

# STAT 436 / 536 - Lecture 2: Time Series Intro

August 29, 2018

## Time Series Basics

- So what is a time series? *measurements taken sequentially in time. In this class, we will typically see fixed time intervals, or discrete time, but you could also see continuous time responses, such as (earthquakes or old faithful eruptions). A time series model can also be referred to as a discrete-time stochastic process.*
- *Time series models, as well as spatial and spatiotemporal frameworks, are essentially just models to handle complicated correlation structure induced by similarity in the responses.*
- What is the purpose of time series analyses?
  - *understand the past and predict the future*
  - *control for correlation structure to make scientific inferences*
  - *exploit temporal correlation in forecasting scenarios*
- What are the defining features of time series data?
  - *measurements close in time are similar*
  - *trends can exist*
  - *seasonal variation or harmonic cycles can exist*

## R Overview

- R will be used frequently in this class. All assignments, labs, and take home exams should be completed using R Markdown to create reproducible reports.
- We will also spend time creating dynamic graphics and use R Shiny.
- For more details on some useful R packages see <https://www.rstudio.com/resources/cheatsheets/>.
  - `lubridate` for times
  - `dplyr` for data manipulation
  - `ggplot2` for graphics
  - `stringr` for string manipulation

## Time Series Objects

-R allows objects to be defined as a time series class using the `ts()` command. The time series object contains:

- the data,
- a start point,
- an end point,
- and the sampling period between observations.

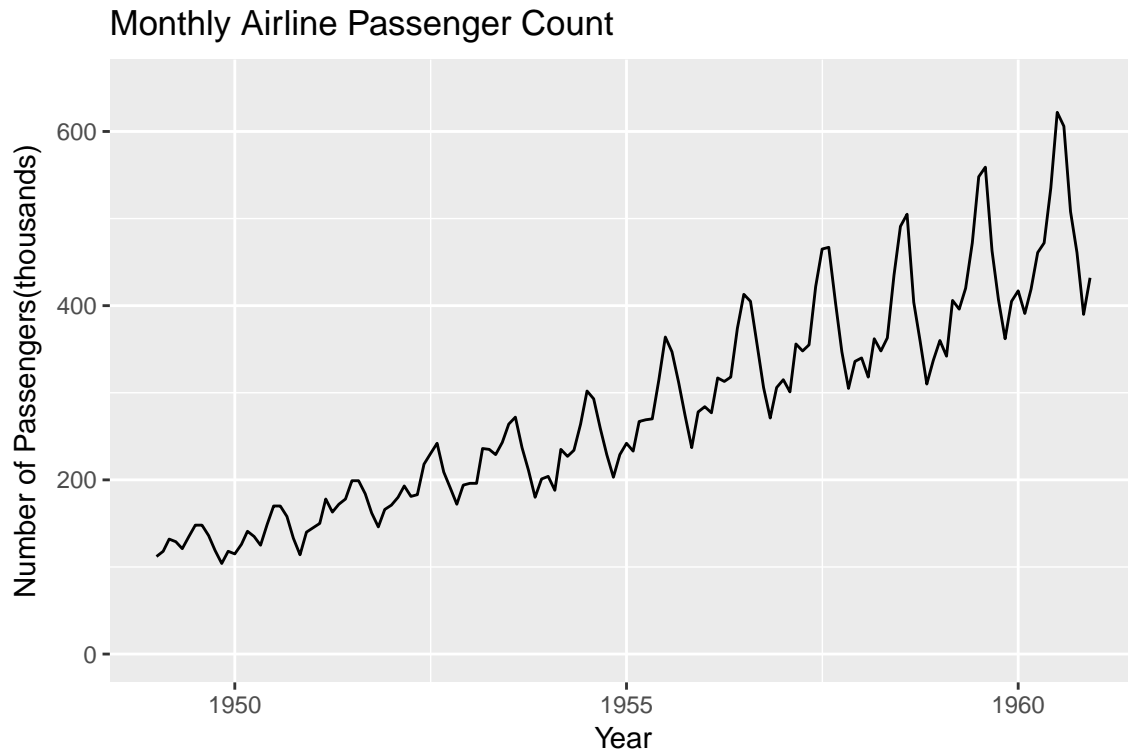
```
library(datasets)
data(AirPassengers)
str(AirPassengers)
```

```
## Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119 ...
```

## Plotting in R

Many time series objects can be directly plotted using the `ggfortify` package.

```
library(ggplot2)
library(ggfortify)
autoplot(AirPassengers) + ylim(0,650) +
  labs(title="Monthly Airline Passenger Count", y="Number of Passengers(thousands)", x= 'Year')
```



