Sequence analysis as a graphical tool for investigating call record data

Olga Maslovskaya, Gabriele Durrant, Peter W.F. Smith

Introduction

- Call record data frequently collected for interviewer-mediated surveys (face-to-face or telephone surveys)

- Surveys collect information from every call of the interviewer to the household

- Often final response outcome analysed

- Here:

    – process leading to response/nonresponse

    – contact sequence as a whole of interest

- Question: how best to analyse and summarise these call attempt histories?

What is a Sequence of Call Attempts?

- Call sequence - string of outcomes of all call attempts

- Example:

House-hold ID	Call 1	Call 2	Call 3	Call 4	Call 5
101	non-contact	non-contact	appointment	interview	-
102	non-contact	non-contact	refusal	-	-
103	non-contact	non-contact	appointment	non-contact	refusal
104	contact	contact	non-contact	non-contact	appointment

Aims and Objectives

- To explore the use of sequence analysis to better understand the complex patterns of call attempts to households

  – For cross-sectional data: first wave of a longitudinal study

  – Across several waves of a longitudinal study: 3 waves

- Sequence analysis per interviewer (e.g. identification of 'unusual' calling behaviour or assessment of interviewer performance)

Data: UK Understanding Society Survey

- Large-scale longitudinal survey (approximately 40,000 responding households in the UK)

- Wave 1: January 2009 – March 2011: All face-to-face interviews

- Wave 2: January 2010 – January 2012: Most face-to-face interviews, some telephone interviews (HH from BHPS)

- Wave 3: January 2011- July 2013: Most face-to-face interviews, some telephone interviews (HH from BHPS) and telephone "mop-up" - to contact adults who could not be contacted by face-to-face interviewers during the main fieldwork period

Software Used

R software package

- Use ***TraMineR*** library to perform Sequence Analysis

- Use ***WeightedCluster*** library to conduct Optimal Matching and then Cluster Analysis

- Use ***Vegan*** library to conduct MDS

Call Record Data Summaries

	Wave 1	Wave 2	Wave 3
Sample size	47,899	25,367	23,005
Distinct sequences	11,143	4,701	4,029
Number of calls	255,778	104,566	91,526
Number of interviewers	741	874	733
Maximum number of calls	30	30	36
Mean	5.34	4.12	3.98
Median	4	3	3
Number of states	6	6	6

States of sequences

1. Non-contact

2. Contact

3. Appointment made

4. Interview done

5. Completed process

6. Any other status (includes refusals, ineligible statuses, etc.)
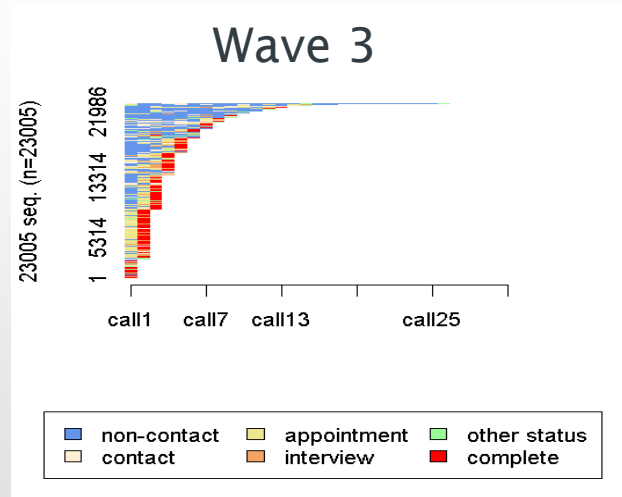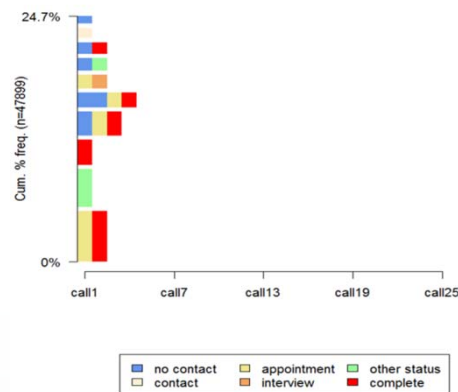
Sequence Index Plots
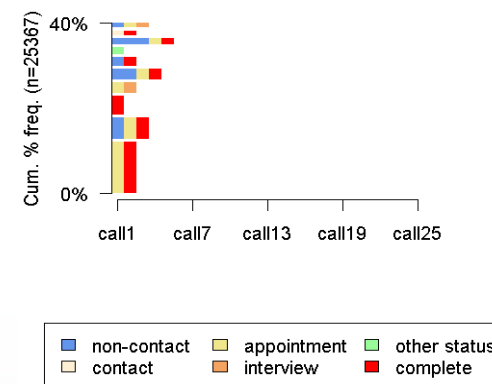

Wave 1
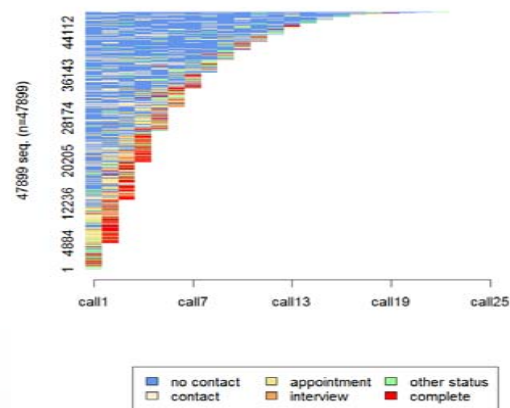
Wave 2
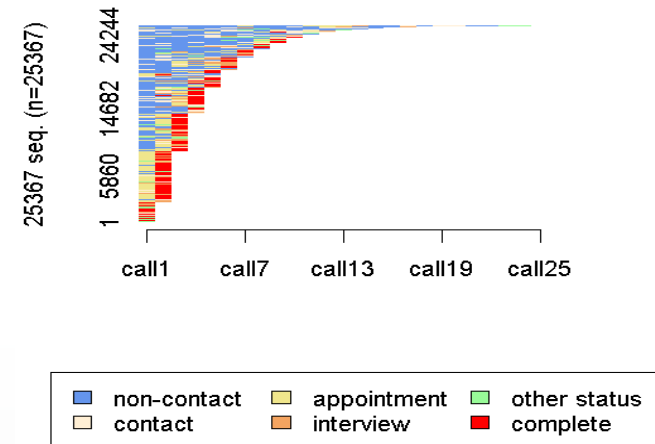
Wave 3

- Patterns are similar

- Waves 2 and 3, more "reds" earlier on in sequences

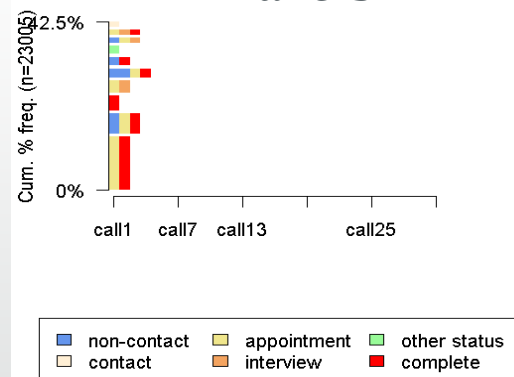# First 10 most frequent sequences

Wave 1

Wave 2

Wave 3

- Similar most frequent sequences with sequence with just "any other status" becoming less frequent over time

Wave 1: Transition Rates

|	To end of sequence	To no contact	To contact	To appointment	To interview	To any other status	To completed
From no contact	0.04	0.65	0.12	0.08	0.01	0.07	0.02
From contact	0.18	0.38	0.20	0.09	0.03	0.09	0.04
From appointment	0.01	0.18	0.06	0.05	0.24	0.05	0.42
From interview	0.53	0.08	0.09	0.04	0.06	0.03	0.17
From any other status	0.56	0.23	0.07	0.02	0.01	0.11	0.01
From completed	0.94	0.01	0.04	0.00	0.00	0.01	0.00

Wave 2: Transition Rates

	To end of sequence	To no contact	To contact	To appointment	To interview	To any other status	To completed
From no contact	0.03	0.57	0.12	0.16	0.02	0.05	0.05
From contact	0.18	0.30	0.19	0.15	0.05	0.05	0.08
From appointment	0.01	0.11	0.04	0.06	0.23	0.02	0.52
From interview	0.57	0.07	0.08	0.04	0.05	0.02	0.18
From any other status	0.54	0.19	0.06	0.06	0.02	0.11	0.03
From completed	0.95	0.01	0.03	0.00	0.00	0.01	0.00

Wave 3: Transition Rates

|	To end of sequence	To no contact	To contact	To appointment	To interview	To any other status	To completed
From no contact	0.04	0.56	0.12	0.17	0.02	0.05	0.05
From contact	0.17	0.30	0.19	0.15	0.06	0.05	0.07
From appointment	0.01	0.10	0.04	0.06	0.25	0.03	0.52
From interview	0.57	0.06	0.06	0.04	0.07	0.02	0.19
From any other status	0.56	0.18	0.06	0.05	0.02	0.10	0.03
From completed	0.99	0.00	0.01	0.00	0.00	0.00	0.00

13

Distance Measure and Clustering

- Aim: identify groups of sequences or summarise sequences

- Construct a distance measure (distance matrix)

- Optimal Matching:

    measures similarity between pairs of sequences by counting the number of operations (substitution, insertion, deletion) required to transform one sequence into another

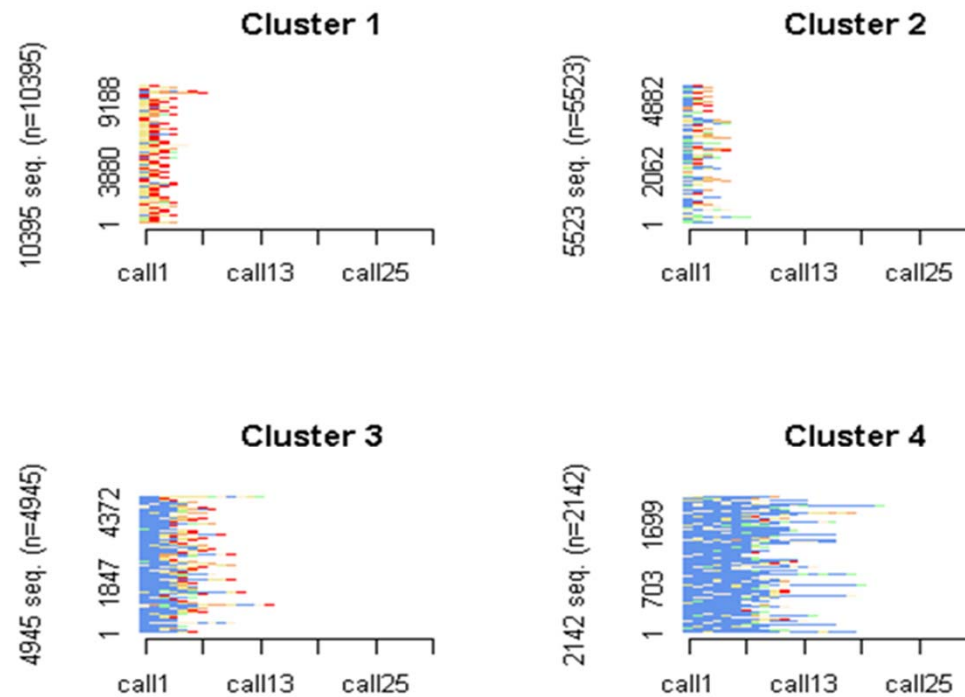- The more operations required the larger the differences between sequences

Wave 1: Sequences by Clusters
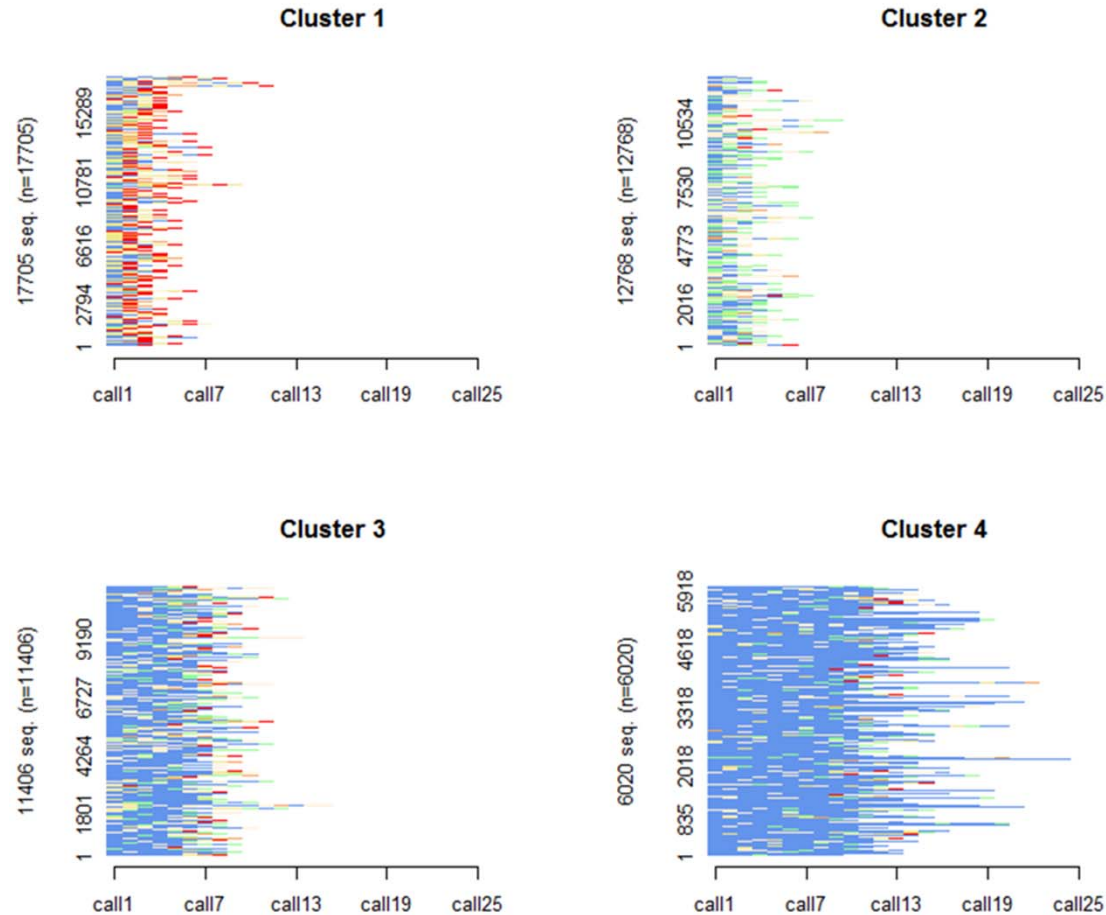
Wave 2: Sequences by Clusters
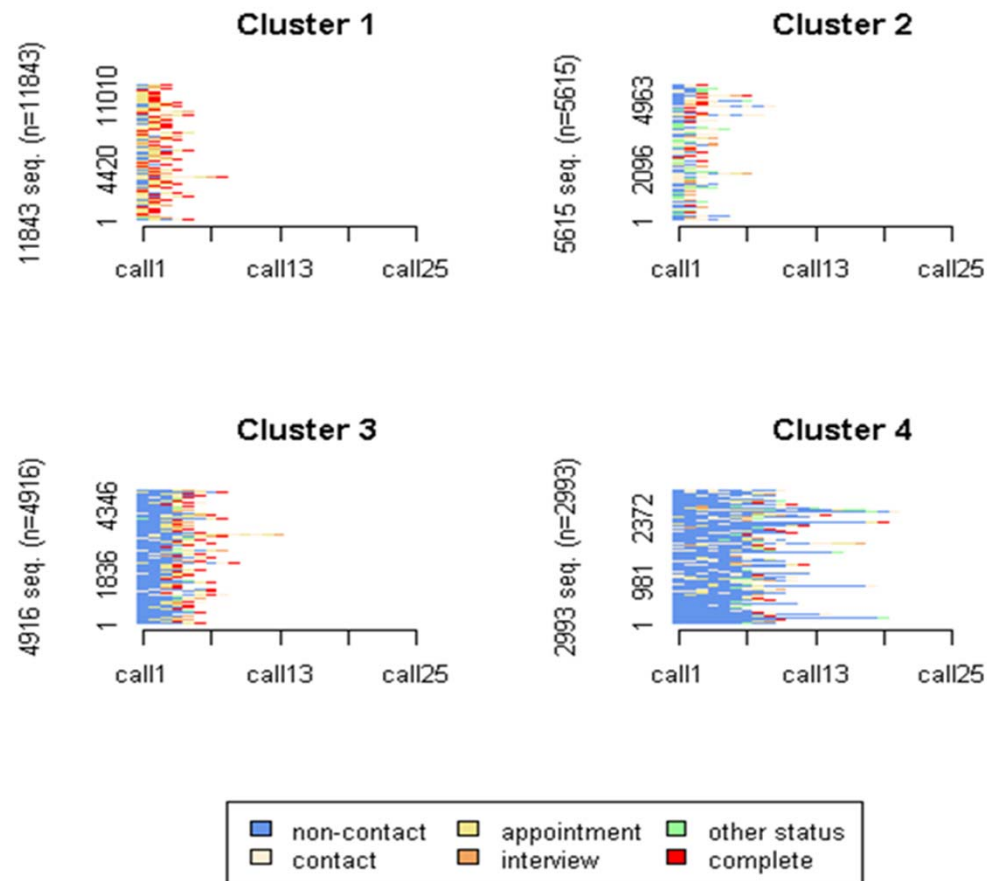