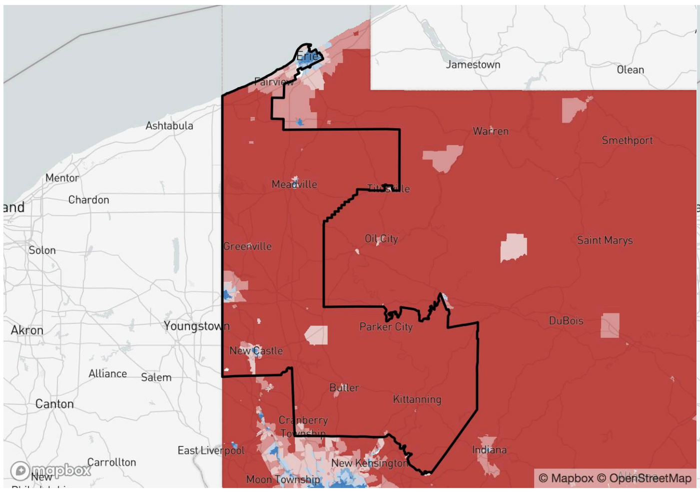
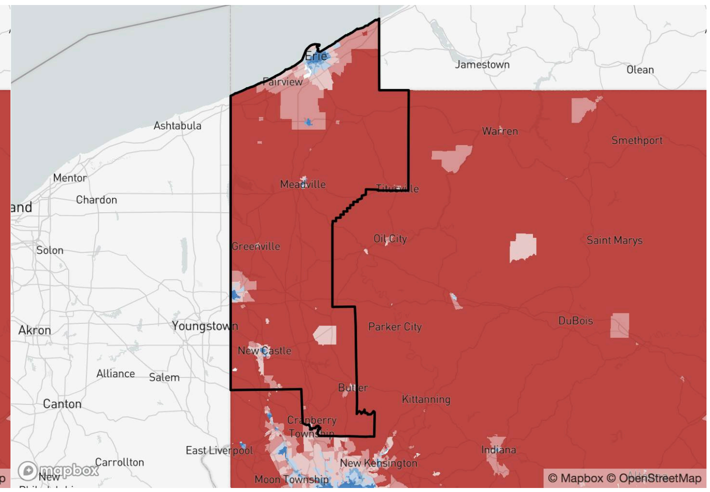
Pennsylvania's 16th congressional district (PA-16) located in north western Pennsylvania is a mostly rural district bordering the two major cities of Cleveland Ohio and Pittsburgh Pennsylvania. Pennsylvania is a majorly contested swing state for the 2024 election having voted for Democrat Barack Obama in 2008 and 2012, Republican Donald Trump in 2016, and Joe Biden in 2020. PA-16 is represented by republican Mike Kelly and has voted consistently Republican. PA-16 was redistricted in 2018 following a state Supreme Court decision redrawing the maps due to gerrymandering slightly modifying PA-03 to become PA-16 making the district more competitive.





Old District 3 New District 17

PA-16 also contains the town of Butler Pennsylvania where on July 13, 2024 a 20 year old gunman snuck an AR-15 style rifle into one of Former President and Republican presidential candidate Donald Trump's rally and shot the Former President. This event combined with the unpredictable nature of Pennsylvania's voting patterns, and the redistricting of Pennsylvania makes PA-16 a must watch district in the upcoming election cycle.

2021 ACS 1-Year Data Profiles provides detailed data on PA-16's social, economic, housing, and demographic characteristics. Some of the most interesting and surprising statistics our group found was PA-16's median income is only $60,000, 88.2% of PA-16 is white, 15.1% of PA-16 has jobs in manufacturing, 41.9% of PA-16 has public health coverage, and 29.3% having a bachelor's degree or higher . PA-16 is characterized as a mostly low income, white, rural area with jobs in low wage jobs manufacturing.

