Time Series Midterm Review

David Josephs

2019-06-28

Contents

Chapter 1

Expected Values, Variance, etc.

1.1 Notation and Some Definitions

In time series, we denote the response variable as X_t, and clearly the explanatory variable as t. To denote a specific point in time, let us say $t = 9$, we would say X_9.

The collection of all X_n is a stochastic process known as a time series.

1.1.1 Realizations and Ensembles

A **realization** is a particular instance of a time series. It is one of the infinitely many possible time series that could have been observed. As a time series is a stochastic process, each one could be entirely different. The **realization** Is the one that was *actually* observed.

Sometimes a realization is the only realization possible. We can theorize what an infinite population of those realizations, but it is often the case that there is only one.

An **ensemble** is the theoretical totality of all possible realizations. It is the *ensemble* of *realizations*. We can think of that as the *population* of realizations. When we are thinking of means and variances, we think of them in terms of the ensemble, even though we may only be able to look at a single realization.

1.2 Population Means and Variances

It is important to note that when dealing with an ensemble of realizations, each time point is allowed to have its own mean and variance.

1.2.1 The Mean of X_t

μ_t is the mean of all possible realizations of X_t *for a fixed t*.

1.2.2 Variance of X_t

$\sigma_t{}^2$ is the variance of all possible realizations of X_t *for a fixed t*

1.3 Estimating Mean and Variance for Multiple Realizations

Let us say we have n realizations of X_t at time t. The average can be then written as

$\hat{\mu}_t = \frac{1}{n} \sum_{i=1}^{n} X_{t_n}$

Variance is calculated the same way. We are making sample means and sample variances. If we assume the population is normally distributed, this allows us to take confidence intervals on the mean of our set of realizations.

However, We have a problem: What to do when we only have one realization?

1.4   Expected Value

In short, the expected value $(E[X])$ of a random variable (RV) denoted by X is the **mean** or more intuitively the *long run average* of the event that variable represents.

1.4.1   Discrete RVs

A discrete RV is an RV in which X cannot take on any value, it has specific values it can exist at and that is it.

The formula for EV of a discrete RV is as follows:

$$E[X] = \sum XP(X) = \mu$$

1.4.2   Continuous RVs

A continuous RV has instead of discrete values a *probability distribution/density function,* $f(x)$.

The formula for EV of a continuous RV is as follows:

$$E[X] = \int_a^b xf(x)\,dx = \mu$$

Note that is directly analagous to the discrete RV, and that a and b can span to $\pm\infty$

Important Note

$$E[X] = \mu$$

That is the expected value of a RV is equivalent to the population mean

1.4.3   Some Rules

Let a and b be constants

$$E[a] = a$$

$$E[aX] = aE[X] = a\mu$$

$$E[aX + b] = aE[X] + E[b] = aE[x] + b = a\mu + b$$

Knowing these rules, lets try out a challenge:

1.4.4   A challenge in Expected Values

Find $E[\bar{x}]$ where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$

E.G find the expected value of an average

1.4.5 A solution

First, we can factor out the 1/n, so we have

$$\frac{1}{n} E\left[\sum_{i=1}^{n} x_i\right]$$

then we can say that the sum is just $x_1 + x_2 + x_3 + ... + x_n$. This would lead us to the conclusion that

$$E[\bar{x}] = \frac{1}{n} \sum_{i=1}^{n} E[x_i]$$

If we let each x_i be a constant then we have:

$$E[\bar{x}] = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

This makes sense as $\bar{x}$ is the average value of x, and so the expected value of x for a very high n is the average value of x

1.5 Variance and Covariance

1.5.1 Variance

The variance is the *dispersion*. Variance is defined as

$$\text{Var}(X) = E\left[(X - \mu)^2\right] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = \sigma^2$$

$$\widehat{\text{Var}(X)} = \sum (x_i - \bar{x})^2$$

1.5.2 Covariance

Assume we have two RVs, X and Y:

$$\text{Cov}(X, Y) = E\left[(X - \mu_x)(Y - \mu_y)\right]$$

$$\widehat{\text{Cov}(X, Y)} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

The variance is, in essence, the sum of cross products.

1.5.3 Some generalizations on covariance:

If y tends to increase with x, then $\text{Cov}(X, Y) > 0$

If Y tends to decrease with X, then $\text{Cov}(X, Y)$ is *Less than 0*

If it appears as a random cloud, then $\text{Cov}(X, Y)$ is approximately 0

1.5.4 A challenge

Let X be a RV. What is $\text{Cov}(X, X)$?

1.5.5   A solution

$$\mathrm{Cov}(X, X) = E\left[(X - \mu_x)(X - \mu_X)\right]$$

We see that the arguments in the left and right parens are exactly the same. Thus, we can rewrite the equation as:

$$\mathrm{Cov}(X, X) = E\left[(X - \mu_x)^2\right] = \mathrm{Var}(X)$$

1.5.6   A note on covariance and correlation

Correlation is covariance but standardized.

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X)\mathrm{SD}(Y)}$$

$$= \frac{E[X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y}$$

$$\widehat{\mathrm{Corr}(X, Y)} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y}$$

Chapter 2

A Brief Discussion of Stationarity

In order for a time series to be considered stationary, it must satisfy *three* conditions:

- Constant Mean with Time

- Constant Variance with Time

- Constant Autocorellation with Time

2.1   Condition One: Constant Mean

Mean is constant with time. That is $E[X_t] = \mu$. Note the lack of a little tiny t in mu! It is indepoent of tine! That is stationarity condition number 1!

Important result: **If we assume constant mean, we can use *all* of the data to estimate the mean**

Here we can see after 5 realizations that the mean is clearly not constant with time

```
library(tswge)
a1 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(3,
    0), vara = 3, plot = F)
b1 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(3,
    0), vara = 3, plot = F)
c1 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(3,
    0), vara = 3, plot = F)
d1 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(3,
    0), vara = 3, plot = F)
e1 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(3,
    0), vara = 3, plot = F)
```
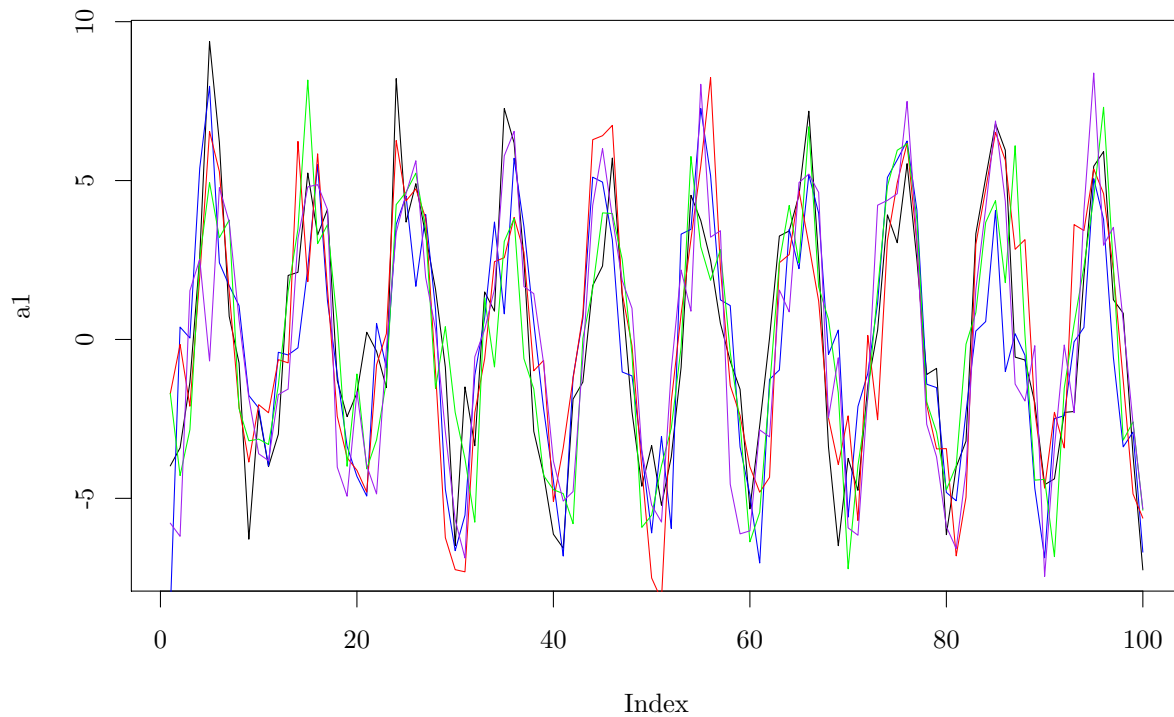
Now lets look at 5 realizations which come from the same realization, which might not have a constant mean:

```
a2 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(runif(1,
    0, 2 * pi), 0), vara = 3, plot = F)
b2 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(runif(1,
    0, 2 * pi), 0), vara = 3, plot = F)
c2 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(runif(1,
    0, 2 * pi), 0), vara = 3, plot = F)
d2 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(runif(1,
    0, 2 * pi), 0), vara = 3, plot = F)
e2 = gen.sigplusnoise.wge(100, coef = c(5, 0), freq = c(0.1, 0), psi = c(runif(1,
    0, 2 * pi), 0), vara = 3, plot = F)
```
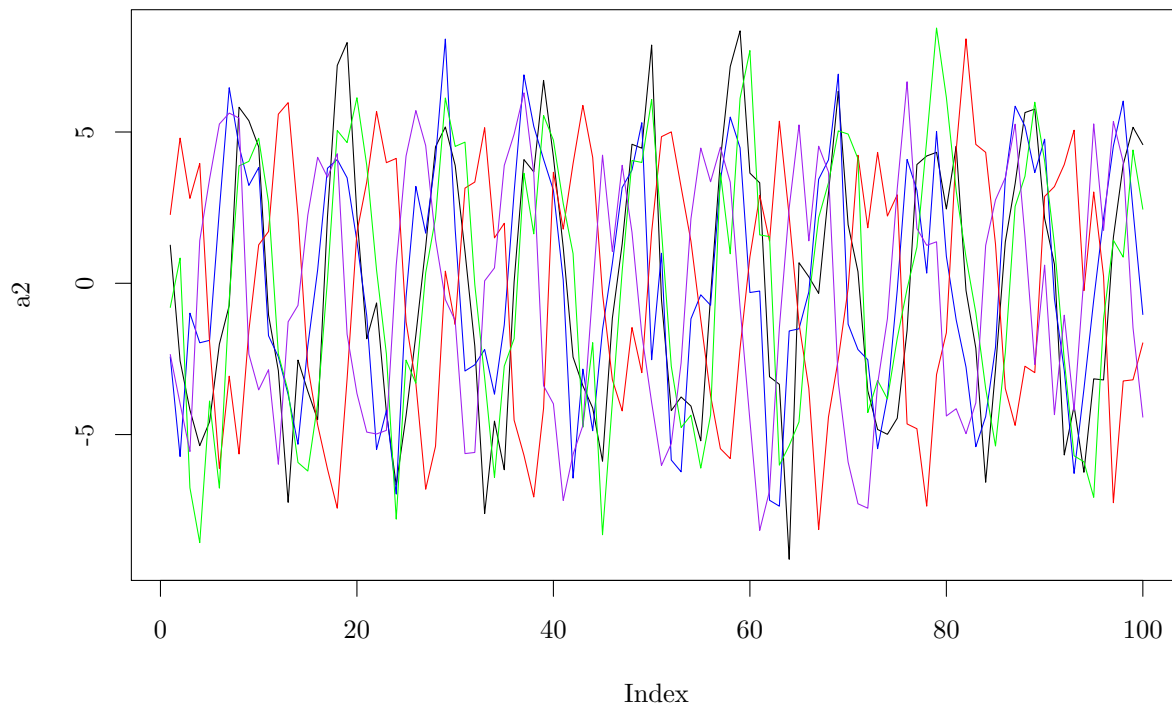
We cannot say for sure whether the mean is constant or not in this case!

2.2 Condition Two: Constant Variance

Variance does not depend on time

Variance is *constant* and finite*

That is :

$$Var[X_t] = \sigma^2 \neq \infty$$

If we can make this assumption, then we can use all of the data to make the variance. This is typically the hardest one to tell, and it will take some practice to see.

2.3 Condition 3: Constant autocorrelation

Correlation of $X_{t_1} and X_{t_2}$ ***only depends on*** $t_2 - t_1$

That is, the correlation between points depends only on how far apart they are in time, not where they are in time.

To describe this mathematically:

Let

$$h = t_2 - t_1$$

then

$$\mathrm{Cor}\left(X_t, X_{t+h}\right) = \rho_h$$

ρ represents the population correlation coefficient.

Chapter 3

Autocorrelation

```r
library(tswge)
library(ggplot2)
acfdf <- function(vec) {
    vacf <- acf(vec, plot = F)
    with(vacf, data.frame(lag, acf))
}
ggacf <- function(vec) {
    ac <- acfdf(vec)
    ggplot(data = ac, aes(x = lag, y = acf)) + geom_hline(aes(yintercept = 0)) +
        geom_segment(mapping = aes(xend = lag, yend = 0))
}

tplot <- function(vec) {
    df <- data.frame(X = vec, t = seq_along(vec))
    ggplot(data = df, aes(x = t, y = X)) + geom_point() + geom_line()
}
```

3.1 Independence

Two events are independent in a time series if the probability that an event at time t occurs in no way depends on the ocurrence of any event in the past or affects any event in the future. Mathematically, this is written as:

$$\text{Independence}: P\left(x_{t+1}|X_t\right) = P\left(X_{t+1}\right)$$

3.2 Serial Dependence / Autocorrelation

3.2.1 A definition

If two events are independent, their corellation is 0 That is if X_t and X_{t+k} are independent, $\rho_{x_t, x_{t+k}} = 0$

*Corollary: If the correlation between two variables is not zero, then they are not **independent***

In other words if $\rho_{x_t, x_{t+k}} \neq 0$, they are **not** independent.

