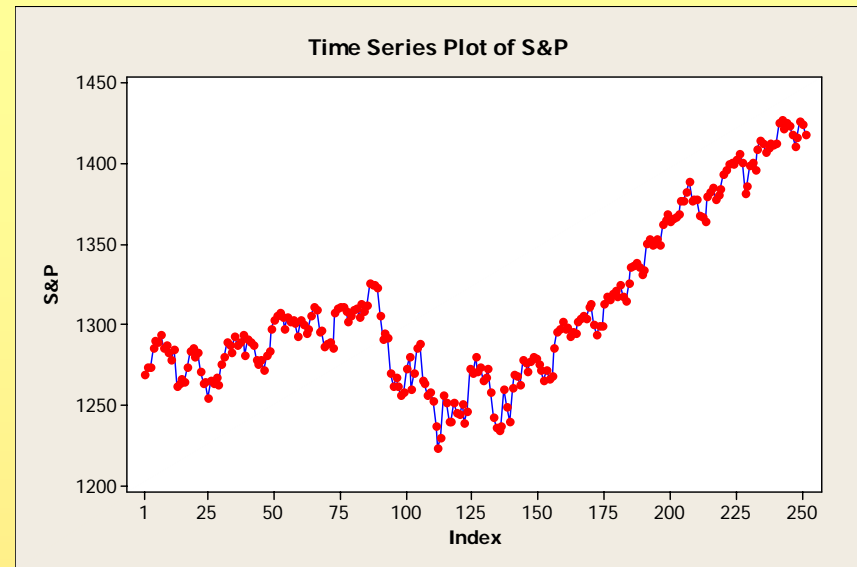
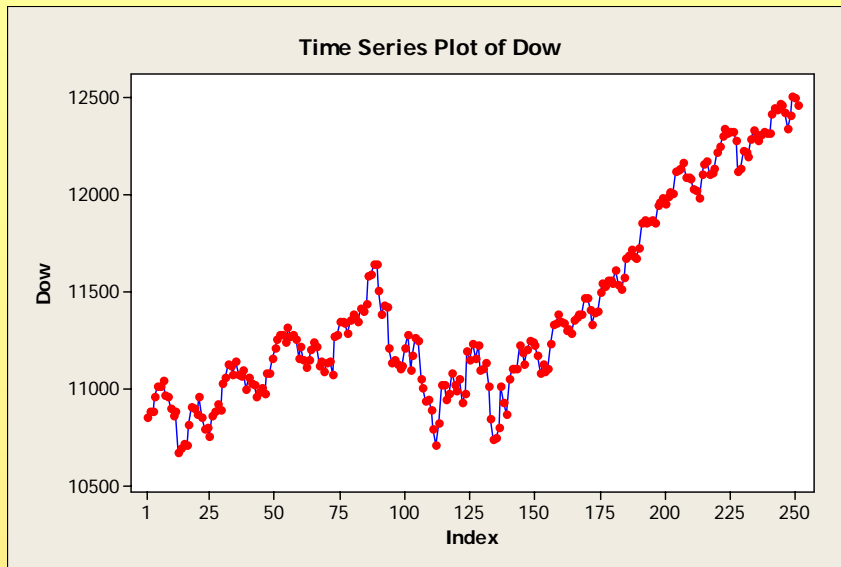


Principal Component Analysis (PCA)

- Principal component analysis (PCA) creates new variables (components) that consist of uncorrelated, linear combinations of the original variables.
- PCA is used to simplify the data structure and still account for as much of the total variation in the original data as possible.

