Temporal and Spatial Correlation

BIO 599

“Everything is related to everything else, but near things are more related than distant things.”
Tobler’s first law of geography

Reminders/Updates

Assignment #2 is graded

Jeremy Forsythe will be here next week

Paper discussion on Thursday:
new paper posted as of yesterday!

Assignment #2 - Feedback

What went well

- Clear ecological questions and good use of linear/mixed models
- Many of you started interpreting coefficients biologically
- Figures and organization improving

What needs help

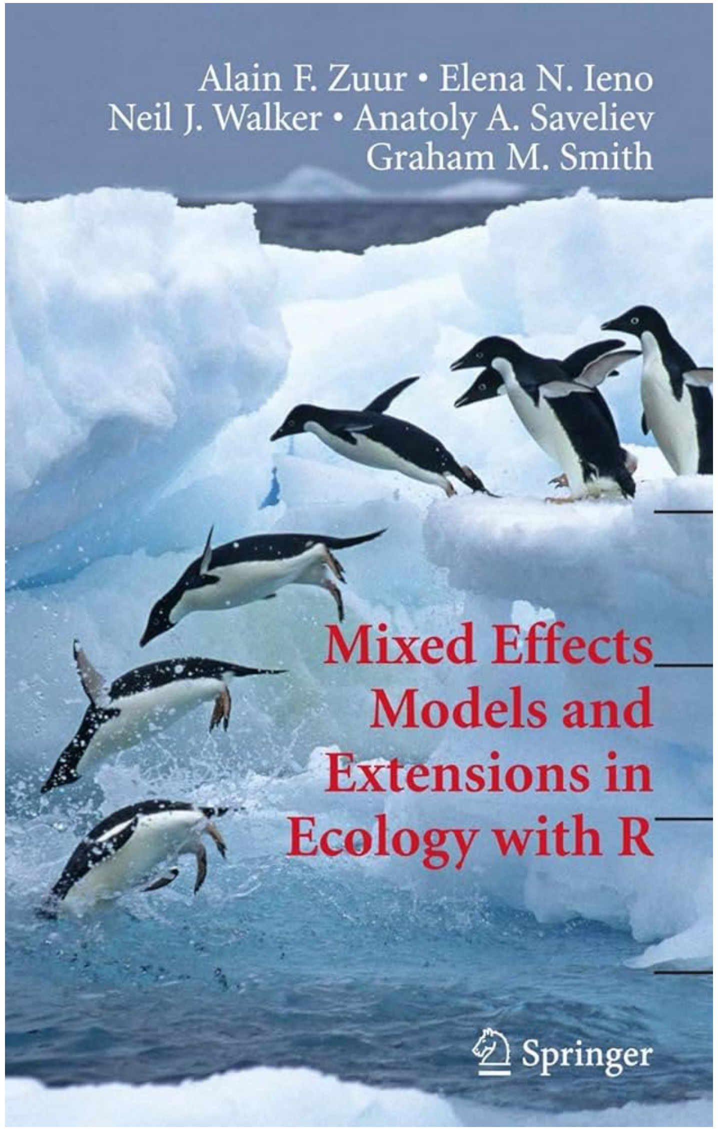
- Hypotheses need clear **direction** and **mechanistic reasoning**
- Choose **random effects and variance structures** *before* reducing fixed effects
- **Interpret** diagnostic plots — don’t just show them
- Figures should display **model predictions** with confidence intervals
- Use **narrative explanations**, not just R code, to answer questions

Outline for today

1. Overview of autocorrelation

2. Temporal Autocorrelation Example

3. Spatial Autocorrelation Example



Problems with linear regression

1. Heterogeneity - GLS
2. Nested data – LME
3. Temporal and Spatial Correlation (today)

Homogeneity of variance

$$y_i = \underbrace{\beta_0 + \beta_1 \times x_i}_{\text{Linearity}} + \varepsilon_i \qquad \varepsilon_i \sim \underbrace{\mathcal{N}(0,\sigma^2)}_{\text{Normality}} \qquad \mathbf{V} = cov = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & \vdots \\ \vdots & \cdots & \sigma^2 & \vdots \\ 0 & \cdots & \cdots & \sigma^2 \end{pmatrix}$$

Zero covariance (=independence)

Assumes ZERO co-variance between observations

autocorrelation: when our obs are
not independent.

What is autocorrelation?

- Autocorrelation in regression analysis refers to the correlation between the residuals (errors) of a regression model at different time points or data points

What happens when we violate the independence assumption with autocorrelation?

- Sample size (n) is the denominator when calculating SE and CIs (↑n = ↓SE & ↓ CI)

- with autocorrelated data each new datapoint is related to a previously collected datapoint and does not bring a full independent datapoint worth of information (90% autocorrelation = 10% new information)

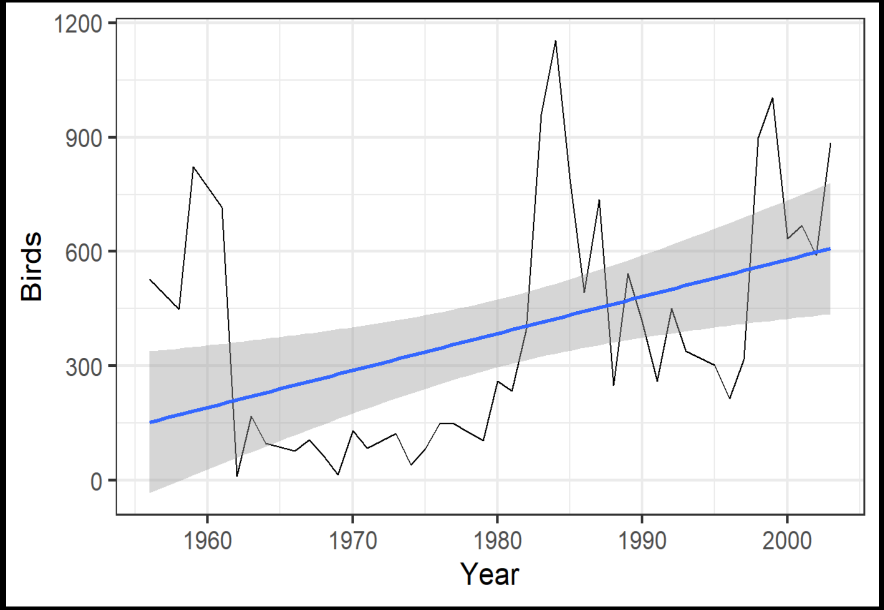
What happens when we violate the independence assumption with autocorrelation?

- Incorrect confidence intervals and p-values
- Biased regression coefficient estimates and significance
- Spurious relationships

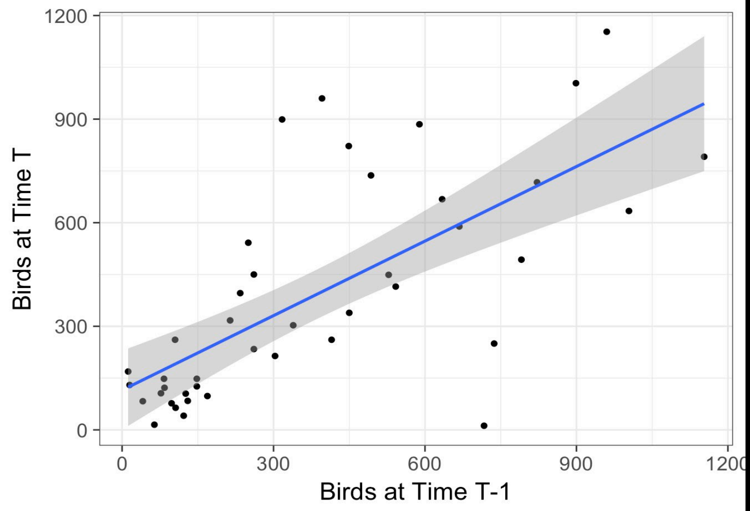
Temporal correlation could falsely generate a trend

2nd measurement is correlated with 1st measurement

Bird abundance is increasing

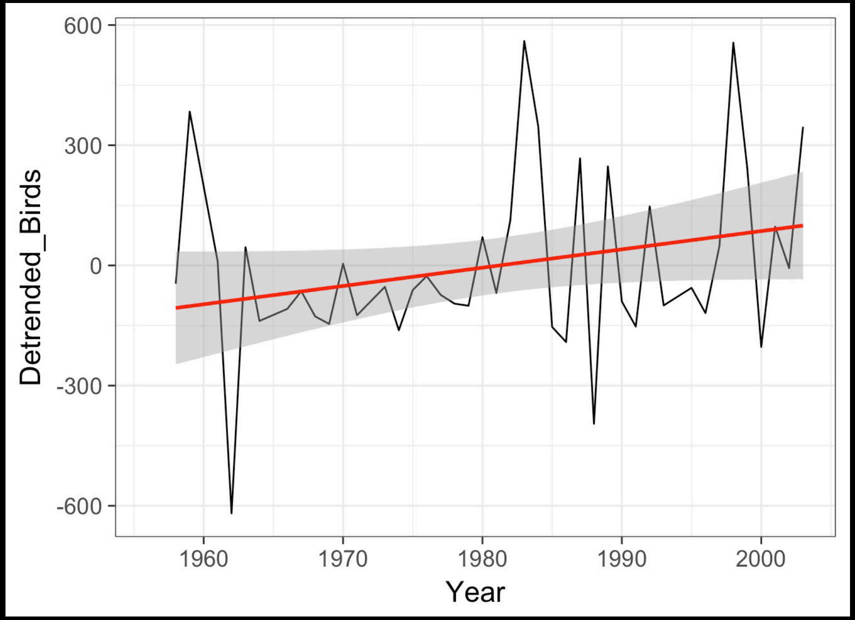


...strong autocorrelation



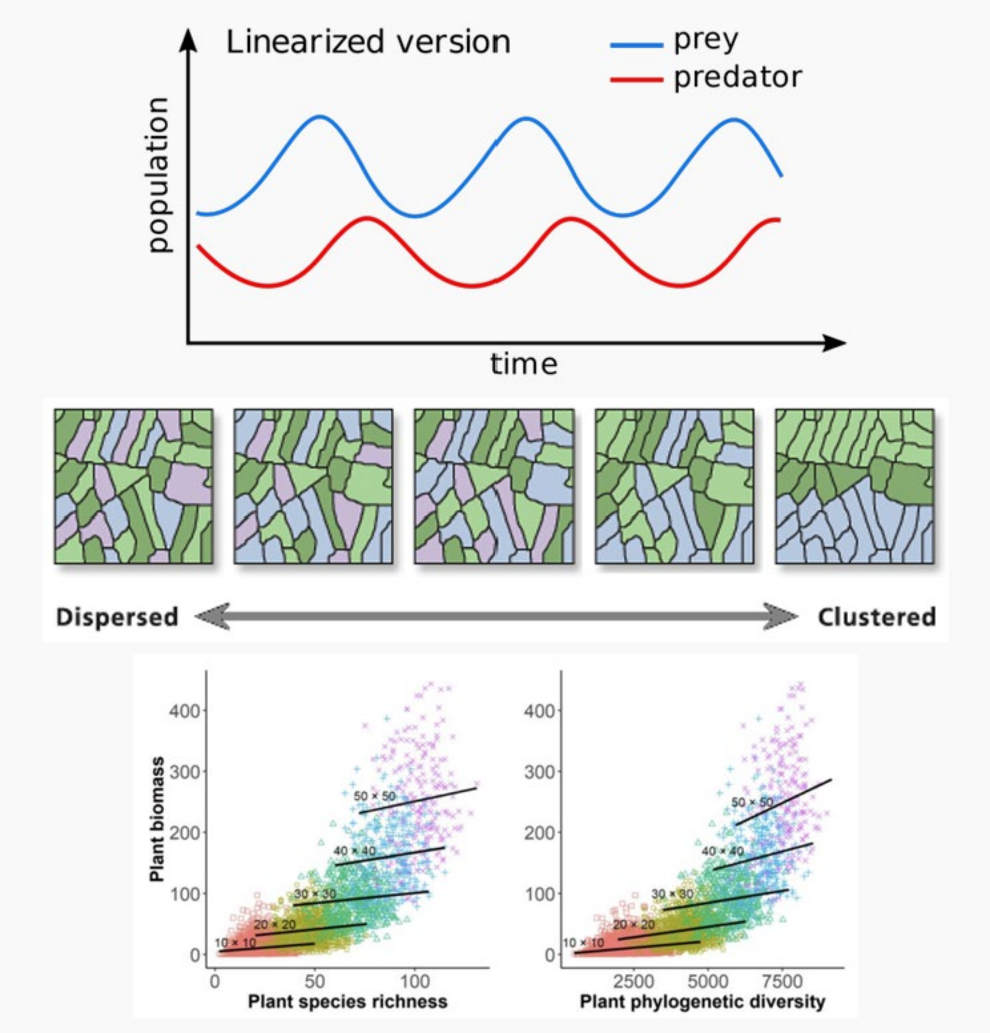
Correlation (ρ) Between T and T-1 = 0.699

Bird abundance is stable



What can lead to autocorrelation?

- Time: periods closer in time are more alike, repeatedly sampling the same replicate
- Space: data that closer together in space are more alike
- Phylogeny: species that are closer together on an evolutionary timescale are more alike



What can lead to autocorrelation?

DISCUSSION:

Think about your own study system.

What type of autocorrelation might be present?

time & space autocorrelation

How can I check for autocorrelation?

1. Plot residuals against any temporal or spatial variable. A pattern that is not random suggests a lack of independence.

2. Use the autocorrelation function (ACF) and pACF on your model residuals (temporal)

3. Make variograms of the residuals (spatial)

4. Add autocorrelation structure into the model and compare to model without using AIC and/or LRT

Choosing a correlation structure – nlme package

Temporal

corAR1
autoregressive process of order 1.
corARMA
autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.
corCAR1
continuous autoregressive process (AR(1) process for a continuous time covariate).
corCompSymm
compound symmetry structure corresponding to a constant correlation.

Spatial

corExp
exponential spatial correlation.
corGaus
Gaussian spatial correlation.
corLin
linear spatial correlation.
corRatio
Rational quadratics spatial correlation.
corSpher
spherical spatial correlation.
corSymm
general correlation matrix, with no additional structure.

Choosing a correlation structure

Schabenberger and Pierce (2002):

‘In our experience it is more important to model the correlation structure in a reasonable and meaningful way rather than to model the correlation structure perfectly’

Other assumptions?

- **Linearity:** The relationship between the dependent variable and the independent variables is linear (for linear models – but can also include autocorrelation in glm, gam, etc.)
- **Independence of errors**: Autocorrelation violates this assumption as it implies a correlation structure among the residuals over time or across observations – but you can account for it!
- **Homoscedasticity**: The variance of the errors is constant across all levels of the independent variables.
- **Normality of errors**: The errors are normally distributed.

Temporal Correlation - example

Fitting a timeseries with GLS

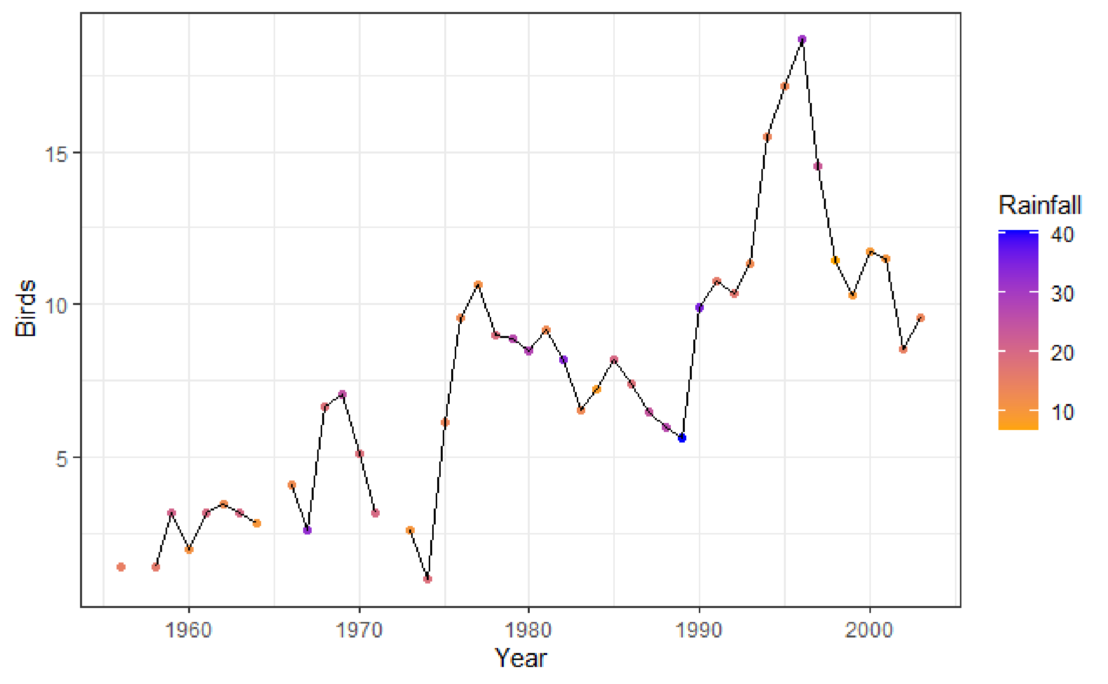
- Apply model without autocorrelation
- Check for autocorrelation
- Add autocorrelation structure

Other correlation structures

Modeling with many time series

Temporal correlation - example

- Hawaii birds dataset - abundance of birds from 1956-2003 as a function of rainfall and year
- For simplicity, start by looking at ONE group of birds: Moorhen on Kauai



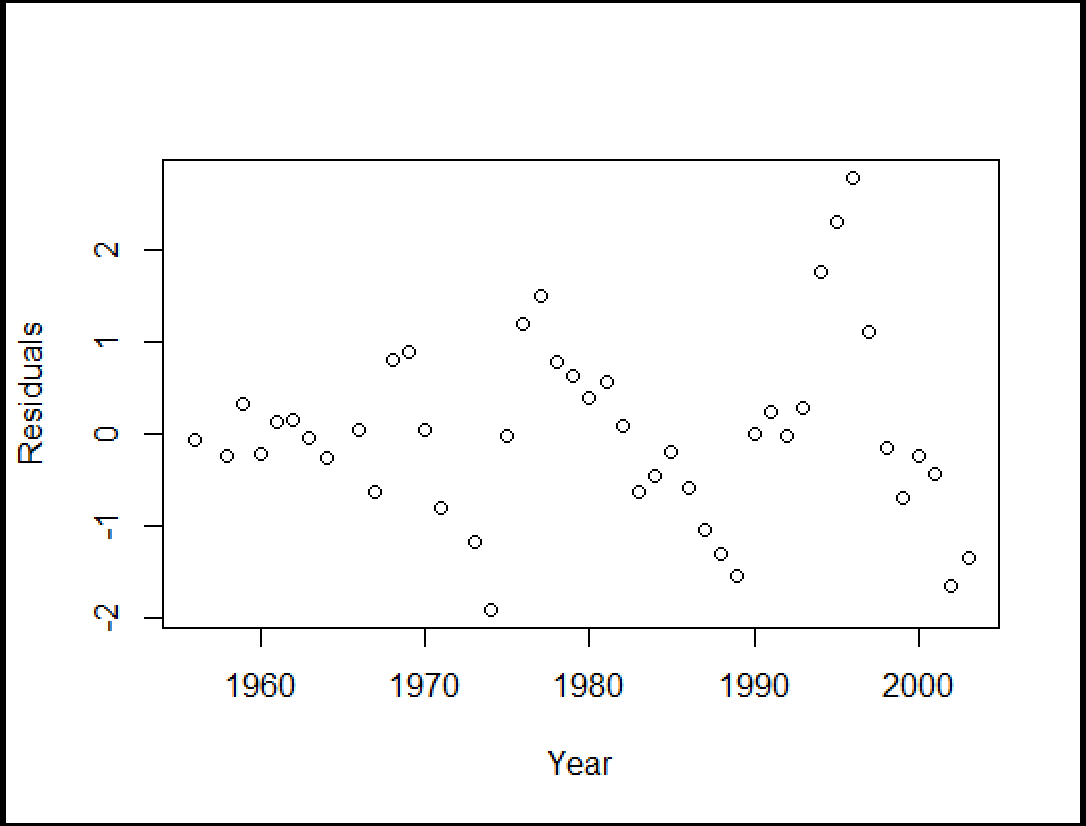
1. Apply the model without auto-correlation

```
birds_lm <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data = Hawaii)
tab_model(birds_lm, show.aic = TRUE, show.stat=TRUE, show.se=TRUE)
```

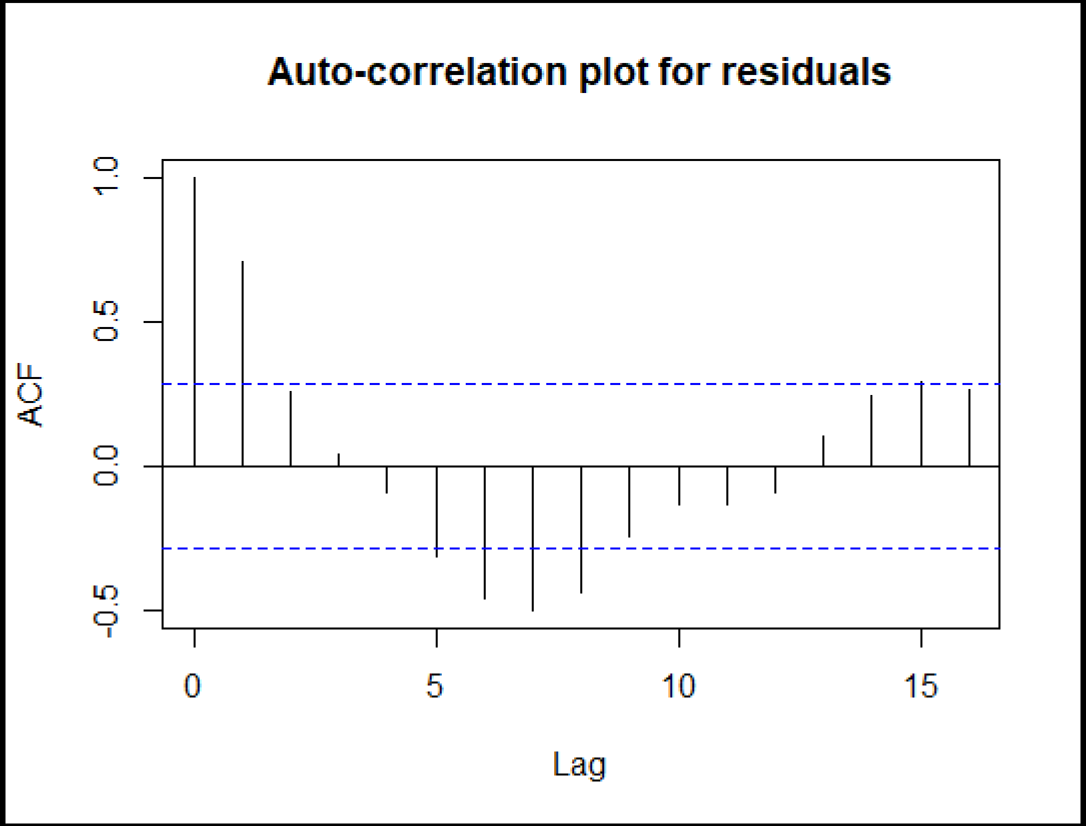
	Birds				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-477.66	56.42	-591.52 – -363.81	-8.47	**<0.001**
Rainfall	0.00	0.05	-0.10 – 0.10	0.02	0.986
Year	0.25	0.03	0.19 – 0.30	8.60	**<0.001**
Observations	45				
R^2	0.638				
AIC	218.886				

2. Check for autocorrelation

Residuals vs. year



Autocorrelation function (ACF)




```
acf(residuals(birds_lm), na.action = na.pass, main="Auto-correlation plot for residuals")
```

3. Add autocorrelation into model

$$cor(\epsilon) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

No covariance (Independent)

$$cor(\epsilon) = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Compound Symmetric Structure:
All points are equally correlated

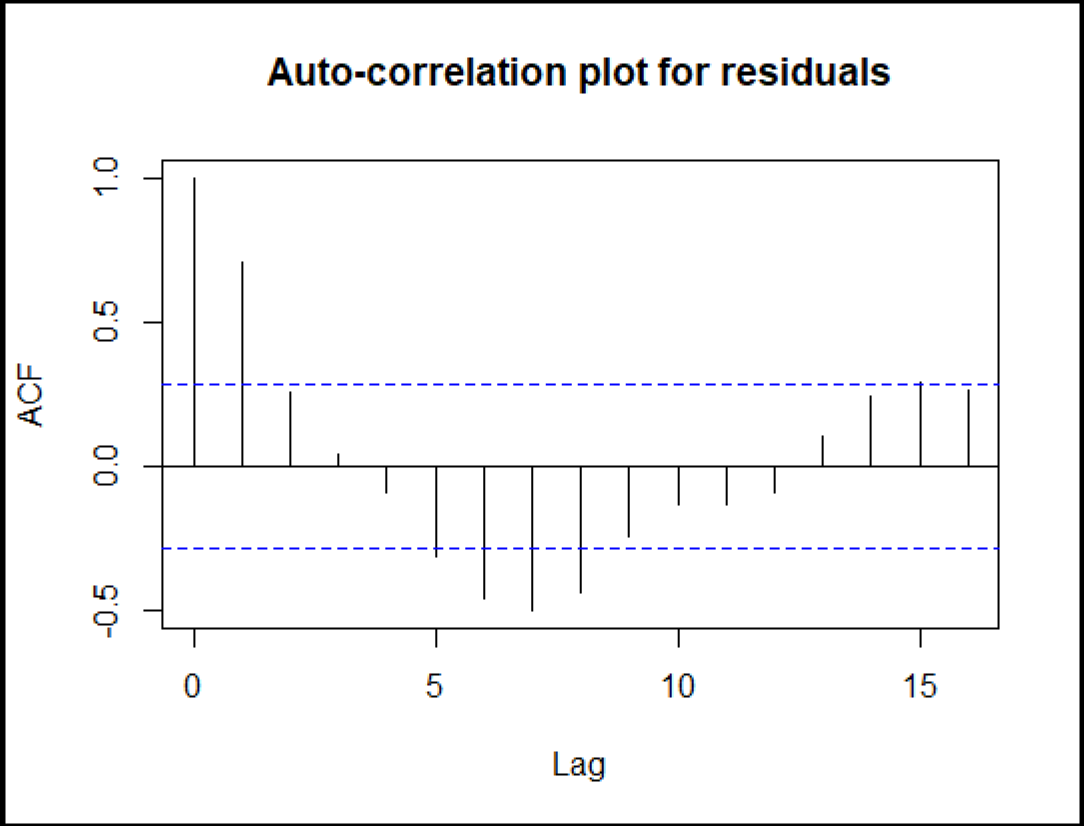
3. Add autocorrelation into model

```
birds_ccs<-gls(Birds ~ Rainfall + Year, na.action = na.omit,
        correlation = corCompSymm(form =~ Year),
        data=Hawaii)
tab_model(birds_ccs, show.aic = TRUE, show.stat=TRUE, show.se=TRUE)
```

	Birds				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-477.66	56.42	-591.52 – -363.81	-8.47	**<0.001**
Rainfall	0.00	0.05	-0.10 – 0.10	0.02	0.986
Year	0.25	0.03	0.19 – 0.30	8.60	**<0.001**
Observations	45				
R^2	0.638				
AIC	220.886				

NO improvement over GLS (OLS) model

But is equally correlated error valid?



NO! We see the dropoff

Autoregressive error structure – AR1

$$\operatorname{cor}(\varepsilon) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{57} \\ \rho & 1 & \rho & \ddots & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 & \rho^3 \\ \rho^3 & \rho^2 & \rho & \ddots & \rho & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & 1 & \rho \\ \rho^{57} & \cdots & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

- the further away two residuals are separated in time, the lower their correlation
- Suppose the correlation between residuals separated by one unit in time is 0.5. If the separation is two units in time, the correlation is $0.5^2 = 0.25$....etc.

3. Add autocorrelation into model

```
birds_ar1<-gls(Birds ~ Rainfall + Year, na.action = na.omit,
      correlation = corAR1(form =~ Year), data = Hawaii)
tab_model(birds_ar1, show.aic = TRUE, show.stat=TRUE, show.se=TRUE, show.re.var=T)
summary(birds_ar1)
```

```
Correlation Structure: ARMA(1,0)
 Formula: ~Year
 Parameter estimate(s):
    Phi1
0.7734303
```

	Birds				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-436.43	138.75	-716.44 – -156.42	-3.15	**0.003**
Rainfall	-0.01	0.03	-0.08 – 0.06	-0.30	0.765
Year	0.22	0.07	0.08 – 0.37	3.20	**0.003**
Observations	45				
R^2	0.627				
AIC	192.595				

Pros and Cons of AR1 structure

Relatively simple and easy to interpret

Only a 1-year lag built in

Need discretely timed sampling intervals

Temporal Correlation - example

Fitting a timeseries with GLS

- Apply model without autocorrelation
- Check for autocorrelation
- Add autocorrelation structure

Other correlation structures

Modeling with many time series

Other correlation structures

Continuous Autoregressive Process

- Uses continuous time
- Easier on data sets with gaps

Autoregressive Moving Average

- Variable autoregressive order (AR1 = 1)
- Incorporate moving average over time

Continuous Autoregressive (CAR) structure

- continuous covariate
- corCAR1 allows unequally spaced observations
- allows for things like missing time points, etc.
- in our example, produces identical results

ARMA structure

- Autoregressive Moving Average – two parts:
  p: the number of auto-regressive parameters (lag)
  q: the number of moving average parameters (average lag)

- The notation ARMA(1, 0) refers to the AR-1 model

- ALL parameters have to be estimated! Values greater than 2 or 3 tend to give error messages

ARMA structure

Two-year AR

```
birds_ar2 <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data=Hawaii,
                 correlation = corARMA(form = ~ Year, p = 2, q=0))
```

Two-year MA

```
birds_ma2 <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data=Hawaii,
                 correlation = corARMA(form = ~ Year, p = 0, q=2))
```

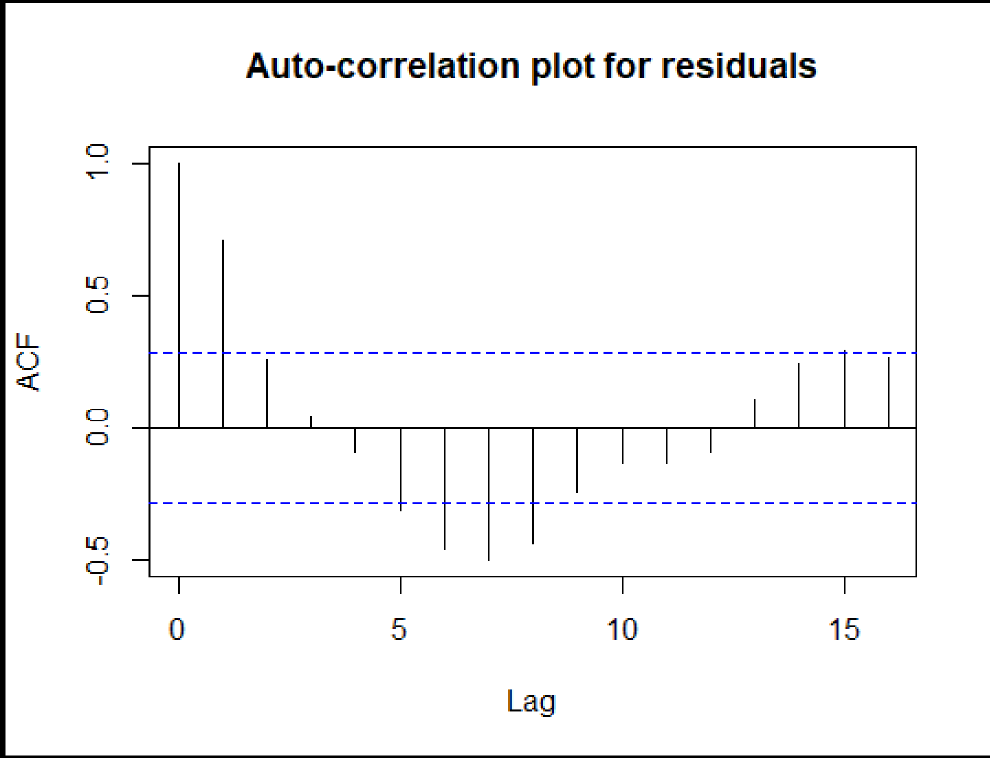
Two-year ARMA

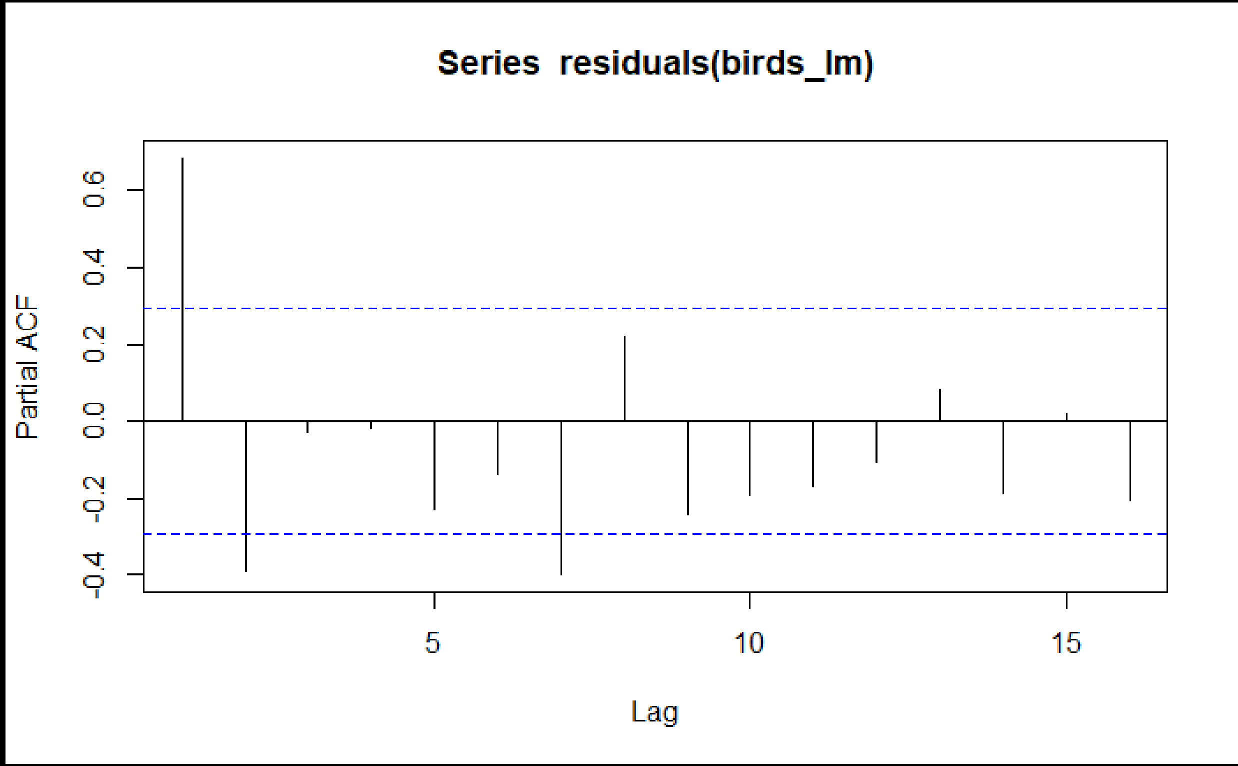
```
birds_arma2 <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data=Hawaii,
                   correlation = corARMA(form = ~ Year, p = 2, q=2))
```

Two-year AR (birs_ar2) has the lowest AIC

Could try all different combinations

AR vs. MA process?
How do I know if my lag is >1 time step?





AR : If ACF function shows long decay, and pACF shows a drop to 0 quickly
MA: If ACF function drops off quickly, but pACF shows a long decay
ARMA: If both ACF and pACF decay gradually

Temporal Correlation - example

Fitting a timeseries with GLS

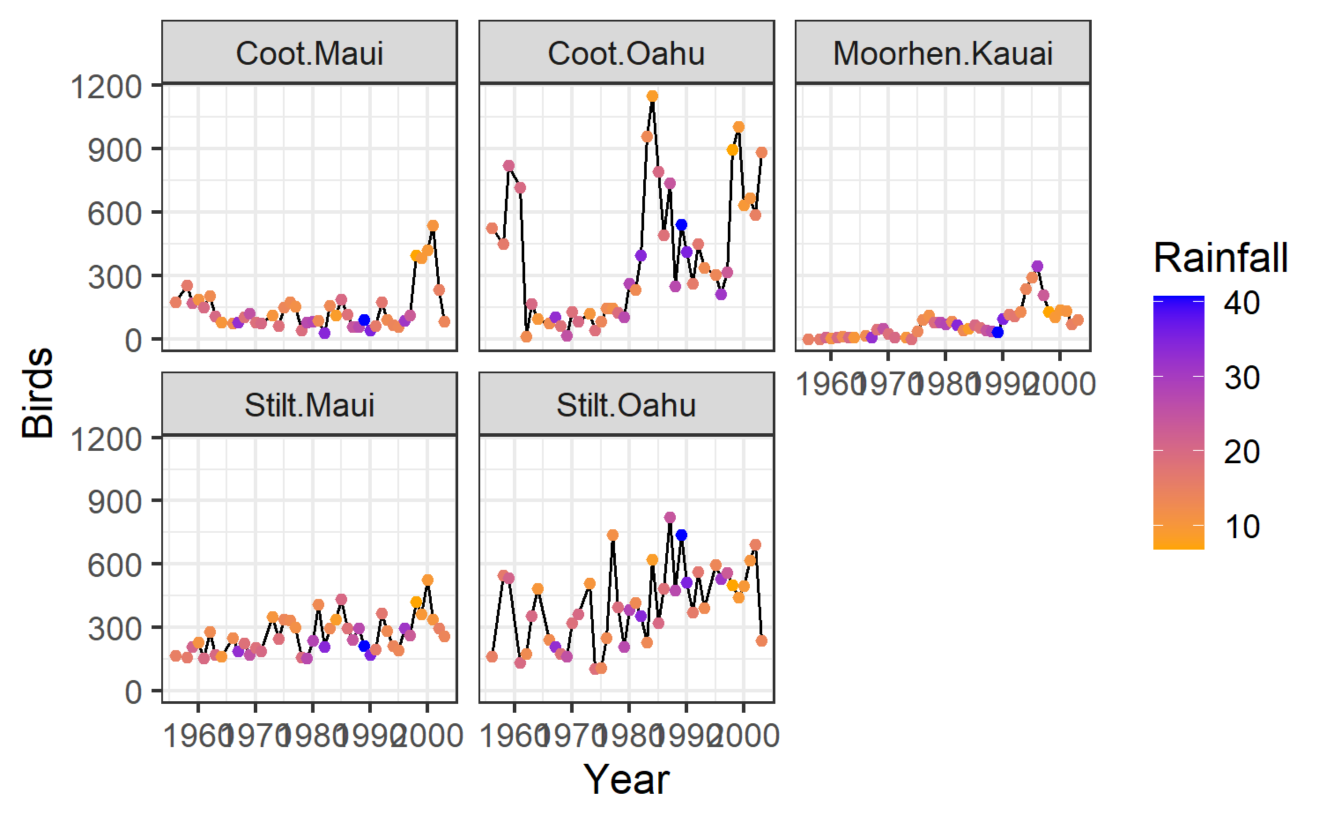
- Apply model without autocorrelation
- Check for autocorrelation
- Add autocorrelation structure

Other correlation structures

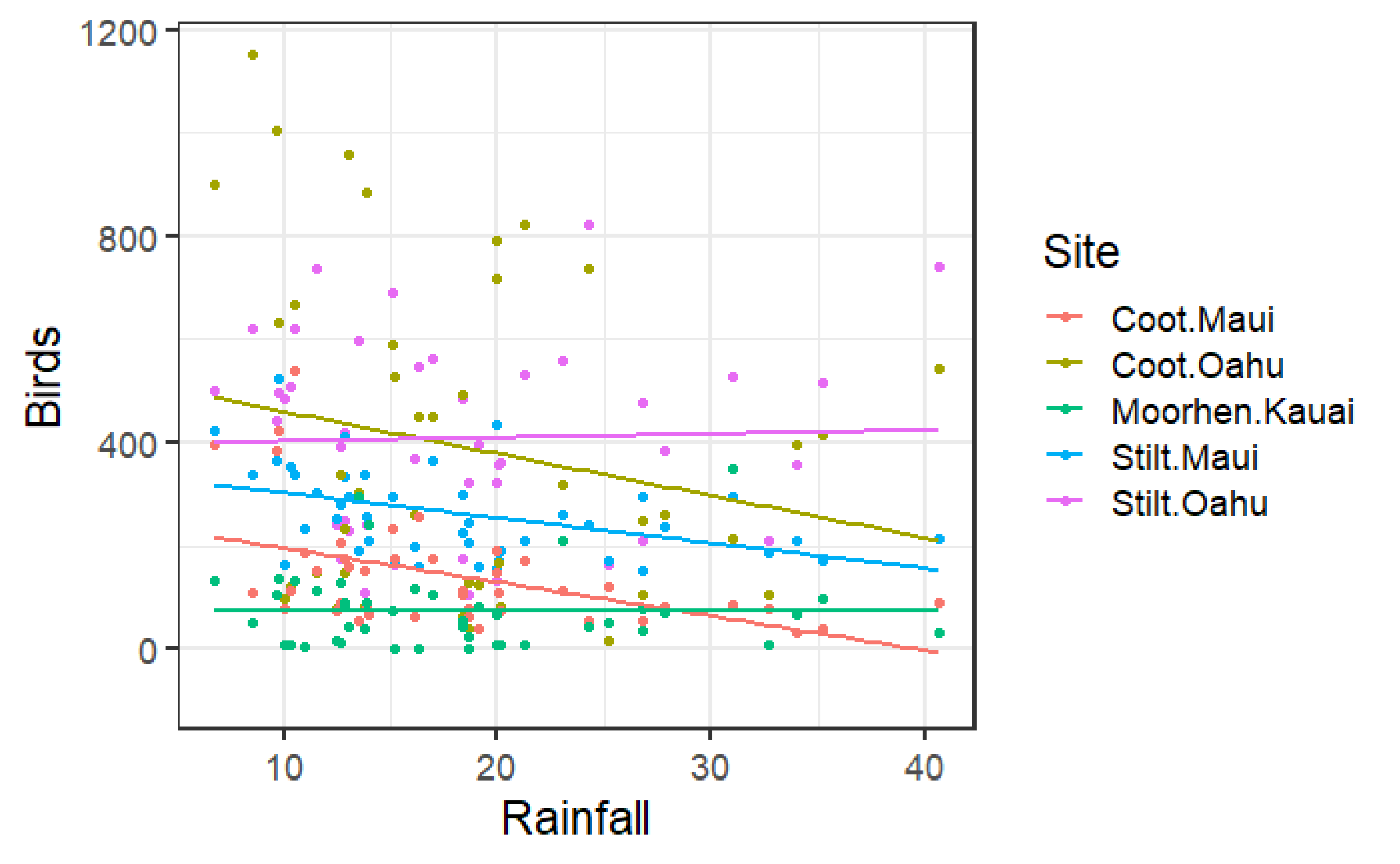
Modeling with many time series

Modeling with many time series (groups)

- Now let’s use the full Birds dataset (Hawaii.csv)



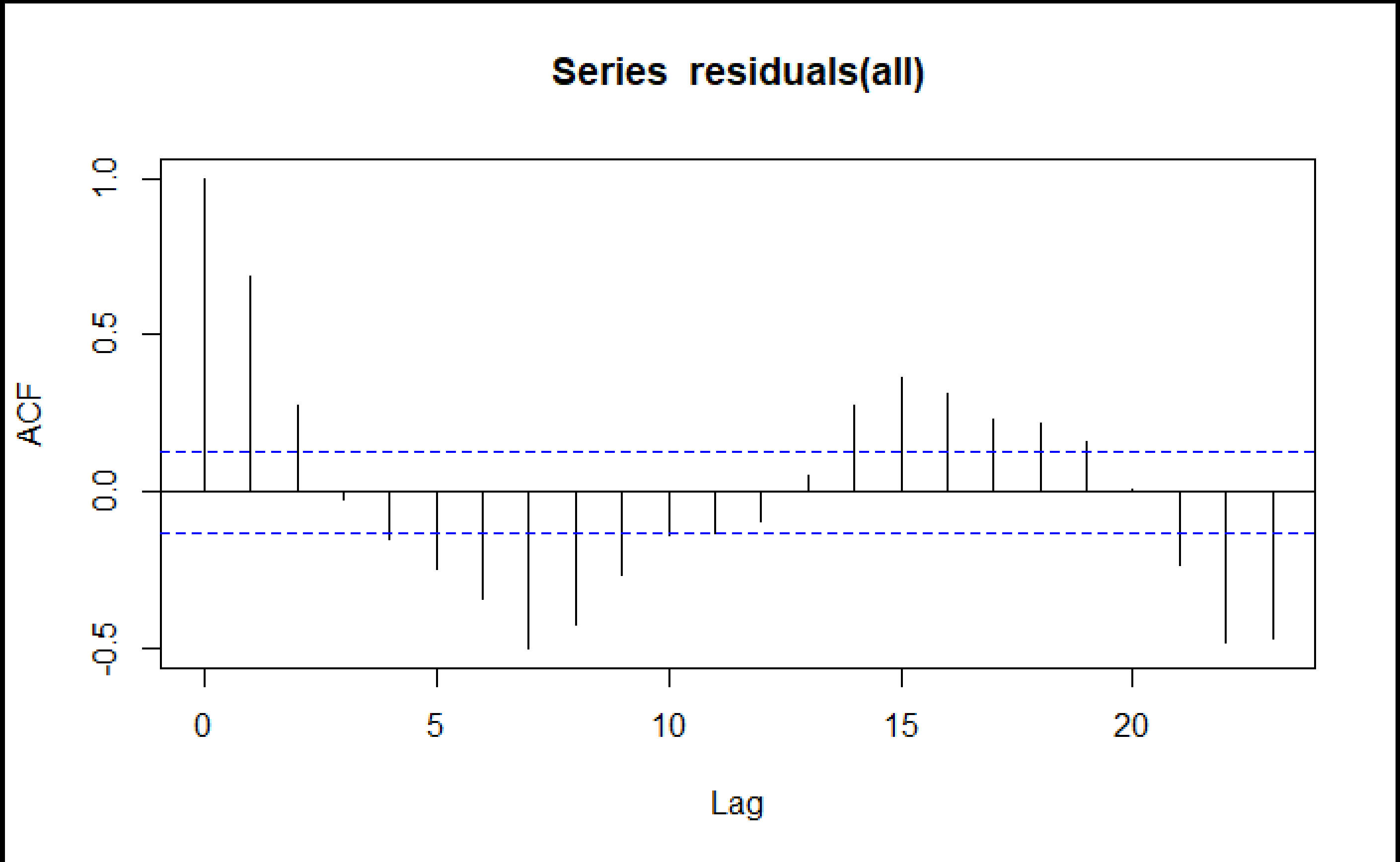
Adding groups – variable relationships?



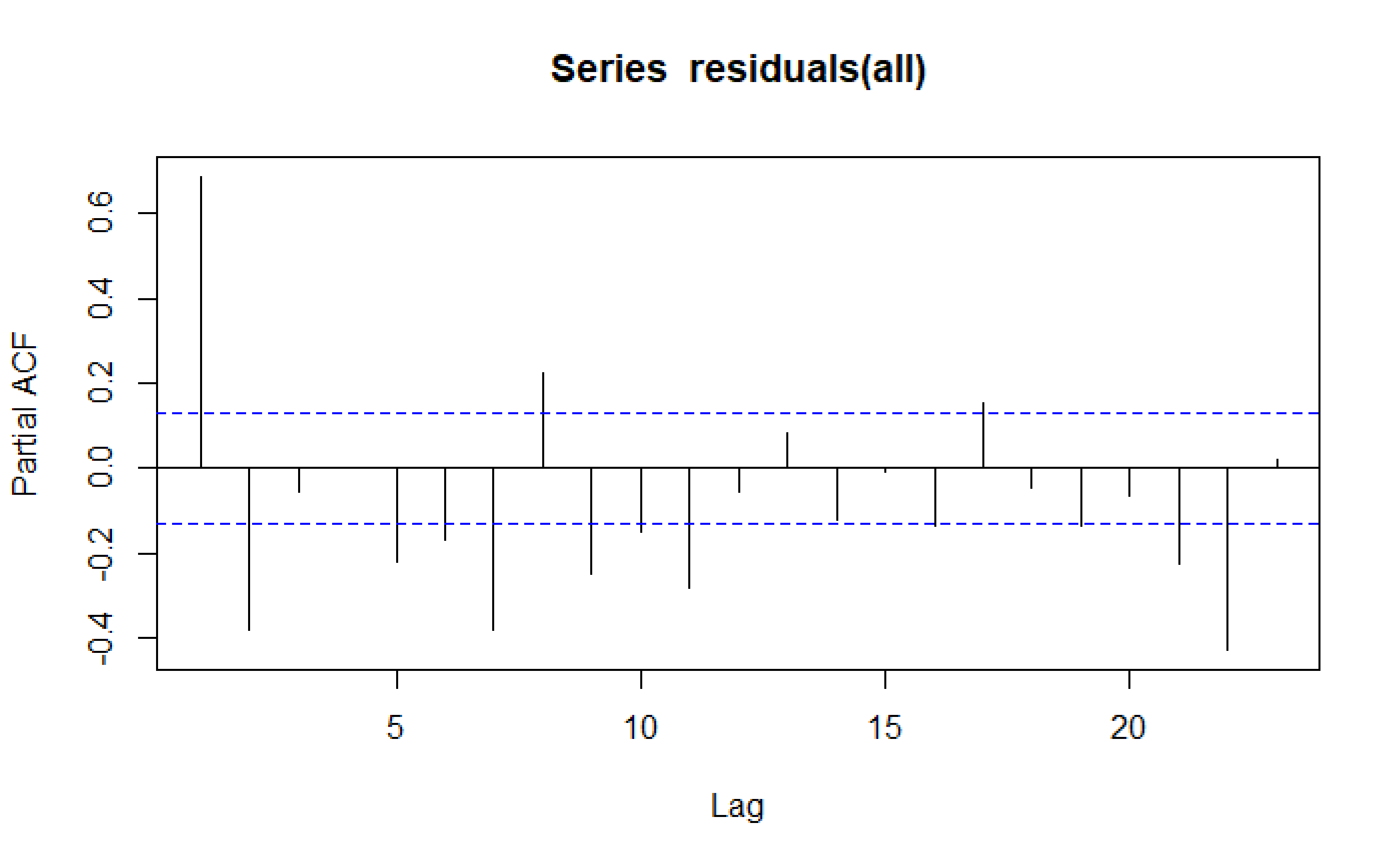
How do we accommodate for between-species heterogeneity?

```
birds_var <- varIdent(form = ~ 1 | Site)

all <- gls(Birds ~ Rainfall*Site + Year,
                data=Hawaii1,
                weights = birds_var,
                na.action=na.omit)

acf(residuals(all))
```

BUT there is still autocorrelation…



Fortunately, it is still probably AR1…

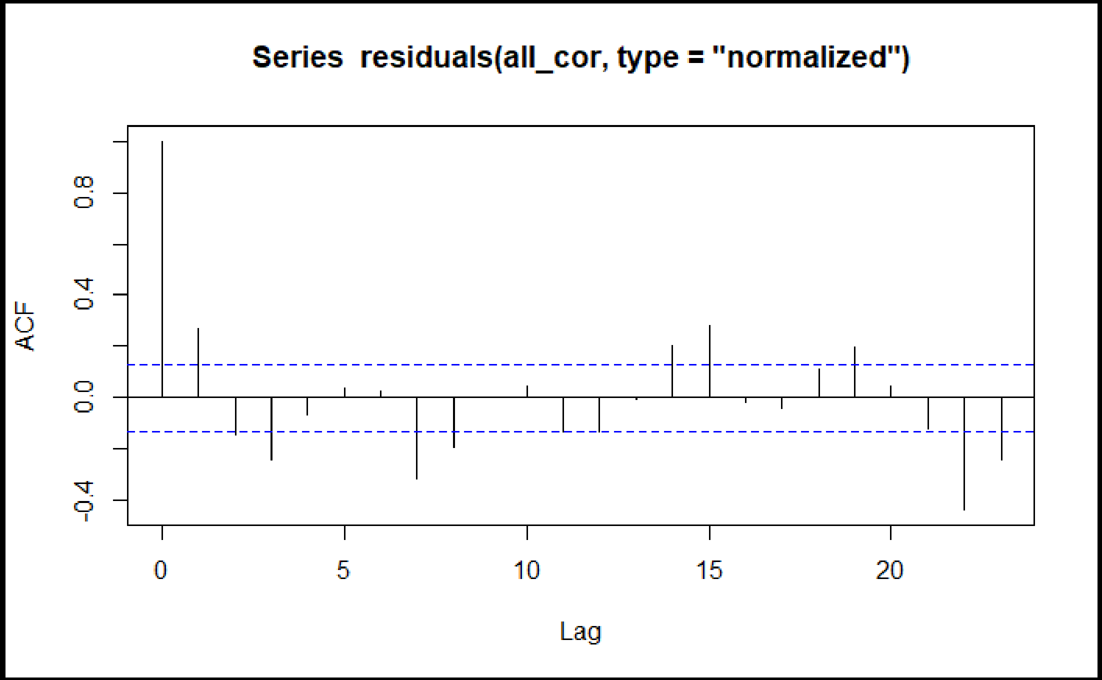


AR1 with groups

```
birds_corAR_site <- corAR1(form = ~ Year | Site)

all_cor <- gls(Birds ~ Rainfall*Site + Year,
                         data=allbirds,
                         weights = birds_var,
                         correlation = birds_corAR_site)

summary(all_cor)
```

Did we need the autocorrelation?

	Model <int>	df <dbl>	AIC <chr>	BIC <chr>	logLik <chr>	Test <fctr>	L.Ratio <chr>	p-value <chr>
all_var	1	16	1093.0613	1146.615	-530.5307			
all_cor	2	17	952.9634	1009.864	-459.4817	1 vs 2	142.0979	<.0001



Temporal autocorrelation: test yourself!

WHAT IS TEMPORAL AUTOCORRELATION, AND WHY IS IT IMPORTANT?

HOW CAN TEMPORAL AUTOCORRELATION BE DETECTED AND DIAGNOSED

WHAT METHODS CAN BE USED TO INCORPORATE TEMPORAL AUTOCORRELATION INTO MODELS?

Spatial autocorrelation - example

1. Fit a GLS
2. Check for spatial independence
   - bubble plot
   - variogram
3. Choose a correlation structure

Spatial autocorrelation - example

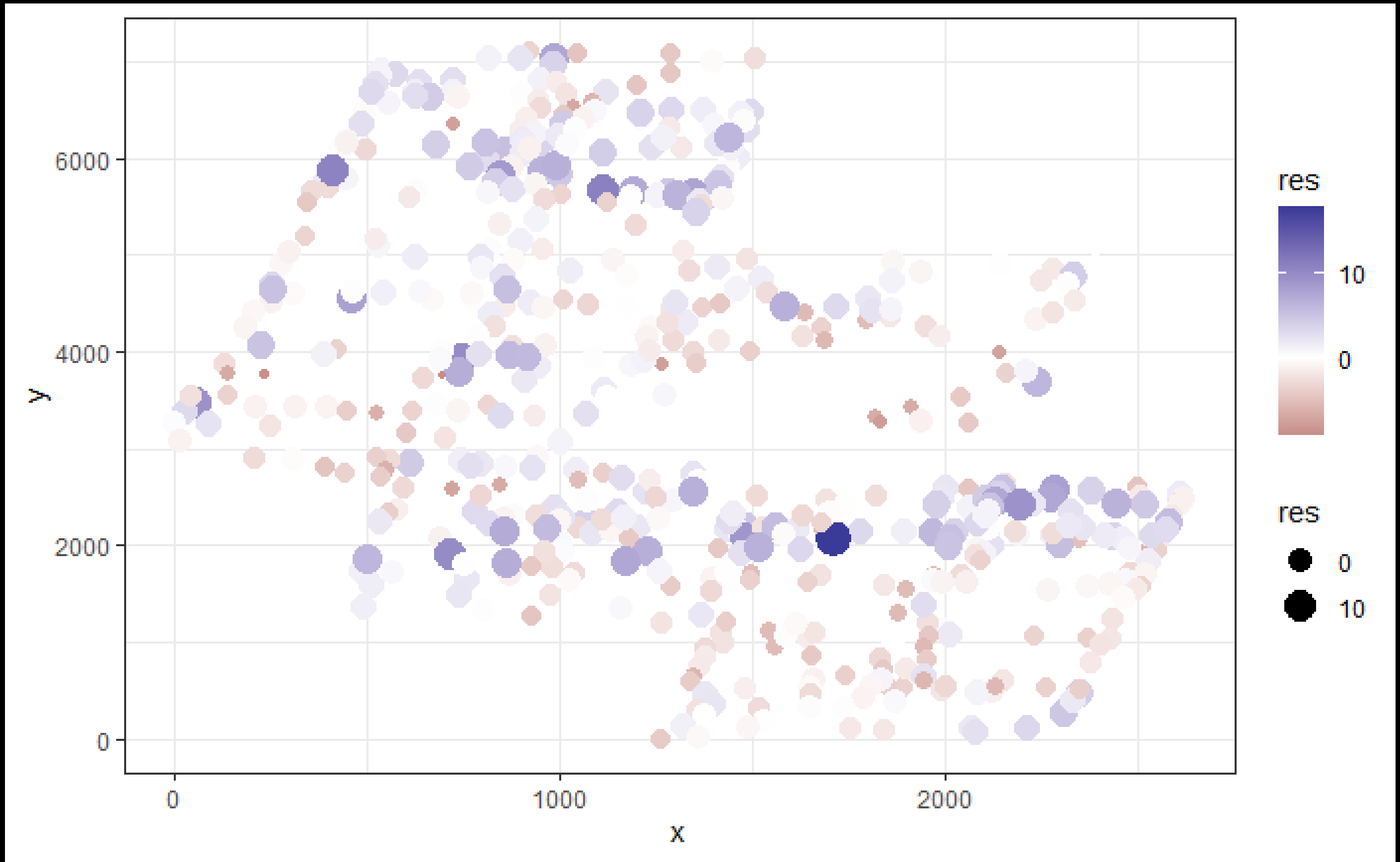
- Boreal forest in Tatarstan, Russia (Boreality.csv)
- Boreal forest species index as a function of wetness

1. Fit a GLS

```
B.lm<-gls(Bor~Wet,data=Boreality)
tab_model(B.lm, show.aic = TRUE, show.stat=TRUE, show.se=TRUE)
```

	Bor				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	18.49	0.38	17.74 – 19.23	48.82	**<0.001**
Wet	165.80	10.60	144.98 – 186.62	15.64	**<0.001**
Observations	533				
R^2	0.315				
AIC	2849.156				

2. Check for independence: bubbleplot

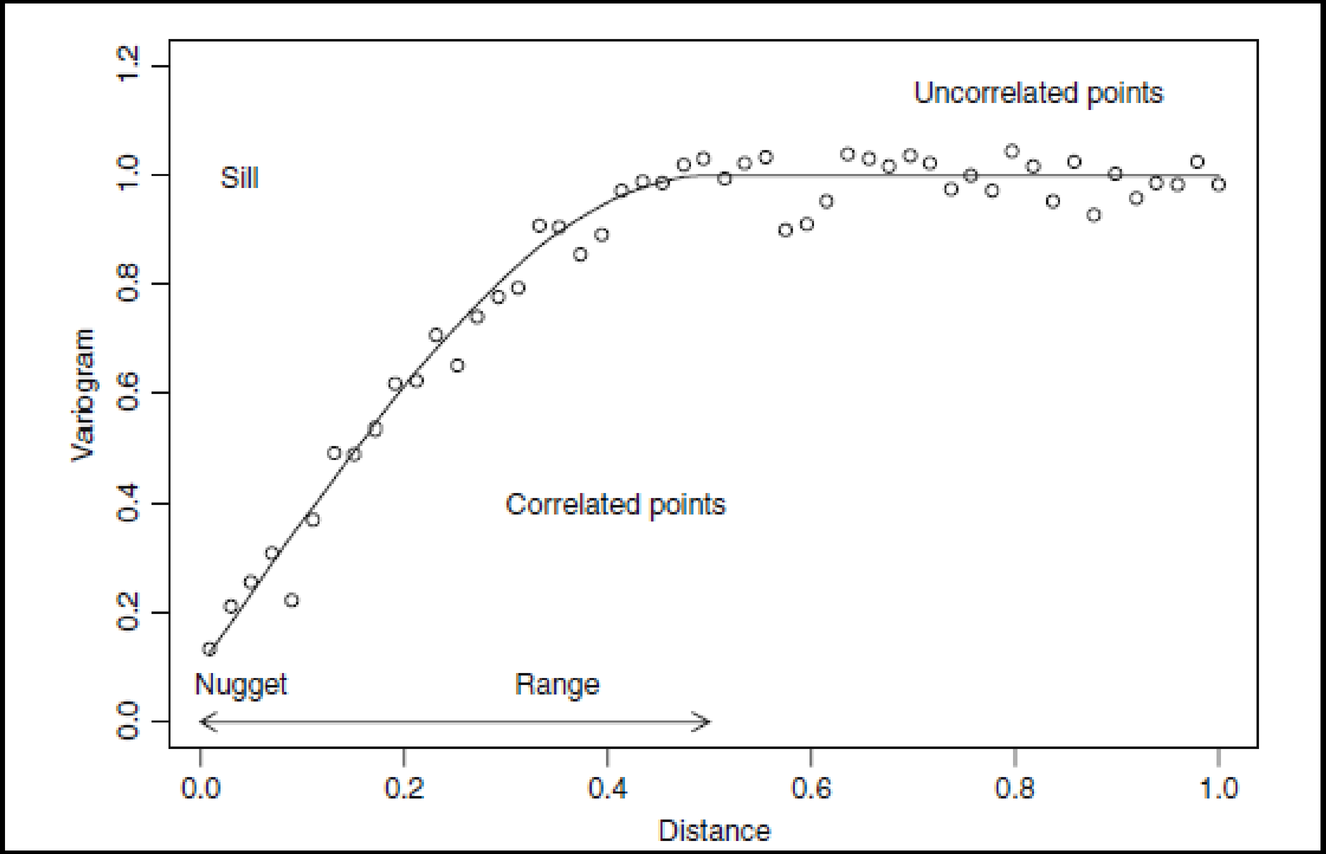



```r
Boreality$res <- as.numeric(residuals(B.lm))
ggplot(Boreality, aes(x, y, colour = res, size=res)) +
  geom_point() + scale_color_gradient2() +theme_bw()
```