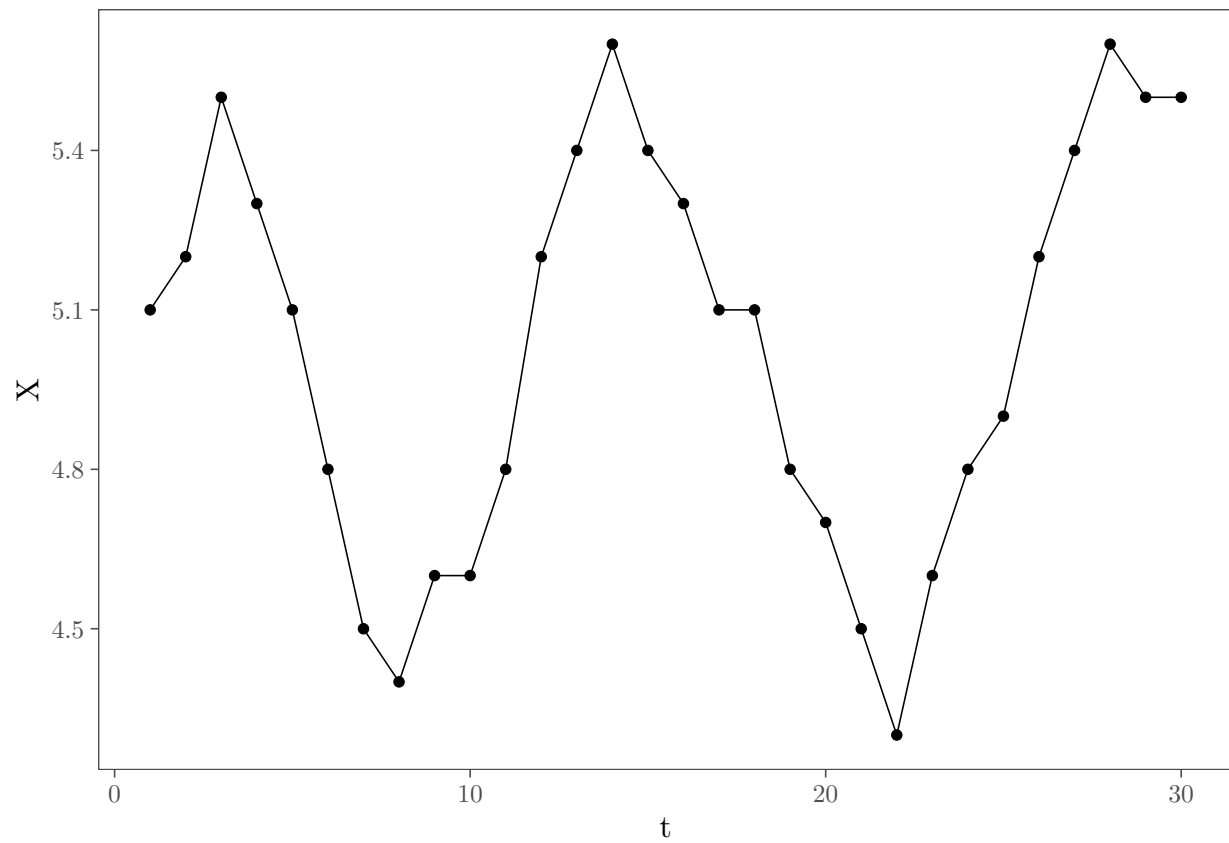### 3.2.2   Autocorrelation Plots

3.2.3   Dependent(ish) data

In time series we look at the **autocorrelation** between X_t and X_{t+1} etc (with t and t+1 it is lag 1 autocorrelation)

For example, visually, let us define a vector, $Y5$ and take its autocorrelation:

```
Y5 <- c(5.1, 5.2, 5.5, 5.3, 5.1, 4.8, 4.5, 4.4, 4.6, 4.6, 4.8, 5.2, 5.4, 5.6,
    5.4, 5.3, 5.1, 5.1, 4.8, 4.7, 4.5, 4.3, 4.6, 4.8, 4.9, 5.2, 5.4, 5.6, 5.5,
    5.5)
knitr::kable(acfdf(Y5))
```

lag	acf
0	1.0000000
1	0.8268482
2	0.5166514
3	0.1125421
4	-0.2687774
5	-0.5538901
6	-0.6935454
7	-0.6979912
8	-0.5505557
9	-0.2752524
10	-0.0056082
11	0.2274804
12	0.3902111
13	0.4573468
14	0.4660549

```
tplot(Y5) + ggthemes::theme_few()
```

```r
ggacf(Y5) + ggthemes::theme_few()
```
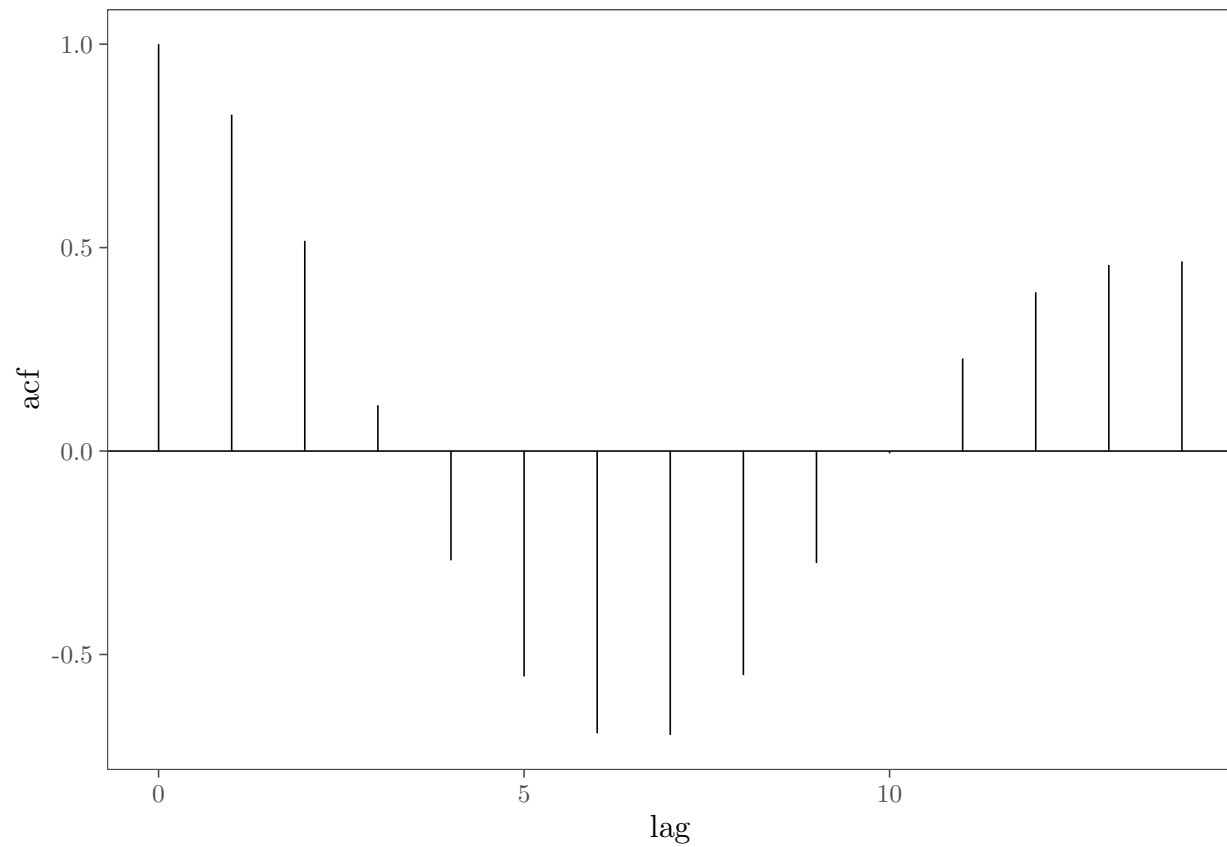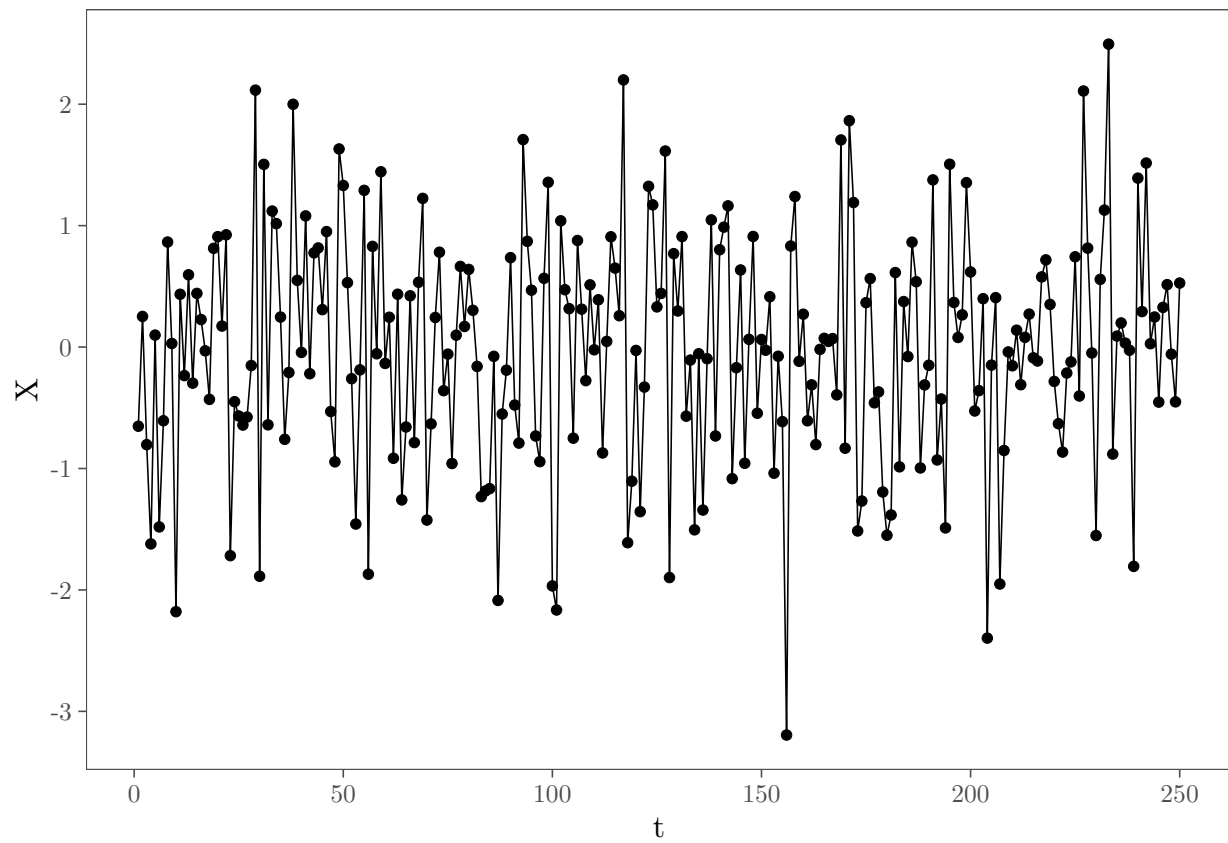
3.2.4   Independent(ish) data!

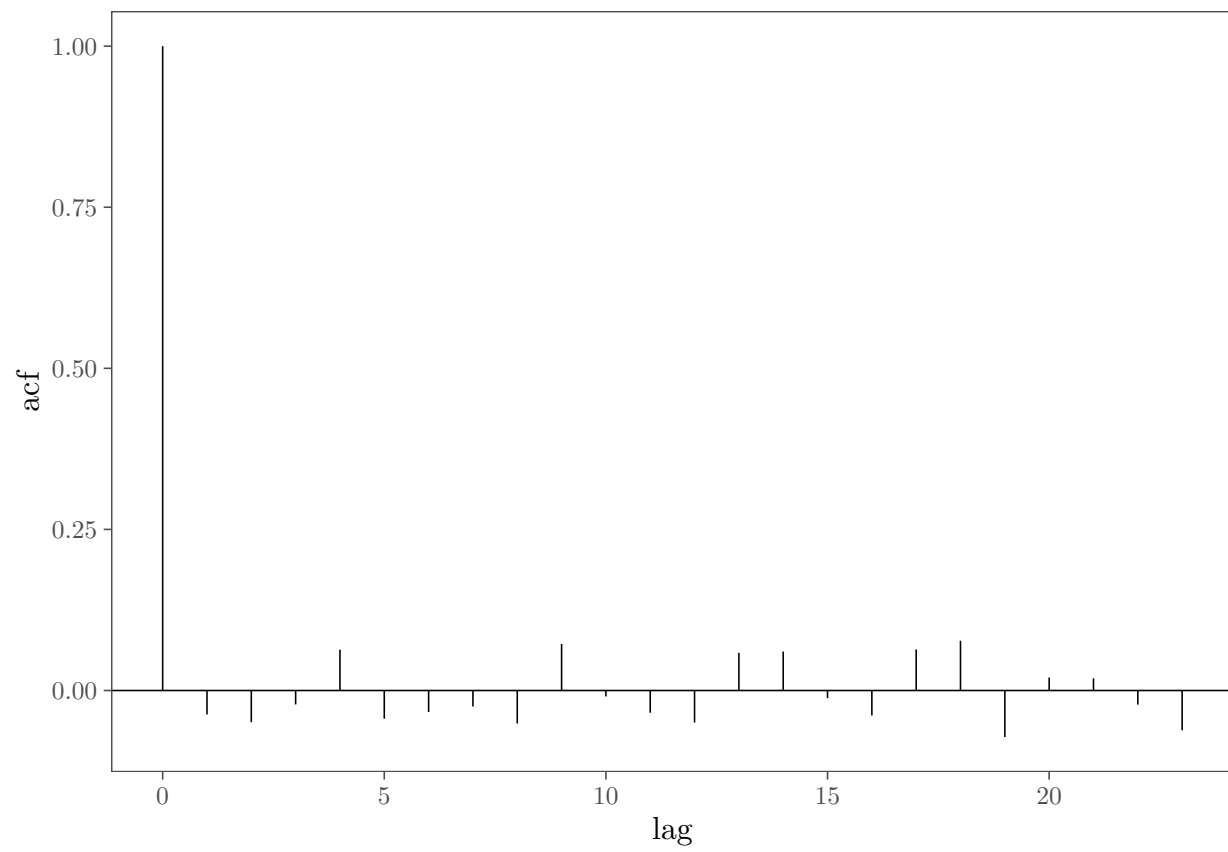Now let us look at autocorrelation of *independent* data

```
xs = gen.arma.wge(n = 250)
```

Let's check it out

```
tplot(xs) + ggthemes::theme_few()
```

```r
ggacf(xs) + ggthemes::theme_few()
```

We see that the autocorrelation is more or less zero

Chapter 4

Autocorrelation Concepts and Notation

4.1 Theoretical concept of ρ

$$\rho = \frac{E\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sigma_X \sigma_Y} = \frac{\mathrm{covariance\,between\,X+y}}{\sigma_X \sigma_Y}$$

In a stationary time series, we have autocorrelation and autocovariance

4.2 autocovariance

$$\gamma_k = E\left[(X_t - \mu)(X_{t+k} - \mu)\right]$$

4.3 autocorrelation

$$\rho_k = \frac{E\left[(X_t - \mu)(X_{t+k} - \mu)\right]}{\sigma_{X_t}\sigma_{X_{t+k}}} = \frac{E\left[(X_t - \mu)(X_{t+k} - \mu)\right]}{\sigma_X^2} = \frac{\gamma_k}{\sigma_X^2}$$

This works because of constant variance and constant mean!

