# Reactions file from Oct 3, 2012 debate

We are looking at the debates reaction data for the Oct 3 2012 presidential debate.

## Load and cleanup

In [35]:
```
import pandas
from pandas.tools.plotting import scatter_matrix
import random
```

Read in the Oct3 2012 debate file.

In [3]:
```
%time reactions = pandas.read_csv('resources_symlink/data/reactions_oct3_4project.csv')
```

```
CPU times: user 1.82 s, sys: 0.16 s, total: 1.98 s
Wall time: 1.99 s
```

The columns are hard to read. Rename.

In [4]:
```
for i,c in enumerate(reactions.columns.tolist()): print c
```

```
UserID
Time
Reaction
<b>25.</b> How are you watching the debate? (It is especially important that we know if you are watching online, so please answer
accurately.)#
<br /><b>10.</b> Economy:#
<br /><b>12.</b> Health Care:#
<br /><b>13.</b> Foreign Policy:#
<br /><b>14.</b> Abortion:#
<br /><b>15.</b> Economy:#
<br /><b>7.</b> Health Care:#
<br /><b>8.</b> Foreign Policy:#
<br /><b>9.</b> Abortion:#
<hr><b>23.</b> Some people don't pay much attention to political campaigns.  How about you? How interested have you been in the political
campaigns (so far) this year?#
<hr><b>24.</b> What specific sources (newspapers, radio shows, television programs, websites) have you been using to get most of your
campaign and election news? Just name a few specific sources you use MOST often.#
<hr><br /><b>16.</b> What is your gender?#
<hr><br /><b>17.</b> What is your age?#
<hr><br /><b>18.</b> What is your family's income?#
<hr><br /><b>19.</b> What is your race?#
<hr><br /><b>20.</b> What is your religion?#
<hr><br /><b>21.</b> If Christian, please be more specific. (Otherwise, skip this question.)#
<hr><br /><b>22.</b> In what state do you live? (Use 2-letter abbreviation, e.g., CA)#
<hr><br /><b>26.</b> If you are watching on TV, exactly how are you watching?#
<hr><br /><b>27.</b> Thinking about the economy, which candidate do you prefer at the moment?#
<hr><br /><b>28.</b> Thinking about foreign policy, which candidate do you prefer at the moment?#
<hr><br /><b>29.</b> Which candidate do you prefer at the moment?#
<hr><br /><b>3.</b> If the election was today, which one of the candidates would you pick?#
<hr><br /><b>4.</b> If you selected a candidate in the question above, how confident are you that this candidate is the best candidate for
president? (Otherwise, skip this question.)#
<hr><br /><b>5.</b> How likely are you to vote in the 2012 presidential election?#
<hr><br /><i>Some of the following questions allow you to use a slider to register your answer anywhere you want along the scale.</i><br />
<br /><b>2.</b> How would you describe your overall views on politics in general?#
<hr><br />Now we're ready to watch the debate! Use the app to respond in real time by selecting a person and then clicking a response
button. If you see a question pop up, answer and then click Next. And don't forget to complete the brief follow-up questions that will
appear once the debate ends. Ready to begin?#
<hr>Please indicate the priority of each of the following issue areas to you personally.<br /><br /><b>6.</b> Immigration:#
<hr>Which political party, if any, do you think is better at dealing with each of these issues?<br /><br /><b>11.</b> Immigration:#
<i>Remember, scroll down to answer all questions, then be sure to click Next to move to the next page of questions.</i><br /><br /><b>1.
</b> Generally speaking, do you think of yourself as being closer to the Democratic Party or the Republican Party?#
This app allows you to react to the 2012 debates. We begin with a short pre-debate survey. Then, during the debate, additional questions
may appear on your screen; answer these questions and then continue using the app. At the end of the debate, you will see a brief set of
follow-up questions.  Please complete these questions -- we want to know what you think!  Students who have been offered credit for
participating MUST complete these follow-up questions (including your name and the Course ID you have been given, but remember all other
responses are anonymous).<br /><br />IMPORTANT: On each page, be sure to scroll down and answer all the questions.  After completing all
the questions on each page (including this one), be sure to click Next.  Note that it might take each page several seconds to load; please
be patient.#
```

