



Introduction

- How much StackOverflow questions can explain what a programmer specializes in.
- Exploring these patterns provide insight of how languages are linked?
- Deduce the buzz words of a question when it is asked in a time efficient manner.

Data Preprocessing

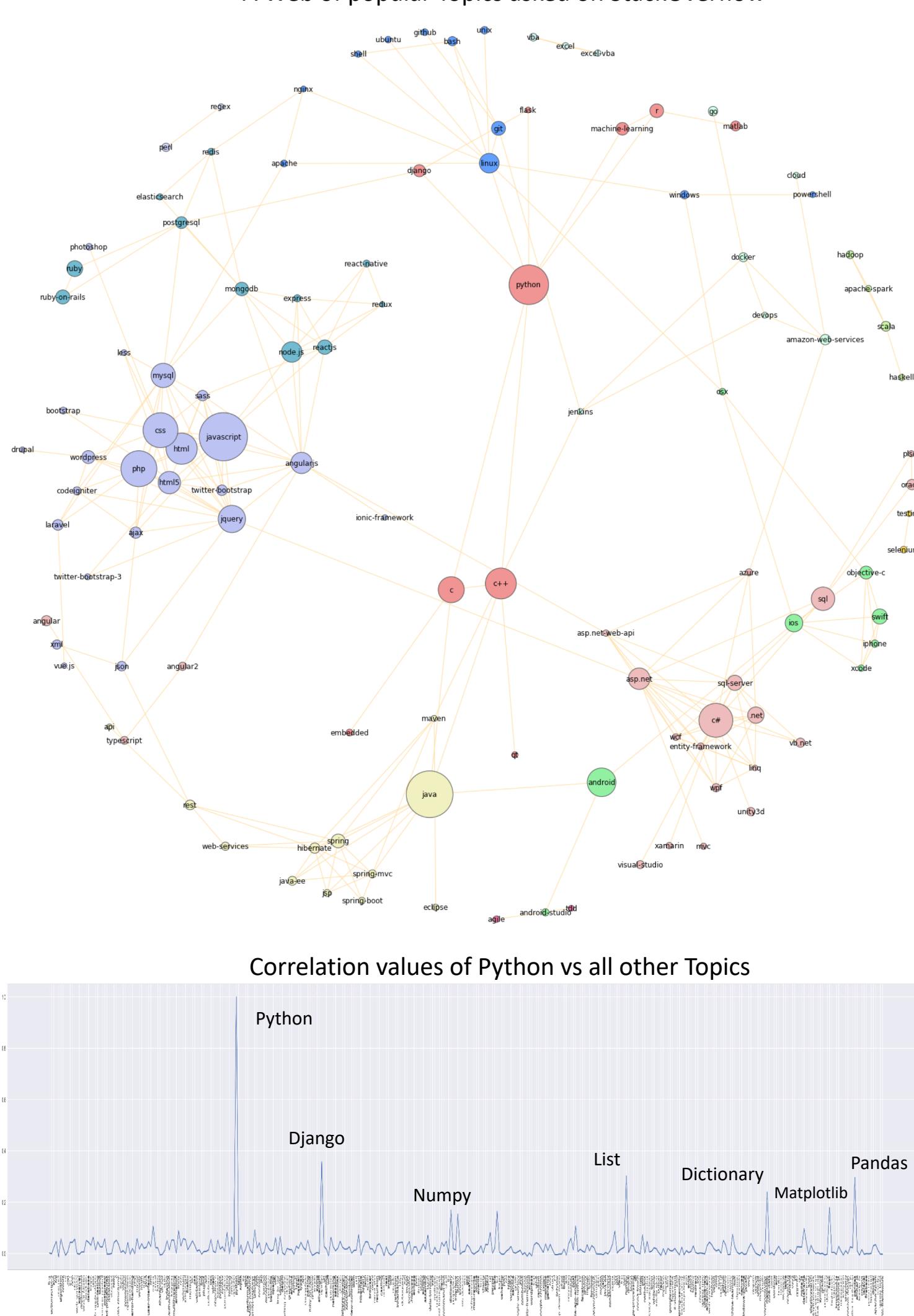
- Downloaded 10% of questions asked on the site.
- Python scripts to remove HTML, 'English' stop words.
- Filtered out uncommon questions.
- Users were profiled as high dimensional points based on a count of Topics addressed in questions

```
<p><a href="http://svnbook.red-bean.com/">Vers... version control subversion good resource sourc...
<p>i wound up using this. It is a kind of a ha... wound using kind hack actually works pretty we...
<p>I've read somewhere the human eye can't dis... read somewhere human eye distinguish less valu...
<p>Yes, I thought about that, but I soon figur... yes thought soon figured another domain specif...
<p><a href="http://www.codeproject.com/Article... oleg shilo script solution code project really...
<p>would be a bit reluctant to use nested cl... would bit reluctant use nested classes created...
```

