Schubert: Disability and Late Style

This is a part of a larger project that employs corpus methods to examine the role of variability as a predictor of a composer's state of health. The specific focus of this study is the effect of syphilis—broadly defined as unwellness—on Franz Schubert's musical output. Measures of scale-degree, interval, and rhythmic variability, are used utilized a binomial logistic regression. Two musical features were significant (P<.01) at predicting the episodes of unwellness in the composer's output, highlighting the link between biographical events and artistic output.

Preparation

Directories and Libraries

```
setwd("/Users/connordavis/Documents/GitHub/schubertWellness/")
schubert_data <- read.csv("schubertData.csv", header=T)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(GGally)
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```

```
library(effects)
```

```
Loading required package: carData
lattice theme set by effectsTheme()
See ?effectsTheme for details.
```

```
set.seed(529)
```

Cleaning

This data set contains musical features from the Schubert melodies throughout his lifetime, along with information pertaining to his wellness/unwellness. Variables include:

- filename - name of the file from the RISM dataset https://rism.digital/
- deutsche_num - these are the cataloging numbers for Schubert's works
- degree_entropy - a measure of variability in melodic scale degree usage.
- intervallic_entropy - a measure of variability in the note-to-note intervalic distances
- npvi - normalized pairwise variability index. This measure of rhythmic variance.
- start_date - the month the composition process began. We have this information because of diaries and letters.
- completion_date - the month the work was complete. We have this information because of diaries and letters.
- wellness - a binary variable of well/unwell. We have this information because of diaries and letters.
- NOTES..to.be.deleted - residual information leftover from compiling the data and biographical events.

```
str(schubert_data)
```

```
'data.frame':   718 obs. of  9 variables:
 $ filename             : chr  "D/B/RISM452509661-1-1-1.thm" "D/B/RISM452509661-1-1-2.thm" "
 $ deutsche_num         : chr  "5" "5" "10" "10" ...
 $ degree_entropy       : num  2.42 1.75 1.86 1.89 2.47 ...
 $ intervallic_entropy  : num  2.25 2.16 1.12 1.66 2.77 ...
 $ npvi                 : num  11.1 48.3 31.6 87.4 59.5 ...
 $ start_date           : chr  "1811.03" "1811.03" "1811.12" "1811.12" ...
 $ completion_date      : chr  "1811.03" "1811.03" "1811.12" "1811.12" ...
 $ wellness             : chr  "well" "well" "well" "well" ...
 $ NOTES..to.be.deleted.: chr  "" "" "" "" ...
```

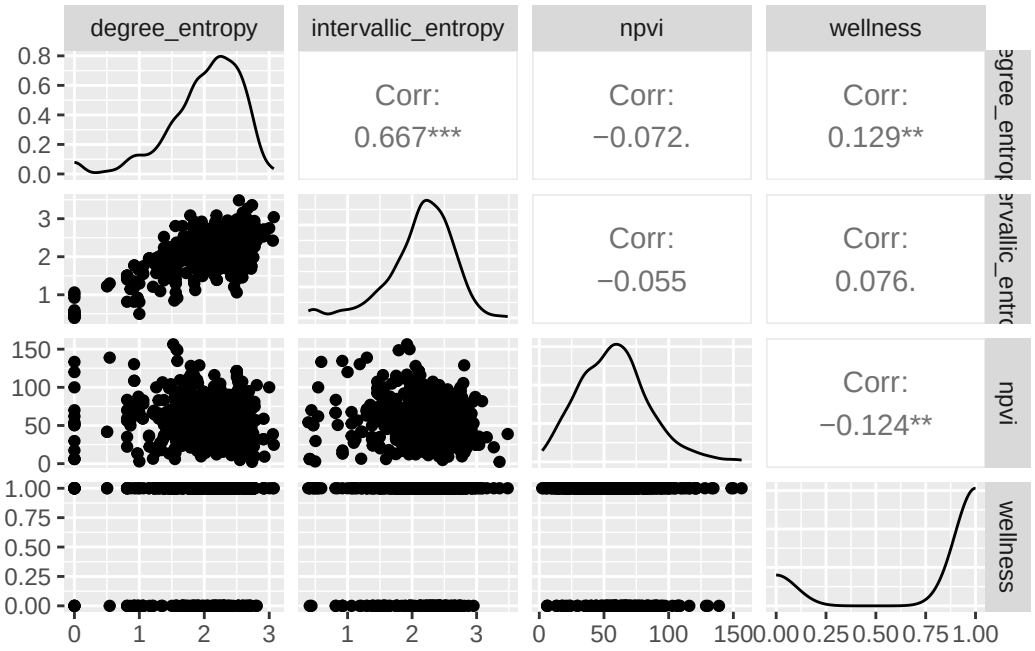
```
schubertDataClean <- schubert_data %>%
  filter(!str_detect(wellness, "uncertain")) %>% #get rid of pieces we aren't certain abou
  select(-NOTES..to.be.deleted., #remove unecessary columns
         -completion_date,
         -deutsche_num,
         -filename) %>%
  filter(!str_detect(start_date, "UNKNOWN")) %>% #remove unknowns from the dates
  drop_na() #drop nas

schubertDataClean$wellness <- if_else(schubertDataClean$wellness == "well", 1, 0) #change
schubertDataClean$start_date <- as.integer(schubertDataClean$start_date) #convert dates to

str(schubertDataClean) #resulting data structures
```

```
'data.frame':   554 obs. of  5 variables:
 $ degree_entropy     : num  2.42 1.75 1.86 1.89 2.47 ...
 $ intervallic_entropy: num  2.25 2.16 1.12 1.66 2.77 ...
 $ npvi               : num  11.1 48.3 31.6 87.4 59.5 ...
 $ start_date         : int  1811 1811 1811 1811 1812 1813 1813 1813 1813 1813 ...
 $ wellness           : num  1 1 1 1 1 1 1 1 1 1 ...
```

Exploring

I use ggpairs() to check for normality and any correlations that might be useful. Variable "degree_entropy" and "intervallic_entropy" are correlated.

```
schubertDataClean %>% #looks like normal disctributions and degree_entropy and intervallic
  select(degree_entropy, intervallic_entropy, npvi, wellness) %>%
  ggpairs()
```



Testing Our Variables

I utilized all of the measures of variability in this example to see which ones are significant in the model. Only "degree_entropy" and "npvi" are significant in the model ($p < .05$ and $p < .01$, respectively).

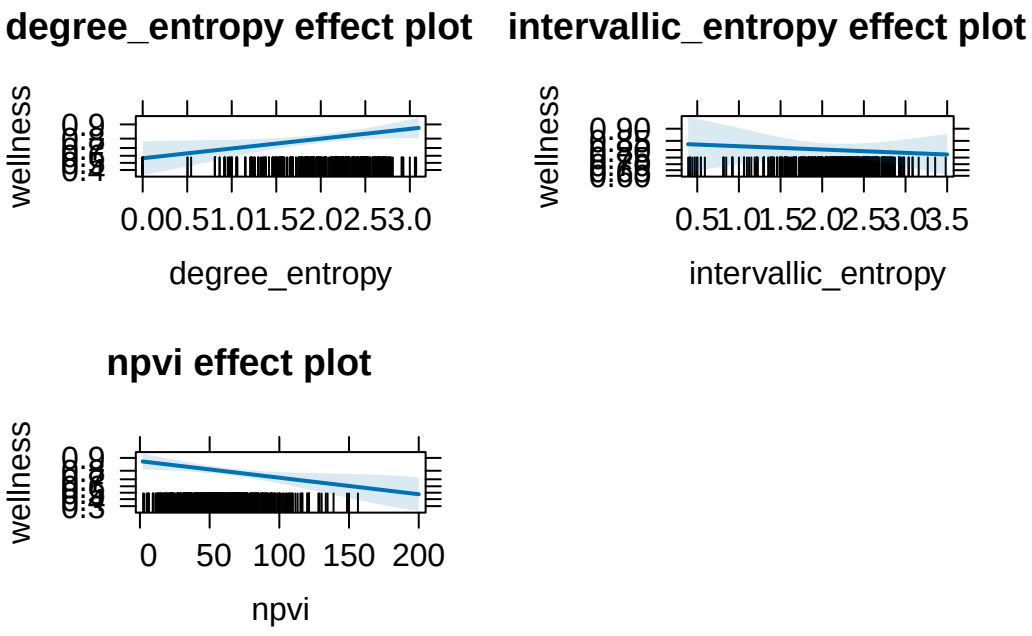
Since we have 2 significant variables, I created the model "glmSchubert" to calculate whether or not "intervallic_entropy" strengthened the model or not.It does not strengthen the model, as shown by the significant ratings of the anova() function and the lower Akaike's Information Criteria (AIC) for "glmSchubert"

```
glmTest <- glm(formula = wellness ~ degree_entropy +
                 intervallic_entropy +
                 npvi,
               data = schubertDataClean,
               family = binomial(link = "logit"))

summary(glmTest) #npvi and degree_entropy are significant at predicting wellness
```

```
Call:
```

```
glm(formula = wellness ~ degree_entropy + intervallic_entropy +
    npvi, family = binomial(link = "logit"), data = schubertDataClean)

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.138213   0.508389   2.239  0.02516 *
degree_entropy       0.558986   0.246322   2.269  0.02325 *
intervallic_entropy -0.126151   0.286897  -0.440  0.66015
npvi                -0.010441   0.003876  -2.694  0.00706 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 565.89  on 553  degrees of freedom
Residual deviance: 549.69  on 550  degrees of freedom
AIC: 557.69

Number of Fisher Scoring iterations: 4
```

```
plot(allEffects(glmTest)) ##to visualize the effect we can see the positive and negative i
```



Testing the Fit and Evaluating Parameters

Since we have 2 significant variables, I created the model "glmSchubert" to calculate whether or not "intervallic_entropy" strengthened the model or not.

It does not strengthen the model, as shown by the significant ratings of the anova() function and the lower Akaike's Information Criteria (AIC) for "glmSchubert". Notice in the ANOVA that I also have a significant chi-squared (p , .01), which is appropriate for a binomial glm since this indicates it is a useful model.

```
glmSchubert <- glm(formula = wellness ~ degree_entropy + #npvi and degree_entropy are sign
                             intervallic_entropy,
                           data = schubertDataClean,
                           family = binomial(link = "logit"))

summary(glmSchubert)
```

```
Call:
glm(formula = wellness ~ degree_entropy + intervallic_entropy,
    family = binomial(link = "logit"), data = schubertDataClean)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           0.4625     0.4271   1.083   0.2789
degree_entropy        0.5841     0.2451   2.383   0.0172 *
intervallic_entropy  -0.1266     0.2844  -0.445   0.6562
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 565.89  on 553  degrees of freedom
Residual deviance: 556.98  on 551  degrees of freedom
AIC: 562.98

Number of Fisher Scoring iterations: 4
```

```
anova(glmTest, glmSchubert, test ="Chisq") #evaluating test against model without npvi. Mo
```

```
Analysis of Deviance Table
```

```
Model 1: wellness ~ degree_entropy + intervallic_entropy + npvi
Model 2: wellness ~ degree_entropy + intervallic_entropy
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       550     549.69
2       551     556.98 -1  -7.2905 0.006932 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
glmTest$aic
```

```
[1] 557.6879
```

```
glmSchubert$aic #just to further check, and glmSchubert is definitely better.
```

```
[1] 562.9783
```

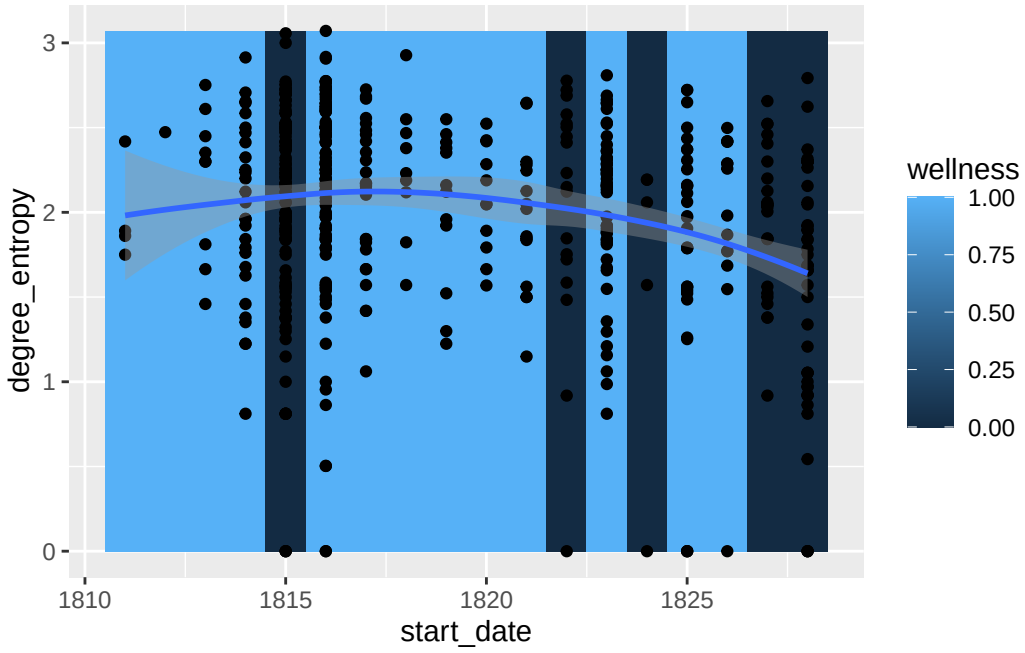
An Elbow

Since these two variables "degree_entropy" and "npvi" are predictors of wellness, we can conclude that musical output does indeed change with a composer's health. This interacts with various ideas of historical hypothoses pertaining to composers "predicting" their own deaths. In reality, this, at least in the case of Schubert, may be better understood as a decline in health which results in a change in style. Interestingly enough, the result is increased variability and disorder in Schubert's compositions as he ages.

The two graphs below capture one picture of this. Not the change in direction with the second onset of illness: this is the date when Schubert developed syphilis. It's particularly present in "npvi", the measure of rhythmic variability in his music.

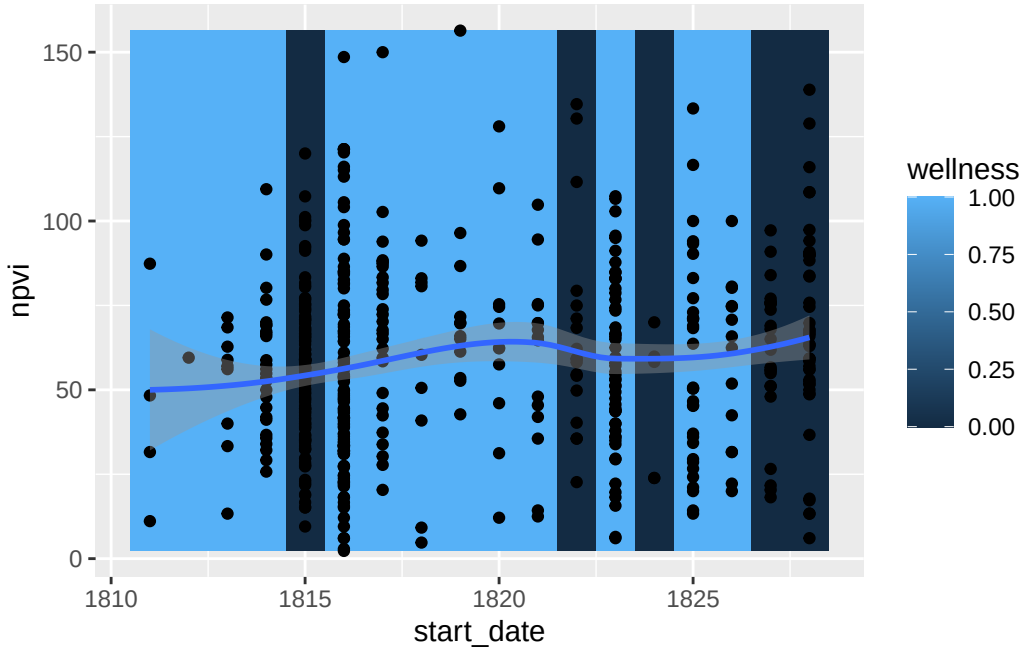
```
ggplot(schubertDataClean, aes(start_date, degree_entropy, wellness)) +
  geom_rect(aes(NULL, NULL,
                xmin=start_date - .5 , xmax=start_date +.5,
                ymin=min(degree_entropy), ymax=max(degree_entropy),
                fill=wellness
                )) +
    geom_point() +
    geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(schubertDataClean, aes(start_date, npvi, wellness)) +
  geom_rect(aes(NULL, NULL,
                xmin=start_date - .5 ,xmax=start_date +.5,
                ymin=min(npvi), ymax=max(npvi),
                fill=wellness)) +
  geom_point() +
  geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```


150
100
50
0
npvi
1810
1815
1820
1825
start_date
wellness
1.00
0.75
0.50
0.25
0.00