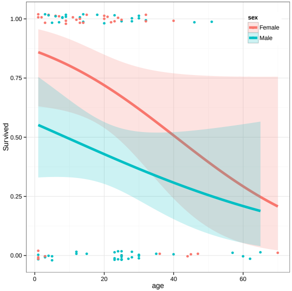
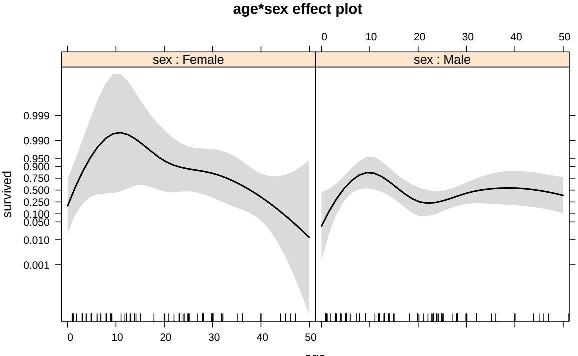
Logistic Regression II

Michael Friendly

Psych 6136

November 9, 2017

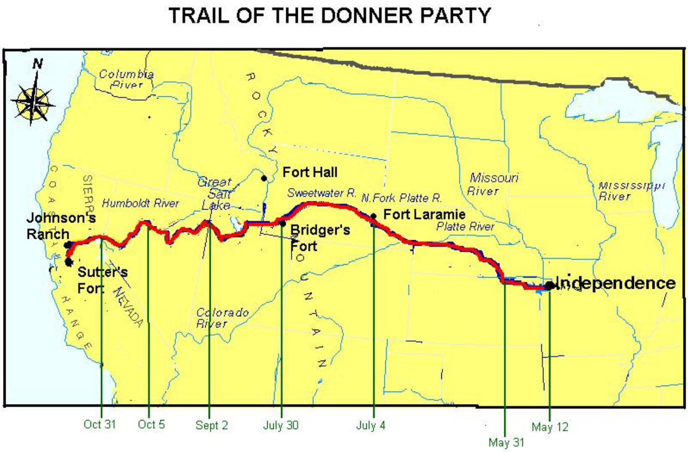




Donner Party: A graphic tale of survival & influence

History:

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

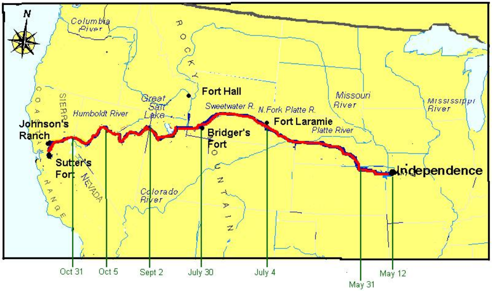


Donner Party: A graphic tale of survival & influence

History:

- "Hasting's Cutoff", untried route through Salt Lake Desert, Wasatch Mtns. (90 people)
- Worst recorded winter: Oct 31 blizzard— Missed by 1 day, stranded at "Truckee Lake" (now Donner's Lake, Reno)
    - Rescue parties sent out ("Dire necessity", "Forelorn hope", ...)
    - Relief parties from CA: 42 survivors (Mar–Apr, '47)

TRAIL OF THE DONNER PARTY



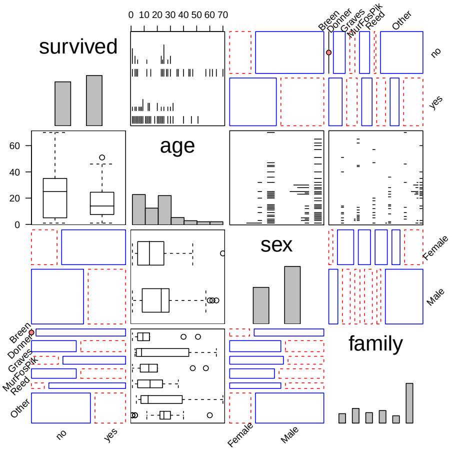
Donner Party: Data

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

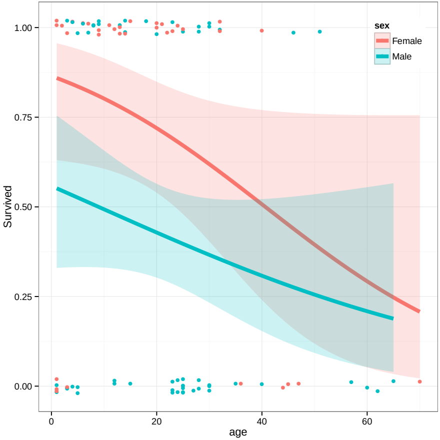
```
library(car)
some(Donner, 12)

##                          family age    sex survived      death
## Breen, Peter              Breen   3   Male      yes       <NA>
## Donner, George           Donner  62   Male       no 1847-03-18
## Donner, Jacob            Donner  65   Male       no 1846-12-21
## Foster, Jeremiah      MurFosPik   1   Male       no 1847-03-13
## Graves, Jonathan         Graves   7   Male      yes       <NA>
## Graves, Mary Ann         Graves  20 Female      yes       <NA>
## Graves, Nancy            Graves   9 Female      yes       <NA>
## McCutchen, Harriet McCutchen   1 Female       no 1847-02-02
## Reed, James                Reed  46   Male      yes       <NA>
## Reed, Thomas Keyes         Reed   4   Male      yes       <NA>
## Reinhardt, Joseph         Other  30   Male       no 1846-12-21
## Wolfinger, Doris       FosdWolf  20 Female      yes       <NA>
```

Overview: a **`gpairs()`** plot

- Binary response: `survived`
- Categorical predictors: `sex`, `family`
- Quantitative predictor: `age`
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?
- This is a generalized pairs plot, with different plots for each pair



Exploratory plots



- Survival decreases with age for both men and women
- Women more likely to survive, particularly the young
- Data is thin at older ages

Using ggplot2

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_point(position = position_jitter(height = 0.02, width = 0))
```

Add conditional linear logistic regressions with `stat_smooth(method="glm")`

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

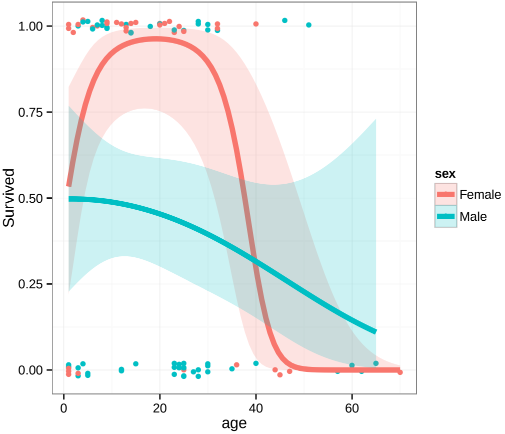
Questions

- 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)`,

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

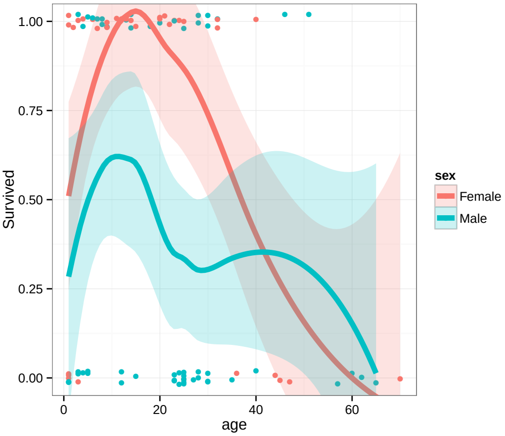
  - Use *natural spline* functions, `ns(age, df)`
  - Use non-parametric smooths, `loess(age, span, degree)`
- Is the relation the same for men and women? i.e., do we need an interaction of age and sex?
  - Allow an interaction of `sex * age` or `sex * f(age)`
  - Test goodness-of-fit relative to the main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

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



Fit separate quadratics for males and females

```
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 males and females

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
```

Fiting 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
```

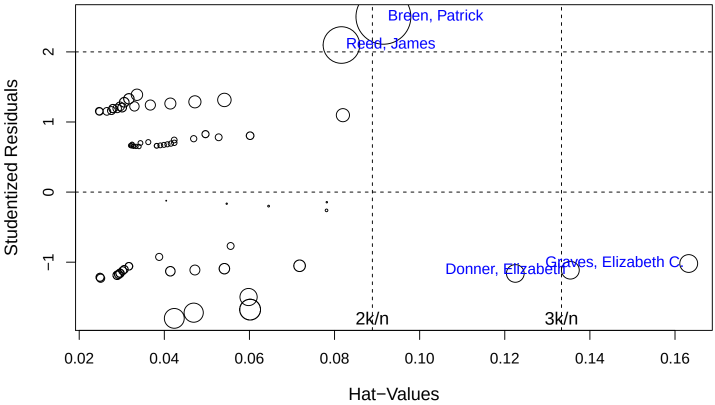
Comparing models

```
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?

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

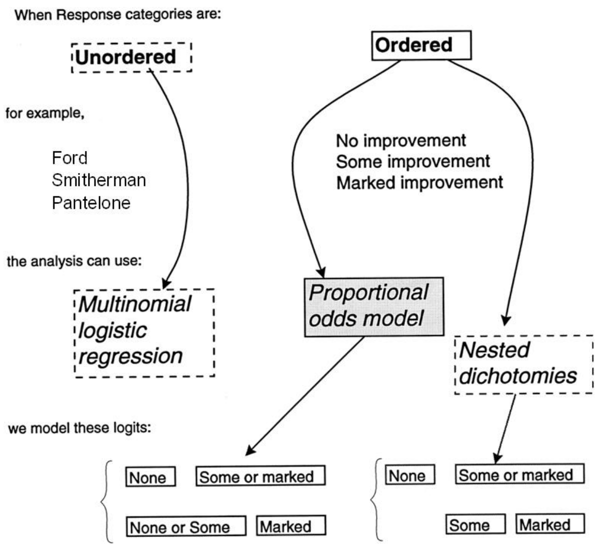


Why are they influential?

```r
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   2.501 0.09148 0.32354
## Donner, Elizabeth      45 Female       no  -1.114 0.13541 0.03409
## Graves, Elizabeth C.   47 Female       no  -1.019 0.16322 0.03418
## Reed, James            46   Male      yes   2.098 0.08162 0.14364
```

- 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



Polytomous responses: Overview

- m categories $\rightarrow$ $(m-1)$ comparisons (logits)
  - One part of the model for each logit
  - Similar to ANOVA where an m-level factor $\rightarrow$ $(m-1)$ contrasts (df)
- **Response categories *unordered***, 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

Polytomous responses: Overview

- **Response categories *ordered***, 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

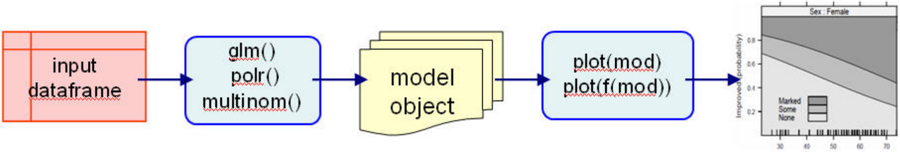
None	Some or Marked
	Some | Marked

    - Model each logit separately
    - G^2 s are additive $\rightarrow$ combined model

Fitting and graphing: Overview

R:

- Model objects contain all necessary information for plotting
- Basic diagnostic plots with `plot(model)`
- Fitted values with **`predict()`**; customize with **`points()`**, **`lines()`**, etc.
- Effect plots most general



Ordinal response: Proportional odds model

Arthritis treatment data:

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

- Model logits for adjacent category cutpoints:

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

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

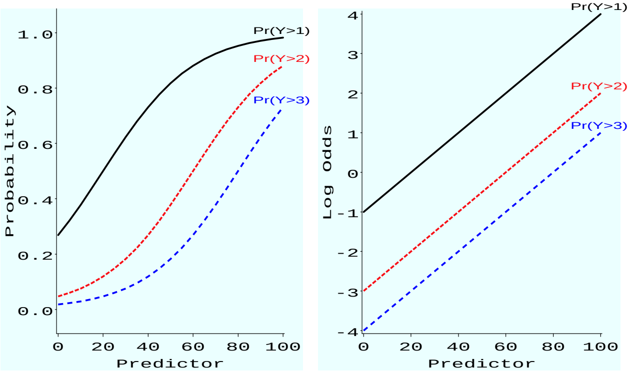
- Consider a logistic regression model for each logit:

$$\text{logit}(\theta_{ij1}) = \alpha_1 + \boldsymbol{x}_{ij}' \boldsymbol{\beta}_1 \qquad \text{None vs. Some/Marked}$$

$$\text{logit}(\theta_{ij2}) = \alpha_2 + \boldsymbol{x}_{ij}' \boldsymbol{\beta}_2 \qquad \text{None/Some vs. Marked}$$

- Proportional odds assumption: regression functions are parallel on the logit scale i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

Proportional Odds Model



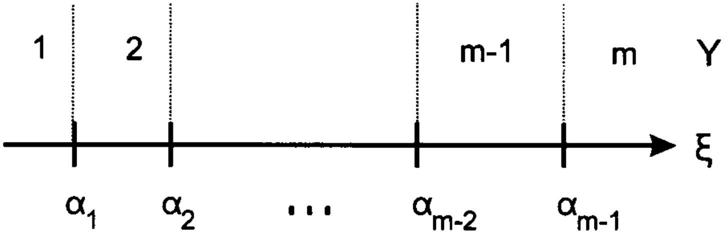
Proportional odds: Latent variable interpretation

A simple motivation for the proportional odds model:

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

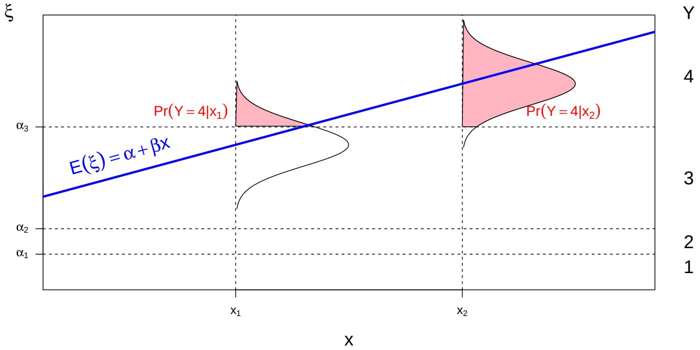
$$\xi_i = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{x}_i + \epsilon_i$$

- 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



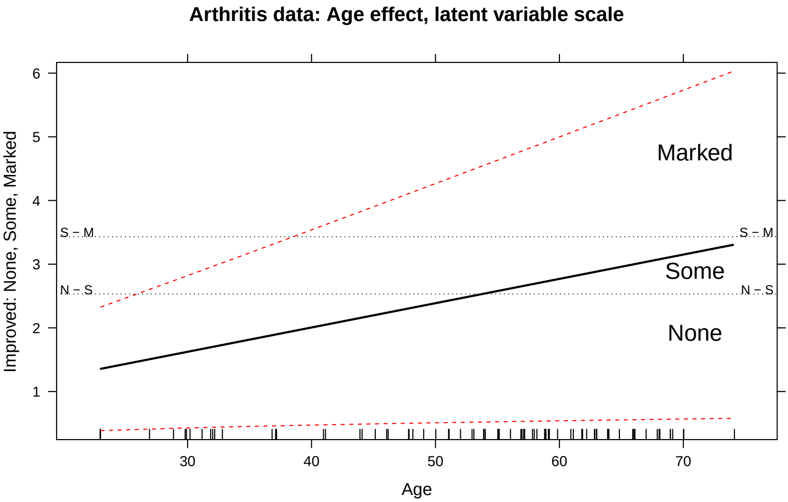
Proportional odds: Latent variable interpretation

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

For the Arthritis data, the relation of improvement to age is shown below (using the effects package)



Proportional odds models in R

- **Fitting:** `polr()` in MASS package

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
```

Fitting:

```
library(MASS)          # for polr()
library(car)           # for Anova()

arth.polr <- polr(Improved ~ Sex + Treatment + Age,
                  data=Arthritis)
summary(arth.polr)
Anova(arth.polr)        # Type II tests
```

The `summary()` function gives standard statistical results:

```
> summary(arth.polr)
```

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

Coefficients:
                     Value  Std. Error   t value
SexMale           -1.25168     0.54636   -2.2909
TreatmentTreated   1.74529     0.47589    3.6674
Age                0.03816     0.01842    2.0722

Intercepts:
            Value   Std. Error t value
None|Some    2.5319  1.0571     2.3952
Some|Marked  3.4309  1.0912     3.1442

Residual Deviance: 145.4579
AIC: 155.4579
```

The car::**Anova()** function gives hypothesis tests for model terms:

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

```
Anova Table (Type II tests)

Response: Improved
          LR Chisq Df Pr(>Chisq)
Sex         5.6880  1  0.0170812 *
Treatment  14.7095  1  0.0001254 ***
Age         4.5715  1  0.0325081 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **anova()** gives Type I (sequential) tests — not usually useful
- Type II (partial) tests control for the effects of all other terms

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

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

- 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))
```

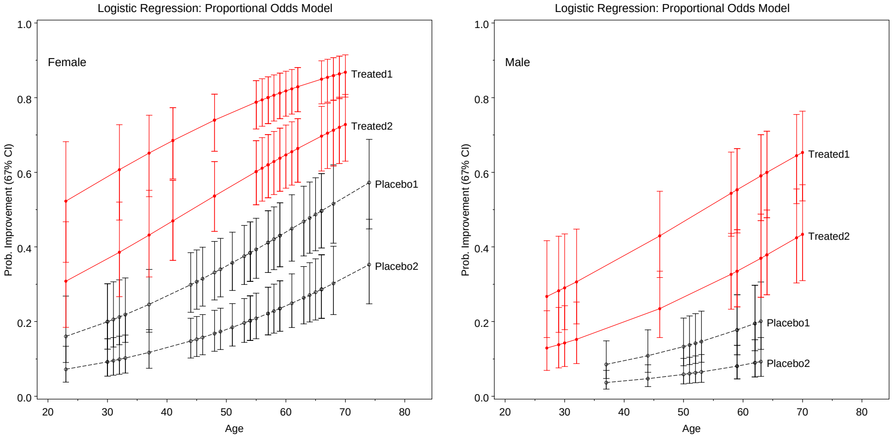
The more general NPO model can be fit using `parallel=FALSE`.

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

The LR test says the PO 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
```

Full-model plot of predicted probabilities:

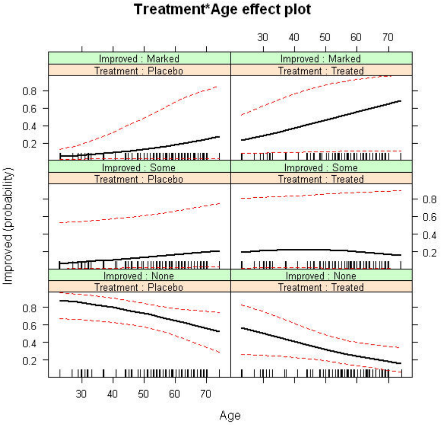


- Intercept1: [Marked , Some] vs. [None]
- Intercept2: [Marked] vs. [Some, None]
- On logit scale, these would be parallel lines
- Effects of age, treatment, sex similar to what we saw before

Proportional odds models in R: Plotting

- **Plotting:** `plot(effect())` in effects package

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

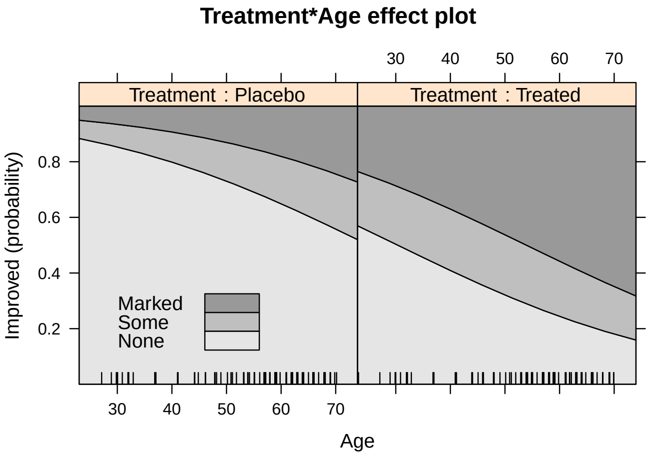


- The default plot shows all details
- But, is harder to compare across treatment and response levels

Proportional odds models in R: Plotting

Making visual comparisons easier:

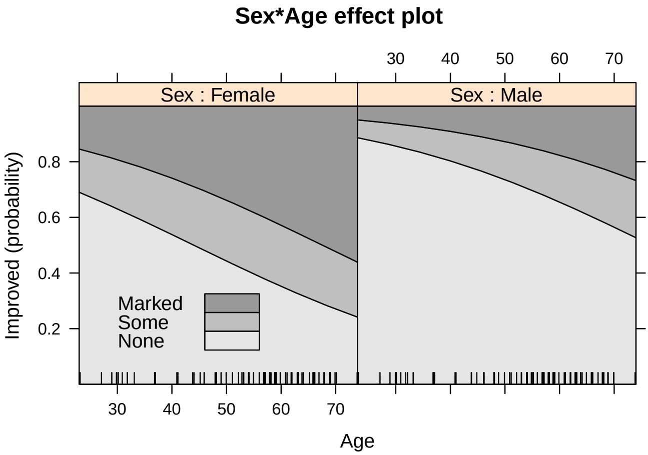
```
> plot(effect("Treatment:Age", arth.polr), style='stacked')
```



Proportional odds models in R: Plotting

Making visual comparisons easier:

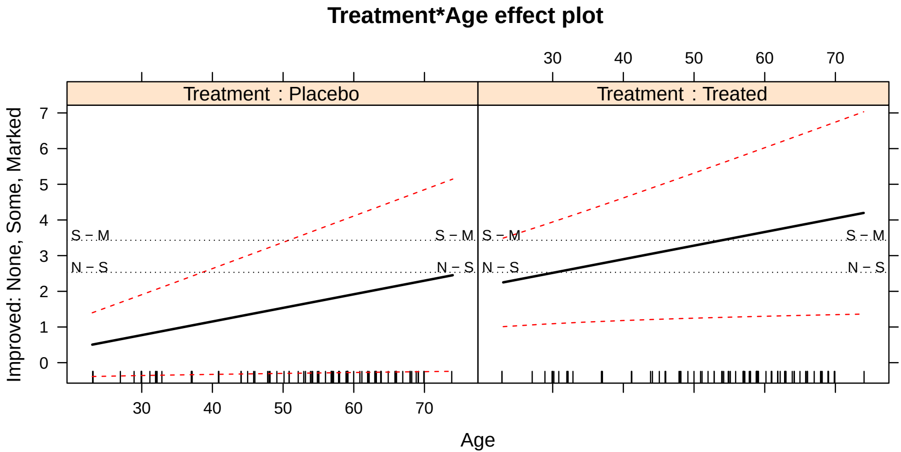
```
> plot(effect("Sex:Age", arth.polr), style='stacked')
```



Proportional odds models in R: Plotting

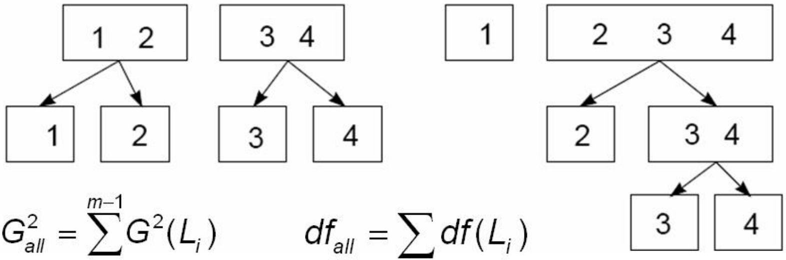
These plots are even simpler on the logit scale, using `latent=TRUE` to show the cutpoints between response categories

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



Polytomous response: 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

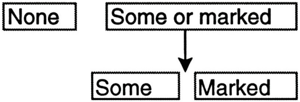


$$G^2_{all} = \sum_i^{m-1} G^2(L_i) \qquad df_{all} = \sum df(L_i)$$

Nested dichotomies: Examples

$m = 3$

Arthritis treatment

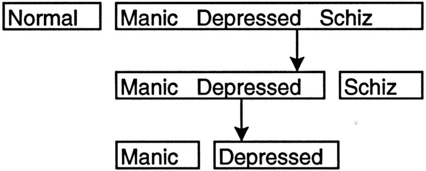


$$L_1 = \log \frac{\pi_1}{\pi_2 + \pi_3}$$

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

$m = 4$

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)

Nested dichotomoies: 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: recoding

In R, first create new variables, `working` and `fulltime`, using the **`recode()`** function in the car:

```
> library(car)    # for data and Anova()
> data(Womenlf)
> Womenlf <- within(Womenlf,{
+   working <- recode(partic, " 'not.work' = 'no'; else = 'yes' ")
+   fulltime <- recode (partic,
+      " 'fulltime' = 'yes'; 'parttime' = 'no'; 'not.work' = NA")})
> some(Womenlf)
```

```
      partic hincome children   region fulltime working
31  not.work      13  present  Ontario     <NA>      no
34  not.work       9   absent  Ontario     <NA>      no
55  parttime       9  present Atlantic       no     yes
86  fulltime      27   absent       BC      yes     yes
96  not.work      17  present  Ontario     <NA>      no
141 not.work      14  present  Ontario     <NA>      no
180 fulltime      13   absent       BC      yes     yes
189 fulltime       9  present Atlantic      yes     yes
234 fulltime       5   absent   Quebec      yes     yes
240 not.work      13  present   Quebec     <NA>      no
```

Nested dichotomies: fitting

Then, fit models for each dichotomy:

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

Some output from `summary(mod.working)`:

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.33583    0.38376   3.481   0.0005 ***
hincome         -0.04231    0.01978  -2.139   0.0324 *
childrenpresent -1.57565    0.29226  -5.391    7e-08 ***
```

Some output from `summary(mod.fulltime)`:

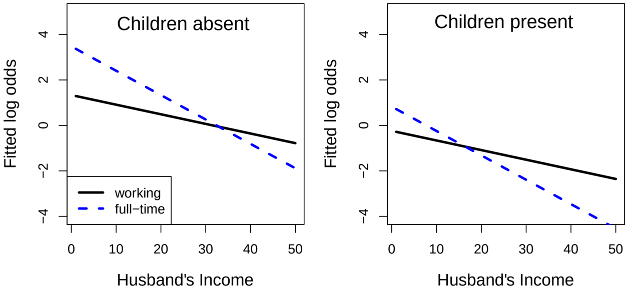
```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.47777    0.76711   4.534 5.80e-06 ***
hincome         -0.10727    0.03915  -2.740  0.00615 **
childrenpresent -2.65146    0.54108  -4.900 9.57e-07 ***
```

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 the **predict()** function.

`type='response'` gives these on the probability scale, whereas `type='link'` (the default) gives these on the logit scale.

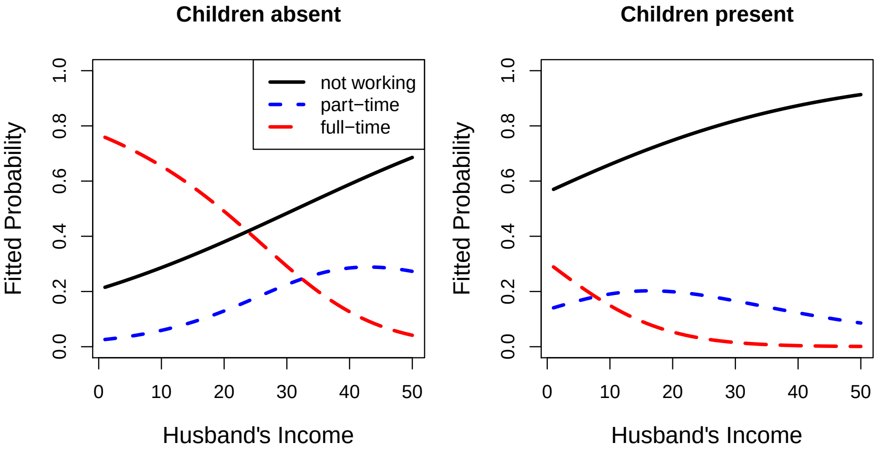
```
> pred <- expand.grid(hincome=1:45, children=c('absent', 'present'))
> # get fitted values for both sub-models
> p.work     <- predict(mod.working, pred, type='response')
> p.fulltime <- predict(mod.fulltime, pred, type='response')
```

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
```

Nested dichotomies in R: plotting

The plot below was produced using the basic R functions `plot()`, `lines()` and `legend()`. See the file `wlf-nested.R` on the course web page for details.



Polytomous response: Generalized Logits

- Models the probabilities of the m response categories as $m-1$ logits comparing each of the first $m-1$ categories to the last (reference) category.
- Logits for any pair of categories can be calculated from the $m-1$ fitted ones.
- With k predictors, $x_1, x_2, \ldots, x_k$, for $j = 1, 2, \ldots, m-1$,

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

  - One set of fitted coefficients, $\boldsymbol{\beta}_j$ for each response category except the last.
  - Each coefficient, β_{hj}, gives the effect on the log odds of a unit change in the predictor x_h that an observation belongs to category j vs. category m.

- Probabilities in response caegories are calculated as:

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

Generalized logit models: Fitting

- In R, the generalized logit model can be fit using the **`multinom()`** function in the nnet
- For interpretation, it is useful to reorder the levels of `partic` so that `not.work` is the baseline level.

```
Womenlf$partic <- ordered(Womenlf$partic,
    levels=c('not.work', 'parttime', 'fulltime'))
library(nnet)
mod.multinom <- multinom(partic ~ hincome + children, data=Womenl
summary(mod.multinom, Wald=TRUE)
Anova(mod.multinom)
```

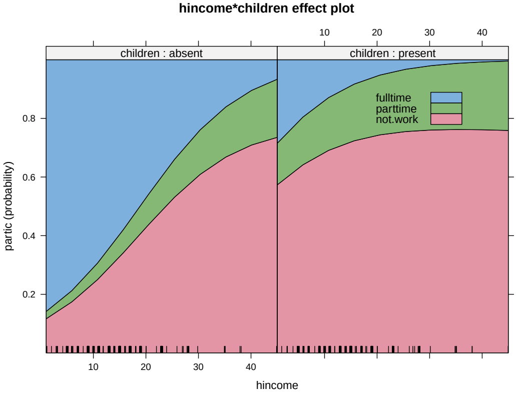
The **`Anova()`** tests are similar to what we got from summing these tests from the two nested dichotomies:

```
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
```

Generalized logit models: Plotting

- As before, it is much easier to interpret a model from a plot than from coefficients, but this is particularly true for polytomous response models
- `style="stacked"` shows cumulative probabilities

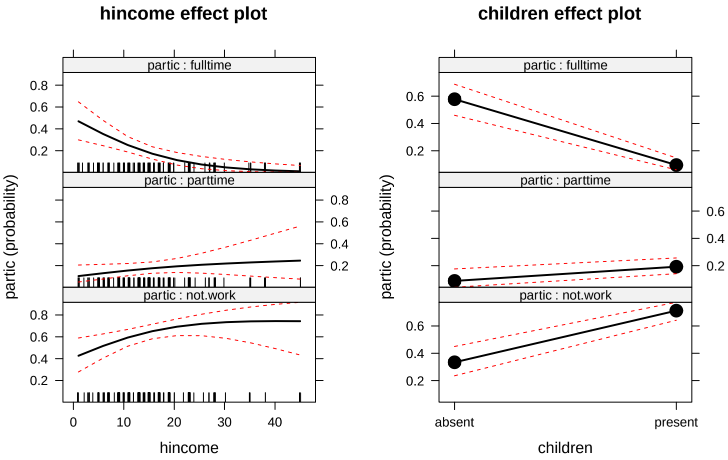
```
library(effects)
plot(effect("hincome*children", mod.multinom), style="stacked")
```



Generalized logit models: Plotting

- You can also view the effects of husband's income and children separately in this main effects model with `plot(allEffects))`.

```
plot(allEffects(mod.multinom), ask=FALSE)
```



Political knowledge & party choice in Britain

Example from Fox & Andersen (2006): Data from 1997 British Election Panel Survey (BEPS)

- **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 using **multinom()** in the nnet package

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

```
Anova 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?

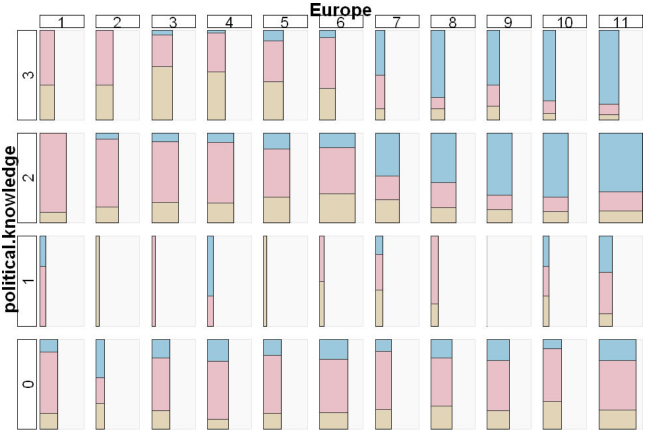
How to understand the *nature* of these effects on party choice?

```
> summary(multinom.mod)
```

```
Call:
multinom(formula = vote ~ age + gender + economic.cond.national +
    economic.cond.household + Blair + Hague + Kennedy + Europe *
    political.knowledge, data = BEPS)

Coefficients:
                 (Intercept)         age gendermale economic.cond.national
Labour               -0.8734 -0.01980     0.1126                 0.5220
Liberal Democrat     -0.7185 -0.01460     0.0914                 0.1451
                 economic.cond.household  Blair   Hague Kennedy    Europe
Labour                          0.178632 0.8236 -0.8684  0.2396 -0.001706
Liberal Democrat                0.007725 0.2779 -0.7808  0.6557  0.068412
                 political.knowledge Europe:political.knowledge
Labour                        0.6583                    -0.1589
Liberal Democrat              1.1602                    -0.1829

Std. Errors:
                 (Intercept)         age gendermale economic.cond.national
Labour                0.6908 0.005364     0.1694                 0.1065
Liberal Democrat      0.7344 0.005643     0.1780                 0.1100
...

Residual Deviance: 2233
AIC: 2277
```

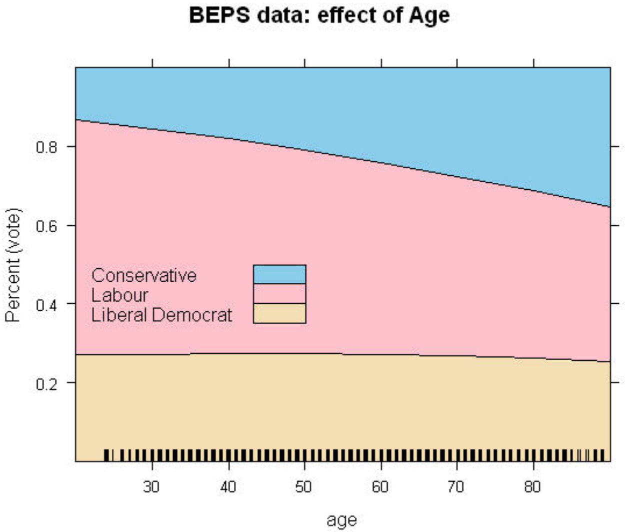
BEPS data: Initial look, relative multiple barcharts

How does party choice— Liberal democrat, Labour, Conservative vary with political knowledge and Europe attitude (high="Eurosceptic")?



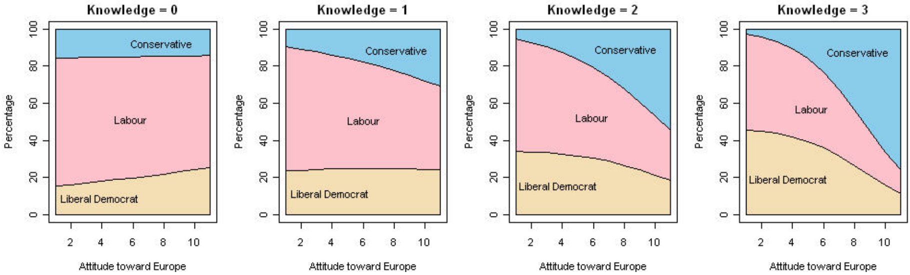
BEPS data: Effect plots to the rescue!

Age effect: Older more likely to vote Conservative



BEPS data: Effect plots to the rescue!

Attitude toward European integration $\times$ political knowledge effect:



- Low knowledge: little relation between attitude and party choice
- As knowledge increases: more Eurosceptic views more likely to support Conservatives
- $\Rightarrow$ 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: Same slopes for all logits
  - Requires proportional odds asumption to be met
  - R: MASS::`polr()`; VGAM::`vglm()`
- **Nested dichotomies**
  - Applies to ordered or unordered categories
  - Fit $m-1$ *separate* independent models $\rightarrow$ Additive χ^2 values
  - R: only need `glm()`
- **Generalized (multinomial) logistic regression**
  - Fit $m-1$ logits as a *single* model
  - Results usually comparable to nested dichotomies
  - R: nnet::`multinom()`