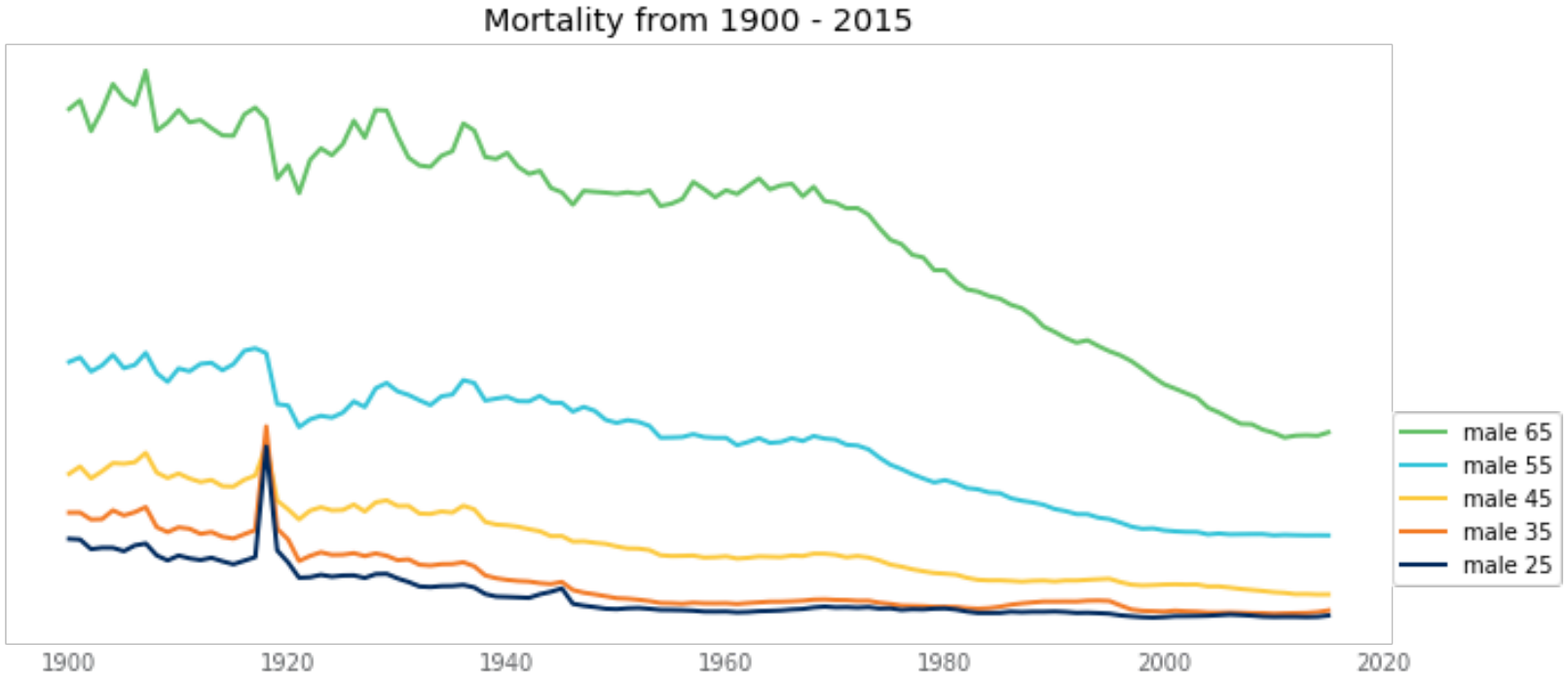
DSI Capstone
US Population
Mortality Analysis

