NYC DATA SCIENCE ACADEMY

Introduction to Time Series

A Primer on Univariate Time Series Analysis

Agenda

- Introduction to Univariate Time Series
- Time Series Plots
- Time Series Models
- Stationary & Diagnostic
- Forecasting /w Lecture Code

Time Series

Univariate

NYC DATA SCIENCE ACADEMY

Time Series - Univariate

	jj
1	0.710000
2	0.630000
3	0.850000
4	0.440000
5	0.610000
6	0.690000
7	0.920000
8	0.550000
9	0.720000
10	0.770000
11	0.920000
12	0.600000
13	0.830000

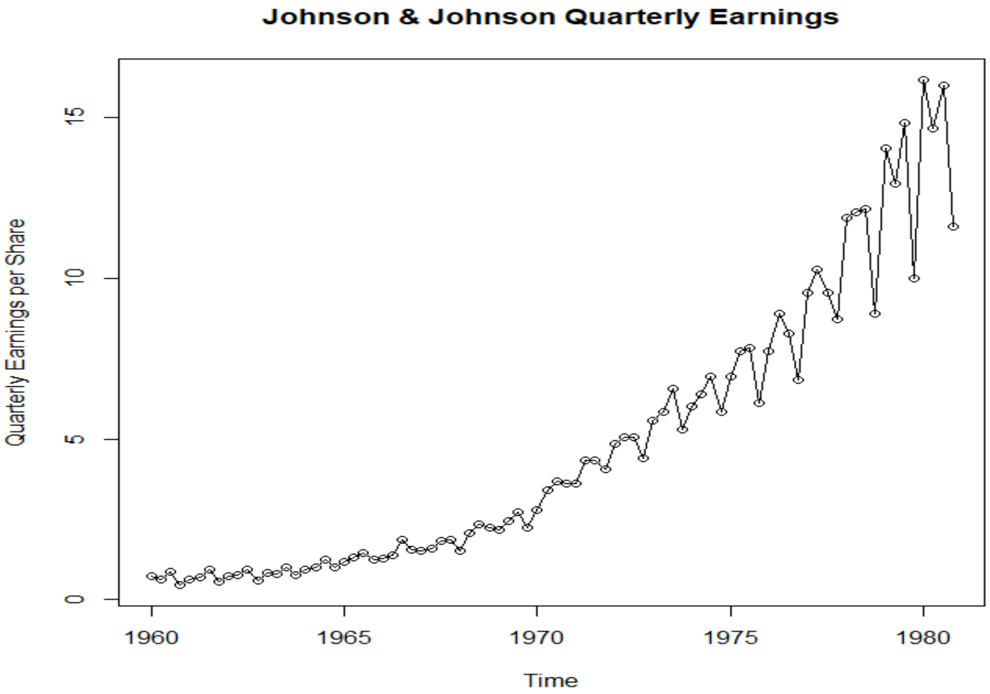
Times series is a form of data that may be encountered by data scientists in their careers. The most common encounter for a time series that is defined by a sequence of data that is tracked at equally spaced, successive, intervals in time. Specifically, we will be interested in examining discrete time-data rather than continuous (such as a sine wave). This lecture will draw from graphs and examples selected from the excellent resources: "*Time Series Analysis and Its Applications*", by Shumway with plots, "*Forecasting: Principles and Practice*" by Hyndman, and tools in R's libraries.

In order to introduce the material from an introductory manner, we will focus on univariate time series where we track a singular observation across each subsequent time interval. An example of univariate time series can be found in the graphic.

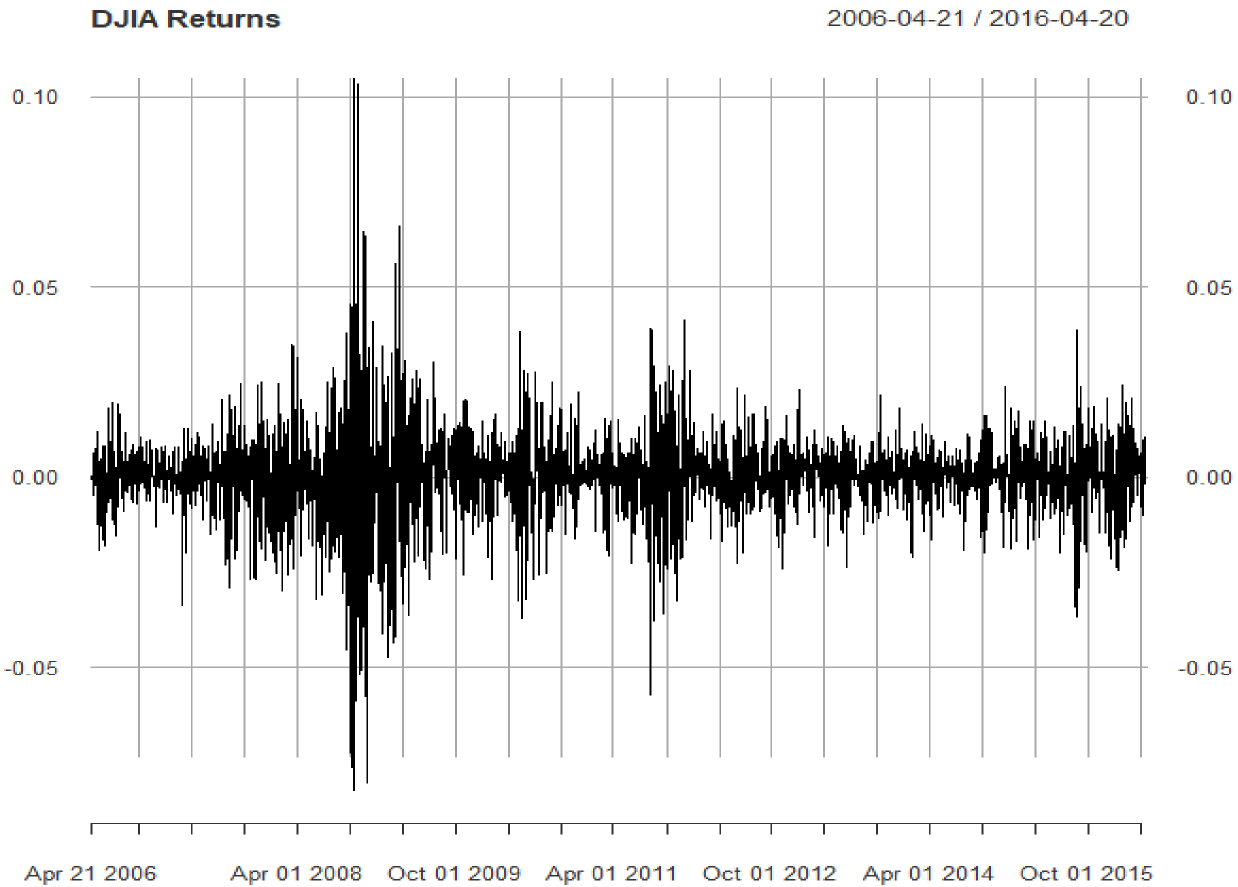
Note: there is no indication from the data itself that this may be time dependent but as this data is sourced from R's Johnson & Johnson Quarterly Earnings, we know this to be a univariate time series.

Time Series Plots

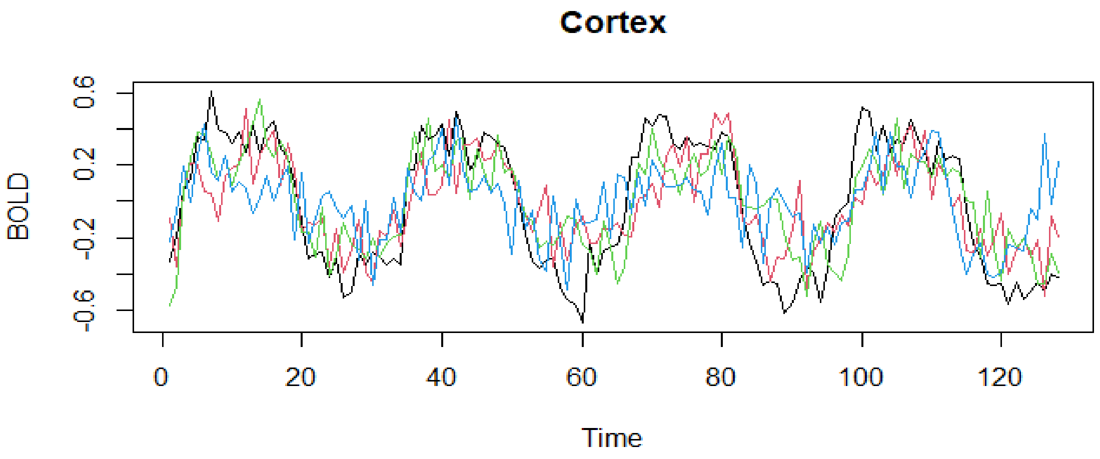
Visualizations serve as a powerful tool for data analysis and time series data is no different. Let’s observe an example:

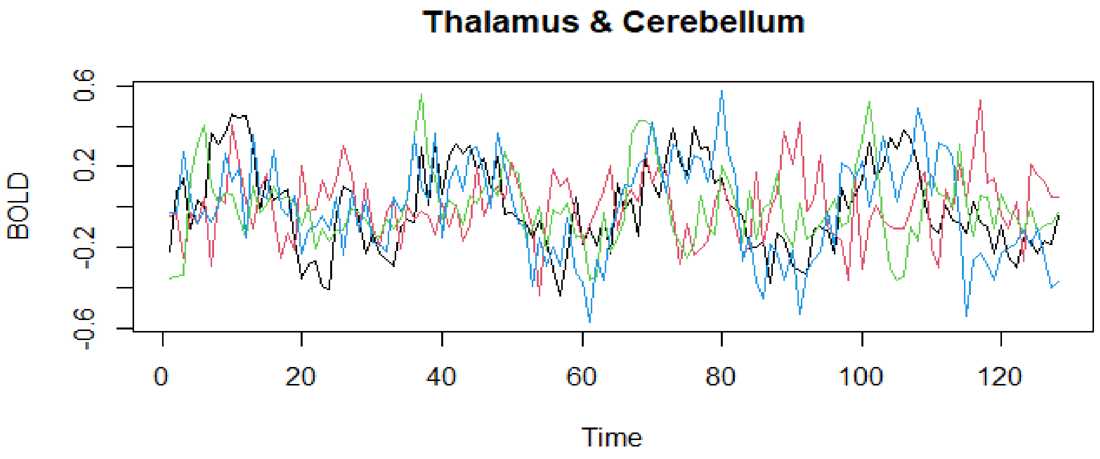


In the above graph, we find that each tick of the x-axis is in quarters. There are other common interval measurements that you may encounter within a dataset such as hours, daily, weekly, monthly, yearly, etc.



Here we have an example of finance data. What can we observe here? Note also that the frequency of the oscillations is daily.

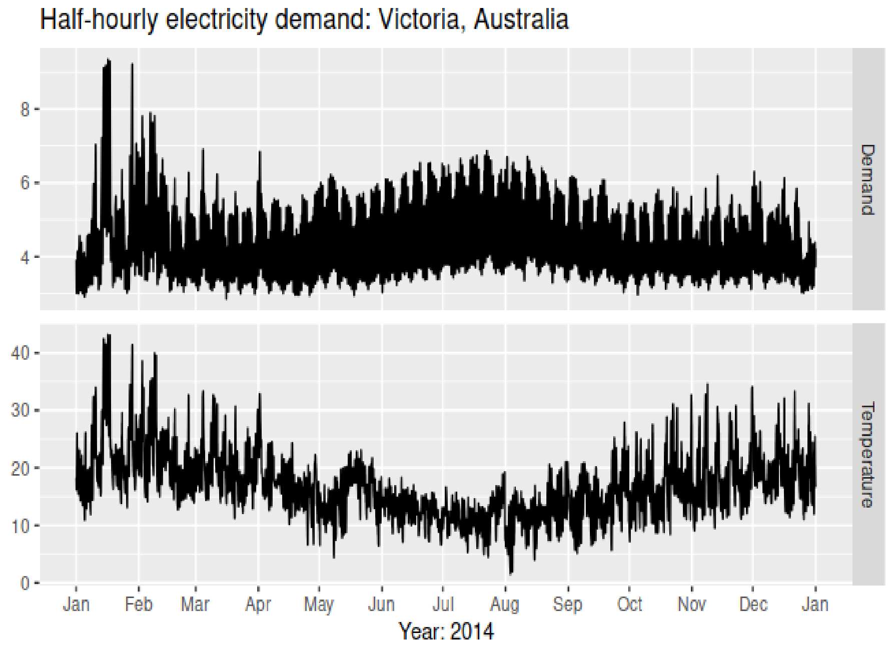


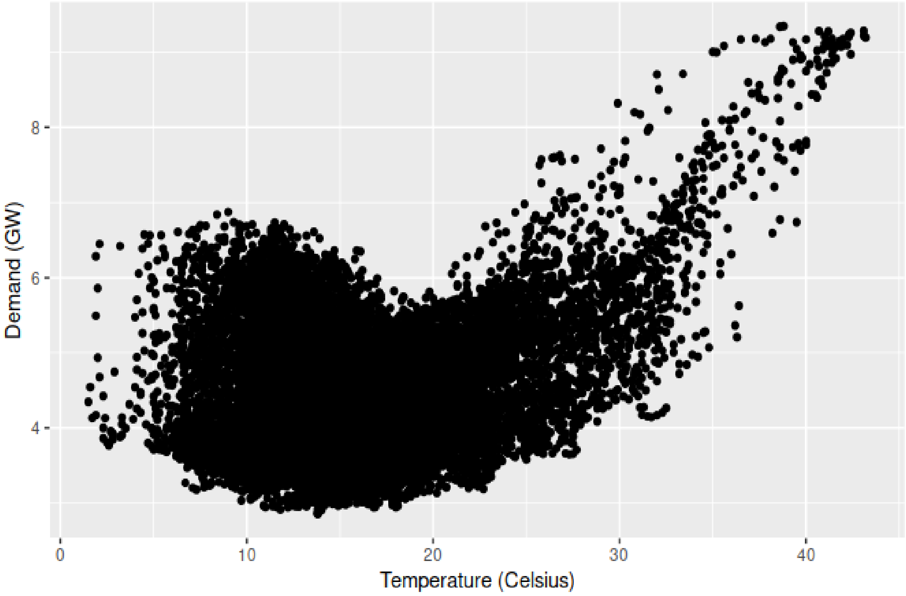


We can see a more novel application of time series. Here we have brain activities sampled every 2 seconds.

Two univariate time series can also be compared against each other to examine if there is a potential relationship between the two. As an example, we can visualize electricity demands and temperature. If we plot a scatterplot of two time series against each other, we observe clearly that past a certain temperature threshold, there is a strong correlation between temperature and electricity demand (people turn on their air conditioning once it gets hot “enough”).

Each of these visualizations allow us to observe some trend, event, cycle, or relationship in the time series for further analysis.





Source: “*Forecasting: Principles and Practice*” by Hyndman

Time Series Models

NYC DATA SCIENCE ACADEMY

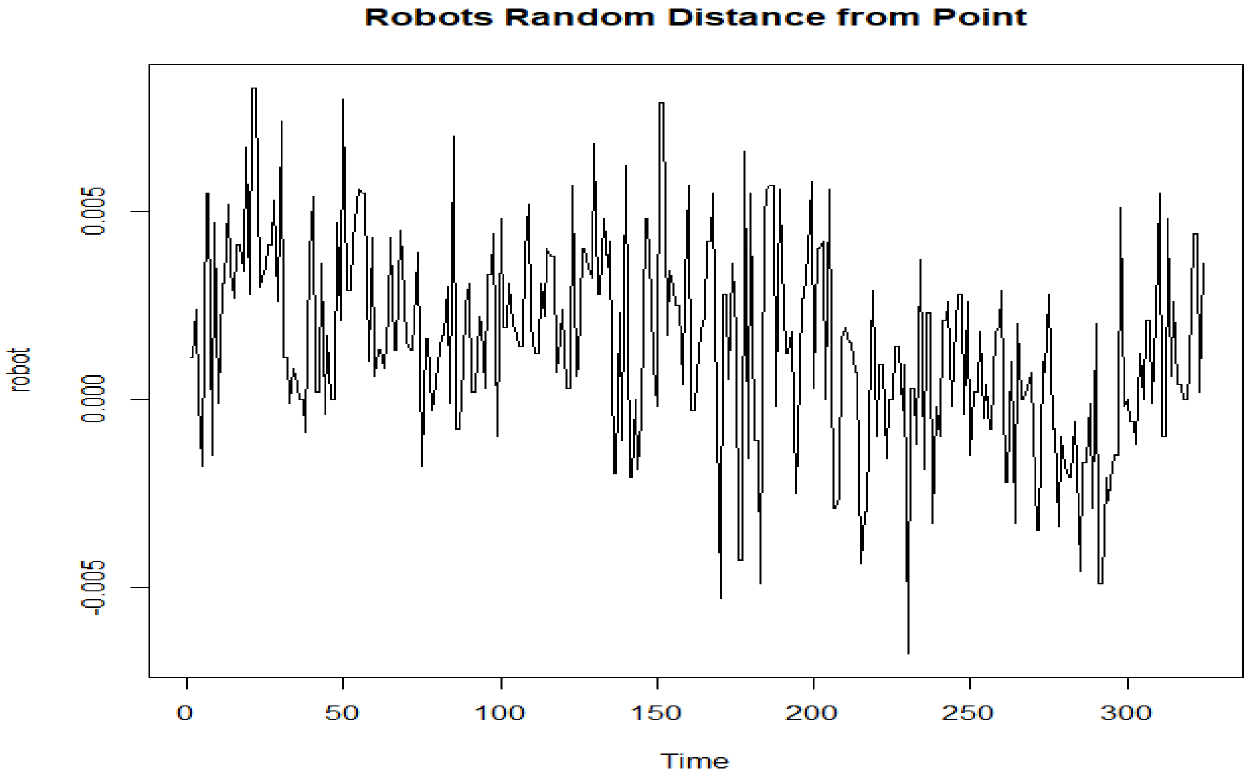
White Noise

A simple time series model is white noise where a collection of uncorrelated random variable that follows the process:

$w_t \sim wn(0, \sigma^2{}_w)$

We can generate a certain type of white noise known as *Gaussian white noise*, where $w_t \sim$ iid $N(0, \sigma^2{}_w)$. An example of this behavior can be seen on the next slide.

White Noise

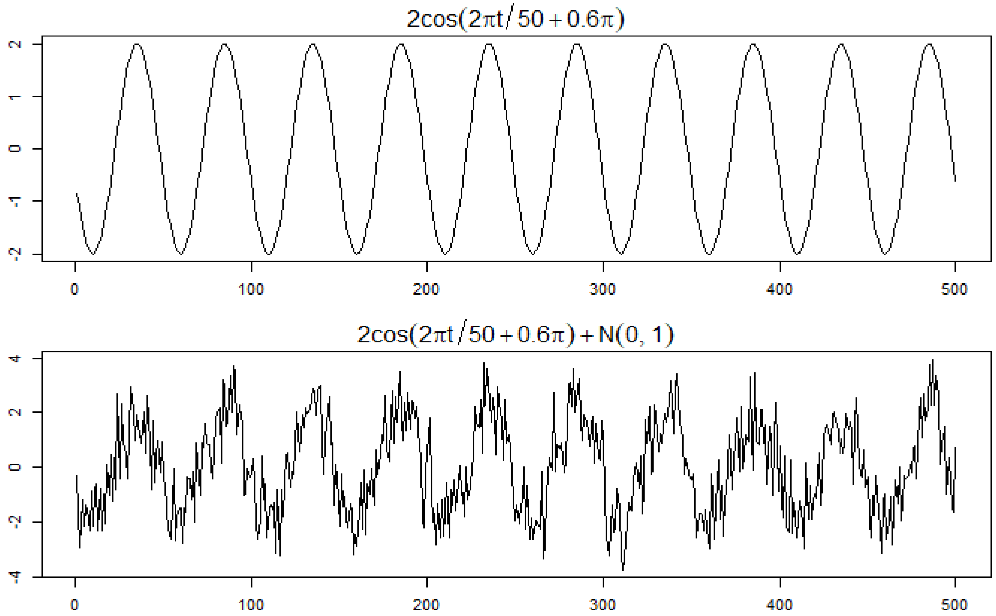

Robots Random Distance from Point
robot
0.005
0.000
-0.005
0
50
100
150
200
250
300
Time

White Noise

- If you observe a time series to be white noise, then it suggests that each time stamp is independent and uncorrelated with any other time stamp in the time series.
- If we observe the time series to be white noise then we can employ classical statistical methods to our data as it’s unaffected by the time element.
- If you decouple white noise from a time series, you can observe a time series process (if any) independent of randomness.

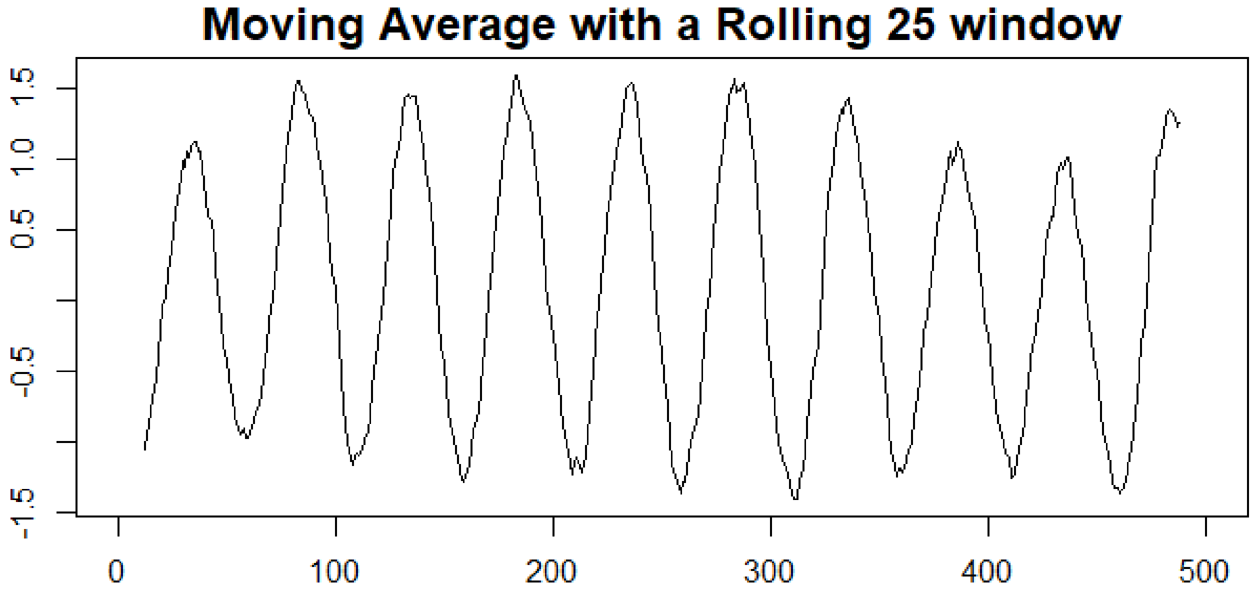
Moving Average

Practitioners may often encounter that white noise may obfuscate the actual signal in their data. As a result, it would be prudent to filter the white noise to examine an underlying signal. A demonstrated example of these two points can be observed by a toy example where we look at a sine wave with an additive white noise.



Moving Average

And then we add a moving average filter ...



Moving Average

Those from engineering may note that this is a common technique in the field of signal processing. More relevant reading can be found here.

Specifically speaking, we encounter moving average as a technique and further studies will reveal there is also a moving average model later in their studies (not covered here due to time constraints).

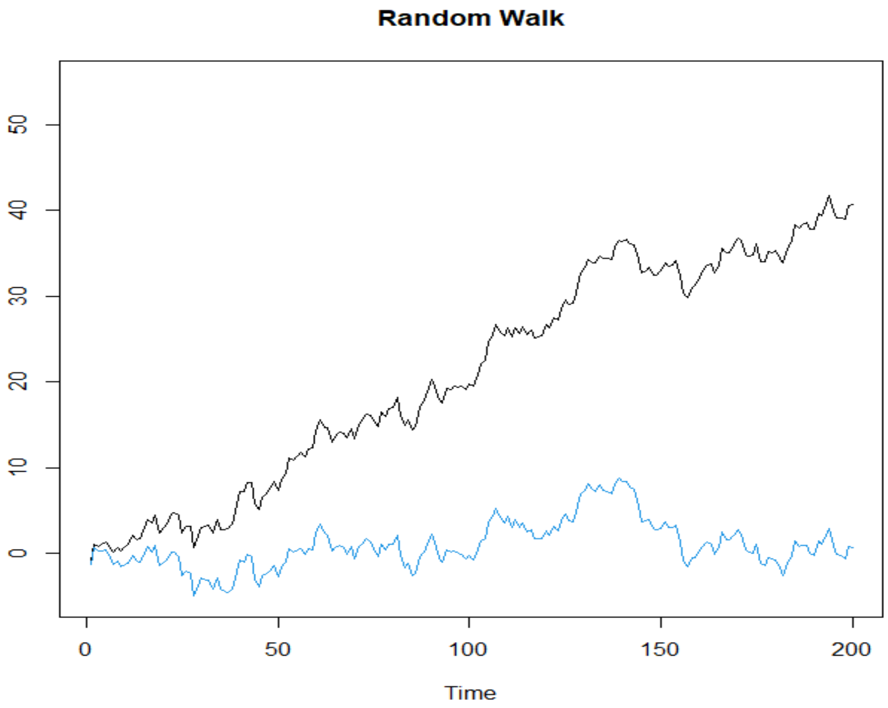
First we will specifically talk about it as a technique which was demonstrated prior with the formula:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j},$$

Note as k are the periods within time *t* we wish to average on and we examine the window of k before and after *t*. The denominator *m* is then = 2k+1 which yields for us a formula that acts as the rolling mean. This example can be found also in the accompanying lecture code. There are also other techniques of moving average such as weighted moving average that we will not discuss here but can be found through additional research.

Random Walk (with Drift)

A random walk (where w_t is Gaussian), both with and without drift, can be coded and plotted with the visualization as follows:



Random Walk (with Drift)

The random walk model represented by the following formula:

$x_t = x_{t-1} + w_t$

What we observe from the formula is that it states at each period, the variable took a random step from the previous period where the step is dictated by the behavior of w_t where w_t is white noise. This signals that if we know the behavior of w_t (such as if it is Gaussian white noise) we can predict our time series as a **cumulative sum of white noise variates**.

We can extend this model to include a certain behavior, specifically "drift" or where we observe that there is a trend in our model that can be represented by adding a constant at each step. With this idea, we can write the **random walk model with drift** as:

$x_t = x_{t-1} + w_t + \delta$

where δ is a constant.

Stationary and Diagnostic

NYC DATA SCIENCE ACADEMY

Stationary

We can think of time series as a particular realization produced by an underlying probability mechanism or a **stochastic process**. Investigating the nature of this process is critical to our examination of whether a time series model is appropriate (analogous to assumptions of a model).

Specifically speaking, our time series must be **stationary** to satisfy our models. In discussion stationarity**,** we are interested in two types:

- Strictly stationary
- Weakly stationary

Stationary

Strictly Stationary assumes that the **distribution** of the random variables of the process **does not change** at any points across the indices of time series. (a proof to examine this would be if the joint probability distribution is time invariant)

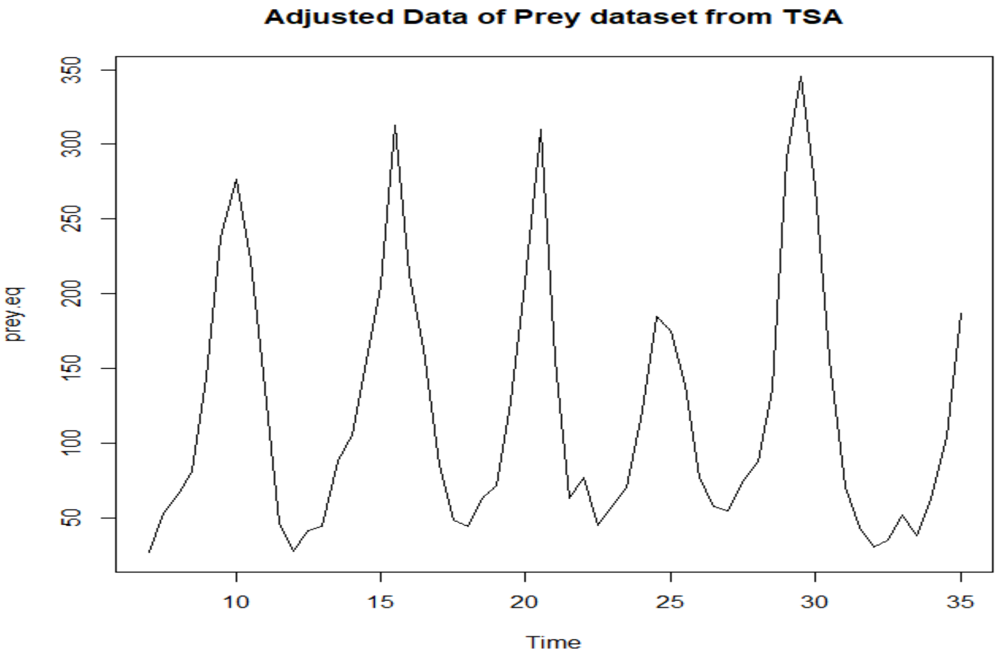
In practice, strictly stationary is difficult to satisfy, so we relax these constraints and it is common to state that a time series is “stationary” if it satisfies the requirements of **weakly stationary.**

Weakly stationary assumptions:

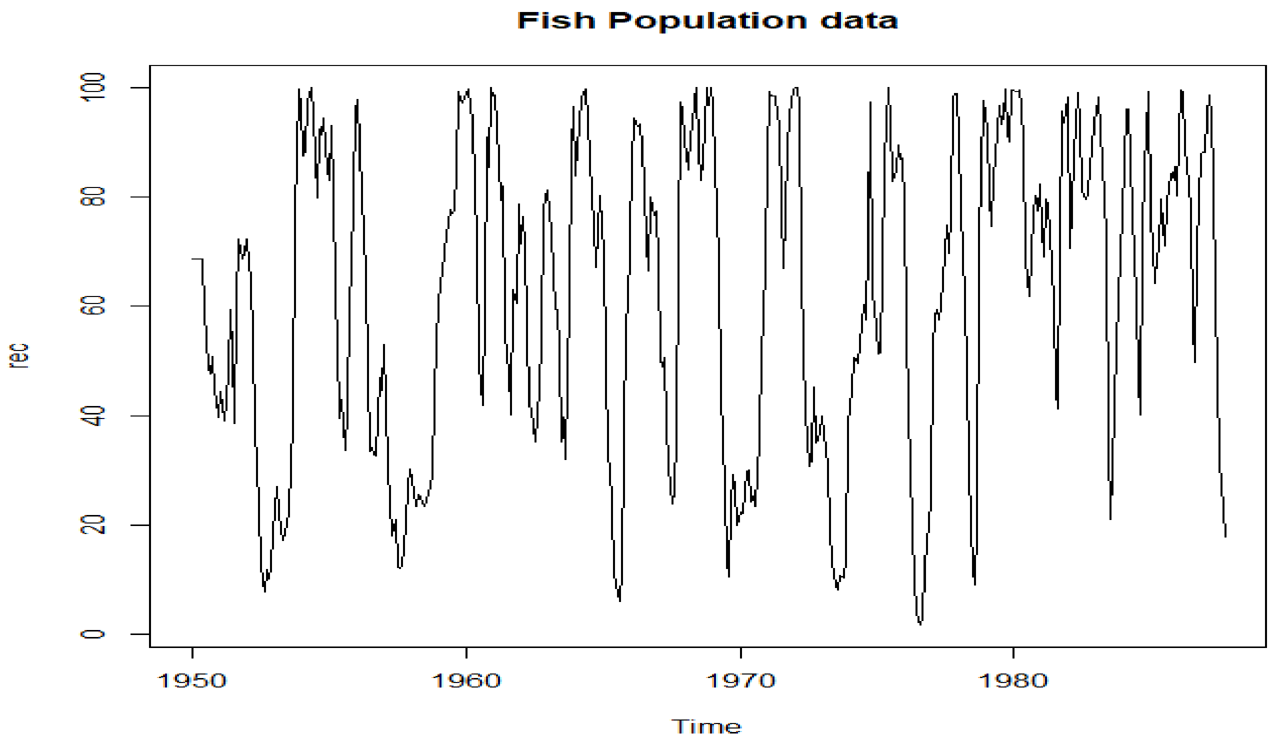
1) The mean of the time series is constant and does not depend on time
2) Variance of the time series is finite
3) (auto)covariance of any two periods (s,t) is dependent **only** on the difference between the two periods (t-s). (Note: this is also why weakly stationary is also referred to as “**covariance stationary**”)

Stationary (plots)

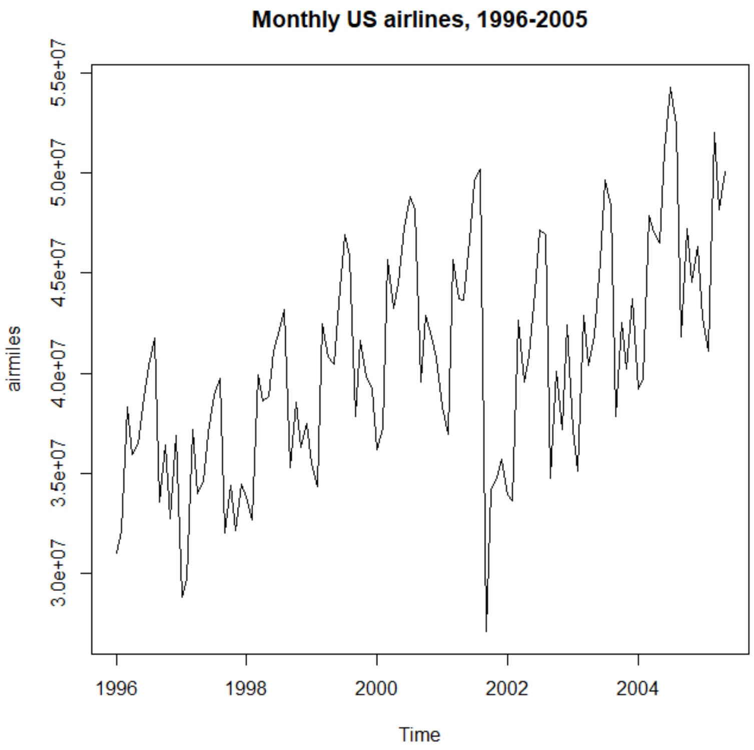
Let's review some plots to see if we can "eye - test" for stationary



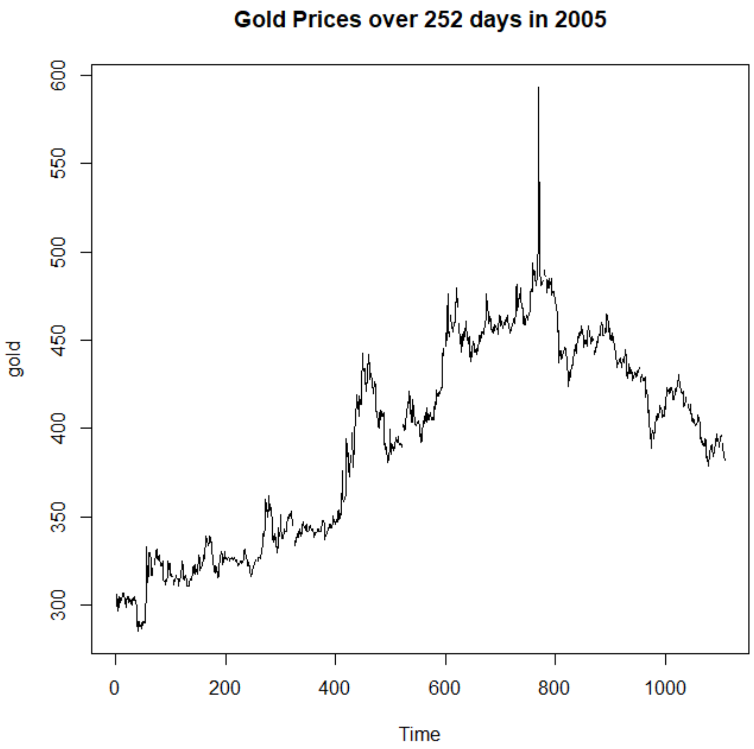
Stationary (plots)



Stationary (plots)



Stationary (plots)



Autocorrelation

The autocorrelation function (ACF) is defined as:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}.$$

In other words, it is the correlation between the value, x_t , at time “t” using only the value, x_s , at time “s”.

Partial Autocorrelation

Suppose, we regress on x_t using x_{t-k} and have an estimate x_t-hat.

If we calculate the correlation of [(x_t - x_{t-k}-hat), (x_t - x_t-hat)], we are performing the partial (auto)correlation function. It can be observed that we "partial out" the linear relationship and examine the correlation between the residuals of x_t and x_{t-k}. A simple example of why this would be useful is suppose we have a time series of x0, x1, x2, ... and assume that this data can be modeled as:

$$x_t = \boldsymbol{\Phi} x_{t-1} + w_t$$

We refer to this later as the **AR(1)** model and we will see how the **PACF** helps inform our AR model. If we examine the relationship between x_t and x_{t-2} , it is important to note that x_t will be dependent on x_{t-2} through x_{t-1}. As a result, our **ACF** will have some effect from the closest signal which will dampen over time. But we are interested in only the relationship between x_t and x_{t-2} so we look to remove the linear dependence of x_{t-1} on x_t and x_{t-2} with this method.

Forecasting

NYC DATA SCIENCE
ACADEMY

Autoregression

Recall that in a linear model, we regress/predict a target using a linear combination of predictors.

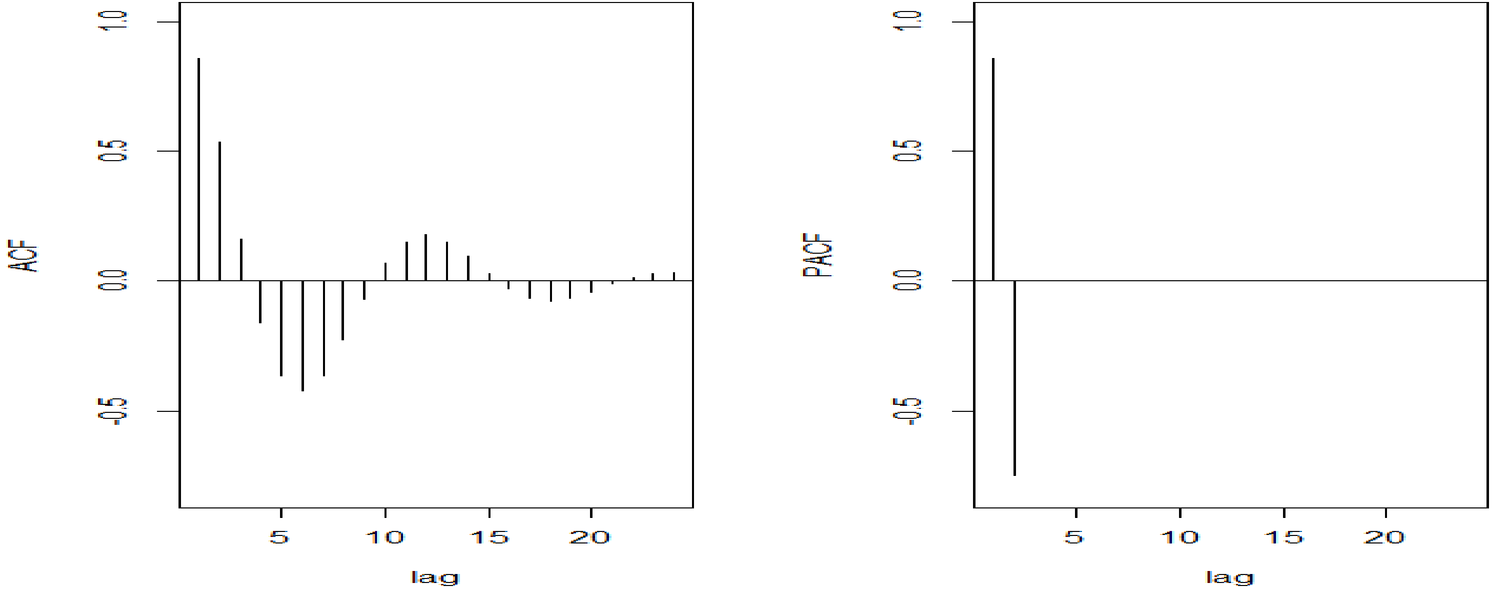
We can extend this to the time series application by using autoregression where our variables are *past* time stamps (lag variables).

Specifically if we look at the prior example, we can use PACF to identify the significant lags that can predict a value at a given period of time.

A useful plot for identifying whether our time series can be modeled by an autoregression model would be to examine the ACF/PACF plot.

ACF/PACF Plots

Based on the prior notes, how can we interpret the following plots?



Forecasting

An aim of analyzing time series is whether we can “predict the future” using the “past”. Specifically we refer to this as “forecasting”.

Since it’s impossible to acquire the future data (if we had it, there’d be no point in the forecast and we’d be seeing into the future…), it’s customary to validate our results by removing a tail end of our time series, then fitting and forecasting with our remaining data and comparing it against the removed “tail” of our time series.

We’ve reviewed the prerequisite concepts for modeling time series, let’s take a look at the lecture code for how to make a forecast a given time series

Main points

- How to analyze time series from visualizations
- Basic time series models to know
- Model Assumptions
- ACF/PACF
- Autoregression for modeling time series data
- Lecture Code