Assignment 3
[20 points]

Instructions

1. You may use python or R. If you are using Python, use Jupyter notebook. If you are using R, use the R notebook by R studio. Both are equivalent.
2. Create a Github private repo where all your assignments and projects would be stored. At an opportune time, you would be asked to share your repo with our evaluation team.
3. Your submission should be sufficiently original to be considered for evaluation.
4. **Submission would include source code, data availability at Github as well as Google classroom submission of the markdown pdf (do not submit any other format on Google classroom).**

Submission date and time

September 2, 2019, by 6 PM.

Q1. Which of these time series are stationary? Explain. [5 points]

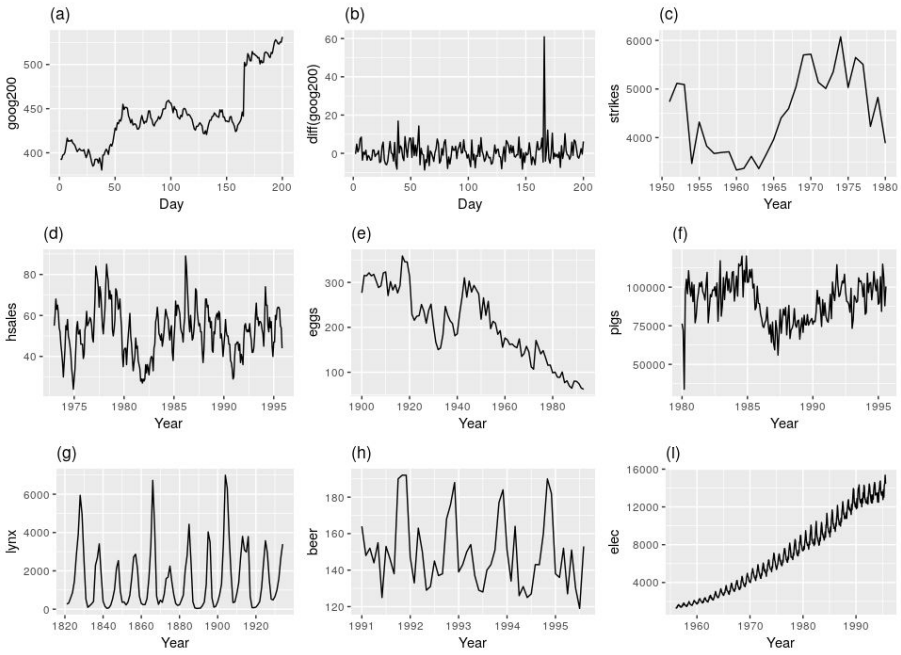

Figure 1. (a) Google stock price for 200 consecutive days; (b) Daily changes in the Google stock price for 200 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

Q2. https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/ covers basic concepts and steps involved in time series data analysis. The key concepts include.

a. time-series smoothing
b. stationarity
c. Autocorrelation
d. acf() and pacf()
e. fitting time series model

Some additional references are:

https://www.kaggle.com/kashnitsky/topic-9-part-1-time-series-analysis-in-python
https://otexts.com/fpp2/stationarity.html

Acquire the past three years worth stock price data of Amazon, Microsoft, and Google.

a. Visualize the time series. Comment on current health of each of the stocks
b. Assess autocorrelation
c. Assess stationarity
d. Fit ARMA and assess the predictability of each stock for an unseen period.

[10 points]

Q3. By now you must be familiar with concepts such as - logistic regression, glm, odds ratio, confidence interval. Download the BreastCancer data from mlbench package in R (https://www.rdocumentation.org/packages/mlbench/versions/2.1-1/topics/BreastCancer). It uses various tumor descriptors to predict if it's benign or malignant. Depending on the distribution of decision variables, use glm function with an appropriate link function to construct a predictive model. Explain the summary statistics returned by glm function. Interpret the odd ratio corresponding to each explanatory variable. Report the 95% CI of the odd ratios.