PCA and Time Series Analysis

Lesson 8

Lesson Overview

- Principal Components
  - Principal Components Analysis
  - Principal Components Regression
- Intro to Time Series Modeling
- Finish Time Series Next Class:
  - Forecasting
  - Time Series Regression
  - Time Series Smoothing and Transforms
  - Backfilling and resampling
  - Rolling Windows
  - Naïve Bayes (Machine Learning!)

Principal Component Analysis

Principal Component Analysis (PCA)

Dimension reduction algorithm

- Convert high to low dimension
- Selects most important features that cause highest variance in output
- Transforms number of possibly correlated variables into a smaller number of uncorrelated variables called principal components

- Used to speed up regression algorithms
  - Helpful for data visualization
  - Noise filtering
  - Feature extraction

PCA

Example: You want to predict the weather in Seattle for June of 2023.

- You have information from the last 200 years
- You have information on climate change
- You have information on volcanos, weather patterns, rainfall, humidity, population growth, traffic
- You have too much information
- Where do we start?

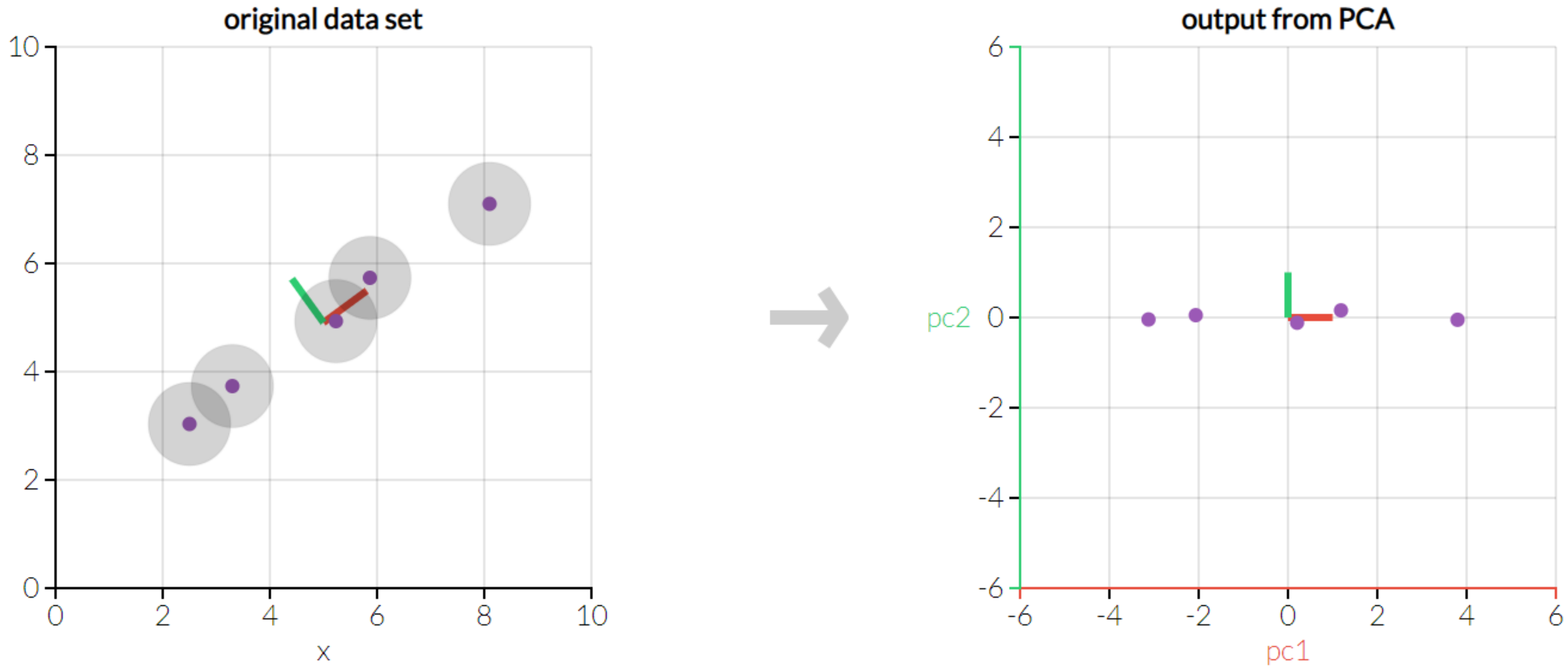
PCA

Dimensionality Reduction

- Reducing the dimension of the feature space
- Feature elimination
  - Simplify model by dropping features that don't add much meaning
  - Could lose valuable information
- Feature extraction
  - Combine features into new variables so we can drop the less meaningful variables while still retaining their info
  - New combined features are independent of each other

Reducing Dimensions

- We choose the matrix directions that have the most meaning (red below) and drop the rest, allowing us to go from many dimensions, to just a few:



How does PCA work? Singular Value Decomposition

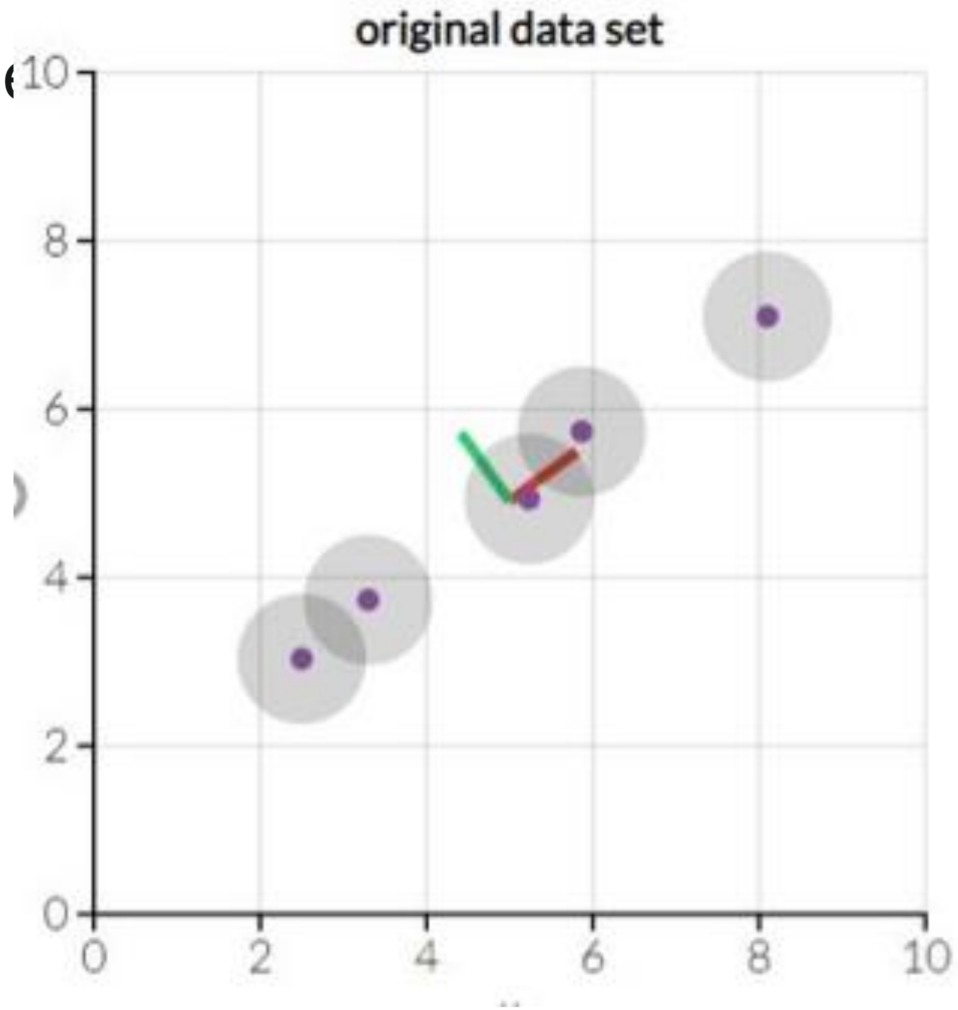
1. Ensure all measurements are on the same scale, or use a correlation matrix

2. **Covariance Matrix:** Calculate a square, symmetric matrix that summarizes the relationship between the variables

3. Determine direction (**eigenvectors**) and magnitude (**eigenvalues**) of the points

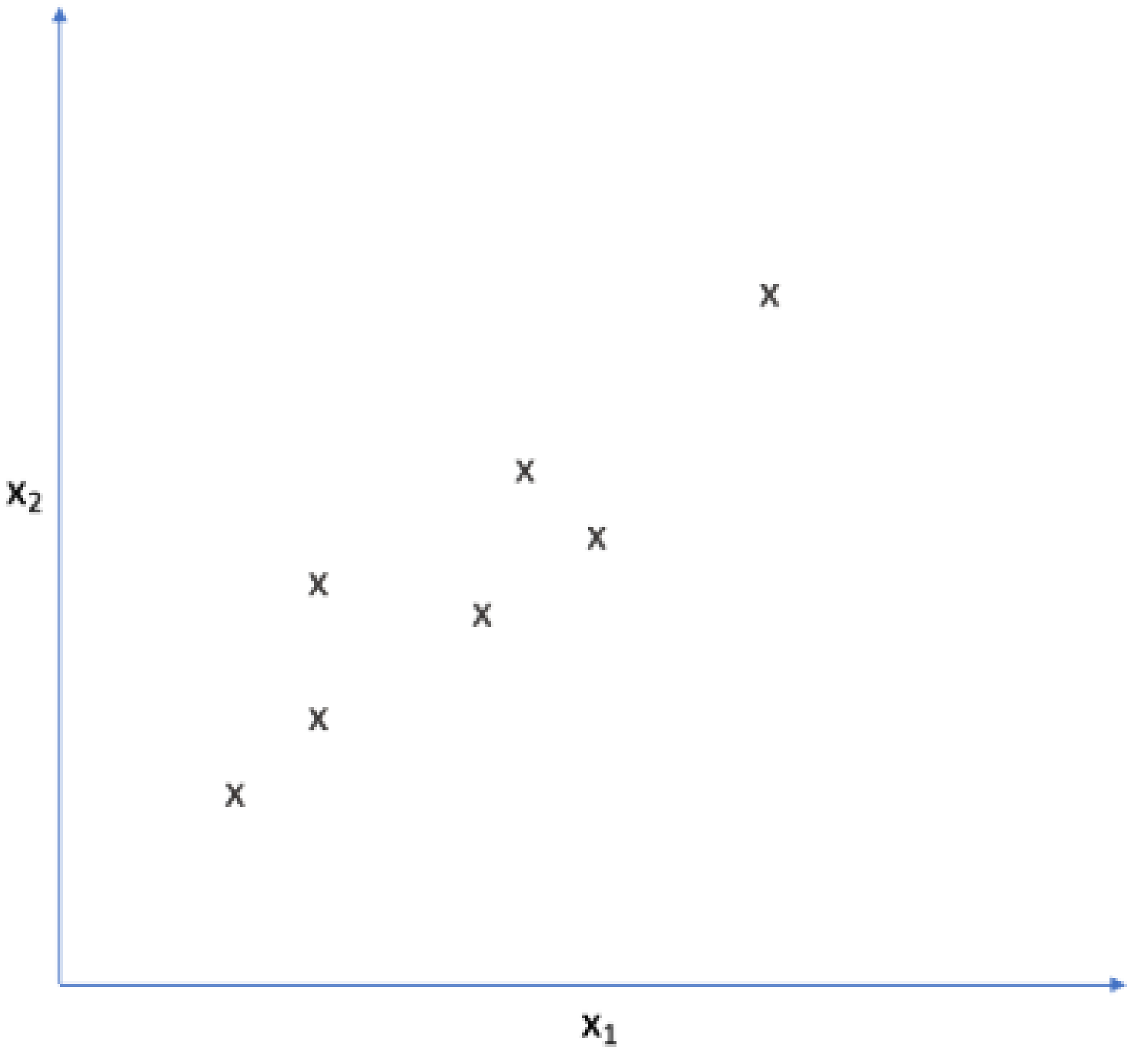
4. Transform data to align with most important direction found in data

- Largest eigenvalues = more important



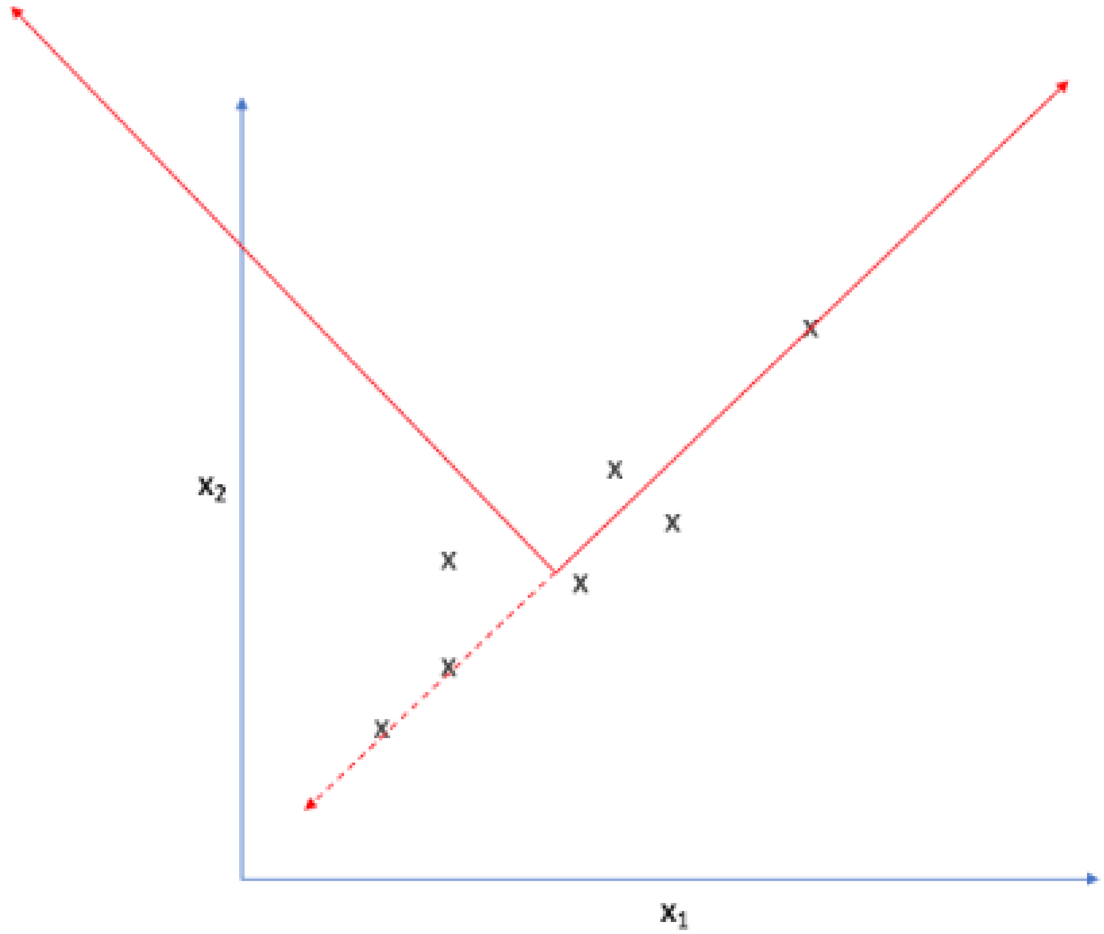
PCA: How it works

- Points on this plot appear to have a roughly linear relationship
  - Requirement for linear models
  - Implies some correlatoin between points



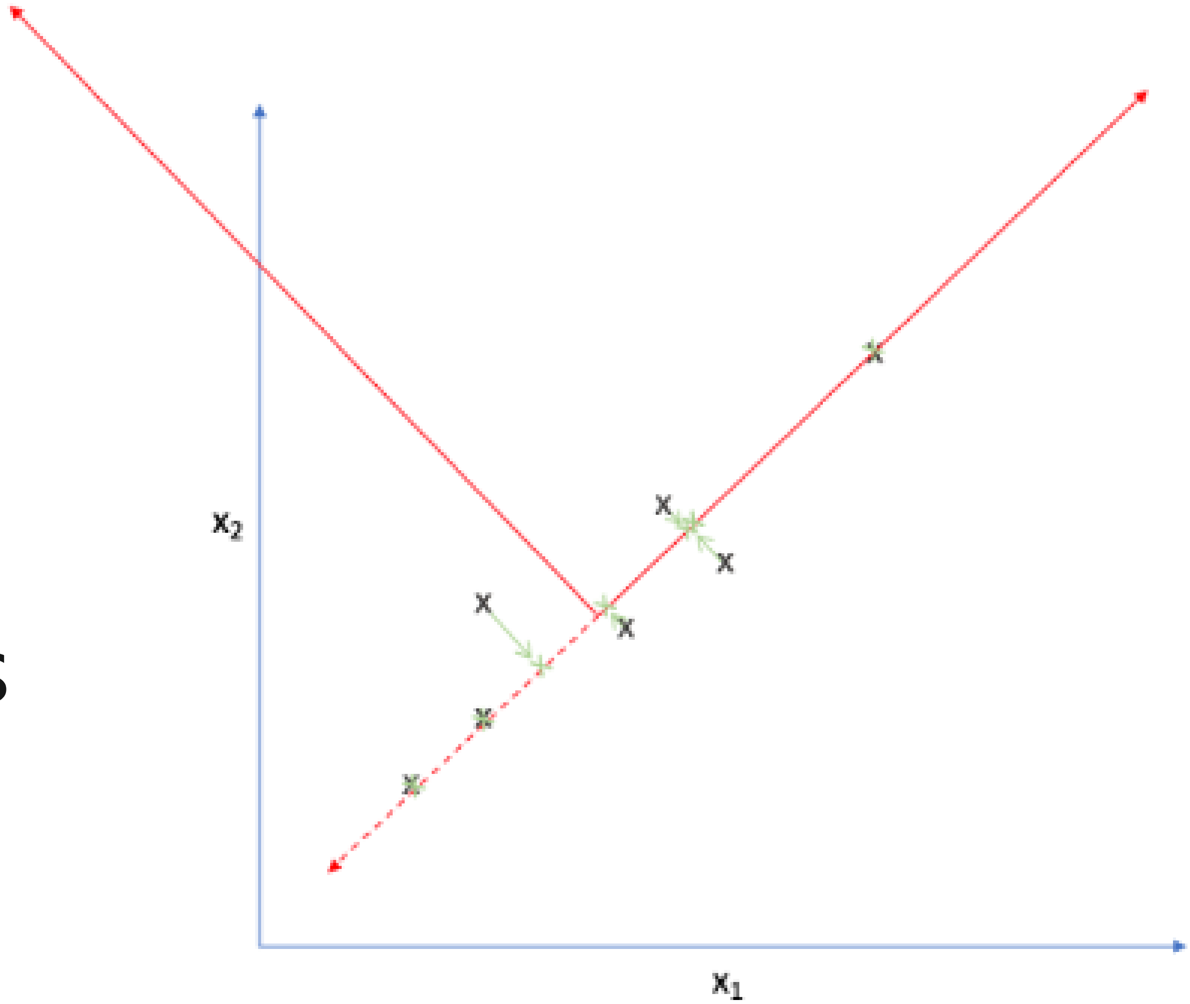
PCA: how it works

- We can reorient the axis to center on the data points
- Direction is perpendicular to original axis



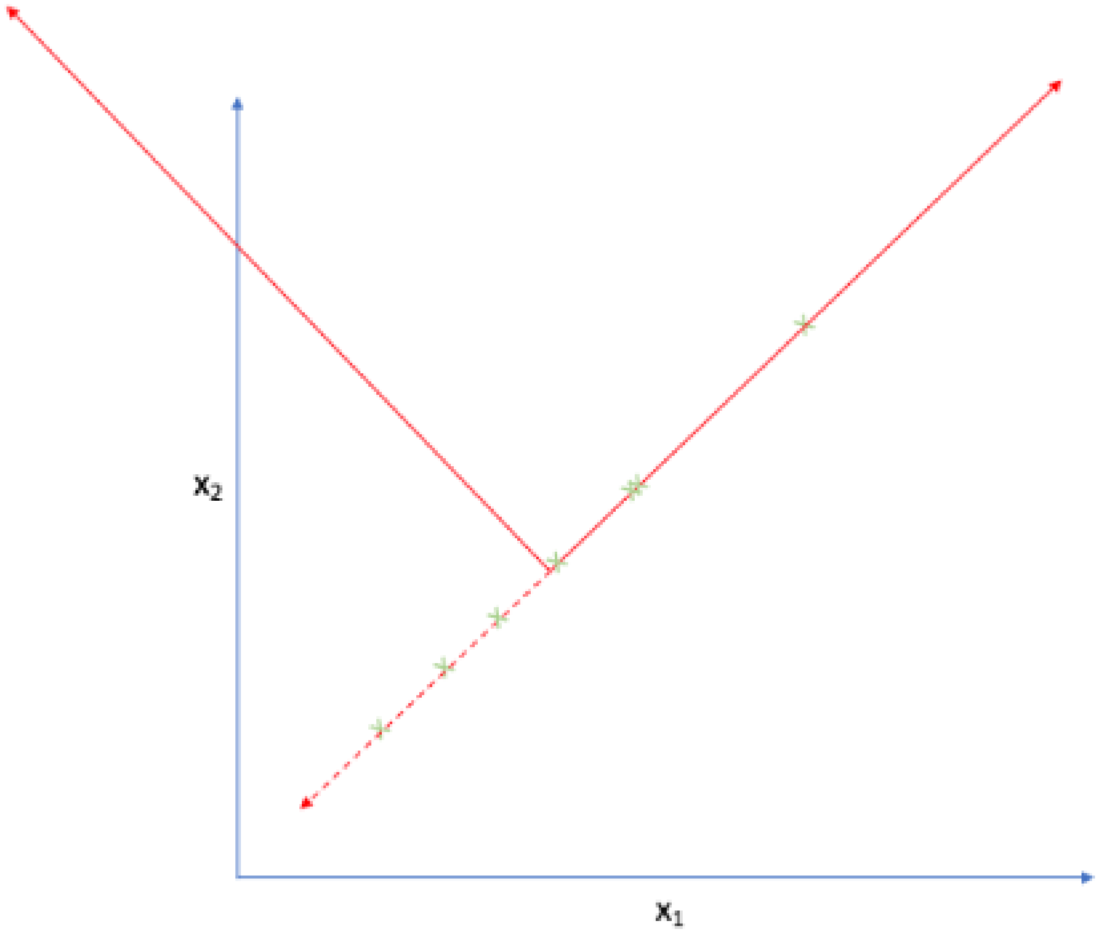
PCA: how it works

- Use linear algebra (singluar value decomposition) to project each point onto the primary axis



PCA: how it works

- We have just reduced the dimension from two to one
- Captures most of the information from our data

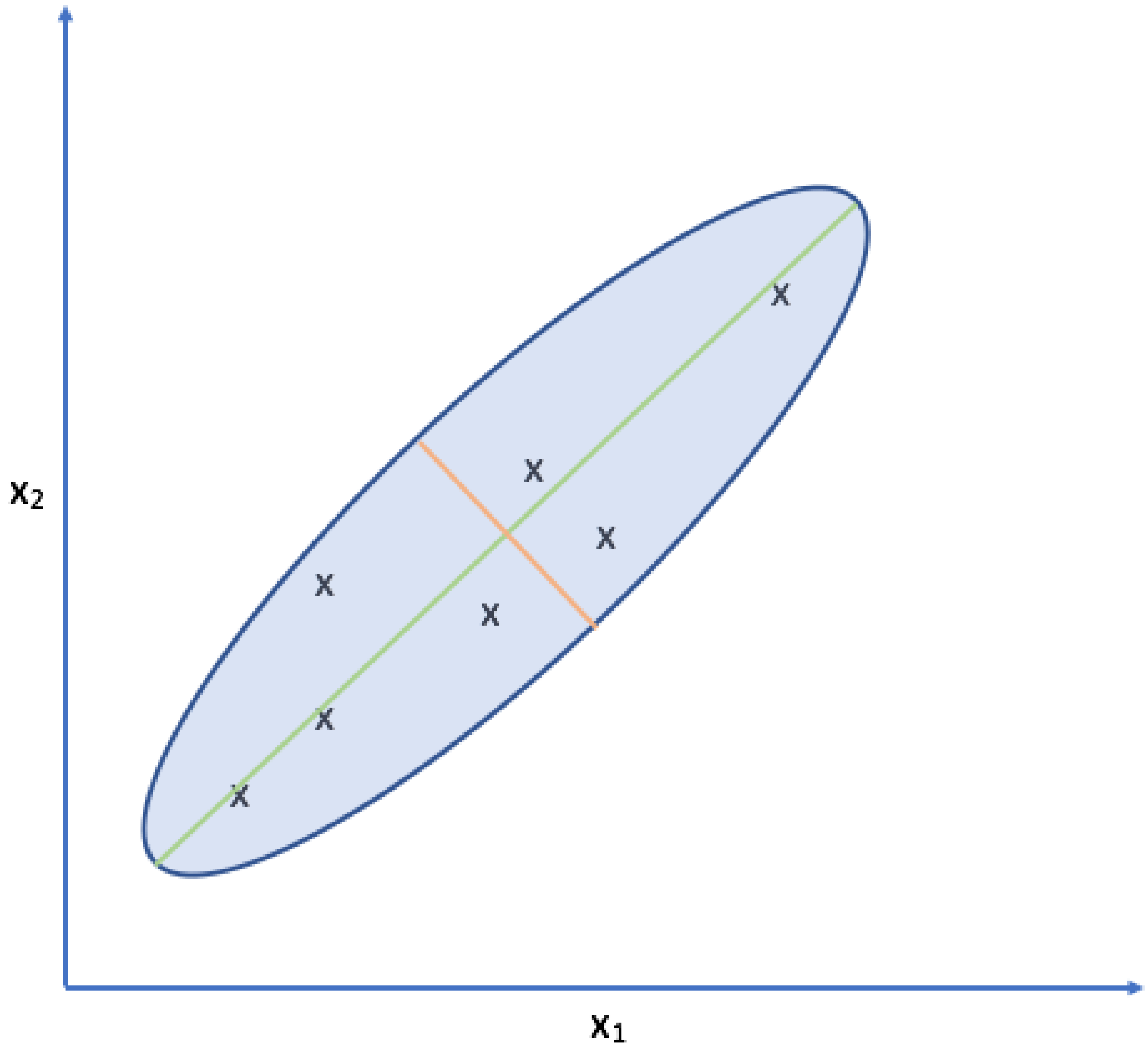


PCA: how it works

- We measure loss as the cumulative distance between our observations and the projected points
- We want to minimize this (aka cost function)

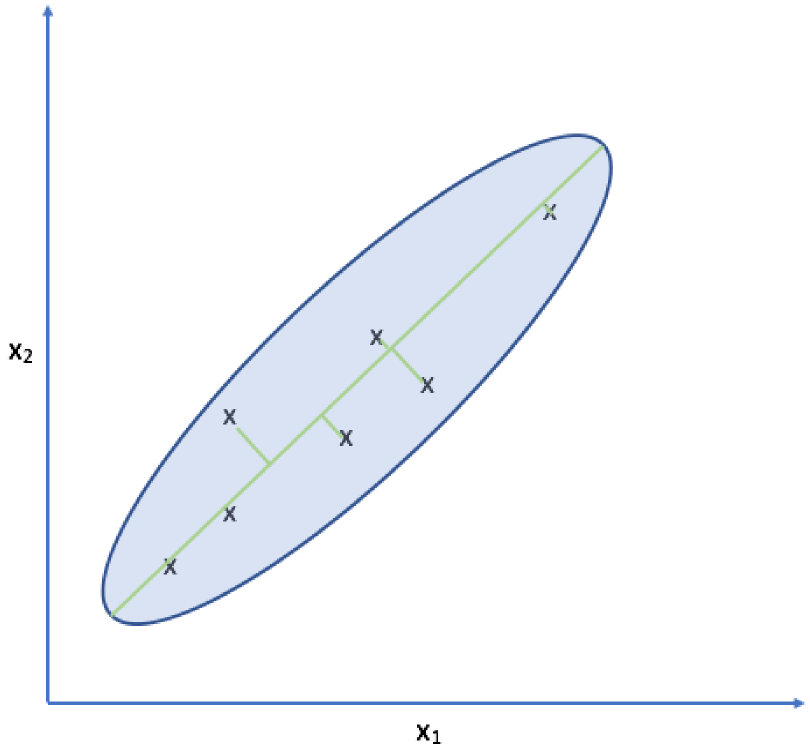
PCA: how it works

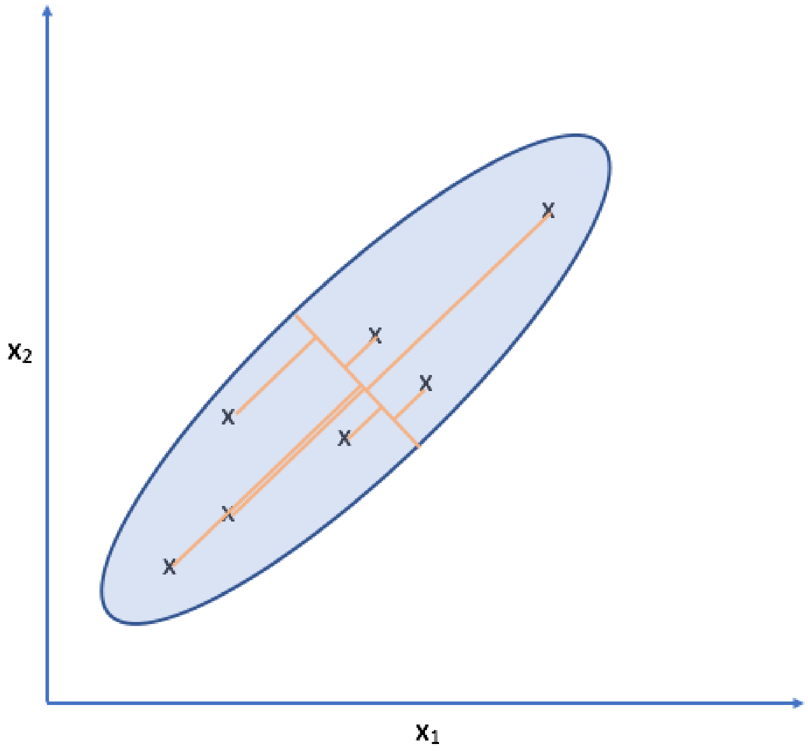
- We use variance to select best direction
- Spread is smaller in green direction than orange
- We'd have to move our points further to get the on the orange axis



PCA: variance

- Vector that can capture maximum variance is the one that can project points to its axis with minimum distance
- Maximum variance minimizes cost (distance)





Goal of PCA

- Find set of k vectors on which we can project our data to reduce the feature-space from n-dimensions to k-dimensions
- Minimize projection error (cost or loss)
    - By finding direction with max variance

Remember the steps?

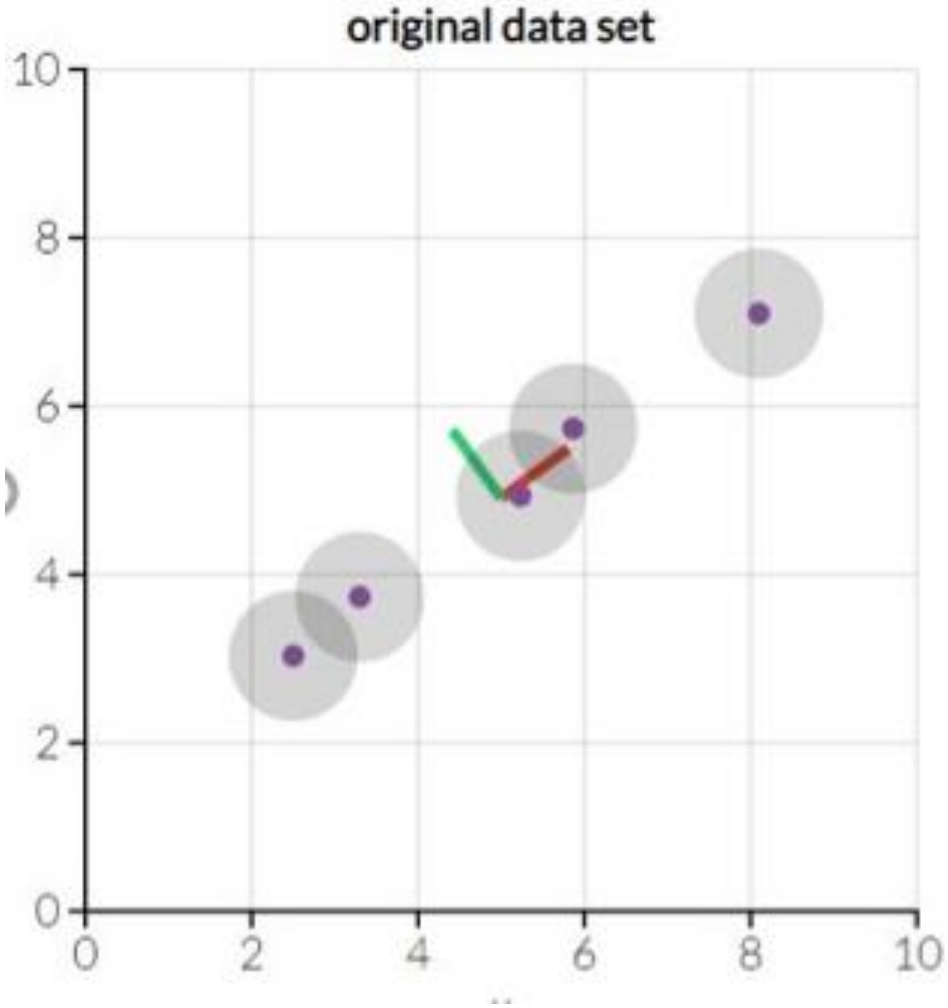
1. Ensure all measurements are on the sam scale, or use a correlation matrix

2. **Covariance Matrix:** Calculate a square, symmetric matrix that summarizes the relationship between the variables

3. Determine direction (**eigenvectors**) and magnitude (**eigenvalues**) of the points

4. Transform data to align with most important direction found in data

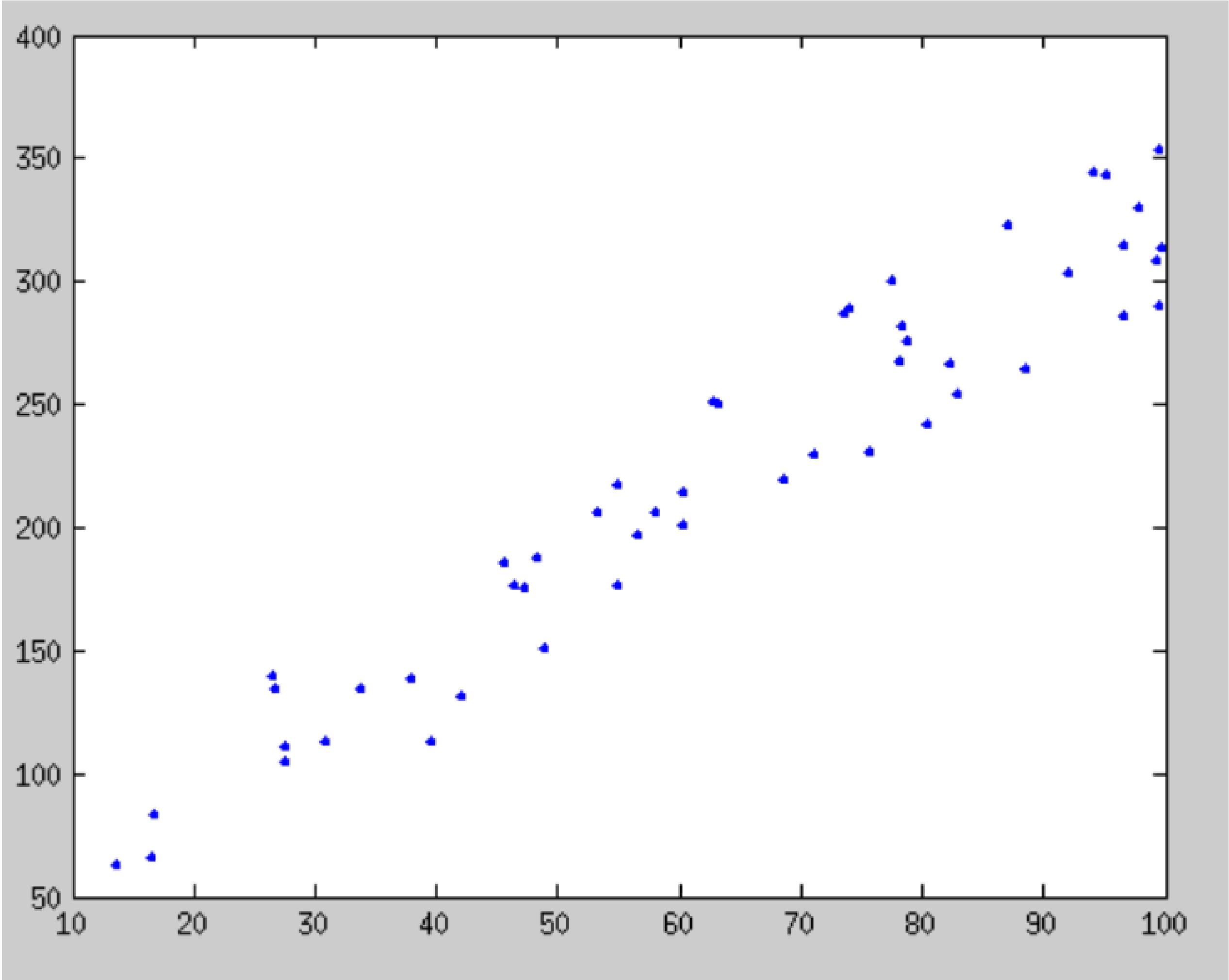
- Largest eigenvalues = more important



Computing the matrix: Singular Value Decomposition (SVD)

- Each point (x_i, y_i) is a pair
- Assume roughly linear relationship:
  - $ax_i + by_i - c = 0$
- Represent as vector a:

$$a\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + c\begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{0}$$



Computing the matrix: Singular Value Decomposition (SVD)

- Variance measures spread of a single variable

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Covariance measures how much two random variables vary together (*below, x-hat is mean of vector x*):

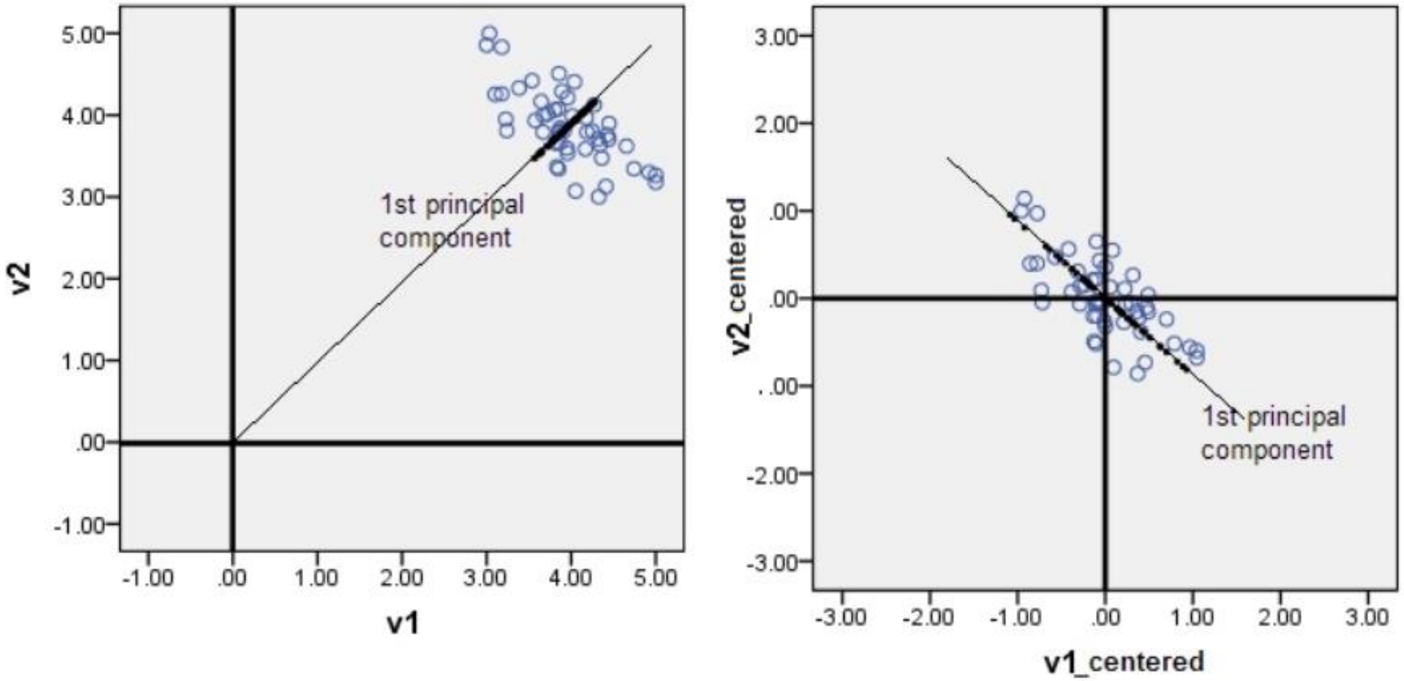
$$\sigma(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Computing the matrix: Singular Value Decomposition (SVD)

$$a\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + c\begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & -1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & -1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{0}$$

- Matrix a is square: $A_{i,j}=\sigma(x_i,x_j)$
- Dimensions = $A \in R_2{}^{d\times d}$ (two dimensional vector space with given square dimension)

Computing the matrix: Singular Value Decomposition (SVD)



- Scale data by subtracting mean from each entry
  - Sets intercept to 0
  - This allows the points to originate from the origin (0) without changing the slope of the line
  - Homogenizes the variance into a linear equation

Computing the matrix: Singular Value Decomposition (SVD)

$$\Sigma = \begin{bmatrix} \operatorname{cov}(x_1, x_1) & \operatorname{cov}(x_1, x_2) & \cdots & \operatorname{cov}(x_1, x_n) \\ \operatorname{cov}(x_2, x_1) & \operatorname{cov}(x_2, x_2) & \cdots & \operatorname{cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(x_n, x_1) & \operatorname{cov}(x_n, x_2) & \cdots & \operatorname{cov}(x_n, x_n) \end{bmatrix}$$

- Compute covariance matrix
  - You can read more on how this is done here:
  - https://towardsdatascience.com/let-us-understand-the-correlation-matrix-and-covariance-matrix-d42e6b643c22
  - https://datascienceplus.com/understanding-the-covariance-matrix/

Computing the matrix: Singular Value Decomposition (SVD)

$$U = \begin{bmatrix} | & | & & | & & | \\ u_1 & u_2 & \cdots & u_k & \cdots & u_n \\ | & | & & | & & | \end{bmatrix}$$

$$U_{reduce} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_k \\ | & | & & | \end{bmatrix}$$

- Compute eigenvectors and eigenvalues
  - Using spectral decomposition of a matrix we create use the properties of a square, symmetric matrix [U,S,V]=svd(Σ) to find the eigenvalues and eigenvectors
  - We then pull out the top k columns of the resulting eigenvector
  - Compute projections onto our new feature space

Interpreting PCA

- Compare our approximation with original values
  - Correlation between eigenvalues and original value
  - Mean of 0
  - Standard deviation = square root of eigenvalue
  - Larger correlation in either direction is better
- **Explained variance**
  - Measure used to evaluate PCA
  - How much of the model's variance did we retain when reducing dimensionality?

Principle Components Regression

- Use when:
  - You have more features than samples
  - Your independent variables show collinearity
- Note on choosing k:
  - Use enough PC's to capture >= 90% of variation
  - Maximize adjusted R^2

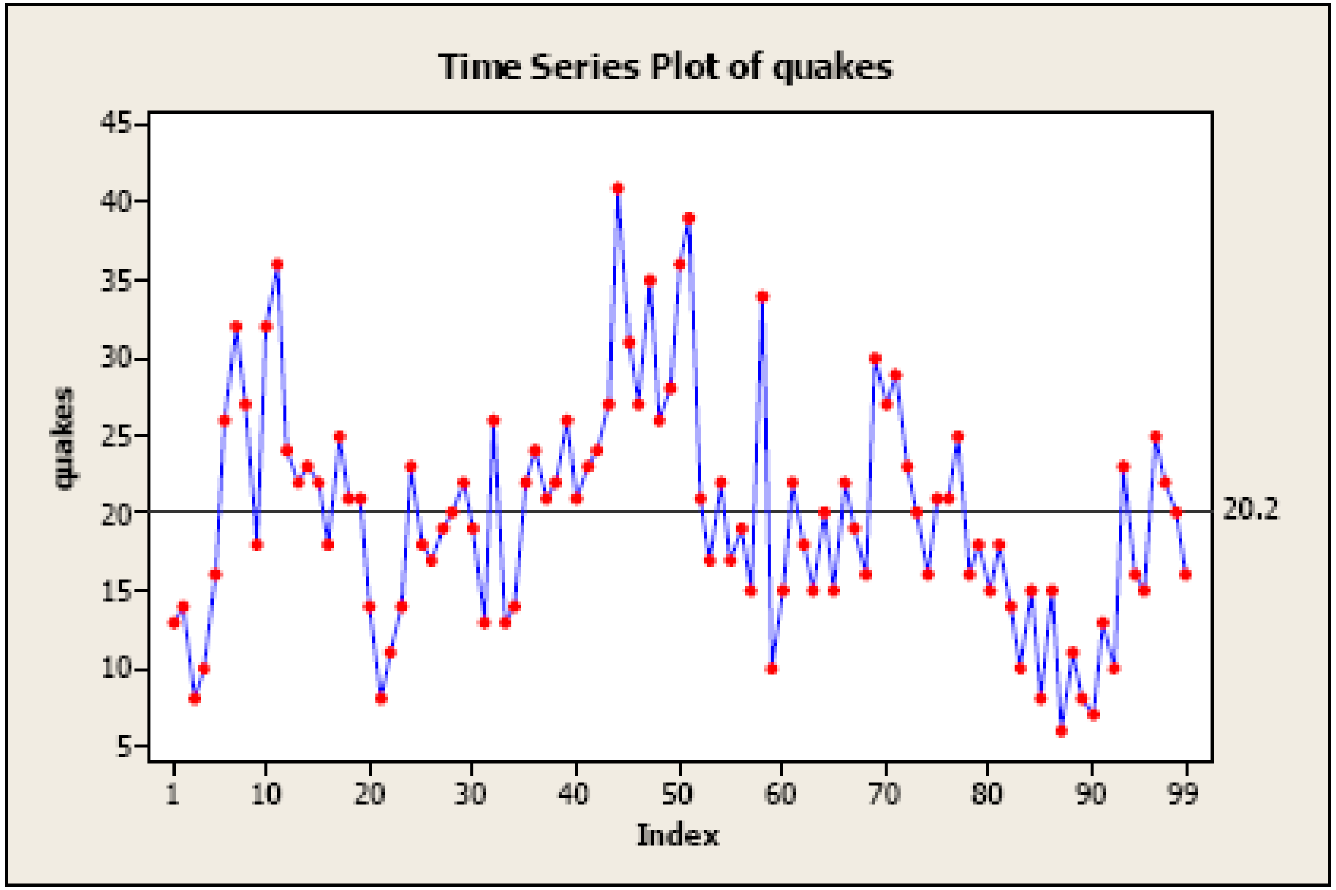
PCR Lab

Time Series Analysis

Time Series Analysis: Overview

- What is a time series and when is it used?
- Time series data structures
- Time series basic concepts:
  - Stationarity
  - White Noise
  - ARIMA Model
- Time series analysis
- Interpreting Time Series Analysis

Time Series Plot: variable plotted over time



Time Series Analysis: What is it?

Time series are a series of data points that occur in chronological order: $Y = (Y_1, Y_2, Y_3, \ldots, Y_T)$

- Values are measured at some point in time
- Can be measured at even intervals
  - And if not, scale to the same interval
- Timestamps record time of occurrence
  - Record a single point in time

Time Series Analysis: When is it used?

Time Series Analysis is used in a wide variety of fields:

- Financial: Daily stock market prices
- Science: Population Growth, Radioactive Decay
- Medicine: Treatment response time
- Signal/Image Processing: De-noising, de-blurring

Objectives:

- Estimate true values in the presence of noise or trends
- Forecasting and predicting the next value

Time Series: Trend, Seasonality and Noise

Time series data is generally described by three terms:

- **Trend**:
    - Is there a pattern to the way it changes over time?
    - Analyze historical data to make predictions.
    - Does not repeat during time interval studied
- **Seasonality:**
    - Does that data have a specific pattern during a specific period of time
    - Retail sales over holiday season
    - Repeats itself at intervals
- **Noise:**
    - Non-systematic components that is nor Trend/Seasonality within the data
    - Random, unpredictable data points
    - Error

Time Series: Data Structures

Time Stamp: Marks a specific point in time

```
Timestamp('2018-01-15 15:45:00')
```

Datetimeindex: Allows us to index our data frame for a time range, allowing us to resample at different frequencies

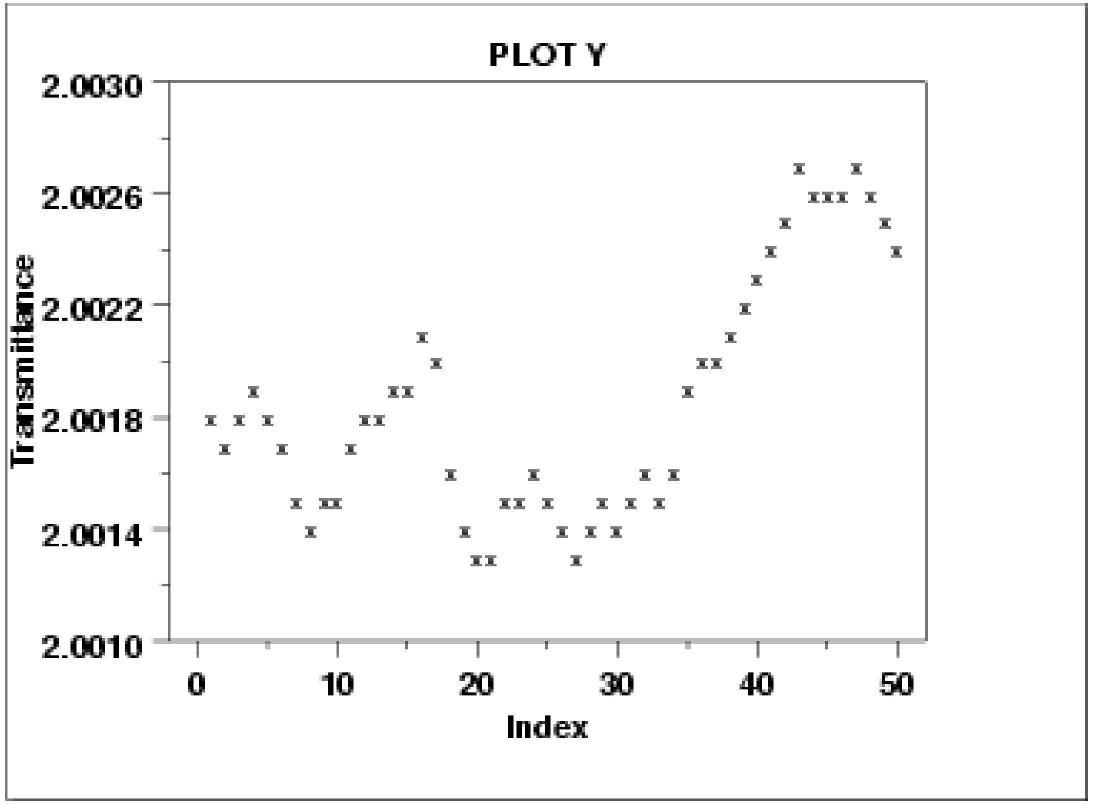
```
DatetimeIndex(['2018-01-01 00:00:00', '2018-01-01 01:00:00',
               '2018-01-01 02:00:00', '2018-01-01 03:00:00',
               '2018-01-01 04:00:00', '2018-01-01 05:00:00',
               '2018-01-01 06:00:00', '2018-01-01 07:00:00',
               '2018-01-01 08:00:00', '2018-01-01 09:00:00',
```

Time Series Analysis: Stationarity

A **stationary** time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time:

- Most time series in nature are not stationary
- Most time series forecasting methods assume that the data can be transformed to stationarity

Run Sequence Plot



Time Series: White Noise

A time series is white noise if its variables are independent and identically distributed with a mean of zero:

- All variables have the same variance
- Each value has zero correlation with any other value in the series
- This means your series is random and you cannot model it
- If your errors from a model are white noise = good

Time Series: Time Domain Models

ARIMA (for Autoregressive Integrated Moving Average):

- Linear model the relates present value of a series to past values and past prediction errors
- The parameters of the ARIMA model are :
  - **p**: The number of lag observations included in the model, also called the lag order.
  - **d**: The number of times that the raw observations are differenced, also called the degree of differencing.
  - **q**: The size of the moving average window, also called the order of moving average
- Learning from past data to predict value of interest
  - Ex: When is Mt. Rainier going to have its next eruption based on past patterns?

Regression

- Model pattern as trend over time

Time Series: Analysis

1. Identify Trends:
   - Rolling Average: for each time point, you take the average of the points on either side of it
     - Number of points specified by window size
     - Average smooths out trend
2. Examine Seasonality
   - Subtract trend from original signal
     - Removes the trend from the timeline to examine seasonality
   - **First order differencing**: look at the difference between a data point and the one before it
   - Helps transform to a stationary series for forecasting

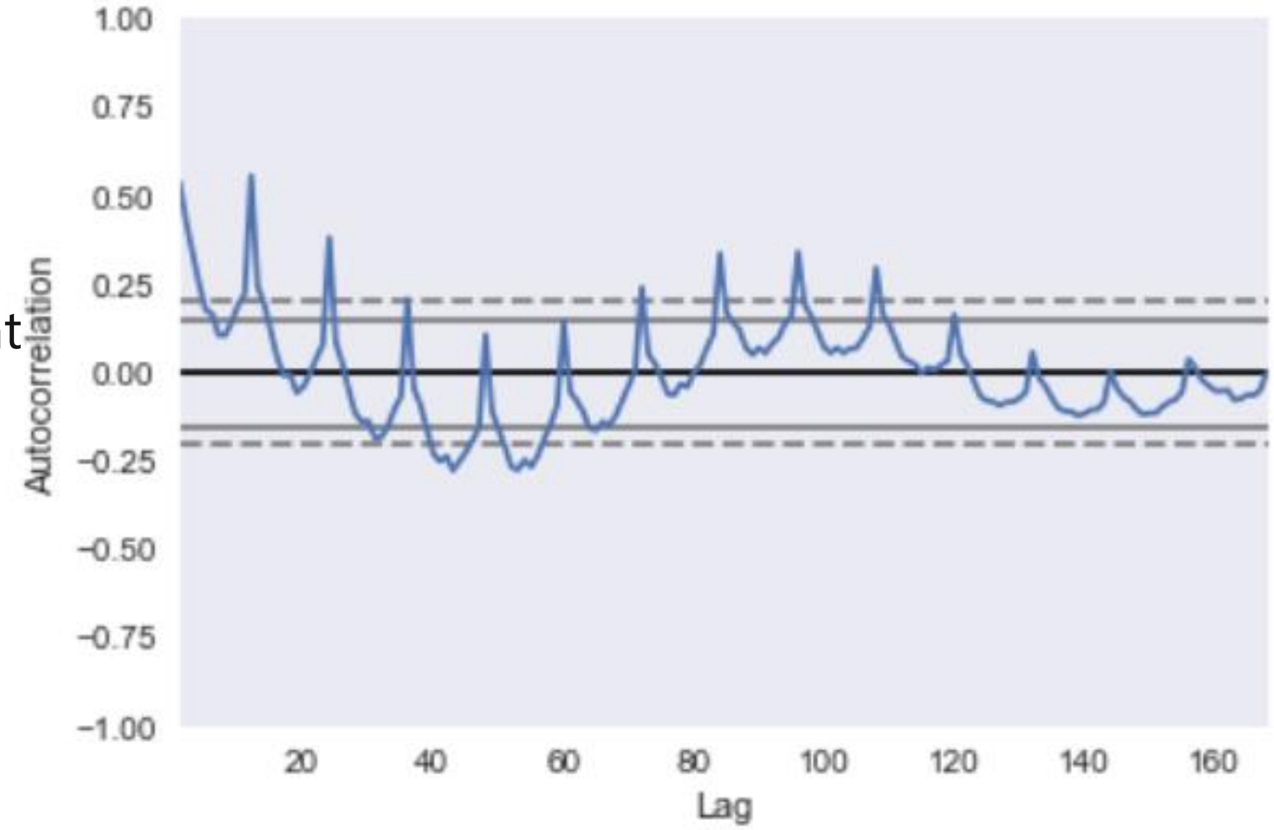
Time Series: Analysis

3. Analyze **periodicity** (does it repeat itself at periodic intervals, such as every few months?)

 – **Autocorrelation**: If there is a cyclic component, you can basically map each cycle onto itself at another point (**Fourier Transform**) in time

 – Calculate correlation coefficient between cycles

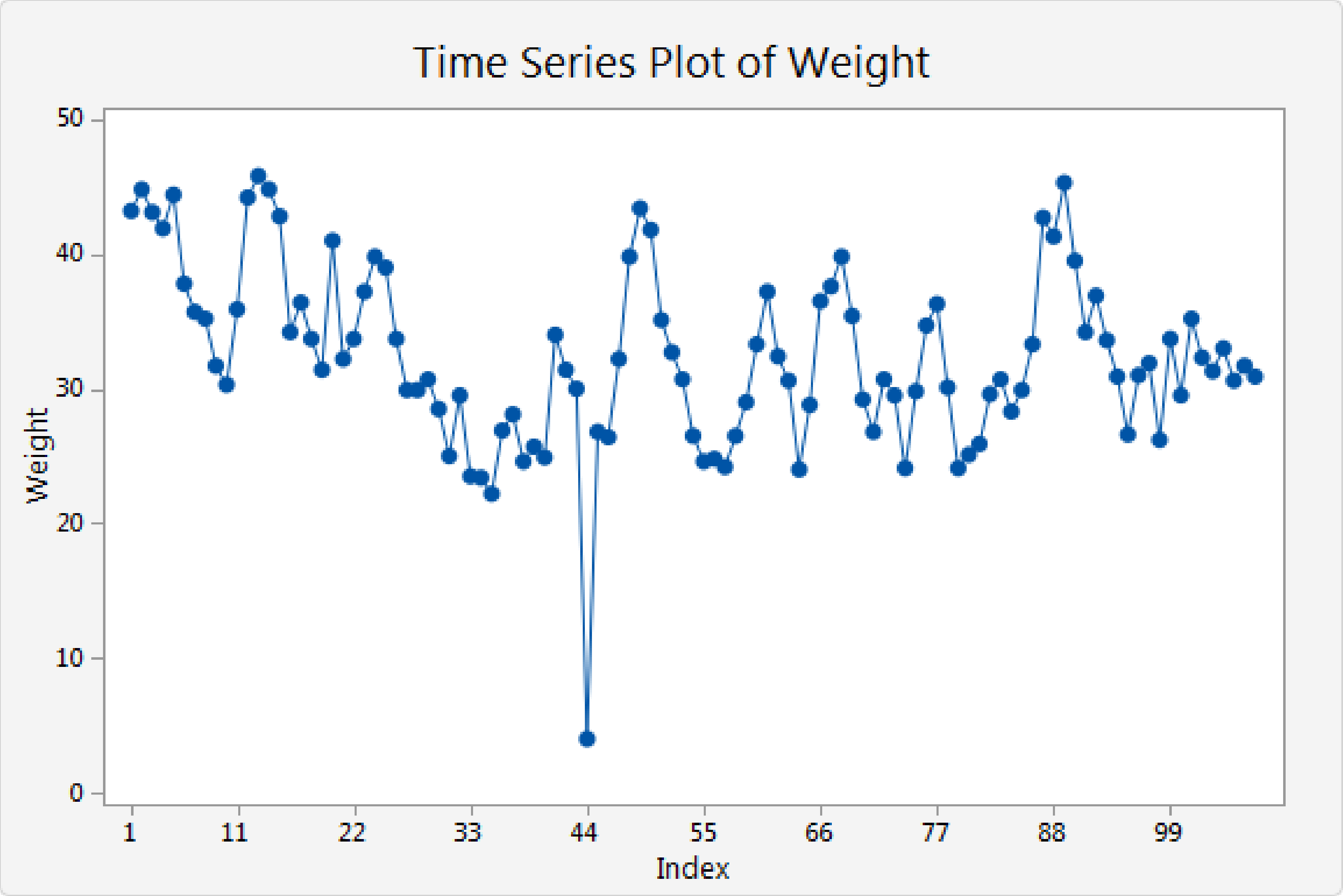
 – Lag: similarity between observations as a function of the **lag** between them



4. Use ARIMA or regression to **forecast** (make a prediction)

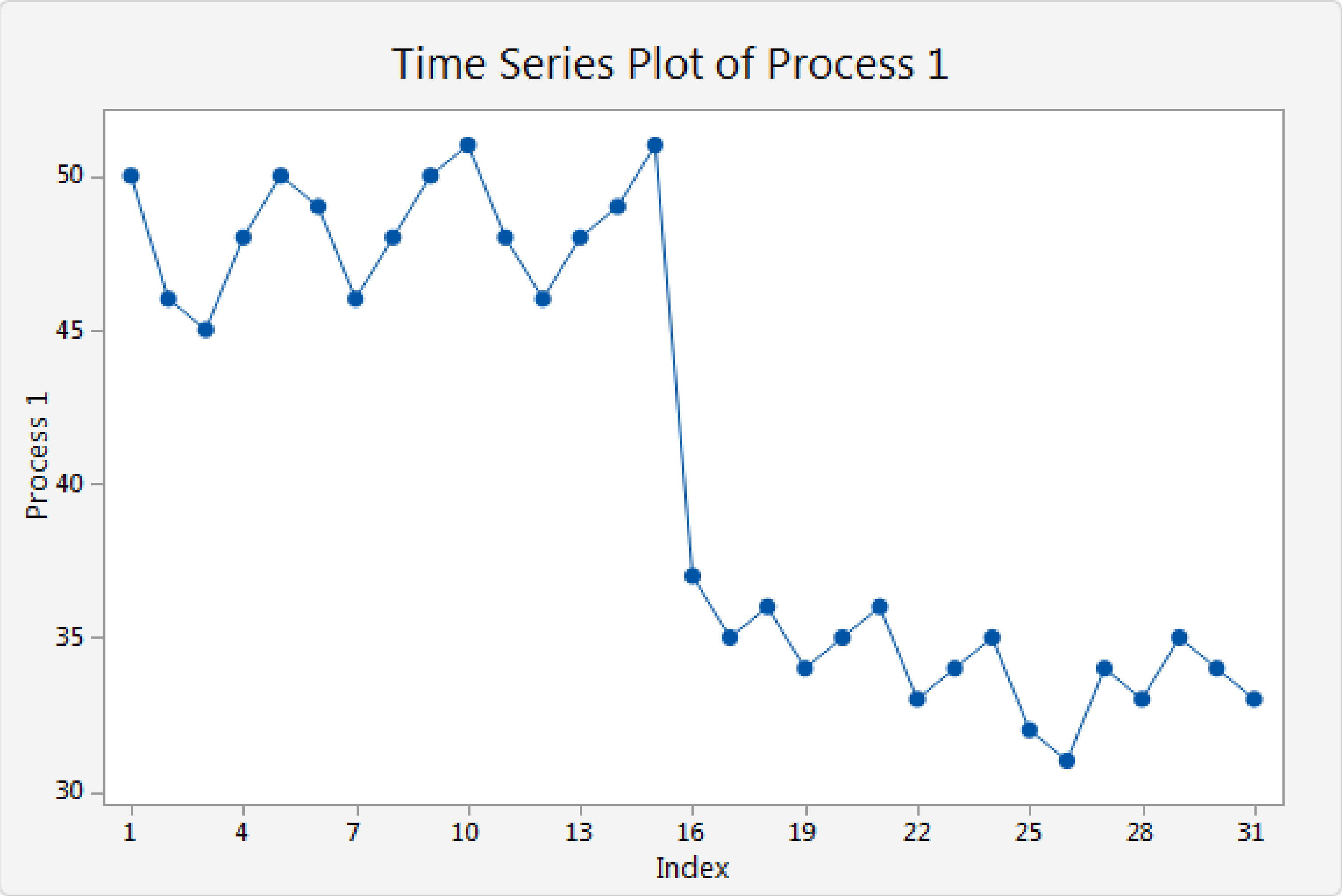
Time Series: Interpreting Results of Analysis

Outliers



Time Series: Interpreting Results of Analysis

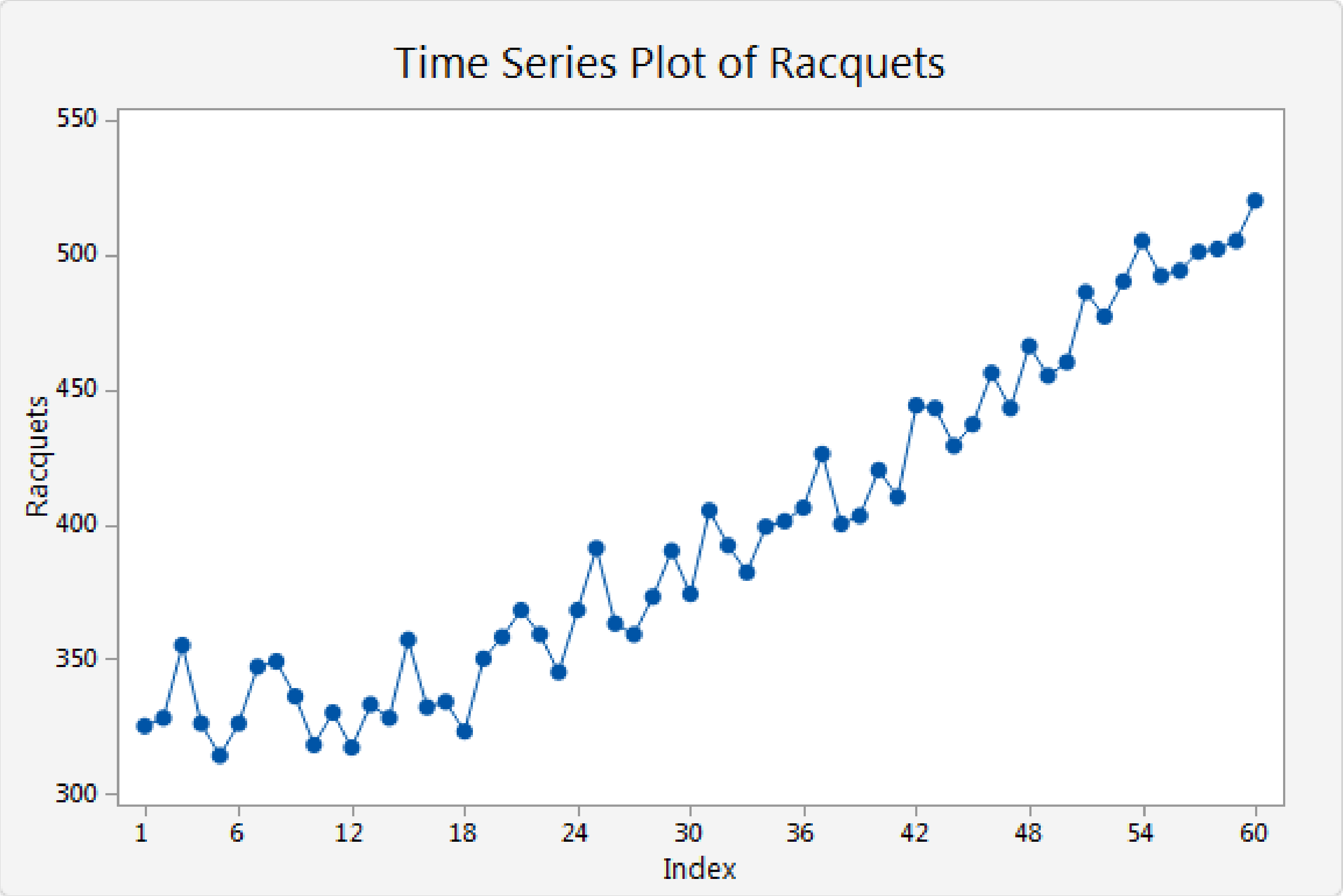
Sudden Shifts



Time Series: Interpreting Results of Analysis

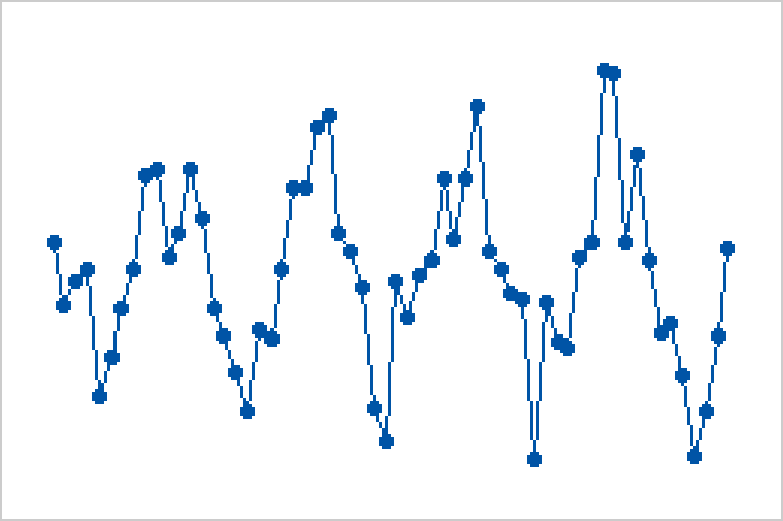
Trends: long term increase or decrease

Ex. Clear upward trend

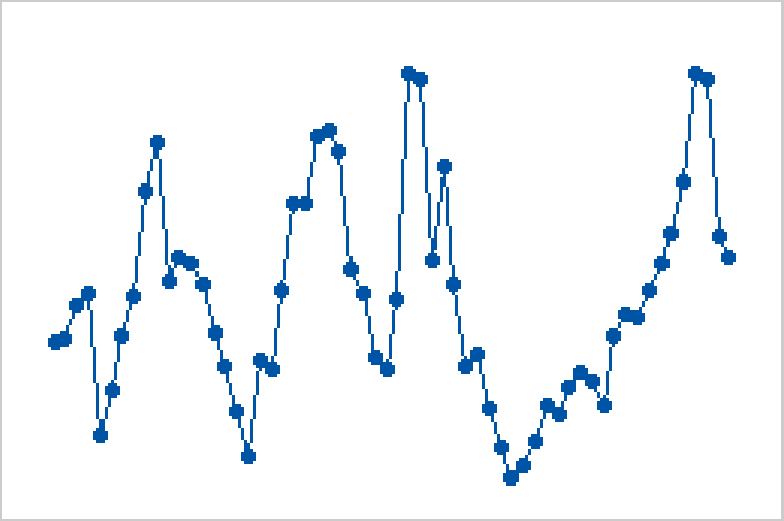


Time Series: Interpreting Results of Analysis

Seasonal Patterns vs Cycles



Seasonal: Patterns repeat

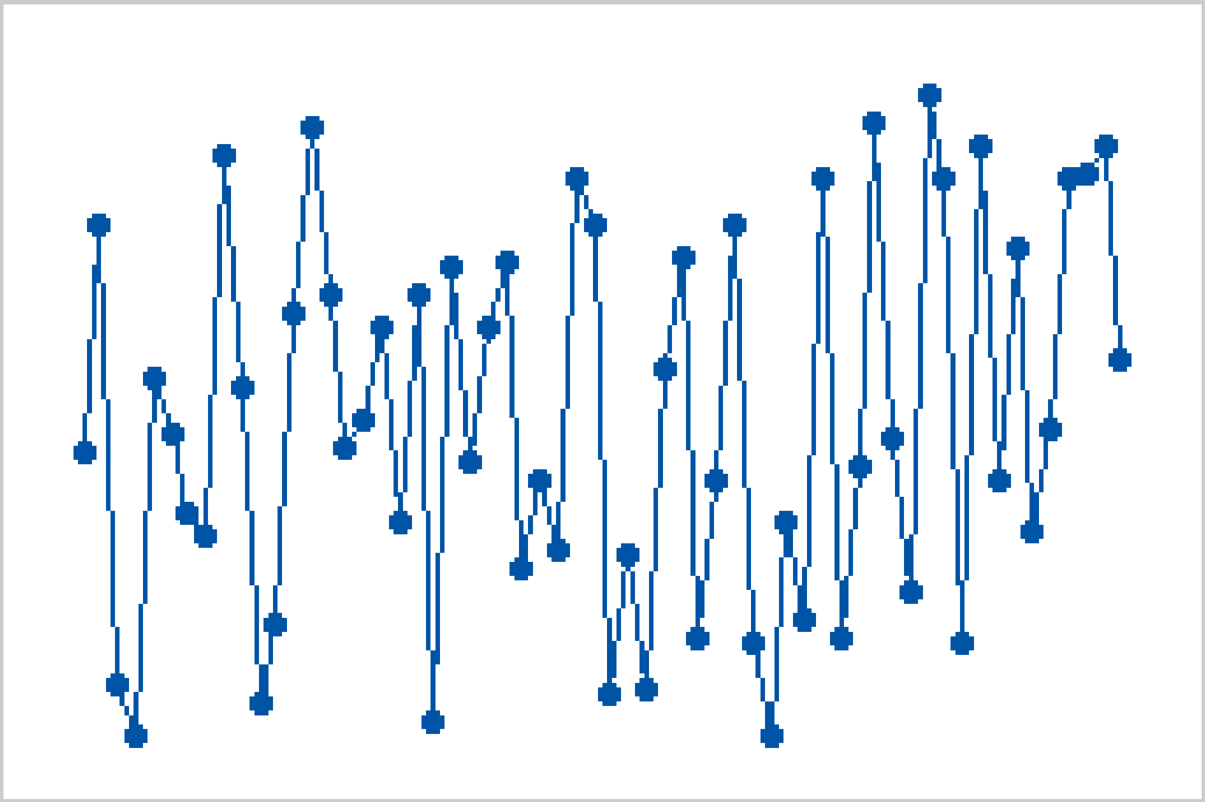


Cycle: do not repeat at regular intervals, but still ups and downs

Time Series: Interpreting Results of Analysis

Randomness

No patterns to be found



Basic Time Series Manipulation Lab