

# Notes from Monday, 1/28/19

ds

## Some Goals and Hypotheses

For this class, we started off with a number of hypotheses.

- H1: entropy between parts is correlated.
- H2: entropy in soprano parts is correlated with amount of suspensions
- H3: npvi is correlated with entropy of specific parts (not a specific hypothesis; I don't know which parts).

We really only got to the first hypothesis, and you will do the next two for Monday.

## Bash Script

Here's the bash script we came up with in class to get the entropy levels for each part. I've commented it somewhat thoroughly.

```
### this first line gives us a header for our output.
echo "bass tenor alto soprano"

## we run our script over every kern file.
for file in *.krn
do
  ### within these kern files, we wanted
  ### to go over every part, so we made which
  ### part a variable that went from 1 to 4.
  ### This was done in a couple of other ways in class.
  for part in `seq 1 4`
  do
    ### here we extract the part for each file,
    ### run the infot tool with the summary option,
    ### and then grep the line that outputs the average
    ### entropy. I used awk to only get the number and
    ### not the preceding text.
    extract -f $part $file | infot -s | grep 'Average' | awk '{print $6}' > $file.$part

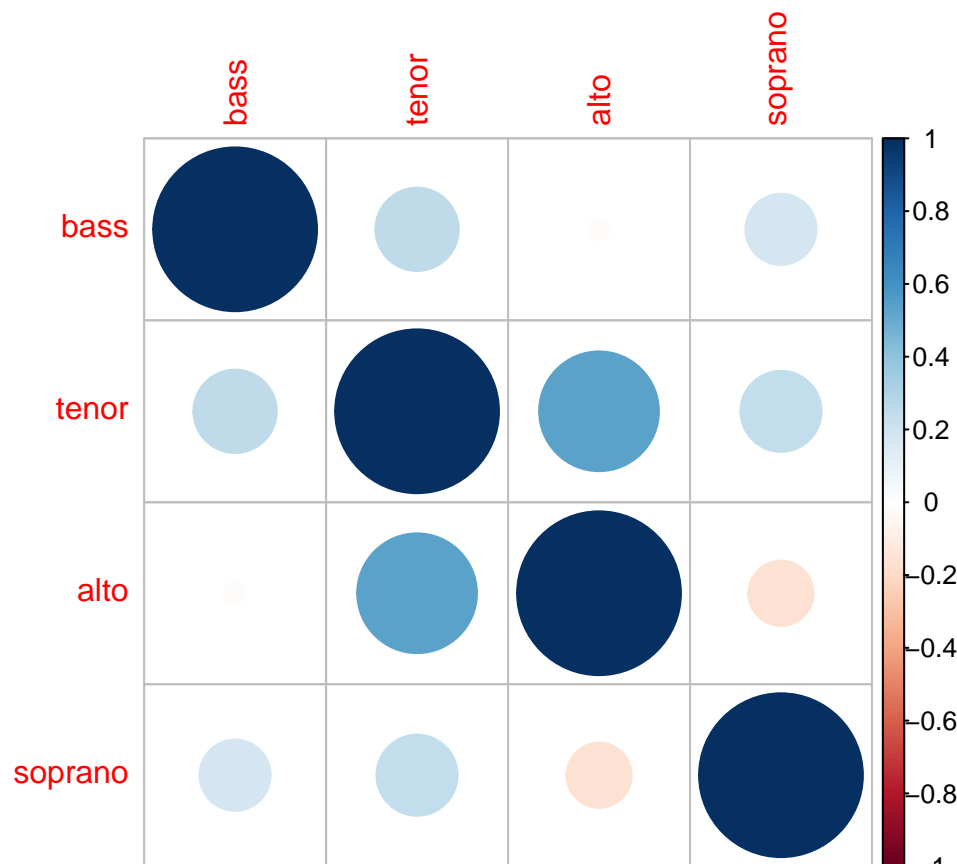
  done
  ### I then have lots of files for
  ### each piece (one for each part). I then paste
  ### them all together to get my output.
  paste $file.1 $file.2 $file.3 $file.4 > $file.info
  # I can use these things to get the points
  # that might relate to H2 and H3.
  npvi=$(./nPVI $file)
  meter=$(grep '\*M.*/' $file | awk '{print $1}')
  sus=$(grep -ch 'sus' $file)
  lnt=$(grep -ch 'lnt' $file)
  unt=$(grep -ch 'unt' $file)
  ant=$(grep -ch 'ant' $file )
done
```

```
## here I clean my folders of all the junky
## files I've creted.
rm -f *.[1-4]
rm -f *.info
```

## Visualizing the data

We might then examine what these correlations look like with the package *corrplot*. First, you run the *cor* function, and then then plot it.

```
setwd("~/corpora/humdrum_scores/Bach/Chorales/chorales.sample/")
my_data <- read.table("my_data.csv", header =T)
library(corrplot)
M<-cor(my_data)
corrplot(M, method="circle")
```



It would seem that alto and tenor parts are correlated in terms of *entropy*, but soprano and alto parts were not. This is quite interesting and, as was mentioned in class, could be related to the fact that the soprano line was traditional, and not written by Bach.

How might you proceed with the other hypotheses? I'd love to see some plots.