Signal Processing

Chaya Hiruncharoenvate

tinyurl.com/
ipstSignalProcessing

About Me

Current:
Senior Officer
Data Modeling and Analytics Team
SEC Thailand (สำนักงาน กลต)

Previous:
Faculty of Informatics, Mahasarakham University
PhD in CS, Georgia Institute of Technology
MS in IT + BS in CS, Carnegie Mellon University

BODYJAM, BODYSTEP Instructor
Mo-Mo Paradise & Sushiro Loyalist
Cat Kidnapper

tinyurl.com/
ipstSignalProcessing

Agenda

- What are signals?
- Signal Data Visualization
- Preprocessing & Filtering Techniques
- Lab (x2)

What are signals?

Signals

- Information that varies over time/space
- Examples
  - Audio wave
  - Temperature readings
  - Financial market data
  - Other?

Time-series Data

- Subset of signals
- Measurements recorded at successive points in time
- Analog (continuous) or Digital (discrete)

Continuous VS Discrete

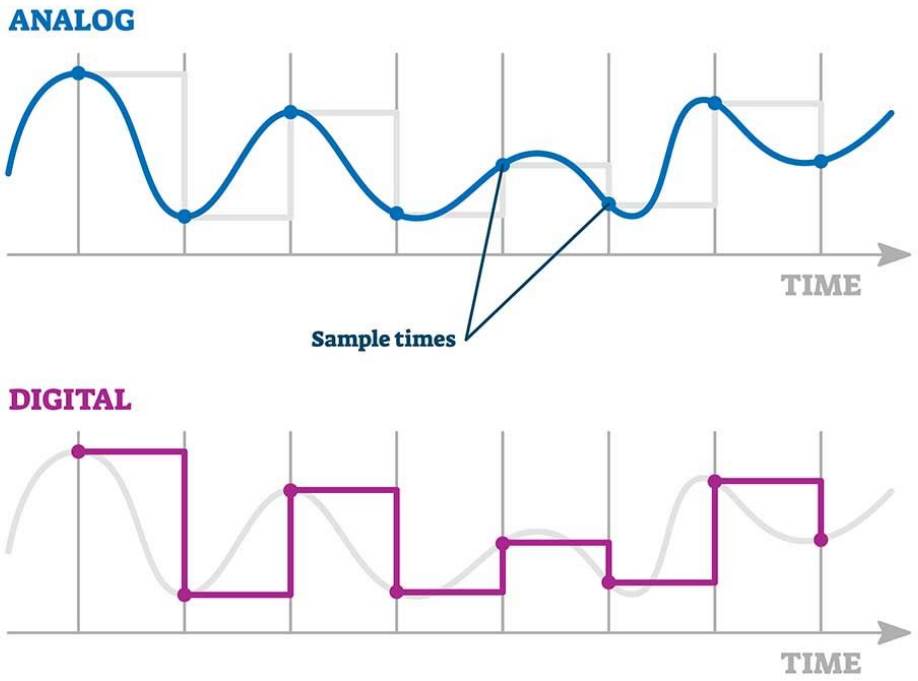
Continuous

- Measured and recorded over a continuous range
- Analog signals
  - Sound waves
  - Temperature measurements (analog thermometers)

Discrete

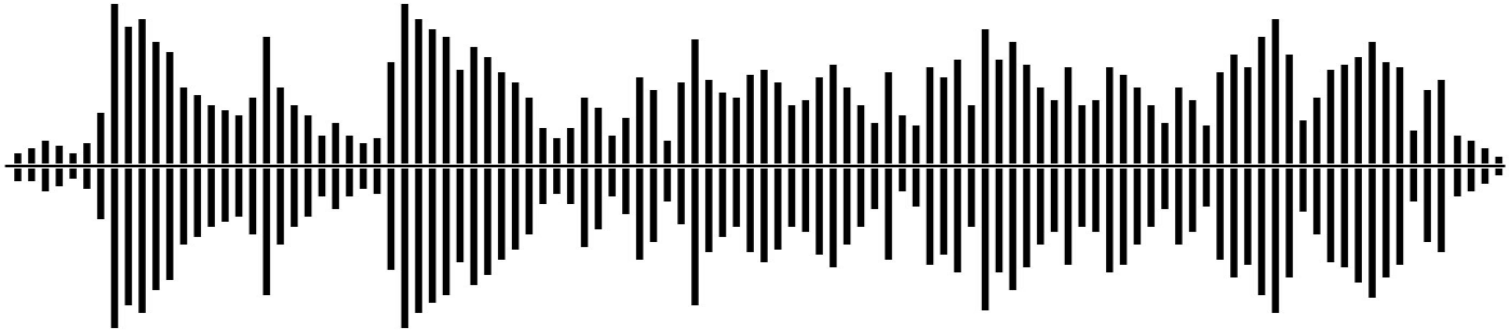
- Recorded at specific, distinct points
- Digital signals
  - IoT measurements
  - Financial market data
- More common

ANALOG VS DIGITAL SIGNAL



Why Signal Processing in Data Science?

- Raw signal data is difficult to interpret
  - Raw sound wave. How can we make sense of it?



Why Signal Processing in Data Science?

- Raw signal data is difficult to interpret
- Extract valuable insights from raw data
  - May be hard to see at first glance
- This information then be used to
  - Make decisions
  - Identify opportunities
  - Solve problems

Signal Processing as Data Preprocessing

- Real-world data is often noisy
- Remove unwanted disturbances, outliers using Signal Processing
- Resulting in a cleaner and more reliable dataset
- Clean data -> Accurate modeling, predictions
- Signal Processing is a core for
  - Time-series forecasting
  - Anomaly detection
  - Image and speech recognition