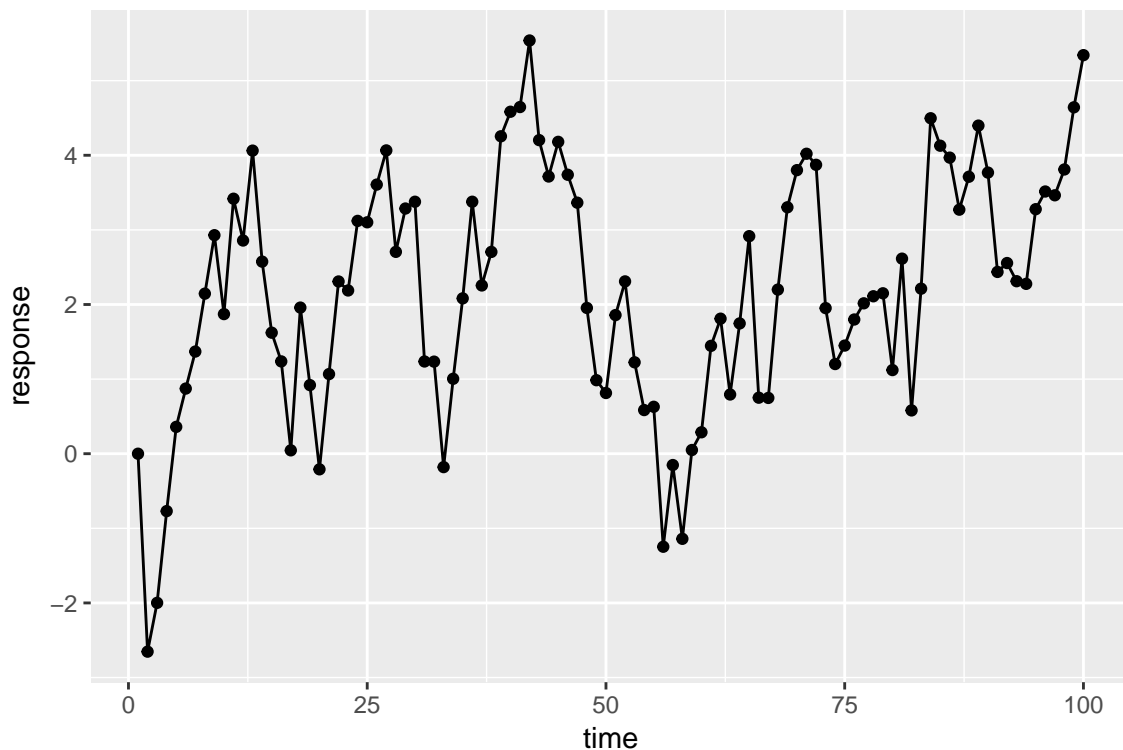
## Simulating discrete-time stochastic process

- We will largely use realistic data to motivate time series models; however, the ability to simulate data with specified characteristics will be useful and illuminate characteristics of time series data.

```
set.seed(08182018)
maxT <- 100
y = rep(0, T)
evolution.var <- 1

# Simulate Data
for (time.pt in 2:maxT){
  y[time.pt] <- y[time.pt - 1] + rnorm(1, 0, evolution.var)
}

# Plot Data
y.dat <- data.frame(response = y, time = 1:maxT)
ggplot(data=y.dat, aes( x=time, y =response)) + geom_line() + geom_point()
```



## Advanced R: with Baltimore Tow

For this example we will work through some basic commands to create a time series plot using a dataset containing information on vehicle towed in Baltimore, Maryland.

For more information on some helpful R packages take a look at the

1. Download the dataset

```
library(readr)
baltimore.tow <- read_csv('http://math.montana.edu/ahoegh/teaching/stat408/datasets/BaltimoreTowing.csv')

## Parsed with column specification:
```

```
## cols(
##   vehicleType = col_character(),
##   vehicleMake = col_character(),
##   vehicleModel = col_character(),
##   receivingDateTime = col_character(),
##   totalPaid = col_character()
## )
```

2. Create a date object

```
library(dplyr)
baltimore.tow <- baltimore.tow %>% mutate(receivingMonth = substr(receivingDateTime,0,2), receivingYear = substr(receivingDateTime,3,4))
```

3. Create an figure

```
library(ggplot2)
ggplot(data = baltimore.tow, aes(receivingMonth) ) + geom_bar(fill = 'dodgerblue4') +
  xlab('Month Vehicle Towed') + ylab('Number of vehicles towed') +
  ggtitle('Count of vehicles towed in Baltimore', subtitle= 'Source: https://data.baltimorecity.gov/Transportation/DOT-Towing/k78j-azhn')
```

