Congressional Voting - EDA

Przewidywanie partii polityka na podstawie zapisów jego głosowań

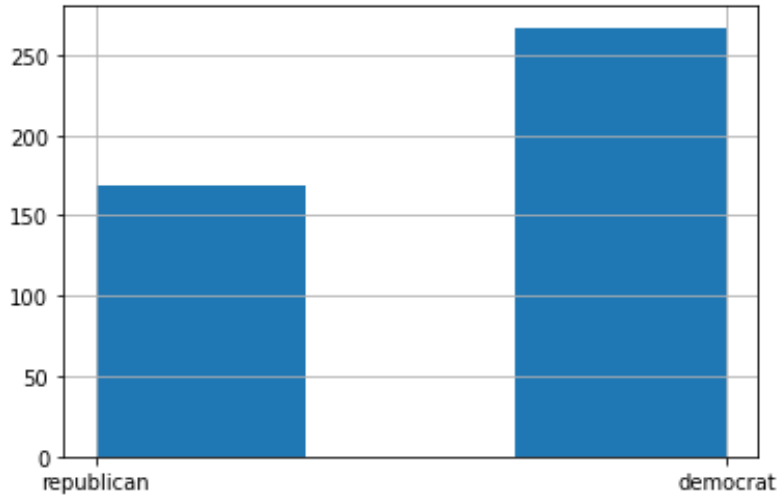
Autorzy: Katarzyna Solawa,
Jan Smoleń

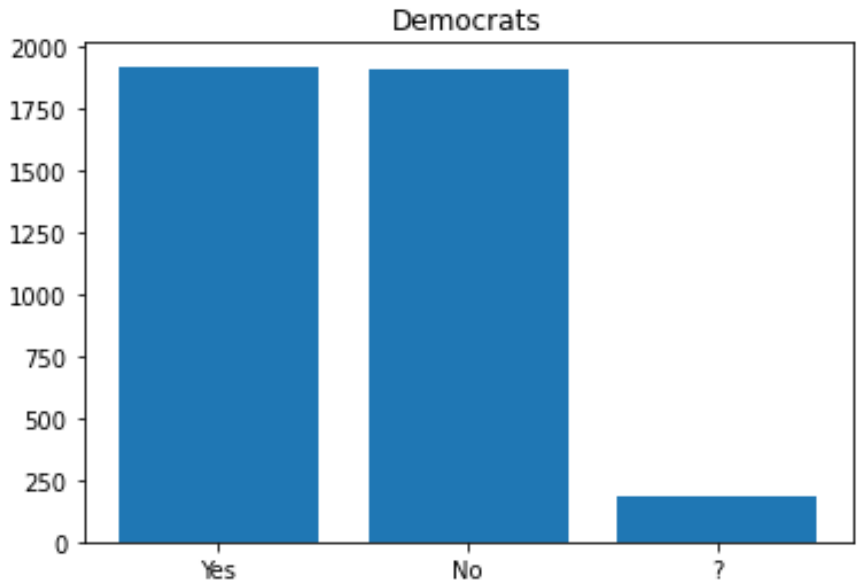
In [2]:
```
df=pd.read_csv("congressional_voting_dataset.csv")
```

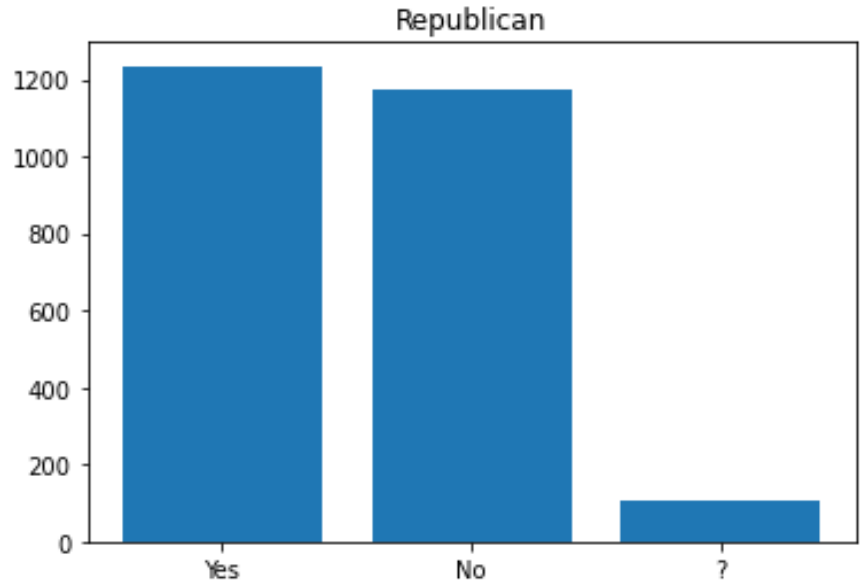
In [3]:
```
df.head()
```

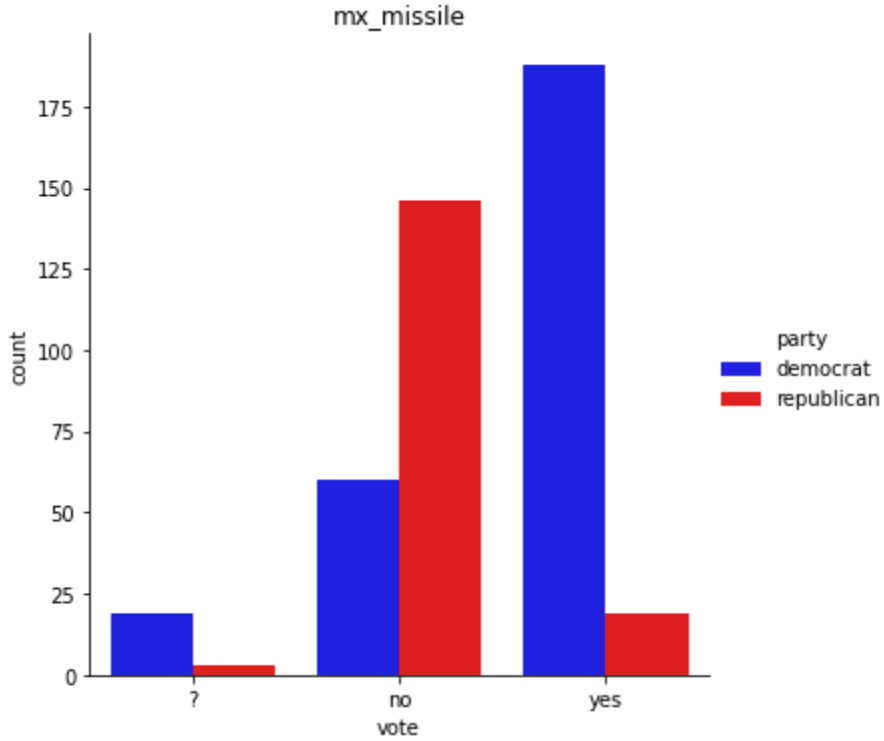
Out[3]:

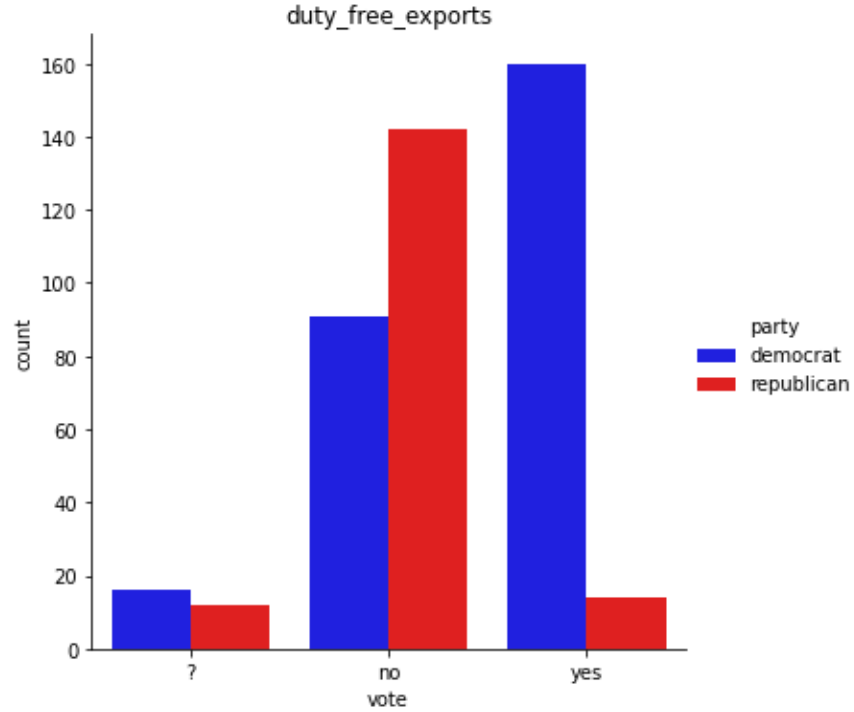
ation	synfuels_corporation_cutback	education_spending	superfund_right_to_sue	crime	duty_free_exports	export_administration_act_south_africa	political_party
y	?	y	y	y	n	y	republican
n	n	y	y	y	n	?	republican
n	y	n	y	y	n	n	democrat
n	y	n	y	n	n	y	democrat
n	y	?	y	y	y	y	democrat
...	...	...	...	...	...	...	...
y	n	y	y	y	n	y	republican
y	n	n	n	n	n	y	democrat
n	y	y	y	y	n	y	republican
?	n	y	y	y	n	y	republican
y	n	y	y	y	?	n	republican

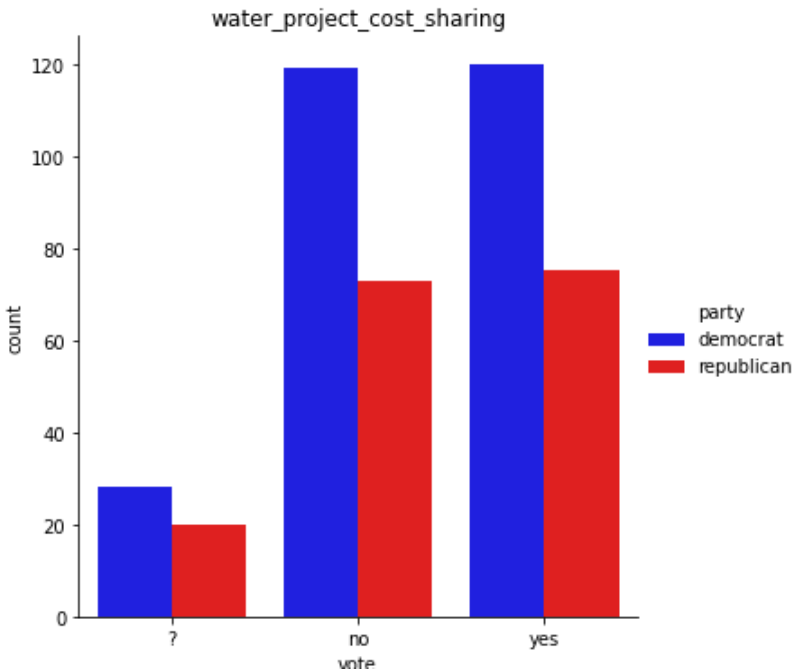

250
200
150
100
50
0
republican
democrat

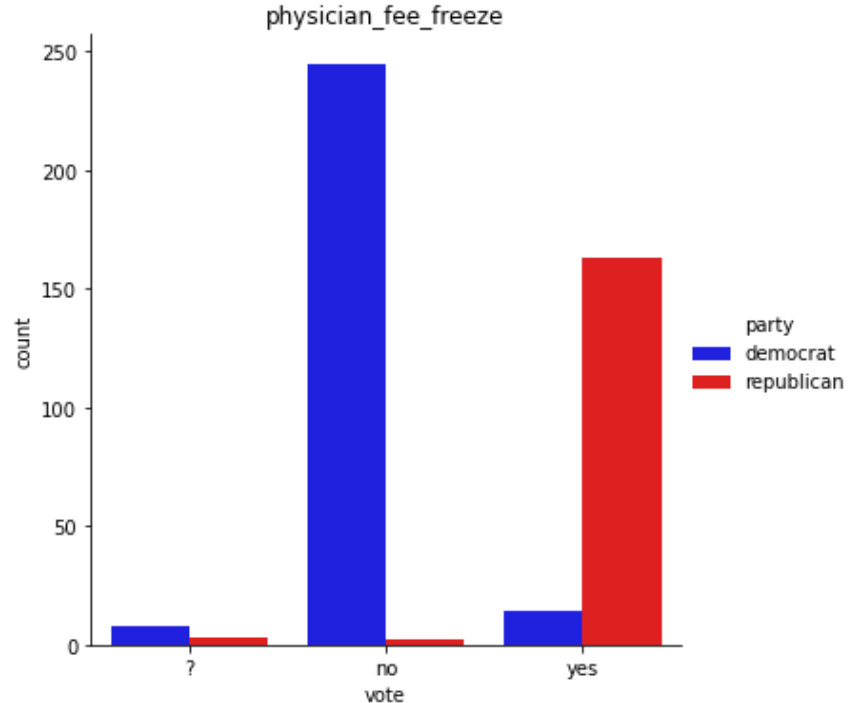

Democrats
2000
1750
1500
1250
1000
750
500
250
0
Yes
No
?

