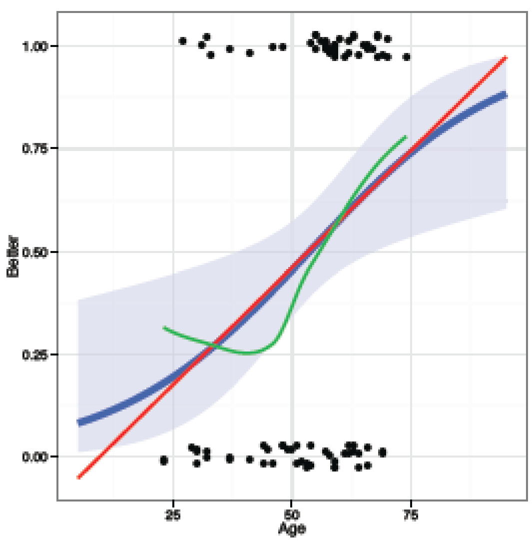
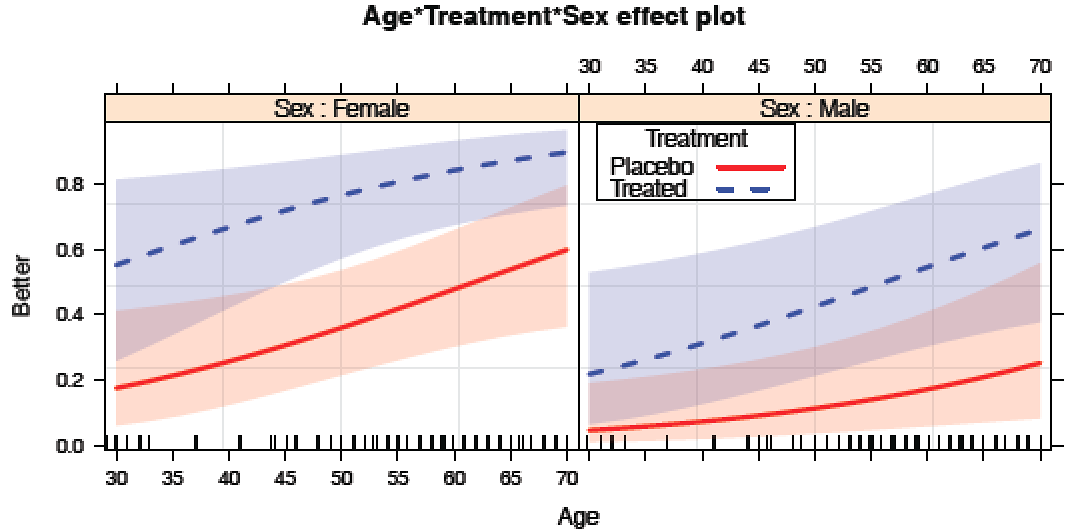




Logistic regression

Psy 6136
Categorical Data
Analysis

Michael Friendly

Psych 6136

http://friendly.github.io/psy6136

Psy 6136
Categorical Data
Analysis

Model-based methods: Overview

Structure

- Explicitly assume some probability distribution for the data, e.g., binomial, Poisson, ...
- Distinguish between the systematic component— explained by the model— and a random component, which is not
- Allow a compact summary of the data in terms of a (hopefully) small number of parameters

Advantages

- Inferences: hypothesis tests *and* confidence intervals
- Can test individual model terms (`anova()`)
- Methods for model selection: adjust balance between goodness-of-fit and parsimony
- Predicted values give model-smoothed summaries for plotting
- $\Longrightarrow$ Interpret the fitted model graphically

loglm() vs. glm()

With **`loglm()`** you can only test overall fit (**`anova()`**) or difference between models (**`Lrstats()`**)

```
> berk.mod1 <- loglm(~ Dept * (Gender + Admit),
data=UCBAdmissions)
> berk.mod2 <- loglm(~(Admit + Dept + Gender)^2,
data=UCBAdmissions)
> anova(berk.mod2)
Call:
loglm(formula = ~(Admit + Dept + Gender)^2, data =
UCBAdmissions)

Statistics:
                   X^2 df P(> X^2)
Likelihood Ratio 20.20  5 0.001144
Pearson          18.82  5 0.00207
```

What we can say:

Even the model with all pairwise associations fits poorly

Comparing models with **anova()** and **LRstats()**

```
> anova(berk.mod1, berk.mod2, test="Chisq")
LR tests for hierarchical log-linear models

Model 1:
 ~Dept * (Gender + Admit)
Model 2:
 ~(Admit + Dept + Gender)^2

          Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)
Model 1      21.74  6
Model 2      20.20  5      1.531         1         0.21593
Saturated     0.00  0     20.204         5         0.00114

> LRstats(berk.mod1, berk.mod2)
Likelihood summary table:
          AIC BIC LR Chisq Df Pr(>Chisq)
berk.mod1 217 238     21.7  6     0.0014 **
berk.mod2 217 240     20.2  5     0.0011 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

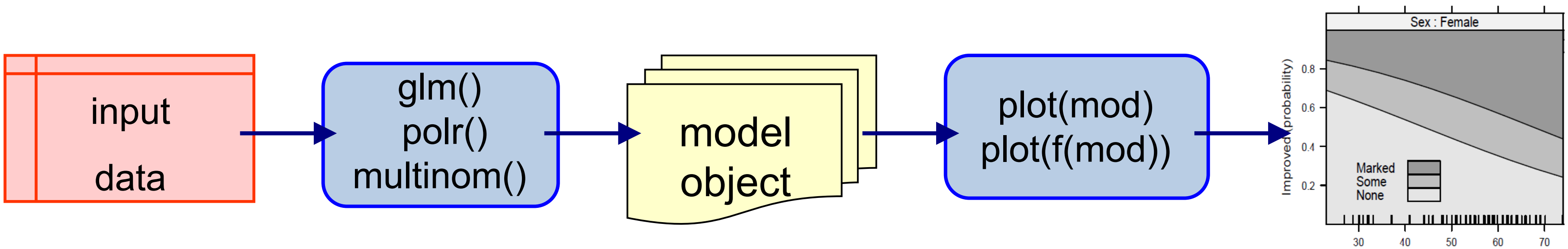
loglm() vs. glm()

With **glm()** you can test individual terms using **anova()** or **car::Anova()**

```
> berkeley <- as.data.frame(UCBAdmissions)
> berk.glm2 <- glm(Freq ~ (Dept+Gender+Admit)^2, data=berkeley,
+                  family="poisson")
> anova(berk.glm2, test="Chisq")
Analysis of Deviance Table

Model: poisson, link: log
Response: Freq

Terms added sequentially (first to last)

             Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                            23       2650
Dept          5      160        18       2491   <2e-16 ***
Gender        1      163        17       2328   <2e-16 ***
Admit         1      230        16       2098   <2e-16 ***
Dept:Gender   5     1221        11        877   <2e-16 ***
Dept:Admit    5      855         6         22   <2e-16 ***
Gender:Admit  1        2         5         20     0.22
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fitting & graphing models: Overview

Object-oriented approach in R:



- Fit model (`obj <- glm(...)`) → a model object
- `print(obj)` and `summary(obj)` → numerical results
- `anova(obj)` and `Anova(obj)` → tests for model terms
- `update(obj)`, `add1(obj)`, `drop1(obj)` for model selection

Plot methods:

- `plot(obj)` often gives diagnostic plots
- Other plot methods:
  - Mosaic plots: `mosaic(obj)` for `"loglm"` and `"glm"` objects
  - Effect plots: `plot(Effect(obj))` for nearly all linear models
  - Influence plots (car): `influencePlot(obj)` for `"glm"` objects

Objects & methods

How this works:

- Model objects have a "class" attribute:
  - `loglm()`: `"loglm"`
  - `glm()`: `c("glm", "lm")` — inherits also from `lm()`
- Class-specific methods have names like `method.class`, e.g., `plot.glm()`, `mosaic.loglm()`
- Generic functions (`print()`, `summary()`, `plot()` ...) call the appropriate method for the class

```
arth.mod <- glm(Better ~ Age + Sex + Treatment, data=Arthritis)
class(arth.mod)

## [1] "glm" "lm"
```

Objects & methods

Methods for “glm” objects

```
> library(MASS); library(vcdExtra)
> methods(class="glm")
 [1] add1              addterm           anova             Anova
 [5] asGnm             assoc             avPlot            avPlot3d
 [9] Boot              bootCase          brief             ceresPlot
[13] coerce            confidenceEllipse confint           Confint
[17] cooks.distance    deviance          drop1             dropterm
[21] effects           extractAIC        family            formula
[25] gamma.shape       influence         initialize        leveragePlot
[29] linearHypothesis  logLik            mcPlot            mmp
[33] model.frame       modFit            mosaic            ncvTest
[37] nobs              predict           print             profile
[41] qqPlot            residualPlot      residualPlots     residuals
[45] rootogram         rstandard         rstudent          S
[49] show              sieve             sigmaHat          slotsFromS3
[53] summary           vcov              weights
see '?methods' for accessing help and source code
```

There are many, many **`plot()`** methods for different types of objects
e.g., **`plot()`** for a "glm" object → **`plot.glm()`**

```
> methods("plot")
 [1] plot,ANY-method      plot,color-method    plot.acf*
 [4] plot.ca*             plot.correspondence* plot.data.frame*
 [7] plot.decomposed.ts*  plot.default         plot.dendrogram*
[10] plot.density*        plot.ecdf            plot.factor*
[13] plot.formula*        plot.function        plot.gnm*
[16] plot.goodfit*        plot.hcl_palettes*   plot.hclust*
[19] plot.histogram*      plot.HLtest*         plot.HoltWinters*
[22] plot.isoreg*         plot.lda*            plot.lm*
[25] plot.loddsratio*     plot.loglm*          plot.mca*
[28] plot.medpolish*      plot.mjca*           plot.mlm*
[31] plot.ppr*            plot.prcomp*         plot.princomp*
[34] plot.profile*        plot.profile.gnm*    plot.profile.nls*
[37] plot.qv*             plot.raster*         plot.ridgelm*
[40] plot.rootogram*      plot.shingle*        plot.spec*
[43] plot.stepfun         plot.stl*            plot.structable*
[46] plot.table*          plot.trellis*        plot.ts
[49] plot.tskernel*       plot.TukeyHSD*       plot.zoo*
see '?methods' for accessing help and source code
```

Modeling approaches: Overview

Association models

- Loglinear models

  (contingency table form)

  [Admit][Gender Dept]

  [Admit Dept][Gender Dept]

  [AdmitDept][AdmitGender][GenderDept]

- Poisson GLMs

  (Frequency data frame)

  Freq ~ Admit + Gender * Dept

  Freq ~ Admit*Dept + Gender*Dept

  Freq ~ Admit*(Dept + Gender) +
  Gender*Dept

- Ordinal variables

  Freq ~ right + left + Diag(right, left)

  Freq ~ right + left + Symm(right, left)

Response models

- Binary response
- Categorical predictors: logit models

  logit(Admit) ~ 1

  logit(Admit) ~ Dept

  logit(Admit) ~ Dept + Gender

- Continuous/mixed predictors
- Logistic regression models

  Pr(Admit) ~ Dept + Gender + Age + GRE

- Polytomous response
- Ordinal: proportional odds model

  Improve ~ Age + Sex + Treatment

- General multinomial model

  WomenWork ~ Kids + HusbandIncome

Logistic regression

Response variable

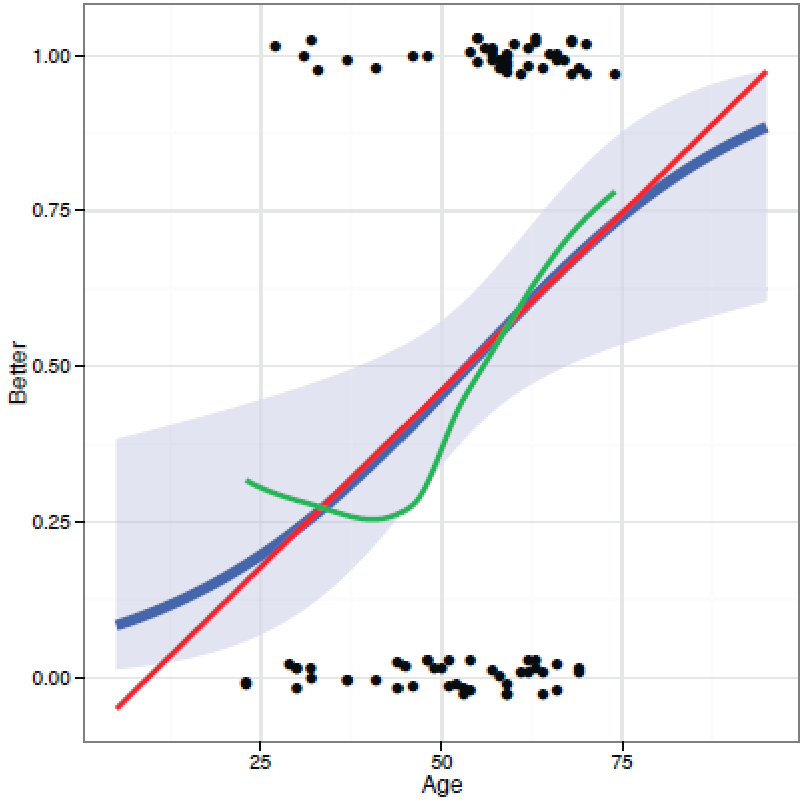
- Binary response: success/failure, vote: yes/no
- Binomial data: x successes in n trials (grouped data)
- Ordinal response: none $<$ some $<$ severe depression
- Polytomous response: vote Liberal, Tory, NDP, Green

Explanatory variables

- Quantitative regressors: age, dose
- Transformed regressors: $\sqrt{\text{age}}$, log(dose)
- Polynomial regressors: age^2, age^3, $\cdots$ (or better: splines)
- Categorical predictors: treatment, sex (dummy variables, contrasts)
- Interaction regessors: treatment $\times$ age, sex $\times$ age

This is exactly the same as in classical ANOVA, regression models

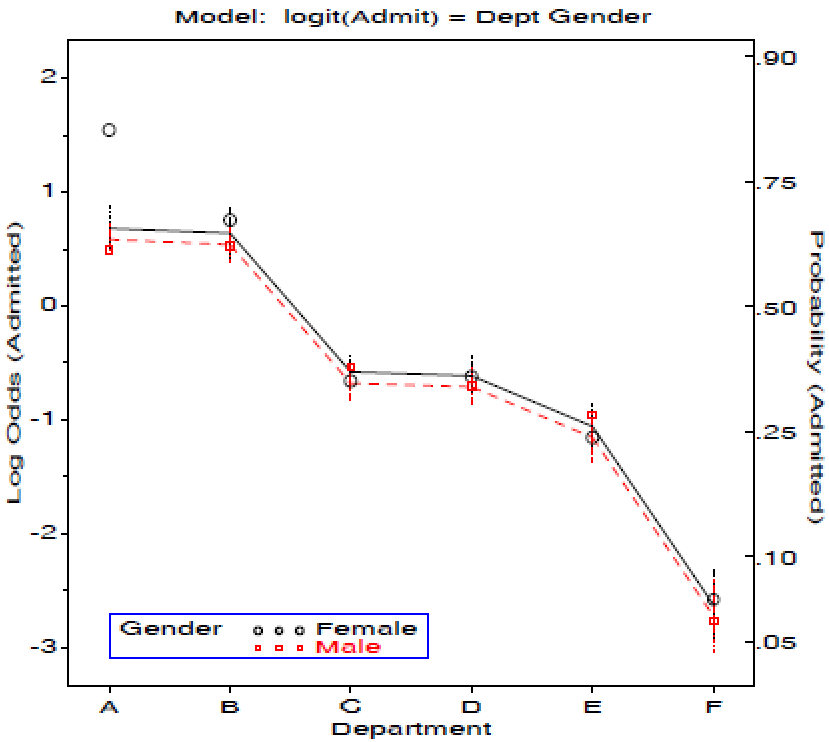
Example: Arthritis treatment



- The response variable, `Improved` is ordinal: `"None" < "Some" < "Marked"`
- A binary logistic model can consider just `Better = (Improved>"None")`
- Other important predictors: Sex, Treatment

- Main Q: how does treatment affect outcome?
- How does this vary with Age and Sex?
- This plot shows the binary observations, with several model-based smoothings

Example: Berkeley admissions



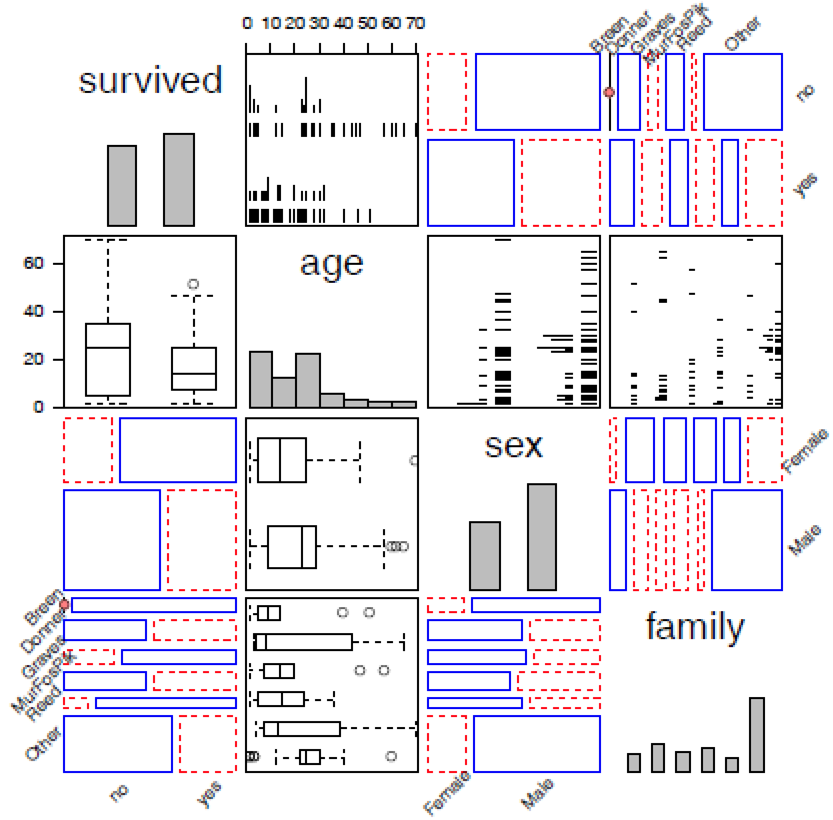
- Admit/Reject can be considered a binomial response for each Dept and Gender
- Logistic regression here is analogous to an ANOVA model, but for log odds(Admit)
- (With categorical predictors, these are often called logit models)

- Every such model has an equivalent loglinear model form.
- This plot shows fitted logits for the main effects model, `Dept + Gender`

Example: Survival in the Donner party

- Binary response: `survived`
- Categorical predictors: `sex`, `family`
- Quantitative predictor: `age`

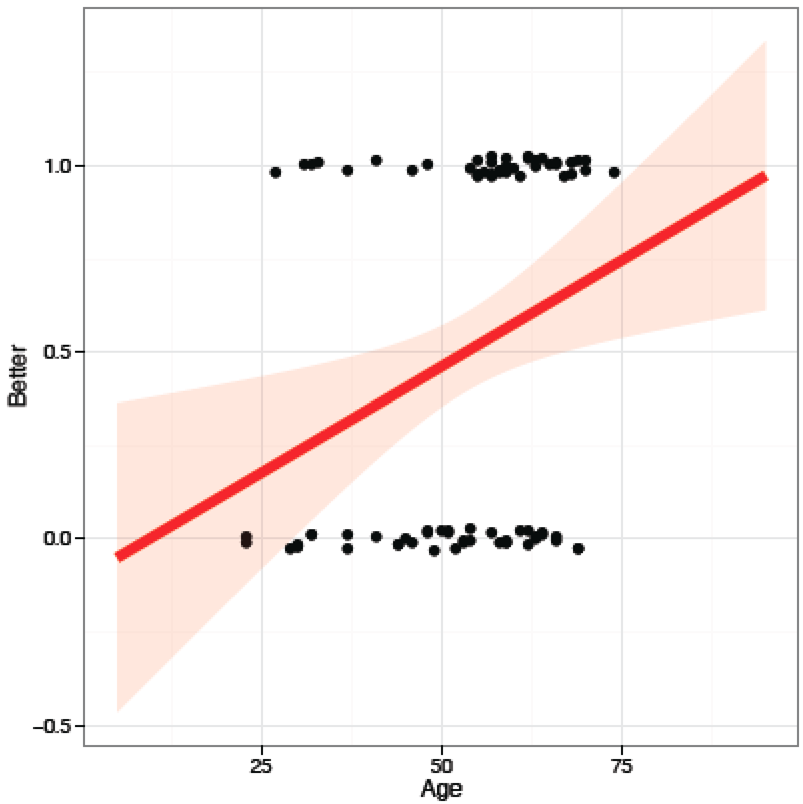
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?
- This is a generalized pairs plot, with different plots for each pair



Binary response: What's wrong with OLS?

- For a binary response, $Y \in (0, 1)$, want to predict $\pi = \Pr(Y = 1 \mid x)$
- A linear probability model uses classical linear regression (OLS)

- Problems:
  - Gives predicted values and CIs outside $0 \leq \pi \leq 1$
  - Homogeneity of variance is violated: $\mathcal{V}(\hat{\pi}) = \hat{\pi}(1 - \hat{\pi}) \neq$ constant
  - Inferences, hypothesis tests are wrong!



Linear regression vs Logistic regression

OLS regression:
- Assume y|x ~ N(0, σ^2)

Logistic regression:
- Assume Pr(y=1|x) ~ binomial(p)

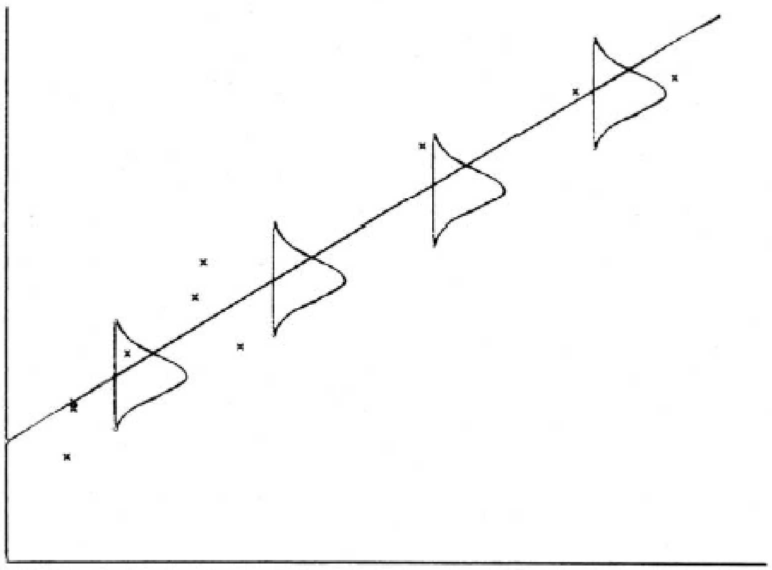


Fig. 2.1. Graphical representation of a simple linear normal regression.

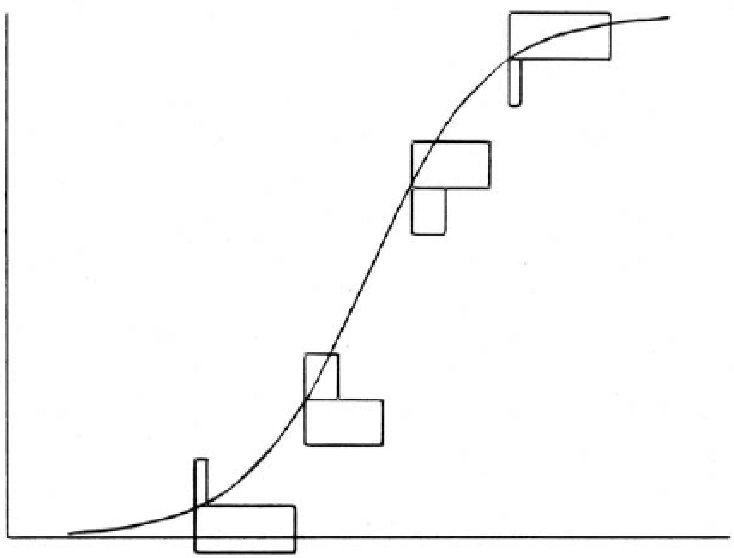


Fig. 2.2. Graphical representation of a simple linear logistic regression.

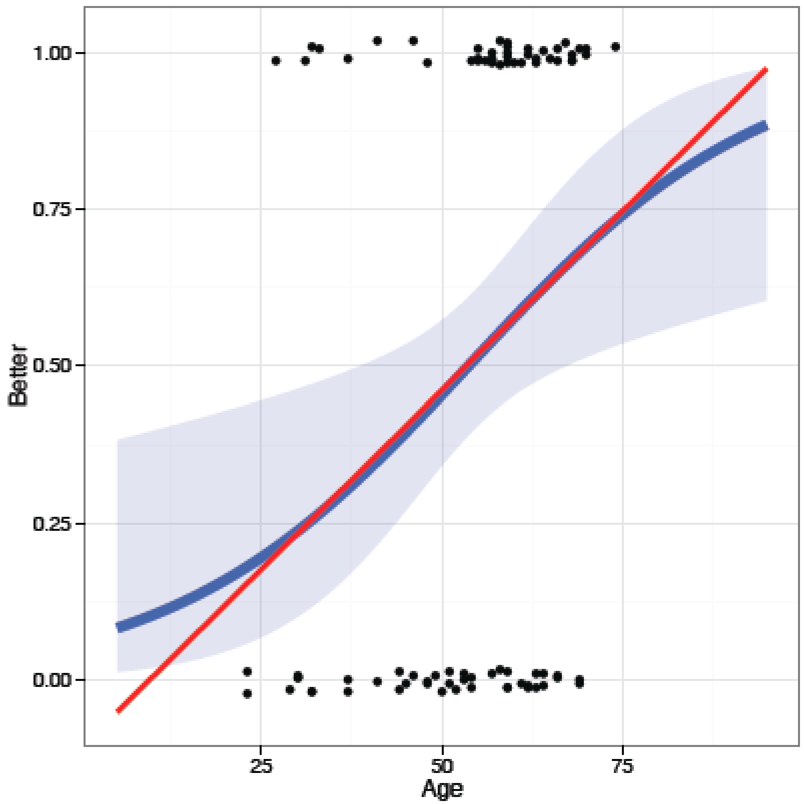
y linear with x
constant residual variance

y ~ logit (x)
non-constant residual variance ~ p (1-p)

Logistic regression

- Logistic regression avoids these problems
- Models $\text{logit}(\pi_i) \equiv \log[\pi/(1-\pi)]$
- logit is interpretable as "log odds" that $Y = 1$

- A related probit model gives very similar results, but is less interpretable
- For $0.2 \leq \pi \leq 0.8$ fitted values are close to those from linear regression.



Logistic regression: One predictor

For a single quantitative predictor, x, the simple linear logistic regression model posits a linear relation between the ***log odds*** (or ***logit***) of $\Pr(Y = 1)$ and x,

$$\text{logit}[\pi(x)] \equiv \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x \ .$$

- When $\beta > 0$, $\pi(x)$ and the log odds increase as x increases; when $\beta < 0$ they decrease with x.
- This model can also be expressed as a model for the probabilities $\pi(x)$

$$\pi(x) = \text{logit}^{-1}[\pi(x)] = \frac{1}{1 + \exp[-(\alpha + \beta x)]}$$

Thinking logistically:

- Model is for the log odds of the marked response, Y = 1
- Can always back transform with logit^{-1} to get probability of Y = 1

Logistic regression: One predictor

The coefficients, α, β of this model have simple interpretations in terms of odds & log odds

$$\text{logit}[\pi(x)] \equiv \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \qquad \text{odds}(Y=1) \equiv \frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x}$$

β is the change in log odds for a unit increase in x

$\rightarrow$The odds of Y=1 are multiplied by e^{β} for each unit increase in x

α is the log odds when x=0

$\rightarrow$The odds of Y=1 when x=0 is e^{α}

In R, use exp(coef(model)) to get these values

Another interpretation: In terms of probability, the slope of the logistic regression curve is $\beta\pi(1-\pi)$

This has the maximum value $\beta/4$ when $\pi = ½$

Logistic regression: Multiple predictors

- For a binary response, $Y \in (0, 1)$, let $\boldsymbol{x}$ be a vector of p regressors, and π_i be the probability, $\Pr(Y = 1 \mid \boldsymbol{x})$.
- The logistic regression model is a linear model for the *log odds*, or *logit* that $Y = 1$, given the values in $\boldsymbol{x}$,

$$\begin{aligned} \text{logit}(\pi_i) \equiv \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \alpha + \boldsymbol{x}_i^\mathsf{T}\beta \\ &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \end{aligned}$$

- An equivalent (non-linear) form of the model may be specified for the probability, π_i, itself,

$$\pi_i = \{1 + \exp(-[\alpha + \boldsymbol{x}_i^\mathsf{T}\beta])\}^{-1}$$

- The logistic model is also a *multiplicative* model for the odds of "success,"

$$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \boldsymbol{x}_i^\mathsf{T}\beta) = \exp(\alpha)\exp(\boldsymbol{x}_i^\mathsf{T}\beta)$$

  Increasing x_{ij} by 1 increases $\text{logit}(\pi_i)$ by β_j, and multiplies the odds by e^{β_j}.

Fitting the logistic regression model

Logistic regression models are the special case of generalized linear models, fit in R using `glm(..., family=binomial)`
For this example, we define **`Better`** as any improvement at all

```
> data(Arthritis, package="vcd")
> Arthritis$Better <- as.numeric(Arthritis$Improved > "None")
```

Fit and print:

```
> (arth.logistic <- glm(Better ~ Age, data=Arthritis, family=binomial))

Call:  glm(formula = Better ~ Age, family = binomial, data = Arthritis)

Coefficients:
(Intercept)          Age
    -2.6421       0.0492

Degrees of Freedom: 83 Total (i.e. Null);  82 Residual
Null Deviance:       116
Residual Deviance: 109    AIC: 113
```

The summary() method gives details and tests of coefficients

```
> summary(arth.logistic)

Call:
glm(formula = Better ~ Age, family = binomial, data = Arthritis)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5106  -1.1277   0.0794   1.0677   1.7611

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.6421     1.0732   -2.46    0.014 *
Age           0.0492     0.0194    2.54    0.011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 116.45  on 83  degrees of freedom
Residual deviance: 109.16  on 82  degrees of freedom
```

Interpreting coefficients

```
> coef(arth.logistic)
(Intercept)        Age
  -2.64207     0.04925
```

```
> exp(coef(arth.logistic))
(Intercept)        Age
   0.07121     1.05048
> exp(10*coef(arth.logistic)[2])
  Age
1.636
```

Interpretations:

- log odds(Better) increase by $\beta = 0.0492$ for each year of age
- odds(Better) multiplied by $e^{\beta} = 1.05$ for each year of age— a 5% increase
- over 10 years, odds(Better) are multiplied by $\exp(10 \times 0.0492) = 1.64$, a 64% increase.
- Pr(Better) increases by $\beta/4 = 0.0123$ for each year (near $\pi = \frac{1}{2}$)

Multiple predictors

The main interest here is the effect of Treatment. Sex and Age are control variables. Fit the main effects model (no interactions):

$$\text{logit}(\pi_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_2 x_{i2}$$

where x_1 is `Age` and x_2 and x_3 are the factors representing `Sex` and `Treatment`, respectively. R uses dummy (0/1) variables for factors.

$$x_2 = \begin{cases} 0 & \text{if Female} \\ 1 & \text{if Male} \end{cases} \qquad x_3 = \begin{cases} 0 & \text{if Placebo} \\ 1 & \text{if Treatment} \end{cases}$$

- α doesn't have a sensible interpretation here. Why?
- β_1: increment in log odds(Better) for each year of age.
- β_2: difference in log odds for male as compared to female.
- β_3: difference in log odds for treated vs. the placebo group

Multiple predictors: Fitting

Fit the main effects model. Use I(Age – 50) to center Age, making interpretable

```
arth.logistic2 <- glm(Better ~ I(Age - 50) + Sex + Treatment,
                      data=Arthritis, family=binomial)
```

lmtest::coeftest() gives just the tests of coefficients provided by summary()

```
> lmtest::coeftest(arth.logistic2)

z test of coefficients:

                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.5781     0.3674   -1.57    0.116
I(Age - 50)        0.0487     0.0207    2.36    0.018 *
SexMale           -1.4878     0.5948   -2.50    0.012 *
TreatmentTreated   1.7598     0.5365    3.28    0.001 **
```

broom::glance() gives model fit statistics

```
> broom::glance(arth.logistic2)
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
          <dbl>   <int>  <dbl> <dbl> <dbl>    <dbl>       <int> <int>
1          116.      83  -46.0  100.  110.     92.1          80    84
```

Interpreting coefficients

```
> cbind(coef=coef(arth.logistic2),
+       OddsRatio=exp(coef(arth.logistic2)),
exp(confint(arth.logistic2)))
Waiting for profiling to be done...
                    coef OddsRatio  2.5 % 97.5 %
(Intercept)      -0.5781     0.561 0.2647  1.132
I(Age - 50)       0.0487     1.050 1.0100  1.096
SexMale          -1.4878     0.226 0.0652  0.689
TreatmentTreated  1.7598     5.811 2.1187 17.727
```

- $\alpha = -0.578$: At age 50, females given placebo have odds(Better) of $e^{-0.578} = 0.56$.
- $\beta_1 = 0.0487$: Each year of age multiplies odds(Better) by $e^{0.0487} = 1.05$, a 5% increase.
- $\beta_2 = -1.49$: Males $e^{-1.49} = 0.26 \times$ less likely to show improvement as females. (Or, females $e^{1.49} = 4.437 \times$ more likely than males.)
- $\beta_3 = 1.76$: Treated $e^{1.76}$=5.81 $\times$ more likely Better than Placebo

Hypothesis testing: Questions

- **Overall test**: How does my model, $\text{logit}(\pi) = \alpha + \mathbf{x}^{\mathsf{T}}\beta$ compare with the null model, $\text{logit}(\pi) = \alpha$?

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

- **One predictor**: Does x_k significantly improve my model? Can it be dropped?

$$H_0 : \beta_k = 0 \quad \text{given other predictors retained}$$

- **Lack of fit**: How does my model compare with a perfect model (saturated model)?

For ANOVA, regression, these tests are carried out using F-tests and t-tests.
In logistic regression (fit by maximum likelihood) we use

- F-tests $\rightarrow$ likelihood ratio G^2 tests
- t-tests $\rightarrow$ Wald z or χ^2 tests

Maximum likelihood estimation

In classical linear models using **`lm()`**, we fit using ordinary least squares.
All **`glm()`** models use maximum likelihood estimation– better properties

- Likelihood, $\mathcal{L} = \Pr(data \mid model)$, as function of model parameters
- For case i,

$$\mathcal{L}_i = \begin{cases} p_i & \text{if } Y = 1 \\ 1 - p_i & \text{if } Y = 0 \end{cases} = p_i^{Y_i}(1 - p_i^{Y_i}) \qquad \text{where} \qquad p_i = 1/(1 + \exp(\mathbf{x}_i \boldsymbol{\beta}))$$

- Under independence, joint likelihood is the product over all cases

$$\mathcal{L} = \prod_i^n p_i^{Y_i}(1 - p_i^{Y_i})$$

- $\Longrightarrow$ Find estimates $\widehat{\beta}$ that maximize $\log \mathcal{L}$. Iterative, but this solves the "estimating equations"

$$\boldsymbol{X}^\mathsf{T} \boldsymbol{y} = \boldsymbol{X}^\mathsf{T} \widehat{\boldsymbol{p}}$$

Overall model tests

Likelihood ratio test (G^2)

- Compare nested models, similar to F tests in OLS
- Let L_1 = maximized value for our model

  $\text{logit}(\pi_i) = \beta_0 + \mathbf{x}^T_i \boldsymbol{\beta}$ w/ k predictors
- Let L_0 = maximized likelihood for the null model

  $\text{logit}(\pi_i) = \beta_0$ under H_0: $\beta_1 = \beta_2 = \ldots = \beta_k$
- Likelihood ratio test statistic:

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = 2(\log L_1 - \log L_0) \sim \chi^2_k$$

Wald tests & confidence intervals

- Analogous to *t*-tests in OLS
- Test H_0: $\beta_i = 0$

$$z = \frac{b_i}{s(b_i)} \sim \mathcal{N}(0,1) \quad \text{or} \quad z^2 \sim \chi_1^2$$

- Confidence interval

$$b_i \pm z_{1-\alpha/2}\, s(b_i)$$

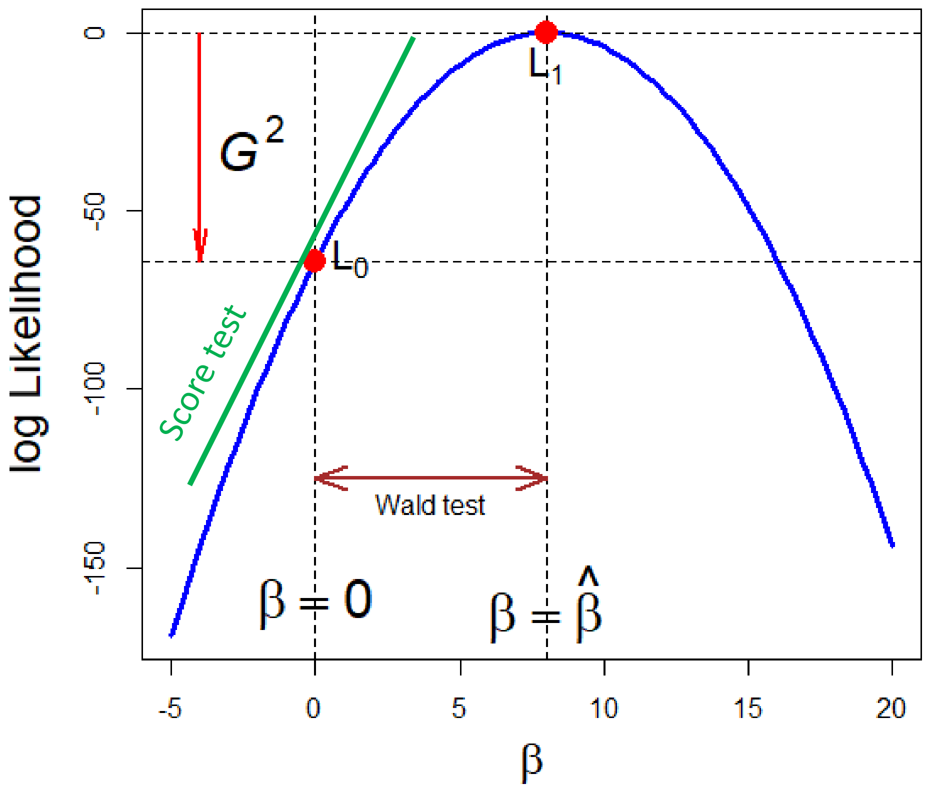
```
> r1 <- lmtest::coeftest(arth.logistic2)
> r2 <- confint(arth.logistic2)
Waiting for profiling to be done...
> cbind(r1, r2)
                 Estimate Std. Error z value Pr(>|z|) 2.5 % 97.5 %
(Intercept)        -0.578      0.367    -1.6    0.116 -1.33  0.124
I(Age - 50)         0.049      0.021     2.4    0.018  0.01  0.092
SexMale            -1.488      0.595    -2.5    0.012 -2.73 -0.372
TreatmentTreated    1.760      0.536     3.3    0.001  0.75  2.875
```

LR, Wald & Score tests

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.3859	3	<.0001
Score	22.0051	3	<.0001
Wald	17.5147	3	0.0006

H0: $\beta_1 = \beta_2 = \beta_3 = 0$



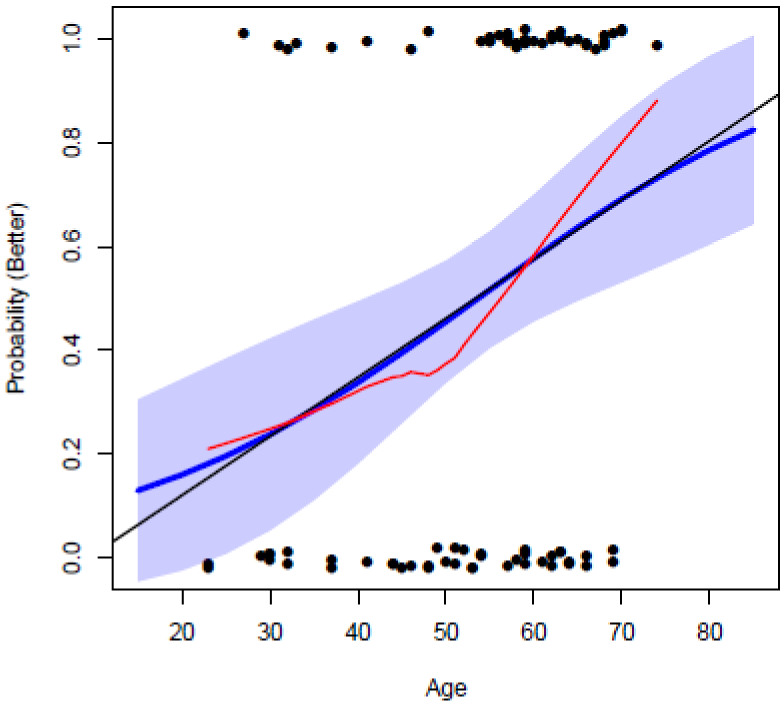
Different ways to measure departure from H_0: $\boldsymbol{\beta} = 0$

- LR test: diff[ce] in log L
- Wald test: $(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^2$
- Score test: slope at $\boldsymbol{\beta} = 0$

Plotting logistic regression data

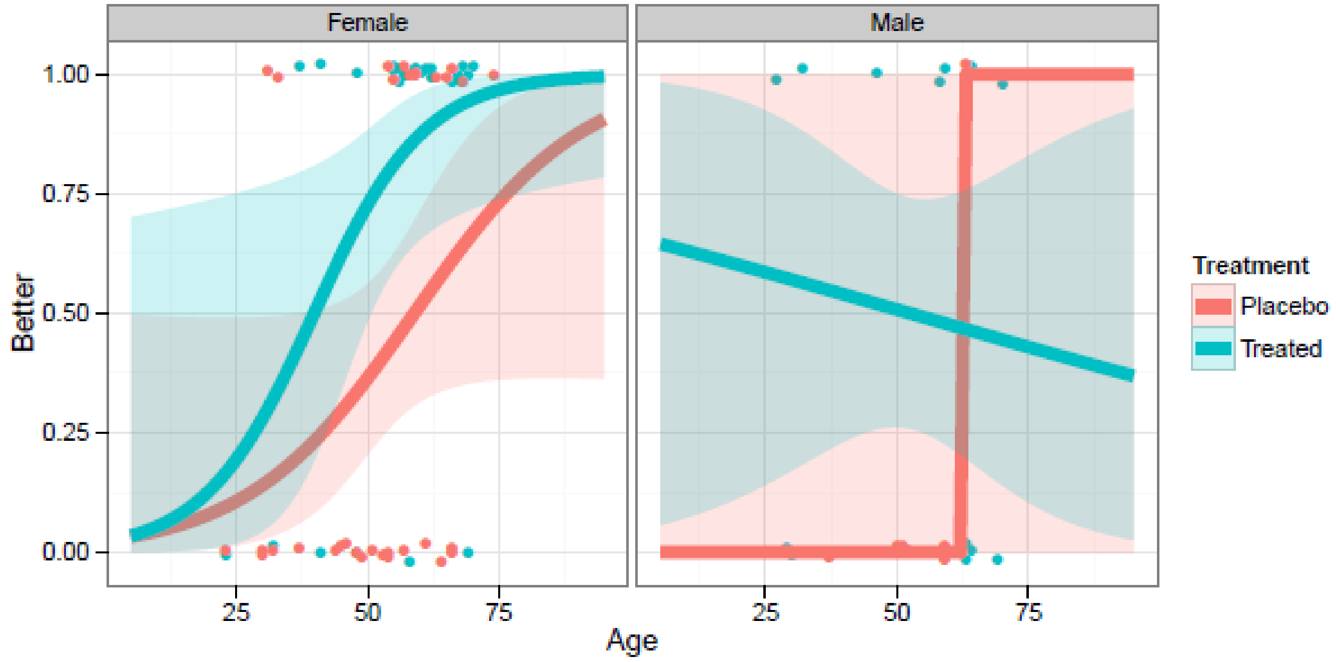
Plotting a binary response together with a fitted logistic model can be difficult because the 0/1 response leads to much overplottting.

- Need to jitter the points
- Useful to show the fitted logistic curve
- Confidence band gives a sense of uncertainty
- Adding a non-parametric (loess) smooth shows possible nonlinearity
- NB: Can plot either on the response scale (probability) or the link scale (logit) where effects are linear



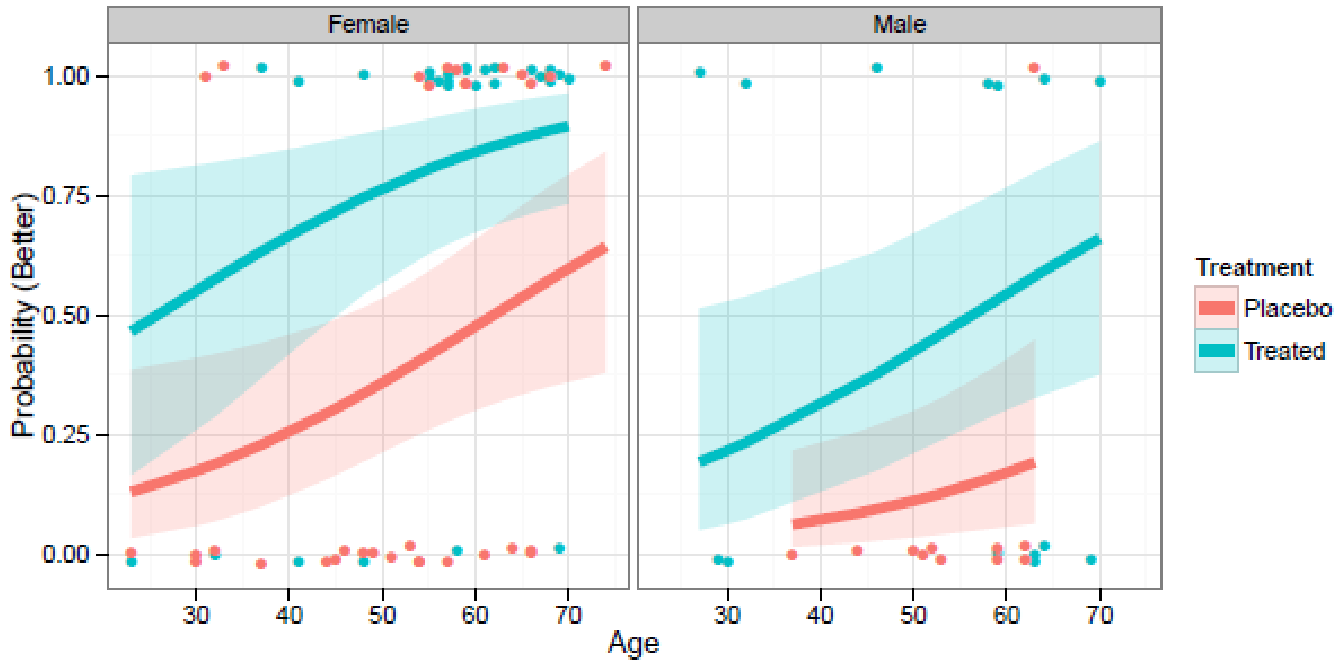
Types of plots

- Conditional plots: Stratified plot of Y or logit(Y) vs. one X, conditioned by other predictors--- only that subset is plotted for each



Types of plots

- Full-model plots: Plot of fitted response surface, showing all effects; usually shown in several panels



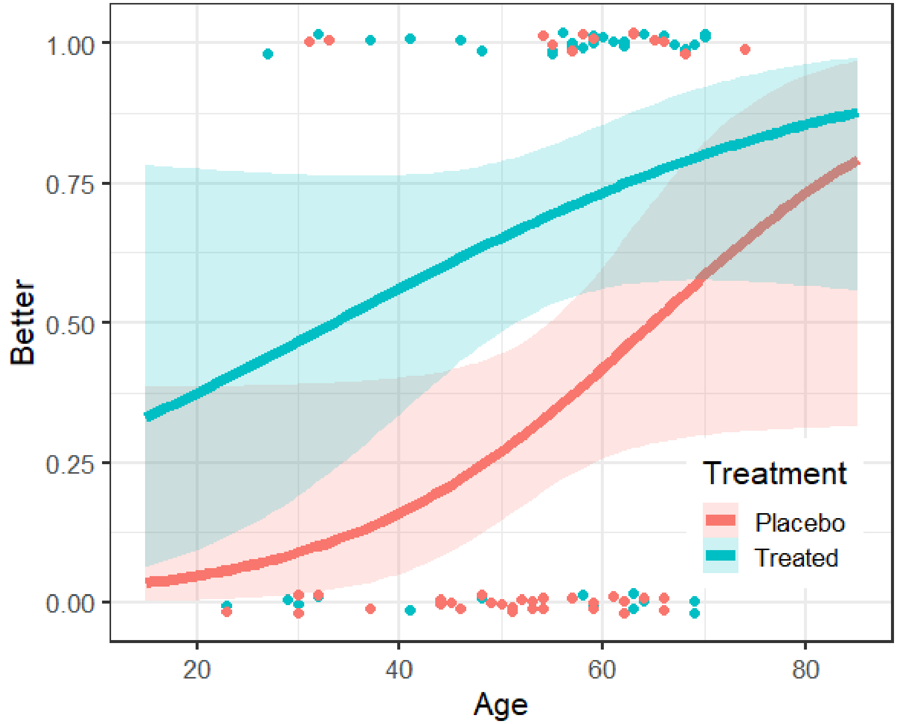
Types of plots

- Effect plots: plots of predicted effects for terms in the model, averaged over predictors not shown in a given plot

Conditional plots with ggplot2

Plot Arthritis data by Treatment, ignoring Sex; overlay fitted logistic reg. lines

```
gg <- ggplot(Arthritis, aes(Age, Better, color=Treatment)) +
  xlim(15, 85) +
  geom_jitter(height = 0.02, width = 0, size=2) +
  stat_smooth(method = "glm", family = binomial, alpha = 0.2,
              aes(fill=Treatment), size=2.5, fullrange=TRUE) +
  theme_bw(base_size = 16) + theme(legend.position = c(.85, .2))
gg   # show the plot
```



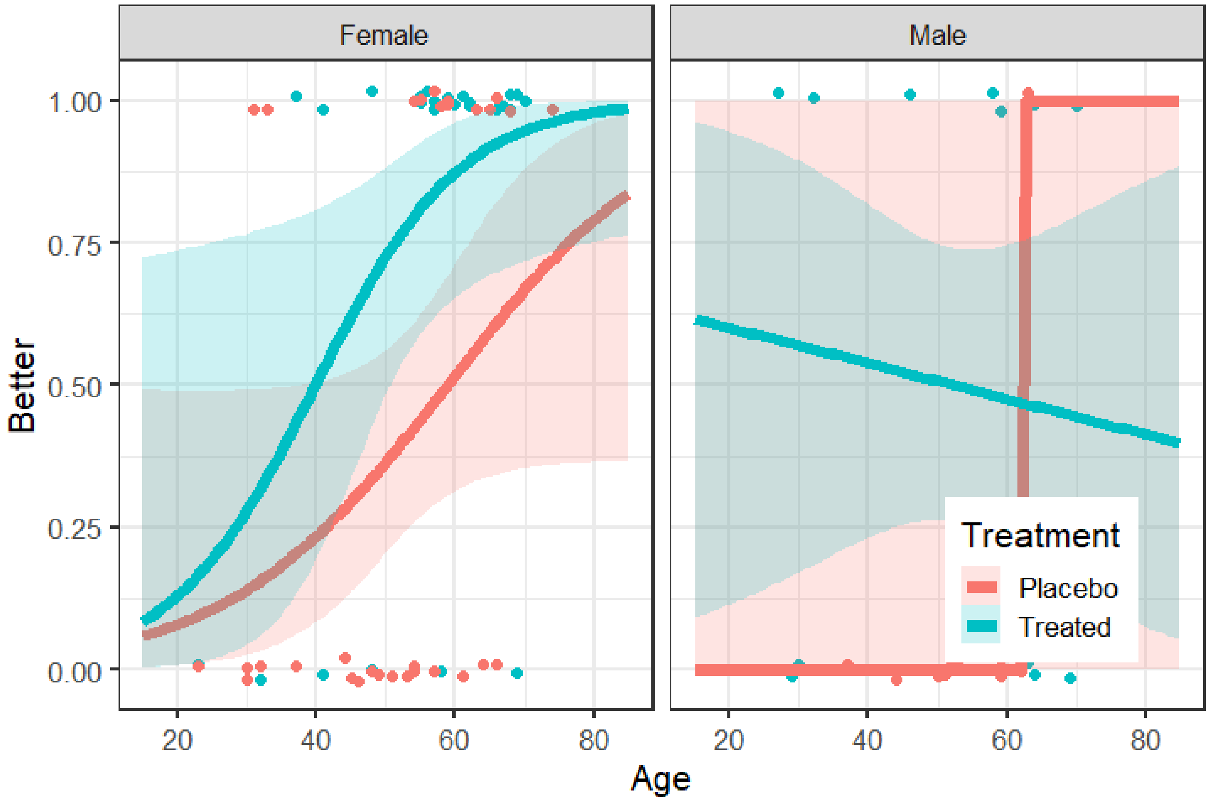
geom_jitter() shows the observations more distinctly

Fitted lines use method="glm", family=binomial

Conditional plots with ggplot2

Can show the conditional plots for M & F, simply by faceting by Sex

```
gg + facet_wrap(~ Sex)
```



Only the data for each Sex is used in each plot

Plotting the data points shows that the data for males is too thin to give good estimates of separate regression

Full-model plots

Full-model plots show the fitted values on the logit scale or on the response scale (probability), usually with confidence bands. This often requires a bit of custom programming.
Steps:

- Obtain fitted values with `predict(model, se.fit=TRUE)`— `type="link"` (logit) is the default
- Can use `type="response"` for probability scale
- Join this to your data (`cbind()`)
- Plot as you like: `plot()`, `ggplot()`, · · ·

```
> arth.fit2 <- cbind(Arthritis,
+                    predict(arth.logistic2, se.fit = TRUE))
> head(arth.fit2[,-9], 4)
  ID Treatment  Sex Age Improved Better   fit se.fit
1 57   Treated Male  27     Some      1 -1.43  0.758
2 46   Treated Male  29     None      0 -1.33  0.728
3 77   Treated Male  30     None      0 -1.28  0.713
4 17   Treated Male  32   Marked      1 -1.18  0.684
```

Plotting with ggplot2

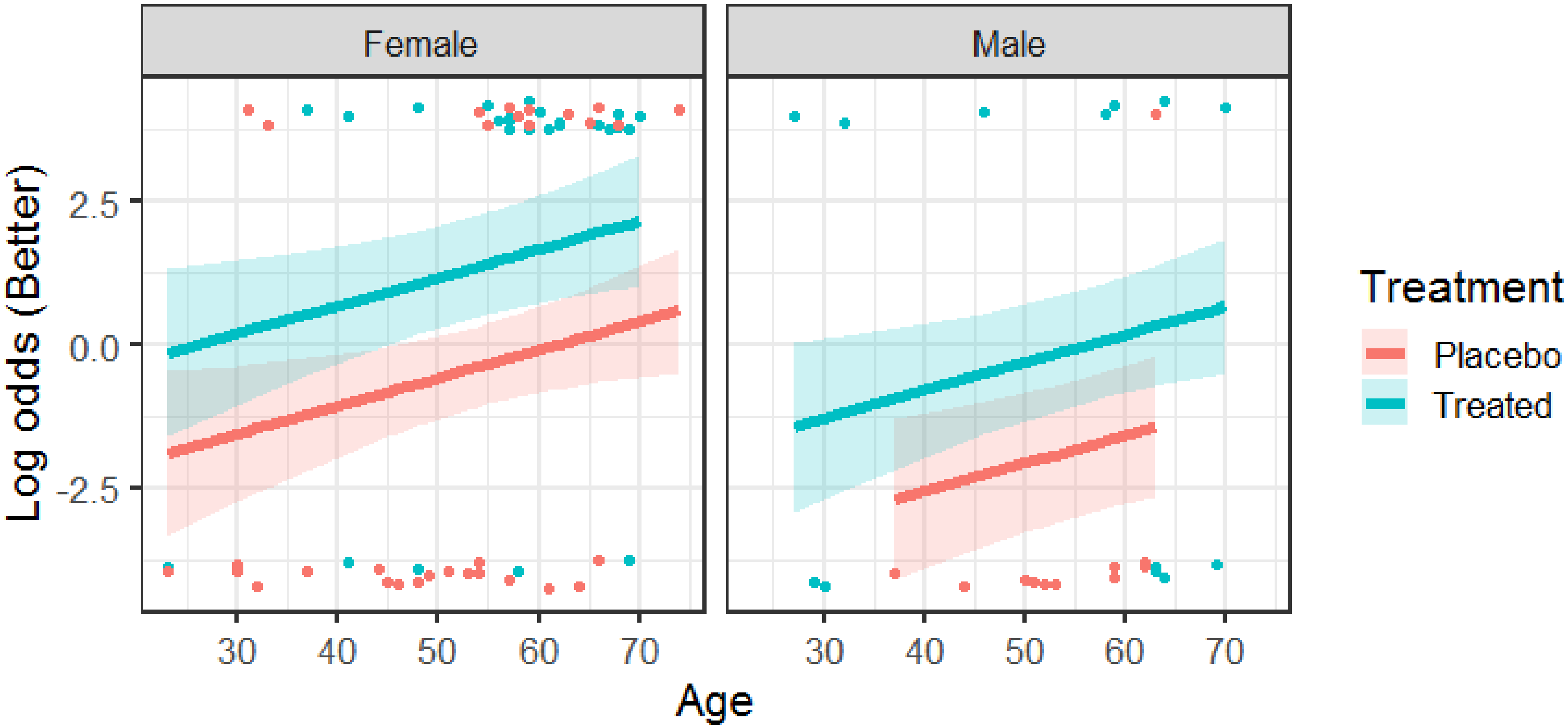
Plot the fitted log odds, confidence band and observations

```
arth.fit2 <- arth.fit2 |>
  mutate(obs = ifelse(Better==0, -4, 4))   # show obs at -4, 4

ggplot( arth.fit2, aes(x=Age, y=fit, color=Treatment)) +
  geom_line(size = 2) +
  geom_ribbon(aes(ymin = fit - 1.96 * se.fit,
                  ymax = fit + 1.96 * se.fit,
                  fill = Treatment), alpha = 0.2,
              color = "transparent") +
  labs(x = "Age", y = "Log odds (Better)") +
  geom_jitter(aes(y=obs), height=0.25, width=0) +
  facet_wrap(~ Sex) +
  theme_bw(base_size = 16)
```

Using color=Treatment gives separate points and lines for the two groups

Full-model plot

Plotting on the logit scale shows the additive effects of age, treatment and sex
NB: easier to compare the treatment groups within the same panel

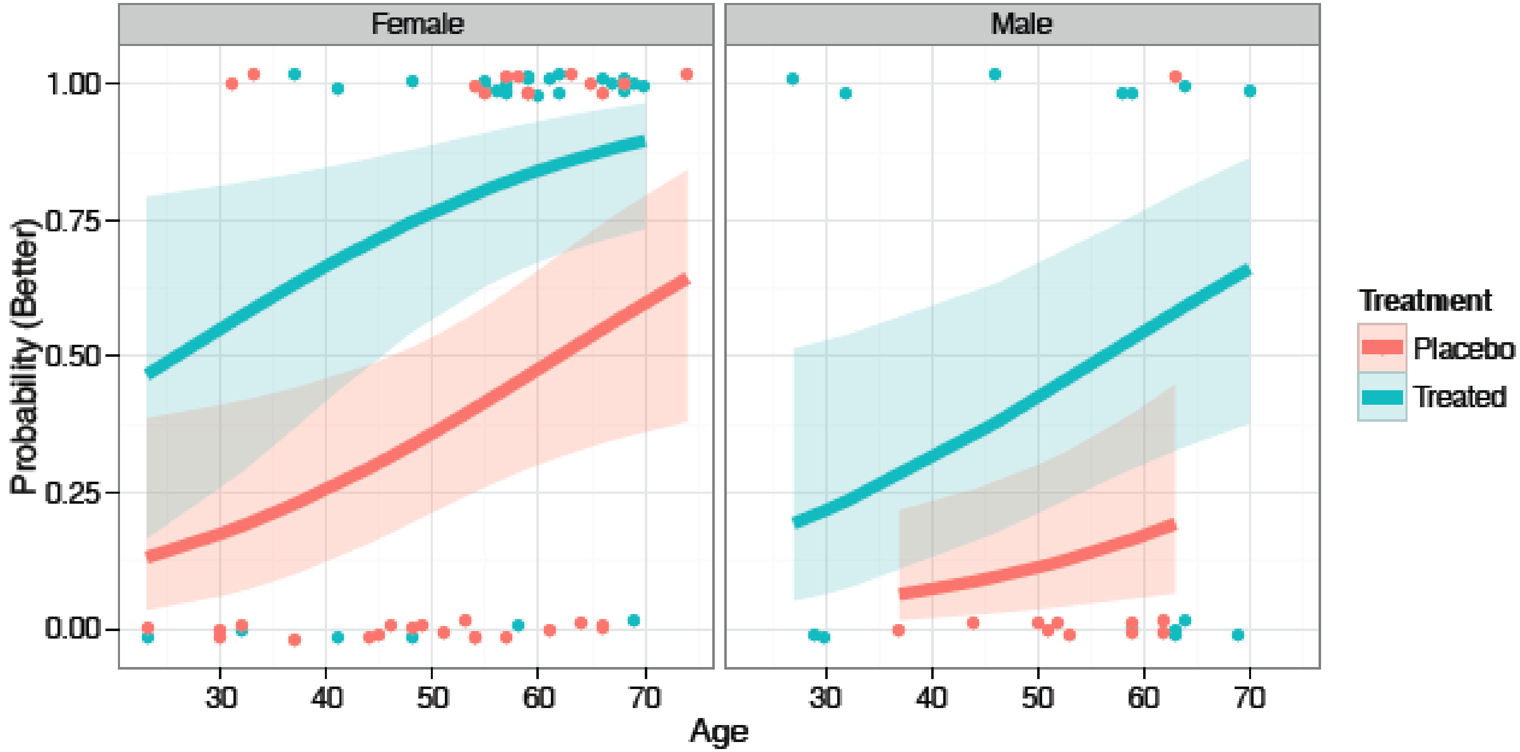


These plots show model uncertainty (confidence bands)
Jittered points show the data

Full-model plot

Plotting on the probability scale may be simpler to interpret
Use **predict(… type = "response")** to get fitted probabilities

```
arth.fit2r <- cbind(Arthritis,
                    predict(arth.logistic2, se.fit = TRUE, type="response"))
```



Models with interactions

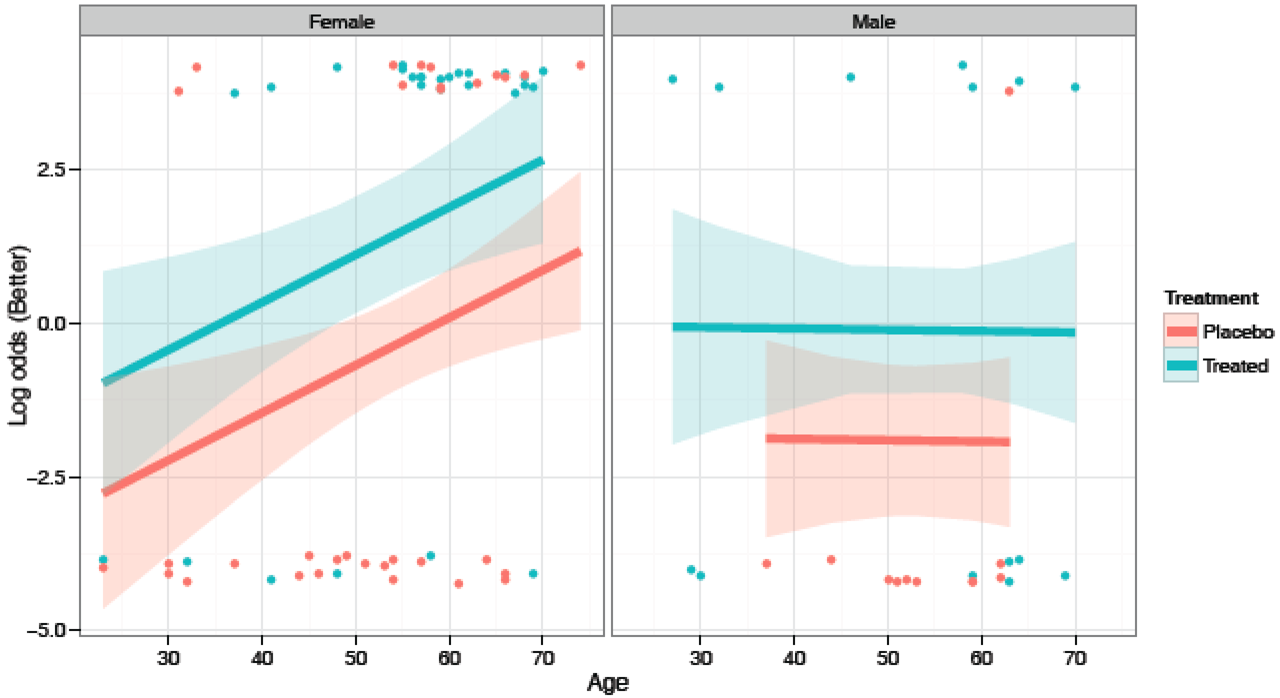
Is the linear effect of age the same for females, males?

- We can test this by adding an interaction of Sex x Age
- **`update()`** makes it easy to add/subtract terms from a model
- **`car::Anova()`** gives partial tests of each term after all others

```
> arth.logistic4 <- update(arth.logistic2, . ~ . + I(Age-50):Sex)
> car::Anova(arth.logistic4)
Analysis of Deviance Table (Type II tests)

Response: Better
                LR Chisq Df Pr(>Chisq)
I(Age - 50)         6.16  1    0.01308 *
Sex                 6.98  1    0.00823 **
Treatment          11.90  1    0.00056 ***
I(Age - 50):Sex     3.42  1    0.06430 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term Age:Sex is not quite significant, but plot the fitted model anyway

Models with interactions



- Only the model changes
- `predict()` automatically incorporates the revised model terms
- Plotting steps remain the same
- This interpretation is quite different!

Effect plots: Basic ideas

Show a given marginal effect, controlling / adjusting for other model effects

Data

	x1	x2	sex	x1x2	y	yhat
1	1	1	F	1	4.73	4.46
2	2	1	M	0	6.10	5.55
3	3	1	F	-1	4.32	4.34
4	1	1	F	1	4.84	4.46
5	2	1	F	0	4.73	4.40
...	...	...	...	...	...	...
29	2	2	M	0	6.10	6.15
30	3	2	F	1	6.71	7.14

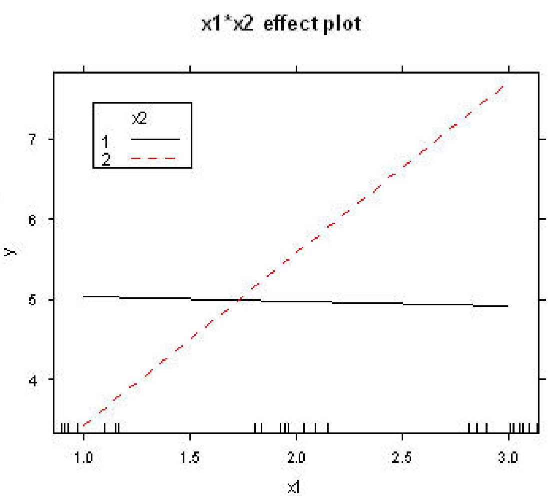
- Fit data: $\mathbf{X}\hat{\boldsymbol{\beta}} \Rightarrow \hat{\mathbf{y}}$
- Score data $\mathbf{X}^*\hat{\boldsymbol{\beta}} \Rightarrow \hat{\mathbf{y}}^*$
  - plot vars: vary over range
  - control vars: fix at means

Score data

	x1	x2	sex	x1:x2	y	yhat*
31	1	1	0.5	1	NA	5.030
32	2	1	0.5	2	NA	4.971
33	3	1	0.5	3	NA	4.912
34	1	2	0.5	2	NA	3.437
35	2	2	0.5	4	NA	5.574
36	3	2	0.5	6	NA	7.710

plot vars (x1, x2) control vars (sex)

plot



Effect plots: Details

- For simple models, full model plots show the complete relation between response and *all predictors*.
- Fox(1987)— For complex models, often wish to plot a specific main effect or interaction (including lower-order relatives)— *controlling for other effects*
  - Fit full model to data with linear predictor (e.g., logit) $\eta = \boldsymbol{X}\beta$ and link function $g(\mu) = \eta \rightarrow$ estimate $\boldsymbol{b}$ of β and covariance matrix $\widehat{V(\boldsymbol{b})}$ of $\boldsymbol{b}$.
  - Construct "score data"
    - Vary each predictor in the term over its' range
    - Fix other predictors at "typical" values (mean, median, proportion in the data)
    - $\rightarrow$ "effect model matrix," $\boldsymbol{X}^*$
  - Use `predict()` on $\boldsymbol{X}^*$
    - Calculate fitted effect values, $\hat{\eta}^* = \boldsymbol{X}^* \boldsymbol{b}$.
    - Standard errors are square roots of $\operatorname{diag} \boldsymbol{X}^* \widehat{V(\boldsymbol{b})} \boldsymbol{X}^{*\mathsf{T}}$
  - Plot $\hat{\eta}^*$, or values transformed back to scale of response, $g^{-1}(\hat{\eta}^*)$.
- *Note*: This provides a general means to visualize interactions in *all* linear and generalized linear models.

Plotting main effects

allEffects() calculates effects for all high-order terms in the model
The response is plotted on the logit scale, but labeled with probabilities

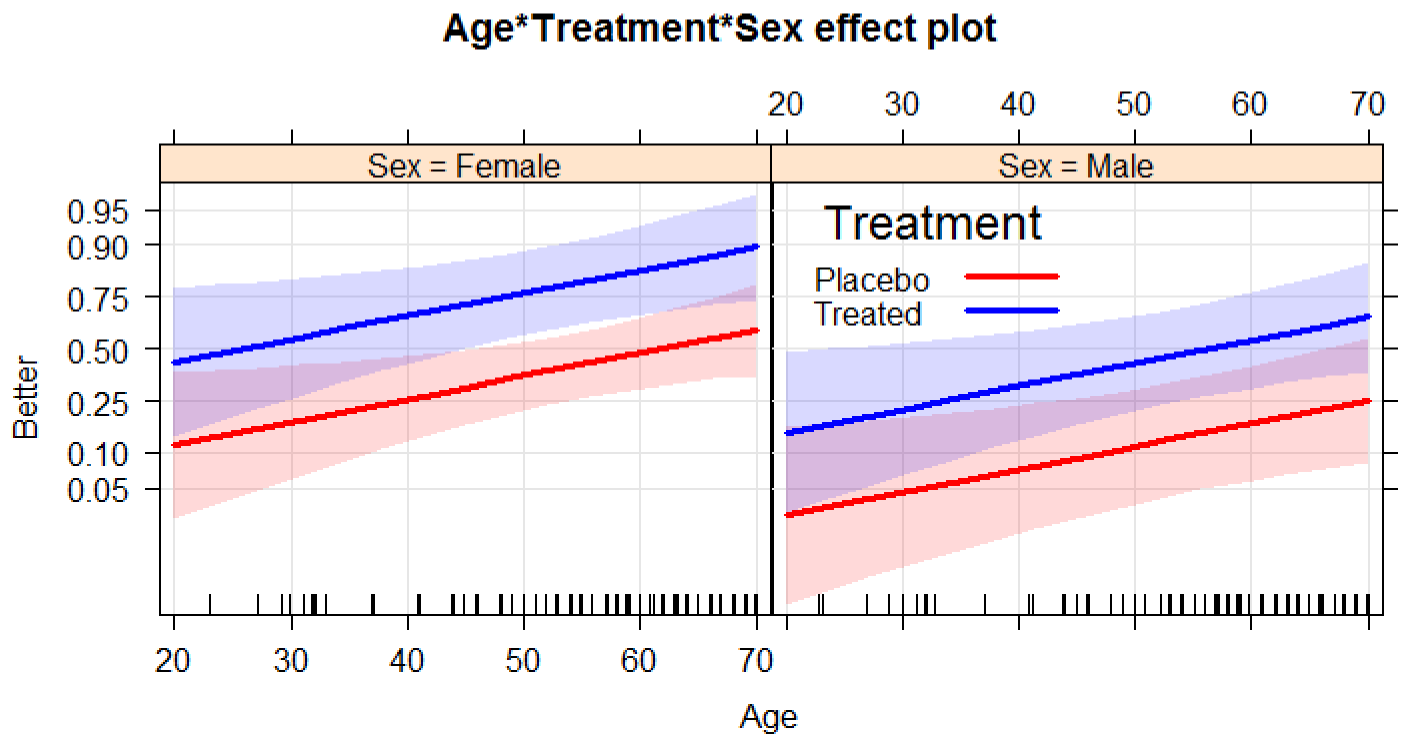
```
library(effects)
arth.eff2 <- allEffects(arth.logistic2)
plot(arth.eff2, rows=1, cols=3, lwd=2)
```

Averaged over:	Sex Treatment	Age Treatment	Age Sex

Full-model plot

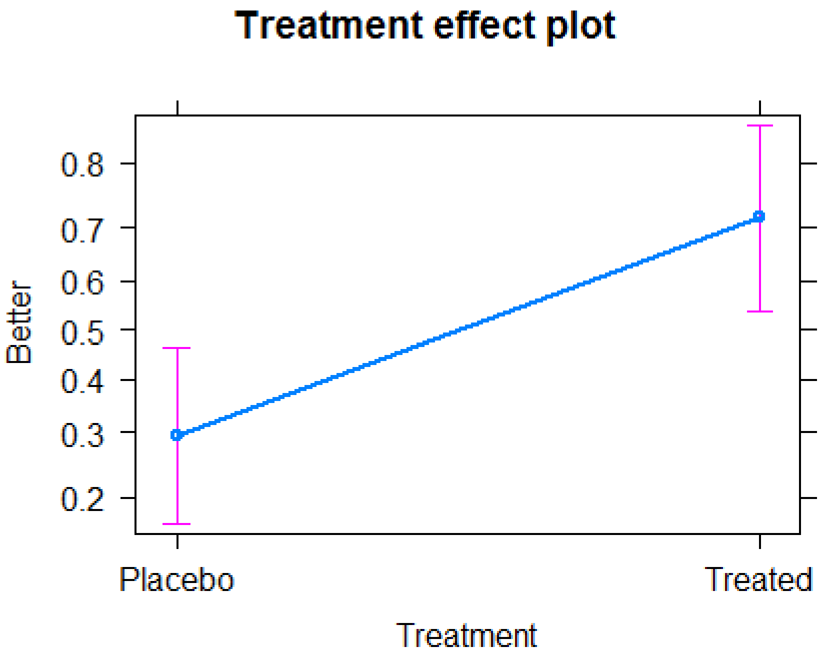
The full-model plot is simply the **`Effect()`** of the highest-order interaction of factors

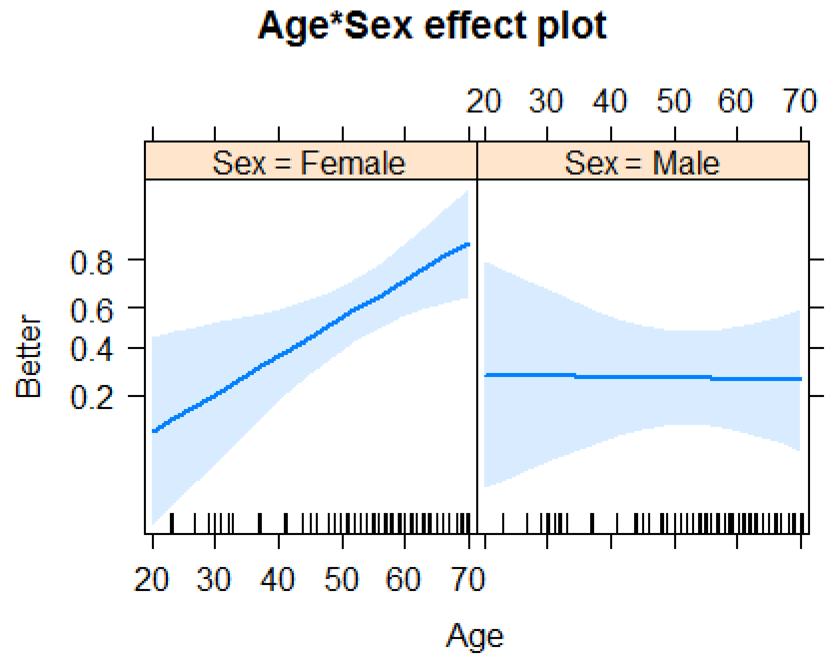
```
arth.full <- Effect(c("Age", "Treatment", "Sex"), arth.logistic2)
plot(arth.full, multiline=TRUE, ci.style="bands",
     colors = c("red", "blue"), lwd=3,   . . .)
```



Model with interaction of Age × Sex

```
arth.eff4 <- allEffects(arth.logistic4)
plot(arth.eff4, lwd=2)
```





Only the high-order terms: Treatment & Age * Sex are shown & need to be interpreted
Q: How would you describe this?

Race & Crime

Toronto Star investigation of racial disparities in treatment by Toronto Police Services

FOI request → > ½ M arrests, 1997—2002

Evidence for racial profiling?

Only look at discretionary charges:

Simple marijuana possession
Non-moving auto infractions

THE SATURDAY STAR

The photo that never was
GARTH WOOLSEY, C3

Also inside . . .

- **Waterfront:** Dreams of what could be, B1, B4-5
- **Hydro woes:** Insulating against price spikes, E1
- **Wheels:** The Bug goes roofless, G1
- **Paul Martin:** The man who would be king, H1
- **Carol Shields:** Her last novel? Unless . . . J1

Periods of rain; windy. High 14 C — October 19, 2002 — thestar.com ONTARIO EDITION

AN INVESTIGATION INTO RACE AND CRIME

JIM RANKIN/TORONTO STAR

SUING POLICE: Jason Burke, falsely accused of dealing drugs during Caribana two years ago, says he was a victim of racial profiling.

Singled out

Star analysis of police crime data shows justice is different for blacks and whites

Telling numbers

Police records show that a black person in Toronto arrested on a single drug possession charge was **less likely** to be released at the scene...

% of each racial group

... and **twice as likely** to be held for a bail hearing, compared to a white person on the same charge.

	White	Black
Released at scene	76.5%	61.8%
Held for bail	7.3%	15.5%

Blacks arrested by Toronto police are treated more harshly than whites, a Toronto Star analysis of crime data shows.

Black people, charged with simple drug possession, are taken to police stations more often than whites facing the same charge.

Once at the station, accused blacks are held overnight, for a bail hearing, at twice the rate of whites.

The Toronto crime data also shows a disproportionate number of black motorists are ticketed for violations that only surface following a traffic stop. This difference, say civil libertarians, community leaders and criminologists, suggests police use racial profiling in deciding whom to pull over.

■ Managing Editor's notebook, A2

The evidence is contained in a massive police database recording more than 480,000 incidents in which an individual was arrested, or ticketed, for an offence dating back to 1996. It included almost 800,000 criminal and other charges. The Star obtained that data through a freedom of information request, marking the first time access to these numbers was granted to anyone outside the police community.

Police are forbidden, by their governing board, from analyzing this data in terms of race, but The Star has no such restriction. The findings provide hard evidence of what blacks have long suspected — race matters in Canadian society especially when dealing with police.

Chief Julian Fantino disputed the findings, saying the colour of a person's skin has nothing to do with how they're treated by his officers.

"We don't treat people different-

Please see Toronto, A12

Chrétien expected to keep cabinet minister

Ethics report has 'wiggle room' to save MacAulay

BY TIM HARPER AND LES WHITTINGTON
OTTAWA BUREAU

OTTAWA — Jean Chrétien receives a report from his ethics counsellor today that is expected to give him enough "wiggle room" to keep his solicitor-general, Lawrence MacAulay, in the federal cabinet.

Ethics counsellor Howard Wilson completed his report and delivered it to the Prime Minister's Office last night, where it was received by Chrétien's chief of staff, Percy Downe.

It was then to be relayed to Chrétien by secure fax to Beirut, where the Prime Minister is attending a summit of French-speaking nations. It was 1:30 a.m. in Beirut when the fax arrived so Chrétien would likely be reading it this morning.

Senior sources said last night that unless there is a surprise in Wilson's report, the Prince Edward Island minister will remain, Chrétien will return to Ottawa and weather the inevitable storm of opposition and media protest and forge ahead with an ethics package by midweek.

Wilson has been investigating whether MacAulay broke ethics guidelines for cabinet ministers in the awarding of a contract and extension worth $100,000 to Everett Roche, a Charlottetown political friend of the solicitor-general's.

Chrétien will not fire MacAulay unless he is given incontrovertible evidence of wrongdoing for two key reasons, sourc-

Please see MacAulay, A8

INSIDE

Case study: Arrests for marijuana

- In Dec. 2002, the *Toronto Star* examined the issue of racial profiling, by analyzing a data base of 600,000+ arrest records from 1997-2002.
- They focused on a subset of arrests for which police action was discretionary, e.g., simple possession of small quantities of marijuana, where the police could:
  - Release the arrestee with a summons – like a parking ticket
  - Bring to police station, hold for bail, … -- harsher treatment
- Response variable: released: “Yes”, “No”
  - Main predictor of interest: skin-colour of arrestee (black, white)
  - Other predictors: year, age, sex, …

Toronto Star meets mosaic displays

How to communicate these results most effectively?

- What is the message? What features are directly comprehensible to the audience?

B SECTION » TORONTO STAR « WEDNESDAY, DECEMBER 11, 2002 ★ thestar.com

Race and Crime

Graphic designer's early attempts

My early attempts

VINCE TALOTTA/TORONTO STAR

York University professor Michael Friendly's expert statistical analysis provided confirmation for the Toronto Star's series on racial profiling by city police.

Man behind the numbers

Racial profiling: Presentation graphic

Together, we created this (nearly) self-explaining infographic

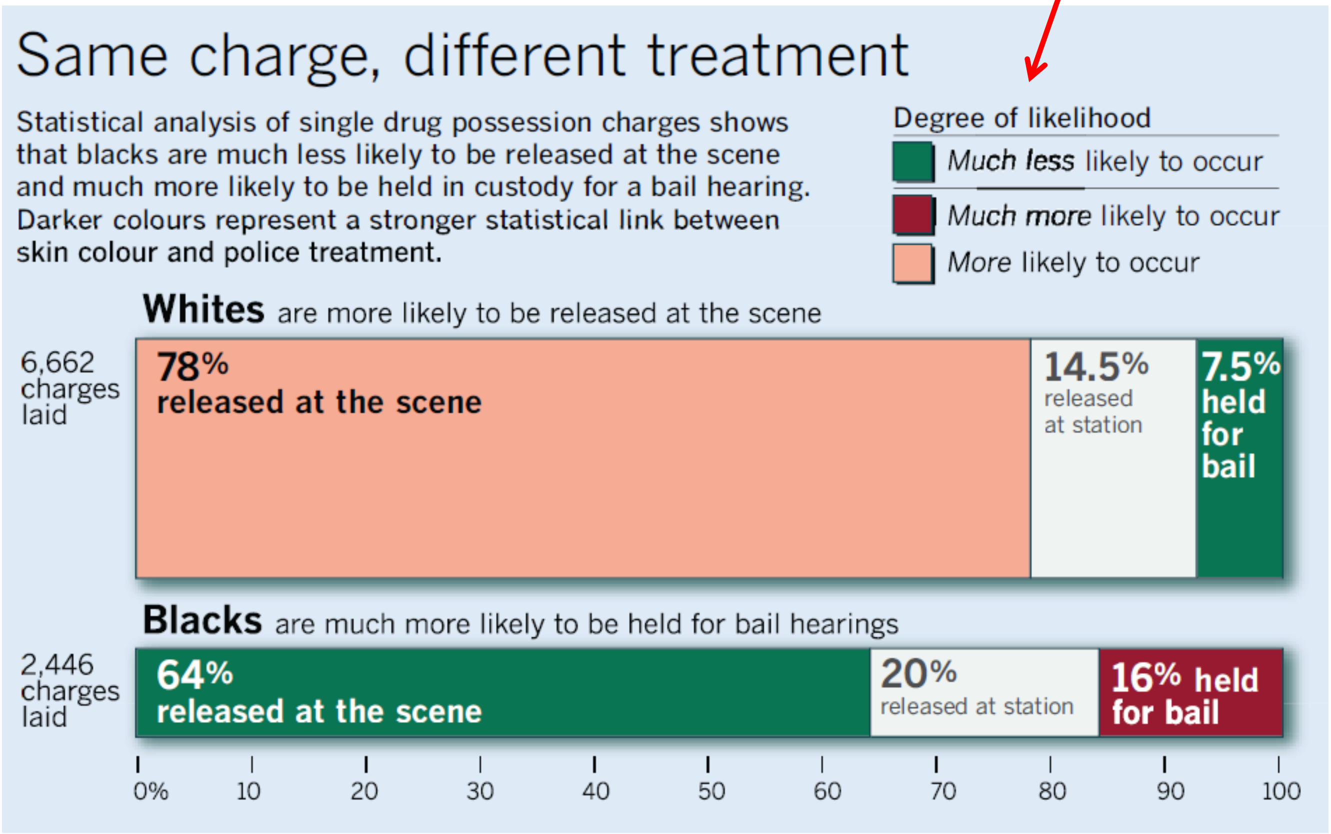
Title gives the main conclusion

Text description gives details

Bar width ~ charges
Divided by % release

numbers shown in the cells

Legend gives a layman's description of shading levels



SOURCE: Toronto police arrest records 1996-2002

Arrests for marijuana: Data

Response variable: released

Control variables:

- year, age, sex
- employed, citizen: Yes, No
- checks: # of police databases (previous arrests, convictions, parole status) where the arrestee's name was found

```
> library(car)         # for Anova()
> data(Arrests, package = "carData")
> some(Arrests)
     released colour year age  sex employed citizen checks
218       Yes  White 2000  24 Male      Yes     Yes      0
1301       No  Black 1999  17 Male      Yes      No      1
1495      Yes  White 1998  23 Male      Yes     Yes      0
1732      Yes  Black 2000  18 Male      Yes     Yes      2
1838      Yes  Black 1997  27 Male       No     Yes      5
2257       No  White 2001  19 Male       No     Yes      2
3100       No  Black 2000  19 Male       No     Yes      4
3843      Yes  White 1999  20 Male      Yes     Yes      0
4580      Yes  Black 1999  26 Male      Yes     Yes      1
4833      Yes  Black 1998  38 Male      Yes     Yes      0
```

Arrests for marijuana: Model

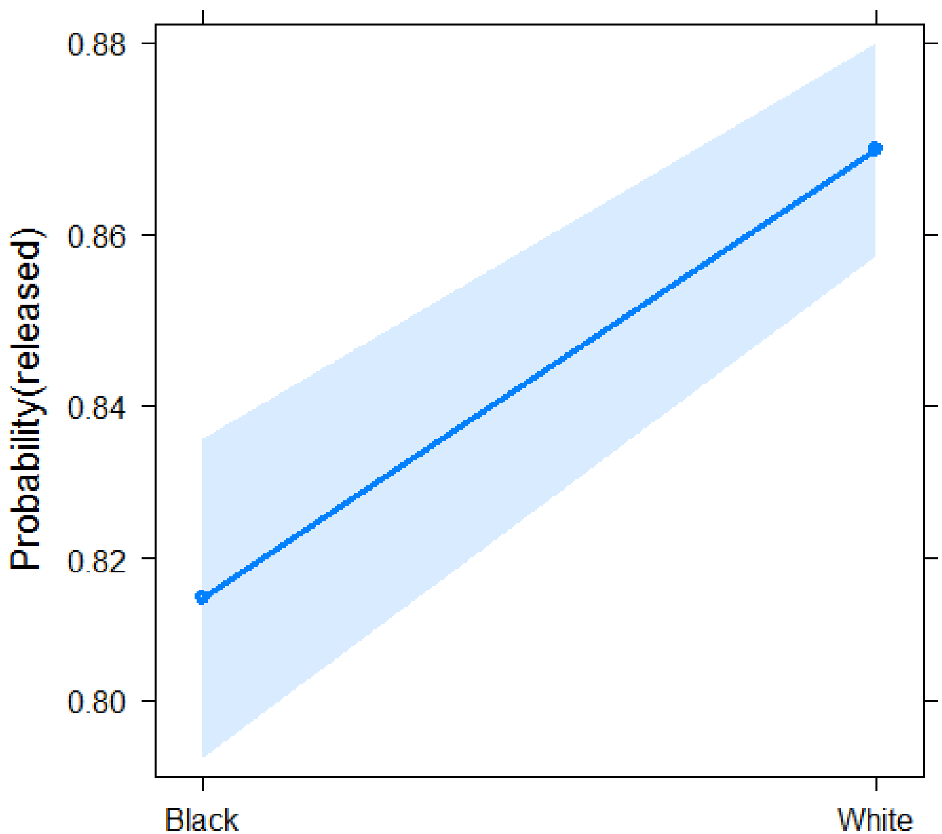
year is numerical. But may be non-linear. Convert to a factor
Fit model with all main effects, but allow interactions of colour:year and colour:age

```
> Arrests$year <- as.factor(Arrests$year)
> arrests.mod <- glm(released ~ employed + citizen + checks +
                     colour*year + colour*age,
                     family=binomial, data=Arrests)
> Anova(arrests.mod)
```

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

Response: released
            LR Chisq Df Pr(>Chisq)
employed        72.7  1    < 2e-16 ***
citizen         25.8  1    3.8e-07 ***
checks         205.2  1    < 2e-16 ***
colour          19.6  1    9.7e-06 ***
year             6.1  5    0.29785
age              0.5  1    0.49827
colour:year     21.7  5    0.00059 ***
colour:age      13.9  1    0.00019 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect plot: Skin colour

```
plot(Effect("colour", arrests.mod), lwd=3, ci.style="bands”, ...)
```



- Effect plot for colour shows average effect controlling (adjusting) for all other factors simultaneously
- (The *Star* analysis controlled for these one at a time.)

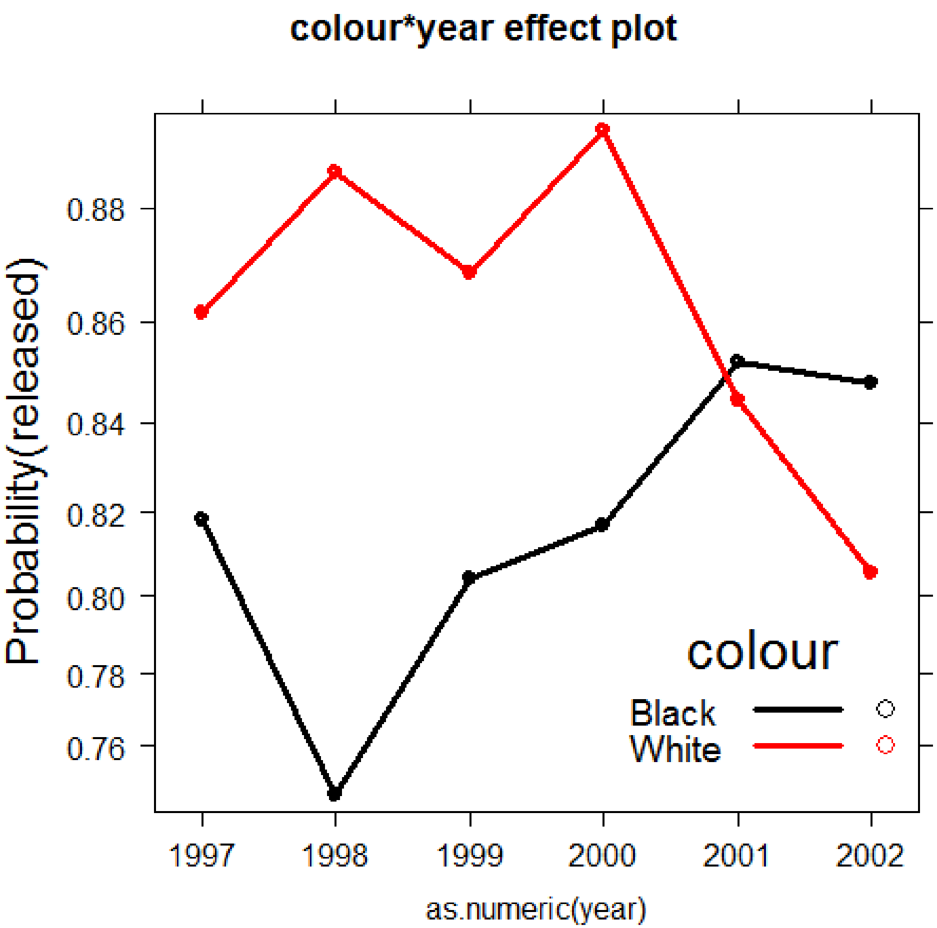
→ Evidence for different treatment of blacks & whites

- Even Francis Nunziata could understand this.
- However, effect smaller than reported by the *Star*

Effect plots: Interactions

The story turned out to be more nuanced than reported by the *Toronto Star*

```
plot(Effect(c("colour","year"), arrests.mod), multiline=TRUE, ...)
```



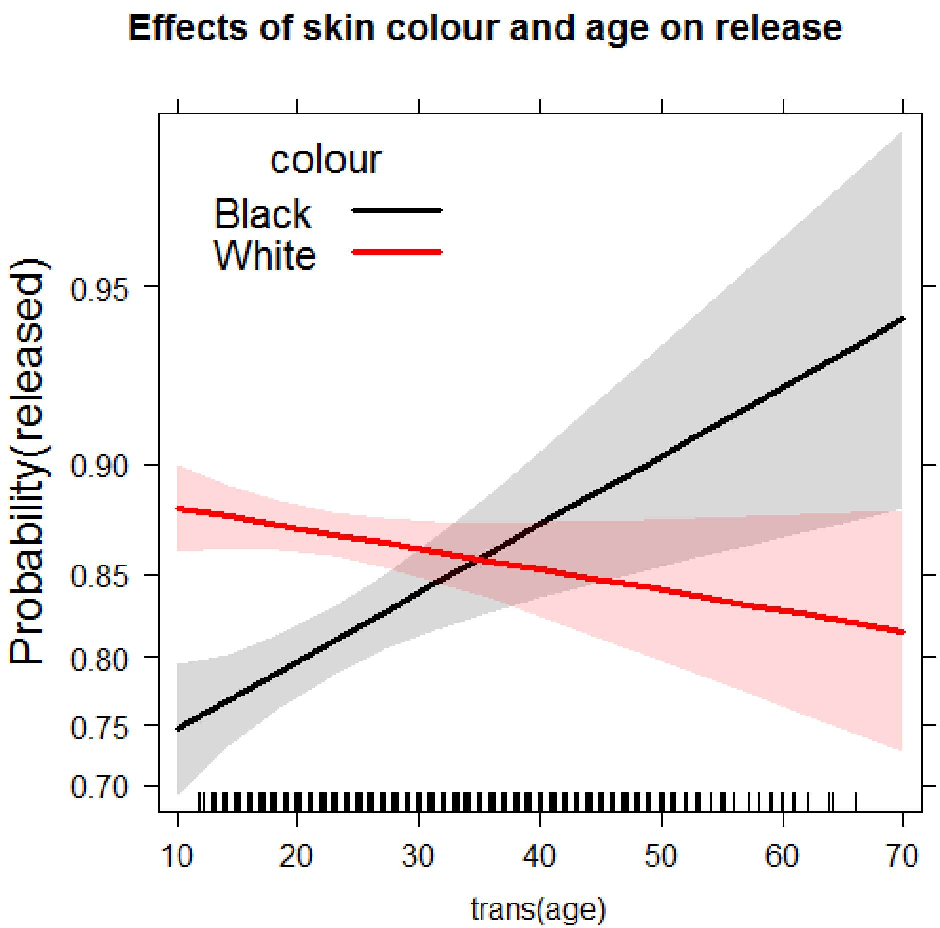
Up to 2000, strong evidence for differential treatment of blacks & whites

Also: evidence to support Police claim of effect of training to reduce racial effects in treatment

Effect plots: Interactions

A more surprising finding …

```
plot(Effect(c("colour","year"), arrests.mod), multiline=TRUE, ...)
```



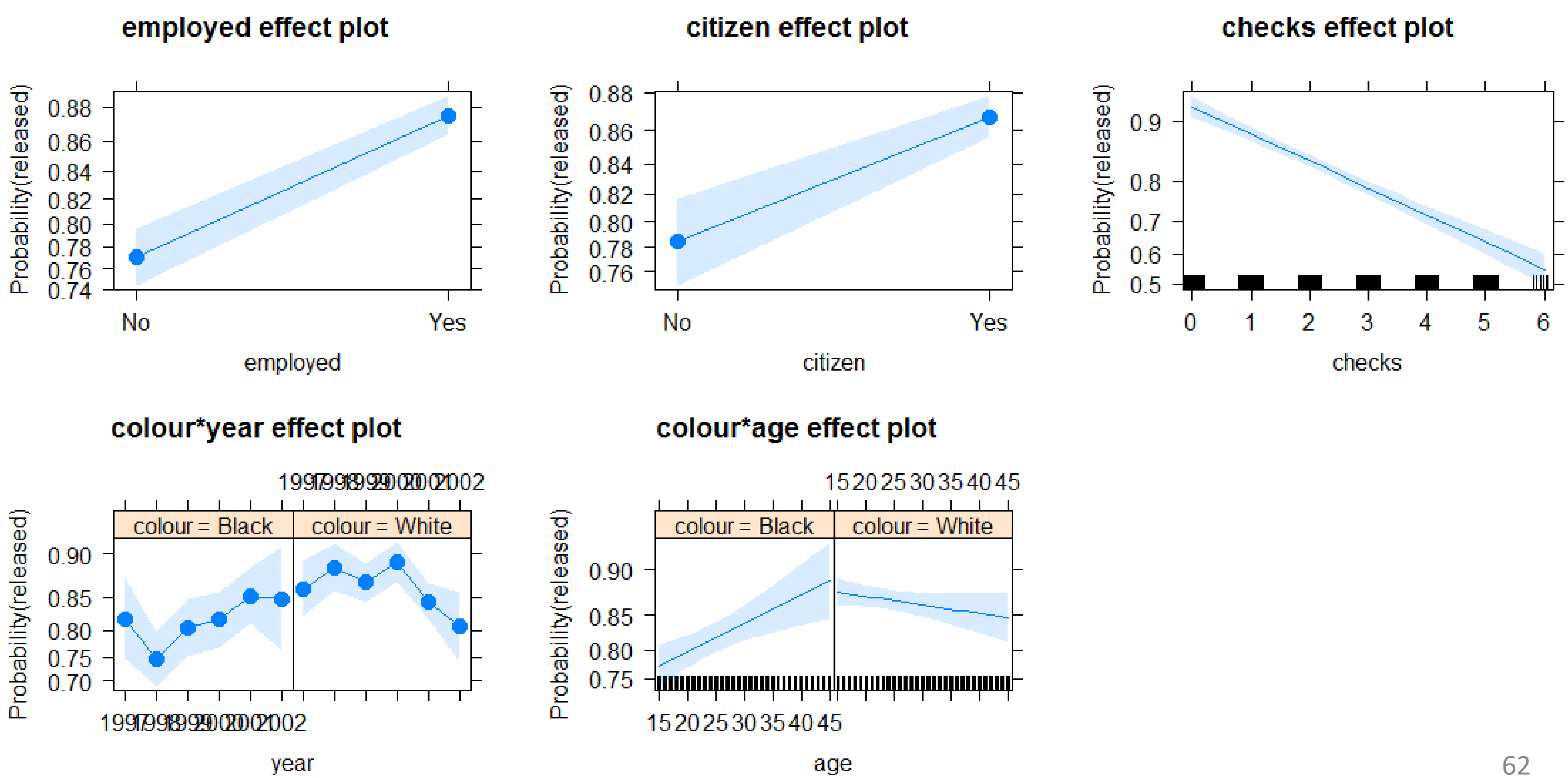
Opposite age effects for blacks & whites:

- Young blacks treated more harshly than young whites
- Older blacks treated less harshly than older whites

Effect plots: allEffects

All high-order terms can be viewed together using plot(allEffects(mod))

```
arrests.effects <- allEffects(arrests.mod,
xlevels=list(age=seq(15,45,5)))
plot(arrests.effects, ylab="Probability(released)", …)
```



Model diagnostics

As in regression and ANOVA, the validity of a logistic regression model is threatened when:

- Important predictors have been omitted from the model
- Predictors assumed to be linear have non-linear effects on $\Pr(Y = 1)$
- Important interactions have been omitted
- A few "wild" observations have a large impact on the fitted model or coefficients

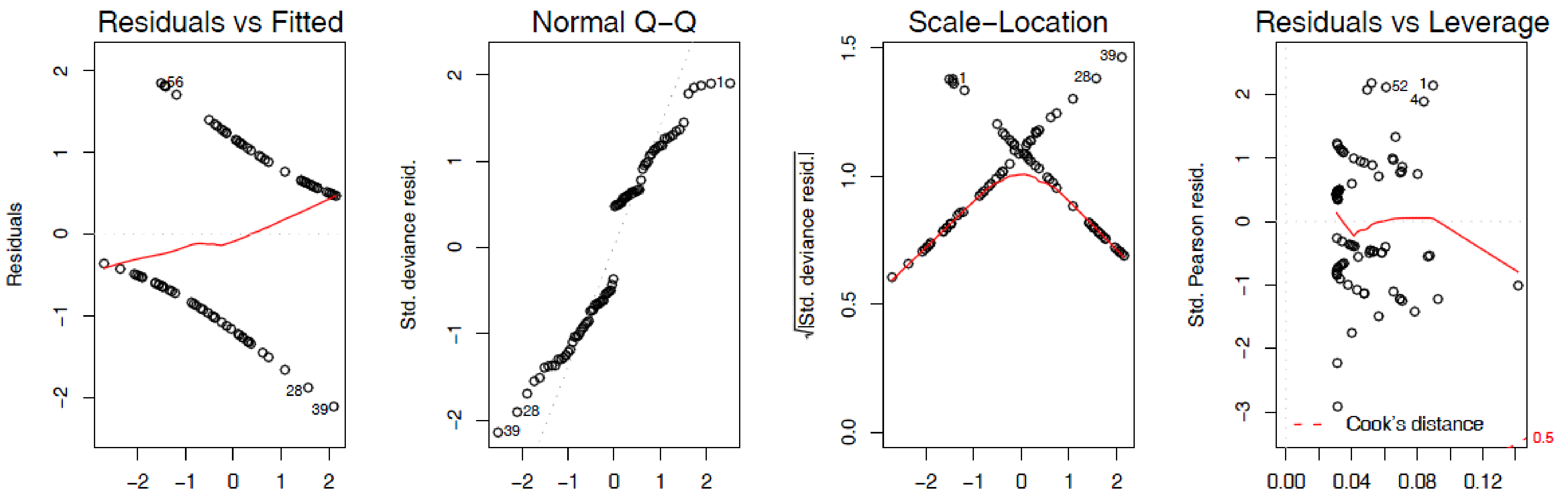
Model specification: Tools and techniques

- Use non-parametric smoothed curves to detect non-linearity
- Consider using polynomial terms ($X^2, X^3, \ldots$) or regression splines (e.g., `ns(X, 3)`)
- Use `update(model, ...)` to test for interactions— formula: $. \sim .^2$

Diagnostic plots in R

In R, plotting a `glm` object gives the "regression quartet" – 4 basic diagnostic plots

```
arth.mod1 <- glm(Better ~ Age + Sex + Treatment, data=Arthritis,
                 family='binomial')
plot(arth.mod1)
```

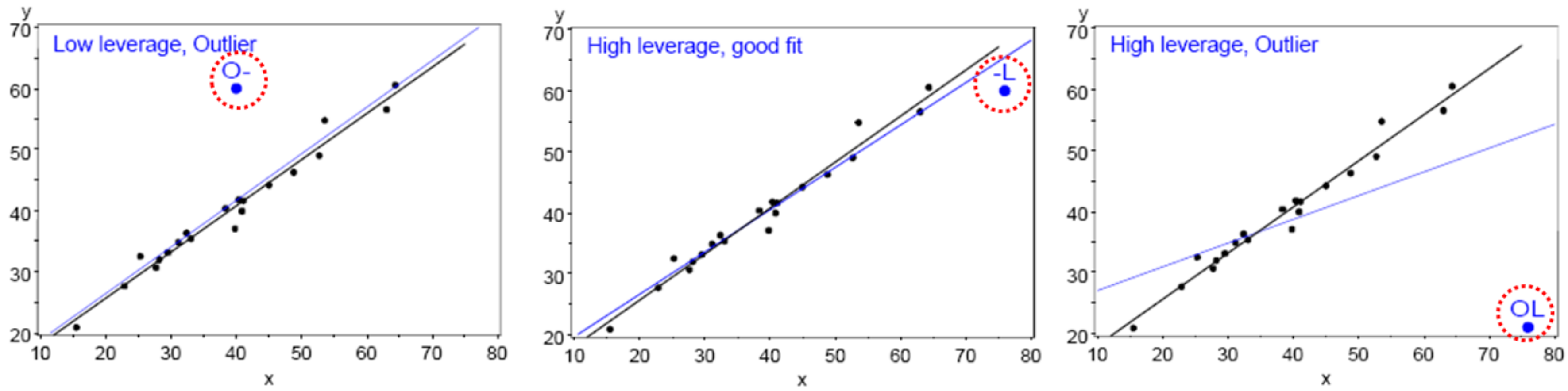


These plots often look peculiar for logistic regression models
Better versions are available in the car package

Unusual data: Leverage & Influence

- "Unusual" observations can have dramatic effects on least-squares estimates in linear models
- Three archetypal cases:
  - Typical X (low leverage), bad fit -- Not much harm
  - Unusual X (high leverage), good fit -- Not much harm
  - Unusual X (high leverage), bad fit -- BAD, BAD, BAD
- Influential observations: unusual in *both* X & Y
- Heuristic formula:

Influence = X leverage × Y residual



Influence plots

Influence (Cook's D) measures impact of individual obs. on coefficients, fitted values

Influence ~ Residual (y -ŷ) × Hat-value $(X - \bar{X})^2$

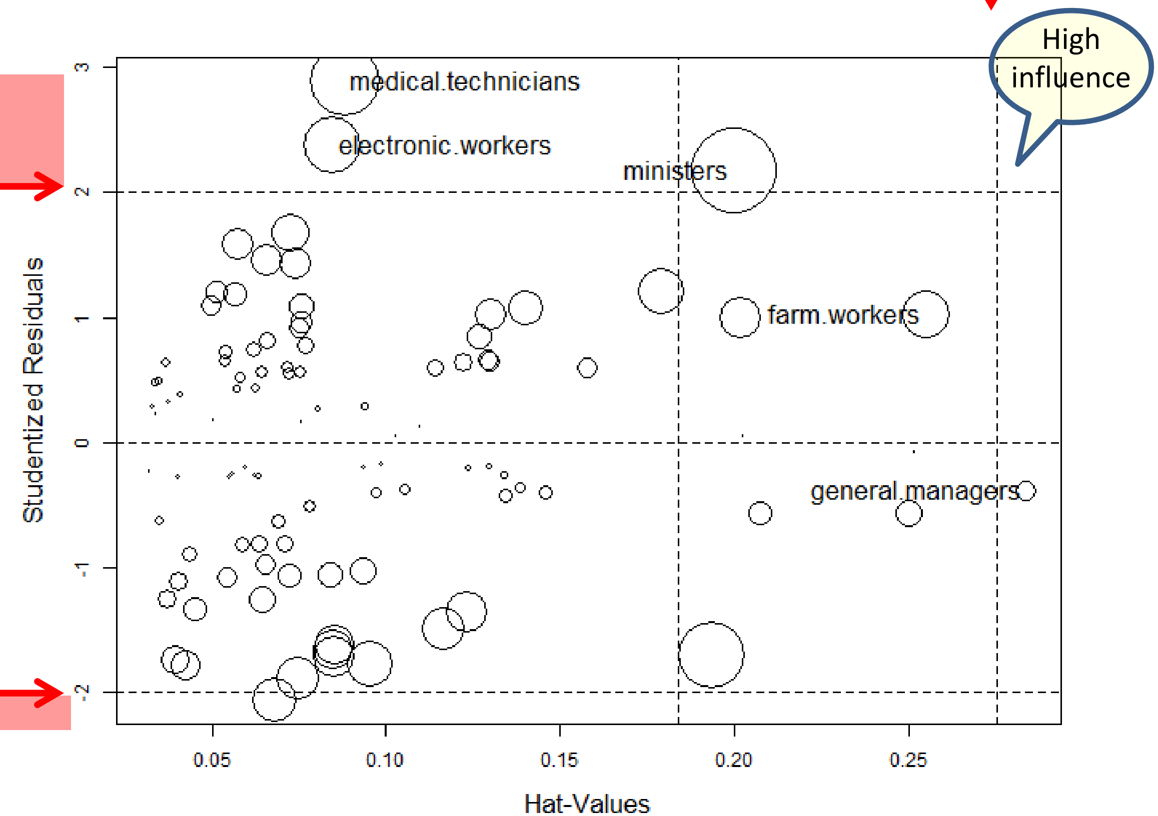
High leverage

Bubble size ~ influence

Bad fit

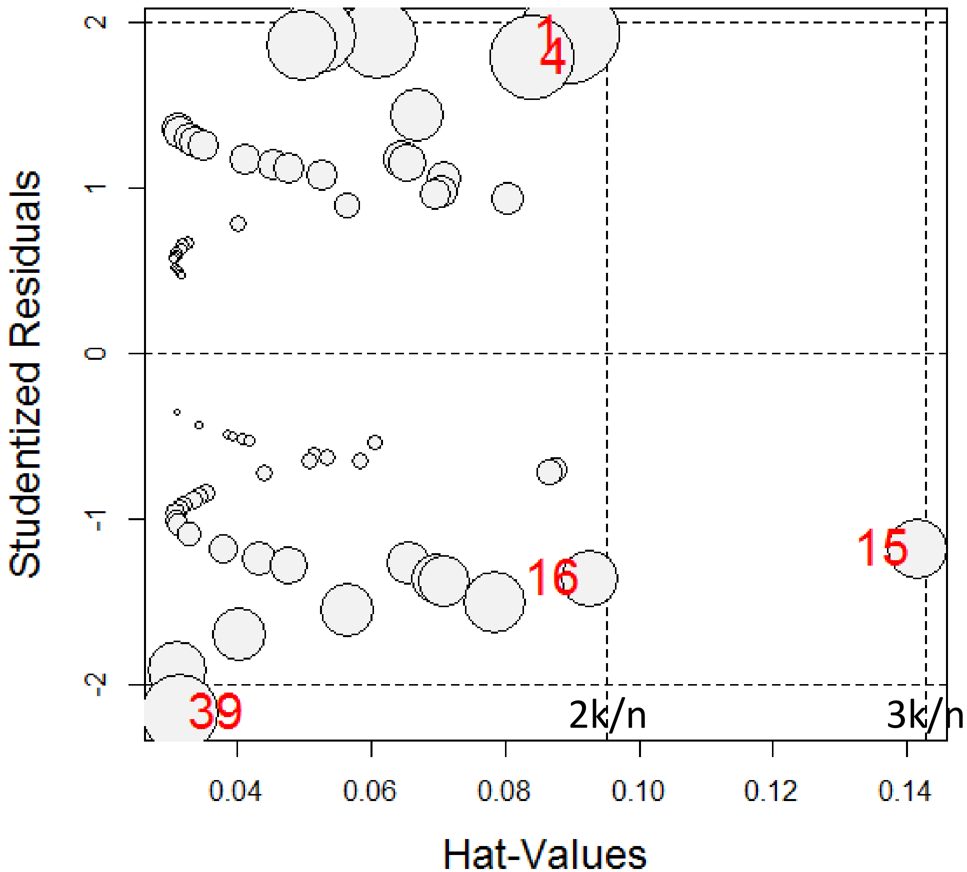
```
influencePlot(mod1)
```

Bad fit



Influence plots in R

```
library(car)
influencePlot(arth.logistic2, …)
```



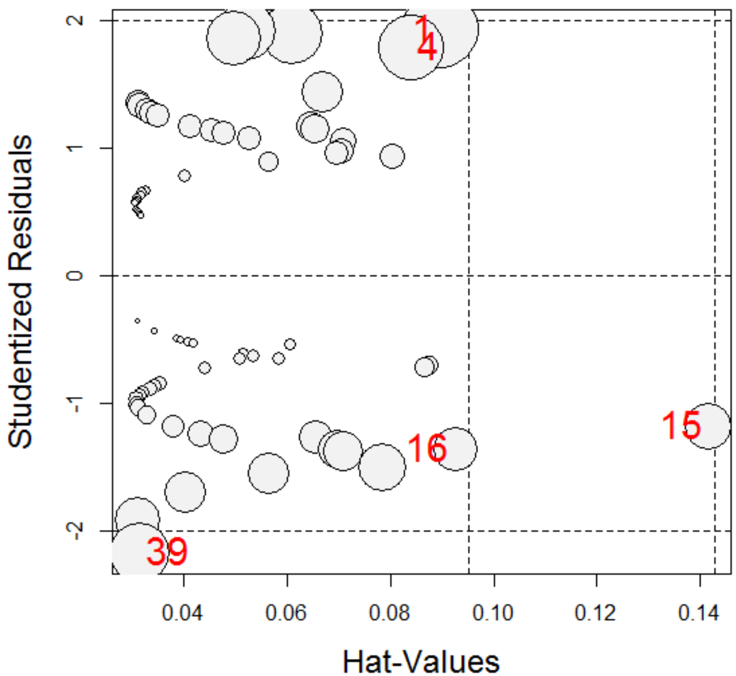
X axis: Leverage ("hat values")
notable values: > 2k/n, 3k/n

Y axis: Studentized residuals

Bubble size ~ Cook's D
(influence on coefficients)

Which cases are influential?

	Treatment	Sex	Age	Better	StudRes	Hat	CookD
1	Treated	Male	27	1	1.92	0.0897	0.1128
4	Treated	Male	32	1	1.79	0.0840	0.0818
15	Treated	Female	23	0	-1.18	0.1416	0.0420
16	Treated	Female	32	0	-1.36	0.0926	0.0381
39	Treated	Female	69	0	-2.17	0.0314	0.0690



case 1: younger male: moderate Hat, better than predicted → large Cook D

case 39: older female: small Hat, but did not improve with treatment

Looking ahead

- Logistic regression models need not always have linear effects– models nonlinear in Xs sometimes useful
- Polytomous outcomes can be handled as well
  - e.g., Improved = {“None”, “Some”, “Marked”}
- If ordinal,
  - the proportional odds model is a simple extension
  - nested dichotomies provides an alternative approach
- Otherwise, multinomial logistic regression is the way

Summary

- Model-based methods provide hypothesis tests, CIs & tests for individual terms
- Logistic regression: A glm() for a binary response
  - linear model for the log odds Pr(Y=1)
  - All similar to classical ANOVA, regression models