In [5]:
```
r2 = reactions.copy()
```

In [6]:
```
r2.columns = [c.strip() for c in """
UserID
Time
Reaction
how_watching_25
economy_priority_10
health_care_party_12
foreign_policy_party_13
abortion_party_14
economy_party_15
health_care_priority_7
foreign_policy_priority_8
abortion_priority_9
```

```
interested_23
news_sources_24
gender_16
age_17
family_income_18
race_19
religion_20
christian_21
state_22
tv_channel_26
economy_candidate_27
foreign_policy_candidate_28
candidate_preferred_29
candidate_choice_3
confidence_in_choice_4
likely_to_vote_5
political_views_2
ready
immigration_priority_6
immigration_party_11
party_1
next
""".split('\n') if not c.strip()==''']
```

Reorder.

In [7]:
```
r3 = r2.reindex(columns=[c.strip() for c in """
UserID
Time
Reaction
party_1
political_views_2
candidate_choice_3
confidence_in_choice_4
likely_to_vote_5
immigration_priority_6
health_care_priority_7
foreign_policy_priority_8
abortion_priority_9
economy_priority_10
immigration_party_11
health_care_party_12
foreign_policy_party_13
abortion_party_14
economy_party_15
gender_16
age_17
family_income_18
race_19
religion_20
christian_21
state_22
interested_23
news_sources_24
how_watching_25
tv_channel_26
economy_candidate_27
foreign_policy_candidate_28
candidate_preferred_29
ready
next
""".split('\n') if not c.strip()==''])
```

In [8]:
```
r3[['UserID','Time','Reaction']].head()
```

Out[8]:

| | UserID | Time | Reaction |
|---|---|---|---|
| 0 | ag1zfnJlYWN0bGFicy00ciwLEgRVc2VyIiJhX2E0Mjc1MD... | Moderator:Agree | 2012-10-04 01:03:50.179 |
| 1 | ag1zfnJlYWN0bGFicy00ciwLEgRVc2VyIiJhX2E0Mjc1MD... | Obama:Spin | 2012-10-04 01:05:17.907 |
| 2 | ag1zfnJlYWN0bGFicy00ciwLEgRVc2VyIiJhX2E0Mjc1MD... | Obama:Dodge | 2012-10-04 01:05:50.337 |
| 3 | ag1zfnJlYWN0bGFicy00ciwLEgRVc2VyIiJhX2E0Mjc1MD... | Obama:Agree | 2012-10-04 01:06:18.192 |
| 4 | ag1zfnJlYWN0bGFicy00ciwLEgRVc2VyIiJhX2E0Mjc1MD... | Romney:Spin | 2012-10-04 01:07:08.096 |

## The questionnaire part

Some of the questionaire questions are difficult to interpret. Also, the answers are duplicated for each reaction.

In [9]:
```
questionnaire = r3[[c.strip() for c in """
UserID
party_1
political_views_2
candidate_choice_3
confidence_in_choice_4
```

```
likely_to_vote_5
immigration_priority_6
health_care_priority_7
foreign_policy_priority_8
abortion_priority_9
economy_priority_10
immigration_party_11
health_care_party_12
foreign_policy_party_13
abortion_party_14
economy_party_15
gender_16
age_17
family_income_18
race_19
religion_20
christian_21
state_22
interested_23
news_sources_24
how_watching_25
tv_channel_26
economy_candidate_27
foreign_policy_candidate_28
candidate_preferred_29
ready
next
""".split('\n') if not c.strip()=='']]
```

In [10]:
```
len(questionnaire)
```

Out[10]: 193286

In [11]:
```
%time q2 = questionnaire.drop_duplicates()
len(q2)
```

```
CPU times: user 1.15 s, sys: 0.07 s, total: 1.21 s
Wall time: 1.22 s
```

Out[11]: 3767

So, there were ~190k reactions by 3767 users.

### Questionnaire: demographics

We look at the demographic parts of the questionnaire.

