Higher-dimensional Data Analysis Using Autocorrelation Wavelets via Julia

Christina Chang
chlchang@ucdavis.edu

Shozen Dan
shodan@ucdavis.edu

University of California, Davis
Faculty Mentor: Naoki Saito

June 21, 2020

Introduction

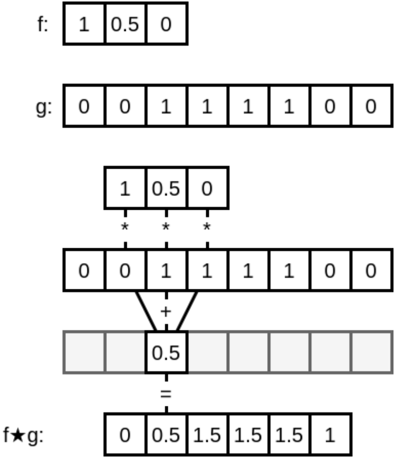
Task:

- Extend the autocorrelation wavelet transform to higher dimensions in the Julia programming language
- Perform denoising experiments on images and analyze multiple time series using autocorrelation wavelets

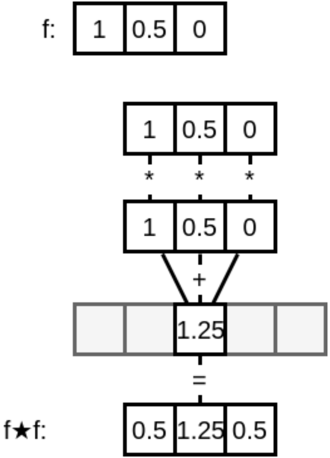
Signals and Filters

- Signal: A function carrying information, often with respect to time
- Filter: A function to extract certain information or feature from a signal (usually represented as a vector). Applied on a signal using sliding dot product
- Signal Operations
  - Cross Correlation
    - Measures the similarity between a filter and an input signal (Note: a filter could be yet another input signal)
  - Autocorrelation
    - Correlation of a signal with a delayed copy of itself
    - Autocorrelation function is symmetric

Discrete Correlation

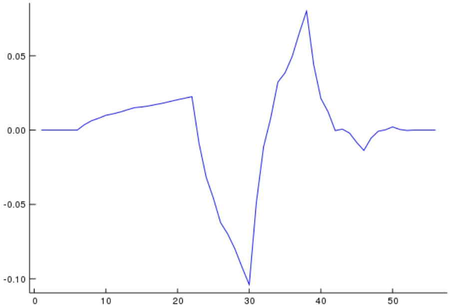


Discrete Autocorrelation

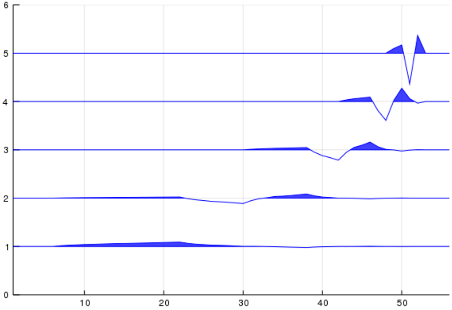


Wavelets

- Wavelet: A function that resembles a single oscillation of a wave
- Wavelets have both frequency and time domain info



Daubechies Wavelet



Daubechies Wavelets of various scales

Wavelet Transforms

- Represent a signal as a linear combination of wavelet basis functions

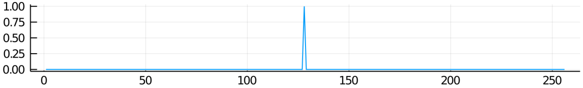
$$f(t) = \sum_{i=1}^{n} a_i \phi_i(t)$$

  where a_i are expansion coefficients and $\phi(t)$ are the basis functions
- Wavelet transform deconstructs the signal using the same wavelet at different scales

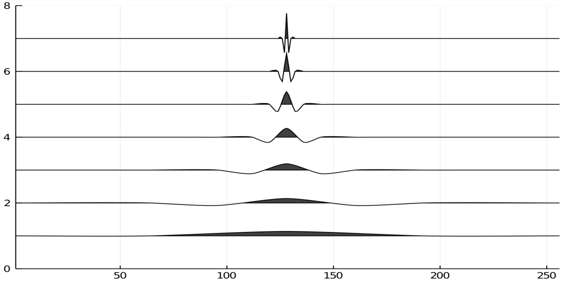
1D Autocorrelation Wavelet Transform

- Redundant wavelet tranform
- Autocorrelation properties: shift-invariant **and** symmetric

Original Signal



Autocorrelation Functions of Wavelets



2D Autocorrelation

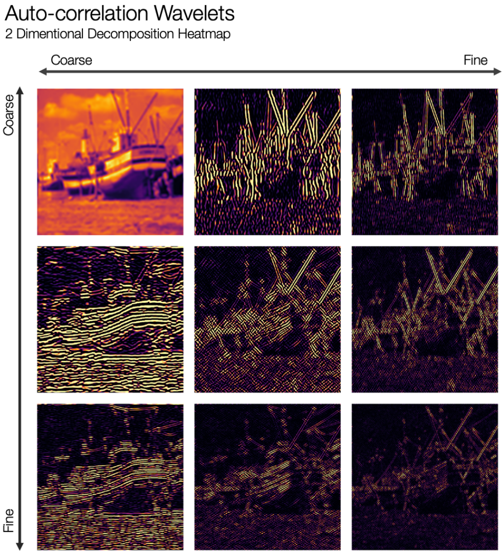


Image De-noising

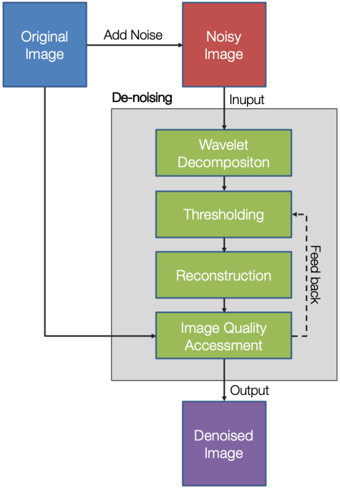
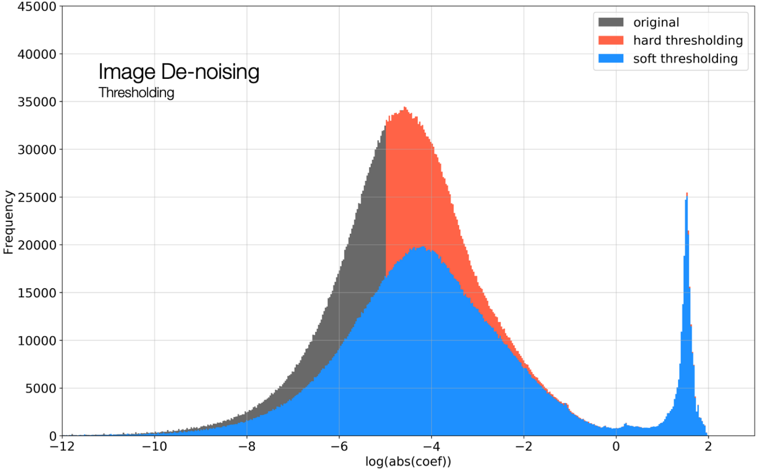


Image De-noising
Flowchart

Image De-noising



Note: This thresholding value is not used in any subsequent plots. It is for illustration purposes only.

De-noising: Peak Signal to Noise Ratio (PSNR)

Peak Signal to Noise Ratio (PSNR)

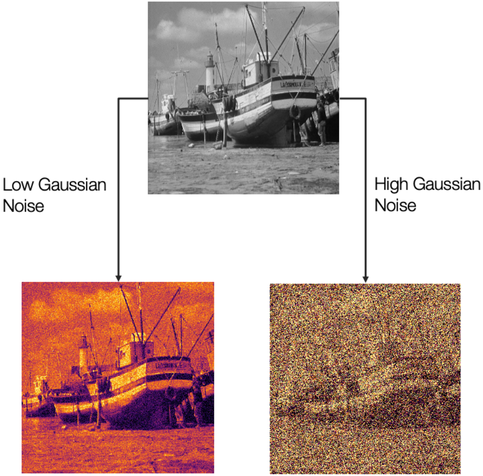
- Ratio between maximum power of a signal and maximum power of corrupting noise
- Quantifies image quality degradation

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{1}$$

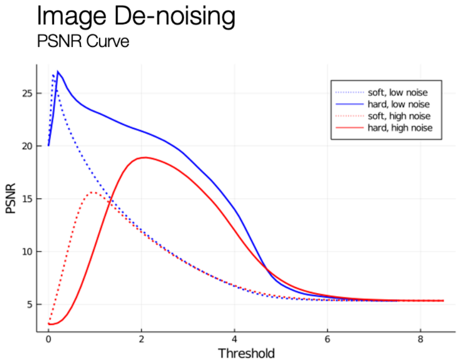
$$PSNR = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{2}$$

- I, K, MAX_I: original image, denoised image and the maximum possible pixel value of I
- The smaller the MSE, the larger the value

De-noising: Peak Signal to Noise Ratio (PSNR)

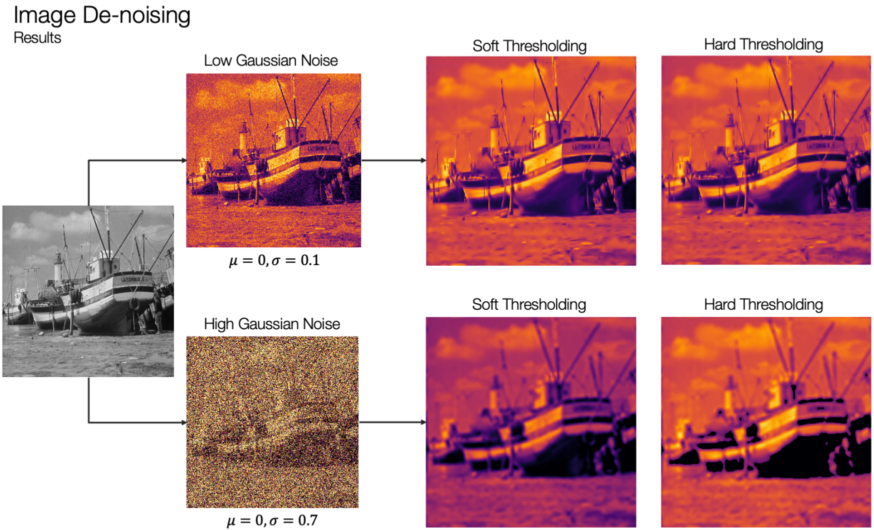


$\mu = 0, \sigma = 0.1$ $\mu = 0, \sigma = 0.7$



Use the threshold corresponding to the largest PSNR value

De-noising: Peak Signal to Noise Ratio (PSNR)

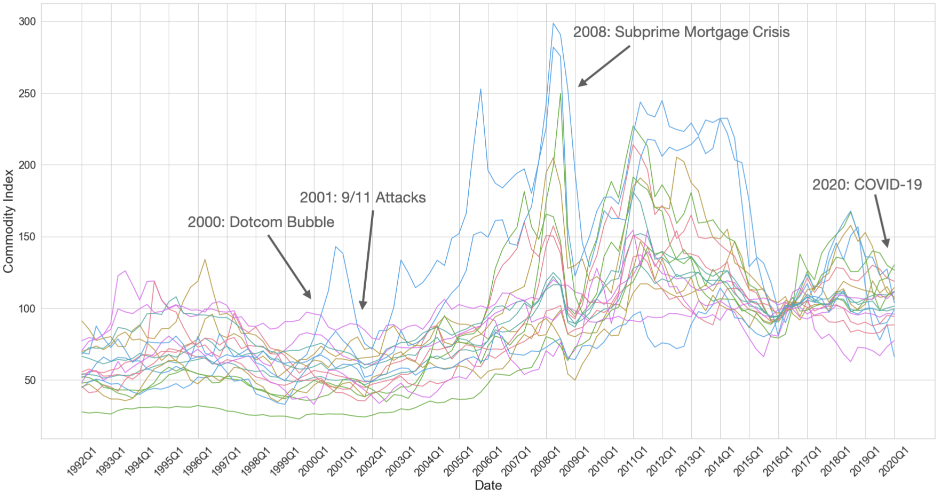


Data Analysis 1: Commodity Prices

- Commodity index: weighted average of the prices of a selected basket of goods relative to their prices in some base year
- Source: International Monetary Fund
  - Includes commodity indices such as Industrial Materials Index, Food Index, Energy Index, etc.
  - Quarterly data which spans from 1992Q1 to 2020Q1
- The indices are based in 2016
  - Example

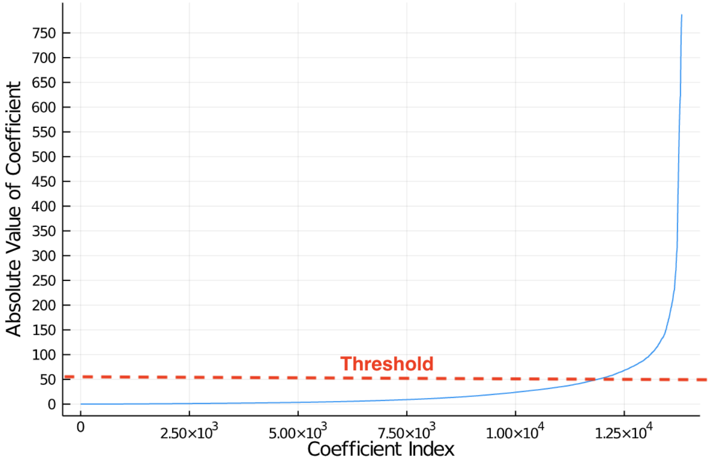
$$\text{Commodity Index for 2020} = \frac{\text{Price of Basket in 2020}}{\text{Price of Basket in 2016}} * 100$$

Data Analysis 1: Commodity Prices



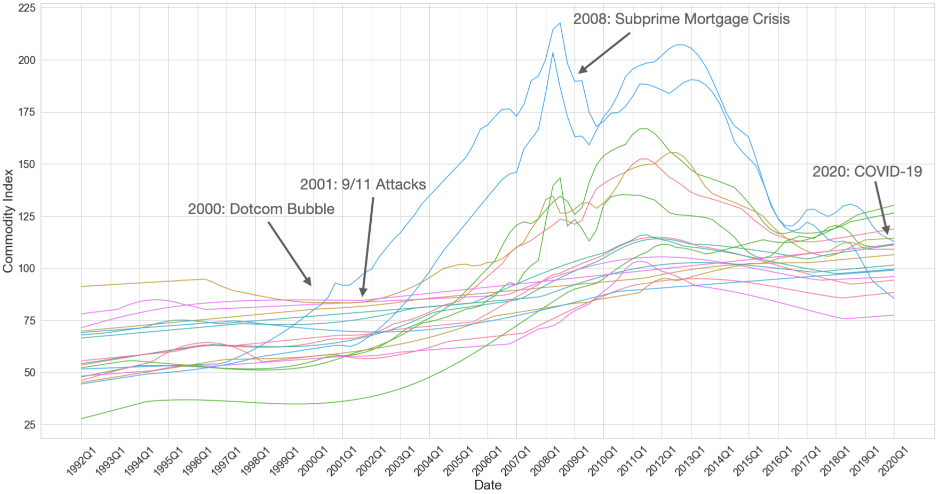
Data Analysis 1: Commodity Prices

- Goal: denoise the data by thresholding the coefficients obtained from the autocorrelation wavelet transform



Data Analysis 1: Commodity Prices

After thresholding:

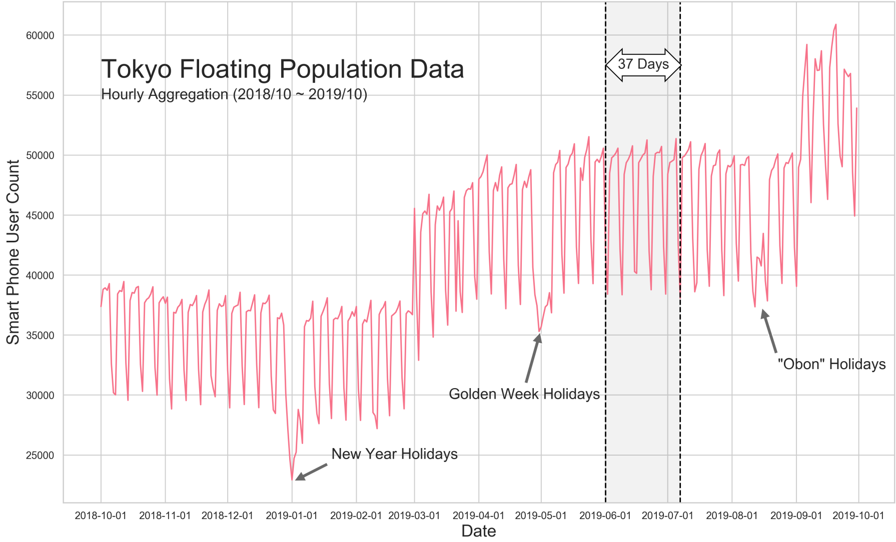


Data Analysis 2: Floating Population

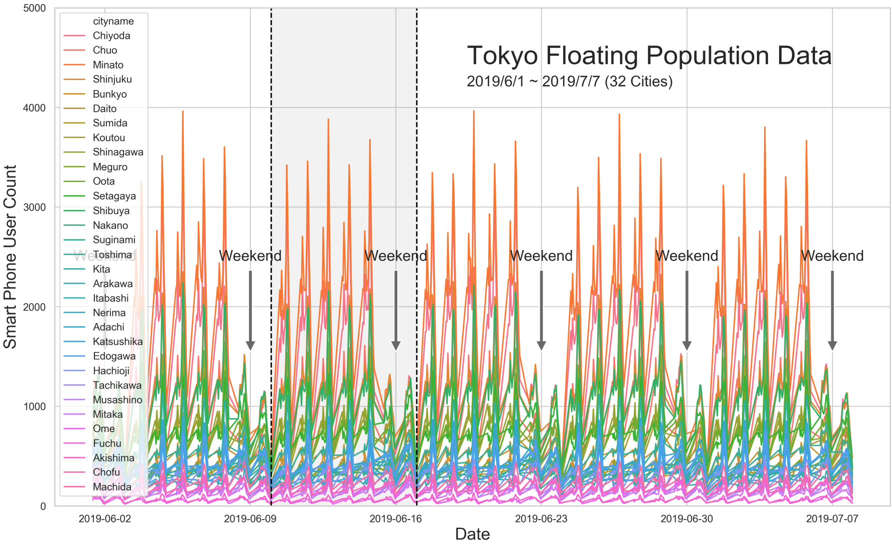
Floating Population Data

- Source: Tokyo Public Transportation Open Data Challenge
  - Agoop: Subsidiary of SoftBank(Largest network provider in Japan)
- Smart Phone Log Data (2018/10 to 2019/10)
  - Date
  - Location
  - Gender
  - Various Settings e.g. Currency, Language, Country
- Goal: Use 2D auto-correlation wavelet decomposition to find trends in the flow of people.

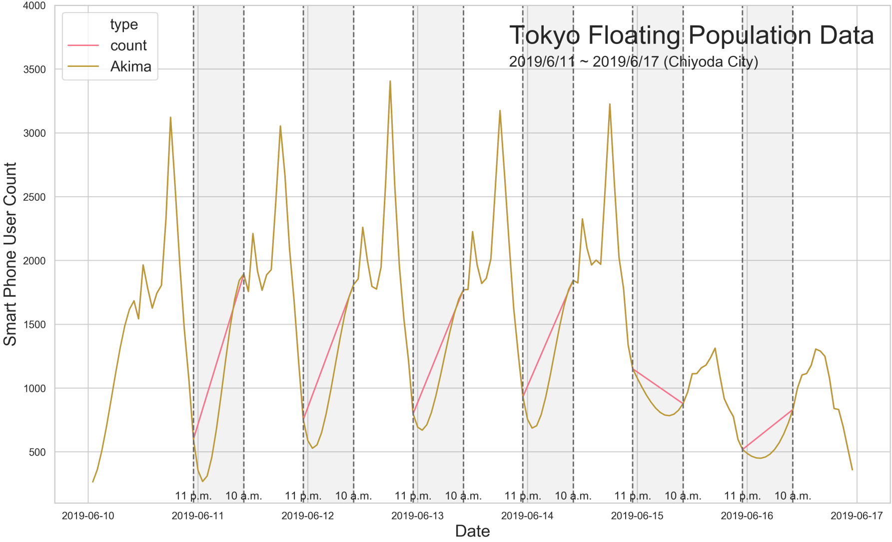
Data Analysis 2: Floating Population



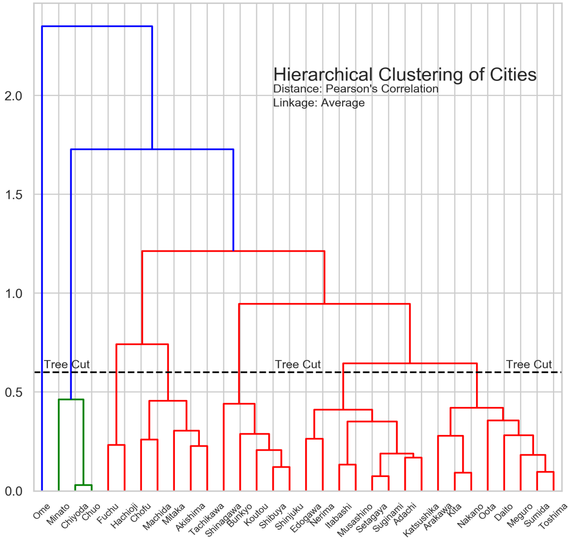
Data Analysis 2: Floating Population



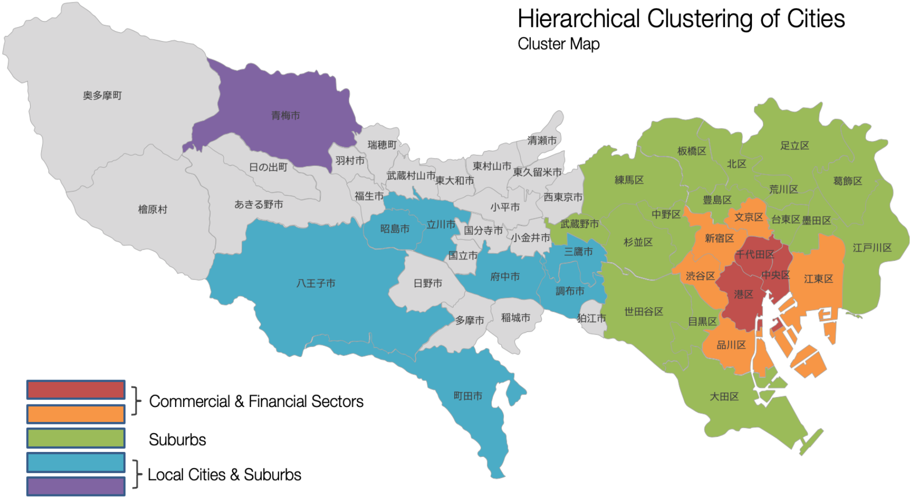
Data Analysis 2: Floating Population



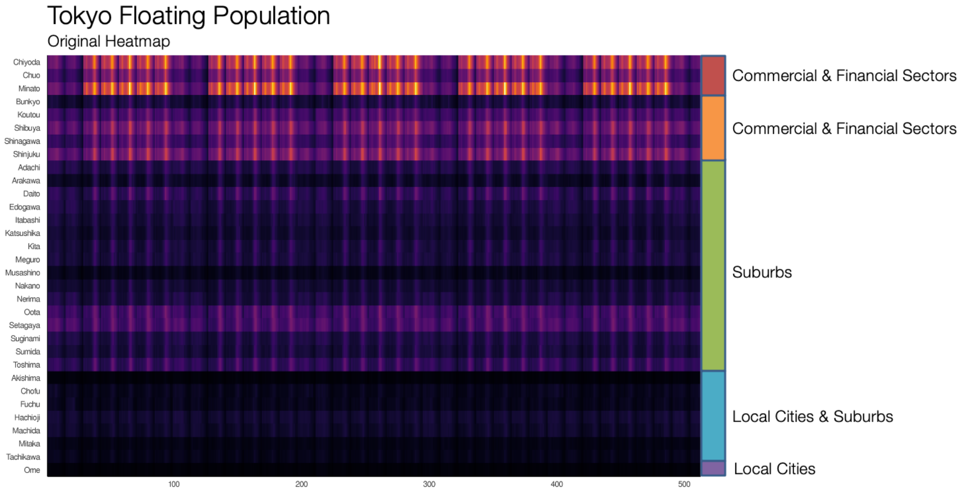
Data Analysis 2: Floating Population



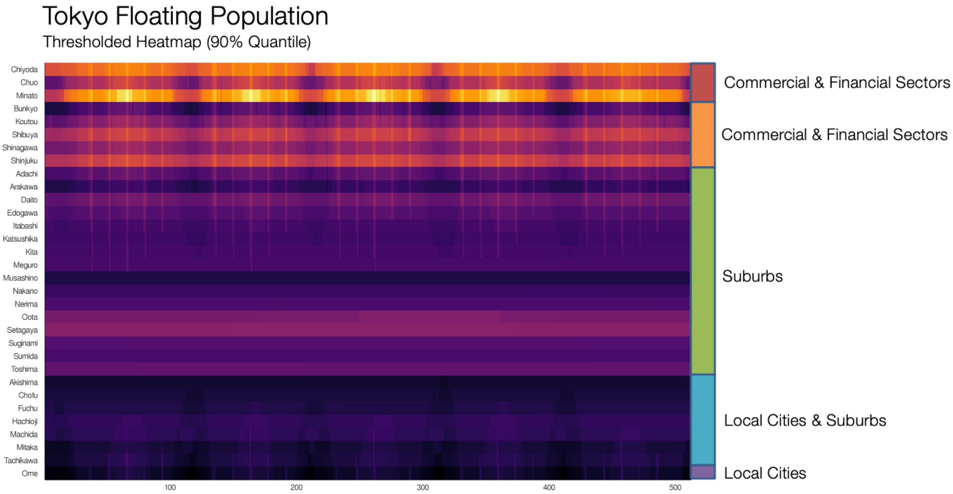
Data Analysis 2: Floating Population



Data Analysis 2: Floating Population



Data Analysis 2: Floating Population



Summary

- Wavelets are functions that contain frequency and time information
- We can represent signals with a basis of wavelets
- The advantages of the autocorrelation wavelet transform is that it is both shift-invariant and symmetric
- Some useful applications of the 2D autocorrelation wavelet transform are denoising and trend analysis