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

In [2]:
```python
data = pd.read_csv("election_train.csv")
data.head()
```

Out[2]:

	Year	State	County	Office	Party	Votes
0	2018	AZ	Apache County	US Senator	Democratic	16298.0
1	2018	AZ	Apache County	US Senator	Republican	7810.0
2	2018	AZ	Cochise County	US Senator	Democratic	17383.0
3	2018	AZ	Cochise County	US Senator	Republican	26929.0
4	2018	AZ	Coconino County	US Senator	Democratic	34240.0

In [3]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2410 entries, 0 to 2409
Data columns (total 6 columns):
Year      2410 non-null int64
State     2410 non-null object
County    2410 non-null object
Office    2410 non-null object
Party     2410 non-null object
Votes     2400 non-null float64
dtypes: float64(1), int64(1), object(4)
memory usage: 113.0+ KB
```

In [4]:
```
data_a_wide = pd.pivot_table(data, index = ['Year', 'State','County','Office'], columns = 'Party',values = 'V
otes', aggfunc = np.sum).reset_index()
print(data_a_wide)
```

Party	Year	State	County	Office	Democratic	Republican
0	2018	AZ	Apache County	US Senator	16298.0	7810.0
1	2018	AZ	Cochise County	US Senator	17383.0	26929.0
2	2018	AZ	Coconino County	US Senator	34240.0	19249.0
3	2018	AZ	Gila County	US Senator	7643.0	12180.0
4	2018	AZ	Graham County	US Senator	3368.0	6870.0
5	2018	AZ	La Paz County	US Senator	1609.0	3265.0
6	2018	AZ	Maricopa County	US Senator	732671.0	672505.0
7	2018	AZ	Mohave County	US Senator	19214.0	50209.0
8	2018	AZ	Navajo County	US Senator	16624.0	18767.0
9	2018	AZ	Pima County	US Senator	221242.0	160550.0
10	2018	AZ	Santa Cruz County	US Senator	9241.0	3828.0
11	2018	AZ	Yavapai County	US Senator	40160.0	65308.0
12	2018	CT	Fairfield County	US Senator	210899.0	131321.0
13	2018	CT	Hartford County	US Senator	203591.0	123864.0
14	2018	CT	Middlesex County	US Senator	42383.0	32836.0
15	2018	CT	New Haven County	US Senator	179714.0	126004.0
16	2018	CT	Tolland County	US Senator	34732.0	28046.0
17	2018	CT	Windham County	US Senator	20490.0	19032.0
18	2018	DE	Sussex County	US Senator	40675.0	50391.0
19	2018	FL	Alachua County	US Senator	74493.0	40599.0
20	2018	FL	Baker County	US Senator	1945.0	8579.0
21	2018	FL	Bay County	US Senator	16723.0	46681.0
22	2018	FL	Bradford County	US Senator	2879.0	7576.0
23	2018	FL	Brevard County	US Senator	121112.0	160305.0
24	2018	FL	Broward County	US Senator	472239.0	211397.0
25	2018	FL	Charlotte County	US Senator	33525.0	52916.0
26	2018	FL	Citrus County	US Senator	22660.0	48008.0
27	2018	FL	Collier County	US Senator	54390.0	101266.0
28	2018	FL	Desoto County	US Senator	3328.0	5503.0
29	2018	FL	Dixie County	US Senator	1322.0	4442.0
...	...	...	...	...	...	...
1175	2018	WV	Pocahontas County	US Senator	1269.0	1411.0
1176	2018	WV	Preston County	US Senator	3686.0	5943.0
1177	2018	WV	Raleigh County	US Senator	10581.0	12620.0
1178	2018	WV	Randolph County	US Senator	4472.0	4017.0
1179	2018	WV	Ritchie County	US Senator	1082.0	1961.0
1180	2018	WV	Roane County	US Senator	2165.0	1899.0
1181	2018	WV	Summers County	US Senator	2069.0	1868.0
1182	2018	WV	Taylor County	US Senator	2376.0	2642.0
1183	2018	WV	Tucker County	US Senator	1469.0	1502.0
1184	2018	WV	Tyler County	US Senator	1065.0	1603.0
1185	2018	WV	Upshur County	US Senator	3102.0	4010.0

```
1186  2018   WV       Wayne County  US Senator    6395.0    5954.0
1187  2018   WV      Wetzel County  US Senator    2518.0    2135.0
1188  2018   WV        Wood County  US Senator   14189.0   13696.0
1189  2018   WV     Wyoming County  US Senator    2607.0    3096.0
1190  2018   WY      Albany County  US Senator    7576.0    6366.0
1191  2018   WY    Campbell County  US Senator    1628.0   11020.0
1192  2018   WY      Carbon County  US Senator    1359.0    3673.0
1193  2018   WY    Converse County  US Senator     834.0    3959.0
1194  2018   WY     Fremont County  US Senator    4734.0    9262.0
1195  2018   WY      Goshen County  US Senator    1020.0    3658.0
1196  2018   WY     Johnson County  US Senator     722.0    3085.0
1197  2018   WY     Lincoln County  US Senator    1152.0    5846.0
1198  2018   WY     Natrona County  US Senator    7285.0   16359.0
1199  2018   WY    Niobrara County  US Senator     144.0     980.0
1200  2018   WY      Platte County  US Senator     801.0    2850.0
1201  2018   WY    Sublette County  US Senator     668.0    2653.0
1202  2018   WY  Sweetwater County  US Senator    3943.0    8577.0
1203  2018   WY       Uinta County  US Senator    1371.0    4713.0
1204  2018   WY    Washakie County  US Senator     588.0    2423.0

[1205 rows x 6 columns]
```

In [5]:

```
data_a_wide.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1205 entries, 0 to 1204
Data columns (total 6 columns):
Year          1205 non-null int64
State         1205 non-null object
County        1205 non-null object
Office        1205 non-null object
Democratic    1205 non-null float64
Republican    1205 non-null float64
dtypes: float64(2), int64(1), object(3)
memory usage: 56.6+ KB
```

In [6]:
```
demo = pd.read_csv("demographics_train.csv")
demo.head()
```

Out[6]:

	State	County	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Percent Age 29 and Under	Percent Age 65 and Older	M Hous In
0	Wisconsin	La Crosse	55063	117538	0	90.537528	1.214075	1.724549	2.976059	51.171536	43.241335	14.702479	
1	Virginia	Alleghany	51005	15919	12705	91.940449	5.207614	1.432251	1.300333	51.077329	31.660280	23.902255	
2	Indiana	Fountain	18045	16741	12750	95.705155	0.400215	2.359477	1.547100	49.770026	35.899887	18.941521	
3	Ohio	Geauga	39055	94020	0	95.837056	1.256116	1.294405	2.578175	50.678579	36.281642	18.028079	
4	Wisconsin	Jackson	55053	20566	15835	86.662453	1.983857	3.082758	1.376058	46.649810	36.292911	17.587280	

In [7]:
```
demo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1216 entries, 0 to 1215
Data columns (total 17 columns):
State                                    1216 non-null object
County                                   1216 non-null object
FIPS                                     1216 non-null int64
Total Population                         1216 non-null int64
Citizen Voting-Age Population            1216 non-null int64
Percent White, not Hispanic or Latino    1216 non-null float64
Percent Black, not Hispanic or Latino    1216 non-null float64
Percent Hispanic or Latino               1216 non-null float64
Percent Foreign Born                     1216 non-null float64
Percent Female                           1216 non-null float64
Percent Age 29 and Under                 1216 non-null float64
Percent Age 65 and Older                 1216 non-null float64
Median Household Income                  1216 non-null int64
Percent Unemployed                       1216 non-null float64
Percent Less than High School Degree     1216 non-null float64
Percent Less than Bachelor's Degree      1216 non-null float64
Percent Rural                            1216 non-null float64
dtypes: float64(11), int64(4), object(2)
memory usage: 161.6+ KB
```

In [8]:
```
state_abbrev = pd.Series({

    'Alabama': 'AL',

    'Alaska': 'AK',

    'Arizona': 'AZ',

    'Arkansas': 'AR',

    'California': 'CA',

    'Colorado': 'CO',

    'Connecticut': 'CT',

    'Delaware': 'DE',

    'District of Columbia': 'DC',

    'Florida': 'FL',

    'Georgia': 'GA',

    'Hawaii': 'HI',

    'Idaho': 'ID',

    'Illinois': 'IL',

    'Indiana': 'IN',

    'Iowa': 'IA',

    'Kansas': 'KS',

    'Kentucky': 'KY',

    'Louisiana': 'LA',

    'Maine': 'ME',
```

```
    'Maryland': 'MD',

    'Massachusetts': 'MA',

    'Michigan': 'MI',

    'Minnesota': 'MN',

    'Mississippi': 'MS',

    'Missouri': 'MO',

    'Montana': 'MT',

    'Nebraska': 'NE',

    'Nevada': 'NV',

    'New Hampshire': 'NH',

    'New Jersey': 'NJ',

    'New Mexico': 'NM',

    'New York': 'NY',

    'North Carolina': 'NC',

    'North Dakota': 'ND',

    'Northern Mariana Islands':'MP',

    'Ohio': 'OH',

    'Oklahoma': 'OK',

    'Oregon': 'OR',

    'Palau': 'PW',

    'Pennsylvania': 'PA',

    'Puerto Rico': 'PR',
```

```
    'Rhode Island': 'RI',

    'South Carolina': 'SC',

    'South Dakota': 'SD',

    'Tennessee': 'TN',

    'Texas': 'TX',

    'Utah': 'UT',

    'Vermont': 'VT',

    'Virgin Islands': 'VI',

    'Virginia': 'VA',

    'Washington': 'WA',

    'West Virginia': 'WV',

    'Wisconsin': 'WI',

    'Wyoming': 'WY',

})
demo["State"] = demo["State"].map(state_abbrev)
demo.sort_values(by = ['State']).head()
```

Out[8]:

	State	County	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Percent Age 29 and Under	Percent Age 65 and Older	Me House Inc
1161	AZ	Pima	4019	1003338	0	53.271579	3.199719	36.105978	12.903428	50.807405	40.087388	17.801778	4
195	AZ	Graham	4009	37529	0	51.461536	1.811932	32.097844	4.385942	46.313518	46.393456	12.315809	4
16	AZ	Yavapai	4025	218586	0	81.159361	0.518331	14.054880	6.456955	51.092476	28.717301	28.272625	4
429	AZ	Navajo	4017	108209	76280	41.927196	0.672772	11.049913	2.914730	49.846131	43.243168	15.745456	3
106	AZ	Santa Cruz	4023	46547	27155	15.274883	0.199798	83.219112	32.644424	52.125808	43.300320	15.895761	3

In [9]:
```
demo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1216 entries, 0 to 1215
Data columns (total 17 columns):
State                                      1216 non-null object
County                                     1216 non-null object
FIPS                                       1216 non-null int64
Total Population                           1216 non-null int64
Citizen Voting-Age Population              1216 non-null int64
Percent White, not Hispanic or Latino      1216 non-null float64
Percent Black, not Hispanic or Latino      1216 non-null float64
Percent Hispanic or Latino                 1216 non-null float64
Percent Foreign Born                       1216 non-null float64
Percent Female                             1216 non-null float64
Percent Age 29 and Under                   1216 non-null float64
Percent Age 65 and Older                   1216 non-null float64
Median Household Income                    1216 non-null int64
Percent Unemployed                         1216 non-null float64
Percent Less than High School Degree       1216 non-null float64
Percent Less than Bachelor's Degree        1216 non-null float64
Percent Rural                              1216 non-null float64
dtypes: float64(11), int64(4), object(2)
memory usage: 161.6+ KB
```

In [10]:
```
data_a_wide['County'] = data_a_wide['County'].str.replace(' County', '')
data_a_wide['County'] = data_a_wide['County'].str.lower()

demo['County'] = demo['County'].str.lower()
```

In [11]:
```
data_a_wide.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1205 entries, 0 to 1204
Data columns (total 6 columns):
Year          1205 non-null int64
State         1205 non-null object
County        1205 non-null object
Office        1205 non-null object
Democratic    1205 non-null float64
Republican    1205 non-null float64
dtypes: float64(2), int64(1), object(3)
memory usage: 56.6+ KB
```

In [12]:
```
#demo['County'].sort_values()
```

In [13]:
```
#data_a_wide['County'].sort_values()
```

In [14]:
```
#demo[['State','County']].sort_values(by = ['State','County']).reset_index()
```

In [15]:
```
# Merge dataset data_a_wide and dataset demo
data_merged = pd.merge(data_a_wide, demo, how = 'inner',
                       on = ['State','County'], sort = True)
#data1 = data_merged.sort_values(by = 'Citizen Voting-Age Population')
#print(data1.head())
print(data_merged)
```

	Year	State	County	Office	Democratic	Republican	FIPS	\
0	2018	AZ	apache	US Senator	16298.0	7810.0	4001	
1	2018	AZ	cochise	US Senator	17383.0	26929.0	4003	
2	2018	AZ	coconino	US Senator	34240.0	19249.0	4005	
3	2018	AZ	gila	US Senator	7643.0	12180.0	4007	
4	2018	AZ	graham	US Senator	3368.0	6870.0	4009	
5	2018	AZ	la paz	US Senator	1609.0	3265.0	4012	
6	2018	AZ	maricopa	US Senator	732671.0	672505.0	4013	
7	2018	AZ	mohave	US Senator	19214.0	50209.0	4015	
8	2018	AZ	navajo	US Senator	16624.0	18767.0	4017	
9	2018	AZ	pima	US Senator	221242.0	160550.0	4019	
10	2018	AZ	santa cruz	US Senator	9241.0	3828.0	4023	
11	2018	AZ	yavapai	US Senator	40160.0	65308.0	4025	
12	2018	CT	fairfield	US Senator	210899.0	131321.0	9001	
13	2018	CT	hartford	US Senator	203591.0	123864.0	9003	
14	2018	CT	middlesex	US Senator	42383.0	32836.0	9007	
15	2018	CT	new haven	US Senator	179714.0	126004.0	9009	
16	2018	CT	tolland	US Senator	34732.0	28046.0	9013	
17	2018	CT	windham	US Senator	20490.0	19032.0	9015	
18	2018	DE	sussex	US Senator	40675.0	50391.0	10005	
19	2018	FL	alachua	US Senator	74493.0	40599.0	12001	
20	2018	FL	baker	US Senator	1945.0	8579.0	12003	
21	2018	FL	bay	US Senator	16723.0	46681.0	12005	
22	2018	FL	bradford	US Senator	2879.0	7576.0	12007	
23	2018	FL	brevard	US Senator	121112.0	160305.0	12009	
24	2018	FL	broward	US Senator	472239.0	211397.0	12011	
25	2018	FL	charlotte	US Senator	33525.0	52916.0	12015	
26	2018	FL	citrus	US Senator	22660.0	48008.0	12017	
27	2018	FL	collier	US Senator	54390.0	101266.0	12021	
28	2018	FL	desoto	US Senator	3328.0	5503.0	12027	
29	2018	FL	dixie	US Senator	1322.0	4442.0	12029	
...	...	...	...	...	...	...	...	
1170	2018	WV	pocahontas	US Senator	1269.0	1411.0	54075	
1171	2018	WV	preston	US Senator	3686.0	5943.0	54077	
1172	2018	WV	raleigh	US Senator	10581.0	12620.0	54081	
1173	2018	WV	randolph	US Senator	4472.0	4017.0	54083	
1174	2018	WV	ritchie	US Senator	1082.0	1961.0	54085	
1175	2018	WV	roane	US Senator	2165.0	1899.0	54087	
1176	2018	WV	summers	US Senator	2069.0	1868.0	54089	
1177	2018	WV	taylor	US Senator	2376.0	2642.0	54091	
1178	2018	WV	tucker	US Senator	1469.0	1502.0	54093	
1179	2018	WV	tyler	US Senator	1065.0	1603.0	54095	
1180	2018	WV	upshur	US Senator	3102.0	4010.0	54097	

```
1181  2018    WV       wayne  US Senator      6395.0      5954.0  54099
1182  2018    WV      wetzel  US Senator      2518.0      2135.0  54103
1183  2018    WV        wood  US Senator     14189.0     13696.0  54107
1184  2018    WV     wyoming  US Senator      2607.0      3096.0  54109
1185  2018    WY      albany  US Senator      7576.0      6366.0  56001
1186  2018    WY    campbell  US Senator      1628.0     11020.0  56005
1187  2018    WY      carbon  US Senator      1359.0      3673.0  56007
1188  2018    WY    converse  US Senator       834.0      3959.0  56009
1189  2018    WY     fremont  US Senator      4734.0      9262.0  56013
1190  2018    WY      goshen  US Senator      1020.0      3658.0  56015
1191  2018    WY     johnson  US Senator       722.0      3085.0  56019
1192  2018    WY     lincoln  US Senator      1152.0      5846.0  56023
1193  2018    WY     natrona  US Senator      7285.0     16359.0  56025
1194  2018    WY    niobrara  US Senator       144.0       980.0  56027
1195  2018    WY      platte  US Senator       801.0      2850.0  56031
1196  2018    WY    sublette  US Senator       668.0      2653.0  56035
1197  2018    WY  sweetwater  US Senator      3943.0      8577.0  56037
1198  2018    WY       uinta  US Senator      1371.0      4713.0  56041
1199  2018    WY    washakie  US Senator       588.0      2423.0  56043

      Total Population  Citizen Voting-Age Population  \
0                72346                              0
1               128177                          92915
2               138064                         104265
3                53179                              0
4                37529                              0
5                20304                          15245
6              4088549                        2723565
7               203629                              0
8               108209                          76280
9              1003338                              0
10               46547                          27155
11              218586                              0
12              941618                              0
13              895699                         644940
14              164438                              0
15              860874                         631715
16              151689                              0
17              117078                              0
18              211224                              0
19              256581                         197720
20               27312                          20415
21              178361                         135795
```

```
22             26919                         0
23            560683                    438510
24           1863780                         0
25            169642                    141230
26            140453                         0
27            348236                         0
28             35134                         0
29             16084                     12890
...              ...                       ...
1170            8620                      7070
1171           33793                     27010
1172           78051                         0
1173           29287                     23555
1174           10044                         0
1175           14513                         0
1176           13325                     10995
1177           16949                     13450
1178            6922                      5690
1179            9000                      7205
1180           24632                         0
1181           41237                     32220
1182           15997                         0
1183           86262                     67640
1184           22537                     17730
1185           37836                     30070
1186           48473                         0
1187           15696                     11335
1188           14223                         0
1189           40683                     30170
1190           13546                         0
1191            8572                      6590
1192           18543                         0
1193           80871                     60415
1194            2498                      1995
1195            8740                      6830
1196           10032                         0
1197           44812                     30565
1198           20893                     14355
1199            8351                         0

