Patterns and Trends Visualization

Introduction to Patterns and Trends Visualization

- Definition of trends in data visualization
- Importance of trend visualization
- Overview of approaches: smoothing and functional form fitting

Understanding Smoothing in Trend Visualization

- Purpose of smoothing: reducing noise while preserving key patterns
- Common smoothing methods: moving averages, LOESS, and splines
- Applications in financial data and time series

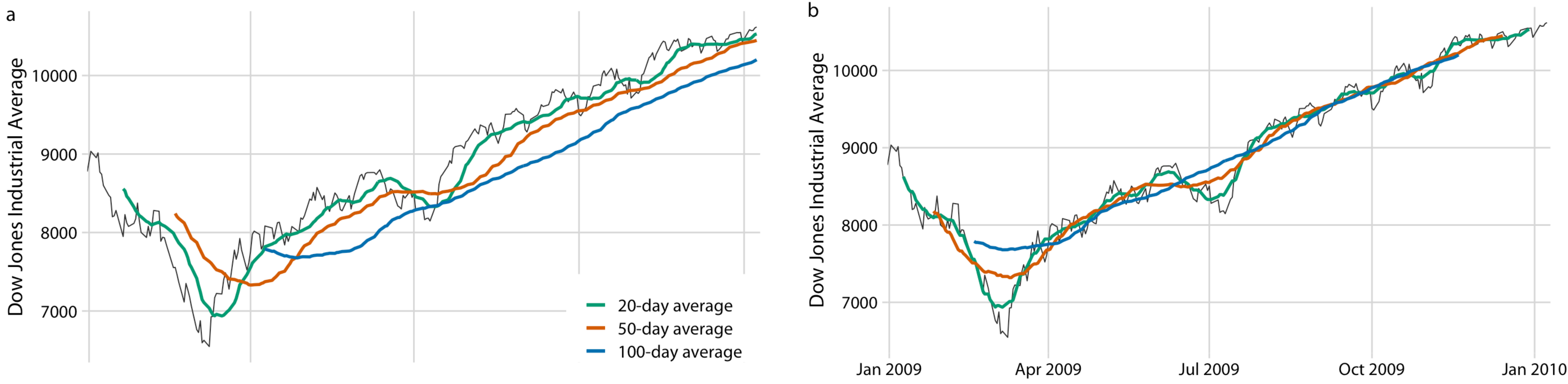
Financial data as time-series

Daily closing values of the Dow Jones Industrial Average for the year 2009. Data source: Yahoo! Finance

Moving Averages for Trend Detection

- Explanation of moving averages
- 20-day, 50-day, and 100-day moving averages
- Moving averages plotted at different positions in the time window

Moving Average



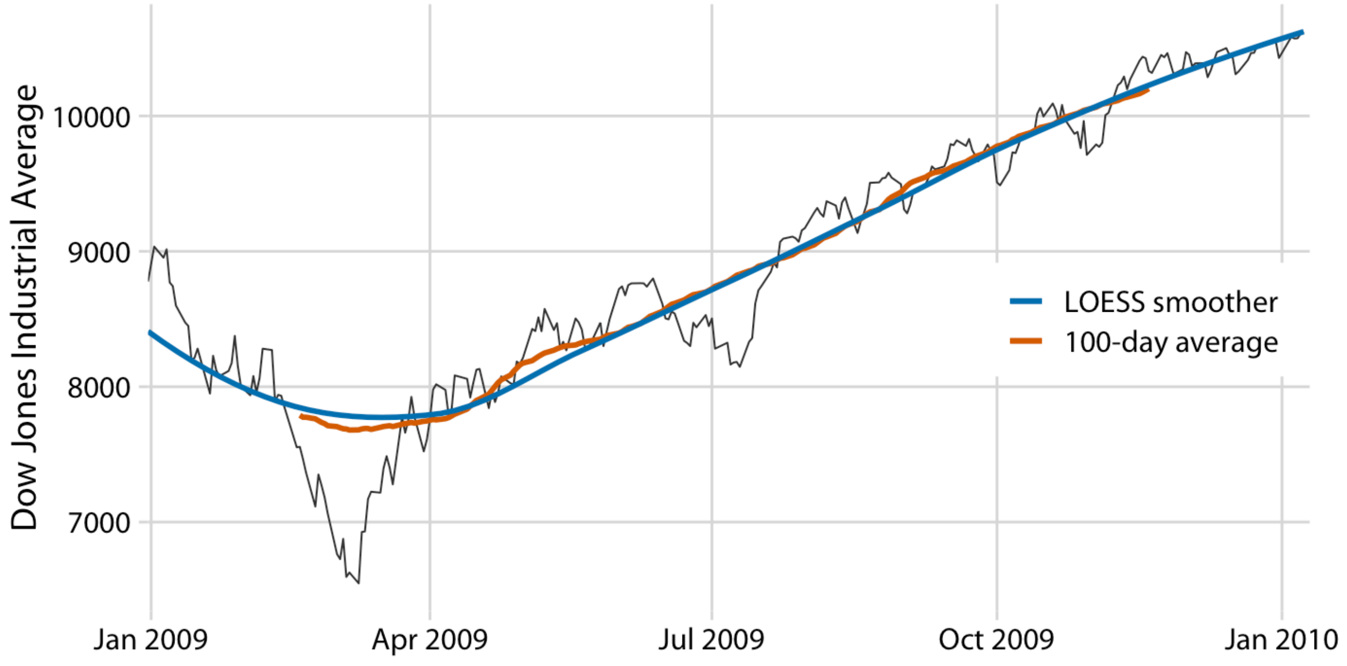
Daily closing values of the Dow Jones Industrial Average in 2009 with 20-day, 50-day, and 100-day moving averages. (a) Moving averages plotted at the end of time windows. (b) Moving averages plotted at the center of time windows.

