

Quiz 3

Brandan Owens and Loan Pham

Q.1 Read in the dataframe "gss_cat.csv".

```
In [46]: #import dataset and tools
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

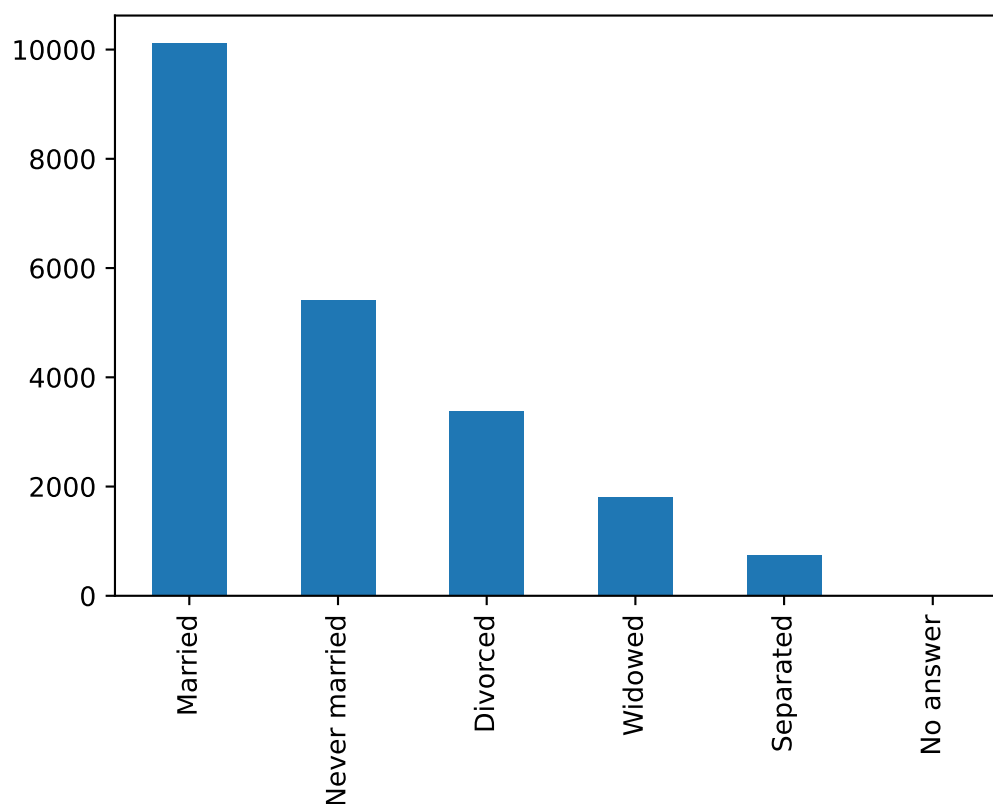
```
In [47]: df = pd.read_csv("../dataFiles/gss_cat.csv")
df.rename(columns={'Unnamed: 0': 'numbers'}, inplace=True)
df.head()
```

```
Out[47]:
```

	numbers	year	marital	age	race	rincome	partyid	relig	denom	tvhours
0	1	2000	Never married	26.0	White	\$8000 to 9999	Ind,near rep	Protestant	Southern baptist	12.0
1	2	2000	Divorced	48.0	White	\$8000 to 9999	Not str republican	Protestant	Baptist-dk which	NaN
2	3	2000	Widowed	67.0	White	Not applicable	Independent	Protestant	No denomination	2.0
3	4	2000	Never married	39.0	White	Not applicable	Ind,near rep	Orthodox-christian	Not applicable	4.0
4	5	2000	Divorced	25.0	White	Not applicable	Not str democrat	None	Not applicable	1.0

