Interpretation of Regression

CORRELATION AND REGRESSION IN R

Ben Baumer

Assistant Professor at Smith College

datacamp

Is that textbook overpriced?

```
head(textbooks)
```

```
  deptAbbr course          ibsn uclaNew amazNew more   diff
1  Am Ind   C170 978-0803272620   27.67   27.95    Y  -0.28
2  Anthro      9 978-0030119194   40.59   31.14    Y   9.45
3  Anthro   135T 978-0300080643   31.68   32.00    Y  -0.32
4  Anthro  191HB 978-0226206813   16.00   11.52    Y   4.48
5  Art His M102K 978-0892365999   18.95   14.21    Y   4.74
6  Art His  118E 978-0394723693   14.95   10.17    Y   4.78
```
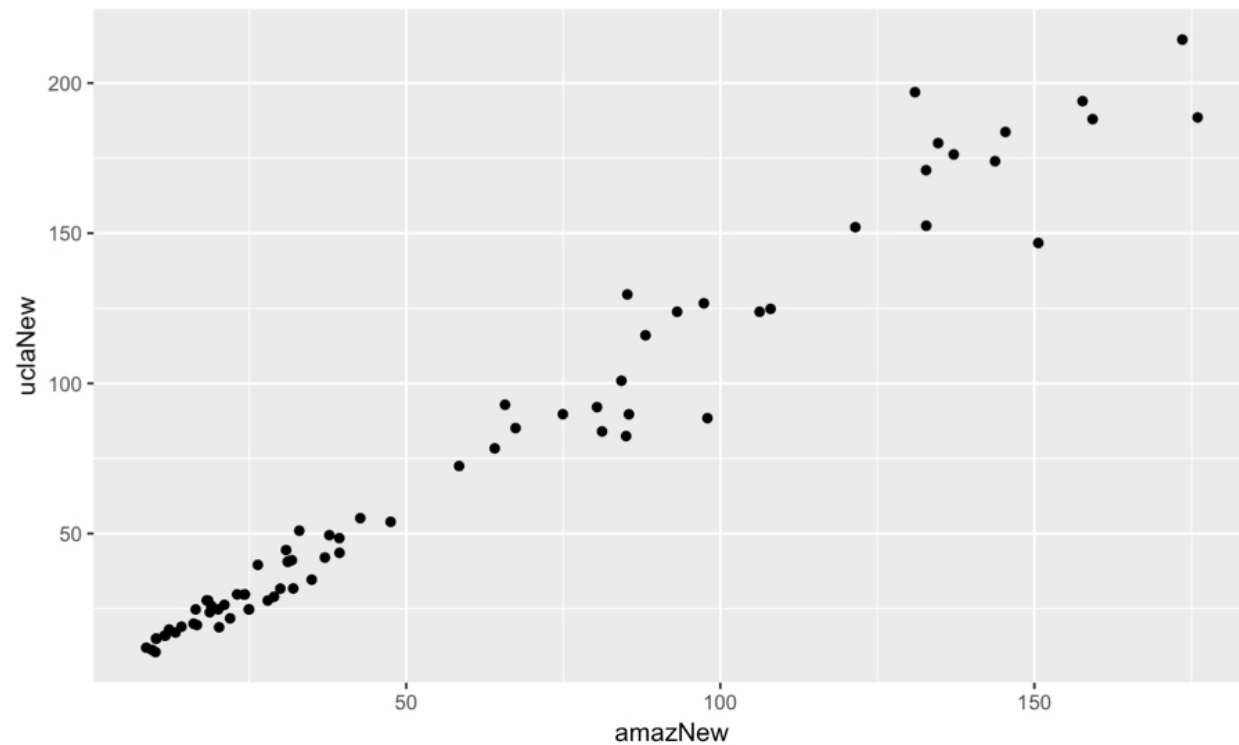
Compared to the course number?

```
textbooks %>%
  mutate(course_number = readr::parse_number(course)) %>%
  ggplot(aes(x = course_number, y = uclaNew)) +
    geom_point()
```
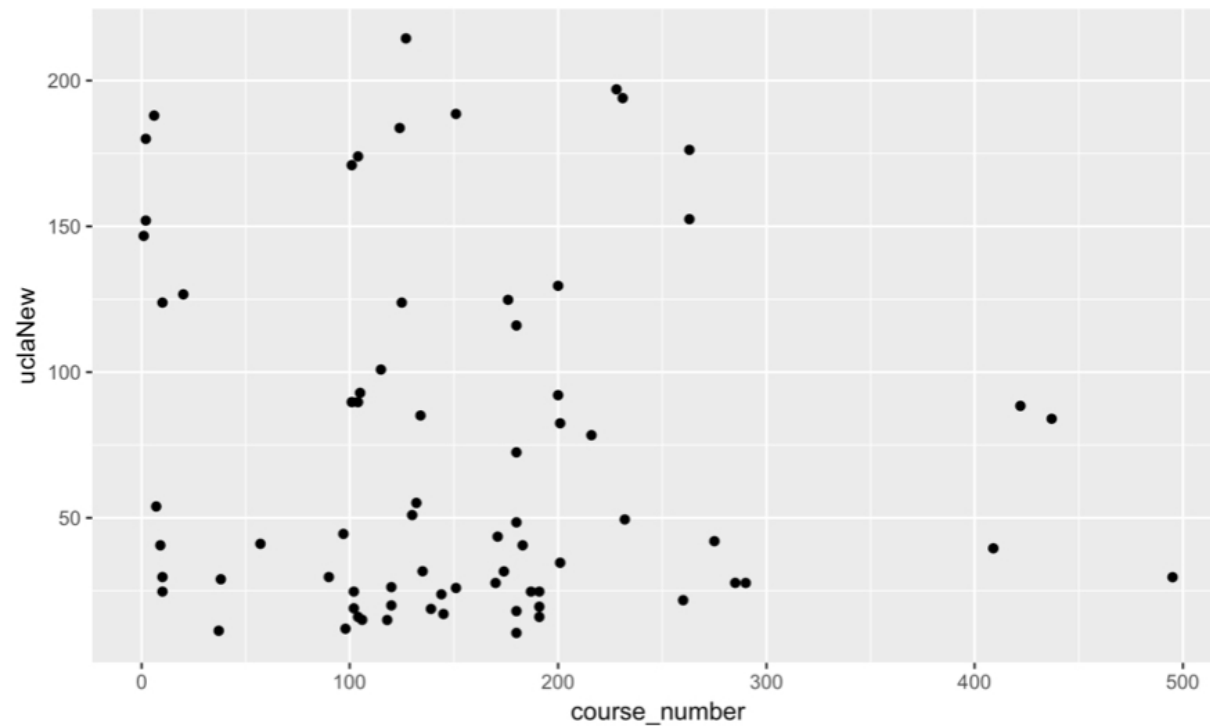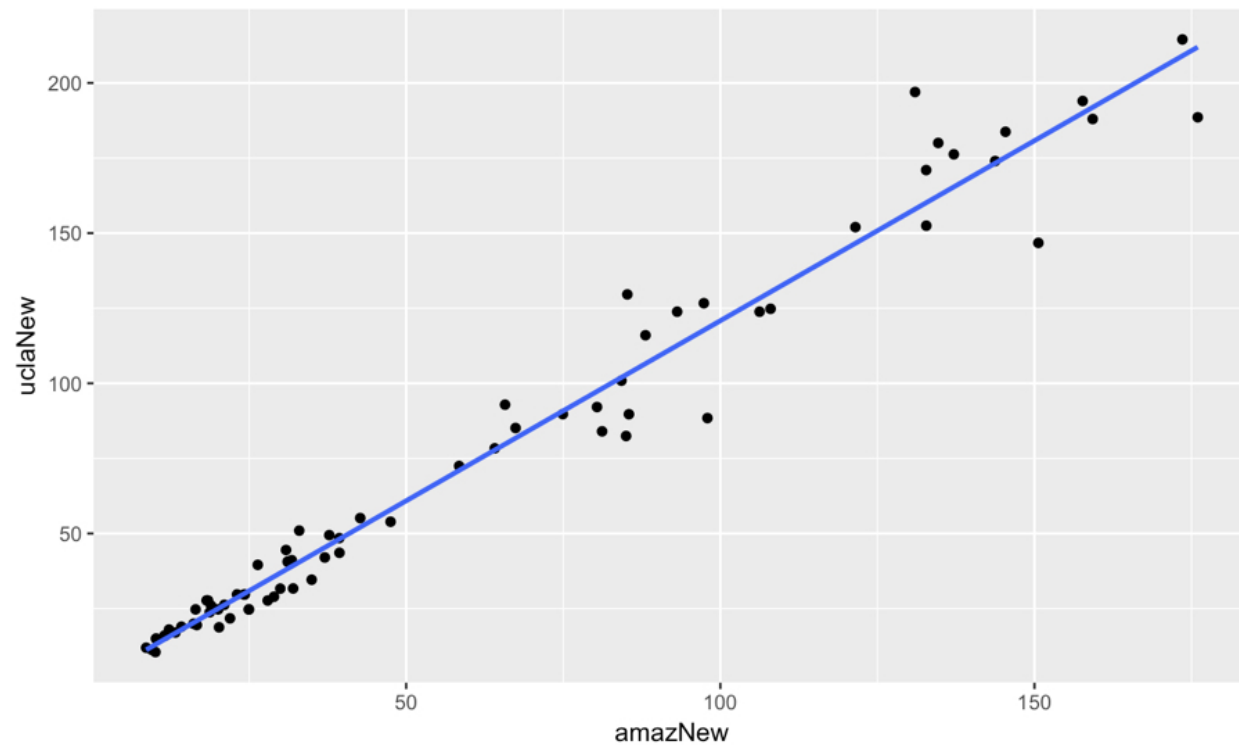
Compared to Amazon?

```
ggplot(data = textbooks, aes(x = amazNew, y = uclaNew)) +
  geom_point()
```

Compared to Amazon?

```
ggplot(data = textbooks, aes(x = amazNew, y = uclaNew)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE)
```

Slope and intercept

```
lm(uclaNew ~ amazNew, data = textbooks)
```

```
Call:
lm(formula = uclaNew ~ amazNew, data = textbooks)


Coefficients:
(Intercept)        amazNew
      0.929          1.199
```

$$\widehat{uclaNew} = 0.929 + 1.199 \cdot amazNew$$

Units and scale

```r
textbooks %>%
  mutate(amazNew_cents = amazNew * 100) %>%
  lm(uclaNew ~ amazNew_cents, data = .)
```

```
Call:
lm(formula = uclaNew ~ amazNew_cents, data = .)


Coefficients:
  (Intercept)   amazNew_cents
       0.929         0.01199
```

Let's practice!

CORRELATION AND REGRESSION IN R

Your linear model object

CORRELATION AND REGRESSION IN R

Ben Baumer
Assistant Professor at Smith College

Is that textbook overpriced?

```r
mod <- lm(uclaNew ~ amazNew, data = textbooks)
class(mod)
```

```
"lm"
```

Print

```
mod
```

```
Call:
lm(formula = uclaNew ~ amazNew, data = textbooks)

Coefficients:
(Intercept)        amazNew
      0.929          1.199
```

Fitted coefficients

```
coef(mod)
```

```
(Intercept)        amazNew
      0.929          1.199
```

Summary

```
summary(mod)
```

```
Call:
lm(formula = uclaNew ~ amazNew, data = textbooks)

Residuals:
   Min     1Q Median     3Q    Max
-34.78  -4.57   0.58   4.01  39.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9290     1.9354    0.48     0.63
amazNew       1.1990     0.0252   47.60   <2e-16

Residual standard error: 10.5 on 71 degrees of freedom
Multiple R-squared:  0.97,   Adjusted R-squared:  0.969
F-statistic: 2.27e+03 on 1 and 71 DF,  p-value: <2e-16
```

Fitted values

```
fitted.values(mod)
```

```
       1        2        3        4        5        6        7        8        9       10
   34.44    38.27    39.30    14.74    17.97    13.12    24.98    20.90   128.32    16.83
      11       12       13       14       15       16       17       18       19       20
   36.84   106.55    23.05    20.68   117.69    57.89    90.77   160.12   146.61   130.42
      21       22       23       24       25       26       27       28       29       30
   14.92    23.64    15.60    27.25    38.27    35.64    20.29    46.19    39.03    40.46
      31       32       33       34       35       36       37       38       39       40
   37.94   102.84    42.83   118.37    98.26    12.32    13.16   162.42   173.29   211.95
      41       42       43       44       45       46       47       48       49       50
   81.53   175.26   209.03   158.00   189.99   165.40    30.84   191.91    28.59    26.16
      51       52       53       54       55       56       57       58       59       60
   52.10    48.13   103.08   112.59    81.74   160.14    30.08    30.84   103.38    13.01
      61       62       63       64       65       66       67       68       69       70
   79.74   101.96    11.24    70.97    97.29    77.77    45.34    25.16    48.10    32.55
      71       72       73
   29.93    23.37    22.77
```

Residuals

```
residuals(mod)
```

```
         1          2          3          4          5          6          7
  -6.77105    2.32413   -7.61701    1.25854    0.98322    1.82719   -0.28093
         8          9         10         11         12         13         14
  -1.40433   -4.48287    0.17228   -5.20906    9.45100    4.61946    4.02348
        15         16         17         18         19         20         21
   8.98228   -3.99352   -1.04014   10.87962    5.39236   -5.62112    1.07869
        22         23         24         25         26         27         28
   2.31195    2.39526   -5.51705    2.32413   -6.69006   -0.34284    3.25873
        29         30         31         32         33         34         35
   2.05677   10.48996    6.55786  -20.39409   -8.23406  -29.95115  -14.26390
        36         37         38         39         40         41         42
  -1.06948    1.84122   17.60753    0.71458  -23.37321  -34.78455    8.48623
        43         44         45         46         47         48         49
   5.47235   39.00185    4.01249   10.85401   -6.14405   -3.90591    1.11007
        50         51         52         53         54         55         56
   0.08405    3.02765   -4.57365   26.51611   11.24803    3.37834   -7.66436
        57         58         59         60         61         62         63
```

broom

```
library(broom)
augment(mod)
```

```
    uclaNew amazNew .fitted .se.fit    .resid    .hat .sigma   .cooksd
1     27.67   27.95   34.44   1.460  -6.77105 0.01944 10.515 4.227e-03
2     40.59   31.14   38.27   1.418   2.32413 0.01834 10.543 4.687e-04
3     31.68   32.00   39.30   1.407  -7.61701 0.01806 10.507 4.955e-03
4     16.00   11.52   14.74   1.721   1.25854 0.02700 10.546 2.059e-04
5     18.95   14.21   17.97   1.674   0.98322 0.02555 10.546 1.186e-04
6     14.95   10.17   13.12   1.745   1.82719 0.02776 10.545 4.469e-04
7     24.70   20.06   24.98   1.577  -0.28093 0.02268 10.547 8.544e-06
8     19.50   16.66   20.90   1.632  -1.40433 0.02430 10.546 2.295e-04
9    123.84  106.25  128.32   1.700  -4.48287 0.02637 10.533 2.548e-03
10    17.00   13.26   16.83   1.690   0.17228 0.02605 10.547 3.716e-06
11    31.63   29.95   36.84   1.433  -5.20906 0.01874 10.528 2.407e-03
12   116.00   88.09  106.55   1.422   9.45100 0.01844 10.485 7.794e-03
13    27.67   18.45   23.05   1.603   4.61946 0.02343 10.532 2.390e-03
14    24.70   16.47   20.68   1.636   4.02348 0.02439 10.536 1.891e-03
15   126.67   97.38  117.69   1.554   8.98228 0.02202 10.491 8.468e-03
```

Let's practice!

CORRELATION AND REGRESSION IN R

Using your linear model

CORRELATION AND REGRESSION IN R

Ben Baumer
Assistant Professor at Smith College

Is that textbook overpriced?

```r
mod <- lm(uclaNew ~ amazNew, data = textbooks)
```

Examining residuals

```r
library(broom)
augment(mod) %>%
  arrange(desc(.resid)) %>%
  head()
```

|	uclaNew	amazNew	.fitted	.se.fit	.resid	.hat	.sigma	.cooksd	.std.resid
1	197.00	131.00	158.00	2.179	39.00	0.04331	9.409	0.32816	3.808
2	129.60	85.20	103.08	1.387	26.52	0.01753	10.051	0.05822	2.554
3	180.03	134.69	162.42	2.257	17.61	0.04644	10.324	0.07219	1.722
4	92.88	65.73	79.74	1.236	13.14	0.01393	10.428	0.01128	1.264
5	123.84	93.13	112.59	1.491	11.25	0.02026	10.459	0.01217	1.085
6	171.00	132.77	160.12	2.216	10.88	0.04479	10.463	0.02649	1.063

Markup

```r
textbooks %>%
  filter(uclaNew == 197)
```

```
  deptAbbr course            ibsn uclaNew amazNew more diff
1    Mgmt    228 978-0073379661     197     131    Y   66
```
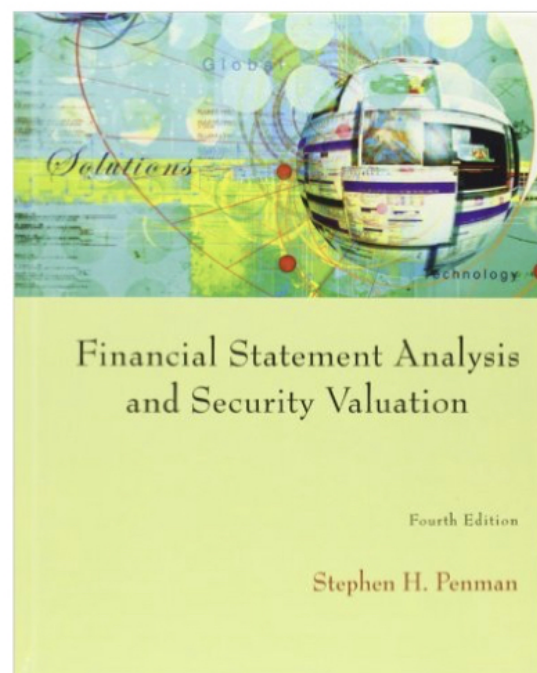
Making predictions
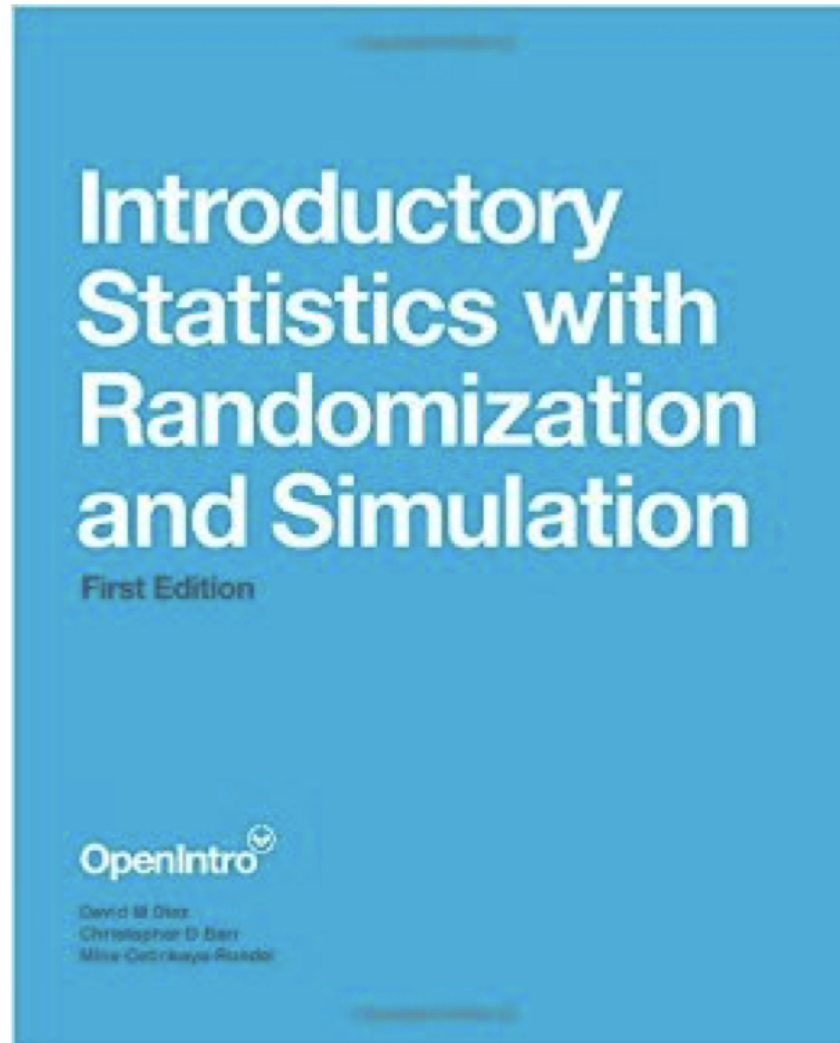
Making predictions

```
predict(   )
```

Making predictions

predict(lm)

Making predictions

predict(lm)

↓

fitted values for existing data

Making predictions

```
predict(lm, newdata)
```

Making predictions

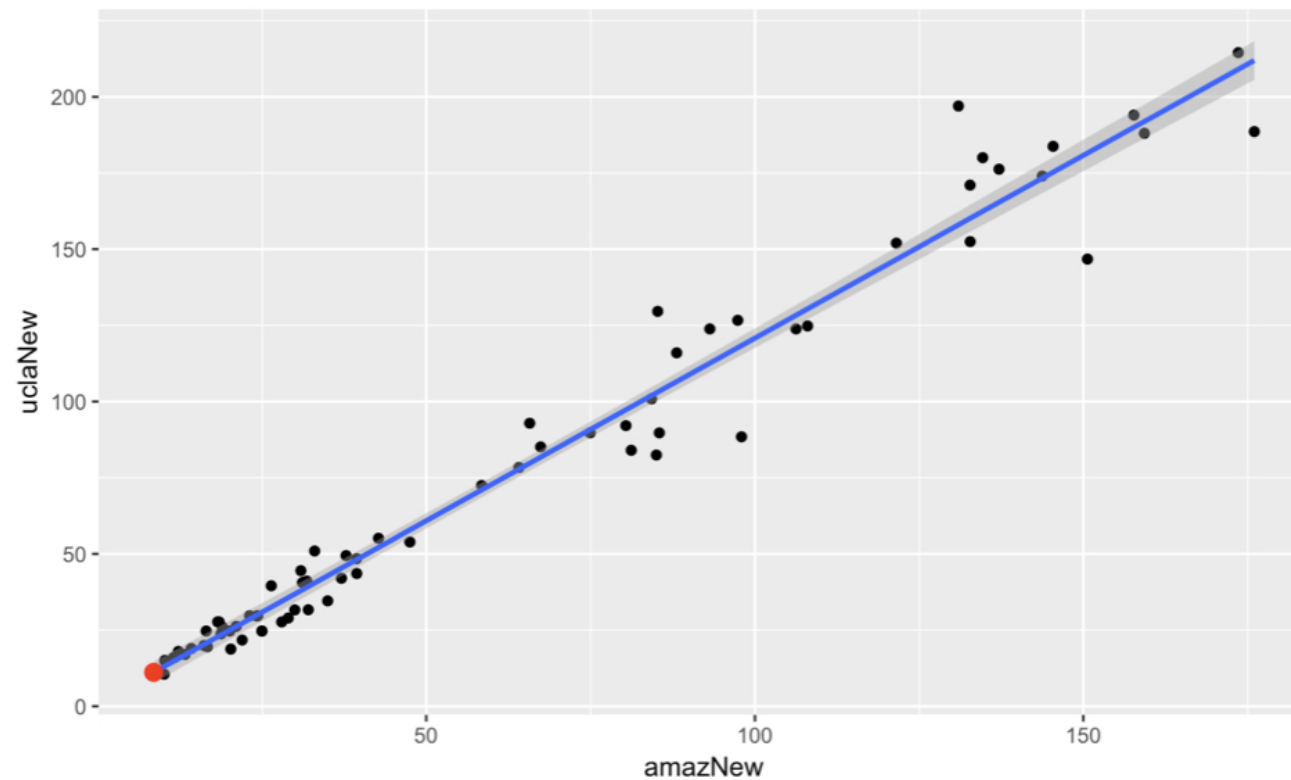predict(lm, newdata)

fitted values for any new data

New data

```r
new_data <- data.frame(amazNew = 8.49)
predict(mod, newdata = new_data)
```
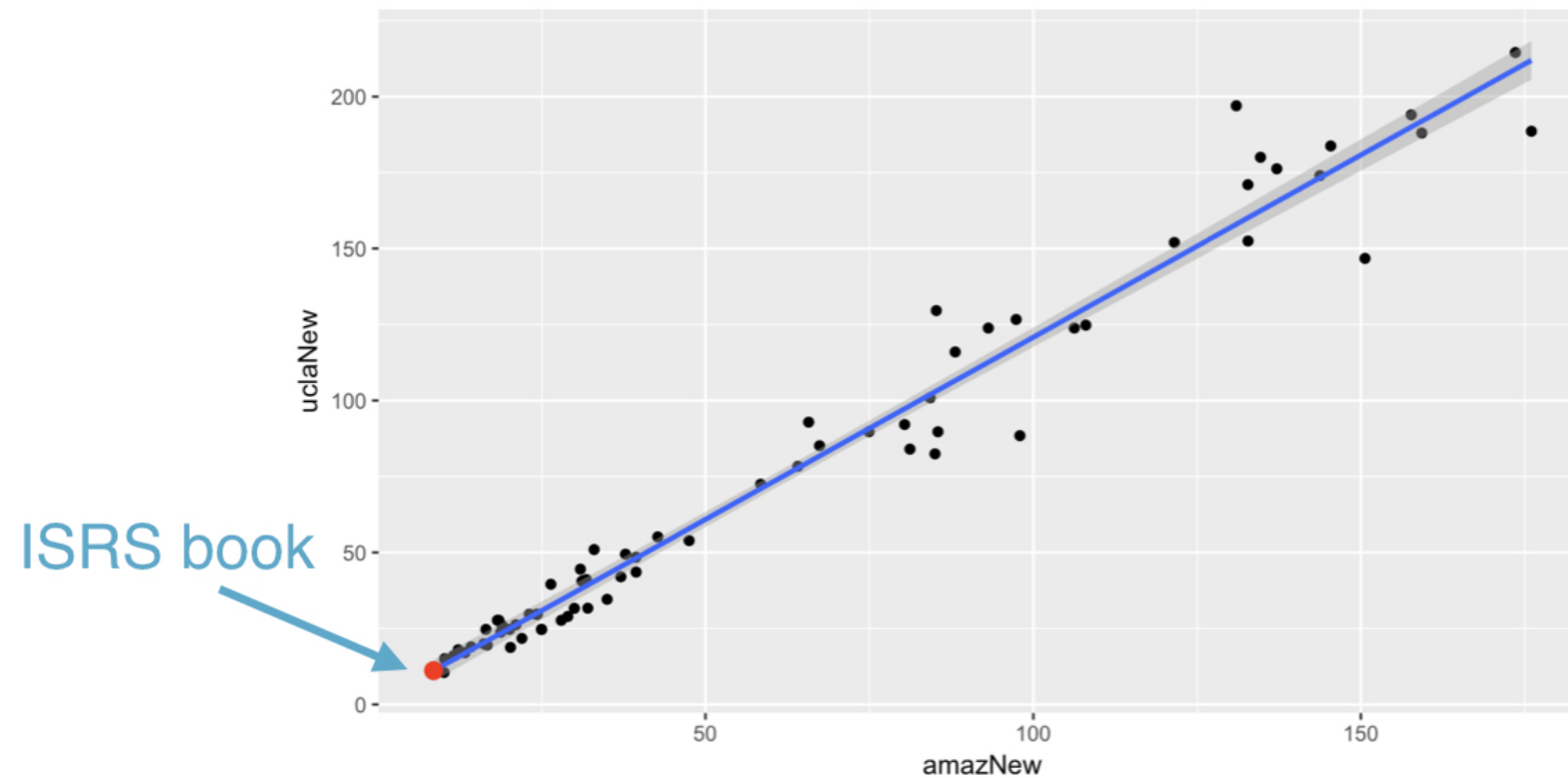
```
       1
11.11
```

Visualize new observations

```r
isrs <- broom::augment(mod, newdata = new_data)
ggplot(data = textbooks, aes(x = amazNew, y = uclaNew)) +
    geom_point() + geom_smooth(method = "lm") +
    geom_point(data = isrs, aes(y = .fitted), size = 3, color = "red")
```

Visualize new observations

```
isrs <- broom::augment(mod, newdata = new_data)
ggplot(data = textbooks, aes(x = amazNew, y = uclaNew)) +
    geom_point() + geom_smooth(method = "lm") +
    geom_point(data = isrs, aes(y = .fitted), size = 3, color = "red")
```

Let's practice!

CORRELATION AND REGRESSION IN R