

ISA 401: Business Intelligence & Data Visualization

18: Charts Used for Time-Series Data

Fadel M. Megahed, PhD

Endres Associate Professor
Farmer School of Business
Miami University

 @FadelMegahed

 fmegahed

 fmegahed@miamioh.edu

 Automated Scheduler for Office Hours

Spring 2024

08 : 00

Non-graded: COVID19 Data Viz

Over the next 8 minutes, use Tableau to answer the following questions based on the [`covid_cases_county.csv`](#) (which can be downloaded from Canvas):

- **What are the total number of cases per county?** Use a suitable map to answer this question.
- **What are the total number of deaths per state?** You can show that using either a table or a map. In case of a map, please show the numbers for each state on the map as well.

Learning Objectives for Today's Class

- Understand main goals behind visualizing time-series data
- Explain the different types of charts for univariate and multivariate time-series

Types of Data Over Time

Cross Sectional Data

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 1: H1B 2020-2022 Data for Senior Data Scientists at Netflix

	START DATE	JOB TITLE	BASE SALARY	LOCATION
1	2021-08-11	SENIOR DATA SCIENTIST	118,955	LOS GATOS, CA
2	2021-06-14	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
3	2021-09-09	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
4	2021-06-14	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
5	2022-01-03	SENIOR DATA SCIENTIST	143,562	LOS GATOS, CA

Showing 1 to 5 of 23 entries

Previous

1

2

3

4

5

Next

Cross Sectional Data

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 2: NBA 2023-2024 Leaders - Top Players in PTS/Game

	Player	Pos	Age	Tm	G	FG			
1	Joel Embiid	C	29	PHI	35	11.7	0.531	0.558	35
2	Luka Dončić	PG	24	DAL	66	11.5	0.489	0.573	33.9
3	Giannis Antetokounmpo	PF	29	MIL	70	11.6	0.616	0.629	30.8
4	Shai Gilgeous-Alexander	PG	25	OKC	71	10.7	0.539	0.571	30.3
5	Jalen Brunson	PG	27	NYK	70	10	0.475	0.54	27.8

Showing 1 to 5 of 719 entries

Source: Data scraped from https://www.basketball-reference.com/leagues/NBA_2024_per_game.html (Basketball Reference) on April 04, 2024 using the `rvest` package. The printing was limited to the selected variables.

Previous

1

2

3

4

5

...

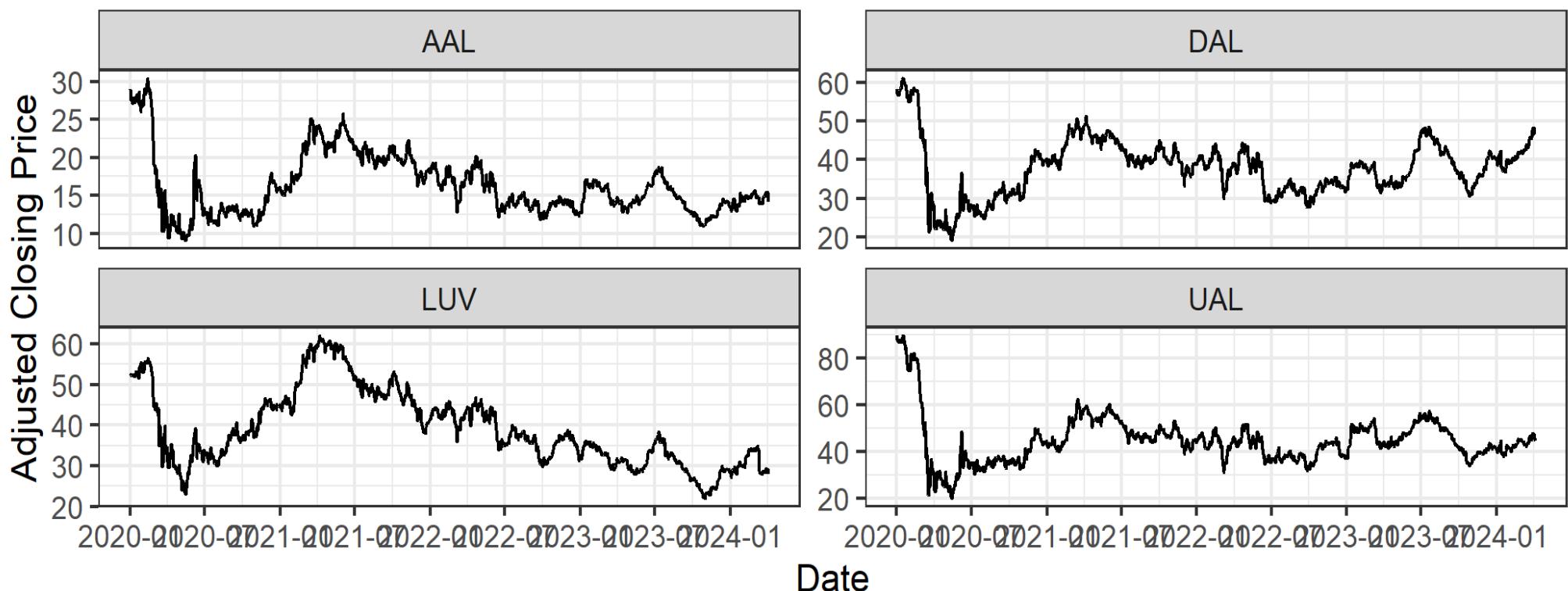
144

Next

Time Series Data

Time Series Data: Comparable measurements recorded on a single (or a few) variables over time (usually a long period of time).

Example 2: Stock prices of U.S. Airlines



Panel Data

Panel Data: Cross sectional measurements (usually many variables) repeated over time (usually over a few time periods).

Example: World Bank's Data

	iso3c◆	date◆	NY.GDP.MKTP.KD.ZG◆	SH.DYN.NMRT◆	SH.HIV.INCD.ZS◆	SH.MED.BED◆
1	CHN	2021	8.4	3.2		
2	CHN	2022	3	3.2		
3	CHN	2023	3	3.2		
4	EGY	2021	3.3	10		
5	EGY	2022	6.6	10		

Showing 1 to 5 of 9 entries

Previous

1

2

Next

Source: Data queried from the [World Bank Data](#) using the `wbstats`  in R. The printed results show a snapshot of 7 variables (out of a much larger panel dataset). You can think of panel data as a cross-sectional dataset with a longitudinal/time component.

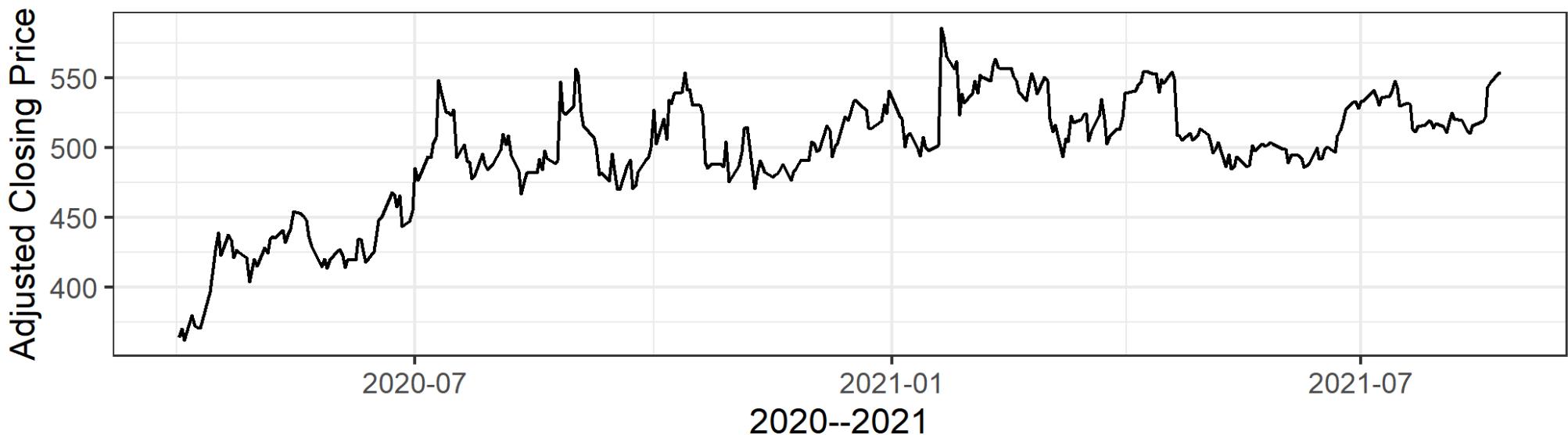
Components of a Time Series

Trend

A **trend** is an increasing or decreasing pattern over time.

Increasing Trend

The meteoric rise of \$NFLX from 2020-04-01 to 2021-08-25

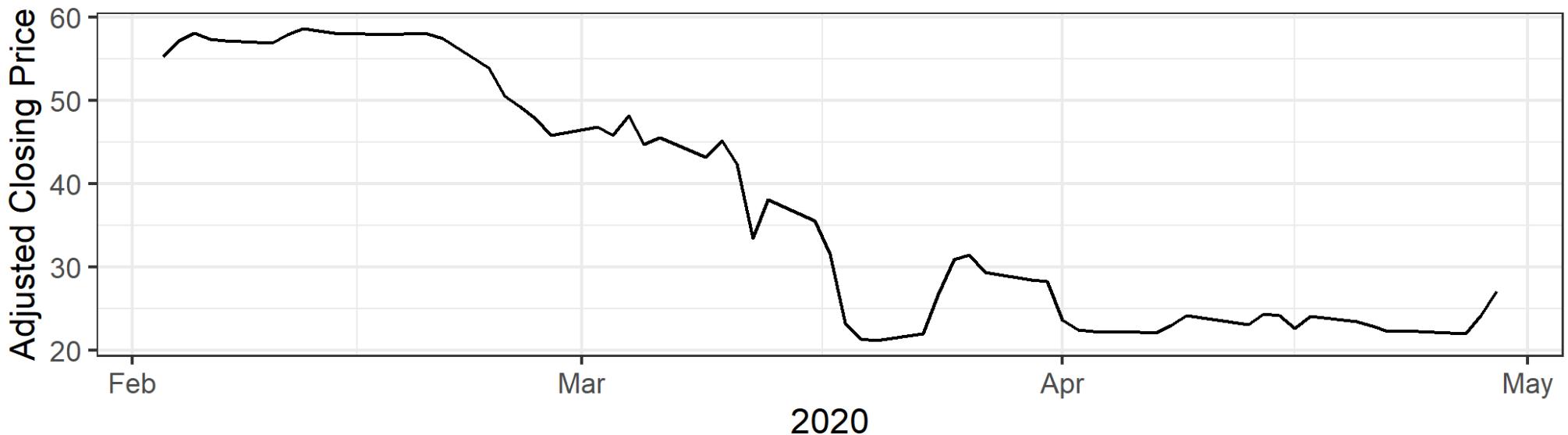


Trend

A **trend** is an increasing or decreasing pattern over time.