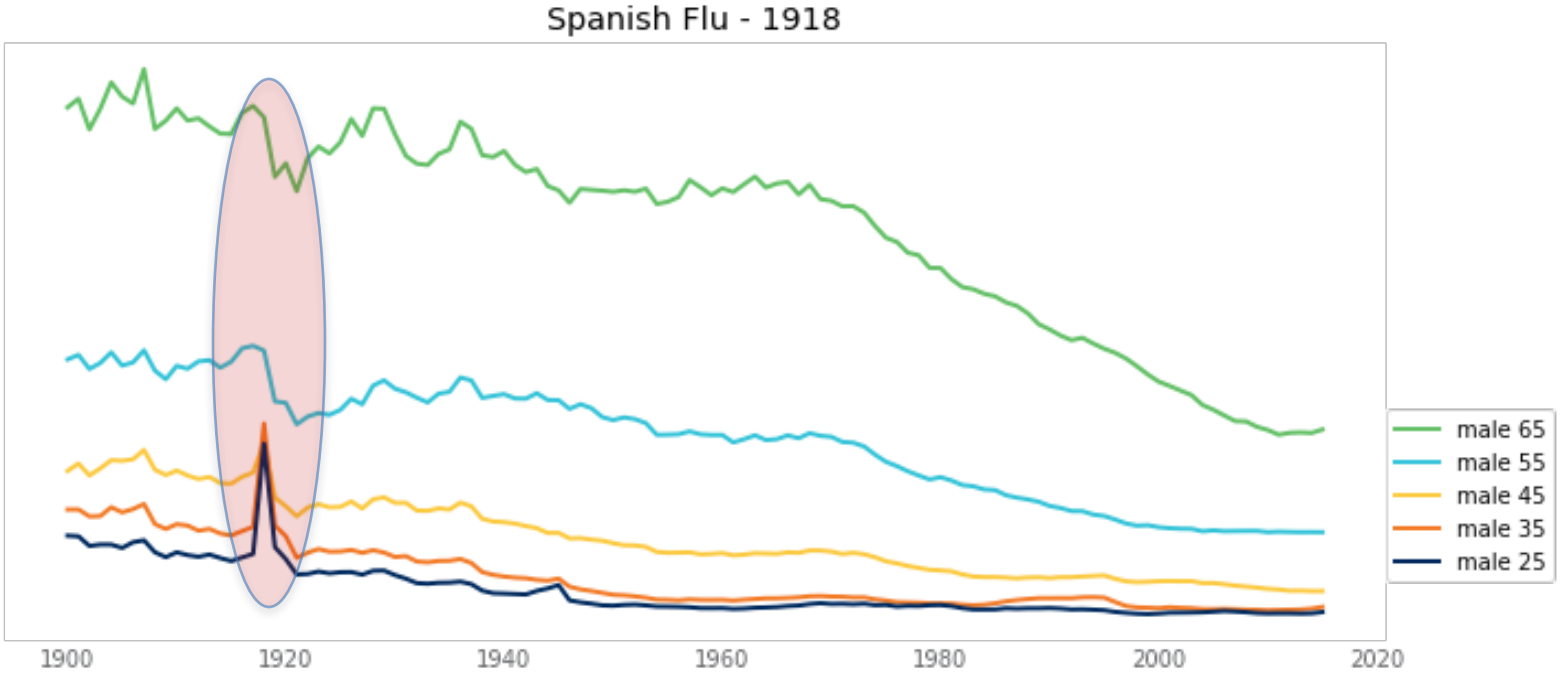
Scott Wright
October 2018

Mortality over time

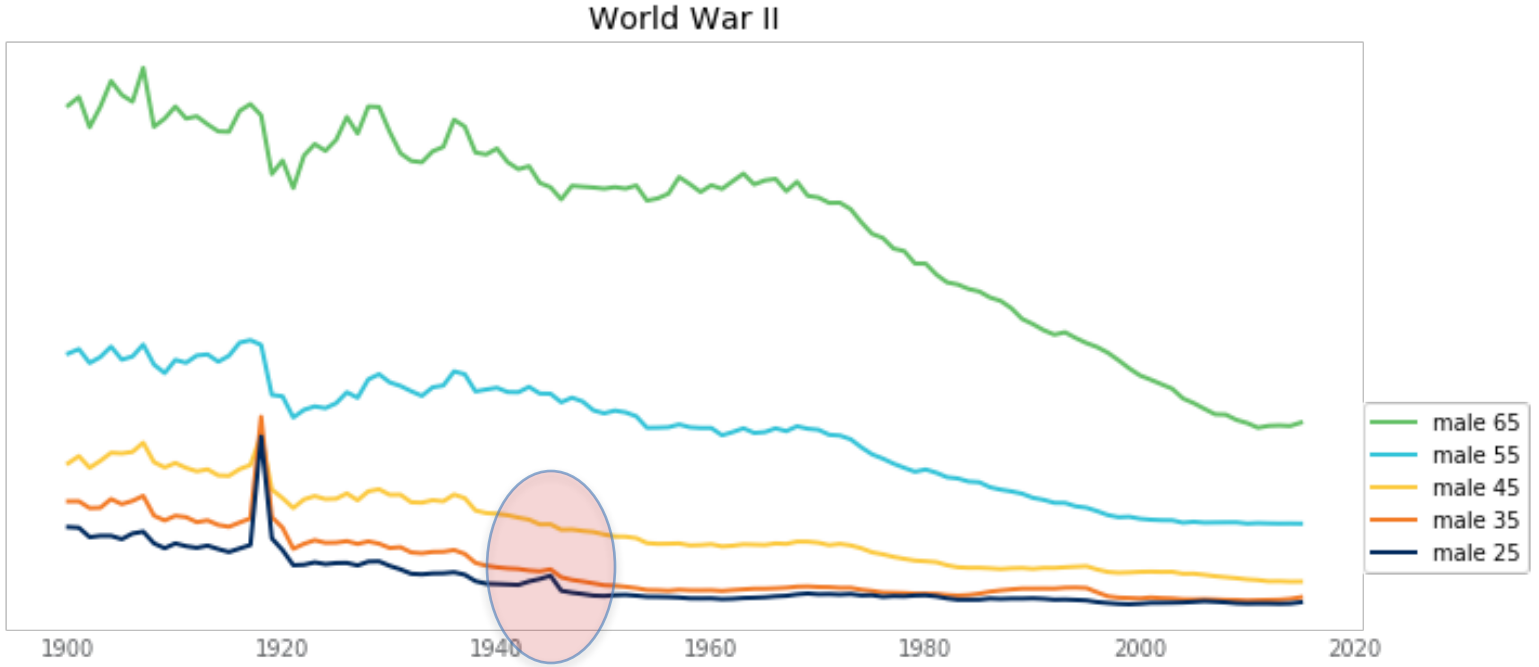


SSA Mortality Rates:
https://www.ssa.gov/OACT/HistEst/Death/2018/DeathProbabilities2018.html

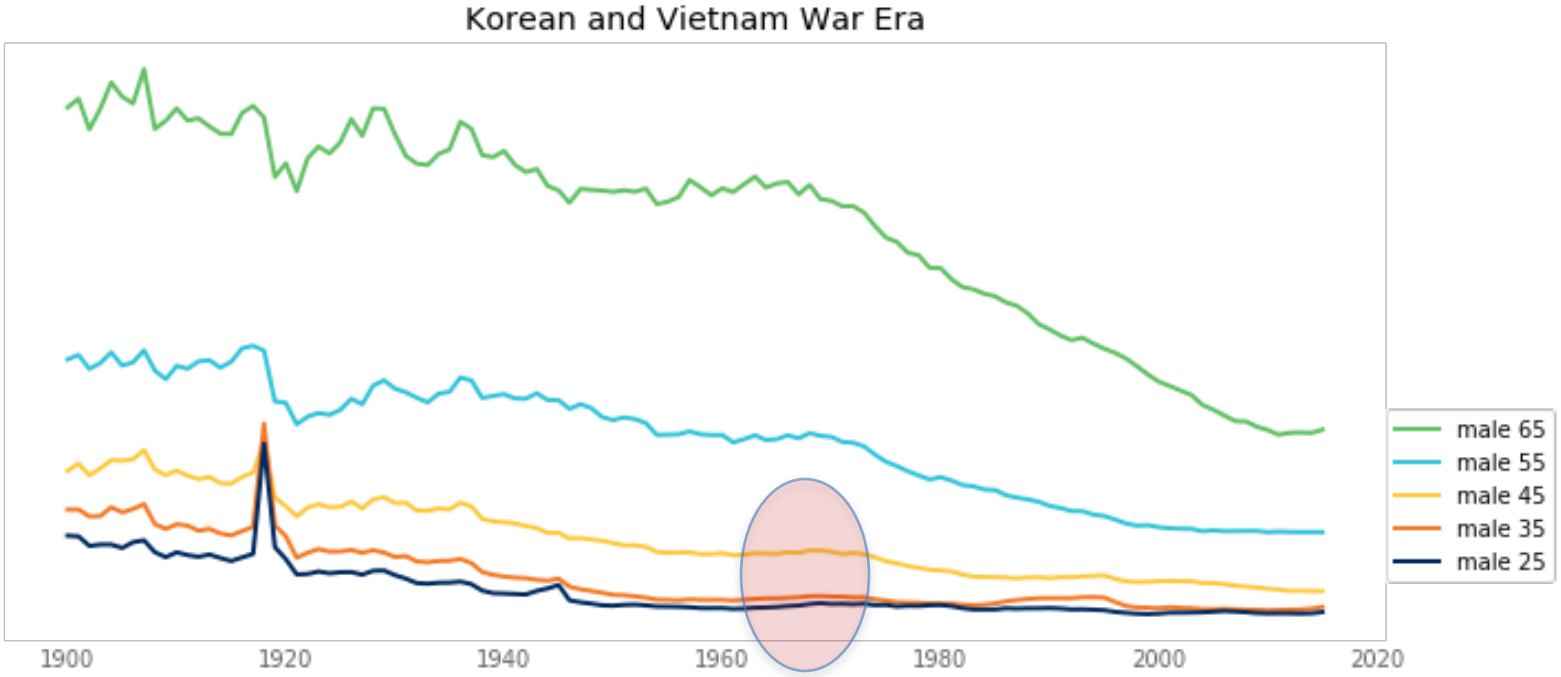
Mortality over time



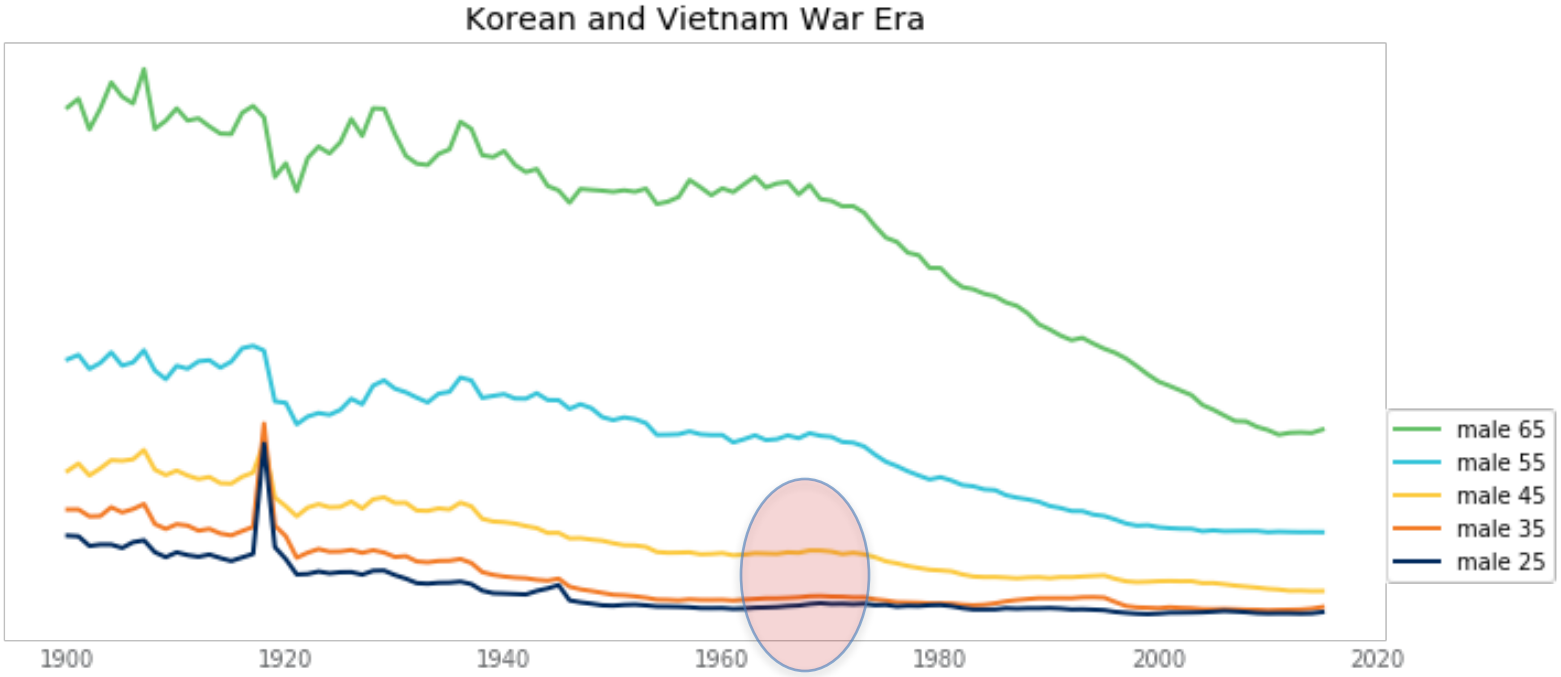
Mortality over time



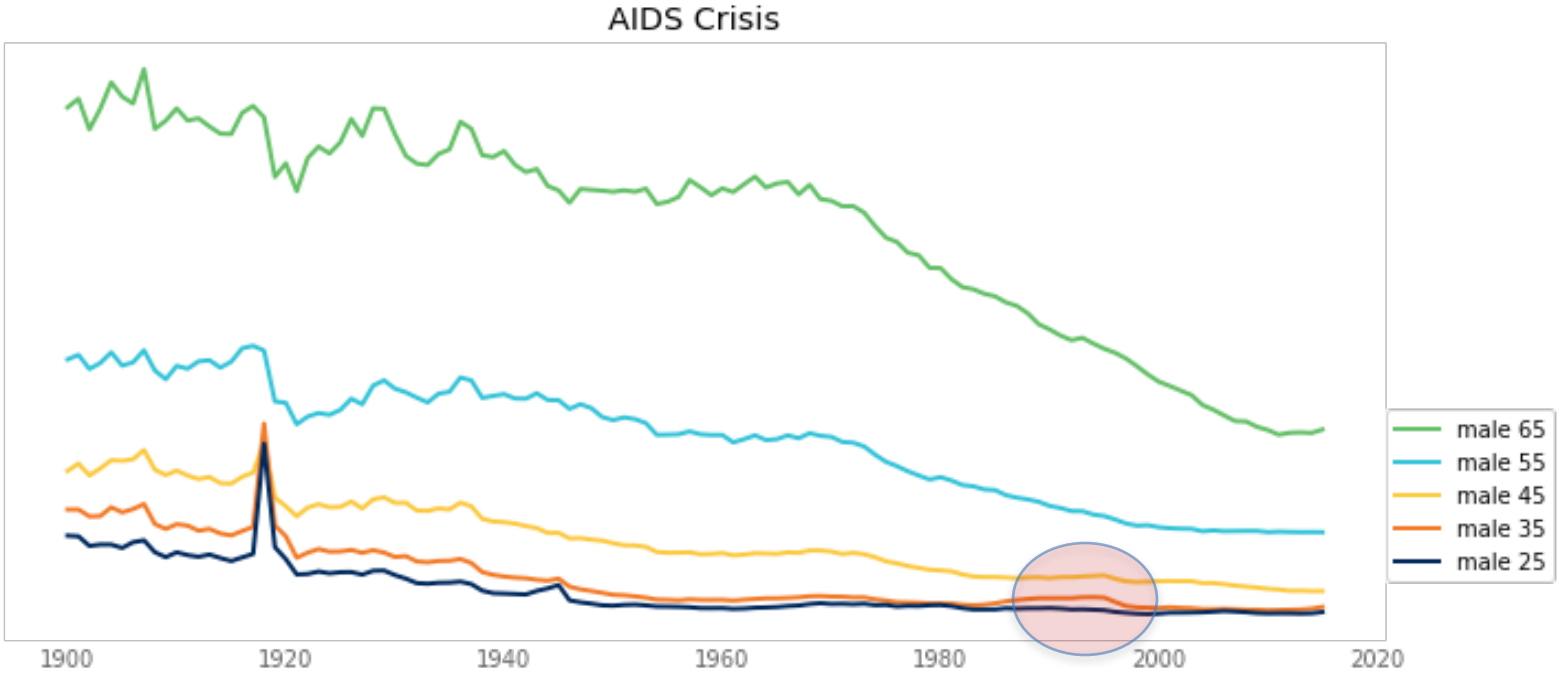
Mortality over time



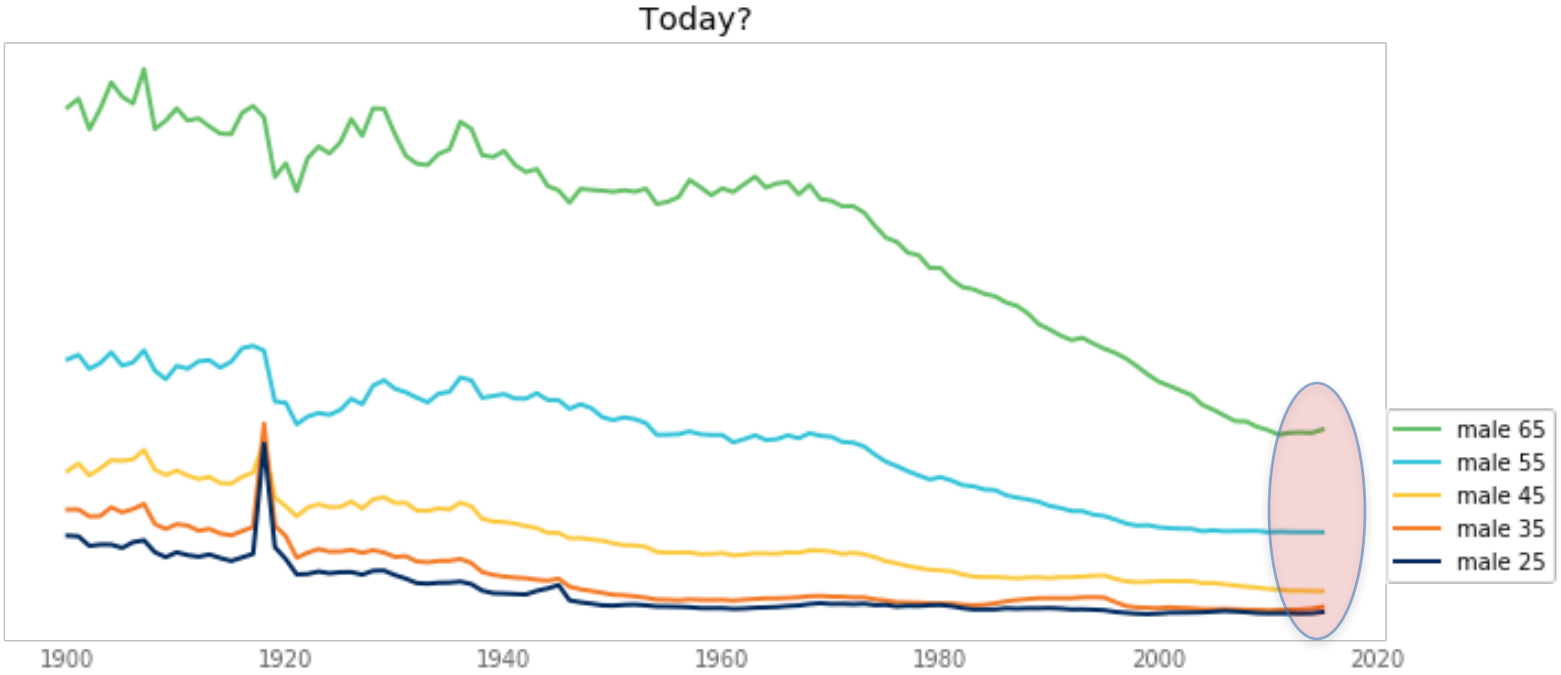
Mortality over time



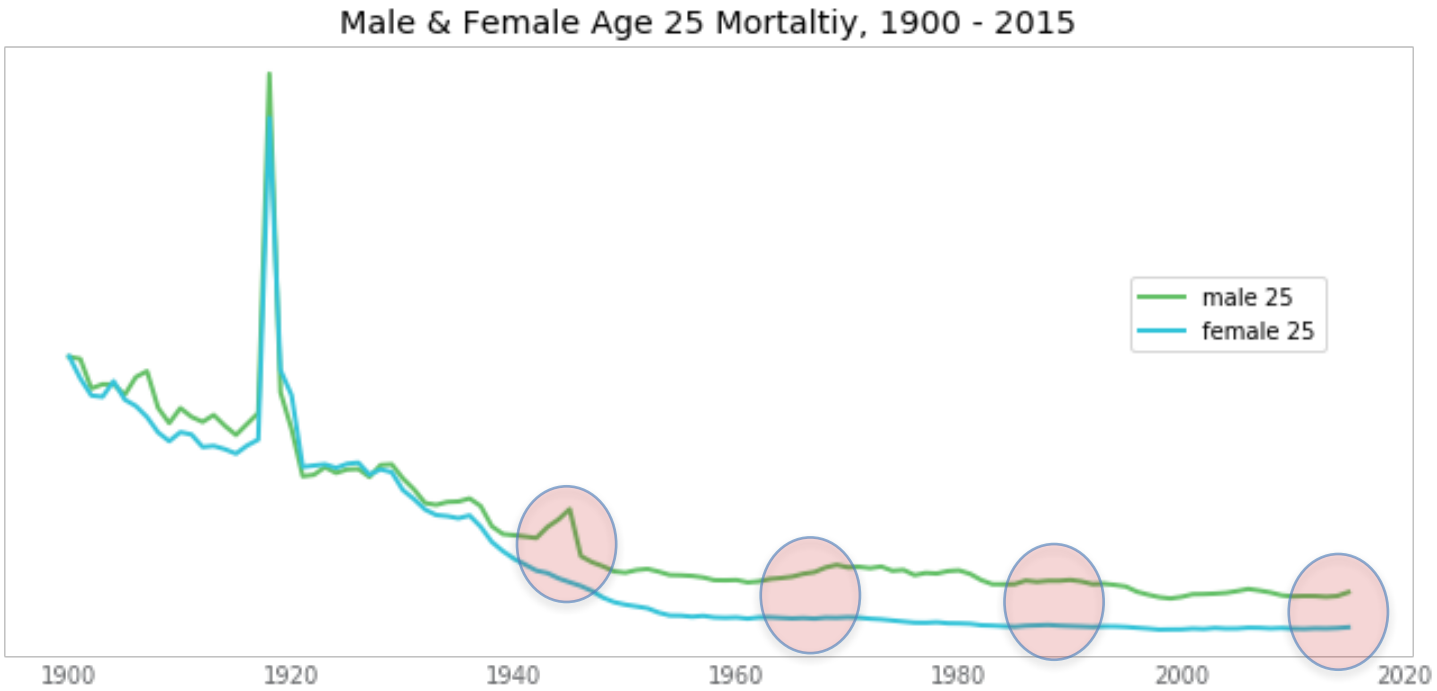
Mortality over time



Mortality over time



Mortality over time



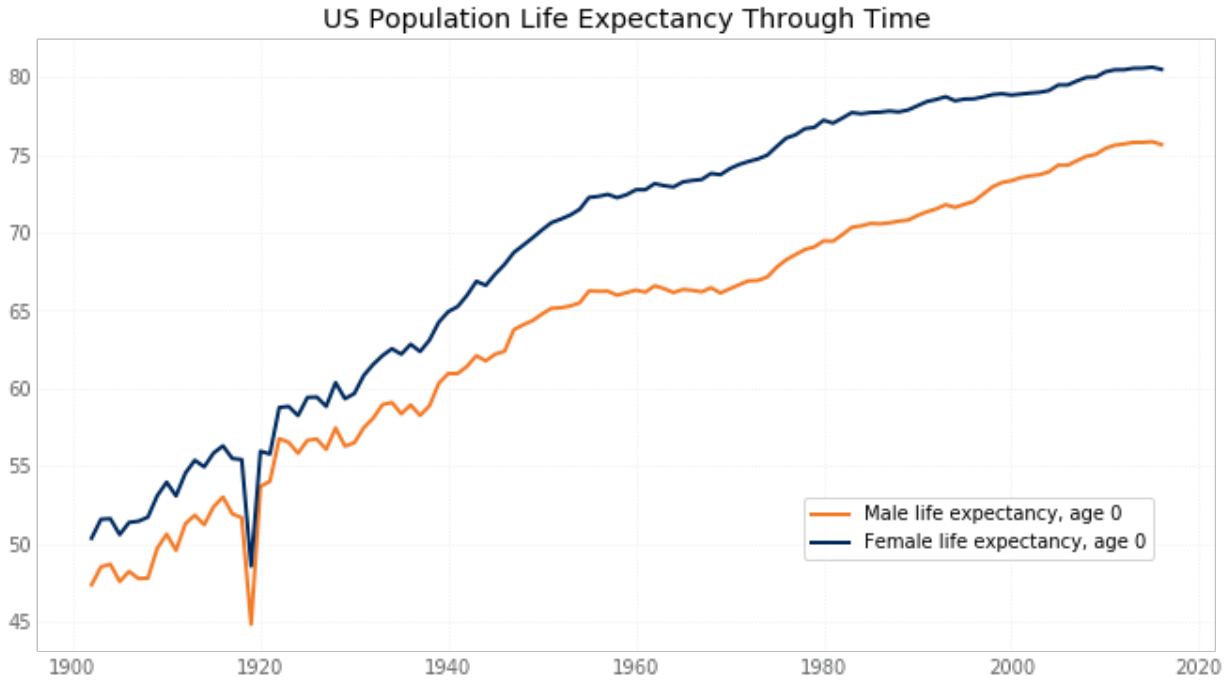
Life expectancy

- Difficult to analyze 100 ages for years of duration for both sexes
- Also, looking at improvement or deterioration (year-over-year percentage change in mortality) is tough
- Instead, look at life expectancy at age 0:

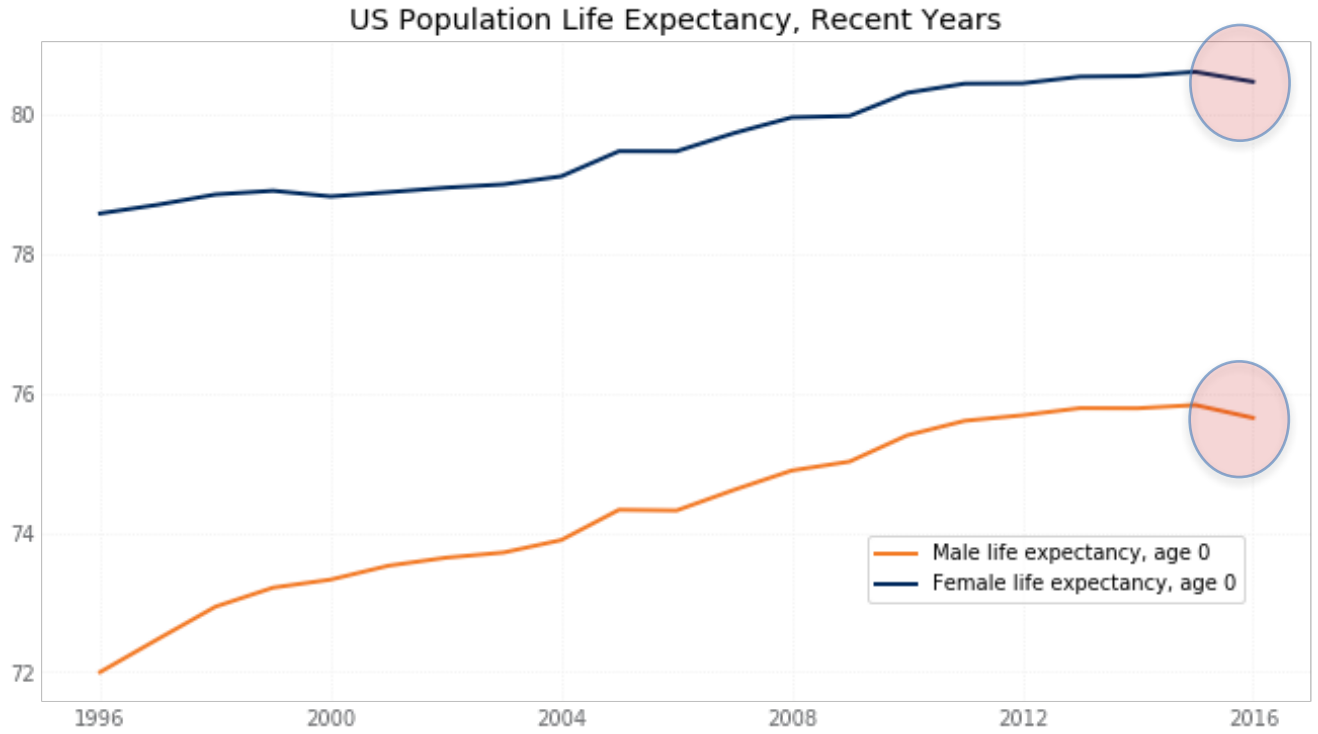
$$e_o = \sum_{t=1}^{\infty} p_0(t)$$

Where $p_x(t)$is the probability someone aged x lives for t years

Life expectancy over time (1900 – 2015)



Life expectancy over time – recent years



Initial question / modeling

- It's an arduous process to compute mortality tables.
- Are there external factors that can help predict life expectancy before tables are completed?

- Looked at an ARIMA model with exogenous variables (features)

Features explored for ARIMA model

Feature collected	Male model	Female model
First difference	[✓]	[✓]
GDP	[ ]	[ ]
GDP per capita	[ ]	[ ]
GDP increase yr over yr	[ ]	[ ]
Household income 40th percentile	[ ]	[ ]
Household income 95th percentile	[ ]	[ ]
Income inequality measure	[✓]	[ ]
Labor participation rate	[✓]	[ ]
Military spending - dollars	[ ]	[ ]
Military spending as a percent of GDP	[ ]	[ ]
Military spending per capita	[ ]	[ ]

Feature collected	Male model	Female model
Military spending yr over yr increase	[✓]	[ ]
Inflation	[ ]	[ ]
Performance of S&P 500	[✓]	[✓]
Return on 3mo T-bill	[ ]	[ ]
Return on 10y T-bond	[ ]	[ ]
Health insurance coverage percent	[ ]	[ ]
Overweight	[✓]	[✓]
Obese	[ ]	[ ]
Severely obese	[ ]	[ ]
Alcohol consumption per capita	[✓]	[✓]
Tobacco usage percent	[ ]	[ ]

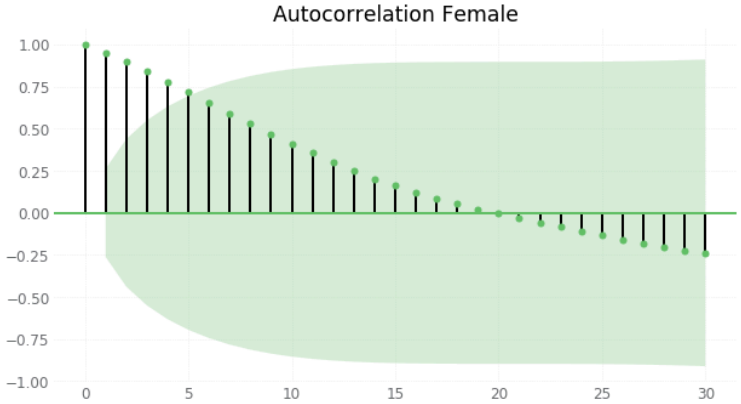
Sources for feature variables

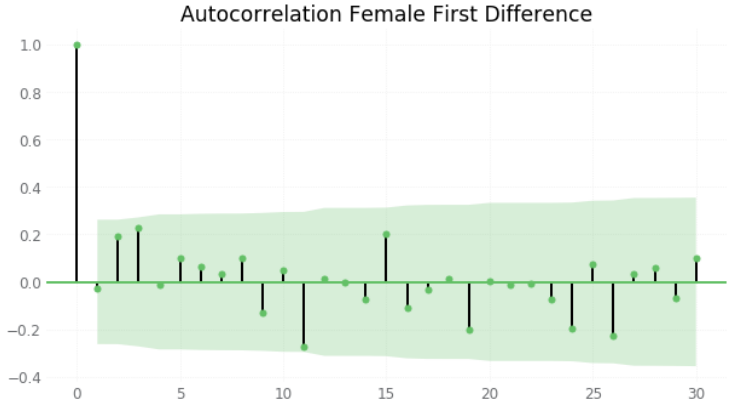
- Mainly US governmental data sources (CDC, Census Bureau, Health & Human Services, Federal Reserve Bank of St. Louis' FRED)
- Also used data from the World Bank
- Main limitation: annual data that only goes back to the 1960s.
- Had to impute values for some missing data (mainly interpolated between observations)

ARIMA model exploration

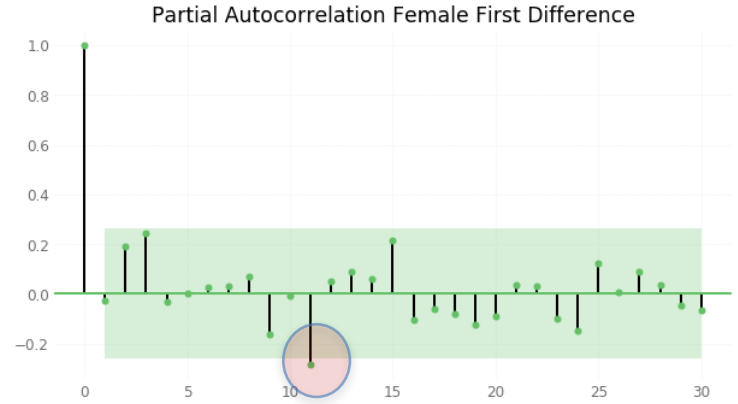
- Life expectancy did not exhibit stationarity
- Dickey-Fuller test showed p-value of 0.95 and 0.22 for male and female, respectively
- First difference did exhibit stationarity at alpha = 0.05 with p-values of 0.013 and 0.002.
- Conclusion ➔ start with ARIMA parameter d = 1.

ACF and PACF Plots





- Both male & female plots were similar (showing female here)
- Based on ACF and PACF plots, initial p and q values set to 0.
- Also added in a duration 11 lag variable of the target



Model Testing and Fit

- Tried many combinations of features to arrive at 7 for the male model and 4 for the female mode
- Hyper-parameter tuning on p, q, and d did not improve the initial ARIMA(0,1,0) model.
- Model fit based on R-squared:

R squared	Male model	Female model
train	99.1%	99.1%
test	97.8%	96.6%

Features explored for ARIMA model

Feature collected	Male model	Female model
First difference	[✓]	[✓]
GDP	[ ]	[ ]
GDP per capita	[ ]	[ ]
GDP increase yr over yr	[ ]	[ ]
Household income 40th percentile	[ ]	[ ]
Household income 95th percentile	[ ]	[ ]
Income inequality measure	[✓]	[ ]
Labor participation rate	[✓]	[ ]
Military spending - dollars	[ ]	[ ]
Military spending as a percent of GDP	[ ]	[ ]
Military spending per capita	[ ]	[ ]

Feature collected	Male model	Female model
Military spending yr over yr increase	[✓]	[ ]
Inflation	[ ]	[ ]
Performance of S&P 500	[✓]	[✓]
Return on 3mo T-bill	[ ]	[ ]
Return on 10y T-bond	[ ]	[ ]
Health insurance coverage percent	[ ]	[ ]
Overweight	[✓]	[✓]
Obese	[ ]	[ ]
Severely obese	[ ]	[ ]
Alcohol consumption per capita	[✓]	[✓]
Tobacco usage percent	[ ]	[ ]

Conclusions

- It is possible to see a relationship between macro-economic variables and life expectancy in the US
- The overall model was only slightly better than a pure auto-regressive model in terms of R-squared fit.
- The question is difficult to answer due to a low number of observations (<60).
- It is worthwhile to continue to look at these variables, but be willing to adapt over time since some male/female diffference could change due to chaning social norms.

Thank you for your time

Questions?