ISA 401/501: Business Intelligence & Data Visualization

19: Charts Used Time-Series, Spatial and Spatiotemporal Data

Fadel M. Megahed, PhD

Enders Associate Professor
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu
Automated Scheduler for Office Hours

Fall 2022

Recap of Assignment 14

Let us discuss why the majority of submitted solutions to Assignment 14 are incorrect. Here is an example of incorrect_submission.

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
- Explain the different types of spatial plots
- Select suitable spatial graphs for different scenarios
- Understand how spatiotemporal plots can help in storytelling (what makes BI special)

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	2021-10-18	SENIOR DATA SCIENTIST	143,562	LOS GATOS, CA
5	2022-07-20	SENIOR DATA SCIENTIST	143,562	LOS GATOS, CA

Showing 1 to 5 of 18 entries Previous 1 2 3 4 Next

Data scraped from https://h1bdata.info/index.php?em=Netflix+Inc&job=Senior+Data+Scientist&city=LOS+GATOS&year=All+Years on October 25, 2022 using the rvest 🤗. The printing was limited to those individuals who started on/after January 01, 2020, with the filters specified in the URL.

Cross Sectional Data

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

Example 2: NBA 2022-2023 Leaders - Top Players in PTS/Game

	Player	Pos	Age	Tm	G	FG	FG%	eFG%	PTS
1	Ja Morant	PG	23	MEM	4	11.5	0.548	0.619	35.3
2	Luka Dončić	PG	23	DAL	2	11	0.478	0.543	33.5
3	Stephen Curry	PG	34	GSW	3	10.3	0.47	0.591	33.3
4	Damian Lillard	PG	32	POR	4	10.5	0.5	0.595	33.3
5	Donovan Mitchell	SG	26	CLE	3	11.3	0.493	0.572	33.3

Showing 1 to 5 of 404 entries

Source: Data scraped from https://www.basketball-reference.com/leagues/NBA_2023_per_game.html{Basketball-Reference on October 25, 2022 using the rvest 🤗. The printing was limited to the selected variables.

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.BE
1	CHN	2019	6	3.7		
2	CHN	2020	2.2	3.5		
3	CHN	2021	8.1	3.5		
4	EGY	2019	5.6	10.7		
5	EGY	2020	3.6	10.3		

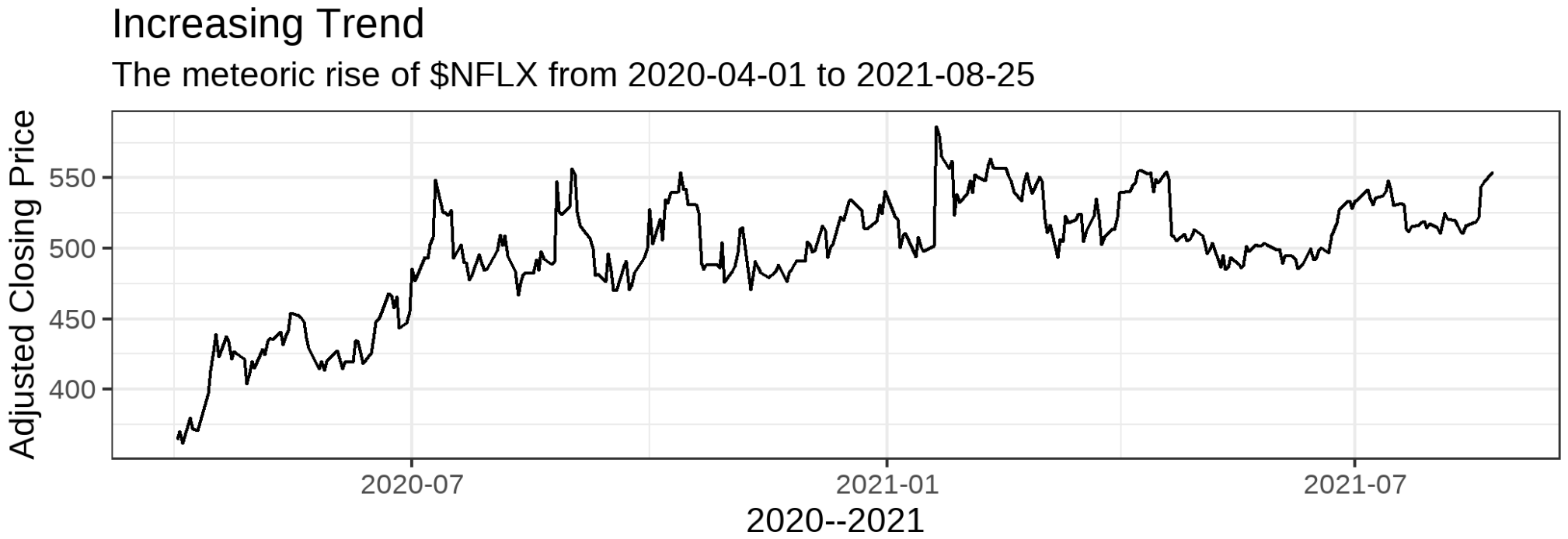
Showing 1 to 5 of 9 entries Previous 1 2 Next

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

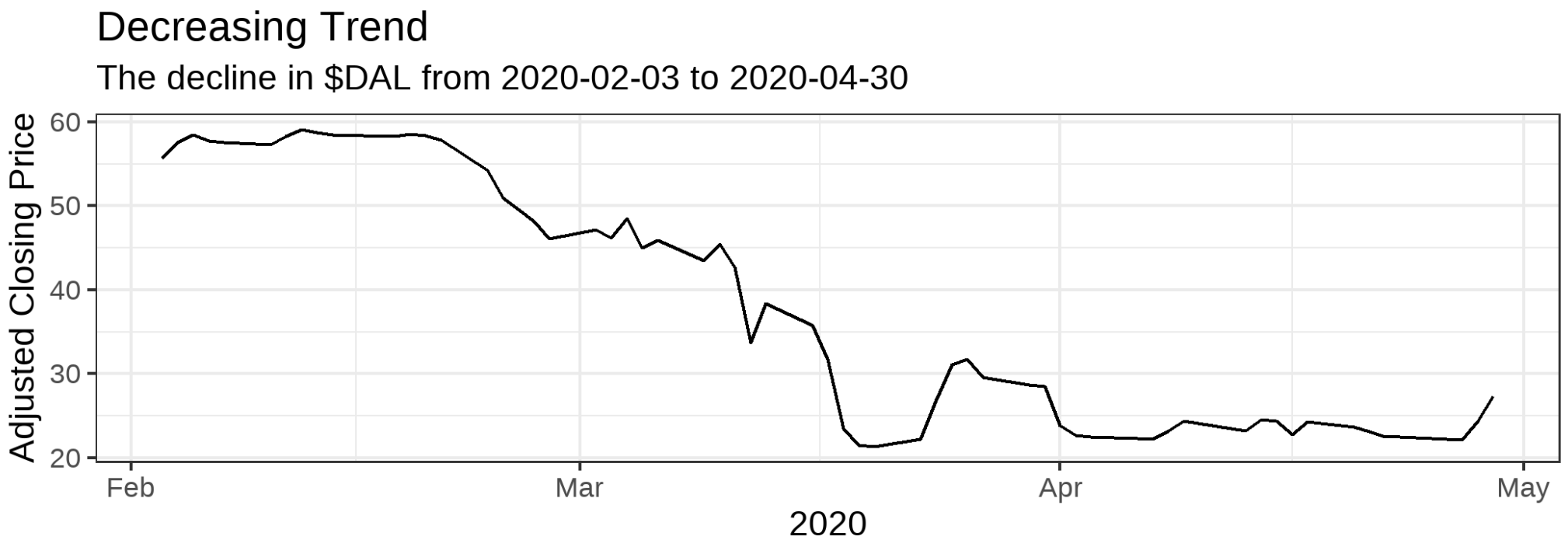
Trend

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



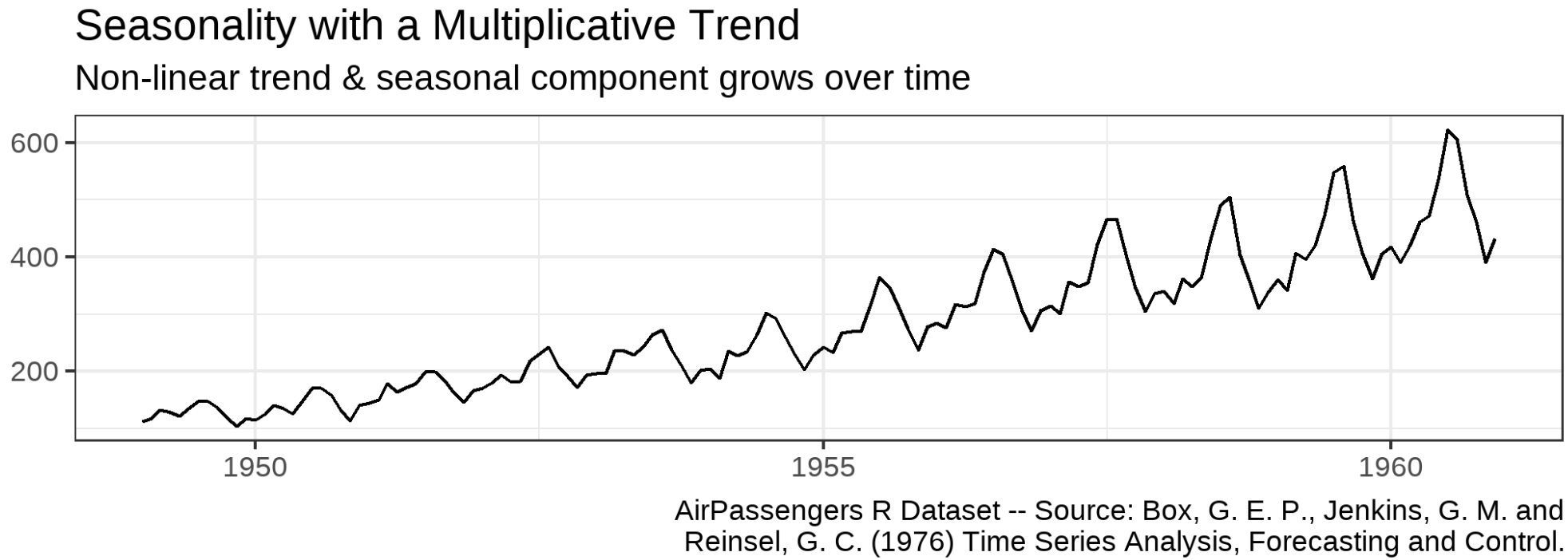
Trend

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



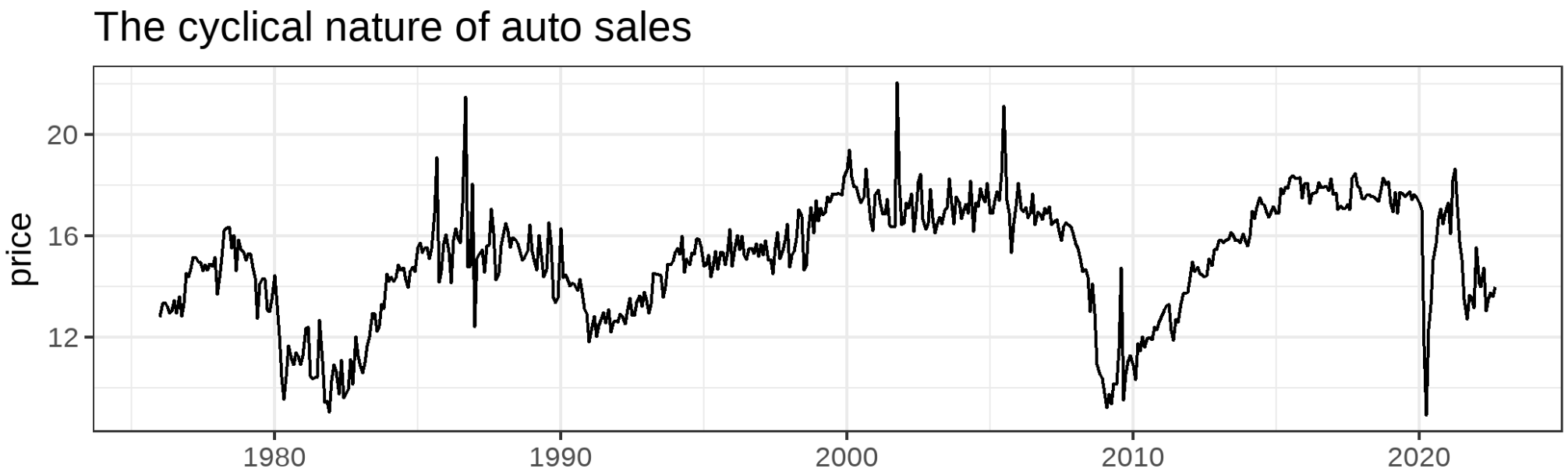
Seasonality

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



Cycle

Cyclical fluctuations are somewhat irregular (unknown duration).



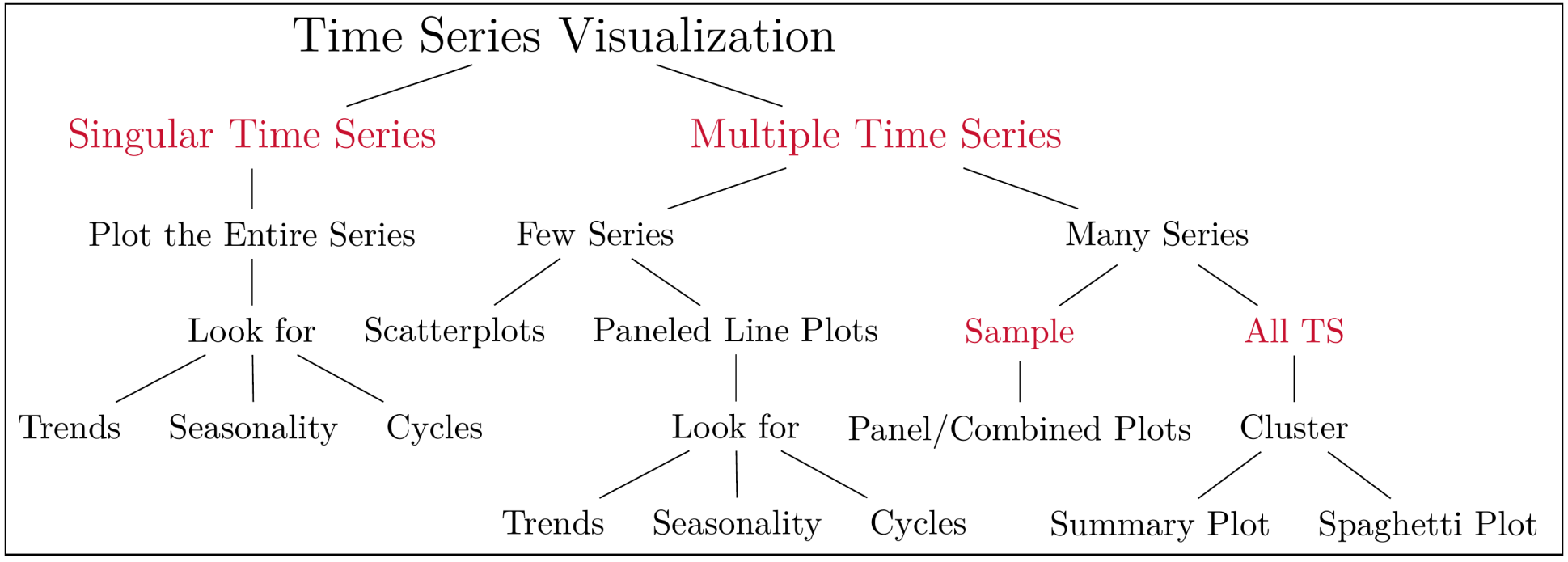
U.S. Bureau of Economic Analysis, Total Vehicle Sales [TOTALSA], retrieved from FRED, Federal Reserve Bank of St. Louis https://fred.stlouisfed.org/series/TOTALSA, on October 25, 2022

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)

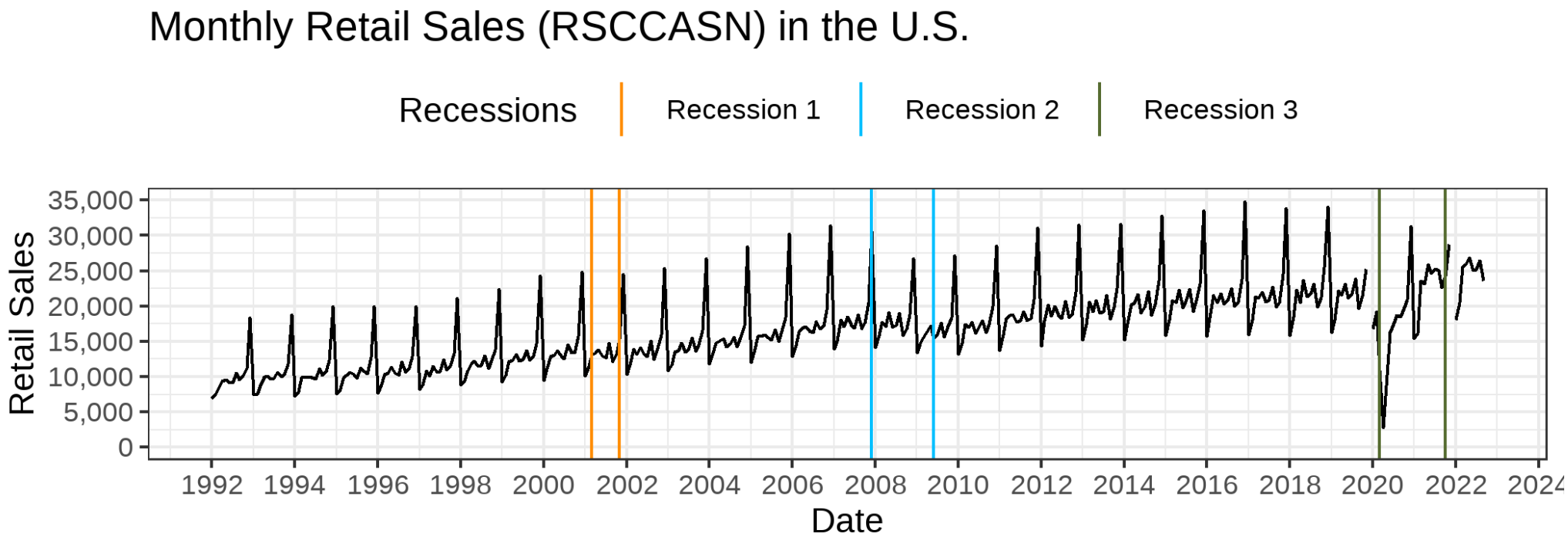
A Structured Approach for Time Series Viz



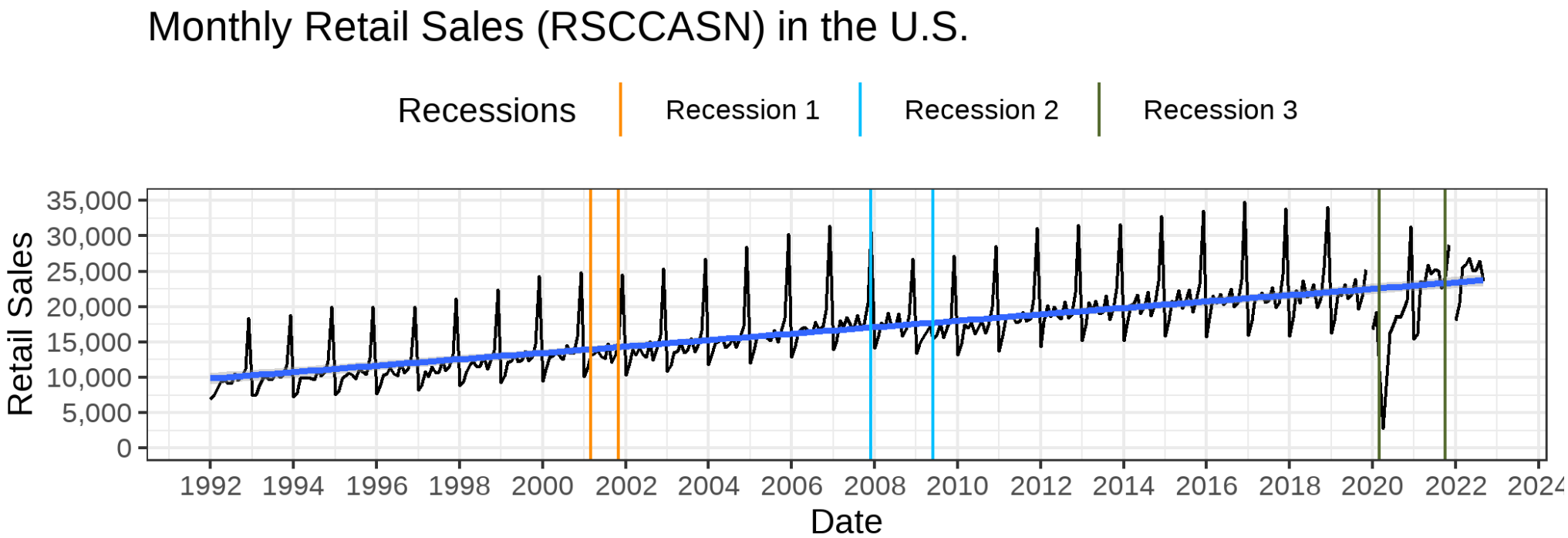
A Potential Framework for Time Series Visualization

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.

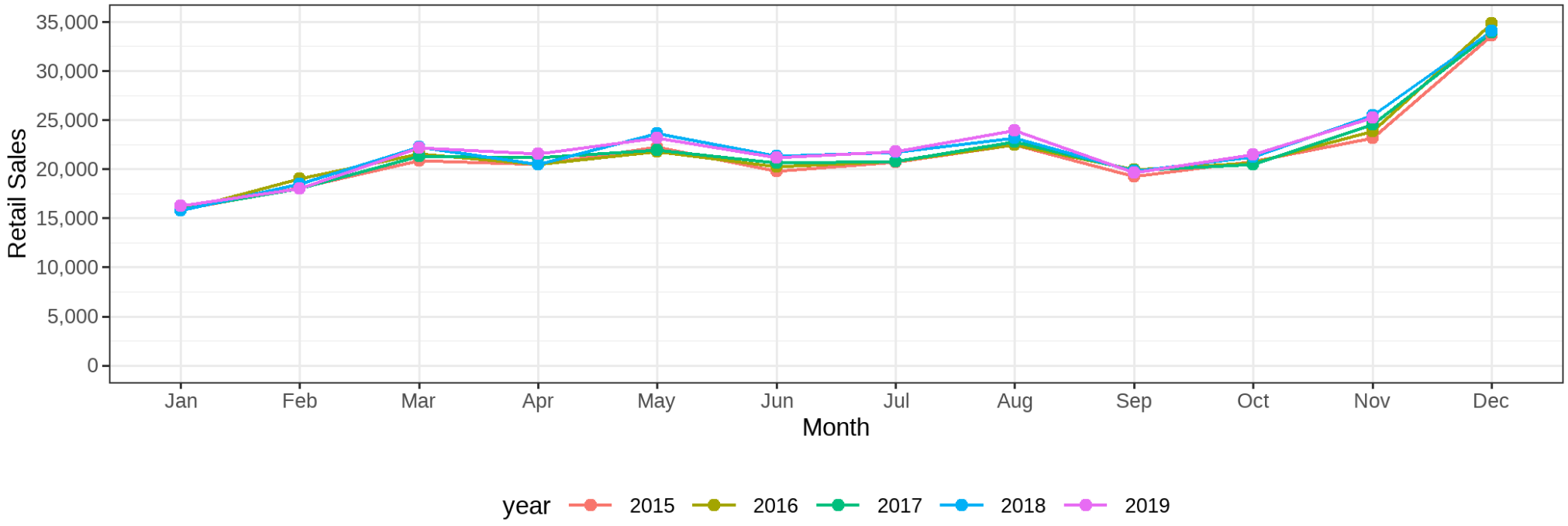
Singular TS


Monthly Retail Sales (RSCCASN) in the U.S.
Recessions
Recession 1
Recession 2
Recession 3
Retail Sales
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
2024
Date

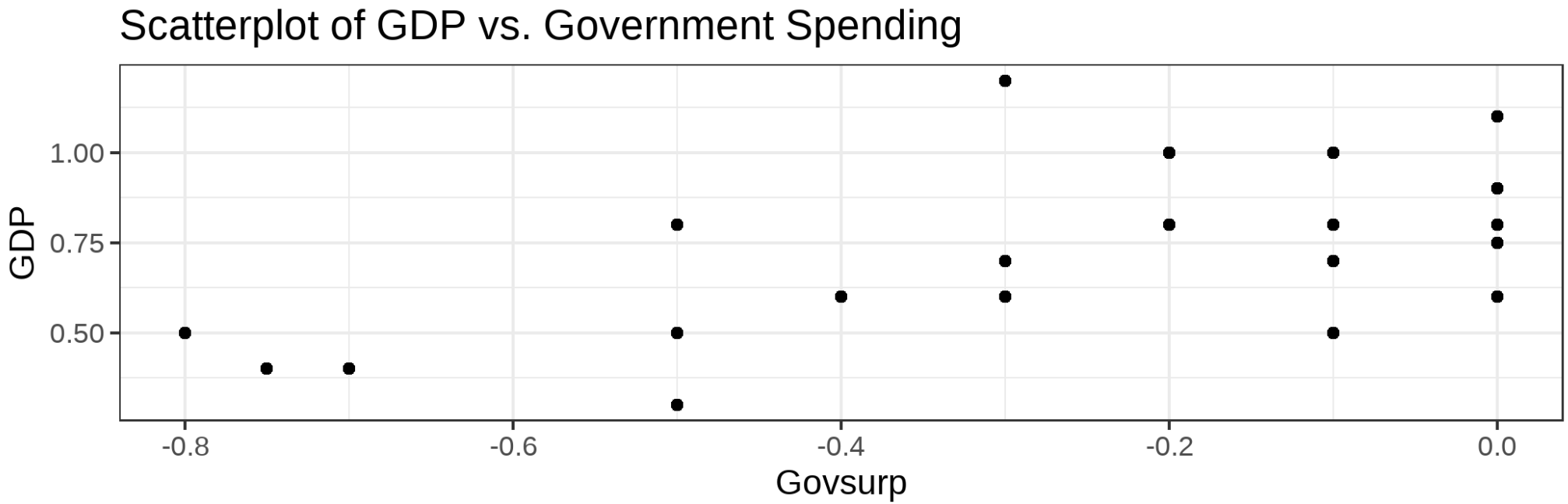
Looking for Trends



Looking for Seasonality

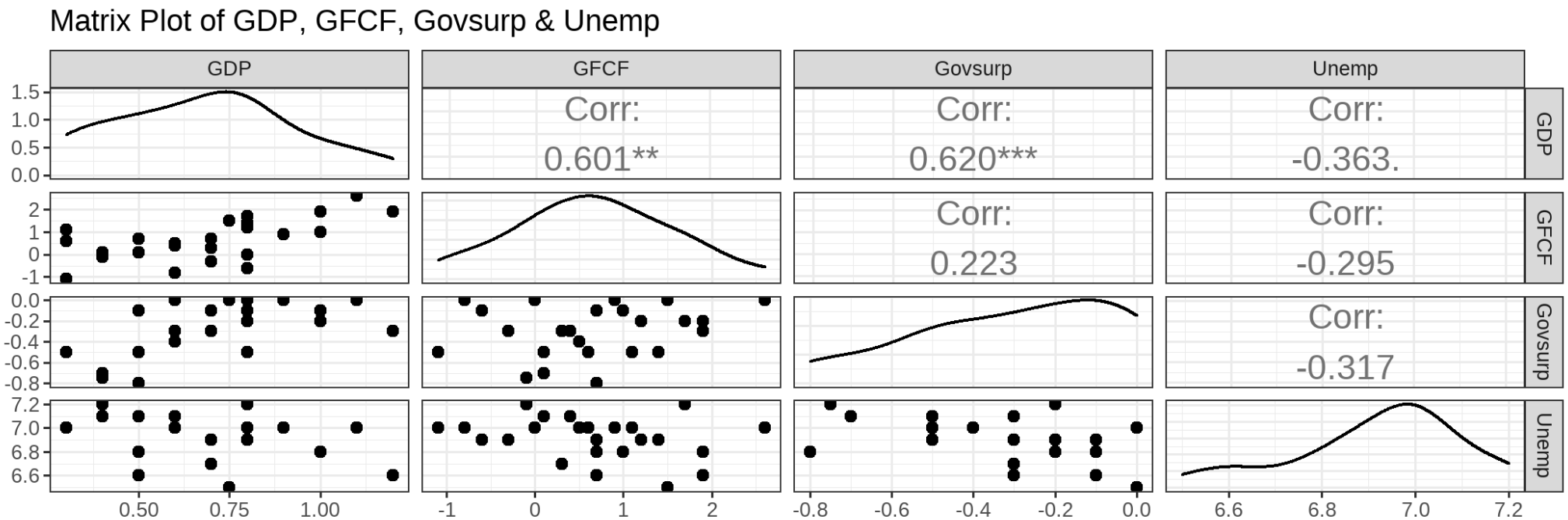


Multiple TS: Scatterplots



Data from Muller-Droge et al. (2016)

Multiple TS: Scatterplot Matrix



Multiple TS: Panel Plots

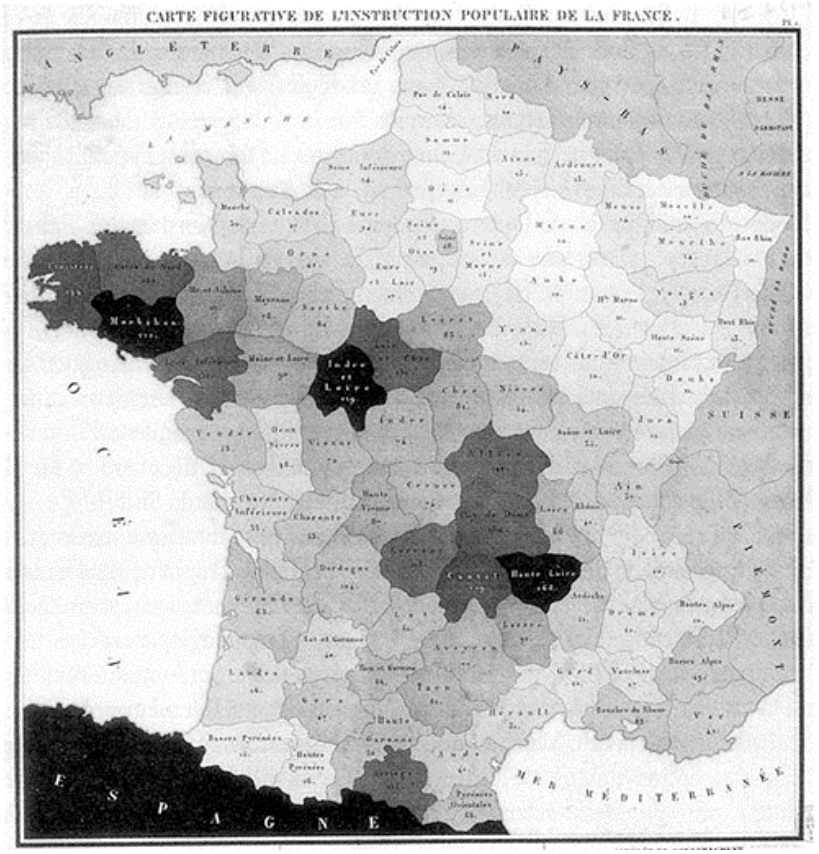
Spaghetti Plots (Often w/ Clustering)

Summary Plots (Often w/ Clustering)

Spatial Plots

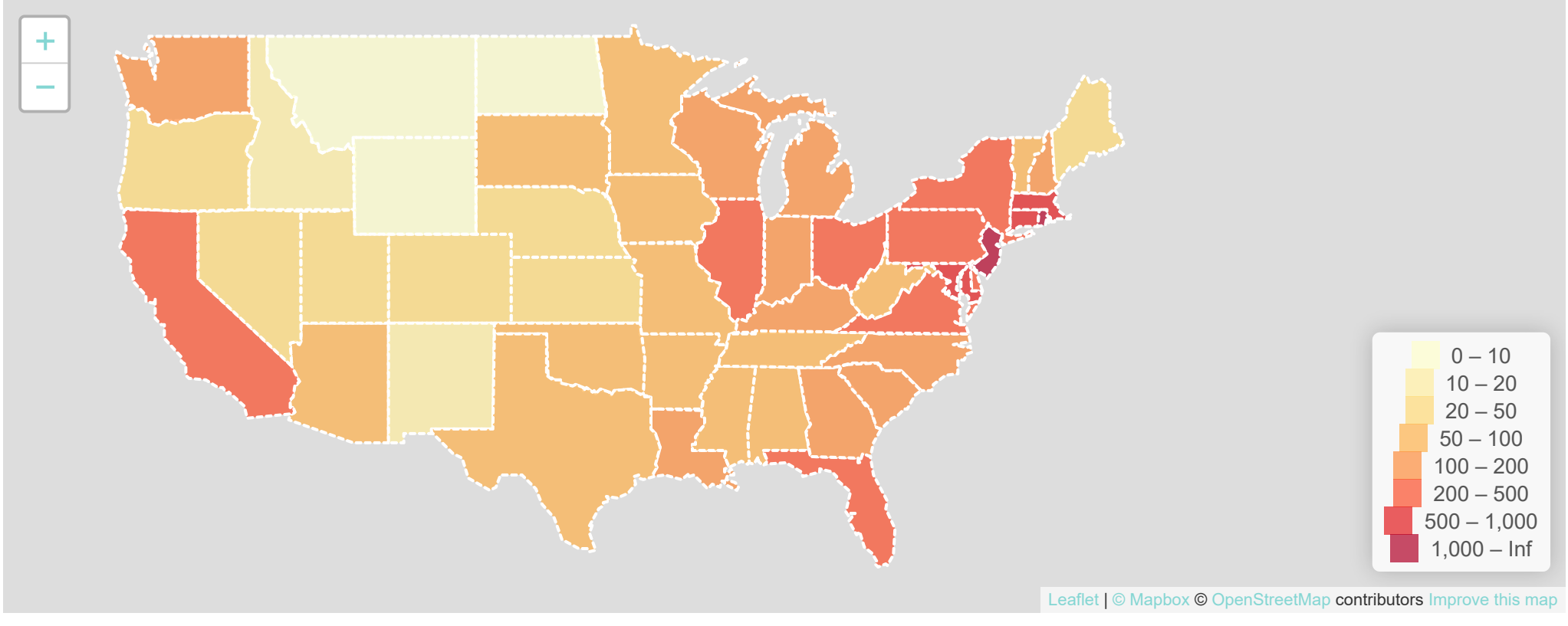
Choropleth Maps

Maps where areas are **shaded, colored, or patterned** relative to a data attribute value.



Choropleth Maps

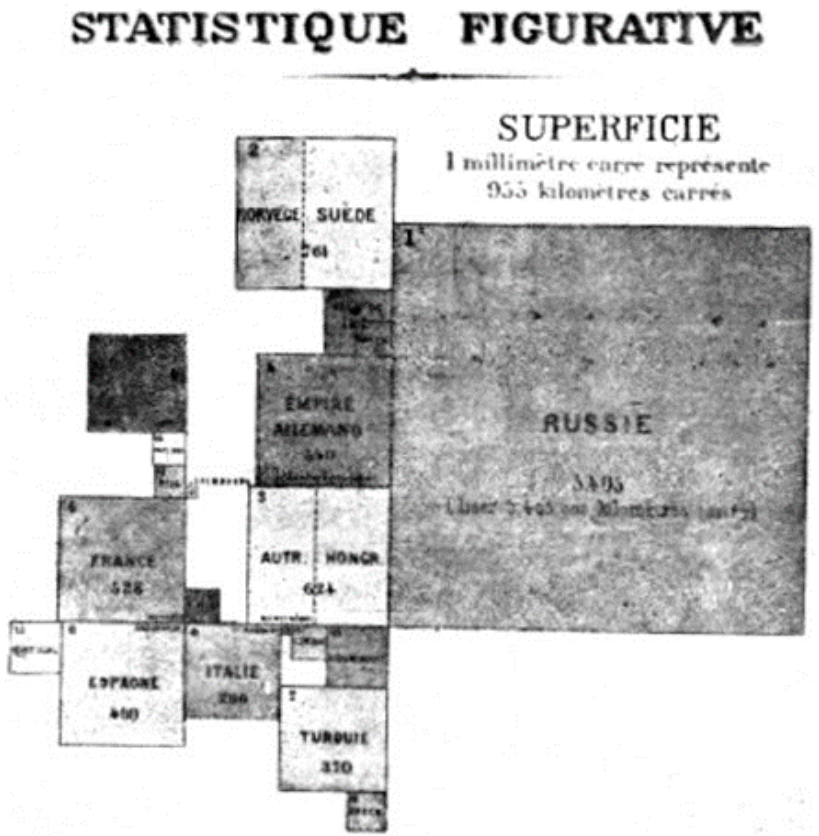
Maps where areas are **shaded, colored, or patterned** relative to a data attribute value.



Population Density in U.S.

Cartograms

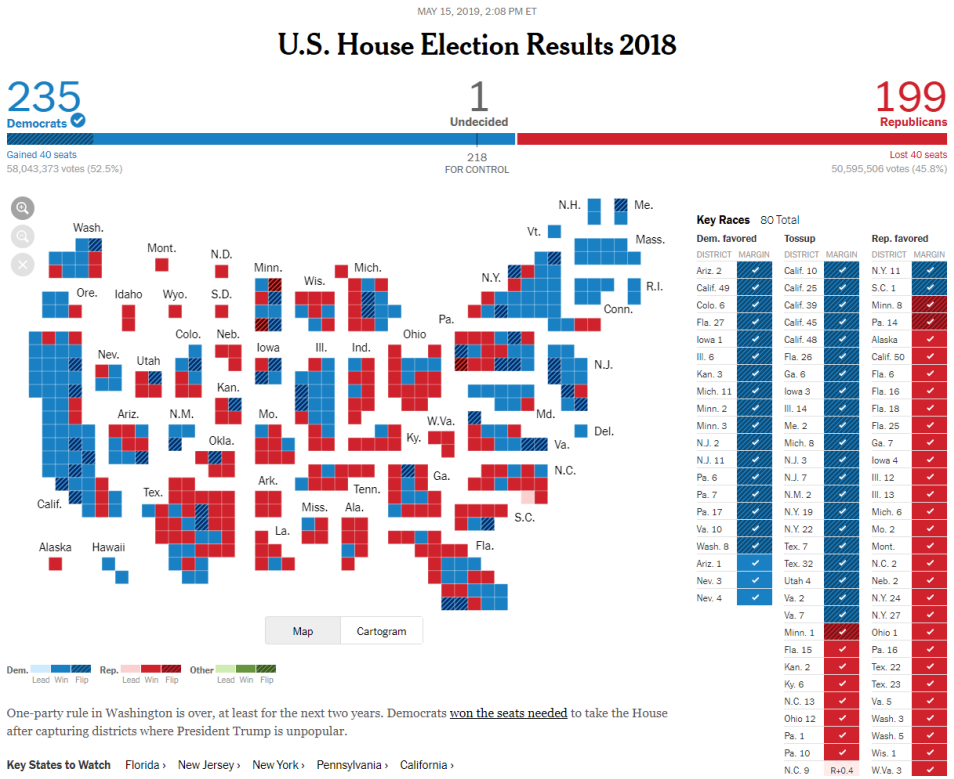
A cartogram is a map in which **areas are scaled and distorted relative to a data attribute value**