      Percent White, not Hispanic or Latino  ...  Percent Hispanic or Latino  \
0                                 18.571863  ...                    5.947806
1                                 56.299492  ...                   34.403208
```

```
2                                54.619597  ...                 13.711033
3                                63.222325  ...                 18.548675
4                                51.461536  ...                 32.097844
5                                58.884949  ...                 26.182033
6                                56.918114  ...                 30.286833
7                                78.252606  ...                 15.708470
8                                41.927196  ...                 11.049913
9                                53.271579  ...                 36.105978
10                               15.274883  ...                 83.219112
11                               81.159361  ...                 14.054880
12                               63.509512  ...                 18.636007
13                               63.252052  ...                 16.897976
14                               84.738929  ...                  5.641032
15                               64.782767  ...                 16.790959
16                               85.892187  ...                  5.029369
17                               83.782606  ...                 10.877364
18                               74.829091  ...                  9.195925
19                               62.460198  ...                  8.994820
20                               82.088459  ...                  2.317663
21                               77.687387  ...                  5.728270
22                               74.787325  ...                  3.796575
23                               75.798803  ...                  9.301691
24                               39.245351  ...                 27.564841
25                               84.643543  ...                  6.632202
26                               88.871010  ...                  5.121998
27                               64.222539  ...                 26.634524
28                               55.040701  ...                 30.688222
29                               85.041035  ...                  3.780154
...                                    ...  ...                       ...
1170                             95.696056  ...                  0.893271
1171                             92.365283  ...                  1.831740
1172                             87.123804  ...                  1.475958
1173                             96.182607  ...                  0.816062
1174                             97.779769  ...                  0.756671
1175                             97.050920  ...                  0.957762
1176                             91.639775  ...                  1.696060
1177                             96.300667  ...                  0.967609
1178                             97.688529  ...                  0.650101
1179                             98.000000  ...                  0.711111
1180                             96.354336  ...                  1.181390
1181                             97.524068  ...                  0.613527
1182                             97.574545  ...                  0.481340
1183                             95.478890  ...                  1.032900
```

```
1184                                   97.475263  ...                  0.257355
1185                                   83.269902  ...                  9.229305
1186                                   87.774637  ...                  8.357230
1187                                   77.924312  ...                 17.647808
1188                                   88.849047  ...                  7.691767
1189                                   70.198855  ...                  6.779244
1190                                   86.409272  ...                 10.519711
1191                                   91.565562  ...                  2.134858
1192                                   92.600982  ...                  4.416761
1193                                   87.026252  ...                  8.198242
1194                                   89.511609  ...                  4.243395
1195                                   89.359268  ...                  7.814645
1196                                   91.646730  ...                  7.814992
1197                                   79.815674  ...                 15.859591
1198                                   87.718375  ...                  8.959939
1199                                   82.397318  ...                 13.962400

      Percent Foreign Born  Percent Female  Percent Age 29 and Under  \
0                 1.719515       50.598513                 45.854643
1                11.458374       49.069646                 37.902276
2                 4.825298       50.581614                 48.946141
3                 4.249798       50.296170                 32.238290
4                 4.385942       46.313518                 46.393456
5                11.372143       48.946020                 28.073286
6                14.729333       50.549278                 41.886620
7                 6.969047       49.676618                 30.485835
8                 2.914730       49.846131                 43.243168
9                12.903428       50.807405                 40.087388
10               32.644424       52.125808                 43.300320
11                6.456955       51.092476                 28.717301
12               21.154120       51.308280                 38.019876
13               15.011070       51.492075                 37.389458
14                7.652732       51.098894                 33.241708
15               12.073428       51.759956                 37.982562
16                6.769113       49.967367                 42.301024
17                5.269991       50.422795                 37.389604
18                6.671117       51.495569                 31.746866
19               10.144555       51.670623                 48.551530
20                1.475542       47.715290                 41.344464
21                5.479898       50.280050                 37.739192
22                2.306921       44.741632                 35.343066
23                8.632685       51.140841                 31.972612
24               32.715718       51.376504                 36.701971
```

```
25          10.185567    51.260891          22.388913
26           5.413911    51.667818          24.581889
27          23.816894    50.881299          29.998909
28          17.504412    43.345477          37.513520
29           2.785377    45.417807          31.236011
...               ...          ...                ...
1170         0.881671    48.364269          29.408353
1171         1.003166    48.477495          33.465511
1172         1.680952    49.720055          35.009161
1173         0.669239    48.714447          34.011678
1174         0.169255    50.398248          32.487057
1175         0.496107    50.768277          33.404534
1176         0.435272    53.696060          28.900563
1177         0.749307    49.247743          33.482801
1178         0.592314    50.245594          29.731292
1179         0.233333    50.211111          31.833333
1180         0.633323    49.837610          37.950633
1181         0.603827    51.378616          34.481170
1182         0.443833    50.997062          33.168719
1183         1.001600    51.756277          34.726763
1184         0.257355    50.547988          33.877623
1185         6.081510    47.756105          55.447193
1186         3.645328    48.008170          45.326264
1187         5.198777    45.839704          39.181957
1188         2.706883    49.933207          38.515081
1189         1.339626    49.907824          39.751247
1190         2.724051    47.091392          35.914661
1191         1.656556    46.966869          32.571162
1192         2.151755    48.773122          38.715418
1193         2.729038    49.421919          40.688257
1194         0.280224    54.443555          36.269015
1195         2.780320    47.711670          32.700229
1196         2.053429    46.949761          36.393541
1197         5.509685    47.824244          44.153352
1198         3.986981    49.327526          43.205858
1199         3.783978    51.359119          34.774279

      Percent Age 65 and Older  Median Household Income  Percent Unemployed  \
0                    13.322091                    32460           15.807433
1                    19.756275                    45383            8.567108
2                    10.873943                    51106            8.238305
3                    26.397638                    40593           12.129932
4                    12.315809                    47422           14.424104
```

5	36.056935	36321	10.599013
6	13.837843	55676	6.808454
7	26.858650	39856	11.680953
8	15.745456	36868	18.525791
9	17.801778	46764	9.214114
10	15.895761	38941	9.749896
11	28.272625	46638	8.525986
12	14.383115	86670	8.211898
13	15.644876	68027	8.237243
14	17.703937	79837	5.257635
15	15.604490	62715	8.545483
16	13.868507	80129	6.267325
17	14.571482	60689	8.375524
18	24.214578	54218	7.108621
19	12.430772	44702	7.020154
20	12.910076	53327	6.832522
21	16.103857	48577	7.475668
22	17.452357	43373	10.563312
23	22.491140	49914	9.278023
24	15.371879	52954	8.710312
25	37.622759	44865	10.088979
26	35.105694	39054	11.648012
27	29.589704	59783	6.408868
28	19.408550	35513	8.291825
29	21.729669	34634	5.306200
...	...	...	...
1170	22.331787	36026	5.685358
1171	17.429645	45221	7.707416
1172	18.099704	41533	7.709258
1173	19.814252	40308	7.425664
1174	20.141378	40850	8.428745
1175	19.506649	34144	10.515990
1176	21.275797	35620	10.729049
1177	18.065963	44371	9.082581
1178	22.594626	43529	8.190184
1179	20.133333	38674	8.458542
1180	18.281098	42240	7.692308
1181	18.289400	38311	9.504391
1182	21.278990	39446	8.472337
1183	18.559737	43944	7.625458
1184	17.846208	35469	12.077958
1185	9.549107	43043	4.579174
1186	6.954387	80822	4.669432

```
1187          13.774210          56972          4.141937
1188          13.668003          66737          5.282284
1189          16.409803          53559          7.344324
1190          20.389783          44883          6.918819
1191          20.496967          54594          4.512276
1192          14.382786          64579          5.618095
1193          12.825364          56983          4.861868
1194          18.254604          40640          0.457875
1195          22.013730          41051          3.901047
1196          13.337321          76004          2.786971
1197           9.417120          68233          5.072255
1198          10.678218          53323          6.390755
1199          19.650341          46212          7.441860

      Percent Less than High School Degree  \
0                                21.758252
1                                13.409171
2                                11.085381
3                                15.729958
4                                14.580797
5                                24.842215
6                                13.051927
7                                16.145850
8                                18.494087
9                                12.252238
10                               25.206726
11                                9.830672
12                               10.103521
13                               10.736314
14                                6.153444
15                               10.503197
16                                6.214203
17                               11.925239
18                               13.792481
19                                7.468059
20                               17.871127
21                               11.320101
22                               23.212996
23                                8.823708
24                               11.688725
25                               10.578890
26                               13.288908
27                               14.321540
```

