything will be done within Anaconda -> http://docs.continuum.io/anaconda/pkg-docs

ABHINAV

1. One of the final visualizations are supposed to look as the one show in the main page of the following web-site (On a National Level)
    a. http://www.electionsinindia.com/default.aspx   Checked

2. Make sure you have had a look at the Visualizations offered by the **Sarah Bird's Pycon talk** - based on data from Africa.   SEEN

3. Data about the various Education parameters per state [among loads others] is available on
   a. http://www.indiastat.com/Education/6370/specimen.aspx

4. **TODO** Find a way to download such tables in CSV format. Making Progress

5. Gonna write up an overview regarding the Current Heading to bring Ma'am up to date, tomorrow.

6. Important Resources regarding Twitter
   a. https://dev.twitter.com/streaming/public
   b. http://tweepy.readthedocs.org/en/v3.4.0/api.html
   c. http://mike.teczno.com/notes/streaming-data-from-twitter.html
   d. http://schoolofdata.org/2013/04/25/social-network-analysis-for-journalists-using-the-twitter-api/
   e. http://sachithdhanushka.blogspot.in/2014/02/mining-twitter-data-using-python.html
   f. http://www.kdnuggets.com/2012/11/best-python-modules-for-data-mining.html
   g. https://dev.twitter.com/web/sign-in/implementing
   h. http://adilmoujahid.com/posts/2014/07/twitter-analytics/
   i. http://www.kdnuggets.com/2012/11/best-python-modules-for-data-mining.html
   j. https://pypi.python.org/pypi/tweetstream/
   k. http://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/

7. **TASK** - **NIHARIKA** - I need you to Google two topics
   a. " **How to find the various stats related to a Bill ??**" -> As in how many MP's were in favor/against.
   b. " **Official Profiles for MP's ??**" -> As in their official ID's where their qualification/ state etc are enumerated.

8. Another Python Library - Flask seems useful as well.

18 August 2015

ABHINAV

1. Decided the General Flow of the entire presentation as well as the project.

2. Will go forward with the Technological stuff as soon as Ma'am approves the direction.

3. **NOTE** The Parameters for the graphs are yet to be decided. Making Progress

NIHARIKA

1. **Examples of various bans happened** recently(to be updated as discussed) :

    a. Ban on beef by Maharashtra Government → March 4
(http://www.thehindu.com/news/cities/mumbai/beef-ban-hc-asks-for-exact-date-of-enforcement-of-new-law/article6982581.ece)
    b. 28 banned swear words in movies by Central Board of film certification. → Feb 11
(http://www.deccanchronicle.com/150213/entertainment-bollywood/article/here%E2%80%99s-list-words-banned-censor-board-india - Some tweets are there too)
    c. Patna High Court banned movie 'Dirty Politics' in Bihar
    d. Bikini banned in Goa → late june around 28 june
(http://www.india.com/loudspeaker/ban-bikinimini-skirts-congress-demands-88709/)
    e. Lingerie and Mannequins → may 18 2013
(http://www.ndtv.com/mumbai-news/after-lingerie-mannequin-ban-mumbai-politicians-want-ban-on-lingerie-ads-too-523626)
    f. fifty shades of grey
    g. BBC Documentary- India's Daughter ---> supposed to be released on Mar 8,
Sun. After being banned by govt. on Tuesday,March 3, released by BBC the very next day .
(http://www.thestar.com/news/world/2015/03/04/bbc-airing-documentary-on-gang-rape-after-controversial-ban-in-india.html)
    h. Karnataka govt. banned all parties where foreigners are invited unless it is under
police vigil.
    i. Maggie banned
    j. Ban on porn -> August 1 or 2
(http://www.ibtimes.com/india-bans-porn-narendra-modis-government-launches-massive-crackdown-adult-websites-2035837)
    k. AIB's roast show on youtube → Feb ½
(http://topsy.com/s?q=AIBroast&window=a)
    l. On word 'Bombay'

2. Have to create a report, to be sent to mam. Done

19 August 2015

ABHINAV

1. The Report has been edited and sent.

2. **TASK** - NIHARIKA - Do pay attention the changes that have been made to the draft. The process of the Progress Log might be done faster in future. Good job, with the draft.   Done

20 August 2015

ABHINAV

1. Haven't received any reply from Anamika Ma'am till now. I have started working with NLTK in any case.

2. **Will need to call her tomorrow (21 August)** if we don't receive any reply by then.