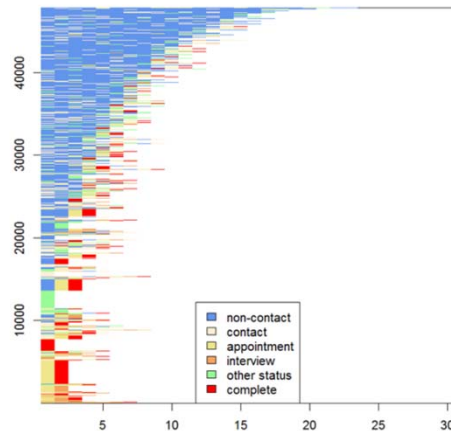
Wave 3: Sequences by Clusters
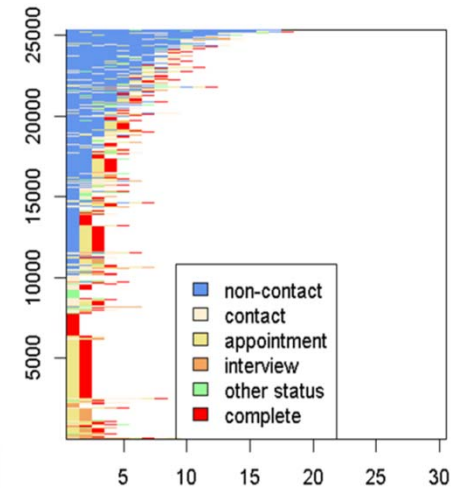
Multidimensional Scaling

- Construct a distance matrix

- Aim:

  - Reveal the structure of the data set by plotting points in one or two dimensions

  - Produce a 2-dimensional map to plot sequences
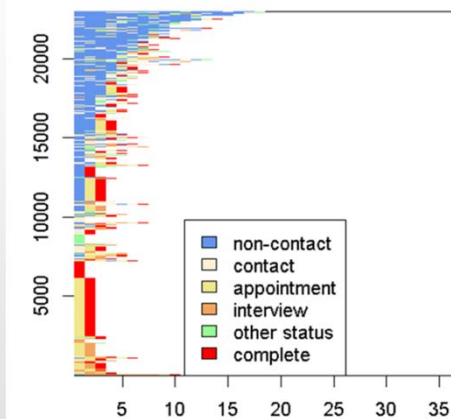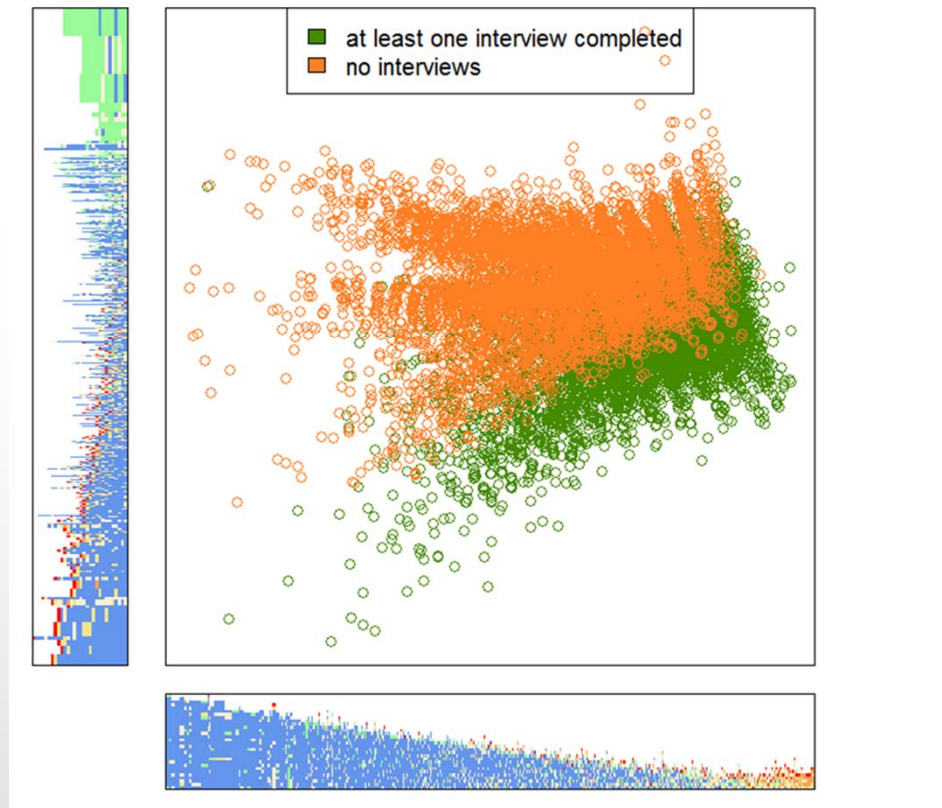
Multidimensional Scaling : First Dimension

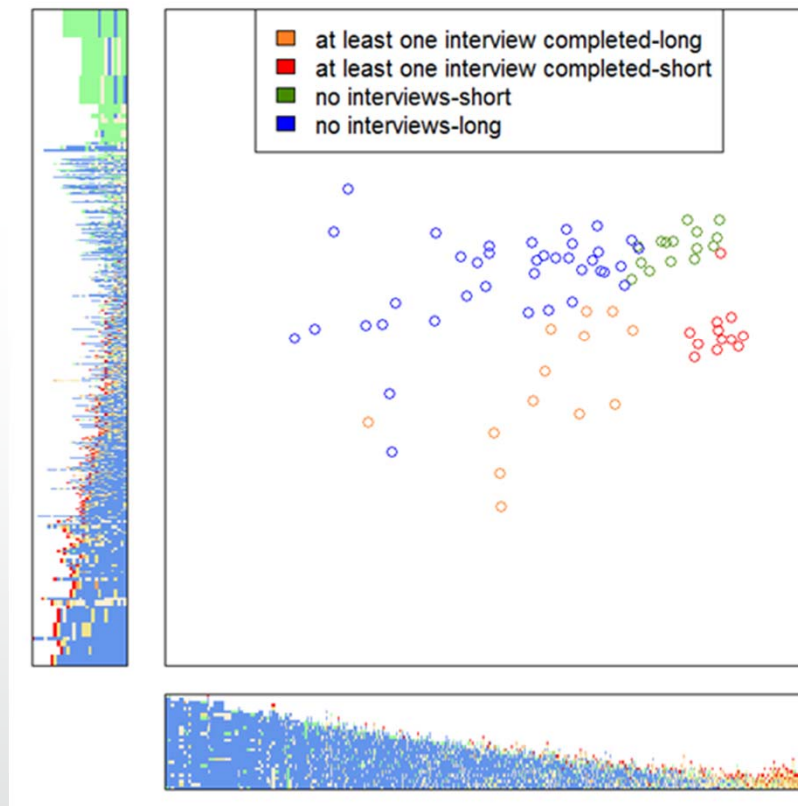- Similar patterns in different waves

- Length and outcome

Multidimensional Scaling: Two Dimensions
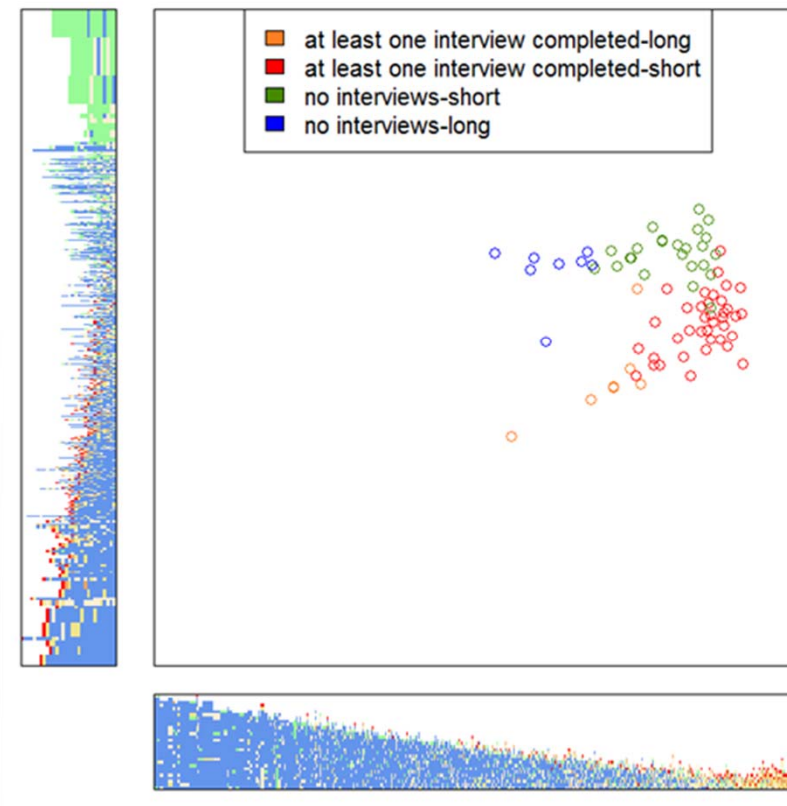
Wave 1

Multidimensional Scaling: per interviewer

Wave 1

Multidimensional Scaling: per interviewer

Wave 1

Future work

- Modelling outcome and length

    - Dependent variables:

        - Length

        - Outcome

        - Outcome * length (4 categories)- short successful, short unsuccessful, long successful, long unsuccessful

- Out-of-sample prediction for the following wave using results from the previous wave

- Comparison of sequence analysis to latent class analysis

Conclusions

- Sequence analysis offers a useful graphical tool and a way of visualising and summarising complex data patterns of complete call attempt histories

- Sequence as a whole, not just summary measures

- Allows the identification of groups with similar sequences using a distance matrix

- Identification of 'unusual ' sequences

- Investigation at interviewer level

- Application in both cross-sectional and longitudinal contexts

- Informs the specification of models to analyse sequences

- Benefits: to inform field process work and survey managers and to inform further modelling work

UNIVERSITY OF Southampton

Southampton Statistical
Sciences Research Institute

Thank you!

om206@soton.ac.uk