Simple Case: Stock Market Data

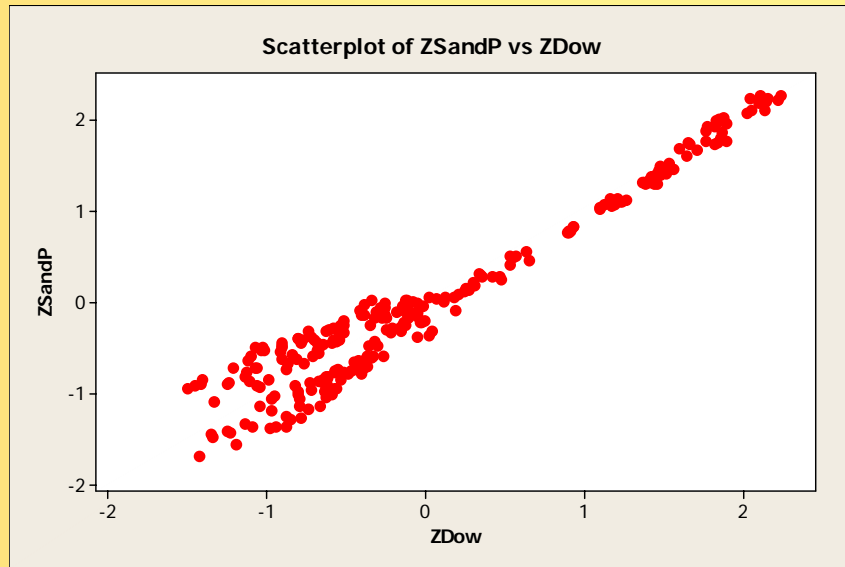


Can the data be reduced to just one linear combinations of the original variables be used without losing much information?

3 Steps for PCA

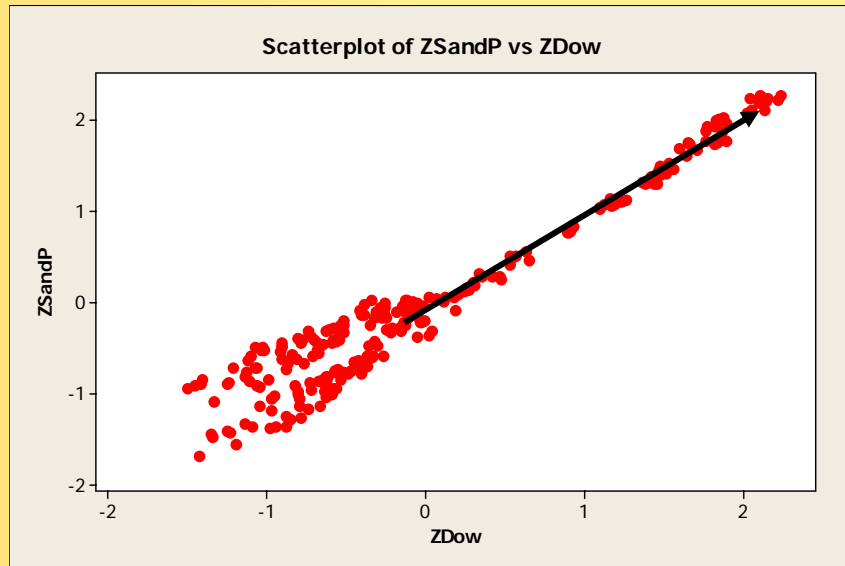
- 1) Calculate the correlation matrix
- 2) Calculate the eigenvectors of the correlation matrix
- 3) Multiply the eigenvectors by the standardized original data. The first principal component (PC1) is a linear combination of the standardized data with the first eigenvector is used as the weights.

