

An Overview on COVID-19 and Racial Disparity & Population Density

...

Kexin Chen, Jordan Paquin

Project Goal

- Investigate COVID-19 data on different scale
- Geovisualization of racial disparity at the state level
- Correlation analysis between population density and number of cases at the county level

Coronavirus Disease 2019 (COVID-19)

CDC > Coronavirus Disease 2019 (COVID-19) > People Who Need Extra Precautions > Others At Risk



Coronavirus Disease 2019
(COVID-19)

Symptoms

Testing +

Prevent Getting Sick +

COVID-19 in Racial and Ethnic Minority Groups

Other Languages

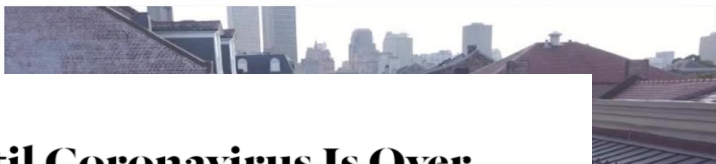
Print Page

The effects of COVID-19 on the health c
current data suggest a disproportionat

Black communities account for disproportionate number of Covid-19 deaths in the US, study finds

By Shelby Lin Erdman

Updated 10:32 AM ET, Wed May 6, 2020



More from CNN



A Dollar Tree worker told a man he needed a mask, so he wiped...



Retired Colorado paramedic dies from coronavirus after

PERSPECTIVE

We Can't Wait Until Coronavirus Is Over to Address Racial Disparities

Early data show worse outcomes for Americans of color from Covid-19. This isn't only a reflection of historic inequality. The response is *creating* inequality, too.

APRIL 22, 2020



JUNIA HOWELL

urban sociologist, University of Pittsburgh

SHARE

TWEET

...

[Recent data](#) from the Centers for Disease Control and Prevention on

Racial disparity

- Health differences between racial and ethnic groups are often due to economic and social conditions that are more common among some racial and ethnic minorities than whites. (CDC)



<https://www.npr.org/sections/coronavirus-live-updates/2020/04/14/834466606/new-survey-highlights-racial-disparities-in-the-coronavirus-pandemic>

<https://www.wusa9.com/article/news/health/coronavirus/african-americans-are-dying-disproportionately-from-coronavirus-states-with-the-worst-disparities/507-3f12c705-5b60-4eab-8d84-6030d13796dd>

Data Collection

- Percentage of confirmed cases by racial group for each state
 - Collected on 4/21
 - Source: State government website
-
- Population percentage by race/ethnicity for each state
 - Source: <https://www.governing.com/gov-data/census/state-minority-population-data-estimates.html>

Sample Data

State	Hispanic (of any race)	Non-Hispanic White	Non-Hispanic Black	Non-Hispanic Asian
Alabama	0.04	0.66	0.27	0.01
Alaska	0.07	0.61	0.03	0.07
Arizona	0.31	0.55	0.04	0.03
Arkansas	0.07	0.72	0.15	0.02
California	0.39	0.37	0.06	0.14
Colorado	0.22	0.68	0.04	0.03
Connecticut	0.16	0.67	0.1	0.05
Delaware	0.09	0.62	0.22	0.04
District of Columbia	0.11	0.37	0.45	0.04
Florida	0.26	0.54	0.15	0.03
Georgia	0.1	0.53	0.31	0.04
Hawaii	0.11	0.22	0.02	0.37
Idaho	0.12	0.82	0.01	0.01
Illinois	0.17	0.61	0.14	0.05
Indiana	0.07	0.79	0.09	0.02
Iowa	0.06	0.86	0.03	0.03
Kansas	0.12	0.76	0.06	0.03
Kentucky	0.04	0.85	0.08	0.01

State	White	Black	Hispanic	Asian	url
Alabama	0.54	0.43	0.04	0.01	https://www.alabamapublichealth.gov/covid19/
Alaska					
Arizona	0.47	0.05	0.25		https://www.azdhs.gov/preparedness/epidemiology/
Arkansas	0.59	0.32		0.01	https://www.healthy.arkansas.gov/program-and-services/
California	0.29	0.07	0.4	0.13	https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Imz.aspx
Colorado	0.55	0.07	0.31	0.03	https://covid19.colorado.gov/case-data
Connecticut	0.53	0.19	0.24	0.02	https://portal.ct.gov/-/media/Coronavirus/COVID-19-Daily-Status-Reports
Delaware					
District of Columbia	0.22	0.58	0.22	0.02	https://coronavirus.dc.gov/release/coronavirus-daily-status-reports
Florida	0.66	0.22	0.4		https://floridadisaster.org/globalassets/covid-19-daily-status-reports
Georgia	0.4	0.53	0.11	0.02	https://dph.georgia.gov/covid-19-daily-status-reports
Hawaii					
Idaho					
Illinois	0.33	0.33	0.23	0.05	http://www.dph.illinois.gov/covid19/covid19-daily-status-reports
Indiana	0.62	0.21	0.11	0.01	https://www.coronavirus.in.gov/2393.htm
Iowa	0.74	0.17	0.24	0.07	https://idph.iowa.gov/Emerging-Health-Issues
Kansas	0.74	0.2	0.25	0.02	https://www.coronavirus.kdheks.gov/160/Covid-19-Daily-Status-Reports
Kentucky					

Geovisualization by ArcGIS Online

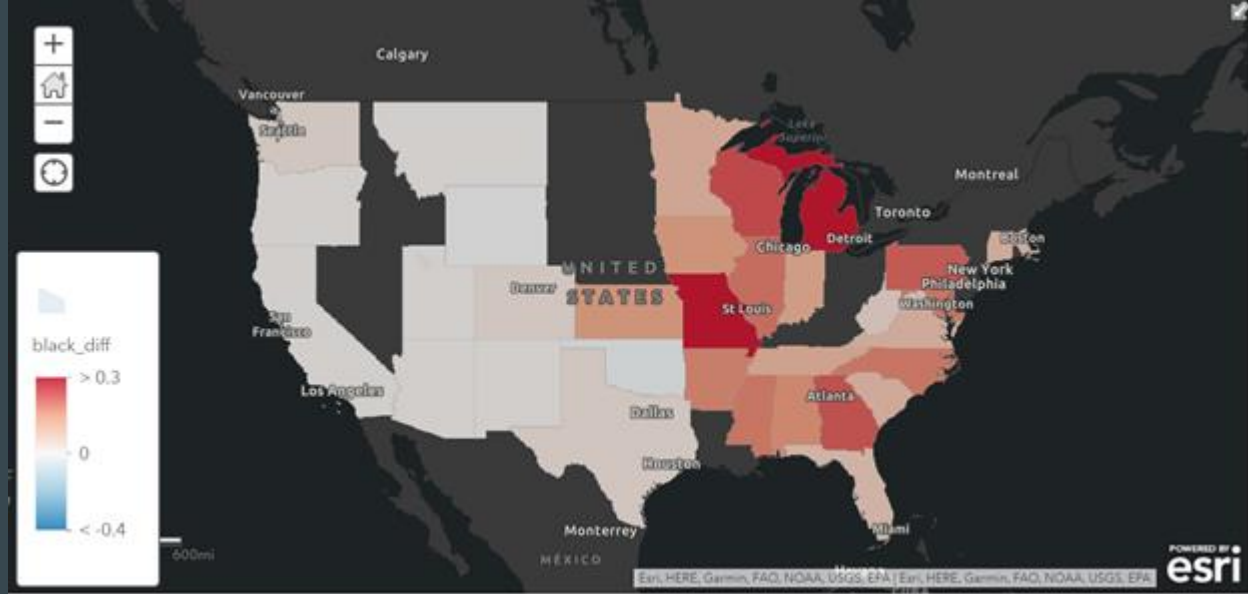
- Aggregate the data using Pandas DataFrame
- Convert data into geojson format
- Import as feature layer into ArcGIS Online

```
In [121]: new_data ={\n    'type': 'FeatureCollection',\n    'features':[\n\n    ]\n}\nwith open("copy.json") as f:\n    data = json.load(f)\n\nfor state in data['features']:\n    feature_dict={}\n    feature_dict['type']='Feature'\n    feature_dict['geometry'] = state['geometry']\n    feature_dict['properties']={}\n    properties = state['properties']\n    state_name = properties["NAME"]\n    if state_name == "Puerto Rico":\n        continue\n\n    for item in race_diff_dict:\n        if item["State"] == state_name:\n            feature_dict['properties'] = item\n            new_data['features'].append(feature_dict)\n\nnew_data['features']\n\n#puertorico
```

```
Out[121]: [{'type': 'Feature',\n  'geometry': {'type': 'MultiPolygon',\n  'coordinates': [[[[[-67.619761, 44.519754],\n    [-67.61541, 44.521973],\n    [-67.587738, 44.516196],\n    [-67.582113, 44.513459],\n    [-67.589259, 44.50084],\n    [-67.590627, 44.49415],\n    [-67.580288, 44.488068],\n    [-67.562651, 44.472104],
```

Black

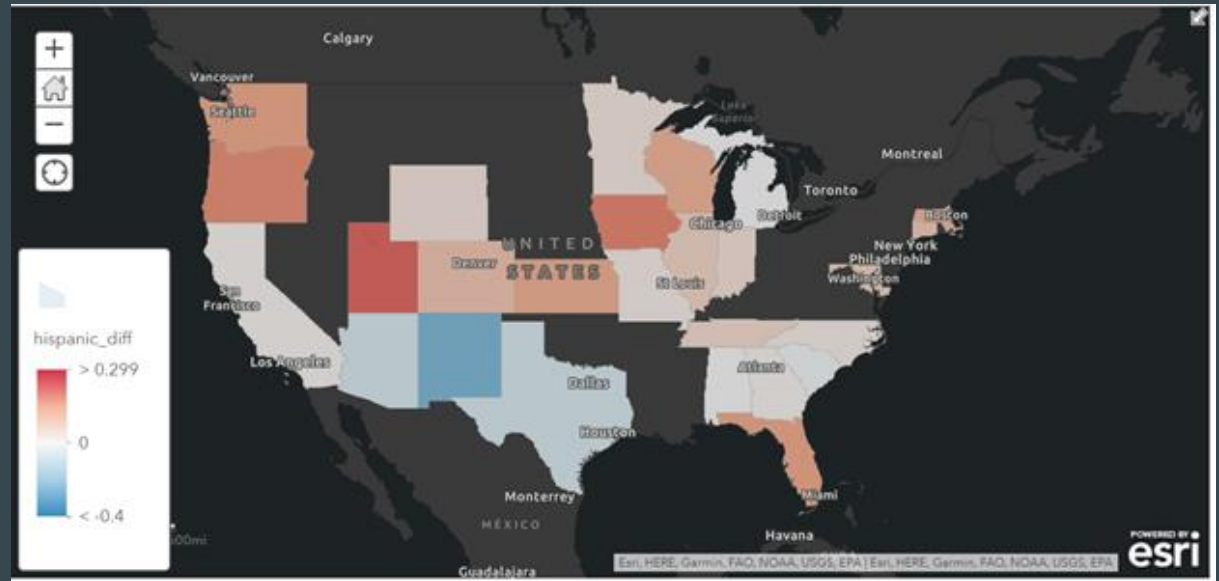
- Black communities account for disproportionate number of cases in the eastern US
- Percentage of cases is higher than its population percentage in 33 states among the 34 states
- In Missouri and Michigan, the percentage of cases is 29% higher than their population percentage in the state



- **Red:** percentage of cases > percentage of population
- **White:** percentage of cases = percentage of population
- **Blue:** percentage of cases < percentage of population

Hispanic

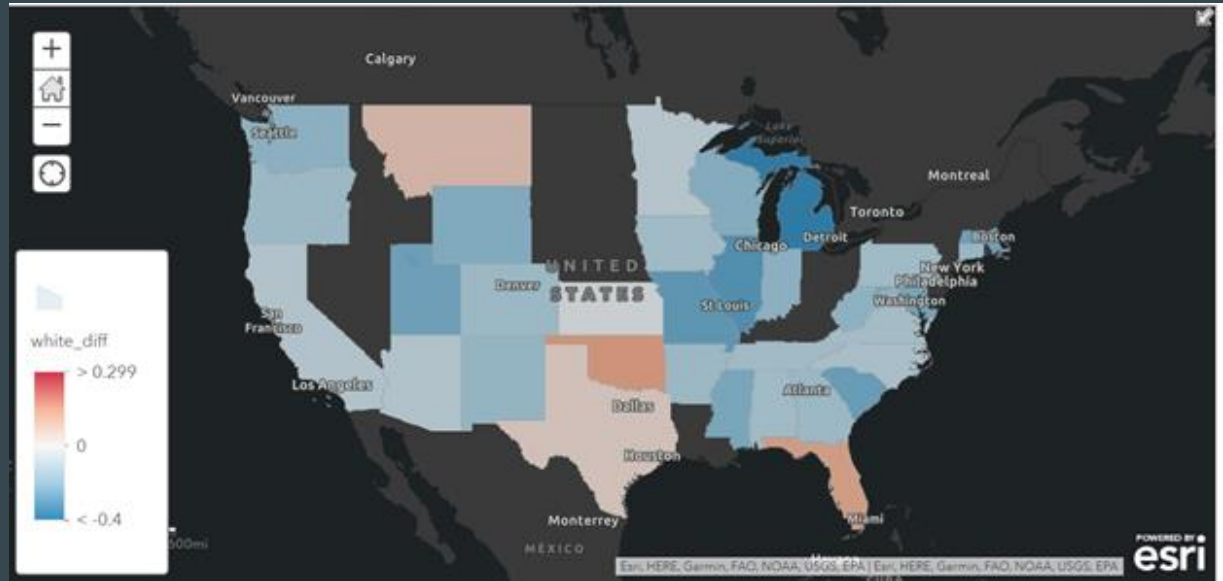
- More missing data than that of the blacks
- Percentage of cases is higher than its population percentage in 21 states among the 27 states
- In Utah, the percentage of cases is 21% higher than its population percentage, which is the highest across all the states.



- **Red:** percentage of cases > percentage of population
- **White:** percentage of cases = percentage of population
- **Blue:** percentage of cases < percentage of population

White

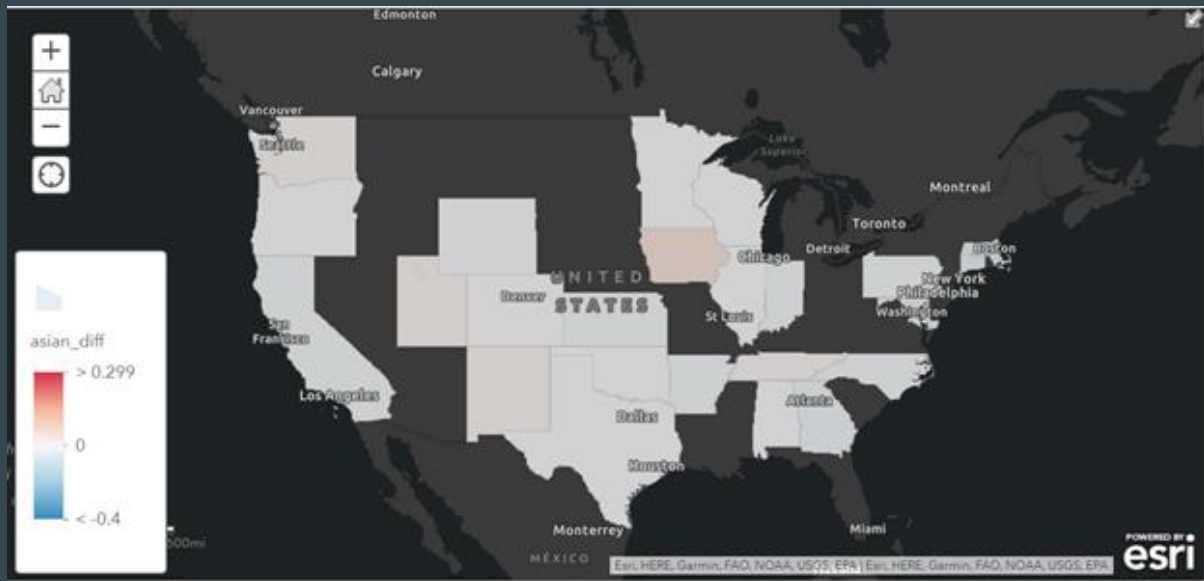
- In most states, the percentage of cases is lower than its population percentage
- The percentage of cases is 2% to 34% lower than its population percentage in the blue states



- **Red:** percentage of cases $>$ percentage of population
- **White:** percentage of cases = percentage of population
- **Blue:** percentage of cases $<$ percentage of population

Asian

- Missing data and uncertainty involved in the data because of the general small population compared to other groups
- Neither overrepresented or underrepresented
- Percentage of cases higher or lower to percentage of population in a small range: 1% - 4%



- **Red:** percentage of cases $>$ percentage of population
- **White:** percentage of cases = percentage of population
- **Blue:** percentage of cases $<$ percentage of population

Conclusion

- Black and Hispanic communities are more vulnerable to the pandemic
- Still miss a lot of data:
 - Some state governments didn't publish racial group data at all
 - States classify race/ethnicity groups differently. Most data of other minority groups like Pacific Islanders, Native Americans, multi-races, etc are not comparable
- Uncertainty: some published data are not refined

Population Density and Number of Cases

Main question:

Is population density a determinant of number of cases or infection rate?

Data Collection

- Total confirmed cases at the county level
 - Data collected on 5/3
 - Source: CSSE at Johns Hopkins University
 - https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports_us
-
- Population density (per square mile) at the county level
 - Source: ArcGIS Hub: USA Counties (Generalized)
 - <https://hub.arcgis.com/datasets/esri::usa-counties-generalized>

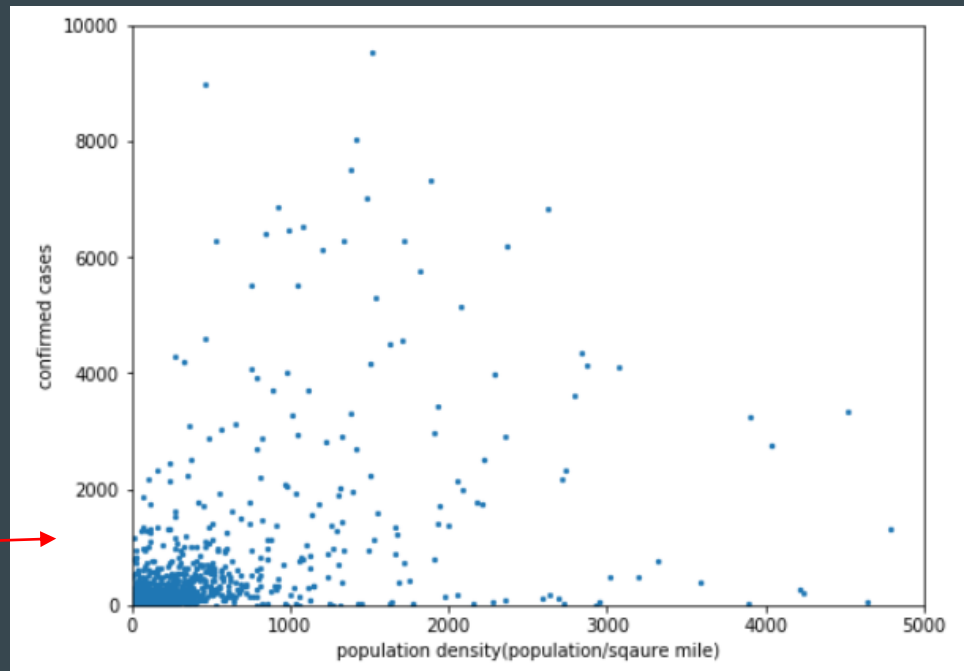
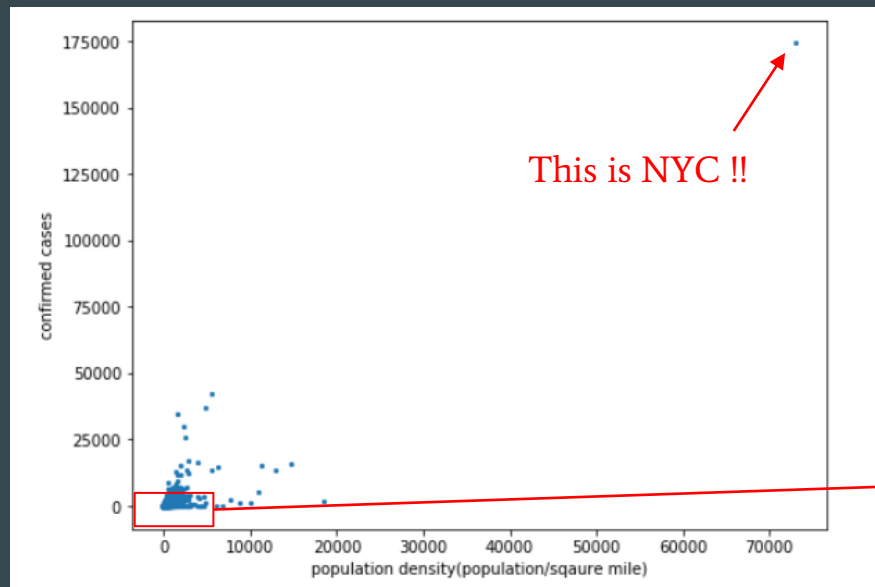
Method: Simple Linear Regression

