Assessing Risk and Fairness: The Role of Statistical Science in Policy

David W. Scott *

December 5, 2000

Abstract

Probability theory provides a mathematical framework for modeling risk. Philosophy considers fundamental questions of the nature and meaning of chance. But it falls upon statistical science to collect and analyze data to estimate risk, influence policy, and make decisions. Insurance provides a compelling case study for notions of fairness and subsidy. This talk will examine the notion of a fair game and consider its application in areas including decision making, social security, medical insurance, and exit polling. What are some of the elements of "fair" policies? A deeper understanding of statistical modeling and evaluation would illuminate subsidies implicit in public policy and would sharpen political debate.

Discussants: Steve Klineberg, Joan Strassmann, Randi Martin.

1 Introduction

Thank you very much for the kind introduction, Jim. I am especially grateful to Franz Brotzen who invited me to give this talk a year ago. One of the great pleasures of my time as a new faculty member was meeting Solomon

*David W. Scott (*scottdw@rice.edu*) is Professor of Statistics, MS-138, Rice University, PO Box 1892 Houston, TX 77005-1892. This paper was presented at the Scientia Lecture Series at Rice University on December 6, 2000. This research was supported in part by the National Science Foundation grant #99-71797

Bochner, founder of the *Scientia* lecture series. Bochner was certainly one of the premier mathematicians of the 20th Century and his work is widely used in statistical theory.

When I wrote my abstract for this talk, the election was still ahead of us. Like many of you on election night, I stayed up to 4:30 a.m. before retiring. The event contained a number of statistical failure and intrigue and while not the focus of this talk, I will try to offer a few comments at the end. However, when I re-read the last phrase in my abstract about "sharpen[ing] the political debate" in the context of events over the past several weeks, I sometimes wonder if the debate isn't already sharp enough.

For the title of my talk, I chose the phrase "statistical science" because it encompasses not only the Department of Statistics proper, but the many groups across this campus for whom statistics is an integral part of research and teaching. Indeed, all three of the discussants today are serious consumers of statistical methodology. The Department of Statistics has a number of joint faculty who come from Psychology, Economics, the Jones School, Education, Political Science, and Electrical Engineering. An equal number of departments could easily be added to this list.

Statistical methods are applied in many marvelous ways. Sophisticated word counting techniques may be used to help determine the authorship of disputed manuscripts. The statistical tools that underlie computer voice recognition are among the amazing examples of the unimaginable power to take in millions of small increments of information and produce text. Data mining the massive data sets commonplace in business, government, and science is a huge and growing endeavor.

Public policy involves a complicated set of incentives and disincentives. Many involve subsidies that are not well-understood by the public. Without a clear picture of the nature and size of such subsidies, informal discussion of policy changes is limited. Universities play an important role in the process of collecting and analyzing data for public dissemination. For example, the Global Warming conference held at the Baker Institute earlier this Fall was enormously successful at illuminating what is known and what is not known about the science of global warming. A simple answer is seldom available.

Today I would like to take a few minutes to explain what data analysis is all about, how to think about risk, and to explore the implications of the notion of fairness in policy, with examples from insurance, elections, and social security. Over the past eight years, I have had the good fortune to serve as co-chair of the fringe benefits committee, chair the faculty club board, and be a member of the Provost's early-retirement committee. I hope to use Rice examples often in hopes that they will mean something to you.

2 Risk

The field of actuarial statistics is devoted to modeling and collecting data to handle risky situations. Probability theory is used to abstract risk, while Statistics uses data to estimate the models. Although such matters can be dry, the physicist Wilson wrote an article in 1979 that tried to promote the idea of "equivalent risks." In particular, he wondered what activities would increase the change of death by "one in a million." Here is a portion of his list. In the accident category, we see that flying 1000 miles in a jet is equivalent to driving 350 miles or 10 miles by bicycle or 6 minutes in a canoe. Someone with a fear of flying won't be convinced, but I imagine there are many individuals here who may have wanted an excuse not to go canoeing.

RISK	OUTCOME
Smoking 1.4 cigarettes	Cancer, heart disease
Living 2 months with a cig. smoker	Cancer, heart disease
Drinking 1/2 liter of wine	Cirrhosis of the liver
Drinking 30 12 oz cans diet soda	Cancer due to saccharin
Drinking Miami's water 1 year	Cancer due to chloroform
Eating 40 Tbsp's peanut butter	Liver cancer aflatoxin B
Eating 100 charcoal broiled steaks	Cancer due to benzopyrene
Spending 1 hour in a coal mine	Black lung disease
Spending 3 hours in a coal mine	Accident
Traveling 6 minutes by canoe	Accident
Traveling 10 miles by bicycle	Accident
Traveling 300 miles by car	Accident
Flying 1000 miles by jet	Accident
Living 2 days in New York or Boston	Air pollution
Flying 6000 miles by jet	Cancer due to cosmic radiation
Living 2 months in Denver	Cancer due to cosmic radiation
Living 2 months in a brick building	Cancer due to nat'l radioactivity
One chest x-ray	Cancer due to radiation
Exposure to nuclear event at 5 miles	Cancer due to radiation

Many items in his list relate to cancer. Smoking 1.4 cigarettes is equivalent to living 2 months with a cigarette smoker or living 2 months in a brick building (natural radioactivity) or living within 5 miles of a nuclear reactor for 50 years. Clearly individual risks versus group risks are perceived differently.

Environmental equivalent risks tabulated include living 2 days in New York City, drinking Miami's tap water for 1 year, or living 2 months in Denver (cosmic radiation at higher altitudes). These are equivalent to the radiation exposure flying 6000 miles in a jet. A colleague at Baylor tells me the Apollo astronauts got a significant dosage and that the Mars astronauts face dangerous risks.

Statistical decision theory is a model for managing risk. Consider this two by two table. In any situation, once you have made a decision, time will tell in which one of two squares you fell. Two squares represent correct decisions while two represent incorrect decisions. Statistical theory indicates how to make the best decision with the data available, and how reliable that decision is.

		PREDICTION	
		A	B
TRUTH	A	√	X
	B	X	√

In the networks' reporting of the results in Florida, we have a most unusual situation. At various times during the first 12 hours after the polls closed, the networks bounced all around the decision table. Over the course of the evening, they predicted each candidate as the winner and believed each candidate was the actual winner.

Florida Election

TRUTH \ PREDICTION	BUSH	GORE	WAIT
BUSH	$\surd$	X	
GORE	X	$\surd$	

A CBS official offered that the network had thought the odds of being wrong when Gore was called the winner in Florida at about 1 in 200. Choosing a threshold for this error probability is a key decision variable. In the general scientific literature, results are accepted when this error probability is only 5%, or 1 chance in 20. When a poll indicates an error of $\pm 3\%$ or $\pm 4\%$, you should translate that to 1 chance in 20 that the poll is off by a greater margin. (This error probability is called the p-value in the jargon of the field.)

This election-calling situation is unusual in that both possible errors can be made to vanish simply by waiting. However, competitive pressure rewards risk-taking and early calls (that are correct!) among the various networks and news organizations.

Some disciplines require a much higher standard of proof. The FDA has a three-stage design that all new drugs must pass through before gaining approval. Clearly the cost of drugs is related to the delays caused by this cautious approach. The long-time equilibrium which has set the tradeoff between drug cost and drug safety has been disturbed by approval of new drugs for AIDS and the desire for a much speedier time-to-market. Decision theory is also useful for balancing multiple criteria.

The most stringent standard may be found in experimental physics, which requires the so-called "five sigma" standard, corresponding to a chance of error of only 1 in 3.5 million. Why so stringent? Well, millions of events are observed in colliders such as the one at Fermilab and the European supercollider at CERN. Indeed, John Bahcall, a particle physicist at Princeton, says in a recent *Science* article (Vol 289, p. 2260, 2000):

Half of all three-sigma results are wrong. Look at history.

The reason is clear. If 1000 experiments are run at the 5% level, you would expect 50 "findings" even if there is no true finding.

What is the current drama at CERN? Years of running experiments to find the Higgs boson have produced results that are significant at about the 3.5-sigma level. But after extending the experiment one month, the Italian director decided to pull the plug in favor of installing new equipment and moving on to other experiments. The Texas supercollider would surely have won the discovery prize (assuming, of course, the Higgs boson is real). A group of Rice Physicists and Statisticians had an important role in the discovery of the top quark at Fermilab a few years ago, using a new nonparametric technique we developed.

3 Fair Games

Fairness has a specific connotation in probability. A game is fair if the stake or cost to play the game is exactly equal to the expected winnings. A 50-50 chance at winning $2 is equivalent to $1 cash in hand. Thus $1 is the fair cost to play the game. However, if I offer you a 50-50 chance to win $1,000,000, you might strongly prefer to have the $500,000 cash (or even $250,000 cash). In the other direction, when a fresh Texas Lottery Game starts up, your expected winnings are only 23.2 cents, much less than the $1 price of a ticket. Yet many Texans enthusiastically play this "game." The odds of choosing all 6 lottery numbers are 1 in 25,827,165 and the smallest prize is $6,000,000. Even when the prize exceeds $100,000,000, the expected winnings are not much above $1 since multiple winners are more likely.

We won't pursue this interesting tangent into utility theory which would explain why individuals willingly take or avoid "fair bets." Gamblers are risk prone whereas most statisticians are risk averse.

4 Data

Collecting data separates our discipline from purely theoretical or philosophical approaches. While data has been collected for some time (recall a census in ancient Egypt or Rome), the analyses were not modern. Indeed, there is no evidence that the mathematicians of Greece or Arabia made the intellectual jump. The honor goes to a little-known London haberdasher, John Graunt, who analyzed the lists of the dead during the plague years of the mid-Seventeenth Century. These Bills of Mortality contained enough information to produce the first life or survival tables. A pre-statistical analysis might be a simple tabulation of the number deaths by cause. Graunt took these lists and re-organized the data into groups and tabulated the results by age. He also made some interpolations that were amateur. Today, we plot such curves as histograms. The following figure shows Graunt's death estimates in the form of a histogram. Graunt's amazing estimate is that 1 in 3 British subjects died before their sixth birthday.

Scientia's founder, Solomon Bochner, devoted much energy to the history of mathematics during his years at Rice. In his book entitled *The Role of Mathematics in the Rise of Science*, he lumps Graunt's work together with Sir William Petty, who has often been suggested to be the true genius behind Graunt's work. Bochner makes two observations that are relevant to the thesis today:

> ... *[they] imparted to our present-day civilization the tendency to statistical enumeration of anything and everything.*

and

> *Statistics* founded on probability *is perhaps the most exclusive characteristic of our civilization since 1600; and it would be difficult to find even a trace of it anywhere before.*

I can further demonstrate the difference between lists, counting, and histograms. I took the transcript of the 90-minute oral arguments this past Friday at the U.S. Supreme Court and listed the number of questions asked by each Justice. I also counted the number of lines of transcript for each question. The first figure shows a bar chart of the number of questions asked for each Justice, sorted by seniority. The less senior Justices do seem to ask fewer questions. However, if we "mine" deeper into these data and look at the fraction of total time each Justice asked questions, we see that the middle 3 Justices dominate the court. A pie chart shows this, too, but not as clearly. Indeed, the fact that Justice Thomas asked no questions could be missed. Finally, we do something a bit different, tabulating the histogram of how long each of the 183 individuals questions lasted. This histogram does not follow the famous bell curve. It also clearly highlights 3 very long questions, with the Justice indicated. Perhaps the significance of these three will be known by now, or soon.

What is the story of mortality today? Most Americans are aware that life expectancy is in the 70's, with women outliving men by a few years, on average. One of the important products of the U.S. census is a detailed decennial mortality table for the U.S. population. The 1997 data is shown in the figure. The average age is 76.1 years. But the details in this histogram deserve some study and understanding. The relative likelihood of death at a particular age is proportional to the height of the histogram curve at that age. Clearly the shape of this curve is also not the famous bell-shaped curve. Features of interest in the diagram include the magnitude of infant mortality, the peak at age 85, the rapid decline thereafter, and the clear excess mortality around age 20, details lost in a simple average summary.

What are the trends? Our next figure compares the 1997 data to the 1980 and 1960 data. Clearly things are improving. A closer look at ages from 0 to 10 shows how dramatically infant mortality has declined during the past 40 years. A closer look at the ages between 75 and 90 shows that the peak age has not increased in the past 20 years, although the probability of living beyond 80 has dramatically increased.

What is the detailed story of men compared to women? Our next figure reveals the answer. The mortality curve for women dominates the male curve at all ages. The average age of death is 79.4 for women but only 73.5 for men, nearly a 6 year difference.

However, if we examine the older years more closely we see that the most common age of death is quite later, 87 versus 82, a difference of 5 years. For those of you who are curious, I also show a plot of the probability of dying at each age (on a log-log scale). The excess mortality for ages from 15-25 is quite dramatic in this figure. The public perception of life expectancy in retirement is quite wrong (off by almost 10 years), which helps explain the difficulty appreciating the long-term problems facing Social Security and Medicare, topics we will return to in a moment.

5 Medical Insurance Premiums

Insurance is a simple statistical scheme that allows individuals to pool their uncertainty into an affordable proposition. Since the individuals are free to choose or decline coverage, most insurance purchased is perceived to be close to being a good value (i.e., “fair”).

Rice University is fairly typical in the rate structure it offers employees. Health insurance premiums are not age-indexed while life insurance premiums are age-indexed. Since Rice heavily subsidizes the actual cost of its health insurance, the fairness issue is not perceived to be in play. My belief is that such a flat system does not prepare individuals for the shock of retirement and coping with the Medicare system’s underpriced structure and the loss of subsidies for supplemental insurance. Most individuals of working age cope with steadily rising property taxes, but current health insurance practices do not prepare individuals for similar steady increases in health expenses. Many individuals routinely spend 25–40% of their total income on home mortgages, which hopefully are fully paid before retirement, but most individuals balk at spending such a similar high percentage on the most important issue during retirement, health. It is not “fair.”

If you go into the open insurance market, you will find insurance companies offering medical coverage with premiums that are adjusted for age, sex, smoking, and number of children. I have performed a small actuarial analysis of what Rice’s GMI/PPO insurance rates would look like if we followed the rate structure of the Unicare Insurance Company (www.unicare.com). How significant would the effect be on older employees, and how much difference would younger employees see? (Now that I am sliding into the “older” category, part of me is reluctant to be so fair as to describe these data.)

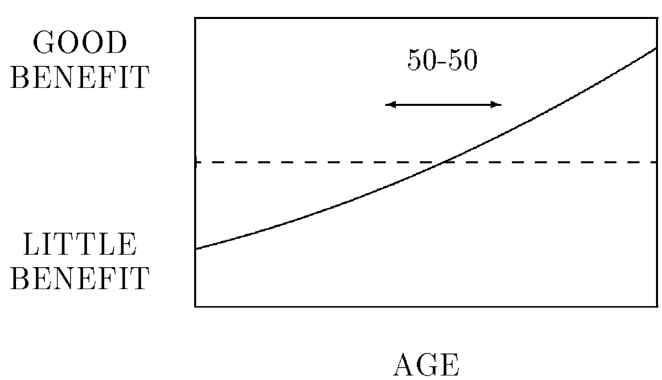
We examine the examples of two hypothetical employees from their twenties until death. One is a female staff member who has no dependents. The second marries after a few years and the family has three children at 2-year intervals. We examine their medical premiums under the 2000–2001 Rice program and under my estimate of an feasible age-indexed version (with the same total of premiums collected). In fact, Rice self-insures its GMI/PPO plan and attempts to beat commercial offerings with the broadest coverage and with no cap on lifetime benefit.

The workforce of Rice University mirrors the nation and is growing older, on average, year by year. This means that the rate increases are larger each year in part because the risk pool is older than average. One benefit if we had age-indexed premiums is that the premium increase of a 25-year-old is not affected by the number of 65-year-olds in the risk pool, and vice-versa.

Clearly, I find it fair to pool individuals by age. The age range is an interesting practical issue. Since access to health insurance even with preexisting conditions is now essentially guaranteed, one might argue that a flat premium independent of age is fair. Indeed, it is fair over one's entire lifetime. But the evidence is clear that many young adults who are self-employed and in good health prefer to take a risk rather than pay a high premium. This group of individuals had a large role in the failure of the Clinton's proposed universal health care in 1993. Time horizon matters. Pooling over 5 year intervals rather than 50 years is both politically acceptable and attracts young adults into choosing coverage immediately. Given the high correlation between age and income, there is no practical impediment to such an approach.

As an aside, I should note that the Unicare policies (unlike Rice) give different rates for men and women. Women's premiums are higher from age 21-50, but men's are higher from age 55-65; however, the difference is always less that 15%. Finally, smokers pay an additional 40% at all ages under Unicare.

Cost-benefit analyses are commonly used to determine coverage in medical plans. A number of recent studies have called into question the benefit of yearly pap smears and breast mammography in the 40–50 year-old group among healthy individuals with no particular family history. This is not surprising. Benefit-cost ratios typically grow slowly with age. If there is a clear benefit among older individuals and a clear non-benefit among younger individuals, then there must be a period of time for individuals in the middle where the benefits are not at all clear.



The new Pap smear protocol indicates the tradeoffs. If a woman has had 3 consecutive annual exams with no problems, then the frequency of exams may be reduced to once every 3 years. With annual exams, an extensive study found that 25 of every 10,000 women showed precancerous lesions. That number (25) increases to only 29 and 33 if the exams are taken every 2 and 3 years. The current advice is that each individual should decide what to do in consultation with one's physician.

I was very fortunate to spend my first three years as a faculty member working for Tony Gorry across the street at Baylor College of Medicine. I had some opportunities to do actuarial projects of internal interest to Baylor. One such project involved predicting the parking revenues for the building across from The Methodist Hospital. Our analysis showed that even a pessimistic estimate of the patient census after one year would provide more than enough parking garage revenue to make the project feasible. But of more relevance was Baylor's decision to self-insure for malpractice. Initially the plan was modeled after Aetna's policy, which places physicians in one of seven risk categories (GP no surgery to Brain Surgeon). When a decision was made to consider "simplifying" the plan by having only 4 risk categories, you could see the forces struggling with "fairness" at work.

6 Social Security

Part of the social security program is to provide a lifetime annuity. Individuals typically accumulate \$100,000-\$300,000 in their Social Security account which is loosely paid in an actuarial fashion. The system is not a good investment vehicle, barely returning 1%; however, the program remains a popular safety net. What if you could withdraw those funds and take them to a private insurance company for an annuity? We can use the same mortality tables for this analysis. To keep it simple, we suppose constant dollars and a real return on investment of 4%. In the following two figures, we see how the annual annuity increases depending on the age of retirement for each \$100,000 turned over to the insurance company. The second figure examines the fair annuity for men and women. Or is it fair to separate men and women? What about other factors at the time of retirement, such as smoking, ethnicity, current health status? The life expectancy of an individual retiring at age 65 is about 16 or 19 years, depending on sex. If you thought your life expectancy was significantly less, you might seriously consider managing your own funds and not buying into an annuity.

Some recent newspaper articles examine the Social Security crisis and what alternatives might improve its life expectancy. The problem is clear. Although the graph does not extend beyond the year the SS Trust fund hits a balance of zero, it clearly is heading south rapidly in 2037. Of more practical importance, the Social Security trust fund starts drawing down in 2020 (and in 2010 for Medicare). Social security annuities grow as retirement is deferred from age 65 to 70, a 5% jump each year. This exceeds the correct actuarial figure of half that. Perhaps the notion of a fixed retirement age could be replaced by a simple table of social security payments as a function of age and nest egg.

7 Election 2000

On November 23, William R. Brody, President of The Johns Hopkins University, wrote in an editorial:

> *While an army of lawyers and a few judges are likely to determine the outcome of this presidential election, I hope that we might engage an armada of statisticians before the next national campaign. They could conduct a comprehensive review of the various voting methods used across the country and the relative accuracy of each. Armed with facts — not opinions — we could then move expeditiously to bring the kinds of quality improvements to voting that we've seen occur in virtually every other facet of contemporary American life. Let's develop a fair and unbiased process that incorporates an understanding of the intrinsic errors of the voting process to prevent a repeat performance of Election 2000.*

I was initially relieved when the networks blamed their missteps on the "models" rather than on the statisticians. But the combination of shifting demographics and an ever-growing non-response rate have lessened the quality and reliability of our polls and predictions.

It is clear that exit poll results influence voters' behavior. Perhaps we will consider moving to a 6 p.m. EST closing with a 24-hour voting period and no exit poll reports. How best to standardize voting machines deserves careful thought. If we buy electronic touch screen machines, can we really do without any recount mechanism? Is there a risk of losing votes due to power failure? Will sufficient controls exist to discourage fraud?

8 Discussion

Whether events are truly random is an active matter of discussion. Is weather random or chaotic or deterministic? Clearly the modelers believe their equations are correct. What is interesting is that statisticians often introduce randomness *on purpose* to avoid biases caused by our ignorance of all relevant factors.

The events in Florida provide a critical data point for those trying to judge the outcome of related policy decisions such as adjusting the census and reforming the electoral college. My colleague, Jim Thompson, will be offering one of the Freshman seminar courses next semester, STAT 100, which will explore the data–policy link further.

The Federal Government is particularly prone to unfair insurance practices. Flood insurance rates that properly reflected the hurricane threat to coastal areas as well as the certain effects of rising seas would discourage unwise growth. However, there is a large constituency of coastal dwellers who would be unhappy about such a jump in insurance rates. Similarly, even in areas known to be high-risk for earthquakes. the Federal Government will be expected to pay hundreds of billions of dollars. Such huge subsidies deserve careful scrutiny.

Collecting data to strengthen decision-making is not always popular as it appears to dilute power. Ross Perot may have had many faults but he was quite correct in calling for pilot programs before committing to a new course. Problems that we have experienced on campus with the registrar's office and even details of the new language requirement could have been largely avoided by more incremental steps. On the positive side, the construction of the two new colleges is the result of a careful survey of the parents of prospective Rice students, who value a guarantee of of an option of on-campus housing for all four years.

On the other hand, there are numerous examples of surveys not able to foresee the big picture. Last year, Sony Corporation celebrated the 30th birthday of the Walkman. Sony CEO explained that repeated customer surveys revealed no interest in such a device, but he decided to take a chance. When a product is revolutionary, then surveys lag real opinion.

Eight years ago, the CEO of Motorola gave a talk at Rice in which he told a similar story about an idea of establishing a global network of 60 communication satellites for a new telephone system. Sadly, that decision has gone bad. Ironically, the CEO was not willing to take the risk that would have ensured the success of the endeavor, namely, going digital. The conservative decision to stick to analog technology doomed the project, with heavy and expensive phones.

What if we attain perfect information and perfect prediction so there is no risk. With no uncertainty, will there be no need for insurance? Ethicists will have to handle these cases.

In 1998, Senator Patrick Daniel Moynihan addressed the Joint Statistical Meetings where I was Program Chair. He reiterated two long-standing ideas he has advocated. The first was welfare reform, not only what has occurred during Clinton's terms, but more pro-family programs. The second was a call for a unified Federal statistical agency, bringing together the many pieces of the Federal statistical effort spread across many cabinet departments (Senate bill 1404 "The Federal Statistical System Act of 1997"). Surprisingly, the House passed such a bill this year, and many thought that the Senate would pass the bill to honor Senator Moynihan. I am told by a Bureau of Labor Statistics senior executive that a single letter from a worried AFL-CIO derailed the bill. Nevertheless, I believe the improvements to Federal data, including the census and current surveys, warrant continued study of this proposal.

I hope that you will each feel more inclined to dig a little deeper into actuarial work and try to understand the appropriateness of the subsidies that are in play.

Naturally, I think this topic is so important that every educated individual should study statistics. But I would like to close with some advice from General George S Patton, who wrote

> *Take calculated risks. That is quite different from being rash.*

Thank you very much.

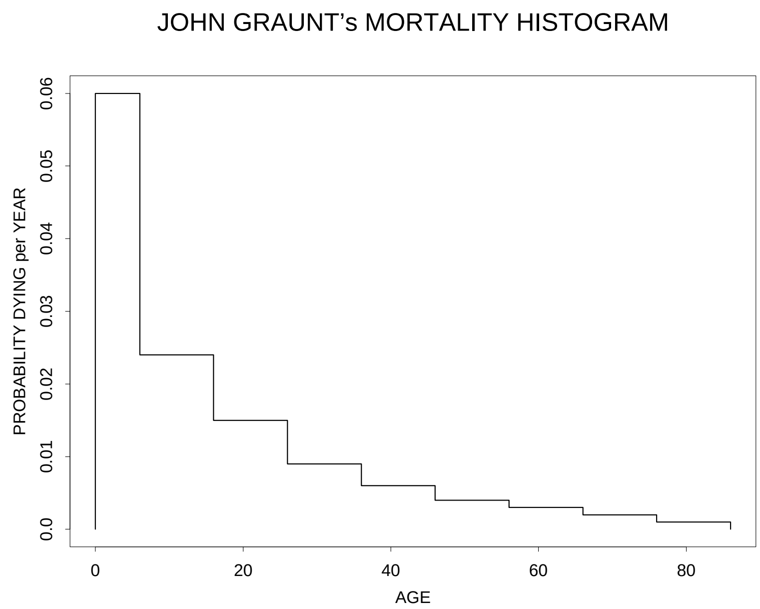


Figure 1: John Graunt's mortality table for London 1661.

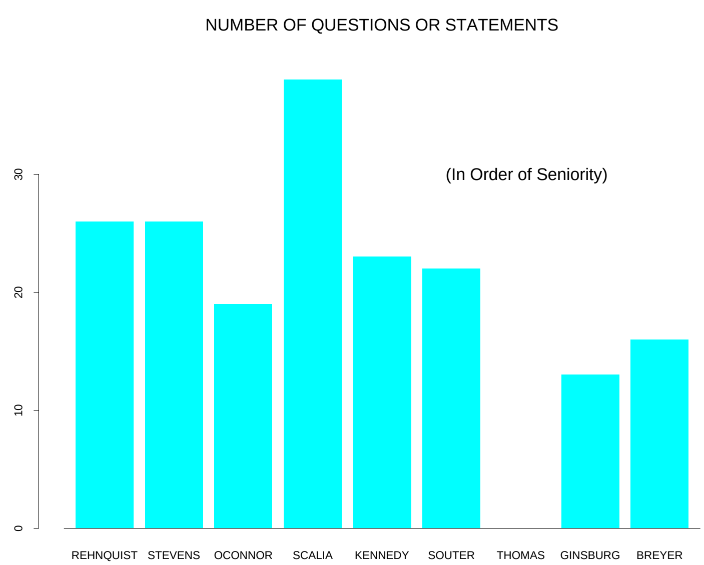


Figure 2: Number of questions asked by each Justice.

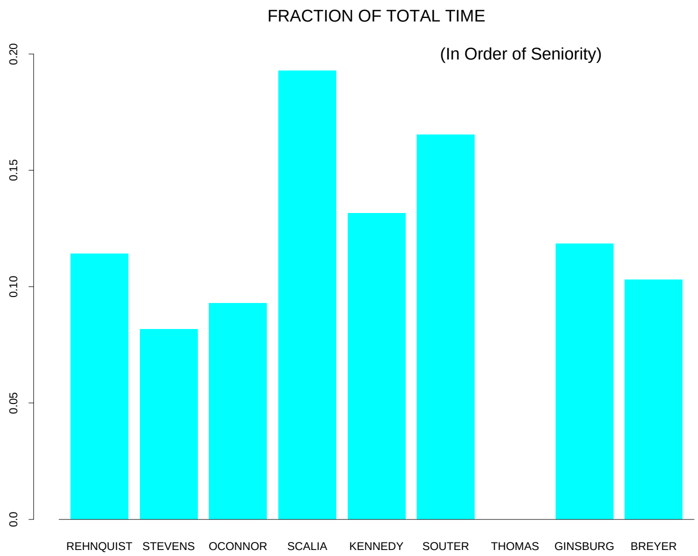


Figure 3: Fraction of time taken by each Justice.

Figure 4: Pie chart of fraction of time taken by each Justice.

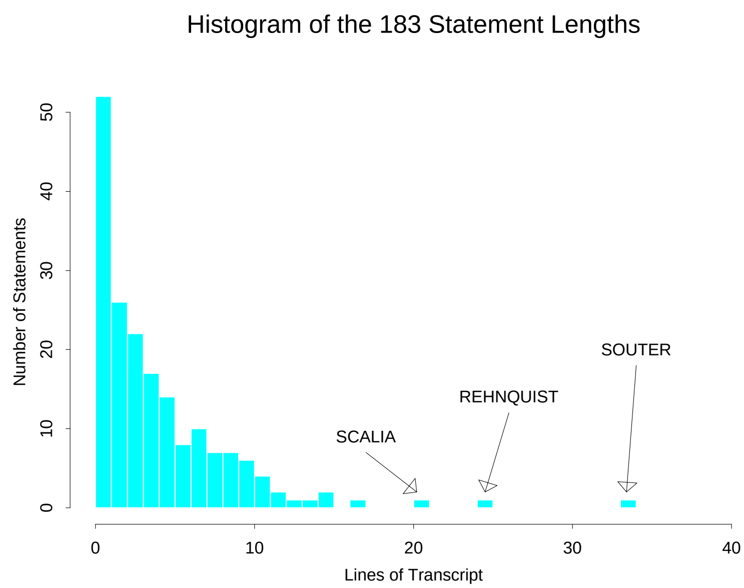


Figure 5: Histogram of length of all 183 questions.

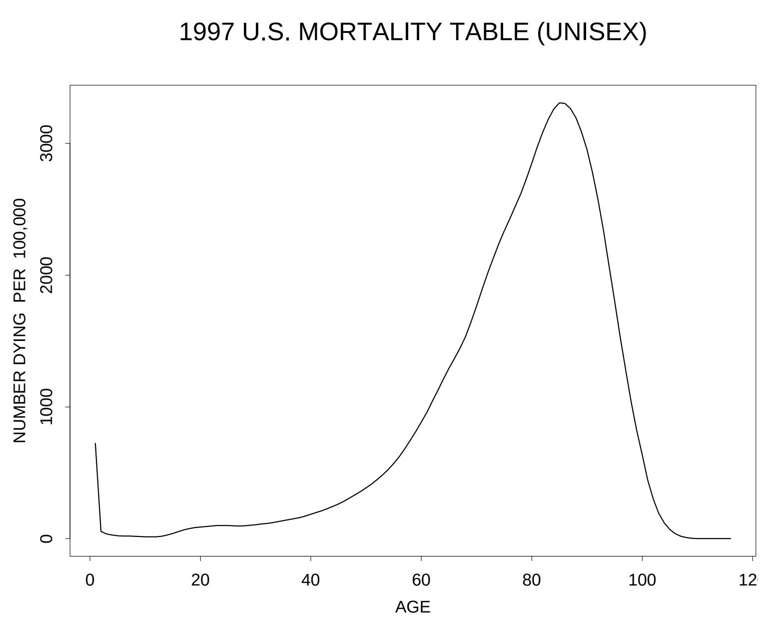


Figure 6: Unisex life table for 1997.

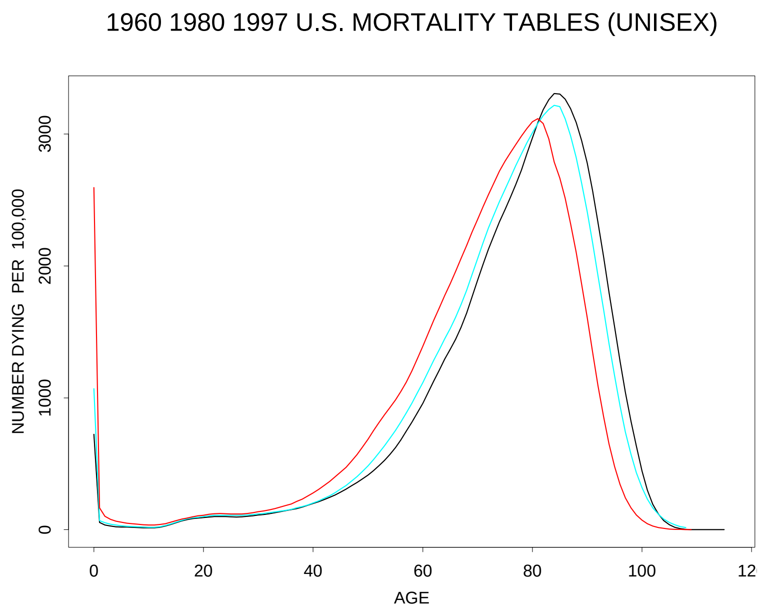


Figure 7: Unisex life tables for 1960, 1980, and 1997.

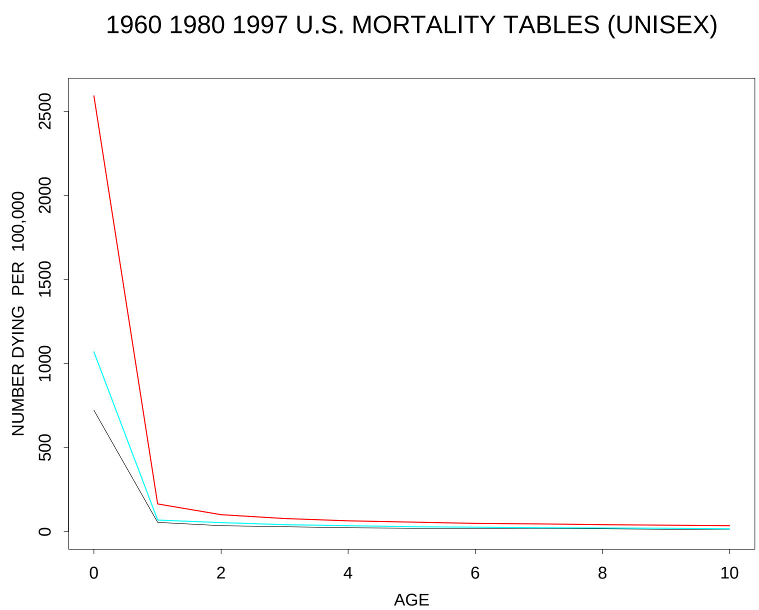


Figure 8: Blowup for ages 0–10.

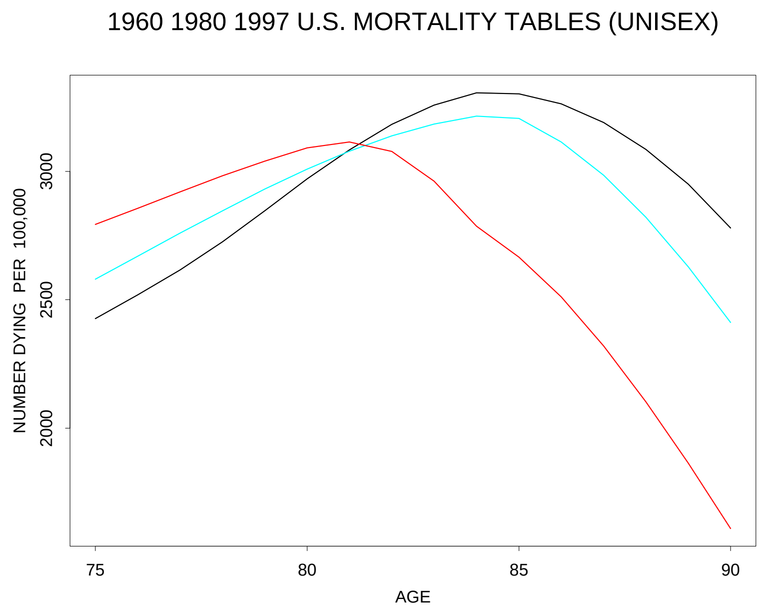


Figure 9: Blowup for ages 75–90.

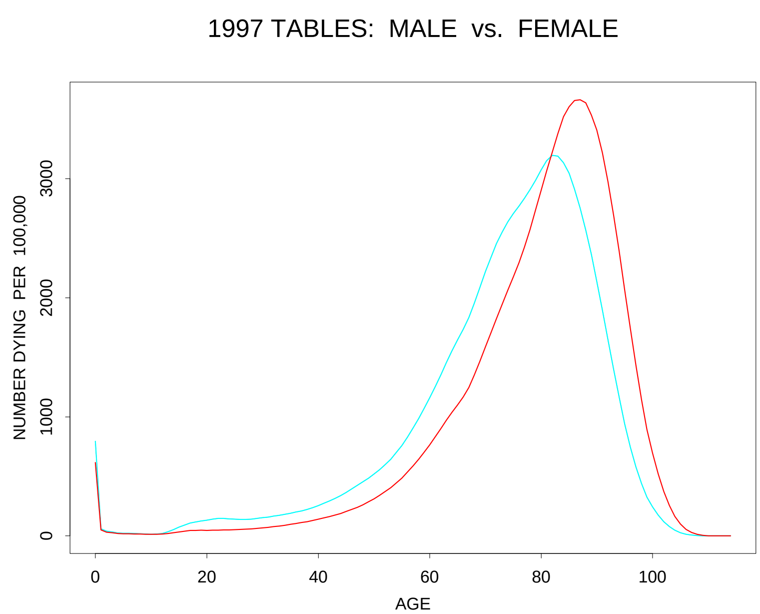


Figure 10: Male/female life table for 1997.

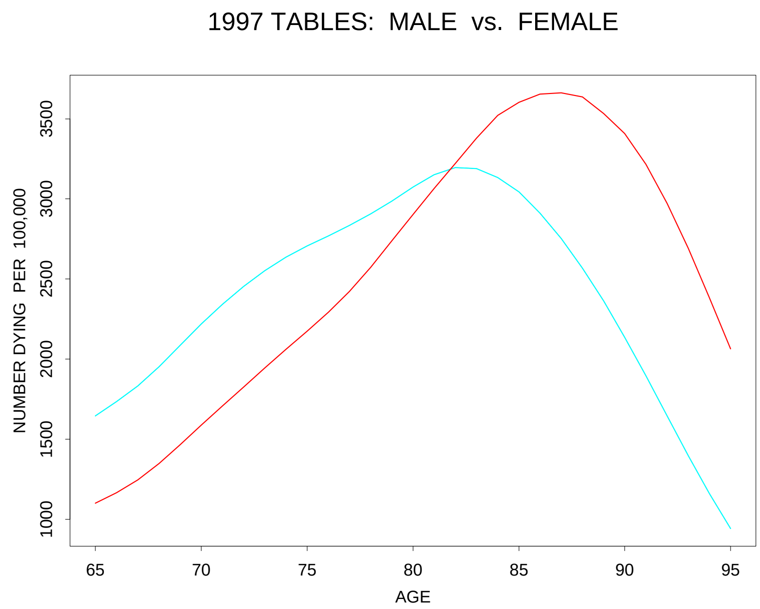


Figure 11: Blowup of years 65-95.

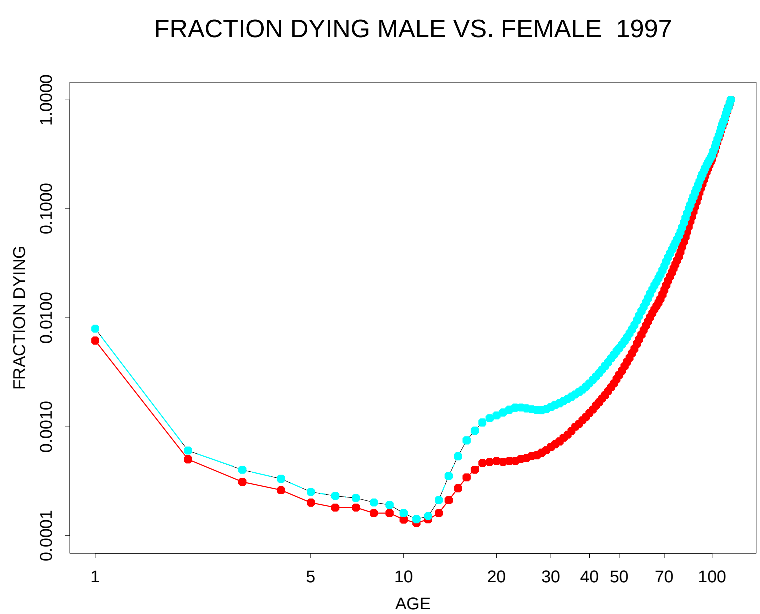


Figure 12: Male/female probability of dying by age.

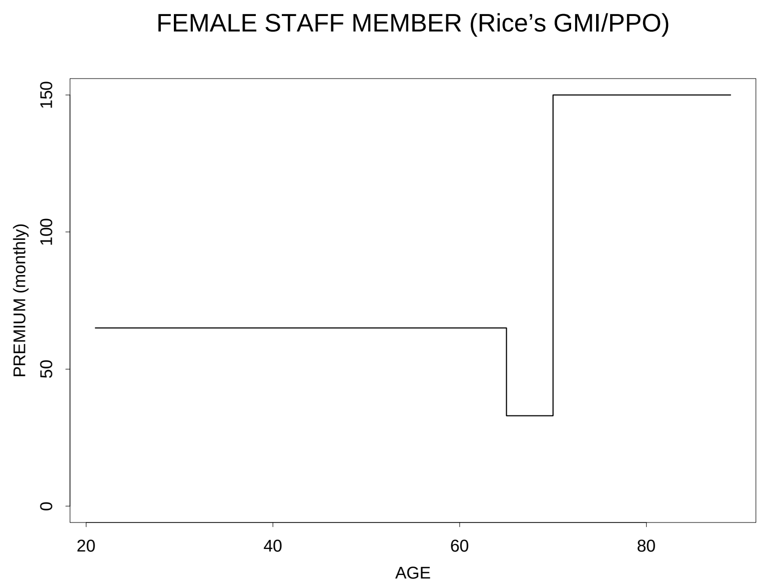


Figure 13: Single female's GMI/PPO premiums.

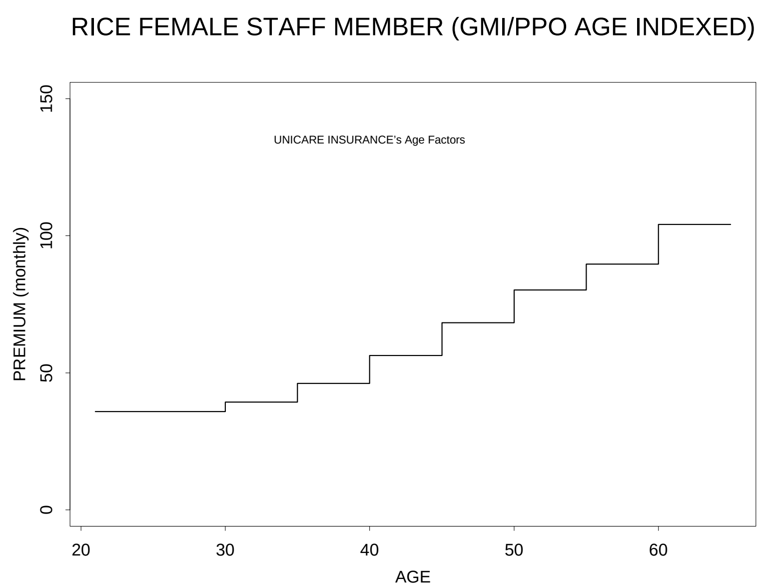


Figure 14: Single female's age-indexed GMI/PPO premiums.

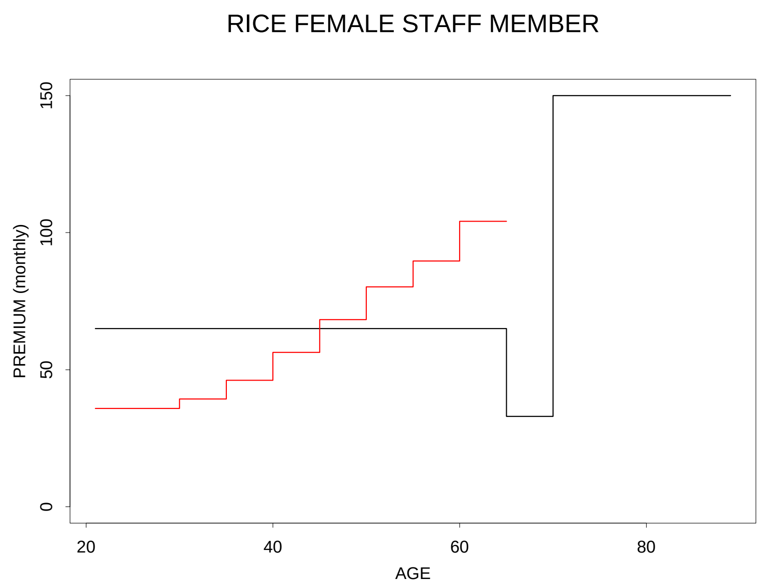


Figure 15: Overlay.

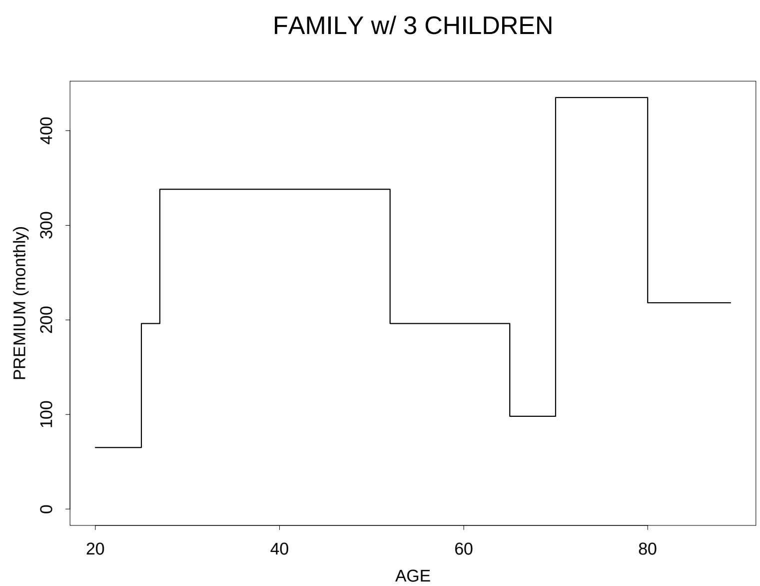


Figure 16: Family's GMI/PPO premiums.

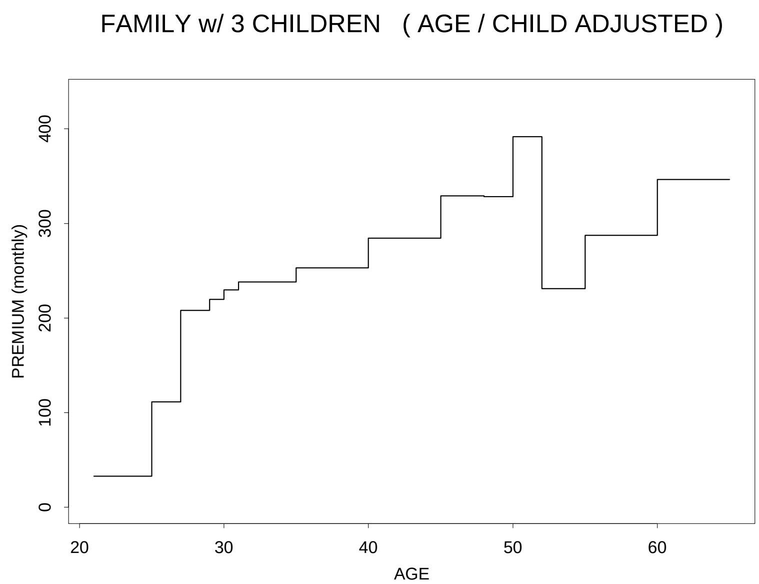


Figure 17: Family's adjusted GMI/PPO premiums.

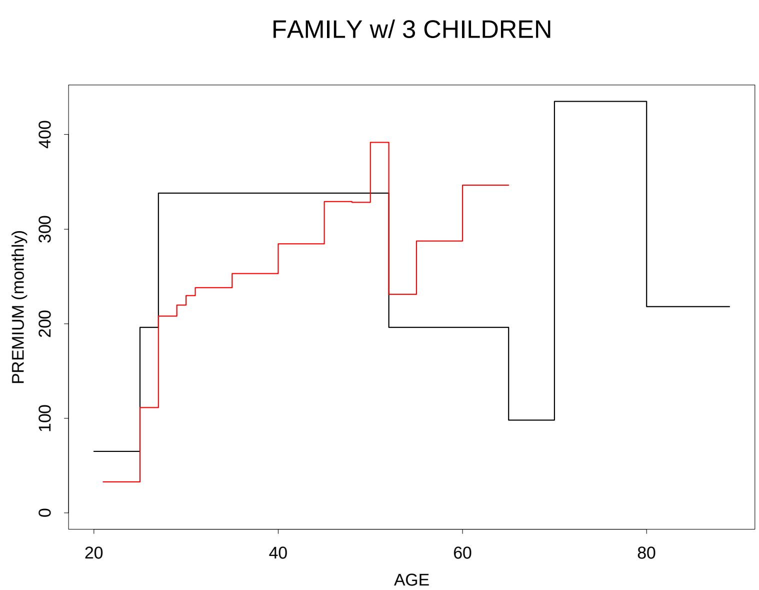


Figure 18: Overlay.