In [12]:
```
demos = q2[[c.strip() for c in """
gender_16
age_17
family_income_18
race_19
religion_20
christian_21
state_22
""".split('\n') if not c.strip()=='']]
```

In [13]:
```
demos.head()
```

Out[13]:

| | gender_16 | age_17 | family_income_18 | race_19 | religion_20 | christian_21 | state_22 |
|---|---|---|---|---|---|---|---|
| **0** | female | 18-24 | $100,000 or above | white/caucasian | christian (see next question) | catholic | pa |
| **62** | male | 18-24 | $25,000 – 49,999 | hispanic | christian (see next question) | catholic | tx |
| **70** | male | 18-24 | $100,000 or above | no answer | none (e.g., atheist, agnostic) | no answer | tx |
| **105** | no answer | no answer | no answer | no answer | no answer | no answer | no answer |
| **121** | female | 18-24 | no answer | white/caucasian | christian (see next question) | catholic | tx |

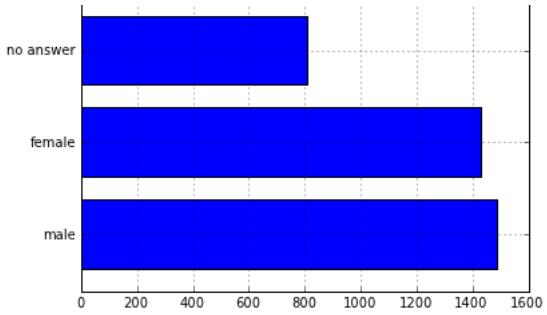
In [14]:
```
demos.describe()
```

Out[14]:

| | gender_16 | age_17 | family_income_18 | race_19 | religion_20 | christian_21 | state_22 |
|---|---|---|---|---|---|---|---|
| **count** | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 |
| **unique** | 3 | 6 | 6 | 6 | 6 | 6 | 86 |
| **top** | male | 18-24 | $100,000 or above | white/caucasian | christian (see next question) | no answer | no answer |
| **freq** | 1485 | 2511 | 1014 | 1890 | 1764 | 2062 | 769 |

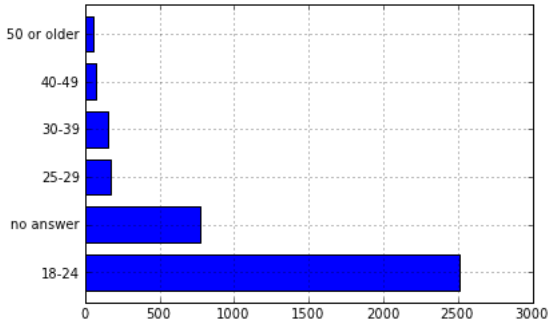
In [15]:
```
demos.gender_16.value_counts().plot(kind='barh')
```

