In [1]:

```
import pandas as pd
import csv
import numpy as np
```

In [2]:

```
cps = pd.read_csv('cps17.csv')
```

```
/Users/apple/anaconda3/lib/python3.6/site-packages/IPython/core/intera
ctiveshell.py:2698: DtypeWarning: Columns (5,14,33,34,37,70,72,87,111,
112,113,114,115,116,119,124,125,126,127,128,129,130,131,132,133,134,14
0,144,153,155,156,160,161,181,183,355,356,382,384,385,386) have mixed
types. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [3]:

```
cps.head(10)
```

Out[3]:

	hrhhid	hrmonth	hryear4	huresplі	hufinal	huspnish	hetenure	hehousut
0	4797110019	6	2017	-1	UNOCCUPIED TENT OR TRAILER SITE	0	-1	UNOCCUPIED TENT SITE OR TRLR SITE
1	110177987986	6	2017	-1	UNOCCUPIED TENT OR TRAILER SITE	0	-1	UNOCCUPIED TENT SITE OR TRLR SITE
2	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
3	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
4	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
5	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
6	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
7	110327856469	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT
8	110339935453	6	2017	-1	259	0	-1	HOUSE, APARTMENT, FLAT
9	110351278134	6	2017	1	CAPI COMPLETE	0	-1	HOUSE, APARTMENT, FLAT

10 rows × 387 columns

Part 1

Question 1

To access the data, we can either download it from NBER http://nber.org/cps-basic/jun17pub.zip (http://nber.org/cps-basic/jun17pub.zip), or from the original United States Census Bureau Website https://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpsbasic (https://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpsbasic) . The dataset is stored in these two website. The original source and curator of the dataset is United States Census Bureau, and it is also curated by NBER.

Question 2

Hirsch, Barry T., and David A. Macpherson. "Union membership and coverage database from the current population survey: Note." *ILR Review* 56, no. 2 (2003): 349-354.

Madrian, Brigitte C., and Lars John Lefgren. "An approach to longitudinally matching Current Population Survey (CPS) respondents." *Journal of Economic and Social Measurement*26, no. 1 (2000): 31-62.

Burkhauser, Richard V., Kenneth A. Couch, and David C. Wittenburg. "A reassessment of the new economics of the minimum wage literature with monthly data from the Current Population Survey." *Journal of Labor Economics* 18, no. 4 (2000): 653-680.

Mellor, Jennifer M., and Jeffrey Milyo. "Income inequality and health status in the United States: evidence from the current population survey." *Journal of Human Resources* (2002): 510-539.

Wewers, Mary Ellen, Frances A. Stillman, Anne M. Hartman, and Donald R. Shopland. "Distribution of daily smokers by stage of change: Current Population Survey results." *Preventive medicine* 36, no. 6 (2003): 710-720.

Question 3

Data was collected by conducting monthly surveys that cover various aspects of a household to a portion of random housing units that satisfies certain conditions in a rotation basis from 1976.

Question 4

In [109]:

```
a = cps['pemaritl'].value_counts()/len(cps)
```

In [110]:

```
b = cps['pespouse'].value_counts()/len(cps)
```

In [111]:

```
c = cps['pesex'].value_counts()/len(cps)
```

In [112]:

```
d = cps['peeduca'].value_counts()/len(cps)
```

In [138]:

```
e = cps['prfamnum'].value_counts()/len(cps)
```

Out[138]:

```
PRIMARY FAMILY MEMBER ONLY    0.657686
NOT A FAMILY MEMBER           0.163069
-1                            0.147575
SUBFAMILY NO. 2 MEMBER        0.030448
SUBFAMILY NO. 3 MEMBER        0.001168
SUBFAMILY NO. 4 MEMBER        0.000054
Name: prfamnum, dtype: float64
```

In [140]:

```
f = cps['penatvty'].value_counts()/len(cps)
f
```

Out[140]:

```
United States                  0.746937
-1                             0.147575
Mexico                         0.026085
India                          0.005964
Philippines                    0.004823
China                          0.004606
Puerto Rico                    0.003546
Vietnam                        0.003148
El Salvador                    0.002999
Germany                        0.002810
Canada                         0.002378
Cuba                           0.002155
Dominican Republic             0.002040
Guatemala                      0.001965
Korea                          0.001878
Columbia                       0.001391
Haiti                          0.001391
Japan                          0.001270
Brazil                         0.001229
Honduras                       0.001202
Jamaica                        0.001182
England                        0.001135
Italy                          0.001087
Pakistan                       0.000912
Russia                         0.000858
Africa, not specified          0.000851
Ukraine                        0.000844
Poland                         0.000838
Peru                           0.000804
Iran                           0.000777
                                  ...
Czechoslovakia                 0.000061
Zimbabwe                       0.000061
U. S. Virgin Islands           0.000061
Croatia                        0.000054
Uruguay                        0.000054
Paraguay                       0.000054
Singapore                      0.000054
Latvia                         0.000047
Fiji                           0.000047
Algeria                        0.000047
Kazakhstan                     0.000047
Zambia                         0.000047
St. Lucia                      0.000041
Czech Republic                 0.000034
St. Kitts--Nevis               0.000034
Azerbaijan                     0.000034
Zaire                          0.000034
Antigua and Barbuda            0.000027
Ivory Coast                    0.000027
Libya                          0.000027
```

```
Tanzania                            0.000027
Norway                              0.000027
Iceland                             0.000020
Serbia                              0.000020
St. Vincent and the Grenadines      0.000020
Bermuda                             0.000014
Northern Ireland                    0.000014
Northern Marianas                   0.000014
Slovakia                            0.000014
Estonia                             0.000007
Name: penatvty, Length: 162, dtype: float64
```

In [115]:

```
g = cps['prcitshp'].value_counts()/len(cps)
```

In [116]:

```
h = cps['pemlr'].value_counts()/len(cps)
```

In [146]:

```
series = [a, b, c, d, e, f, g, h]
df1 = None
df2 = None
df3 = None
df4 = None
df5 = None
df6 = None
df7 = None
df8 = None
dfs = [df1, df2, df3, df4, df5, df6, df7, df8]
vals = ['Marital Status', 'Num of spouse', 'Sex', 'Edu', 'Family size', 'Birth count
       'Citizenship', 'Labor force participation']
for i in range(0, 8):
    arrays = [vals[i] * len(series[i]),
              series[i].index]
    tuples = list(zip(*arrays))
    index = pd.MultiIndex.from_tuples(tuples, names=['Varible', 'Value'])
    dfs[i] = pd.Series(list(series[i].values), index=index)
result = pd.concat(dfs)
```

In [147]:

```
result
```

Out[147]:

Varible	Value	
M	MARRIED - SPOUSE PRESENT	0.353533
a	-1	0.305814
r	NEVER MARRIED	0.202791
i	DIVORCED	0.073453
t	WIDOWED	0.042430
a	SEPARATED	0.012036
l	MARRIED - SPOUSE ABSENT	0.009942
N	NO SPOUSE	0.646467
u	2	0.171147
m	1	0.169668
	3	0.007153
o	4	0.003026
f	5	0.001317
	6	0.000621
s	7	0.000297
p	8	0.000176
o	9	0.000068
u	10	0.000034
s	11	0.000014
e	14	0.000007
N	15	0.000007
S	FEMALE	0.438685
e	MALE	0.413741
x	-1	0.147575
E	-1	0.305814
d	HIGH SCHOOL GRAD-DIPLOMA OR EQUIV (GED)	0.1

```
96395
u       BACHELOR'S DEGREE (EX: BA, AB, BS)                          0.1
33046
E       SOME COLLEGE BUT NO DEGREE                                  0.1
21354
d       MASTER'S DEGREE (EX: MA, MS, MEng, MEd, MSW)                0.0
55919
u       ASSOCIATE DEGREE-ACADEMIC PROGRAM                           0.0
36656

...
t       St. Kitts--Nevis                                            0.0
00034
h       Azerbaijan                                                  0.0
00034
        Zaire                                                       0.0
00034
c       Antigua and Barbuda                                         0.0
00027
o       Ivory Coast                                                 0.0
00027
u       Libya                                                       0.0
00027
n       Tanzania                                                    0.0
00027
t       Norway                                                      0.0
00027
r       Iceland                                                     0.0
00020
y       Serbia                                                      0.0
00020
B       St. Vincent and the Grenadines                              0.0
00020
i       Bermuda                                                     0.0
00014
r       Northern Ireland                                            0.0
00014
t       Northern Marianas                                           0.0
00014
h       Slovakia                                                    0.0
00014
        Estonia                                                     0.0
00007
C       NATIVE, BORN IN THE UNITED STATES                           0.7
47011
i       -1                                                          0.1
47575
t       FOREIGN BORN, NOT A CITIZEN OF THE UNITED STATES            0.0
50428
i       FOREIGN BORN, U.S. CITIZEN BY NATURALIZATION                0.0
44551
z       NATIVE, BORN ABROAD OF AMERICAN PARENT OR PARENTS           0.0
06437
e       NATIVE, BORN IN PUERTO RICO OR OTHER U.S. ISLAND AREAS      0.0
03999
L       EMPLOYED-AT WORK                                            0.3
86298
a       -1                                                          0.3
```