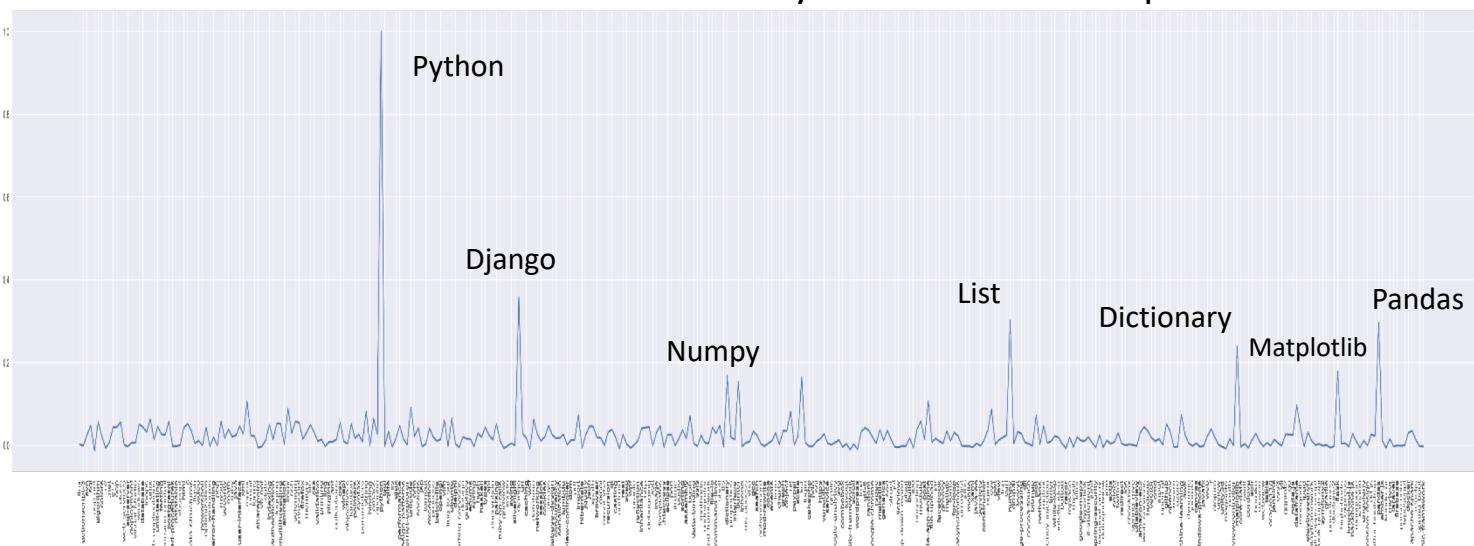
Topic Correlation

- By keeping track of Topics which were asked together in questions, a network relating languages and frameworks was created.
- Individual Topics were analyzed vs. all other Topics.
- Distribution of User question counts.