Out[15]: <matplotlib.axes.AxesSubplot at 0x9a3d330>



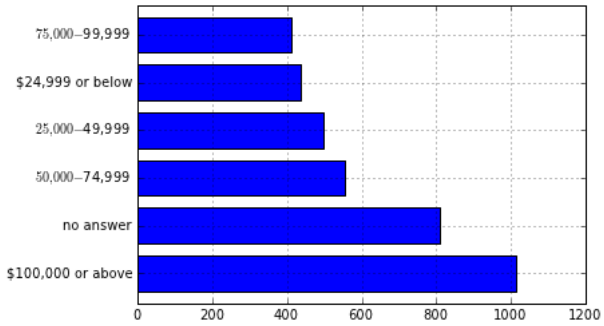

In [16]:
```
demos.age_17.value_counts().plot(kind='barh')
```

Out[16]: <matplotlib.axes.AxesSubplot at 0x88299f0>



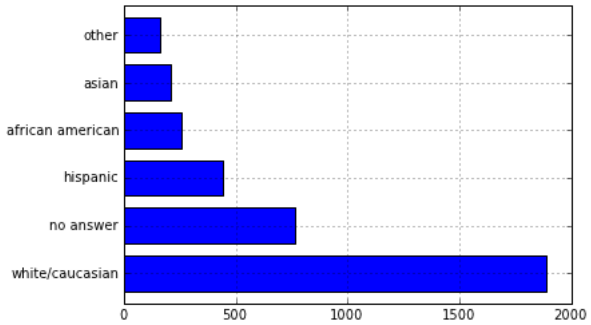

In [17]:
```
demos.family_income_18.value_counts().plot(kind='barh')
```

Out[17]: <matplotlib.axes.AxesSubplot at 0x7f74b70>



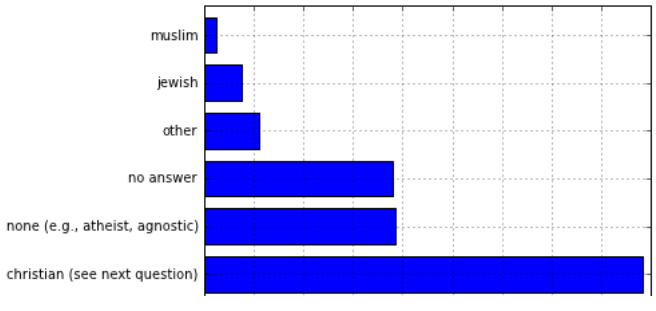

In [18]:
```
demos.race_19.value_counts().plot(kind='barh')
```

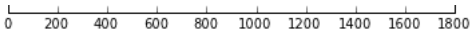
Out[18]: <matplotlib.axes.AxesSubplot at 0x7d65e10>




In [19]:
```
demos.religion_20.value_counts().plot(kind='barh')
```

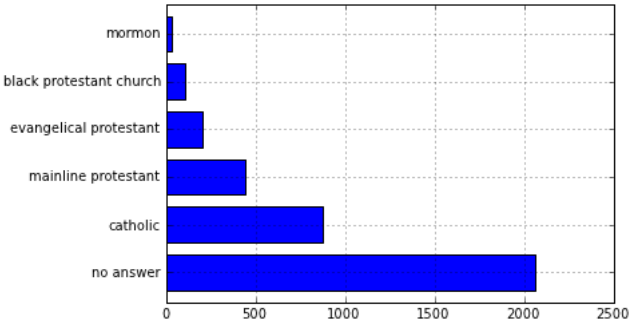
Out[19]: <matplotlib.axes.AxesSubplot at 0x7d322d0>





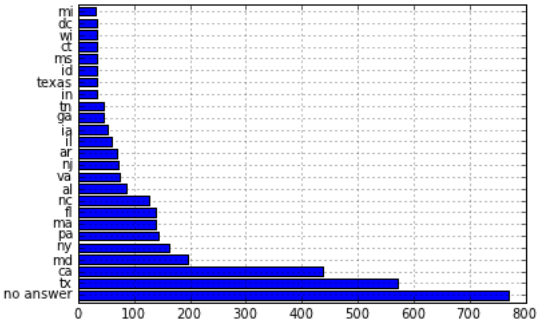

In [20]:
```
demos.christian_21.value_counts().plot(kind='barh')
```

Out[20]: <matplotlib.axes.AxesSubplot at 0x4e56090>




In [21]:
```
demos.state_22.value_counts()[:25].plot(kind='barh')
```

Out[21]: <matplotlib.axes.AxesSubplot at 0x547ddf0>




### Questionnaire: political

Now the part of the questionnaire with political questions.

In [22]:
```
pols = questionnaire[[c.strip() for c in """
party_1
political_views_2
candidate_choice_3
confidence_in_choice_4
likely_to_vote_5
immigration_priority_6
health_care_priority_7
foreign_policy_priority_8
abortion_priority_9
economy_priority_10
immigration_party_11
health_care_party_12
foreign_policy_party_13
abortion_party_14
economy_party_15
interested_23
news_sources_24
economy_candidate_27
foreign_policy_candidate_28
candidate_preferred_29
""".split('\n') if not c.strip()=='']]
```

In [23]:
```
pols.head()
```

Out[23]:

| | party_1 | political_views_2 | candidate_choice_3 | confidence_in_choice_4 | likely_to_vote_5 | immigration_priority_6 | health_care_priority_7 | foreign_policy_priority_8 |
|---|---|---|---|---|---|---|---|---|
| **0** | closest to republican party | 73 | romney | 100 | 100 | 72 | 99 | 100 |
| **1** | closest to republican party | 73 | romney | 100 | 100 | 72 | 99 | 100 |
| **2** | closest to republican party | 73 | romney | 100 | 100 | 72 | 99 | 100 |
| **3** | closest to republican party | 73 | romney | 100 | 100 | 72 | 99 | 100 |
| | closest to | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | republican party | 73 | romney | 100 | 100 | 72 | 99 | 100 |