Simple Case: Stock Market Data



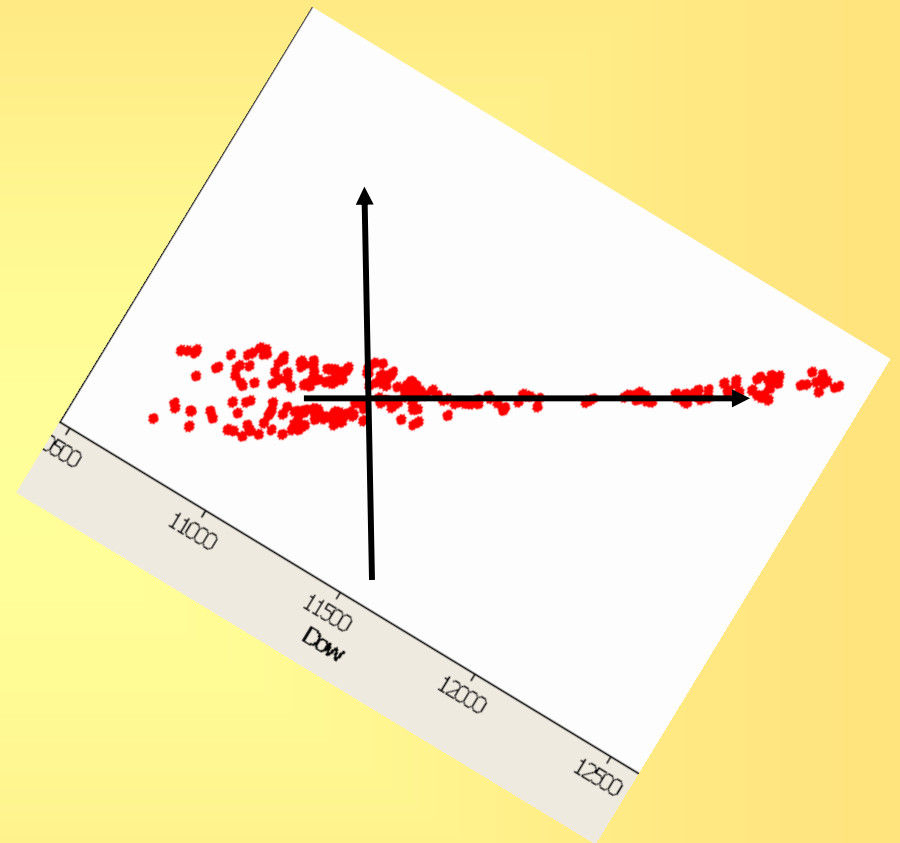
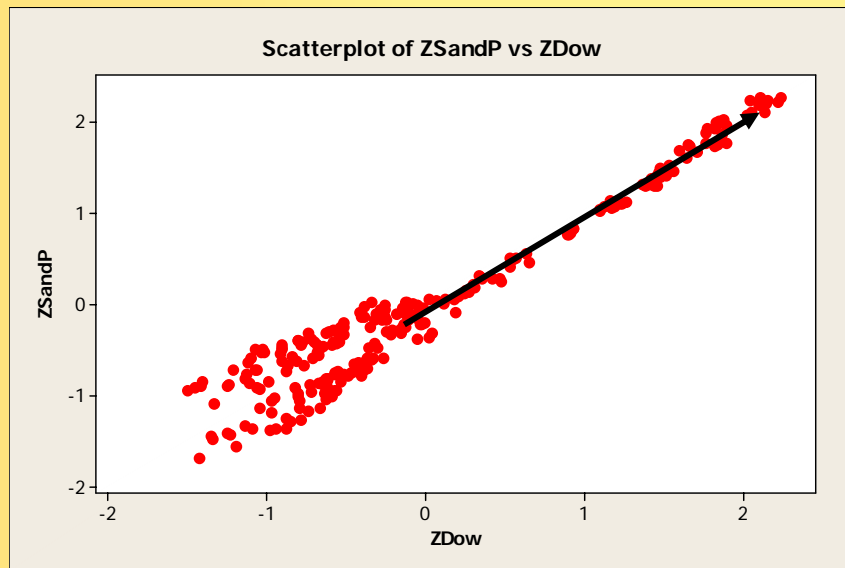
Standardized closing values of 2006 Dow Index vs
2006 S&P 500