```
In [48]: # (a) create a bar plot for "marital". It should look like this:
df['marital'].value_counts().plot.bar()
```

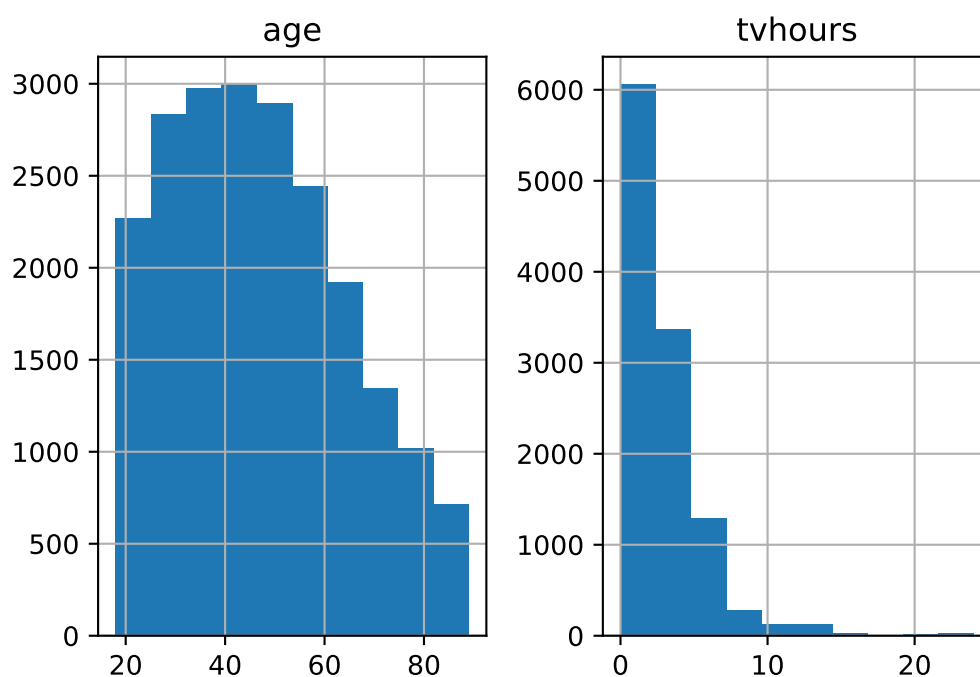
```
Out[48]: <AxesSubplot:>
```



In [49]:

```
# (b) Create a histogram for "age" on the Left, then a histogram for "tvhours" on the r
plt.figure(figsize=(10,5))
df.hist(column=["age", "tvhours"])
plt.show()
```

<Figure size 720x360 with 0 Axes>



In [50]:

```
# (c) You will create a line plot to see how the proportions of people identified as De
# change over time.
```

