

Stream Crossing Risk Model

Chris Hoover

April 2, 2019

Model Specification

Available data on stream crossing risk is provided from John Ladd's JMT survey as a rating from 0-5 describing perceived difficulty of all stream crossings during the individual's trip. The stream crossing risk model will use this rating as the outcome, classifying any rating >1 as "risky" and any rating ≤ 1 as not risky (to generate binary outcome). Covariates in the model will include individual hiker characteristics including: [] and an estimate of

```
daily_snow_melt <- load_csv_from_googledrive("12qFuFl2Vj3jhJiGjjdviqT6xKw3oprZn")
person_crossings <- load_csv_from_googledrive("12P_z60jhEPqWLpYr2311w3fFZlMAzSbh")
person_covariates <- load_csv_from_googledrive("1CmPS7bhIHSsy-6dkjfsSuZmYoYXcG0e8")

merged_data <- person_crossings %>%
  left_join(person_covariates, by = c("person_id" = "UniqueID")) %>%
  left_join(daily_snow_melt, by = c("crossing_name" = "watershed",
                                   "crossing_date" = "Date"))

model_data <- merged_data %>%
  group_by(person_id) %>%
  summarise(overall_stream_challenge = first(overall_stream_challenge),
            peak_SWE_melt = max(SWE_melt),
            mean_SWE_melt = mean(SWE_melt),
            peak_melt_crossing = ifelse(peak_SWE_melt == 0, "Not Applicable",
                                       crossing_name[which(SWE_melt == peak_SWE_melt)]),
            peak_SWE = max(SWE),
            mean_SWE = mean(SWE),
            peak_SWE_crossing = ifelse(peak_SWE == 0, "Not Applicable",
                                       crossing_name[which(SWE == peak_SWE)]),
            gender = first(`Gender (compiled and best evidence)`),
            age = first(`Age at entry TH`),
            height_in = as.numeric(first(`What is your height? feet_y`))*12 +
                        as.numeric(first(`What is your height? inches_y`)),
            weight = first(`What did you weigh at the start and end of your hike? (Estimates are accepted)`),
            MPD = first(`Calc mean MPD`),
            StartDate = as.Date(first(StartDate)),
            start_week = week(StartDate)) %>%
  mutate(crossing_difficulty_binary = ifelse(overall_stream_challenge <= 1, 0, 1))

table(model_data$overall_stream_challenge)

##
##    0    1    2    3    4    5
## 313 443  95  23   3   2

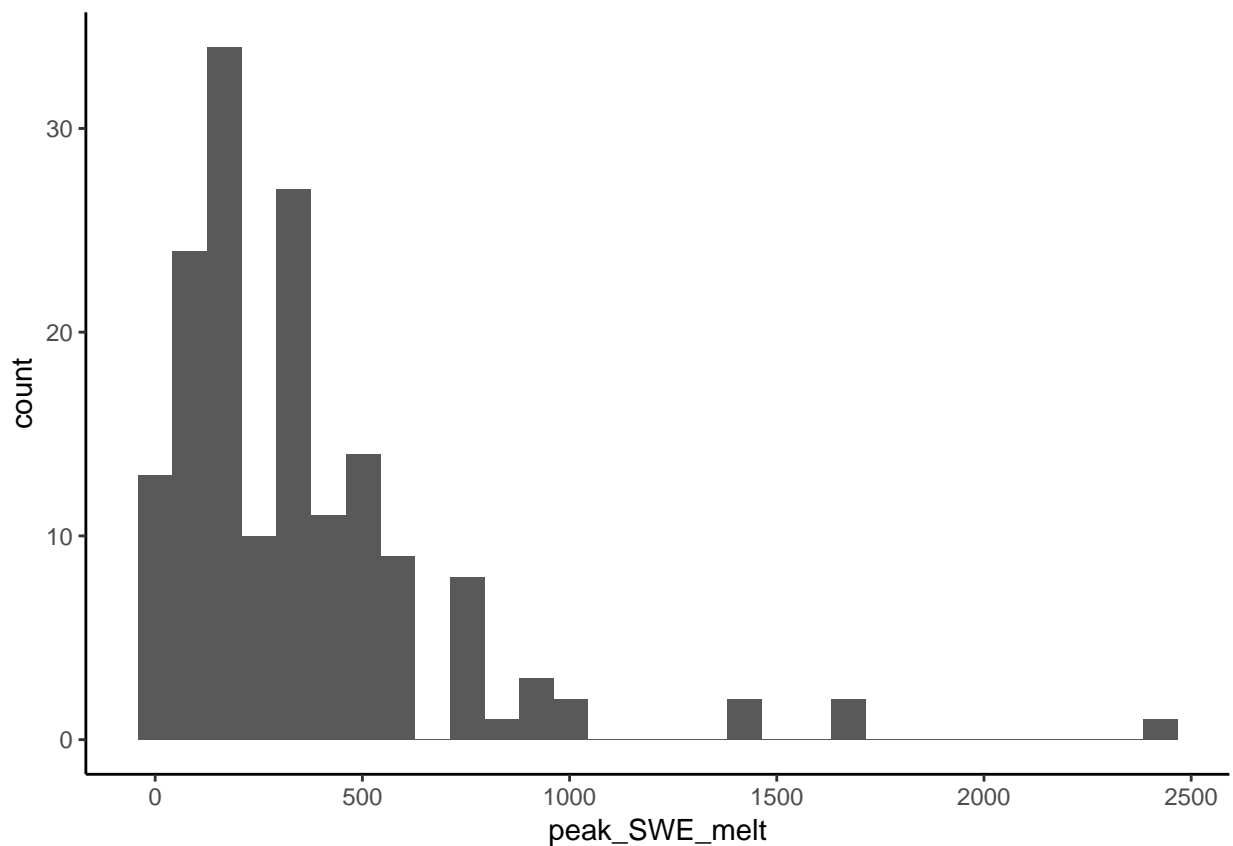
table(model_data$peak_SWE_crossing)
```

```
##
##      Arctic Lake outlet  Arrowhead Lake outlet      Bear Creek
##              3              3              2
##      Bubbs Creek        Deer Creek      Evolution Creek
##              4              16              7
##      Evolution Lake inlet  Helen Lake outlet  Hilgard Branch
##              5              3              21
##      Ireland Creek        Mott Creek      Not Applicable
##              61              2              712
## Silver Pass Creek lower  South Fork Kings River  Tyndall Creek
##              2              28              6
##      White Fork        Wright Creek
##              2              2
```

Individuals whose peak snow melt across all crossings was 0 (implying there was no snow melt) were labeled as “Not Applicable”. Among those who did experience snow melt at crossings, most appear to experience it at Ireland Creek, Hilgard Branch, South Fork Kings, Bear, and Bubbs creeks

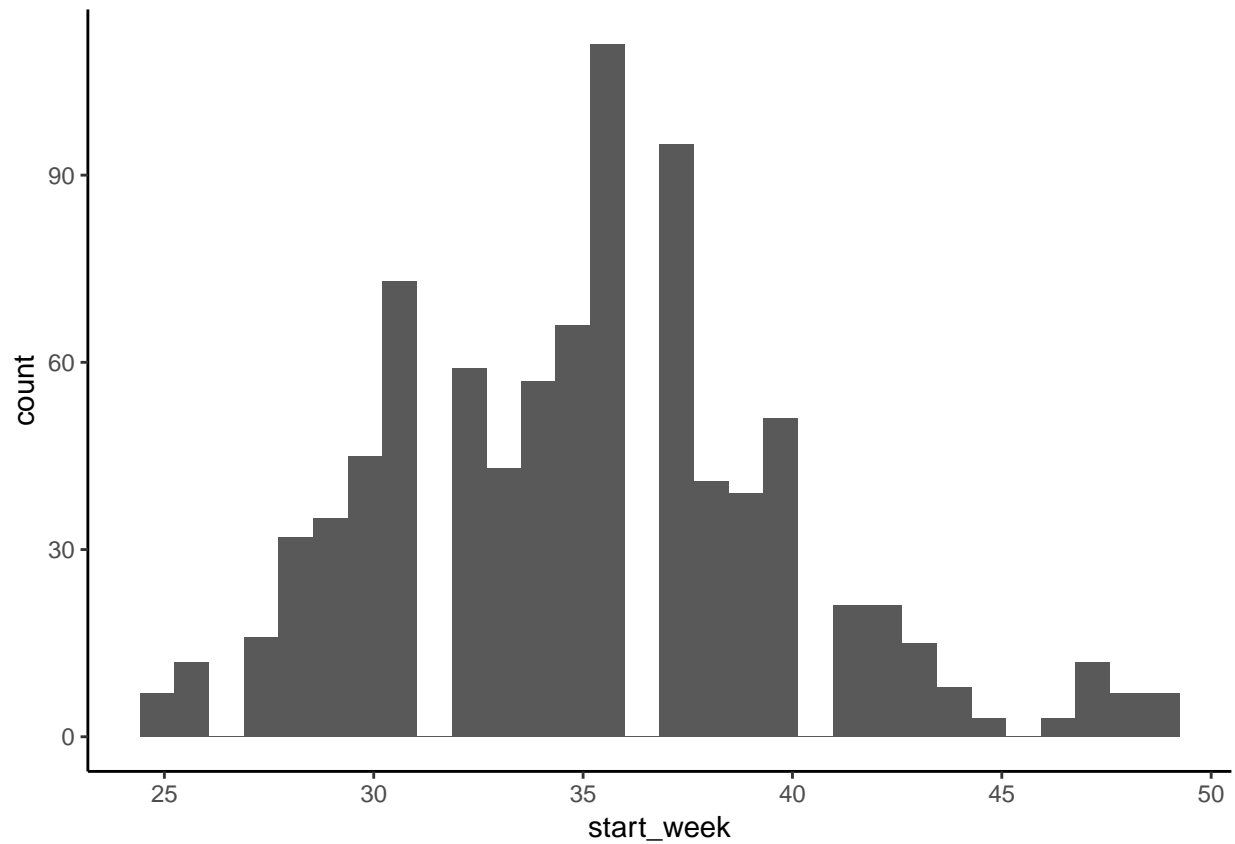
```
model_data %>%
  filter(peak_SWE_melt > 0) %>%
  ggplot(aes(peak_SWE_melt)) + geom_histogram() + theme_classic()
```

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



```
model_data %>%
  ggplot(aes(start_week)) + geom_histogram() + theme_classic()
```

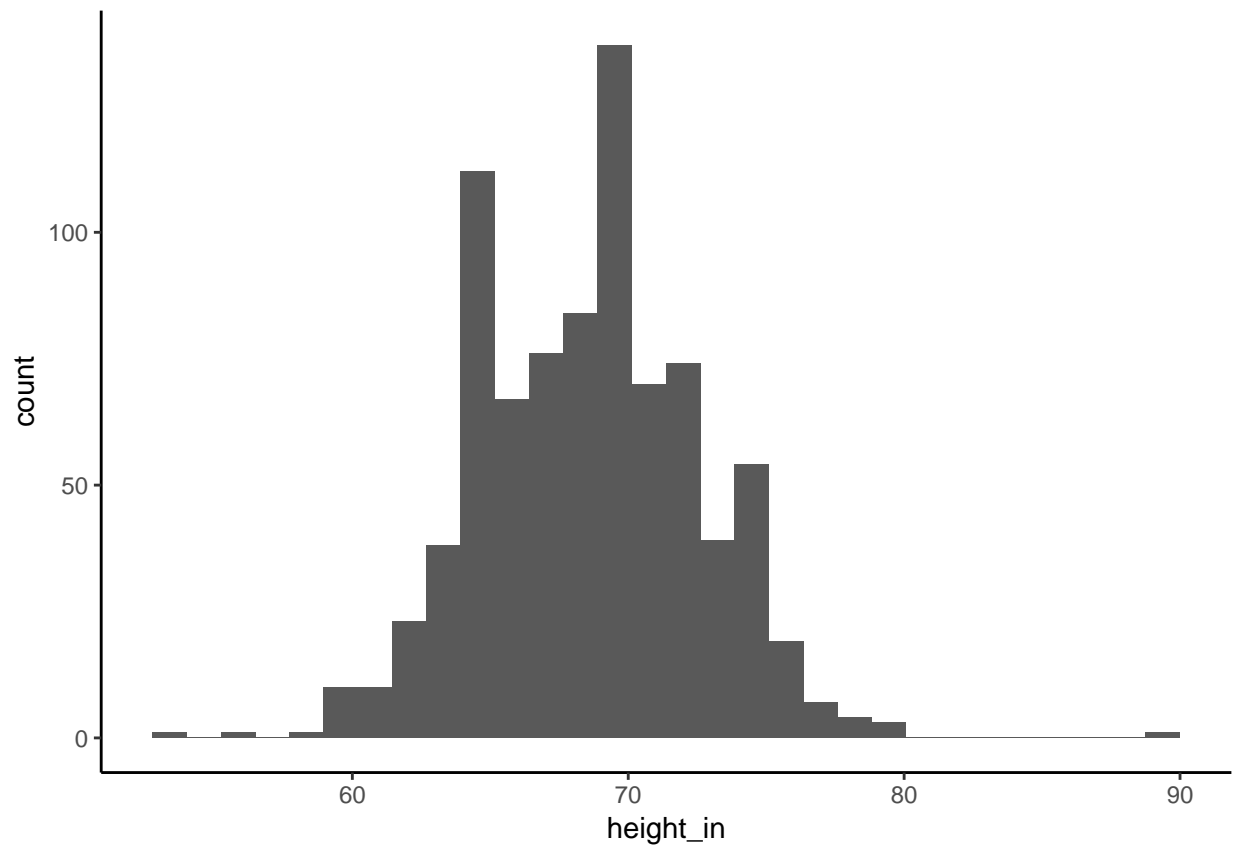
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
model_data %>%
  ggplot(aes(height_in)) + geom_histogram() + theme_classic()
```

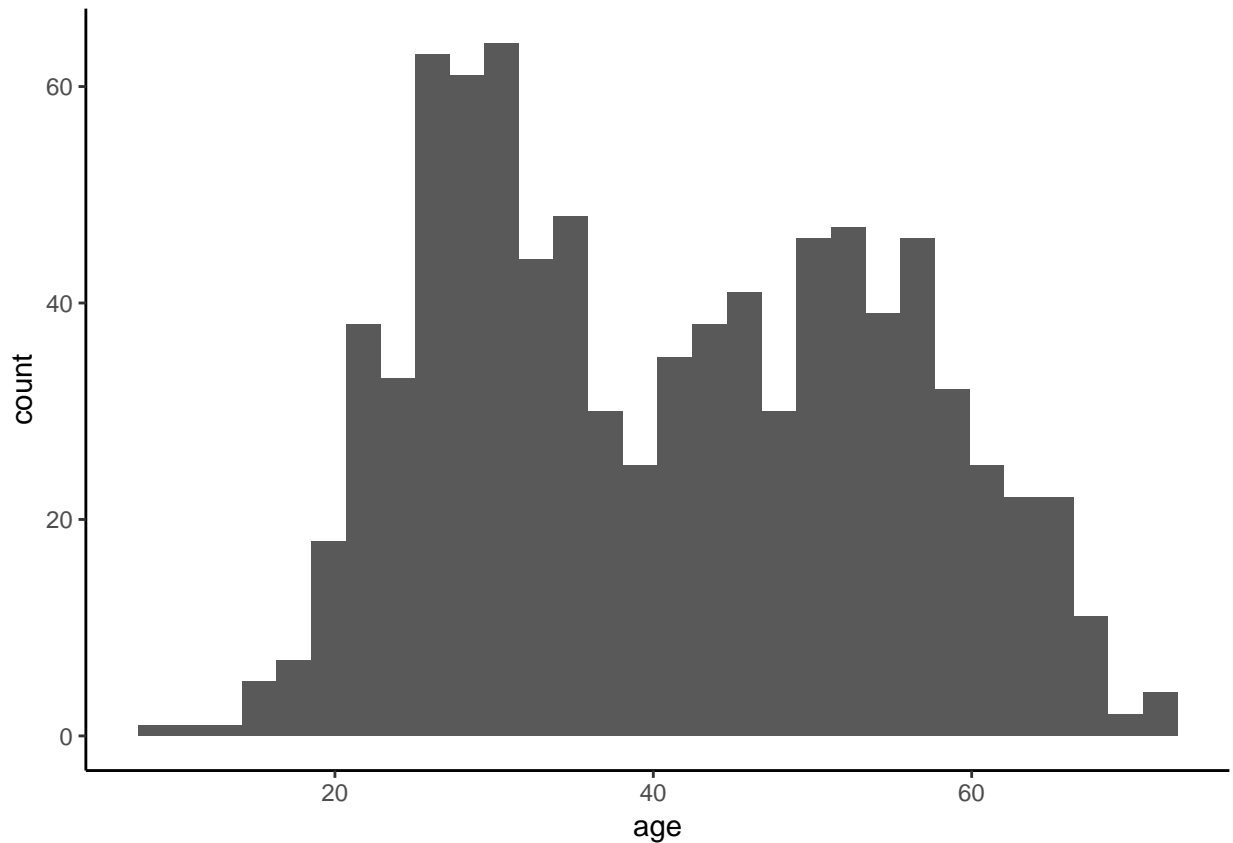
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 48 rows containing non-finite values (stat_bin).
```



```
model_data %>%  
  ggplot(aes(age)) + geom_histogram() + theme_classic()
```

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



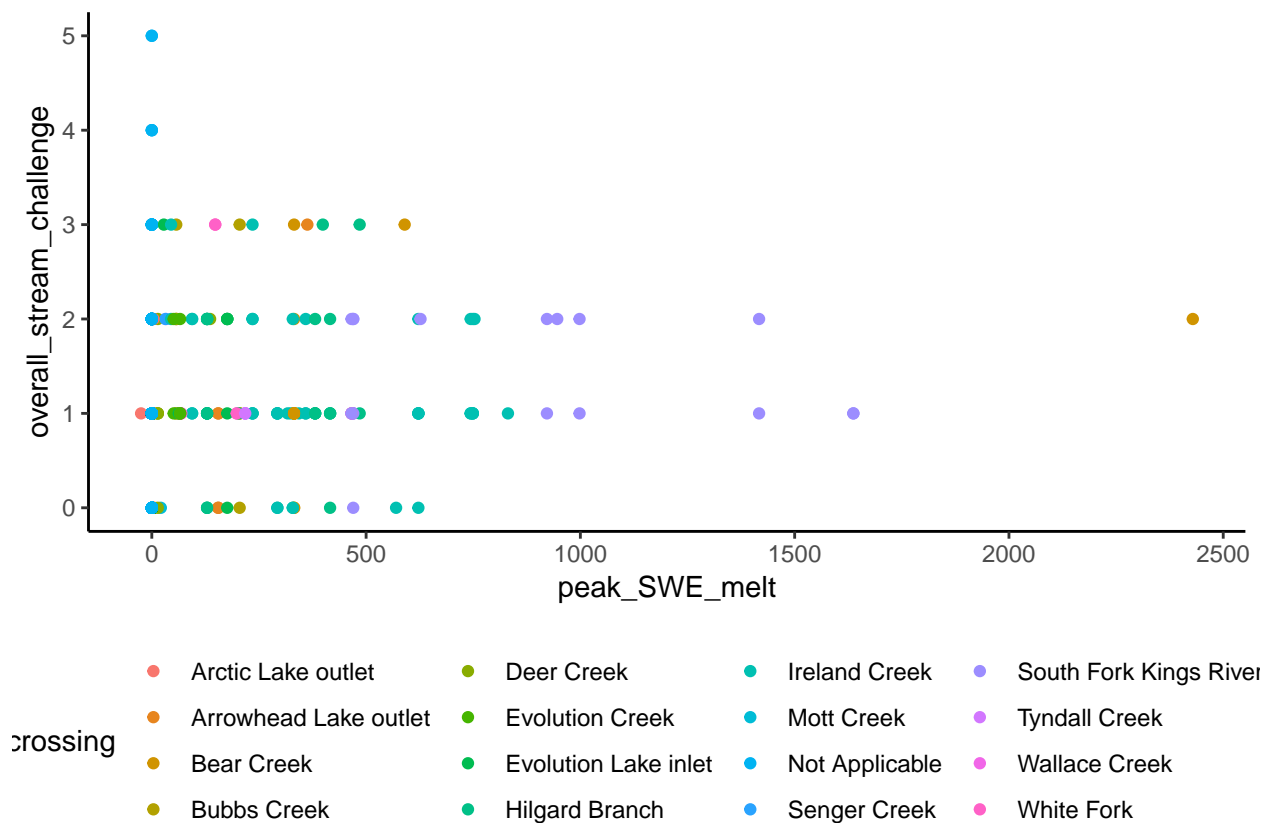
```
crossing_difficulty_bin_mod <- glm(crossing_difficulty_binary ~ peak_SWE_melt,
                                   family = "binomial", data = model_data)
summary(crossing_difficulty_bin_mod)
```

```
##
## Call:
## glm(formula = crossing_difficulty_binary ~ peak_SWE_melt, family = "binomial",
##      data = model_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6808  -0.5091  -0.5091  -0.5091   2.0530
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.9778279  0.1070191 -18.481  <2e-16 ***
## peak_SWE_melt  0.0019006  0.0004121   4.612   4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 711.71  on 878  degrees of freedom
## Residual deviance: 689.10  on 877  degrees of freedom
## AIC: 693.1
```

```
##
## Number of Fisher Scoring iterations: 4
```

Very small, but highly significant increase in odds of rating streams as >1 associated with increased now melt experienced. Let's look at a linear model, which assumes that rating 0-5 is continuous (which we know isn't true) but is worth checking out regardless.

```
model_data %>%
  ggplot(aes(x = peak_SWE_melt, y = overall_stream_challenge, col = peak_melt_crossing)) +
  geom_point() +
  theme_classic() +
  theme(legend.position = "bottom")
```



```
crossing_difficulty_lin_mod <- lm(overall_stream_challenge ~ peak_SWE_melt,
  data = model_data)
summary(crossing_difficulty_lin_mod)
```

```
##
## Call:
## lm(formula = overall_stream_challenge ~ peak_SWE_melt, data = model_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2587 -0.7752  0.2248  0.2248  4.2248
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7751899  0.0270002  28.711 < 2e-16 ***
## peak_SWE_melt 0.0007774  0.0001305   5.959 3.67e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7633 on 877 degrees of freedom
## Multiple R-squared:  0.03892,    Adjusted R-squared:  0.03782
## F-statistic: 35.51 on 1 and 877 DF,  p-value: 3.667e-09

crossing_difficulty_bin_mod_adj <- glm(crossing_difficulty_binary ~ peak_SWE_melt + gender + age + height_in + weight + start_week,
                                     family = "binomial", data = model_data)
summary(crossing_difficulty_bin_mod_adj)

##
## Call:
## glm(formula = crossing_difficulty_binary ~ peak_SWE_melt + gender +
##      age + height_in + weight + start_week, family = "binomial",
##      data = model_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3972  -0.5771  -0.4690  -0.3719   2.6408
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.559e+00  2.651e+00   2.851 0.004352 **
## peak_SWE_melt  1.083e-03  4.379e-04   2.473 0.013397 *
## genderM        5.429e-02  2.832e-01   0.192 0.848009
## age           -6.623e-05  7.824e-03  -0.008 0.993246
## height_in     -1.003e-01  4.075e-02  -2.460 0.013876 *
## weight        2.806e-03  4.069e-03   0.690 0.490453
## start_week    -9.066e-02  2.614e-02  -3.468 0.000525 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 668.79  on 820  degrees of freedom
## Residual deviance: 628.06  on 814  degrees of freedom
## (58 observations deleted due to missingness)
## AIC: 642.06
##
## Number of Fisher Scoring iterations: 5

crossing_difficulty_lin_mod_adj <- lm(overall_stream_challenge ~ peak_SWE_melt + gender + age + height_in + weight + start_week,
                                     data = model_data)
summary(crossing_difficulty_lin_mod_adj)

##
## Call:
## lm(formula = overall_stream_challenge ~ peak_SWE_melt + gender +
```

```

##      age + height_in + weight + start_week, data = model_data)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.2077 -0.7154  0.1123   0.2902   4.2087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9345306   0.6564227   4.470 8.91e-06 ***
## peak_SWE_melt  0.0005288   0.0001442   3.667 0.000261 ***
## genderM        0.0260216   0.0737936   0.353 0.724460
## age           -0.0002409   0.0019897  -0.121 0.903653
## height_in     -0.0210236   0.0099563  -2.112 0.035026 *
## weight         0.0006165   0.0010080   0.612 0.540957
## start_week    -0.0230049   0.0060904  -3.777 0.000170 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7613 on 814 degrees of freedom
## (58 observations deleted due to missingness)
## Multiple R-squared:  0.05775,    Adjusted R-squared:  0.0508
## F-statistic: 8.314 on 6 and 814 DF,  p-value: 9.237e-09

```