

# Linking the reactions to the corpus

We want to link the individual reactions from the Oct 3 2012 presidential debate to the corpus for the debate.

```
In [1]: import pandas as pd
        from pandas.tools.plotting import scatter_matrix
        import random
        import numpy as np
        import reactions
```

## Reactions data

Read in the Oct3 2012 debate reactions file and get the reactions part.

```
In [2]: %time parts = reactions.split_reactions_file('data/reactions_oct3_4project.csv')

CPU times: user 4.26 s, sys: 0.36 s, total: 4.61 s
Wall time: 4.62 s

In [3]: r = parts['reactions'].copy()
```

Add a *start* column to hold the time of the reaction; we will use this to merge with the corpus file later.

```
In [4]: r['start'] = r.Time.apply(lambda t: pd.datetime.time(t))

In [5]: r.head(2)
```

Out[5]:

	UserID	Time	Reaction_who	Reaction_what	start
0	ag1zfnJIYWN0bGFicy00ciwLEgRVc2VyliJhX2E0Mjc1MD...	2012-10-04 01:03:50.179000	Moderator	Agree	01:03:50.179000
1	ag1zfnJIYWN0bGFicy00ciwLEgRVc2VyliJhX2E0Mjc1MD...	2012-10-04 01:05:17.907000	Obama	Spin	01:05:17.907000

## Transcript corpus

Load the transcript file.

```
In [6]: c = pd.read_csv('corpora/oct3_coded_transcript_sync.csv')

Add a start column like we did for the reactions data.

In [7]: c['start'] = pd.to_datetime(c["Sync'd start"]).apply(lambda t: pd.datetime.time(t))

In [8]: c.head(2)
```

Out[8]:

	Sync'd start	Sync'd end	Speaker	Transcript	QuestionTopic	Topic	Frame	Tone	start
0	1:02:01	1:02:06	0	Good evening from the Magness Arena at the Uni...	99	9	9	0	01:02:01
1	1:02:06	1:02:09	0	I'm Jim Lehrer of the PBS NewsHour,	99	9	9	0	01:02:06

## Merge

The *start* column will be shared, but the times will be out of order, so sort it on *start*.

```
In [9]: m = c.append(r)

In [10]: m = m.sort(columns='start')

In [11]: m.head(2)
```

Out[11]:

	Frame	QuestionTopic	Reaction_what	Reaction_who	Speaker	Sync'd end	Sync'd start	Time	Tone	Topic	Transcript	UserID
56965	NaN	NaN	Disagree	Moderator	NaN	NaN	NaN	2012-10-04 00:30:00.429000	NaN	NaN	NaN	ag1zfnJIYWN0bGFicy01ciwLEgRVc2Vyli.
92001	NaN	NaN	Agree	Obama	NaN	NaN	NaN	2012-10-04 00:30:02.337000	NaN	NaN	NaN	ag1zfnJIYWN0bGFicy0xciwLEgRVc2Vyli.

Assuming the times are already synchronized and that reactions happen *after* a statement is made, we will assign reactions between two transcript entries to the *prior* entry. We do this by filling forward the blank columns of the table.

```
In [12]: m = m.fillna(method='ffill')
```

It appears that the app accepted clicks a 1/2 hour before the start of the debate (we will see that the same happened for a 1/2 hour at the end as well). Let's get rid of those reactions at the start, which contain NAs.

```
In [13]: m2 = m.dropna()
```

```
In [14]: print len(m), len(m2), len(m)-len(m2)

194458 193033 1425
```

So that removed ~1400 reactions.

Now notice that the last reaction happened at ~3pm and the last transcript record happened at ~2:30p. Let's get rid of those really late reactions.

```
In [15]: m2[["Sync'd start", "Time", "start", "UserID"]].tail(1)
```

Out[15]:

	Sync'd start	Time	start	UserID
141777	2:32:41	2012-10-04 02:59:59.396000	02:59:59.396000	ag1zfnJIYWN0bGFicy0yciwLEgRVc2VyliJhX2NIOTc1M2...

```
In [16]: m3 = m2[m2.start < pd.datetime.time(pd.to_datetime('2:33'))]
```

```
In [17]: m3[["Sync'd start", "Time", "start", "UserID"]].tail(1)
```

Out[17]:

	Sync'd start	Time	start	UserID
191634	2:32:41	2012-10-04 02:32:59.840000	02:32:59.840000	ag1zfnJIYWN0bGFicy0zciwLEgRVc2VyliJhXzBkNGVIZT...

## Turns

Let's find the turns, i.e. the contiguous records where the same speaker was speaking.

```
In [18]: m3['turn'] = (m3.Speaker.shift(1) != m3.Speaker).astype(int).cumsum()
```

```
In [19]: m3[['Speaker', 'Transcript', 'start', 'turn', 'Reaction_what', 'Reaction_who']].head(5)
```

Out[19]:

	Speaker	Transcript	start	turn	Reaction_what	Reaction_who
0	0	Good evening from the Magness Arena at the Uni...	01:02:01	1	Agree	Moderator
56861	0	Good evening from the Magness Arena at the Uni...	01:02:01.401000	1	Disagree	Moderator
60192	0	Good evening from the Magness Arena at the Uni...	01:02:02.011000	1	Disagree	Romney
117568	0	Good evening from the Magness Arena at the Uni...	01:02:02.268000	1	Agree	Moderator
48989	0	Good evening from the Magness Arena at the Uni...	01:02:02.482000	1	Disagree	Moderator

```
In [21]: m3[['Speaker', 'Transcript', 'start', 'turn', 'Reaction_what', 'Reaction_who']].tail(5)
```

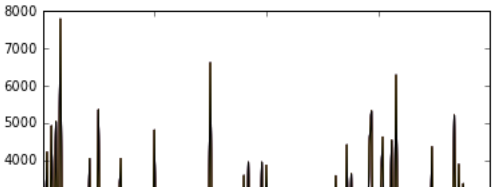
Out[21]:

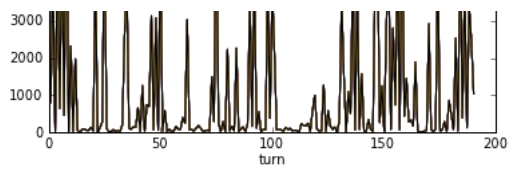
	Speaker	Transcript	start	turn	Reaction_what	Reaction_who
188581	0	Thank you, and good night.	02:32:58.844000	190	Disagree	Romney
178766	0	Thank you, and good night.	02:32:58.890000	190	Agree	Moderator
167646	0	Thank you, and good night.	02:32:59.280000	190	Agree	Romney
68397	0	Thank you, and good night.	02:32:59.726000	190	Disagree	Romney
191634	0	Thank you, and good night.	02:32:59.840000	190	Agree	Romney

It looks like early turns got lot of turns, then there was perhaps some dull spots, then more attention at the end.

```
In [22]: m3.groupby('turn').count().plot(legend=False)
```

```
Out[22]: <matplotlib.axes.AxesSubplot at 0x7aa8b90>
```





In [ ]: