McMaster University

Mathematics
and Statistics

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

Mathematics 4MB3/6MB3
Mathematical Biology

Instructor: David Earn

Lecture 8
Epidemic Data
Wednesday 24 January 2018

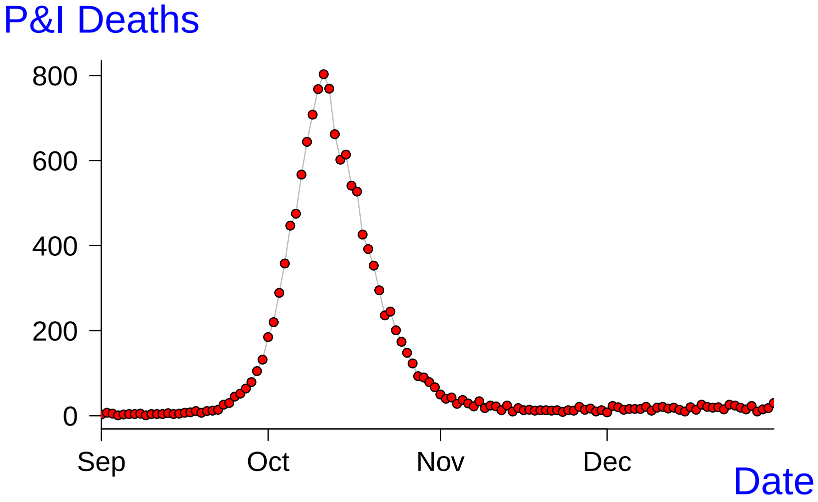
Announcements

- Thanks everyone for doing the contributions survey for Assignment 1.
- Don't stress about the ratings about each other's contributions. The issue is whether some group members did not pull their weight. If somebody didn't try and others had to pick up the slack, that person should be penalized. I will not penalize somebody because they tried but felt they didn't contribute as much to the final document as they could have. Do try to even out the work across the assignments.
- Make sure everyone in your group gets a chance to be in control of the LaTeX for one assignment.

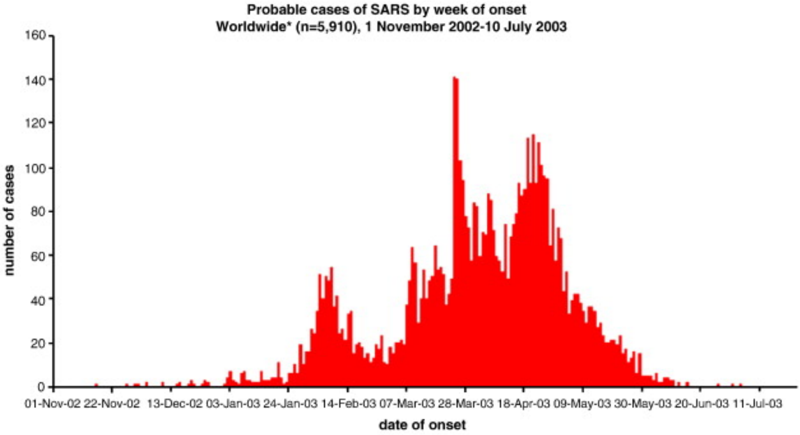
More Announcements!

- **Assignment 2:**
  Due Monday 5 February 2018 in class (and by e-mail) at 11:30am.
- **Midterm test:**
  - *Date:* week of 5–9 March? or 12–16 March?
  - *Time:* TBA
  - *Location:* TBA

P&I Mortality, Philadelphia, 1918

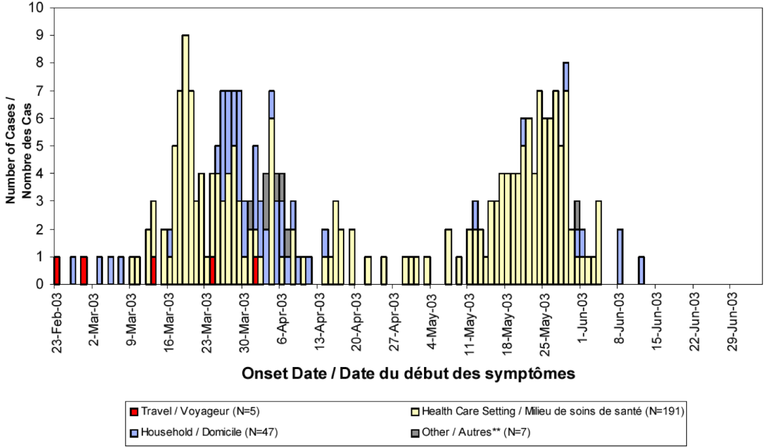


SARS in 2003 (Worldwide)



*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)

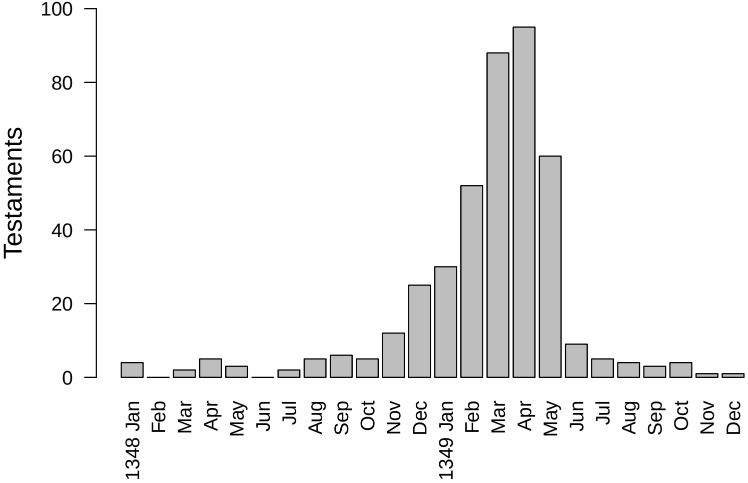


$N = 249$ (of 250 reported)

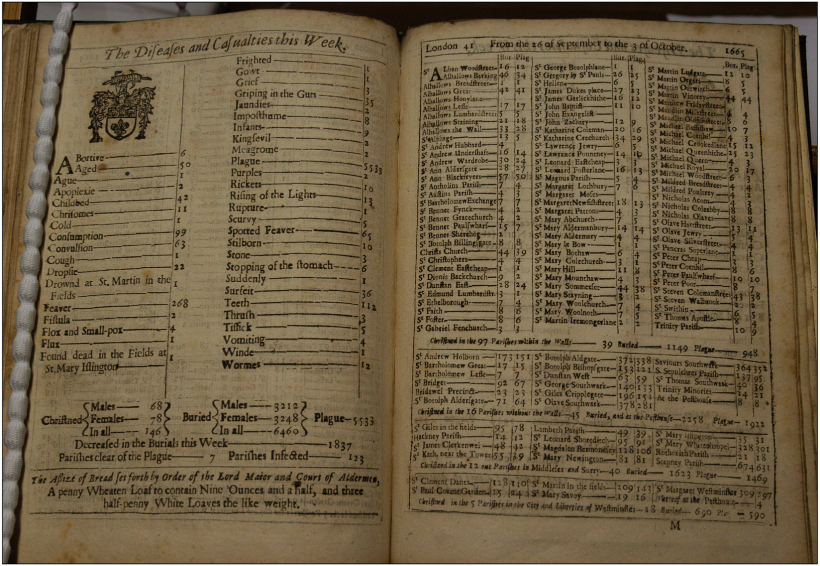
Some SARS Facts

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

The Black Death in London, England, 1348–1349



London Bill of Mortality, 26 Sept to 3 Oct 1665



The Diſeaſes and Caſualties this Week.

Abortive	6
Aged	50
Ague	1
Apoplexie	2
Childbed	42
Chriſomes	11
Cold	1
Conſumption	99
Convulſion	63
Cough	1
Dropſie	22
Drownd at St. Martin in the Fields	1
Feaver	268
Fiſtula	2
Flox and Small-pox	4
Flux	1
Found dead in the Fields at St. Mary Iſlington	1
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	10
Riſing of the Lights	13
Rupture	1
Scurvy	5
Spotted Feaver	65
Stilborn	10
Stone	3
Stopping of the ſtomach	6
Suddenly	1
Surfeit	36
Teeth	112
Thruſh	3
Tiſſick	5
Vomiting	4
Winde	1
Wormes	12

Chriſtned: Males 68, Females 78, In all 146
Buried: Males 3212, Females 3248, In all 6460
Plague 5533

Decreaſed in the Burials this Week 1837

Pariſhes clear of the Plague 7 Pariſhes Infected 123

The Aſsize of Bread ſet forth by Order of the Lord Maior and Court of Aldermen, A penny Wheaten Loaf to contain Nine Ounces and a half, and three half-penny White Loaves the like weight.

London 41 From the 26 of September to the 3 of October. 1665

	Bur.	Plag.		Bur.	Plag.		Bur.	Plag.
St Alban Woodſtreet	16	12	St George Botolphlane	1	1	St Martin Ludgate	12	10
Alhallows Barking	46	34	St Gregory by St Pauls	26	25	St Martin Orgars	8	5
Alhallows Breadſtreet	1	1	St Hellen	6	5	St Martin Outwitch	6	5
Alhallows Great	42	41	St James Dukes place	27	23	St Martin Vintrey	44	44
Alhallows Honylane			St James Garlickhithe	16	12	St Matthew Fridayſtreet		
Alhallows Leſſe	17	17	St John Baptiſt	11	10	St Maudlin Milkſtreet	4	4
Alhallows Lumbardſtreet	5	5	St John Evangeliſt			St Maudlin Oldfiſhſtreet	8	6
Alhallows Staining	21	18	St John Zachary	12	9	St Michael Baſſiſhaw	10	7
Alhallows the Wall	33	28	St Katharine Coleman	20	16	St Michael Cornhil	4	3
St Alphage	13	5	St Katharine Crechurch	34	29	St Michael Crookedlane	15	12
St Andrew Hubbard	4		St Lawrence Jewry	6	5	St Michael Queenhithe	25	23
St Andrew Underſhaft	16	14	St Lawrence Pountney	14	10	St Michael Quern	4	3
St Andrew Wardrobe	30	24	St Leonard Eaſtcheap	3	3	St Michael Royal	20	17
St Ann Alderſgate	28	27	St Leonard Foſterlane	16	13	St Michael Woodſtreet	6	3
St Ann Blackfryers	57	50	St Magnus Pariſh	5	4	St Mildred Breadſtreet	4	4
St Antholins Pariſh	7	4	St Margaret Lothbury	7	6	St Mildred Poultrey	4	2
St Auſtins Pariſh	4	3	St Margaret Moſes			St Nicholas Acons	4	3
St Bartholomew Exchange	7	7	St Margaret Newfiſhſtreet	18	13	St Nicholas Coleabby	8	8
St Bennet Fynck	4	2	St Margaret Pattons	4	3	St Nicholas Olaves	8	8
St Bennet Gracechurch	4	2	St Mary Abchurch	7	5	St Olave Hartſtreet	13	11
St Bennet Paulſwharf	15	7	St Mary Aldermanbury	14	14	St Olave Jewry	5	4
St Bennet Sherehog	2		St Mary Aldermary	4	4	St Olave Silverſtreet	4	4
St Botolph Billingſgate	8	8	St Mary le Bow	1	1	St Pancras Soperlane	1	1
Chriſts Church	44	39	St Mary Bothaw	6	4	St Peter Cheap	3	3
St Chriſtophers	4	4	St Mary Colechurch	3	1	St Peter Cornhil	8	6
St Clement Eaſtcheap	1	1	St Mary Hill	11	8	St Peter Paulſwharf	10	10
St Dionis Backchurch	9	2	St Mary Mounthaw	4	3	St Peter Poor	8	7
St Dunſtan Eaſt	28	24	St Mary Sommerſet	44	38	St Steven Colemanſtreet	43	38
St Edmund Lumbardſtr.	3	1	St Mary Stayning	3	2	St Steven Walbrook	2	2
St Ethelborough	7	4	St Mary Woolchurch	7	4	St Swithin	6	5
St Faith	8	6	St Mary Woolnoth	7	5	St Thomas Apoſtle	8	4
St Foſter	8	6	St Martin Iremongerlane	2	2	Trinity Pariſh	10	9
St Gabriel Fenchurch	3	3						