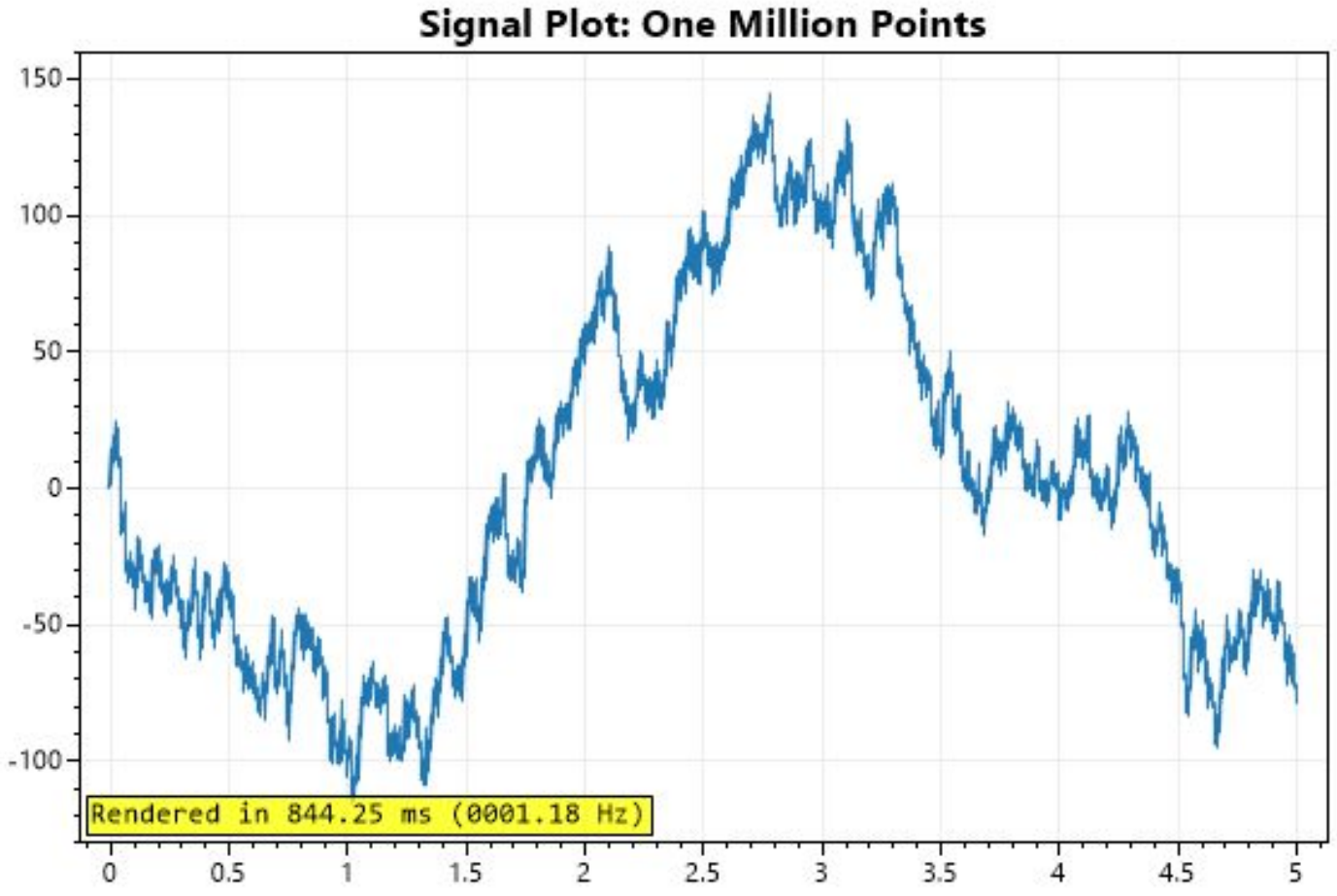
Signal Data Visualization

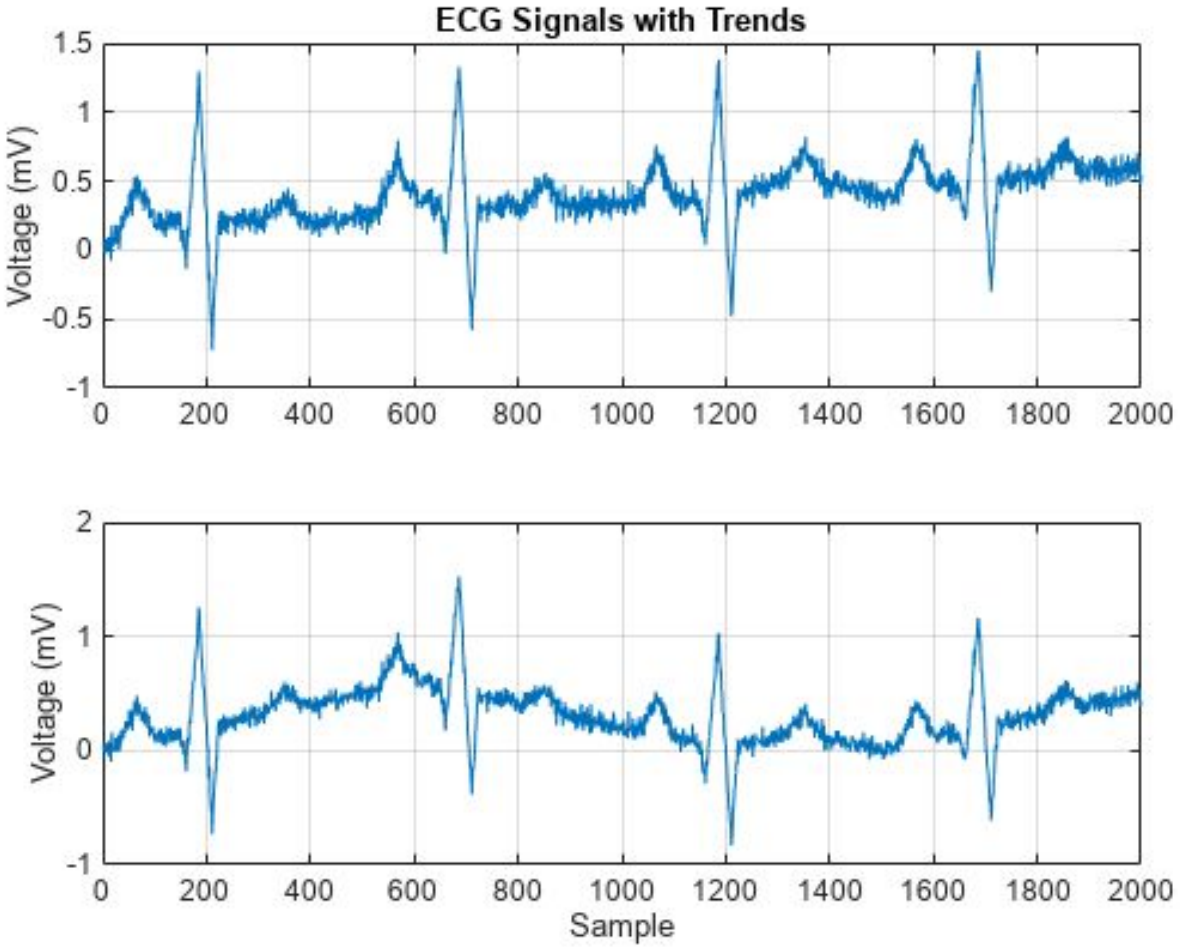
Importance of visualizations

- Signals are best understood through visuals
- A few plots are generally used
  - Line plots
  - Box plots

Line plot

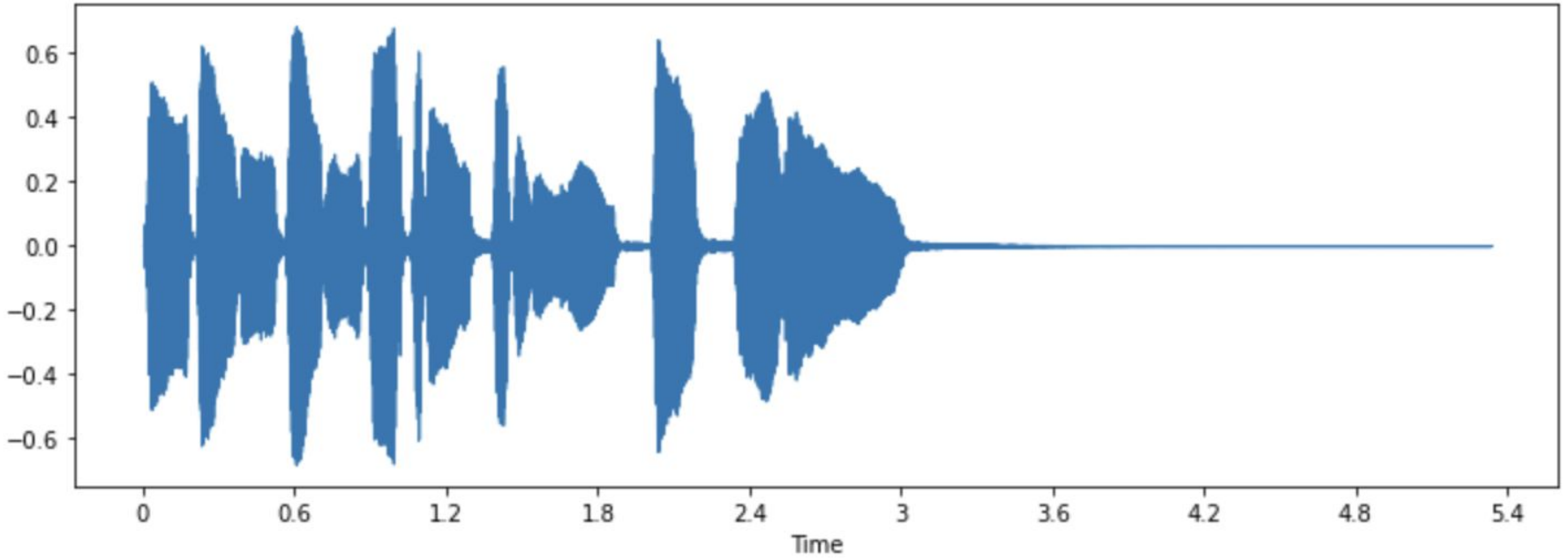
- X-axis: sequential metric (time/sample number/meters/lot number)
- Y-axis: signal value
- Provide understandings for trends and fluctuations


Signal Plot: One Million Points
150
100
50
0
-50
-100
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Rendered in 844.25 ms (0001.18 Hz)


ECG Signals with Trends
Voltage (mV)
1.5
1
0.5
0
-0.5
-1
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Voltage (mV)
2
1
0
-1
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Sample

Audio waveform

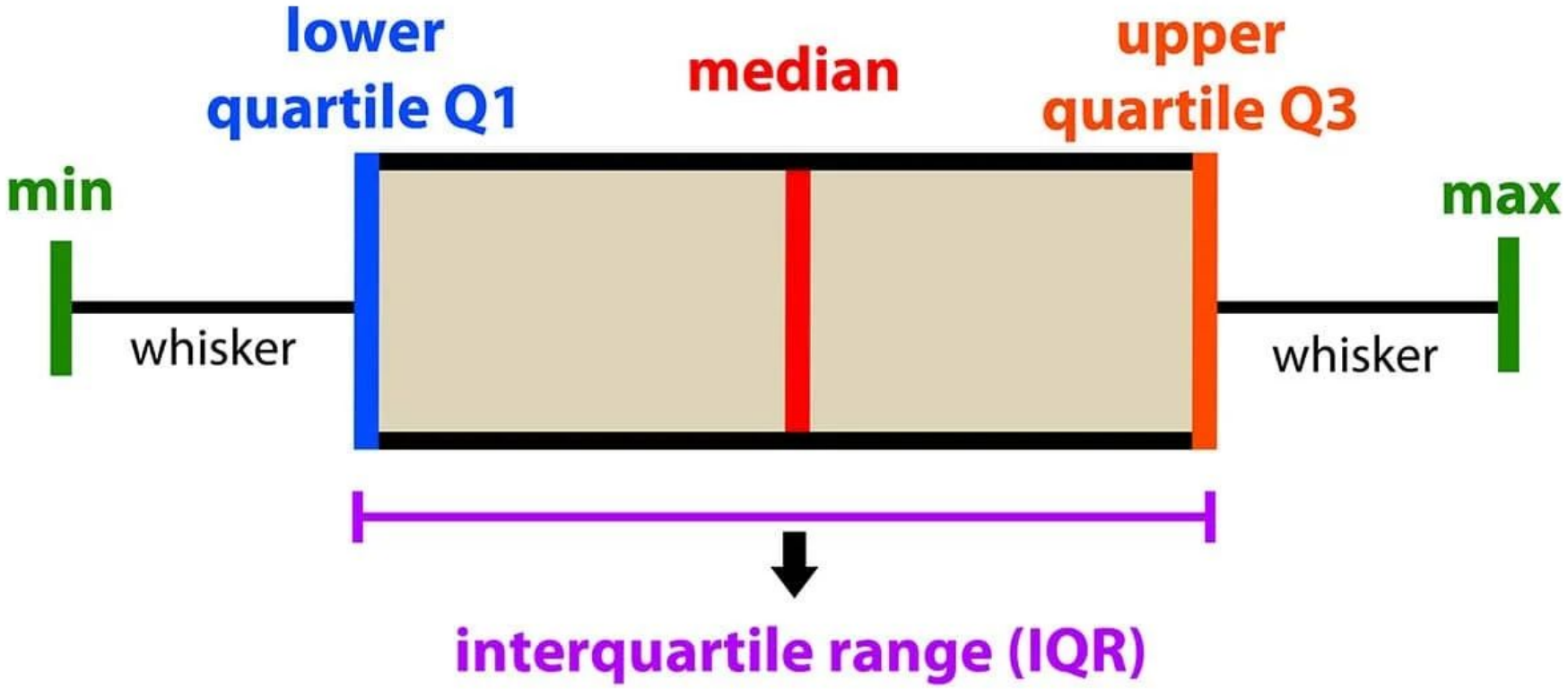
- Specific type of line plot
- Visualizing sound
- X-axis: Amplitude (loudness)
- Y-axis: time sample

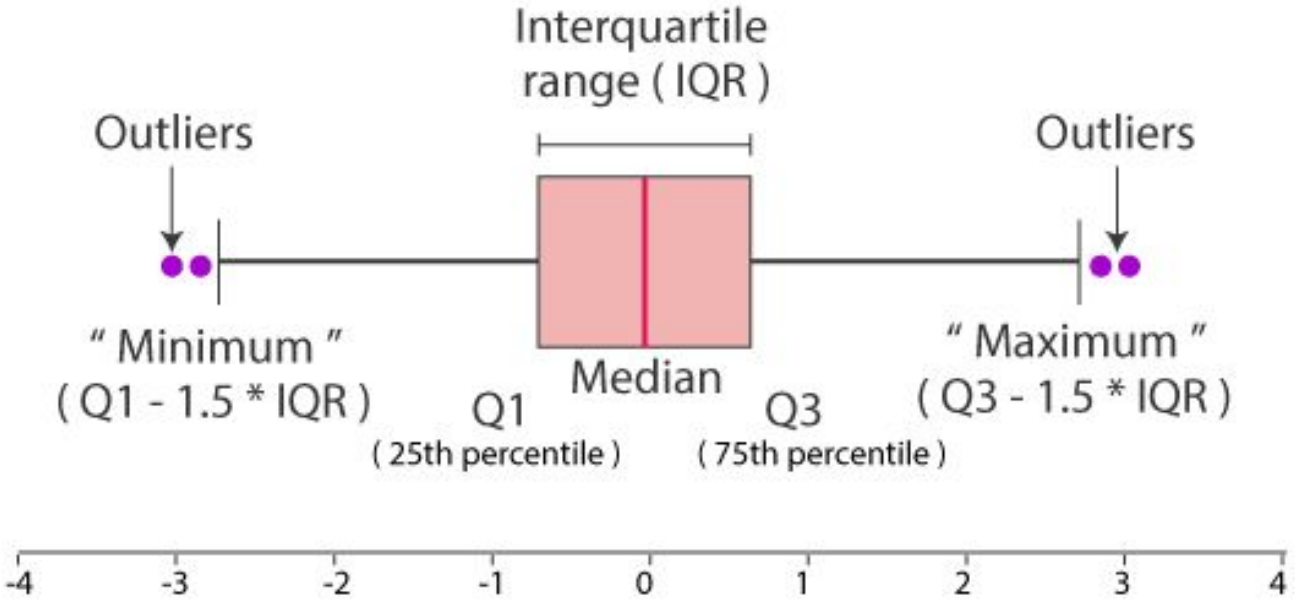

0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
5.4
Time

Box plot

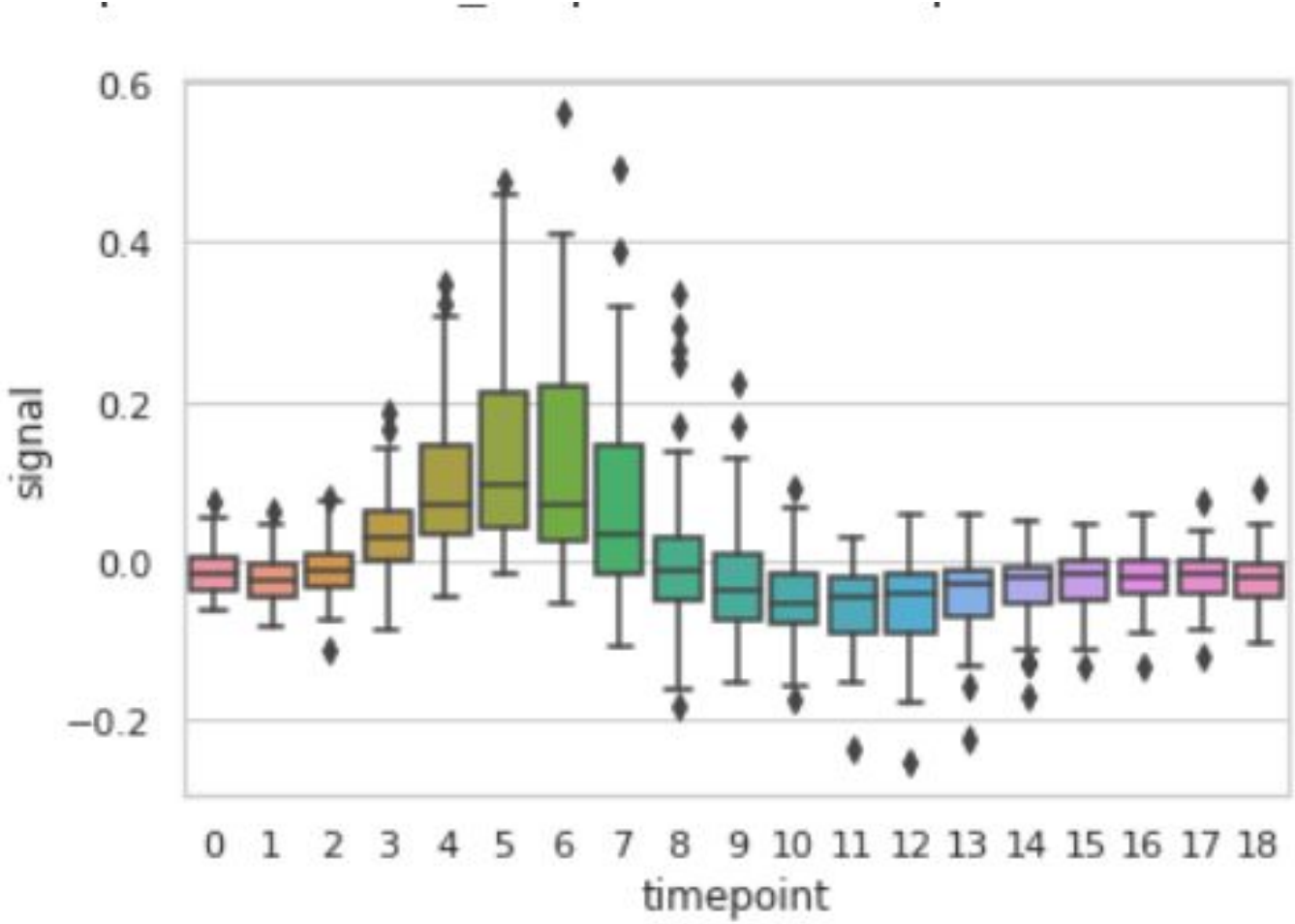
- Show several statistics for each time sample
  - “Min”
  - 1st quartile
  - Median
  - 3rd quartile
  - “Max”
  - Outliers

introduction to data analysis: Box Plot



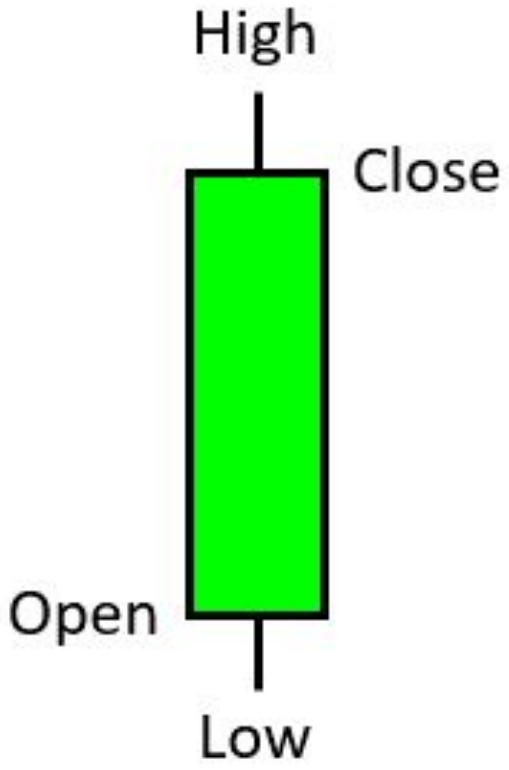


Different parts of boxplot

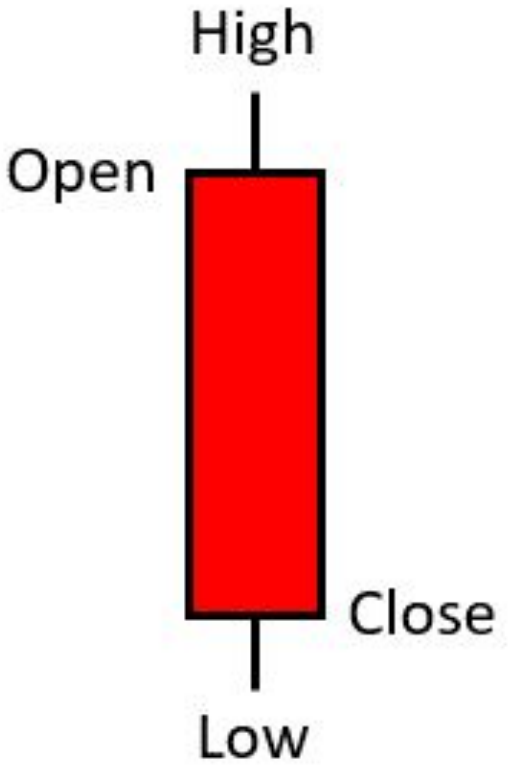

0.6
0.4
0.2
0.0
−0.2
signal
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
timepoint

Candlestick plot

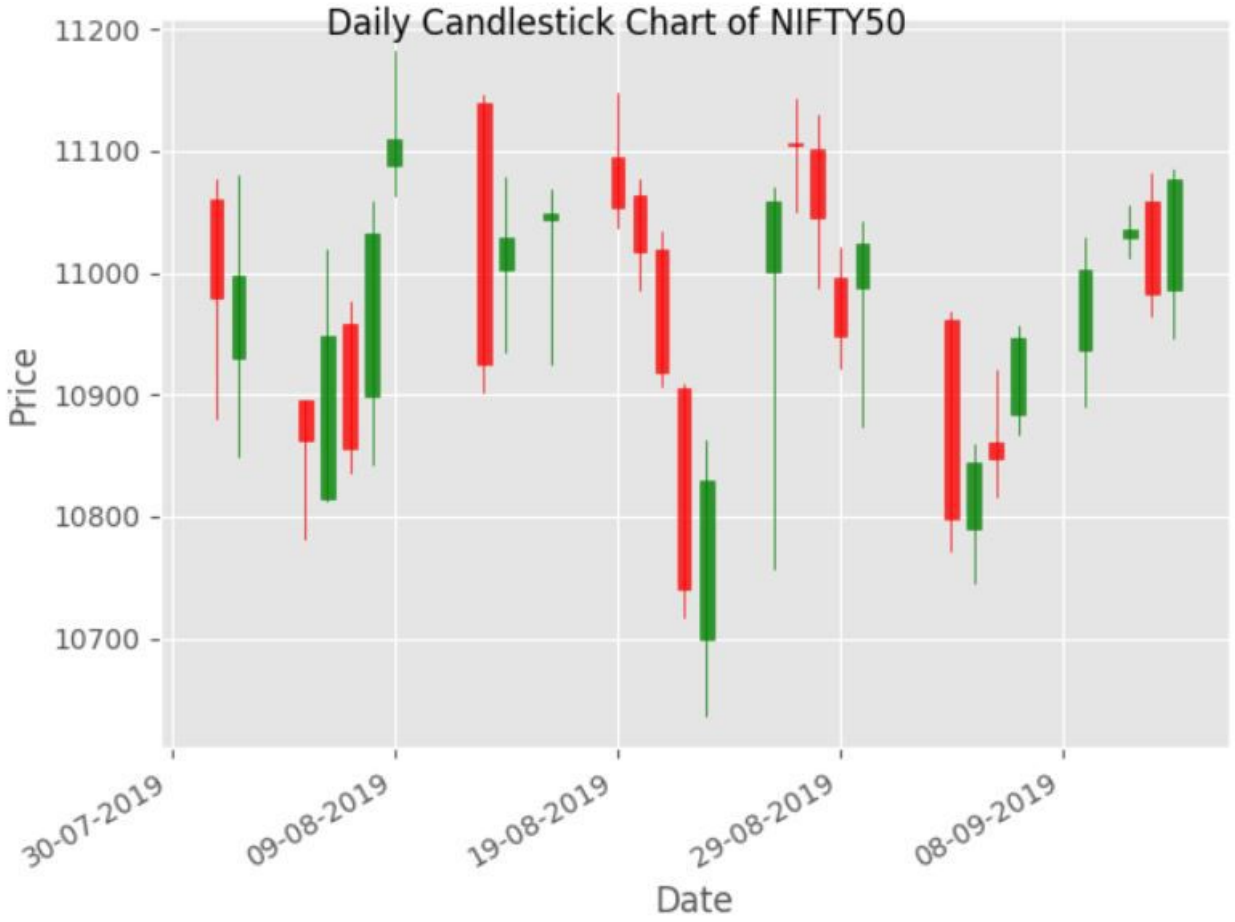
- Special type of box plot to visualize financial market data
- Each box shows
  - Opening price
  - Closing price
  - Highest price
  - Lowest price
  - Closing price > Opening price? (color)



Bullish candlestick



Bearish candlestick


Daily Candlestick Chart of NIFTY50
11200
11100
11000
10900
10800
10700
Price
30-07-2019
09-08-2019
19-08-2019
29-08-2019
08-09-2019
Date


British Pound / Japanese Yen, 1D,
142.000
140.000
138.000
136.000
134.000
132.000
130.000
128.000
126.000
Jun
Jul
Aug
Sep
Oct

Lab 1

Preprocessing & Filtering Techniques

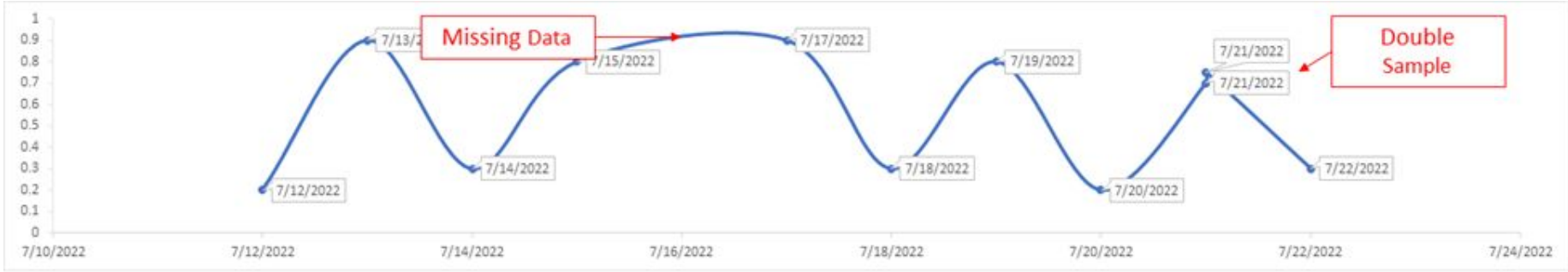
Preprocessing & Filtering Techniques

- Dealing with irregular or missing data
- Handling noise and outliers

Dealing with Irregular or Missing Data

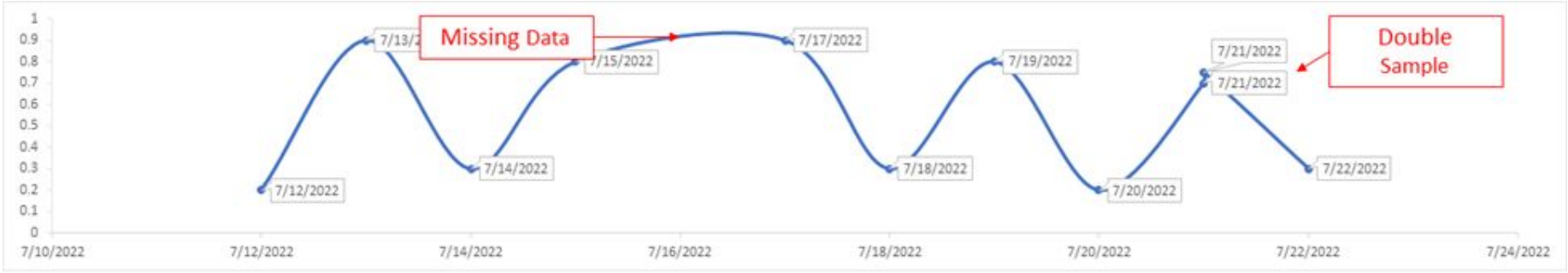
Dealing with Irregular or Missing Data

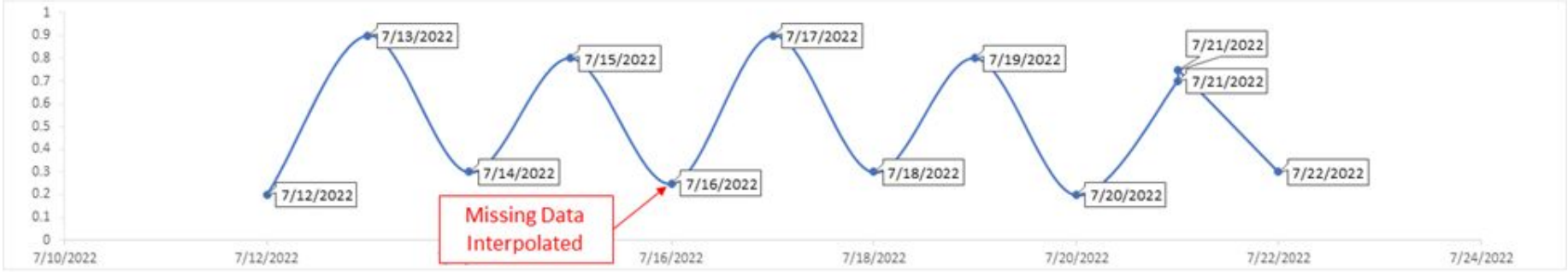
- What to do when sampling intervals are irregular?
- How do we handle missing data points?



Interpolation

- Technique for handling missing data
- “Estimate” the value of the missing data point
- Common techniques
  - Linear interpolation
  - Spline interpolation
  - Time-based interpolation

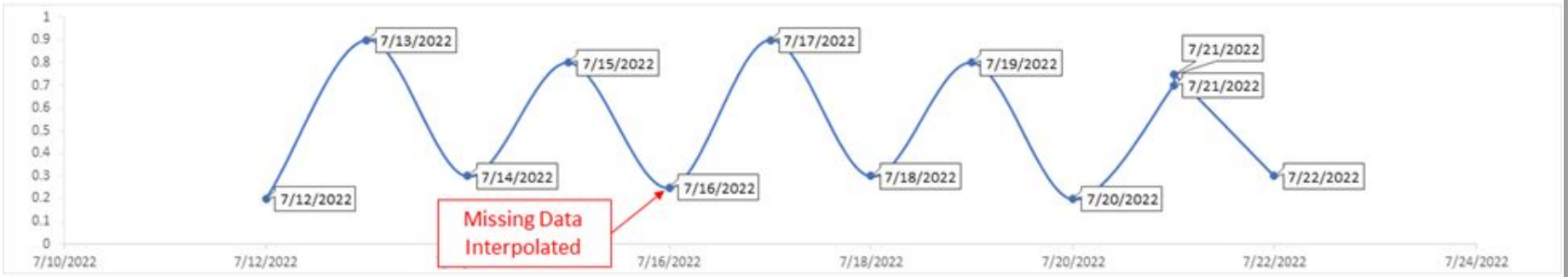

Missing Data
Double Sample
7/12/2022
7/13/2022
7/14/2022
7/15/2022
7/17/2022
7/18/2022
7/19/2022
7/20/2022
7/21/2022
7/21/2022
7/22/2022
7/10/2022
7/12/2022
7/14/2022
7/16/2022
7/18/2022
7/20/2022
7/22/2022
7/24/2022

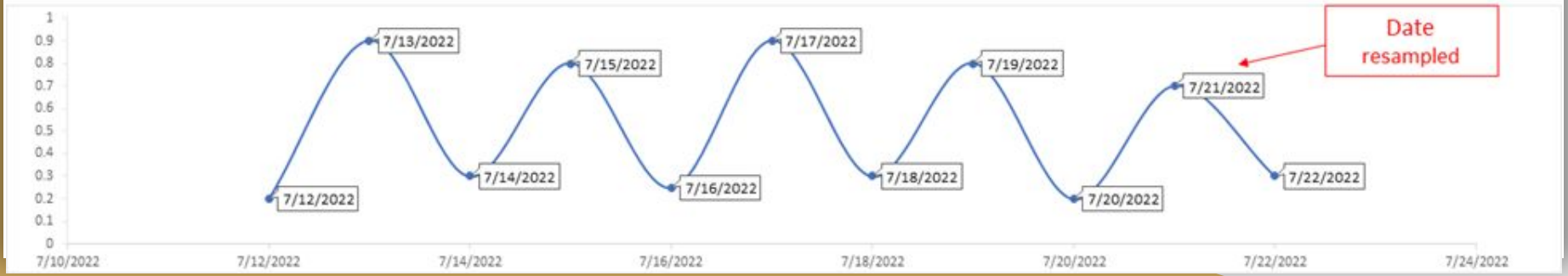

7/12/2022
7/13/2022
7/14/2022
7/15/2022
7/16/2022
7/17/2022
7/18/2022
7/19/2022
7/20/2022
7/21/2022
7/21/2022
7/22/2022
Missing Data Interpolated
7/10/2022
7/12/2022
7/16/2022
7/18/2022
7/20/2022
7/22/2022
7/24/2022

Resampling

- Technique used to standardize the intervals of the data
- Can be
  - Upsampling – increase the frequency of data points
  - Downsampling – decrease the frequency of data points


Missing Data
Double Sample
7/12/2022
7/13/2022
7/14/2022
7/15/2022
7/17/2022
7/18/2022
7/19/2022
7/20/2022
7/21/2022
7/21/2022
7/22/2022
7/10/2022
7/12/2022
7/14/2022
7/16/2022
7/18/2022
7/20/2022
7/22/2022
7/24/2022

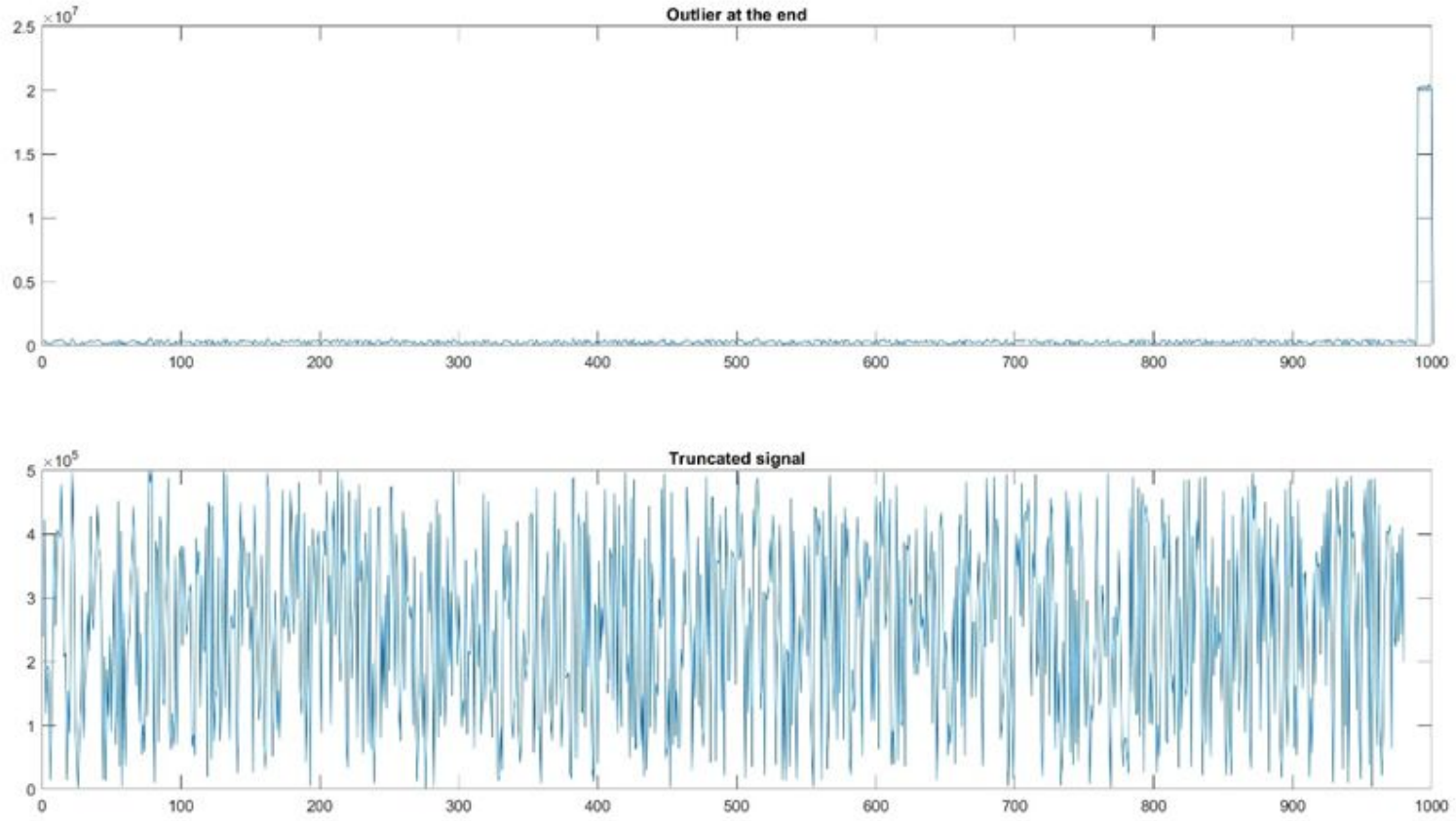

7/12/2022
7/13/2022
7/14/2022
7/15/2022
7/16/2022
7/17/2022
7/18/2022
7/19/2022
7/20/2022
7/21/2022
7/21/2022
7/22/2022
Missing Data Interpolated
7/10/2022
7/12/2022
7/16/2022
7/18/2022
7/20/2022
7/22/2022
7/24/2022

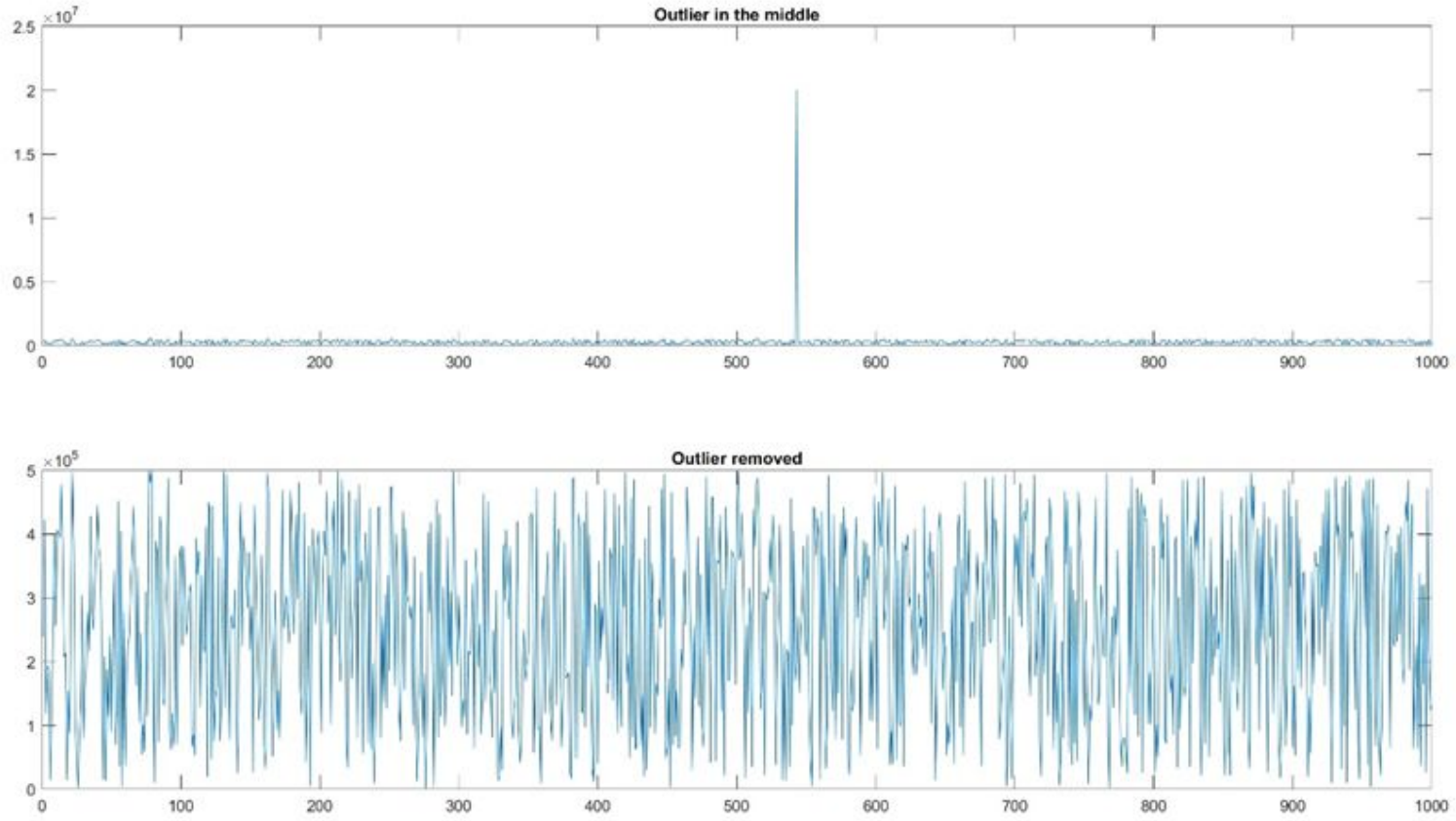

Date resampled
7/12/2022
7/13/2022
7/14/2022
7/15/2022
7/16/2022
7/17/2022
7/18/2022
7/19/2022
7/20/2022
7/21/2022
7/22/2022
7/10/2022
7/12/2022
7/14/2022
7/16/2022
7/18/2022
7/20/2022
7/22/2022
7/24/2022

Handling Noise and Outliers

Handling Noise and Outliers

- Noise = random fluctuations in the data
- Outliers = points that significantly deviate from the overall patterns
- Handle by
  - Data truncation – capping extreme values at a certain threshold
  - Treat outliers separately
  - Smoothing


Outlier at the end
×10^7
2.5
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
1000
Truncated signal
×10^5
5
4
3
2
1
0
0
100
200
300
400
500
600
700
800
900
1000

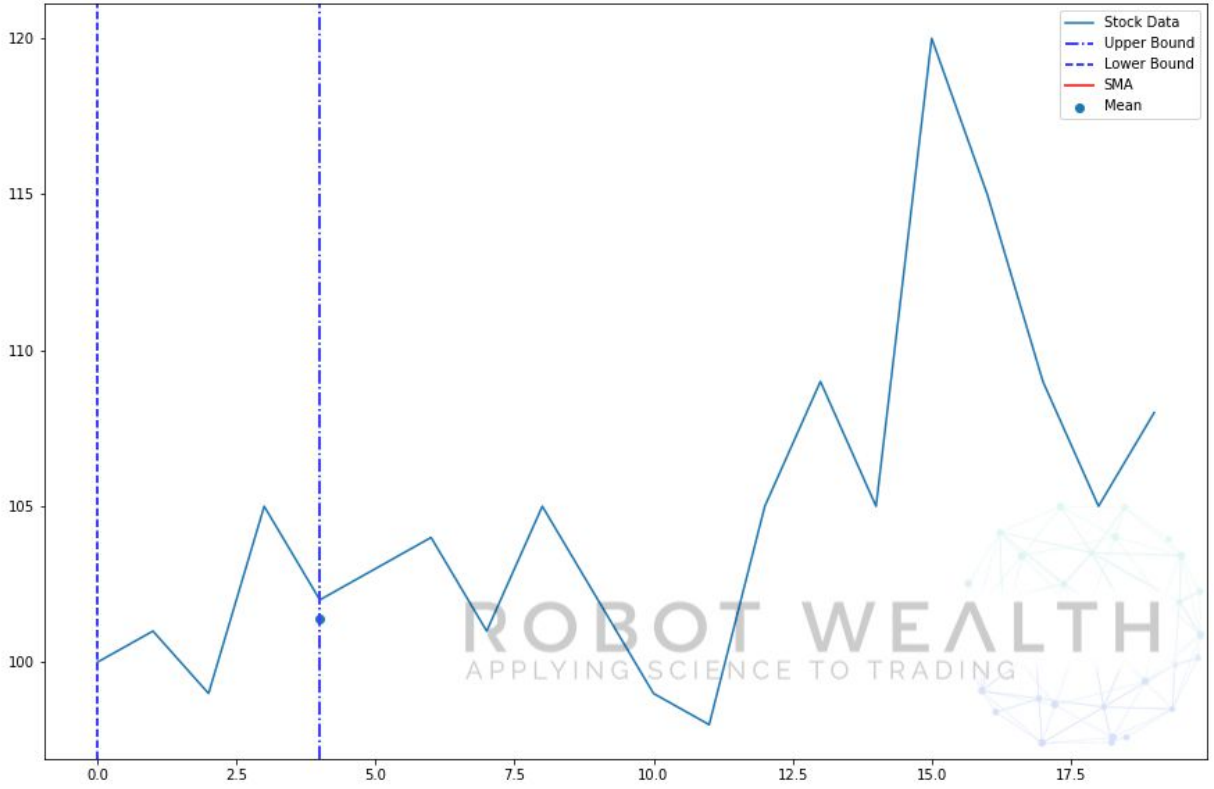

Outlier in the middle
×10^7
2.5
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
1000
Outlier removed
×10^5
5
4
3
2
1
0
0
100
200
300
400
500
600
700
800
900
1000

Data Smoothing

- Reduce the impact of random fluctuations
- Help reveal long-term trends while suppressing short term noise
- Methods
  - Rolling windows and moving averages
  - Low-pass, High-pass, Band-pass filters

Rolling window

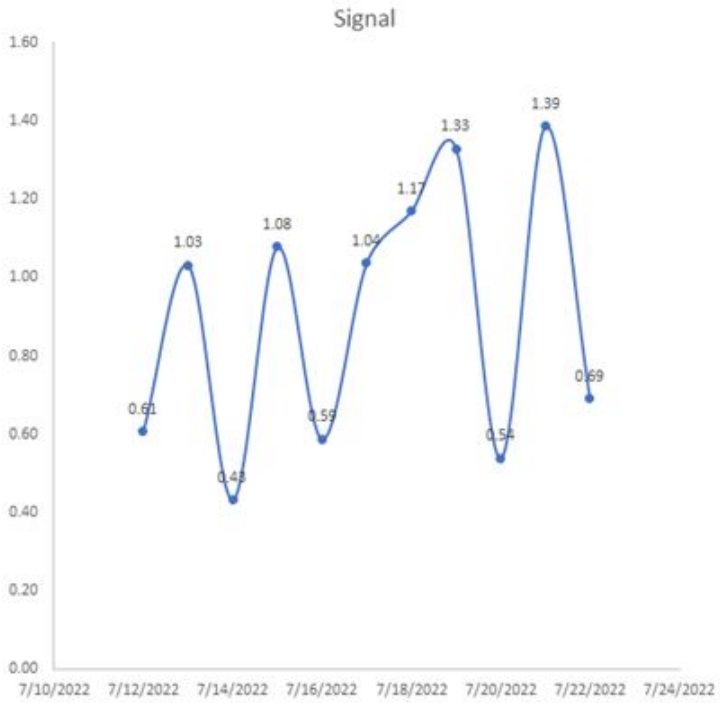
- Compute a specific statistics (mean, sd, etc) over a sliding window of data points

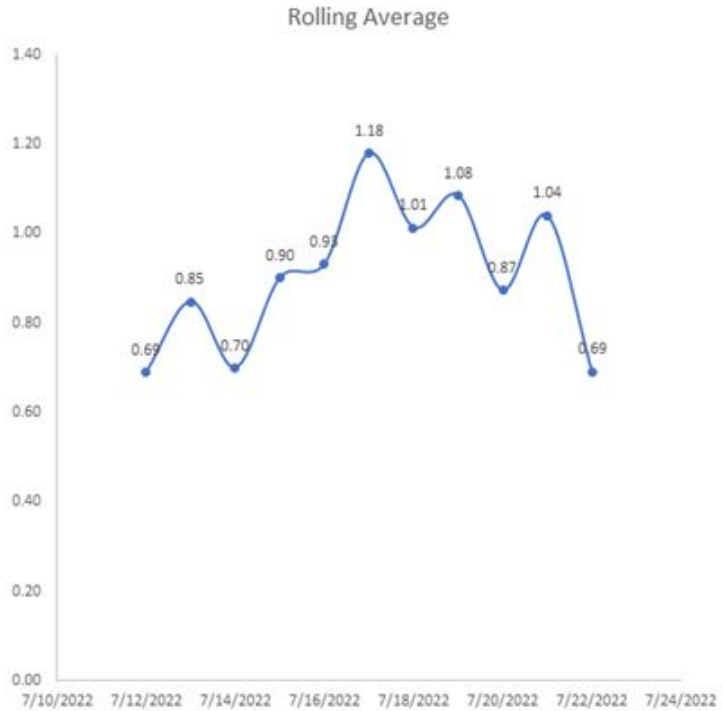

Stock Data
Upper Bound
Lower Bound
SMA
Mean
120
115
110
105
100
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
ROBOT WEALTH
APPLYING SCIENCE TO TRADING

Moving average

- An application of rolling windows – using mean as the statistic to compute

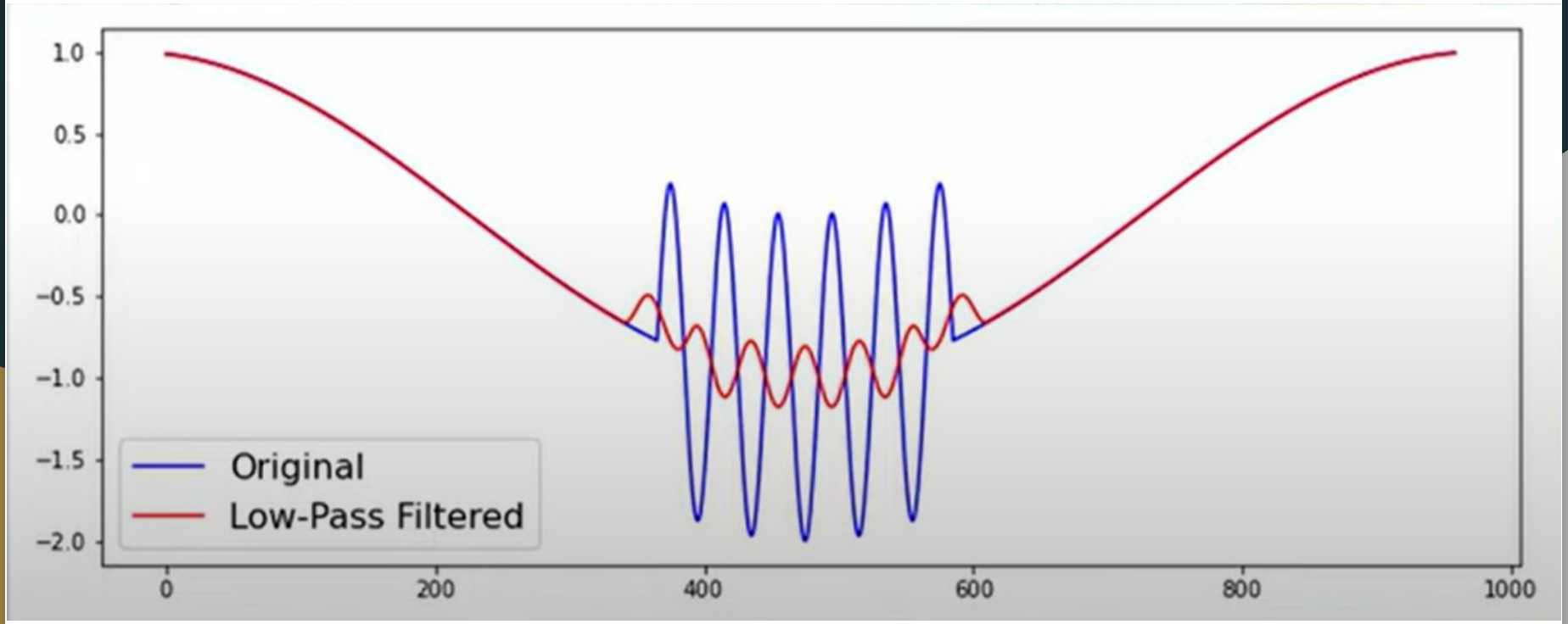
Window size = 3





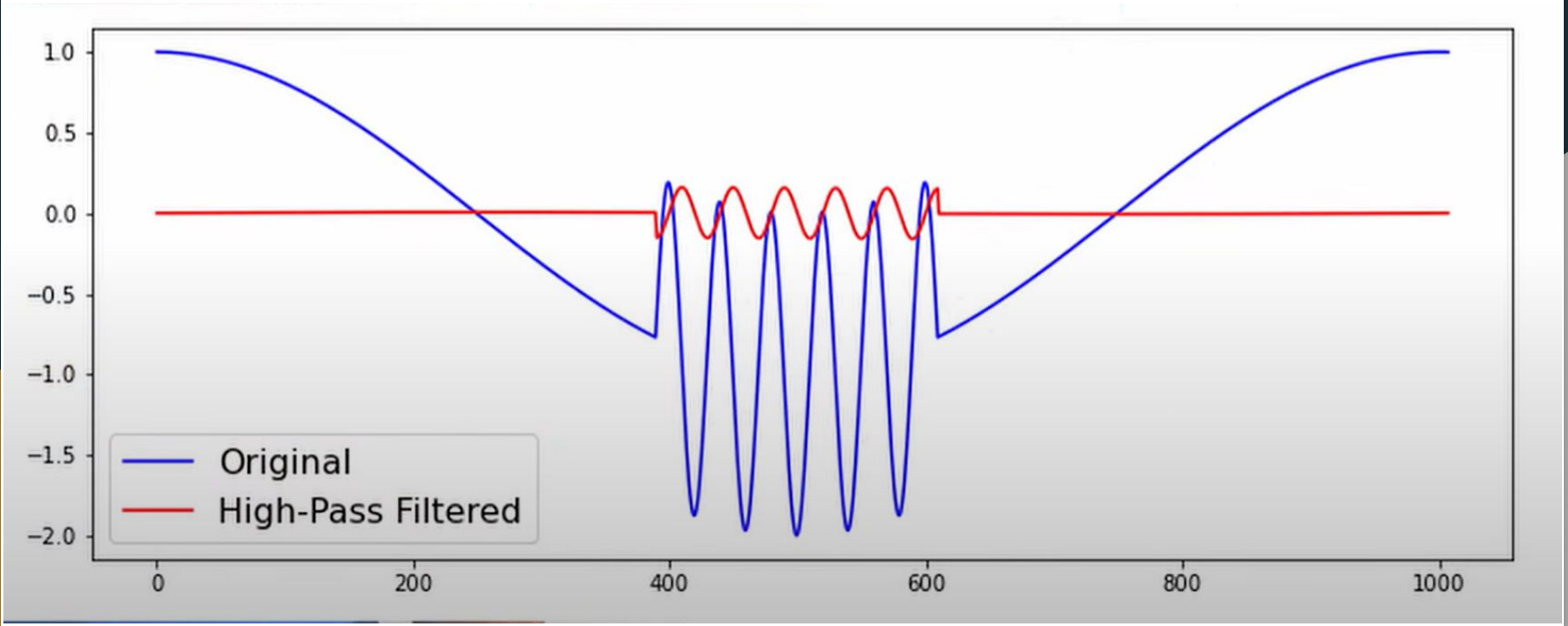
Low-pass filter

- Filter that allow low-frequency signals to pass while “attenuating” higher frequency components
- Useful for removing nose and retain slow-changing trends
- Moving average is a type of low-pass filter


1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
200
400
600
800
1000
Original
Low-Pass Filtered

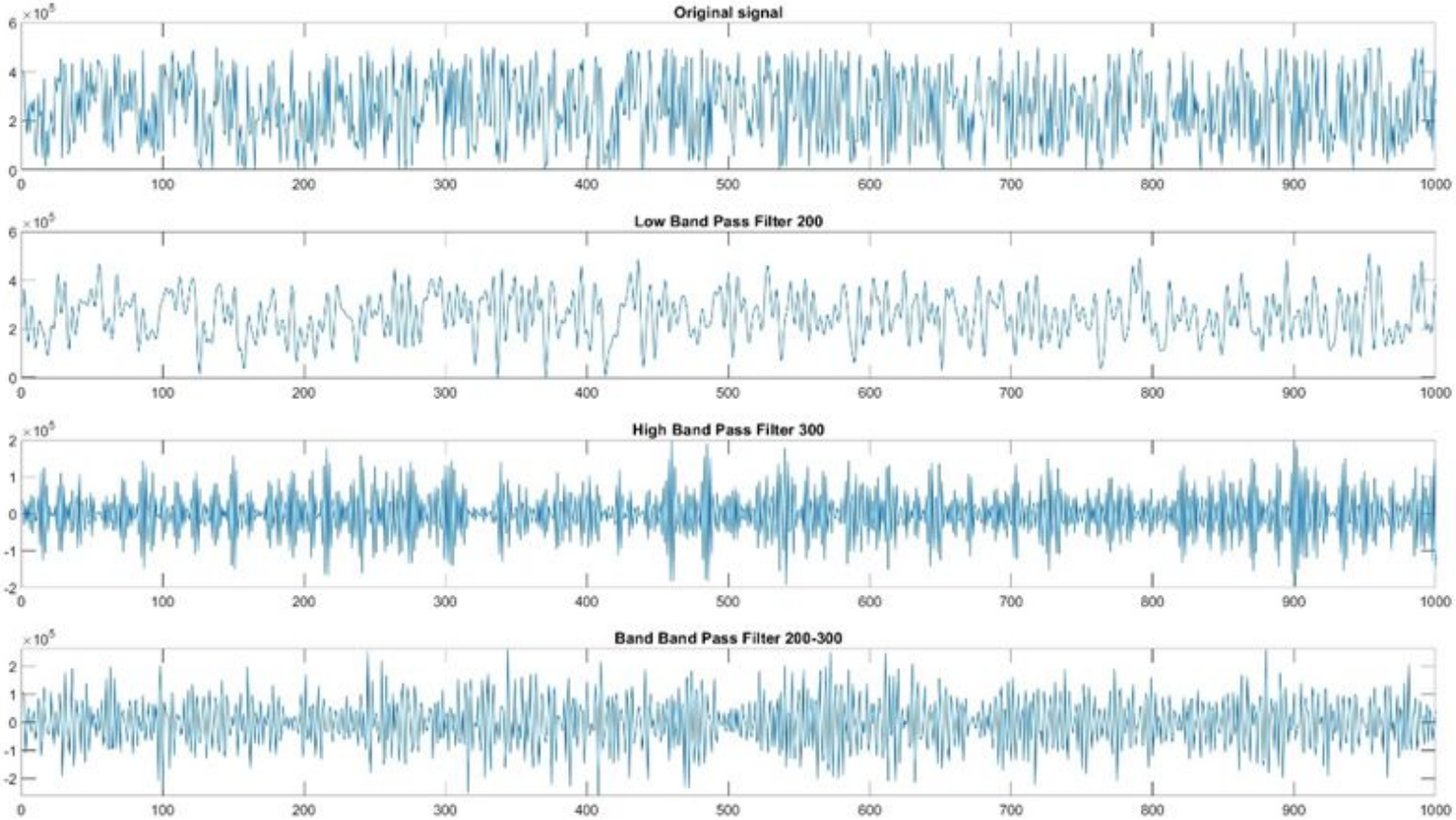
High-pass filter

- Permit higher-frequency signal to pass
- Filter out low-frequency components
- Highlight short-duration events or sudden changes in the time series


1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
200
400
600
800
1000
Original
High-Pass Filtered

Band-pass filter

- Allow signals with specific frequency band to pass while blocking others
- Useful for application where specific frequency ranges contain relevant information
  - Audio processing – extract frequency corresponding to human speech


Original signal
Low Band Pass Filter 200
High Band Pass Filter 300
Band Band Pass Filter 200-300

Lab 2

Questions?