3. Have *received the Go-Ahead* for the project.

4. Made the computer system (Linux) ready for the project.NLTK wasn't working on laptop, not at least with it's full functionality.

5. *Made no progress today, whatsoever - A day wasted!!*

6. Will work on understanding Bokeh-servers today. Hope to make first dynamic vidaĵeto on the web page.

21 August 2015

NIHARIKA **** Pay attention** to the Indentation and Color coding we are using. Stick to them. Takes a bit extra time but trust me, gets easier to skim documents later and **neat and visually appealing work is an Art.**

NIHARIKA

1. Checked the Sarah's Talk and other websites mentioned.
2. Installed Emacs and Anaconda

**http://ubuntuhandbook.org/index.php/2014/10/emacs-24-4-released-install-in-ubuntu-14-04/

3. Regarding :
    a. " **How to find the various stats related to a Bill ??**" -> As in how many MP's
**were in favor/against. Yes.
    b. " **Official Profiles for MP's ??**" -> As in their official ID's where their qualification/ state etc are enumerated. Yes.

** Considering the relevant sources, I guess it's pretty hard to find which MLA particularly were in favor or not. But to get some data, we have to extract information by running a python script (or may be something else) on following sites.

**http://www.prsindia.org/downloads/bills-pending-in-parliament/
**http://www.prsindia.org/downloads/bills-passed/2015/

4. Made a sketch for basic git commands using following urls:

**https://help.github.com/articles/good-resources-for-learning-git-and-github/

**https://try.github.io

**http://git-scm.com/about

**https://training.github.com/classes/

**http://marklodato.github.io/visual-git-guide/index-en.html

**http://rogerdudler.github.io/git-guide/ (Go through this, really good.)

**https://training.github.com/kit/downloads/github-git-cheat-sheet.pdf (pdf for git cheat-sheet,

**better to use than my bad handwritten note. Will send them anyway,by mail.)

5. Which IDE did you use to run Anaconda? Oh, it's called **Spyder** but you can also use it as a simple **IPython web-page based process.** IDE's are basically for advanced coding processes like Code-Factoring , Simplifying Large Projects, Debugging etc.

Might be a **good idea to pin-it-up on the Ubuntu sidebar** once you have them running via the Terminal. Clarified

ABHINAV

1. Made a sketch for the chunking of the entire process so that we have various stations with a clear input and a clear output.

2. **TASK - ABHINAV** Send the Chunk-up scheme Images to Niharika. DONE - You should be having an invitation to the NOTES folder in where 8 images have been uploaded. **I forgot to serialize them**, so take your hints from the phone's naming scheme. Will keep in mind next time. DONE.

3. **TASK - NIHARIKA** You need to make an *Account on GitHub* so that I may share the status of the Code and the various Prototypes with you at all times. My ID is, of course ,<< abhi18av >>.Okay!

   NOPE, that doesn't work as the private repository needs an upgraded paid account. But,what's that? Clarified

   Will figure out the solution using other methods but I think that being able to read the working-prototype (rather than development prototype!) would only help your cause.

4. **SUGGESTION - NIHARIKA** out You can get your feet wet again with Python programming. The very basics as of now but as we progress I would like you to contribute in the Coding department as well. It's better to develop the various chunks of the Code in parallel. Well obviously, did I ever implied that I don't want to contribute in the coding department? Okay!

It would be a plus point if you understand what has been done with the Code and would be of immense advantage during the presentation as well.

22 August 2015

ABHINAV

1. After much thought and effort I got the Bokeh Server working on the Ubuntu machine as well. The steps I followed were
   a. To update the Anaconda System [most libraries have been advanced]
      i. Change the permission of folders in anaconda3 folder.
      ii. get to the sudo mode
      iii. conda update conda

   b. Move to the examples folder in anaconda3 folder
   c. bokeh-server xyz-file
      i. http://brocktane.net/blog/bokeh-server-example/
      ii. http://bokeh.pydata.org/en/latest/docs/user_guide/server.html

   d. open up your browser and run the code file that after slight modifications
   e. **Folder shared** with you << code >>
   f. **Install** the <<Drive Notepad>> app on Google Drive to be able to share code as such.

2. Will commence with the actual Bokeh experimentation. Done

3. **NIHARIKA** Updates regarding the Screenflow and the Plots.

4. **SUGGESTION** - **NIHARIKA** Head over to the following link and use the visualizations to aid the screenflow.
   http://nbviewer.ipython.org/github/bokeh/bokeh-notebooks/blob/master/index.ipynb

5. *This is how our final work is to Look*.

