

donnees_anonymisees

April 22, 2022

1 Données anonymisées

Ces données ont été générées par une quarantaine de personnes et sont mises à disposition une fois anonymisées.

```
[1]: from jupyterhelper import add_notebook_menu
      add_notebook_menu()
```

```
[1]: <IPython.core.display.HTML object>
```

```
[2]: %matplotlib inline
```

1.1 Observations brutes

```
[3]: import pandas
      df = pandas.read_csv("logs/qcm100.txt", encoding="utf-8")
      df.head()
```

```
[3]:  -a0  -a1  -a2  events  game  ml_french_qcm-0-a0  \
0  NaN  NaN  NaN    NaN    NaN    NaN
1  NaN  NaN  NaN -a0,on  simple_french_qcm    NaN
2  NaN  NaN  NaN    NaN    NaN    NaN
3  NaN  NaN  NaN -a2,on  simple_french_qcm    NaN
4  NaN  NaN  NaN    NaN    NaN    NaN

      ml_french_qcm-0-a2  ml_french_qcm-0-a3  ml_french_qcm-0-b  \
0                NaN                NaN                NaN
1                NaN                NaN                NaN
2                NaN                NaN                NaN
3                NaN                NaN                NaN
4                NaN                NaN                NaN

      ml_french_qcm-0-duration  ...  simple_french_qcm-7-a1  \
0                NaN  ...                NaN
1                NaN  ...                NaN
2                NaN  ...                NaN
3                NaN  ...                NaN
4                NaN  ...                NaN

      simple_french_qcm-7-a2  simple_french_qcm-7-b  simple_french_qcm-7-duration  \
0                NaN                NaN                NaN
```

```

1           NaN           NaN           NaN
2           NaN           NaN           NaN
3           NaN           NaN           NaN
4           NaN           NaN           NaN

simple_french_qcm-7-nbvisit simple_french_qcm-8-ANS simple_french_qcm-8-b \
0           NaN           NaN           NaN
1           NaN           NaN           NaN
2           NaN           NaN           NaN
3           NaN           NaN           NaN
4           NaN           NaN           NaN

simple_french_qcm-8-duration simple_french_qcm-8-nbvisit \
0           NaN           NaN
1           NaN           NaN
2           NaN           NaN
3           NaN           NaN
4           NaN           NaN

time
0 2018-12-12 17:56:29.989
1 2018-12-12 17:56:34.255
2 2018-12-12 17:56:34.302
3 2018-12-12 17:56:37.645
4 2018-12-12 17:56:37.677

```

[5 rows x 168 columns]

1.2 Préparation des données

```
[4]: df2 = df[df.qtime == 'end'].copy()
df2.head().T.head()
```

```
[4]:
-a0           1           3           5 \
-a0           NaN           NaN           NaN
-a1           NaN           NaN           NaN
-a2           NaN           NaN           NaN
events      -a0,on      -a2,on      -a2,on
game  simple_french_qcm  simple_french_qcm  simple_french_qcm

           7           9
-a0           NaN           NaN
-a1           NaN           NaN
-a2           NaN           NaN
events  -a2,on-a2,on-a3,on-a2,on      -a2,on
game      simple_french_qcm  simple_french_qcm

```

1.3 Premier jeu : simple_french_qcm

```
[5]: person_id = ['person_id']
```

```
[6]: cols = person_id + [c for c in df2.columns if "simple_french_qcm" in c and '-a' in c]
df_question = df2[cols]
cols = person_id + [c for c in df2.columns if "simple_french_qcm" in c and '-b' in c]
df_bouton = df2[cols]
cols = person_id + [c for c in df2.columns if "simple_french_qcm" in c and '-nb' in c]
df_visit = df2[cols]
cols = person_id + [c for c in df2.columns if "simple_french_qcm" in c and '-ANS' in c]
df_ans = df2[cols]
cols = person_id + [c for c in df2.columns if "simple_french_qcm" in c and '-dur' in c]
df_dur = df2[cols]
```

```
[7]: df_question.shape
```

```
[7]: (465, 31)
```

```
[8]: df_dur.head().T
```

```
[8]:
person_id                c241c15008614ea67480
simple_french_qcm-0-duration  0 days 00:00:04.266000000
simple_french_qcm-1-duration                NaN
simple_french_qcm-2-duration                NaN
simple_french_qcm-3-duration                NaN
simple_french_qcm-4-duration                NaN
simple_french_qcm-5-duration                NaN
simple_french_qcm-6-duration                NaN
simple_french_qcm-7-duration                NaN
simple_french_qcm-8-duration                NaN

person_id                32606f02fa0df6aac111
simple_french_qcm-0-duration                NaN
simple_french_qcm-1-duration  0 days 00:00:03.343000000
simple_french_qcm-2-duration                NaN
simple_french_qcm-3-duration                NaN
simple_french_qcm-4-duration                NaN
simple_french_qcm-5-duration                NaN
simple_french_qcm-6-duration                NaN
simple_french_qcm-7-duration                NaN
simple_french_qcm-8-duration                NaN

person_id                32606f02fa0df6aac111
simple_french_qcm-0-duration                NaN
simple_french_qcm-1-duration                NaN
simple_french_qcm-2-duration  0 days 00:00:06.750000000
simple_french_qcm-3-duration                NaN
simple_french_qcm-4-duration                NaN
simple_french_qcm-5-duration                NaN
simple_french_qcm-6-duration                NaN
simple_french_qcm-7-duration                NaN
simple_french_qcm-8-duration                NaN
```

```
7 \
```

```

person_id          32606f02fa0df6aac111
simple_french_qcm-0-duration      NaN
simple_french_qcm-1-duration      NaN
simple_french_qcm-2-duration      NaN
simple_french_qcm-3-duration      1 days 00:00:00.000000000
simple_french_qcm-4-duration      NaN
simple_french_qcm-5-duration      NaN
simple_french_qcm-6-duration      NaN
simple_french_qcm-7-duration      NaN
simple_french_qcm-8-duration      NaN

```

9

```

person_id          32606f02fa0df6aac111
simple_french_qcm-0-duration      NaN
simple_french_qcm-1-duration      NaN
simple_french_qcm-2-duration      NaN
simple_french_qcm-3-duration      NaN
simple_french_qcm-4-duration      0 days 00:00:09.688000000
simple_french_qcm-5-duration      NaN
simple_french_qcm-6-duration      NaN
simple_french_qcm-7-duration      NaN
simple_french_qcm-8-duration      NaN

```

```

[9]: import numpy

def aggotnan_serie(values):
    res = []
    for v in values:
        if isinstance(v, float) and numpy.isnan(v):
            continue
        if pandas.isnull(v):
            continue
        if v in ('ok', 'on'):
            v = 1
        elif v == 'skip':
            v = 1000
        res.append(v)
    if len(res) > 0:
        if isinstance(res[0], str):
            r = ",".join(str(_) for _ in res)
        else:
            if len(res) == 1:
                r = res[0]
            else:
                try:
                    r = sum(res)
                except:
                    r = 0
    else:
        r = numpy.nan
    return r

```

```

def aggotnan(values):
    if isinstance(values, pandas.core.series.Series):
        r = aggotnan_serie(values)
        return r
    else:
        res = []
        for col in values.columns:
            val = list(values[col])
            res.append(aggotnan_serie(val))
        df = pandas.DataFrame(res, columns)
        return df

gr_ans = df_ans.groupby("person_id").agg(aggotnan)
gr_ans

```

[9]:

simple_french_qcm-8-ANS

person_id	
0c3a2f07b272478e4cd8	Thalès !
0cf4e678b9586d5b0e37	Pythagore
0fa2c1f9be1abc2c1d0b	Thalès
10fb62c55bc7a6a5cda2	
14db30224001f714a266	Thales
26310ba7a08a6b652fad	T
2c987776b9f528f4eeec	Pythagore
2f2bc77792dacadebd46	NaN
30035b23562d5b2df36f	
32606f02fa0df6aac111	NaN
33e5a65b534574acfb6e	
33eada5fee0e4231d402	Pythagore
3747fb5233006d3805da	
43637590b1591346cfc7	Pythagore
4a09796a11ccc59ef66a	
4b2cc1988785a7526ca1	Astérix et Cléopâtre
50670df89db35a7ae946	Pythagore
5102d6da7b6ed2c2690c	Thales
5a24afa1f3a84660b6d7	Pythagore
5a617c2783baa779527f	NaN
69fc4cf871b5db6fca3a	NaN
6a5bf88f39630d05d66c	Thales
753a507a9205e4c196fa	Prout
75a639141c325f39a368	
75b801f977bc69f8a34b	
8a8c40ad28eb1206efd5	
8e09880e917ca37a12dd	Pythagore, il était pété et il s'est exclamé ...
8f84e052a4513ed0dd80	Pythagore
93087555ec15d71b0da1	euler
afe4be0559208b09e577	Thalès
b0316cecc64ba1c1bd87	Thales
b07297f205b22a5d42d5	Pythagore
b0d3a8fd2b67439fba05	Pythagore
c241c15008614ea67480	thalès
c25deb2067e5a89b0223	Pythagore
d084aaa43b1793c01dd2	Pythagore

d228121baddcfea28ec2
d36d229a6a41414c6b26
dc1f7980a1e746469861
e2cf842295864a4d2620
eed744add9c2737812ad

Pythagore
Pythagore

Pytagore
Foucault

```
[10]: gr_dur = df_dur.groupby("person_id").agg(aggnotnan)
      gr_dur.head().T
```

```
[10]: person_id          0c3a2f07b272478e4cd8 \
      simple_french_qcm-0-duration  0 days 00:00:05.094000000
      simple_french_qcm-1-duration  0 days 00:00:04.359000000
      simple_french_qcm-2-duration  0 days 00:00:26.656000000
      simple_french_qcm-3-duration  0 days 00:00:08.125000000
      simple_french_qcm-4-duration  0 days 00:01:29.203000000
      simple_french_qcm-5-duration  0 days 00:00:08.047000000
      simple_french_qcm-6-duration  0 days 00:00:18.359000000
      simple_french_qcm-7-duration  0 days 00:00:23.953000000
      simple_french_qcm-8-duration  0 days 00:00:07.641000000

      person_id          0cf4e678b9586d5b0e37
      \
      simple_french_qcm-0-duration  1 days 00:00:00.000000000,1 days 00:00:00.0000...
      simple_french_qcm-1-duration          0 days 00:00:09.323000000
      simple_french_qcm-2-duration          0 days 00:00:05.733000000
      simple_french_qcm-3-duration          0 days 00:00:05.321000000
      simple_french_qcm-4-duration          0 days 00:00:07.680000000
      simple_french_qcm-5-duration          0 days 00:00:08.045000000
      simple_french_qcm-6-duration          0 days 00:00:04.377000000
      simple_french_qcm-7-duration          0 days 00:00:11.580000000
      simple_french_qcm-8-duration          0 days 00:00:06.796000000

      person_id          0fa2c1f9be1abc2c1d0b \
      simple_french_qcm-0-duration  0 days 00:00:03.546000000
      simple_french_qcm-1-duration  0 days 00:00:04.949000000
      simple_french_qcm-2-duration  0 days 00:00:58.690000000
      simple_french_qcm-3-duration  0 days 00:00:24.203000000
      simple_french_qcm-4-duration  0 days 00:00:09.975000000
      simple_french_qcm-5-duration  0 days 00:00:16.402000000
      simple_french_qcm-6-duration  0 days 00:01:13.287000000
      simple_french_qcm-7-duration  0 days 00:00:10.718000000
      simple_french_qcm-8-duration  0 days 00:00:08.438000000

      person_id          10fb62c55bc7a6a5cda2
      \
      simple_french_qcm-0-duration  0 days 00:00:04.912000000,0 days 00:00:07.5230...
      simple_french_qcm-1-duration  1 days 00:00:00.000000000,1 days 00:00:00.0000...
      simple_french_qcm-2-duration  0 days 00:00:31.779000000,0 days 00:00:34.1990...
      simple_french_qcm-3-duration  0 days 00:00:26.622000000,0 days 00:00:28.4660...
      simple_french_qcm-4-duration  0 days 00:00:13.814000000,0 days 00:00:15.0330...
      simple_french_qcm-5-duration  0 days 00:00:14.796000000,0 days 00:00:15.8870...
      simple_french_qcm-6-duration  0 days 00:00:26.275000000,0 days 00:00:28.6000...
      simple_french_qcm-7-duration          0 days 00:00:01.156000000
```

```

simple_french_qcm-8-duration          0 days 00:00:01.813000000

person_id                             14db30224001f714a266
simple_french_qcm-0-duration          0 days 00:00:08.688000000
simple_french_qcm-1-duration          1 days 00:00:00.000000000
simple_french_qcm-2-duration          0 days 00:00:15.349000000
simple_french_qcm-3-duration          1 days 00:00:00.000000000
simple_french_qcm-4-duration          0 days 00:01:08.922000000
simple_french_qcm-5-duration          1 days 00:00:00.000000000
simple_french_qcm-6-duration          0 days 00:00:34.538000000
simple_french_qcm-7-duration          0 days 00:00:22.348000000
simple_french_qcm-8-duration          0 days 00:00:47.971000000

