Physical Therapy

Team: Yongrong Chai, Keliang Xu, Jack Carbaugh

Summary

Client: Kerry Costello from BU SAR Movement and Applied Imaging Lab

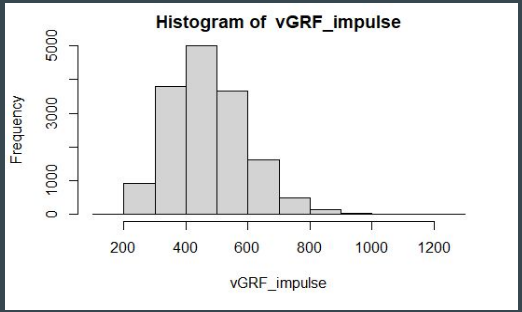
The client is studying osteoporosis of the knee by imaging subjects' walking motion over a series of pressure plates.

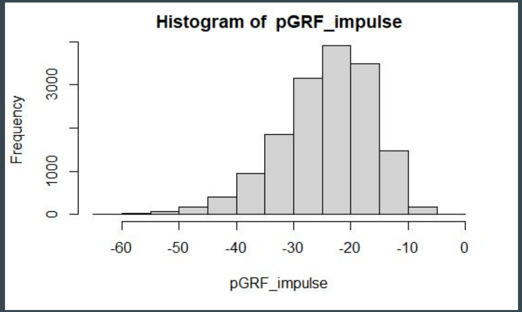
This process produces a lot of time series and discrete data, and the client would like to know if there are any noticeable patterns in the data, if we can use dimension reduction to locate the most important elements.

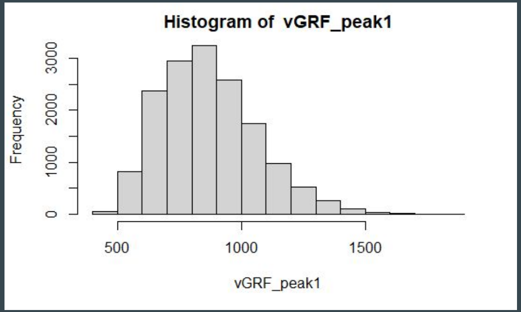
Histograms of the discrete data

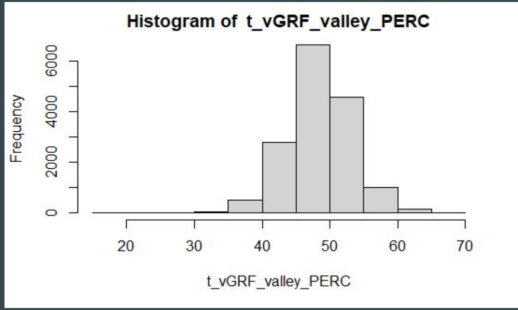
Most discrete variables appear to be normally distributed, with some being skewed

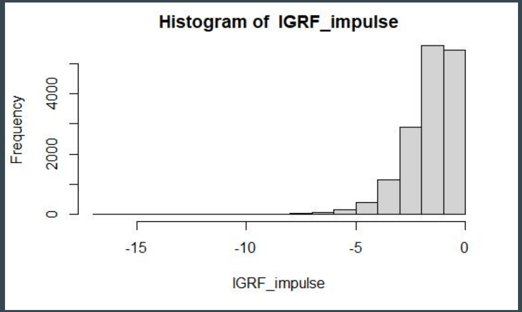
A few variables likely need to be log or logit transformed

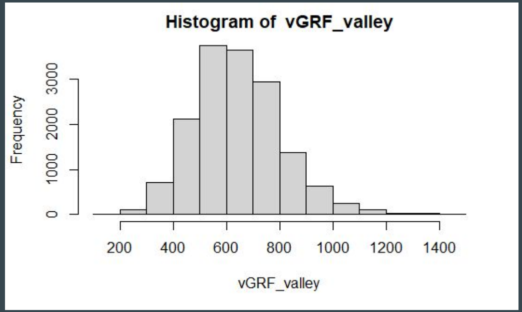




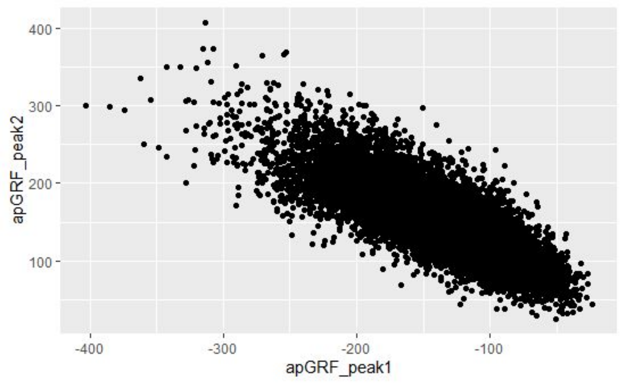


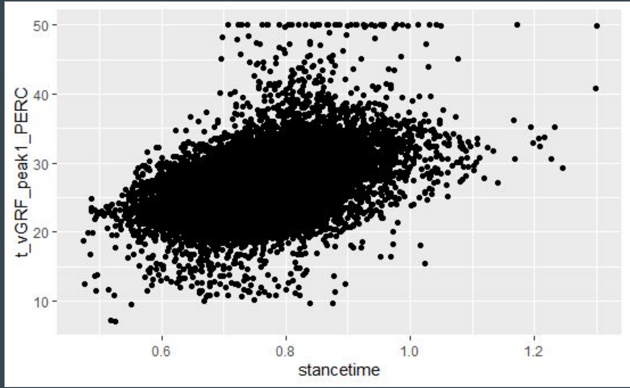




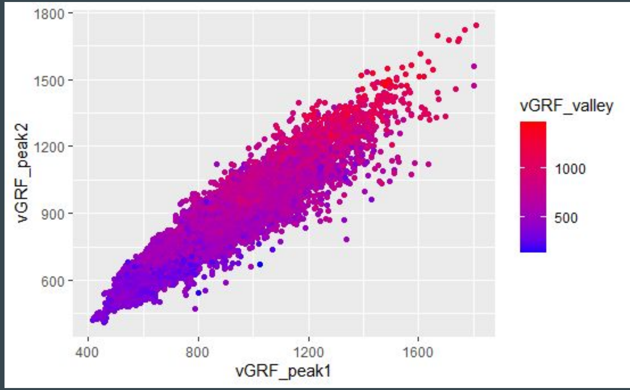


Scatter Plots of Discrete Data

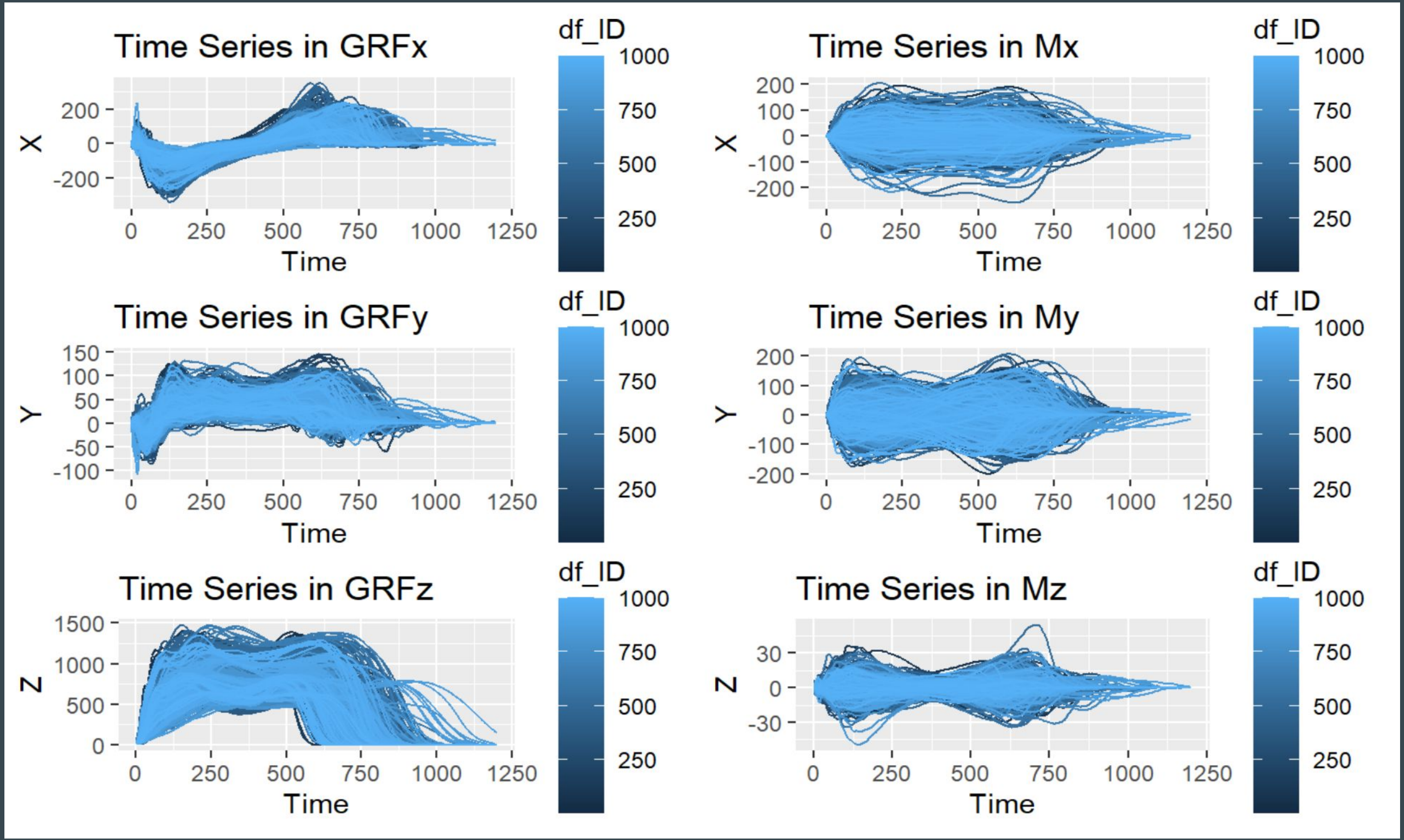




Not many scatter plots showed clear or strong relationships. Those that did mostly involved related peak or valley movements.



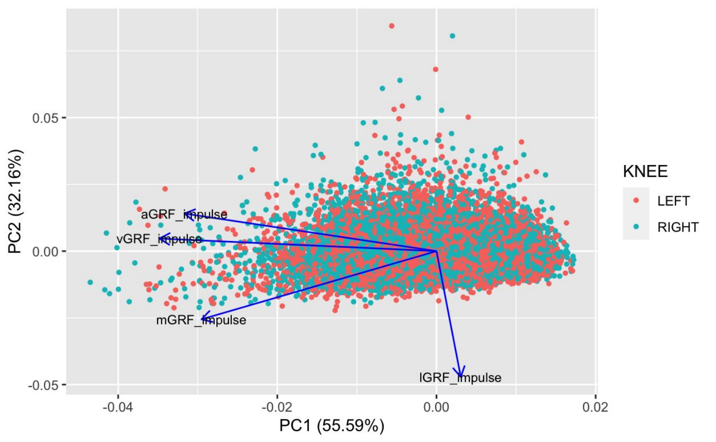
Plot time series data as line plots

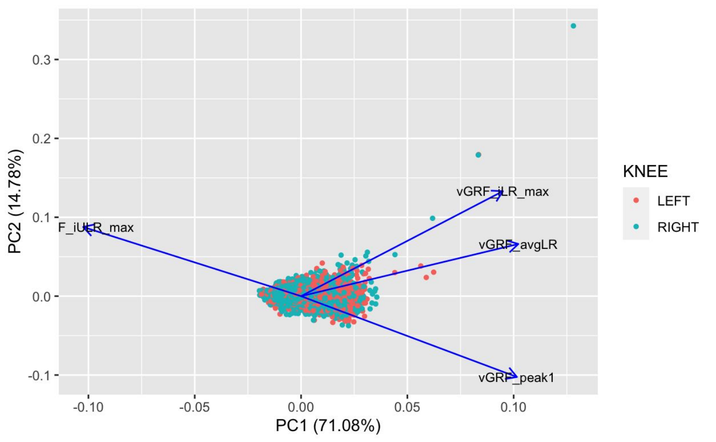


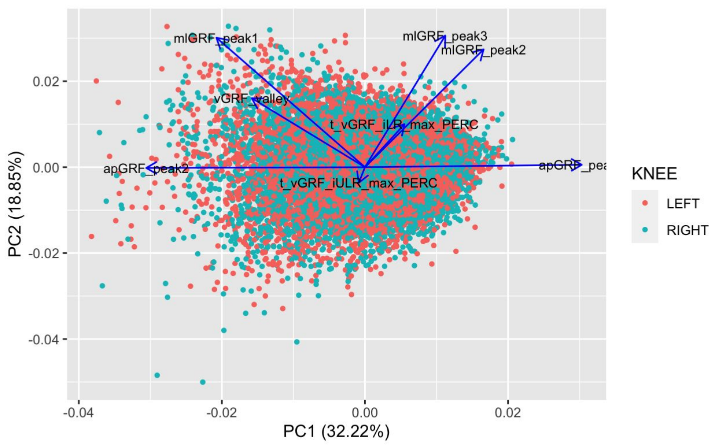
GRFx,y,z:
force in x,y,z

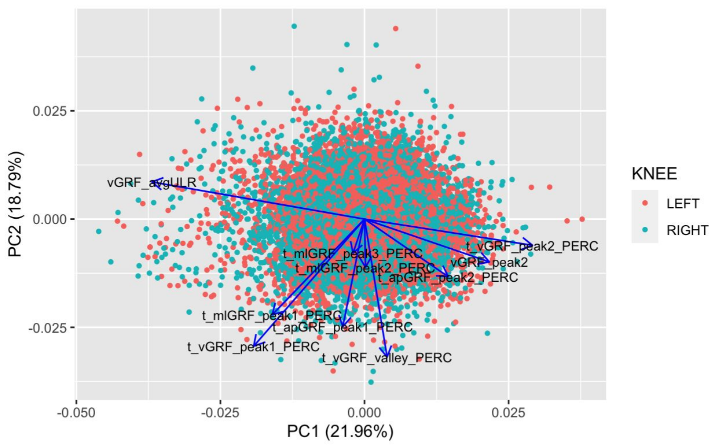
Mx,y,z:
moment in x,y,z

PCA Visualization









Four units:
N/S
N
N*S
%stance

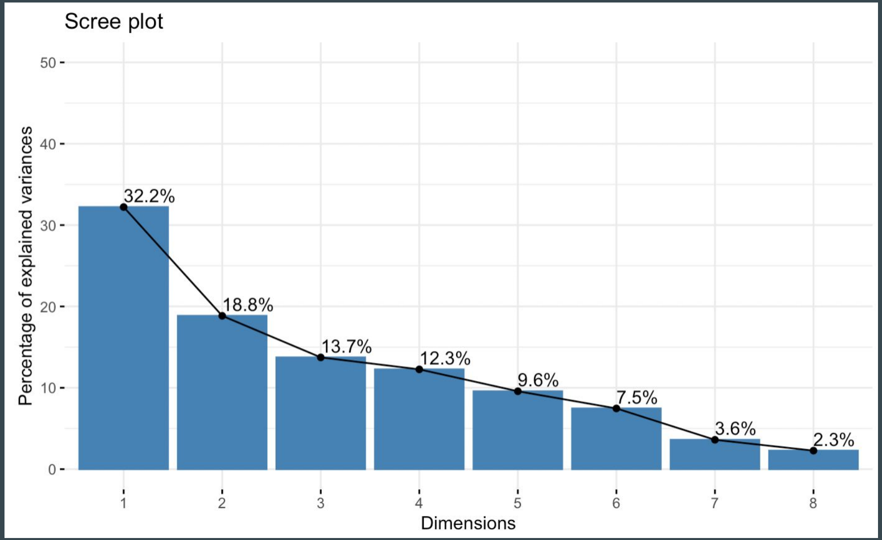
PCA - Interesting Finds

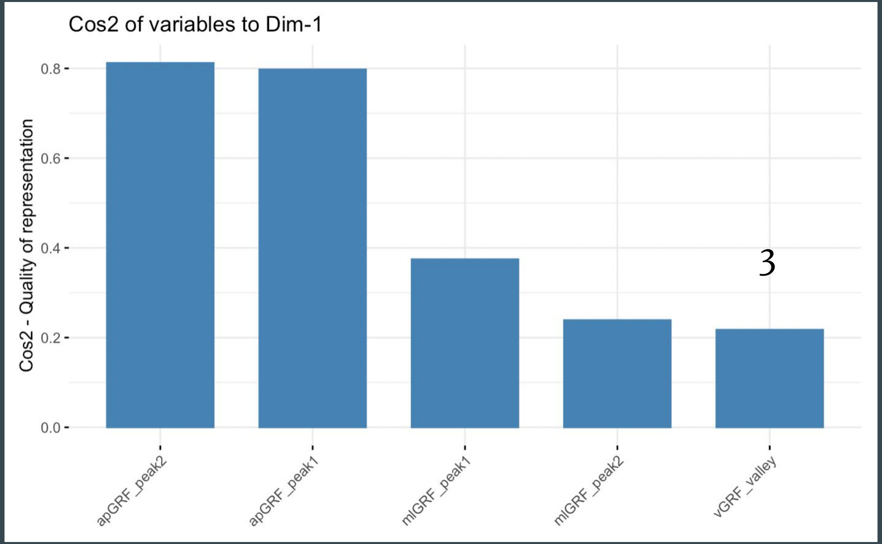
```
pca.impulse.summary$importance #first2 - 0.555930+0.321630
pca.unloading.summary$importance #first2 - 0.710820+0.1478000
pca.peak.summary$importance #first5 - 0.322190+0.188500+0.137390+0.1226800+0.0957700
pca.time.summary$importance #first6
```

	PC1	PC2	PC3	PC4
Standard deviation	1.491218	1.134256	0.5719785	0.4032038
Proportion of Variance	0.555930	0.321630	0.0817900	0.0406400
Cumulative Proportion	0.555930	0.877570	0.9593600	1.0000000

PCA

1


Scree plot
Percentage of explained variances
Dimensions
32.2%
18.8%
13.7%
12.3%
9.6%
7.5%
3.6%
2.3%
50
40
30
20
10
0
1
2
3
4
5
6
7
8


Cos2 of variables to Dim-1
Cos2 - Quality of representation
0.8
0.6
0.4
0.2
0.0
apGRF_peak2
apGRF_peak1
mlGRF_peak1
mlGRF_peak2
vGRF_valley

3

Important Factors

vGRF_avgLR
vGRF_iLR_max

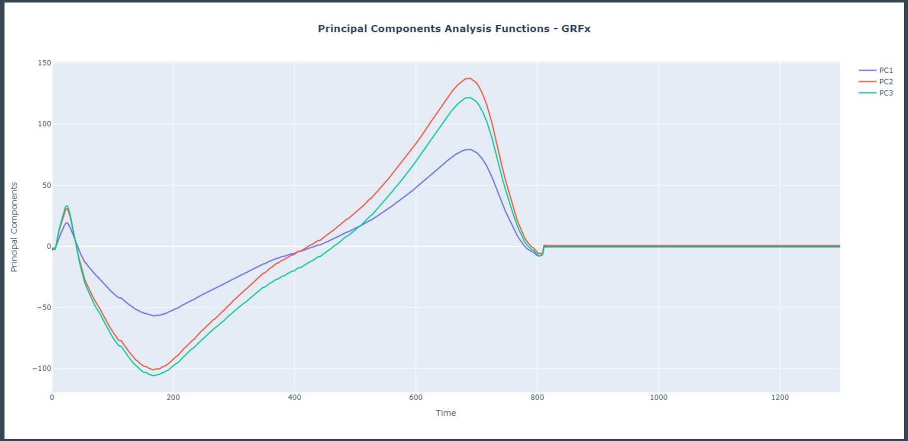
vGRF_impulse
mGRF_impulse

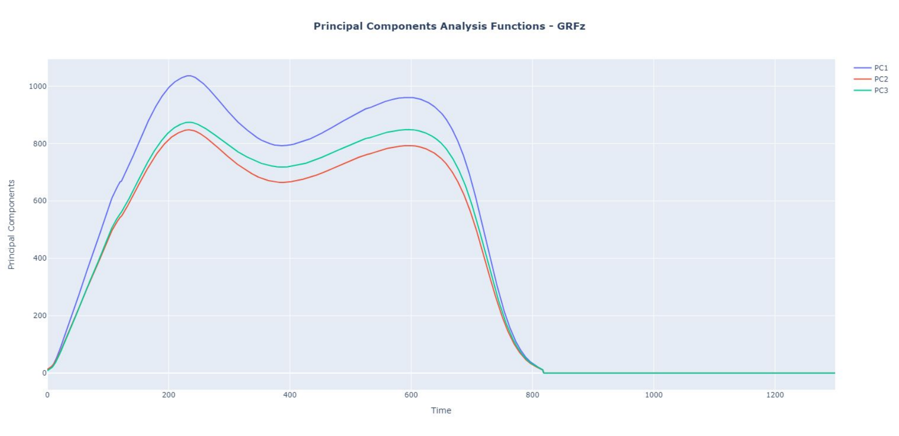
vGRF_avgULR
vGRF_peak2
t_vGRF_peak1_PERC
t_vGRF_peak2_PERC
t_vGRF_valley_PERC
t_apGRF_peak1_PERC

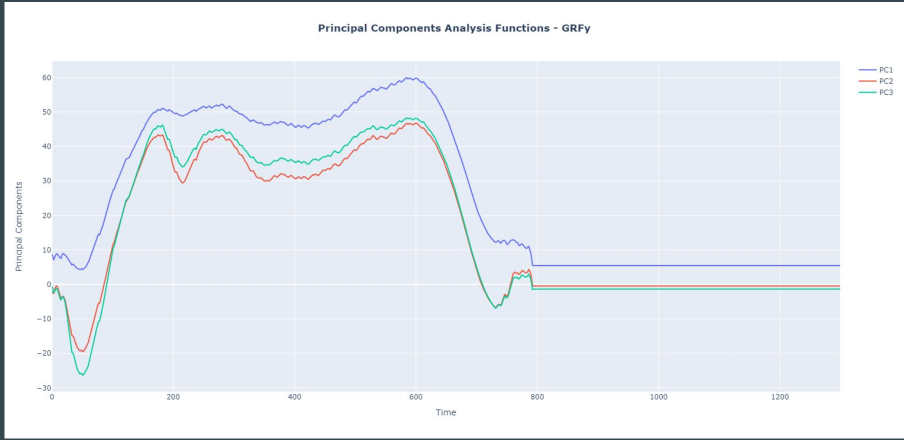
vGRF_valley
apGRF_peak1
apGRF_peak2
mlGRF_peak1
mlGRF_peak2

Baseline Model - Functional PCA

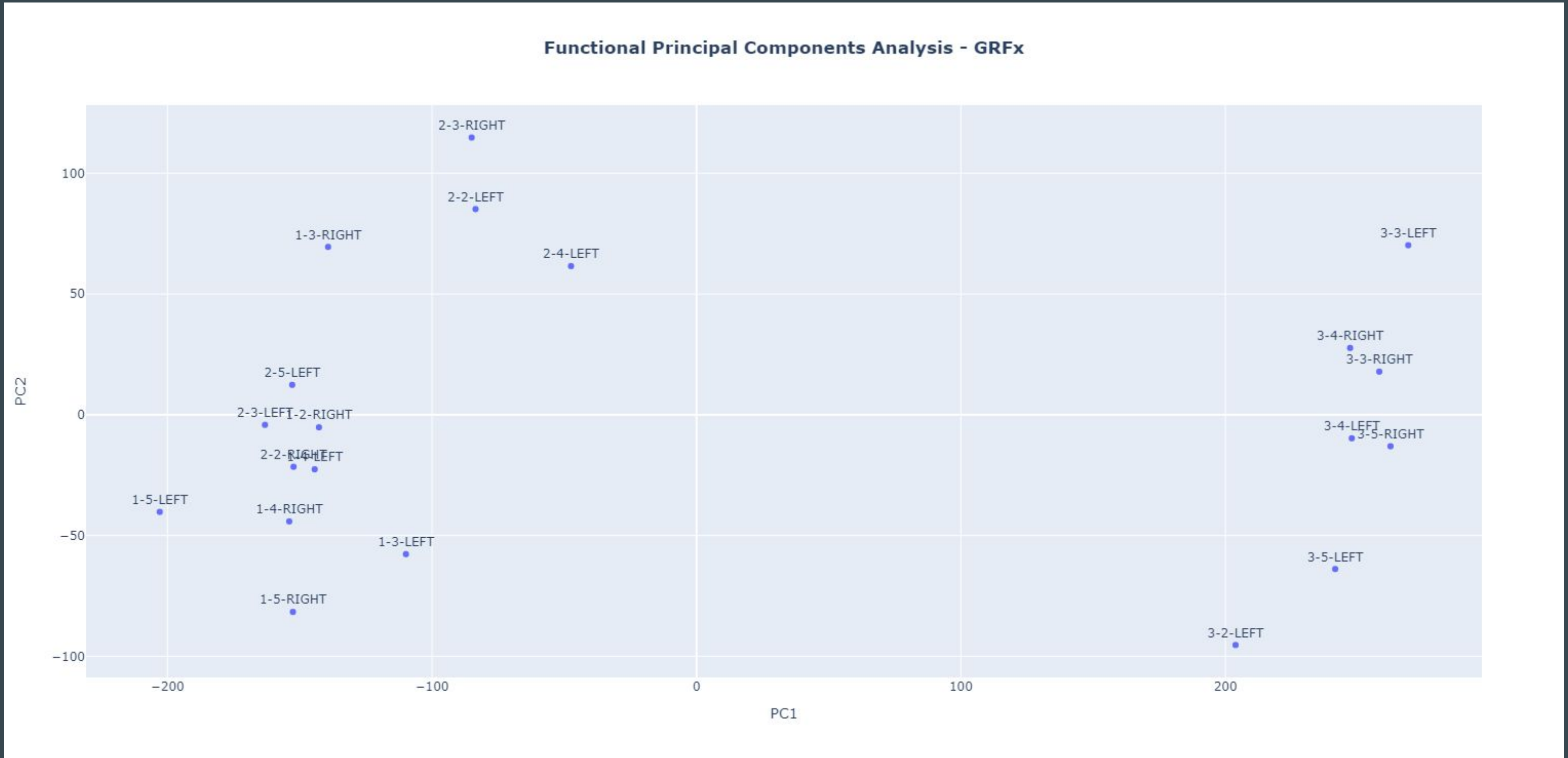
Started by running FPCA on GRFx, GRFy, and GRFz. Computing was slow, so we started with a small sample



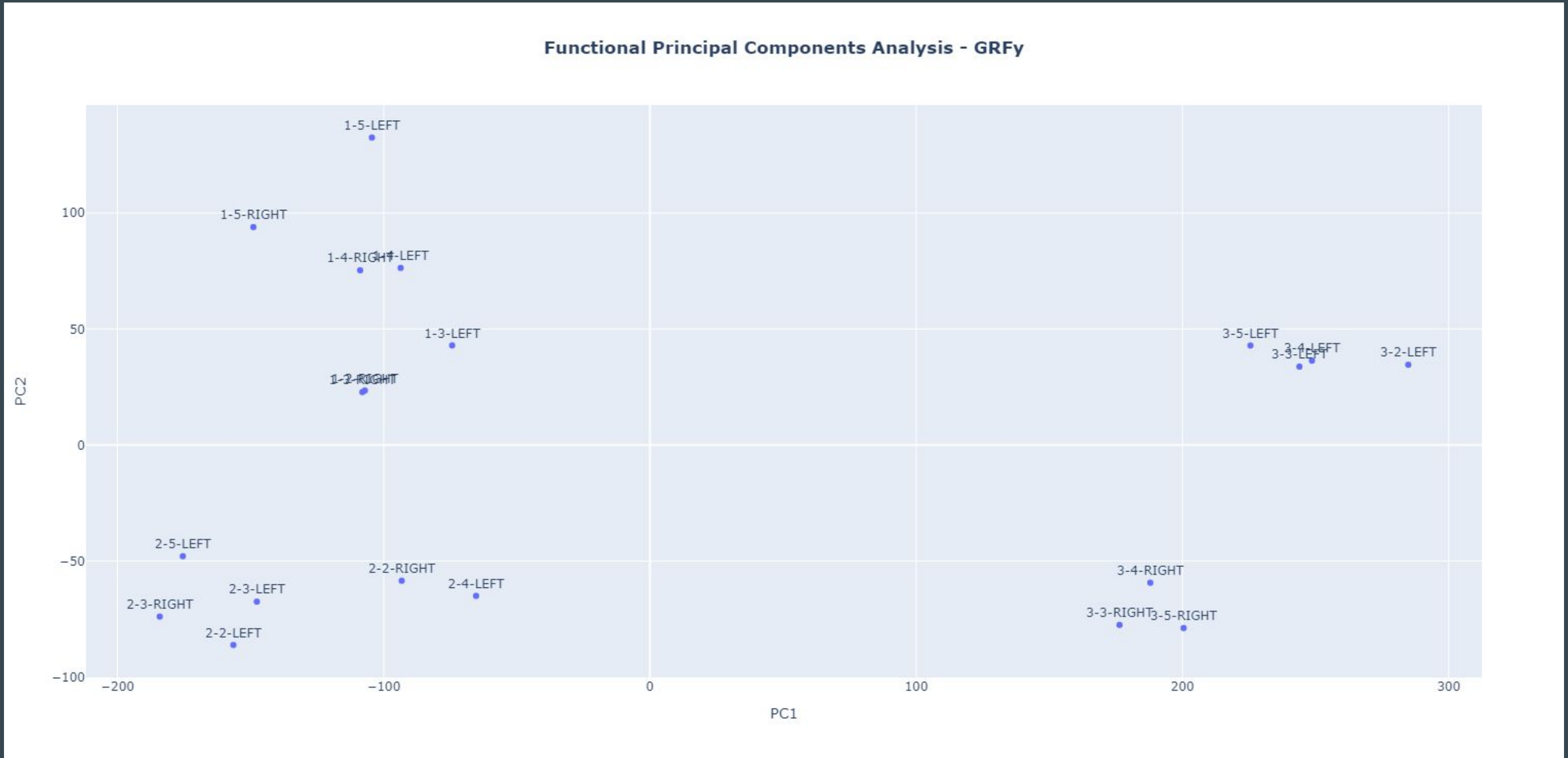




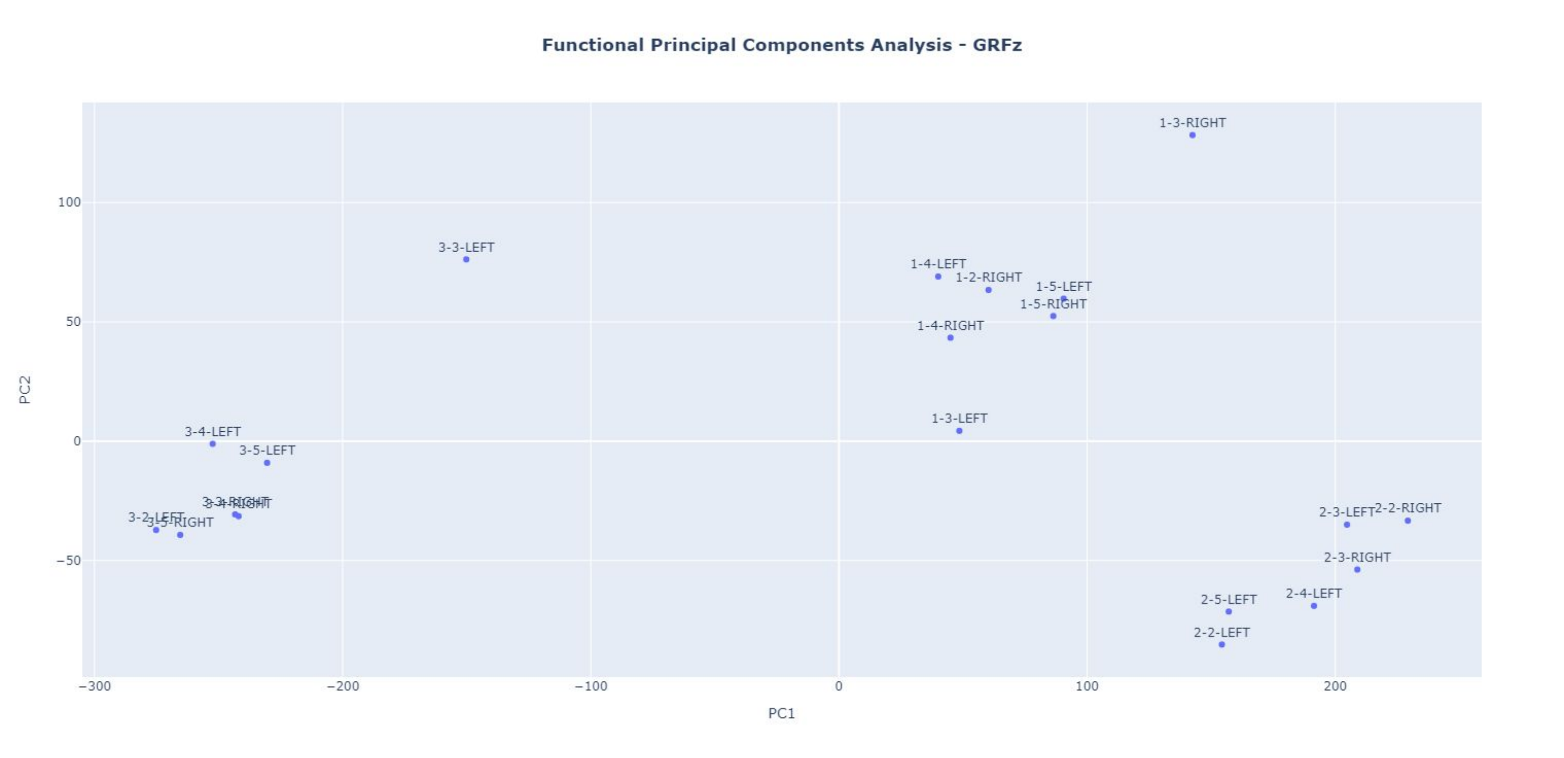
Plotting PCs for Observations: GRFx



Plotting PCs for Observations: GRFy



Plotting PCs for Observations: GRFz



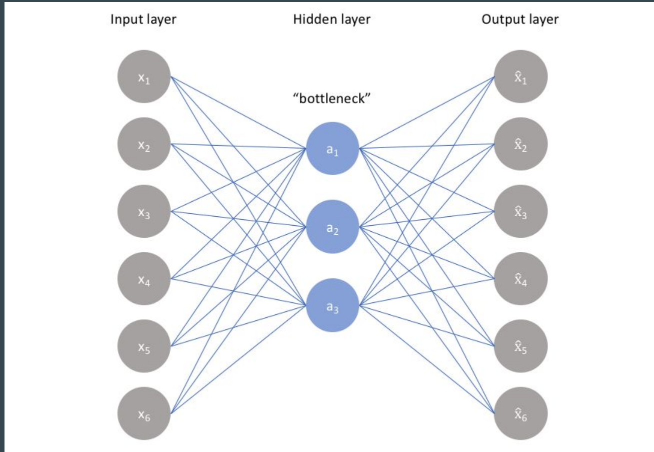
Autoencoder

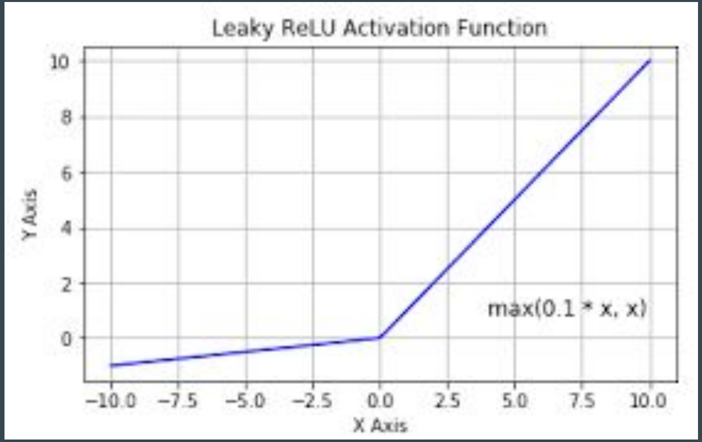
Vanilla Autoencoder used on the non-normalized multivariate time series (GRFx, GRFy, GRFz, Mx, My, Mz, COPx, COPy)

Autoencoders are neural networks that reduce the dimensionality of the data with the purpose of reconstruction

First scaled each variable to 0 - 1 scale to avoid skewed results.

Keras - One layer used; Leaky Relu activation. Dimensionality reduced down to 100 variables.

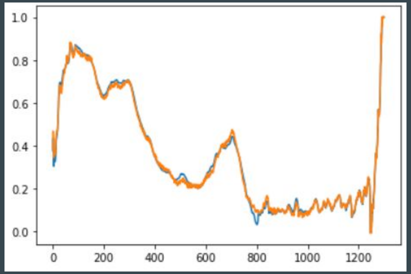




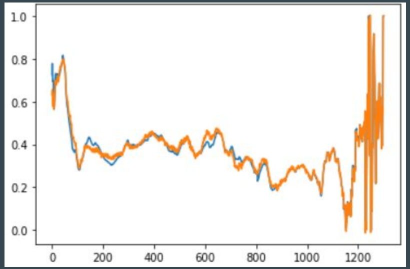
Autoencoder Reconstruction Results

Sample Reconstruction results: (Blue = actual, Orange = autoencoder result)

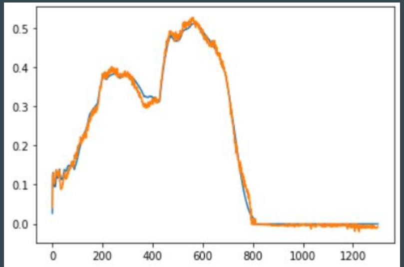
GRFx



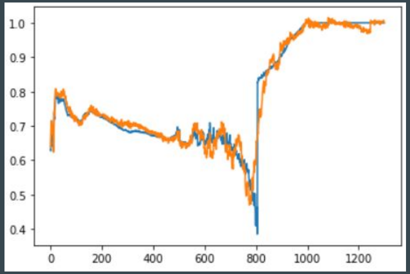
GRFy



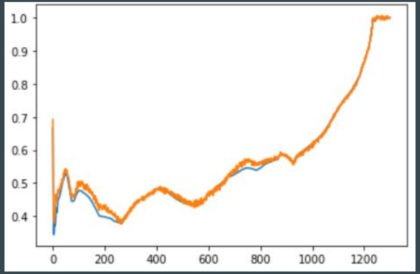
GRFz



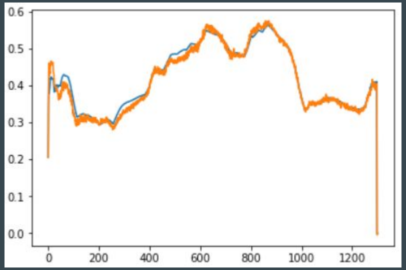
COPx



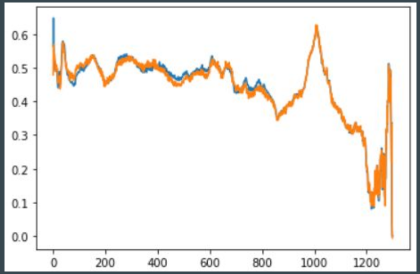
Mx



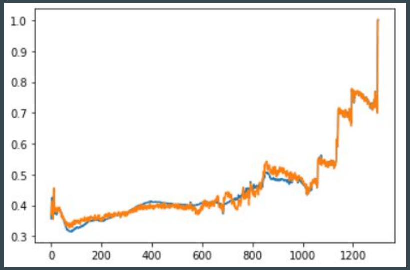
My



Mz



COPy

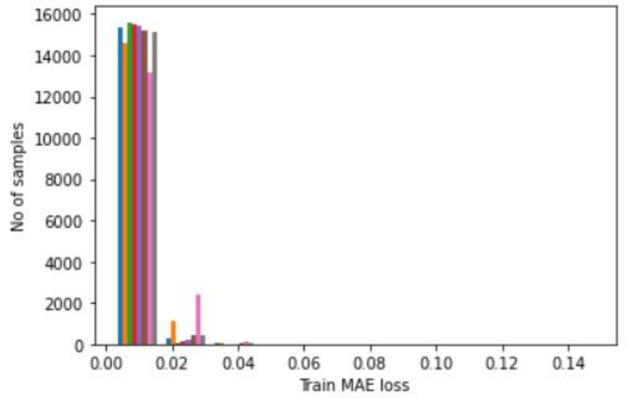


Autoencoder Anomaly Analysis

We can observe sequences that the autoencoder had trouble recreating, as these may be anomalies to the average.

Some subjects had a large amount of error for multiple of the variables

Used loss > 0.03, but can use tighter/more specific bound. Gave 215 high-error samples

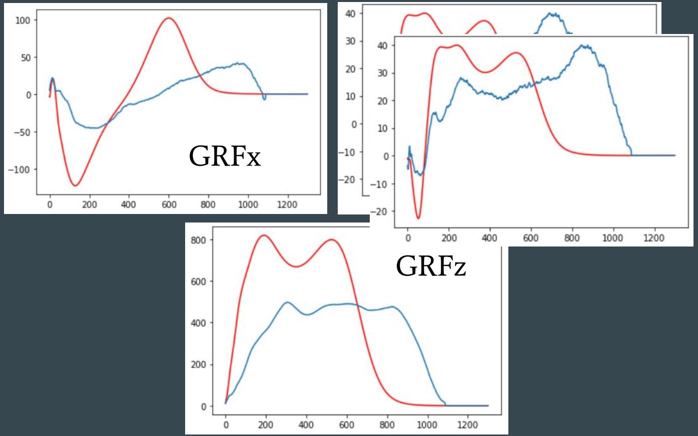


797, 798, 800 -> Subject 123

1803 - 1806 -> Subject 273

3789 - 3792 -> Subject 572

Example: 797



Summary

EDA of time series and discrete data

Important factors according to PCA

The autoencoder - Reduce the data down to 100 dimensions

The reconstruction - Locate anomalies in the dataset