AN INVESTIGATION INTO THE CORRELATION BETWEEN COVID19-RELATED DEATHS IN 2020 AND THE LEADING CAUSES OF DEATH IN THE UNITED STATES FROM 2015-2019

CREDITS TO:

Maryam AliakBari

Farzin Valiloo

ABOUT THE SUBJECT

Let’s start with the subject …

The planning and development of health policies should be based on:

- Precise measurement and understanding of correct indices
- Past and current profiles
- Prediction of change over time

-- (Defo, 2014)

REVIEW

The **top 10** causes of death accounted for **55%** deaths worldwide in **2019**

(World Health Organization, 2020).

In early **2020** Covid caught the world off-guard

(Centre for Disaster Protection, Clarke, 2022).

Underlying death causes can contribute to the mortality rate of Covid-19

(USA Center for Disease Control and Prevention)

PROBLEM STATEMENT

“Is there a difference in mortality rate of Covid-19 among the states with different highest causes of death in prior years to 2020?”

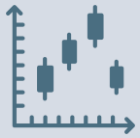
RESEARCH MAIN OBJECTIVES

1. Exploring the select causes of death between 2014-2019 in the United States
2. Understanding the differences in patterns and distributions
3. Comparing the real data with the findings (mainly the effect of Covid-19)
4. Proposing actions influencing the discovered trends and patterns

ABOUT THE METHODOLOGY

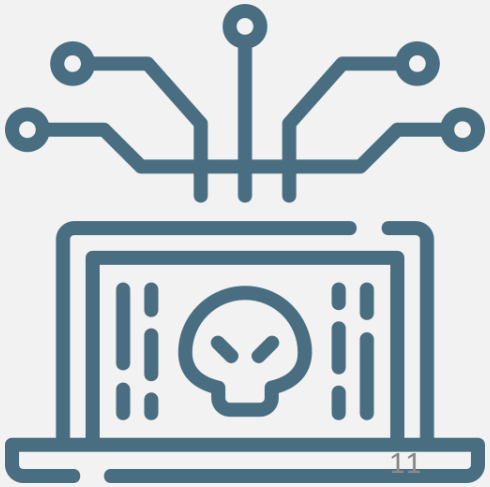
Let’s see the methods utilized …

LET'S REVIEW THE COURSE OF ACTION WE'VE FOLLOWED

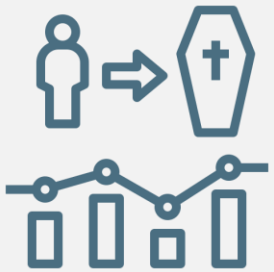
STEPS	OBJECTIVES	TOOLS
Data Exploration	▪ Finding inconsistencies (missing data, excessive columns, wrongly formatted data , etc.)	▪ MS Excel ▪ Python ▪ Tableau
Data Cleaning	▪ Addressing the potential issues	▪ Python
Descriptive Analysis	▪ Getting the clear understanding of the data ▪ Determining required transitions and complimentary data	▪ Python ▪ Tableau
Data Manipulation	▪ Merging and joining all data	Python's Libraries ▪ Pandas ▪ NumPy
 **Analysis**	▪ Statistical hypotheses testing ▪ Machine Learning	Python's Libraries ▪ SciPy ▪ PyCaret
Conclusion	▪ Discussing the results ▪ Implications & Limitations ▪ Future recommendations	

ABOUT THE DATA

Let’s talk about the data …



DATASETS



US WEEKLY DEATHS

Weekly No. of 13 top death causes in all US states from 2014 - 2019

- 16,903 Rows
- 2 Categorical Features
- 17 Numeric Features

(Source: U.S. government, data.gov)

COVID 19 DEATHS

Weekly No. of 13 death causes in all US states from 2020 - 2022

- 3623 Rows
- 2 Categorical Features
- 19 Numeric Features

(Source: U.S. government, data.gov)

US POPULATION 2010-2019

Estimation of population in each US state for 2010-2019 No. rows

- 51 Rows
- 1 Categorical Features
- 19 Numeric Features

(Source: U.S. Census Bureau, census.gov)

US POPULATION 2020

Estimation of population in each US state for 2020

- 52 Rows
- 1 Categorical Features
- 3 Numeric Features

(Source: populationu.com)

DATA EXPLORATION & CLEANING

MAIN DATASETS ISSUES

- Numbers stored as texts
- Comma sign inside numerical values
- Different state names in different datasets
- Missing values
- Unnormalized form of the raw data (based on population of the states)

ABOUT THE ANALYSIS

Let’s start the analysis …

DATA MANIPULATION

ISSUES WITH THE MAIN DATA SET:

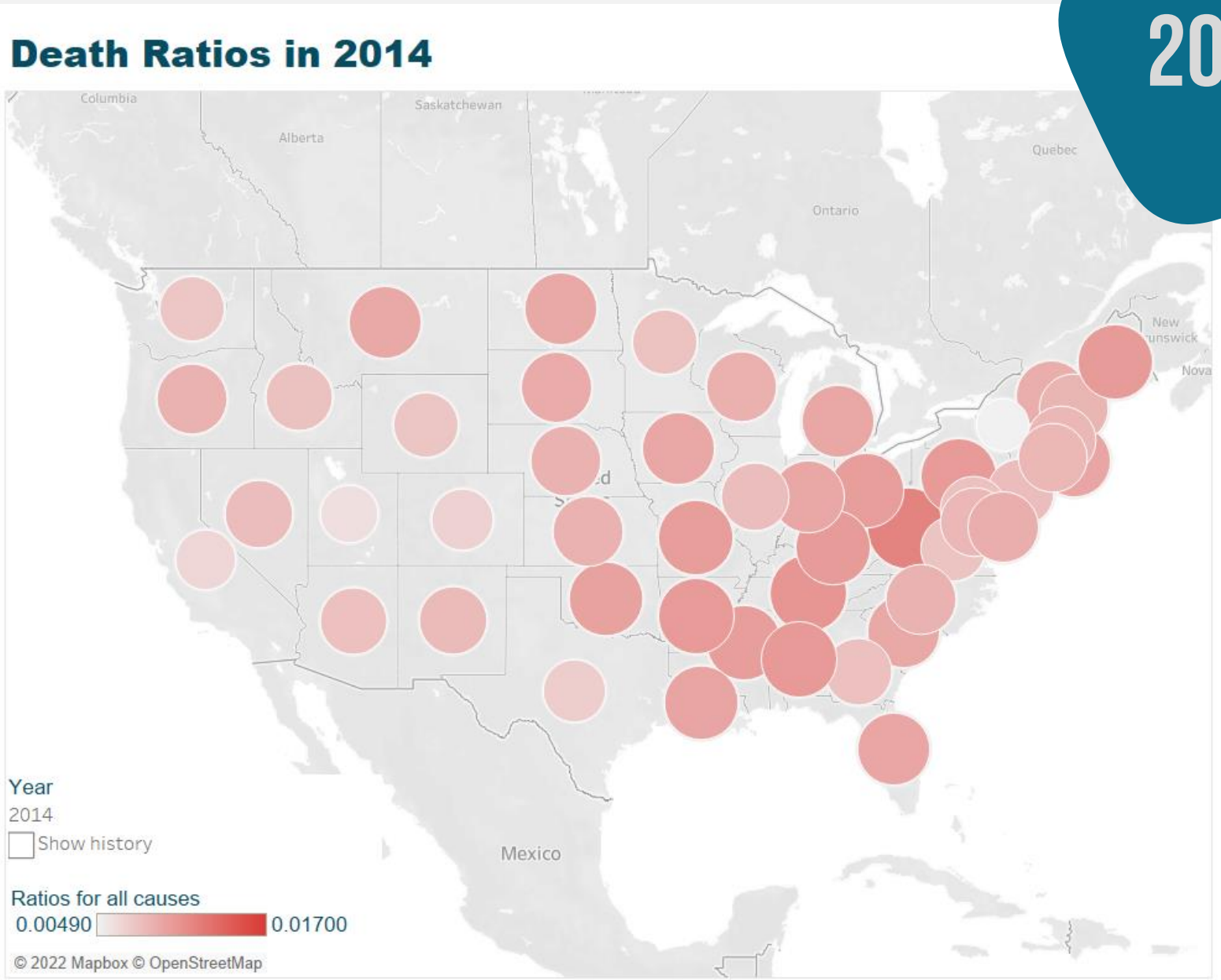
- Not comparable (normalized)
- On a weekly basis

SOLUTIONS

- Add population data (yearly)
- Aggregate data (yearly)
- Aggregate the data for the years 2014-2019
- Ranking death causes in each state based on the normalized data
- Extracting top (n) causes in each state

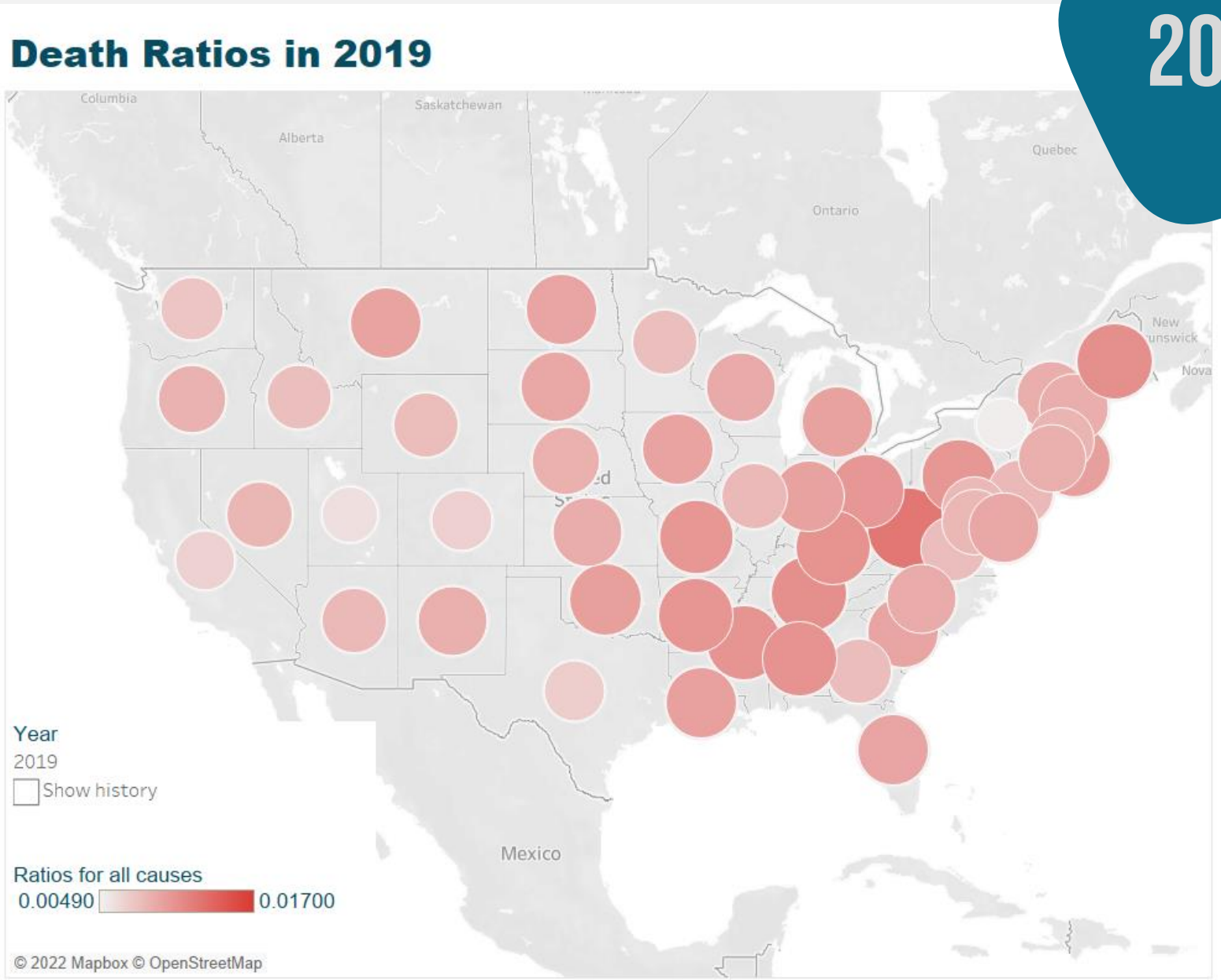
DESCRIPTIVE ANALYSIS

2014



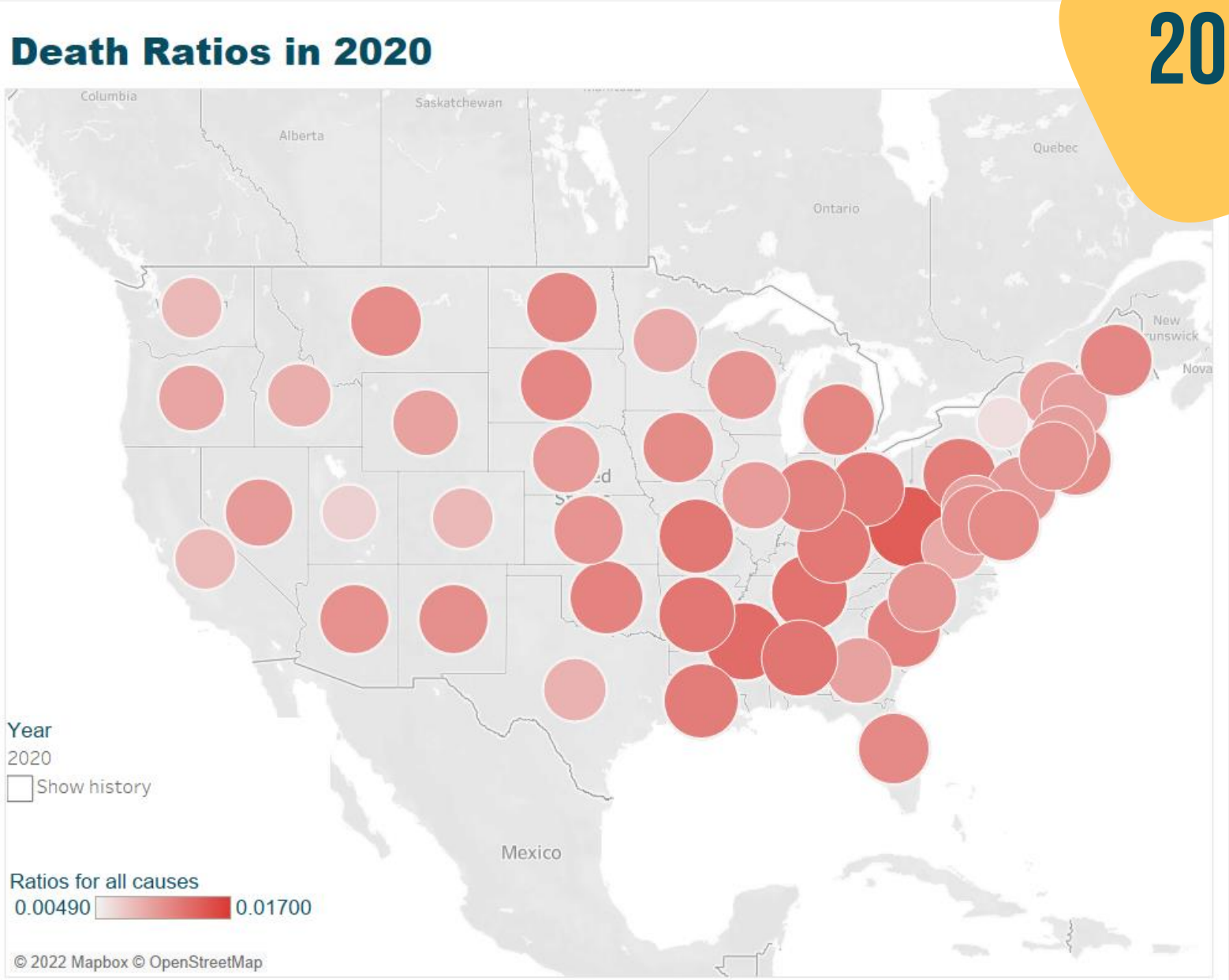
DESCRIPTIVE ANALYSIS

2019

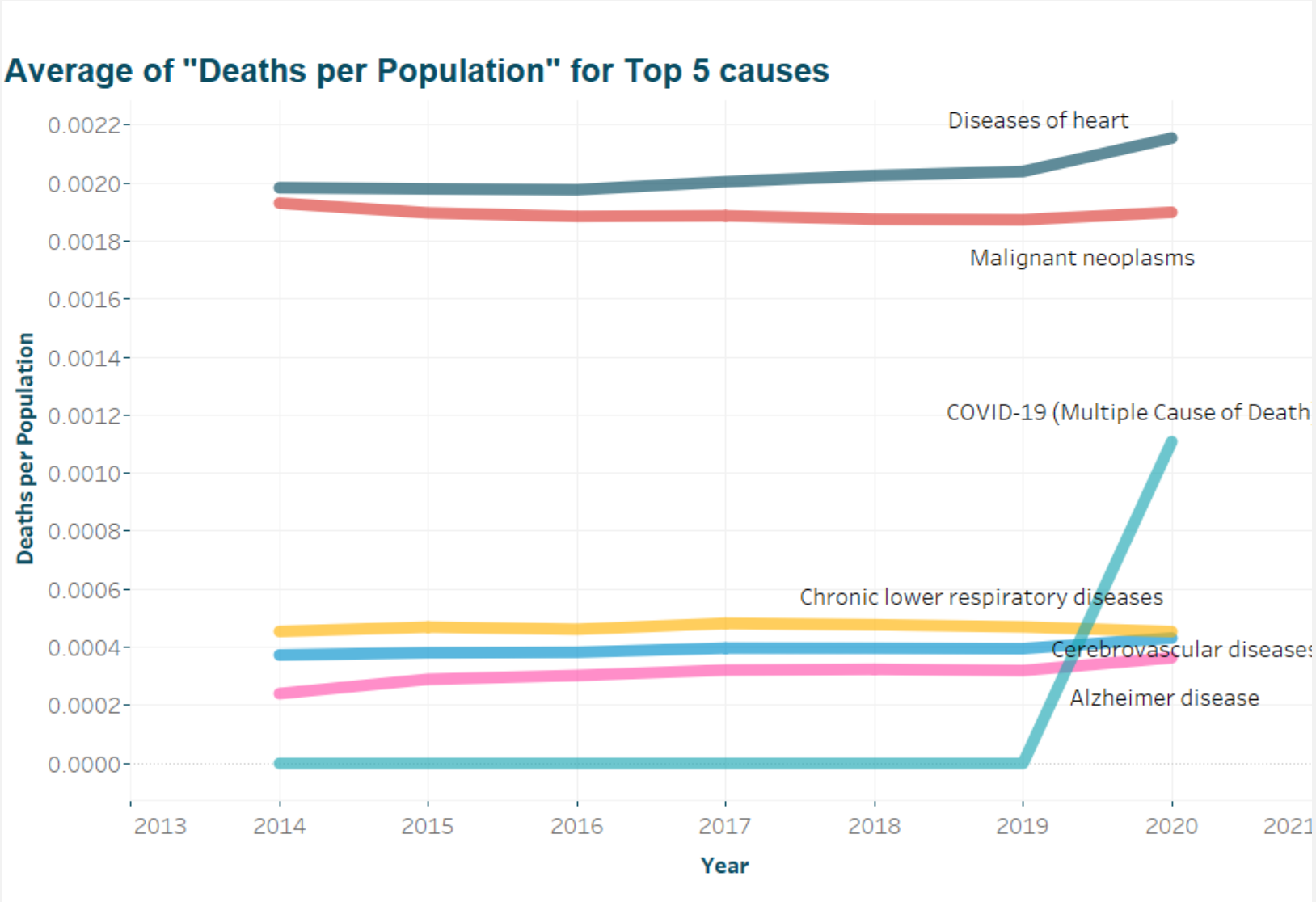


DESCRIPTIVE ANALYSIS

2020



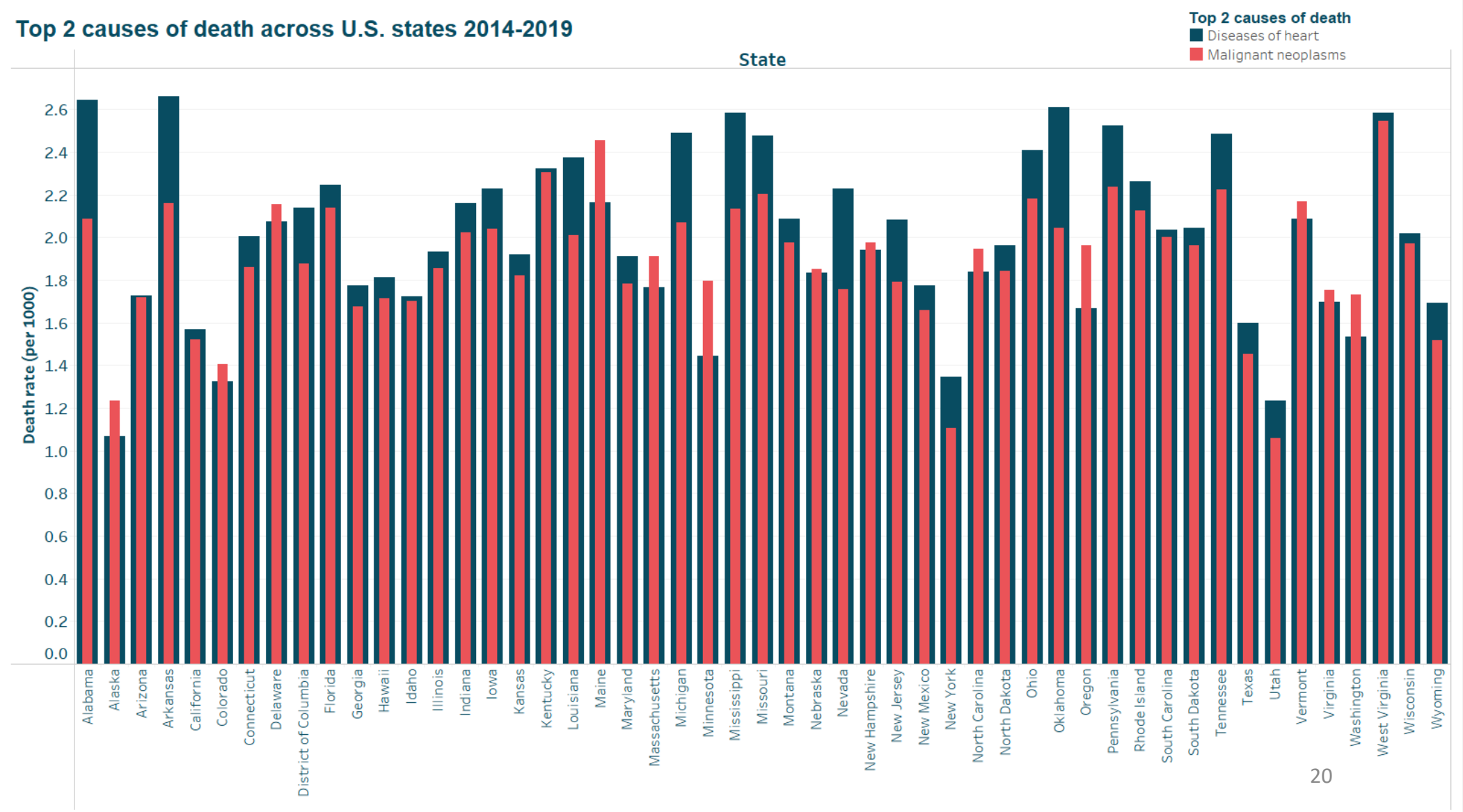
DESCRIPTIVE ANALYSIS CONT'D



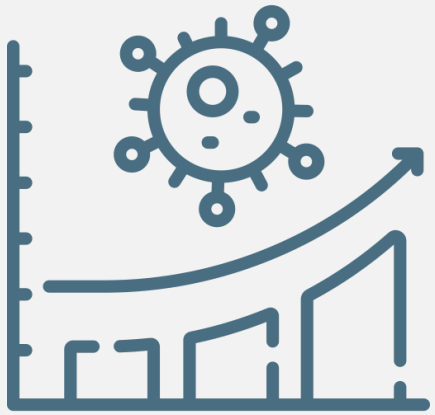
The trend of top 5 death causes and Covid-19 between 2014 and 2020

DESCRIPTIVE ANALYSIS CONT'D

See how the top 2 causes of death distribute among all the states



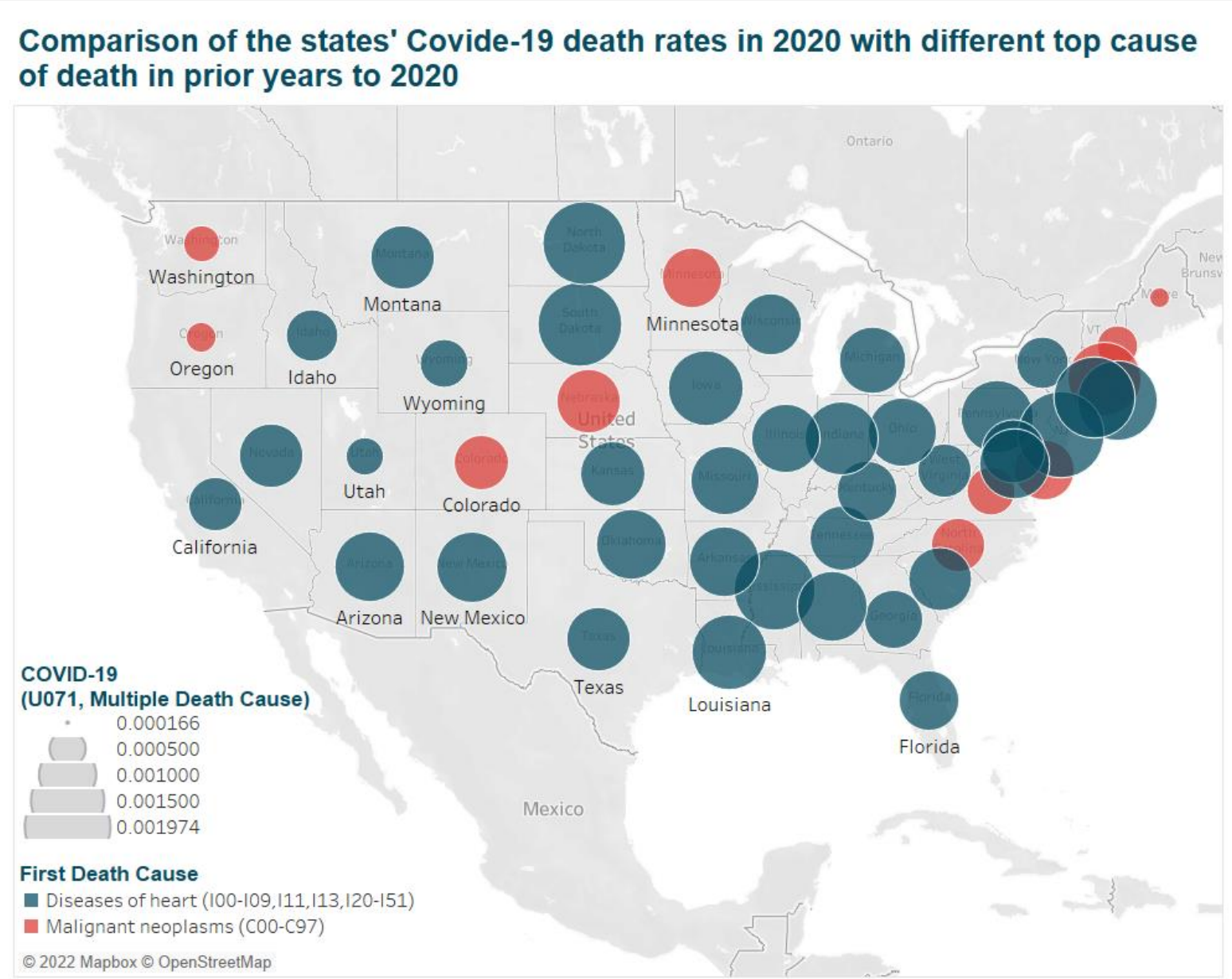
DATA EXPLORATION & CLEANING



INSIGHTS FROM DESCRIPTIVE ANALYSIS

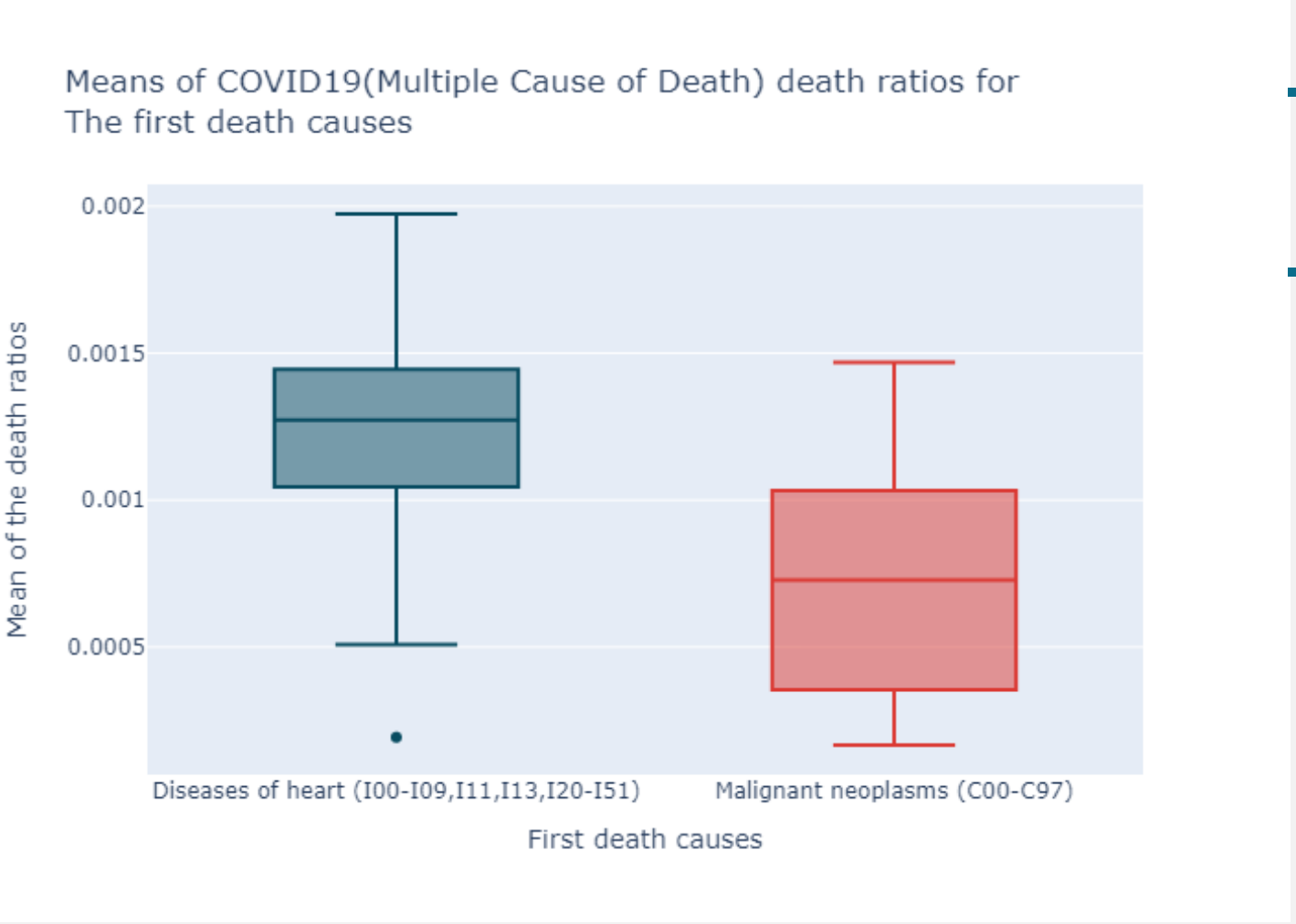
- Two first cause of deaths in every state:
  - Disease of heart
  - Malignant neoplasms
- Two groups of states based on their top cause of death

STATISTICAL ANALYSIS



See the death rate of Covid-19 in different groups of states

STATISTICAL ANALYSIS



GROUP A

States which their TOP death cause is **"Diseases of heart"**,

GROUP B

States whose top death cause is **"Malignant neoplasms"**

H0

There is no difference in the means of death rates caused by COVID19 (Multiple Cause of Death)

H1

The mean of death rates caused by COVID19 (Multiple Cause of Death) in is higher in states group A than that of group B

P-value <0.01 So we can reject the null value in alpha = 99 %

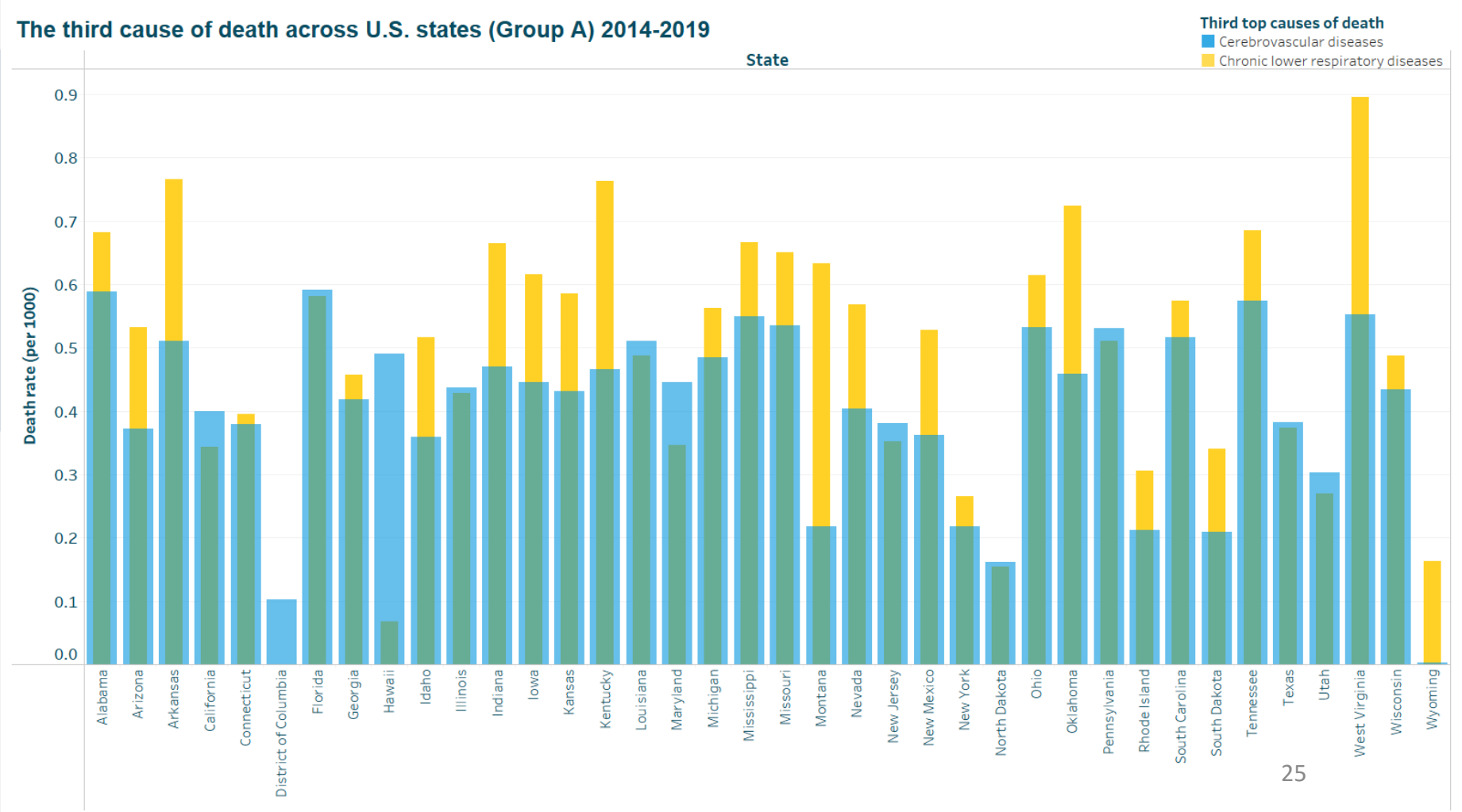
MORE ANALYSIS

about the Second toHowp cause?

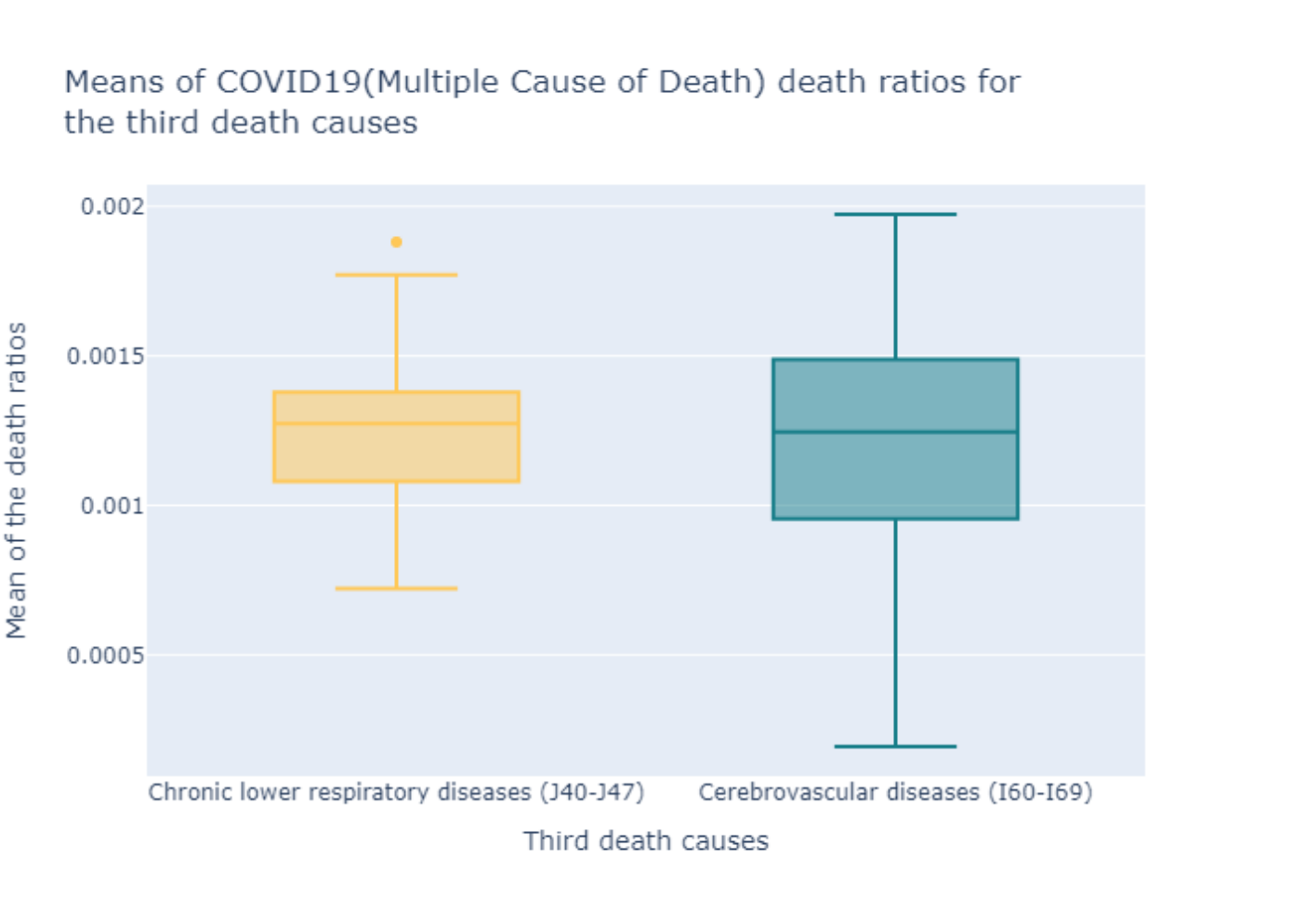
No need! As they are the same as top first.

STATISTICAL ANALYSIS

How about the Third top cause?



STATISTICAL ANALYSIS



GROUP AC

States which their Third death cause is **"Chronic lower respiratory diseases (J40-J47)"**

GROUP AD

States which their Third death cause is **"Cerebrovascular diseases (I60-I69)"**

H0 *There is no difference in the means of death rates caused by COVID19 (Multiple Cause of Death)*

H1 *The mean of death rates caused by COVID19(Multiple Cause of Death) in is not equal in states group AC than and group AD*

P-value > 0.01 so we cannot reject the null hypotheses in alpha = 99 %

MORE ANALYSIS

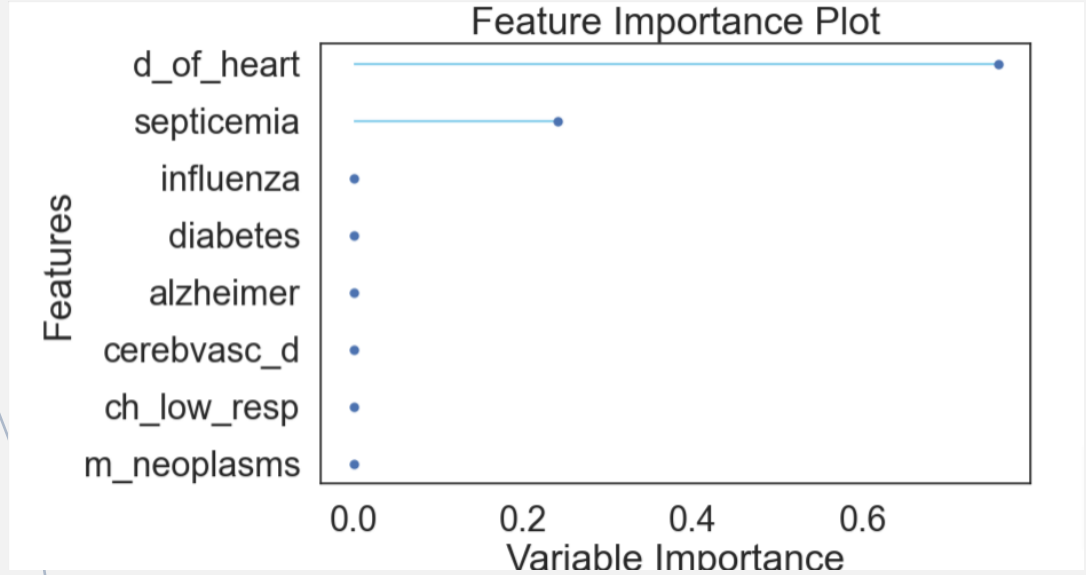
How about other causes of death?

- Machine learning to fit a model

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
dt	Decision Tree Regressor	0.2616	0.1175	0.3161	-0.1487	0.1632	0.3816	0.0110
rf	Random Forest Regressor	0.2620	0.1240	0.3265	-0.1628	0.1620	0.4016	0.2690
knn	K Neighbors Regressor	0.3130	0.1638	0.3622	-0.2866	0.1835	0.5006	0.0390
omp	Orthogonal Matching Pursuit	0.2906	0.1570	0.3444	-0.3080	0.1714	0.4039	0.0120
br	Bayesian Ridge	0.2896	0.1681	0.3554	-0.4799	0.1807	0.4784	0.0090
et	Extra Trees Regressor	0.2834	0.1394	0.3521	-0.5277	0.1742	0.4068	0.2300
lasso	Lasso Regression	0.3299	0.1793	0.3924	-0.5417	0.1969	0.5230	0.0120
lightgbm	Light Gradient Boosting Machine	0.3299	0.1793	0.3924	-0.5417	0.1969	0.5230	0.0140
llar	Lasso Least Angle Regression	0.3299	0.1793	0.3924	-0.5417	0.1969	0.5230	0.0090
en	Elastic Net	0.3299	0.1793	0.3924	-0.5417	0.1969	0.5230	0.0110
dummy	Dummy Regressor	0.3299	0.1793	0.3924	-0.5417	0.1969	0.5230	0.0080
ada	AdaBoost Regressor	0.2664	0.1302	0.3378	-0.5525	0.1661	0.3940	0.0440
par	Passive Aggressive Regressor	0.2967	0.1500	0.3499	-0.5843	0.1860	0.3854	0.0120
gbr	Gradient Boosting Regressor	0.2711	0.1219	0.3361	-0.5981	0.1677	0.4054	0.0270
huber	Huber Regressor	0.2952	0.1788	0.3721	-0.6166	0.1870	0.4960	0.0130
ridge	Ridge Regression	0.3029	0.1761	0.3715	-0.8075	0.1872	0.4760	0.0100
lr	Linear Regression	0.3174	0.1930	0.3903	-1.0924	0.1955	0.4953	0.6570
lar	Least Angle Regression	0.3174	0.1930	0.3903	-1.0924	0.1955	0.4953	0.0100

DECISION TREE TUNED

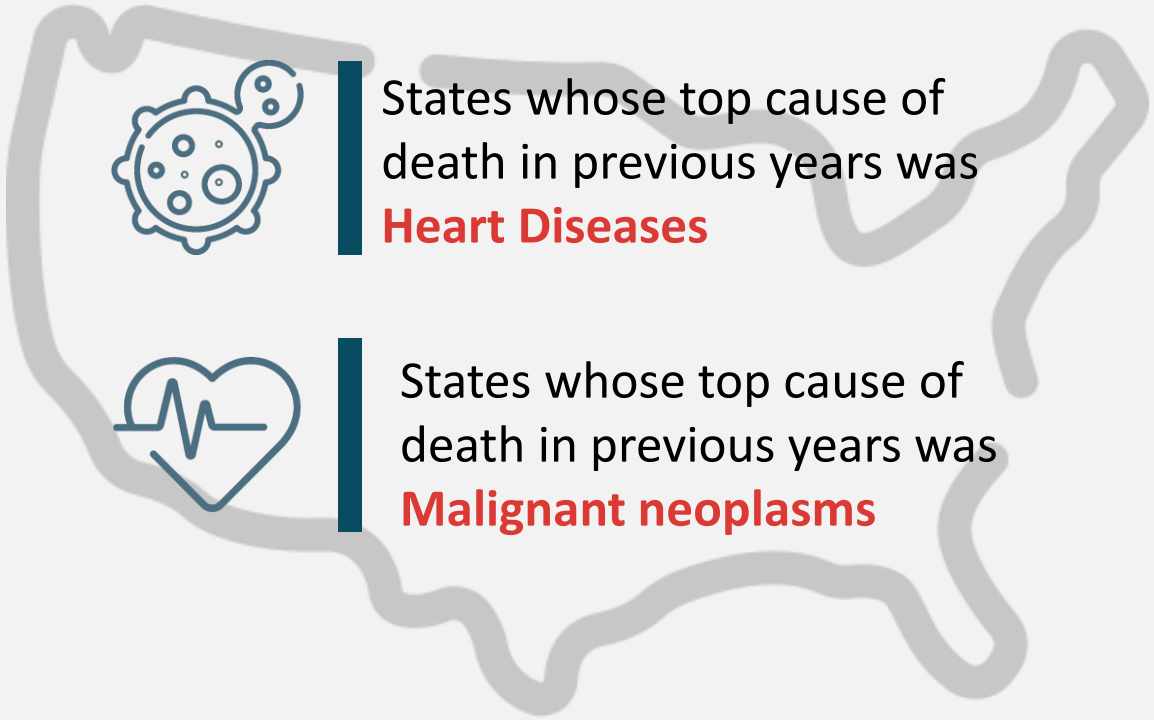
Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.1743	0.0337	0.1837	0.2044	0.0997	0.2333
1	0.3395	0.2485	0.4985	-0.1689	0.3019	1.3330
2	0.2288	0.1018	0.3191	-0.4481	0.1483	0.1694
3	0.3291	0.1793	0.4234	0.0795	0.2368	0.6312
4	0.2807	0.1285	0.3585	0.3922	0.1647	0.2905
5	0.1025	0.0144	0.1200	0.3132	0.0531	0.0783
6	0.4009	0.2337	0.4834	-0.0312	0.2034	0.2873
7	0.1665	0.0393	0.1983	-0.1115	0.0961	0.1669
8	0.4618	0.2480	0.4980	0.5015	0.2603	0.9306
9	0.1952	0.0445	0.2109	-0.8227	0.1074	0.1782
Mean	0.2679	0.1272	0.3294	-0.0092	0.1672	0.4299
Std	0.1084	0.0895	0.1367	0.3834	0.0771	0.3885



ABOUT THE RESULTS & CONCLUSION

So, what does this analysis implicate?

There is a significant difference between Covid caused death rates of these two groups of states:



States whose top cause of death in previous years was **Heart Diseases**

States whose top cause of death in previous years was **Malignant neoplasms**

Although ML algorithms did not produce perfect results, but **Feature importance analysis** was consistent with findings in the statistical analysis

IMPLICATIONS

Covid is far from over:

To warn states in group A

Among group A states, few states that had lower death:

Their systems and actions can be studied

LIMITATIONS

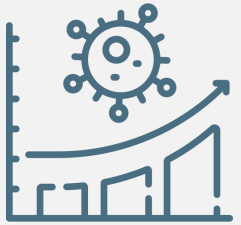
- Final Transformed Dataset is too small for a Machine-learning model
- Nature of this study needed many steps of data transformation and aggregation
- The purpose was to get a grasp of the feature importance
  - Problem itself
  - Limited available data

RECOMMENDATIONS

To whom it may concern…

RECOMMENDATIONS FOR FUTURE WORKS

Adding more detailed complimentary datasets (enriching)



Tracking Covid behavior and comparing the results in upcoming years

Comparing with other regions using the same methods

RESOURCES

Our study couldn’t have been done without …

TOOLS

Tableau 2022.3

SciPy 1.7.3

MS Office 365

Python 3.7.11

Pandas 1.3.5

NumPy 1.21.5

PyCaret 2.3.10

Matplotlib 3.5.3

REFERENCES

1) *Center for Disease Control and Prevention (CDC) . (n.d.). Excess Deaths Associated with COVID-19. Retrieved from cdc.gov: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm#dashboard*
2) *Clarke, D. (2022, April 9). Covid-19 caught the world off guard - pandemics must never surprise us again. Retrieved from disasterprotection.org: https://www.disasterprotection.org/blogs/covid-19-caught-the-world-off-guard-pandemics-must-never-surprise-us-again*
3) *DEFO, B. K. (2014). Beyond the 'transition' frameworks: the cross-continuum of health, disease and mortality framework. Retrieved from Global health action: https://doi.org/10.3402/gha.v7.24804*
4) *Harvard University, School of Public Health. (n.d.). The latest on the coronavirus. Retrieved from hsph.harvard.edu: https://www.hsph.harvard.edu/news/hsph-in-the-news/the-latest-on-the-coronavirus/*
5) *populationu.com. (n.d.). US States by Population. Retrieved from populationu.com: https://www.populationu.com/gen/us-states-by-population*
6) *U.S. Census Bureau. (n.d.). State Population Totals and Components of Change: 2010-2019. Retrieved from census.gov: https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html*
7) *U.S. Government. (n.d.). The home of the U.S. Government's open data. Retrieved from Data.gov: https://catalog.data.gov/dataset/weekly-counts-of-deaths-by-state-and-select-causes-2014-2018*
8) *World Health Organization. (2020, December 09). The top 10 causes of death. Retrieved from www.who.int: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death*
9) *For picture & photos: https://www.flaticon.com/*

THANKS
&, … Stay Safe!