```
# (i) First, recode the variable "partyid"
party_newgroup = {"partyid" :
  {"No answer":"other",
   "Don't know":"other",
   "Other party":"other",
   "Strong republican":"rep",
   "Not str republican":"rep",
   "Ind,near rep":"ind",
   "Independent":"ind",
   "Ind,near dem":"ind",
   "Not str democrat":"dem",
   "Strong democrat":"dem"}}
df.replace(to_replace=party_newgroup, inplace=True)
df.head(45)
```

Out[50]:

	numbers	year	marital	age	race	rincome	partyid	relig	denom	tvhours
0	1	2000	Never married	26.0	White	\$8000 to 9999	ind	Protestant	Southern baptist	12.0
1	2	2000	Divorced	48.0	White	\$8000 to 9999	rep	Protestant	Baptist-dk which	NaN
2	3	2000	Widowed	67.0	White	Not applicable	ind	Protestant	No denomination	2.0
3	4	2000	Never married	39.0	White	Not applicable	ind	Orthodox-christian	Not applicable	4.0
4	5	2000	Divorced	25.0	White	Not applicable	dem	None	Not applicable	1.0
5	6	2000	Married	25.0	White	\$20000 - 24999	dem	Protestant	Southern baptist	NaN
6	7	2000	Never married	36.0	White	\$25000 or more	rep	Christian	Not applicable	3.0
7	8	2000	Divorced	44.0	White	\$7000 to 7999	ind	Protestant	Lutheran-mo synod	NaN
8	9	2000	Married	44.0	White	\$25000 or more	dem	Protestant	Other	0.0
9	10	2000	Married	47.0	White	\$25000 or more	rep	Protestant	Southern baptist	3.0
10	11	2000	Married	53.0	White	\$25000 or more	dem	Protestant	Other	2.0
11	12	2000	Married	52.0	White	\$25000 or more	ind	None	Not applicable	NaN
12	13	2000	Married	52.0	White	\$25000 or more	dem	Protestant	Southern baptist	1.0
13	14	2000	Married	51.0	White	\$25000 or more	rep	Protestant	United methodist	NaN
14	15	2000	Divorced	52.0	White	\$25000 or more	ind	None	Not applicable	1.0
15	16	2000	Married	40.0	Black	\$25000 or more	dem	Protestant	Baptist-dk which	7.0

	numbers	year	marital	age	race	rincome	partyid	relig	denom	tvhours
16	17	2000	Widowed	77.0	White	Not applicable	rep	Jewish	Not applicable	NaN
17	18	2000	Never married	44.0	White	\$25000 or more	ind	None	Not applicable	3.0
18	19	2000	Married	40.0	White	\$10000 - 14999	dem	Catholic	Not applicable	3.0
19	20	2000	Married	45.0	Black	Not applicable	ind	Protestant	United methodist	NaN
20	21	2000	Married	48.0	White	\$25000 or more	ind	Catholic	Not applicable	1.0
21	22	2000	Married	49.0	White	Refused	rep	Protestant	United methodist	2.0
22	23	2000	Never married	19.0	White	Not applicable	ind	None	Not applicable	2.0
23	24	2000	Widowed	54.0	White	\$25000 or more	ind	Christian	Not applicable	1.0
24	25	2000	Widowed	82.0	White	Not applicable	dem	Protestant	Other	3.0
25	26	2000	Widowed	83.0	White	Not applicable	dem	Protestant	Episcopal	NaN
26	27	2000	Widowed	89.0	White	Not applicable	dem	Protestant	Other lutheran	4.0
27	28	2000	Widowed	88.0	White	Not applicable	rep	Protestant	Afr meth ep zion	NaN
28	29	2000	Divorced	72.0	White	Not applicable	dem	Protestant	Southern baptist	7.0
29	30	2000	Widowed	82.0	White	Not applicable	ind	Protestant	Am bapt ch in usa	NaN
30	31	2000	Widowed	89.0	White	Not applicable	ind	Protestant	Other methodist	3.0
31	32	2000	Married	34.0	White	Refused	dem	Catholic	Not applicable	NaN
32	33	2000	Married	55.0	Black	Not applicable	dem	Protestant	Other	1.0
33	34	2000	Never married	37.0	White	\$25000 or more	ind	None	Not applicable	NaN
34	35	2000	Married	22.0	White	\$10000 - 14999	ind	Protestant	Other	NaN
35	36	2000	Never married	33.0	Other	\$20000 - 24999	dem	Christian	Not applicable	NaN
36	37	2000	Never married	37.0	White	\$25000 or more	ind	None	Not applicable	1.0

	numbers	year	marital	age	race	rincome	partyid	relig	denom	tvhours
37	38	2000	Never married	43.0	White	\$20000 - 24999	ind	Catholic	Not applicable	4.0
38	39	2000	Never married	29.0	White	\$25000 or more	ind	Catholic	Not applicable	2.0
39	40	2000	Never married	57.0	White	Refused	rep	Protestant	Southern baptist	1.0
40	41	2000	Married	31.0	White	\$20000 - 24999	ind	None	Not applicable	2.0
41	42	2000	Married	45.0	White	\$25000 or more	ind	Catholic	Not applicable	NaN
42	43	2000	Divorced	36.0	Black	\$15000 - 19999	dem	None	Not applicable	NaN
43	44	2000	Never married	52.0	White	Not applicable	ind	Other	Not applicable	1.0
44	45	2000	Never married	26.0	White	\$25000 or more	ind	None	Not applicable	1.0

In [51]:

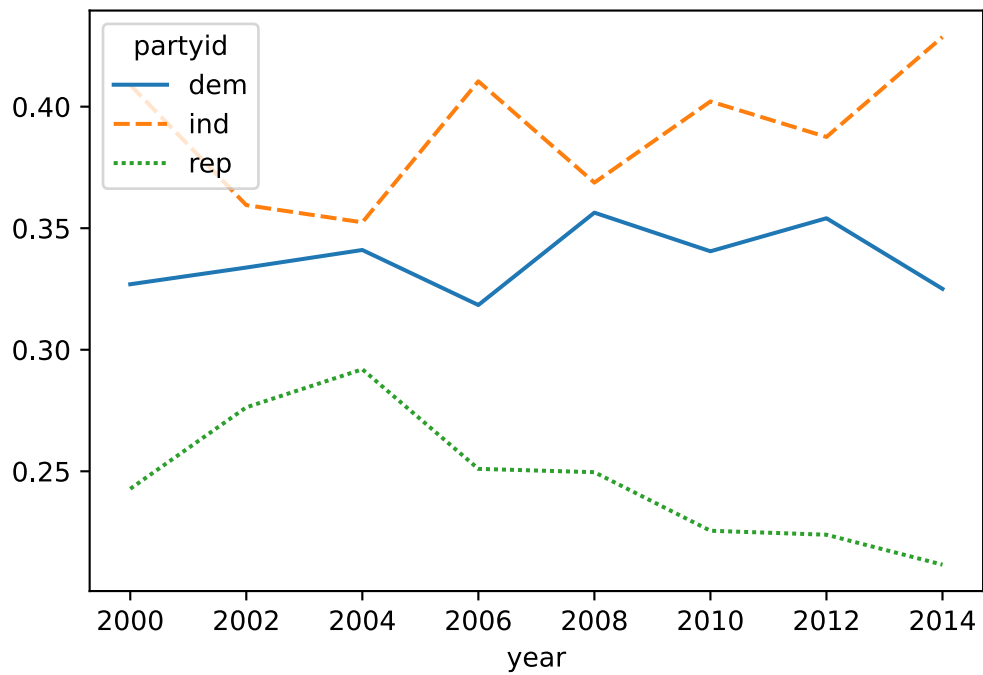
```
# (ii) then, create a table which shows the proportions of people identified  
# as Democrat, Republican, Independent and Other for each year.
```

