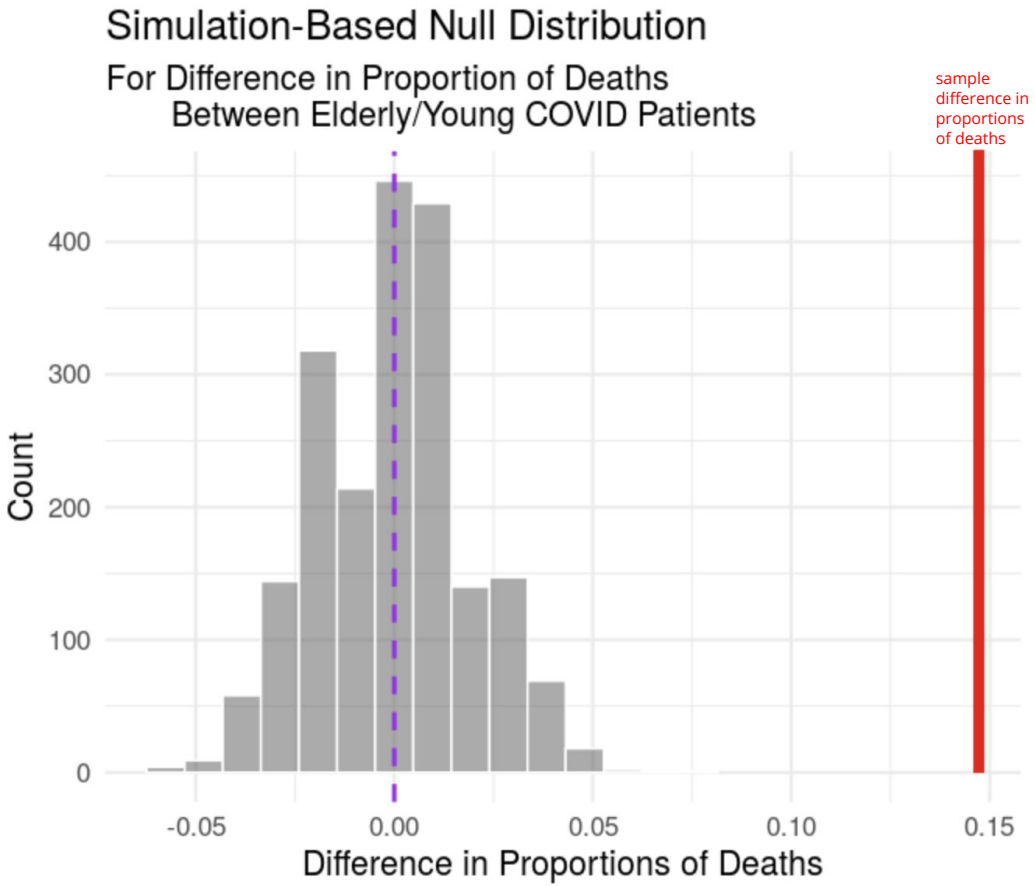
Research Question & Dataset

As many reports initially claimed that older COVID-19 patients experience more severe symptoms, we wanted to explore the relationship between age and survival status. In doing so, we investigated the impact of several factors on this relationship; these factors include gender, country of origin, if the patient was from Wuhan, and whether the patient visited Wuhan.

- We used a dataset from the early stages of the outbreak, specifically of 1085 observations from 1/20/2020 to 2/15/2020, in which each observation is a case of a patient who tested positive for COVID-19.
- The variables of the dataset include those mentioned above, in addition to the date the case was reported, city, date of symptom onset, start and end dates of exposure to virus, among others.

Hypothesis Test for Age

- One-sided hypothesis test for age variable
  - Divided patients into “old” and “young” age groups based whether they were over or under 60 years of age, respectively, as the mean age of deaths in our sample was 68
- **Null hypothesis**:
- $H_0 : p_{old} = p_{young}$
- **Alternative Hypothesis:**
- $H_A : p_{old} > p_{young}$
- where p_{old} = the true proportion of deaths in patients 60 years or older and p_{young} = the true proportion of deaths in patients younger than 60
- **P-value** = 0. Thus, we have sufficient evidence to reject the null hypothesis at the α = 0.05 level



Conclusions

1. Age of the patient did bear impact on the survival status of the patient
    - From our hypothesis test for age, we there is significant evidence to support that the death rate due to COVID-19 of patients aged 60 or older is higher than that of patients younger than 60

2. The gender of the patient and whether they lived in Wuhan also influenced the survival status of patients, but on a very small scale

3. We were unable to reach significant conclusions for the relation of death outcome to other explanatory variables, mainly due to the limitations of the data set and the challenges of working with data from a current pandemic