Your Project Title (be descriptive and creative!)

You can either put a subtitle here, or delete this line

Team members names, tutorial groups, and group number (if applicable)

April 1, 2021

Question 1

What is the correlation between a member's age and sex and the total amount they spent at RCYC facilities?
Motivation: our goal is to help better adapt RCYC's facilities to the members' needs. Then, the first research question...

- explores whether age, sex, or both can predict a member's spending
- can reveal interesting patterns in spending

How does this help?

Based on these data, the managers of RCYC can make an informed decision on which facilities to invest in more

- perhaps in combination with other available, external data, such as an age group's preferred activities.
- for instance, if there is a correlation between age and spending, the managers can decide whether to focus efforts on retaining old members or recruiting new members, depending on the trend.

Statistical methods

To analyze the relationship, we use a linear regression model with two interacting predictors. We assume the conditions—linearity of relation, constancy of variance, independence of observation, and normality of residuals. Also, we create a column called `total_spending`—the sum of the separate accounts—and filter for 2017 values.

Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i$$

Where y_i is the predicted total spending in 2017, x_{1i} is the age of the member, x_{2i} is the sex of the member—where 0 is female and 1 is male.

Hypotheses

Let $\alpha = 0.05$.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \qquad H_A : \text{at least one } \beta_i \neq 0$$

In simple terms, unless the data shows otherwise, we expect that there is no linear relationship between age and spending or sex and spending.

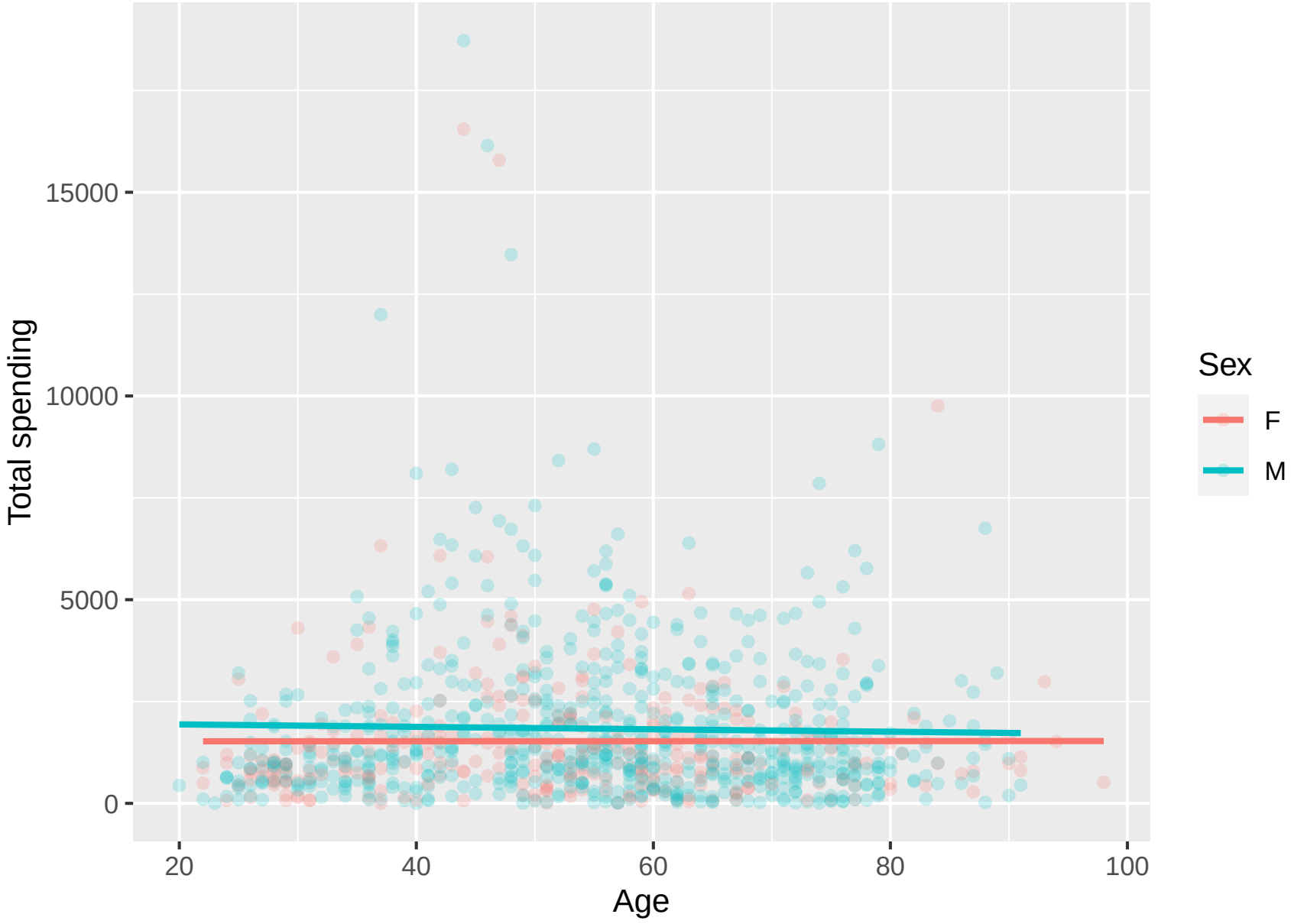

Figure 1: Relationship between member's age, sex, and total spending (2017)

Results

We obtain the following:

predictor	R^2	p-value
Age	0.0001070233	7.503991e-01
Sex	0.005616849	2.101313e-02
Age and Sex	0.00607189	7.033530e-01

With this low of R^2 and this high of p-values, we see that the linear relationship is extremely weak and further fail to reject the null hypothesis. Thus, we do not have sufficient evidence to conclude that at least one of Age or Sex can predict the total spending. In simple terms, **there most likely is not a correlation between age or sex and spending.**

Note: we see some interesting data points for members in their mid-forties—just when midlife crises begin. Purely speculativel—as we would need more analyses to come to a conclusion—the club may wish to advertise more to those of middle-age.