Complex Interactions

Introduction

- Simple interactions
  - Between continuous variables
  - Including categorical variables
- Dealing with non-linear associations
- Splines
- Non-parametric regression

Height and Weight Data

```
require(readr)
require(dplyr)
require(magrittr)

howell1 <- read_delim('https://raw.githubusercontent.com/rm
    mutate(sex = ifelse(male, 'male', 'female'))
```

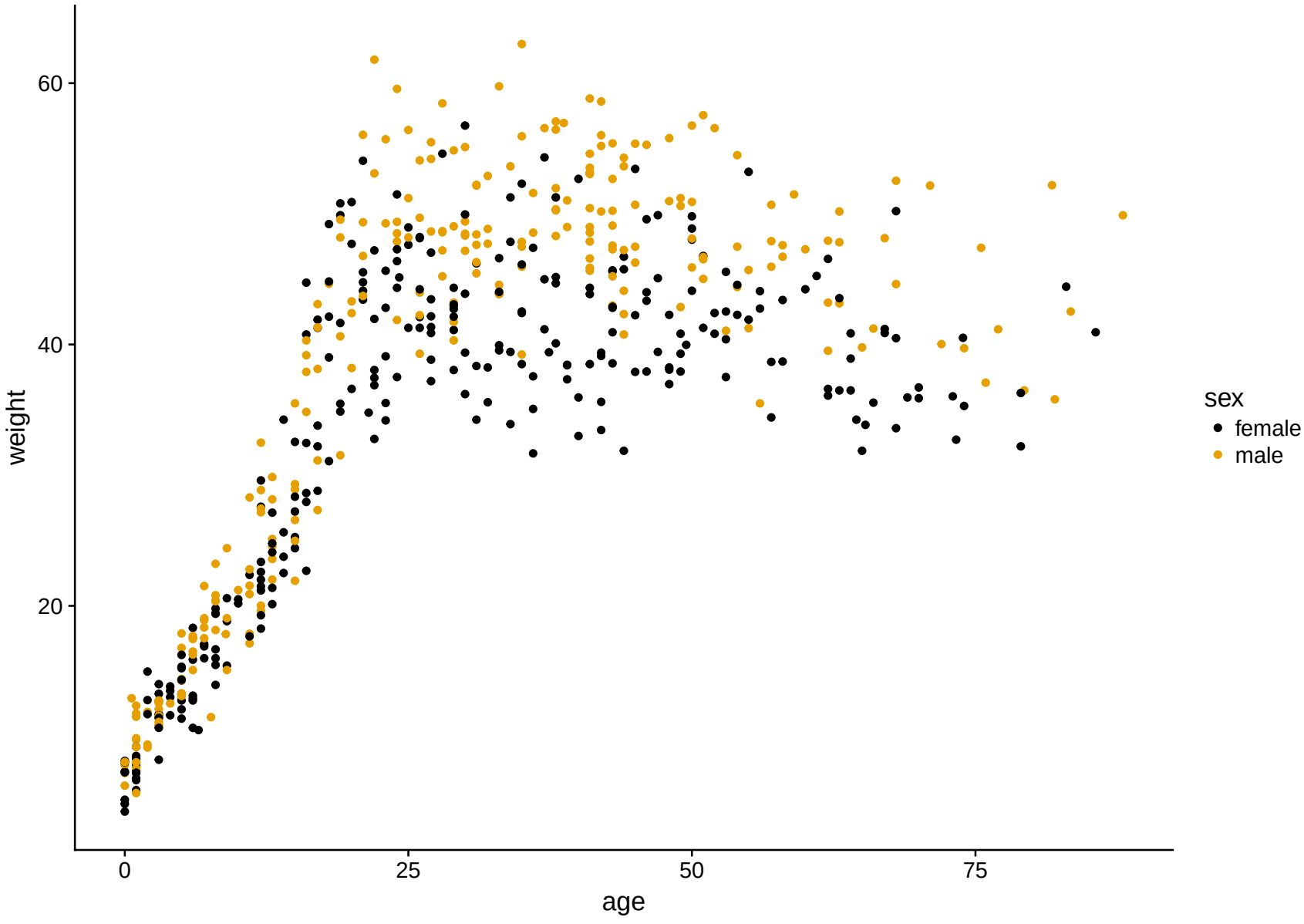
```
## Parsed with column specification:
## cols(
##   height = col_double(),
##   weight = col_double(),
##   age = col_double(),
##   male = col_integer()
## )
```

(data set available at https://github.com/rmcelreath/rethinking)

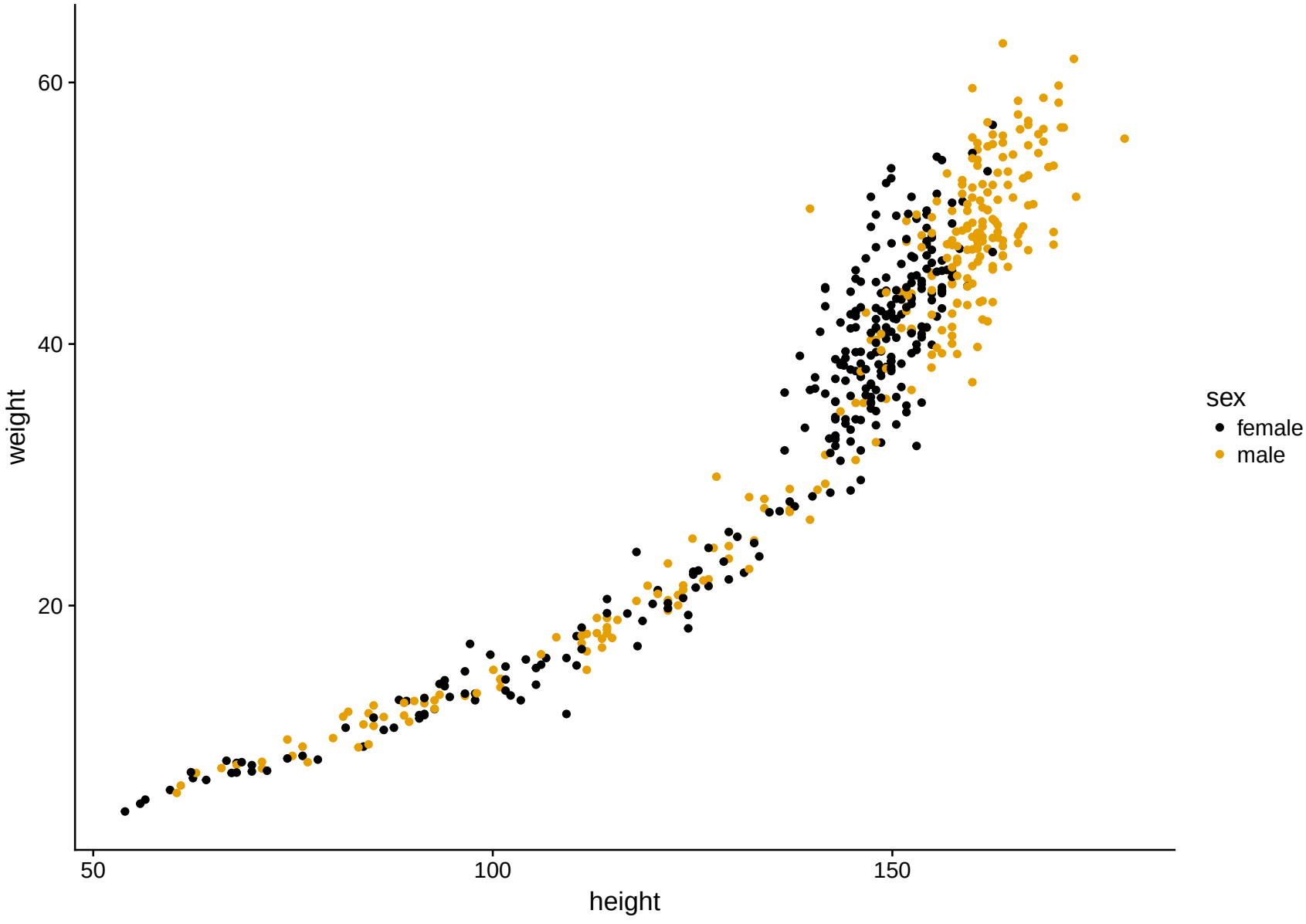
Height and Weight Data

```
## # A tibble: 544 x 5
##    height weight   age  male sex
##     <dbl>  <dbl> <dbl> <int> <chr>
##  1    152   47.8  63.0     1 male
##  2    140   36.5  63.0     0 female
##  3    137   31.9  65.0     0 female
##  4    157   53.0  41.0     1 male
##  5    145   41.3  51.0     0 female
##  6    164   63.0  35.0     1 male
##  7    149   38.2  32.0     0 female
##  8    169   55.5  27.0     1 male
##  9    148   34.9  19.0     0 female
## 10    165   54.5  54.0     1 male
## # ... with 534 more rows
```

Simple Linear Model



Simple Linear Model

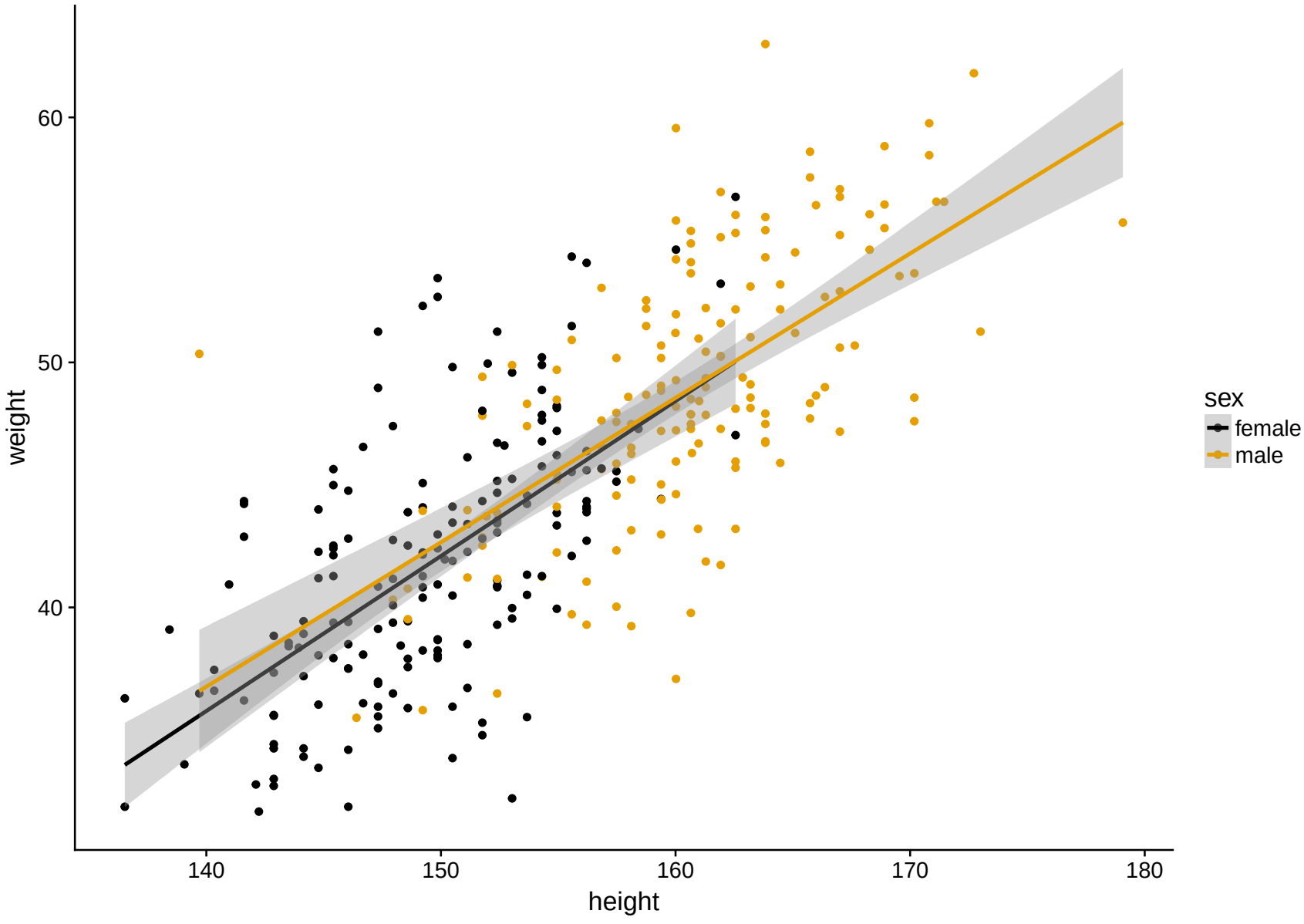


Simple Linear Model

$$weight_i = \beta_0 + \beta_1 * height_i + \beta_2 * male_i + \varepsilon_i$$

- `weight`: outcome / dependent variable
- β_0, β_1, and β_2: regression coefficients
- `height` and `male`: predictors / independent variables
- ε: error term,
- `i` subscript: specific sample

Adding Interactions



Adding Interactions

$$weight_i = \beta_0 + \beta_1 * height_i + \beta_2 * male_i + \beta_3 * male_i * height_i + \varepsilon_i$$

- β_0: `weight` for females who are 0 cm tall
- β_1: mean difference in `weight` for males, compared to females
- β_2: slope of the line describing change in female `weight` as `height` changes
- β_3: the difference in the slope for males, compared to the slope of the line for females

Adding Interactions

```
model <- lm(weight ~ height + male + male:height,
            data = filter(howell1, age > 20))

require(broom)
tidy(model) %>%
    print(digits = 3)
```

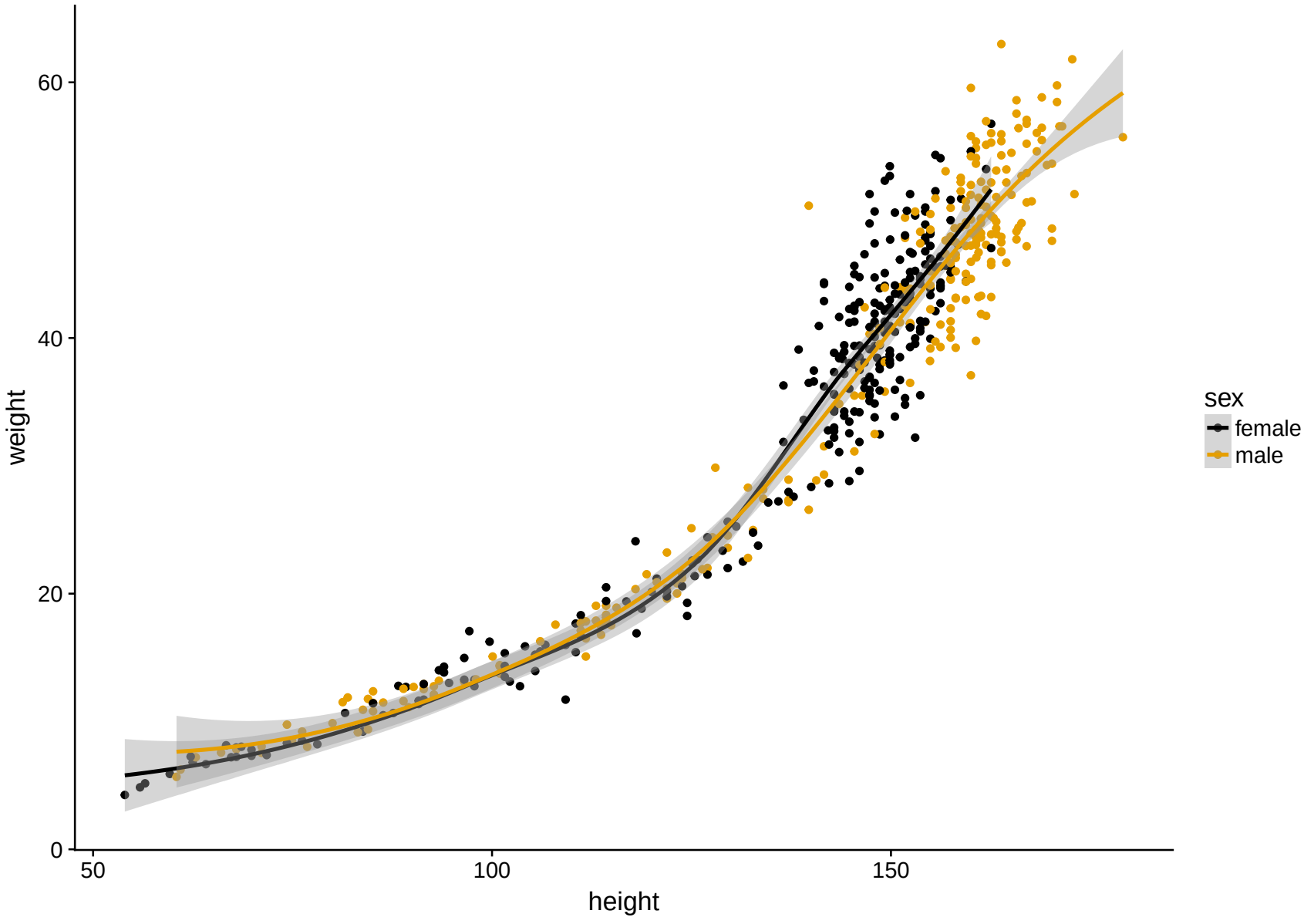
```
##          term estimate std.error statistic  p.value
## 1 (Intercept) -52.7968    9.4301    -5.599 4.58e-08
## 2      height   0.6326    0.0631    10.032 7.95e-21
## 3        male   7.0458   13.2496     0.532 5.95e-01
## 4 height:male  -0.0432    0.0856    -0.505 6.14e-01
```

Model Interpretation

$$
\begin{aligned}
\mathrm{E}(\textit{weight} \mid \textit{male} = 0) &= \beta_0 + \beta_1 * \textit{height} + \cancel{\beta_2 * 0} + \cancel{\beta_3 * 0 * \textit{height}} \\
&= -52.8 + 0.633 * \textit{height}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\textit{weight} \mid \textit{male} = 1) &= \beta_0 + \beta_1 * \textit{height} + \beta_2 * 1 + \beta_3 * 1 * \textit{height} \\
&= (-52.8 - 7.0) + (0.633 - 0.043) * \textit{height} \\
&= -59.7 + 0.59 * \textit{height}
\end{aligned}
$$

Non-linear Associations



Non-linear Associations

```
howell1 <- mutate(howell1, lweight = log(weight))

log_model <- lm(lweight ~ height*male,
                data = howell1)

tidy(log_model) %>%
    print(digits = 3)
```

```
##          term estimate std.error statistic  p.value
## 1 (Intercept)  0.50749  0.032442     15.64 9.17e-46
## 2      height  0.02138  0.000237     90.36 0.00e+00
## 3        male  0.16583  0.046006      3.60 3.42e-04
## 4 height:male -0.00148  0.000326     -4.55 6.62e-06
```

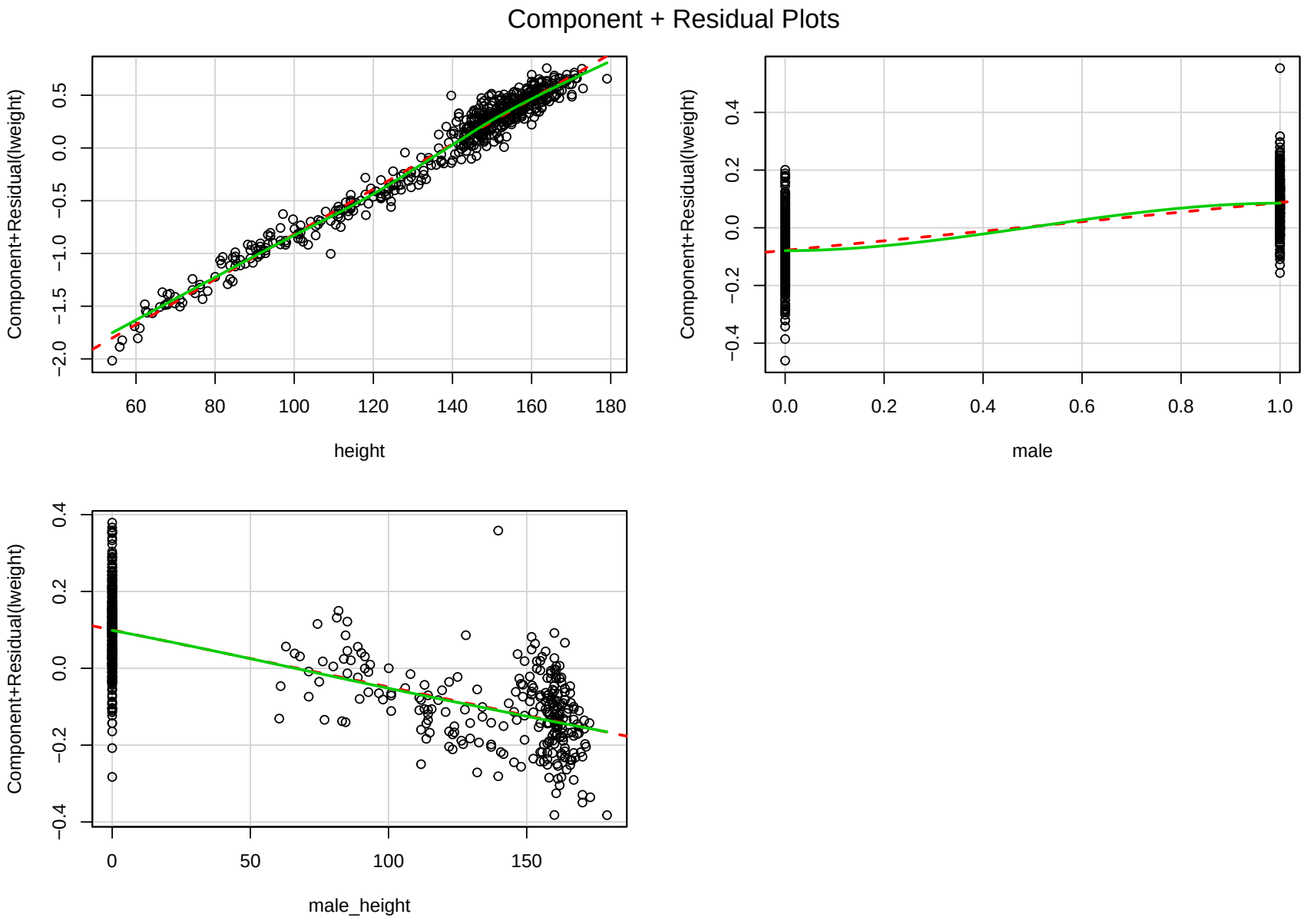
Model Interpretation

$$
\begin{aligned}
\mathrm{E}(\log(\textit{weight}) \mid \textit{male} = 0) &= \beta_0 + \beta_1 * \textit{height} + \cancel{\beta_2 * 0} + \cancel{\beta_3 * 0 * \textit{ht}} \\
&= 0.507 + 0.021 * \textit{height}
\end{aligned}
$$

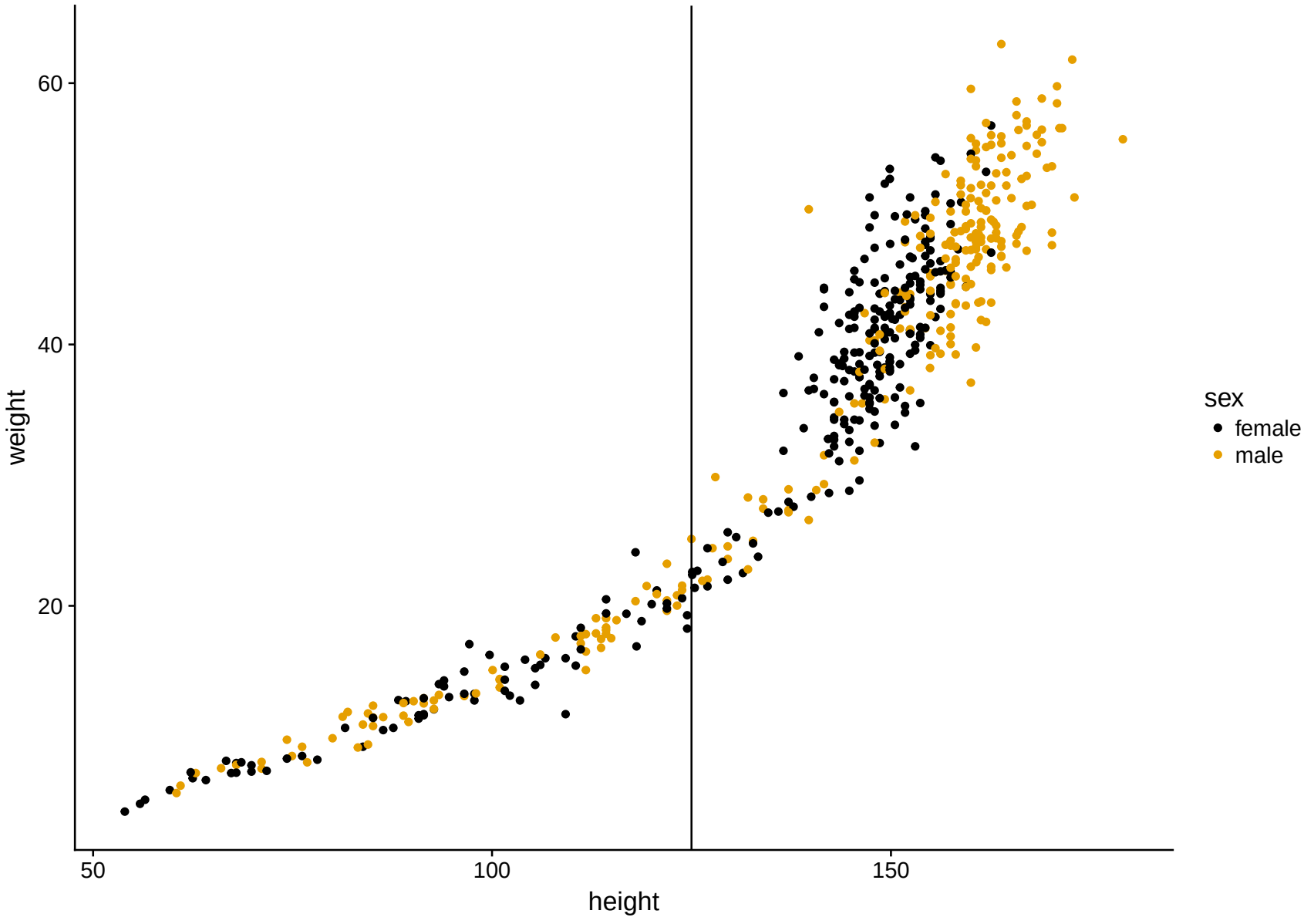
$$
\begin{aligned}
\mathrm{E}(\log(\textit{weight}) \mid \textit{male} = 1) &= \beta_0 + \beta_1 * \textit{height} + \beta_2 * 1 + \beta_3 * 1 * \textit{ht} \\
&= (0.507 + 0.166) + (0.021 - 0.001) * \textit{ht} \\
&= 0.673 + 0.020 * \textit{height}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\textit{weight} \mid \textit{male} = 1, \textit{height} = 145) &= e^{0.673 + 0.02 * 145} \\
&= 35.6 \textit{kg}
\end{aligned}
$$

Model Checks



Splines



Splines

```
knot <- 125
howell1 <- mutate(howell1,
                  height_spline = ifelse(height > knot, he

spline_model <- lm(weight ~ height + height_spline,
                   data = howell1)
```

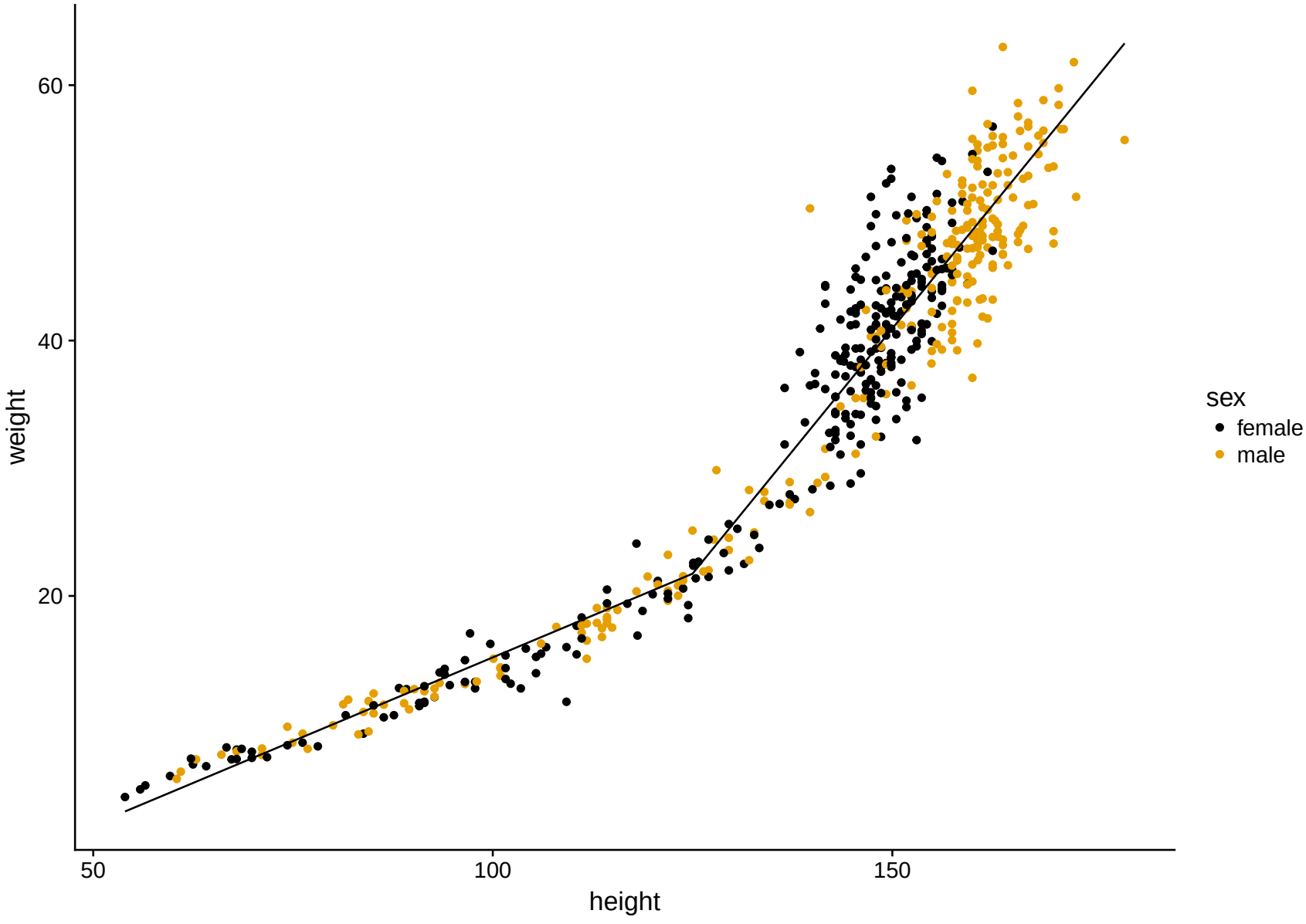
```
##             term estimate std.error statistic  p.value
## 1    (Intercept)  -11.035    1.4320     -7.71 6.23e-14
## 2         height    0.262    0.0136     19.27 2.10e-63
## 3 height_spline    0.506    0.0259     19.57 6.94e-65
```

Splines

$$
\begin{aligned}
\mathrm{E}(\mathit{weight} \mid \mathit{ht} \leq 125) &= \beta_0 + \beta_1 * \mathit{ht} + \beta_2 * \cancel{\min(0, \mathit{ht} - 125)} \\
&= -9.8 + 0.247 * \mathit{ht}
\end{aligned}
$$

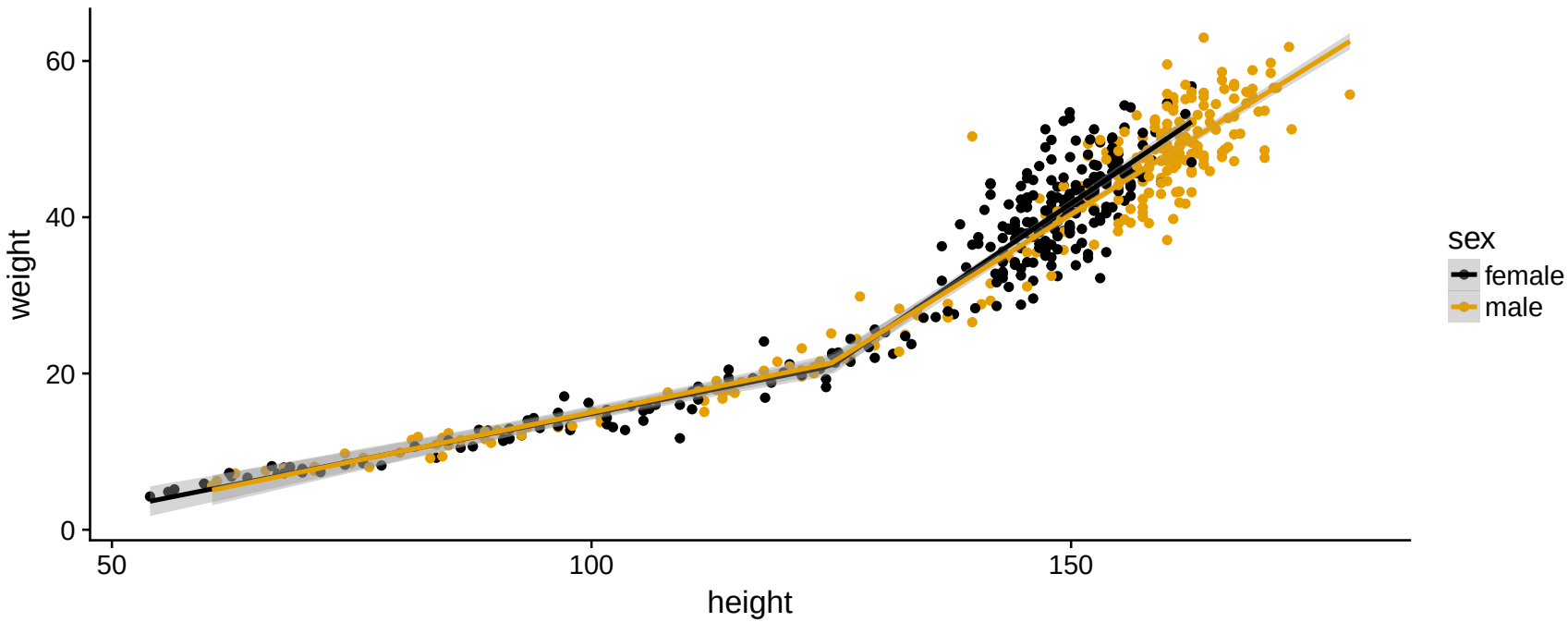
$$
\begin{aligned}
\mathrm{E}(\mathit{weight} \mid \mathit{ht} > 125) &= \beta_0 + \beta_1 * \mathit{ht} + \beta_2 * \min(0, \mathit{ht} - 125) \\
&= \beta_0 + \beta_1 * \mathit{ht} + (\beta_2 * \mathit{ht} - \beta_2 * 125) \\
&= (-9.8 - 0.508 * 125) + (0.247 + 0.508) * \mathit{ht} \\
&= -73.3 + 0.755 * \mathit{ht}
\end{aligned}
$$

Splines



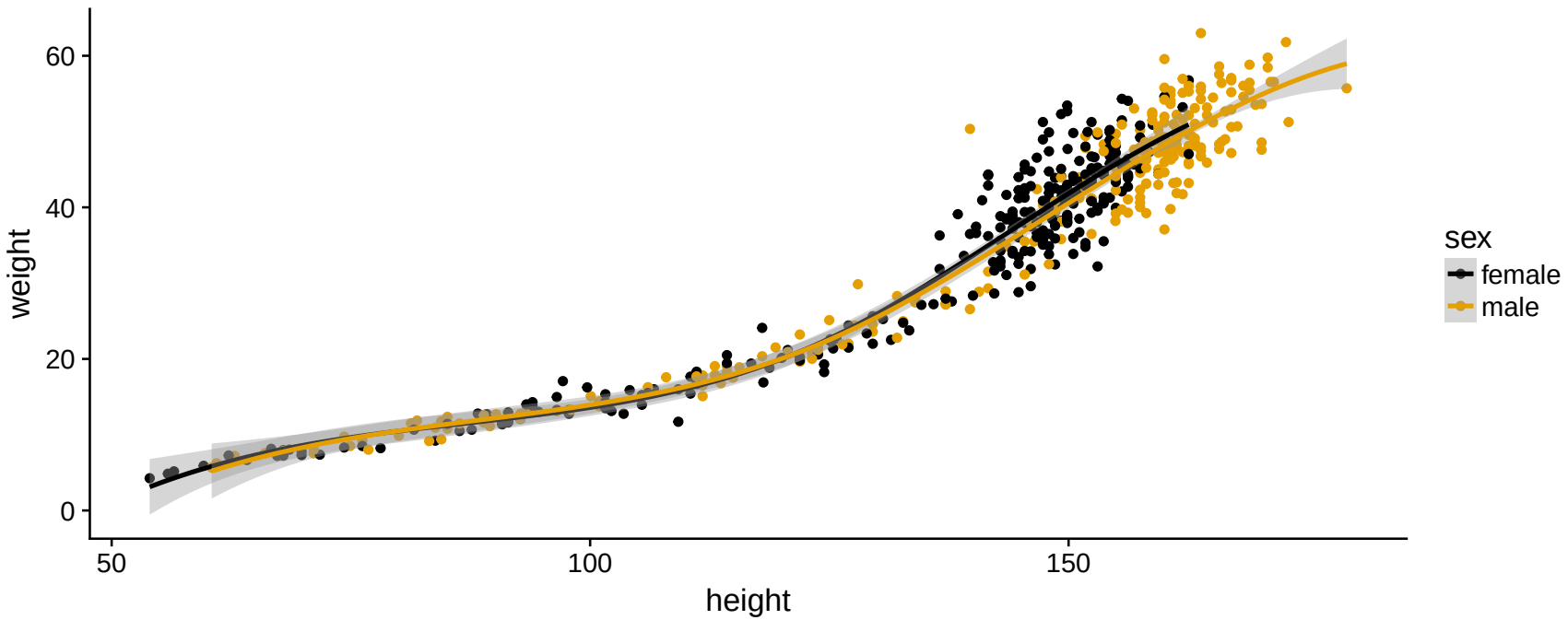
Splines

```
require(splines)
ggplot(howell1, aes(height, weight, color=sex)) +
    geom_point() +
    scale_color_manual(values = cbbPalette) +
    geom_smooth(method = 'lm',
                formula = y ~ bs(x, knots=125, degree=1))
```



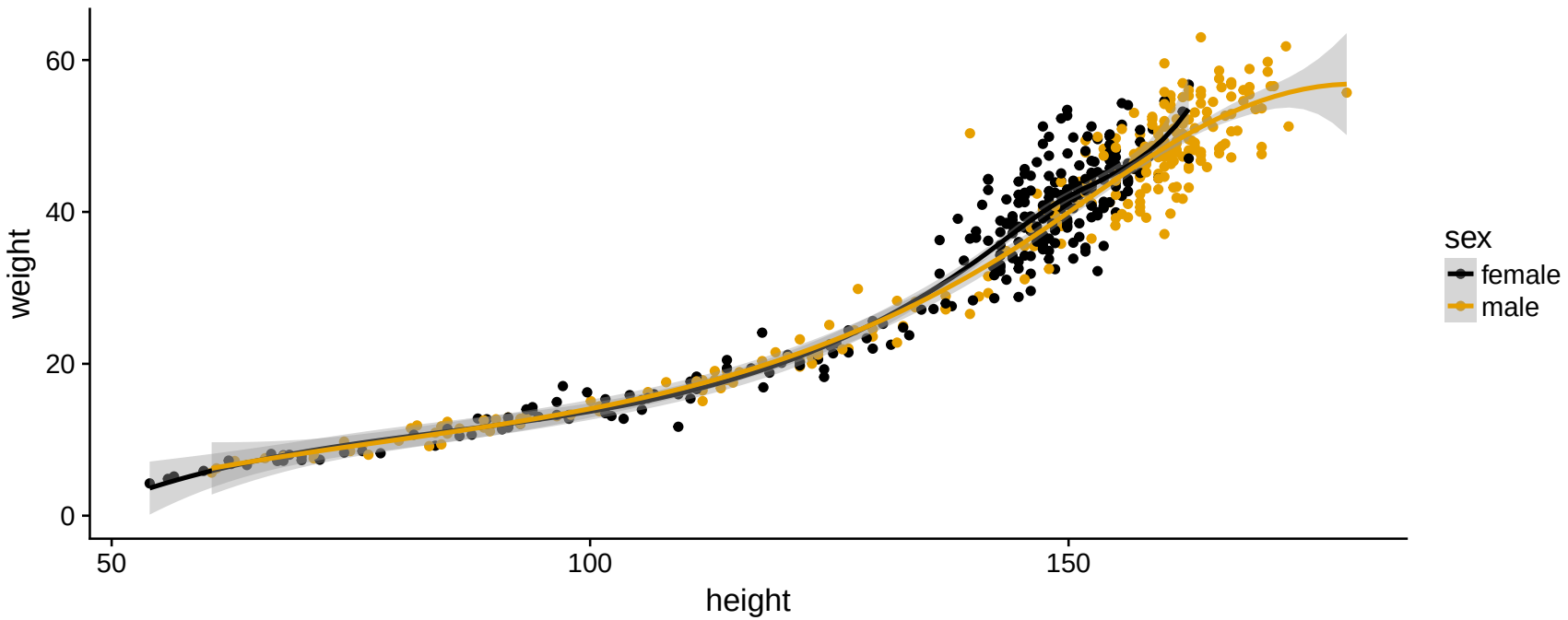
Splines

```
ggplot(howell1, aes(height, weight, color=sex)) +
    geom_point() +
    scale_color_manual(values = cbbPalette) +
    geom_smooth(method = 'lm',
                formula = y ~ bs(x, knots = 125))
```



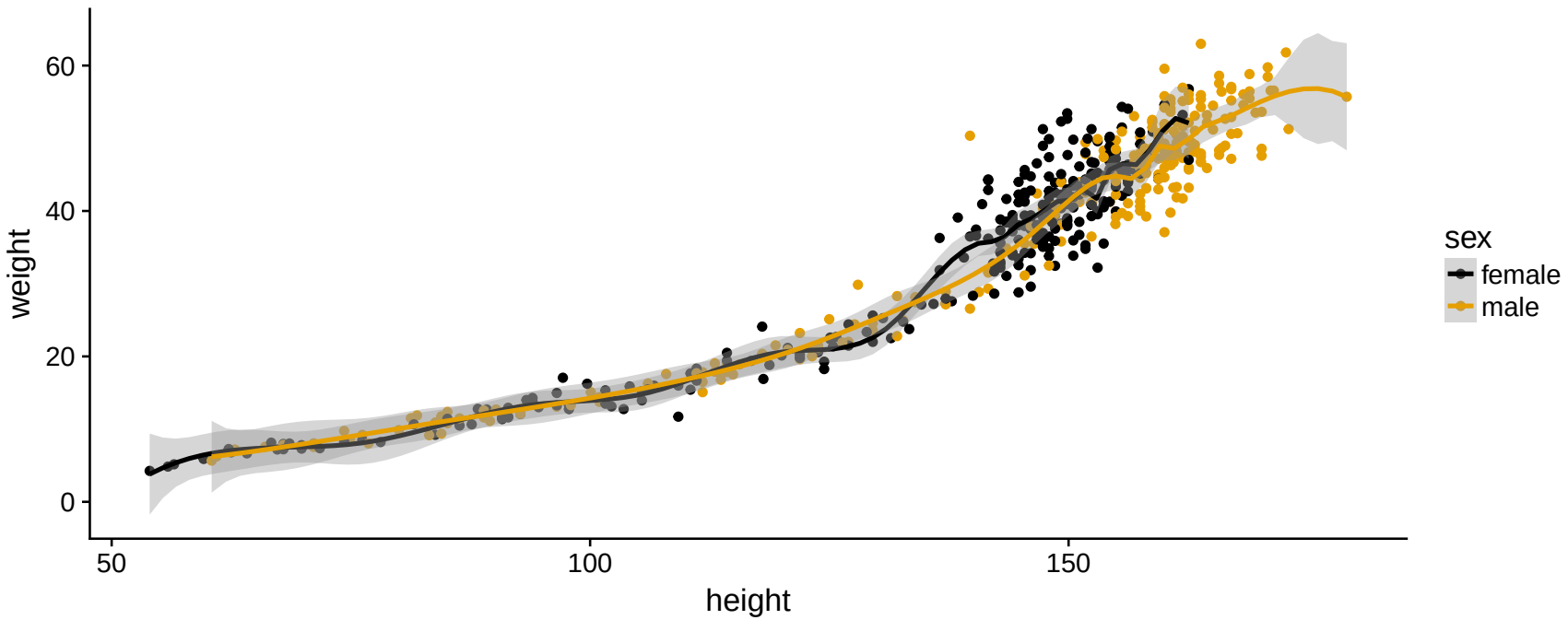
Splines

```
ggplot(howell1, aes(height, weight, color=sex)) +
    geom_point() +
    scale_color_manual(values = cbbPalette) +
    geom_smooth(method = 'lm',
                formula = y ~ bs(x, df = 5))
```



Splines

```
ggplot(howell1, aes(height, weight, color=sex)) +
    geom_point() +
    scale_color_manual(values = cbbPalette) +
    geom_smooth(method = 'lm',
                formula = y ~ bs(x, df = 20))
```



Non-parametric Regression

```
ggplot(howell1, aes(height, weight, color=sex)) +
    geom_point() +
    scale_color_manual(values = cbbPalette) +
    geom_smooth(method = 'loess')
```