The First Cartogram – Emile Levasseur, 1868

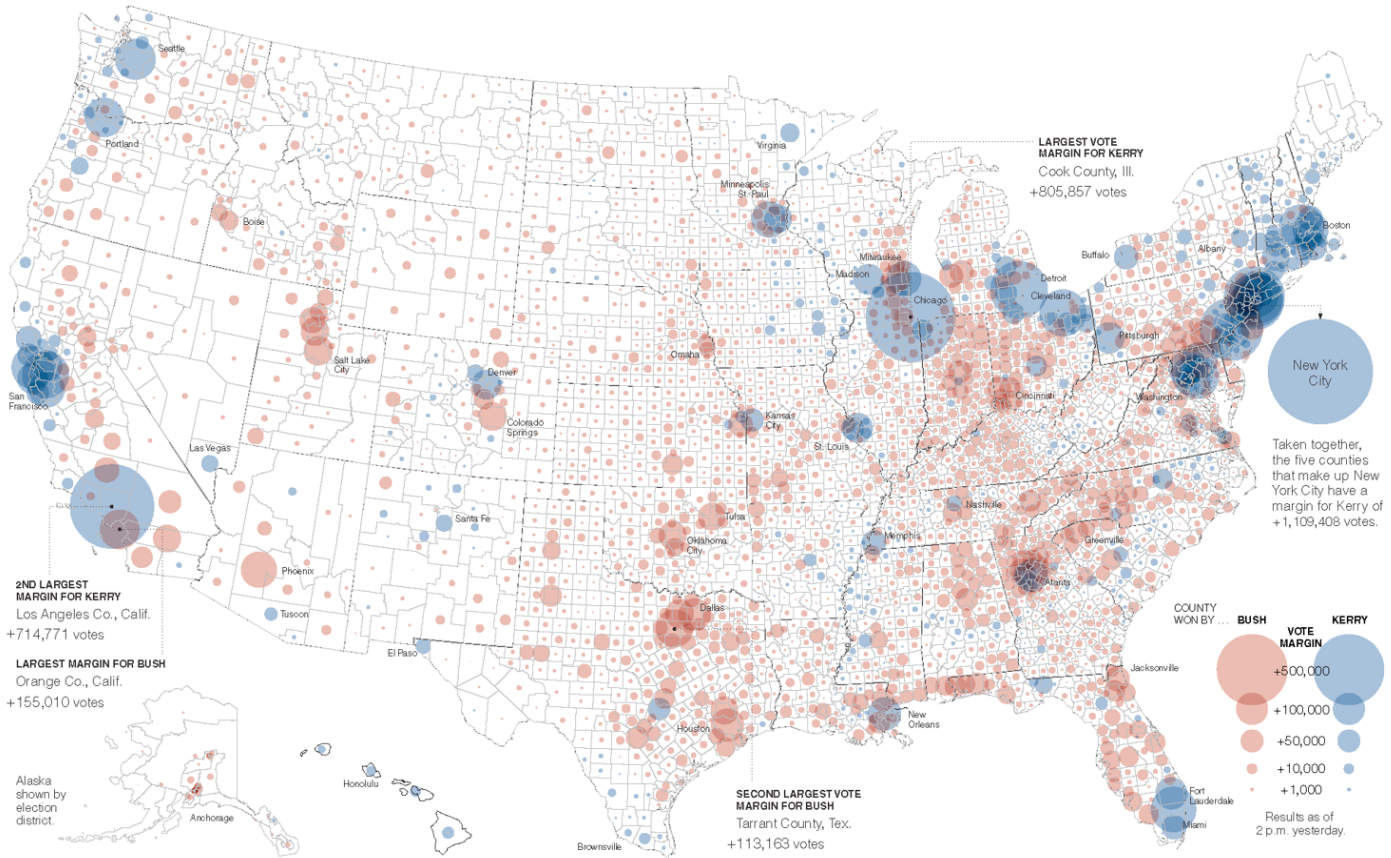
Cartograms

A cartogram is a map in which **areas are scaled and distorted relative to a data attribute value**



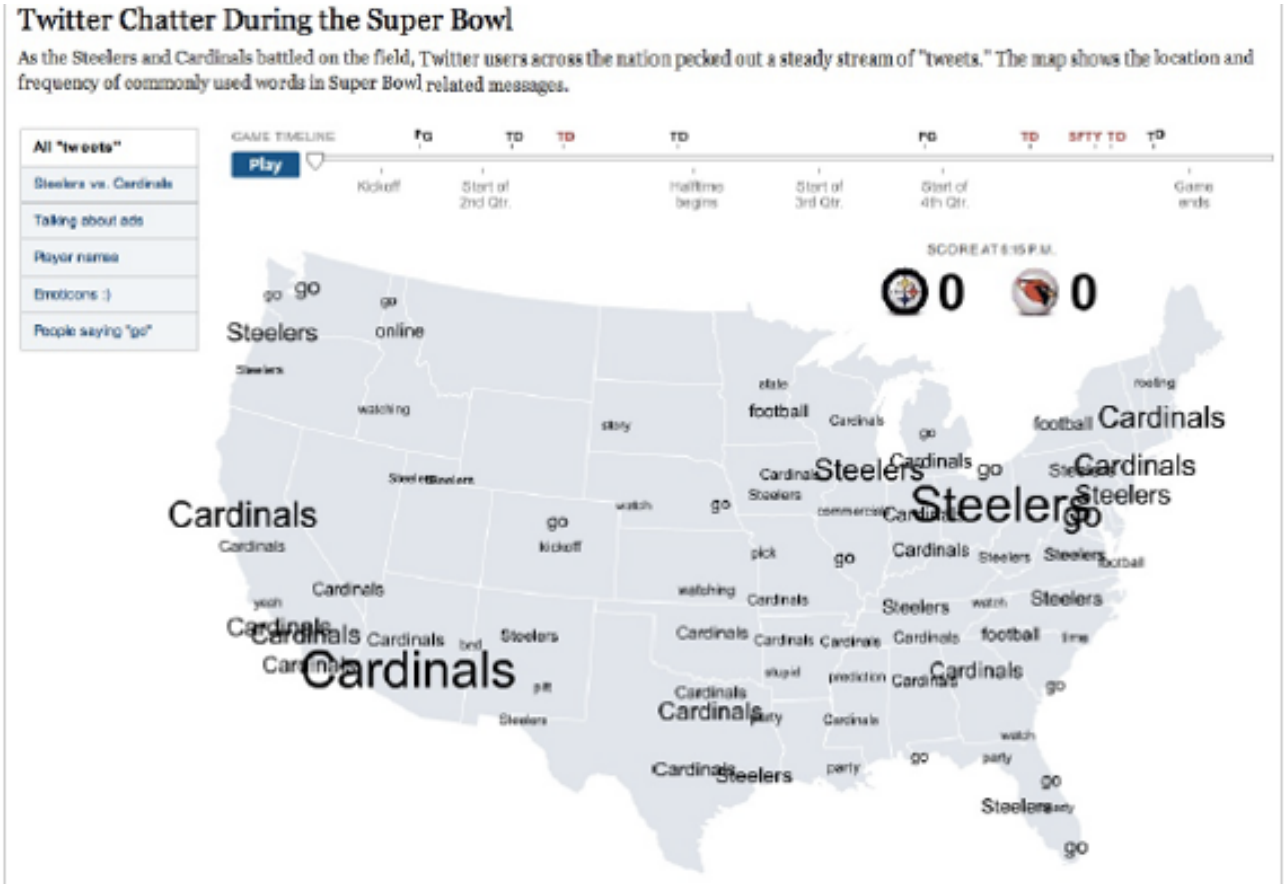
The NYT's U.S. House Election Results 2018

Proportional Symbols Map



The NYT's U.S. Coverage of the Bush Vs Kerry Presidential Elections

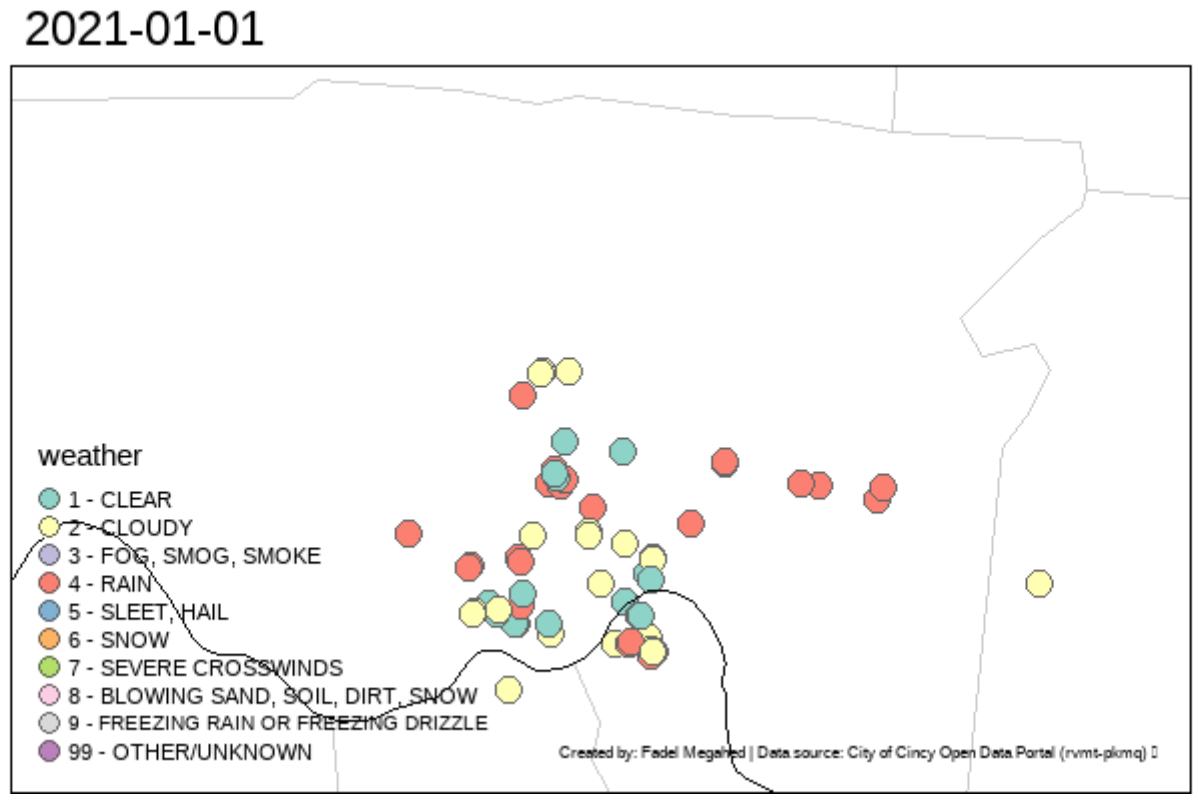
Proportional Symbols Map



The NYT's U.S. Coverage of 2009 Super Bowl

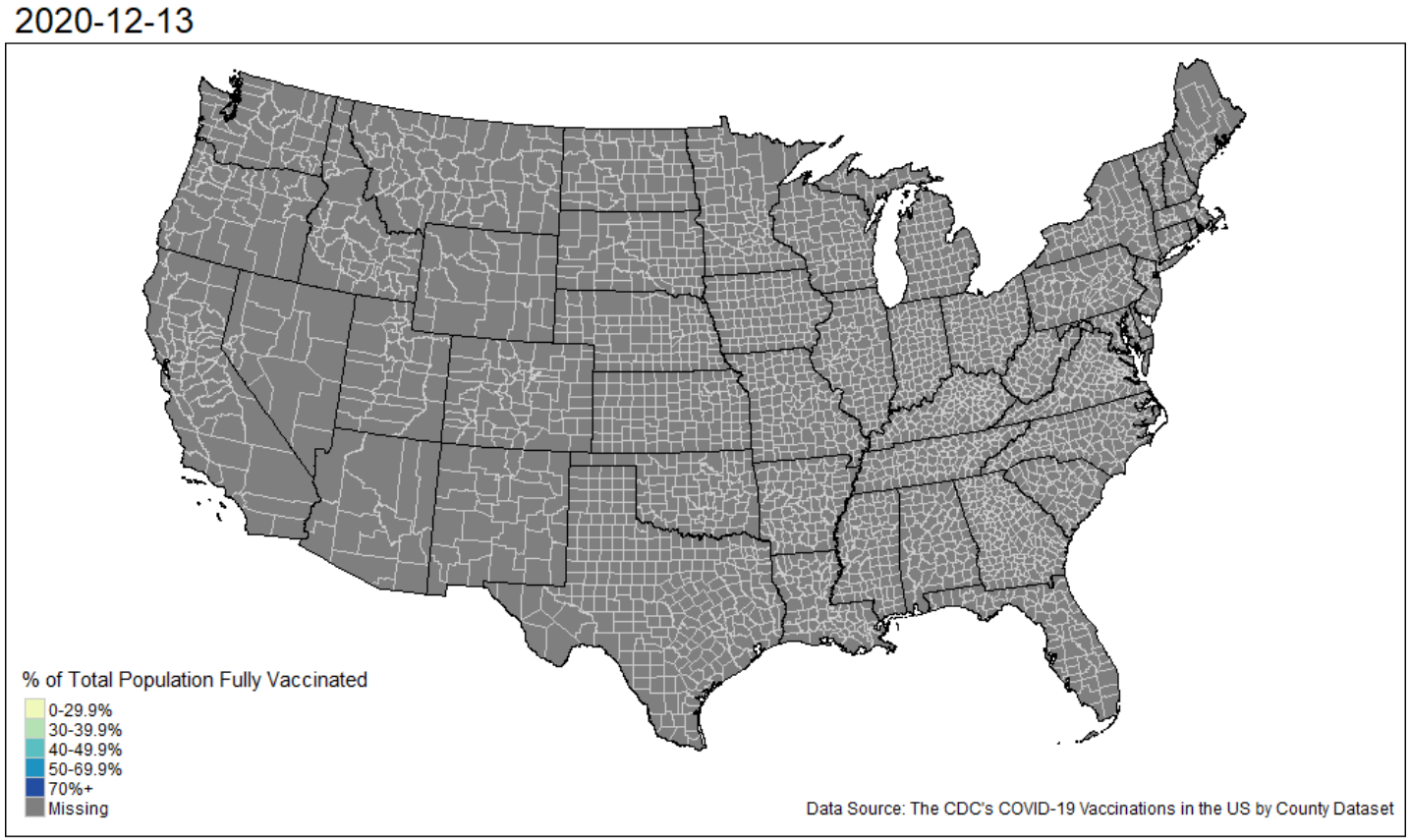
Spatiotemporal Maps

Spatiotemporal Maps



A visualization of car crashes in the Cincinnati metro area

Spatiotemporal Maps



Software Demo

Exploring the Cincy Crashes Dataset

Let us use Tableau to explore the cincy_2021_crashes.csv, where we will create the following:

- A calculated field titled `unique_count`
- A plot of the total number of unique crashes per day
- A table of number of unique crashes by week day
- A table of number of unique crashes by week day
- An animated symbols map

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
- Explain the different types of spatial plots
- Select suitable spatial graphs for different scenarios
- Understand how spatiotemporal plots can help in storytelling (what makes BI special)