411 Assignment 2 thoughts:

* Highly unbalanced classes
* In paper, they ignore the Q AAMI class since “it is practically nonexistent”.
* Only MLII provided by all records (so let’s use that?)
* Perform some sort of pre-processing (in the paper they use baseline removal, others use high-frequency noise filtering too).
* In the paper, they use the labels for segmentation, as they focus on classification.
* In the paper, they standardize the data
* In the paper, they add weights to the classes to help deal with imbalance.
* Can’t use accuracy (imbalanced classes). In the paper they use jk index (which we could use to allow for comparison).
* <https://archive.physionet.org/physiobank/annotations.shtml> annotations explained
* <https://i.imgur.com/mFdm9U9.png> standard classes from annotations
* <https://www.scopus.com/inward/record.url?eid=2-s2.0-0034953193&partnerID=10&rel=R3.0.0> more on the dataset
* <https://github.com/mondejar/ecg-classification> for a bunch of code based on the paper, kind of very helpful lol…

Progress:

* Data processing script
* Do we use all samples, or just ones with corresponding annotations?
  + Plotted both
* Baseline removal
  + Frequency is 360Hz, so the 200ms will have a window of:
    - 360\*0.2 = 72 + 1 (to make it odd)
  + And the 600ms will have a window of:
    - 360\*0.6 = 216 + 1 (to make it odd)
* Segmentation
  + In practice, an algorithm would have to be used to detect QRS
  + In the paper, they use a window of size 180 (+/- 90 centered on annotation)
  + Let’s see if we can get to a similar number ourselves…
    - Nope, ended up with 137 :’(
    - So I’ll just copy the paper again…
  + ~~Centered windows on peak (which can be slightly different than the annotation, i.e., the annotation is in a trough) with search of +/- 20~~
  + Put into sensible csv format