Simple Case: Stock Market Data



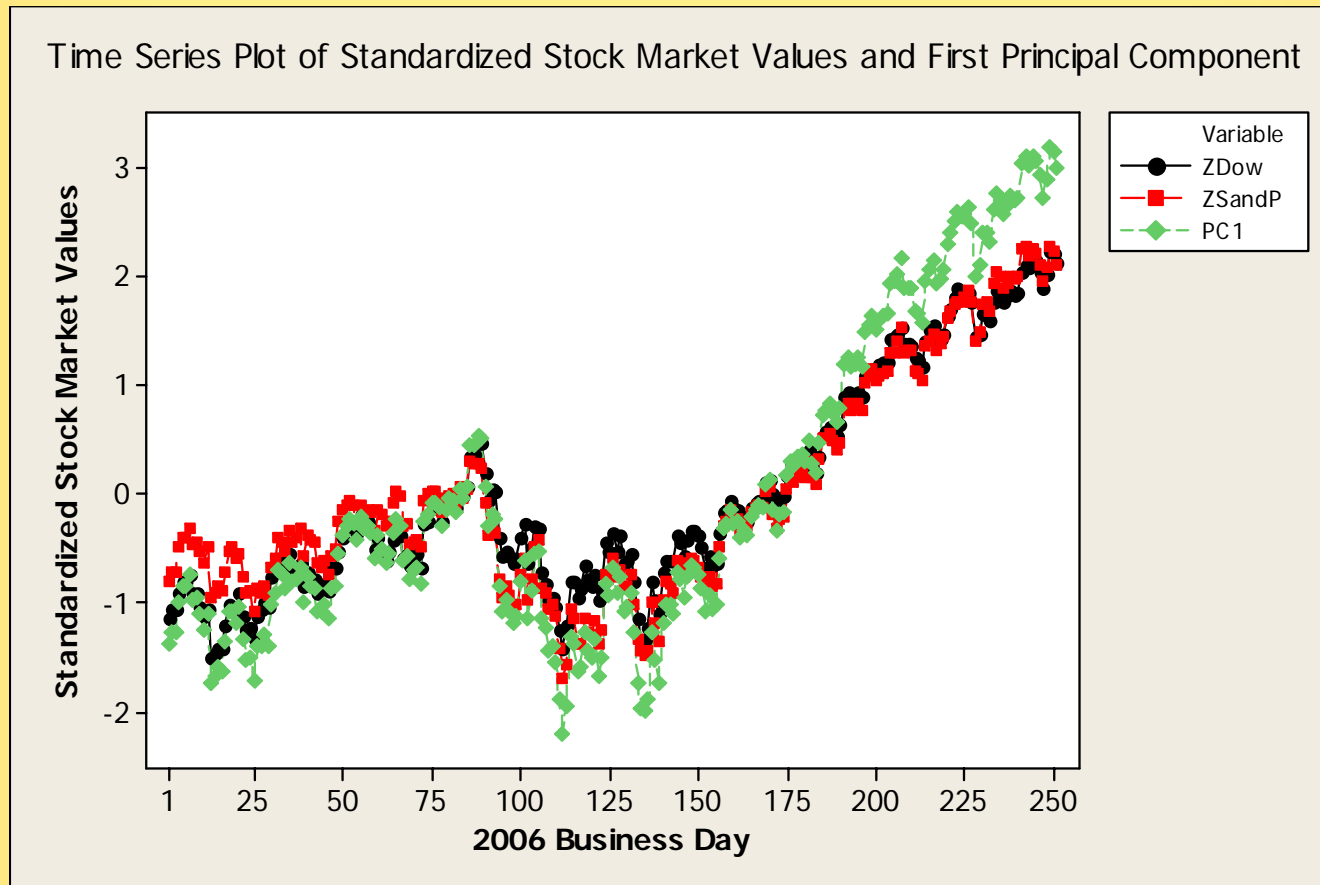
Direction of first principal component (the first eigenvalue).

Simple Case: Stock Market Data



Rotating the data to the first principal component. PC1 is a linear combination of the standardized data with the first eigenvector is used as the weights.

Simple Case: Stock Market Data



LAB: Principal Component Analysis in Environmental Studies

The Debate Over Statistical Techniques Used in the Derivation of the Global Warming Hockey Stick Graph