23 August 2015

ABHINAV

1. The Web-drive has been brought up to date.

2. *Made HUGE progress regarding the Bokeh Library* and experimented a lot with the various elements that we can alter within the Web-Page. Happy With the Progress Made;P

3. **TASK - ABHINAV** Things which are yet to be done
   a. How to open a new web-page (computed) when clicked on a certain element of a <<vidajxeto>> Done
   b. Prepare the entire screenflow with the various screen elements (elementary ones) and get the output-prototype running. Making Progress. Done

4. **TASK - NIHARIKA** I have uploaded a bunch of HTML files <<Bokeh-samples>>. Do run them on your computer and let me know that they are dynamic over there as well.

24 August 2015

ABHINAV

1. On the Visualization end most of the research-experimentation work has been done. Though, we need to ascertain whether we Really need CALLBACKS for our plots? As this facility would need a working knowledge of JavaScript as well - which increases the total time-cost by a couple of weeks as the Doom's Day scenario.

2. I am of the opinion that we should get done with this project as soon as possible. I have to focus on the Next Project;P

3. Bokeh already provides us with much of the Dynamic Nature out-of-the-box. And I have figured out ways to link the various plots together. Done

4. *Prototype - 1 is gonna take a lot Less time than what I initially anticipated*. In a couple of days I shall put up the *web-of-resulting-HTML's* in the Code folder.

5. *All that would remain after the completion of this stage* would be to move ahead with the NLTK-Twitter and to fetch the Statistical data from the various web-sites and to prepare the actual data-bed for the visualizations.

6. No Updates from you, **NIHARIKA ??**

7. **TASK - ABHINAV** Put up a working HTML-file-web for as the prototype.

   a. The Charts/Plots are gonna be based on dummy data.

b. The Leitmotiv is to get the Presentation part working and then to fine-tune it slightly as and when required.

c. The Project's Source Code is to be distributed in a bunch of Source files which combine together in a coherent whole. As in, One-for-all and All-for-one. Think More

26 August 2015

ABHINAV

1. Made progress today with the basic connectivity and the screen interactivity.

2. The final file structure was also elaborated.

3. Currently working hard to overcome the HDF5 data format manipulation so as to use the Geo-Spatial data to produce the Map of the country and it's various states.

4. It's this map that's the real hurdle at the moment. Thus I have moved this map on ToDo list and working on the other simpler aspects of HTML presentation.

5. Need to commence using the actual data once the statistics and parameter we are to use have been elaborated upon further.

27 August 2015

ABHINAV

1. Various useful sources for the Geo-Spatial data (for producing the Chloropleth)
   a. http://gadm.org/
   b. http://community.tableau.com/thread/136772
   c. http://geocommons.com/overlays/73828
   d. https://github.com/datameet/india-election-data
   e.

2. Still not quite comfortable with the Data Formats.

3. Apart from the Chloropeth (Statewise division) most plots can easily be produced but again we need to learn how to deal with Big Data anyhow.

4. **TASK - NIHARIKA** Have a look at this link for the various parameters regarding the official statistics

a. https://data.gov.in/

5. After Much surfing, researching, reading the manual and books, I FINALLY HAVE THE MAP OF DELHI !!

6. Good Progress made today.

7. Though my points are in CSV format which slows the entire process down as there number of points (latitudes and longitudes) are HUGE. Need to learn how to port it to HDF5 format and then to base the workflow upon that form.

8. **SUGGESTION** - **NIHARIKA** This is how you can change the permissions to the folder and get the entire Tool Chain Working -
    a. Navigate to the anaconda3 folder (don't open it)
    b. RIght click and << open terminal >>
    c. ls -l
    d. notice the permission along the << anaconda3 >> folder
    e. <<drwxrwxrwx>> or << drwx ----- >> is the goal!
    f. If the permission is different , run
    g. sudo chmod 700 -R anaconda3
    h. OR
    i. sudo chmod 777 -R anaconda3
    j. now run << ls -l >> again.
    k. after this we need to update the anaconda version and packages
    l. sudo gedit
    m. close the gedit (it is just to reach the root user mode again, in case you closed the shell)
    n. conda update anaconda
    o. Now open up spyder and notice the version of Python and Anaconda. It should say the following in the Ipython console
    p. << Python 3.4.3 |Anaconda 2.3.0 (64-bit)| (default, Jun 4 2015, 15:29:08) >>

