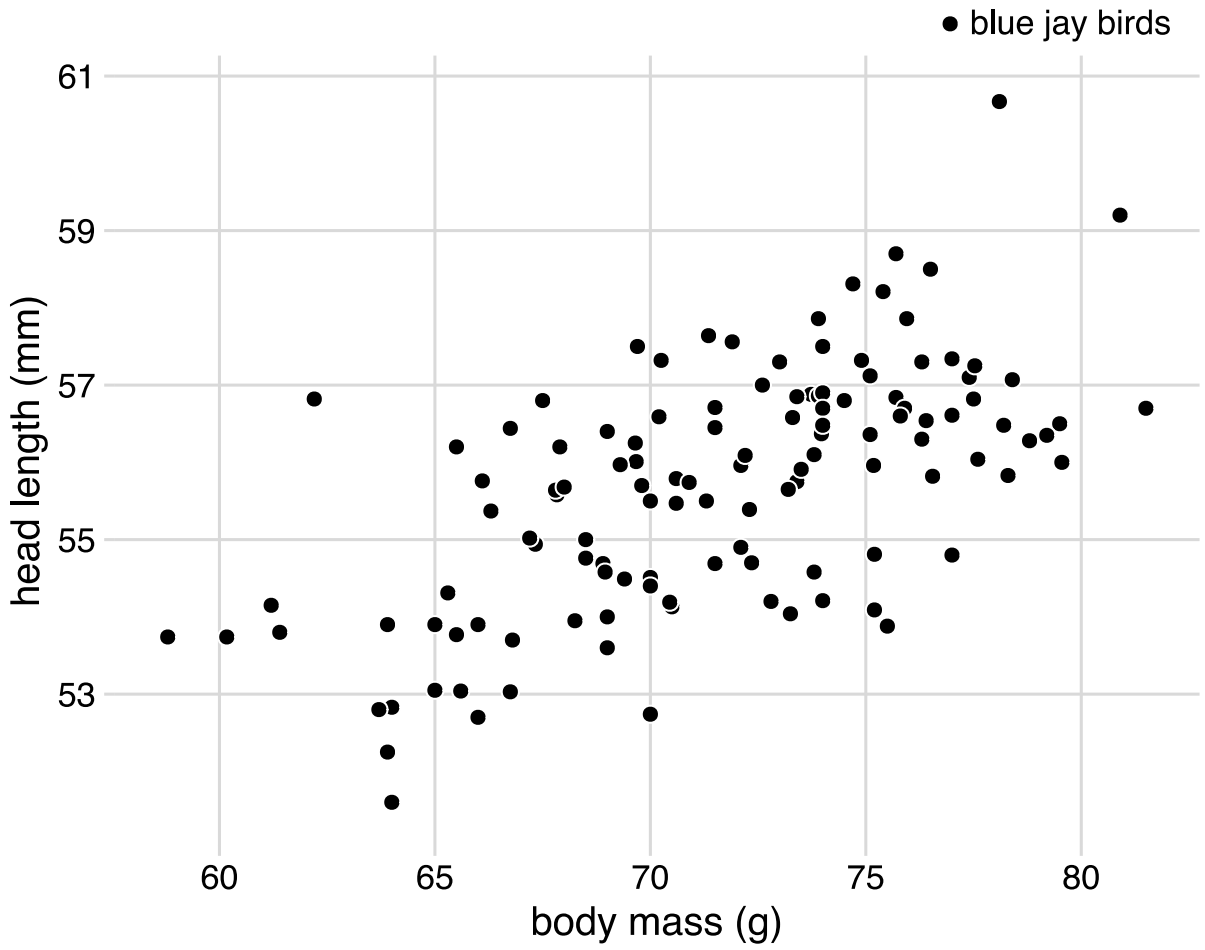
Visualizing trends

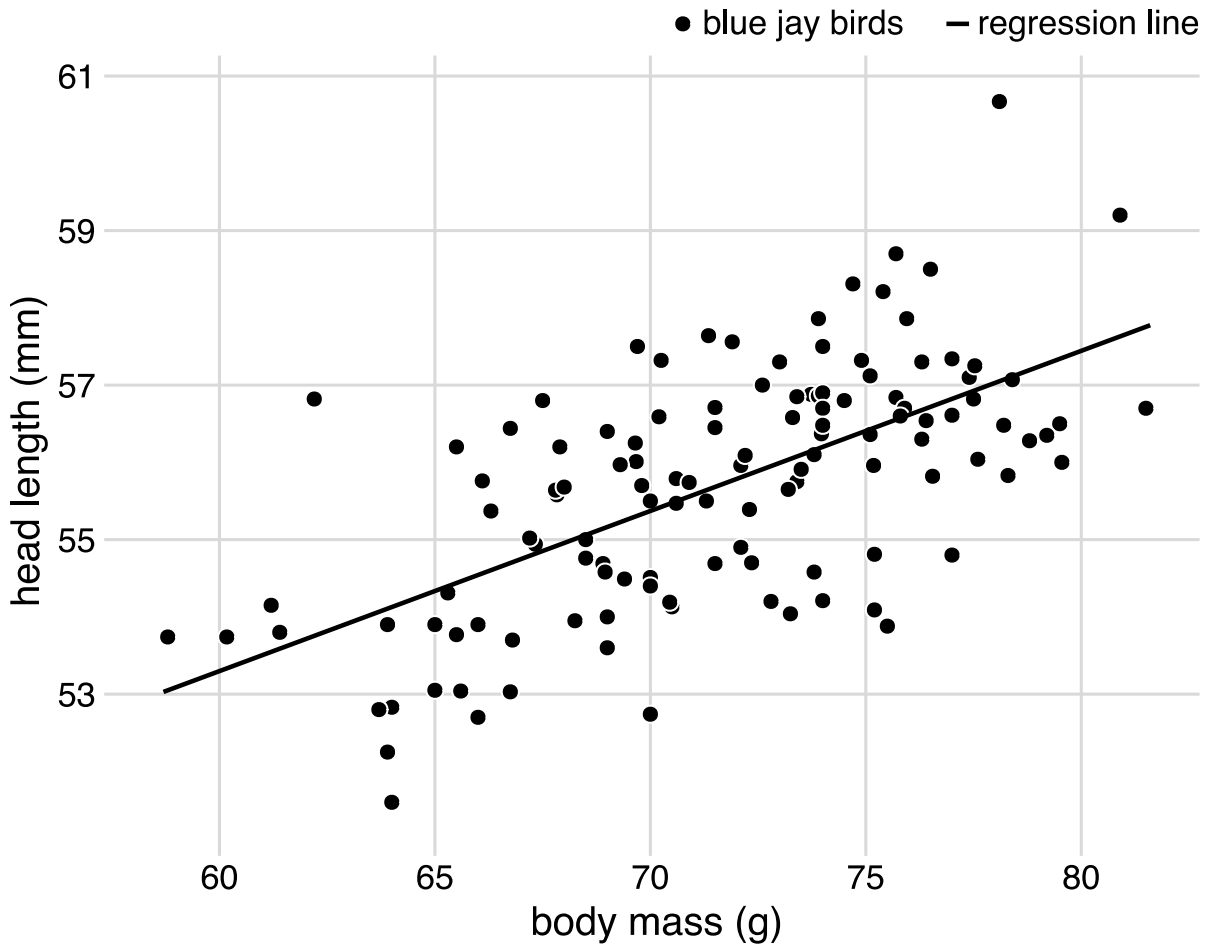
Claus O. Wilke

last updated: 2021-03-09

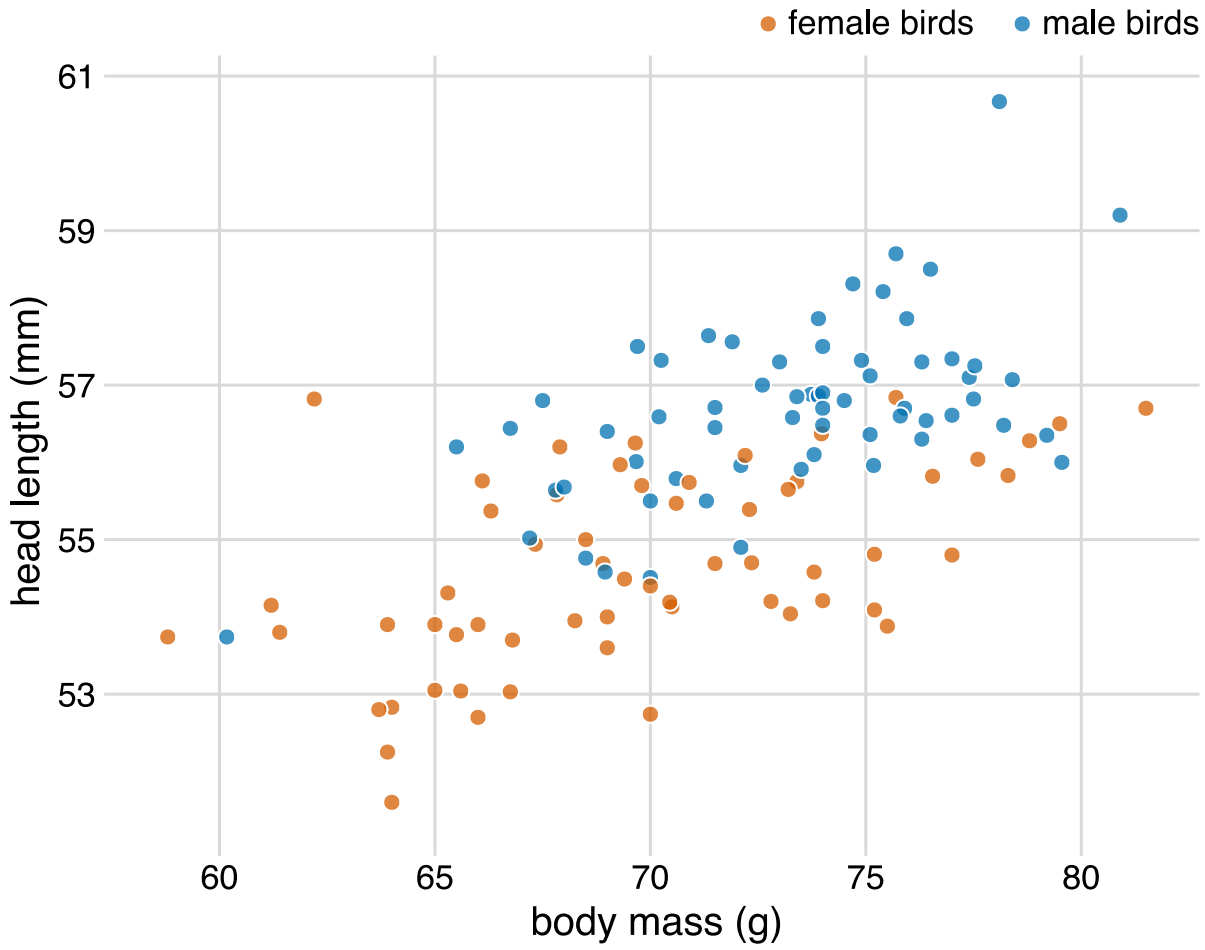
We visualize linear trends with regression lines



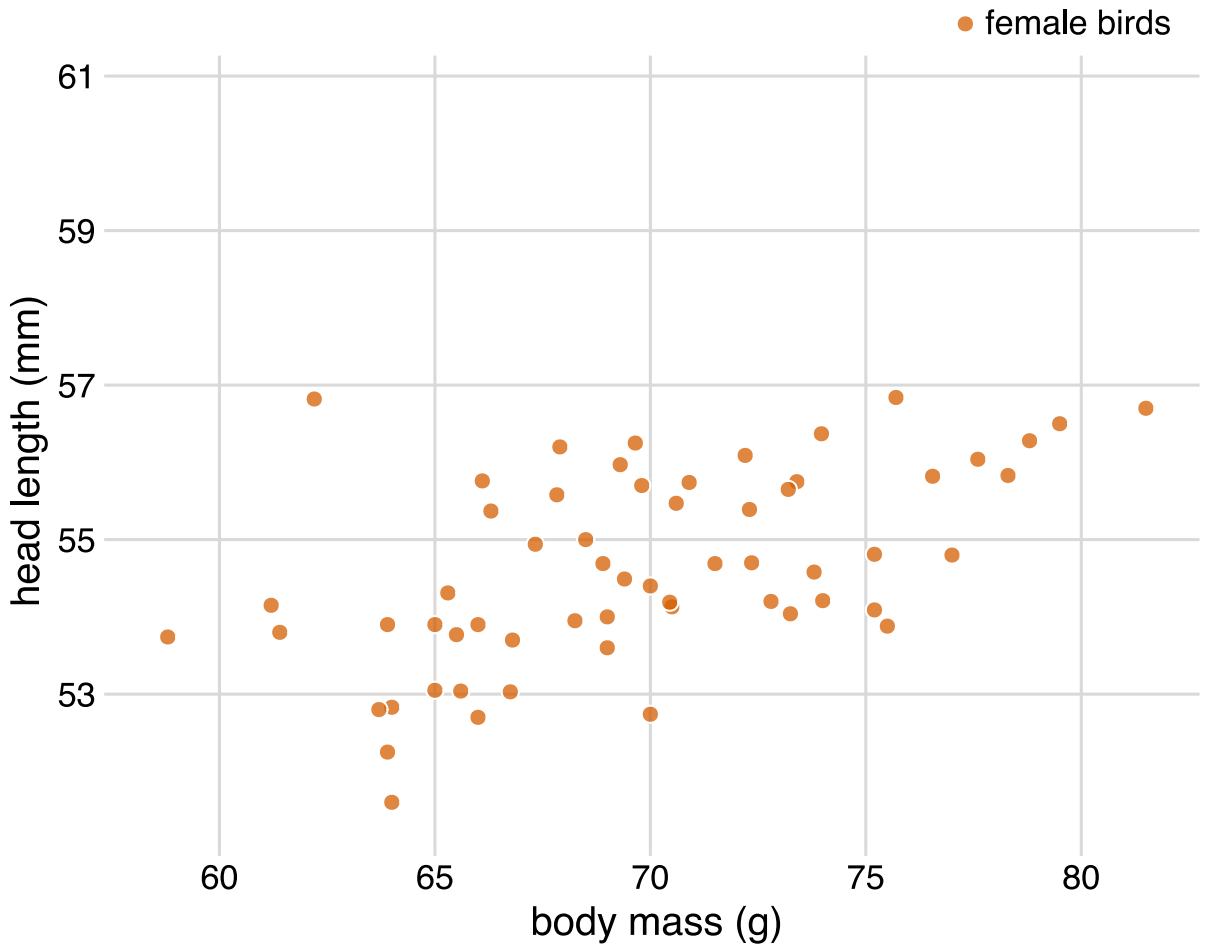
We visualize linear trends with regression lines



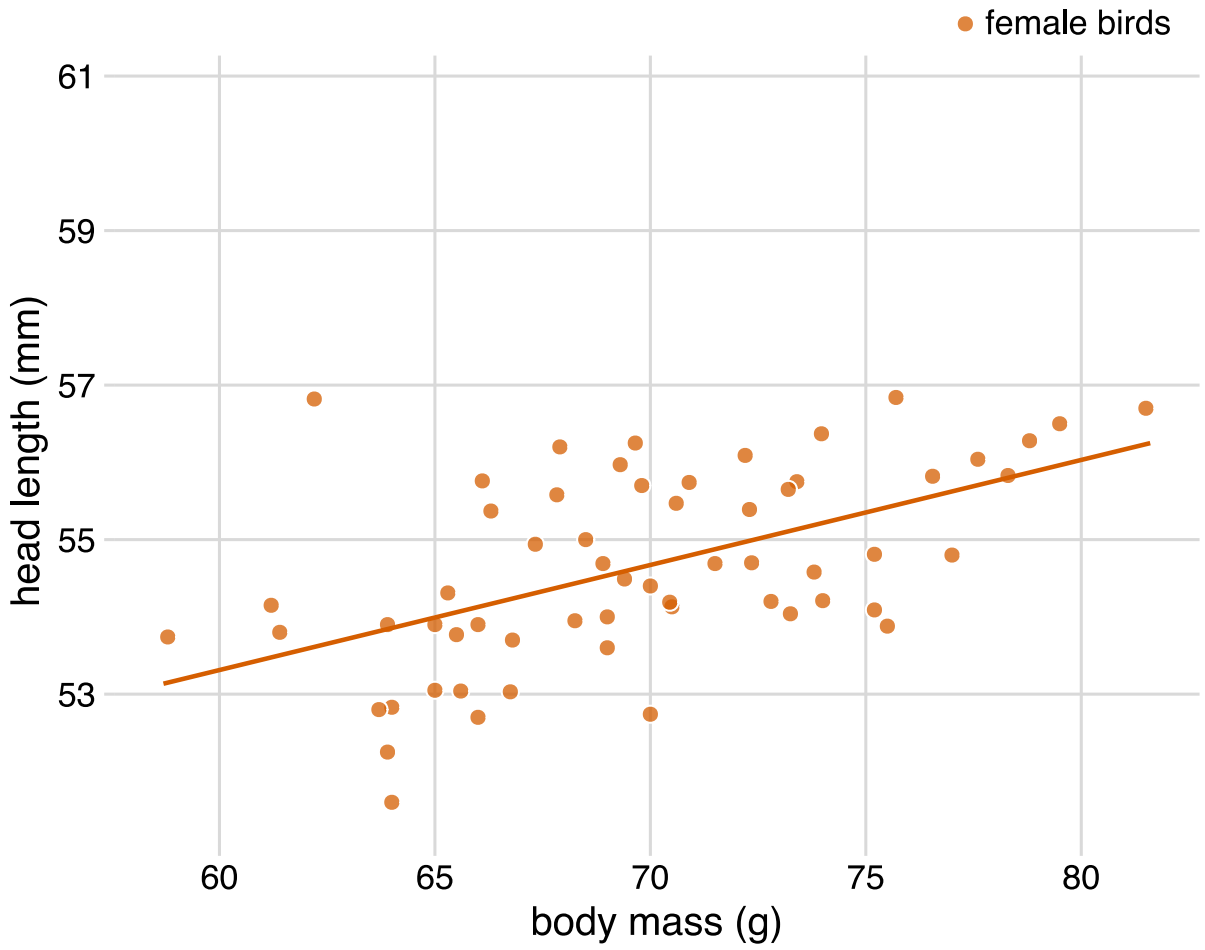
We visualize linear trends with regression lines



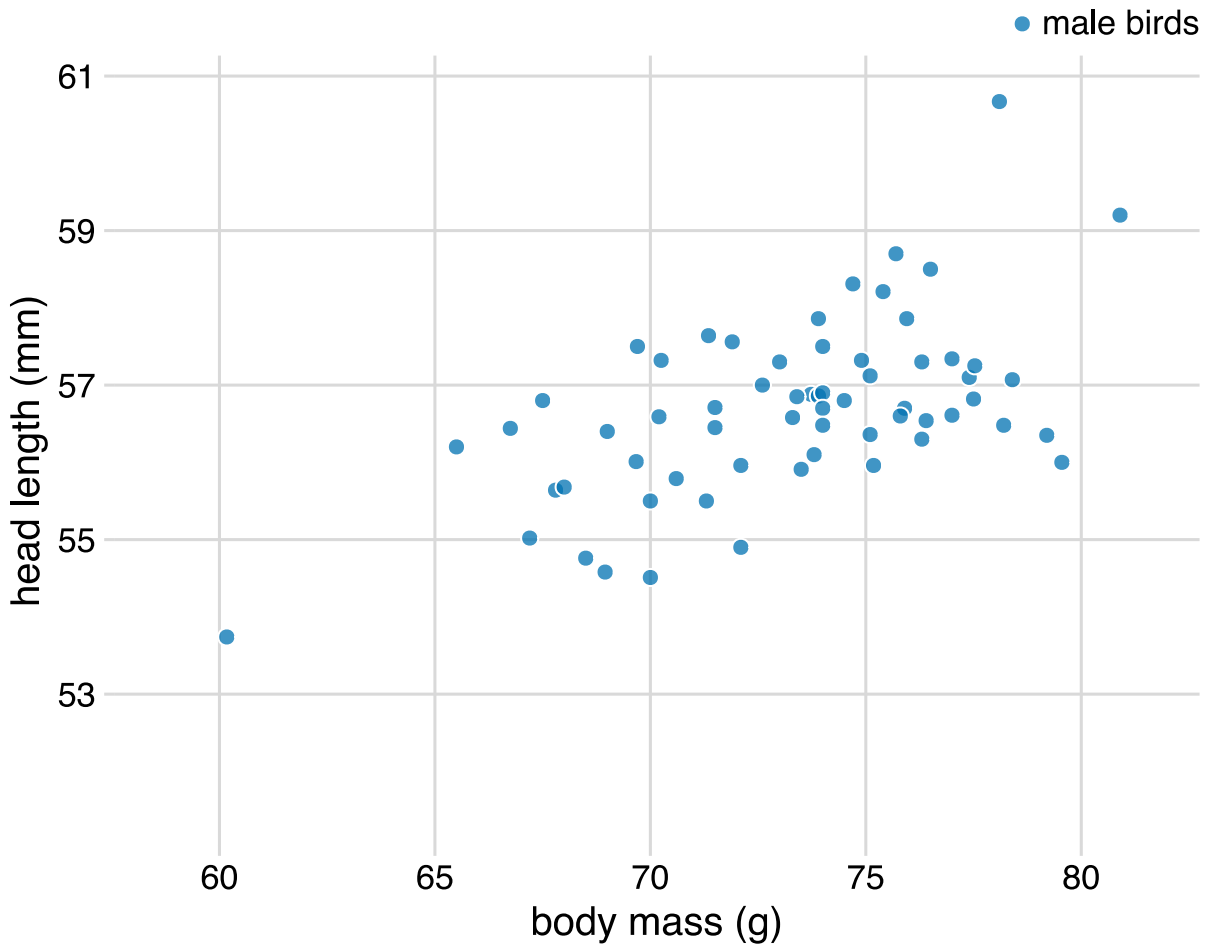
We visualize linear trends with regression lines



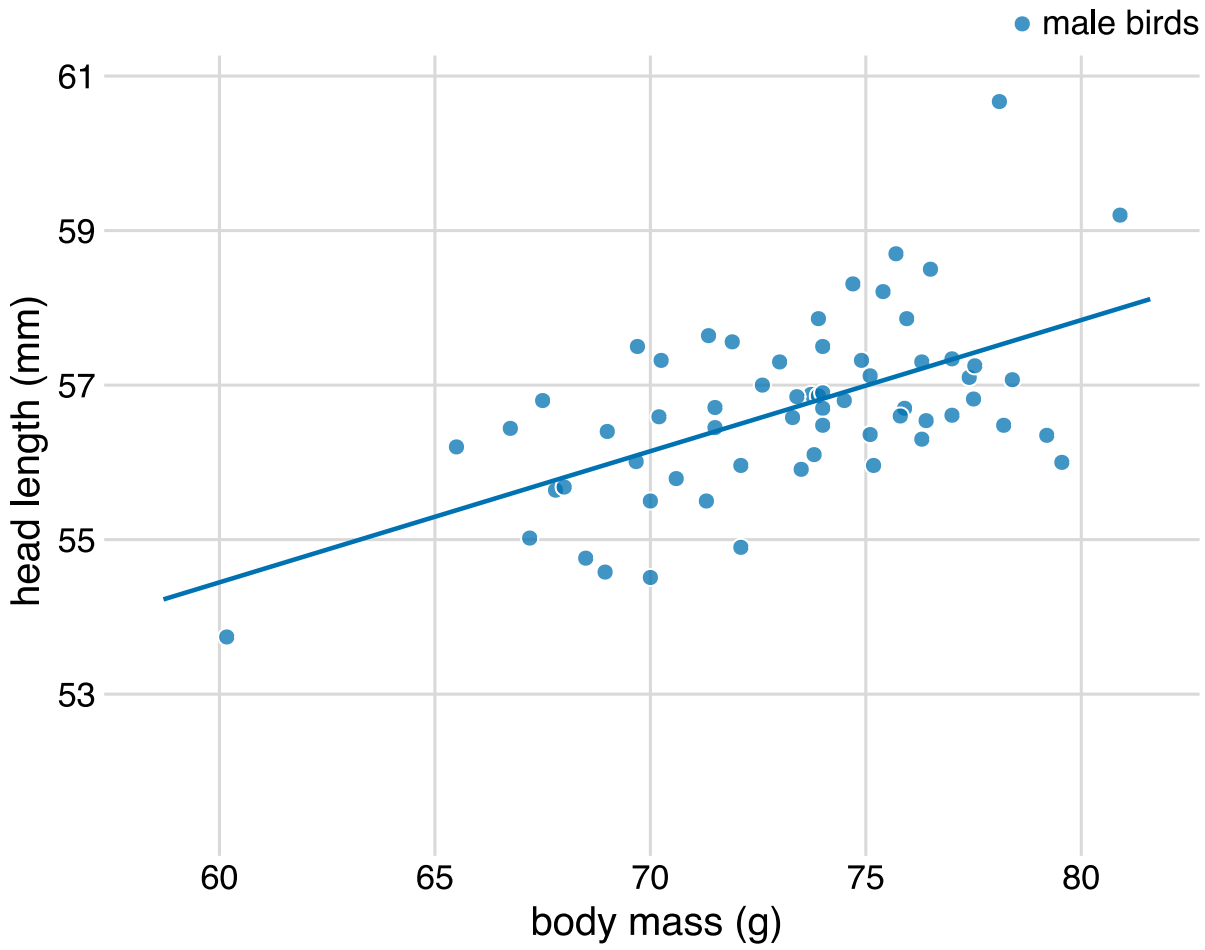
We visualize linear trends with regression lines



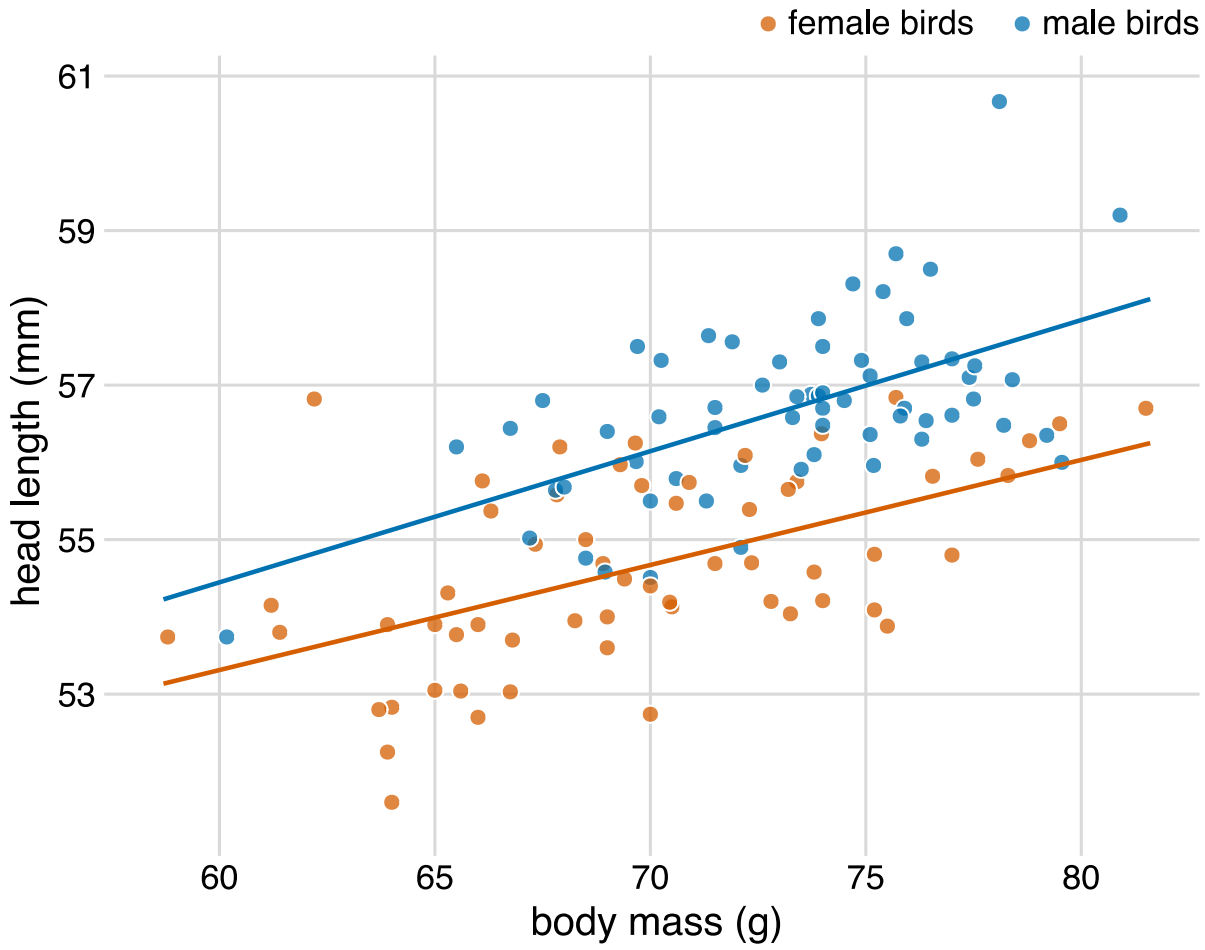
We visualize linear trends with regression lines



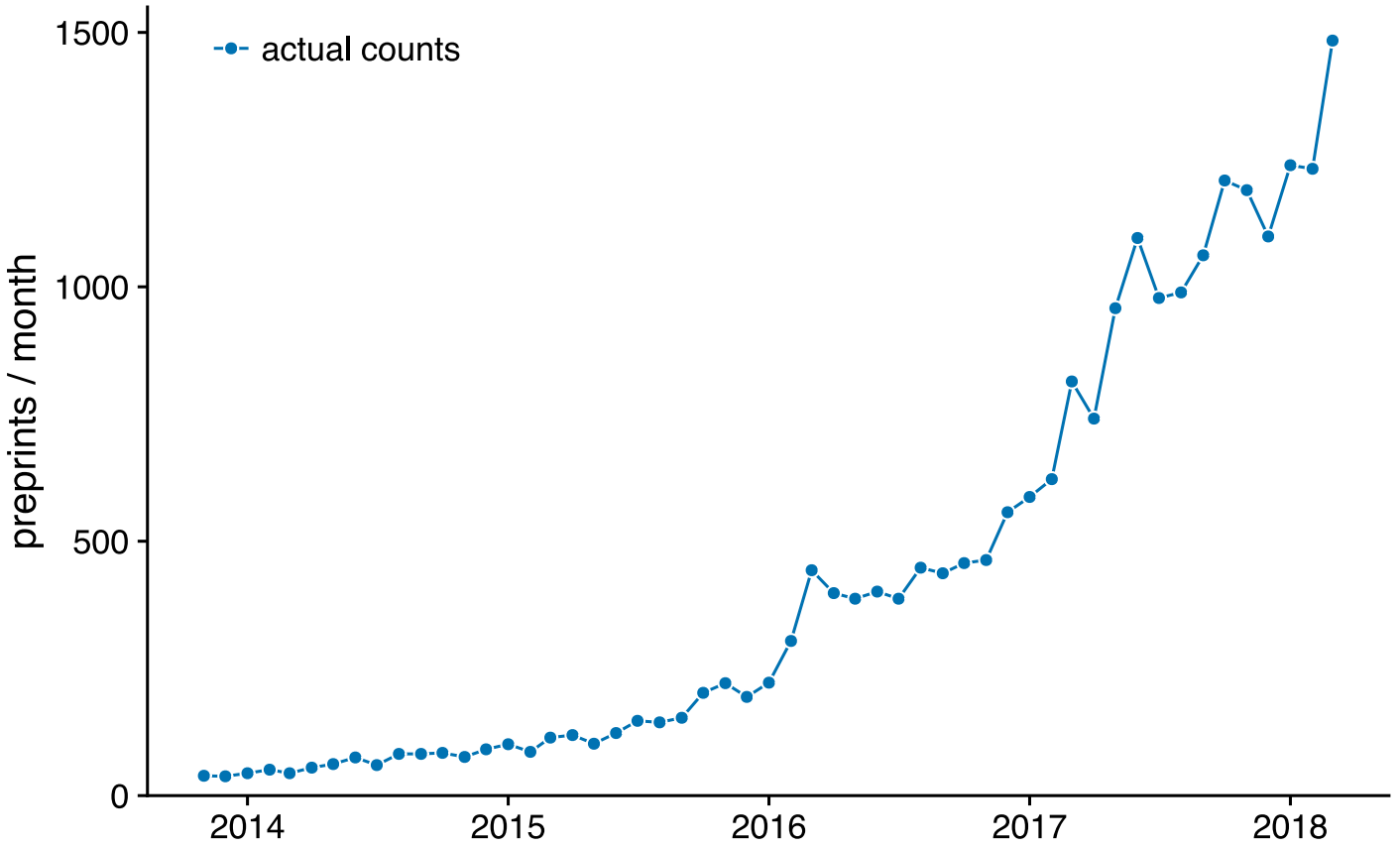
We visualize linear trends with regression lines



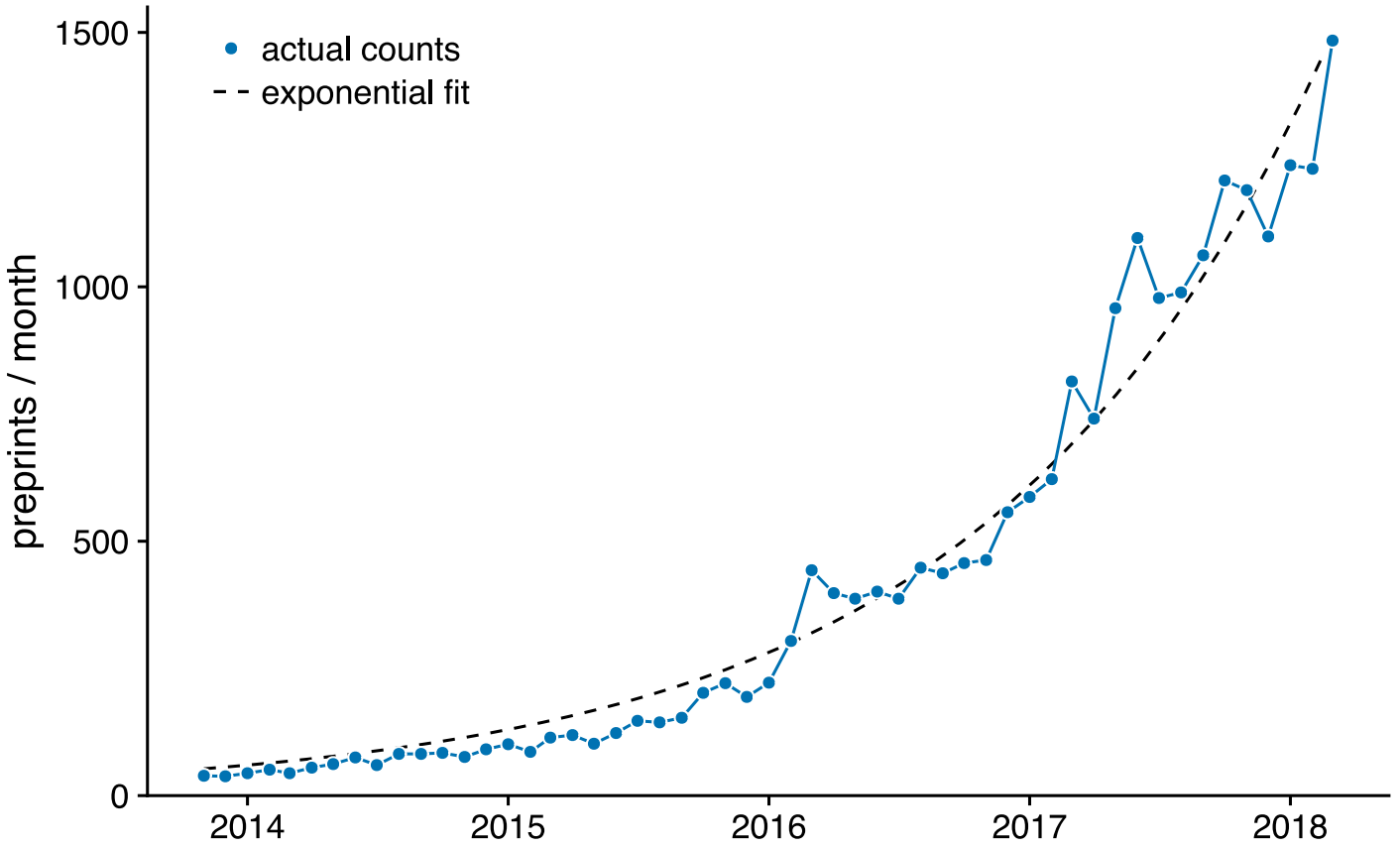
We visualize linear trends with regression lines



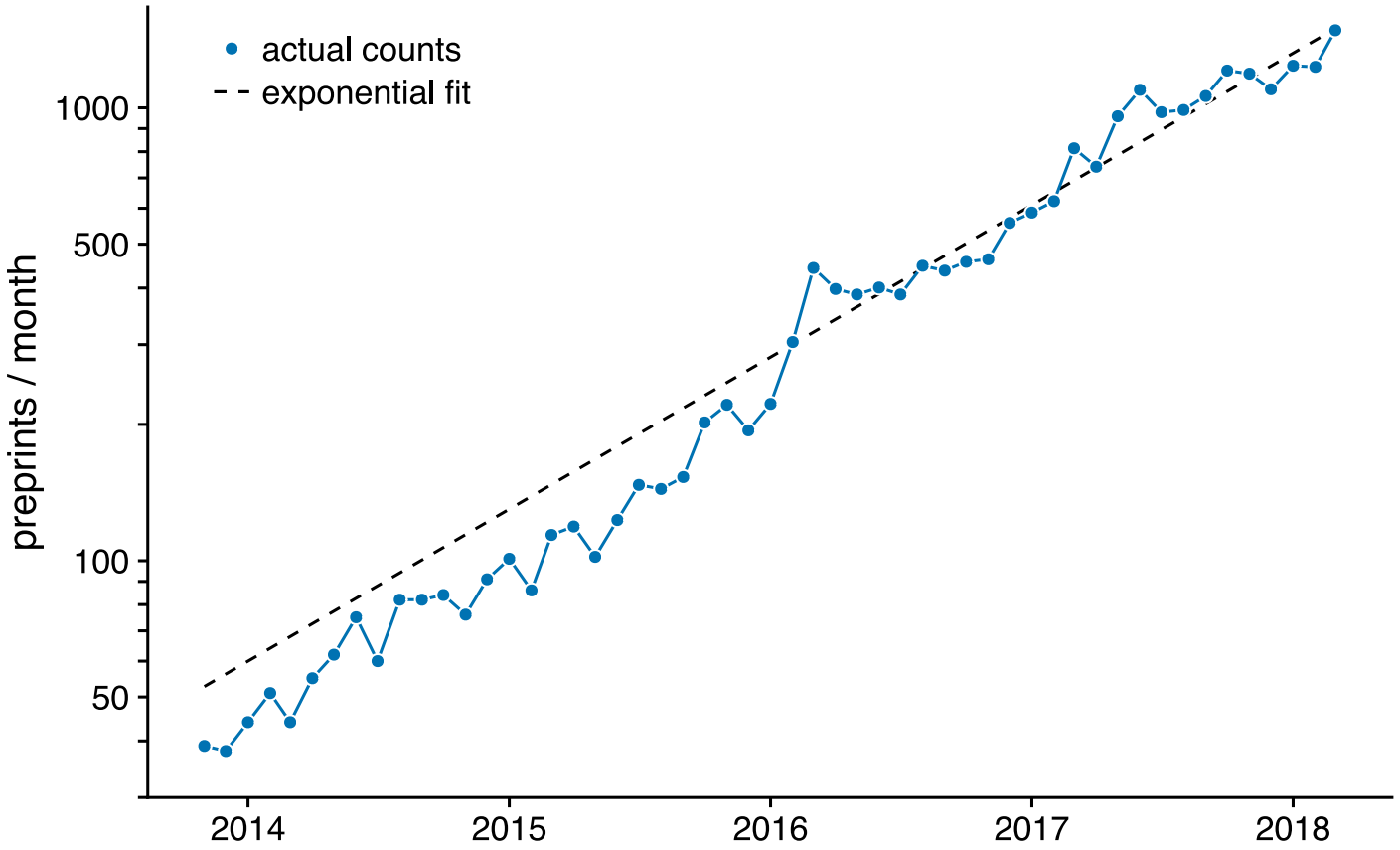
Exponential trends are linear trends on a log scale



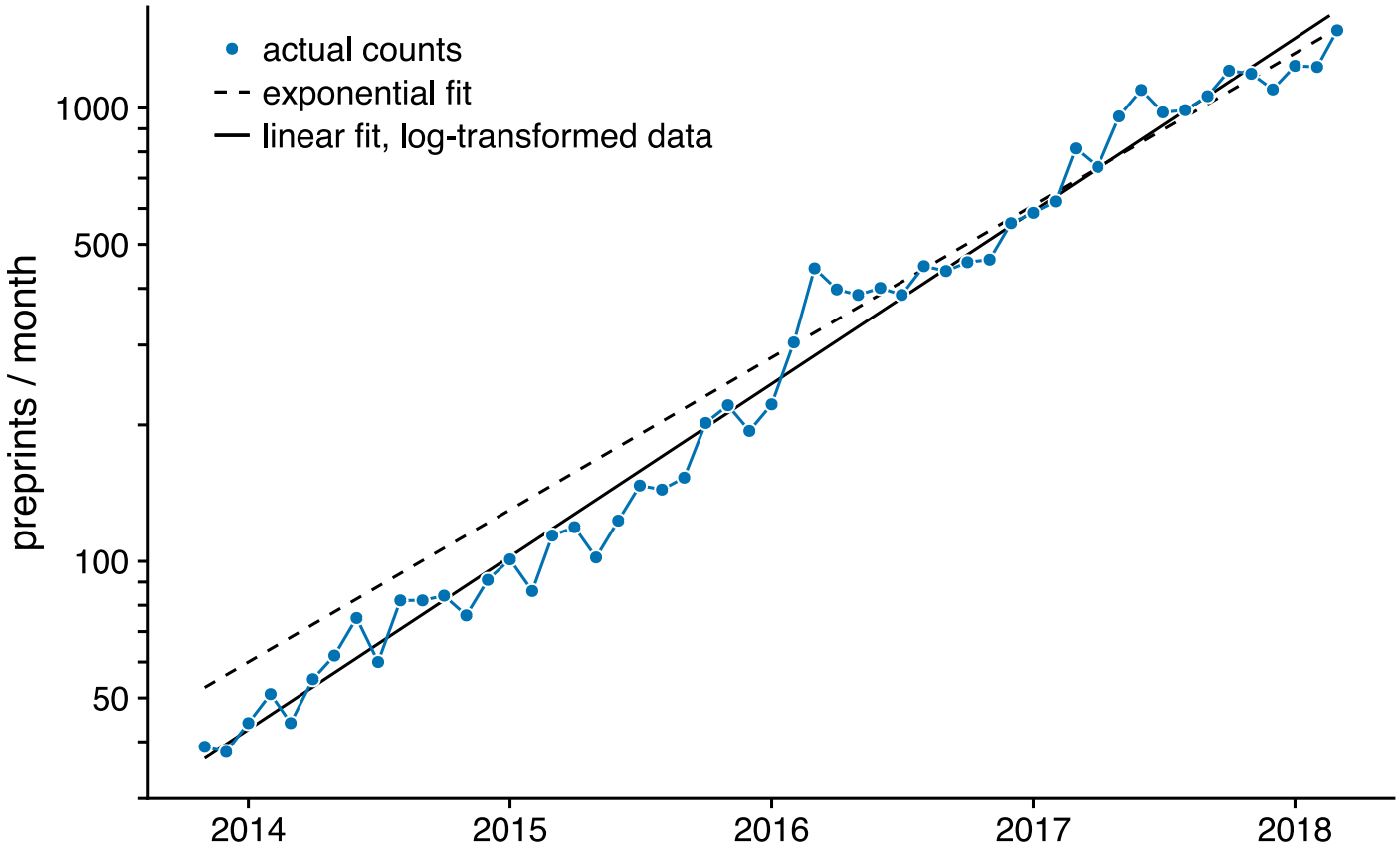
Exponential trends are linear trends on a log scale



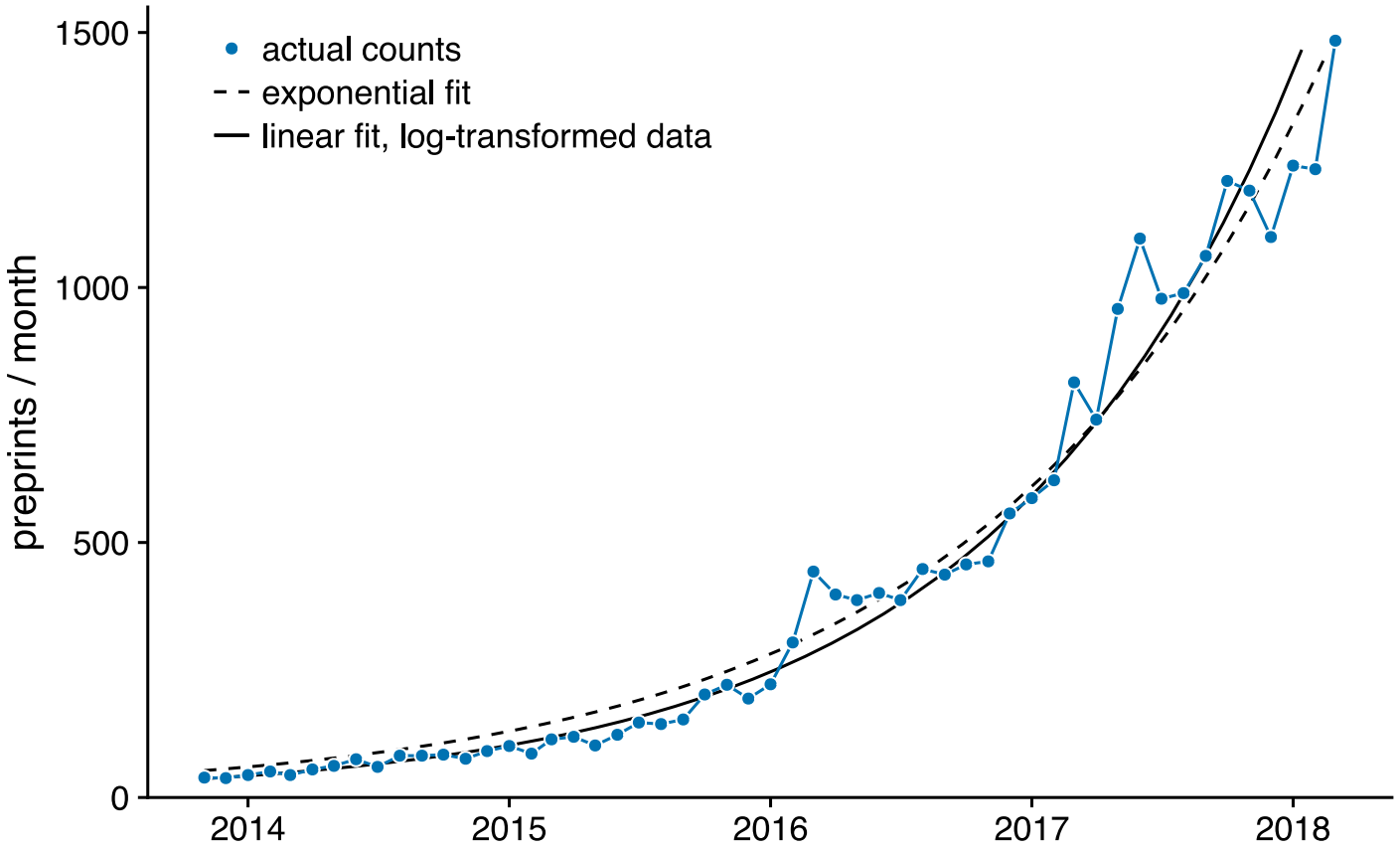
Exponential trends are linear trends on a log scale



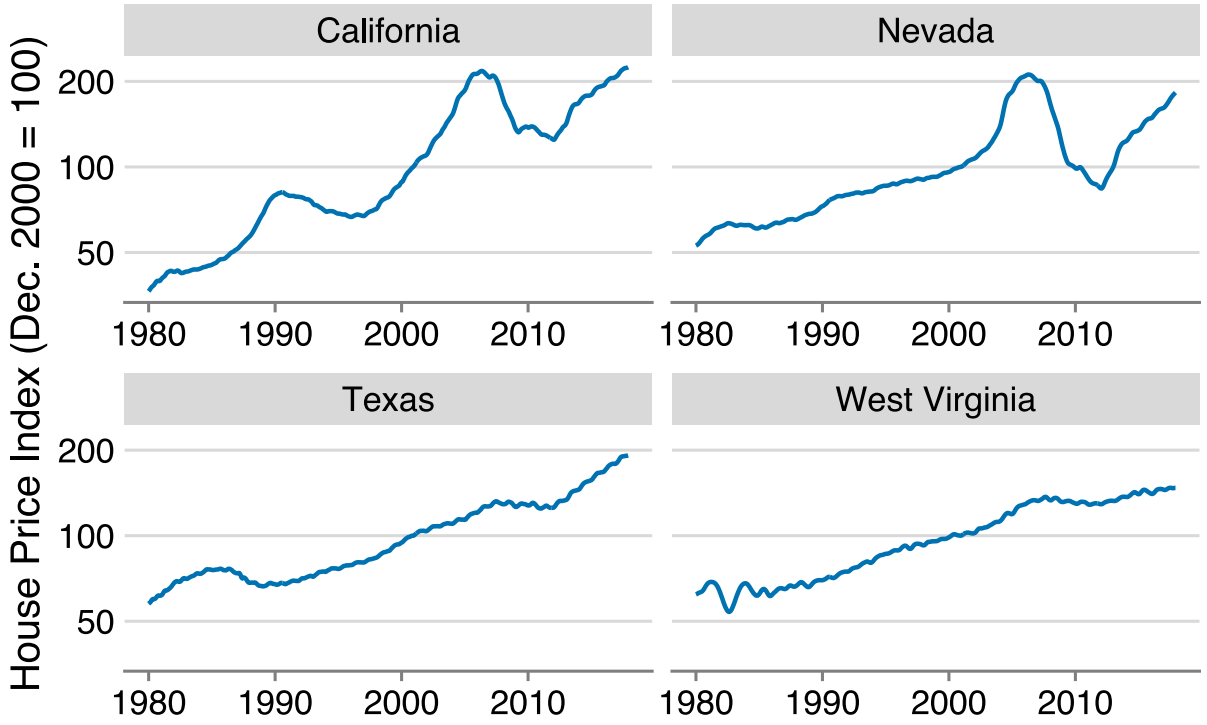
Exponential trends are linear trends on a log scale



Exponential trends are linear trends on a log scale



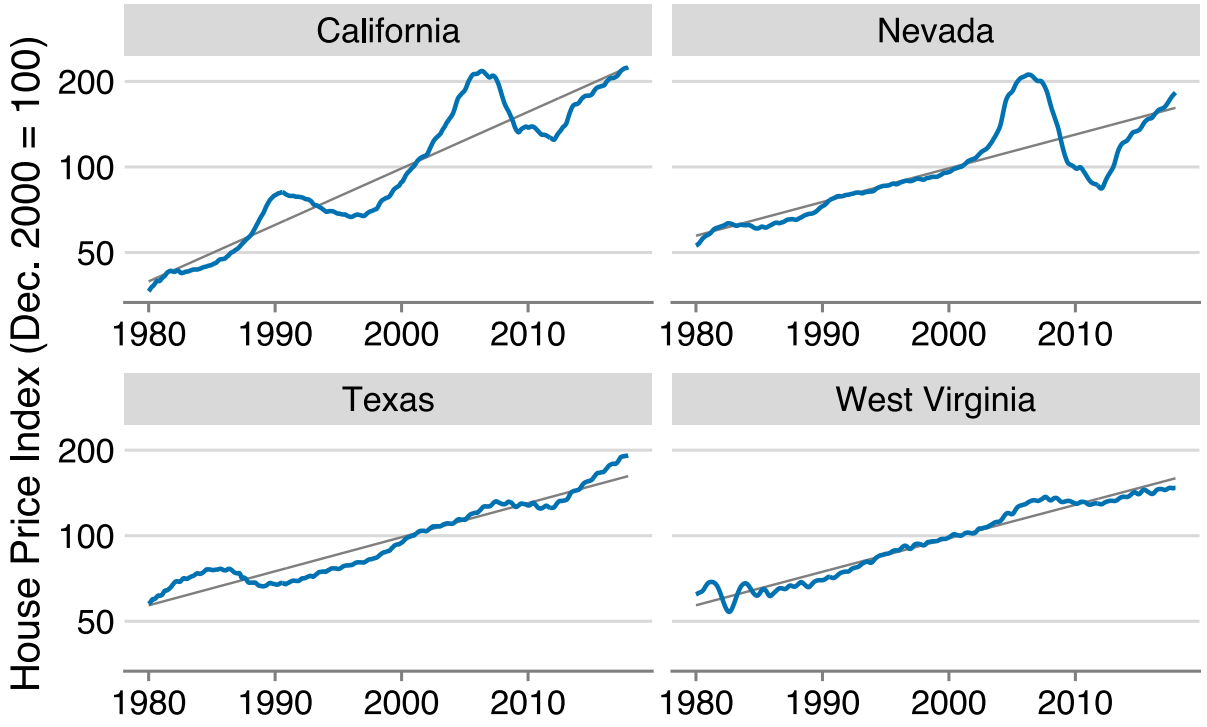
Detrending: Removing the underlying trend



Did housing prices in California decline substantially from 1990 to 1998?

Did housing prices in West Virginia recover by 2017?

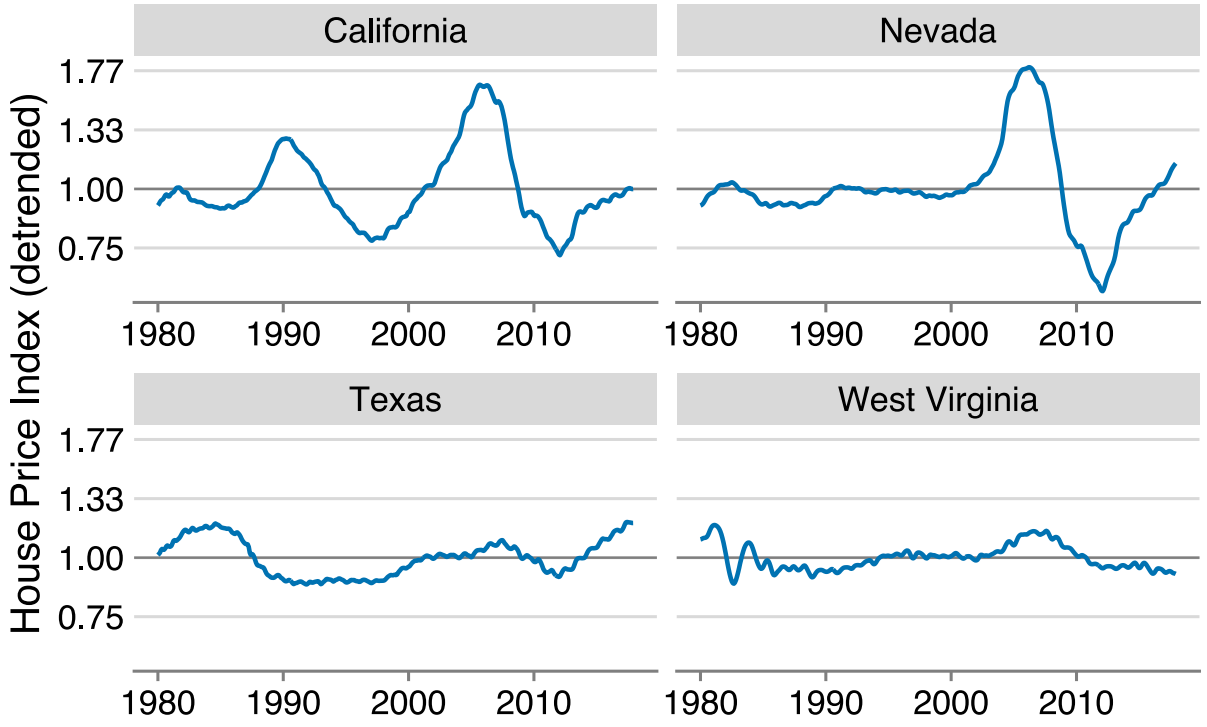
Detrending: Removing the underlying trend



Did housing prices in California decline substantially from 1990 to 1998?

Did housing prices in West Virginia recover by 2017?

Detrending: Removing the underlying trend

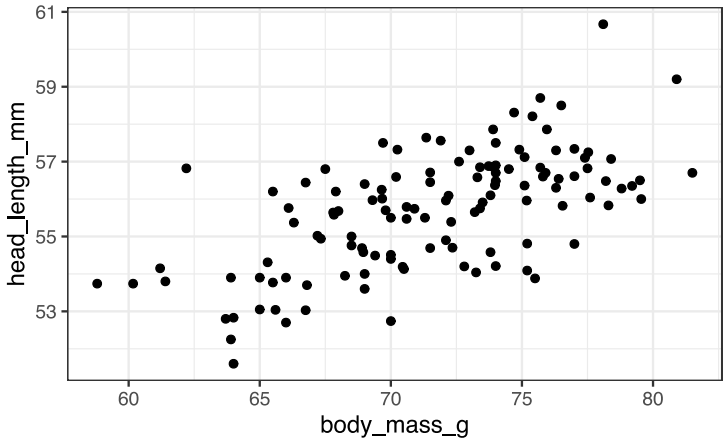


Did housing prices in California decline substantially from 1990 to 1998? — yes

Creating trendlines in ggplot2

We add trend lines with `geom_smooth()`

```
ggplot(blue_jays) +
  aes(body_mass_g, head_length_mm
  geom_point() + theme_bw(14)
```

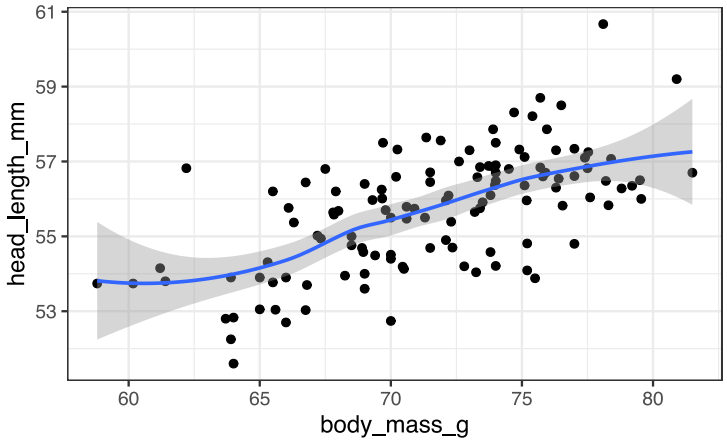


Scatter plot only

We add trend lines with `geom_smooth()`

```
ggplot(blue_jays) +
  aes(body_mass_g, head_length_mm
  geom_point() + theme_bw(14) +
  geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and form
```

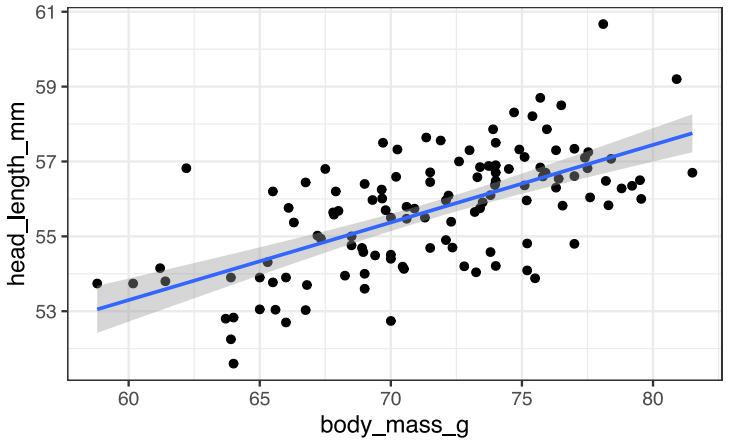


Scatter plot with loess smooth

We add trend lines with `geom_smooth()`

```
ggplot(blue_jays) +
  aes(body_mass_g, head_length_mm
  geom_point() + theme_bw(14) +
  geom_smooth(
    # smooth using linear model
    method = "lm"
  )
```

```
`geom_smooth()` using formula 'y ~ x'
```

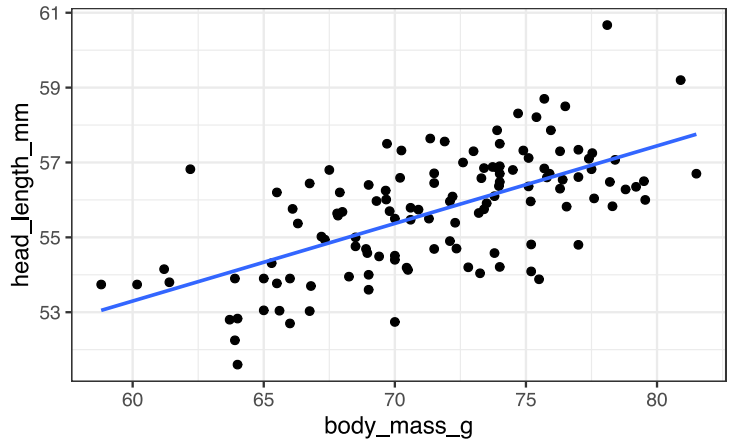


Scatter plot with linear regression

We add trend lines with `geom_smooth()`

```
ggplot(blue_jays) +
  aes(body_mass_g, head_length_mm
  geom_point() + theme_bw(14) +
  geom_smooth(
    # smooth using linear model
    method = "lm",
    # suppress confidence band
    se = FALSE
  )
```

```
`geom_smooth()` using formula 'y ~ x'
```

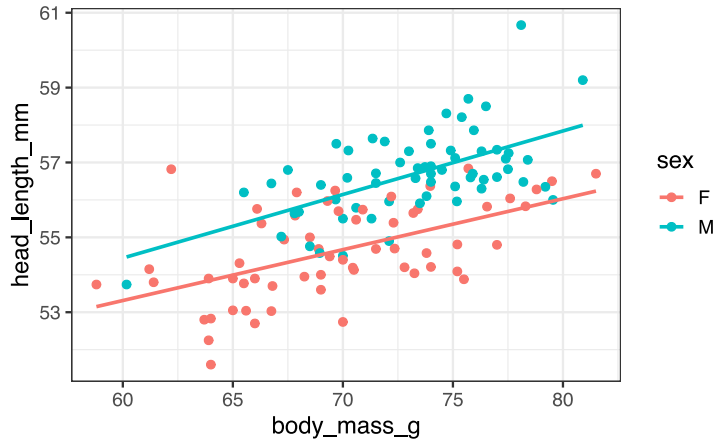


Scatter plot with linear regression, no confidence band

We add trend lines with `geom_smooth()`

```
ggplot(blue_jays) +
  aes(
    body_mass_g, head_length_mm,
    color = sex
  ) +
  geom_point() + theme_bw(14) +
  geom_smooth(
    # smooth using linear model
    method = "lm",
    # suppress confidence band
    se = FALSE
  )
```

```
`geom_smooth()` using formula 'y ~ x'
```



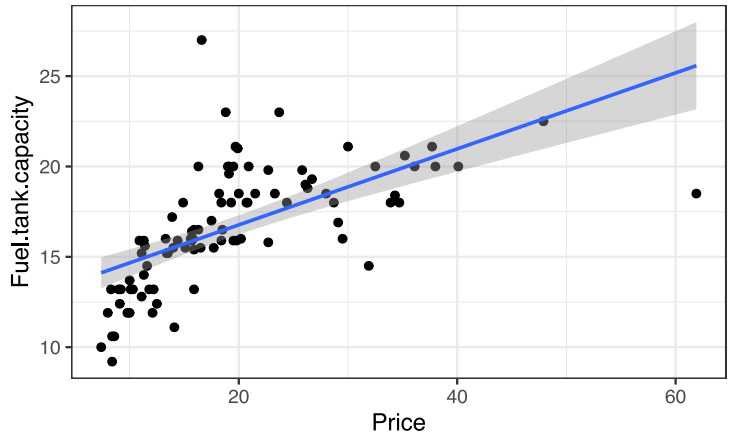
Scatter plot with linear regression by sex

Linear regression can be nonsensical

Example: Fuel-tank capacity versus price in cars

```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  geom_smooth(method = "lm")
```

```
`geom_smooth()` using formula 'y ~ x'
```

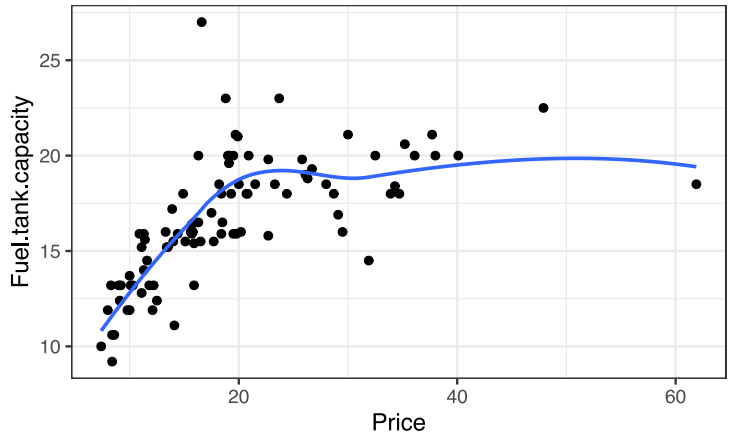


Do more expensive cars have a larger fuel tank?

Example: Fuel-tank capacity versus price in cars

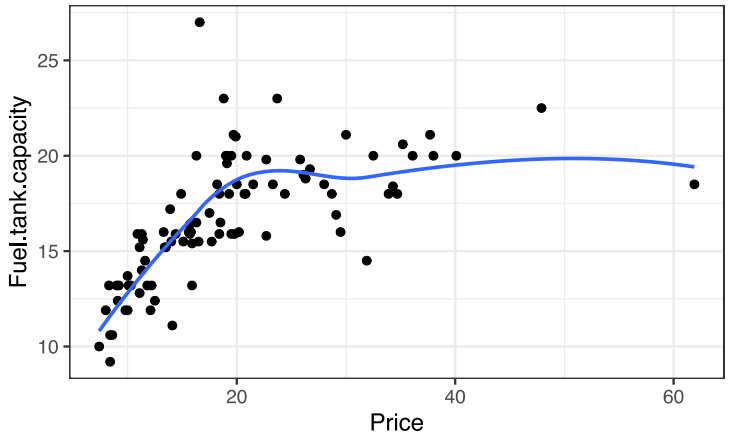
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # default: loess smoothing
  geom_smooth(
    se = FALSE
  )
```

```
`geom_smooth()` using method = 'loess' and form
```



Example: Fuel-tank capacity versus price in cars

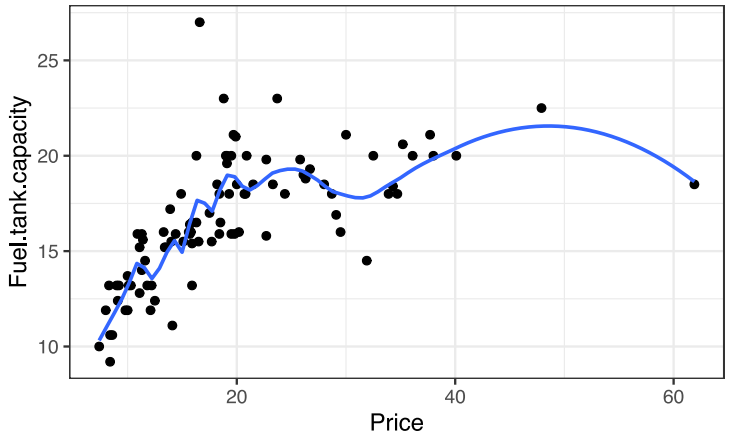
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # loess smoothing
  geom_smooth(
    se = FALSE,
    method = "loess",
    formula = y ~ x
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

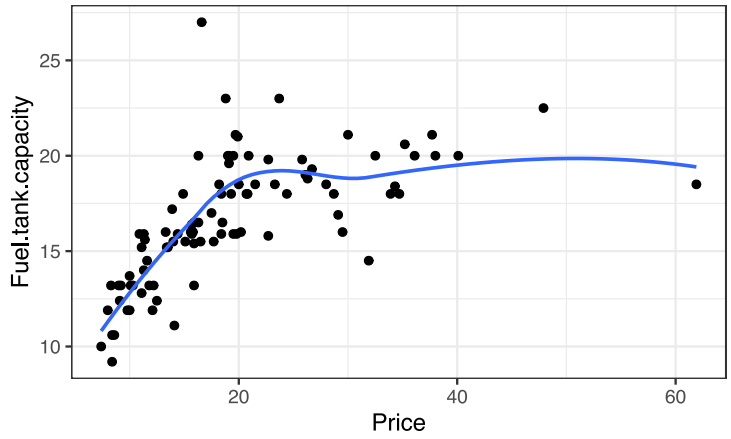
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # loess smoothing
  geom_smooth(
    se = FALSE,
    method = "loess",
    formula = y ~ x,
    span = 0.25
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

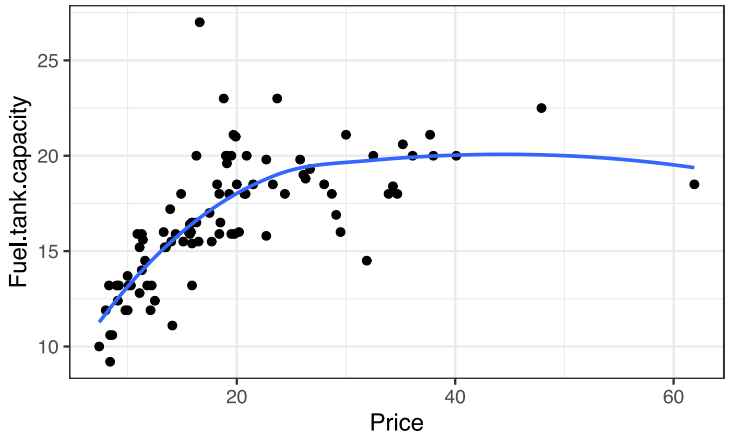
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # loess smoothing
  geom_smooth(
    se = FALSE,
    method = "loess",
    formula = y ~ x,
    span = 0.75 # default value
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

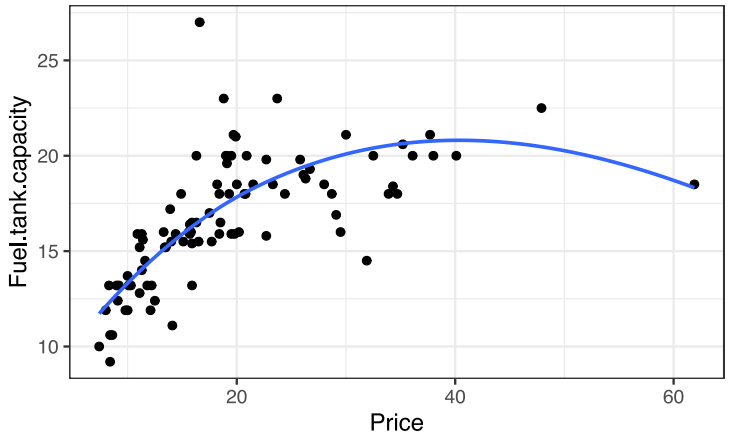
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # loess smoothing
  geom_smooth(
    se = FALSE,
    method = "loess",
    formula = y ~ x,
    span = 1.0
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

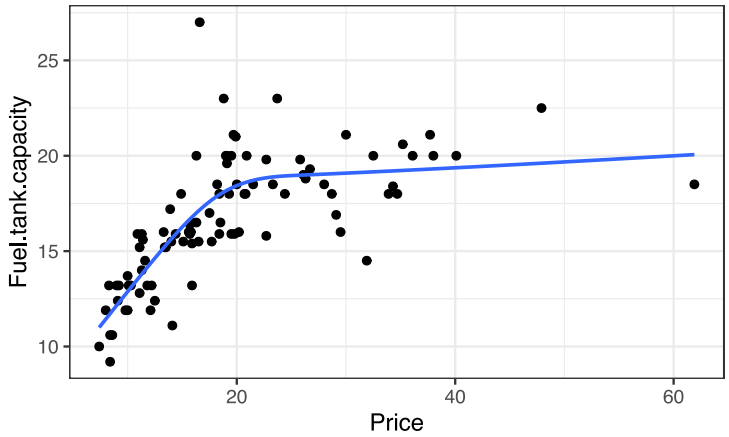
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # loess smoothing
  geom_smooth(
    se = FALSE,
    method = "loess",
    formula = y ~ x,
    span = 1.5
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

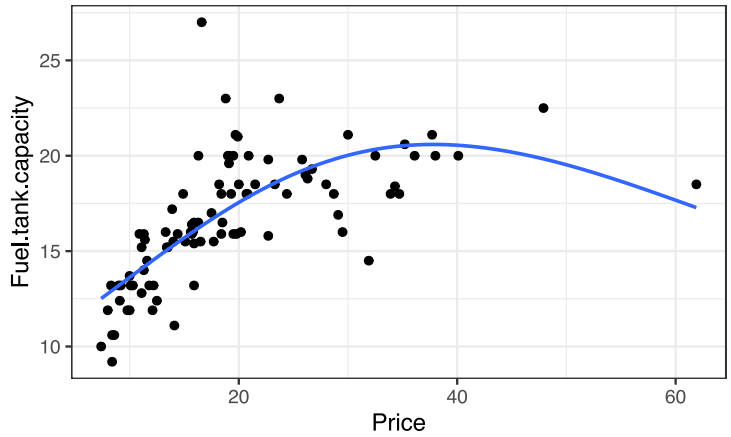
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # cubic spline, 5 knots
  geom_smooth(
    se = FALSE,
    method = "gam",
    formula = y ~ s(x, k = 5, bs
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

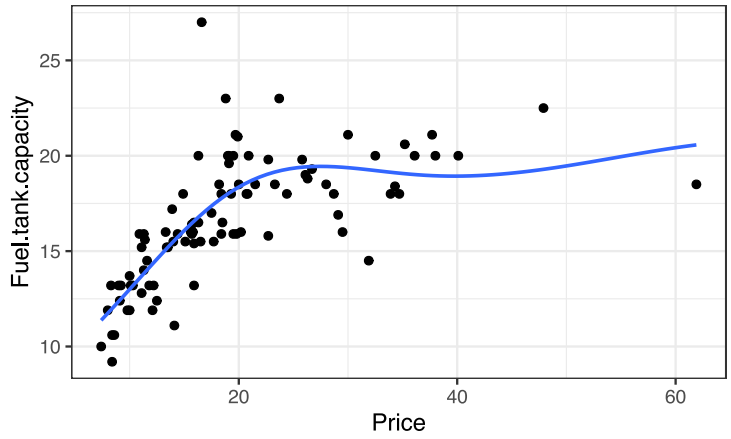
```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # thin-plate spline, 3 knots
  geom_smooth(
    se = FALSE,
    method = "gam",
    formula = y ~ s(x, k = 3)
  )
```



Caution: Exact shape of smoothing line depends on method details

Example: Fuel-tank capacity versus price in cars

```
ggplot(cars93) +
  aes(x = Price, y = Fuel.tank.ca
  geom_point() + theme_bw(14) +
  # Gaussian process spline, 6 kn
  geom_smooth(
    se = FALSE,
    method = "gam",
    formula = y ~ s(x, k = 6, bs
  )
```



Caution: Exact shape of smoothing line depends on method details

Smoothing lines are particularly unreliable near their endpoints

Further reading

- Fundamentals of Data Visualization: Chapter 14: Visualizing trends
- Data Visualization—A Practical Introduction: Chapter 6: Work with models
- **ggplot2** reference documentation: `geom_smooth()`
- **mgcv** reference documentation (for gam smoothing): pdf document