ISA 444: Business Forecasting

03 - Time Series Plots

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Spring 2021

Outline

What we Covered Last Week

Main Learning Outcomes

☑ **Describe course objectives & structure.**

☑ **Describe what do we mean by forecasting and Explain the PIVASE framework**

☑**Explain the differences between cross sectional, time series, and panel datasets.**

☑ **Identify and describe the basic components of a time series including trends, seasonal components, and cycles.**

A Few Comments on Assignments 01 and 02 [1]

1. I have received several emails from you pertaining to both assignments. I believe I have answered all of them, either by directly responding to your emails or by meeting with you during office hours. **As a reminder, I use the Quiz functionality in Canvas to capitalize on the auto-grading; however, these are assignments and not quizzes and hence, you can ask me questions.**
2. There were 4-5 students who continued to have issues with package installations. In a couple of instances, I was able to help the students directly. In the other instances, the work around was to use the FSB Virtual Desktop to run our code.
    a. As a reminder, if you were to use the FSB Virtual Desktop, please ensure that all your code is saved on the M Drive so you do not lose that information. Additionally, you may need to install packages every time (not 100% sure on that), if your R profile does not carry over from one session to the next.

A Few Comments on Assignments 01 and 02 [2]

b. This information should not be new to you as it was clearly highlighted in the STUDENTS START HERE: First Steps page on Canvas, which is part of the Week 0: Getting Started module.

3. For every assignment, the correct answers (with code and/or references) are made available to you. So please check them to learn from any potential mistakes made. Note that any code that I provide captures **ONE** approach to solving the problem at hand, and there are typically **a large number of ways** that you can approach the problem.

4. The assignments are an indispensable learning component of this course. Some of them will challenge you, and that is okay. **You are here to be challenged, develop new skills, and learn new methods.**

Learning Objectives for Today's Class

Main Learning Outcomes

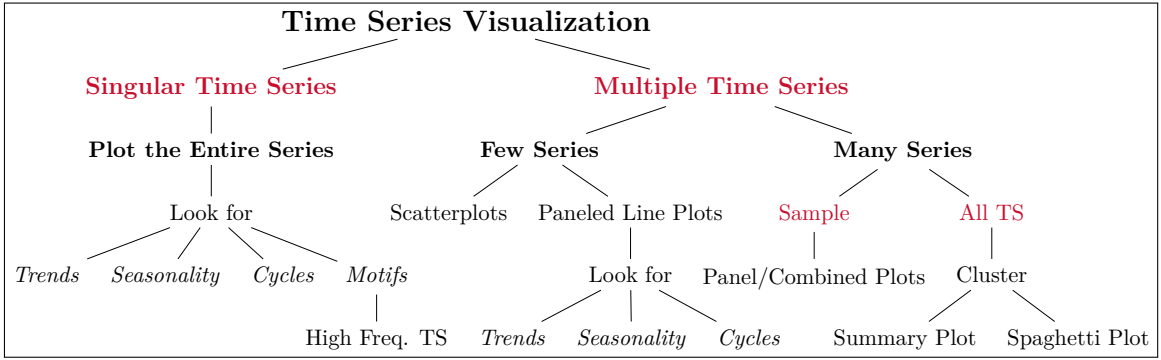
- **Explain different goals for visualizing time series data**
- **Identify an appropriate chart for a specific time series data visualization goal**
- **Use software to construct charts of interest**

Outline

Exploratory vs Confirmatory Visualization Goals

Visualizations can be used to explore data, to confirm a hypothesis, or to manipulate a viewer. . . . In ***exploratory visualization the user does not necessarily know what he is looking for****. This creates a dynamic scenario in which interaction is critical. . . . In a* ***confirmatory visualization, the user has a hypothesis that needs to be tested****. This scenario is more stable and predictable. System parameters are often predetermined. — (Grinstein and Ward 2001, 22)*

A Structured Approach for Time Series Visualization

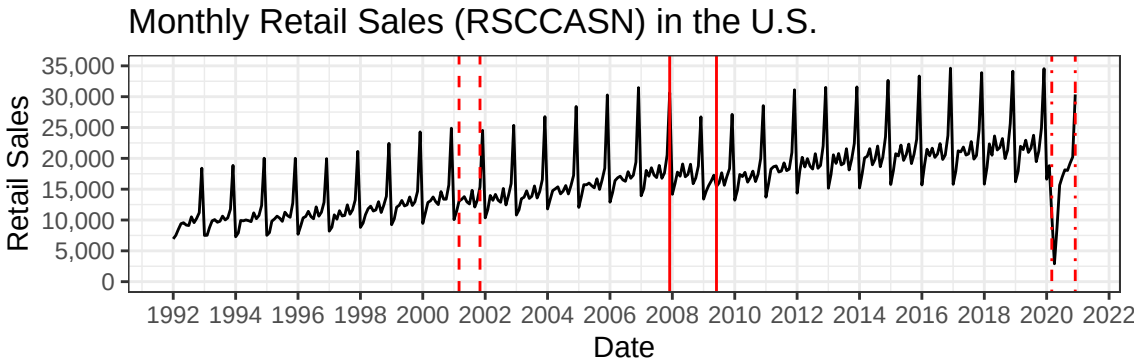


A Potential Framework for Time Series Visualization.[1]

[1]This is my best attempt to improve on the general advice provided in the previous slide. Many of the suggestions, presented in this flow chart, stem from my past and current research/consulting collaborations. They are by no means a comprehensive list of everything that you can do.

Outline

Plot the Entire Time Series [1]



Question 1: Ignoring the economic downturns in between each pair of red lines, how would you characterize this time series? Please provide your answer at www.menti.com.

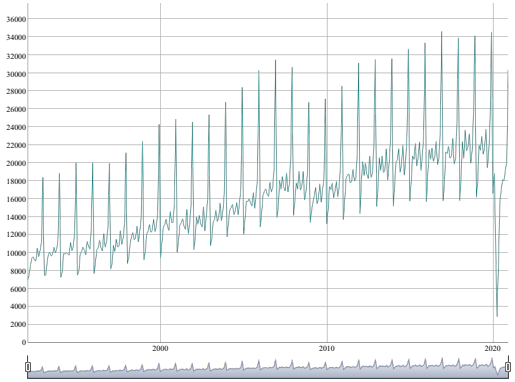
Plot the Entire Time Series [2]

Question 2: Compared with the visual in the last side, please:

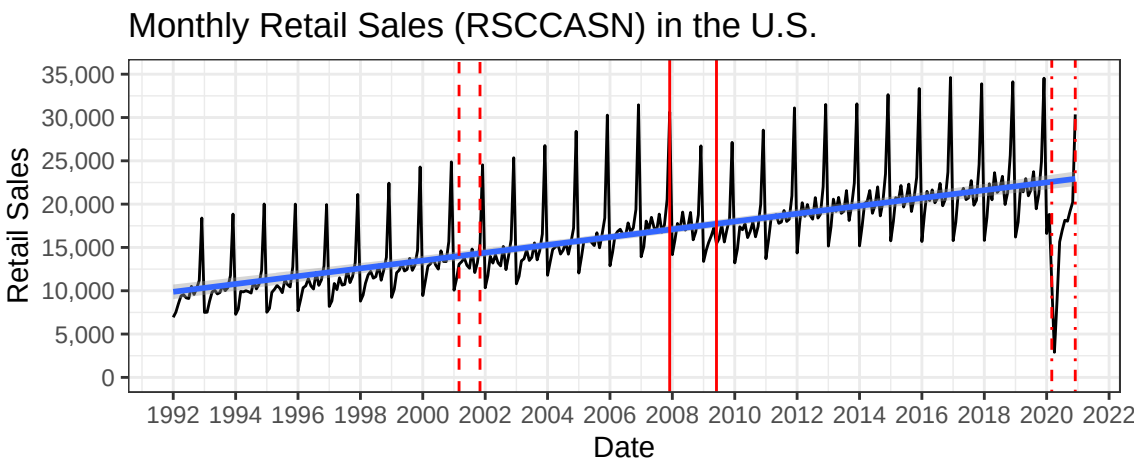
- Highlight 1-2 reasons why the visual provided by FRED is superior. The FRED chart can be accessed at https://fred.stlouisfed.org/series/RSCCASN.
- Please provide your response at www.menti.com.

Interactive Time Series Plots in R

Using the Markdown template, let us create an interactive plot for the retailSales data. The chart below represents a static version of one of two charts, which we will create in our knitted HTML document.

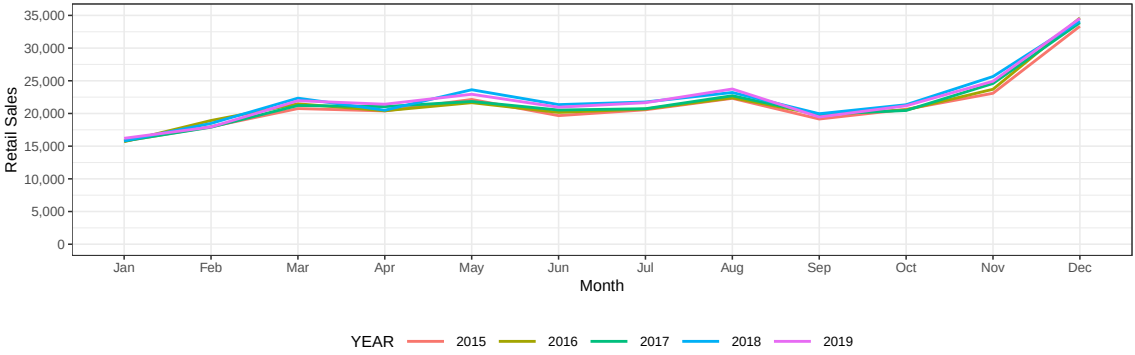


Looking for Trends



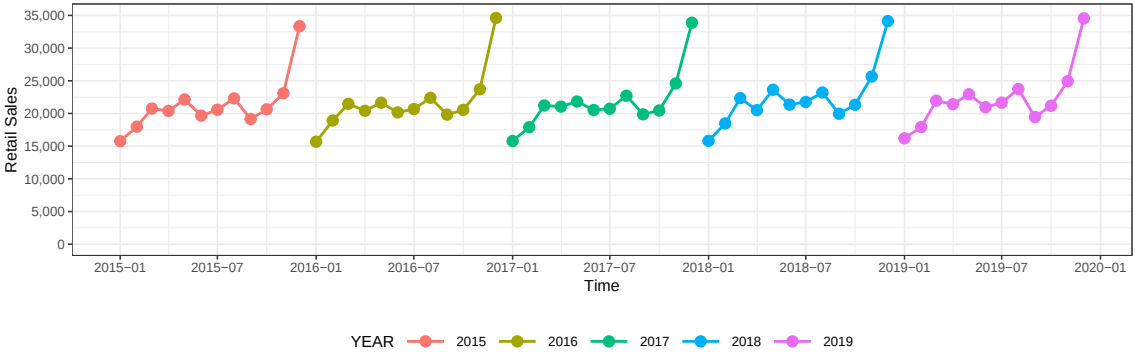
Looking for Seasonality [1]

In Section 2.2.1 of our textbook, the authors presented two approaches for considering seasonality. We can replicate them for our data. For simplicity, let us use the data from 2015-2019 (post housing crisis but before COVID impacting the retailSales).

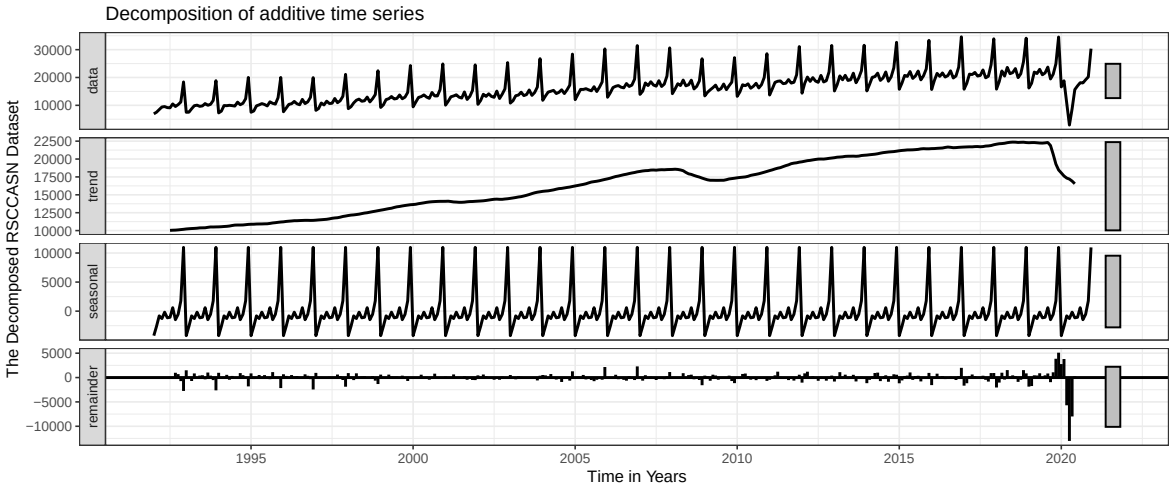


Looking for Seasonality [2]

In Section 2.2.1 of our textbook, the authors presented two approaches for considering seasonality. We can replicate them for our data. For simplicity, let us use the data from 2015-2019 (post housing crisis but before COVID impacting the retailSales).



A Statistical Decomposition of the Entire Time Series



Looking for Motifs: Definitions

What do we mean by motif discovery?

Motifs allow us to cluster subsequences of a time series. It is a popular (unsupervised) learning approach, where patterns are automatically detected in time series.

To illustrate the idea, let us start with a text example.[2] If you have parallel texts, then over time you can learn a dictionary with high accuracy.

. . . And God said, “Let there be light”; and there was light. And God saw the light, that it was good. And God . . .

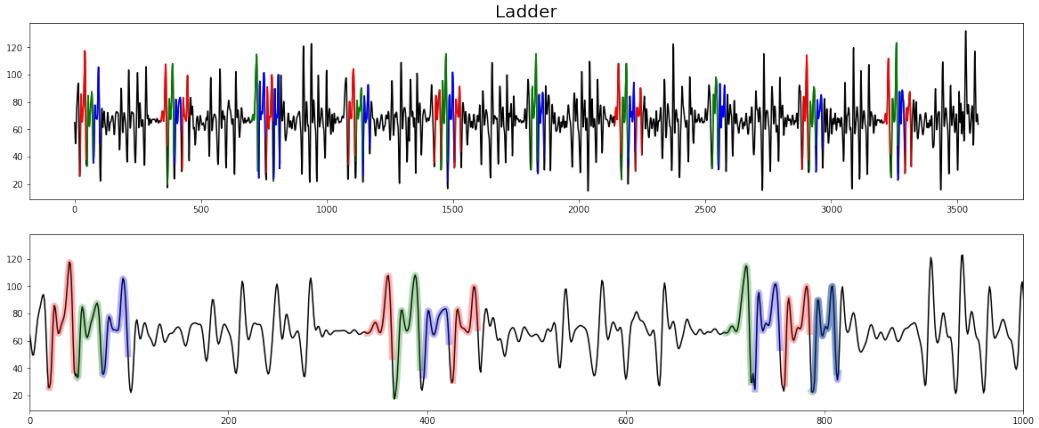
. . . Y dijo Dios: Sea la luz; Y fue la luz. y vio Dios que la luz era buena . Y llamó Dios a . . .

Note the mapping is non-linear, the learning algorithms in this domain are non-trivial.

Suppose however that the unknown “language’ ’ is not discrete, but real-valued time series? **In this case, repeated pattern discovery can help.**

[2] Mueen and Keogh (2015). “Finding Repeated Structure in Time Series: Algorithms and Applications”. http://www.cs.unm.edu/~mueen/Tutorial/SDM2015Tutorial2.pdf.

Looking for Motifs based on Wearable Sensors Data



Three dominant motifs discovered in an electrical utility application.[3]

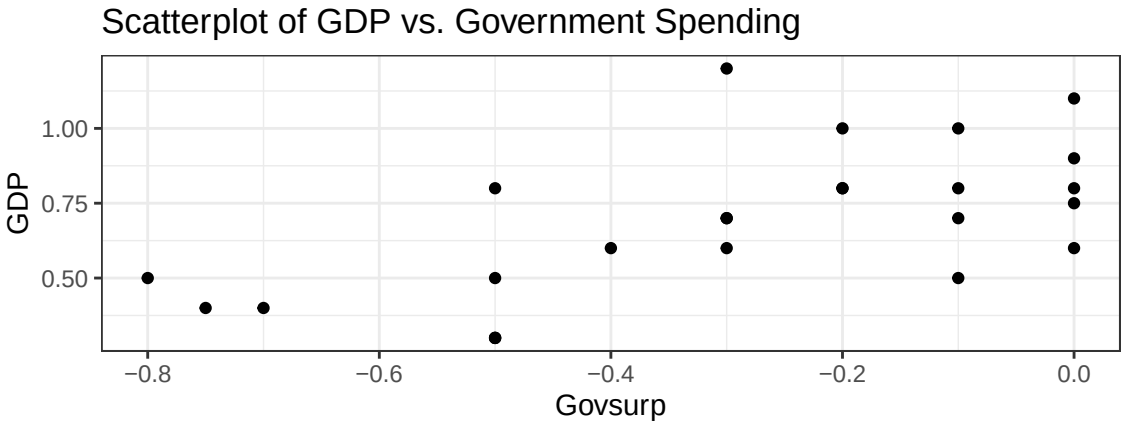
[3] Joint Work with GE Research and the University at Buffalo.

Outline

Scatterplots [1]

- **Scatterplots** are frequently used to visualize the correlation between two continuous variables.
- In this Example, we will be using the German_Forecast Data. The file can be downloaded to your working directory, using the `download.file()` from base R.
- Note that the data is an xlsx file, which would require us to use the `read_excel()` from the readxl package.
- We will remake the plot of GDP vs Govsurp (Figure 2.4 in our textbook) using R. As noted in the chapter, the figure was created using Minitab for the book.
- The plot using the `ggplot()` is shown in the next slide. We will recreate it in class.

Scatterplots [2]



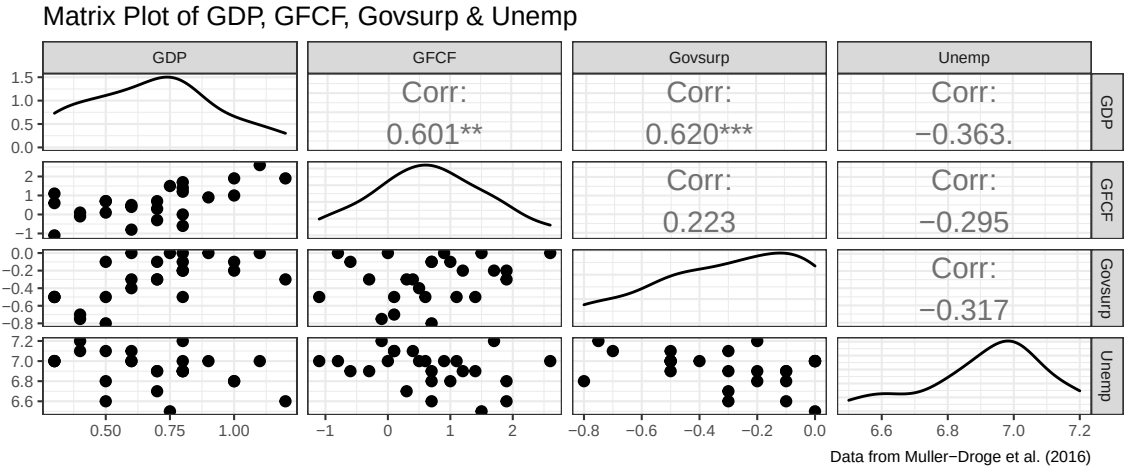
Data from Muller−Droge et al. (2016)

Scatterplot Matrix / Generalized Pairs Plots [1]

- Extending scatterplots for when we have more than two variables.[4]
- Can be easily created in R using the `ggpairs()` from the GGally package.

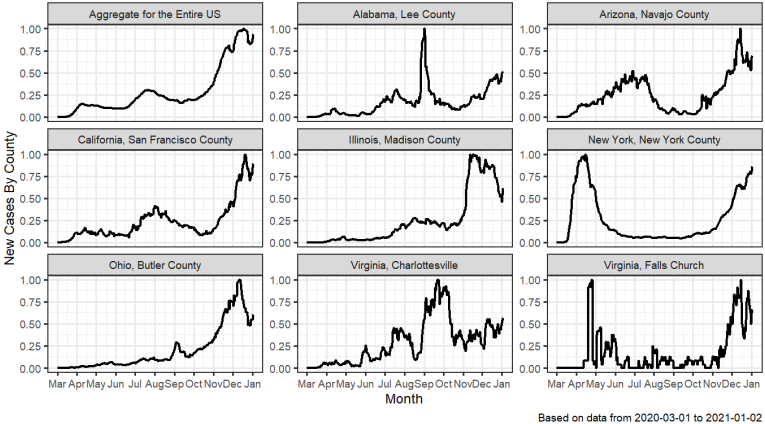
[4] John W Emerson, Walton A Green, Barret Schloerke, Jason Crowley, Dianne Cook, Heike Hofmann, Hadley Wickham. The Generalized Pairs Plot. Journal of Computational and Graphical Statistics, vol. 22, no. 1, pp. 79–91, 2012. Click here to access paper.

Scatterplot Matrix / Generalized Pairs Plots [2]



Data from Muller−Droge et al. (2016)

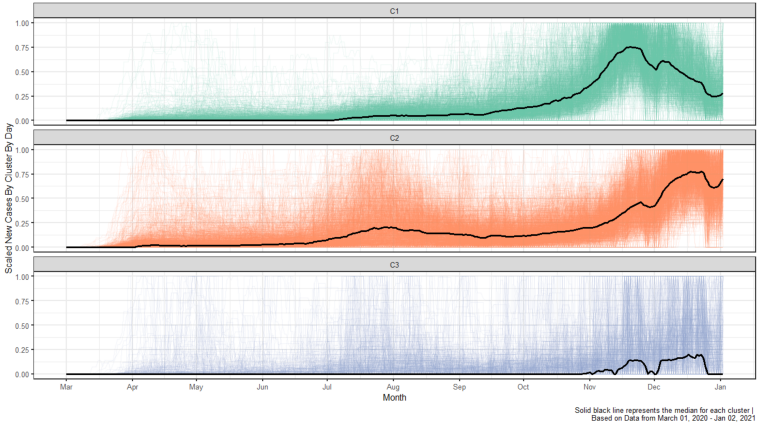
Panel Plots



New COVID-19 Cases in the United States.[5]

[5] Joint Work with Saint Louis University.

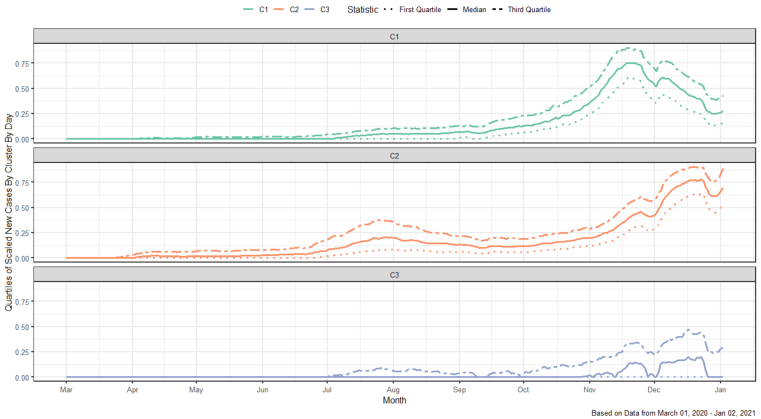
Clustering of COVID-19 New Cases: Spaghetti Plot



Spaghetti Plot of the 3 Major Clusters of COVID-19 Time-Series Profile Shapes.[6]

[6] Joint Work with Saint Louis University.

Clustering of COVID-19 New Cases: Summary Plot



Summary Plot of the 3 Major Clusters of COVID-19 Time-Series Profile Shapes.[7]

[7] Joint Work with Saint Louis University.

Outline

Summary of Main Points

Main Learning Outcomes

- **Explain different goals for visualizing time series data**
- **Identify an appropriate chart for a specific time series data visualization goal**
- **Use software to construct charts of interest**

Things to Do

- **Recommended:** Thoroughly read Chapter 2.1-2.3 of our book.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- **Required:** Complete the graded assignment.

Graded Assignment 03: Evaluating your Understanding

Please go to Canvas (click here) and answer the four questions. **Due February 04, 2021 [11:59 PM, Ohio local time].**

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 03. To reinforce your understanding of the covered material, I also suggest reading up to and including Chapter 2.3 of the book.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment.
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the Remote Desktop)

References

Grinstein, Georges G, and Matthew O Ward. 2001. "Introduction to Data Visualization." In *Information Visualization in Data Mining and Knowledge Discovery*, edited by Usama Fayyad, Georges G Grinstein, and Andreas Wierse, 21–45. San Francisco, CA: Morgan Kaufmann Publishers.

ISA 444: Business Forecasting

03 - Time Series Plots

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Spring 2021