9. And you are all set as far as the Environment is concerned.

10. Points about *the meeting held today* and the outcomes from the meeting.

11. Stage 1 is almost over, which involved
    a. Figuring out the basic visualizing scheme
    b. The databases and resources we are to use
    c. The various skills we are gonna use and how

12. **TASK ASSIGNMENT - ABHINAV**
    a. Figure out how the Format Conversion works for maps as well as the other databases
    b. Get the India Chloropleth working
    c. Screen-Flow for a particular instance (One-Ban)

    d.   Get started with the Regular Expressions

    e.   Figure out how to do Effective Source Code chunking

13. **TASK ASSIGNMENT -** NIHARIKA
    a.   Download the actual datasets. DONE
    b.   Move forward with the NLTK and Twitter API's
    c.   Design the SCREEN-FLOW as per the major scheme discussed in the meeting. DONE

14. NIHARIKA Once you have decided on the screens and parameters, **send me a mock up using Google Drawings!** Check Mail!
        Use actual Graphs from Bokeh-Gallery and tag them using the text facility in Google Drawing.

15. **TODO** The entire expression is to be converted into a Book and look up suitable CreativeCommons license. Post up all the code on GItHub.

28 August 2015

ABHINAV

1. Apart from exploring the Native Python CSV module, no progress made.

NIHARIKA

1. Reinstalled Anaconda for the python3

29 August 2015 (Rakshabandhan)

ABHINAV

1. Not much progress made today either - will do my piece of work tomorrow.

NIHARIKA

1. Same here- No progress made today

30 August 2015

ABHINAV

1. The CSV data file for the Chloropleth map needs to be shrinked down to about 10 Mb's [Extreme Case estimate] from it' current size which is 97 Mb's !!

2. Sarah Bird's Africa Plot needs only about 1.1 Mb on the disk.

3. Commencing with the other Regular Expressions, though for the current project we don't need to dwell in too deep as NLTK is going to do most of our work.

4. Regular Expressions are still important though as many database manipulation techniques make a goodly use of RE.

5. The sheer number of points causing a bit of problem. To delete the far too many surplus points I need to learn more about CSV file manipulation from within the Python environment and then repeat the entire process of Data Conversion and the Plotting of those Data Points via Bokeh!

6. Much of it is done, but a final push is necessary to get the entire map right. As we don't really need to go below the various Intra-State-Level divisions.

7. ***Sarah's database has only 60 points per state as opposed to 100000 in the one we are working with.***

NIHARIKA

1. Initiated with the screen flow.

1 September 2015

ABHINAV

1. I am lagging behind on my tasks and to change the Scenery up a bit, I am gonna work with the NLTK and the Twitter API to get the numbers out of the text.

2. More links regarding the Twitter Analysis [and Geospatial data]
   a. http://www.laurentluce.com/posts/api-to-access-the-cambridge-city-geospatial-data/
   b. http://www.laurentluce.com/posts/python-twitter-statistics-and-the-2012-french-presidential-election/
   c. http://code.runnable.com/UqCbQNqXMkkMAALL/twitter-sentiment-analysis-for-python

     d.   http://www.cs.cornell.edu/people/pabo/movie-review-data/
     e.

3. Installed the Tweepy module in ubuntu using the following commands
   a. sudo gedit
   b. # then close the gedit window
   c. pip install tweepy

4. The *GITHUB repo* for the project is under the name [the-visual-verdict](#).

5. Making progress with the GIT workflow and there is gonna be quite some data which doesn't need to be published on the GitHub repo. Thus more work needs to be done on << git ignore >> command.

6. Have made a twitter account and registered the project under the name << the-visual-verdict >> and have obtained the required credentials.

7. Better to divide the process of NTLK analysis on Twitter into multiple chunks
   a. Learn to use the NLTK on << simple tweet text >>
   b. Isolate the << base text >> from the meta-data jungle in the raw json format.
   c. Working with the json data from website mentioned on **31 August 2015 // ABHINAV // 2 // c**

8. Have started to work with the << ZSH + GIT >> environment, **NIHARIKA** you should be able to keep an update regarding the progress made with the code in the project.

9. I have shelved all other activities in the meantime. It is my utmost desire to get it all done before the mid-sems. ***Dragging simple things only end up making us more of a procrastinator - not an ideal to strive for, I believe.***

10. **SUGGESTION** - **NIHARIKA**
    a. Make sure you ***get going with your repository*** as well.
    b. install << **gitg** >> from the Ubuntu repository - will help keep track of things visually.

2 September 2015

ABHINAV

1. Oh, scraping data from twitter for specific profiles is quite easy!

2. Now I too am catching up with the the PythonProgramming.net tutorials on NLTK, though I am doing only the last ones - 19, 20, 21

3. From what I gather, the entire NLTK series is about 212 minutes long, **NIHARIKA**, it took you a lot longer than it should have!

4. Been getting a few errors as the tutorial is really based on Python2 rather than Python3 so tweaking the code a bit was necessary.

5. **SUGGESTION** - **NIHARIKA** Ran into *2 errors and solved them* thus -

   a. **401**
      i. The reason for this error is really the imprecise time settings on the machine
      ii. << sudo ntpdate ntp.ubuntu.com >>
      iii. This command syncs the system clock to the Internet's clock.

   b. **Can't convert 'bytes' object to str implicitly**
      i. After much surfing I came across the minute thread on github forums
         1. Look for comment by Cozos
         2. https://github.com/tweepy/tweepy/issues/615
         3. Made the suggested changes to the <<streaming.py>> file
         4. ~wrkspc/anaconda/the-visual-word/twitter-prac/tweepy-master/ tweepy/streaming.py
         5. Then go back to the <<tweepy-master>> folder
         6. Re-install the <<tweepy>> module by running the <<setup.py>> file as follows
         7. python setup.py install
         8. Used the code from Pythonprogramming.net code in Spyder
         9. Entered the app credentials (keys)
         10. ***Und wir sind da! Now Spyder is singing tweets like a canary***

6. Further progress needs to be made to the code regarding the <<filter>> function that has been defined, which solves the problem of location based tweets.

7. Good Strides taken today! Happy;P

8. **TASK ASSIGNMENT - NIHARIKA** Screen Flow needs more work. **???**

NIHARIKA

1. Tried to access Twitter's API using developer's account. Had some issues while registering Mobile No. and unable to access it.

2. Recently upgraded, one can only access API after registering mobile no.

I guess, scraping is a better idea. Right!
Will work on that in case the << tweepy >> doesn't deliver.

Though, now that I have got it working - it's all a matter of *going through the documentation.*

3. Looked up and surfed a bit for MP break down per ban. But, didn't achieved anything concrete yet. I have a feeling that even this kind of an important data *would be placed in obscurity somewhere*. Keep the search going *or find good Alternatives.*

3rd September 2015

NIHARIKA
**NIHARIKA

1. Finally found something tangible on MP Ban: (from Lok Sabha Debates)
   http://164.100.47.132/LssNew/psearch/Debatetextsearch16.aspx
   http://164.100.47.132/LssNew/psearch/DebateAdvSearch15.aspx

2. From Rajya Sabha Debates

http://rajyasabha.nic.in/# Click on Debates -> Official -> Search

http://rsdebate.nic.in/simple-search?sort_by=dc.date.debatedate_dt&order=DESC&query=ban&submit=Go

3. I guess, time has finally come to select/choose particular bans now. Sure, but wait till tomorrow. I'll mail you something then use that overview for the data.

ABHINAV

1. This is useful stuff, **Good Job**.

2. **TASK ASSIGNMENT - ABHINAV** Things are starting to get a bit messy with all these libraries, code and resource around. Need to organize all the data and to commit all code over to Github.

4 September 2015

ABHINAV

1. The second progress report was completed, to be sent to ma'am tomorrow.

2. Gonna organize the source code files for the various aspects of the Project.

3. This project shouldn't take too long to be functional now.

5th September 2015

NIHARIKA

1)Lok Sabha

Got a pdf containing specific details that might help huge :
http://164.100.47.132/LssNew/Members/infobulletin.pdf

Contains information on :
1. Distribution of Age Group of members with a span of five years
2. Average Age of Lok Sabha Members
3. Different Age-Groups of First- time members
4. Educational Background of members
5. Occupational Background of members
6. First-time elected members(state-wise)
7. Women members of the 16th Lok sabha (state and party-wise)
8. Educational and Occupational background of women members

2) http://www.prsindia.org/parliamenttrack/about-parliament/the-indian-parliament-198/
   Helps in bill tracking, MP tracking and parliament tracking

3)Tracking an MP: (Will help a lot)

http://www.prsindia.org/mptrack/rajyasabha
http://www.prsindia.org/mptrack/16loksabha/

Ex. - UP http://www.prsindia.org/mptrack/16loksabha/list?state=Uttar%20Pradesh

-->We can even download data.
-->Can sort by state, party and age. Other factors :
   a. attend Parliament regularly
   b. ask more questions
   c. participate regularly in debates
   d. introduce private member bills

4)Four ways in which the decision of the house can be ascertained either in favour of or against. (more detail in other doc)
Going by this, we can at least figure out who initiated the bill regarding a particular ban.

