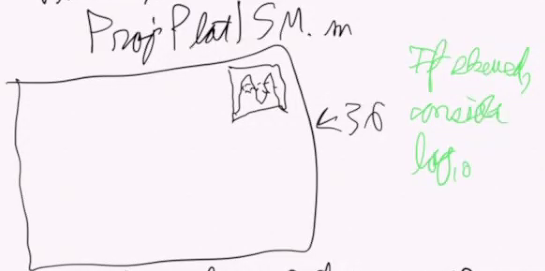
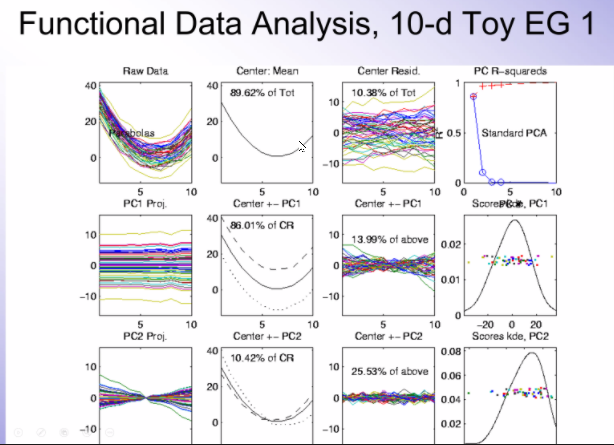
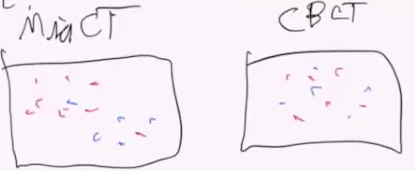
## Consulting notes 2020/8/31

Paper notes: Automatic quantification framework to detect cracks in teeth, Proc SPIE Int Soc Opt Eng, 2018

* Fig3. What’s the rationale behind multiple level HP/LP fiter?
* Fig4. What is Simoncelli Mother Wavelet?
* Fig6. Since PC1 seems to help classify healthy vs. cracked teeth, What about visualization of PCA 1 weights?
* Can ask Jack Prothero for help

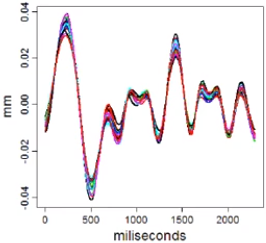
Plan:

* 2 datasets from the same 36 teeth (n=36, 22 with cracks, 14 healthy)
  + MicroCT
  + High solution CBCT
* For each tooth, has size of connected pixels (there are false positives)
* Treat distributions as data object
  + <http://marron.web.unc.edu/stor-881-object-oriented-data-analysis-fall-2019/>
  + Visualize all 36 distributions (projplot1SM)
    - 
  + 4. Using matlab
  + 7. Slide 28 marginal distribution plot
    - Make a big panel of plots: each one with set of dots, and 36 of those
  + See if there is a difference between crack and healthy teeth
    - How distribution compare
    - 1. Represent each by “quantile function”, using cquantSM.m
    - 2. Visualize using PCA via curdatSM.m and scatplotSM.m
  + Slide 55: probability density as data objects
    - Quantile: best modes of variation
      * Shift will be mean variance
      * Tilt will be variance
      * STOR 881:
    - Jive (joint individual variation) method how things work together and separately (H&E correlate with genomics), individual variation. 12/3/2019 slide 80
      * 
      * Use different color to represent healthy vs. crack
  + Hope:
    - microCT
    - CBCT (maybe more overlap) – over-processed, might be hard to see differences
    - 
  + Find DWD direction between classes (healthy vs. crack)
    - DWD is better way of (direction in the space to separate the 2 classes)
    - Plot scatplot SM use DWD as input and orthogonal PC
    - Quantify use diproperm hypothesis test (no.4 slide, P29 DiProPerm)
    - Instead of using P-value, use z score, (how many std is the data from the null distribution mean)
    - 1 is there a stat sig. how diff is it (compare z scores or p-values)

Todo:

1. Install package from <http://marron.web.unc.edu/sample-page/marrons-matlab-software/>
2. Go through dataset example LungCancer2011, comment relevant parts
3. Get teeth dataset, load in as object format
4. Plot 1: Distribution of 36 teeth
5. Plot 2: cquantSM

## Online video 2020/9/16

* <https://www.youtube.com/watch?v=SUp_Nq8NwfE>
* Functional data refer to curve data as a function of time
  + Often a large number of related quantities
  + Treat each replication as a single observation (an object)
* Features
  + Quantity
  + Frequency
  + Smoothness – most important
  + These data describe a process that changes smoothing, and continuously over time
  + Functional data analysis = analysis of data that are functions – domain is usually time, but can be space, energy, frequency, etc.
  + Functional data is often complicated
    - so not easily described by mathematical formulae (parametric form)
    - variation between replications even harder to describe
    - 
* Characteristics
  + Data are measurements of smooth processes over time
  + We usually do not want to make parametric assumptions about those processes
  + Often have multiple measurements of the same process
  + We are interested in describing the variation of processes
  + Frequently, collected data have high resolution and low noise
  + Can be applied to any estimate of a smooth process
  + 