ACS's Profile on PA-16 is extremely detailed and provides lots of data that probably would not be useful in analyzing how PA-16 is going to be voting in the upcoming election. Our group picked 13 of what we thought would be the most interesting, useful, and statistics best representing PA-16 to consider.

District	White	Bachelor's degree or higher	Median household income (dollars)	Manufacturing	Educational services, and health care and social assistance	$200,000 or more	Mean household income (dollars)	With public coverage	Foreign born	English only	With a broadband Internet subscription	Management, business, science, and arts occupations	Worked from home
PA16	88.2	29.3	60630	15.1	25.8	5.5	80392	41.9	2.7	95.5	87.0	38.5	12.7

In order to identify similar districts to PA-16 we decided to look at other PA districts believing that looking at far away districts will not capture the unique swing nature that is PA. When we compare these 13 characteristics with all other PA districts, We find there are many wealthy urban PA districts that are unlike PA-16 such as PA-04, PA-05 or high immigrant districts like PA-03, PA-02, but there were also quite a few similar such as PA-13 or PA-14.

	District	White	Bachelor's degree or higher	Median household income (dollars)	Manufacturing	Educational services, and health care and social assistance	$200,000 or more	Mean household income (dollars)	With public coverage	Foreign born	English only	With a broadband Internet subscription	Management, business, science, and arts occupations	Worked from home
0	PA01	81.6	44.6	100136	13.1	25.3	17.0	128656	31.0	10.2	86.6	93.7	49.9	23.6
1	PA02	37.6	26.6	52293	8.0	31.6	5.4	74045	49.5	19.8	64.4	87.6	39.1	17.6
2	PA03	32.7	43.7	54392	5.8	34.2	7.6	81947	40.7	9.9	87.3	87.5	51.6	31.1
3	PA04	76.9	48.7	99271	13.6	25.2	18.1	133608	30.1	8.9	88.4	93.2	51.9	26.5
4	PA05	60.1	42.7	75243	8.6	28.9	13.9	111836	36.0	12.2	84.7	91.3	47.1	22.3
5	PA06	71.3	47.9	94356	12.7	21.8	17.3	126385	32.1	10.3	82.4	92.6	50.2	24.7
6	PA07	70.1	31.3	71407	13.6	25.2	8.4	95193	37.0	9.9	80.4	88.7	39.0	16.6
7	PA08	75.9	27.9	63058	11.1	26.5	4.8	78887	43.1	7.9	86.7	88.2	37.7	12.4
8	PA09	88.7	23.2	62659	15.2	24.0	4.7	79802	40.8	2.3	94.3	85.3	34.3	10.8
9	PA10	73.7	33.6	72359	10.2	23.6	6.3	91469	36.8	7.1	88.8	90.5	41.5	18.4
10	PA11	83.4	28.9	75875	15.9	23.0	7.9	95093	34.3	5.1	87.4	88.5	36.7	13.5
11	PA12	74.3	40.6	61514	8.0	30.2	7.7	86404	37.9	6.1	92.4	89.6	47.6	25.3
12	PA13	91.3	21.9	60754	13.3	25.5	4.3	78345	40.6	1.7	94.9	85.7	32.8	10.3
13	PA14	91.1	26.0	58075	12.0	24.1	4.5	77629	42.8	1.5	96.9	86.0	36.0	12.9
14	PA15	91.5	25.2	57945	15.8	28.7	4.6	75600	40.7	2.6	94.3	84.3	36.4	10.9
15	PA16	88.2	29.3	60630	15.1	25.8	5.5	80392	41.9	2.7	95.5	87.0	38.5	12.7
16	PA17	83.0	45.2	77984	8.6	26.2	11.1	105966	35.1	5.0	93.8	90.9	50.1	24.0

In order to fairly rank these 16 districts similarity to PA-16 we decided to score each vector cosine similarity. But first normalized the vectors using a min max scaler fixing the range on each category to be between [0,1]. This gave us a score between 0,1 for each district based on how similar it was to PA-16. Based on our results PA-9 and PA-14 were the two most similar districts to PA-16, having scored cosine similarity scores of .9752 and .9751. Indicating one of these districts is probably not more similar then the other

1 PA16 1.0
2 PA9 0.9752
3 PA14 0.9751
4 PA15 0.9702
5 PA13 0.9645
6 PA8 0.9498
7 PA11 0.9209
8 PA7 0.8788
9 PA10 0.8666
10 PA12 0.7825
11 PA17 0.7564
12 PA1 0.6778
13 PA5 0.6600
14 PA4 0.6583
15 PA6 0.6304
16 PA3 0.5231
17 PA2 0.4622

Looking back at our data it makes sense why these two districts scored highly. Both districts have median incomes only $3000 away from PA-16, PA-9 is 88.7% white 2.3% foreign born and 94.3% english only, which is very similar to PA14 which is 91% white 1.5% foreign born and 96.9% english only. With PA-16 being 88.2% white 2.7% foreign born and 95.5%

english only. PA-16, PA-9 and PA-14’s demographic, social, and economic characteristics are all very similar.

District	White	Bachelor's degree or higher	Median household income (dollars)	Manufacturing	Educational services, and health care and social assistance	$200,000 or more	Mean household income (dollars)	With public coverage	Foreign born	English only	With a broadband Internet subscription	Management, business, science, and arts occupations	Worked from home
PA16	88.2	29.3	60630	15.1	25.8	5.5	80392	41.9	2.7	95.5	87.0	38.5	12.7
PA09	88.7	23.2	62659	15.2	24.0	4.7	79802	40.8	2.3	94.3	85.3	34.3	10.8
PA14	91.1	26.0	58075	12.0	24.1	4.5	77629	42.8	1.5	96.9	86.0	36.0	12.9

Geographically PA-14 similarly borders Pittsburgh and Ohio. While PA-9 is on the eastern side bordering close to Philadelphia and New York. Both districts have historically voted republican.

links
https://ballotpedia.org/Pennsylvania%27s_16th_Congressional_District

https://www.nytimes.com/interactive/2018/02/19/upshot/pennsylvania-new-house-districts-gerrymandering.html

https://www.nytimes.com/article/shooting-trump-rally.html

https://www.census.gov/acs/www/data/congressional-and-state-legislative-districts/

In [1]:
```
import pandas as pd
dps = {}
for i in range(2,6):
    dp_temp = {}
    for x in range(1,18):
        filename = "DP%02d_PA%02d.xlsx" %(i,x)
        dp_temp["PA%02d" %x] = pd.read_excel(filename)
        dps["DP%02d" %i] = dp_temp
    print("Imported DP0"+str(i))
```

```
Imported DP02
Imported DP03
Imported DP04
Imported DP05
```

In [2]:
```
def getDataVector(table, row , col):
    row = row-2
#     print(dps[table]["PA01"]["Table ID: "+table][row])
    rtn = []
    for x in range(1,18):
        df = dps[table]["PA%02d" %x]
#         print("PA%02d" % x ,"\t",df["Unnamed: "+str(1)][row])
        rtn.append(df["Unnamed: "+str(col)][row])
    return dps[table]["PA01"]["Table ID: "+table][row] , rtn
```

In [3]:
```
vectors = []
things = [ ("DP05" , 52 , 3) , ("DP02" , 93 , 3) ,("DP03",84,1) , ("DP03",53,3) ,("
#            White                  HS              median income  Manufacturing
         ("DP03" , 83 , 3), ("DP03" , 85 , 1) , ("DP03" , 128 , 3),("DP02" , 130 ,
#           200+                 Meanincome              public health       foreign bo
        ("DP02" , 204 , 3) , ("DP03" , 43 , 3) , ("DP03" , 37 , 3)]
#               internet            job                 work from home
for table , row , col in things:
    vectors.append(getDataVector(table,row,col))


df = pd.DataFrame({
    "District" : ["PA%02d" %x for x in range(1,18)],
    **{vectors[x][0] : vectors[x][1] for x in range(0,len(vectors))}
})
df
```

Out[3]:

	District	White	Bachelor's degree or higher	Median household income (dollars)	Manufacturing	Educational services, and health care and social assistance	$200,000 or more	Me househo incor (dolla
0	PA01	81.6	44.6	100136	13.1	25.3	17.0	1286
1	PA02	37.6	26.6	52293	8.0	31.6	5.4	740
2	PA03	32.7	43.7	54392	5.8	34.2	7.6	819
3	PA04	76.9	48.7	99271	13.6	25.2	18.1	1336
4	PA05	60.1	42.7	75243	8.6	28.9	13.9	1118
5	PA06	71.3	47.9	94356	12.7	21.8	17.3	1263
6	PA07	70.1	31.3	71407	13.6	25.2	8.4	951
7	PA08	75.9	27.9	63058	11.1	26.5	4.8	788
8	PA09	88.7	23.2	62659	15.2	24.0	4.7	798
9	PA10	73.7	33.6	72359	10.2	23.6	6.3	914
10	PA11	83.4	28.9	75875	15.9	23.0	7.9	950
11	PA12	74.3	40.6	61514	8.0	30.2	7.7	864
12	PA13	91.3	21.9	60754	13.3	25.5	4.3	783
13	PA14	91.1	26.0	58075	12.0	24.1	4.5	776
14	PA15	91.5	25.2	57945	15.8	28.7	4.6	756
15	PA16	88.2	29.3	60630	15.1	25.8	5.5	803
16	PA17	83.0	45.2	77984	8.6	26.2	11.1	1059

In [4]:

```
import numpy as np
from sklearn.preprocessing import minmax_scale
from sklearn.metrics.pairwise import cosine_similarity
vector_scaled = np.array([minmax_scale(x[1]) for x in  vectors])
len(vector_scaled)
# for i in range(1,18):
#     print(cosine_similarity(vector_scaled[15]))
print(vector_scaled.T.shape)
Final_scores = []
for i in range(1,18):
    v = vector_scaled.T[i-1]
    Final_scores .append( ("PA" + str(i) , cosine_similarity(v.reshape(1, -1) , vec
# for v, in vector_scaled.T:
#     Final_scores .append( cosine_similarity(v.reshape(1, -1) , vector_scaled.T[15
Final_scores.sort(key = lambda x:-x[1])
print(Final_scores[0])
```

```
print(Final_scores[1])
print(Final_scores[2])
print(Final_scores[3])
```

```
(17, 13)
('PA16', 1.0)
('PA9', 0.9752702518856211)
('PA14', 0.9751573173135198)
('PA15', 0.9702141915411692)
```

In [5]:
```
[print(x,y) for x,y in Final_scores]
pass
```

```
PA16 1.0
PA9 0.9752702518856211
PA14 0.9751573173135198
PA15 0.9702141915411692
PA13 0.964508563094101
PA8 0.9498110836484239
PA11 0.9209997948393922
PA7 0.8788752760900211
PA10 0.866657660286038
PA12 0.7825011365202779
PA17 0.7564135871413133
PA1 0.6778928336232144
PA5 0.6600740668636202
PA4 0.6583414632706031
PA6 0.6304461161623047
PA3 0.5231973897839859
PA2 0.4622191969095474
```

In [6]:
```
df.iloc[[15,8,13]]
```

Out[6]:

	District	White	Bachelor's degree or higher	Median household income (dollars)	Manufacturing	Educational services, and health care and social assistance	$200,000 or more	Me househ inco (dolla
15	PA16	88.2	29.3	60630	15.1	25.8	5.5	803
8	PA09	88.7	23.2	62659	15.2	24.0	4.7	798
13	PA14	91.1	26.0	58075	12.0	24.1	4.5	776