ABHINAV

1. Great Job, **NIHARIKA** - Sure now we can put the MP based <<vidajxeto>> [Eo for visualets!].

2. We already have the state parameters somewhere in the above links.

3. I think that it's best to invest some time in learning databases so that we are able to isolate the various instances related to a particular parameter - though I wish to keep it as simple as possible, this Database approach is a fallback policy.

4. The website mentioned previously on   **17 August 2015 // ABHINAV // 3 // a** . is useless for our purposes as it is from a private limited company and one needs to have a password etc to access the data sets.

5. Found the following site for various state related parameters -
   http://www.censusindia.gov.in/2011census/population_enumeration.html

6. ***This feels Great! Things are falling into place again. I think the only major thing that is left to do are***

   a. *VISUALIZATION*

      i.   Define clearly the Parameters

      ii.   Learn how to correlate data of the state with that of the MP

      iii. It's all forming itself in my mind - shouldn't take long now.

   b. *TWITTER - NLTK*

      i.   Need to chunk out numbers from the Tweets.

      ii.   We obviously don't see the reactions from a previous date - as the API offers live-feed. Need to find a way around that.

      iii. Need to use the <<filter>> function to sort out the data from various states regarding various bans - or simplify our criterias.

      iv.   **NIHARIKA** Look Up

1. https://dev.twitter.com/streaming/overview/request-parameters

2. http://thoughtfaucet.com/search-twitter-by-location/examples/

3. Shall we use hashtags - look into it.

4. http://tech.firstpost.com/news-analysis/now-you-can-look-for-tweets-by-specific-date-using-twitters-advanced-search-220903.html

5. http://www.digisecrets.com/social-media/5-best-sites-to-search-old-tweets/

7. **TASK ASSIGNMENT - ABHINAV**

   a. Need to get the Datasets right.

   b. Get the Visualization Prototype working.

   c. Get the repository in order.

   My tasks are pretty much the same, I seem to have spent much time trying to set up the environment right for the various needs. Haven't read the Manuals for a long time now - will do so soon.

8. **TASK ASSIGNMENT - NIHARIKA**

   a. NLTK and Twitter is your job for the time being
   b. Focus on exploring the various fine-tuning that you can make using the <<filter>> function and make a summary of your experimentation.

   c. Better to start saving up dedicated Scripts files so that we may integrate them all quickly.

   d. Put your work up on GitHub.

6 September 2015

ABHINAV

1. Haven't received any reply from Ma'am regarding the Project. **NIHARIKA** Would you please call her up on Monday in case there is no class - if there is then we'll go and meet her.

2. And this baby has come alive! Created the visualization screen using the Dummy data.

3. Two of the Major goals were achieved during this stage -

   a. Arranging multiple plots on the same html page

   b. Linking a plot to another so that when clicked on a particular part it opens up another plot - Doing this for the locally stored file was a bit round-about, got it done anyway.

4. **NIHARIKA** Would you like to have those HTMLs and Code ?

5. Next up the ***main thing left in Visualisation*** is to get the ***Chloropleth*** working. The skills required to do this include ***the ability to handle CSV and HDF5 formats*** as well. It's my Main priority now.

6. I need to reduce (drastically so) the number of points per state.

7. **NIHARIKA** Progress made on the NLTK and Twitter analysis. **??**

8. ***Made Immense progress regarding the ScreenFlow - Database - Folder Structure as well the Graphics that we are gonna use per screen.***

9. Will give you the overview to bring you up to date once we meet up tomorrow [7 September 2015]

10. Now the Implementation Drudgery Begins!

7 September 2015

BOTH

1. Met ma'am and updated her regarding the project's progress and the direction in which, we are moving ahead.

2. Updated code, NLP corpuses [in NLTK package] on Niharika's laptop

3. Created a task flow for the next week...

4. Discussed regarding screenflow, file structure and divided tasks...

8 September 2015

NIHARIKA

1. Got stuck trying to run the tweepy package yesterday. Thought, it was an error due to the <ntpdate> command. But,it turned out that i was wrong. After scrolling through multiple web-pages ,i realized time was already synced(as i mentioned earlier).

   a. That **401** error was due to an extra space while copying keys and thus, the mishap!

   b. After rectifying that, it shifted back to the **Can't convert 'bytes' object to str implicitly' error.** Analyzed the Ipython console for the description for error, figured out that the <streaming.py> file need to be updated into the anaconda folder.

