Weather Data and Time Series Emotions Among High School Students

Marty Moesta

5/4/2017

Introduction

The rising awareness of developing and understanding emotional intelligence in pre-college education has created a large opportunity for data scientists to uncover ways in which student's emotions are affected and the dynamic nature of their existence. For this project, I intend to address two separate questions: *How do emotions in high school students at time t affect their emotions at time t+1?* & *How does the weather affect student's emotions at a given time period?* Insights gained from these analyses can affect the way teachers think about their students emotions, and further improve and optomize pre-college education.

The Dataset and Programming Language

In order to find a dataset that would help in answering the above questions, I collaborated with Julia Moeller in the Emotional Intelligence department. With her help, I procured a dataset of emotional responses from 472 different high school students in the Connecticut area using the Emotional Sampling Method (ESM) (18,610 responses in total). Students were prompted to record their emotions three times per school day on their cell phone from May 9, 2016 to June 10, 2016. The average student age was 15.8, with 71.2% of sample responses coming from females. Grade levels are fairly equally represented, with the exception of 13.1% of responses coming from high school seniors (average: 25%). 13.8% of students came from low-income schools, and 86.2% came from middle-to-high income schools. Students were offered a $40 Amazon gift card if they participated in more than 90% of the available surveys. Responses were to the question: On a scale of 1 to 4, how (Emotion) do you feel in this moment? For more information on the dataset, please see the full R Markdown script.

In answering the second part of my question, I needed to obtain detailed weather reports. To achieve this, I requested exports from the National Oceanic and Atmospheric Association (NOAA) website: https://www.ncdc.noaa.gov/wct/. In addition, I retrieved humidity and visibility measures from Weather Underground: https://www.wunderground.com/. I chose to examine weather from Meriden Markham Municipal Airport and Sikorsky Municipal Airport, as one of those locations fell within 15 miles from all of the schools surveyed.

All data analysis is performed using the R statistical computing language, with RStudio as an IDE.

NOTE: For brevity's sake, some non-essential code will not be included in this document. However, a full RMarkdown script containing all code will come attached to this report.

Part 1: Time-Series Emotional Management

Before I would be able to run any analysis on the datset, substantial data cleaning had to be performed, in order to get the data into an operable state. One row of data is equivalent to one survey result.

```
## Reading in data from SPSS file and subsetting with columns that are necessary
## install.packages("foreign")
library(foreign) #Library that helps to read in SPSS files
options(warn=-1) #Non-vital warnings are supressed
x <- read.spss(file = "dataset.sav", to.data.frame = T)
y <- x[x$BG_OR_ESM=="ESM",]
```

```
beep <- y[,709:740]
beep[6,]  # An example of a row of data
```

```
##     BG_OR_ESM EXPERIENCESAMPLINGVARIABLESBELOW Participant    Date
## 519       ESM                              NA       19442 5/10/16
##           Day        Time         Session_Name Responded Completed_Session
## 519 Tuesday   10:32:39_AM Day_survey                   1                 1
##     Session_Instance Location Act_School Act_Home Act_Other LessonSubj
## 519                1        1          2       NA        NA          3
##     enthusiast_ESM happy_ESM interested_ESM curious_ESM calm_ESM
## 519              2         4              4           1        3
##     relaxed_ESM frustrated_ESM anxious_ESM afraid_ESM tired_ESM sad_ESM
## 519           2              2           2          2         2       2
##     bored_ESM stressed_ESM challenge_ESM skills_ESM choice interact1
## 519         3            2             3          3      3         1
```

Next, emotions are anonymized A-P, and a columns A2-P2 are intialized. A2-P2 will contain a participant's emotion A-P at time *t+1*. The **date_coded** column of the data frame will be used shortly to identify the relative date of an individuals series of responses.

Some entries were taken in the evening, however since the focus is to examine student's emotions in an education setting, those entries are removed.

Next, the data frame is split into a list, sorted by their participant ID. This ID is a way to collate all of the ESM responses from a given individual.

```
beep2 <- split(beep,beep$Participant)
```

The loop below fills in the **date_coded** column of the data frame, starting with '1' for each survey response recorded on the day a participant began the survey. This transormation just makes it easier to calculate *t vs. t+1* correlations. An example of date v. date coded is found below.

```
## Coding dates
M <- length(beep2) # Number of participants, 472
for(i in 1:M){
  N <- dim(beep2[[i]])[1] # Number of responses from an individual 'i'
  temp <- 1 # To signify 1st entry from a survey
  init <- as.numeric(beep2[[i]]$Date[1]) # The first date that someone responds to a survey
  for(j in 1:N){
    tempdate <- as.numeric(beep2[[i]]$Date[j])
    if(tempdate > init){ #If the date on entry j is the next day
      init <- tempdate # The counter 'init' become the new date
      temp <- temp + 1
    }
    beep2[[i]]$date_coded[j] <- temp
  }
}
beep2[[4]]$Date[1:10]
```

```
##  [1] 5/10/16 5/10/16 5/10/16 5/10/16 5/11/16 5/11/16 5/12/16 5/12/16
##  [9] 5/13/16 5/13/16
## 28 Levels:         5/10/16 5/11/16 5/12/16 5/13/16 5/16/16 ... 6/9/16
```

```
beep2[[4]]$date_coded[1:10]
```

```
##  [1] 1 1 1 1 2 2 3 3 4 4
```

Looping through each participant and then their survey responses, A2-P2 values are recorded. If consecutive

entries occur on different days, then A2-P2 are not recorded, since the interest is focused more on the within-day relationship between emotions. Including inter-day emotional carryover could pose problems to any underlying conclusions the data may present.

```
## Loop for filling in emotions at time t+1
for(i in 1:M){ # Number of participants
  N <- dim(beep2[[i]])[1] # Number of responses from participant 'i'
  if(N > 1){ # Cannot record emotion at t+1 if there is only one response
    for(j in 1:(N-1)){
      if(beep2[[i]]$date_coded[j]==
         beep2[[i]]$date_coded[j+1]){ # If the next session is in the same day
        beep2[[i]][j,35:50] <-
          beep2[[i]][(j+1),16:31]  # Todays A2-P2 equals next session's A-P
      }
    }
  }
}
```

In order to interpret how well certain emotions are at predicting others, a coefficient matrix is created. This matrix records all beta coefficient values of the multiple regression:

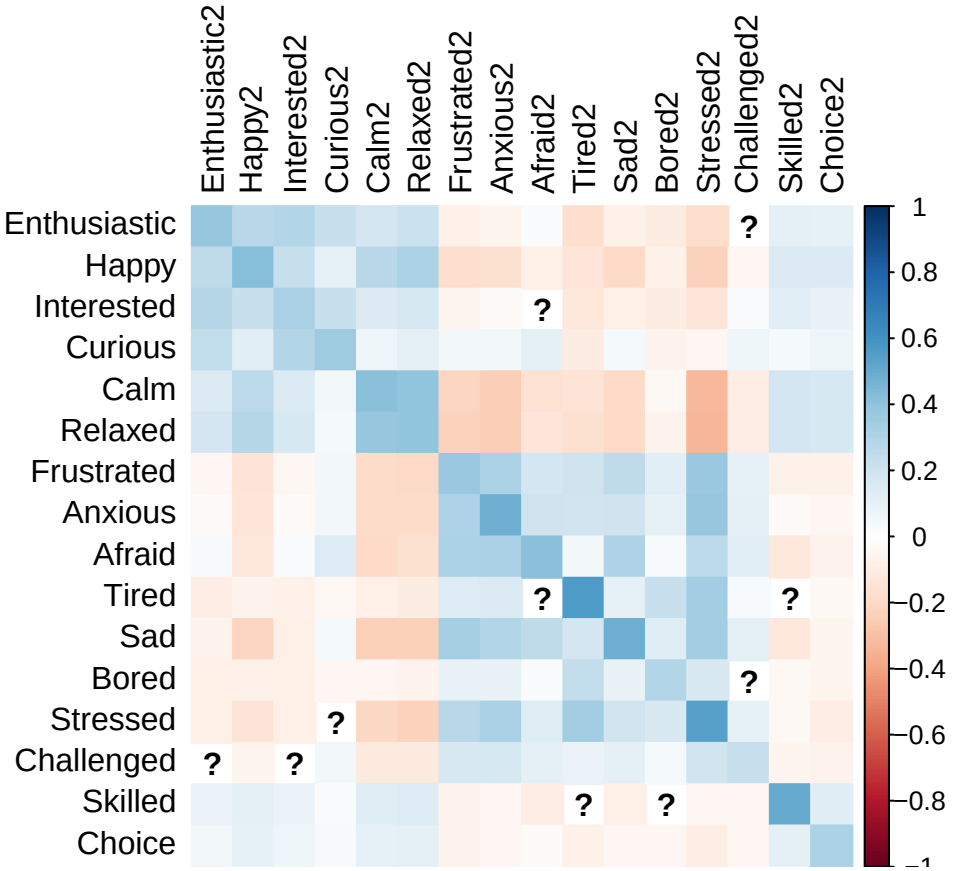
$$E_{t+1} = \beta E_t + \gamma E_t + C$$

for all emotion combinations A-P. The coefficient gamma represents the "participant" variable and is used here to help account for any variability among particpants. Coefficients with a significance level below 0.05 are removed.

```
##              Enthusiastic2 Happy2 Interested2 Curious2  Calm2 Relaxed2
## Enthusiastic         0.384  0.272       0.295    0.221  0.186    0.212
## Happy                0.257  0.423       0.239    0.109  0.275    0.312
## Interested           0.283  0.229       0.329    0.229  0.160    0.179
## Curious              0.248  0.125       0.299    0.360  0.077    0.111
## Calm                 0.157  0.266       0.157    0.058  0.416    0.400
## Relaxed              0.182  0.282       0.172    0.048  0.381    0.396
## Frustrated          -0.040 -0.148      -0.043    0.056 -0.189   -0.195
## Anxious             -0.028 -0.130      -0.026    0.060 -0.184   -0.189
## Afraid               0.036 -0.124       0.027    0.145 -0.192   -0.168
## Tired               -0.090 -0.066      -0.074   -0.037 -0.090   -0.101
## Sad                 -0.067 -0.215      -0.084    0.049 -0.234   -0.239
## Bored               -0.088 -0.075      -0.087   -0.042 -0.045   -0.065
## Stressed            -0.088 -0.141      -0.090       NA -0.208   -0.226
## Challenged              NA -0.058          NA    0.053 -0.113   -0.114
## Skilled              0.080  0.116       0.086    0.030  0.137    0.142
## Choice               0.060  0.106       0.061    0.031  0.101    0.116
##              Frustrated2 Anxious2 Afraid2 Tired2   Sad2 Bored2 Stressed2
## Enthusiastic      -0.071   -0.056   0.023 -0.171 -0.078 -0.109    -0.176
## Happy             -0.170   -0.165  -0.086 -0.131 -0.195 -0.079    -0.221
## Interested        -0.057   -0.026      NA -0.128 -0.080 -0.102    -0.134
## Curious            0.059    0.055   0.114 -0.102  0.042 -0.070    -0.043
## Calm              -0.217   -0.246  -0.150 -0.150 -0.200 -0.039    -0.329
## Relaxed           -0.225   -0.241  -0.132 -0.162 -0.197 -0.069    -0.330
## Frustrated         0.379    0.313   0.185  0.200  0.255  0.129     0.374
## Anxious            0.301    0.483   0.201  0.191  0.203  0.107     0.384
## Afraid             0.312    0.325   0.411  0.051  0.308  0.036     0.260
## Tired              0.147    0.155      NA  0.568  0.105  0.224     0.345
## Sad                0.337    0.299   0.256  0.183  0.481  0.132     0.345
```

```
## Bored              0.100    0.099   0.024  0.249  0.098  0.298     0.161
## Stressed           0.278    0.323   0.133  0.348  0.200  0.161     0.548
## Challenged         0.166    0.177   0.114  0.081  0.116  0.044     0.199
## Skilled           -0.064   -0.040  -0.092     NA -0.081     NA    -0.042
## Choice            -0.068   -0.044  -0.029 -0.072 -0.042 -0.048    -0.095
##               Challenged2 Skilled2 Choice2
## Enthusiastic           NA    0.112   0.101
## Happy              -0.043    0.157   0.150
## Interested          0.027    0.120   0.098
## Curious             0.080    0.047   0.072
## Calm               -0.099    0.184   0.171
## Relaxed            -0.095    0.185   0.174
## Frustrated          0.105   -0.072  -0.076
## Anxious             0.118   -0.026  -0.049
## Afraid              0.128   -0.129  -0.070
## Tired               0.038       NA  -0.036
## Sad                 0.117   -0.128  -0.058
## Bored                  NA   -0.031  -0.057
## Stressed            0.103   -0.037  -0.093
## Challenged          0.223   -0.054  -0.060
## Skilled            -0.045    0.507   0.138
## Choice             -0.041    0.113   0.316
```

These coefficients are now used to create a network analysis of emotions in time series. The network was created using the corrplot package in R. The rows are emotions at time t and the columns are emotions at time $t+1$



What conclusions can we draw from this analysis? First, it seems that the auto-regressions (predicting an emotion by itself in the previous time period) have the strongest beta values. In particular, negative emotions like sadness, tiredness, and stress seem to have the highest beta values when predicting themselves in the next time period. Symmetry is also evident between the first 13 rows and columns, as a rough outline of four quadrants emerges. Where the symmetry fails to persist, however, is when students are asked about their levels of curiousity. Curiousity in the current period has a positive coefficient for predicting fear and sadness in the next. In addition, feelings of anxiousness, frustration, and fear in the current period have a positive relationship to curiousity in the next. Although this finding needs replication in other studies to substantiate itself, this could signal to educators to develop processes that help channel student's anxiety or frustration into creativity and curiousity.

Part 2: Emotions and Weather

The next exploration hopes to glean insights about emotion dynamics at different types of weather. It is well-documented, however, that emotions on an individual level can be caused by many things, so we must proceed with caution as we attempt to draw our conclusions.

Before the data can be analyzed, the weather dataset must be merged with the ESM dataset. In order to merge weather data and ESM data, each ESM row was given an ID to determine which airport they would be sourcing the data from. This ID was based off of school. Columns were then created in the original ESM dataset to bring in the weather data. Weather & ESM data would be matched by date, and then empty weather columns in the ESM datset would source their values from the weather data at a gien airport at a given date.

Now that the data is cleaned and merged it is time to begin exploration on the dataset. This set contains fourteen emotion measures and over twenty weather variables per observation, so we must hone our efforts. A two-pronged exploration approach is taken: numerical and visual. Using the mean emotion score as an initial measurement parameter, tables of means are constructed for each emotion at different: 1) average daily temperatures 2) daily humidity levels 3) daily visibility levels and 4) daily precipitation levels (rain). Then using the ggplot2 graphics package, these means are laid out on four separate plots in an effort to narrow the lens of focus.

Below is an example of how the means are calculated and laid out at the various temperatures. Code for other weather parameters is included in the R markdown.

```
cols <- 17
count <- 1
temps <- unique(esm$tempavg)  # All the unique temperatures recorded
sort(temps)
```

```
##  [1] 53 57 58 59 61 62 63 65 66 67 69 71 72 73 74
```

```
em_avg_temp <- matrix(ncol = cols,nrow=length(temps)) # Creating empty matrix
for(i in temps){
  em_avg_temp[count,1] <- i # First row is the temperature
  em_avg_temp[count,2] <- length(which(esm$tempavg==i)) # Second is # of observations
  em_avg_temp[count,3:cols] <-
    colMeans(esm[esm$tempavg==i,724:738], na.rm=T) #Column means are taken from subsetted data
  count <- count+1 # Move onto the next row in matrix
}
em_avg_temp <- em_avg_temp[order(em_avg_temp[,1]),] # Order matrix by ascending temp
em_avg_temp_df <- as.data.frame(em_avg_temp) # Make it a dataframe
colnames(em_avg_temp_df) <- c("tempavg","num_obs",names(esm)[724:738]) #Naming columns
```

Temperature Mean Table

```
##    tempavg num_obs enthusiast_ESM happy_ESM interested_ESM curious_ESM
## 1       53     956       1.911392  2.368201       2.025175    1.780726
## 2       57     659       2.056962  2.443975       2.309979    2.010638
## 3       58    2904       1.976007  2.420172       2.047576    1.774457
## 4       59    1970       1.976422  2.407202       2.085357    1.869958
## 5       61    1005       2.013106  2.458442       2.096859    1.822222
## 6       62    2885       1.871264  2.316007       1.994480    1.768310
## 7       63     748       1.914035  2.356766       2.000000    1.773852
## 8       65     970       1.951589  2.375000       2.013575    1.760968
## 9       66      58       2.068182  2.720930       2.227273    1.674419
## 10      67     982       1.900815  2.324288       2.027100    1.758152
## 11      69    1803       2.000740  2.513694       2.062268    1.705403
## 12      71     901       1.931193  2.294656       2.007657    1.749618
## 13      72     595       1.911162  2.311364       1.911162    1.746575
## 14      73    1938       1.929300  2.319756       1.970708    1.747613
## 15      74     236       1.942529  2.261364       1.887640    1.685393
##    calm_ESM relaxed_ESM frustrated_ESM anxious_ESM afraid_ESM tired_ESM
## 1  2.461752    2.301543       1.906031    2.021008   1.441176  2.895105
## 2  2.614894    2.458422       1.867238    2.072034   1.407249  2.963907
## 3  2.476902    2.374319       1.909379    1.992290   1.482556  2.771041
## 4  2.476784    2.371349       1.896600    2.013889   1.478834  2.785863
## 5  2.449673    2.392157       1.789267    1.938320   1.473890  2.644444
## 6  2.356057    2.245509       1.941962    2.050138   1.528302  2.821445
## 7  2.338028    2.317460       1.992970    2.126538   1.640845  2.670194
## 8  2.330827    2.268477       2.007519    2.058559   1.652108  2.621988
## 9  2.744186    2.795455       1.581395    1.909091   1.651163  2.181818
## 10 2.361789    2.282993       1.881954    1.994573   1.489824  2.728997
## 11 2.548889    2.549556       1.765185    1.910569   1.446006  2.585799
## 12 2.315549    2.303817       2.015267    2.085366   1.623476  2.611280
## 13 2.364253    2.304545       1.974828    2.078475   1.675057  2.625000
## 14 2.336283    2.273224       1.971311    2.065395   1.569775  2.699320
## 15 2.375000    2.261364       1.704545    1.943820   1.488636  2.348315
##     sad_ESM bored_ESM stressed_ESM challenge_ESM skills_ESM
## 1  1.626928  2.406425     2.395775      1.841880   2.589158
## 2  1.615711  2.419831     2.513800      1.817391   2.676087
## 3  1.626304  2.335601     2.362443      1.834328   2.593664
## 4  1.674548  2.367177     2.382536      1.905063   2.606892
## 5  1.566406  2.258486     2.273560      1.828229   2.666223
## 6  1.678309  2.410764     2.469238      1.874534   2.610462
## 7  1.680141  2.277385     2.505300      1.996396   2.693841
## 8  1.666163  2.255287     2.457831      1.981538   2.706790
## 9  1.431818  2.023810     1.840909      1.627907   2.418605
## 10 1.659864  2.359079     2.369418      1.951923   2.663912
## 11 1.594235  2.172694     2.218519      1.713855   2.726244
## 12 1.776758  2.287023     2.435976      1.911491   2.721617
## 13 1.766440  2.309795     2.430524      1.941725   2.646370
## 14 1.716814  2.353741     2.437926      1.859710   2.663209
## 15 1.516854  2.284091     2.227273      1.875000   2.988636
```

Visibility Mean Table

```
##   visiblity num_obs enthusiast_ESM happy_ESM interested_ESM curious_ESM
## 1         5     834       1.983713  2.526829       2.034202    1.700813
## 2         6     982       1.900815  2.324288       2.027100    1.758152
## 3         7     819       1.988764  2.511218       2.072464    1.768000
## 4         8     595       1.911162  2.311364       1.911162    1.746575
## 5         9    2644       1.921209  2.377293       2.022829    1.743411
## 6        10   12736       1.952121  2.372334       2.042442    1.797711
##   calm_ESM relaxed_ESM frustrated_ESM anxious_ESM afraid_ESM tired_ESM
## 1 2.627642    2.631494       1.697561    1.816260   1.428339  2.531707
## 2 2.361789    2.282993       1.881954    1.994573   1.489824  2.728997
## 3 2.547352    2.461415       1.861736    1.980769   1.426282  2.790735
## 4 2.364253    2.304545       1.974828    2.078475   1.675057  2.625000
## 5 2.371585    2.345084       1.919603    2.027737   1.527282  2.759682
## 6 2.418945    2.322128       1.922632    2.033782   1.523407  2.738039
##    sad_ESM bored_ESM stressed_ESM challenge_ESM skills_ESM
## 1 1.595779  2.066775     2.144951      1.618421   2.761120
## 2 1.659864  2.359079     2.369418      1.951923   2.663912
## 3 1.589085  2.343548     2.334936      1.847154   2.636808
## 4 1.766440  2.309795     2.430524      1.941725   2.646370
## 5 1.641229  2.326227     2.404762      1.876138   2.667343
## 6 1.665918  2.345192     2.408187      1.869844   2.635495
```

Humidity Mean Table

```
##    humidity num_obs enthusiast_ESM happy_ESM interested_ESM curious_ESM
## 1        42     158       1.880342  2.448276       2.051282    1.732759
## 2        43     956       1.911392  2.368201       2.025175    1.780726
## 3        47     798       1.942244  2.411184       2.108731    1.888704
## 4        51     120       2.000000  2.200000       2.388889    2.285714
## 5        54     659       2.056962  2.443975       2.309979    2.010638
## 6        57     235       1.942529  2.261364       1.887640    1.685393
## 7        58    2924       1.966216  2.386384       2.054554    1.788618
## 8        59    1005       2.013106  2.458442       2.096859    1.822222
## 9        61    1052       2.001248  2.413233       2.053885    1.837703
## 10       62    1028       1.944730  2.367137       1.978149    1.750643
## 11       64    2102       1.875661  2.286847       1.939615    1.735411
## 12       65     982       1.900815  2.324288       2.027100    1.758152
## 13       67    1496       1.923147  2.301370       1.968864    1.748399
## 14       70       4       1.000000  1.000000       1.000000    1.000000
## 15       74     969       2.014925  2.502717       2.085714    1.709239
## 16       76     794       1.971572  2.385382       2.038333    1.780936
## 17       83    1027       1.823980  2.266242       1.973316    1.755754
## 18       84     834       1.983713  2.526829       2.034202    1.700813
## 19       86     819       1.988764  2.511218       2.072464    1.768000
## 20       87     648       1.935614  2.366935       2.008114    1.774848
##    calm_ESM relaxed_ESM frustrated_ESM anxious_ESM afraid_ESM tired_ESM
## 1  2.534483    2.547009       1.801724    2.008547   1.632479  2.296610
## 2  2.461752    2.301543       1.906031    2.021008   1.441176  2.895105
## 3  2.479407    2.365289       1.910744    2.039604   1.466227  2.892916
## 4  2.470588    2.588235       1.970588    2.314286   1.558824  3.114286
## 5  2.614894    2.458422       1.867238    2.072034   1.407249  2.963907
```

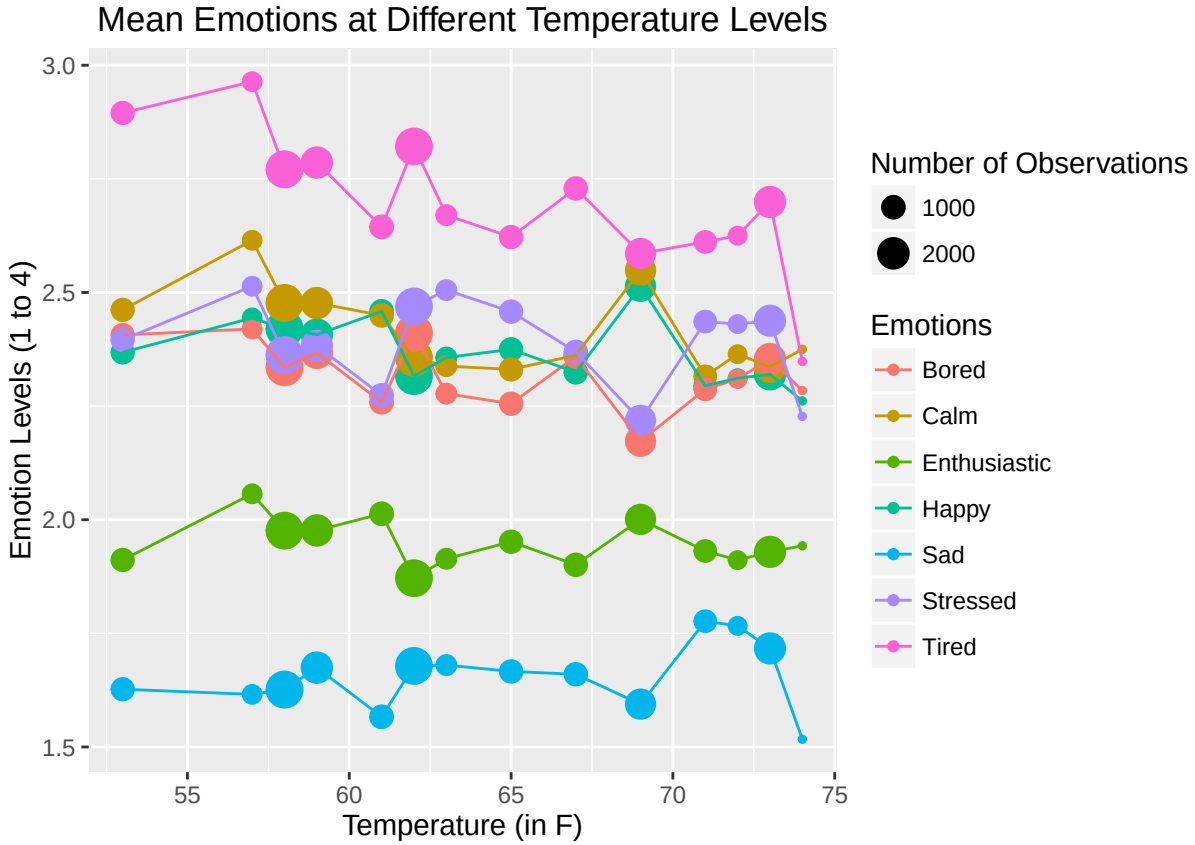
```
## 6  2.375000    2.261364       1.704545    1.943820   1.488636  2.348315
## 7  2.420839    2.308631       1.919530    2.038357   1.536134  2.692377
## 8  2.449673    2.392157       1.789267    1.938320   1.473890  2.644444
## 9  2.475062    2.366708       1.882793    1.981227   1.485000  2.690387
## 10 2.483290    2.353925       1.909091    1.939433   1.442308  2.871630
## 11 2.305703    2.225914       1.994020    2.116711   1.595096  2.746190
## 12 2.361789    2.282993       1.881954    1.994573   1.489824  2.728997
## 13 2.335155    2.304110       1.999084    2.082577   1.644099  2.616788
## 14 1.000000    1.000000       1.000000    2.000000   1.000000  3.000000
## 15 2.482993    2.480978       1.821769    1.989160   1.460705  2.630936
## 16 2.330017    2.270000       2.023256    2.046434   1.635607  2.635607
## 17 2.295019    2.242347       1.959184    1.996178   1.516582  2.908163
## 18 2.627642    2.631494       1.697561    1.816260   1.428339  2.531707
## 19 2.547352    2.461415       1.861736    1.980769   1.426282  2.790735
## 20 2.327273    2.305668       2.002016    2.135081   1.643725  2.716024
##     sad_ESM bored_ESM stressed_ESM challenge_ESM skills_ESM
## 1  1.521368  2.139130     2.162393      1.815789   2.684211
## 2  1.626928  2.406425     2.395775      1.841880   2.589158
## 3  1.656198  2.425041     2.505766      1.899160   2.579832
## 4  1.800000  2.382353     2.371429      1.970588   2.470588
## 5  1.615711  2.419831     2.513800      1.817391   2.676087
## 6  1.516854  2.284091     2.227273      1.875000   2.988636
## 7  1.642309  2.338295     2.419748      1.840037   2.609153
## 8  1.566406  2.258486     2.273560      1.828229   2.666223
## 9  1.682957  2.322459     2.289638      1.906683   2.633039
## 10 1.652510  2.374359     2.367609      1.865711   2.551499
## 11 1.747020  2.387823     2.477800      1.871553   2.652554
## 12 1.659864  2.359079     2.369418      1.951923   2.663912
## 13 1.772603  2.296161     2.433790      1.923579   2.691589
## 14 1.000000  1.000000     1.000000      1.000000   1.000000
## 15 1.592944  2.260459     2.279891      1.794444   2.696801
## 16 1.689482  2.265442     2.454243      1.989796   2.699659
## 17 1.652229  2.412516     2.444727      1.870801   2.637306
## 18 1.595779  2.066775     2.144951      1.618421   2.761120
## 19 1.589085  2.343548     2.334936      1.847154   2.636808
## 20 1.695565  2.288032     2.527383      2.006198   2.671518
```

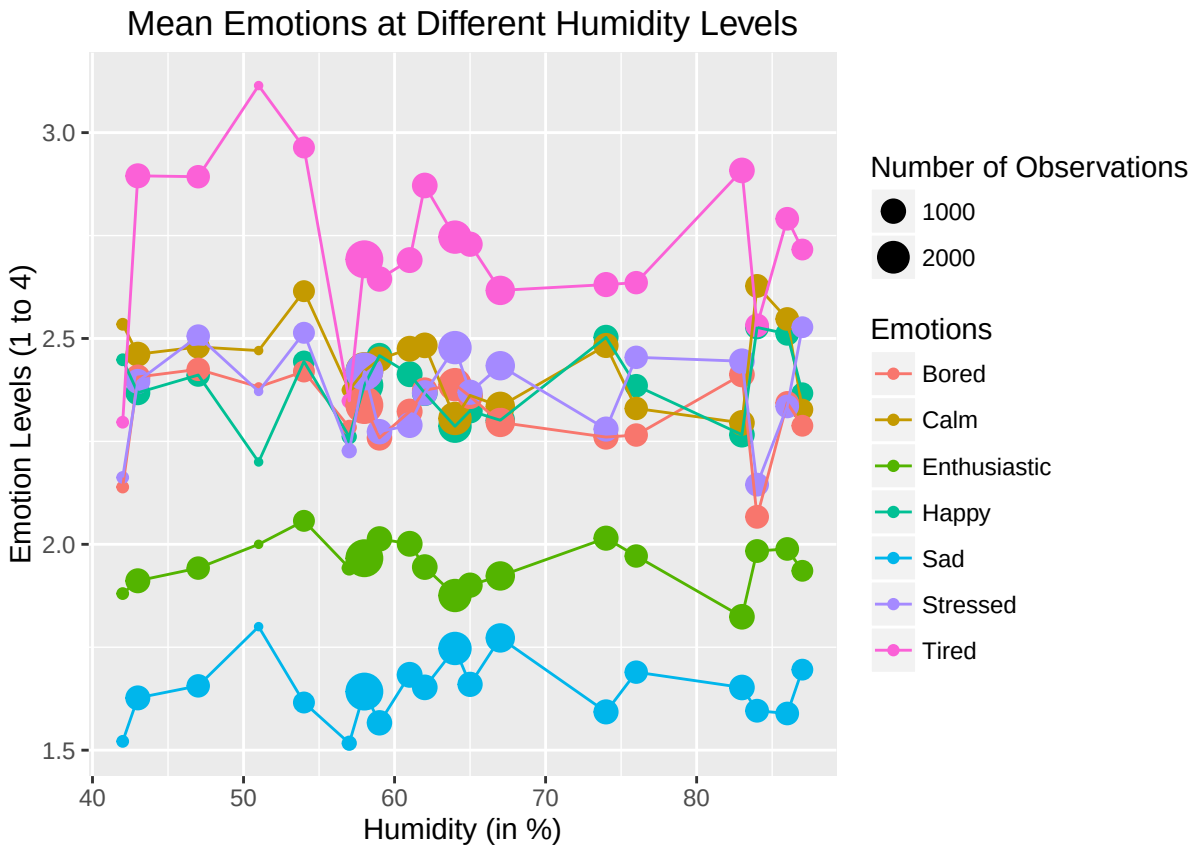
Precipitation Mean Table

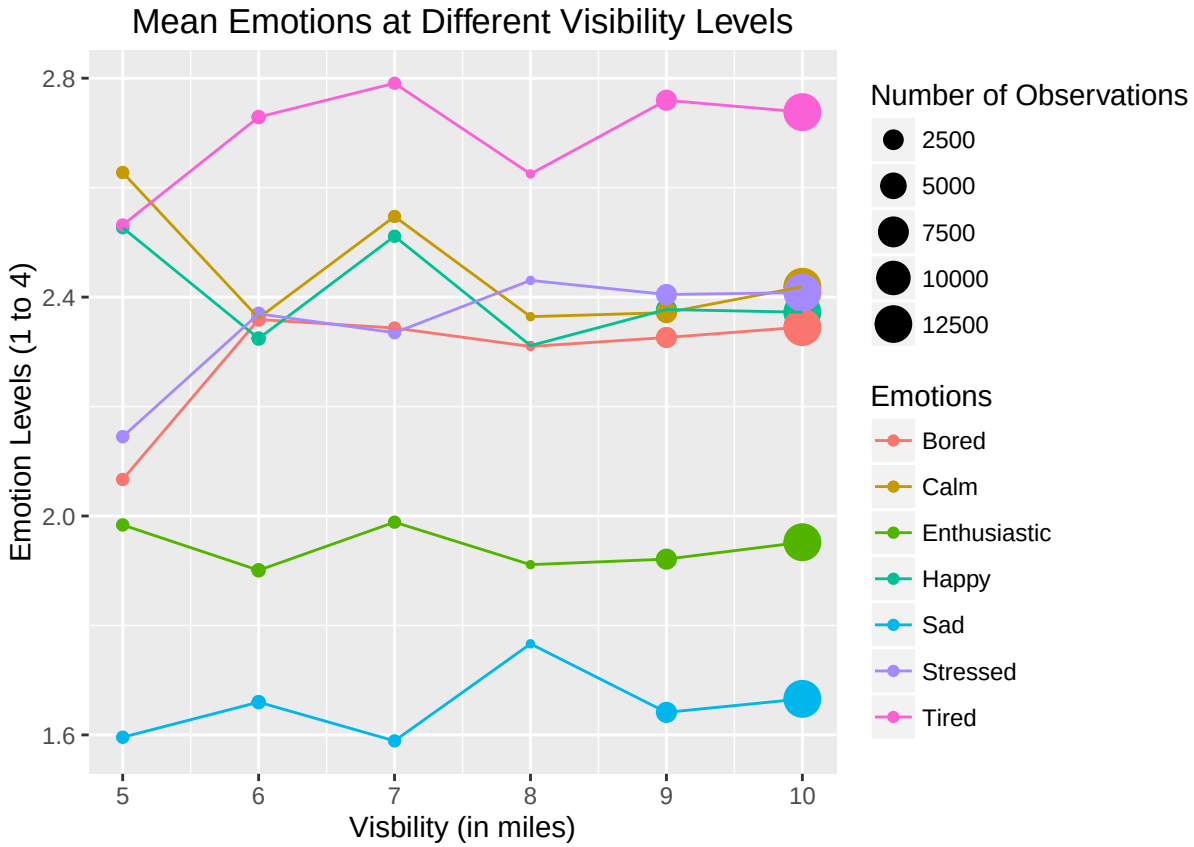
```
##   rain num_obs enthusiast_ESM happy_ESM interested_ESM curious_ESM
## 1 0.00   14120       1.958708  2.382047       2.045216    1.792077
## 2 0.04     986       1.862903  2.288978       1.975741    1.739247
## 3 0.07     176       1.761905  2.274194       1.777778    1.571429
## 4 0.09     648       1.935614  2.366935       2.008114    1.774848
## 5 0.17     819       1.988764  2.511218       2.072464    1.768000
## 6 0.34    1027       1.823980  2.266242       1.973316    1.755754
## 7 1.23     834       1.983713  2.526829       2.034202    1.700813
##   calm_ESM relaxed_ESM frustrated_ESM anxious_ESM afraid_ESM tired_ESM
## 1 2.426771    2.340034       1.911188    2.022445   1.516681  2.724816
## 2 2.289367    2.184388       1.979784    2.125337   1.587366  2.763122
## 3 2.338710    2.253968       1.857143    2.174603   1.809524  2.492063
## 4 2.327273    2.305668       2.002016    2.135081   1.643725  2.716024
## 5 2.547352    2.461415       1.861736    1.980769   1.426282  2.790735
```

```
## 6 2.295019    2.242347       1.959184    1.996178   1.516582  2.908163
## 7 2.627642    2.631494       1.697561    1.816260   1.428339  2.531707
##    sad_ESM bored_ESM stressed_ESM challenge_ESM skills_ESM
## 1 1.659432  2.334448     2.390747      1.872620   2.642934
## 2 1.757047  2.420699     2.493960      1.877551   2.614130
## 3 1.444444  2.158730     2.492063      1.903226   2.774194
## 4 1.695565  2.288032     2.527383      2.006198   2.671518
## 5 1.589085  2.343548     2.334936      1.847154   2.636808
## 6 1.652229  2.412516     2.444727      1.870801   2.637306
## 7 1.595779  2.066775     2.144951      1.618421   2.761120
```

Now that mean tables have been calculated, scatterplots are created to better analyze interesting trends.

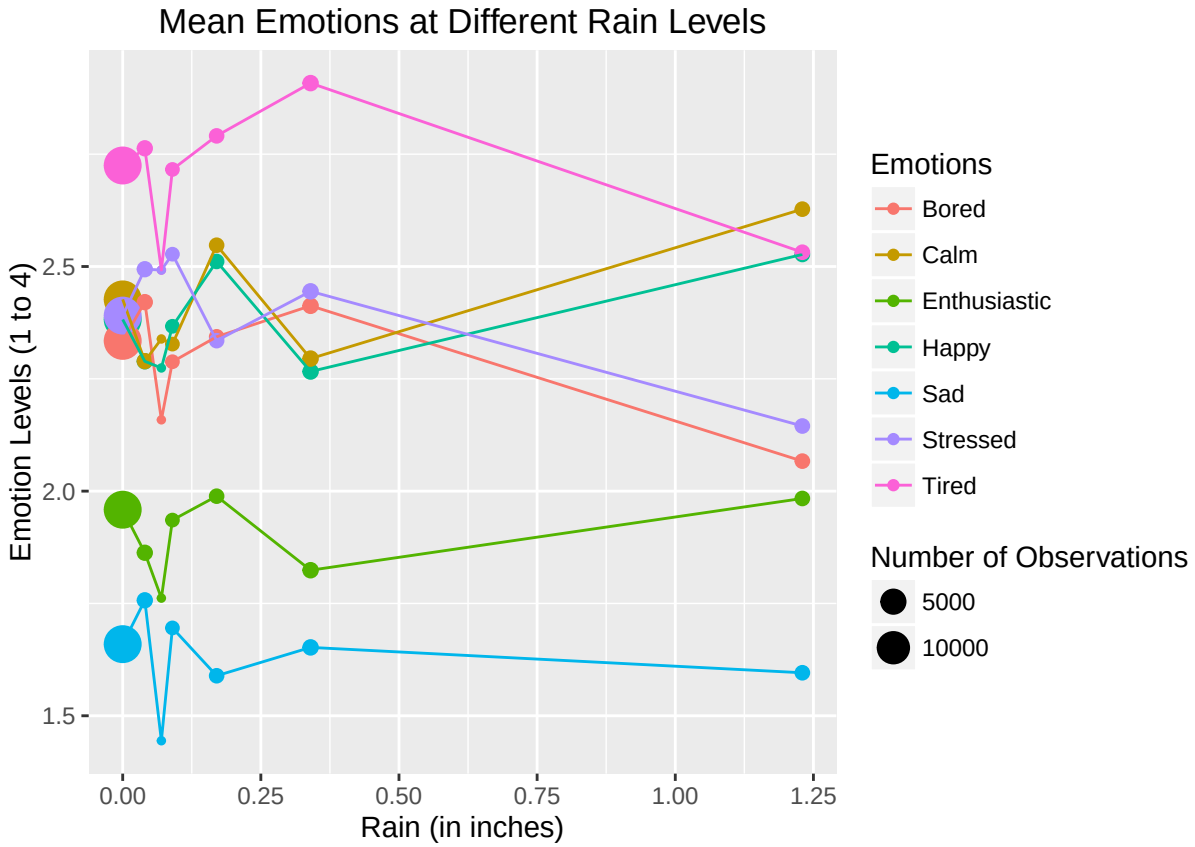



Mean Emotions at Different Humidity Levels
Emotion Levels (1 to 4)
3.0
2.5
2.0
1.5
40
50
60
70
80
Humidity (in %)
Number of Observations
1000
2000
Emotions
Bored
Calm
Enthusiastic
Happy
Sad
Stressed
Tired


Mean Emotions at Different Visibility Levels
Emotion Levels (1 to 4)
2.8
2.4
2.0
1.6
5
6
7
8
9
10
Visbility (in miles)
Number of Observations
2500
5000
7500
10000
12500
Emotions
Bored
Calm
Enthusiastic
Happy
Sad
Stressed
Tired

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```



This exploration provides many opportunities worth pursuing, however, two will be considered here: 1) The seemingly indirect relationship between "tiredness" and temperature, and 2) The relationship between high levels calmness and enthusiasm in juxtaposition to low levels of stress and boredom as visibility decreases and amount of precipitation increases.

2.1 Tiredness v. Temperature

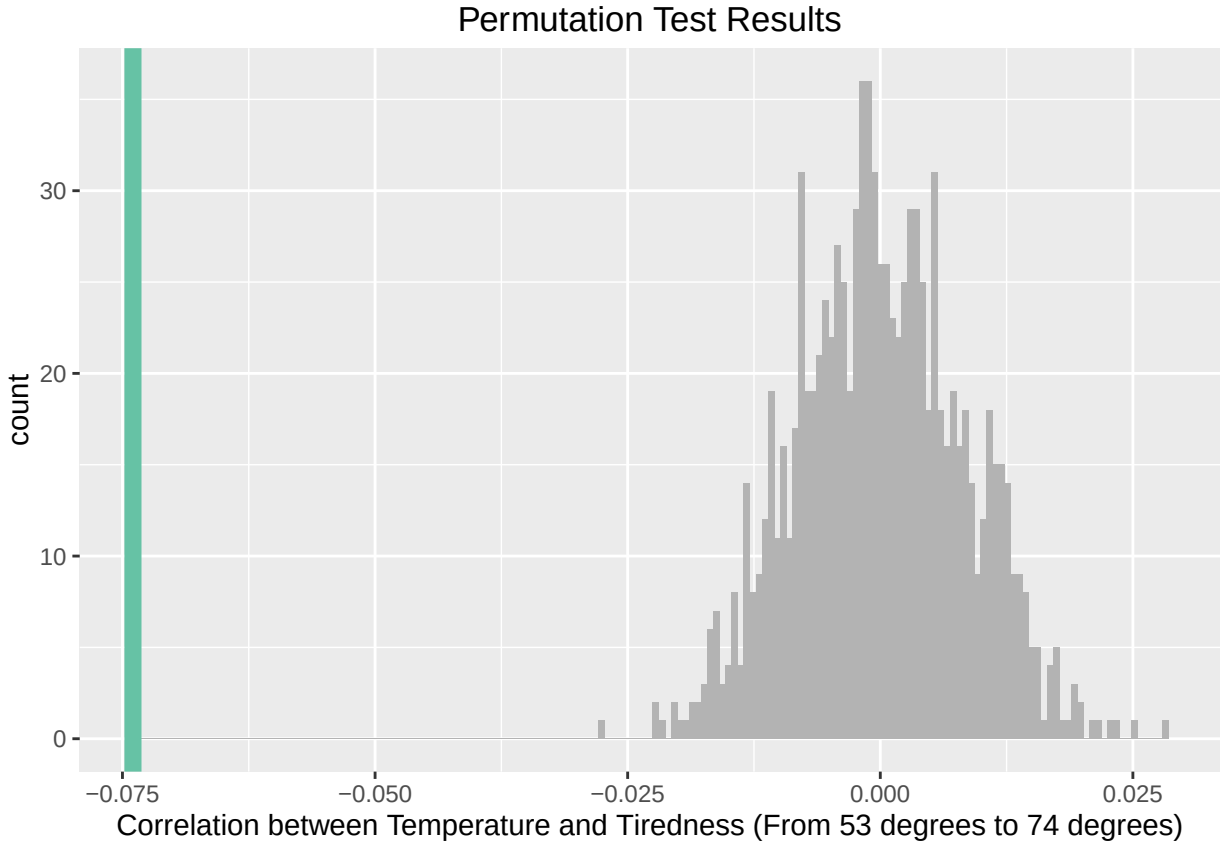
Revisiting the temperature vs. emotions plot, it seems that levels of tiredness generally decrease with an increase in temperature, but is this a truly valuable finding? Is it anything more than random noise in the data? Could it be caused by something other than an increase in temperature? The following section will examine these questions.

Before trying to uncover lurking variables in the data, one must determine if this finding is within in the realm of what we could expect from the data if it were truly random. In order to investigate this, a permutation test is conducted, examining the correlation between the two variables in question. The permutation test will rearrange the vector of responses to the question "How tired do you feel right now?" without replacement, while leaving the vector of temperatures intact. Then the correlation will be measured between the random vector of tiredness scores and their given termperatures. This process will be permuted 1000 times, at which point the distribution of correlations with be compared to the within-groups correlation of the actual data.

```
a <- esm$tired_ESM
b <- esm$tempavg
c <- esm$Participant
alpha <- data.frame(a,b,c)
pack <- statsBy(alpha,group = "c")
s_star <- pack$rwg[2]
s_star  # Within-group correlation
```

```
## [1] -0.07395933
```

Generally speaking, a correlation measures how strongly pairs of varaibles are related. A correlation of -0.07 may seem small, but it is important to interpret this within the context of what would be an expected correlation if the data was truly random. There are many ways to calculate a correlation. When dealing with survey data from many participants, two types of correlations emerge: within-group correlation and between-group correlation. The former measures the relationship between temperature and tiredness *within* individuals across situations. The latter accounts for differences in personality *between* individuals, and reflects the components of emotions that remain stable across a situation but differ from person to person. Since the investigation wants to examine how individuals vary *across situations*, the within-group correlation is the important statistic. When running a permutation test, since responses are randomized across participants, one no longer has to account for between-group correlation and can just use the default Pearson Correlation.

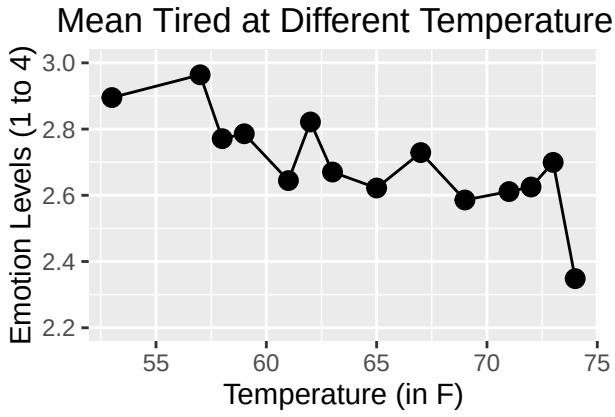


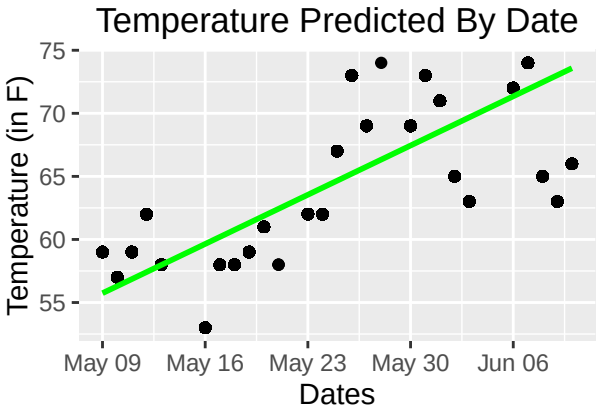
In the figure above, the green line is the test statistic, while the histogram to its right represents the distribution of correlations one could expect to find if the distribution of the data was truly random. Since the actual correlation falls well outside the scope of simulated correlations, there is evidence that this trend is more than mere "noise" in the data.

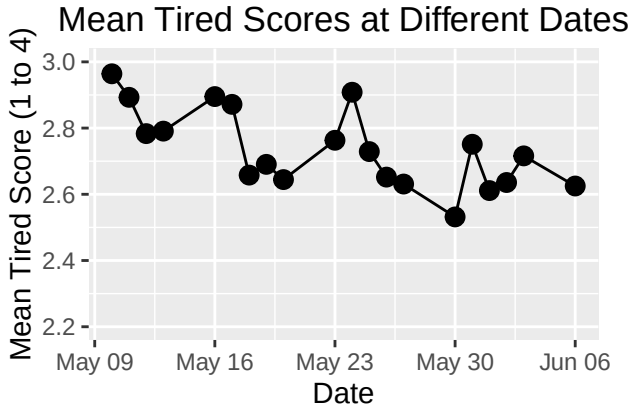
As mentioned earlier, emotions, among high-school students, can be caused by a variety of factors, and this dataset does not provide the entire picture. However, there are still ways of examining other phenomena

that, under the surface, could be the real cause for the trend we are seeing.

A consideration must be made to examine *date* as a lurking variable in this analysis. It could be possible, that this decrease in level of tiredness could be because students are getting towards the end of school, and are getting excited for summer, which also reflects itself in rising temperatures.







The plots above show that the trend found among tired means at different temperatures is similar to the means at different days (as students get closer and closer to the end of school). However, the variables may confound one another, as the linear regression on the right shows a positive relationship between dates and temperatures (with R^2 = 0.58). To investigate further, regressions on the means at given temperatures and dates are performed.

```
## Regressions of temp and date trying to predict avg level of tired per day
fit_temp <- lm(tired_ESM ~ tempavg, data = em_avg_temp_df)
## Predicting mean_tired scores based off of temperature
fit_temp2 <- lm(tired_ESM ~ tempavg + datecol, data = em_avg_temp_df)
summary(fit_temp)$r.squared
```

```
## [1] 0.6156617
```

```
fit_date <- lm(tired_scores ~ dates + tempers , data = tired_df)
# Accounting for within day differences in temperature
fit_date2 <- lm(tired_scores ~ dates, data = tired_df)
# Day predicting tired means

summary(fit_date)$r.squared
```

```
## [1] 0.4941182
```

```
fit_temp$coefficients
```

```
## (Intercept)     tempavg
##  3.85310066 -0.01788796
```

```
fit_temp2$coefficients
```

```
##   (Intercept)       tempavg       datecol
## 165.471710578  -0.006962726  -0.009578696
```

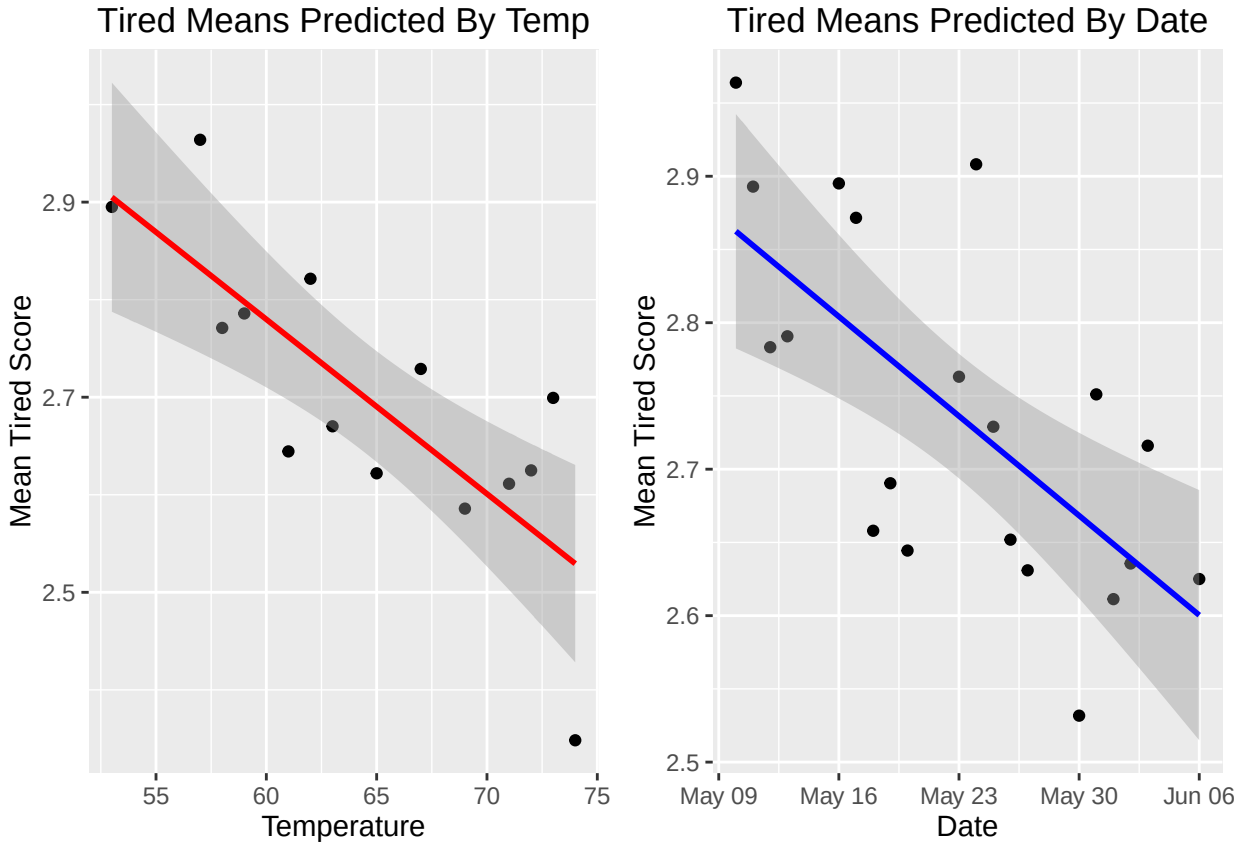
```
fit_date2$coefficients
```

```
##   (Intercept)         dates
## 167.222675391  -0.009707651
```

```
fit_date$coefficients
```

```
##   (Intercept)         dates       tempers
## 108.176684238  -0.006199445  -0.006242424
```

The purpose of performing these regressions is not necessarily to create a model that predicts emotional scores to a satisfactory level. Instead, operating on the assumption that emotions on the individual level are complicated to anticipate, the models help examine the predictive power of the variables in question. First, examining the R^2 for both models, the *fit_temp* model has a higher value. That means that more of variance in the data is explained by the *fit_temp* model over the *fit_date* model. When the models begin to control for the lurking coefficients, it hampers their respective strength. Both parameters lose some of their predictive power when controlling for the other variable, which could mean that *both* temperature and date are affecting levels of tiredness.



Using the *stat_smooth* method in the ggplot package in R, the graphs above show a narrower margin of error for the temperature regression than the date regression, as what was expected by earlier findings with the R^2.

So what conclusions can be drawn from this analysis? First, what is occuring in the dataset is more than just

due to random variance that can be expected during sampling, as made evident by the permutation test. So if something special is going on here, what is causing it? Regression analysis would point to temperature and date each having some effect on how tired a student reports he or she is feeling, however, the two are entangled in such a fashion that makes it difficult to determine exactly which one has the stronger effect.

Part 2: Rainy Day Emotions

When examining the means at different visibility and precipitation levels, one may find an interesting (possibly counter-intuitive) trend: students show higher levels of calm and enthusiasm and lower levels of stress and boredom when visibility is low (cloudy) and precipitation is high. Before going further it important to make sure that the low visibility and high precipitation are occuring on the same day(s).

```
##
##          0  0.04  0.07  0.09  0.17  0.34  1.23
##   5      0     0     0     0     0     0   834
##   6    982     0     0     0     0     0     0
##   7      0     0     0     0   819     0     0
##   8    595     0     0     0     0     0     0
##   9    969     0     0   648     0  1027     0
##   10 11574   986   176     0     0     0     0
```

Good, it looks like the day of low visibility occurs on the same day as high rain levels. In addition, having a high n number of observations (834) lowers the margin of error in the sample and lets one make conclusions about this finding with greater certainty.

Next, the data is grouped into two categories: 1) Med/High visibility, little to no rain and 2) Low visibility, moderate/heavy rain. Grouping the data into two categories provides several advantages. For one, a two-sample t-test is conducted on the populations for all the different emotions.

Calmness t-test

```
#Significant
t.test(esm$calm_ESM[esm$visibility==5],esm$calm_ESM[esm$visibility!=5])
```

```
##
##  Welch Two Sample t-test
##
## data:  esm$calm_ESM[esm$visibility == 5] and esm$calm_ESM[esm$visibility != 5]
## t = 5.958, df = 677.22, p-value = 4.106e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1440760 0.2857154
## sample estimates:
## mean of x mean of y
##  2.627642  2.412747
```

Enthusiasm t-test p-value

```
## [1] 0.2847787
```

Stress t-test p-value

```
## [1] 1.111781e-11
```

Bored t-test p-value

```
## [1] 1.73371e-09
```

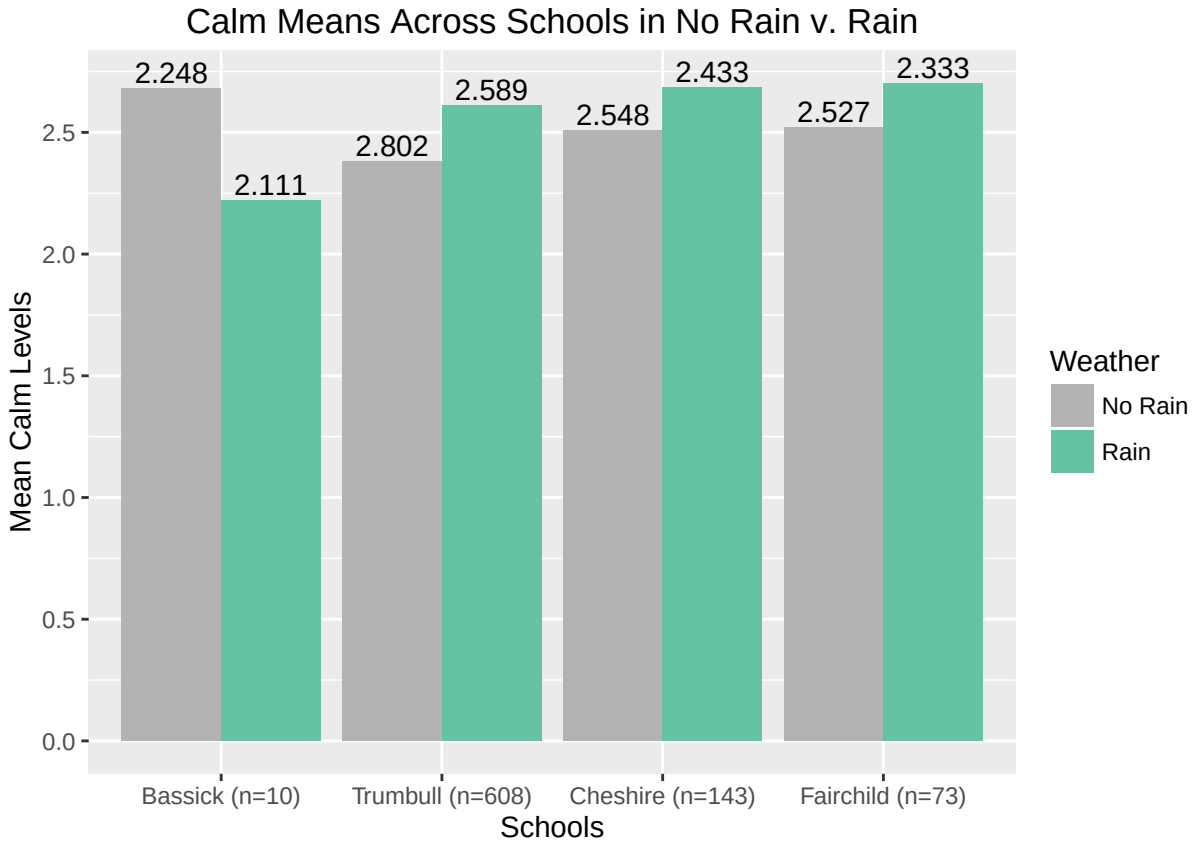
The data can be subset by visibility==5 or precip==1.23 since all of the observations for each overlap with one another. The t-test confirms some of the original suscipicions. The two-sample test determines whether or not the two group means are truly different from one another. The test produces a 95% confidence interval of the difference between the means. If this interval contains zero, then it can be said with 95% confidence that the sample means may not be different. This would result in a p-value greater than 0.05, the significance level. The p-value can be interpreted as the odds that a dataset with these group means would appear, given that the sample means are the same. If the p-value is below 0.05, then it can be argued that the sample means are not the same.

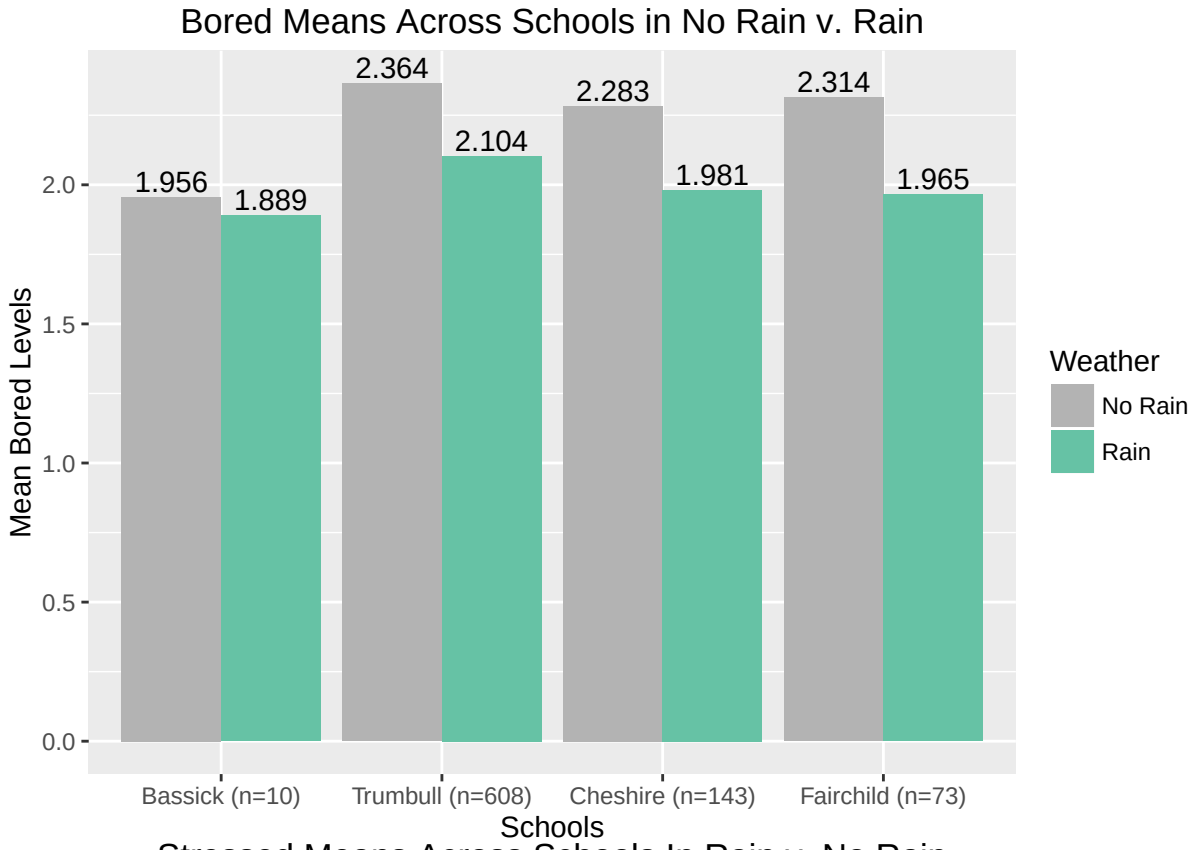
The tests show that group means for stress, boredom and calmness are significant, but enthusiasm is not, so it will be removed from further analysis.

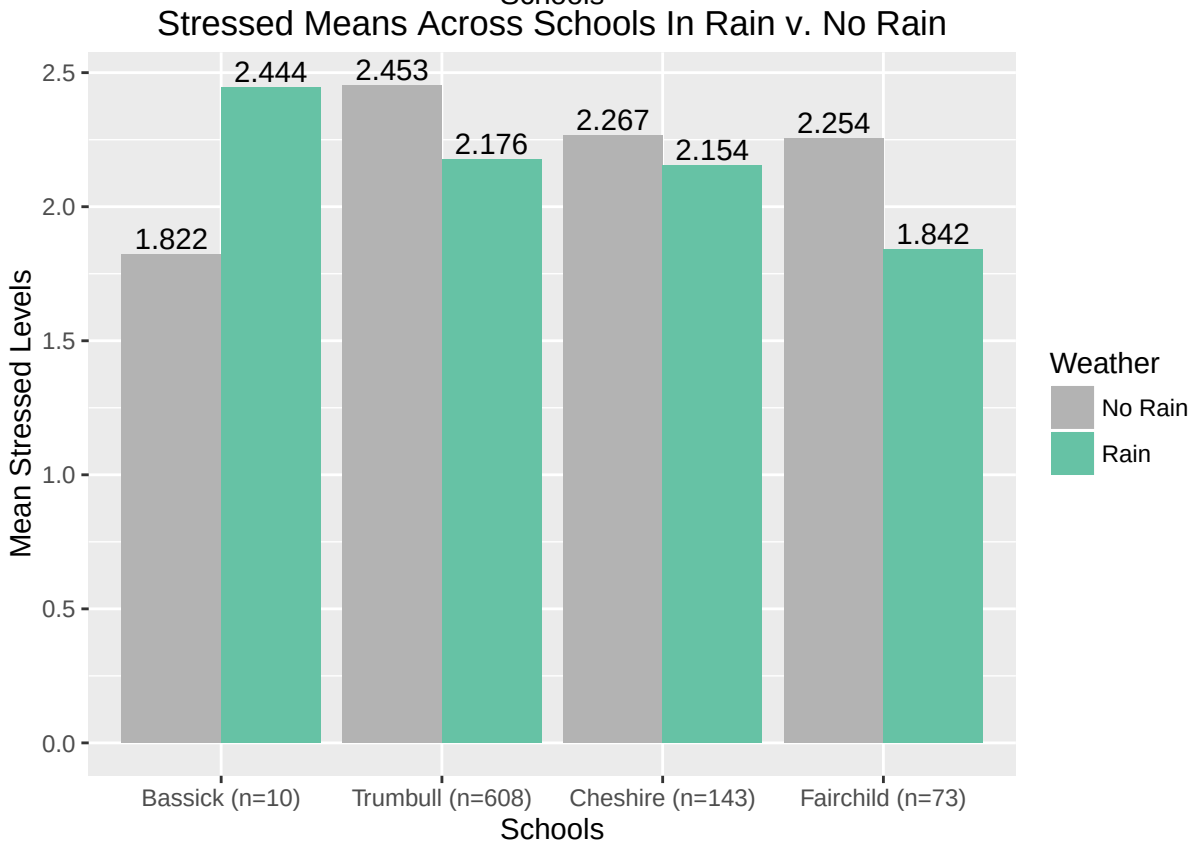
Since the rainy day data is collated from only one day, some concerns arise. Could the change in mood be caused by something else that happened that day at school? One way to investigate this is to determine if preliminary findings are consistent with results *across schools*. Below is a table of number of responses from each school on the rainy day (Code is in the RMarkdown):

```
##
##   Bassick  Trumbull  Cheshire Fairchild
##        10       608       143        73
```

It looks like Trumbull, Cheshire, and Fairchild will have enough a high enough n to feel confident about rainy day emotion levels, however Bassick only had 10 respondents, so results there are not as reliable, however they will still be included. The graphs below investigate the differences in emotions among schools in rainy day vs. no rainy day.







All non-Bassick mean differences are significant at the 0.05 level except for calmness levels in Fairchild and

stress levels in Cheshire. See the R Markdown for all relevant t-tests. These graphs and tests show, that in large part, the trends in calmness, stress, and boredom are consistent across schools, which helps respond to the concern that these findings could be due to something else going on at school that day. If that were the case, something simlar would have had to occur across other schools in Connecticut, which is more unlikely.

The t tests performed above compared students who took the survey during the rain to all students in a given school. In order to generalize the findings found to *all* students across *every* school, it must be shown that there is nothing out of the ordinary about the group of students who responded during the rainy day in comparison to the rest of the students at their school. If it can be shown that the students who responded during the rainy day generally feel the same levels of calmness, stress and boredom as all other students during the non-rainy days, the conclusions made here can be generalized to high school students across Connecticut.

P Values for Calmness at Cheshire, Farichild, and Trumbull

```
## [1] 0.2084072
```

```
## [1] 0.5350166
```

```
## [1] 0.6037533
```

P Values for Stress at Cheshire, Farichild, and Trumbull

```
## [1] 0.7601945
```

```
## [1] 0.7524109
```

```
## [1] 0.6979869
```

P Values for Boredom at Cheshire, Farichild, and Trumbull

```
## [1] 0.6321865
```

```
## [1] 0.5323635
```

```
## [1] 0.0670668
```

As shown above, all p-values are above the significance level 0.05. This means that it cannot be said, with any degree of certainty, that the calmness, stress, and boredom levels of students who took the survey during the rainy day are different than all of the survey respondents at a given school. More generally, it can be said that the population of students who took the survey during the rainy day are a fair representation of the students at a school, and since findings were fairly consistent across schools (with two exceptions), one could make the arguement for the initial findings being indicitive of Connecticut high school students.

This finding, if true, has implications in many classrooms. It may seem intuitive that students posses a generally lower emotional disposition during days of "bad weather", however this finding challenges that and claims that students are calmer and less stressed during rainy days, but also less bored, and therefore primed better for learning and challenges. It may also mean challenging other commonly held beliefs about emotions and weather.

Further Steps

A lot of the conlusions made by this analysis have limitations. For one, findings from this study may not generalize to national education, since students from different climates are not included. For example, consider the way students in Washington react to rain vs. students in Arizona. In addition, although grade levels are fairly equally represented, genders are not. Finally, as has been already mentioned, emotions are caused by many different external and internal factors, and limitations of this dataset make it difficult to control for all of them.

This analysis does, however, provide more direction for further studies. Exploratory analysis, substantiated by further statistical tests, have shown that there is some relationship between tiredness and temperature within

a certain temperature range, as well as calmness, boredom, and stress (and possibly enthusiasm) during rainy days. Further studies could focus on these scenarios, as well as record survey data from multiple states, in order to defend findings that could become the building blocks for change in national education policy.