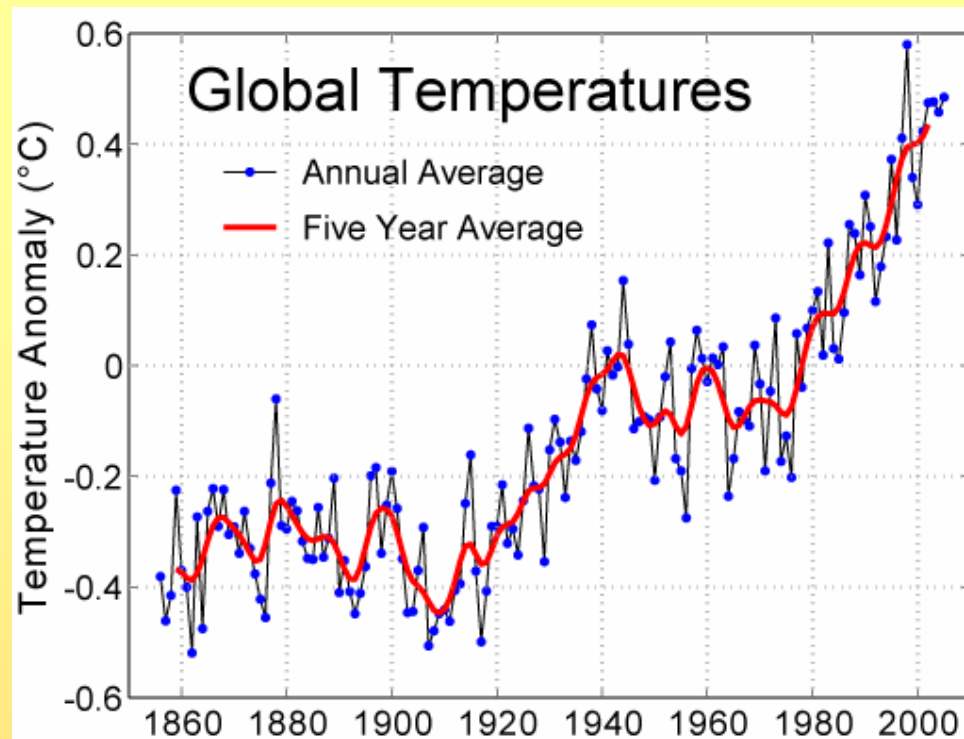


Figure 1: The instrumental record of global average temperatures.

The Hockey Stick Graph

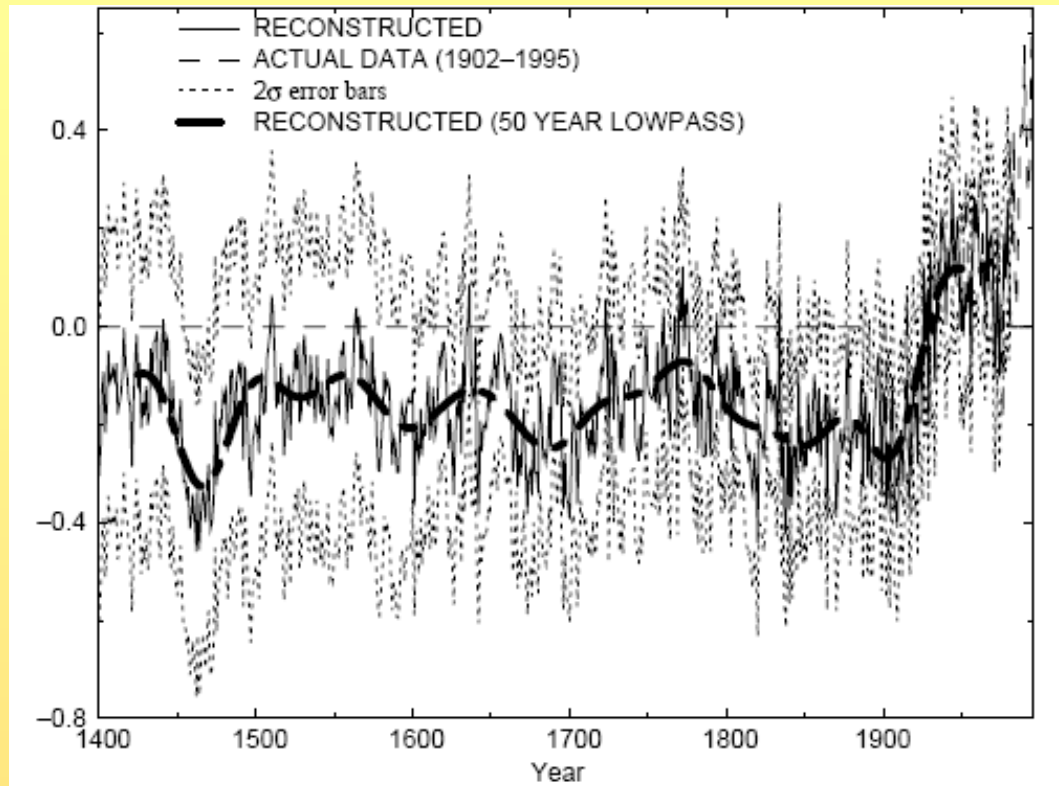


Figure 2: Mann's 1998 Hockey Stick Graph

The Hockey Stick Graph

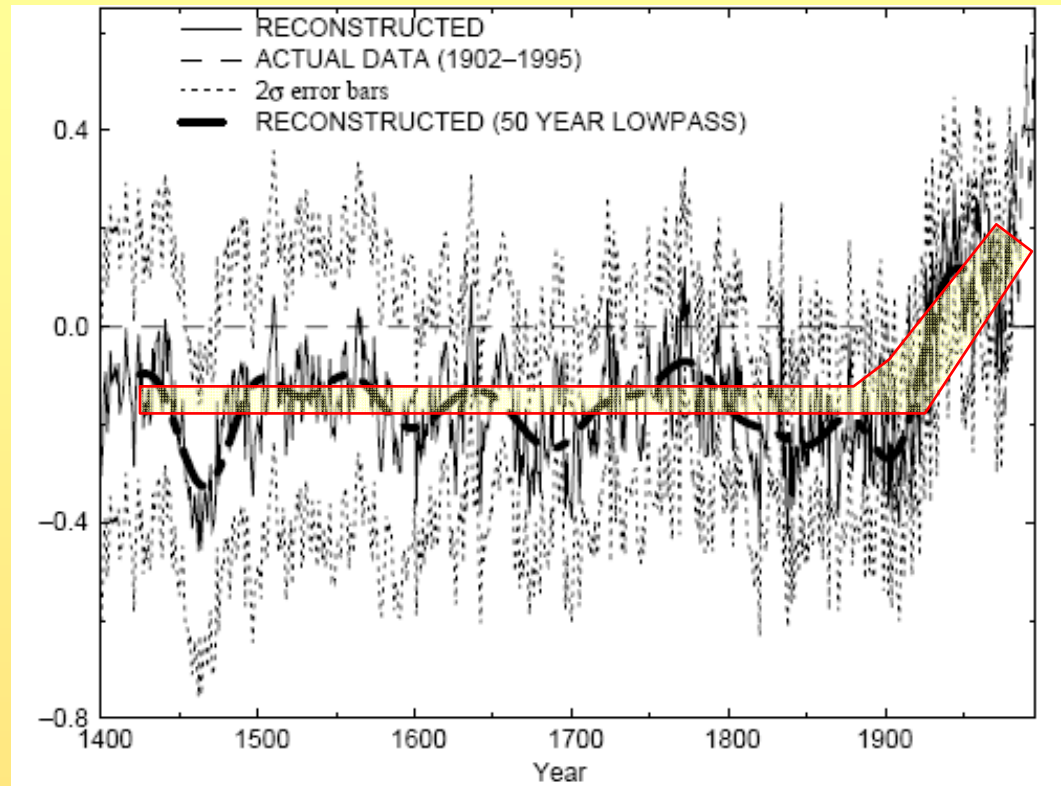
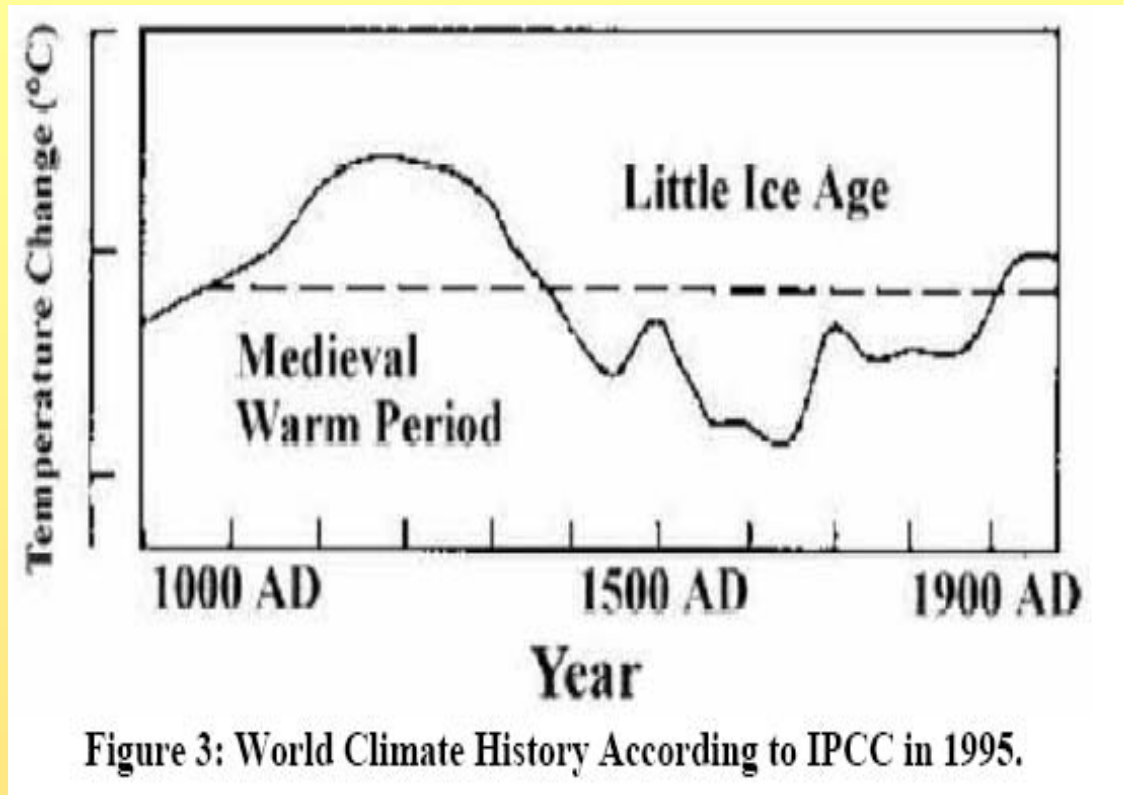


Figure 2: Mann's 1998 Hockey Stick Graph

The Hockey Stick Graph



The Hockey Stick Graph

- In 1998 Mann, Bradley, and Hughes (MBH) used a modified PCA to reduce 70 series of proxy data to one principal component (PC1).
- MBH's graph was widely used as evidence of global warming.
- In 2003 McIntyre and McKittrick (MM), claimed that the graph was not correct – but had a significant amount of trouble getting published.
- In 2005 MM published a simulation study that showed that MBH's modified PCA technique would consistently result in a hockey stick shape.
- In 2006 Ed Wegman provided an ad-hoc committee report to congress on the “Hockey Stick Global Climate Reconstruction”, <http://www.heartland.org/pdf/19383.pdf> .

The Hockey Stick Graph

- MBH used data from 1400-1980, 581 observations for each of the 70 proxy variables (tree ring data)
- Each variable would typically be standardized by the following formula:

$$\frac{X - \mu_{[1400:1980]}}{S_{[1400:1980]}}$$

- MBH use a ‘decentered’ standardization:

$$\frac{X - \mu_{[1902:1980]}}{S_{[1902:1980]}}$$

- What is the mean and standard deviation of a ‘decentered’ variable?
- How will this impact principal component analysis?

Simulation Study of the Hockey Stick Graph

Questions 1 and 2: Generate a matrix of random AR(1) data.

AR(1) data follows the general pattern of tree ring growth in many trees.

Question 3: Standardize the data matrix

Question 4: Perform PCA on a random AR(1) matrix with 70 series.

Question 5: Write a function that repeats question 4 ten times.

Question 6: Write a function that repeats question 5, but uses a ‘decentered’ standardization.

Does it look like ‘hockey stick’ shaped graphs occur more often with decentered data? Can we conduct a more thorough simulation study?

Proposed Course

Week 1: Review of Statistics 101

Lab: Making connections between the two sample t-test, ANOVA, and regression

Week 2-3: Randomization Tests/Nonparametric Tests

Activity: Westvaco discrimination case

Week 4-6: Multiple Regression

Intro Lab: How much is your car worth?

Lab: Population control and economic growth

Week 7-9: Designing an Experiment

Intro Lab: Weight gain in pigs

Lab: Perfection- reaction time tests

Week 10-12: Principal Component Analysis

Intro Lab: Stock market values

Lab: Global warming and the hockey stick graph

Week 13 and 14: Final Projects