

# Correlation Structure

## Lecture 07.2: Autocorrelation

Lauren Sullivan

Module: Linear, Nonlinear, and Mixed Effects Models

# Readings

## Required for class:

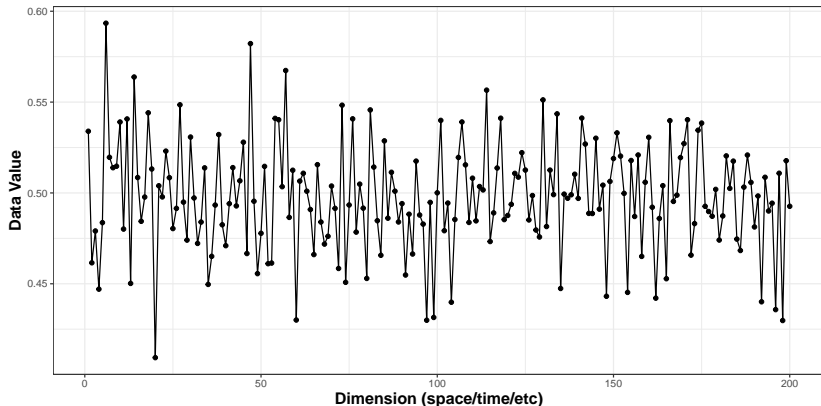
- ▶ NA

## Optional:

- ▶ M. Clark Mixed Models - Extensions for Residual Structure
- ▶ Dormann et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*
- ▶ J. Shah 2019. Time Series - ARMA Models in R

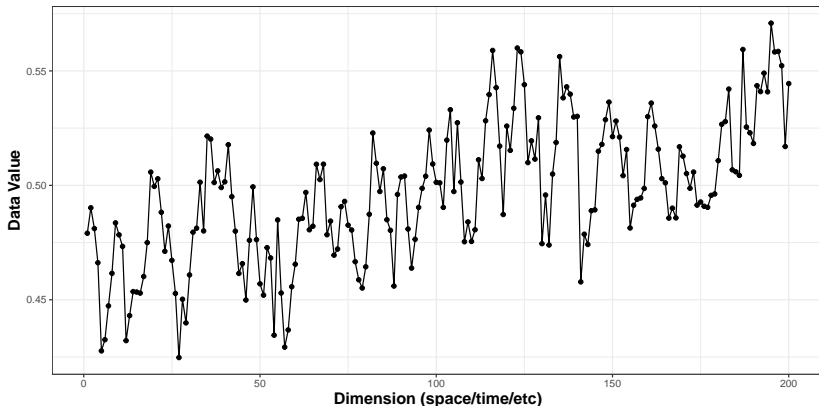
# Autocorrelation

When data are not correlated, and there is no covariance between points, then there is no autocorrelation in data. Data are said to be “random” and this is also called white noise. This is the assumption of most models.



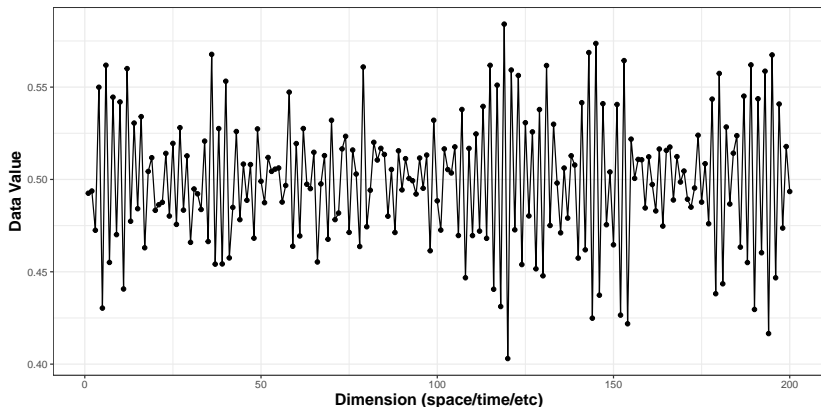
# Positive Autocorrelation

If data are positively autocorrelated, that means that data that are close in time/space/etc are more likely to be **similar**. This is also called red noise.



# Negative Autocorrelation

If data are negatively autocorrelated, that means that data that are close in time/space/etc are more likely to be **dissimilar**. This is also called blue noise.



# Autocorrelation

In our work, we will most likely have to account for positive autocorrelation, as it is common for data to be correlated through space or time. So when analyzing data through time (or space) we need to think about how to account for this autocorrelation.

- ▶ Average daily temperature
- ▶ Likelihood of rainfall events
- ▶ Moisture content in the soil
- ▶ Disease spread across the country

# ARMA models

A common to deal with autocorrelation is through autoregressive (AR) and moving-average (MA) models termed ARMA.

- ▶ **AR** *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations. (Makes predictions using previous values of the dependent variable)
- ▶ **MA** *Moving Average*. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. (Makes predictions using the series mean and previous errors)

Basic autocorrelation models we will think about will be either autoregressive **AR**, or autoregressive moving-average **ARMA** models.

# Stationarity

AR and MA models assume **stationarity** of your time-series data (or space-series). There are three requirements of stationarity

1. The mean of the series should be constant, not a function of time/space
2. The variance of the series should be homoscedastic (not a function of time/space)
3. The covariance of the  $i_{th}$  term and the  $(i + m)_{th}$  term should not be a function of time.

