Modeling the ACFT

Code

3/29/2023

Load the data and packages.

Code

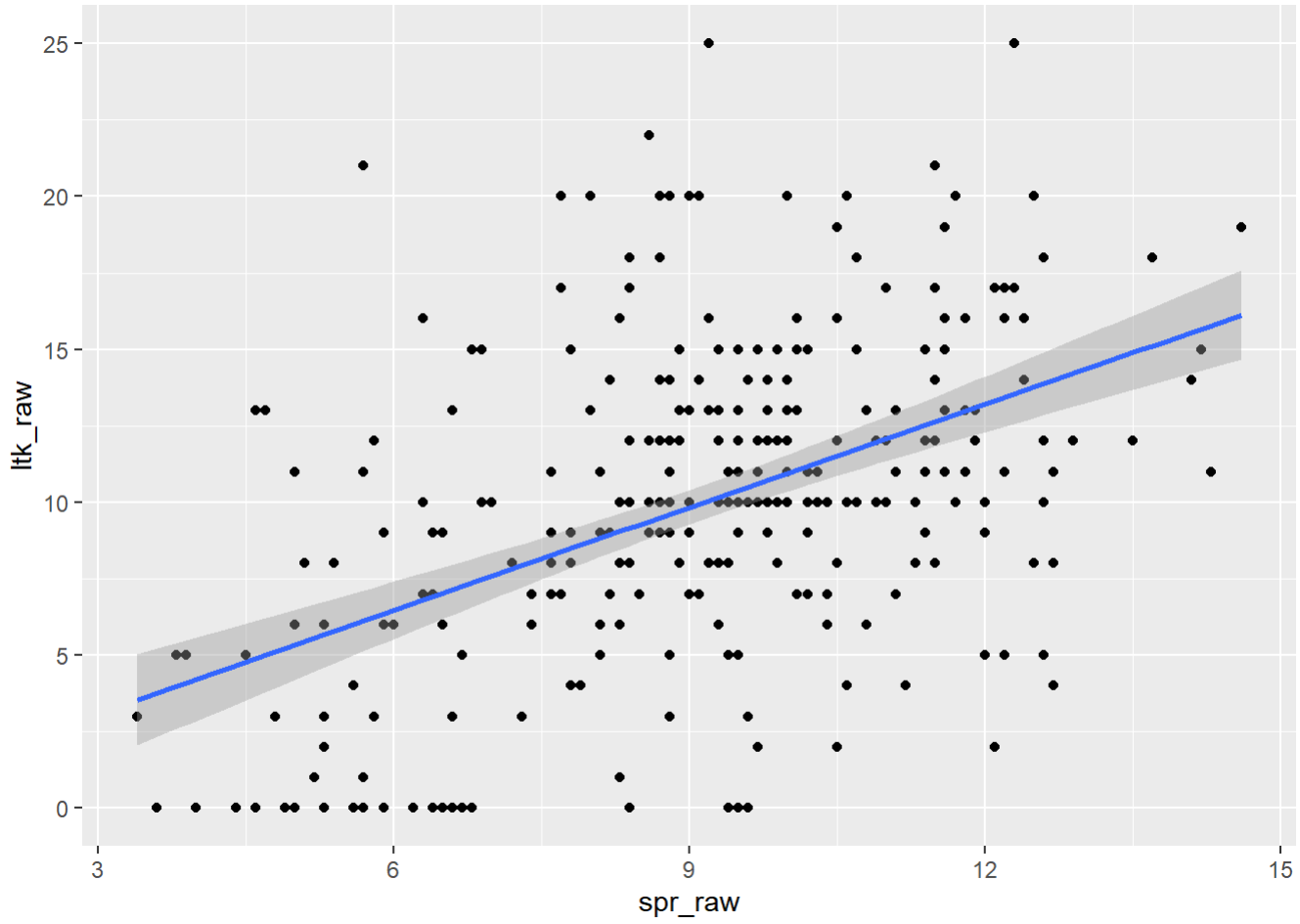
Is there an association between the Standing Power Throw (SPT) and Leg Tuck (LTK) raw scores?

Causal Diagram: SPT $\rightarrow$ LTK

1. Write the equation for the linear model with beta's.
   $\hat{LTK} = \beta_0 + \beta_1 * SPT$
2. Give the hypotheses in symbols and words.
   $H_0 : \beta_1 = 0$ "No association between SPT and LTK raw scores."
   $H_A : \beta_1 \neq 0$ "There is an association between SPT and LTK raw scores."
3. Plot the association, including the regression line. Is there anything unusual about the data? (Hint: You may need to remove an outlier.)

Code

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



1. Create a linear model to test your hypotheses.

Code

```
##
## Call:
## lm(formula = ltk_raw ~ spr_raw, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3131  -2.9451  -0.2205   2.8304  14.9427
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2713     1.1653  -0.233    0.816
## spr_raw       1.1227     0.1245   9.015   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.825 on 290 degrees of freedom
## Multiple R-squared:  0.2189, Adjusted R-squared:  0.2162
## F-statistic: 81.28 on 1 and 290 DF,  p-value: < 2.2e-16
```

1. Give the updated linear model with the coefficients determined by linear regression.

$$\hat{LTK} = -0.271 + 1.12 * SPT$$

2. Interpret the intercept term.
   The predicted number of LTK's for someone with an SPT of 0m is -0.271 reps. This doesn't have any real world significance because we are extrapolating beyond our data.

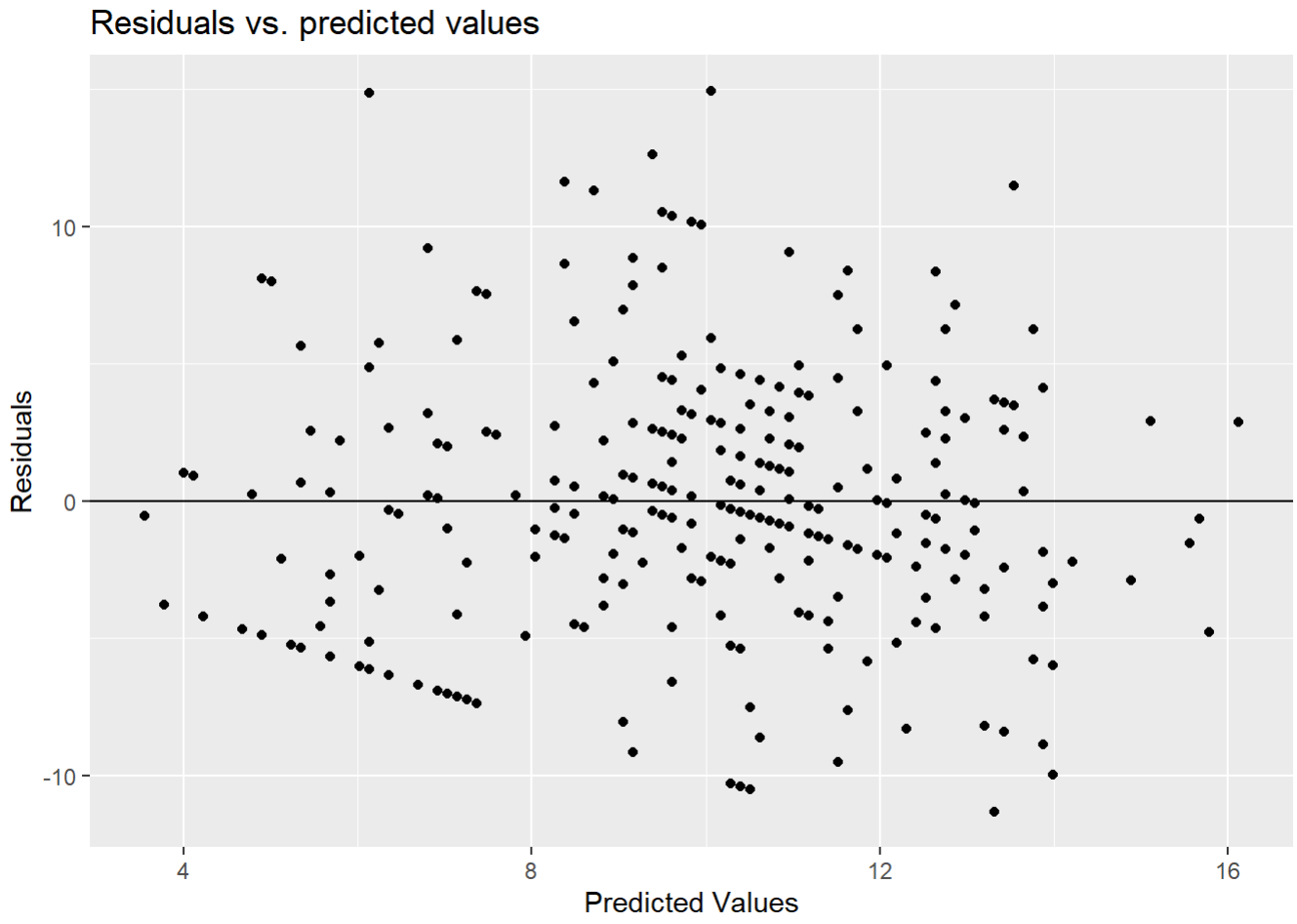
3. Interpret the slope term.
   For every 1 meter increase in SPT, we expect a 1.12 rep increase in leg tucks.

4. Conclusion of your hypothesis test at a 5% significance level.
   We are looking at the p-value associated with the SPT coefficient ($< 2 * 10^{-16}$).
   $\text{p-value} < \alpha$, so we reject H_0 in favor of the alternative. There is a positive association between SPT and LTK.

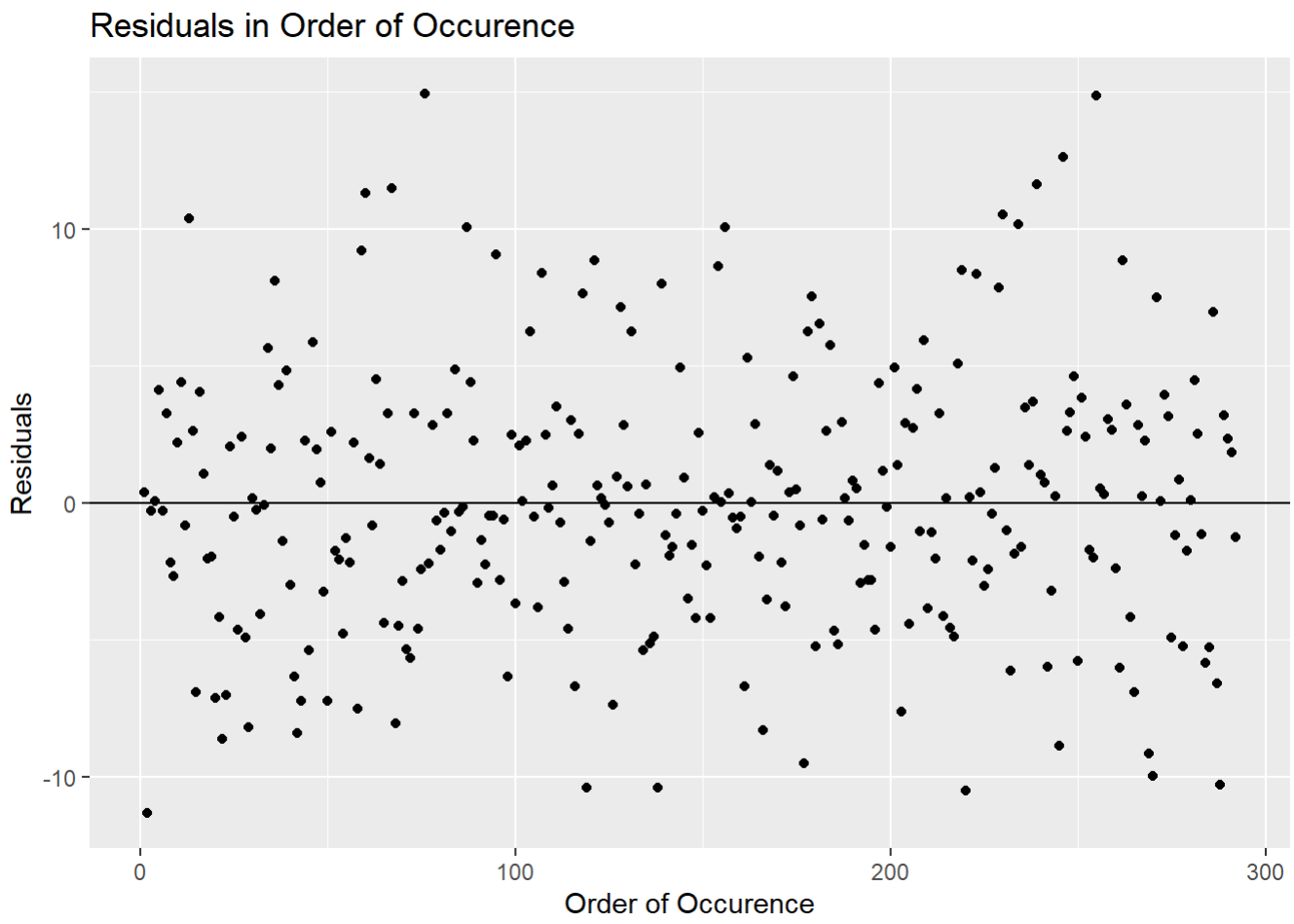
5. Interpret the R-squared value.
   Our model accounts for 21.9% of the variance in LTK reps.

6. Check the 4 Validity Conditions (L.I.N.E.). Are any not met?

Code

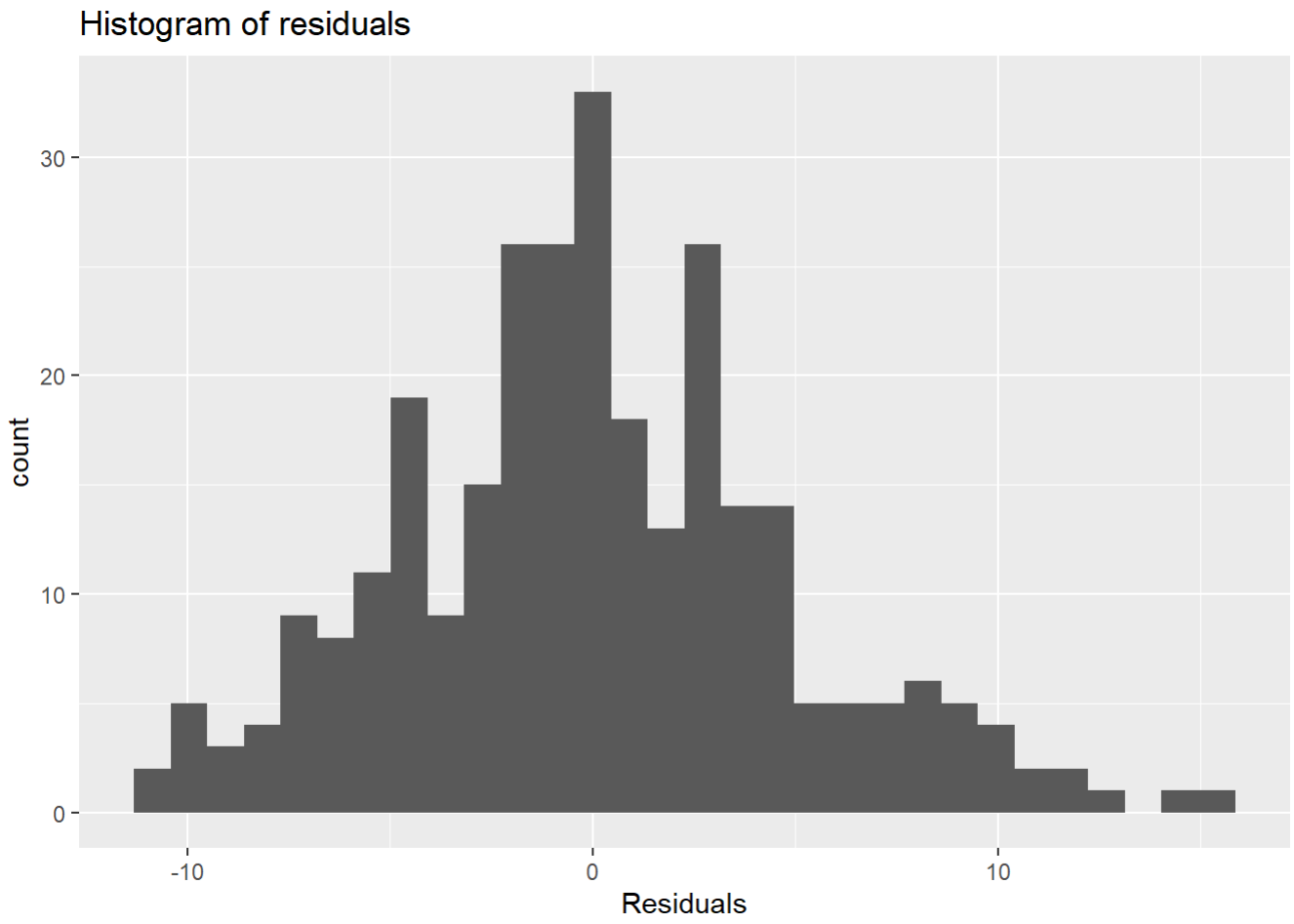


Code



Code

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



All four validity conditions are met.

1. Can we generalize these results to a larger population? Why or why not?
   No, this is not a random sample of cadets. There is the potential for bias.

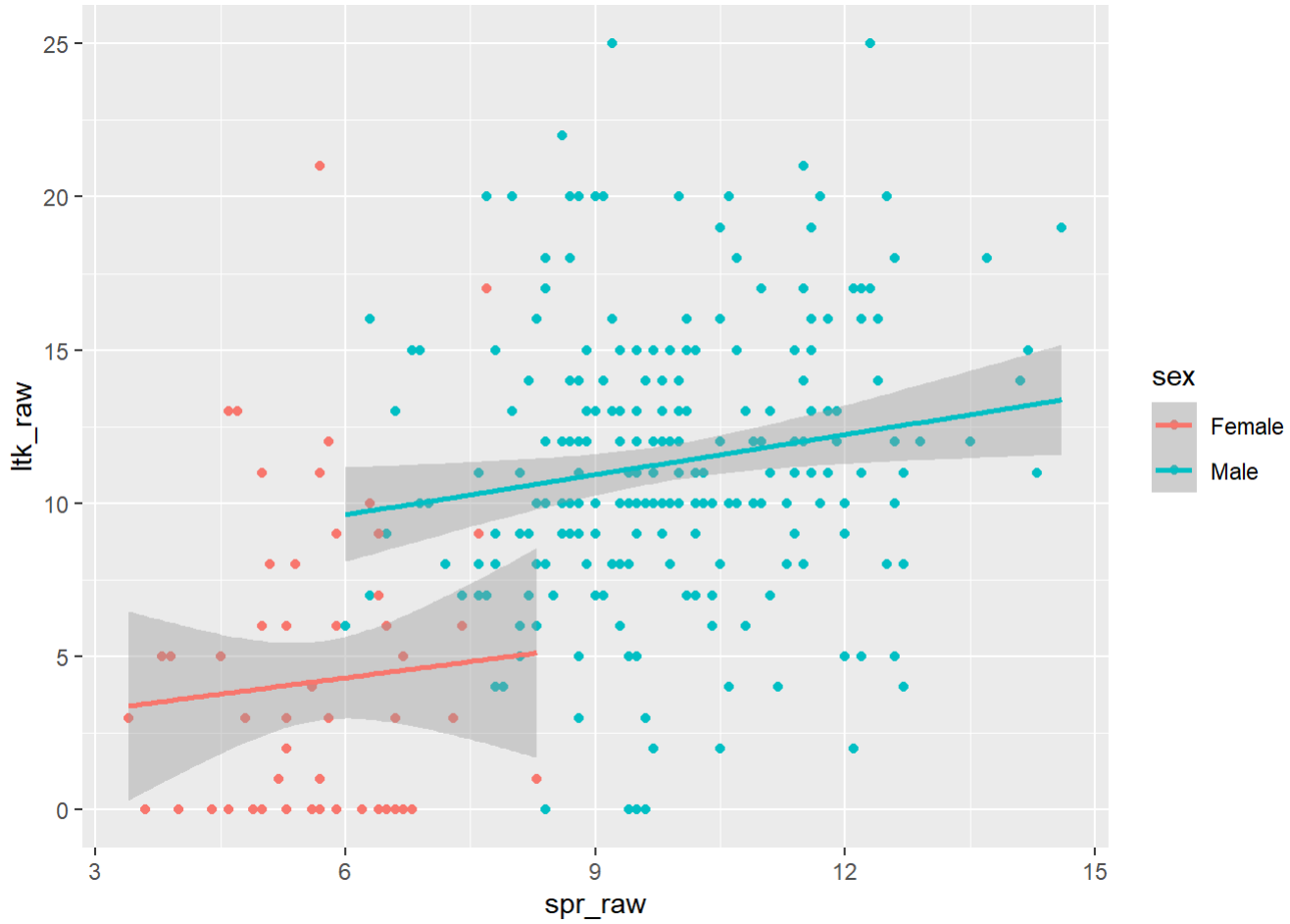
2. Can we determine causation? Why or why not.
   No, this is an observational study and not an experiment, therefore subjects were not randomly assigned to explanatory variable groups. There may be confounding variables influencing the associations.

Does sex change the association between the Standing Power Throw (SPT) and Leg Tuck (LTK) raw scores?

1. Draw the updated causal diagram.
   SPT $\rightarrow$ LTK plus and arrow from sex to both SPT and LTK
2. Write the equation for the linear model with beta's.
   $\hat{LTK} = \beta_0 + \beta_1 * SPT + \beta_2 * sex + \beta_3 * SPT * sex$
3. Give the hypotheses in symbols and words.
   $H_0 : \beta_3 = 0$ "Sex does not change the association between SPT and LTK." $H_A : \beta_3 \neq 0$ "Sex does change the association between SPT and LTK."
4. Plot the association, including the regression line.

Code

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



1. Create a linear model to test your hypotheses.

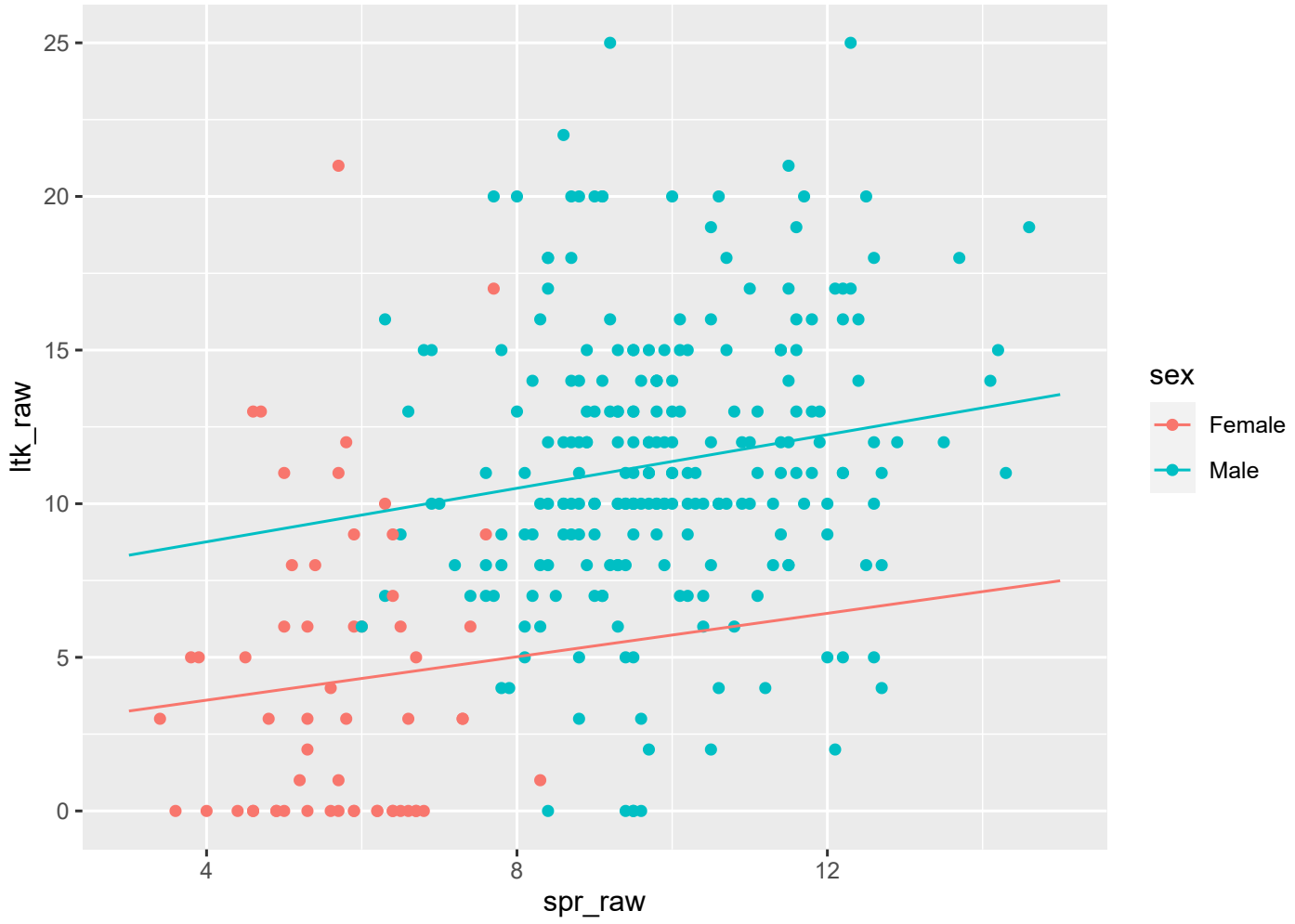
Code

```
##
## Call:
## lm(formula = ltk_raw ~ spr_raw * sex, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2001  -3.0803  -0.7206   2.7345  16.7943
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.19184    3.33436   0.657    0.511
## spr_raw          0.35331    0.57485   0.615    0.539
## sexMale          4.82354    3.82656   1.261    0.208
## spr_raw:sexMale  0.08261    0.60446   0.137    0.891
##
## Residual standard error: 4.624 on 288 degrees of freedom
## Multiple R-squared:  0.2878, Adjusted R-squared:  0.2803
## F-statistic: 38.79 on 3 and 288 DF,  p-value: < 2.2e-16
```

Code

```
## Warning: package 'ggiraphExtra' was built under R version 4.2.3
```

Code



1. Give the updated linear model with the coefficients determined by linear regression.
   $\hat{LTK} = 2.19 + 0.353 * SPT + 4.82 * sexMale + 0.0826 * SPT * sex$

2. Which level of *sex* is included in the base model?
   Female. By default the base level is the level that comes first alphabetically (ex. female before male). To change the base level for a categorical variable use the **relevel** function.
   Example: `df$sex = relevel(as.factor(df$sex), "Male")`

3. Interpret the intercept term.
   The predicted number of LTK's for a female cadet with an SPT of 0m is 2.19 reps.

4. Interpret the base slope term.
   For female cadets, for every additional SPT meter, we predict a 0.353 rep increase in LTK's, after adjusting for sex.

5. Interpret the base term for *sex*.
   For male cadets, we have a vertical adjustment of 4.82 LTK reps at the y-intercept.

6. Interpret the interaction term.
   For male cadets, we expect a 0.0826 increase in the base slope of 0.353. This means that for male cadets, for every additional SPT meter, we expect a (0.353 + 0.0826 = 0.4356) increase in LTK's.

We can think of the interaction term as a slope adjustment that is only turned on for male cadets.

1. Calculate the LTK reps for a male cadet with an SPT of 6m.

Code

```
##        1
## 9.630845
```

1. Calculate the LTK reps for a female cadet with an SPT of 6m.

Code

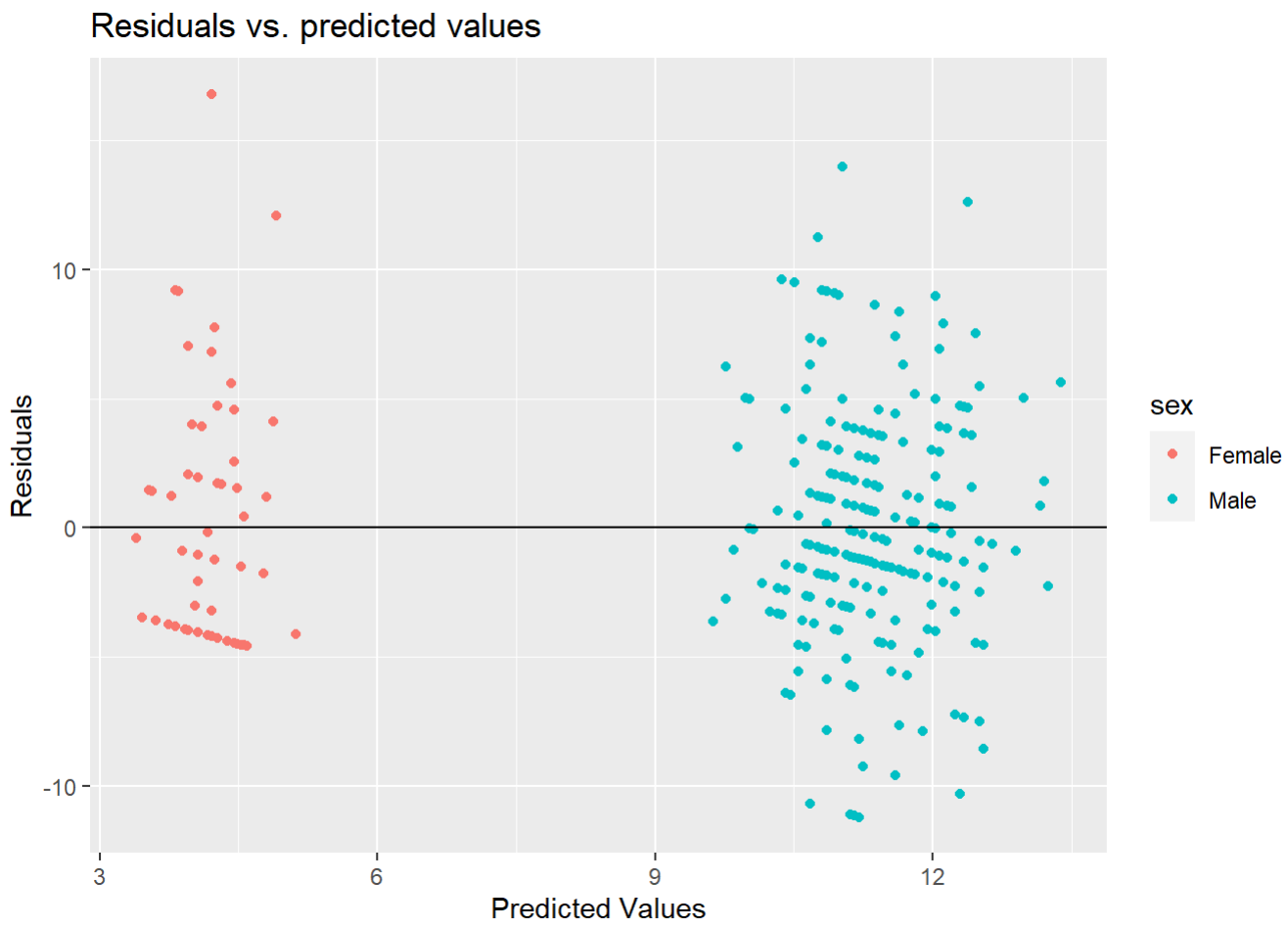
```
##       1
## 4.31167
```

1. Conclusion of your hypothesis test at a 5% significance level.
   We are looking at the p-value associated with the interaction term (0.891).
   $\text{p-value} \geq \alpha$ so we fail to reject H_0. It is plausible that sex does not affect the association between SPT and LTK raw scores.

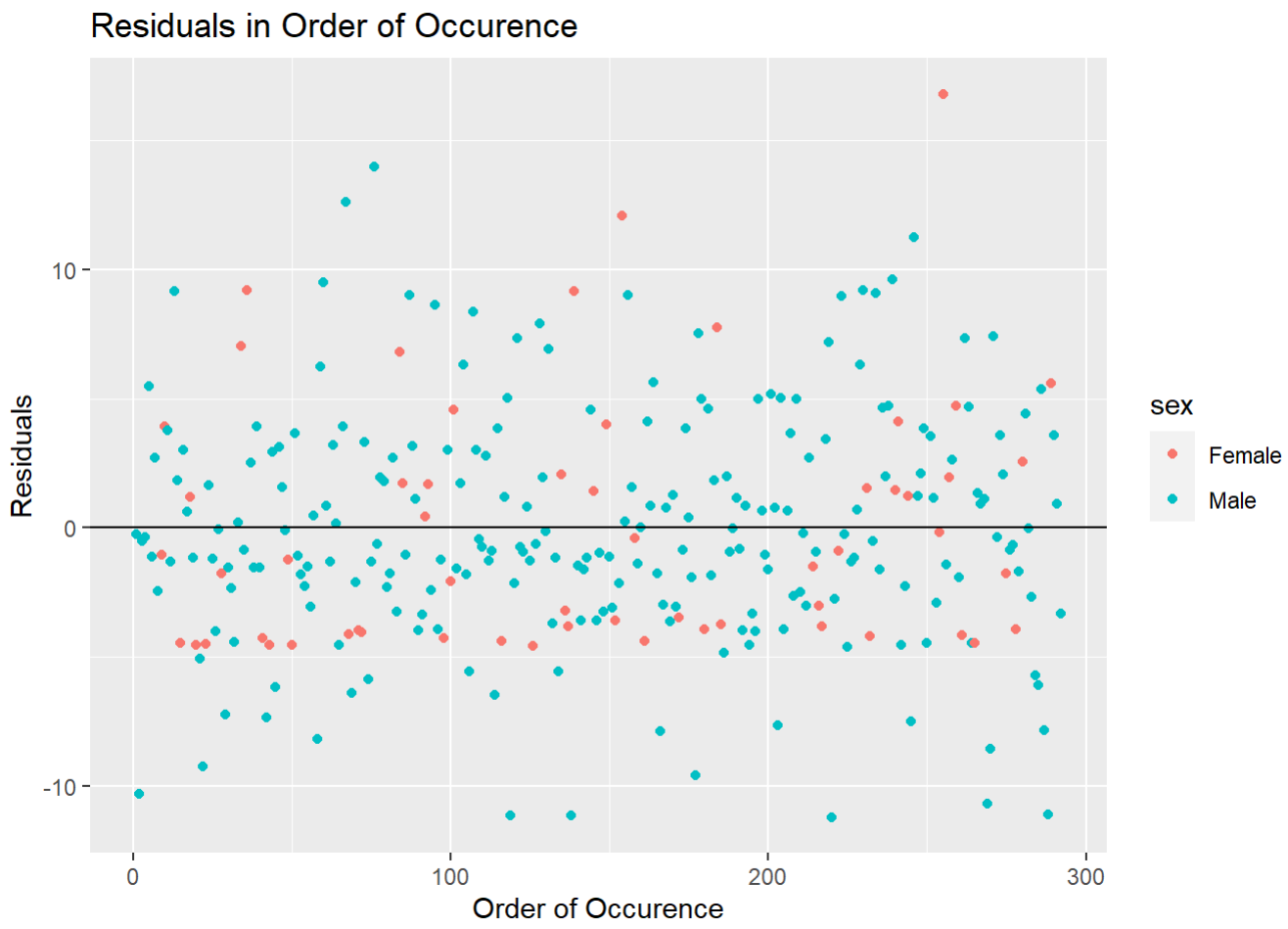
2. Interpret the R-squared value.
   Our model accounts for 28.8% of the variation in LTK reps.

3. Check the 4 Validity Conditions (L.I.N.E.). Are any not met?

Code

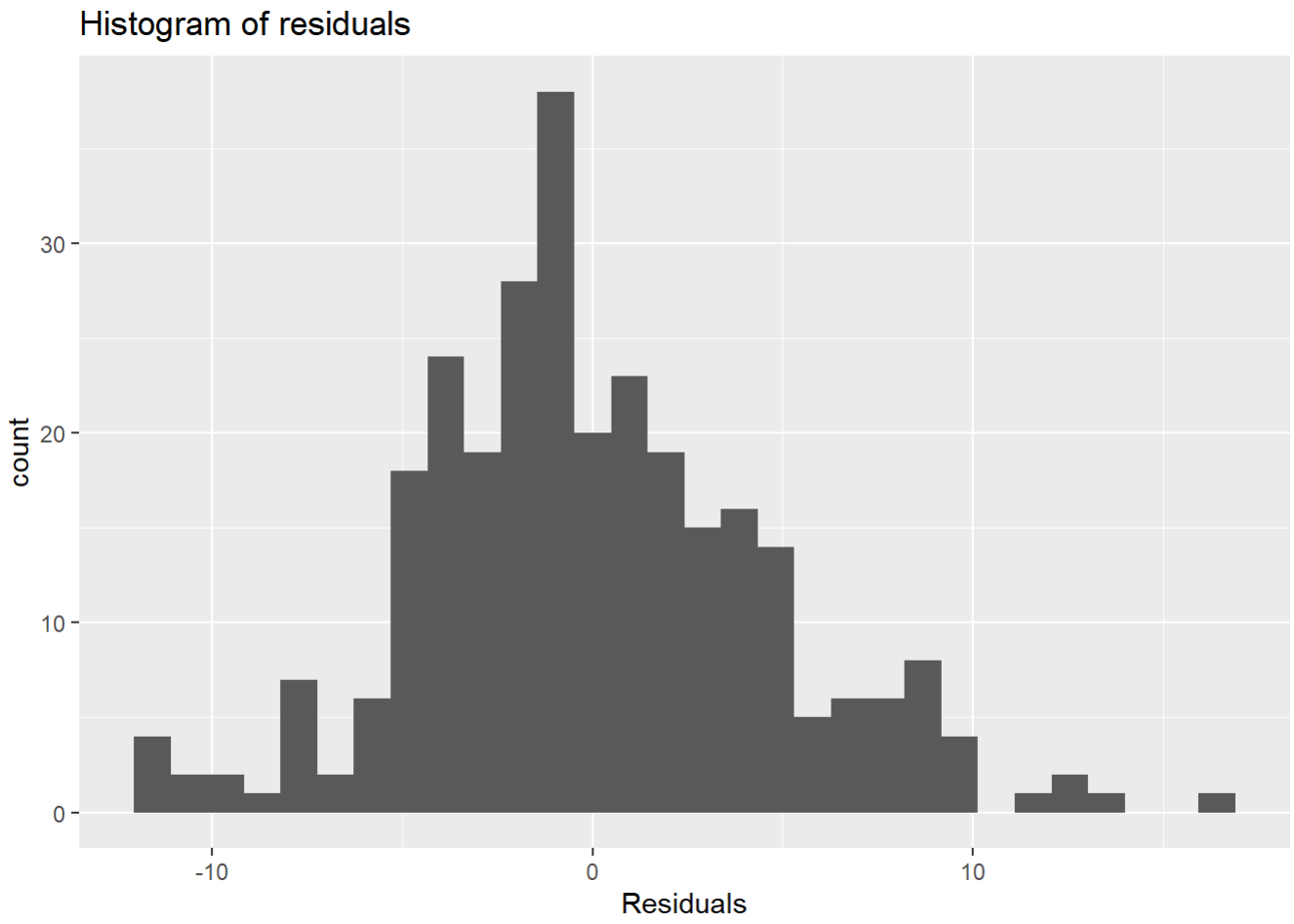

Residuals vs. predicted values
Residuals
10
0
-10
3
6
9
12
Predicted Values
sex
Female
Male

Code



Code

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



All four validity conditions are met.