A straight line is needed to show no spatial correlation but a pattern

2. Check for independence: variogram

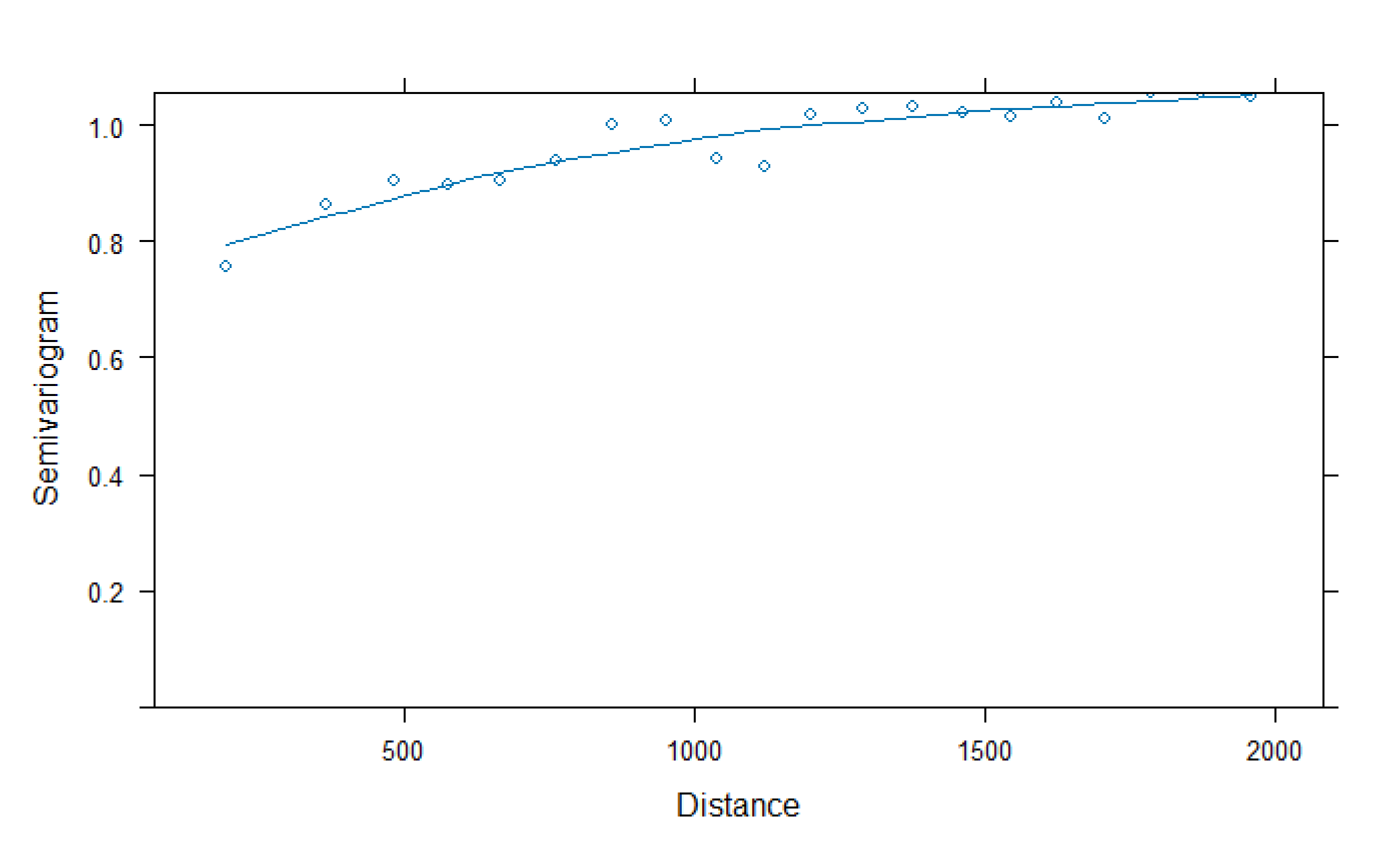
shows spatial correlation



The variogram is calculated by measuring the variability between pairs of observations at different distances and plotting these semi-variance values against distances

Small values = residuals are correlated, large values = uncorrelated

2. Check for independence: variogram




```
B1.gls<-gls(Bor ~ Wet, data = Boreality)
var1<-Variogram(B1.gls,form=~x+y,robust=TRUE,maxDist=2000,
             resType="pearson")
plot(var1)
```

3. Add correlation structure

corExp: correlation decreases exponentially with distance

corGaus: correlation decreases exponentially with the squared distance

corLin: correlation decreases linearly with distance

corRatio: exponential-like behavior at short distances and linear-like behavior at long distances

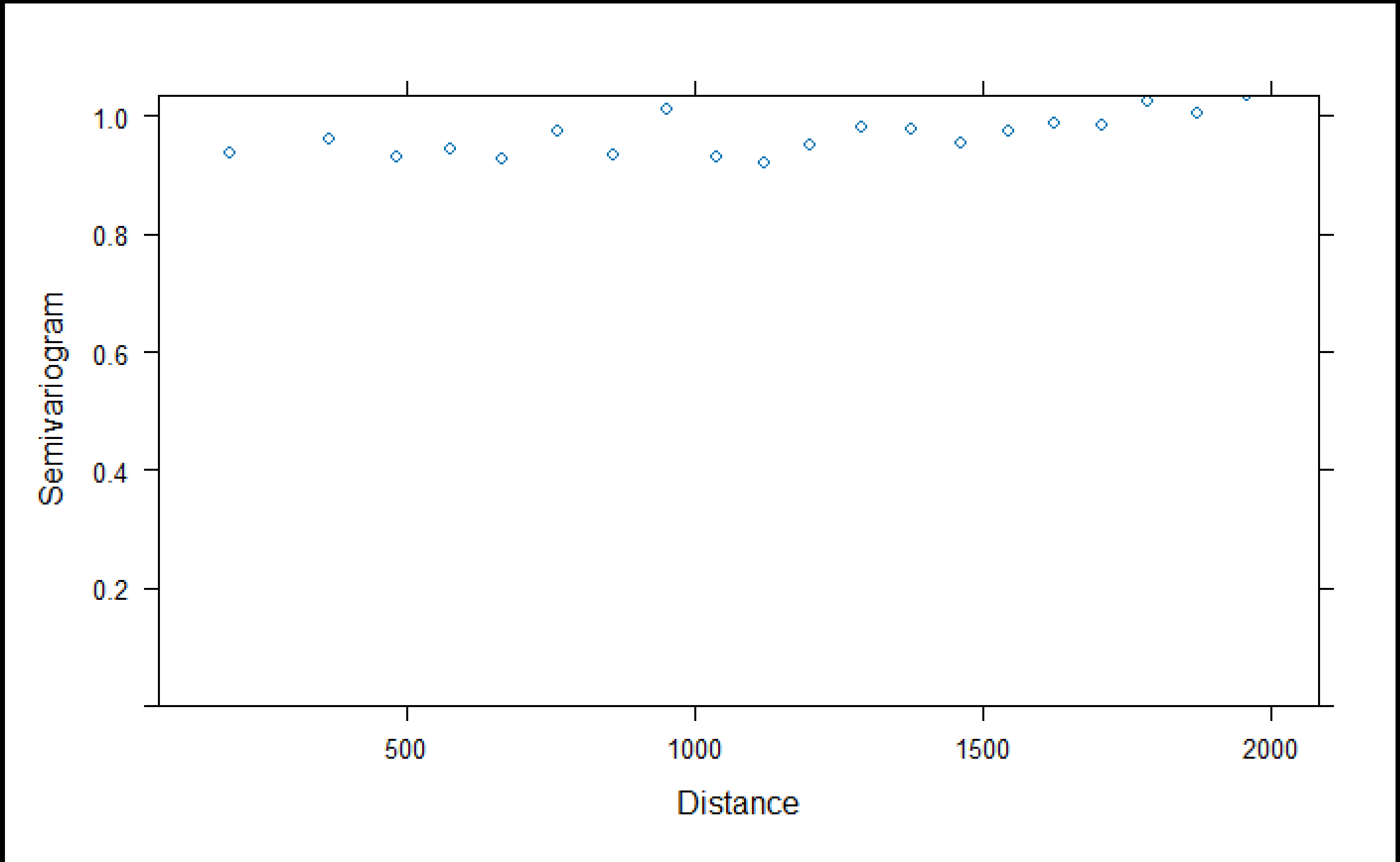
corSpher: correlation decreases with distance until it reaches a certain range, beyond which observations are uncorrelated.

3. Add correlation structure

```
B1A<-gls(Bor ~ Wet,correlation=corSpher(form=~x+y,nugget=T),data=Boreality)
B1B<-gls(Bor ~ Wet,correlation=corLin(form=~x+y,nugget=T),data=Boreality)
B1C<-gls(Bor ~ Wet,correlation=corRatio(form=~x+y,nugget=T),data=Boreality)
B1D<-gls(Bor ~ Wet,correlation=corGaus(form=~x+y,nugget=T),data=Boreality)
B1E<-gls(Bor ~ Wet,correlation=corExp(form=~x+y,nugget=T),data=Boreality)
AIC(B1.gls,B1A,B1B,B1C,B1D,B1E)
```

Model	Object	df	AIC
No correlation	B1	3	2844.54
corSpher	B1A	5	2737.01
corLin	B1B	5	2848.51
corRatio	B1C	5	2732.93
corGaus	B1D	5	2736.29
corExp	B1E	5	2732.22

4. Check for independence: again!



Autocorrelation summary

Temporal autocorrelation involves the correlation between observations of a variable at different time points, while spatial autocorrelation pertains to the correlation between observations at neighboring spatial locations

Plotting residuals versus spatial or temporal data, variograms, and autocorrelation plots aid in identifying and visualizing autocorrelation

The choice of correlation structure is less important than the choice to include it
- include something, even if it is not perfect!

Don’t forget to check model assumptions!