

```
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

import pandas as pd
confirm_dth = pd.read_csv('/content/drive/MyDrive/kaggle/confirmed-covid-19-deaths-in-us-by-state')
state = pd.read_csv('/content/drive/MyDrive/kaggle/covid-19-state-level-data copy.csv')
mobility = pd.read_csv('/content/drive/MyDrive/kaggle/us-mobility copy.csv')
```

confirm\_dth

```
display(confirm_dth.head())
```

→ Unnamed:  
0 county\_fips county\_name state\_name state\_fips date deaths lat long

0	0	0	Statewide Unallocated	AL	1	2020-01-22	0	NaN	NaN
1	1	0	Statewide Unallocated	AL	1	2020-01-23	0	NaN	NaN
2	2	0	Statewide Unallocated	AL	1	2020-01-24	0	NaN	NaN

```
print(confirm_dth[confirm_dth['county_fips']==0]['state_name'].describe())
print(confirm_dth[confirm_dth['county_fips']==0]['state_name'].value_counts())
```

```
count    9400
unique     50
top      AL
freq    188
Name: state_name, dtype: object
AL    188
PA    188
NV    188
NH    188
NJ    188
NM    188
NY    188
NC    188
ND    188
OH    188
OK    188
OR    188
RI    188
AK    188
SC    188
```

```
SD    188
TN    188
TX    188
UT    188
VT    188
VA    188
WA    188
WV    188
WI    188
NE    188
MT    188
MO    188
MS    188
AZ    188
AR    188
CA    188
CO    188
CT    188
DE    188
FL    188
GA    188
HI    188
ID    188
IL    188
IN    188
IA    188
KS    188
KY    188
LA    188
ME    188
MD    188
MA    188
MI    188
MN    188
WY    188
```

Name: state\_name, dtype: int64

```
print(confirm_dth['county_fips'].value_counts())
print(confirm_dth['county_fips'].describe())
print(confirm_dth['county_fips'].isnull().sum())
```

```
0        9400
39085    188
39089    188
39091    188
39093    188
...
21111    188
21113    188
21115    188
21117    188
56045    188
Name: county_fips, Length: 3146, dtype: int64
count    600660.000000
mean     29882.221283
```

```
std      15522.061935
min      0.000000
25%     18097.000000
50%     29123.000000
75%     45055.000000
max     56045.000000
Name: county_fips, dtype: float64
0
```

```
print(confirm_dth['county_name'].value_counts())
print(confirm_dth['county_name'].describe())
print(confirm_dth['county_name'].isnull().sum())
```

```
Statewide Unallocated    9400
Washington County       5640
Jefferson County         4700
Franklin County          4512
Lincoln County           4324
...
Androscoggin County     188
Winn Parish              188
West Feliciana Parish   188
West Carroll Parish      188
Weston County             188
Name: county_name, Length: 1882, dtype: int64
count                  600660
unique                 1882
top       Statewide Unallocated
freq                  9400
Name: county_name, dtype: object
0
```

```
print(confirm_dth['state_name'].value_counts())
print(confirm_dth['state_name'].describe())
print(confirm_dth['state_name'].isnull().sum())
```

```
TX      47940
GA      30080
VA      25192
KY      22748
MO      21808
KS      19928
IL      19364
NC      18988
IA      18800
TN      18048
NE      17672
IN      17484
OH      16732
MN      16544
MI      15792
MS      15604
OK      14664
AR      14288
```

```
WI      13724
PA      12784
AL      12784
FL      12784
SD      12596
CO      12220
LA      12220
NY      12032
CA      11280
MT      10716
WV      10528
ND      10152
SC      8836
ID      8460
WA      7520
OR      6956
NM      6392
AK      5828
UT      5640
MD      4700
WY      4512
NJ      4136
NV      3384
ME      3196
AZ      3008
VT      2820
MA      2820
NH      2068
CT      1692
RI      1128
HI      1128
DE      752
DC      188
Name: state_name, dtype: int64
count      600660
unique      51
top        TX
freq      47940
Name: state_name, dtype: object
0
```

```
print(confirm_dth['state_fips'].value_counts())
print(confirm_dth['state_fips'].describe())
print(confirm_dth['state_fips'].isnull().sum())
```

```
48      47940
13      30080
51      25192
21      22748
29      21808
20      19928
17      19364
37      18988
19      18800
47      18048
```

```
31    17672
18    17484
39    16732
27    16544
26    15792
28    15604
40    14664
5     14288
55    13724
42    12784
1     12784
12    12784
46    12596
8     12220
22    12220
36    12032
6     11280
30    10716
54    10528
38    10152
45    8836
16    8460
53    7520
41    6956
35    6392
2     5828
49    5640
24    4700
56    4512
34    4136
32    3384
23    3196
4     3008
50    2820
25    2820
33    2068
9     1692
44    1128
15    1128
10    752
11    188
```

```
Name: state_fips, dtype: int64
count    600660.000000
mean      30.250391
std       15.157627
min       1.000000
25%      18.000000
50%      30.000000
75%      45.000000
99%      85.000000
```

```
print(confirm_dth['long'].value_counts())
print(confirm_dth['long'].describe())
print(confirm_dth['long'].isnull().sum())
```

```
-86.644082      188
-82.481858      188
-82.150835      188
```

```
-83.654686    188
-83.401783    188
...
-84.582520    188
-82.826432    188
-84.532412    188
-82.949811    188
-104.567488   188
Name: long, Length: 3142, dtype: int64
count      590696.000000
mean       -92.268406
std        12.860422
min       -164.035380
25%       -98.238509
50%       -90.400244
75%       -83.436796
max       -67.628135
Name: long, dtype: float64
9964
```

```
print(confirm_dth['geometry'].value_counts())
print(confirm_dth['geometry'].describe())
print(confirm_dth['geometry'].isnull().sum())
```

```
POINT (-86.64408227 32.53952745)    188
POINT (-82.48185785 40.09136236)    188
POINT (-82.15083537 41.29553751)    188
POINT (-83.65468618 41.62101218)    188
POINT (-83.40178317 39.89381073)    188
...
POINT (-84.58252003 37.87119786)    188
POINT (-82.8264322 37.85106355)    188
POINT (-84.53241199 38.93950319)    188
POINT (-82.94981134 37.35388197)    188
POINT (-104.5674881 43.83961191)    188
Name: geometry, Length: 3142, dtype: int64
count                  590696
unique                 3142
top       POINT (-86.64408227 32.53952745)
freq                   188
Name: geometry, dtype: object
9964
```

```
confirm_dth.rename(columns={"date": "dates", "long": "longitude"}, inplace=True)
```

```
confirm_dth = confirm_dth[confirm_dth['county_fips'] != 0]
display(confirm_dth.head())
# confirm_dth.shape
```

	Unnamed: 0	county_fips	county_name	state_name	state_fips	dates	deaths
188	188	1001	Autauga County	AL	1	2020- 01-22	0 32
189	189	1001	Autauga County	AL	1	2020- 01-23	0 32

state

```
display(state.head())
```

	Unnamed: 0	date	state	fips	cases	deaths
0	0	2020-01-21	Washington	53	1	0
1	1	2020-01-22	Washington	53	1	0
2	2	2020-01-23	Washington	53	1	0
3	3	2020-01-24	Illinois	17	1	0
4	4	2020-01-24	Washington	53	1	0

```
print(state['state'].value_counts())
print(state['state'].describe())
print(state['state'].isnull().sum())
```

Washington	190
Illinois	187
California	186
Arizona	185
Massachusetts	179
Wisconsin	175
Texas	168
Nebraska	163
Utah	155
Oregon	152
New York	150
Rhode Island	150

```
Florida          150
Georgia          149
New Hampshire   149
North Carolina   148
New Jersey       147
Colorado          146
Maryland          146
Nevada           146
Tennessee         146
Minnesota         145
Pennsylvania      145
Oklahoma          145
South Carolina    145
Kentucky          145
Hawaii            145
Indiana           145
District of Columbia 144
Kansas             144
Missouri          144
Vermont           144
Virginia          144
Iowa               143
Connecticut        143
Louisiana         142
Ohio               142
Michigan          141
South Dakota      141
Wyoming            140
New Mexico         140
North Dakota      140
Delaware           140
Mississippi        140
Arkansas           140
Alaska              139
Maine              139
Alabama            138
Idaho              138
Montana            138
Puerto Rico        138
Virgin Islands     137
Guam               136
West Virginia      134
Northern Mariana Islands 123
Name: state, dtype: int64
count              8154
unique              55
```

```
print(state['fips'].value_counts())
print(state['fips'].describe())
print(state['fips'].isnull().sum())
```

```
53    190
17    187
6     186
4     185
```

```
25    179
55    175
48    168
31    163
49    155
41    152
36    150
44    150
12    150
13    149
33    149
37    148
34    147
8     146
24    146
32    146
47    146
27    145
42    145
40    145
45    145
21    145
15    145
18    145
11    144
20    144
29    144
50    144
51    144
19    143
9     143
22    142
39    142
26    141
46    141
56    140
35    140
38    140
10    140
28    140
5     140
2     139
23    139
1     138
16    138
30    138
72    138
78    137
66    136
54    134
69    123
Name: fips, dtype: int64
count    8154.000000
mean      21.761110
```

```
print(state['cases'].value_counts())
```

```
print(state['cases'].describe())
print(state['cases'].isnull().sum())

1      216
2      113
3      38
6      35
14     34
...
7352    1
12912   1
25676   1
83168   1
54114   1
Name: cases, Length: 6097, dtype: int64
count      8154.000000
mean       28152.917709
std        59724.328733
min        1.000000
25%       716.250000
50%       6169.000000
75%       28682.500000
max       474951.000000
Name: cases, dtype: float64
0
```

```
print(state['deaths'].value_counts())
print(state['deaths'].describe())
print(state['deaths'].isnull().sum())

0      956
6      233
2      210
1      163
17     93
...
918     1
1375    1
5555    1
2155    1
916     1
Name: deaths, Length: 2572, dtype: int64
count      8154.000000
mean       1355.825484
std        3730.562247
min        0.000000
25%       12.000000
50%       174.000000
75%       988.750000
max       32333.000000
Name: deaths, dtype: float64
0
```

## Mobility

```
display(mobility.head())
```

	state	county	date	retail	grocery_and_pharmacy	parks	transit_stations	wo
0	Total	Total	2020-02-15	6.0		2.0	15.0	3.0
1	Total	Total	2020-02-16	7.0		1.0	16.0	2.0
2	Total	Total	2020-02-17	6.0		0.0	28.0	-9.0

```
print(mobility['state'].value_counts())
print(mobility['state'].describe())
print(mobility['state'].isnull().sum())
```

Texas	11858
Georgia	8208
Virginia	7308
Kentucky	5808
Missouri	5617
North Carolina	5573
Illinois	5285
Indiana	5210
Iowa	5164
Tennessee	5153
Ohio	5047
Michigan	4505
Minnesota	4470
Mississippi	4245
Wisconsin	4001
Arkansas	3892
Florida	3837
Alabama	3825
Oklahoma	3760
Pennsylvania	3726
New York	3534
Louisiana	3520
Kansas	3371
California	3231
Nebraska	2689
Colorado	2682
South Carolina	2651
West Virginia	2544
Washington	2049
Idaho	2028
Oregon	1946
Montana	1691
New Mexico	1686
South Dakota	1535

```
Maryland          1425
Utah             1392
New Jersey       1254
Wyoming          1167
North Dakota    1162
Maine            969
Arizona          911
Massachusetts   855
Vermont          800
Nevada           763
New Hampshire   627
Alaska            604
Connecticut      513
Rhode Island     342
Hawaii            285
Delaware          228
District of Columbia  57
Total              57
Name: state, dtype: int64
count      155060
unique      52
top        Texas
freq       11858
```

```
mobility = mobility[mobility['state'] != 'Total']
```

```
print(mobility[mobility['county']=='Total'].shape)
print(mobility['county'].value_counts())
print(mobility['county'].describe())
print(mobility['county'].isnull().sum())
```

```
(2907, 9)
Total          2907
Washington County 1622
Jefferson County 1347
Franklin County 1248
Jackson County 1145
...
Antelope County  2
Wolfe County     1
Real County      1
Wirt County      1
Wheeler County   1
Name: county, Length: 1711, dtype: int64
count      155003
unique      1711
top        Total
freq       2907
Name: county, dtype: object
0
```

```
# mobility.replace('Total', None)
```

```
# import seaborn as sns
# sns.countplot(x="county", data=mobility)

#Mode of each county per state
temp = mobility.groupby(['state'])['county'].agg(lambda x: pd.Series.mode(x)[0]).to_frame()
temp.rename(columns={"state": "state_mode", "county": "county_mode"}, inplace=True)
temp
```

**county\_mode****state**

<b>Alabama</b>	Autauga County
<b>Alaska</b>	Anchorage
<b>Arizona</b>	Apache County
<b>Arkansas</b>	Arkansas County

```
mobility=mobility.merge(temp, how = 'inner', on = 'state')
```

<b>Colorado</b>	Adams County
-----------------	--------------

```
#Replaced value 'Total' with each states most common county
```

```
import numpy as np
```

```
mobility['county'] = np.where(mobility['county'] != 'Total', mobility['county'], mobility['co
```

```
mobility
```

	state	county	date	retail	grocery_and_pharmacy	parks	transit_stati
<b>0</b>	Alabama	Autauga County	2020-02-15	5.0		2.0	39.0
<b>1</b>	Alabama	Autauga County	2020-02-16	0.0		-2.0	-7.0
<b>2</b>	Alabama	Autauga County	2020-02-17	3.0		0.0	17.0
<b>3</b>	Alabama	Autauga County	2020-02-18	-4.0		-3.0	-11.0
<b>4</b>	Alabama	Autauga County	2020-02-19	4.0		1.0	6.0
...	...	...	...	...	...	...	...
<b>154998</b>	Wyoming	Weston County	2020-04-06	NaN		NaN	NaN
<b>154999</b>	Wyoming	Weston County	2020-04-07	NaN		NaN	NaN

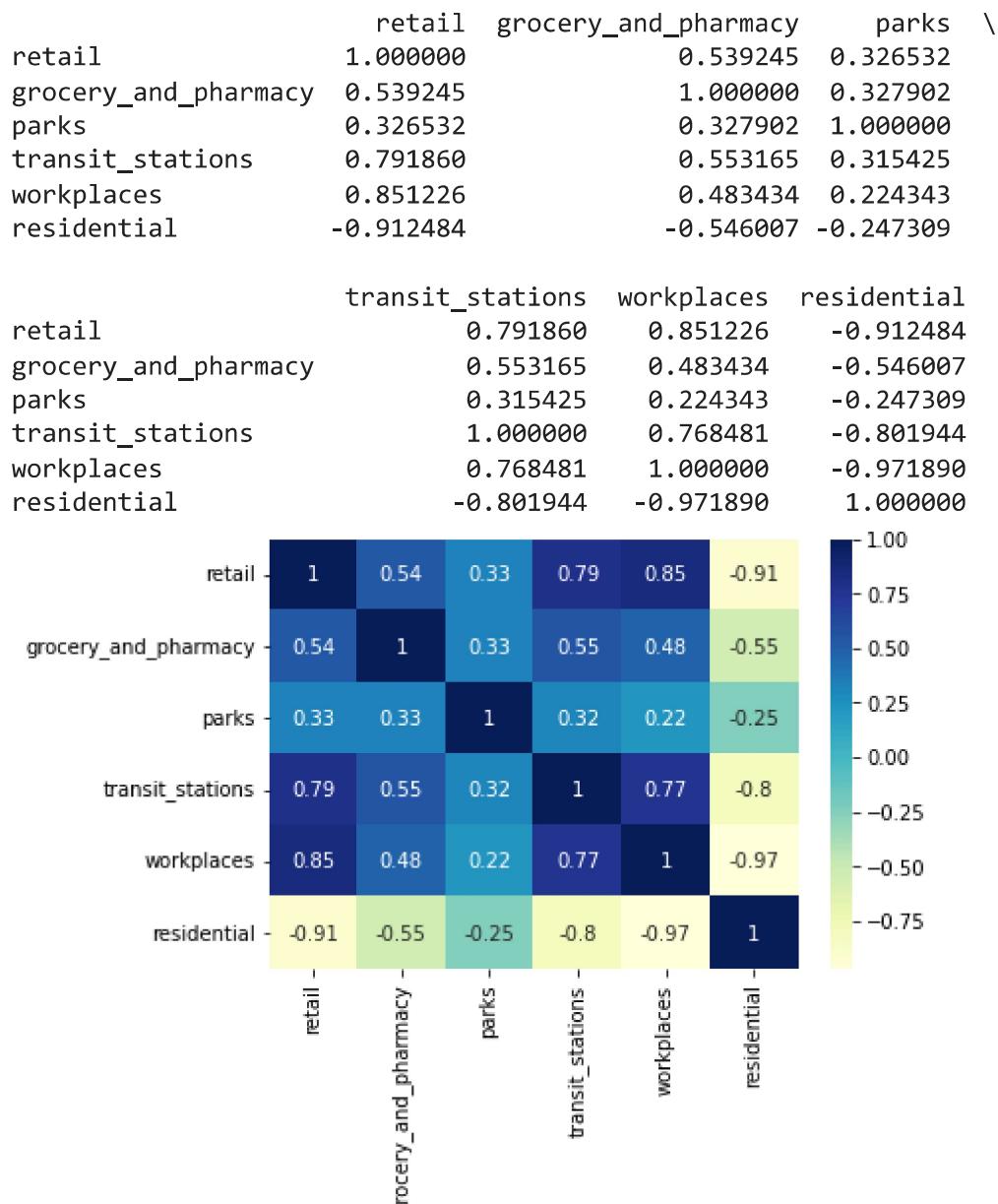
```

import matplotlib.pyplot as mp
import seaborn as sb
print(mobility.corr())

# plotting correlation heatmap
dataplot = sb.heatmap(mobility.corr(), cmap="YlGnBu", annot=True)

# displaying heatmap
mp.show()

```



```

print(mobility['retail'].value_counts())
print(mobility['retail'].describe())

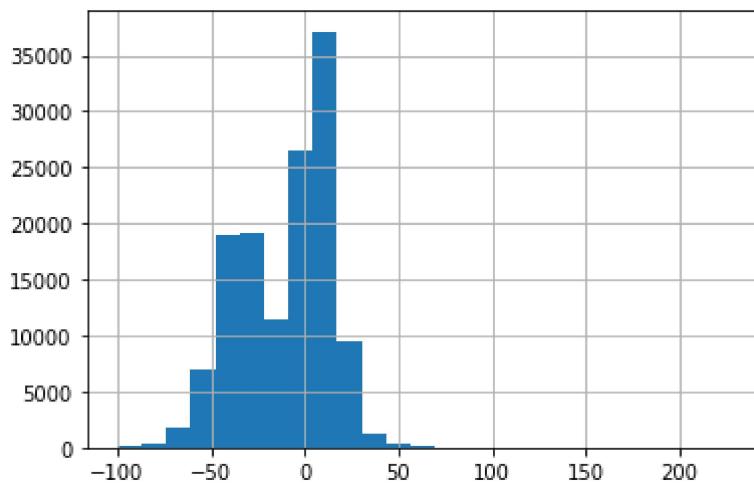
```

```
print(mobility['retail'].isnull().sum())
```

```
5.0      3638
6.0      3608
4.0      3565
7.0      3538
8.0      3511
...
92.0      1
90.0      1
151.0     1
-95.0     1
226.0     1
Name: retail, Length: 225, dtype: int64
count    133336.00000
mean     -10.089991
std      24.316575
min     -100.000000
25%     -31.000000
50%     -2.000000
75%      9.000000
max      226.000000
Name: retail, dtype: float64
21667
```

```
mobility['retail'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f172a0a63d0>
```



```
mobility['retail'].fillna(-10,inplace=True)#-10 is the mean
```

```
print(mobility['grocery_and_pharmacy'].value_counts())
print(mobility['grocery_and_pharmacy'].describe())
print(mobility['grocery_and_pharmacy'].isnull().sum())
```

```

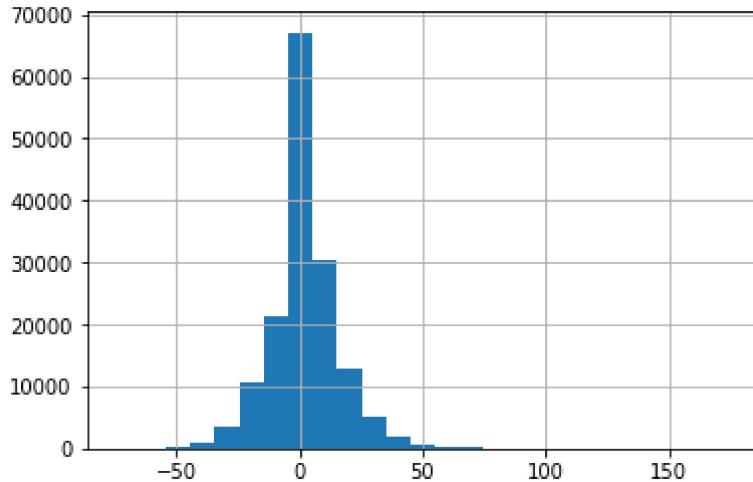
0.0      4666
2.0      4577
4.0      4540
3.0      4412
1.0      4251
...
-69.0     1
110.0     1
-71.0     1
132.0     1
120.0     1
Name: grocery_and_pharmacy, Length: 198, dtype: int64
count    128965.000000
mean      2.505649
std       15.369929
min      -74.000000
25%      -6.000000
50%      2.000000
75%      11.000000
max      174.000000
Name: grocery_and_pharmacy, dtype: float64
26038

```

```
mobility['grocery_and_pharmacy'].fillna(2.5,inplace=True)#2.5 is the mean
```

```
mobility['grocery_and_pharmacy'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f172a037590>
```



```

print(mobility['parks'].value_counts())
print(mobility['parks'].describe())
print(mobility['parks'].isnull().sum())

```

```

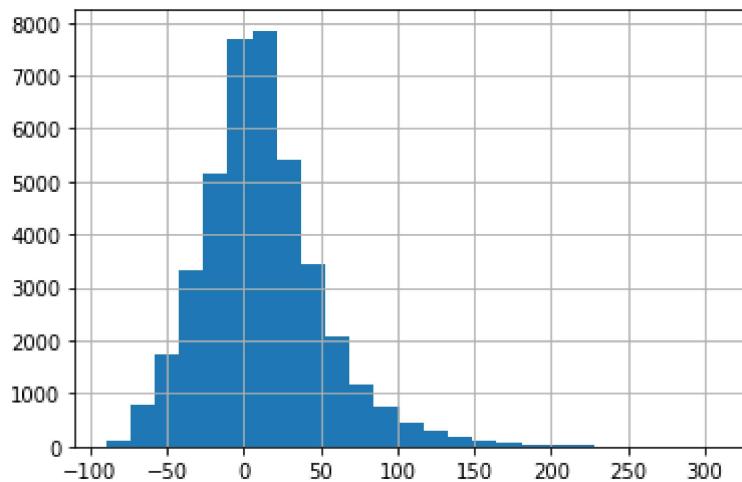
0.0      662
4.0      610
-2.0     586
6.0      581
12.0     567

```

```
...
220.0      1
196.0      1
192.0      1
233.0      1
308.0      1
Name: parks, Length: 313, dtype: int64
count    40554.000000
mean     11.886398
std      39.380488
min     -90.000000
25%     -13.000000
50%      8.000000
75%     32.000000
max     308.000000
Name: parks, dtype: float64
114449
```

```
mobility['parks'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1729f32250>
```



```
mobility['parks'].fillna(11.88,inplace=True)#11.88 is the mean
```

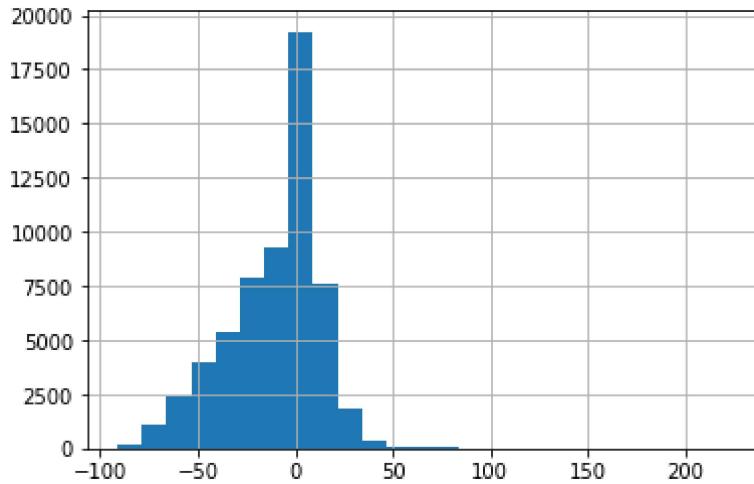
```
print(mobility['transit_stations'].value_counts())
print(mobility['transit_stations'].describe())
print(mobility['transit_stations'].isnull().sum())
```

```
2.0      1848
4.0      1681
0.0      1645
3.0      1643
5.0      1619
...
80.0      1
181.0      1
185.0      1
98.0      1
85.0      1
```

```
Name: transit_stations, Length: 189, dtype: int64
count    59489.000000
mean     -10.422465
std      23.646143
min     -91.000000
25%     -25.000000
50%     -4.000000
75%      6.000000
max     222.000000
Name: transit_stations, dtype: float64
95514
```

```
mobility['transit_stations'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1729eb5210>
```



```
mobility['transit_stations'].fillna(-10,inplace=True)#-10 is the mean
```

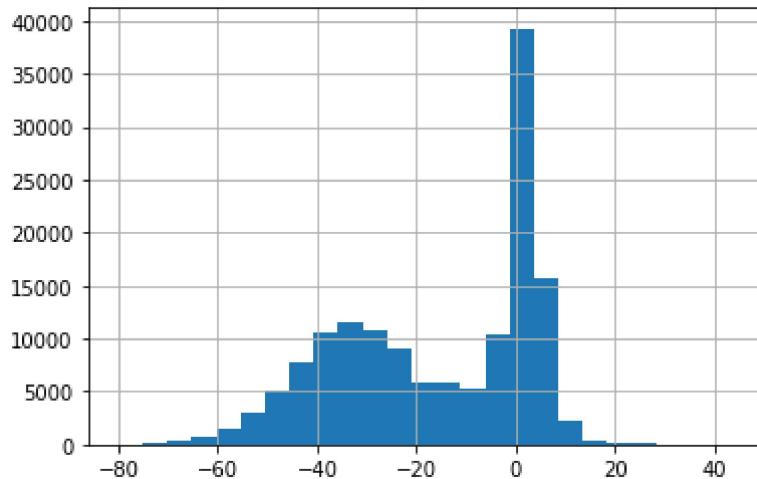
```
print(mobility['workplaces'].value_counts())
print(mobility['workplaces'].describe())
print(mobility['workplaces'].isnull().sum())
```

```
2.0      10425
1.0      8836
3.0      8507
0.0      7089
4.0      6194
...
38.0      1
33.0      1
32.0      1
42.0      1
37.0      1
Name: workplaces, Length: 122, dtype: int64
count    145974.000000
mean     -15.799451
std      19.308954
min     -80.000000
25%     -32.000000
```

```
50%      -11.000000
75%       2.000000
max       43.000000
Name: workplaces, dtype: float64
9029
```

```
mobility['workplaces'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1729dc57d0>
```



```
mobility['workplaces'].fillna(-15,inplace=True)#-15 is the mean
```

```
print(mobility['residential'].value_counts())
print(mobility['residential'].describe())
print(mobility['residential'].isnull().sum())
```

0.0	11125
-1.0	10827
1.0	5181
-2.0	4755
2.0	2577
13.0	2295
15.0	2265
14.0	2256
16.0	2150
12.0	2063
17.0	2054
11.0	1955
18.0	1845
3.0	1809
10.0	1695
19.0	1599
9.0	1439
4.0	1418
20.0	1403
-3.0	1364
5.0	1198
21.0	1191

```
8.0      1167
6.0      1015
7.0      983
22.0     952
23.0     774
24.0     565
25.0     447
26.0     336
27.0     279
-4.0     231
28.0     189
29.0     136
30.0     125
31.0     92
32.0     54
-5.0     53
33.0     45
35.0     27
34.0     26
36.0     16
-6.0     14
-7.0      7
37.0      2
38.0      2
Name: residential, dtype: int64
count    72001.000000
mean      6.706462
std       8.482861
min      -7.000000
25%      0.000000
50%      2.000000
75%      14.000000
max      38.000000
Name: residential, dtype: float64
83002
```

```
mobility['residential'].fillna(7,inplace=True)#7 is the mean
```

```
mobility['residential'].hist(bins=25)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f172ce39a90>
```

```
# confirm_dth=confirm_dth.drop(['Unnamed: 0'], axis = 1)
# state=state.drop(['Unnamed: 0'], axis = 1)
```

A horizontal number line starting at 0 and ending at 600,000. The line is divided into six major segments by vertical tick marks. The first segment ends at 100,000, the second at 200,000, the third at 300,000, the fourth at 400,000, the fifth at 500,000, and the sixth at 600,000. A solid blue rectangular bar is placed exactly on the tick mark for 500,000.

e-click (or enter) to edit

```
# confirm_dth.to_csv('confirmed-covid-19-deaths-in-us-by-state-and-county_cleaned.csv')
# state.to_csv('covid-19-state-level-data_cleaned.csv')
# mobility.to_csv('us-mobility_cleaned.csv')
```

$\sigma$  |  |  $\sigma$

```
confirm dth.head()
```

Unnamed: 0 county\_fips county\_name state\_name state\_fips dates deaths

**188** 188 1001 Autauga County AL 1 2020-01-22 0 32

**189** 189 1001 Autauga County AL 1 2020-01-23 0 32

◀ ▶

mobility.head()

	state	county	date	retail	grocery_and_pharmacy	parks	transit_stations
0	Alabama	Autauga County	2020-02-15	5.0		2.0	39.0
1	Alabama	Autauga County	2020-02-16	0.0		-2.0	-7.0

... 44 rows omitted

```
state.head()  
state.shape
```

```
(8154, 6)
```