```
28                                 29.475750
29                                 22.236087
...                                      ...
1170                               15.241010
1171                               16.522582
1172                               15.658951
1173                               17.197844
1174                               18.066123
1175                               21.849308
1176                               17.191498
1177                               14.098280
1178                               12.702390
1179                               12.429292
1180                               16.125150
1181                               20.628495
1182                               16.947084
1183                               10.370080
1184                               22.638690
1185                                4.210167
1186                                8.330054
1187                                9.579879
1188                                9.758393
1189                                8.537172
1190                                8.390574
1191                                5.105750
1192                                6.949996
1193                                8.332255
1194                               11.054422
1195                                9.675889
1196                                4.658830
1197                                9.314606
1198                               10.361224
1199                               12.577108

      Percent Less than Bachelor's Degree  Percent Rural
0                               88.941063      74.061076
1                               76.837055      36.301067
2                               65.791439      31.466066
3                               82.262624      41.062000
4                               86.675944      46.437399
5                               89.563407      56.327786
6                               69.031137       2.363800
7                               88.121178      22.963644
```

8	85.507970	54.138242
9	69.199391	7.523491
10	77.506775	26.883172
11	74.458362	33.197178
12	53.637538	4.584061
13	63.133913	5.408640
14	59.071541	24.540066
15	65.711800	3.639981
16	61.140250	38.207884
17	76.143316	49.761036
18	76.115066	41.303609
19	58.485040	21.193437
20	87.235709	59.491057
21	77.466127	11.999858
22	88.220732	75.529453
23	72.271467	5.074387
24	68.984269	0.016933
25	78.396083	8.884347
26	82.154355	34.518820
27	66.118970	8.485942
28	89.960743	46.242327
29	93.602862	77.024723
...	...	...
1170	83.794950	100.000000
1171	85.424578	90.450477
1172	81.850342	39.258677
1173	81.700397	62.383948
1174	89.611288	100.000000
1175	88.316085	79.994640
1176	85.219445	72.126086
1177	81.476652	59.704054
1178	85.794140	100.000000
1179	86.255924	91.051260
1180	81.438026	64.001814
1181	87.121342	65.186789
1182	89.251155	53.856359
1183	79.738786	26.774461
1184	91.787560	88.884687
1185	52.136794	11.939723
1186	80.935983	29.072464
1187	79.681613	41.630469
1188	84.468152	55.360370
1189	77.648670	51.424370

```
1190                         77.036880      45.995924
1191                         74.537343      49.025557
1192                         79.631291      82.729482
1193                         77.786814      14.449304
1194                         81.009070     100.000000
1195                         80.300395      58.647744
1196                         75.645069     100.000000
1197                         78.628507      10.916313
1198                         81.793082      43.095937
1199                         78.923920      35.954529

[1200 rows x 21 columns]
```

In [16]:

```
data_merged.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1200 entries, 0 to 1199
Data columns (total 21 columns):
Year                                       1200 non-null int64
State                                      1200 non-null object
County                                     1200 non-null object
Office                                     1200 non-null object
Democratic                                 1200 non-null float64
Republican                                 1200 non-null float64
FIPS                                       1200 non-null int64
Total Population                           1200 non-null int64
Citizen Voting-Age Population              1200 non-null int64
Percent White, not Hispanic or Latino      1200 non-null float64
Percent Black, not Hispanic or Latino      1200 non-null float64
Percent Hispanic or Latino                 1200 non-null float64
Percent Foreign Born                       1200 non-null float64
Percent Female                             1200 non-null float64
Percent Age 29 and Under                   1200 non-null float64
Percent Age 65 and Older                   1200 non-null float64
Median Household Income                    1200 non-null int64
Percent Unemployed                         1200 non-null float64
Percent Less than High School Degree       1200 non-null float64
Percent Less than Bachelor's Degree        1200 non-null float64
Percent Rural                              1200 non-null float64
dtypes: float64(13), int64(5), object(3)
memory usage: 206.2+ KB
```

In [17]:
```
#Task3
# There are 21 attributes in the merged dataset.
#The attributes "Year", "Office" are the irrelevant attributes.
#But there are no redundant attributes"
#We are removing those irrelevant attributes
```

In [18]:
```
#Task3
data_merged = data_merged.drop(['Year', 'Office'], axis=1)
data_merged.head()
```

Out[18]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Perc Age Un
0	AZ	apache	16298.0	7810.0	4001	72346	0	18.571863	0.486551	5.947806	1.719515	50.598513	45.854
1	AZ	cochise	17383.0	26929.0	4003	128177	92915	56.299492	3.714395	34.403208	11.458374	49.069646	37.902
2	AZ	coconino	34240.0	19249.0	4005	138064	104265	54.619597	1.342855	13.711033	4.825298	50.581614	48.946
3	AZ	gila	7643.0	12180.0	4007	53179	0	63.222325	0.552850	18.548675	4.249798	50.296170	32.238
4	AZ	graham	3368.0	6870.0	4009	37529	0	51.461536	1.811932	32.097844	4.385942	46.313518	46.393

In [19]:
```
#Task 4
data_merged.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1200 entries, 0 to 1199
Data columns (total 19 columns):
State                                   1200 non-null object
County                                  1200 non-null object
Democratic                              1200 non-null float64
Republican                              1200 non-null float64
FIPS                                    1200 non-null int64
Total Population                        1200 non-null int64
Citizen Voting-Age Population           1200 non-null int64
Percent White, not Hispanic or Latino   1200 non-null float64
Percent Black, not Hispanic or Latino   1200 non-null float64
Percent Hispanic or Latino              1200 non-null float64
Percent Foreign Born                    1200 non-null float64
Percent Female                          1200 non-null float64
Percent Age 29 and Under                1200 non-null float64
Percent Age 65 and Older                1200 non-null float64
Median Household Income                 1200 non-null int64
Percent Unemployed                      1200 non-null float64
Percent Less than High School Degree    1200 non-null float64
Percent Less than Bachelor's Degree     1200 non-null float64
Percent Rural                           1200 non-null float64
dtypes: float64(13), int64(4), object(2)
memory usage: 187.5+ KB
```

In [20]:
```
#Task 4

#Yes. There are many zeroes (missing values) in the Citizen Voting-Age Population. Replacing them with null v
alues  . .
```

In [21]:
```
#Task 4
#data_merged["Citizen Voting-Age Population"][data_merged["Citizen Voting-Age Population"] == 0] = 'nan'
data_merged["Citizen Voting-Age Population"].replace(0, np.nan, inplace=True)
data_merged.head()
```

Out[21]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Perc Age Un
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	1.719515	50.598513	45.854
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	11.458374	49.069646	37.902
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	4.825298	50.581614	48.946
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	4.249798	50.296170	32.238
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	4.385942	46.313518	46.393

In [22]:
```
#Task 5
data_merged['Party'] = (data_merged["Democratic"] > data_merged["Republican"]).astype("int64")
data_merged.head()
```

Out[22]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Perc Age Un
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	1.719515	50.598513	45.854
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	11.458374	49.069646	37.902
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	4.825298	50.581614	48.946
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	4.249798	50.296170	32.238
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	4.385942	46.313518	46.393

In [23]:
```
#Task 6
democratic = data_merged["Total Population"][data_merged["Party"] == 1]
print(democratic.mean())
republic = data_merged["Total Population"][data_merged["Party"] == 0]
print(republic.mean())
```

```
300998.3169230769
53974.214857142855
```

In [24]:
```
# TASK 6
from scipy import stats
[stat, pvalue] = stats.ttest_ind(democratic, republic, equal_var = False)
print(round(stat,3))
print(pvalue)
```

```
8.001
2.0965719353509958e-14
```

In [25]:
```
#Task 6
# As pvalue is less than the given significance level = 0.05,
# we can reject the null hypothesis that the population mean of democratic and republic counties are equal.
```

In [26]:
```
#Task 7
democratic_MHI = data_merged["Median Household Income"][data_merged["Party"] == 1]
print(democratic_MHI.mean())
republic_MHI = data_merged["Median Household Income"][data_merged["Party"] == 0]
print(republic_MHI.mean())
```

```
53798.732307692306
48724.15085714286
```

In [27]:
```
# TASK 7
[stat, pvalue] = stats.ttest_ind(democratic_MHI, republic_MHI, equal_var = False)
print(round(stat,3))
print(pvalue)
```

```
5.507
6.173239891230373e-08
```

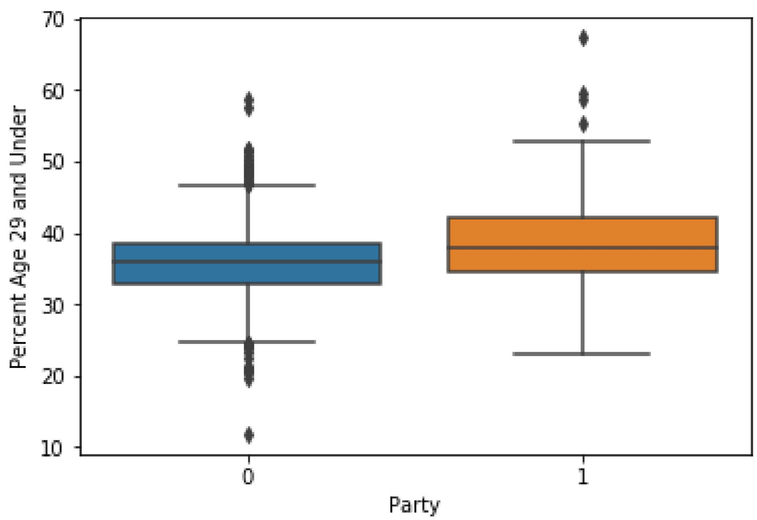
In [28]:
```
#Task 7
# As pvalue is less than the given significance level = 0.05,
# we can reject the null hypothesis that the mean Median Household Income of democratic and republic counties
are equal.
```

In [29]:
```
# Task 8
import matplotlib.pyplot as plt
import seaborn as sns
print(data_merged.groupby("Party").describe()["Citizen Voting-Age Population"])
CVAP = sns.boxplot(x = data_merged["Party"], y = data_merged["Citizen Voting-Age Population"], data = data_me
rged)
print(CVAP)
```

