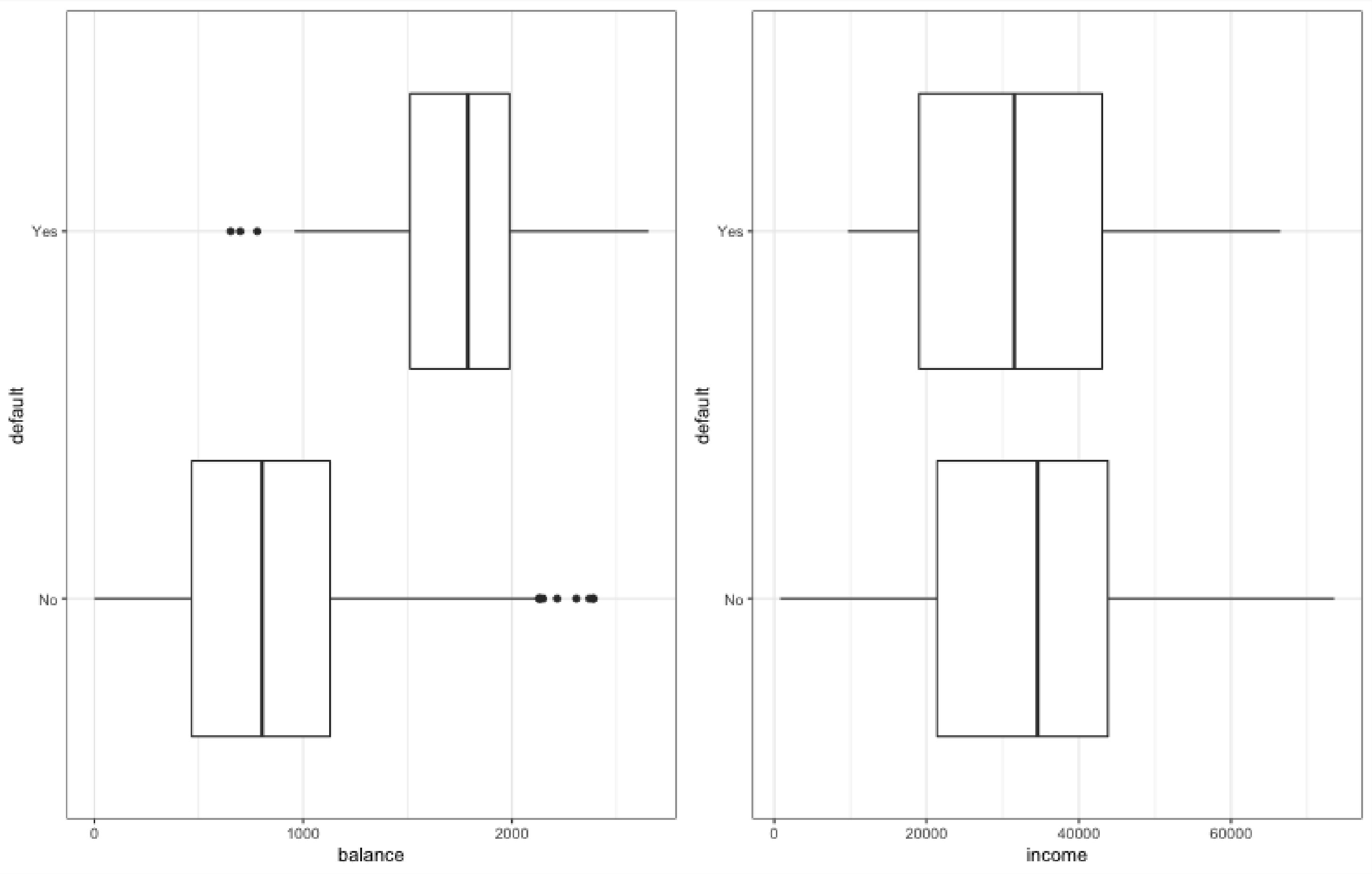
Logistic Regression

Example: Credit Default

```r
library(ISLR)
data(Default)
head(Default)
```

```
##   default student   balance    income
## 1      No      No  729.5265 44361.625
## 2      No     Yes  817.1804 12106.135
## 3      No      No 1073.5492 31767.139
## 4      No      No  529.2506 35704.494
## 5      No      No  785.6559 38463.496
## 6      No     Yes  919.5885  7491.559
```

Exploratory Data Analysis



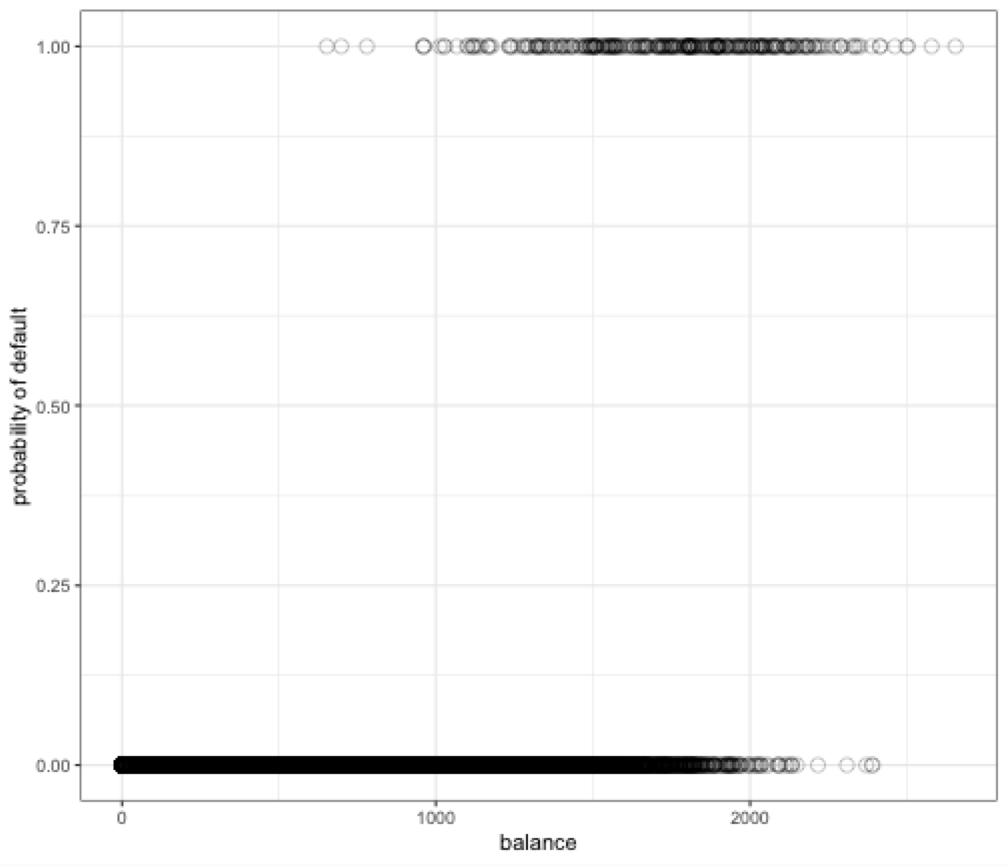
Model Fitting

```r
m1 <- glm(default ~ balance,
          data = Default,
          family = binomial)
coef(m1)
```

```
##   (Intercept)       balance
## -10.651330614   0.005498917
```

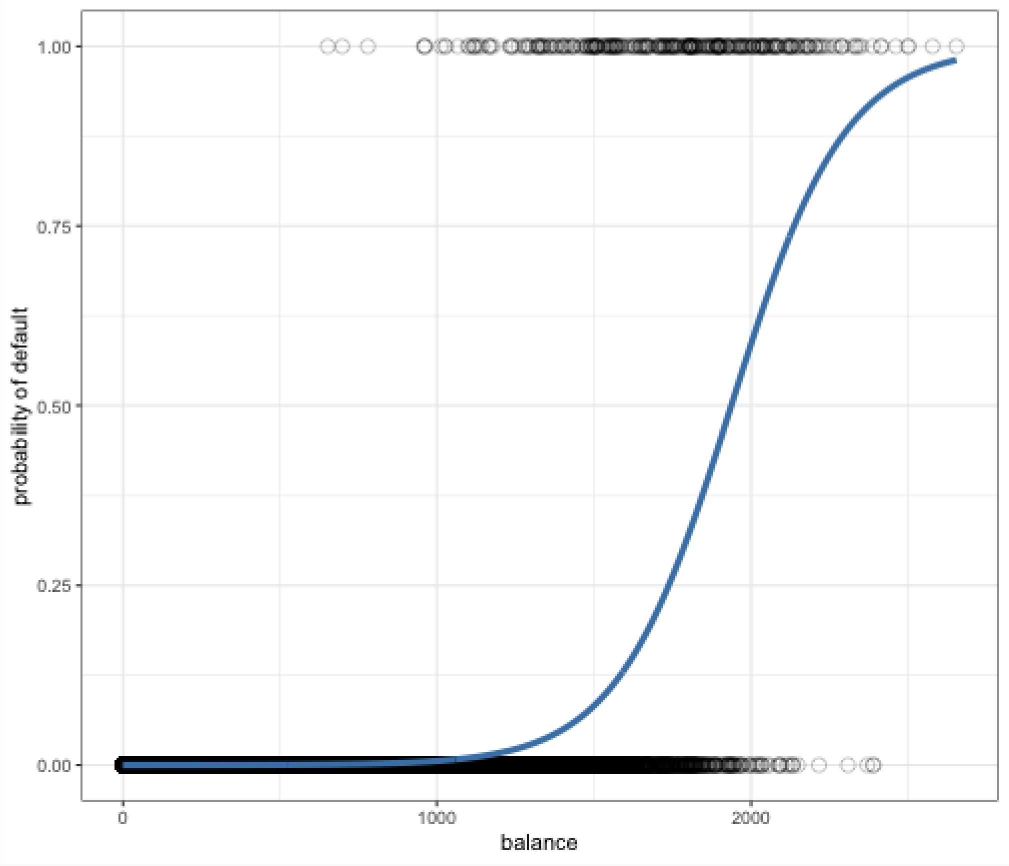
Logistic Model

$$P(Y=1|X=x_i) = \frac{1}{1+e^{-(-10.65+0.0055x_i)}}$$



Logistic Model

$$P(Y = 1|X = x_i) = \frac{1}{1 + e^{-(-10.65 + 0.0055 x_i)}}$$



Logistic Model Coefficients

```
summary(m1)$coef
```

```
##                 Estimate    Std. Error   z value      Pr(>|z|)
## (Intercept) -10.651330614 0.3611573721 -29.49221 3.623124e-191
## balance       0.005498917 0.0002203702  24.95309 1.976602e-137
```

Where did those SEs come from?

The Likelihood Function

48 male bank supervisors were asked to assume the role of the personnel director of a bank and were given a personnel file to judge whether the person should be promoted to a branch manager position. The files given to the participants were identical, except that half of them indicated the candidate was male and the other half indicated the candidate was female. These files were randomly assigned to the supervisiors. For each supervisor we recorded the gender associated with the assigned file and the promotion decision.

	promoted	**not promoted**
male	18	6
female	14	10

Is this data consistent with the claim that females are unfairly discriminated against in promotion decisions? What statistical method would you use to make that determination?

A model for promotion

	promoted	not promoted	p(promoted)
male	18	6	18/24 = .75
female	14	10	14/24 = .58

1. Each decision was independent.
2. All males were promoted with the same probability p_M.
3. All females were promoted with the same probability p_F.

$$Y \sim \text{binomial}(n = 24, p = p_M)$$
$$X \sim \text{binomial}(n = 24, p = p_F)$$

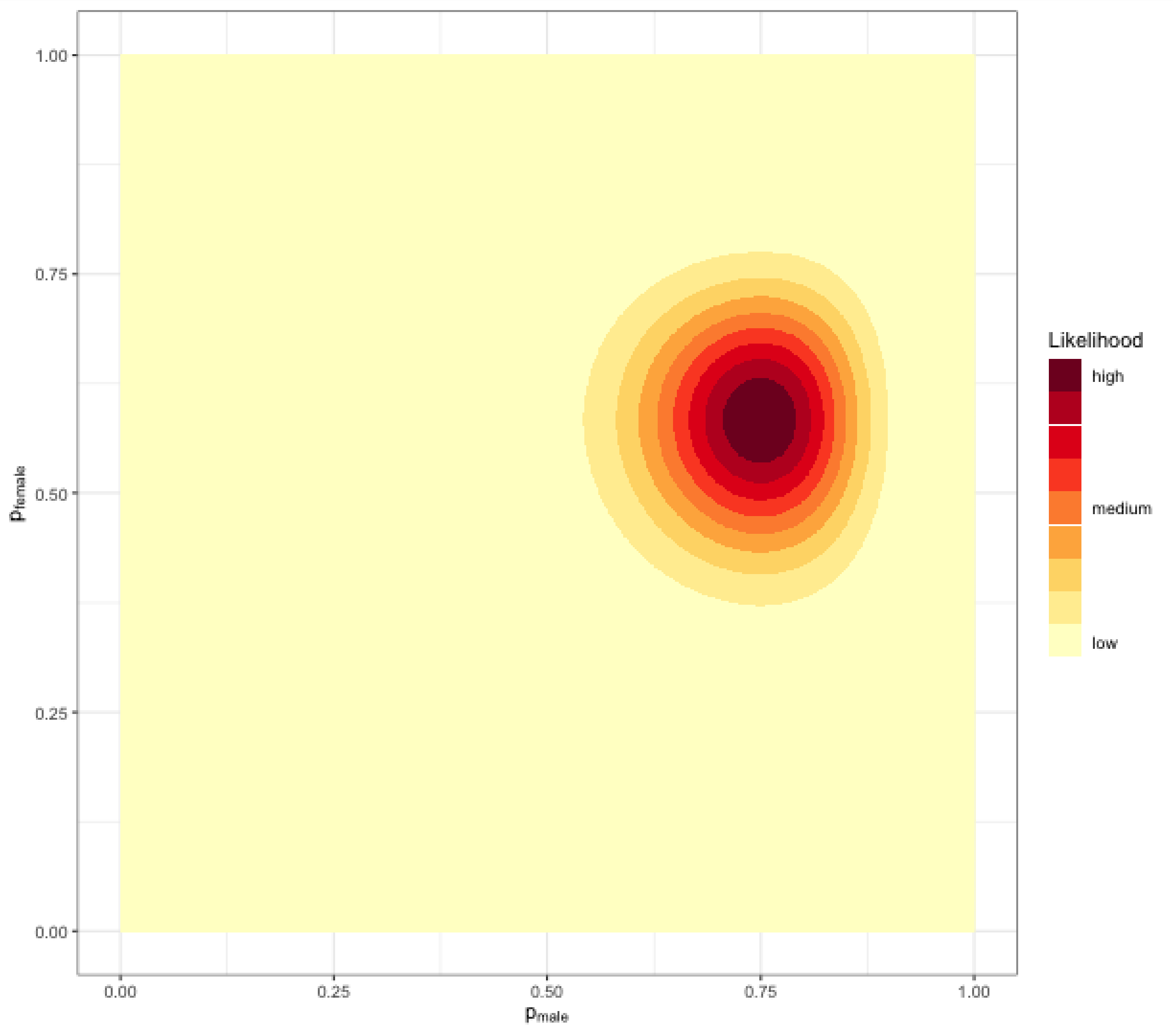
From Probability to Likelihood

$$P(y, x|n, p_M, p_F) = \binom{n}{y} p_M^y (1 - p_M)^{n-y} \binom{n}{x} p_F^x (1 - p_F)^{n-x}$$

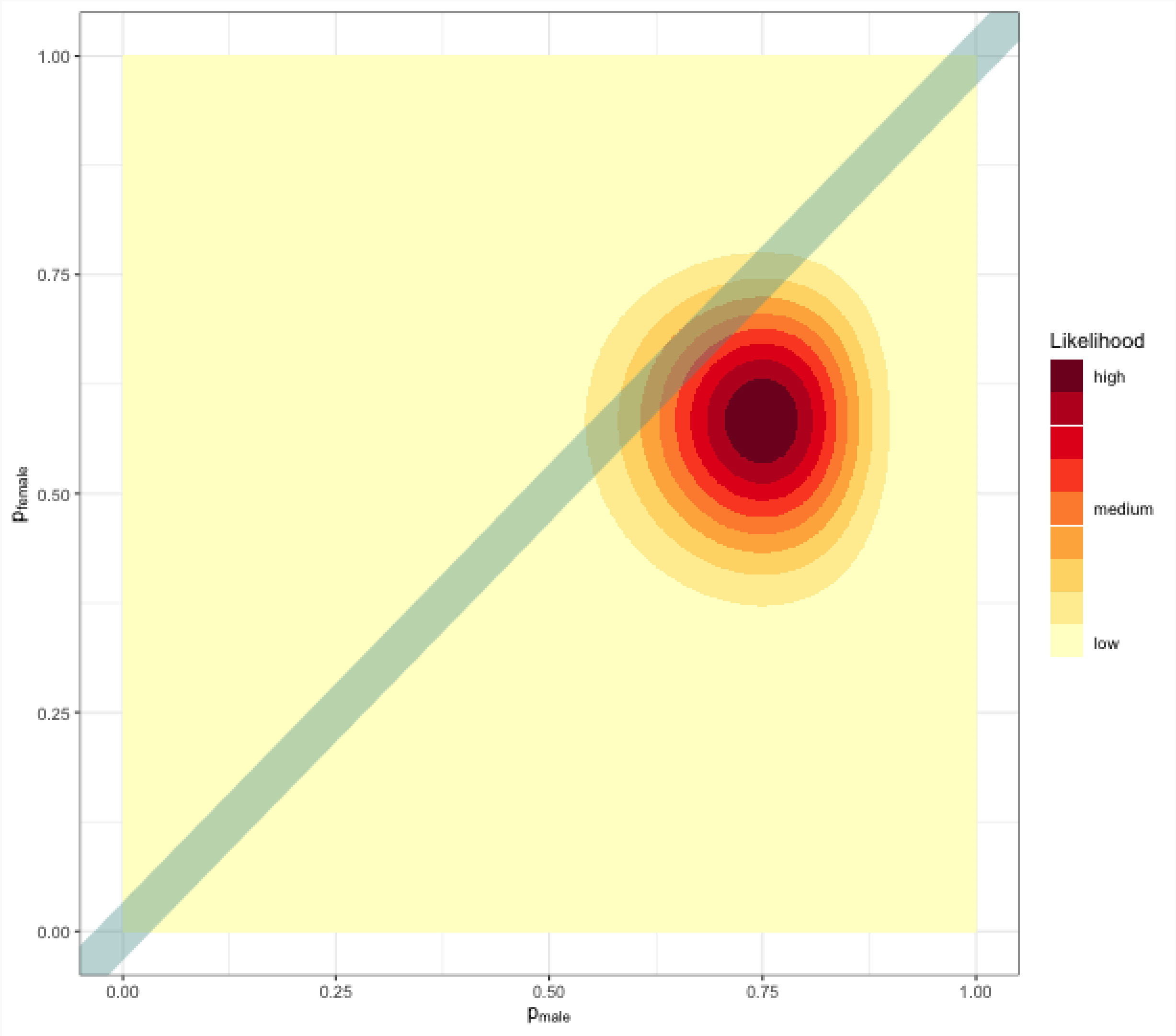
vs.

$$L(p_M, p_F|n, y, x) = \binom{n}{y} {p_M}^y (1 - p_M)^{n-y} \binom{n}{x} {p_F}^x (1 - p_F)^{n-x}$$

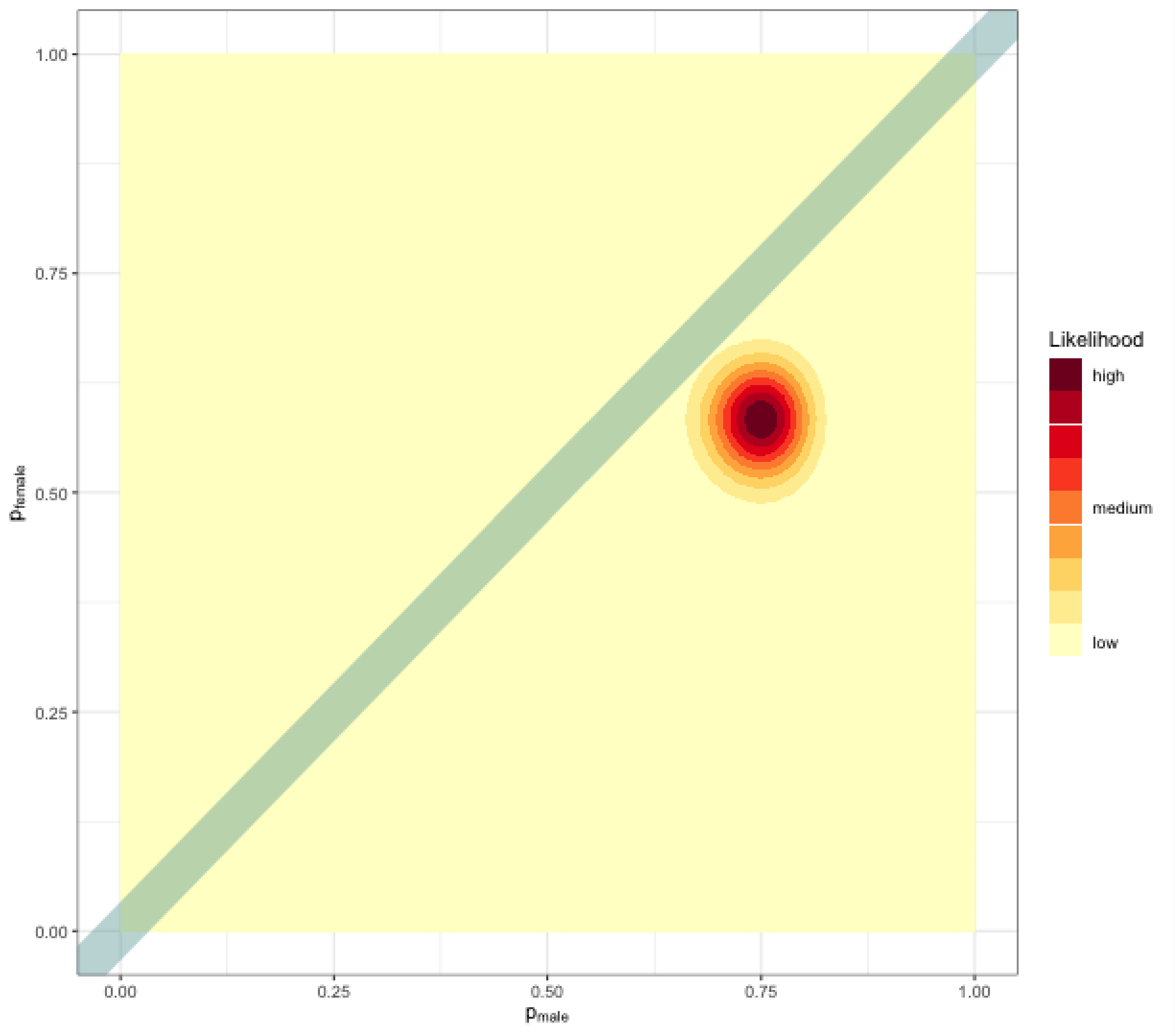
The Likelihood Function



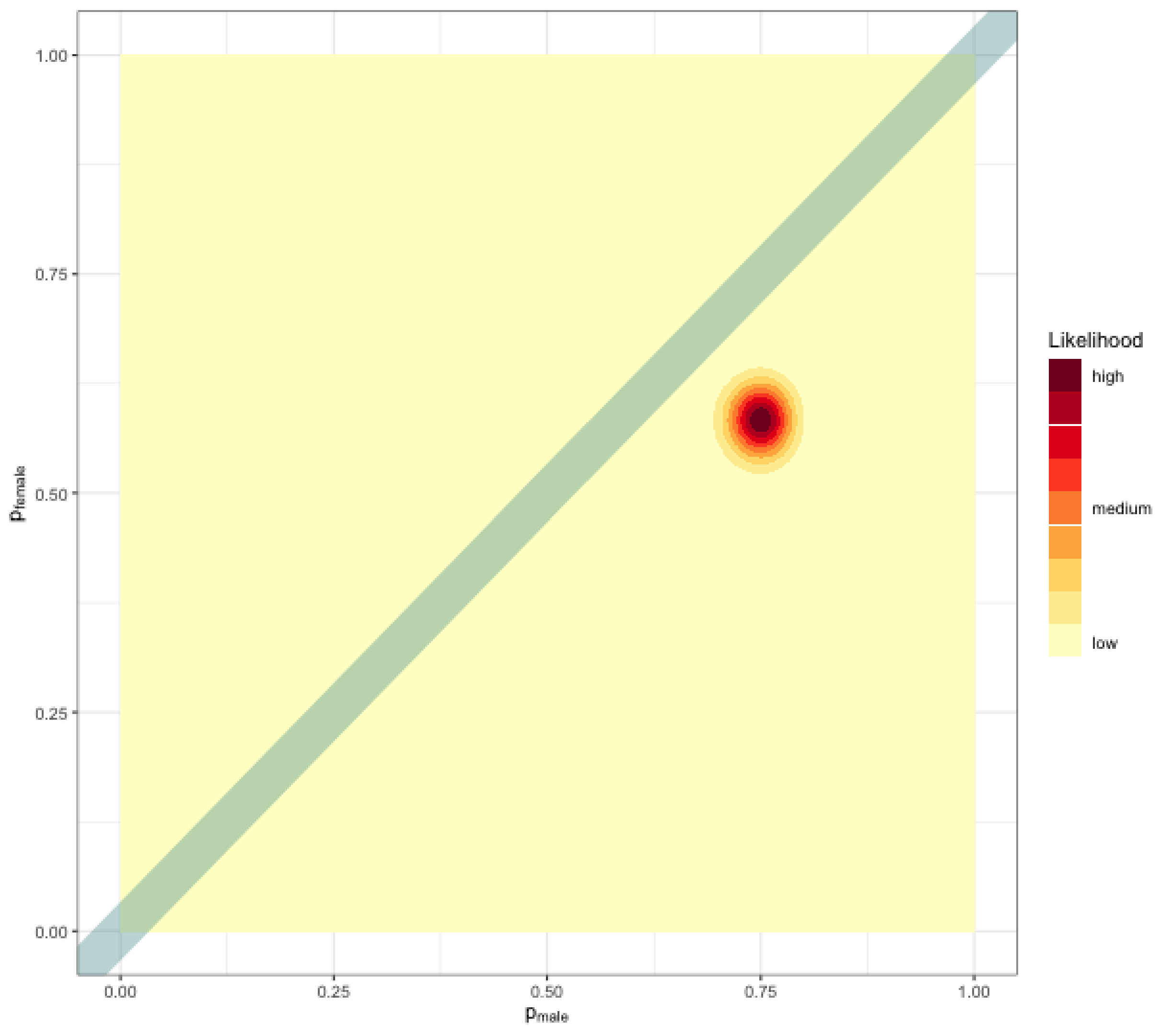
The Likelihood Function



Likelihood Function, more data

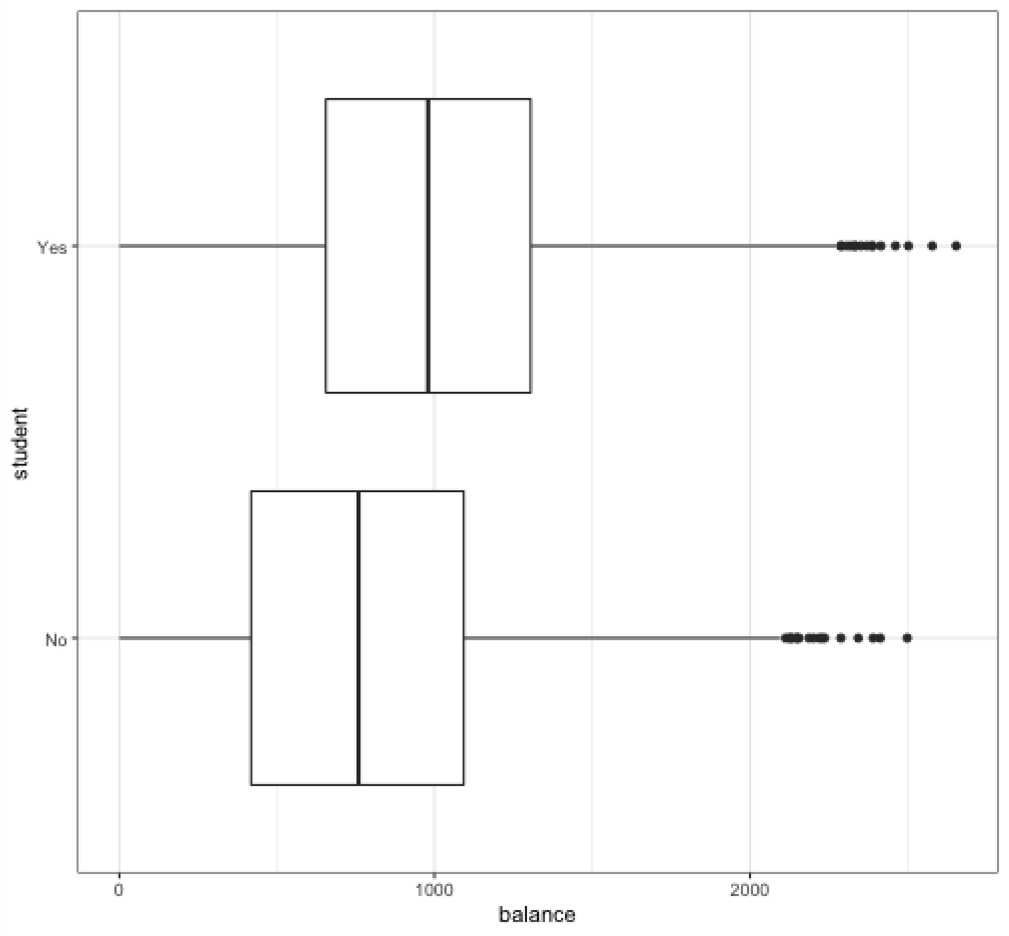


Likelihood Function, even more data



Multiple Logisitic Regression

Add student as a predictor?



Multiple Logistic Model

```r
m2 <- glm(default ~ balance + student,
          data = Default,
          family = binomial)
summary(m2)$coef
```

```
##                 Estimate  Std. Error    z value      Pr(>|z|)
## (Intercept) -10.749495878 0.369191361 -29.116326 2.230782e-186
## balance       0.005738104 0.000231847  24.749526 3.136911e-135
## studentYes   -0.714877620 0.147519010  -4.846003  1.259734e-06
```

What's going on?

Multiple Logistic Model, cont.