In [1]:
```
import pandas as pd
import numpy as np
import scipy.stats as st
from scipy.stats import ttest_ind
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.figure_factory as ff
#rendering purposes for pdf
from IPython.display import Image
```

In [2]:
```
election_data = pd.read_csv("election_train.csv")
demo_data = pd.read_csv("demographics_train.csv")
```

In [3]:
```
# Task 1
#Reshape (pivot) dataset from long format to wide format)
election_data_r = pd.pivot_table(election_data, values = 'Votes',index = ['Year', 'State','County','Office'], columns ='Party',aggfunc=np.sum).reset_in
dex()
election_data_r.head()
```

Out[3]:

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

In [4]:
```
#Convert to lower case for handling inconsistencies
demo_data['County']=demo_data['County'].str.lower()
election_data_r['County']=election_data_r['County'].str.lower()
```

In [5]:
```
#Stripping extra spaces from column 'County'
election_data_r['County']=election_data_r['County'].str.replace(' county','')
```

In [6]:
```
#Change States to Abbreviation in Demo_data
change_values = {'Alabama' : 'AL' ,'Alaska' : 'AK','Arizona' : 'AZ' ,
                 'Arkansas' : 'AR' ,'California' : 'CA' ,'Colorado' : 'CO' ,
                 'Connecticut' : 'CT' ,'Delaware' : 'DE' ,'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' ,'Ohio' : 'OH' ,'Oklahoma' : 'OK' ,'Oregon' : 'OR' ,'Pennsylvania' : 'PA'
                 ,'Rhode Island' : 'RI' ,'South Carolina' : 'SC' ,'South Dakota' : 'SD' ,'Tennessee' : 'TN' ,
                 'Texas' : 'TX' ,'Utah' : 'UT' ,'Vermont' : 'VT' ,'Virginia' : 'VA' ,'Washington' : 'WA' ,
                 'West Virginia' : 'WV' ,
                 'Wisconsin' : 'WI' ,'Wyoming' : 'WY' }

demo_data['State'] = demo_data['State'].map(change_values)
```

In [7]:
```python
election_data_r.head()
```

Out[7]:

Party	Year	State	County	Office	Democratic	Republican
0	2018	AZ	apache	US Senator	16298.0	7810.0
1	2018	AZ	cochise	US Senator	17383.0	26929.0
2	2018	AZ	coconino	US Senator	34240.0	19249.0
3	2018	AZ	gila	US Senator	7643.0	12180.0
4	2018	AZ	graham	US Senator	3368.0	6870.0