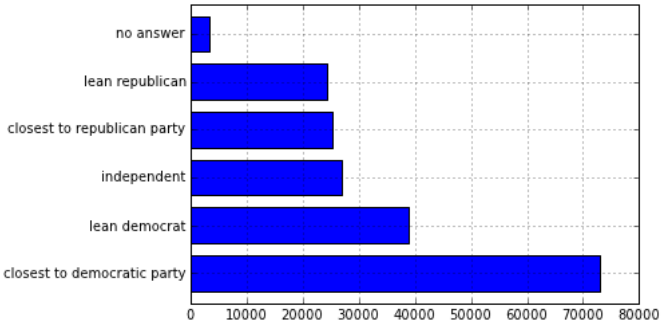
In [24]:
```
pols.describe()
```

Out[24]:

| | political_views_2 | confidence_in_choice_4 | likely_to_vote_5 | immigration_priority_6 | health_care_priority_7 | foreign_policy_priority_8 | abortion_priority_9 | econor |
|---|---|---|---|---|---|---|---|---|
| **count** | 191841.000000 | 191841.000000 | 191841.000000 | 191841.000000 | 191841.000000 | 191841.000000 | 191841.000000 | 191841 |
| **mean** | 40.546567 | 75.755011 | 87.692302 | 63.013485 | 77.507128 | 75.696139 | 60.767985 | 86.688 |
| **std** | 25.575280 | 22.215299 | 26.482706 | 23.840604 | 20.024258 | 19.675739 | 31.344281 | 15.996 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 |
| **25%** | 21.000000 | 59.000000 | 95.000000 | 50.000000 | 66.000000 | 64.000000 | 42.000000 | 78.000 |
| **50%** | 39.000000 | 79.000000 | 100.000000 | 65.000000 | 79.000000 | 77.000000 | 64.000000 | 93.000 |
| **75%** | 55.000000 | 99.000000 | 100.000000 | 79.000000 | 97.000000 | 94.000000 | 88.000000 | 100.00 |
| **max** | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.00 |

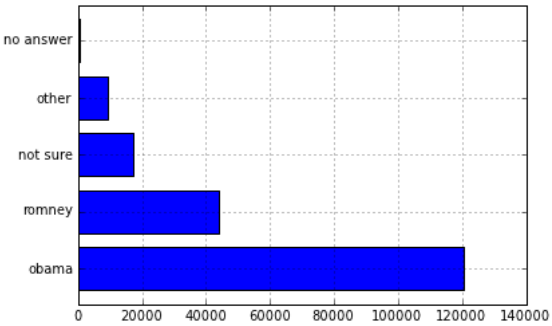
In [25]:
```
pols.party_1.value_counts().plot(kind='barh')
```

