Adding a numerical explanatory variable

MULTIPLE AND LOGISTIC REGRESSION IN R

Ben Baumer
Instructor

Adding a second numeric explanatory variable

- Mathematical:

$$\hat{bwt} = \hat{\beta}_0 + \hat{\beta}_1 \cdot gestation + \hat{\beta}_2 \cdot age$$

- Syntactical:

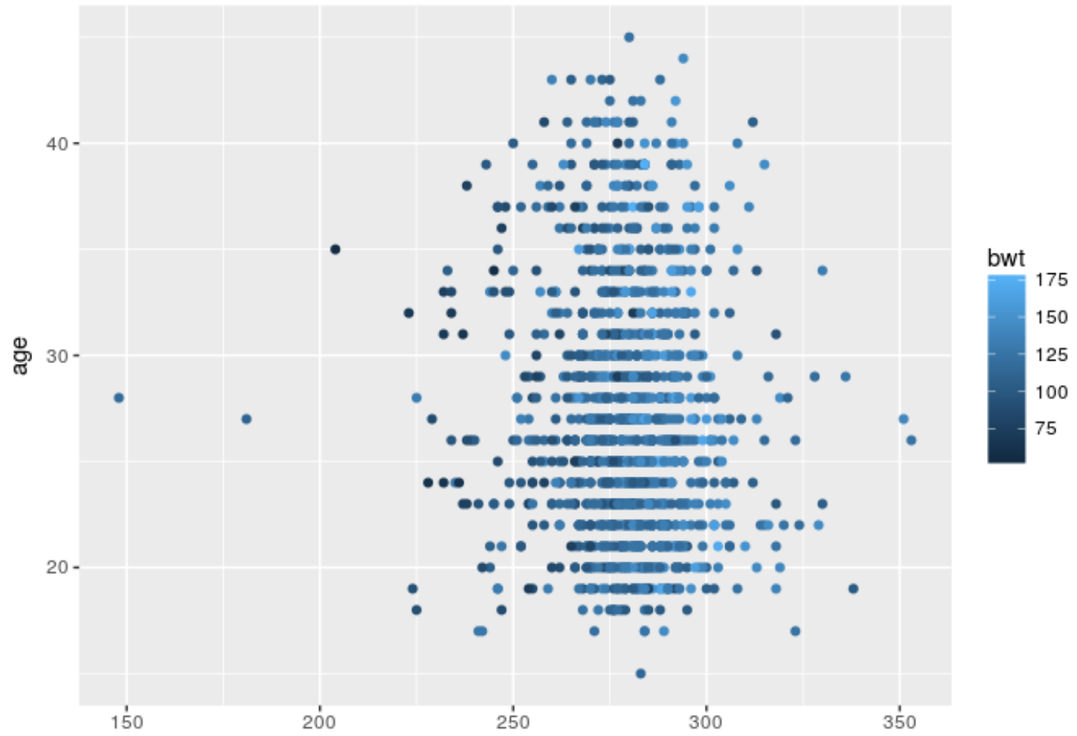
```
lm(bwt ~ gestation + age, data = babies)
```

No longer a 2D problem

```
# doesn't work
ggplot(data = babies, aes(x = gestation, y = age, z = bwt)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

Data space is 3D

```
data_space <- ggplot(babies, aes(x = gestation, y = age)) +
  geom_point(aes(color = bwt))
data_space
```



Tiling the plane

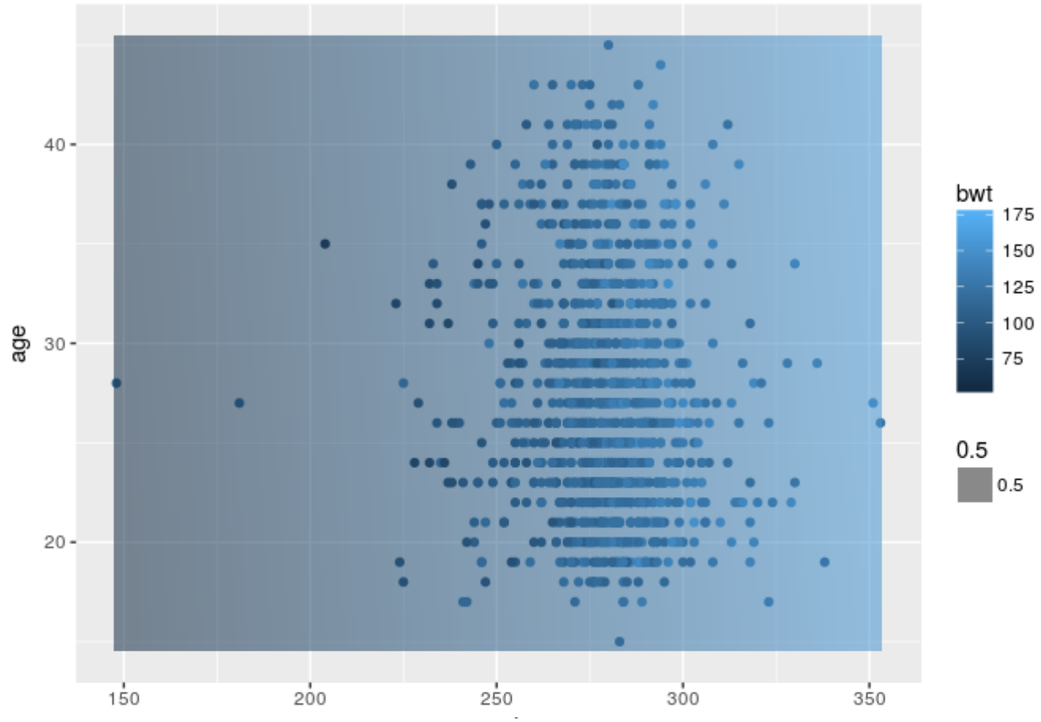
```
grid <- babies %>%
  data_grid(
    gestation = seq_range(gestation, by = 1),
    age = seq_range(age, by = 1)
  )

mod <- lm(bwt ~ gestation + age, data = babies)


bwt_hats <- augment(mod, newdata = grid)
```

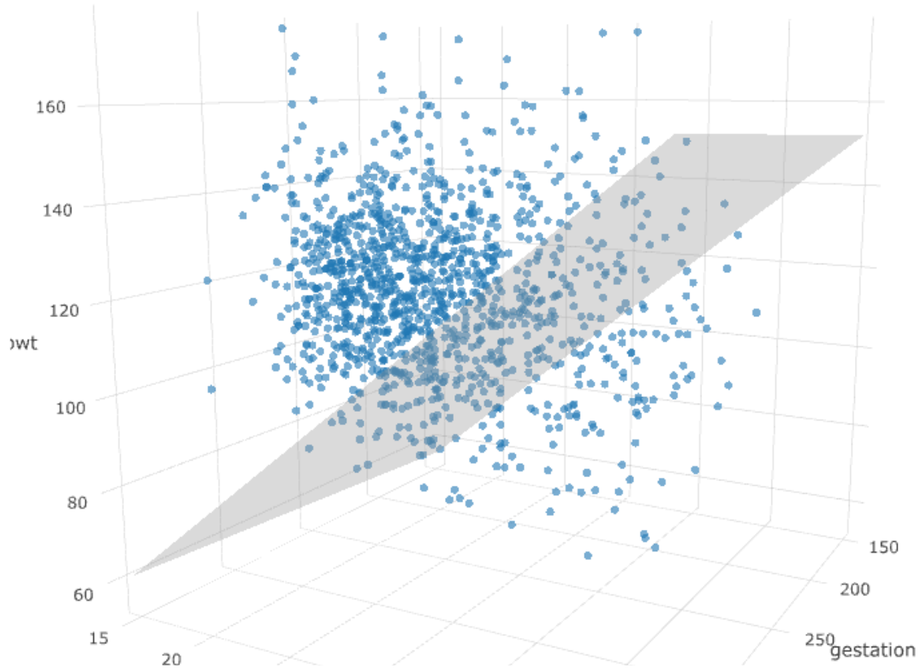
Tiles in the data space

```
data_space +
  geom_tile(data = bwt_hats, aes(fill = .fitted, alpha = 0.5)) +
  scale_fill_continuous("bwt", limits = range(babies$bwt))
```



3D visualization

```
plot_ly(data = babies, z = ~bwt, x = ~gestation, y = ~age, opacity = 0.6) %>%
  add_markers(text = ~case, marker = list(size = 2)) %>%
  add_surface(x = ~x, y = ~y, z = ~plane, showscale = FALSE,
              cmax = 1, surfacecolor = color1, colorscale = col1)
```



Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

Conditional interpretation of coefficients

MULTIPLE AND LOGISTIC REGRESSION IN R

Ben Baumer
Instructor

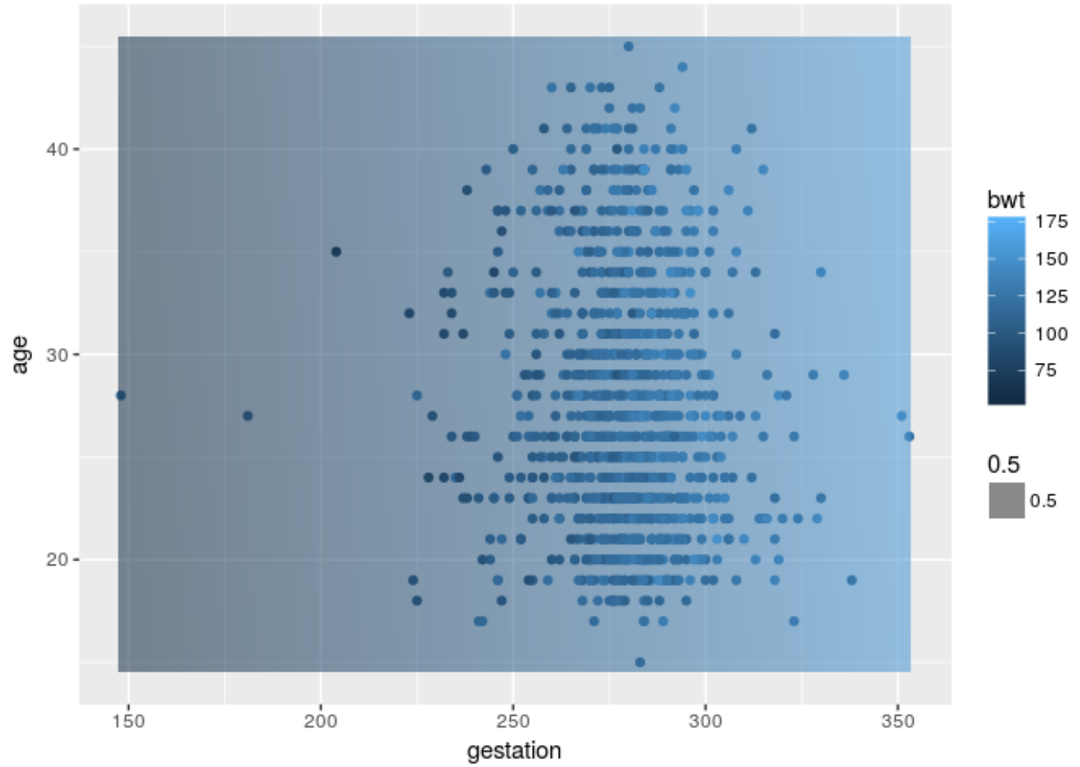
Two slope coefficients

```
lm(bwt ~ gestation + age, data = babies)
```

```
## Coefficients:
## (Intercept)    gestation          age
##    -15.5226       0.4676       0.1657
```

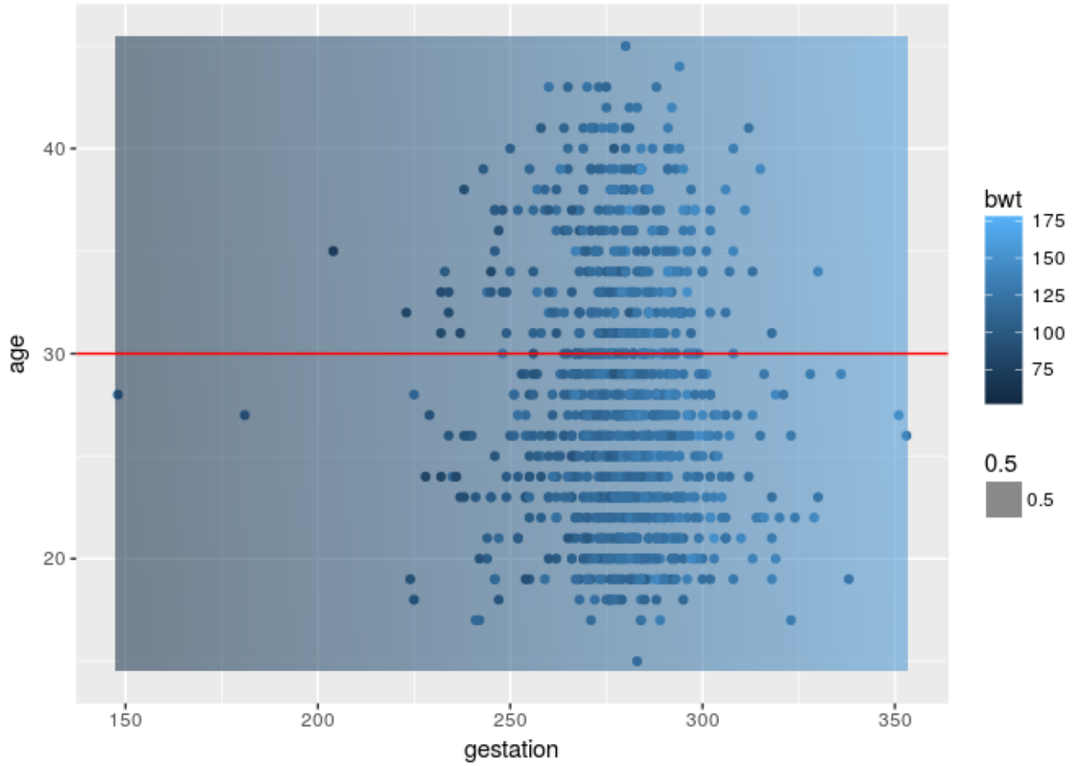
Tiled plane

```
model_space
```



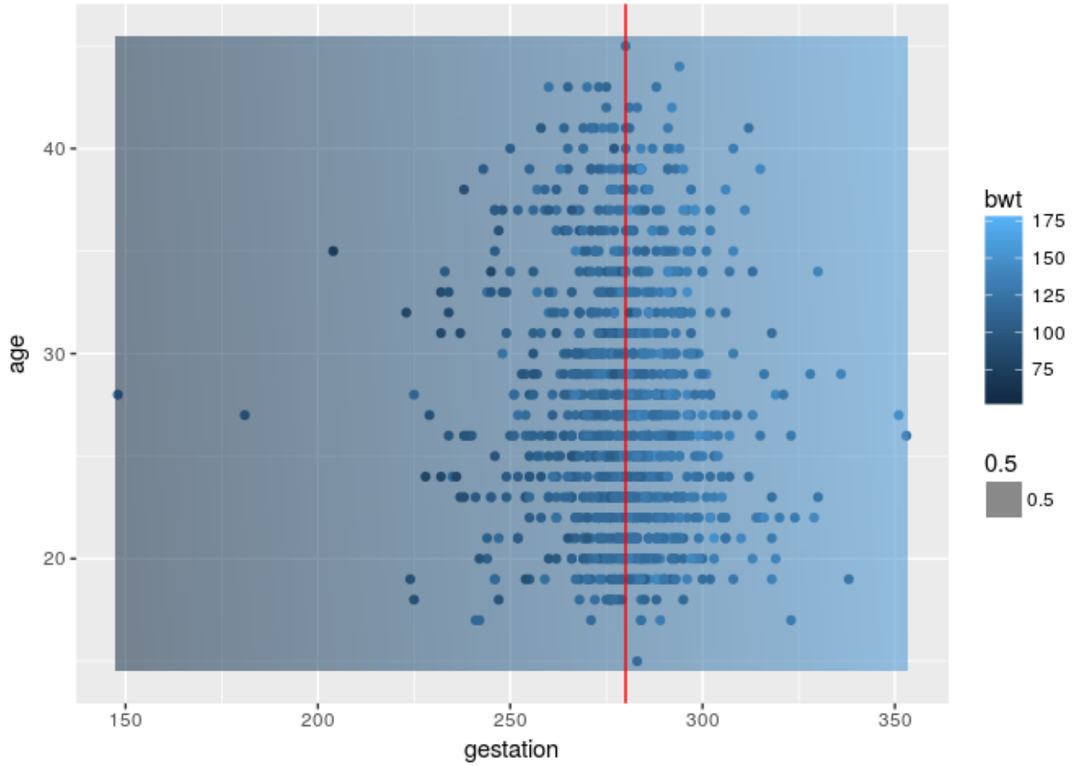
Tiled plane plus first slope

```
model_space +
  geom_hline(yintercept = 30, color = "red")
```



Tiled plane plus second slope

```
model_space +
  geom_vline(xintercept = 280, color = "red")
```



Coefficient interpretation

```
lm(bwt ~ gestation + age, data = babies)
```

```
## Coefficients:
## (Intercept)    gestation          age
##    -15.5226       0.4676       0.1657
```

Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

Adding a third (categorical) variable

MULTIPLE AND LOGISTIC REGRESSION IN R

Ben Baumer
Instructor

How could we forget about smoking?

- Mathematical:

$$\hat{bwt} = \hat{\beta}_0 + \hat{\beta}_1 \cdot gestation + \hat{\beta}_2 \cdot age + \hat{\beta}_3 \cdot smoke$$

- Syntactical:

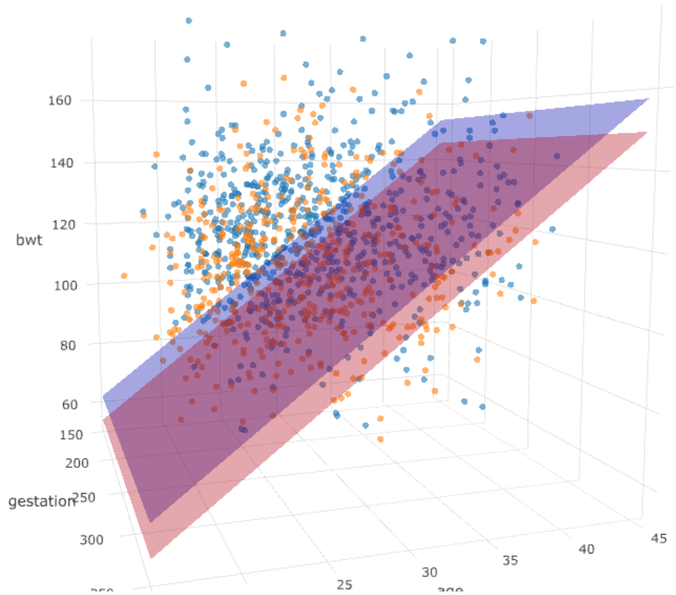
```
lm(bwt ~ gestation + age + smoke, data = babies
```

Geometry

- 1 numeric + 1 categorical:
  - parallel lines
- 2 numeric:
  - a plane
- 2 numeric + 1 categorical:
  - parallel planes!

Drawing parallel planes in 3D

```
plot_ly(data = babies, z = ~bwt, x = ~gestation, y = ~age, opacity = 0.6) %>%
  add_markers(color = ~factor(smoke), text = ~case, marker = list(size = 2)) %>
  add_surface(x = ~x, y = ~y, z = ~plane0, showscale = FALSE,
              cmin = 0, cmax = 1, surfacecolor = color1, colorscale = col1) %>%
  add_surface(x = ~x, y = ~y, z = ~plane1, showscale = FALSE,
              cmin = 0, cmax = 1, surfacecolor = color2, colorscale = col1)
```



Coefficient interpretation

```
lm(bwt ~ gestation + age, data = babies)
```

```
## Coefficients:
## (Intercept)    gestation          age
##    -15.5226       0.4676       0.1657
```

```
lm(bwt ~ gestation + age + smoke, data = babies)
```

```
## Coefficients:
## (Intercept)    gestation          age        smoke
##     -4.6037       0.4455       0.1069      -8.0143
```

Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

Higher dimensions

MULTIPLE AND LOGISTIC REGRESSION IN R

Ben Baumer
Instructor

Adding more variables

- Mathematical:

$$\hat{bwt} = \hat{\beta}_0 + \hat{\beta}_1 \cdot gestation + \hat{\beta}_2 \cdot age + \hat{\beta}_3 \cdot smoke +$$

$$+ \hat{\beta}_4 \cdot height + \hat{\beta}_5 \cdot weight + \hat{\beta}_6 \cdot parity$$

- Syntactical:

```
lm(bwt ~ gestation + age + smoke + height + weight + parity,
             data = babies)
```

- Syntactical (same model, but note order of coefficients)

```
lm(bwt ~ . - case, data = babies)
```

Higher dimensional geometry

- (Parallel) hyperplanes, etc.

Interpretation in large models

```
lm(bwt ~ gestation + age + smoke + height + weight + parity,
                  data = babies)
```

```
## Coefficients:
## (Intercept)    gestation          age        smoke       heig
##   -80.41085      0.44398     -0.00895     -8.40073      1.154
##      weight       parity
##     0.05017     -3.32720
```

Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R