

## Introduction

- a *time series* is a set of observations made sequentially in time (but techniques from time series analysis often applied to observations made sequentially in depth, along a transect, etc.)
- denote set by either  $\{x_t : t \in T\}$  or  $\{x(t) : t \in T\}$
- $x_t$  or  $x(t)$  is observation indexed by  $t$ , where  $t$  is a value we can use to figure out the actual time (e.g.,  $t = 50$  might mean the 50th day of year 2020, i.e., Feb. 19th)
- $T$  is the index set, which typically is either
  - discrete:  $T = \{1, 2, \dots, n\}$  or  $T = \{\dots, -1, 0, 1, \dots\} \stackrel{\text{def}}{=} \mathbb{Z}$  (set of all integers – here ‘ $\stackrel{\text{def}}{=}$ ’ is ‘equal by definition’); or
  - continuous:  $T = [a, b]$  (a closed interval such as the unit interval  $[0, 1]$ ) or  $T = (-\infty, \infty) \stackrel{\text{def}}{=} \mathbb{R}$  (set of all real numbers)

# Classification of Time Series: I

- by index set  $T$  (discrete or continuous): ‘discrete parameter’ time series & ‘continuous parameter’ time series (‘parameter’ used here rather than ‘time’ to avoid overworking ‘time’!)
  - discrete parameter time series are typically – but not always – equally spaced (some are ‘gappy’ – stock market prices due to market closures on weekends, holidays etc.)
  - discrete parameter time series might be intrinsically discrete (number of planes departing Seatac every day) or might arise from underlying continuous parameter time series via
    - \* sampling (instantaneous temperature each day at noon)
    - \* aggregation (daily average temperature)
    - \* extrema (daily maximum temperature)

## Classification of Time Series: II

- by number of values recorded
  - one value: univariate
  - more than one value: vector-valued or multivariate
    - \* special case: bivariate time series (two values)

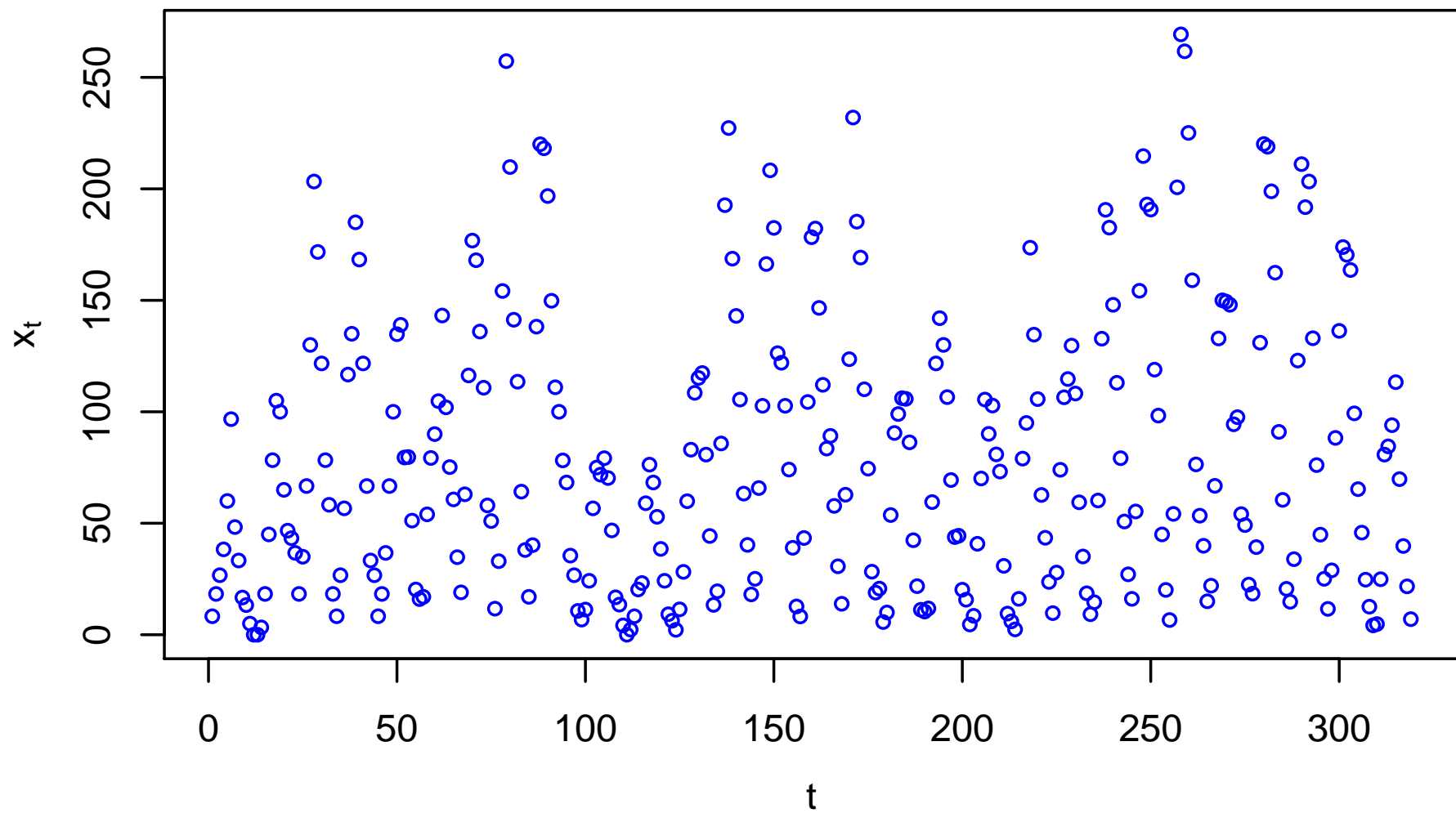
## Classification of Time Series: III

- by type of value  $x_t$  can assume
  - real-valued: a value in  $\mathbb{R}$  or subset thereof (temperatures)
  - complex-valued (applications in electrical engineering)
  - nonnegative integer (whale population)
  - categorical (outcome of scheduled football match: win, lose, tie, forfeit, cancellation)
  - binary-valued (coding for two categories)
  - circular (wind direction)

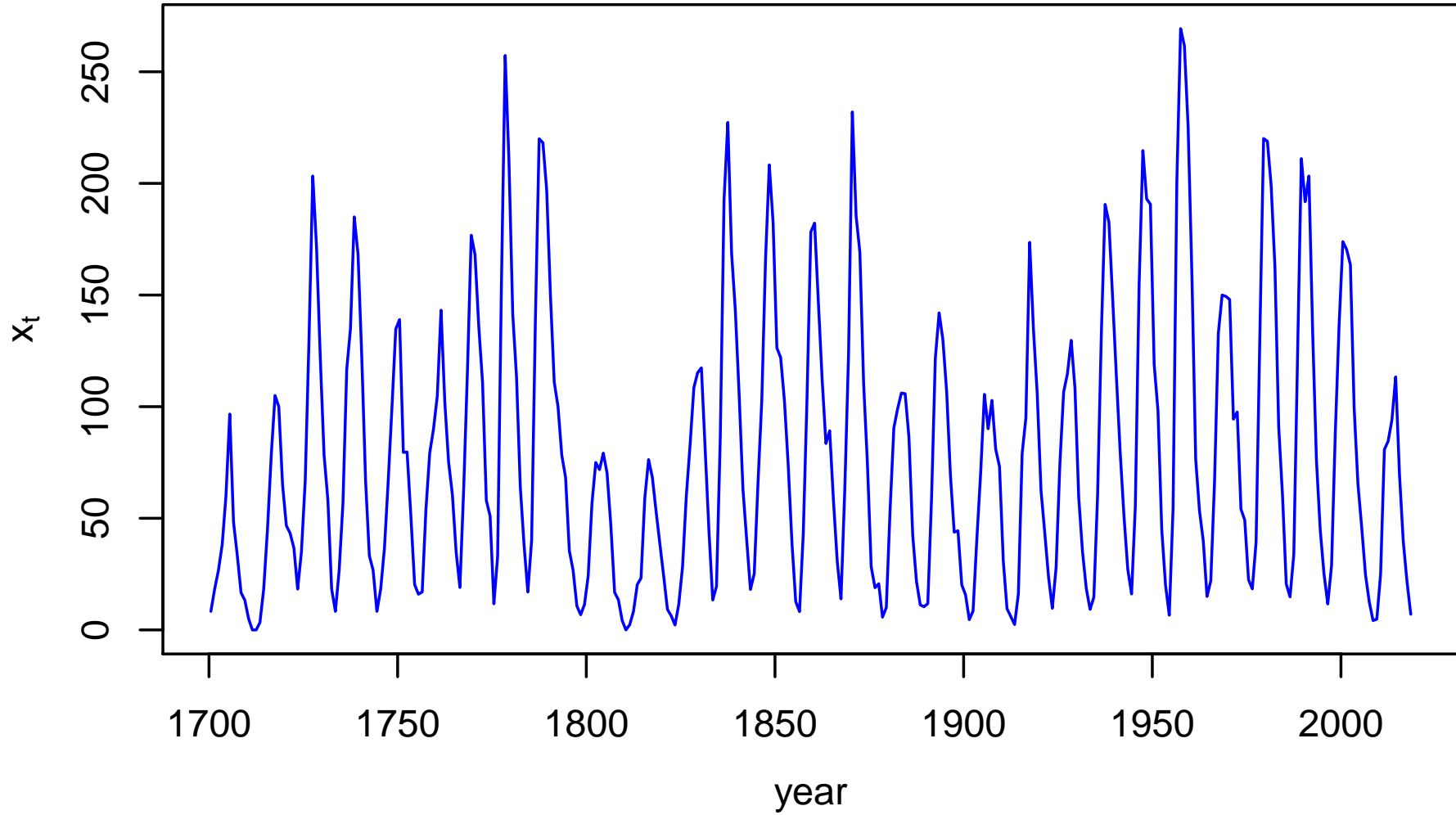
## Classification of Time Series: IV

- by predictability of  $x_t$ 
  - deterministic (predictable perfectly or to a high degree)
    - \* horizontal displacement of a pendulum in a vacuum as a function of time
    - \* tidal levels as a function of time (but not individual waves)
  - stochastic (inherent degree of unpredictability)
    - \* stock market prices
    - \* physical phenomenon about which we do not have enough theory/information to fully predict (Yule's pea shooter)
- emphasis here: discrete parameter time series  $\{x_t : t \in T\}$ , equally spaced, univariate (to start with), real-valued, stochastic in nature
- let's look at some examples of 'real-world' time series

