

Homework 4 - COVID-19 Study and Framingham Study

Aaron Tom Neil Song Ben Zhao Luke Melcher Tanvi Kigga

Due: July 26 10pm

1 Case Study 1: COVID-19 Case Study

1.1 Background

The outbreak of the novel Corona virus disease 2019 (COVID-19) [was declared a public health emergency of international concern by the World Health Organization \(WHO\) on January 30, 2020](#). Upwards of [112 million cases have been confirmed worldwide, with nearly 2.5 million associated deaths](#). Within the US alone, there have been [over 500,000 deaths and upwards of 28 million cases reported](#). Governments around the world have implemented and suggested a number of policies to lessen the spread of the pandemic, including mask-wearing requirements, travel restrictions, business and school closures, and even stay-at-home orders. The global pandemic has impacted the lives of individuals in countless ways, and though many countries have begun vaccinating individuals, the long-term impact of the virus remains unclear.

The impact of COVID-19 on a given segment of the population appears to vary drastically based on the socioeconomic characteristics of the segment. In particular, differing rates of infection and fatalities have been reported among different [racial groups](#), [age groups](#), and [socioeconomic groups](#). One of the most important metrics for determining the impact of the pandemic is the death rate, which is the proportion of people within the total population that die due to the disease.

We assemble this dataset for our research with the goal to investigate the effectiveness of lockdown on flattening the COVID curve. We provide a portion of the cleaned dataset for this case study.

There are two main goals for this case study.

1. We show the development of COVID cases and COVID-related death at state level.
2. We try to figure out what county-level demographic and policy interventions are associated with mortality rate in the US. We try to construct models to find possible factors related to county-level COVID-19 mortality rates.

Remark: please keep track with the most updated version of this write-up.

1.2 Data Summary

The data comes from several different sources:

1. [County-level infection and fatality data](#) - This dataset gives daily cumulative numbers on infection and fatality for each county.
 - [NYC data](#)
2. [County-level socioeconomic data](#) - The following are the four relevant datasets from this site.

- i. Income - Poverty level and household income.
 - ii. Jobs - Employment type, rate, and change.
 - iii. People - Population size, density, education level, race, age, household size, and migration rates.
 - iv. County Classifications - Type of county (rural or urban on a rural-urban continuum scale).
3. [Intervention Policy Data](#) - This dataset is a manually compiled list of the dates that interventions/lockdown policies were implemented and lifted at the county level.

1.3 Two Main Datasets

In this case study, we use the following two cleaned data:

- **covid_county.csv**: County-level socioeconomic information that combines the above-mentioned 4 datasets: Income (Poverty level and household income), Jobs (Employment type, rate, and change), People (Population size, density, education level, race, age, household size, and migration rates), County Classifications
- **covid_rates.csv**: Daily cumulative numbers on infection and fatality for each county

Among all data, the unique identifier of county is FIPS.

First read in the data.

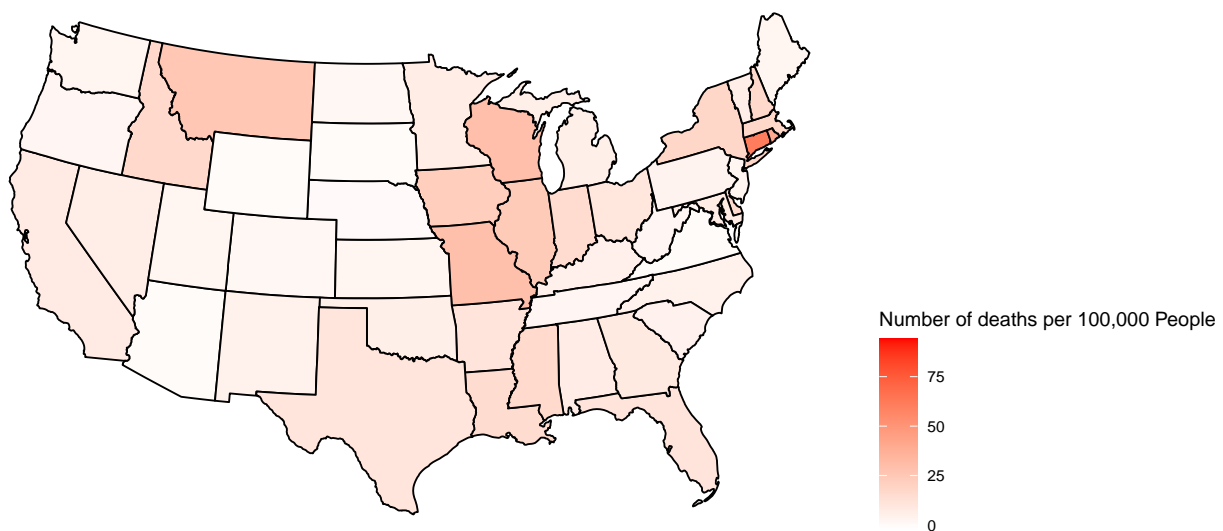
The detailed description of variables is in **Appendix: Data description**. Please get familiar with the variables.

