## Correlation Between Vaccination Rates and Covid Case, Hospitalization and Death Rates

Using data from *The New York Times'* online Covid tracking tool (https://www.nytimes.com/interactive/2021/us/covid-cases.html) from January 10, 2022, this notebook takes a high-level look at correlations between vaccination rates and case, hospitalization and death rates in the 50 U.S. states and Washington, D.C. It finds no meaningful correlation between vaccination rates and hospitalization and death rates, and a weak *positive* correlation with case rates, suggesting a higher case rate among the vaccinated. This analysis is subject to many caveats, including that it reflects just a single point in time and that the data is not adjusted for demographics and underlying health factors. But it is consistent with several recent studies finding little or no association between vaccination rates and Covid case, hospitalization and death rates.

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
pd.options.display.max_rows = 999
pd.set_option('display.max_columns', 500)
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

In [2]:
```
df=pd.read_excel("covid_vacc_correl_01_10_22.xls", index_col="Unnamed: 0")
df.head()
```

Out[2]:

| | Cases DAILY AVG. | Cases per 100000 | 14-day CHANGE | HOSPITALIZED DAILY AVG. | HOSPITALIZED per 100000 | HOSPITALIZED 14-day CHANGE | DEA D |
|---|---|---|---|---|---|---|---|
| **United States** | 677243 | 204 | 2.15 | 131361 | 40 | 0.82 | 15 |
| **Rhode Island ›** | 4371 | 413 | 2.16 | 435 | 41 | 0.58 | |
| **New York ›** | 74182 | 381 | 1.28 | 12082 | 62 | 1.18 | |
| **New Jersey ›** | 31512 | 355 | 1.25 | 5960 | 67 | 1.39 | |
| **Massachusetts ›** | 19902 | 289 | 1.78 | 2556 | 37 | 0.91 | |

In [3]:
```
df.columns = df.columns.str.strip().str.lower().str.replace(' ', '_').str.repl
ace('(', '').str.replace(')', '')
df.head()
```

Out[3]:

| | cases_daily_avg. | cases_per_100000 | 14-day_change | hospitalized_daily_avg. | hospit |
|---|---|---|---|---|---|
| **United States** | 677243 | 204 | 2.15 | 131361 | |
| **Rhode Island ›** | 4371 | 413 | 2.16 | 435 | |
| **New York ›** | 74182 | 381 | 1.28 | 12082 | |
| **New Jersey ›** | 31512 | 355 | 1.25 | 5960 | |
| **Massachusetts ›** | 19902 | 289 | 1.78 | 2556 | |

In [4]:
```
df.columns
```

Out[4]:
```
Index(['cases_daily_avg.', 'cases_per_100000', '14-day_change',
       'hospitalized_daily_avg.', 'hospitalized_per_100000',
       'hospitalized_14-day_change', 'deaths_daily_avg.', 'deaths_per_10000
0',
       'fully_vaccinated'],
      dtype='object')
```

In [5]:
```
df.head()
```

Out[5]:

| | cases_daily_avg. | cases_per_100000 | 14-day_change | hospitalized_daily_avg. | hospit |
|---|---|---|---|---|---|
| **United States** | 677243 | 204 | 2.15 | 131361 | |
| **Rhode Island ›** | 4371 | 413 | 2.16 | 435 | |
| **New York ›** | 74182 | 381 | 1.28 | 12082 | |
| **New Jersey ›** | 31512 | 355 | 1.25 | 5960 | |
| **Massachusetts ›** | 19902 | 289 | 1.78 | 2556 | |

In [6]:
```
df
```

Out[6]:

| | cases_daily_avg. | cases_per_100000 | 14-day_change | hospitalized_daily_avg. | hospit |
|---|---|---|---|---|---|
| **United States** | 677243 | 204 | 2.15 | 131361 | |
| **Rhode Island ›** | 4371 | 413 | 2.16 | 435 | |
| **New York ›** | 74182 | 381 | 1.28 | 12082 | |
| **New Jersey ›** | 31512 | 355 | 1.25 | 5960 | |
| **Massachusetts ›** | 19902 | 289 | 1.78 | 2556 | |
| **Washington, D.C. ›** | 1969 | 279 | 0.5 | 861 | |
| **Delaware ›** | 2663 | 274 | 1.59 | 600 | |
| **Florida ›** | 58336 | 272 | 2.27 | 8641 | |
| **Hawaii ›** | 3409 | 241 | 1.51 | 242 | |
| **Illinois ›** | 29077 | 229 | 1.34 | 6826 | |
| **Texas ›** | 65087 | 224 | 6.78 | 9765 | |
| **Connecticut ›** | 7940 | 223 | 1.72 | 1878 | |
| **Maryland ›** | 13392 | 222 | 1.13 | 3461 | |
| **Louisiana ›** | 9967 | 214 | 5.46 | 1414 | |
| **Pennsylvania ›** | 26858 | 210 | 1.93 | 7085 | |
| **South Carolina ›** | 10419 | 202 | 6.52 | 1569 | |
| **Alabama ›** | 9870 | 201 | 5.22 | 1617 | |
| **Arkansas ›** | 6053 | 201 | 5.26 | 927 | |
| **Missouri ›** | 12127 | 198 | 3.95 | 3111 | |
| **Kansas ›** | 5738 | 197 | 4.18 | 1053 | |
| **Vermont ›** | 1223 | 196 | 1.92 | 98 | |
| **Kentucky ›** | 8620 | 193 | 5.78 | 1936 | |
| **California ›** | 74709 | 189 | 4.05 | 9757 | |
| **Mississippi ›** | 5351 | 180 | 7.02 | 1001 | |
| **Utah ›** | 5630 | 176 | 4.61 | 550 | |
| **Virginia ›** | 14645 | 172 | 1.85 | 3339 | |
| **Arizona ›** | 12439 | 171 | 3.77 | 2708 | |
| **Georgia ›** | 18092 | 170 | 2.11 | 5073 | |
| **West Virginia ›** | 2982 | 166 | 1.78 | 806 | |
| **Ohio ›** | 19358 | 166 | 0.6 | 6976 | |
| **Tennessee ›** | 11193 | 164 | 2.9 | 2494 | |
| **Michigan ›** | 16048 | 161 | 1.46 | 4609 | |
| **Wisconsin ›** | 9343 | 160 | 1.23 | 2241 | |
| **Indiana ›** | 10323 | 153 | 1.41 | 3274 | |

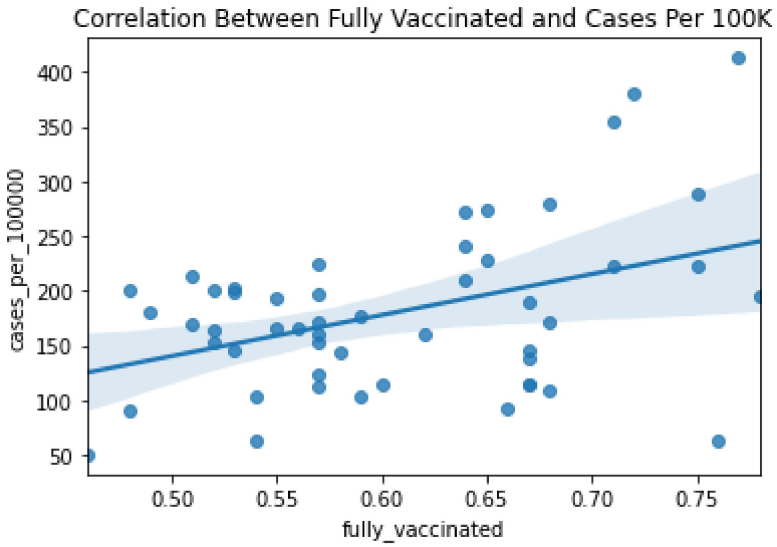
| | cases_daily_avg. | cases_per_100000 | 14-day_change | hospitalized_daily_avg. | hospit |
|---|---|---|---|---|---|
| **North Carolina ›** | 15925 | 152 | 3.19 | 3804 | |
| **New Hampshire ›** | 1988 | 146 | 0.67 | 394 | |
| **North Dakota ›** | 1114 | 146 | 2.91 | 226 | |
| **South Dakota ›** | 1276 | 144 | 2.69 | 296 | |
| **Colorado ›** | 7963 | 138 | 1.77 | 1447 | |
| **Alaska ›** | 903 | 123 | 5.04 | 75 | |
| **Nebraska ›** | 2209 | 114 | 1.85 | 595 | |
| **Oregon ›** | 4805 | 114 | 4.31 | 679 | |
| **New Mexico ›** | 2384 | 114 | 1.37 | 598 | |
| **Nevada ›** | 3464 | 112 | 2.4 | 1283 | |
| **Washington ›** | 8257 | 108 | 2.49 | 1726 | |
| **Oklahoma ›** | 4094 | 103 | 2.45 | 1255 | |
| **Iowa ›** | 3250 | 103 | 1.16 | 869 | |
| **Minnesota ›** | 5232 | 93 | 0.76 | 1585 | |
| **Wyoming ›** | 527 | 91 | 3.83 | 71 | |
| **Montana ›** | 676 | 63 | 3.26 | 156 | |
| **Maine ›** | 830 | 62 | –10% | 418 | |
| **Idaho ›** | 896 | 50 | 1.69 | 299 | |

In [7]:
```
df.drop(['United States'], axis=0, inplace=True)
```

In [8]:
```
corr_vac_and_cases=df['fully_vaccinated'].corr(df['cases_per_100000'])
print("The correlation between vaccination rate and case rate is:  ",corr_vac_
and_cases)
```

```
The correlation between vaccination rate and case rate is:   0.41749233614606
396
```

In [9]:
```
# use the function regplot to make a scatterplot
sns.regplot(y=df['cases_per_100000'], x=df['fully_vaccinated']).set_title("Cor
relation Between Fully Vaccinated and Cases Per 100K")
plt.show()
```



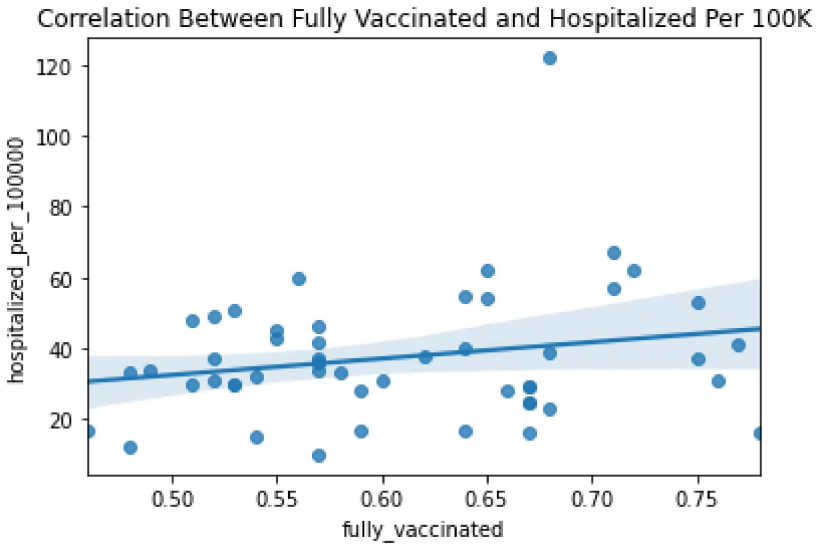

This means that there is a weak association between vaccination rate and *increased* number of cases.

Below, the data show no meaningful correlation between vaccination rates and hospitalization and death rates.

In [10]:
```
corr_vac_and_hospitalized=df['hospitalized_per_100000'].corr(df['fully_vaccina
ted'])
print("The correlation between vaccination rate and hospitalization rate is:
 ", corr_vac_and_hospitalized)
```

```
The correlation between vaccination rate and hospitalization rate is:   0.212
8437125507713
```

In [11]:
```
# use the function regplot to make a scatterplot
sns.regplot(y=df['hospitalized_per_100000'], x=df['fully_vaccinated']).set_tit
le("Correlation Between Fully Vaccinated and Hospitalized Per 100K")
plt.show()
```




In [12]:
```
corr_vac_and_deaths=df['deaths_per_100000'].corr(df['fully_vaccinated'])
print("The correlation between vaccination rate and death rate is:  ",corr_vac
_and_deaths)
```

```
The correlation between vaccination rate and death rate is:   0.0609696284148
4327
```

In [13]:
```
# use the function regplot to make a scatterplot
sns.regplot(y=df['deaths_per_100000'], x=df['fully_vaccinated']).set_title("Co
rrelation Between Fully Vaccinated and Deaths Per 100K")
plt.show()
```