Time Series

Explain vrs. Predict

Explanation is the goal of "time series **analysis**"

- Models are based on causal argument
- Models are not "black-box"

Forecasting (our focus) seeks to **predict** future values

Time Series Components

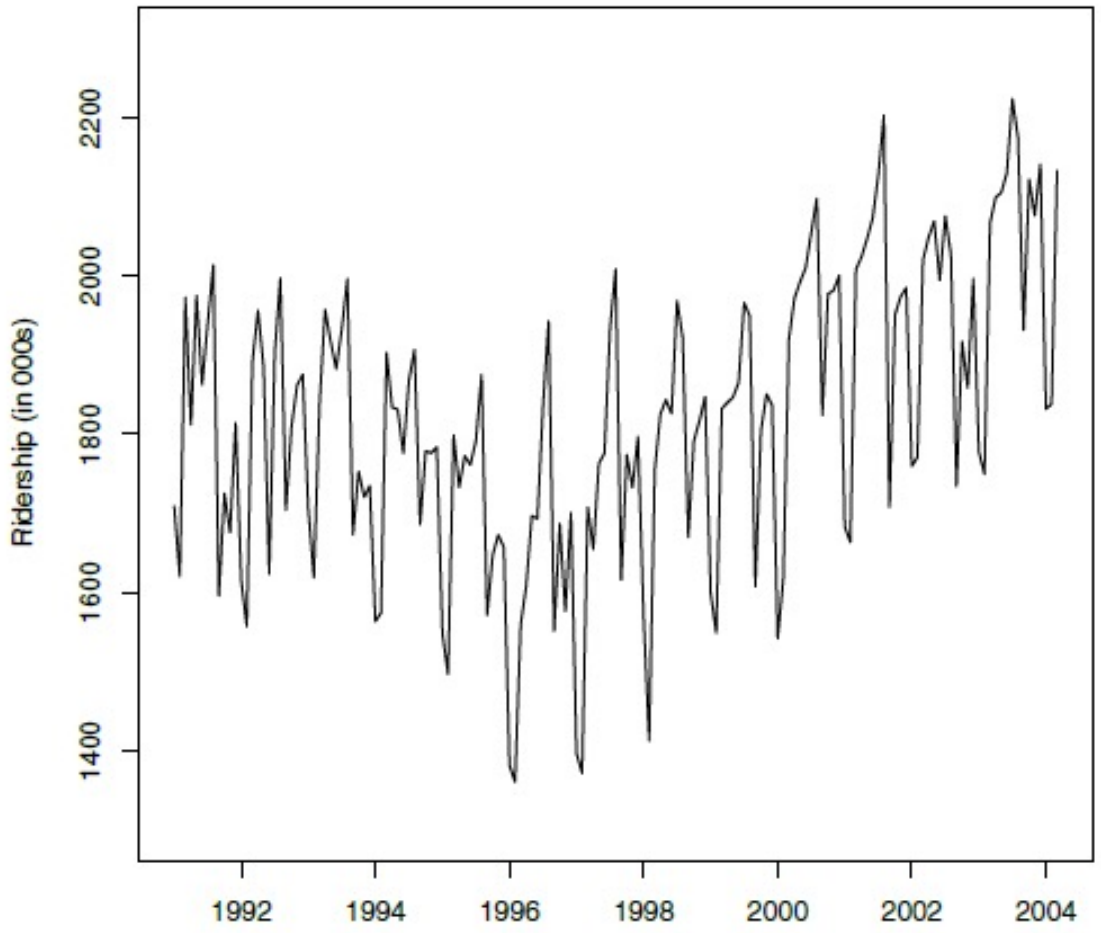
Level

Trend

Seasonality

Noise

Amtrak Ridership



Zoom to 3 years (1997-1999)

Seasonality* appears:

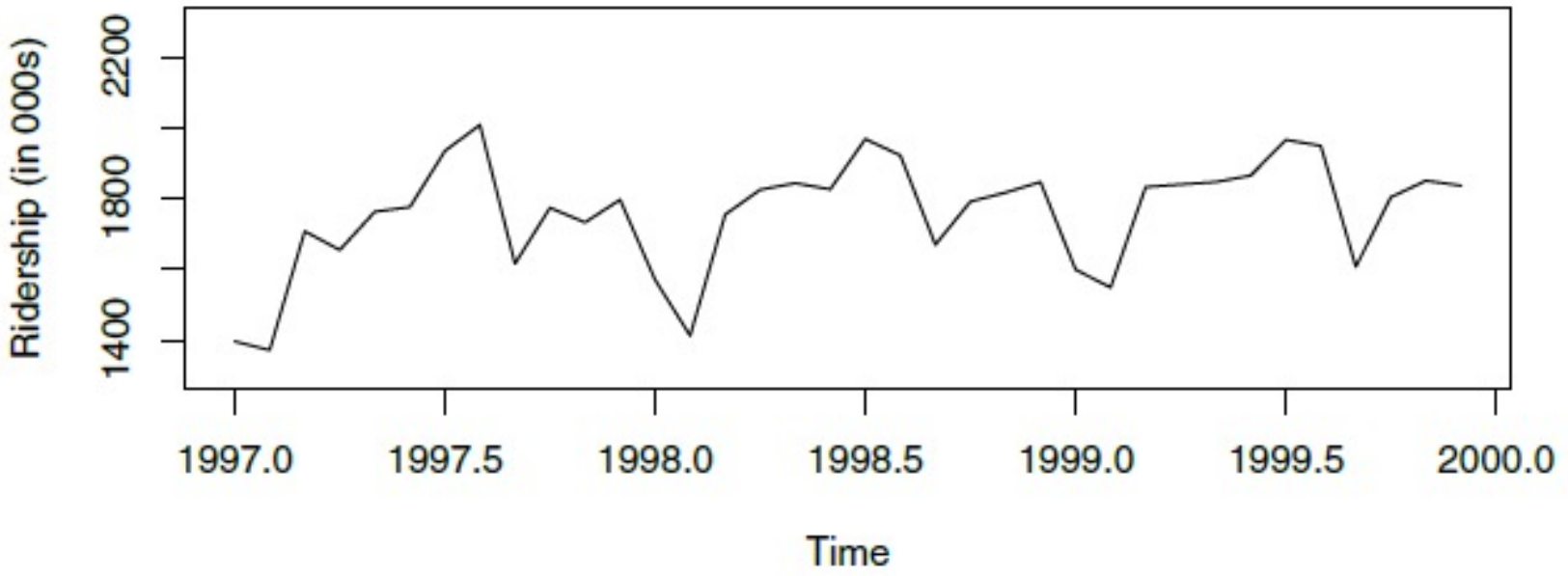
Each year traffic peaks in summer

Noise:

Departure from the general level that is neither trend nor seasonality

* Don't confuse the time series term "season," which is the period over which a cyclical pattern repeats (e.g. a year), with the standard English seasons of the year (fall, winter, etc.)

Amtrak Ridership – zoom to 3-years



Partitioning

Divide data into training portion and validation portion

Test model on the validation portion

Performance can be assessed against the “naïve benchmark” – *naïve forecast* is simply the most recent value in the time series

Partitioning is not random

Random partitioning would leave holes in the data, which causes problems

- Forecasting methods assume regular sequential data

Instead of random selection, divide data into two parts

- Train on early data
- Validate on later data

Key Takeaways

Focus is to predict (not describe/explain)

Four components

- Level
- Trend
- Seasonality
- Noise

Partition data by dividing into early/late