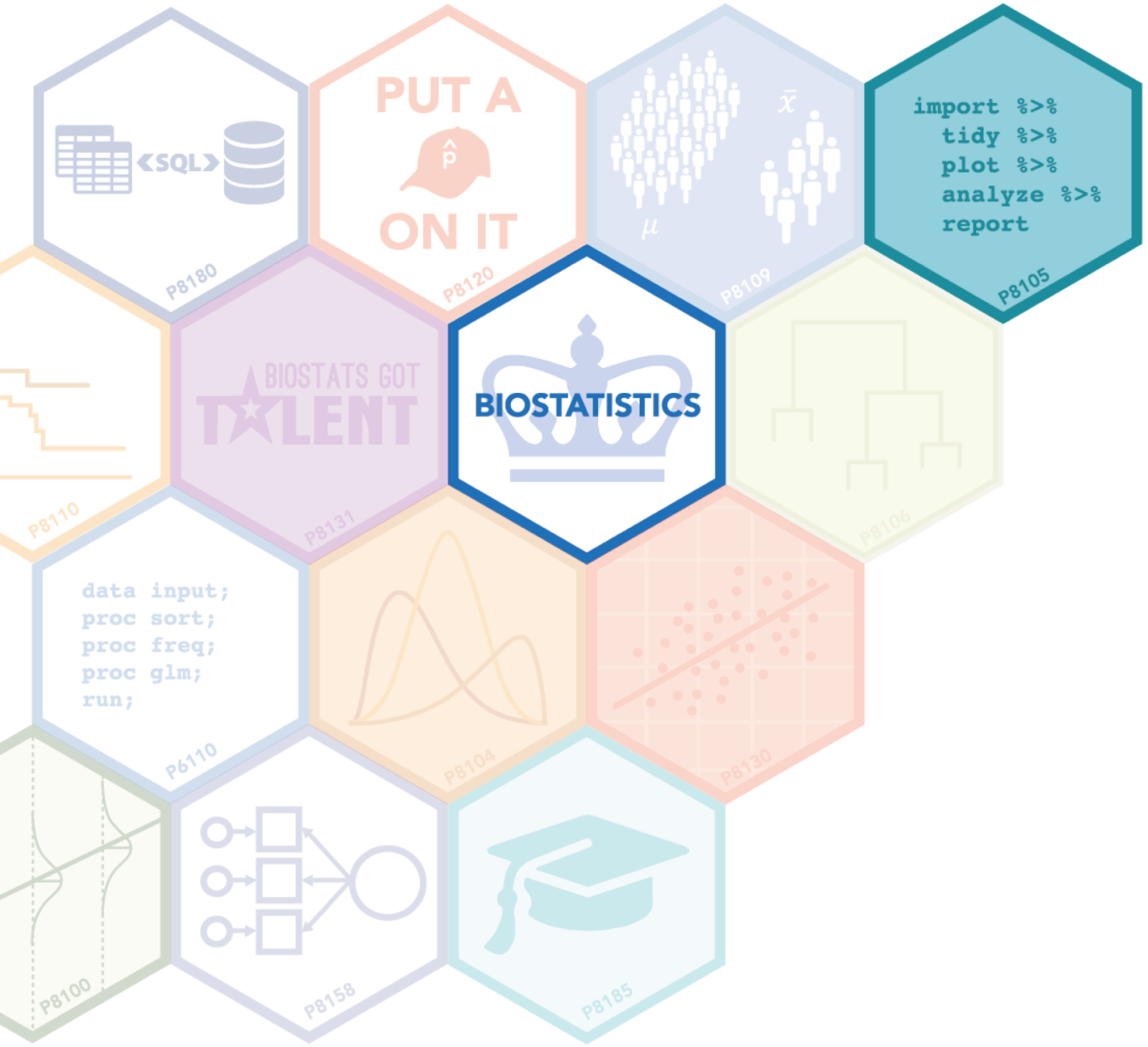


NOTATION FOR FDA

Jeff Goldsmith, PhD
Department of Biostatistics

What are "Functional Data"?

Tentative definition:

Observations on subjects that you can imagine as $X_i(t_i)$, where t_i is continuous

Functional notation is conceptual; observations are made on a finite discrete grid.

Characteristics of FD

- High dimensional
- Temporal and/or spatial structure
- Interpretability across subject domains

Discretization

- Conceptually, we regard functional data as being defined on a continuum, e.g. Xi(t), 0 < t <1
- In practice, functional data are observed at a finite number of points
- Observation grid is often regular and dense – many observations for each subject, all over a common collection of time points
  - Minute of the day
- At each observation point t, Xi(t) has a distribution

Summaries of FD

- Suppose we have functional data $\{X_i(t),\, t \in [0,1],\, i = 1, \ldots, n\}$

- Mean: $\mu(t) = E\left[X_i(t)\right]$

- The mean is itself functional
- Typically, we assume that the mean is smooth.
- “Raw" estimator is sample mean:

$$\frac{1}{n}\sum X_i(t)$$

- A typical estimator of $\mu(t)$ would be a smoothed version of this

Summaries of FD

- Suppose we have functional data $\{X_i(t),\, t \in [0,1],\, i = 1, \ldots, n\}$
- Variance: $\Sigma(s,t) = \mathrm{Cov}(X(s), X(t)) = E\left[(X(s) - \mu(s))(X(t) - \mu(t))\right]$
- This is a (two-dimensional) surface
- “Raw" estimator is sample covariance:

$$\hat{\Sigma}(s,t) = \mathrm{Cov}(X_i(s), X_i(t))$$

- Would need to smooth this as well.

Data displays

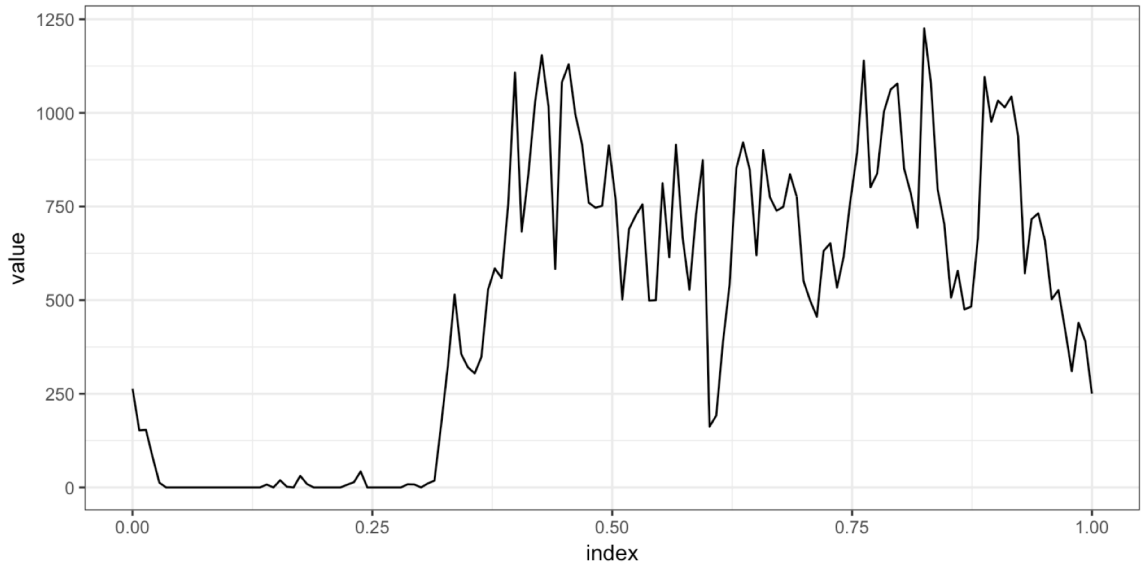
- Spaghetti plots
- Rainbow plots
- Lasagna plots

Switch to code

Code

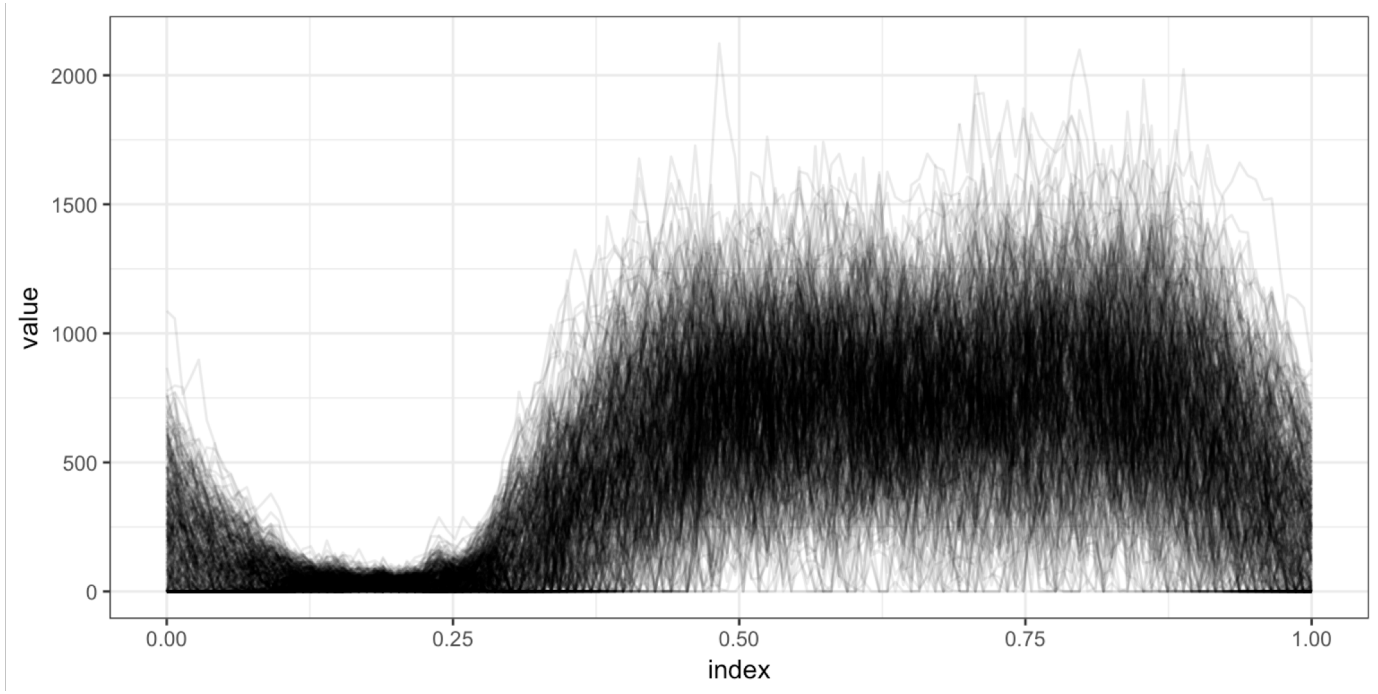
```
load("./DataCode/HeadStart.RDA")
```

```
as_refundObj(accel) %>%
    filter(id == 1) %>%
    ggplot(aes(x = index, y = value)) + geom_path()
```



Code

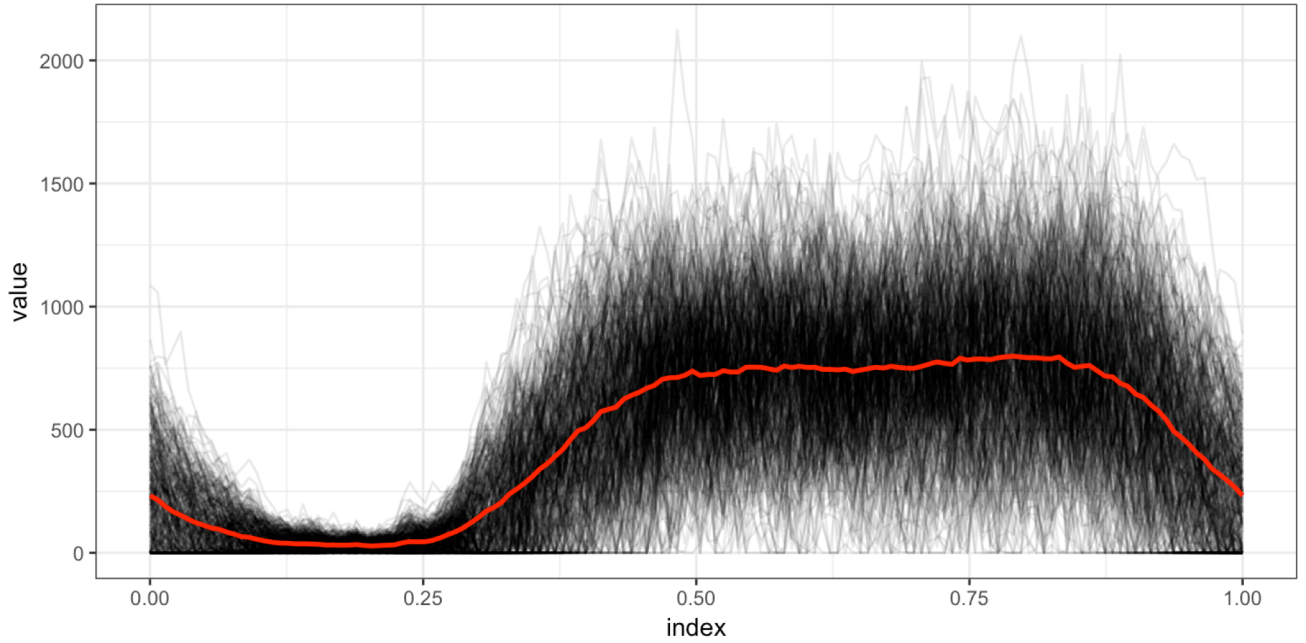
```
as_refundObj(accel) %>%
    ggplot(aes(x = index, y = value, group = id)) + geom_path(alpha = .1)
```



Code

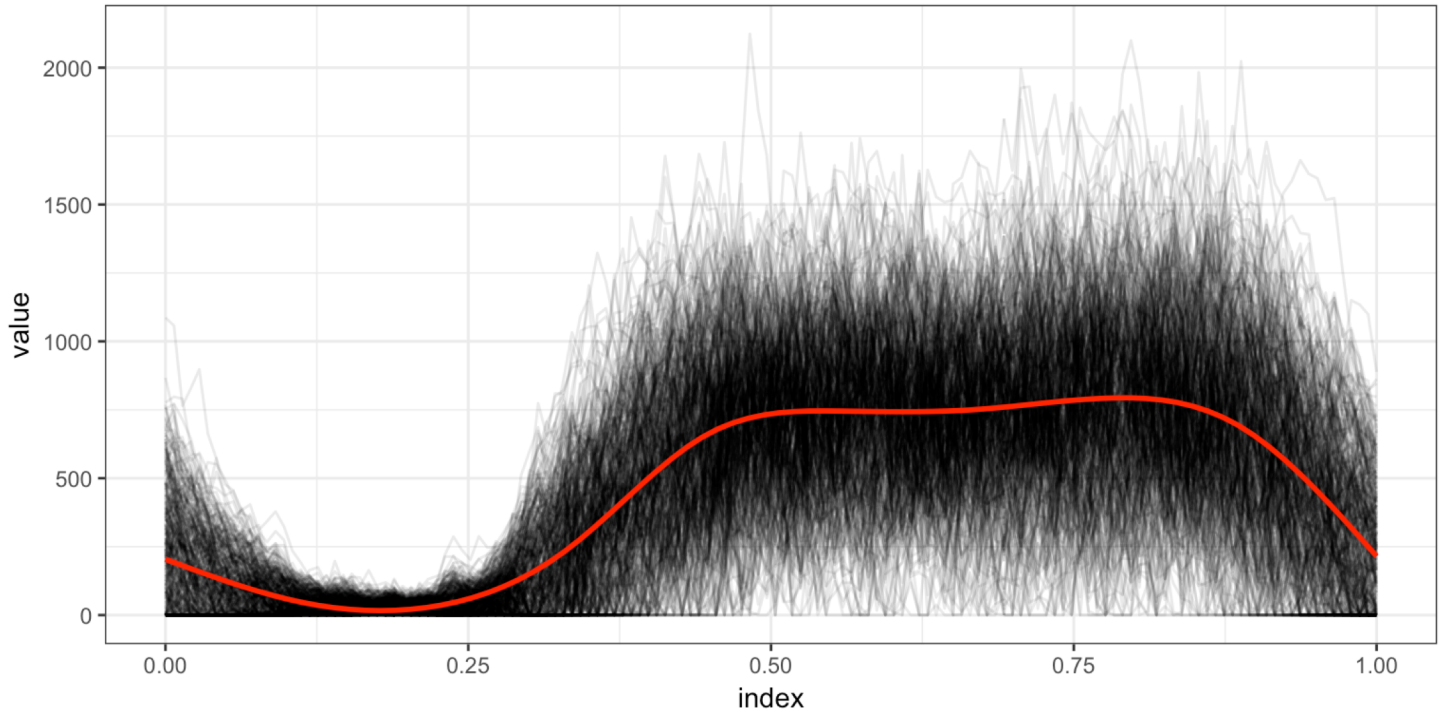
```
pw_mean = as_refundObj(accel) %>%
    group_by(index) %>%
    summarize(pw_mean = mean(value))

as_refundObj(accel) %>%
    ggplot(aes(x = index, y = value, group = id)) + geom_path(alpha = .1) +
    geom_path(data = pw_mean, aes(y = pw_mean, group = NULL), color = "red", size = 1)
```



Code

```
as_refundObj(accel) %>%
    ggplot(aes(x = index, y = value, group = id)) + geom_path(alpha = .1) +
    geom_smooth(aes(group = NULL), color = "red", size = 1)
## `geom_smooth()` using method = 'gam'
```



Code

```
as_refundObj(accel) %>%
    left_join(dplyr::select(covariate_data, id, BMIZ)) %>%
    ggplot(aes(x = index, y = value, group = id, color = BMIZ)) + geom_path(alpha = .5) +
  scale_colour_gradientn(colours = c("red","yellow","green","lightblue","darkblue"),
                         values = c(1.0, 0.6, 0.55, 0.45, 0.4, 0))
## Joining, by = "id"
```