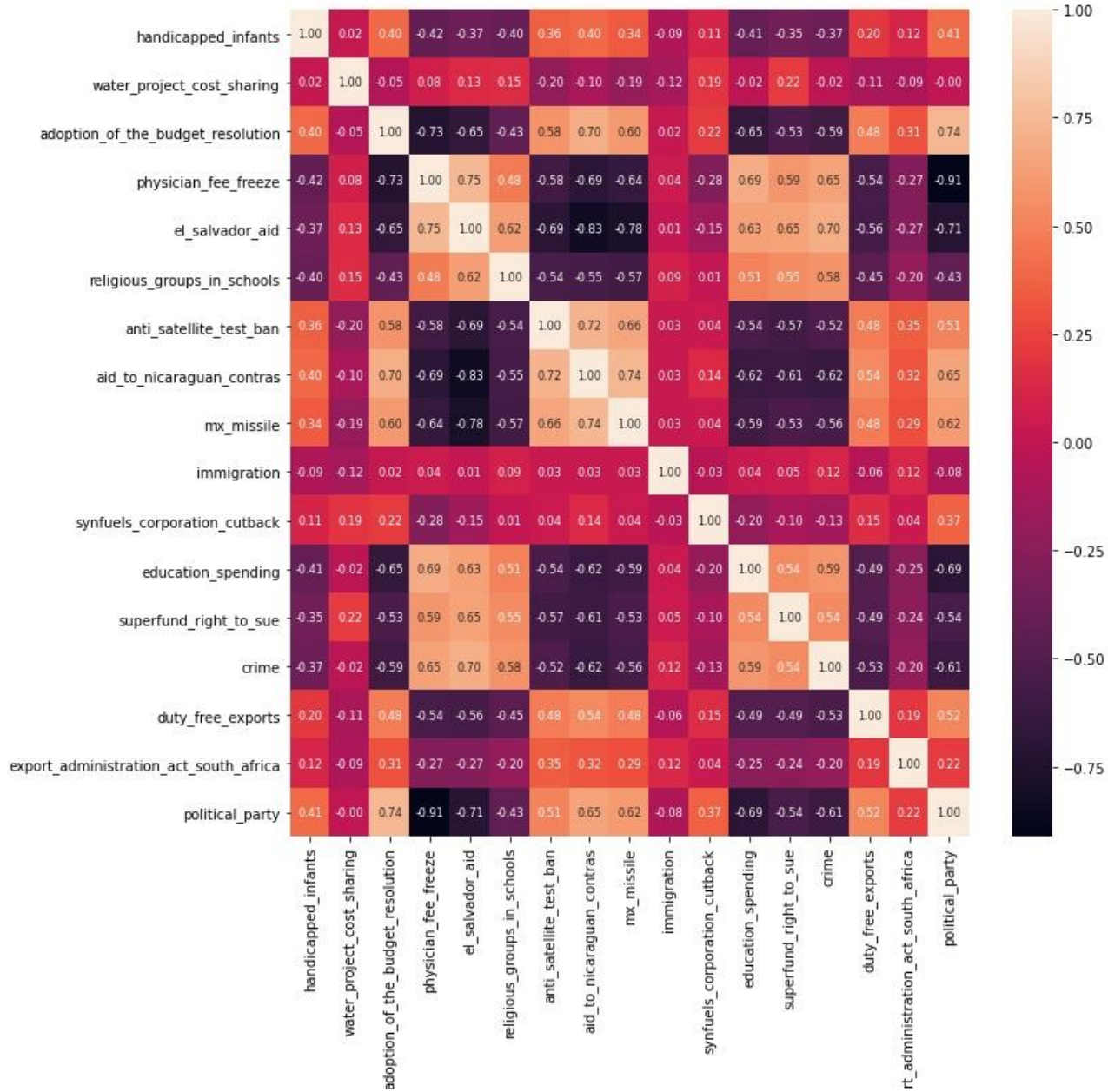

Republican
1200
1000
800
600
400
200
0
Yes
No
?