Limitations of Moving Averages

- Shortened time series due to missing data at edges
- Lagging effect in moving averages
- Wiggles and instability in highly volatile datasets

LOESS (Locally Estimated Scatterplot Smoothing)

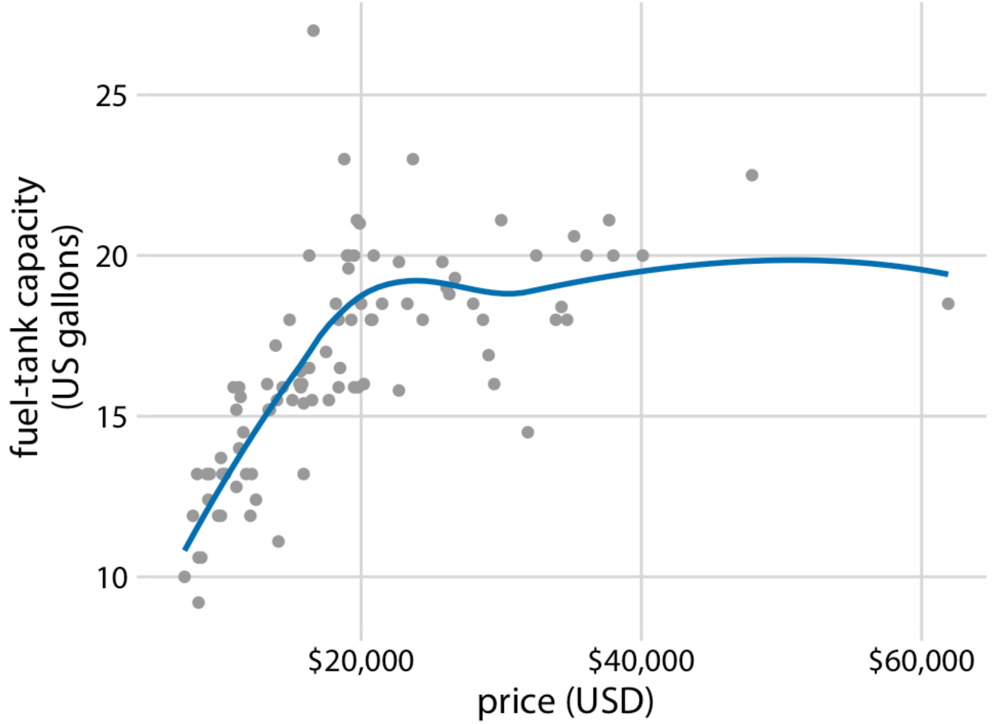
- Definition and advantages of LOESS
- Application in stock market and price trend analysis
- Comparison of LOESS vs. moving averages



Comparison of LOESS fit to the 100-day moving average for Dow Jones data. LOESS shows a nearly identical trend but is smoother and extends across the entire data range.

Application of LOESS in Scatter Plots

- Example: Fuel-tank capacity vs. car price
- LOESS curve interpretation
- Benefits of using LOESS over simple averaging techniques



Fuel-tank capacity vs. price for 93 cars from the 1993 model year. Each dot represents a car, with a LOESS smooth (solid line) showing a linear increase up to ~$20,000, then leveling off.

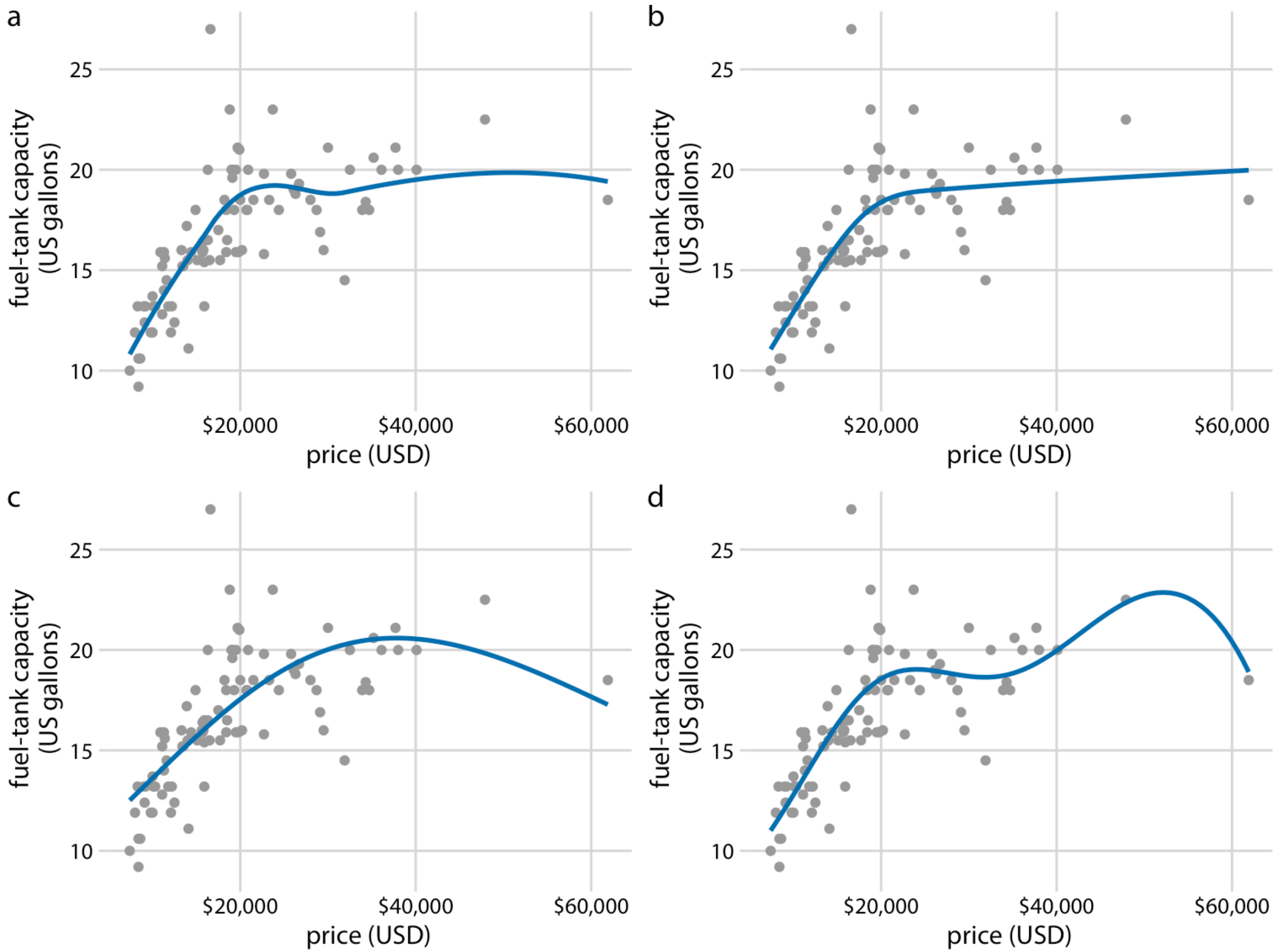
Spline Models for Smoothing

- Definition of splines and their flexibility
- Types of splines: cubic, B-splines, thin-plate splines, Gaussian process splines
- Influence of knots on smoothing effectiveness

Comparison of Different Smoothing Models

- Visualization of different smoothing methods applied to the same data
- Strengths and weaknesses of LOESS, cubic splines, and Gaussian process splines

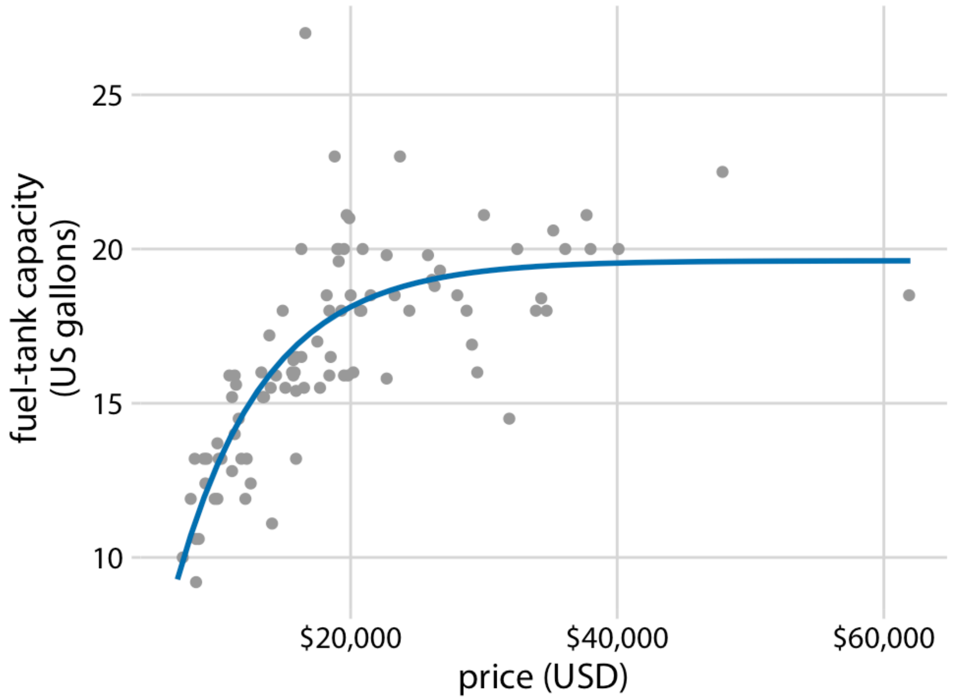
Comparing different smoothing models



Different smoothing models display widely different behaviors, in particular near the boundaries of the data. (a) LOESS. (b) Cubic regression splines with 5 knots. (c) Thin-plate regression spline with 3 knots. (d) Gaussian process spline with 6 knots.

Functional Form-Based Trend Fitting

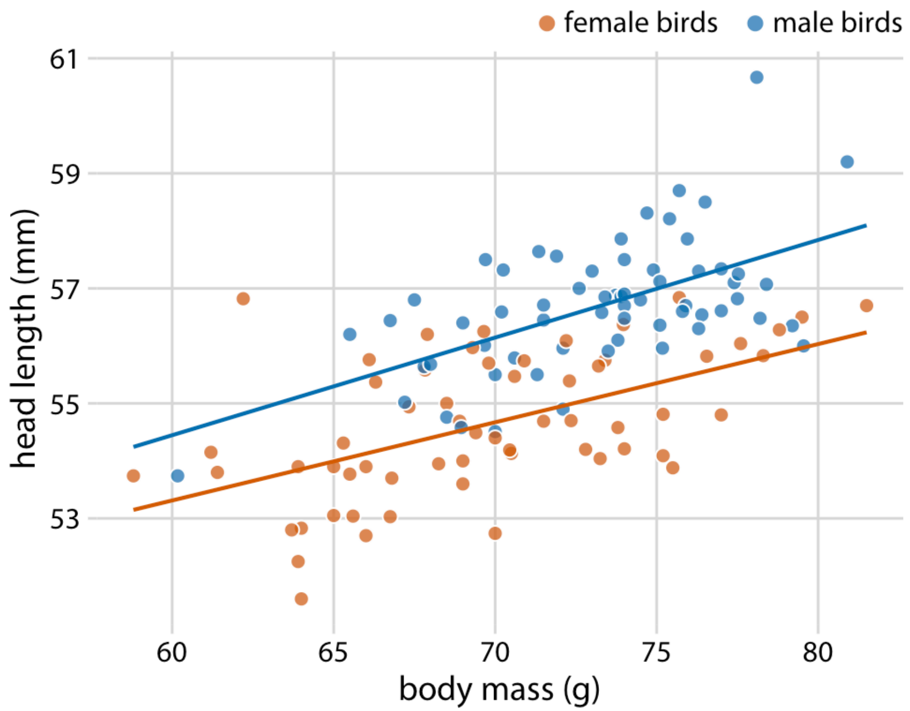
- Advantages over general-purpose smoothers
- Explicit analytical models for trend representation
- Example function: y = A - B exp(-mx)



Fuel-tank data represented with an explicit analytical model. The solid line corresponds to a least-squares fit of the formula y=A−B exp(−mx) to the data. Fitted parameters are A=19.6, B=29.2, m=0.00015.

Linear Trends in Data

- Example: Blue jays' head length vs. body mass
- Importance of adding linear trend lines to scatter plots
- Interpretation of slope and intercept values



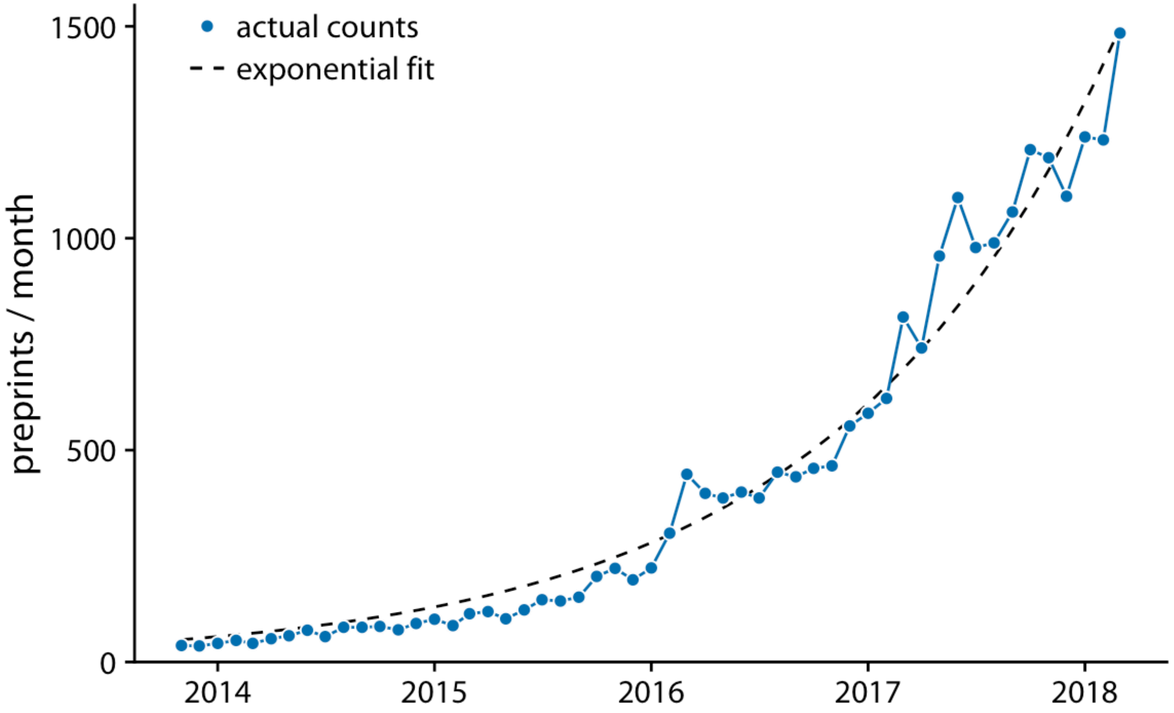
Head length versus body mass for 123 blue jays. The birds' sex is indicated by color. This has linear trend lines on top of the individual data points.

Non-Linear Trends and Transformations

- When to use nonlinear models
- Log transformations for linearizing exponential relationships
- Example: bioRxiv monthly submission trends

Exponential and Log-Linear Models

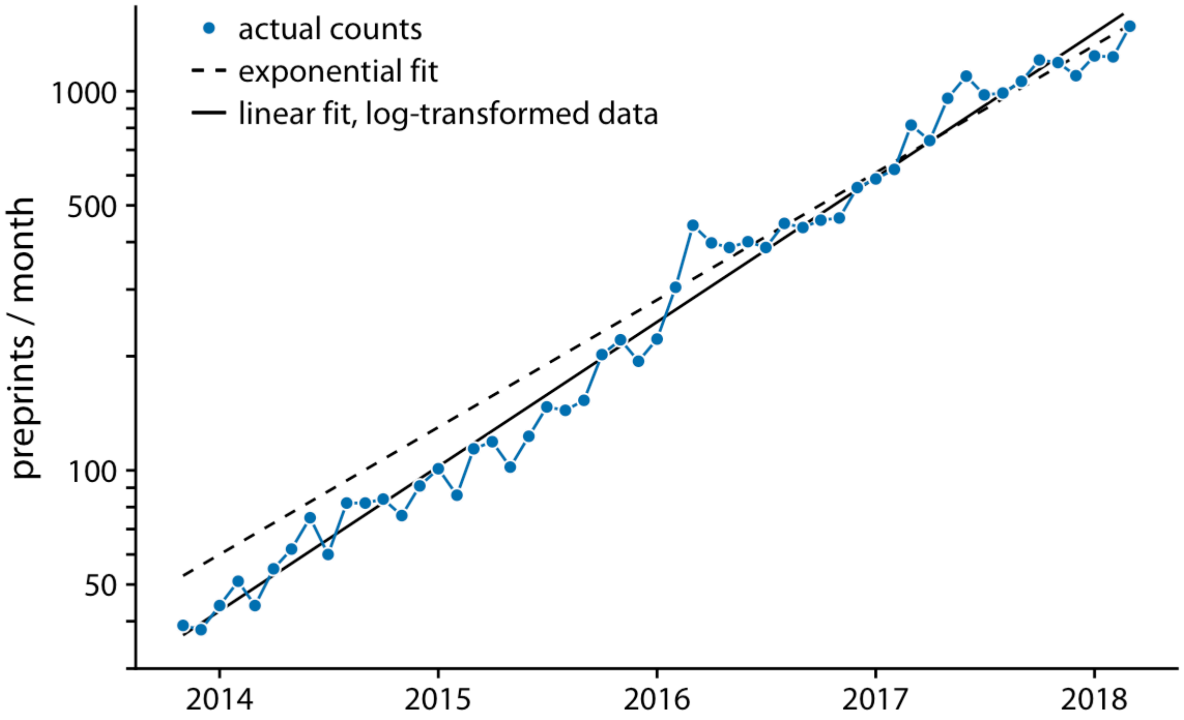
- Fitting exponential curves to data
- Limitations of exponential fits
- Log-linear transformation for improved trend detection



Monthly submissions to the preprint server bioRxiv. The solid blue line represents the actual monthly preprint counts and the dashed black line represents an exponential fit to the data, $y=60 \exp[0.77(x-2014)]$.

Log-Log and Linear-Log Plots

- Use cases of different transformations
- Identifying power laws and exponential growth
- Real-world applications in economic and scientific data