1. Population density and total confirmed cases
2. Population density and infection rate (cases per 10000 people)

FIPS	Admin2	Province	Country	Last_Update	Lat	Long	Confirmed	Deaths	Recovered	Active	Combined_Key
45001	Abbeville	South Carolina	US	#####	34.22333	-82.4617	33	0	0	0	33 Abbeville, South Carolina, US
22001	Acadia	Louisiana	US	#####	30.29506	-92.4142	134	10	0	0	124 Acadia, Louisiana, US
51001	Accomack	Virginia	US	#####	37.76707	-75.6323	400	6	0	0	394 Accomack, Virginia, US
16001	Ada	Idaho	US	#####	43.45266	-116.242	705	17	0	0	688 Ada, Idaho, US
19001	Adair	Iowa	US	#####	41.33076	-94.4711	1	0	0	0	1 Adair, Iowa, US
21001	Adair	Kentucky	US	#####	37.1046	-85.2813	81	13	0	0	68 Adair, Kentucky, US
29001	Adair	Missouri	US	#####	40.19059	-92.6008	12	0	0	0	12 Adair, Missouri, US
40001	Adair	Oklahoma	US	#####	35.88494	-94.6586	65	3	0	0	62 Adair, Oklahoma, US
8001	Adams	Colorado	US	#####	39.87432	-104.336	1780	68	0	0	1712 Adams, Colorado, US
16003	Adams	Idaho	US	#####	44.89334	-116.455	3	0	0	0	3 Adams, Idaho, US
17001	Adams	Illinois	US	#####	39.98816	-91.1879	40	1	0	0	39 Adams, Illinois, US
18001	Adams	Indiana	US	#####	40.74577	-84.9367	8	1	0	0	7 Adams, Indiana, US
28001	Adams	Mississippi	US	#####	31.4767	-91.3533	148	9	0	0	139 Adams, Mississippi, US
31001	Adams	Nebraska	US	#####	40.52449	-98.5012	203	3	0	0	200 Adams, Nebraska, US
39001	Adams	Ohio	US	#####	38.84541	-83.4719	5	0	0	0	5 Adams, Ohio, US
42001	Adams	Pennsylvania	US	#####	39.8714	-77.2161	145	4	0	0	141 Adams, Pennsylvania, US
53001	Adams	Washington	US	#####	46.983	-118.56	48	0	0	0	48 Adams, Washington, US
55001	Adams	Wisconsin	US	#####	43.96975	-89.7678	4	1	0	0	3 Adams, Wisconsin, US
50001	Addison	Vermont	US	#####	44.03217	-73.1413	61	2	0	0	59 Addison, Vermont, US

Sample data from JHU Github

Pop Density and Total Cases: county level



Pop Density and Total Cases

- Dependent variable: population density
- Independent variable: total cases
- Coefficient of determination: 0.804
- Seems like population density is an important factor of total cases ?

```
In [113]: ▶ x=merged['POP_SQMI'].values.reshape(-1,1)
           y=merged['Confirmed'].values.reshape(-1,1)
           x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1, random_state=0)

In [114]: ▶ regressor = LinearRegression()
           regressor.fit(x_train, y_train)

Out[114]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

In [115]: ▶ r2=regressor.score(x_train,y_train)
           r2

Out[115]: 0.8039924239177635

In [116]: ▶ print(regressor.intercept_)
           print(regressor.coef_)

[-146.18061991]
[[2.11862376]]
```