(Also updating file is stored in a zip folder. Thus extracting it and then making changes in it , won't really help. One have to open the zip folder from the terminal and then make updations in the <streaming.py> and then update it). And, it worked.

ABHINAV

1. Uploaded the Images for the proposed screen-flow and the file-structure in the << NOTES>> folder.

2. Will make more progress with the Real-Databases management.

3. The solution to tweepy problem is below.

cozos commented on Jul 20

In streaming.py:

I changed line 161 to

```
self._buffer += self._stream.read(read_len).decode('UTF-8', 'ignore')
```

and line 171 to

```
self._buffer += self._stream.read(self._chunk_size).decode('UTF-8', 'ignore')
```

and then reinstalled via `python3 setup.py install` on my local copy of tweepy.

9 September 2015

ABHINAV

1. Made progress with the CSV manipulation - but not there yet!

2. Have figured out the ScreenFlow, it does involve a lot many alternatives. We can be certain of the fact that the final Screens wouldn't involve so many graphs.

3. Less-is-More

4. CSV problem has been solved!

5. Useful keywords regarding the solution of the problem
   a. << unpacking>>
   b. << list comprehension >>
   c. << OrderedDict >>
   d. Conversion of << str >> objects to << int >>

6. Links regarding the Solutions
   a. https://developmentality.wordpress.com/2012/03/30/three-ways-of-creating-dictionaries-in-python/

   b. http://www.python-course.eu/python3_dictionaries.php

10 September 2015

ABHINAV

1. Another Good Day at the CSV lab. Can manipulate the file using the simple Python constructs and thus it's all quite sensible once again.

2. Have made tremendous progress in reading the values from CSV and then using them with Bokeh.

3. Can get the Goa plot working but there needs to be some fine-tuning required for the plots of other states. Trying hard to reduce the number of points we need for the plots.

11 September 2015

ABHINAV

1. Sent the Book request to Ma'am - a single book for now. Let's see what happen.

2. Will collect all the relevant Databases files today.

3. *Need Updates* **NIHARIKA** -
   a. Progress regarding the MP voting breakdown per vote.**?**

   b. And which votes have you decided upon**?**

26 September 2015

ABHINAV

1. Will recommence working on the project tomorrow onwards.

2. Today I caught up
   a. with all the work that has been done so far
   b. The workplan that was decided before the exams set in.
   c. The re-routing we might need to do in order to counter NA data.

27 September 2015

ABHINAV

1. Have converted the information from the PDF about current session, into and Excel Sheet.

2. It has been placed in the <<Sentiment Analysis>> set.

3. Will do my best to complete the Screens with the data that we have finalized so far.

4. Have downloaded the Basic Datasets for every Screen.

29 September 2015

ABHINAV

1. Held a small meeting today at College and the tasks updates were exchanged.

1 October 2015

ABHINAV

1. Apart from the Economy-per-state data all the data sets have been sorted out and been collected as per the following themes (each with multiple sub-categories) :-

    a. Education

    b. Population

2. Resources for Economic Data

    a. http://planningcommission.nic.in/data/datatable/index.php?data=datatab
    b. http://www.isec.ac.in/databank-weblinks.htm
    c. http://mospi.nic.in/Mospi_New/site/India_Statistics.aspx
    d. https://data.gov.in/catalog/gross-state-domestic-product-factor-cost-constant-prices#web_catalog_tabs_block_10
    e. https://www.rbi.org.in/scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20Economy
    f. http://dbie.rbi.org.in/DBIE/dbie.rbi?site=home

3. With these considerations the DataSets have grown immensely complex and thus the number of states that we shall represent on the Chloropleth has been reconfigured to 10

    a. Jammu and Kashmir
    b. Assam
    c. Delhi
    d. Bengal
    e. Bihar
    f. Gujarat
    g. Andhra Pradesh
    h. Maharashtra
    i. Uttar Pradesh
    j. Punjab

2 October 2015

NIHARIKA

1. Some insight into filter attribute on 'location'

    a. https://dev.twitter.com/streaming/overview/request-parameters#locations
    b. http://support.gnip.com/articles/filtering-twitter-data-by-location.html
    c. http://thoughtfaucet.com/search-twitter-by-location/
    d. https://twittercommunity.com/t/80-reduction-in-tweets-with-coordinate-data/36605/1
    e. https://www.mapbox.com/blog/twitter-map-every-tweet/

   Bounding box:
   → http://boundingbox.klokantech.com/
   → http://www.mapsofindia.com/lat_long/maharashtra/

ABHINAV

1. << The Senate >> will be ready to ship tomorrow.