```
08820
b        NOT IN LABOR FORCE-RETIRED                                0.1
32870
o        NOT IN LABOR FORCE-OTHER                                  0.0
93189
r        NOT IN LABOR FORCE-DISABLED                               0.0
39344
         EMPLOYED-ABSENT                                           0.0
21053
f        UNEMPLOYED-LOOKING                                        0.0
15893
o        UNEMPLOYED-ON LAYOFF                                      0.0
02533
Length: 223, dtype: float64
```

Question 5

In [24]:

```
import matplotlib.pyplot as plt
```

In [48]:

```
labor = cps['pemlr']
```

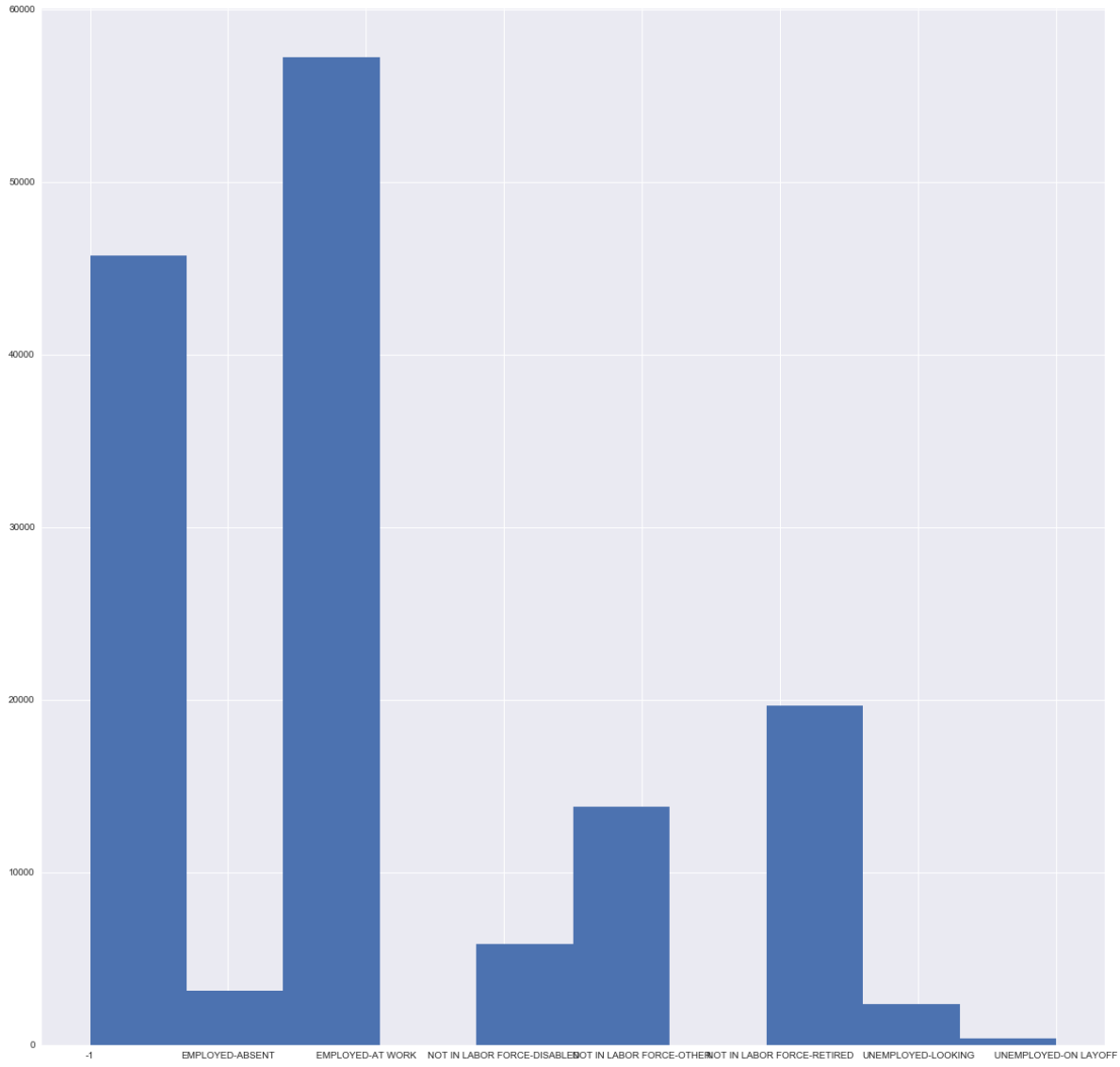
Visulization of labor force participation distribution

In [47]:

```
labor.hist(figsize=(20, 20))
```

Out[47]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x10d652be0>
```



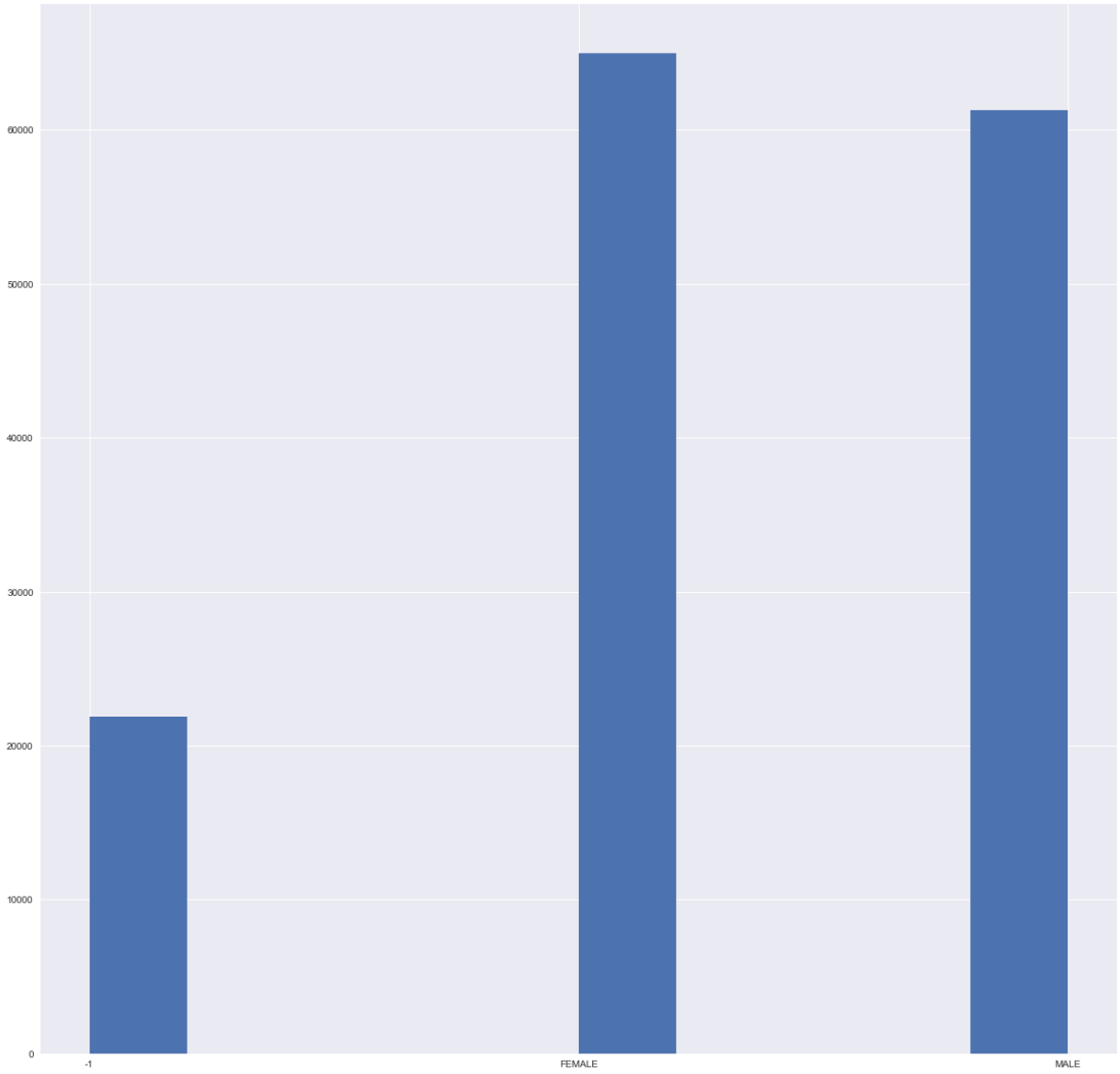
Visulization of gender distribution

In [149]:

```
sex = cps['pesex']
sex.hist(figsize=(20, 20))
```

Out[149]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x147b84630>
```



Question 6

Labor force participation distribution for male and female

In [51]:

```
female = cps[cps.pesex == "FEMALE"]
male = cps[cps.pesex == 'MALE']
```

In [54]:

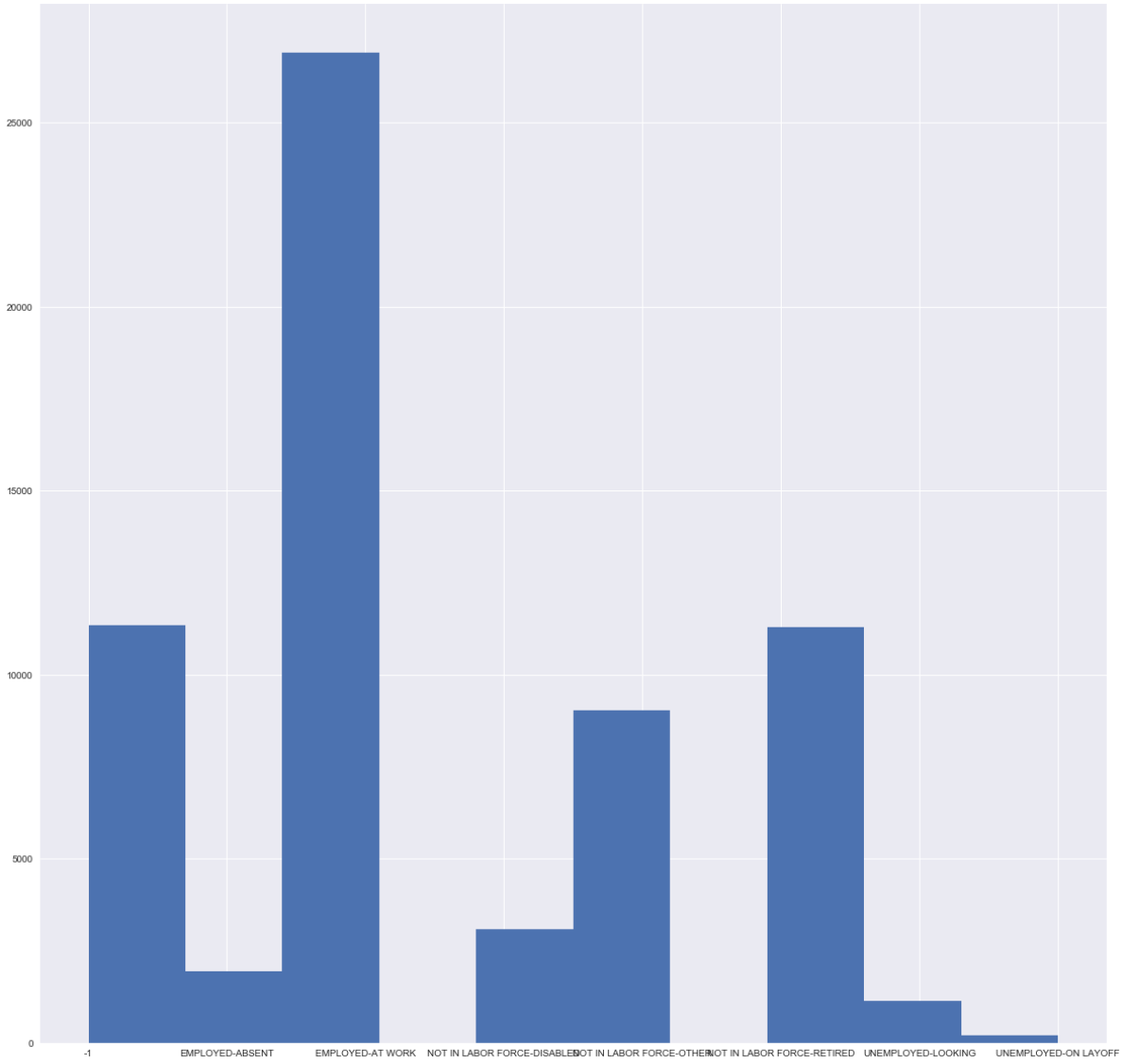
```
female_labor = female['pemlr']
male_labor = male['pemlr']
```

In [59]:

```
female_labor.hist(figsize=(20, 20))
```

Out[59]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x14a58def0>
```

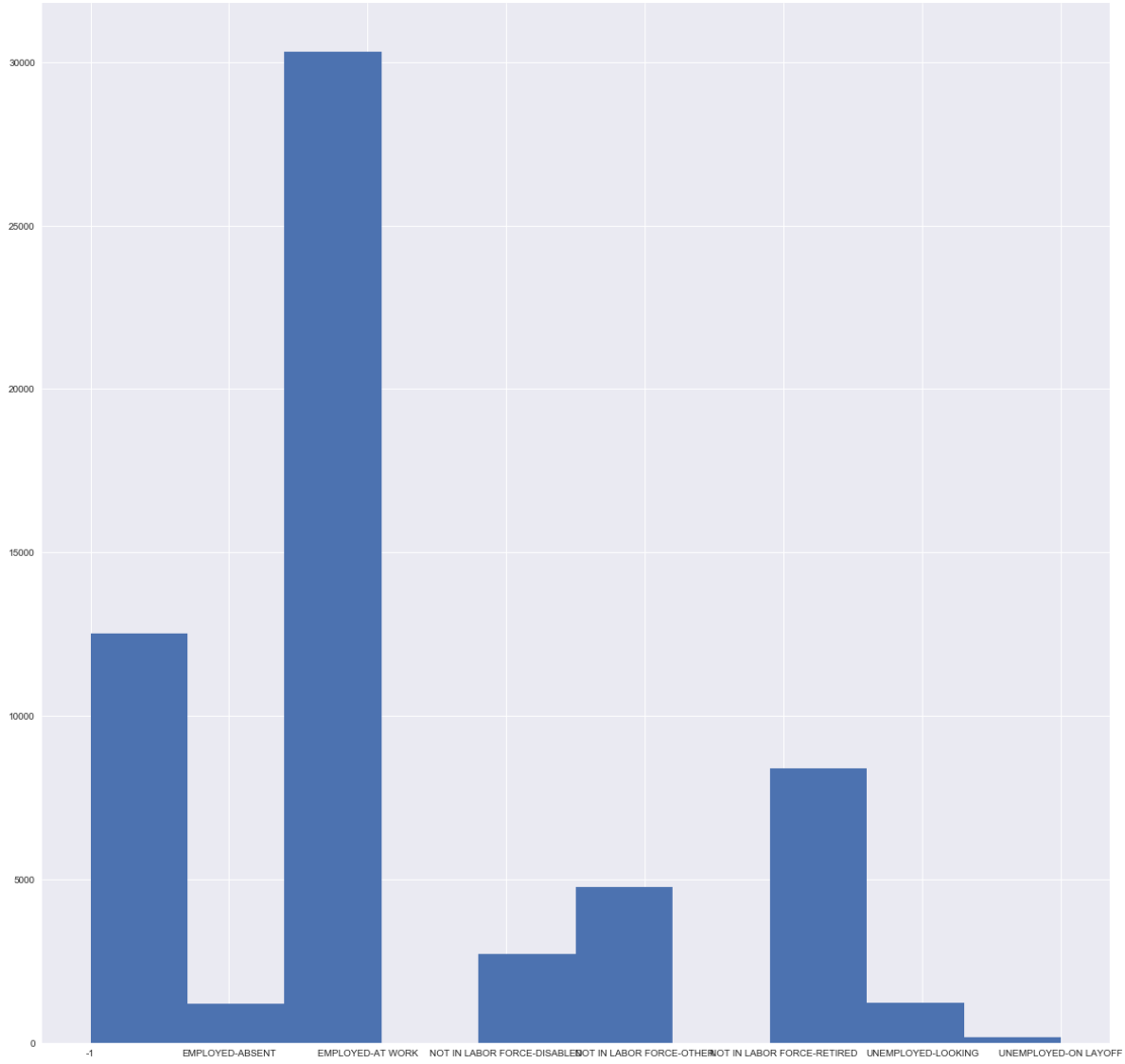


In [60]:

```
male_labor.hist(figsize=(20, 20))
```

Out[60]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x14ba2eb38>
```



Question 7

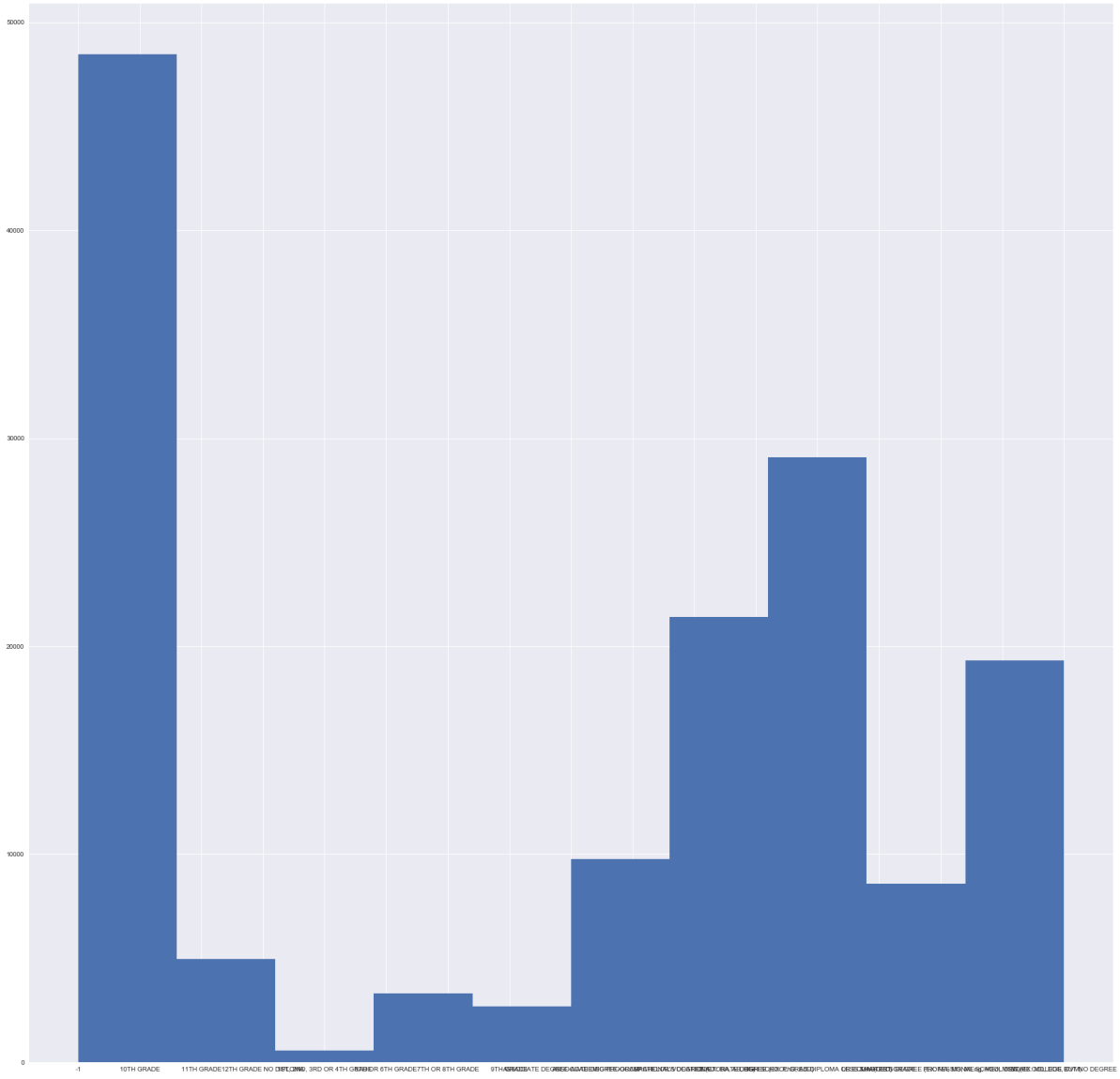
Distribition of education

In [62]:

```
cps['peeduca'].hist(figsize=(30, 30))
```

Out[62]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x14ce6b7b8>
```