```
party = df.groupby(['year', 'partyid'])['partyid'].count().unstack()
party['sum'] = party.sum(axis = 1)
party['dem'] = party['dem'] / party['sum']
party['ind'] = party['ind'] / party['sum']
party['other'] = party['other'] / party['sum']
party['rep'] = party['rep'] / party['sum']
party.drop(['sum'], axis=1, inplace = True)
```

In [52]:

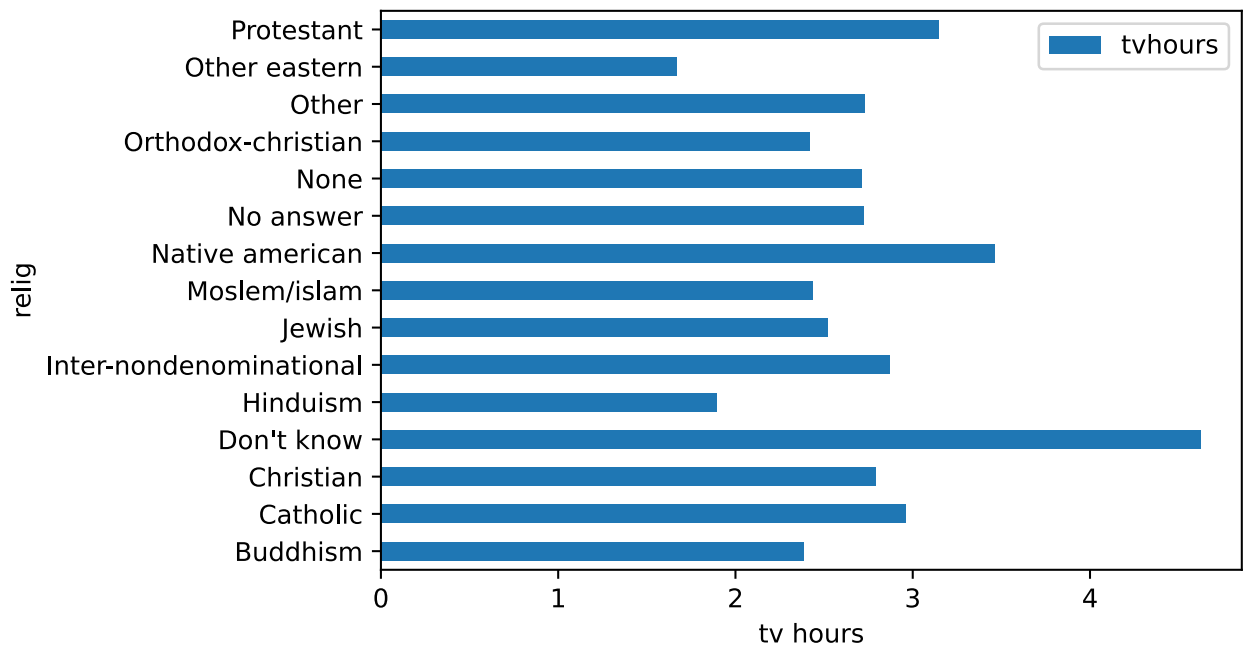
```
# (iii) Create a line plot to see how the proportions of people identified  
# as Democrat, Republican, and Independent changed over time.
party.drop(['other'], axis=1, inplace=True)
sns.lineplot(data=party)
```

Out[52]: <AxesSubplot:xlabel='year'>



In [53]:

```
# Finally, create a plot to show the average hours of tv viewing for each religion. The
grouped_df = df.groupby("relig").mean().reset_index()
new_table = grouped_df[['relig', 'tvhours']].set_index('relig').plot.barh().set_xlabel(
```



In []: