Handling Missing Data & Imbalanced Data

Handling Missing Data

1. Imputation: fill in missing data based on observations about the entire dataset
    1. Forward fill: fill in the last known value (tidyverse code: https://tidyr.tidyverse.org/reference/fill.html)
    2. Backward fill (lookahead) - fill in data backward in time/ do not use this method to predict future
    3. Moving average: using past values to predict missing future values.
2. Interpolation: fill in missing data by estimating it using neighboring data points
3. Deletion of affected time period: chose time periods that have no missing data

Imbalanced data

https://wellsr.com/python/upsampling-and-downsampling-imbalanced-data-in-python/

1. Upsampling (increasing the sample size)
2. Downsampling

Cohomology Codes:

circular coordinate (cc)

- compress a topological signal (select smaller set of vertices (call landmark points)
    - subset of the whole set (calculation is quicker); simpler model
        - random selection
        - process called maxmin (deterministic) furthest from the previous n points > forces each new point to come from the unoccupied space
            - disadvantage: prone the outliers first because it is biased towards outliers
- hsv_map (hue saturation)
- DL (selected number of point (row) * random number totals number of points)
    - where it's a 5*400 matrix, where the nth row is 0 because distance to itself
- Intrinsic and extrinsic dimension

- Circle in a plane (2D) extrinsic way of drawing a circle is 2d
- Intrinsic dimension of a circle is one dimension - radian