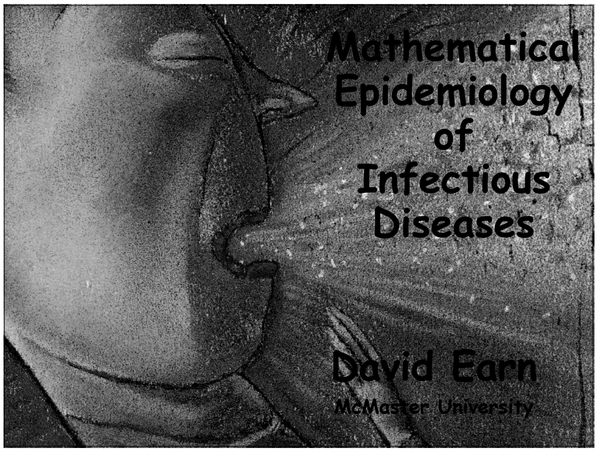
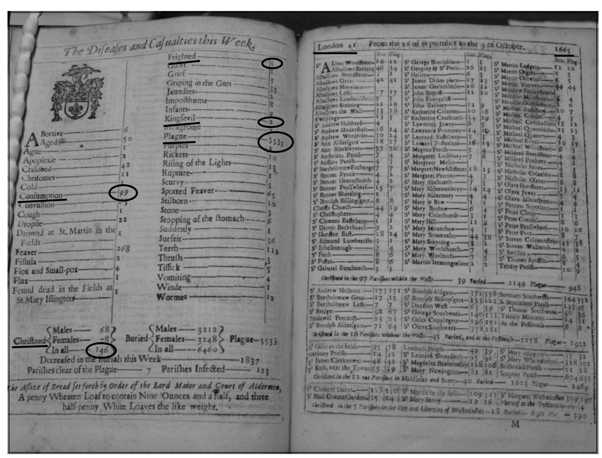
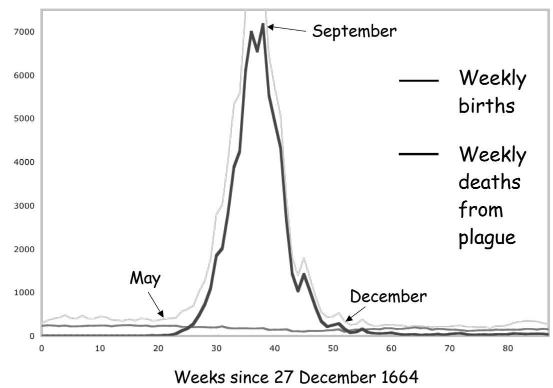


- Background
- Describing epidemics
- Modelling epidemics
- Predicting epidemics
- Manipulating epidemics

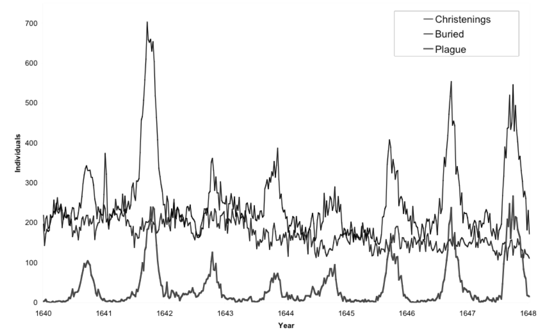
- **Background**
- Describing epidemics
- Modelling epidemics
- Predicting epidemics
- Manipulating epidemics



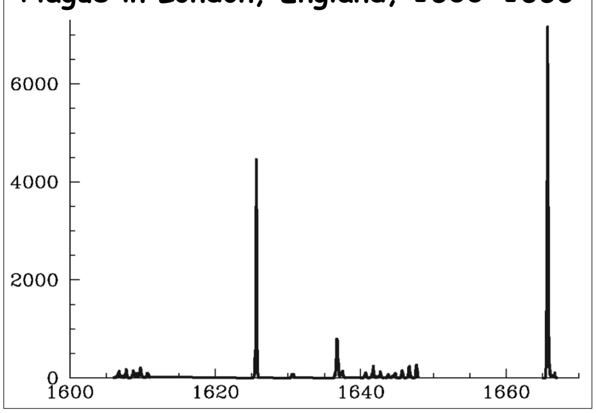
The Great Plague of London, England, 1665



Plague in London, England, 1640-1648

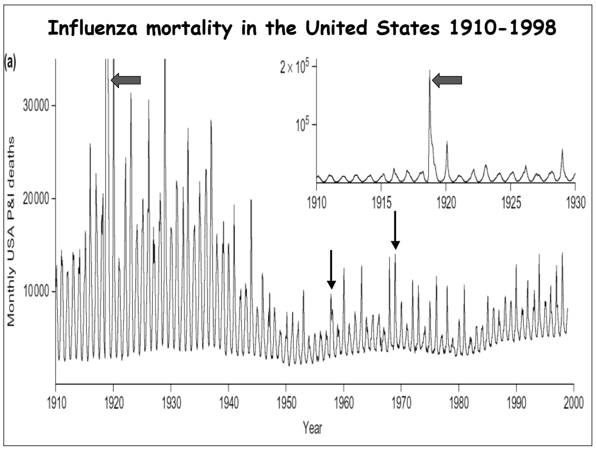


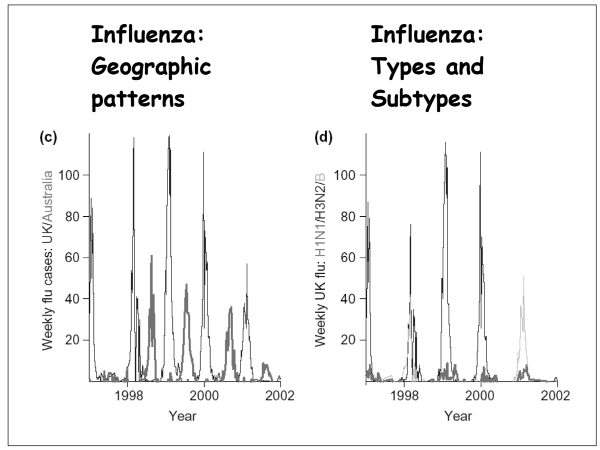
Plague in London, England, 1600-1666

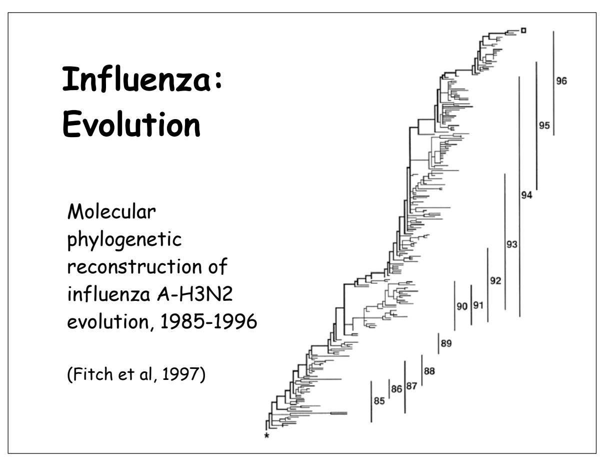


Plague Facts

- Severe (bubonic?) plague epidemics recorded from Roman times to early 1900s
- 1/3 of population of Europe killed by plague of 1348 (it took 300 years to for the population to reach the same level)
- **Spatial data** for Great Plague of 1665…
- Still a concern: rodent reservoir, antibiotic-resistant strains…


Influenza mortality in the United States 1910-1998
(a)
Monthly USA P&I deaths
Year

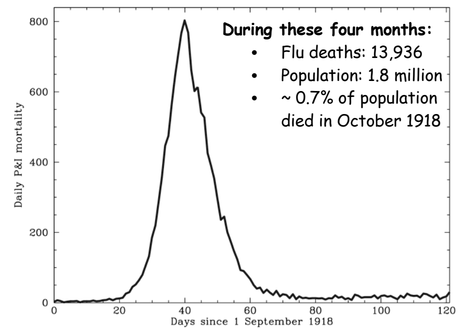

Influenza: Geographic patterns
Influenza: Types and Subtypes
(c)
(d)
Weekly flu cases: UK/Australia
Weekly UK flu: H1N1/H3N2/B
Year
Year


Influenza: Evolution
Molecular phylogenetic reconstruction of influenza A-H3N2 evolution, 1985-1996
(Fitch et al, 1997)

Flu Facts

- Annual influenza epidemics are a major stress on healthcare systems worldwide
- ~30,000 deaths in the United States attributed to influenza every year (mostly in people >65 years old)
- Individual-level mortality data available for the US since 1979, in Canada since 1951
- Can never be eradicated: reservoir in aquatic birds… constant threat of new emergence/pandemic…

Daily influenza mortality in Philadelphia, Sept to Dec 1918

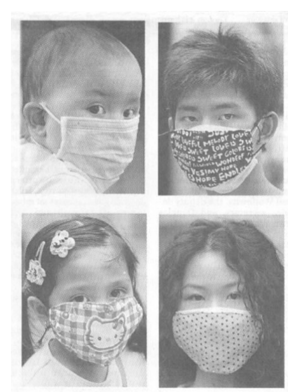


Pandemic Flu Facts

- 20-100 million deaths in 1918 pandemic

- Less severe pandemics in 1957 and 1968
- A new flu pandemic could occur any time

What next?



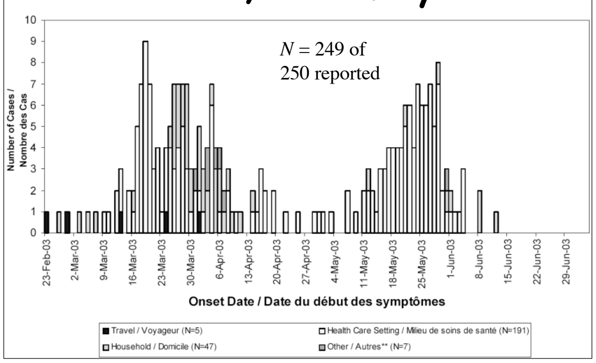
The answer lies here…

Metropole Hotel (Hong Kong)

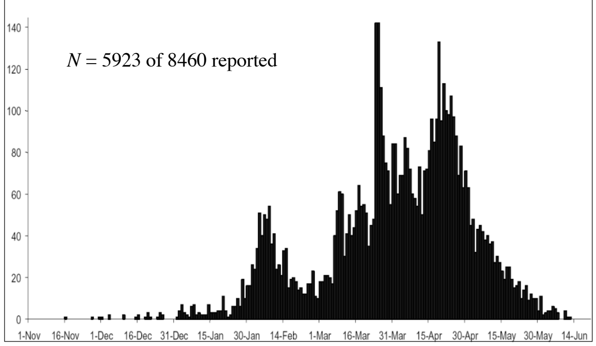
And here…

Downtown Toronto, April 2003

Daily SARS cases in Canada by onset date, to 4 July 2003



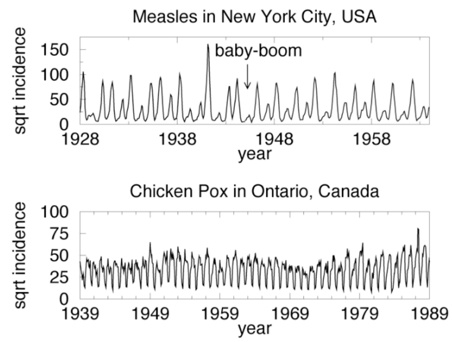
Daily SARS cases worldwide by onset date, to 16 June 2003



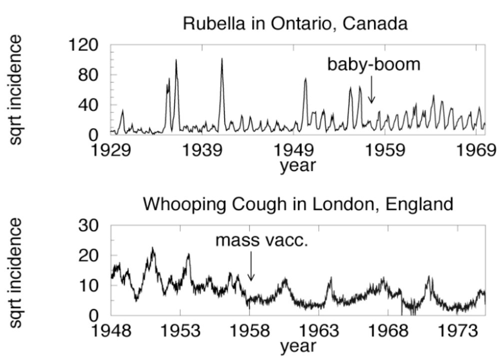
SARS Facts

- **High case fatality rate**
  - 1918 flu: < 3%
  - SARS: > 10% (and most others require hospitalization)
- **Long hospital stay times**
  - Mean time from admission to discharge or death ~25 days in Hong Kong (Donnelly et al 2003)
- **As of 26 September 2003:**
  - 8098 probable cases, 774 deaths

Childhood Diseases

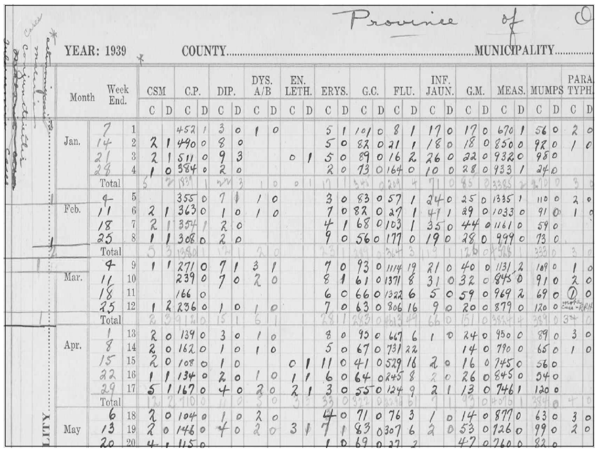


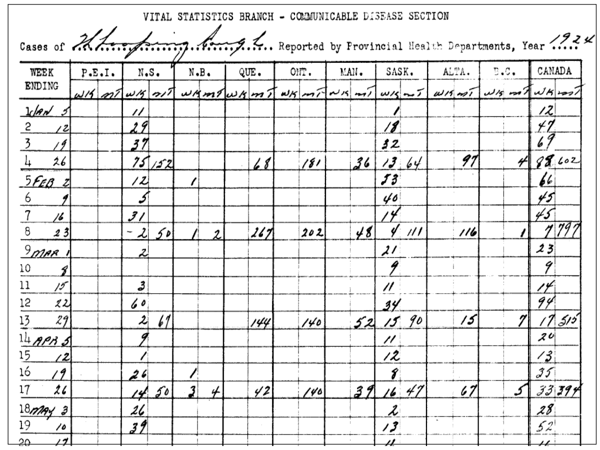
More Childhood diseases…



Canadian Data Sources

- **Ontario Ministry of Health**
  - Weekly notifications, 1939-1989, aggregated for the whole province (county and municipality level spreadsheets were destroyed in the 1990s, except for two years)…
- **Manitoba Health**
  - DBS weekly notifications spreadsheet for 1958
- **Statistics Canada**
  - Weekly/monthly notifications, by province, since 1924
  - Incomplete in parts, primarily because notification practices varied over time and space
- **Tiny bits from Quebec, Alberta, BC**





Childhood Disease Facts (Definition)

- **Short incubation period**
  - << 10 years (typically a few days)
- **High transmission rate**
  - mean age at infection is during childhood
- **Lifelong immunity**
  - Can ignore evolution of pathogen
  - Vaccine does not need to be updated
  - Eradication possible in principle (if no non-human reservoir)

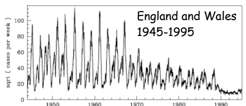
Measles Facts

- 30-40 million cases and ~750,000 measles deaths occur each year.
- Measles accounts for 46% of the 1.7 million annual deaths due to vaccine-preventable diseases.

http://www.unicef.org/measles/factsheet.htm
http://www.measlesinitiative.org/

- Costs ~$1 per measles vaccination.

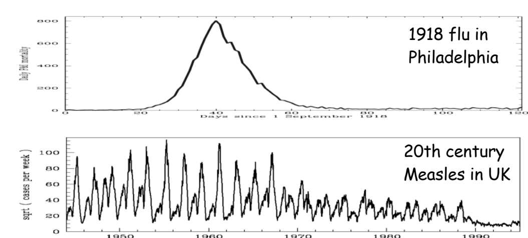
Measles Epidemics



- Understand past patterns
- Predict future patterns
- Manipulate future patterns
- Develop vaccination strategy that can…

- Background
- **Describing epidemics**
- Modelling epidemics
- Predicting epidemics
- Manipulating epidemics

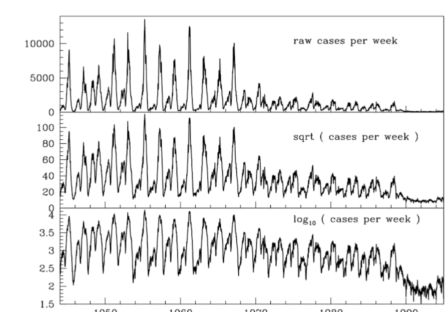
The Time Plot



- For single epidemic: excellent descriptor
- For recurrent epidemics: still best starting point, but often hard to interpret

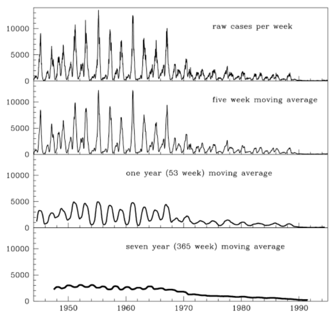
Time Plots of transformed data

- Reveal some hidden aspects of time series



Time Plots of smoothed data

- Reveal trends clouded by noise or seasonality



Moving average

$$X_t \rightarrow \frac{1}{2a+1}\sum_{i=-a}^{a} X_{t+i}$$

General linear filter

$$X_t \rightarrow \sum_{i=-\infty}^{\infty} \lambda_i X_{t+i}$$

Autocorrelation

- Recall usual **correlation coefficient** for pairs of observations

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

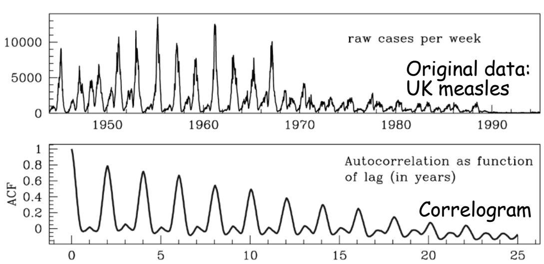
- **Autocorrelation at lag *k*** is correlation coefficient for observations that are k steps apart in time

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

- Sheds light on nature of serial dependence in a time series

Correlogram

- Plot of autocorrelation as a function of lag



- Peaks in the correlogram correspond to periodicities in the original time series

Spectral density

- Another way to identify periodicities in the data
- Suppose we express the data as a Fourier series:

$$x_t = a_0 + \sum_{p=1}^{(N/2)-1} [a_p \cos\omega_p t + b_p \sin\omega_p t] + a_{N/2}\cos\pi t$$

where $\omega_p = \frac{2\pi p}{N}$

- Then $a_0 = \bar{x}$

$$a_p = \frac{2}{N}\left[\sum x_t \cos\omega_p t\right]$$

$$b_p = \frac{2}{N}\left[\sum x_t \sin\omega_p t\right]$$

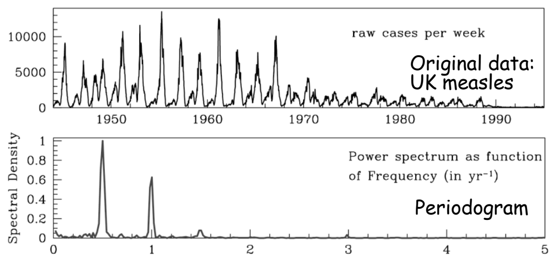
$$a_{N/2} = \tfrac{1}{N}\sum (-1)^t x_t$$

- The estimated ***power spectral density at frequency*** ω_p is:

$$I(\omega_p) = \frac{N}{4\pi}\left(a_p^2 + b_p^2\right)$$

Periodogram

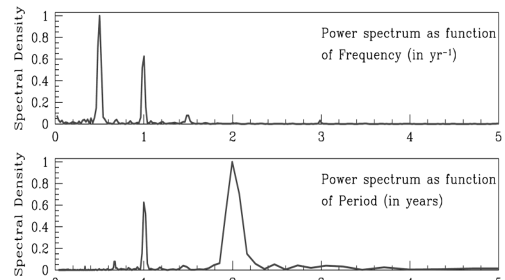
- Plot of estimated spectral density as function of frequency ω_p



- Peaks in the periodogram correspond to periodicities in the original time series

Period Periodogram

- Often helpful to plot periodogram as a function of period

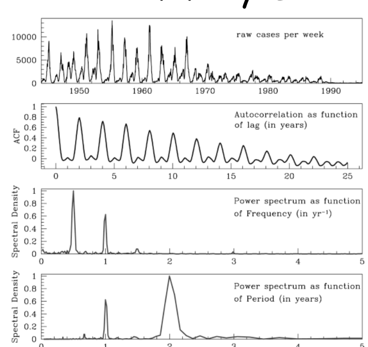


Properties of Periodogram

- Periodogram is discrete Fourier transform of correlogram
  - Same information in autocorrelation and power spectrum
  - Periodogram generally easier to interpret than correlogram
  - Convenient to calculate periodogram via correlogram
- Total area under the Periodogram is equal to variance of the time series
  - $I(\omega_p)$ is the proportion of the variance associated with ω_p
- Periodogram is really an estimator of the "true" (continuous) power spectrum
  - Precision of autocorrelation coefficients decreases with lag because series is finite
  - When Fourier transforming correlogram, commonly linearly filter with, e.g., *Tukey window:*

$$\lambda_k = \frac{1}{2}\left(1+\cos\frac{\pi k}{M}\right) \qquad k = 0,1,\ldots,M$$

Summary Description



Raw data

Correlogram

Frequency Periodogram

Period Periodogram

Problem

- Examine measles dynamics in New York City (data file `nycmeas.dat`) via:
  - Time plots of raw and transformed data
  - Autocorrelation / correlogram
  - Power spectral density / periodogram
- Apply these methods to segments of the time series that look different by eye
  - Is there evidence for: Frequency components not evident by eye? Changes in the frequency structure over long time scales? If so, why might this have occurred?
- What if you remove trend and/or seasonality?

When you have time…

- More sophisticated spectral methods exist.
- ***Wavelet analysis*** provides a method for frequency decomposition that is local in time, so you can see changes in the spectrum over time without having to identify distinct temporal segments yourself.
- If you're ambitious… explore the New York City measles time series using wavelet analysis, e.g., via the matlab wavelet toolbox.