# State Side: Income, the Heart, & Covid

By: Eddy Salgado, Jack Farnsworth, Nikki Ransom and Carla Lorente

# Problem:

How does the current makeup of the United States affect heart disease rates, covid cases, and covid deaths?

# Deep Dive:

## Questions:

- What region of the country has the highest rate of covid cases, covid deaths and heart disease?
- How did covid cases and covid deaths progress over time?
- How does political affiliation affect vaccination, covid deaths, and heart disease?
- What is the correlation between median income, vaccination rates and covid deaths?
- How did heart disease rates in the years previous to COVID  affect the number of covid deaths in 2020?

# Steps:

**Data:**

First we needed to gather our data sources and merge into a useable data frame.

**Analysis**

Using our data we began working through the analyses we wanted to perform to answer our question.

**Merge**

Once our analyses were ran we needed to merge to GitHub.

**Finalize**

As a team run through our notebook and code to see where we might have made mistakes and checked out everyone's analyses.

# Data Gathering and Cleaning

# Data Gathering

Combined 6 datasets, all by county

- COVID-19 cases and deaths (NY Times)
  - Cumulative over time
- Census data (CDC)
  - Population
  - Household Income
- Heart disease rate (CDC)
- Stroke rate (CDC)
- Vaccination rates (CDC)
  - Only considered two doses
- Election results (Harvard)
  - Assigned political party based on 2020 election

**Census Dataframe**

| | County | Name | Population | Median Age | Household Income |
|---|---|---|---|---|---|
| **0** | 173 | Sedgwick County, Kansas | 512064.0 | 35.2 | 54974.0 |
| **1** | 157 | Republic County, Kansas | 4686.0 | 51.1 | 48022.0 |
| **2** | 065 | Graham County, Kansas | 2545.0 | 51.9 | 40769.0 |
| **3** | 045 | Douglas County, Kansas | 119319.0 | 29.5 | 55832.0 |
| **4** | 179 | Sheridan County, Kansas | 2506.0 | 44.3 | 56071.0 |
| **...** | ... | ... | ... | ... | ... |
| **3215** | 003 | Adams County, Idaho | 4019.0 | 54.2 | 45319.0 |
| **3216** | 053 | Jerome County, Idaho | 23431.0 | 32.7 | 49306.0 |
| **3217** | 061 | Lewis County, Idaho | 3845.0 | 48.3 | 41326.0 |
| **3218** | 073 | Owyhee County, Idaho | 11455.0 | 38.4 | 40430.0 |
| **3219** | 021 | Boundary County, Idaho | 11549.0 | 43.5 | 43507.0 |

3220 rows × 5 columns

# Data Cleaning

- One of the challenges was figuring out how to merge the datasets into a single, clean one.
- We needed to make  a county name column that was consistent across our datasets.

```
# Defining function for cleaning up data from chronicdata.cdc.gov
def makehealthcsv(x, colname):

    # Filter
    x = x[(x['Stratification1'] == 'Overall') &  (x['Stratification2'] == 'Overall')]

    # Use state abbreviation dictionary and create county name column to be consistent with other dataframes
    x = x[x["LocationAbbr"].isin(us_state_abbrev)]
    x['county'] = x['LocationDesc'] + ', ' + x['LocationAbbr'].apply(lambda x: us_state_abbrev[x])
    x['county'] = x['county'].str.replace(' County', '').str.replace(' Parish', '')

    # Drop missing data, duplicates, and rename column of interest
    x = x[['county','Data_Value']].dropna()
    x.rename(columns = {'Data_Value' : colname}, inplace=True)
    x.drop_duplicates('county', inplace=True)
    x.set_index('county', inplace=True)

    return x
```

# Final Merge and Clean

```
# Concatenate above dataframes
merged_df = pd.concat([covid_total, vaccine, census, heart_disease, stroke, party_df], join='inner', axis=1)
merged_df.dropna(inplace=True)
merged_df.reset_index(inplace=True)

# Split county and state name into separate columns for future analysis
merged_df[['County','State']] = merged_df.county.str.split(", ",expand=True)
merged_df.drop(columns=['county'], inplace=True)

# New columns for cases and death by capita
merged_df['Cases per Capita'] = merged_df['cases']/merged_df['Population']
merged_df['Deaths per Capita'] = merged_df['deaths']/merged_df['Population']

# Rename columns and save to csv
merged_df.rename(columns = {'cases' : 'Cases', 'deaths' : 'Deaths', 'party': 'Party'}, inplace=True)
merged_df.to_csv('resources/full_data.csv', index=False)

# Filtering out very small and large counties by population and save in separate csv
merged_df.drop(merged_df.loc[merged_df['Population'] < 20000].index, inplace = True)
merged_df.drop(merged_df.loc[merged_df['Population'] > 200000].index, inplace = True)
merged_df.to_csv('resources/data.csv', index=False)

merged_df
```

# Final Dataframe

| | Cases | Deaths | Vaccination Rate | Population | Median Age | Household Income | Heart Disease | Stroke | Party | County | State | Cases per Capita | Deaths per Capita |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 6104 | 57.0 | 39.2 | 24657.0 | 43.8 | 36685.0 | 321.0 | 78.6 | REPUBLICAN | Abbeville | South Carolina | 0.247556 | 0.002312 |
| **1** | 14951 | 269.0 | 51.8 | 62568.0 | 36.2 | 41177.0 | 476.2 | 93.6 | REPUBLICAN | Acadia | Louisiana | 0.238956 | 0.004299 |
| **2** | 6569 | 90.0 | 70.7 | 32742.0 | 45.9 | 43210.0 | 411.2 | 92.1 | REPUBLICAN | Accomack | Virginia | 0.200629 | 0.002749 |
| **6** | 5535 | 56.0 | 44.0 | 25325.0 | 27.7 | 40046.0 | 335.1 | 80.7 | REPUBLICAN | Adair | Missouri | 0.218559 | 0.002211 |
| **7** | 6678 | 65.0 | 34.8 | 22113.0 | 37.6 | 32986.0 | 575.0 | 62.7 | REPUBLICAN | Adair | Oklahoma | 0.301994 | 0.002939 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2983** | 7537 | 95.0 | 47.5 | 27974.0 | 35.7 | 31402.0 | 368.6 | 119.6 | DEMOCRAT | Yazoo | Mississippi | 0.269429 | 0.003396 |
| **2984** | 6386 | 94.0 | 47.6 | 21573.0 | 39.3 | 42361.0 | 472.9 | 86.2 | REPUBLICAN | Yell | Arkansas | 0.296018 | 0.004357 |
| **2986** | 38255 | 486.0 | 50.5 | 157816.0 | 38.2 | 59117.0 | 322.4 | 70.9 | REPUBLICAN | Yellowstone | Montana | 0.242403 | 0.003080 |
| **2993** | 9037 | 88.0 | 65.3 | 67587.0 | 39.5 | 90367.0 | 233.0 | 72.0 | REPUBLICAN | York | Virginia | 0.133709 | 0.001302 |
| **2995** | 15811 | 104.0 | 51.1 | 75493.0 | 32.5 | 52624.0 | 370.1 | 87.2 | REPUBLICAN | Yuba | California | 0.209437 | 0.001378 |

1446 rows × 13 columns

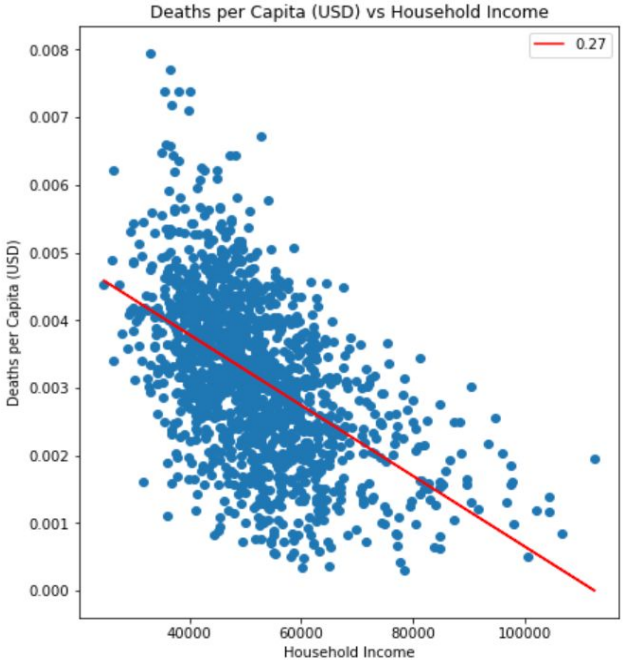
# Main Correlations

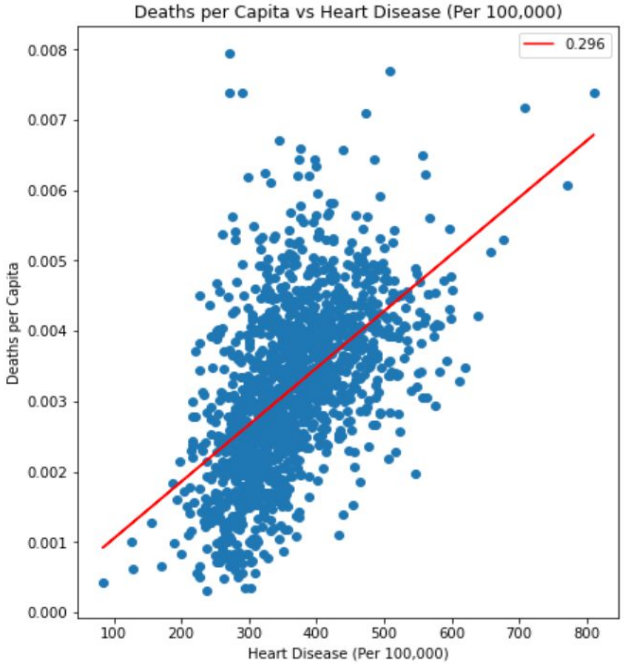
# Correlation Matrix

| | Cases | Deaths | Vaccination Rate | Population | Median Age | Household Income | Heart Disease | Stroke | Cases per Capita | Deaths per Capita |
|---|---|---|---|---|---|---|---|---|---|---|
| **Cases** | 1.000000 | 0.822636 | 0.207994 | 0.950363 | -0.238787 | 0.270933 | -0.155044 | -0.065259 | 0.257009 | -0.104763 |
| **Deaths** | 0.822636 | 1.000000 | 0.049213 | 0.799732 | -0.084372 | 0.026746 | 0.073631 | 0.085283 | 0.175210 | 0.317244 |
| **Vaccination Rate** | 0.207994 | 0.049213 | 1.000000 | 0.258447 | 0.140084 | 0.348115 | -0.449743 | -0.346423 | -0.124366 | -0.399787 |
| **Population** | 0.950363 | 0.799732 | 0.258447 | 1.000000 | -0.172629 | 0.332495 | -0.207696 | -0.108684 | 0.003292 | -0.198430 |
| **Median Age** | -0.238787 | -0.084372 | 0.140084 | -0.172629 | 1.000000 | -0.027420 | -0.075867 | -0.104189 | -0.299938 | 0.079298 |
| **Household Income** | 0.270933 | 0.026746 | 0.348115 | 0.332495 | -0.027420 | 1.000000 | -0.534294 | -0.439737 | -0.135538 | -0.519665 |
| **Heart Disease** | -0.155044 | 0.073631 | -0.449743 | -0.207696 | -0.075867 | -0.534294 | 1.000000 | 0.516215 | 0.197634 | 0.544322 |
| **Stroke** | -0.065259 | 0.085283 | -0.346423 | -0.108684 | -0.104189 | -0.439737 | 0.516215 | 1.000000 | 0.144171 | 0.400362 |
| **Cases per Capita** | 0.257009 | 0.175210 | -0.124366 | 0.003292 | -0.299938 | -0.135538 | 0.197634 | 0.144171 | 1.000000 | 0.335218 |
| **Deaths per Capita** | -0.104763 | 0.317244 | -0.399787 | -0.198430 | 0.079298 | -0.519665 | 0.544322 | 0.400362 | 0.335218 | 1.000000 |

# Effects of Income and Health

Household income and heart disease rates are two of the strongest factors related to COVID-19 deaths per capita.



# Data Analysis by Region within US

# Covid Cases by Region

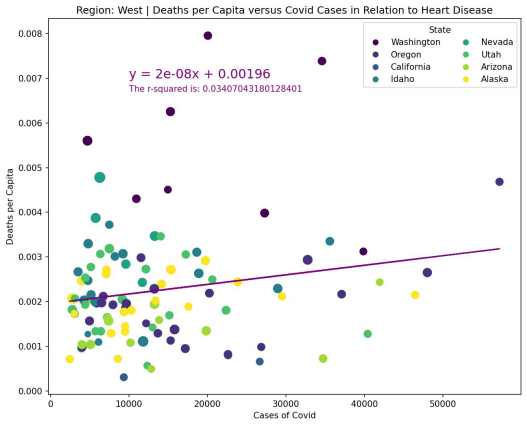
```
#Sorting through states and relegating to correct region
#West Region
West_df = merged_df.loc[(merged_df["State"]=="Washington") | (merged_df["State"]=="Oregon")
                        |(merged_df["State"]=="California") | (merged_df["State"]=="Idaho")
                        | (merged_df["State"]=="Nevada") | (merged_df["State"]=="Utah")
                        | (merged_df["State"]=="Arizona") | (merged_df["State"]=="Alaska")
                        | (merged_df["State"]=="Hawaii"),:]

#Plains Region
Plains_df = merged_df.loc[(merged_df["State"]=="Montana") | (merged_df["State"]=="North Dakota")
                        |(merged_df["State"]=="South Dakota") | (merged_df["State"]=="Wyoming")
                        | (merged_df["State"]=="Nebraska") | (merged_df["State"]=="Colorado")
                        | (merged_df["State"]=="Oklahoma") | (merged_df["State"]=="Kansas")
                        | (merged_df["State"]=="Texas") | (merged_df["State"]=="New Mexico"),:]

#Midwest Region
Midwest_df = merged_df.loc[(merged_df["State"]=="Minnesota") | (merged_df["State"]=="Wisconsin")
                        |(merged_df["State"]=="Michigan") | (merged_df["State"]=="Iowa")
                        | (merged_df["State"]=="Illinois") | (merged_df["State"]=="Indiana")
                        | (merged_df["State"]=="Ohio") | (merged_df["State"]=="Missouri")
                        | (merged_df["State"]=="Kentucky"),:]

#Southeast Region
Southeast_df = merged_df.loc[(merged_df["State"]=="Arkansas") | (merged_df["State"]=="Louisiana")
                        |(merged_df["State"]=="Tennessee") | (merged_df["State"]=="Mississippi")
                        | (merged_df["State"]=="Alabama") | (merged_df["State"]=="Georgia")
                        | (merged_df["State"]=="Florida") | (merged_df["State"]=="North Carolina")
                        | (merged_df["State"]=="South Carolina"),:]
#Northeast Region of US
Northeast_df = merged_df.loc[(merged_df["State"]=="Maine") | (merged_df["State"]=="New Hampshire")
                        |(merged_df["State"]=="Vermont") | (merged_df["State"]=="New York")
                        | (merged_df["State"]=="Massachusetts") | (merged_df["State"]=="Connecticut")
                        | (merged_df["State"]=="Rhode Island") | (merged_df["State"]=="New Jersey")
                        | (merged_df["State"]=="Pennsylvania") | (merged_df["State"]=="Delaware")
                        | (merged_df["State"]=="West Virginia") | (merged_df["State"]=="Virginia")
                        | (merged_df["State"]=="Maryland"),:]
```

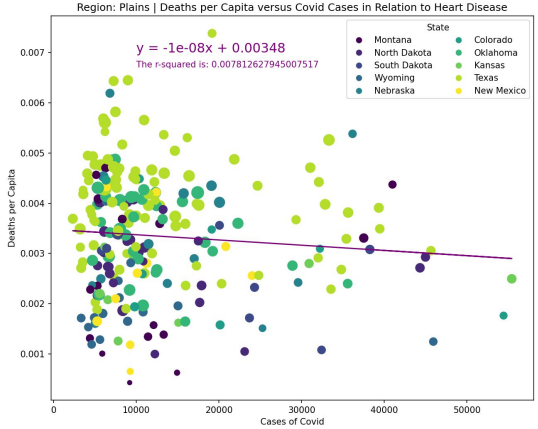
# Covid Cases by Region

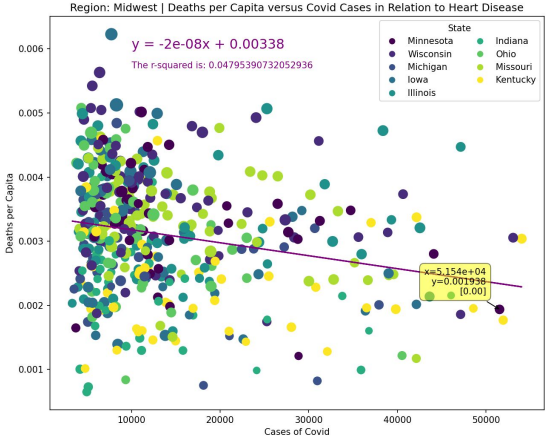
```
title = ["West", "Plains", "Midwest", "Southeast", "Northeast"]
State_list = [West_df,Plains_df,Midwest_df,Southeast_df,Northeast_df]
labels = [West, Plains, Midwest, Southeast, Northeast]
holder = 0

for index in range(len(State_list)):
    plt.figure(figsize=(8,6))
    plt.xlabel("Cases of Covid")
    plt.ylabel("Deaths per Capita")
    (slope, intercept, rvalue, pvalue, stderr) = linregress(State_list[index]["Cases"], State_list[inde
    regress_values = State_list[index]["Cases"] * slope + intercept
    line_eq = "y = " + str(round(slope,8)) + "x + " + str(round(intercept,5))
    Scatter2 = plt.scatter(State_list[index]["Cases"],State_list[index]["Deaths per Capita"],s=State_li
            c=State_list[index].State.astype('category').cat.codes)
    plt.plot(State_list[index]["Cases"],regress_values,color = "purple")
    plt.annotate(line_eq,(10000,.006),fontsize=15,color = "purple")
    plt.annotate(f"The r-squared is: {rvalue**2}",(10000,.0057),fontsize=10,color = "purple")
    #Hover Event
    mplcursors.cursor(Scatter2, hover=True)
    plt.legend(loc="upper right", ncol= 2, handles=Scatter2.legend_elements()[0],
           labels=labels[index],
           title="State")
    plt.title (f"Region:  {title[holder]} | Deaths per Capita versus Covid Cases in Relation to Heart D
    plt.savefig(f'output_images/cases_vs_death-{title[holder]}.png')
    holder = holder + 1
```

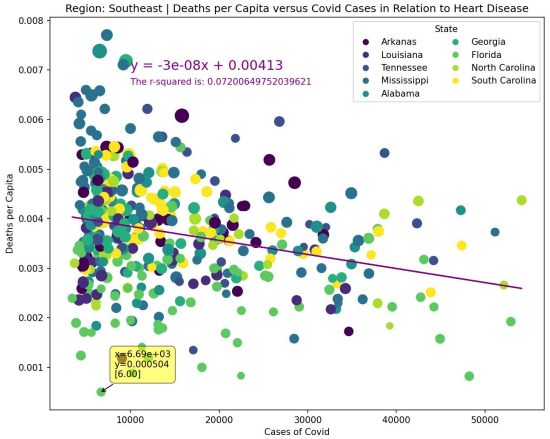
# Regional Correlation

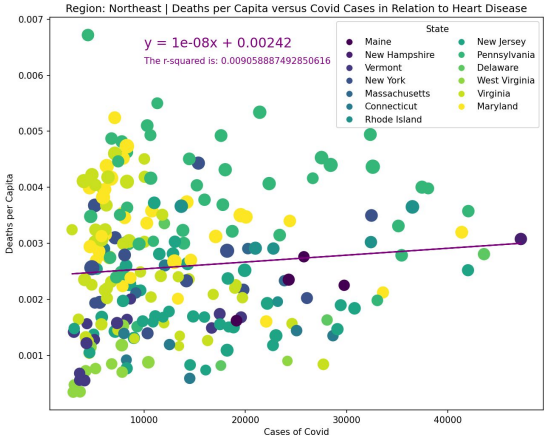
- Little to no correlation between counties, cases of covid, and the deaths per capita



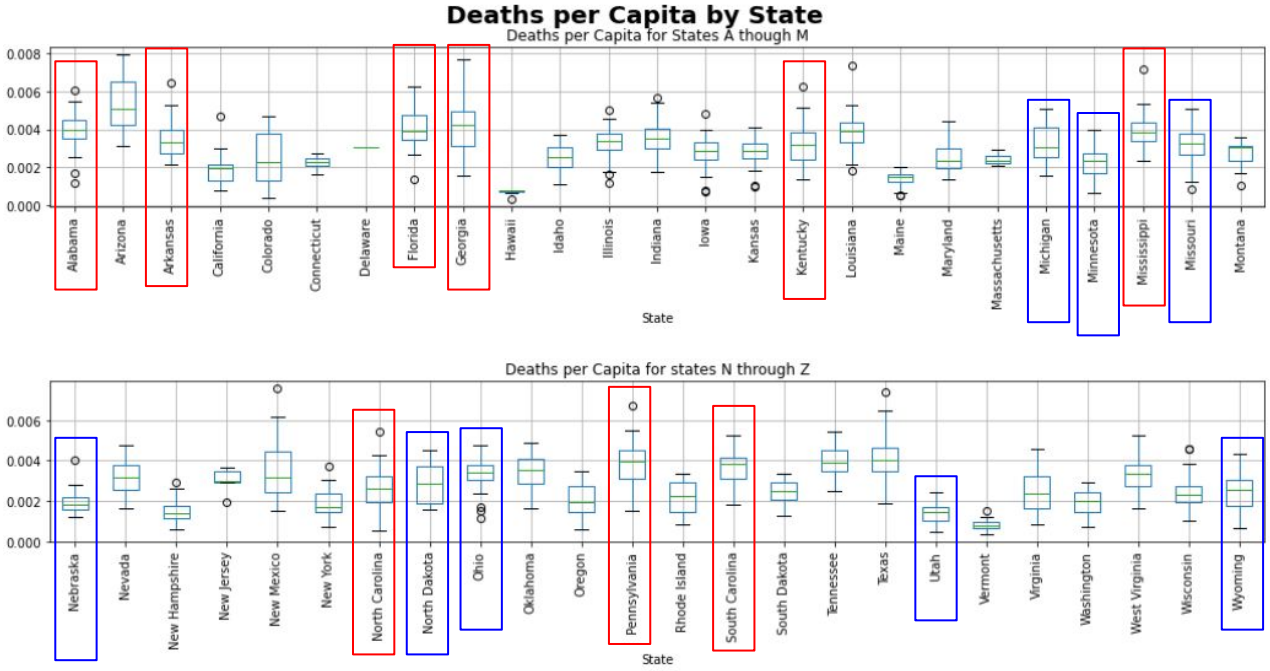



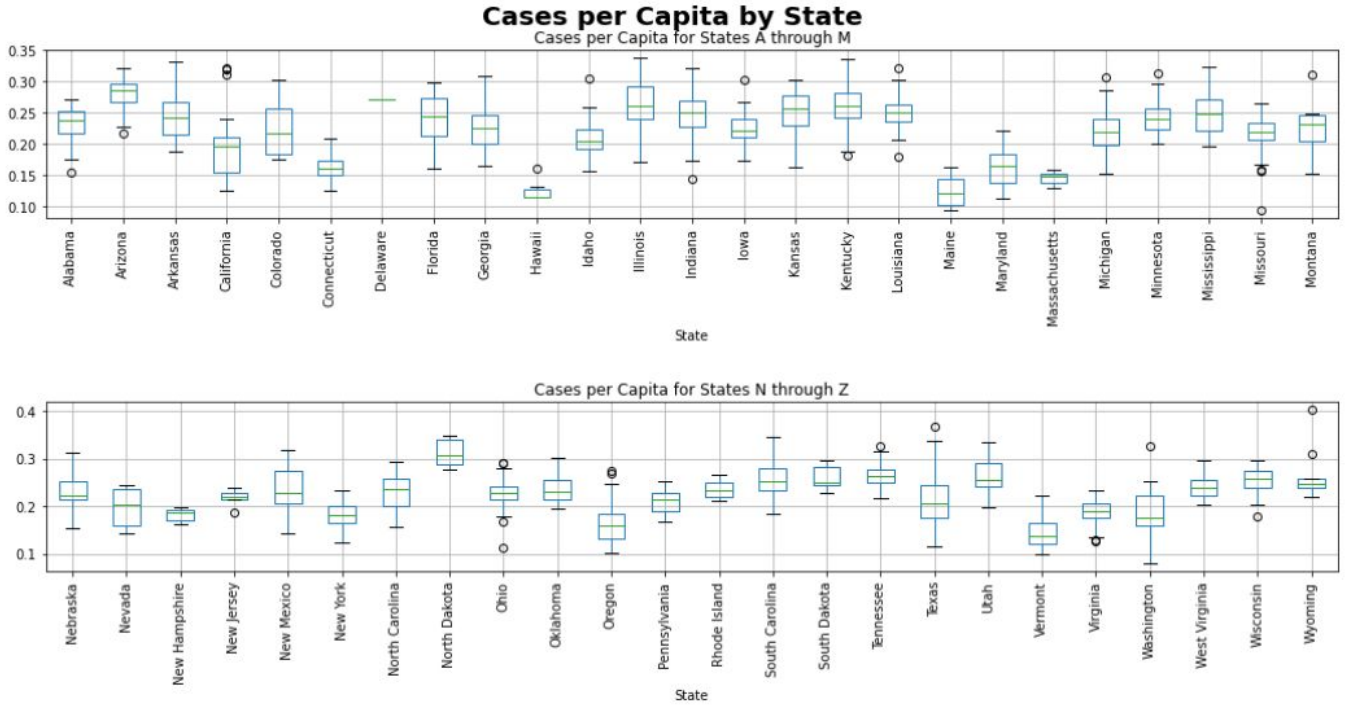
# Data Analysis by Counties and States

# COVID Deaths per Capita By State

- **East coast states** have higher average deaths per capita that Mid East coast states

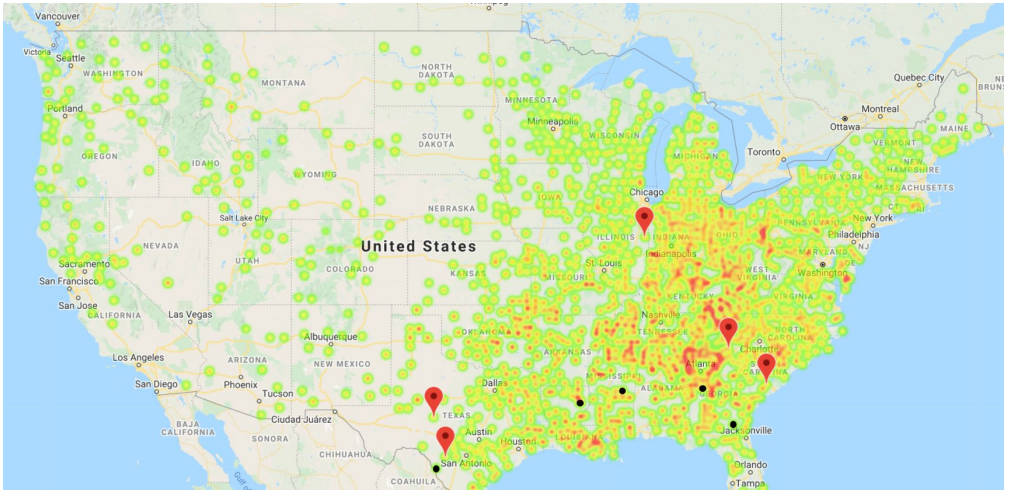


# COVID Cases per Capita By State

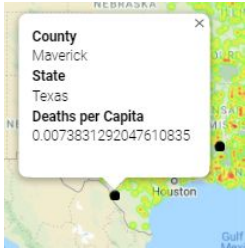


# Heatmap

- API Interaction with Gmaps to create heatmap based on Heart Disease
- Max intensity = 600 out of 100,000 population (0.6% of population)
- Similarity to heatmaps of COVID Deaths per Capita and COVID cases per capita




## Black Dots - Top 5 Counties with COVID Deaths per Capita




| | County | State | Cases per Capita |
|---|---|---|---|
| 0 | Upson | Georgia | 0.280096 |
| 1 | Maverick | Texas | 0.330706 |
| 2 | Franklin | Louisiana | 0.321277 |
| 3 | Neshoba | Mississippi | 0.323087 |
| 4 | Ware | Georgia | 0.258490 |

## Pins - Top 5 Counties with COVID Cases per Capita


| | County | State | Cases per Capita |
|---|---|---|---|
| 0 | Uvalde | Texas | 0.366804 |
| 1 | Pickens | South Carolina | 0.346113 |
| 2 | Vermilion | Illinois | 0.337878 |
| 3 | Tom Green | Texas | 0.335867 |
| 4 | Dorchester | South Carolina | 0.335265 |

# Heatmap

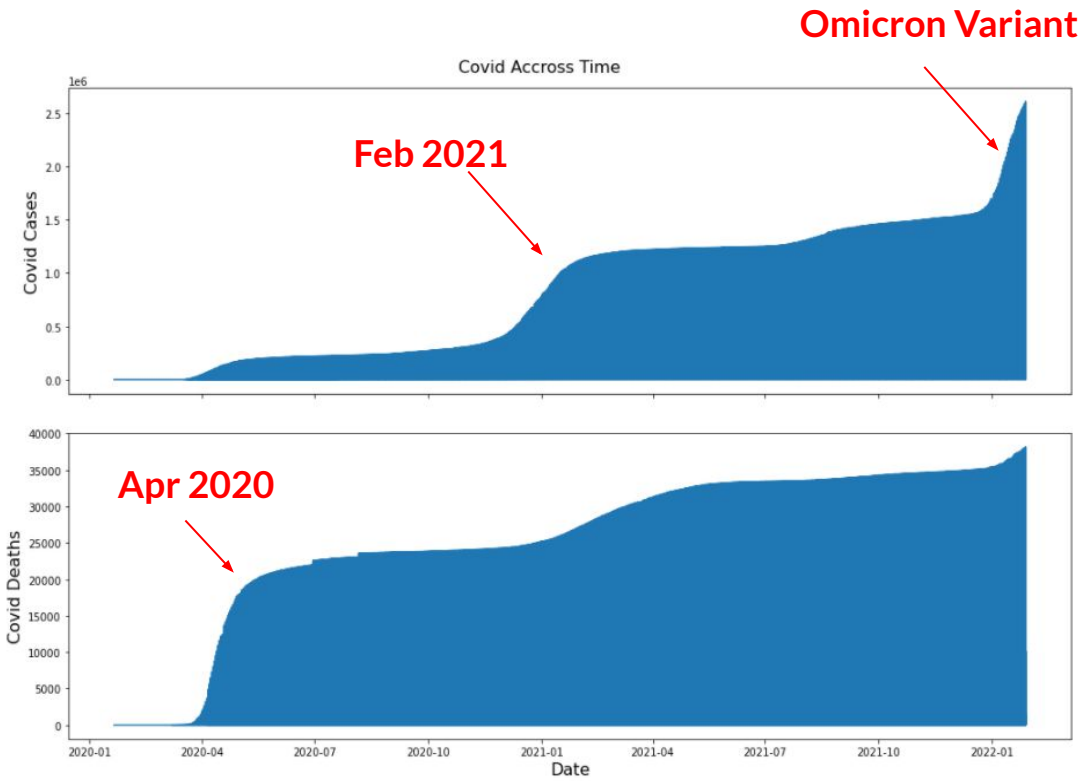
```
def create_map(weights, max_intensity, locations_for_marker_layer, info_box_for_marker_layer, locations_for_symbol_layer, info_box_for_symbol_layer):
    figure = gmaps.figure()
    locations = merged_df[["Y_Latitude", "X_Longitude"]]
    heat_layer = gmaps.heatmap_layer(locations, weights = weights, dissipating = False, max_intensity = max_intensity,point_radius = 0.3)
    figure.add_layer(heat_layer)

    # Add marker layer ontop of heat map
    markers = gmaps.marker_layer(locations_sorted_df_by_cases, info_box_content = info_box_for_marker_layer)
    figure.add_layer(markers)
    symbol_layer = gmaps.symbol_layer(locations_for_symbol_layer, info_box_content=info_box_for_symbol_layer)
    figure.add_layer(symbol_layer)

    return figure
```

```
# Using the template add the County/State to the heatmap
info_box_template_cases_per_capita = """
<dl>
<dt>County</dt><dd>{County}</dd>
<dt>State</dt><dd>{State}</dd>
<dt>Cases per Capita</dt><dd>{Cases per Capita}</dd>
</dl>
"""

box_template_cases = [info_box_template_cases_per_capita.format(**row) for index, row in sorted_df_by_cases.iterrows()]
```

# COVID Trend Across Time

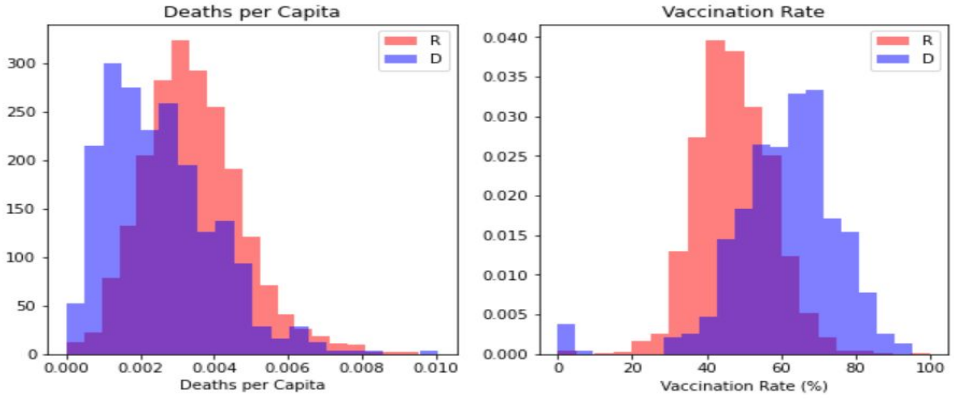
# COVID Cases Across Time

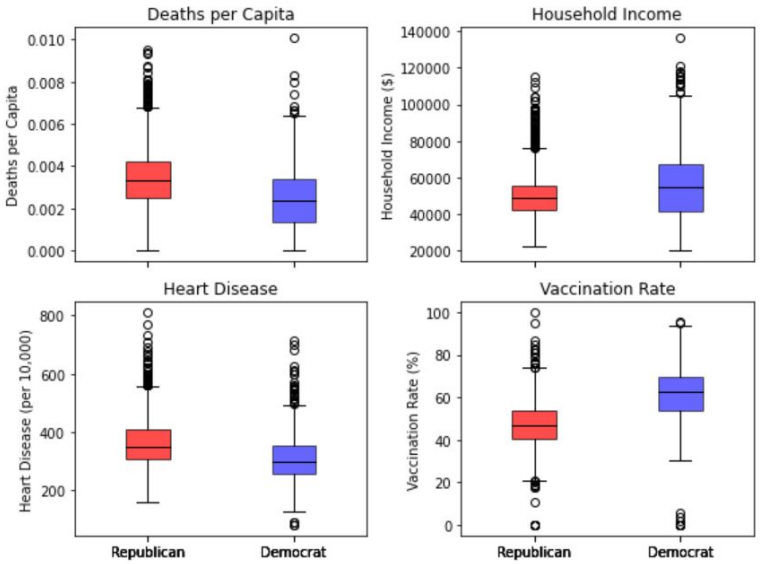
- **Observation**: Different spikes of covid cases vs covid deaths
- **Possible Explanations**:
  1. Vaccination started in early 2021 (even though cases continued to increase, covid deaths remained stable)
  2. Type of Variant



# Political Affiliation Correlations

# Red vs. Blue




| | t-statistic | p-value |
|---|---|---|
| **Deaths per Capita** | 11.988907 | 4.341622e-30 |
| **Household Income** | -6.971634 | 9.006186e-12 |
| **Heart Disease** | 10.603520 | 2.649415e-24 |
| **Vaccination Rate** | -20.152412 | 6.060113e-69 |

# Correlation between Median Income, COVID deaths, and vaccination rates

# Anova Tests for Vaccination Rates & Median Income

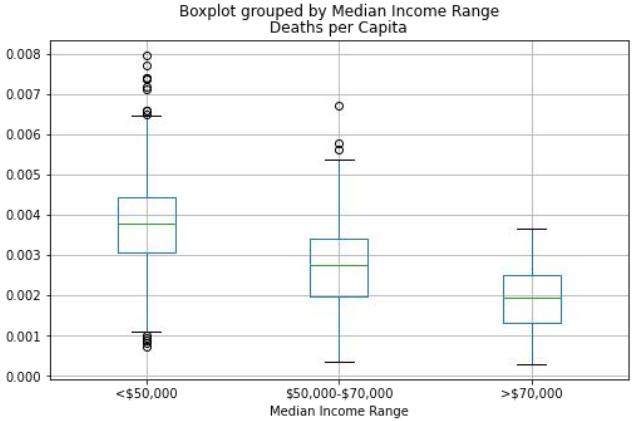
- Binned Data

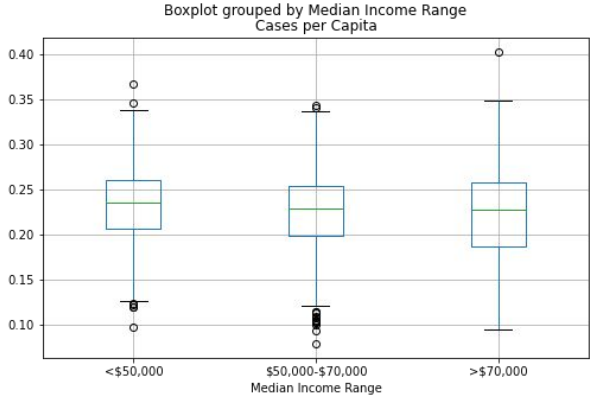
| Deaths | Vaccination Rate | Population | Median Age | Household Income | Heart Disease | Stroke | Party | County | State | Cases per Capita | Deaths per Capita | Median Income Range | Vaccination Rate Range |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57.0 | 39.2 | 24657.0 | 43.8 | 36685.0 | 321.0 | 78.6 | REPUBLICAN | Abbeville | South Carolina | 0.247556 | 0.002312 | <$50,000 | 20%-40% |
| 269.0 | 51.8 | 62568.0 | 36.2 | 41177.0 | 476.2 | 93.6 | REPUBLICAN | Acadia | Louisiana | 0.238956 | 0.004299 | <$50,000 | 40%-60% |
| 90.0 | 70.7 | 32742.0 | 45.9 | 43210.0 | 411.2 | 92.1 | REPUBLICAN | Accomack | Virginia | 0.200629 | 0.002749 | <$50,000 | 60%-80% |
| 56.0 | 44.0 | 25325.0 | 27.7 | 40046.0 | 335.1 | 80.7 | REPUBLICAN | Adair | Missouri | 0.218559 | 0.002211 | <$50,000 | 40%-60% |
| 65.0 | 34.8 | 22113.0 | 37.6 | 32986.0 | 575.0 | 62.7 | REPUBLICAN | Adair | Oklahoma | 0.301994 | 0.002939 | <$50,000 | 20%-40% |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95.0 | 47.5 | 27974.0 | 35.7 | 31402.0 | 368.6 | 119.6 | DEMOCRAT | Yazoo | Mississippi | 0.269429 | 0.003396 | <$50,000 | 40%-60% |
| 94.0 | 47.6 | 21573.0 | 39.3 | 42361.0 | 472.9 | 86.2 | REPUBLICAN | Yell | Arkansas | 0.296018 | 0.004357 | <$50,000 | 40%-60% |
| 486.0 | 50.5 | 157816.0 | 38.2 | 59117.0 | 322.4 | 70.9 | REPUBLICAN | Yellowstone | Montana | 0.242403 | 0.003080 | $50,000-$70,000 | 40%-60% |
| 88.0 | 65.3 | 67587.0 | 39.5 | 90367.0 | 233.0 | 72.0 | REPUBLICAN | York | Virginia | 0.133709 | 0.001302 | >$75,000 | 60%-80% |
| 104.0 | 51.1 | 75493.0 | 32.5 | 52624.0 | 370.1 | 87.2 | REPUBLICAN | Yuba | California | 0.209437 | 0.001378 | $50,000-$70,000 | 40%-60% |

```
#Binning income ranges to do anova analysis

bins = [-float('inf'), 50000, 70000, float('inf')]
income_ranges = ['<$50,000', "\$50,000-\$70,000", ">$75,000"]
data["Median Income Range"] = pd.cut(data["Household Income"], bins, labels=income_ranges, include_lowest=True)
data
```

# Binned Income Boxplot

- ANOVA test returned strongly rejects null hypothesis
- P-value for deaths is ~2.05*$e^{-90}$
- P-value for deaths is ~.0005

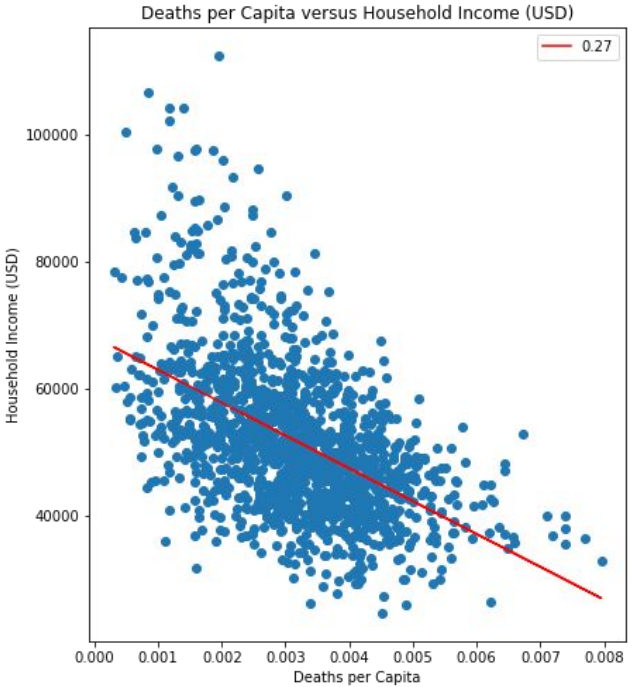


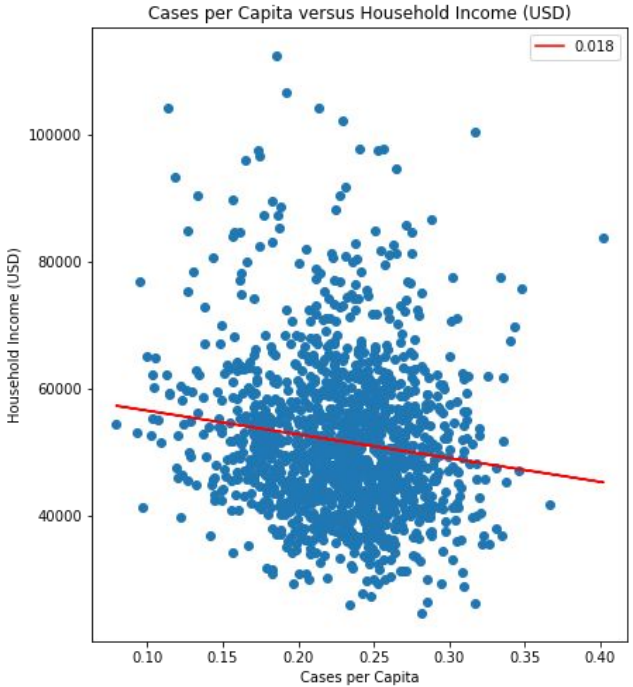




```
# Extract individual groups and perform ANOVA test
income_groups = []
for i in income_ranges:
    income_groups.append(data[data["Median Income Range"] == i]["Deaths per Capita"])
_, p = stats.f_oneway(*income_groups)
print(f"The null hypothesis' pvalue for income ranges vs deaths per capita is {p}")
income_groups = []
for i in income_ranges:
    income_groups.append(data[data["Median Income Range"] == i]["Cases per Capita"])
_, p = stats.f_oneway(*income_groups)
print(f"The null hypothesis' pvalue for income ranges vs cases per capita is {p}")
```

```
The null hypothesis' pvalue for income ranges vs deaths per capita is 2.053183505933904e-90
The null hypothesis' pvalue for income ranges vs cases per capita is 0.0005751076241736748
```

# Scatter Plots and $R^2$ Value

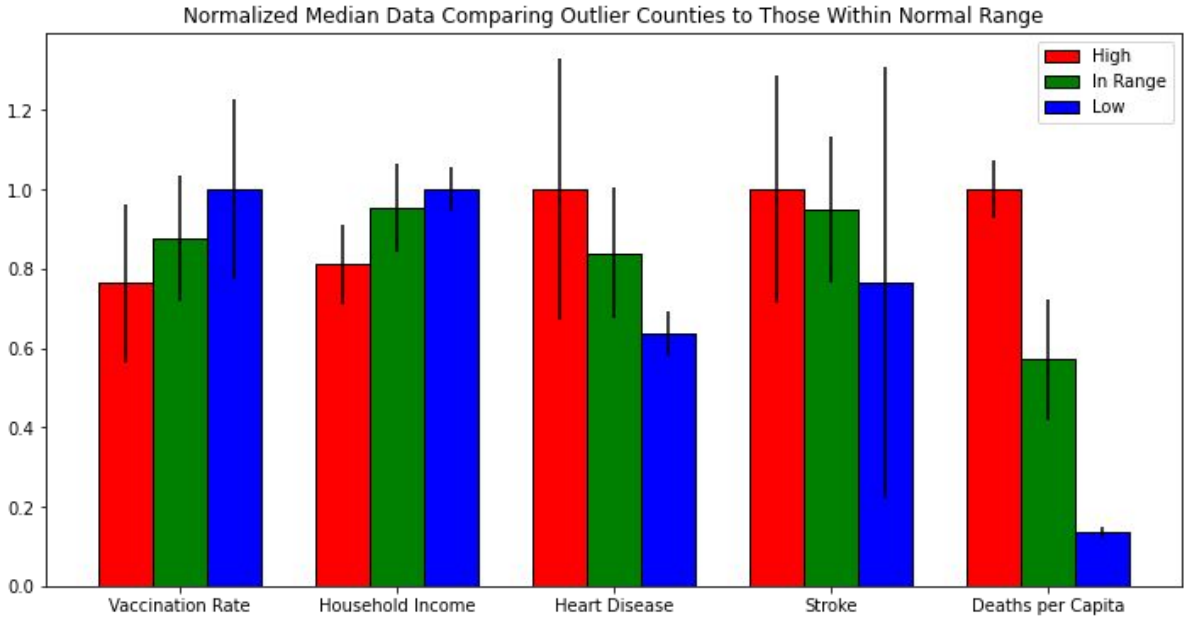


# Analysis of Outlier Counties

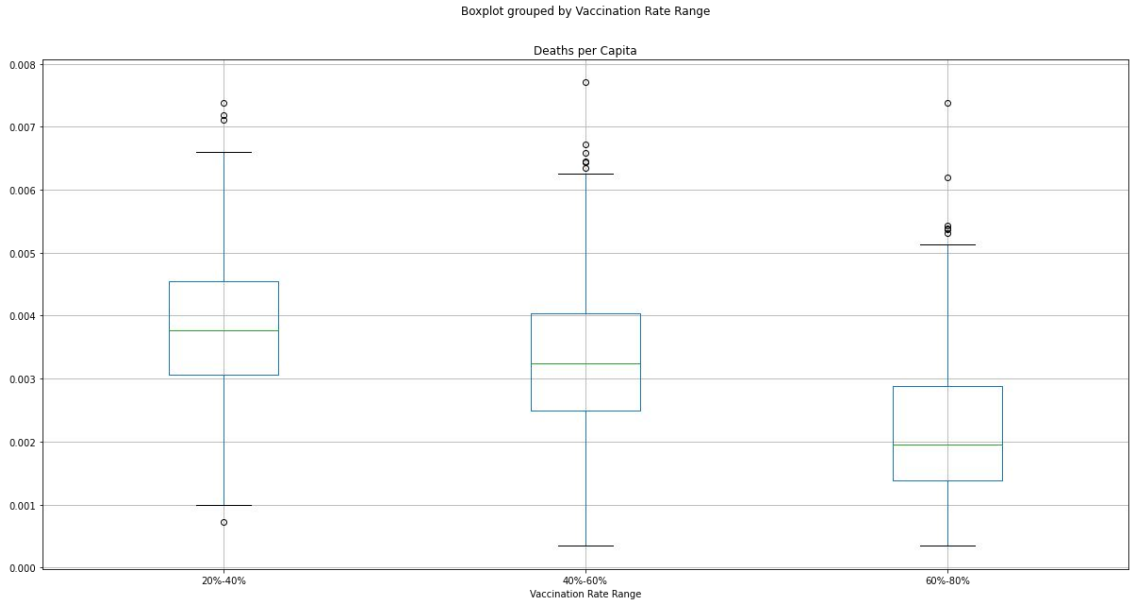
- Data normalization
- Understanding error bars
- A good indicator that income and heart disease in particular are account for many early deaths

| | Vaccination Rate | Household Income | Heart Disease | Stroke | County | State | Deaths per Capita | Median Income Range | Vaccination Rate Range |
|---|---|---|---|---|---|---|---|---|---|
| **163** | 35.9 | 49348.0 | 301.8 | 57.2 | Caledonia | Vermont | 0.000723 | <$50,000 | 20%-40% |
| **604** | 64.3 | 45528.0 | 308.4 | 166.6 | Humboldt | California | 0.000950 | <$50,000 | 60%-80% |
| **1034** | 52.3 | 44315.0 | 321.5 | 103.6 | Pitt | North Carolina | 0.000823 | <$50,000 | 40%-60% |
| **1099** | 38.2 | 49910.0 | 269.8 | 66.5 | Riley | Kansas | 0.000996 | <$50,000 | 20%-40% |
| **1371** | 54.3 | 45268.0 | 260.2 | 54.7 | Watauga | North Carolina | 0.000887 | <$50,000 | 40%-60% |



# Binned Vaccination Boxplot

- Vaccination Rates also reject the null Hypothesis but not as strongly




```
In [8]: #Performing ANOVA test for vaccination rates
vax_groups = []
for i in vaccination_rates:
    vax_groups.append(data[data["Vaccination Rate Range"] == i]["Deaths per Capita"])
_, p = stats.f_oneway(*vax_groups)
print(f"The null hypothesis' pvalue for vaccination rates is {p}")
```

```
The null hypothesis' pvalue for vaccination rates is 9.010006665976374e-50
```

# Issues with Vaccination Rate Data

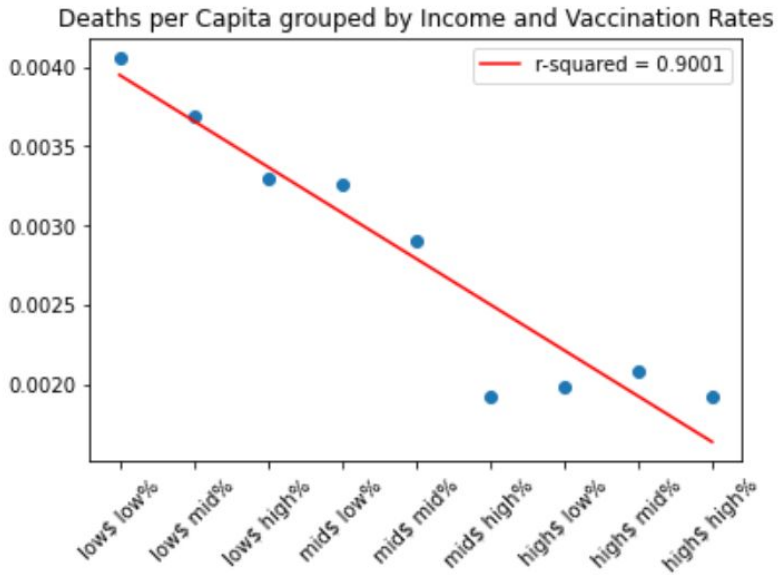
- Vaccination Rates correlate highly with Household Income as well as Heart Disease.
- In fact, Heart Disease itself is highly correlated with household income
- Vaccination Rates are taken at the end of covid, this data measures deaths from before a vaccine was created!
- Better to measure deaths from a given time point until now.
- So does this mean the data has nothing to say about vaccination?

| | Cases | Deaths | Vaccination Rate | Population | Median Age | Household Income | Heart Disease | Stroke | Cases per Capita | Deaths per Capita |
|---|---|---|---|---|---|---|---|---|---|---|
| **Cases** | 1.000000 | 0.822636 | 0.207994 | 0.950363 | -0.238787 | 0.270933 | -0.155044 | -0.065259 | 0.257009 | -0.104763 |
| **Deaths** | 0.822636 | 1.000000 | 0.049213 | 0.799732 | -0.084372 | 0.026746 | 0.073631 | 0.085283 | 0.175210 | 0.317244 |
| **Vaccination Rate** | 0.207994 | 0.049213 | 1.000000 | 0.258447 | 0.140084 | 0.348115 | -0.449743 | -0.346423 | -0.124366 | -0.399787 |
| **Population** | 0.950363 | 0.799732 | 0.258447 | 1.000000 | -0.172629 | 0.332495 | -0.207696 | -0.108684 | 0.003292 | -0.198430 |
| **Median Age** | -0.238787 | -0.084372 | 0.140084 | -0.172629 | 1.000000 | -0.027420 | -0.075867 | -0.104189 | -0.299938 | 0.079298 |
| **Household Income** | 0.270933 | 0.026746 | 0.348115 | 0.332495 | -0.027420 | 1.000000 | -0.534294 | -0.439737 | -0.135538 | -0.519665 |
| **Heart Disease** | -0.155044 | 0.073631 | -0.449743 | -0.207696 | -0.075867 | -0.534294 | 1.000000 | 0.516215 | 0.197634 | 0.544322 |
| **Stroke** | -0.065259 | 0.085283 | -0.346423 | -0.108684 | -0.104189 | -0.439737 | 0.516215 | 1.000000 | 0.144171 | 0.400362 |
| **Cases per Capita** | 0.257009 | 0.175210 | -0.124366 | 0.003292 | -0.299938 | -0.135538 | 0.197634 | 0.144171 | 1.000000 | 0.335218 |
| **Deaths per Capita** | -0.104763 | 0.317244 | -0.399787 | -0.198430 | 0.079298 | -0.519665 | 0.544322 | 0.400362 | 0.335218 | 1.000000 |

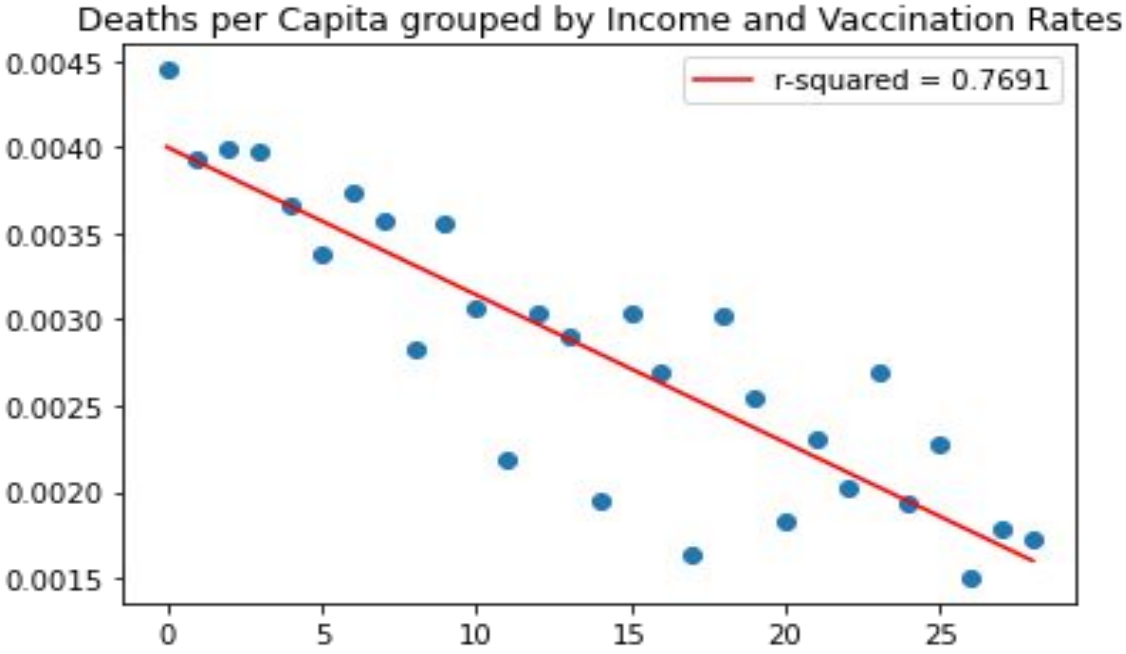
# Broad Visualisation of Incomes and Vaccination Rate

- This chart certainly disagrees!
- Primary driver is still income
- What are broader implications?

| Median Income Range | Vaccination Rate Range | Deaths per Capita |
|---|---|---|
| <$50,000 | 20%-40% | 0.004063 |
| | 40%-60% | 0.003690 |
| | 60%-80% | 0.003298 |
| $50,000-$70,000 | 20%-40% | 0.003262 |
| | 40%-60% | 0.002910 |
| | 60%-80% | 0.001927 |
| >$75,000 | 20%-40% | 0.001982 |
| | 40%-60% | 0.002085 |
| | 60%-80% | 0.001920 |



# Vaccination Rates Broken into Finer Income Groups



# Summary and Next Steps

# Summary of Findings

- What region of the country has the highest rate of covid cases, covid deaths and heart disease?
  - East coast states had higher previous heart disease problems prior to the pandemic, and thus higher number of covid cases and covid deaths in 2020-2022 (possible limitation of
  - data points in the west coast affecting the results)
- How did covid cases and covid deaths progressed over time?
  - Proportion of deaths to covid cases has decreased since the introduction of the vaccine early 2021
- How does political affiliation affect vaccination, covid deaths, and heart disease?
  - Republican counties have significantly lower vaccination rates, higher rates of heart disease, lower income, and higher COVID death rates
- What is the correlation between Median Income, Vaccination Rates and Covid Deaths?
  - States with disparate incomes had similar rates of infection, but vastly different rates of death by covid
- How did heart disease rates in the years previous to COVID  affect the number of covid deaths in 2020?
  - When accounting for income, high rates of heart disease made you significantly more likely to have an abnormally high rate of death

# Post Mortem

## DIFFICULTIES

- Limitation of columns on selected datasets -> Need to include additional datasets
- Limitation of data in states in the Midwest (most states have less than 20 counties)
- Division of the states in groups (west vs midwest vs east states; red vs blue states)
- Accounting for total deaths made vaccination data less reliable
- High cross correlations can make it difficult to understand root causes

## NEXT STEPS

- Study more in depth how vaccination affected covid deaths