

# Introduction to Data Visualization

Visualizing Time Series and  
Other Functions of an  
Independent Variable  
&  
Trends

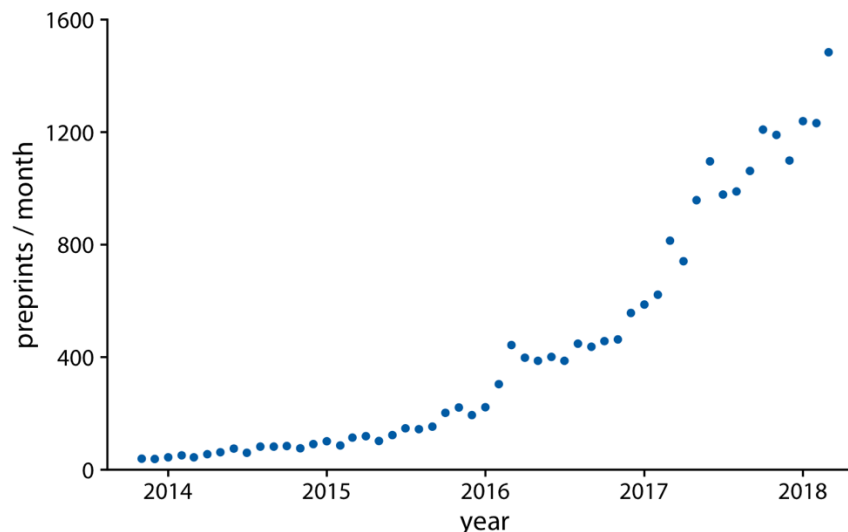
**Halil Bisgin, Ph.D.**

## Time Series & Independent Variables

- What if one of the variables is time that imposes order?
- We frequently want to visualize this temporal order, and we do so with line graphs.
- We can have other variables such as dose that impose order.

## Individual Time Series

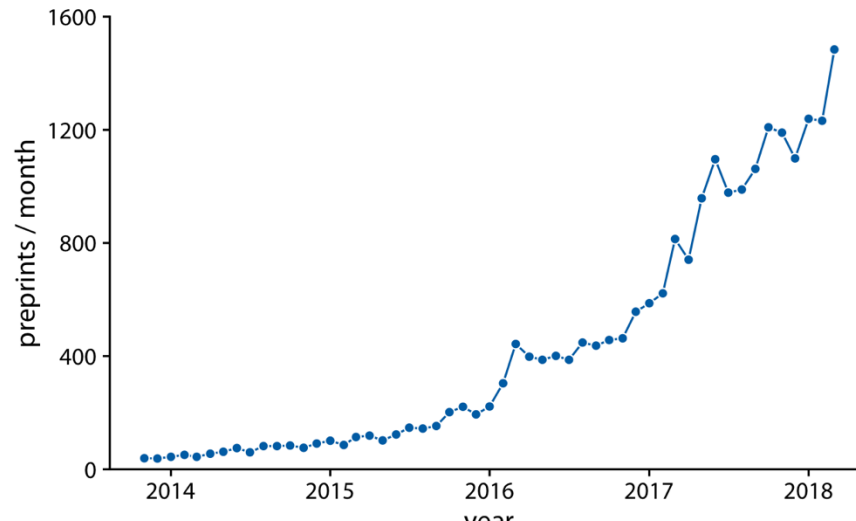
- Let's consider the pattern of monthly preprint submissions in biology.
  - *Preprints are scientific articles that researchers post online before formal peer review and publication in a scientific journal*
- Visualize this growth by making a form of scatterplot



- Not a regular scatterplot
- Evenly placed x values

## Individual Time Series-Line Graph

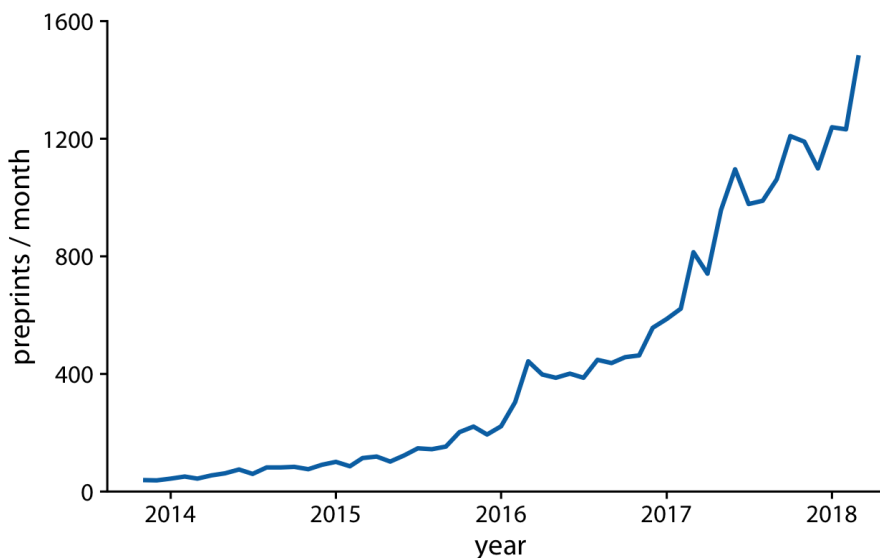
- We can visually emphasize the order by connecting neighboring points with lines.
- Such a plot is called a line graph.
  - lines do not represent observed data
  - +may help with perception when the points are spaced far apart



lines are meant as a guide to the eye

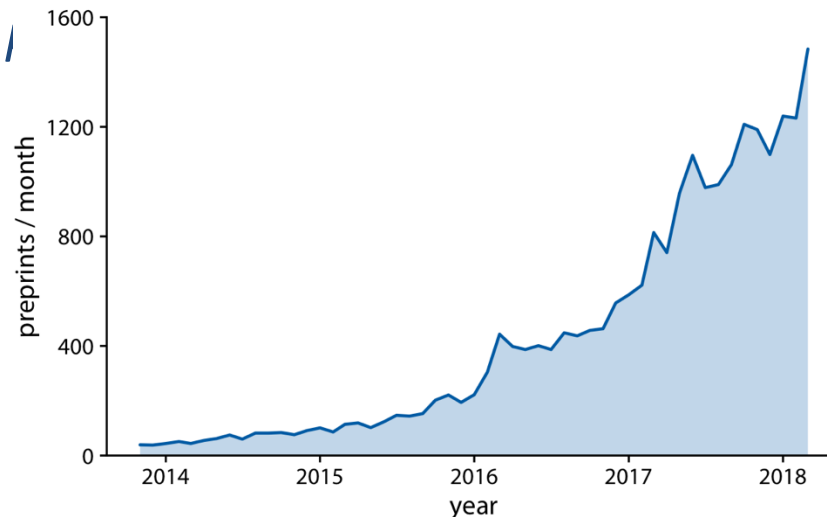
# Individual Time Series-Line Graph

- You can use lines to represent time and omit dots.
  - *Places more emphasis on the overall trend in the data and less on individual observations.*
  - *A figure without dots is also visually less busy*
  - *The denser the time series, the less important it is to show individual observations with dots*



- `seaborn.lineplot`
- `geom_line`

- We can also fill the area under the curve with a solid color.
  - *Further emphasizes the overarching trend in the data, because it visually separates the area above the curve from the area below.*
  - *Only valid if the y axis starts at zero, so that the height of the shaded area at each time point represents the data value at that time*

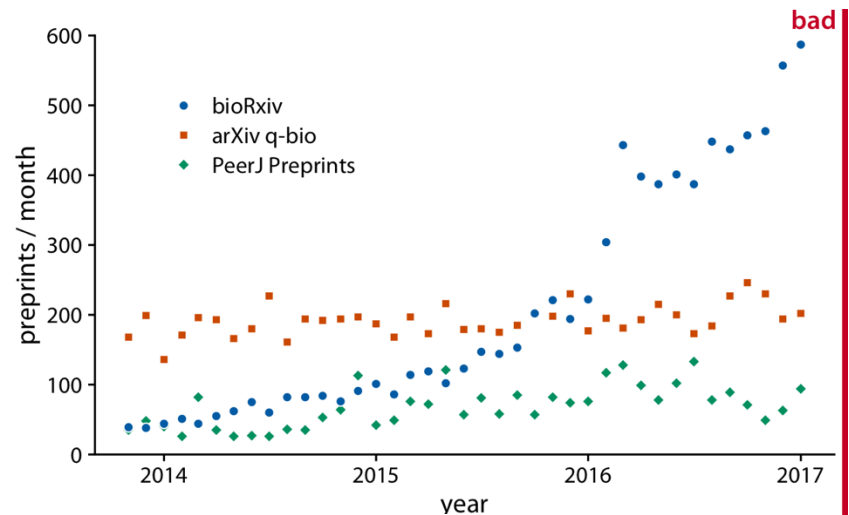


# Multiple Time Series and Dose–Response Curves

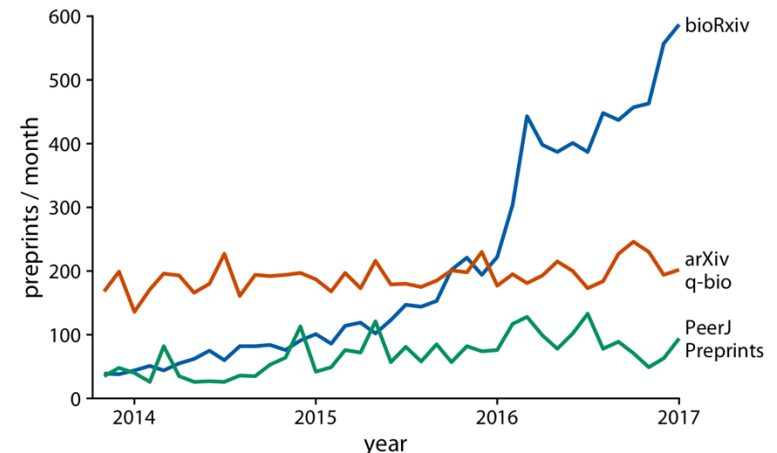
- Multiple time courses that we want to show at once



## Scatter Plot

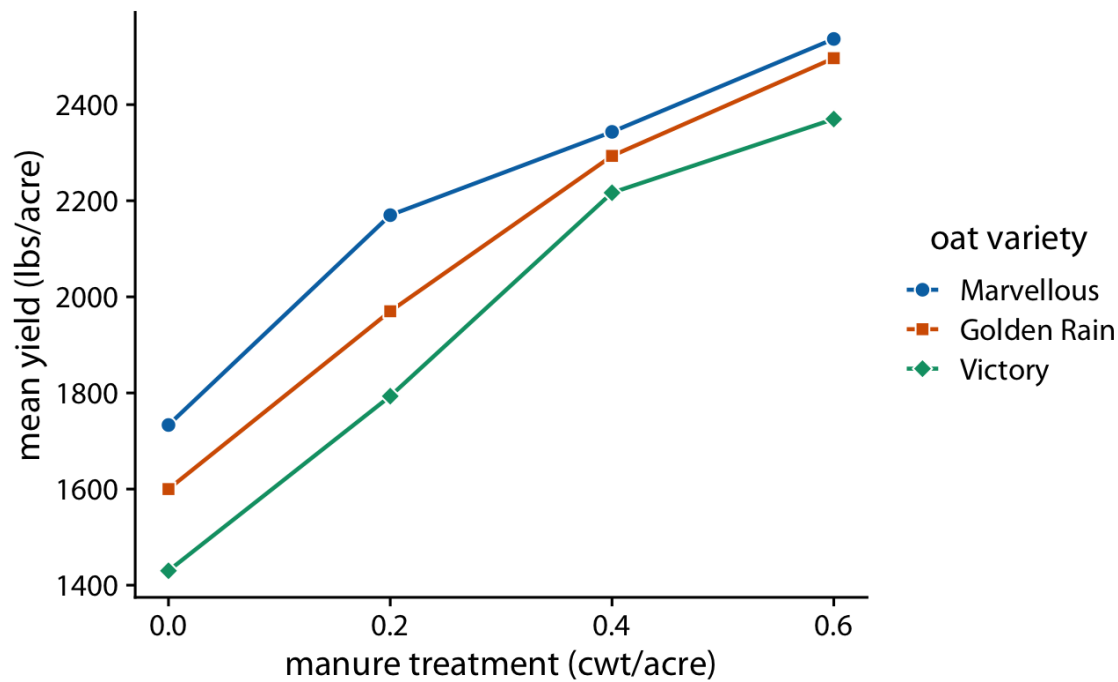


## Line Plot



## Dose Response Curves

- Line plots are not limited to time.
- Whenever the data points have a natural order

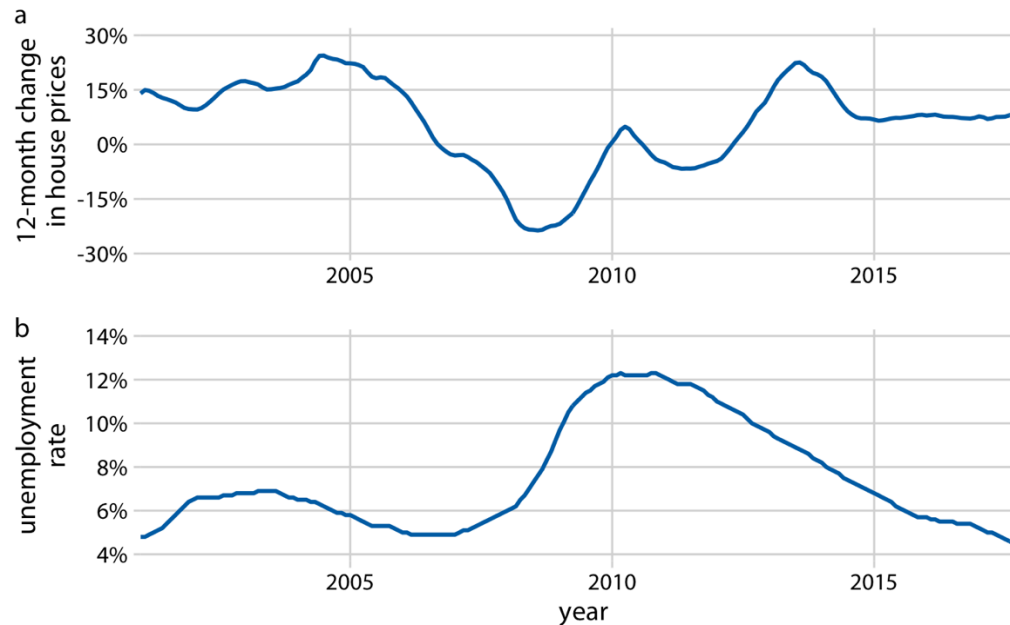


mean yield of oat varieties after fertilization with manure



# Time Series of Two or More Response Variables

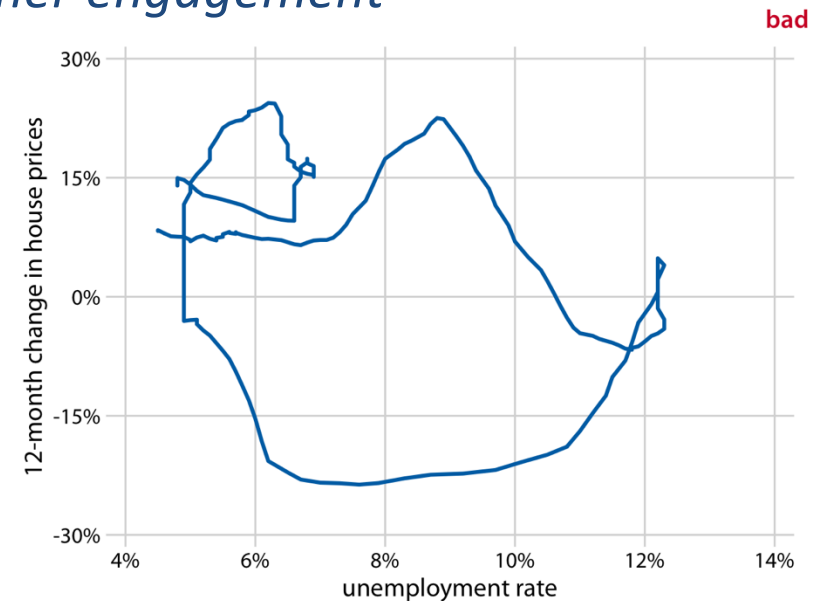
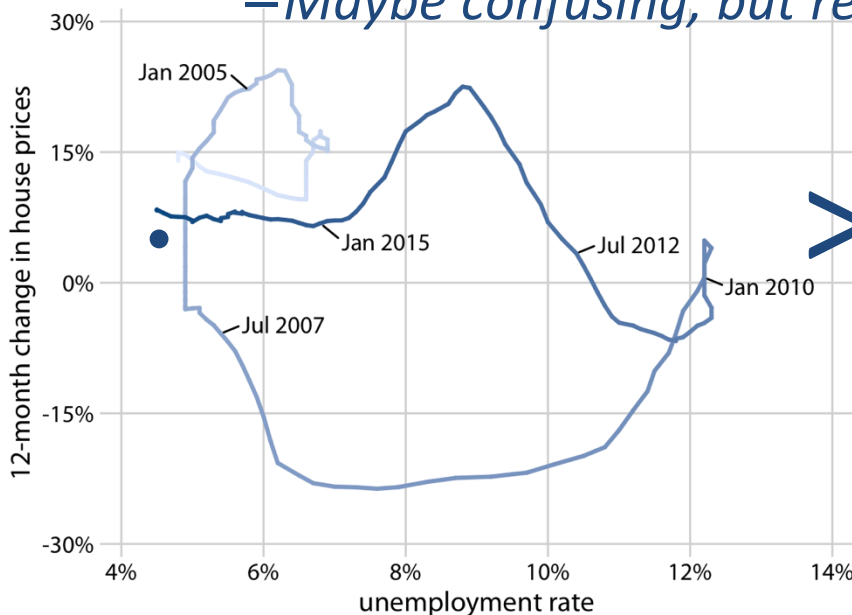
- Not usual to have more than one response variable.
  - *Change in house prices from the previous 12 months as it relates to the unemployment rate.*



Can we incorporate temporal change better?

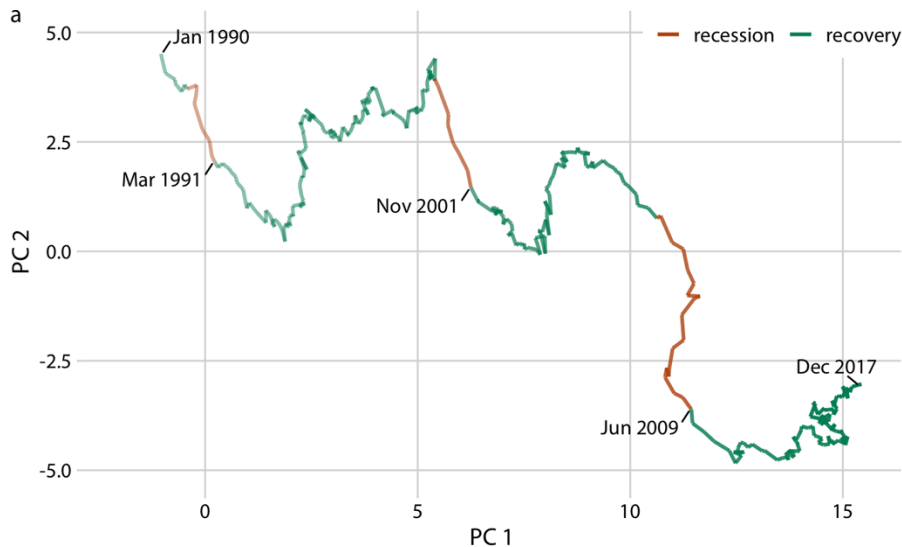
## Connected scatterplot

- Draw a path that leads from the earliest time point to the latest (phase portrait).
  - *Darker shades represent more recent months.*
  - *The anticorrelation between the change in house prices and the unemployment rate causes to form two counterclockwise circles*
  - *Maybe confusing, but result in higher engagement*

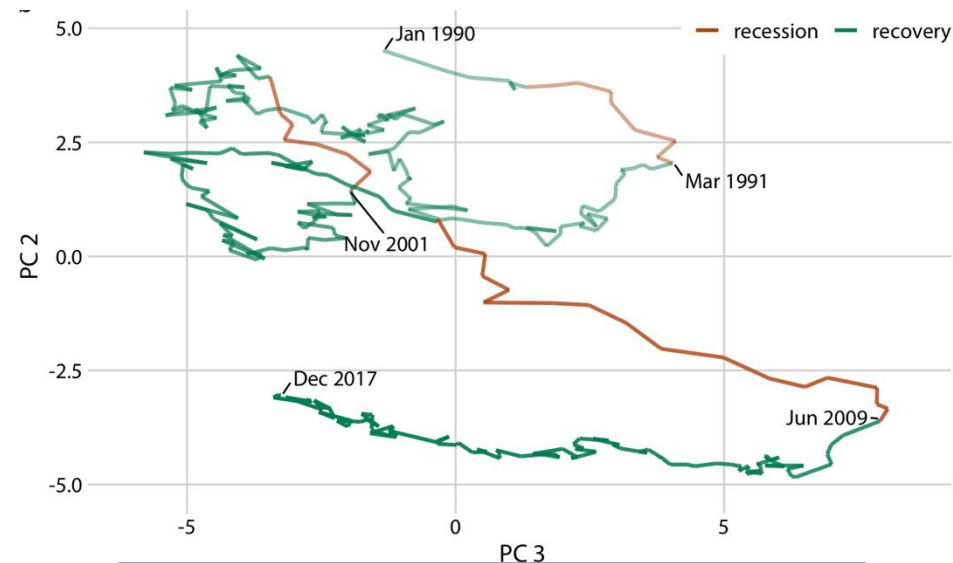


## Connected Scatterplot w/ Higher Dimensions

- The trick is to apply dimension reduction first
- We can then draw a connected scatterplot in the dimension-reduced space using PCs.
  - E.g., Monthly observations of over 100 macroeconomic indicators, provided by the Federal Reserve Bank of St. Louis.



Recession associated w/ PC2



Recoveries associated w/PC3

## Visualizing Trends

- We are often interested in the overarching trend.
- Drawing the trend on top of or instead of the actual data points, usually in the form of a straight or curved line helps.

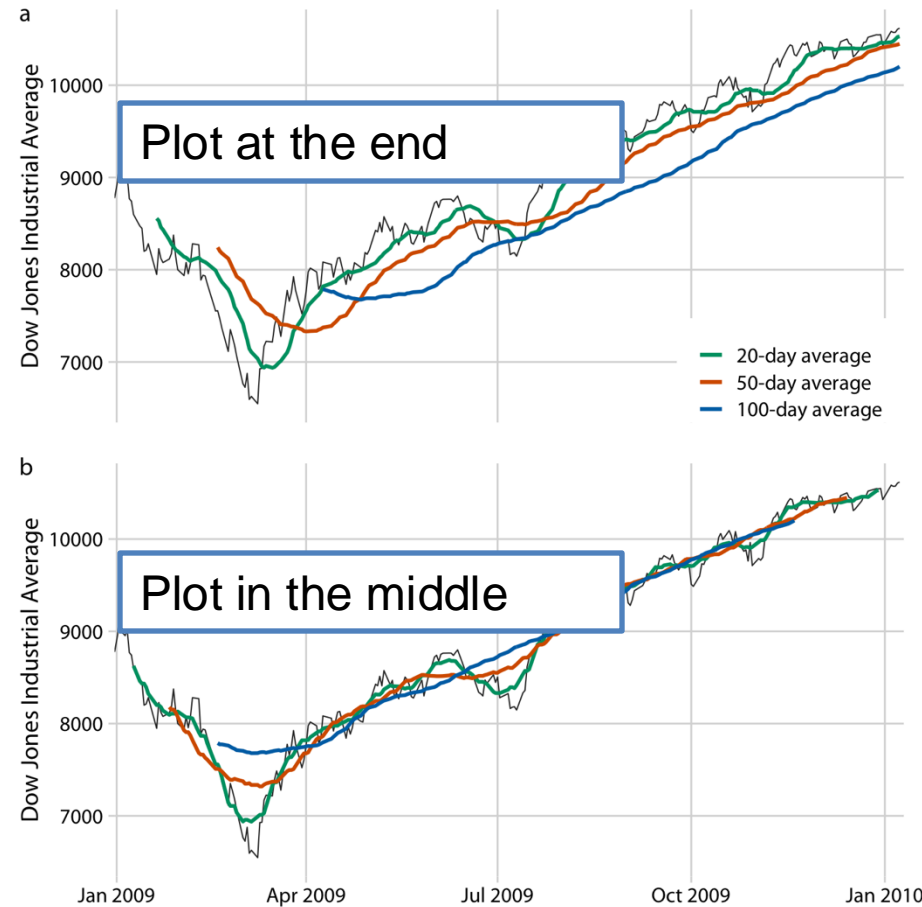
## Visualizing Trends-Smoothing

- How can we visualize these longer-term trends while deemphasizing the less important short-term fluctuations?
- We can smooth by using methods such as *moving averages*.



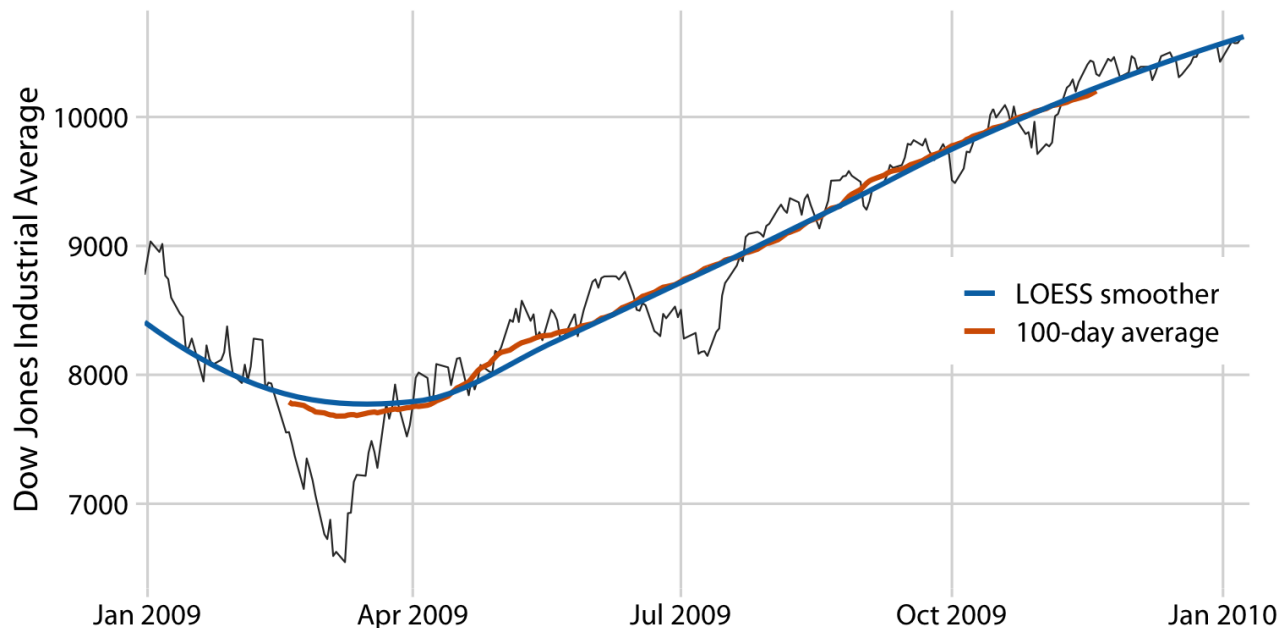
# Visualizing Trends-Smoothing

- Take a time window, say the first 20 days in the time series
- Calculate the average price over these 20 days
- Move the time window by one day, so it now spans the 2nd to 21st days.



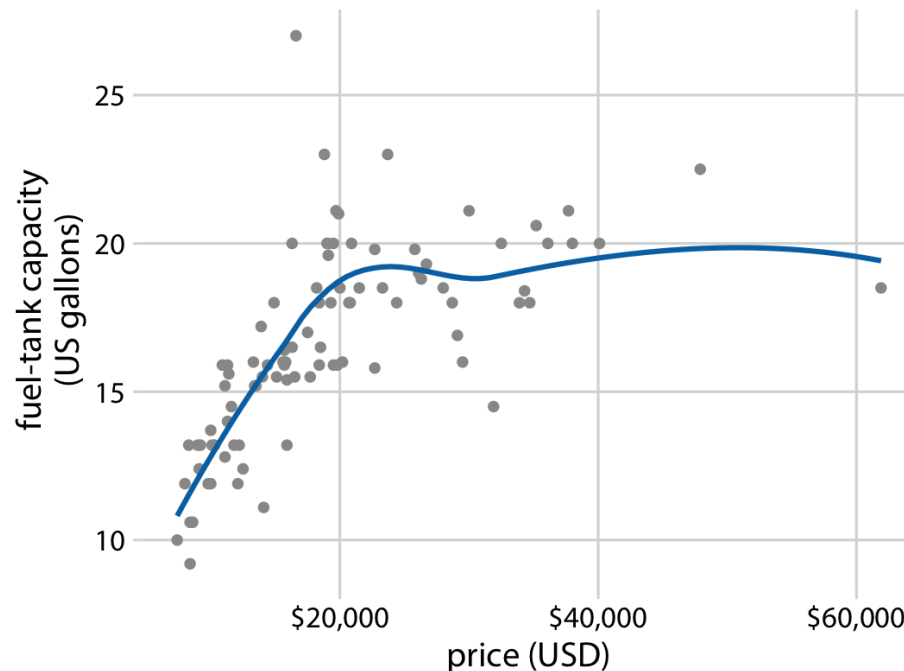
## Visualizing Trends-LOESS

- Smoothed curves are shorter and have missing parts.
- Smoothed curves are not necessarily smooth.
- Locally estimated scatterplot smoothing (LOESS) fits low-degree polynomials to subsets of the data.



## Visualizing Trends-LOESS

- LOESS is not limited to time series. It can be applied to arbitrary scatterplots, as is apparent from its name, locally estimated scatterplot smoothing.



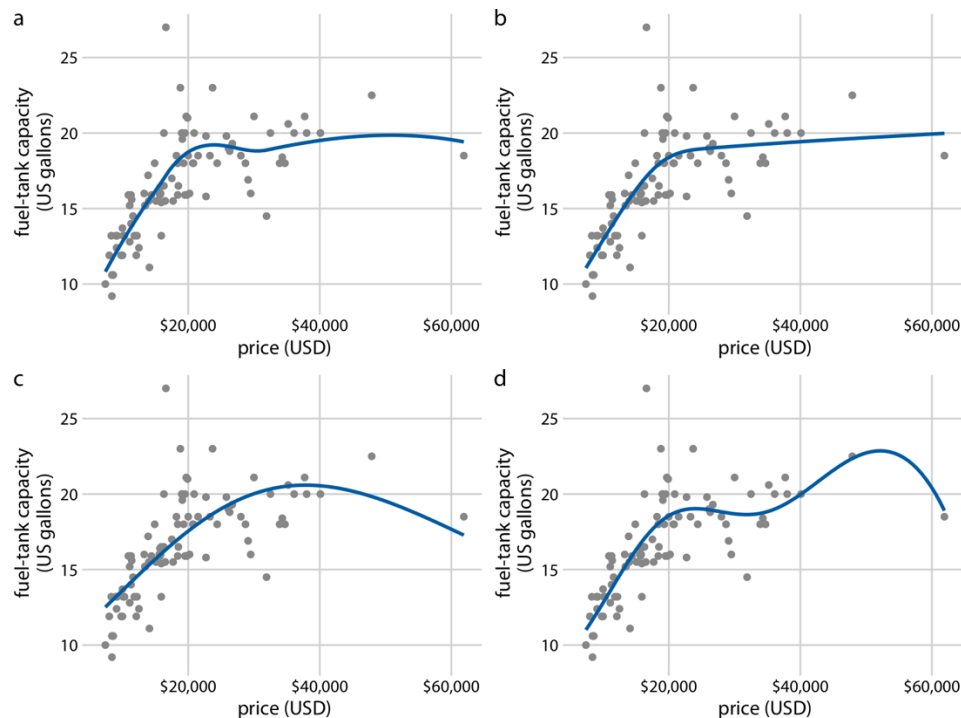


## Visualizing Trends-Spline Models

- LOESS requires the fitting of many separate regression models. This makes it slow for large datasets.
- A spline is a piecewise polynomial function that is highly flexible yet always looks smooth.
- Knots are used to create small spline segments.
- Tools offer splines as a smoothing parameter.

## Visualizing Trends-Warning!!!

- Be careful when interpreting the results from a smoothing function. The same dataset can be smoothed in many different ways.

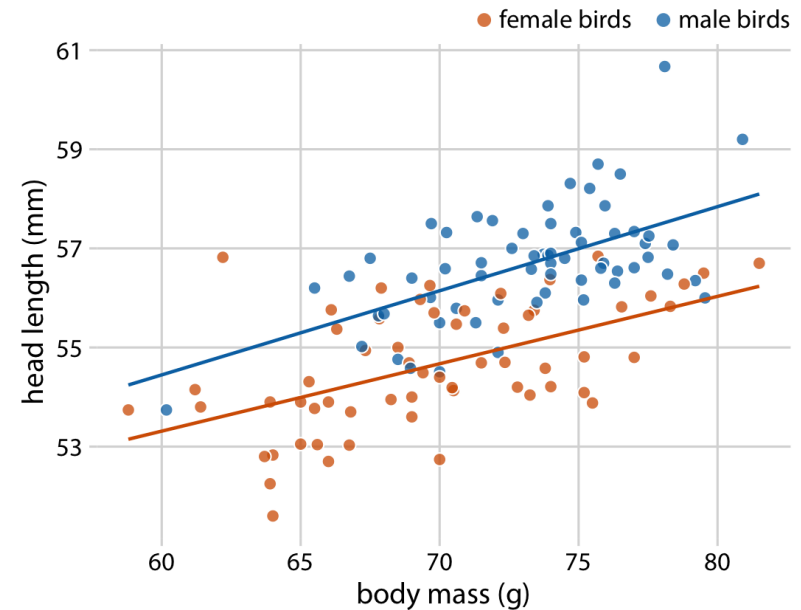
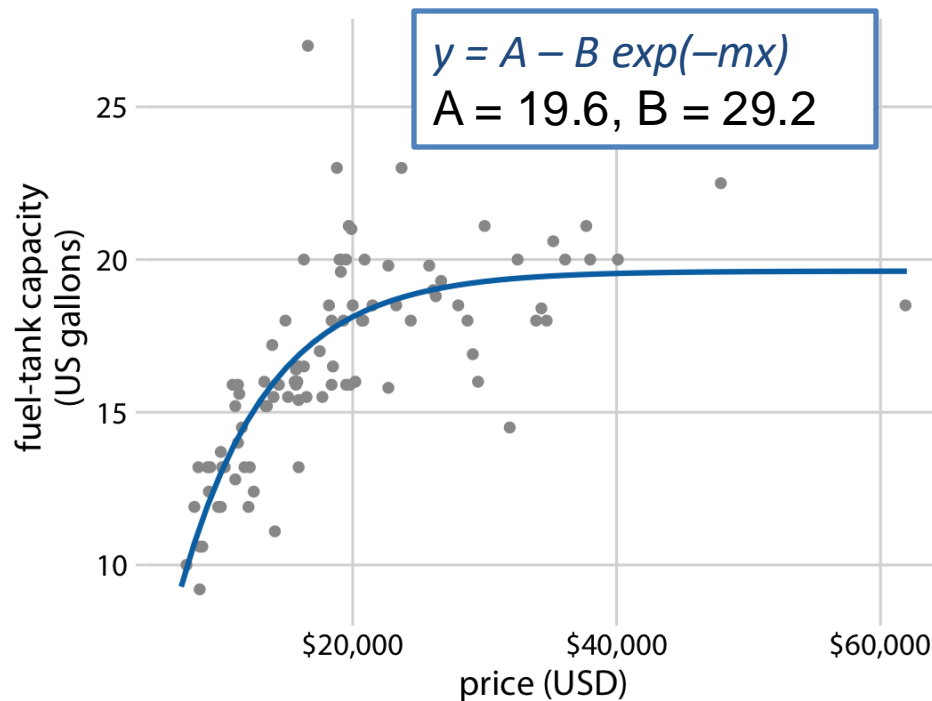


# Showing Trends with a Defined Functional Form

- The behavior of general-purpose smoothers can be somewhat unpredictable.
- They do not provide parameter estimates that have a meaningful interpretation.
- We can try to fit a curve based on generic functions.
  - *Fuel price – tank capacity:  $y = A - B \exp(-mx)$*
  - *Blue jay head - mass:  $y = A + mx$ .*

# Showing Trends with a Defined Functional Form

- Examples

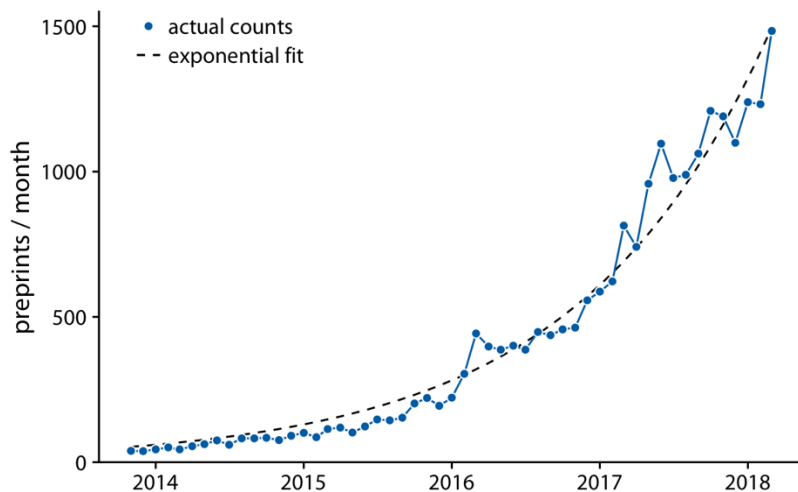


# Showing Trends with a Defined Functional Form

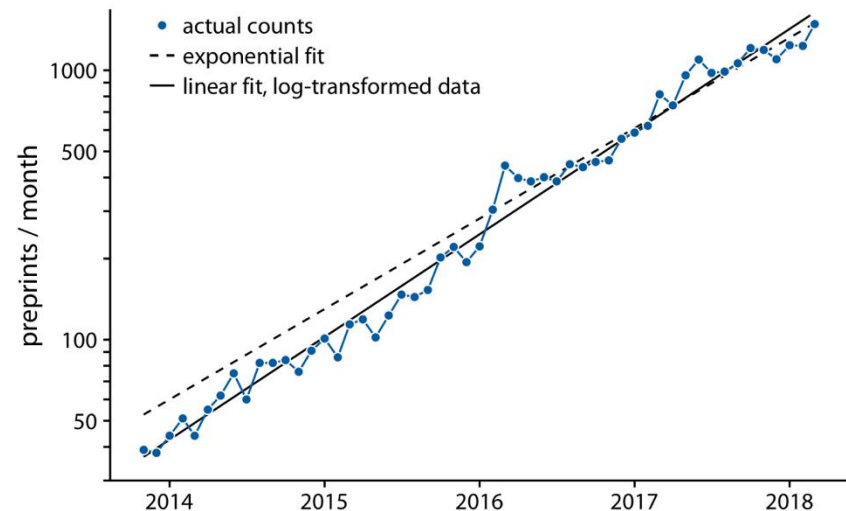
- When you have data in the form of exponential function, you can try log and fit a linear line.  

$$-y = A \exp(mx) \rightarrow \log(y) = \log(A) + mx$$
- Avoid exponential fits and instead use linear fits on log-transformed data.

$$y = 60 \exp[0.77(x - 2014)]$$



$$y = 43 \exp[0.88(x - 2014)]$$



# Detrending and Time-Series Decomposition

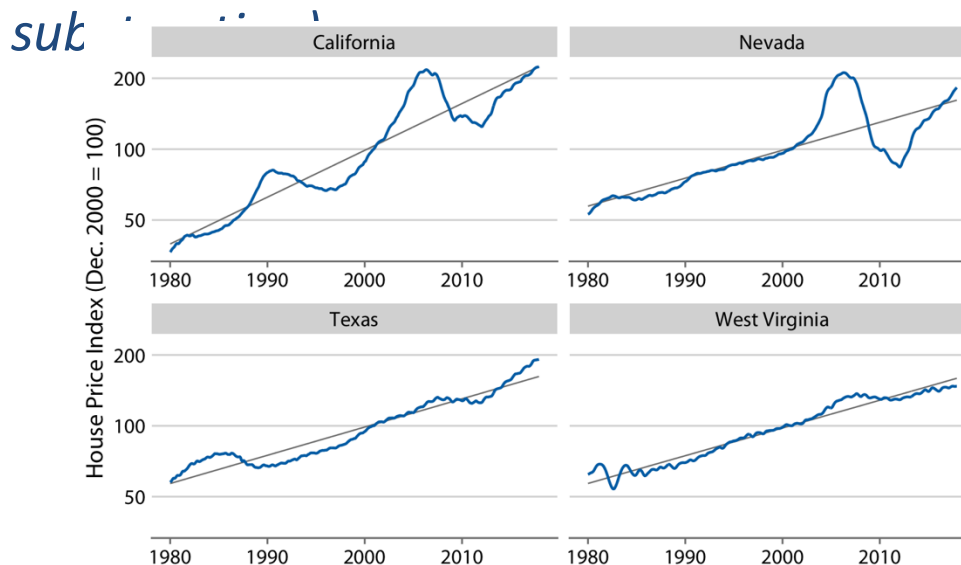
- It may be useful to specifically highlight any notable deviations.

– *Housing prices: Bubbles*

– *Logarithmic y axis*

– *Grey lines are trends*

– *Divide values by the points on the lines (log scale it's*



# Detrending and Time-Series Decomposition

- Shows the bubbles now:
  - *California experienced two housing bubbles, around 1990 and in the mid-2000s*

