1. How do we discover that there are bot users in the data?

Telling facts:

a) We analysed 710637 tweets on 2016.6.21, which is 2 days before referendum

b) users are more activate in 6.21 than average day (see mean values, median, 75% quantile)

In bot identifier file, find:

```
data = pd.merge(data, df, how='outer', on=['user_id_str'])

data.describe()
```

	twitter_count_160621	tweets_per_day
mean	42.407495	15.538318
Median (50% quantile)	7.000000	4.136082
75% quantile	13.157115	35.000000

Users were sends 3 times as many tweets as they were in their average.

c) One user tweets 1417 tweets on that day. This is impossible for human to tweet. That's approximately 1 minute per tweet in 24 hours. (60 * 60 * 24) /1417 = 60.9

d) In 2016.6, 4215 new users have registered. This is above the value of registration of all month in 2014, 2015, and 2016.

In bot identifier file:

Data: created_count (pandas dataframe with 3 columns: year, month, count)

```
created_count = pd.DataFrame({'count' : idf.groupby(['user_year', 'user_month']).size()}).reset_index()
created_count
```

	user_year	user_month	count
0	2006	7	5
1	2006	8	4
2	2006	9	8

Visualization: Heatmap

Conclusion: there are bots in the data, and they have distorted the data (in particular, the opinions that real-people have).

2. What is the number of tweets sent by bot and number of tweets sent by human on 6.21

In bot_analysis and human_analysis file:

Data: proportion (float)

Visualization: pie chart, “proportion : proportion”

3. Where are bots coming from?

   In bot_analysis file:

   Data: geo_list (list of tuple, [(country code, frequency)])

   Visualization: World map (choropleth map)

4. What did the data say? What did the bots say? What did human say?

   a) Content of tweets

      We use most frequent (top 50) hashtags:

      In bot_analysis/human_analysis/160621_analysis file:

      Data:

      hashtags_dist.most_common(50) (list of tuple)

      leave_tags, remain_tags (list of string)

      Visualization: word cloud or treemap. Font size or Area could match the frequency of the word, and leave, remain can map to different colour.

   b) Sentiment variation

      i. Sentiment – the intensity of the subjective opinion on Brexit **(actually, we do not care about the sentiment value is positive or negative)**

      In bot_analysis/human_analysis/160621_analysis file:

      Data: sentiment_hour (json like object, i.e. python dictionary, with hour as keys, sentiment outputs (average 5 minutes) as values)

      Visualization: time series line graph

      ii. the intensity of the subjective opinion in a single number

      Data: sentiment_variation (float), standard deviation of the sentiment across 24 hours