Note that in a stationary time series:

$$\sigma_X^2 = 1 E[(X_t - \mu)^2] = E[(X_t - \mu)(X_t - \mu)] = \gamma_0$$

Therefore:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

4.4 Stationary Covariance

Correlation is not affected by where, only by how far apart, that is:

$$\rho_h = \mathrm{Cor}\left(X_t, X_{t+h}\right)$$

Let us try it out with this little snippet:

```r
ggsplitacf <- function(vec) {
    h1 <- vec[1:(length(vec)/2)]
    h2 <- vec[(length(vec)/2):length(vec)]
    first <- ggacf(vec) + ggthemes::theme_few()
    second <- ggacf(h1) + ggthemes::theme_few()
    third <- ggacf(h2) + ggthemes::theme_few()
    cowplot::plot_grid(first, second, third, labels = c("original", "first half",
        "second half"), nrow = 2, align = "v")
}
```

```r
Realize = gen.arma.wge(500, 0.95, 0, plot = T, sn = 784)
```
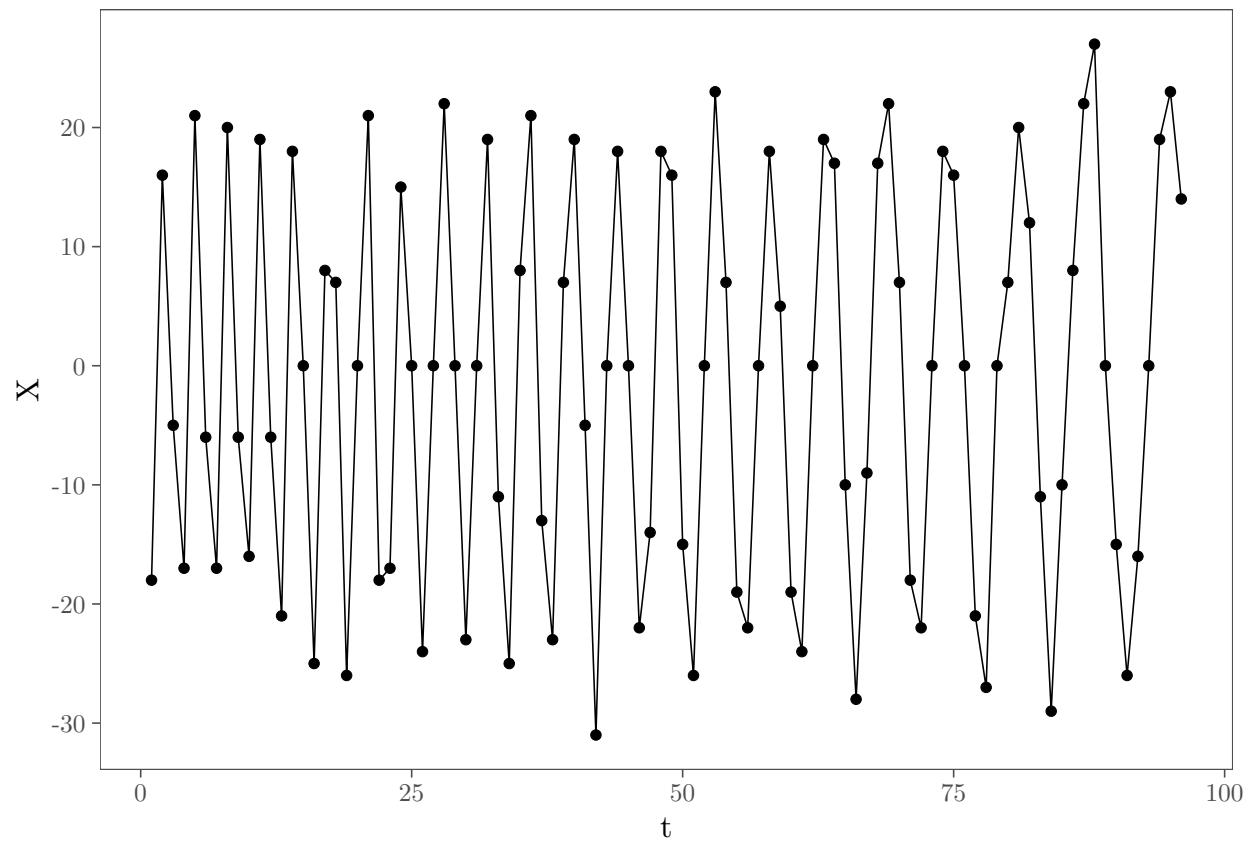
```r
ggsplitacf(Realize)
```
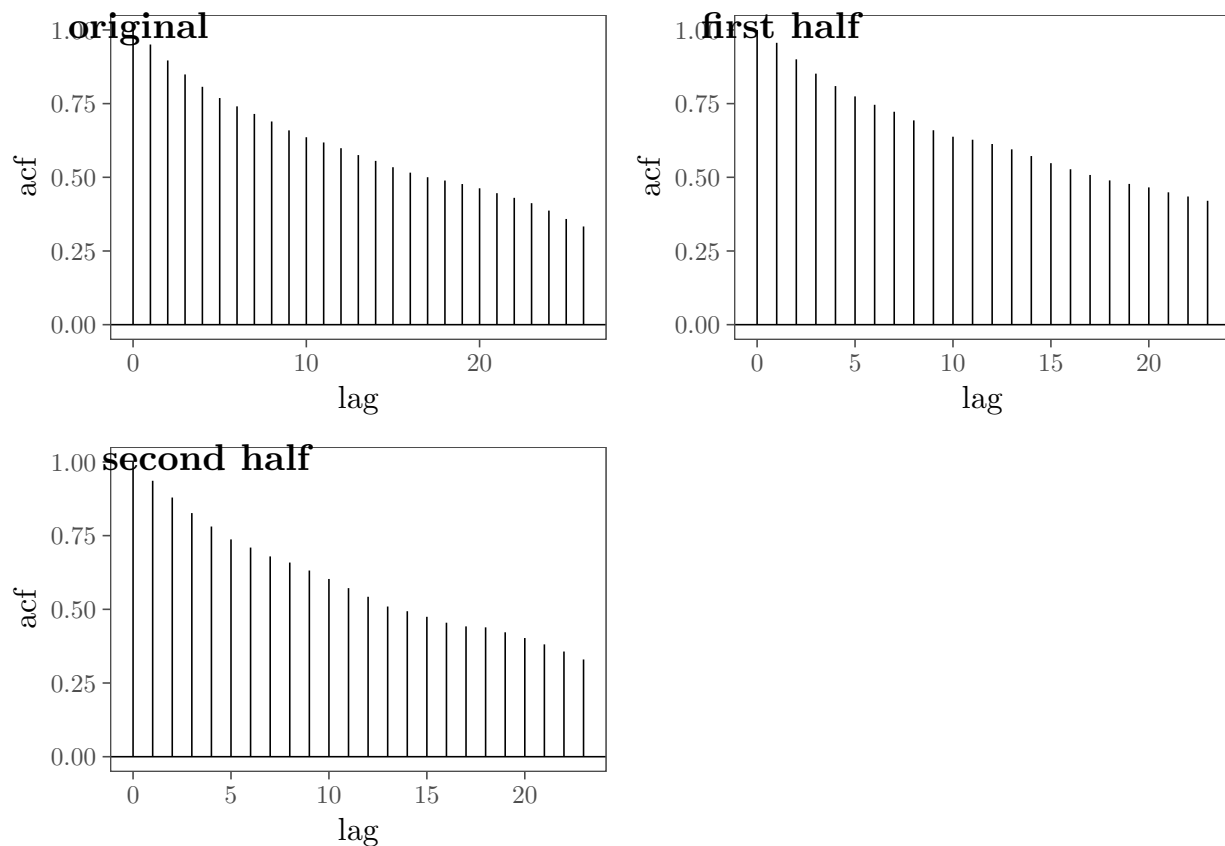

This looks pretty stationary to me! Let's try with some sample data:
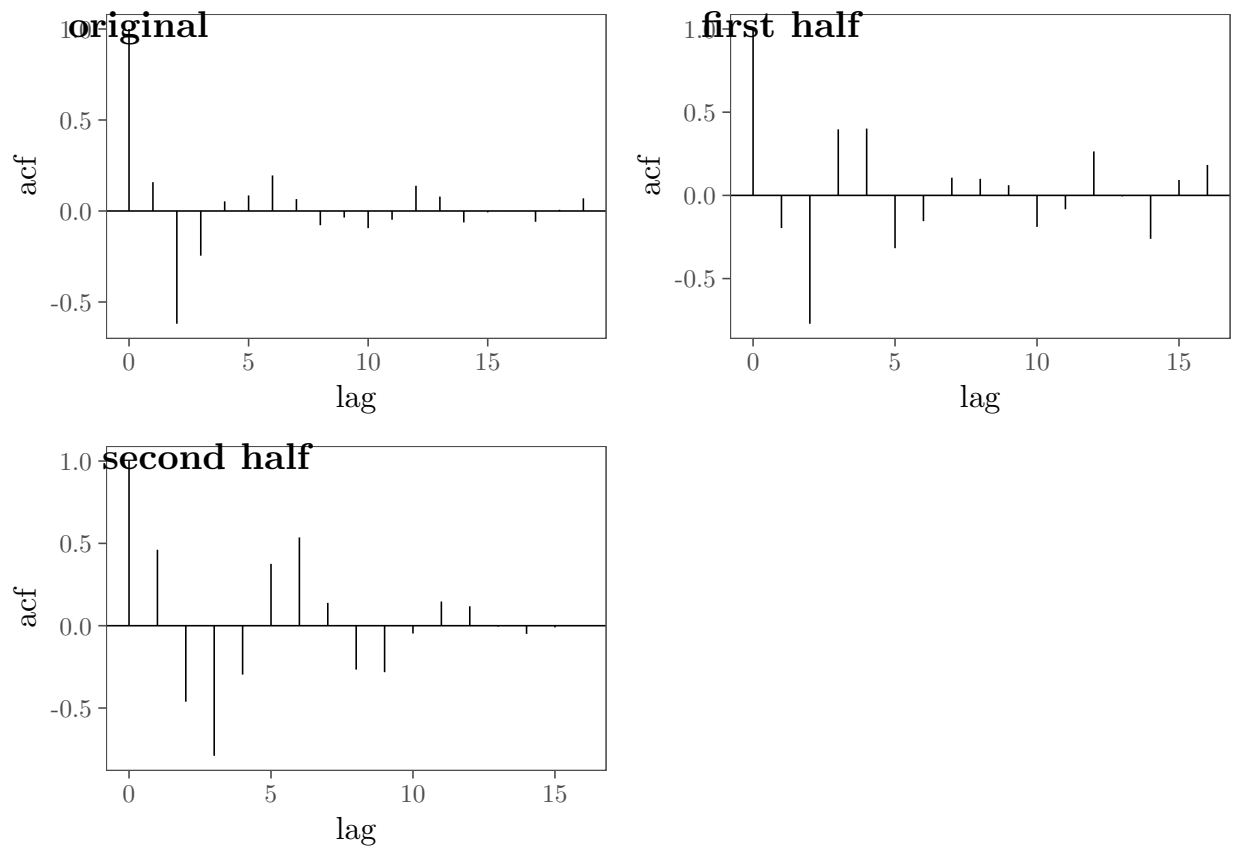
```r
data("noctula")
tplot(noctula) + ggthemes::theme_few()
```
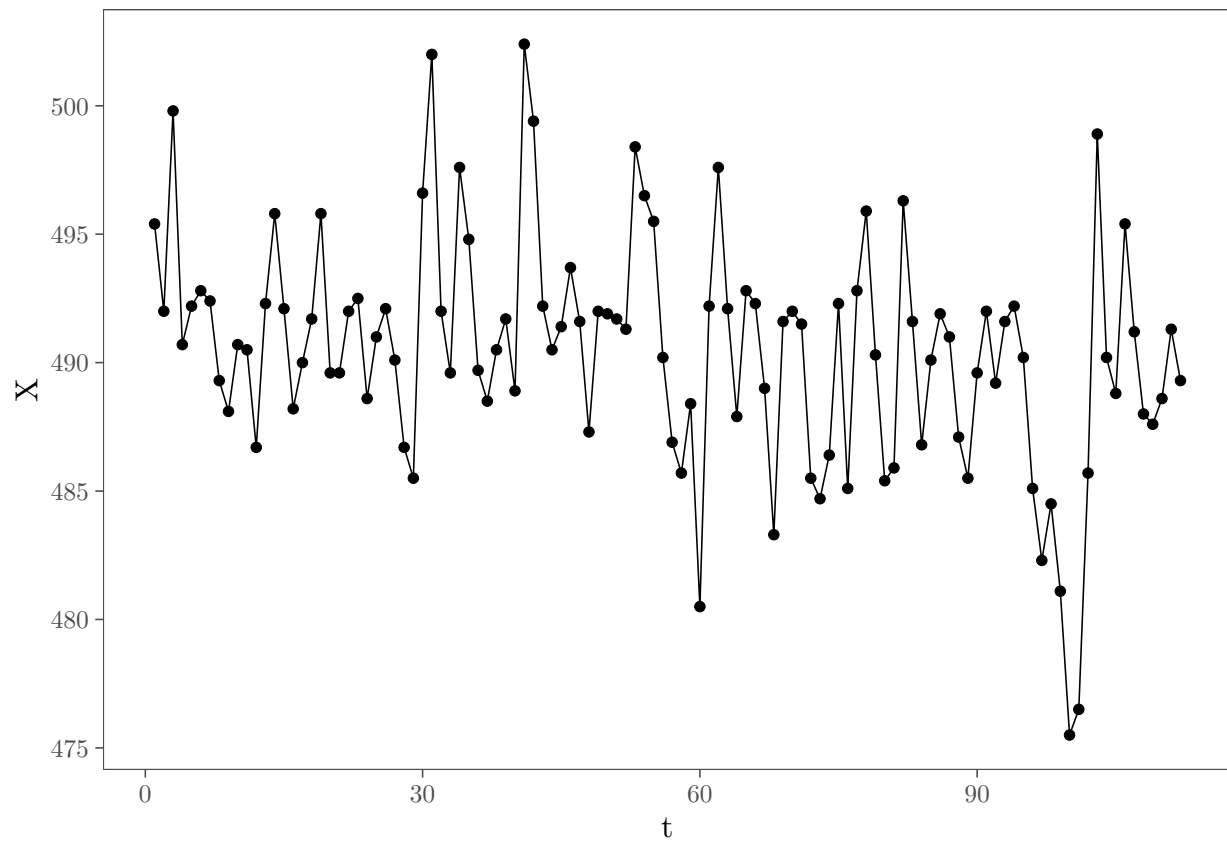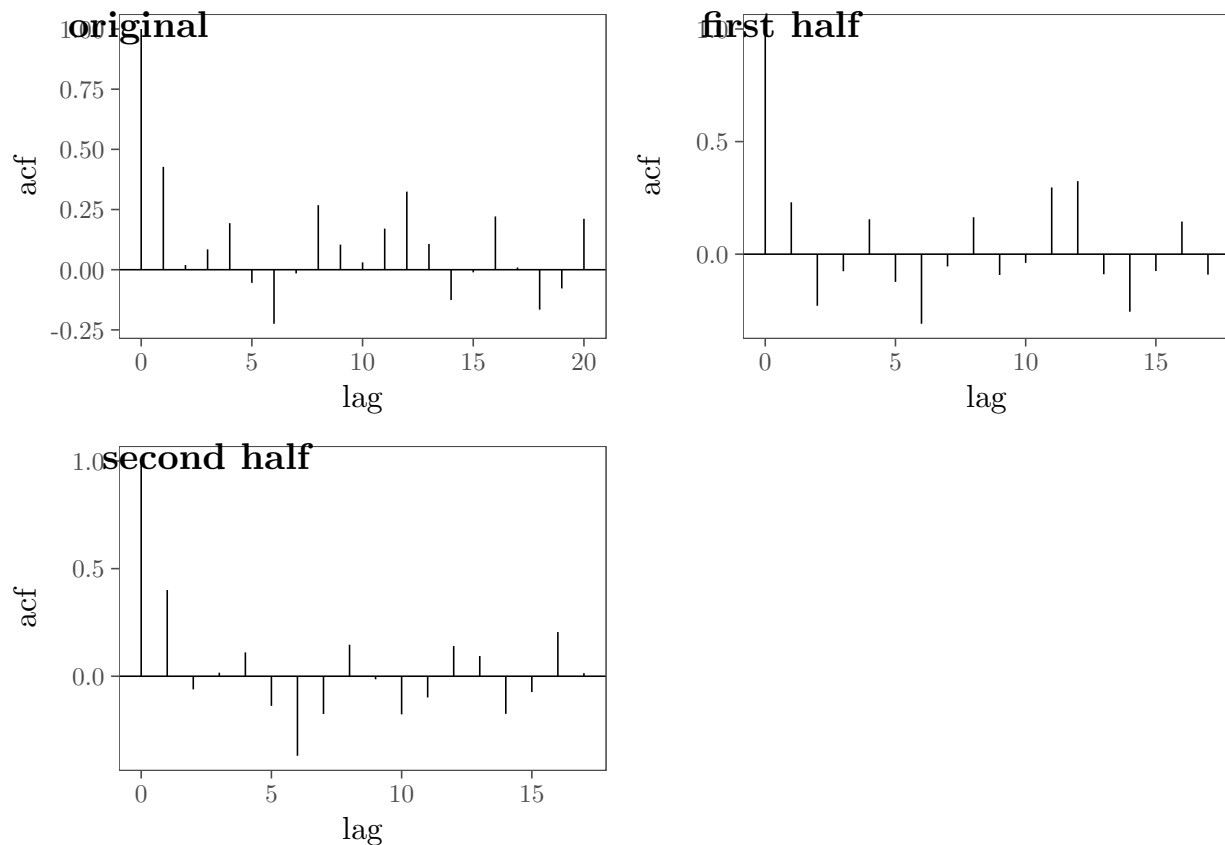
```
ggsplitacf(noctula)
```

So in this case our mean probably looks constant, our variance could be constant as well, however the ACFs are clearly different. Let's look at a different dataset:

```
data(lavon)
tplot(lavon) + ggthemes::theme_few()
```

```
ggsplitacf(lavon)
```

In this case, our mean looks maybe to be constant around 495, our variance does not appear to be constant, and our ACF looks pretty good, however I would say overall this probably is not stationary, due to the variance and slightly off ACF.