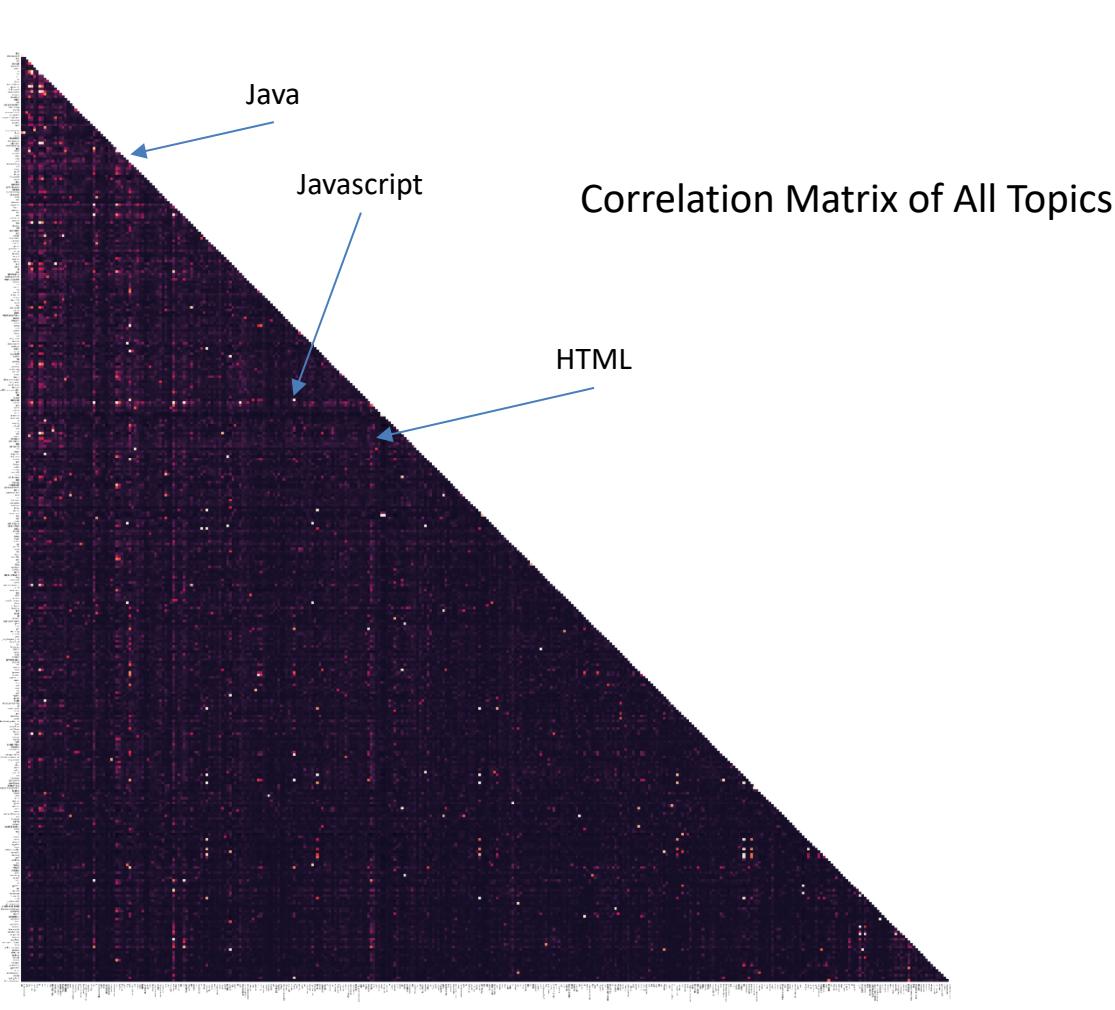
A Web of popular Topics asked on StackOverflow