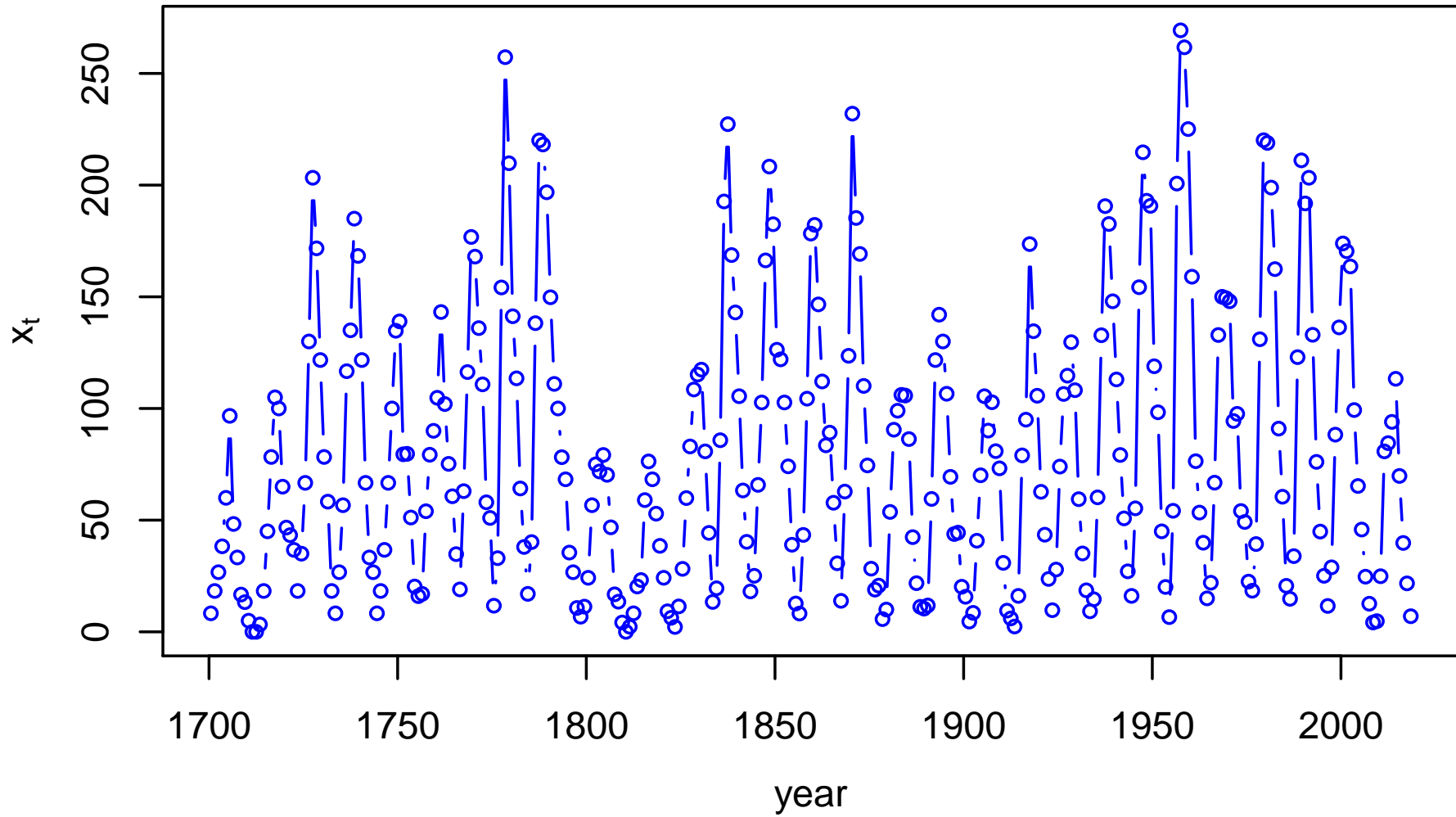
# Sunspots (1700–2018)



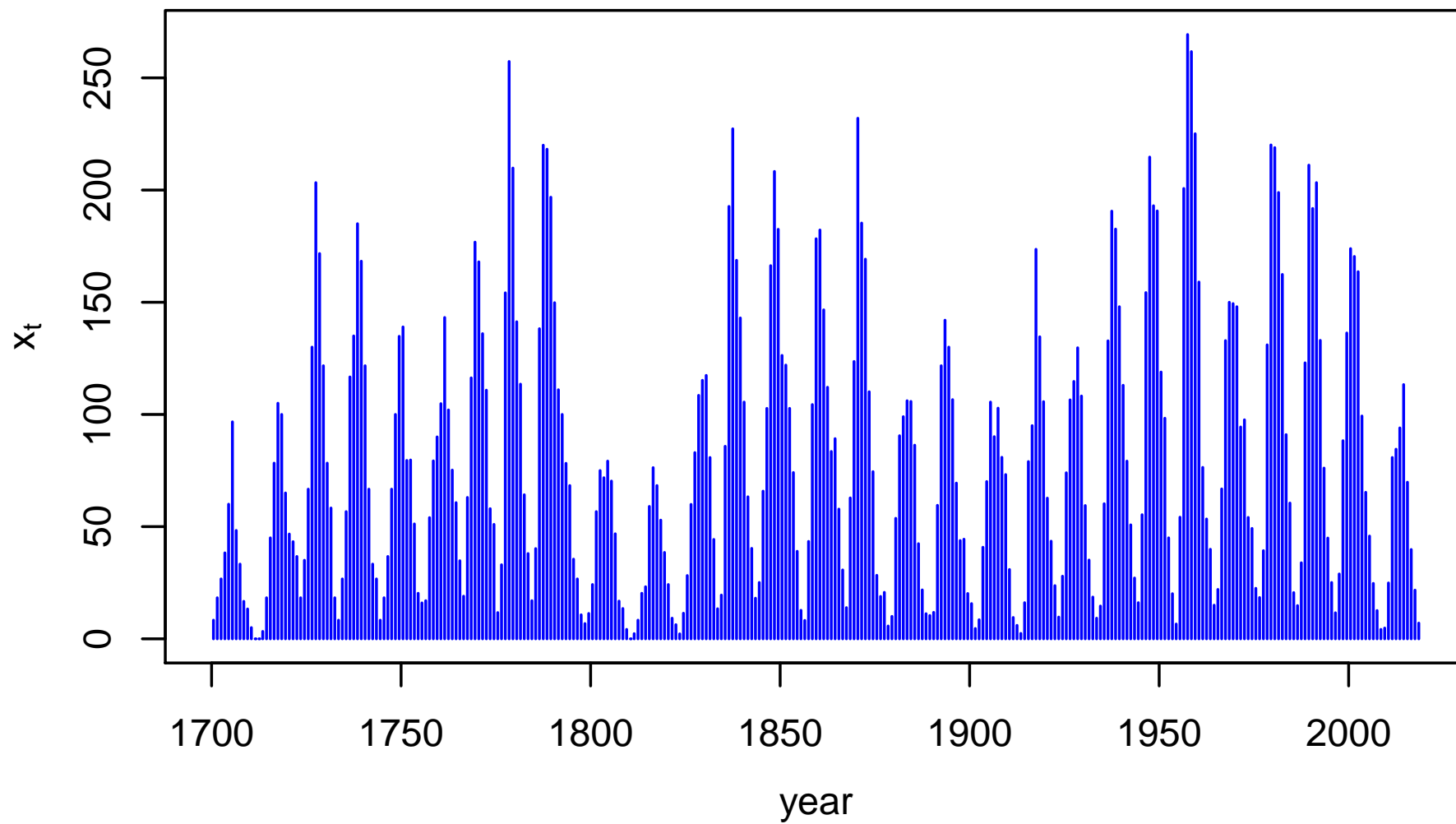
# Sunspots (1700–2018)



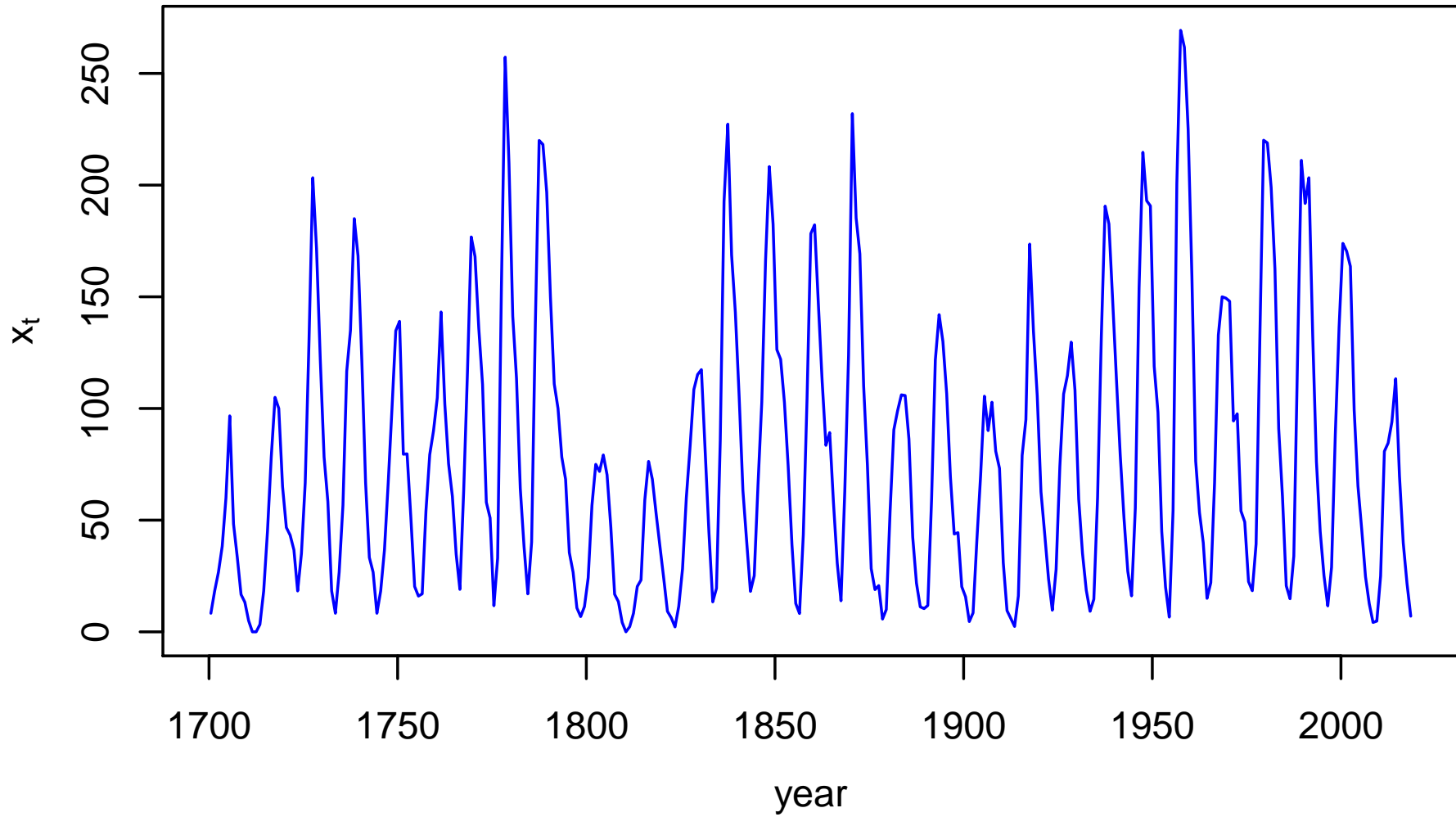
# Sunspots (1700–2018)



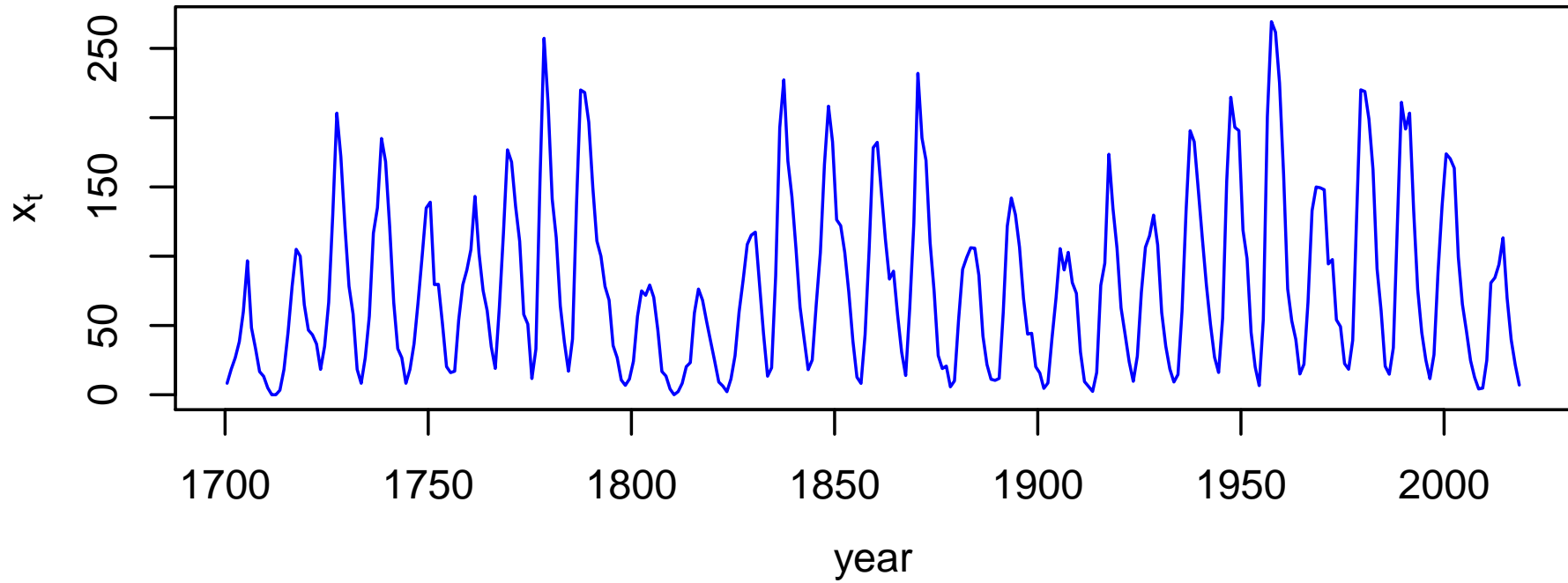
# Sunspots (1700–2018)



# Sunspots (1700–2018)



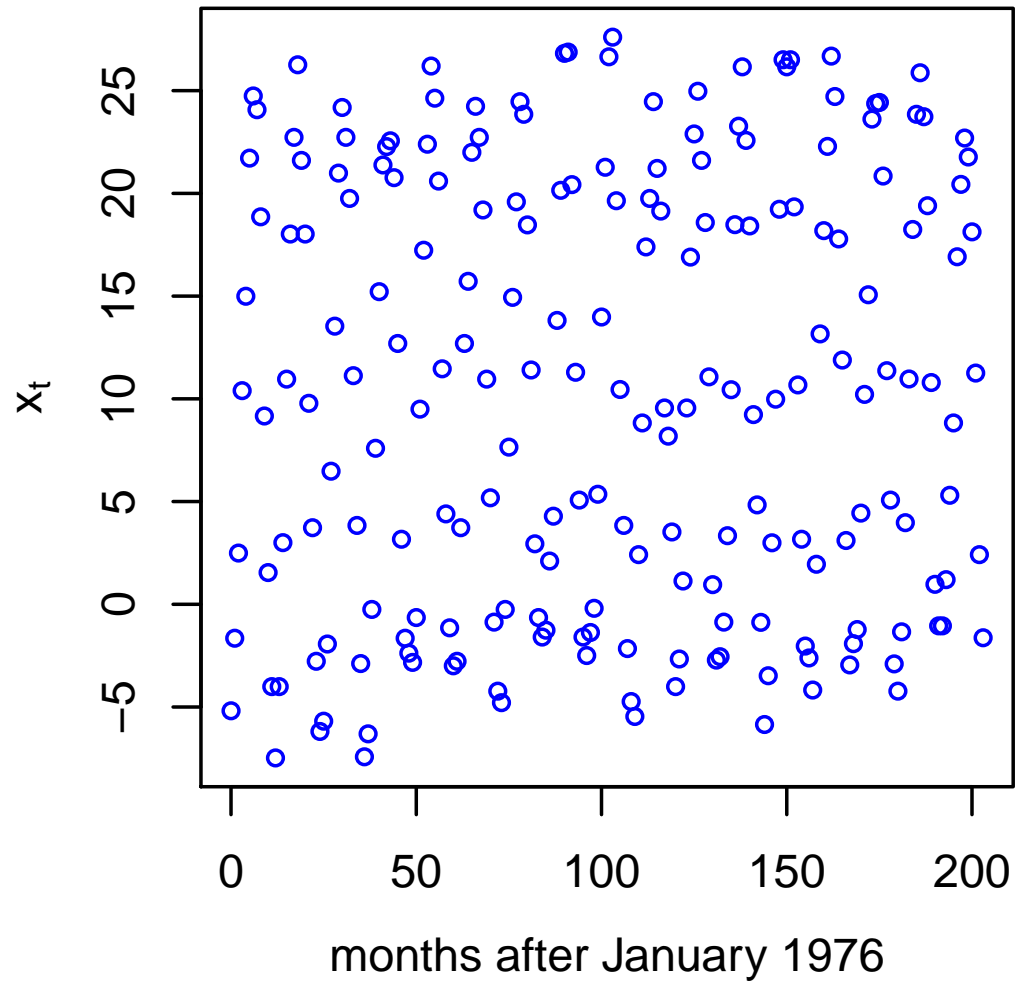
# Sunspots (1700–2018)



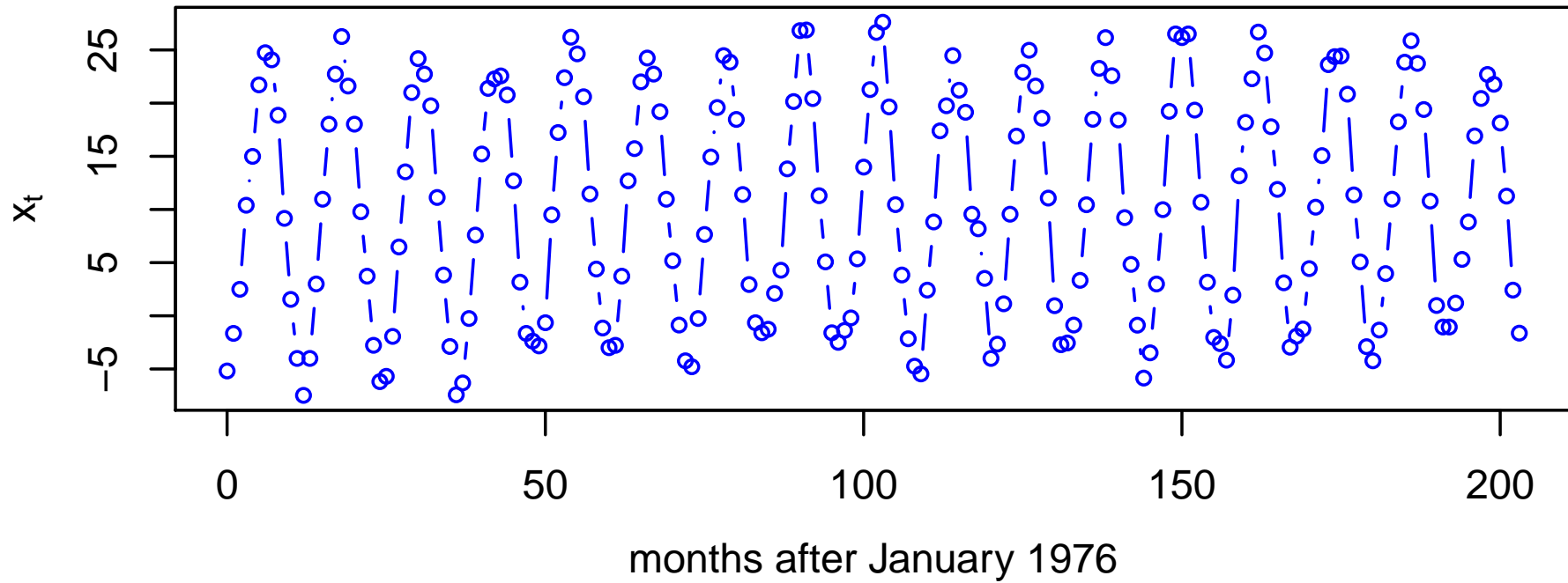
## Sunspots (1700–2018)

- index set is discrete parameter and equally spaced a year apart
- univariate time series
- $x_t$  can assume nonnegative rational values (average of daily counts)
- of interest to understand cycles for comparison with physical theories
- forecasting future values also of interest (in principle, if we knew enough about the sun, should be able to predict the number of spots)

# Monthly Soil Temperature Versus Month Count



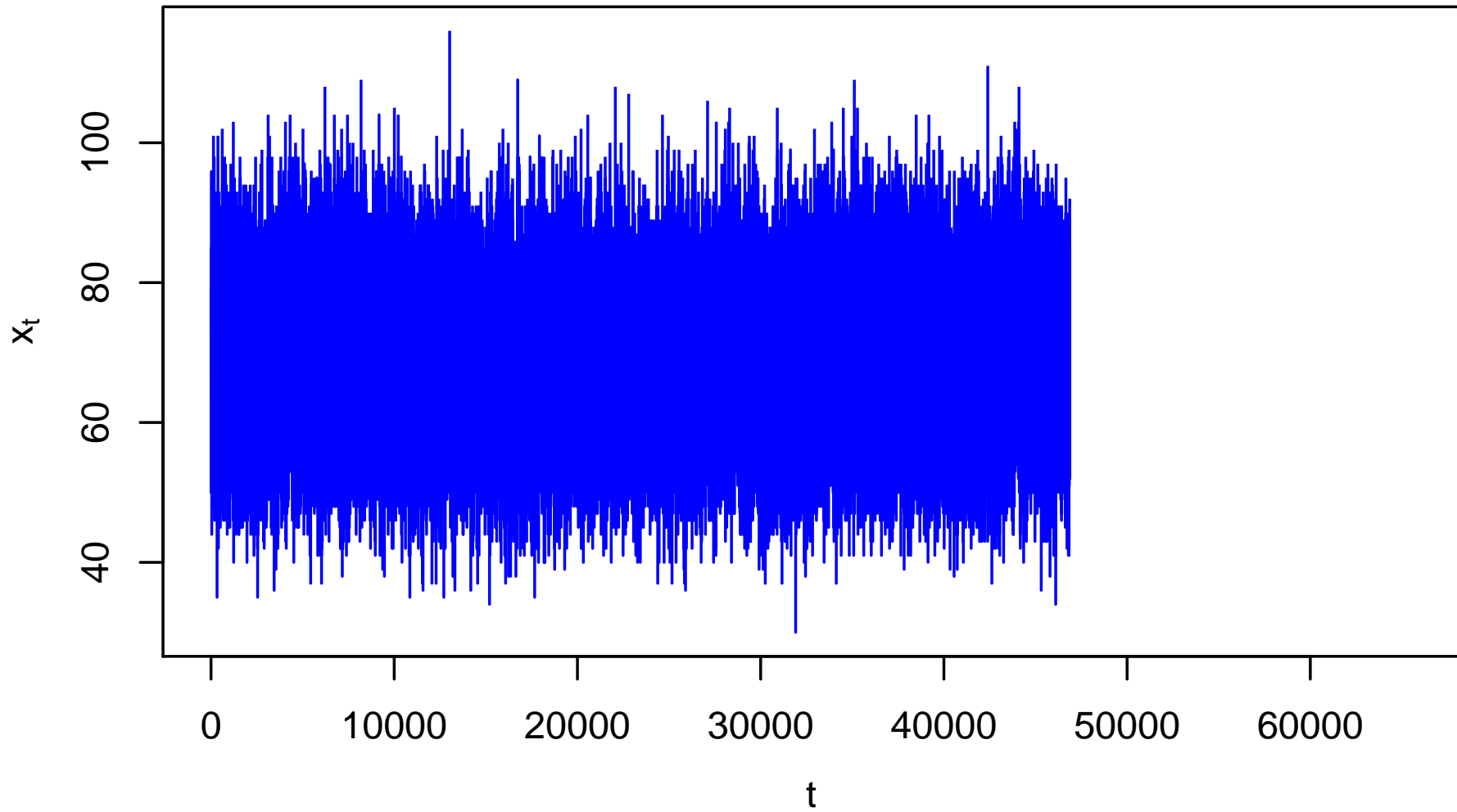
# Monthly Soil Temperature Versus Month Count



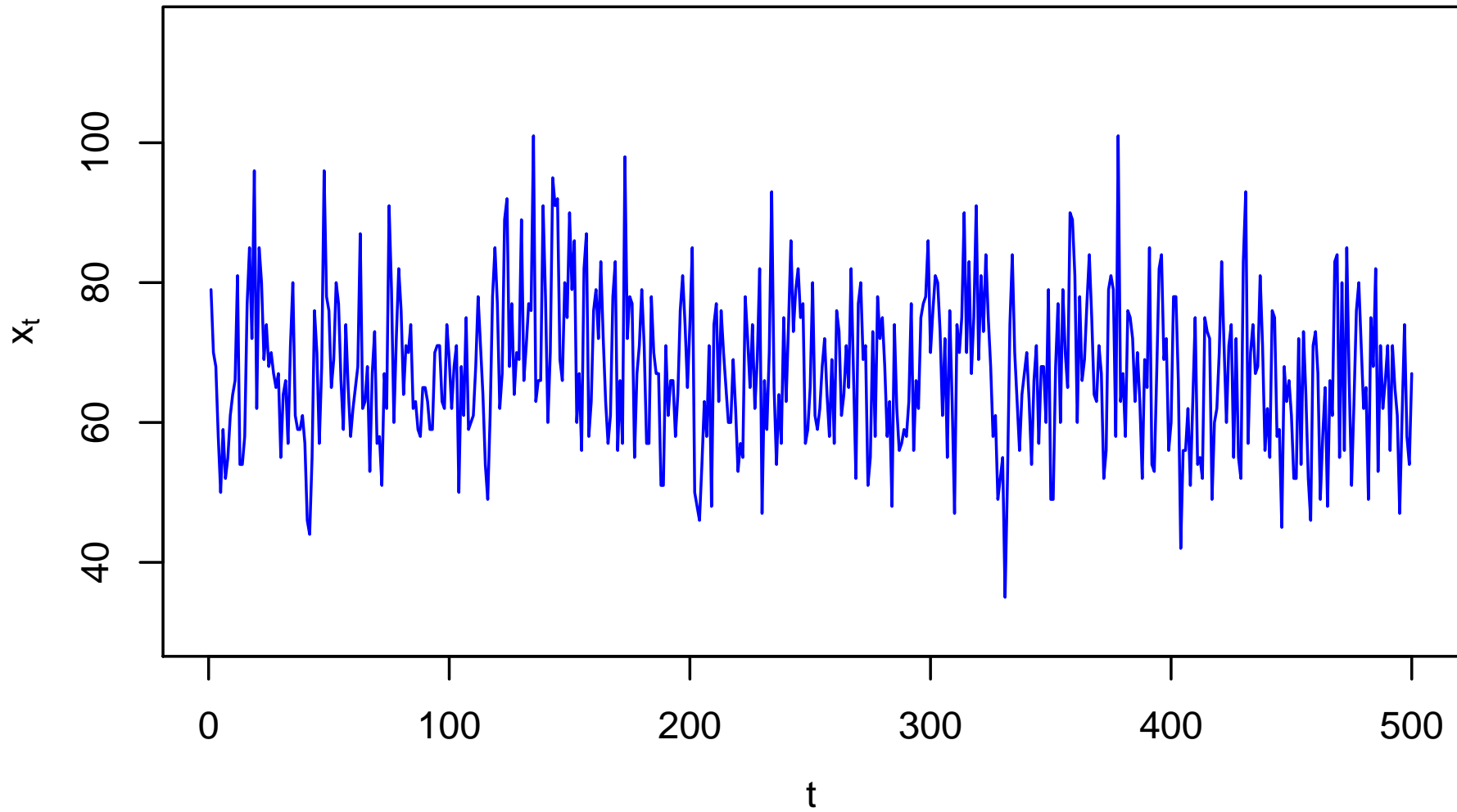
## Monthly Soil Temperatures in Mitchell, Nebraska

- index set is discrete parameter & equally spaced (ignoring fact that months have different lengths)
- univariate time series
- $x_t$  is average monthly temperature in centigrade (at depth of 20 cm) – values can't be below  $-273.15^\circ$  (absolute zero)
- strong seasonal component, but examining series for a long-term trend might be of interest

# Galactic X-Ray Source GX 5-1



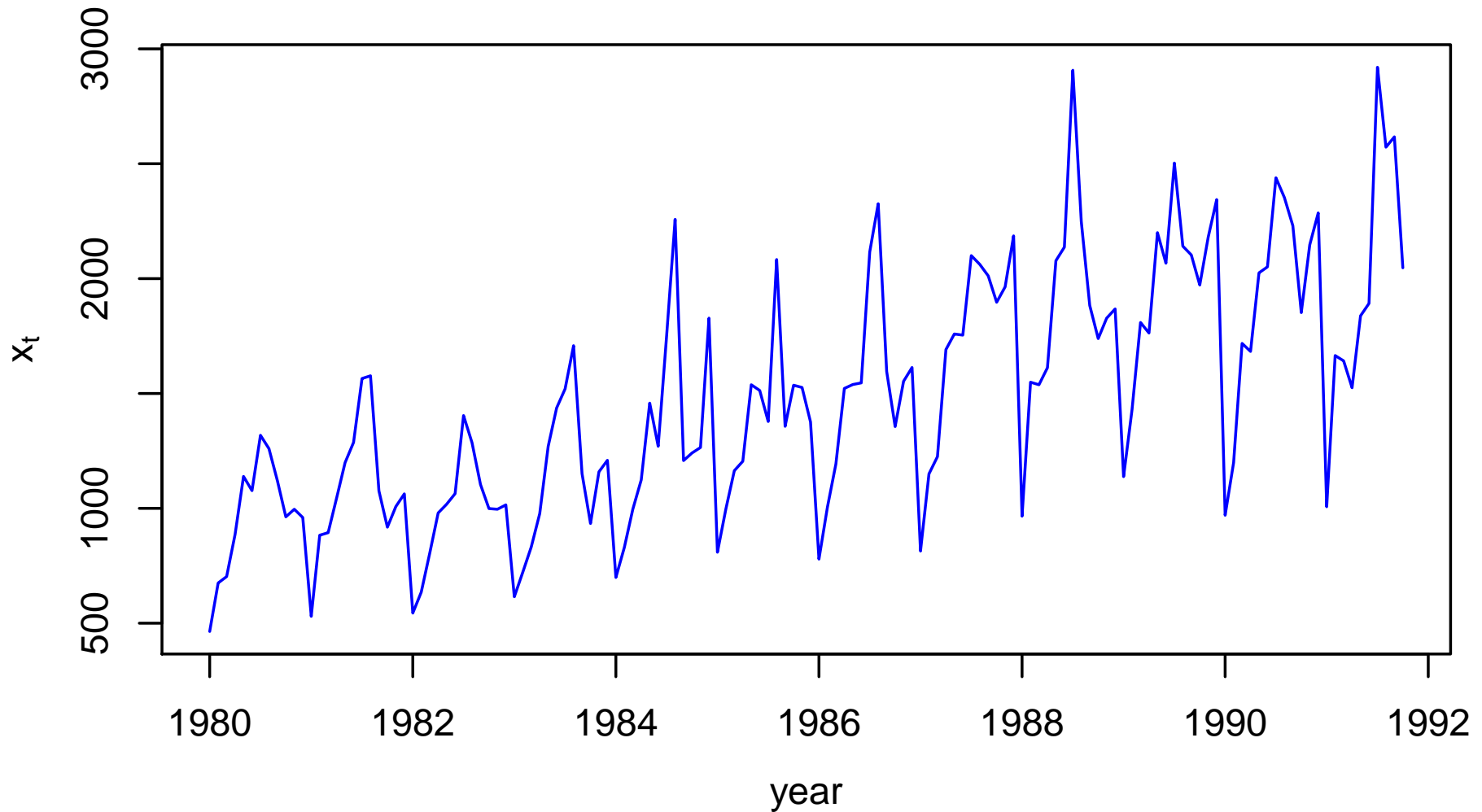
# Galactic X-Ray Source GX 5-1



## Galactic X-Ray Source GX 5–1

- index set is discrete parameter and equally spaced (128 observations per second measured by Ginga satellite observatory)
- univariate time series
- $x_t$  can assume nonnegative integer values (counts within 1/128-sec intervals)
- understanding patterns of interest for comparison with physical theories

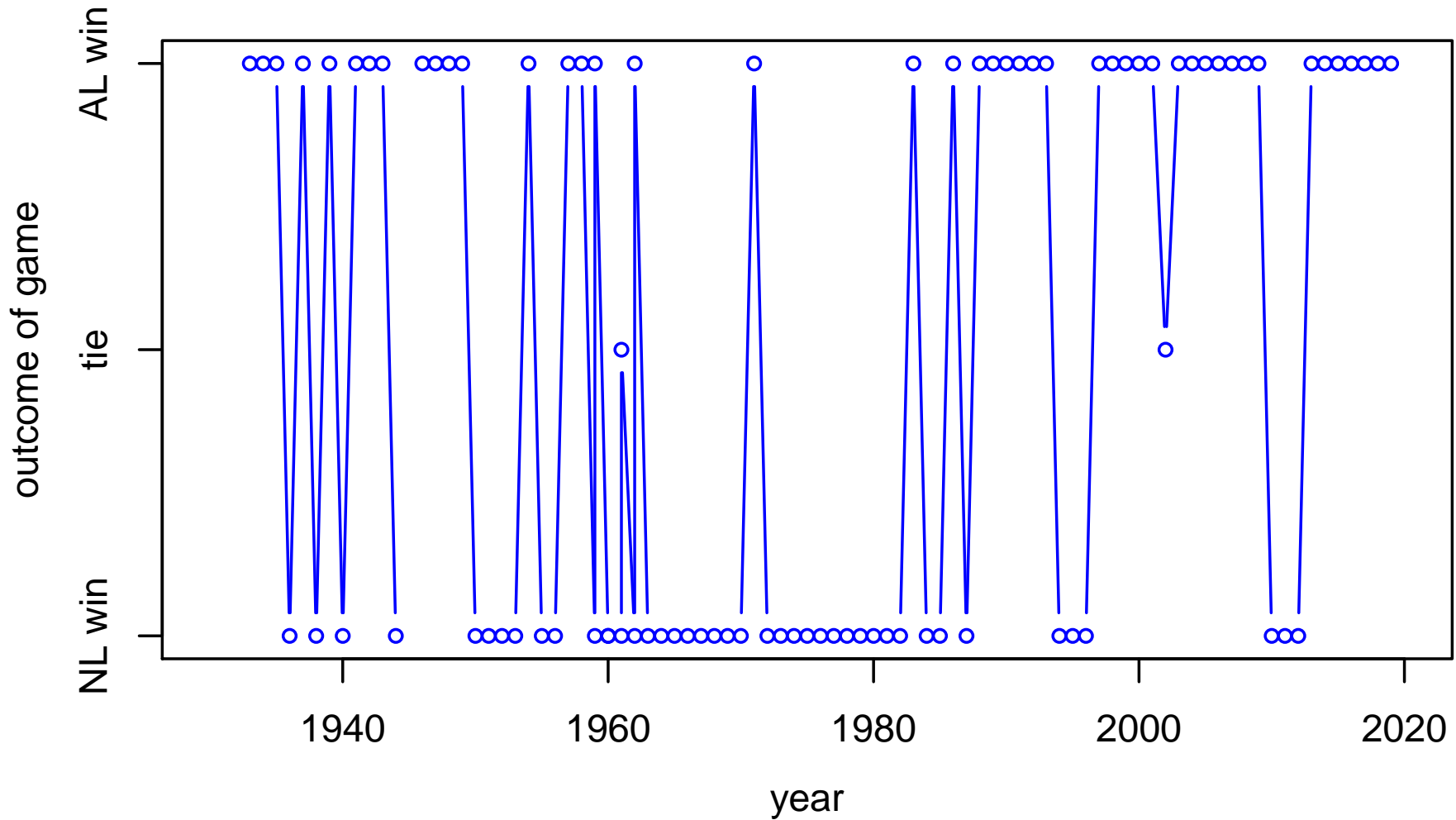
# Australian Monthly Red Wine Sales



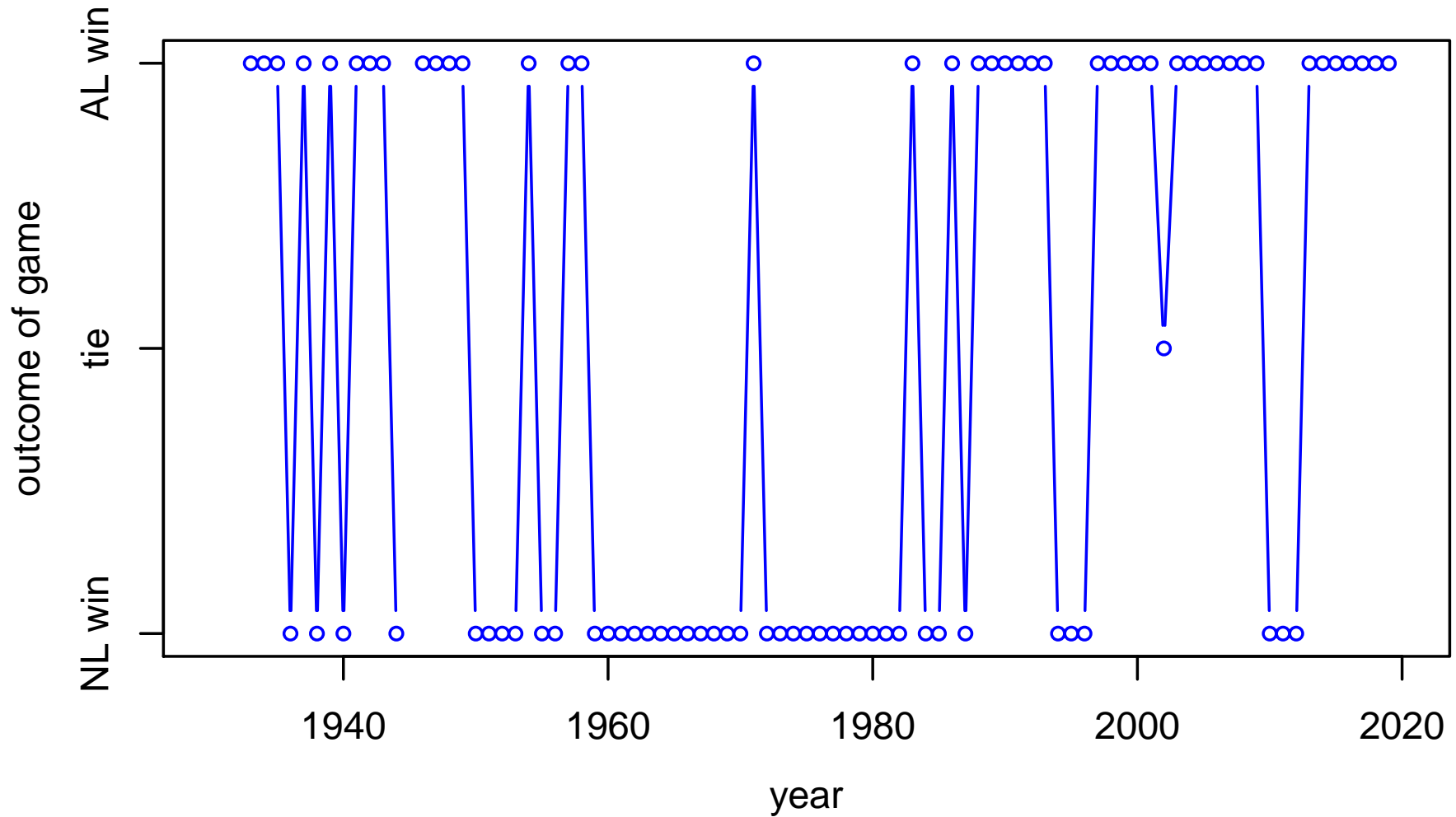
## Australian Monthly Red Wine Sales

- index set is discrete parameter and equally spaced (again ignoring fact that months don't have same number of days)
- univariate time series
- $x_t$  can assume nonnegative real values (kiloliters)
- might be of interest to compare seasonal patterns with what goes on in the Northern Hemisphere (does Christmas in summer rather than in winter matter?)
- also of interest to forecast this series (perhaps looking at external factors such as average temperature/rainfall in growing areas, economic conditions, etc.)

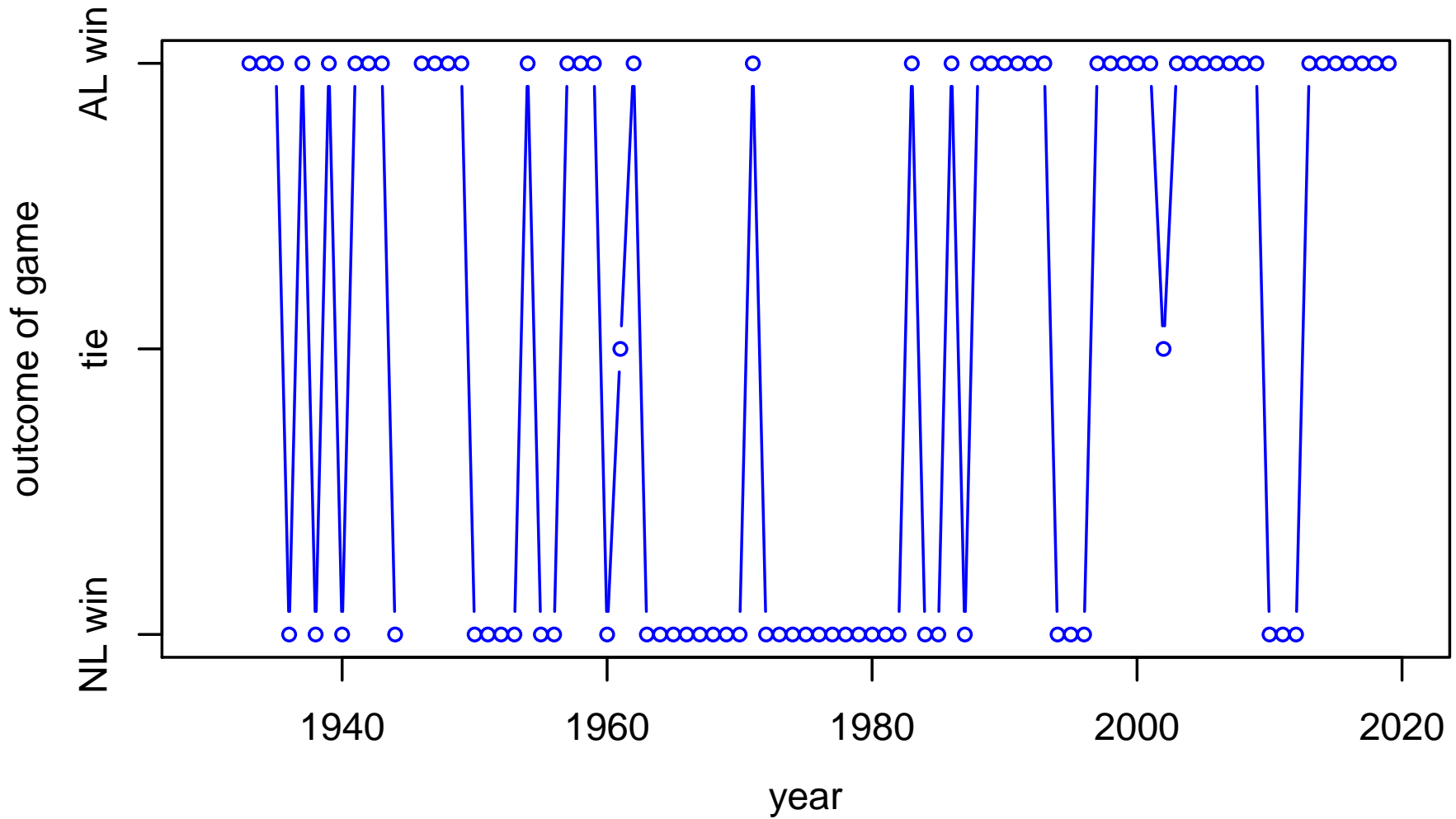
# All-Star Baseball Games (45 AL; 43 NL; 2 ties)



# All-Star Baseball Games (First in Each Year)



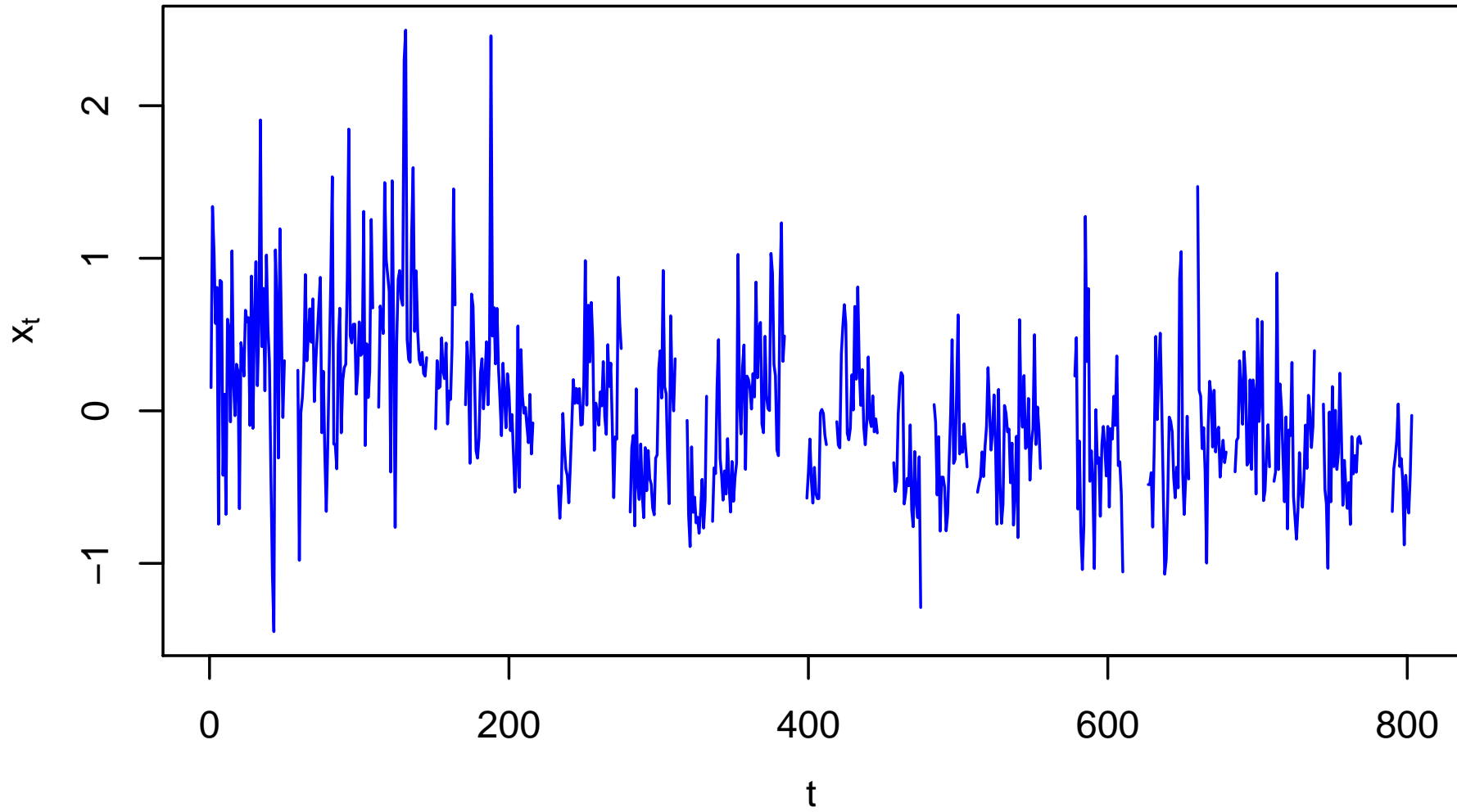
# All-Star Baseball Games (Last in Each Year)



