Identifying Signature Features of Epidemic Diseases in 19th Century All-cause Mortality Data

Rasmus Kristoffer Pedersen

Postdoc, PandemiX Center
Dept. Science and Environment,
Roskilde University, Denmark
Email: rakrpe@ruc.dk

Joint work with
Mathias Mølbak Ingholt, Maarten van Wijhe,
Viggo Andreasen & Lone Simonsen

Epidemics 9, November 30th, 2023

Roskilde University

PANDEMIX CENTER

Danmarks Grundforskningsfond
Danish National Research Foundation

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.
- Our response to threats depend on our experience with past threats.

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.
- Our response to threats depend on our experience with past threats.
  - SARS-CoV-2 - Experiences from influenza pandemics and SARS-CoV-1.

Introduction

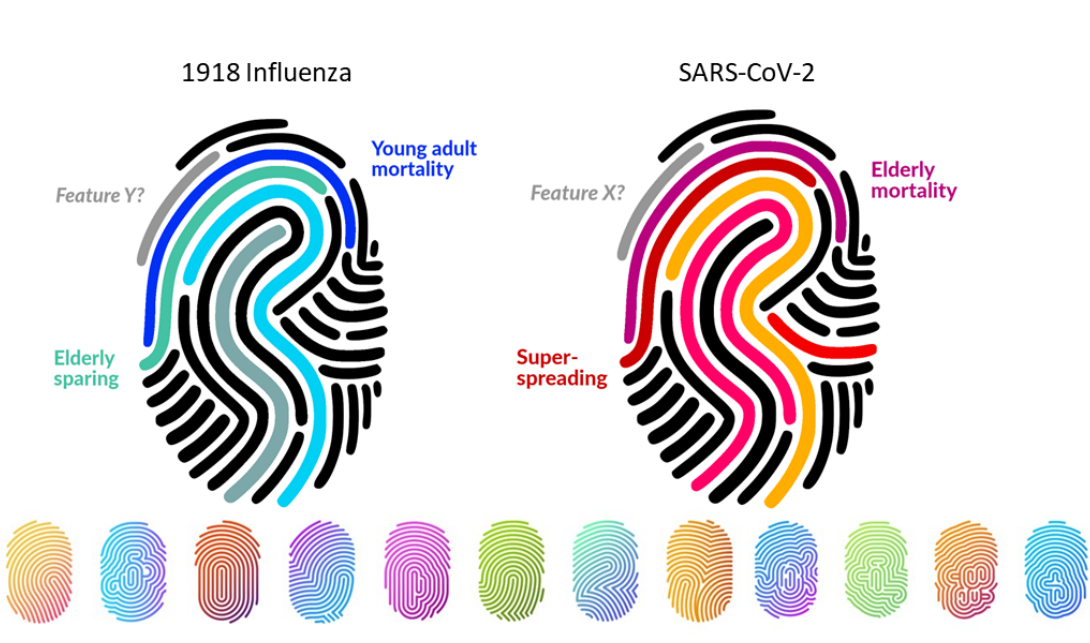
- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.
- Our response to threats depend on our experience with past threats.
  - SARS-CoV-2 - Experiences from influenza pandemics and SARS-CoV-1.
  - MPox - Experiences from small pox.

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.
- Our response to threats depend on our experience with past threats.
  - SARS-CoV-2 - Experiences from influenza pandemics and SARS-CoV-1.
  - MPox - Experiences from small pox.
  - Scarlet fever - Experiences from historical scarlet fever.

Introduction

- Lack of data is a challenge for surveillance of emerging diseases.
- The epidemics and pandemics of recent years may only be a subset of potential threats to consider.
- Our response to threats depend on our experience with past threats.
  - SARS-CoV-2 - Experiences from influenza pandemics and SARS-CoV-1.
  - MPox - Experiences from small pox.
  - Scarlet fever - Experiences from historical scarlet fever.
- Analyzing historical pandemics is a way to study infectious diseases in more details than the study of modern pandemics alone allow for.


1918 Influenza
Young adult mortality
Feature Y?
Elderly sparing
SARS-CoV-2
Feature X?
Elderly mortality
Super-spreading

Overview of talk

In this talk, I will talk about:

- Our recent study of a unique dataset of all-cause mortality.

Overview of talk

In this talk, I will talk about:

- Our recent study of a unique dataset of all-cause mortality.
- Mortality baseline calculations.

Overview of talk

In this talk, I will talk about:

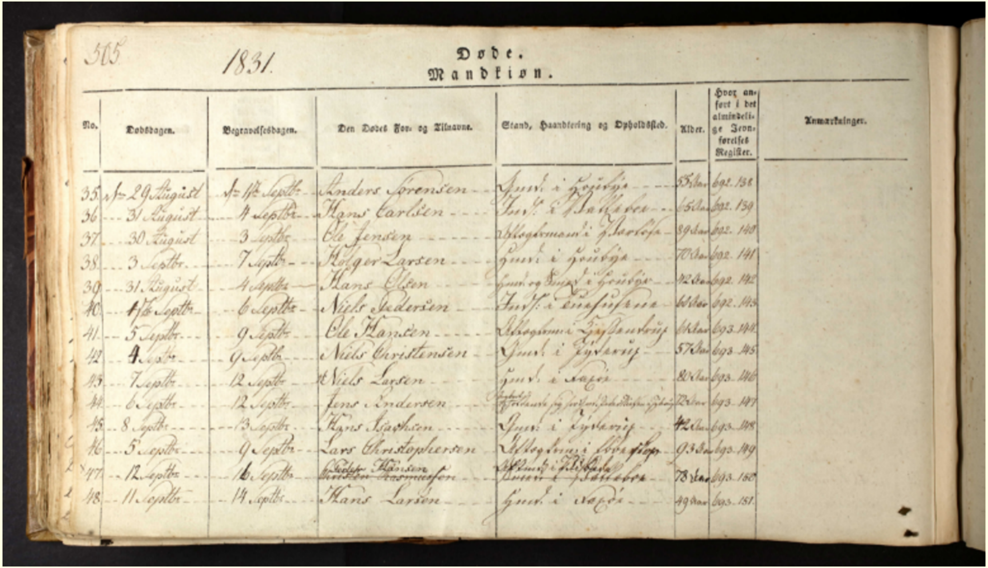
- Our recent study of a unique dataset of all-cause mortality.
- Mortality baseline calculations.
- Analyzing age patterns to classify epidemics.

Overview of talk

In this talk, I will talk about:

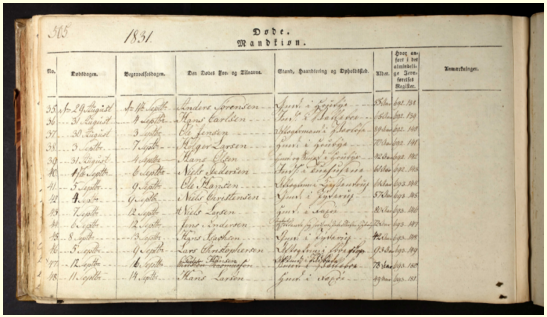
- Our recent study of a unique dataset of all-cause mortality.
- Mortality baseline calculations.
- Analyzing age patterns to classify epidemics.
- Reflect on what we can learn through historical epidemiology.

Data source



Data source

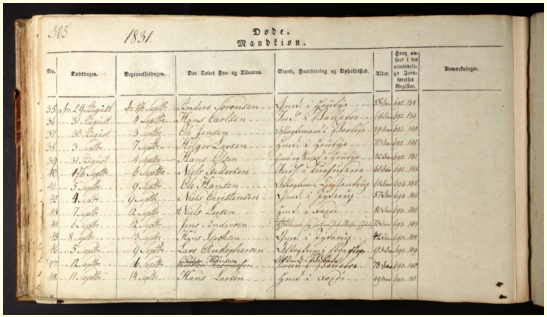
- Parish registers for Danish church parishes
  between 1815-1915



Scan of parish register for "Fakse" parish.

Data source

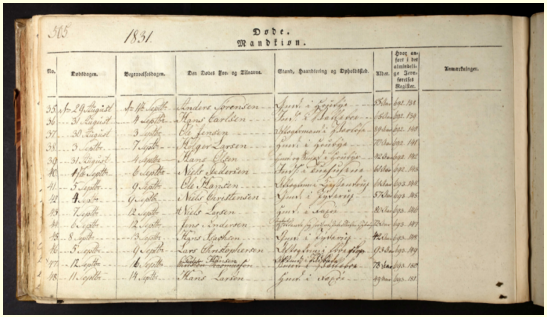
- Parish registers for Danish church parishes
  between 1815-1915
- Approximately 4 million burials



Scan of parish register for "Fakse" parish.

Data source

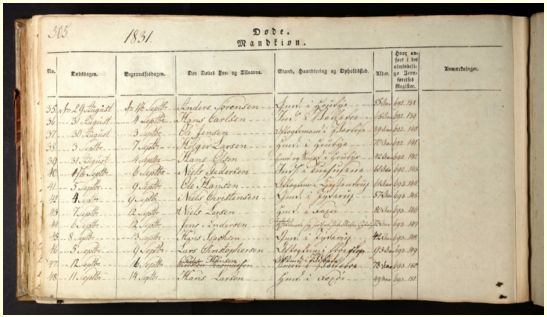
- Parish registers for Danish church parishes
  between 1815-1915
- Approximately 4 million burials
- Individual level information
  Includes date, age, sex and parish



Scan of parish register for "Fakse" parish.

