Doing Data Science
Unit 12
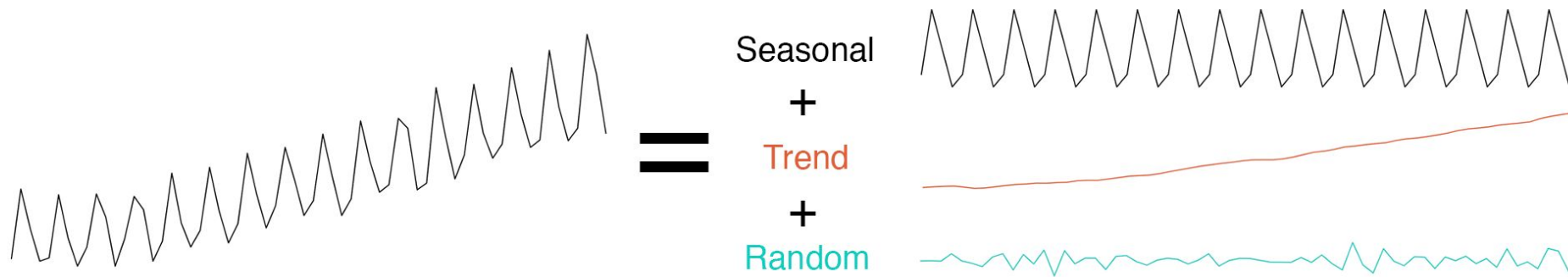
Faizan Javed
Data Science @ SMU

Admin notes

Live Session Unit 10 assignment due next Monday

Live Session Unit 11 assignment due 1 hr before Live Session 14

Case Study 2 by Wednesday (planned)

# Components of a Time Series ref anomaly

Seasonality: how things change within a specific period, e.g., year, month, week, day. Example: US unemployment tends to decrease during end of the year holiday season (should adjust for this to detect overall trend). Can be detected using Fourier transforms

Trend: how things are changing overall, e.g., a website gaining in popularity should have a general trend of increasing number of visitors per month.

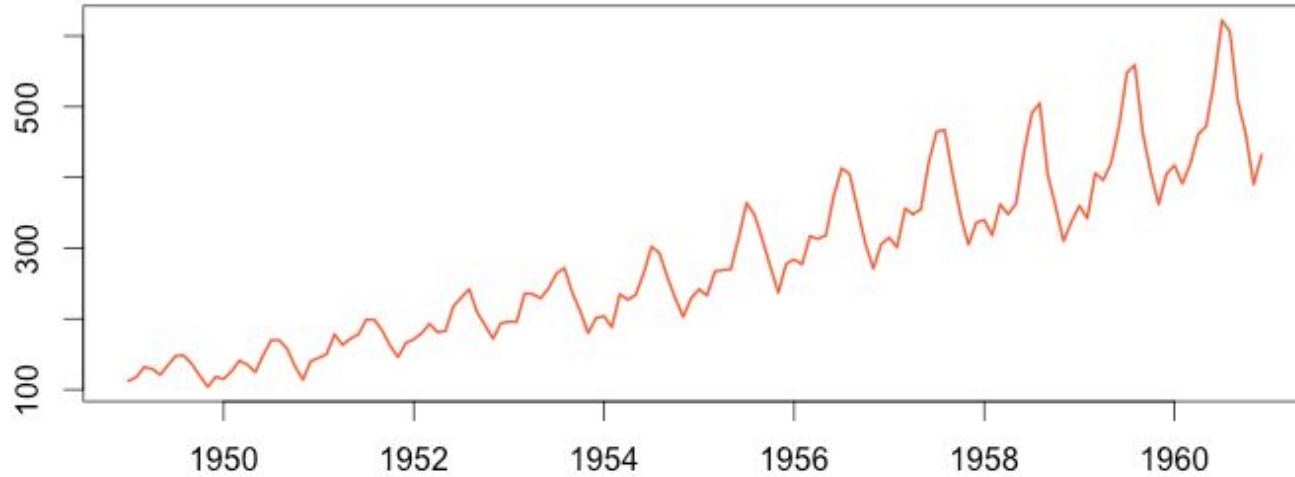**Random/residuals/noise/remainder:** activity not explained by trends or seasonality

Additive and Multiplicative decompositions
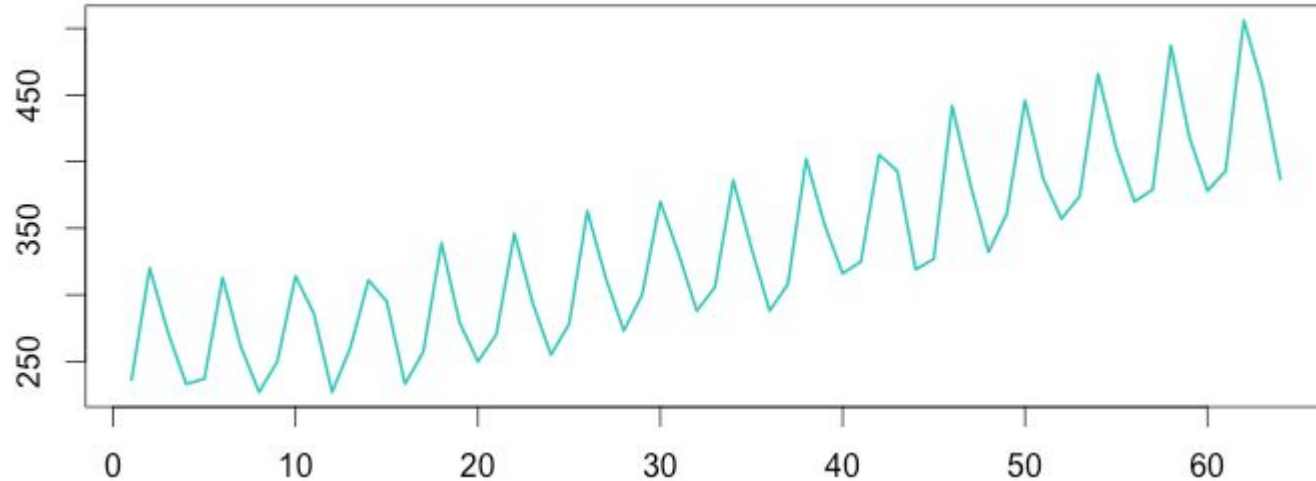
E.g., Australian beer production. Constant seasonal variation. Doesnt change when trend/time series increases.

Additive model:

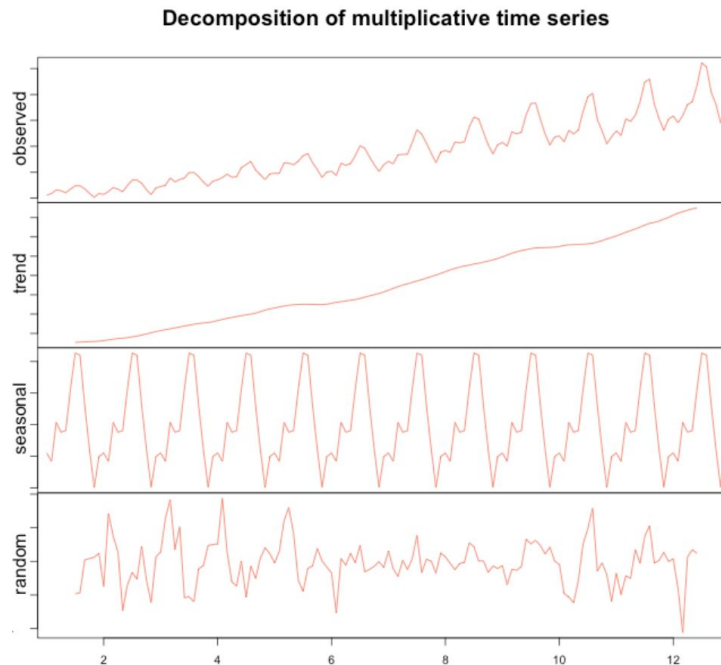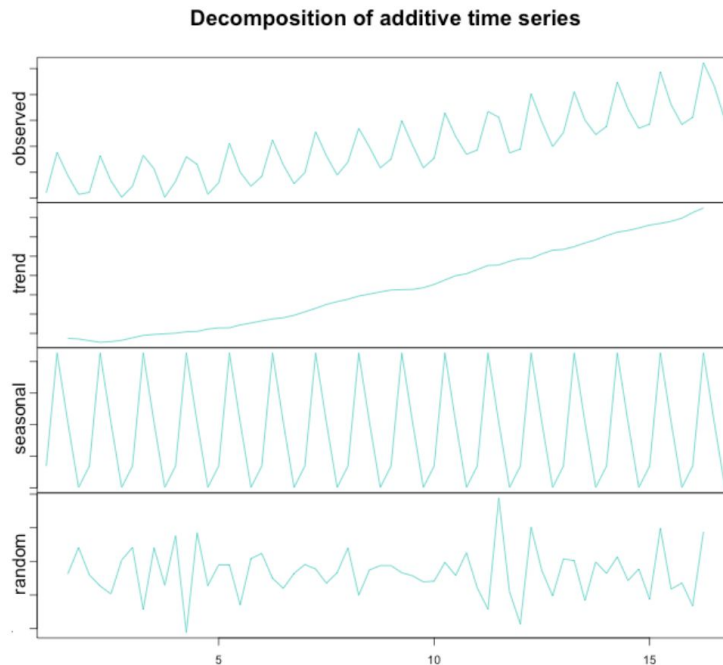Time Series = Seasonal + Trend + Random

E.g., Airline passenger number. Seasonal variation increases with time series increase.

Multiplicative model:

Time Series = Seasonal * Trend * Random

Decompose time series in R

Decomposition of additive time series

Decomposition of multiplicative time series

Use the **Decompose(..)** function or the **STL(...)** function.

Decompose(..) assumes that seasonality component repeats from year to year.
STL might be more robust, while Decompose is easier to use.

Forecasting and smoothing/filtering

Forecasting: predict future values based on previous values and patterns. Fit a model then predict! See the **ses** exponential smoothing forecast function

Smoothing/filtering: removing the seasonal effects of a time series by using a smoothing method, see **HoltWinters(x,alpha,beta,gamma)** function.

It's a zoo in here! ts (base R package) , zoo and xts
https://cran.r-project.org/web/packages/xts/vignettes/xts.pdf
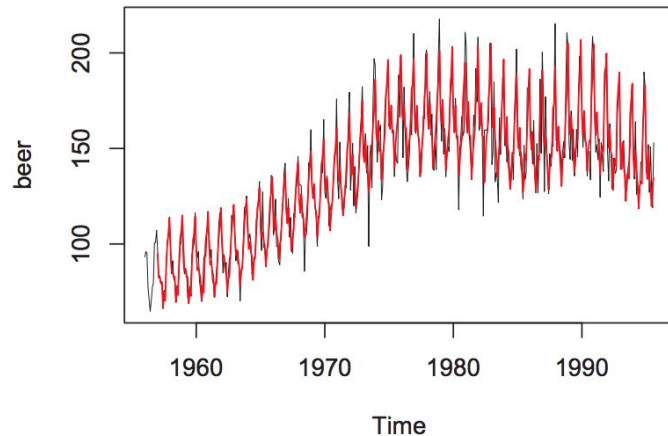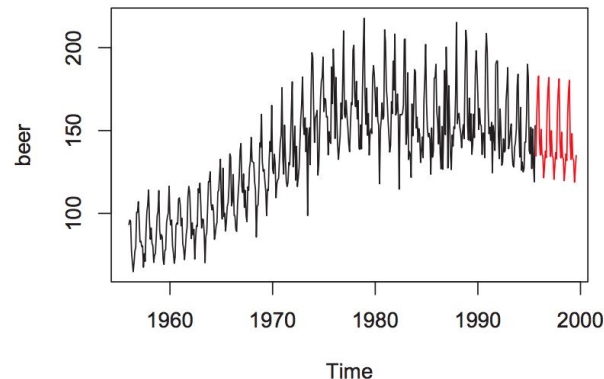
Lookback windows

Simple average: arithmetic average of all previous data points.

Strict rolling/"sliding" window: average of all previous **n** data points

Continuous lookback window/weighted moving average:
Values in the sliding window are given different values
"exponentially downweight" older data,
"Information gradually fades away from people's memories"

Single Exponential Smoothing (SES): weighted average where we consider all data points assigning them exponentially smaller weights as we go back in time converging to zero.

Alpha = "decay" or smoothing factor → jagged lines are smoothed out.

Subsetting time series

Window() (Time Windows): extracts the subset of the object x observed between the times start and end

stocks ← window(NYSE, 1980) #stock price subset from 1980 onwards

Subset() : more robust than Window, allows subsetting by index values and extraction of values of a specific season or subset of seasons in an year.

plot(subset(gas,month="November"))
subset(woolyrnq,quarter=3)
subset(USAccts, start=49)

dygraphs for R http://rstudio.github.io/dygraphs/
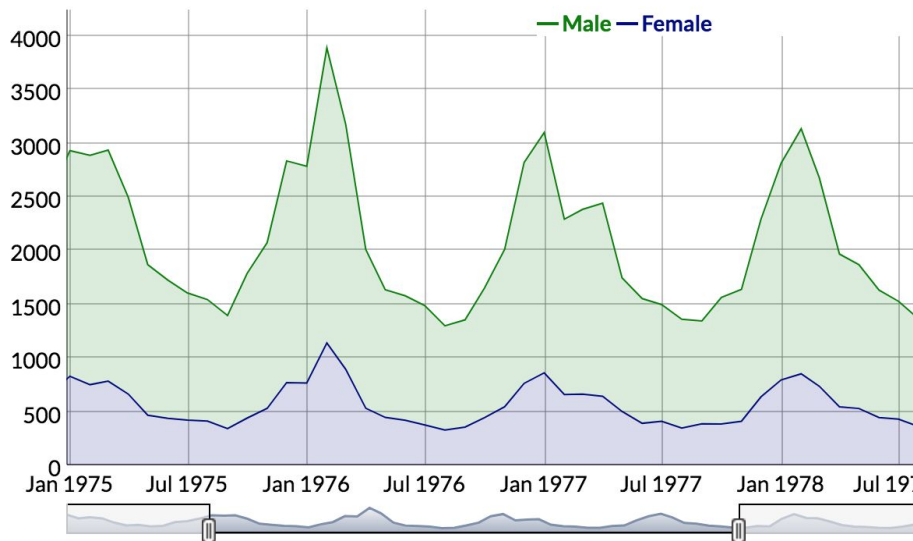
R interface to the dygraphs JavaScript library.

Rich facilities for charting time series & financial data in R

Xts and xts-like objects in interactive charts

Synchronize and zoom-in on time series

Relatively responsive

Can handle big datasets

Unit 12 demo.

What did you learn today?

Questions?