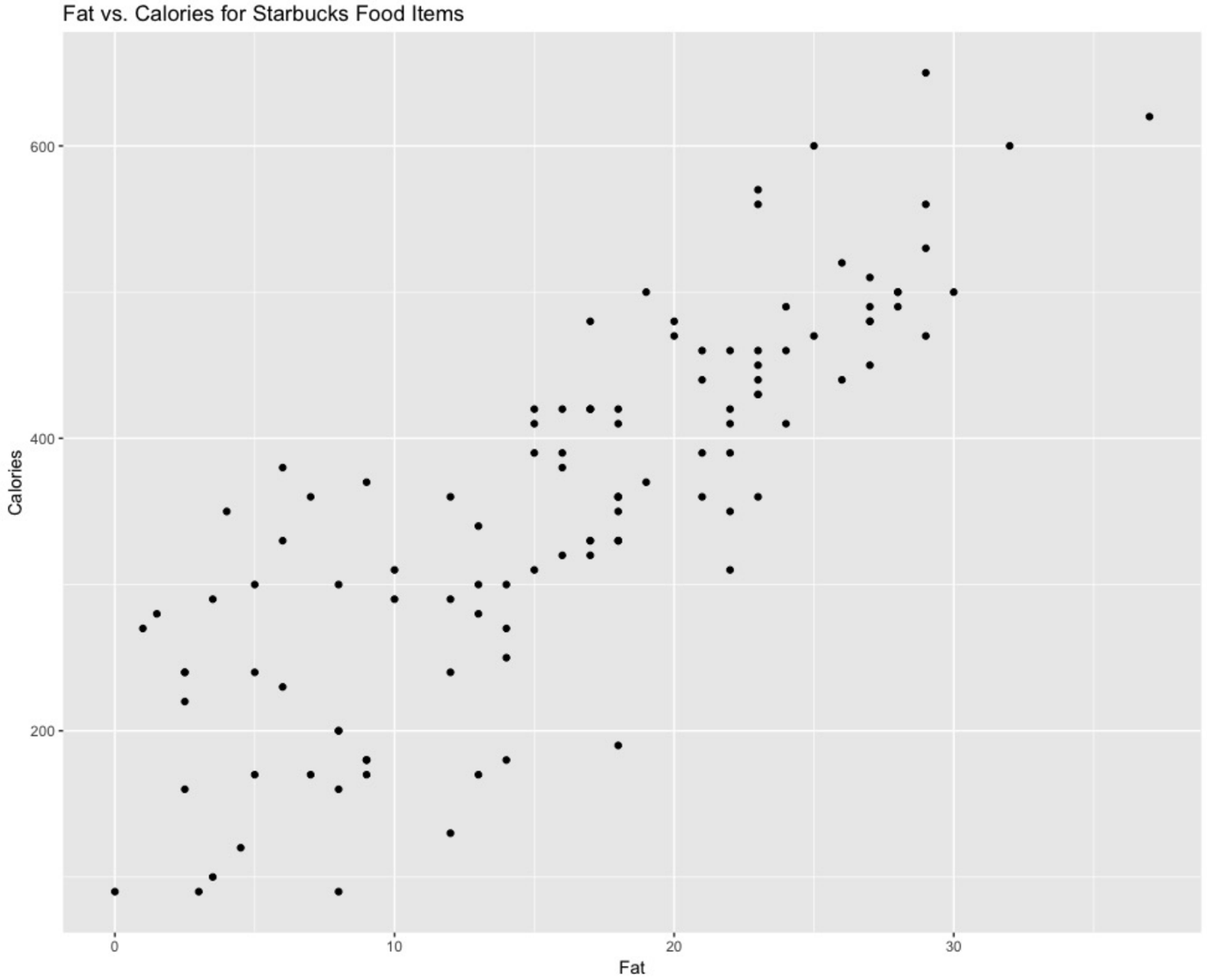
Welcome to the course!

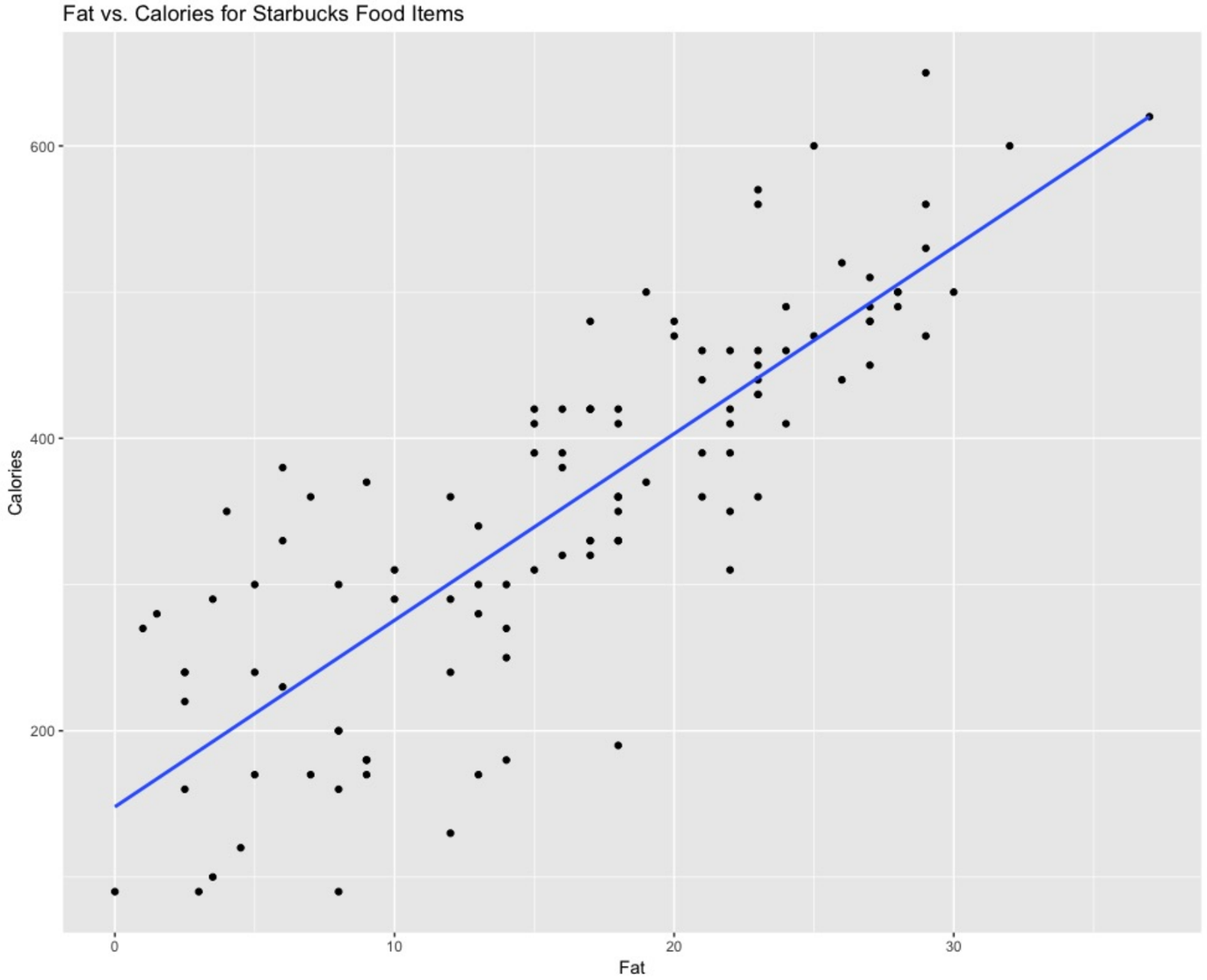
INFERENCE FOR LINEAR REGRESSION IN R

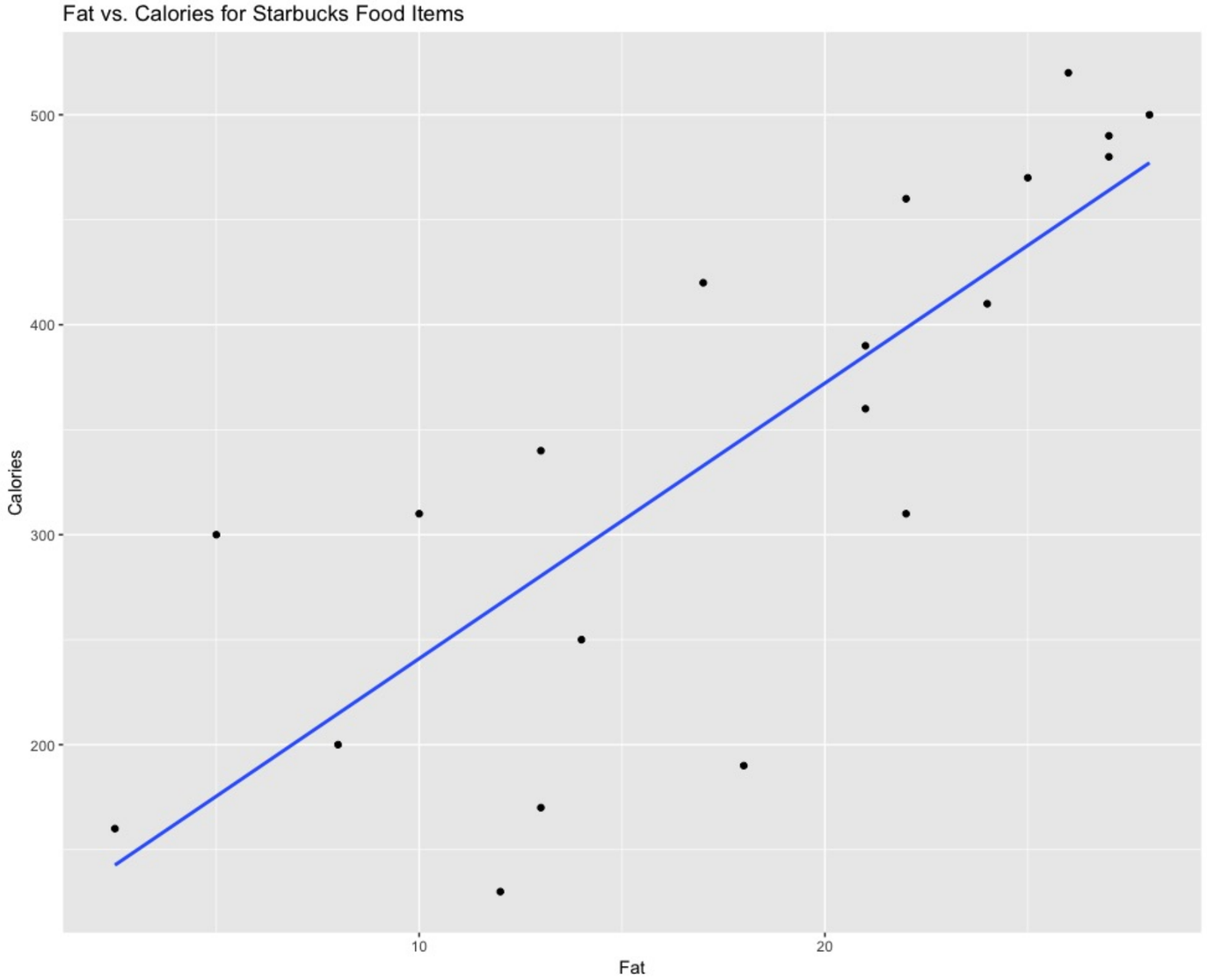
Jo Hardin
Professor, Pomona College

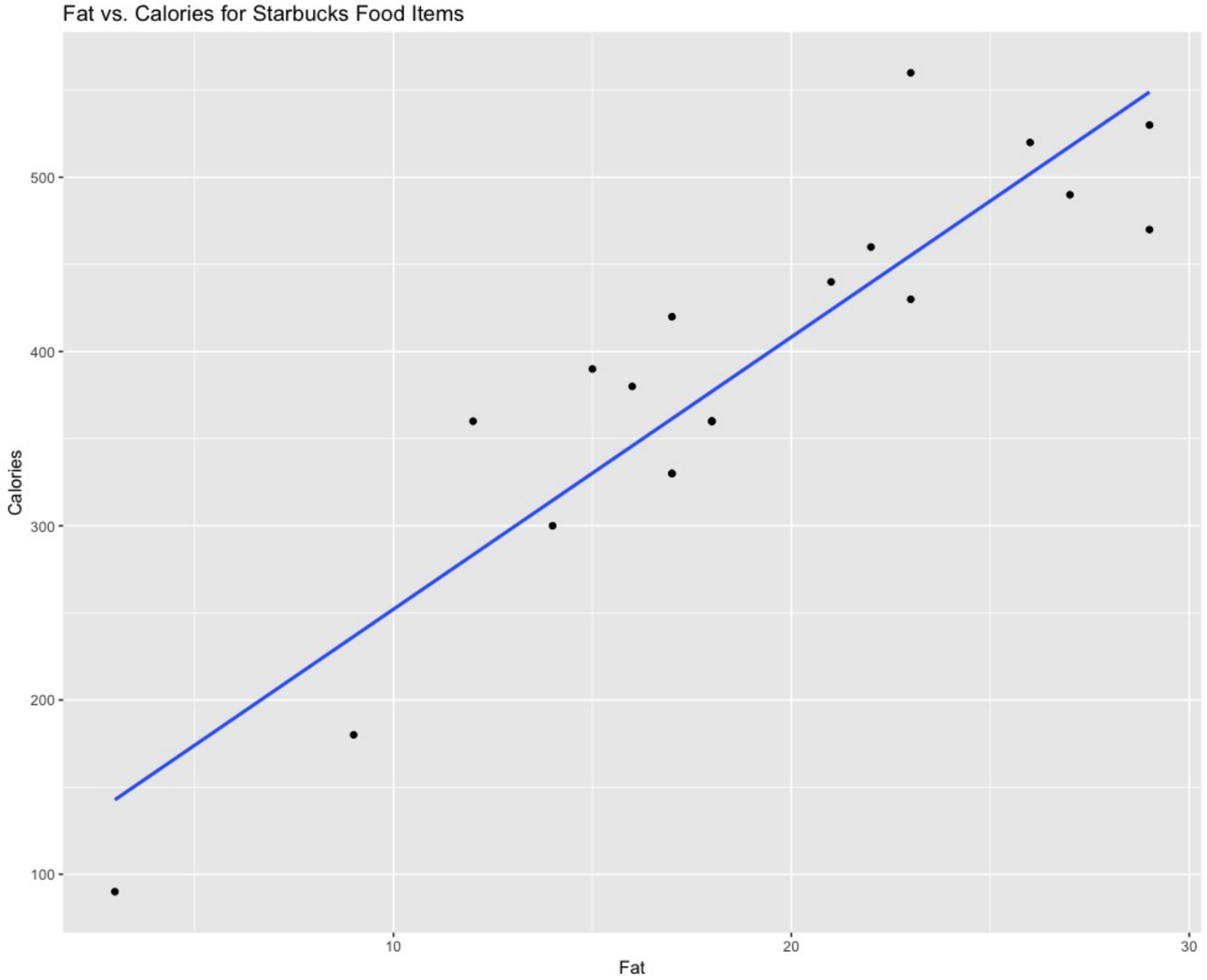
In this course you will

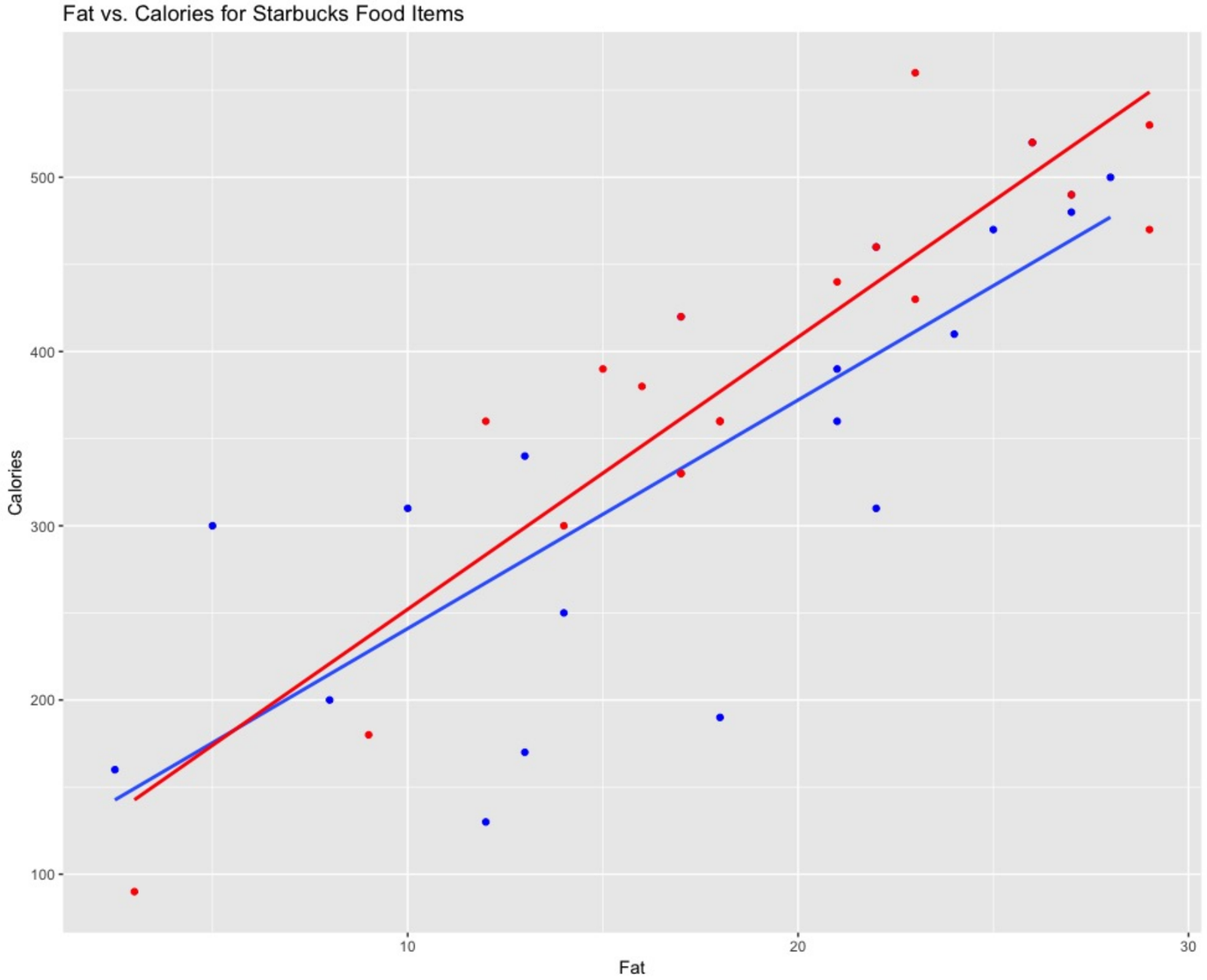
- Make inferential claims about models.
- Use least squares estimation.
- Create confidence intervals for the slope.


Fat vs. Calories for Starbucks Food Items
Calories
600
400
200
0
10
20
30
Fat


Fat vs. Calories for Starbucks Food Items
Calories
600
400
200
0
10
20
30
Fat

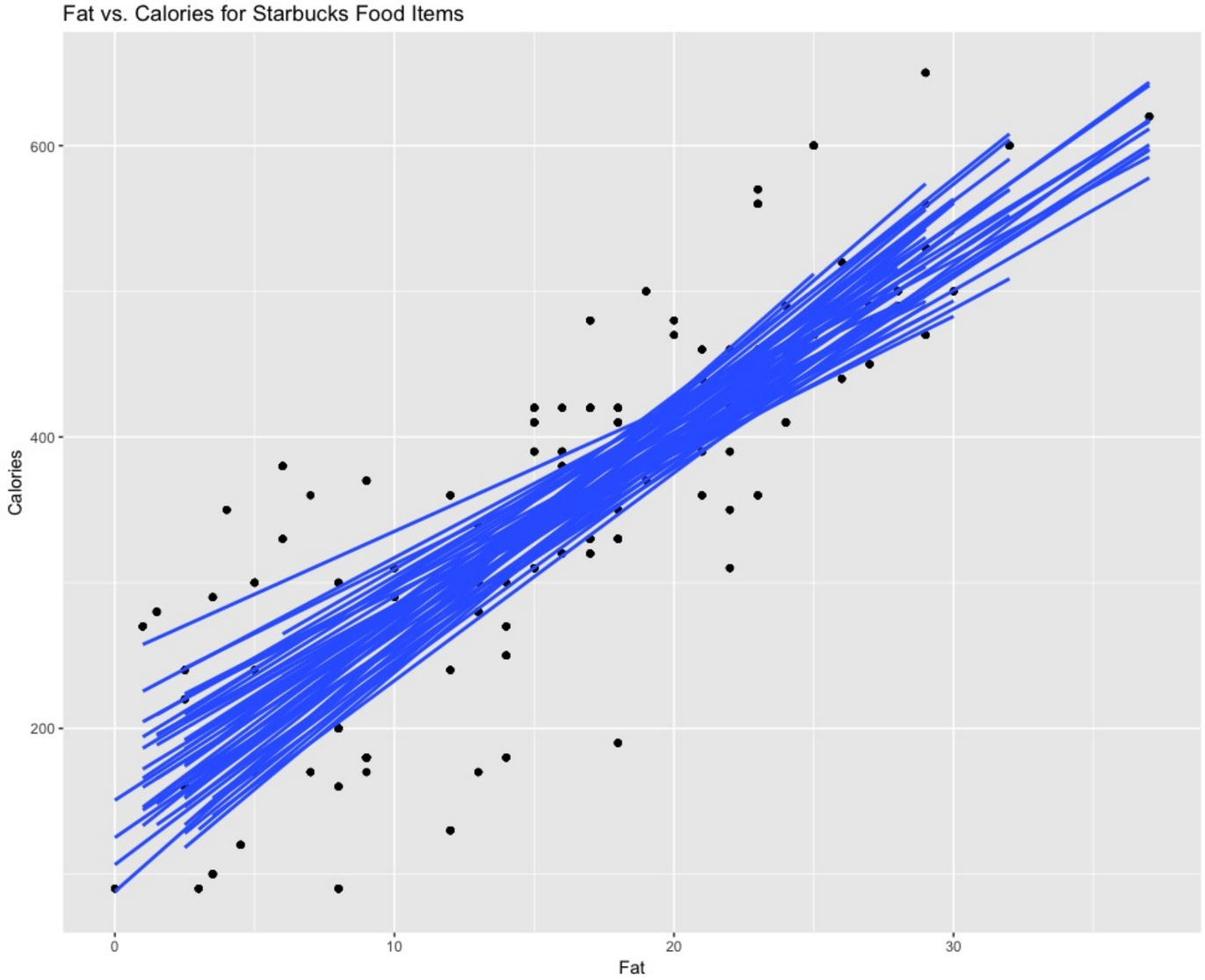

Fat vs. Calories for Starbucks Food Items
Calories
500
400
300
200
10
20
Fat

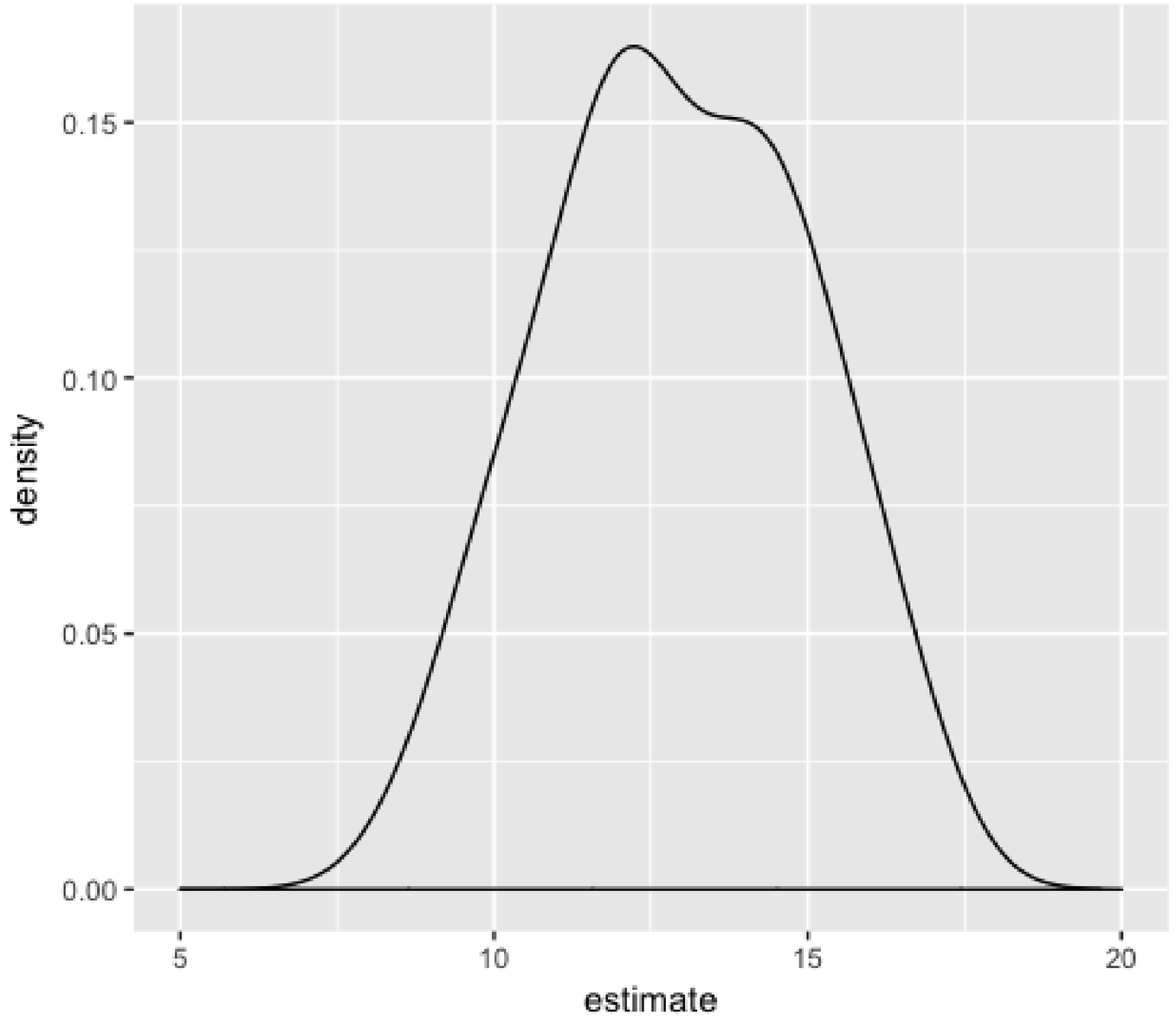

Fat vs. Calories for Starbucks Food Items
Calories
500
400
300
200
100
10
20
30
Fat


Fat vs. Calories for Starbucks Food Items
Calories
500
400
300
200
100
10
20
30
Fat

Sampling variability

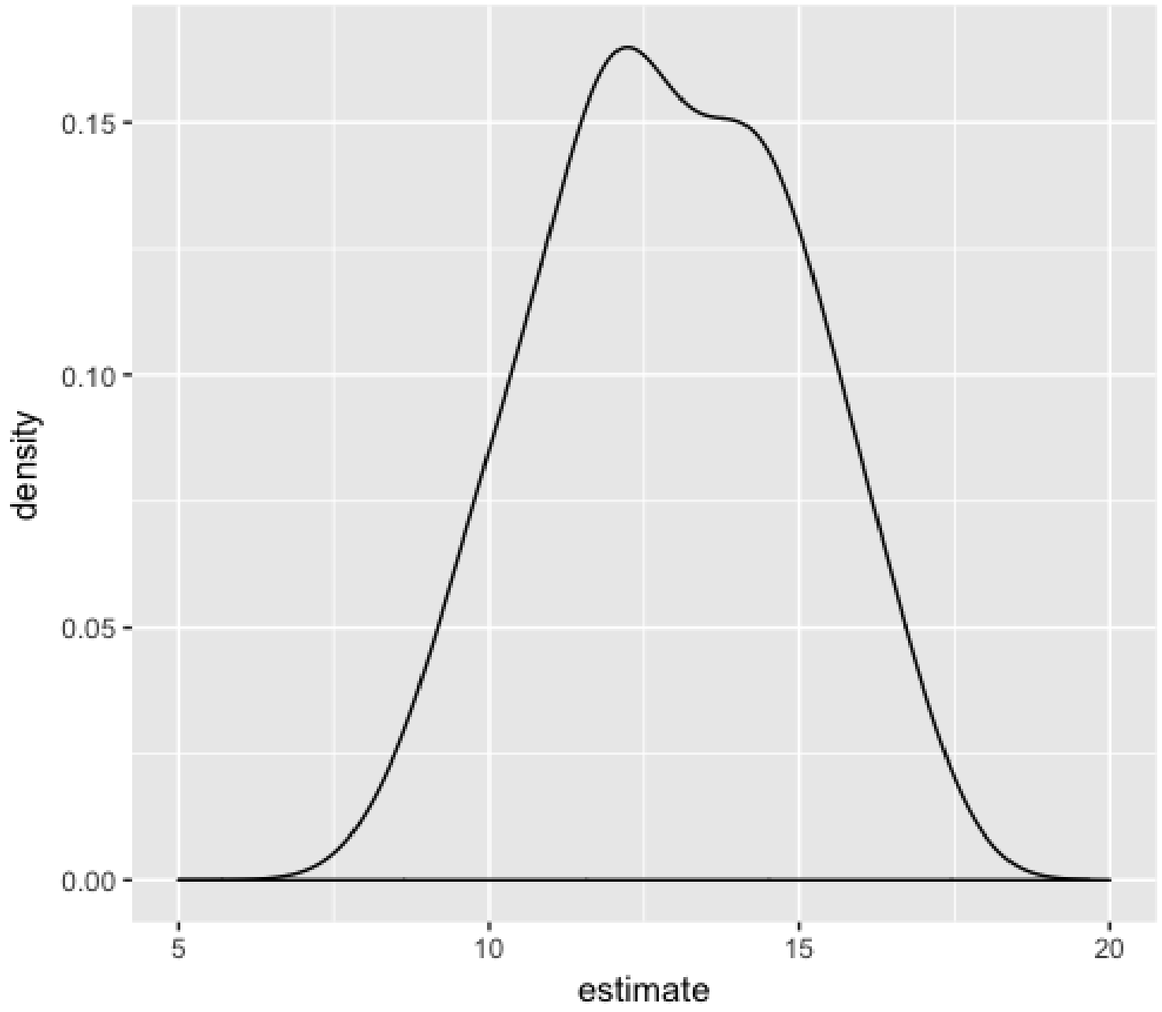
- Variability in the regression line

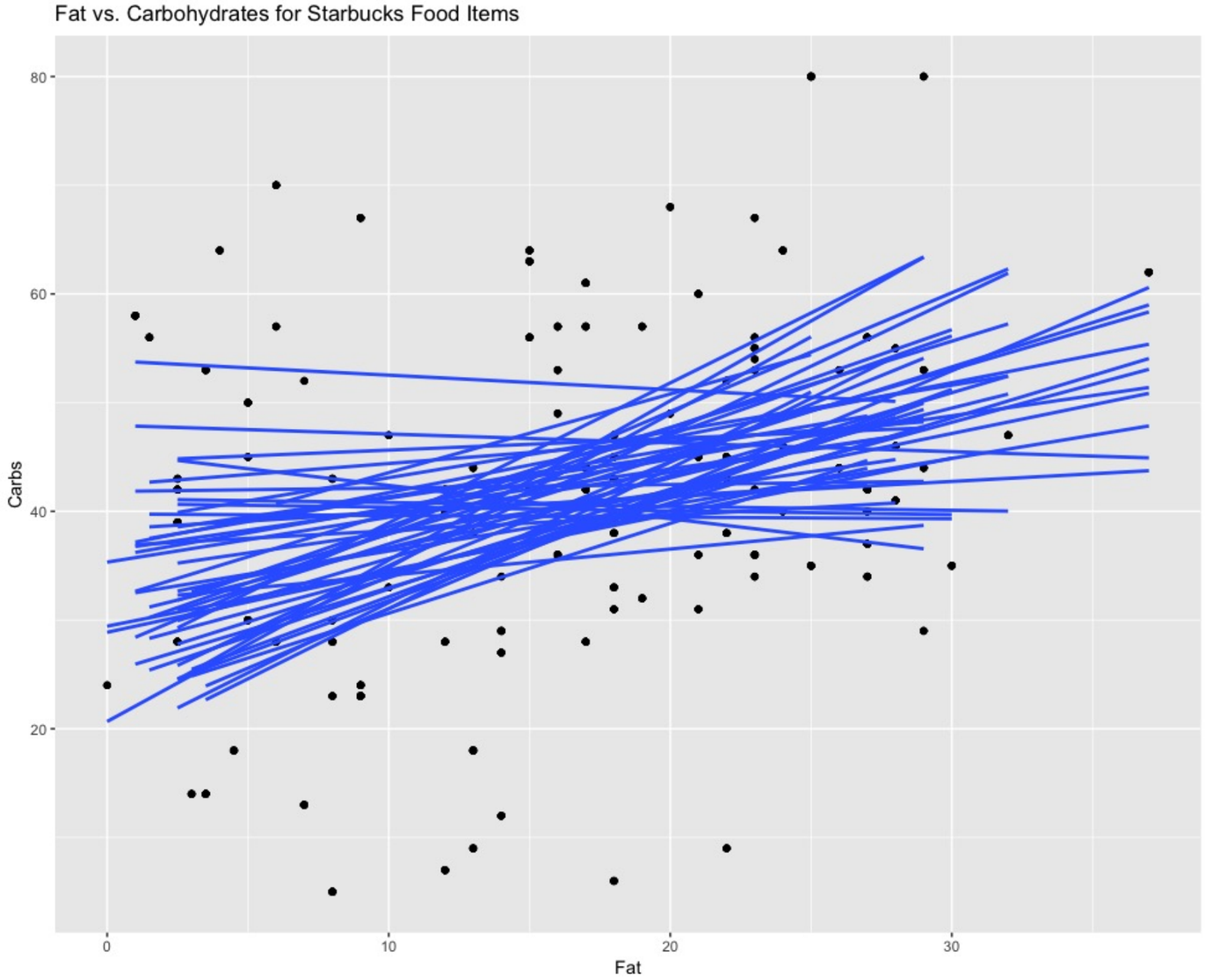

Fat vs. Calories for Starbucks Food Items
Calories
600
400
200
0
10
20
30
Fat

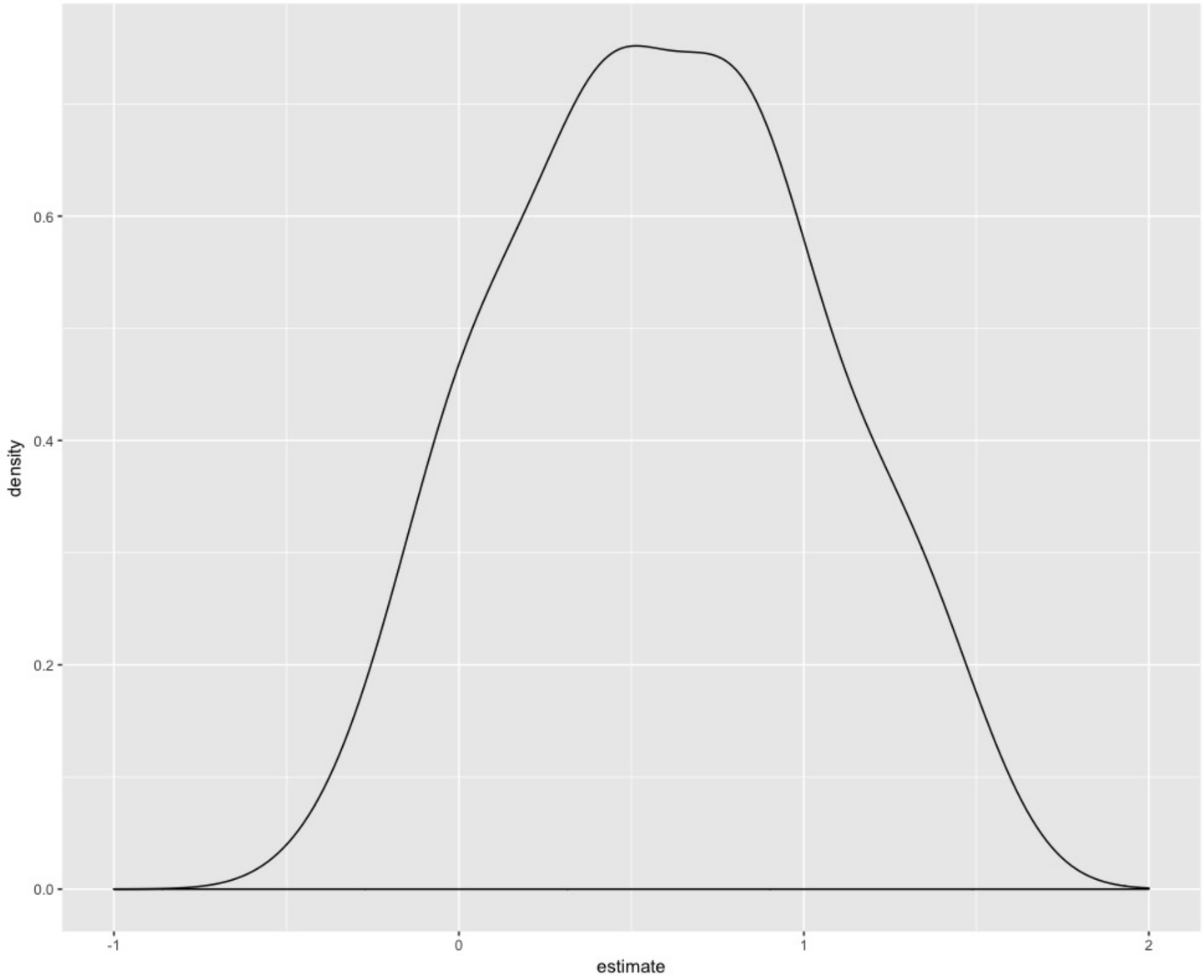

0.15
0.10
0.05
0.00
density
5
10
15
20
estimate

Interpret the density plot

- Slopes between 8 and 17
- None close to zero
- Strong evidence the association is positive

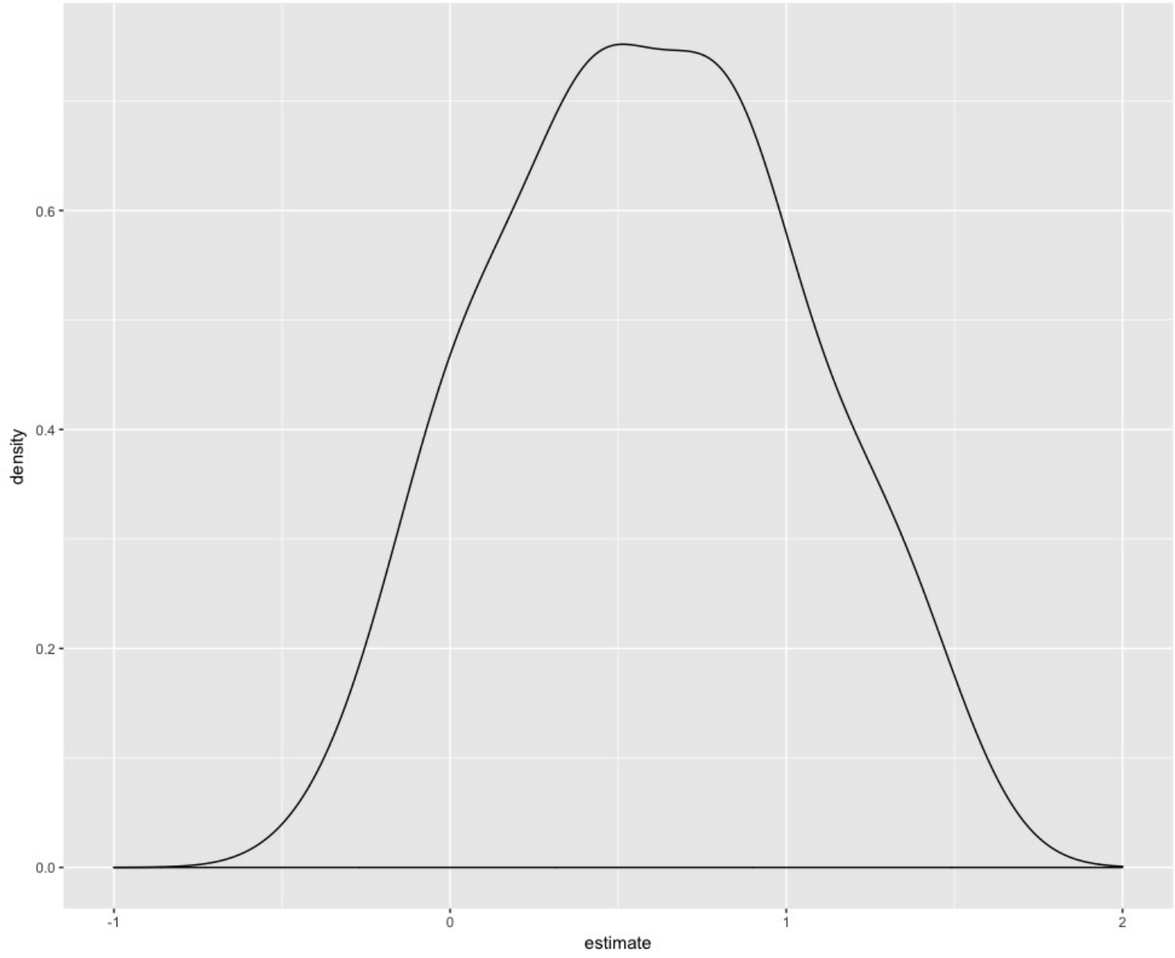



Fat vs. Carbohydrates for Starbucks Food Items
Carbs
80
60
40
20
0
10
20
30
Fat


density
0.6
0.4
0.2
0.0
-1
0
1
2
estimate

Interpreting the density plot

- Some slopes close to zero
- High variability
- We can't make any conclusions



Let's practice!

INFERENCE FOR LINEAR REGRESSION IN R

Research question

INFERENCE FOR LINEAR REGRESSION IN R

Jo Hardin
Professor, Pomona College

Consider possible research questions for the Starbucks data

- Are protein and carbohydrates linearly associated in the population? (two-sided research question)
- Are protein and carbohydrates linearly associated in a **positive** direction in the population? (one-sided research question)

```
head(starbucks)
```

```
A tibble: 6 x 6
                                 Item Calories   Fat Carbs
                                <chr>    <int> <dbl> <int>
1                        Chonga Bagel      300     5    50
2                        8-Grain Roll      380     6    70
3                      Almond Croissant      410    22    45
4                         Apple Fritter      460    23    56
5                      Banana Nut Bread      420    22    52
6 Blueberry Muffin with Yogurt and Honey      380    16    53
# ... with 2 more variables: Fiber <int>, Protein <int>
```

```
summary(lm(Carbs ~ Protein, data = starbucks))
```

```
Call:
lm(formula = Carbs ~ Protein, data = starbucks)
Residuals:
    Min      1Q  Median      3Q     Max
-35.360 -11.019   0.125   9.970  35.640
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.1116     2.4680   15.04   <2e-16 ***
Protein       0.3815     0.1734    2.20   0.0299 *
 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
```

```
lm(Carbs ~ Protein, data = starbucks) %>% tidy()
```

```
         term   estimate std.error statistic      p.value
1 (Intercept) 37.1116401 2.4680349 15.036919 1.539345e-28
2     Protein  0.3814696 0.1734226  2.199654 2.990434e-02
```

```
Call:
lm(formula = Carbs ~ Protein,
   data = starbucks)

Residuals:
    Min      1Q  Median      3Q     Max
-35.360 -11.019   0.125   9.970  35.640

Coefficients:
```

```
summary(lm(Carbs ~ Protein,
        data = starbucks))
```

```
            Estimate Std. Error
(Intercept)  37.1116     2.4680
Protein       0.3815     0.1734
            t value Pr(>|t|)
(Intercept)   15.04   <2e-16 ***
Protein        2.20   0.0299 *
--
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Std. Error
2.4680
0.1734
```

```
lm(Carbs ~ Protein,
   data = starbucks) %>%
  tidy()
```

```
std.error
2.4680349
0.1734226
```

```
Call:
lm(formula = Carbs ~ Protein,
   data = starbucks)

Residuals:
    Min      1Q  Median      3Q     Max
-35.360 -11.019   0.125   9.970  35.640

Coefficients:
```

```
summary(lm(Carbs ~ Protein,
           data = starbucks))
```

```
t value
15.04
2.20
```

```
            Estimate Std. Error
(Intercept)  37.1116     2.4680
Protein       0.3815     0.1734
            t value Pr(>|t|)
(Intercept)   15.04   <2e-16 ***
Protein        2.20   0.0299 *
--
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm(Carbs ~ Protein,
   data = starbucks) %>%
  tidy()
```

```
statistic
15.036919
 2.199654
```

```
Call:
lm(formula = Carbs ~ Protein,
   data = starbucks)

Residuals:
    Min      1Q  Median      3Q     Max
-35.360 -11.019   0.125   9.970  35.640

Coefficients:
```

```
summary(lm(Carbs ~ Protein,
        data = starbucks))
```

```
            Estimate Std. Error
(Intercept)  37.1116     2.4680
Protein       0.3815     0.1734
            t value Pr(>|t|)
(Intercept)   15.04   <2e-16 ***
Protein        2.20   0.0299 *
--
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Pr(>|t|)
<2e-16 ***
0.0299 *
```

```
lm(Carbs ~ Protein,
   data = starbucks) %>%
  tidy()
```

```
p.value
1.539345e-28
2.990434e-02
```

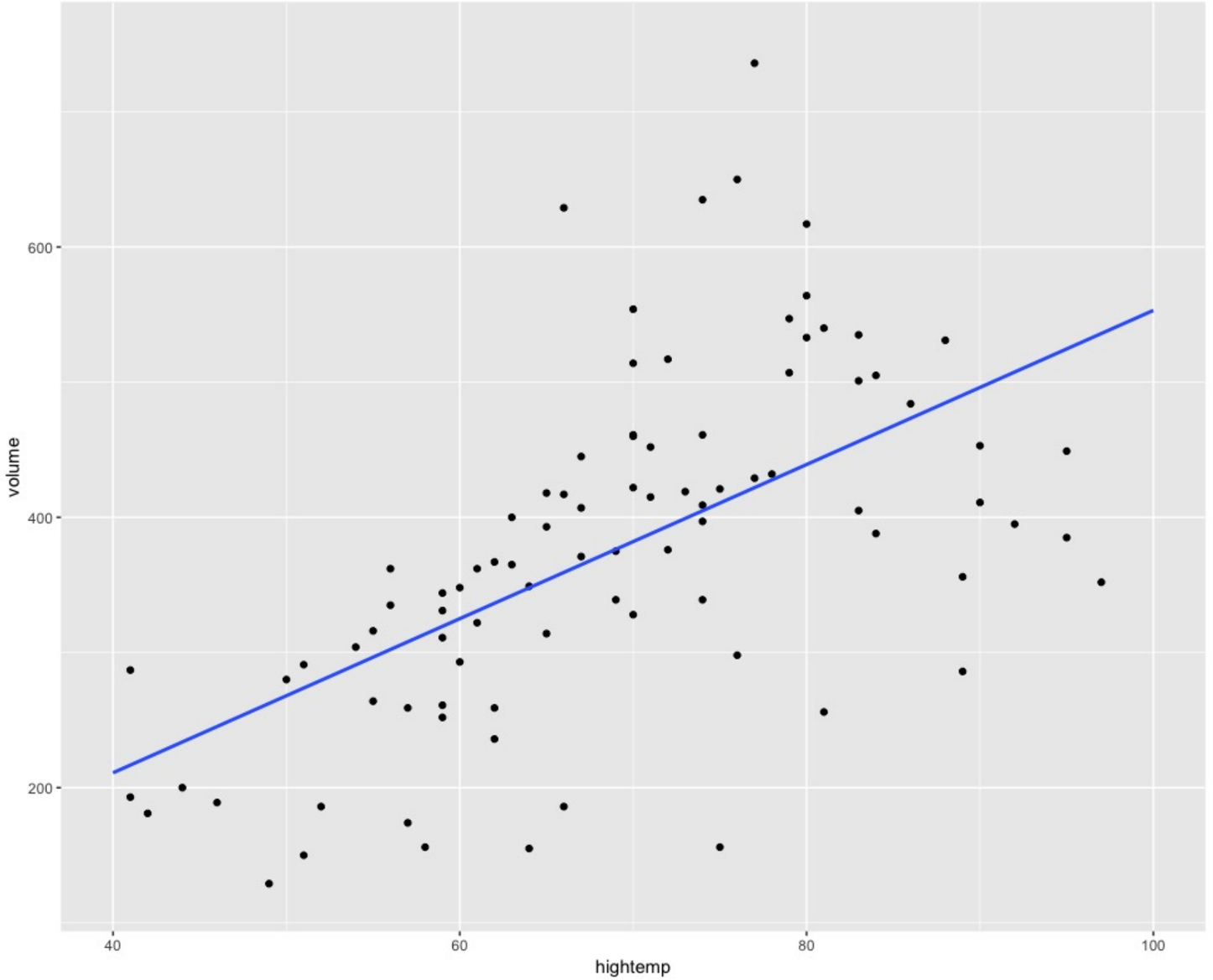
Let's practice!

INFERENCE FOR LINEAR REGRESSION IN R

Variability of coefficients

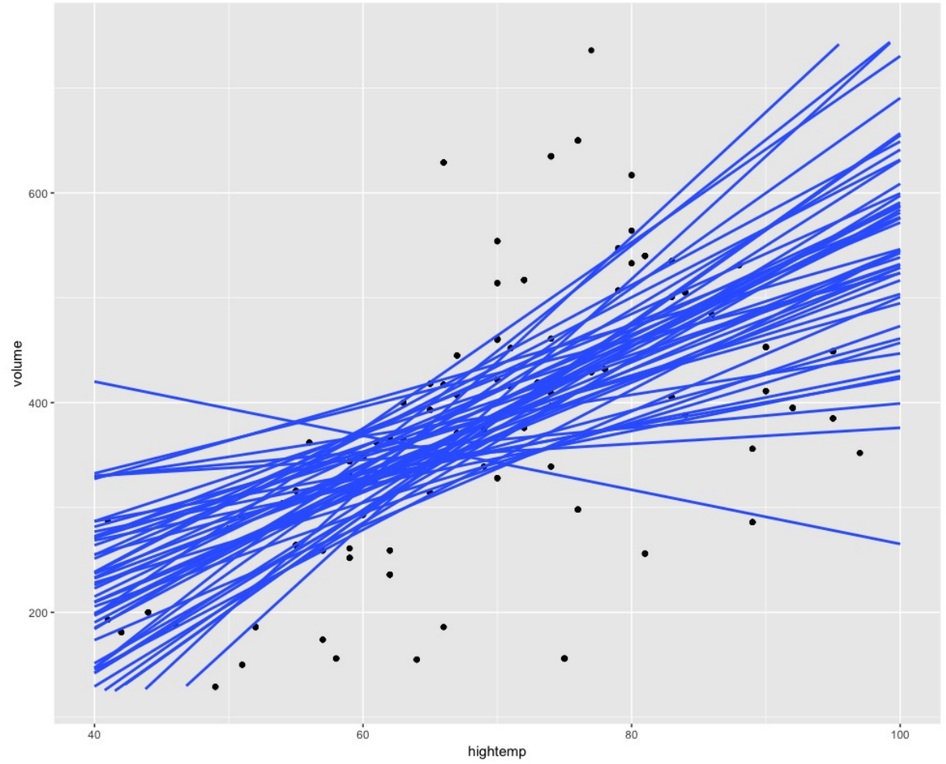
INFERENCE FOR LINEAR REGRESSION IN R

Jo Hardin
Professor, Pomona College

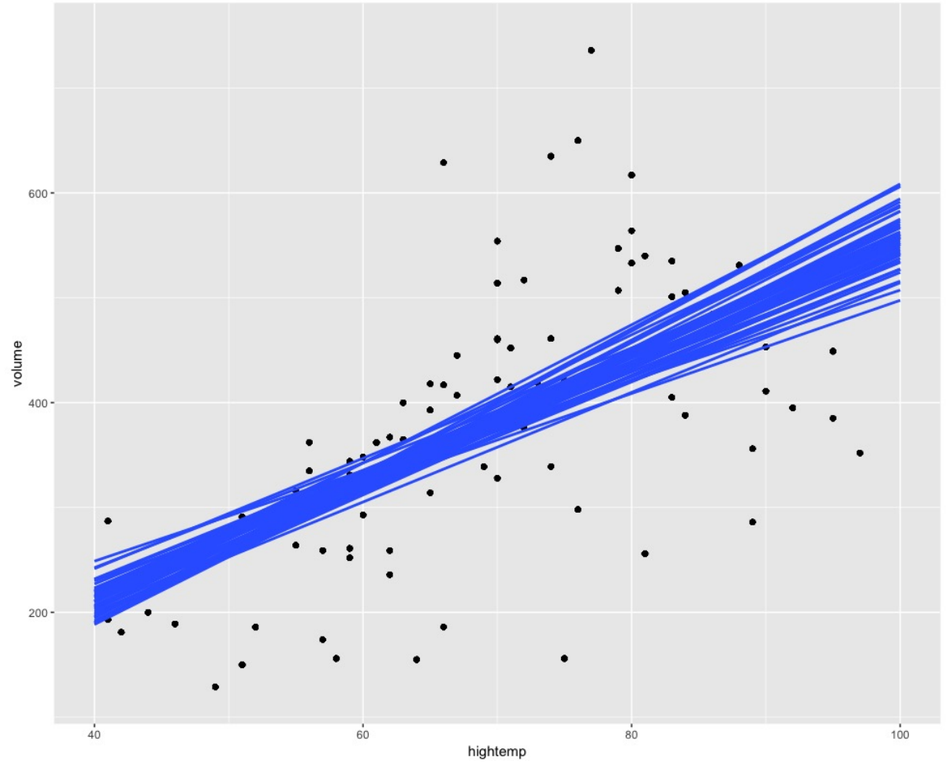

volume
600
400
200
40
60
80
100
hightemp

RailTrails - a change in sample size

n=10

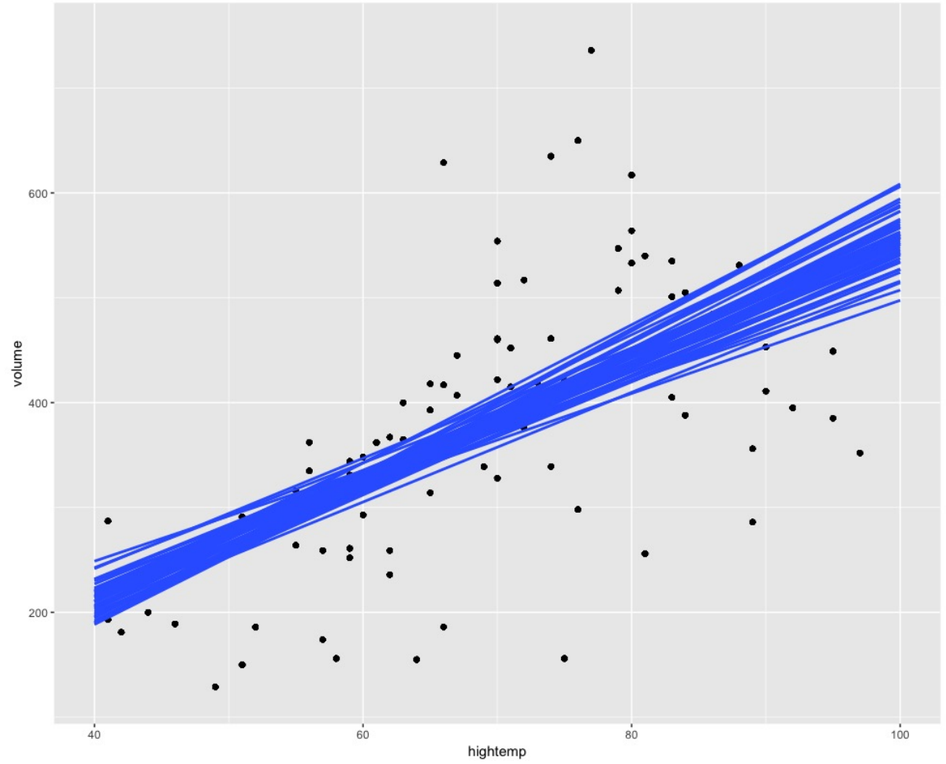


n=50

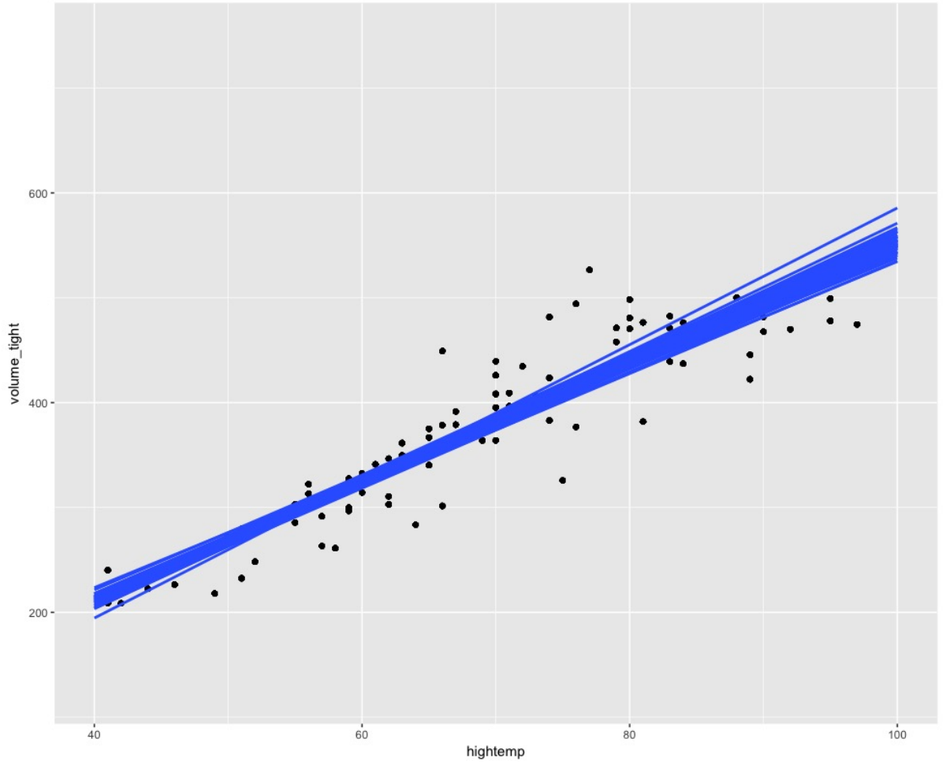


RailTrails - less variability around the line

n=50, original data

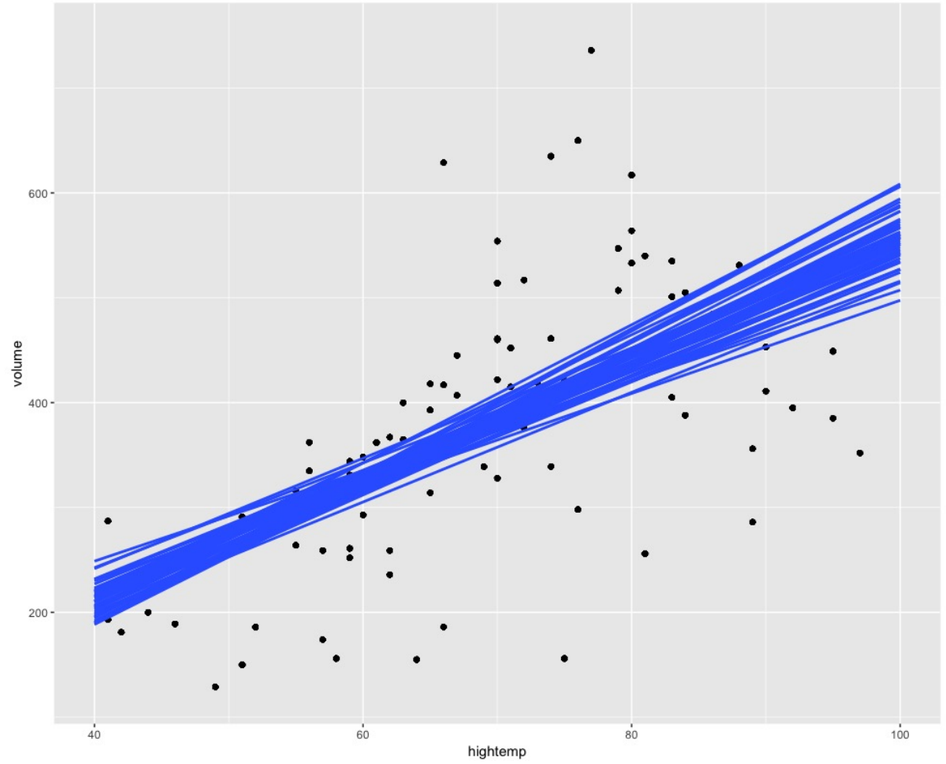


n=50, tighter data

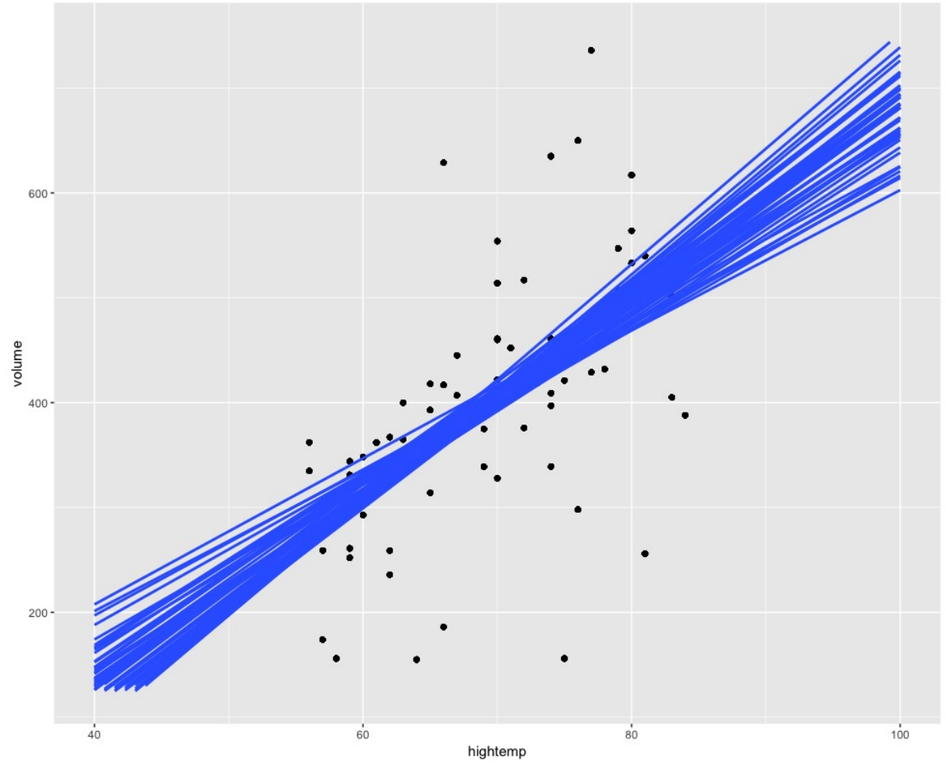


RailTrails - less variability in the x direction

n=50, original data



n=50, less data in x-direction



Let's practice!

INFERENCE FOR LINEAR REGRESSION IN R