Stream Crossing Risk Model

Chris Hoover

April 2, 2019

Risk Model from Survey Data

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: weight, height, age, gender and start date

```
daily_snow_melt <- load_csv_from_googledrive("12qFuFl2Vj3jhJiGjjdviqT6xKw3oprZn")
person_crossings <- load_csv_from_googledrive("12P_z6OjhEPqWLpYr23l1w3fFZlMAzSbh")
person_covariates <- load_csv_from_googledrive("1CmPS7bhIHSsy-6dkjfsSuZmYoYXcGOe8")

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 accepta
            MPD = first(`Calc mean MPD`),
            Group_Size = first(`GrpSize AQ19  / Not counting any friends you first met on your hike, wha
           # Experience = first(),
            StartDate = as.Date(first(StartDate)),
            start_week = week(StartDate)) %>%
  mutate(crossing_difficulty_binary = ifelse(overall_stream_challenge <= 1, 0, 1),
         crossing_difficulty_prop = overall_stream_challenge / 5)
```

```
table(model_data$overall_stream_challenge)
```

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

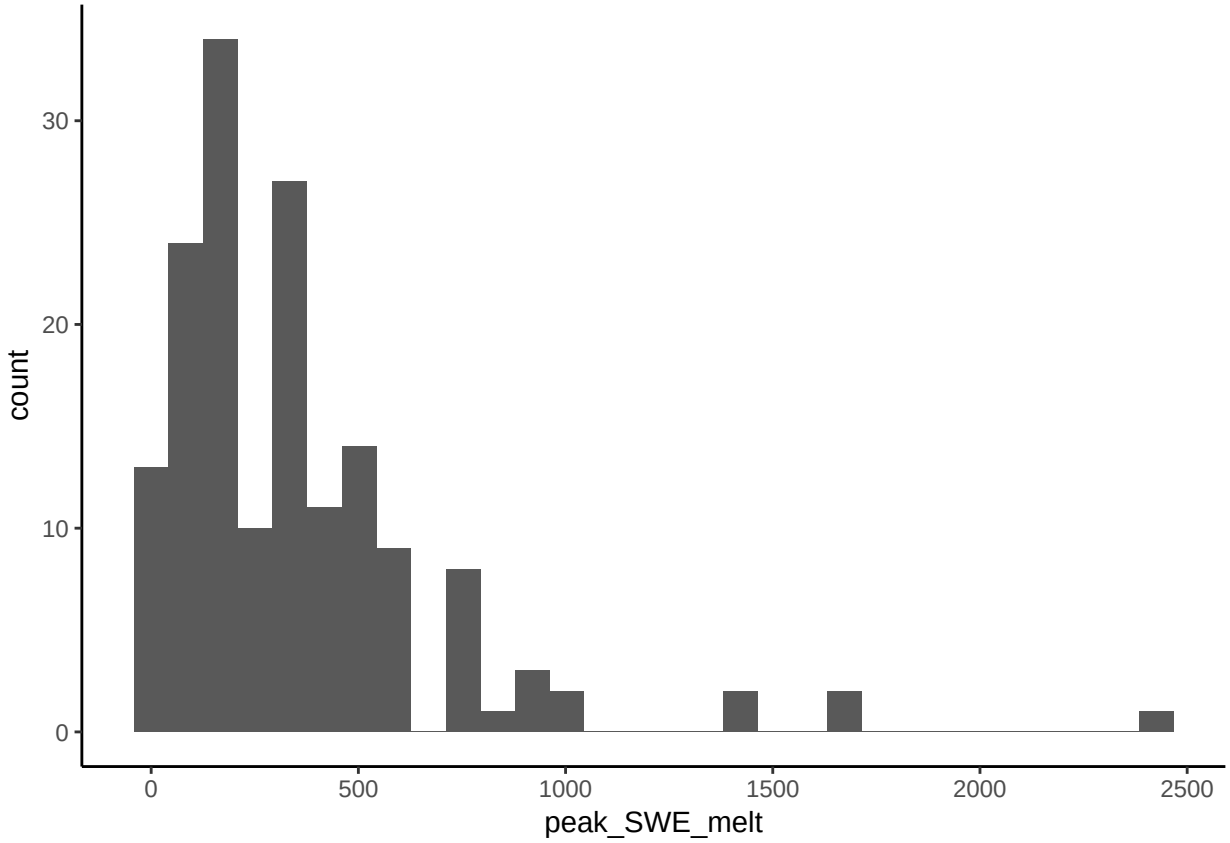
```
table(model_data$peak_SWE_crossing)
```

```
##
##      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

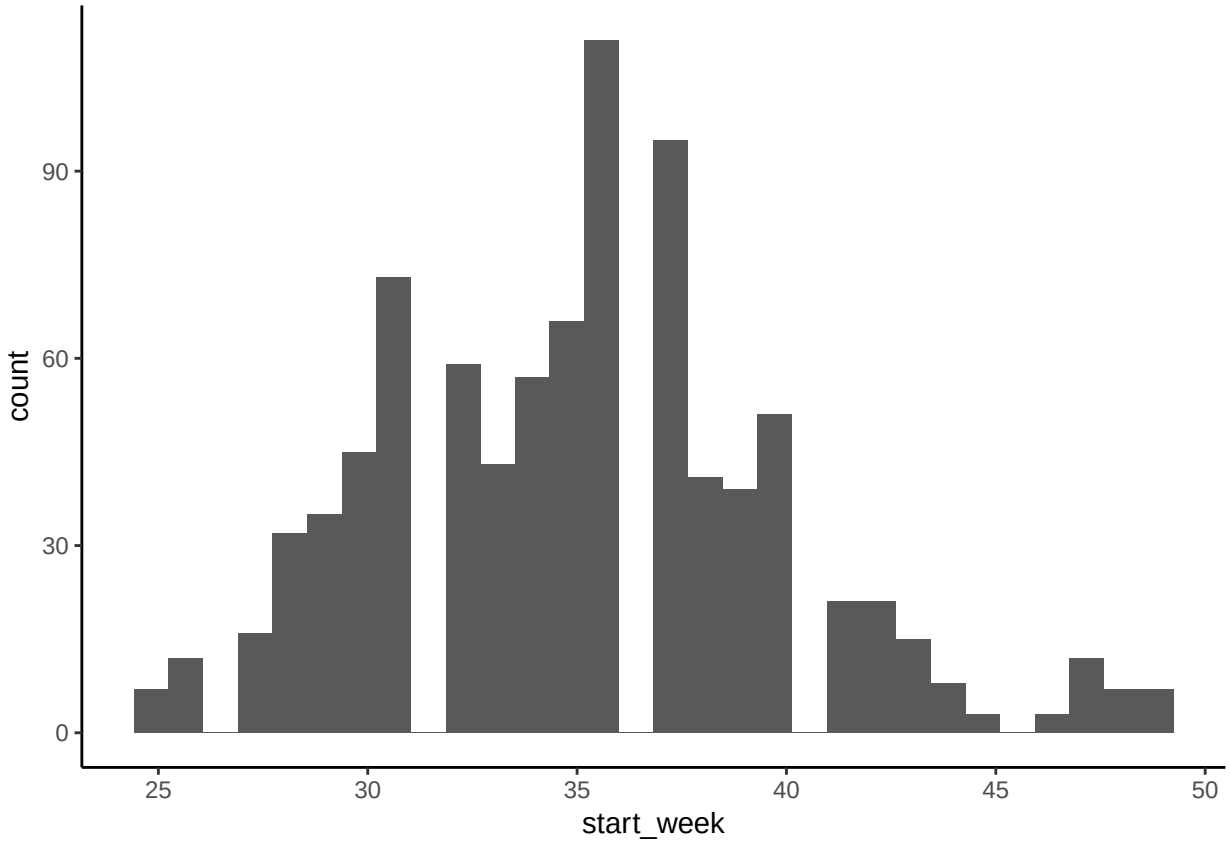
```
model_data %>%
  filter(peak_SWE_melt > 0) %>%
  ggplot(aes(peak_SWE_melt)) + geom_histogram() + theme_classic()
```

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



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

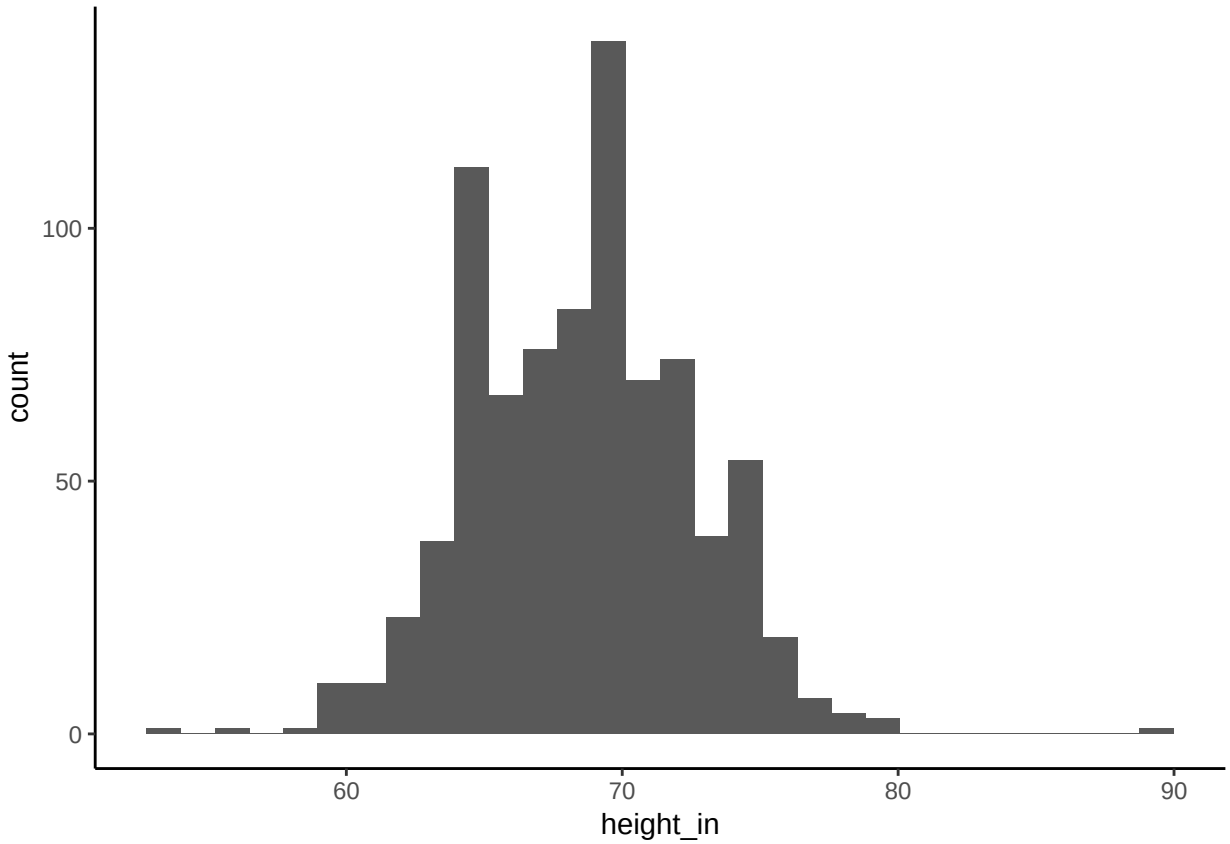
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

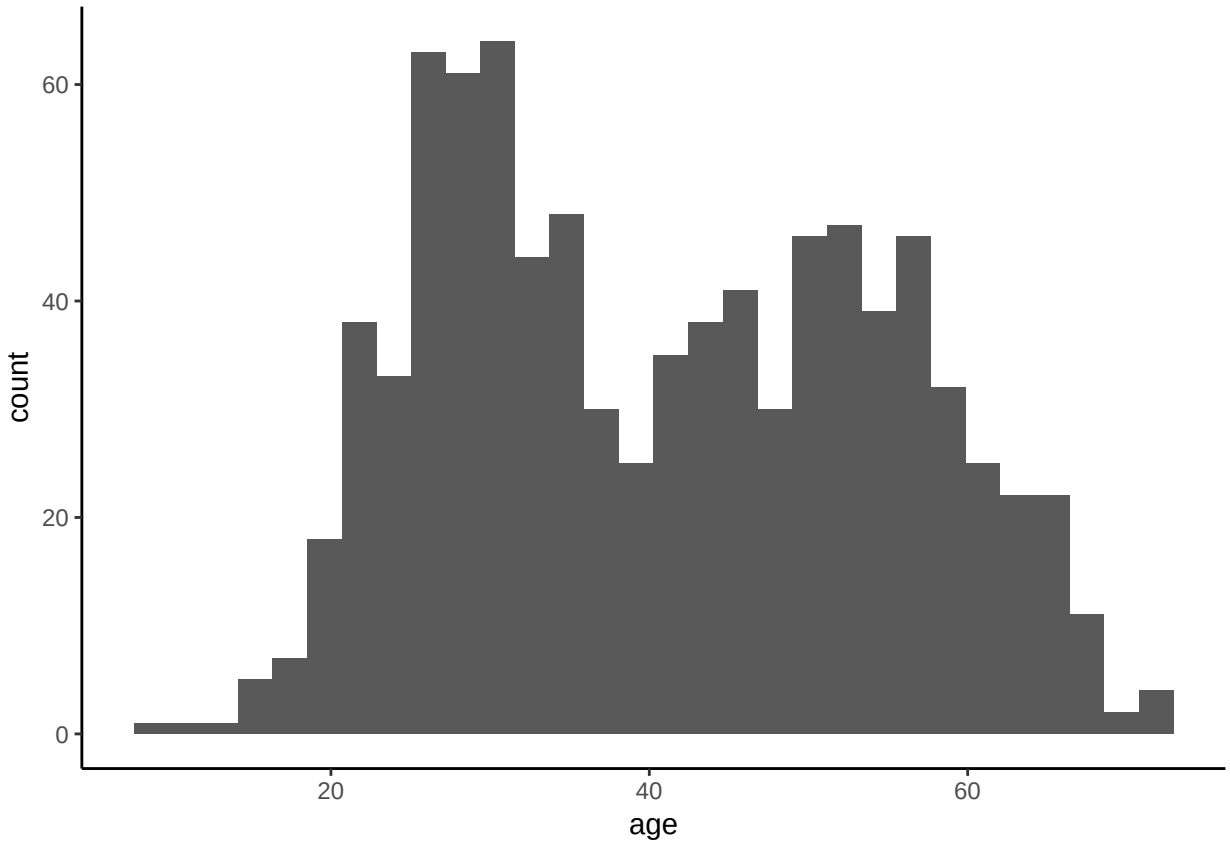
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## 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`.
```



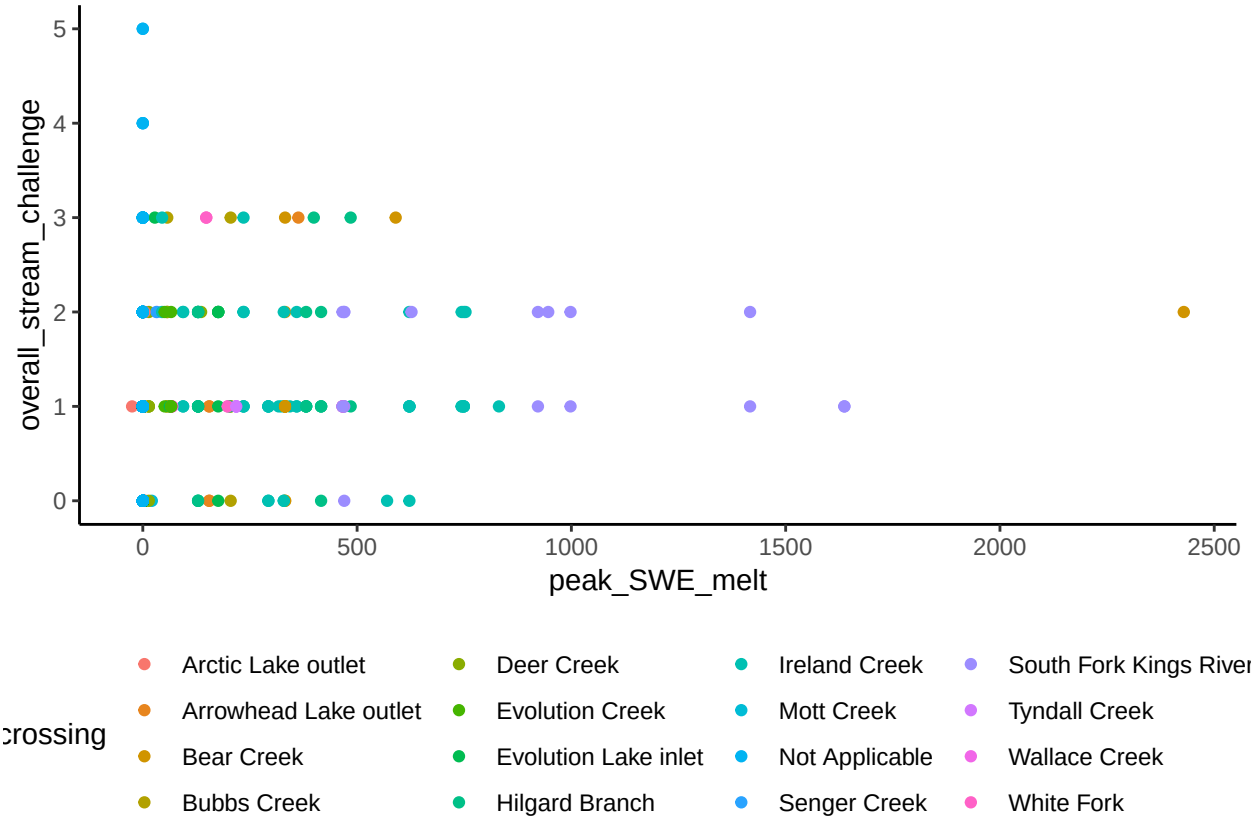

```
crossing_difficulty_bin_mod <- glm(crossing_difficulty_binary ~ peak_SWE_melt,
                                   family = "binomial", data = model_data)
  summary(crossing_difficulty_bin_mod)
```

```
##
## 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 + heigh
                                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 +
                                      MPD + Group_Size,
                                    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 + MPD + Group_Size,
##     data = model_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2686 -0.7045  0.0992  0.3011  4.1436
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.6783135  0.6669312   4.016 6.47e-05 ***
## peak_SWE_melt  0.0005127  0.0001442   3.555 0.000400 ***
## genderM        0.0076741  0.0742976   0.103 0.917759
## age            0.0007096  0.0020344   0.349 0.727344
## height_in     -0.0233988  0.0100117  -2.337 0.019674 *
## weight         0.0009818  0.0010208   0.962 0.336434
## start_week    -0.0217757  0.0061382  -3.548 0.000411 ***
## MPD            0.0193082  0.0087626   2.203 0.027840 *
## Group_Size     0.0218038  0.0164202   1.328 0.184595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7602 on 811 degrees of freedom
##   (59 observations deleted due to missingness)
## Multiple R-squared:  0.06407,    Adjusted R-squared:  0.05484
## F-statistic:  6.94 on 8 and 811 DF,  p-value: 7.219e-09
```

```
save(crossing_difficulty_lin_mod_adj, file = "risk_model_object.Rdata")
```

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

```
##
## Call:
## lm(formula = overall_stream_challenge ~ scale(peak_SWE_melt) +
##     gender + scale(age) + scale(height_in) + scale(weight) +
##     scale(start_week) + scale(MPD) + scale(Group_Size), data = model_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2686 -0.7045  0.0992  0.3011  4.1436
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.818200   0.048081  17.017  < 2e-16 ***
## scale(peak_SWE_melt)  0.101233   0.028479   3.555 0.000400 ***
## genderM               0.007674   0.074298   0.103 0.917759
```

```
## scale(age)            0.009721   0.027872   0.349 0.727344
## scale(height_in)     -0.093370   0.039950  -2.337 0.019674 *
## scale(weight)         0.036183   0.037620   0.962 0.336434
## scale(start_week)    -0.103040   0.029045  -3.548 0.000411 ***
## scale(MPD)            0.064177   0.029125   2.203 0.027840 *
## scale(Group_Size)     0.036579   0.027547   1.328 0.184595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7602 on 811 degrees of freedom
##   (59 observations deleted due to missingness)
## Multiple R-squared:  0.06407,    Adjusted R-squared:  0.05484
## F-statistic:  6.94 on 8 and 811 DF,  p-value: 7.219e-09
```

Temp-melt model

In order to do any sort of predictions, have to be able to relate input variables to variables that are forecasted (like temperature). Attampe here to model snow melt in watersheds as a function of antecedent SWE and temperature

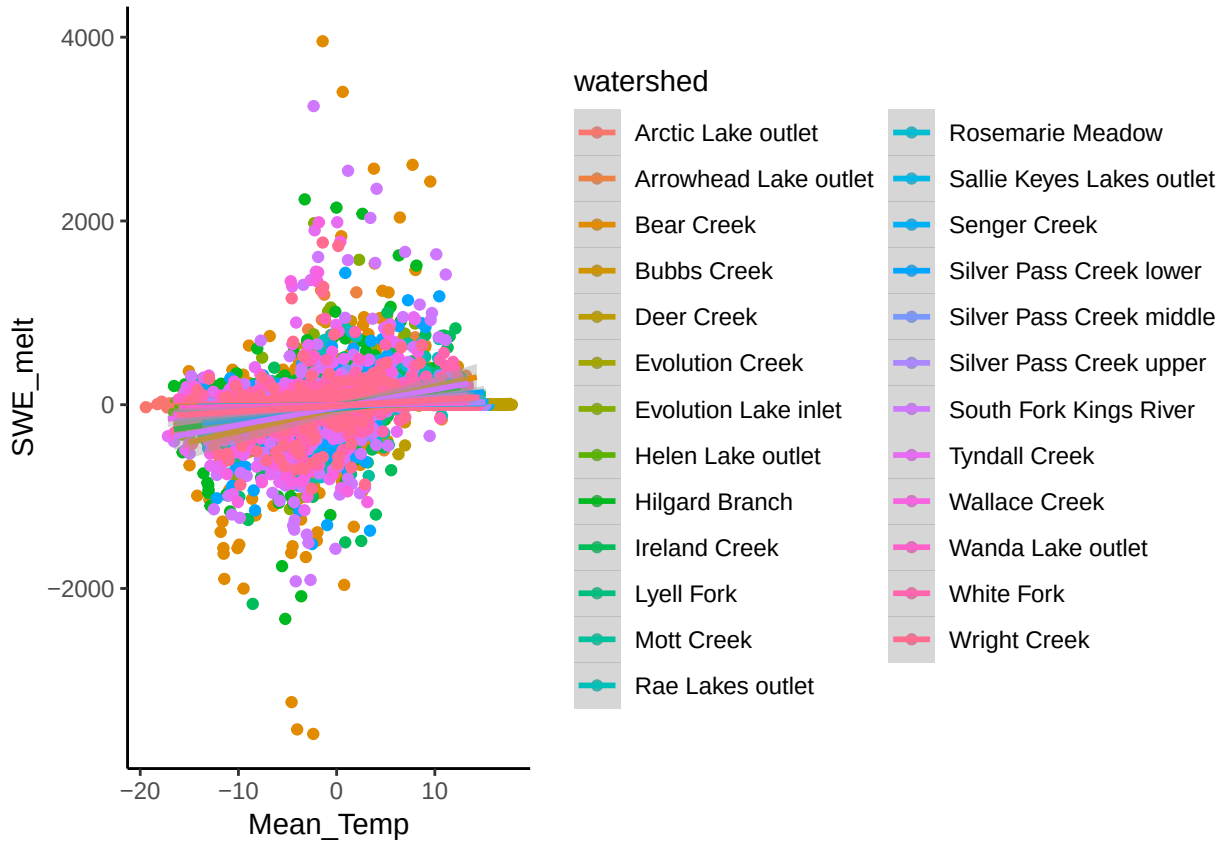
```
daily_temp <- load_csv_from_googledrive("1GAmIE-5zCUjrLJebs4G-u9mT0qniL6B4")
```

```
## File downloaded:
##   * jmt_watersheds_mean_temp_2015_long.csv
## Saved locally as:
##   * C:\Users\CHRIS_~1\AppData\Local\Temp\RtmpWcN2PH\file2d087eb5d6c.csv
```

```
## Parsed with column specification:
## cols(
##   Date = col_date(format = ""),
##   watershed = col_character(),
##   Mean_Temp = col_double()
## )
```

```
temp_melt_merge <- daily_temp %>%
  left_join(daily_snow_melt %>% filter(Date <= ymd("2016-01-01")), by = c("Date" = "Date",
                                                                          "watershed" = "watershed"))

temp_melt_merge %>% ggplot(aes(x = Mean_Temp, y = SWE_melt, col = watershed)) +
  theme_classic() +
  geom_point() +
  stat_smooth(method = "lm")
```




```
melt_glm <- lm(SWE_melt ~ SWE + Mean_Temp + watershed, data = temp_melt_merge)
  summary(melt_glm)
```

```
##
## Call:
## lm(formula = SWE_melt ~ SWE + Mean_Temp + watershed, data = temp_melt_merge)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3509.2   -60.7    -7.7    62.3  3981.8
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      6.815284  13.612265   0.501  0.61661
## SWE                             -0.005011   0.001937  -2.587  0.00971
## Mean_Temp                        6.408629   0.417463  15.351  < 2e-16
## watershedArrowhead Lake outlet -13.165396  19.519411  -0.674  0.50003
## watershedBear Creek             -7.481143  20.350357  -0.368  0.71317
## watershedBubbs Creek           -14.733405  19.393434  -0.760  0.44745
## watershedDeer Creek            -41.063081  19.508485  -2.105  0.03533
## watershedEvolution Creek       -18.950905  19.317339  -0.981  0.32660
## watershedEvolution Lake inlet   -9.459037  19.313543  -0.490  0.62431
## watershedHelen Lake outlet      -9.048403  19.254075  -0.470  0.63840
## watershedHilgard Branch         -2.338383  19.653485  -0.119  0.90529
## watershedIreland Creek         -20.191687  20.469374  -0.986  0.32395
```

```
## watershedLyell Fork                -19.086388  19.410486  -0.983  0.32548
## watershedMott Creek                -20.818155  19.374234  -1.075  0.28261
## watershedRae Lakes outlet          -15.158239  19.365808  -0.783  0.43380
## watershedRosemarie Meadow          -20.160650  19.397610  -1.039  0.29868
## watershedSallie Keyes Lakes outlet -22.157341  19.296730  -1.148  0.25090
## watershedSenger Creek              -24.544686  19.331961  -1.270  0.20424
## watershedSilver Pass Creek lower   -24.009320  19.603299  -1.225  0.22070
## watershedSilver Pass Creek middle  -24.105314  19.317947  -1.248  0.21213
## watershedSilver Pass Creek upper   -23.692877  19.309017  -1.227  0.21984
## watershedSouth Fork Kings River     -3.541736  20.270809  -0.175  0.86130
## watershedTyndall Creek              -6.310493  19.437038  -0.325  0.74544
## watershedWallace Creek              -9.249978  19.400361  -0.477  0.63352
## watershedWanda Lake outlet          -9.312508  19.254986  -0.484  0.62865
## watershedWhite Fork                -15.378442  19.445022  -0.791  0.42904
## watershedWright Creek              -11.894990  19.377908  -0.614  0.53933
##
## (Intercept)
## SWE                                **
## Mean_Temp                          ***
## watershedArrowhead Lake outlet
## watershedBear Creek
## watershedBubbs Creek
## watershedDeer Creek                *
## watershedEvolution Creek
## watershedEvolution Lake inlet
## watershedHelen Lake outlet
## watershedHilgard Branch
## watershedIreland Creek
## watershedLyell Fork
## watershedMott Creek
## watershedRae Lakes outlet
## watershedRosemarie Meadow
## watershedSallie Keyes Lakes outlet
## watershedSenger Creek
## watershedSilver Pass Creek lower
## watershedSilver Pass Creek middle
## watershedSilver Pass Creek upper
## watershedSouth Fork Kings River
## watershedTyndall Creek
## watershedWallace Creek
## watershedWanda Lake outlet
## watershedWhite Fork
## watershedWright Creek
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 259.6 on 9073 degrees of freedom
## Multiple R-squared:  0.03414,    Adjusted R-squared:  0.03137
## F-statistic: 12.33 on 26 and 9073 DF,  p-value: < 2.2e-16
```

```r
# Looks like no significant differences between watersheds, so stick with simpler GLM

melt_glm2 <- lm(SWE_melt ~ SWE + Mean_Temp, data = temp_melt_merge)
  summary(melt_glm2)
```

```
##
## Call:
## lm(formula = SWE_melt ~ SWE + Mean_Temp, data = temp_melt_merge)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3507.1   -59.0    -7.9    61.1  3988.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.910199   3.463621  -2.573  0.01011 *
## SWE         -0.004510   0.001611  -2.800  0.00511 **
## Mean_Temp    6.259199   0.401184  15.602  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 259.4 on 9097 degrees of freedom
## Multiple R-squared:  0.03304,    Adjusted R-squared:  0.03283
## F-statistic: 155.4 on 2 and 9097 DF,  p-value: < 2.2e-16
```

Now see how well model predicts daily snow melt in next year

```
daily_temp2016 <- load_csv_from_googledrive("10_0uB8Qxv-AOZbI0MjjHgSnmeP7397D2")
```

```
## File downloaded:
##   * jmt_watersheds_mean_temp_2016_long.csv
## Saved locally as:
##   * C:\Users\CHRIS_~1\AppData\Local\Temp\RtmpWcN2PH\file2d08584c9d.csv

## Parsed with column specification:
## cols(
##   Date = col_date(format = ""),
##   watershed = col_character(),
##   Mean_Temp = col_double()
## )
```

```
daily_snow_melt2016 <- load_csv_from_googledrive("10ATgwIxJ8cL_A9JcG9N3YajH8E4E087y")
```

```
## File downloaded:
##   * jmt_watersheds_SWE_2016_long.csv
## Saved locally as:
##   * C:\Users\CHRIS_~1\AppData\Local\Temp\RtmpWcN2PH\file2d0854704ce6.csv

## Parsed with column specification:
## cols(
##   Date = col_date(format = ""),
##   watershed = col_character(),
##   SWE = col_double(),
##   SWE_melt = col_double()
## )
```

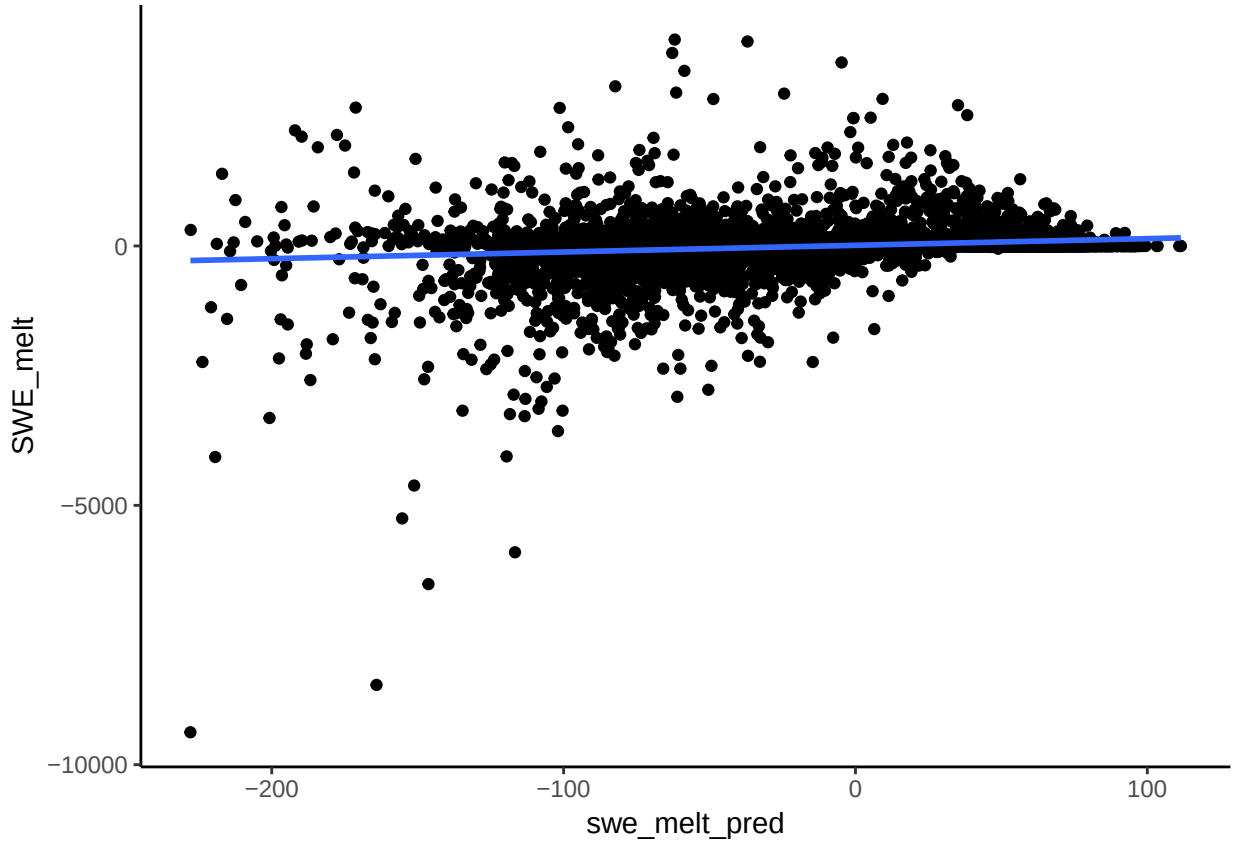
```
temp_melt_merge2016 <- daily_temp2016 %>%
  left_join(daily_snow_melt2016, by = c("Date" = "Date",
                                        "watershed" = "watershed"))

temp_melt_merge2016 <- temp_melt_merge2016 %>%
  mutate(swe_melt_pred = predict.lm(melt_glm2, newdata = temp_melt_merge2016))

temp_melt_merge2016 %>%
  ggplot(aes(x = swe_melt_pred, y = SWE_melt)) +
    theme_classic() +
    geom_point() +
    stat_smooth(method = "lm")
```

```
## Warning: Removed 25 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```




```
temp_melt_merge2016 %>%

  ggplot(aes(x = Date, y = SWE_melt)) +
    theme_classic() +
    geom_line() +
    geom_line(aes(x = Date, y = swe_melt_pred), lty = 2) +
    facet_wrap(~watershed)
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```