Class Participation 1

You are asked to predict the annual number of visitors to Zion National Park for the next five years so the Park Service can plan and prepare for the future using data from the US National Park Service.

Your analysis should include:

- Data: The US National Park Service collects an impressive amount of data on the number of visitors to our national parks. For this assignment use the National Park Service Visitor Use Statistics webpage at `https://irma.nps.gov/Stats/` for data from a selected national park.

  Download the **Annual Park Recreation Visitation** data and write documented reproducible code to create an appropriate R dataframe called `zion`. Change the units of `RecreationVisitors` to 'millions of visitors.'

  The column `RecreationVisitors` has commas at thousands and millions to make the table easier to read. R reads these as factors (not numbers) but the values can be converted to numeric using

  ```
  zion$RecreationVisitors<-as.numeric(gsub(',','',zion$RecreationVisitors))
  ```

- EDA: Plot the time series for annual visitors (in millions).
- Analysis:
  - Is there evidence of a multiplicative (not additive) time series model? What could you do to the data to better comply with time series model assumptions?
  - Is there evidence of a non-constant mean change? What are possible causes for this effect? How does filtering to the 'recent past' improve agreement with the constant mean change assumption and reflect the future Zion tourism demand?
  - After transforming and filtering the data to the 'recent past,' fit an ARIMA(1,1,1) model and report the parameter estimates and standard errors.
  - Forecast the annual visitors (in millions) for the next 5 years. Report a table of predictions and 95% prediction intervals.
  - Create a publication quality graphic showing the historical values and the 5 year predictions (with uncertainty) of annual visitors (in millions).

- Research Task and Data Features that Match Analysis Strengths:
- Analysis Weaknesses:

CHALLENGE: Provide a research task and find data with the same characteristics as this assignment.

Class Participation 2

You are asked to predict the **monthly** US residential energy consumption for the next 2 years for the 'Short-Term Energy Outlook' using data from the US Department of Energy. Use a model that captures the pattern within a year due to heating and air conditioning in response to temperature variation.

Your analysis should include:

- Data: The US Department of Energy collects an impressive amount of data to support the energy industries. For this assignment use the monthly total energy consumed by the US residential sector since 1991. The webpage `https://www.eia.gov/totalenergy/data/browser/?tbl=T02.02#/?f=M` contains monthly web tables but an easier approach to getting the data into R uses the result of the 'Download' button.

  The following code help creates `energy` from the CSV download file, selects the data in column TERCBUS (in trillion Btu), filters to current data, and removes the annual total rows (coded 'month 13').

```
# US residential energy consumption
# http://www.eia.gov/totalenergy/data/monthly/index.cfm#consumption
data1<-read.csv("http://www.eia.gov/totalenergy/data/browser/csv.cfm?tbl=T02.01")

# subset to TERCBUS Total Energy Consumed by the Residential Sector
data2<-subset(data1,MSN=="TERCBUS")

# subset to "your lifetime"
data3<-subset(data2,data2$YYYYMM>199100)

# remove yearly total (coded "month 13", every 13th obs)
data4<-subset(data3,data3$YYYYMM%%100 != 13)

energy<-data4$Value
```

- EDA: Plot the time series for US residential energy consumption (in trillion Btu). Does it appear there is a monthly pattern? If so, explain the possible business/economic 'cause' for this effect.

- Analysis:
  - Fit an ARIMA$(1,1,1) \times (1,1,1)_{12}$ model and report the parameter estimates and standard errors.
  - Forecast the US residential energy consumption for the next 2 years. Report a table of predictions and 95% prediction intervals.
  - Create a publication quality graphic showing the historical values and the 2 year predictions (with uncertainty) of US residential energy consumption.
- Research Task and Data Features that Match Analysis Strengths:
- Analysis Weaknesses:

CHALLENGE: Provide a research task and find data with the same characteristics as this assignment.

Homework

1. ARIMA time series models make assumptions about the possible trend, stationarity, and distribution of the data. For each of the following national parks:

   - Download the Annual Park Recreation Visitation data and write documented reproducible code to create an appropriate R dataframe.
   - Plot the time series for annual visitors (in millions).
   - Identify a significant feature of the time series that violates the ARIMA model assumptions. Research why the data exhibits this unusual feature.

   (a) Washington Monument

   (b) Jefferson Memorial

   (c) Wright Bros Memorial

   (d) Acadia
   Note: You may find the following article by Mark Good of the *Mount Desert Islander* interesting:
   `http://www.mdislander.com/acadia-centennial-2/visitation-not-visitors-tallied-park`

2. Choose a national park that you have some interest in (one you've been to or would like to visit) from the following list:

   - Valley Forge
   - Lincoln Memorial
   - Theodore Roosevelt Island
   - LBJ Memorial Grove on the Potomac
   - Ft McHenry
   - Blue Ridge Parkway
   - Timpanogos Cave
   - Golden Spike
   - Capitol Reef
   - Arches

- Bryce Canyon
- Muir Woods

(a) Where is it? Why did you go or why do you want to?

(b) Write a paragraph about the creation of the national park.

(c) Download the Annual Park Recreation Visitation data and write documented reproducible code to create an appropriate R dataframe.

(d) Plot the time series for visitation (in millions) and comment on the validity of the ARIMA model assumptions.

(e) If you feel filtering to the 'recent past' or a transformation is necessary, explain what you observed and demonstrate why it is better after filtering and/or transforming.

(f) Fit an ARIMA(1,1,1) model and report the parameter estimates and standard errors.

(g) Forecast the annual visitors (in millions) for the next 5 years. Report a table of predictions and 95% prediction intervals.

(h) Create a publication quality graphic showing the historical values and the 5 year predictions (with uncertainty) of annual visitors (in millions).