A7: Project Report
DATA 512
Laura Thriftwood
December 14, 2021

Introduction

The COVID-19 pandemic disrupted human behavior at a magnitude previously unseen in modern history. While much of the impact was negative, there were a number of positive impacts resulting from the large-scale change to our daily lives. We saw a massive shift to remote work, the implementation of lockdowns and stay-at-home orders, and restrictions on travel. This resulted in a decrease in the number of people commuting to work and an overall reduction in traffic volume. Ideally, drastically fewer cars on the road would result in fewer traffic accidents. However, it's possible that fewer cars on the road would increase traveling speed or other behavioral changes that could increase the severity of crashes.

These are the human-centered assumptions about the impact of the COVID-19 pandemic that I explored through the lens of my assigned county, Salt Lake County, Utah.

Background/Related Work

While there was consistent agreement about the overall reduction in traffic volume at the start of the pandemic, research and news articles provided conflicting information on the effect of the reduction in traffic on both crash frequency and outcome. Many articles such as the ones titled "Research finds increase in car crashes with decrease in traffic during pandemic" (Bachman, 2020) and "Lower Volumes, Higher Speeds: Changes to Crash Type, Timing, and Severity on Urban Roads from COVID-19 Stay-at-Home Policies" (Stiles et al., 2021) suggest that there was an increase in both the number of traffic crashes and the number of fatalities. They suggest that the increase in severity of crashes is a result of increases in speeding, reckless driving, substance abuse, and/or other reckless behaviors like failure to wear a seatbelt.

In a paper titled "Global impact of COVID-19 pandemic on road traffic collisions" the authors conducted an analysis of the impact of reduced road traffic from COVID-19 lockdowns and stay-at-home orders on road traffic collisions around the world. They reviewed the incidence, patterns, injury severity, management, and outcomes of road traffic collisions (Yasin et al., 2021). They concluded that the pandemic was associated with a significant reduction in road traffic collisions globally, with 32 of the 36 countries included in the study seeing a reduction in road deaths as well.

My review of these conflicting accounts of the impact of the pandemic on crash behavior influenced my decision to examine traffic crash frequency and traffic crash severity separately.

A7: Project Report
DATA 512
Laura Thriftwood
December 14, 2021

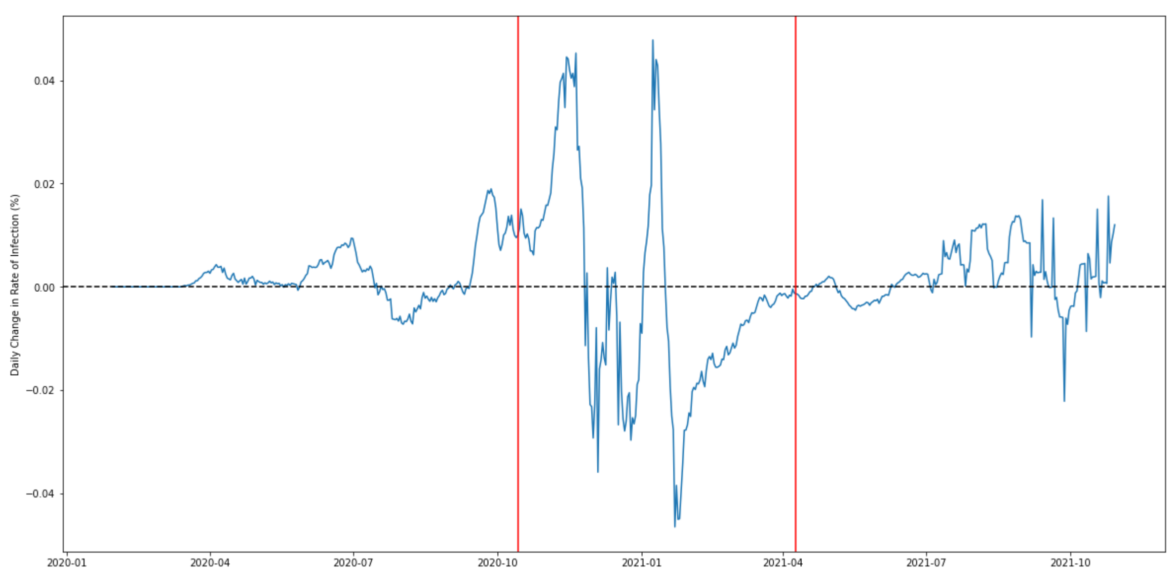
Methodology

Many confounding factors influence changes to human behavior. Something as simple as "driving behavior" could be influenced by a number of possible factors - traffic volume, stress, mental health, changes in the use of public transportation - many of which do not come with available datasets to include in such an analysis. For this reason, I wanted to keep the focus of my methods targeted and simple, rather than try to account for all possible influencing factors. I relied on a combination of time series analysis, data visualizations, and hypothesis testing. My analysis was done in 4 phases as outlined below.

Phase 1

The analysis in Phase 1 reflects the Common Analysis in A4 and explored the relationship between the implementation and removal of county mask mandates on the change in rate of infection in Salt Lake County from 1/22/2020 to 10/29/2021, the period for which we have available case counts from the John Hopkins dataset. Daily infection rate represents the proportion of the population at risk who are considered to be actively infected, assuming a 14 day infection period. "Population at risk" estimates the number of people in the county who are not currently in active infection, including those who were previously infected. In order to smooth the volatility and reduce noise, I first used a rolling 7 day average for the daily counts of confirmed cases and used these figures to calculate the infection rate. Salt Lake County implemented a mask mandate on 10/14/2020 which remained in effect until 4/9/2021. The visualization below captures the outcome of that analysis.

Change in Rate of Infection, Salt Lake County, Utah



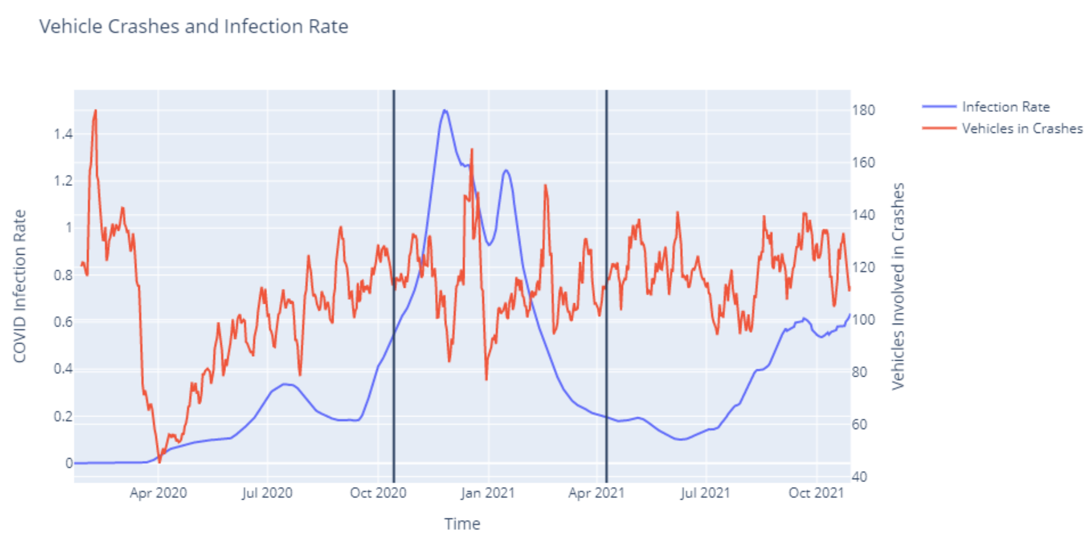
Red lines denote the period of time with a mask mandate

A7: Project Report
DATA 512
Laura Thriftwood
December 14, 2021

Phase 2

In this phase my goal was to examine whether the mask mandate behavior correlated with vehicle crash behavior. The dataset I used comes from the Utah Department of Public Safety, which provides a Raw Crash Data in a .csv file that is updated daily. As stated on their website, “[t]he data for the Utah Crash Summary is derived from Utah crash reports. These reports are completed by law enforcement officers throughout the state who investigate crash scenes on public roadways. Information is collected when a crash involves injuries, deaths, or at least $2,500 property damage. Crash reports are uploaded daily to the Utah Department of Public Safety data warehouse for central collection. Additional information is collected on fatal crashes at the Department of Public Safety’s Highway Safety Office and compiled into the Fatality Analysis Reporting System (FARS) database. FARS is a national data system containing data on all fatal traffic crashes in the U.S.” (Utah Department of Public Safety, 2021)

I started with a plot of the daily infection rate (rather than the change in infection rate) from the previous analysis and overlaid a plot of the number of vehicles involved in crashes in the same time period. As a single crash incident could involve more than one vehicle, I used the number of vehicles involved in crashes as I felt it better captured the human impact of each crash incident. The visualization is provided below.



By using dual y axes, I was able to intuit visually that the crash behavior did not correlate with changes in mask mandate or infection rates. However, I sought statistical confirmation of this assumption and ran a One-Way ANOVA F-Test to compare the group means of the three groups - Pre-mandate, Mandate, and Post-Mandate.

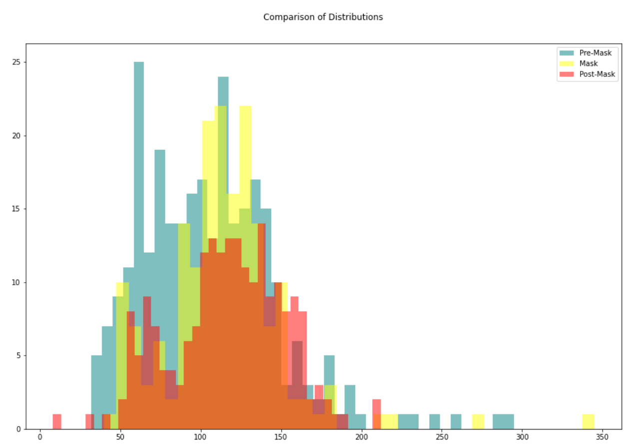
Assumptions

Independent samples

I used random sampling to get a sample from each population using df.sample in Python.

Each sample is from a normally distributed population

A comparison of the three population distributions is below showing roughly normal distributions (if slightly right skewed).



The population standard deviations of the groups are all equal

	pre_mask	mask	post_mask
sample mean	103.8	109.96	118.26
sample st_dev	43.4	41.1	38.1
sample n	50	50	50

After confirming that the assumptions of the test were met, I defined the null hypothesis as:

$$\mathbf{H_0 : \mu_A = \mu_B = \mu_C}$$

The alternative hypothesis is therefore defined as:

H_a : the three populations means are not all equal

where μ_A, μ_B, and μ_C are the mean number of vehicles involved in crash incidents in the county for the three time periods. I used a sample size of 50 and a significance level of 0.05. The results of the test are below.

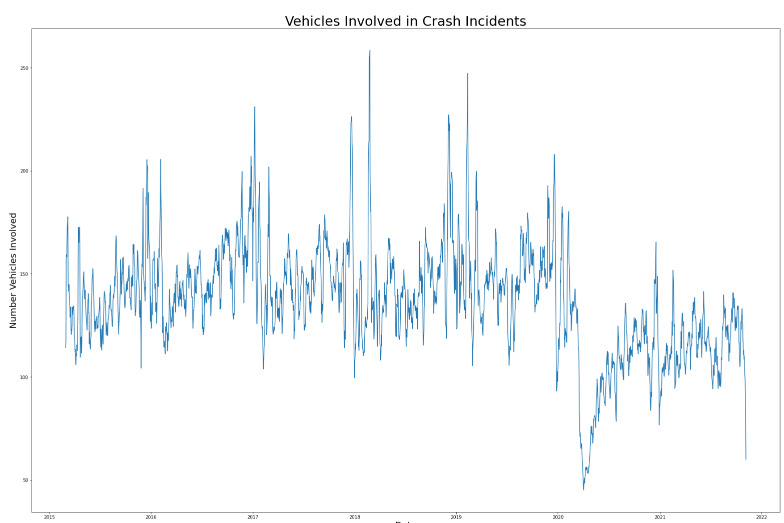
F-Value: 6.0128765
The p-value at a significance level of 0.05 is: 0.00258
We reject the null hypothesis that the three population means are equal.

The results suggest that at least one of the population means is not equal to the other two. This is not surprising, but it would be a mistake to conclude from this that this relates to the implementation and removal of mask mandates in the county. It was clear from the earlier visualization that there was a large drop in crashes in the pre-mandate group, with fairly steady crash behavior in the mandate and post-mandate groups. This hypothesis test does not capture that important behavior change. As such, I opted to ignore the influence of mask mandates entirely, and instead focus on examining the changes in crash behavior as it related to the pandemic more generally.

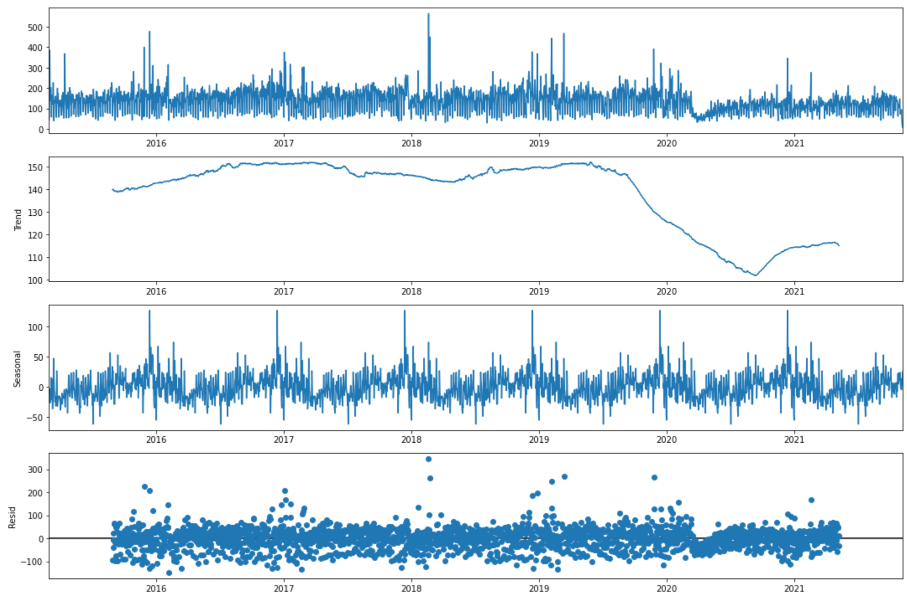
Phase 3

In this phase, I instead examined the change in crash behavior before and after the start of the pandemic. It should hold that an event as impactful as the pandemic would result in a statistically significant change in everyday human behavior, in this case represented by vehicle crashes.

I used 3/1/2020 as the date that delineated pre- and post-COVID populations as the first recorded case in Salt Lake County was on 3/13/2020 to which I add a roughly 14-day incubation period (Centers for Disease Control and Prevention, 2021). I also chose to restrict the pre-COVID population to the 5 years prior to the pandemic to ensure the comparison captures changes to recent behavior. Visually, the delineation of the two testing populations is clear in the below plot of the number of vehicles involved in crash incidents from 3/1/2015 to 11/6/2021.



It is clear from this plot that the time series of crash data displays some seasonality that might hide trends. I used a fairly naive method of decomposing the elements of the data with the *seasonal_decompose* function of *statsmodels*. The result of this decomposition is visualized below.



By removing the seasonality and analyzing the individual components of the time series, we can clearly see the trend in the second figure indicating a significant decrease in crash incidents in the same period of time as the pandemic. This supports the structure of the analysis as laid out earlier.

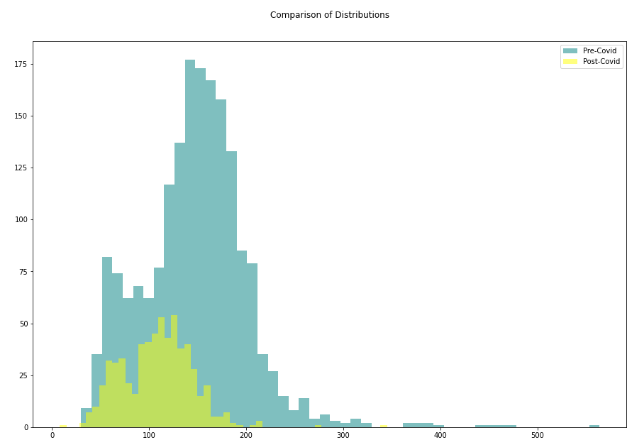
For this hypothesis test I used a Two-Sample Z-Test to compare population means. I first confirmed that the following assumptions were met:

Independent Samples

I used random sampling to get a sample from each population using df.sample in Python.

Populations are normally distributed

A comparison of the three population distributions is below showing roughly normal distributions (if slightly right skewed).



Population standard deviations are known

These values are known and provided below.

	pre-covid	post-covid
st_dev	54.52	36.41

The null hypothesis is defined as:

$$\mathbf{H_0 : \mu_A = \mu_B}$$

The alternative hypothesis is therefore defined as:

$$\mathbf{H_a : \mu_A \neq \mu_B}$$

where μ_A, and μ_B are the mean number of vehicles involved in crash incidents in the county for time period A (5 years prior to the pandemic) and period B (since the start of the pandemic). I used a sample size of 100 from each population and looked at a significance level of 0.025 as it was a two-tailed test. The results of the test are below.

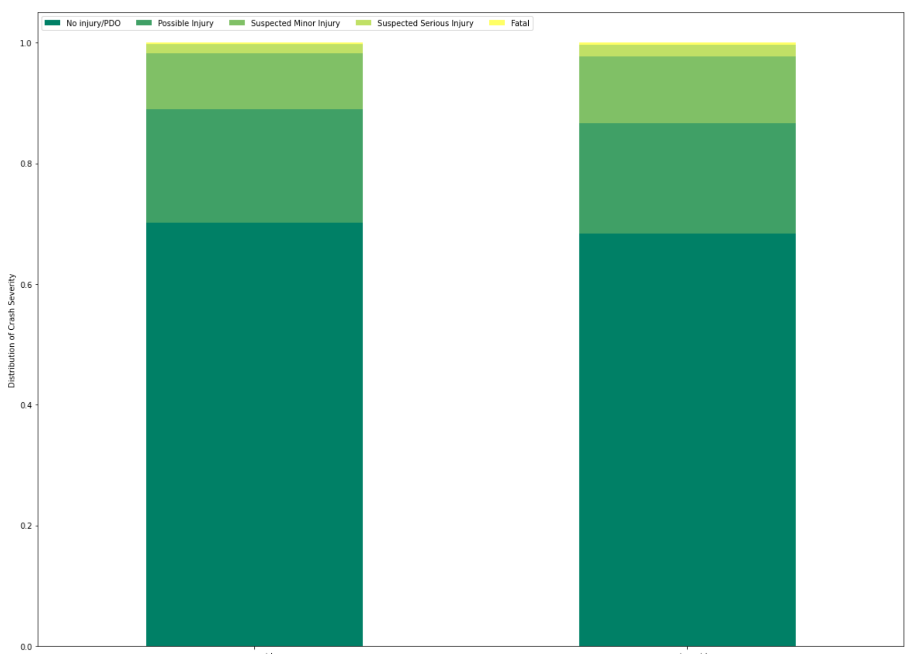
Z Statistic: 16.1722
The p-value at a significance level of 0.025 is: 7.917e-59
We reject the null hypothesis that the two population means are equal.

There is enough evidence to suggest that the mean number of vehicles involved in crashes prior to the start of the pandemic differs in a statistically significant way from the mean number of vehicles involved in crashes after the start of the pandemic.

Phase 4

Finally, I wanted to explore any potential changes in the severity of crashes. As noted, this metric may reflect changes in driving behavior, not simply driving volume. Given the vast number of potential factors that influence driving behavior (speed, mental health, substance abuse, destination, overall health) I wanted to keep my analysis simple. Using the same two populations as identified in Phase 3, I analyzed the makeup of KABCO Crash Severity metrics included in the dataset for each incident. The KABCO Crash Severity is a scale established by the Federal Highway Safety Administration (FHWA) to evaluate the severity of auto collision injuries (US Department of Transportation, 2017). As can be seen by the below table and visualization comparing the proportion of crashes in each category below, there does not appear to be a significant difference in the proportion of fatal crashes during the pandemic.

	No injury/PDO	Possible Injury	Suspected Minor Injury	Suspected Serious Injury	Fatal
pre_covid	0.701058	0.188291	0.092916	0.015282	0.002453
post_covid	0.684071	0.182238	0.110370	0.019800	0.003521



Focusing specifically on the proportion of fatal crashes, we do see an increase from 0.25% to 0.35%, with 333 crash events receiving this rating in the 5 years prior to the pandemic and 122 crash events with Fatal ratings in the 1 ½ years after the start of the pandemic.

Findings

To summarize the findings of my analysis, the overall decrease in traffic volume after the start of the pandemic resulted in a statistically significant decrease in the number of vehicles involved in crashes. However, I did not find evidence to suggest a significant difference in the severity of crashes between the two time periods. There was a slight increase in the proportion of crash incidents that received a Fatal rating from 0.25% prior to the pandemic to 0.35% after the start of the pandemic, but it was not significant enough to attribute to the change in behavior resulting from the pandemic.

Discussion/Implications

These findings are important as they provide objective analysis of two commonly held beliefs and claims about changes in vehicle crash behavior. The first claim is that there were fewer car crashes during the pandemic. Through time series analysis and hypothesis testing, I was able to provide objective evidence that supports this claim. The other claim is that crashes occurring during the pandemic were more severe and resulted in more fatalities than crashes that occurred prior to the pandemic. While I found that there was a slight increase in the proportion of crashes that likely resulted in a fatality, the overall profile of crash severity did not appear to differ significantly during the pandemic in Salt Lake County, UT.

Vehicle crash behavior directly impacts other areas of human life, such as healthcare provider utilization and infrastructure repair spending. Additionally, vehicle crash behavior is directly impacted by factors relating to human behavior, such as mental health and substance abuse. It is an area of research with many direct ties to the human condition and is therefore worthy of study.

Future research in this area could examine the combination of features in the dataset related to vehicle crashes, such as road conditions or time of day. As other literature notes, there are many possible sources of pandemic-related behavioral changes that could contribute to a change in driving behavior, such as mental health and substance abuse. A broader analysis of those contributing factors could provide interesting insight through the lens of available data. These insights have the potential to support data-driven decisions about public health policy or resource allocation.

Limitations

The primary limitation of this analysis is that it is limited to my specific assigned county of Salt Lake County, Utah. The impact of the pandemic on crash behavior in this county may not be representative of the impact in other areas of the state or country. It is the most populous county in the state and houses the state capital of Salt Lake City, so the profile of the county is much different from other Utah counties. It would be more reasonable to compare it with counties in other states with similar populations housing a metropolitan city. As mentioned previously in this analysis, there are a variety of factors that influence driving behavior on an individual and societal level. These factors when coupled with state-to-state differences in policy and protocols in response to the pandemic would minimize the value of any conclusions of such an analysis.

There were two important limitations of the crash dataset used in this analysis. First, while the dataset does include information on time of day and road conditions, many factors that influence the severity of a crash can't be captured by first responders at the scene, such as speed. It also lacks details on the number of passengers in the vehicles involved, making extrapolation challenging. The second limitation of the data is in the KABCO Crash Severity metric used in the analysis. Even though the metric is established by a federal agency and is used consistently for all crashes, it relies on determinations made on scene by the responding officer and may not be an accurate accounting of outcomes for each driver and/or passenger involved.

Finally, while the assumptions required for the One-Way ANOVA F-Test in Phase 2 were met, the assumption of equal variances within the groups is highly dependent on which points were included in the random sampling for the pre-mask group. This is the group that contains crash data that spans the onset of the pandemic and sees a large drop in crash frequency. This limitation is one of the reasons I elected to modify the hypothesis test in Phase 3 to achieve higher confidence in the results.

Conclusion

The goal of my analysis was to examine the impact of the COVID-19 pandemic on vehicle crash behavior in Salt Lake County, UT. I relied on a combination of time series analysis, data visualizations, and hypothesis testing. I first confirmed that the overall decrease in traffic volume after the start of the pandemic resulted in a statistically significant decrease in the number of vehicles involved in crashes. While I found that there was a slight increase in the proportion of crashes that resulted in a fatality, the overall profile of crash severity did not appear to differ significantly during the pandemic in Salt Lake County, UT. As vehicle crash behavior is influenced by the human condition and has a direct impact on societal resources, it serves as a valuable available metric for current and future research.

References

Bachman, D. (2020, November 11). *Research finds increase in car crashes with decrease in traffic during pandemic*. Eastern Connecticut State University. Retrieved December 01, 2021, from https://www.easternct.edu/news/_stories-and-releases/2020/11-november/research-finds-increase-in-car-crashes-with-decrease-in-traffic-during-pandemic.html

Centers for Disease Control and Prevention. (2021, 9 20). *Coronavirus Disease 2019 (COVID-19)*. Disease or Condition of the Week. https://www.cdc.gov/dotw/covid-19/index.html

Stiles, J., Kar, A., Lee, J., & Miller, H. J. (2021, 09 28). Lower Volumes, Higher Speeds: Changes to Crash Type, Timing, and Severity on Urban Roads from COVID-19 Stay-at-Home Policies. *Transportation Research Record: Journal of the Transportation Research Board*. 10.1177/03611981211044454

US Department of Transportation. (2017, 1 19). *KABCO Injury Classification Scale and Definitions*. Federal Highway Administration. https://safety.fhwa.dot.gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf

Utah Department of Public Safety. (2021). *Utah Crash Summary*. Crash Data and Statistics. https://highwaysafety.utah.gov/crash-data/

Yasin, Y. J., Grivna, M., & Abu-Zidan, F. M. (2021). Global impact of COVID-19 pandemic on road traffic collisions. *World Journal of Emergency Surgery*, *16*(51), 1-14. 10.1186/s13017-021-00395-8

Data Sources

Utah Department of Public Safety - Utah Crash Summary
https://udps.numetric.net/utah-crash-summary#/?view_id=7

The RAW_us_confirmed_cases.csv file from the Kaggle repository of John Hopkins University COVID-19 data
https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

The CDC dataset of masking mandates by county
https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

The New York Times mask compliance survey data
https://github.com/nytimes/covid-19-data/tree/master/mask-use