```

```
[11]: gr_question = df_question.groupby("person_id").agg(aggnotnan)
      gr_question.head().T
```

```

[11]: person_id          0c3a2f07b272478e4cd8  0cf4e678b9586d5b0e37  \
simple_french_qcm-0-a0          1.0          NaN
simple_french_qcm-0-a1          NaN          NaN
simple_french_qcm-0-a2          NaN          2.0
simple_french_qcm-1-a1          NaN          NaN
simple_french_qcm-1-a2          1.0          1.0
simple_french_qcm-2-a0          NaN          NaN
simple_french_qcm-2-a1          NaN          NaN
simple_french_qcm-2-a2          NaN          NaN
simple_french_qcm-2-a3          1.0          1.0
simple_french_qcm-3-a0          NaN          1.0
simple_french_qcm-3-a2          NaN          NaN
simple_french_qcm-3-a3          1.0          NaN
simple_french_qcm-3-a4          NaN          NaN
simple_french_qcm-4-a0          NaN          NaN
simple_french_qcm-4-a1          NaN          NaN
simple_french_qcm-4-a2          1.0          1.0
simple_french_qcm-4-a3          NaN          NaN
simple_french_qcm-5-a0          1.0          NaN
simple_french_qcm-5-a1          1.0          1.0
simple_french_qcm-5-a2          1.0          1.0
simple_french_qcm-5-a4          1.0          NaN
simple_french_qcm-6-a0          1.0          NaN
simple_french_qcm-6-a1          NaN          1.0
simple_french_qcm-6-a2          1.0          NaN
simple_french_qcm-6-a3          NaN          NaN
simple_french_qcm-6-a4          NaN          NaN
simple_french_qcm-6-a5          NaN          NaN
simple_french_qcm-7-a0          NaN          NaN
simple_french_qcm-7-a1          NaN          NaN
simple_french_qcm-7-a2          1.0          1.0

person_id          0fa2c1f9be1abc2c1d0b  10fb62c55bc7a6a5cda2  \
simple_french_qcm-0-a0          NaN          1.0
simple_french_qcm-0-a1          1.0          NaN
simple_french_qcm-0-a2          NaN          NaN
simple_french_qcm-1-a1          NaN          NaN

```

simple_french_qcm-1-a2	1.0	NaN
simple_french_qcm-2-a0	NaN	NaN
simple_french_qcm-2-a1	NaN	NaN
simple_french_qcm-2-a2	1.0	1.0
simple_french_qcm-2-a3	NaN	NaN
simple_french_qcm-3-a0	NaN	NaN
simple_french_qcm-3-a2	1.0	NaN
simple_french_qcm-3-a3	NaN	NaN
simple_french_qcm-3-a4	NaN	NaN
simple_french_qcm-4-a0	NaN	1.0
simple_french_qcm-4-a1	1.0	NaN
simple_french_qcm-4-a2	NaN	NaN
simple_french_qcm-4-a3	NaN	NaN
simple_french_qcm-5-a0	1.0	NaN
simple_french_qcm-5-a1	1.0	1.0
simple_french_qcm-5-a2	1.0	NaN
simple_french_qcm-5-a4	NaN	NaN
simple_french_qcm-6-a0	NaN	NaN
simple_french_qcm-6-a1	NaN	NaN
simple_french_qcm-6-a2	NaN	1.0
simple_french_qcm-6-a3	1.0	NaN
simple_french_qcm-6-a4	NaN	NaN
simple_french_qcm-6-a5	NaN	NaN
simple_french_qcm-7-a0	NaN	NaN
simple_french_qcm-7-a1	NaN	NaN
simple_french_qcm-7-a2	1.0	NaN

person_id	14db30224001f714a266
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-a1	1.0
simple_french_qcm-0-a2	NaN
simple_french_qcm-1-a1	1.0
simple_french_qcm-1-a2	NaN
simple_french_qcm-2-a0	NaN
simple_french_qcm-2-a1	NaN
simple_french_qcm-2-a2	1.0
simple_french_qcm-2-a3	NaN
simple_french_qcm-3-a0	1.0
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-a4	NaN
simple_french_qcm-4-a0	1.0
simple_french_qcm-4-a1	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-a3	NaN
simple_french_qcm-5-a0	1.0
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-a4	NaN
simple_french_qcm-6-a0	NaN
simple_french_qcm-6-a1	NaN
simple_french_qcm-6-a2	1.0
simple_french_qcm-6-a3	NaN


```

simple_french_qcm-6-a4      NaN
simple_french_qcm-6-a5      NaN
simple_french_qcm-7-a0      NaN
simple_french_qcm-7-a1      NaN
simple_french_qcm-7-a2      1.0

```

```
[12]: gr_question.shape
```

```
[12]: (41, 30)
```

```
[13]: gr_bouton = df_bouton.groupby("person_id").agg(aggnotnan)
gr_bouton.head().T
```

```
[13]: person_id      0c3a2f07b272478e4cd8  0cf4e678b9586d5b0e37  \
simple_french_qcm-0-b      1.0      1001.0
simple_french_qcm-1-b      1.0      1.0
simple_french_qcm-2-b      1.0      1.0
simple_french_qcm-3-b      1.0      1.0
simple_french_qcm-4-b      1.0      1.0
simple_french_qcm-5-b      1.0      1.0
simple_french_qcm-6-b      1.0      1.0
simple_french_qcm-7-b      1.0      1.0
simple_french_qcm-8-b      1.0      1.0
```

```

person_id      0fa2c1f9be1abc2c1d0b  10fb62c55bc7a6a5cda2  \
simple_french_qcm-0-b      1.0      1001.0
simple_french_qcm-1-b      1.0      2000.0
simple_french_qcm-2-b      1.0      2000.0
simple_french_qcm-3-b      1.0      2000.0
simple_french_qcm-4-b      1.0      2000.0
simple_french_qcm-5-b      1.0      1001.0
simple_french_qcm-6-b      1.0      2000.0
simple_french_qcm-7-b      1.0      1000.0
simple_french_qcm-8-b      1.0      1000.0

```

```

person_id      14db30224001f714a266
simple_french_qcm-0-b      1.0
simple_french_qcm-1-b      1.0
simple_french_qcm-2-b      1.0
simple_french_qcm-3-b      1.0
simple_french_qcm-4-b      1.0
simple_french_qcm-5-b      1.0
simple_french_qcm-6-b      1.0
simple_french_qcm-7-b      1.0
simple_french_qcm-8-b      1.0

```

```
[14]: gr_visit = df_visit.groupby("person_id").agg(aggnotnan)
gr_visit.head().T
```

```
[14]: person_id      0c3a2f07b272478e4cd8  0cf4e678b9586d5b0e37  \
simple_french_qcm-0-nbvisit      1.0      2.0
simple_french_qcm-1-nbvisit      1.0      2.0
simple_french_qcm-2-nbvisit      1.0      1.0
simple_french_qcm-3-nbvisit      1.0      1.0

```

simple_french_qcm-4-nbvisit	1.0	1.0
simple_french_qcm-5-nbvisit	1.0	1.0
simple_french_qcm-6-nbvisit	1.0	1.0
simple_french_qcm-7-nbvisit	1.0	1.0
simple_french_qcm-8-nbvisit	1.0	1.0
person_id	0fa2c1f9be1abc2c1d0b	10fb62c55bc7a6a5cda2 \
simple_french_qcm-0-nbvisit	1.0	3.0
simple_french_qcm-1-nbvisit	1.0	4.0
simple_french_qcm-2-nbvisit	1.0	3.0
simple_french_qcm-3-nbvisit	1.0	3.0
simple_french_qcm-4-nbvisit	1.0	3.0
simple_french_qcm-5-nbvisit	1.0	3.0
simple_french_qcm-6-nbvisit	1.0	3.5
simple_french_qcm-7-nbvisit	1.0	1.5
simple_french_qcm-8-nbvisit	1.0	1.0
person_id	14db30224001f714a266	
simple_french_qcm-0-nbvisit	1.0	
simple_french_qcm-1-nbvisit	1.5	
simple_french_qcm-2-nbvisit	1.0	
simple_french_qcm-3-nbvisit	1.5	
simple_french_qcm-4-nbvisit	1.0	
simple_french_qcm-5-nbvisit	1.5	
simple_french_qcm-6-nbvisit	1.0	
simple_french_qcm-7-nbvisit	1.0	
simple_french_qcm-8-nbvisit	1.0	

1.3.1 Histogrammes

```
[15]: nonan_question = gr_question.fillna(0)
```

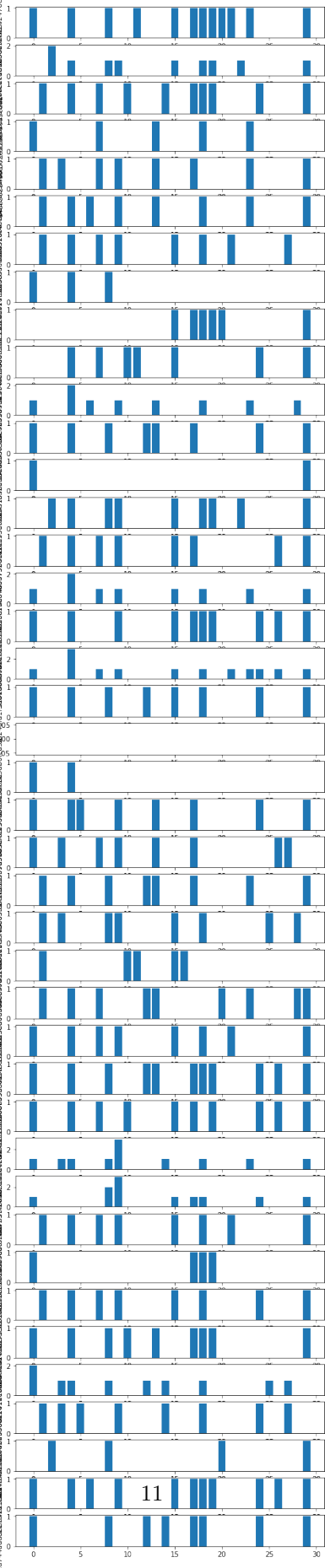
```
[16]: nonan_question.shape
```

```
[16]: (41, 30)
```

```
[17]: import matplotlib.pyplot as plt

fig, ax = plt.subplots(nonan_question.shape[0], 1,
                        figsize=(8, nonan_question.shape[0]))
for i in range(0, nonan_question.shape[0]):
    ax[i].set_ylabel(nonan_question.index[i])
    ax[i].bar(list(range(nonan_question.shape[1])),
               nonan_question.iloc[i,:])
```

Set1: 2783aa77957f1
 ec74a2c9c2e2f2f1a47e1a7482866a18281c6b5c5c3897290715a9786d09766c03969758644815998857718152803968605521322721288c9f9a8d33886656525070072660416598a4c2818863a1b2c2747864c8



1.3.2 ACP

```
[18]: nonan_question.shape
```

```
[18]: (41, 30)
```

```
[19]: from sklearn.decomposition import PCA
      acp = PCA(n_components=2, svd_solver="arpack")
      acp.fit(nonan_question)
```

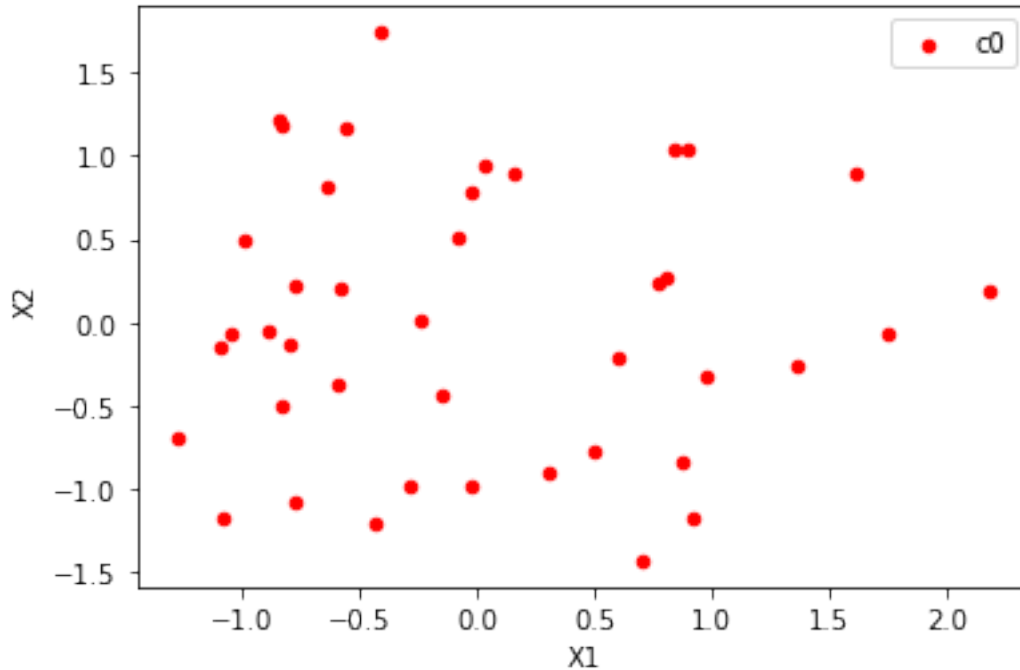
```
[19]: PCA(n_components=2, svd_solver='arpack')
```

```
[20]: coord = acp.transform(nonan_question)
      data = pandas.DataFrame(data=coord, columns=['X1', 'X2'], index=nonan_question.index)
      data["cluster"] = 0
      data.head()
```

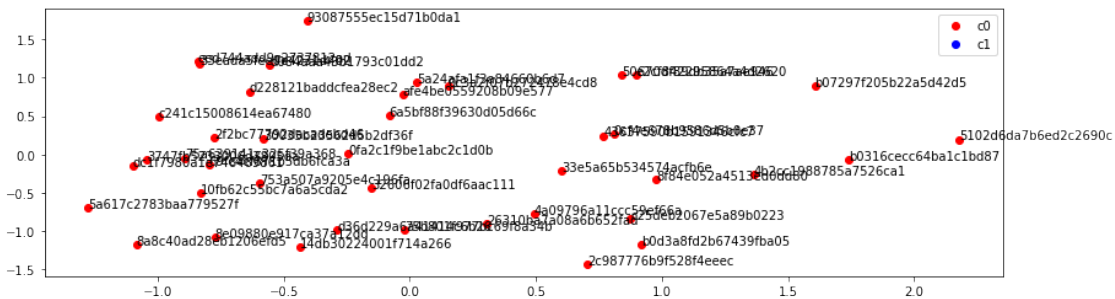
```
[20]:
```

	X1	X2	cluster
person_id			
0c3a2f07b272478e4cd8	0.151681	0.892031	0
0cf4e678b9586d5b0e37	0.809105	0.267815	0
0fa2c1f9be1abc2c1d0b	-0.243414	0.005855	0
10fb62c55bc7a6a5cda2	-0.830133	-0.506903	0
14db30224001f714a266	-0.433964	-1.207925	0

```
[21]: import matplotlib.pyplot as plt
      fig, ax = plt.subplots(1, 1)
      colors = ['red', 'blue', 'orange', 'green']
      for i in range(0, 1):
          data[data.cluster==i].plot(x="X1", y="X2",
                                     kind="scatter",
                                     ax=ax, label="c%d" % i,
                                     color=colors[i])
```



```
[22]: data["weight"] = 10
# Plot miles per gallon against horsepower with other semantics
fig, ax = plt.subplots(1, 1, figsize=(14, 4))
ax.plot(data.loc[data.cluster == 0, 'X1'].values,
        data.loc[data.cluster == 0, 'X2'].values, "ro", label="c0")
ax.plot(data.loc[data.cluster == 1, 'X1'].values,
        data.loc[data.cluster == 1, 'X2'].values, "bo", label="c1")
ind = list(data.index)
for i in range(0, data.shape[0]):
    ax.text(data.iloc[i, 0], data.iloc[i, 1], ind[i])
ax.legend();
```



1.4 Second jeu de données

```
[23]: df2 = df[df.qtime == 'end'].copy()
      person_id = ['person_id']
```

```
[24]: cols = person_id + [c for c in df2.columns if "ml_french" in c and '-a' in c]
      df2_question = df2[cols]
```

```
[25]: import missingno as msno
      msno.matrix(df2_question, figsize=(15,3));
```



```
[26]: cols = person_id + [c for c in df2.columns if "ml_french" in c and '-b' in c]
      df2_bouton = df2[cols]
      cols = person_id + [c for c in df2.columns if "ml_french" in c and '-nb' in c]
      df2_visit = df2[cols]
      cols = person_id + [c for c in df2.columns if "ml_french" in c and '-dur' in c]
      df2_dur = df2[cols]
```

```
[27]: gr2_question = df2_question.groupby("person_id").agg(aggotnan)
      gr2_question.T
```

```
[27]: person_id          0c3a2f07b272478e4cd8  0cf4e678b9586d5b0e37  \
ml_french_qcm-0-a0          1.0          NaN
ml_french_qcm-0-a2          NaN          NaN
ml_french_qcm-0-a3          NaN          NaN
ml_french_qcm-1-a1          NaN          NaN
ml_french_qcm-1-a2          NaN          NaN
ml_french_qcm-1-a3          1.0          NaN
ml_french_qcm-10-a0         NaN          NaN
ml_french_qcm-10-a1         1.0          NaN
ml_french_qcm-10-a3         1.0          NaN
ml_french_qcm-11-a0         NaN          NaN
ml_french_qcm-11-a1         1.0          NaN
ml_french_qcm-11-a2         1.0          NaN
ml_french_qcm-11-a3         1.0          NaN
ml_french_qcm-12-a0         NaN          NaN
ml_french_qcm-12-a1         NaN          NaN
ml_french_qcm-12-a2         NaN          NaN
ml_french_qcm-12-a3         1.0          NaN
ml_french_qcm-13-a0         NaN          NaN
ml_french_qcm-13-a2         NaN          NaN
ml_french_qcm-13-a3         1.0          NaN
ml_french_qcm-14-a1         1.0          NaN
```

ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	1.0	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	1.0	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	1.0	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	1.0	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	1.0	NaN
ml_french_qcm-7-a6	1.0	NaN
ml_french_qcm-8-a1	1.0	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	1.0	NaN
ml_french_qcm-9-a1	1.0	NaN
ml_french_qcm-9-a2	1.0	NaN
ml_french_qcm-9-a3	NaN	NaN

person_id	0fa2c1f9be1abc2c1d0b	10fb62c55bc7a6a5cda2	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	NaN	
ml_french_qcm-0-a3	1.0	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	1.0	NaN	
ml_french_qcm-10-a0	NaN	NaN	
ml_french_qcm-10-a1	1.0	NaN	
ml_french_qcm-10-a3	NaN	NaN	
ml_french_qcm-11-a0	NaN	NaN	
ml_french_qcm-11-a1	NaN	NaN	
ml_french_qcm-11-a2	1.0	NaN	
ml_french_qcm-11-a3	1.0	NaN	
ml_french_qcm-12-a0	NaN	NaN	
ml_french_qcm-12-a1	1.0	NaN	
ml_french_qcm-12-a2	NaN	NaN	
ml_french_qcm-12-a3	NaN	NaN	
ml_french_qcm-13-a0	NaN	NaN	
ml_french_qcm-13-a2	1.0	NaN	

ml_french_qcm-13-a3	1.0	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	1.0	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	1.0	NaN
ml_french_qcm-3-a4	1.0	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	1.0	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	1.0	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	1.0	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	1.0	NaN
ml_french_qcm-7-a4	1.0	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	1.0	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	1.0	NaN
ml_french_qcm-9-a3	NaN	NaN

person_id	14db30224001f714a266	26310ba7a08a6b652fad	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	NaN	
ml_french_qcm-0-a3	NaN	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	NaN	NaN	
ml_french_qcm-10-a0	NaN	NaN	
ml_french_qcm-10-a1	NaN	NaN	
ml_french_qcm-10-a3	NaN	NaN	
ml_french_qcm-11-a0	NaN	NaN	
ml_french_qcm-11-a1	NaN	NaN	
ml_french_qcm-11-a2	NaN	NaN	
ml_french_qcm-11-a3	NaN	NaN	
ml_french_qcm-12-a0	NaN	NaN	
ml_french_qcm-12-a1	NaN	NaN	
ml_french_qcm-12-a2	NaN	NaN	
ml_french_qcm-12-a3	NaN	NaN	

ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN

person_id	2c987776b9f528f4eeec	2f2bc77792dacadebd46	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	1.0	
ml_french_qcm-0-a3	NaN	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	NaN	1.0	
ml_french_qcm-10-a0	NaN	NaN	
ml_french_qcm-10-a1	NaN	NaN	
ml_french_qcm-10-a3	NaN	1.0	
ml_french_qcm-11-a0	NaN	NaN	
ml_french_qcm-11-a1	NaN	NaN	
ml_french_qcm-11-a2	NaN	2.0	
ml_french_qcm-11-a3	NaN	NaN	
ml_french_qcm-12-a0	NaN	NaN	
ml_french_qcm-12-a1	NaN	NaN	

ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	1.0
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	1.0
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	1.0
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	1.0
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	1.0
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	1.0
ml_french_qcm-5-a0	NaN	1.0
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	1.0
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	1.0
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	1.0
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	1.0

person_id	30035b23562d5b2df36f	32606f02fa0df6aac111	...	\
ml_french_qcm-0-a0	NaN	NaN	...	
ml_french_qcm-0-a2	NaN	NaN	...	
ml_french_qcm-0-a3	NaN	NaN	...	
ml_french_qcm-1-a1	NaN	NaN	...	
ml_french_qcm-1-a2	NaN	NaN	...	
ml_french_qcm-1-a3	NaN	NaN	...	
ml_french_qcm-10-a0	NaN	NaN	...	
ml_french_qcm-10-a1	NaN	NaN	...	
ml_french_qcm-10-a3	NaN	NaN	...	
ml_french_qcm-11-a0	NaN	NaN	...	
ml_french_qcm-11-a1	NaN	NaN	...	
ml_french_qcm-11-a2	NaN	NaN	...	
ml_french_qcm-11-a3	NaN	NaN	...	

ml_french_qcm-12-a0	NaN	NaN	...
ml_french_qcm-12-a1	NaN	NaN	...
ml_french_qcm-12-a2	NaN	NaN	...
ml_french_qcm-12-a3	NaN	NaN	...
ml_french_qcm-13-a0	NaN	NaN	...
ml_french_qcm-13-a2	NaN	NaN	...
ml_french_qcm-13-a3	NaN	NaN	...
ml_french_qcm-14-a1	NaN	NaN	...
ml_french_qcm-14-a3	NaN	NaN	...
ml_french_qcm-2-a0	NaN	NaN	...
ml_french_qcm-2-a2	NaN	NaN	...
ml_french_qcm-2-a3	NaN	NaN	...
ml_french_qcm-3-a1	NaN	NaN	...
ml_french_qcm-3-a2	NaN	NaN	...
ml_french_qcm-3-a3	NaN	NaN	...
ml_french_qcm-3-a4	NaN	NaN	...
ml_french_qcm-3-a5	NaN	NaN	...
ml_french_qcm-4-a0	NaN	NaN	...
ml_french_qcm-4-a1	NaN	NaN	...
ml_french_qcm-4-a2	NaN	NaN	...
ml_french_qcm-4-a4	NaN	NaN	...
ml_french_qcm-5-a0	NaN	NaN	...
ml_french_qcm-5-a1	NaN	NaN	...
ml_french_qcm-5-a2	NaN	NaN	...
ml_french_qcm-5-a4	NaN	NaN	...
ml_french_qcm-6-a0	NaN	NaN	...
ml_french_qcm-6-a1	NaN	NaN	...
ml_french_qcm-6-a2	NaN	NaN	...
ml_french_qcm-6-a3	NaN	NaN	...
ml_french_qcm-7-a3	NaN	NaN	...
ml_french_qcm-7-a4	NaN	NaN	...
ml_french_qcm-7-a5	NaN	NaN	...
ml_french_qcm-7-a6	NaN	NaN	...
ml_french_qcm-8-a1	NaN	NaN	...
ml_french_qcm-8-a2	NaN	NaN	...
ml_french_qcm-8-a3	NaN	NaN	...
ml_french_qcm-9-a0	NaN	NaN	...
ml_french_qcm-9-a1	NaN	NaN	...
ml_french_qcm-9-a2	1.0	NaN	...
ml_french_qcm-9-a3	NaN	NaN	...
person_id	b07297f205b22a5d42d5	b0d3a8fd2b67439fba05	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	NaN	
ml_french_qcm-0-a3	NaN	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	NaN	NaN	
ml_french_qcm-10-a0	NaN	NaN	
ml_french_qcm-10-a1	NaN	NaN	
ml_french_qcm-10-a3	NaN	NaN	
ml_french_qcm-11-a0	NaN	NaN	
ml_french_qcm-11-a1	NaN	NaN	

ml_french_qcm-11-a2	NaN	NaN
ml_french_qcm-11-a3	NaN	NaN
ml_french_qcm-12-a0	NaN	NaN
ml_french_qcm-12-a1	NaN	NaN
ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	NaN
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN

person_id	c241c15008614ea67480	c25deb2067e5a89b0223	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	NaN	
ml_french_qcm-0-a3	NaN	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	NaN	NaN	
ml_french_qcm-10-a0	NaN	NaN	
ml_french_qcm-10-a1	NaN	NaN	
ml_french_qcm-10-a3	NaN	NaN	

ml_french_qcm-11-a0	NaN	NaN
ml_french_qcm-11-a1	NaN	NaN
ml_french_qcm-11-a2	NaN	NaN
ml_french_qcm-11-a3	NaN	NaN
ml_french_qcm-12-a0	NaN	NaN
ml_french_qcm-12-a1	NaN	NaN
ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	NaN
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN

person_id	d084aaa43b1793c01dd2	d228121baddcfea28ec2	\
ml_french_qcm-0-a0	NaN	NaN	
ml_french_qcm-0-a2	NaN	NaN	
ml_french_qcm-0-a3	NaN	NaN	
ml_french_qcm-1-a1	NaN	NaN	
ml_french_qcm-1-a2	NaN	NaN	
ml_french_qcm-1-a3	NaN	NaN	
ml_french_qcm-10-a0	NaN	NaN	

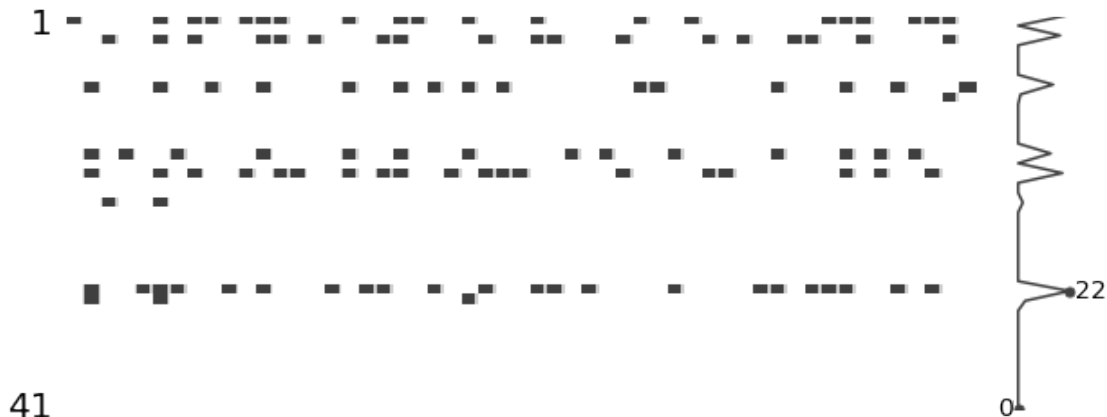
ml_french_qcm-10-a1	NaN	NaN
ml_french_qcm-10-a3	NaN	NaN
ml_french_qcm-11-a0	NaN	NaN
ml_french_qcm-11-a1	NaN	NaN
ml_french_qcm-11-a2	NaN	NaN
ml_french_qcm-11-a3	NaN	NaN
ml_french_qcm-12-a0	NaN	NaN
ml_french_qcm-12-a1	NaN	NaN
ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	NaN
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN
person_id	d36d229a6a41414c6b26	dc1f7980a1e746469861 \
ml_french_qcm-0-a0	NaN	NaN
ml_french_qcm-0-a2	NaN	NaN
ml_french_qcm-0-a3	NaN	NaN
ml_french_qcm-1-a1	NaN	NaN
ml_french_qcm-1-a2	NaN	NaN

ml_french_qcm-1-a3	NaN	NaN
ml_french_qcm-10-a0	NaN	NaN
ml_french_qcm-10-a1	NaN	NaN
ml_french_qcm-10-a3	NaN	NaN
ml_french_qcm-11-a0	NaN	NaN
ml_french_qcm-11-a1	NaN	NaN
ml_french_qcm-11-a2	NaN	NaN
ml_french_qcm-11-a3	NaN	NaN
ml_french_qcm-12-a0	NaN	NaN
ml_french_qcm-12-a1	NaN	NaN
ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	NaN
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN
person_id	e2cf842295864a4d2620	eed744add9c2737812ad
ml_french_qcm-0-a0	NaN	NaN
ml_french_qcm-0-a2	NaN	NaN
ml_french_qcm-0-a3	NaN	NaN

ml_french_qcm-1-a1	NaN	NaN
ml_french_qcm-1-a2	NaN	NaN
ml_french_qcm-1-a3	NaN	NaN
ml_french_qcm-10-a0	NaN	NaN
ml_french_qcm-10-a1	NaN	NaN
ml_french_qcm-10-a3	NaN	NaN
ml_french_qcm-11-a0	NaN	NaN
ml_french_qcm-11-a1	NaN	NaN
ml_french_qcm-11-a2	NaN	NaN
ml_french_qcm-11-a3	NaN	NaN
ml_french_qcm-12-a0	NaN	NaN
ml_french_qcm-12-a1	NaN	NaN
ml_french_qcm-12-a2	NaN	NaN
ml_french_qcm-12-a3	NaN	NaN
ml_french_qcm-13-a0	NaN	NaN
ml_french_qcm-13-a2	NaN	NaN
ml_french_qcm-13-a3	NaN	NaN
ml_french_qcm-14-a1	NaN	NaN
ml_french_qcm-14-a3	NaN	NaN
ml_french_qcm-2-a0	NaN	NaN
ml_french_qcm-2-a2	NaN	NaN
ml_french_qcm-2-a3	NaN	NaN
ml_french_qcm-3-a1	NaN	NaN
ml_french_qcm-3-a2	NaN	NaN
ml_french_qcm-3-a3	NaN	NaN
ml_french_qcm-3-a4	NaN	NaN
ml_french_qcm-3-a5	NaN	NaN
ml_french_qcm-4-a0	NaN	NaN
ml_french_qcm-4-a1	NaN	NaN
ml_french_qcm-4-a2	NaN	NaN
ml_french_qcm-4-a4	NaN	NaN
ml_french_qcm-5-a0	NaN	NaN
ml_french_qcm-5-a1	NaN	NaN
ml_french_qcm-5-a2	NaN	NaN
ml_french_qcm-5-a4	NaN	NaN
ml_french_qcm-6-a0	NaN	NaN
ml_french_qcm-6-a1	NaN	NaN
ml_french_qcm-6-a2	NaN	NaN
ml_french_qcm-6-a3	NaN	NaN
ml_french_qcm-7-a3	NaN	NaN
ml_french_qcm-7-a4	NaN	NaN
ml_french_qcm-7-a5	NaN	NaN
ml_french_qcm-7-a6	NaN	NaN
ml_french_qcm-8-a1	NaN	NaN
ml_french_qcm-8-a2	NaN	NaN
ml_french_qcm-8-a3	NaN	NaN
ml_french_qcm-9-a0	NaN	NaN
ml_french_qcm-9-a1	NaN	NaN
ml_french_qcm-9-a2	NaN	NaN
ml_french_qcm-9-a3	NaN	NaN

[53 rows x 41 columns]


```
[28]: msno.matrix(gr2_question, figsize=(10,4));
```



1.4.1 ACP

```
[29]: gr2_question.shape
```

```
[29]: (41, 53)
```

```
[30]: nonan2_question = gr2_question.fillna(0)
```

```
[31]: from sklearn.decomposition import PCA
acp = PCA(n_components=2, svd_solver="arpark")
acp.fit(nonan2_question)
```

```
[31]: PCA(n_components=2, svd_solver='arpark')
```

```
[32]: coord = acp.transform(nonan2_question)
data = pandas.DataFrame(data=coord, columns=['X1', 'X2'], index=nonan_question.index)
data["cluster"] = 0
data
```

```
[32]:
```

	X1	X2	cluster
person_id			
0c3a2f07b272478e4cd8	2.887376	0.633808	0
0cf4e678b9586d5b0e37	-0.453284	-0.002668	0
0fa2c1f9be1abc2c1d0b	2.192238	2.514812	0
10fb62c55bc7a6a5cda2	-0.453284	-0.002668	0
14db30224001f714a266	-0.453284	-0.002668	0
26310ba7a08a6b652fad	-0.453284	-0.002668	0
2c987776b9f528f4eeec	-0.453284	-0.002668	0
2f2bc77792dacadebd46	2.600291	-1.841381	0
30035b23562d5b2df36f	-0.344912	0.188055	0
32606f02fa0df6aac111	-0.453284	-0.002668	0
33e5a65b534574acfb6e	-0.453284	-0.002668	0
33eada5fee0e4231d402	-0.453284	-0.002668	0

```

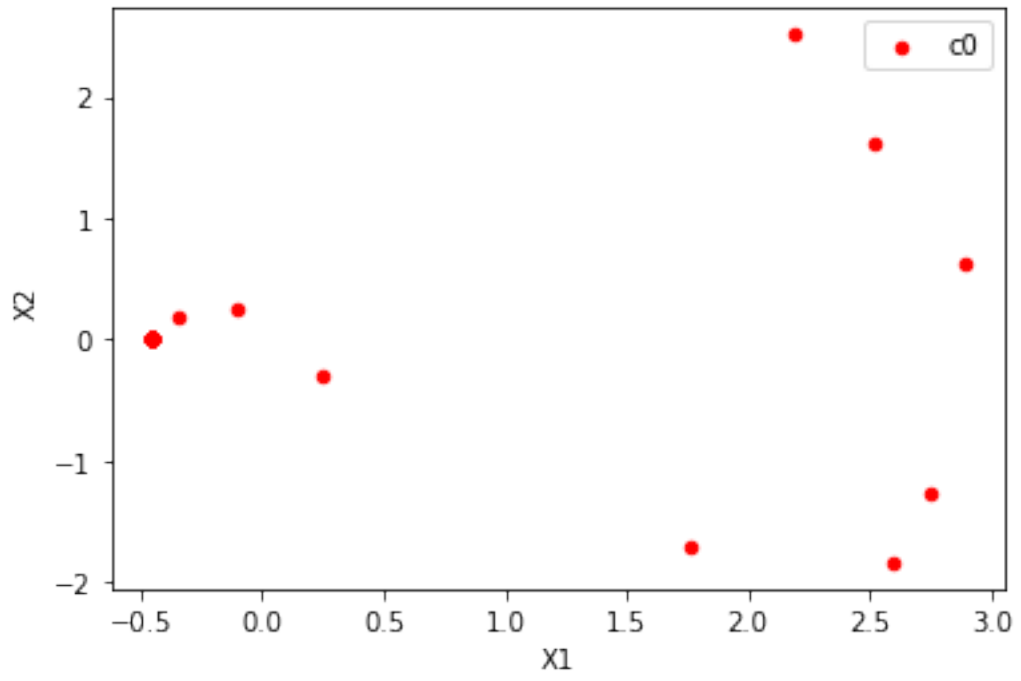
3747fb5233006d3805da -0.453284 -0.002668 0
43637590b1591346cfc7 -0.453284 -0.002668 0
4a09796a11ccc59ef66a 1.761605 -1.720518 0
4b2cc1988785a7526ca1 -0.453284 -0.002668 0
50670df89db35a7ae946 2.517077 1.622700 0
5102d6da7b6ed2c2690c -0.453284 -0.002668 0
5a24afa1f3a84660b6d7 -0.453284 -0.002668 0
5a617c2783baa779527f -0.105955 0.247710 0
69fc4cf871b5db6fca3a -0.453284 -0.002668 0
6a5bf88f39630d05d66c -0.453284 -0.002668 0
753a507a9205e4c196fa -0.453284 -0.002668 0
75a639141c325f39a368 -0.453284 -0.002668 0
75b801f977bc69f8a34b -0.453284 -0.002668 0
8a8c40ad28eb1206efd5 -0.453284 -0.002668 0
8e09880e917ca37a12dd -0.453284 -0.002668 0
8f84e052a4513ed0dd80 -0.453284 -0.002668 0
93087555ec15d71b0da1 2.753499 -1.265432 0
afe4be0559208b09e577 0.243858 -0.294390 0
b0316cecc64ba1c1bd87 -0.453284 -0.002668 0
b07297f205b22a5d42d5 -0.453284 -0.002668 0
b0d3a8fd2b67439fba05 -0.453284 -0.002668 0
c241c15008614ea67480 -0.453284 -0.002668 0
c25deb2067e5a89b0223 -0.453284 -0.002668 0
d084aaa43b1793c01dd2 -0.453284 -0.002668 0
d228121baddcfea28ec2 -0.453284 -0.002668 0
d36d229a6a41414c6b26 -0.453284 -0.002668 0
dc1f7980a1e746469861 -0.453284 -0.002668 0
e2cf842295864a4d2620 -0.453284 -0.002668 0
eed744add9c2737812ad -0.453284 -0.002668 0

```

```

[33]: import matplotlib.pyplot as plt
fig, ax = plt.subplots(1, 1)
colors = ['red', 'blue', 'orange', 'green']
for i in range(0, 1):
    data[data.cluster==i].plot(x="X1", y="X2",
                               kind="scatter",
                               ax=ax, label="c%d" % i,
                               color=colors[i])

```



[34] :