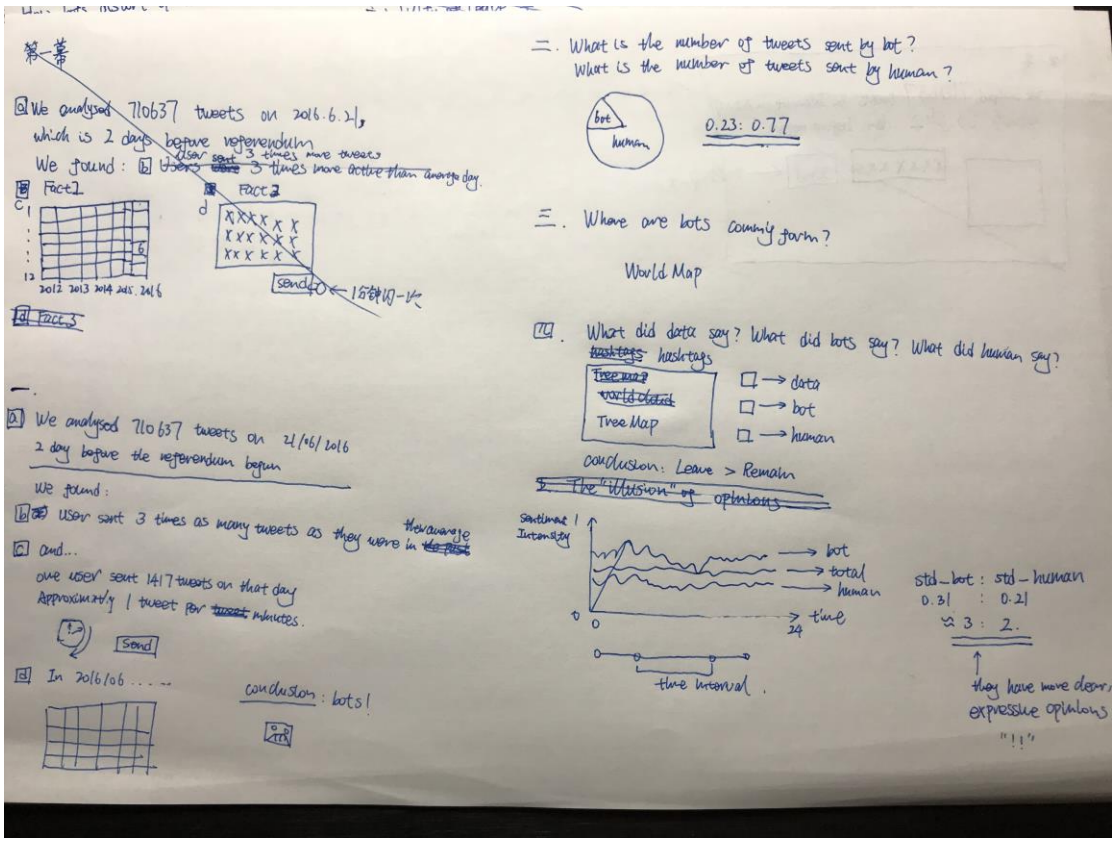
      Visualization: in text

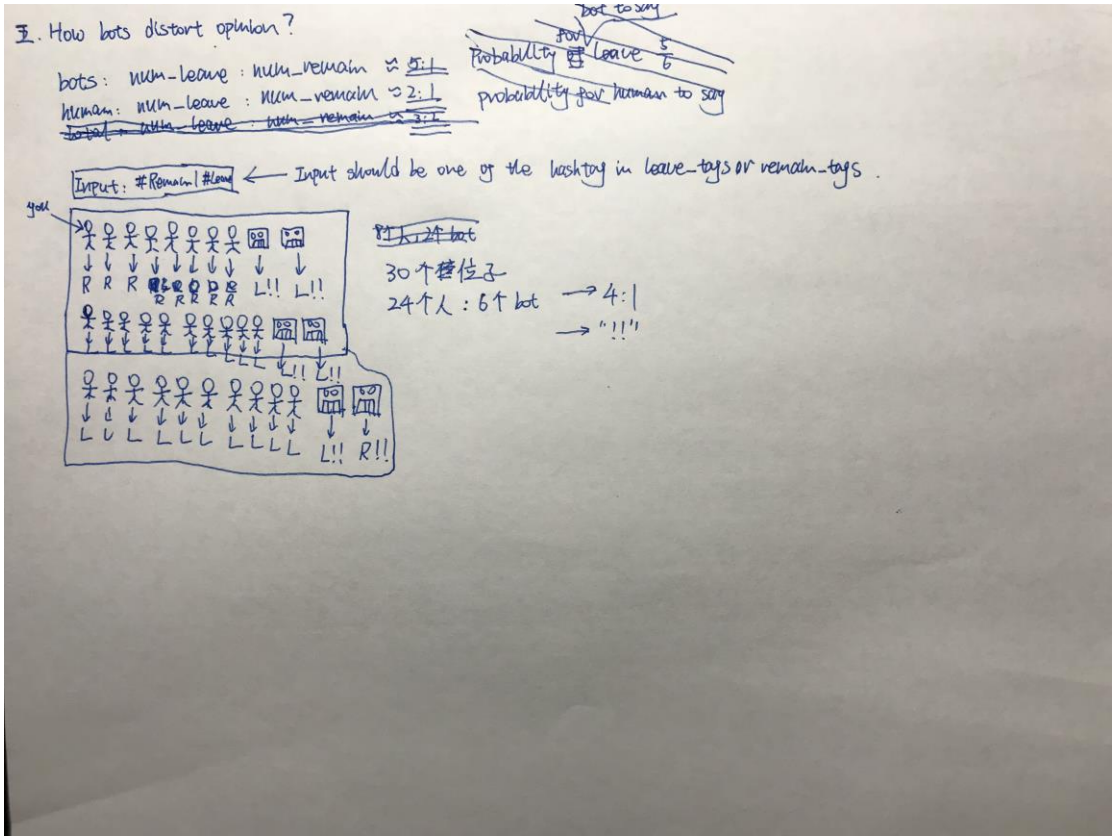
5. How bots distort the opinion?

   In bot_analysis/human_analysis file:

   Data: num_leave, num_remain (float), leave_tags, remain_tags (list of string)

   Visualization: interactive game – user input a hashtag in leave_tags or remain_tags to express his/her opinion. According to the ratio of num_leave, num_remain, other icons express their opinions (leave or remain) about referendum.


第一幕
We analysed 710637 tweets on 2016.6.24,
which is 2 days before referendum
We found:
Fact1
Fact2
Fact3
2012 2013 2014 2015 2016
send
15秒10一次
一.
We analysed 710637 tweets on 24/06/2016
2 day before the referendum begun
We found:
User sent 3 times as many tweets as they were in the average
and...
one user sent 1417 tweets on that day
Approximately 1 tweet per minutes.
Send
In 2016/06 ...
conclusion: bots!
二. What is the number of tweets sent by bot?
What is the number of tweets sent by human?
bot
human
0.23: 0.77
三. Where are bots coming from?
World Map
四. What did data say? What did bots say? What did human say?
hashtags
Tree Map
data
bot
human
conclusion: Leave > Remain
Sentiment / Intensity
bot
total
human
0
24
time
time interval.
std_bot : std_human
0.3 : 0.2
≈ 3 : 2.
they have more clear, expressive opinions
"!!"


五. How bots distort opinion?
bots: num_leave : num_remain ≈ 5:1
human: num_leave : num_remain ≈ 2:1
bot to say
for
Probability for Leave 5/6
probability for human to say
Input: #Remain | #Leave
Input should be one of the hashtag in leave_tags or remain_tags.
you
30个替位子
24个人：6个bot
→ 4:1
→ "!!"

6. (optional) How do we identified tweets that were sent by bots?

```
data['authenticity'] = np.where((((data['twitter_count_160621'] - data['tweets_per_day'])>20) & ((data['tweets_per_d
y']>50) | (data['twitter_count_160621']>35))), False, True)
```

a) User who send tweets far more than human can do (> 35 tweets per day) (approximately 1.48 tweets per hour)

b) User who sent tweets much more activate in 6.21 than their average tweet per day (difference > 20)