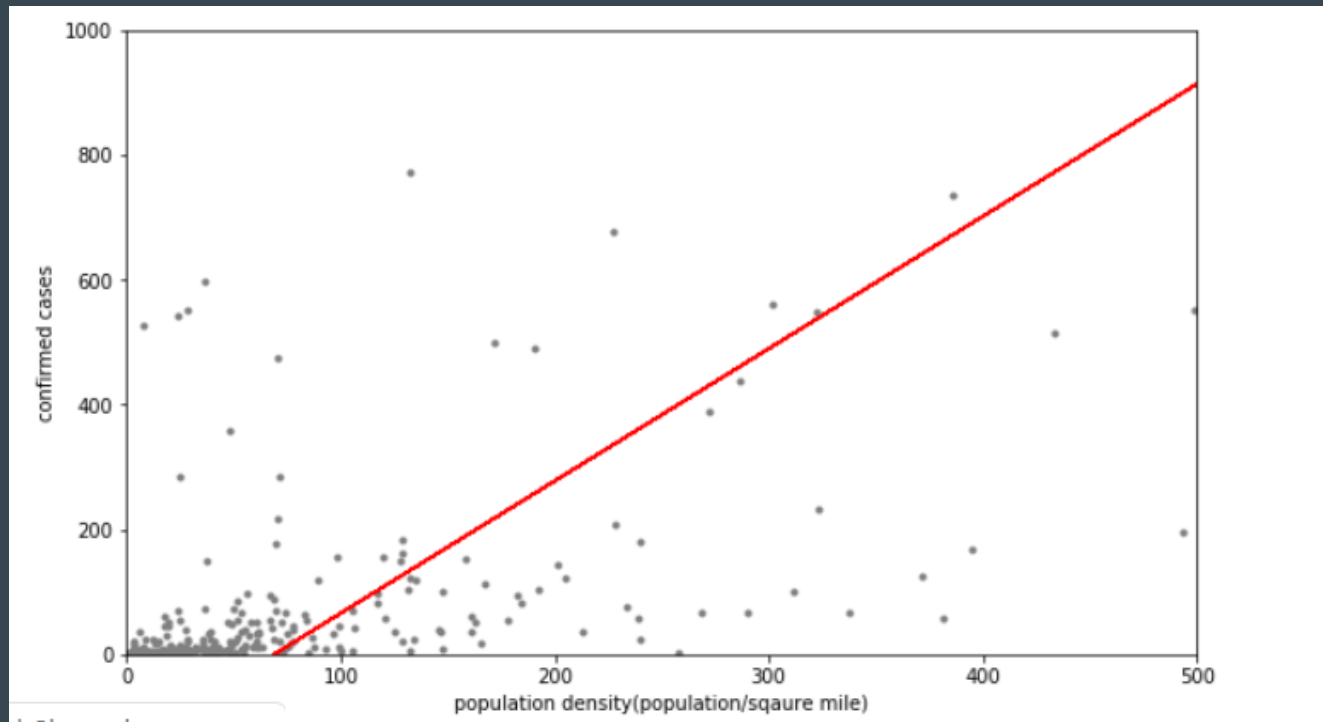
However...

Pop Density and Total Cases

- Train the model with 90% of the data
- Use the model to predict total cases by the rest 10% of the data
- Predicted values of total cases don't match with the actual values

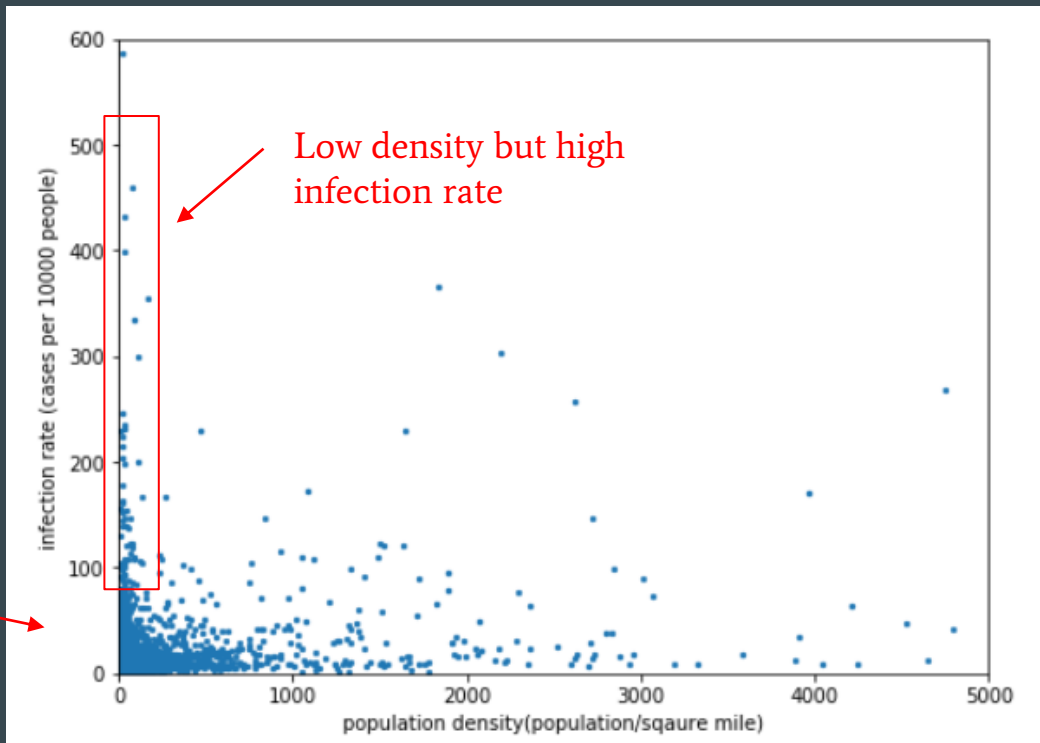
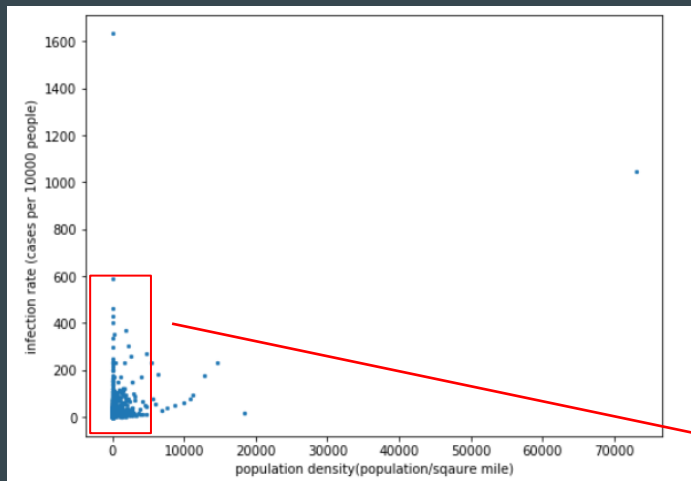
	Actual	Predicted
0	14	-103.384420
1	5	-138.977299
2	1	-131.773978
3	17	-117.155474
4	3	-57.622147
...
281	11	64.410582
282	5	-108.045392
283	25	138.774276
284	388	429.449457
285	67	-33.469836

Pop Density and Total Cases



- Only few points fall on the best-fit line
- Population density may have a weak linear relationship with total cases
- Other factors matter: base population of the county, mobility of people, rurality of the county...

Pop Density and Infection Rate (cases per 10000 people)



Pop Density and Infection Rate

- Dependent variable: population density
- Independent variable: infection rate
- Coefficient of determination: 0.189
- Infection rate is not dependent on population density
- Many counties with low population density have high infection rate

Pop Density and Infection Rate

The two areas of infection rate > 1000 cases per 10000 people

	FIPS	STATE_NAME	NAME	POP_SQMI	cases per 10000	Confirmed	POPULATION	Lat	Long_
1877	36061.0	New York	New York	73032.2	1047.407961	174331	1664404	40.767273	-73.971526
2566	47169.0	Tennessee	Trousdale	70.7	1631.910766	1346	8248	36.390262	-86.160880

The spike of cases in Trousdale is because of an outbreak in prison !

Very weak relationship between population density and infection rate

Potential hotspots of Covid-19?

Places with
population density <
100 people per
square mile and
infection rate > 200
cases per 10000
people

STATE_NAME	NAME	POP_SQMI	cases per 10000	Confirmed	POPULATION
Georgia	Early	20.7	204.634648	219	10702
Georgia	Randolph	17.7	214.191853	163	7610
Nebraska	Dawson	23.7	224.352353	543	24203
Idaho	Blaine	8.1	229.476406	497	21658
Kansas	Ford	32.1	231.035265	815	35276
Kansas	Seward	36.2	235.704744	547	23207
Iowa	Louisa	26.9	246.200338	277	11251
Indiana	Cass	93.0	335.130372	1293	38582
Tennessee	Bledsoe	36.8	399.198932	598	14980
Minnesota	Nobles	30.2	431.172882	940	21801
Nebraska	Dakota	79.6	460.288385	980	21291
Arkansas	Lincoln	24.5	586.388929	822	14018
Tennessee	Trousdale	70.7	1631.910766	1346	8248

Thank you !