4.5 Stationary Time Series: Putting it Together

4.5.1 Workflow:

- Step 1: Check condition 1
    - see if the assumption is more or less met
- Step 2: Check condition two
    - same thing
- Step 3: Check for condition three

If one of the things is iffy, and the other two arent, we are going to be just fine.

Chapter 5

Practical Autocorrelation

5.1 Estimation!

5.1.1 $\rho_k \to 0$

For a stationary time series, if the autocorrelation approaches zero, then :

A Single realization lets us estimate mean, variance, and autovvariance!

Remember that $\rho_k = \frac{\gamma_k}{\gamma_0}$, and we use n-k pairs to calculate it (the summation)

5.1.1.1 Mean

Just calculate the mean normally for this case.

5.1.1.2 Variance

$$\text{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{\mid k \mid}{n}\right) \rho_k$$

σ^2 is calculated as normal, we will see rhok next!

remember !

Now it is time for some code!

```
library(glue)
```

```
##
## Attaching package: 'glue'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
xbar <- function(xs) {
    mean(xs)
}

ghat_zero <- function(xs) {
    summand <- (xs - xbar(xs))^2
    mean(summand)
```

```r
}

ghat_one <- function(xs) {
    lhs <- xs[1:(length(xs) - 1)] - xbar(xs)
    rhs <- xs[2:length(xs)] - xbar(xs)
    summand <- lhs * rhs
    summate <- sum(summand)
    summate/length(xs)

}
rhohat_zero <- 1
rhohat_one <- function(xs) {
    ghat_one(xs)/ghat_zero(xs)
}
v <- c(76, 70, 66, 60, 70, 72, 76, 80)
xbar(v)
```

```
## [1] 71.25
```

```r
ghat_zero(v)
```

```
## [1] 34.9375
```

```r
ghat_one(v)
```

```
## [1] 14.74219
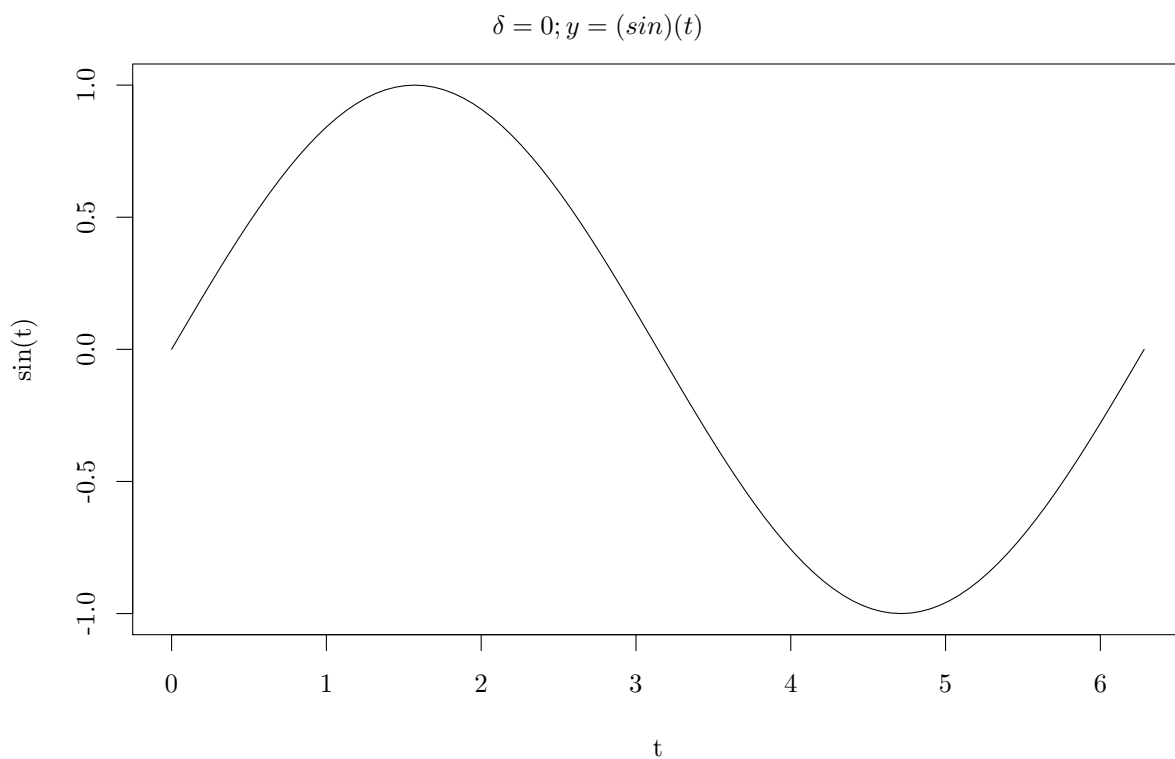```

```r
rhohat_one(v)
```

```
## [1] 0.4219589
```

Chapter 6

The Frequency Domain

6.1 Review of sine and cosine

6.1.1 Phase shifts

This is what a sine wave looks like

$$\delta = 0; y = (sin)(t)$$
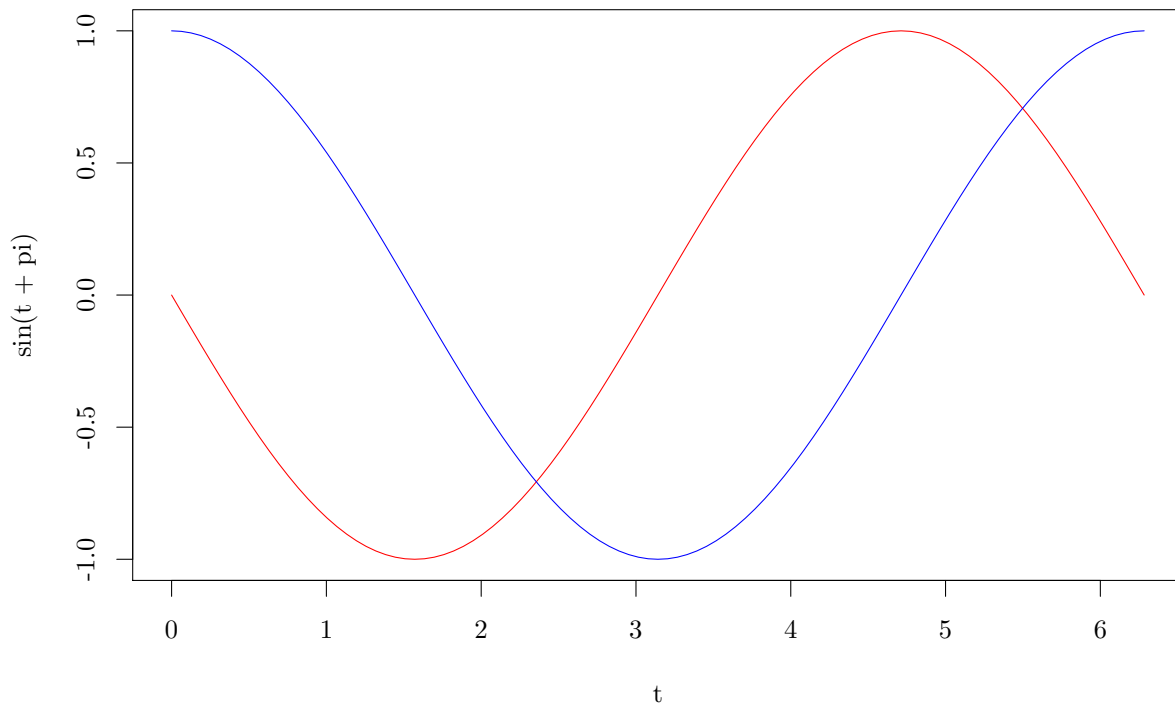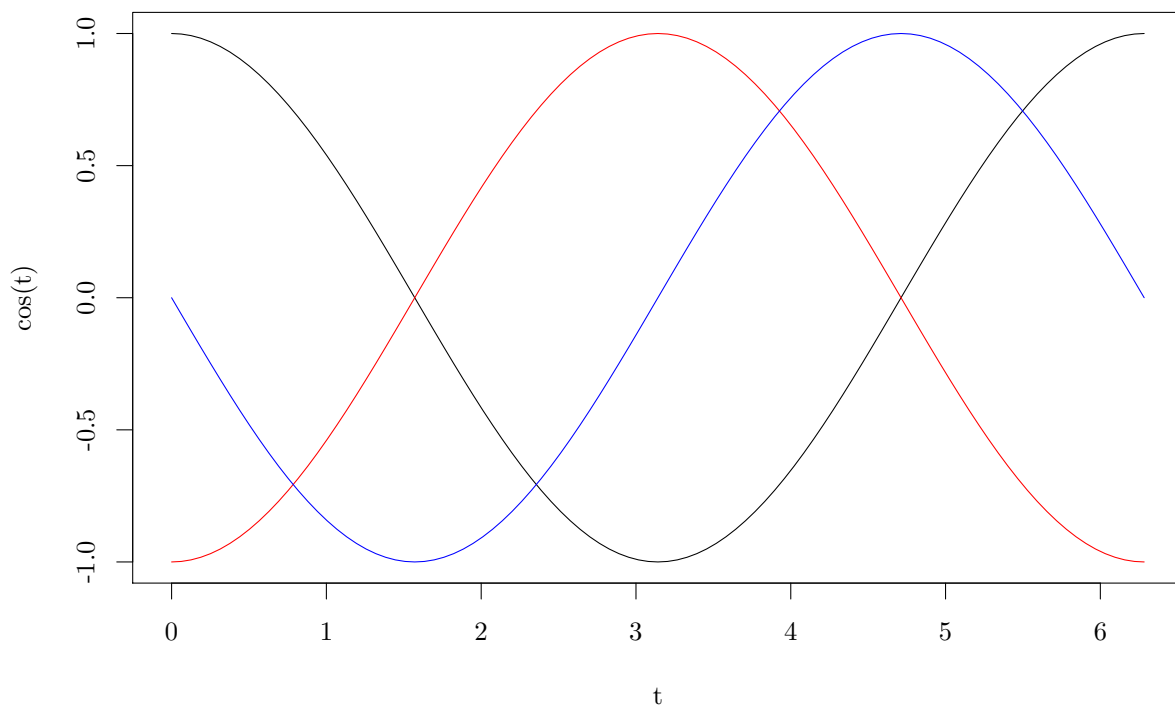
We can phase shift the sin wave by adding

RED:

$$\delta = \pi; y = \sin(t + \pi)$$

BLUE:

$$\delta = \frac{\pi}{2}; y = \sin(t + \frac{\pi}{2})$$

Now cosine

6.2   Frequency and Period

Let us say $f(x)$ is a periodic function. If $f(x)$ has period p, then p is the smallest value S.T.

$$f(x) = f(x + kp)$$

Cool

6.3 psueudo-periodicity

Data are *psuedo-periodic* with period p if p is the smallest value such that a cycle appears to repeat itself.

6.4 Aperiodicity

Aperiodic data is data in which no p exists.

6.5 Frequency

Frequency is calculated as : $\frac{1}{p}$

It symbolizes the number of cycles per unit(number of periods per unit)

6.6 Frequency Terminology

Frequency of a trigonometruc function:

Let $f(x) = \sin{(2\pi ft)}$

Then we have that :

$$p = 1/f$$

$$\nu = f$$

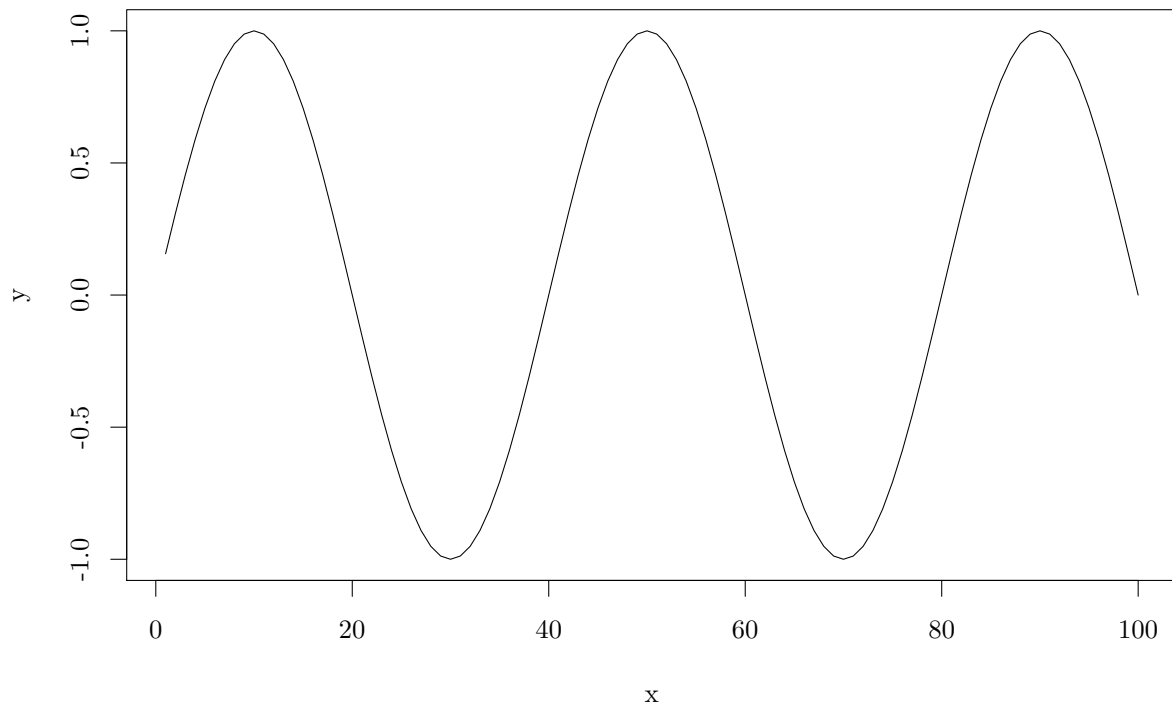Consider:

$$\sin(Bt + C)$$

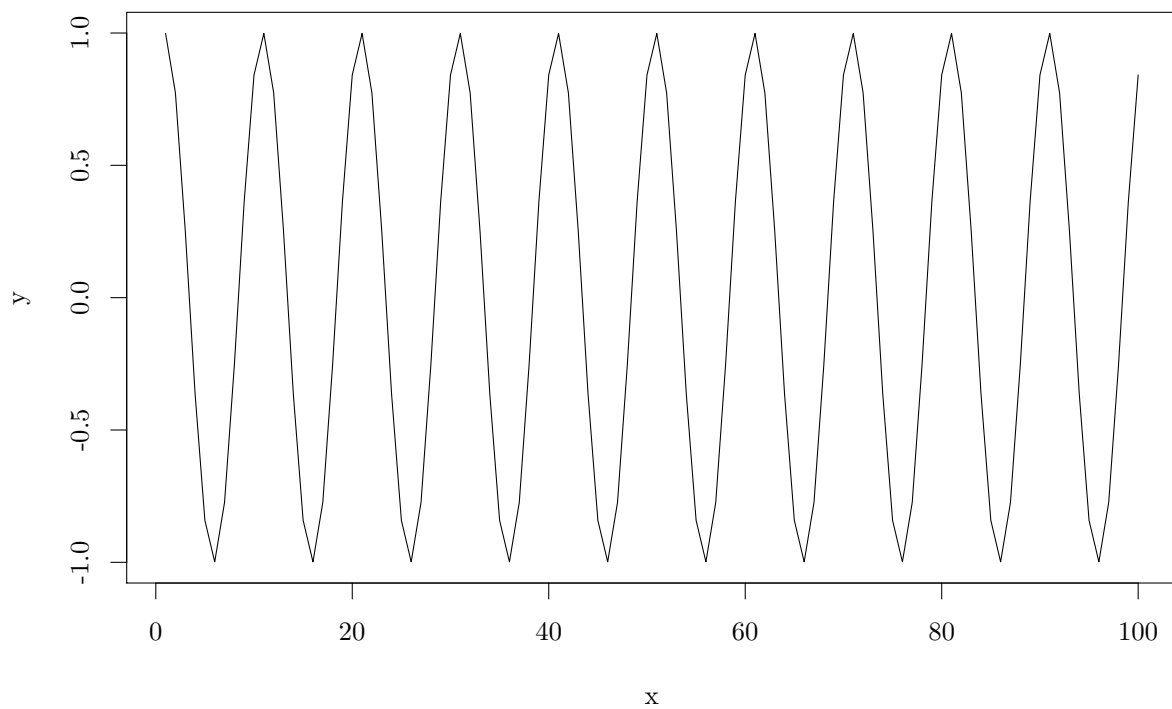Then

$$p = \frac{2\pi}{B}$$

$$\nu = \frac{B}{2\pi}$$

6.7 Composite sine function

```
t <- seq(1, 100, length = 100)
y1 <- sin(2 * pi * 0.025 * t)
y2 <- sin(2 * pi * 0.1 * t + 1)
y3 <- sin(2 * pi * 0.15 * t + 2.5)
lineplotter <- function(x, y) {
    plot(x, y, type = "l")
}

lineplotter(t, y1)
```
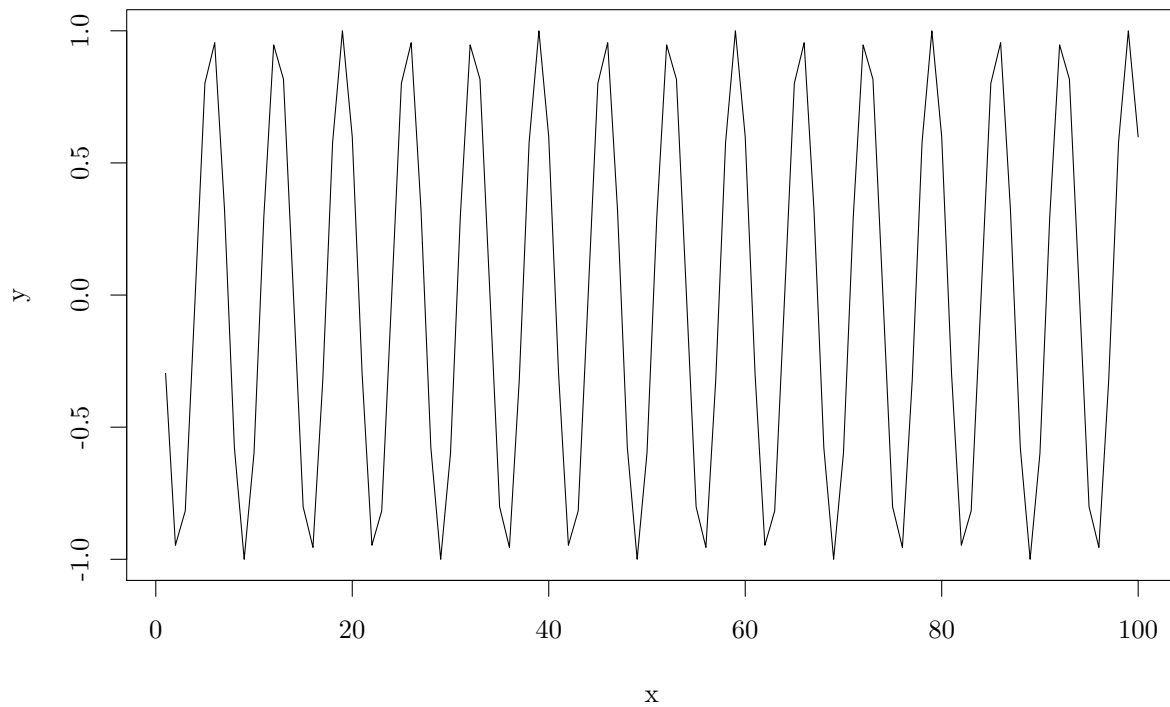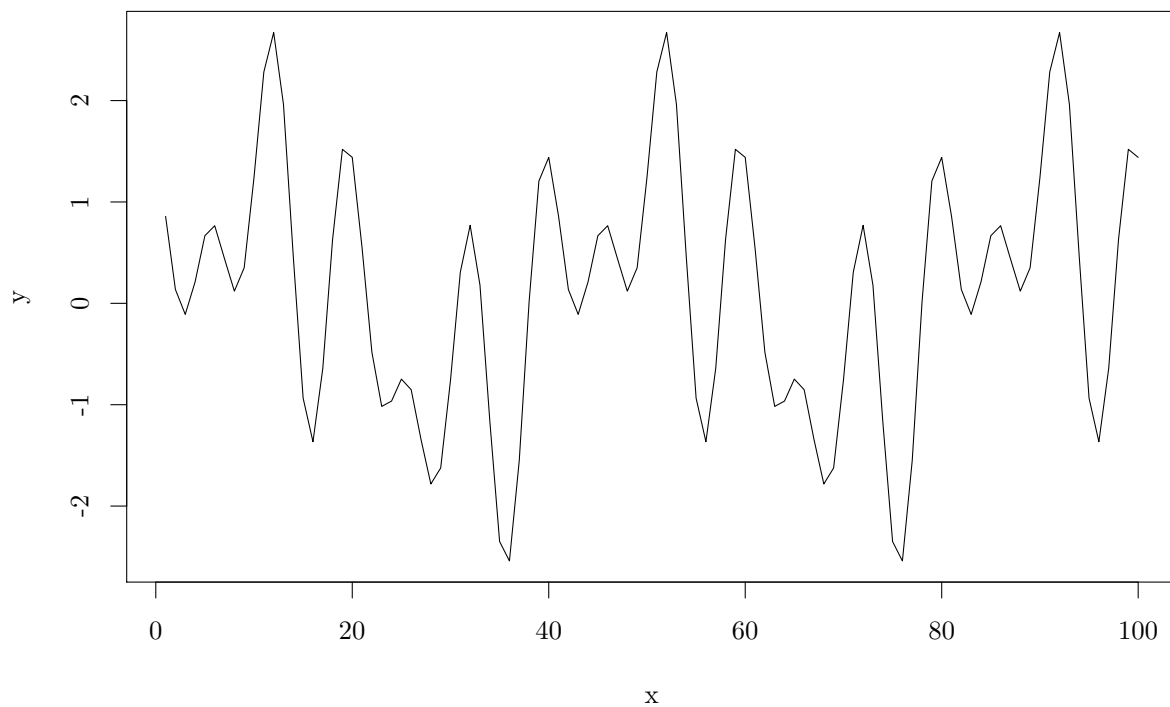
```
lineplotter(t, y2)
```



```
lineplotter(t, y3)
```

```
lineplotter(t, y1 + y2 + y3)
```


this is a signal with three frequencies!

6.8   Fundamental idea

A lot of realizations are made up of components of various frequencies, and some do not have frequency at all. Uncovering the frequency content of a dataset (or lack thereof) allows us to better understand the process generating the data.

6.8.1   A Brief Look at Fourier Series

When we do spectral analysis, we are generally going to be decomposing functions into sine and cosine terms, in order to better analyze the frequency of the data.

6.8.1.1   Fourier Series

You should just know this by this point, if you dont, google it. Too lazy to tex up the equations. Basically, we can decompose any function into a bunch of sines and cosines. It is pretty rad. The important conclusion of the existence of fourier series is that we can represent non cyclic functions as a bunch of sines and cosines, which will allow us to do spectral analysis of them.

Chapter 7

Spectral Density

A tool to identify the frequency content of a (stationary) time series.

7.1 Derivation of spectral density of white noise

$$S_x(f) = 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos\left(2\pi f k\right), \mid f \mid = 0.5$$

For white noise, $\rho_0 = 1$, if $k \neq 0$, $\rho_k = 0$

This tells us that for white noise:

$$S_x(f) = 1 + 2 * 0 = 1$$

In other words: **All frequencies are equally present in white noise**

7.2 Estimating Spectral Density from Real Data

With real data, we don't have an infinite amount, and we do not know what the true autocorrelation is. So we are probably going to have to make some estimates.

Lets first estimate autocorrelation:

$$S_x(f) = 1 + 2 \sum_{k=1}^{\infty} \widehat{\rho_k} \cos\left(2\pi f k\right), \mid f \mid = 0.5$$

Now lets deal with the infinity, which is killing us. We know that:

$$\hat{\rho}_1 = \frac{\frac{1}{n} \sum_{t=1}^{n} -1(X_t - \bar{X})(X_{t+1} - \bar{X})}{\hat{\gamma}_0}$$

And (check yourself to see the math works)

$$\hat{\rho}_{n-1} = (X_1 - \bar{X})(X_n - \bar{(x)})/(n\hat{\gamma}_0)$$

We can then rewrite the sample spectral density/sample spectrum as!:

$$S_x(f) = 1 + 2 \sum_{k=1}^{n-1} \widehat{\rho_k} \cos\left(2\pi f k\right), \mid f \mid = 0.5$$

However, this estimate of autocorrelation cannot be very good, since it is estimated using one cross product/pair. Thats like estimating the mean with one value!

So, lets stop this sum somewhere more reasonable.

$$\widehat{S_x(f)} = 1 + 2\sum_{k=1}^{M} \widehat{\rho_k}\cos\left(2\pi fk\right), \mid f \mid = 0.5$$

Where M is less than $n - 1$. This lets us have a bit more freedom to decide what M should be.

Let us now rewrite the estimate of spectral density as:

$$\widehat{S_x(f)} = \lambda_0 1 + 2\sum_{k=1}^{M} \lambda_k\widehat{\rho_k}\cos\left(2\pi fk\right), \mid f \mid = 0.5$$

Where $\mid \lambda_k \mid$ decreases as k increases. With this, we have M minimizing and limiting the impact of sample autocorrelations when k gets pretty close to n, and lambda further decreases the impact of the poor sample autocorrelations at high values of k. These are called *window functions*. In our case, we will use what is called the Parzen window. Other windows (every statistician in the 60's had their own window) include: the Tukey window, the Bartlett window, etc. We call this the the smooth spectral density estimator. M is frequently chosen to be $2\sqrt{n}$
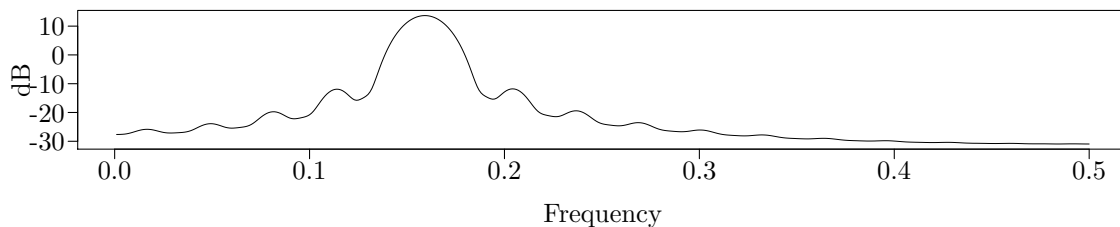
7.3   Spectral Density in the Wild

First, lets look at a simple example:

```r
x <- 1:1000
y <- sin(x)
par(mfrow = c(2, 1))
curve(sin(x), 1, 1000)
parzen.wge(y)
```
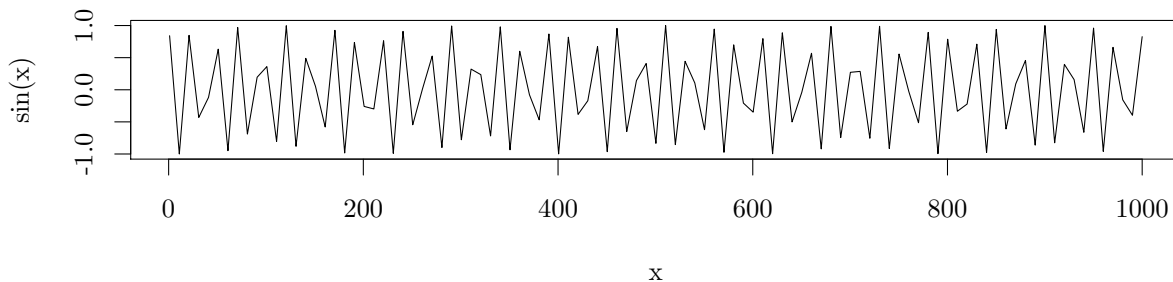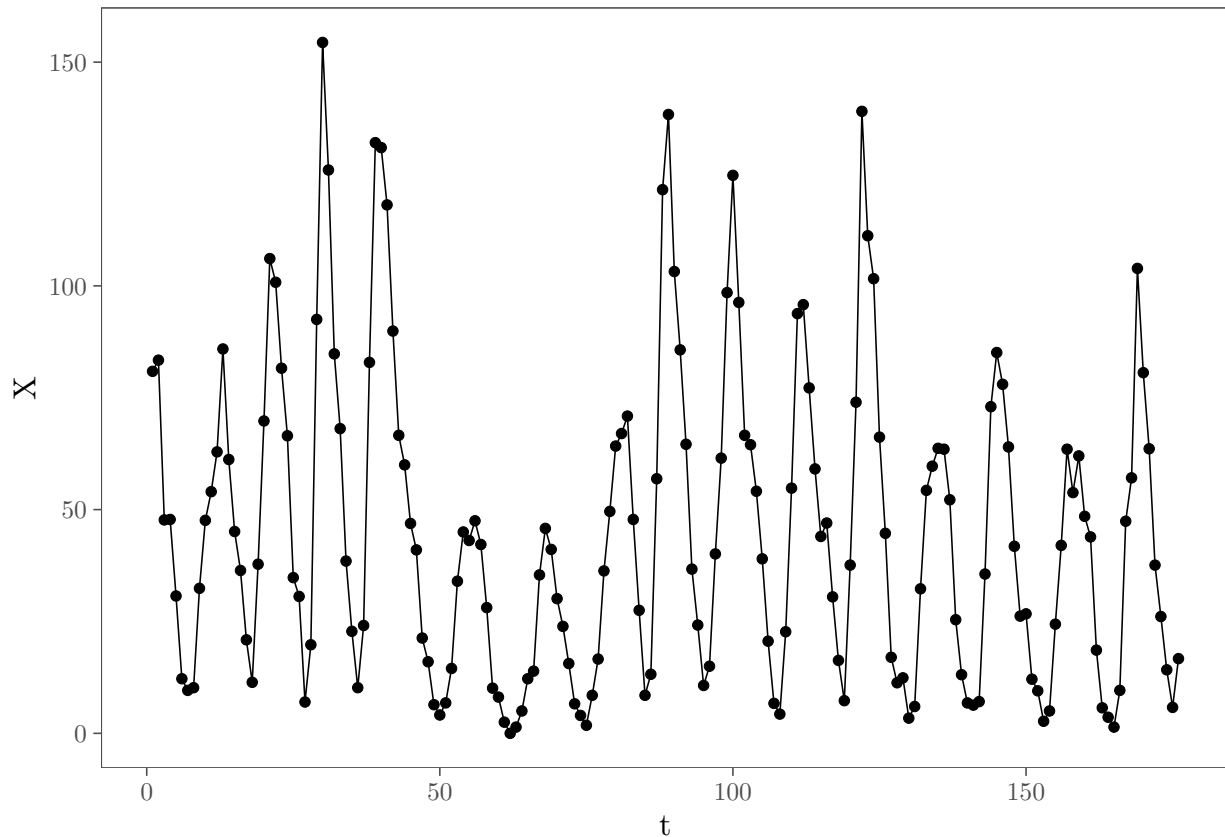

Parzen Window Truncation point: M = 63

Shockingly, we have a frequency peak at roughly $\frac{1}{2\pi} = 0.1591549$. If I did a better job of generating the data, it would be a super shark peak and nothing else, however I did not. Lets look at sunspot data now:
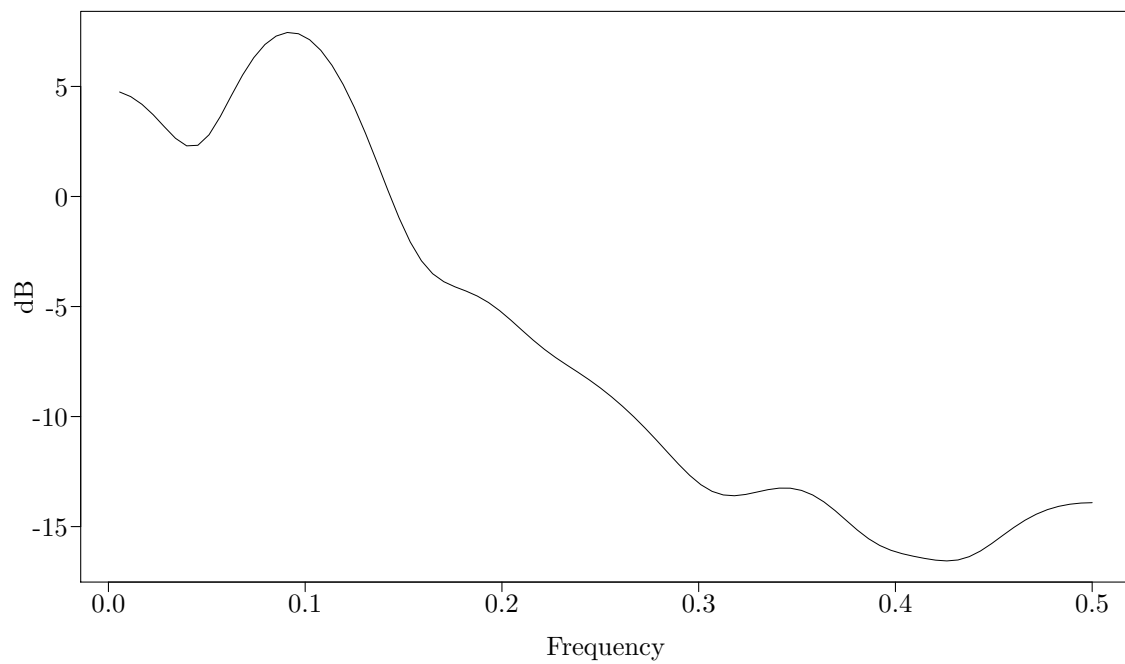
```r
par(mfrow = c(1, 1))
data(sunspot.classic)
tplot(sunspot.classic) + ggthemes::theme_few()
parzen.wge(sunspot.classic)
```
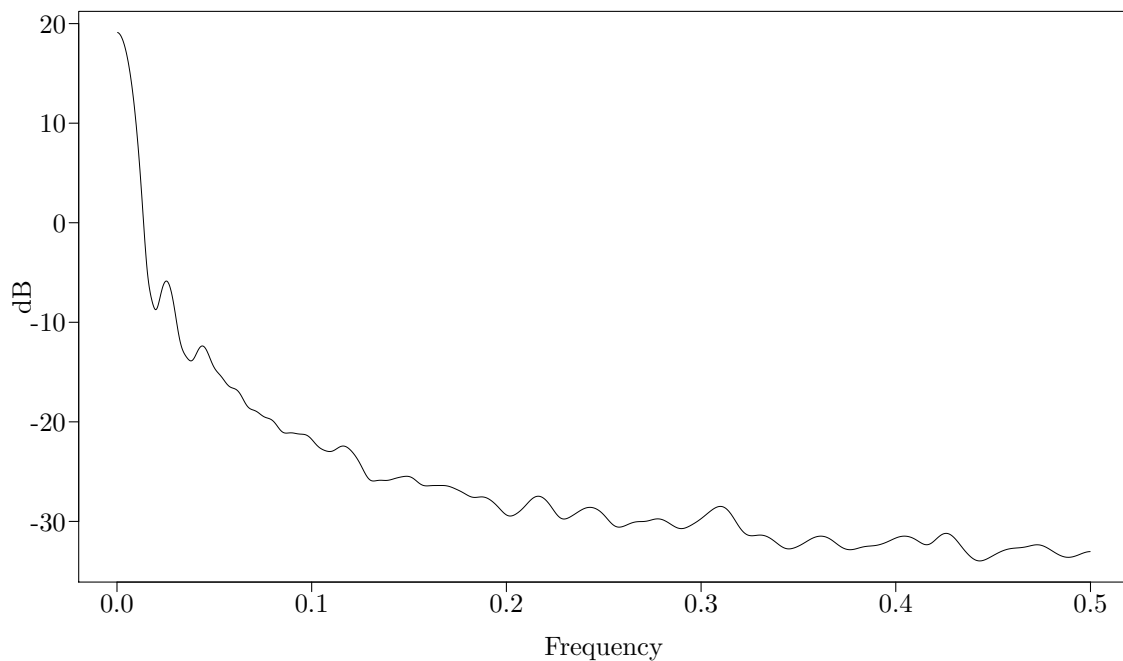

Parzen Window Truncation point: M = 26

It looks like we have a frequency peak at about 0.1, which makes sense because the classic sunspot psuedoperiod is about 11 years. Lets look at some aperiodic stock data now:

```
getSymbols("GOOG", src = "yahoo")
plot(GOOG[, "GOOG.Close"])
parzen.wge(GOOG$GOOG.Close)
```


GOOG[, "GOOG.Close"]   2007-01-03 / 2019-06-27

Parzen Window Truncation point: M = 112

Sadly, there appears to be no real periodicity in google's stock price, no free money for us :(. The peak at 0 means that the data is either aperiodic or it has not been long enough to show a period.

7.4   Why do we do spectral density from -0.5 to 0.5?

Time series data is taken in integers. What is the shortest distance between two integeres that we can complete a period in? It is 2. This means the highest obesrvable frequency is 0.5. Interestingly, if you look at sin(2pi.3t) and sin(2pi1.3t) at the integers, we will have exactly the same value. This means that even if the frequency is higher, we will have it be aliased to be exactly the same as the lower frequency function:

```
x <- 0:100
y1 <- sin(2 * pi * 0.3 * x)
y2 <- sin(2 * pi * 1.3 * x)
round(y2 - y1, 2)
```

```
##    [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   [71] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

This is called *aliasing*. These two functions will have the same spectral densities. 0.5 is known as the *Nyquist Frequency*

```
knitr::include_graphics("fig/windows.png")
```

1. Bartlett (triangular) window

$$\lambda_k = 1 - \left(\frac{k}{M}\right), \quad k = 0, 1, \ldots, M$$
$$= 0, \quad k > M$$

2. Tukey window

$$\lambda_k = \frac{1}{2}\left[1 + \cos\left(\frac{k\pi}{M}\right)\right], \quad k = 0, 1, \ldots, M$$
$$= 0, \quad k > M.$$

3. Parzen window

$$\lambda_k = 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{|k|}{M}\right)^3, \quad 0 < k \leq \frac{M}{2}$$
$$= 2\left[1 - \left(\frac{|k|}{M}\right)\right]^3, \quad M/2 < k \leq M$$
$$= 0, \quad k > M.$$

A note on plots of spectral density: Normally we do this in a logarithmic scale, so we can see whats going on.

Chapter 8

AR(1) Models and Filtering

8.1 Algebra review

Be able to find the roots of a 2nd order polynomial

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$z = a + bi$$

$$i = \sqrt{-1}$$

a is real b is imaginary

$$z^* = a - bi$$

z^* is complex conjugate

It is basically a vector so absolute value is just the magnitude

Unit circle is the values in the complex plane in which the magnitude of z is equal to one. If Z is a complex number with magnitude greater than one, it is outside the unit circle.

```r
quad_form <- function(a, b, c) {
    rad <- b^2 - 4 * a * c
    if (is.complex(rad) || all(rad >= 0)) {
        rad <- sqrt(rad)
    } else {
        rad <- sqrt(as.complex(rad))
    }
    round(cbind(-b - rad, -b + rad)/(2 * a), 3)
}

quad_form(c(1, 4), c(1, -5), c(6, 2))
```

```
##                [,1]          [,2]
## [1,] -0.500-2.398i -0.500+2.398i
## [2,]  0.625-0.331i  0.625+0.331i
```

```
# [,1] [,2] [1,] -0.500-2.398i -0.500+2.398i [2,] 0.625-0.331i 0.625+0.331i
```

8.2 Linear filters

A filter will turn Z_t into X_t

for example:

$$\text{difference} : X_t = Z_t - Z_{t-1}$$

Moving average also a filter (smoother)

8.2.1 Example

Consider $Z\_1 = 8 , Z\_2 =14, Z\_3 = 14, Z\_4 = 7$

If we apply the difference filter to this, we have that:

$$X_2 = 6, X_3 = 0, X_4 = 7$$

Note: the differenced data are a realization of length n-1

We can use the difference filter to remove the wandering behjaviior

8.2.2 5 point moving average

We can only get a realization of n-4. (n-nopoints-1). What does the 5 point moving average do (it averages a point and two points ahead and two behind). This filter can filter out some frequencies

8.3 Types of filters:

- Low pass filter
    - Filters out high frequency
    - Such as 5 point moving average, it smooths
- High pass filter
    - Leaves high freq but removes low freq
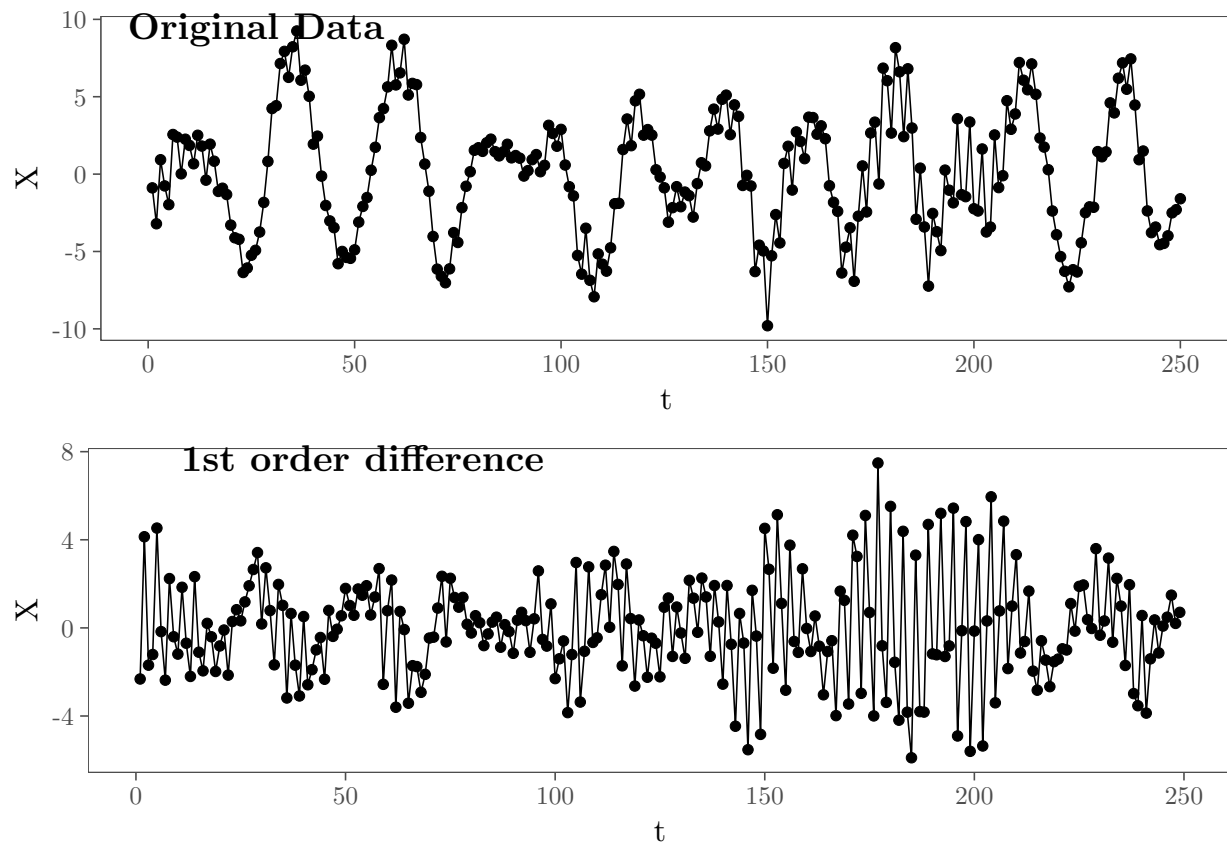   - Such as differencing

8.3.1 An example in R

```
mafun <- function(xs, n) {
    stats::filter(xs, rep(1, n))/n
}
dfun <- function(xs, n) {
    diff(xs, lag = n)
}
th <- ggthemes::theme_few()
library(tswge)
data(fig1.21a)
ma <- mafun(fig1.21a, 5)
d <- dfun(fig1.21a, 1)
fp <- tplot(fig1.21a) + th
mp <- tplot(ma) + th
dp <- tplot(d) + th
```
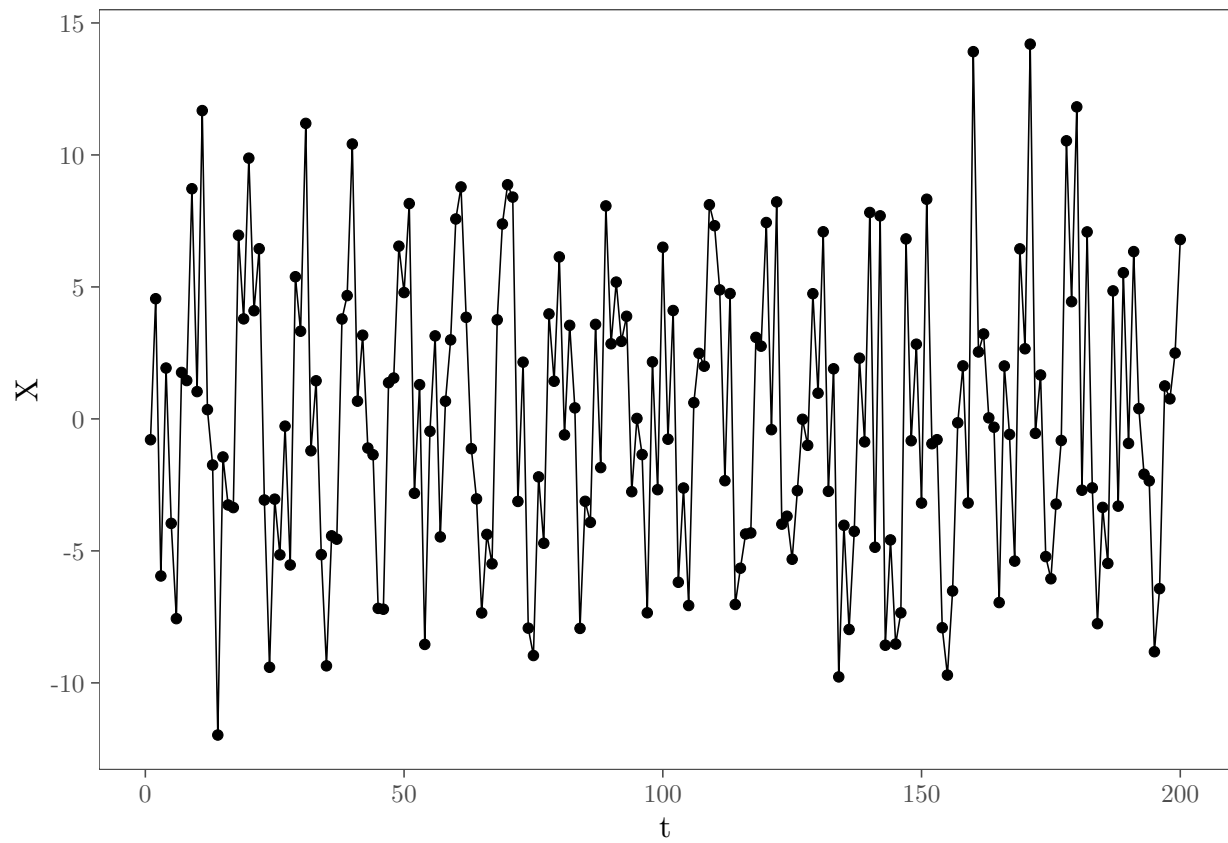
And here lets view the 5 point moving average
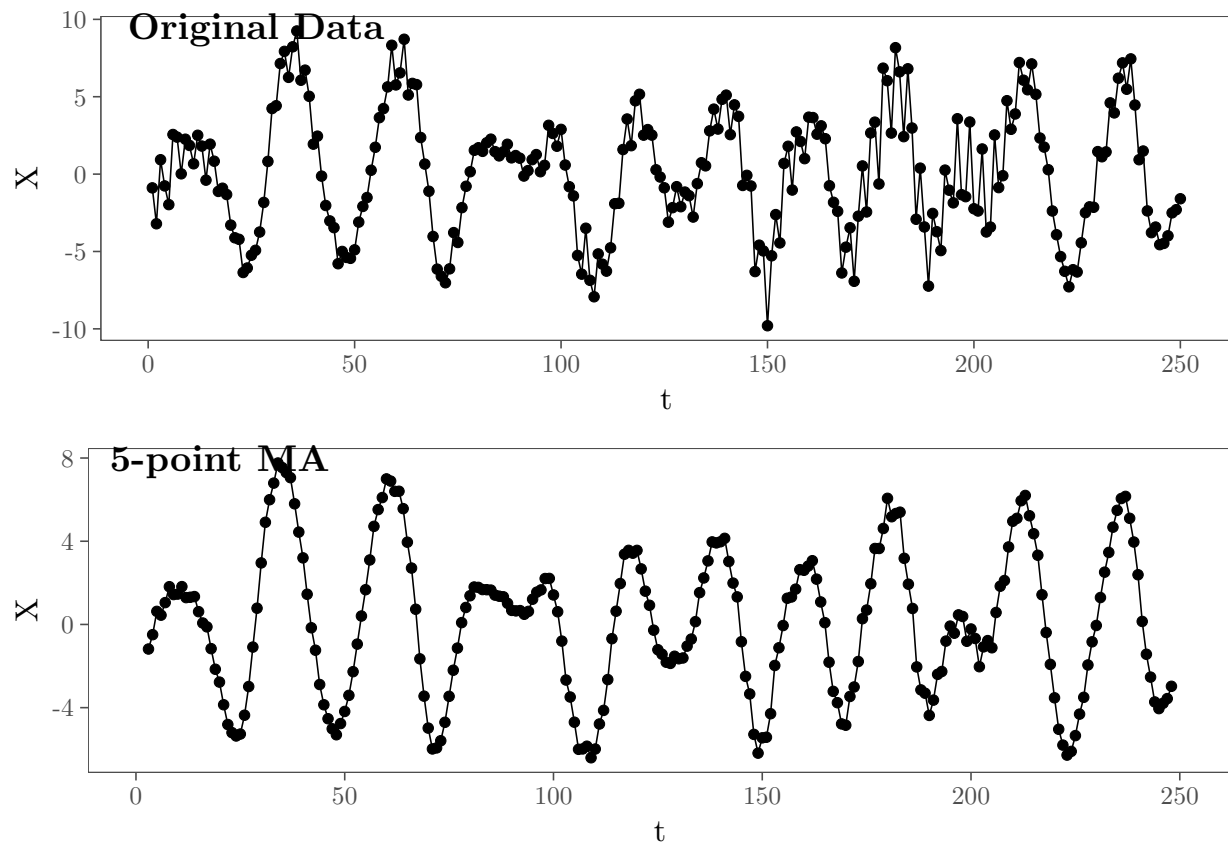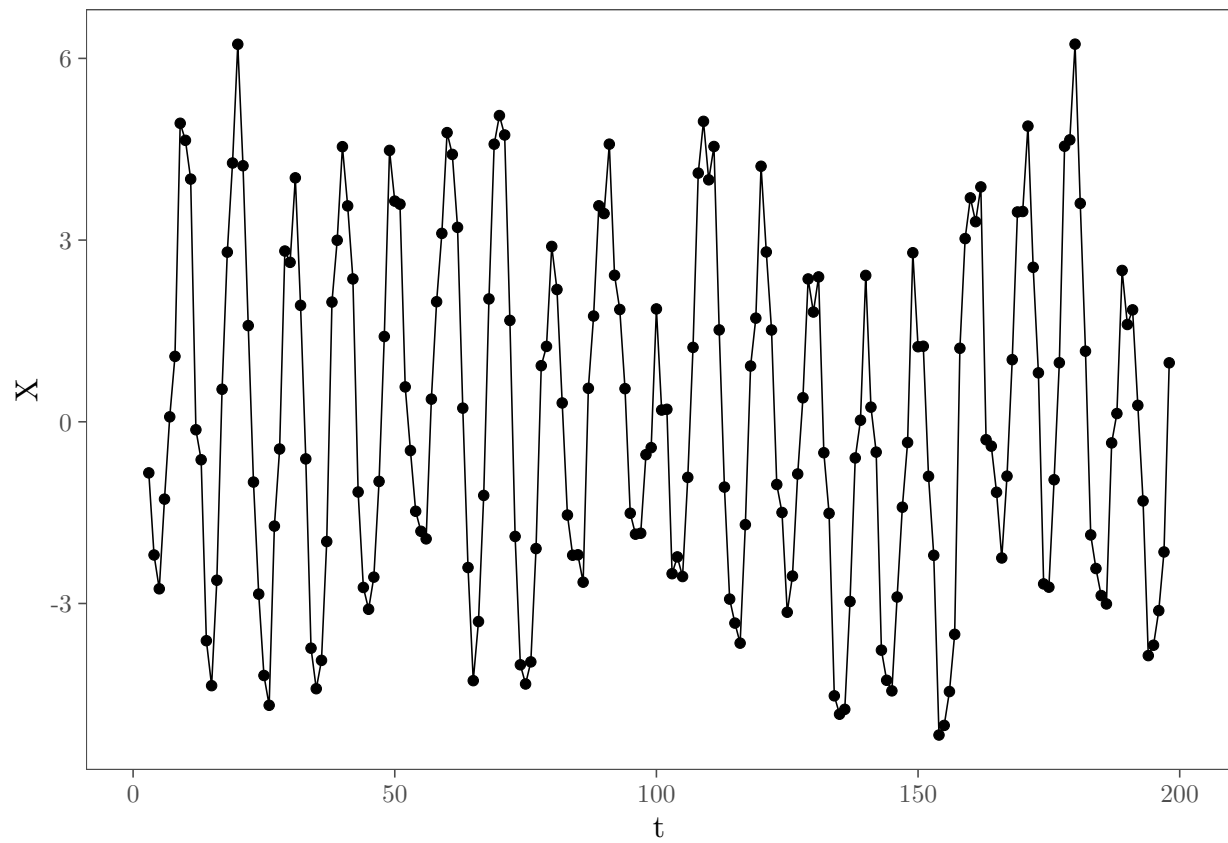
And now a look at the difference filter

```r
rlz <- gen.sigplusnoise.wge(200, coef = c(5, 3), freq = c(0.1, 0.45), vara = 10,
    sn = 1, plot = F)

pfun <- function(x, n, l) {
    fp <- tplot(x) + th
    mp <- tplot(mafun(x, n)) + th
    dp <- tplot(dfun(x, l)) + th
    list(original = fp, MA = mp, dif = dp)
}
plts <- pfun(rlz, 5, 1)
plts$original
```
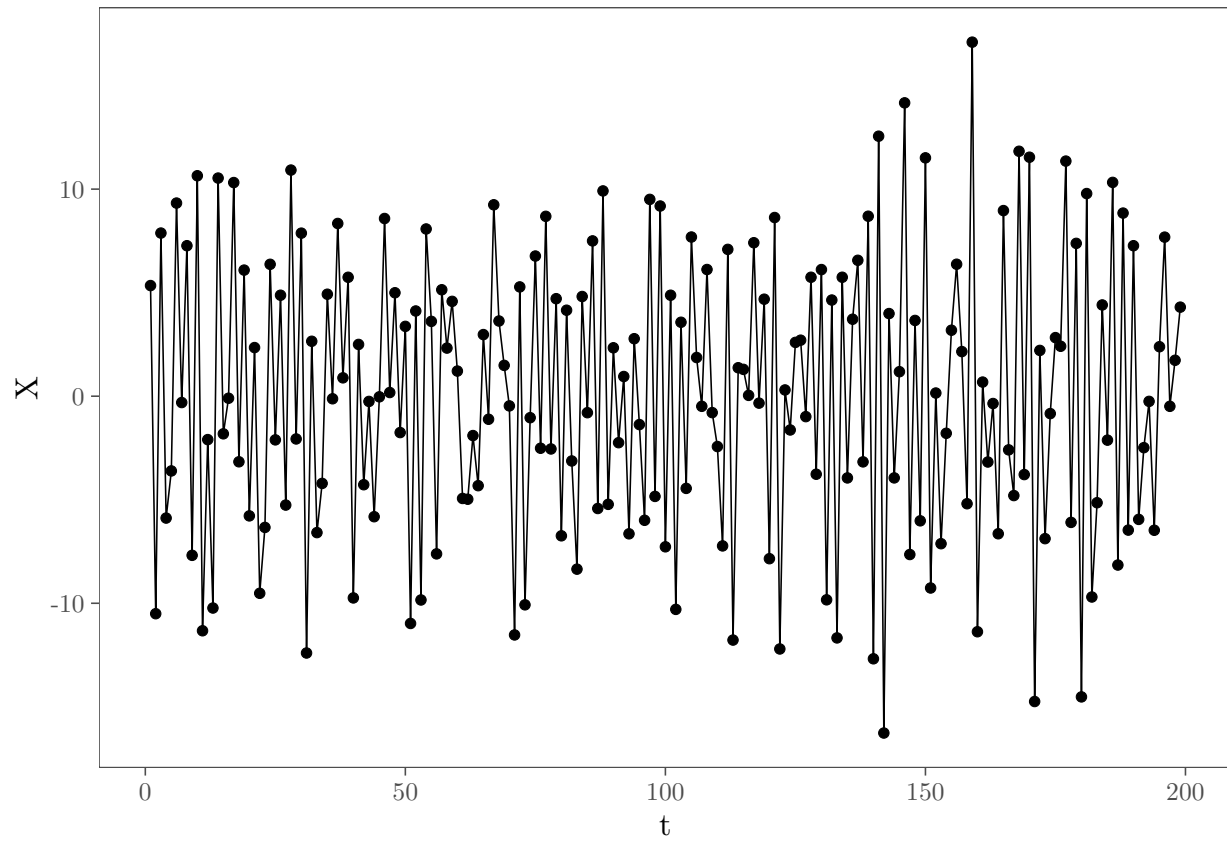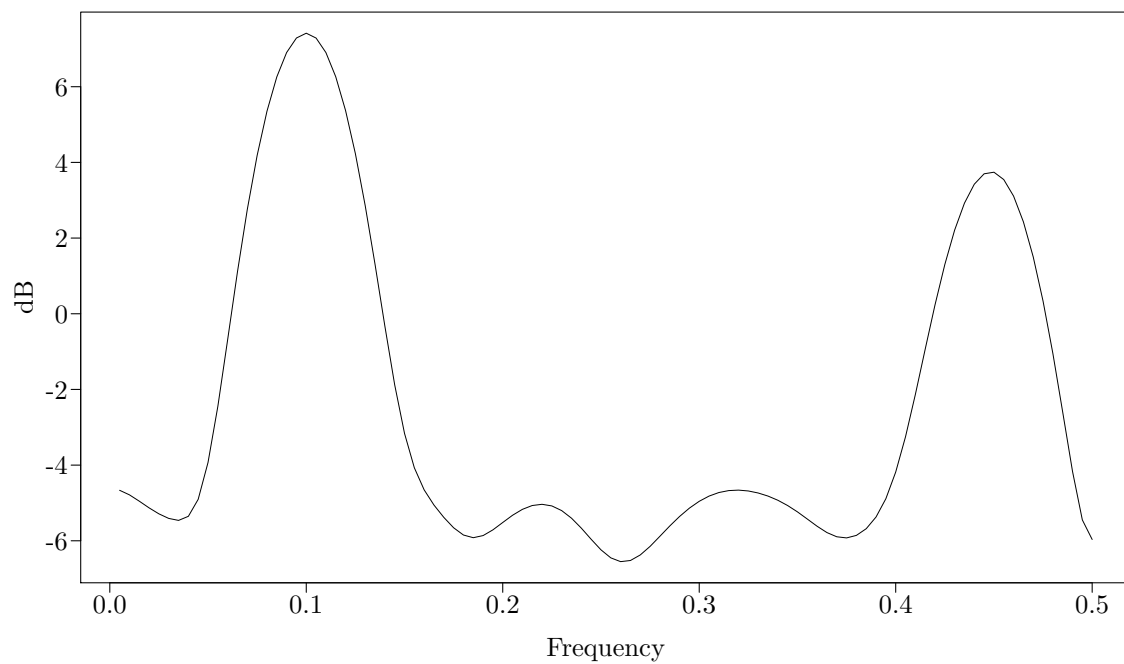
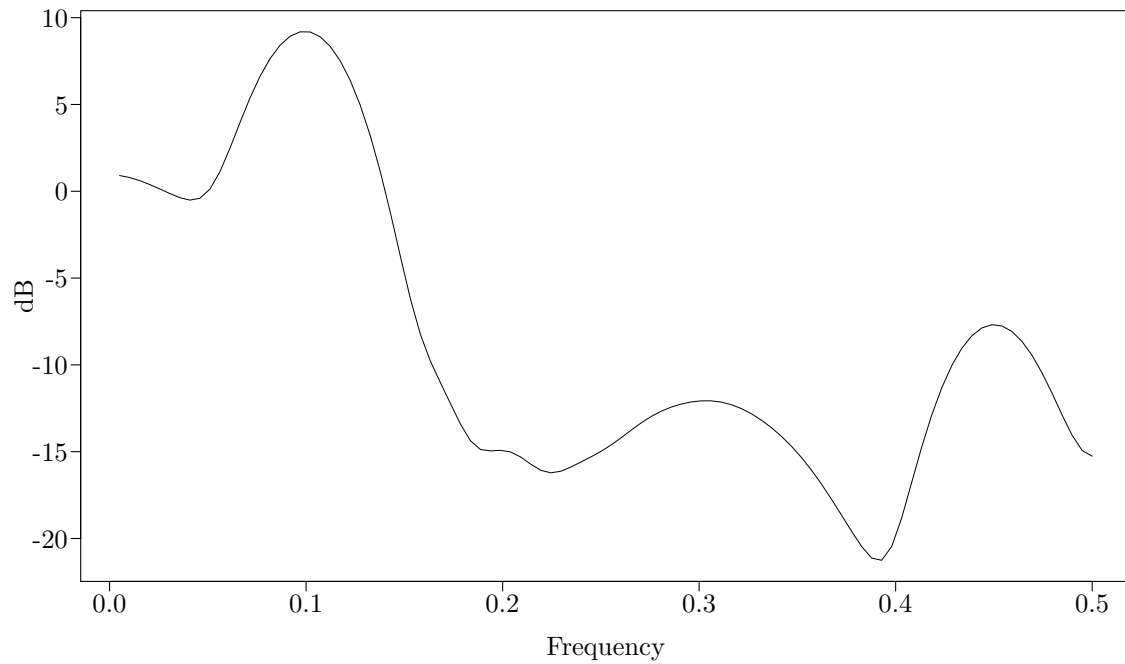```
plts$MA
```