## All-Star Baseball Games (1933–2019)

- index set is discrete parameter and equally spaced (almost!), but gappy (no game in 1945)
  - ‘almost’ because there were two games in each of 1959–1962
- univariate time series
- $x_t$  is categorical, but, if two ties eliminated, could regard as binary-valued
- 45 American League (AL) wins & 43 National League (NL) wins, but win/loss patterns seem to have streaks inconsistent with assumption of independence of outcomes (more later!)
- gamblers would be interested in forecasting this series!

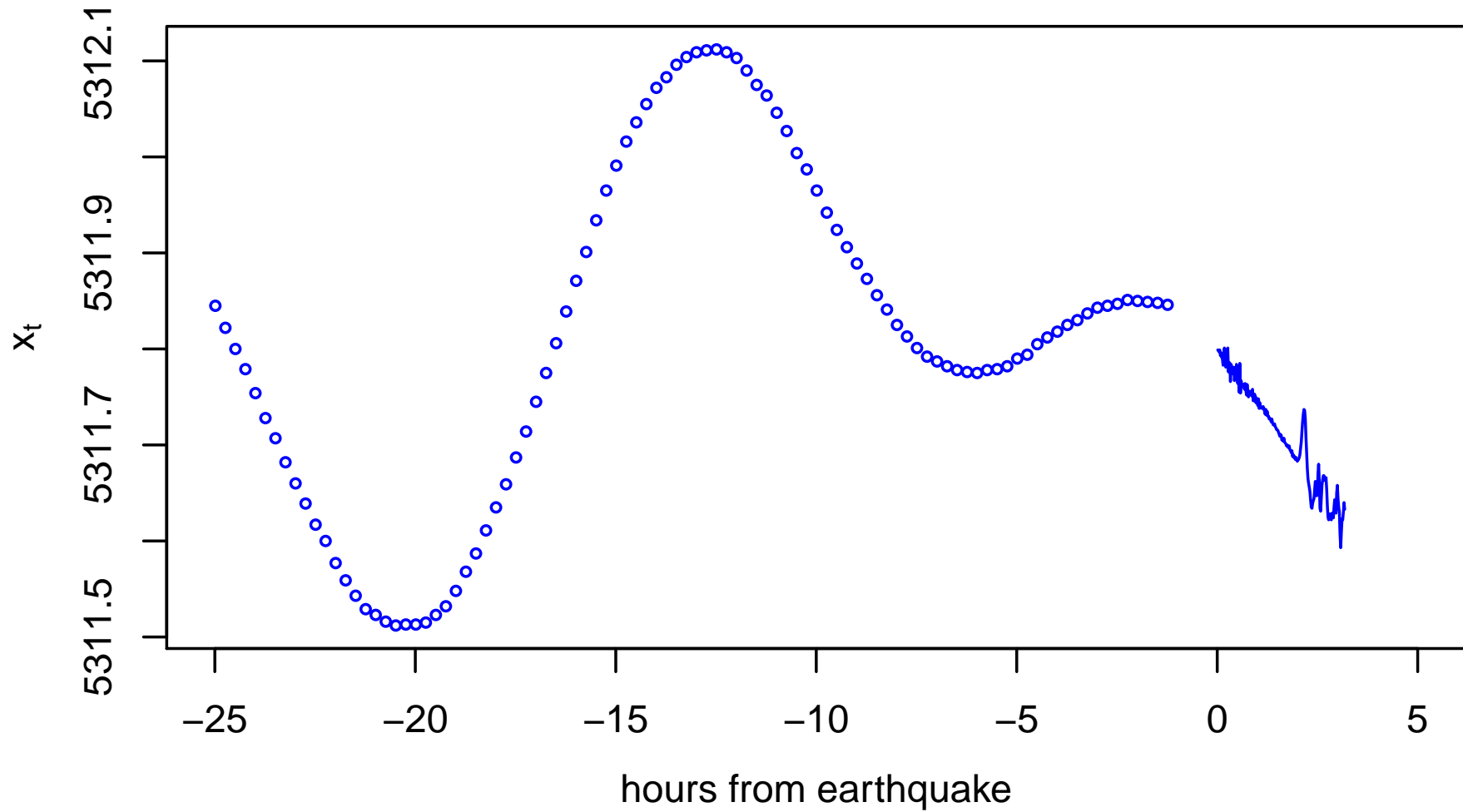
# Sea-Ice Profile



## Sea-Ice Profile

- index set is discrete parameter and equally spaced, but gappy
- univariate time series, but here ‘time’ is distance along a transect
- $x_t$  is real-valued (meters)
- characterizing roughness of series is of interest for comparison with physical models
- filling in the missing observations is also of interest

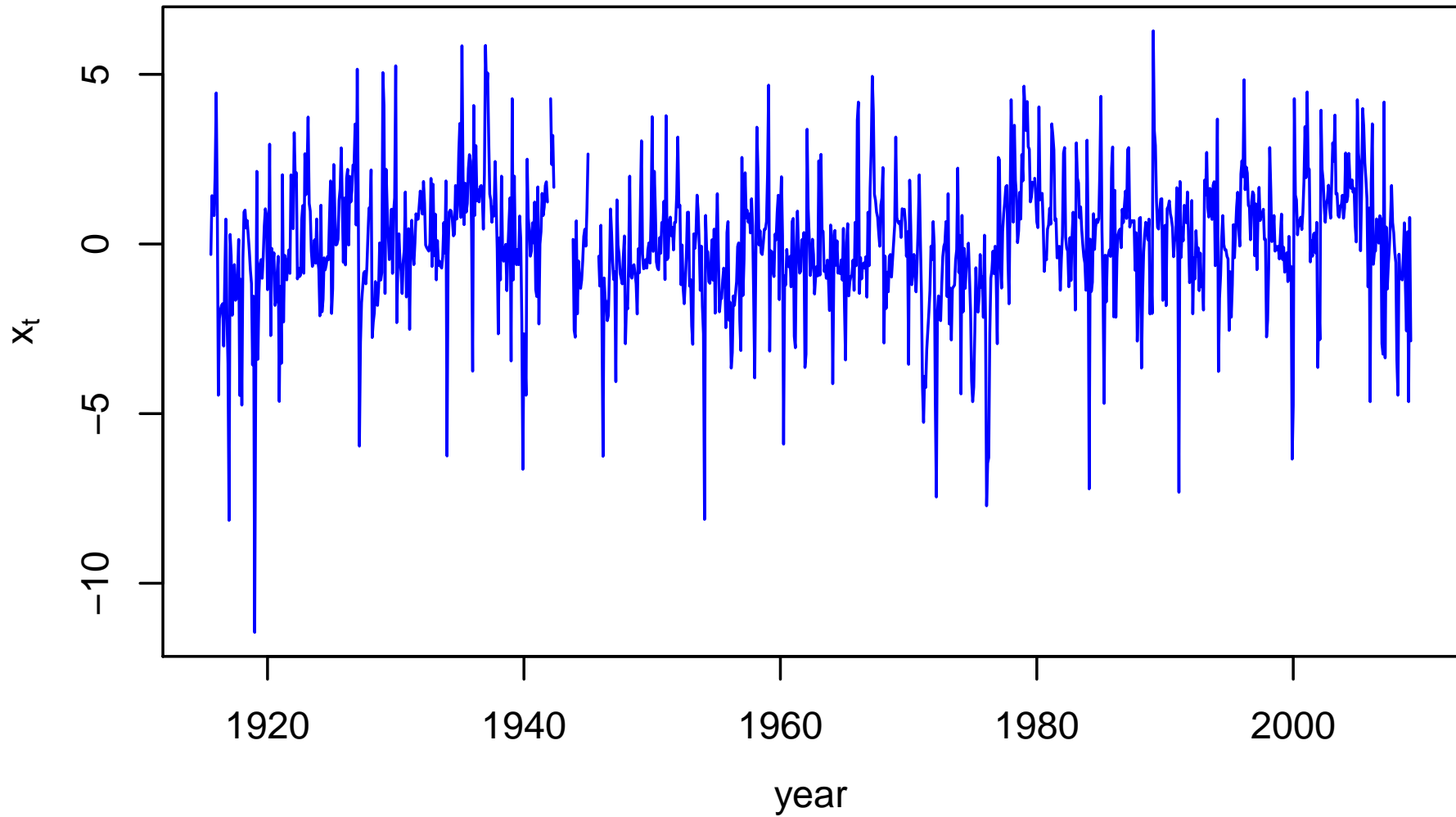
# Pressure Measurements from DART<sup>®</sup> Buoy



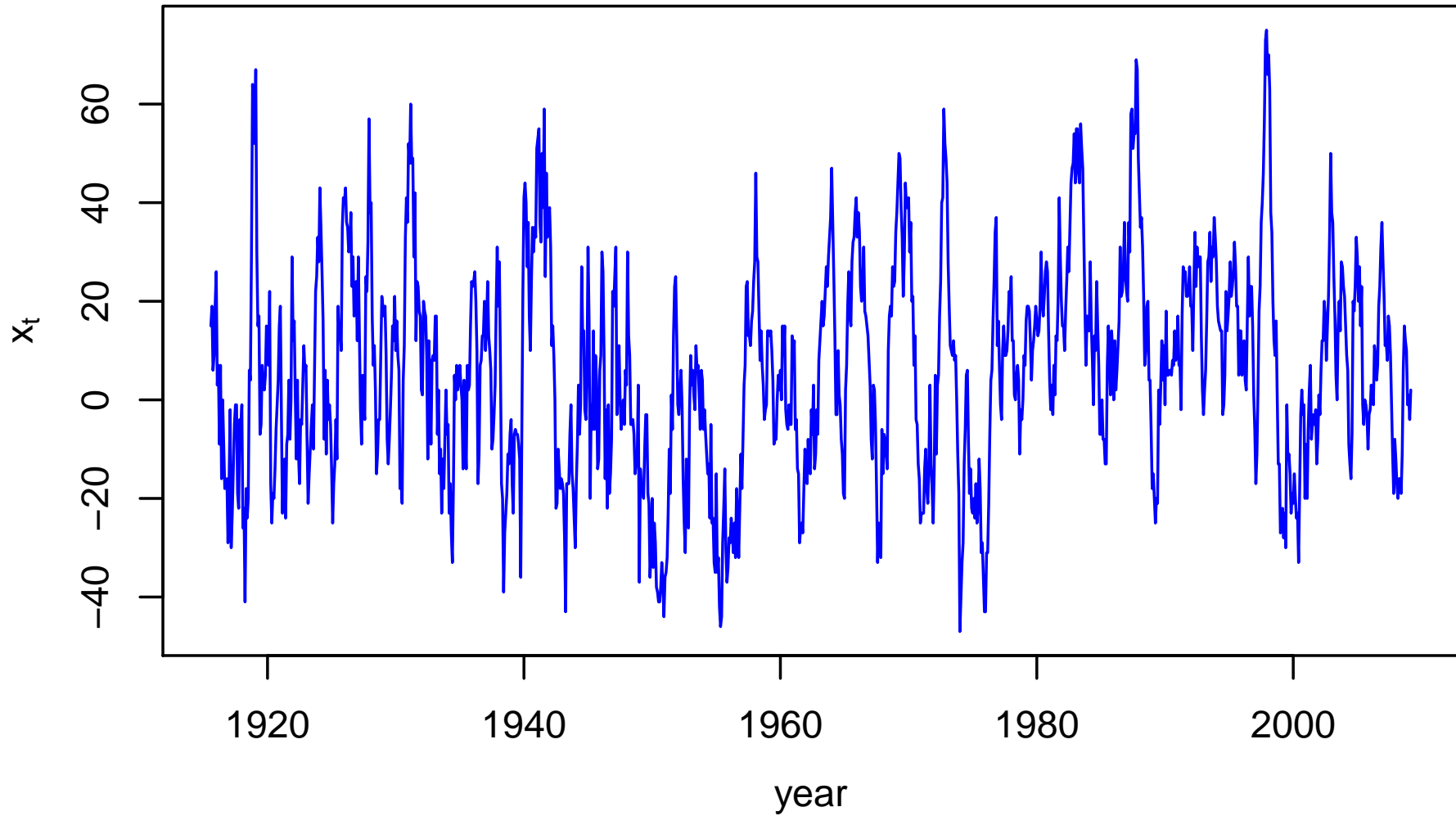
## Pressure Measurements from DART<sup>®</sup> Buoy

- index set of underlying measurements is discrete parameter and equally spaced (one value every 15 seconds), but these values only become available after a delay of up to 2 or 3 years
- one 15-sec value is available every 15 minutes during ambient conditions (four values per hour)
- when triggered by a tsunami-generating earthquake, four-point (1 minute) averages are available at 1 minute intervals
- univariate time series
- $x_t$  is real-valued (height of water column above instrument)
- of interest to remove variations due to tides ('detiding')
- after detiding, of interest to characterize nature of residuals from fits of detided data to numerical models

# St Paul Temperatures



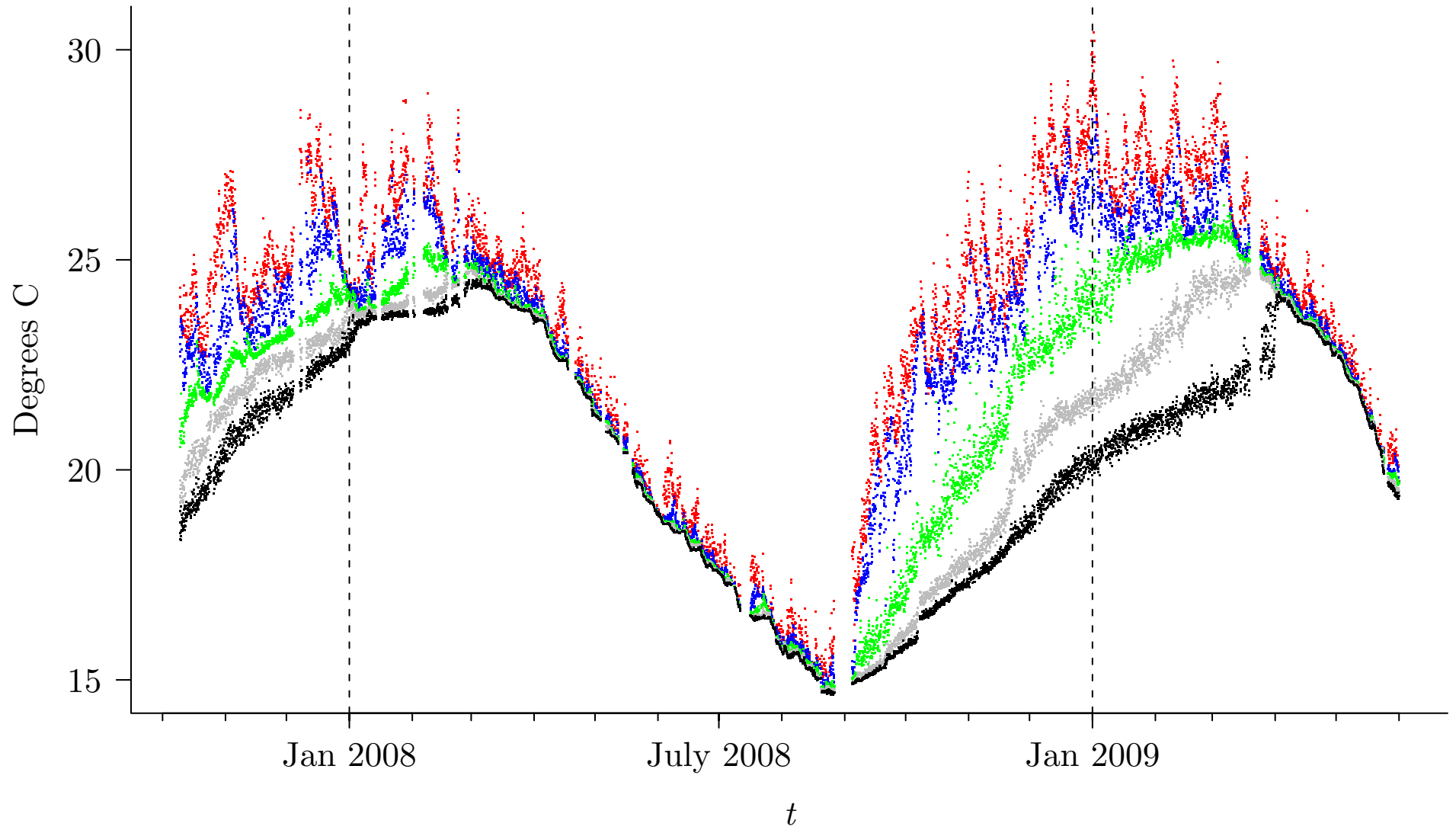
# ENSO Index



## St Paul Temperatures and ENSO Index

- St Paul is island in Bering Sea (west of mainland Alaska and north of chain of Aleutian Islands; population 479 as of 2010)
- ENSO is abbreviation for ‘El Niño/Southern Oscillation’, which is a cyclic warming/cooling of eastern and central Pacific (‘El Niño’ is oceanic component; ‘Southern Oscillation’ is atmospheric component; ENSO is coupled system)
- index set is discrete parameter and equally spaced (one value every month), but temperature observations are gappy
- bivariate time series
- $x_t$  is real-valued
- of interest to deduce the relationship between the two variates

$x_t$  at Depths of 1, 5, 10, 15 & 20 Meters



## Temperature Measurements at Wivenhoe Dam

- index set is discrete parameter and equally spaced (almost), but gappy
- multivariate time series, with depth relationship between variates
- $x_t$  can assume real values (degrees centigrade, but negative values highly unlikely!)
- daily, subannual and annual patterns are of interest for monitoring water quality
- of interest to use components of this series as covariates to predict, e.g., algae blooms

## Uses/Objectives for Time Series Analysis

- forecasting/prediction (possible due to nonindependence of adjacent values in series – not living in ‘independent and identically distributed’ (IID) world!)
- compare/contrast two series (how are they related? – e.g., input/output to some system)
- data description (how does series vary?) and exploration (is there anything unusual?)
- process control (e.g., reservoir draining in response to rainfall)
- hypothesis testing
- will advocate data description/statistical approach, but other approaches are in use (see, e.g., article on ‘technical analysis’ in Wikipedia and critique by J.A. Paulos in his book *A Mathematician Plays the Stock Market*, 2003)

## Reference

- G.U. Yule (1927), 'On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers,' *Phil. Trans. (A)*, **226**, pp. 267–98