```
       count           mean            std     min      25%      50%  \
Party
0      386.0   39270.012953   64700.902510    60.0   8212.5  19780.0
1      134.0  175827.425373  319459.705034  1575.0  18617.5  54337.5

             75%        max
Party
0       38796.25   460215.0
1      163812.50  2723565.0
AxesSubplot(0.125,0.11;0.775x0.77)
```

In [30]:
```
#Task 8
print(data_merged.groupby("Party").describe()["Percent Age 29 and Under"])
ageLT29 = sns.boxplot(x ="Party", y = "Percent Age 29 and Under", data = data_merged)
print(ageLT29)
```

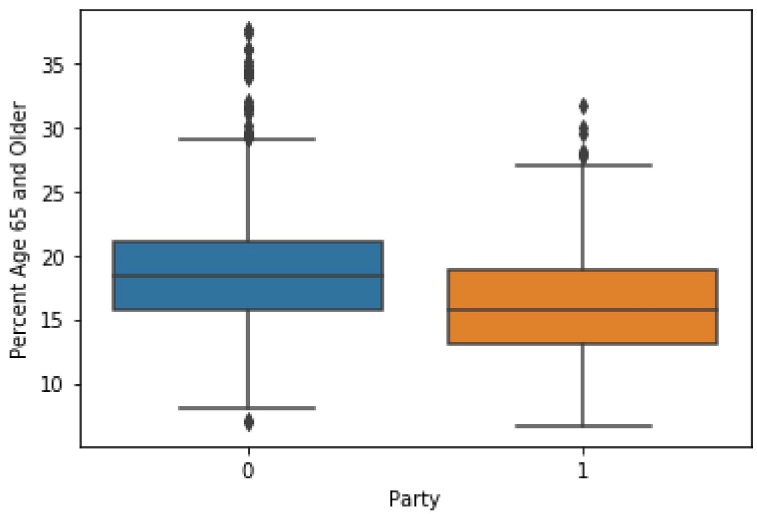
```
       count       mean       std        min        25%        50%        75%  \
Party
0      875.0  36.020984  5.179824  11.842105  33.003249  35.864651  38.548722
1      325.0  38.726959  6.252786  23.156452  34.488444  38.074151  42.161162

             max
Party
0      58.749116
1      67.367823
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [31]:
```
#Task 8
# The age group 29 and under has not much effect on the voting. The difference between the medians for both p
arties is not significant.
```

In [32]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Age 65 and Older"])
ageGT65 = sns.boxplot(x ="Party", y = "Percent Age 65 and Older", data = data_merged)
print(ageGT65)
```

```
       count       mean       std       min        25%        50%        75%  \
Party
0      875.0  18.814997  4.733641  6.954387  15.781389  18.377039  21.109296
1      325.0  16.194826  4.282422  6.653188  13.106233  15.698087  18.806426

             max
Party
0      37.622759
1      31.642106
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [33]:
```
#Task 8
# The age group 65 and over has not much effect on the voting. The difference between the medians for both pa
rties is not significant.
```

In [34]:

```
# Task 8

data_merged["Percent Age between 30 and 64 inclusive"] = 100 - data_merged["Percent Age 29 and Under"] - data
_merged["Percent Age 65 and Older"]
data_merged.head()
```

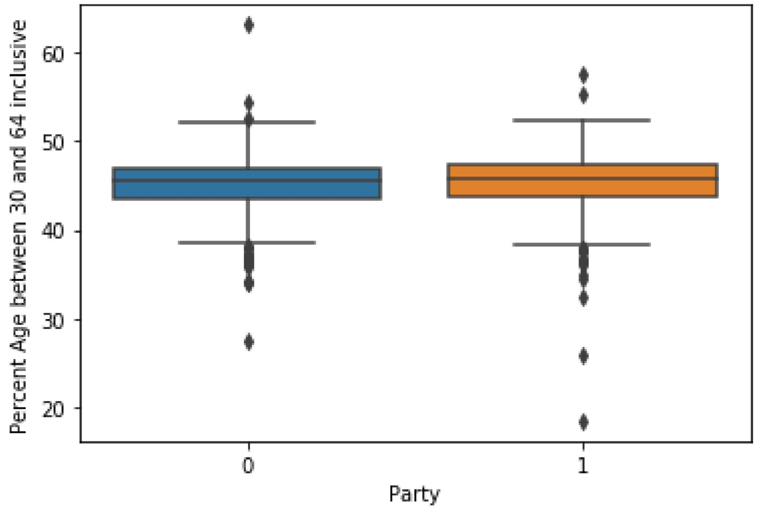
Out[34]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Percent Female	Percent Age 29 and Under	ar
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	50.598513	45.854643	13
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	49.069646	37.902276	19
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	50.581614	48.946141	10
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	50.296170	32.238290	26
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	46.313518	46.393456	12

5 rows × 21 columns

In [35]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Age between 30 and 64 inclusive"])
ageBET30and64 = sns.boxplot(x ="Party", y = "Percent Age between 30 and 64 inclusive", data = data_merged)
print(ageBET30and64)
```

```
       count       mean       std        min        25%        50%        75%  \
Party
0      875.0  45.164018  2.909028  27.421759  43.506682  45.554480  46.980099
1      325.0  45.078214  3.907598  18.433769  43.741937  45.817819  47.448269

             max
Party
0      63.157895
1      57.478906
AxesSubplot(0.125,0.125;0.775x0.755)
```

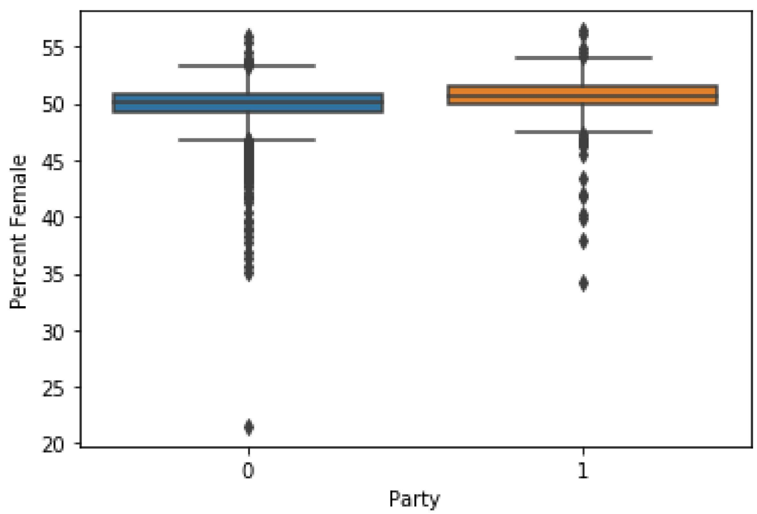


In [36]:
```
#Task 8
# The age group Age between 30 and 64 inclusive has almost no effect on the voting. The difference between th
e medians for both parties is almost zero
```

In [37]:
```
# Task 8
# Plot histogram
#heights = data_merged[["Party","Percent Age 65 and Older"]]
#plt.title('Height histogram')
#plt.xlabel('Party')
#plt.ylabel('Percent Age 65 and Older')
#plt.hist(heights['Party'], bins = 10)
```

In [38]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Female"])
genderF = sns.boxplot(x ="Party", y = "Percent Female", data = data_merged)
print(genderF)
```

```
       count       mean       std        min        25%        50%        75%  \
Party
0      875.0  49.617156  2.447883  21.513413  49.207916  50.174456  50.827181
1      325.0  50.385433  2.149359  34.245291  49.854280  50.653830  51.492075

             max
Party
0      55.885023
1      56.418468
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [39]:
```
data_merged["Percent Male"] = 100 - data_merged["Percent Female"]
data_merged.head()
```

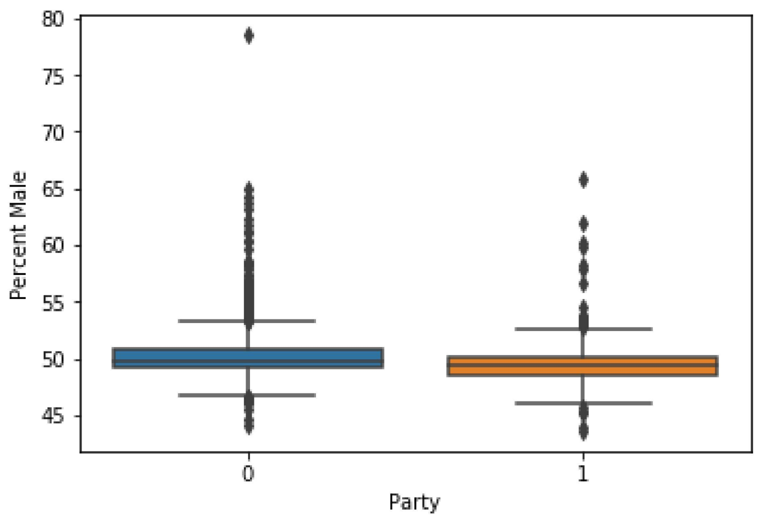
Out[39]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Percent Age 29 and Under	Percent Age 65 and Older	Hc
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	45.854643	13.322091	
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	37.902276	19.756275	
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	48.946141	10.873943	
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	32.238290	26.397638	
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	46.393456	12.315809	

5 rows × 22 columns

In [40]:
```python
print(data_merged.groupby("Party").describe()["Percent Male"])
genderM = sns.boxplot(x ="Party", y = "Percent Male", data = data_merged)
print(genderM)
```

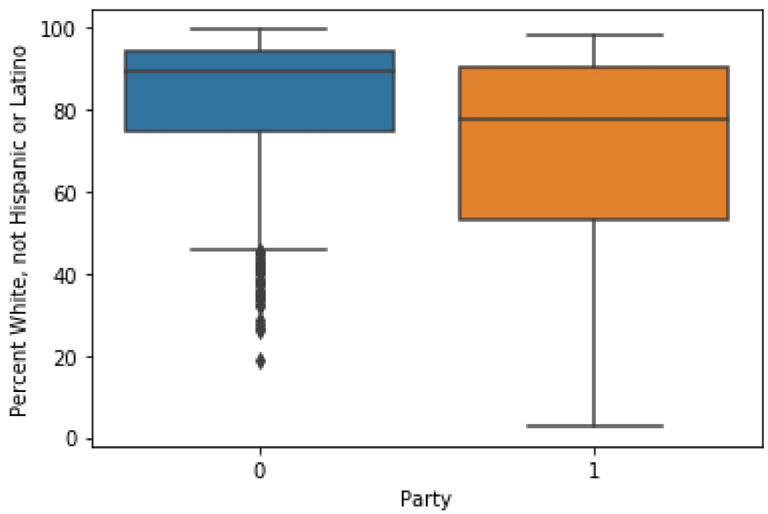
```
       count       mean       std        min        25%        50%        75%  \
Party
0      875.0  50.382844  2.447883  44.114977  49.172819  49.825544  50.792084
1      325.0  49.614567  2.149359  43.581532  48.507925  49.346170  50.145720

