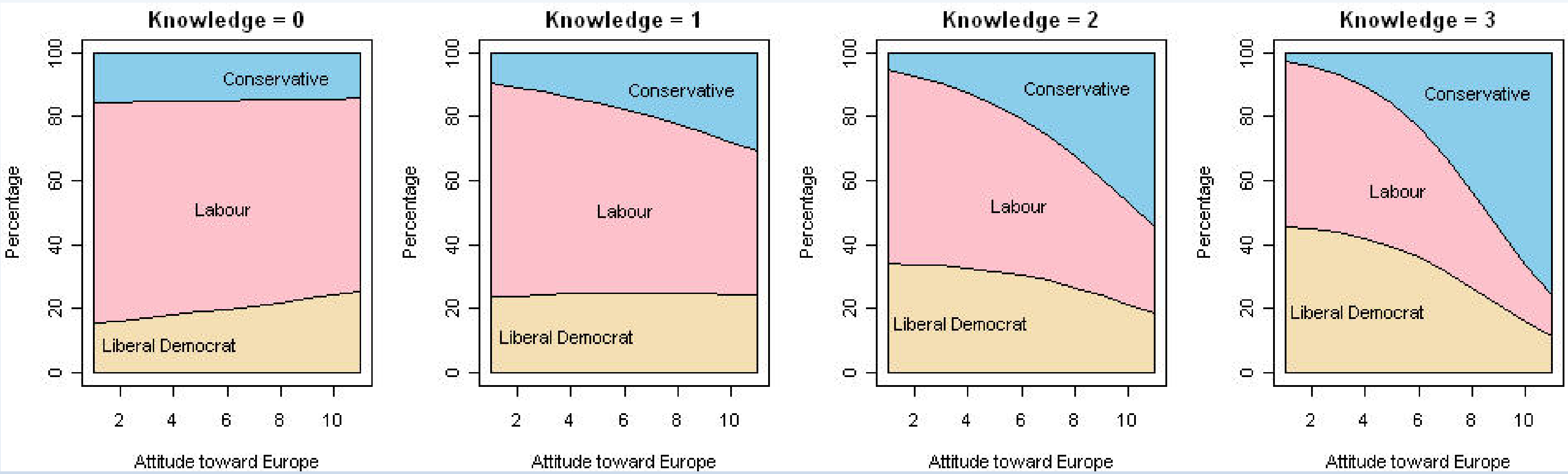


Logistic regression: Extensions

Michael Friendly

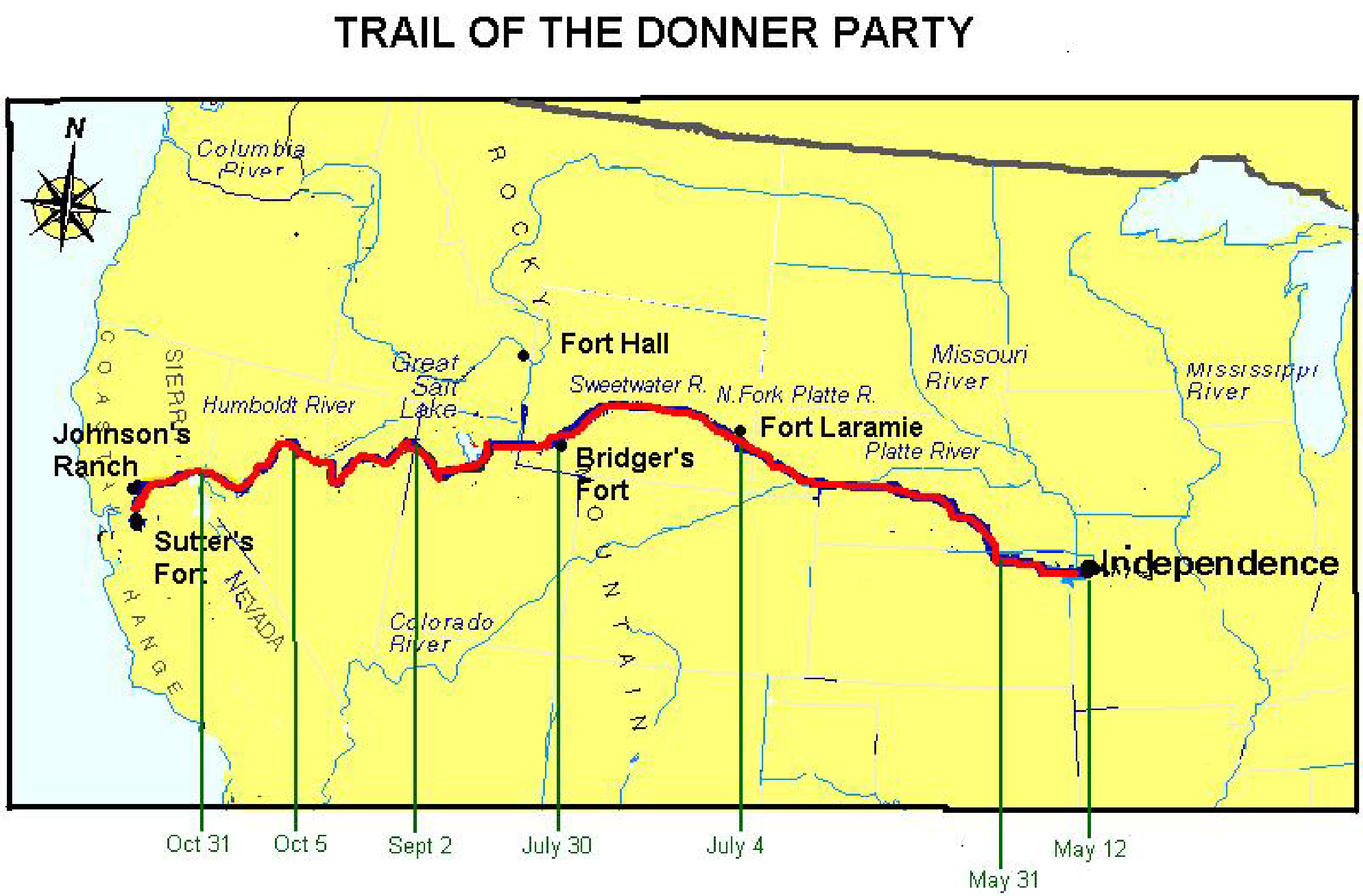
Psych 6136

http://friendly.github.io/psy6136

Donner party: A graphic tale of survival & influence

History:

- Apr—May, 1846: Donner/Reed families set out from Springfield, IL to CA
- July: Reach Bridger's Fort WY: 87 people, 23 wagons

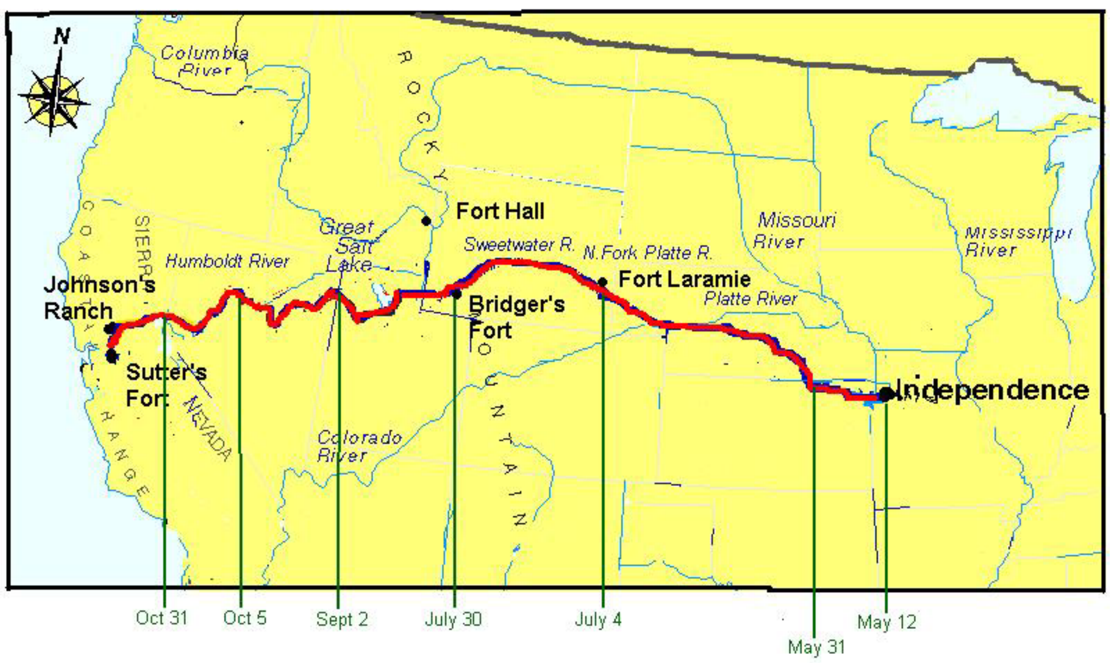


Donner party: A graphic tale of survival & influence

History:

- "Hastings cutoff": an untried route through Salt Lake desert (90 people)
- Worst recorded winter: Oct 31 blizzard; stranded at Truckee Lake (nr Reno)
  - Rescue parties sent out ("Dire necessity", "Forelorn hope", …)
  - Relief parties from CA: 42 survivors (Mar—Apr 1847)

TRAIL OF THE DONNER PARTY



Who lived? Who died?

Can we explain w/ logistic regression?

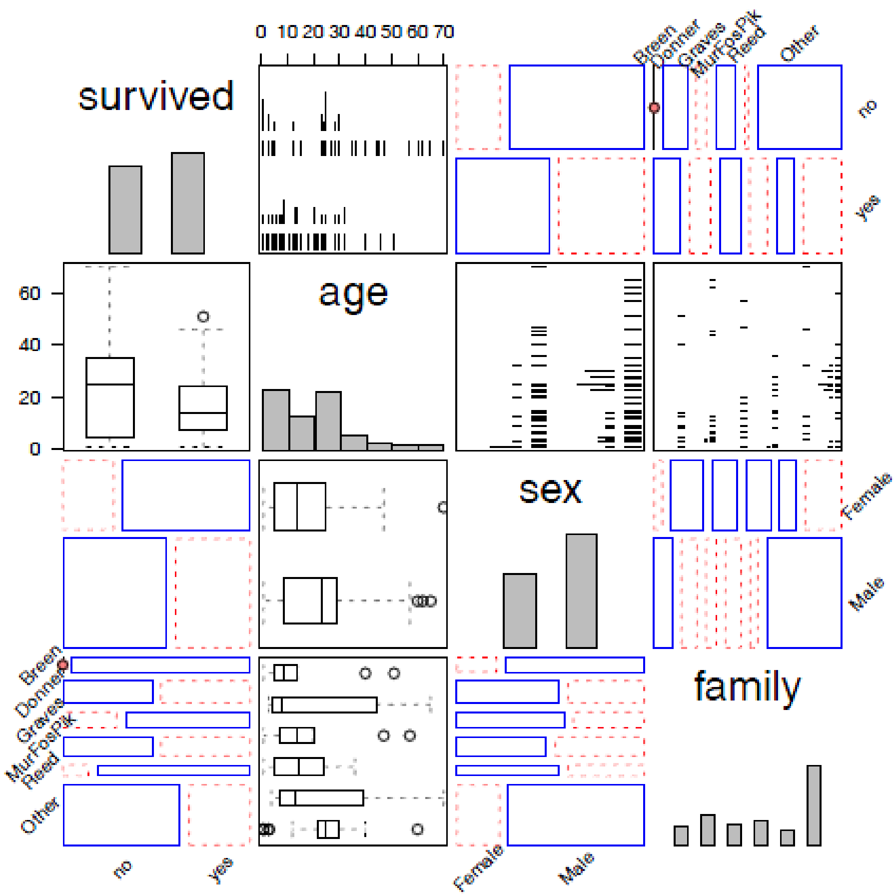
Donner party: Data

```
> data("Donner", package="vcdExtra")
> Donner$survived <- factor(Donner$survived,
                            labels=c("no", "yes"))

> car::some(Donner, 8)
                       family age    sex survived      death
Breen, Peter            Breen   3   Male      yes       <NA>
Donner, Jacob          Donner  65   Male       no 1846-12-21
Foster, Jeremiah    MurFosPik   1   Male       no 1847-03-13
Graves, Nancy          Graves   9 Female      yes       <NA>
McCutchen, Harriet McCutchen   1 Female       no 1847-02-02
Reed, James              Reed  46   Male      yes       <NA>
Reinhardt, Joseph       Other  30   Male       no 1846-12-21
Wolfinger, Doris     FosdWolf  20 Female      yes       <NA>
```

Overview: a gpairs() plot

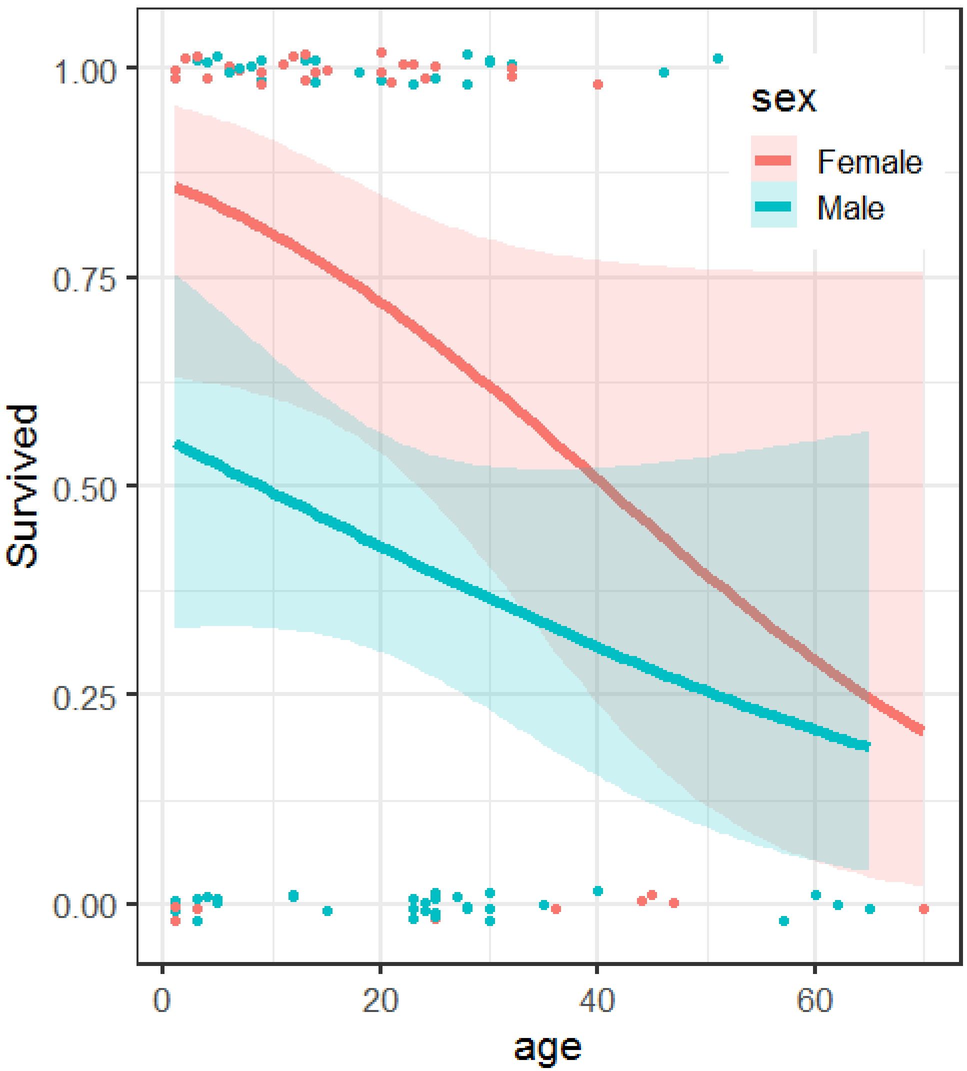
A generalized pairs plot uses different plot types for pairs of continuous, discrete variables



- Binary response: `survived`
- Categorical predictors: `sex`, `family`
- Quantitative predictor: `age`
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?

Exploratory plots

Before fitting models, it is useful to explore the data with conditional ggplots



Survival decreases with age for both men and women

Women more likely to survive, particularly the young

Conf. bands show the data is thin at older ages

Using ggplot

Basic plot: survived vs. age, colored by sex, with jittered points

```
gg <- ggplot(Donner,
             aes(age, as.numeric(survived=="yes"), color=sex)) +
  ylab("Survived") +
  geom_jitter(height = 0.02, width = 0)
```

To this we can add conditional logistic fits using **`stat_smooth(method="glm")`**

```
gg + stat_smooth(method = "glm",
                 method.args = list(family = binomial),
                 formula = y ~ x,
                 alpha = 0.2, size=2, aes(fill = sex)) +
  theme_bw(base_size = 16) +
  theme(legend.position = c(.85, .85))
```

Questions

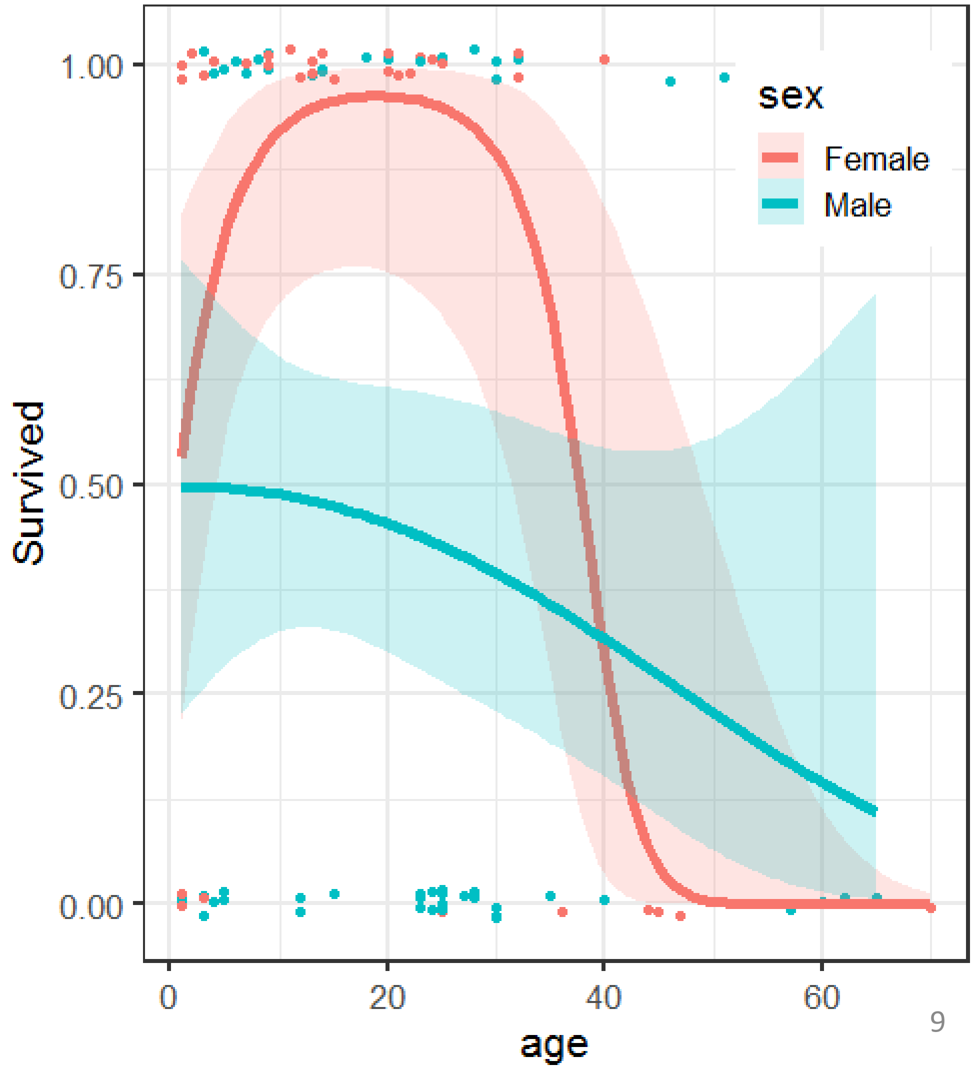
- Is the relation of survival to age well expressed as a linear logistic regression model?
  - Allow a quadratic or higher power using poly(age,2), poly(age,3)

$$
\begin{aligned}
\text{logit}(\pi_i) &= \alpha + \beta_1 x_i + \beta_2 x_i^2 \\
\text{logit}(\pi_i) &= \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3
\end{aligned}
$$

  - Use natural spline functions: ns(age, df)
  - Use non-parametric smooths: loess(age, span, degree)
- Is the relation the same for men & women?
  - Allow an interaction of sex * age or sex * f(age)
  - Test goodness of fit relative to the main effects model

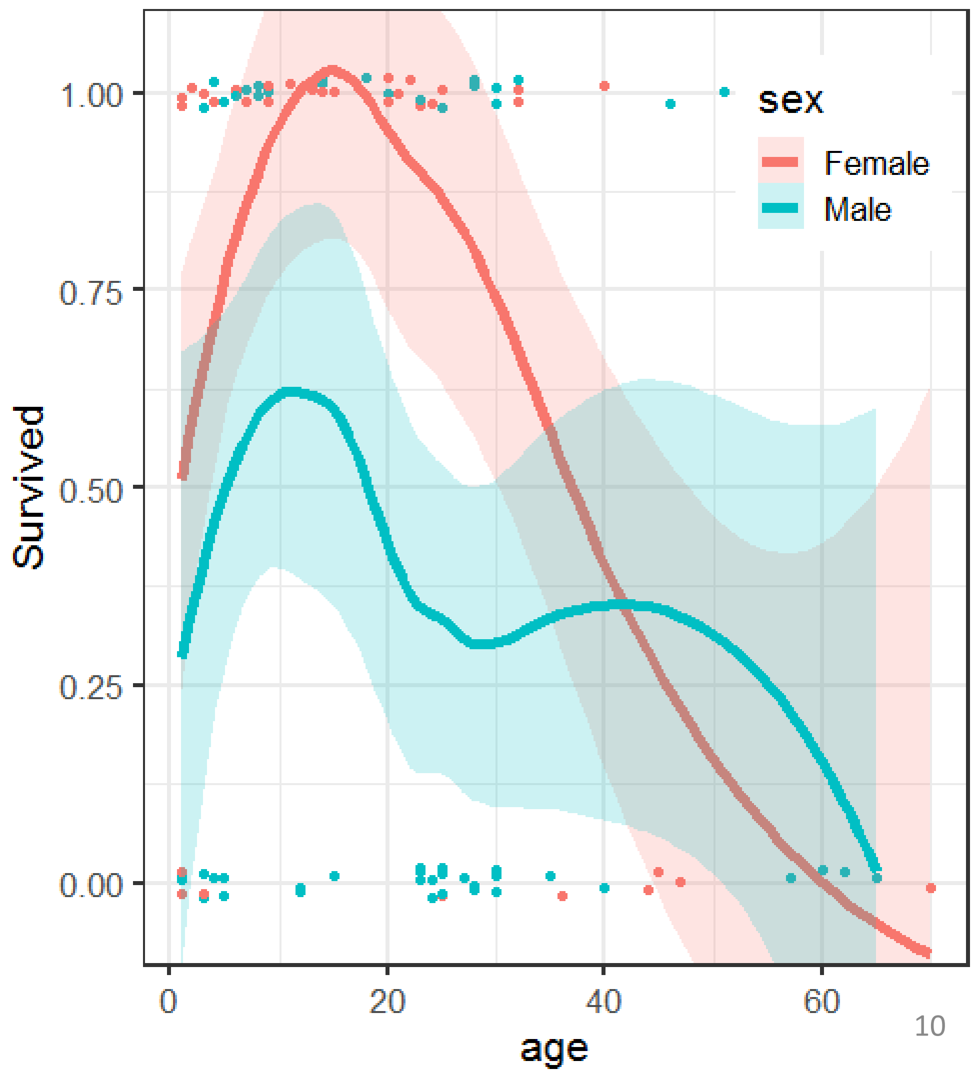
```
gg + stat_smooth(method = "glm",
          method.args = list(family = binomial),
          formula = y ~ poly(x,2), alpha = 0.2, size=2, aes(fill = sex)) + …
```

Fit separate quadratics for M & F



```
gg + stat_smooth(method = "loess", span=0.9,
            alpha = 0.2, size=2,
            aes(fill = sex)) + coord_cartesian(ylim=c(-.05,1.05)) +
```

Fit separate loess smooths for M & F



Fitting models

Models with linear effect of age:

```
> donner.mod1 <- glm(survived ~ age + sex,
                      data=Donner, family=binomial)
> donner.mod2 <- glm(survived ~ age * sex,
                       data=Donner, family=binomial)
> Anova(donner.mod2)
Analysis of Deviance Table (Type II tests)

Response: survived
        LR Chisq Df Pr(>Chisq)
age         5.52  1     0.0188 *
sex         6.73  1     0.0095 **
age:sex     0.40  1     0.5269
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fitting models

Models with quadratic effect of age:

```
> donner.mod3 <- glm(survived ~ poly(age,2) + sex,
                     data=Donner, family=binomial)
> donner.mod4 <- glm(survived ~ poly(age,2) * sex,
                     data=Donner, family=binomial)
> Anova(donner.mod4)
Analysis of Deviance Table (Type II tests)

Response: survived
                 LR Chisq Df Pr(>Chisq)
poly(age, 2)         9.91  2     0.0070 **
sex                  8.09  1     0.0044 **
poly(age, 2):sex     8.93  2     0.0115 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing models

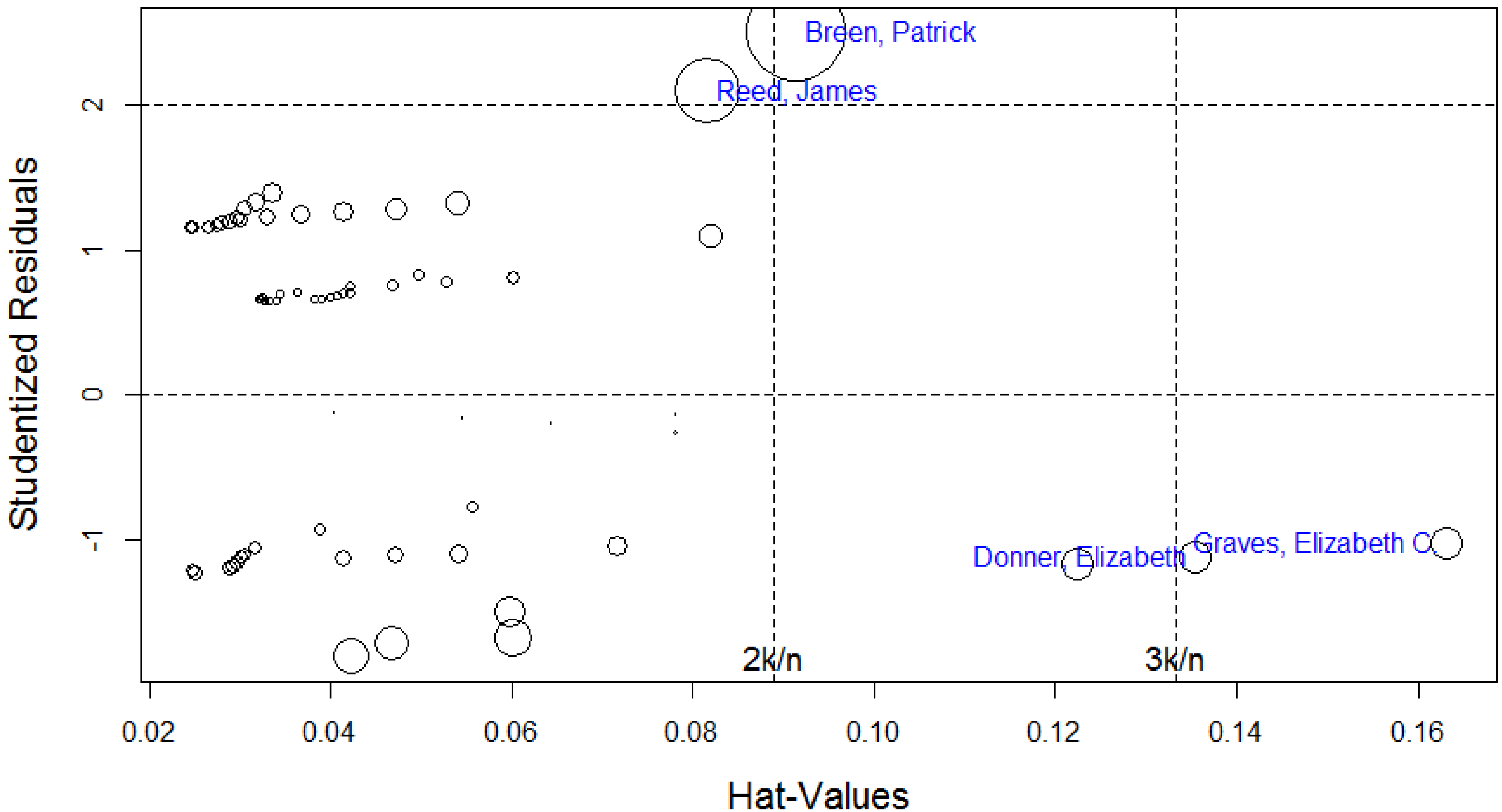
These models are only nested in pairs. We can compare them using AIC & $\Delta\chi^2$

```
> library(vcdExtra)
> LRstats(donner.mod1, donner.mod2, donner.mod3, donner.mod4)
Likelihood summary table:
            AIC BIC LR Chisq Df Pr(>Chisq)
donner.mod1 117 125    111.1 87      0.042 *
donner.mod2 119 129    110.7 86      0.038 *
donner.mod3 115 125    106.7 86      0.064 .    ✓
donner.mod4 110 125     97.8 84      0.144
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	linear	non-linear	$\Delta\chi^2$	p-value	
additive	111.128	106.731	4.396	0.036	✓
non-additive	110.727	97.799	12.928	0.000	✓
$\Delta\chi^2$	0.400	8.932			
p-value	0.527	0.003			
		✓			

Who was influential?

```
res <- influencePlot(donner.mod3, id = list(col="blue", n=2), scale=8)
```

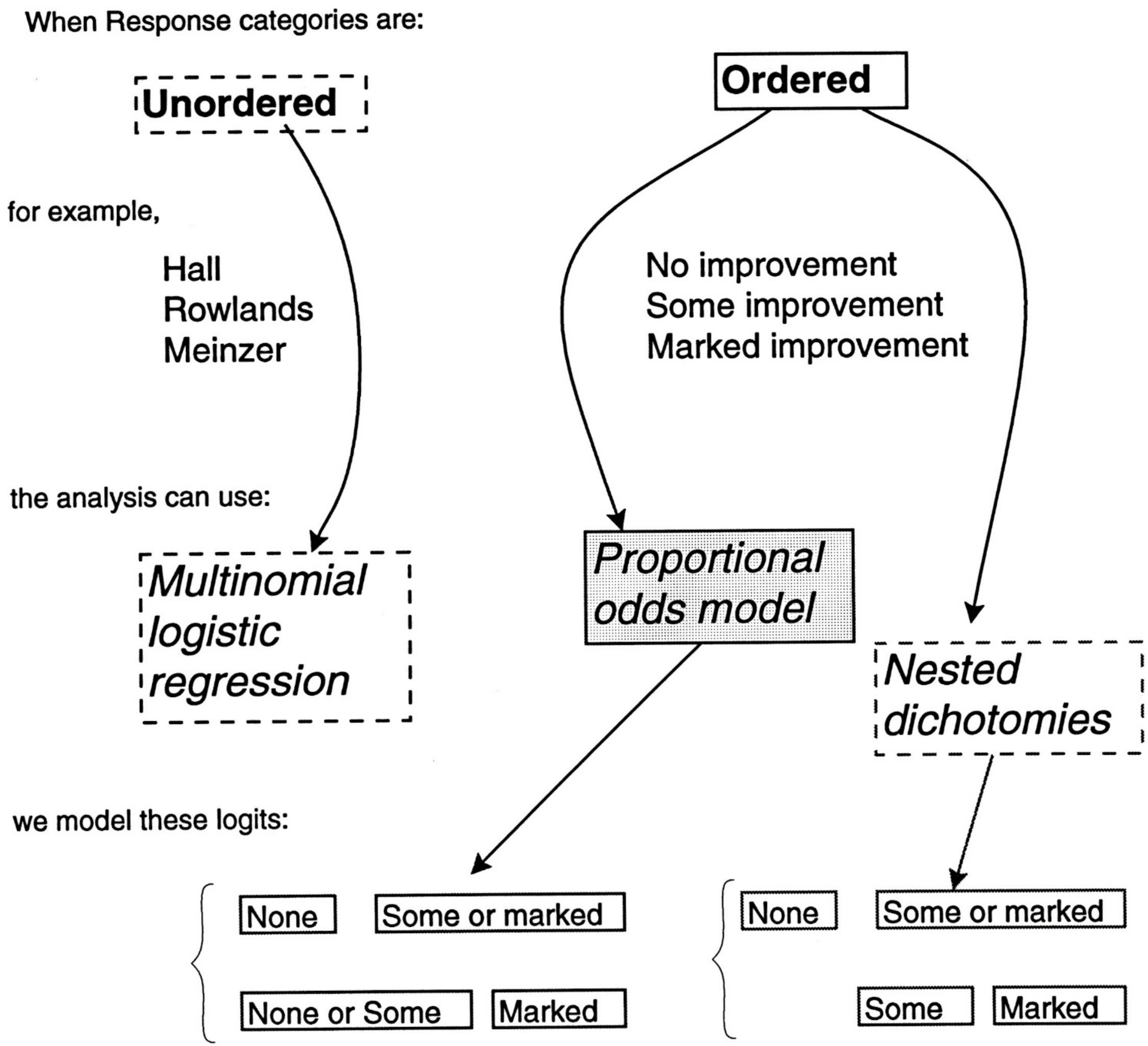


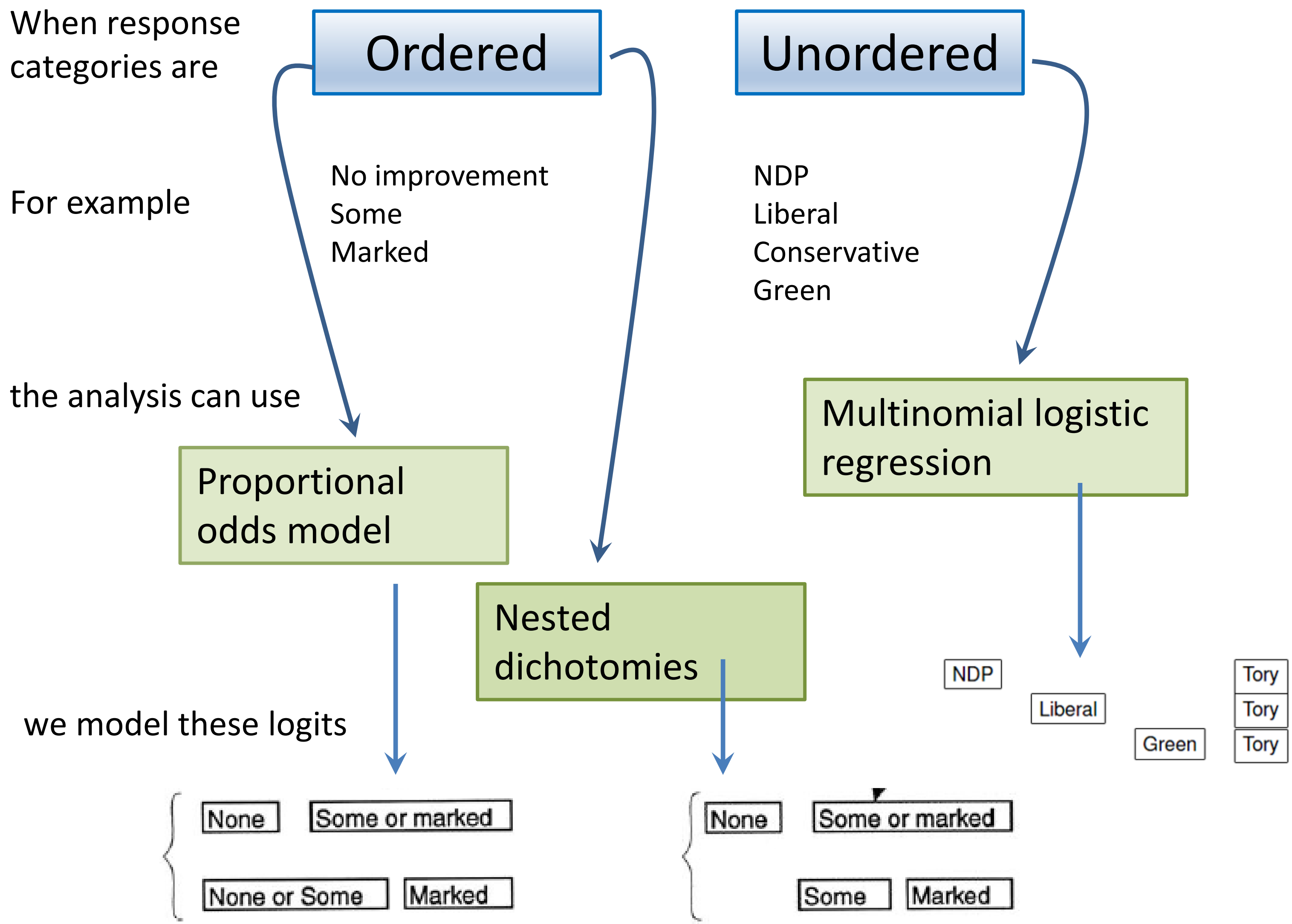
Why were they influential?

```
> idx <- which(rownames(Donner) %in% rownames(res))
> # show data together with diagnostics
> cbind(Donner[idx,2:4], res)
                     age    sex survived StudRes    Hat  CookD
Breen, Patrick        51   Male      yes    2.50 0.0915 0.3235
Donner, Elizabeth     45 Female       no   -1.11 0.1354 0.0341
Graves, Elizabeth C.  47 Female       no   -1.02 0.1632 0.0342
Reed, James           46   Male      yes    2.10 0.0816 0.1436
```

- Patrick Breen, James Reed: Older men who survived
- Elizabeth Donner, Elizabeth Graves: Older women who died
- Moral lessons of this story:
  - Don't try to cross the Donner Pass in late October; if you do, bring lots of food
  - Plots of fitted models show *only* what is included in the model
  - Discrete data often need smoothing (or non-linear terms) to see the pattern
  - Always examine model diagnostics — preferably graphic

Polytomous responses: Overview




When response categories are
Ordered
Unordered
For example
No improvement
Some
Marked
NDP
Liberal
Conservative
Green
the analysis can use
Proportional odds model
Nested dichotomies
Multinomial logistic regression
we model these logits
None
Some or marked
None or Some
Marked
None
Some or marked
Some
Marked
NDP
Liberal
Green
Tory
Tory
Tory

Polytomous responses: Ordered

Polytomous responses

- *m* categories → (*m*-1) comparisons (logits)
- One part of the model for each logit
- Similar to ANOVA where an *m*-level factor → (*m*-1) contrasts (df)

Ordered response categories, e.g., None, Some, Marked improvement

- Proportional odds model
  - Uses adjacent-category logits

None	Some or Marked
None or Some	Marked

  - Assumes slopes are equal for all $m - 1$ logits; only intercepts vary
  - R: `polr()` in MASS
- Nested dichotomies

None	Some or Marked
	Some | Marked

  - Model each logit separately
  - G^2 s are additive → combined model

Polytomous responses: Unordered

Unordered response categories, e.g., vote: NDP, Liberal, Green, Tory

- Multinomial logistic regression
  - Fits $m - 1$ logistic models for logits of category $i = 1, 2, \ldots m - 1$ vs. category m
  - e.g.,

NDP	Tory
Liberal	Tory
Green	Tory

  - This is the most general approach
  - R: `multinom()` function in nnet
- Can also use nested dichotomies

Proportional odds model

Arthritis treatment data:

		Improvement			
Sex	Treatment	None	Some	Marked	Total
F	Active	6	5	16	27
F	Placebo	19	7	6	32
M	Active	7	2	5	14
M	Placebo	10	0	1	11

The proportional odds model uses logits for (m-1) = 2 adjacent category cutpoints

$$\text{logit}\,(\theta_{ij1}) = \log \frac{\pi_{ij1}}{\pi_{ij2} + \pi_{ij3}} = \text{logit (None vs. [Some or Marked])}$$

$$\text{logit}\,(\theta_{ij2}) = \log \frac{\pi_{ij1} + \pi_{ij2}}{\pi_{ij3}} = \text{logit ([None or Some] vs. Marked)}$$

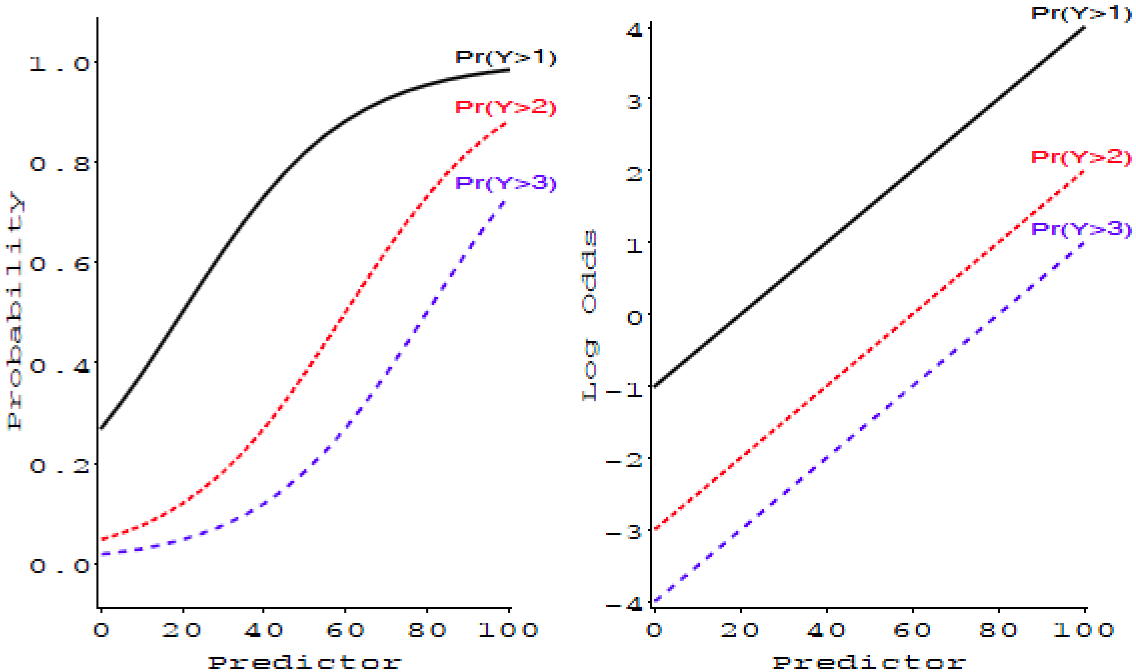
- Consider a logistic regression model for each logit:

$$\text{logit}(\theta_{ij1}) = \alpha_1 + \boldsymbol{x}_{ij}' \beta_1 \qquad \text{None vs. Some/Marked}$$

$$\text{logit}(\theta_{ij2}) = \alpha_2 + \boldsymbol{x}_{ij}' \beta_2 \qquad \text{None/Some vs. Marked}$$

- Proportional odds assumption: regression functions are parallel on the logit scale i.e., $\beta_1 = \beta_2$.

Proportional Odds Model



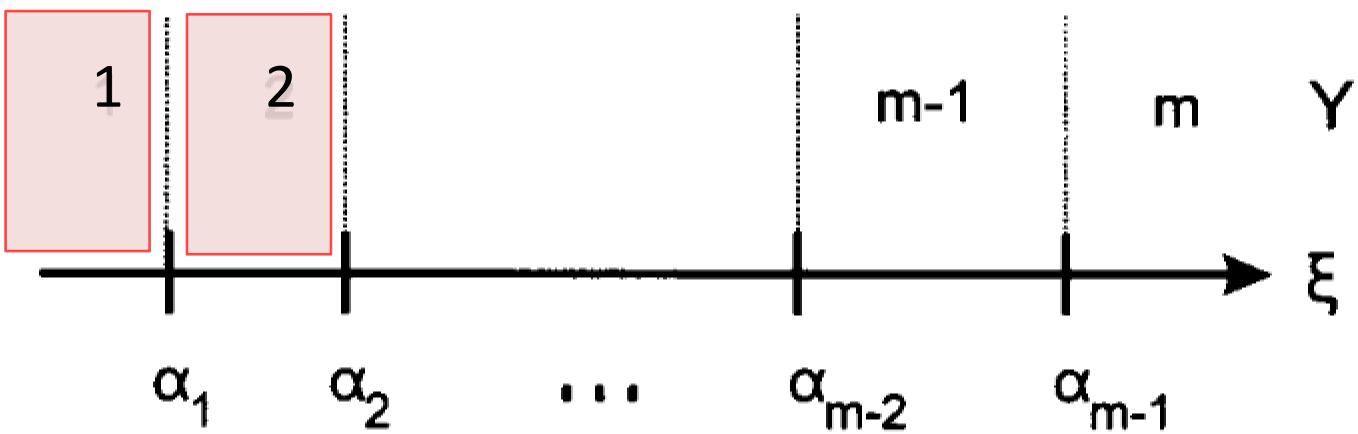
Proportional odds: Latent variable interpretation

A simple motivation for the proportional odds model:

- Imagine a continuous, but *unobserved* response, ξ, a linear function of predictors

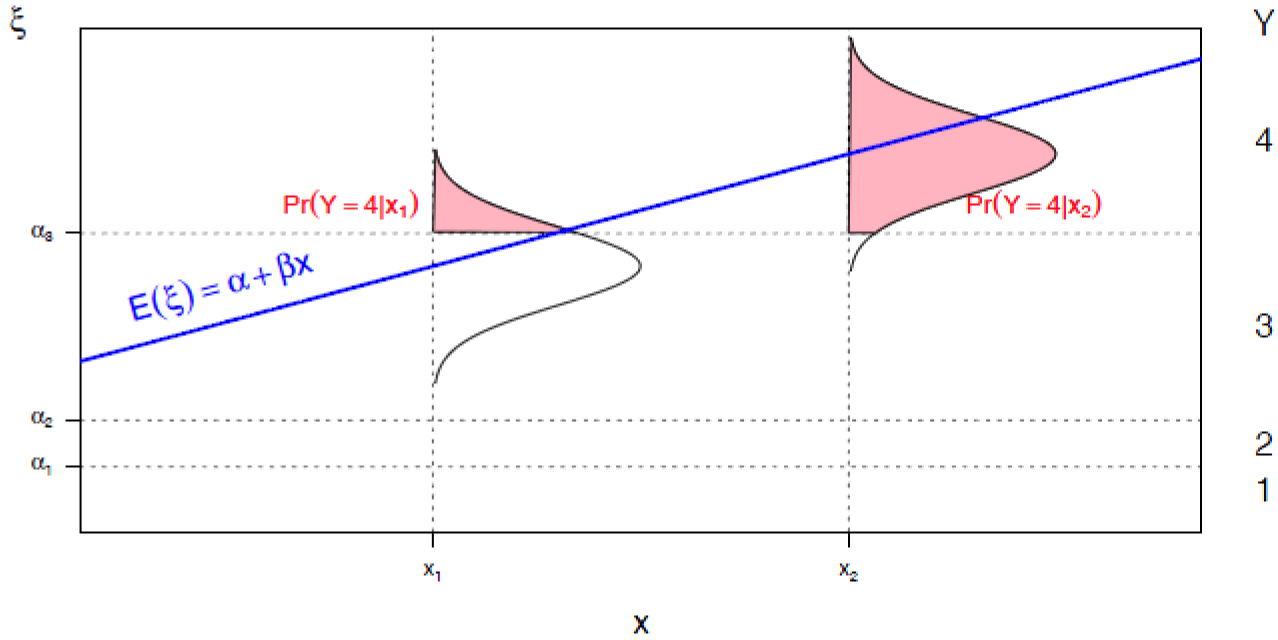
$$\xi_i = \beta^{\mathsf{T}} \boldsymbol{x}_i + \epsilon_i$$

- The *observed* response, Y, is discrete, according to some *unknown* thresholds, $\alpha_1 < \alpha_2, < \cdots < \alpha_{m-1}$
- That is, the response, $Y = i$ if $\alpha_i \leq \xi_i < \alpha_{i+1}$
- Thus, intercepts in the proportional odds model $\sim$ thresholds on ξ



Proportional odds: Latent variable interpretation

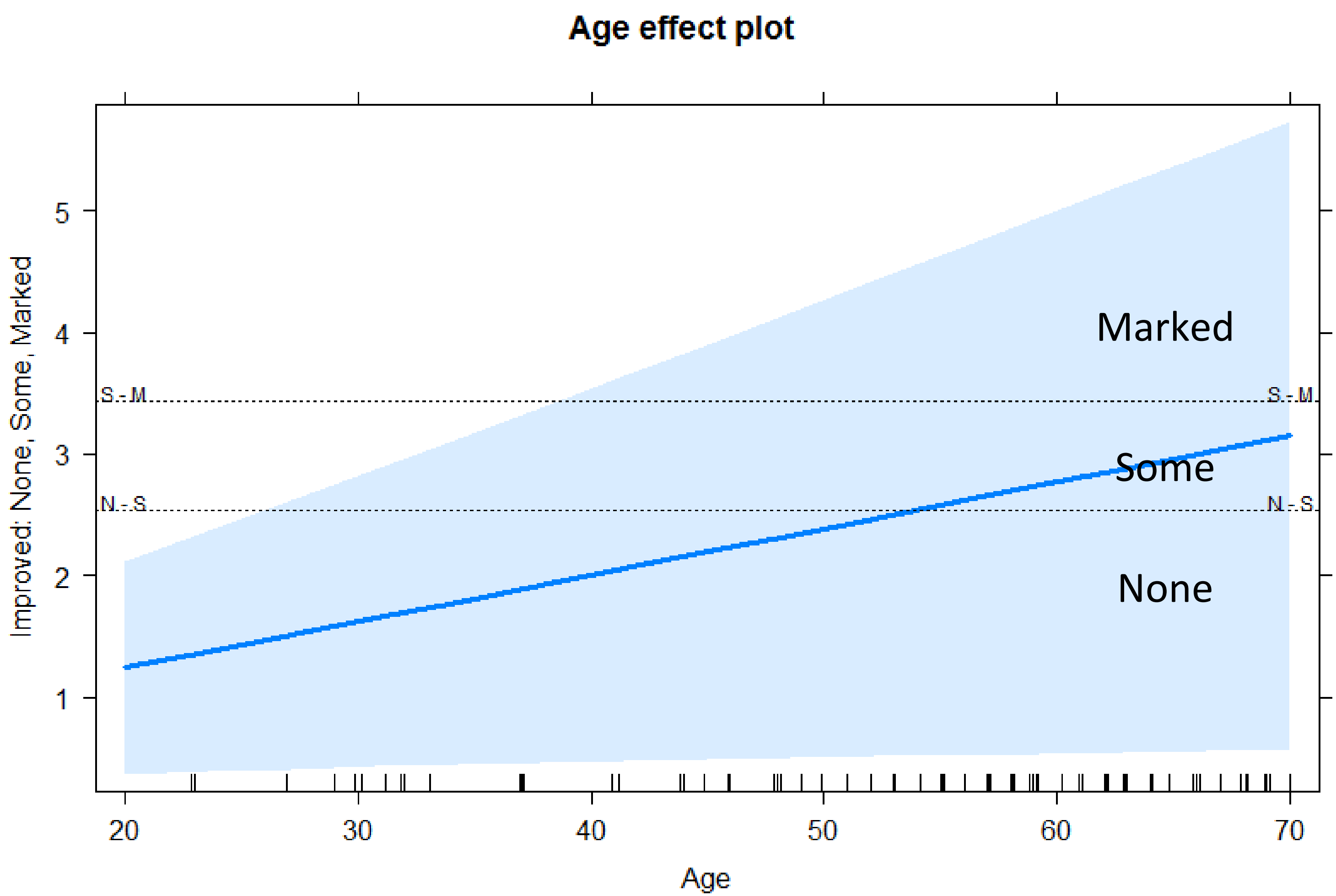
We can visualize the relation of the latent variable ξ to the observed response Y, for two values, x_1 and x_2, of a single predictor, X as shown below:



Proportional odds: Latent variable interpretation

Plotting the effect of Age on the latent variable scale

```
plot(effect("Age", mod = arth.polr, latent = TRUE))
```



Fitting the proportional odds model

The response Improved has been defined as an ordered factor

```
> data(Arthritis, package = "vcd")
> head(Arthritis$Improved)
[1] Some   None   None   Marked Marked Marked
Levels: None < Some < Marked
```

Fit the model with MASS::polr()

```
> arth.polr <- polr(Improved ~ Sex + Treatment + Age,
                    data = Arthritis)

> summary(arth.polr)    # for coefficients
> Anova(arth.polr)      # Type II tests
```

summary() gives the standard statistical results

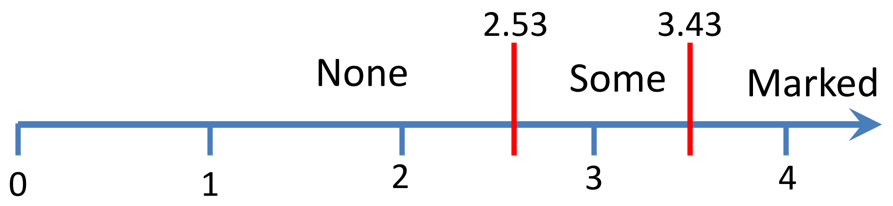
```
> summary(arth.polr)     # for coefficients

Call:
polr(formula = Improved ~ Sex + Treatment + Age, data = Arthritis)

Coefficients:
                   Value Std. Error t value
SexMale          -1.2517     0.5464   -2.29
TreatmentTreated  1.7453     0.4759    3.67
Age               0.0382     0.0184    2.07

Intercepts:
            Value  Std. Error t value
None|Some    2.532  1.057       2.395
Some|Marked  3.431  1.091       3.144

Residual Deviance: 145.46
AIC: 155.46
```



car::Anova() gives hypothesis tests for the model terms

```
> Anova(arth.polr)       # Type II tests
Analysis of Deviance Table (Type II tests)

Response: Improved
          LR Chisq Df Pr(>Chisq)
Sex           5.69  1    0.01708 *
Treatment    14.71  1    0.00013 ***
Age           4.57  1    0.03251 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Type II tests are partial tests, controlling for the effects of all other terms
- e.g., G^2 (Sex | Treatment, Age), G^2 (Treatment | Age, Sex)
- NB: anova() gives only Type I (sequential) tests – not usually useful

Testing the proportional odds assumption

- The PO model is valid only when the slopes are equal for all predictors
- This can be tested by comparing this model to the generalized logit NPO model

$$\begin{aligned} \text{PO}: \quad L_j &= \alpha_j + \mathbf{x}^\mathsf{T}\beta \qquad j = 1, \ldots, m-1 \qquad (1) \\ \text{NPO}: \quad L_j &= \alpha_j + \mathbf{x}^\mathsf{T}\beta_j \qquad j = 1, \ldots, m-1 \qquad (2) \end{aligned}$$

- A likelihood ratio test requires fitting both models calculating $\Delta G^2 = G^2_{\text{NPO}} - G^2_{\text{PO}}$ with p df.
- This can be done using `vglm()` in the VGAM package
- The rms package provides a visual assessment, plotting the conditional mean $E(X \mid Y)$ of a given predictor, X, at each level of the ordered response Y.
- If the response behaves ordinally in relation to X, these means should be strictly increasing or decreasing with Y.

Testing the proportional odds assumption

In VGAM, the PO model is fit using **`family = cumulative(parallel=TRUE)`**

```
library(VGAM)
arth.po <- vglm(Improved ~ Sex + Treatment + Age, data=Arthritis,
                family = cumulative(parallel=TRUE))
```

The more general NPO model is fit using **`parallel=FALSE`**

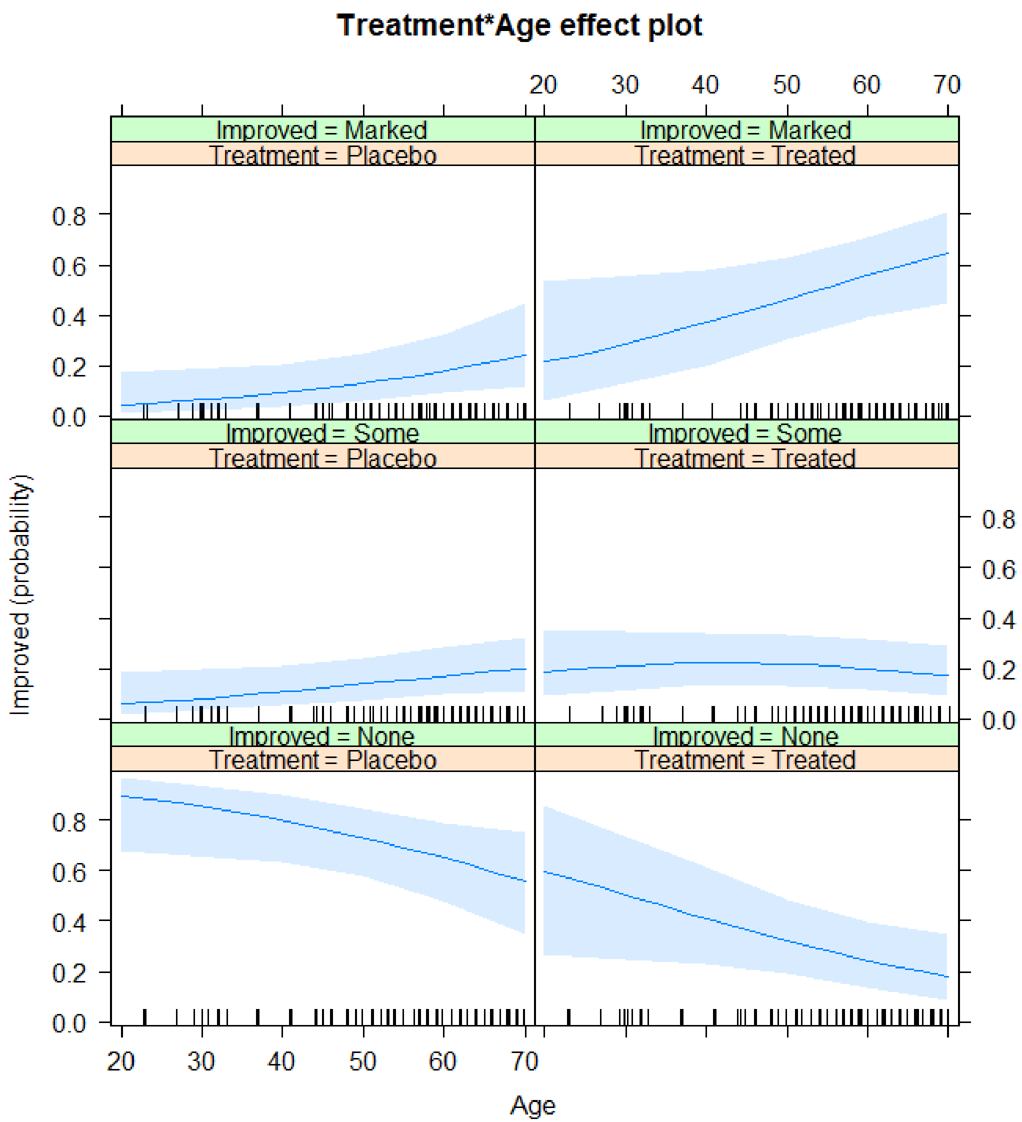
```
arth.npo <- vglm(Improved ~ Sex + Treatment + Age, data=Arthritis,
                 family = cumulative(parallel=FALSE))
```

The LR test indicates that the proportional odds model is OK

```
> VGAM::lrtest(arth.npo, arth.po)
Likelihood ratio test

Model 1: Improved ~ Sex + Treatment + Age
Model 2: Improved ~ Sex + Treatment + Age
  #Df LogLik Df Chisq Pr(>Chisq)
1 160  -71.8
2 163  -72.7  3  1.88        0.6
```

Plotting effects in the PO model



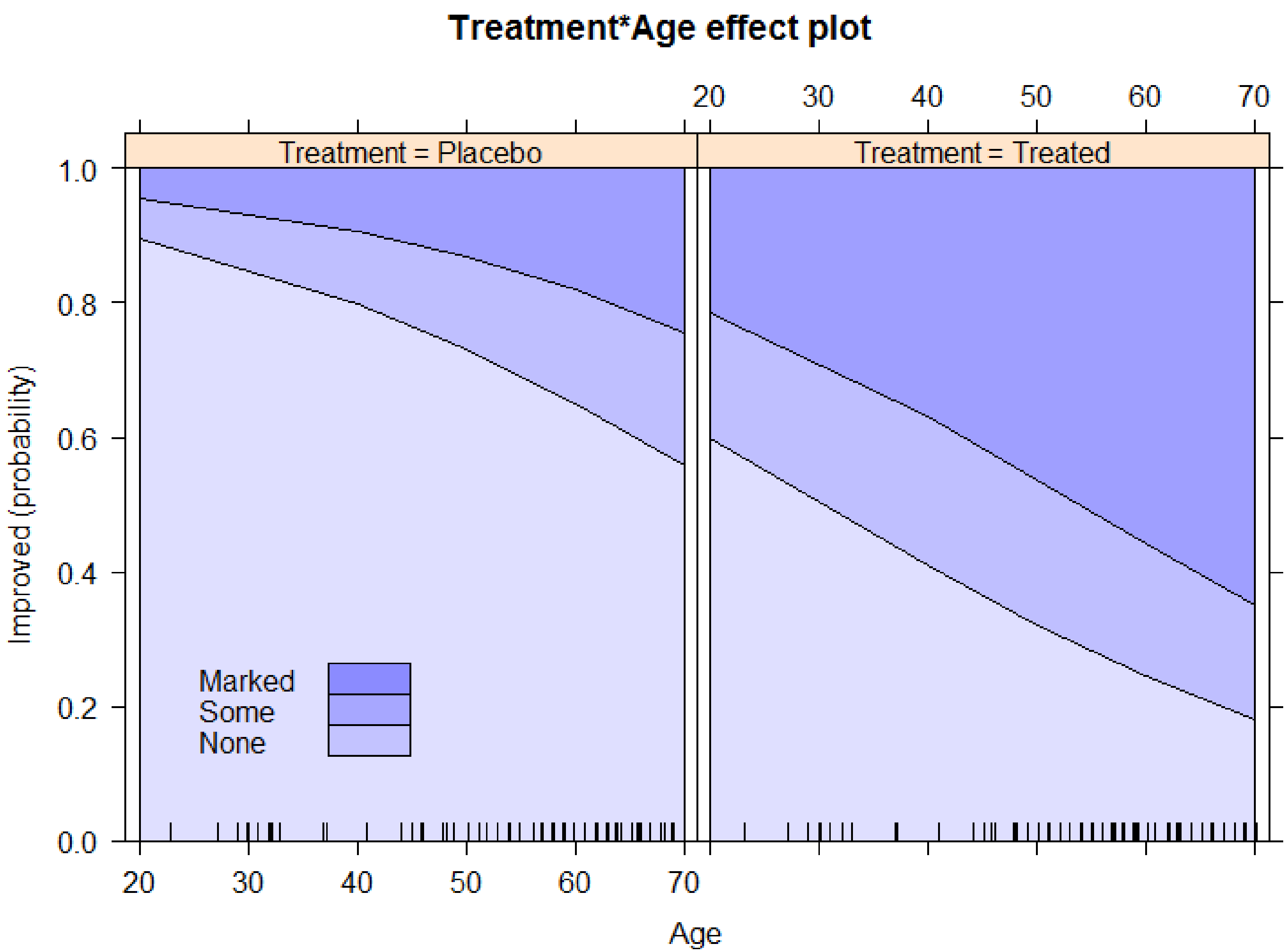

```
library(effects)
plot(effect("Treatment:Age",
      arth.polr))
```

The default style shows separate curves for the response categories

Difficult to compare these in different panels

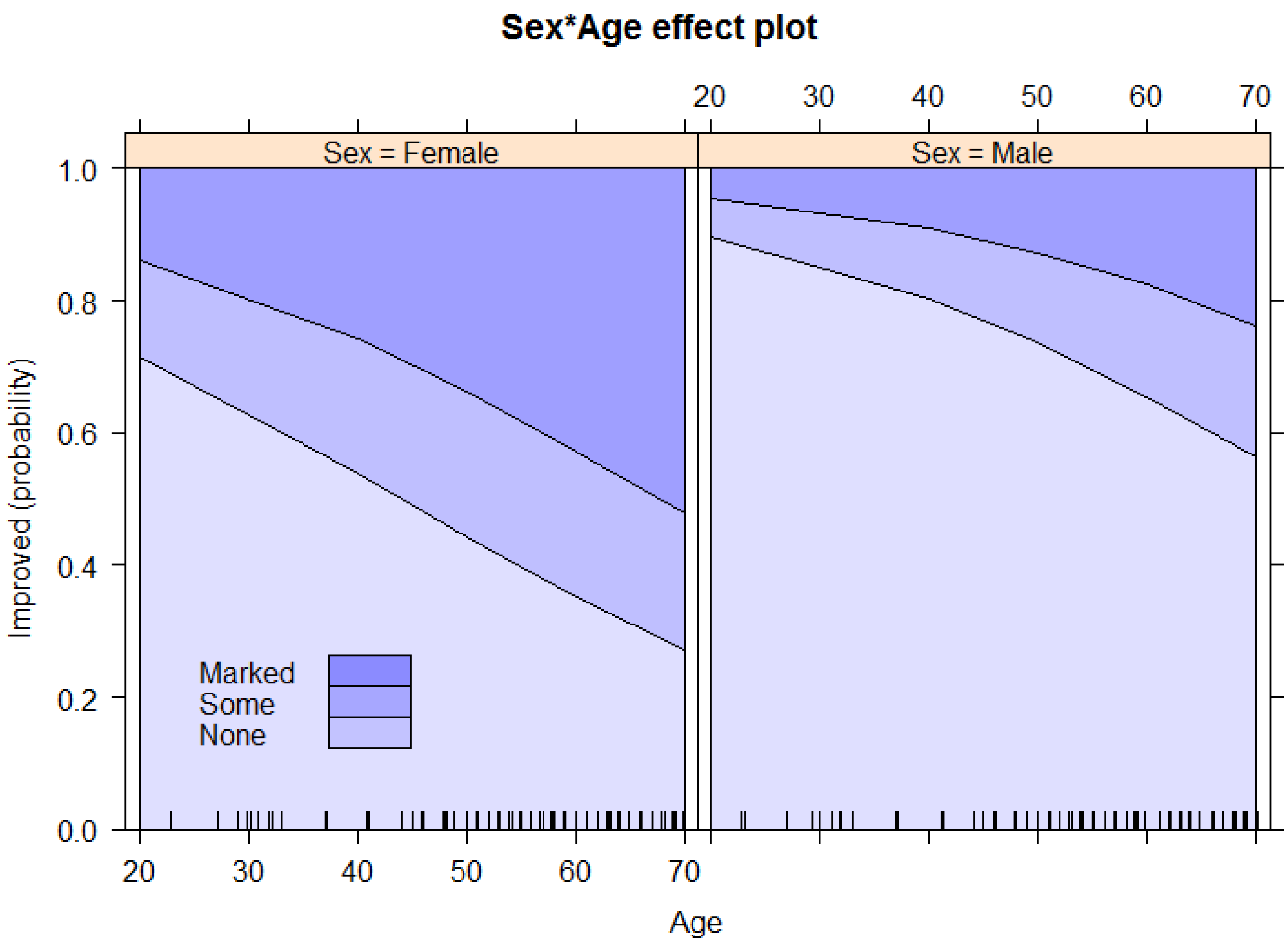
Visual comparisons are easier when the response levels are "stacked"

```
plot(effect("Treatment:Age", arth.polr), style='stacked',
     colors=scales::alpha("blue", alpha = (1:3)/8) )
```



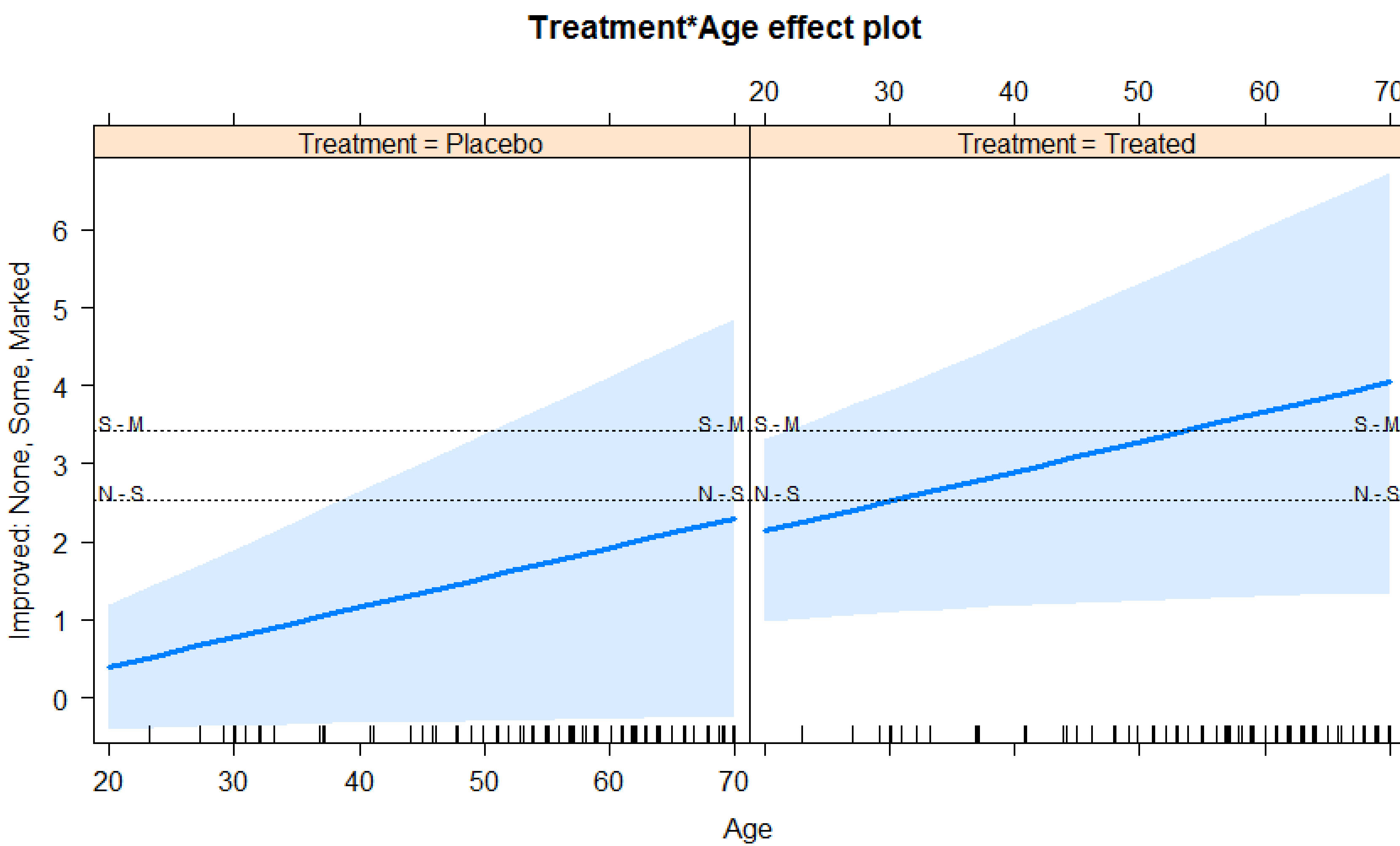
Visual comparisons are easier when the response levels are "stacked"

```
plot(effect("Sex:Age", arth.polr), style='stacked',
      colors=scales::alpha("blue", alpha = (1:3)/8) )
```



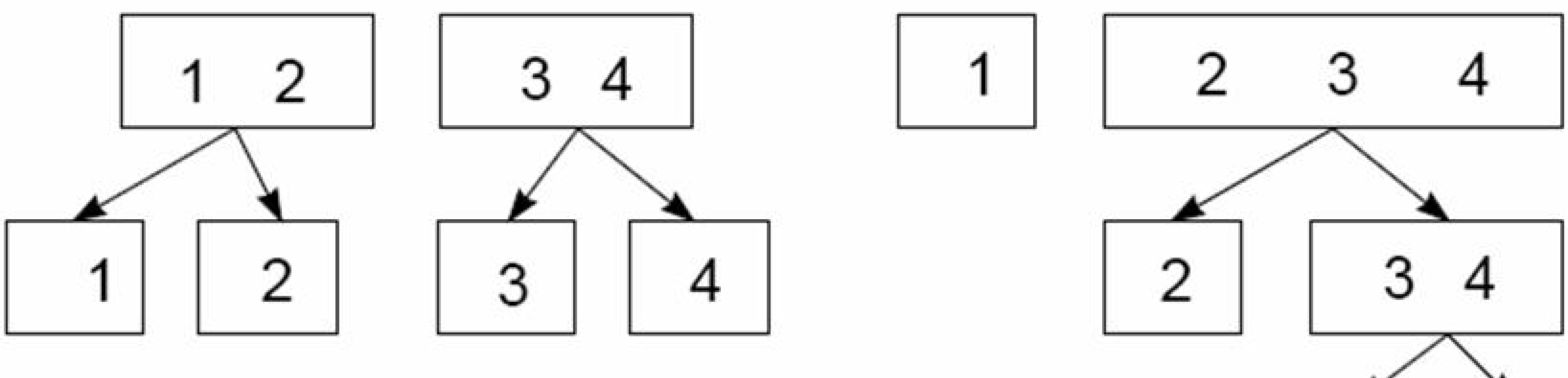
These plots are even simpler on the logit scale, using latent = TRUE to show the cutpoints between adjacent categories

```
plot(effect("Treatment:Age", arth.polr, latent = TRUE))
```



Nested dichotomies

- m categories $\rightarrow$ $(m-1)$ comparisons (logits)
- If these are formulated as $(m-1)$ nested dichotomies:
  - Each dichotomy can be fit using the familiar binary-response logistic model,
  - the $m-1$ models will be statistically independent (G^2 statistics will be additive)
  - (Need some extra work to summarize these as a single, combined model)
- This allows the slopes to differ for each logit

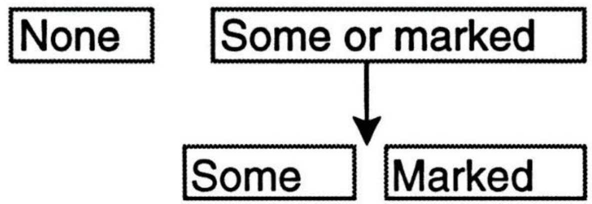


$$G^2_{all} = \sum_{i}^{m-1} G^2(L_i) \qquad df_{all} = \sum df(L_i)$$

Nested dichotomies: Examples

$m = 3$

Arthritis treatment

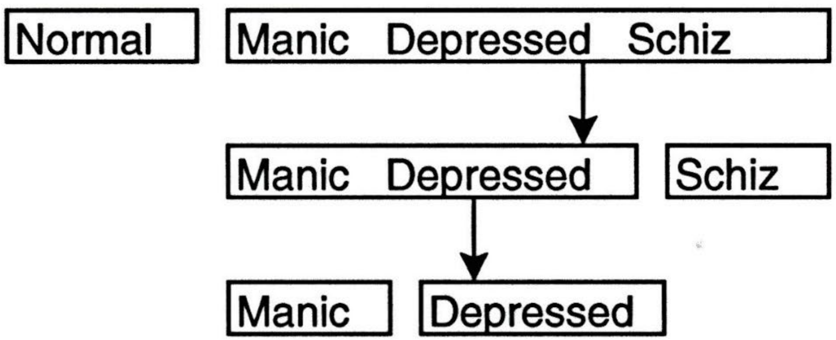


$$L_1 = log \frac{\pi_1}{\pi_2 + \pi_3}$$

$$L_2 = log \frac{\pi_2}{\pi_3}$$

$m = 4$

Psychiatric diagnosis



$$L_1 = log \frac{\pi_1}{\pi_2 + \pi_3 + \pi_4}$$

$$L_2 = log \frac{\pi_4}{\pi_2 + \pi_3}$$

$$L_3 = log \frac{\pi_2}{\pi_3}$$

Example: Women's Labour-force participation

Data: *Social Change in Canada Project*, York ISR, car::`Womenlf` data

- **Response:** not working outside the home (n=155), working part-time (n=42) or working full-time (n=66)
- Model as two nested dichotomies:
  - Working (n=106) vs. NotWorking (n=155)
  - Working full-time (n=66) vs. working part-time (n=42).

L_1:	not working	part-time, full-time	
L_2:		part-time	full-time

- **Predictors:**
  - Children? — 1 or more minor-aged children
  - Husband's Income — in \$1000s
  - Region of Canada (not considered here)

	partic	hincome	children	region
31	not.work	13	present	Ontario
51	parttime	10	present	Prairie
74	not.work	17	present	Ontario
108	not.work	19	present	Ontario
131	parttime	19	present	Ontario
161	not.work	15	present	Ontario
178	fulltime	13	absent	Ontario

Nested dichotomies: Recoding

In R, need to create new variables, working and fulltime.

```
> library(dplyr)
> Womenlf <- Womenlf |>
    mutate(working = ifelse(partic=="not.work", 0, 1)) |>
    mutate(fulltime = case_when(
      working & partic == "fulltime" ~ 1,
      working & partic == "parttime" ~ 0)
    )
> some(Womenlf, 8)
      partic hincome children   region working fulltime
76  parttime      38  present  Ontario       1        0
93  parttime       9  present  Ontario       1        0
101 fulltime      11   absent Atlantic       1        1
107 not.work      13  present  Prairie       0       NA
109 not.work      19  present Atlantic       0       NA
157 parttime      15  present       BC       1        0
220 fulltime      16   absent   Quebec       1        1
249 not.work      23   absent   Quebec       0       NA
```

Nested dichotomies: Fitting

Then, fit separate models for each dichotomy:

```
Womenlf <- within(Womenlf, contrasts(children)<- 'contr.treatment')
mod.working <- glm(working ~ hincome + children, family=binomial, data=Womenlf)
mod.fulltime <- glm(fulltime ~ hincome + children, family=binomial, data=Womenlf)
```

Some output from summary(mod.working)

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.3358     0.3838    3.48   0.0005 ***
hincome          -0.0423     0.0198   -2.14   0.0324 *
childrenpresent  -1.5756     0.2923   -5.39    7e-08 ***
```

Some output from summary(mod.fulltime)

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.4778     0.7671    4.53  5.8e-06 ***
hincome          -0.1073     0.0392   -2.74   0.0061 **
childrenpresent  -2.6515     0.5411   -4.90  9.6e-07 ***
```

Nested dichotomies: Combined tests

- Nested dichotomies $\rightarrow \chi^2$ tests and df for the separate logits are independent
- $\rightarrow$ add, to give tests for the full *m*-level response (manually)

Global tests of BETA=0

Test	Response	ChiSq	DF	Prob ChiSq
Likelihood Ratio	working	36.4184	2	<.0001
	fulltime	39.8468	2	<.0001
	ALL	**76.2652**	**4**	**<.0001**

Wald tests for each coefficient:

Wald tests of maximum likelihood estimates

Variable	Response	WaldChiSq	DF	Prob ChiSq
Intercept	working	12.1164	1	0.0005
	fulltime	20.5536	1	<.0001
	ALL	**32.6700**	**2**	**<.0001**
children	working	29.0650	1	<.0001
	fulltime	24.0134	1	<.0001
	ALL	**53.0784**	**2**	**<.0001**
husinc	working	4.5750	1	0.0324
	fulltime	7.5062	1	0.0061
	ALL	**12.0813**	**2**	**0.0024**

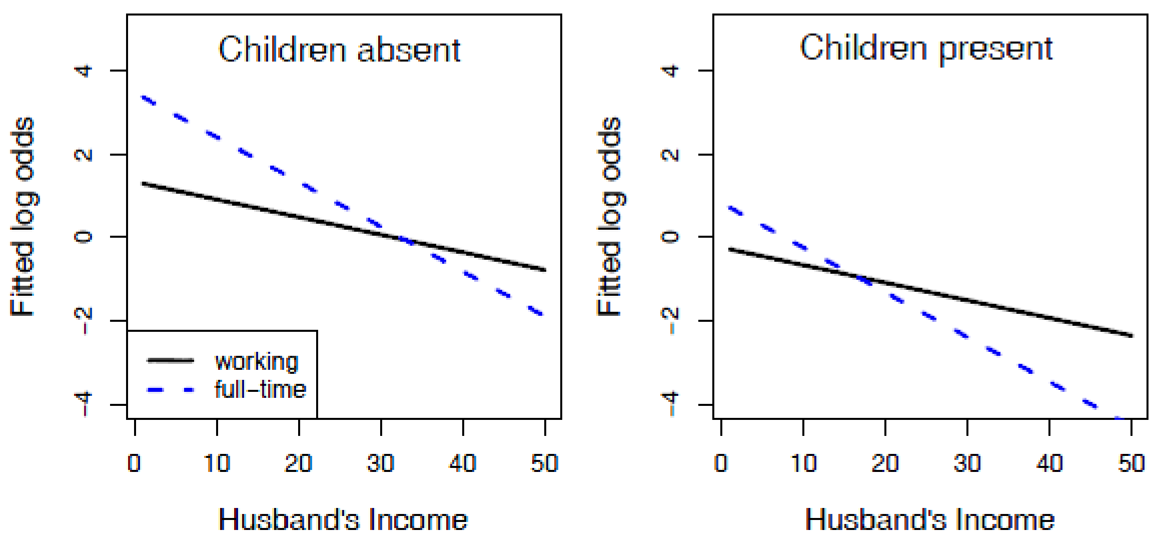
Nested dichotomies: Interpretation

Write out the predictions for the two logits, and compare coefficients:

$$\log\left(\frac{\Pr(\text{working})}{\Pr(\text{not working})}\right) = 1.336 - 0.042\,\text{H\$} - 1.576\,\text{kids}$$

$$\log\left(\frac{\Pr(\text{fulltime})}{\Pr(\text{parttime})}\right) = 3.478 - 0.107\,\text{H\$} - 2.652\,\text{kids}$$

Better yet, plot the predicted log odds for these equations:



Nested dichotomies: Plotting

For plotting, calculate the predicted probabilities (or logits) over a grid of combinations of the predictors in each sub-model, using **`predict()`**.

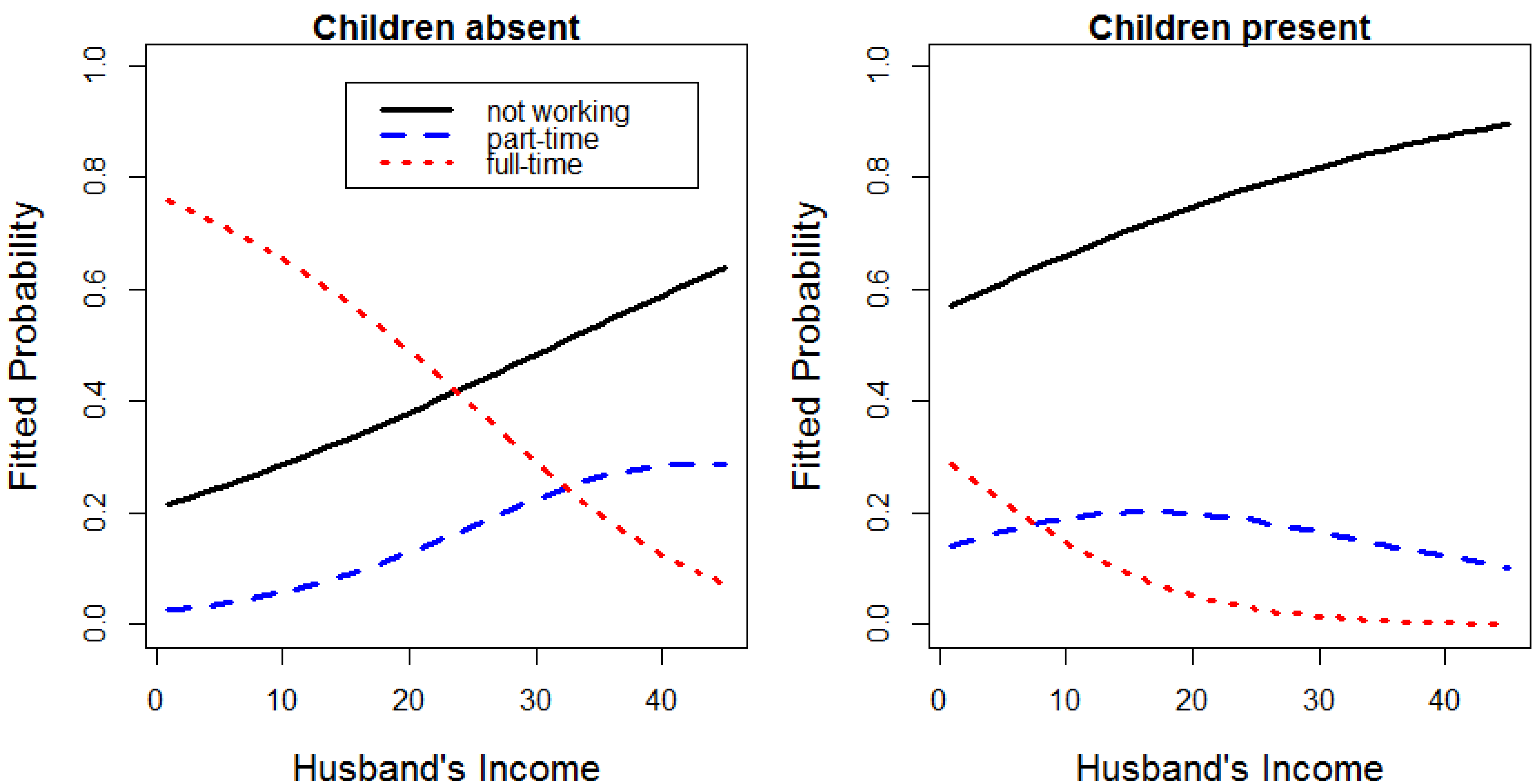
- `type = "response"` gives these on the probability scale
- `type = "link"` (default) gives these on the logit scale

```
predictors <- expand.grid(hincome=1:45, children=c('absent', 'present'))
# get fitted values for both sub-models
p.work     <- predict(mod.working, predictors, type='response')
p.fulltime <- predict(mod.fulltime, predictors, type='response')
```

The fitted value for the fulltime dichotomy is conditional on working outside the home; multiplying by the probability of working gives the unconditional probability.

```
p.full <- p.work * p.fulltime
p.part <- p.work * (1 - p.fulltime)
p.not  <- 1 - p.work
```

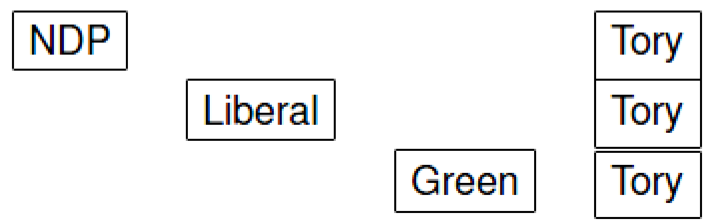
This plot is produced using base R functions plot(), lines() and legend()
See the file: wlf-nested.R on the course web page for details



Multinomial logistic regression

- Multinomial logistic regression models the probabilities of *m* response categories as (*m*-1) logits
  - Typically, these compare each of the first *m*-1 categories to the last (reference) category: 1 vs. *m*, 2 vs. *m*, … *m*-1 vs. *m*

e.g., vote for
(m = 4)



  - Logits for any pair of categories can be calculated from the *m*-1 fitted ones

Multinomial logistic regression

- with k predictors, x_1, x_2, …, x_k and for j=1, 2, …, m-1, the model fits separate slopes for each logit

$$\begin{aligned} L_{jm} \equiv \log\left(\frac{\pi_{ij}}{\pi_{im}}\right) &= \beta_{0j} + \beta_{1j}\, x_{i1} + \beta_{2j}\, x_{i2} + \cdots + \beta_{kj}\, x_{ik} \\ &= \beta_j^{\mathsf{T}} \boldsymbol{x}_i \end{aligned}$$

  - One set of coefficients, $\boldsymbol{\beta}_j$ for each response category except the last
  - Each coefficient, β_{hj}, gives effect on log odds that response is j vs. m, for a one unit change in the predictor $\boldsymbol{x}_h$
- Probabilities in response categories are calculated as

$$\pi_{ij} = \frac{\exp(\beta_j^{\mathsf{T}} \boldsymbol{x}_i)}{\sum_{j=1}^{m-1} \exp(\beta_j^{\mathsf{T}} \boldsymbol{x}_i)} \;, j = 1, \ldots, m-1 \,; \qquad \pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij}$$

Fitting multinomial regression models

Fit the multinomial model using nnet::multinom()
For ease of interpretation, make not.work the reference category

```
> Womenlf$partic <- relevel(Womenlf$partic, ref="not.work")
> library(nnet)
> wlf.multinom <- multinom(partic ~ hincome + children,
                           data=Womenlf, Hess=TRUE)
```

The **Anova()** tests are similar to what we got from summing these tests from the two nested dichotomies

```
> Anova(wlf.multinom)
Analysis of Deviance Table (Type II tests)

Response: partic
         LR Chisq Df Pr(>Chisq)
hincome      15.2  2    0.00051 ***
children     63.6  2    1.6e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting coefficients

As before, interpret coefficients as increments in log odds or exp(coef) as multiples

```
> coef(wlf.multinom)
         (Intercept)  hincome childrenpresent
parttime     -1.43  0.00689          0.0215
fulltime      1.98 -0.09723         -2.5586
```

```
> exp(coef(wlf.multinom))
         (Intercept) hincome childrenpresent
parttime     0.239   1.007           1.0217
fulltime     7.263   0.907           0.0774
```

$$\log\left(\frac{\Pr(\text{parttime})}{\Pr(\text{notworking})}\right) = -1.43 + 0.0069\ \text{H\$} - 0.215\ \text{kids}$$

$$\log\left(\frac{\Pr(\text{fulltime})}{\Pr(\text{notworking})}\right) = 1.98 - 0.097\ \text{H\$} - 2.55\ \text{kids}$$

Each 1000$ of husband's income:

- Increases log odds of parttime by 0.0069; multiplies odds by 1.007 (+0.7%)
- Decreases log odds of fulltime by 0.097; multiplies odds by 0.091 (-9%)

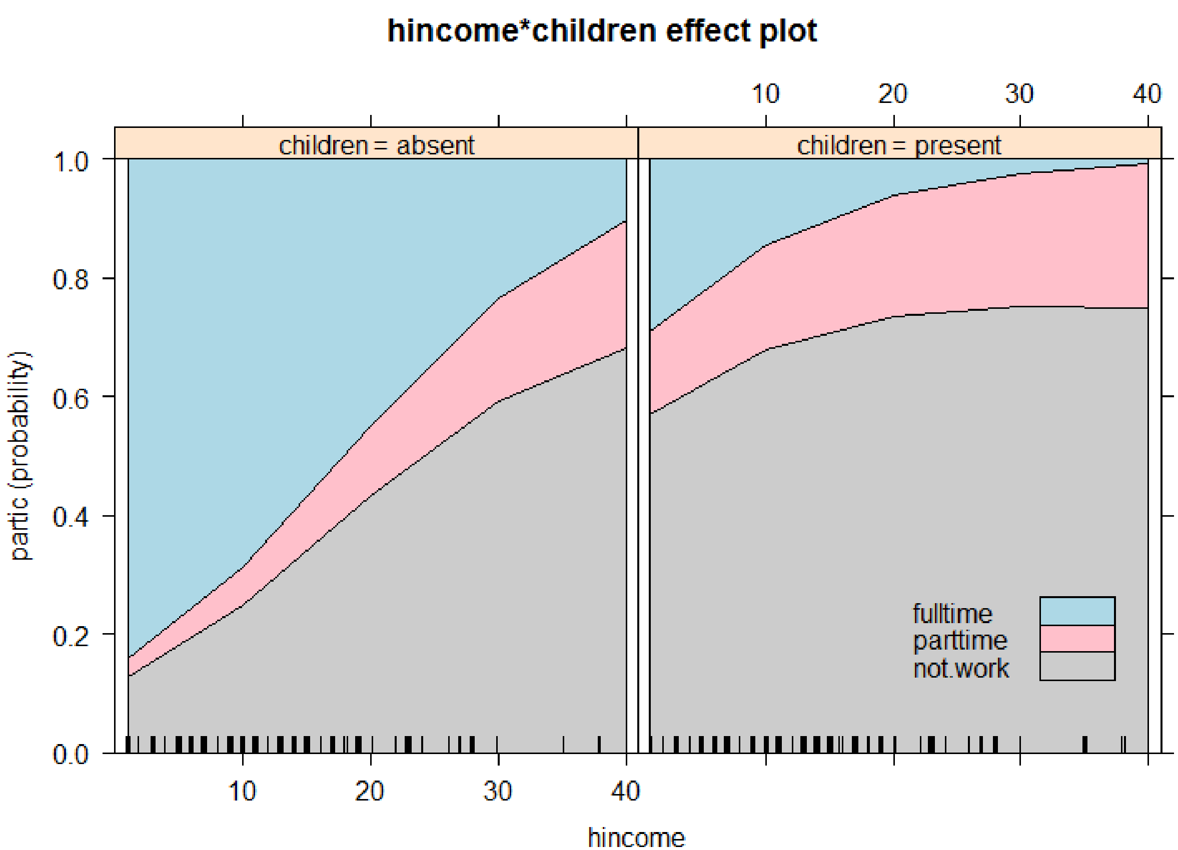
Having young children:

- Increases odds of parttime by 0.0215; multiplies odds by 1.0217 (+2%)
- Decreases odds of fulltime by 2.559; multiplies odds by 0.0774 (-92%)

Multinomial models: Plotting

Much easier to interpret a model from a plot, but even more so for polytomous response models

```
library(effects)
plot(Effect(c("hincome", "children"), wlf.multinom), style = "stacked")
```



For multinomial model, style="stacked" plots cumulative probs.

Multinomial models: Plotting

An alternative is to plot the predicted probabilities of each level of participation over a grid of predictor values for husband's income and children.

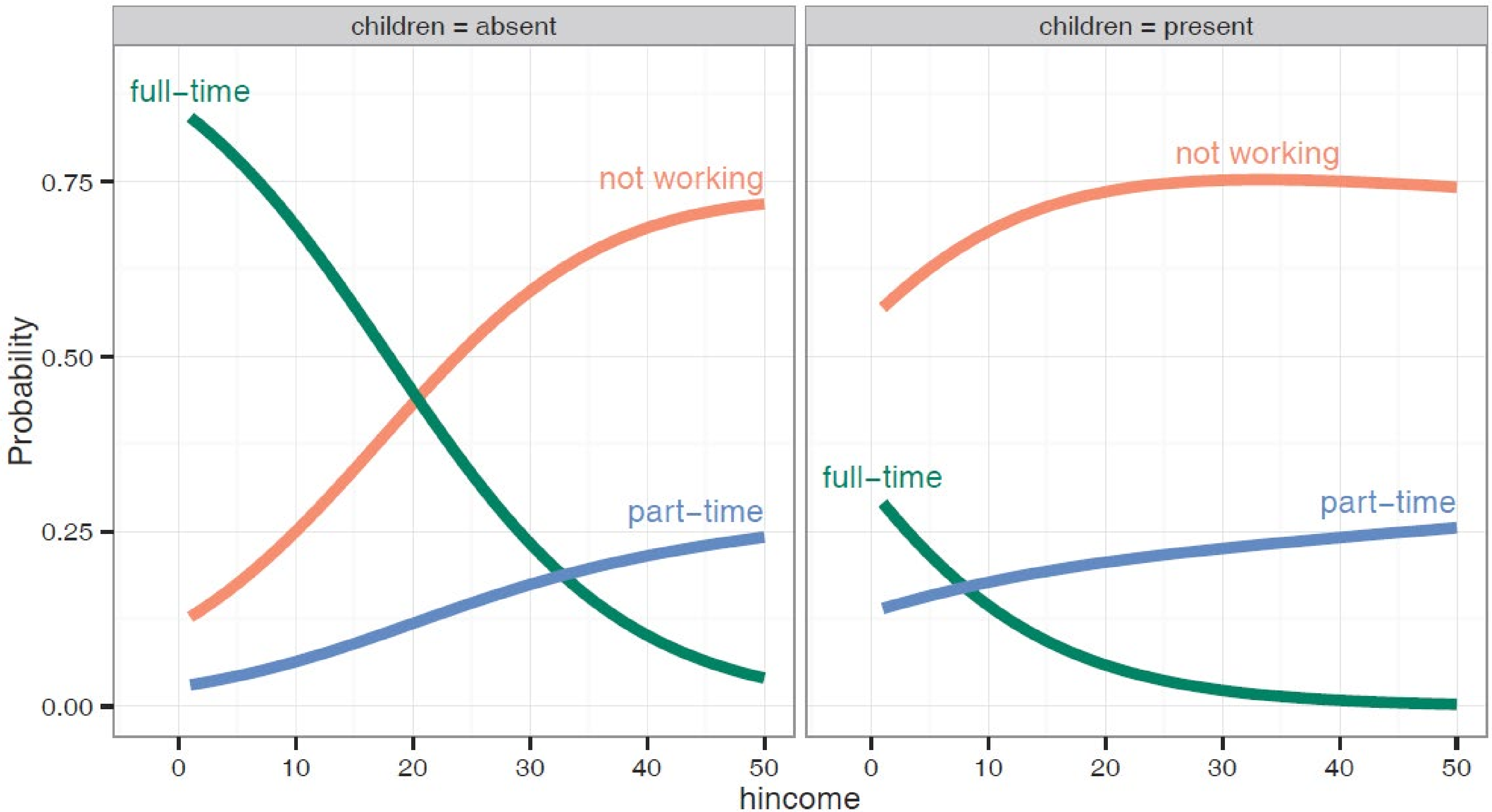
```
> predictors <- expand.grid(hincome=1:50, children=c('absent', 'present'))
> fit <- data.frame(predictors,
+                   predict(wlf.multinom, predictors, type='probs'))
> fit |> filter(hincome %in% c(10, 25, 40))    # show a few observations
   hincome children not.work parttime fulltime
10      10   absent    0.250   0.0639  0.68627
25      25   absent    0.520   0.1475  0.33233
40      40   absent    0.683   0.2150  0.10157
60      10  present    0.678   0.1773  0.14427
75      25  present    0.747   0.2164  0.03693
90      40  present    0.750   0.2411  0.00863
```

We want to plot predicted probability vs. hincome, with separate curves for levels of participation. To do this we need to reshape the fit data from wide to long

```
plotdat <- fit |>
  gather(key="Level", value="Probability", not.work:fulltime)
```

Now, plot Probability ~ hincome, with separate curves for Level of partic

```
library(directlabels)
gg <- ggplot(plotdat, aes(x = hincome, y = Probability, colour = Level)) +
  geom_line(size=1.5) + facet_grid(~ children, labeller = label_both)
direct.label(gg, list("top.bumptwice", dl.trans(y = y + 0.2)))
```



A larger example: BEPS data

Political knowledge & party choice in Britain

Example from Fox & Anderson (2006); data from 1997 British Election Panel (BEPS), N=1325

- **Response**: Party choice— Liberal democrat, Labour, Conservative
- **Predictors**
  - Europe: 11-point scale of attitude toward European integration (high="Eurosceptic")
  - Political knowledge: knowledge of party platforms on European integration ("low"=0–3="high")
  - Others: Age, Gender, perception of economic conditions, evaluation of party leaders (Blair, Hague, Kennedy)– 1:5 scale

- **Model**:
  - Main effects of Age, Gender, economic conditions (national, household)
  - Main effects of evaluation of party leaders
  - Interaction of attitude toward European integration with political knowledge

BEPS data: Fitting

Fit a model with main effects and an interaction of Europe * political knowledge

```
library(car)       # for Anova()
library(nnet)      # for multinom()
data(BEPS, package = "carData")
BEPS.mod <- multinom(vote ~ age + gender + economic.cond.national +
                economic.cond.household + Blair + Hague + Kennedy +
                Europe*political.knowledge, data=BEPS)
Anova(BEPS.mod)
```

```
Analysis of Deviance Table (Type II tests)

Response: vote
                           LR Chisq Df Pr(>Chisq)
age                            13.9  2    0.00097 ***
gender                          0.5  2    0.79726
economic.cond.national         30.6  2    2.3e-07 ***
economic.cond.household         5.7  2    0.05926 .
Blair                         135.4  2    < 2e-16 ***
Hague                         166.8  2    < 2e-16 ***
Kennedy                        68.9  2    1.1e-15 ***
Europe                         78.0  2    < 2e-16 ***
political.knowledge            55.6  2    8.6e-13 ***
Europe:political.knowledge     50.8  2    9.3e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

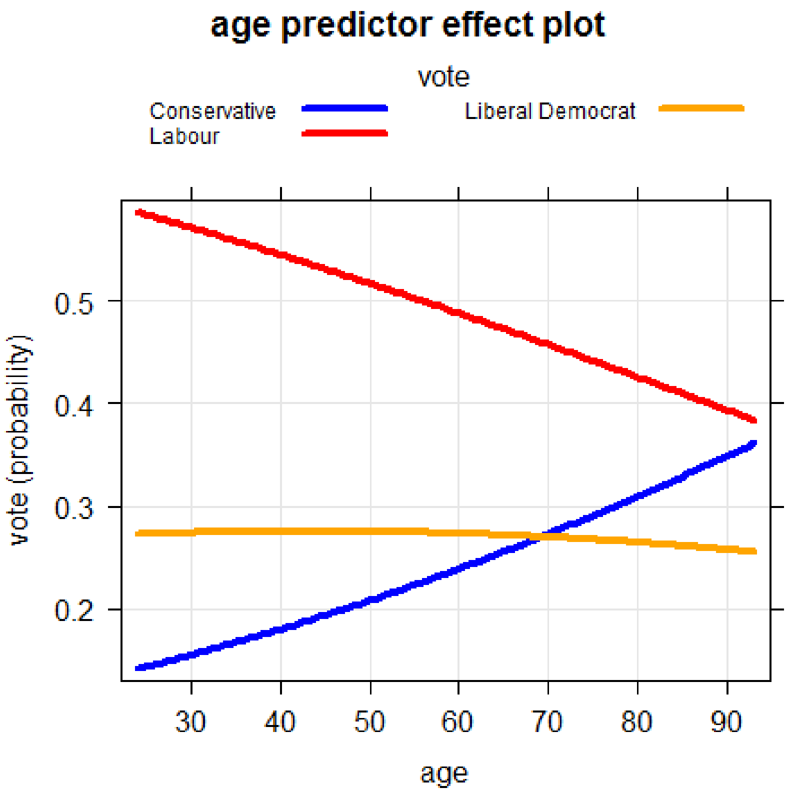
BEPS data: Interpretation?

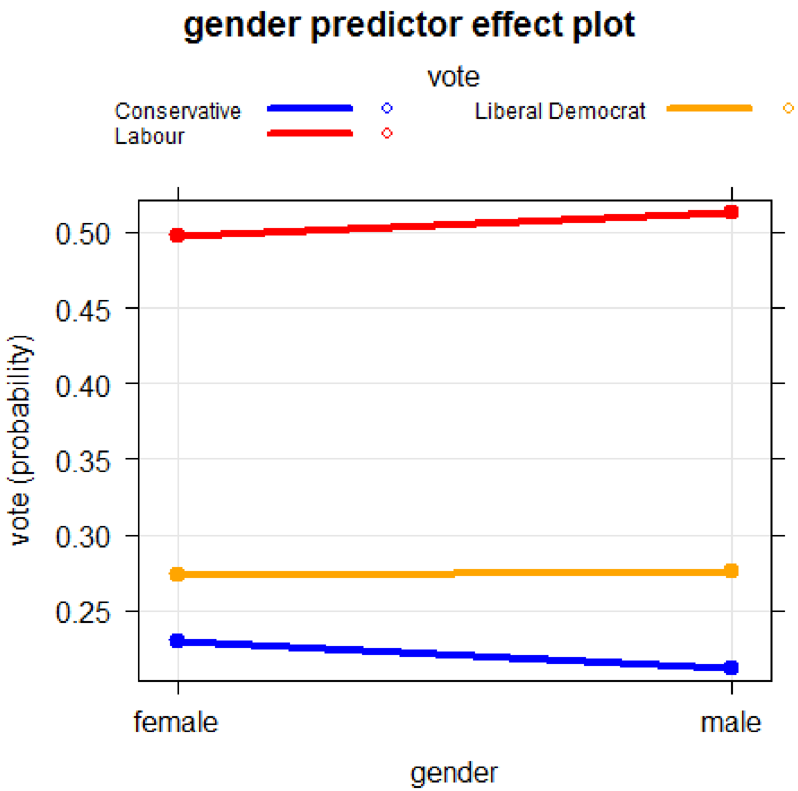
Coefficients give log odds relative of party choice relative to Conservatives
How to understand the nature of these effects?

```
> coef(BEPS.mod)
                 (Intercept)     age gendermale economic.cond.national
Labour                -0.873 -0.0198     0.1126                  0.522
Liberal Democrat      -0.718 -0.0146     0.0914                  0.145
                 economic.cond.household Blair  Hague Kennedy   Europe
Labour                           0.17863 0.824 -0.868   0.240 -0.00171
Liberal Democrat                 0.00773 0.278 -0.781   0.656  0.06841
                 political.knowledge Europe:political.knowledge
Labour                         0.658                     -0.159
Liberal Democrat               1.160                     -0.183
```

BEPS data: Effect plots

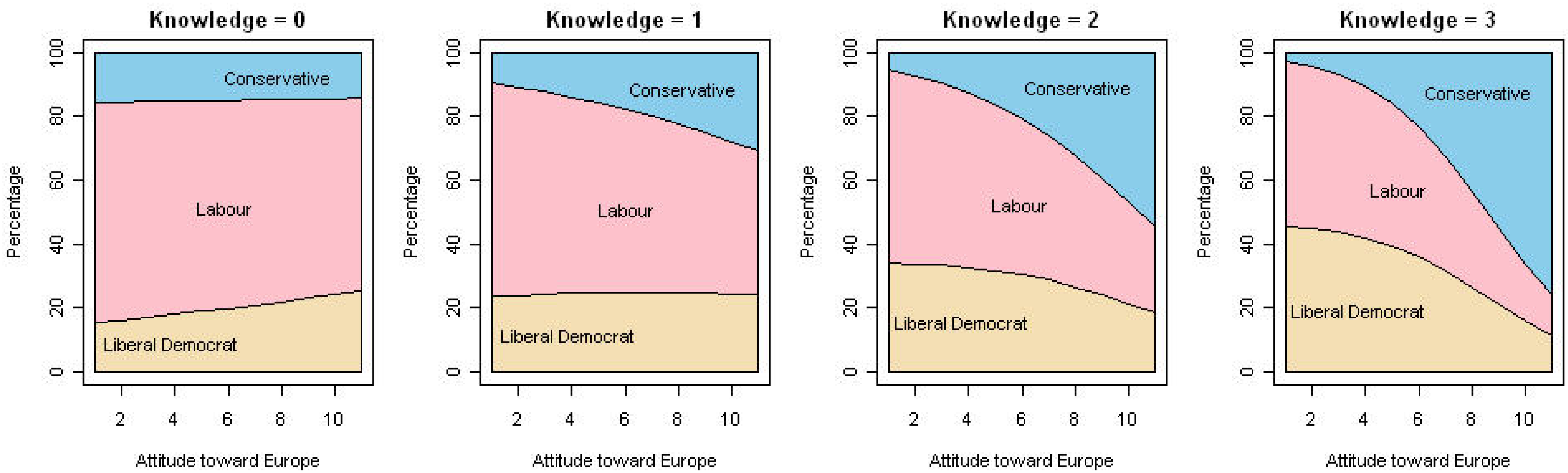
```
plot(predictorEffects(BEPS.mod, ~ age + gender),
    lattice=list(key.args=list(rows=1)),
    lines=list(multiline=TRUE, col=c("blue", "red", "orange")))
```





BEPS data: Effect plots

Examine the interaction between political knowledge and attitude toward European integration



- Low knowledge: little relation between attitude and party choice
- As knowledge increases: more Eurosceptic view $\rightarrow$ more likely to support Conservatives
- Detailed understanding of complex models depends strongly on visualization!

Summary

- Polytomous responses
  - m response categories $\rightarrow$ (m-1) comparisons (logits)
  - Different models for ordered vs. unordered categories
- Proportional odds model
  - Simplest approach for ordered categories
  - Assumes same slopes for all logits
    - Fit with MASS::polr()
    - Test PO assumption with VGAM::vglm()
- Nested dichotomies
  - Applies to ordered or unordered categories
  - Fit m – 1 separate independent models $\rightarrow$ Additive $\chi 2$ values
- Multinomial logistic regression
  - Fit m – 1 logits as a single model
  - Results usually comparable to nested dichotomies, but diff interpretation
  - R: nnet::multinom()