MEASUREMENT ERROR 2

JAMES SAXON

To Do

- Refusals.
- Make the likelihood fit for NDI data.

Sanity Checks

(1) Are there false positives? Yes, note pre-1966 interviews in Figure 1.
(2) Zombies. Are any respondents not interviewed due to death and subsequently interviewed? See R Code 1 – none are found. Still, this may be self-fulfilling if the NLS doesn't try to interview dead people.
(3) Evolution of false positives over time. Look for men interviewed at dates later than the one on their matched death certificate Table 1.
(4) Certified Sawyers. Do we observe death certificates dated more than two years after a man has been not-interviewed due to death? Yes – apparently quite a lot, Table 2.
    - What is the rate if the NDI and SSA death years match? Lower – but it still ain't nothin. See column, 'NDI & SS.' And 9-digit SSN match, the last column (Full SS).
    - What is the interview type in 1990 for people who died before 1988? See Table 3. Did any windows (or their proxies) respond to the interview in 1990 for subjects supposed to have died after 1992? See R Code 2. There are ten such cases for the NDI dates.

R Code 1: Ghost check: interview after non-interview due to death. None are found.

```
cn <- colnames(d)
fni <- cn[grep("fNonInt", cn)][1:11]
```

Date: March 18, 2017.

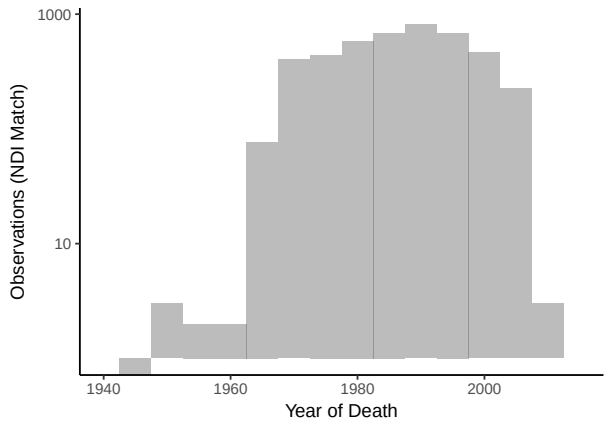


Figure 1. Death year by NDI. Survey started in 1966, so deaths before 1966 are errors of one flavor or another.

```
for (a in seq_along(1:10)) {

  fniA <- fni[a]
  fniB <- fni[a+1]

  zombies <- (d[[fniA]] == "DECEASED") & (d[[fniB]] != "DECEASED")
  if (sum(zombies)) {
    print(sprintf("Zombies in %s to %s!!!", fniA, fniB))
    print(d[zombies,"IDENTIFICATION CODE, 66"])
  }
}
```

	NDI	SS	Cert	Int
1967	10	11	6	9
1968	10	14	7	8
1969	6	9	5	6
1971	5	11	3	6
1973	5	7	4	6
1975	1	6	2	5
1976	0	5	2	4
1978	1	5	0	2
1980	0	2	0	0
1981	2	4	0	1
1983	3	10	1	2

TABLE 1. Interviewed after death by match: ghosts.

	NDI	SS	NDI & SS	Int	Cert	Full SS
1967	1	0	0	1	0	0
1968	3	1	1	2	0	0
1969	8	3	2	3	0	0
1971	12	4	4	5	1	1
1973	15	4	5	5	1	1
1975	21	6	6	5	1	1
1976	26	8	7	4	2	2
1978	28	9	7	4	2	2
1980	32	11	9	5	3	3
1981	34	12	10	6	3	3
1983	31	12	10	4	2	2

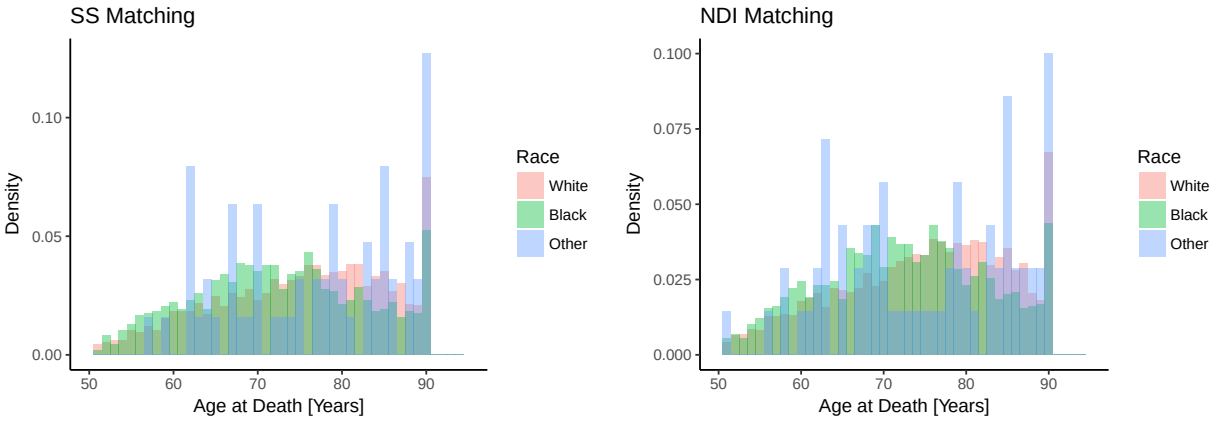
TABLE 2. Death certificate dated more than two years after non-interview due to death: Tom Sawyers.

R Code 2: Check for respondents noted to have died after 1992 but whose widow (or their proxy) responded in 1990.

```
colSums(d[d$fResp_1990 %in% c("Widow", "Widow Proxy"),
        c("ndiDeathYear","ssDeathYear","certDeathYear","int90DeathYear")] > 1992,
        na.rm = T)
```

	Non-interview	Respondent	Widow	Widow Proxy
NDI	436	3	1057	696
SS	359	1	983	626
Cert	299	0	866	554
Int90	433	0	1073	724

TABLE 3. Interview type in 1990, for repondents recorded to have died before 1988.



(A) SSA (B) NDI

FIGURE 2. Age at death by race from the SSA and NDI matches.

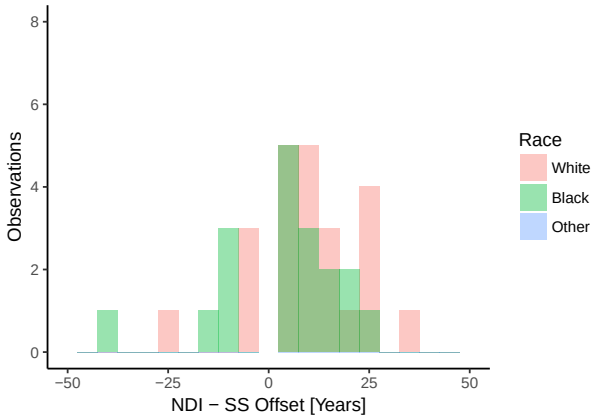
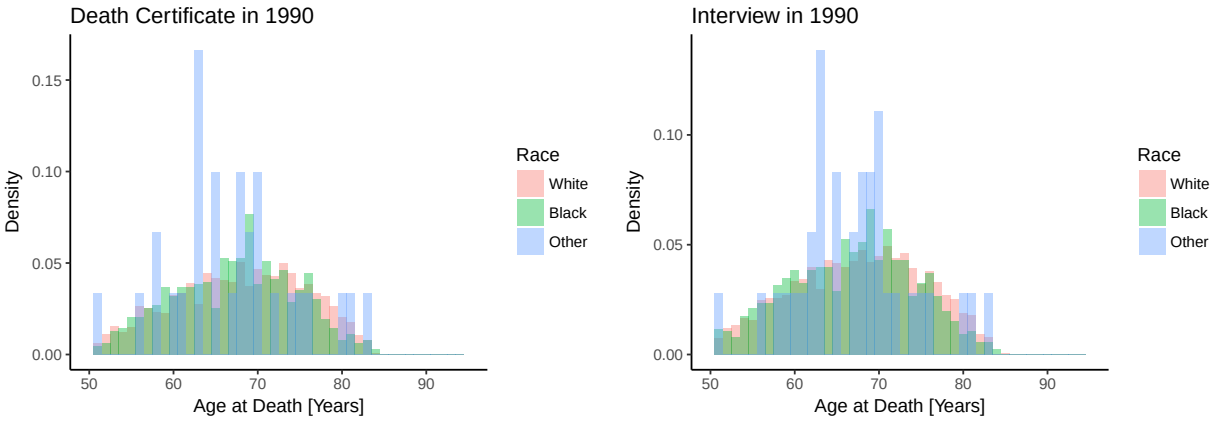


FIGURE 3. Difference in ages between SSA and NDI match.

```
##   ndiDeathYear    ssDeathYear  certDeathYear int90DeathYear
##             10              0              0              0
```



(A) Interview (B) Certificate

FIGURE 4. Age at death by race from 1990 interview and death certificate.

Covariates

Predictive measures: life expectancy for conditions. Also list these for ghosts.

- Wealth: `TOTAL FAMILY INCOME, YY (SUMMATION)`, `INCOME FROM WAGES & SALARY IN YY - R` (66, 67, 69, 71, 73, 75, 76, 78, 80)
  - Use year for constant age, then adjust wage to year X??
  - Or by quantiles? Still doesn't capture different respondents at different points in wage trajectory.
- Smoking: 1990 R (current R0627700 and R0628100) and W (R0719600). Weird selection here, since R xor W must live to 1990. Could do years or number of cigarettes smoked instead?
- Heavy alcohol: frequency in lifetime, recorded in 1990 by R (any R0628600, days R0628700, quantity R0628800) and W (any R0720400, days R0720500, quantity R0720600). Calling heavy alcohol (totally arbitrarily) $\geqslant 3$ days a week AND $\geqslant 5$ drinks per drinking day.
- BMI in 73 and 90 (W = R0258700, H = R0258600). BMI = $\text{kg}/\text{m}^2 = 703.27 \times \text{lbs}/\text{in}^2$.
- Ethnicity by foreign language spoken at home (R0228000): 0 None, 1 Spanish, 2 Germanic, 3 Other Romance, 4 Slavic, 5 Other.
- Which health problems are predictive? Do both "conditions experienced as a problem," and "health limits X." The annoying thing about these variables is that it's the universe, and not the responses themselves, that give the first-order flag. So I am comparing "yes I experience it as a (manageable) problem" to "-4, no it's not registering" as opposed to "not manageable" v. "manageable."

Likelihood from NDI Data

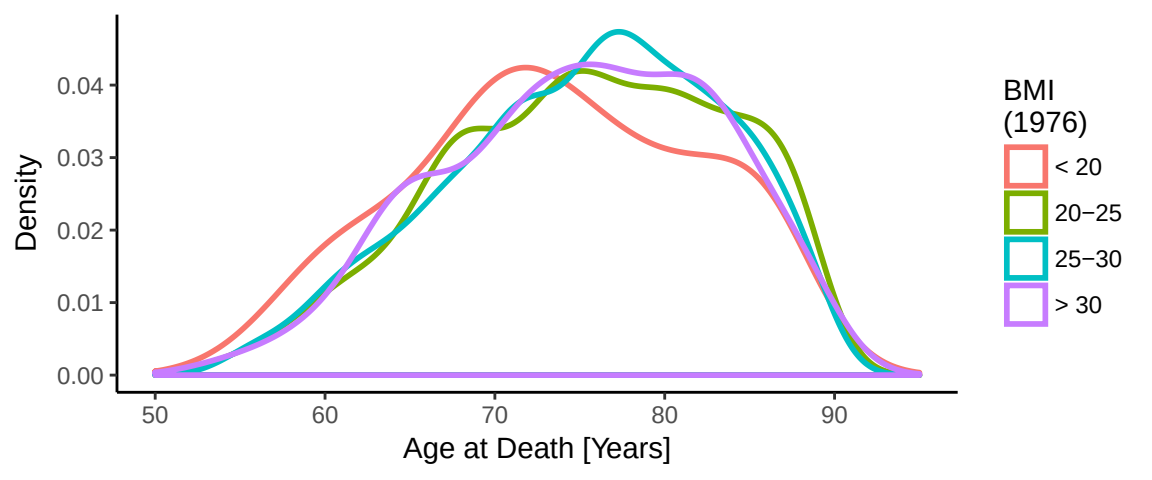
In the original paper we expressed the componenent for a single person i dying at age a_i with terminal/maximum observations A_T, by:

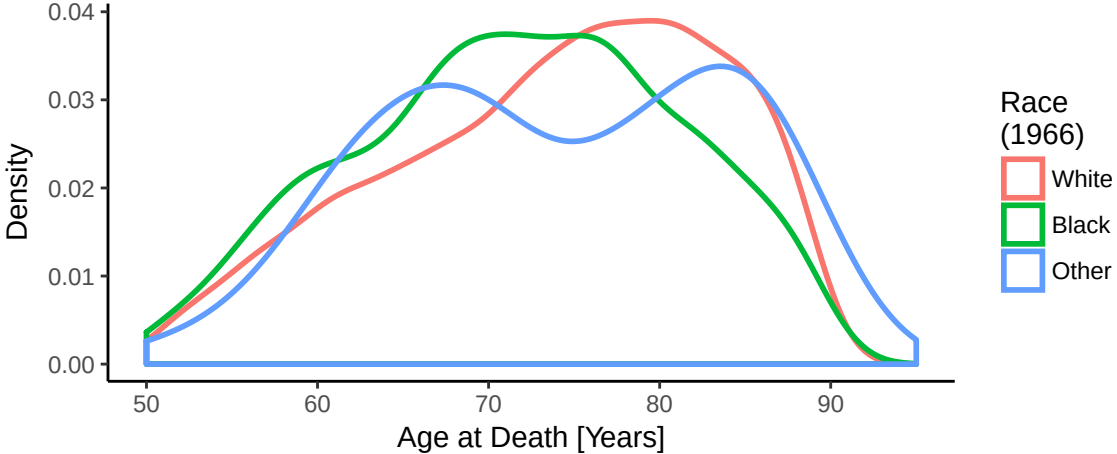
$$\mathcal{L}_i = \begin{cases} \log h(a_i) + \sum_{a=1}^{a_i-1} \log(1-h(a_i)) & a_i < A_T \\ \log \sum_{a=1}^{A_T} \log(1-h(a_i)) & a_i \geqslant A_T \end{cases}$$

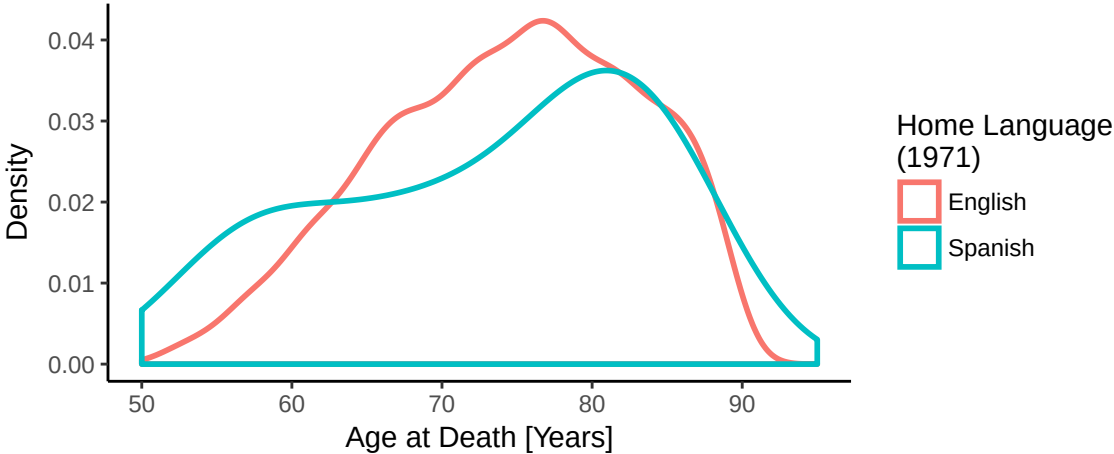
Where one could either parameterize over all hazards, $h(a) = e^{-\lambda+\beta a/A_T}$ or just let the per age hazards float, $h(a) = h_a$. And then the total was the sum over subjects, $\sum_i \mathcal{L}_i$.

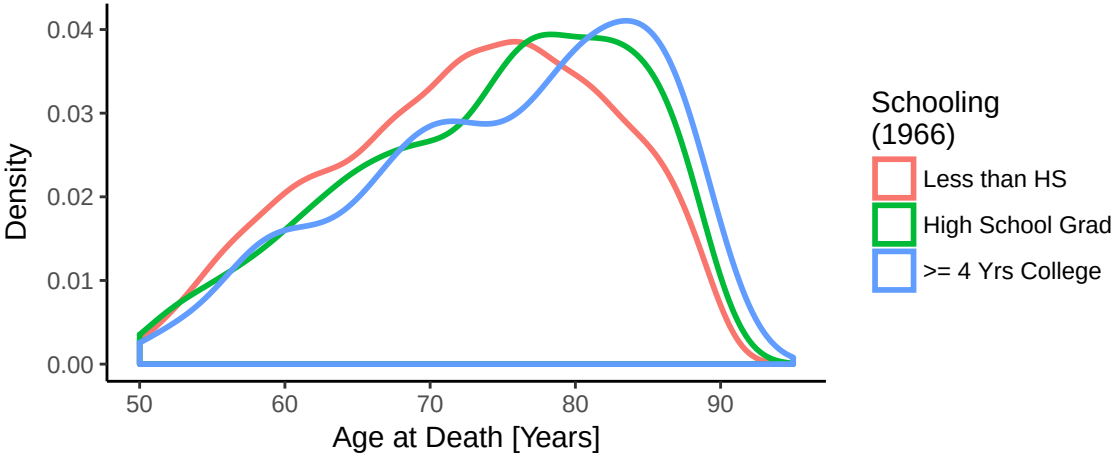
Let's think about the NDI data:

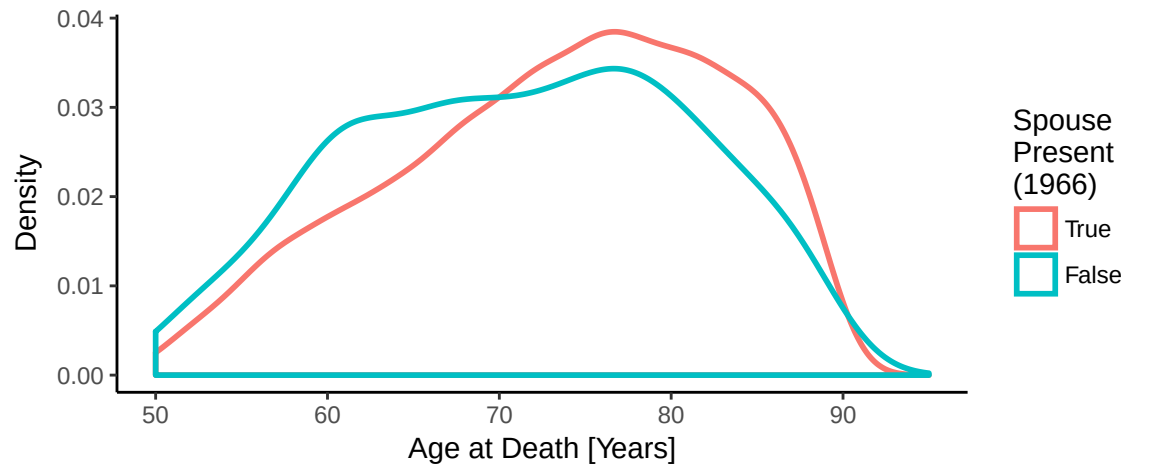
- What do the -5 (non-interview), -4 (valid skip) and -2 (don't know) mean in this context?
- N.B. the sharp spikes at the top-coded ages of 90 (see Figure 2). These presumably consist of (a) failure to match and mismatch, (b) people who died at 90 and (c) people who died at > 90.
- The thing that irks me is that as the likelihood is written, it feels like we're fitting a string of observations when we actually just have one (no match or match age). $\Rightarrow$ Read papers on this from Dan. Is it just that the hazard is 0 if you're dead, and $\log(1) = 0$ so there's no contribution to the likelihood?

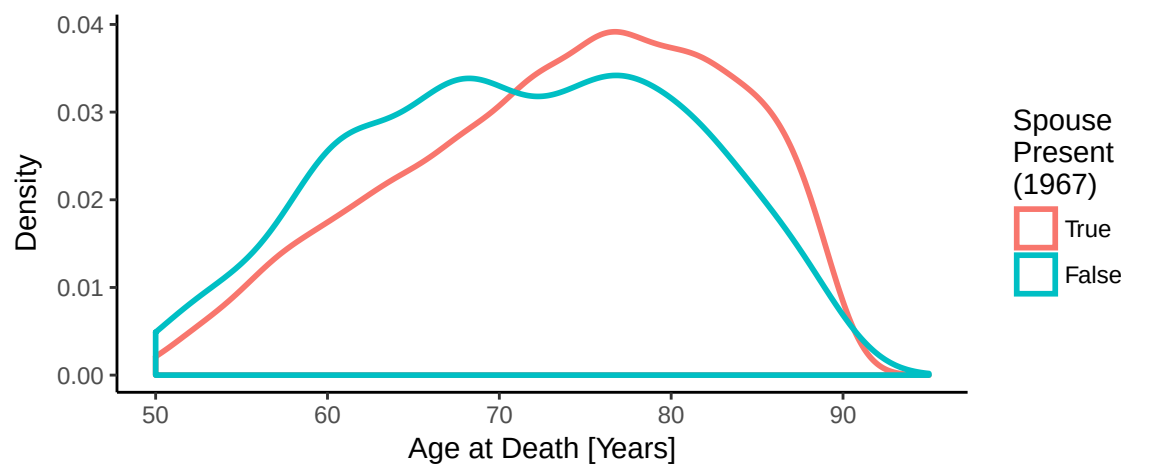

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
BMI
(1976)
< 20
20–25
25–30
> 30

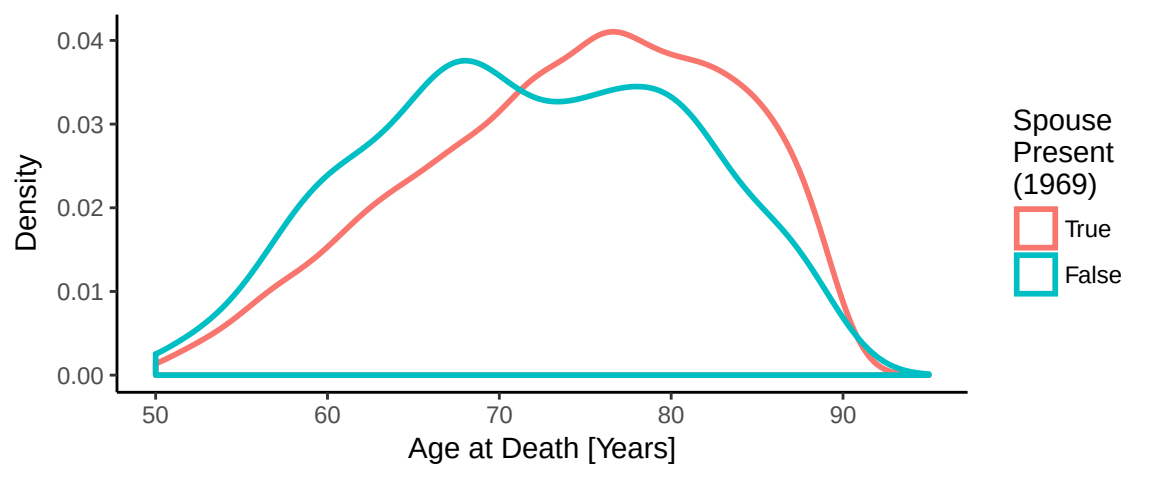

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Race
(1966)
White
Black
Other

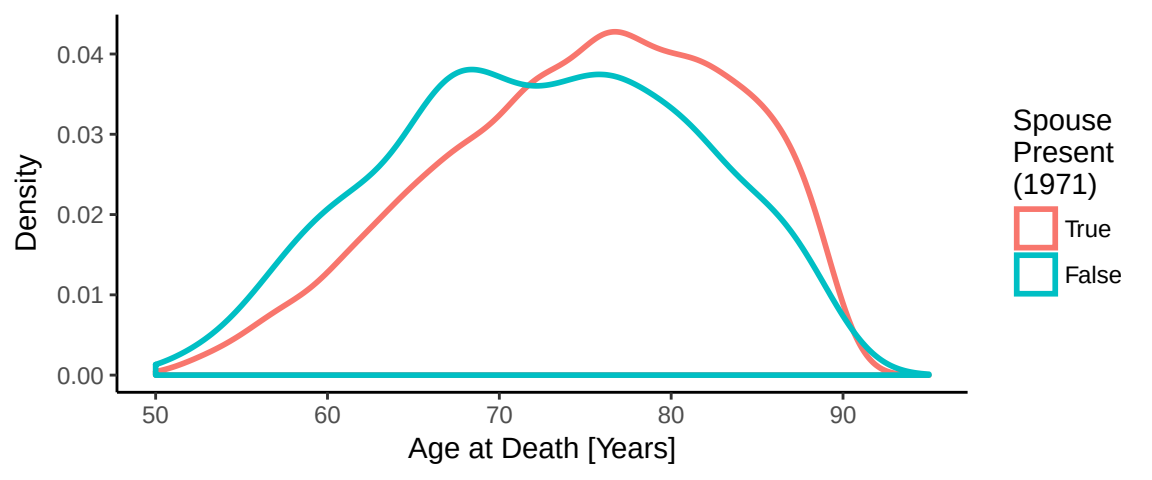

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Home Language
(1971)
English
Spanish

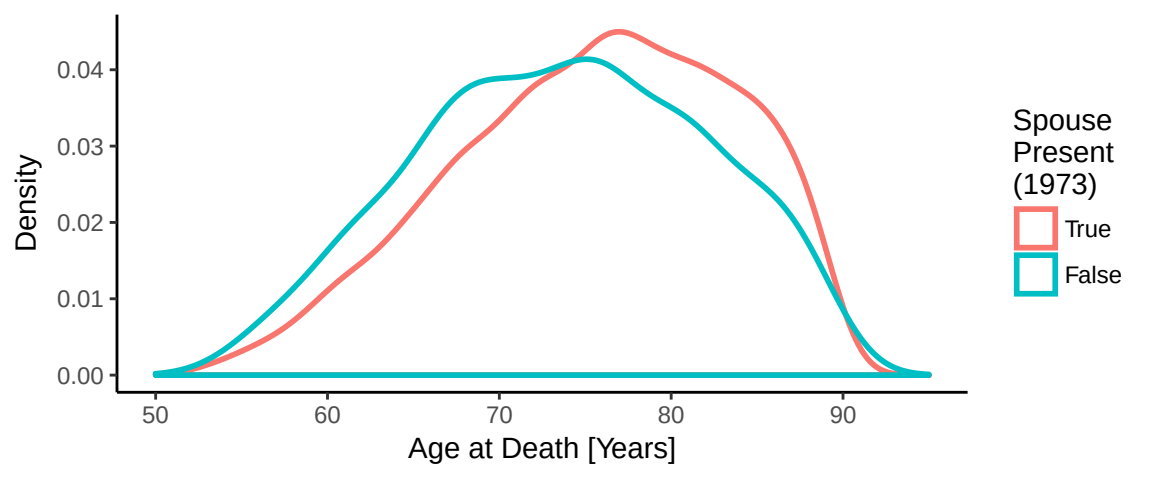

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Schooling
(1966)
Less than HS
High School Grad
>= 4 Yrs College

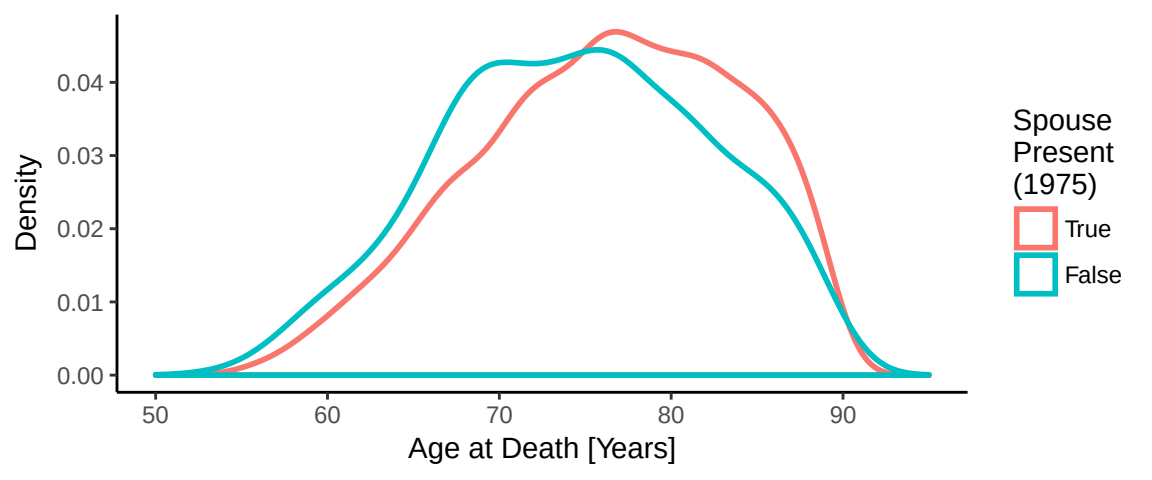

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1966)
True
False

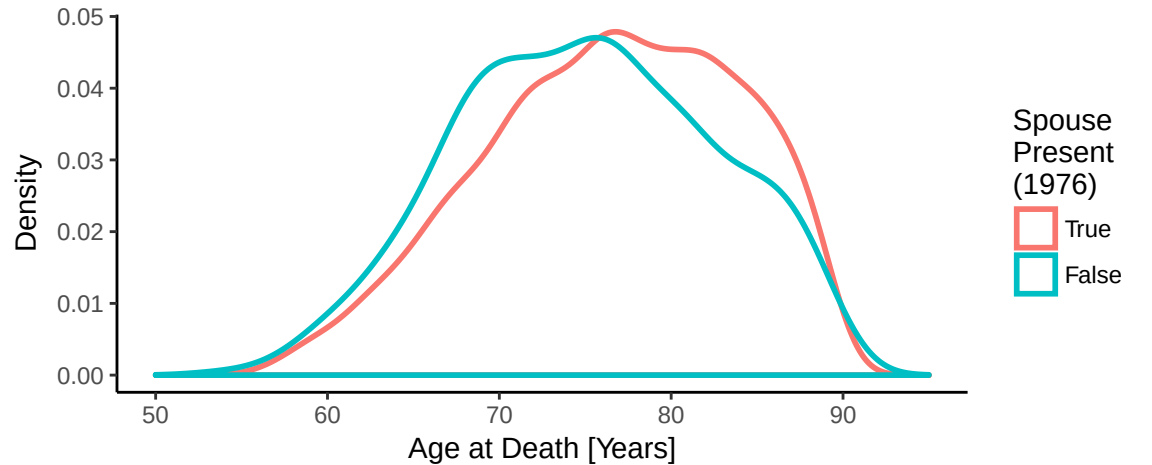

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1967)
True
False

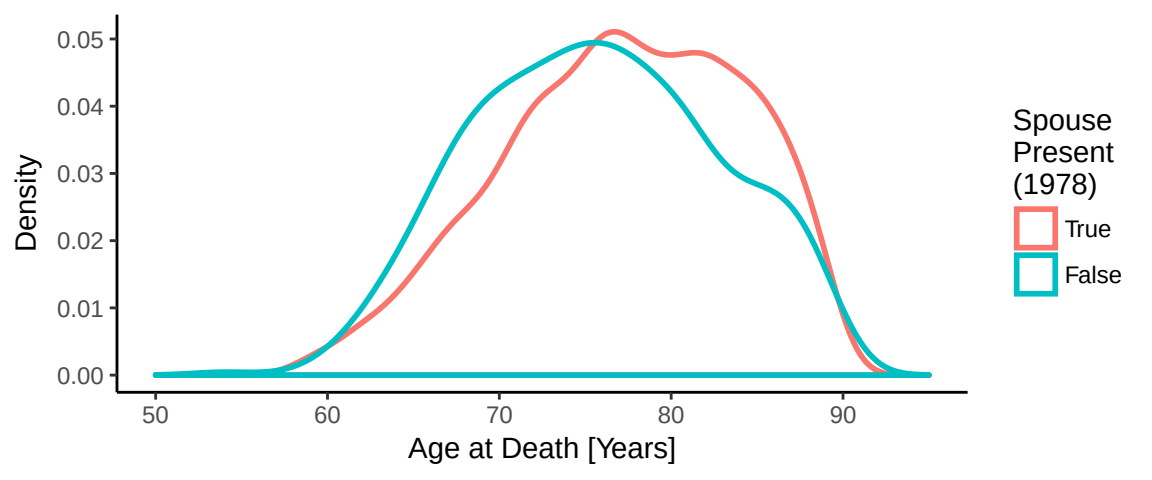

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1969)
True
False

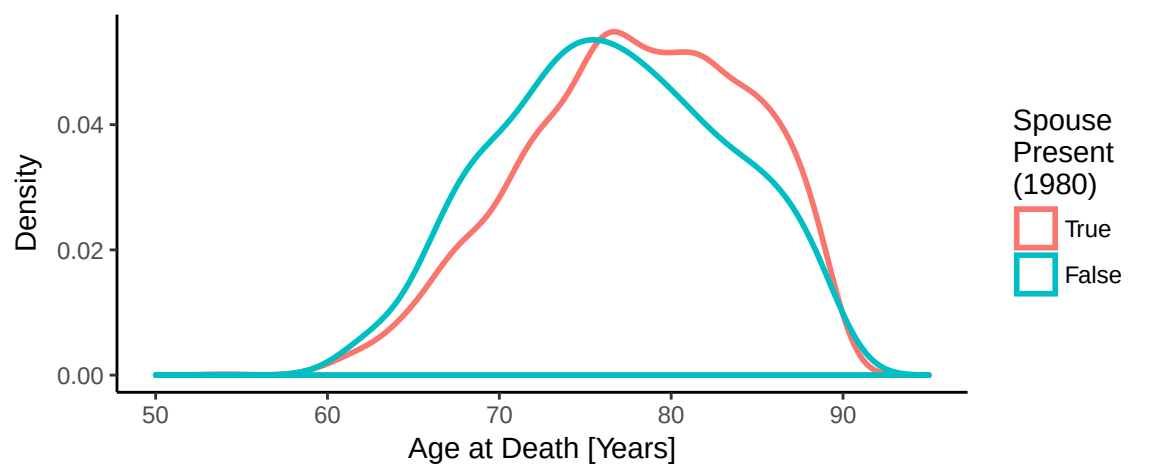

0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1971)
True
False

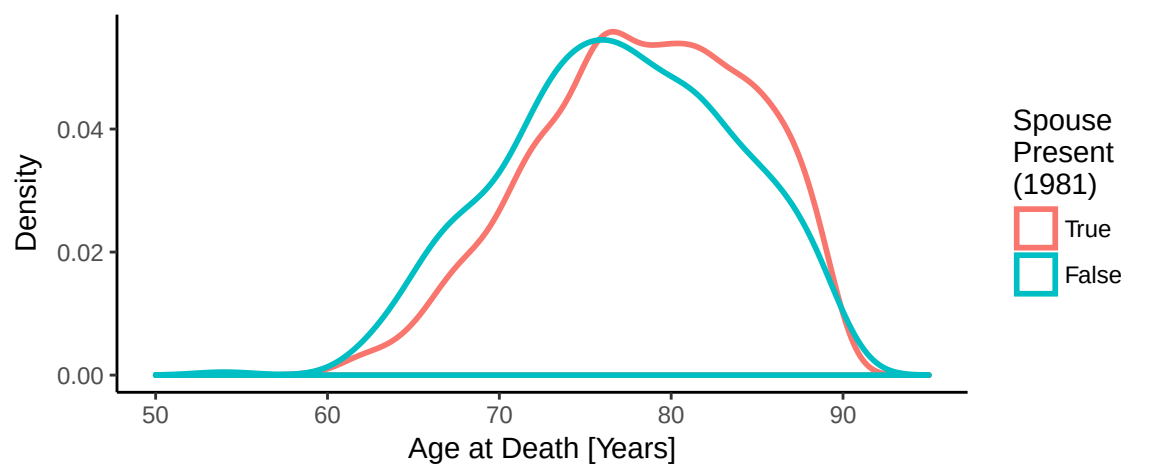

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Spouse Present (1973)
True
False

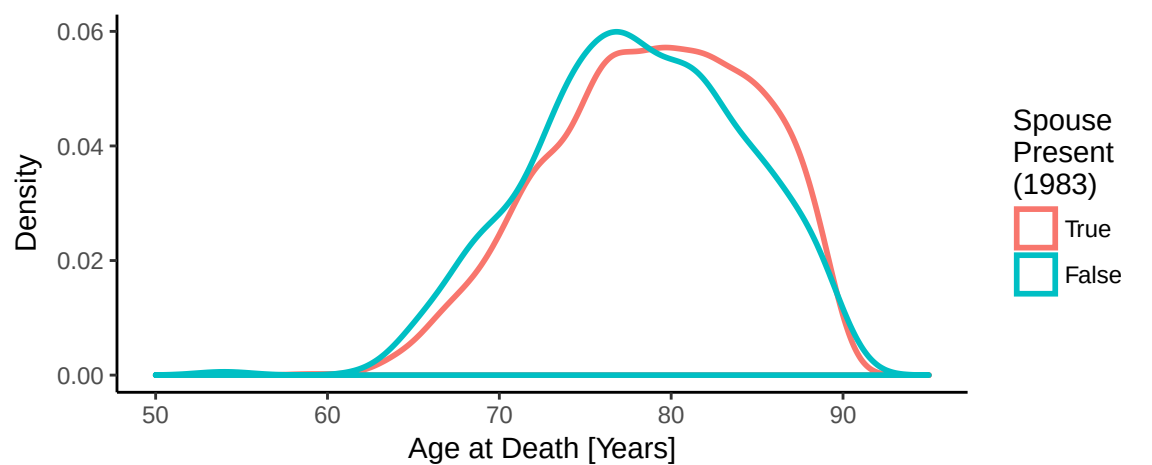

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Spouse Present (1975)
True
False

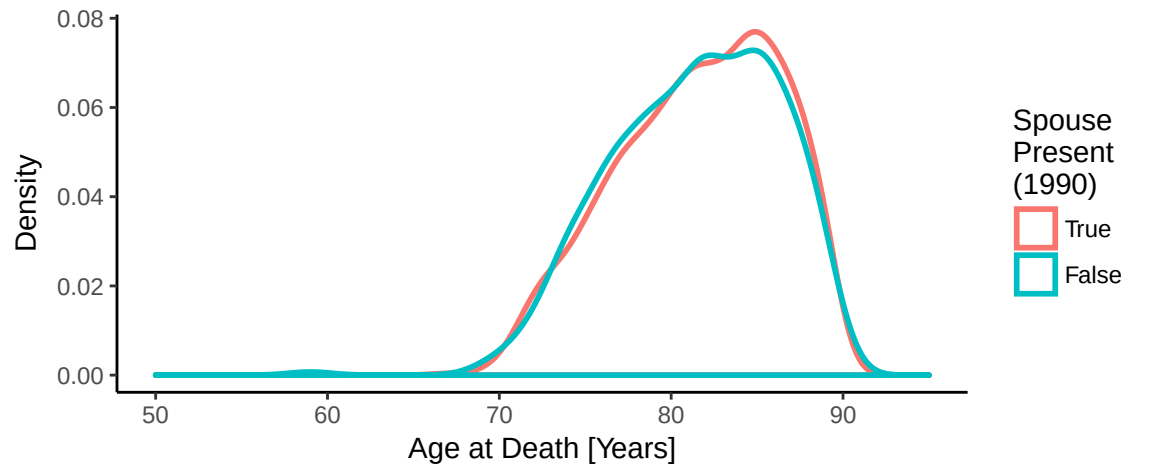

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1976)
True
False

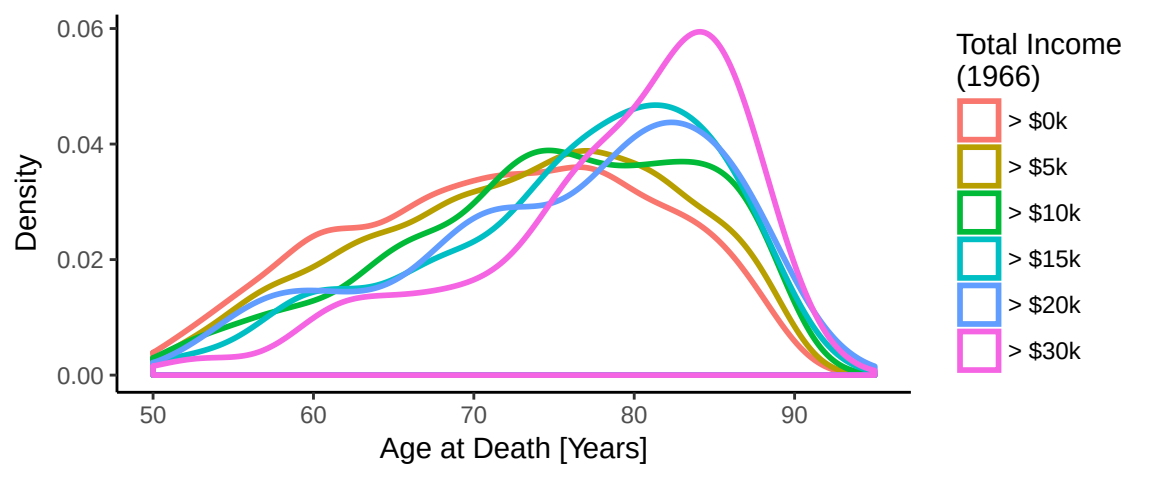

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1978)
True
False

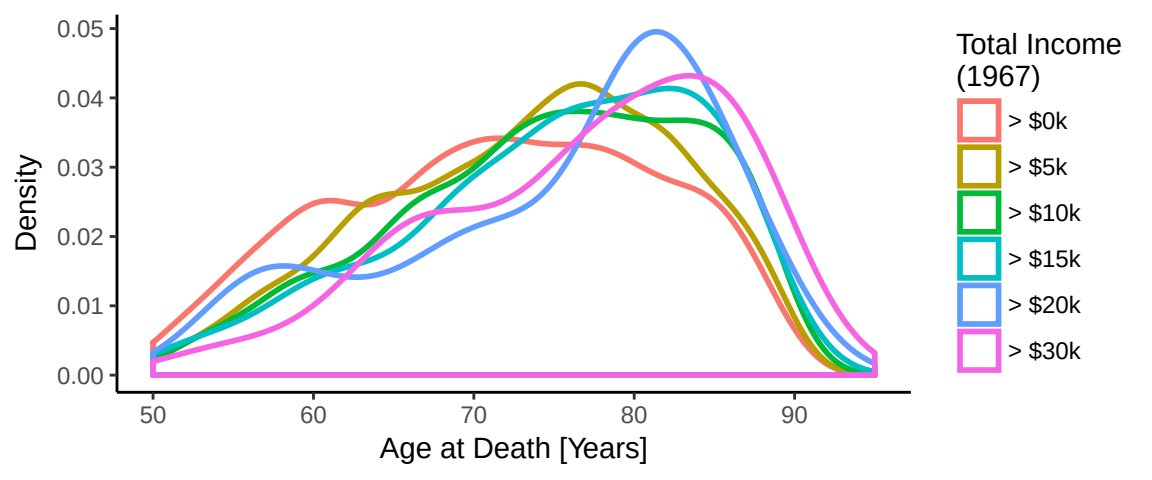

Density
0.04
0.02
0.00
50
60
70
80
90
Age at Death [Years]
Spouse Present (1980)
True
False

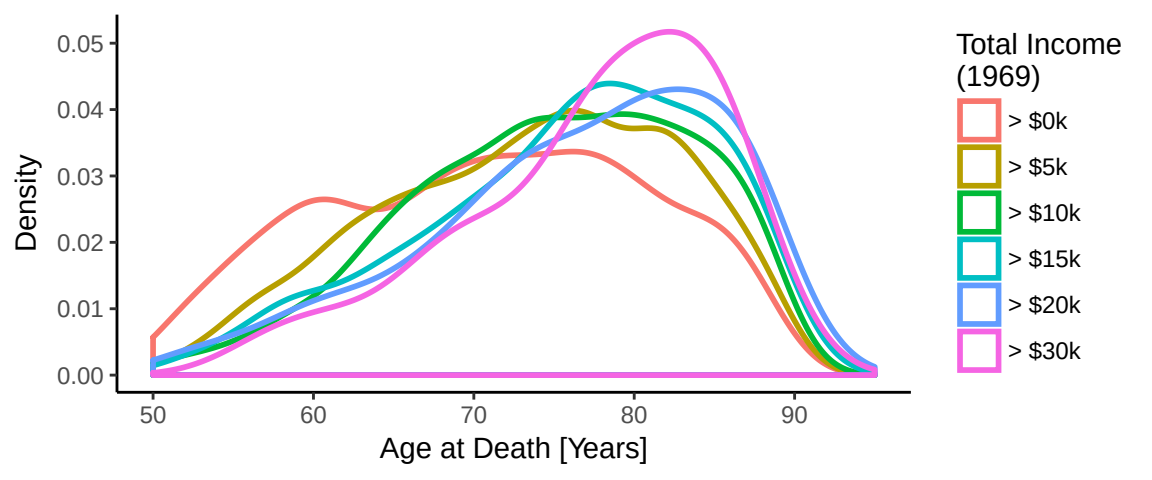

Density
0.04
0.02
0.00
50
60
70
80
90
Age at Death [Years]
Spouse Present (1981)
True
False

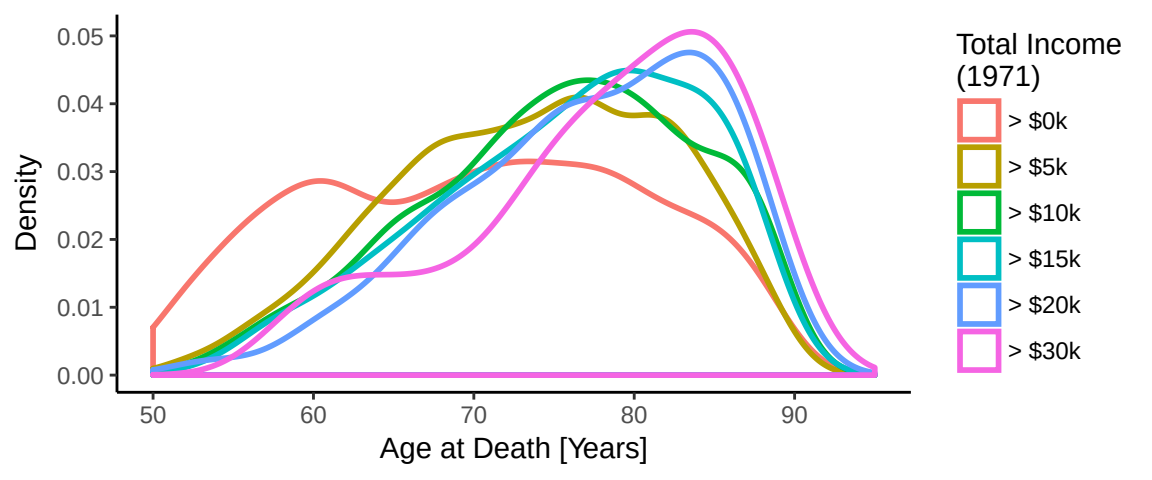

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1983)
True
False

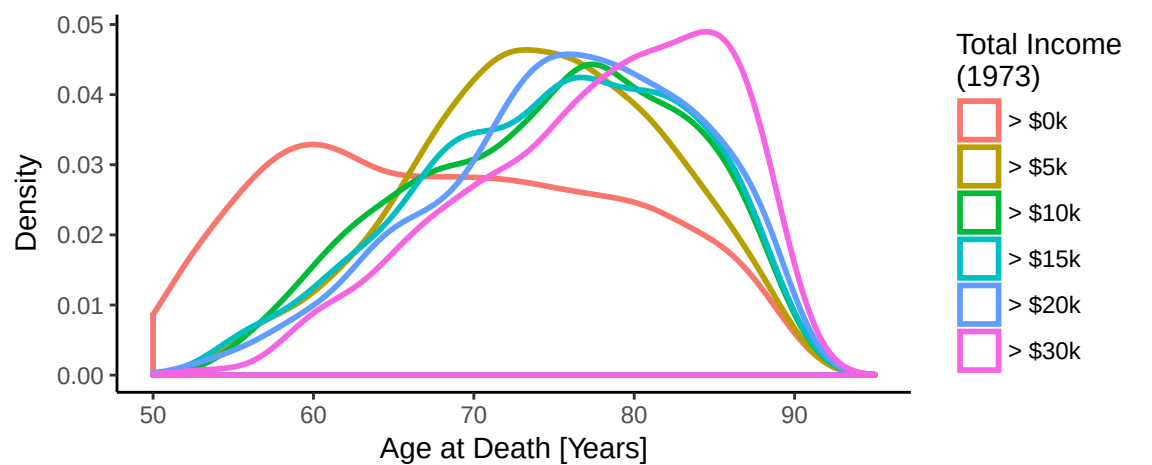

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Spouse Present (1990)
True
False

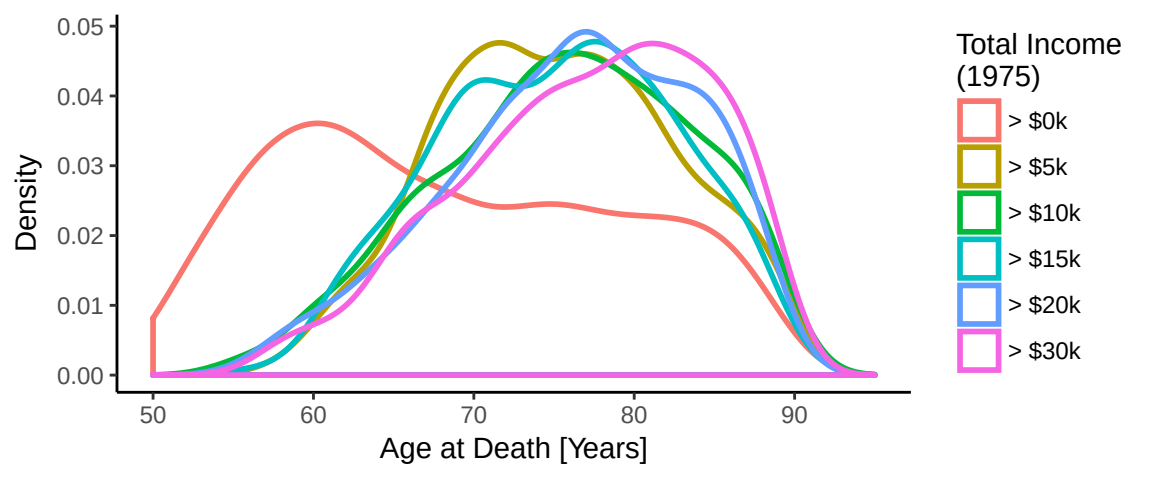

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1966)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

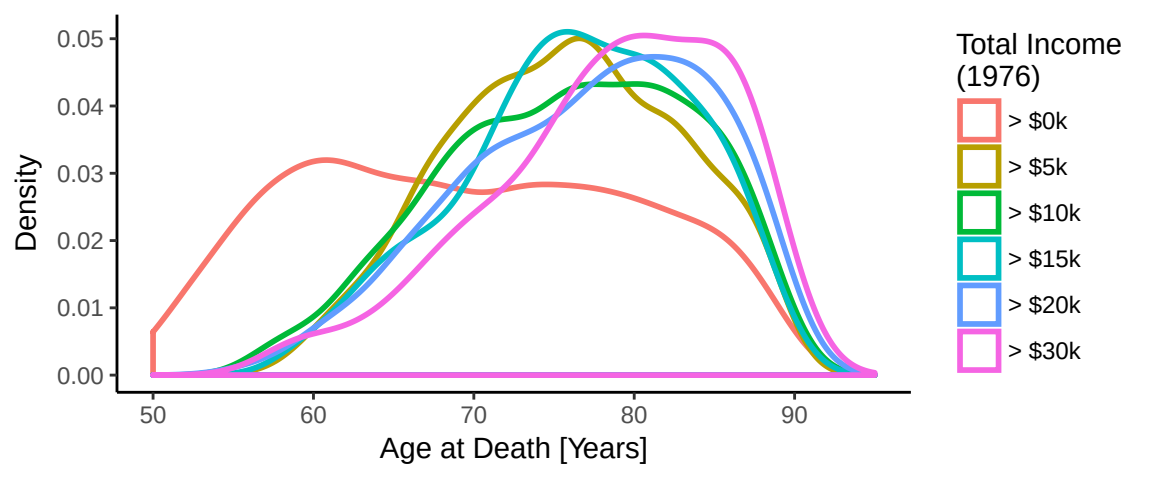

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1967)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

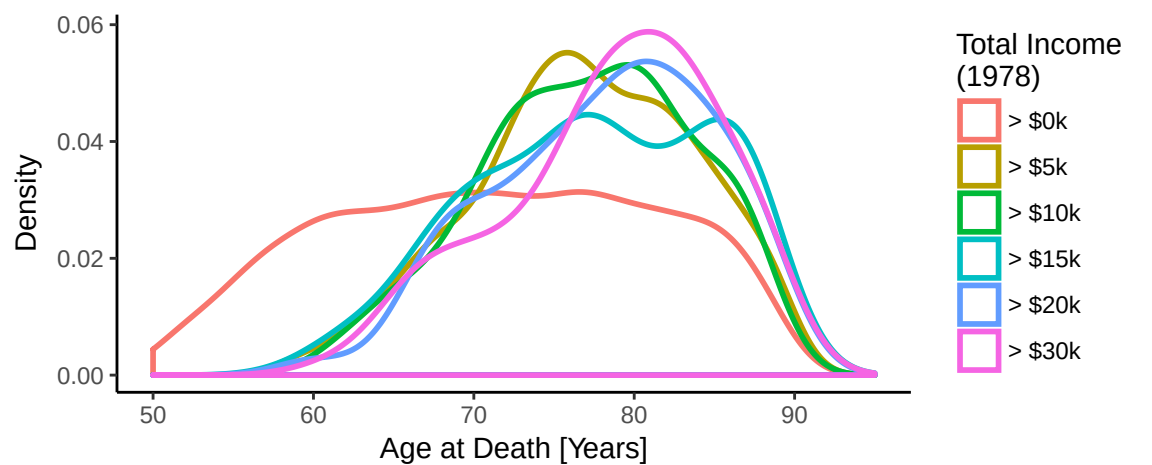

Density
0.05
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Total Income (1969)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

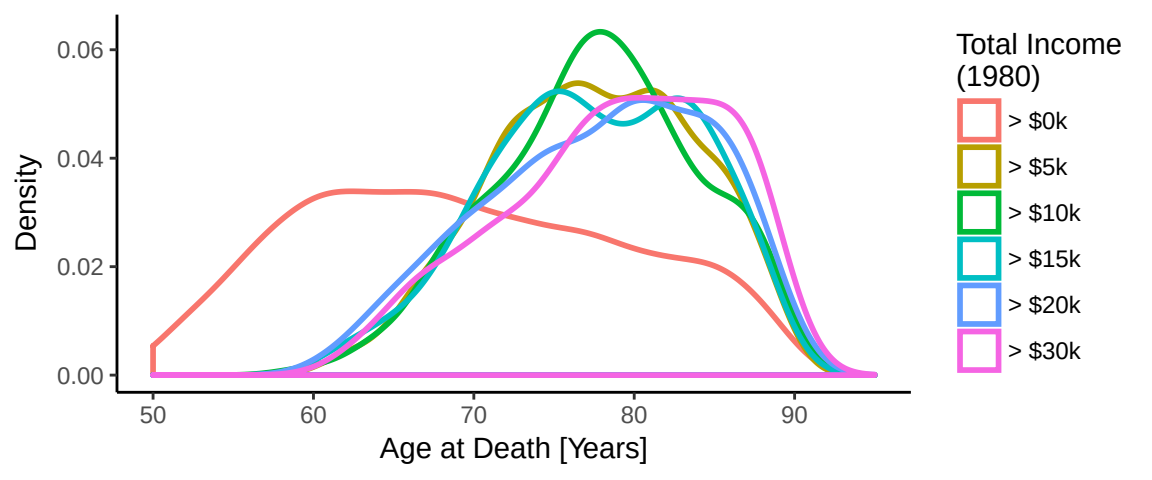

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1971)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

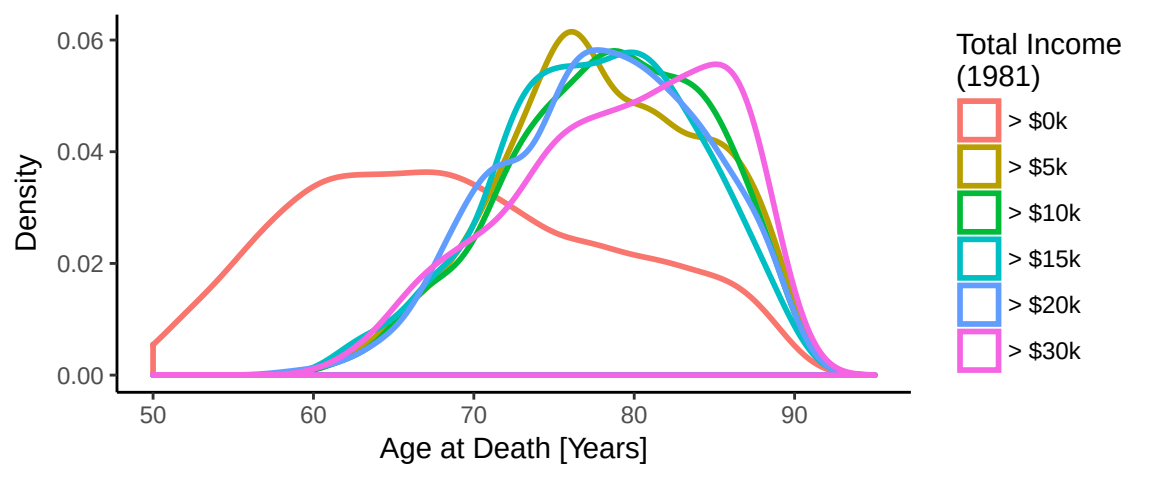

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1973)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

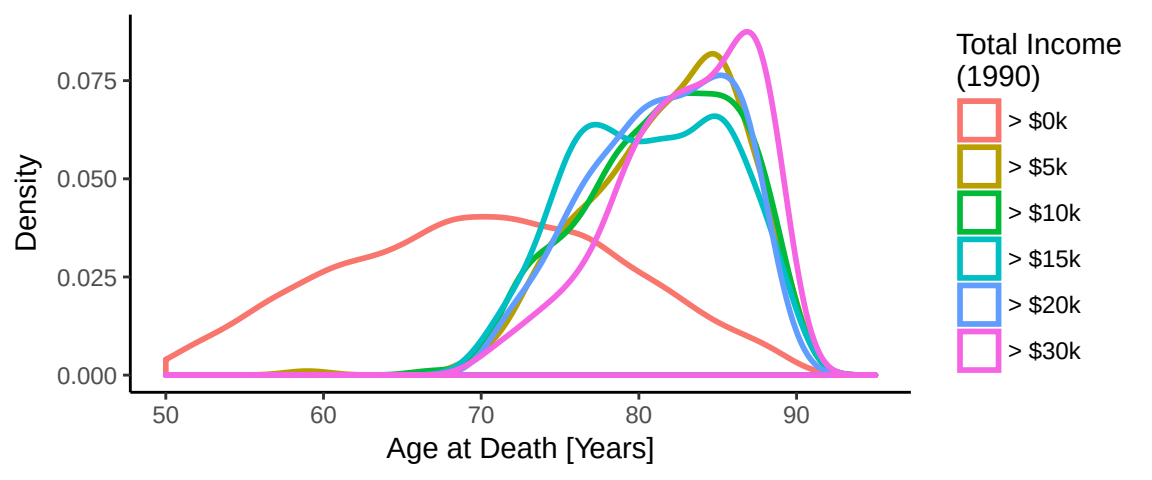

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1975)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

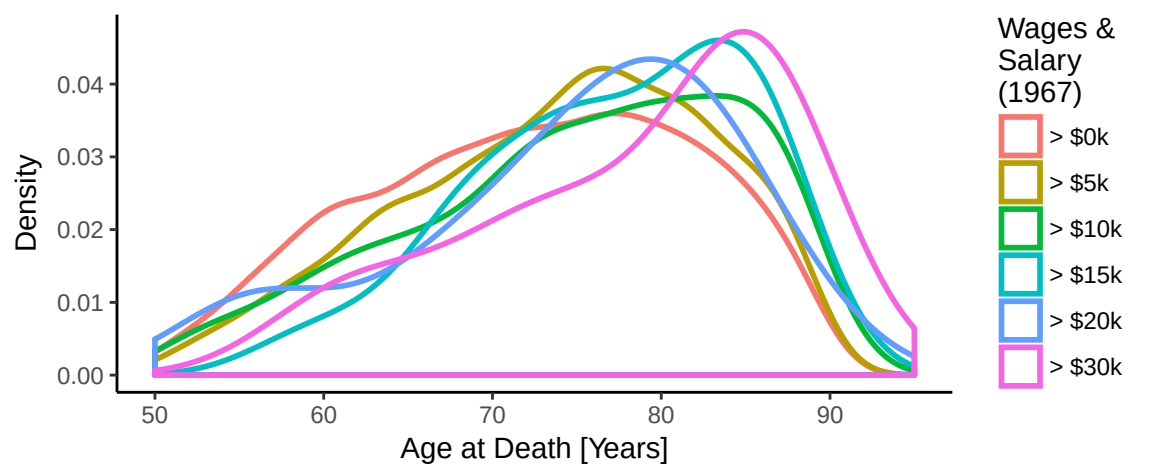

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1976)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

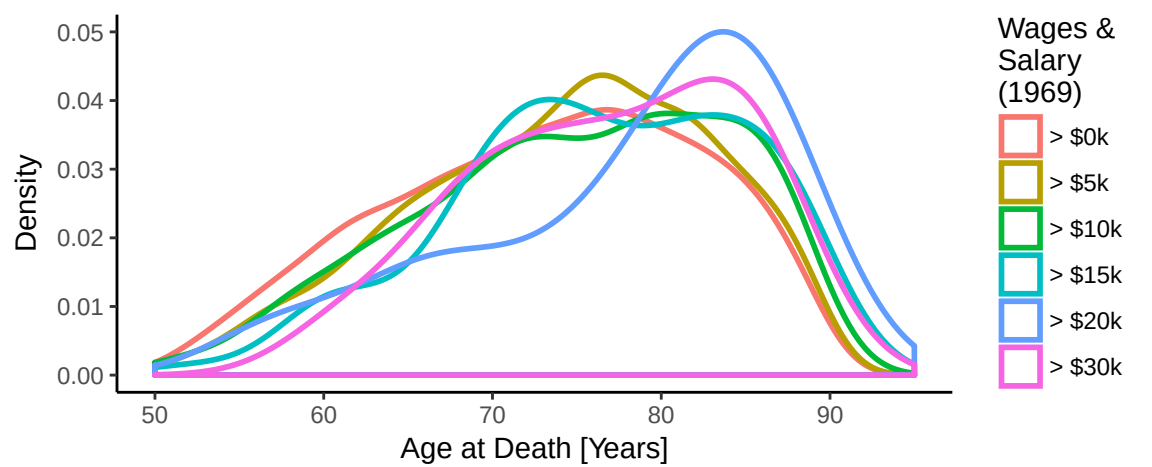

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income
(1978)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

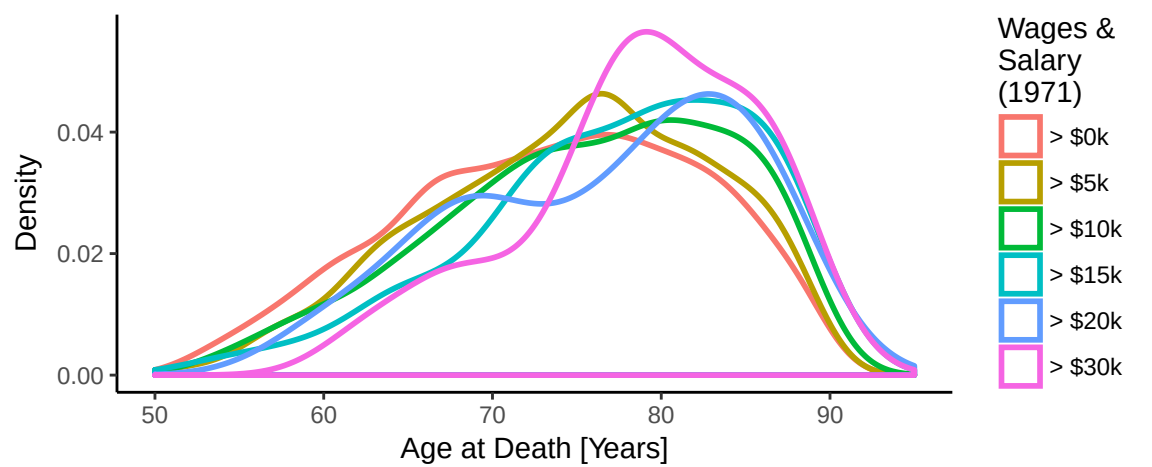

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1980)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

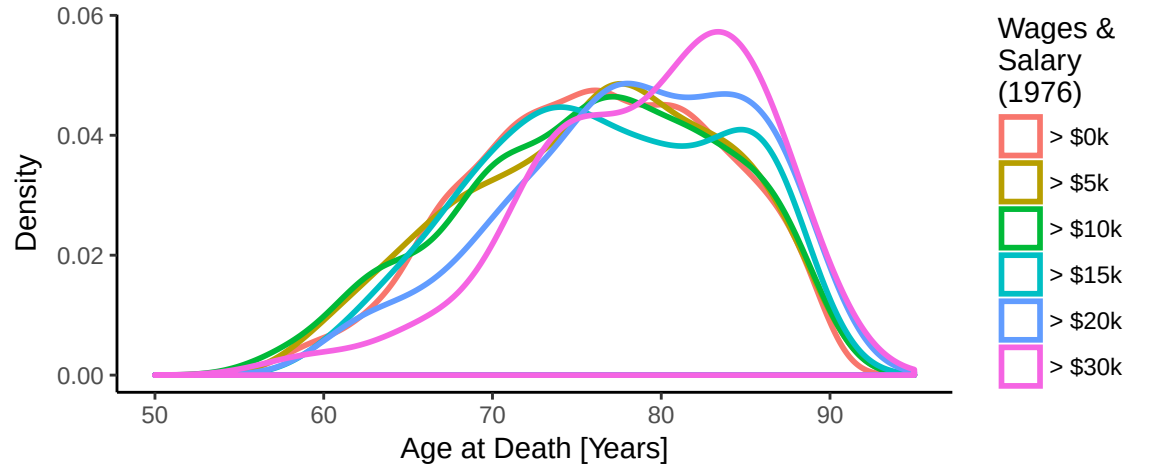

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Total Income (1981)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k


Density
0.075
0.050
0.025
0.000
50
60
70
80
90
Age at Death [Years]
Total Income (1990)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

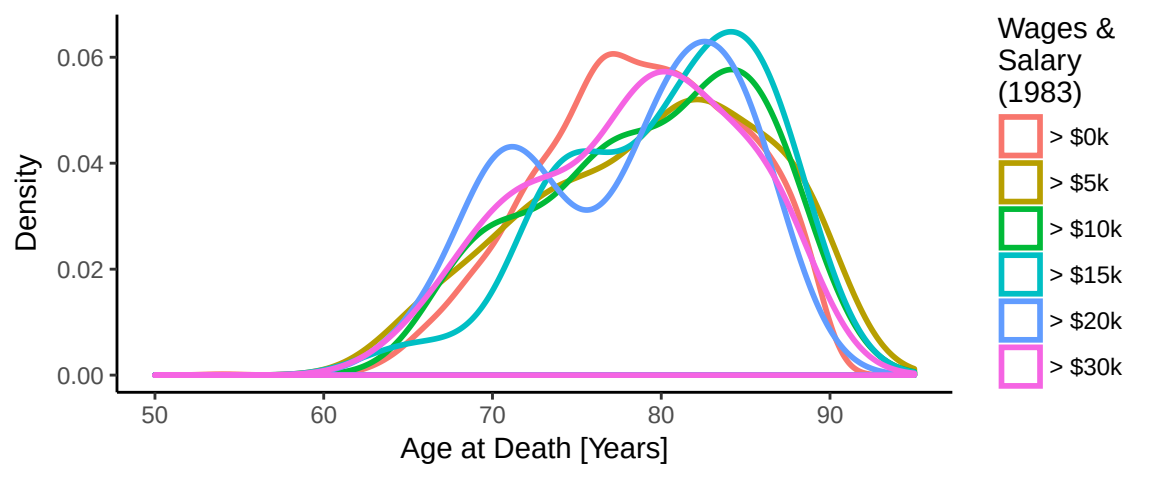

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1967)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

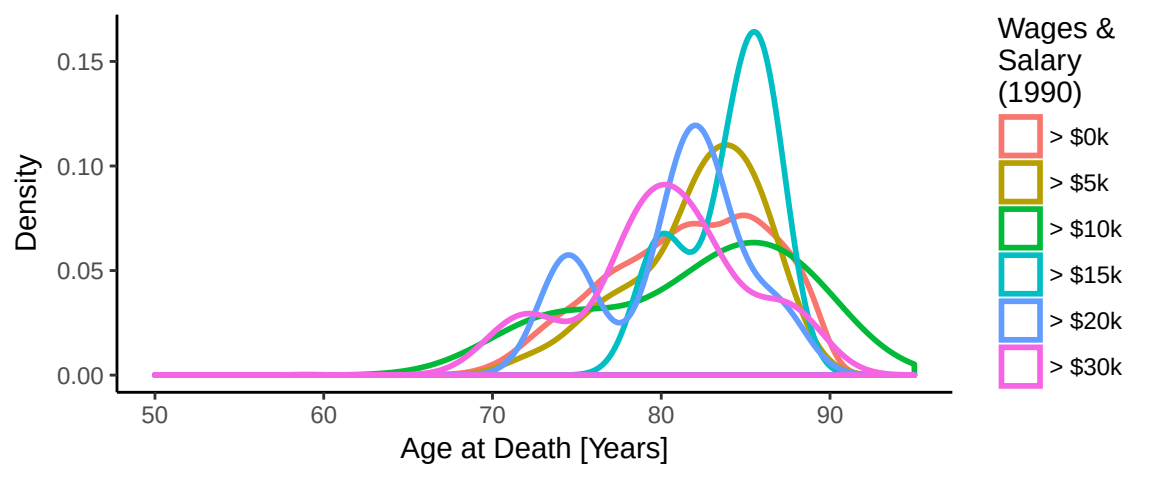

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1969)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

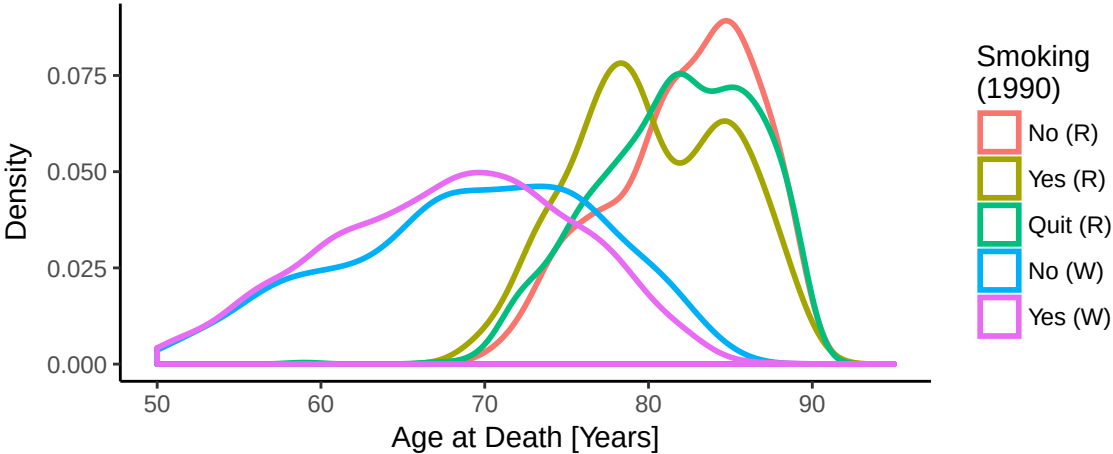

Density
0.04
0.02
0.00
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1971)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

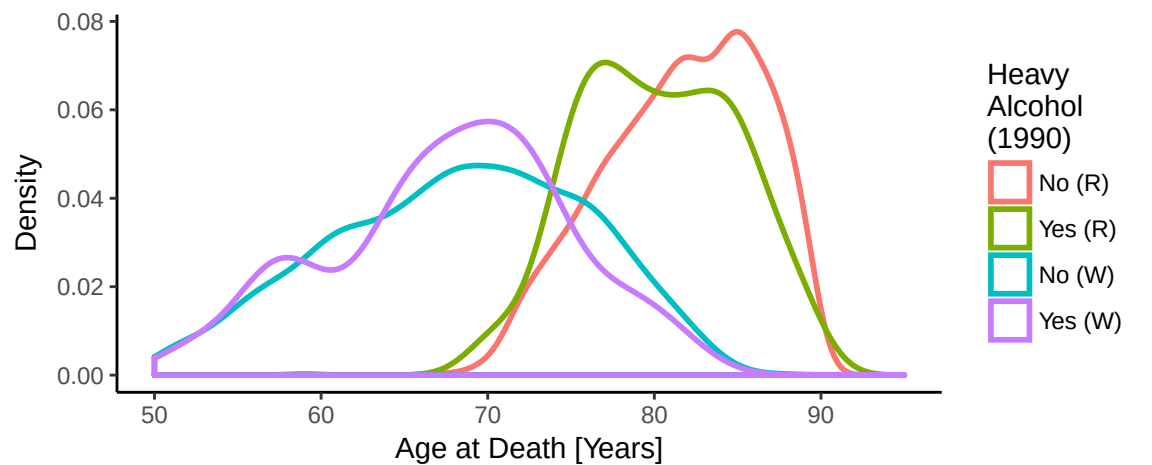

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1976)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

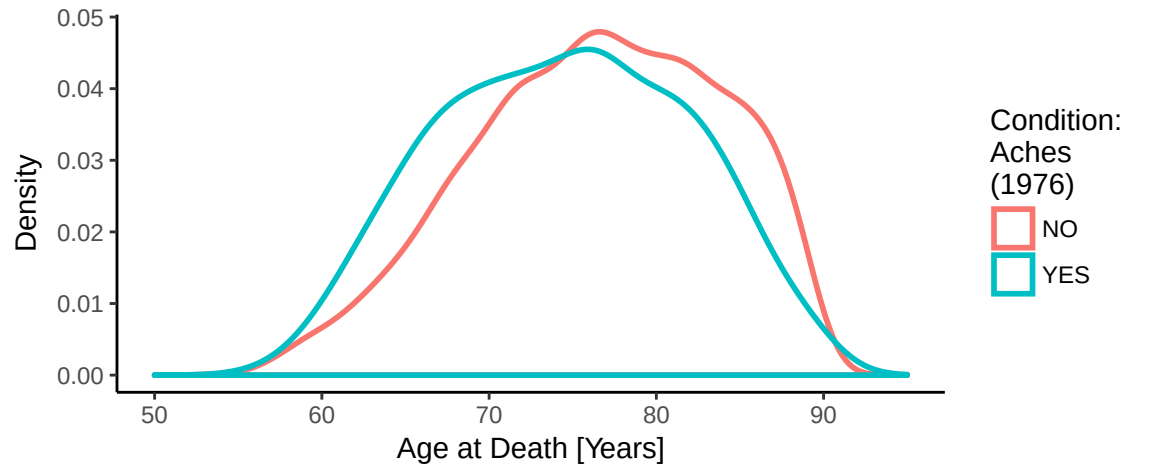

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1981)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

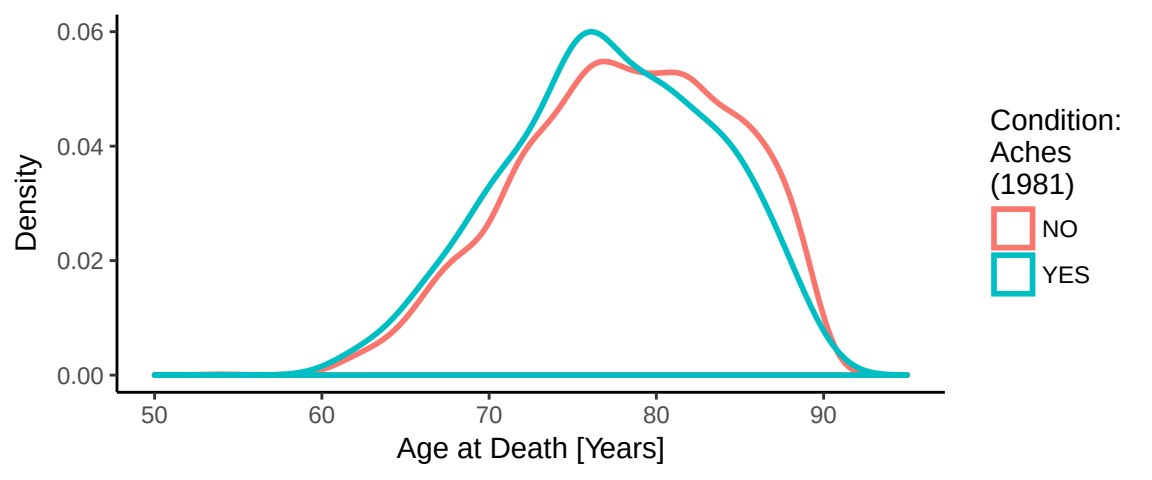

Density
0.06
0.04
0.02
0.00
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1983)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

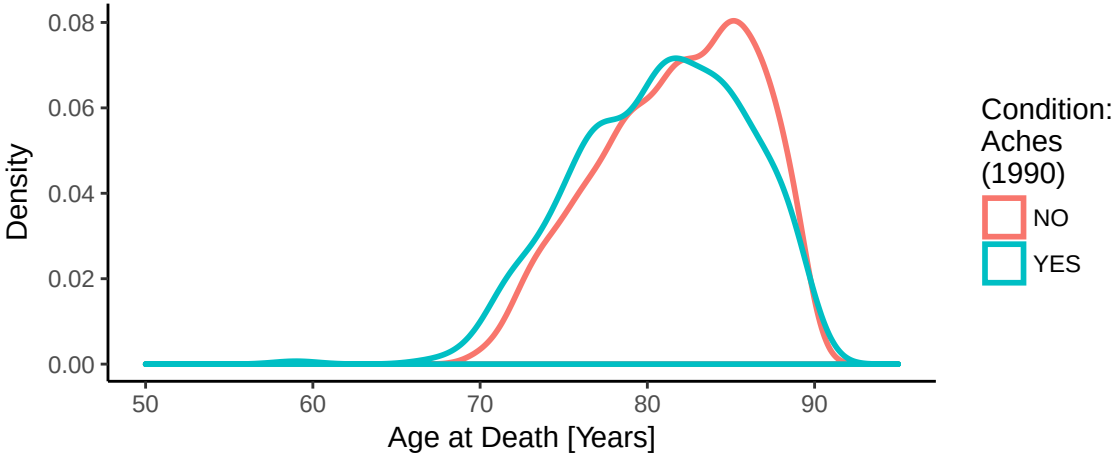

Density
0.15
0.10
0.05
0.00
50
60
70
80
90
Age at Death [Years]
Wages & Salary (1990)
> $0k
> $5k
> $10k
> $15k
> $20k
> $30k

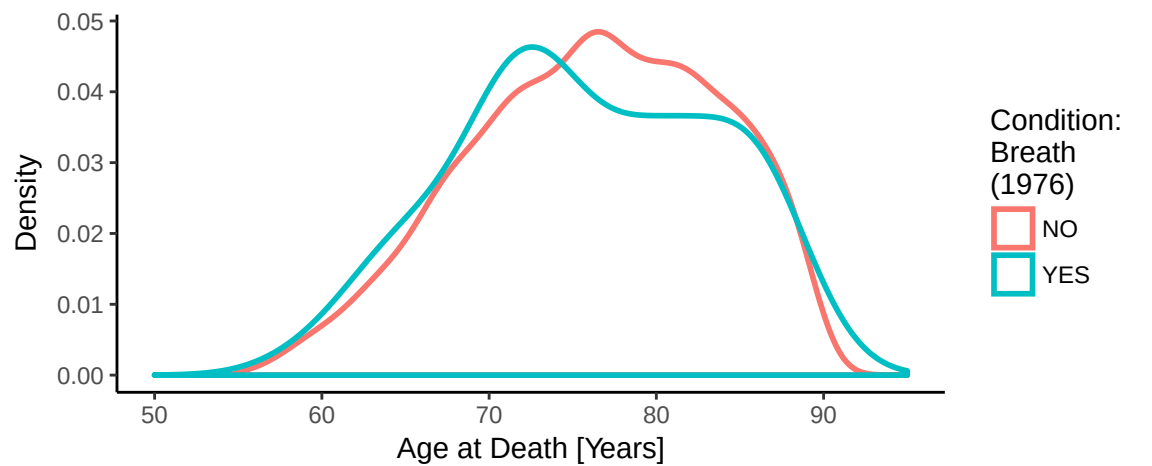

Density
0.075
0.050
0.025
0.000
50
60
70
80
90
Age at Death [Years]
Smoking (1990)
No (R)
Yes (R)
Quit (R)
No (W)
Yes (W)

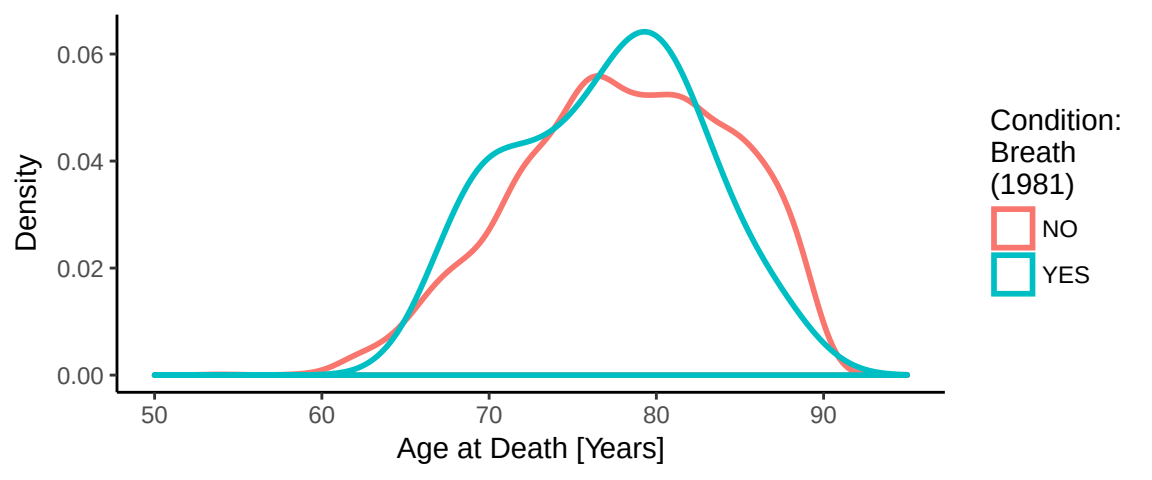

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Heavy Alcohol (1990)
No (R)
Yes (R)
No (W)
Yes (W)

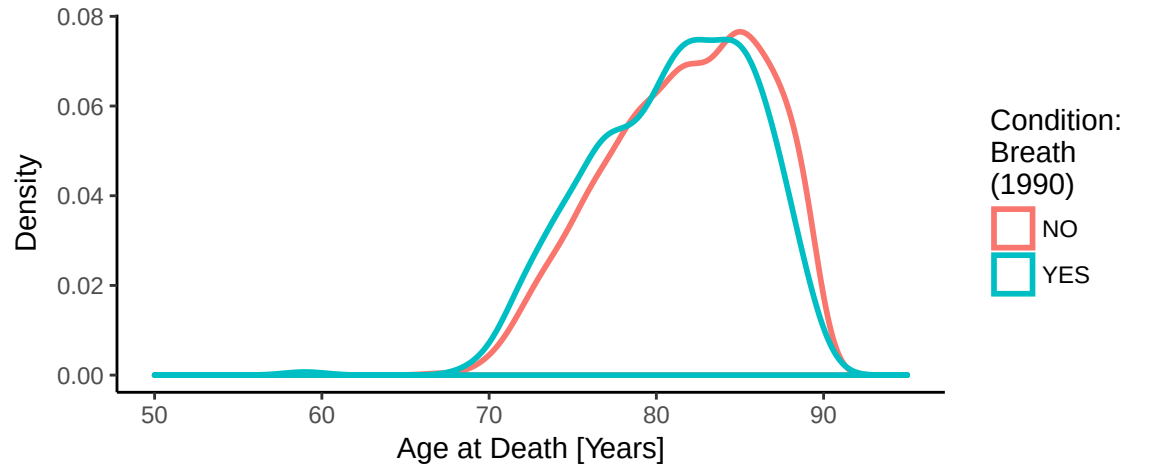

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Aches
(1976)
NO
YES

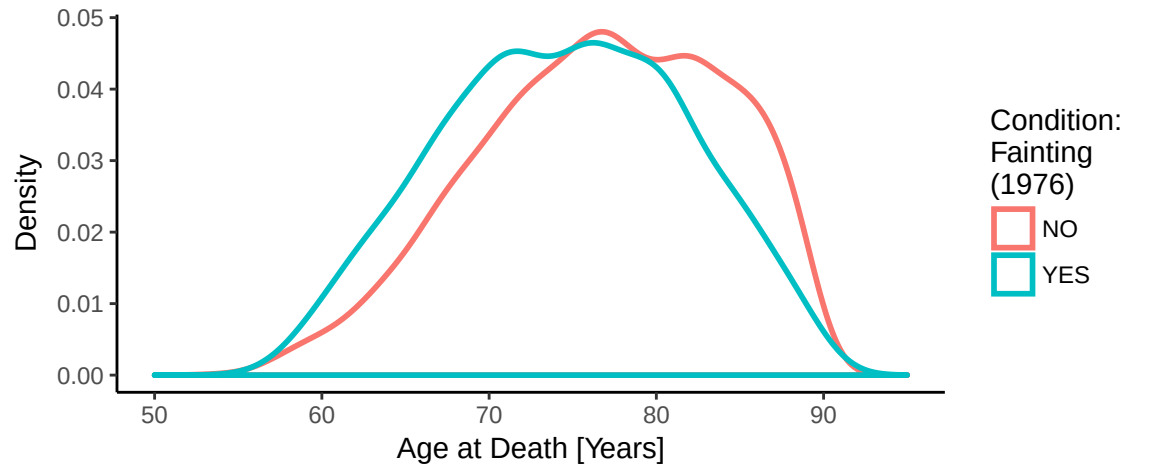

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Aches
(1981)
NO
YES

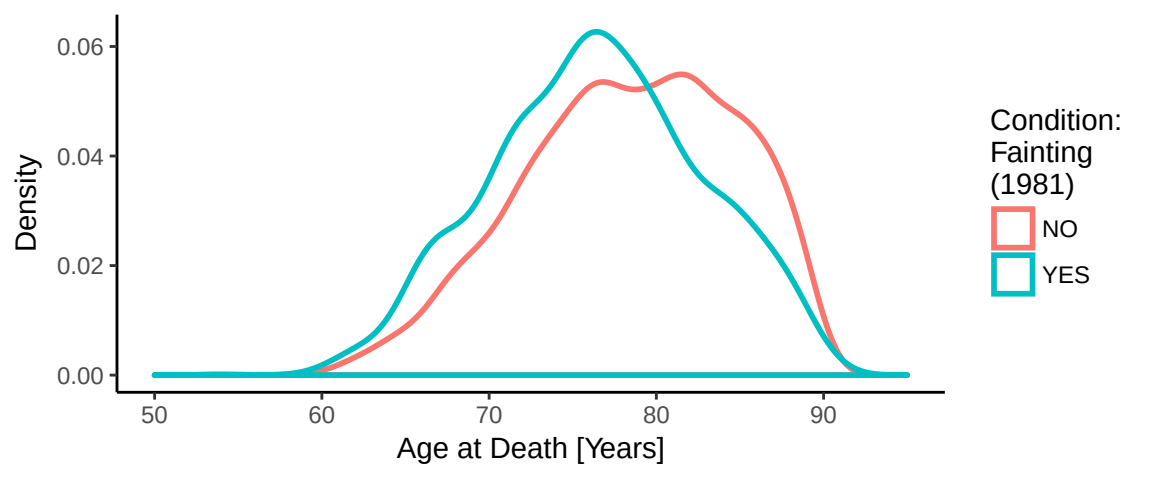

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Aches
(1990)
NO
YES

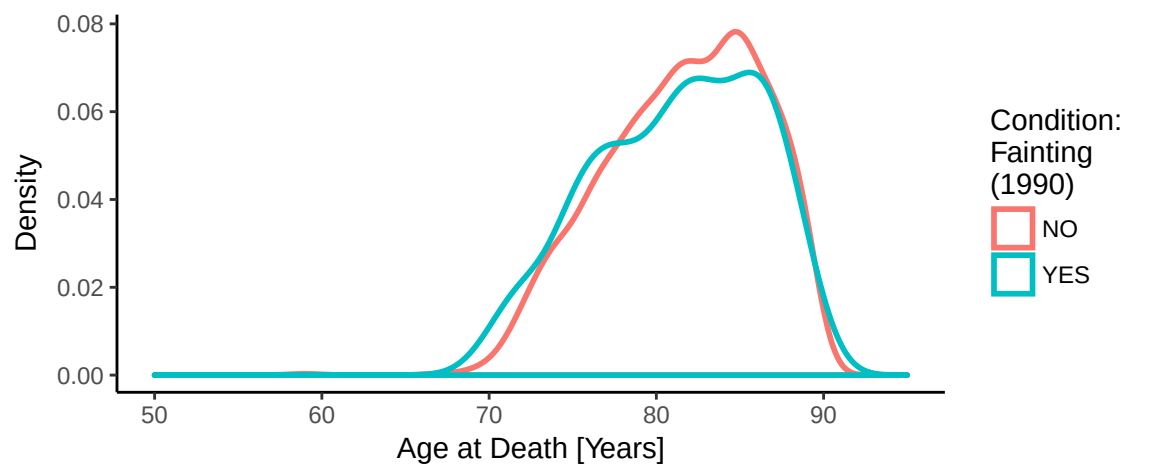

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Breath
(1976)
NO
YES

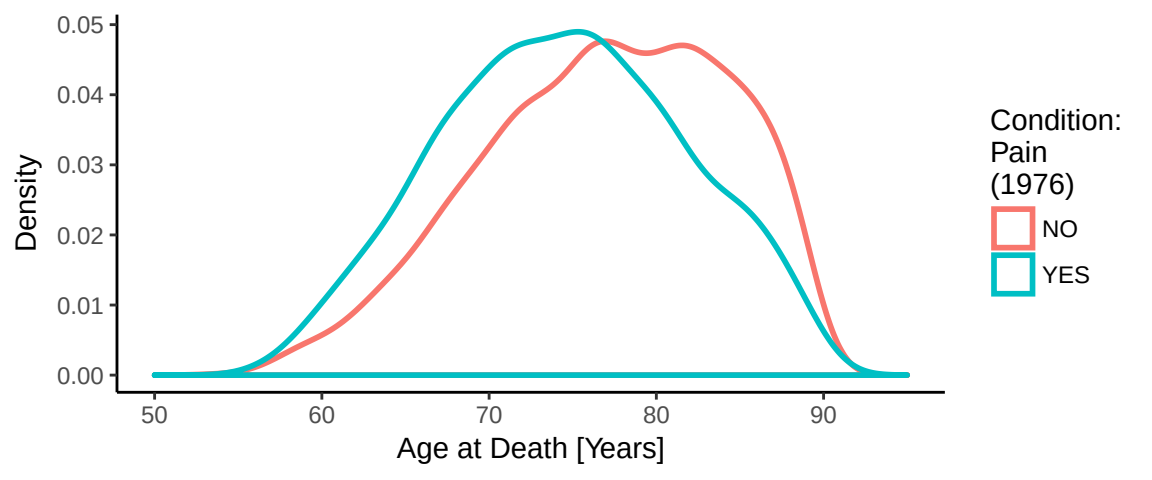

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Breath
(1981)
NO
YES

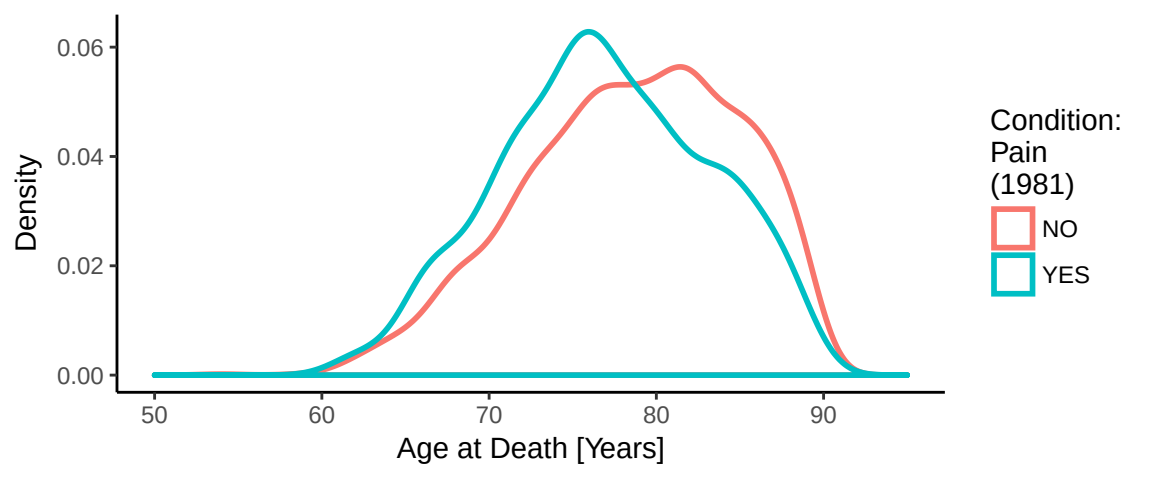

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Breath
(1990)
NO
YES

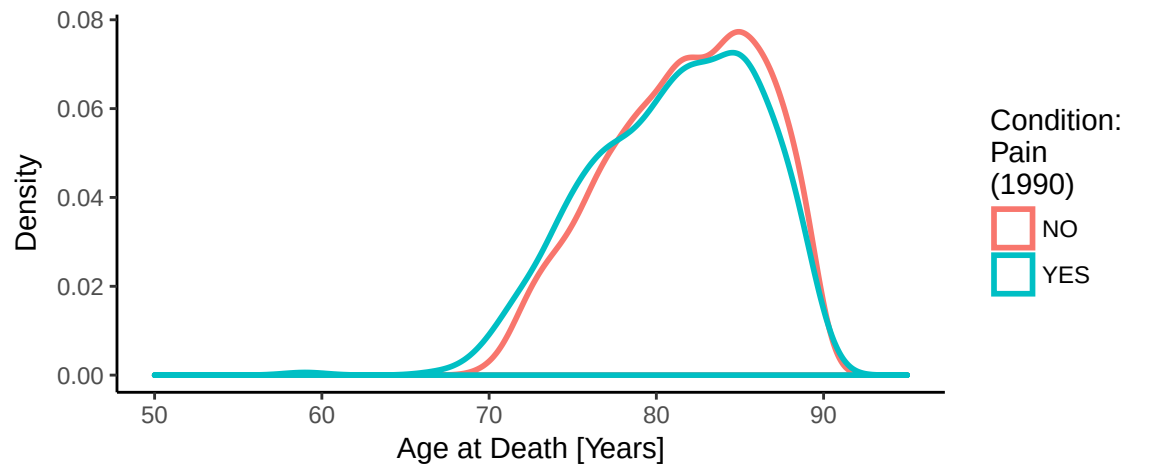

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Fainting
(1976)
NO
YES

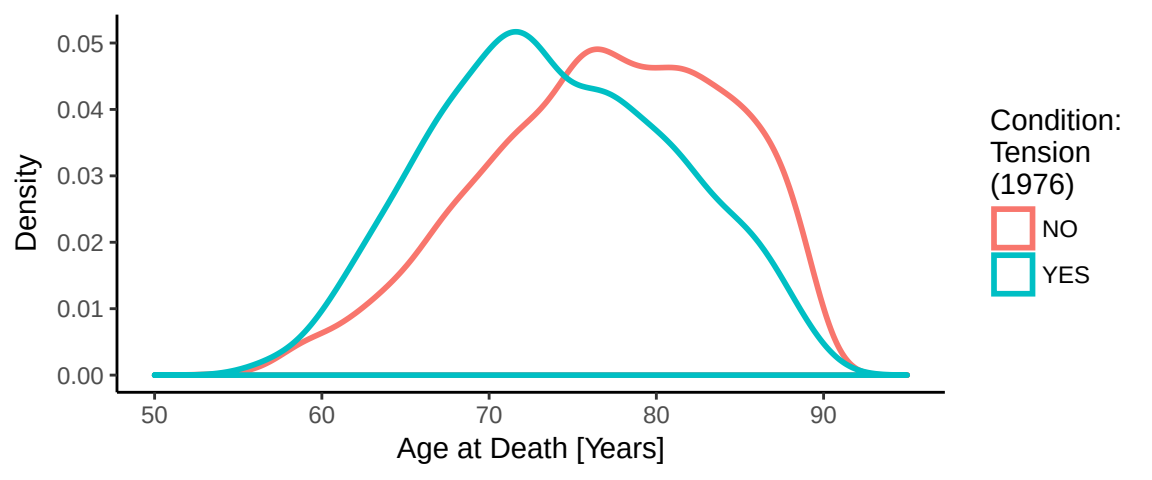

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Fainting
(1981)
NO
YES

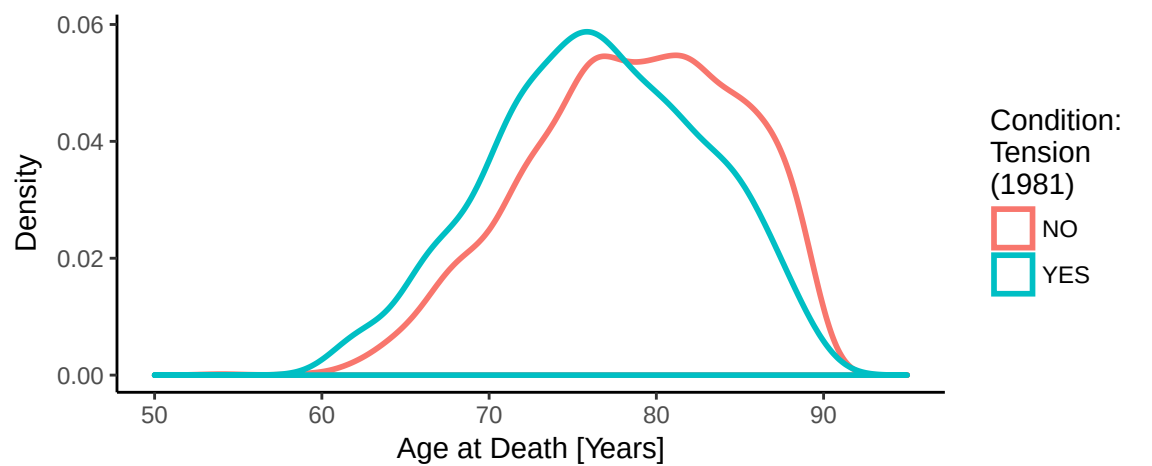

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Fainting
(1990)
NO
YES

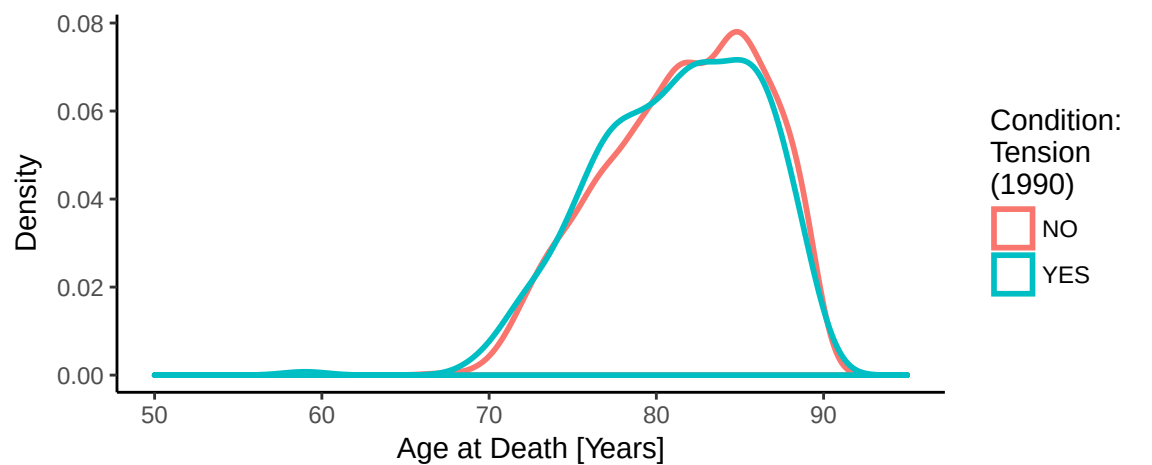

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Pain
(1976)
NO
YES

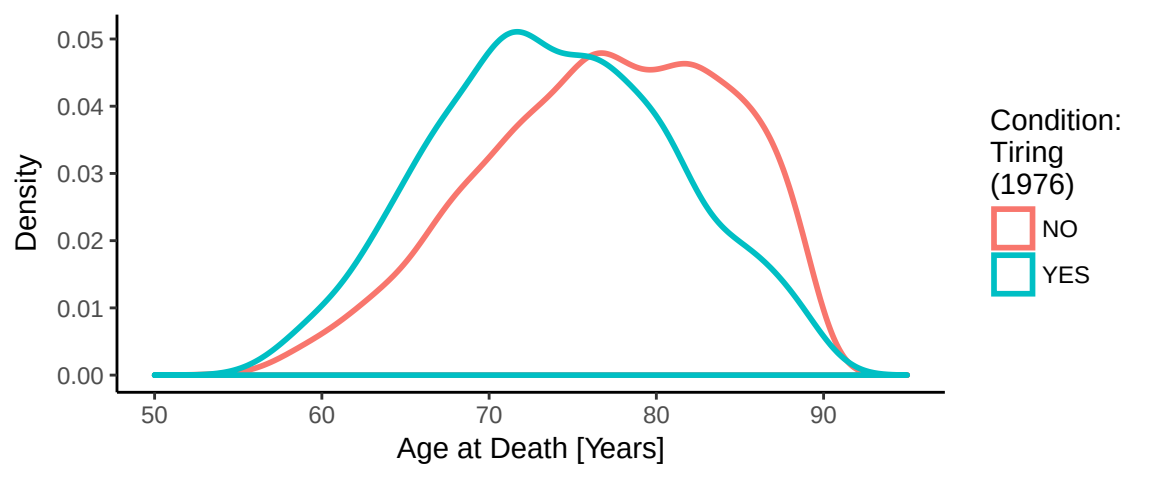

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Pain
(1981)
NO
YES

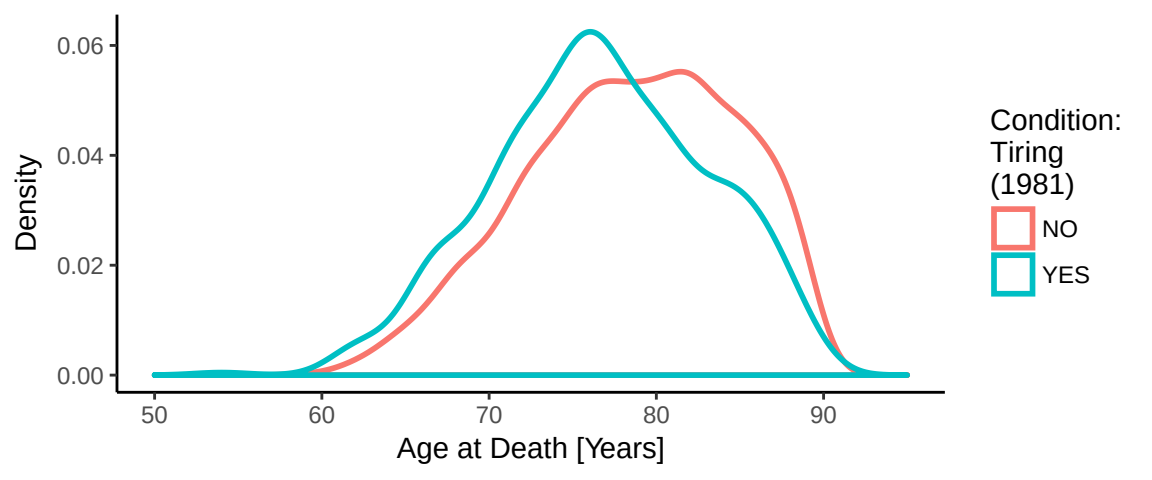

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Pain
(1990)
NO
YES

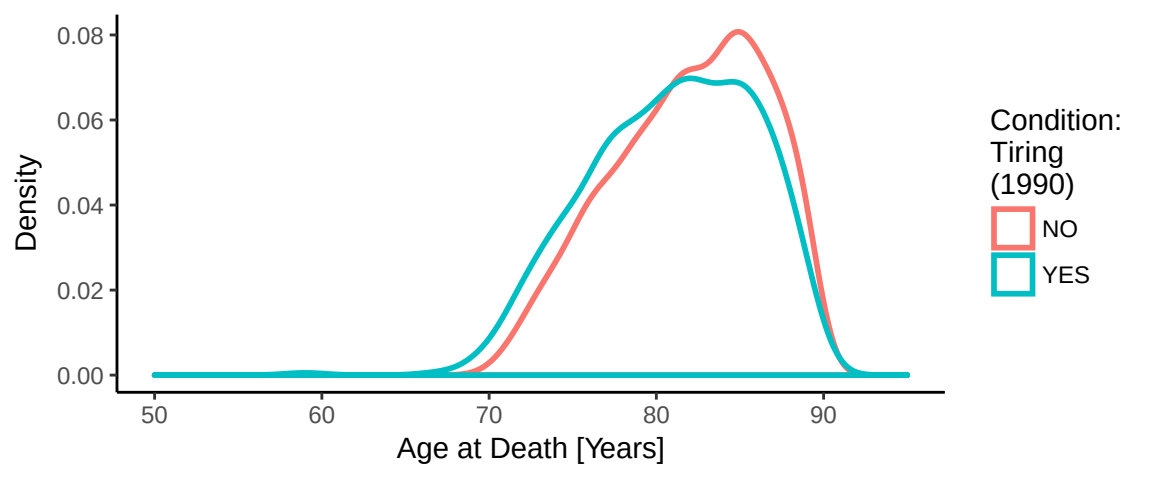

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tension
(1976)
NO
YES

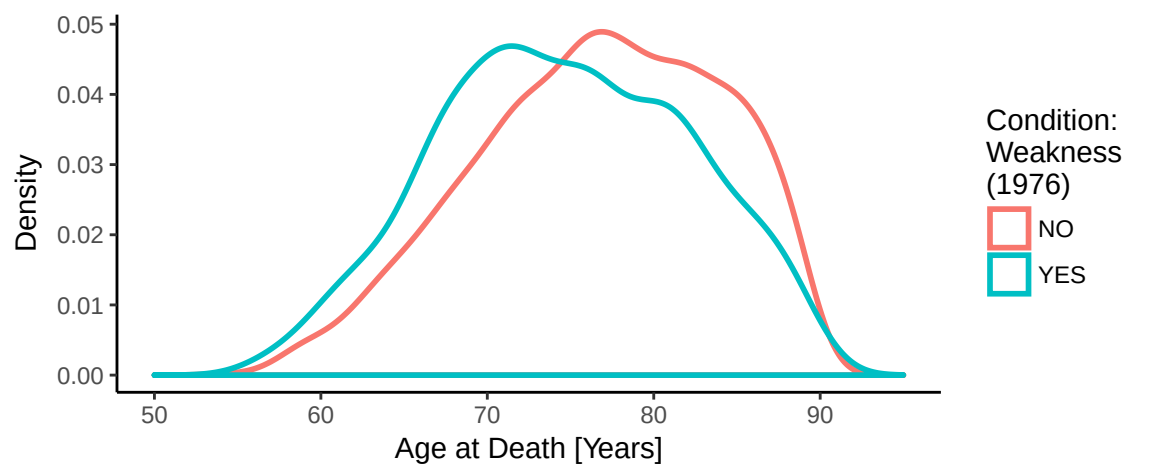

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tension
(1981)
NO
YES

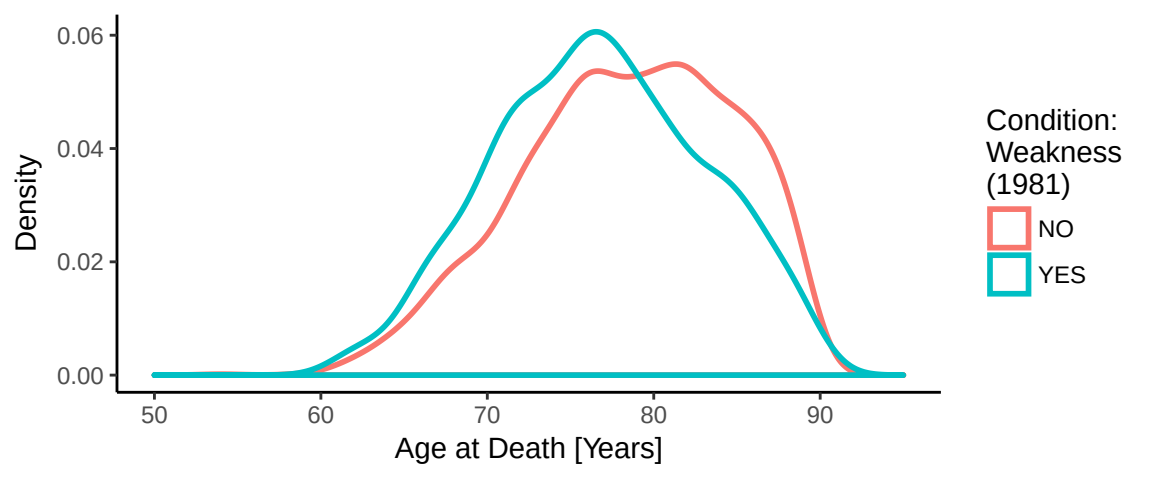

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tension
(1990)
NO
YES

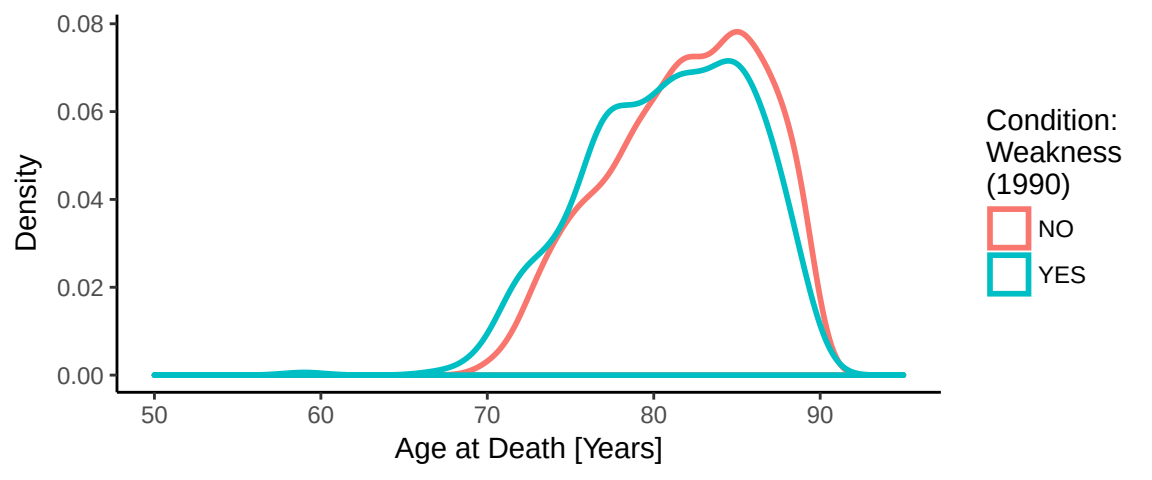

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tiring
(1976)
NO
YES

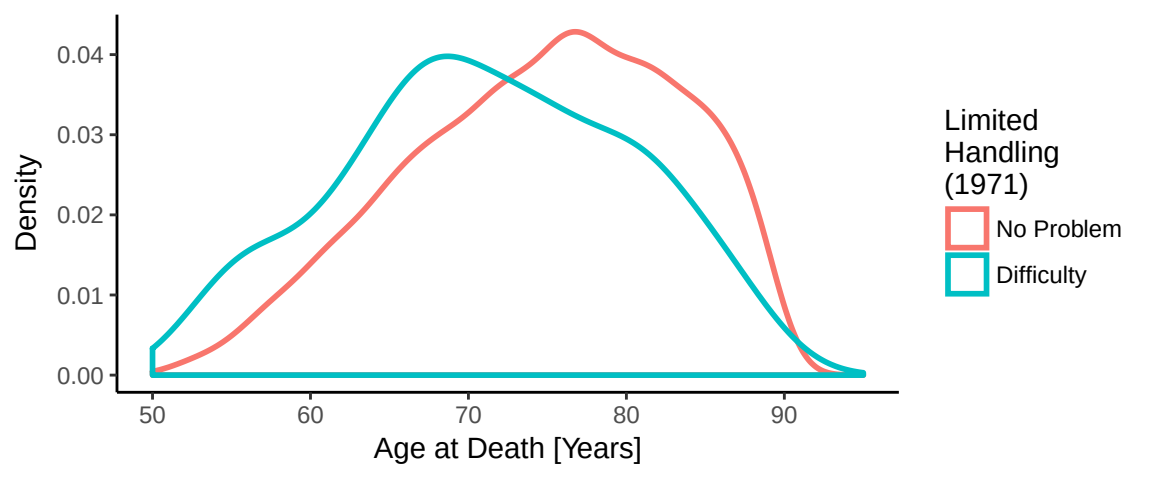

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tiring
(1981)
NO
YES

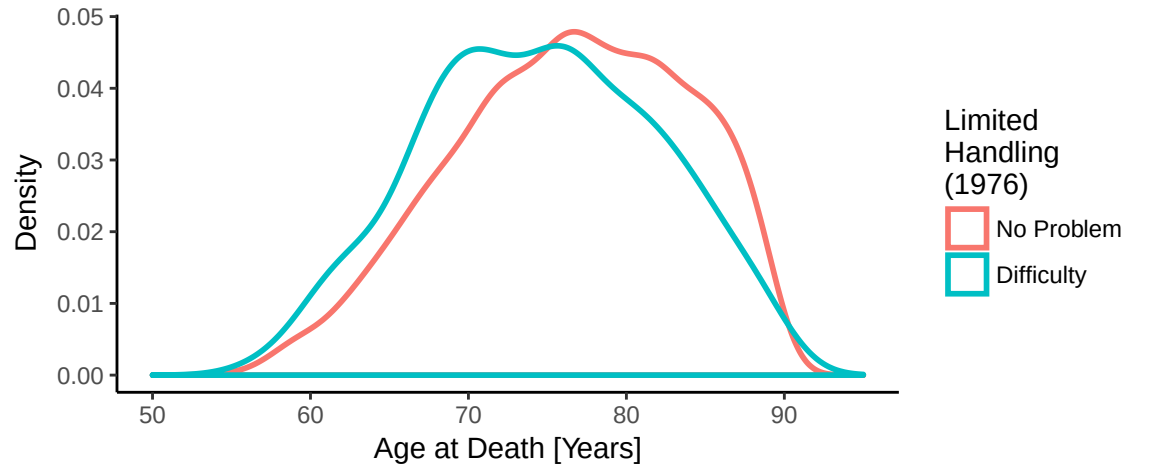

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Tiring
(1990)
NO
YES

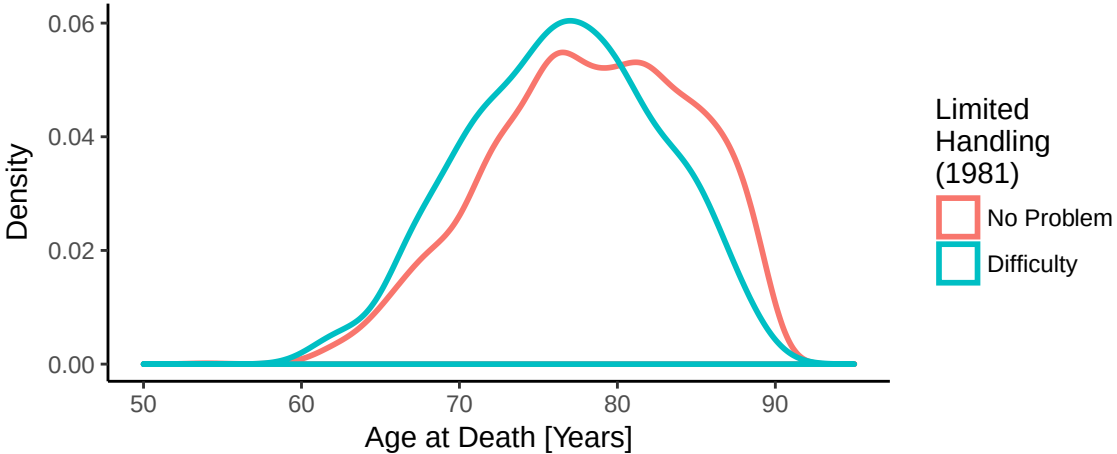

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Weakness
(1976)
NO
YES

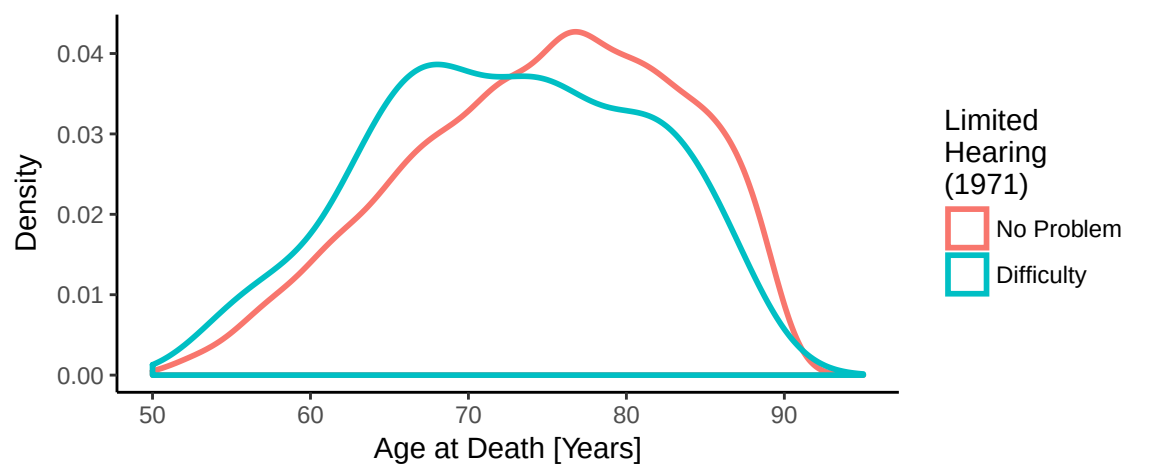

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Weakness
(1981)
NO
YES

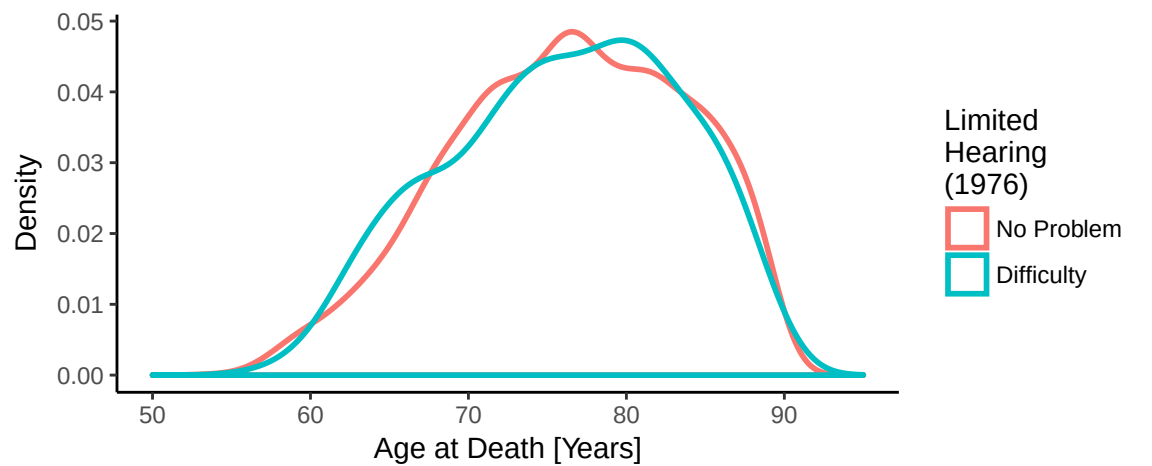

0.08
0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Condition:
Weakness
(1990)
NO
YES

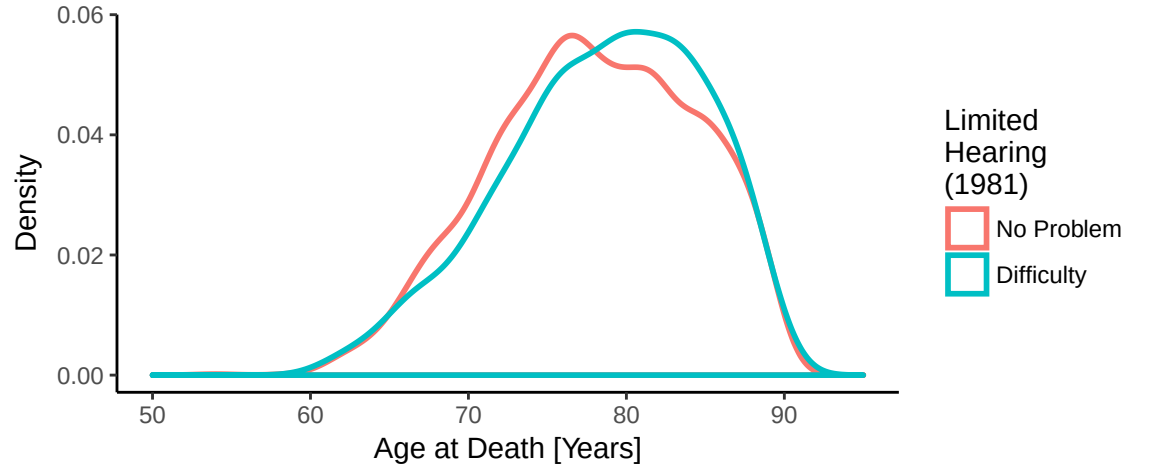

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Handling (1971)
No Problem
Difficulty

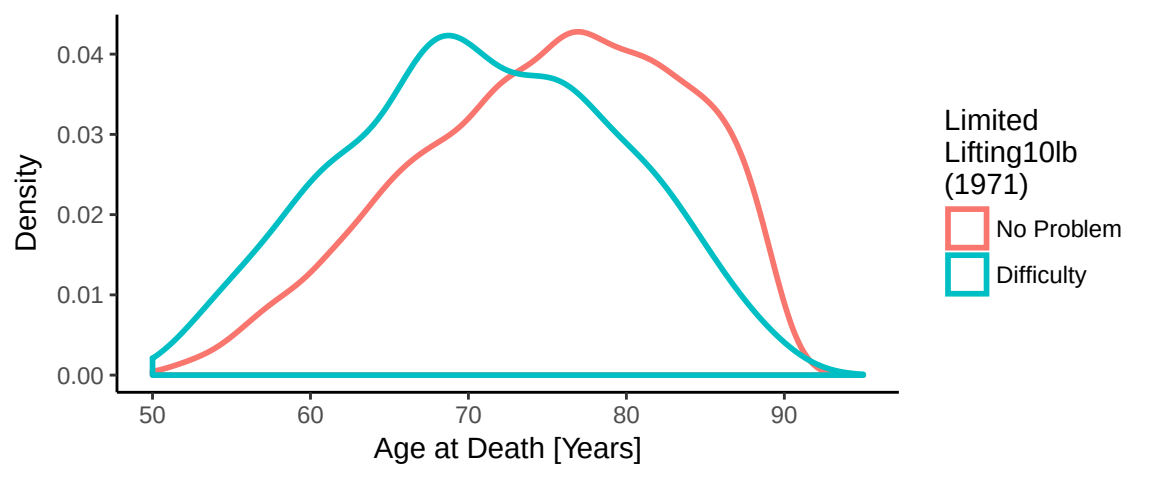

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Handling (1976)
No Problem
Difficulty

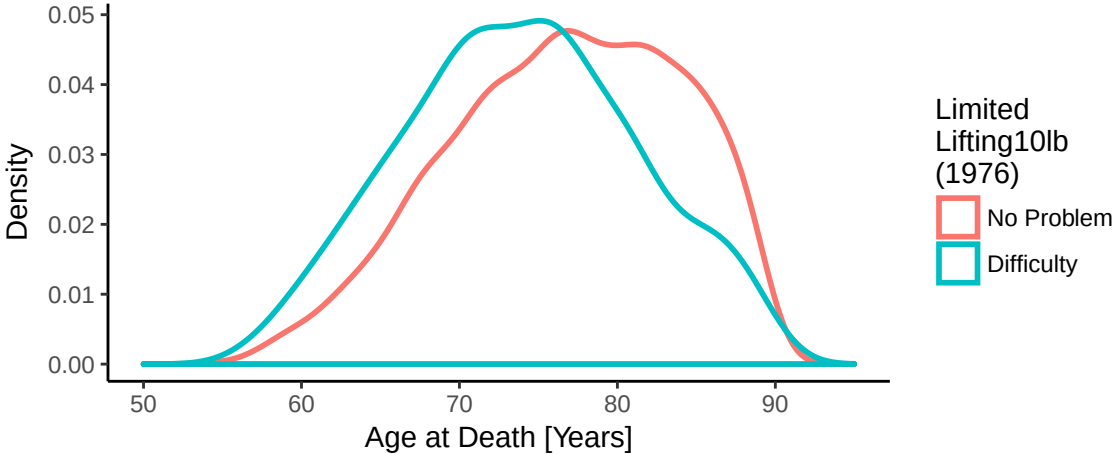

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Handling (1981)
No Problem
Difficulty

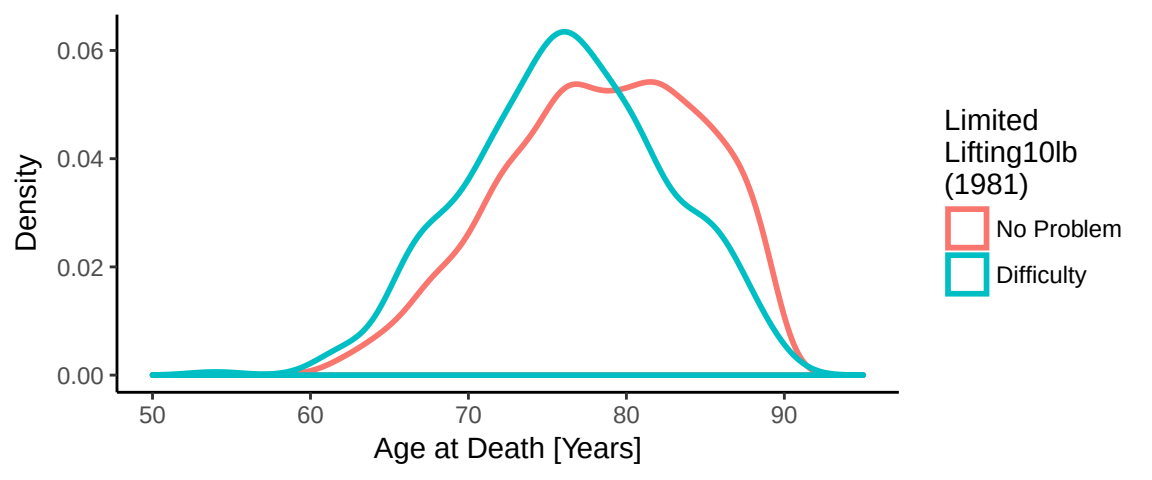

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Hearing (1971)
No Problem
Difficulty

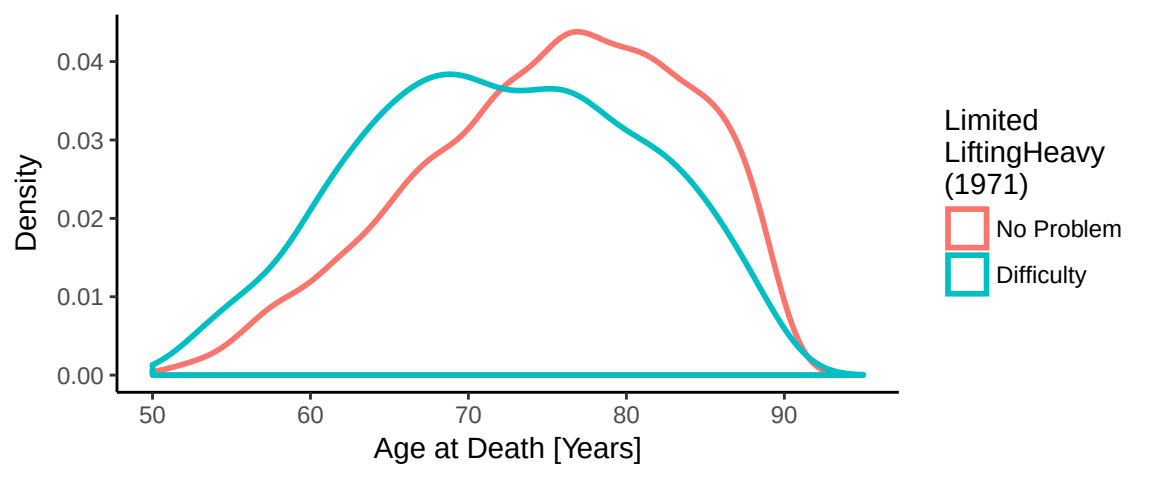

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Hearing (1976)
No Problem
Difficulty

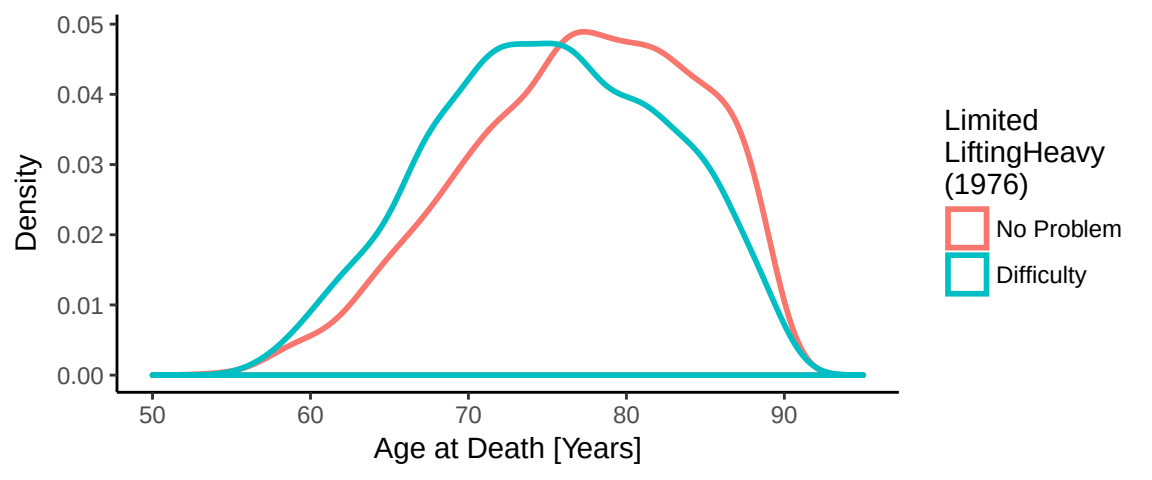

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Hearing (1981)
No Problem
Difficulty

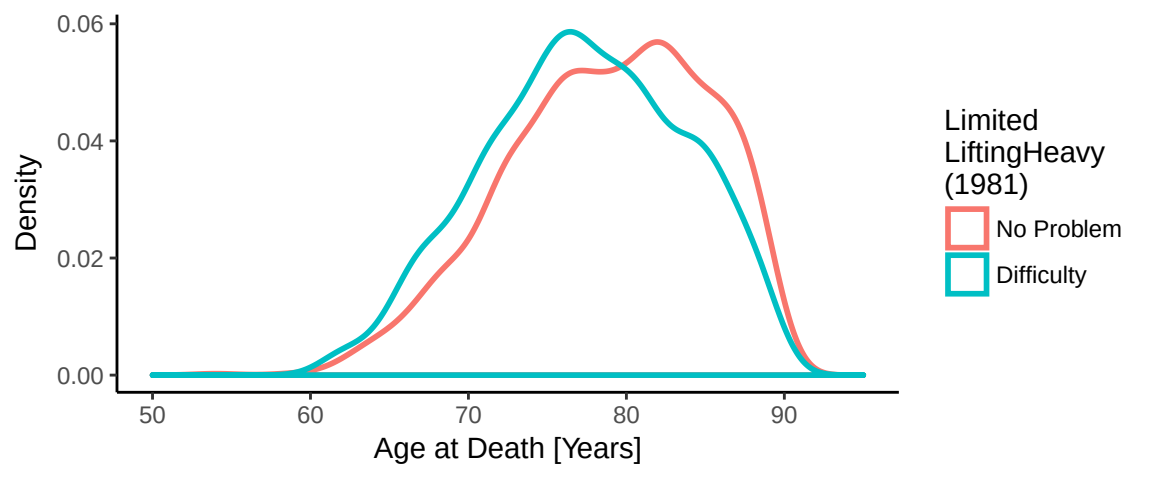

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Lifting10lb (1971)
No Problem
Difficulty

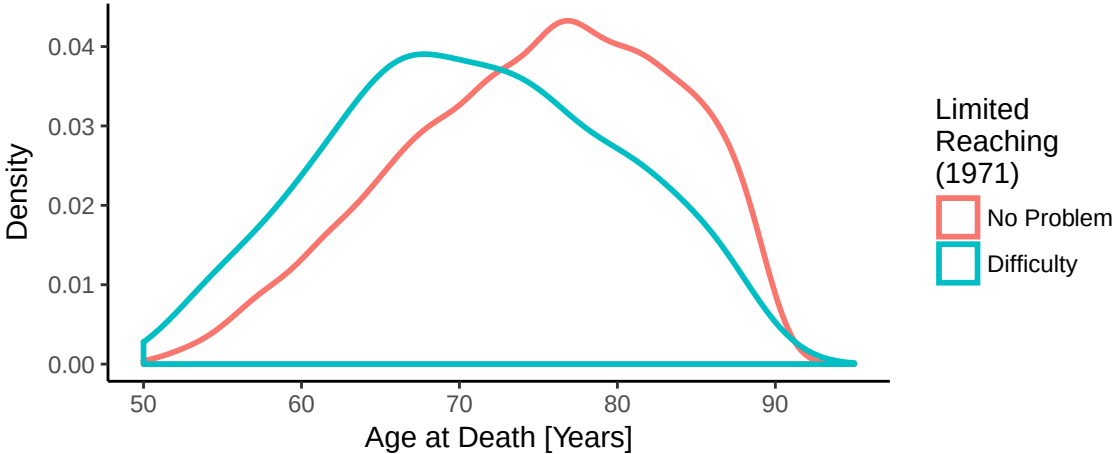

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Lifting10lb (1976)
No Problem
Difficulty

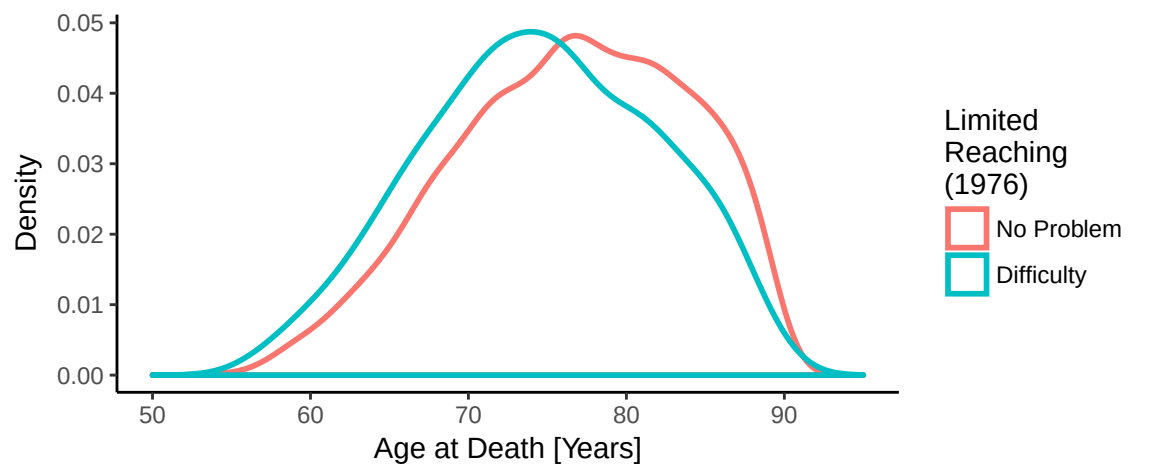

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Lifting10lb (1981)
No Problem
Difficulty

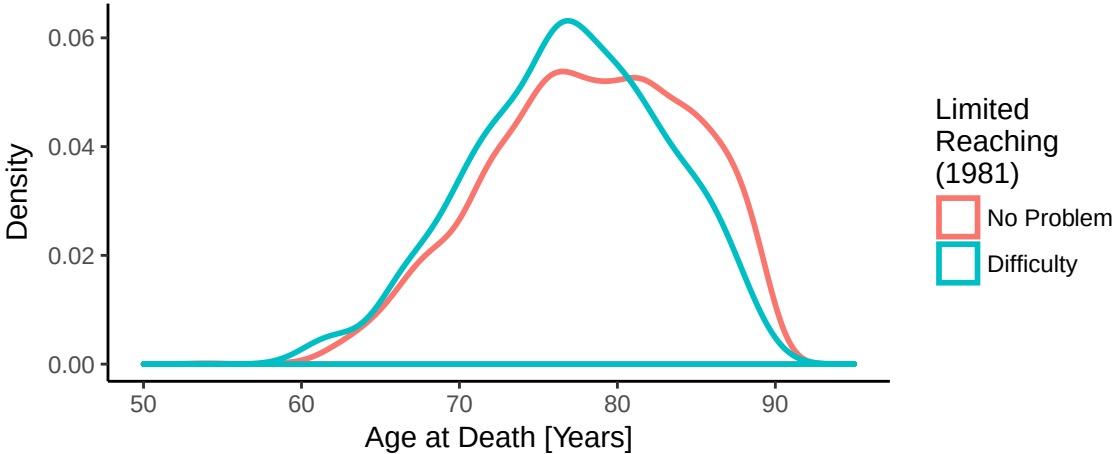

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited
LiftingHeavy
(1971)
No Problem
Difficulty

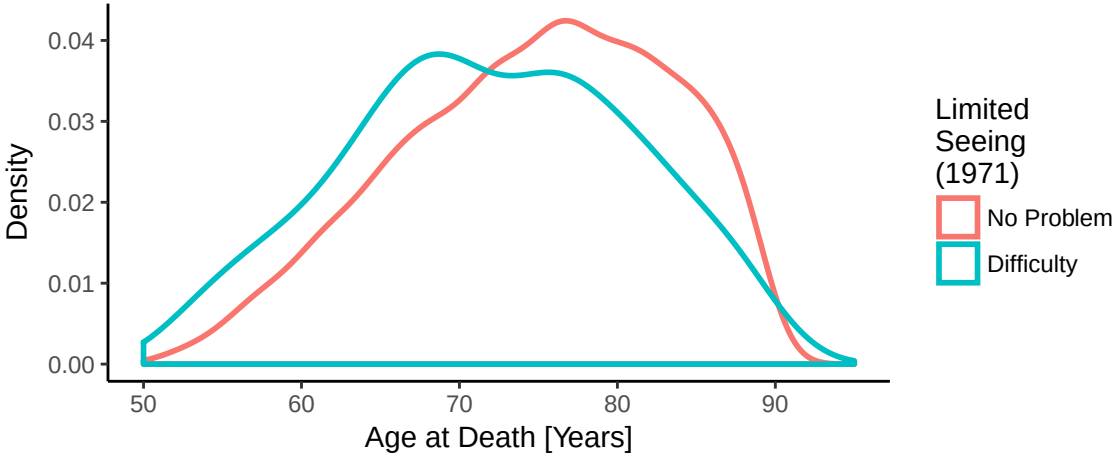

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited LiftingHeavy (1976)
No Problem
Difficulty

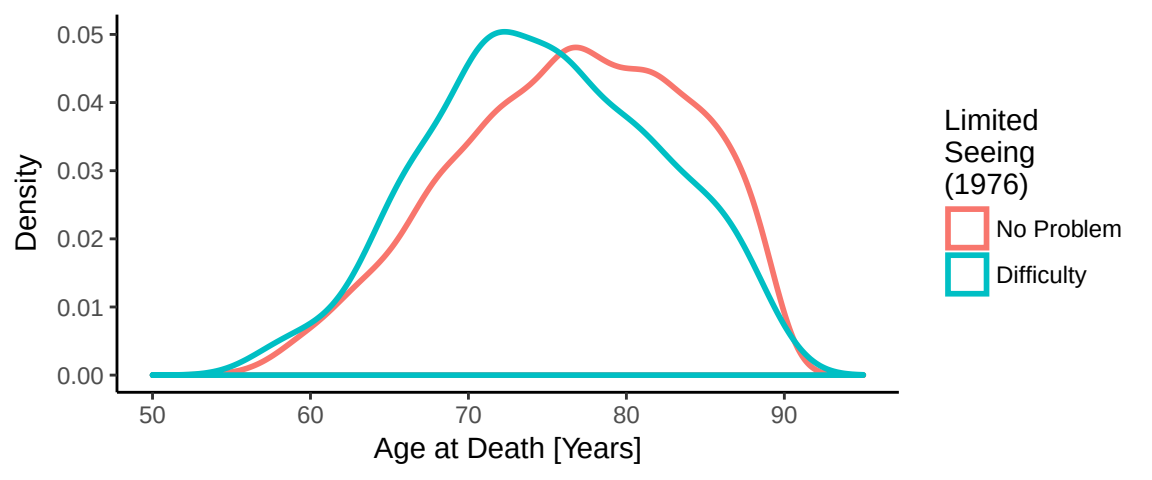

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited
LiftingHeavy
(1981)
No Problem
Difficulty

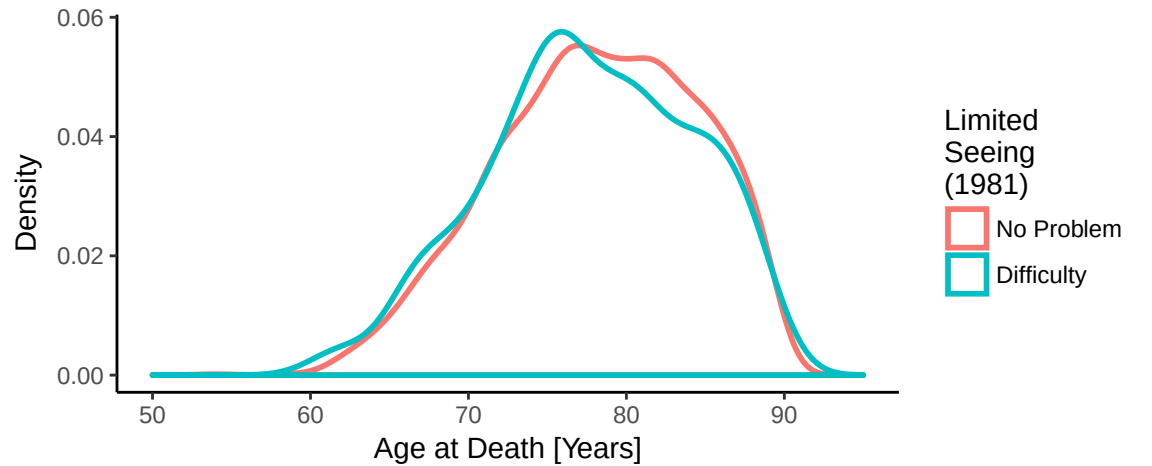

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Reaching (1971)
No Problem
Difficulty

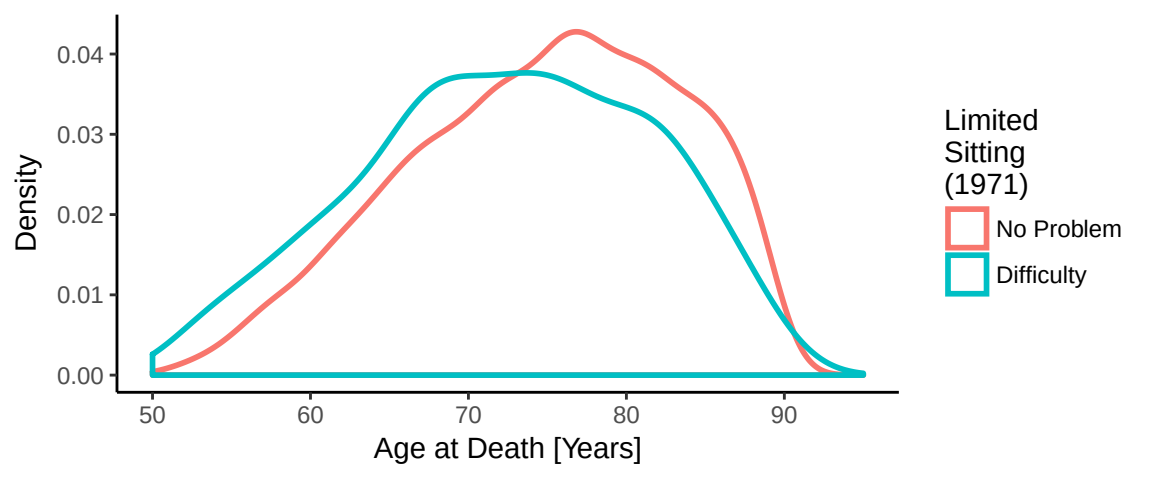

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited
Reaching
(1976)
No Problem
Difficulty

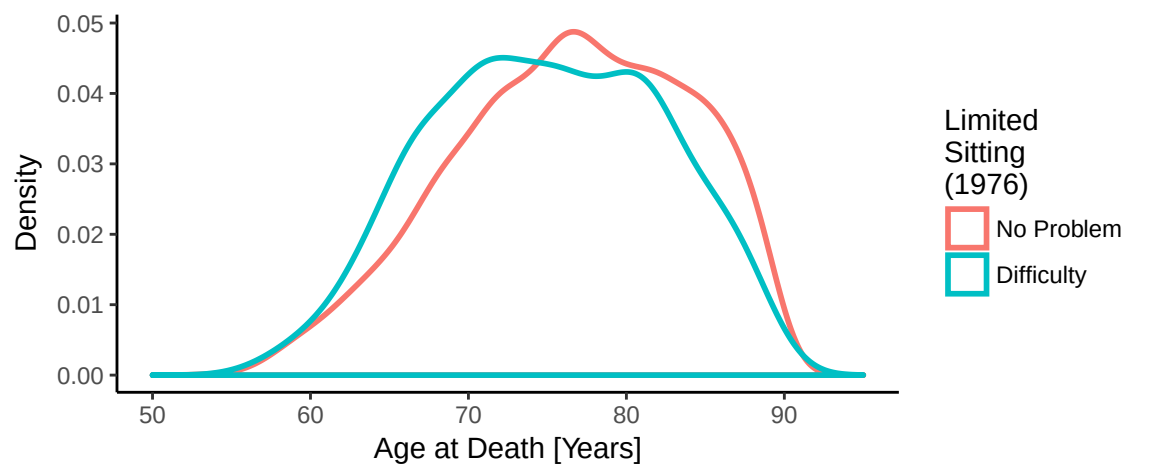

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Reaching (1981)
No Problem
Difficulty

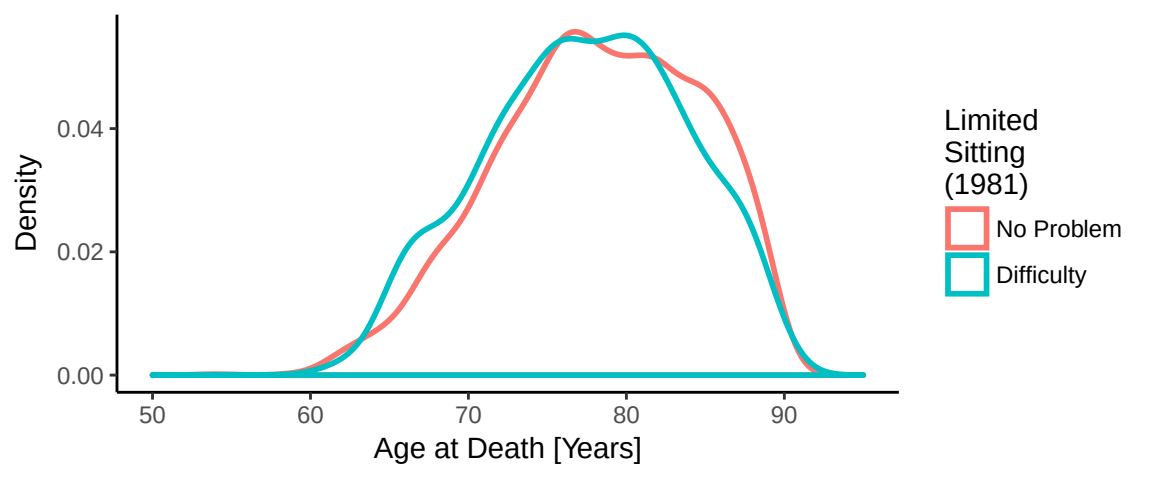

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Seeing (1971)
No Problem
Difficulty

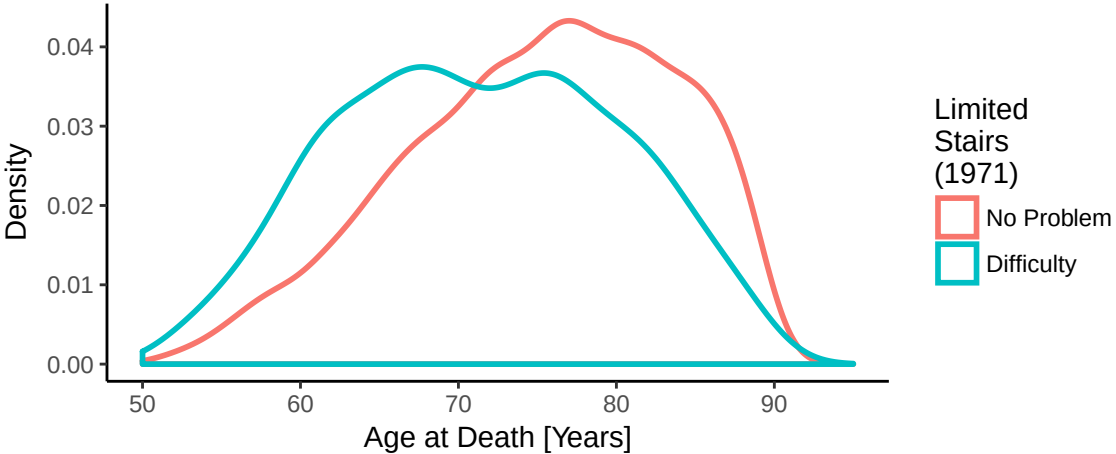

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Seeing (1976)
No Problem
Difficulty

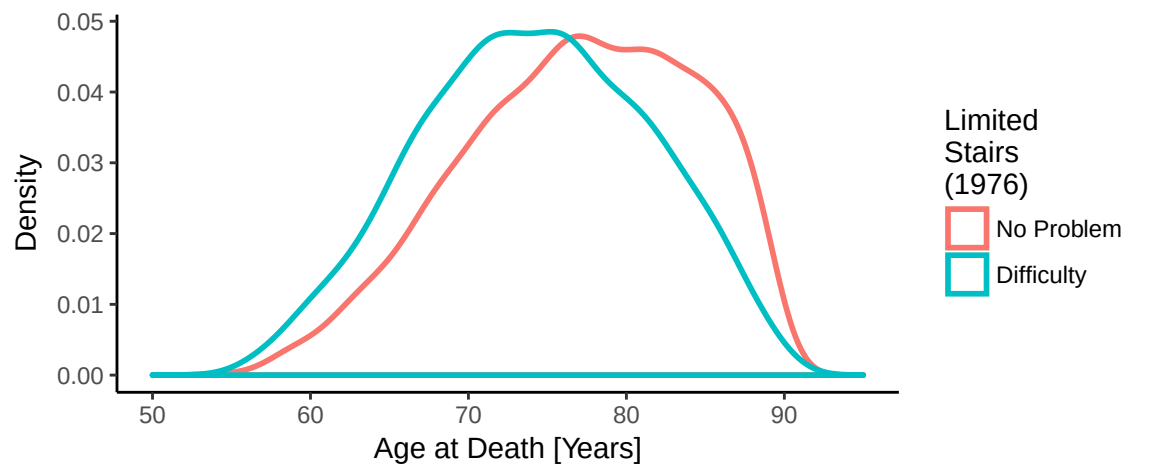

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Seeing (1981)
No Problem
Difficulty

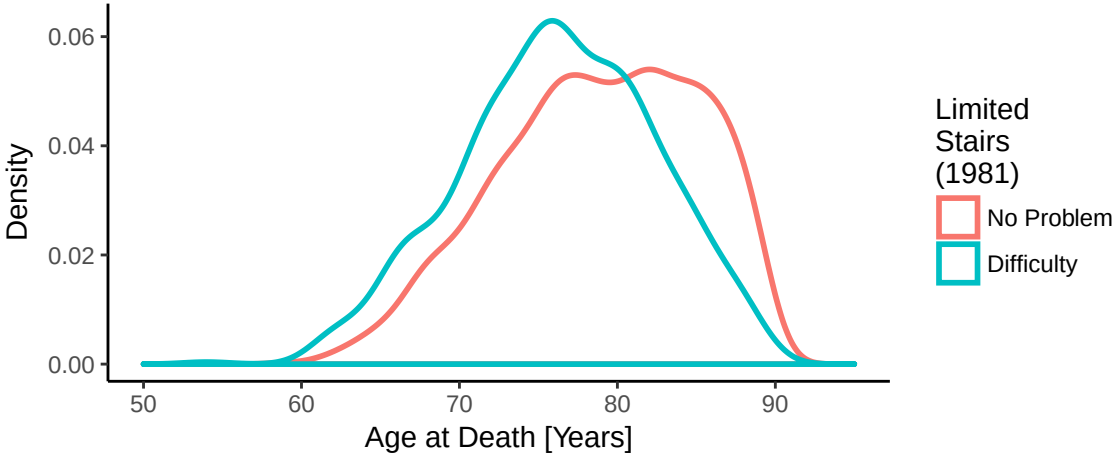

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Sitting (1971)
No Problem
Difficulty

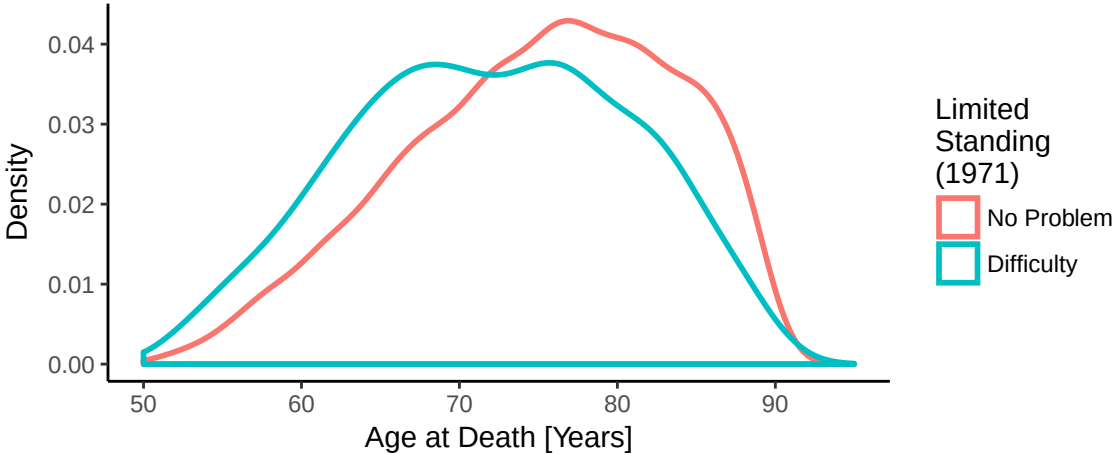

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited
Sitting
(1976)
No Problem
Difficulty

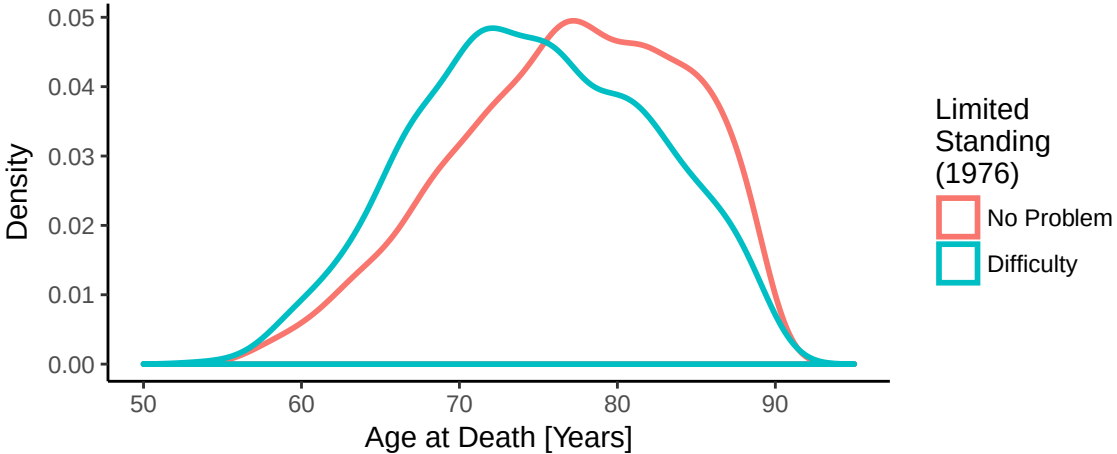

Density
0.04
0.02
0.00
50
60
70
80
90
Age at Death [Years]
Limited Sitting (1981)
No Problem
Difficulty

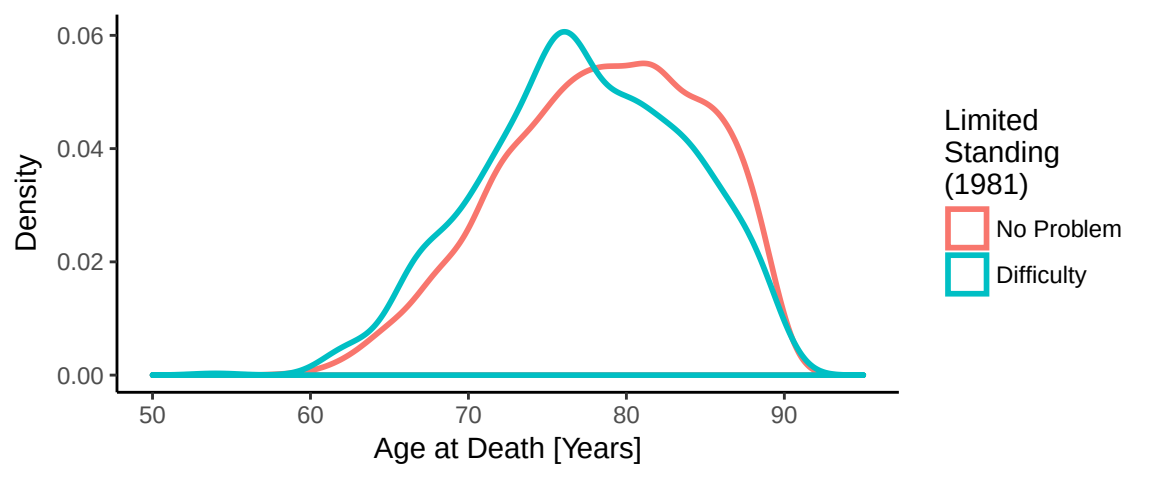

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Stairs (1971)
No Problem
Difficulty

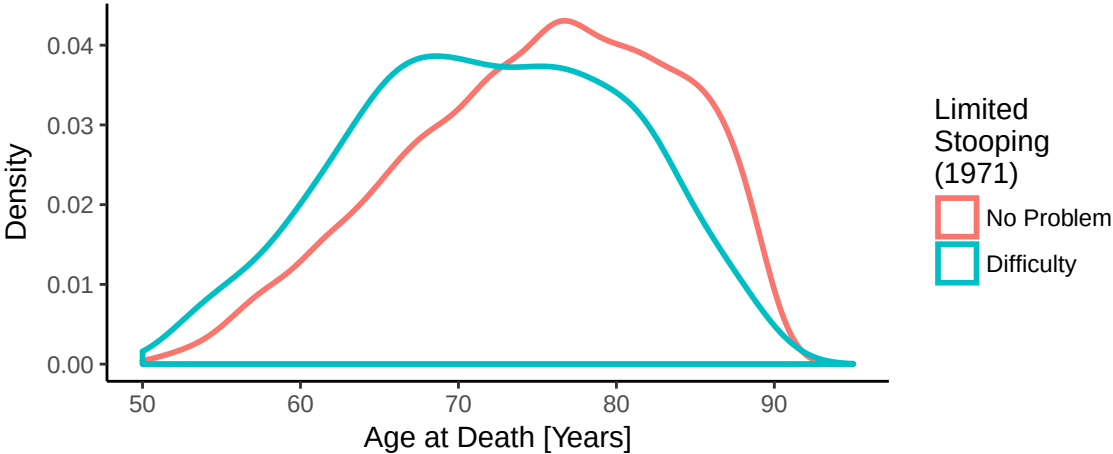

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited
Stairs
(1976)
No Problem
Difficulty

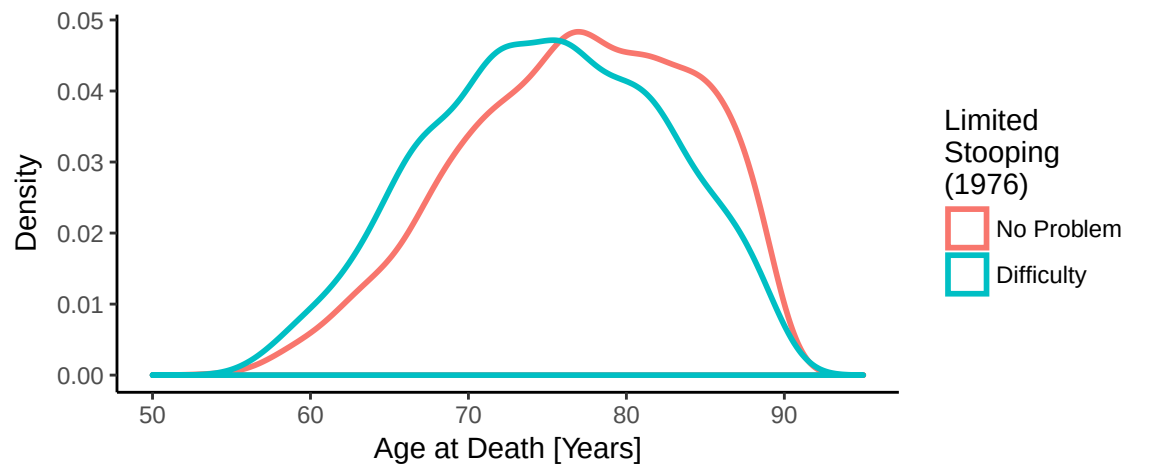

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Stairs (1981)
No Problem
Difficulty

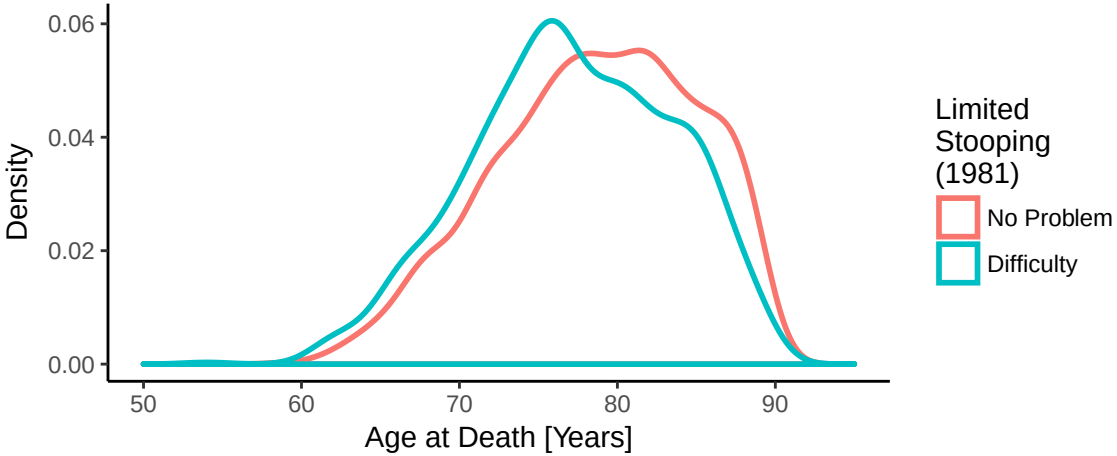

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Standing (1971)
No Problem
Difficulty

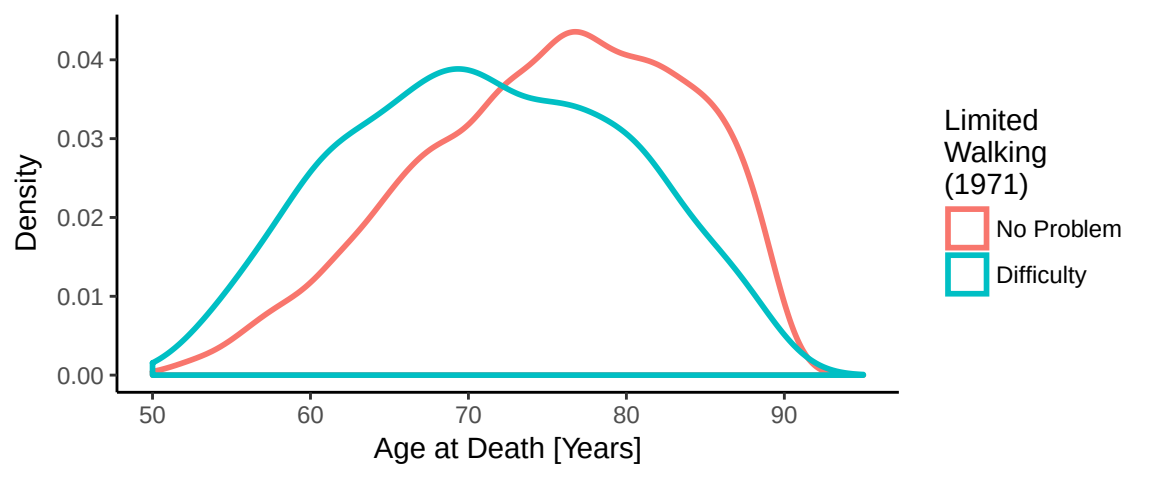

0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Standing (1976)
No Problem
Difficulty

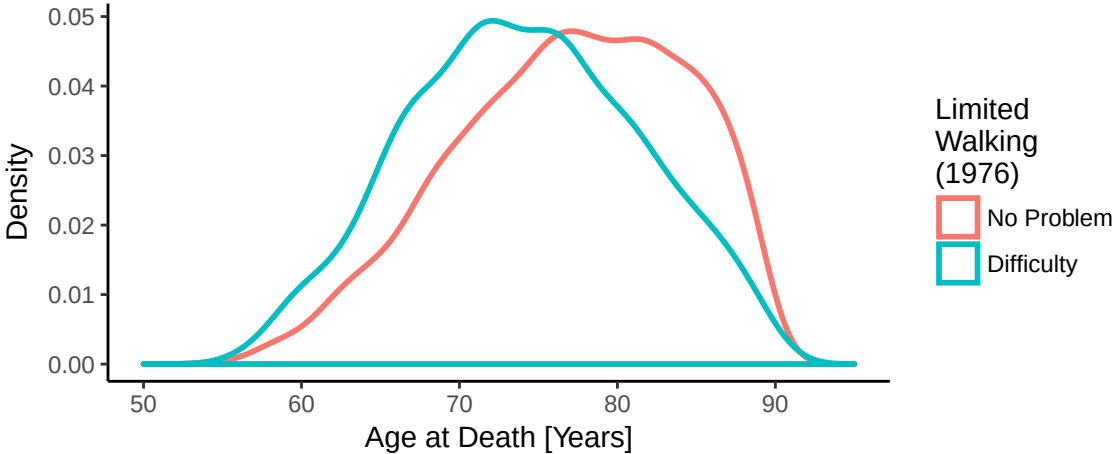

0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Standing (1981)
No Problem
Difficulty

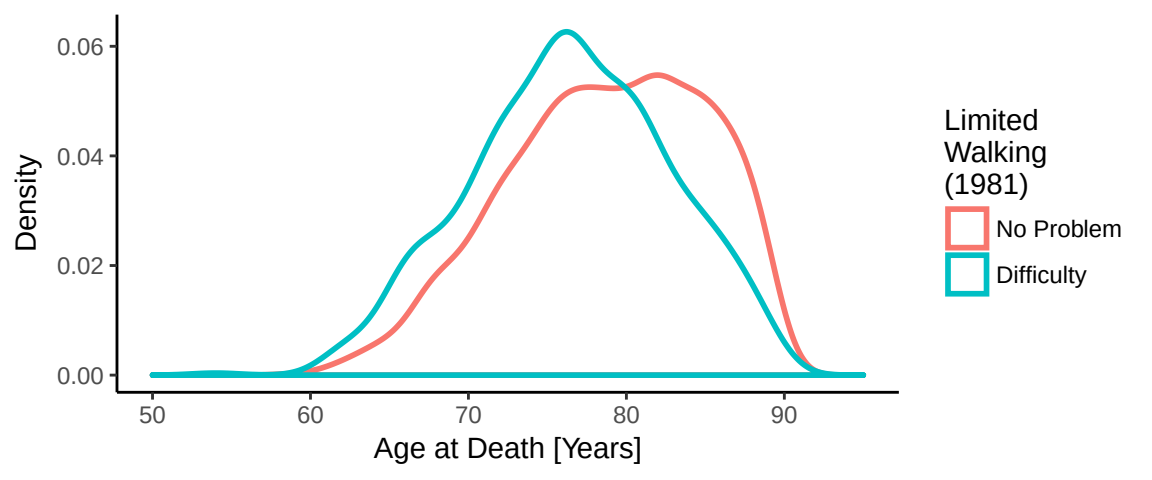

Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Stooping (1971)
No Problem
Difficulty


0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Stooping (1976)
No Problem
Difficulty


0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Stooping (1981)
No Problem
Difficulty


Density
0.04
0.03
0.02
0.01
0.00
50
60
70
80
90
Age at Death [Years]
Limited Walking (1971)
No Problem
Difficulty


0.05
0.04
0.03
0.02
0.01
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Walking (1976)
No Problem
Difficulty


0.06
0.04
0.02
0.00
Density
50
60
70
80
90
Age at Death [Years]
Limited Walking (1981)
No Problem
Difficulty