In [8]:
```python
demo_data.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	Median Household Income	Percent Unemployed	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Rural
0	WI	la crosse	55063	117538	0	90.537528	1.214075	1.724549	2.976059	51.171536	43.241335	14.702479	51477	4.796952	5.474767	67.529757	16.827753
1	VA	alleghany	51005	15919	12705	91.940449	5.207614	1.432251	1.300333	51.077329	31.660280	23.902255	45538	4.560986	15.537543	83.711604	52.393846
2	IN	fountain	18045	16741	12750	95.705155	0.400215	2.359477	1.547100	49.770026	35.899887	18.941521	45924	7.978789	12.032155	85.538940	65.951276
3	OH	geauga	39055	94020	0	95.837056	1.256116	1.294405	2.578175	50.678579	36.281642	18.028079	74165	4.036902	8.928599	62.730824	63.968990
4	WI	jackson	55053	20566	15835	86.662453	1.983857	3.082758	1.376058	46.649810	36.292911	17.587280	49608	5.569698	11.792912	86.129256	72.238251

In [9]:
```python
#Task 2
merge_data=pd.merge(election_data_r,demo_data,how='inner', left_on=['State','County'],right_on=['State','County'])
```

In [10]:
```
merge_data
```

Out[10]:

	Year	State	County	Office	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	...	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Percent Age 29 and Under	Percent Age 65 and Older	Median Household Income	Percent Unemployed	Percent Less than High School Degree
0	2018	AZ	apache	US Senator	16298.0	7810.0	4001	72346	0	18.571863	...	5.947806	1.719515	50.598513	45.854643	13.322091	32460	15.807433	21.758252
1	2018	AZ	cochise	US Senator	17383.0	26929.0	4003	128177	92915	56.299492	...	34.403208	11.458374	49.069646	37.902276	19.756275	45383	8.567108	13.409171
2	2018	AZ	coconino	US Senator	34240.0	19249.0	4005	138064	104265	54.619597	...	13.711033	4.825298	50.581614	48.946141	10.873943	51106	8.238305	11.085381
3	2018	AZ	gila	US Senator	7643.0	12180.0	4007	53179	0	63.222325	...	18.548675	4.249798	50.296170	32.238290	26.397638	40593	12.129932	15.729958
4	2018	AZ	graham	US Senator	3368.0	6870.0	4009	37529	0	51.461536	...	32.097844	4.385942	46.313518	46.393456	12.315809	47422	14.424104	14.580797
5	2018	AZ	la paz	US Senator	1609.0	3265.0	4012	20304	15245	58.884949	...	26.182033	11.372143	48.946020	28.073286	36.056935	36321	10.599013	24.842215
6	2018	AZ	maricopa	US Senator	732671.0	672505.0	4013	4088549	2723565	56.918114	...	30.286833	14.729333	50.549278	41.886620	13.837843	55676	6.808454	13.051927
7	2018	AZ	mohave	US Senator	19214.0	50209.0	4015	203629	0	78.252606	...	15.708470	6.969047	49.676618	30.485835	26.858650	39856	11.680953	16.145850
8	2018	AZ	navajo	US Senator	16624.0	18767.0	4017	108209	76280	41.927196	...	11.049913	2.914730	49.846131	43.243168	15.745456	36868	18.525791	18.494087
9	2018	AZ	pima	US Senator	221242.0	160550.0	4019	1003338	0	53.271579	...	36.105978	12.903428	50.807405	40.087388	17.801778	46764	9.214114	12.252238
10	2018	AZ	santa cruz	US Senator	9241.0	3828.0	4023	46547	27155	15.274883	...	83.219112	32.644424	52.125808	43.300320	15.895761	38941	9.749896	25.206726
11	2018	AZ	yavapai	US Senator	40160.0	65308.0	4025	218586	0	81.159361	...	14.054880	6.456955	51.092476	28.717301	28.272625	46638	8.525986	9.830672
12	2018	CT	fairfield	US Senator	210899.0	131321.0	9001	941618	0	63.509512	...	18.636007	21.154120	51.308280	38.019876	14.383115	86670	8.211898	10.103521
13	2018	CT	hartford	US Senator	203591.0	123864.0	9003	895699	644940	63.252052	...	16.897976	15.011070	51.492075	37.389458	15.644876	68027	8.237243	10.736314
14	2018	CT	middlesex	US Senator	42383.0	32836.0	9007	164438	0	84.738929	...	5.641032	7.652732	51.098894	33.241708	17.703937	79837	5.257635	6.153444
15	2018	CT	new haven	US Senator	179714.0	126004.0	9009	860874	631715	64.782767	...	16.790959	12.073428	51.759956	37.982562	15.604490	62715	8.545483	10.503197
16	2018	CT	tolland	US Senator	34732.0	28046.0	9013	151689	0	85.892187	...	5.029369	6.769113	49.967367	42.301024	13.868507	80129	6.267325	6.214203
17	2018	CT	windham	US Senator	20490.0	19032.0	9015	117078	0	83.782606	...	10.877364	5.269991	50.422795	37.389604	14.571482	60689	8.375524	11.925239
18	2018	DE	sussex	US Senator	40675.0	50391.0	10005	211224	0	74.829091	...	9.195925	6.671117	51.495569	31.746866	24.214578	54218	7.108621	13.792481
19	2018	FL	alachua	US Senator	74493.0	40599.0	12001	256581	197720	62.460198	...	8.994820	10.144555	51.670623	48.551530	12.430772	44702	7.020154	7.468059
20	2018	FL	baker	US Senator	1945.0	8579.0	12003	27312	20415	82.088459	...	2.317663	1.475542	47.715290	41.344464	12.910076	53327	6.832522	17.871127
21	2018	FL	bay	US Senator	16723.0	46681.0	12005	178361	135795	77.687387	...	5.728270	5.479898	50.280050	37.739192	16.103857	48577	7.475668	11.320101
22	2018	FL	bradford	US Senator	2879.0	7576.0	12007	26919	0	74.787325	...	3.796575	2.306921	44.741632	35.343066	17.452357	43373	10.563312	23.212996

	Year	State	County	Office	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	...	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Percent Age 29 and Under	Percent Age 65 and Older	Median Household Income	Percent Unemployed	Percent Less than High School Degree
23	2018	FL	brevard	US Senator	121112.0	160305.0	12009	560683	438510	75.798803	...	9.301691	8.632685	51.140841	31.972612	22.491140	49914	9.278023	8.823708
24	2018	FL	broward	US Senator	472239.0	211397.0	12011	1863780	0	39.245351	...	27.564841	32.715718	51.376504	36.701971	15.371879	52954	8.710312	11.688725
25	2018	FL	charlotte	US Senator	33525.0	52916.0	12015	169642	141230	84.643543	...	6.632202	10.185567	51.260891	22.388913	37.622759	44865	10.088979	10.578890
26	2018	FL	citrus	US Senator	22660.0	48008.0	12017	140453	0	88.871010	...	5.121998	5.413911	51.667818	24.581889	35.105694	39054	11.648012	13.288908
27	2018	FL	collier	US Senator	54390.0	101266.0	12021	348236	0	64.222539	...	26.634524	23.816894	50.881299	29.998909	29.589704	59783	6.408868	14.321540
28	2018	FL	desoto	US Senator	3328.0	5503.0	12027	35134	0	55.040701	...	30.688222	17.504412	43.345477	37.513520	19.408550	35513	8.291825	29.475750
29	2018	FL	dixie	US Senator	1322.0	4442.0	12029	16084	12890	85.041035	...	3.780154	2.785377	45.417807	31.236011	21.729669	34634	5.306200	22.236087
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1170	2018	WV	pocahontas	US Senator	1269.0	1411.0	54075	8620	7070	95.696056	...	0.893271	0.881671	48.364269	29.408353	22.331787	36026	5.685358	15.241010
1171	2018	WV	preston	US Senator	3686.0	5943.0	54077	33793	27010	92.365283	...	1.831740	1.003166	48.477495	33.465511	17.429645	45221	7.707416	16.522582
1172	2018	WV	raleigh	US Senator	10581.0	12620.0	54081	78051	0	87.123804	...	1.475958	1.680952	49.720055	35.009161	18.099704	41533	7.709258	15.658951
1173	2018	WV	randolph	US Senator	4472.0	4017.0	54083	29287	23555	96.182607	...	0.816062	0.669239	48.714447	34.011678	19.814252	40308	7.425664	17.197844
1174	2018	WV	ritchie	US Senator	1082.0	1961.0	54085	10044	0	97.779769	...	0.756671	0.169255	50.398248	32.487057	20.141378	40850	8.428745	18.066123
1175	2018	WV	roane	US Senator	2165.0	1899.0	54087	14513	0	97.050920	...	0.957762	0.496107	50.768277	33.404534	19.506649	34144	10.515990	21.849308
1176	2018	WV	summers	US Senator	2069.0	1868.0	54089	13325	10995	91.639775	...	1.696060	0.435272	53.696060	28.900563	21.275797	35620	10.729049	17.191498
1177	2018	WV	taylor	US Senator	2376.0	2642.0	54091	16949	13450	96.300667	...	0.967609	0.749307	49.247743	33.482801	18.065963	44371	9.082581	14.098280
1178	2018	WV	tucker	US Senator	1469.0	1502.0	54093	6922	5690	97.688529	...	0.650101	0.592314	50.245594	29.731292	22.594626	43529	8.190184	12.702390
1179	2018	WV	tyler	US Senator	1065.0	1603.0	54095	9000	7205	98.000000	...	0.711111	0.233333	50.211111	31.833333	20.133333	38674	8.458542	12.429292
1180	2018	WV	upshur	US Senator	3102.0	4010.0	54097	24632	0	96.354336	...	1.181390	0.633323	49.837610	37.950633	18.281098	42240	7.692308	16.125150
1181	2018	WV	wayne	US Senator	6395.0	5954.0	54099	41237	32220	97.524068	...	0.613527	0.603827	51.378616	34.481170	18.289400	38311	9.504391	20.628495
1182	2018	WV	wetzel	US Senator	2518.0	2135.0	54103	15997	0	97.574545	...	0.481340	0.443833	50.997062	33.168719	21.278990	39446	8.472337	16.947084
1183	2018	WV	wood	US Senator	14189.0	13696.0	54107	86262	67640	95.478890	...	1.032900	1.001600	51.756277	34.726763	18.559737	43944	7.625458	10.370080
1184	2018	WV	wyoming	US Senator	2607.0	3096.0	54109	22537	17730	97.475263	...	0.257355	0.257355	50.547988	33.877623	17.846208	35469	12.077958	22.638690
1185	2018	WY	albany	US Senator	7576.0	6366.0	56001	37836	30070	83.269902	...	9.229305	6.081510	47.756105	55.447193	9.549107	43043	4.579174	4.210167

	Year	State	County	Office	Democratic	Republican	FIPS	Total Population	Citizen Voting-Age Population	Percent White, not Hispanic or Latino	...	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Percent Age 29 and Under	Percent Age 65 and Older	Median Household Income	Percent Unemployed	Percent Less than High School Degree
1186	2018	WY	campbell	US Senator	1628.0	11020.0	56005	48473	0	87.774637	...	8.357230	3.645328	48.008170	45.326264	6.954387	80822	4.669432	8.330054
1187	2018	WY	carbon	US Senator	1359.0	3673.0	56007	15696	11335	77.924312	...	17.647808	5.198777	45.839704	39.181957	13.774210	56972	4.141937	9.579879
1188	2018	WY	converse	US Senator	834.0	3959.0	56009	14223	0	88.849047	...	7.691767	2.706883	49.933207	38.515081	13.668003	66737	5.282284	9.758393
1189	2018	WY	fremont	US Senator	4734.0	9262.0	56013	40683	30170	70.198855	...	6.779244	1.339626	49.907824	39.751247	16.409803	53559	7.344324	8.537172
1190	2018	WY	goshen	US Senator	1020.0	3658.0	56015	13546	0	86.409272	...	10.519711	2.724051	47.091392	35.914661	20.389783	44883	6.918819	8.390574
1191	2018	WY	johnson	US Senator	722.0	3085.0	56019	8572	6590	91.565562	...	2.134858	1.656556	46.966869	32.571162	20.496967	54594	4.512276	5.105750
1192	2018	WY	lincoln	US Senator	1152.0	5846.0	56023	18543	0	92.600982	...	4.416761	2.151755	48.773122	38.715418	14.382786	64579	5.618095	6.949996
1193	2018	WY	natrona	US Senator	7285.0	16359.0	56025	80871	60415	87.026252	...	8.198242	2.729038	49.421919	40.688257	12.825364	56983	4.861868	8.332255
1194	2018	WY	niobrara	US Senator	144.0	980.0	56027	2498	1995	89.511609	...	4.243395	0.280224	54.443555	36.269015	18.254604	40640	0.457875	11.054422
1195	2018	WY	platte	US Senator	801.0	2850.0	56031	8740	6830	89.359268	...	7.814645	2.780320	47.711670	32.700229	22.013730	41051	3.901047	9.675889
1196	2018	WY	sublette	US Senator	668.0	2653.0	56035	10032	0	91.646730	...	7.814992	2.053429	46.949761	36.393541	13.337321	76004	2.786971	4.658830
1197	2018	WY	sweetwater	US Senator	3943.0	8577.0	56037	44812	30565	79.815674	...	15.859591	5.509685	47.824244	44.153352	9.417120	68233	5.072255	9.314606
1198	2018	WY	uinta	US Senator	1371.0	4713.0	56041	20893	14355	87.718375	...	8.959939	3.986981	49.327526	43.205858	10.678218	53323	6.390755	10.361224
1199	2018	WY	washakie	US Senator	588.0	2423.0	56043	8351	0	82.397318	...	13.962400	3.783978	51.359119	34.774279	19.650341	46212	7.441860	12.577108

1200 rows × 21 columns

In [11]:
```
merge_data.shape
```

Out[11]:
```
(1200, 21)
```

In [12]:
```
merge_data.dtypes
```

Out[12]:
```
Year                                       int64
State                                     object
County                                    object
Office                                    object
Democratic                               float64
Republican                               float64
FIPS                                       int64
Total Population                           int64
Citizen Voting-Age Population              int64
Percent White, not Hispanic or Latino    float64
Percent Black, not Hispanic or Latino    float64
Percent Hispanic or Latino               float64
Percent Foreign Born                     float64
Percent Female                           float64
Percent Age 29 and Under                 float64
Percent Age 65 and Older                 float64
Median Household Income                    int64
Percent Unemployed                       float64
Percent Less than High School Degree     float64
Percent Less than Bachelor's Degree      float64
Percent Rural                            float64
dtype: object
```

In [13]:
```
#Task 3
#There are 21 variable in the data set
#Type of these variables are int64, Object,float64
#The column 'Year' and 'Office' are irrelevant variable, since they don't provide any important information.
#The only information it gives is that data is for 2018 US Senator office
#The total citizen voting age population variable is also removed because it doesnt offer any valuable information to work
#with and also is missing values in a lot of counties making less valuable
#They are no redundant variables
#We can drop the irrelevant variables(Year and Office)
#merge_data=merge_data.drop(['Year','Office'],axis=1)
```

In [14]:
```
#Task 4
merge_data.info()
(merge_data==0).sum()
```

```
<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
```

Out[14]:
```
Year                                       0
State                                      0
County                                     0
Office                                     0
Democratic                                 5
Republican                                 5
FIPS                                       0
Total Population                           0
Citizen Voting-Age Population            680
Percent White, not Hispanic or Latino      0
Percent Black, not Hispanic or Latino     45
Percent Hispanic or Latino                 5
Percent Foreign Born                       3
Percent Female                             0
Percent Age 29 and Under                   0
Percent Age 65 and Older                   0
Median Household Income                    0
Percent Unemployed                         3
Percent Less than High School Degree       0
Percent Less than Bachelor's Degree        0
Percent Rural                             19
dtype: int64
```

In [15]:
```
#Yes, the merged data has missing values (missing values in every columns)
#Since the number is low compared to the size of the dataset, we can drop these values
merge_data=merge_data.drop(['Year','Office', 'Citizen Voting-Age Population'],axis=1)
#merge_data=merge_data.drop("Citizen Voting-Age Population", axis = 2)
```

In [16]:
```
merge_data.head(10)
```

Out[16]:

	State	County	Democratic	Republican	FIPS	Total 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	Median Household Income	Percent Unemployed	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Rural
0	AZ	apache	16298.0	7810.0	4001	72346	18.571863	0.486551	5.947806	1.719515	50.598513	45.854643	13.322091	32460	15.807433	21.758252	88.941063	74.061076
1	AZ	cochise	17383.0	26929.0	4003	128177	56.299492	3.714395	34.403208	11.458374	49.069646	37.902276	19.756275	45383	8.567108	13.409171	76.837055	36.301067
2	AZ	coconino	34240.0	19249.0	4005	138064	54.619597	1.342855	13.711033	4.825298	50.581614	48.946141	10.873943	51106	8.238305	11.085381	65.791439	31.466066
3	AZ	gila	7643.0	12180.0	4007	53179	63.222325	0.552850	18.548675	4.249798	50.296170	32.238290	26.397638	40593	12.129932	15.729958	82.262624	41.062000
4	AZ	graham	3368.0	6870.0	4009	37529	51.461536	1.811932	32.097844	4.385942	46.313518	46.393456	12.315809	47422	14.424104	14.580797	86.675944	46.437399
5	AZ	la paz	1609.0	3265.0	4012	20304	58.884949	0.379236	26.182033	11.372143	48.946020	28.073286	36.056935	36321	10.599013	24.842215	89.563407	56.327786
6	AZ	maricopa	732671.0	672505.0	4013	4088549	56.918114	5.013612	30.286833	14.729333	50.549278	41.886620	13.837843	55676	6.808454	13.051927	69.031137	2.363800
7	AZ	mohave	19214.0	50209.0	4015	203629	78.252606	0.951731	15.708470	6.969047	49.676618	30.485835	26.858650	39856	11.680953	16.145850	88.121178	22.963644
8	AZ	navajo	16624.0	18767.0	4017	108209	41.927196	0.672772	11.049913	2.914730	49.846131	43.243168	15.745456	36868	18.525791	18.494087	85.507970	54.138242
9	AZ	pima	221242.0	160550.0	4019	1003338	53.271579	3.199719	36.105978	12.903428	50.807405	40.087388	17.801778	46764	9.214114	12.252238	69.199391	7.523491

In [17]:
```
#Drop rows where Democratic and republican votes are zero
merge_data=merge_data[merge_data.Democratic!=0]
merge_data=merge_data[merge_data.Republican!=0]
```

In [18]:
```
merge_data.shape
```

Out[18]: (1195, 18)

In [19]:
```
#Task 5
merge_data['Party'] = np.where(merge_data['Democratic']>merge_data['Republican'], 1,0)
```

In [20]:
```python
#merge_data=merge_data.drop(['Party'],axis=1)
merge_data.head(15)
```

Out[20]:

	State	County	Democratic	Republican	FIPS	Total 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	Median Household Income	Percent Unemployed	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Rural	Party
	AZ	apache	16298.0	7810.0	4001	72346	18.571863	0.486551	5.947806	1.719515	50.598513	45.854643	13.322091	32460	15.807433	21.758252	88.941063	74.061076	1
	AZ	cochise	17383.0	26929.0	4003	128177	56.299492	3.714395	34.403208	11.458374	49.069646	37.902276	19.756275	45383	8.567108	13.409171	76.837055	36.301067	0
	AZ	coconino	34240.0	19249.0	4005	138064	54.619597	1.342855	13.711033	4.825298	50.581614	48.946141	10.873943	51106	8.238305	11.085381	65.791439	31.466066	1
	AZ	gila	7643.0	12180.0	4007	53179	63.222325	0.552850	18.548675	4.249798	50.296170	32.238290	26.397638	40593	12.129932	15.729958	82.262624	41.062000	0
	AZ	graham	3368.0	6870.0	4009	37529	51.461536	1.811932	32.097844	4.385942	46.313518	46.393456	12.315809	47422	14.424104	14.580797	86.675944	46.437399	0
	AZ	la paz	1609.0	3265.0	4012	20304	58.884949	0.379236	26.182033	11.372143	48.946020	28.073286	36.056935	36321	10.599013	24.842215	89.563407	56.327786	0
	AZ	maricopa	732671.0	672505.0	4013	4088549	56.918114	5.013612	30.286833	14.729333	50.549278	41.886620	13.837843	55676	6.808454	13.051927	69.031137	2.363800	1
	AZ	mohave	19214.0	50209.0	4015	203629	78.252606	0.951731	15.708470	6.969047	49.676618	30.485835	26.858650	39856	11.680953	16.145850	88.121178	22.963644	0
	AZ	navajo	16624.0	18767.0	4017	108209	41.927196	0.672772	11.049913	2.914730	49.846131	43.243168	15.745456	36868	18.525791	18.494087	85.507970	54.138242	0
	AZ	pima	221242.0	160550.0	4019	1003338	53.271579	3.199719	36.105978	12.903428	50.807405	40.087388	17.801778	46764	9.214114	12.252238	69.199391	7.523491	1
	AZ	santa cruz	9241.0	3828.0	4023	46547	15.274883	0.199798	83.219112	32.644424	52.125808	43.300320	15.895761	38941	9.749896	25.206726	77.506775	26.883172	1
	AZ	yavapai	40160.0	65308.0	4025	218586	81.159361	0.518331	14.054880	6.456955	51.092476	28.717301	28.272625	46638	8.525986	9.830672	74.458362	33.197178	0
	CT	fairfield	210899.0	131321.0	9001	941618	63.509512	10.456151	18.636007	21.154120	51.308280	38.019876	14.383115	86670	8.211898	10.103521	53.637538	4.584061	1
	CT	hartford	203591.0	123864.0	9003	895699	63.252052	12.657935	16.897976	15.011070	51.492075	37.389458	15.644876	68027	8.237243	10.736314	63.133913	5.408640	1
	CT	middlesex	42383.0	32836.0	9007	164438	84.738929	4.664980	5.641032	7.652732	51.098894	33.241708	17.703937	79837	5.257635	6.153444	59.071541	24.540066	1

In [21]:
```python
##*********************** Task 6**********************************
data_a = merge_data.groupby('Party')['Total Population'].mean()
#data_a = merge_data.groupby('Party').mean()
demo_county=merge_data['Total Population'][merge_data['Party']==1]
rep_county=merge_data['Total Population'][merge_data['Party']==0]
demo_mean=demo_county.mean()
rep_mean=rep_county.mean()
data_a.head()
```

Out[21]:
```
Party
0     53864.672414
1    300998.316923
Name: Total Population, dtype: float64
```

In [22]:
```python
[ttest_stat,pvalue]=ttest_ind(demo_county,rep_county, equal_var=False)
print('ttest_stat',ttest_stat)
print(pvalue)
```

```
ttest_stat 8.004638577960957
2.0478717602973023e-14
```

In [23]:
```
#st.ttest,pval = st.ttest_ind(demo, Rep)
print("p-value",pvalue)
if pvalue < 0.05:
  print("reject null hypothesis")
else:
  print("accept null hypothesis")
```

```
p-value 2.0478717602973023e-14
reject null hypothesis
```

In [24]:
```
#Population of Democratic counties is higher than republican counties
```

In [25]:
```
#*********************** Task 7**********************************
demo_MHI=merge_data['Median Household Income'][merge_data['Party']==1]
rep_MHI=merge_data['Median Household Income'][merge_data['Party']==0]
demo_MHI_mean=demo_MHI.mean()
rep_MHI_mean=rep_MHI.mean()
print('Democratic MHI', demo_MHI_mean)
print('Republican MHI', rep_MHI_mean)
#MHI = merge_data.groupby('Party')['Median Household Income'].mean()

#//checking purposes
#0  48817.001159
#1    54201.577558
#data_b = merge_data.groupby('Party')['Median Household Income'].mean()
#grp = MHI.groupby('Party')
#grp1 = grp.get_group(0)
#grp1.head()
#grp = MHI.groupby('Party')
#grp2 = grp.get_group(1)
#grp2.head()
```

```
Democratic MHI 53798.732307692306
Republican MHI 48746.81954022989
```

In [26]:
```
#Median Household Income for Democrats is higher than Republicans
```

In [27]:
```
#Democrat and Republican means are both higher than our P-val therefore we must reject
ttest,pval = st.ttest_ind(demo_MHI,rep_MHI,equal_var=False)
print('ttest_stat',ttest)
print("p-value",pval)
if pval < 0.05:
  print("reject null hypothesis")
else:
  print("accept null hypothesis")
```

```
ttest_stat 5.479141589767387
p-value 7.149437363182572e-08
reject null hypothesis
```

In [28]:
```
#Median household Income
#p-value 7.149437363182598e-08
#reject null hypothesis
```

In [29]:
```
merge_data.describe()
```

Out[29]:

FIPS	Total 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	Median Household Income	Percent Unemployed	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Rural	Party
1195.000000	1.195000e+03	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000	1195.000000
38283.497071	1.210768e+05	79.128457	5.563599	10.509368	5.076938	49.836106	36.745805	18.112059	50120.770711	6.539559	13.431087	78.613134	55.890266	0.271967
13009.708572	3.189417e+05	19.756652	9.290601	15.787826	6.065171	2.379191	5.622903	4.759789	12302.863971	2.773478	6.426123	9.181991	32.105077	0.445159
4001.000000	7.600000e+01	2.776702	0.000000	0.000000	0.000000	21.513413	11.842105	6.653188	21190.000000	0.000000	2.134454	26.335440	0.000000	0.000000
27144.000000	1.212200e+04	70.187494	0.536934	1.819735	1.466434	49.369757	33.283841	15.060018	42114.500000	4.717913	8.914888	74.406739	29.694051	0.000000
39137.000000	3.265300e+04	86.809245	1.607865	3.880463	2.865353	50.315399	36.393541	17.705904	48407.000000	6.419141	11.840739	81.009070	56.182090	0.000000
48417.000000	8.587900e+04	93.868714	6.440764	10.994025	6.325443	50.974363	39.435814	20.506315	54998.000000	8.115772	16.760684	85.003634	83.228082	1.000000
56043.000000	4.434257e+06	99.627329	63.953279	95.479801	52.229868	56.418468	67.367823	37.622759	125672.000000	18.771186	49.673777	97.014925	100.000000	1.000000

In [30]:
```
grpP = merge_data.groupby('Party')
grpFemale = grpP['Percent Female'].mean()
grpFemale
```

Out[30]:
```
Party
0    49.630898
1    50.385433
Name: Percent Female, dtype: float64
```

In [31]:
```
grpMalerep = 100 - grpFemale[0]
grpMalerep
```

Out[31]:
```
50.369102366229924
```

In [32]:
```
grpMaledem = 100 - grpFemale[1]
grpMaledem
```

Out[32]:
```
49.61456664655384
```

In [33]:
```
#Average percantage of voters Age under 29 for counties that voted Dem and Rep
grpAge29U = grpP['Percent Age 29 and Under'].mean()
grpAge29U
```

Out[33]:
```
Party
0    36.005719
1    38.726959
Name: Percent Age 29 and Under, dtype: float64
```

In [34]:
```
#Average percantage of voters Age over 65 for counties that voted Dem and Rep
grpAge65O = grpP['Percent Age 65 and Older'].mean()
grpAge65O
```

Out[34]:
```
Party
0    18.828267
1    16.194826
Name: Percent Age 65 and Older, dtype: float64
```

In [35]:
```python
#Average percentage of voters Age Over 29 and Under 65 for Rep Counties
grpAge29OU65rep = 100 - (grpAge65O[0] + grpAge29U[0])
grpAge29OU65rep
```
Out[35]:
```
45.166014679095426
```

In [36]:
```python
#Average percentage of voters Age Over 29 and Under 65 for Dem Counties
grpAge29OU65dem = 100 - (grpAge65O[1] + grpAge29U[1])
grpAge29OU65dem
```
Out[36]:
```
45.07821419864001
```

In [37]:
```python
#Race
grpP.head(1)
grpWhite = grpP['Percent White, not Hispanic or Latino'].mean()
grpWhite
```
Out[37]:
```
Party
0    82.656646
1    69.683766
Name: Percent White, not Hispanic or Latino, dtype: float64
```

In [38]:
```python
grpBlack = grpP['Percent Black, not Hispanic or Latino'].mean()
grpBlack
```
Out[38]:
```
Party
0    4.189241
1    9.242649
Name: Percent Black, not Hispanic or Latino, dtype: float64
```

In [39]:
```python
grpHispanic = grpP['Percent Hispanic or Latino'].mean()
grpHispanic
```
Out[39]:
```
Party
0     9.733094
1    12.587391
Name: Percent Hispanic or Latino, dtype: float64
```

In [40]:
```python
grpLessHS = grpP['Percent Less than High School Degree'].mean()
grpLessHS
```
Out[40]:
```
Party
0    14.009112
1    11.883760
Name: Percent Less than High School Degree, dtype: float64
```

In [41]:
```python
grpLessBD = grpP["Percent Less than Bachelor's Degree"].mean()
grpLessBD
```
Out[41]:
```
Party
0    81.095427
1    71.968225
Name: Percent Less than Bachelor's Degree, dtype: float64
```

In [42]:
```python
grpMoreBDrep = 100 - grpLessBD[0] #Rep with BD or more
grpMoreBDrep
```
Out[42]:
```
18.904572714586266
```

In [43]:
```
grpMoreBDdem = 100 - grpLessBD[1] #Dem with BD or more
grpMoreBDdem
```

Out[43]:
```
28.03177501178463
```

In [44]:
```
#grpP.plot(kind='bar',x='Party',y='Count')
#grpP.head()
#grpP.plot(kind='bar',x='Democratic')
#grpP.show()
```

In [45]:

```
##Task 8
merge_data.groupby('Party').describe(include=[np.number]).transpose()
```

Out[45]:

	Party	0	1
Democratic	**count**	870.000000	325.000000
	mean	7926.549425	71193.172308
	std	17538.649168	125306.803889
	min	6.000000	521.000000
	25%	951.500000	5242.000000
	50%	2807.500000	18159.000000
	75%	7010.750000	72677.000000
	max	215190.000000	881802.000000
Republican	**count**	870.000000	325.000000
	mean	12644.403448	41322.861538
	std	22601.266060	74689.108440
	min	46.000000	220.000000
	25%	2544.000000	3611.000000
	50%	5932.500000	12348.000000
	75%	12632.750000	46403.000000
	max	219990.000000	672505.000000
FIPS	**count**	870.000000	325.000000
	mean	38714.074713	37130.873846
	std	12658.615292	13860.571592
	min	4003.000000	4001.000000
	25%	30073.500000	27027.000000
	50%	42040.000000	36103.000000
	75%	48342.500000	51095.000000
	max	56043.000000	56001.000000
Total Population	**count**	870.000000	325.000000
	mean	53864.672414	300998.316923
	std	94192.572794	553600.025712
	min	76.000000	1969.000000
	25%	9559.500000	23645.000000
	50%	25465.000000	82049.000000
...	**...**	...	...
Percent Unemployed	**std**	2.766060	2.763816
	min	0.000000	0.313234
	25%	4.558432	5.074594
	50%	6.376338	6.617676
	75%	8.069736	8.234271
	max	18.525791	18.771186

	Party	0	1
Percent Less than High School Degree	**count**	870.000000	325.000000
	mean	14.009112	11.883760
	std	6.303126	6.505613
	min	2.134454	3.215803
	25%	9.662491	7.893714
	50%	12.572435	10.370080
	75%	17.447168	13.637059
	max	47.812773	49.673777
Percent Less than Bachelor's Degree	**count**	870.000000	325.000000
	mean	81.095427	71.968225
	std	6.815537	11.192404
	min	43.419470	26.335440
	25%	78.108424	65.711800
	50%	82.406700	72.736143
	75%	85.546272	79.903653
	max	97.014925	94.849957
Percent Rural	**count**	870.000000	325.000000
	mean	63.274485	36.123281
	std	28.766224	32.259481
	min	0.000000	0.000000
	25%	40.738057	5.928800
	50%	63.415088	26.862739
	75%	91.701077	60.670737
	max	100.000000	100.000000

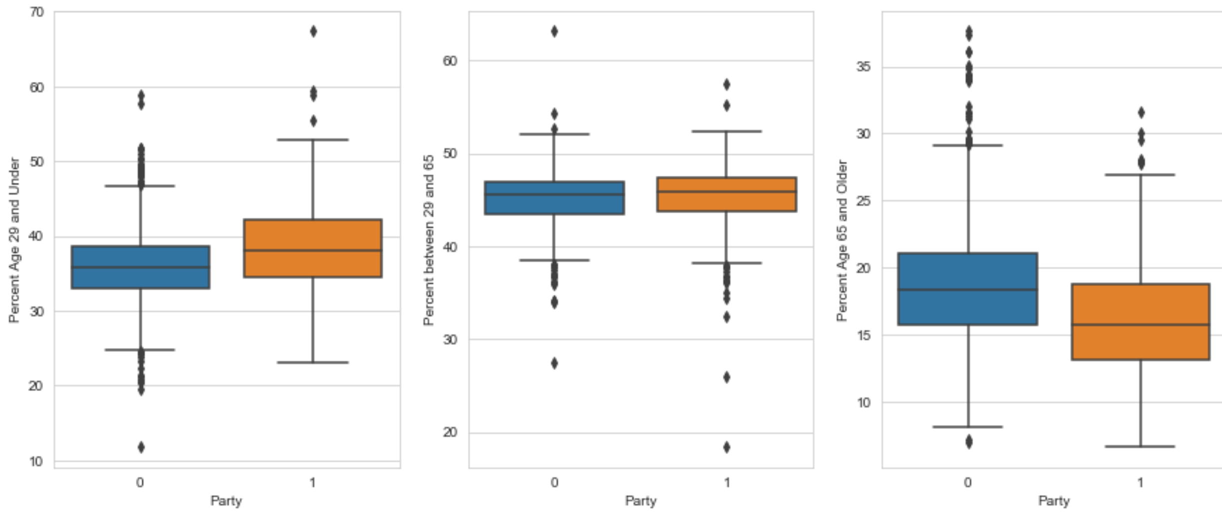
128 rows × 2 columns

In [46]:

```
sns.set_style("whitegrid")
fig,axes = plt.subplots(1,3, figsize = (15,6))
sns.boxplot(x='Party',y='Percent Age 29 and Under',data = merge_data,ax= axes[0])
Percent29less=merge_data['Percent Age 29 and Under']
Percent65older=merge_data['Percent Age 65 and Older']
Percent29and65=100-(Percent29less+Percent65older)
ylabel=sns.boxplot(x='Party',y=Percent29and65,data = merge_data,ax= axes[1])
ylabel.set(ylabel='Percent between 29 and 65')
sns.boxplot(x='Party',y='Percent Age 65 and Older',data = merge_data,ax= axes[2])
```

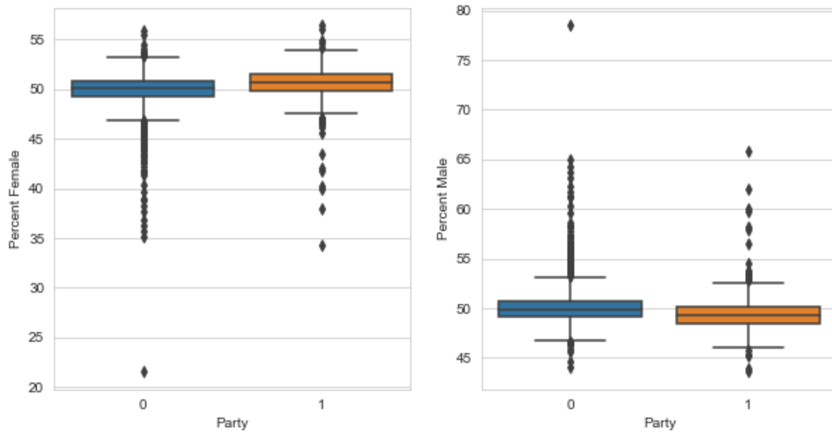
Out[46]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c2c61f2e8>
```



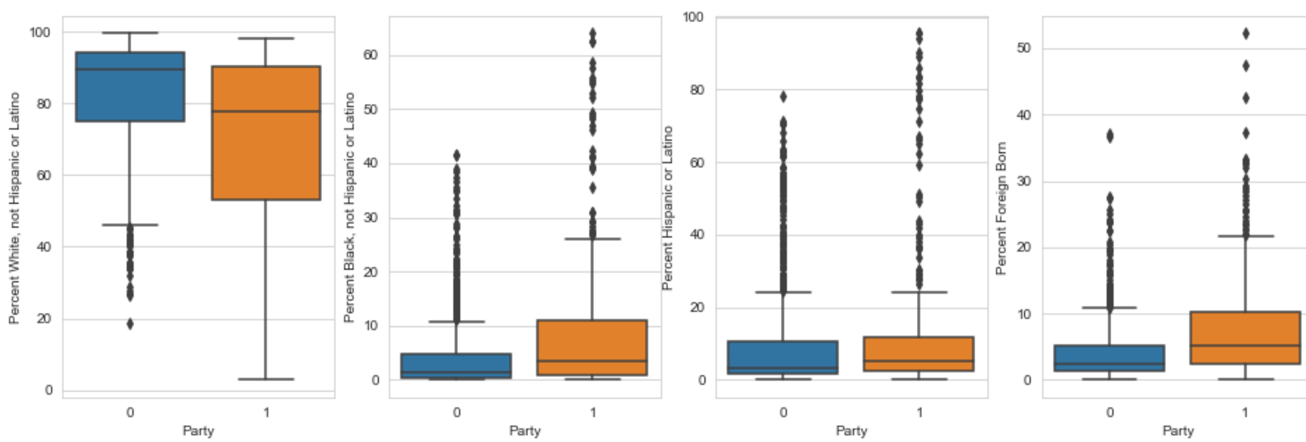
In [47]:
```
female=merge_data['Percent Female']
male=100-female
fig,axes = plt.subplots(1,2, figsize = (10,5))
sns.boxplot(x='Party',y='Percent Female',data = merge_data,ax= axes[0])
ylabel=sns.boxplot(x='Party',y=male,data = merge_data,ax= axes[1])
ylabel.set(ylabel='Percent Male')
```

Out[47]:
```
[Text(0, 0.5, 'Percent Male')]
```



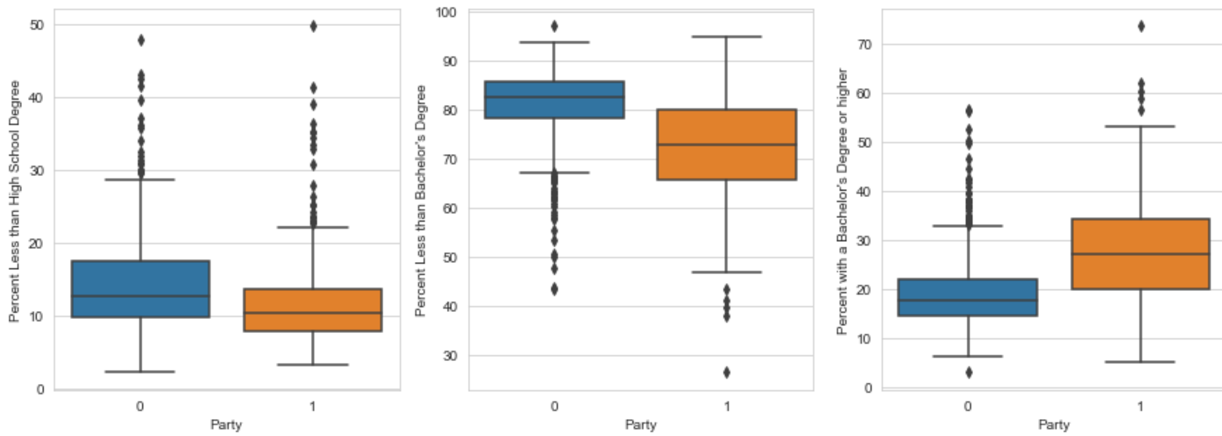
In [48]:
```
fig,axes = plt.subplots(1,4, figsize = (16,5))
sns.boxplot(x='Party', y='Percent White, not Hispanic or Latino',data = merge_data,ax=axes[0])
sns.boxplot(x='Party', y='Percent Black, not Hispanic or Latino',data = merge_data,ax=axes[1])
sns.boxplot(x='Party', y='Percent Hispanic or Latino',data = merge_data,ax=axes[2])
sns.boxplot(x='Party', y='Percent Foreign Born',data = merge_data,ax=axes[3])
```

Out[48]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c2cc98e10>
```



In [49]:
```
fig,axes = plt.subplots(1,3, figsize = (15,5))
sns.boxplot(x='Party',y='Percent Less than High School Degree',data = merge_data,ax= axes[0])
sns.boxplot(x='Party',y="Percent Less than Bachelor's Degree",data = merge_data,ax= axes[1])
c = sns.boxplot(x='Party',y=100-merge_data["Percent Less than Bachelor's Degree"],data = merge_data,ax= axes[2])
c.set_ylabel("Percent with a Bachelor's Degree or higher")
```

Out[49]:
```
Text(0, 0.5, "Percent with a Bachelor's Degree or higher")
```

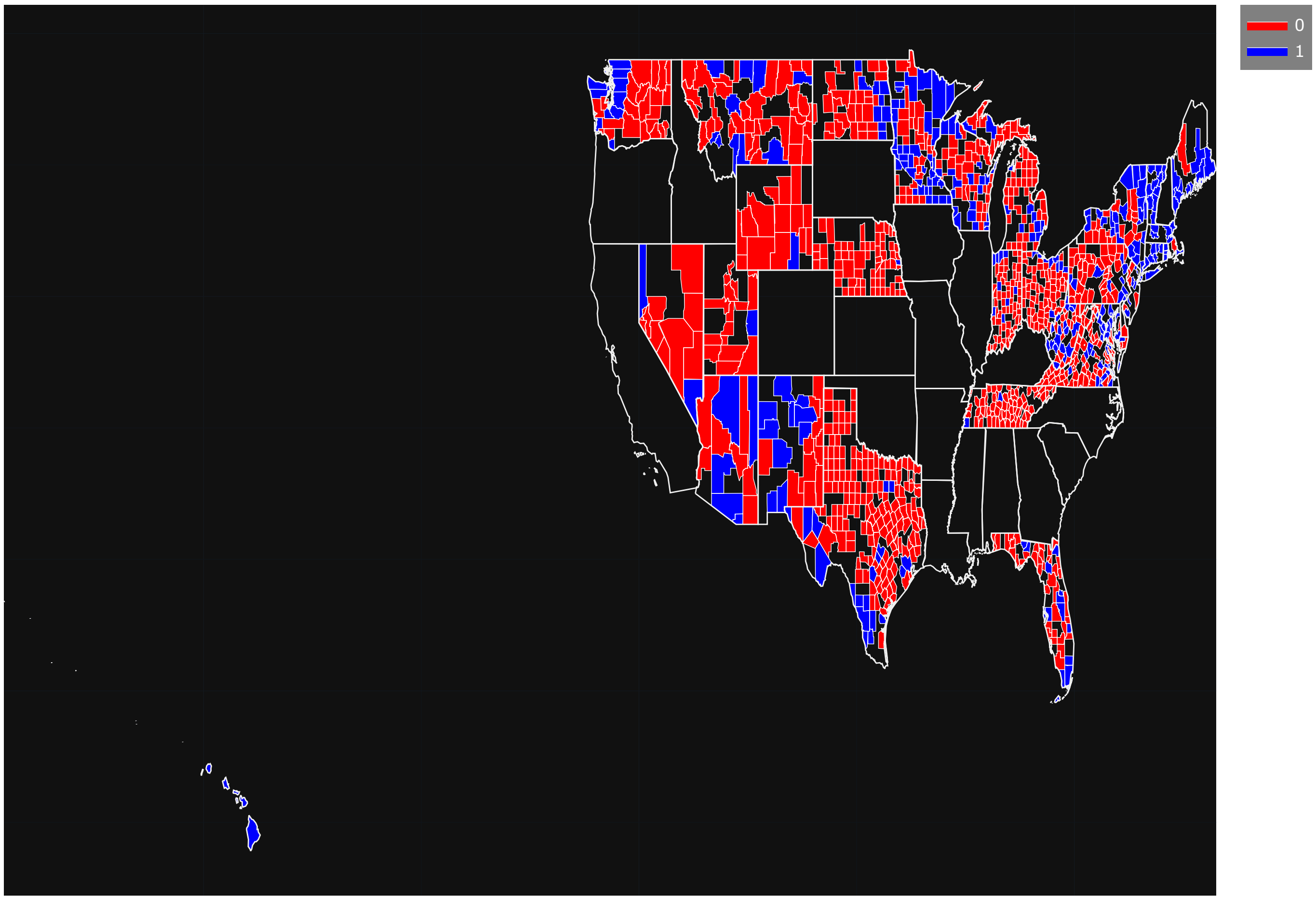


Task 8: The conclusions we make from the variables is the Republicans win counties where there is a high White Demographic than the Democrats.Although both relatively win counties where the population is highly white, but the Republicans seem to win the whiter the county is. Whereas, the Democrats win Counties with higher black and foreign born populated counties compared to the Republicans. In the case of the hispanic population it is to close to say wether a higher hispanic population can help either party win a county. In the case of gender the data is very tight along party lines but their is a small diffrence seen in the boxplots and that Democrats primarly win where there is a greater female population and Republicans win where there is a greater male population. Now in the case of age we can conclude that people aged 29 and under tend to vote Democrats as is why Democrats win counties with a younger population that votes and people aged 65 and over vote Republican more and that is why Republicans win counties with the older populations but as far as the ages in between it is too close to say wether they lean more Democrat or Republican as the stats are far to close to say wether they tend to vote a certain way, although very minimal if observed closely on the boxplots you can see that Democrats to win that age just by a very minute margin but not enough to say they will in the future. Lastly in the case of education we can conclude based on our data and plots that counties with higher populations with people with less than a high school degree and a Bachelors degree will vote Republican more whereas in counties with a high population with more than a Bachelors degree will vote Democrat which makes sense as seen in other data reports that republicans do well in the deep south and the deep south has a higher than average highschool dropout rates than in other parts of ther country. So with these box plots it helps us visualize how certain demographics can influence an election before the election just based on how certain counties with a high population of such demographic will vote a certain way.

Task 9 I believe that Education and Age are the most important as seen in the differnce in the plots and data is more marginal than other varaibles like race and gender where the data isnt strong enough to support wether a county will be Democrat or Republican although with age and education we can see it and also agree based on how election turnouts go in the case of education as seen in todays politcal landscape the less educated tend to be more Republican seen in the south and rural areas like Nebraska compared to more educated areas like California and New York that tend to vote Democrat but this is mostly on the basis that education plays a big role on an individual to decide how to vote it makes them more concise on what they are voting for and know who not to vote for like a compulsive liar or somebody who doesnt understand basic science and lifestlye neccessities. In the case of Age it makes sense that Democrats outwin the Republicans in this sector based on the parties pillars. Young voters tend to be more for free education as this is the age they are going to school so what will be better than getting to go to school without worrying about the finances of it or student loans. Democrats are keen also in more progressive ideas like same-sex marriage, legalization of recreational cannabis, and womens reproductive rights. Whereas the senior populations vote more Republican as that is the party of more conservative and traditional values that older people cherish more not to mention this is the age where they are retired and are more economically vulnerable because changes in their society can affect their livelihood and even threatend their current way of life. But the in between age of these two groups is what really decides whether a county will vote a certain way and perhaps with more through information we can also investigate whether if say there is a high young population and in between age population with a extremly tiny population will this indirecty make the middle population vote Democrat via influence of the larger younger population or will it have the other effect and vice versa.

In [ ]:
```python
#********************************Task 10***************************

fips = merge_data['FIPS']
values=merge_data['Party']

colorscale = [
    'rgb(255,0,0)',
    'rgb(0,0,255)'
]

fig=ff.create_choropleth(
                         fips=fips,colorscale = colorscale,values=values,scope=['USA', 'HI'],
                         county_outline={'color':'white','width':0.5},
                        )
fig.layout = dict(
    height = 700,
    paper_bgcolor = 'GRAY',
    margin = dict(t = 25, b = 0, l = 0, r = 0),
    font = dict(color = 'WHITE', size = 12),
    mapbox = dict(zoom = 3,)
    )

fig.layout['title'] = 'Party WINS by County'
fig.layout.template= 'plotly_dark'


fig.show()
#The next folowing line of code is only for rendering purposes
fig.write_image("images/fig1.pdf")
```

```
//anaconda3/lib/python3.7/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 [ ]:

In [ ]:

In [ ]: