

Data Processing: Impute using Redistricting

Group 42: Manish Vuyyuru, Victor Sheng, Elise Penn, Yajaira Gonzalez

Uses population overlap files generated in the overlap_area program to impute specified values into different years.

```
In [ ]: 1 import pandas as pd
        2 import pickle
        3 import shapely
```

```
In [ ]: 1 #to_correct = 'Datasets/data_FEC_NATIONALPOLL_2004_2018.p'
        2 to_correct = 'Datasets/data_FEC_NATIONALPOLL_DEMOGRAPHICSIMPUTED_2004_2018.p'
        3
        4 overlap = pickle.load(open('Datasets/all_overlap_data.p', 'rb'))
        5 dataset = pickle.load(open(to_correct, 'rb'))
        6 dataset.shape
```

```
In [ ]: 1 dataset[dataset['year'] == 2012].shape
```

```
In [ ]: 1 #drop redistricted districts
        2 new_dataset_name = to_correct[:-2] + '_REDISTRICTDROP.p'
        3 border_change = overlap['border_change']
        4 border_change = border_change[~border_change.isnull()]
        5 border_change = border_change[border_change == 1] #1: border changed, 0: border not
        6 for cand_index in dataset.index.values:
        7     if cand_index in border_change.index:
        8         dataset = dataset[dataset.index != cand_index]
        9 dataset.shape
```

```
In [ ]: 1 dataset[dataset['year'] == 2012].shape
```

```
In [ ]: 1 centroids = overlap['centroid']
        2 for cand_index in dataset.index.values:
        3     lon, lat = centroids[centroids.index == cand_index].values[0][0]
        4     dataset.loc[dataset.index == cand_index, 'longitude'] = lon
        5     dataset.loc[dataset.index == cand_index, 'latitude'] = lat
        6 dataset.shape, dataset.columns
```

```
In [ ]: 1 dataset[dataset['year']!=2018].isnull().values.any()
```

```
In [ ]: 1 pickle.dump(dataset, open(new_dataset_name, 'wb'))
```

```
In [ ]: 1
```