In [69]:
```
#Prepare and explore a dataset containing results for the 2018 United States S
enate elections and demographic information for United States counties
import pandas as pd
import numpy as np
import scipy.stats as st
import math
import plotly.figure_factory as ff
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
election = pd.read_csv("election_train.csv")
```

In [70]:
```
election.head()
```

Out[70]:

	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 [71]:
```
election.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
```

TASK 01

In [72]:
```
#task 1 : Reshape dataset election_train from long format to wide format
election_reshaped = pd.pivot_table(election,values="Votes",index=["Year","State","County","Office"],columns="Party",aggfunc=np.sum).reset_index()
election_reshaped
```

Out[72]:

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

Party	Year	State	County	Office	Democratic	Republican
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 × 6 columns

TASK 02

In [73]:
```
#task 2 :  Merge reshaped dataset election_train with dataset demographics_tra
in.
election_reshaped.County = election_reshaped.County.str.replace('County','')
election_reshaped.County = election_reshaped.County.str.strip()
demographics = pd.read_csv("demographics_train.csv")
shortforms = {
    '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',
```

```
}
demographics.State = demographics.State.map(shortforms)
demographics.County = demographics.County.str.strip()
election_reshaped.State = election_reshaped.State.str.strip()
election_reshaped.County = election_reshaped.County.str.lower()
demographics.County = demographics.County.str.lower()
data_merged = demographics.merge(election_reshaped,on=['County','State'],how=
'inner')
data_merged
```

Out[73]:

	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	Perce Forei Bo
0	WI	la crosse	55063	117538	0	90.537528	1.214075	1.724549	2.9760
1	VA	alleghany	51005	15919	12705	91.940449	5.207614	1.432251	1.3003
2	IN	fountain	18045	16741	12750	95.705155	0.400215	2.359477	1.5471
3	OH	geauga	39055	94020	0	95.837056	1.256116	1.294405	2.5781
4	WI	jackson	55053	20566	15835	86.662453	1.983857	3.082758	1.3760
5	TX	baylor	48023	3639	0	86.644683	1.841165	8.353943	2.4732
6	NE	madison	31119	35125	24885	81.249822	1.155872	14.217794	6.7843
7	HI	hawaii	15001	193680	0	30.401694	0.547811	12.405514	11.0037
8	TN	henry	47079	32291	25285	87.662197	8.599919	2.201852	1.5608
9	MI	oceana	26127	26152	18930	82.486999	1.131845	14.419547	5.5789
10	NE	pierce	31139	7179	5385	96.893718	0.222872	1.587965	0.7800
11	TX	jack	48237	8866	6535	78.411911	4.376269	15.880893	5.5492
12	FL	walton	12131	61528	47490	84.447731	4.950592	5.888376	5.7599
13	VA	washington	51191	54562	0	95.484037	1.268282	1.414904	1.6110
14	FL	escambia	12033	309574	0	65.219624	21.532816	5.389988	4.7474
15	TX	wheeler	48483	5642	3785	68.397731	2.658632	26.462247	9.5179
16	AZ	yavapai	4025	218586	0	81.159361	0.518331	14.054880	6.4569
17	NE	loup	31115	542	435	97.970480	0.000000	0.000000	0.0000
18	MI	antrim	26009	23215	0	95.179841	0.323067	1.955632	2.0159
19	MN	wabasha	27157	21327	16385	94.926619	0.150045	2.874291	1.3550
20	NV	elko	32007	52029	34740	67.354744	1.060947	24.067347	8.6028

	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	Perce Forei Bo
21	WI	dodge	55027	88404	0	90.629383	2.085878	4.485091	1.8845
22	MN	lake of the woods	27077	3901	0	93.975904	0.102538	1.230454	1.5637
23	WV	tucker	54093	6922	5690	97.688529	0.173360	0.650101	0.5923
24	VA	lexington city	51678	7036	6335	80.542922	8.996589	4.178511	3.9368
25	FL	gilchrist	12041	17033	0	87.295250	5.988376	5.336699	2.9061
26	TN	claiborne	47025	31701	25350	95.495410	1.050440	1.113530	1.1640
27	CT	middlesex	9007	164438	0	84.738929	4.664980	5.641032	7.6527
28	UT	sevier	49041	20913	0	92.244059	0.344283	4.882131	1.9891
29	MA	hampshire	25015	161035	0	84.610799	2.683268	5.253516	7.8461
...	...	...	...	...	...	...	...	...	
1170	WV	wood	54107	86262	67640	95.478890	1.123322	1.032900	1.0016
1171	FL	orange	12095	1256055	854605	42.656970	19.981211	29.338126	20.1573
1172	ND	walsh	38099	10995	8280	85.657117	0.300136	10.759436	3.9199
1173	IN	white	18181	24265	0	90.245209	0.461570	7.554090	4.3148
1174	MT	carter	30011	1295	0	96.602317	0.000000	1.389961	0.0772
1175	IN	pulaski	18131	12910	9870	94.771495	0.209140	2.788536	0.8055
1176	WA	lewis	53041	75724	56900	84.489726	0.662934	9.629708	4.5124
1177	PA	greene	42059	37669	0	92.598689	4.149300	1.422921	0.8096
1178	TX	yoakum	48501	8316	0	34.271284	0.000000	63.552189	25.6373
1179	MT	pondera	30073	6166	0	80.960104	0.389231	1.881284	2.1894
1180	ND	dickey	38021	5160	0	93.720930	1.472868	3.333333	1.7635
1181	TX	refugio	48391	7315	0	42.761449	4.743677	49.323308	2.7341

	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	Perce Forei Bo
1182	VA	smyth	51173	31513	24960	94.376924	2.068987	1.948402	1.2566
1183	VA	charlotte	51037	12232	9540	67.282538	30.657292	0.948332	0.6621
1184	MI	grand traverse	26055	90715	70710	92.673758	1.234636	2.621397	2.4218
1185	IN	rush	18139	16873	0	96.325490	1.149766	1.428317	1.0608
1186	MT	deer lodge	30023	9176	0	90.747602	0.348736	3.073235	1.8744
1187	OH	washington	39167	61154	0	95.364163	1.100500	1.013834	1.4308
1188	WI	portage	55097	70551	0	91.815850	0.687446	2.951057	3.1197
1189	NY	monroe	36055	749236	0	71.329327	14.497969	8.120272	8.4607
1190	FL	broward	12011	1863780	0	39.245351	27.214639	27.564841	32.7157
1191	WA	wahkiakum	53069	4051	0	89.681560	0.123426	4.640829	3.5053
1192	NY	tioga	36107	49649	38525	95.454088	0.713005	1.695905	2.2296
1193	MT	carbon	30009	10340	0	94.787234	0.009671	2.379110	1.2765
1194	TX	johnson	48251	157544	0	74.343041	2.487559	19.762733	6.1386
1195	MT	lincoln	30053	19268	15640	93.351671	0.057089	2.678015	2.3458
1196	OH	tuscarawas	39157	92579	70485	95.155489	0.804718	2.349345	1.6504
1197	MI	newaygo	26123	47957	0	90.716684	1.317847	5.728048	1.9788
1198	TN	lauderdale	47097	27261	0	60.456330	34.789626	2.380690	1.7570
1199	TX	sabine	48403	10367	0	86.341275	7.080158	3.839105	0.7523

1200 rows × 21 columns

TASK 03

In [74]:
```
#task 3: Explore the merged dataset
data_merged.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1200 entries, 0 to 1199
Data columns (total 21 columns):
State                                    1200 non-null object
County                                   1200 non-null object
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
Year                                     1200 non-null int64
Office                                   1200 non-null object
Democratic                               1200 non-null float64
Republican                               1200 non-null float64
dtypes: float64(13), int64(5), object(3)
memory usage: 206.2+ KB
```

Task 3 Questions

Q1: How many variables does the dataset have?

Answer: There are 21 variables present in the merged dataset

Q2: What is the type of these variables?

Answer: The types of the variables - {object, int64, float64}

Q3: Are there any irrelevant or redundant variables?

Answer: Irrelavant variables are Office and Year. Redundant variable in the dataset is Citizen Voting-Age Population

Q4: how will you deal with these variables?

Answer: We can drop the variable in the dataset or we can ignore it

In [75]:
```
# Dropping the irrelavant variables from the merged data set
data_merged = data_merged.drop(['Office', 'Year'],axis=1)
```

In [76]:
```
data_merged.head()
```

Out[76]:

	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	P F
0	WI	la crosse	55063	117538	0	90.537528	1.214075	1.724549	2.976059	51.1
1	VA	alleghany	51005	15919	12705	91.940449	5.207614	1.432251	1.300333	51.0
2	IN	fountain	18045	16741	12750	95.705155	0.400215	2.359477	1.547100	49.7
3	OH	geauga	39055	94020	0	95.837056	1.256116	1.294405	2.578175	50.6
4	WI	jackson	55053	20566	15835	86.662453	1.983857	3.082758	1.376058	46.6

Task 4 Questions

Q1: Are there any missing values?

Answer: Yes, there are missing values in the dataset. The missing values are assigned with a value '0' variable 'Citizen Voting-Age Population' has missing values more than 50%.
Some of the observations have both Democratic votes and Republic votes as '0'

Q2: If so, how will you deal with these values?

Answer: we can deal with the missing values by dropping, replacing, estimating and ignoring In this case, we are dropping a variable and few observations. The variable 'Citizen Voting-Age Population' is dropped because it has missing values more than 50% and we are dropping 5 observations with both Democratic votes and Republican votes as '0'

In [77]:
```
# Dropping the column "Citizen Voting-Age Population" since there are more obs
ervations with the missing values
data_merged = data_merged.drop(['Citizen Voting-Age Population'],axis=1)
```

In [78]:

```
# Drop the observations with both democratic votes and republican votes as '0'
# We dont have observations with democratic votes as 0 and republic votes as n
on zero. Also we dont have observations with democratic votes as non zero and
 republican votes as 0.
# So we can directly drop observations with democratic votes as 0 which will a
lso drop those observations with republican votes as 0
data_merged = data_merged[data_merged.Democratic != 0]
data_merged.head()
```

Out[78]:

	State	County	FIPS	Total Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percent Female	Pe A (
0	WI	la crosse	55063	117538	90.537528	1.214075	1.724549	2.976059	51.171536	43.24
1	VA	alleghany	51005	15919	91.940449	5.207614	1.432251	1.300333	51.077329	31.66
2	IN	fountain	18045	16741	95.705155	0.400215	2.359477	1.547100	49.770026	35.89
3	OH	geauga	39055	94020	95.837056	1.256116	1.294405	2.578175	50.678579	36.28
4	WI	jackson	55053	20566	86.662453	1.983857	3.082758	1.376058	46.649810	36.29

TASK 05

In [79]:
```
#Task 5 : Creating a new variable 'Party' based on the votes of Democratic and Republican

data_merged["Party"] = np.where(data_merged['Democratic'] > data_merged['Republican'], 1, 0)
data_merged
```

Out[79]:

	State	County	FIPS	Total Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percer Femal
0	WI	la crosse	55063	117538	90.537528	1.214075	1.724549	2.976059	51.17153
1	VA	alleghany	51005	15919	91.940449	5.207614	1.432251	1.300333	51.07732
2	IN	fountain	18045	16741	95.705155	0.400215	2.359477	1.547100	49.77002
3	OH	geauga	39055	94020	95.837056	1.256116	1.294405	2.578175	50.67857
4	WI	jackson	55053	20566	86.662453	1.983857	3.082758	1.376058	46.64981
5	TX	baylor	48023	3639	86.644683	1.841165	8.353943	2.473207	51.66254
6	NE	madison	31119	35125	81.249822	1.155872	14.217794	6.784342	50.44839
7	HI	hawaii	15001	193680	30.401694	0.547811	12.405514	11.003717	50.14301
8	TN	henry	47079	32291	87.662197	8.599919	2.201852	1.560806	51.44157
9	MI	oceana	26127	26152	82.486999	1.131845	14.419547	5.578923	49.39584
10	NE	pierce	31139	7179	96.893718	0.222872	1.587965	0.780053	49.65872
11	TX	jack	48237	8866	78.411911	4.376269	15.880893	5.549289	43.18745
12	FL	walton	12131	61528	84.447731	4.950592	5.888376	5.759979	49.34988
13	VA	washington	51191	54562	95.484037	1.268282	1.414904	1.611011	50.50034
14	FL	escambia	12033	309574	65.219624	21.532816	5.389988	4.747492	50.29944
15	TX	wheeler	48483	5642	68.397731	2.658632	26.462247	9.517901	49.20241
16	AZ	yavapai	4025	218586	81.159361	0.518331	14.054880	6.456955	51.09247
17	NE	loup	31115	542	97.970480	0.000000	0.000000	0.000000	52.39852
18	MI	antrim	26009	23215	95.179841	0.323067	1.955632	2.015938	50.27353
19	MN	wabasha	27157	21327	94.926619	0.150045	2.874291	1.355090	50.17114
20	NV	elko	32007	52029	67.354744	1.060947	24.067347	8.602895	47.94826
21	WI	dodge	55027	88404	90.629383	2.085878	4.485091	1.884530	47.78856
22	MN	lake of the woods	27077	3901	93.975904	0.102538	1.230454	1.563702	47.96206
23	WV	tucker	54093	6922	97.688529	0.173360	0.650101	0.592314	50.24559
24	VA	lexington city	51678	7036	80.542922	8.996589	4.178511	3.936896	43.44798
25	FL	gilchrist	12041	17033	87.295250	5.988376	5.336699	2.906123	47.39036
26	TN	claiborne	47025	31701	95.495410	1.050440	1.113530	1.164001	51.20027
27	CT	middlesex	9007	164438	84.738929	4.664980	5.641032	7.652732	51.09889
28	UT	sevier	49041	20913	92.244059	0.344283	4.882131	1.989193	48.89303
29	MA	hampshire	25015	161035	84.610799	2.683268	5.253516	7.846120	53.23066
...	...	...	...	...	...	...	...	...	.
1170	WV	wood	54107	86262	95.478890	1.123322	1.032900	1.001600	51.75627

	State	County	FIPS	Total Population	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino	Percent Foreign Born	Percer Femal
1171	FL	orange	12095	1256055	42.656970	19.981211	29.338126	20.157318	50.82755
1172	ND	walsh	38099	10995	85.657117	0.300136	10.759436	3.919964	48.72214
1173	IN	white	18181	24265	90.245209	0.461570	7.554090	4.314857	50.29466
1174	MT	carter	30011	1295	96.602317	0.000000	1.389961	0.077220	51.73745
1175	IN	pulaski	18131	12910	94.771495	0.209140	2.788536	0.805577	49.31061
1176	WA	lewis	53041	75724	84.489726	0.662934	9.629708	4.512440	50.15847
1177	PA	greene	42059	37669	92.598689	4.149300	1.422921	0.809684	48.08463
1178	TX	yoakum	48501	8316	34.271284	0.000000	63.552189	25.637326	49.15824
1179	MT	pondera	30073	6166	80.960104	0.389231	1.881284	2.189426	49.72429
1180	ND	dickey	38021	5160	93.720930	1.472868	3.333333	1.763566	49.80620
1181	TX	refugio	48391	7315	42.761449	4.743677	49.323308	2.734108	50.79972
1182	VA	smyth	51173	31513	94.376924	2.068987	1.948402	1.256624	50.42363
1183	VA	charlotte	51037	12232	67.282538	30.657292	0.948332	0.662198	49.95912
1184	MI	grand traverse	26055	90715	92.673758	1.234636	2.621397	2.421871	50.83062
1185	IN	rush	18139	16873	96.325490	1.149766	1.428317	1.060866	50.97493
1186	MT	deer lodge	30023	9176	90.747602	0.348736	3.073235	1.874455	46.78509
1187	OH	washington	39167	61154	95.364163	1.100500	1.013834	1.430814	50.59685
1188	WI	portage	55097	70551	91.815850	0.687446	2.951057	3.119729	49.84621
1189	NY	monroe	36055	749236	71.329327	14.497969	8.120272	8.460752	51.69412
1190	FL	broward	12011	1863780	39.245351	27.214639	27.564841	32.715718	51.37650
1191	WA	wahkiakum	53069	4051	89.681560	0.123426	4.640829	3.505307	51.24660
1192	NY	tioga	36107	49649	95.454088	0.713005	1.695905	2.229652	50.55489
1193	MT	carbon	30009	10340	94.787234	0.009671	2.379110	1.276596	49.99032
1194	TX	johnson	48251	157544	74.343041	2.487559	19.762733	6.138603	50.16884
1195	MT	lincoln	30053	19268	93.351671	0.057089	2.678015	2.345858	49.97405
1196	OH	tuscarawas	39157	92579	95.155489	0.804718	2.349345	1.650482	50.82362
1197	MI	newaygo	26123	47957	90.716684	1.317847	5.728048	1.978856	49.65698
1198	TN	lauderdale	47097	27261	60.456330	34.789626	2.380690	1.757089	47.73485
1199	TX	sabine	48403	10367	86.341275	7.080158	3.839105	0.752387	50.50641

1195 rows × 19 columns

TASK 06

In [80]:
```
#Task 6: Compute the mean population for Democratic counties and Republican co
unties
data_democratic = data_merged.groupby('Party').get_group(1)
mean_democratic = data_democratic['Total Population'].mean()
mean_democratic
```

Out[80]: 300998.3169230769

In [81]:
```
data_republican = data_merged.groupby('Party').get_group(0)
mean_republican = data_republican['Total Population'].mean()
mean_republican
```

Out[81]: 53864.6724137931

Task 6 Questions

Q1: Which one is higher?

Answer: Democratic population mean is higher

In [82]:
```
# Perform a hypothesis test to determine whether this difference is statistica
lly significant at the α = 0. 05 significance level
# We consider the null hypothesis as mean(democratic-Total population) = mean
(republican-Total population)
# we cansider the alternate hypothesis as mean(democratic-Total population) !=
mean(republican-Total population)
# So, the hypothesis test is a two-tailed test

[statistic, pvalue] = st.ttest_ind(data_democratic['Total Population'], data_r
epublican['Total Population'], equal_var = False)
print("T statistic is", statistic)
print("pvalue is", pvalue)
```

```
T statistic is 8.004638577960957
pvalue is 2.0478717602973023e-14
```

Q2: What is the result of the test?

Answer: The pvalue of the hypothesis test is lesser than the significant value. we reject the null hypothesis.

Q3: What conclusion do you make from this result?

Answer: The mean of the population of Democratic is different from the mean of the population of Republican

TASK 07

In [83]:
```
#task 7: Compute the mean median household income for Democratic counties and
 Republican counties.

mean_MHI_democratic = data_democratic['Median Household Income'].mean()
mean_MHI_democratic
```

Out[83]: 53798.732307692306

In [84]:
```
mean_MHI_republican = data_republican['Median Household Income'].mean()
mean_MHI_republican
```

Out[84]:
```
48746.81954022989
```

Task 07 Questions

Q1: Which one is higher?

Answer: Democratic median household income mean is higher Republican median household income mean

In [85]:
```
# Perform a hypothesis test to determine whether this difference is statistically significant at the α = 0.05 significance level
# We consider the null hypothesis as mean(democratic-Median Household Income) = mean(republican-Median Household Income)
# we cansider the alternate hypothesis as mean(democratic-Median Household Income) != mean(republican-Median Household Income)
# So, the hypothesis test is a two-tailed test

[statistic, pvalue] = st.ttest_ind(data_democratic['Median Household Income'], data_republican['Median Household Income'], equal_var = False)
print("T statistic is", statistic)
print("pvalue is", pvalue)
```

```
T statistic is 5.479141589767388
pvalue is 7.149437363182572e-08
```

Q2: What is the result of the test?

Answer: The pvalue of the hypothesis test is lesser than the significant value. we reject the null hypothesis.

Q3: What conclusion do you make from this result?

Answer: The mean of the median household income of Democratic is different from the mean of the median household income of Republican

TASK 08

In [86]:
```
# Descriptive Statistics of Democratic counties and Republican counties in ter
ms of age, gender,race and ethnicity, and education

data_democratic.insert(9, 'Percent Male', 100 - data_democratic['Percent Femal
e'], allow_duplicates="yes")
data_republican.insert(9, 'Percent Male', 100 - data_republican['Percent Femal
e'], allow_duplicates="yes")
data_democratic.insert(16, 'Percent Bachelor\'s Degree or Higher', 100 - data_
democratic['Percent Less than Bachelor\'s Degree'], allow_duplicates="yes")
data_republican.insert(16, 'Percent Bachelor\'s Degree or Higher', 100 - data_
republican['Percent Less than Bachelor\'s Degree'], allow_duplicates="yes")
data_democratic.insert(11, 'Percent Age 29_above and 65_below', 100 - (data_de
mocratic['Percent Age 29 and Under']+data_democratic['Percent Age 65 and Olde
r']), allow_duplicates="yes")
data_republican.insert(11, 'Percent Age 29_above and 65_below', 100 - (data_re
publican['Percent Age 29 and Under']+data_republican['Percent Age 65 and Olde
r']), allow_duplicates="yes")
```

In [87]:
```
data_democratic[['Percent Age 29 and Under', 'Percent Age 29_above and 65_belo
w', 'Percent Age 65 and Older']].describe()
```

Out[87]:

	Percent Age 29 and Under	Percent Age 29_above and 65_below	Percent Age 65 and Older
count	325.000000	325.000000	325.000000
mean	38.726959	45.078214	16.194826
std	6.252786	3.907598	4.282422
min	23.156452	18.433769	6.653188
25%	34.488444	43.741937	13.106233
50%	38.074151	45.817819	15.698087
75%	42.161162	47.448269	18.806426
max	67.367823	57.478906	31.642106

In [88]:
```
data_democratic[['Percent Male', 'Percent Female']].describe()
```

Out[88]:

	Percent Male	Percent Female
count	325.000000	325.000000
mean	49.614567	50.385433
std	2.149359	2.149359
min	43.581532	34.245291
25%	48.507925	49.854280
50%	49.346170	50.653830
75%	50.145720	51.492075
max	65.754709	56.418468

In [89]:
```
data_democratic[['Percent Less than High School Degree',
                 'Percent Less than Bachelor\'s Degree',
                'Percent Bachelor\'s Degree or Higher']].describe()
```

Out[89]:

	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Bachelor's Degree or Higher
count	325.000000	325.000000	325.000000
mean	11.883760	71.968225	28.031775
std	6.505613	11.192404	11.192404
min	3.215803	26.335440	5.150043
25%	7.893714	65.711800	20.096347
50%	10.370080	72.736143	27.263857
75%	13.637059	79.903653	34.288200
max	49.673777	94.849957	73.664560

In [90]:
```
data_democratic[['Percent White, not Hispanic or Latino',
                 'Percent Black, not Hispanic or Latino',
                'Percent Hispanic or Latino']].describe()
```

Out[90]:

	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino
count	325.000000	325.000000	325.000000
mean	69.683766	9.242649	12.587391
std	24.981502	13.351340	19.575030
min	2.776702	0.000000	0.193349
25%	53.271579	0.839103	2.531017
50%	77.786090	3.485992	5.039747
75%	90.300749	11.058843	11.857116
max	98.063495	63.953279	95.479801

In [91]:
```
data_republican[['Percent Age 29 and Under', 'Percent Age 29_above and 65_below', 'Percent Age 65 and Older']].describe()
```

Out[91]:

	Percent Age 29 and Under	Percent Age 29_above and 65_below	Percent Age 65 and Older
count	870.000000	870.000000	870.000000
mean	36.005719	45.166015	18.828267
std	5.181522	2.910264	4.733155
min	11.842105	27.421759	6.954387
25%	32.983652	43.522522	15.784982
50%	35.846532	45.553295	18.377896
75%	38.539787	46.975771	21.112847
max	58.749116	63.157895	37.622759

In [92]:
```
data_republican[['Percent Male', 'Percent Female']].describe()
```

Out[92]:

	Percent Male	Percent Female
count	870.000000	870.000000
mean	50.369102	49.630898
std	2.429013	2.429013
min	44.114977	21.513413
25%	49.170230	49.222905
50%	49.823208	50.176792
75%	50.777095	50.829770
max	78.486587	55.885023

In [93]:
```
data_republican[['Percent Less than High School Degree',
                 'Percent Less than Bachelor\'s Degree',
                'Percent Bachelor\'s Degree or Higher']].describe()
```

Out[93]:

	Percent Less than High School Degree	Percent Less than Bachelor's Degree	Percent Bachelor's Degree or Higher
count	870.000000	870.000000	870.000000
mean	14.009112	81.095427	18.904573
std	6.303126	6.815537	6.815537
min	2.134454	43.419470	2.985075
25%	9.662491	78.108424	14.453728
50%	12.572435	82.406700	17.593300
75%	17.447168	85.546272	21.891576
max	47.812773	97.014925	56.580530

In [94]:
```
data_republican[['Percent White, not Hispanic or Latino',
                 'Percent Black, not Hispanic or Latino',
                'Percent Hispanic or Latino']].describe()
```

Out[94]:

	Percent White, not Hispanic or Latino	Percent Black, not Hispanic or Latino	Percent Hispanic or Latino
count	870.000000	870.000000	870.000000
mean	82.656646	4.189241	9.733094
std	16.056122	6.721695	14.049576
min	18.758977	0.000000	0.000000
25%	75.016397	0.460419	1.704539
50%	89.434849	1.318311	3.427435
75%	94.466596	4.753831	10.709696
max	99.627329	41.563041	78.397012

In [95]:
```
#Plots Drawn on age, gender, race and ethinicity, and education for Democaratic and Republican counties and then compared

#Box plot on gender. Comparing each gender from democratic and republican counties
y0 = 100 - data_democratic['Percent Female']
y1 = data_democratic['Percent Female']
y2 = 100 - data_republican['Percent Female']
y3 = data_republican['Percent Female']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Males (% of total pop) Democratics',marker_color = 'skyblue'))
fig.add_trace(go.Box(y=y2, name = 'Males (% of total pop) Republicans',marker_color = 'darkred'))
fig.add_trace(go.Box(y=y1, name='Females (% of total pop) Democratics',marker_color = 'blue'))
fig.add_trace(go.Box(y=y3, name = 'Females (% of total pop) Republicans',marker_color = 'red'))

fig.show()
```

In [96]:
```
#Box plot on gender. Comparision between genders in both democratic and republ
ican counties
y0 = 100 - data_democratic['Percent Female']
y1 = data_democratic['Percent Female']
y2 = 100 - data_republican['Percent Female']
y3 = data_republican['Percent Female']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Males (% of total pop) Dems',marker_color =
'skyblue'))
fig.add_trace(go.Box(y=y1, name = 'females (% of total pop) Dems',marker_color
= 'blue'))
fig.add_trace(go.Box(y=y2, name='Males (% of total pop) Republicans',marker_co
lor = 'red'))
fig.add_trace(go.Box(y=y3, name = 'Females (% of total pop) Republicans',marke
r_color = 'darkred'))

fig.show()
```

In [97]:
```
#Camparision between ages in both Democratic and Republican counties

y0 = data_democratic['Percent Age 29 and Under']
y1 = 100 - (data_democratic['Percent Age 29 and Under'] + data_democratic['Per
cent Age 65 and Older'])
y2 = data_democratic['Percent Age 65 and Older']
y3 = data_republican['Percent Age 29 and Under']
y4 = 100 - (data_republican['Percent Age 29 and Under'] + data_republican['Per
cent Age 65 and Older'])
y5 = data_republican['Percent Age 65 and Older']

fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Age >=29 Dems % of population',marker_color =
'darkblue'))
fig.add_trace(go.Box(y=y1, name = 'Age 30-64 Dems % of population',marker_colo
r = 'skyblue'))
fig.add_trace(go.Box(y=y2, name = 'Age 65=< Dems % of population',marker_color
= 'blue'))
fig.add_trace(go.Box(y=y3, name = 'Age >=29 Repub % of population',marker_colo
r = 'indianred'))
fig.add_trace(go.Box(y=y4, name = 'Age 30-64 Repub % of population',marker_col
or = 'red'))
fig.add_trace(go.Box(y=y5, name = 'Age 65=< Repub % of population',marker_colo
r = 'darkred'))
fig.show()
```

In [98]:

```
#Camparing each age group from Democratic data with each age group of Republican data
y0 = data_democratic['Percent Age 29 and Under']
y1 = 100 - (data_democratic['Percent Age 29 and Under'] + data_democratic['Percent Age 65 and Older'])
y2 = data_democratic['Percent Age 65 and Older']
y3 = data_republican['Percent Age 29 and Under']
y4 = 100 - (data_republican['Percent Age 29 and Under'] + data_republican['Percent Age 65 and Older'])
y5 = data_republican['Percent Age 65 and Older']

fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Age >=29 Dems % of population',marker_color = 'darkblue'))
fig.add_trace(go.Box(y=y3, name = 'Age >=29 Repub % of population',marker_color = 'indianred'))
fig.add_trace(go.Box(y=y1, name = 'Age 30-64 Dems % of population',marker_color = 'skyblue'))
fig.add_trace(go.Box(y=y4, name = 'Age 30-64 Repub % of population',marker_color = 'red'))
fig.add_trace(go.Box(y=y2, name = 'Age 65=< Dems % of population',marker_color = 'blue'))
fig.add_trace(go.Box(y=y5, name = 'Age 65=< Repub % of population',marker_color = 'darkred'))
fig.show()
```

In [99]:
```
#Camparision between race and ethinicity in both Democratic and Republican cou
nties
y0 = data_democratic['Percent White, not Hispanic or Latino']
y1 = data_democratic['Percent Black, not Hispanic or Latino']
y2 = data_democratic['Percent Hispanic or Latino']
y3 = data_republican['Percent White, not Hispanic or Latino']
y4 = data_republican['Percent Black, not Hispanic or Latino']
y5 = data_republican['Percent Hispanic or Latino']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Percent White (% of total pop) Dems',marker_c
olor = 'blue'))
fig.add_trace(go.Box(y=y1, name = 'Percent Black (% of total pop) Dems',marker
_color = 'skyblue'))
fig.add_trace(go.Box(y=y2, name = 'Percent Hispanic (% of total pop) Dems',mar
ker_color = 'darkblue'))
fig.add_trace(go.Box(y=y3, name='Percent White (% of total pop) Repubs',marker
_color = 'red'))
fig.add_trace(go.Box(y=y4, name = 'Percent Black (% of total pop) Repubs',mark
er_color = 'indianred'))
fig.add_trace(go.Box(y=y5, name = 'Percent Hispanic (% of total pop) Repubs',m
arker_color = 'darkred'))
fig.show()
```

In [100]:
```
#Camparing each race and ethinicity group from Democratic data with each race
 and ethinicity group of Republican data
y0 = data_democratic['Percent White, not Hispanic or Latino']
y1 = data_democratic['Percent Black, not Hispanic or Latino']
y2 = data_democratic['Percent Hispanic or Latino']
y3 = data_republican['Percent White, not Hispanic or Latino']
y4 = data_republican['Percent Black, not Hispanic or Latino']
y5 = data_republican['Percent Hispanic or Latino']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Percent White (% of total pop) Dems',
marker_color = 'blue'))
fig.add_trace(go.Box(y=y3, name='Percent White (% of total pop) Repubs',
marker_color = 'red'))
fig.add_trace(go.Box(y=y1, name = 'Percent Black (% of total pop) Dems',
marker_color = 'skyblue'))
fig.add_trace(go.Box(y=y4, name = 'Percent Black (% of total pop) Repubs',
marker_color = 'indianred'))
fig.add_trace(go.Box(y=y2, name = 'Percent Hispanic (% of total pop) Dems',
marker_color = 'darkblue'))
fig.add_trace(go.Box(y=y5, name = 'Percent Hispanic (% of total pop) Repubs',
marker_color = 'darkred'))
fig.show()
```

In [101]:

```
#Camparision between education level in both Democratic and Republican counties
y0 = data_democratic['Percent Less than High School Degree']
y1 = data_democratic['Percent Less than Bachelor\'s Degree']
y2 = 100 - data_democratic['Percent Less than Bachelor\'s Degree']
y3 = data_republican['Percent Less than High School Degree']
y4 = data_republican['Percent Less than Bachelor\'s Degree']
y5 = 100 - data_republican['Percent Less than Bachelor\'s Degree']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Percent < High School Dems',
marker_color = 'blue'))
fig.add_trace(go.Box(y=y1, name = 'Percent < Bachelor\'s Degree Dems',
marker_color = 'skyblue'))
fig.add_trace(go.Box(y=y2, name = 'Percent > Bachelor\'s Degree Dems',
marker_color = 'darkblue'))
fig.add_trace(go.Box(y=y3, name='Percent < High School Repubs',
marker_color = 'indianred'))
fig.add_trace(go.Box(y=y4, name = 'Percent < Bachelor\'s Degree Repubs',
marker_color = 'red'))
fig.add_trace(go.Box(y=y5, name = 'Percent > Bachelor\'s Degree Repubs',
marker_color = 'darkred'))
fig.show()
```

In [102]:
```
#Camparing each education level group from Democratic data with each education
level group of Republican data
y0 = data_democratic['Percent Less than High School Degree']
y1 = data_democratic['Percent Less than Bachelor\'s Degree']
y2 = 100 - data_democratic['Percent Less than Bachelor\'s Degree']
y3 = data_republican['Percent Less than High School Degree']
y4 = data_republican['Percent Less than Bachelor\'s Degree']
y5 = 100 - data_republican['Percent Less than Bachelor\'s Degree']
fig = go.Figure()
fig.add_trace(go.Box(y=y0, name='Percent < High School Dems',marker_color = 'b
lue'))
fig.add_trace(go.Box(y=y3, name='Percent < High School Repubs',marker_color =
'indianred'))
fig.add_trace(go.Box(y=y1, name = 'Percent < Bachelor\'s Degree Dems',marker_c
olor = 'skyblue'))
fig.add_trace(go.Box(y=y4, name = 'Percent < Bachelor\'s Degree Repubs',marker
_color = 'red'))
fig.add_trace(go.Box(y=y2, name = 'Percent > Bachelor\'s Degree Dems',marker_c
olor = 'darkblue'))
fig.add_trace(go.Box(y=y5, name = 'Percent > Bachelor\'s Degree Repubs',marker
_color = 'darkred'))
fig.show()
```

TASK 09 The important variables are the ‘Percent White, not Hispanic or Latino’, ‘Percent Less than Bachelor's Degree’.

Percent White, not Hispanic or Latino: The percent white population is more in republican counties than in democratic counties. (nearly 90% in republican and 77% in democrartic)

Percent Less than Bachelor's Degree: The median percentage of population less than bachelors degree is more in Republican counties than Democratics (nearly 83% in republic and nearly 70% in democratic) The median percentage of population greater than bachelors degree is more in democratics than republican(nearly 30% in democratic and nearly 18% in republican)

So, we consider these variables play a crucial role in deciding the win of the parties. So county with nearly 90% white can be considered as republican otherwise democratic.

On the basis of education, county with nearly 30% and more of population with > bachelors degree is considered as democratic otherwise republican

TASK 10

In [103]:
```
#Create a map of Democratic counties and Republican counties using the countie
s’ FIPS codes

change_values = {1: 'Democratic', 0: 'Republican'}
data_merged['Party'] = data_merged['Party'].map(change_values)
```

In [104]:
```
fips = data_merged['FIPS']

values = data_merged['Party']

fig = ff.create_choropleth(fips=fips, values=values, legend_title='Counties',title='Democratic and Republican Counties in the United States of America')
fig.layout.template = None
fig.show()
```

```
C:\Users\Bharath\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 [ ]: