Global Trends in Mortality of Prominent Individuals

Enora Cattanach, Rachael Gu, Jack Nugent, and Garrett Pinkston
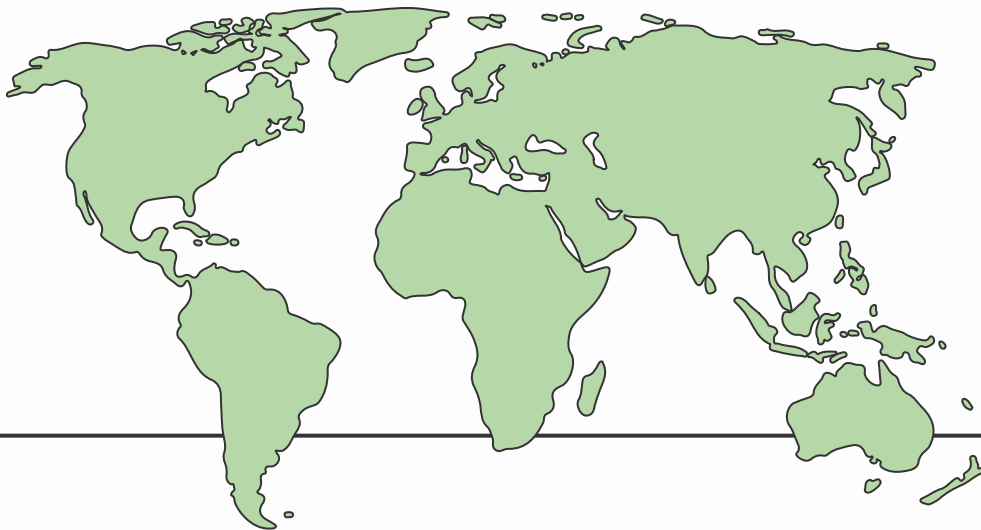
Table of contents

Motivation

Our goal is to explain **who lives longer, why, and how** longevity in the past century has changed with the many recent advancements in technology and medicine

1. How do gender, region, and occupation individually influence survival?

2. How do these factors jointly affect mortality risk?

3. How have major historical events shaped cohort-level survival patterns?

1

Data and Methodology

Tools to Address Research Questions

Dataset (Brief History of Human Time)
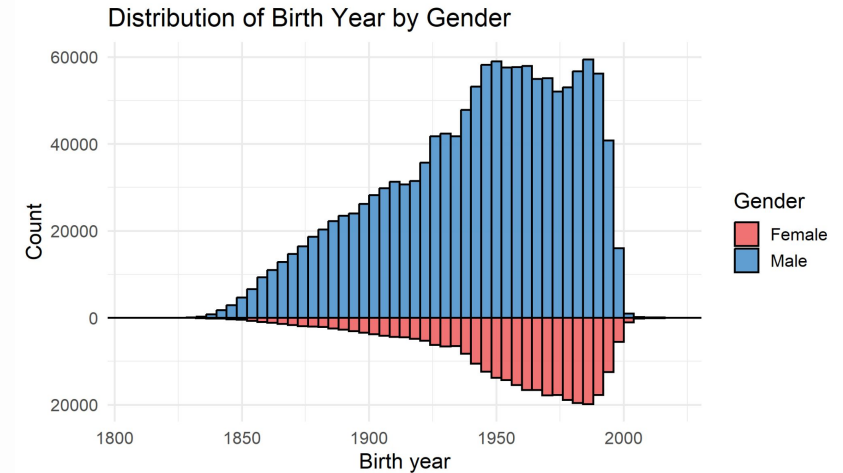
Data Filtering

- Kept (~75% of total individuals)
  - Born >= 1925 and no death year (assumed to be alive)

- Included anyone who was alive at any point in the last 100 years, rather than only those born in the last 100 years
  - **Rationale:** maximize sample size and reduce censoring

Data Cleaning

- Determined most reliable death year

- Removed
  - Missing birth years
  - Born < 1925 and no death year (likely deceased)
  - Birth year > death year (negative lifespan)

Demographics

- Over 1.6 million people who died in 1925 or later
  - 634,850 still alive (37.9% of data)
- Sample is 81.5% men, 18.5% women
- Birth year ranged from 1810 to 2018
- Athletes tend to be born more recently
- Across most occupations and regions, the women tend to be younger than the men

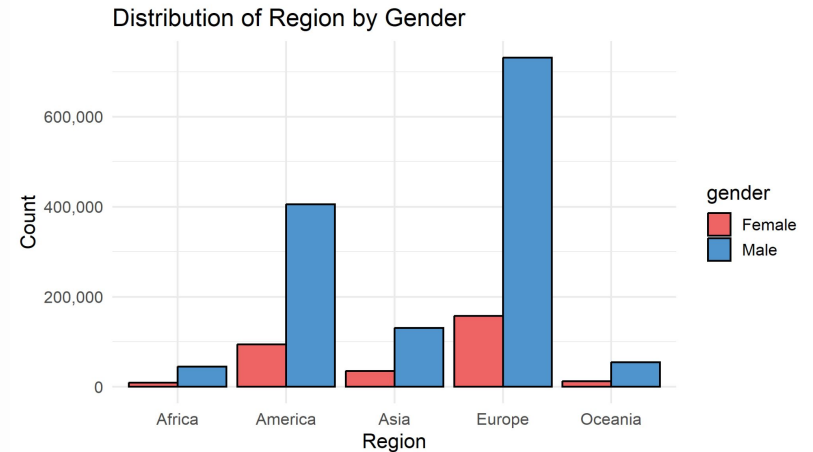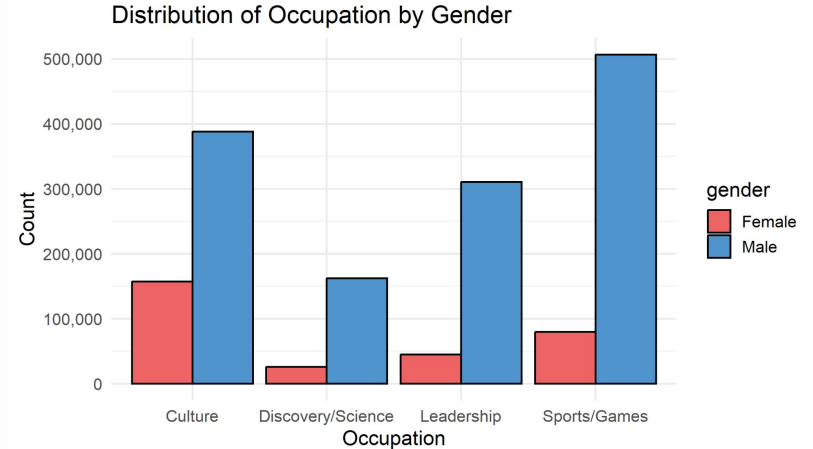
Distribution of Birth Year by Gender

Demographics

- **Occupation:**
  - Sports/Games: 34.5%
  - Culture: 32%
  - Leadership: 21%
  - Science: 11%
  - Other: 1.5%
- **Regions:**
  - Europe: 53%
  - America: 30%
  - Asia: 10%
  - Oceania: 4%
  - Africa: 3%

Distribution of Occupation by Gender

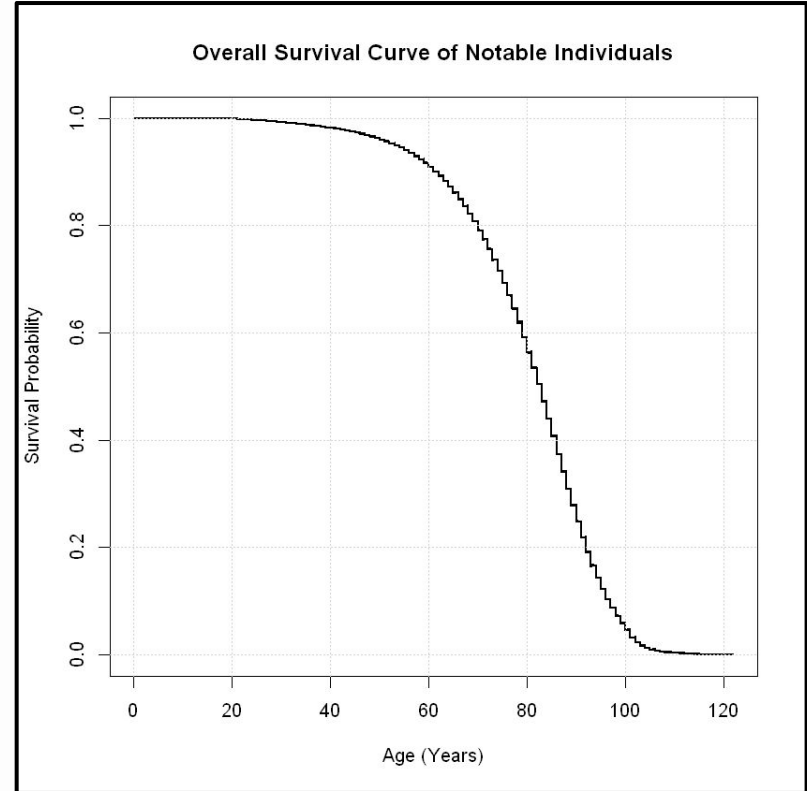
Distribution of Region by Gender

2

Results

EDA, Survival Modeling, and Mortality Trends

Question 1

How do gender, region, and occupation influence the survival time of notable individuals over the past 100 years?
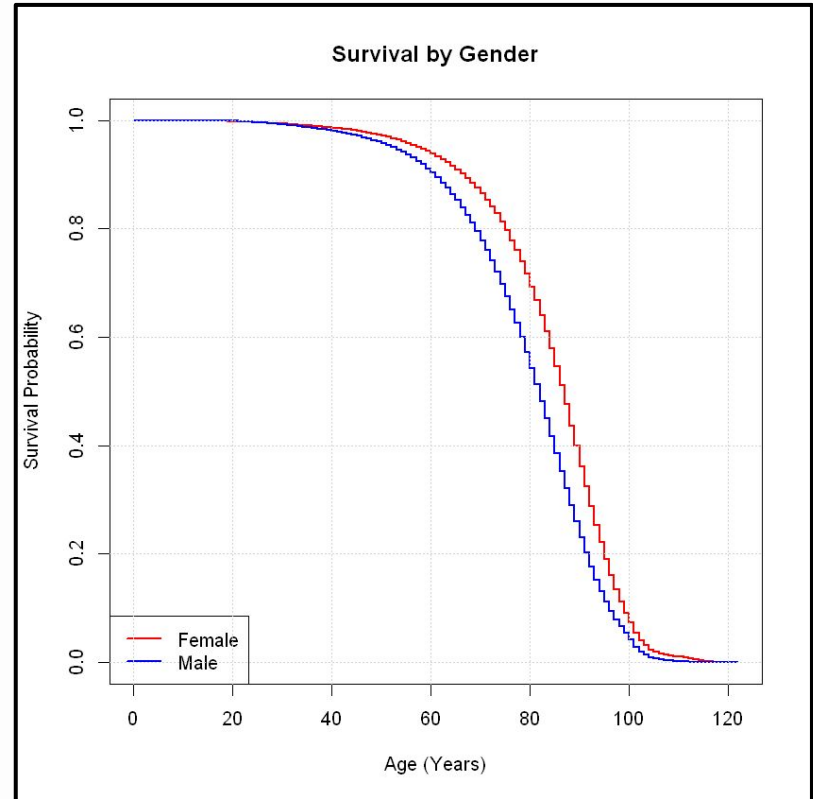
Overall Survival Curve

- **Early Stability (Ages 0-40)**
  - Survival probability is close to 1 showing low mortality

- **Steep Decline (Ages 50-100)**

  - Survival probability drops sharply indicating many deaths

- **Late Age (Ages 100+)**

  - Survival probability is near 0 indicating very few live past 100
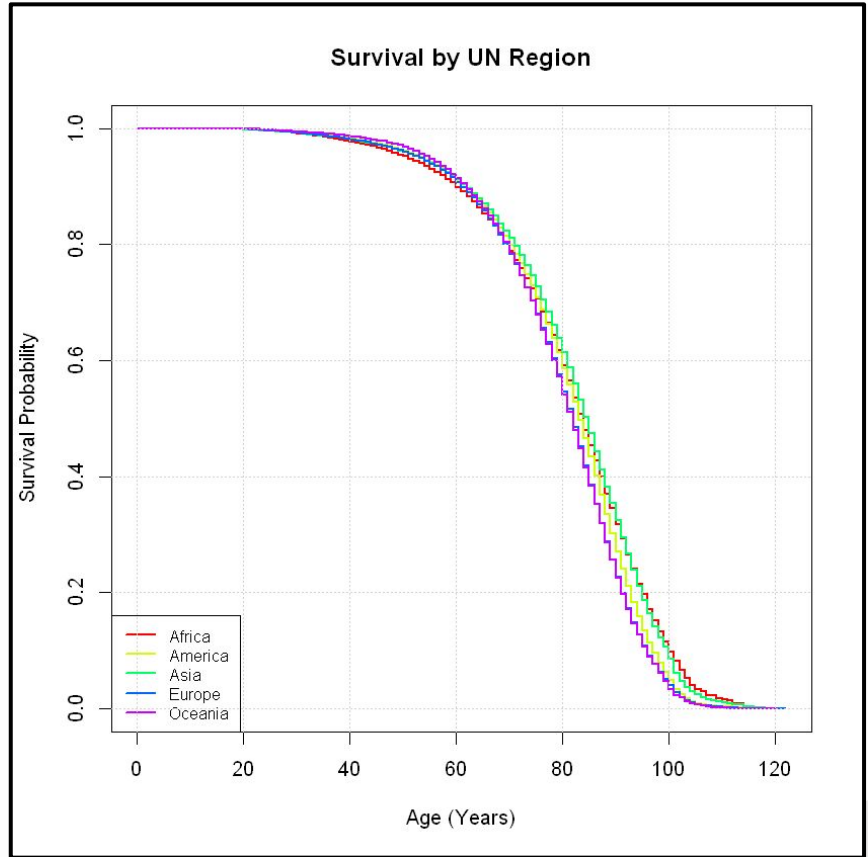

Overall Survival Curve of Notable Individuals

Gender KM Curve

- **Gender Differences**
  - Women survive longer between ages 50-100
  - Curves overlap around at very young and very old ages
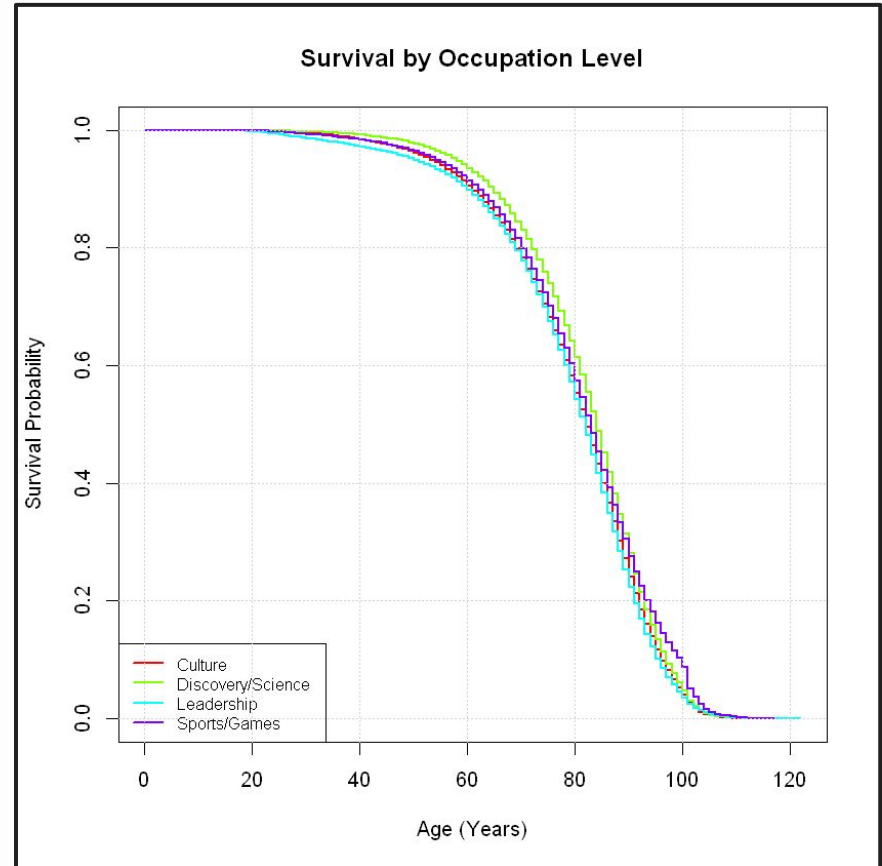
Survival by Gender

UN Region

- **Regional Differences**
  - Differences emerge between ages 50-90
  - The differences between regions are less pronounced than those between genders (overlapping curves)

Survival by UN Region

Occupation

- **Occupation Differences**
  - The differences between occupations are less pronounced than those between genders (overlapping curves)

Survival by Occupation Level

Question 2

How do these factors jointly affect mortality risk?

Multivariate Cox Model

Key Findings

- **Gender Effect**
  - Males have **~53% higher** hazard of death than females (at any given moment)

- **Region Effects**
  - **Africa:** ~6% lower hazard
  - **Asia:** ~13% lower hazard
  - **Europe:** ~11% higher hazard
  - **Oceania:** ~14% higher hazard

- **Occupation Effects**
  - **Science/Discovery:** ~18% lower hazard
  - **Sports/Games:** ~16% lower hazard
  - **Leadership:** No meaningful difference

Covariate	Hazard Ratio	95% CI
Male	1.53	(1.51 - 1.54)
Africa	0.94	(0.93 - 0.96)
Asia	0.87	(0.86 - 0.88)
Europe	1.11	(1.11-1.12)
Oceania	1.14	(1.12-1.15)
Science	0.82	(0.81-0.83)
Sports	0.84	(0.84-0.85)
Leadership	1.00	(0.99-1.01)

As compared to a baseline of Female, American, and "Culture" for occupation

Proportional Hazards Assumption Check

What We Tested

Cox Model Assumption:
- Does each variable affect mortality consistently across age?

What We Found

- No, effects change over time
- Gender, region, and occupation **do not impact mortality the same** at all ages
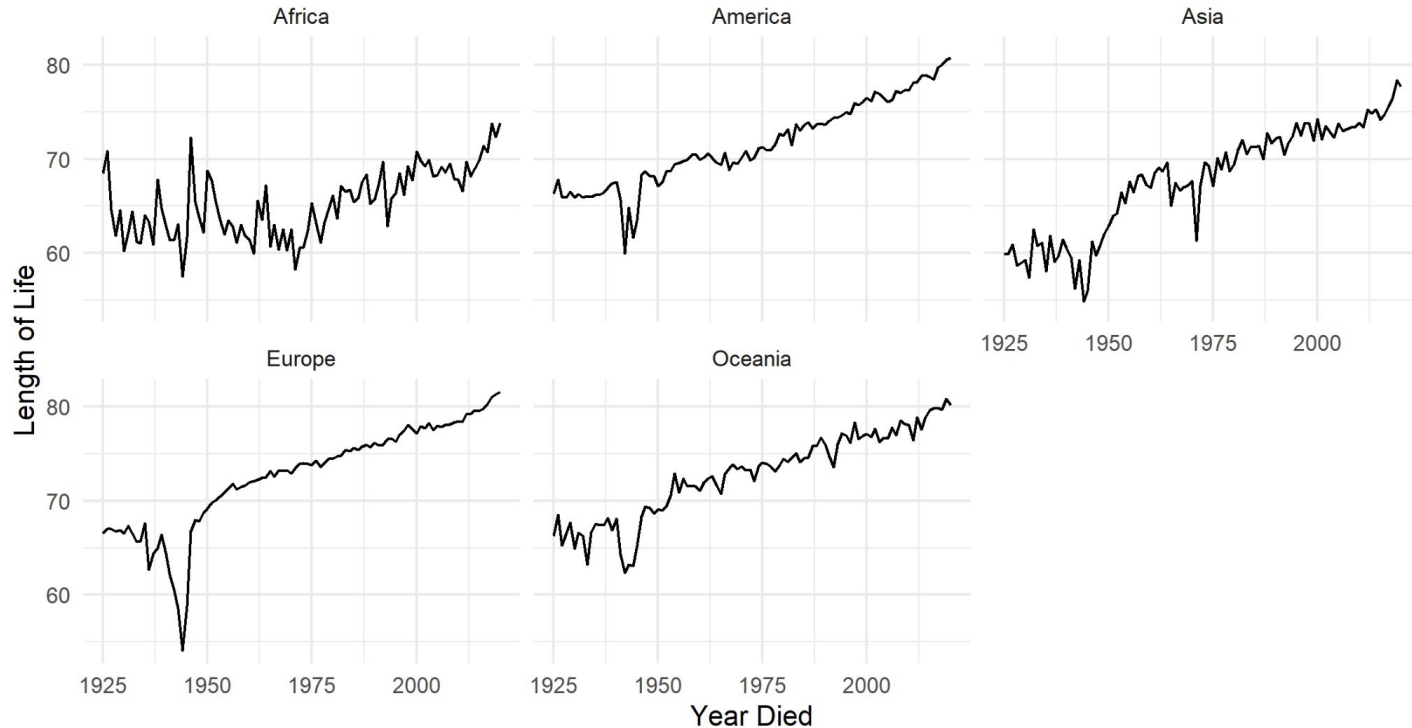
What This Means

- Hazard ratios reflect **overall lifetime patterns**, not age-specific effects

- This is expected in historical data: the influence of gender, region, and occupations have evolved and shifted over time

Question 3

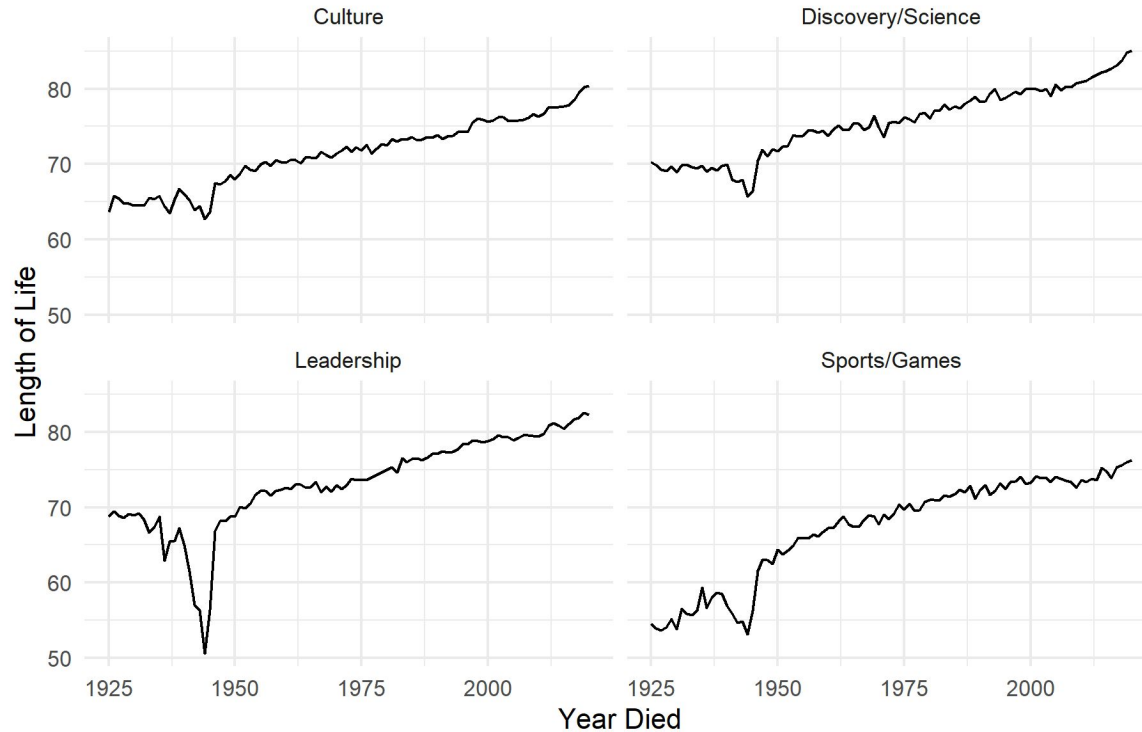How have major historical events, such as wars, impacted survival time?

Average Age of Death Over Time

Mean Lifespan by Year and Region

Average Age of Death Over Time

Mean Lifespan by Year and Occupation

Conclusion

Main Takeaways

Conclusion

Q1

- Visually, gender has the biggest effect over region and occupation for survival curves

Q2

- Cox PH model hazard ratios are dependent on time, where men had the highest hazard rate across any group
- And should be interpreted as average effects over a lifetime

Q3

- Wars had the biggest impact on survival. Lifespan has been increasing over time, but major global wars and famine decrease average life expectancy.

Limitations

Unknown Death Dates

Assumed those born >= 1925 with missing death dates are alive → potential lifespan overestimation

Bias

Data is primarily European and American

Broad Categories for Region and Occupation

Do not account for nuances in regions and occupations

Thank You!

Do you have any questions?

APPENDIX

Schoenfeld Residual Plots