1.4 COVID death trend (Homework: question i-vi)

- i) For each month in 2020, plot the monthly deaths per 100k heatmap by state on US map. Use the same color range across months. (Hints: Set `limits` argument in `scale_fill_gradient()` to fix the color scale. Use `facet_wrap()` to plot by month.)

state death rate per month

Continental US States



Important Remark:

1. Using the following R chunk, we processed the data into the right format `covid_monthly` for you :) Specifically, we use `lubridate::month()` and `lubridate::year()` to extract month and year from date, use `tidyr::complete(state, month, fill = list(new_case_per100k = NA))` to complete the missing months with no cases.
2. Use the data `covid_monthly` after running the R chunk below to create the heatmap

1.5 COVID factor

We now try to build a good parsimonious model to find possible factors related to death rate on county level. Let us not take time series into account for the moment and use the total number as of *Feb 1, 2021*. (Hint: use `dplyr::filter()`.)

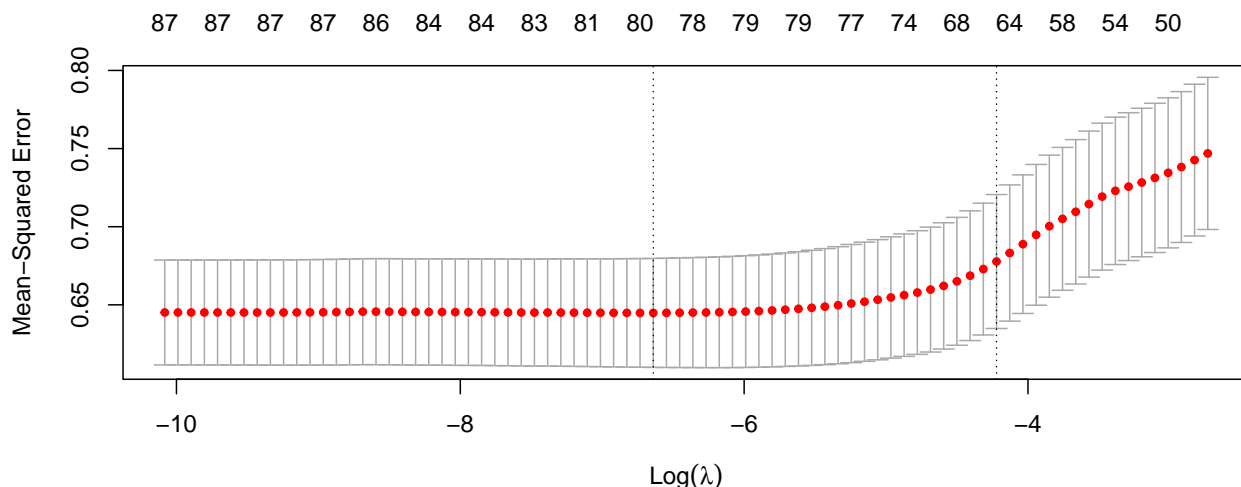
- ii) Create the response variable `total_death_per100k` as the total of number of COVID deaths per 100k by *Feb 1, 2021*. We suggest to take log transformation as `log_total_death_per100k = log(total_death_per100k + 1)`. Merge `total_death_per100k` to `county_data` for the following analysis. (Hint: check the R function `dplyr::join()` or `data.table::merge()`)
- iii) Select possible variables in `county_data` as covariates. We provide `county_data_sub`, a subset variables from `county_data`, for you to get started. Please add any potential variables as you wish.

Report missing values in your final subset of variables. In the following analysis, you may ignore the missing values.

In the final subset of variables, we remove the “county” variable, as we already have a FIPS code for each county. Some counties may be named the same, thus affecting the linear model when we try to create it. The rest of the values may all have a factor that we would like to consider. We also added the “log_death_rate” variable in order to gain a more accurate data variable that isn’t skewed to the left.

- iv) Use LASSO to choose a parsimonious model with all available sensible county-level information. Use `set.seed(15)`. **Force in State** in the process. Why we need to force in State?

We need to force in State because without adding the penalty factor, all the variables from the data set would be included, but we are only interested in the state and each state’s county data. Thus, we have to add a penalty (that being the number 1) to each state and the penalty 0 to each non-state variable. In turn, when we run `cv.glmnet` we can return a model that only contains the State variables.



- v) Use `cv.glmnet()` to perform LASSO and land on a parsimonious model. You can use `lambda.min`, `lambda.1se`, or any sensible λ . Follow the chunk below and use `set.seed(15)`. **Force in State** in the process using the `penalty.factor` argument in `cv.glmnet()`. Why we need to force in State?
- vi) If necessary, reduce the model from v) to a final model with all variables being significant at 0.05 level. Use `Anova()` for backward selection. Again always keep **State** in the model during the process. Are the linear model assumptions all reasonably met?

Now we will be removing the variables with very high P value.

Call:

```
lm(formula = log_death_rate ~ PovertyAllAgesPct + PerCapitaInc +
    UnempRate2019 + PctEmpFIRE + PctEmpConstruction + PctEmpTrans +
    PctEmpTrade + PctEmpAgriculture + PctEmpManufacturing + PctEmpServices +
    PopDensity2010 + Age65AndOlderPct2010 + Under18Pct2010 +
    Ed2HSDiplomaOnlyPct + Ed3SomeCollegePct + Ed5CollegePlusPct +
    NetMigrationRate1019 + NaturalChangeRate1019 + WhiteNonHispanicPct2010 +
    NativeAmericanNonHispanicPct2010 + HispanicPct2010 + RuralUrbanContinuumCode2013 +
    Perpov_1980_0711 + HiAmenity + State, data = covid_county_0201.sub)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.690	-0.262	0.059	0.368	3.686

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.95e+00	6.69e-01	5.90	3.9e-09	***
PovertyAllAgesPct	1.31e-02	5.81e-03	2.26	0.02369	*
PerCapitaInc	-9.25e-06	5.96e-06	-1.55	0.12075	
UnempRate2019	-6.40e-02	1.67e-02	-3.83	0.00013	***
PctEmpFIRE	3.76e-02	1.07e-02	3.52	0.00043	***
PctEmpConstruction	-2.88e-02	7.80e-03	-3.69	0.00023	***
PctEmpTrans	1.82e-02	9.13e-03	2.00	0.04577	*
PctEmpTrade	1.61e-02	6.67e-03	2.41	0.01620	*
PctEmpAgriculture	-2.19e-02	5.21e-03	-4.21	2.6e-05	***
PctEmpManufacturing	1.68e-02	4.38e-03	3.83	0.00013	***
PctEmpServices	1.59e-02	5.10e-03	3.12	0.00185	**
PopDensity2010	2.08e-05	9.27e-06	2.24	0.02499	*
Age65AndOlderPct2010	4.54e-02	8.14e-03	5.58	2.7e-08	***
Under18Pct2010	5.74e-02	8.31e-03	6.90	6.1e-12	***
Ed2HSDiplomaOnlyPct	-9.66e-03	5.57e-03	-1.74	0.08276	.
Ed3SomeCollegePct	-3.32e-02	6.45e-03	-5.14	2.9e-07	***
Ed5CollegePlusPct	-2.04e-02	5.42e-03	-3.77	0.00017	***
NetMigrationRate1019	-1.41e-02	2.64e-03	-5.36	9.1e-08	***
NaturalChangeRate1019	-4.47e-02	9.96e-03	-4.49	7.6e-06	***
WhiteNonHispanicPct2010	-3.91e-03	1.79e-03	-2.19	0.02863	*
NativeAmericanNonHispanicPct2010	4.61e-03	3.22e-03	1.43	0.15239	
HispanicPct2010	7.26e-03	2.51e-03	2.90	0.00382	**
RuralUrbanContinuumCode2013	-3.06e-02	8.66e-03	-3.53	0.00042	***
Perpov_1980_0711	-1.90e-01	6.38e-02	-2.97	0.00300	**
HiAmenity	-1.19e-01	4.81e-02	-2.47	0.01366	*
StateAR	1.55e-01	1.36e-01	1.14	0.25369	

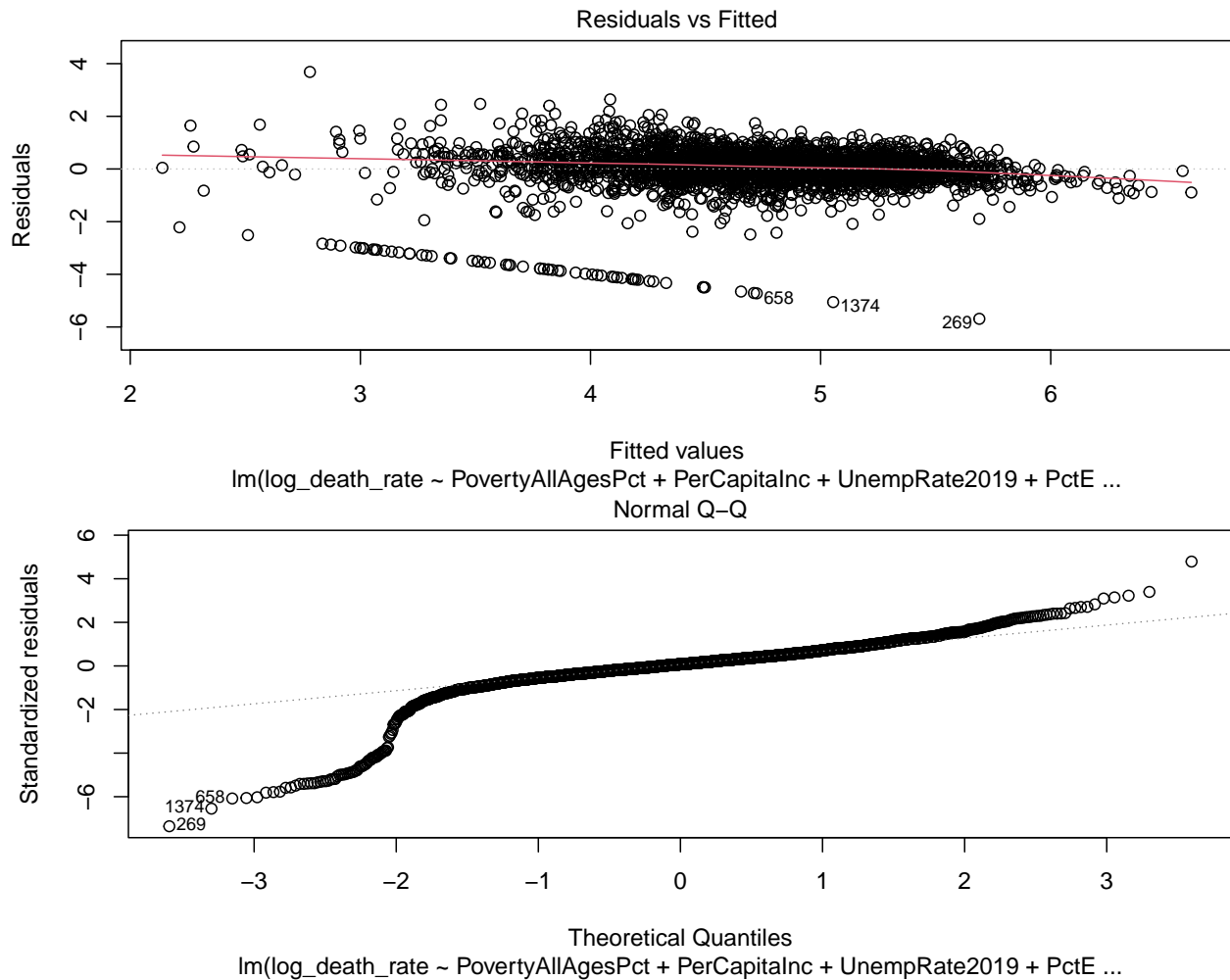
StateAZ	4.75e-01	2.46e-01	1.93	0.05359	.
StateCA	-5.14e-01	1.68e-01	-3.07	0.00217	**
StateCO	-1.51e-01	1.57e-01	-0.96	0.33614	
StateCT	3.25e-01	3.04e-01	1.07	0.28590	
StateDC	8.83e-01	8.05e-01	1.10	0.27308	
StateDE	-4.99e-02	4.66e-01	-0.11	0.91470	
StateFL	5.96e-02	1.49e-01	0.40	0.68821	
StateGA	1.38e-01	1.16e-01	1.19	0.23326	
StateIA	3.09e-01	1.37e-01	2.25	0.02448	*
StateID	-1.13e-01	1.69e-01	-0.67	0.50338	
StateIL	3.11e-01	1.39e-01	2.24	0.02492	*
StateIN	-1.04e-02	1.37e-01	-0.08	0.93996	
StateKS	-5.60e-01	1.39e-01	-4.02	5.9e-05	***
StateKY	-5.95e-01	1.33e-01	-4.47	8.1e-06	***
StateLA	5.00e-01	1.47e-01	3.40	0.00068	***
StateMA	1.45e-01	2.42e-01	0.60	0.54917	
StateMD	9.65e-02	1.98e-01	0.49	0.62655	
StateME	-1.30e+00	2.26e-01	-5.77	8.9e-09	***
StateMI	-4.15e-03	1.44e-01	-0.03	0.97702	
StateMN	-3.81e-02	1.44e-01	-0.26	0.79216	
StateMO	-2.94e-01	1.31e-01	-2.25	0.02466	*
StateMS	4.74e-01	1.39e-01	3.42	0.00064	***
StateMT	7.27e-01	1.64e-01	4.44	9.2e-06	***
StateNC	-3.01e-01	1.29e-01	-2.34	0.01951	*
StateND	9.77e-01	1.71e-01	5.72	1.2e-08	***
StateNE	-3.27e-01	1.45e-01	-2.25	0.02441	*
StateNH	-7.71e-01	2.74e-01	-2.81	0.00498	**
StateNJ	4.08e-01	2.12e-01	1.93	0.05367	.
StateNM	-4.09e-01	2.00e-01	-2.05	0.04061	*
StateNV	-2.99e-01	2.32e-01	-1.29	0.19650	
StateNY	-3.02e-01	1.56e-01	-1.94	0.05287	.
StateOH	-4.15e-01	1.42e-01	-2.92	0.00353	**
StateOK	-3.20e-01	1.41e-01	-2.27	0.02356	*
StateOR	-6.13e-01	1.79e-01	-3.42	0.00063	***
StatePA	1.44e-01	1.56e-01	0.92	0.35610	
StateRI	-1.04e-01	3.72e-01	-0.28	0.77929	
StateSC	-5.49e-02	1.52e-01	-0.36	0.71732	
StateSD	8.74e-01	1.59e-01	5.49	4.3e-08	***
StateTN	1.30e-01	1.31e-01	0.99	0.32259	
StateTX	2.40e-01	1.28e-01	1.88	0.05982	.
StateUT	-8.05e-01	1.99e-01	-4.05	5.3e-05	***
StateVA	-4.34e-01	1.24e-01	-3.51	0.00046	***
StateVT	-2.06e+00	2.41e-01	-8.55	< 2e-16	***
StateWA	-5.49e-01	1.79e-01	-3.07	0.00215	**
StateWI	-4.94e-02	1.47e-01	-0.34	0.73648	
StateWV	-5.24e-01	1.64e-01	-3.20	0.00141	**
StateWY	4.93e-01	2.09e-01	2.36	0.01822	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.783 on 3032 degrees of freedom

Multiple R-squared: 0.372, Adjusted R-squared: 0.357

F-statistic: 24.9 on 72 and 3032 DF, p-value: <2e-16



The various assumption of the linear model: We can say the model is homoscedastic and linear as there are equal amounts of points above and below the line. However, with an exception of some abnormalities of the large line at the bottom. This could be due to over fitting. However in the second plot (normal qq): most of the points are not falling on the curve. Therefore, we can conclude that not all assumptions of a linear model are reasonably met.

- vi) Based on your final model, summarize your findings. In particular, summarize the state effect controlling for others. Provide recommendations to policy makers to reduce COVID death rate.

Based on the final model, which the base is a state, some other states when tested against the base have 0 correlation. Other states have high correlations to the base such as California. There are also other variables such as poverty, unemployment, percentage of employees in certain jobs, ages 65 and older, ages under 18, education college percentage, that have a significant affect on the covid death rate/catch based on the base. Recommendations would be to monitor older age people, warn against younger people going out, and warn workforce employees with those specific jobs. People in poverty or unemployed have a high correlation with the covid death rate based on the base.

2 Case Study 2: Framingham heart disease study

2.1 Background

We will continue to use the Framingham Data (`Framingham.dat`) so that you are already familiar with the data and the variables. All the results are obtained through training data.

Liz is a patient with the following readings: AGE=50, SEX=FEMALE, SBP=110, DBP=80, CHOL=180, FRW=105, CIG=0. We would be interested to predict Liz's outcome in heart disease.

To keep our answers consistent, use a subset of the data, and exclude anyone with a missing entry. For your convenience, we've loaded it here together with a brief summary about the data.

We note that this dataset contains 311 people diagnosed with heart disease and 1095 without heart disease.

After a quick cleaning up here is a summary about the data:

HD	AGE	SEX	SBP	DBP	CHOL	FRW	
0:1086	Min. :45.0	FEMALE:730	Min. : 90	Min. : 50.0	Min. : 96	Min. : 52	M
1: 307	1st Qu.:48.0	MALE :663	1st Qu.:130	1st Qu.: 80.0	1st Qu.:200	1st Qu.: 94	1
	Median :52.0		Median :142	Median : 90.0	Median :230	Median :103	M
	Mean :52.4		Mean :148	Mean : 90.2	Mean :235	Mean :105	M
	3rd Qu.:56.0		3rd Qu.:160	3rd Qu.: 98.0	3rd Qu.:264	3rd Qu.:114	3
	Max. :62.0		Max. :300	Max. :160.0	Max. :430	Max. :222	M

2.2 Identify risk factors for Heart.Disease (Homework: question viii-x)

- viii. Start with a full model using all the covariates. Use backward selection to get a model with all variables significant at 0.1 level. Summarize your final model.

```
Call:
glm(formula = HD ~ AGE + SEX + SBP + CHOL + CIG, family = binomial,
     data = hd_data.f)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.754  -0.729  -0.554  -0.344   2.447
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.70228    0.92683  -9.39 < 2e-16 ***
AGE           0.06136    0.01475   4.16 3.2e-05 ***
SEXMALE      0.88575    0.15579   5.69 1.3e-08 ***
SBP           0.01709    0.00237   7.20 5.9e-13 ***
CHOL          0.00440    0.00150   2.93 0.0033 **
CIG           0.01136    0.00606   1.87 0.0608 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1469.3 on 1392 degrees of freedom
Residual deviance: 1345.5 on 1387 degrees of freedom
```

AIC: 1358

Number of Fisher Scoring iterations: 4

- ix. Based on the model in viii, calculate the mis-classification error (MCE) with $1/2$ and $1/3$ as the threshold respectively. Which threshold achieves a smaller MCE?

A threshold with $1/2$ has a smaller MCE.

- x. Let predict whether Liz has heart disease or not using the model in viii. Use $1/3$ as the threshold to be conservative. Why we would rather be conservative in this case? Could you come up with a better metric to measure the cost of mis-classification?

Liz is predicted to not have heart disease, as the predicted probability 0.0491 is lower than 0.333 . We would rather be conservative because false negative is much more expensive than false positive. A better metric to measure the cost of mis-classification is optimization over false positive and false negative cost.

END

Appendix: COVID data description

A detailed summary of the variables in each data set follows:

Infection and fatality data

- date: Date
- county: County name
- state: State name
- fips: County code that uniquely identifies a county
- cases: Number of cumulative COVID-19 infections
- deaths: Number of cumulative COVID-19 deaths

Socioeconomic demographics

Income: Poverty level and household income

- PovertyUnder18Pct: Poverty rate for children age 0-17, 2018
- Deep_Pov_All: Deep poverty, 2014-18
- Deep_Pov_Children: Deep poverty for children, 2014-18
- PovertyAllAgesPct: Poverty rate, 2018
- MedHHInc: Median household income, 2018 (In 2018 dollars)
- PerCapitaInc: Per capita income in the past 12 months (In 2018 inflation adjusted dollars), 2014-18
- PovertyAllAgesNum: Number of people of all ages in poverty, 2018
- PovertyUnder18Num: Number of people age 0-17 in poverty, 2018

Jobs: Employment type, rate, and change

- UnempRate2007-2019: Unemployment rate, 2007-2019
- NumEmployed2007-2019: Employed, 2007-2019
- NumUnemployed2007-2019: Unemployed, 2007-2019
- PctEmpChange1019: Percent employment change, 2010-19
- PctEmpChange1819: Percent employment change, 2018-19
- PctEmpChange0719: Percent employment change, 2007-19
- PctEmpChange0710: Percent employment change, 2007-10
- NumCivEmployed: Civilian employed population 16 years and over, 2014-18
- NumCivLaborforce2007-2019: Civilian labor force, 2007-2019
- PctEmpFIRE: Percent of the civilian labor force 16 and over employed in finance and insurance, and real estate and rental and leasing, 2014-18
- PctEmpConstruction: Percent of the civilian labor force 16 and over employed in construction, 2014-18
- PctEmpTrans: Percent of the civilian labor force 16 and over employed in transportation, warehousing and utilities, 2014-18
- PctEmpMining: Percent of the civilian labor force 16 and over employed in mining, quarrying, oil and gas extraction, 2014-18
- PctEmpTrade: Percent of the civilian labor force 16 and over employed in wholesale and retail trade, 2014-18
- PctEmpInformation: Percent of the civilian labor force 16 and over employed in information services, 2014-18
- PctEmpAgriculture: Percent of the civilian labor force 16 and over employed in agriculture, forestry, fishing, and hunting, 2014-18
- PctEmpManufacturing: Percent of the civilian labor force 16 and over employed in manufacturing, 2014-18
- PctEmpServices: Percent of the civilian labor force 16 and over employed in services, 2014-18
- PctEmpGovt: Percent of the civilian labor force 16 and over employed in public administration, 2014-18

People: Population size, density, education level, race, age, household size, and migration rates

- PopDensity2010: Population density, 2010
- LandAreaSQMiles2010: Land area in square miles, 2010
- TotalHH: Total number of households, 2014-18
- TotalOccHU: Total number of occupied housing units, 2014-18
- AvgHHSize: Average household size, 2014-18
- OwnHomeNum: Number of owner occupied housing units, 2014-18
- OwnHomePct: Percent of owner occupied housing units, 2014-18
- NonEnglishHHPct: Percent of non-English speaking households of total households, 2014-18
- HH65PlusAlonePct: Percent of persons 65 or older living alone, 2014-18
- FemaleHHPct: Percent of female headed family households of total households, 2014-18
- FemaleHHNum: Number of female headed family households, 2014-18
- NonEnglishHHNum: Number of non-English speaking households, 2014-18
- HH65PlusAloneNum: Number of persons 65 years or older living alone, 2014-18
- Age65AndOlderPct2010: Percent of population 65 or older, 2010
- Age65AndOlderNum2010: Population 65 years or older, 2010
- TotalPop25Plus: Total population 25 and older, 2014-18 - 5-year average
- Under18Pct2010: Percent of population under age 18, 2010
- Under18Num2010: Population under age 18, 2010
- Ed1LessThanHSPct: Percent of persons with no high school diploma or GED, adults 25 and over, 2014-18
- Ed2HSDiplomaOnlyPct: Percent of persons with a high school diploma or GED only, adults 25 and over, 2014-18
- Ed3SomeCollegePct: Percent of persons with some college experience, adults 25 and over, 2014-18
- Ed4AssocDegreePct: Percent of persons with an associate's degree, adults 25 and over, 2014-18

- Ed5CollegePlusPct: Percent of persons with a 4-year college degree or more, adults 25 and over, 2014-18
- Ed1LessThanHSNum: No high school, adults 25 and over, 2014-18
- Ed2HSDiplomaOnlyNum: High school only, adults 25 and over, 2014-18
- Ed3SomeCollegeNum: Some college experience, adults 25 and over, 2014-18
- Ed4AssocDegreeNum: Number of persons with an associate's degree, adults 25 and over, 2014-18
- Ed5CollegePlusNum: College degree 4-years or more, adults 25 and over, 2014-18
- ForeignBornPct: Percent of total population foreign born, 2014-18
- ForeignBornEuropePct: Percent of persons born in Europe, 2014-18
- ForeignBornMexPct: Percent of persons born in Mexico, 2014-18
- ForeignBornCentralSouthAmPct: Percent of persons born in Central or South America, 2014-18
- ForeignBornAsiaPct: Percent of persons born in Asia, 2014-18
- ForeignBornCaribPct: Percent of persons born in the Caribbean, 2014-18
- ForeignBornAfricaPct: Percent of persons born in Africa, 2014-18
- ForeignBornNum: Number of people foreign born, 2014-18
- ForeignBornCentralSouthAmNum: Number of persons born in Central or South America, 2014-18
- ForeignBornEuropeNum: Number of persons born in Europe, 2014-18
- ForeignBornMexNum: Number of persons born in Mexico, 2014-18
- ForeignBornAfricaNum: Number of persons born in Africa, 2014-18
- ForeignBornAsiaNum: Number of persons born in Asia, 2014-18
- ForeignBornCaribNum: Number of persons born in the Caribbean, 2014-18
- Net_International_Migration_Rate_2010_2019: Net international migration rate, 2010-19
- Net_International_Migration_2010_2019: Net international migration, 2010-19
- Net_International_Migration_2000_2010: Net international migration, 2000-10

- Immigration_Rate_2000_2010: Net international migration rate, 2000-10
- NetMigrationRate0010: Net migration rate, 2000-10
- NetMigrationRate1019: Net migration rate, 2010-19
- NetMigrationNum0010: Net migration, 2000-10
- NetMigration1019: Net Migration, 2010-19
- NaturalChangeRate1019: Natural population change rate, 2010-19
- NaturalChangeRate0010: Natural population change rate, 2000-10
- NaturalChangeNum0010: Natural change, 2000-10
- NaturalChange1019: Natural population change, 2010-19
- TotalPop2010: Population size 4/1/2010 Census
- TotalPopEst2010: Population size 7/1/2010
- TotalPopEst2011: Population size 7/1/2011
- TotalPopEst2012: Population size 7/1/2012
- TotalPopEst2013: Population size 7/1/2013
- TotalPopEst2014: Population size 7/1/2014
- TotalPopEst2015: Population size 7/1/2015
- TotalPopEst2016: Population size 7/1/2016
- TotalPopEst2017: Population size 7/1/2017
- TotalPopEst2018: Population size 7/1/2018
- TotalPopEst2019: Population size 7/1/2019
- TotalPopACS: Total population, 2014-18 - 5-year average
- TotalPopEstBase2010: County Population estimate base 4/1/2010
- NonHispanicAsianPopChangeRate0010: Population change rate Non-Hispanic Asian, 2000-10
- PopChangeRate1819: Population change rate, 2018-19
- PopChangeRate1019: Population change rate, 2010-19
- PopChangeRate0010: Population change rate, 2000-10

- NonHispanicNativeAmericanPopChangeRate0010: Population change rate Non-Hispanic Native American, 2000-10
- HispanicPopChangeRate0010: Population change rate Hispanic, 2000-10
- MultipleRacePopChangeRate0010: Population change rate multiple race, 2000-10
- NonHispanicWhitePopChangeRate0010: Population change rate Non-Hispanic White, 2000-10
- NonHispanicBlackPopChangeRate0010: Population change rate Non-Hispanic African American, 2000-10
- MultipleRacePct2010: Percent multiple race, 2010
- WhiteNonHispanicPct2010: Percent Non-Hispanic White, 2010
- NativeAmericanNonHispanicPct2010: Percent Non-Hispanic Native American, 2010
- BlackNonHispanicPct2010: Percent Non-Hispanic African American, 2010
- AsianNonHispanicPct2010: Percent Non-Hispanic Asian, 2010
- HispanicPct2010: Percent Hispanic, 2010
- MultipleRaceNum2010: Population size multiple race, 2010
- WhiteNonHispanicNum2010: Population size Non-Hispanic White, 2010
- BlackNonHispanicNum2010: Population size Non-Hispanic African American, 2010
- NativeAmericanNonHispanicNum2010: Population size Non-Hispanic Native American, 2010
- AsianNonHispanicNum2010: Population size Non-Hispanic Asian, 2010
- HispanicNum2010: Population size Hispanic, 2010

County classifications: Type of county (rural or urban on a rural-urban continuum scale)

- Type_2015_Recreation_NO: Recreation counties, 2015 edition
- Type_2015_Farming_NO: Farming-dependent counties, 2015 edition
- Type_2015_Mining_NO: Mining-dependent counties, 2015 edition

- Type_2015_Government_NO: Federal/State government-dependent counties, 2015 edition
- Type_2015_Update: County typology economic types, 2015 edition
- Type_2015_Manufacturing_NO: Manufacturing-dependent counties, 2015 edition
- Type_2015_Nonspecialized_NO: Nonspecialized counties, 2015 edition
- RecreationDependent2000: Nonmetro recreation-dependent, 1997-00
- ManufacturingDependent2000: Manufacturing-dependent, 1998-00
- FarmDependent2003: Farm-dependent, 1998-00
- EconomicDependence2000: Economic dependence, 1998-00
- RuralUrbanContinuumCode2003: Rural-urban continuum code, 2003
- UrbanInfluenceCode2003: Urban influence code, 2003
- RuralUrbanContinuumCode2013: Rural-urban continuum code, 2013
- UrbanInfluenceCode2013: Urban influence code, 2013
- Noncore2013: Nonmetro noncore, outside Micropolitan and Metropolitan, 2013
- Micropolitan2013: Micropolitan, 2013
- Nonmetro2013: Nonmetro, 2013
- Metro2013: Metro, 2013
- Metro_Adjacent2013: Nonmetro, adjacent to metro area, 2013
- Noncore2003: Nonmetro noncore, outside Micropolitan and Metropolitan, 2003
- Micropolitan2003: Micropolitan, 2003
- Metro2003: Metro, 2003
- Nonmetro2003: Nonmetro, 2003
- NonmetroNotAdj2003: Nonmetro, nonadjacent to metro area, 2003
- NonmetroAdj2003: Nonmetro, adjacent to metro area, 2003

- Oil_Gas_Change: Change in the value of onshore oil and natural gas production, 2000-11
- Gas_Change: Change in the value of onshore natural gas production, 2000-11
- Oil_Change: Change in the value of onshore oil production, 2000-11
- Hipov: High poverty counties, 2014-18
- Perpov_1980_0711: Persistent poverty counties, 2015 edition
- PersistentChildPoverty_1980_2011: Persistent child poverty counties, 2015 edition
- PersistentChildPoverty2004: Persistent child poverty counties, 2004
- PersistentPoverty2000: Persistent poverty counties, 2004
- Low_Education_2015_update: Low education counties, 2015 edition
- LowEducation2000: Low education, 2000
- HiCreativeClass2000: Creative class, 2000
- HiAmenity: High natural amenities
- RetirementDestination2000: Retirement destination, 1990-00
- Low_Employment_2015_update: Low employment counties, 2015 edition
- Population_loss_2015_update: Population loss counties, 2015 edition
- Retirement_Destination_2015_Update: Retirement destination counties, 2015 edition