Lobster

Gao

2023-10-12

Import libraries

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidymodels)
```

```
## -- Attaching packages --------------------------------------- tidymodels 1.1.1 --
## v broom        1.0.5     v rsample      1.2.0
## v dials        1.2.0     v tune         1.1.2
## v infer        1.0.5     v workflows    1.1.3
## v modeldata    1.2.0     v workflowsets 1.0.1
## v parsnip      1.1.1     v yardstick    1.2.0
## v recipes      1.0.8
## -- Conflicts ------------------------------------------ tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Learn how to get started at https://www.tidymodels.org/start/
```

```
library(ggforce)
library(yardstick)
```

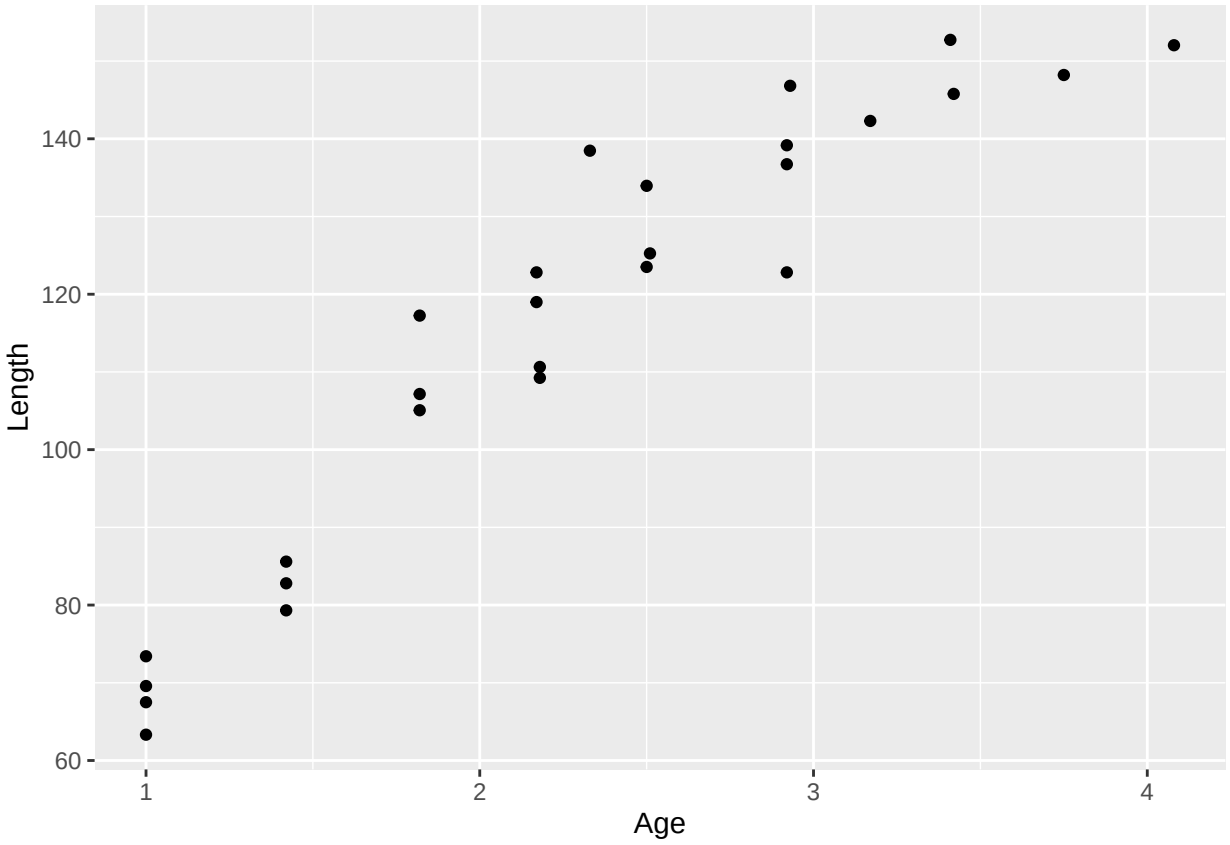
Import Lobsters data

```
lobsters <- read_csv("Lobsters.csv") %>% as_tibble()
```

```
## Rows: 27 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (2): Length, Age
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Create a plot

```
ggplot(data = lobsters) + geom_point(aes(x = Age, y = Length))
```



Create a linear regression model

```
model <- lm(Length ~ Age, data = lobsters)
model
```

```
##
## Call:
## lm(formula = Length ~ Age, data = lobsters)
##
## Coefficients:
## (Intercept)          Age
##       45.76        30.51
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = Length ~ Age, data = lobsters)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.1991  -6.4895  -0.1763   5.0956  21.6209
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   45.764      5.299   8.636 5.67e-09 ***
## Age           30.509      2.168  14.075 2.19e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.699 on 25 degrees of freedom
## Multiple R-squared:  0.8879, Adjusted R-squared:  0.8835
## F-statistic: 198.1 on 1 and 25 DF,  p-value: 2.19e-13
```

```r
lobsters$sqrtLength <- lobsters$Length^0.5

model2 <- lm(sqrtLength ~ Age, data = lobsters)
model2
```

```
##
## Call:
## lm(formula = sqrtLength ~ Age, data = lobsters)
##
## Coefficients:
## (Intercept)          Age
##       7.295        1.473
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = sqrtLength ~ Age, data = lobsters)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97240 -0.35598  0.01368  0.32635  1.04156
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.2946     0.2783   26.21  < 2e-16 ***
## Age           1.4726     0.1138   12.94 1.41e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5093 on 25 degrees of freedom
```

```
## Multiple R-squared:  0.8701, Adjusted R-squared:  0.8649
## F-statistic: 167.4 on 1 and 25 DF,  p-value: 1.408e-12
```

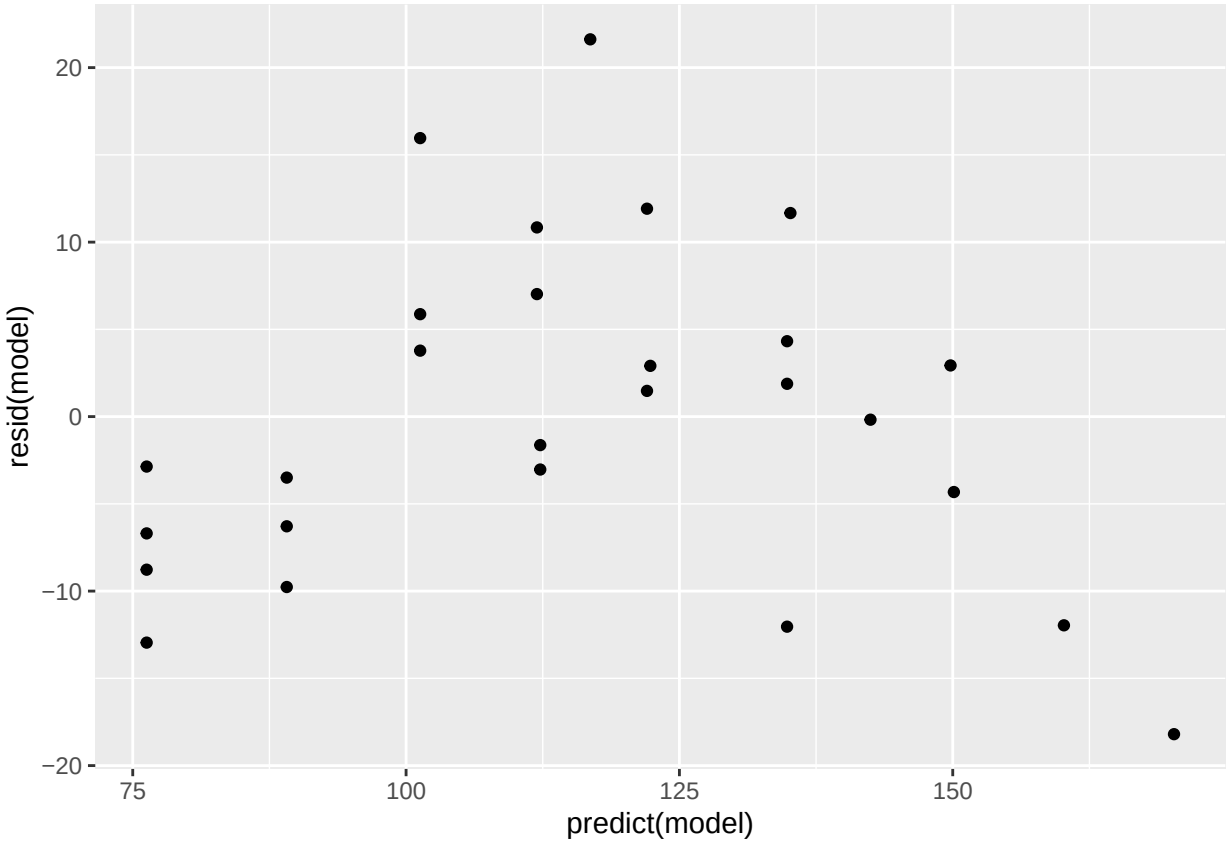
```
lobsters$Length2 <- lobsters$Length^2

model3 <- lm(Length2 ~ Age, data = lobsters)
model3
```

```
##
## Call:
## lm(formula = Length2 ~ Age, data = lobsters)
##
## Coefficients:
## (Intercept)          Age
##       -1329         6758
```

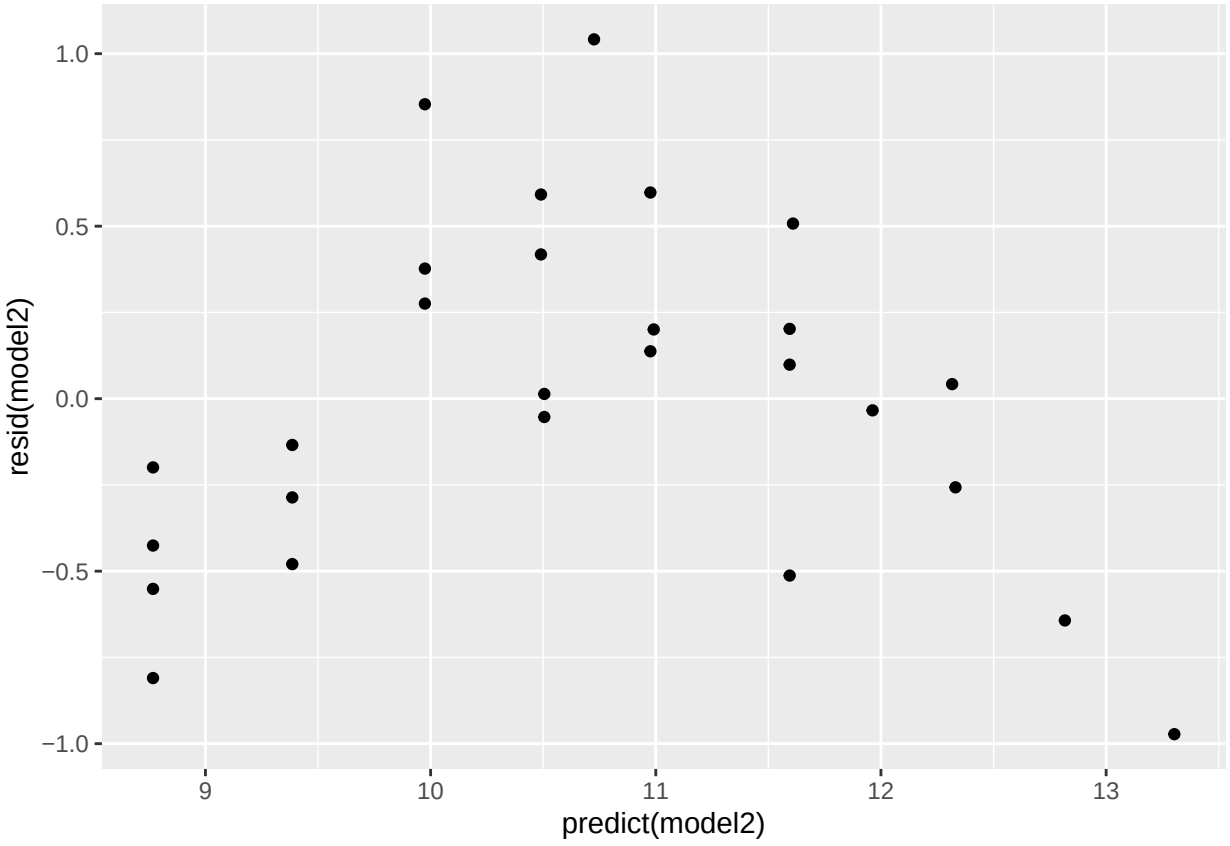
```
summary(model3)
```

```
##
## Call:
## lm(formula = Length2 ~ Age, data = lobsters)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3321.9 -1287.0   -40.2   893.4  4756.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1328.5     1047.8  -1.268    0.216
## Age           6757.8      428.6  15.769 1.69e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1918 on 25 degrees of freedom
## Multiple R-squared:  0.9086, Adjusted R-squared:  0.905
## F-statistic: 248.6 on 1 and 25 DF,  p-value: 1.687e-14
```

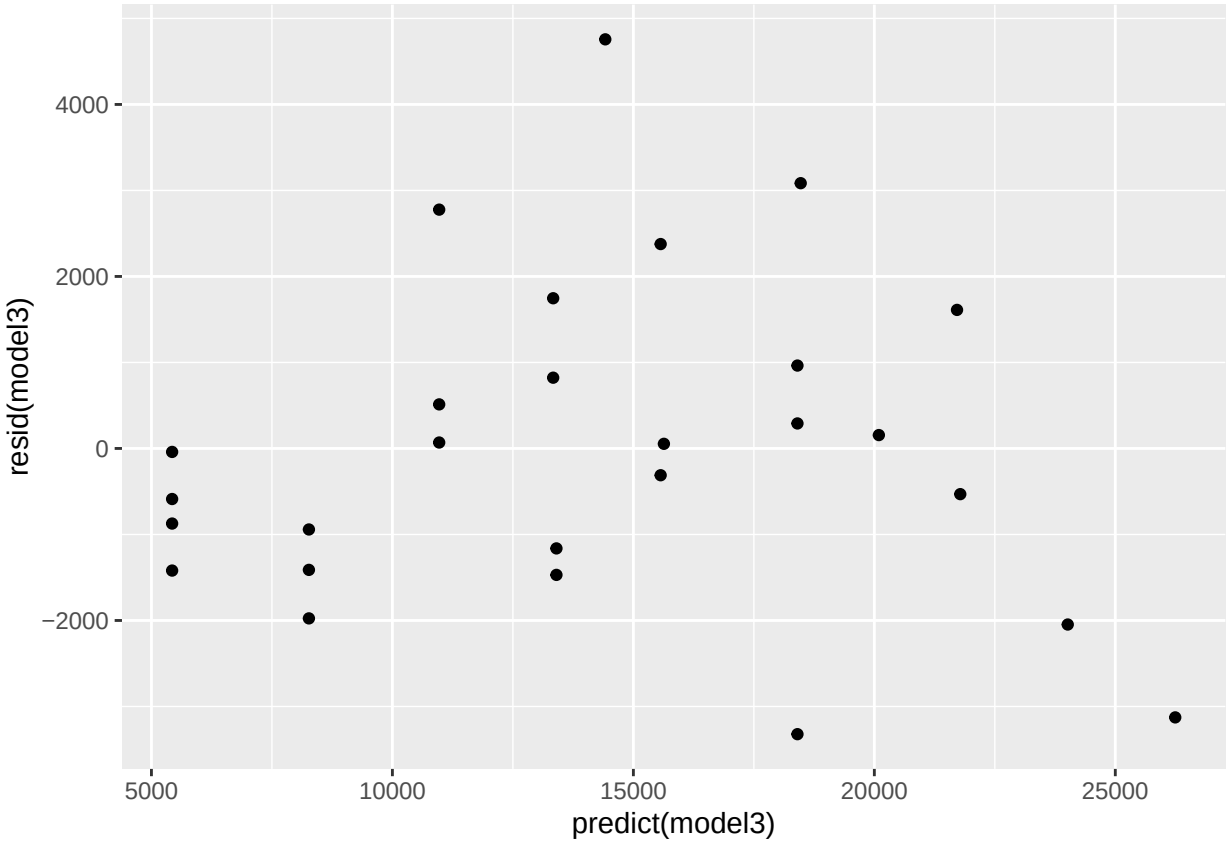
```
ggplot(lobsters) + geom_point(aes(x=predict(model), y=resid(model)))
```




```
ggplot(lobsters) + geom_point(aes(x=predict(model2), y=resid(model2)))
```




```
ggplot(lobsters) + geom_point(aes(x=predict(model3), y=resid(model3)))
```


4000
2000
0
−2000
resid(model3)
5000
10000
15000
20000
25000
predict(model3)