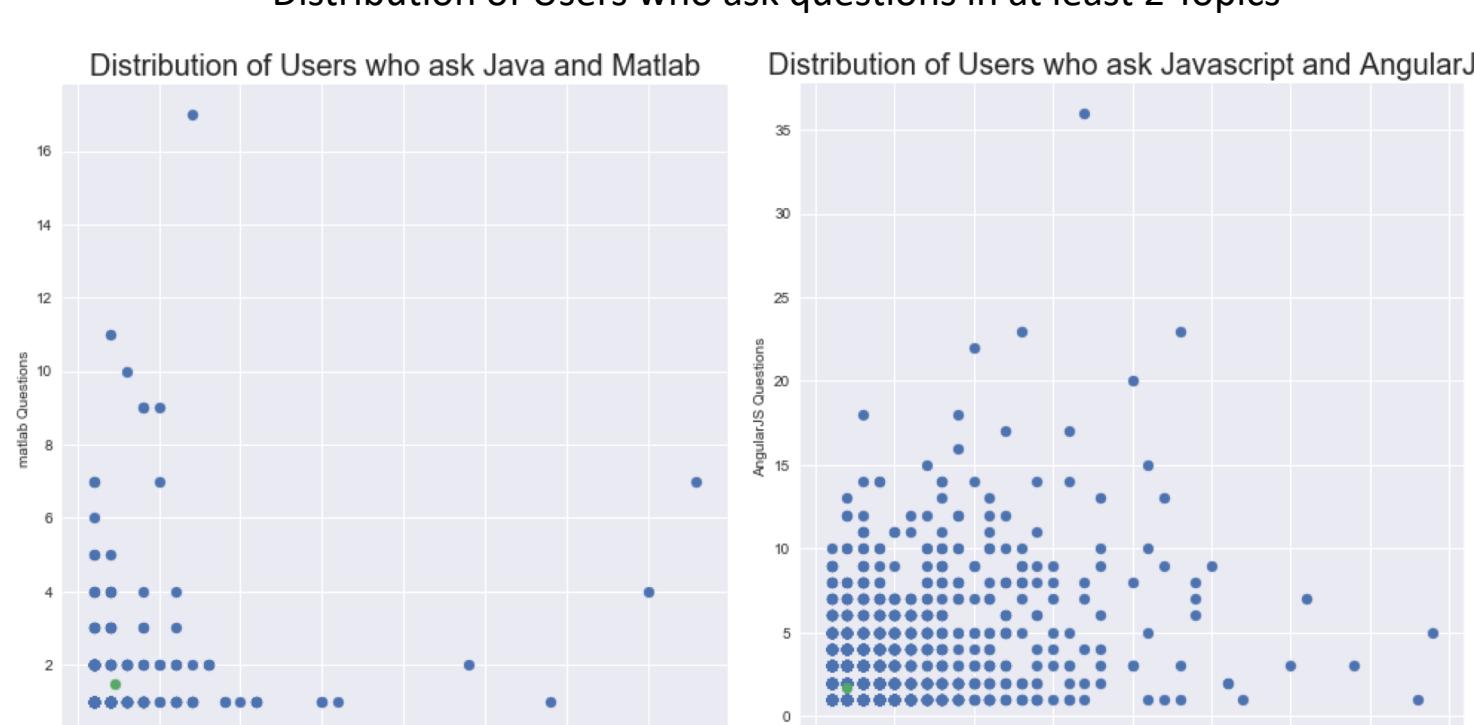
Correlation values of Python vs all other Topics



- Each User's questions and their Topics were used to profile the Users.
- Correlation Matrix used to pinpoint "Hand-in-Hand" Topics

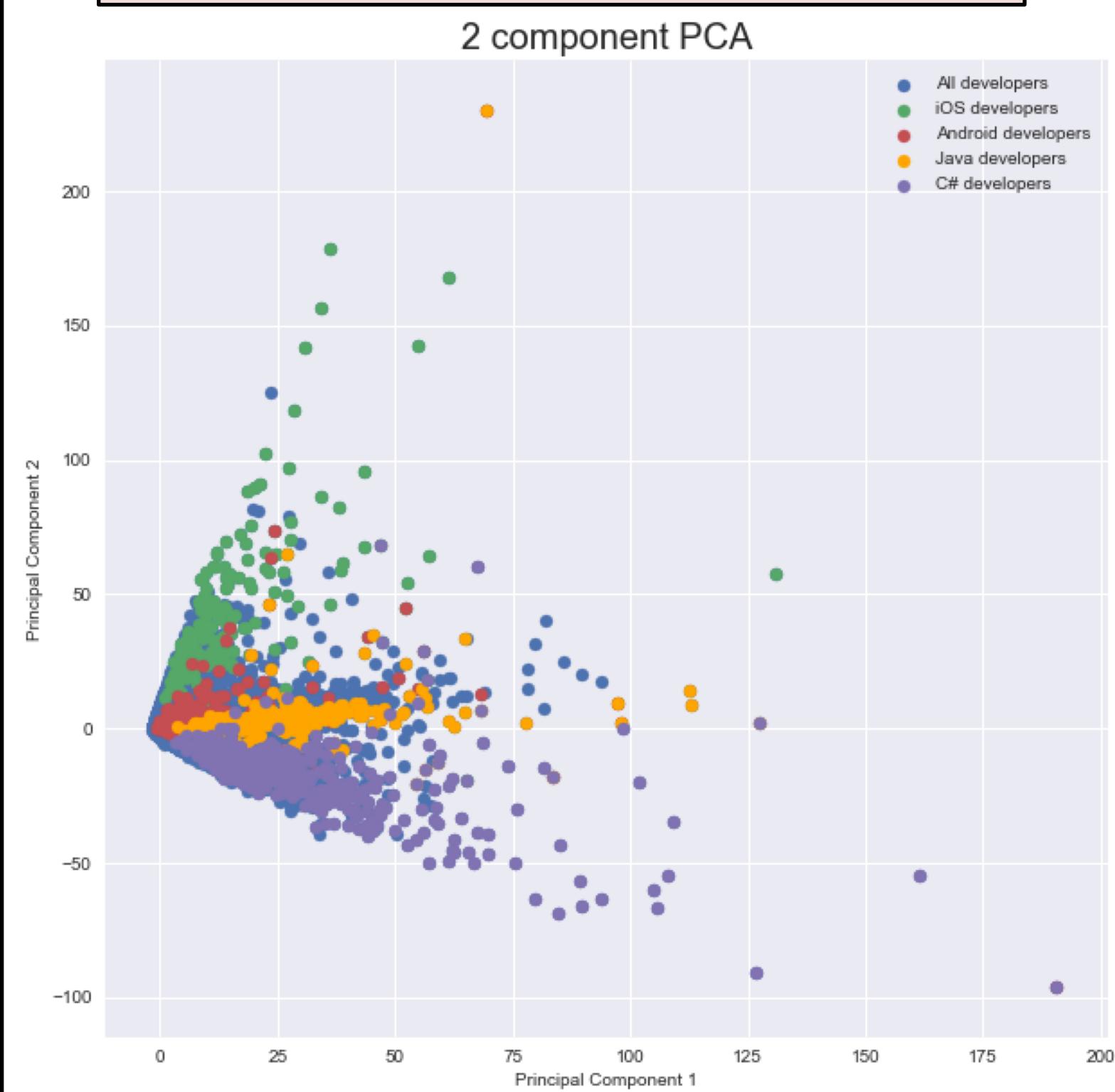


Distribution of Users who ask questions in at least 2 Topics



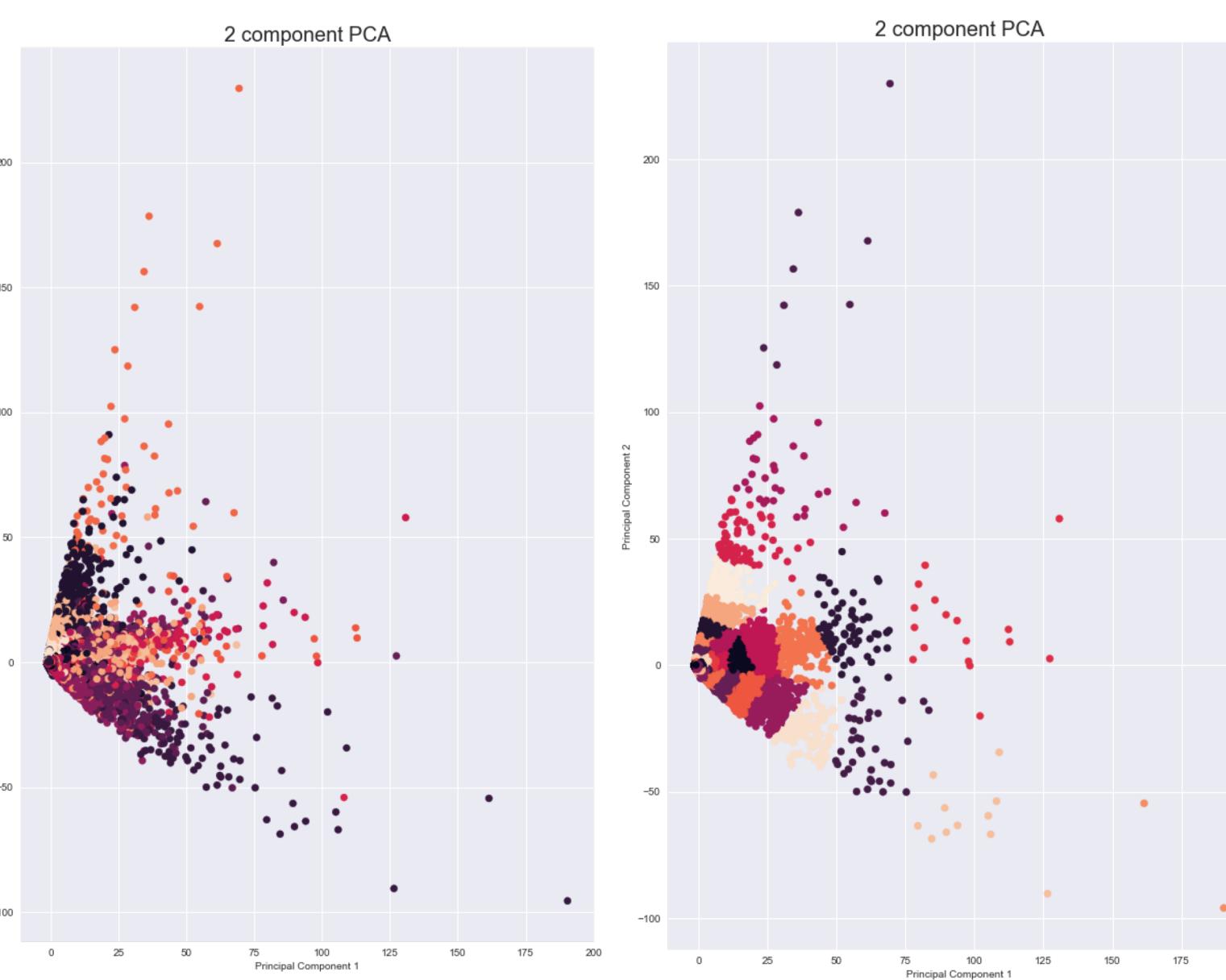
Principal Component Analysis

- The high-dimensional User points went through dimensional reduction via PCA
- Using only principal components σ_1 and σ_2



K-means clustering

- The high-dimensional User points went through dimensional reduction via PCA
- Using only principal components σ_1 and σ_2
- Clustering dimensional reduction vs. After reduction



- Clustering after PCA -> Not Clear Cut results.
- Clustering before PCA -> Unclear clusters, but clear results

Cluster First	PCA First
PHP, MySQL	PHP, Version Control, Math
Android, Java	Android, Ruby, Date, Indexing
Javascript, AngularJS, JQuery, HTML, node.js, CSS, PHP, Java	Javascript, Logging, Authentication, Sharepoint, File, Collections, Firefox, Selenium, SSH
Java, Spring, Hibernate	Java, Performance, Ajax, Unicode, Orm, Jsp, Parameters, Hash, Api
iOS, Objective-C	iOS, Haskell, Types, Tomcat, Exception, Sockets, Import, Dom, Pdf, Xpath, Excel
C#, Asp.net, Net, SQL-server, Cisual Studio	C#, Recursion, Rest, Soap, Labda, Assembly, Perl, HTTP, Perl, Sorting
C++, C, OpenCV	C++, Properties, iPhone, Search, Ms-Access, Search, Text, Random, CakePhp
Python, Django, Numpy, Pandas	Python, Module, Encoding, Build, Makefile, Struct, Curl, Ngninx

Buzz Word Detection

- TF-IDF weightage based buzz words.
- Compared sample questions with all questions asked in same topic and other random questions.
- Collected the top 5 highest weighted words in a question

AngularJS - How do I submit a form to a controller on the server?