Chriſtned in the 97 Pariſhes within the Walls 39 *Buried* 1149 *Plague* 948

St Andrew Holborn	173	151	St Botolph Aldgate	372	338	Saviours Southwark	364	352
St Bartholomew Great	17	15	St Botolph Biſhopſgate	153	121	S. Sepulchres Pariſh	137	95
St Bartholomew Leſſe	7	7	St Dunſtan Weſt	63	59	St Thomas Southwark	40	36
St Bridget	92	67	St George Southwark	140	133	Trinity Minories	24	21
Bridewel Precinct	23	23	St Giles Cripplegate	196	151	At the Peſthouſe	8	8
St Botolph Alderſgate	71	64	St Olave Southwark	378	281			

Chriſtned in the 16 Pariſhes without the Walls 45 *Buried, and at the Peſthouſe* 2258 *Plague* 1922

St Giles in the fields	95	78	Lambeth Pariſh	49	39	St Mary Iſlington	35	31
Hackney Pariſh	14	12	St Leonard Shoreditch	95	91	St Mary Whitechappel	328	301
St James Clerkenwel	48	42	St Magdalen Bermondſey	128	106	Rotherith Pariſh	21	18
St Kath. near the Tower	55	39	St Mary Newington	81	81	Stepney Pariſh	674	631

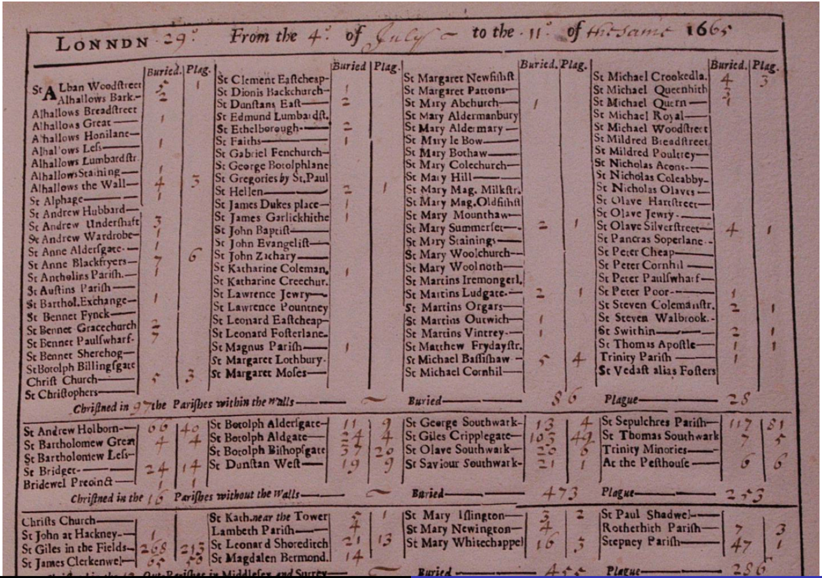
Chriſtned in the 12 out Pariſhes in Middleſex and Surry 40 *Buried* 1623 *Plague* 1469

St Clement Danes	128	110	St Martin in the fields	209	143	St Margaret Weſtminſter	309	297
St Paul Covent Garden	25	24	St Mary Savoy	19	16	*whereof at the* Peſthouſe		4

Chriſtned in the 5 Pariſhes in the City and Liberties of Weſtminſter 18 *Buried* 690 *Plague* 590

M

Mortality Bills are typically handwritten



LONNDN 29° From the 4° of July to the 11° of the same 1665

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

Chriſtned in 97 the Pariſhes within the Walls ~ Buried 86 Plague 28

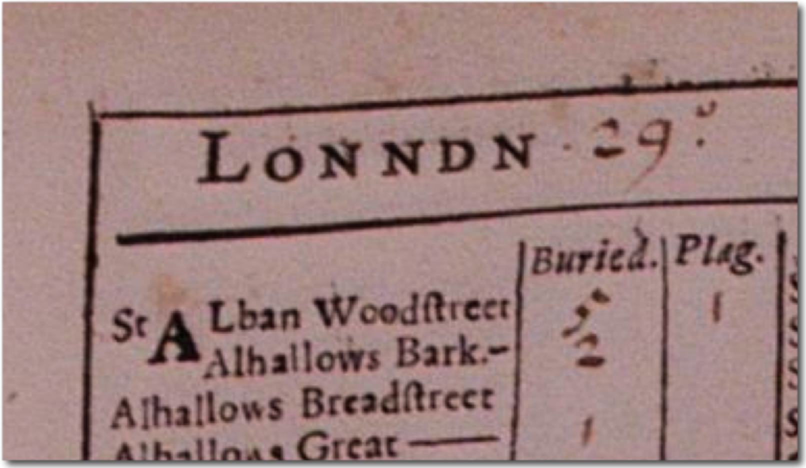
	Buried	Plag.		Buried	Plag.		Buried	Plag.		Buried	Plag.
St Andrew Holborn	66	40	St Botolph Alderſgate	11	9	St George Southwark	13	4	St Sepulchres Pariſh	117	81
St Bartholomew Great	4	4	St Botolph Aldgate	24	4	St Giles Cripplegate	103	49	St Thomas Southwark	7	5
St Bartholomew Leſs			St Botolph Biſhopſgate	37	20	St Olave Southwark	20	6	Trinity Minories		
St Bridget	24	14	St Dunſtan Weſt	19	9	St Saviour Southwark	21	1	At the Peſthouſe	6	6
Bridewel Precinct	1	1									

Chriſtned in the 16 Pariſhes without the Walls ~ Buried 473 Plague 253

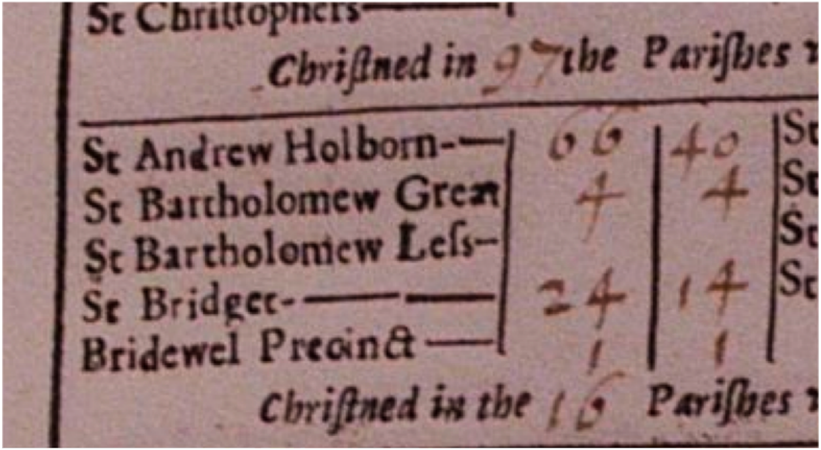
	Buried	Plag.		Buried	Plag.		Buried	Plag.		Buried	Plag.
Chriſts Church			St Kath. near the Tower	5	1	St Mary Iſlington	3	2	St Paul Shadwel		
St John at Hackney	1		Lambeth Pariſh	4		St Mary Newington	4		Rotherhith Pariſh	7	3
St Giles in the Fields	268	213	St Leonard Shoreditch	21	13	St Mary Whitechappel	16	5	Stepney Pariſh	47	1
St James Clerkenwel	65	50	St Magdalen Bermond.	14							

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

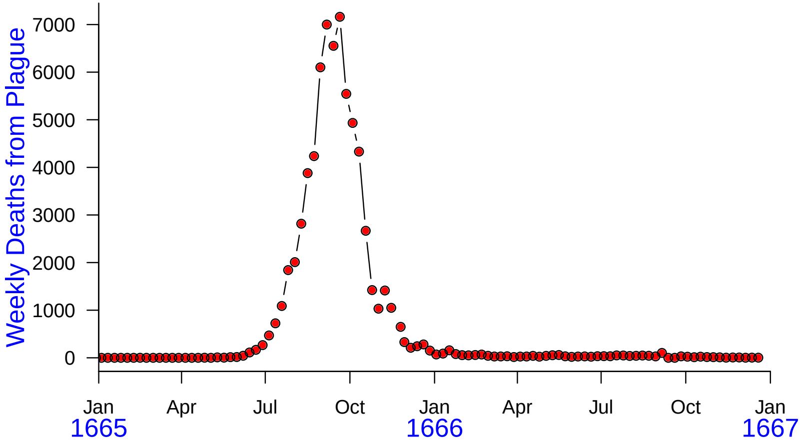
But handwriting is usually very clear



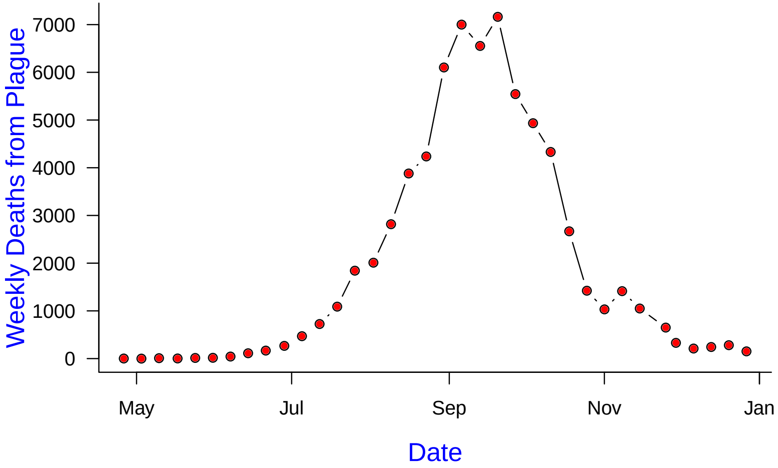
But handwriting is usually very clear



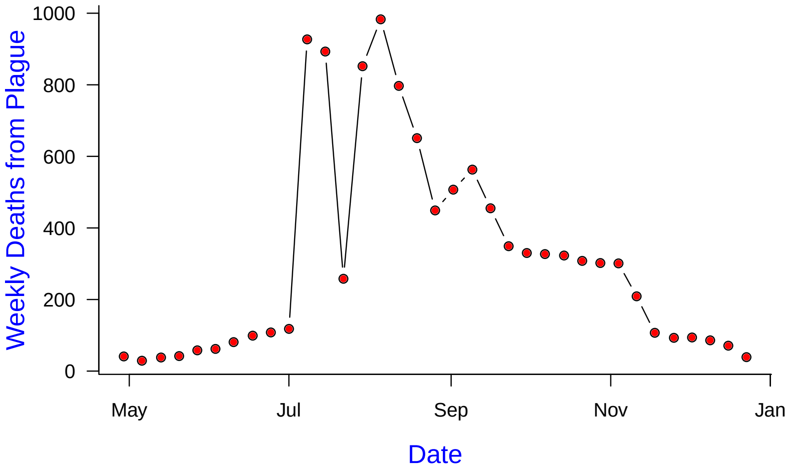
The Great Plague of London, 1665



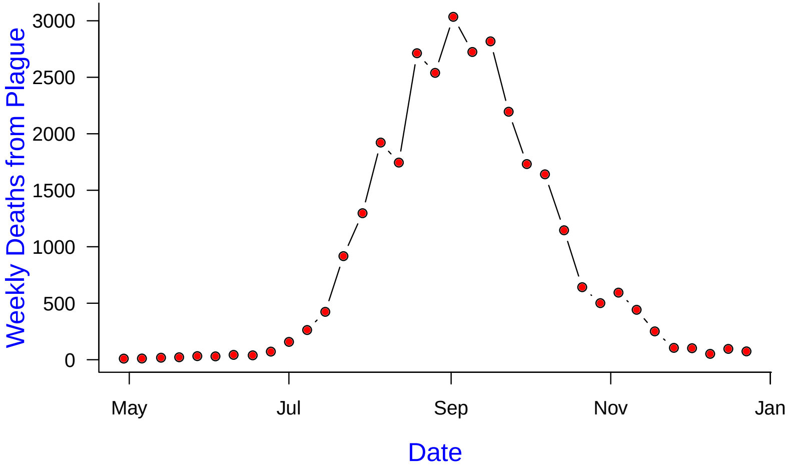
The Great Plague of London, 1665



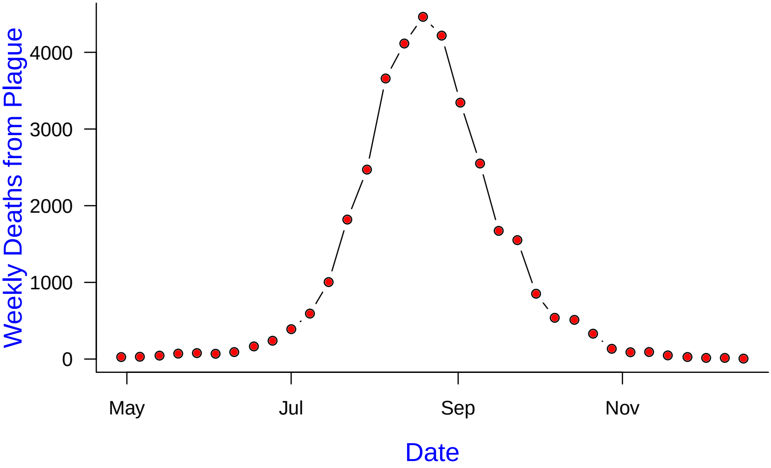
London Plague of 1593



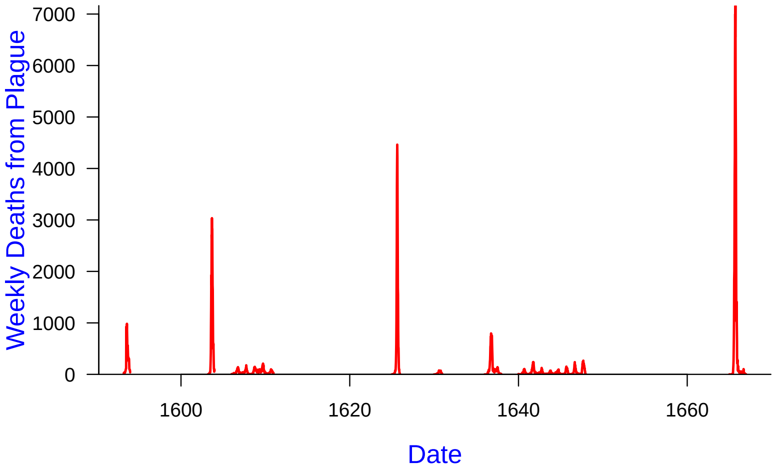
London Plague of 1603



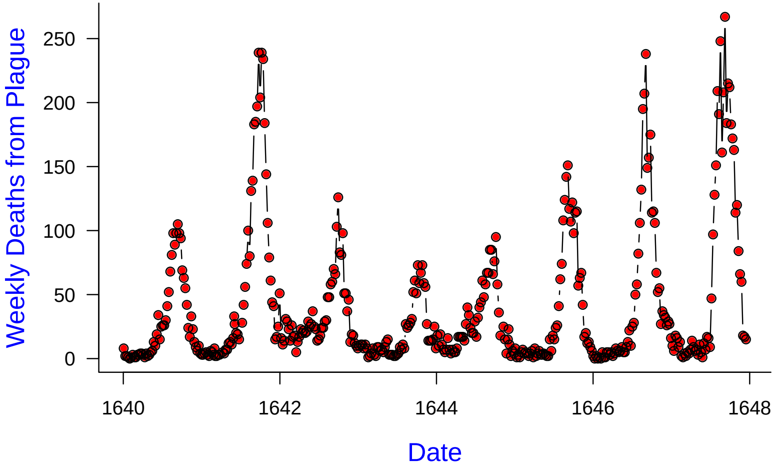
London Plague of 1625



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

McMaster University

Mathematics and Statistics

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

Mathematics 4MB3/6MB3 Mathematical Biology

Instructor: David Earn

Lecture 9
Epidemic Data II
Friday 26 Jan 2018

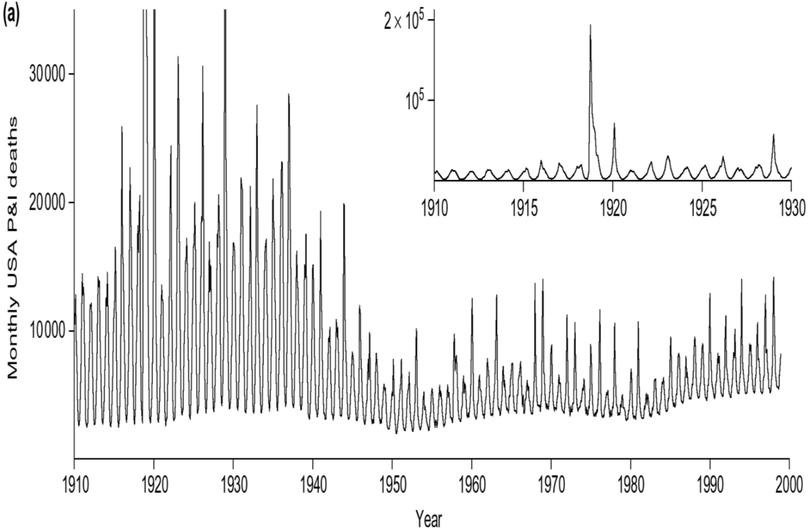
Announcements

- **Assignment 2:**
  Due Monday 5 February 2018 in class (and by e-mail) at 11:30am.
- **Midterm test:** We agreed on:
  - *Date:* Thursday 8 March 2018
  - *Time:* 7:00pm to 9:00pm
  - *Location:* TBA

Visualization of entire course of the Great Plague

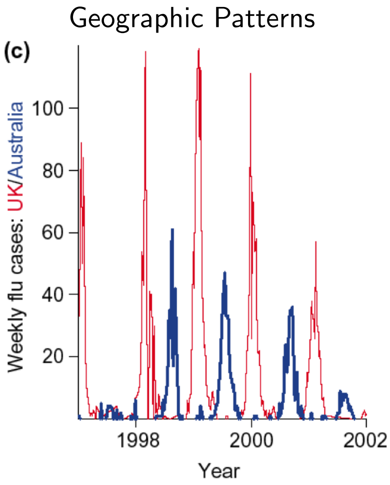
- What happenned after initial spatial spread?
- Visualize full spatial epidemic structure
- Show magnitude of epidemic in each parish with cylinder.
- Epidemic Visualization (`EpiVis`) software by Junling Ma.

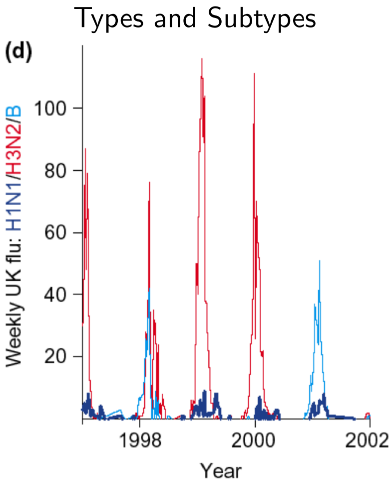
P&I mortality in U.S.A., 1910–1998



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

Influenza Incidence Patterns (lab confirmed)

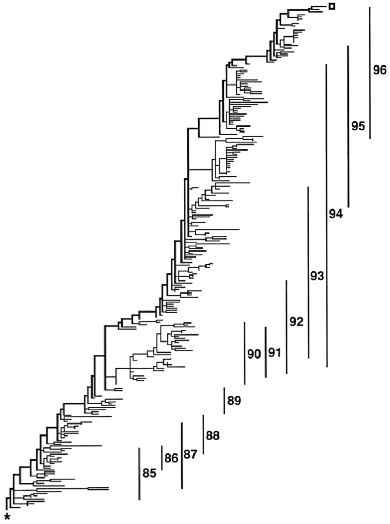




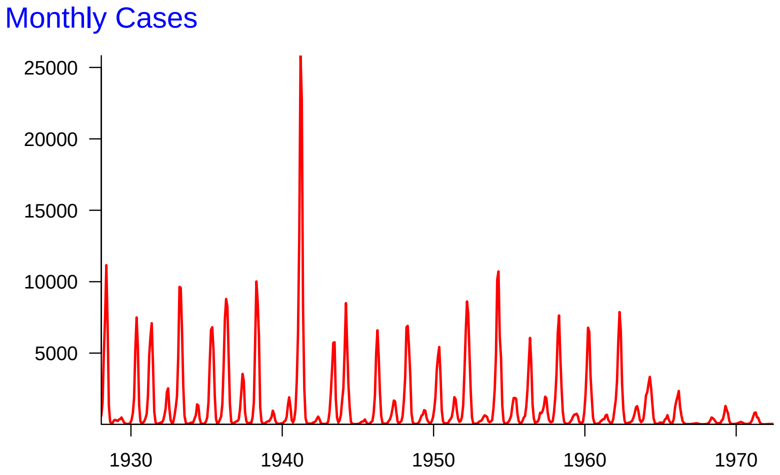
Earn, Dushoff & Levin 2002, *Trends in Ecology and Evolution* **17**, 334–340

Influenza Evolution

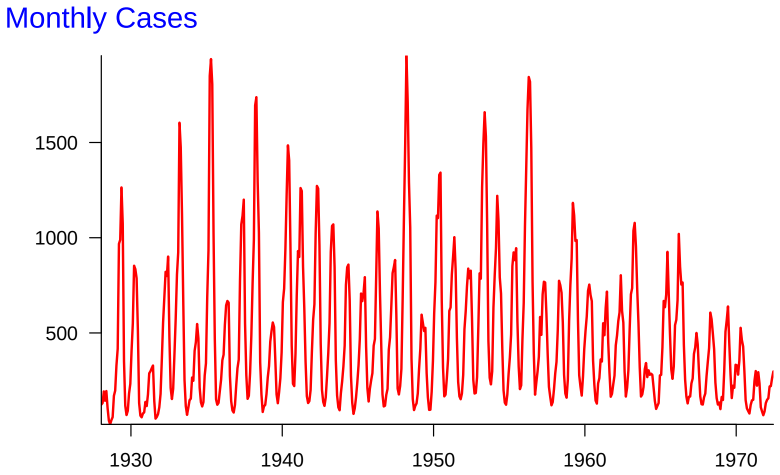
Molecular phylogenetic reconstruction of influenza A/H3N2 evolution, 1985–1996 (Fitch *et al.* 1997)



Measles in New York City, 1928–1972

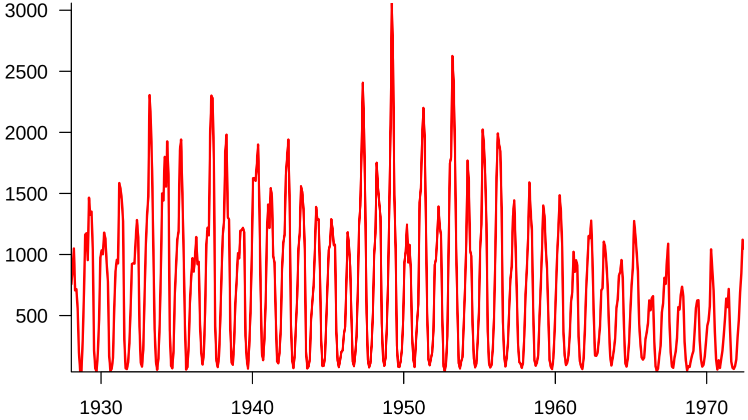


Mumps in New York City, 1928–1972

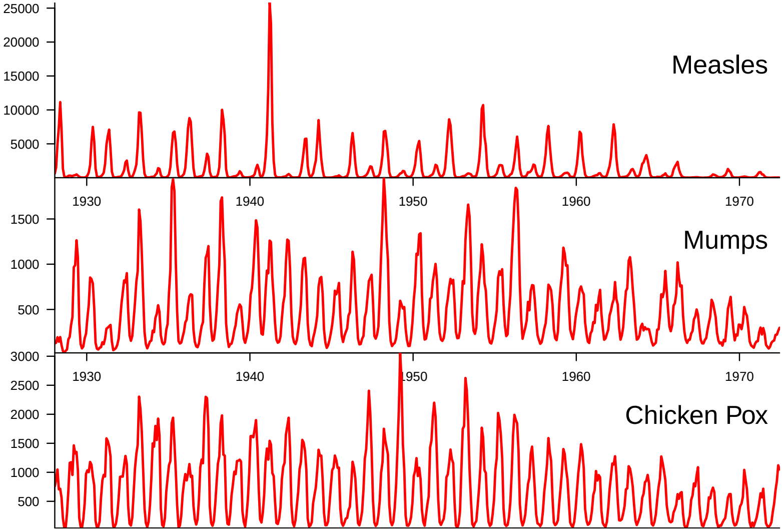


Chicken Pox in New York City, 1928–1972

Monthly Cases

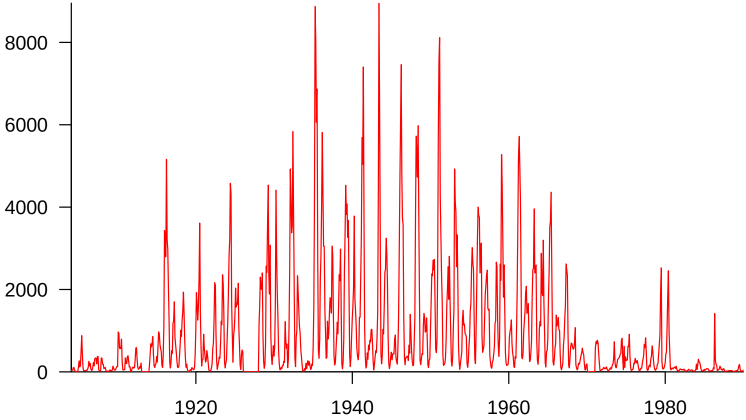


Childhood diseases in New York City, 1928–1972



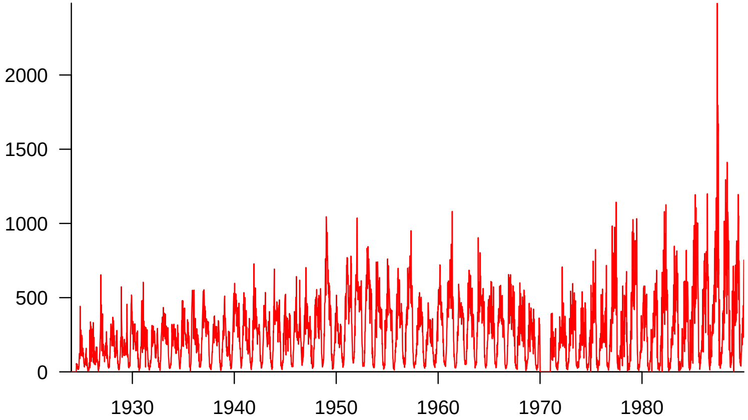
Measles in Ontario, 1904–1989

Monthly Cases



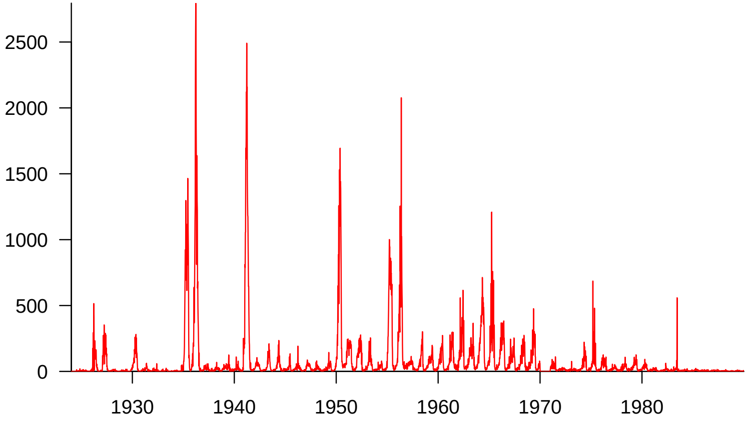
Chicken Pox in Ontario, 1924–1989

Monthly Cases



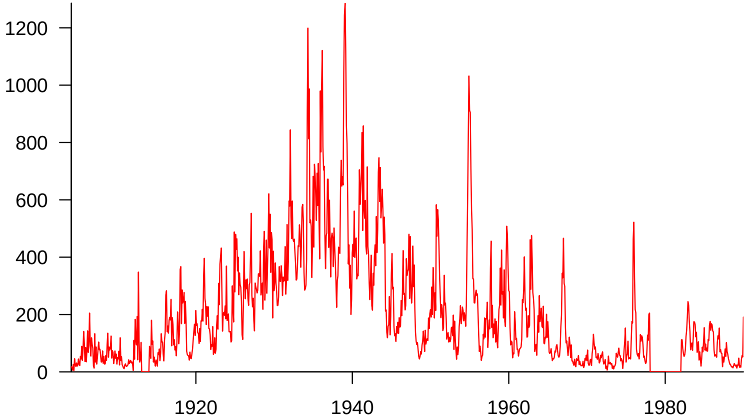
Rubella in Ontario, 1924–1989

Weekly Cases

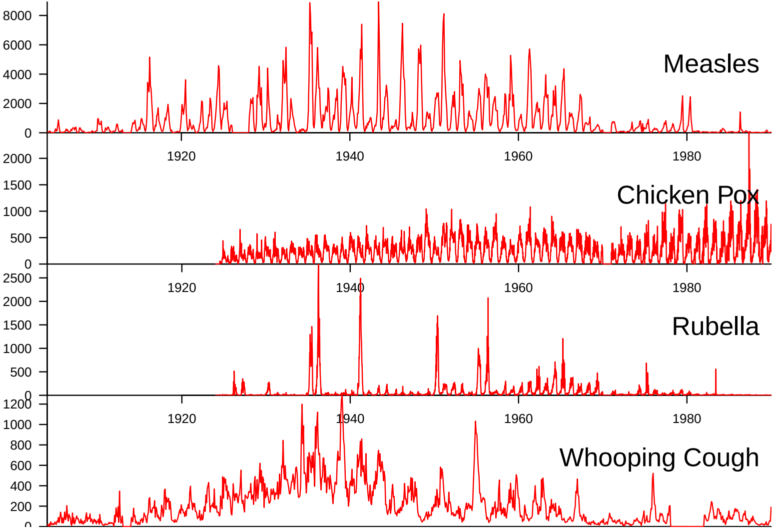


Whooping Cough in Ontario, 1904–1989

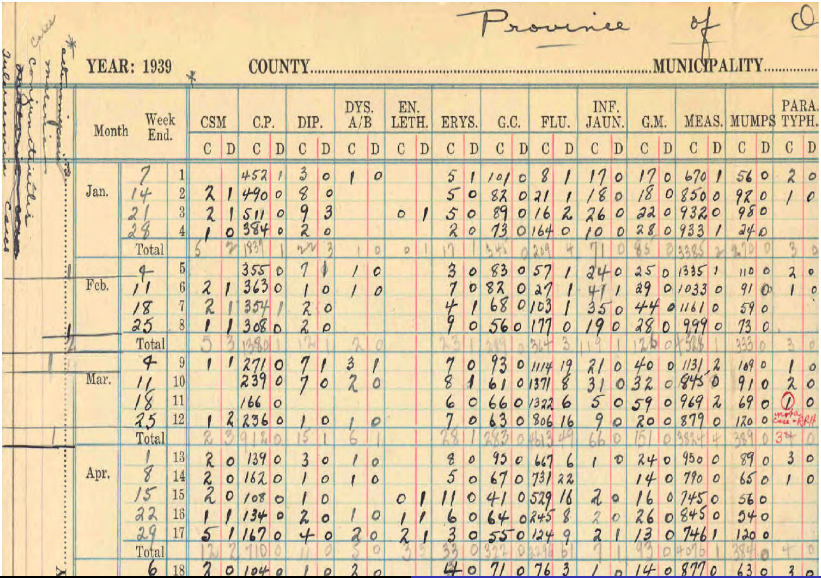
Monthly Cases



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	1839	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		
	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	

Mathematics
and Statistics

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

Mathematics 4MB3/6MB3
Mathematical Biology

Instructor: David Earn

Lecture 10
Epidemic Data III
Monday 29 Jan 2018

Announcements

- Comment from TA on Assignment 1:
  "For a few of the groups, I would recommend that they look over the work that their group members have done. Question 2c and 2d in particular were closely related and there were a few obvious cases where the students had not communicated with each other."

- **Assignment 2:**
  Due Monday 5 February 2018 in class (and by e-mail) at 11:30am.

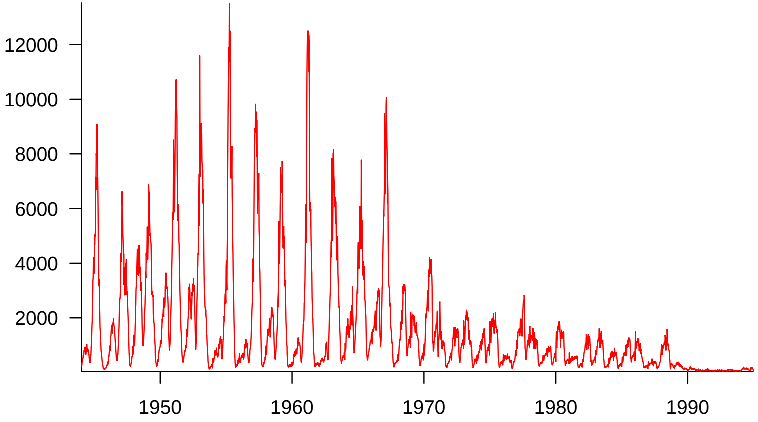
- **Midterm test:** We agreed on:
  - *Date:* Thursday 8 March 2018
  - *Time:* 7:00pm to 9:00pm
  - *Location:* TBA

Recurrent epidemics of childhood infections

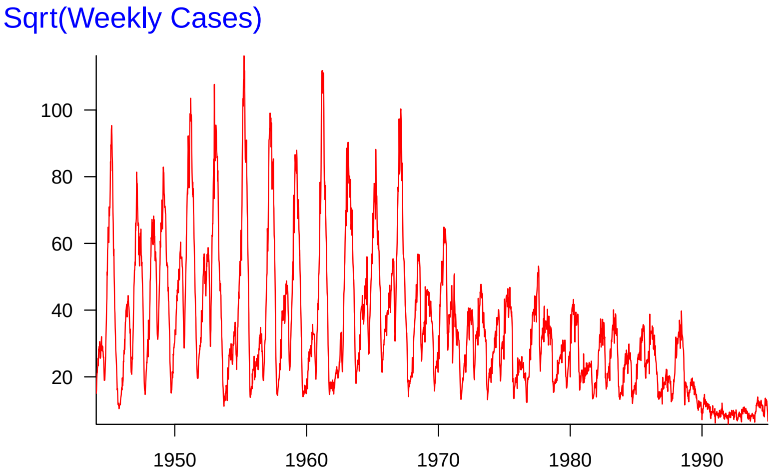
- Childhood diseases in New York City, 1928–1972
- Childhood diseases in Ontario, 1904–1989

Measles incidence in England and Wales, 1944–1995

Weekly Cases



Measles incidence in England and Wales, 1944–1995



Why study measles epidemics?

- $\sim 90,000$ children died from measles in 2016.
- 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

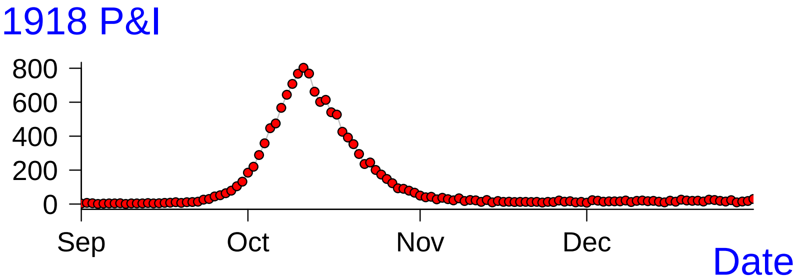
- Excellent data for human infectious diseases
  - Models can be tested!

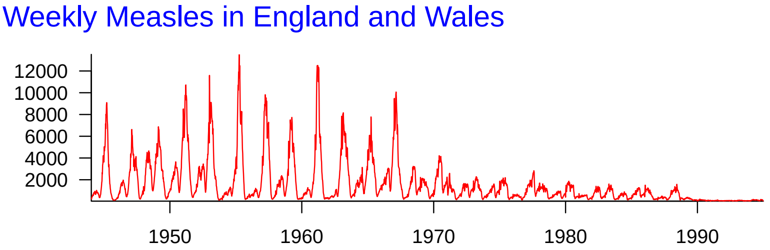
Modelling population dynamics childhood infections

- The basic SIR model cannot explain recurrent epidemics.
- What should we do?. . .
  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

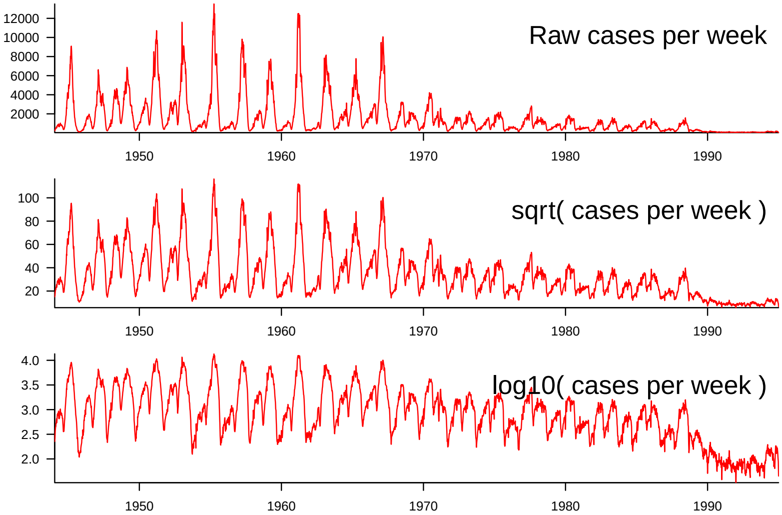
Time Plots of Temporal Epidemic Patterns





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:*

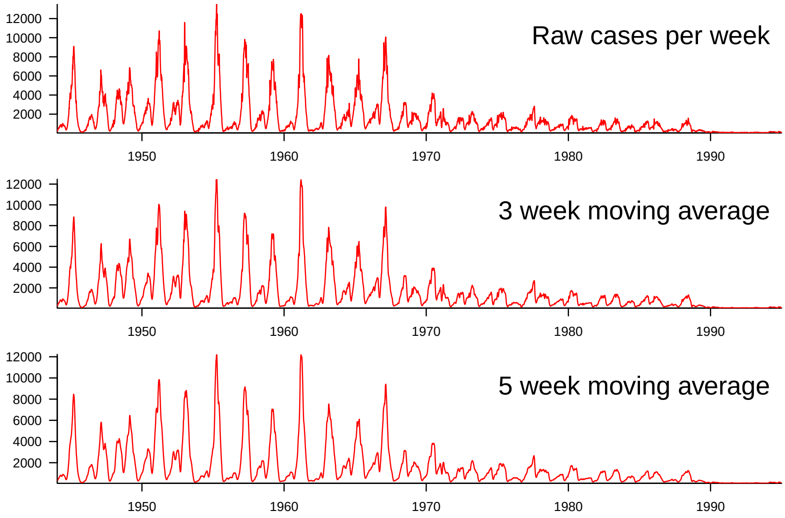
$$x_t \to \frac{1}{2a+1} \sum_{i=-a}^{a} x_{t+i}$$

  - Replace original data points x_t with averages of nearby points.
- *Linear filter:*

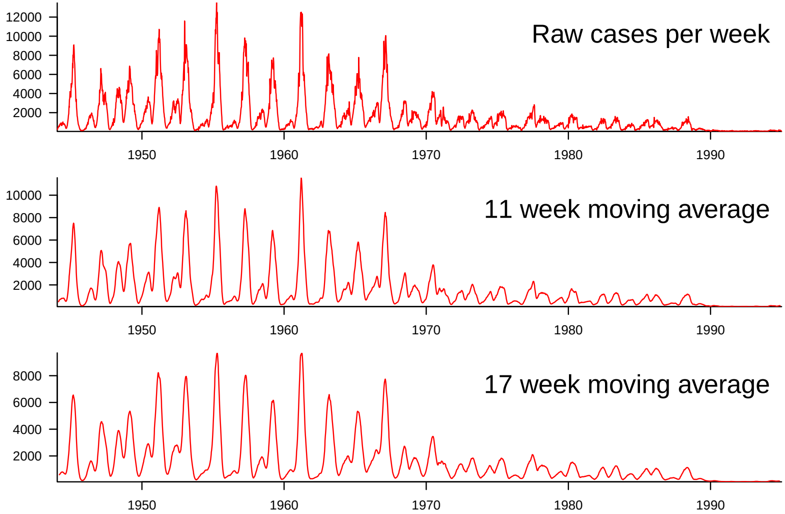
$$x_t \to \sum_{i=-\infty}^{\infty} \lambda_i x_{t+i}$$

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

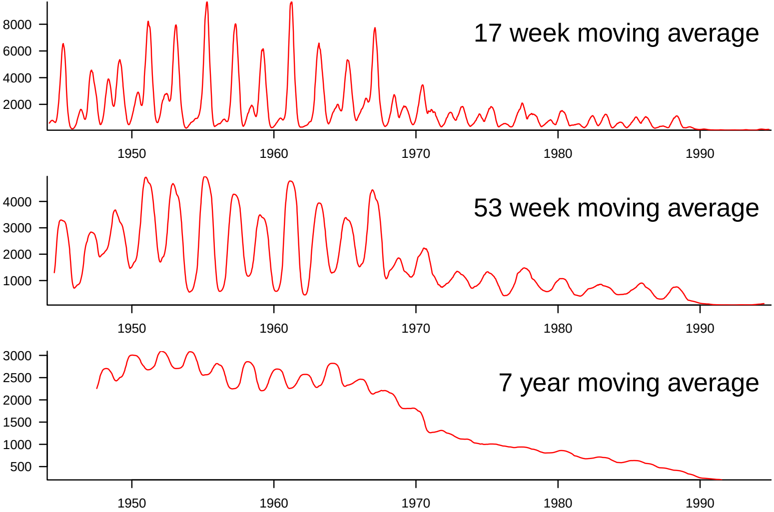
Times Plots of Smoothed Data



Times Plots of Smoothed Data



Times Plots of Smoothed Data



Correlation

- Recurrent epidemics $\implies$ number of cases now is correlated with number of cases in the past and the future.
- Given N pairs of observations of different quantities, $\{(x_i, y_i) : i = 1, \dots, N\}$, 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.

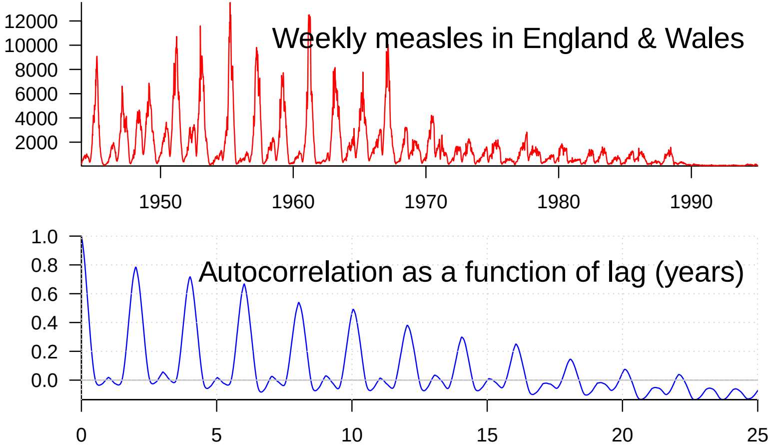
Autocorrelation

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

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

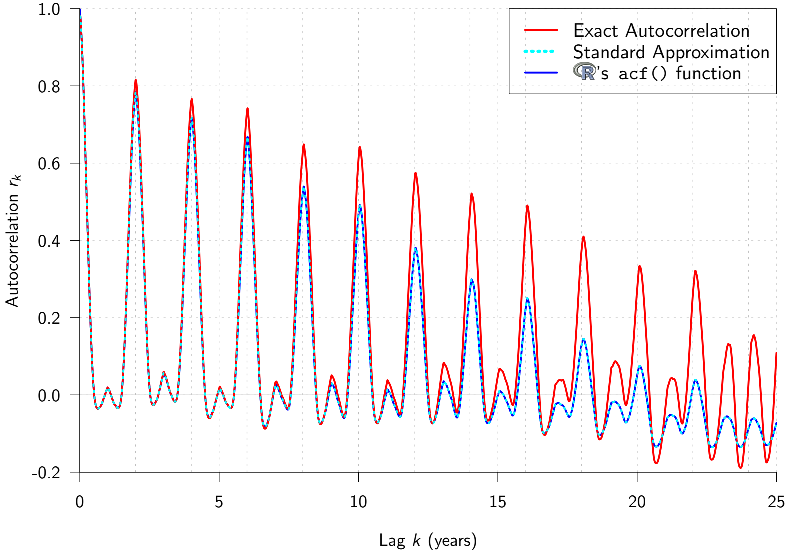
- Approximation of r_k is worse for larger lags k
- Plot of autocorrelation r_k as a function of lag k is called the *correlogram*.

Correlogram



- Peaks in correlogram $\implies$ 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.

Mathematics and Statistics

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

Mathematics 4MB3/6MB3 Mathematical Biology

Instructor: David Earn

Lecture 11
Epidemic Data Tools
Wednesday 31 Jan 2018

Announcements

- **Assignment 2:**
  Due Monday 5 February 2018 in class (and by e-mail) at 11:30am.
- **Midterm test:** We agreed on:
  - *Date:* Thursday 8 March 2018
  - *Time:* 7:00pm to 9:00pm
  - *Location:* TBA

Please consider. . .

5 minute *Student Respiratory Illness Survey:*

https://surveys.mcmaster.ca/limesurvey2/index.php/893454

> *Please complete this anonymous survey to help us monitor the patterns of respiratory illness, over-the-counter drug use, and social contact within the McMaster community. There are no risks to filling out this survey, and your participation is voluntary. You do not need to answer any questions that make you uncomfortable, and all information provided will be kept strictly confidential. Thanks for participating.*
>
> *–Dr. Marek Smieja (Infectious Diseases)*

Last time. . .

- Statistical description of time series:
  time plot, moving average
- Correlation coefficient: properties
- Autocorrelation
- Correlogram
- Exact vs. approximate autocorrelation
- Power spectral density (PSD)

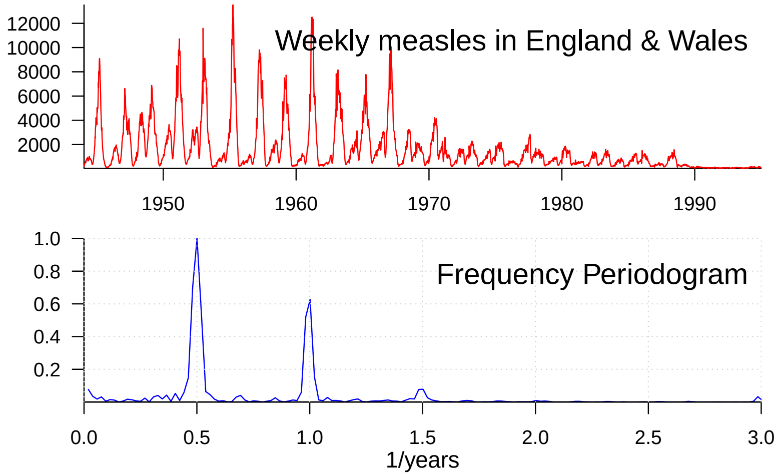
Spectral Density

- 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:

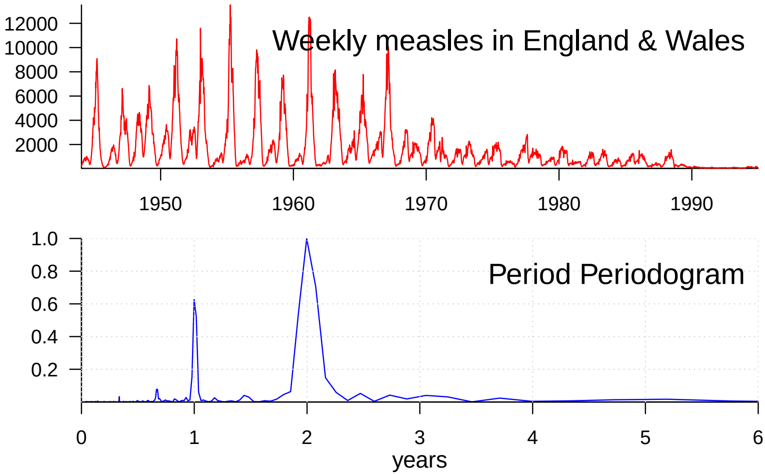
$$I(\omega_p) = \frac{1}{\pi}\left(r_0 + 2\sum_{k=1}^{N-1} r_k \cos \omega_p k\right)$$

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

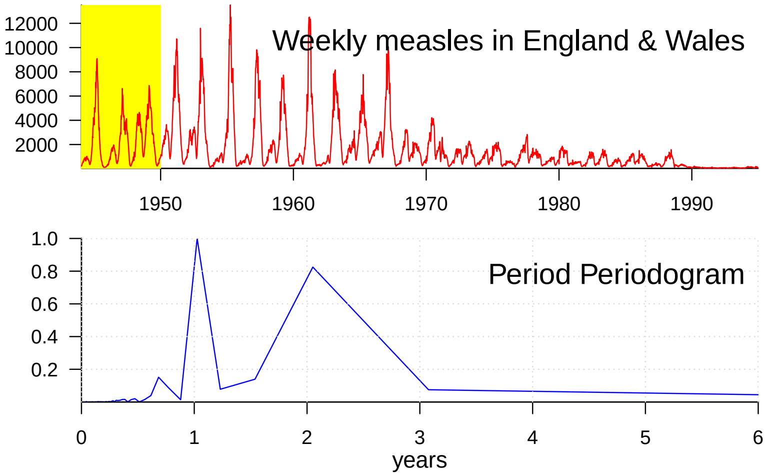
Spectral Density



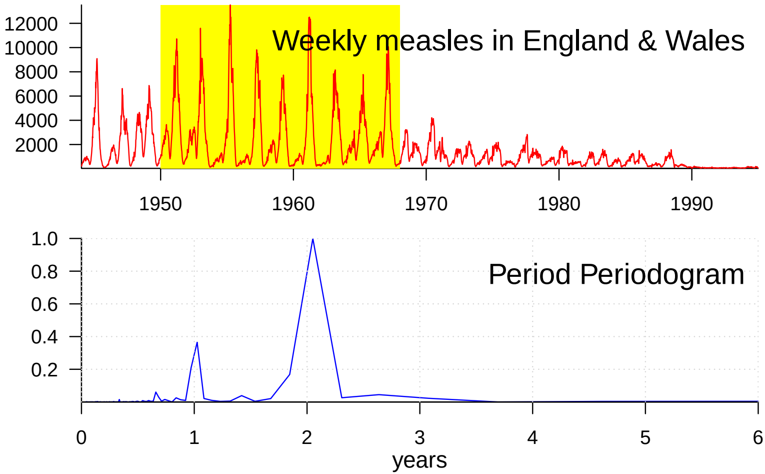
Spectral Density



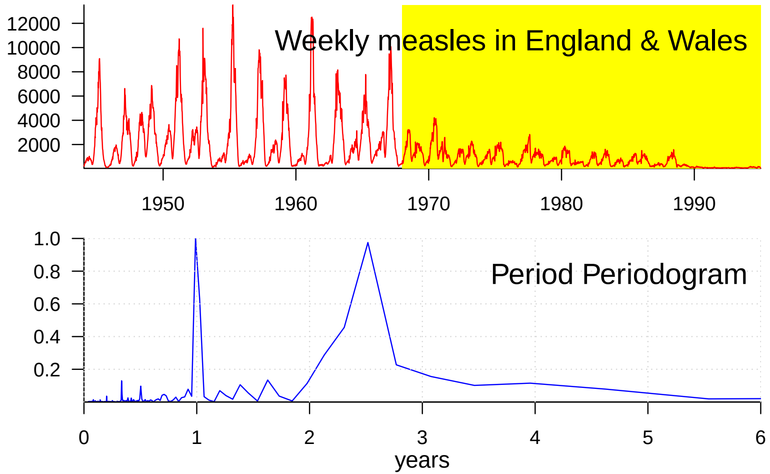
Spectral Density of Temporal Segments



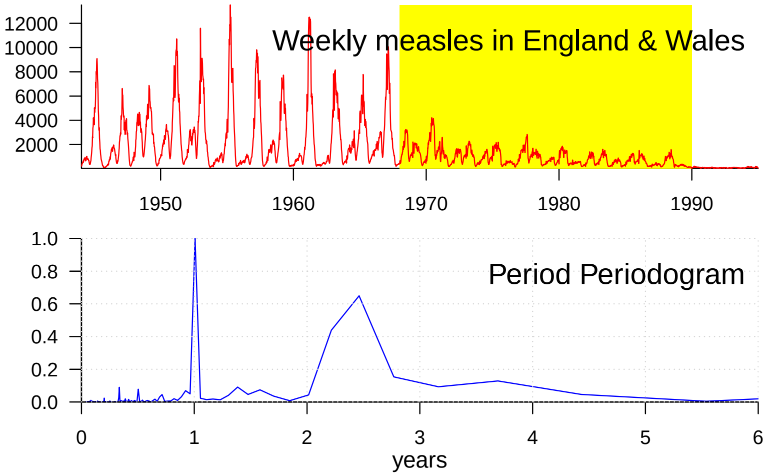
Spectral Density of Temporal Segments



Spectral Density of Temporal Segments



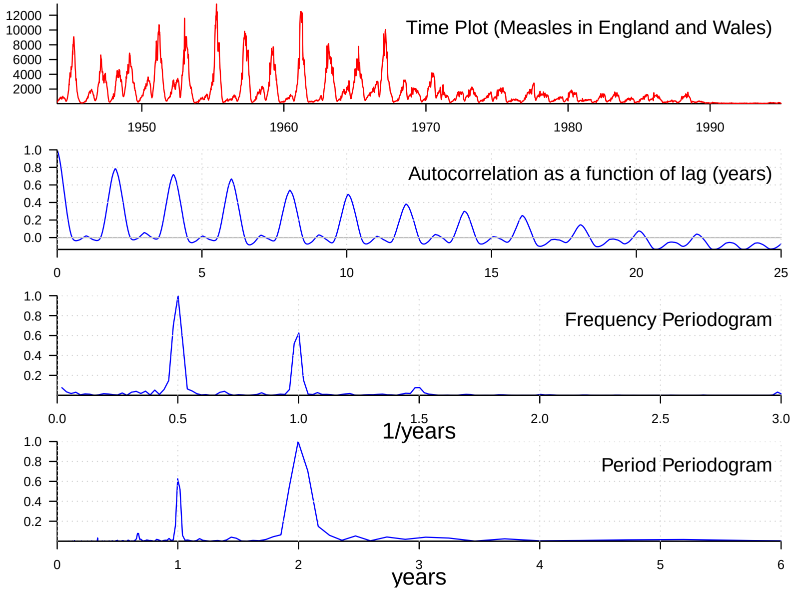
Spectral Density of Temporal Segments



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



Spectral Density of Temporal Segments

- Pre-war measles
- Post-war pre-vaccination measles
- Vacinnation era measles
- Vacinnation era measles until 1990

Time series analysis functions

R has built-in tools for time series analysis:

- Time plot: `plot()` *etc.*
- Linear filter (*e.g.*, moving average): `filter()`
- Correlogram (auto-correlation function): `acf()`
- Periodogram (power spectrum): `spectrum()`

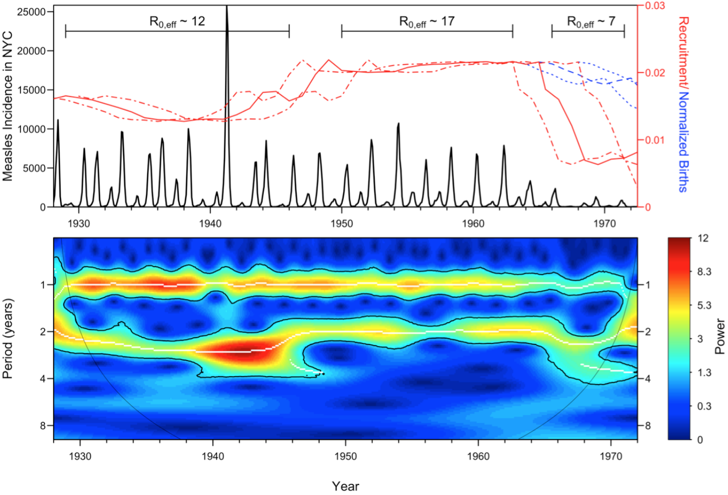
You will use all of these functions in **Assignment 4.**

More sophisticated spectral method

- Traditional power spectrum measures frequency content of entire time series.

- Wavelet decomposition is local in time.
  - Reveals changes in the spectrum over time without having to identify distinct temporal segments yourself.
  - Nice intro to wavelet analysis of time series:
    Torrence and Compo (1998) "A Practical Guide to Wavelet Analysis" *Bulletin of the American Meteorological Society* **79**, 61–78
  - $\exists$ R packages for wavelet analysis of time series (*e.g.*, WaveletComp, wavelets), and at least one book on wavelet methods in R

Wavelet Spectrum of Monthly Measles in New York City



Krylova & Earn 2013, *J. R. Soc. Interface* **10**, 20130098

Wavelet Spectrum of Weekly Measles in New York City

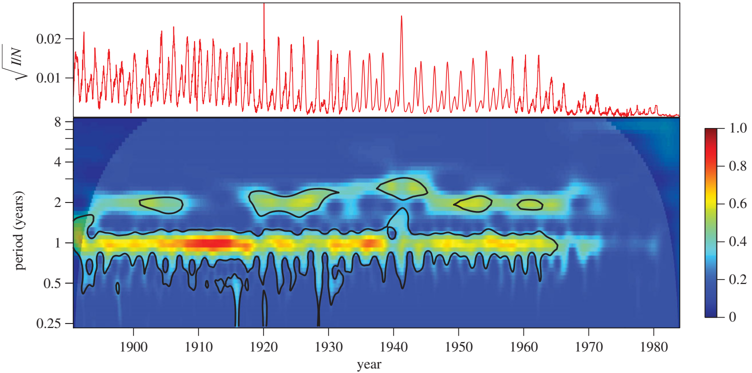


Figure 5. Observed measles dynamics in NYC from 1891 to 1984. (*a*) Square root of measles case reports, normalized by total concurrent population. (*b*) Colour depth plot of a continuous wavelet transform of the square root of normalized observed NYC measles cases (colour warmth scales with spectral power and 95% significance contours are shown in black). Shaded regions in the upper left and right indicate the cone of influence.

Hempel & Earn 2015, *J. R. Soc. Interface* **12**, 20150024

Statistical Modelling of Time Series

Statistical Modelling of Time Series

- Imagine time series $\{X_t\}$ is generated by random processes.
- Simplest case: X_t (number of cases at time t) is simply a random variable with a known distribution,

$$X_t = \mu + Z_t \qquad (*)$$

  where $\mu =$ time average number of cases
  and $\{Z_t\} =$ sequence of random variables with zero mean.
- Might be a reasonable model for importation of new, infectious individuals into a focal community.
- Bad model for epidemics: ignores transmission from one individual to another.
  - There must be a correlation between the number of individuals in the focal community who are infected now and the number who will be infected in the near future.

Statistical Modelling of Time Series: AR and MA

- So, imagine that that successive data points in $\{X_t\}$ are correlated.
- For example, perhaps the data are generated by an *autoregressive (AR) process*:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + Z_t \,,$$

where the α_i are constants that determine the degree of correlation along the time series.

- Alternatively, the data might be generated by a *moving average (MA) process*:

$$X_t - \mu = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \,,$$

where the β_i are constants that define a weighted average.

Statistical Modelling of Time Series: ARMA

- More generally, the data might be generated by an *autoregressive moving average "ARMA(p,q)" process*:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) \\ + \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \,.$$

Statistical Modelling of Time Series: ARIMA

- Finally, an *autoregressive integrated moving average "ARIMA(p,d,q)" model* includes weighted differences of the time series:

$$\begin{aligned} X_t - \mu = {} & \alpha_1(X_{t-1}-\mu) + \alpha_2(X_{t-2}-\mu) + \cdots + \alpha_p(X_{t-p}-\mu) \\ & + \gamma_1(X_{t-1}-X_{t-2}) + \gamma_2(X_{t-2}-X_{t-3}) + \cdots \\ & + \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}\,. \end{aligned}$$

  - The "I" in ARIMA refers to the original time series X_t, which is an "integrated" version of the differenced time series.
  - Technically, an ARIMA model is just an ARMA model with differently labelled coefficients, but explicit differences are often helpful conceptually (*e.g.*, they can "stationarize" a time series).

What kind of process generated our data?

- *How can we tell if our data were generated by such a process? Can we identify an AR(p), MA(q) or ARMA(p, q) process?*

- Compare time plots of these processes with time plot of our data? (Comparison by eye often challenging/unreliable.)
- Compare autocorrelation functions (correlograms) of these processes with correlogram of our data? (Better.)
- Compare power spectra (periodograms) of these processes with periodogram of our data? (Even better.)
- Compare wavelet spectra of these processes with wavelet spectrum of our data? (Better yet.)

Statistical Modelling of Time Series: ARMA fitting

- Looking at the power spectra of ARMA models would be instructive.
- But is there a better approach to discovering if an ARMA model could explain our data?
- Find the *best fit* ARMA parameters by minimizing the residual sum of squares. *e.g.*, for an AR model, minimize:

$$S = \sum_{t=p+1}^{N} \left[(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \cdots - \alpha_p(x_{t-p} - \mu) \right]^2 .$$

- More generally, we can find the best fit parameters of an ARIMA(p, d, q) model
  - Non-trivial, but there are standard methods
- Compare models with Akaike Information Criterion (AIC), which penalizes models that have more parameters
  - See Earn (2009) review article for more discussion of this.

McMaster University

Mathematics and Statistics

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

Mathematics 4MB3/6MB3 Mathematical Biology

Instructor: David Earn

Lecture 12
Epidemic Data Tools II
Friday 2 Feb 2018

Announcements

- **Assignment 2:**
  Due Monday 5 February 2018 in class (and by e-mail) at 11:30am.
- **Midterm test:** We agreed on:
  - *Date:* Thursday 8 March 2018
  - *Time:* 7:00pm to 9:00pm
  - *Location:* BSB-B154

Time series tools discussed so far. . .

- Statistical description of time series:
  time plot, moving average, correlation coefficient, autocorrelation, correlogram, power spectral density (PSD), periodogram, wavelet spectrum
- Time series models:
  AR, MA, ARMA, ARIMA

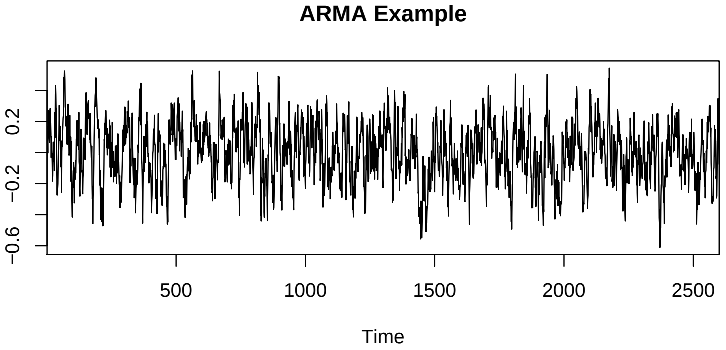
Statistical Modelling of Time Series

How to do it in R . . .

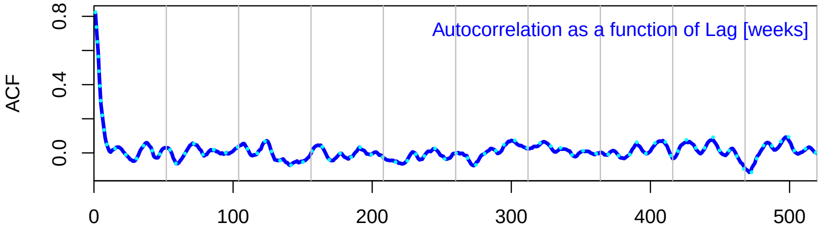
- Simulate any ARIMA(p, d, q) model with `arima.sim()`
- Fit an AR model to a time series with `ar()`
- Fit an ARIMA model to a time series with `arima()`
- Alternatively, there are specialized time series modelling packages.

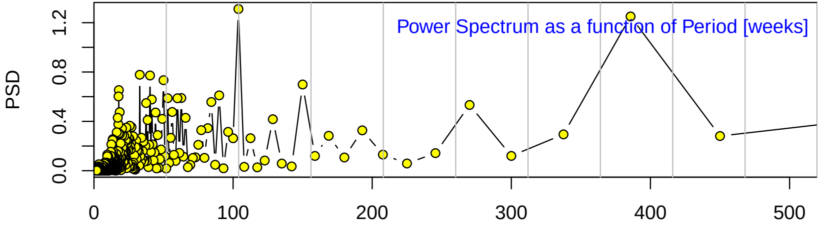
ARMA Example (50 years of weekly data)

```
my.model <- list(ar=c(1,-0.5,0.5,-0.25),ma=c(-0.25,0.5))
my.sim <- arima.sim(n=52*50,model=my.model,sd=0.1)
plot(my.sim,main="ARMA Example",ylab="",xaxs="i")
```



ARMA Example (ACF and PSD up to 10 year lag)





Statistical Modelling of Time Series: Forecasting

- Once we have a fitted model, we can then use it to *forecast* future observations
- *Validate* this procedure by using part of the data to fit the model and then forecast the remainder of the data (*cf.* cross-validation)

- How successful is this likely to be for an infectious disease time series?
  - Conceivably good for chicken pox in NYC.
  - Not for measles. . . at least not for the main patterns. . .

Statistical Modelling of Time Series: Limitations

- Best to remove mean, trend and seasonality before fitting an ARMA model
  - But this means we will remove the aspects of the data about which we care most!
- The fitted parameters of an ARMA model have no obvious biological meaning
  - The model completely ignores any understanding we have of infectious disease transmission
- Statistical models use the time series itself to parameterize an ARMA (or more general) process
  - It would be better to have a model that we can parameterize from independently collected data and then see if that model can explain the observed time series

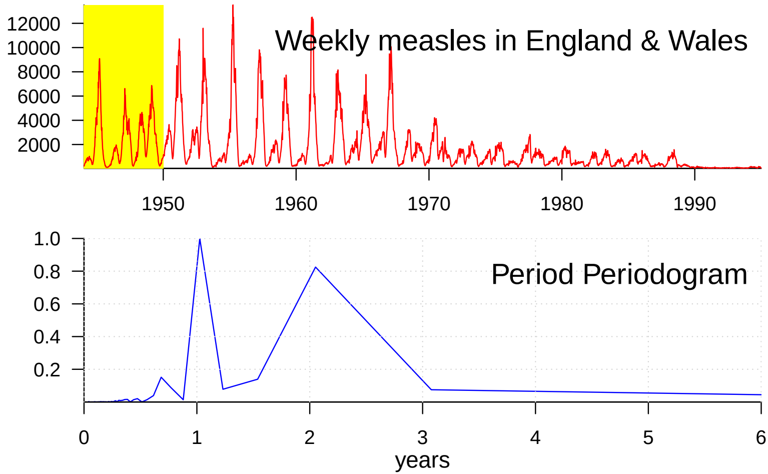
Mechanistic Mathematical Modelling

- SIR and all that...
- Takes into account transmission process...

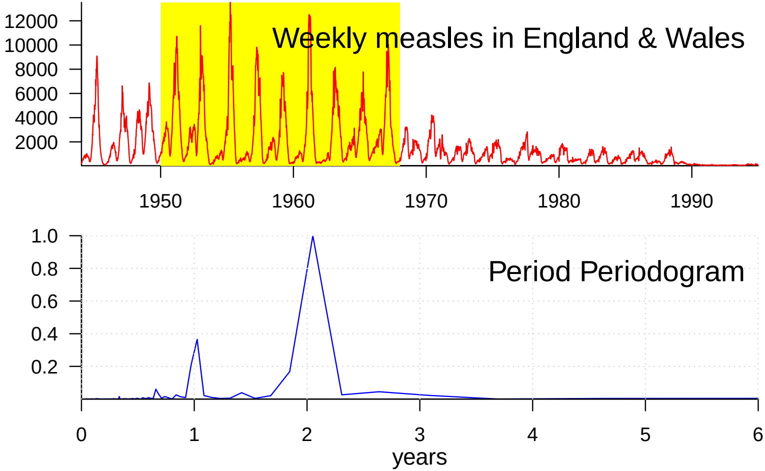
- So why did we just spend time talking about statistical modelling?
  - Important to be familiar with time series models that are in common use.
  - Helps us appreciate the value of mechanistic modelling.
  - Some processes that affect disease dynamics might be better modelled as ARMA or similar processes.
    - Weather (*e.g.*, perhaps model $\beta = \beta(t)$ as an ARMA process)
    - Immigration
  - Ruling out an ARMA model (or at least one with a modest number of parameters) is a step towards finding a good model.

Mechanistic Modelling of Recurrent Epidemics

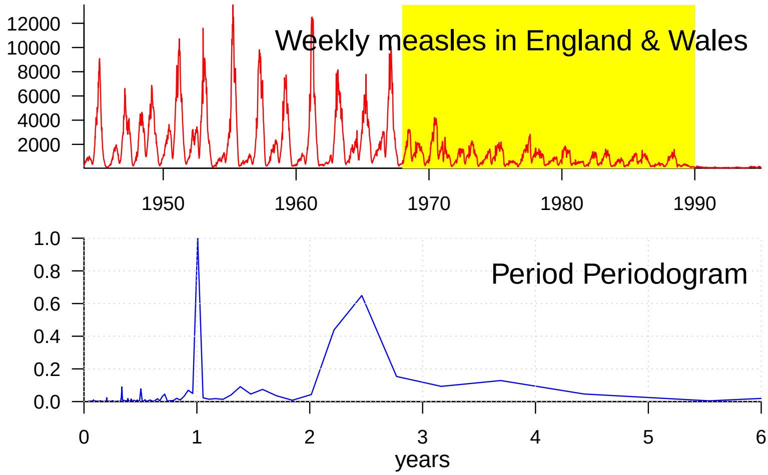
What causes changes in frequency content over time?



What causes changes in frequency content over time?

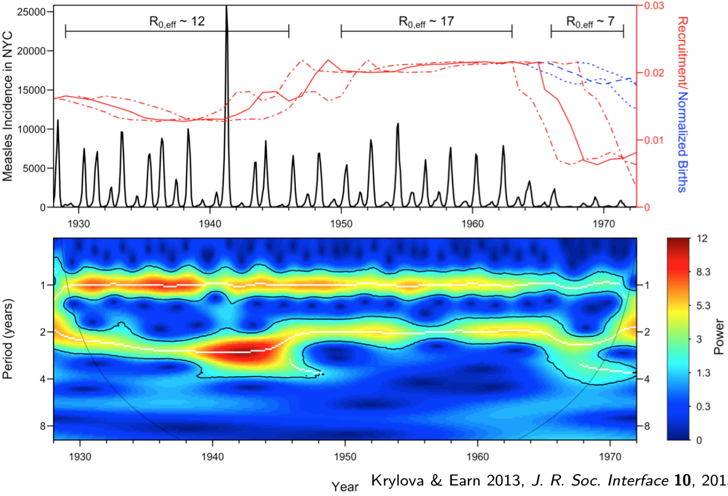


What causes changes in frequency content over time?



What causes changes in frequency content over time?

Measles in New York City

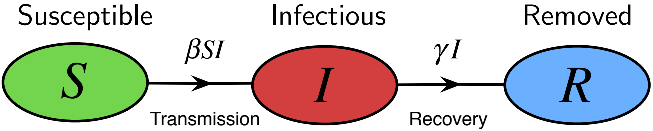


Krylova & Earn 2013, *J. R. Soc. Interface* **10**, 20130098

Mechanistic Epidemic Modelling: Principles

- Consider the biological mechanisms involved in disease transmission and spread
- Model mechanisms and infer their effects
- Start as simple as possible!
- Rule out simple models by comparing results with observed time series of incidence or mortality
- Add complexity one step at a time, so key mechanisms can be identified
- Ideally converge on simplest possible model that can explain observed patterns

The SIR model: Flow Chart and Parameters



$$\frac{dS}{dt} = -\beta SI$$

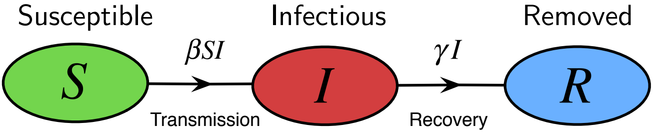
$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Parameters:**
  - Transmission rate β
  - Recovery rate γ

    (or Removal rate)

The SIR model: Derived Parameters



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Derived Parameters:**
  - Initial growth rate $\beta - \gamma$
  - Mean infectious period $\frac{1}{\gamma}$
  - Basic Reproduction Number

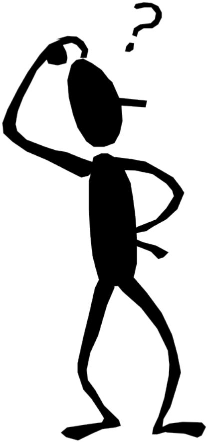
$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

Basic SIR Model: Important Results

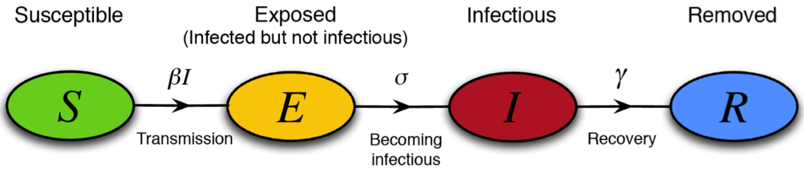
- Epidemic occurs if and only if $\mathcal{R}_0 > 1$
- Exact solution for phase portrait
- Single epidemic, then disease disappears
- Exact formula for final size as a function of $\mathcal{R}_0$

- Cannot explain diseases that persist
- Cannot explain recurrent cycles of epidemics

What are we missing?

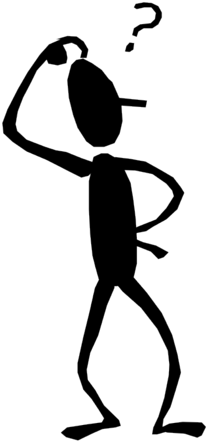


SEIR Model: flow chart

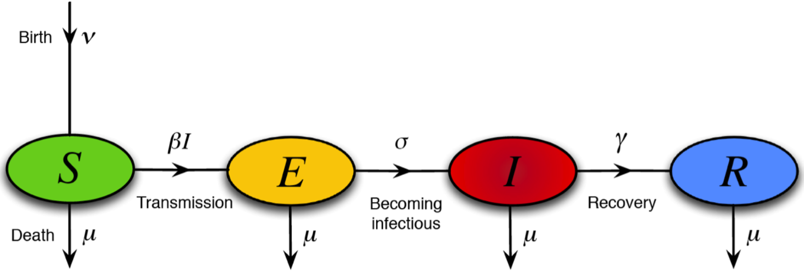


- Introduces only one new parameter (σ)
- Mean latent period ($1/\sigma$) can often be estimated
- But... effect of inclusion of exposed class usually small

What are we still missing?



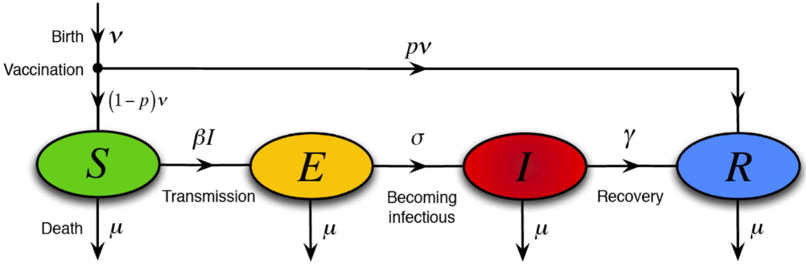
SEIR Model with vital dynamics: flow chart



New Parameters:

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Mean latent period ($1/\sigma$)

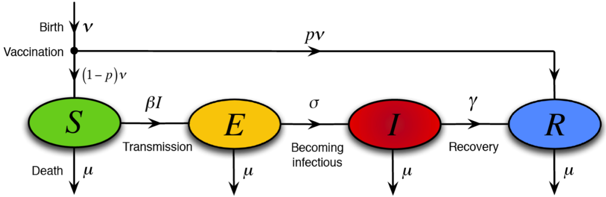
SEIR with vital dynamics and vaccination: flow chart



New Parameters:

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Mean latent period ($1/\sigma$)
- Proportion vaccinated (p)

SEIR with vital dynamics and vaccination: Equations



$$\frac{dS}{dt} = \nu(1-p) - \beta SI - \mu S$$

$$\frac{dE}{dt} = \beta SI - \sigma E - \mu E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I$$

$$\frac{dR}{dt} = \nu p + \gamma I - \mu R$$

- Birth rate (ν for natality)
- Death rate (μ for mortality)
- Proportion vaccinated (p)

- Transmission rate (β)
- Mean latent period ($1/\sigma$)
- Mean infectious period ($1/\gamma$)

SEIR with vital dynamics and vaccination: Analysis

- $\mathcal{R}_0$?
  - Biological derivation: (assuming $\nu = \mu$ and $p = 0$)
    $\mathcal{R}_0 = \beta \times \frac{\sigma}{\sigma+\mu} \times \frac{1}{\gamma+\mu} \quad \simeq \frac{\beta}{\gamma} \quad \because \frac{1}{\mu} \gg \max\left(\frac{1}{\sigma}, \frac{1}{\gamma}\right)$
  - Mathematical derivation:
    $\mathcal{R}_0 = 1$ is stability boundary
- Final size ? Not well defined (because of continuous source of new susceptibles).
- Equilibria ?