

# Why Branches are Hard

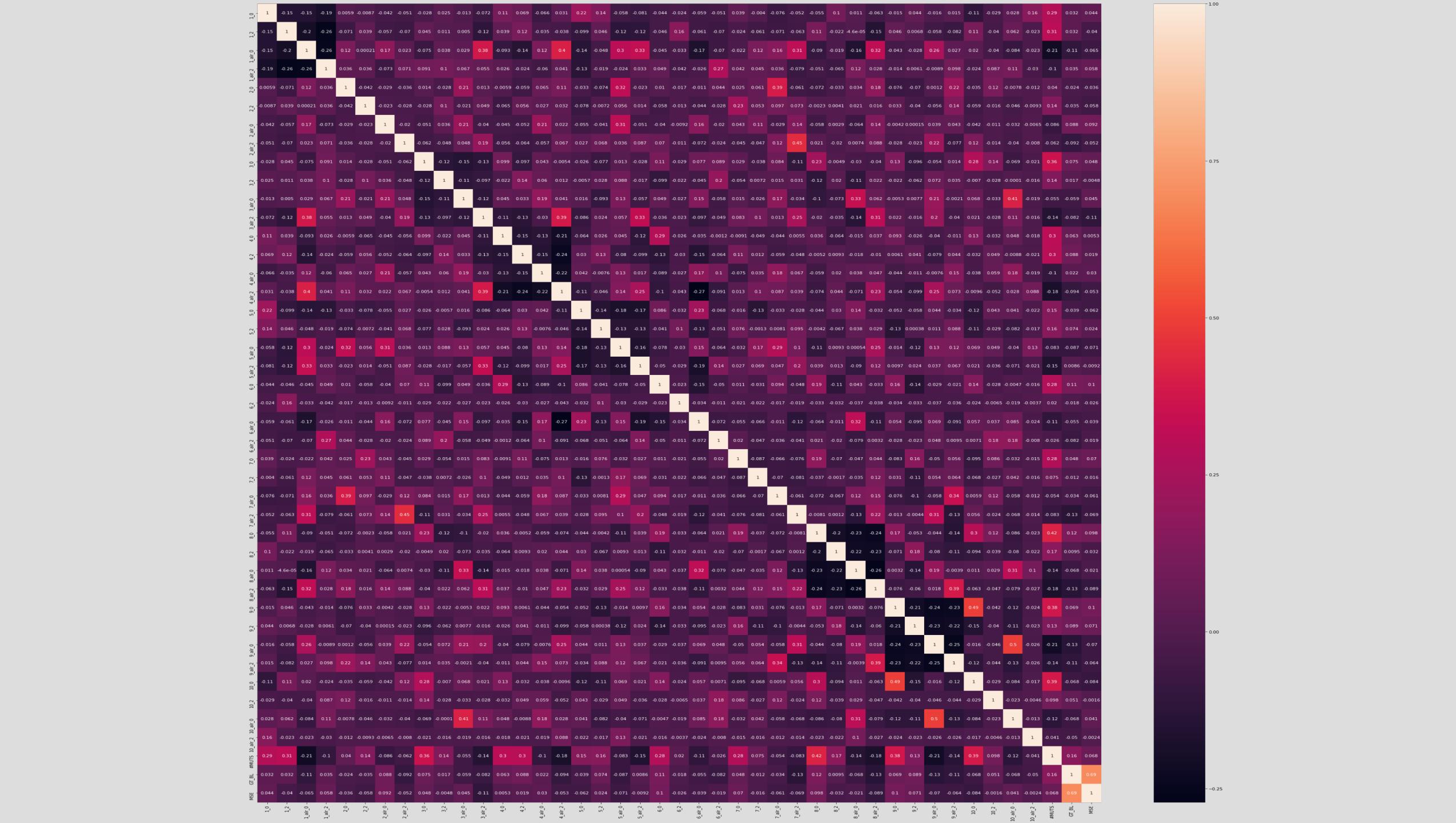
A tale told through graphs

# Some Tips and Tricks

- All findings are derived for pair data from the training set (not trees)
- Branch lengths are normalized.
- Contents:
  1.  $BL \sim Mut$
  2.  $BL \sim Mut + [Mutation\ 1 \rightarrow j\ at\ Trit\ i]$  for  $i \in [1, 10]$  and  $j \in \{0, 2\}$
  3. Heatmap of variables:  
$$['1' \rightarrow '0', '1' \rightarrow '2', 'already\_0', 'already\_2'] * 10 := [i\_0, i\_2, i\_alr\_0, i\_alr\_2]$$
for trit i  
+ [#Mutations, BL, MSE from FFN prediction to GT\_BL]
  4. TLDR on my progress







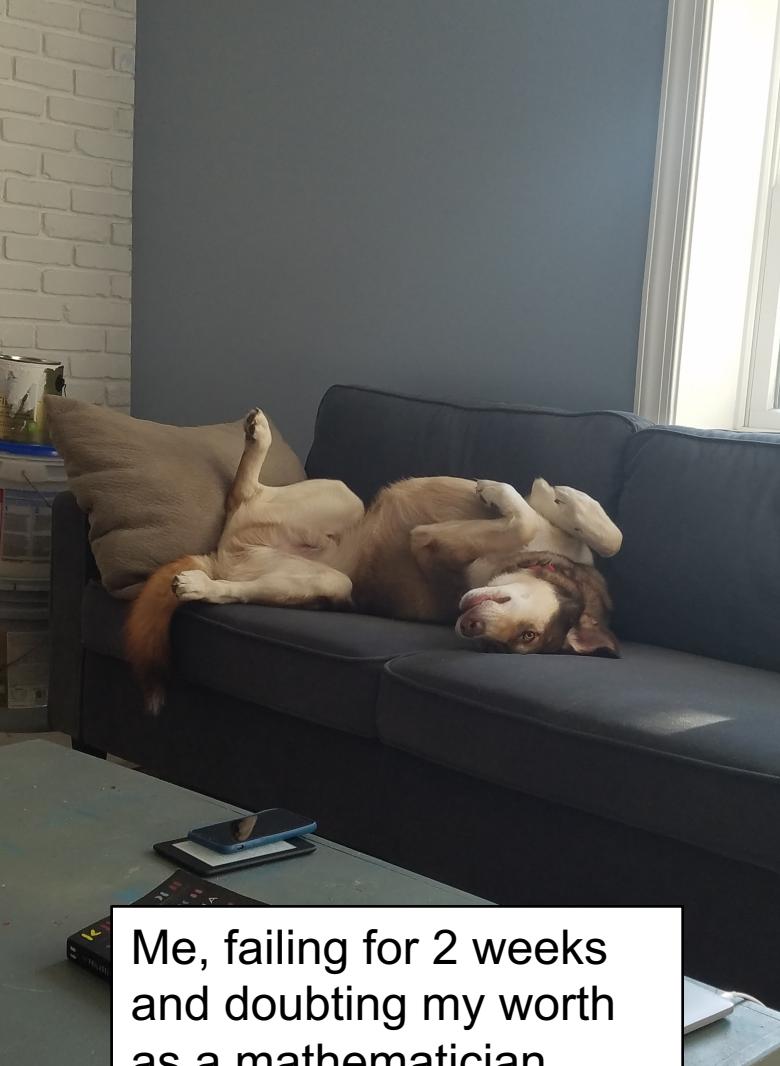
# TLDR on Current State

- Error minimizing method did not work
- Max likelihood method did not work
- Current application is a Feed Forward NN with an average L2 loss of  $0.031 \pm 0.053$  on the test data
  - Architecture based on complexity of problem
  - Input is mutation data, data on previously mutated trits and # of mutations (41 variables)
- Variance largely attributed to cases where the network predicts a shorter than actual BL

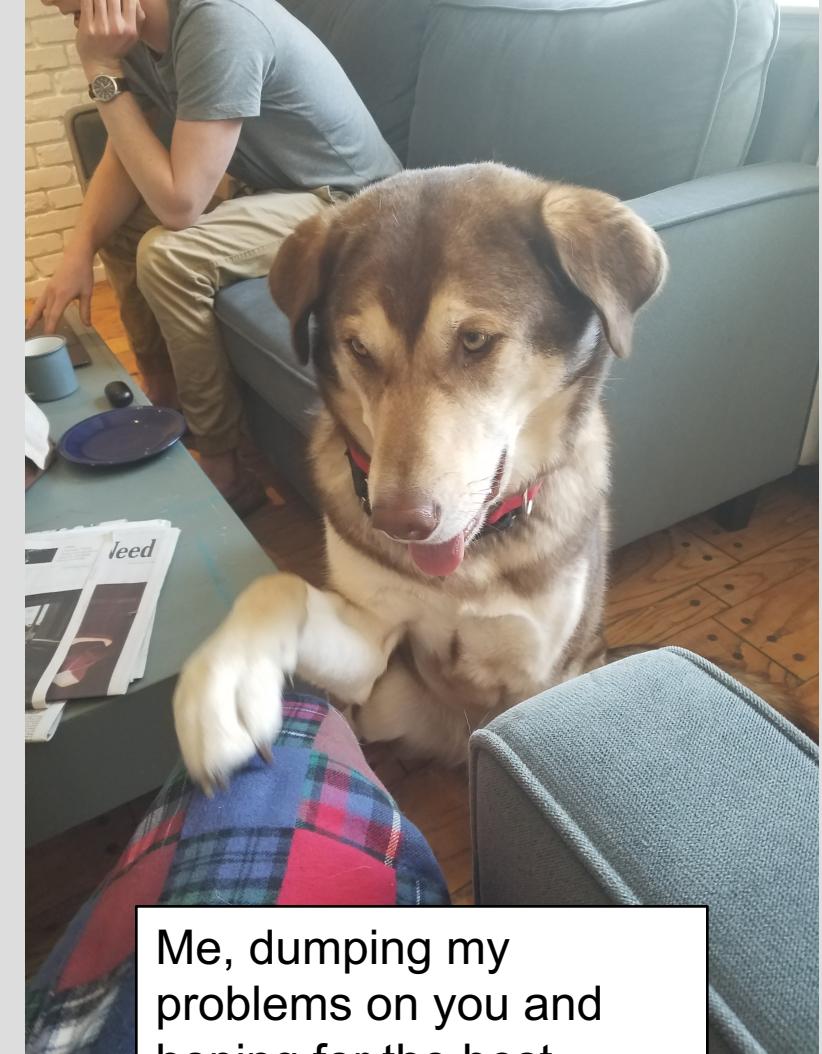
# My Journey as Told by Auggie



Me, excited to easily derive lengths from ancestry data



Me, failing for 2 weeks and doubting my worth as a mathematician



Me, dumping my problems on you and hoping for the best

**Thanks for coming to my Ted Talk!**