Data source

- Parish registers for Danish church parishes
  between 1815-1915
- Approximately 4 million burials
- Individual level information
  Includes date, age, sex and parish
- Property of the Danish National Archives, but digitized and transcribed by *Ancestry*



Scan of parish register for "Fakse" parish.

Data cleaning and managing

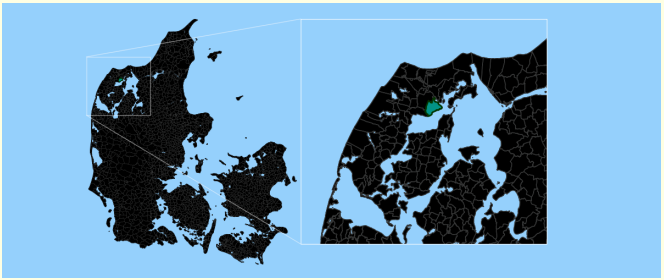
Date of burial	Name	Age	Gender	Amt	Sogn
1857-01-02	Ane Kirstine Christensen	2	Female	Thisted Amt	Jannerup Sogn (Thisted Amt)
1857-01-02	Birthe Marie Christensen.	0	Female	Thisted Amt	Hundborg Sogn
1857-01-02	Ane Marie Mortensdatter	81	Female	Thisted Amt	Vejerslev Sogn (Mors)
1857-01-02	Gjertrud Jensdatter	82	Female	Thisted Amt	Thisted Sogn
1857-01-02	Karen Christensdatter Wiilsbøll	52	Female	Thisted Amt	Vester Vandet Sogn
1857-01-02	Karen Marie Jensen	21	Female	Thisted Amt	Sennels Sogn
1857-01-02	Ane Christensdatter Krogh	76	Female	Thisted Amt	Vester Vandet Sogn
1857-01-03	Anders Hansen Tüylfang	79	Male	Thisted Amt	Gøttrup Sogn
1857-01-03	Marcus Christensen	6	Male	Thisted Amt	Hunstrup Sogn
1857-01-04	Maren Jensen	0	Female	Thisted Amt	Hillerslev Sogn (Thisted Amt)
1857-01-04	Thomas Jensen	0	Male	Thisted Amt	Flade Sogn (Thisted Amt)
1857-01-04	Niels Madsen Thÿstrup	74	Male	Thisted Amt	Skjoldborg Sogn
1857-01-04	Poul Pedersen	72	Male	Thisted Amt	Villerslev Sogn
1857-01-04	Oline Christine Christensen	1	Female	Thisted Amt	Kollerup Sogn (Thisted Amt)
1857-01-04	Maren Cathrine Nielsen	2	Female	Thisted Amt	Kollerup Sogn (Thisted Amt)

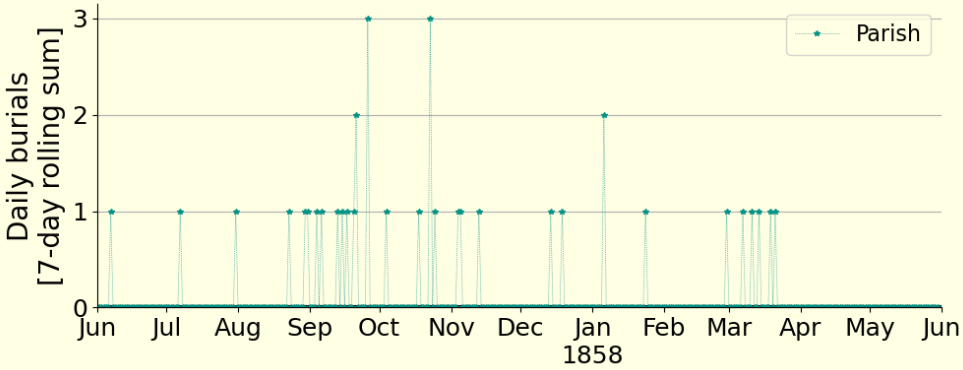
Data cleaning and managing

Date of burial	Name	Age	Gender	Amt	Sogn
1857-01-02	Ane Kirstine Christensen	2	Female	Thisted Amt	Jannerup Sogn (Thisted Amt)
1857-01-02	Birthe Marie Christensen.	0	Female	Thisted Amt	Hundborg Sogn
1857-01-02	Ane Marie Mortensdatter	81	Female	Thisted Amt	Vejerslev Sogn (Mors)
1857-01-02	Gjertrud Jensdatter	82	Female	Thisted Amt	Thisted Sogn
1857-01-02	Karen Christensdatter Willsbøll	52	Female	Thisted Amt	Vester Vandet Sogn
1857-01-02	Karen Marie Jensen	21	Female	Thisted Amt	Sennels Sogn
1857-01-02	Ane Christensdatter Krogh	76	Female	Thisted Amt	Vester Vandet Sogn
1857-01-03	Anders Hansen Tüylfang	79	Male	Thisted Amt	Gøttrup Sogn
1857-01-03	Marcus Christensen	6	Male	Thisted Amt	Hunstrup Sogn
1857-01-04	Maren Jensen	0	Female	Thisted Amt	Hillerslev Sogn (Thisted Amt)
1857-01-04	Thomas Jensen	0	Male	Thisted Amt	Flade Sogn (Thisted Amt)
1857-01-04	Niels Madsen Thjstrup	74	Male	Thisted Amt	Skjoldborg Sogn
1857-01-04	Poul Pedersen	72	Male	Thisted Amt	Villerslev Sogn
1857-01-04	Oline Christine Christensen	1	Female	Thisted Amt	Kollerup Sogn (Thisted Amt)
1857-01-04	Maren Cathrine Nielsen	2	Female	Thisted Amt	Kollerup Sogn (Thisted Amt)

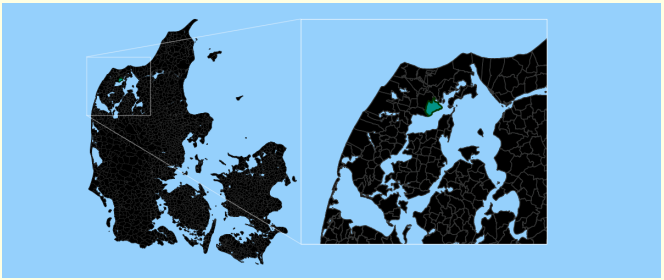
- Geographical resolution:
  - Individual parishes
  - Shire (groups of 5 to 10 parishes)
  - Counties (groups of 5 to 10 shires)
- Temporal resolution:
  - Daily
  - Weekly
  - Monthly
  - Yearly

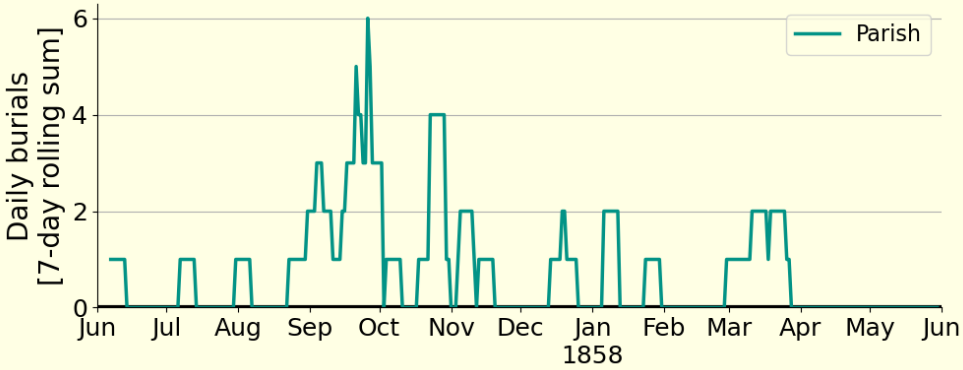
Data cleaning and managing



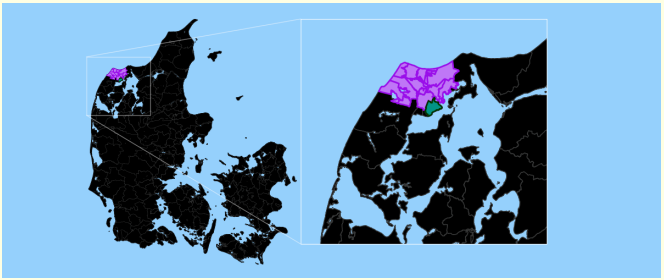


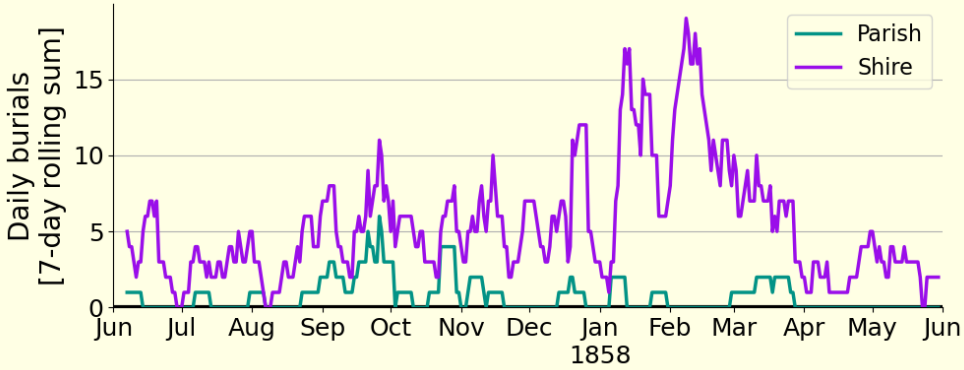
Data cleaning and managing



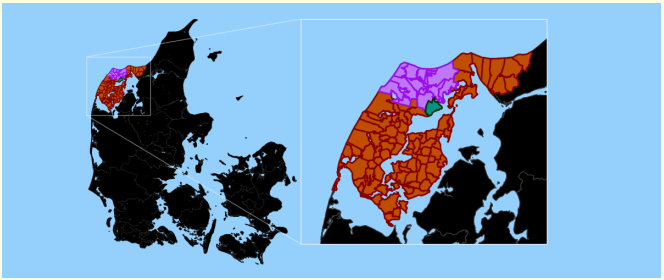


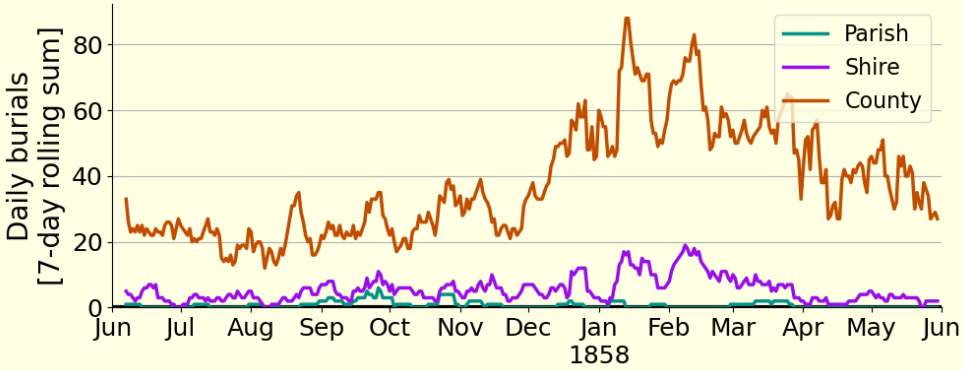
Data cleaning and managing





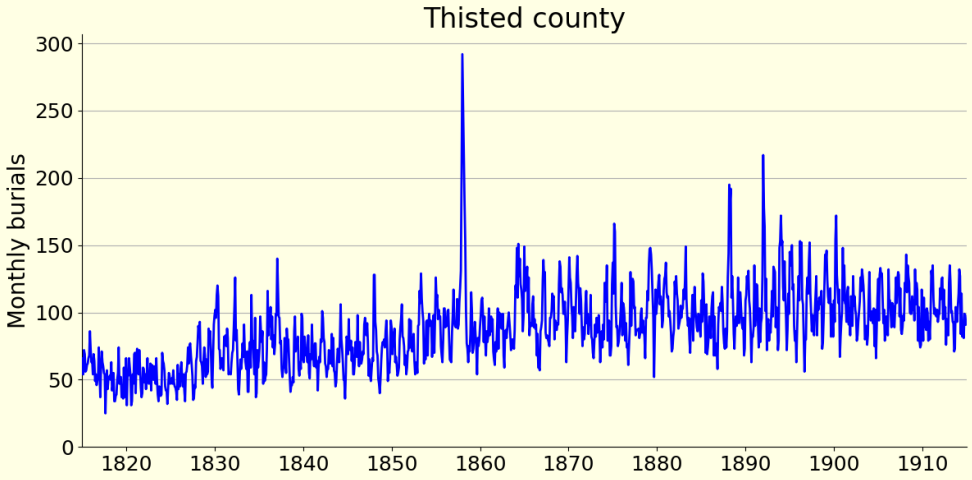
Data cleaning and managing



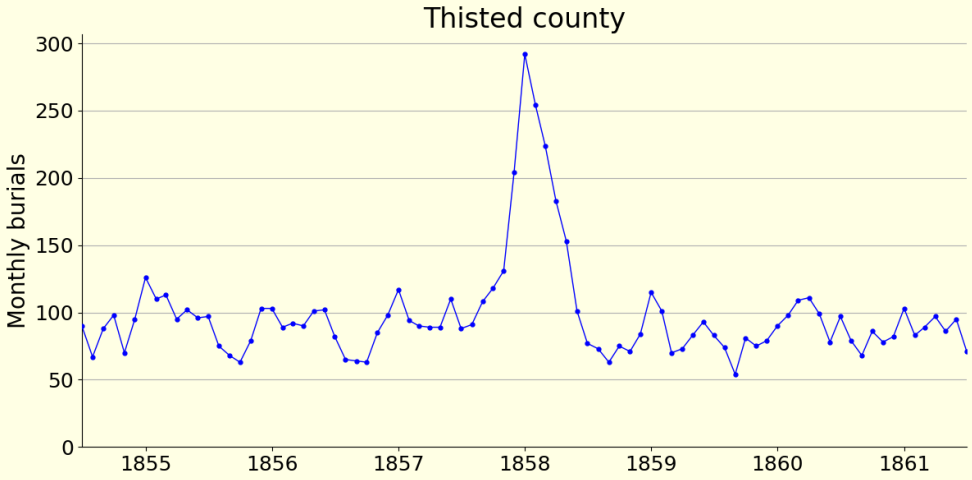


PANDEMIX CENTER

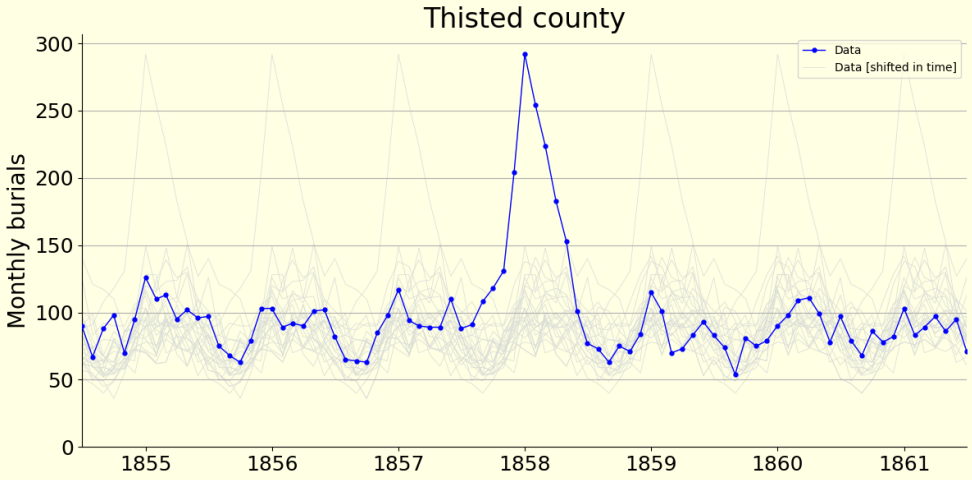
Calculating the mortality baseline



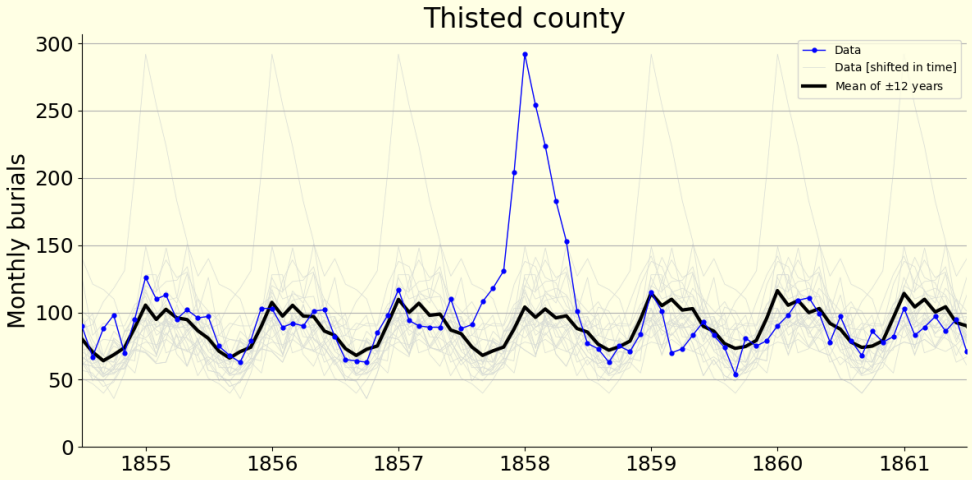
Calculating the mortality baseline



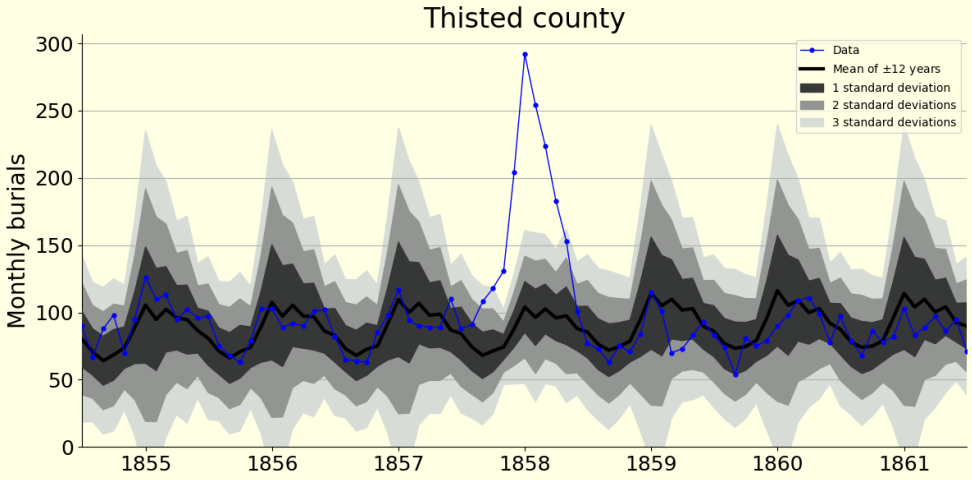
Calculating the mortality baseline



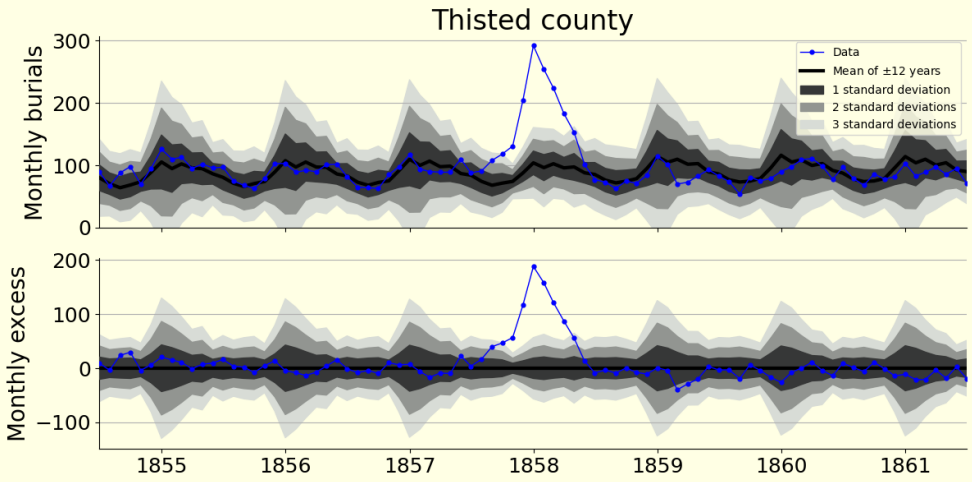
Calculating the mortality baseline



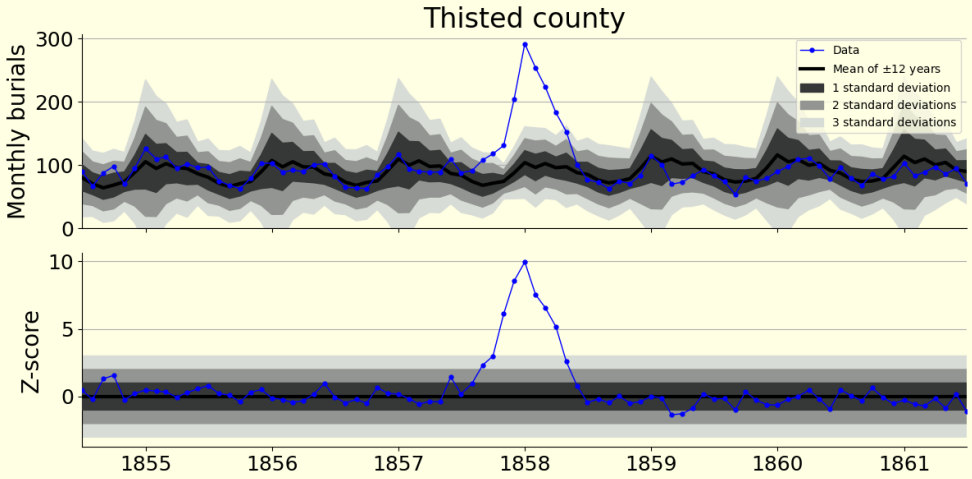
Calculating the mortality baseline



Calculating the mortality baseline



Calculating the mortality baseline

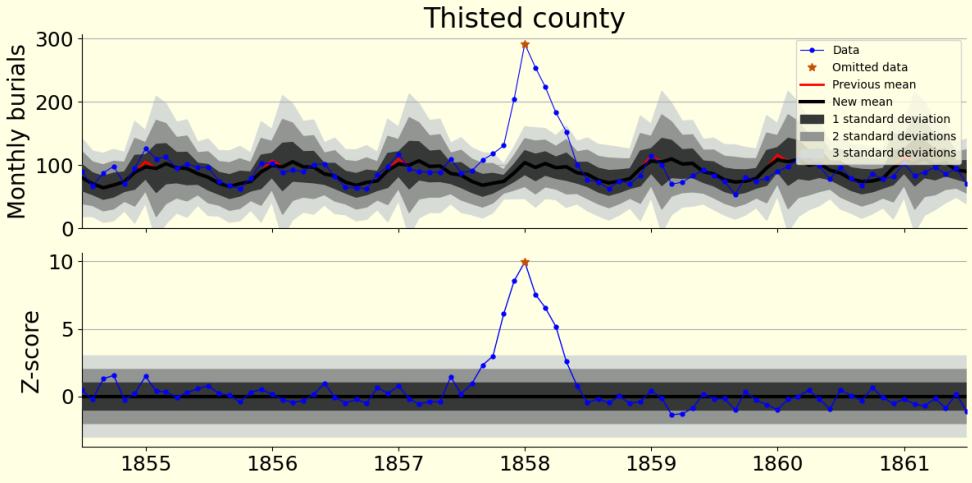


Hinde[1] uses the definition of months with a Z-score above two and three as "mortality crisis" and "severe mortality crisis", respectively

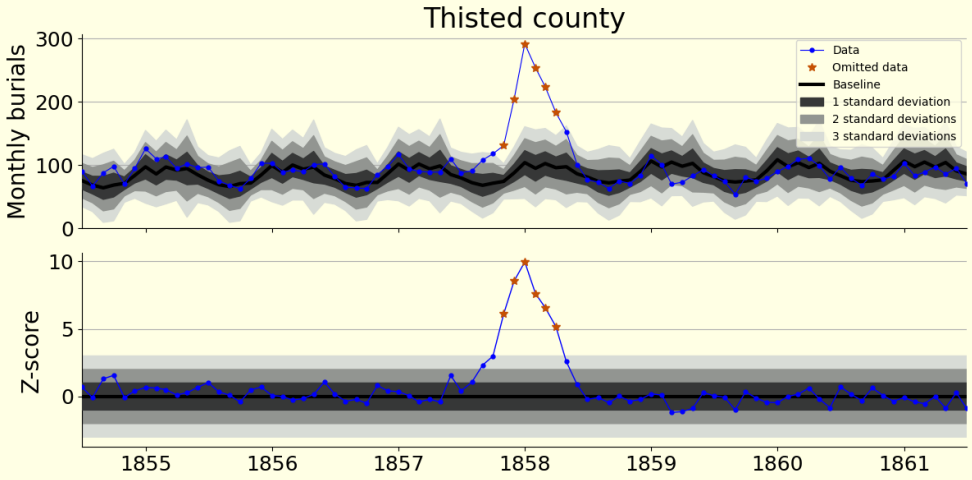
[1] A. Hinde (2010) *"A review of methods for identifying mortality 'crises' using parish record data"* - Local Population Studies

PANDEMIX CENTER

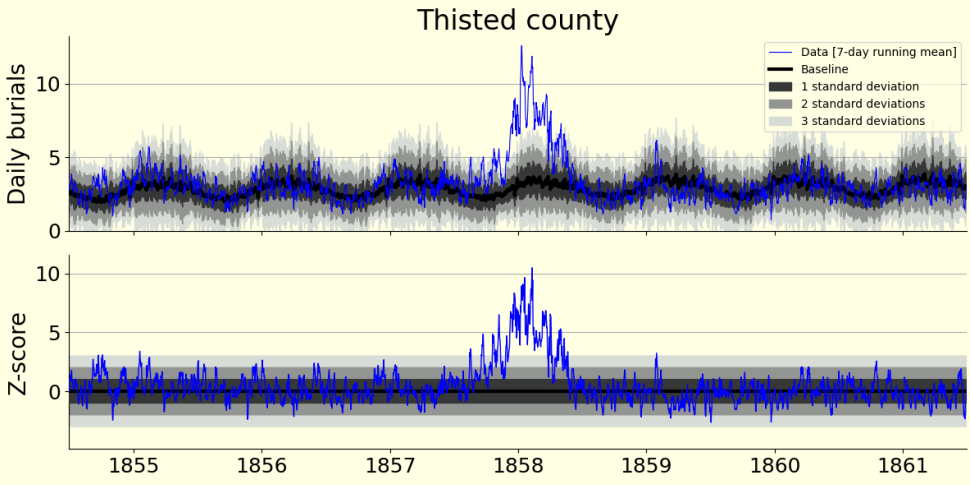
Calculating the mortality baseline



Calculating the mortality baseline

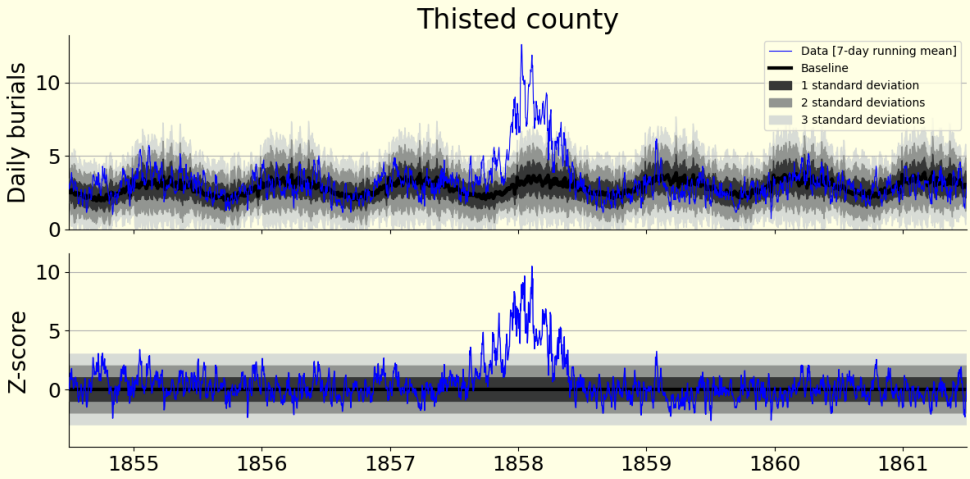


Calculating the mortality baseline



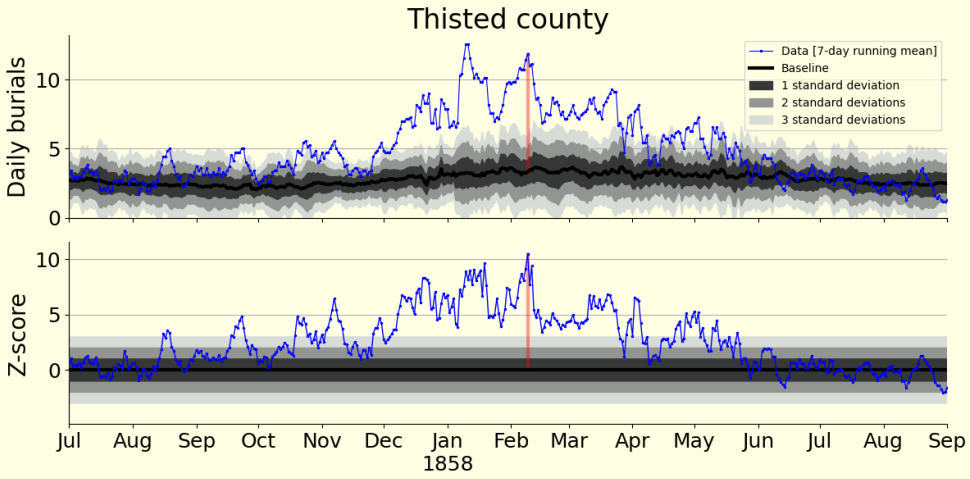
PANDEMIX CENTER

Identification of “mortality crises”



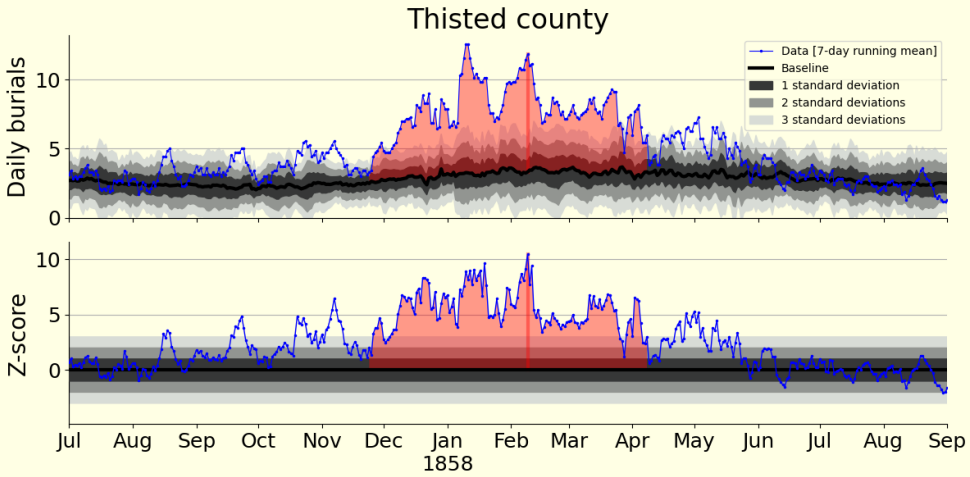
We wish to identify periods continuous excess mortality.

Identification of “mortality crises”



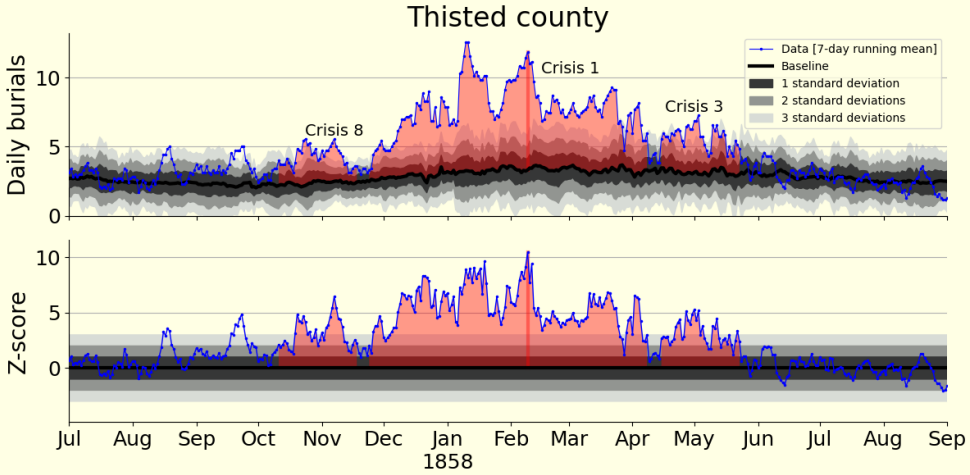
Starting from the date with most burials...

Identification of “mortality crises”



... we group all days with Z-score above three, until the Z-score drops below two for *four* days or more.

Identification of “mortality crises”



All crises with at least *seven* days above the Z-score threshold of three are considered “mortality crises”.

Brief return to the "signature features"

Using this methodology, we identify 320 mortality crises in Denmark between 1820 and 1910.

Brief return to the "signature features"

Using this methodology, we identify 320 mortality crises in Denmark between 1820 and 1910.
For each crisis, we can determine:

- Significant excess mortality.
  Enough to cause excess burials on a county level
  (Population-sizes $\approx 100,000$).

Brief return to the "signature features"

Using this methodology, we identify 320 mortality crises in Denmark between 1820 and 1910.
For each crisis, we can determine:

- Significant excess mortality.
  Enough to cause excess burials on a county level
  (Population-sizes $\approx 100,000$).
- Timing and seasonality.
  e.g. "peaking in winter" or "late summer"

Brief return to the "signature features"

Using this methodology, we identify 320 mortality crises in Denmark between 1820 and 1910.
For each crisis, we can determine:

- Significant excess mortality.
  Enough to cause excess burials on a county level
  (Population-sizes $\approx 100,000$).
- Timing and seasonality.
  e.g. "peaking in winter" or "late summer"
- Duration

  e.g. "lasting two months"

Brief return to the "signature features"

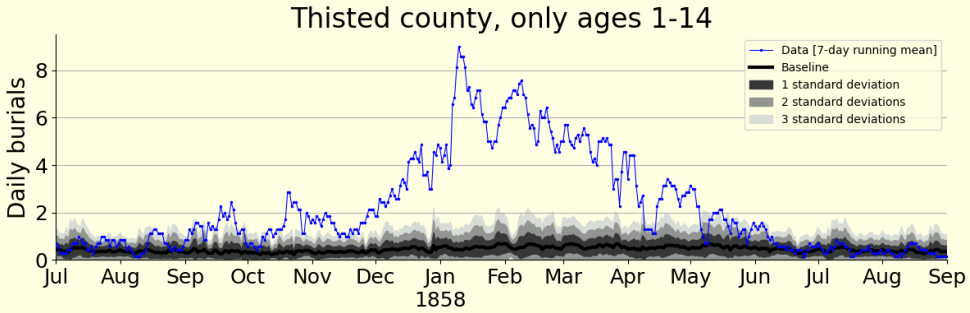
Using this methodology, we identify 320 mortality crises in Denmark between 1820 and 1910.
For each crisis, we can determine:

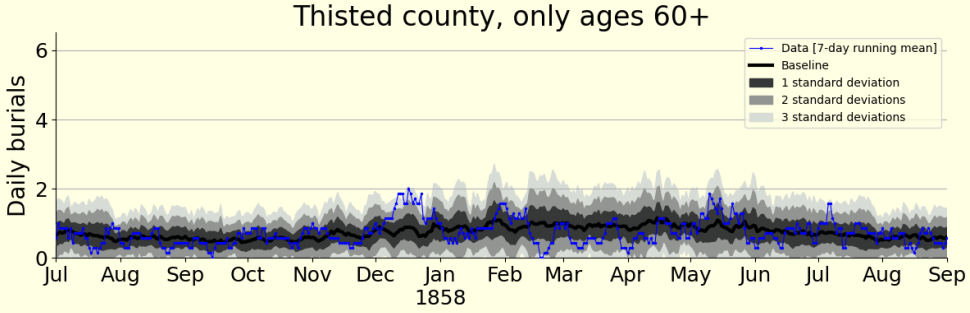
- Significant excess mortality.
  Enough to cause excess burials on a county level
  (Population-sizes $\approx 100,000$).
- Timing and seasonality.
  e.g. "peaking in winter" or "late summer"
- Duration

  e.g. "lasting two months"

But we also have data on age.

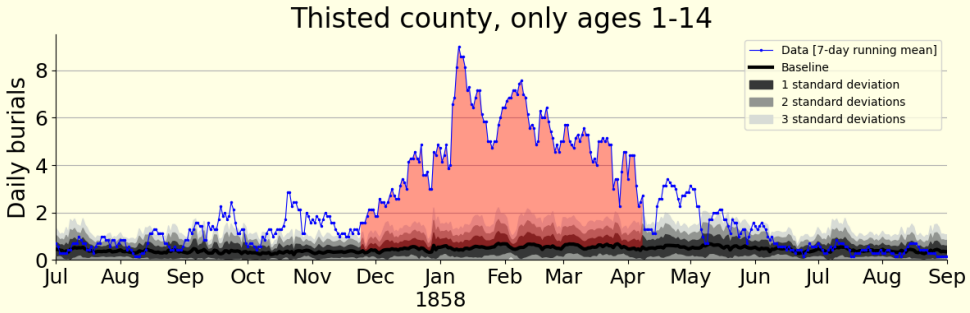
Analyzing age-patterns

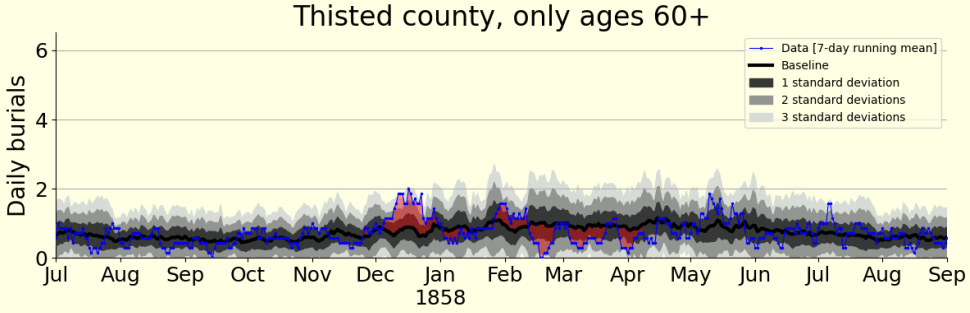




PANDEMIX CENTER

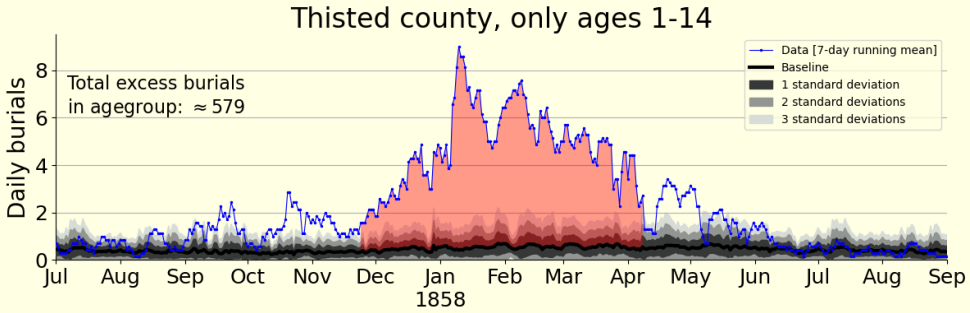
Analyzing age-patterns

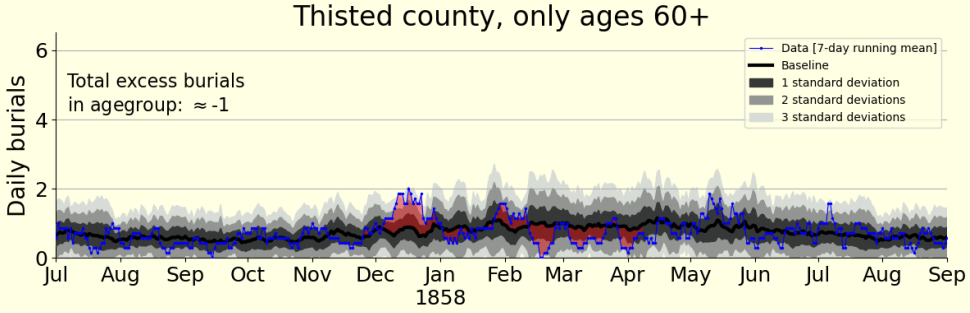




PANDEMIX CENTER

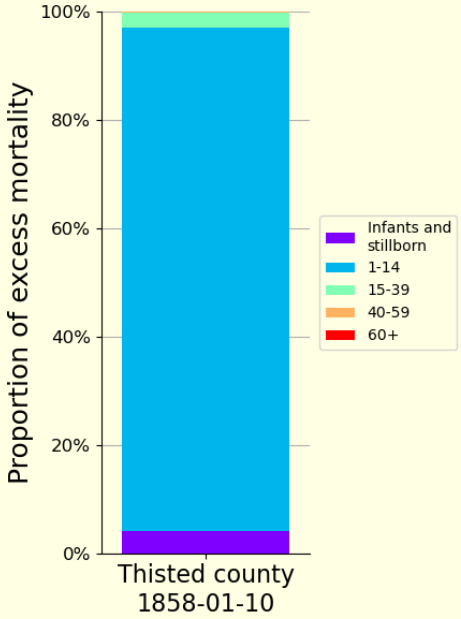
Analyzing age-patterns





PANDEMIX CENTER

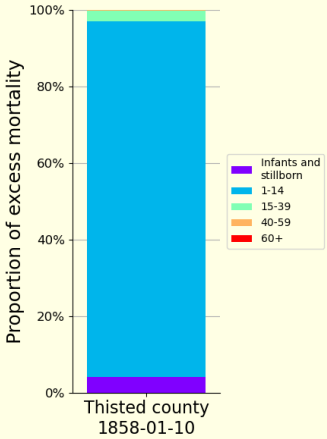
Analyzing age-patterns



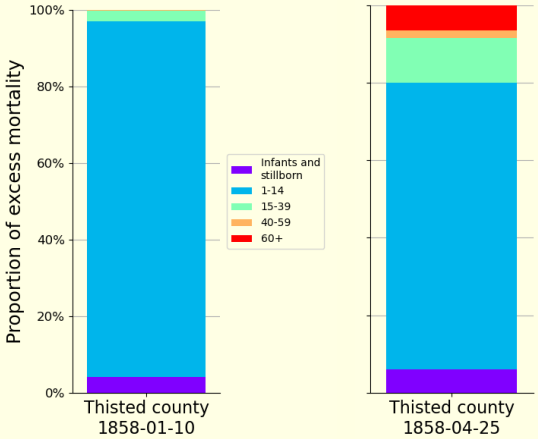
Age-specific excess mortality in this period:

- Age group "60+": <1%
- Age group "40-59": <1%
- Age group "15-39": 3%
- Age group "1-14": 93%
- Age group "Below 1 year": 4%

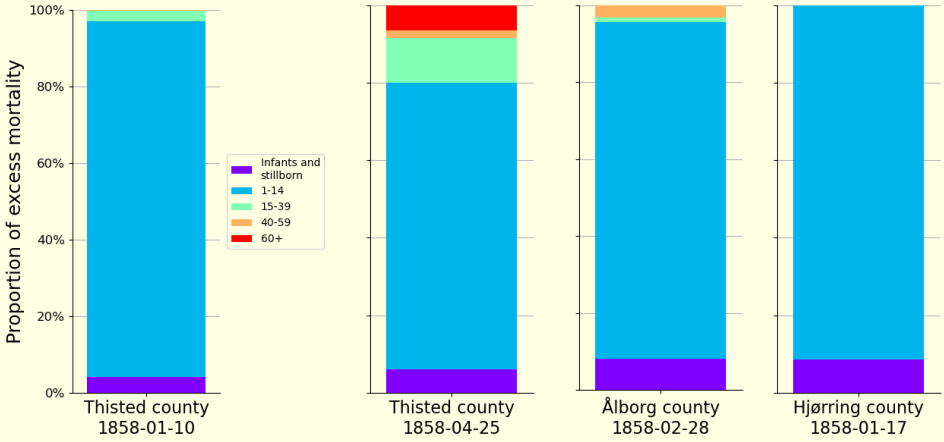
Mortality crises with comparable age patterns



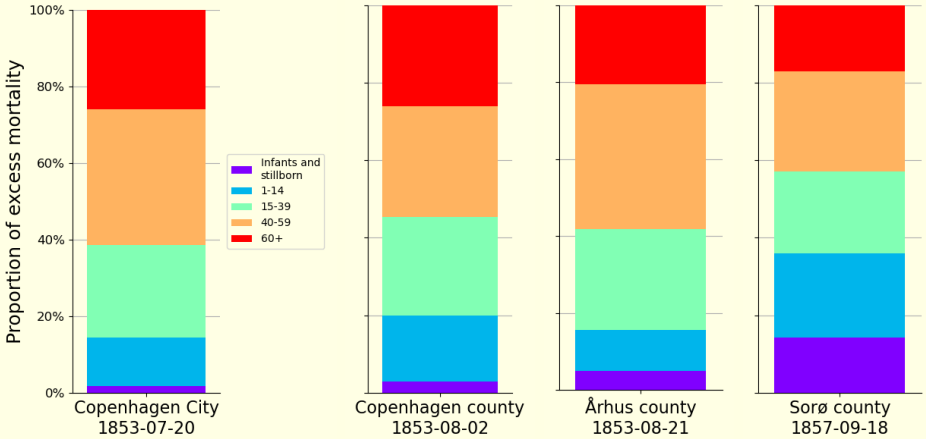
Mortality crises with comparable age patterns



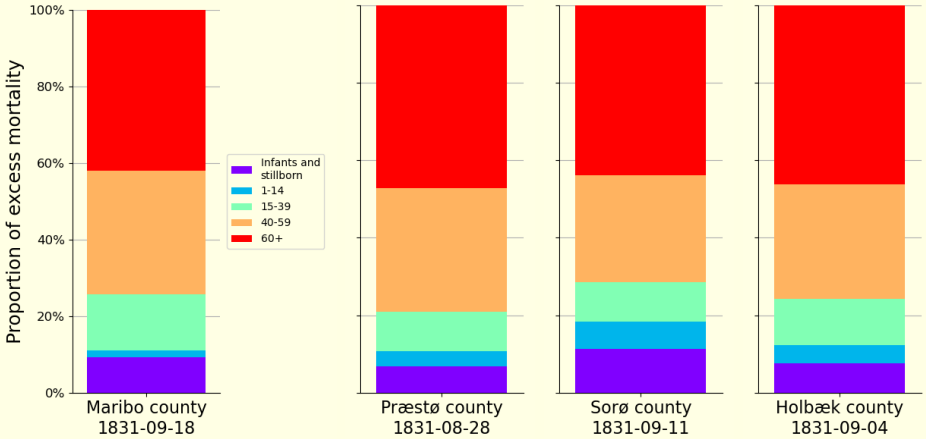
Mortality crises with comparable age patterns



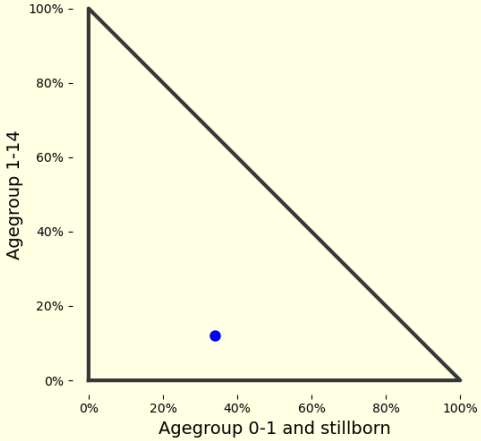
Mortality crises with comparable age patterns

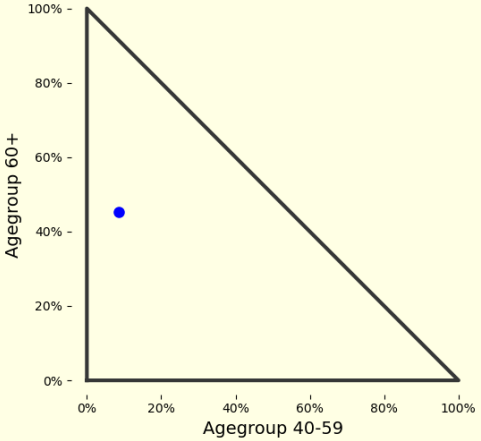


Mortality crises with comparable age patterns



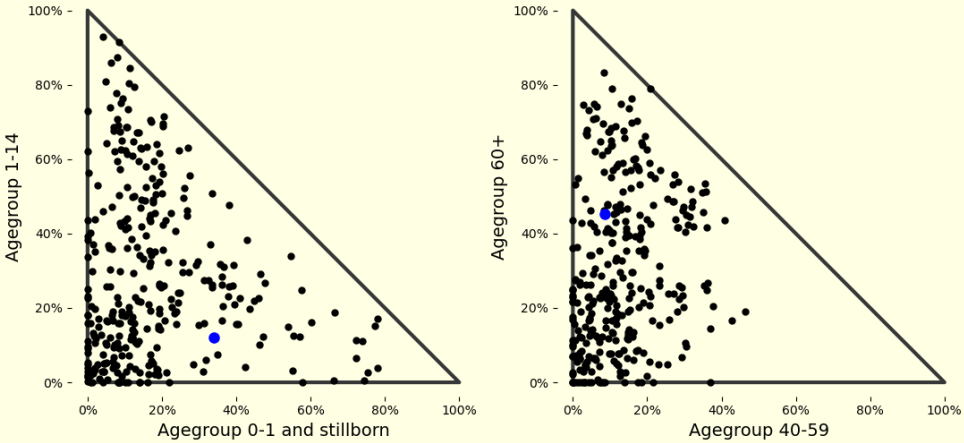
Mortality crises with comparable age patterns





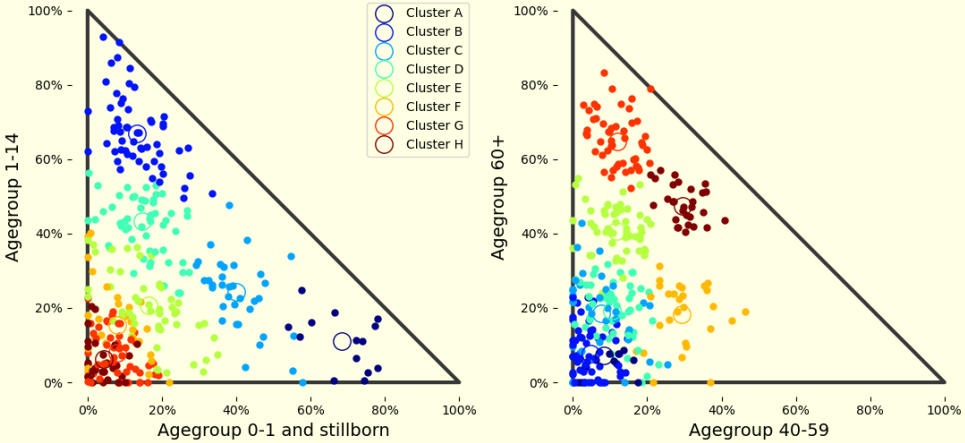
(Age group "15-39" not shown here)

Mortality crises with comparable age patterns



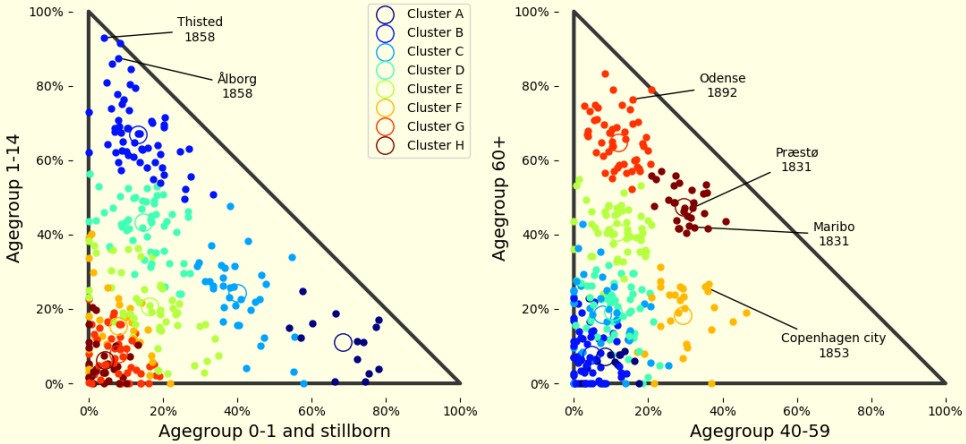
Adding the other 319 mortality crises identified.

Mortality crises with comparable age patterns



Gaussian mixture modelling on full five-dimensional data.

Mortality crises with comparable age patterns



Gaussian mixture modelling on full five-dimensional data.

All epidemics and pandemics in 19th century Denmark

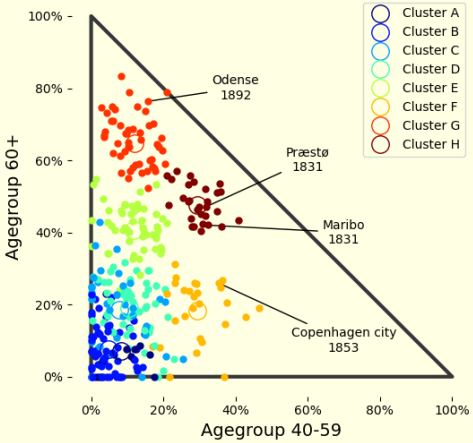
Going through all major mortality crises, and crossreferencing with historical records of epidemics:

Disease	*Timing*	*Total excess*	*Age structure*
Cholera	Late summer, 1853 and 1857	4588	Adults Cluster "F"
Scarlet fever	Winter 1857/1858	2323	Children (1-15) Cluster "B"
"Harvest" epidemic[1]	Late summer, 1825-1831	11539	Adults Cluster "H" and "F"
Pandemic influenza	1892 and 1900	9532	Elderly Cluster "G"

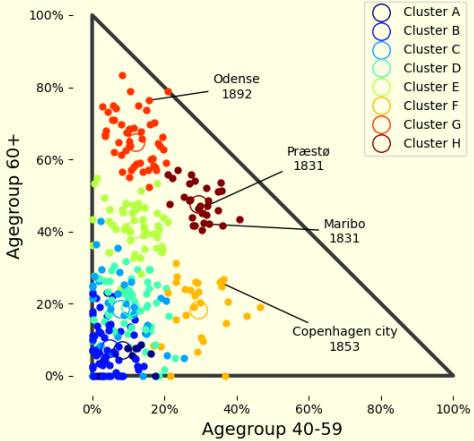
And other epidemics as well as mortality crises unrelated to disease, e.g. war.

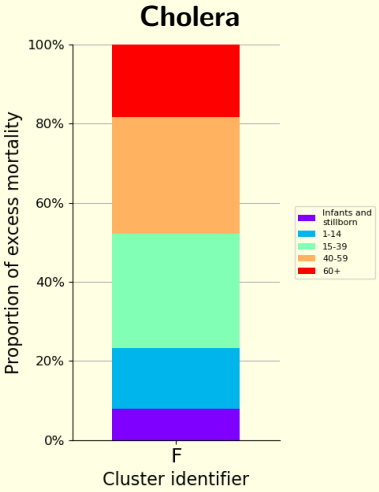
[1] The cause of this mortality crisis was probably a subsistence crisis as well as a range of diseases. Discussed in detail in Ingholt (2022) *Scandinavian Journal of History*

The signature features of certain diseases

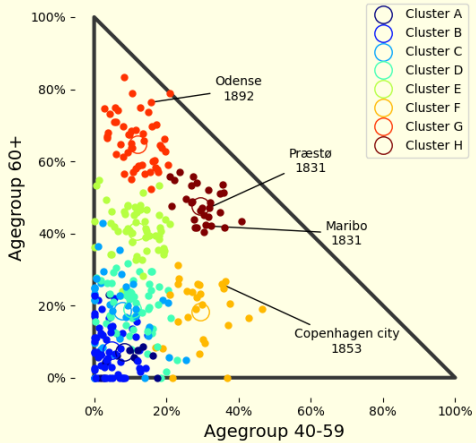


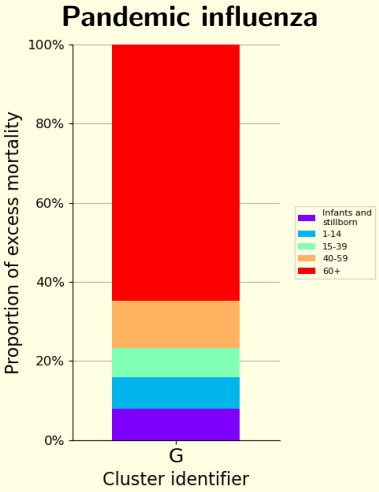
The signature features of certain diseases



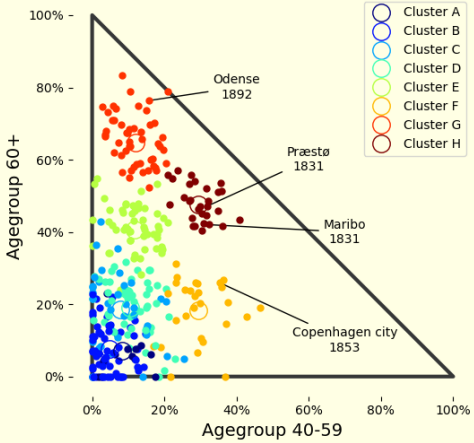


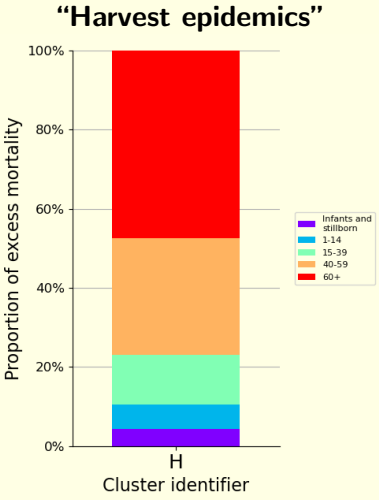
The signature features of certain diseases



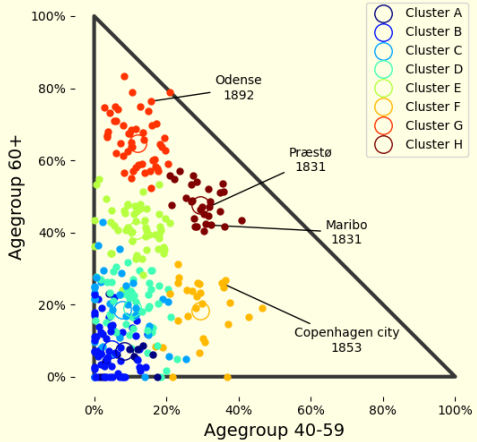


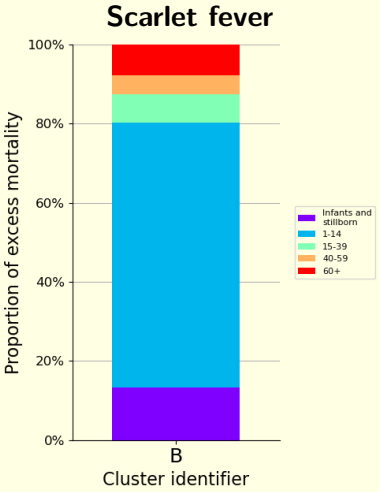
The signature features of certain diseases





The signature features of certain diseases





Summary

- We determine mortality baselines on county-level and estimate excess mortality.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:
  - Timing/Seasonality.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:
  - Timing/Seasonality.
  - Duration.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:
  - Timing/Seasonality.
  - Duration.
  - Geography.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:
  - Timing/Seasonality.
  - Duration.
  - Geography.
  - Age-pattern.

Summary

- We determine mortality baselines on county-level and estimate excess mortality.
- We identify 320 major mortality crises in 19th century Denmark.
- For each crisis, we determine signature features:
  - Timing/Seasonality.
  - Duration.
  - Geography.
  - Age-pattern.
- By comparing these features and validating with historical sources, we are able to determine groups of mortality crises with the same etiology, and estimate the total number of excess deaths during specific epidemics.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.
- Similar methods could be applied to modern data.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.
- Similar methods could be applied to modern data.
  - Mortality baseline calculations.
    Available online soon, as both Python and R package.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.
- Similar methods could be applied to modern data.
  - Mortality baseline calculations.
    Available online soon, as both Python and R package.
  - Quantitative comparison of age-patterns in modern all-cause mortality data.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.
- Similar methods could be applied to modern data.
  - Mortality baseline calculations.
    Available online soon, as both Python and R package.
  - Quantitative comparison of age-patterns in modern all-cause mortality data.
- Understanding changes through history of the mortality of specific diseases may help us understand the disease today.

Final comments

- Despite demographic differences between 19th century Denmark, the identified age patterns may still be relevant in modern data.
- Similar methods could be applied to modern data.
  - Mortality baseline calculations.
    Available online soon, as both Python and R package.
  - Quantitative comparison of age-patterns in modern all-cause mortality data.
- Understanding changes through history of the mortality of specific diseases may help us understand the disease today.
- As more historical data becomes transcribed, e.g. thanks to improved OCR, similar studies of other countries will become possible.

Thank you for your attention.

Feel free to email me with questions or comments

Website:

rasmuspedersen.com

Email:

rakrpe@ruc.dk

"*Identifying Signature Features of Epidemic Diseases in 19th Century All-cause Mortality Data*"
Pedersen RK, Ingholt MM, van Wijhe M, Andreasen V & Simonsen L

Roskilde University

PANDEMIX CENTER

Danmarks Grundforskningsfond
Danish National Research Foundation

PANDEMIX CENTER