Out[25]: <matplotlib.axes.AxesSubplot at 0x5639bb0>



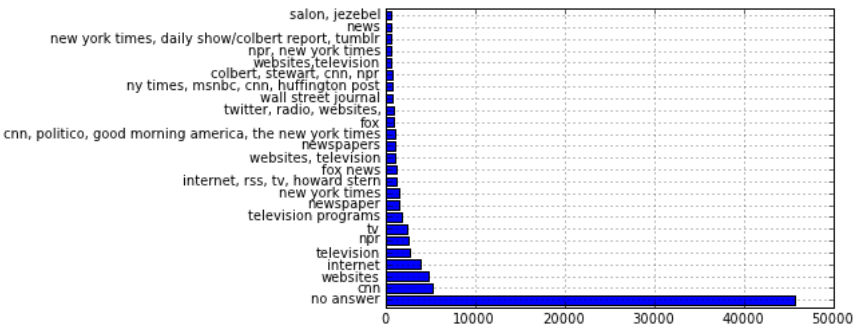

In [26]:
```
pols.candidate_choice_3.value_counts().plot(kind='barh')
```

Out[26]: <matplotlib.axes.AxesSubplot at 0x581c070>




In [27]:
```
pols.news_sources_24.value_counts()[:25].plot(kind='barh')
```

Out[27]: <matplotlib.axes.AxesSubplot at 0x5693870>



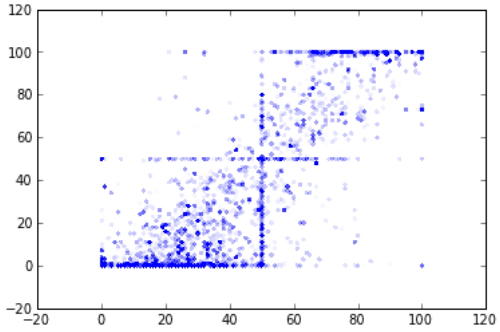

## Interpreting questionnaire answers that are given with a slider

#### What does a high or low value in the party-specific slider questions mean?

In [52]:
```
%time scatter(pols.political_views_2, pols.candidate_preferred_29, alpha=1.0/400)
```

```
CPU times: user 1.24 s, sys: 0.01 s, total: 1.25 s
Wall time: 1.25 s
```

Out[52]: <matplotlib.collections.PathCollection at 0x1df82f50>




Clearly the political views slider correlates with the candidate slider, which suggests that the conservative choice is always on one side and the liberal on the other, for example.

In [61]:
```
p2 = pols[[c.strip() for c in """
party_1
political_views_2
foreign_policy_party_13
abortion_party_14
economy_party_15
economy_candidate_27
foreign_policy_candidate_28
candidate_preferred_29
""".split('\n') if not c.strip()=='']].groupby('party_1')
```

In [62]:
```
p2.head(2)
```

Out[62]:

| | | party_1 | political_views_2 | foreign_policy_party_13 | abortion_party_14 | economy_party_15 | economy_candidate_27 | foreign_policy_candidate_28 |
|---|---|---|---|---|---|---|---|---|
| **party_1** | | | | | | | | |
| **closest to democratic party** | **62** | closest to democratic party | 20 | 23 | 50 | 27 | NaN | 28 |
| | **63** | closest to democratic party | 20 | 23 | 50 | 27 | NaN | 28 |
| **closest to republican party** | **0** | closest to republican party | 73 | 100 | 56 | 100 | NaN | 100 |
| | **1** | closest to republican party | 73 | 100 | 56 | 100 | NaN | 100 |
| **independent** | **581** | independent | 23 | 5 | 5 | 4 | NaN | 50 |
| | **582** | independent | 23 | 5 | 5 | 4 | NaN | 50 |
| **lean democrat** | **105** | lean democrat | 34 | 31 | 10 | 21 | NaN | 0 |
| | **106** | lean democrat | 34 | 31 | 10 | 21 | NaN | 0 |
| **lean republican** | **168** | lean republican | 63 | 78 | 46 | 90 | NaN | 81 |
| | **169** | lean republican | 63 | 78 | 46 | 90 | NaN | 81 |
| **no answer** | **121** | no answer | 58 | 79 | 0 | 27 | NaN | 50 |
| | **122** | no answer | 58 | 79 | 0 | 27 | NaN | 50 |

In [63]:
```
p2.aggregate(np.average)
```

Out[63]:

| | political_views_2 | foreign_policy_party_13 | abortion_party_14 | economy_party_15 | economy_candidate_27 | foreign_policy_candidate_28 | candidate_preferr |
|---|---|---|---|---|---|---|---|
| **party_1** | | | | | | | |
| **closest to democratic party** | 20.332649 | 24.421116 | 14.377450 | 21.843162 | NaN | NaN | NaN |
| **closest to republican party** | 79.671523 | 80.305494 | 68.528184 | 87.919181 | NaN | NaN | NaN |
| **independent** | 46.119535 | 48.697849 | 40.619739 | 52.364417 | NaN | NaN | NaN |
| **lean democrat** | 35.226116 | 36.128599 | 24.834810 | 36.727172 | NaN | NaN | NaN |
| **lean republican** | 62.788921 | 64.867993 | 47.917152 | 74.031268 | NaN | NaN | NaN |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **no answer** | 42.362053 | 43.073551 | 32.219754 | 42.798259 | NaN | NaN | NaN |

Clearly, the slider was set such that the high number (close to 100) was the republican choice and the low number (close to 0) was the democratic choice.