3 Epidemic Data and Time Series Tools

Mathematics and Statistics

$$\int_M d\omega = \int_{\partial M} \omega$$

Mathematics 4MB3/6MB3
Mathematical Biology

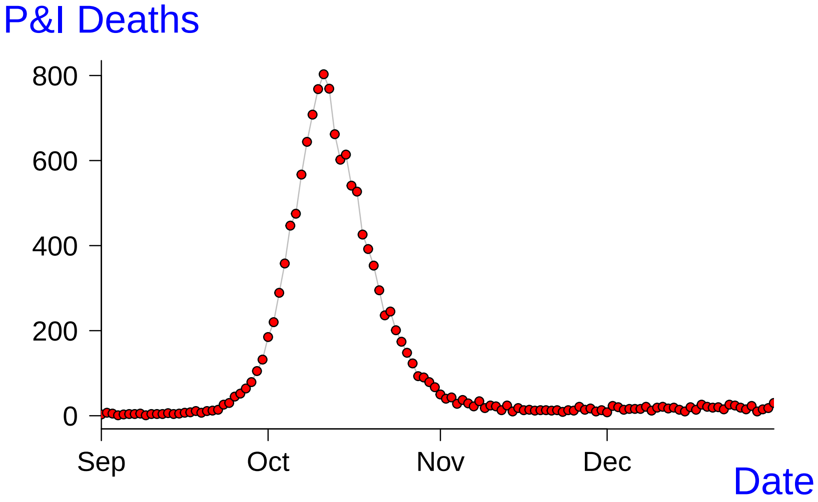
Instructor: David Earn

Lecture 3
Epidemic Data and Time Series Tools
Tuesday 17 September 2024

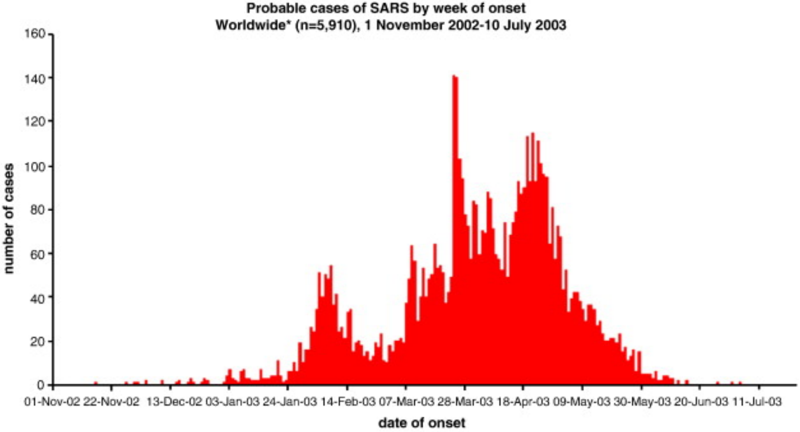
Announcements

- Next week's lecture will be recorded in advance and posted on the Echo 360 page for this course.
  - Live Q&A, either in last hour of scheduled class or at a mutually convenient time.
- Assignments:
  - Assignment 1 due 23 Sep 2024 (next Monday)
  - Assignment 2 due 7 Oct 2024
    (good to work on before class on 1 Oct 2024)
- Class on 1 Oct 2024 will be given by Mikael Jagan
  (install `epigrowthfit` before that class)
- Lecture on 8 Oct 2024 will pre-recorded and posted

P&I Mortality, Philadelphia, 1918

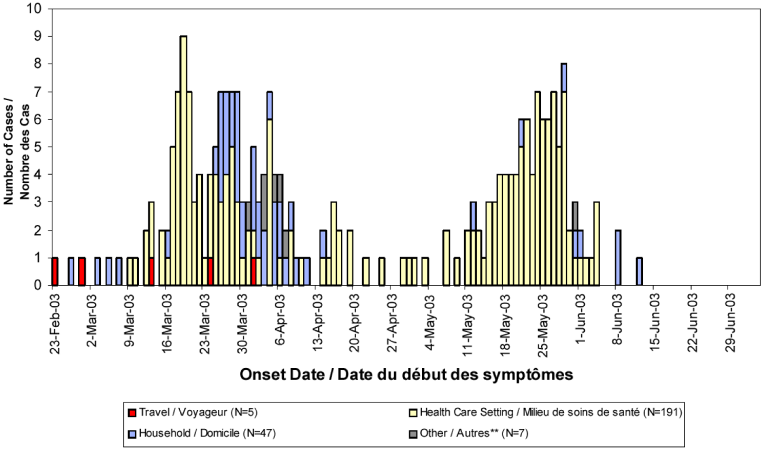


SARS in 2003 (Worldwide)



*This graph does not include 2,527 probable cases of SARS (2,521 from Beijing, China), for whom no dates of onset are currently available.

SARS in 2003 (Toronto)



$N = 249$ (of 250 reported)

Some SARS Facts

- High case fatality
  - 1918 flu $< 3\%$
  - SARS $> 10\%$
- Long hospital stays
  - Mean time from admission to discharge or death: ~ 25 days in Hong Kong
- 8098 probable cases, 774 deaths
- How bad would it have been if it had not been controlled?

COVID (ancestral) hospitalization and survival in Ontario

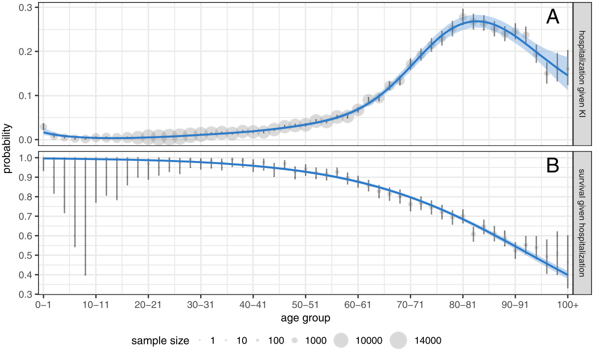
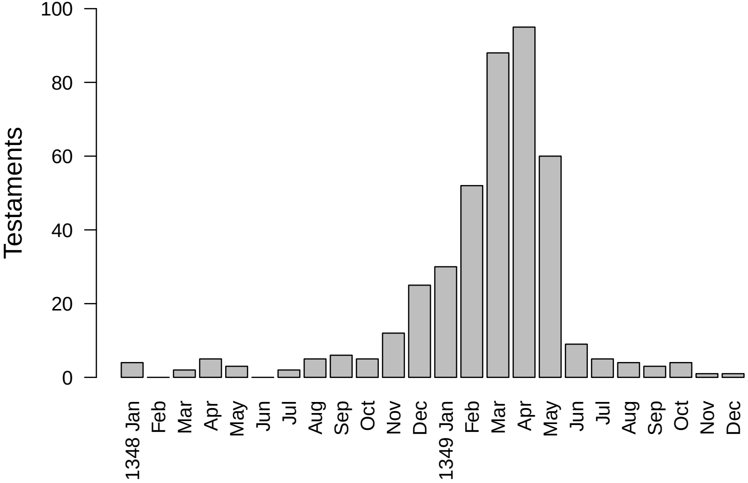


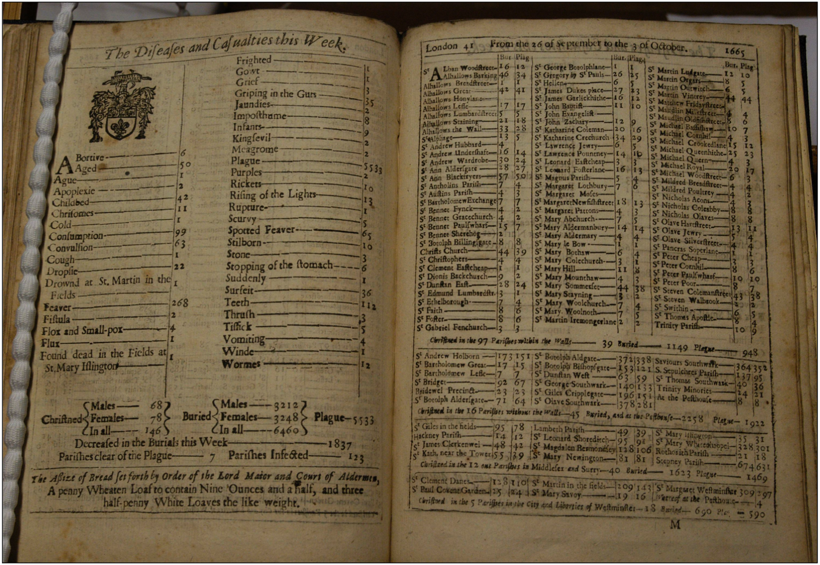
Fig. 4 Age-dependent COVID-19 hospitalization probability for known SARS-CoV-2 infection (panel **a**) and survival probability for hospitalized patients (panel **b**) in Ontario. We give age-by-age estimates of each probability (points; 95% exact binomial confidence intervals given by vertical lines), where point area is proportional to age-specific sample size. We additionally provide more precise estimates of these probabilities under stricter assumptions, modelling the hospitalization probability using a generalized additive model and the survival probability using a generalized linear model (curves; 95% confidence bands given by shaded regions). See "Methods" section for details

Papst et al (2021), BMC Public Health, 21:706

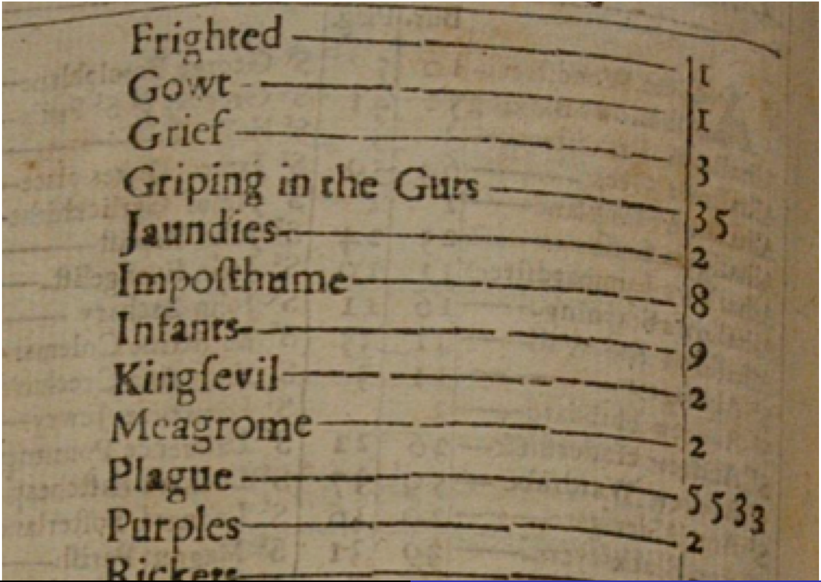
The Black Death in London, England, 1348–1349



London Bill of Mortality, 26 Sept to 3 Oct 1665

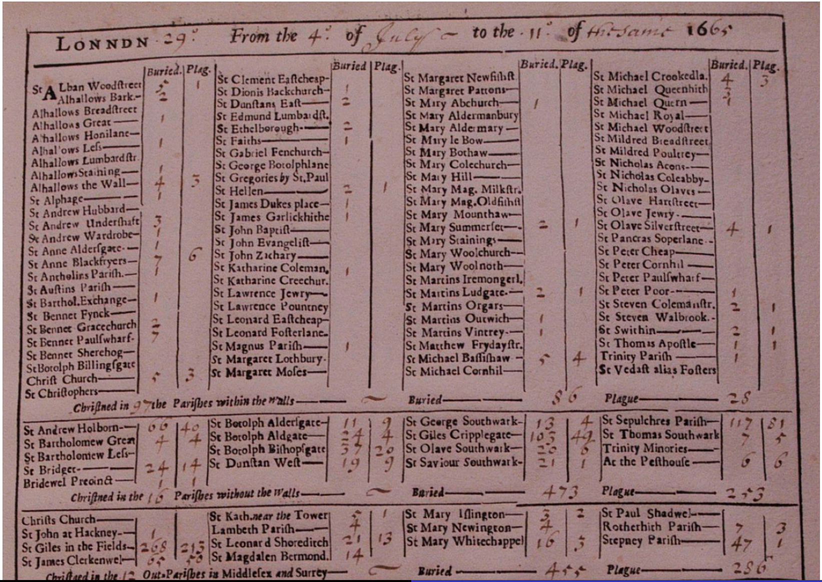


London Bill of Mortality, 26 Sept to 3 Oct 1665



Frighted	1
Gowt	1
Grief	3
Griping in the Guts	35
Jaundies	2
Imposthume	8
Infants	9
Kingsevil	2
Meagrome	2
Plague	5533
Purples	2
Rickets	

Mortality Bills are typically handwritten



LONNDN · 29° From the 4° of July — to the · 11° of the Same 1665

	Buried	Plag.
St Alban Woodstreet	5	1
Alhallows Bark.	2	
Alhallows Breadstreet		
Alhallows Great	1	
Alhallows Honilane		
Alhallows Less	1	
Alhallows Lumbardstr.		
Alhallows Staining	1	
Alhallows the Wall	4	3
St Alphage	1	
St Andrew Hubbard		
St Andrew Undershaft	3	
St Andrew Wardrobe	1	
St Anne Aldersgate	1	
St Anne Blackfryers	7	6
St Antholins Parish	1	
St Austins Parish		
St Barthol. Exchange	1	
St Bennet Fynck		
St Bennet Gracechurch	2	
St Bennet Paulswharf	7	
St Bennet Sherehog		
St Botolph Billingsgate		
Christ Church	5	3
St Christophers		
St Clement Eastcheap		
St Dionis Backchurch	1	
St Dunstans East	2	
St Edmund Lumbardst.		
St Ethelborough	2	
St Faiths	1	
St Gabriel Fenchurch		
St George Botolphlane		
St Gregories by St. Paul		
St Hellen	2	1
St James Dukes place	1	
St James Garlickhithe	1	
St John Baptist		
St John Evangelist		
St John Zachary		
St Katharine Coleman	1	
St Katharine Creechur.		
St Lawrence Jewry		
St Lawrence Pountney		
St Leonard Eastcheap		
St Leonard Fosterlane		
St Magnus Parish	1	
St Margaret Lothbury		
St Margaret Moses		
St Margaret Newfishst		
St Margaret Pattons		
St Mary Abchurch	1	
St Mary Aldermanbury		
St Mary Aldermary		
St Mary le Bow		
St Mary Bothaw		
St Mary Colechurch		
St Mary Hill		
St Mary Mag. Milkstr.		
St Mary Mag. Oldfishst		
St Mary Mounthaw		
St Mary Summerset	2	1
St Mary Stainings		
St Mary Woolchurch		
St Mary Woolnoth		
St Martins Iremongerl.		
St Martins Ludgate	2	1
St Martins Orgars		
St Martins Outwich	1	
St Martins Vintrey	1	
St Matthew Frydaystr.		
St Michael Bassishaw	5	4
St Michael Cornhil		
St Michael Crookedla.	4	3
St Michael Queenhith	3	
St Michael Quern	1	
St Michael Royal		
St Michael Woodstreet		
St Mildred Breadstreet		
St Mildred Poultrey		
St Nicholas Acons		
St Nicholas Coleabby		
St Nicholas Olaves		
St Olave Hartstreet		
St Olave Jewry		
St Olave Silverstreet	4	1
St Pancras Soperlane		
St Peter Cheap		
St Peter Cornhil		
St Peter Paulswharf		
St Peter Poor	1	
St Steven Colemanstr.	2	1
St Steven Walbrook.		
St Swithin	2	1
St Thomas Apostle	1	1
Trinity Parish	1	
St Vedast alias Fosters		

Christned in 97 *the Parishes within the Walls* — *Buried* — 86 *Plague* — 28

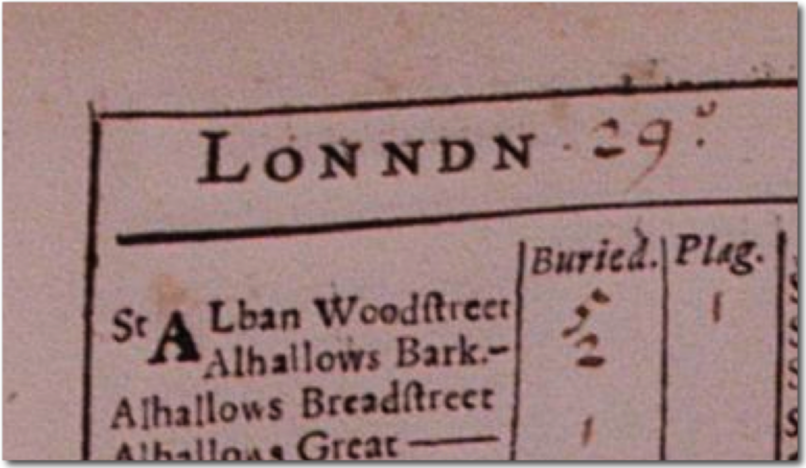
	Buried	Plag.
St Andrew Holborn	66	40
St Bartholomew Great	4	4
St Bartholomew Less		
St Bridget	24	14
Bridewel Precinct	1	1
St Botolph Aldersgate	11	9
St Botolph Aldgate	24	4
St Botolph Bishopsgate	37	20
St Dunstan West	19	9
St George Southwark	13	4
St Giles Cripplegate	103	42
St Olave Southwark	20	6
St Saviour Southwark	21	1
St Sepulchres Parish	117	81
St Thomas Southwark	7	5
Trinity Minories		
At the Pesthouse	6	6

Christned in the 16 *Parishes without the Walls* — *Buried* — 473 *Plague* — 2[illegible]3

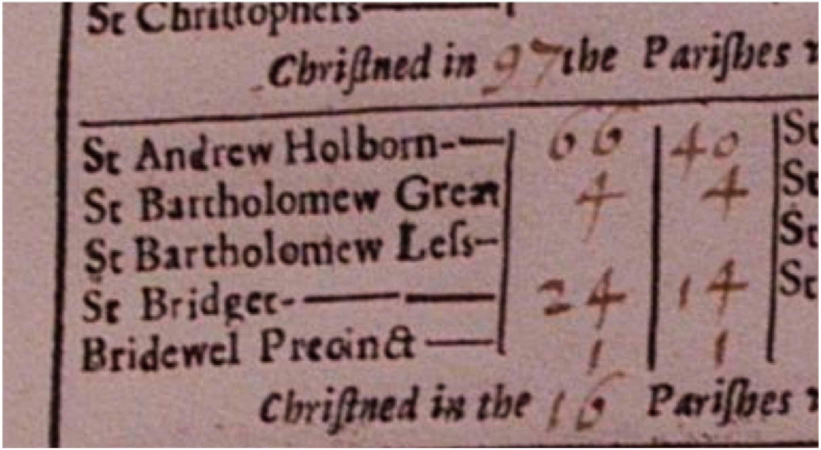
	Buried	Plag.
Christs Church		
St John at Hackney	1	
St Giles in the Fields	268	213
St James Clerkenwel	65	58
St Kath. near the Tower	4	1
Lambeth Parish		
St Leonard Shoreditch	21	13
St Magdalen Bermond.	14	
St Mary Islington	3	2
St Mary Newington	4	
St Mary Whitechappel	16	5
St Paul Shadwel		
Rotherhith Parish	7	3
Stepney Parish	47	1

Christned in the 12 *Out-Parishes in Middlesex and Surrey* — *Buried* — 455 *Plague* — 256

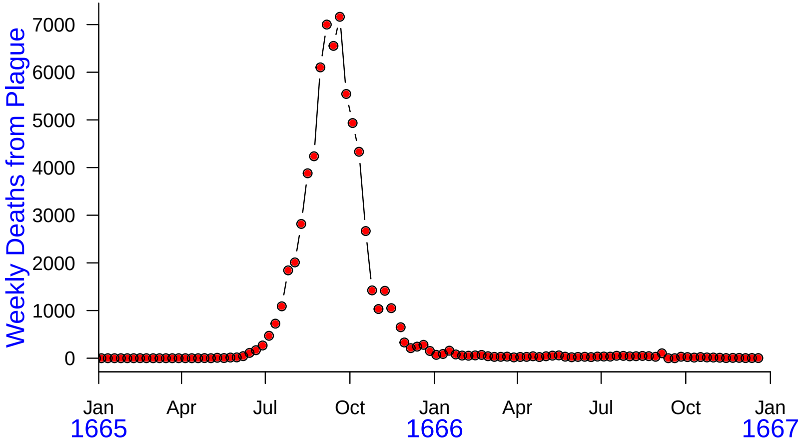
But handwriting is usually very clear



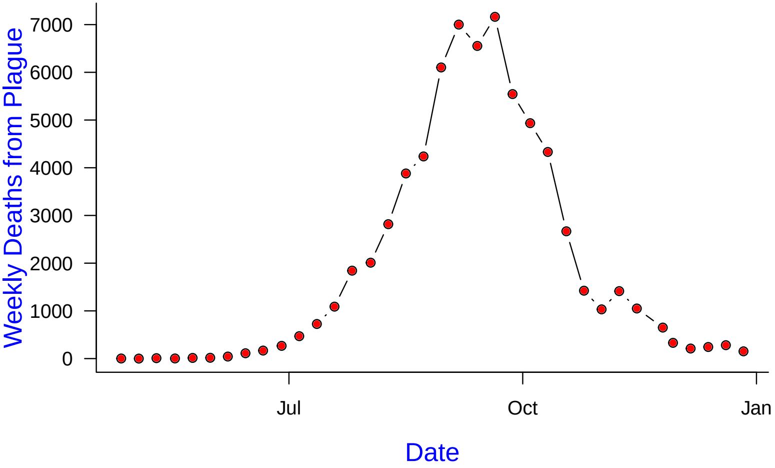
But handwriting is usually very clear



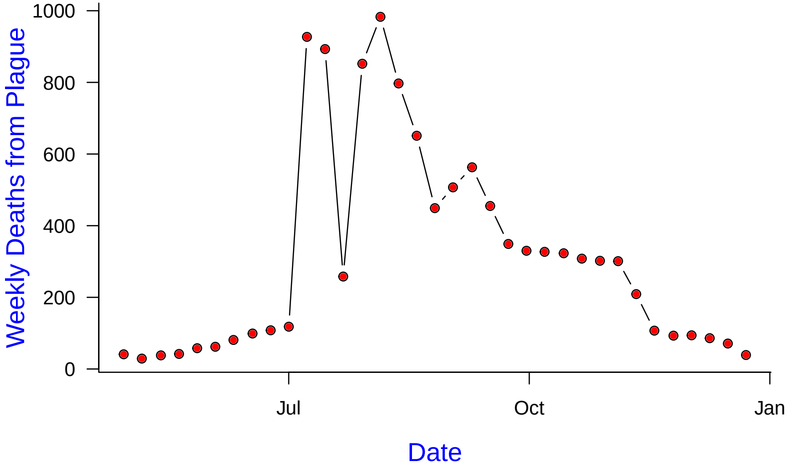
The Great Plague of London, 1665



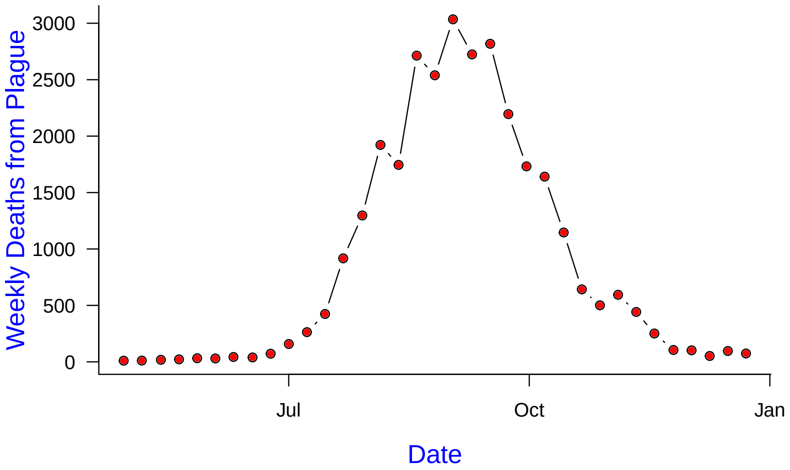
The Great Plague of London, 1665



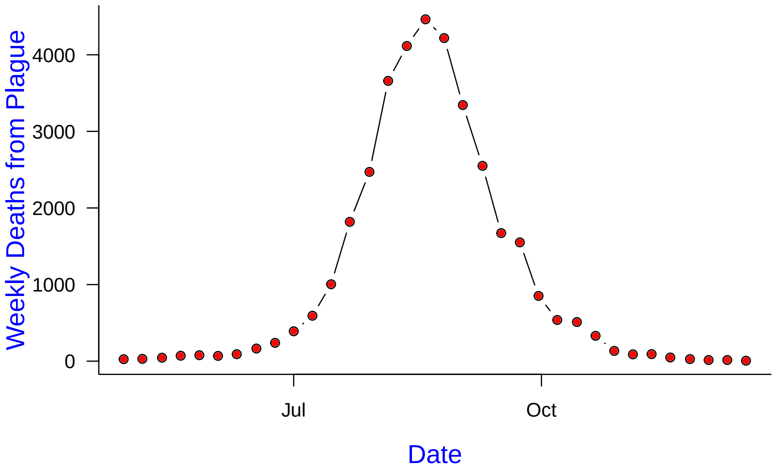
London Plague of 1593



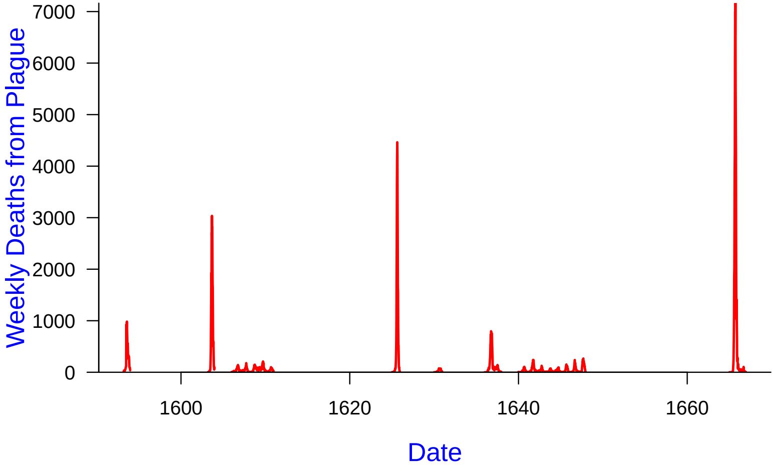
London Plague of 1603



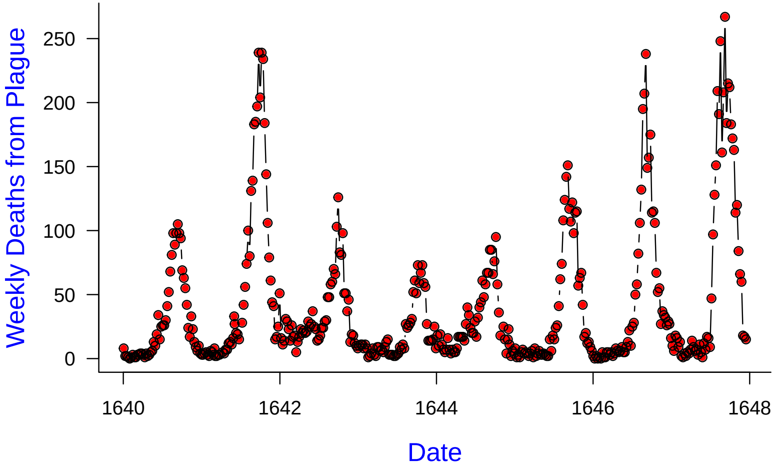
London Plague of 1625



Weekly Deaths from Plague in London, 1592–1666



Weekly Plague in London, 1640–1648



Some Plague Facts

- Plague epidemics recorded from Roman times to early 1900s.
- $\gtrsim 1/3$ Europe's population died in "Black Death" of 1348
  - ~ 300 years for the population to reach the same level.
- Recently (2011) established (at McMaster!) that the pathogen that caused The Black Death was *Yersinia pestis*

  [Bos *et al.* 2011, ***Nature*** **478**, 506–510]
- More recently (2014) established (again at McMaster!) that the pathogen that caused The Plague of Justinian (541–543 AD) was *Yersinia pestis*

  [Wagner *et al.* 2014, ***Lancet Infectious Diseases*** **14**, 319–326]
- *Y. pestis* still a concern?
  Yes: Rodent reservoir, antibiotic-resistant strains, bioterrorism
- **Spatial data** for any plagues? Yes, for London in 1665...

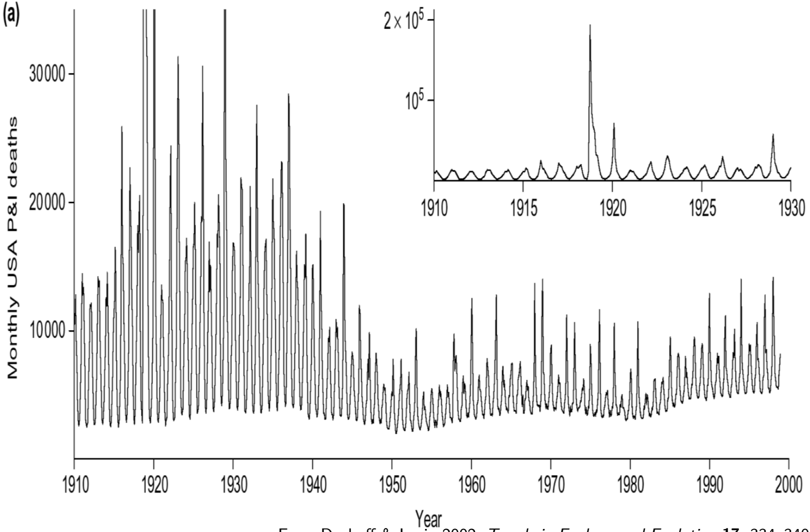
Visualization of spatial structure of Great Plague

- GIS encoding of parish boundaries
- Overlay parish boundaries on more modern map for reference
- Colour parishes as they become infected
- Is there evidence for spatial spread or was the spatial pattern random?
- DE low-tech animation...
- CBC high-tech animation...
  - *The Nature of Things*, 21 August 2014. http://www.cbc.ca/natureofthings/episodes/secrets-in-the-bones-the-hunt-for-the-black-death-killer

Visualization of entire course of the Great Plague

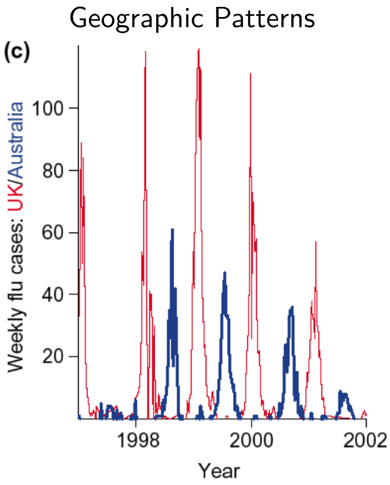
- What happenned after initial spatial spread?
- Visualize full spatial epidemic structure
- Show magnitude of epidemic in each parish with cylinder.
- Epidemic Visualization (`EpiVis`) software by Junling Ma.

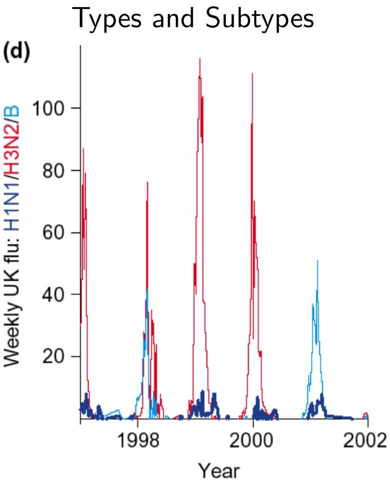
P&I mortality in U.S.A., 1910–1998



Earn, Dushoff & Levin 2002, *Trends in Ecology and Evolution* **17**, 334–340

Influenza Incidence Patterns (lab confirmed)

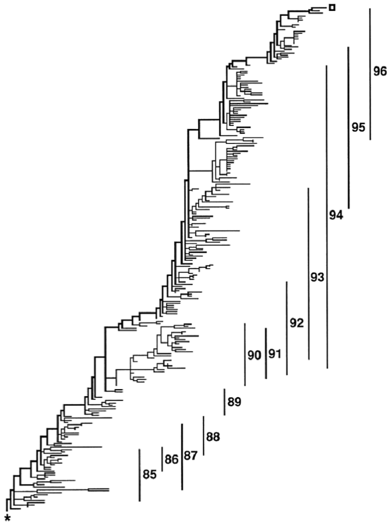




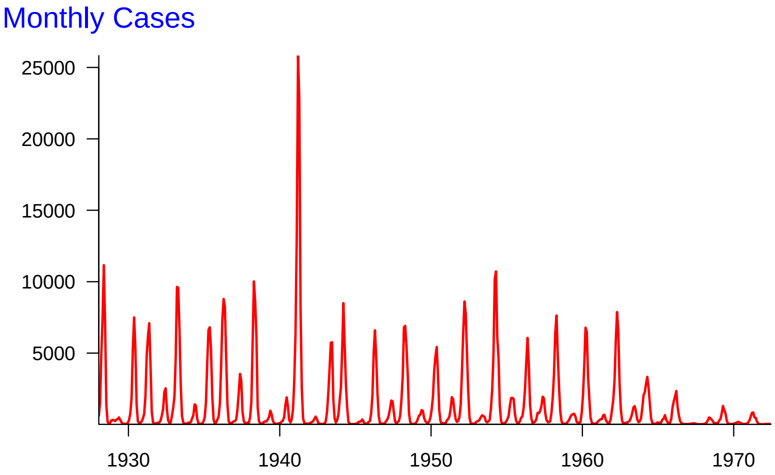
Earn, Dushoff & Levin 2002, *Trends in Ecology and Evolution* **17**, 334–340

Influenza Evolution

Molecular phylogenetic reconstruction of influenza A/H3N2 evolution, 1985–1996 (Fitch *et al.* 1997)

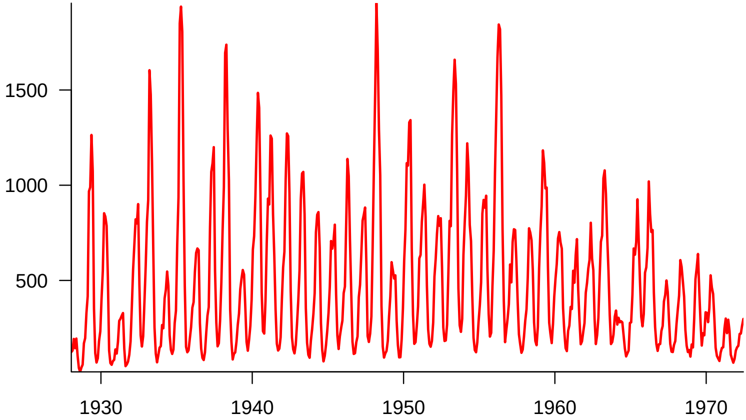


Measles in New York City, 1928–1972



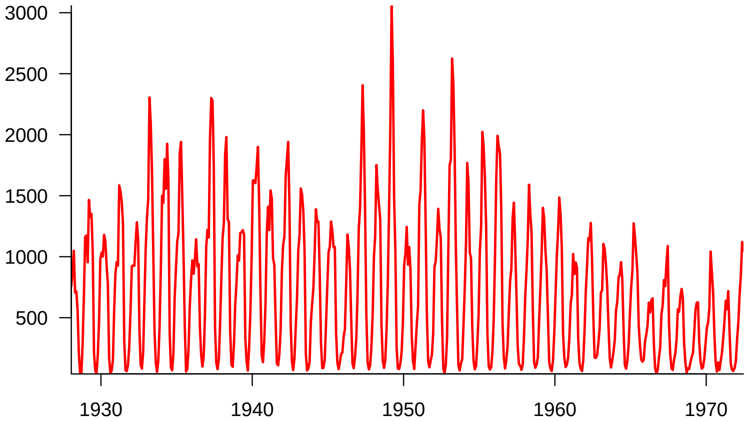
Mumps in New York City, 1928–1972

Monthly Cases

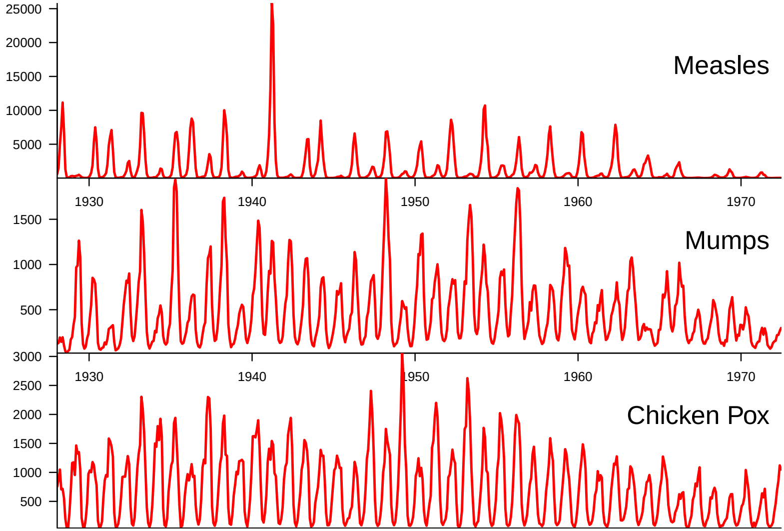


Chicken Pox in New York City, 1928–1972

Monthly Cases

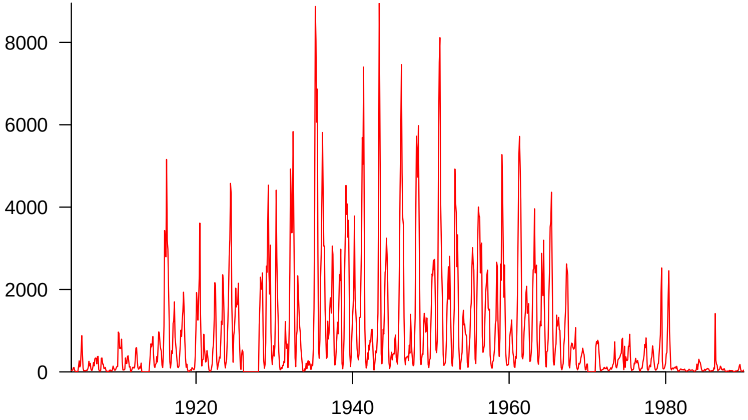


Childhood diseases in New York City, 1928–1972



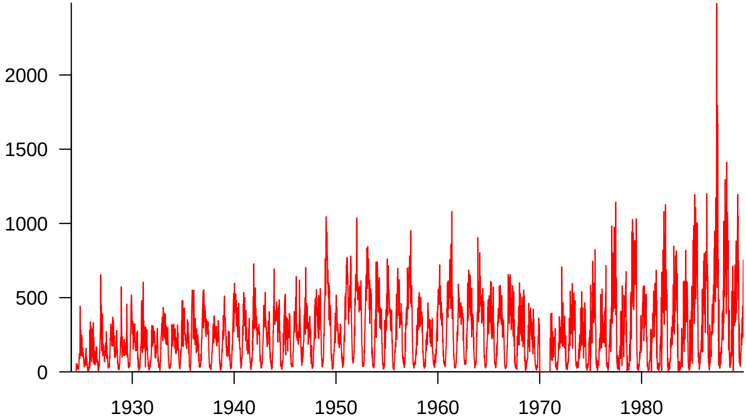
Measles in Ontario, 1904–1989

Monthly Cases



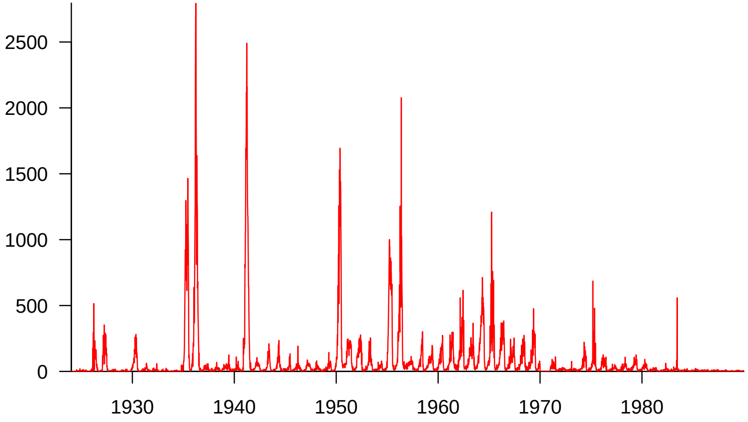
Chicken Pox in Ontario, 1924–1989

Monthly Cases



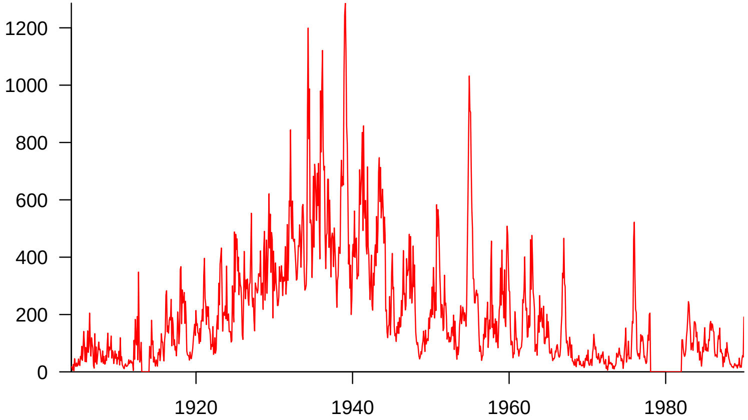
Rubella in Ontario, 1924–1989

Weekly Cases

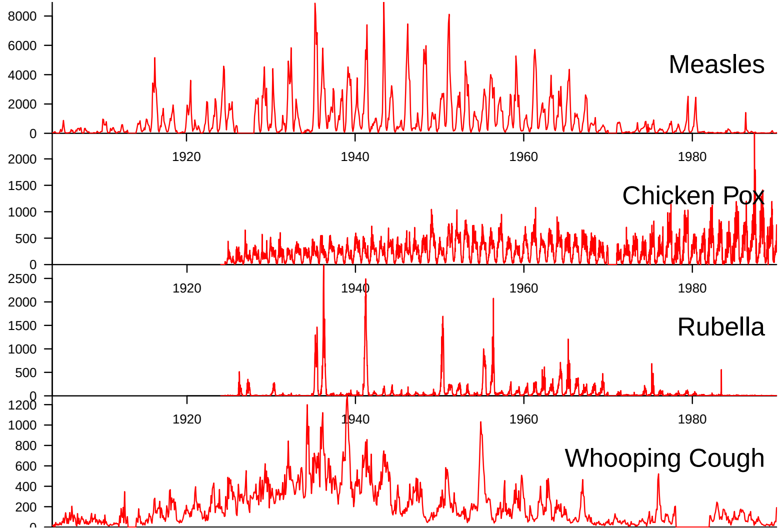


Whooping Cough in Ontario, 1904–1989

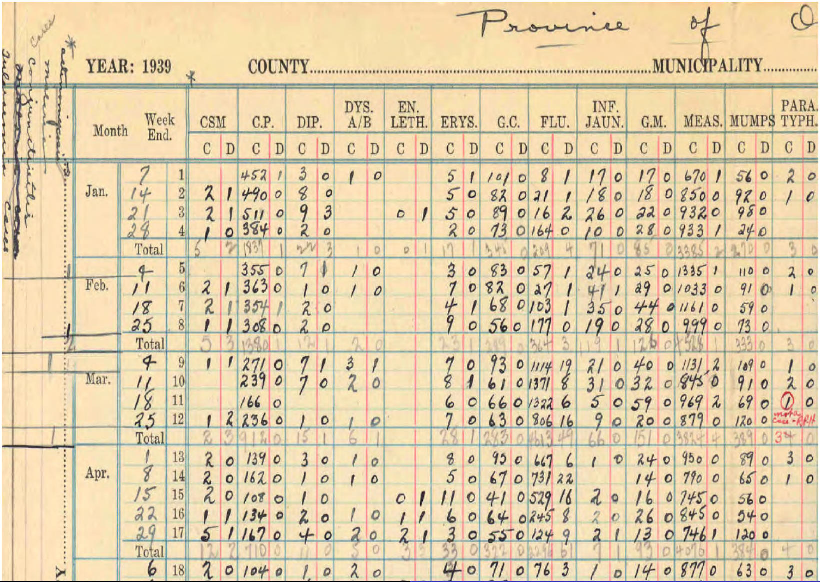
Monthly Cases



Childhood diseases in Ontario, 1904–1989



Ontario Disease Notification Data



Province of O

YEAR: 1939 COUNTY.................................... MUNICIPALITY..............

Month	Week End.		CSM C	CSM D	C.P. C	C.P. D	DIP. C	DIP. D	DYS. A/B C	DYS. A/B D	EN. LETH. C	EN. LETH. D	ERYS. C	ERYS. D	G.C. C	G.C. D	FLU. C	FLU. D	INF. JAUN. C	INF. JAUN. D	G.M. C	G.M. D	MEAS. C	MEAS. D	MUMPS C	MUMPS D	PARA. TYPH. C	PARA. TYPH. D
Jan.	7	1			452	1	3	0	1	0			5	1	101	0	8	1	17	0	17	0	670	1	56	0	2	0
	14	2	2	1	490	0	8	0					5	0	82	0	21	1	18	0	18	0	850	0	92	0	1	0
	21	3	2	1	511	0	9	3			0	1	5	0	89	0	16	2	26	0	22	0	932	0	98	0		
	28	4	1	0	384	0	2	0					2	0	73	0	164	0	10	0	28	0	933	1	24	0		
	Total		5	2	1837	1	22	3	1	0	0	1	17	1	345	0	209	4	71	0	85	0	3385	2	270	0	3	0
Feb.	4	5			355	0	7	1	1	0			3	0	83	0	57	1	24	0	25	0	1335	1	110	0	2	0
	11	6	2	1	363	0	1	0	1	0			7	0	82	0	27	1	41	1	29	0	1033	0	91	0	1	0
	18	7	2	1	354	1	2	0					4	1	68	0	103	1	35	0	44	0	1161	0	59	0		
	25	8	1	1	308	0	2	0					9	0	56	0	177	0	19	0	28	0	999	0	73	0		
	Total		5	3	1380	1	12	1	2	0			23	1	289	0	364	3	119	1	126	0	4528	1	333	0	3	0
Mar.	4	9	1	1	271	0	7	1	3	1			7	0	93	0	1114	19	21	0	40	0	1131	2	109	0	1	0
	11	10			239	0	7	0	2	0			8	1	61	0	1371	8	31	0	32	0	845	0	91	0	2	0
	18	11			166	0							6	0	66	0	1322	6	5	0	59	0	969	2	69	0	1	0
	25	12	1	2	236	0	1	0	1	0			7	0	63	0	806	16	9	0	20	0	879	0	120	0	meta case - RPH	
	Total		2	3	912	0	15	1	6	1			28	1	283	0	4613	49	66	0	151	0	3824	4	389	0	3*	0
Apr.	1	13	2	0	139	0	3	0	1	0			8	0	95	0	667	6	1	0	24	0	950	0	89	0	3	0
	8	14	2	0	162	0	1	0	1	0			5	0	67	0	731	22			14	0	790	0	65	0	1	0
	15	15	2	0	108	0	1	0			0	1	11	0	41	0	529	16	2	0	16	0	745	0	56	0		
	22	16	1	1	134	0	2	0	1	0	1	1	6	0	64	0	245	8	2	0	26	0	845	0	54	0		
	29	17	5	1	167	0	4	0	2	0	2	1	3	0	55	0	124	9	2	1	13	0	746	1	120	0		
	Total		12	2	710	0	11	0	5	0	3	3	33	0	322	0	2296	61	7	1	93	0	4076	1	384	0	4	0
	6	18	2	0	104	0	1	0	2	0			4	0	71	0	76	3	1	0	14	0	877	0	63	0	3	0

Dominion Bureau of Statistics Disease Notification Data

VITAL STATISTICS BRANCH - COMMUNICABLE DISEASE SECTION

Cases of Whooping Cough Reported by Provincial Health Departments, Year 1924

WEEK ENDING	P.E.I.		N.S.		N.B.		QUE.		ONT.		MAN.		SASK.		ALTA.		B.C.		CANADA	
	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt	wk	mt
1 Jan 5			11										1						12	
2 12			29										18						47	
3 19			37										32						69	
4 26			75	152				68		181		36	13	64		97		4	88	602
5 Feb 2			12		1								53						66	
6 9			5										40						45	
7 16			31										14						45	
8 23			-2	50	1	2		267		202		48	4	111		116		1	7	797
9 Mar 1			2										21						23	
10 8													9						9	
11 15			3										11						14	
12 22			60										34						94	
13 29			2	67				144		140		52	15	90		15		7	17	515
14 Apr 5			9										11						20	
15 12			1										12						13	
16 19			26		1								8						35	
17 26			14	50	3	4		42		140		39	16	47		67		5	33	394
18 May 3			26										2						28	

All Historical Canadian Infectious Disease Data

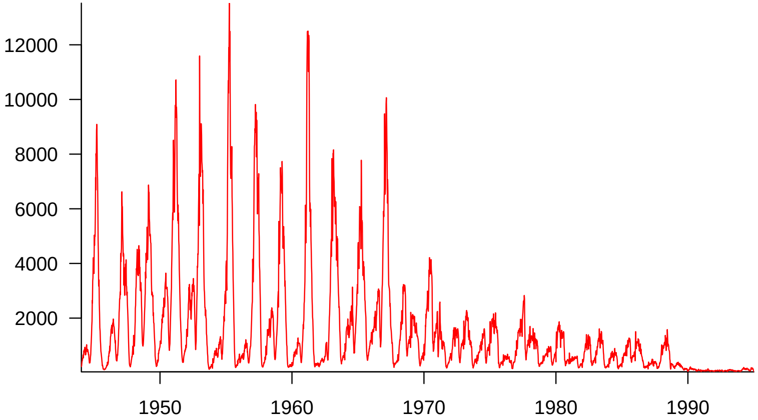
https://canmod.net/digitization/

Recurrent epidemics of childhood infections

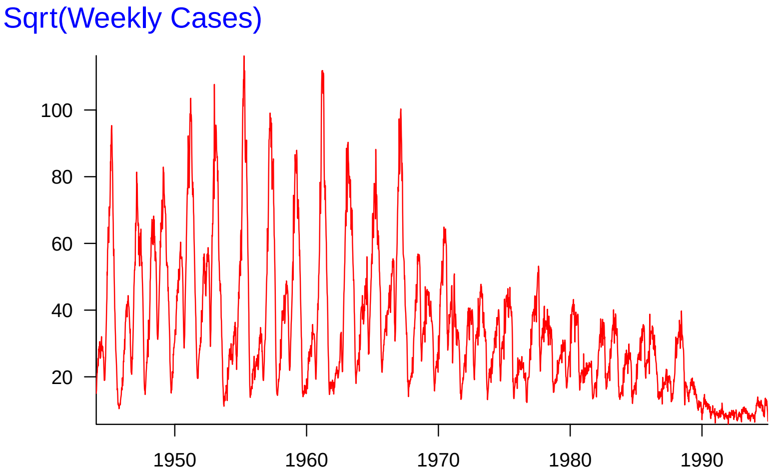
- Childhood diseases in New York City, 1928–1972
- Childhood diseases in Ontario, 1904–1989

Measles incidence in England and Wales, 1944–1995

Weekly Cases



Measles incidence in England and Wales, 1944–1995



Why study measles epidemics?

- $\sim 140,000$ annual deaths from measles
- A major cause of *vaccine-preventable* deaths.
- Potential impact in developed countries during vaccine scares (e.g., MMR scare in UK in 1990s).

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

Other reasons to model infectious disease epidemics

- Mathematical models make hypotheses and inferences precise
  - Give better advice to policymakers
  - Make better predictions
- Host-pathogen dynamics are important aspects of ecosystem dynamics
  - Infectious disease models more likely to be successful than predator-prey models
- Excellent data for human infectious diseases
  - Models can be tested!

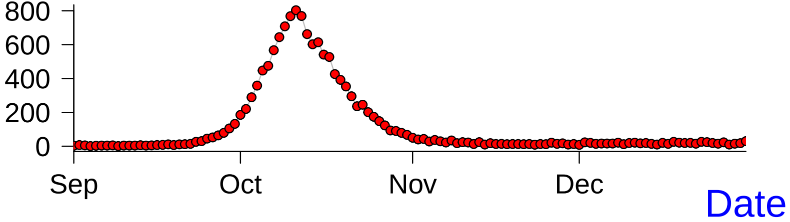
Modelling population dynamics of childhood infections

- The basic SIR model cannot explain recurrent epidemics.
- What should we do?. . . The usual options:
  1. Get depressed, drop the course.
  2. Keep developing models until we can explain recurrent epidemics.
- First, let's talk about tools that allow us to make our questions about time series data more precise.

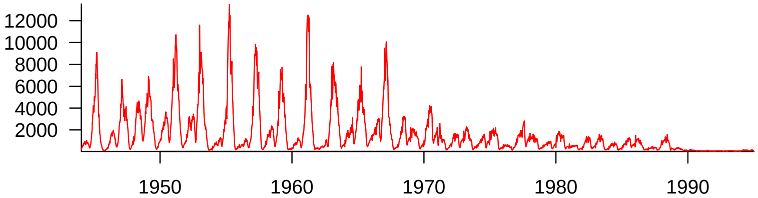
Epidemic Data Analysis

Time Plots of Temporal Epidemic Patterns

1918 P&I

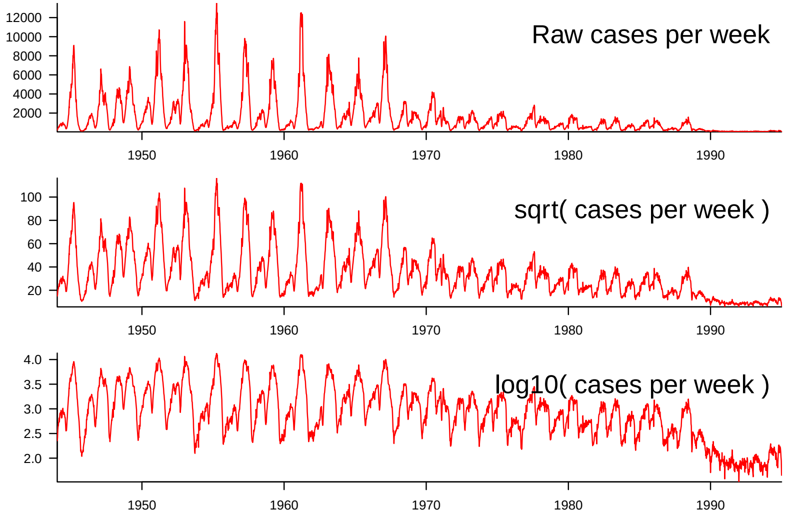


Weekly Measles in England and Wales



Time Plots of Transformed Data

- Reveal unobvious aspects of time series



Times Plots of Smoothed Data

- Reveal trends clouded by noise or seasonality

- *Moving Average:*

$$x_t \to \frac{1}{2a+1} \sum_{i=-a}^{a} x_{t+i}$$

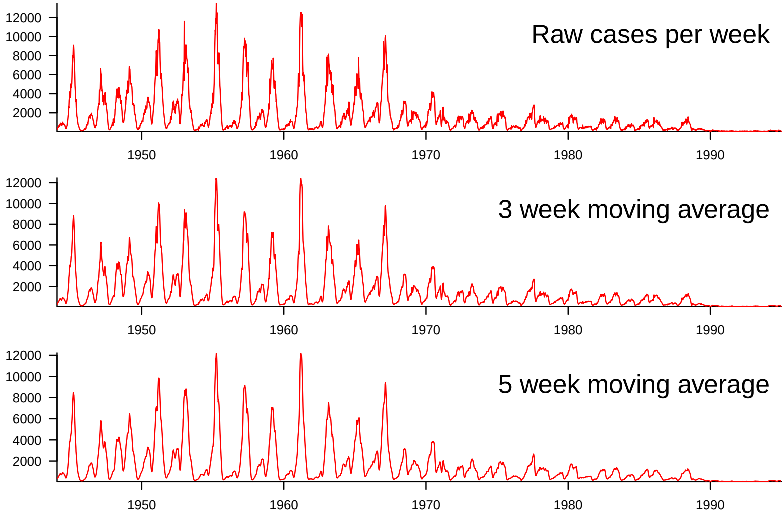
  - Replace original data points x_t with averages of nearby points.

- *Linear filter:*

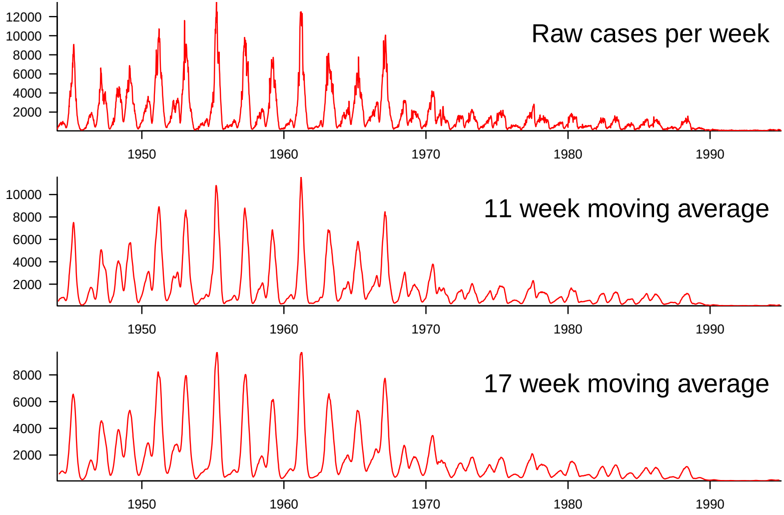
$$x_t \to \sum_{i=-\infty}^{\infty} \lambda_i x_{t+i}$$

  - Generalization of moving average.
  - *Weights* λ_i can be nonlinear functions of i.

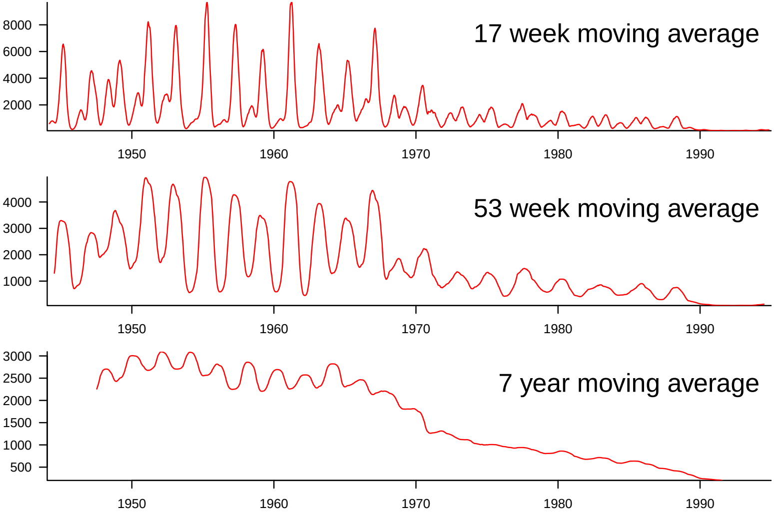
Times Plots of Smoothed Data



Times Plots of Smoothed Data



Times Plots of Smoothed Data



Correlation

- Recurrent epidemics $\Longrightarrow$ number of cases now is correlated with number of cases in the past and the future.
- Given N pairs of observations of different quantities, $\{(x_i, y_i) : i = 1, \ldots, N\}$, the *correlation coefficient* is defined to be

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the means of $\{x_i\}$ and $\{y_i\}$, respectively.

Correlation

Properties of the correlation coefficient:

- $-1 \leq r \leq 1$ (Proof? Cauchy-Schwarz inequality)
- $r = 1 \iff$ all points lie on a line with positive slope ("complete positive correlation")
- $r = -1 \iff$ all points lie on a line with negative slope ("complete negative correlation")
- $r \simeq 0 \implies$ "uncorrelated"
- *Interpretation:* r^2 is the proportion of the variance in y explained by a linear function of x.

Derivations and discussions:

- MathWorld on r^2, Wikipedia on r^2
- Wikipedia on general coefficient of determination

Autocorrelation

- Given a single sequence of observations $\{x_t : t = 1, \ldots, N\}$, we can compute the correlation of each observation with the observation k time steps in the future.

- Thus, we consider the pairs of observations $\{(x_t, x_{k+t}) : t = 1, \ldots, N-k\}$ and define the *autocorrelation coefficient at lag k* to be

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

  where $\bar{x}_{1,N-k}$ and $\bar{x}_{k+1,N}$ are the means of first and last $N-k$ observations, respectively.

Autocorrelation

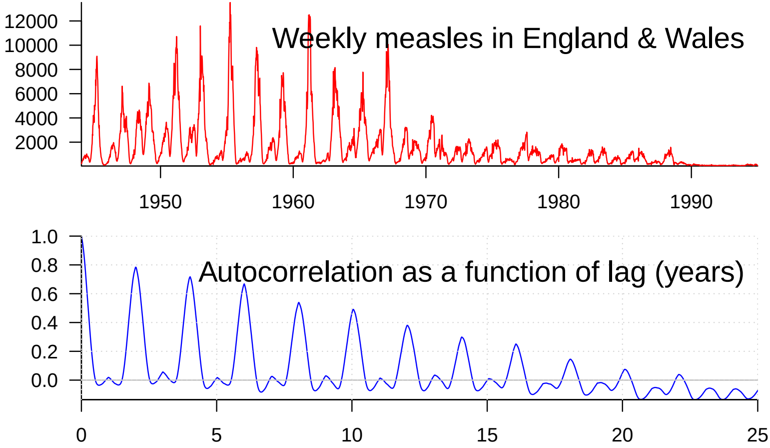
- If number of observations N is large and lag $k \ll N$ then

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

- Approximation of r_k is worse for larger lags k

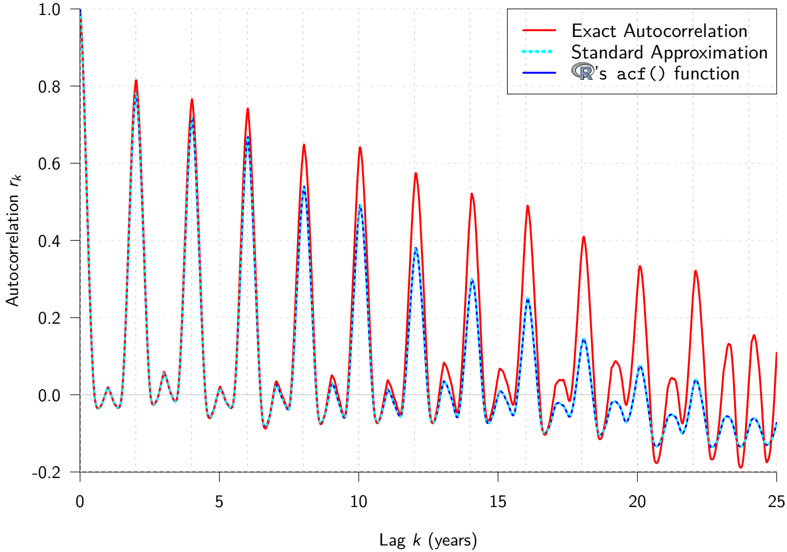
- Plot of autocorrelation r_k as a function of lag k is called the *correlogram*.

Correlogram



- Peaks in correlogram $\Longrightarrow$ periodicities in original time series.
- Correlograms of temporal segments are often informative.

Correlogram: exact vs. approximate r_k



Spectral Density

- Can we compute the dominant periods in the time series? (Rather than estimating them by eye from the correlogram.)
- Express the time series as a Fourier series:

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

where $\omega_p = 2\pi p/N$.

- Compute the *Fourier coefficients* $\{a_p\}$, $\{b_p\}$ by taking inner products with $\cos \omega_p t$ and $\sin \omega_p t$.

Spectral Density

- Fourier coefficients of x_t are:

$$a_0 = \bar{x} = \frac{1}{N}\sum_t x_t \,,$$

$$a_p = \frac{2}{N}\sum_t x_t \cos\omega_p t \,, \qquad b_p = \frac{2}{N}\sum_t x_t \sin\omega_p t \,,$$

$$a_{N/2} = \frac{1}{N}\sum_t (-1)^t x_t \,,$$

where sum is over observation times.

- Estimated power spectral density (PSD) at frequency ω_p is*:

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

*The normalization by $N/4\pi$ is the convention chosen by Chatfield (2004, "Analysis of Time Series: An Introduction"). Other normalization conventions are also in common use.

Spectral Density

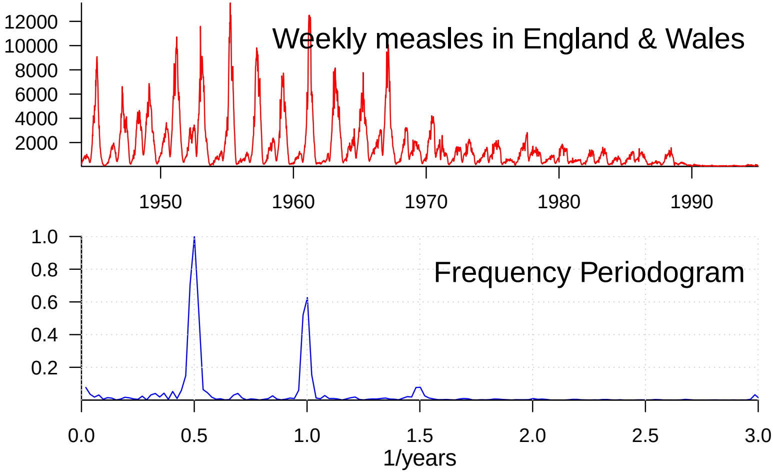
- There are many different ways to express the power spectral density (aka *power spectrum*).

- Most common/useful equivalence is that the power spectrum is the discrete Fourier transform of the correlogram:

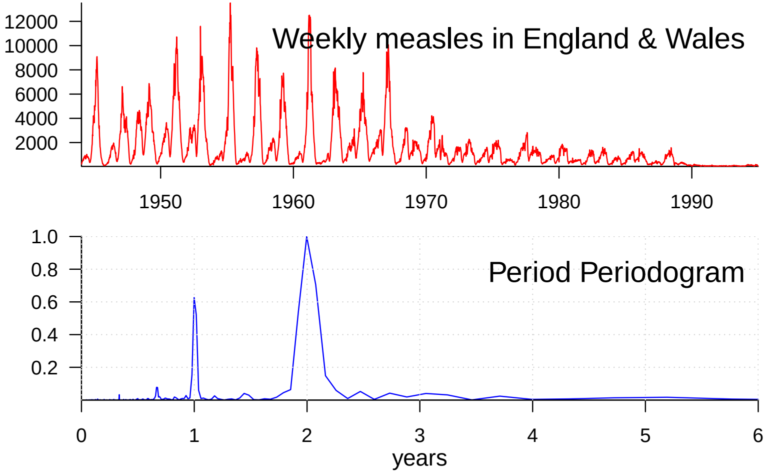
$$I(\omega_p) = \frac{1}{\pi}\left(r_0 + 2\sum_{k=1}^{N-1} r_k \cos \omega_p k\right)$$

- Plot of estimated power spectrum as a function of frequency ω_p is called the *frequency periodogram* or just the *periodogram*.

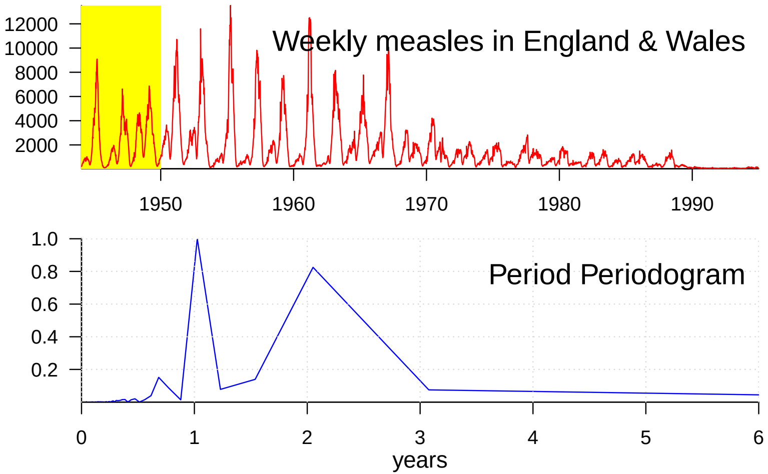
Spectral Density



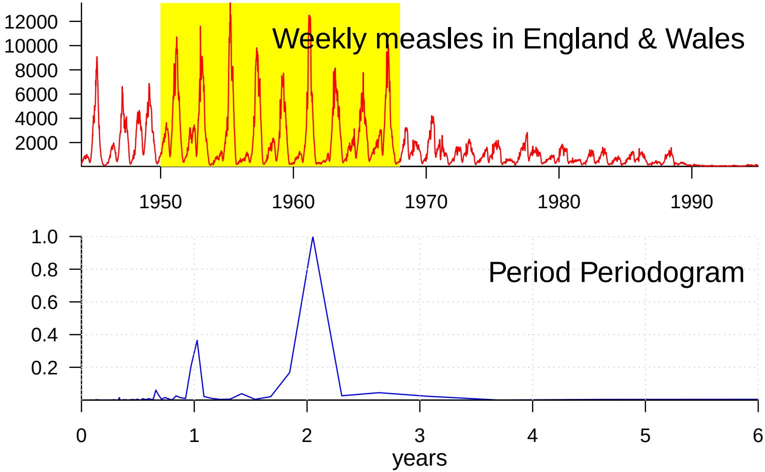
Spectral Density



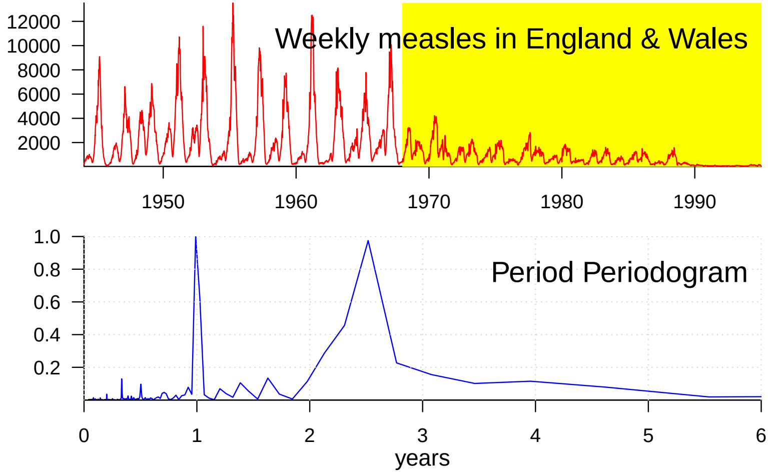
Spectral Density of Temporal Segments



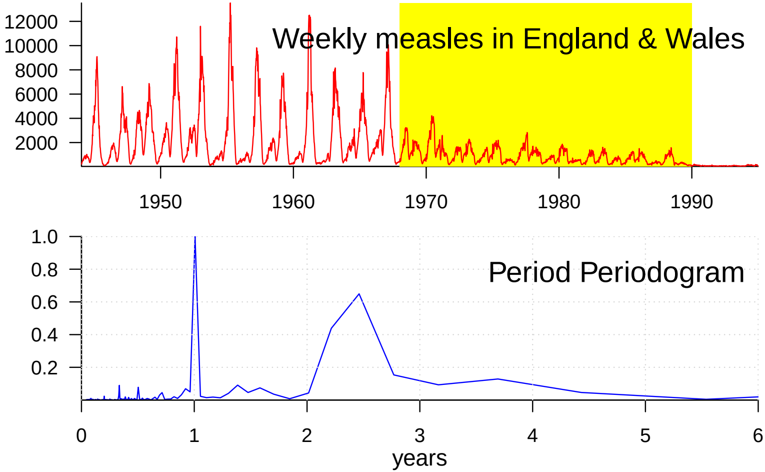
Spectral Density of Temporal Segments



Spectral Density of Temporal Segments



Spectral Density of Temporal Segments



Spectral Density Properties

- Periodogram is discrete Fourier transform of correlogram
- Same information in correlogram and periodogram
- Periodogram usually easier to interpret
- In R, calculate power spectrum with `spectrum()`
- The power spectrum $I(\omega_p)$ partitions the variance in the time series with respect to frequency ω_p.
  - Parseval's theorem implies $\frac{1}{N}\sum_t (x_t - \bar{x})^2 = \frac{1}{2\pi N}\sum_{p>0} I(\omega_p)$. But $\frac{1}{N}\sum_t (x_t - \bar{x})^2 = \mathrm{Var}\{x_t\}$, hence $I(\omega_p)/(2\pi N)$ is the proportion of the variance in the time series associated with period $2\pi/\omega_p$. [For details, see Chatfield (2004).]

Basic Time Series Analysis of Epidemic Data