             max
Party
0      78.486587
1      65.754709
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [41]:
```python
#Task 8
# The gender has almost no effect on the voting. The difference between the medians for both parties is almos
t zero
```

In [42]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent White, not Hispanic or Latino"])
percentwhite = sns.boxplot(x ="Party", y = "Percent White, not Hispanic or Latino", data = data_merged)
print(percentwhite)
```

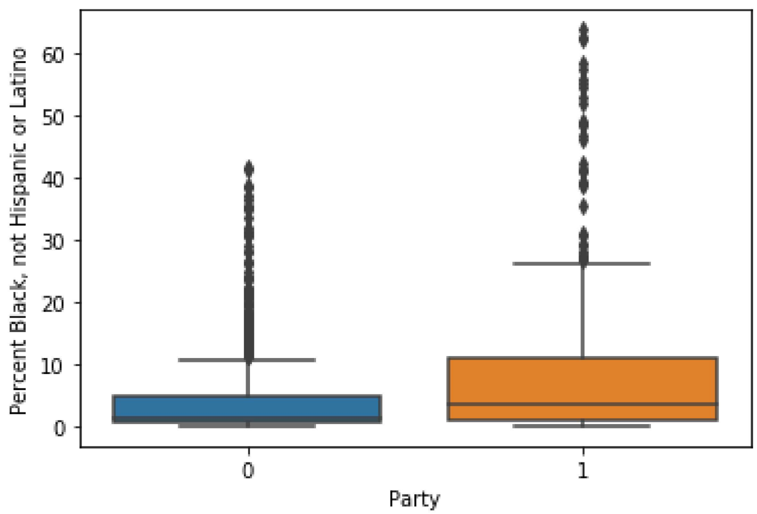
```
       count       mean        std        min        25%        50%  \
Party
0      875.0  82.597026  16.134097  18.758977  74.960538  89.418396
1      325.0  69.683766  24.981502   2.776702  53.271579  77.786090

             75%        max
Party
0      94.468872  99.627329
1      90.300749  98.063495
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [43]:
```
#Task 8
# The attribute Percent White, not Hispanic or Latino has significant effect on the voting. The difference be
tween the medians for both parties is significant.
```

In [44]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Black, not Hispanic or Latino"])
percentblack = sns.boxplot(x ="Party", y = "Percent Black, not Hispanic or Latino", data = data_merged)
print(percentblack)
```

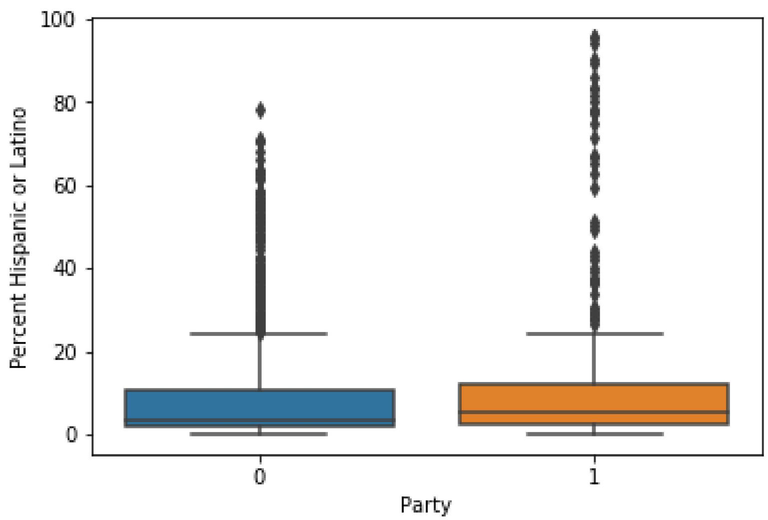
```
       count      mean        std  min       25%       50%        75%  \
Party
0      875.0  4.182092   6.706383  0.0  0.460803  1.318775   4.750447
1      325.0  9.242649  13.351340  0.0  0.839103  3.485992  11.058843

             max
Party
0      41.563041
1      63.953279
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [45]:
```
#Task 8
# The attribute Percent Black, not Hispanic or Latino has very little effect on the voting. The difference be
tween the medians for both parties is not that much.
```

In [46]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Hispanic or Latino"])
percentHL = sns.boxplot(x ="Party", y = "Percent Hispanic or Latino", data = data_merged)
print(percentHL)
```

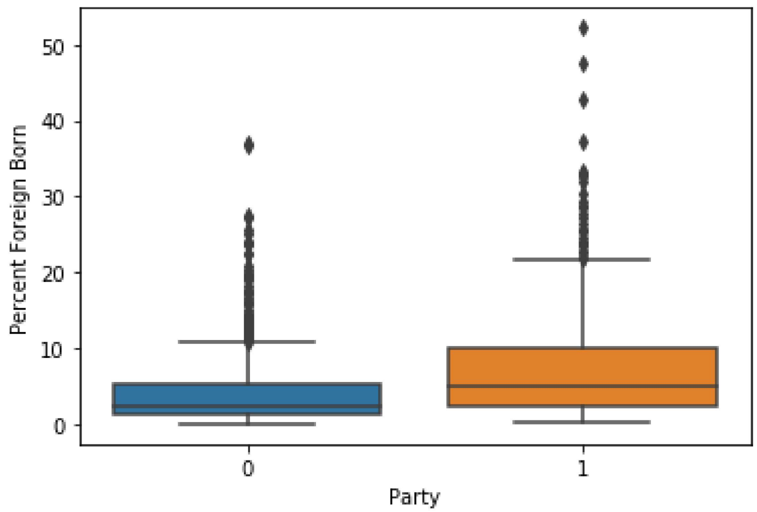
```
       count       mean        std       min       25%       50%        75%  \
Party
0      875.0   9.801825  14.144003  0.000000  1.704640  3.440794  10.785963
1      325.0  12.587391  19.575030  0.193349  2.531017  5.039747  11.857116

             max
Party
0      78.397012
1      95.479801
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [47]:
```
#Task 8
# The attribute Percent Hispanic or Latino has very little effect on the voting. The difference between the m
edians for both parties is not that much.
```

In [48]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Foreign Born"])
percentFB = sns.boxplot(x ="Party", y = "Percent Foreign Born", data = data_merged)
print(percentFB)
```

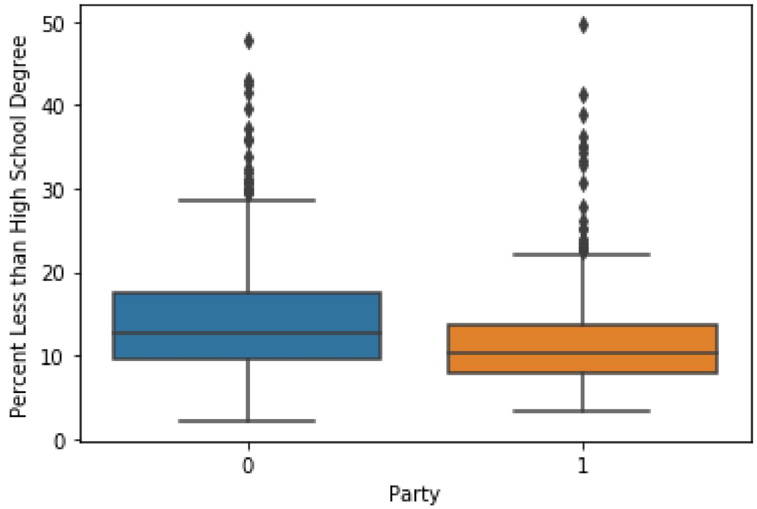
```
       count      mean       std       min       25%       50%        75%  \
Party
0      875.0  3.989607  4.497946  0.000000  1.320845  2.326782   5.139964
1      325.0  7.986330  8.330740  0.179769  2.470508  5.105490  10.144555

             max
Party
0      37.058317
1      52.229868
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [49]:
```
#Task 8
# The attribute Percent Foreign Born has very little effect on the voting. The difference between the medians
for both parties is not that much.
```

In [50]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Less than High School Degree"])
percentLHS = sns.boxplot(x ="Party", y = "Percent Less than High School Degree", data = data_merged)
print(percentLHS)
```

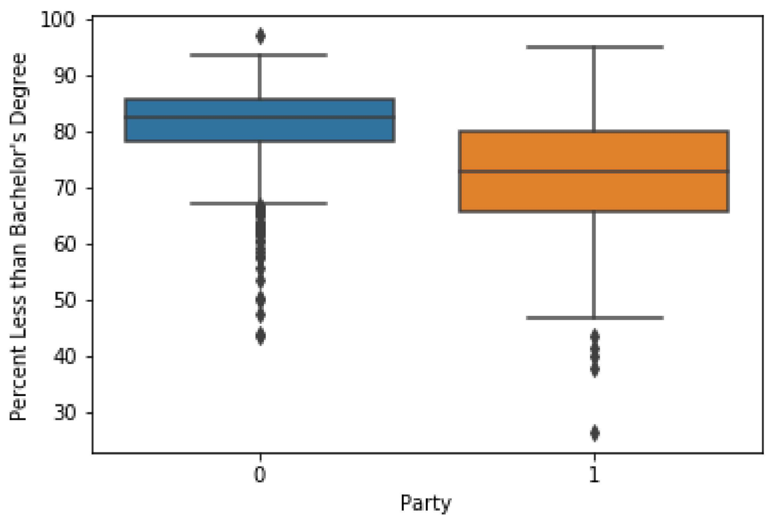
```
       count       mean       std       min       25%        50%        75%  \
Party
0      875.0  14.029195  6.319875  2.134454  9.666957  12.577108  17.489907
1      325.0  11.883760  6.505613  3.215803  7.893714  10.370080  13.637059

             max
Party
0      47.812773
1      49.673777
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [51]:
```
#Task 8
# The attribute Percent Less than High School Degree has very little effect on the voting. The difference bet
ween the medians for both parties is not that much.
```

In [52]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent Less than Bachelor's Degree"])
percentLBD = sns.boxplot(x ="Party", y = "Percent Less than Bachelor's Degree", data = data_merged)
print(percentLBD)
```

```
       count       mean        std       min        25%        50%        75%  \
Party
0      875.0  81.103128   6.842667  43.41947  78.108767  82.409455  85.561291
1      325.0  71.968225  11.192404  26.33544  65.711800  72.736143  79.903653

             max
Party
0      97.014925
1      94.849957
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [53]:
```
#Task 8
# The attribute Percent Less than Bachelor's Degree has significant effect on the voting. The difference betw
een the medians for both parties is significant.
```

In [54]:
```python
data_merged["Percent more than High School Degree"] = 100 - data_merged["Percent Less than High School Degre
e"]
data_merged.head()
```

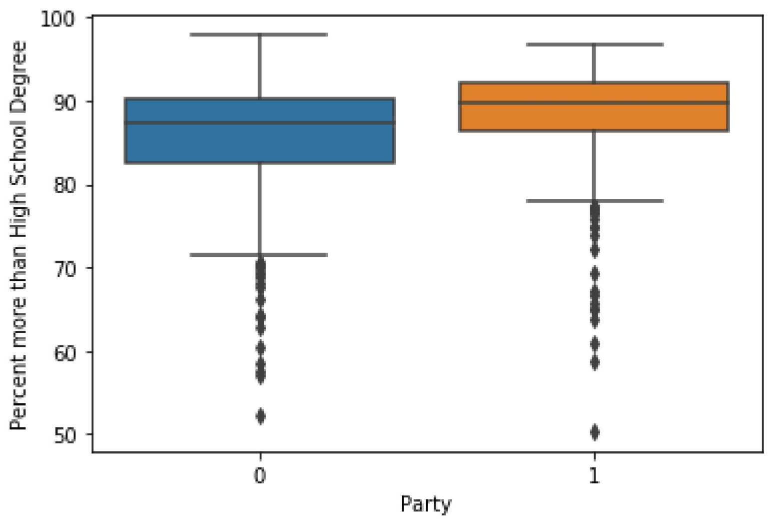
Out[54]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Percent Age 65 and Older	Median Household Income	U
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	13.322091	32460	
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	19.756275	45383	
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	10.873943	51106	
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	26.397638	40593	
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	12.315809	47422	

5 rows × 23 columns

In [55]:
```
print(data_merged.groupby("Party").describe()["Percent more than High School Degree"])
percentMBD = sns.boxplot(x ="Party", y = "Percent more than High School Degree", data = data_merged)
print(percentMBD)
```

```
       count       mean       std        min        25%        50%        75%  \
Party
0      875.0  85.970805  6.319875  52.187227  82.510093  87.422892  90.333043
1      325.0  88.116240  6.505613  50.326223  86.362941  89.629920  92.106286

             max
Party
0      97.865546
1      96.784197
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [56]:
```
data_merged["Percent more than Bachelor's Degree"] = 100 - data_merged["Percent Less than Bachelor's Degree"]
data_merged.head()
```

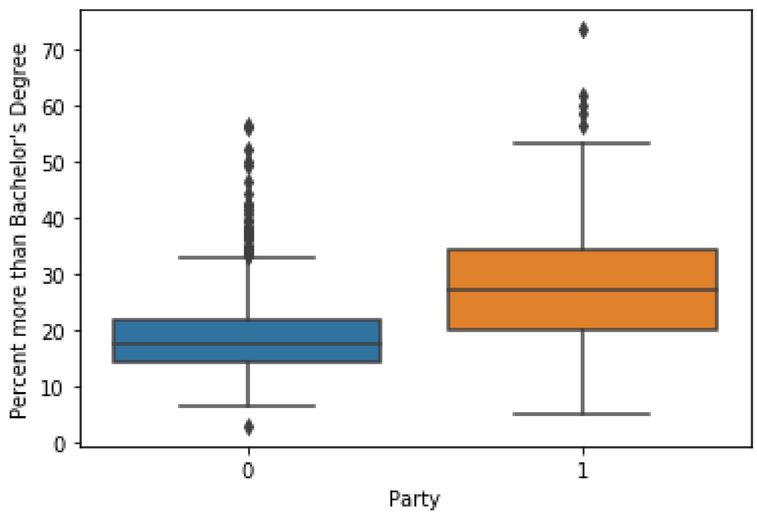
Out[56]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Median Household Income	Percent Unemployed
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	32460	15.807433
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	45383	8.567108
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	51106	8.238305
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	40593	12.129932
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	47422	14.424104

5 rows × 24 columns

In [57]:
```
# TASK 8
print(data_merged.groupby("Party").describe()["Percent more than Bachelor's Degree"])
percentMBD = sns.boxplot(x ="Party", y = "Percent more than Bachelor's Degree", data = data_merged)
print(percentMBD)
```

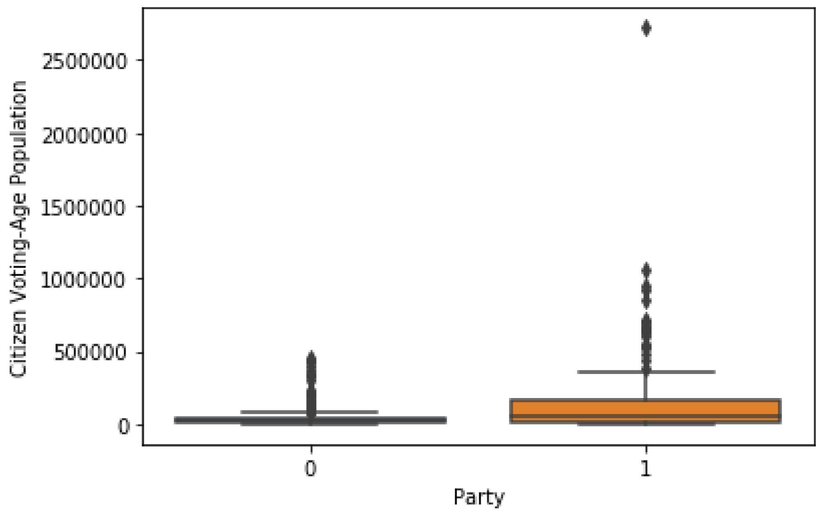
```
       count       mean        std       min        25%        50%        75%  \
Party
0      875.0  18.896872   6.842667  2.985075  14.438709  17.590545  21.891233
1      325.0  28.031775  11.192404  5.150043  20.096347  27.263857  34.288200

            max
Party
0      56.58053
1      73.66456
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [58]:
```
#Task 8
# The attribute Percent more than Bachelor's Degree has significant effect on the voting. The difference betw
een the medians for both parties is significant.
```

In [59]:
```
# Task 8
print(data_merged.groupby("Party").describe()["Citizen Voting-Age Population"])
percentMBD = sns.boxplot(x ="Party", y = "Citizen Voting-Age Population", data = data_merged)
print(percentMBD)
```

```
       count           mean            std     min      25%      50%  \
Party
0      386.0   39270.012953   64700.902510    60.0   8212.5  19780.0
1      134.0  175827.425373  319459.705034  1575.0  18617.5  54337.5

             75%        max
Party
0       38796.25   460215.0
1      163812.50  2723565.0
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [60]:
```
# Task 8
data_merged["Citizen Voting-Age Population voted"] = data_merged["Democratic"] + data_merged["Republican"]
data_merged.head()
```

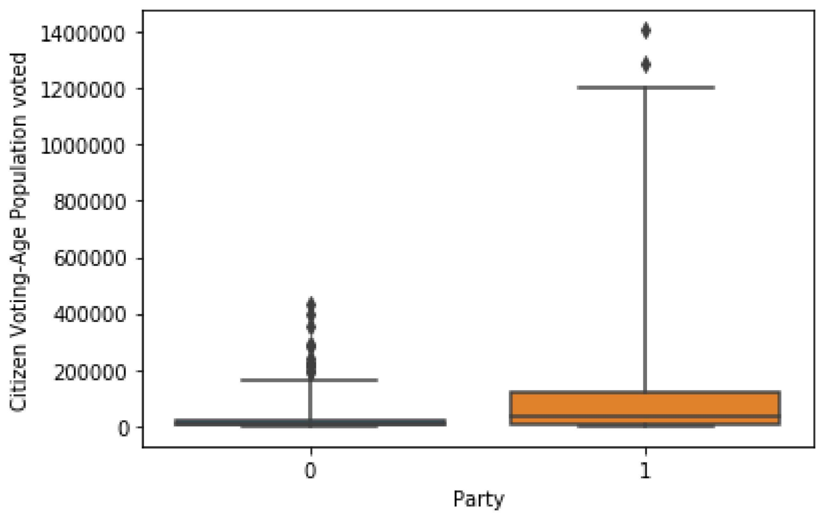
Out[60]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Percent Unemployed	Percent Less than High School Degree
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	15.807433	21.758252
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	8.567108	13.409171
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	8.238305	11.085381
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	12.129932	15.729958
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	14.424104	14.580797

5 rows × 25 columns

In [61]:
```
# Task 8
print(data_merged.groupby("Party").describe()["Citizen Voting-Age Population voted"])
percentMBD = sns.boxplot(x ="Party", y = "Citizen Voting-Age Population voted", data = data_merged, whis = 10
)
print(percentMBD)
```

```
       count          mean            std    min     25%      50%       75%  \
Party
0      875.0   20453.404571   39598.706398    0.0  3581.0   8850.0   19010.5
1      325.0  112516.033846  195429.662071  818.0  8899.0  32443.0  123590.0

             max
Party
0       435180.0
1      1405176.0
AxesSubplot(0.125,0.125;0.775x0.755)
```



In [62]:
```
#Task 9

# The variables "Percent White, not Hispanic or Latino", "Percent Less than Bachelor's Degree" and "Percent more than Bachelor's Degree"
# are more important to determine whether a county is labeled as Democratic or Republican because of the folowing reasons:
# 1) In the plot of "Percent White, not Hispanic or Latino", nearly 656 democratic counties have White (not Hispanic or Latino population) more than 75 percent.
   # And the median of the White (not Hispanic or Latino population) percent for the democratic counties is greater by 12 percent.
# 2) In the plot of "Percent Less than Bachelor's Degree", nearly 656 democratic counties have more than 78 percent of people with less than a bachelor's degree.
   # And the median of the Percent Less than Bachelor's Degree for the democratic counties is greater by 10 percent.
# 3) In the plot of "Percent More than Bachelor's Degree", nearly 244 republican counties have more than 27 percent of people with more than bachelor’s degrees.
   # And the median of the Percent More than Bachelor's Degree for the republican counties is greater by 10 percent.
```

In [63]:
```
change_values = pd.Series({1 : "Demo", 0 : "Rep"})
data_merged["Party"] = data_merged["Party"].map(change_values)
data_merged.head()
```

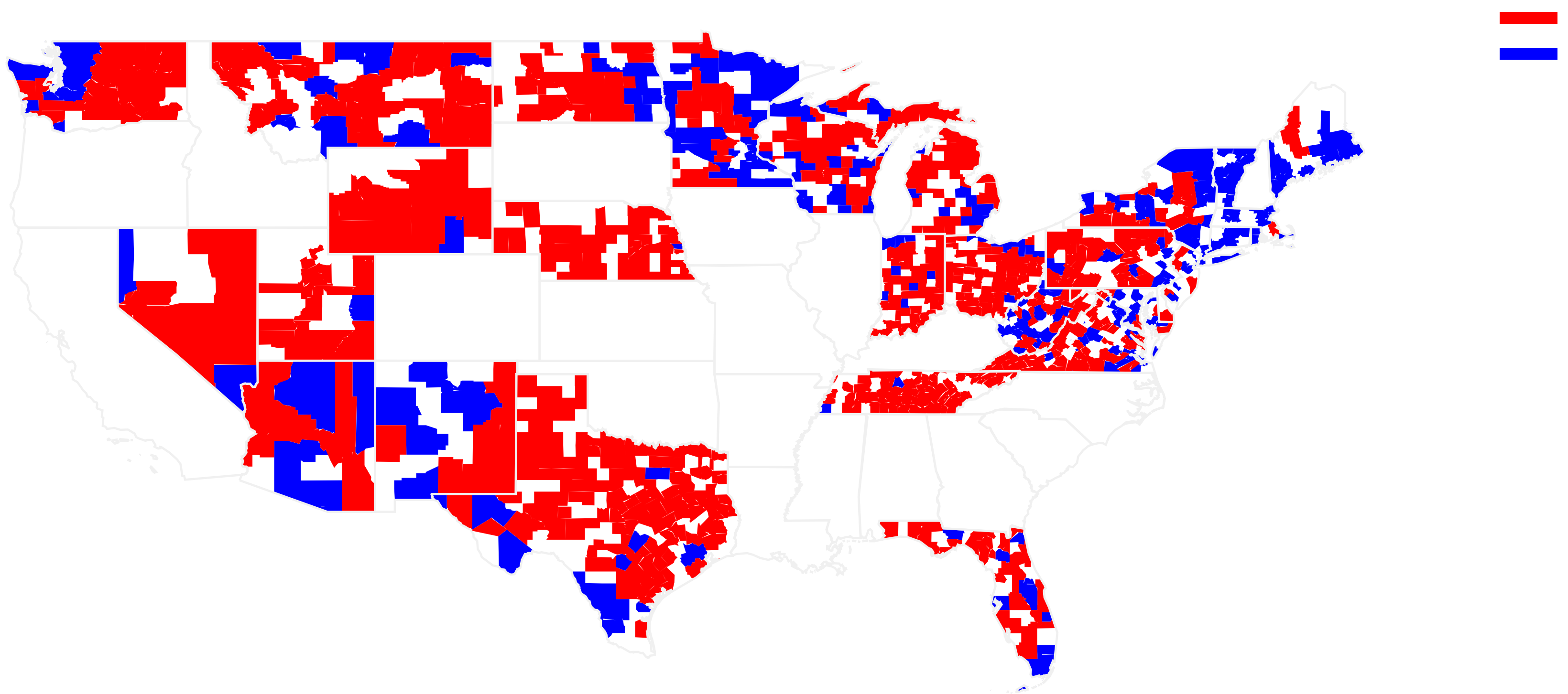
Out[63]:

	State	County	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	...	Percent Unemployed	Percent Less than High School Degree
0	AZ	apache	16298.0	7810.0	4001	72346	NaN	18.571863	0.486551	5.947806	...	15.807433	21.758252
1	AZ	cochise	17383.0	26929.0	4003	128177	92915.0	56.299492	3.714395	34.403208	...	8.567108	13.409171
2	AZ	coconino	34240.0	19249.0	4005	138064	104265.0	54.619597	1.342855	13.711033	...	8.238305	11.085381
3	AZ	gila	7643.0	12180.0	4007	53179	NaN	63.222325	0.552850	18.548675	...	12.129932	15.729958
4	AZ	graham	3368.0	6870.0	4009	37529	NaN	51.461536	1.811932	32.097844	...	14.424104	14.580797

5 rows × 25 columns

In [64]:
```
import plotly as py
import plotly.graph_objs as go
import plotly.figure_factory as ff

fips =data_merged['FIPS'].astype(str).tolist()
values = data_merged['Party'].tolist()
colorscale = ['rgb(0,0,255)', 'rgb(255,0,0)'
             ]
fig = ff.create_choropleth(fips=fips, values=values, colorscale = colorscale)
fig.layout.template = None
fig.show()
```

```
C:\Users\kalya\Anaconda3\lib\site-packages\pandas\core\frame.py:6692: FutureWarning:

Sorting because non-concatenation axis is not aligned. A future version
of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.
```



In [ ]: