

Time Series Visualization

Principles, Importance, and Basic Techniques

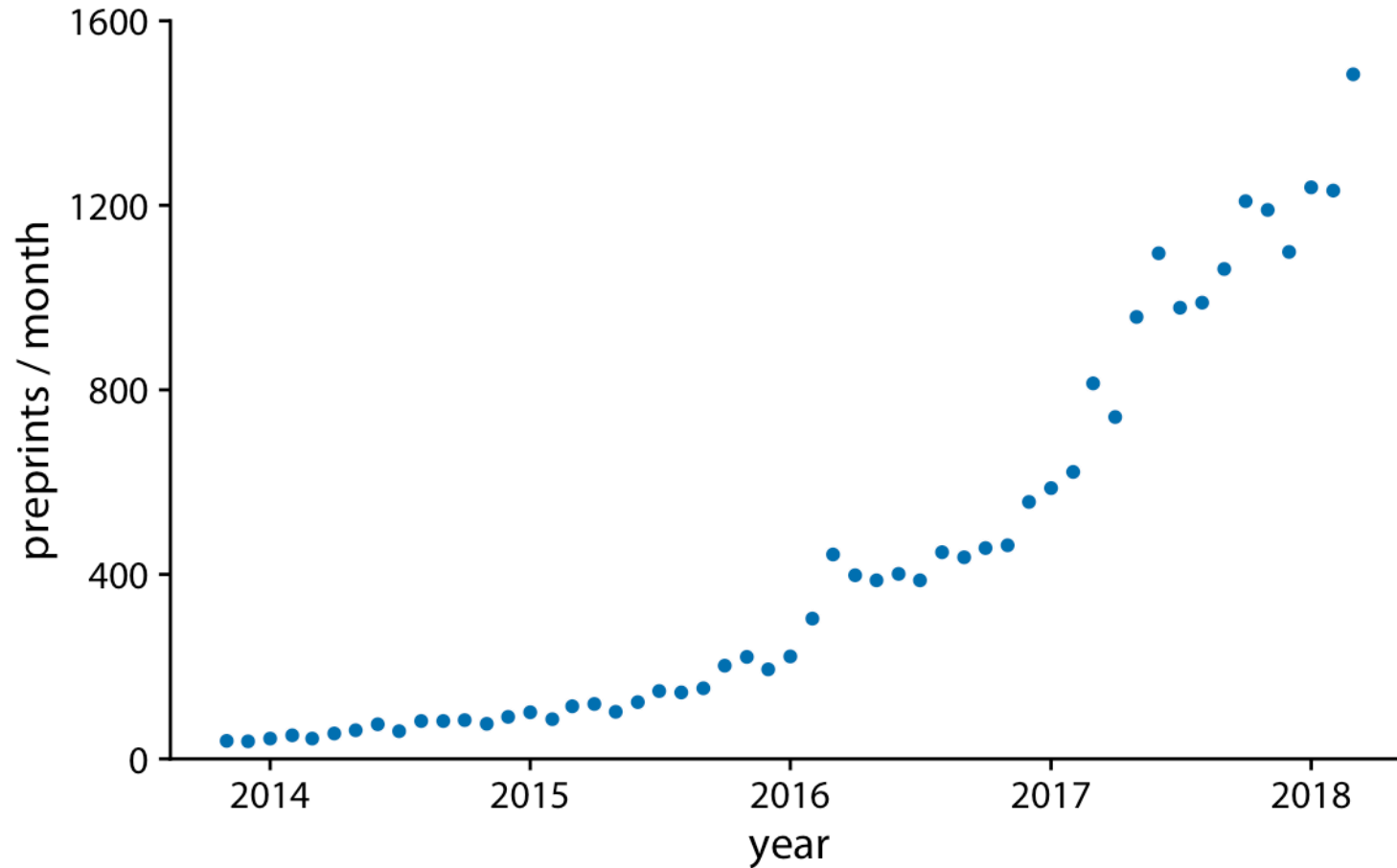
What is Time Series Data?

- Data collected over time at regular intervals.
- Examples: Stock prices, weather data, website traffic.
- Time series visualization involves plotting data points over time to identify trends, patterns, and anomalies.
- Line graphs are commonly used for this purpose, but other visualization techniques can also be applied depending on the dataset and objectives.

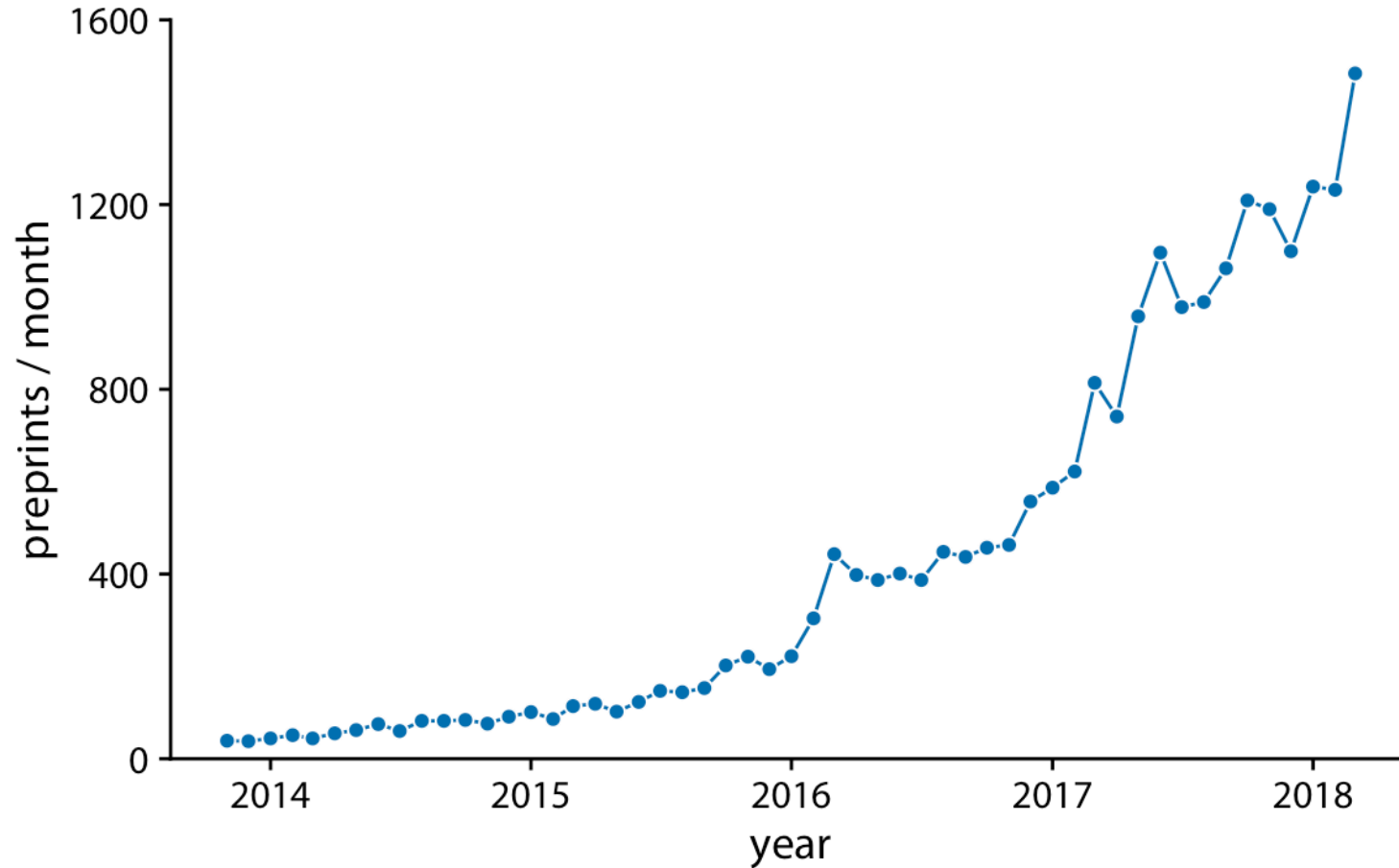
Individual Time Series

- Data points in a time series have an inherent order, usually determined by time.
- Scatter plots with evenly spaced dots can represent time series data, but connecting these dots with lines enhances readability.
- Example: Monthly preprint submissions to bioRxiv (2014–2018)

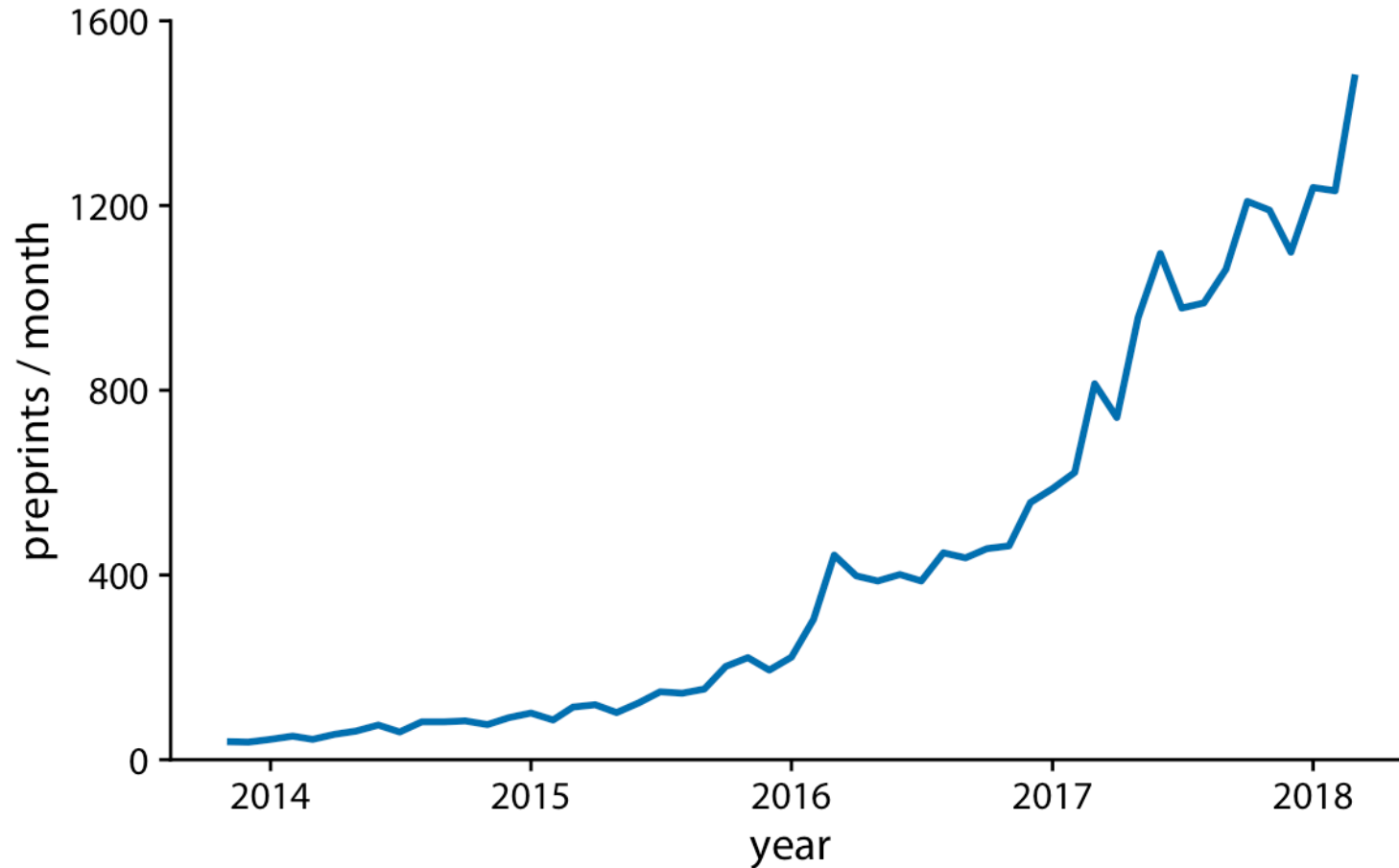
Scatter plot of submission counts per month



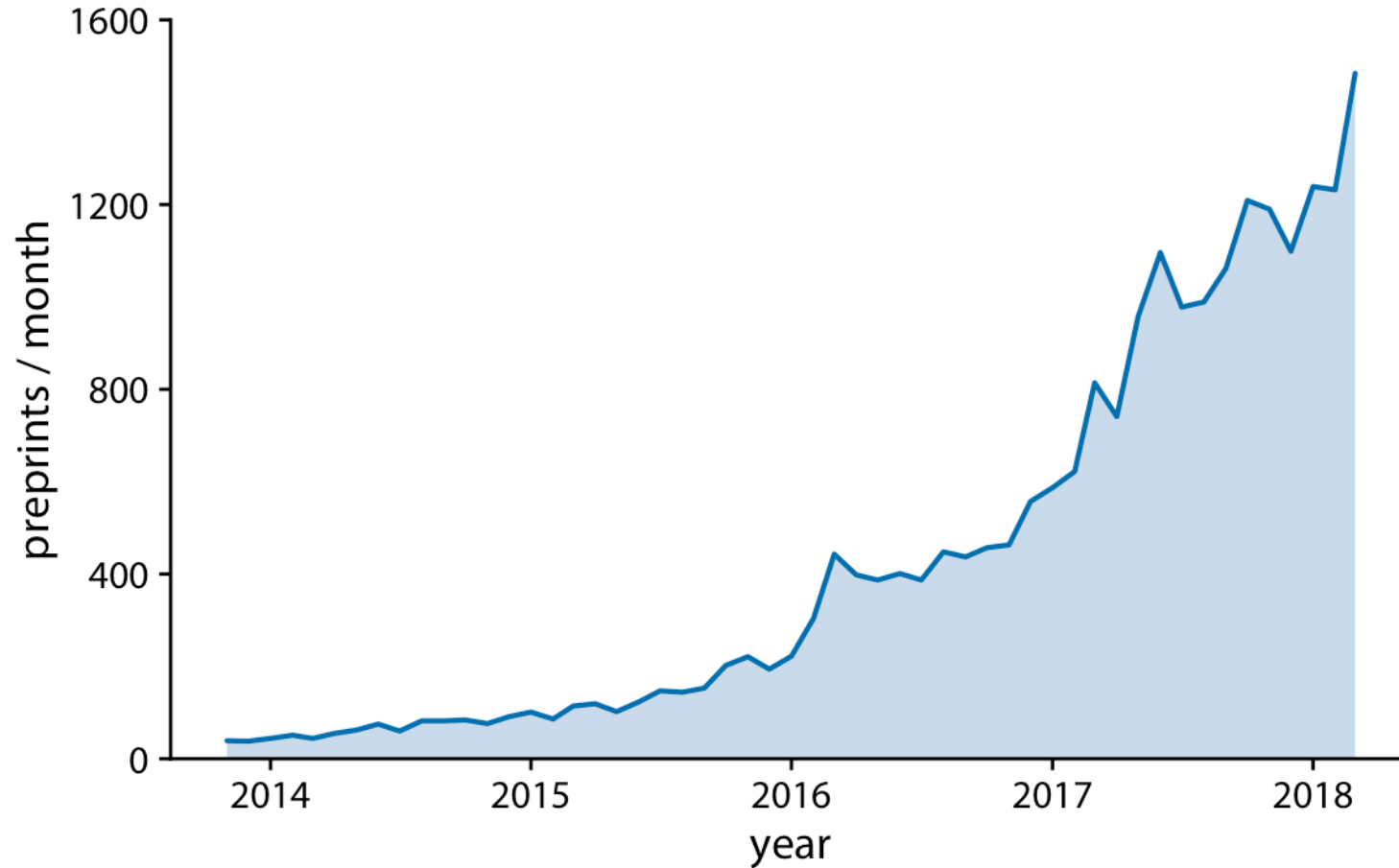
Line graph emphasizing temporal order



Line graph without dots to highlight trends



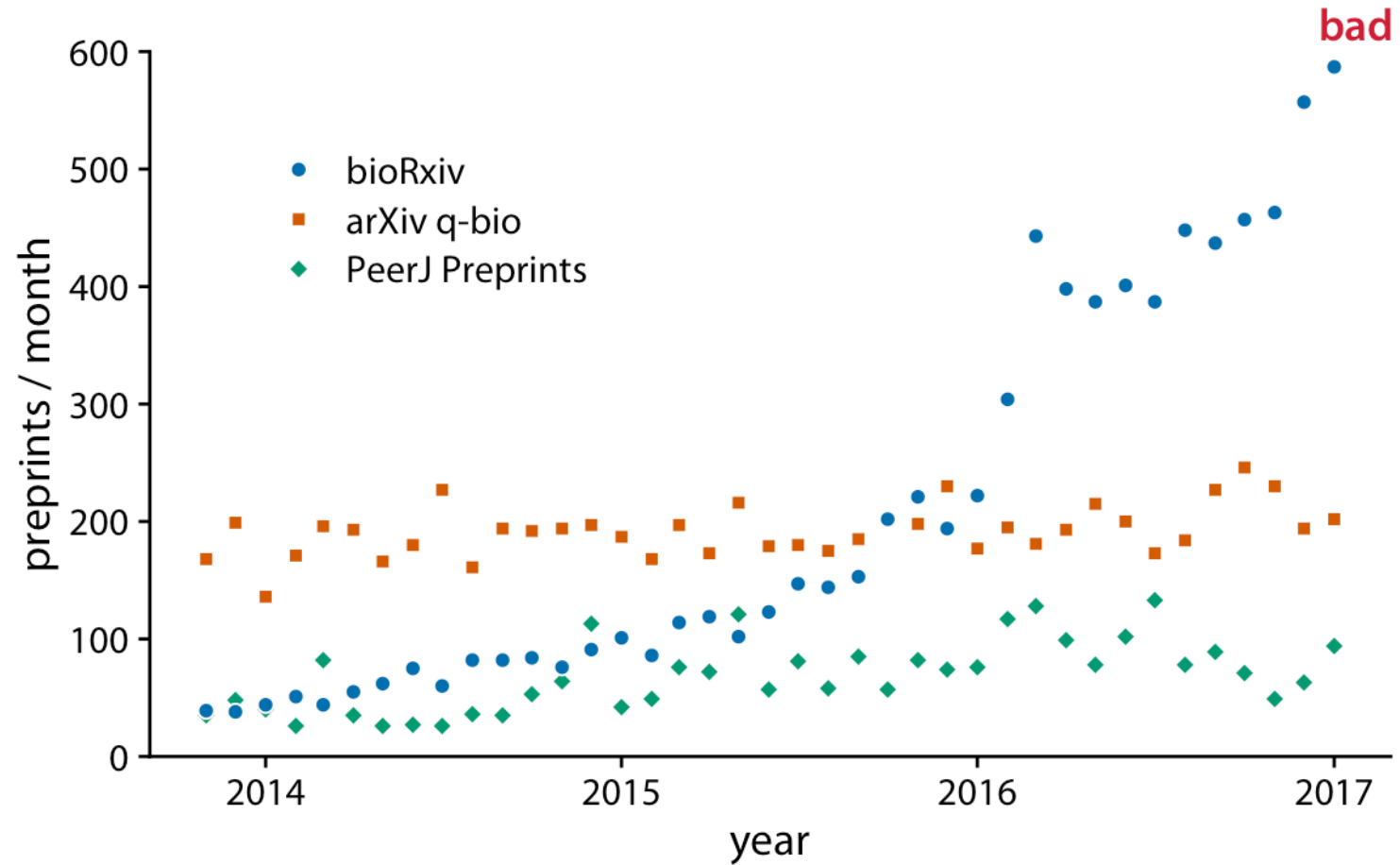
Area-under-curve filled graph



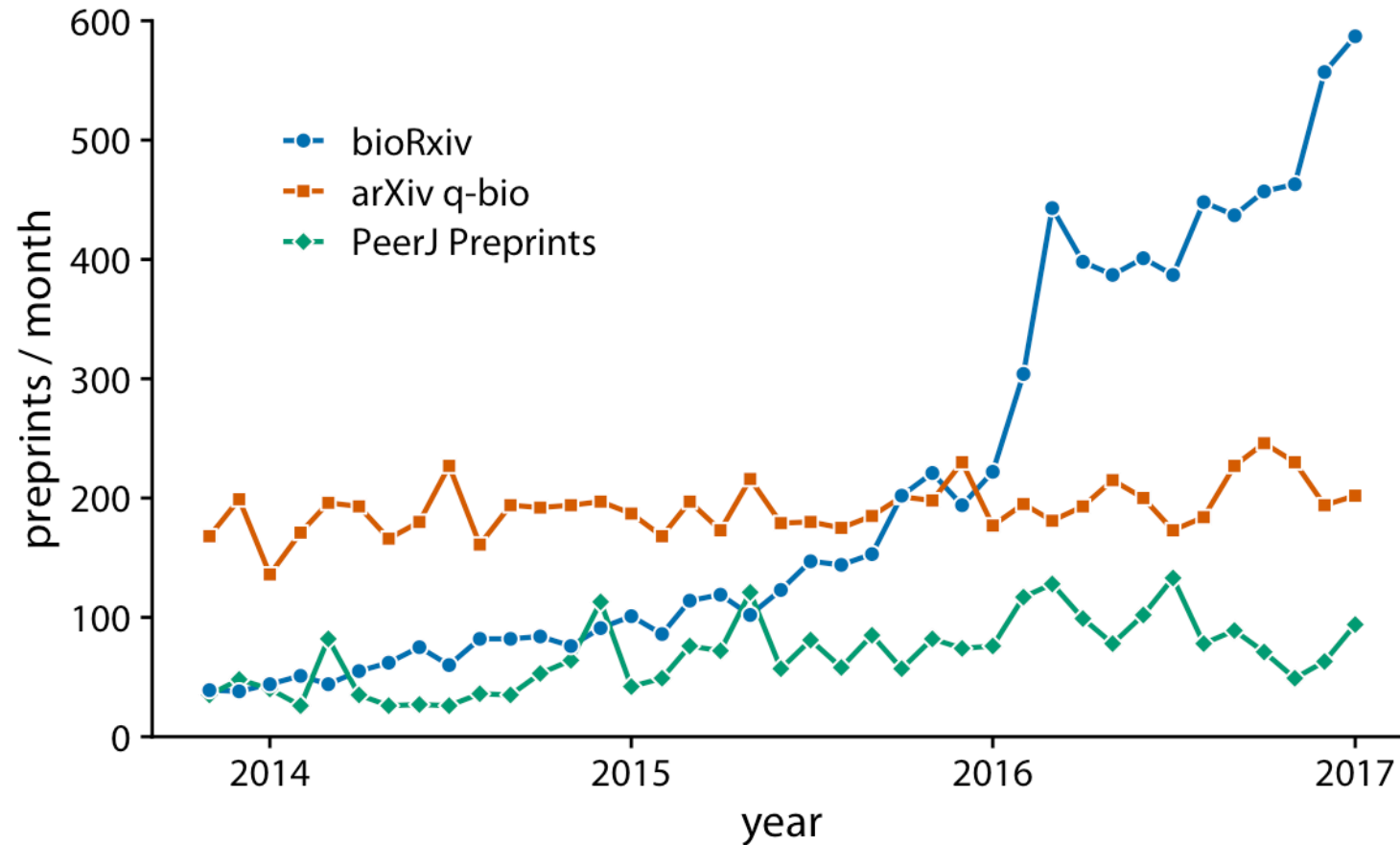
Multiple Time Series and Dose-Response Curves

- When visualizing multiple time series, careful design choices help reduce clutter and improve interpretation.

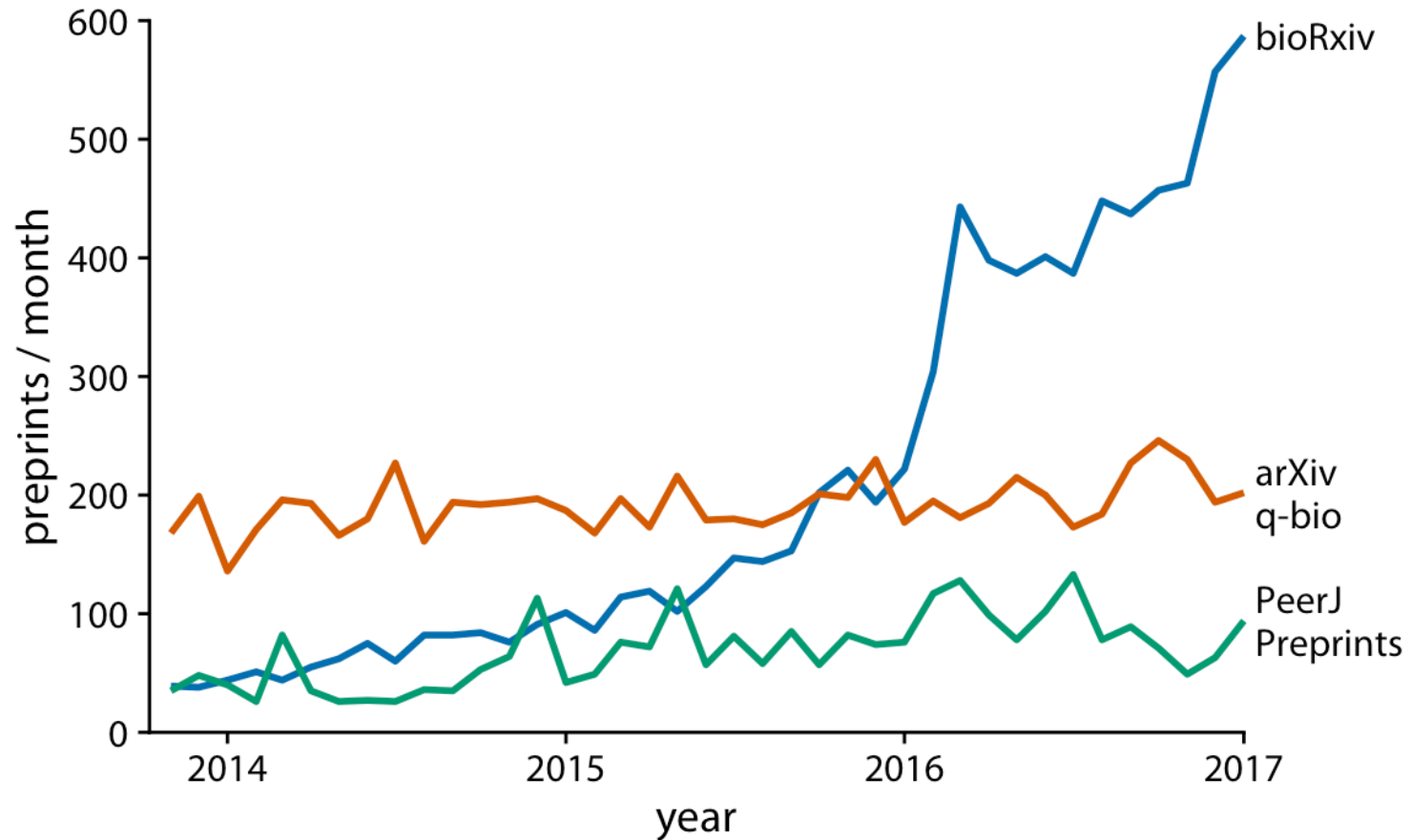
Scatter plot of monthly submissions to 3 preprint servers



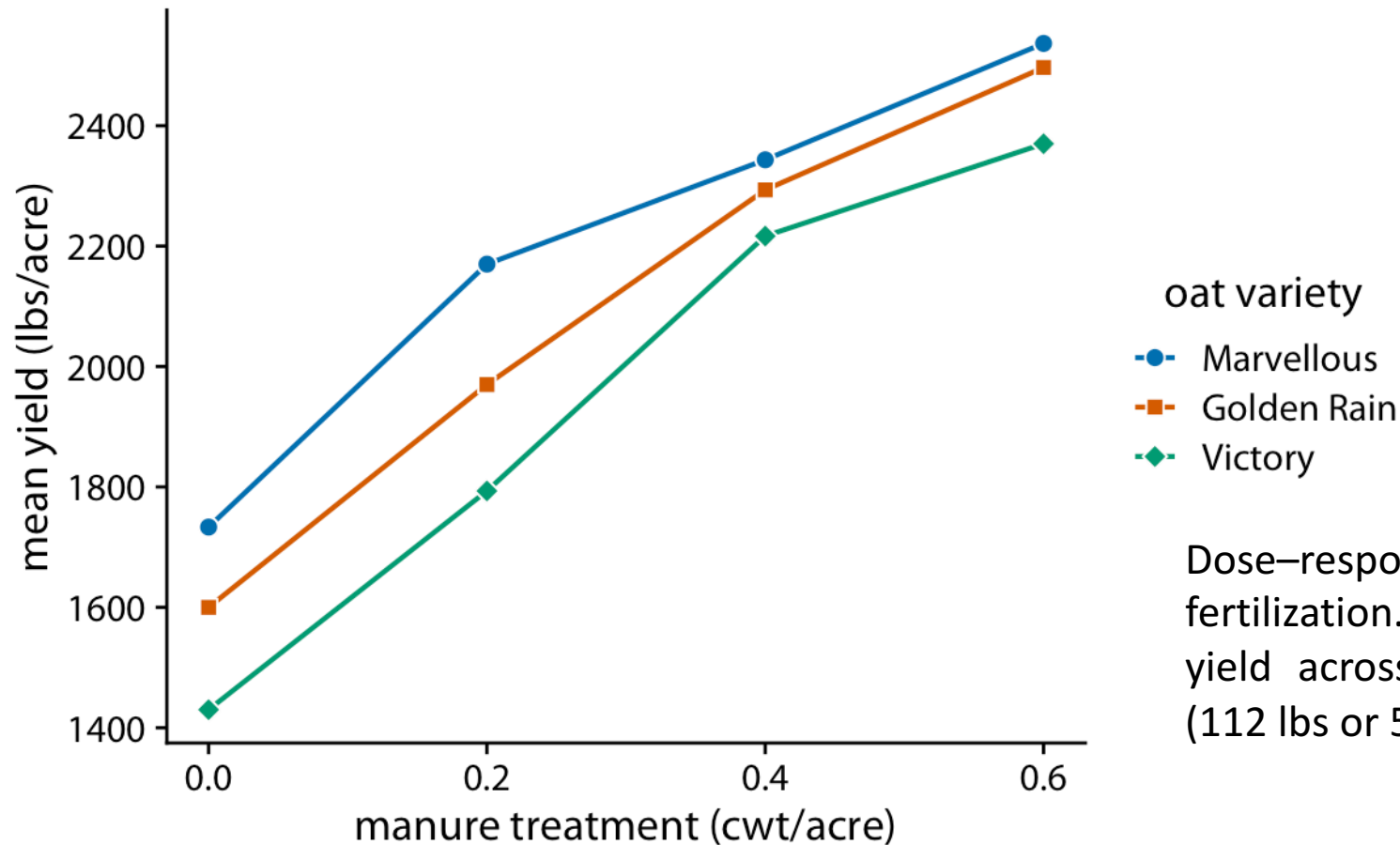
Line graph connecting dots improves readability



Directly labeling lines instead of using a legend reduces cognitive load



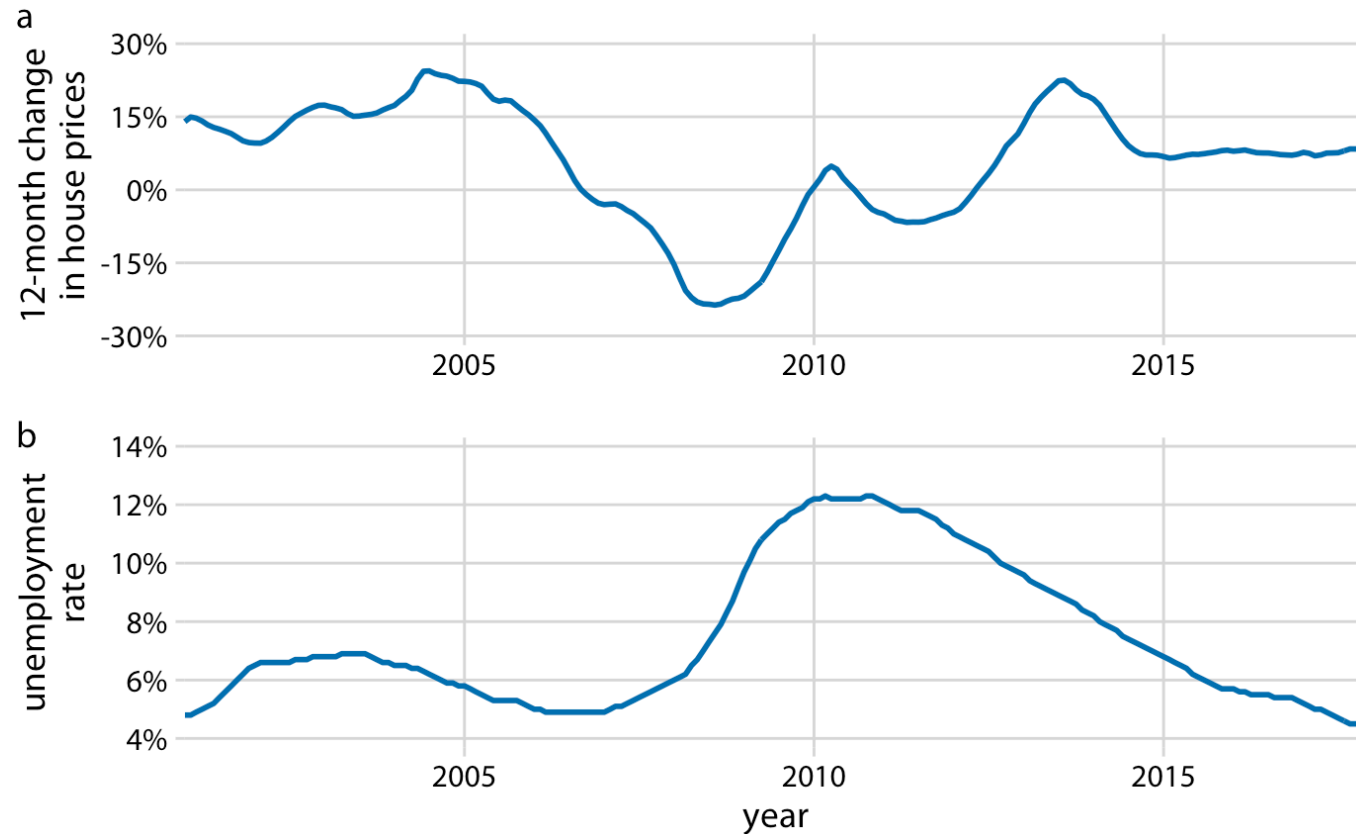
Line graphs can also depict non-temporal relationships



Dose–response curve of mean oat yield after manure fertilization. Manure provides nitrogen, increasing yield across varieties. Application measured in cwt (112 lbs or 50.8 kg) per acre. Data source: Yates.

Time Series of Two or More Response Variables

- Stacked Line Graphs: Separate graphs for each variable (e.g., house price changes and unemployment rate in Figure 9).

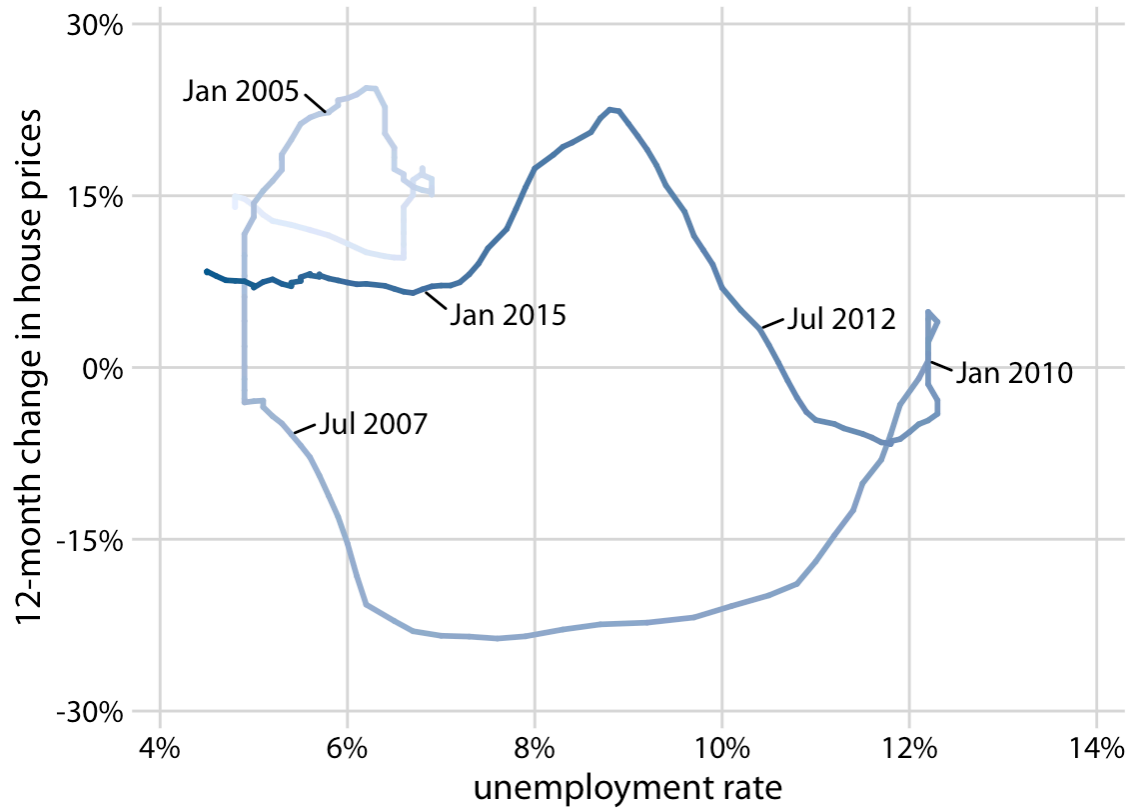


12-month change in house prices (a) and unemployment rate (b) from Jan. 2001 to Dec. 2017.

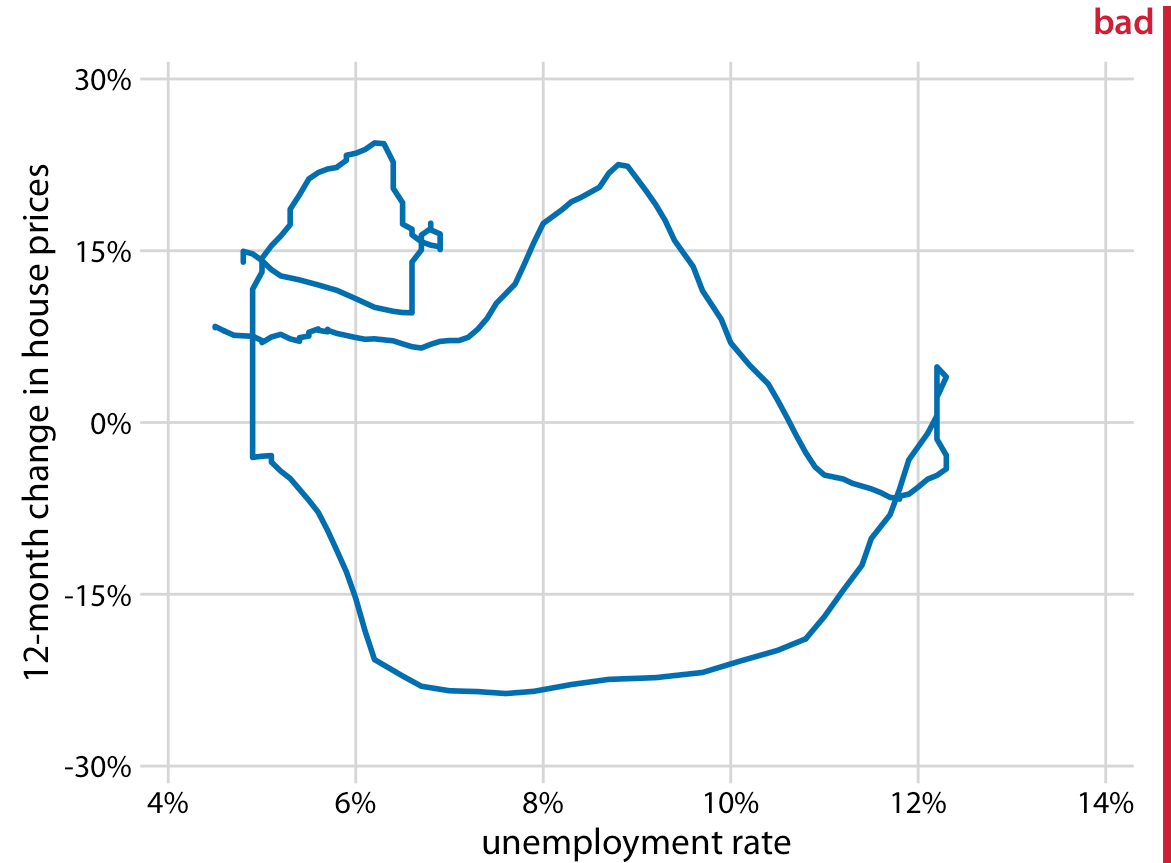
Connected Scatter Plots

- Plotting one variable against another with lines connecting time-sequenced points.
- Example: House price changes vs. unemployment rate showing counter-clockwise cycles.

Lines connecting time-sequenced points



12-month change in house prices vs. unemployment as a connected scatter plot. Darker shades indicate recent months, forming two counterclockwise loops.

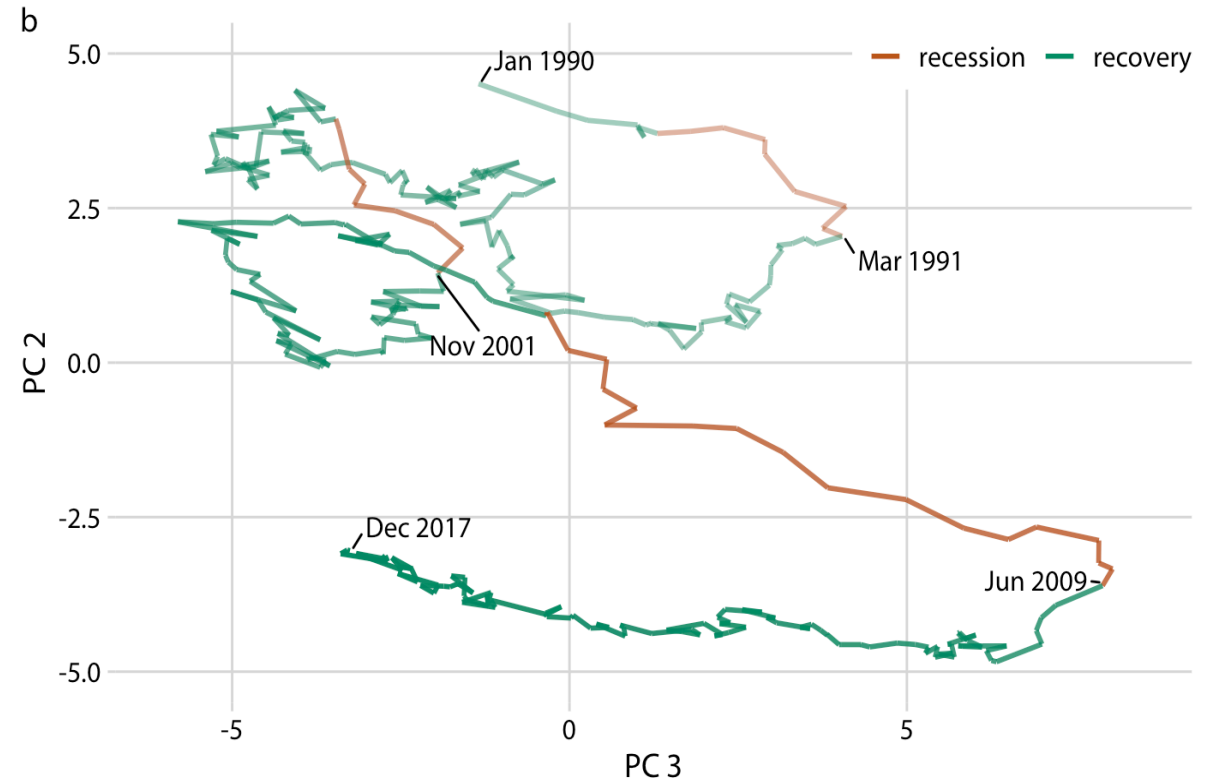
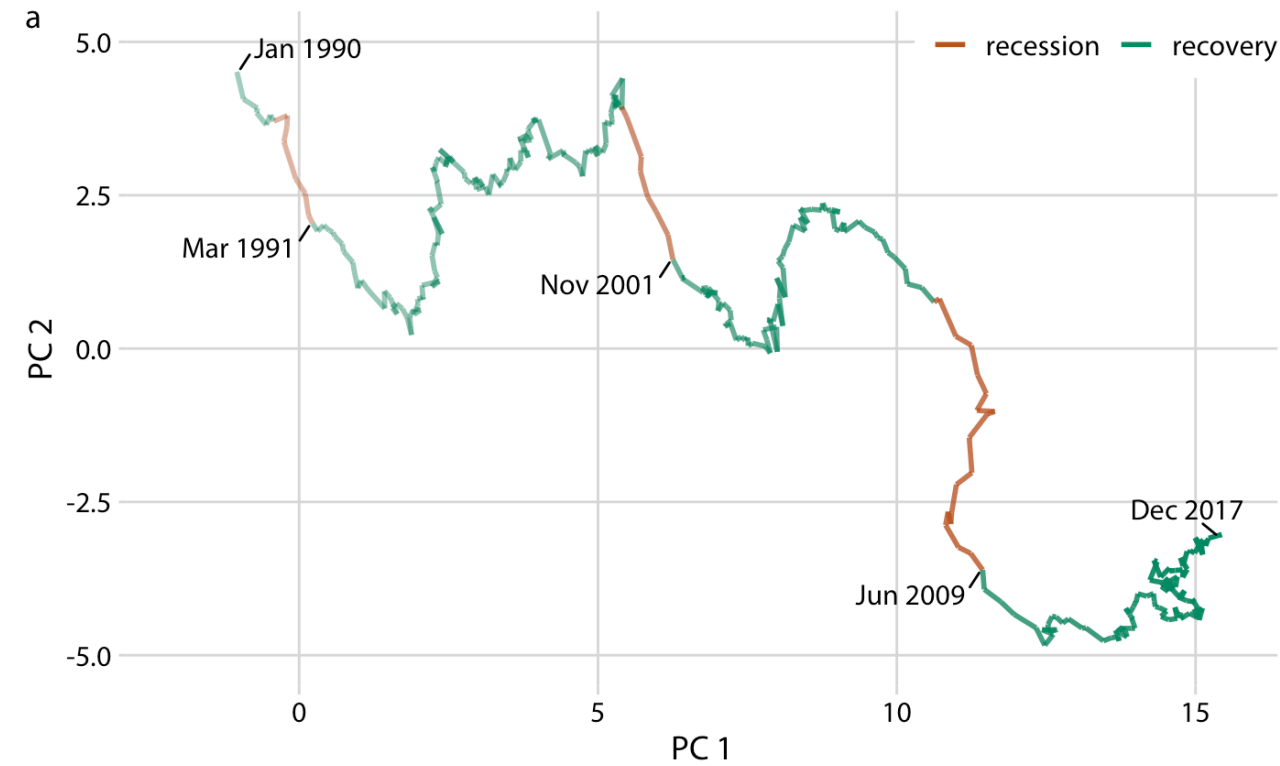


Without date markers and color shading, the data's direction and rate of change are unclear.

High-Dimensional Time Series Visualization

- When datasets have multiple variables, dimension reduction techniques like Principal Component Analysis (PCA) can help.
- Example: Macroeconomic indicators visualized using PCA and connected scatter plots.
- Color coding and labeled end points enhance interpretability.

Visualizing high-dimensional time series



The path shows the movement of 100+ macroeconomic indicators (1990–2017), with recessions and recoveries color-coded. End points of three recessions are labeled.

Challenges

- Data granularity and missing values.
- Overlapping lines in multivariate time series.
- If not properly annotated, direction and speed of changes can be lost.

Advantages

- Helps identify cyclical relationships that may be harder to spot in line graphs.
- Identifying trends, patterns, and seasonality.
- Making informed decisions based on historical data.

Guidelines for Effective Time Series Visualization

- Avoid overplotting by reducing data density or using transparency.
- Line graphs remain the dominant method for visualizing time series data, but modifications such as direct labeling, area fills, and connected scatter plots improve clarity.
- Choosing the right visualization method depends on the dataset's structure, the number of variables, and the story one wants to convey.