Introduction:

1. Is a recession coming? Looking at time series graphs
   1. Questions to ask: What makes a time series different from other data we have seen?
   2. Slides 🡺 time has to be plotted on the X-axis. It violates the idd tenet.
2. Let’s take a look at a hurricane forecast
   1. Questions to ask: Where is that hurricane/storm heading? How do we know where it’s going?
   2. Forecasting has probabilistic nature. What’s the probability that it will be here? Or here? 95% confidence interval; 90%? 80%?

That’s the nature of this class. Two goals: 1) understand structure of time series and 2) use that structure to forecast what will happen with some level of confidence.

Decomposing Time Series

1. Base R handles basic time series data.
2. There’s also the ts package.

Let’s take a look at the time series plot for AP. What do you see?

What are some things you see?

1) Is there a trend? What is no trend? I draw a line through it and it's a straight line with mean = 0

2) The series seems to be repeating itself. There's seasonality

Seasonality that is trending!!

3) The variance is getting bigger and bigger. My peaks are getting higher; My valleys are getting bigger too. Nonconstant variance. Do I have a cyclical nature? Let's take a look at the US business cycle.

If I can find all the underlying time structures, then I can predict with some level of certainty what will happen.

What I just did to the time series is to "decompose" it.

What non-fancy/basic forecasting models can I deploy?

Four basic forecasting methods:

-Average/Mean

-Naive ==> Miss out on seasonality/trend/cyclical

-Seasonal Naive ==> Replicate the seasonality in the last year ==> Miss out on trend/cyclical

-Drift ==> Allow the forecast to increase or decrease over time, where the amount of change over time (drift) is set to be

the average change seen in historical data. Essentially drawing a line between the first and last observation and letting it "drift" forward.

Be careful to the “window” of your time series. Data range matters a great deal in the forecasts you produce!

Measuring model performance

Train/test: 80/20

100% accuracy means actual = forecast. Error (residual) is 0.

Rules for residuals:

Mean = 0

Uncorrelated

Constant variance

Normally distributed

We can use many measures of performance: RMSE is the standard.

Let’s take a look at another package that handles time series

Zoo

lag --> using yesterday's value to set as today's value (This is the Naive method)

Finding a lagged structure

Because what is happening today is contigent on what happened yesterday!