Step 2 - Read ACS (Assignment 1)

November 15, 2021

Table of Contents

[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

[1]:
```
import pandas as pd
import numpy as np
```

[2]:
```
acs = pd.read_csv('https://bitbucket.org/jimcody/sampledata/raw/
 ↪fb57313245ea48958efbb81a84f614bbdc765b33/acs2017.csv')
acs.columns= acs.columns.str.lower()
acs.head()
```

[2]:

	countyid	state	county	totalpop	men	women	hispanic	\
0	1001	Alabama	Autauga County	55036	26899	28137	2.7	
1	1003	Alabama	Baldwin County	203360	99527	103833	4.4	
2	1005	Alabama	Barbour County	26201	13976	12225	4.2	
3	1007	Alabama	Bibb County	22580	12251	10329	2.4	
4	1009	Alabama	Blount County	57667	28490	29177	9.0	

	white	black	native	asian	pacific	votingagecitizen	income	\
0	75.4	18.9	0.3	0.9	0.0	41016	55317	
1	83.1	9.5	0.8	0.7	0.0	155376	52562	
2	45.7	47.8	0.2	0.6	0.0	20269	33368	
3	74.6	22.0	0.4	0.0	0.0	17662	43404	
4	87.4	1.5	0.3	0.1	0.0	42513	47412	

	incomepercap	poverty	childpoverty	unemployment
0	27824	13.7	20.1	5.2
1	29364	11.8	16.1	5.5
2	17561	27.2	44.9	12.4
3	20911	15.2	26.6	8.2
4	22021	15.6	25.4	4.9

[3]:
```
acs_state = acs.groupby('state').agg({'totalpop':'sum','men':'sum',
```

```
                                        'women':'sum','income':'mean','poverty':
↪'mean'}).reset_index()
acs_state['income'] = acs_state['income'].round(2)
acs_state['poverty'] = acs_state['poverty'].round(2)

acs_state.head()

#acs_state = acs.groupby('state')['totalpop','men','women'].sum().reset_index()
#df_new = df.groupby(['col1', 'col2'])["col3", "col4"].sum()
```

[3]:

	state	totalpop	men	women	income	poverty
0	Alabama	4850771	2350806	2499965	40271.40	20.95
1	Alaska	738565	386319	352246	66425.45	13.56
2	Arizona	6809946	3385055	3424891	45816.67	20.19
3	Arkansas	2977944	1461651	1516293	39025.08	20.42
4	California	38982847	19366579	19616268	61046.76	15.74

[4]:
```
# Calculate gender majority
# If Men > Women then 'Male' else 'Female'

def majority(acs_state):
    if acs_state['men']> acs_state['women']: val = 'Male'
    else: val = 'Female'
    return val
```

[5]:
```
acs_state['gender_majority'] = acs_state.apply(majority, axis = 1)
acs_state.head()
```

[5]:

	state	totalpop	men	women	income	poverty	gender_majority
0	Alabama	4850771	2350806	2499965	40271.40	20.95	Female
1	Alaska	738565	386319	352246	66425.45	13.56	Male
2	Arizona	6809946	3385055	3424891	45816.67	20.19	Female
3	Arkansas	2977944	1461651	1516293	39025.08	20.42	Female
4	California	38982847	19366579	19616268	61046.76	15.74	Female

[6]:
```
acs_state.to_csv('state_acs', index = False)
```

1 Read from database

[7]:
```
import sqlalchemy
from sqlalchemy.sql import select
from sqlalchemy import create_engine
import psycopg2
import pandas as pd

db_string = 'postgresql://doadmin:pgwkRxvGIX1vmzvZ@diabetes-do-user-10225574-0.
↪b.db.ondigitalocean.com:25060/diabetes'
```

```
db = create_engine(db_string)
```

[8]:
```
states = pd.read_sql("""
           select * from state
            """, con = db)
states.head()
```

[8]:

	name	abbr	region_code	region
0	Alabama	AL	11	Southeast
1	Alaska	AK	12	West
2	American Samoa	AS	13	West
3	Arizona	AZ	10	Plains
4	Arkansas	AR	14	Southeast

[ ]:
```

```