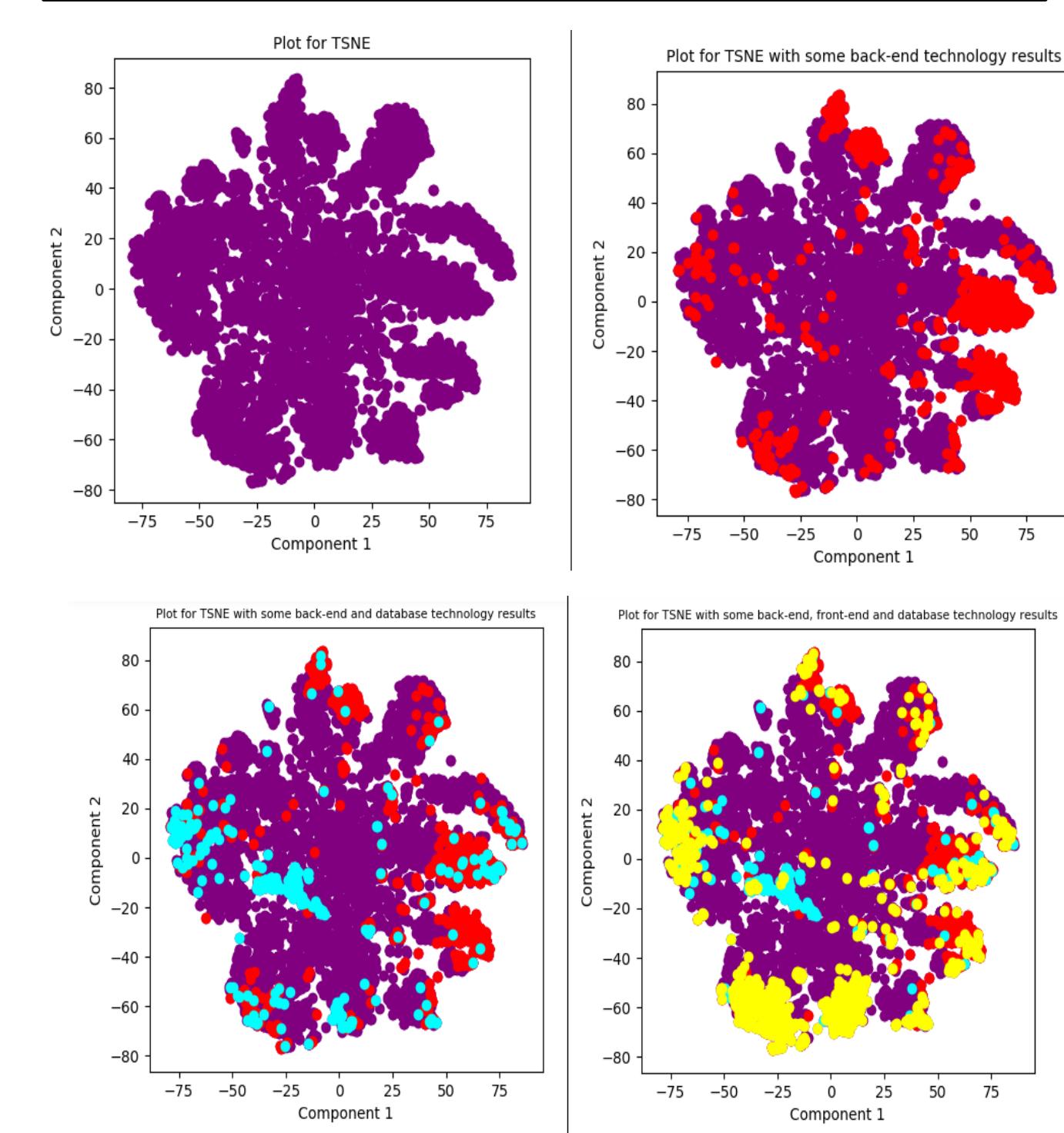
The cookbook form examples on the AngularJS site only save the state on the client. How do I submit to the server?
 Alternatively, how do I use jQuery's `form.submit()` on the `form` in the `ng:click="save()"` function?
 Edit - Found 2 ways to do this (I also removed the HTML markup I pasted before - just refer to the advanced `form` cookbook example for the source)
 1. <http://webpac2.r0t3.org:3000/conferenceWork> (by Dobrica Pavlinusic) to go the AngularJS way using a resource to send the data to the `server` in JSON format. I had issues with that on the `server` side - AngularJS was sending it fine but grails were mangling it (according to firebug and request content-length). I need to look into this more. How do I change the content-type in angular for a resource method like `$save()`?
 2. Put a `form` in and use a `submit` button. Since I am not doing a single page web app, I used this method. Most validations were on the client and a few more on the `server` which was sufficient for me.

Just putting this here so that someone else can use this for possible solutions and best approach.

['cookbook', 'submit', 'form', 'server', 'webpac2']