mx_missile
count
vote
?
no
yes
party
democrat
republican


duty_free_exports
count
vote
?
no
yes
party
democrat
republican


water_project_cost_sharing
count
vote
?
no
yes
party
democrat
republican

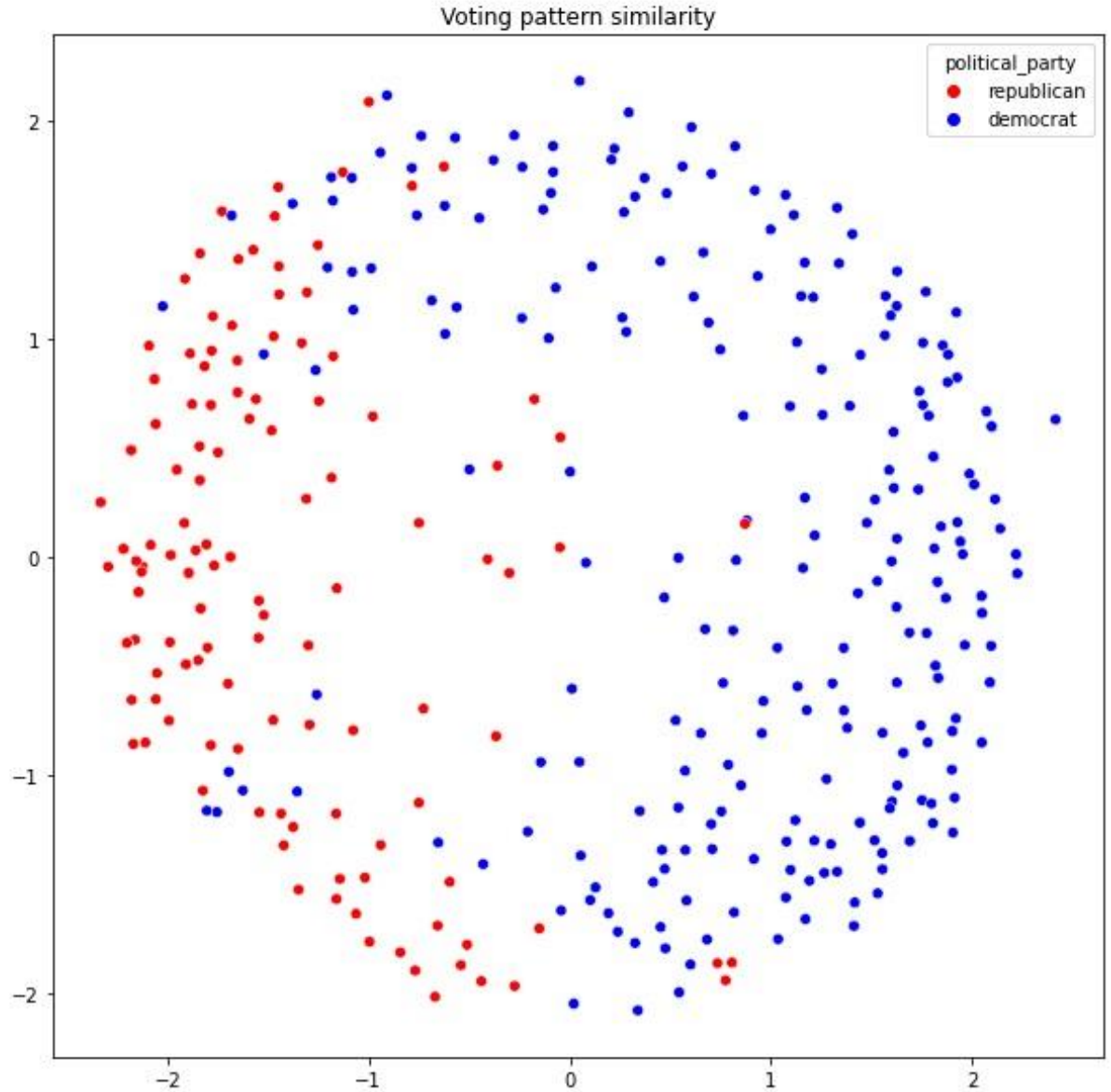

physician_fee_freeze
count
vote
?
no
yes
party
democrat
republican


handicapped_infants 1.00 0.02 0.40 -0.42 -0.37 -0.40 0.36 0.40 0.34 -0.09 0.11 -0.41 -0.35 -0.37 0.20 0.12 0.41
water_project_cost_sharing 0.02 1.00 -0.05 0.08 0.13 0.15 -0.20 -0.10 -0.19 -0.12 0.19 -0.02 0.22 -0.02 -0.11 -0.09 -0.00
adoption_of_the_budget_resolution 0.40 -0.05 1.00 -0.73 -0.65 -0.43 0.58 0.70 0.60 0.02 0.22 -0.65 -0.53 -0.59 0.48 0.31 0.74
physician_fee_freeze -0.42 0.08 -0.73 1.00 0.75 0.48 -0.58 -0.69 -0.64 0.04 -0.28 0.69 0.59 0.65 -0.54 -0.27 -0.91
el_salvador_aid -0.37 0.13 -0.65 0.75 1.00 0.62 -0.69 -0.83 -0.78 0.01 -0.15 0.63 0.65 0.70 -0.56 -0.27 -0.71
religious_groups_in_schools -0.40 0.15 -0.43 0.48 0.62 1.00 -0.54 -0.55 -0.57 0.09 0.01 0.51 0.55 0.58 -0.45 -0.20 -0.43
anti_satellite_test_ban 0.36 -0.20 0.58 -0.58 -0.69 -0.54 1.00 0.72 0.66 0.03 0.04 -0.54 -0.57 -0.52 0.48 0.35 0.51
aid_to_nicaraguan_contras 0.40 -0.10 0.70 -0.69 -0.83 -0.55 0.72 1.00 0.74 0.03 0.14 -0.62 -0.61 -0.62 0.54 0.32 0.65
mx_missile 0.34 -0.19 0.60 -0.64 -0.78 -0.57 0.66 0.74 1.00 0.03 0.04 -0.59 -0.53 -0.56 0.48 0.29 0.62
immigration -0.09 -0.12 0.02 0.04 0.01 0.09 0.03 0.03 0.03 1.00 -0.03 0.04 0.05 0.12 -0.06 0.12 -0.08
synfuels_corporation_cutback 0.11 0.19 0.22 -0.28 -0.15 0.01 0.04 0.14 0.04 -0.03 1.00 -0.20 -0.10 -0.13 0.15 0.04 0.37
education_spending -0.41 -0.02 -0.65 0.69 0.63 0.51 -0.54 -0.62 -0.59 0.04 -0.20 1.00 0.54 0.59 -0.49 -0.25 -0.69
superfund_right_to_sue -0.35 0.22 -0.53 0.59 0.65 0.55 -0.57 -0.61 -0.53 0.05 -0.10 0.54 1.00 0.54 -0.49 -0.24 -0.54
crime -0.37 -0.02 -0.59 0.65 0.70 0.58 -0.52 -0.62 -0.56 0.12 -0.13 0.59 0.54 1.00 -0.53 -0.20 -0.61
duty_free_exports 0.20 -0.11 0.48 -0.54 -0.56 -0.45 0.48 0.54 0.48 -0.06 0.15 -0.49 -0.49 -0.53 1.00 0.19 0.52
export_administration_act_south_africa 0.12 -0.09 0.31 -0.27 -0.27 -0.20 0.35 0.32 0.29 0.12 0.04 -0.25 -0.24 -0.20 0.19 1.00 0.22
political_party 0.41 -0.00 0.74 -0.91 -0.71 -0.43 0.51 0.65 0.62 -0.08 0.37 -0.69 -0.54 -0.61 0.52 0.22 1.00
handicapped_infants
water_project_cost_sharing
adoption_of_the_budget_resolution
physician_fee_freeze
el_salvador_aid
religious_groups_in_schools
anti_satellite_test_ban
aid_to_nicaraguan_contras
mx_missile
immigration
synfuels_corporation_cutback
education_spending
superfund_right_to_sue
crime
duty_free_exports
rt_administration_act_south_africa
political_party
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

In [41]:
```
adist=sklearn.metrics.pairwise_distances(df.drop(["political_party"], axis=1))
adist
```

Out[41]:
```
array([[0.        , 1.22474487, 2.17944947, ..., 1.22474487, 1.5       ,
        1.22474487],
       [1.22474487, 0.        , 1.93649167, ..., 1.22474487, 1.5       ,
        1.22474487],
       [2.17944947, 1.93649167, 0.        , ..., 1.93649167, 2.54950976,
        2.17944947],
       ...,
       [1.22474487, 1.22474487, 1.93649167, ..., 0.        , 1.5       ,
        1.87082869],
       [1.5       , 1.5       , 2.54950976, ..., 1.5       , 0.        ,
        1.80277564],
       [1.22474487, 1.22474487, 2.17944947, ..., 1.87082869, 1.80277564,
        0.        ]])
```

In [42]:
```
df["political_party"]=df["political_party"].replace(0, "republican")
df["political_party"]=df["political_party"].replace(1, "democrat")
adist=np.array(adist)
mds = manifold.MDS(n_components=2, dissimilarity="precomputed", random_state=6)
results = mds.fit(adist)
coords = results.embedding_
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(
    coords[:, 0], coords[:, 1], marker = 'o', hue=df["political_party"], palette=["red", "blue"]
    )
ax.set_title("Voting pattern similarity")
```


Voting pattern similarity
political_party
republican
democrat
2
1
0
−1
−2
−2
−1
0
1
2