Lab 4

* Considerations
  + Describe all steps building up to final model
  + Introduction
    - What could be the motivation for this exercise?
      * Conduct a monthly, 11-month ahead forecast of the 2015 series
    - What is our sampling interval?
      * Monthly, 1990 - 2015
    - Hypothesis – Expect MA or AR? Seasonality?
      * AR – Spike changes state of system/doesn’t go away
      * MA – Random shock works its way out of system quickly
  + TSEDA –
    - Setup ts with appropriate timestamps
    - Split into Training and Test sets
      * Training = 1990 - Dec 2014
      * Test = 2015
    - Plot the Timeseries
    - Autocorrelation (ACF/PACF)
      * MA
        + ACF drops to zero after p’
        + PACF drops slowly
      * AR
        + ACF drops slowly
        + PACF drops rapidly after q’
      * Look also at seasonality
    - Series Stationarity
      * Decompose() for:
        + Trend

Linear

Non-Linear

* + - * + Seasonality

Stationary in the Mean

Each season has its own mean value

* + - * Moving Walk
        + Dickey Fuller
      * Visualize Differenced Data
        + Apply Filter for underlying trend
  + Transformations Required for Stationarity
    - Trend in data?
      * Subtract trend
      * Apply difference (d)
    - Season in data?
      * Subtract seasonality
      * If seasons non-stationary, apply difference (D) (Is this correct?)
    - Variance constant?
      * Log transform
    - Evaluate with Filtered Plot
  + Model Selection
    - Options Available
      * Linear/Seasonality Decomposition with ARMA of residuals
      * Smoothing with ARMA of residuals
      * SARIMA
    - Evaluation
      * Model Roots
        + Stationary?
        + Invertible?
      * Residual Analysis
        + Error Series Plot

Should approximate white noise series (zero mean, independent)

Homoskedastic

Multi-panel plot

* + - * + Error Series Autocorrelation (ACF/PACF)

No spikes

Ljung-box test

* + - * AIC (two types)/BIC
      * Mean Absolute Percentage Error (MAPE)
      * SSEs of out of sample testing
        + Visualize SSE over time, which performs better short-term vs. long-term?
        + What is the most appropriate timeframe?
      * Cross-validation for time-series
    - Summary
      * For-Loop to calculate AIC/BIC/SSE/MAPE, etc.
      * Put in Table and visualize
    - Reflection
      * Is there an intuitive explanation for the chosen model?
      * Is it parsimonious?