2. Next up are the << State >> screens.

3. The focus is on
    a. The first prototype
    b. Getting the Job done using the Simplest Graphs (Less innovation for now)
    c. Figure out the arrangement of Source Code - no need be typed twice.

4. **NIHARIKA** - I haven't made any progress regarding < Location > parameter you asked for, been caught up with the Code. Will figure out those problems, in case they are still there by next monday. Done. Good!

5. Love the new changes the Google people have added to the Drive!

6. Right now, I am working on Variations of Graphs which we can use to represent the same data. Keeping it simple.

7. There seem to be a certain limited number of variations which the multiple family of graphs represent.

8. Of Course, these factors can be combined in various ways to express < Comparative > nature of data. But the overall Themes of Variations are the same.

15 October 2015

NIHARIKA

1. Some reflection on **Why can't one retrieve old tweets from the Twitter's Streaming API**?
   → https://github.com/geduldig/TwitterAPI/issues/14
   →
   http://stackoverflow.com/questions/5695564/get-all-past-tweets-via-twitter-streaming-api
   → https://dev.twitter.com/rest/public/rate-limiting
   → http://searchengineland.com/where-have-all-the-old-tweets-gone-33579

ABHINAV

1. Had a meeting with ma'am explaining the delay in the project, on 15 October 2015.

2. I haven't had done much to make a Genuine leap forward in the project for a while now.
3. The Twitter + NLTK portion is the one that's we wait eagerly for and once it's done we shall move forward with the prototype as soon as possible.

4. The goal of the Project has shifted to Get It Done, as of now.

4 November 2015

ABHINAV

1. All the source code (.py) files have been ported to the IPython based workflow.

2. Now working on making connection and correlating various aspects of data together.

5 November 2015

ABHINAV

1. All the related datasets wrt to 10 states have been isolated and are ready to be put in the visualets.

2. Initially the downloaded files from the main site were not working, rather incompatible with the tools available - no surprize there really. It's Microsoft!!

3. Had to convert them all to CSV using the Google Drive and then had to reformat them as per the set of states we have intended to focus upon.

4. Then the next step was to change them all to Pandas DataFrame objects, which could further be passed to the Bokeh Objects.

5. Next step is to customize various properties of the visualets.

6 November 2015

ABHINAV

1. Moving forward by the inches. Tomorrow, I might explore the Sphinx documentation library as there is no PDF reference for Bokeh. Need to have the PDFs at hand.

2. Topsy.com link metioned by Niharika for access to old tweets.

7 November 2015

ABHINAV

1. Made progress regarding the structure of Bokeh. There are many High Level options to make the job easier but for most tasks the Fundamental constructors contained in << plotting>> and <<models>> submenu are the most vital.

2. My approach for the moment is to get the alpha out, which means that the kinds of graphs displayed in the various pages might be repetitive - here we can improvise with limited built-in factors available per << Chart>>

3. Post alpha, I shall work exclusively with the Basic Building Blocks of the Glyphs.

4. Remark - MatplotLib is much, much easier to understand as one doesn't has to deal with the HTML and rendering part of the Interactive Bokeh Plots. But the experience with Bokeh is an awesome and enlightening experience in its own right.

5. Spent much time trying to get the Bokeh source code converted to PDFs. Beats me...for now.

8 November 2015

ABHINAV

1. Made progress with the Documentation as far as understanding the basic of building ones own personalized docs is concerned.

2. Explored <<Seaborn>> for improving the default << Style >> in << Matplotlib>>.

3. Shouldn't take long to have the alpha roll out.

13 November 2015

ABHINAV

1. After many compromises made and trying out a lot many things, the end Jupyter notebook was exported in a HTML format and sent to Anamika ma'am and Niharika.

2. We are gonna explain why we changed our entire toolkit for this semester and, honestly are relying on the Graces of Jupyter to gloss over quite a few details.

3. There is still some fine-tuning which can be done for notebooks as the exam have been shifted to after the theory exams.