Topological Data Analysis for Health Data
Introduction and Opportunities

Pawel Dlotko and Simon Rudkin
Rationale for Topological Data Analysis

• We learn a lot from seeing data – Limit on how many dimensions we can see at any one time

• Data conveys message through it’s shape – How do you describe the shape of data?

• Topology is the study of shape – seeing what is actually there

The review also found:

- Lower numbers of deaths at the start of 2020, potentially due to a mild winter, may have increased the age and vulnerability of the population

- The highest mortality rates were in Cardiff and Vale health board areas, while the lowest rates of infection were in Hywel Dda health board and Powys

- Older people, black, Asian and ethnic minority people, and those from deprived areas had the highest death rates

- Men have consistently higher mortality rates across all ethnic backgrounds

The report warns that, in order to save lives in the event of a future wave, early recognition of resurgence of infection is “critical”.

Example from www.bbc.co.uk taken on 20th July 2020
https://www.bbc.co.uk/news/uk-wales-53449371
• Common approach of viewing trends, correlations and associations
• Case 1 age and incidences of Covid-19 requiring hospital
• Simple headline about older people having more cases
• Younger cases not unheard of but do not fit relation
• Cause of differential might be insignificant overall (e.g. use of playground swings)
• Common approach of viewing trends, correlations and associations

• Case 2 age, Income and incidences of Covid-19

• Colour by number of cases requiring treatment – Red (low) to Purple (high)

• See interesting low income young group
Two variables commonly expressed as cross-tabulation requiring creation of categories. May still use colour to see patterns. Correlation valuable where we cannot prove causation – familiarity with notion of shape. What if correlation changes as axes change?
Topological Data Analysis Ball Mapper

• Creates a visualization of data which is topologically faithful
• Abstract representation of multiple dimensions – seeing more variables than we could before
• Understand data as if it were a scatter plot coloured by outcome
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
- Cover data in “balls” of fixed radius
- Connectivity where points in overlap

![Diagram of Ball Mapper](image-url)
• Creates a visualization of data which is topologically faithful
• Abstract representation of multiple dimensions – seeing more variables than we could before
• Understand data as if it were a scatter plot coloured by outcome
• Cover data in “balls” of fixed radius
• Connectivity where points in overlap
• Size according to points in ball
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
- Cover data in “balls” of fixed radius
- Connectivity where points in overlap
- Size according to points in ball
- Colour according to outcomes
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
- Cover data in “balls” of fixed radius
- Connectivity where points in overlap
- Size according to points in ball
- Colour according to outcomes
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
- Cover data in “balls” of fixed radius
- Connectivity where points in overlap
- Size according to points in ball
- Colour according to outcomes
Topological Data Analysis Ball Mapper

- Creates a visualization of data which is topologically faithful
- Abstract representation of multiple dimensions – seeing more variables than we could before
- Understand data as if it were a scatter plot coloured by outcome
- Cover data in “balls” of fixed radius
- Connectivity where points in overlap
- Size according to points in ball
- Colour according to outcomes
- Makes sense where we cannot see the shape
Example of Topological Data Analysis: Covid 19

- 6 Axes - Hourly wage, hours worked, GDP, GVA, population density and age
- Coloured by proportion of cases within each ball
- Perform dynamic analysis to show evolution of Covid-19
- Concentration of cases in economically similar areas
- Spread to commuter areas very limited

Example from Dlotko and Rudkin (2020) Visualising the Evolution of English Covid-19 Cases with Topological Data Analysis Ball Mapper
Example of Topological Data Analysis: Brexit

• 16 Axes based on 2011 census characteristics
• Coloured by average Leave vote
• Concentration of Leave versus dispersal of Remain

Example of Topological Data Analysis: Brexit

- 16 Axes based on 2011 census characteristics
- Coloured by average Leave vote
- Concentration of Leave versus dispersal of Remain
- Residuals from OLS model – proportion bigger than 4 percentage points
- Different models for different parts of the space justified – partition and local model?

Example of Topological Data Analysis: Blood Biomarkers

- Example from Brazilian data shared on data science platforms
- 20 different blood characteristics tested on admission
- Coloured by probability of requiring ICU
- Strong correlation shown by narrow band like shape – exactly like scatter plot
- Admission probabilities are high in two parts of space

Data: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/discussion/139347
Topological Data Analysis and Health in Wales

**Axes**
- Demographics
- Medical records
- Local area characteristics

**Outcomes**
- Value of healthcare
- Clinical outcomes
- Infection rates

**Questions**
- Covid-19 impact?
- Care accessibility?
- Targets for policy intervention?

Data must have sufficiently many different ordinal values

Outcome usually average within ball (e.g. average infection rate)

Priorities from the field
Summary of Topological Data Analysis

- View data from multiple dimensions – Ball Mapper is a tool to answer many questions
- Recognises that the effect of variables changes across the space of other variables
- Capture dynamics on static maps – an early warning system
- All data contributes to shape – don’t discard information

The review also found:

- Lower numbers of deaths at the start of 2020, potentially due to a mild winter, may have increased the age and vulnerability of the population
- The highest mortality rates were in Cardiff and Vale health board areas, while the lowest rates of infection were in Hywel Dda health board and Powys
- Older people, black, Asian and ethnic minority people, and those from deprived areas had the highest death rates
- Men have consistently higher mortality rates across all ethnic backgrounds

The report warns that, in order to save lives in the event of a future wave, early recognition of resurgence of infection is "critical".

Example from [www.bbc.co.uk](http://www.bbc.co.uk) taken on 20th July 2020
[https://www.bbc.co.uk/news/uk-wales-53449371](https://www.bbc.co.uk/news/uk-wales-53449371)