

EDA Week 8





THE ONLY ACCEPTABLE PIE CHART

Agenda

- Quick wrap up: Time series
- Midterm Day 1: Grades/Review
- Midterm Day 2: Point reduction announcement
- Schedule for the semester, what is left?
- Final Projects Overview

What should you know about time series?

- You could study time series analysis for a very long time.
- This class aims to:
 1. Teach you how to identify time series structure
 2. Teach you how to examine time series data in R
 3. Teach you how to describe time series data
 4. Give you a foundation for later so that you know time series data requires special modeling considerations
- Note: today's lecture heavily relies on examples from the book "Forecasting: Principles and Practice" I highly recommend it!

Time series object

- A time series can be thought of as a list of numbers, along with some information about what times those numbers were recorded.

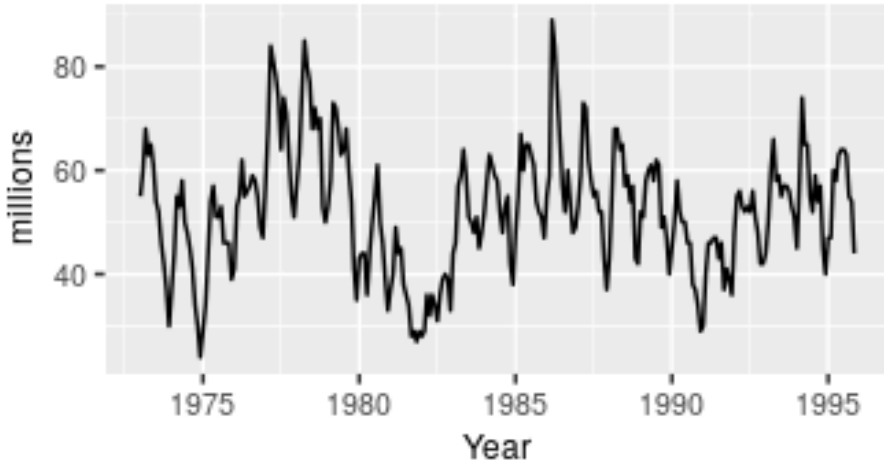
The “frequency” is the number of observations before the seasonal pattern repeats.¹ When using the `ts()` function in R, the following choices should be used.

| Data | frequency |
|-----------|-----------|
| Annual | 1 |
| Quarterly | 4 |
| Monthly | 12 |
| Weekly | 52 |

Describing Time Series Data

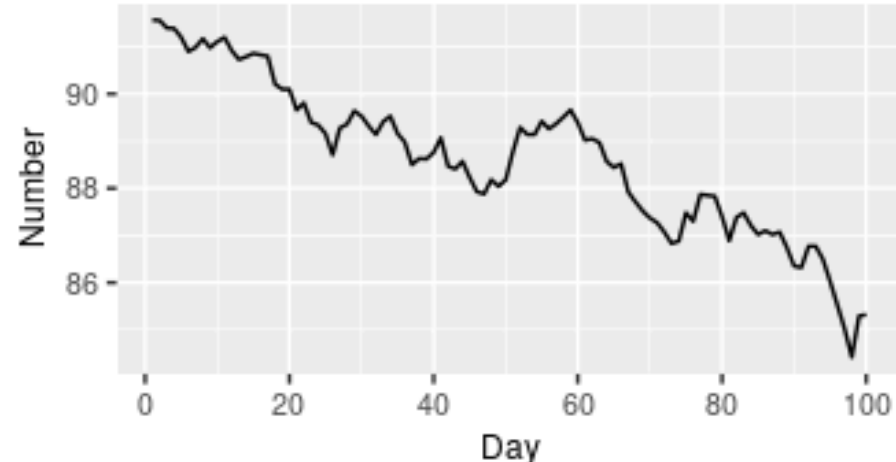
- Trend
 - A trend exists when there is a long-term increase or decrease in the data
 - A trend does not have to be linear
- Seasonal pattern
 - This occurs when a time series is affected by seasonal factors such as the day of the week. Seasonality is a fixed and known frequency.
 - These are associated with some aspect of the calendar
- Cyclic
 - A cycle occurs when the data exhibit a rise and fall that are not due to a fixed frequency.

Sales of new one-family houses, USA



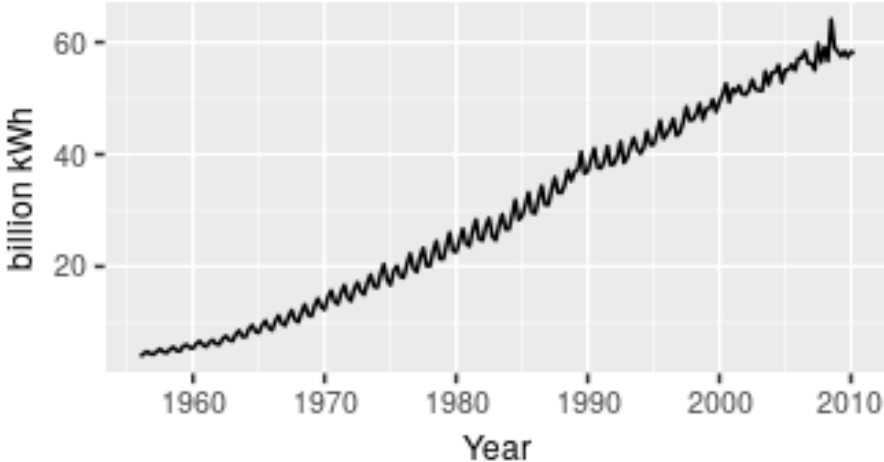
Seasonal (year)
Cyclic (6-10 years)

US treasury bill contracts



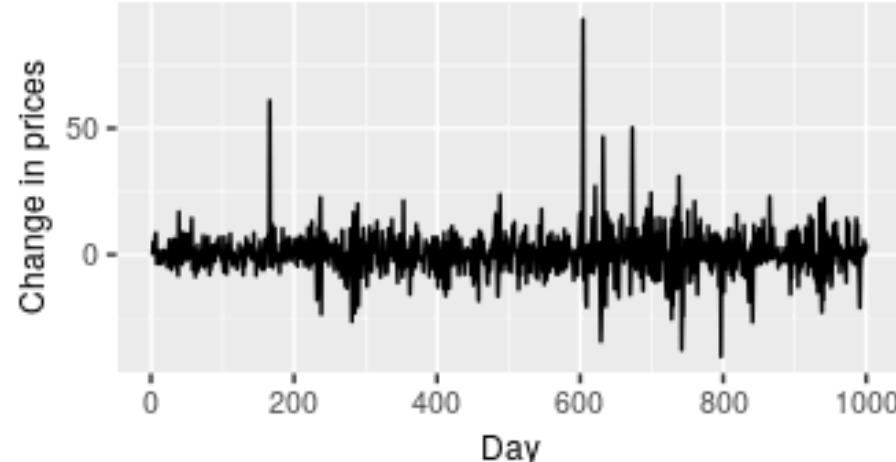
Downward
trend

Australian quarterly electricity production



Increasing trend
Seasonality

Google daily changes in closing stock price



No trend,
seasonality or
cyclic behavior

Lag

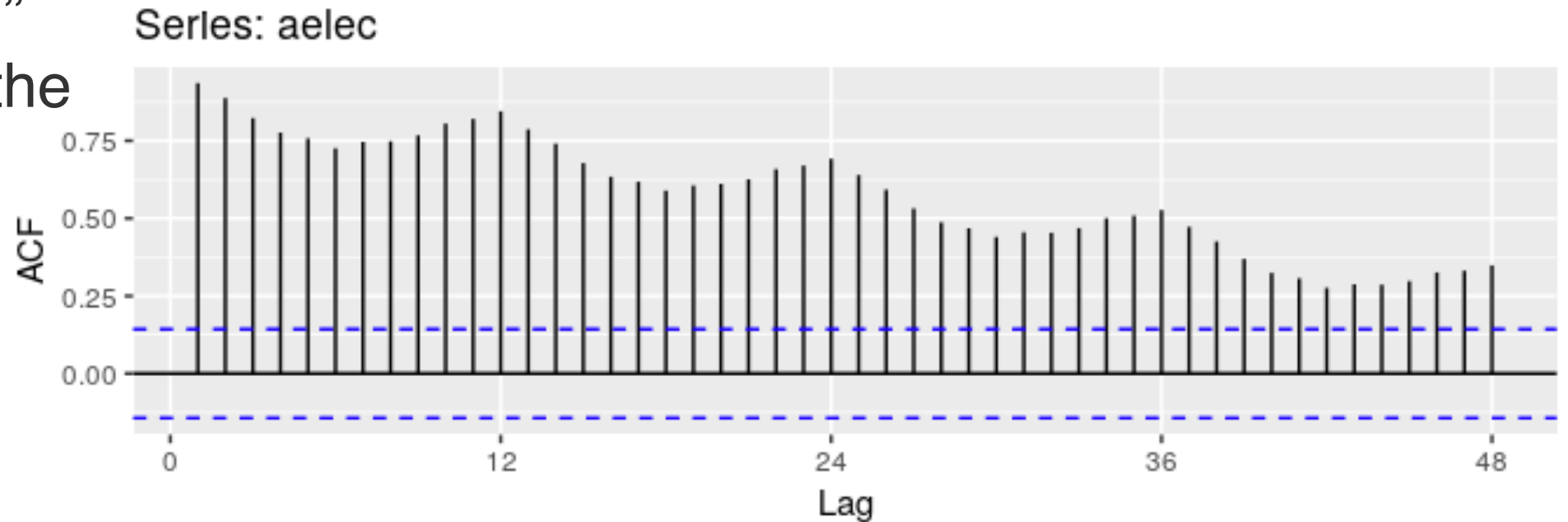
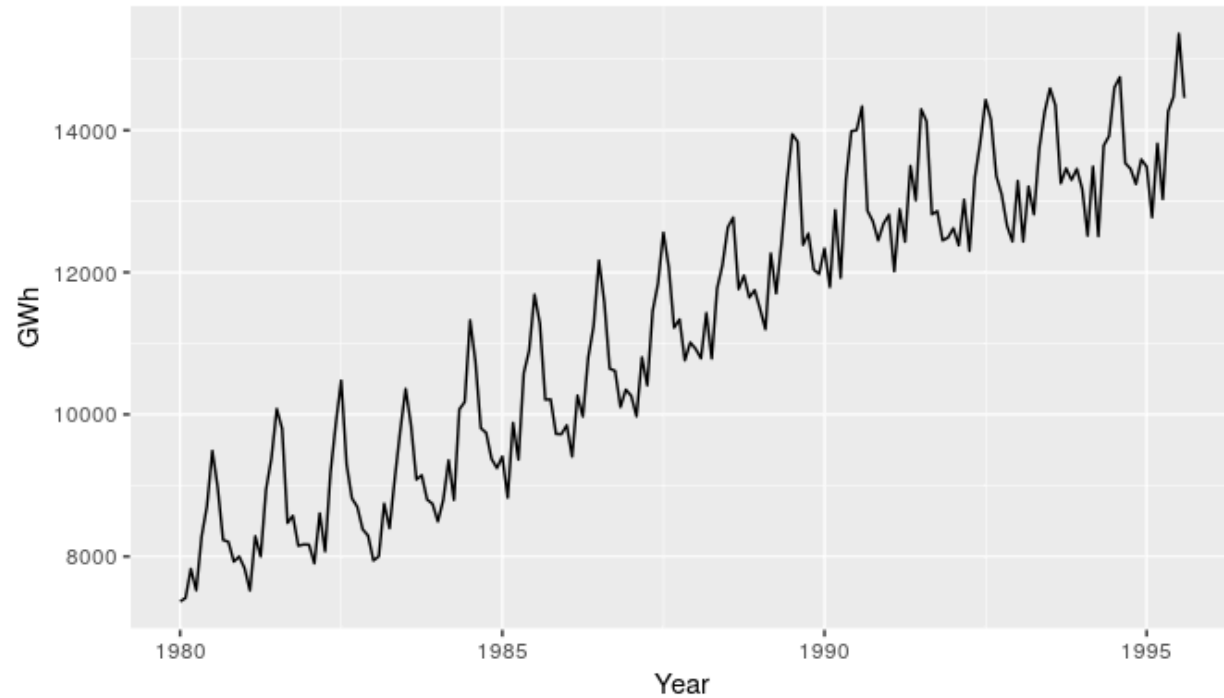
- a period of time between one event or phenomenon and another.
- “there was a time lag between the commission of the crime and its reporting to the police”
- often used in online gaming! But can be generally used to describe many gaps between an event or action and a latter effect from that event or action.
- if $t_0 = \text{now}$, lags can occur at any given t_{-1}, t_{-2} etc.

Autocorrelation

- autocorrelation measures the linear relationship between *lagged values* of a time series.
- ACF: autocorrelation function
 - When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in size.
 - So the ACF of trended time series tend to have positive values that slowly decrease as the lags increase.

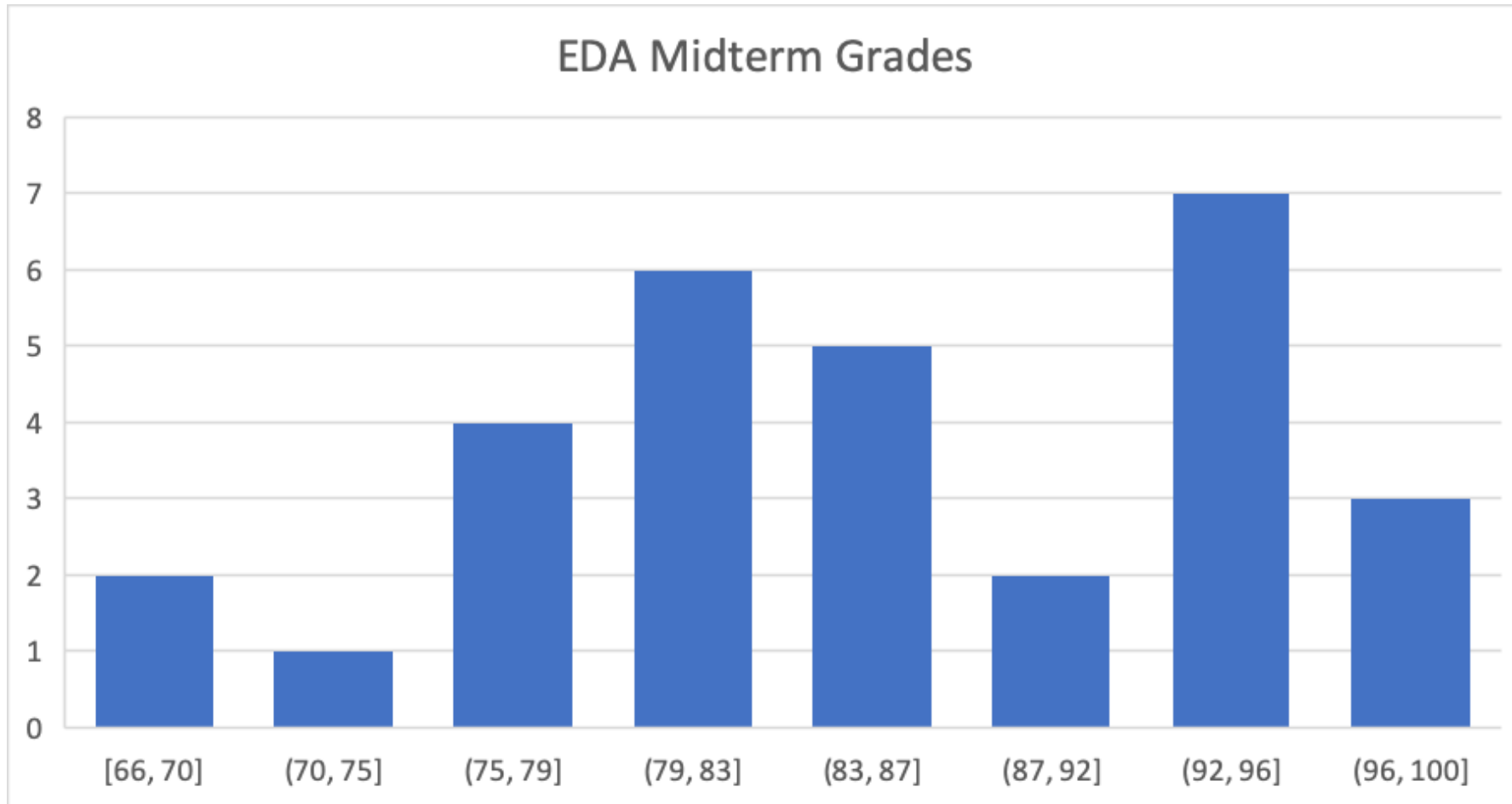
ACF

The slow decrease in the ACF as the lags increase is due to the trend, while the “scalloped” shape is due the seasonality.



More resources

- R markdown lecture on course website
- Forecasting: Principles and Practice
 - <https://otexts.com/fpp2/>
- Tutorial:
<https://www.rdocumentation.org/packages/forecast/versions/8.9/vignettes/JSS2008.Rmd>



← C+ and lower B- and higher →

How would you improve this graphic?

Point reduction announcement

Midterm Day 2



| Time completed | Point change |
|--------------------------------------|--|
| On time (3:54 or earlier) | Good job! Thank you. I know who you are. |
| 2 Minutes late | Grace period, no point reduction. |
| For every 1 minute after 3:56 | You will be penalized by .5 points. |
| For example, if you finished at 3:58 | Your exam grade = total score - 2 points |

Agenda

- ~~Quick wrap up: Time series~~
- ~~Midterm Day 1: Grades/Review~~
- ~~Midterm Day 2: Point reduction announcement~~
- Schedule for the semester, what is left?
- Final Projects Overview