Decreasing Trend

The decline in \$DAL from 2020-02-03 to 2020-04-30

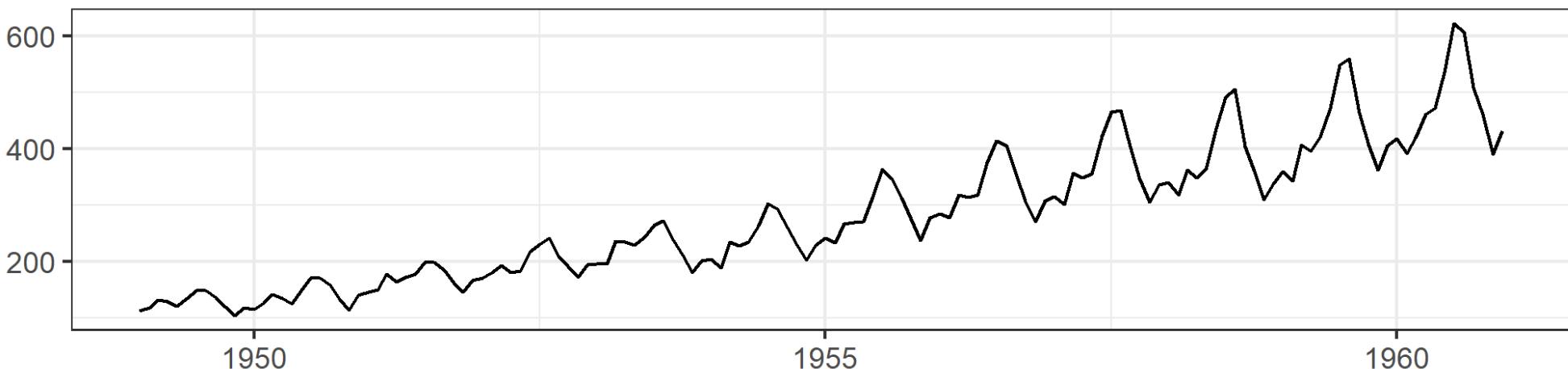


Seasonality

Seasonality refers to the property of a time series that displays REGULAR patterns that repeat at a constant frequency (m).

Seasonality with a Multiplicative Trend

Non-linear trend & seasonal component grows over time

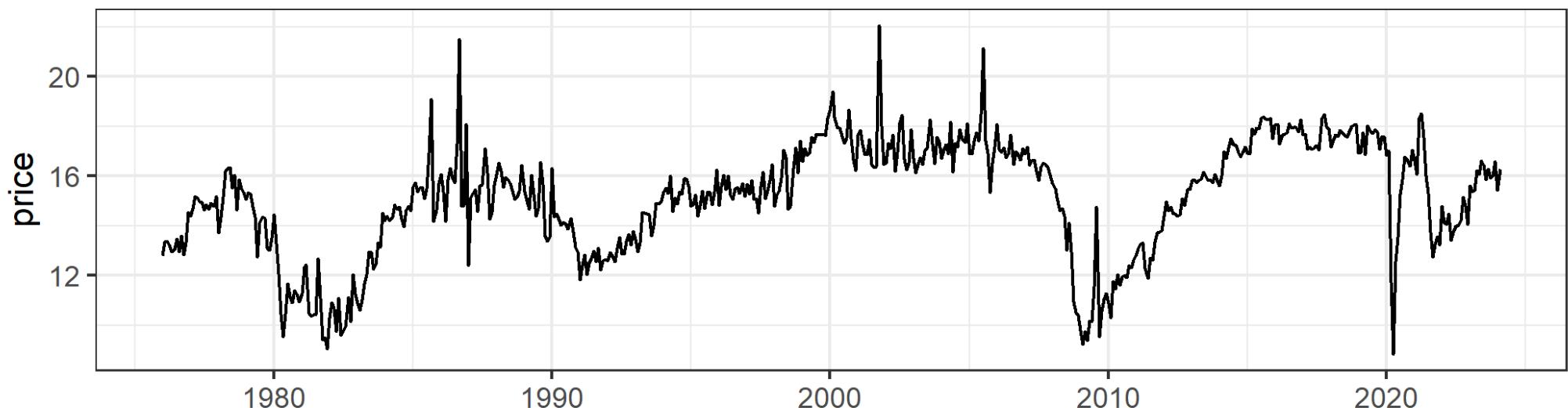


AirPassengers R Dataset -- Source: Box, G. E. P., Jenkins, G. M. and
Reinsel, G. C. (1976) Time Series Analysis, Forecasting and Control.

Cycle

Cyclical fluctuations are somewhat irregular (unknown duration).

The cyclical nature of auto sales



U.S. Bureau of Economic Analysis, Total Vehicle Sales [TOTALSA], retrieved from FRED, Federal Reserve Bank of St. Louis
<https://fred.stlouisfed.org/series/TOTALSA>, on April 04, 2024

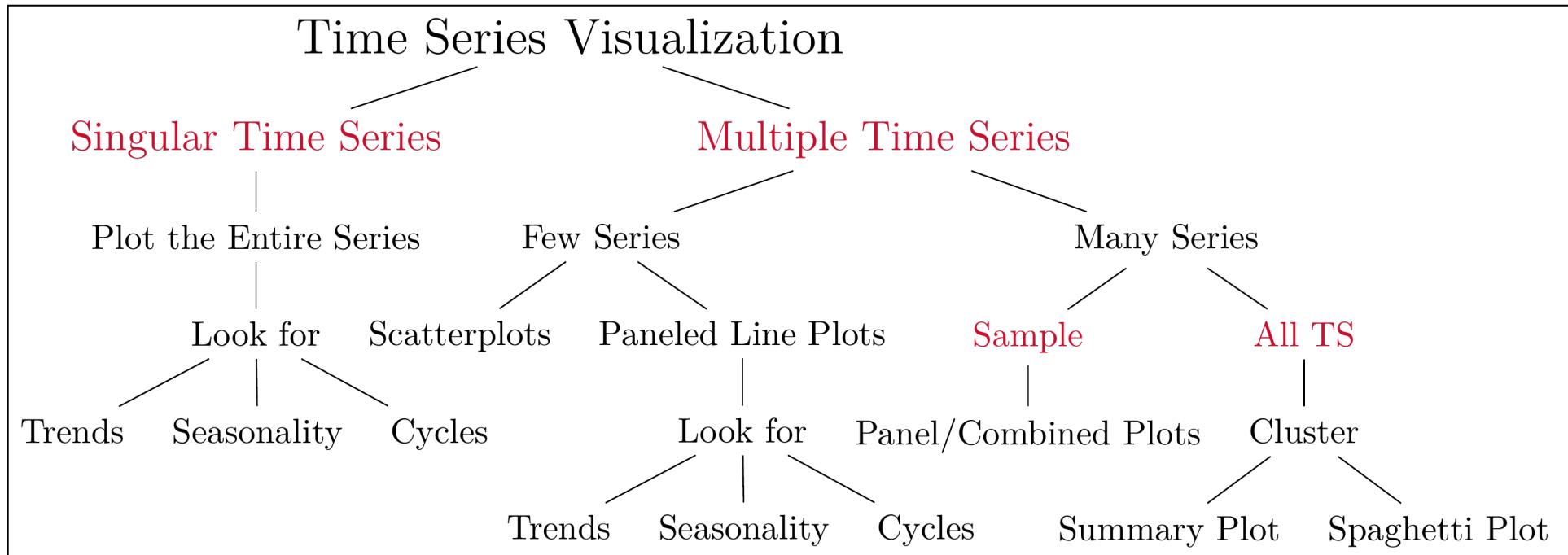
The Goals Behind Visualizing (Time Series) Data

Exploratory vs Confirmatory Viz 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 Viz

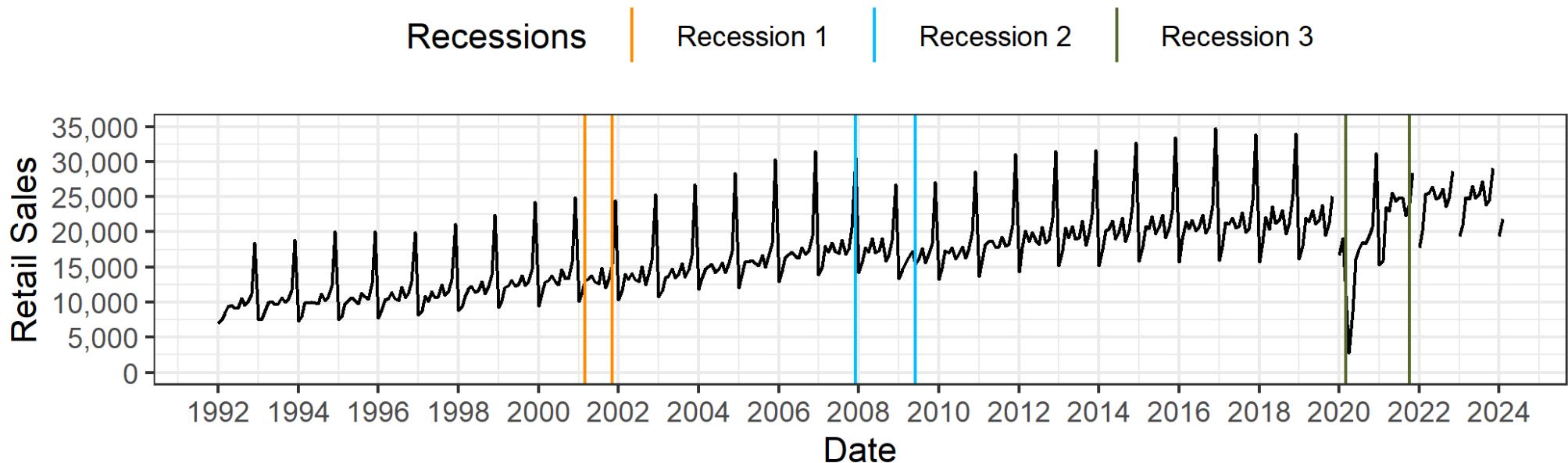


A Potential Framework for Time Series Visualization

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.

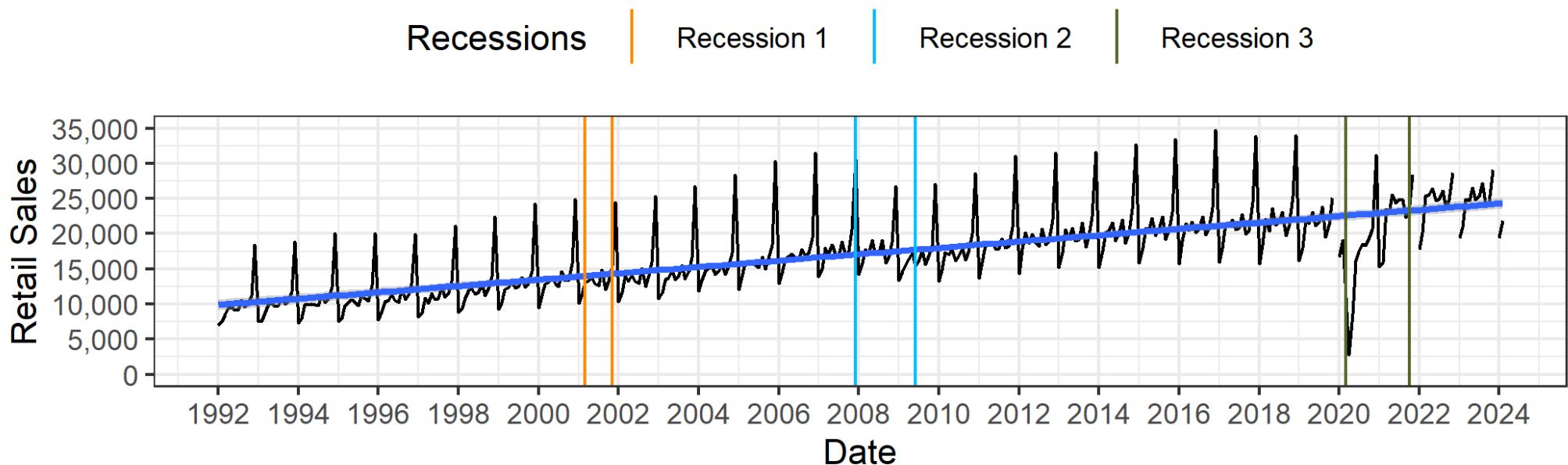
Singular TS

Monthly Retail Sales (RSCCASN) in the U.S.

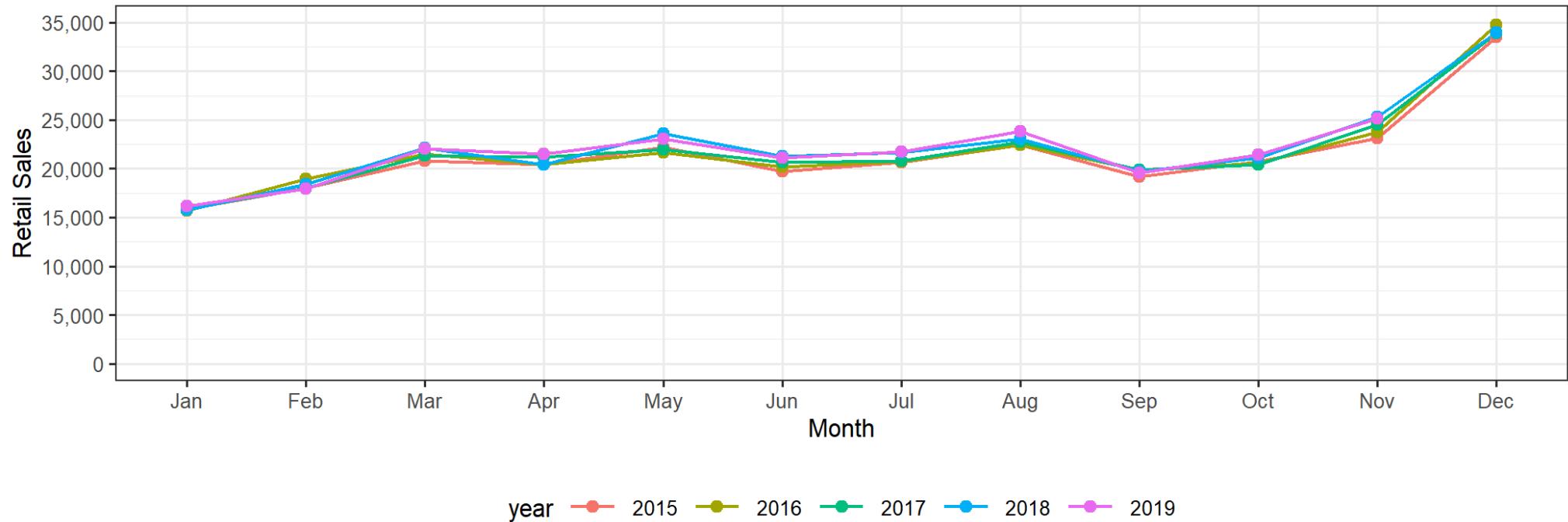


Looking for Trends

Monthly Retail Sales (RSCCASN) in the U.S.