```
plts$dif
```

```
parzen.wge(rlz)
```

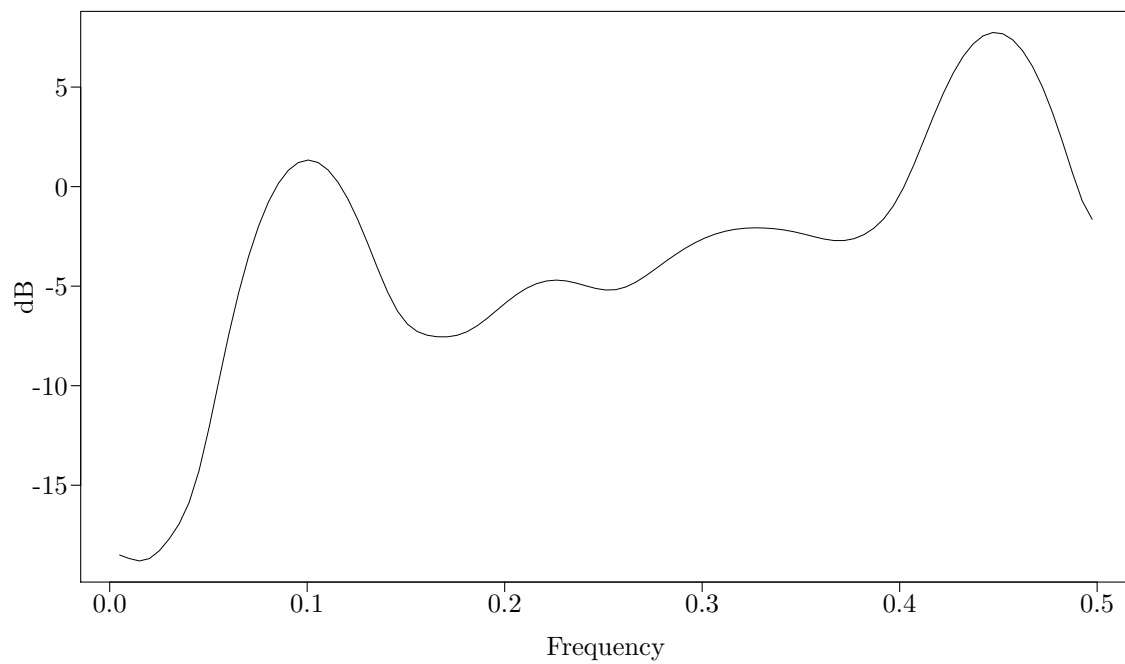**Parzen Window Truncation point: M = 28**

```
parzen.wge(na.omit(mafun(rlz, 5)))
```

Parzen Window Truncation point: M = 28


```
parzen.wge(na.omit(dfun(rlz, 1)))
```

Parzen Window Truncation point: M = 28

8.4 GLPs

a GLP is a linear filter with a *white noise input*

We take white noise input and put it through a filter and get something that isnt white noise.

$$\sum_{j=0}^{\infty} \psi_j a_{t-j} = X_t - \mu$$

GLP's are an infinite sum of white noise terms. This might be weird now but later on this concept will return.

The ψ_j's are called psi weights, these will be useful again later. AR, MA, and ARMA are all special cases of GLP's, and will be useful when we study confidence intervals.

8.5 AR(1) Intro

AR(p) in which p = 1

We will go through well known forms of the AR(1) model.

$$X_t = \beta + \phi_1 X_{t-1} + a_t$$

$$\beta = (1 - \phi_1)\mu$$

Beta is moving average constant

We are saying that the vluse of x depends on some constant, the previous value of x, and some random white noise. Looks a lot like regression except there is a wierd variable

we can rewirte tis as

$$X_t = (1 - \phi_1)\mu + \phi_1 X_{t-1} + a_t$$

An are 1 process is stationary iff the magnitude of ϕ_1 is less than 1

We will deal with this a lot in the near future

8.5.1 AR(1) math zone

$$E\left[X_t\right] = \mu?$$
$$E\left[X_t\right] = E\left[1 - \phi_1\mu\right] + E\left[\phi_1 X_{t-1}\right] + E[a_t]$$

We can rewrite this as
$$E\left[X_t\right] = 1 - \phi_1\mu + \phi_1 E\left[X_t\right] + 0$$

$$E\left[X_t\right]\left(1 - \phi_1\right) = 1 - \phi_1\mu$$
$$E\left[X_t\right] = \mu$$

Mean does not depend on T

The variance also does not, and if phi1 is less than one variance is finite

$$\sigma_X^2 = \frac{\sigma_a^2}{1 - \phi_1^2}$$

and rhok is phi1 to the k

$$\rho_k = \phi_1^k, k \geq 0$$

Spectral Density of AR(1) also does not depend on time, it just monotonically increases or decreases depending on phi1:

$$S_X(f) = \frac{\sigma_a^2}{\sigma_X^2} \left(\frac{1}{| 1 - \phi_1 e^{-2\pi if} |^2} \right)$$

8.5.2   the zero mean form of AR1

with zero mean,
$$X_t = \phi_1 X_{t-1} + a_t$$

OR

$$X_t - \phi_1 X_{t-1}$$

A

8.5.3   AR1 with positive phi

remember we have

$$\rho_k = \phi_1^k$$

With positive ϕ_1, we have:

realizations are wandering and aperiodic

Autocorrelations are damped exponentials

Spectral density Peaks at zero.

Higher values = stronger versions of these characterstics

8.5.4   AR1 with negative phi

We have:

realization are oscillating

*Autocorrelations* are damped and alternating(negative to a power you fool)

Spectral density[*] has a peak at f = 0.5 (cycle length of 2)

Higher magnitude = stronger characteristics

8.5.5   Nonstationary

if phi = 1 or phi> 1, we have that realizations are from a nonstationary process. with it equal to one, it looks ok, and is actualy a special ARIMA model. WIth it > 1, we have crazy explosive realizations. Check this out:

```
nonstarma <- function(n, phi) {
    x <- rep(0, n)
    a <- rnorm(n)
    x[1:n] <- 0
    for (k in 2:n) {
        x[k] = phi * x[k - 1] + a[k]
    }
    tplot(x)
}

nonstarma(n = 50, phi = 1.5) + th
```
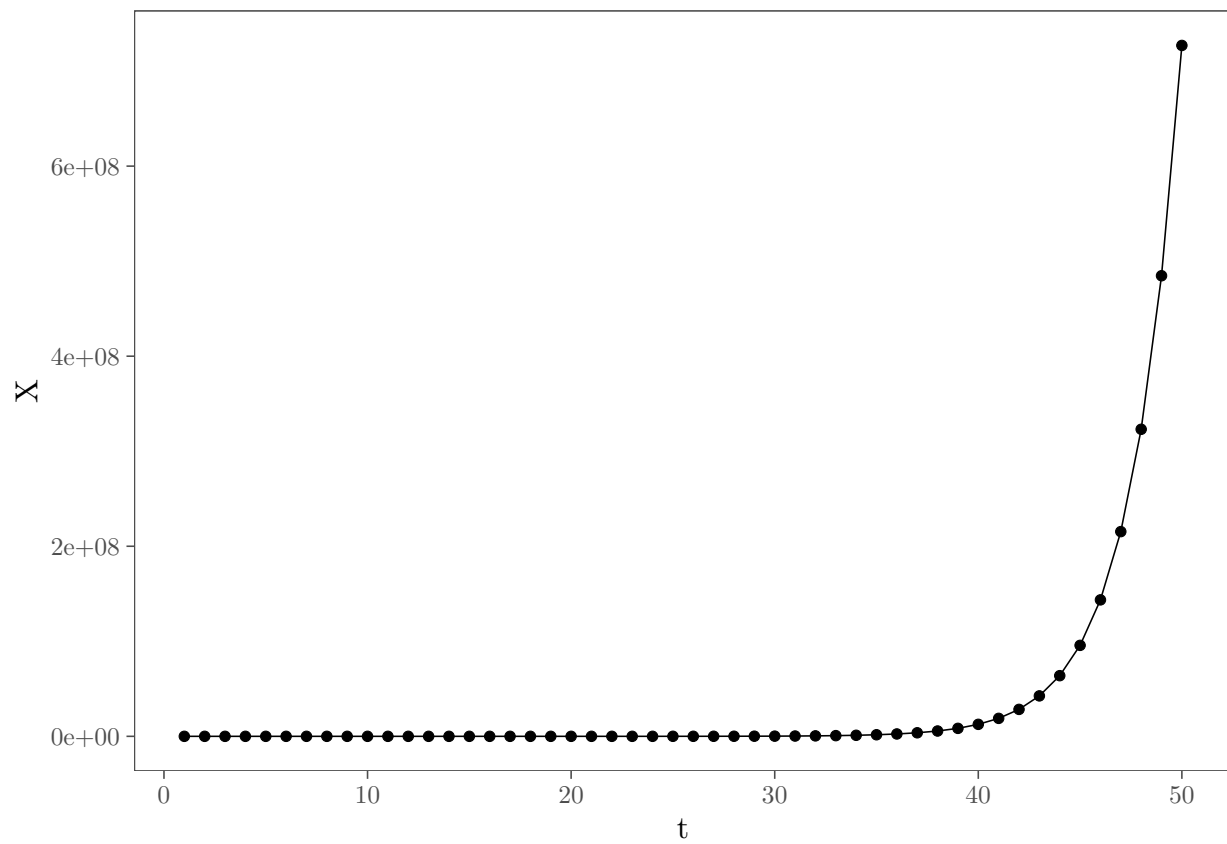
Chapter 9

Characteristic Equations for AR(1) Models

Consider the zero-mean form of an AR(1) model

$$X_t - \phi_1 X_{t-1} = a_t$$

We can rewtite in operator form as

$$(1 - \phi_1 B)X_t = a_t$$

Where B is an operator that turns X_t into X_{t-1}. We can then rewrite this akgebraically (for solving purposes) as a characteristic equation:

$$1 - \phi_1 Z = 0$$

.

Now recall that AR(1) is stationary $iff \mid \phi_1 \mid$ is less than 1. Lets solve the characteristic equation for the root now:

$$r = z = \frac{1}{\phi_1}$$

If X_t is stationary , $\mid r \mid$ is greater than one. This does not feel important now, but when we get to higher order time series this will save us a lot of thinking. Lets look at a numerical example just to make sure we got it:

$$X_t = -1.2X_{t-1} + a_t$$
$$X_t + 1.2X_{t-1} = a_t$$
$$(1 + 1.2B)X_t = a_t$$
$$1 + 1.2Z = 0$$
$$r = z = -\frac{1}{1.2} = -0.8333333$$

Note that in this case, $\phi_1 = -1.2$\$. This is a weird thing with this notation, get used to it.