

CORRELATION IN CAUSE OF DEATH

Derek Gorthy, Bryce Melvin, Michael Condon

Questions We Wanted to Answer

- Can we reproduce the findings of the CDC and other primary sources?
- Can we replicate the results of studies conducted by secondary analysts?
- Can we provide evidence for our unconventional cause of death assumptions?
 - Friday the 13th
 - Full moon
 - Stock market performance


Data Preprocessing

- Cleaning unnecessary attributes
 - Python script
- Date estimation
 - Previous examples
 - Uniformly-random distribution
- Adding date-dependent attributes

Tools Used

- Python
- R
- Github

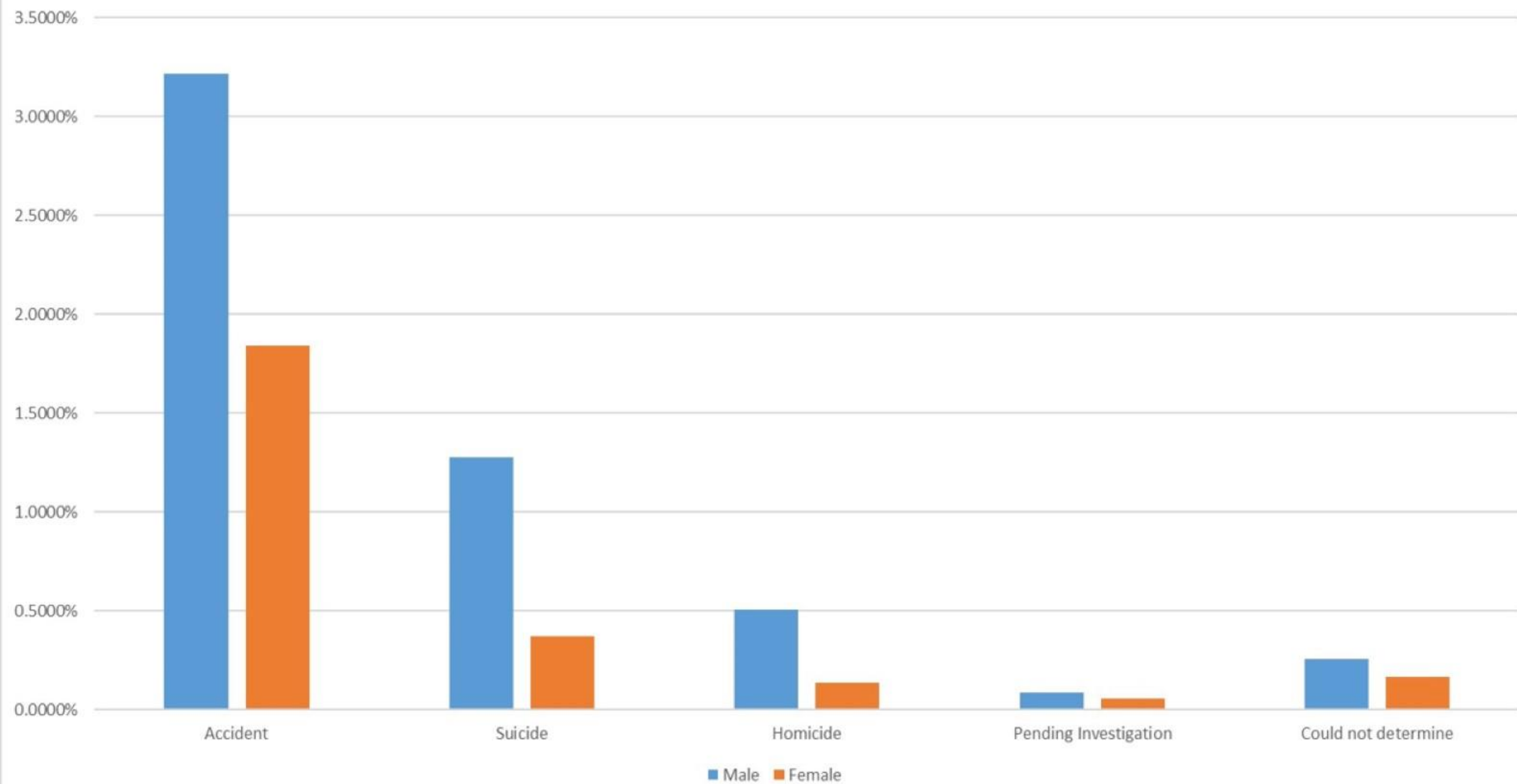
Data Classification

- Basic correlations calculations
 - Bayesian technique
 - Does a full moon influence total daily suicides?
 - Total daily suicides given that it is a full moon
- 

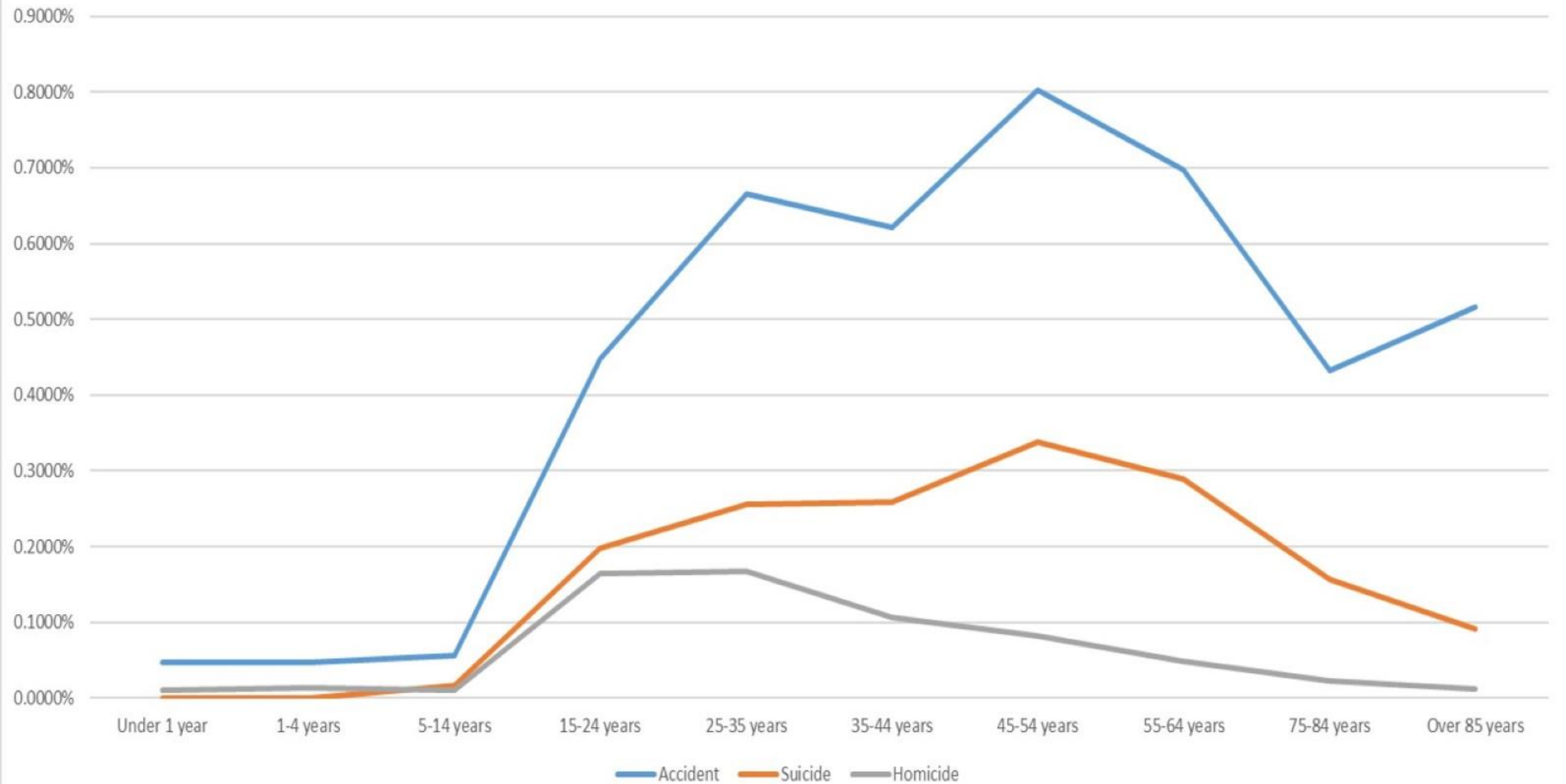
Interesting Findings...

- Were able to reproduce the CDC's results

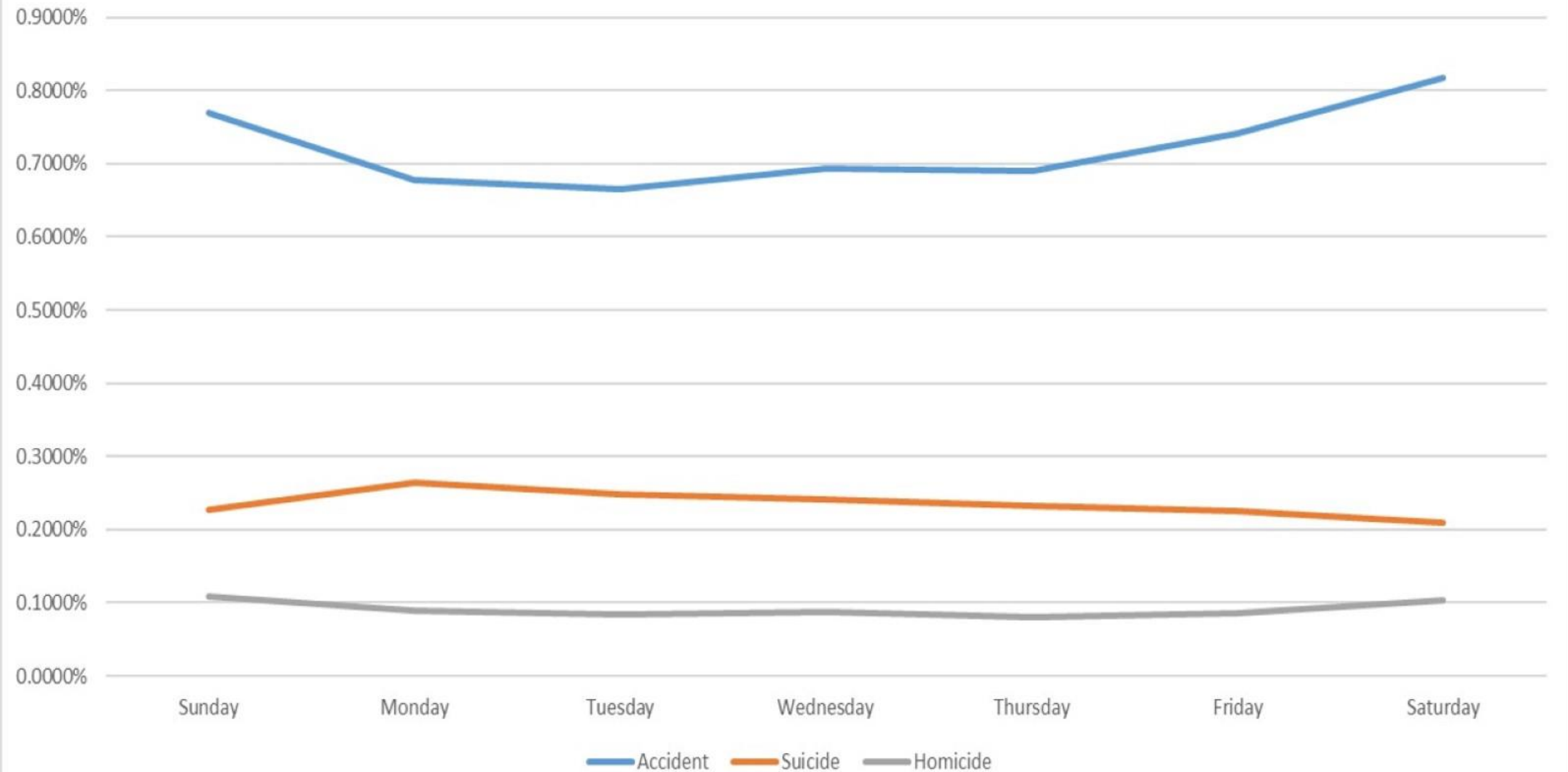
Cause of Death by Sex



Cause of Death given Age

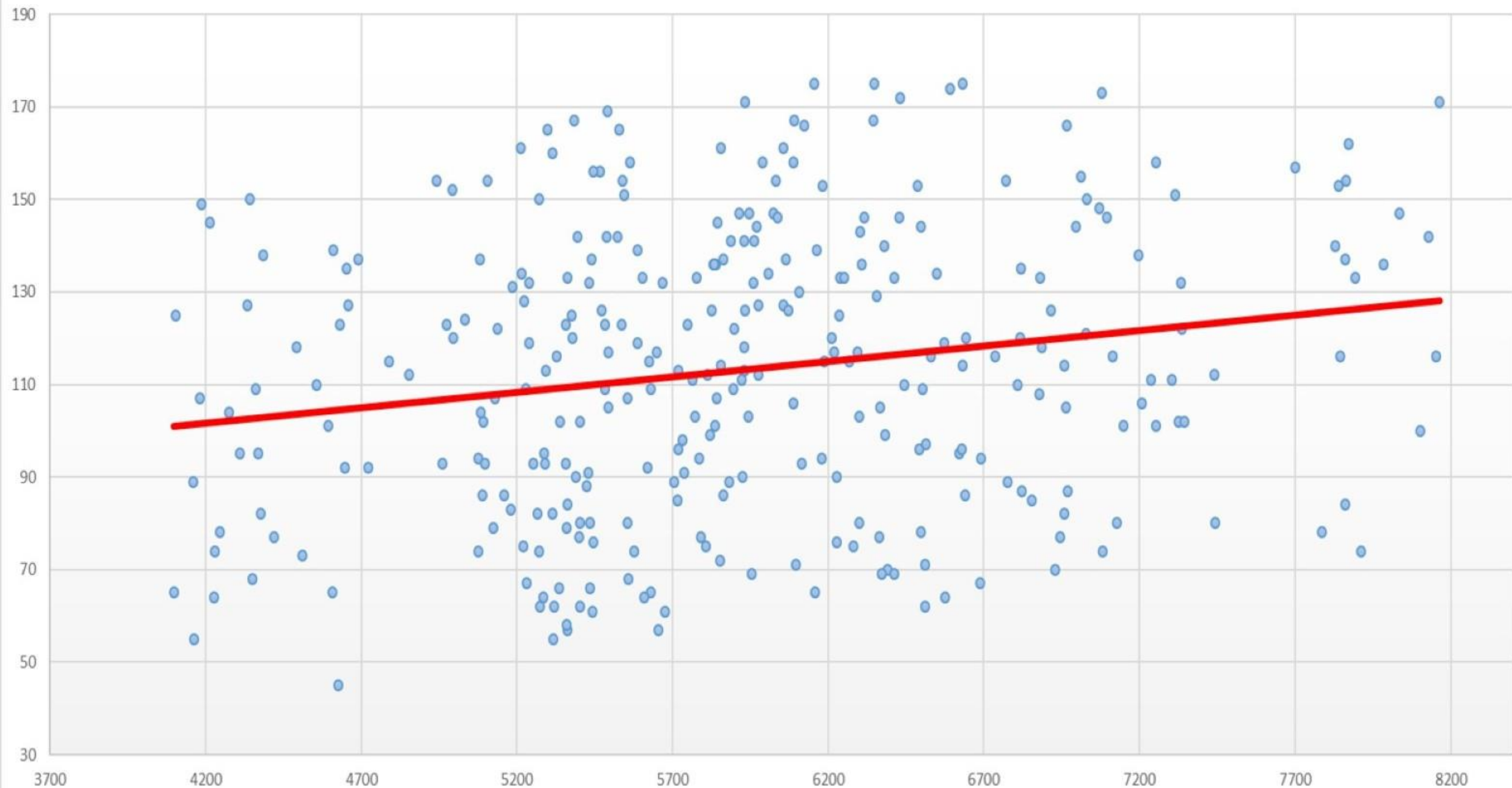


Cause of Death by Day of Week

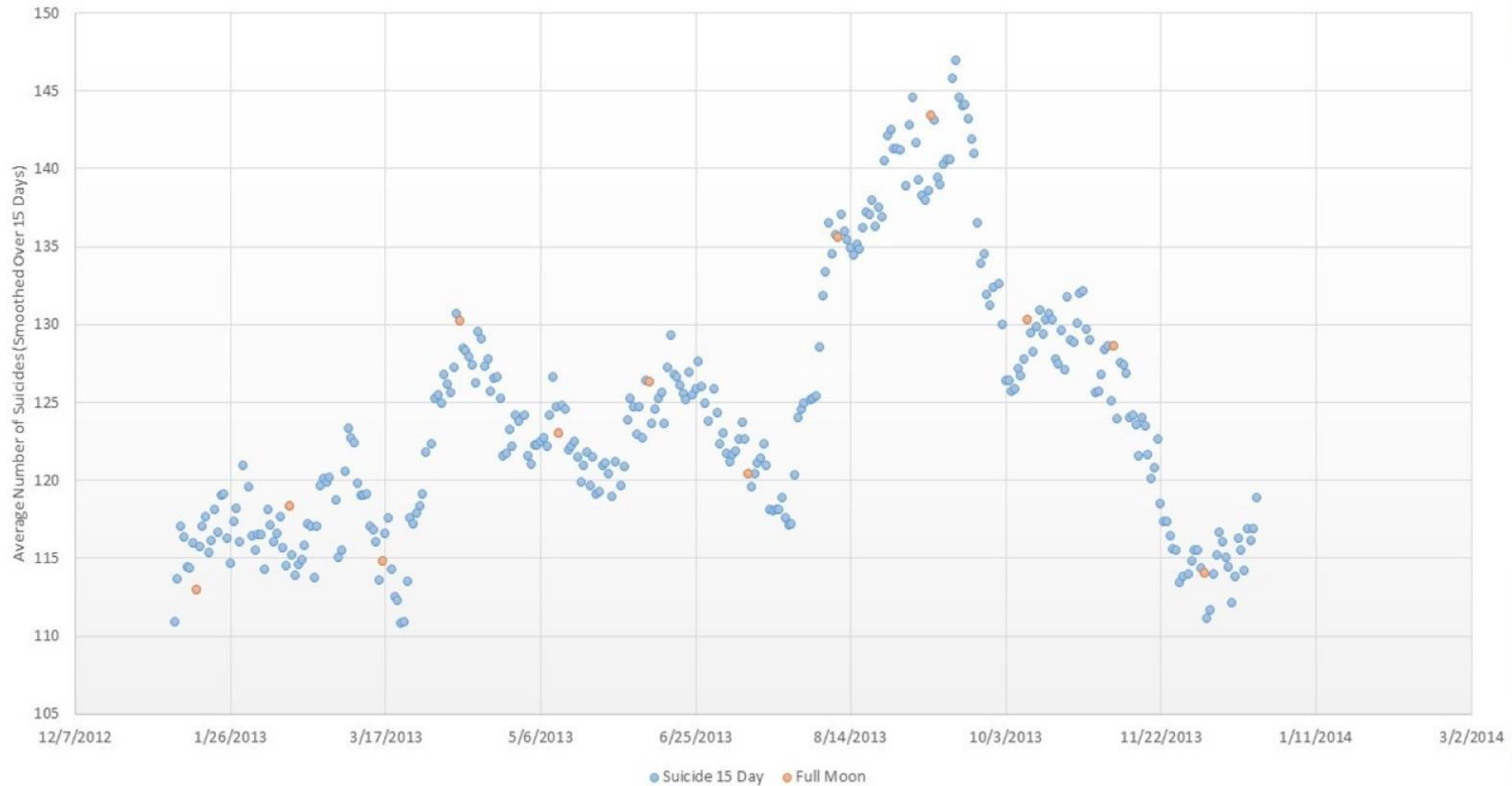


Sun Spot Number vs. Natural Deaths (removed outliers)

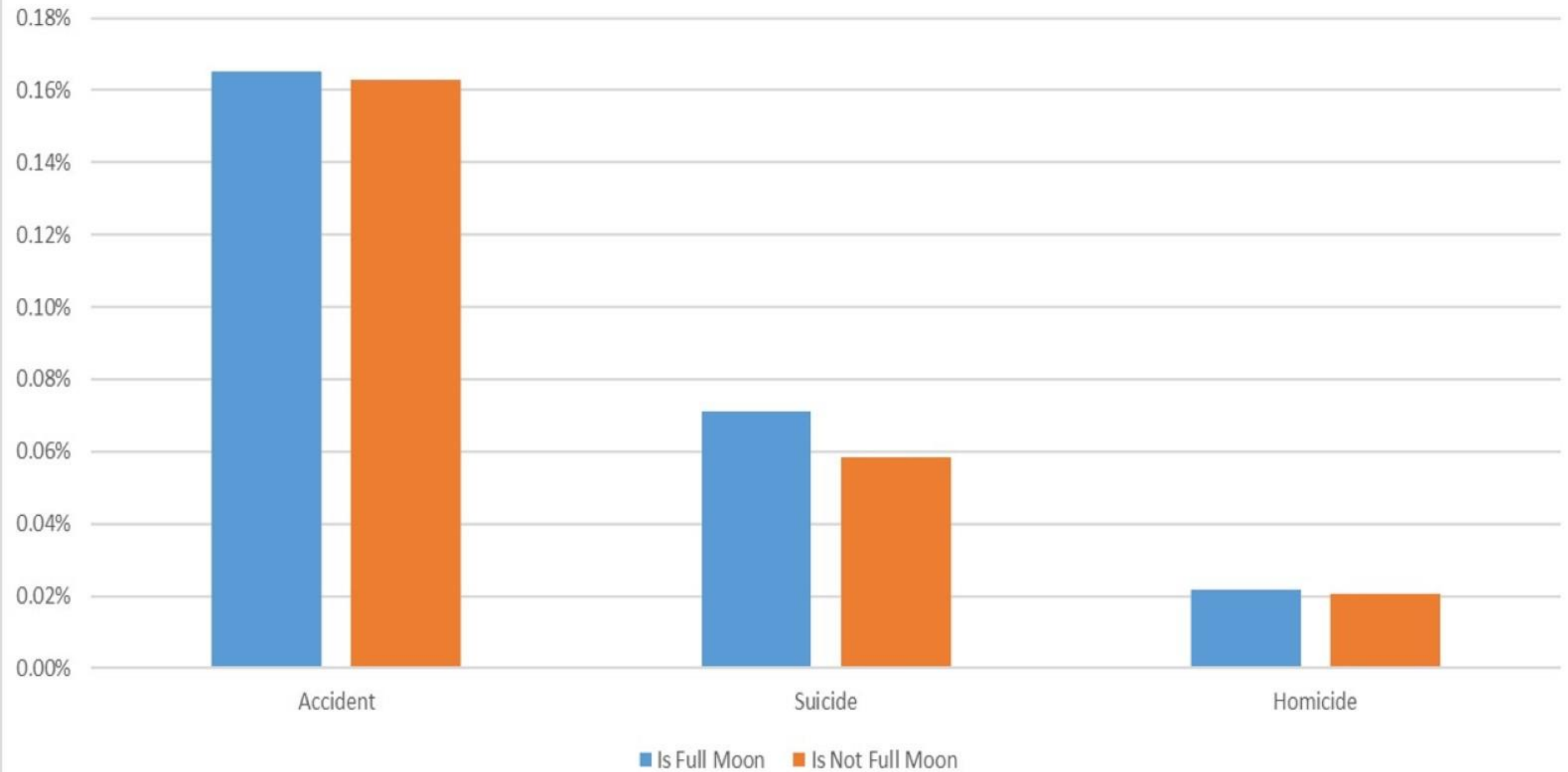
Correlation Coefficient = 0.1963



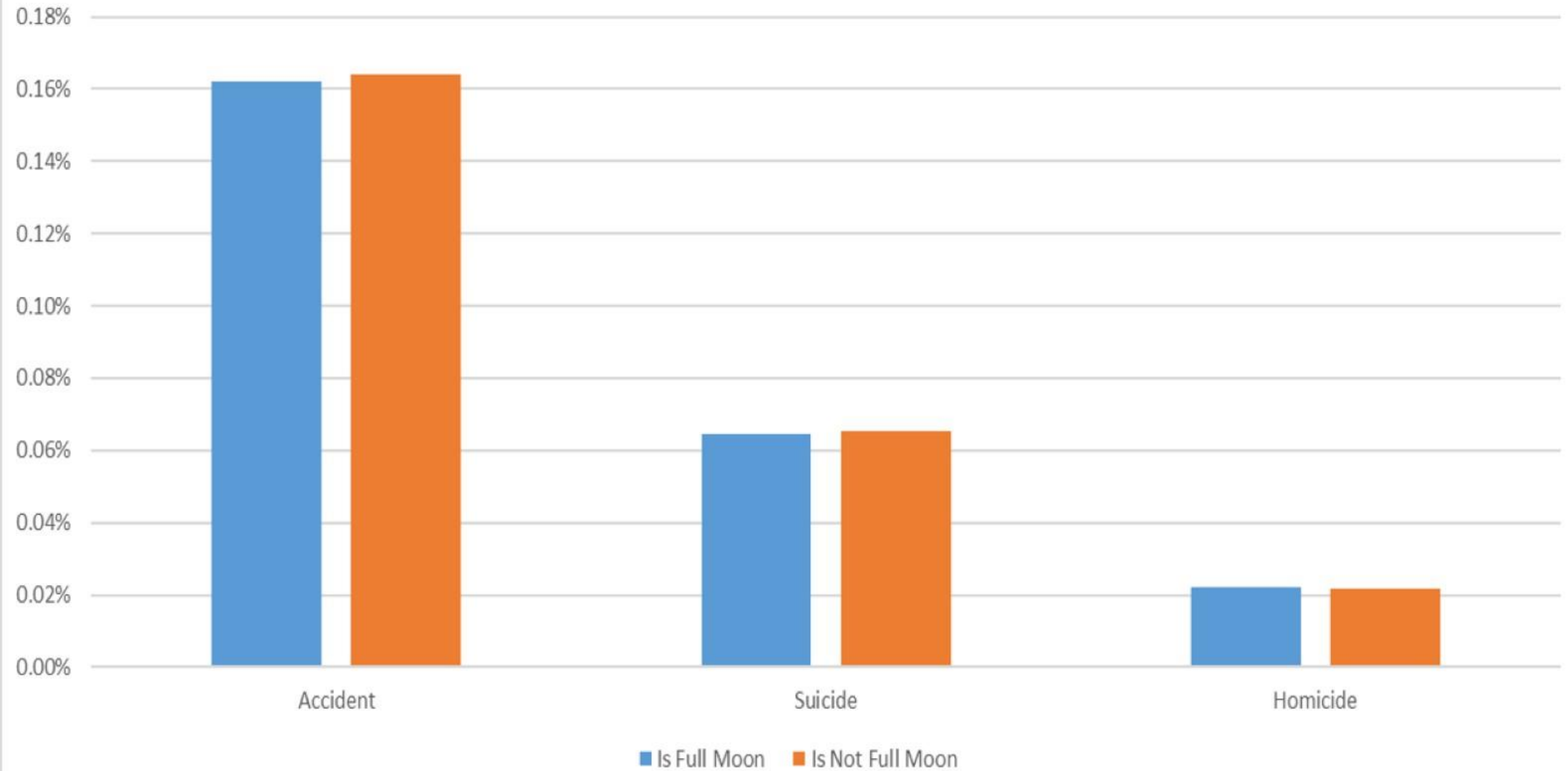
15-Day Smoothed Suicide over Entire Year



Cause of Death given Full Moon



Cause of Death given Full Moon (Supplementary Set)



Applications

- Given different granularity, identify the most useful measure(s) and remove the rest
- Identifying the distribution of the estimated attribute value
- Running sensitivity analysis to check results