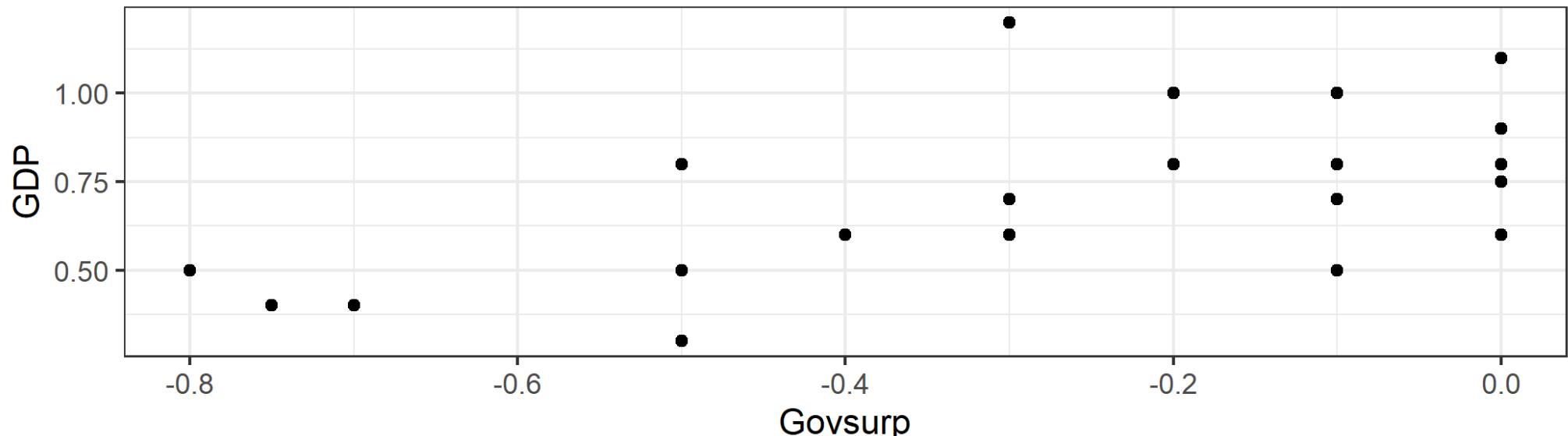


Looking for Seasonality



Multiple TS: Scatterplots

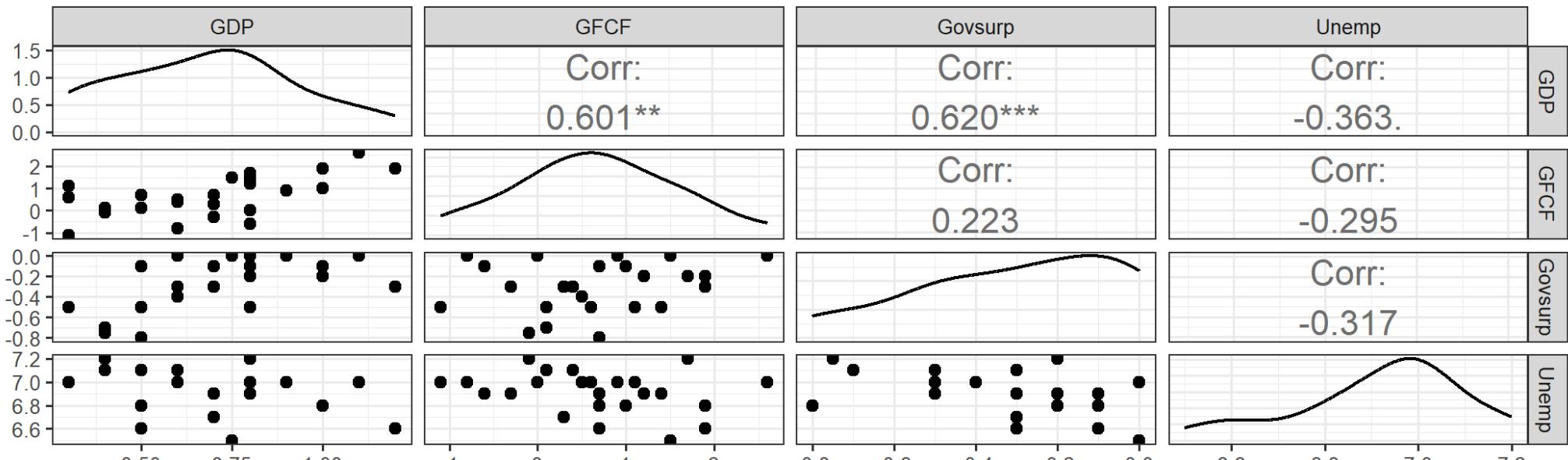
Scatterplot of GDP vs. Government Spending



Data from Muller-Droge et al. (2016)

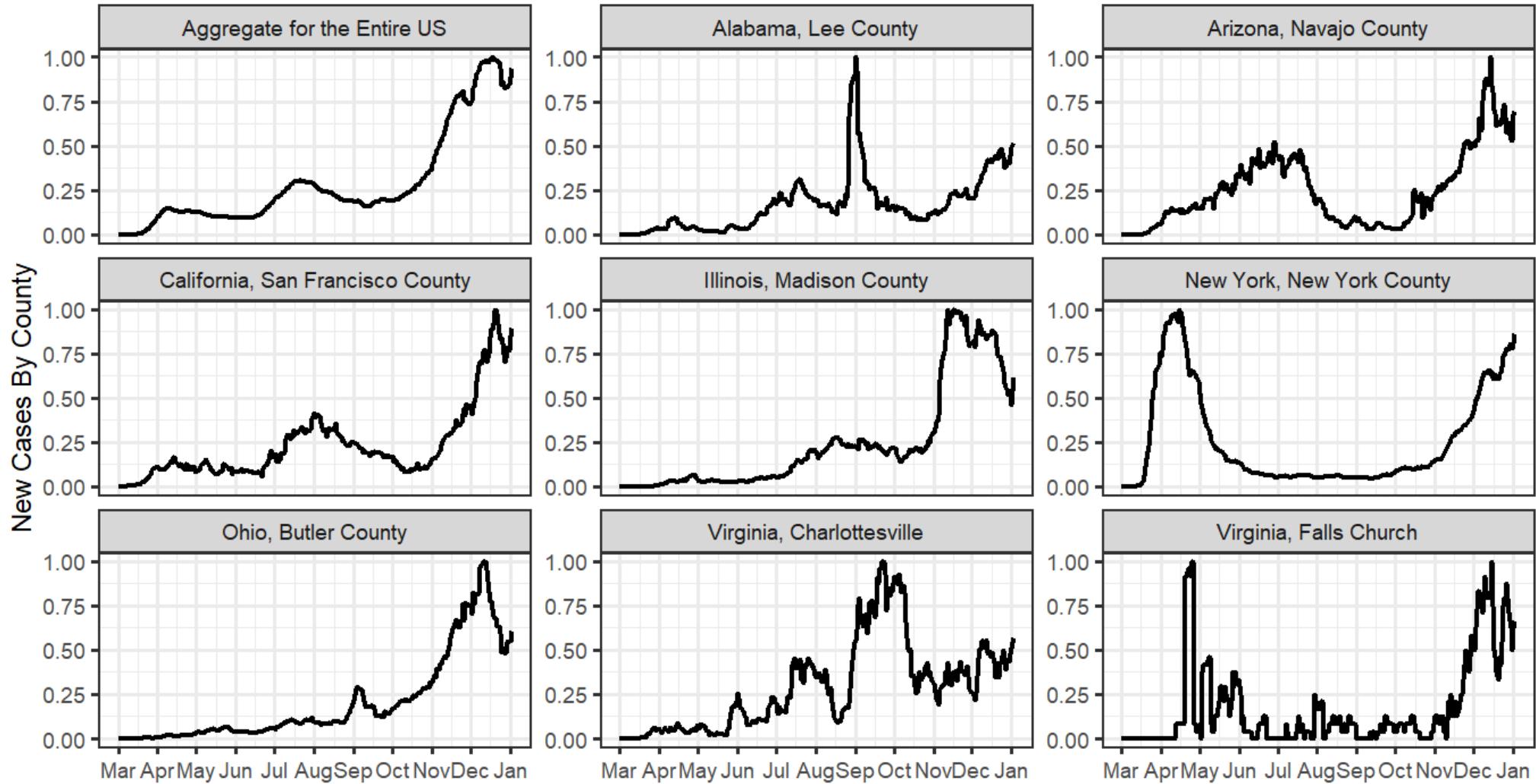
Multiple TS: Scatterplot Matrix

Matrix Plot of GDP, GFCF, Govsurp & Unemp

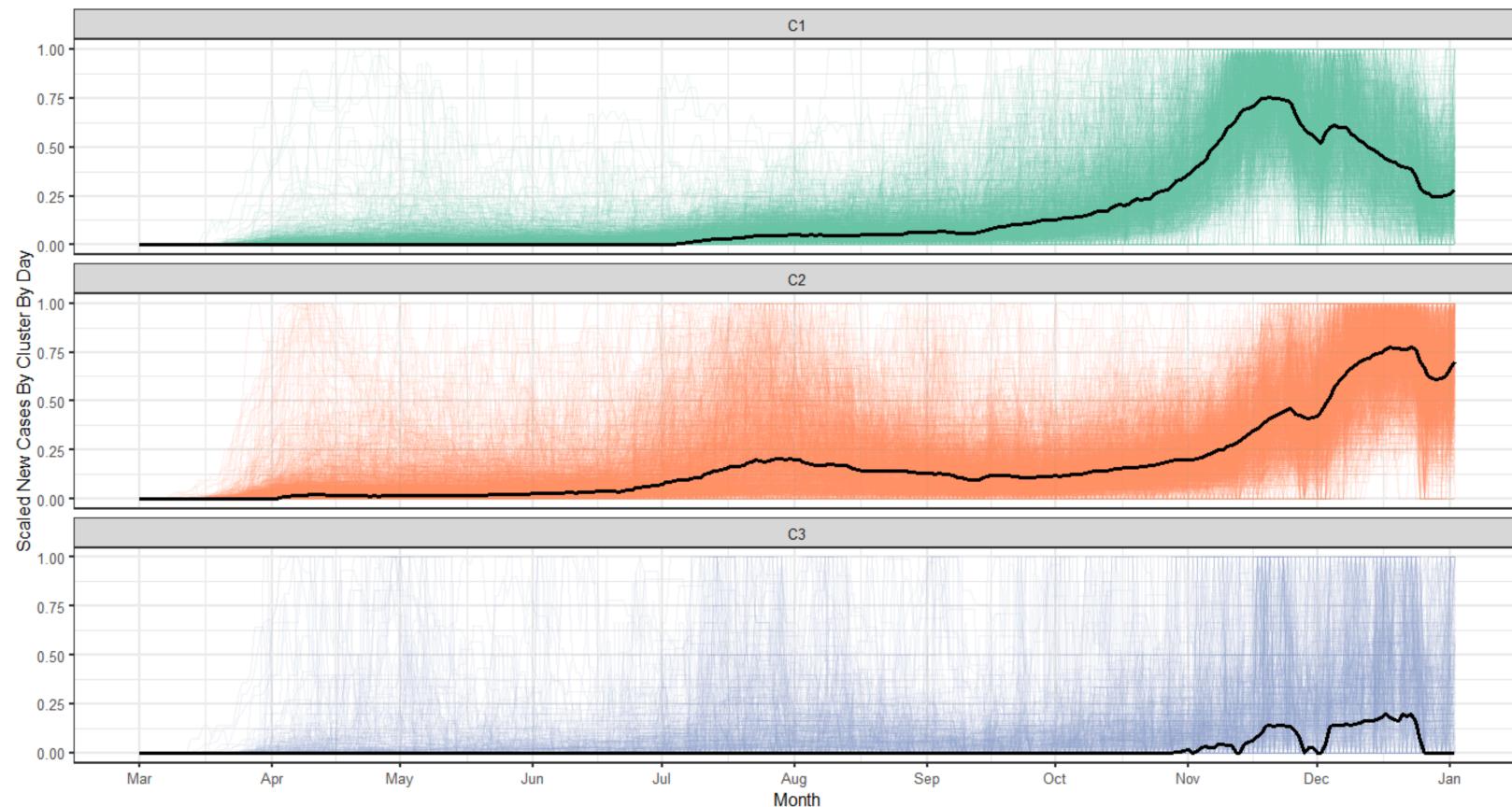


Data from Muller-Droge et al. (2016)

Multiple TS: Panel Plots

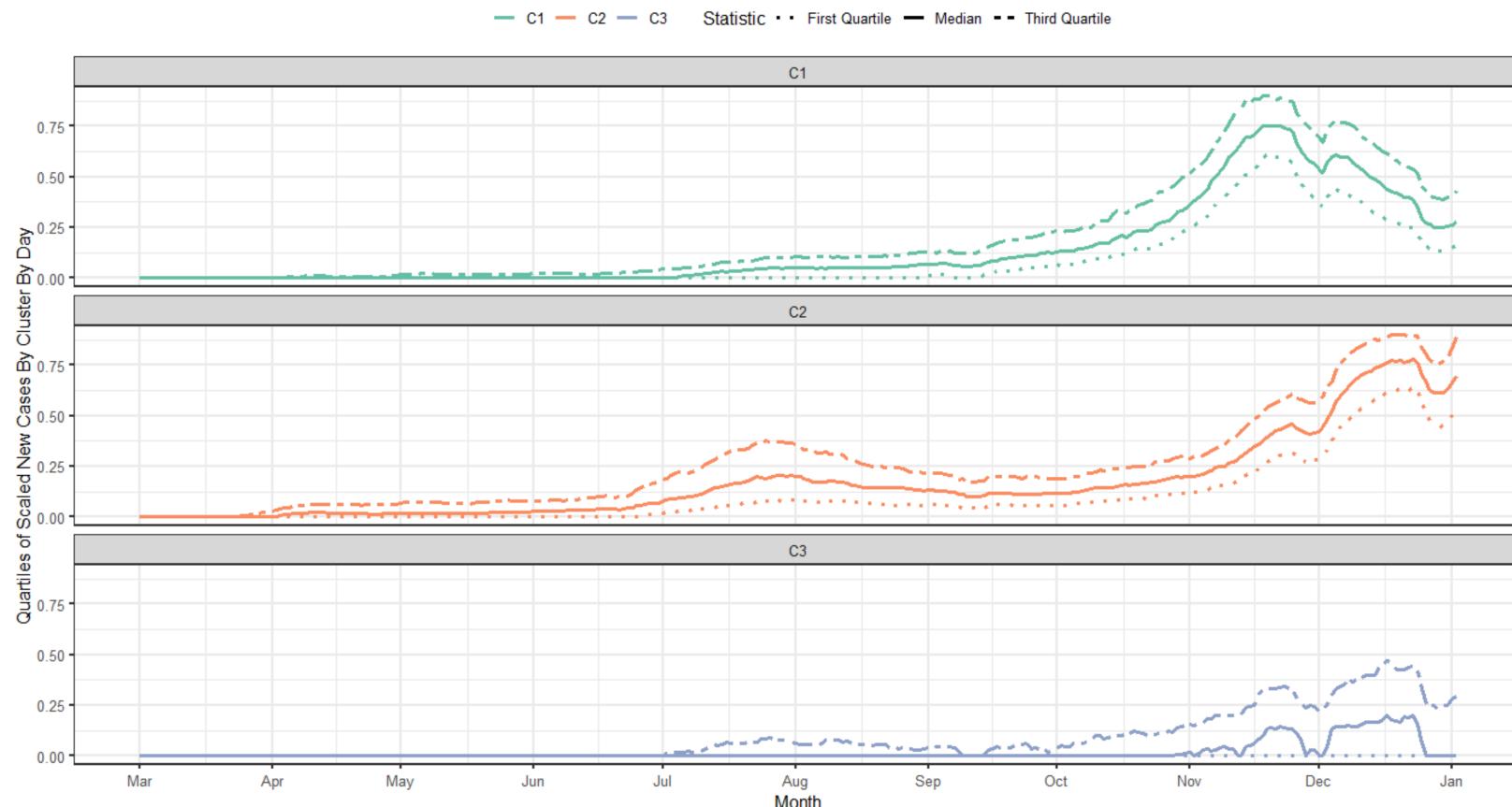


Spaghetti Plots (Often w/ Clustering)



Solid black line represents the median for each cluster.
Based on Data from March 01, 2020 - Jan 02, 2021

Summary Plots (Often w/ Clustering)



Recap

Summary of Main Points

- Understand main goals behind visualizing time-series data
- Explain the different types of charts for univariate and multivariate time-series