**Case Assignment #4**

We are going to use ARIMA to model the seasonal flu.

1. Extract historical flu data (positive cases over time) – you can choose to model flu patterns at the national or regional/state-level
2. Build an ARIMA model; determine the appropriate values for (p,d,q)
3. How well does your model perform on validation data? (Note: you’ll need to create a training and validation set to measure forecast accuracy)
4. Provide analysis to support your determinations

**Data source notes**

* Beginning in the 2015-16 season, reports from public health and clinical laboratories were presented separately in the weekly influenza update, FluView.
* Data from clinical Introduction About the Data laboratories include the weekly total number of specimens tested, the number of positive influenza tests, and the percent positive by influenza type.
* Data presented from public health laboratories include the weekly total number of specimens tested, the number of positive influenza tests, and the number by influenza virus type, influenza A subtype, and influenza B lineage
* Are these two sources inclusive or exclusive?

**Data Resources**

**From CDC:**

**FluView Interactive (Landing Page)**

<https://www.cdc.gov/flu/weekly/fluviewinteractive.htm>

**FluView Interactive (Application)**

<https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>

**FluView Application Quick Reference Guide**

<https://gis.cdc.gov/grasp/fluview/FluViewPhase2QuickReferenceGuide.pdf>

**cdcfluview** (<https://cran.r-project.org/web/packages/cdcfluview/cdcfluview.pdf>)

**From WHO:**

**FluNet Database (Landing Page)**

<https://www.who.int/influenza/gisrs_laboratory/flunet/en/>

**Download influenza laboratory surveillance data from any week (Database)**

<https://apps.who.int/flumart/Default?ReportNo=12>

**Supplemental Resources**

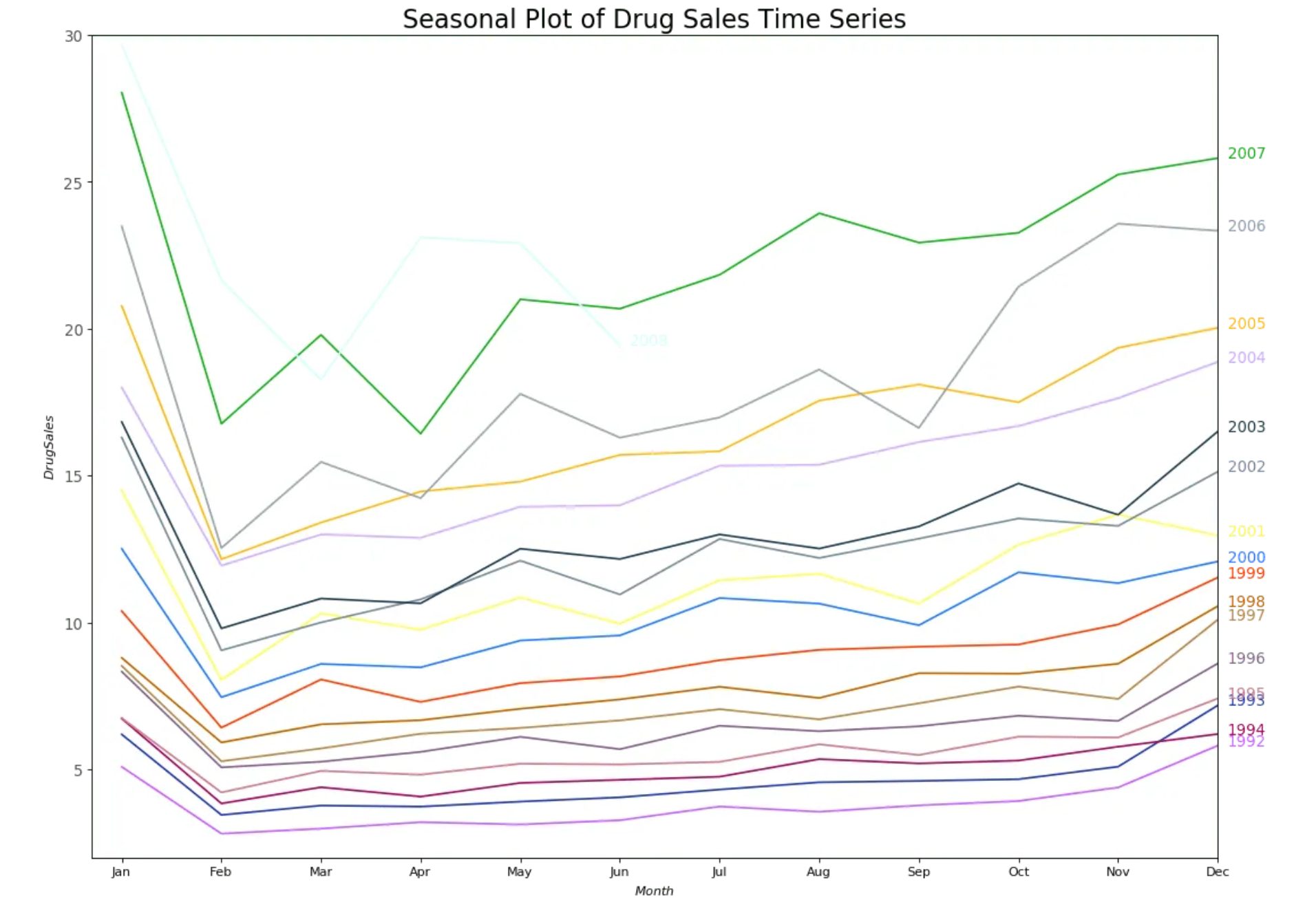
**Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011**

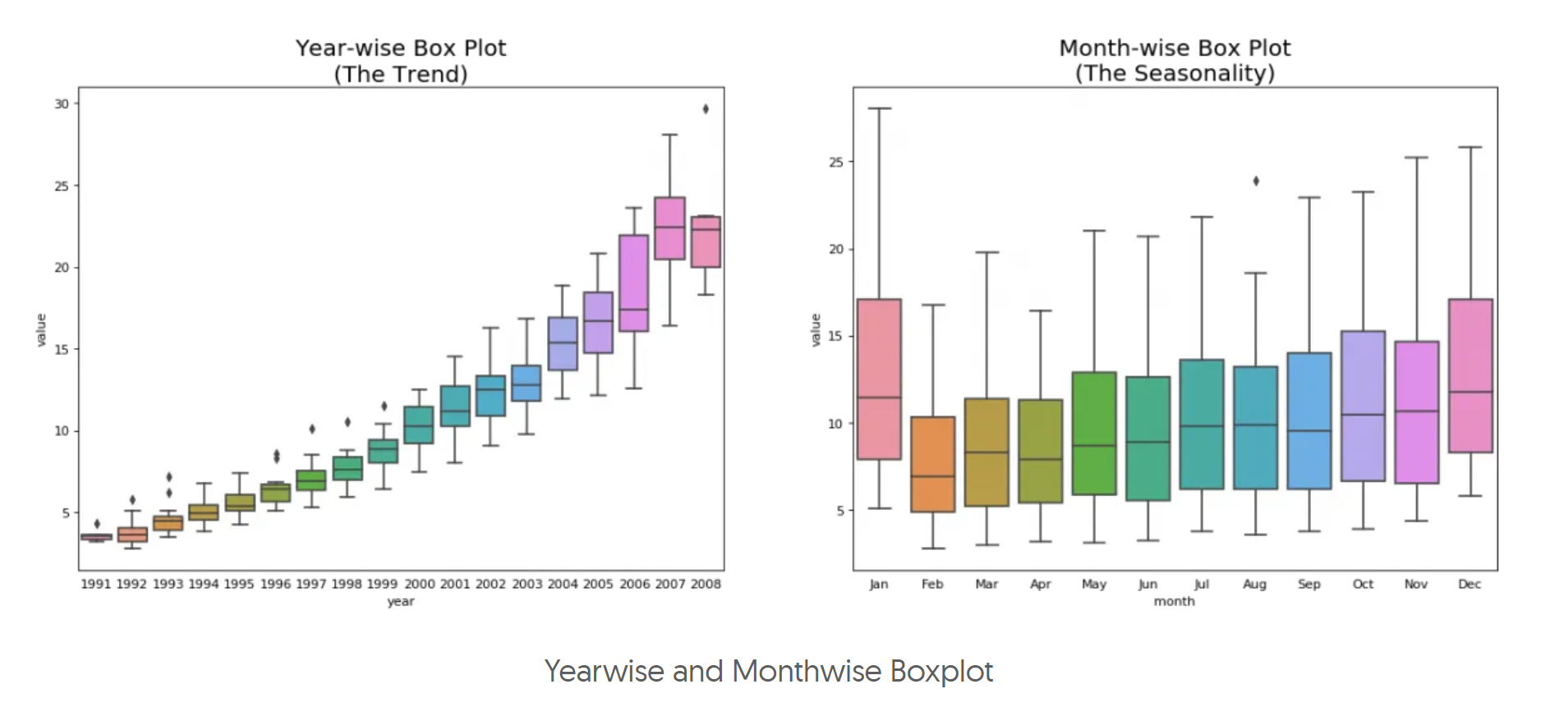
<https://journals.lww.com/md-journal/fulltext/2016/06280/time_series_analysis_of_influenza_incidence_in.15.aspx>

**ARIMA Model – Complete Guide to Time Series Forecasting in Python**

<https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

NW Notes – could use following plots for EDA





Concepts:

Cyclical, not necessarily at consistent/regular intervals

Vs

Seasonal,

Dickey-Fuller is a Unit Root Test

Impute Strategies:

Backward fill

Linear Interpolation

Quadratic Interpolation

Means of nearest neighbors

Means of seasonal counterparts

.. test with mean square error ..

<https://www.machinelearningplus.com/time-series/time-series-analysis-python/>

NW notes - framework from article:

- read data, calc basic summary

- check time series data cycle, plot raw

- decompose time series data

- test for stationarity

- fit model

- forecast

OR

- visualize time series

- stationarize series

- Plot ACF/PACF find optimal parameters

- Build ARIMA model

- Forecast/make predictions

p-value > 0.05: Fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary.p-value <= 0.05: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.