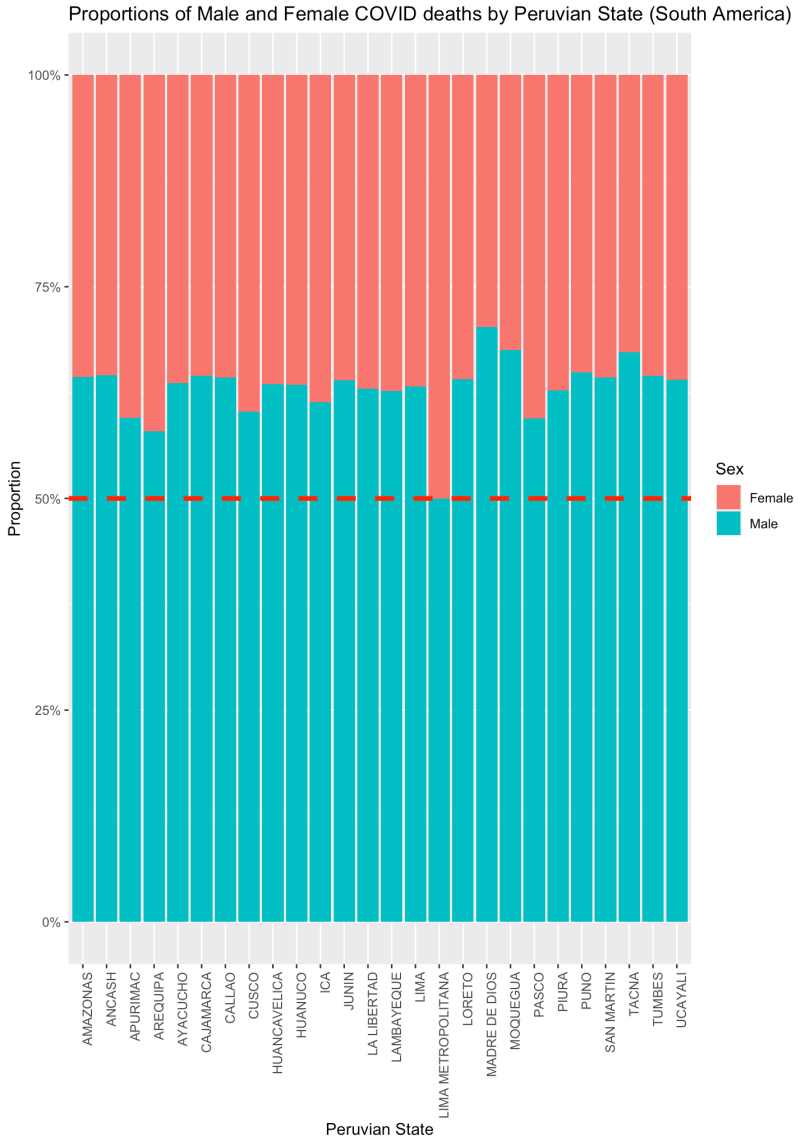

We investigated the distribution of COVID-19 deaths across different states in Peru and compared the death rates between males and females (On the left, you can see a small visualization of our dataset). Our hypotheses were as follows, with a significance level of $\alpha = 0.05$:

H_0 – *There is no difference in deaths by sex amongst each Peruvian state.*
H_a – *There is a difference in deaths by sex amongst each Peruvian state.*

After cleaning the data by removing mislabeled and irrelevant entries, leaving only male and female death counts for the analysis, a contingency table was created to summarize deaths by sex across each state. Afterwards a chi-squared contingency test was done with 25 degrees of freedom yielding a p-value of $3.7e^{-40}$.

Since this p-value is significantly lower than our significance level, the null hypothesis was rejected, concluding that *there are indeed differences in how COVID-19 affected males and females across different regions of Peru*.

Source: https://www.datosabiertos.gob.pe/dataset/fallecidos-por-covid-19-ministerio-de-salud-minsa

Assignment 2 for Data Science in Bioinformatics
Grace Rico
Emilio Sopprani
Gonzalo Quintanilla
Mark Hegedus