The Socio-Economic Demographic Impact upon COVID-19 Infection & Death Rates in the United States

Jennifer Johnson

Math 217 – Biostatistics

Final Project Draft Report

Date: 8 May 2020

Table of Contents

Part I – Introduction

A. Research Focus

In human history, infectious diseases have been the cause of the most deaths, more than wars, more than cancers.[1] In December of 2019 Wuhan, China experienced an outbreak of a new virus, COVID-19, which rapidly developed into a pandemic as defined by the World Health Organization ("WHO").[2] COVID-19 is defined as:

> "A mild to severe respiratory illness that is caused by a coronavirus (*Severe acute respiratory syndrome coronavirus 2* of the genus *Betacoronavirus*), is transmitted chiefly by contact with infection material (such as respiratory droplets) or with objects or surfaces contaminated by the causative virus, and is characterized especially by fever, cough, and shortness of breath and may progress to pneumonia and respiratory failure."[3]

Due to the rapid spread of COVID-19, the WHO declared a global health emergency on January 30, 2020.[4] Since it's origination the virus has sickened nearly 4.3 million people across at least 177 countries with many more estimated to have carried the disease asymptomatically. At least 294,046 people have died.[5] Current medical experts anticipate that the disease will continue to spread over then next 18 months to 2 years, with multiple waves of infection and illness.[6] The pandemic could last potentially longer depending on the effectiveness of the global mitigation response measures.[7] See Figure 1 below for the impact upon mitigation response measures upon infection numbers and timeline.[8]

---

[1] Bryan Walsh, *Covid-19: The history of pandemics*, BBC FUTURE, March 25, 2020, *available at* https://www.bbc.com/future/article/20200325-covid-19-the-history-of-pandemics

[2] D.B. Taylor, *How the Coronavirus Pandemic Unfolded: a Timeline*, NYTIMES, April 28, 2020, *available at* nytimes.com/article/coronavirus-timeline.html

[3] *COVID-19*, MERRIAM-WEBSTER DICTIONARY, *available at* https://www.merriam-webster.com/dictionary/COVID-19.

[4] *Rf. Note 2.*

[5] Coronavirus (COVID-19), WHO, *as of 14 May, 2020, available at* https://covid19.who.int/

[6] Yasemin Saplakoglu, *Conoravirus pandemic could last over 18 months, according to a federal plan*, LIVE SCIENCE, March 21, 2020, *available at* https://www.livescience.com/coronavirus-pandemic-could-last-18-months.html

[7] A good report on Community Strategy measures for pre-pandemic mitigation in the US with an early presentation of the "flattening the curve" graph on page 18:
Interim Pre-pandemic Planning Guidance: Community Strategy for Pandemic Influenza Mitigation in the United States – Early, Targeted, Layered Use of Nonpharmaceutical Interventions, CDC & DEPT. OF HEALTH & HUMAN SERVS., February 2007, *available at* https://stacks.cdc.gov/view/cdc/11425.

[8] Dr. Sabine L. van Elsland, Ryan O'Hare, *COVID-19: Imperial researchers model likely impact of public health measures*, IMPERIAL COLLEGE LONDON NEWS, March 17, 2020, *available at* https://www.imperial.ac.uk/news/196234/covid19-imperial-researchers-model-likely-impact/.

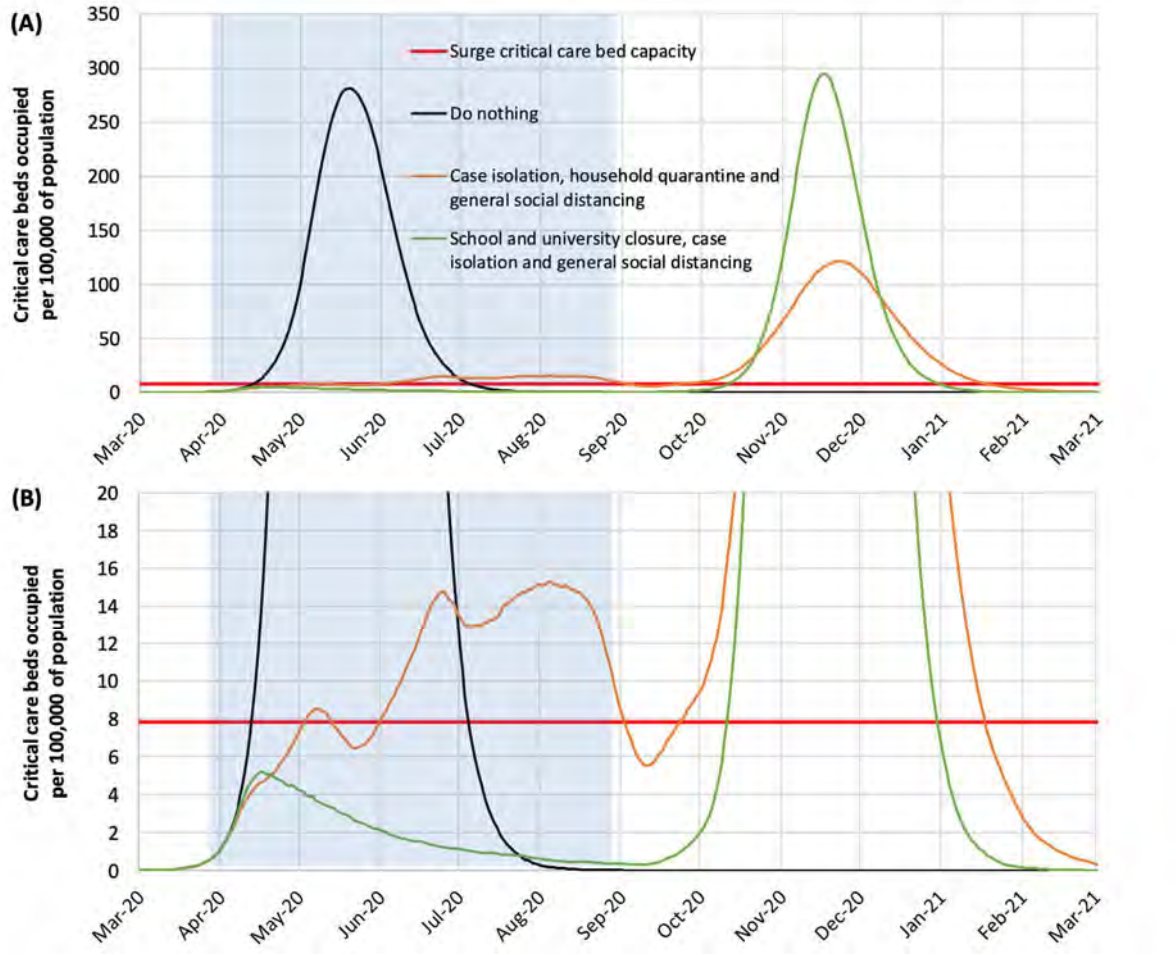


Fig. 1. Flattening the curve: The report highlights suppression strategy scenarios for the UK showing intensive care unit (ICU) bed requirements. The black line shows the unmitigated epidemic. Green shows a suppression strategy incorporating closure of schools and universities, case isolation and population-wide social distancing beginning in late March 2020. The orange line shows a containment strategy incorporating case isolation, household quarantine and population-wide social distancing. The red line is the estimated surge ICU bed capacity in GB. The blue shading shows the 5-month period in which these interventions are assumed to remain in place. (B) shows the same data as in panel (A) but zoomed in on the lower levels of the graph. (Source: WHO collaborating Centre / MRC GIDA / J-IDEA)

Current priorities in reducing the impact of the virus are the development of a vaccine and more advanced testing measures, including antibody tests, and stimulating an economic in the midst of an unprecedented halt to human movement.[9] Unfortunately, this is a novel disease, and researchers are still working to understanding the biology behind the disease, with numerous unknowns remaining including the nuances of transmission, infection development, and cases of death in young, otherwise healthy individuals, while others remain asymptomatic. Researchers seek a vaccine, treatments that minimize symptoms occurrence and severity and ultimately a cure. Experts warn that the development of a vaccine will likely take more than 18 months,[10] with infection and death rates expected to continue to grow.

[999] *Scientists Race for a Coronavirus Vaccine*, NYTIMES, April 27, 2020, *available at* https://www.nytimes.com/2020/04/27/world/coronavirus-world-pandemic.html?searchResultPosition=10.

[10] Nsikan Akpan, *Why a coronavirus vaccine could take way longer than a year*, NATIONAL GEOGRAPHIC, SCIENCE, April 10, 2020, *available at* https://www.nationalgeographic.com/science/2020/04/why-coronavirus-vaccine-could-take-way-longer-than-a-year/

This paper was intended to contain a review of available open source federal and state government data in the United States regarding the impact of socio-economic demographics upon infection and mortality rates from COVID-19 thus far. Due to the lack of available data and information on socio-economic groups across the country from a federal governmental source, individual state data sets will be used as case studies to illustrate the COVID-19's infection, hospitalization and mortality rates filtered through demographic information available on race & ethnicity, age, and gender.

Part II - Background

A. COVID-19: Current Context and Historical Import

This paper is being written in the midst of the economic shut-down caused by the COVID-19 crisis, where people around the world have been predominately confined to their homes for nearly two months in an effort to contain the spread of the disease. The 21st Century has not yet experienced a worldwide life-altering event like the mass infection and resultant societal shut down that has arisen from COVID-19. Our world has been forever changed and it is expected that daily human operation and systems will be dramatically reshaped after this pandemic has been contained.[11]

Historically, there are few notable events that have changed the world as we know it so dramatically that they create profound lasting changes in individual societies, let alone the world. Some examples of past revolutionizing events in the 20th Century include the Spanish Flu of 1918, the Great Depression, World War I and World War II, and the 9/11 terrorist attacks on the World Trade Centers in New York City.[12] No event in modern times has shut down the global economy and human transportation systems to the extent that COVID-19 has, and the ripple effects from this period are yet to be fully realized, particularly in fields such as medical research, health care, education, the global supply chain, food supply, employment type and availability, urban planning and public transportation systems, among many others.

B. Infection and Mortality Rates: Disparate Impact within Socio-Economic Classifications

One aspect of major historical moments, particularly crises,[13] that remains universal, yet often ignored or unaddressed in manner that substantively changes the systematic frameworks that support it, is the disparate negative impact upon marginalized and economically disadvantaged

---

[11] *Coronavirus Will Change the World Permanently. Here's How*., POLITICO MAGAZINE, *available at* https://www.politico.com/news/magazine/2020/03/19/coronavirus-effect-economy-life-society-analysis-covid-135579.

[12] *Events that changed the world*, BIOGRAPHY ONLINE, https://www.biographyonline.net/events/changed-world.html; Kelly Knauer, *History's Greatest Events: 100 Turning Points that Changed the World: An Illustrated Journey*, TIME, Oct. 12, 2010.

[13] Carmin Chappell, *Climate Change in the US will hurt poor people the most, according to a bombshell federal report*, CNBC, Nov. 26, 2018, *available at* https://www.cnbc.com/2018/11/26/climate-change-will-hurt-poor-people-the-most-federal-report.html; Campbell Robertson, *Racially Disparate Views of New Orleans's Recover After Hurricane Katrina*, NYTIMES, 25 Aug. 2015, *available at* https://www.nytimes.com/2015/08/25/us/hurricane-katrina-new-orleans-recovery-ten-years-opinion.html; Nazila Ghanea, James A. Goldston, Mumtaz Lalani and Preti Taneja, *Religions Minorities in a post-9/11 world*, State of the World's Minorities and Indigenous Peoples 2010, MINORITY RIGHTS GROUP INTERNATIONAL, *available at* https://minorityrights.org/wp-content/uploads/old-site-downloads/download-842-Religious-minorities-in-a-post-911-world.pdf

populations. At a time when the US economy is at its most economically divided,[14] with economic inequality at levels not experienced in the US since the pre-Great Depression roaring 20s,[15] and with a large number of the population uninsured, COVID-19 is reported to be disproportionately impacting low-income communities and minorities in both infection rates and mortality rates as well as via the economic consequences of the shutdown. Lower-income workers are disproportionately represented in jobs classified as "essential" thus increasing these workers' potential exposure to the virus, and has already resulted in increased infection rates among them.[16] This paper will explore the statistical data available to test the validity of these reports. If a disparate impact is determined given a statistical analysis of the available data, then the extent of the impact will be further evaluated.

The CDC has acknowledged the import of socio-economic factors upon COVID-19 risk factors with the following statement on their website:

> *"Because COVID-19 is a new disease, more work is needed to better understand the risk factors for severe illness or complications. Potential risk factors that have been identified to date include:*
>
> - *Age*
> - *Race/ethnicity*
> - *Gender*
> - *Some medical conditions*
> - *Use of certain medications*
> - *Poverty and crowding*
> - *Certain occupations*
> - *Pregnancy*
>
> *Additional research will help us confirm if these are risk factors for severe COVID-19 illness and determine if there are other factors that increase a person's risk."*[17]

C. U.S. COVID-19 Timeline

The first death from the coronavirus in the US was confirmed on the 29th of February, 2020 in the state of Washington.[18] However, later testing revealed that the first US deaths occurred in early February, meaning that infection were present in the US in January.[19] There is speculation that the

---

[14] Juliana Menasce Horowitz, Ruth Igielnik and Rakesh Kochhar, *Trends in Income and Wealth Inequality*, PEW RESEARCH CENTER, January 9, 2020, *available at* https://www.pewsocialtrends.org/2020/01/09/trends-in-income-and-wealth-inequality/.

[15] Jesse Colombo, *America's Wealth Inequality Is At Roaring Twenties Levels*, FORBES, Feb. 28, 2019, *available at* https://www.forbes.com/sites/jessecolombo/2019/02/28/americas-wealth-inequality-is-at-roaring-twenties-levels/#7a4131972a9c.

[16] Abha Bhattarai, *'It feels like a war zone': As more of them die, grocery workers increasingly fear showing up at work*, WASHINGTON POST, April 12, 2020, *available at* https://www.washingtonpost.com/business/2020/04/12/grocery-worker-fear-death-coronavirus/.

[17] CDC, *Assessing Risk Factors*, *available at* https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/assessing-risk-factors.html.

[18] Dakin Andone, Jamie Gumbrecht and Michael Nedelman, *First death from coronavirus in the US confirmed in Washington State*, CNN, 29 Feb. 2020, *available at* https://www.cnn.com/2020/02/29/health/us-coronavirus-saturday/index.html.

[19] Sharon Bernstein, Kanishka Singh, *Coronavirus circulated in U.S. weeks earlier than thought, mistaken for flu, health officials say*, REUTERS, April 22, 2020, *available at* https://www.reuters.com/article/us-health-coronavirus-usa-california-idUSKCN2240Z6; Thomas Fuller and Mike Baker, *Coronavirus Death in California*

earliest COVID-19 cases were mistaken for the flu. In the United States alone, as of May 13, 2020, the Center for Disease Control ("CDC") estimates the total COVID-19 cases at 1,364,061 and the total deaths at 82,246.[20]

D. Definitions

Coronavirus has swept across the globe with the ferocity of a tornado. It has moved through our world with a speed and force that introduced a new way of existing and thinking about our systematic structures. New vocabularies have been introduced and become commonplace throughout the society due to the novel aspects of this unique crisis. Daily syntax has incorporated phrases such as "flattening the curve", "herd immunity", "social distancing", and "antibody resistance". Statistics and data science have played a central role in the visual and verbal communication surrounding the disease, particularly in reference to governmental and medical mitigation and response activities. Hence, more technical language has been used as daily information sessions for the public are held by medical and government officials. Below are a few definitions that are newly in frequent use due to COVID-19:[21]

- **COVID-19**:
- **Outbreak**: a sudden rise in the incidence of a disease.
- **Epidemic**: the rapid spreading of a disease among a region or certain population.
- **Pandemic**: an epidemic that has become a worldwide phenomenon.
- **Incubation period**: the time between infection and showing symptoms of illness. (An incubation period of 2-14 days is estimated for the coronavirus with symptoms most commonly showing at about five days after infection, according to the WHO).
- **Self-quarantine**: usually a voluntary confinement of oneself by staying at home.
- **Stay at home/shelter in place**: A governmental order used to limit residents from leaving their homes, except out of necessity for tasks like buying groceries or seeking medical care.
- **Non-essential/Essential**: Businesses, services and employees defined as essential are allowed to continue operations, while "non-essential" counterparts are required to close or remain at home.
- **Asymptomatic**: not exhibiting symptoms while having a disease.
- **Community spread**: the spread of an illness for which the source of infection is unknown.
- **Morbidity rate**: a measure of how many people have an illness relative to the population.
- **Mortality rate**: a measure of how many people have died because of an illness, also relative to the population

---

Came Weeks Before First Known U.S. Death, NYTIMES, April 22, 2020, *available at* https://www.nytimes.com/2020/04/22/us/coronavirus-first-united-states-death.html.

[20] *Cases in the U.S.,* CENTER FOR DISEASE CONTROL, May 3, 2020, *available at* https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. *Refer* fn.1 for a more complete global timeline.

[21] For a more in depth discussion of the COVID-19 specific definition refer to the following: Katy Steinmetz, *Coronavirus: A Glossary of Terms to Help You Understand the Unfolding Crisis*, TIME, March 23, 2020, *available at* time.com/5798684/coronavirus-glossary-definitions/; Nick Coltrain, *What is "social distancing"? What does it mean to "shelter in place"? A glossary of COVID-19 terms*, DES MOINES REGISTER, March 11, 2020, *available at* Desmoinesregister.com/story/news/health/2020/03/11/coronavirus-in-america-covid-definitions-presumptive-isolation-quarantine-isolation-cdc-who/5012006002/

- **Social distancing**: courses of action designed to limit when and where people gather. During the COVID-19 outbreak a distance of at least 6 feet from other people has been recommended.
- **Confidence Interval**:
- **Flatten the curve**: the curve specific to COVID-19 is the rate at which people become infected with the novel coronavirus. It is preferred to have a low and long curve than a high and narrow curve. A high and narrow curve would exceed the capacity of the health care system resulting in more people getting sick and dying overall. The most common image in circulation is below:[22]

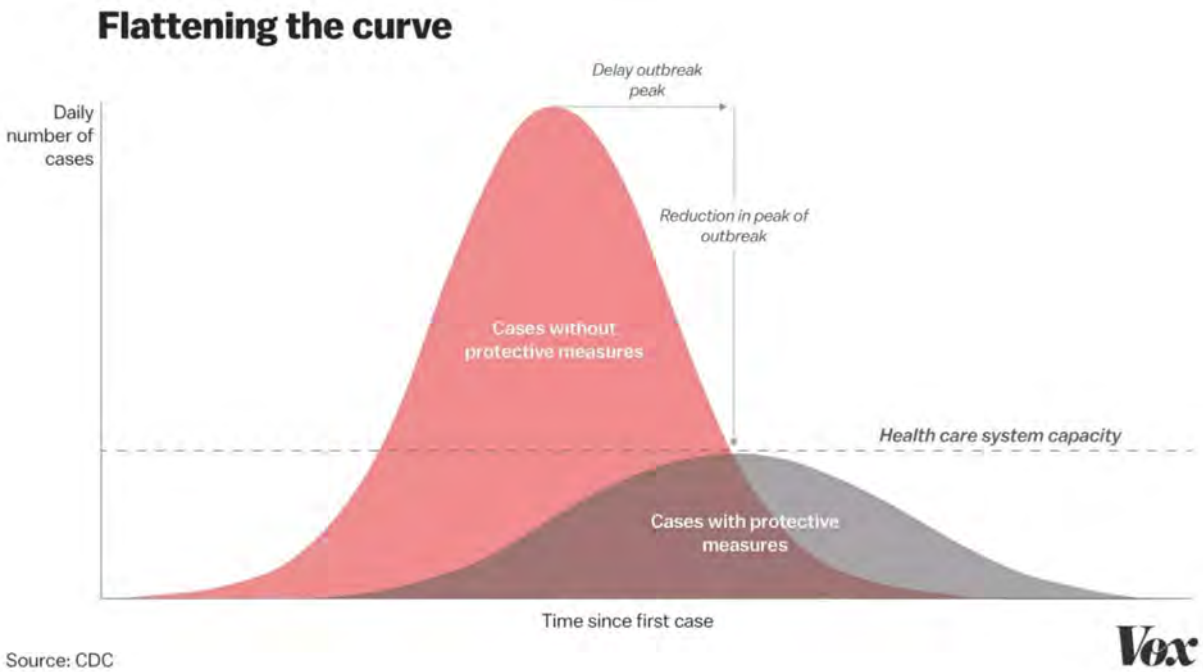


Figure 1 (Flattening the Curve from Vox)[23]

- **Zoonotic disease**: a disease which exists in animals but can also infect humans. It is believed that COVID-19 originated in an animal and was spread to humans.
- **Vaccine**: any preparation used as a preventive inoculation to confer immunity against a specific disease, usually employing an innocuous form of the disease agent, as killed or weakened bacteria or viruses, to stimulate antibody production.[24]
- **Herd immunity**: when enough people in a given population have taken an effective vaccine, the community develops herd, or group, immunity.
- **PPE**: Personal protective equipment, such as face masks, face shields, gloves, isolation gowns and N95 respirators. According to the CDC there has been a continuous shortage of PPE throughout the pandemic leaving many medical professionals unprotected and creating bidding wars between state and international governments.

[22] Brandon Spektor, *Coronavirus: What is 'flattening the curve,' and will it work?,* LIVESCIENCE, March16, 2020, *available at* https://www.livescience.com/coronavirus-flatten-the-curve.html.
[23] *Flattening the Curve Won't End the Pandemic*, FINANCIAL LITERACY FOR PHYSICIANS, *available at* https://www.fiphysician.com/flatten-the-curve/.
[24] Vaccine definition, *available at* https://www.dictionary.com/browse/vaccine?s=t

E. Current Context: COVID-19 Data Collection in the US

1. Overwhelmed Systems Result in a Lack of Reliable Available Data

The rapid and widespread infection rate of this disease has resulted in medical and governmental systems being overwhelmed and frantic. Every resource is being expended to keep systems from drowning, to various levels of success. Hospital workers report working incredibly long hours, with insufficient PPE (leading to high health care worker infection rates)[25] and other required resources at their facilities. Pop-up hospitals have been constructed around the country to help meet demand at the same time the economy is nose-diving and some hospitals and governments already on the edge of financial constraints are faced with maximum demand with minimal resources.[26] This context is important to the logistics of data gathering. In this time literal life or death decision are being made based off of the available data analysis. Individuals are making decisions based on the data presented at daily presidential or governor briefings and are trying to make sense of this crises through their understanding of graphs and statistical graphics projecting the disease trajectory. Everyone is looking for answers in the midst of an unknown future. Enter data projection at center stage.

However, with health care and governmental systems traditionally responsible for collecting and inputting data overwhelmed and understaffed, the reliability of the data available may be undermined. The accuracy of the reporting may be at question as well, particularly when we are aware that the current cases are underreported due to a lack of available COVID-19 tests, and with a growing number of people shown to be positive and asymptomatic, and an exhausted staff that may neglect the administrative duties at the end of an otherwise harrowing shift.[27] Additionally, nuances within the context for the data, the limitation of the models being used, and the educational background required for accurate interpretation of the numbers are being overlooked.

2. International, National, State and Local Data Collection

The WHO has been the primary international leader to collect and disseminate information internationally on COVID-19. However, the data presented by the WHO is based on confirmed cases and deaths within each country and does not examine more detailed demographic breakdowns, they do have a global database of current research and academic publications related to COVID-19, unfortunately despite having hundreds of articles on COVID-19, advanced searches found not one of them focused on the socio-economic impact of the pandemic.

The CDC is the primary authority within the United States at a federal level for data collection and analysis on COVID-19. Initially, the CDC was not collecting information on any demographic breakdown until there was a demand by democratic senators to collect data on racial disparities in

---

[25] Newsday Staff*, Nearly 1,200 hospital staff on Long Island have contracted COVID-19*, NEWSDAY, April 13, 2020, *available at* https://www.newsday.com/news/health/coronavirus/hospital-covid-19-nurses-1.43778185.

[26] *Photos: How the world is mobilizing to make pop-up hospitals*, LOS ANGELES TIMES, March 31, 2020*, available at* https://www.latimes.com/world-nation/story/2020-03-31/construction-of-hospitals-around-the-world-ramps-up-to-handle-influx-of-coronavirus-patients

[27] Hollie Silverman, *More than 370 workers at a pork plant in Missouri tested positive for coronavirus. All were asymptomatic*, CNN, May 4, 2020, *available at* https://www.cnn.com/2020/05/04/us/triumph-foods-outbreak-missouri/index.html.

America's coronavirus response.[28] The CDC is now providing an open dataset on COVID-19 and race. This dataset will be used as the case study for racial statistical breakdowns in Part III.

Various state Departments of Health have provided open datasets with further demographic information. All of the state Departments of Health provide demographic information, however not all states provide open source access to their datasets to run independent analyses. Of the few states that do provide open dataset access, their available data is often limited, providing information on only one demographic group. Hence, Part III will contain case studies from Connecticut on gender and the CDC on race and ethnicity, with summary statistics and graphs provided by other individual states. There were no open datasets available conducting income-level information.

A number of datasets have been collated by private companies, universities and other organization in an attempt to create more comprehensive datasets and data-based competitions have been created to aid in the efforts to further understand the impacts of this virus.[29] These datasets continue to evolve and will likely provide the most comprehensive open datasets in the long term but at this point are still in their creation phase.

3. ***Incomplete Information***

Although many States have been analyzing and reporting on the basic demographics of age, race, and gender, there are still many other areas remaining with a severe lack of information. For example, how is COVID-19 disproportionately impacting farmers, health care workers, "essential workers", morticians, rural/urban populations, immigrants, domestic violence rates, etc.? Without this information being gathered at this point, much of this information will never be available and we will base future studies off of observational studies with uncontrolled variables, if this information is analyzed at all.

It is assumed that the deaths related to COVID-19 are also severely underrepresented in the data for a number of reasons. First, persons not seeking medical treatment may be passing at home and never having their deaths recording as COVID-19 related because they were never tested nor received treatment. Since there have been recorded transmission of COVID-19 from a corpse to a living human there are some cities and states that are testing for COVID-19 post-mortem, however many are still not, in part due to the lack of available tests. Further, since COVID-19 is a novel disease, the tests are also new and have not had the time to be refined. One study conducted by the Cleveland Clinic found the current COVID-19 tests have a false-negative rate of 14.8%.[30]

Second, comorbidity factors can influence the recorded reason for deaths. For example, some patients have been presenting classic symptoms of a heart attack on electrocardiograms but upon closer examination it was determined that more than half of the patients did not have a blockage in

[28] Jason Silverstein, *Democrats demand data on racial disparities in America's coronavirus response*, CBS NEWS, March 31, 2020, *available at* https://www.cbsnews.com/news/democrats-demand-data-on-racial-disparities-in-americas-coronavirus-response/

[29] NYTimes Covid-19 dataset; COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

[30] Rob Stein, *Study Raises Question About False Negatives From Quick COVID-19 Test*, NPR MORNING EDITION, April 21, 2020, *available at* https://www.npr.org/sections/health-shots/2020/04/21/838794281/study-raises-questions-about-false-negatives-from-quick-covid-19-test.

a major artery, the typical cause of a heart attack. They were later determined to be the result of COVID-19 infection.[31]

Third, a large number of people infected with COVID-19 have no symptoms.[32] Many people will have been infected with COVID-19 and never know it. However, there may be later complications that develop from them having been exposed to the disease that will not be recorded as such.

This paper will conduct the analysis based on the information available, with acknowledgement that the data could represent an underestimate of the true population numbers.

Part III – Data & Analysis Methodology

A. **Research Focus for the Dataset Analysis**

1. Are different socio-economic groups being disproportionately sickened, hospitalized, and dying due to COVID-19?
2. If so, which groups are disproportionately dying from COVID-19, and
3. **to what extent** are socio-economic demographic factors impacting the COVID-19 infection, hospitalization and mortality rates in the United States?

B. **Parameters Explored in the Datasets**

There is an overall lack of socio-economic data being recorded at this time. Race & Ethnicity, Gender and Age are the only socio-economic categories with available COVID-19 datasets. Hence, the analysis will be limited to these three categories.

Within these categories the investigation will focus upon percentage of the overall population compared to the percentage of people infected, hospitalized and killed as a result of COVID-19.

C. About the Data

1. Race & Ethnicity

The datasets utilized for this analysis were created within the general context described in Part II.E. of this paper. The dataset on race & ethnicity was produced by the CDC and reviews data for 18 different states reporting on race. The dataset provides information based on State for three categories: 1. Distribution of COVID deaths (%), 2. Weighted distribution of population (%), and 3. Unweighted distribution of population (%). The CDC provides the follow information about the data:

> *The Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET) conducts population-based surveillance for laboratory-confirmed COVID-19-associated hospitalizations in children (persons younger than 18 years) and adults. The current network covers nearly 100 counties in the 10 Emerging Infections Program (EIP) states (CA, CO, CT, GA, MD, MN, NM, NY, OR, and TN) and four additional*

[31] E.J. Mundell, *It Looks Like a Classic Heart Attack, But in COVID-19 Patients Other Issues May be at Play*, U.S. NEWS & WORLD REPORT, April 20, 2020, available at https://www.usnews.com/news/health-news/articles/2020-04-20/it-looks-like-a-classic-heart-attack-but-in-covid-19-patients-other-issues-may-be-at-play.

[32] Bruce Y. Lee, Study: 17.9% of People With COVID-19 Coronavirus Had No Symptoms, Forbes, March 18, 2020, available at https://www.forbes.com/sites/brucelee/2020/03/18/what-percentage-have-covid-19-coronavirus-but-do-not-know-it/#2d3a9b6b7e90

states through the Influenza Hospitalization Surveillance Project (IA, MI, OH, and UT). The network represents approximately 10% of US population (~32 million people).

Cases are identified by reviewing hospital, laboratory, and admission databases and infection control logs for patients hospitalized with a documented positive SARS-CoV-2 test.

Data gathered are used to estimate age-specific hospitalization rates on a weekly basis and describe characteristics of persons hospitalized with COVID-19. Laboratory confirmation is dependent on clinician-ordered SARS-CoV-2 testing. Therefore, the rates provided are likely to be underestimated as COVID-19-associated hospitalizations can be missed due to test availability and provider or facility testing practices.

COVID-NET hospitalization data are preliminary and subject to change as more data become available. In particular, case counts and rates for recent hospital admissions are subject to lag. As data are received each week, prior case counts and rates are updated accordingly.[33]

Selected information and statistics from individual state Departments of Health, including Maryland, will be provided in the Race Analysis section below.

2. Gender

The CDC is not collecting information based on Gender. Many of the individual state Department's of Health are also not providing information based on gender for COVID-19. Some states are providing summary statistics based on gender, such as New York State, California, New York City, and Maryland. That summary information is available in the Gender Analysis section below. Connecticut's is the one state that was found to provide an open dataset with information from their Department of Health on COVID-19 and Gender. That information is also summary though, providing simply a running tally of infection and death rates by the day.

Connecticut provides the following information about the dataset:

Laboratory-confirmed cases of COVID-19 that have been reported among Connecticut residents, broken down by gender. All data in this report are preliminary; data for previous dates will be updated as new reports are received and data errors are corrected. Hospitalization data were collected by the Connecticut Hospital Association. Deaths reported to the either the Office of the Chief Medical Examiner (OCME) or Department of Public Health (DPH) are included in the daily COVID-19 update.

Data are reported daily, with timestamps indicated in the daily briefings posted at: portal.ct.gov/coronavirus. Data are subject to future revision as reporting changes.[34]

---

[33] *COVID-NET: COVID-19-Associated Hospitalization Surveillance Network, Centers for Disease Control and Prevention, available at https://gis.cdc.gov/grasp/COVIDNet/COVID19_5.html* Accessed on May *6, 2020.*

[34] Connecticut Department of Health, COVID-19 confirmed cases and deaths by gender, *available at* https://data.ct.gov/Health-and-Human-Services/COVID-19-confirmed-cases-and-deaths-by-gender/qa53-fghg

3. Age

The CDC is not providing access to information about age at this time. Many states are providing summary statistics, tables and graphs based on age. Selected examples will be provided in the Age Analysis section below, including information from Maryland's Department of Health. However, there are few available datasets with this information. One dataset from the Washington State Department of Health was found with the breakdown of infection and death rates by age per county.

Washington State provides the following notes on their Data:

> ***Note on the county and unassigned data:*** *This data changes rapidly as labs conduct tests and discover new cases. Labs assign those cases to a county. Counties or the Department of Health then determine the appropriate county of jurisdiction. Those don't always match initially. We're working to reduce the "unassigned" number to 0. Contact the local health department for county specific information.*
>
> ***Note on the deaths:*** *Some deaths may be reported by health care providers, medical examiners/coroners, local health departments, or others before they are included in the statewide count. It takes longer for the state to announce deaths because they are often reported first to the local health department and then to us.*
>
> ***Note on the number of infections:*** *Public health experts agree that the true number of people who have been infected with COVID-19 in Washington greatly exceeds the number of COVID-19 infections that have been laboratory-confirmed. It is very difficult to know exactly how many people in Washington have been infected to date since most people with COVID-19 experience mild illness and the ability to get tested is still not widely available.*[35]

Due to the categorization of ages by range – this dataset was ultimately not used in this report but may be useful for future studies.

D. Analysis

1. Race & Ethnicity

Eleven years ago, there was a pandemic outbreak of H1N1 otherwise known as "swine flu". In the US, H1N1 had a disproportionate and severe impact upon communities of color. Compared to white patients, nonwhite patients became sicker faster, recovered more slowly and died at higher rates.[36]

Some of the key reasons why the previous outbreak impacted communities of color more severely were identified by epidemiologists as[37]:

F. Nonwhite workers had less access to sick leave, with Hispanic workers having the least access
G. Nonwhite workers had more public-facing jobs
H. They are three times as likely as white workers to live in apartments instead of homes
I. They are twice as reliant on public transportation

---

[35] Washington State Department of Health, COVID-19 data, *available at* https://www.doh.wa.gov/Emergencies/Coronavirus

[36] Sidney Fussell, *The H1N1 Crisis Predicted Covid-19's Toll on Black Americans*, WIRED, May 6, 2020, *available at* https://www.wired.com/story/h1n1-crisis-predicted-covid-19-toll-black-americans/.

[37] *Id. (all bullet points below are referenced from the previous citation).*

J. They were more likely to be immunocompromised
K. Their environments contributed to elevated risk of other health problems including hypertension, heart disease, and asthma that aggravated the flu's symptoms.
L. They are less likely to have access to vaccines and testing.

Unfortunately, it appears that access to this knowledge did not help the US prepare for the Covid-19 pandemic. Some examples that have been reported include[38]:

M. Washington, D.C. where Black residents account for 45 percent of but make up almost 80 percent of COVID-19 fatalities;
N. Michigan, Black residents account for14 percent of the population but 40 percent of COVID-19 fatalities;
O. Richmond, VA where 40 percent of residents are Black, yet account for all but one of the fatalities.
P. Mission district of San Francisco where Latinx peoples make up less than half of the population but account for 95% of the positive COVID-19 tests.

This section reviews the impact of race upon COVID-19 infection, hospitalization and death rates, based on the current available data.

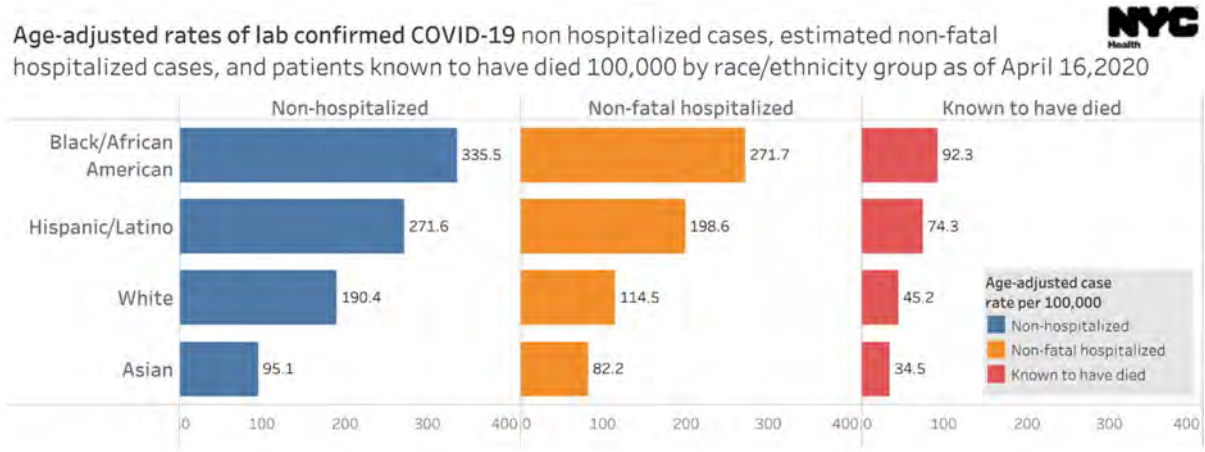
a. Infection

Demographic characteristics of COVID-19 cases in the U.S.

	No. of cases (% of total)						
Age group (years)	< 18	18-44	45-64	65-74	75+	Unknown	Total
Totals	19,601	308,970	299,786	93,018	106,681	54,793	882,849
Race missing/unspecified	13,221 (67.5%)	194,433 (62.9%)	176,183 (58.8%)	45,469 (48.9%)	43,952 (41.2%)	18,399 (33.6%)	491,657 (55.7%)
Race specified	6,380 (32.5%)	114,537 (37.1%)	123,603 (41.2%)	47,549 (51.1%)	62,729 (58.8%)	36,394 (66.4%)	391,192 (44.3%)
Among those with race specified							
American Indian or Alaska Native	145 (2.3%)	1,211 (1.1%)	1,008 (0.8%)	276 (0.6%)	184 (0.3%)	141 (0.4%)	2,965 (0.8%)
Asian	296 (4.6%)	5,968 (5.2%)	6,499 (5.3%)	2,151 (4.5%)	2,299 (3.7%)	1,737 (4.8%)	18,950 (4.8%)
Black or African American	1,475 (23.1%)	31,444 (27.5%)	38,166 (30.9%)	15,040 (31.6%)	13,397 (21.4%)	11,734 (32.2%)	111,256 (28.4%)
Native Hawaiian or other Pacific Islander	45 (0.7%)	475 (0.4%)	355 (0.3%)	108 (0.2%)	69 (0.1%)	78 (0.2%)	1,130 (0.3%)
White	3,158 (49.5%)	56,838 (49.6%)	60,887 (49.3%)	25,306 (53.2%)	42,068 (67.1%)	15,956 (43.8%)	204,213 (52.2%)
Multiple/other	1,261 (19.8%)	18,601 (16.2%)	16,688 (13.5%)	4,668 (9.8%)	4,712 (7.5%)	6,748 (18.5%)	52,678 (13.5%)
Ethnicity missing/unspecified	12,891 (65.8%)	201,060 (65.1%)	187,075 (62.4%)	51,865 (55.8%)	54,991 (51.5%)	21,863 (39.9%)	529,745 (60.0%)
Ethnicity specified	6,710 (34.2%)	107,910 (34.9%)	112,711 (37.6%)	41,153 (44.2%)	51,690 (48.5%)	32,930 (60.1%)	353,104 (40.0%)
Among those with ethnicity specified							
Hispanic/Latino	2,662 (39.7%)	34,230 (31.7%)	28,465 (25.3%)	7,272 (17.7%)	6,509 (12.6%)	10,403 (31.6%)	89,541 (25.4%)
Non-Hispanic/Latino	4,048 (60.3%)	73,680 (68.3%)	84,246 (74.7%)	33,881 (82.3%)	45,181 (87.4%)	22,527 (68.4%)	263,563 (74.6%)

https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html
These numbers produced by the CDC were gathered from 55 U.S.-affiliated jurisdictions currently reporting cases of COVID-19. This includes 50 states, District of Columbia, Guam, the Northern Mariana Islands, Puerto Rico, and the U.S Virgin Islands.

[38] *Id. (all bullet points below are referenced from the previous citation).*



https://www.vox.com/coronavirus-covid19/2020/4/18/21226225/coronavirus-black-cdc-infection

Chi-Square Test

```r
# Chi Square test for the probability of infection from COVID-19 in the US based on Race
null.probs = c(76.5/108.2, 13.4/108.2, 18.3/108.2)
CDCcases = c(52.20, 28.4, 25.4)
chisq.test(CDCcases, p=null.probs)
```

```
        Chi-squared test for given probabilities

data:  CDCcases
X-squared = 27.785, df = 2, p-value = 9.26e-07
```

The Chi-square test applied to the summary statistics from the CDC, resulting in a p-value ≈ 0, indicates there is very strong evidence that race and ethnicity are positively related to COVID-19 infection throughout the US.

The Odds Ratio – US

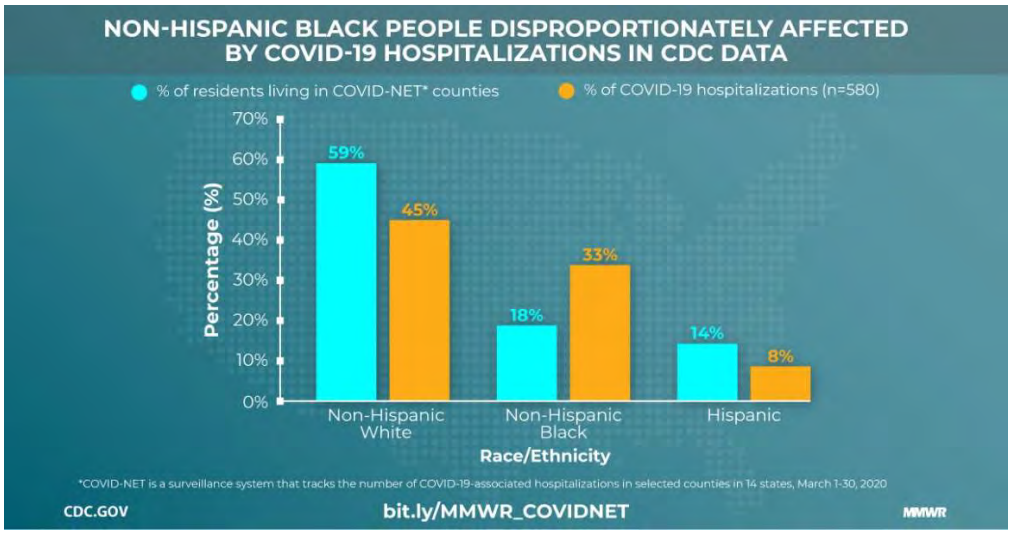
	Odds Ratio	B	B'				
		White	Black	Total			
A	Covid 19	204213.00	111256.00	315469.00			
A'	No Covid 19	235986123.57	41260646.09	277246769.66	Population Ratio	5.71	There are 5.7 times more white people than African Americans in the US
		236190336.57	41371902.09	277562238.66	Odds Ratio Cases	3.12	African Americans are 3.12 times more likely to get Covid-19 than whites

	Odds Ratio	B	B'				
		White	Hispanic				
A	Covid 19	204213.00	89541.00	293754.00	Population Ratio	4.18	There are 4.18 times more white people than Latinx people in the US
A'	No Covid 19	235986123.57	56410892.45	292397016.02	Odds Ratio Cases	1.83	Latinx people are 1.83 times more likely to get Covid-19 than whites
		236190336.57	56500433.45	292690770.02			

The Odds Ratio calculations based off of the summary statistics provided by the CDC support the conclusion that there is strong evidence to suggest race and ethnicity are related to COVID-19 infection and death rates throughout the United States.

b. Hospitalization

Summary statistics provided by the CDC are shown in the graph below indicating the hospitalization rates based on race and ethnicity.



https://www.cdc.gov/coronavirus/2019-ncov/covid-data/data-visualization.htm

c. Death – California, New York and Maryland

California

COVID-19 Race and Ethnicity Data

All Cases and Deaths associated with COVID-19 by Race and Ethnicity

May 4, 2020

Race/Ethnicity	No. Cases	Percent Cases	No. Deaths	Percent Deaths	Percent CA population
Latino	18,087	48	740	34.6	38.9
White	9,724	25.8	744	34.8	36.6
Asian	4,383	11.6	361	16.9	15.4
African American/Black	2,363	6.3	220	10.3	6
Multi-Race	324	0.9	8	0.4	2.2
American Indian or Alaska Native	72	0.2	7	0.3	0.5
Native Hawaiian and other Pacific Islander	420	1.1	20	0.9	0.3
Other	2,275	6	36	1.7	0
Total with data	37,648	100	2,136	100	100

Cases: 56212 total, 18564 (33%) missing race/ethnicity

Death: 2221 total, 85 (4%) missing race/ethnicity

***Census data does not include 'other race' category**

Proportions of Cases and Deaths by Race and Ethnicity Among Ages 0-17

Race/Ethnicity	No. Cases	Percent Cases	No. Deaths	Percent Deaths	Percent CA population
Latino	776	75.2	0	0	47.9
White	122	11.8	0	0	29.2
Asian	51	4.9	0	0	12.7
African American/Black	22	2.1	0	0	5.4
Multi-Race	9	0.9	0	0	4
American Indian or Alaska Native	4	0.4	0	0	0.4
Native Hawaiian and other Pacific Islander	1	0.1	0	0	0.3
Other	47	4.6	0	0	0
Total	1032	100	0	0	100

Cases: 1773 total, 741 (42%) unknown race/ethnicity

Deaths: NA total; NA (NA%) unknown race/ethnicity

***Census data does not include 'other race' category**

Proportions of Cases and Deaths by Race and Ethnicity Among Ages 18-49

Race/Ethnicity	No. Cases	Percent Cases	No. Deaths	Percent Deaths	Percent CA population
Latino	10,198	55.5	94	69.1	43.5
White	3,702	20.1	11	8.1	31.2
Asian	1,914	10.4	10	7.4	15.9
African American/Black	963	5.2	18	13.2	6.3
Multi-Race	185	1	0	0	2.2
American Indian or Alaska Native	40	0.2	2	1.5	0.6
Native Hawaiian and other Pacific Islander	210	1.1	1	0.7	0.4
Other	1,172	6.4	0	0	0
Total	18,384	100	136	100	100

Cases: 27556 total, 9172 (33%) unknown race/ethnicity

Deaths: 139 total, 6 (4%) unknown race/ethnicity

***Census data does not include 'other race' category**

Proportions of Cases and Deaths by Race and Ethnicity Among Ages 50-64

Race/Ethnicity	No. Cases	Percent Cases	No. Deaths	Percent Deaths	Percent CA population
Latino	4,582	46.2	159	49.1	32.2
White	2,571	25.9	74	22.8	43
Asian	1,227	12.4	44	13.6	16.3
African American/Black	683	6.9	36	11.1	6.4
Multi-Race	78	0.8	1	0.3	1.2
American Indian or Alaska Native	15	0.2	1	0.3	0.5
Native Hawaiian and other Pacific Islander	128	1.3	3	0.9	0.4
Other	637	6.4	6	1.9	0
Total	9,921	100	324	100	100

Cases: 14366 total, 4445 (31%) unknown race/ethnicity

Deaths: 342 total, 18 (5%) unknown race/ethnicity

***Census data does not include 'other race' category**

Proportions of Cases and Deaths by Race and Ethnicity Among Ages 65+

Race/Ethnicity	No. Cases	Percent Cases	No. Deaths	Percent Deaths	Percent CA population
Latino	2,524	30.5	487	29.1	21.2
White	3,323	40.1	659	39.3	54.8
Asian	1,182	14.3	307	18.3	16.9
African American/Black	695	8.4	166	9.9	5.3
Multi-Race	51	0.6	7	0.4	1
American Indian or Alaska Native	13	0.2	4	0.2	0.5
Native Hawaiian and other Pacific Islander	81	1	16	1	0.3
Other	417	5	30	1.8	0
Total	8,286	100	1,676	100	100

Cases: 12425 total, 4139 (33%) unknown race/ethnicity

Deaths: 1738 total, 62 (4%) unknown race/ethnicity

***Census data does not include 'other race' category**

https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/Race-Ethnicity.aspx

Chi -Square Test - California

```r
# Chi Square test for The probability of deaths from COVID-19 in CA based on Race
null.probs = c(38.9/100, 36.6/100, 15.4/100, 6/100, 2.2/100, 0.5/100, 0.3/100, 0.1/100)
CAdeaths = c(34.6, 34.8, 16.9, 10.3, 0.4, 0.3, 0.9, 1.7)
chisq.test(CAdeaths, p=null.probs)
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  CAdeaths
X-squared = 32.176, df = 7, p-value = 3.767e-05
```

```r
# Chi Square test for the probability of contracting COVID-19 in CA based on Race
null.probs = c(38.9/100, 36.6/100, 15.4/100, 6/100, 2.2/100, 0.5/100, 0.3/100, 0.1/100)
CAcases = c(48, 25.8, 11.6, 6.3, 0.9, 0.2, 1.1, 6)
chisq.test(CAcases, p=null.probs)
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  CAcases
X-squared = 357.81, df = 7, p-value < 2.2e-16
```

The Chi-square test applied to the summary statistics from the California Department of Health resulting in a p-values ≈ 0, indicate there is very strong evidence that race and ethnicity are positively related to both COVID-19 infection and deaths in California.

The Odds Ratio – California

CA	Odds Ratio	B	B'			
		White	Black			
A	Covid 19 cases	9724.00	2363.00	Population Ratio	6.10	6 times more white people than African Americans
A'	No Covid 19 cases	14451749.62	2368370.38	Odds Ratio: cases	1.48	African Americans are 1.5 times as likely to have Covid-19
		14461473.62	2370733.38			

CA	Odds Ratio	B	B'			
		White	Hispanic			
A	Covid 19 cases	9724.00	18087.00	Population Ratio	0.94	0.94 times as many white people as Latinx in CA
A'	No Covid 19 cases	14451749.62	15352167.75	Odds Ratio: cases	1.75	Latinex are 1.75 times as likely as white people to have Covid in CA
		14461473.62	15370254.75			

CA	Odds Ratio	B	B'			
		White	Black			
A	Covid 19 death	744.00	220.00	Population Ratio	6.10	6 times more white people than African Americans
A'	No Covid 19 death	14460729.62	2370513.38	Odds Ratio: deaths	1.80	African Americans are 1.8 times as likely to die of Covid-19
		14461473.62	2370733.38			

CA	Odds Ratio	B	B'			
		White	Hispanic	Population Ratio	0.94	There are .94 times as many whites as Latinex in CA
A	Covid 19 death	744.00	740.00	Odds Ratio: deaths	0.94	Latinex are .94 times as likely as whites to die of Covid-19
A'	No Covid 19 death	14460729.62	15369514.75			
		14461473.62	15370254.75			

The Odds Ratio calculations based off of the summary statistics provided by the California Department of Health support the conclusion that there is strong evidence to suggest race and ethnicity are related to COVID-19 infection and death rates in CA.

New York

The following summary statistics are provided by the State of New York Department of Health.

Testing data as of: 5/04/2020 Midnight
Testing data last updated on: 5/5/2020
(Updated daily before 2 PM)

Fatalities by Race/Ethnicity *Data is preliminary. With 99% reporting, below is the breakdown for NYS excluding NYC. With 63% reporting, below is the breakdown for NYC as provided by NYCDOHMH.*

Click to see NYS excl. NYC age-adjusted rate

Race/Ethnicity	NYC	NYS Excl. NYC
Hispanic	34% (29% of population)	14% (12% of population)
Black	28% (22% of population)	18% (9% of population)
White	27% (32% of population)	60% (74% of population)
Asian	7% (14% of population)	4% (4% of population)
Other	4% (3% of population)	4% (1% of population)

https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Fatalities?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n#/views/NYS%2dCOVID19%2dTracker/NYSDOHCOVID%2d19Tracker%2dMap?%253Aembed=yes&%253Atoolbar=no

Chi -Square Test – New York State & New York City

```r
# Chi Square test for the probability of dying from COVID-19 in New York State based on Race
null.probs = c(12/100, 9/100, 74/100, 4/100, 1/100)
NYSdeaths = c(14, 18, 60, 4, 4)
chisq.test(NYSdeaths, p=null.probs)
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  NYSdeaths
X-squared = 20.982, df = 4, p-value = 0.0003193
```

```r
# Chi Square test for the probability of dying from COVID-19 in New York City based on Race
null.probs = c(29/100, 22/100, 32/100, 14/100, 3/100)
NYCdeaths = c(34, 28, 27, 7, 4)
chisq.test(NYCdeaths, p=null.probs)
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  NYCdeaths
X-squared = 7.113, df = 4, p-value = 0.13
```

The chi-square test with a resulting in a p-value ≈ 0 suggests there is **very strong evidence** that Race & Ethnicity have a statistically significant impact upon COVID-19 Death rates In New York State. However, the chi-square test resulting in a p-value of 0.13 suggests there is **no evidence** that Race & Ethnicity have a statistically significant impact upon COVID-19 Death rates in **New York City**.

The Odds Ratio – New York State & New York City

NYC	Odds Ratio	B	B'
		White	Black
A	Covid 19 death	3846.96	3989.44
A'	No Covid 19 death	2683752.40	1843735.12
		2687599.36	1847724.56

Population Ratio	1.45	1.45 times more white people than African Americans in NYC
Odds Ratio: Deaths	1.51	African Americans are more 1.5 times more likely to die of Covid-19 in NYC

NYC	Odds Ratio	B	B'
		White	Hispanic
A	Covid 19 death	3846.96	4844.32
A'	No Covid 19 death	2683752.40	2430792.60
		2687599.36	2435636.92

Population Ratio	1.10	1.1 times more white people than Latinx people in NYC
Odds Ratio: Deaths	1.39	Latinx people are 1.4 times more likely to die of Covid-19 in NYC than White people

NYS	Odds Ratio	B	B'
		White	Black
A	Covid 19 death	15963.00	4788.90
A'	No Covid 19 death	14379672.14	1746031.59
		14395635.14	1750820.49

Population Ratio	8.22	8 times more white people than African Americans in NYS
Odds Ratio: Deaths	2.47	African Americans are are 2.5 times more likely than white people to die of Covid-19 in NYS

NYS	Odds Ratio	B	B'
		White	Hispanic
A	Covid 19 death	15963.00	3724.70
A'	No Covid 19 death	14379672.14	2330702.62
		14395635.14	2334427.32

Population Ratio	6.17	6 times more white people than Latinx people in NYS
Odds Ratio: Deaths	1.44	Latinx people are 1.4 times more likely to die of Covid-19 than white people in NYS

The Odds Ratio calculations based off of the summary statistics provided by the New York State and New York City Departments of Health support the conclusion that there is strong evidence to suggest race and ethnicity are related to COVID-19 infection and death rates in New York State, however they also suggest that race and ethnicity are related to COVID-19 infection and death rates in New York City as well.

Maryland

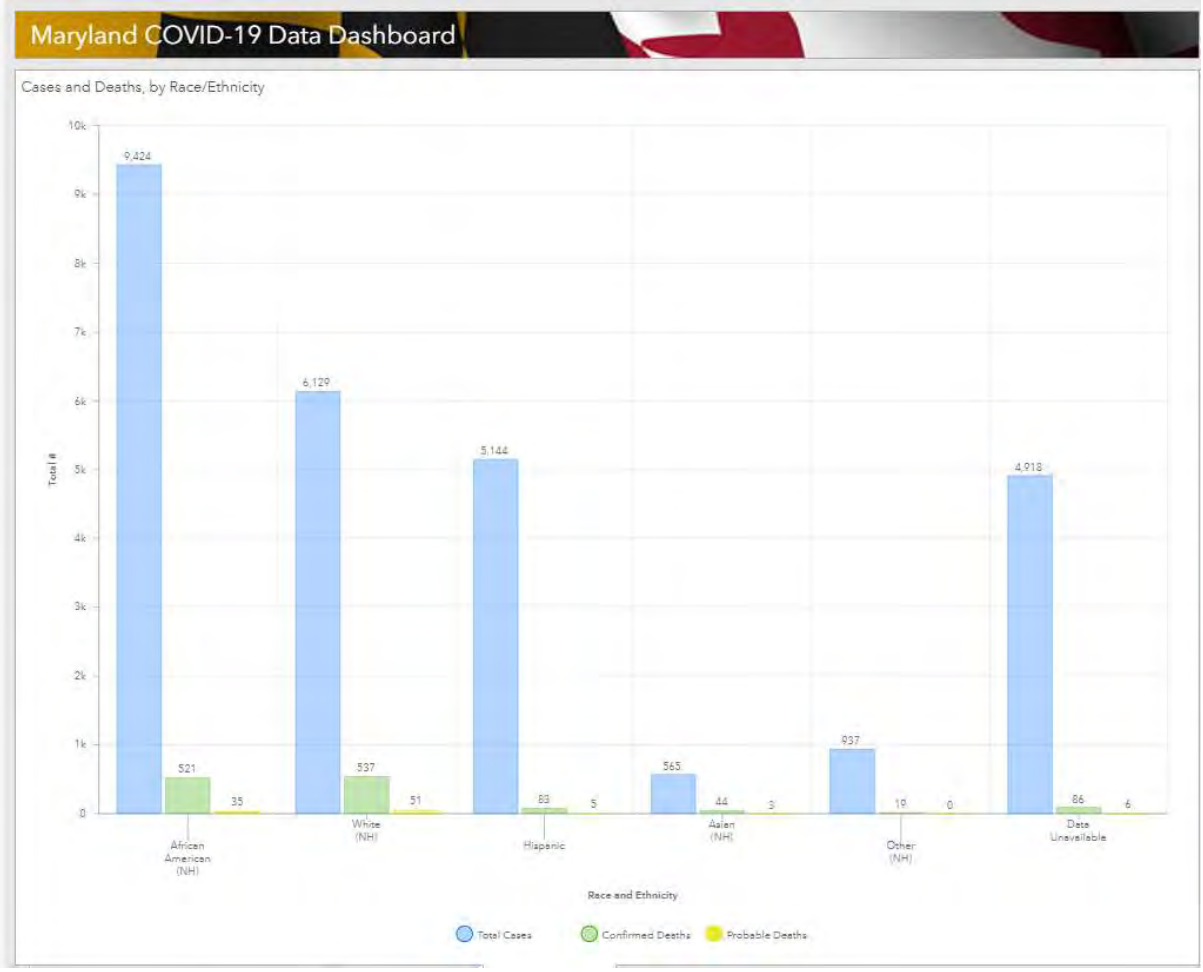
The following summary statistics are provided by the State of Maryland Department of Health.

Maryland COVID-19 Infection & Death Rates & Population Percentages Based on Race & Ethnicity					
Race/Ethnicity	Cases	% of Total Cases	Deaths	% of Total Deaths	Population %
African-American	9,424	34.75%	521	40.39%	30.90
Asian	565	2.08%	44	3.41%	6.70
White	6129	22.60%	537	41.63%	50.50
Hispanic	5144	18.97%	83	6.43%	10.40
Other	937	3.46%	19	1.47%	1.50
Data Not Available	4918	18.14%	86	6.67%	**
Total	27,117	100.00%	1290	100.00%	100.00

By Race and Ethnicity

Race/Ethnicity	Cases	Deaths	
African-American (NH)	9,424	(521)	35*
Asian (NH)	565	(44)	3*
White (NH)	6,129	(537)	51*
Hispanic	5,144	(83)	5*
Other (NH)	937	(19)	
Data not available	4,918	(86)	6*

https://coronavirus.maryland.gov/



MD Census information

Race and Hispanic Origin	
White alone, percent	58.8%
Black or African American alone, percent (a)	30.9%
American Indian and Alaska Native alone, percent (a)	0.6%
Asian alone, percent (a)	6.7%
Native Hawaiian and Other Pacific Islander alone, percent (a)	0.1%
Two or More Races, percent	2.9%
Hispanic or Latino, percent (b)	10.4%
White alone, not Hispanic or Latino, percent	50.5%

Chi-Square Test – Maryland

```
106 ```{r}
107 # Chi Square test for the probability of dying from COVID-19 in Maryland based on Race
108 null.probs = c(30.9/100, 6.7/100, 50.5/100, 10.4/100, 0/100, 1.5/100)
109 MDdeaths = c(40.39, 3.41, 41.63, 6.43, 1.47, 6.67)
110 chisq.test(MDdeaths, p=null.probs)
111 ```
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  MDdeaths
X-squared = Inf, df = 5, p-value < 2.2e-16
```

```
112
113 ```{r}
114 # Chi Square test for the probability of infection from COVID-19 in Maryland based on Race
115 null.probs = c(30.9/100, 6.7/100, 50.5/100, 10.4/100, 0/100, 1.5/100)
116 MDcases = c(34.75, 2.08, 22.60, 18.97, 3.46, 18.14)
117 chisq.test(MDcases, p=null.probs)
118 ```
```

```
Chi-squared approximation may be incorrect
        Chi-squared test for given probabilities

data:  MDcases
X-squared = Inf, df = 5, p-value < 2.2e-16
```

There is very strong evidence that Race & Ethnicity have a statistically significant impact upon COVID-19 Infection and Death rates In Maryland

The Odds Ratio – Maryland

MD	Odds Ratio	B	B'			
		White	Black			
A	Covid 19 cases	6129.00	9424.00	Population Ratio	1.63	1.63 times more white people than African Americans
A'	No Covid 19 cases	3065844.58	1870258.84	Odds Ratio: cases	2.52	African Americans are 2.5 times as likely to have Covid-19
		3071973.58	1879682.84			
MD	Odds Ratio	B	B'			
		White	Hispanic			
A	Covid 19 cases	6129.00	5144.00	Population Ratio	4.86	4.86 times as many white people as Latinx in MD
A'	No Covid 19 cases	3065844.58	627500.06	Odds Ratio: cases	4.10	Latinex are 4.1 times as likely as white people to have Covid in MD
		3071973.58	632644.06			
MD	Odds Ratio	B	B'			
		White	Black			
A	Covid 19 deaths	537.00	521.00	Population Ratio	1.63	1.63 times more white people than African Americans
A'	No Covid 19 deaths	3071436.58	1879161.84	Odds Ratio: deaths	1.59	African Americans are 1.6 times as likely to die of Covid-19
		3071973.58	1879682.84			
MD	Odds Ratio	B	B'			
		White	Hispanic			
A	Covid 19 deaths	537.00	83.00	Population Ratio	4.86	4.86 times as many white people as Latinx in MD
A'	No Covid 19 deaths	3071436.58	632561.06	Odds Ratio: deaths	0.75	Latinex are 0.75 times as likely as white people to die of Covid in MD

The Odds Ratio calculations based off of the summary statistics provided by the Maryland Department of Health support the conclusion that there is strong evidence to suggest race and ethnicity are related to COVID-19 infection and death rates in MD.

2. Gender

The CDC is not collected COVID-19 data based on gender at this time. The same is true for many individual state departments of health. Policies and public health efforts are not focusing upon the gendered impacts of the pandemic. Internationally governments and global health institutions, such as the WHO, are not providing analysis of the outbreak with a gender variable considered. Yet, understanding that diseases impact men and women differently is necessary to understand the effects of the health emergency, to craft tailored biological and medical solution as well as creating effective and equitable policies and intervention on a community level.[39]

Emerging data suggests that although the infection rate appears fairly equally distributed, more men are being hospitalized and dying due to COVID-19. This could potentially be due to "sex-based immunological or gendered differences, such as patterns and prevalence of smoking."[40] There are also many gendered social implications for how the disease impacts the day-to-day lives of men and women differently.

This section reviews the impact of gender upon COVID-19 infection, hospitalization and death rates, based on the current available data. Summary statistics provided by New York City, New York State and Maryland are provided below. There is an analysis of an open dataset from the State of Connecticut department of Health regarding gender's relationship to COVID-19 cases and fatalities.

a. Infection - New York City

Note: Due to the small number of cases among transgender and gender-nonconforming people, data on those cases are not included in this table at this time.

https://www1.nyc.gov/site/doh/covid/covid-19-data.page

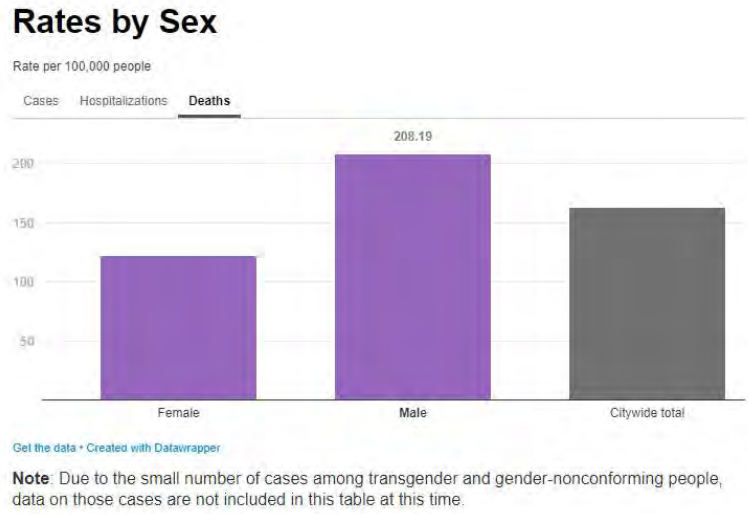
[39] Clare Wenham, Julia Smith, Rosemary Morgan, *COVID-19: the gendered impacts of the outbreak*, THE LANCET, Vol. 395, Issue 10227, P.846-848, March 14, 2020, *available at* https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30526-2/fulltext.

[40] *Id.*

b. Hospitalization - New York City

https://www1.nyc.gov/site/doh/covid/covid-19-data.page

c. Death - New York City



https://www1.nyc.gov/site/doh/covid/covid-19-data.page

New York State

Fatalities by Sex	
Grand Total	19,645 (100.0%)
Female	7,940 (40.4%)
Male	11,696 (59.6%)
Unknown	9 (0.0%)

https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Fatalities?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n#/views/NYS%2dCOVID19%2dTracker/NYSDOHCOVID%2d19Tracker%2dMap?%253Aembed=yes&%253Atoolbar=no

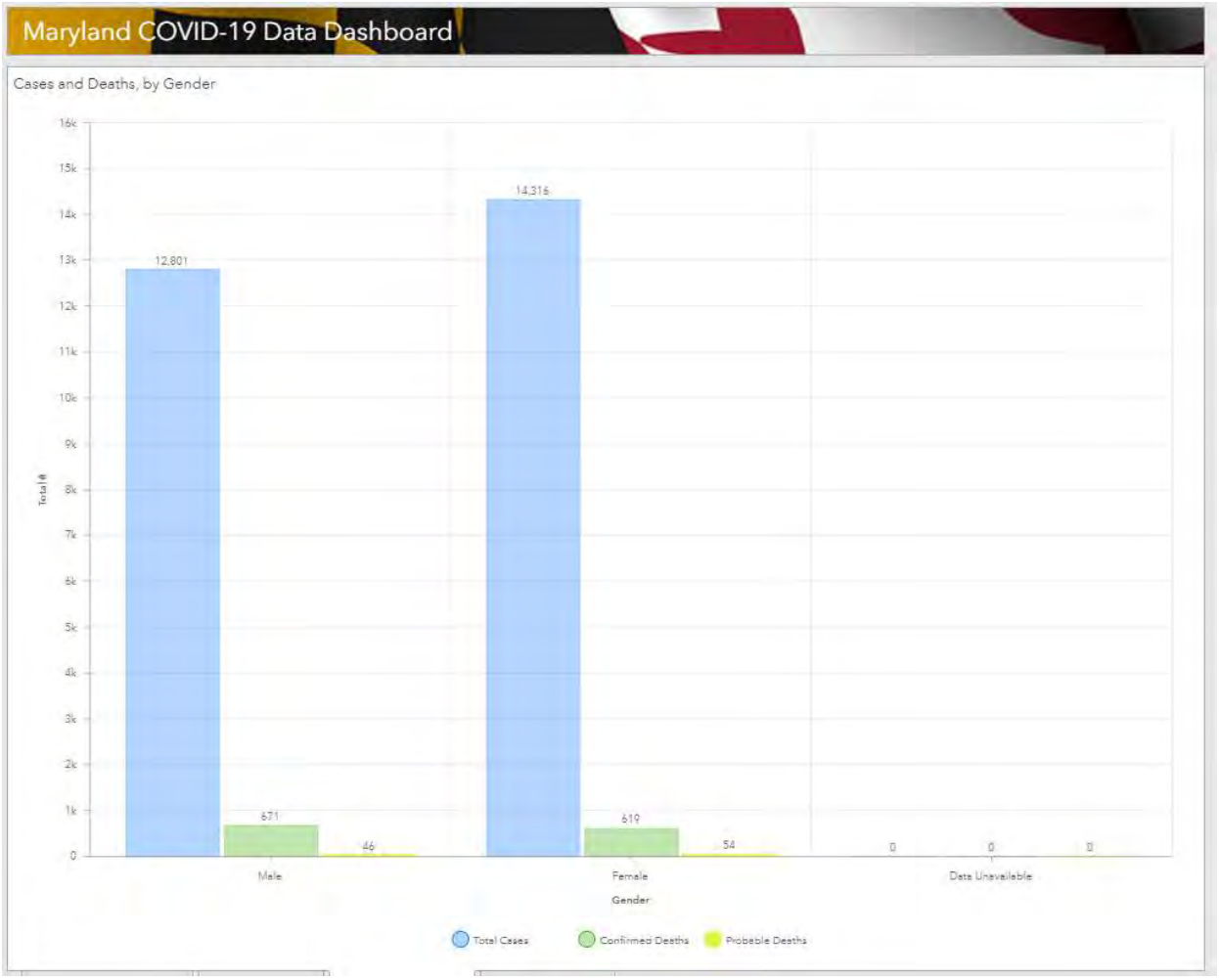
Maryland

By Age Range and Gender		
Age/Gender	Cases	Deaths
Female	14,316	(619)
Male	12,801	(671)

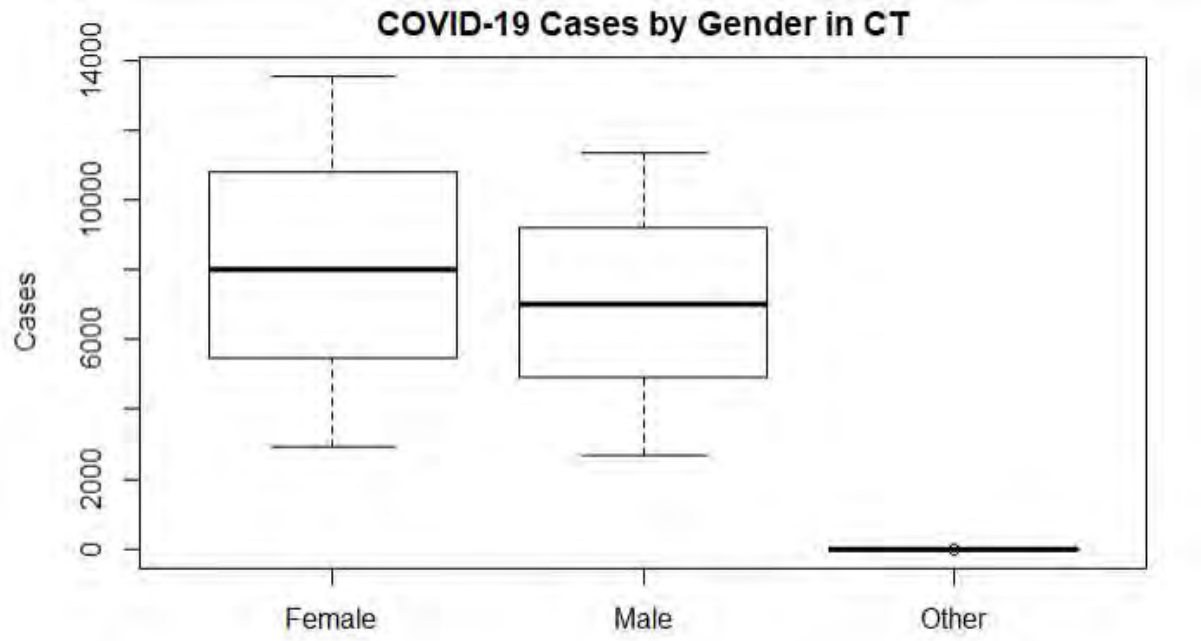
https://coronavirus.maryland.gov/

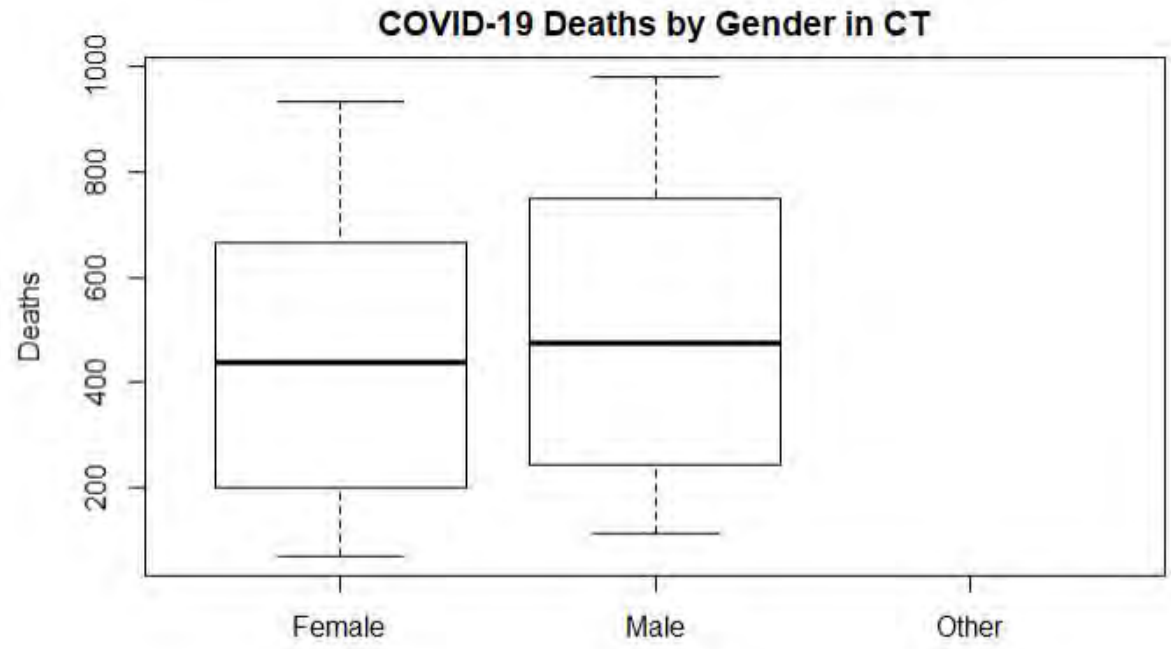
Age and Sex	
Persons under 5 years, percent	6.0%
Persons under 18 years, percent	22.2%
Persons 65 years and over, percent	15.4%
Female persons, percent	51.5%

census.gov/quickfacts/MD



Connecticut Dataset





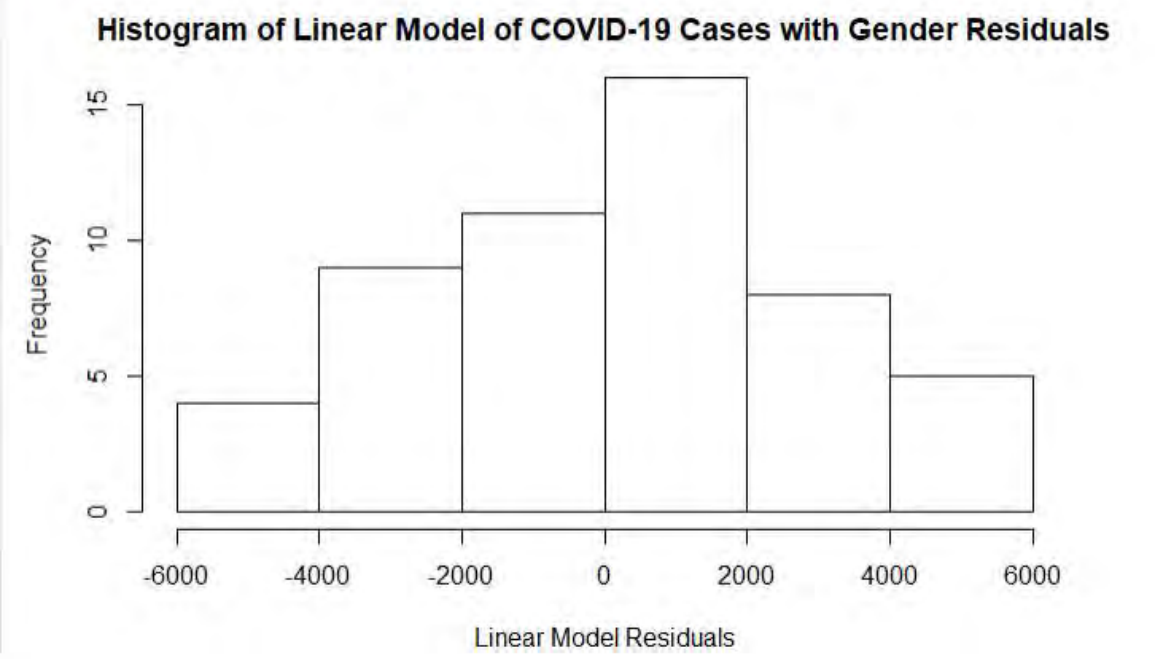
The boxplots above indicate that gender does have an impact upon COVID-19 cases in CT, yet may have a lesser impact upon COVID-19 deaths in CT. A Linear Model analysis was conducted. The results are below.

```r
#Linear model of cases ~ gender
lm_casesg <- lm(cases ~ gender, data = gender2)
sum_lm_casesg <- summary(lm_casesg)
sum_lm_casesg$coefficients
```

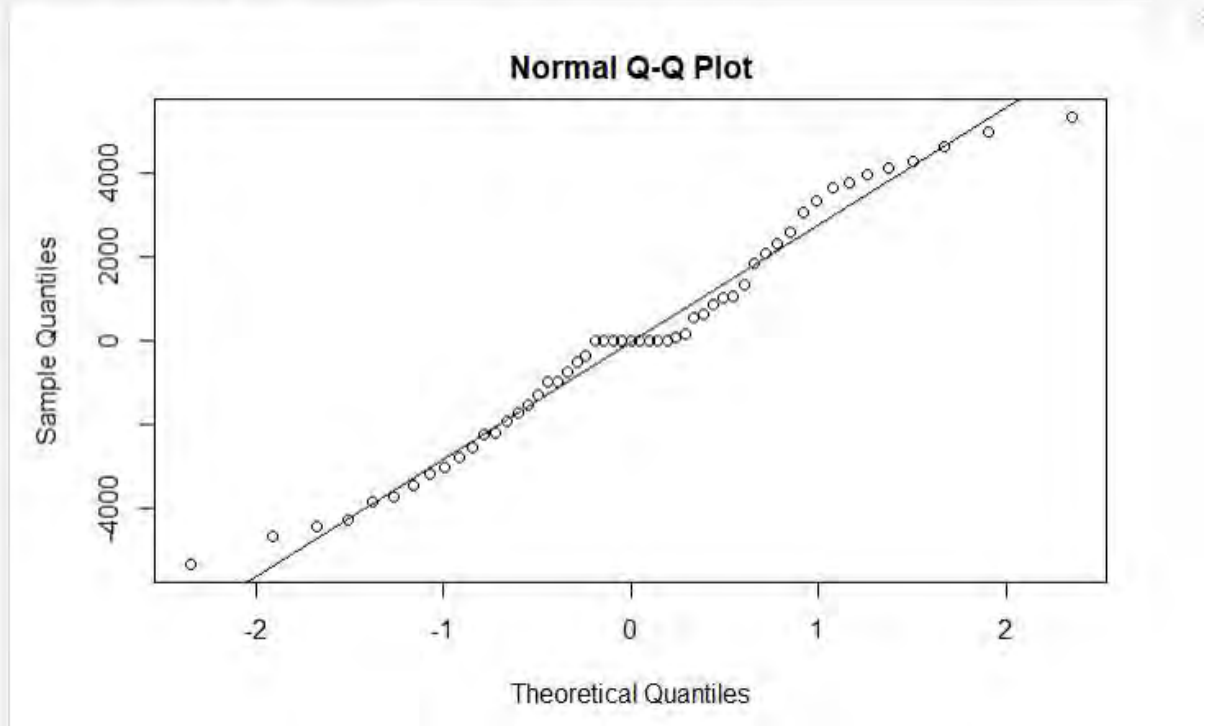
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8220.000	589.0691	13.954221	7.166287e-19
genderMale	-1123.091	833.0695	-1.348136	1.836907e-01
genderOther	-8218.222	1093.2659	-7.517130	9.345055e-10

The resulting p-values of Intercept (Gender Female) p-value: ≈ 0 indicates that gender could have an impact upon the number of cases, however the Gender Male p-value was 0.18, which suggests the contrary. The Gender Other p-value: ≈ 0 will not be considered since n=2 and was not included in the early or later part of the dataset. The residuals for this linear model appear normal though, suggesting a good fit for the data.

```r
hist(lm_cases$residuals)
hist(lm_casesg$residuals, main = "Histogram of Linear Model of COVID-19 Cases with Gender Residuals", xlab = "Linear Model Residuals")
```



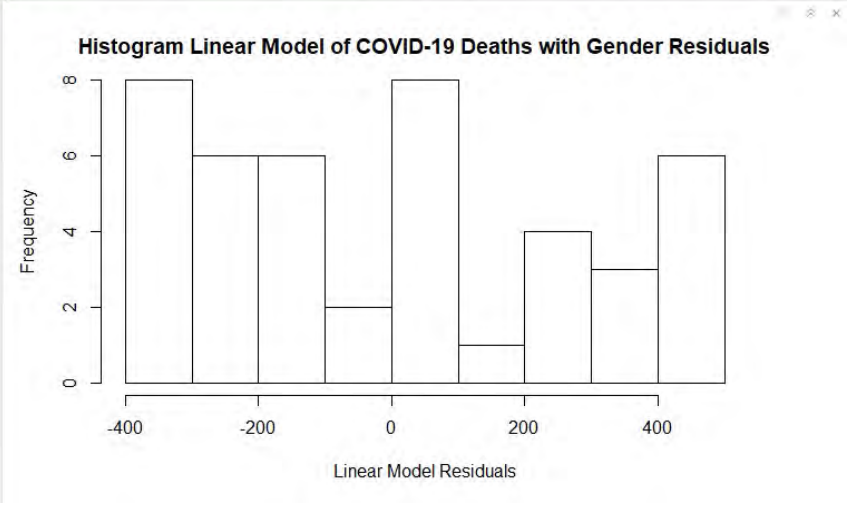
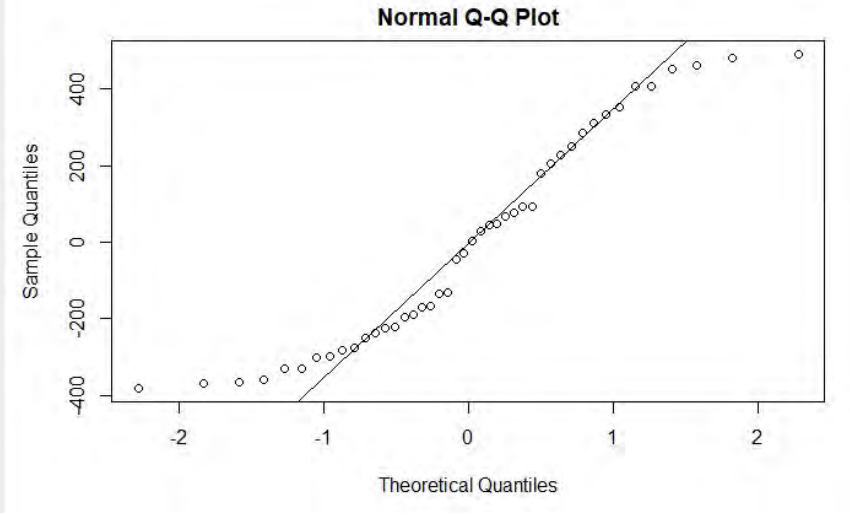
```
qqnorm(lm_casesg$residuals)
qqline(lm_casesg$residuals)
```

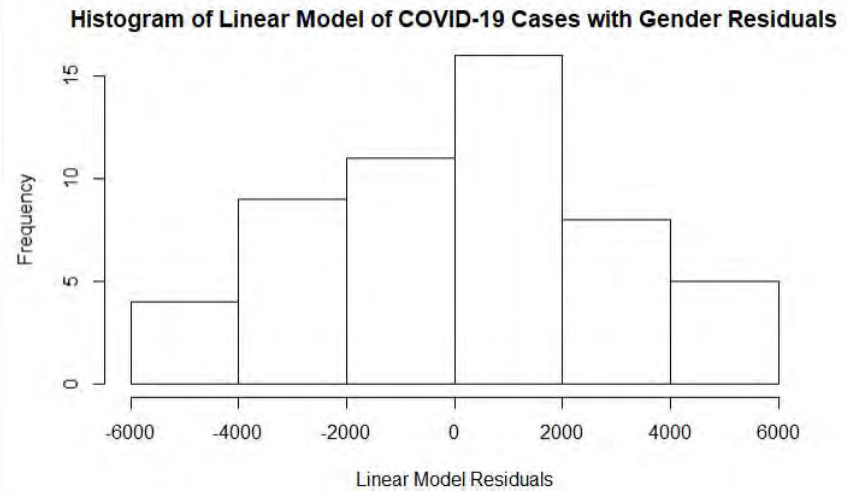


The linear model regarding the impact of gender upon COVID-19 deaths provided similar results. The Intercept (Gender Female) p-value was ≈ 0, suggesting a relationship, however the Gender Male p-value was 0.5 suggesting there is no relationship.

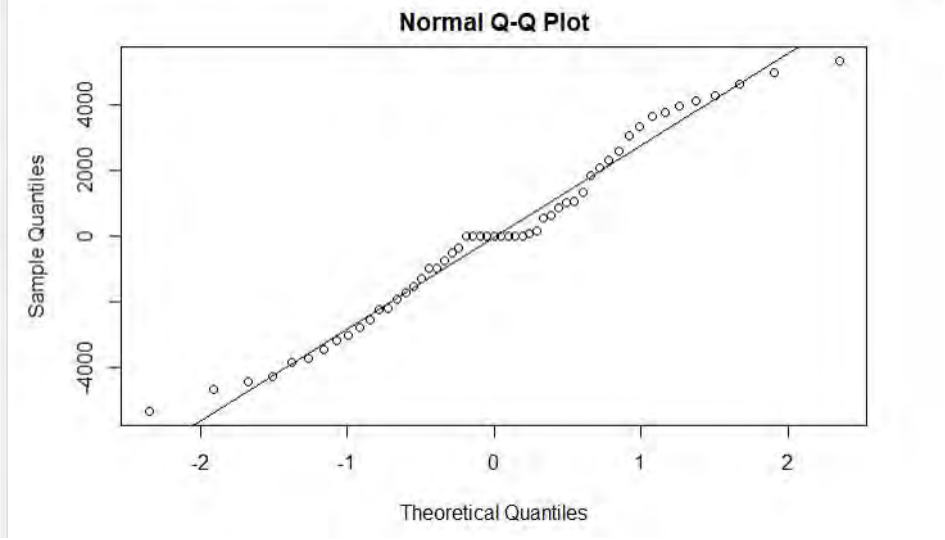
```
```{r}
#Linear model of deaths ~ gender
lm_deathsg <- lm(deaths ~ gender, data = gender2)
sum_lm_deathsg <- summary(lm_deathsg)
sum_lm_deathsg$coefficients
```

```
             Estimate Std. Error   t value    Pr(>|t|)
(Intercept) 439.09091   60.25356 7.2873858 5.65763e-09
genderMale   58.13636   85.21140 0.6822604 4.98820e-01
```

However, given the non-normal residuals this model may not be the best fit for this data and further analysis is required.
```

```
hist(lm_deathsg$residuals, main = "Histogram Linear Model of COVID-19 Deaths with Gender
Residuals", xlab = "Linear Model Residuals")
```




```
```{r}
qqnorm(lm_deathsg$residuals)
qqline(lm_deathsg$residuals)
```
```




```
```{r}
hist(lm_cases$residuals)
hist(lm_casesg$residuals, main = "Histogram of Linear Model of COVID-19 Cases with Gender
Residuals", xlab = "Linear Model Residuals")
```
```



```
qqnorm(lm_casesg$residuals)
qqline(lm_casesg$residuals)
```



Overall, the data does suggest that gender has an impact upon COVID-19 infection and mortality rates, but further analysis is required to determine the extent, particularly given the seemingly contradictory p-value results for Males and Females in the linear model results. Much further testing is required to determine possible explanatory variables.

Some possible explanations for the increased death rate in males, and some recorded higher infection rates in females, that require verification through further research include:

1. Males engage in riskier behavior[41] which lead to health complications such as a history with smoking;[42]
2. Biological differences between men and women create a stronger immune system in women, with the additional X chromosome and less testosterone (testosterone can weak the immune system);[43]
3. Males may be waiting longer before receiving treatment (hence have a higher mortality rate);[44]
4. Females are more likely to be working in roles currently considered "essential" such as in the health care industry, in grocery stores, and other service roles (hence the higher infection rate).[45]

---

[41] Nathaniel Scharping, *The Reason More Men Die from COVID-19 May Lie in their Chromosomes*, DISCOVER MAGAZINE, April 13, 2020, *available at* https://www.discovermagazine.com/health/the-reason-more-men-die-from-covid-19-may-lie-in-their-chromosomes

[42] Hannah Devlin, *Men are much more likely to die from coronavirus – but why?*, THE GUARDIAN, April 16, 2020, *available at* https://www.theguardian.com/world/2020/mar/26/men-are-much-more-likely-to-die-from-coronavirus-but-why

[43] *Refer Note 41.*

[44] *Refer Note 42.*

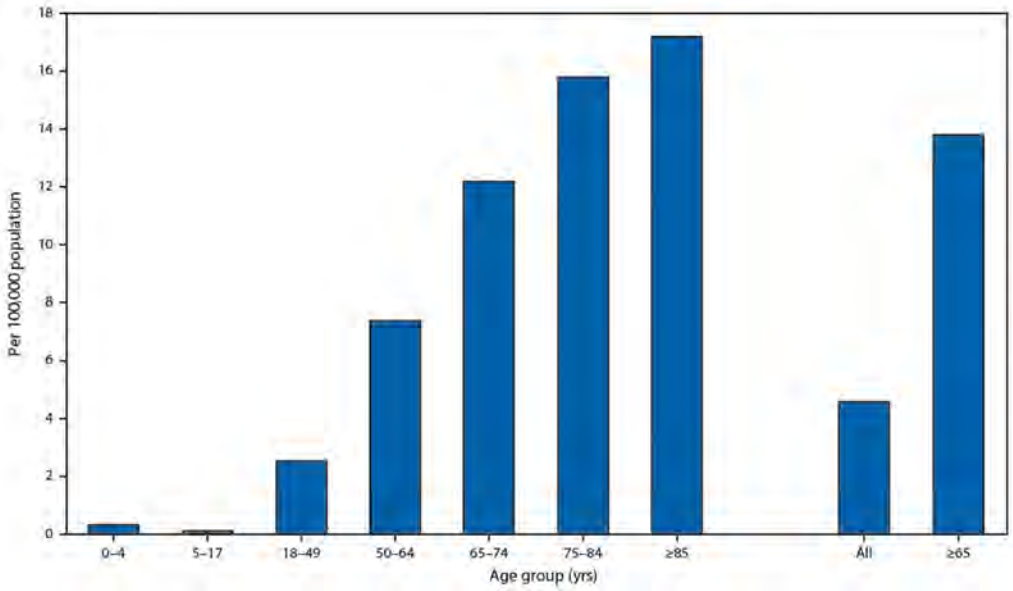
[45] Campbell Robertson & Robert Gebeloff, *How Millions of Women became the Most Essential Workers in America*, THE NEW YORK TIMES, April 18, 2020, *available at* https://www.nytimes.com/2020/04/18/us/coronavirus-women-essential-workers.html

3. Age

There was not a sufficient dataset available for independent analysis of the data on age as it is related to COVID-19 deaths and infections. However, the summary statistics provided by the federal government via the CDC and the individual state governments, including Maryland, indicate a strong relationship, particularly in regard to the number of deaths from COVID-19.

a. Infection – United States CDC

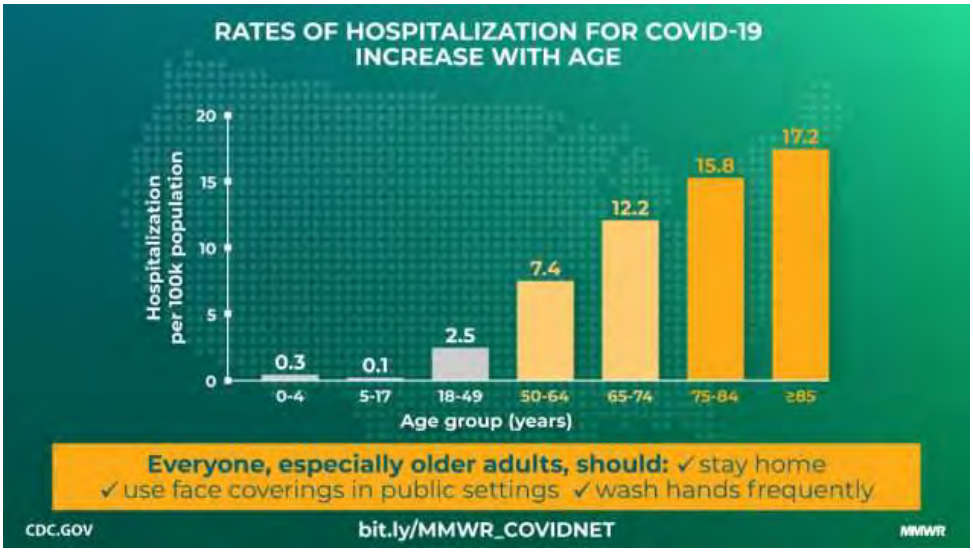
The below chart from the CDC shows the number of COVID-19 infection per 100,000 people categorized by age group below. There appears to be a positive relationship between number of infections and an increase in age.



https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm?s_cid=mm6915e3_e

b. Hospitalization - United States CDC

Similarly, the below chart from the CDC shows the number of COVID-19 hospitalization per 100,000 people categorized by age group below. There appears to be a positive relationship between number of hospitalizations and an increase in age.



c. Death

Individual states provide summary statistics on the COVID-19 death rate as categorized by age. Data from New York State and Maryland is provided below as case studies.

New York State

Fatalities by Age Group

Age Group	Fatality Count	%
Grand Total	19,645	100.0%
0 to 9	3	0.0%
10 to 19	8	0.0%
20 to 29	68	0.4%
30 to 39	281	1.4%
40 to 49	715	3.6%
50 to 59	1,960	10.0%
60 to 69	3,956	20.1%
70 to 79	5,244	26.7%
80 to 89	4,986	25.4%
90 & Over	2,390	12.2%
Unknown	34	0.2%

https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Fatalities?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n#/views/NYS%2dCOVID19%2dTracker/NYSDOHCOVID%2d19Tracker%2dMap?%253Aembed=yes&%253Atoolbar=no

Maryland

The figures provided by the State of Maryland Department of Health below show the dramatic impact of age upon COVID-19 infection and death rates in Maryland. While persons over 65 account for only 15.4 percent of the population persons over 60 account for 30.58 percent of infection and an incredible 88.07 percent of COVID-19 deaths.

Maryland COVID-19 Infection & Death Rates & Population Percentages Based on Age

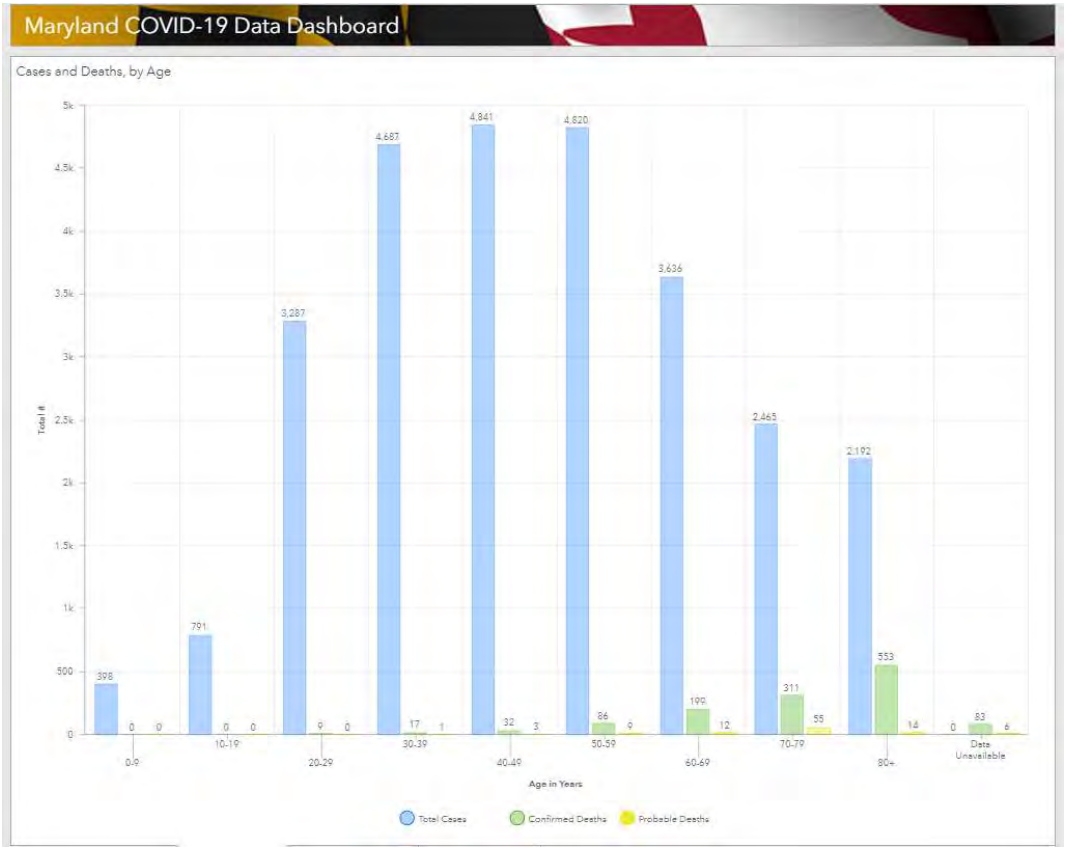
Age	Cases	% of Total Cases	Total % ≥ 60	Deaths	% of Total Deaths	Total % ≥ 60	Population %
A 0-9	398	1.47%		0	0.00%		**
A 10-19	791	2.92%		0	0.00%		**
A20-29	3287	12.12%		9	0.75%		**
A 30-39	4687	17.28%		17	1.41%		**
A 40-49	4841	17.85%		32	2.65%		**
A 50-59	4820	17.77%		86	7.13%		**
A 60-69	3636	13.41%		199	16.49%		Persons Over 65
A 70-79	2465	9.09%	30.58%	311	25.77%	88.07%	15.40%
A 80+	2192	8.08%		553	45.82%		
Total	27,117	100.00%		1207	100.00%		

By Age Range and Gender			
Age/Gender	Cases	Deaths	
0-9	398		
10-19	791		
20-29	3,287	(9)	
30-39	4,687	(17)	1*
40-49	4,841	(32)	3*
50-59	4,820	(86)	9*
60-69	3,636	(199)	12*
70-79	2,465	(311)	14*
80+	2,192	(553)	55*
Data not available		(83)	6*

https://coronavirus.maryland.gov/

Age and Sex	
Persons under 5 years, percent	⚠ 6.0%
Persons under 18 years, percent	⚠ 22.2%
Persons 65 years and over, percent	⚠ 15.4%
Female persons, percent	⚠ 51.5%

census.gov/quickfacts/MD



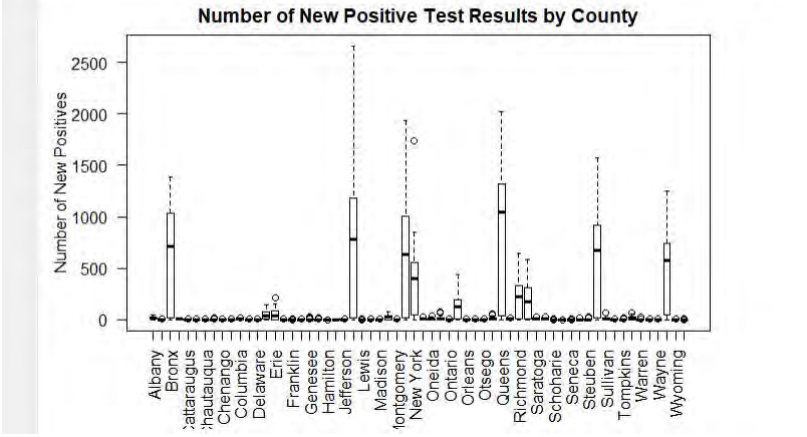
https://coronavirus.maryland.gov/

E. Limitations of the Data

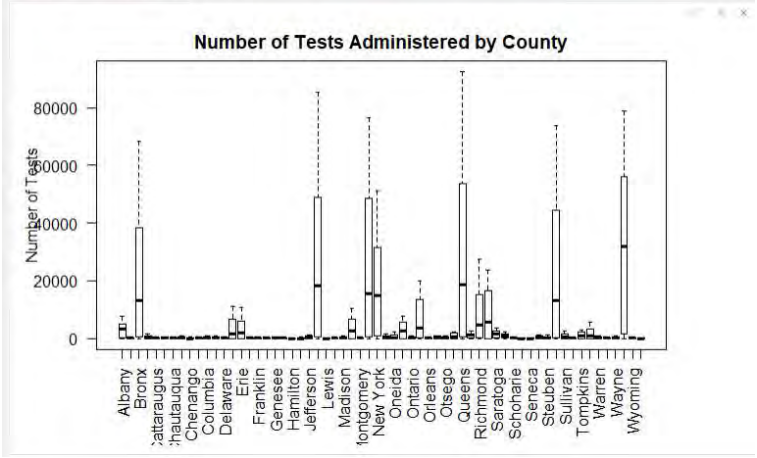
1. Testing

New York State has made a dataset open and available to the public regarding testing rates and results throughout New York States categorized by county and date. The following analysis is based off of that dataset.

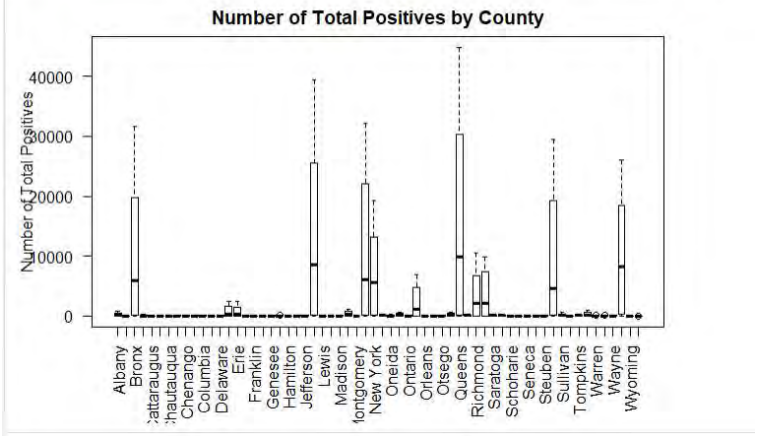
```{r}
plot(new.pos ~ county, las = 2, main = "Number of New Positive Test Results by County", ylab =
"Number of New Positives", xlab = "")
```




```{r}
plot(num.tests ~ county, las = 2, main = "Number of Tests Administered by County", ylab =
"Number of Tests", xlab = "")
```




```{r}
plot(num.pos ~ county, las = 2, main = "Number of Total Positives by County", ylab = "Number of
Total Positives", xlab = "")
```

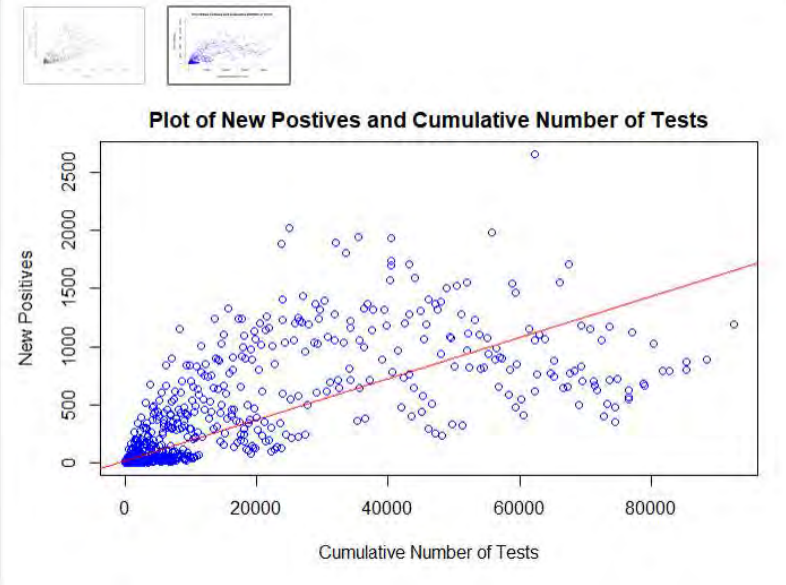


The plots above make clear that geography impacts the number of tests available, which also impacts the rate data for infection and deaths. A linear model was produced to further explore these relationships. The results are below.

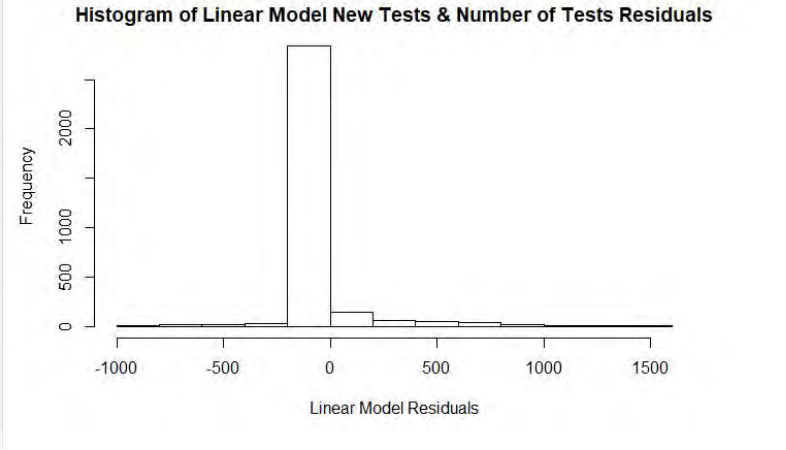
```r
# Linear model of new positives ~ county + totalnumberoftestsperformed + date
new.pos.lm <- lm(new.pos ~  county + num.tests + date, data = covidNY2)
sum_new.pos.lm <- summary(new.pos.lm)
sum_new.pos.lm$coefficients
summary(new.pos.lm)
```

```
Call:
lm(formula = new.pos ~ county + num.tests + date, data = covidNY2)

Residuals:
    Min      1Q  Median      3Q     Max
-798.76   -9.88    1.48    8.95 1486.13

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -7.884e+03  3.261e+03  -2.418  0.01568 *
countyAllegany   2.129e+01  2.711e+01   0.785  0.43241
countyBronx      3.887e+02  2.763e+01  14.068  < 2e-16 ***
countyBroome     2.126e+01  2.711e+01   0.784  0.43293
countyCattaraugus 2.098e+01  2.711e+01   0.774  0.43911
countyCayuga     2.019e+01  2.711e+01   0.745  0.45643
countyChautauqua 2.053e+01  2.711e+01   0.757  0.44903
countyChemung    1.978e+01  2.711e+01   0.730  0.46570
countyChenango   2.209e+01  2.711e+01   0.815  0.41537
countyClinton    2.098e+01  2.711e+01   0.774  0.43913
countyColumbia   2.058e+01  2.711e+01   0.759  0.44794
countyCortland   1.886e+01  2.711e+01   0.696  0.48672
countyDelaware   2.151e+01  2.711e+01   0.793  0.42764
countyDutchess   2.925e+01  2.710e+01   1.079  0.28051
countyErie       3.339e+01  2.710e+01   1.232  0.21794
countyEssex      2.182e+01  2.711e+01   0.805  0.42107
countyFranklin   2.141e+01  2.711e+01   0.790  0.42972
countyFulton     2.102e+01  2.711e+01   0.775  0.43817
countyGenesee    2.218e+01  2.711e+01   0.818  0.41349
countyGreene     2.123e+01  2.711e+01   0.783  0.43373
countyHamilton   2.225e+01  2.712e+01   0.821  0.41198
countyHerkimer   2.153e+01  2.711e+01   0.794  0.42728
countyJefferson  1.832e+01  2.711e+01   0.676  0.49936
countyKings      4.609e+02  2.808e+01  16.412  < 2e-16 ***
countyLewis      2.163e+01  2.711e+01   0.798  0.42501
countyLivingston 2.125e+01  2.711e+01   0.784  0.43334
countyMadison    2.072e+01  2.711e+01   0.764  0.44478
countyMonroe     3.118e+00  2.710e+01   0.115  0.90840
countyMontgomery 2.136e+01  2.711e+01   0.788  0.43093
countyNassau     3.463e+02  2.792e+01  12.403  < 2e-16 ***
countyNew York   1.819e+02  2.749e+01   6.619 4.23e-11 ***
countyNiagara    2.298e+01  2.711e+01   0.847  0.39679
countyOneida     2.084e+01  2.711e+01   0.769  0.44203
countyOnondaga   7.034e-02  2.710e+01   0.003  0.99793
countyOntario    1.974e+01  2.711e+01   0.728  0.46658
countyOrange     7.521e+01  2.712e+01   2.773  0.00558 **
countyOrleans    2.197e+01  2.711e+01   0.810  0.41775
countyOswego     1.878e+01  2.711e+01   0.693  0.48855
countyOtsego     2.036e+01  2.711e+01   0.751  0.45264
countyPutnam     2.376e+01  2.711e+01   0.877  0.38074
countyQueens     5.449e+02  2.825e+01  19.292  < 2e-16 ***
countyRensselaer 1.414e+01  2.711e+01   0.522  0.60189
countyRichmond   1.316e+02  2.714e+01   4.849 1.30e-06 ***
countyRockland   1.137e+02  2.714e+01   4.190 2.87e-05 ***
countySaratoga   1.001e+01  2.710e+01   0.370  0.71177
countySchenectady 1.583e+01  2.711e+01   0.584  0.55914
countySchoharie  2.138e+01  2.711e+01   0.789  0.43044
countySchuyler   2.160e+01  2.711e+01   0.797  0.42564
countySeneca     2.158e+01  2.711e+01   0.796  0.42615
countySt. Lawrence 1.993e+01  2.711e+01   0.735  0.46239
countySteuben    2.159e+01  2.711e+01   0.796  0.42582
countySuffolk    3.217e+02  2.777e+01  11.584  < 2e-16 ***
countySullivan   2.292e+01  2.711e+01   0.846  0.39787
countyTioga      2.136e+01  2.711e+01   0.788  0.43083
countyTompkins   1.174e+01  2.711e+01   0.433  0.66502
countyUlster     1.975e+01  2.710e+01   0.729  0.46611
countyWarren     2.026e+01  2.711e+01   0.747  0.45505
countyWashington 2.111e+01  2.711e+01   0.779  0.43631
countyWayne      1.989e+01  2.711e+01   0.734  0.46315
countyWestchester 1.570e+02  2.843e+01   5.523 3.59e-08 ***
countyWyoming    2.172e+01  2.711e+01   0.801  0.42318
countyYates      2.188e+01  2.711e+01   0.807  0.41985
num.tests        1.172e-02  3.073e-04  38.148  < 2e-16 ***
date             4.285e-01  1.778e-01   2.411  0.01599 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 138.2 on 3160 degrees of freedom
Multiple R-squared:  0.7306,    Adjusted R-squared:  0.7252
F-statistic:   136 on 63 and 3160 DF,  p-value: < 2.2e-16
```

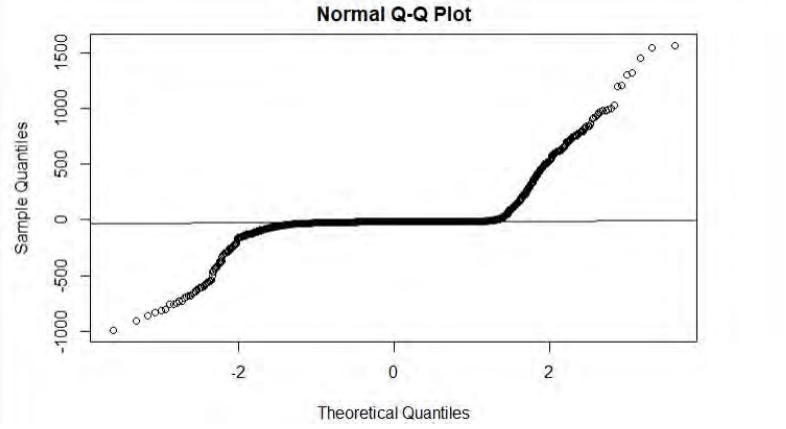
```{r}
plot(new.pos, num.tests)
plotlmnew <- lm(new.pos ~ num.tests, data = covidNY2)
plot(jitter(new.pos) ~ num.tests, main = "Plot of New Postives and Cumulative Number of Tests"
ylab = "New Positives", xlab = "Cumulative Number of Tests", col = "blue")
abline(plotlmnew, col = "red")
```




```{r}
hist(plotlmnew$residuals, main = "Histogram of Linear Model New Tests & Number of Tests
Residuals", xlab = "Linear Model Residuals")
```




```{r}
qqnorm(plotlmnew$residuals)
qqline(plotlmnew$residuals)
```



The linear model p-value of ≈ 0 suggests there is very strong evidence that County within New York is statistically significant for new positive tests for COVID-19. What is not clear is whether proportionally there are more infections in these areas, or whether there are simply more tests available. Further the residuals follow a non-normal pattern suggesting the model may be unfit for the data.

2. ***Bias***

Governmental departments are providing the data and particularly in the US the COVID-19 data has been heavily politicized.

3. ***Definition & Demographic Assignment in the data***

particularly in regard to racial classifications.

4. ***Overwhelmed Systems***

The data is collected in an overwhelmed and hectic environment within systematic frameworks at a breaking point.

5. ***Novel - This disease is very recently evolved and has ballooned in a very short period of time. There is much we do not yet know and understand about the disease, particularly in regard to timeline of the disease which could impact the assessed numbers of cases and deaths due to the disease.***

F. <u>Lessons Learned</u>

1. ***Investigate the details of the dataset prior to basing a study around it.***

It may seem self-explanatory, but it would have been extremely helpful if I had explored the data in much more detail prior to selecting the particular datasets I chose.

2. ***The government, public health organizations and policymakers often use simplified statistics out of context as a basis of decision making – particularly during a crisis.***

On the morning of May 6, 2020, the Washington D.C. NPR station, WAMU had an interview with an academic who was describing the background for the statistical models being used by the federal government to discuss COVID-19, make predictions and formulate policies. He stated that we should be very wary to explore these models without an understanding of the context and nuances that must be utilized to interpret the findings. If we do not explore the details of the model, there is great room for errors in interpretation, on top of the errors and noise within the models already!

3. ***There is a great need for greater education in the general population around data and statistics given their prevalent use in today's world and decision-making.***

There are society-changing decisions that are being made currently based off of models many decision-makers do not understand. It seems that a background is statistics should be a higher priority within our education system, given how influential they are upon our lives.

I recognize that the summary statistics I have used in this paper from various government sources fall prey to the same lack of ability to analyze the simplified numbers and graphics without a proper exploration of the data that was used to produce them.

4. ***Always test the effect of individual variables in a multi-variable data set.***

In particular, when plotting residuals from linear models with multiple variables, it is important to test the impact of each variable upon the results, particularly the residuals. One or two strongly correlated variables could cause a residual plot to appear normal for a different variable. This specifically refers to the Gender analysis in R where the residuals appeared normal until two variables were removed, which left the final variable, gender, to have abnormal residuals plots.

5. ***The Chi-square test null.probs must sum to 1 and there must be 2 or more variables***

Part IV. Conclusions

Systems are overwhelmed and disoriented, on a socio-structural level as well as an individual cognitive level, as the global community attempts to make sense of this new world, and craft intelligent anticipatory responses with educated foresight and minimal damage. There are unfortunately many lessons learned from previous outbreaks that have not been acted upon to reduce the socio-economic impact of COVID-19. This unique moment in time, however, does provide us with the opportunities to actual begin to make some changes to reduce the disparate impact of crisis events and global pandemics upon particularly groups based on their socio-economic status. Let us hope that these statistics will be considered, taken seriously, and acted upon so that this fatally unequal history is not repeated again.

More data is required to be collected at this time. Better quality data is required with more nuanced variables being considered. We need broad thinkers who can visualize the impact of this disease upon various segments of our society in order to craft appropriate solutions.

Race & Ethnicity

The data available and my analyses have shown very strong evidence that race and ethnicity impact both the infection and mortality rates due to COVID-19. The extent of this relationship is highly correlated to geographic area and local demographics, which suggests that different solutions are required in different areas. Interestingly, the Latinx population has incredibly high infection rates, with, proportionately speaking, incredibly low death rates. This insight could contain valuable information about cures and treatments that researchers should explore. Overall, we as a society need to do a better job of taking care of people of colour in our communities and begin to implement changes that reduce the disproportionate risk that people of colour face in this country. Data science can help, but we need more data and better data. We also need more people of color in data science to provide insights and focus as to possible explanatory variables of the disparate impact and help craft solutions based off of skilled and insightful interpretation of the data.

Gender

Historically, gender is an understudied variable in medical sciences. Yet, at this point in time we know that the unique differences in male, female, and intersexed bodies have drastic impacts upon susceptibility to certain diseases and effectiveness of vaccines and treatments. The data available and my analysis have found that COVID-19 is no different. We must understand the gendered

dynamics of this disease in order develop appropriate treatments and ultimately, to save lives. Any medical solution that is not taking gender into account given the current infection and death rates is arguably negligent.

Age

There is no available open dataset of age-specific information, not based on an age range, that was found. The nuances of the impact of age and this disease must be explored further. From all summary statistic data available, age appears highly correlated to COVID-19 deaths. We must collect further nuanced data to investigate this information and not just be satisfied with the easy conclusion that older people will die more from the disease. What measures can we take to prevent that? Might solutions include interdisciplinary discussion with urban planners and health care workers regarding how to desegregated the aged population from the rest of society? This presents an incredible opportunity for our society to initiate a renewed sense of valuation for the elderly in our society. Instead of viewing the only option as a dichotomy between sacrificing their lives or having a productive economy, perhaps we could find a solution that would better *protect* their lives which might help contribute to the economy. Times of crises create windows of opportunities to think differently, and more intelligently. Let us use this moment in time to honor our elderly and find better solution so they can continue to contribute their rich perspectives to our world. Data can help us do that.

Final Conclusion

In sum, COVID-19 will forever leave its mark on the global memory and structure of our world. We as a community are able to create the narrative that COVID-19 is remembered by. Are we going to re-enter the world in the same general condition as we were existing in two months ago? With the same inequalities and the same risks? Or are we going to take advantage of this idiosyncratic time to reflect upon how we can do it differently? How we can reorganize our societies to be better for all? Data can be the greatest asset we have in sculpting our future and redefining our values. Now the choice is in our hands how to wield the force that data becomes in these extraordinary and unprecedented times. Let us use that power to eliminate the impact of socio-economic status upon infection, life and death that we have seen with COVID-19, before the next pandemic strikes.