Rafting

Gao

2023-10-11

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()
## * Use suppressPackageStartupMessages() to eliminate package startup messages
```

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

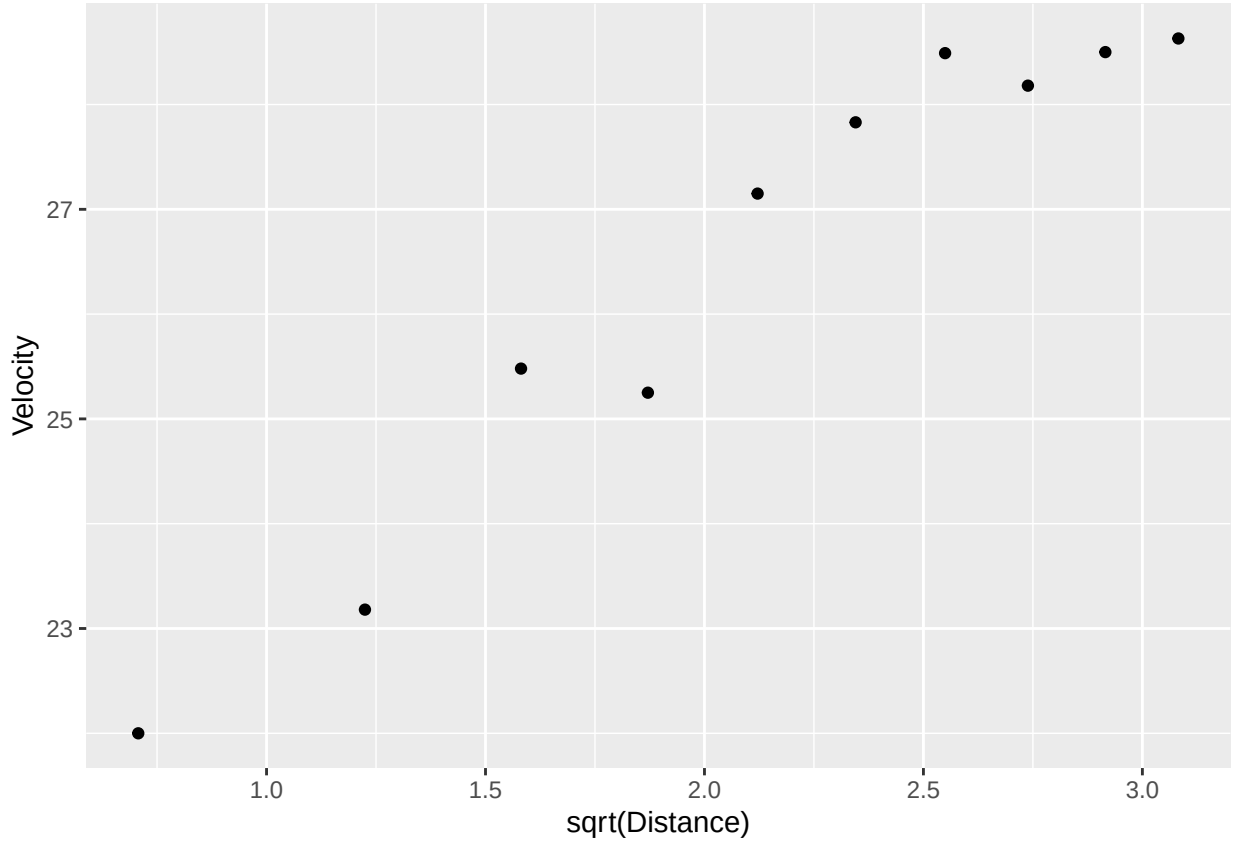
Import the data

```
## [1] "/Users/andrewgao/Documents/GitHub/Advanced-Data-Science/Gao/Unit 3"
```

```
## Rows: 10 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (2): Distance, Velocity
##
## 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 = rafting) + geom_point(aes(x = sqrt(Distance), y = Velocity))
```



Create a linear regression model

```
model <- lm(Velocity ~ Distance, data = rafting)
model
```

```
##
## Call:
## lm(formula = Velocity ~ Distance, data = rafting)
##
## Coefficients:
## (Intercept)     Distance
##     22.8081       0.7322
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = Velocity ~ Distance, data = rafting)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1742 -0.6777 -0.1201  0.9024  1.0471
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.8081     0.6001  38.005 2.52e-10 ***
## Distance      0.7322     0.1041   7.035 0.000109 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9453 on 8 degrees of freedom
## Multiple R-squared:  0.8608, Adjusted R-squared:  0.8435
## F-statistic: 49.49 on 1 and 8 DF,  p-value: 0.0001088
```

```r
rafting$sqrtDistance <- rafting$Distance^0.5

model2 <- lm(Velocity ~ sqrtDistance, data = rafting)
model2
```

```
##
## Call:
## lm(formula = Velocity ~ sqrtDistance, data = rafting)
##
## Coefficients:
##  (Intercept)  sqrtDistance
##       20.110         3.009
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = Velocity ~ sqrtDistance, data = rafting)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7530 -0.4618 -0.2034  0.6466  0.7096
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   20.1102     0.6097   32.99 7.78e-10 ***
## sqrtDistance   3.0085     0.2726   11.03 4.05e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6292 on 8 degrees of freedom
```

```
## Multiple R-squared:  0.9383, Adjusted R-squared:  0.9306
## F-statistic: 121.8 on 1 and 8 DF,  p-value: 4.052e-06
```

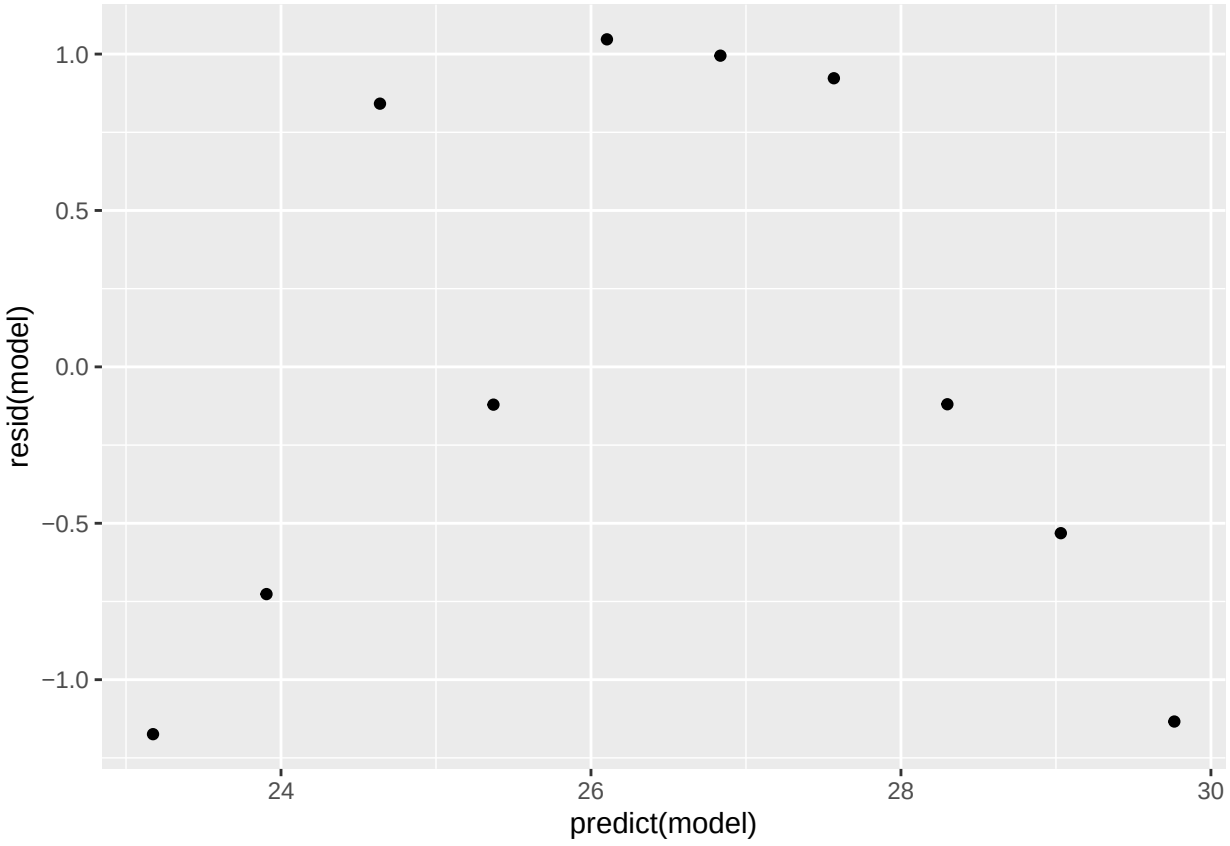
```
rafting$Velocity2 <- rafting$Velocity^2

model3 <- lm(Velocity2 ~ Distance, data = rafting)
model3
```

```
##
## Call:
## lm(formula = Velocity2 ~ Distance, data = rafting)
##
## Coefficients:
## (Intercept)     Distance
##      516.96        37.76
```

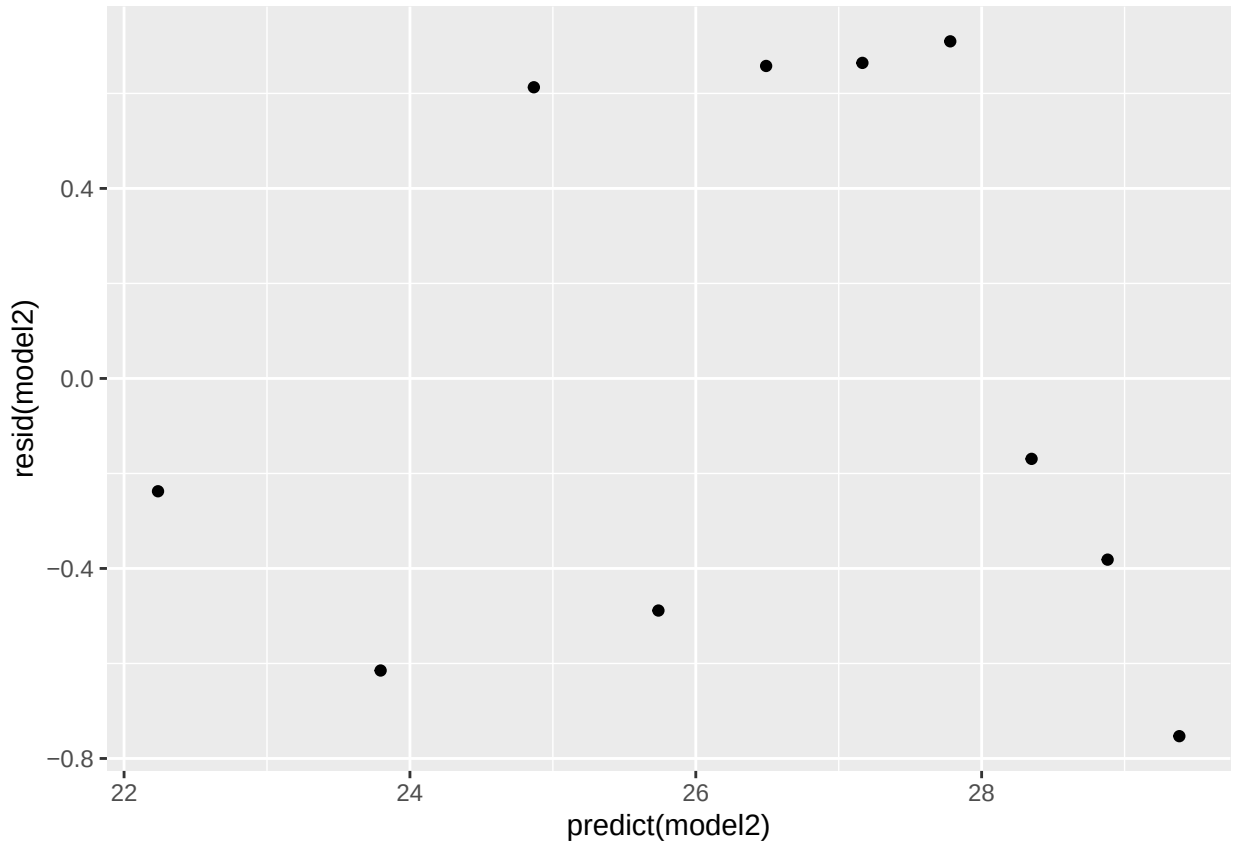
```
summary(model3)
```

```
##
## Call:
## lm(formula = Velocity2 ~ Distance, data = rafting)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -55.972 -33.625  -8.786  46.444  50.255
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  516.964     29.130  17.747 1.04e-07 ***
## Distance      37.756      5.052   7.474 7.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.89 on 8 degrees of freedom
## Multiple R-squared:  0.8747, Adjusted R-squared:  0.8591
## F-statistic: 55.86 on 1 and 8 DF,  p-value: 7.101e-05
```

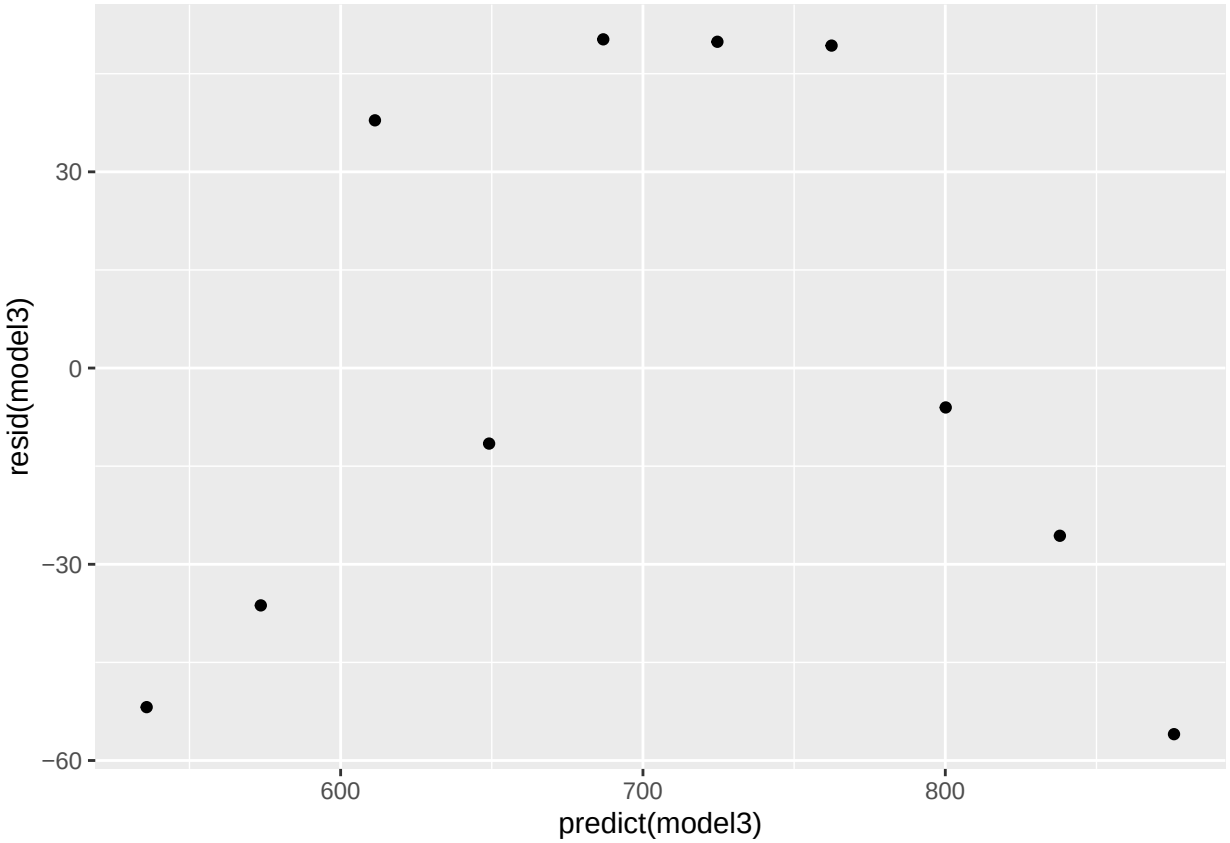
```
ggplot(rafting) + geom_point(aes(x=predict(model), y=resid(model)))
```




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




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


30
0
−30
−60
resid(model3)
600
700
800
predict(model3)