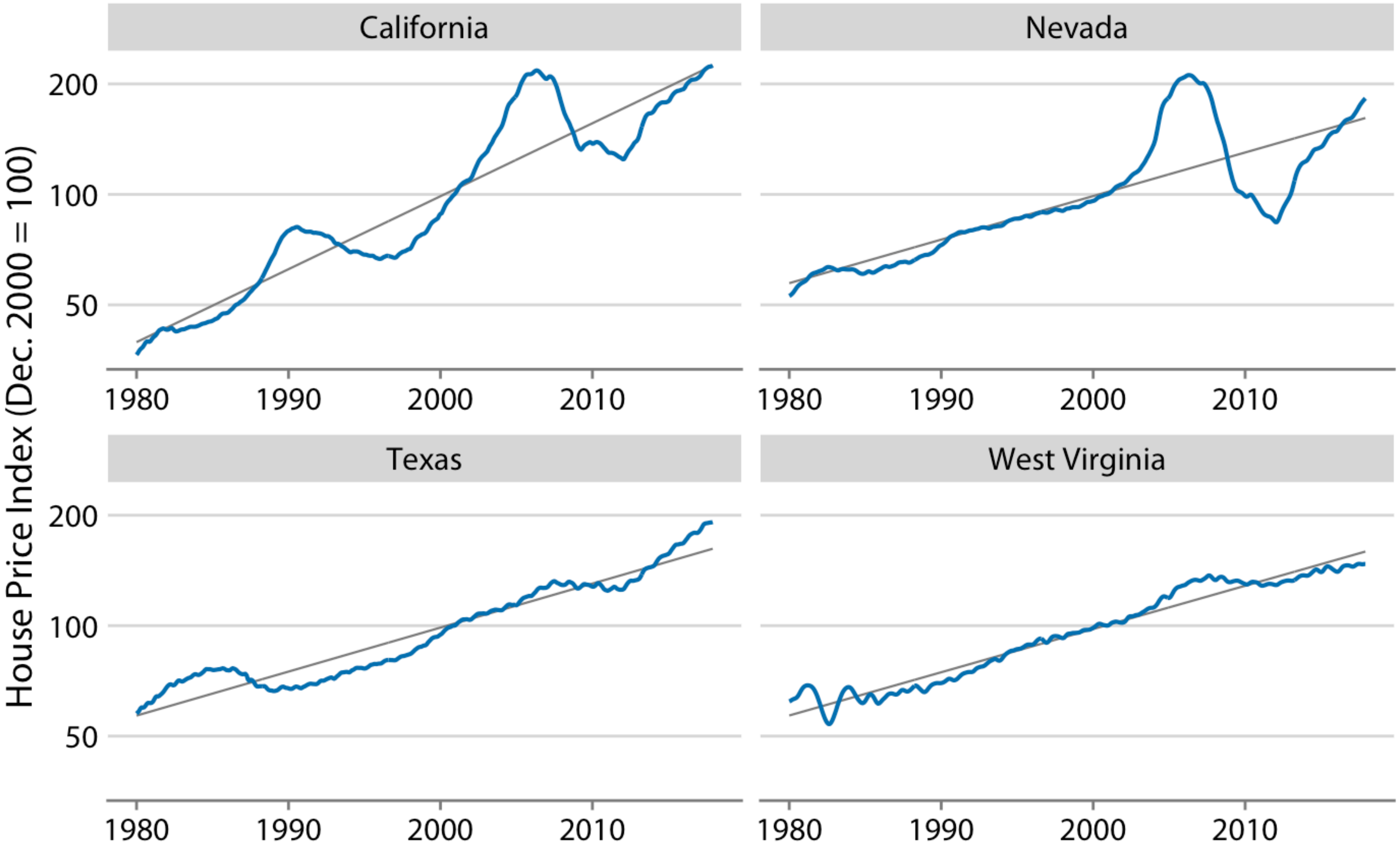


Monthly submissions to the preprint server bioRxiv, shown on a log scale. The solid blue line represents the actual monthly preprint counts, the dashed black line represents the exponential fit, and the solid black line represents a linear fit to log-transformed data, corresponding to $y=43 \exp[0.88(x-2014)]$.

Detrending and Time-Series Decomposition

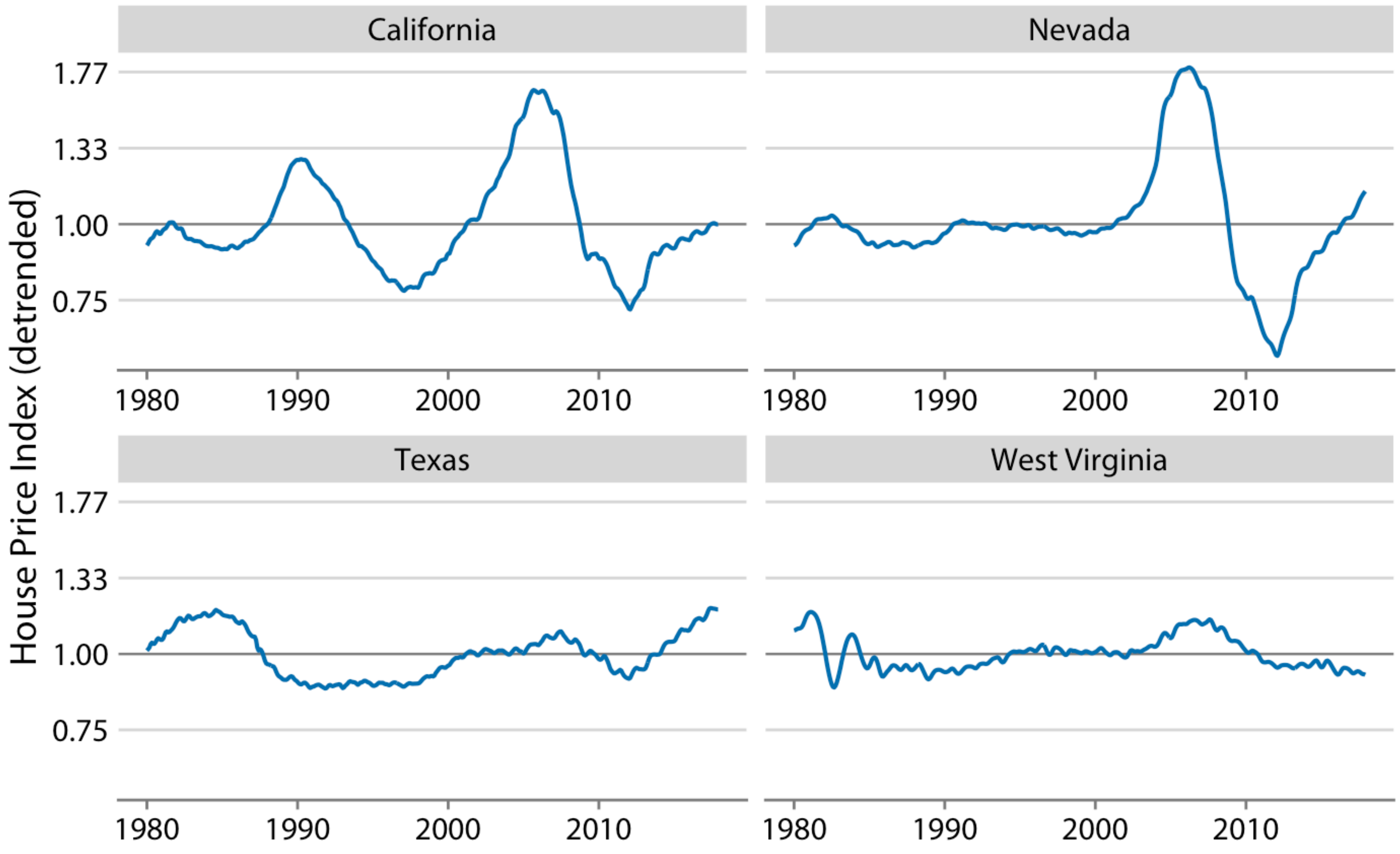
- **Detrending**: Removes long-term trends to highlight notable deviations
- Demonstration using the **Freddie Mac House Price Index** for U.S. states (California, Nevada, Texas, West Virginia)
- Overlaid housing bubbles show boom and bust cycles in California and Nevada
- **Logarithmic y-axis** used to emphasize exponential growth patterns

House price index with long-term trends and price fluctuations



Freddie Mac House Price Index (1980–2017) for California, Nevada, Texas, and West Virginia. The unitless index tracks relative house prices, scaled to 100 in December 2000. Blue lines show monthly fluctuations, while gray lines indicate long-term trends. Logarithmic y-axes highlight exponential growth.

Detrended house prices emphasizing housing bubbles

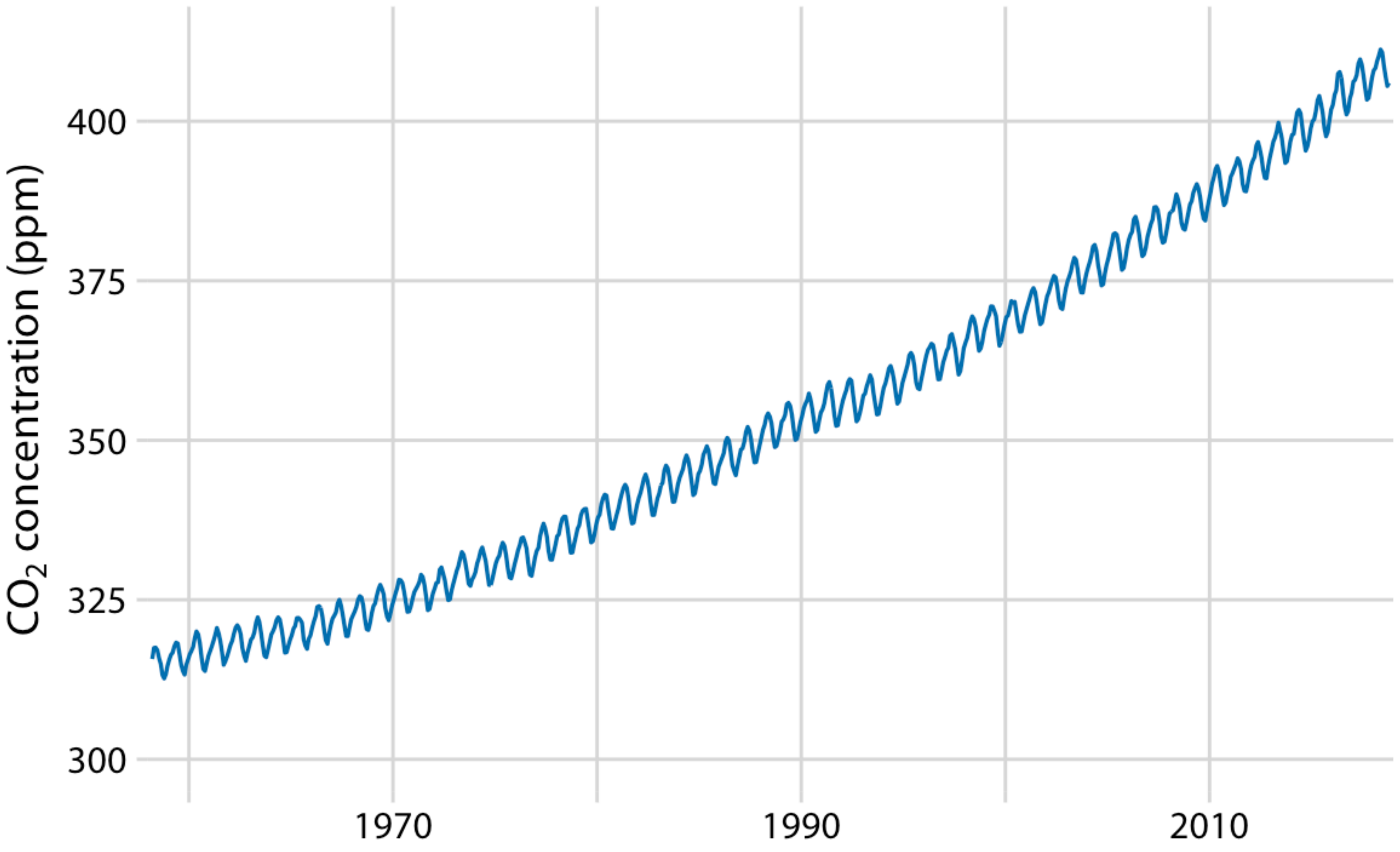


Detrended Freddie Mac House Price Index, calculated by dividing actual values by the long-term trend. California saw housing bubbles around 1990 and the mid-2000s, while Nevada had one in the mid-2000s. Texas and West Virginia showed little deviation.

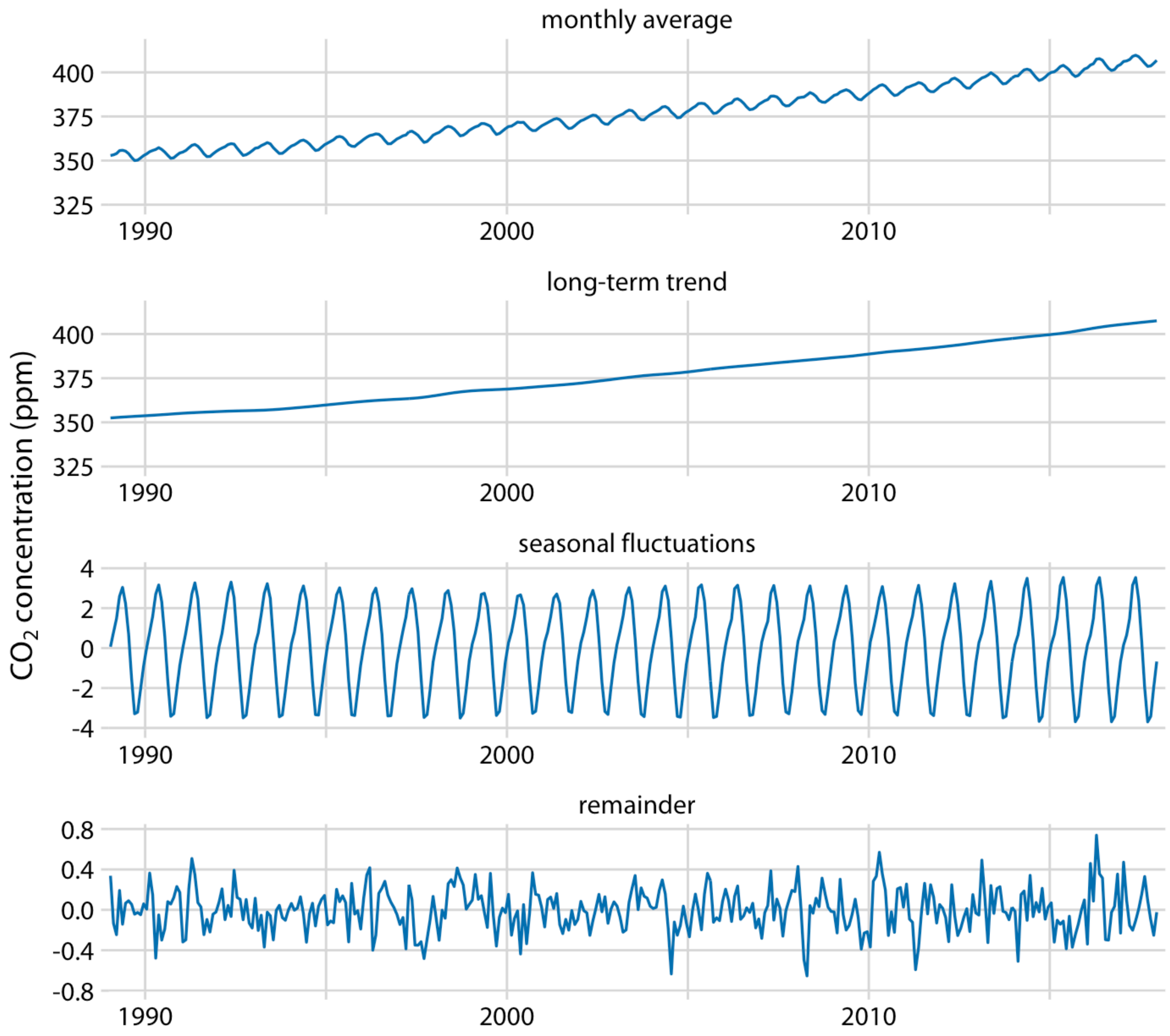
Time-Series Decomposition

- Decompose data into **long-term trend**, **seasonal fluctuations**, and **remainder** (random noise, external events)
- **Keeling Curve** as an example: CO2 changes over time

The Keeling curve showing long-term CO2 increase with seasonal fluctuations



Decomposition showing trend, seasonal fluctuations, and noise



Time-series decomposition of the Keeling curve, showing the monthly average, long-term trend, seasonal fluctuations, and remainder (random noise). The last 30 years are highlighted to emphasize annual patterns.

Conclusion and Best Practices

- Key takeaways from trend visualization techniques
- Avoiding misinterpretation of trends
- Importance of exploratory data visualization before applying trends