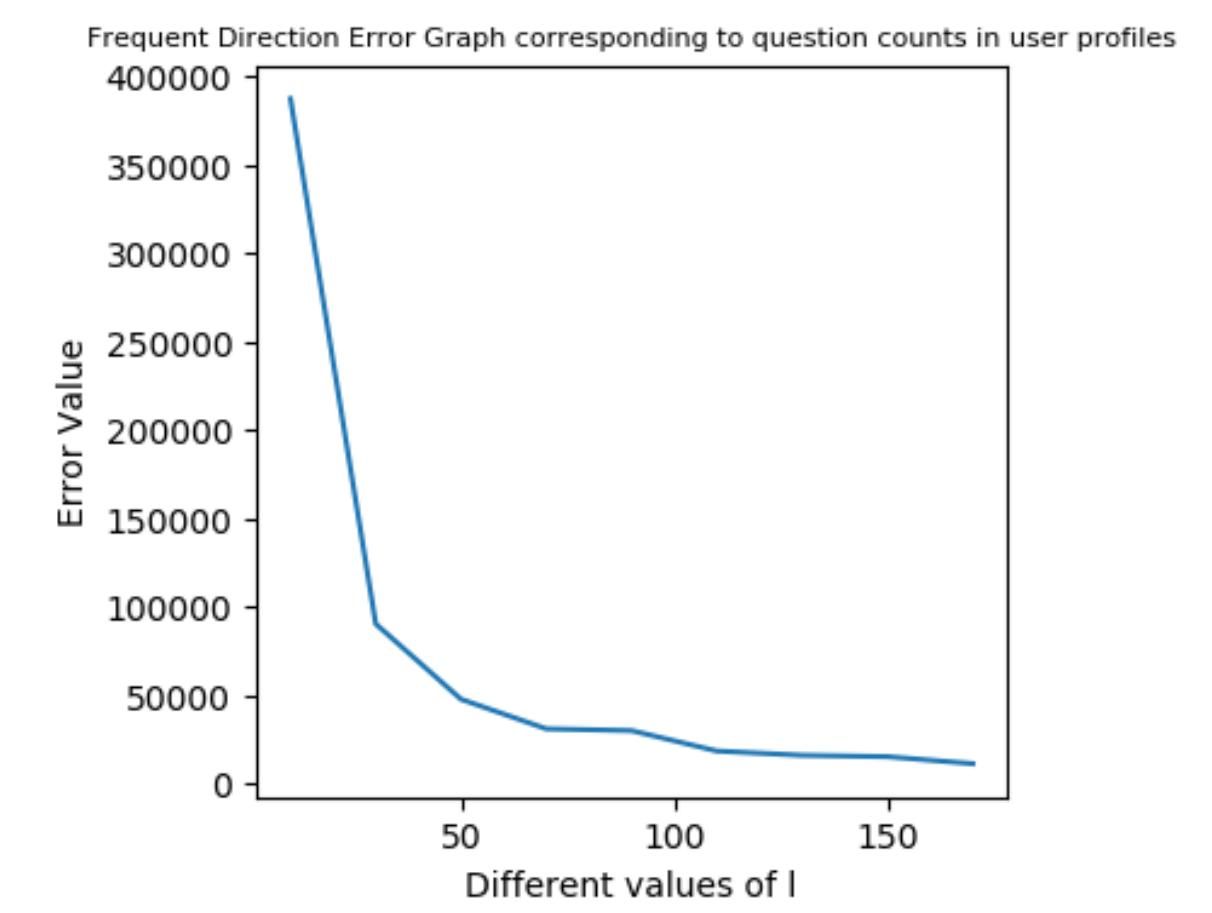
t-SNE

- Converting high-dimensional Euclidian distance to conditional probabilities representing similarities.
- Set of 12,000 data points were computed in one go.
- Several trials with different set of points were run.
- No clear intuition of what each cluster contained.



Frequent Directions

- Streaming algorithm to detect frequent directions.
- Directions magnitudes could tell which tags go together or do not.



- Following are the results obtained by frequent directions on $l=35$ so as to draw a comparison between clustering and frequent directions.

Positive Directions	Negative Directions
asp.net, c#, HTML, php, Java, Javascript, CSS, jquery, Android	N/A
php, Java, Python, Objective-C	asp.net, c#, .net, Visual Studio
asp.net, C#, Java	php, mySQL, Ruby, Javascript, CSS, Python, ajax, Ruby on Rails, AngularJS
php, Java	Ruby, Objective-C, Ruby on Rails, iPhone, iOS
Eclipse, Hibernate, Django, Swing, Ruby on Rails, Spring, Python, Multithreading, Java, C++	iOS, iPhone, Android, jquery, Objecctive-C, Javascript, php
Ruby, Ruby on Rails, Android	C++, Objective-C, Python, Xcode, iPhone, iOS, Swift, iPad
Java, Spring	C++, Python, Django, Pandas
JavaScript, jQuery, AngularJS	SQL, C++, C#, php, Ruby on Rails

- Variance of all the Columns throughout all the directions used was to see which languages are "polar"

