

An Analysis of the Differential Impacts of COVID-19 Mortality in the United States

Roilan Iglesias and Katie Pheysey

Mentored by
Dr. Joseph Cavanaugh
and Jacob Seedorff

ISIB Program sponsored by the National Heart Lung and Blood
Institute (NHLBI), grant # HL-147231



Background - COVID-19

- The SARS-CoV-2 virus emerged in late 2019 in Wuhan, China, causing a global pandemic¹
- The virus can spread through airborne water droplets, as well as contact with contaminated surfaces²
- Since January 2020, there have been over 34 million cases and around 613,000 deaths reported in the United States³
- Previous analyses have shown differences in mortality rates between races and jurisdictions⁴

1 https://en.wikipedia.org/wiki/COVID-19_pandemic#United_States

2 <https://psychcentral.com/coronavirus/background-history-of-the-coronavirus-covid-19#1>

3 https://en.wikipedia.org/wiki/COVID-19_pandemic#United_States

4 <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-deaths.html>



Background - CDC Database

- Publicly available database, continuously updated
- De-identified data
- 27.1 million cases as of 7/14/2021
- 19 variables; including demographic and geographic variables, as well as disease variables such as presence of underlying conditions and hospitalization status
- Link:
<https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4>



Analytic Goals

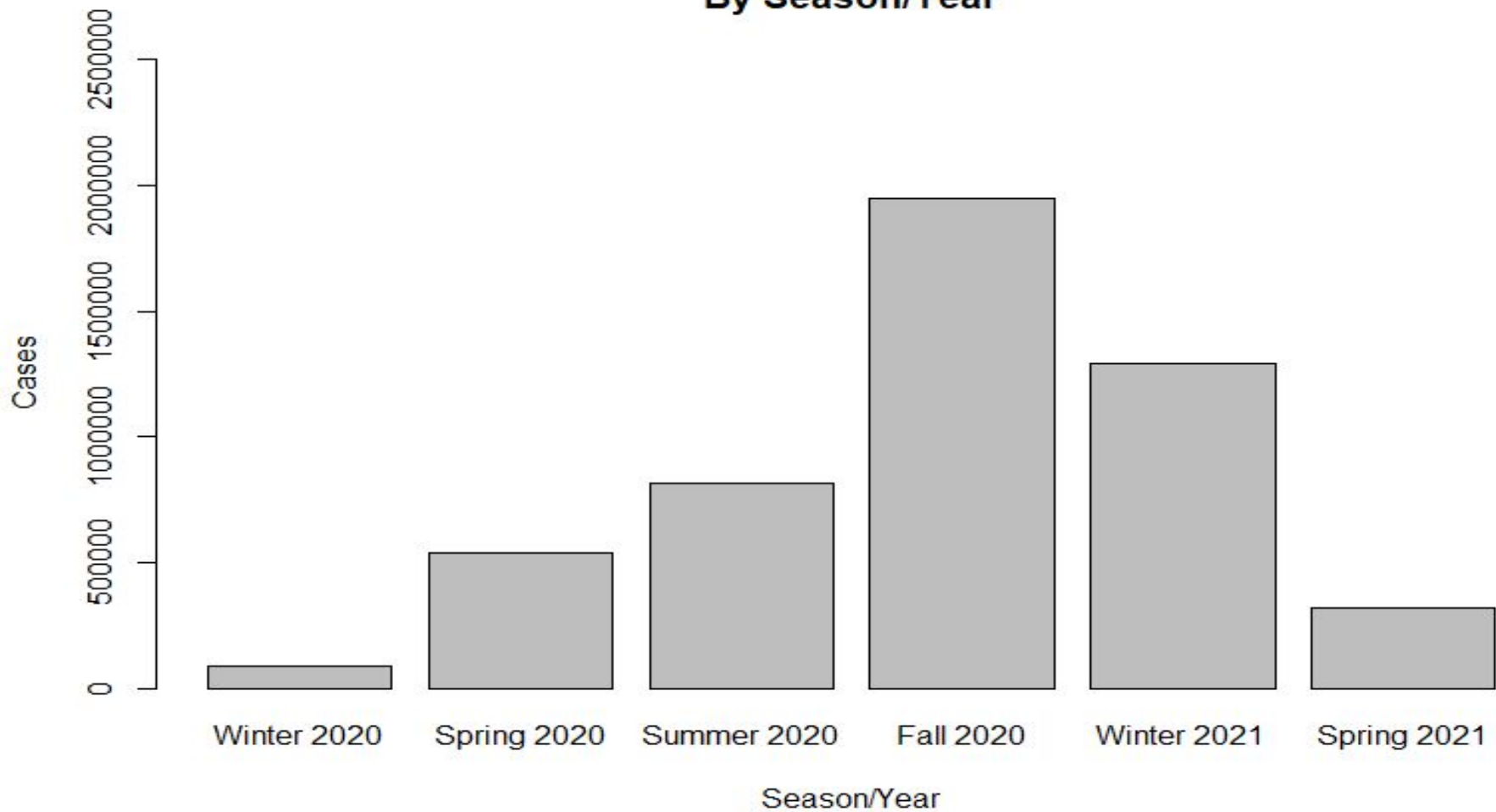
- Quantitatively characterize the relative risks of mortality among individuals diagnosed with COVID-19 based on seven (7) variables: season/year, geographic region, age group, sex, race, hospitalization status, and presence of underlying conditions
- Methods used: frequency distributions, Wilson/score 95% confidence intervals for risk of mortality within various subgroups, multivariable logistic regression



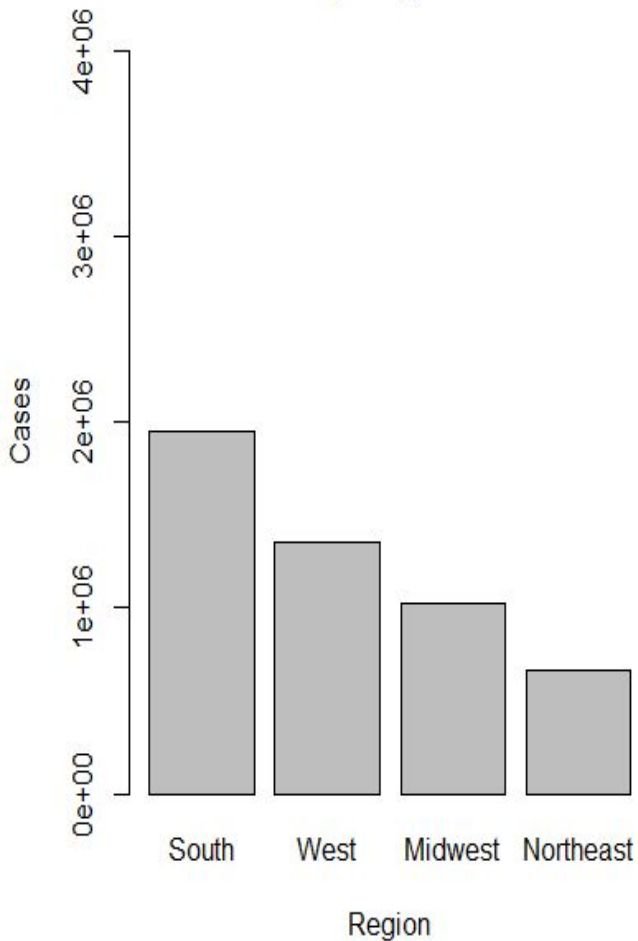
Results

We begin by inspecting the frequency distributions for each of the variables of interest

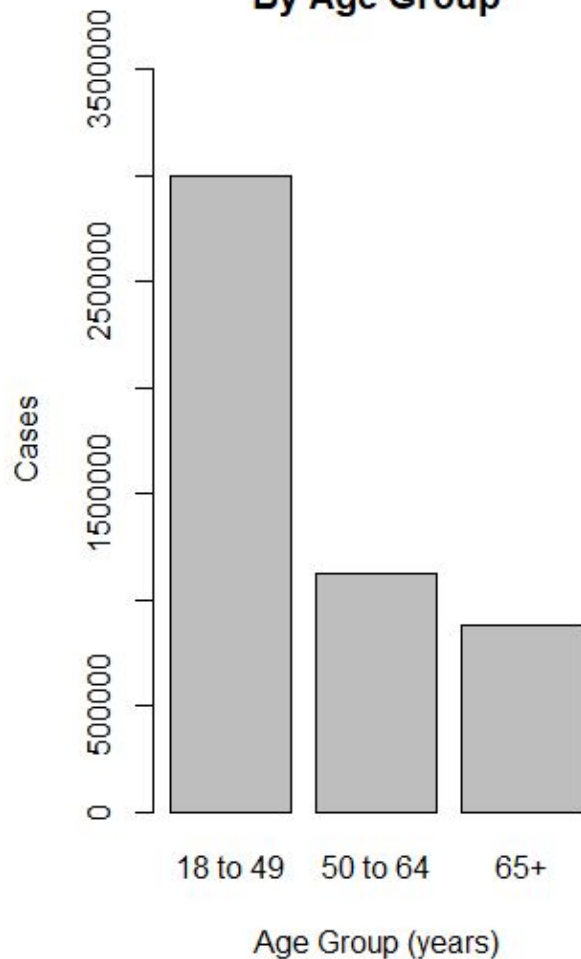
By Season/Year



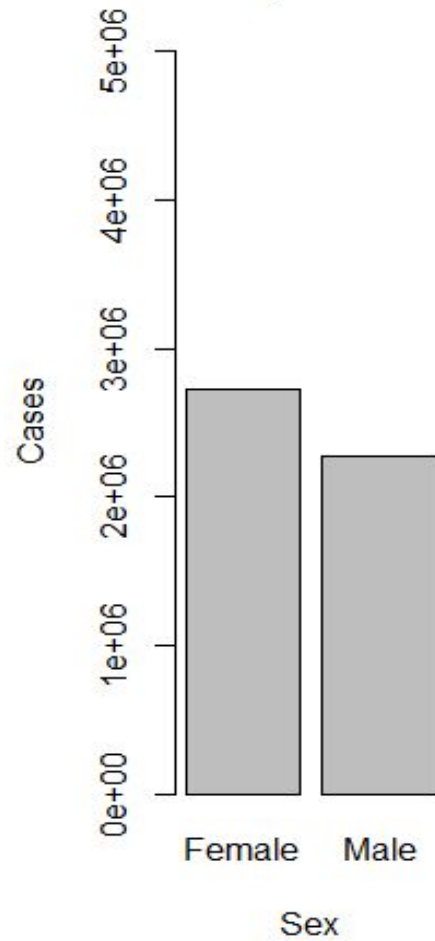
By Region



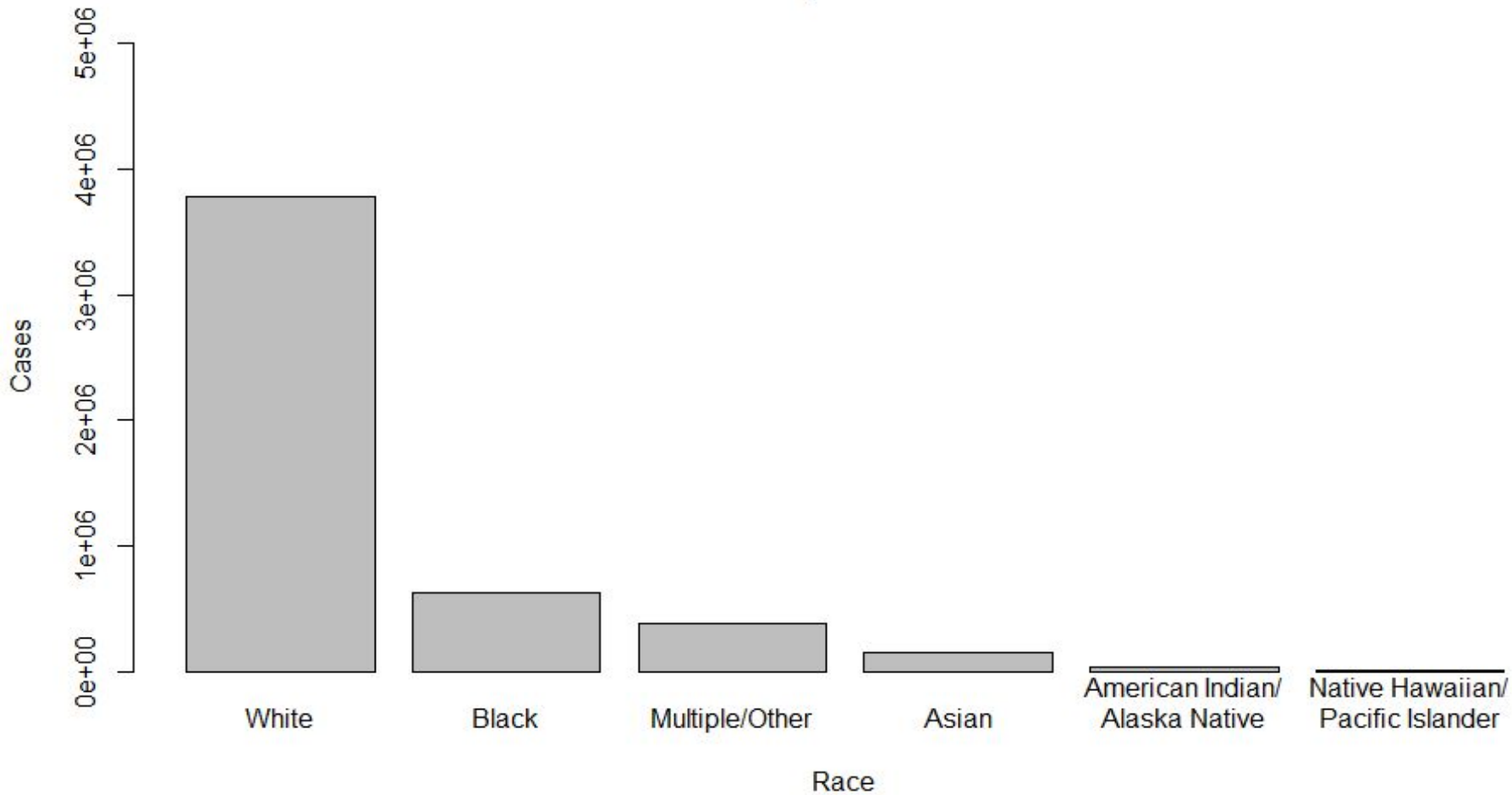
By Age Group



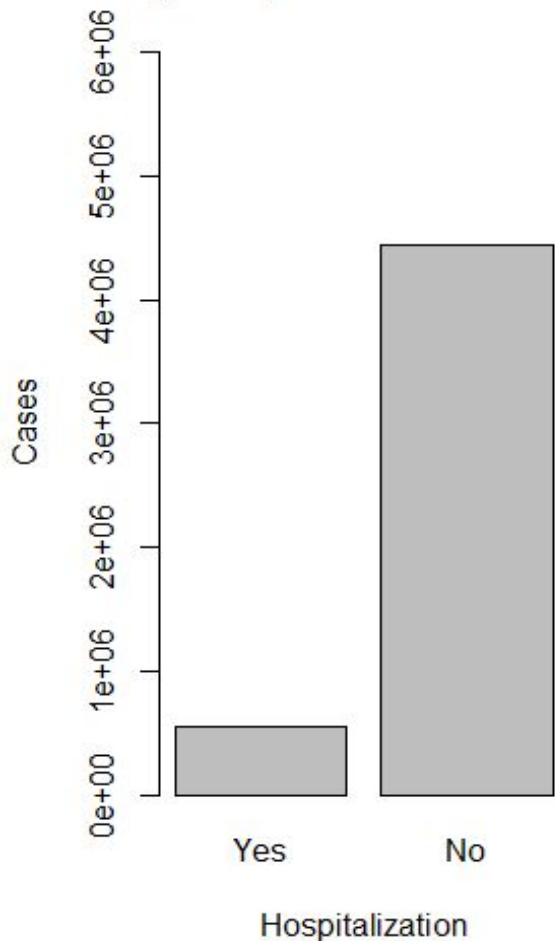
By Sex



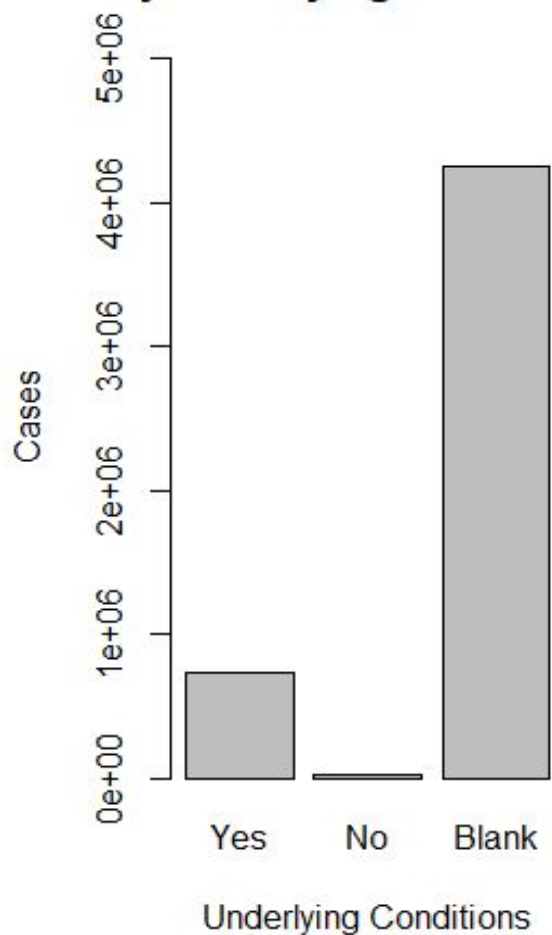
By Race



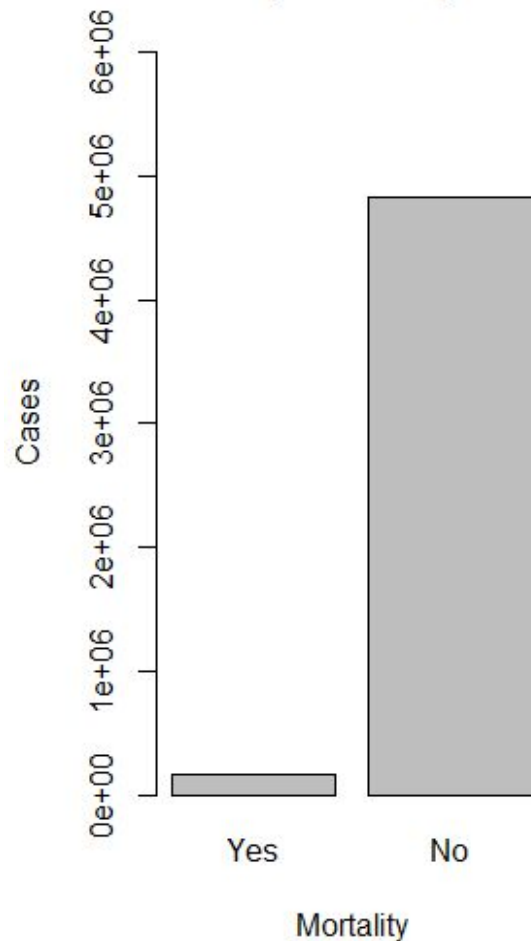
By Hospitalization Status



By Underlying Conditions



By Mortality

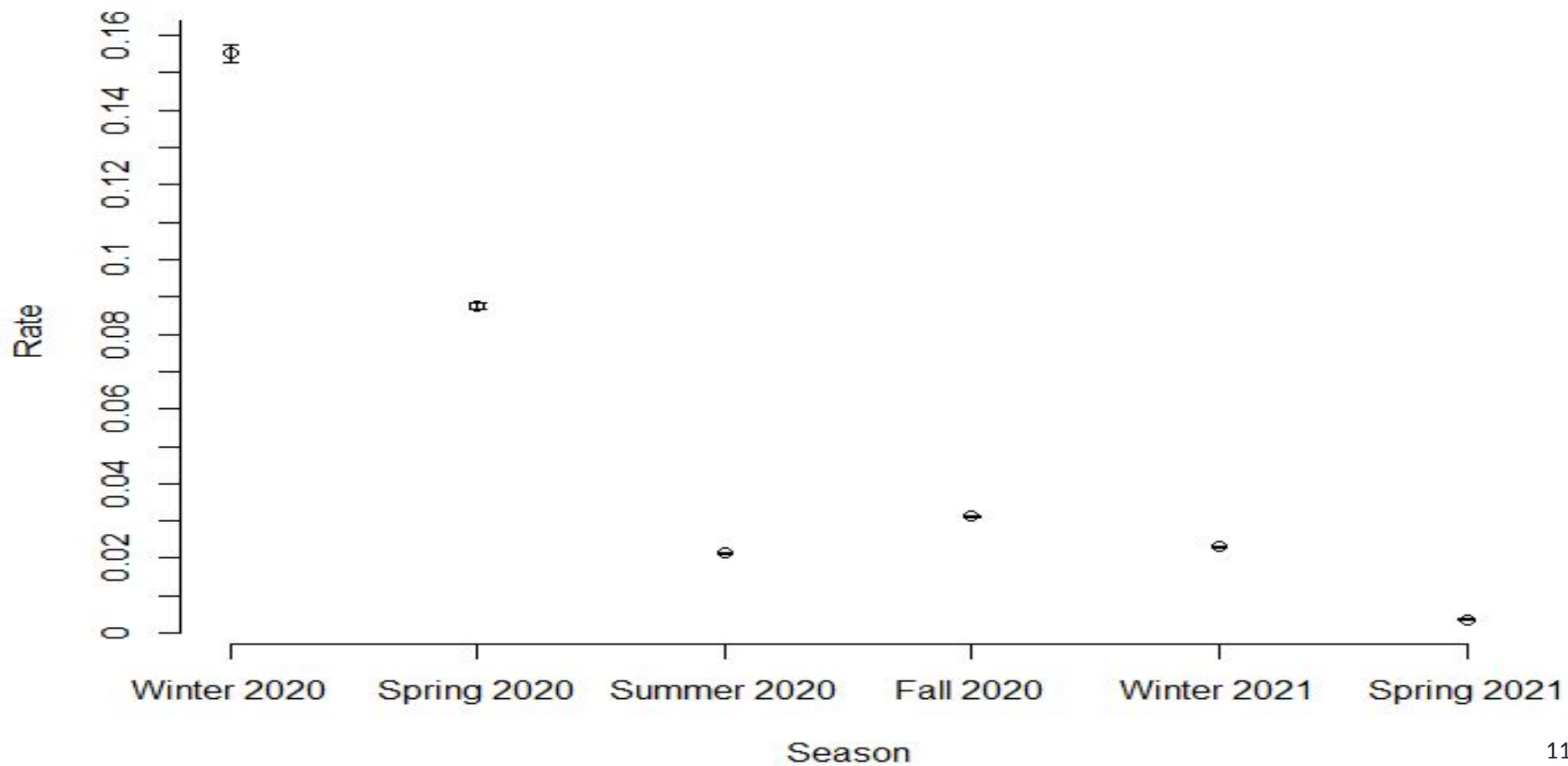




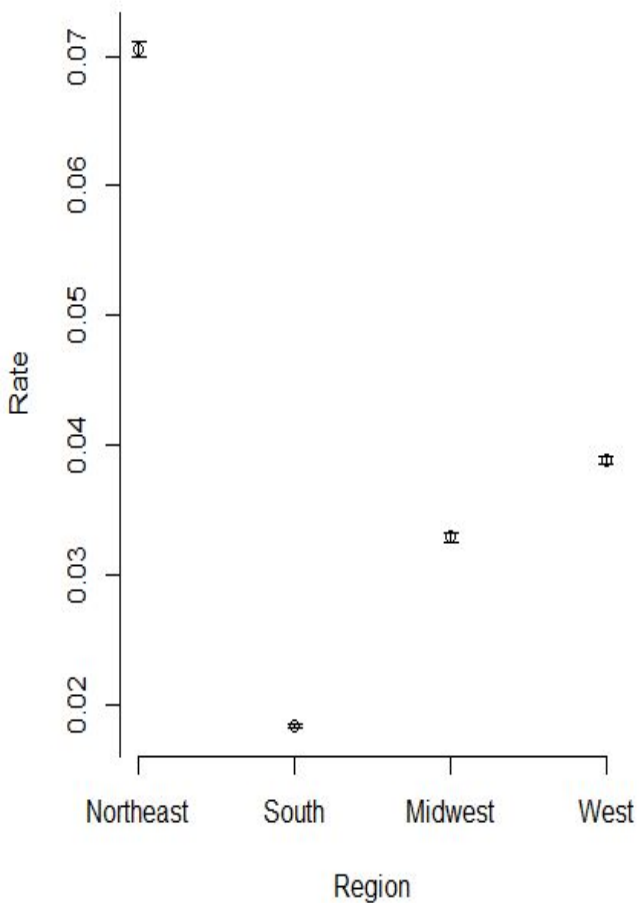
Results

We next inspect 95% confidence intervals for the risk of mortality for the subgroups defined by each of the variables of interest

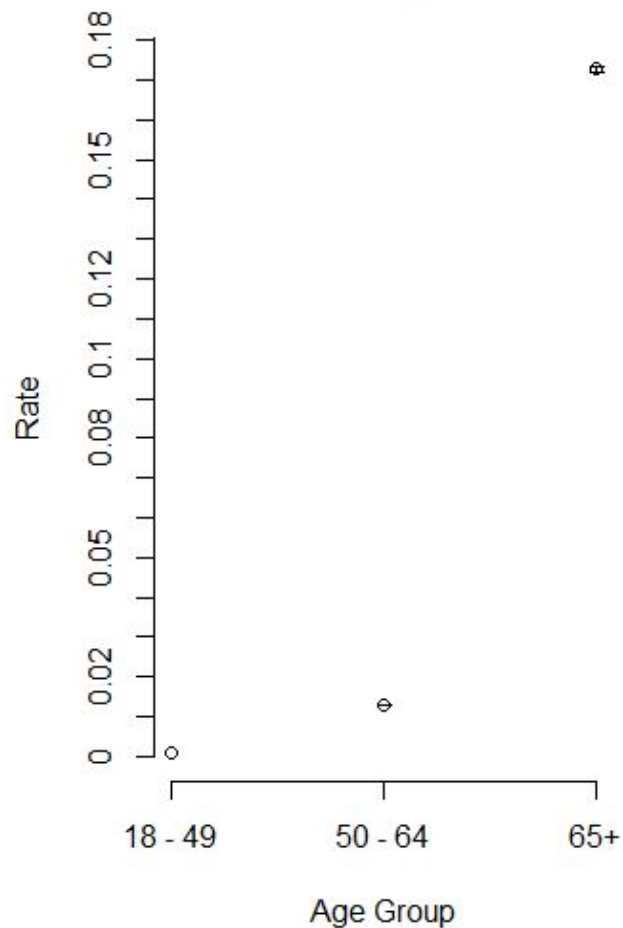
Mortality Rate by Season



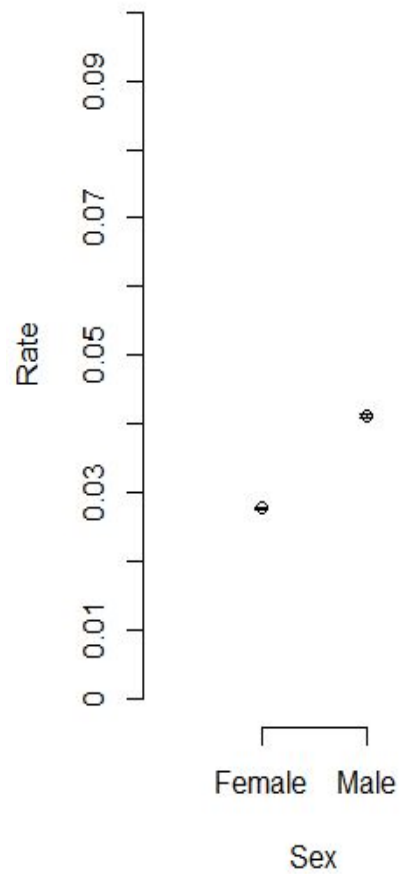
Mortality Rate by Region



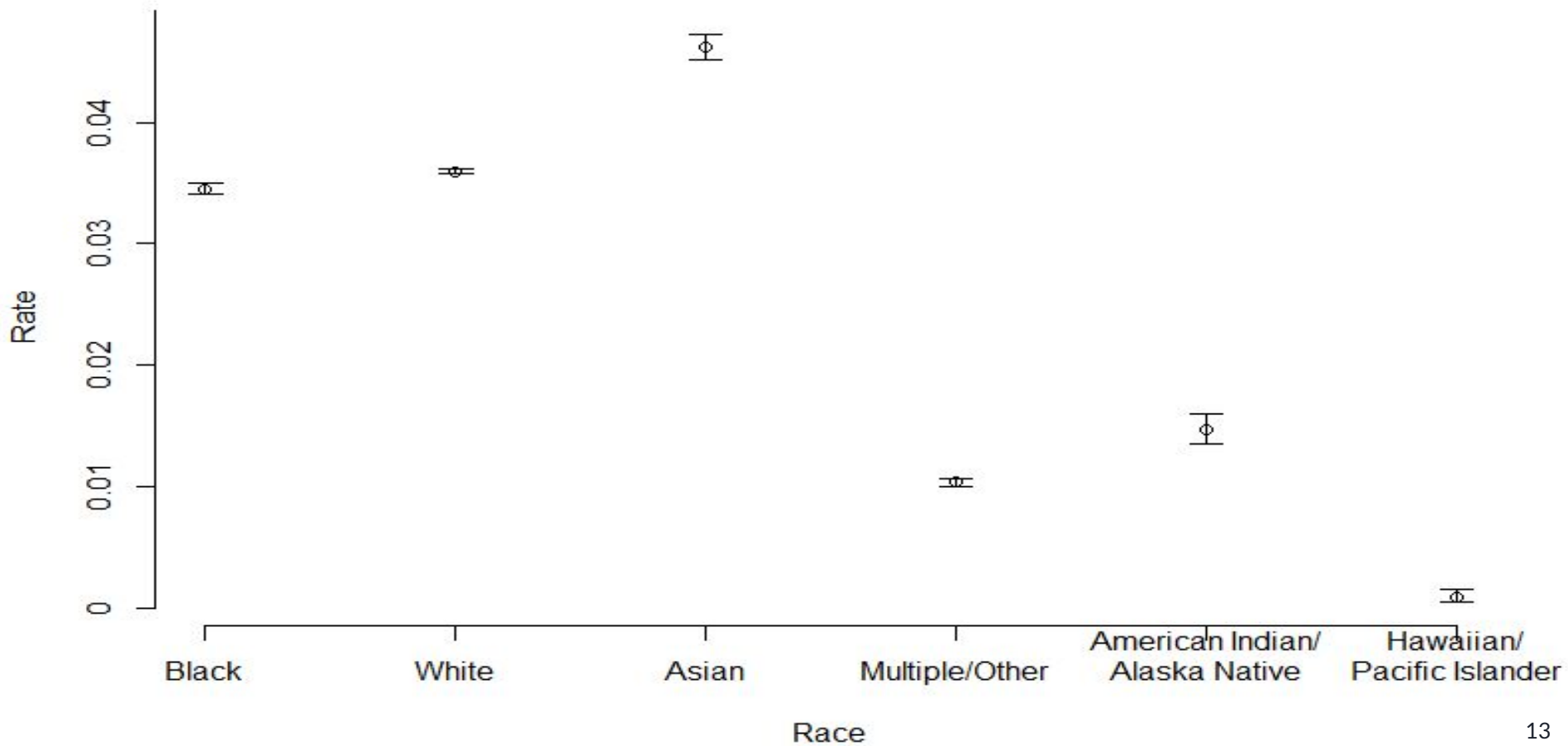
Mortality Rate by Age Group



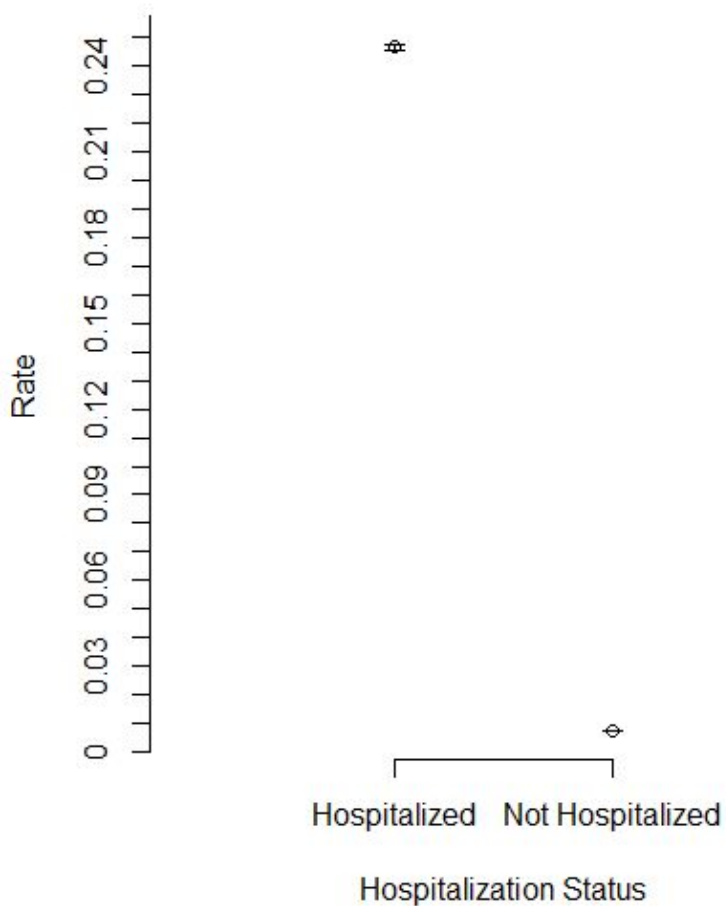
Mortality Rate by Sex



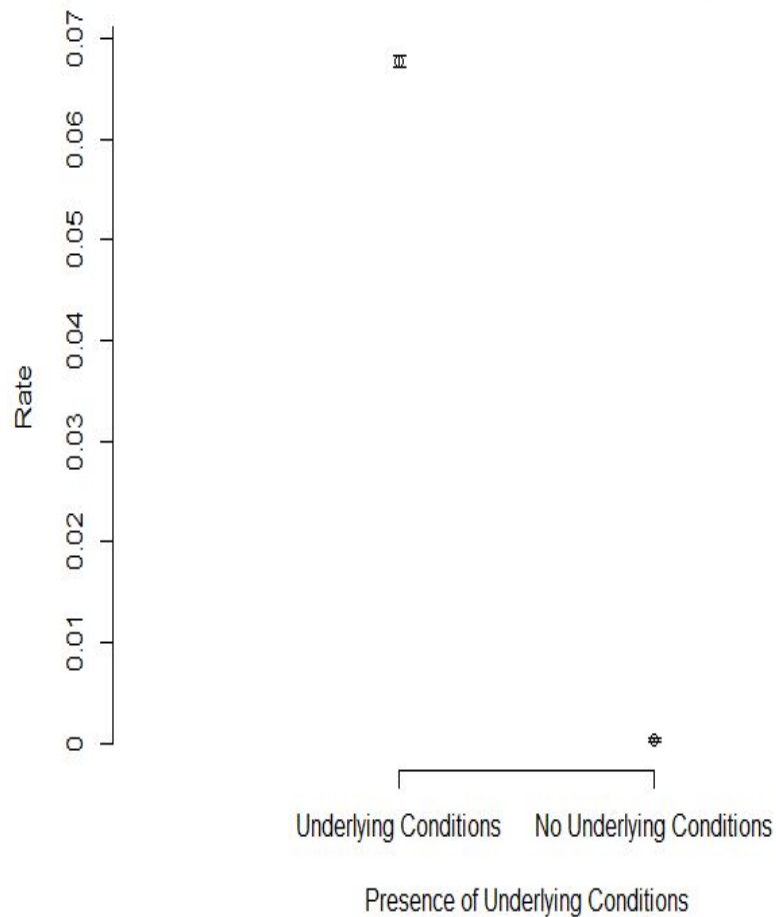
Mortality Rate by Race



Mortality Rate by Hospitalization Status



Mortality Rate by Presence of Underlying Conditions





Logistic Regression Results

First, we will inspect the odds ratio estimates based on the fitted multivariable logistic regression model



Multivariable Logistic Regression Modelling Results

Variable	Odds Ratio	95% Odds Ratio CI
Season/Year		
Fall 2020	5.67	(5.33, 6.02)
Spring 2020	13.65	(12.83, 14.52)
Summer 2020	5.15	(4.84, 5.49)
Winter 2020	13.90	(13.01, 14.84)
Winter 2021	4.43	(4.16, 4.71)
Spring 2021	Ref	Ref
Region		
Midwest	1.20	(1.18, 1.22)
Northeast	1.62	(1.59, 1.65)
West	1.72	(1.69, 1.75)
South	Ref	Ref



Multivariable Logistic Regression Modelling Results

Variable	Odds Ratio	95% Odds Ratio CI
Age Group		
18 - 49	Ref	Ref
50 - 64	9.09	(8.70, 9.49)
65+	90.50	(86.90, 94.25)
Sex		
Female	Ref	Ref
Male	1.37	(1.36, 1.39)



Multivariable Logistic Regression Modelling Results

Variable	Odds Ratio	95% Odds Ratio CI
Race		
American Indian/Alaska Native	0.95	(0.86, 1.06)
Asian	1.02	(0.98, 1.05)
Black	0.88	(0.87, 0.90)
Multiple/Other	0.41	(0.39, 0.42)
Native Hawaiian/Other Pacific Islander	0.03	(0.02, 0.06)
White	Ref	Ref




Multivariable Logistic Regression Modelling Results

Variable	Odds Ratio	95% Odds Ratio CI
Hospitalization Status		
Yes	12.71	(12.54, 12.88)
No	Ref	Ref
Underlying Conditions		
Blank	22.13	(8.30, 59.02)
Yes	26.91	(10.09, 71.76)
No	Ref	Ref



Logistic Regression Results

Second, we will inspect estimated risks of mortality based on hypothetical individual profiles



Estimated Risk of Mortality Based on Hypothetical Individual Profiles

Profile 1: Highest Risk

Estimated Risk of Mortality: 0.71 (71%)

Season/Year: Winter 2020

Region: West

Age: 68

Sex: Male

Race: Asian

Hospitalized: Yes

Underlying Conditions: Yes



Estimated Risk of Mortality Based on Hypothetical Individual Profiles

Profile 2: Lowest Risk

Estimated Risk of Mortality: 0.00000256 (0.000003%)

Season/Year: Spring 2021

Region: South

Age: 32

Sex: Female

Race: Native Hawaiian

Hospitalized: No

Underlying Conditions: No



Estimated Risk of Mortality Based on Hypothetical Individual Profiles

Profile 3: Medium Risk

Estimated Risk of Mortality: 0.00158 (0.16%)

Season/Year: Fall 2020

Region: Northeast

Age: 54

Sex: Male

Race: Black

Hospitalized: Yes

Underlying Conditions: No



Estimated Risk of Mortality Based on Hypothetical Individual Profiles

Profile 4: Medium Risk

Estimated Risk of Mortality: 0.0011 (0.11%)

Season/Year: Spring 2020

Region: Midwest

Age: 20

Sex: Female

Race: White

Hospitalized: No

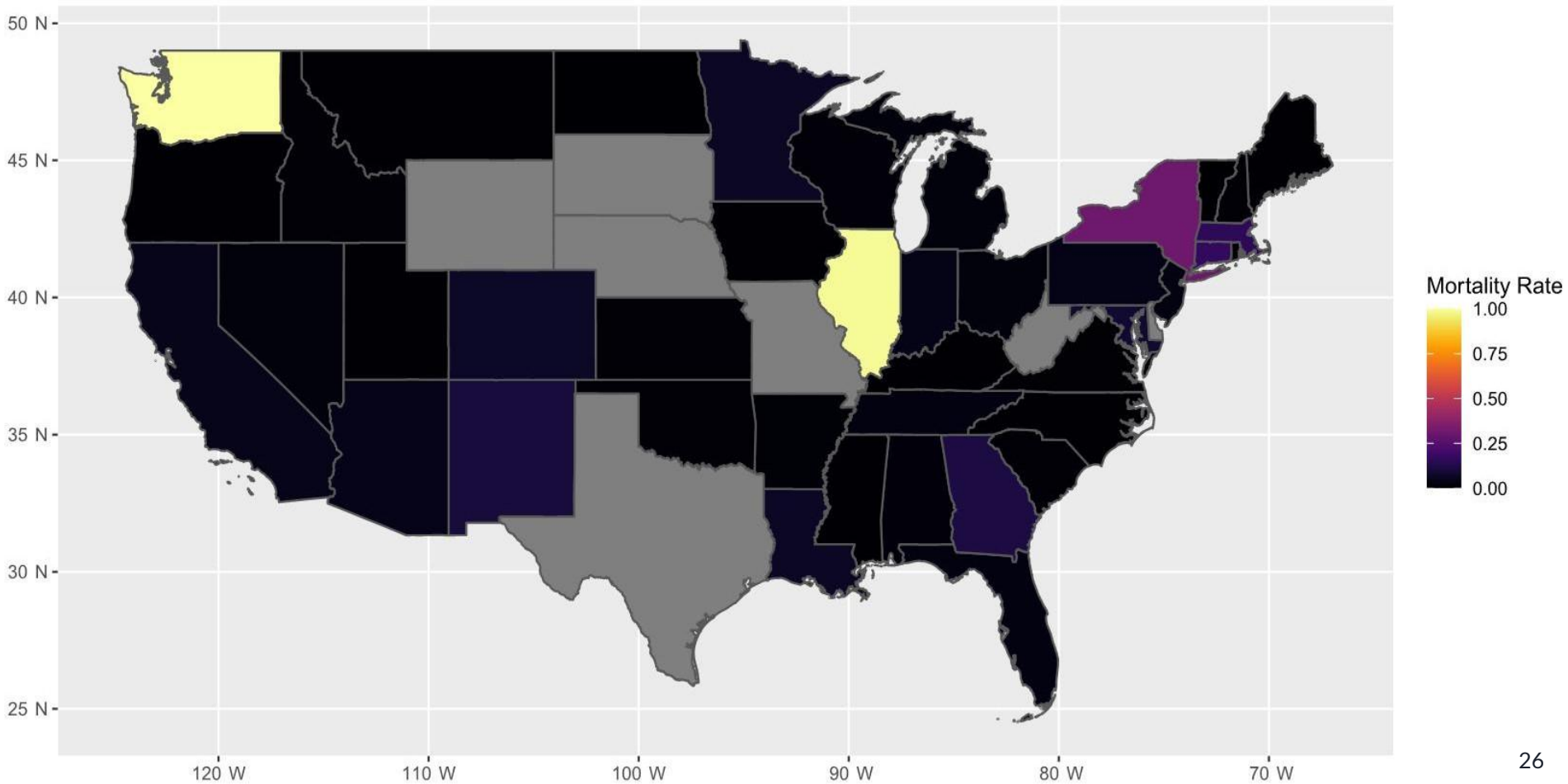
Underlying Conditions: Yes



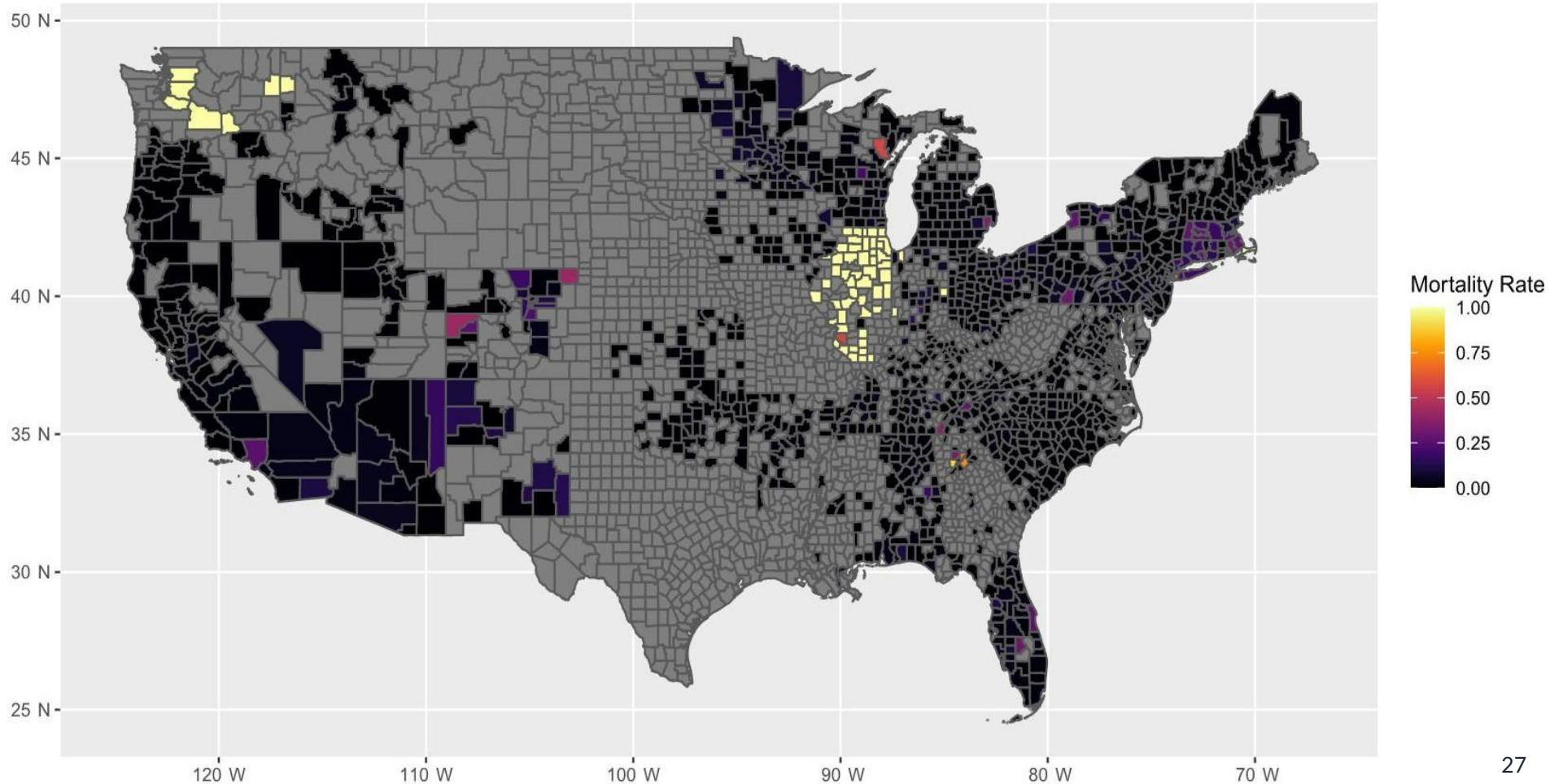
Limitations

- Main limitation: CDC database is incomplete
 - Different recording methods between counties → no data for some variables in some counties
 - No data from many counties → CDC database may therefore lead to biased results

Mortality Rate by State



Mortality Rate by County





Limitations

- Elimination of missing/unknown values for variables of interest reduced our dataset to around 5 million records → possible bias induced by missingness
 - Example: No deaths in 0-17 year age group within our dataset, though we know there were deaths within that age group
 - In the CDC database, around 50% of cases in 0-17 year age group had missing mortality values; no deaths reported in the remaining 50%
 - 0-17 age group therefore eliminated from inferential analyses



Acknowledgements


- Dr. Joseph Cavanaugh
- Jacob Seedorff
- Dr. Gideon Zamba
- Terry Kirk



Thank you!

Questions?

Race (via Census)	Proportion	Race (via CDC database)	Proportion of cases
White	72.5%	White	76.5%
Black	12.7%	Black	12.8%
Asian	5.5%	Asian	3.1%
Multiple/Other	3.3%	Multiple/Other	7.6%
American Indian/Alaska Native	0.8%	American Indian/Alaska Native	0.8%
Native Hawaiian/Other Pacific Islander	0.2%	Native Hawaiian/Other Pacific Islander	0.2%



Region (via Census)	Proportion	Region (via our database)	Proportion of cases
South	38.3%	South	39.0%
West	23.9%	West	27.2%
Midwest	20.8%	Midwest	20.5%
Northeast	17.1%	Northeast	13.3%

Census percentages from https://www.census.gov/popclock/print.php?component=growth&image=//www.census.gov/popclock/share/images/growth_1561939200.png