If data are not stationary, you can detrend (e.g use `log()`). We will assume all our data is stationary for this course.



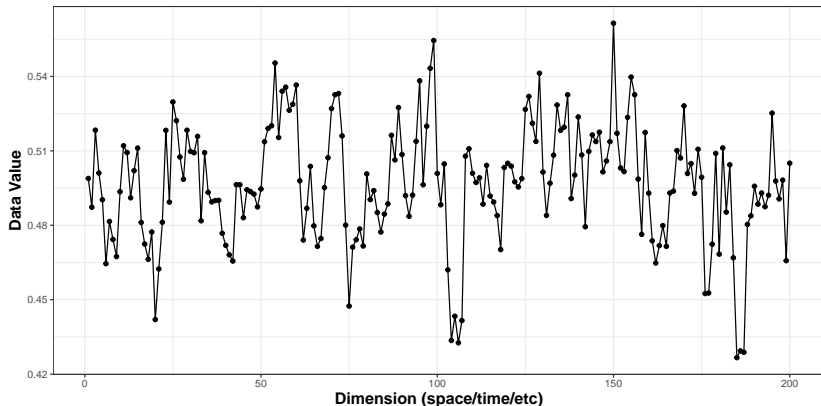
## Note about Autocorrelation in Time/Space

There are many many great ways to investigate time/space series data, such as using the current time series to forecast future predictions of the data. ([See this](#) for useful R functions for forecasting with time series data).

In this course we will only be covering the very basics of accounting for autocorrelation in time/space. We will talk about how to take this correlation into account within a mixed effects model.

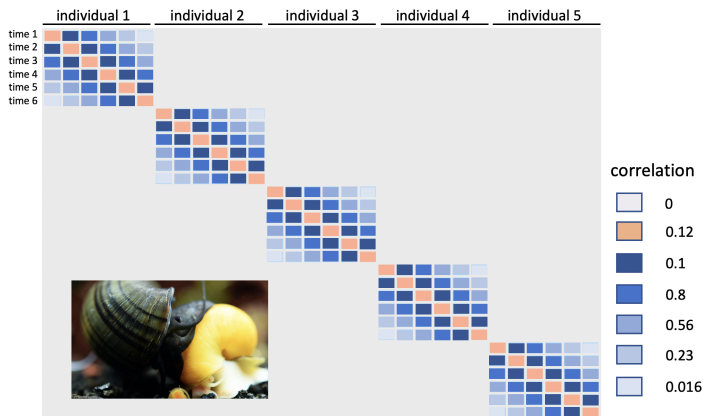
# Positive Autocorrelation

Ok, so let's get back to our positive autocorrelation. Let's consider the case where data that are closer in time are more likely to be similar than those farther in time. This is identical for space, FYI but we will use time as an example.



# Revisiting the Snail Example

So if we go back to our Mystery Snail example, we remember that there were 6 measurements of biomass consumption through time per individual. It may be more likely that the measurements closer in time are more likely to be more similar to each other than those farther in time.



## Correlations with Mixed Models in `lme()`

You can look at [this page](#) for documentation on how to specify the `correlation` argument, and [this page](#) for the different type of classes of correlations. Here you will most likely want to use `corAR1` or `corARMA`. We will use the former here, but try the later on your own.

```
mod_corr <- lme(biomass ~ sex + month,  
               data = snails,  
               random = ~1|individual,  
               correlation = corAR1(form = ~month))
```

# Model Output

```
## Linear mixed-effects model fit by REML
##   Data: snails
##       AIC      BIC    logLik
##   4178.27 4208.795 -2083.135
##
## Random effects:
##   Formula: ~1 | individual
##           (Intercept) Residual
## StdDev:      1.012856  1.36648
##
## Correlation Structure: AR(1)
##   Formula: ~month | individual
##   Parameter estimate(s):
##       Phi
##   0.4183476
## Fixed effects:  biomass ~ sex + month
##               Value Std.Error DF t-value p-value
## (Intercept) 12.795193 0.15535686 999 82.36001 0e+00
## sexmale     -0.729291 0.18122413 198 -4.02425 1e-04
## month        0.535718 0.02639684 999 20.29477 0e+00
## Correlation:
##       (Intr) sexmal
## sexmale -0.554
## month   -0.595  0.000
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -2.95734092 -0.58804004 -0.01897382  0.61183122  2.76106338
##
## Number of Observations: 1200
## Number of Groups: 200
```

# Phylogeny

Taking into account the phylogeny of your data is another form of correlation. It's more complicated because we need to think more deeply about matrices and how to work with them. So we will come back to this in our multivariate unit.

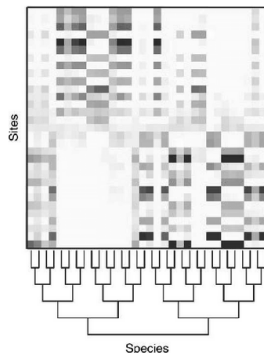


Figure 1: From: Ives and Helmus 2011

## Additional thoughts

If you have a correlation structure or weights similar to what we have discussed this week that you want to use for a linear model without random effects, use `gls()`. [RDocumentation here](#).