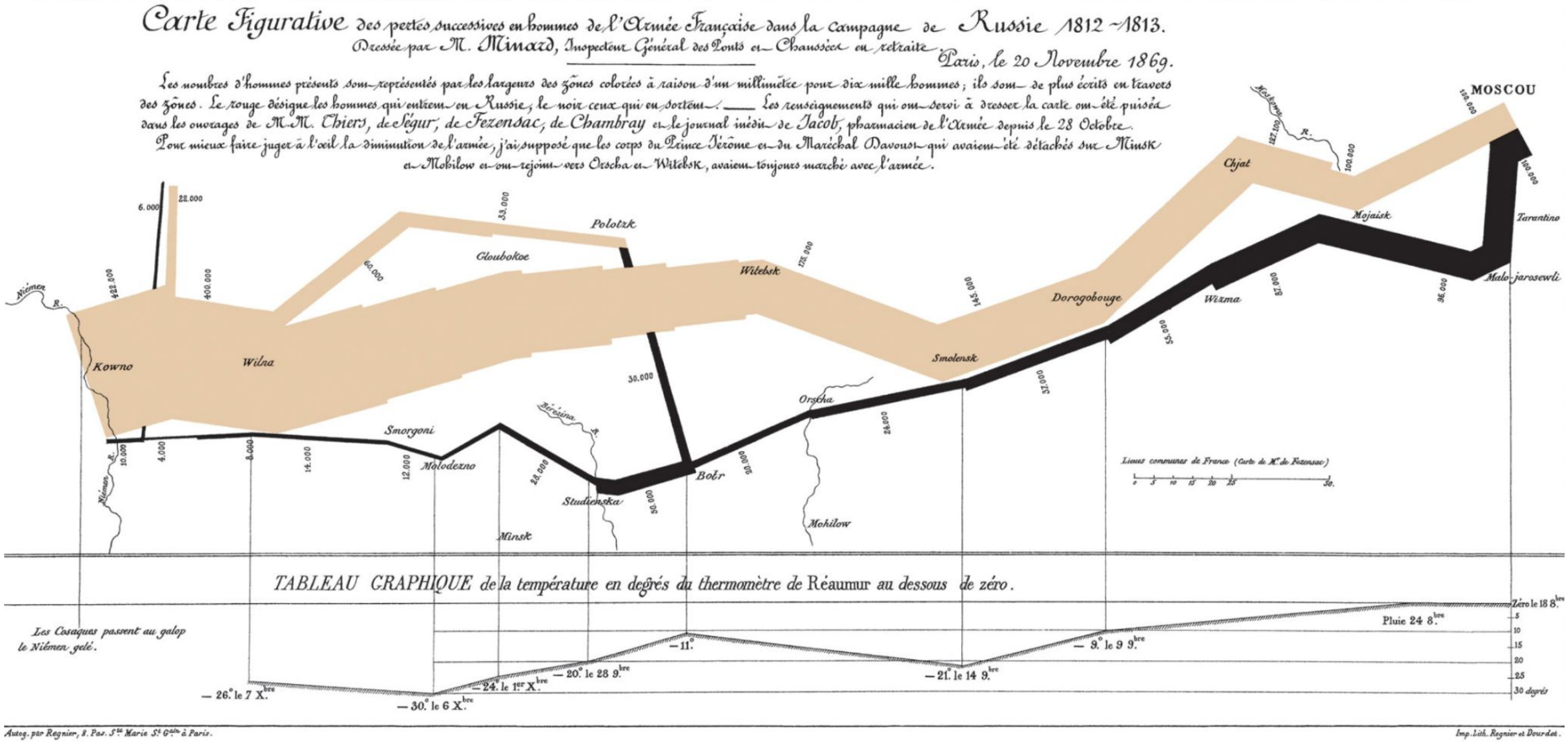
Week 4
Time Series Unit

Elements of Data Visualization

Feb 2, 2026

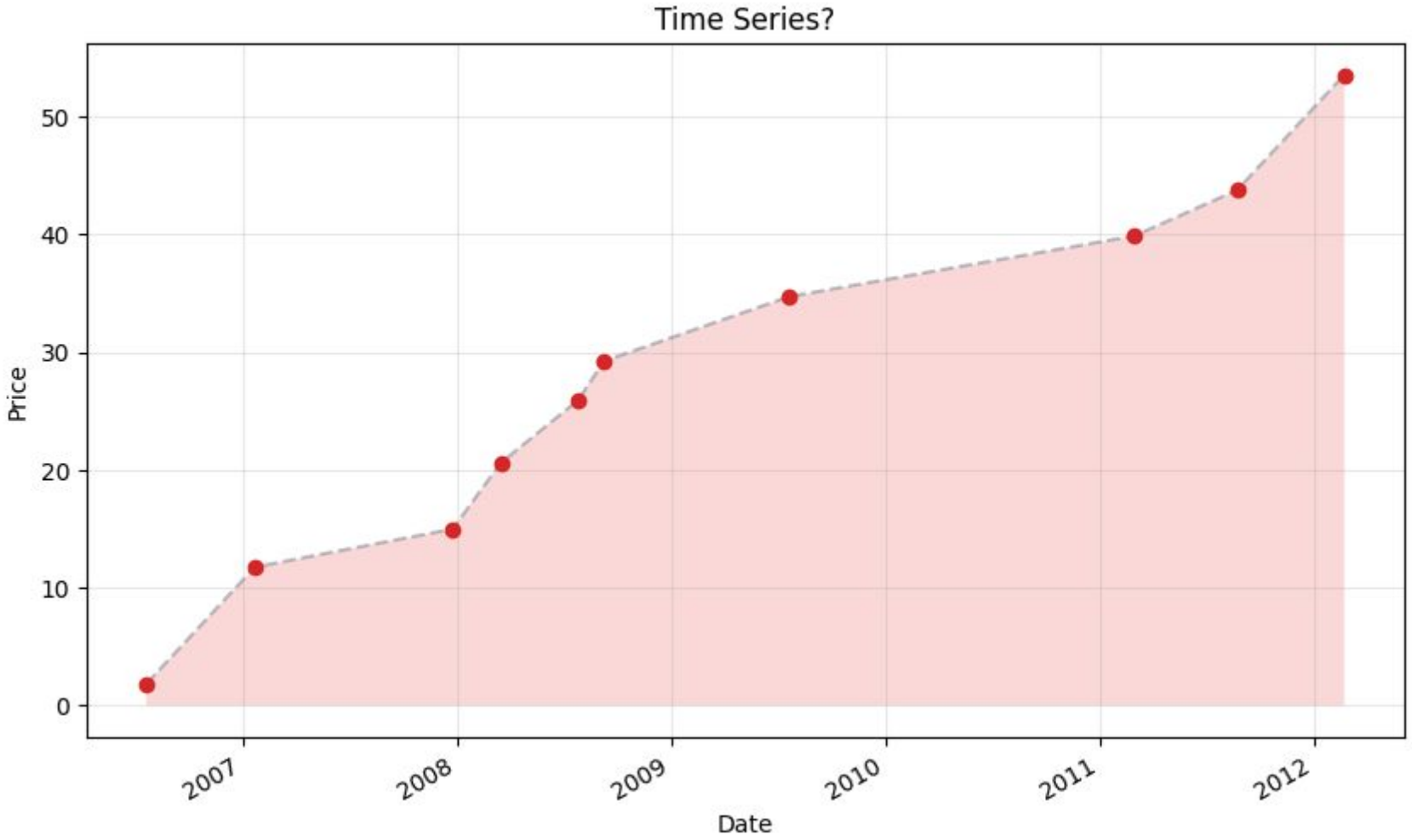

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.
Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de MM. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.
MOSCOU
Kowno
Wilna
Smorgoni
Molodezno
Minsk
Studienska
Bobr
Mohilow
Orscha
Smolensk
Dorogobouge
Wixma
Chjat
Mojaisk
Malo-jarosewli
Tarentino
Polotzk
Gloubokoe
Witebsk
Niemen R.
Beresina R.
Lieues communes de France (Carte de M.r de Fezensac)
TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
— 26.° le 7 X.bre
— 30.° le 6 X.bre
— 24.° le 1.er X.bre
— 20.° le 28 9.bre
— 11.°
— 21.° le 14 9.bre
— 9.° le 9 9.bre
Pluie 24 8.bre
Zéro le 18 8.bre
Autog. par Regnier, 8. Pas. St.te Marie St. Gath. à Paris.
Imp. Lith. Regnier et Dourdet.

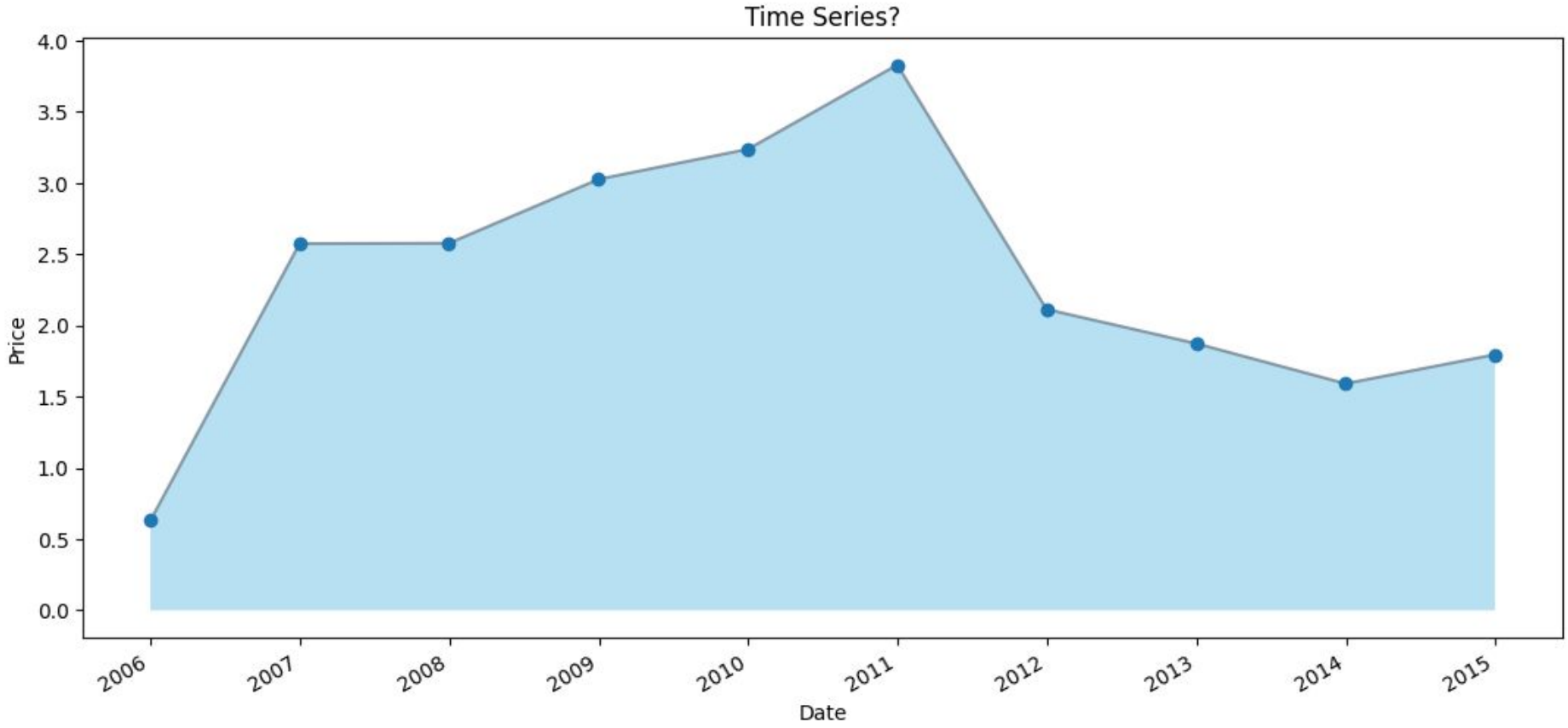
[Source]

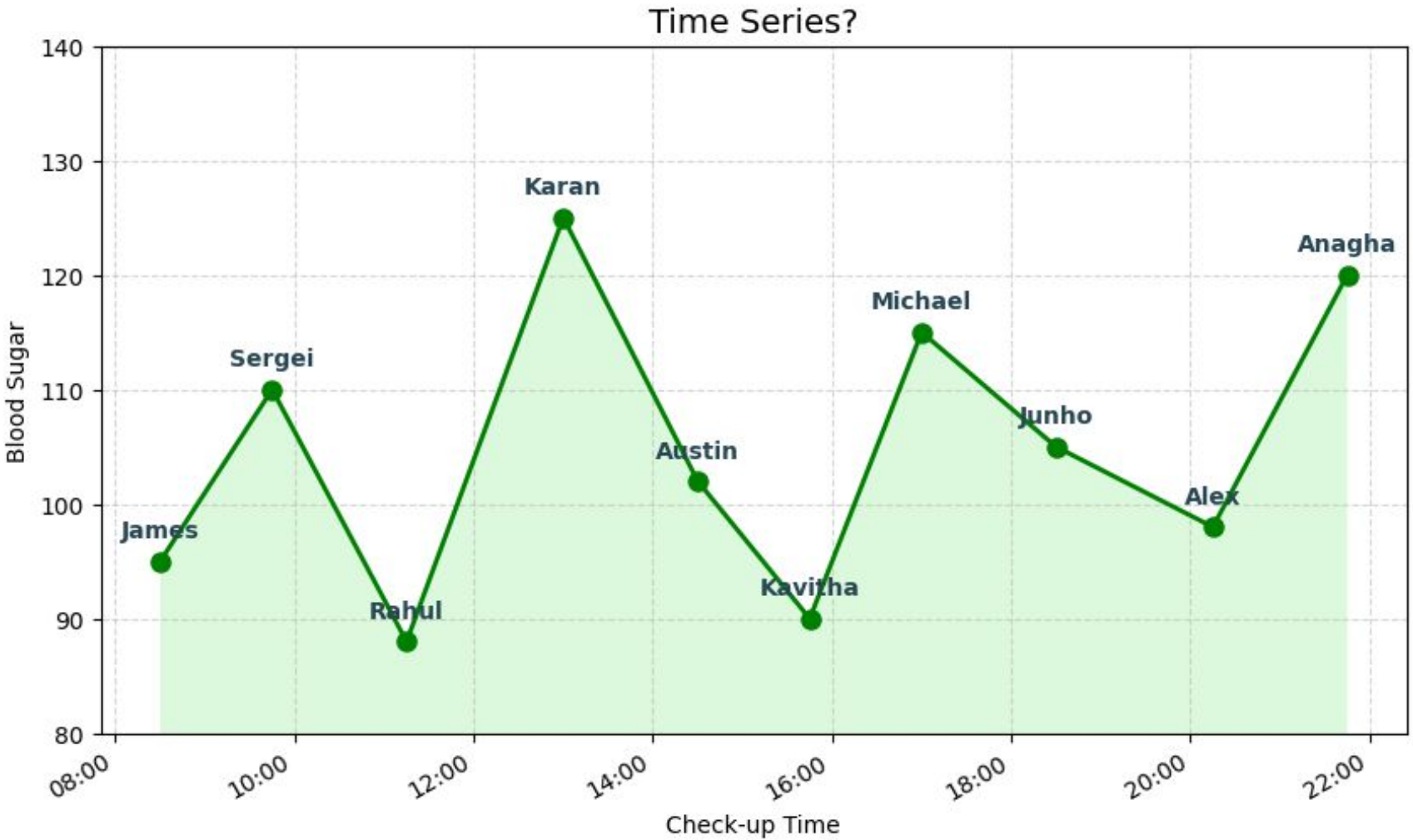
Definition

A time series dataset contains an **ordered** sequence of values of a variable at **equally spaced** time intervals. The data points measured over time may have an internal structure and form patterns and **trends**, which may be **seasonal**.

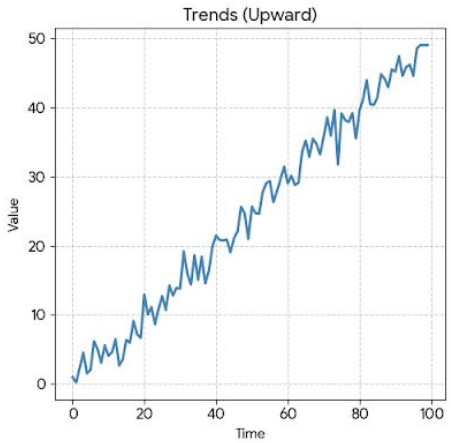
Source: NIST handbook of statistical methods.


Time Series?
Price
Date
0
10
20
30
40
50
2007
2008
2009
2010
2011
2012

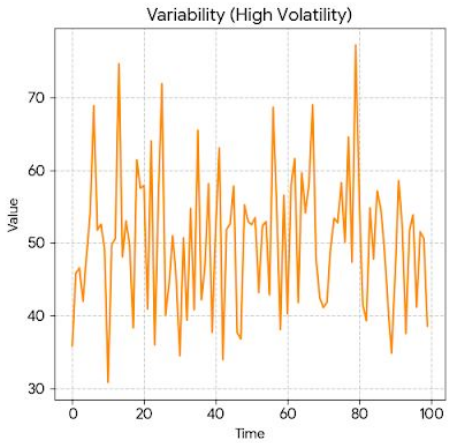

Time Series?
Price
Date
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015


Time Series?
Blood Sugar
Check-up Time
140
130
120
110
100
90
80
08:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
James
Sergei
Rahul
Karan
Austin
Kavitha
Michael
Junho
Alex
Anagha

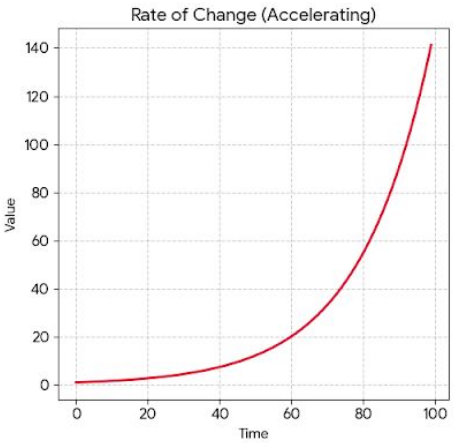
Time Series Properties



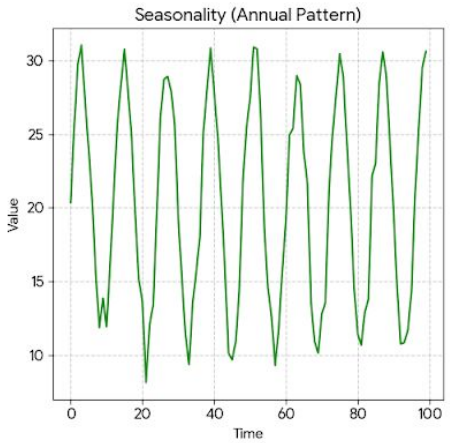
Overall tendency of series of values to increase, decrease, or remain stable over a period of time.



Average degree of change from one point in time to the next.

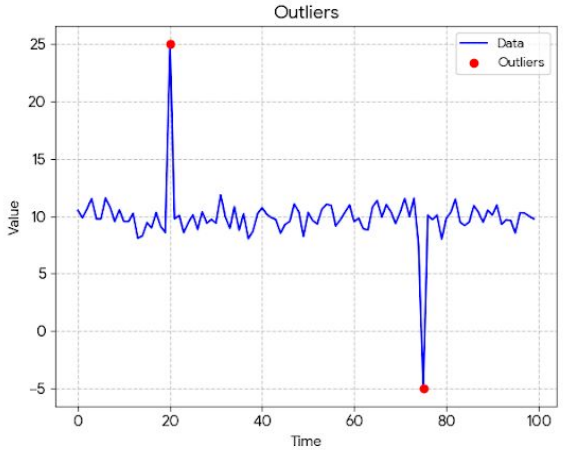


Percentage of difference between one value and the next.

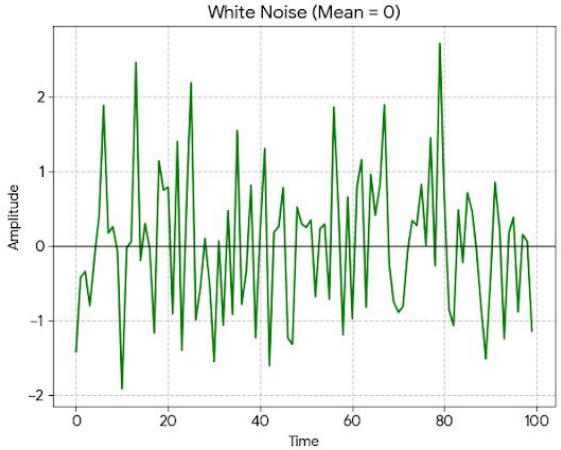


Patterns that repeat at regular intervals that correspond to calendar events.

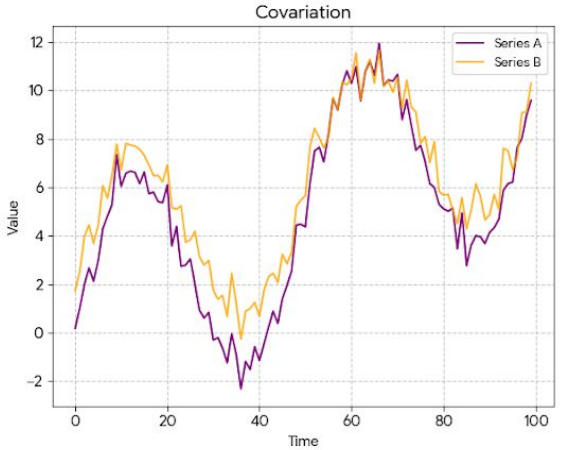
Time Series Properties



Values that fall outside the norm.

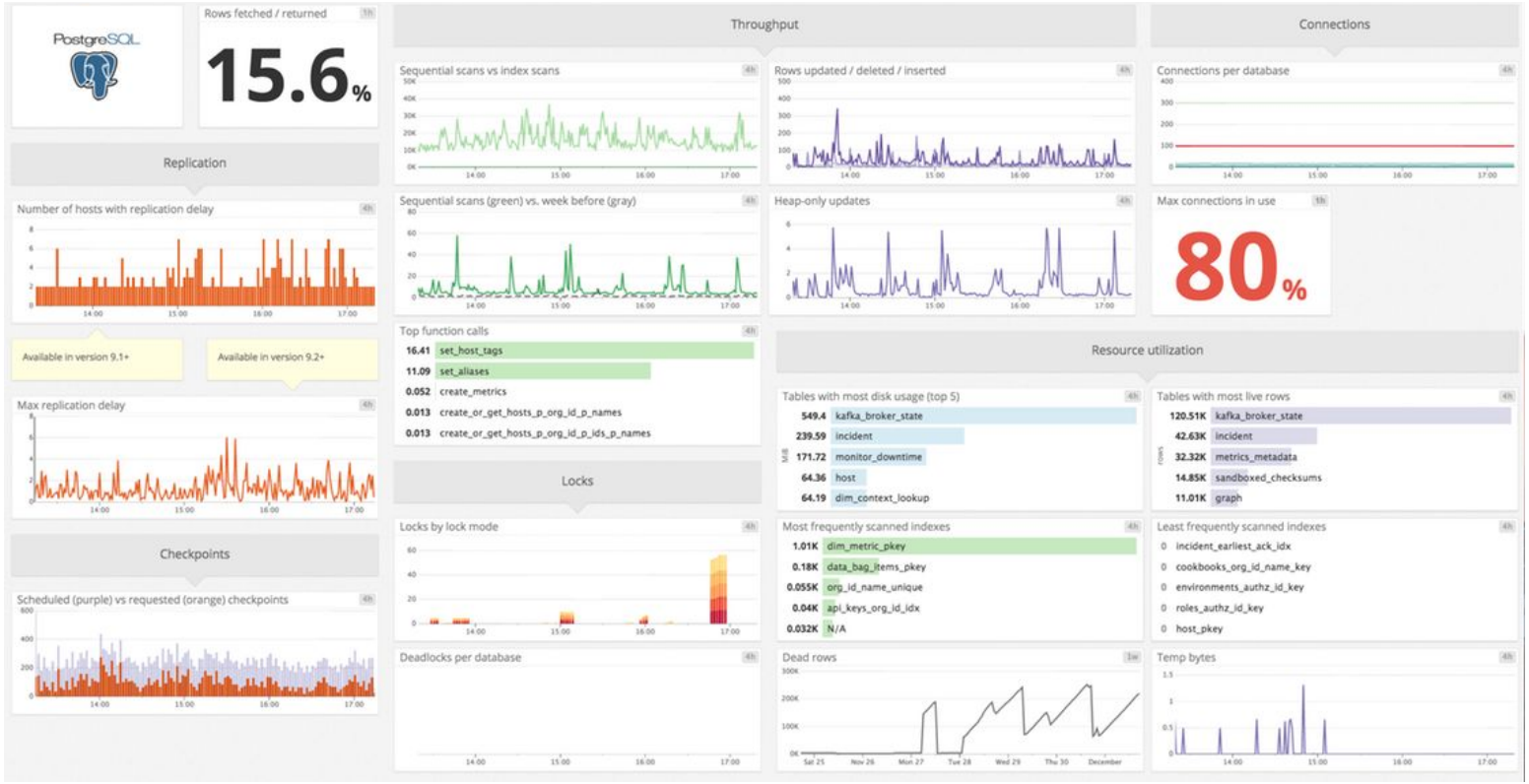


Randomness in a time series, with a mean value equal to zero.

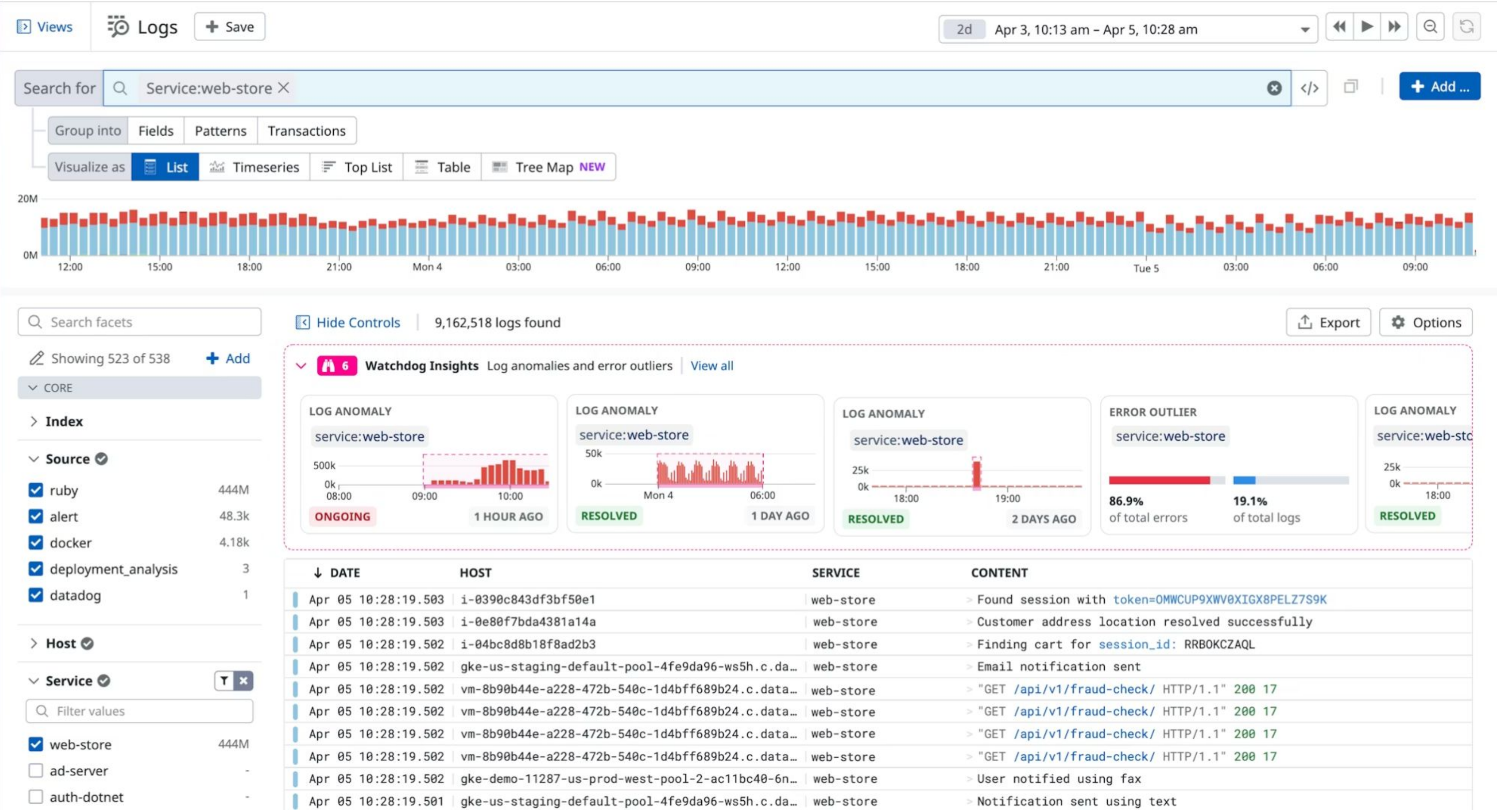


Changes in a time series are reflected in another, immediately or later, and in the same or different direction.

Practical Use Case: Crisis Diagnosis

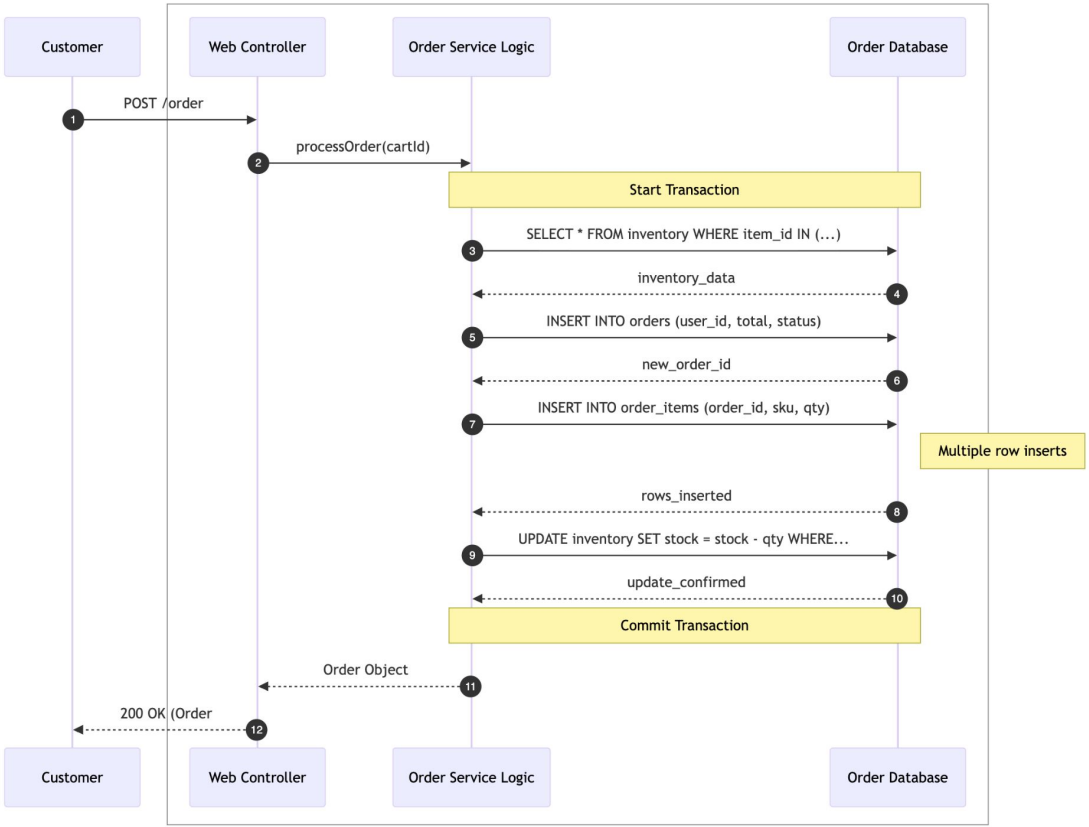


Practical Use Case: Anomaly Detection

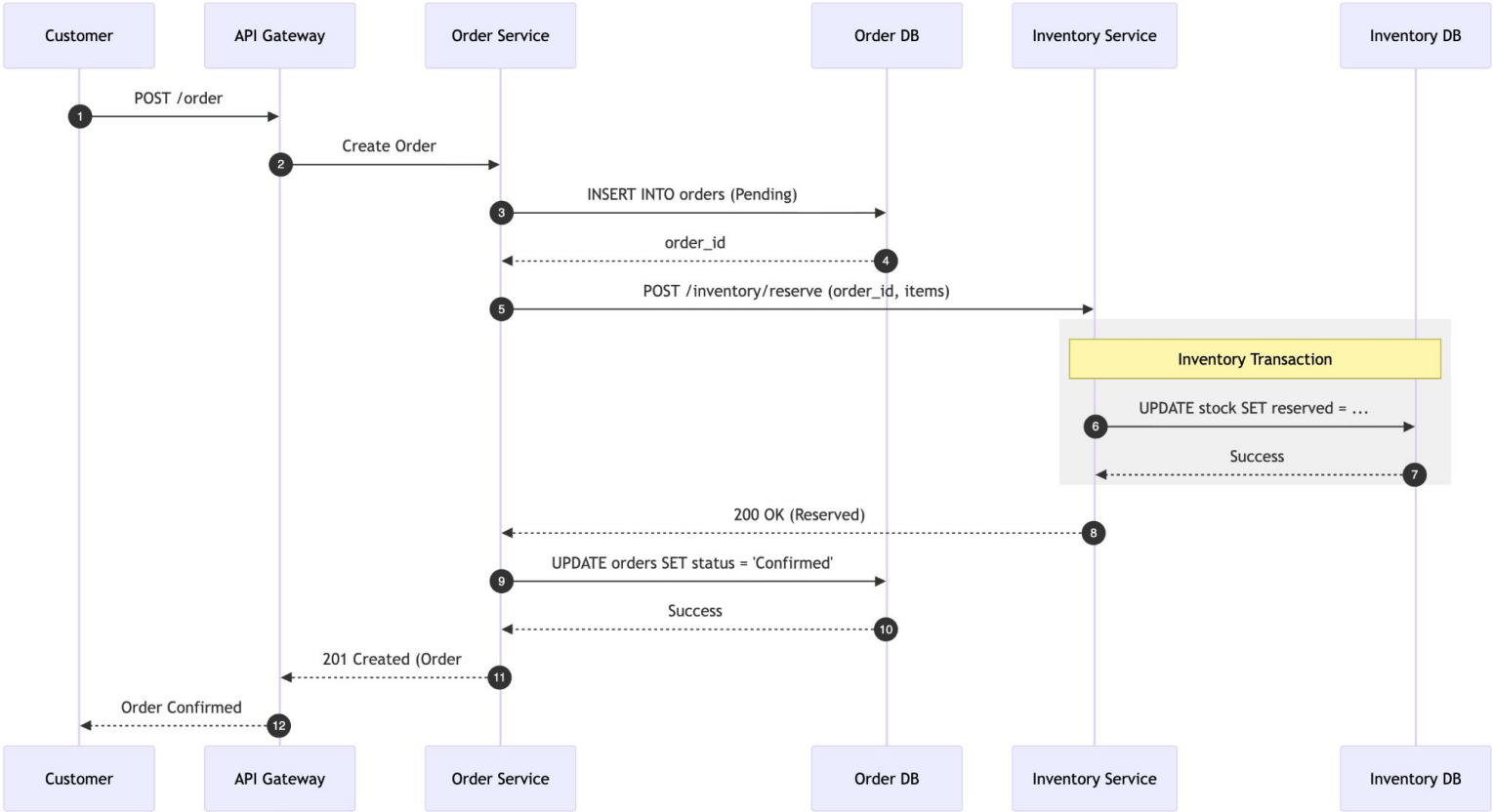


Practical Use Case: Capacity Planning

Practical Use Case: System Design

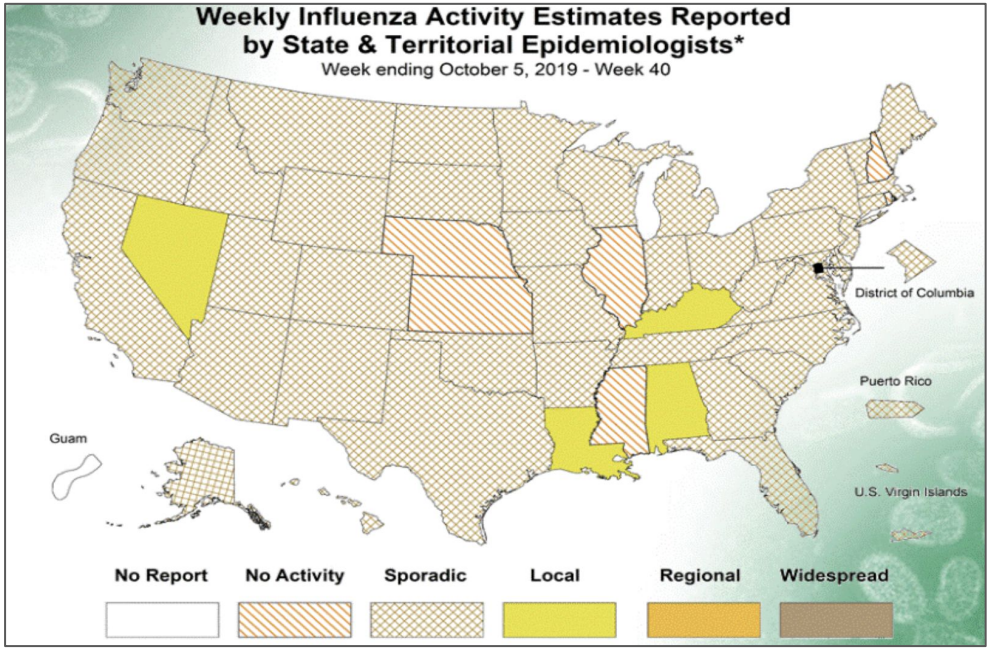


Practical Use Case: System Design



Time Series Study

Primary Dataset: CDC Weekly US Influenza Activity Estimates Report



Time Period: Oct. 2005 - May 2025

- Data file 1: Oct 2005 - May 2020
- Data file 2: Oct 2020 - May 2025

Notable fields: *state, activity estimate, week number,* and *season*.

Complexity: Files use different types of measurements. Geographic spread versus clinical intensity of the disease.

Next step: obtain guidance from CDC on compatibility of their measurements.

Supplemental Datasets: Google Trends

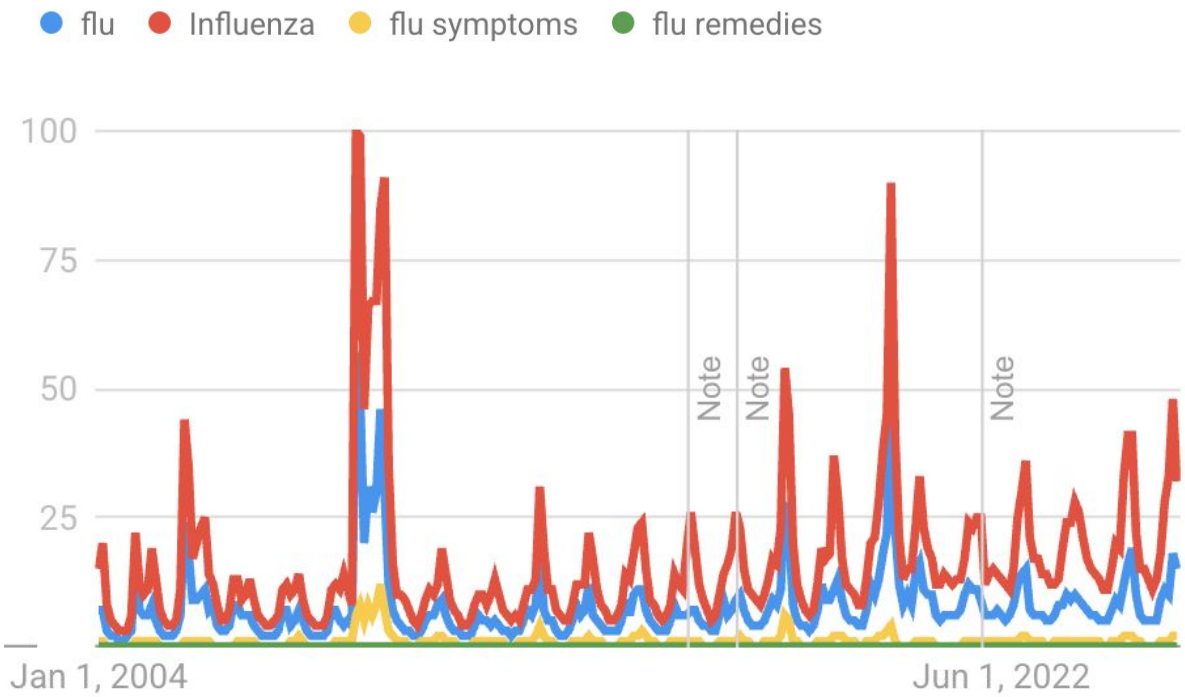
Search Terms:

- Flu
- Flu symptoms
- Flu remedies

Disease:

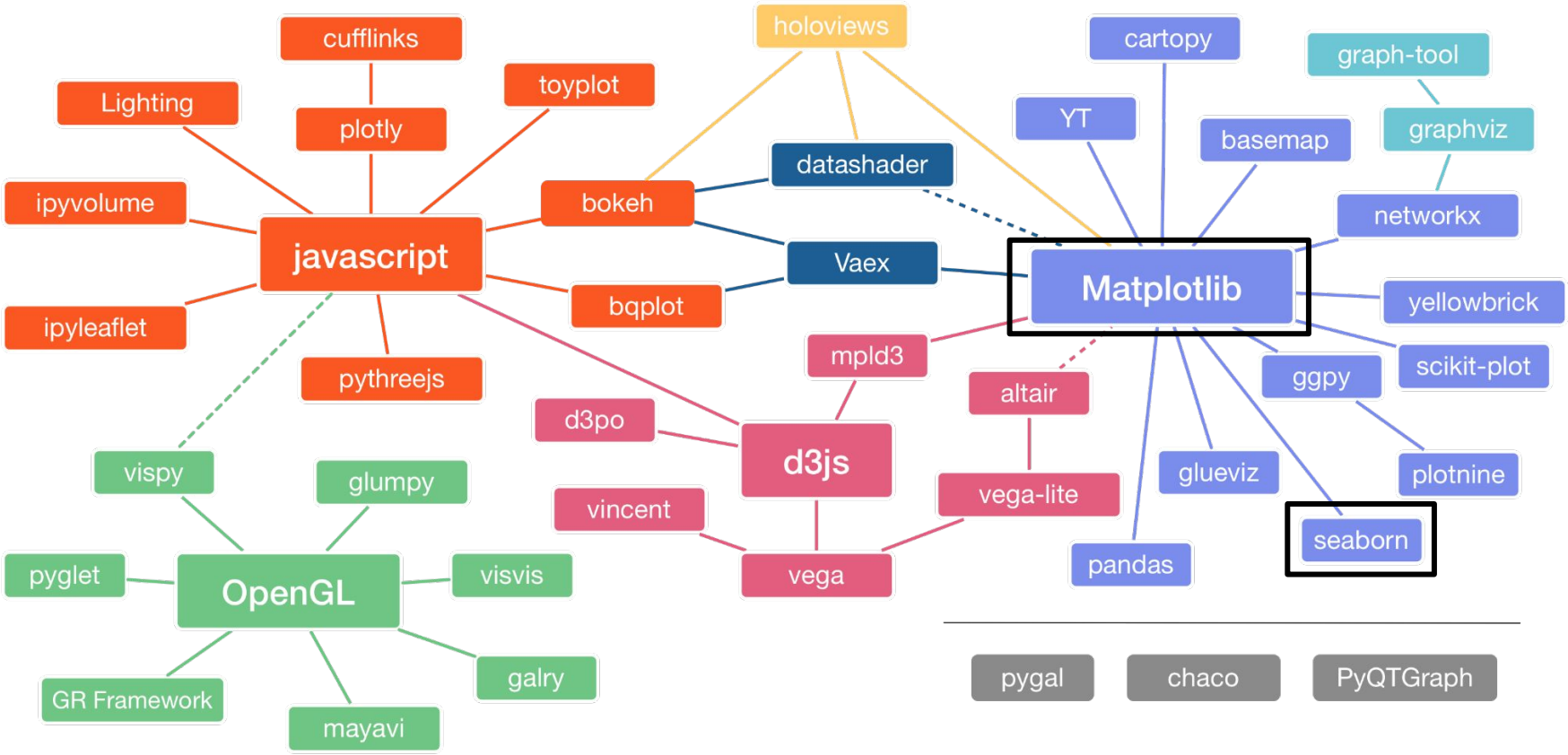
- Influenza

Time Period:
Jan 2004 - Dec 2025



Guiding Questions

- What do flu activity estimates trends look like from 2005 - 2025?
  - Across US states and territories
  - For a specific state (e.g. Texas)
- Do all states and territories seem to have similar trends over the years?
- Have flu activity estimates become more or less widespread from 2005 - 2025?
- Have flu activity estimates remained relatively stable from 2005 - 2025?
- Do any years stand out for having more widespread flu activity estimates?
- Which states have had the most widespread flu activity estimates?
- Which states have had the least widespread flu activity estimates?
- Do neighboring states and territories have similar flu activity estimates?
- Do Google Search trends follow similar patterns to flu activity estimates?



Source: https://pyviz.org/overviews/index.html

Python libraries we'll be working with

- **Pandas:** common data manipulation and analysis operations on numerical tables and time series.
- **Matplotlib:** static, animated, and interactive visualizations.
- **Seaborn:** high-level interface for plotting statistical graphs:
  - countplot() visualize the data distribution of categorical variables.
  - displot() visualize the data distribution of numerical variables.

Let’s dive in:

Load and examine the data

Working with time fields

- Common challenges:
  - Joining several datasets that use different date formatting
  - Converting a set of points from one time zone to another
  - Aggregating data that’s been sampled daily, resulting in weekly aggregates

- Solution: format timestamps
  - Transform time data to datetime type
  - Datetime can either be naive or aware

Handling missing data

- Common challenges:
  - Human error in recording data
  - Human decision not to record data
  - Input pipeline failure

- Solutions:
  - Imputation
  - Interpolation
  - Removal

Adjusting data sampling frequency

- Common challenges:
  - Join two sets of points that have been sampled at different intervals
  - Sampling granularity isn’t appropriate for your analysis (e.g. daily instead of hourly samples)
  - Dealing with seasonal data, but care for a narrower time interval (e.g. summer months instead of every month)

- Solutions:
  - Downsampling
  - Upsampling

Let’s dive in:

Time series: basic tasks