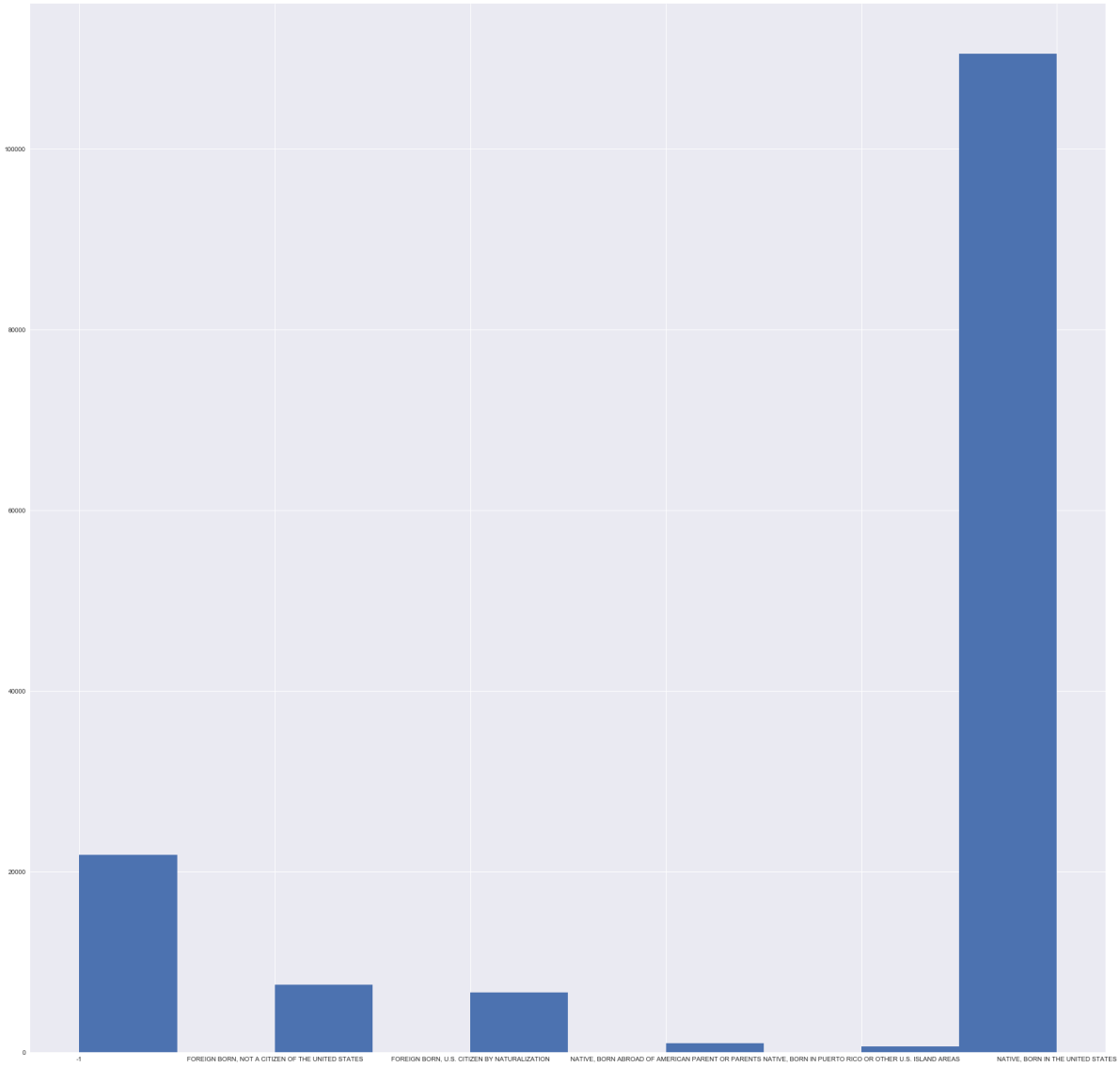
Distribition of citizenship

In [63]:

```
cps['prcitshp'].hist(figsize=(30, 30))
```

Out[63]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x10fb52748>
```



References:

National Bureau of Economic Research. http://nber.org/cps-basic/January_2017_Record_Layout.txt (http://nber.org/cps-basic/January_2017_Record_Layout.txt)
National Bureau of Economic Research. http://nber.org/data/cps_basic.html (http://nber.org/data/cps_basic.html)
United States Census Bureau. https://www.census.gov/programs-surveys/cps.html (https://www.census.gov/programs-surveys/cps.html)

In [ ]:

```

```